UNIVERSITAT DE BARCELONA

Mechanistic insights into substrate-assisted catalysis in glycosidases by means of QM/MM molecular dynamics

Juan Coines Lopez-Nieto

UNIVERSITAT DE BARCELONA

FACULTAT DE QUÍMICA

Departament de Química Inorgànica i Orgànica

Secció de Química Orgànica

Programa de Doctorat en Química Orgànica

Mechanistic insights into substrate-assisted catalysis in glycosidases by means of QM/MM molecular dynamics.

Memòria de recerca presentada per **Joan Coines Lopez-Nieto** per tal d'optar al títol de Doctor per la Universitat de Barcelona.

Barcelona, desembre de 2020

Noms, cognoms i signatura del doctorant:

Joan Coines Lopez-Nieto

Noms, cognoms i signatura de la directora i tutora:

Dra. Carme Rovira Virgili

"Il piacere più nobile è la gioia di comprendere."

Leonardo Da Vinci

"Sometimes science is more art than science."

Rick Sanchez

List of Abbreviations and Symbols

Abbreviation/Symbol	**Full word**
a.u.	Atomic Units
C1	Anomeric Carbon
CAZymes	Carbohydrate-Active enzymes
CP	Car-Parrinello
CVs	Collective Variables
DFT	Density Functional Theory
ESP	Electrostatic potential derived atomic charges
fs	Femtosecond
FES	Free Energy Surface
Gal	Galactose
GH	Glycoside Hydrolase
Glc	Glucose
Glc-ox	Glucose oxazoline
Glc-ox^{+}	Glucose oxazolinium ion
GlcA	Glucuronic acid
GlcNAc	*N*-acetyl-glucosamine
GT	Glycosyltransferase
K	Kelvin Degrees
kcal/mol	kilocalorie per mole
Man	Mannose
MC	Michaelis Complex
MD	Molecular Dynamics
MM	Molecular Mechanics
ns	Nanosecond
O5	Pyranic Oxygen
O_g	Glycosidic Oxygen
O_x	Oxygen from the *N*-acetyl group of GlcNAc
PBE	Perdew Burke Ernzerhof exchange correlation functional
PDB	Protein Data Bank
ps	Picosecond
QM	Quantum Mechanics
QM/MM	Quantum Mechanics / Molecular Mechanics
RMSD	Root Mean Square Deviation
Ry	Rydberg
TS	Transition State
Å	Ångström
Δt	Time Step
μ	Fictitious Electron Mass

Summary

Carbohydrates are ubiquitous molecules on Earth and of major importance. They are involved in key biological processes such as cellular recognition, signalling or post-translational modification of proteins. This has a direct impact on infectious diseases and other pathologies such as cancer. Moreover, carbohydrates have attracted great interest in biotechnology because of their use as biofuels or their beneficial effect in the modification of therapeutic proteins. Consequently, the so-called Carbohydrate-Active Enzymes (CAZymes), the enzymes responsible to form, degrade and modify sugar molecules, have gained attention of the scientific community. Glycoside hydrolases (GHs) are enzymes that catalyse the cleavage of glycosidic bond in carbohydrates and glycoconjugates. The understanding of their mechanism of action is fundamental to design specific inhibitors for these enzymes, as well as engineering them for novel catalytic purposes . In this Thesis, we investigated a specific type of GHs, those that degrade *N*-acetylated carbohydrates. Some of them operate via substrate-assisted mechanisms, in which the substrate actively participates in catalysis. Several questions regarding these mechanisms remained unknown, such as the protonation state of the reaction intermediate, a sugar-oxazoline or an oxazolinium ion, and the role of active site residues in catalysis. Computer simulations based on the hybrid quantum mechanics/molecular mechanics (QM/MM) approach, in combination with enhanced-sampling methods, are powerful tools to investigate catalytic processes in GHs. The main goal of this work is to provide an in-depth comprehension of substrate-assisted reaction mechanisms in GHs. Such complex task requires a multi-scale approach, considering both the size of the system under investigation and the time scale of the chemical reactions taking place in the active site.

The Thesis is organized as follows:

Chapter 1 consist of a general introduction about carbohydrates, GHs, and their catalytic mechanisms, including substrate-assisted catalysis. Open questions and the objectives to tackle them are listed.

Chapter 2 deals with the methodology used through this work. Specifically, we provide a description of QM/MM and metadynamics techniques.

Chapter 3 reports our investigation of the reaction mechanism of *Sm*ChiB, a chitinase from family GH18. We unravelled its catalytic mechanism, paying special attention to the reaction

intermediate, whose nature remained controversial. The results are compared with a GH56 hyaluronidase that exhibits similar active site features.

Chapter 4 reports the conversion of a GH into a phosphorylase by a single point mutation. We rationalize this novel modification in a GH84 *O*-GlcNAcase, evaluating both substrate accommodation and reactivity, in the context of substrate-assisted catalysis.

Chapter 5 reports our study on the possibility of amide tautomerization in GH catalysis. The occurrence of asparagine tautomerization is investigated in GH85 ENGases, enzymes that operate via the substrate-assisted mechanism to degrade *N*-glycans in proteins.

Chapter 6 reports a study of the conformational properties of pyranose rings. By means of sugar puckering coordinates, we explore the conformational space of carbohydrates involved in substrate-assisted catalysis and their related inhibitors.

Chapter 7 lists the most relevant conclusions of this work.

Contents

1. General introduction.

1.1. Carbohydrates

Carbohydrates[1] (also referred as saccharides or sugars) are defined as polyhydroxy aldehydes or polyhydroxy ketones with the empirical formula $C_m(H_2O)_n$ (where *m* and *n* might differ) ubiquitously distributed on Earth. For instance, cellulose, a polysaccharide mainly found in plants which consists of linear chains of glucose units, is the most abundant organic polymer with total production of 10^{11}–10^{12} tons/year.[2] Even though, carbohydrates have been extensively studied not only because of their ubiquity, but also because they have a pivotal role in living organisms. Contribution to cellular signaling, energy storage or being structural components of relevant biomolecules such as coenzymes or nucleic acids are clear examples of central biological processes in which carbohydrates are deeply involved.[3–5]

1.1.1. Classification, nomenclature, and structure

The classification of carbohydrates can follow several criteria. For instance, carbohydrates are classified depending on the number of sugar units in which they are composed of. In terms of nomenclature, the word "saccharide" follows a numerical prefix: monosaccharide for sugars of just one carbohydrate moiety, disaccharide for two, etc. In addition, carbohydrates formed of between three and ten sugar units are considered as oligosaccharides, whereas if they are composed of more than ten are kwon as polysaccharides (Figure 1.1a). Just considering a single monosaccharide, one could use its number of carbon atoms to classify it. For example, pentoses, hexoses and heptoses are carbohydrates that contain five, six and seven carbon atoms, respectively.

The configuration of the asymmetric centre is used as another classification criterium. D- and L- notation allows to differentiate sugar enantiomers, such as D-glucose and L-glucose (Figure 1.1b). Most of natural carbohydrates are D-saccharides. The aldehyde or ketone group from a saccharide of five or more carbons reacts with one of its hydroxyl groups in order to form a cyclic hemiacetal or hemiketal, respectively. Depending on which hydroxyl reacts, different sugar cycles can be generated. This process creates a new stereogenic centre, which is referred to as the anomeric carbon (C1). Two possible anomers might form from the previous process. In the case of D-glucose (Figure 1.1c), α-D-glucose and β-D-glucose which differ in the configuration of its anomeric carbon and exhibit different presence in equilibrium (36% and 64%, respectively).[6] This tendency is partially dictated by the anomeric effect,[7] a stereoelectronic effect in which there is a stabilizing interaction (hyperconjugation) between the unshared electron pair on the pyranose oxygen and the σ* orbital of the C1-O1 bond.

Furthermore, monosaccharides are linked at different positions, which is then considered in the nomenclature of glycosidic bonds together with the anomeric configuration (β-1,4 or α-1,3, for example). The stereochemical variety of hydroxyl groups, in addition to their replacement by other substituents such as acetamido or sulphate groups (Figure 1.1a), and the conformational space that glycosidic bonds and sugar rings can explore, generate a myriad of carbohydrate structures.

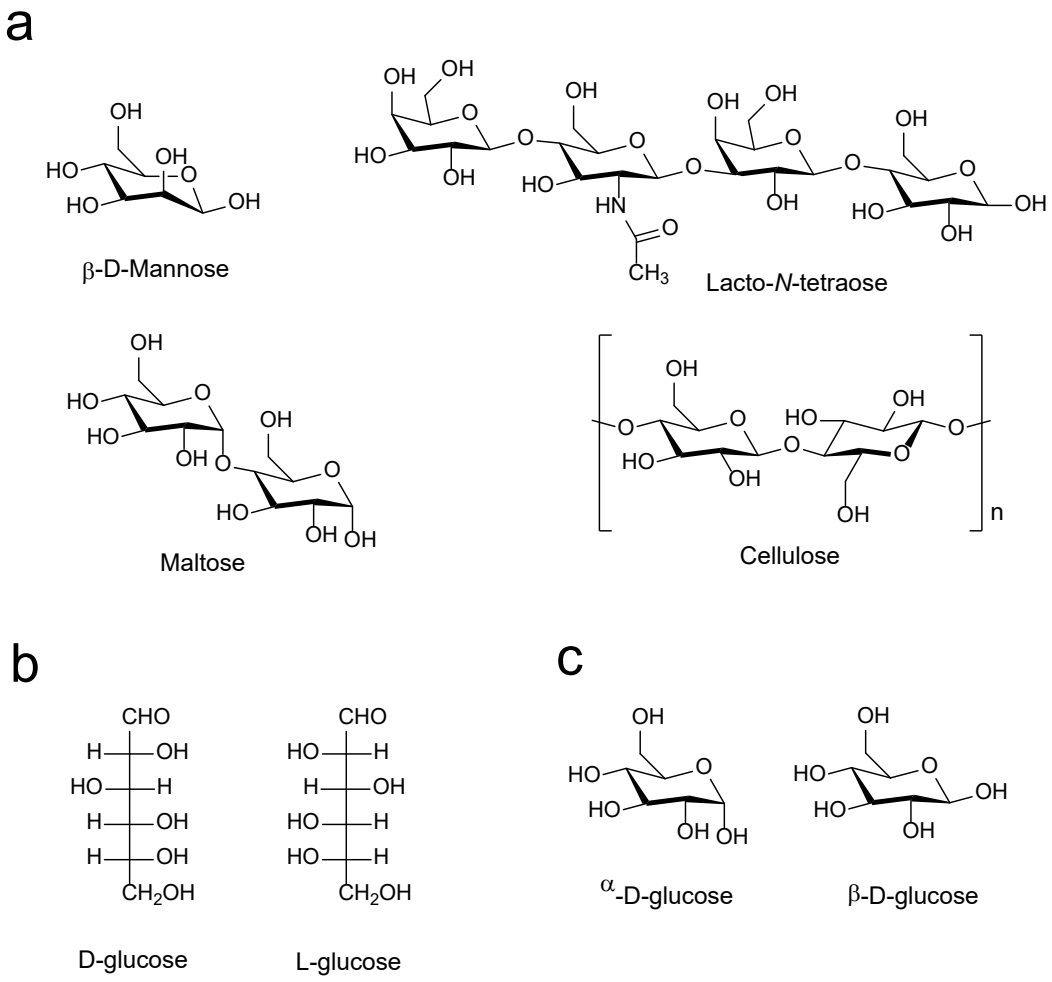

Figure 1.1. Chemical structures of carbohydrates. (a) Mannose, Maltose Glc(α-1,4)Glc, Lacto-*N*-tetraose Gal(α-1,4)GlcNAc(β-1,3)Gal(β-1,4)Glc and cellulose $(Glc(\beta\text{-}1,4)Glc)_n$, as examples of a monosaccharide, disaccharide, oligosaccharide, and polysaccharide, respectively. (b) Fisher projections of L- and D-enantiomers of glucose and (c) both β and α anomeric configurations of D-glucose.

1.1.2. Puckering conformations of carbohydrates

Of particular importance for catalysis is the conformation that sugar rings adopt (section 1.2.3). The rotation of their internal bonds provide significant flexibility and alter the axiality of the hydroxyl groups (or other substituents). Six-membered rings — pyranoses studied in this Thesis — can exhibit up to 38 different canonical conformations, whose nomenclature is regulated by the IUPAC.[8] Specifically, there are five types of conformations: chair, boat, skew-boat, half-chair, and envelope, represented by the capital letters *C*, *B*, *S*, *H* and *E*, respectively (Figure 1.2). For each conformation, a four atoms — at least — reference plane from the pyranose ring is considered. The rest of the atoms, the ones lying outside the plane, are designated by superscripts and subscripts if they are above or below the reference plane, respectively. The conformation of a pyranose ring can be quantitatively and unequivocally described by the puckering coordinates established by Cremer and Pople,[9] which will be used and further discussed along this Thesis.

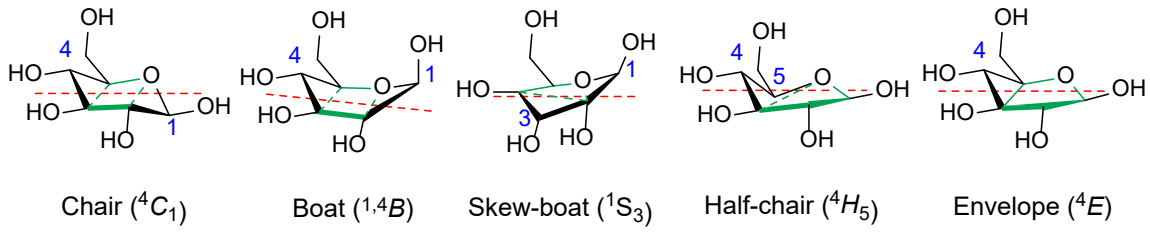

Figure 1.2. Examples of the five canonical conformations of pyranose rings, and in particular β-D-glucose, to show the axiality changes in its hydroxyl groups from one conformation to another. The reference plane is coloured in green and atomic indexes of atoms out of the plane to indicate a specific conformation atom are coloured in blue.

1.2. Glycosidases

In order to deal with the vast chemical space of sugars, Nature requires an extensive enzymatic toolbox known as Carbohydrate Active Enzymes (CAZymes).[10] CAZymes selectively degrade, form, and modify carbohydrates and glycoconjugates. Depending on their function, they are classified in different modules, such as glycoside hydrolases (GHs) or glycosyltransferases (GTs), which catalyse the hydrolytic breakdown and formation of glycosidic bonds, respectively.

GHs, or glycosidases, are enzymes that generally catalyse the hydrolysis of glycosidic linkages in carbohydrates and glycosides.[11] Since saccharides play a crucial role in biology, it is clear that the enzymatic machinery in charge of their breakdown is of major importance as well. Indeed, sugar molecules have a myriad of biotechnological and industrial applications, ranging from biofuels production from biomass,[12] food industry[13] to chemoenzymatic synThesis.[14] Furthermore, several GHs have a pivotal role in human health, since their deficiency and lack of activity can cause pathological conditions such as Gaucher's,[15] Pompe's,[16] Fabry's[17] or Krabbe's[18] diseases. Therefore, the study of these enzymes is paramount to understand diseases and design new therapeutic strategies.

1.2.1. Classification

According to their sequence, GHs and CAZymes in general are systematically classified in several families and curated in the CAZy database (www.cazy.org, 168 GH families reported in September 2020).[19,20] Usually, enzymes that belong to the same family not only exhibit sequence similarity, but also act on similar substrates and share function. This is translated into similar recognition regions from the enzyme for certain type of substrates, equivalent catalytic residues to perform catalysis and thus use of the same reaction mechanism. From a functional point of view, GHs can also be classified according the Enzyme Commission, depending on the substrate they act on.[21] Moreover, if the cleavage takes place at the latest terminal sugar moiety from a carbohydrate chain, GHs are further classified as *exo*-, whereas if catalysis is performed at other regions of the chain as *endo*-. Once the substrate has been hydrolysed, if a GH continues degrading the carbohydrate chain is classified as a *processive* enzyme, while if the substrate is liberated after one catalytic cycle as *non-processive*.

Taking GH active site topology as a criterium, three types can be distinguished: pocket, cleft and tunnel (Figure 1.3).[22] This structural feature is closely related with the type of substrates degraded by GHs and how they perform catalysis. For instance, pocket active sites are usually

present in *exo*-acting enzymes to optimally recognize terminal carbohydrates such as monosaccharides or the extremity of polysaccharides with several free chains. In contrast, a tunnel architecture is frequently employed by *processive endo*-acting GHs to release part of the cleaved product while remaining attached to the carbohydrate chain in order to execute more catalytic cycles, thereby processivity. Cleft active sites allow random placement of different carbohydrate moieties in polymeric substrates and is commonly found in *endo*-acting GHs. Considering that GHs can break glycosidic bonds found in a wide variety of carbohydrate substrates, and that their binding modes may differ from each other, it is necessary to establish a consistent and proper nomenclature for the different active site regions. Active sites are divided in several subsites, each named as *-n* and *+m* (being *n* and *m* integers) and taking as a reference the exact position where the glycosidic bond is hydrolysed (between *-1* and *+1* subsites). Negative and positive subsites are assigned to the *non-reducing* and *reducing* terminal sugars, respectively.

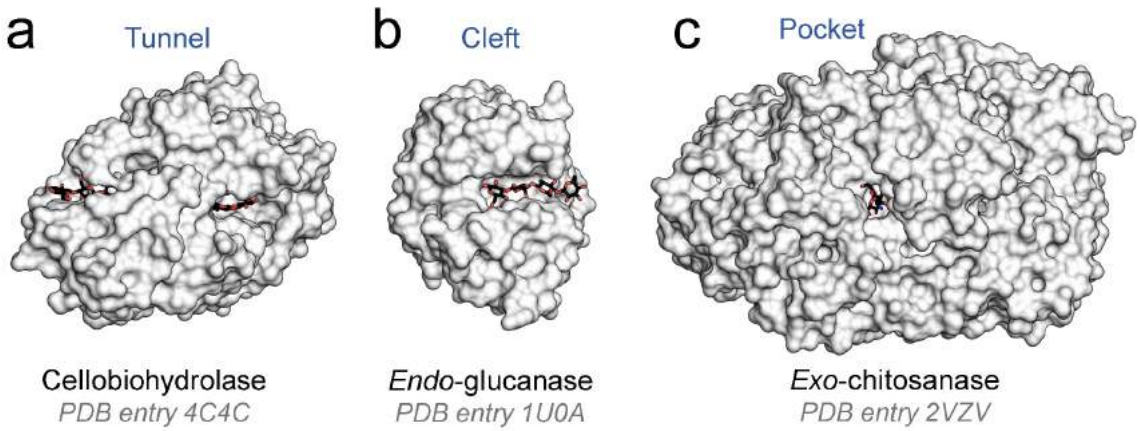

Figure 1.3. Topology of GH active sites. Carbohydrates are shown as black sticks and GHs as light grey surfaces. (a) Tunnel topology of a cellobiohydrolase with a cellononaose fiber complexed inside (PDB entry 4C4C).[23] (b) *Endo*-glucanase with a β-glucan tetrasaccharide bound to its *-4*, *-3*, *-2 and -1* subsites, shaping a cleft (PDB entry 1U0A).[24] (c) *Exo*-chitosanase with a chitosan disaccharide buried in its small pocket active site (PDB entry 2VZV).[25]

Finally, GHs are classified on the basis of their reaction mechanism, one of the most important criterium, as discussed in the following section.

1.2.2. Reaction mechanisms

The reaction mechanism of GHs is common within families and generally consists in a bimolecular nucleophilic substitution (S_N2) where two essential acidic residues catalyse hydrolysis: the general acid/base (proton donor) and the nucleophile (general base).[26] Depending on the stereochemical outcome — if the anomeric configuration is retained or inverted — GHs are classified as *retaining* or *inverting* enzymes. The canonical reaction mechanisms in both cases, the so-called classical Koshland mechanisms,[27] are shown in Figure 1.3. While retaining GHs operate via a double displacement mechanism involving a covalent glycosyl-enzyme intermediate (Figure 1.4a), inverting GHs follow a single displacement mechanism (Figure 1.4b). Each reaction step in both mechanisms involves an oxocarbenium ion-like transition state (TS, Figure 1c), characterized by development of positive anomeric charge, sp^2 hybridization, and ring distortion. The positive charge developed at the anomeric carbon is partially stabilized by electron donation from the pyranose oxygen. Sugar distortion during — and towards to — catalysis helps the reaction to proceed. This will be further discussed in Section 1.2.3.

In both retaining and inverting enzymes, the acid/base residue is similarly positioned, pointing towards the oxygen atom of the glycosidic bond about to be cleaved (between *-1* and *+1* subsites). Nevertheless, the general base position differs, which is strongly related to how the reaction evolves. In inverting enzymes, the catalytic residues are separated by a distance of approximately 10 Å. This allows the accommodation of a water molecule between the nucleophile residue and the anomeric carbon. The single displacement occurs when the acid/base protonates the glycosidic bond promoting the departure of the leaving group, and the nucleophile activates a water molecule that attacks the anomeric carbon. The obtained product exhibits a net inversion of the anomeric configuration. Contrarily, in retaining enzymes the catalytic residues are separated by an average distance of 5.5 Å. In this case, the reaction mechanism comprises two steps, deglycosylation and glycosylation. During the first reaction step (glycosylation), the general acid protonates the oxygen of the scissile glycosidic bond, whereas the nucleophile attacks the anomeric carbon, forming a covalent glycosyl-enzyme intermediate. In the second step (deglycosylation), the acid/base acts now as a base and deprotonates an incoming water molecule that performs the nucleophilic attack on the anomeric carbon to form the final product with net anomeric retention.

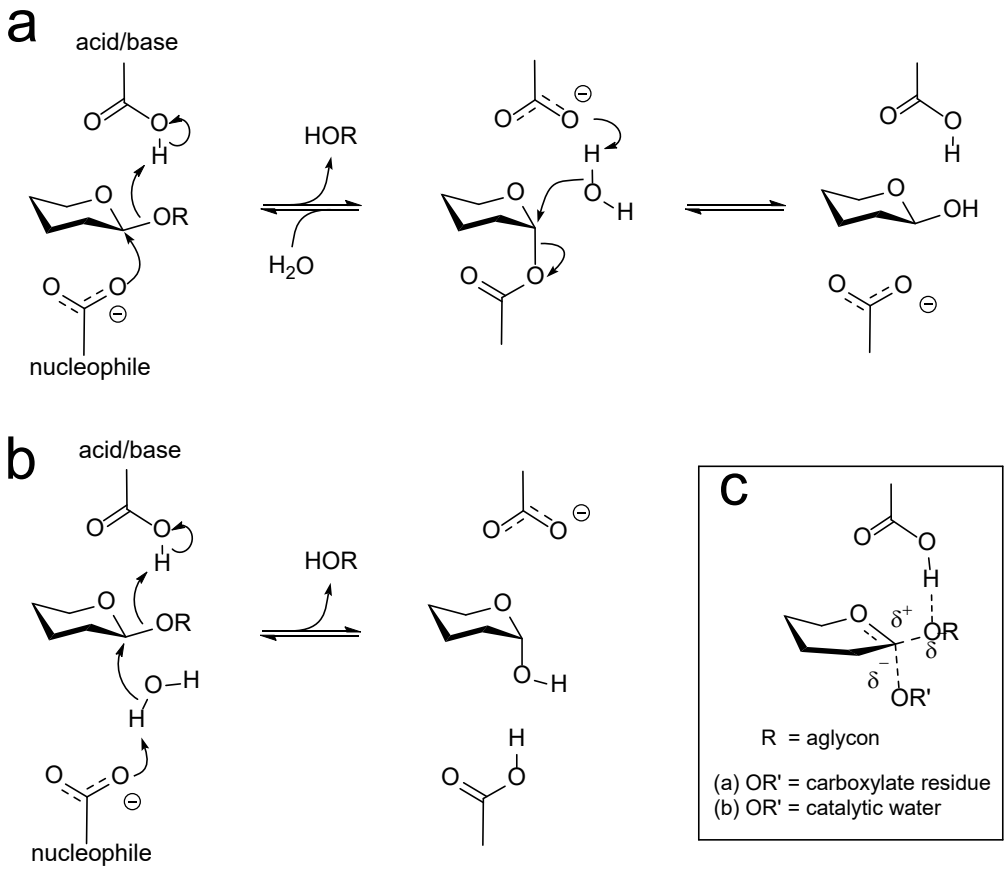

Figure 1.4. General reaction mechanisms of β-glycosidases. (a) Retaining classical double displacement. (b) Inverting single displacement. (c) Oxocarbenium ion–like transition state of the first half-reaction. The nucleophilic oxygen that attacks the anomeric carbon is part of a distinct chemical group in (a) or (b). Dashed lines indicate partially broken/formed bonds.

Enticingly, several GH families use mechanisms far beyond from the classical Koshland mechanisms.[28] For instance, family GH4 glucosidases catalyze redox and elimination reaction steps involving NAD^+ and Mn^{2+} as cofactors, whereas family GH33 sialidases follow an elimination and hydration mechanism involving a tyrosine residue as a nucleophile. Another striking example is the case of GH109 enzymes, which operate using a similar mechanism to GH4 enzymes, but without requiring a metal cofactor. A very recent study of one of these GHs showed that exhibits both α- retaining and β-inverting activities.[29] This remarkable exception escapes from the paradigm of sharing a common stereochemistry for substrates and products within the same GH family.

Likewise, some GHs that degrade 2-*N*-acetylated carbohydrates use a modification of the canonical retaining mechanism. In the first step, the carbonyl group oxygen of the *N*-acetyl sugar at the *-1* subsite, rather than an enzymatic residue, acts as a nucleophile forming a

bicyclic reaction intermediate, not covalently bound to the enzyme. This is the so-called *substrate-assisted* mechanism, which is the central issue investigated in this Thesis (in-depth introduced in Section 1.3). Even though, this feature is not unique to these GHs, since it has been recently demonstrated that in GH99 endo-α-mannosidases a hydroxyl group from mannose attacks the anomeric carbon to form an epoxide reaction intermediate.[30]

1.2.3. Catalytic itinerary

The reaction mechanism Last past half century, Phillips early proposed that conformational changes in the pyranose rings might play a role in glycosidase catalysis on the basis of X-ray crystallography.[31] He solved the very first enzyme X-ray structure in 1965, from Lysozyme,[32] a GH that degrades peptidoglycans. Over time, more insightful X-ray crystallographic investigations have been carried out,[33] elucidating novel GH structures and, in conjunction with computational techniques,[34] allowed to map and rationalize the conformational itineraries that carbohydrates follow during catalysis and its critical importance.[35]

Nowadays it is well-known that sugar ring distortion yields electronic and structural features required in the TS of reactions catalysed by GHs. These provide the aforementioned oxocarbenium ion-like character: (i) coplanarity of C5, O5, C1, and C2 atoms, (ii) sp^2 hybridization, (iii) positive anomeric charge development and (iv) C1···O5 distance shrinkage. Hence, the conformational space for a TS is delimited to eight specific canonical conformations: two half-chairs, two boats and four envelopes (Figure 1.5a). Furthermore, in the Michaelis complex (MC, i.e., when the substrate is bound in the enzyme active site before reacting), the sugar at *-1* subsite typically does not shape a relaxed $^4C_1$ chair conformation, the most stable in solution. Instead, the carbohydrate distorts to favour the reaction by orienting axially the leaving group (aglycon) preventing steric hinderance for the nucleophilic attack and placing the glycosidic oxygen close to the acid/base residue. Therefore, the substrate is pre-activated for the enzymatic reaction. This was first observed in a X-ray study of a retaining β-glucanase.[36] Later, Biarnés et al. confirmed the $^1S_3$ conformation by computational methods in the MC of another retaining β-glucanase.[37] In parallel, they elegantly demonstrated that this pre-activation is an intrinsic property of sugar molecules.[38–40] By analysing isolated monosaccharides, they found that albeit the most stable conformation generally is a chair, other low energy conformations match the ones observed in MC X-ray structures. This indicates that GHs have evolved to accommodate distorted — and thus pre-activated — carbohydrates in order to catalyse their hydrolysis efficiently.

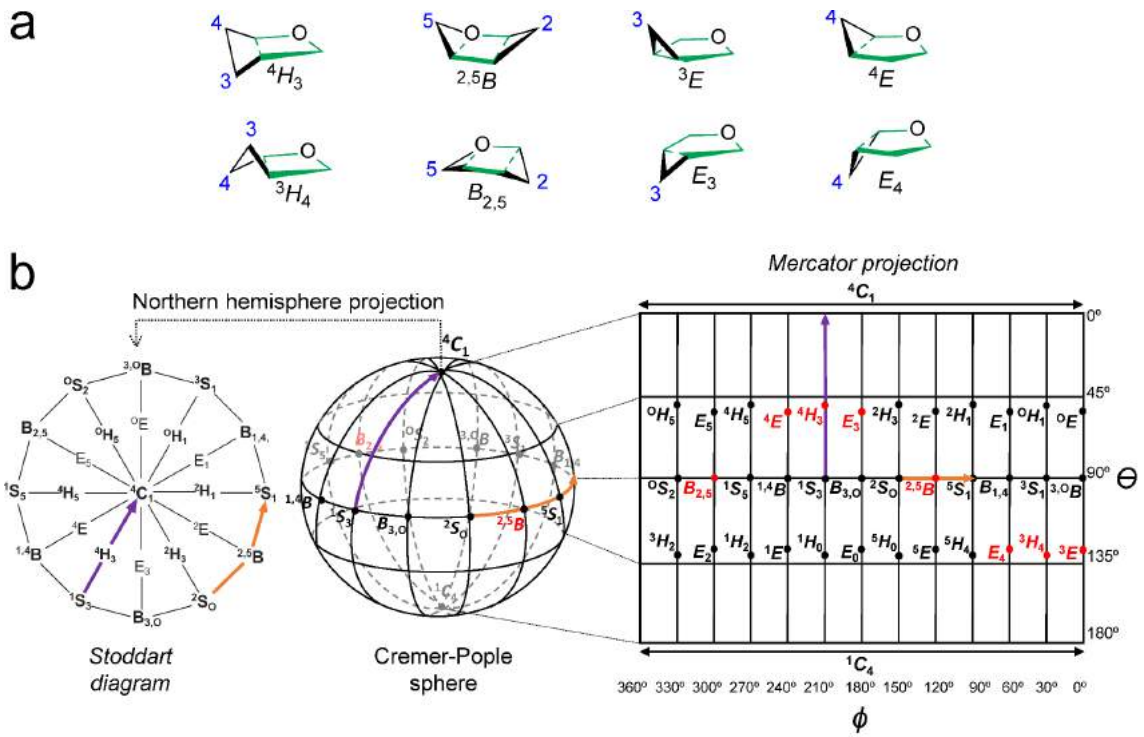

Figure 1.5. Puckering coordinates of pyranose rings. (a) TS-like conformers. The reference plane and atomic indexes of atoms out of the plane are coloured in green and blue, respectively. (b) Stoddart (left) and Mercator (right) representations of the Cremer-Pople sphere (middle). The catalytic itineraries $^{1}S_{3} \rightarrow [^{4}H_{3}]^{\ddagger} \rightarrow {}^{4}C_{1}$ and $^{2}S_{O} \rightarrow [^{2,5}B]^{\ddagger} \rightarrow {}^{5}S_{1}$ correspond to GH families GH5 and GH43, which are shown in purple and orange, respectively. Picture adapted from reference 35.

The understanding of conformational catalytic itineraries is crucial when designing TS-analogue-specific inhibitors and activity-based probes for GHs, since it allows to gain mechanistic insights and target these enzymes.[41–43] In the case of retaining GHs, the itinerary normally refers to the glycosylation step, which is generally the rate-determining one, particularly for sugar-like substrates. Stoddart introduced a diagram that connects all the possible pyranose conformations (Figure 1.5b).[44] This corresponds to the projection of the spherical puckering coordinates established by Cremer-Pople,[9] which distinguish unequivocally the conformation of any six-membered ring. All the possible conformers are located on the external shell of the puckering volume formed by these polar coordinates. The Stoddart diagram facilitates the visualization of the conformational pathway in a continuous space and, therefore, the interpretation of the GH catalytic itineraries. Even though, the conformational itinerary might not be a perfect longitudinal (or latitudinal) pathway across the Cremer-Pople sphere. Furthermore, despite it is expected that the catalytic itinerary is unique and conserved within a GH family, very recent examples demonstrated that certain exo-GHs are able to catalyse hydrolysis via two distinct conformational pathways, even starting from an undistorted MC.[45]

1.3. The substrate-assisted mechanism

1.3.1. Early hints of substrate-assisted catalysis in glycosidases

Historically, strong debates concerned glycosidase catalysis, especially regarding retaining GHs. Probably the most important was whereas they performed catalysis via long-lived oxocarbenium ion intermediate or, in contrast, through an enzyme-substrate covalent intermediate. The latter scenario was consolidated when Vocadlo et al. trapped this covalent intermediate by means of X-ray crystallography in 2001.[46] In the same line, the equivalent controversy arose in retaining GHs that catalyse carbohydrates that possess an acetamido group at position 2. However, there was an additional mystery in this case: no residues were obvious candidates that could serve to stabilize neither a putative oxocarbenium ion intermediate nor bond to the anomeric carbon. This implied that such enigmatic intermediate must be stabilised in another way.

In the mid-nineties, the X-ray structure of the plant enzyme hevamine from *Hevea brasiliensis*, a GH from family 18 with chitinase activity, was elucidated in complex with the inhibitor allosamidin.[47] Parallelly, another study demonstrated that allosamidin and Glc-thiazoline compounds were able to strongly inhibit the family GH20 *N*-acetyl-β-hexosaminidase from the plant *Canavalia ensiformis*.[48] Both GH enzymes degrade sugars *N*-acetylated at position 2, such as *N*-acetyl-D-glucosamine (GlcNAc, Figure 1.6). Inspection of the chemical structures of both inhibitors (Figure 1.6) suggested that they reassemble the putative reaction intermediate if the carbonyl oxygen of the *N*-acetyl group from the sugar at subsite *-1* would perform an intramolecular attack to its anomeric carbon. This phenomenon is known as *anchimeric assistance* or *neighbouring group participation* and it has been commonly observed and employed in organic chemistry. For instance, the spontaneous hydrolysis of methyl β-D-GlcNAc is 1000-fold enhanced over the hydrolysis of methyl β-D-glucose;[49] and the synthesis of β-D-GlcNAc derivatives evolves through a stable oxazoline intermediate.[50] Moreover, the reaction intermediate might be stabilized by the carboxylate residue that interacts with the nitrogen atom of these inhibitors, which explains its anomalous position (i.e. not nearby the anomeric carbon) and clearly differs from the role of the nucleophile in the classical Koshland mechanism. This, together with the nature of allosamidin and GlcNAc-thiazoline inhibitors, served as a first evidence to prove the substrate-assisted mechanism, which involves *anchimeric assistance* or *neighbouring group participation*.

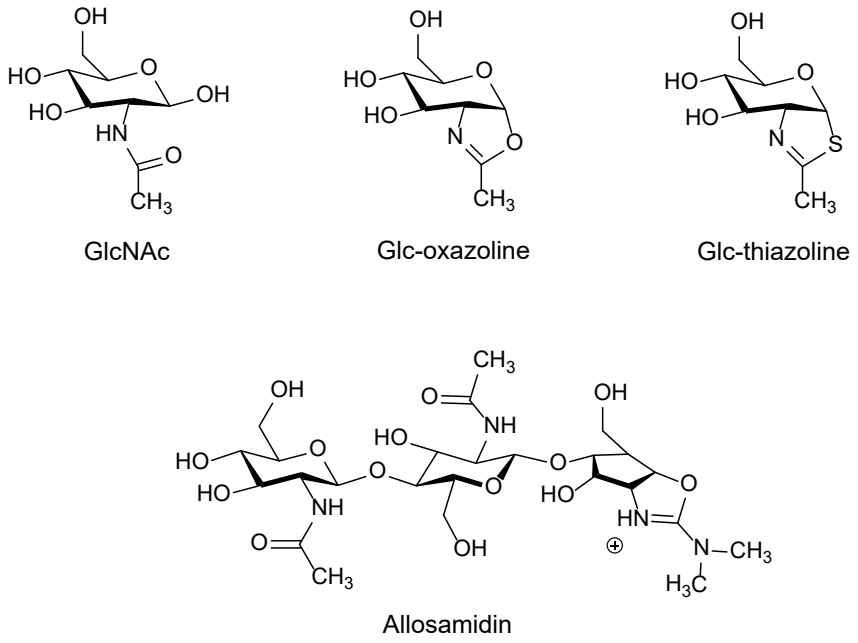

Figure 1.6. Chemical structures of carbohydrates related with substrate-assisted catalysis. From left to right, *N*-acetyl-D-glucosamine (GlcNAc), glucose oxazoline as the reaction intermediate if the intramolecular nucleophilic attack occurs (Glc-oxazoline) and its inhibitors derivatives (glucose thiazoline and allosamidin).

1.3.2. Reaction mechanism

The substrate-assisted mechanism (Figure 1.7) mainly differs from the classical Koshland retaining mechanism in the first step. Instead of a carboxylate residue (Asp/Glu), it is the substrate itself that acts as the nucleophile. In particular, the carbonyl oxygen of the 2-acetamido group of the carbohydrate at *-1* subsite attacks via anchimeric assistance its anomeric carbon. While this occurs, the acid/base residue protonates the glycosidic bond facilitating the departure of the aglycon part of the substrate, as in the classical retaining mechanism. This leads to the formation of a bicyclic glucose oxazoline/oxazolinium ion (Glc-ox/Glc-ox^{+}) reaction intermediate (discussed further below). Afterwards, the reaction intermediate is hydrolysed during the second step by an incoming water, which is activated by the acid/base residue, similarly to the classical Koshland mechanism. The TS of the reaction reassembles the one of the classical mechanisms (Figure 1.4c), but in this case the attacking oxygen is the one from the carbonyl group of the substrate.

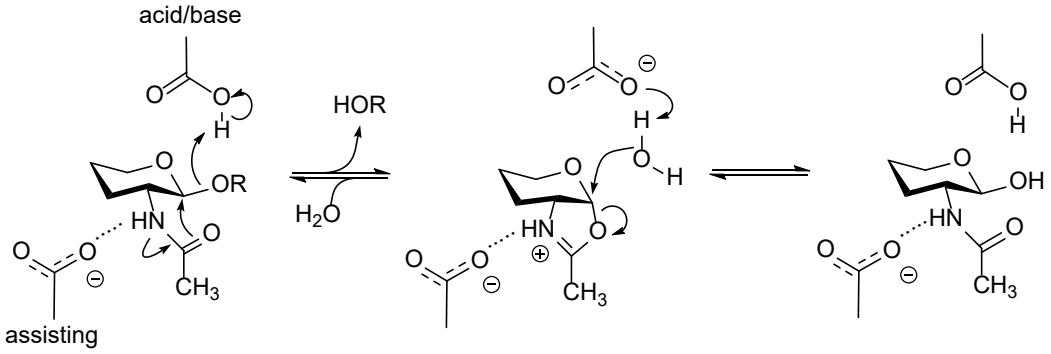

Figure 1.7. General substrate-assisted mechanism of β-glycosidases. The transition state reassembles the one shown in Figure 1.4c, but OR' being the carbonyl group of from the 2-acetamido of the substrate. Note that the reaction intermediate is depicted as a positively charged oxazolinium ion for the sake of simplicity, but this will not always be the case depending on the GH family.

Besides the acid/base residue, another conserved carboxylate residue in the active site of these enzymes, generally an Asp, participates in catalysis. That residue is known as the assisting residue and interacts with — and induces a positive charge at — the nitrogen atom from the *N*-acetyl group of the substrate. In turn, a negative charge is induced at the nearby oxygen atom from the carbonyl group, now being able to attack the anomeric carbon. The mentioned polarization effect allows to use a carbonyl group as an excellent substitute for a carboxylate, even if the latter is intrinsically a better nucleophile. Moreover, the assisting residue continues interacting with the nitrogen atom from the substrate at the reaction intermediate, playing a stabilizing role. Substitution of this residue decreases significantly the enzyme activity, but never as much as the mutations of the acid/base.[51-54]

The reaction intermediate (middle species in Figure 1.7) is characterized by the presence of an oxazoline ring. Oxazolines[55] consist of a five-membered heterocyclic molecules that have a double bound, most commonly located between the nitrogen atom and the carbon atom adjacent to the oxygen atom (i.e. 2-oxazoline). If an oxazoline molecule is protonated (at its nitrogen atom), it leads to an oxazolinium ion. This develops a positive charge, which is delocalized between the nitrogen atom and the adjacent carbon atom, leading to two possible resonant structures. However, the positive charge is most likely located at the nitrogen atom because this precludes the formation of a carbocation and generates a double bound, thereby gaining stability. The protonation state of this species will depend on its specific environment and interactions.

In the MC, the pK_a of the 2-acetamido group of the substrate is estimated as the amide group (pK_a value of 15), which is then decreased during the cyclization to form the reaction intermediate. Indeed, computational calculations of the Glc-ox$^+$ in solution predicted that exhibits a pK_a of 7.7.[56] This value is very similar to the one obtained for the allosamidin inhibitor (pK_a value of 7.4).[57] Since the carboxylate-bearing assisting residue interacts with the nitrogen atom of the GlcNAc interacts with the (acidic, pK_a of 3.7) assisting residue, one would expect that the substrate would remain protonated. However, depending on the specific environment which surely perturbs these pK_a values, and considering the changes likely occurring during the reaction, either of the two protonation states of the reaction intermediate is feasible. Hence, besides promoting the nucleophilic attack in the first step, the assisting residue stabilizes the reaction intermediate, either by acting as a general base subtracting the proton of the 2-acetamido group and forming a glucose oxazoline intermediate (Glc-ox), or by electrostatically stabilising the glucose oxazolinium ion intermediate (Glc-ox$^+$). This issue will be carefully addressed along this doctoral Thesis.

1.3.3. Substrate-assisted GH families

GH families that follow the substrate-assisted mechanism are GH18, GH20, GH56, GH84, GH85 and GH123 (listed in Table 1.1). While just some of them belong to GH-K clan (GH18, GH20 and GH85), all poses a $(\beta/\alpha)_8$ catalytic domain. They catalyse the cleavage of saccharides that contain an *N*-acetyl group at position 2 such as GlcNAc, which is present in chitin, *O*- and *N*-glycans or hyaluronic acid, for example. Thus, chitinases, *O*-GlcNAcases, or hyaluronidases are few examples of these enzymes, which have been studied in this doctoral Thesis. Nevertheless, not all enzymes that degrade substrates possessing a 2-acetamido group use this mechanism. For instance, enzymes from families GH3 and GH22 utilize a classical retaining mechanism with the participation of nucleophile residue. Other enzymes, such as chitinases from family GH19, employ an inverting mechanism. Hence, Nature has developed different mechanistic strategies to efficiently catalyse *N*-acetylated carbohydrates.

Table 1.1. GH families which follow the substrate-assisted mechanism. * Reaction mechanism not known, inferred by analogy with related families.

Family	Activity	Acid/Base	Assisting	Insertion
GH18	Chitinase, endo-β-*N*-GlcNAcase	Glu	Asp	Yes
GH20	Exo-β-*N*-GlcNAcase, β-*N*-GalNacase, β-6-SO_3-*N*-GlcNAcase, exo-lacto-*N*-biosidase	Glu	Asp	No
GH25*	Lysozyme	Glu	Asp	Yes
GH56*	Hyaluronidase	Glu	Asp	Yes
GH84	*O*-GlcNAcase	Asp	Asp	No
GH85	Endo-β-*N*-GlcNAcase	Glu	Asn	Yes
GH123	β-*N*-GalNAcase	Glu	Asp	No

A major distinction between GHs that follow this reaction mechanism is the nature and the location in the protein sequence of the catalytic dyad (acid/base and assisting residues). Table 1.1 illustrates that two main groups can be distinguished: whereas there is an insertion between the catalytic residues (families GH18, GH25, GH56 and GH85) or not (families GH20, GH84 and GH123). That is, if they are separated in the sequence or adjacent, respectively. This subtle difference is translated into a distinct conformation in the active site of these enzymes. If there is an insertion, the backbone can bend and both catalytic residues interact (Figure 1.8a), while when they are adjacent in the sequence they do not (Figure 1.8b). This very likely plays a role in catalysis since this interaction may alter the pK_a of the catalytic dyad and ultimately influence on the intermediate of the reaction.

A remarkable exception is the case of enzymes from family GH85. In all the other GH families that perform substrate-assisted catalysis, the assisting residue is an Asp, while in GH85 enzymes an Asn. This striking difference has been further investigated in chapter 5. Likewise, family GH123 enzymes are the only ones that degrade GalNAc containing substrates at *-1* subsite (glycosphingolipids), instead of GlcNAc. Apparently, this would not imply a mechanistic distinction in the context of substrate-assisted catalysis.

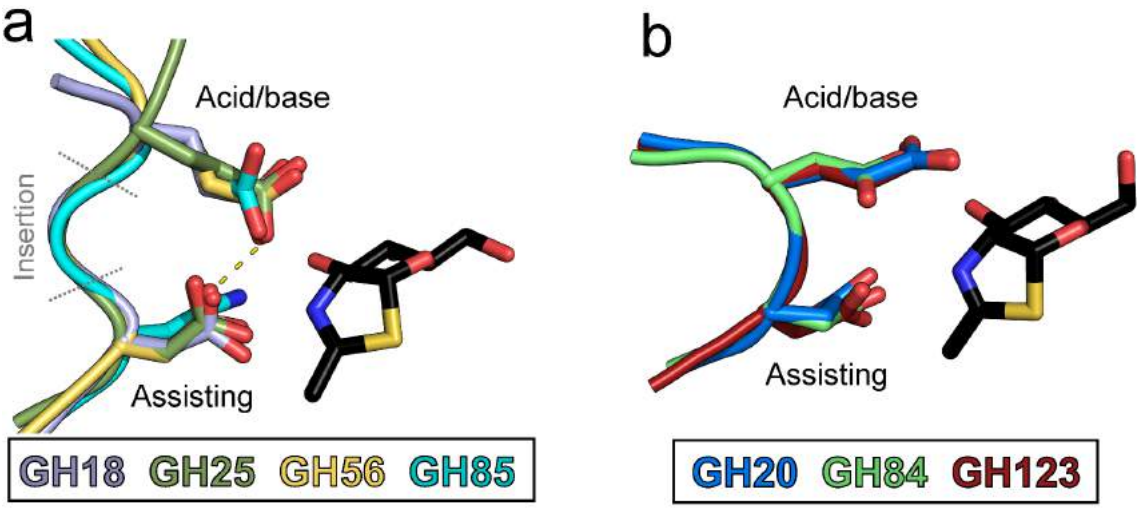

Figure 1.8. Structural alignment of the catalytic dyad within GH families that use substrate-assisted catalysis. (a) Enzymes with an insertion between the catalytic residues. This allows the backbone to bend and ultimately establish an interaction between these residues (dashed yellow line). The assisting residue in GH85 enzymes (cyan) stands out because is the only Asn, in comparison with an Asp in the other families. (b) GHs that possess their catalytic dyad consecutive in their sequence. Note the exception of the shorter acid/base residue in GH84 enzymes (lime green), which is an Asp instead of a Glu. X-ray structures used here for GH18, GH20, GH84, GH85 and GH123 were trapped with the reaction intermediate — or its thiazoline analog (black sticks) — but GH25 and GH56 structures had other substrates bound in their active sites. PDB entries: GH18, 1E6Z;[58] GH20, 5A6A;[59] GH25, 1JFX;[60] GH56, 1FCV;[61] GH84, 2CHN;[53] GH85, 2W92;[62] and GH123, 5L7V.[63]

1.4. Open questions

Despite the current knowledge of substrate-assisted glycosidase catalysis, a detailed mechanistic description of this atypical reaction mechanism, especially the factors controlling the nature of the reaction intermediate, is lacking. In particular, the following questions are yet not solved: are there any mechanistic differences between substrate-assisted GHs with an insertion in their sequence between the catalytic dyad in comparison with those GHs residues being adjacent in the sequence? Do these differences alter the nature (e.g. structure and protonation state) of the reaction intermediate?

As discussed previously (section 1.2.2), both Glc-ox and Glc-ox$^+$ protonation states are mechanistically feasible, and the literature is plenty of examples where both terms are used indistinctly, even within the same GH family. For instance, whereas in GH84 enzymes it was assumed that the substrate in the reaction intermediate features a neutral Glc-ox,[64] Glc-ox$^+$ is reported for GH18 chitinases.[65] In other GH families acting by substrate-assisted catalysis, the intermediate has been either described as Glc-ox$^+$,[56,66,67] Glc-ox,[41,68] or both.[62,69] Thus, a full characterization of the reaction mechanism in GHs that use substrate-assisted catalysis is needed. Moreover, since the reaction intermediate in these GH families is not covalently bound to the enzyme, as occurs in the classical retaining mechanism, it is difficult to understand if the designed inhibitors for these GHs are true mimics of the TS of the reaction. Thus, inhibition of substrate-assisted catalysis requires further investigation.

Here we have tackled the above general questions by studying GHs that use substrate-assisted but differ in aspects such as the type of residues forming the catalytic dyad and their position in the enzyme sequence. In addition, the isolated enzyme substrate has been investigated to infer their intrinsic properties and their possible connection with those of commonly used substrate-assisted GH inhibitors. The previous results that support the specific questions we plan to address are the following:

- Previous theoretical investigations of chitinases from family GH18 concluded that the reaction intermediate featured a positively charged Glc-ox$^+$. Since these enzymes have an Asp-X-Glu catalytic dyad that interacts, and the alternative scenario was not fully addressed, we revisited this issue (chapter 3).

- A GH84 hydrolase was converted into a phosphorylase by means of a single point mutation. Which are the molecular determinants of such transformation? Moreover, GH84 *O*-GlcNAcases have two adjacent Asp as catalytic residues that do not interact, in contrast

with GH18 enzymes. Will this modify the nature of the reaction intermediate or the mechanism? (chapter 4).

- ENGases from family GH85 are expected to use substrate-assisted catalysis, although they feature an Asn assisting residue, in contrast with an Asp as in all the other families. Both Glc-ox and Glc-ox$^+$ reaction intermediates are also compatible with the available X-ray structures, but one of these possibilities involves Asn tautomerization. Is the active site environment able to trigger such unusual Asn tautomerization? (chapter 5).

- The conformation of the molecules involved in the substrate-assisted mechanism has not been characterized yet. Studies of these compounds, together with their derived inhibitors, would help to understand both mechanistic insights and the inhibition of these GHs and their specificity. Are the intrinsic properties of a certain inhibitor similar to the TS or the reaction intermediate? Is there any reason behind inhibitor specificities within these GH families?

1.5. Objectives

In this Thesis we have aimed to unveil the catalytic mysteries around the substrate-assisted mechanism in GHs. Particularly, we have unravelled the reaction mechanism of four GHs encompassing enzymes catalytic dyads with and without insertion in the sequence; and investigated a GH85 enzyme as a clear exception in terms of the nature of the assisting residue (Asn instead of Asp). Finally, we studied the isolated compounds of the mechanism in conjunction with derived inhibitors to understand their specificity. Thus, the following specific objectives were pursued:

- Elucidate the two-steps reaction mechanism of GH18 chitinases and determine the nature of its reaction intermediate. Investigate its relatedness with a GH56 enzyme, whose catalytic dyad is equivalent and its reaction mechanism completely unknown.

- Disclose the conversion of a GH84 *O*-GlcNAcase into a phosphorylase by a single point mutation, while investigating the nature of the reaction intermediate in a substrate-assisted GH that has both catalytic residues adjacent in the sequence.

- Unveil the reaction mechanism of a GH85 ENGase considering the striking possibility of Asn tautomerization along the reaction.

- Evaluate the conformational properties of the isolated compounds involved in the substrate-assisted mechanism and their derived inhibitors.

2. Methodology: theoretical and computational chemistry.

2.1. Bridging calculations and experiments

Understanding enzymes and other biomolecular systems is a daunting task. A multidisciplinary effort has been made to properly describe at atomistic level the molecular mechanisms in which these proteins participate. In this context, simulations are broadly used to interpret, design, and predict experiments and, in turn, models in which calculations are based are continuously being refined with experimental data (Figure 2.1).[70] While in GHs investigations both X-ray crystallography and NMR generate crucial structural information of these enzymes, computer simulations can capture complex reaction coordinates, electronic reorganizations and protein/sugar flexibility that ultimately provide an in-depth understanding of GH mechanisms.[34]

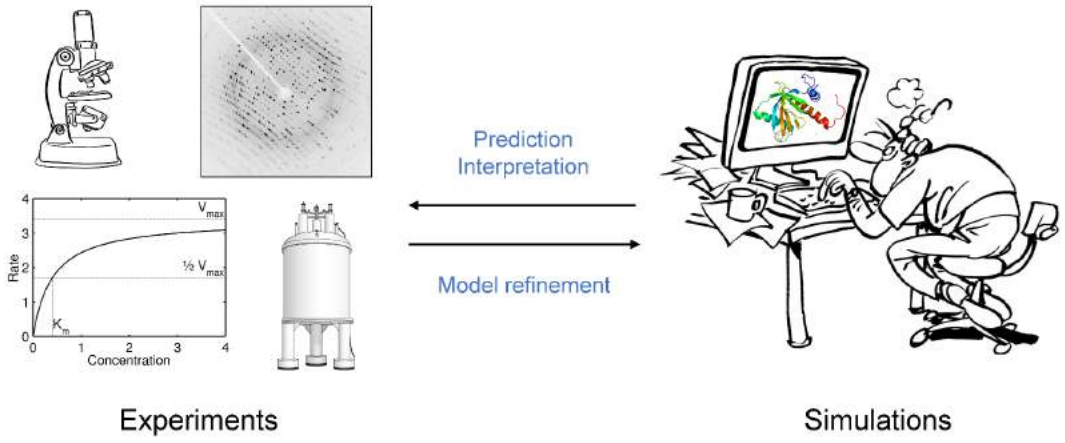

Figure 2.1. Interchange between experimental methods such as kinetics enzymology, X-ray crystallography or NMR; and theoretical techniques such as the ones employed in this Thesis.

One of the main goals of molecular simulations is to estimate thermodynamic properties of realistic systems. Features of interest not only include macroscopic properties (density, heat capacity, etc.), but also microscopic ones such as free energy differences associated with different processes under study. Even though one can think that thermodynamics apply at macroscopic level and statistical mechanics at microscopic scale, it is important to consider that the thermodynamics laws still hold on average regardless the length scale. For instance, in this Thesis we compared our GH simulations with the macroscopic observable k_{cat}, a constant that describes the turnover rate of an enzyme. We employed the transition state theory to obtain a free energy barrier from this experimental constant.[71]

In any molecular simulation, a particle-based description of the system under study is built and then propagated by either probabilistic or deterministic rules to generate a trajectory describing its evolution along the calculation. For each stored configuration of the system (snapshot) one can calculate several relevant properties and compute their average over the course of the entire simulation. Molecular Dynamics (MD) and Monte Carlo (MC) are the two main general techniques to propagate the system. While in MD methods the equations of motion are numerically integrated to produce a trajectory of the system, in MC techniques probabilistic rules are employed to generate a new configuration form the previous one in order to produce a sequence of states. MD simulations can provide structural, dynamic, and thermodynamic properties of the system, whereas MC calculations can be used to calculate structural and thermodynamic properties but not dynamics, since they lack any concept of time. All calculations performed in the work of this Thesis are based on MD methods.

2.2. Molecular dynamics

MD is a computational method for analyzing the physical movement of atoms and obtain relevant properties for a certain system.[72] Given an initial structure, usually obtained by X-ray crystallography or NMR, the system evolves by numerically integrating the Newton's equations of motion where forces between particles and their potential energies are computed at different levels of theory. Several algorithms can perform such integration (e.g. Verlet,[73] velocity Verlet or leapfrog)[74]. The particles of the system can be described by different underlying physical theories: in Molecular Mechanics (MM, Section 2.2.1), molecules are considered as atoms described by an electric charge and a potential energy function with a large number of parameters. Quantum mechanics (QM, Section 2.2.2) allows to describe the electrons explicitly and thus the energy of the system is computed by solving the electronic structure of molecules. From the initial coordinates, initial velocities are assigned according the Maxwell-Boltzmann distribution at a certain temperature, which provides an initial velocity in a certain direction for each particle. Afterwards, the force that acts on each particle of the system is computed from the potential function for each time step of the simulation in order to generate new coordinates and new velocities. The timestep (Δt) is a parameter that corresponds to the integration step of the equations of motion and determine the time resolution of the MD simulation. Its value is typically chosen to one tenth of the period of the fastest molecular motion, which usually is related with stretching of H···O bonds.

2.2.1. Molecular Mechanics

In Molecular Mechanics (MM), the description of molecules is based on the contributions of processes such as bond stretching or their torsions, without considering explicitly the electronic structure. Each atom is treated as a point particle with a punctual charge and thus generally no polarization nor electronic reorganizations are taken into account. This is accomplished by using a large set of parameters, obtained either by quantum calculations or experimental data, that describe the potential energy of the system under investigation. The energy of the system is computed according its ideal energy, which is characterized by the parameters employed in the calculation. These parameters, referred to as force field, include two types of interactions: bonded and non-bonded:

$$E_{total}^{MM} = E_{bonded} + E_{non-bonded} \qquad \text{Equation 1}$$

In turn, these types are divided in bond, angle, and torsion interactions (bonded), and Van der Waals and electrostatic as non-bonded interactions. Bond terms describe the interactions between atoms covalently bonded as

$$E_{bonds} = \sum_{bonds} K_b \cdot (r - r_{eq})^2 \qquad \text{Equation 2}$$

$$E_{angles} = \sum_{angles} K_\theta \cdot (\theta - \theta_{eq})^2 \qquad \text{Equation 3}$$

$$E_{torsions} = \sum_{bonds} \frac{V_n}{2} \cdot [1 + \cos(n\emptyset - \gamma)] \qquad \text{Equation 4}$$

where r and θ are the distance and angle in a certain instant, respectively; r_{eq} and θ_{eq} their respective equilibrium value. K_b and K_θ are the vibrational constants. Torsions include the number of n torsional energy barriers and their height (V_n), together with the phase angle γ for torsional angles.

Van der Waals interactions are described through a Lennard-Jones potential where both attractive and repulsive interactions are considered (R_{ij}^6 and R_{ij}^{12}, respectively). While attractive forces are long-range, repulsive forces act on decreased distances, where R_{ij} is the separation between atoms i and j. A_{ij} and B_{ij} are the Lennard-Jones potential parameters.

$$E_{Van\ der\ Waals} = \sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \qquad \text{Equation 5}$$

$E_{Van\ der\ Waals}$ is zero when the distance between atoms i and j is infinite. As this distance is reduced, the energy exhibits a minimum at the equilibrium distance, favoured by the attractive forces contribution. If the distance is reduced even more, the energy increases drastically because of repulsion at shorter distances.

Electrostatics interactions consider the atom charges q_i and q_j for each nuclei, generally called partial charges. The energy for this type of interactions is given by the Coulomb law:

$$E_{electrostatics} = \sum_{i<j} \frac{q_i \cdot q_j}{\varepsilon \cdot R_{ij}} \qquad \text{Equation 6}$$

The parameters specified in a force field are specific for each atom type, which considers their different chemical state. This allows to differentiate if, for example, a carbon atom is located within an aromatic system or it has a tetrahedral sp^3 hybridization. There are several protein force fields, such as ff14SB,[75] CHARMM[76] or OPLS.[77] Moreover, there are specific force fields for other molecules relevant in this work such as carbohydrates (GLYCAM[78] or CHARMM[79]) or several water models (TIP3P, TIP4P, etc)[80]. The use of force fields allows transferability for MD simulations, since the parameters obtained for small molecules such as amino acids can be employed in different macromolecules that contain them, such as proteins. Furthermore, MM description permits to simulate huge biological systems of hundreds of thousands of atoms at microsecond scale. Indeed, the MD calculation at nanosecond scale of the entire structure of the influenza virus (~160 million atoms) has been a recent milestone.[81] However, force fields show certain limitation, such as that parameters for all the molecules involved in a calculation are required, which is not always the case. Additionally, the goodness of the parameters employed can alter the accuracy of the results, as occurs in the case of phosphate molecules.[82] Another important limitation of MM is that no bond breaking/forming is allowed, except when reactive force fields are employed. This means that the chemistry of the system is constant along the entire simulation, which includes the initial protonation state for each chemical group. However, techniques that consider the pH dependencies of titratable amino acids in MD calculations have been developed to solve this issue, such as constant pH MD.[83]

Here we employed the AMBER software[84] to perform MD with MM level description, hereafter called classical MD, in order to equilibrate GHs in solvation at 300 K before carrying out more complex calculations.

2.2.2. Quantum Mechanics

Treating particles with quantum mechanics (QM) in simulations allows to accurately describe their electronic structure, thereby enabling to capture mechanistic details in chemical reactions or other processes that require electronic reorganizations.

The starting point of QM description is the time-dependent Schrödinger equation, from which the time-independent formula can be obtained and used to predict the energy and wave function of the stationary state of the system:

$$E\Psi = H\Psi \qquad \text{Equation 7}$$

where $E\Psi$ is the energy of the wave function Ψ, and $H\Psi$ its Hamiltonian operator. The probability of locating a particle is determined by the product of the wave function and its complex conjugate. The wave function depends both on the movement of the nuclei and the electrons. To reduce the complexity when calculating the energy of the system, Born and Oppenheimer proposed to separate the degrees of freedom of nuclei and electrons

$$\Psi_{total} = \Psi_{nuclei} \cdot \Psi_{electrons} \qquad \text{Equation 8}$$

2.2.2.1. Density Functional Theory

Since the computational cost of computing all the degrees of freedom in practical systems scales drastically, Density functional theory (DFT) was proposed to overcome such bottleneck within chemical accuracy. DFT offers a scheme to calculate the total energy of a polyatomic system from its atomic coordinates, exhibiting a fair compromise between computational cost and accuracy, compared with other QM methods. According DFT, the energy of the fundamental state for an electronic system where electrons interact between them and are under an effective potential, is a unique function of the electron density $\rho(\mathrm{r})$. This observable can be computed by minimizing the energy functional respect the density

$$E^{DFT} = \min_{\rho(\mathrm{r})}(E[\rho(\mathrm{r})])$$ Equation 9

The procedure for an MD method based on DFT starts by solving the Kohn-Sham equations[85] through an iterative process from the initial monoelectronic orbitals $\Psi_i(r)$, minimizing the density functional to obtain the total energy. At this point, the gradient of the nuclear coordinates of the system is evaluated considering the mass of each atom and the equations of motion. From the new nuclear positions, the process is iteratively repeated, calculating again the electronic structure. This method is known as Born-Oppenheimer MD.

The limiting step for such calculation is the minimization of the functional and solving the Kohn-Sham equations. In order to tack this issue, Car and Parrinello developed in 1985 a MD method which overcomes this problem (introduced in the next section). It is worth mentioning that the results of DFT calculations strongly depend on the choice of the functional employed, so usually a benchmark work is required to assure that a given functional is able to properly describe the interactions in a particular system. In this Thesis we employed the Perdew-Burke-Ernzerhof (PBE) generalized gradient-corrected functional,[86,87] which has been successfully applied in investigations of GHs in our group.[88] Furthermore, it has been demonstrated that this DFT functional provides accurate results in terms of energetics and conformations in carbohydrate systems (within chemical accuracy).[89]

2.2.2.2. Car-Parrinello Molecular Dynamics

The Car-Parrinello MD (CPMD)[90] is an *ab initio* technique that combines DFT and MD. The main feature of this type of calculation is that the Kohn-Sham orbitals are included in the equations of motion as new degrees of freedom so that both nuclei and electrons degrees of freedom evolve simultaneously. The advantage of this approach is that minimization of the density functional is not required for each MD step and thus the computational cost of the entire simulation is significantly reduced. However, the Δt has to be smaller (0.1-0.2 fs) in order to maintain adiabaticity and describe properly the motions of the system.

The electronic degrees of freedom are included as the coefficient expansion of the Kohn-Sham orbitals in a planewave basis set as

$$\Psi_i = \sum_j c_{ij} \cdot e^{i\,G_j\,r}$$ Equation 10

where G_j are the reciprocal lattice vectors and c_{ij} their corresponding coefficients of the planewaves expansion, which vary as the equations of motion are integrated along the simulation. The CPMD method is relies on the use of an extended Lagrangian where the electronic degrees of freedom are described as classic particles

$$\mathcal{L} = E_N^{kin} + E_{el}^{kin} - E^{KS} + \sum_{ij} \Lambda_{ij} \left(\int d\mathrm{r}\, \Psi_i^*(\mathrm{r}) \Psi_j(\mathrm{r}) - \delta_{ij} \right) \qquad \text{Equation 11}$$

where E^{KS} is the potential energy computed from DFT and Λ_{ij} the Lagrange multipliers to ensure orthonormality of the orbitals during the calculations. The kinetic energy E_{el}^{kin} can be separated in

$$E_N^{kin} = \sum_N \frac{1}{2} M_N R_N \qquad \text{Equation 12}$$

where M_N and R_N are the mass and position of nuclei, respectively. The fictious electron kinetic energy E_{el}^{kin} associated to the electronic system $\Psi_i(r)$ is

$$E_{el}^{kin} = \sum \mu \int d\mathrm{r}\, |\Psi_i(\mathrm{r})|^2 \qquad \text{Equation 13}$$

The fictitious mass μ is a parameter that controls the time scale for the electronic movement. Finally, the total energy of the system will be determined by

$$E_{tot}^{CP} = E_N^{kin} + E_{el}^{kin} - E^{KS} \qquad \text{Equation 14}$$

The CPMD procedure (Figure 2.2) consists in first minimizing the density functional and afterwards integrating the equations of motion to obtain the nuclei and electronic coordinates (Kohn-Sham orbitals in the case of the electronic system), which evolve over the potential energy surface of the fundamental state of the system. With that, we avoid computing the minimization of the density functional in the next MD step, reducing the computational cost of the simulation.

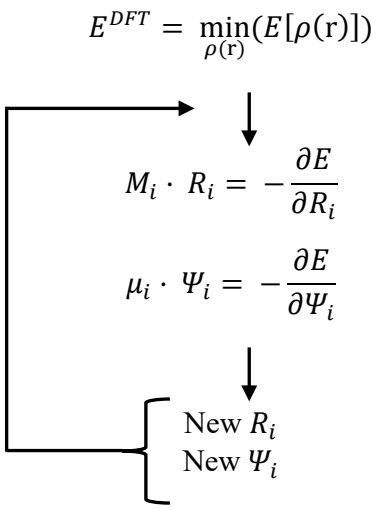

Figure 2.2. Scheme of a CPMD calculation.

Both μ and Δt parameters significantly affect the stability of a CPMD calculation. Because of that, these have to be tuned and tested in order to maintain system adiabaticity.[91] As explained above, Δt depends on the higher vibrational frequencies, which in this case are those associated to the electronic degrees of freedom. To maintain the system adiabatic, the nuclear and the electronic movements cannot interchange energy (i.e. their vibrational frequencies should be separated). With that, the system oscillates around the fundamental state where the electronic system does not separate from the Born-Oppenheimer surface. In CPMD, the highest vibrational frequency depends on μ: as μ increases, the highest vibrational frequency decreases. Typically, values between 500 and 1100 atomic units are used for this parameter.

2.3. Hybrid methods Quantum Mechanics / Molecular dynamics

Simulations of enzymatic catalysis requires highly accurate methods to compute chemical reactions within complex biological systems. This means that both the electronic structure and reorganization of atoms involved in the reaction — those belonging to catalytic residues and substrates — should be described at QM level. Additionally, the entire system (rest of the protein and solvent) is required in the calculation, described at MM level, to provide a realistic environment. Thus, there is an obvious bottleneck when investigating enzymes with computational techniques: describing at QM level events that occur in huge biological systems (around 50.000-120.000 atoms).

To overcome such limitations, quantum mechanics/ molecular mechanics (QM/MM) hybrid methods were developed,[92-94] which combine both QM and MM strengths. Introduced by Warshel and Levitt in 1976,[92] who along Karplus[94] were recognized with the 2013 Nobel Prize in Chemistry "for the development of multiscale models for complex chemical systems." In this type of techniques, the system is separated into two different regions (Figure 2.3). The QM region, where the enzymatic reaction takes places (usually the active site and atoms involved in the quantum process) and the MM part, which includes the rest of the system (protein and solvent). These parts are described at QM and MM level of theory, respectively.

While it is straightforward that atoms within the same region should be described by each respective level of theory, the interactions between QM and MM parts can vary. Depending on how these interactions are treated, the energy of the system is described by two type of coupling schemes: subtractive and additive. The former computes the QM/MM potential energy of the system by adding the energy from the QM part at QM level to the entire system energy computed at MM level. Afterwards the energy of the atoms of the QM region calculated at MM level is subtracted. This approach exhibits several drawbacks such as that a flexible force field description is required for the QM part and that polarization of the electron density by the MM part is not considered.

Additive schemes obtain the QM/MM potential energy from by adding the QM part computed at QM level of theory, the MM part at its corresponding force field description and finally QM-MM coupling terms. In this Thesis we employed the additive scheme developed by Laio et al.,[95] where the QM region is treated at DFT level (CPMD) and the MM with either Gromos96 or Amber force-fields

$$E^{QM/MM} = E^{QM} + E^{MM} + E^{QM-MM} \qquad \text{Equation 15}$$

The E^{QM-MM} term can be divided in bonded and non-bonded terms, as explained above.

$$E^{QM-MM} = E^{QM-MM}_{bonded} + E^{QM-MM}_{non-bonded} \qquad \text{Equation 16}$$

2.3.1. Bonded $E^{QM\text{-}MM}$ terms

Special attention must be paid when the frontier between the QM and the MM regions cut a covalent bond. The adjacent QM atoms would be unsaturated thereby showing one or more unpaired electrons in the quantum subsystem. This is the case when GHs are under investigation and the sidechain of the catalytic residues need to be described quantum mechanically.

Several approaches have been developed to saturate these covalent bonds: monovalent capping atoms (usually hydrogens), link atoms pseudopotentials or optimized effective core potentials. In this Thesis we cut carbon-carbon bonds by using capping hydrogens with scaled down pseudopotentials and the required valence charge, so that the distance of this type of bonds is correctly described. The rest of bonded E^{QM-MM} interactions such as bond, angles or torsions involving at least two QM atoms are described from the force-field.

2.3.2. Non-bonded $E^{QM\text{-}MM}$ terms

The non-bonded E^{QM-MM} energy is described by both steric and electrostatic terms, which include Van der Waals and electrostatic interactions, respectively. While the steric terms are described at MM level of theory by a Lennard-Jones potential the electrostatic interactions should take into account both MM charges and the quantum charge density of the QM region.

There are different electrostatics coupling schemes: mechanical embedding, polarized embedding, or electrostatic embedding. While the mechanical embedding scheme does consider the MM charges influence on the QM part, the polarized scheme requires polarized force fields, which are not always available. In this Thesis we used the electrostatic embedding scheme, where the QM part is polarized by the MM atoms. The electrostatic embedding scheme exhibits two main problems: the electron spill-out and the computational cost. The former consists in the occurrence of an over-polarization the electron density of the QM region by the point charges of an MM atom nearby the boundary (attraction or repulsion). In order to avoid this artifact, smeared-out charges for the MM atoms are used instead of the typical point charges. In CPMD, this is computed by a hierarchical electrostatic embedding approach, which also solves the second problem by not computing explicitly all the electrostatic interactions of the system. The hierarchical approach splits the electrostatics interactions between QM and MM parts in three contributions: short-range, medium-range and long-range, which are defined considering the NN, MIX and RESP subregion cut-offs (Figure 2.3).

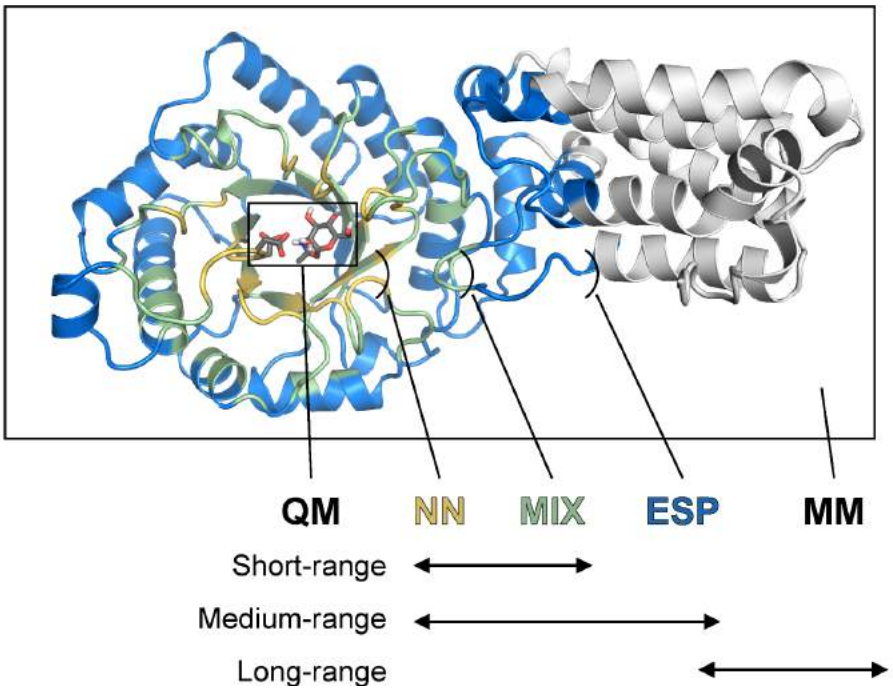

Figure 2.3. Schematic representation of the QM/MM description employed in this Thesis (CPMD). QM atoms are represented as black sticks. The rest of the system form the MM region. The electrostatic embedding scheme splits the MM region in several subregions by using NN, MIX and RESP cut-offs (coloured as orange, green and blue, respectively). The electrostatic treatment of the system implies three different type of interactions (short-, medium- and long-range). Short-range and medium-range description overlap since depending on how much strong the point charge of a certain atom is, this will be described by one approach or the other.

The short-range term involves the nearest neighbor atoms (NN region). This is, MM atoms at a certain cut-off between the QM region and the following MM subregion (MIX). Among them, if the point charge is < -0.1 e or > 0.1 e, the Coulomb interaction is described considering smeared charges to avoid the electron spill-out. The quantum charge density is mapped to a real space grid in the QM region that is evaluated explicitly with the MM point charges by the Coulomb law. Medium-range interactions include MM atom localized between the MIX and ESP subregions, and NN atoms that has not so strong point charges (i.e. > -0.1 e or < 0.1 e). Here, the charge density is represented by the electrostatic potential-based D-(R)ESP charges centered on the QM atoms, reproducing approximately the nuclear point charges in the QM region. Finally, the long-range term approximates the interaction between the rest of the MM atoms with the quantum charge density by expanding the latter in terms of multipoles up to quadrupolar order. Overall, this approach allows to properly describe the electrostatic interactions between the QM and the MM regions without compromising the computational cost of the calculation.

2.4. Metadynamics

MD is a powerful technique that has become a standard tool and had a huge impact in several branches of science, from physics to material science, from chemistry to biology. Nevertheless, MD suffers from certain limitations such as the timescale required to observe a certain phenomenon (i.e. visit all the energetically relevant configurations for a certain system). There are particular circumstances in which the free energy surface (FES) exhibits several local minima separated by energy barriers. To explore the entire FES of the process of interest, the system needs to escape from the initial minimum by surpassing those energy barriers. Protein folding or chemical reactions — including enzymatic catalysis — are examples of biochemical processes in which the *time scale problem* needs to be addressed and considered. For example, a small denatured protein requires 10^{-4} seconds approximately to explore its conformational FES and find its folded state. A rough estimation of the computational cost for observing a single folding event results in approximately one year of dedicated time on a petaflop machine. Clearly, this is not practical and thus a large variety of methods have been developed to tackle this issue, the so-called enhanced sampling techniques.[96] Among them, one can find umbrella sampling, metadynamics or the more recent variational enhanced sampling. In this work we employed metadynamics to enhance the sampling when studying GHs, as it has been successfully used in the past.[88]

Metadynamics is a MD-based technique aimed at enhancing the sampling of the configurational space and at estimating the FES.[97,98] This method is based on the reduction of the phase space dimensionality to a smaller set of collective variables (CVs), which include the essential modes that are associated with the transitions in the analyzed process. CVs can be defined as functions — generally nonlinear — of the atomic coordinates $R: s(R) = (s_1(R), s_2(R), \dots, s_d(R))$. Considering the equilibrium distribution of the CVs as

$$P(s) \int dR\, \delta[s - s(R)]P(R) \qquad \text{Equation 17}$$

where $P(R)$ is the Boltzmann distribution, and its corresponding FES is defined as

$$F(s) = -\frac{1}{\beta} \log P(s) \qquad \text{Equation 18}$$

$$F(s) = -\frac{1}{\beta}\lim_{t\to\infty}\log N(s,t) \qquad \text{Equation 19}$$

where $\beta = (k_B T)^{-1}$ and $N(s,t) = \int_t^0 dt'\,\delta\,[s - s\big(R(t')\big)]/\int_t^0 dt'$ is a normalized histogram accumulated in an unbiased ensemble. In this case the system would remain stuck fluctuating around one particular state. To escape from this local equilibrium state, a bias potential $V(s(R))$ is considered, which depends on $s(R)$ and thus allows to enhance the sampling at the CV space. The dynamics evolution of the system is altered by the addition of this external bias potential to discourage the system from revisiting configurations that have already been sampled. With this, the CVs are distributed according to

$$P_V(s) = \int dR\;\delta[s - s(R)]P_V(R) \qquad \text{Equation 20}$$

where $P_V(s)$ is the Boltzmann distribution of the biased simulation. In the case of metadynamics, the bias potential is defined as small repulsive Gaussian-like terms. At time t

$$V_G(s,t) = \int_0^t dt'\omega\,\exp(-\sum_{i=1}^{d}\frac{(s_i(R) - s_i(R(t')))^2}{2\sigma_i^2}), \qquad \text{Equation 21}$$

where σ_i is the Gaussian width for the *i*th CV and ω an energy rate composed of the Gaussian height W and a deposition stride τ_G

$$\omega = \frac{W}{\tau_G} \qquad \text{Equation 22}$$

Gaussian potentials are added in the regions of the CV space that have already been explored in order to effectively enhance the sampling and visit new states of the FES (Figure 2.4). The negative of the sum of the added potentials corresponds to the underlying FES on the CVs space

$$V_G(s, t\to\infty) = F(s), \qquad \text{Equation 23}$$

It is worth mention that in metadynamics no *a priori* knowledge of the FES is required. However, the selection of the CVs for a certain study is usually the most critical aspect of this technique. The chosen CVs should fulfill the following requirements: (i) they are explicit functions of the atomic positions, (ii) able to distinguish the different metastable states of the system and (iii) include all the slow degrees of freedom. Under those conditions, the FES will contain all the relevant information about the system, such as the location and relative stability of the metastable states, as well as the free energy barriers separating them.[98] Once the CVs are fluctuating in a diffusive regime, convergence is reached. In the case of enzymatic reactions, the re-crossing criteria was applied (i.e. the simulation was stopped when the systems re-crossed the TS returning to the reactants), as recommended for chemical reactions.[99]

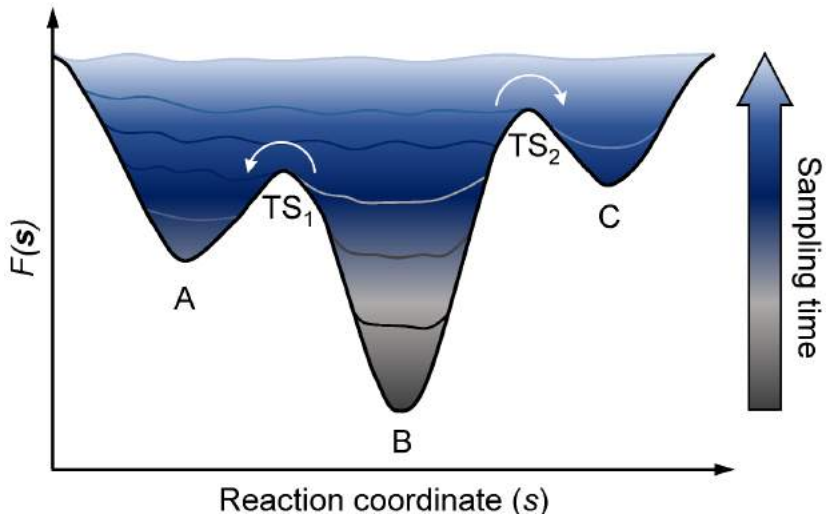

Figure 2.4. Metadynamics calculation example in a one-dimensional model potential. The underlying potential (thick black line) is progressively filled with the Gaussians deposited along the trajectory. The simulation starts from the global minimum B. Afterwards, the system crosses the first transition state (TS_1), being able to visit the local minimum A. The system fluctuates in a diffusive regime between A and B till the second transition state (TS_2) is overcome and all minima are explored. The sum of the underlying potential and of the metadynamics bias is shown according to different times (coloured thin lines).

Extensions and variants of the metadynamics algorithm have been further developed, such as well-tempered metadynamics,[100] bias-exchange metadynamics[101] or the use of multiple-walkers.[102] In this Thesis we employed the PLUMED 2 plugin[103] patched to CPMD software in order to drive both sugar conformational exploration (puckering coordinates as CVs) and enzymatic reactions (distances or distance differences as CVs).

2.5. Molecular Docking

Molecular docking is a computational procedure whose main goal is to efficiently predict the noncovalent binding of small molecules (ligand) in macromolecules (receptor) and estimate its associated energy. As in MM, neither the protonation state of residues nor their atomic charges — thus polarization effects — change during the binding process. Moreover, the receptor is described as rigid, although with the choice of considering few residues as flexible. While a selected set of covalent bonds from the ligand can be rotatable, covalent lengths and angles are constant. Molecular docking generally depends on two components. A search algorithm that drives the conformational search by following the second component, the scoring function, which attempts to approximate the standard chemical potentials of the system. These functions can be physics-based or include empirical parameters, among other possibilities.

In this work we used molecular docking to accommodate missing carbohydrates (ligand) in the initial structure of GHs (receptor) obtained by X-ray crystallography, and thus the binding energies were not evaluated as the main criterium to select a binding pose. Instead, the prior biochemical knowledge in GH catalysis was followed, since our goal was to construct the most reliable GH-carbohydrate complex and the method used here has a standard error of ~ 3 kcal/mol. Moreover, each complex has been further submitted to MD calculations. To perform molecular docking, we employed the graphical interface AutoDockTools[104] to process both the receptor and the ligand; while calculations were carried out using the software Autodock Vina.[105] Vina uses the following scoring function

$$c = \sum_{i<j} f_{ti\,tj}(r_{ij}) \qquad \text{Equation 24}$$

where a type t_i is assigned to each atom i, and a set of interaction functions $f_{ti\ tj}$ of the interatomic distance r_{ij} is defined. Thus, c corresponds to the sum of inter- and intramolecular contributions. In particular, the sum of steric interactions, identical for all atom pairs, hydrophobic interaction between hydrophobic atoms, and, where applicable, hydrogen bonding. Hydrogen atoms are not considered explicitly.

Vina code explores the searching space via an iterated local search global optimizer in order to find the global minimum of c. This algorithm includes mutation and a local optimization, with each step being accepted or rejected in accordance with the Metropolis criterion. The local optimization is performed considering not only the value of the scoring function but also its gradient, which provides a sense of direction from a single evaluation.

2.6. General Procedure

The general approach followed in this Thesis consisted in starting the calculations from a GH-carbohydrate substrate complex obtained either by X-ray crystallography or by Molecular Docking (Autodock Vina program). Then, each system under investigation was equilibrated with classical MD (i.e. at MM level of theory) at 300 K in solution by using the AMBER software. This was evaluated by following the root mean square deviation (RMSD) of the protein backbone, which should remain stable at 1.5-2.5 Å when the system is equilibrated. From a snapshot of the end of the classical MD, we initiated the QM/MM treatment of the system by using the CPMD program. After minimizing the density functional within the CPMD method, we optimized the geometry of the system with annealing of the nuclei until reaching a maximal component of the nuclear gradient of $5 \cdot 10^{-4}$ a.u. Several ps of QM/MM MD were undertaken to equilibrate each system at 300 K. Different fictitious electron mass were tested for each particular system in order to ensure adiabaticity. From the equilibrated system (i.e. RMSD of protein backbone reaching a plateau), we launched a metadynamics calculation to investigate the process of interest, such as enzymatic reactions. We employed PLUMED 2 patched to CPMD software. This procedure is described and discussed in a publication from our group.[91]

3. The reaction mechanism of GH18 chitinases: elucidating the nature of the reaction intermediate

Abstract: Chitinases are glycosidases that are responsible of hydrolysing chitin and have gained great interest because of their industrial and medical applications. GH18 chitinases follow a substrate-assisted mechanism in which the 2-acetamido group of the *N*-acetylglucosamine (GlcNAc) located at the *-1* subsite reacts with the sugar anomeric carbon, forming an intermediate which was commonly described as a glucose oxazolinium ion (Glc-ox$^+$). By means of QM/MM metadynamics simulations on chitinase B from *Serratia marcescens* (*Sm*ChiB), we examined the entire reaction mechanism, showing that the reaction intermediate unequivocally features a neutral glucose oxazoline (Glc-ox) in a $^4C_1/^4H_5$ conformation, with a Glc-ox$^+$ being formed only on the pathway toward the reaction products. The protonation events that orchestrate catalysis take place by virtue of the *Sm*ChiB active site architecture, where the sugar and the catalytic residues (Asp142 and Glu144) form a well-defined hydrogen bond network. These observations can be extended to other glycosidases that follow the substrate-assisted mechanism and exhibit an insertion in the sequence between the catalytic residues, as is shown here for a hyaluronidase from family GH56.

Publication: J. Coines, M. Alfonso-Prieto, X. Biarnés, A. Planas and C. Rovira. Oxazoline or Oxazolinium Ion? The Protonation State and Conformation of the Reaction Intermediate of Chitinase Enzymes Revisited. *Chem. Eur. J.*, 2018, 24, 19258–19265

3.1. Introduction

Chitin is a crystalline homopolymer composed of *N*-acetyl-D-glucosamine (GlcNAc) units (Figure 3.1a). Despite being the second most abundant polysaccharide in Nature after cellulose, it does not accumulate in the environment because of the action of chitinases, which act in synergy with other enzymes such as chitin deacetylases or lytic polysaccharide monooxygenases to fully decompose chitin. Chitinases (EC 3.2.1.14) are glycosidases that catalyse the hydrolysis of β-1,4 linkages between GlcNAc units of chitin. Bacteria, fungi, insects, plants and animals possess these hydrolytic enzymes, carrying out several functions such as the exoskeleton remodelling in crustaceans and insects, or participating in the immunological response against chitin-containing pathogens in humans.[106]

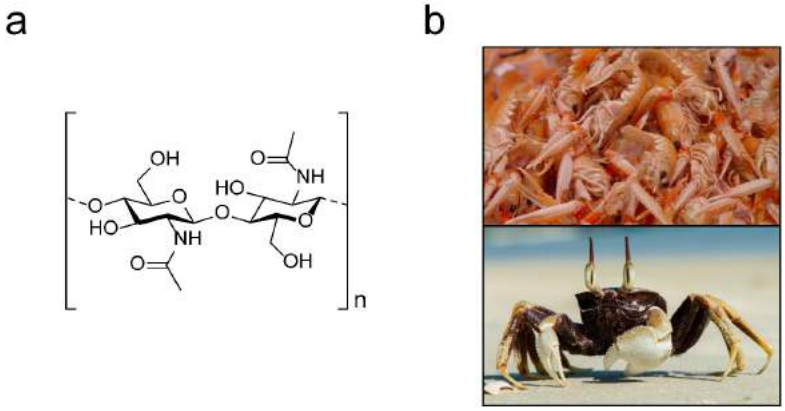

Figure 3.1. (a) Chitin chemical structure, composed of GlcNAc bound through β-1,4 linkages. (b) Natural sources of chitin. Norway lobster (*Nephrops norvegicus*, top) and horned ghost crab (*Ocypode ceratophthalma*, bottom).

Because of the potential applications of chitinases in industry and medicine, research efforts have been devoted towards comprehending and employing them. From an industrial point of view, they can be used to convert chitin-containing biomass from marine waste into high-quality products (chitosans and chitooligosaccharides) through processes with low environmental impacts.[107,108] These bioactive molecules exhibit diverse medical, pharmaceutical, cosmetics, and food applications.[109-111] In addition, human chitinases — and their structurally related but catalytically inactive, chitolectins — are involved in bacterial infections and inflammatory processes such as asthma and cancer, and thus are potential pharmaceutical targets.[112,113] Moreover, chitinase inhibitors are also being investigated for their use as insecticides, fungicides and antimalarial drugs.[114-116]

Chitinases have been classified in GH18 and GH19 families. Whereas GH19 chitinases are inverting enzymes, GH18 chitinases cleave their substrates with retention of the anomeric configuration, following a substrate-assisted mechanism (Figure 3.2). Here we focused on family GH18 enzymes and, specifically, chitinase B from *Serratia marcescens* (*Sm*ChiB). This bacterium produces a battery of enzymes that allow efficient chitin degradation.[117] Among them, *Sm*ChiB has *exo* activity and cleaves the nonreducing ends of chitin chains with processivity, due to the tunnel topology of the active site and the interactions provided by numerous aromatic residues. Moreover, it can degrade chitooligosaccharides of three, four, five and six units of GlcNAc, producing chitobiase.[118] *Sm*ChiB possesses three domains (Figure 3.3): a GH18 catalytic domain (residues 16–425), a linker sequence (residues 426–451) and a carbohydrate binding module CBM5 (residues 451–499).[119] The active site comprises six subsites (*-3, -2, -1, +1, +2* and *+3*), and contains the Asp-X-Asp-X-Glu catalytic motif, which corresponds to Asp140, Asp142 and Glu144 in *Sm*ChiB.

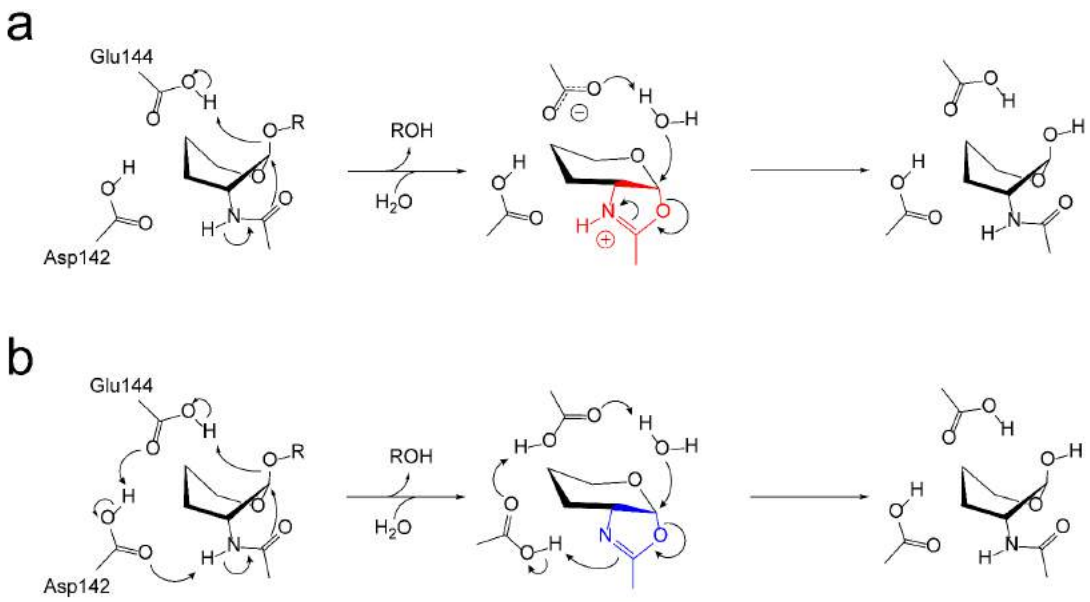

Figure 3.2. The two-steps substrate-assisted reaction mechanism of *Sm*ChiB considering a (a) Glc-ox$^+$ (oxazolinium ion in red) reaction intermediate, as previously assumed; or (b) a neutral Glc-ox (oxazoline in blue), as obtained in this work. From left to right: Michaelis complex, reaction intermediate and products.

While Glu144 acts as the acid/base residue during the substrate-assisted catalysis performed by *Sm*ChiB, Asp142 (assisting residue) interacts with the nitrogen atom of the 2-acetamido group of GlcNAc at *-1* subsite (Figure 3.2). On one hand, Asp142 promotes the nucleophilic attack in the first step of the enzymatic reaction and, on the other, it stabilizes the reaction

intermediate, either by acting as a general base subtracting the proton of the acetamido group and forming a Glc-ox intermediate, or by electrostatically stabilising the Glc-ox$^+$. Although the role of most active site residues has been determined by structural,[52,58,120-122] kinetic[51,120] and computational studies[51,123], the nature of such reaction intermediate remains still controversial.

Considering that the assisting and the acid/base residues of *Sm*ChiB are linked by a hydrogen bond, as inferred from the observed distance between oxygen atoms from their carboxylate groups in several X-ray structures, it is expected that they share a proton.[58] This interaction would increase the pK_a of the assisting residue with respect to the situation when both residues are not linked or they are in solution,[124] and thus favour a Glc-ox intermediate instead. However, it was surprising to discover that all experimental and theoretical studies to date describe the reaction intermediate as Glc-ox$^+$ (Figure 3.2a).[58,65,117,125,126] While X-ray crystallography cannot discern both intermediates, previous theoretical studies of *Sm*ChiB excluded a Glc-ox intermediate because of its high energy barrier upon hydrolysis. However, the constructed model that would evolve through a Glc-ox intermediate lacks one hydrogen atom between Asp142 and Glu144, neglecting the crucial hydrogen bond that these residues form.[65]

Here the entire reaction mechanism of *Sm*ChiB has been investigated, paying special attention to the nature of its reaction intermediate and considering the results obtained in previous studies. Our results demonstrate that, unlike previously assumed, the reaction intermediate of *Sm*ChiB contains a neutral Glc-ox, which evolves towards reaction products using an effective proton shuttle mechanism through the two catalytic residues (Figure 3.2b). These results can probably be extended to other glycosidases that follow a substrate-assisted mechanism and exhibit a similar hydrogen bond network. This happens in similar GHs having an insertion between the catalytic residues in their sequence, as will be shown here for hyaluronidases from family GH56. These results expand the understanding of GH18 enzymes, which has also served as starting point of novel mechanistic studies of this family, such as chitinase A from *Serratia marcescens* (*Sm*ChiA).[127]

3.2. Results and discussion

3.2.1. Modeling the reaction intermediate of *Sm*ChiB

The calculations were started from the coordinates of the crystal structure of *Sm*ChiB trapped in the reaction intermediate state (PDB entry 1E6Z).[58] This structure contains two conformations for the Asp142 residue with an occupancy of 0.50 each. One of these is not interacting with the oxazoline/oxazolinium ion, and thus was discarded for modeling the reaction intermediate. In the other configuration, Asp142 interacts with the Glc-ox/ox$^+$ and thus was considered as reactive. This configuration was also observed in the crystal structure of the Michaelis complex (PDB entry 1E6N) and consistent with previous mechanistic studies.[65,126] The reaction intermediate was reconstructed by replacing the trapped monosaccharide Glc-ox ligand with a diacetylchitobiose-oxazoline (GlcNAc-Glc-ox), occupying the *-1* and *-2* subsites of *Sm*ChiB (see computational details).

Two models were generated, considering either a GlcNAc-Glc-ox or a GlcNAc-Glc-ox$^+$ at the *-1* and *-2* subsites (Figure 3.3). The protonation of the assisting residue (Asp142) and the catalytic acid/base (Glu144) was set according to the most likely hydrogen-bond interactions formed by these residues, consistently with the crystallographic structure. Therefore, both models contain a hydrogen between the oxazoline nitrogen and Asp142, and a second proton between both catalytic residues, which maintain their orientation required for a productive reaction. Indeed, the latter hydrogen was the one neglected in previous theoretical studies, where Asp142 was ionized and the oxazoline was excluded as reaction intermediate.[65] Each system was equilibrated at 300 K by 45 ns of classical MD, till the RMSD of the protein backbone reached a plateau. All the protein-ligand interactions remained stable, except the hydrogen bond between Tyr214 and the oxygen from the oxazoline ring in the Glc-ox$^+$ simulation, which was disrupted after 6 ns. However, this interaction was recovered after 39 ns, showing that Tyr214 exhibits two alternative conformations, as was also observed in previous studies.[65] Interestingly, a water molecule was found to sit above the anomeric carbon in both models along the simulation, forming a hydrogen bond with Glu144, and thus being well oriented for nucleophilic attack (Figure 3.3).

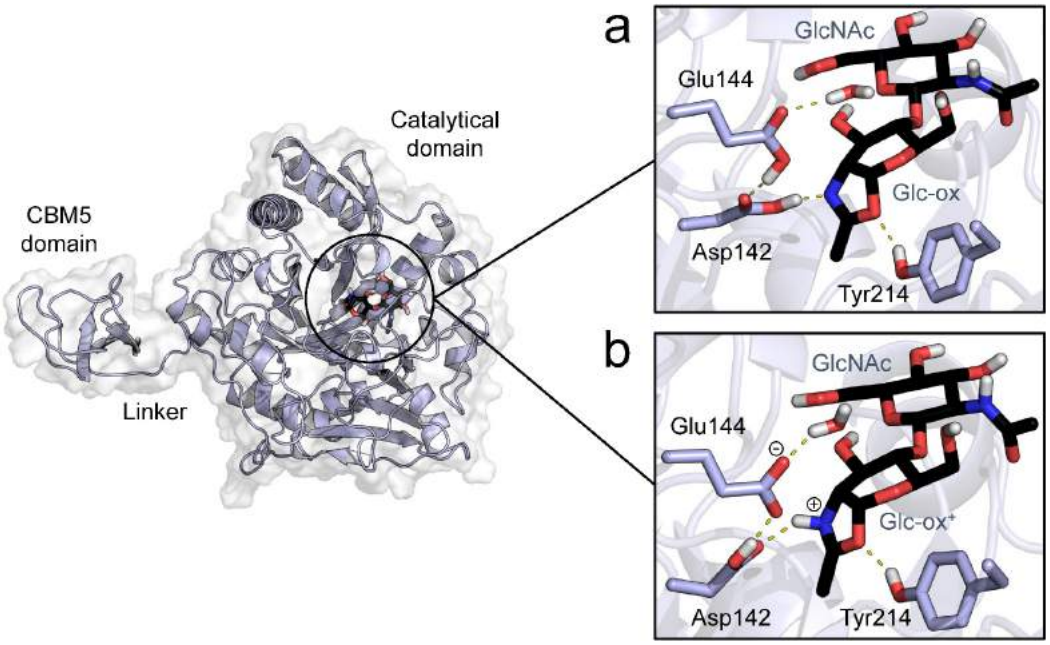

Figure 3.3. Structure of the reaction intermediate of *Sm*ChiB obtained from classical MD simulations considering either (a) a neutral oxazoline (GlcNAc-Glc-ox) or (b) an oxazolinium ion (GlcNAc-Glc-ox$^+$). Hydrogen bonds are shown as dashed yellow lines. Non polar hydrogens are omitted for the sake of clarity.

Afterwards, QM/MM structure optimizations were performed, including in the quantum region the disaccharide, the side chains of the catalytic residues and the nucleophilic water. The Glc-ox model (Figure 3.3a) was found to be stable, exhibiting a conformation of the pyranose ring between $^4C_1$ and $^4H_5$. On the contrary, the Glc-ox$^+$ model (Figure 3.3b) was unstable: two proton transfer events took place, one proton from the oxazoline nitrogen (towards Asp142) and a second from Asp142 (towards Glu144), evolving to the same configuration as the Glc-ox model. This state was further maintained during the subsequent QM/MM MD simulations (Figure S3.1). Calculations using alternative DFT functionals (Figure S3.2), as well as simulations starting from a restrained oxazolinium ion state (Figure S3.3), further confirmed these results.

To verify the above results, we computed the free energy difference between oxazoline and oxazolinium ion states by means of QM/MM metadynamics, starting from the equilibrated Glc-ox model (Figure 3.4). The FES shows that Glc-ox$^+$ is not a stable species in the active site of *Sm*ChiB, being ~ 5 kcal/mol higher in energy than Glc-ox. Altogether, these results show that the reaction intermediate of *Sm*ChiB contains a neutral Glc-ox at the *-1* subsite, rather than a Glc-ox$^+$.

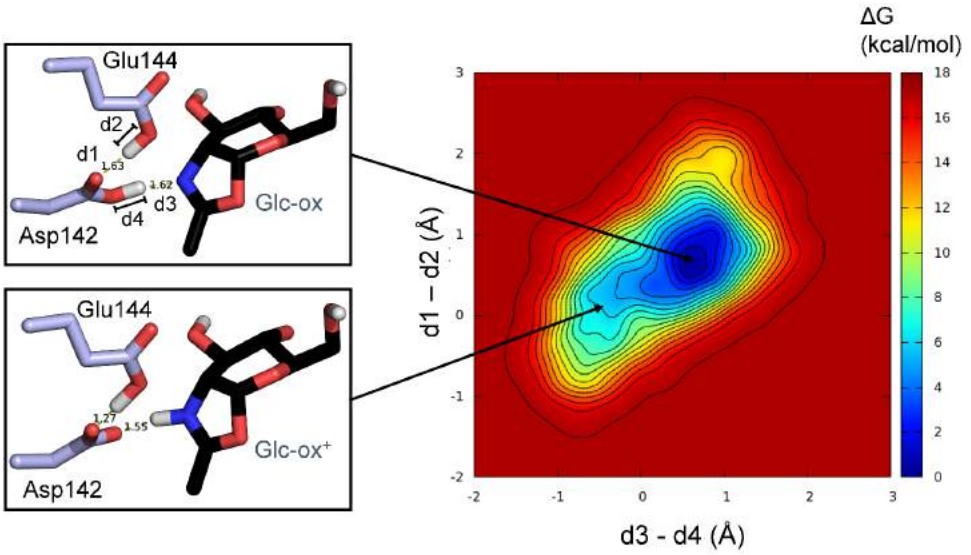

Figure 3.4. Reconstructed free energy surface (FES) from the metadynamics simulation using as collective variables the combinations of distances involving the oxazoline proton and the proton shared between Asp142 and Glu144. Contour lines are separated by 1 kcal/mol. Only polar hydrogen atoms are shown for clarity. Contour lines are at 1 kcal/mol.

After publication of our results,[128] a QM(DFT)/MM metadynamics investigation of the reaction mechanism of chitinase A from *Serratia marcescens* (*Sm*ChiA)[127] appeared in which the authors conclude that both Glc-ox and Gl-ox$^+$ reaction intermediates are isoenergetic and interconvertible at room temperature. This is surprising because *Sm*ChiA and *Sm*ChiB have a very similar active site: a catalytic DXDXE motif and a Tyr residue that interacts with the carbonyl oxygen of the substrate (Tyr390 in *Sm*ChiA equivalent to Tyr214 in *Sm*ChiB). The authors argued that the different protonation state of the reaction intermediate between *Sm*ChiB and *Sm*ChiA may arise from the different pK_a of their respective assisting residues (Asp142 and Asp313 in *Sm*ChiB and *Sm*ChiA, respectively). Indeed this was investigated in another theoretical study by others on the complexes of both enzymes with allosamidin, a chitinase inhibitor that mimics the reaction intermediate.[129] The authors reported that Asp313 in *Sm*ChiB exhibits a pK_a of 16.0, whereas Asp142 (the corresponding assisting residue) in *Sm*ChiA has a much lower pK_a (10.3), suggesting that the substrate of the former is more likely to maintain its initial proton (Figure 3.2a). These high pK_a values for an Asp are likely caused by the second Asp in the DXDXE motif, which has a very low pK_a in both enzymes (< 0.0). Therefore, although it is clear that the reaction intermediate of *Sm*ChiB features a neutral oxazoline (Glc-ox), it seems that small changes in the delicate hydrogen bond network that connects the substrate and both catalytic residues might modify the pK_a of the assisting residue and ultimately the final energetic differences of both possible protonation states of the substrate.

3.2.2. Conformational free energy landscape of the Glc-ox intermediate

The conformation of the reaction intermediate is an aspect of interest in GH mechanistic studies, since it gives idea of the catalytic itinerary of the *-1* sugar during catalysis, which is a fingerprint of GH family. Thus, we investigated conformation of the pyranose ring of Glc-ox at *-1* subsite of *Sm*ChiB by QM/MM metadynamics and using the Cartesian puckering coordinates q_x and q_y as collective variables (CVs).[38]

The simulations show that the Glc-ox reaction intermediate remains unprotonated along most of the metadynamics trajectory, showing that its protonation state does not depend on its ring conformation. The computed FES displays two minima: a global minimum that corresponds to a $^4C_1/^4H_5$ conformation; and a secondary minimum whose conformation is $^1S_3$ and 3.4 kcal/mol higher in energy than the former (Figure 3.5).

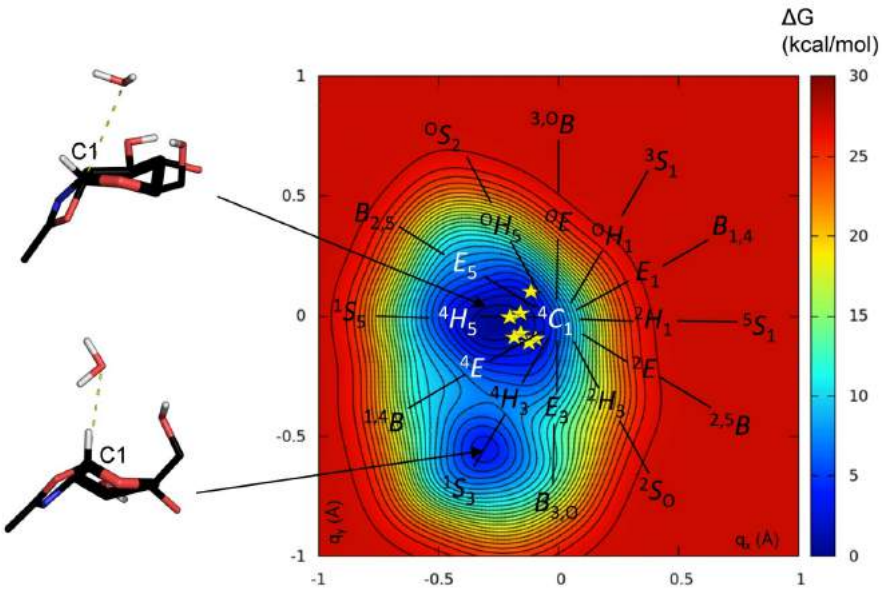

Figure 3.5. Conformational free energy surface (FES) of the Glc-ox pyranose ring at the *-1* subsite in the reaction intermediate of *Sm*ChiB. Contour lines are at 1 kcal/mol. The FES has been annotated (yellow stars) with the conformations observed in X-ray structures of reaction intermediates (oxazolines or their analogues) bound to GHs that follow a substrate-assisted reaction mechanism (Table S3.1).

The most stable conformation, $^4C_1/^4H_5$, is consistent with the observed conformation in our QM/MM MD results and also the calculations from the recent study of *Sm*ChiA, computed at DFT level of theory as well.[127] This is in agreement with the conformations observed in the X-ray structures of oxazolines — or their analogues — bound to other GHs (yellow stars

in Figure 3.5), which lie on the global minimum of the conformational FES. It is worth to mention that previous theoretical studies of *Sm*ChiB using semiempirical methods reported that a $^4C_1$ or $B_{3,O}$ conformation for the (Glc-ox$^+$ in this case) reaction intermediate.[65,126] However, it is well-known that semiempirical methods do not describe satisfactorily sugar conformations, whereas the PBE functional used here provides very good results.[89]

The $^1S_3$ secondary minimum in the conformational FES is not expected to be catalytically competent because it is less stable than $^4C_1/^4H_5$ and exhibits an axial C1···H1, which hinders the nucleophilic attack of the catalytic water. In contrast, $^4C_1/^4H_5$ conformation can accommodate the water molecule in a suitable orientation for nucleophilic attack. Thus, the reactivity of the Glc-ox reaction intermediate in *Sm*ChiB was analysed using the $^4C_1/^4H_5$ conformation as starting one.

3.2.3. Reaction mechanism of *Sm*ChiB: deglycosylation

To elucidate the catalytic mechanism and further understand the occurrence of a neutral Glc-ox as a reaction intermediate, the hydrolysis of the GlcNAc-Glc-ox (deglycosylation) has been investigated by QM/MM metadynamics. Two CVs were employed to study the reaction. CV_1 was defined as the distance between the anomeric carbon (C1) and the oxygen of the nucleophilic water (CV_1=C1···O_w), and thus accounts for the water nucleophilic attack. CV_2 was defined as the distance between the oxazoline nitrogen atom and the hydrogen atom of the Asp142 (CV_2=N_x···H_x). Therefore, this variable discriminates between Glc-ox and Glc-ox$^+$. The reconstructed FES displays two minima that correspond to the initial GlcNAc-Glc-ox intermediate (**I**), and the products of the reaction (**P**, 5.6 kcal/mol lower in energy than **I**), in which a $(GlcNAc)_2$ disaccharide is formed (Figure 3.6).

The reaction initiates with the approach of the catalytic water to C1. Simultaneously, N_x is protonated by Asp142, and thus the sugar adopts the Glc-ox$^+$ configuration, as evidenced in the change in CV_2 (N_x···H_x distance, from 1.61±0.04 Å at **I** to 1.15±0.03 Å at **I'**, Table S.3.2). The Glc-ox$^+$ configuration (**I'**) corresponds to a shoulder on the reaction pathway, being about 2 kcal/mol above the initial Glc-ox state (**I**). Only when the water molecule is close enough to C1 (3.08±0.05 Å), Glc-ox$^+$ can be formed, indicating that the water approach triggers Glc-ox protonation on the pathway towards the **TS**. Although there is a small difference in free energy between Glc-ox and Glc-ox$^+$, it is important to remark that Glc-ox$^+$ is not a minimum along the reaction pathway and thus is not a stable species. This is consistent with the previous equilibrium QM/MM MD calculations, where the lifetime of Glc-ox$^+$ was almost null.

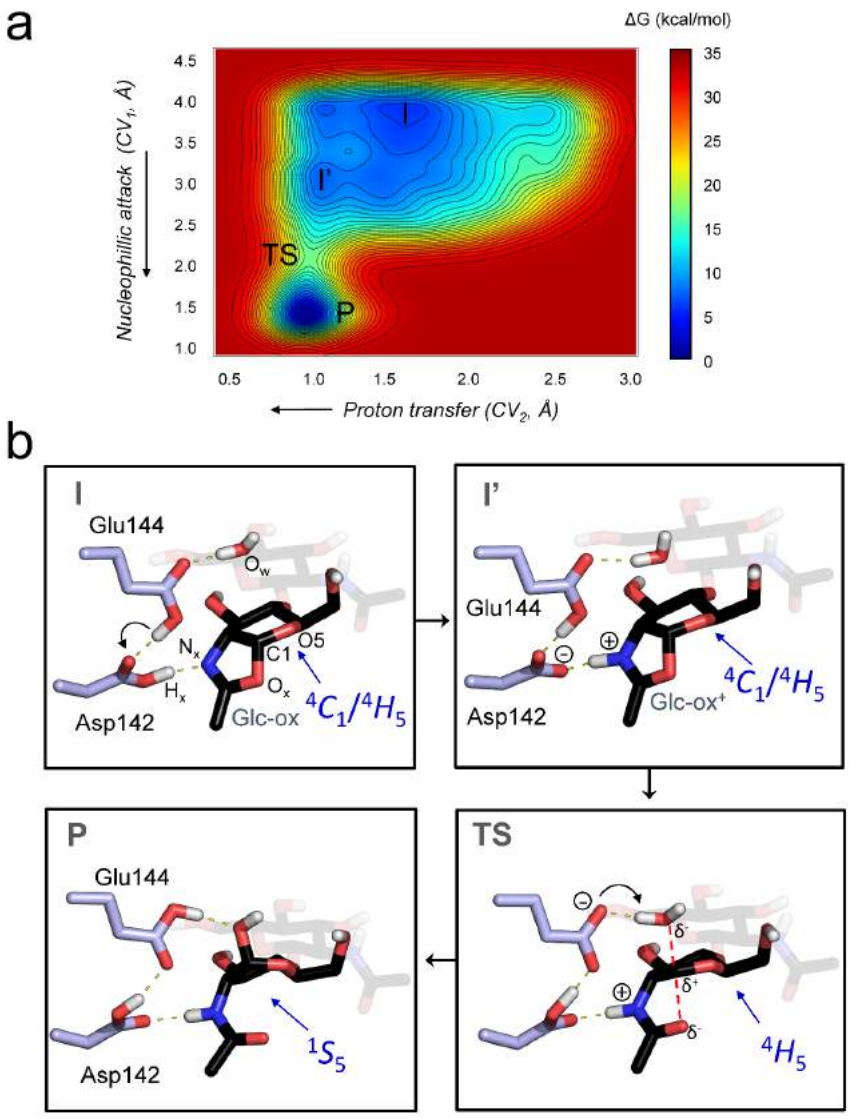

Figure 3.6. Deglycosylation step of SmChiB. (a) Free energy surface (FES) of the hydrolysis reaction of the oxazoline intermediate in *Sm*ChiB. Isolines at 1 kcal/mol. (b) Average structures corresponding to relevant configurations along the reaction coordinate (I=Glc-ox; I'=Glc-ox$^+$, TS=transition state, P=reaction products). The GlcNAc unit at the *-2* subsite is shown as transparent for clarity. Bonds being formed/broken are indicated with a dashed red line and relevant hydrogen bonds are displayed as dashed yellow lines. Only polar hydrogen atoms are shown for the sake of clarity.

The reaction **TS** is reached when the catalytic water further approaches C1. Similar bond distances for C1···O$_w$ and C1···O$_x$ are observed at this point (2.11±0.02 Å and 2.23±0.21 Å, respectively), indicating that these bonds are concertedly being formed and broken, respectively. The computed free energy barrier for the reaction (11.6 kcal/mol) is in agreement with experimental kinetic data (k_{cat}=28 s^{-1}; $\Delta G^{\ddagger}$=15.6 kcal/mol).[130] Since glycosylation was expected to be the rate-limiting step of the reaction mechanism, the free

energy barrier for deglycosylation (first step) should be lower than the experimentally measured rate. Finally, the acid/base residue (Glu144) fully deprotonates the water molecule, which forms a covalent bond with C1, producing the new $(GlcNAc)_2$ (**P**). The reaction products are further stabilized by the strengthening of the interaction between the oxygen from the GlcNAc carbonyl group (O_x) and the hydroxyl group of Tyr214 (from 2.00±0.15 Å at **I** to 1.88±0.07 Å at **P**). This supports a stabilizing role for this residue, as discussed in previous studies.[51,65]

The conformations of the Glc-ox pyranose ring at the *-1* subsite follows a $^4C_1/^4H_5 \rightarrow [^4H_5]^{\ddagger} \rightarrow {}^1S_5$ itinerary during hydrolysis (Figure S3.4). Surprisingly, the $^4H_5$ conformation at the **TS** does not correspond to any of the canonical conformations that fulfill the electronic and structural requirements for an oxocarbenium ion,[35] even though it is not far from ^4E, which does. This is probably caused by the structural constraints imposed by the oxazoline five-membered ring, which limits the torsion around the C1···C2 bond and thus the conformations available. As expected in GHs, the **TS** exhibits a clear oxocarbenium ion-like character: the dihedral angle formed by atoms C1, C2, C5 and O5 is close to planarity (13.5±9.6 degrees, Table S3.3), the C1···O5 distance shrinks with respect to the reactants state (from 1.38±0.03 Å to 1.32±0.02 Å) and the anomeric charge increases by 0.16e (Table S3.4).

Interestingly, a dynamic hydrogen-bond network involving the oxazoline, Asp142 and Glu144 is revealed during the enzymatic reaction. This well-defined network assists catalysis by means of three consecutive proton-transfer events, which are highly dependent among them and are initiated only when the water molecule approaches to C1. Specifically, a proton is transferred from Asp142 to the N_x (at **I'**), from Glu144 to Asp142 (at **TS**), and finally from the catalytic water to Glu144 (at **P**).

3.2.4. Reaction mechanism of *Sm*ChiB: glycosylation

To further explore the mechanism by which the enzyme employs this hydrogen-bond network in catalysis, we also investigated the first reaction step (glycosylation). To this aim, we first modelled the enzyme Michaelis complex, where a $(GlcNAc)_5$ is bound to *-2*, *-1*, *+1*, *+2* and *+3* subsites of *Sm*ChiB (starting from structure PDB entry 1E6N) with classical MD (60 ns), followed by QM/MM metadynamics simulations of the first step of the enzymatic reaction, which leads to the reaction intermediate (I). In this case, the two CVs used to drive the reaction assessed both the nucleophilic attack of the carbonyl group and the protonation of the glycosidic bond (see computational details). Therefore, none of these CV included any

proton transfer between the carbonyl group from the *-1* GlcNAc, Asp142 and Glu144. This allowed us to evaluate if the proton shuttle observed in the deglycosylation reaction would occur here without enhancing its sampling.

The results show that the reaction intermediate formed upon substrate-assisted glycosidic bond cleavage is again a neutral oxazoline in a $^{4}C_1/^{4}H_5$ conformation. The three proton transfer events previously observed when modeling the second reaction step were observed also here in a reverse manner (Figure 3.7). In particular, first Glu144 protonates the glycosidic bond, which at the same time is being disrupted). After crossing the **TS**, Asp142 transfers its proton to Glu144 to act as a base deprotonating the nitrogen from the acetamido group of the GlcNAc at *-1* subsite. With that, the Glc-ox reaction intermediate is formed (**I**).

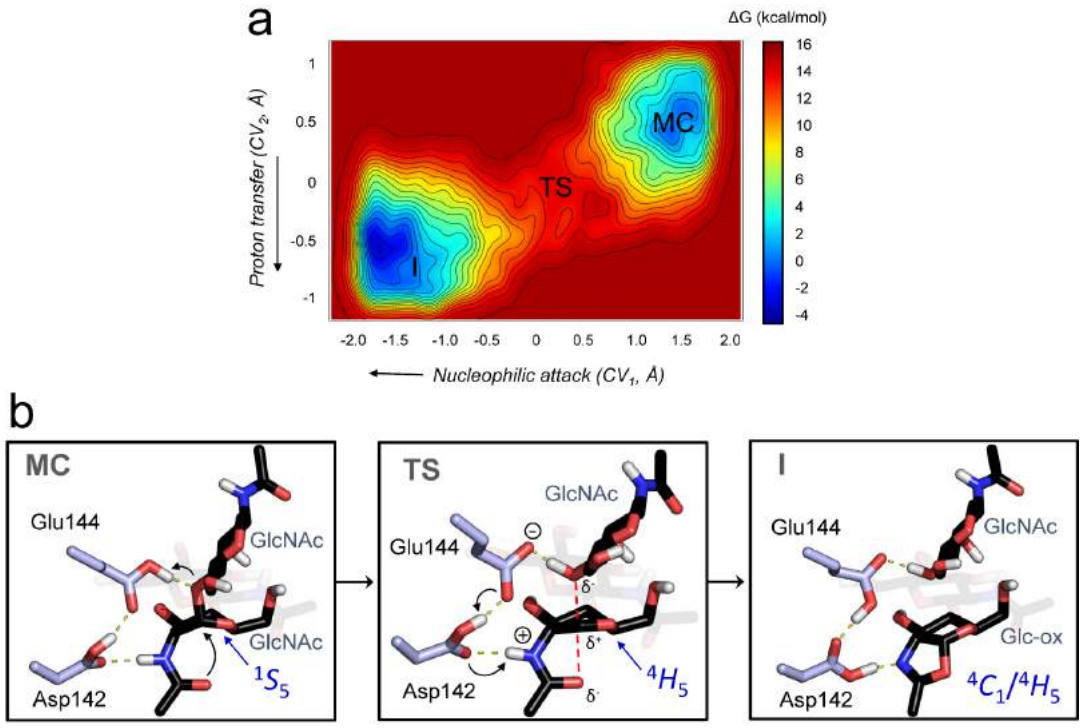

Figure 3.7. Glycosylation step of *Sm*ChiB. (a) FES reconstructed from the metadynamics simulation of the glycosylation reaction. Isolines at 1 kcal/mol. (b) Average structures corresponding to relevant configurations along the reaction coordinate (MC= Michaelis complex, TS=transition state, I= Glc-ox). The GlcNAc unit at the *-2* subsite is shown as transparent for clarity. Likewise, GlcNAc units at the *+2* and *+3* subsites are not shown. Hydrogen atoms have been also omitted, except the ones attached to heteroatoms. Bonds being broken/formed are indicated with a dashed red line.

The free energy barrier obtained for the first reaction step was 13.7 kcal/mol, in agreement with the value estimated from the experimental reaction rate (k_{cat}=28 s^{-1}; $\Delta G^{\ddagger}$=15.6 kcal/mol). On the contrary, in the recent theoretical investigation of *Sm*ChiA[127] the obtained energy barrier (9.3 kcal/mol) was considerably lower than the corresponding estimated value (k_{cat}=33 s^{-1}; $\Delta G^{\ddagger}$=15.5 kcal/mol).

In summary, our results indicate that Glc-ox^{+} is not stable in *Sm*ChiB — unlike previously assumed — and the reaction intermediate should be described as Glc-ox. We also conclude that the conformation of the sugar ring at the reaction intermediate lies between $^{4}C_{1}$ and $^{4}H_{5}$, in agreement with structures observed in experimental studies, but in contrast to the previously reported in previous studies using semiempirical methods.[126] The complete mechanism of catalysis obtained in this work exemplifies how reversible proton transfer events can assist an enzymatic reaction by either stabilizing the intermediate (GlcNAc deprotonation) or activating it (GlcNAc protonation) for efficient catalysis. Studies in other GHs that follow a substrate-assisted mechanism are in progress in our group to assess whether the conclusions obtained for *Sm*ChiB can be extended to other type of active sites.

3.2.5. Family GH56 hyaluronidases

One might think that the proton shuttle previously discussed in *Sm*ChiB catalysis (and observed in *Sm*ChiA by others) can be extended to other glycosidases that also follow a substrate-assisted mechanism. In order to investigate this, we studied the hyaluronidase from bee venom (*Bv*Hya) that belongs to family GH56 (EC 3.2.1.35).[61] This enzyme hydrolyzes hyaluronic acid, a polymer found in several tissues such as human skin, that is composed of D-glucuronic acid (GlcA) and GlcNAc linked via alternative β-1,4 and β-1,3 glycosidic bonds. Hya has in its active site Asp111 and Glu113 as assisting and acid/base residues, respectively (Figure 3.8a), analogous to Asp142 and Glu144 in *Sm*ChiB. *By*Hya also features an active site tyrosine, Tyr227 (Figure 3.8c), which interacts with the carbonyl oxygen of the *-1* GlcNAc substrate in the same manner as Tyr214 in *Sm*ChiB.

We modelled the reaction products of Hya starting from the X-ray structure of a GH56 enzyme (PDB entry 1FCV),[61] solved with a hyaluronic acid tetramer (GlcA-GlcNAc)$_2$ bound at *-4*, *-3*, *-2* and *-1* subsites (Figure 3.8a). This was the only structure of a GH56 enzyme with a carbohydrate substrate. When analyzing it, we found that the equatorial orientation of the O1 atom of the hydroxyl group of the GlcNAc at the *-1* subsite was incorrectly assigned. Taking into account the reported $^{1,4}B$ conformation of the *-1* ring and that all glycosidic bonds

in hyaluronic acid are β, this hydroxyl group should exhibit an axial orientation. Therefore, we corrected this in our model.

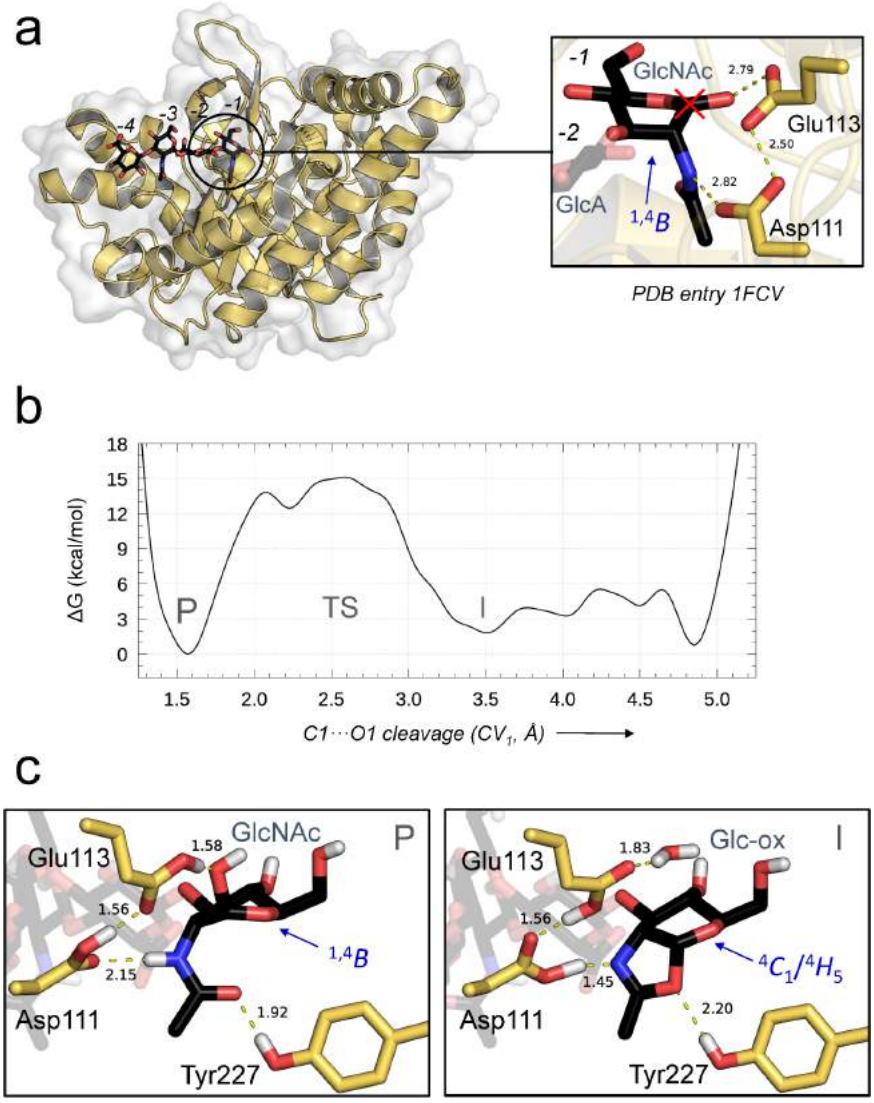

Figure 3.8. Bee venom Hyaluronidase from family GH56. (a) Overview of the X-ray structure (PDB entry 1FCV) of a hyaluronic acid tetramer bound to Hya (left) and its *-1* subsite zoomed (right). The red cross indicates the wrongly assessed stereochemistry. (b) Free energy profile reconstructed from the metadynamics simulation of the reverse deglycosylation reaction. (c) Average structures corresponding to the products and intermediate of the reaction (P=products, I= Glc-ox). The GlcA and GlcNAc units at the *-2*, *-3* and *-4* subsites are shown as transparent for the sake of clarity. Nonpolar hydrogens have been omitted.

We equilibrated the hyaluronic acid tetramer complexed to Hya with 75 ns of classical MD, considering the correct β configuration of the substrate and both Asp111 and Glu113 protonated. After QM/MM MD equilibration (2.5 ps), we analyzed the reverse second reaction step by using metadynamics and the distance between the anomeric carbon (C1) and the oxygen of its hydroxyl group (O1) as CV (CV_1=C1···O1), equivalent to C1···O_w distance in the second reaction step in *Sm*ChiB. Therefore, starting from the reaction products (GlcNAc), we obtained the reaction intermediate in its neutral form (Glx-ox) featuring a $^4C_1/^4H_5$ conformation, as in *Sm*ChiB (Figure 3.8b).

Even though the reaction was driven by only one CV that does not account for any of the possible proton transfer events, all occurred analogously to the reverse deglycosylation reaction in *Sm*ChiB (Figure 3.8c). Moreover, this calculation reinforces the idea of that the approach of the catalytic water — or here the disruption of the C1···O1 bond — triggers the proton shuttle between the water molecule, Glu113, Asp111 and the oxazoline ring. The maximum energy value along the free energy profile is located at CV_1=2.58 Å. However, there is another saddle point 1.3 kcal/mol lower in energy at CV_1=2.07 Å, resembling the TS in the two-dimensional FES in *Sm*ChiB (C1···O_w = 2.11 Å, Figure 3.6). The TS is not unambiguously identified because of the use of just one CV. The free energy barrier may vary between 12-14 kcal/mol, which is in the same scale as the second reaction in *Sm*ChiB. We used only one CV in this calculation because the main objective was to reconstruct and characterize the reaction intermediate.

In conclusion, we show that the proton shuttle machinery present in *Sm*ChiB that enables the occurrence of a neutral oxazoline (Glc-ox) as a reaction intermediate is transferable to other glycosidases that have a Asp-X-Glu catalytic machinery, not only within GH18 family but also GH56.

3.3. Conclusions

- The reaction intermediate in *Sm*ChiB features a neutral Glc-ox, in contradiction with what was previously reported in the literature (a positively charged Glc-ox$^+$).

- The rate-limiting step of the overall reaction catalysed by *Sm*ChiB is the glycosylation step. Both glycosylation and deglycosylation reactions are influenced by a well-defined hydrogen bond network that connects the nucleophilic water with the oxazoline nitrogen through the two catalytic residues (Asp142 and Glu144), allowing the occurrence of a Glc-ox$^+$ only *on the pathway* of the glycosylation/deglycosylation reactions.

- The catalytic features observed in *Sm*ChiB, a member of family 18, are transferable to other GH families which also have an Asp-X-Glu catalytic motif. Indeed, we obtained a neutral Glc-ox reaction intermediate in *Bv*Hya, a member of family 56.

3.4. Computational details

System preparation

To model the reaction mechanism of in *Sm*ChiB, the crystal structure of the reaction intermediate and the Michaelis complex were used as initial structures (PDB entries 1E6Z and 1E6N, respectively).[58] In the case of the reaction intermediate, the X-ray structure of *Sm*ChiB contains a putatively trapped Glc-oxazoline/oxazolinium molecule at the *-1* subsite. However, the electron density of the carbohydrate is very poor and the sugar at *-2* subsite is missing. Thus, we introduced in the *-1* and *-2* subsites the disaccharide GlcNAc-Glc-ox obtained from another crystal structure (PDB entry 4HMD)[125], by means of molecular docking using Autodock Vina.[105] The resulting binding pose overlapped the Glc-ox moiety of the disaccharide ligand with the Glc-ox molecule trapped by crystallography. The assisting residue Asp142 exhibits two conformations with 0.5 occupancy each in the reaction intermediate, but was modelled considering only the catalytically competent one, in which is interacting with both the substrate and the acid/base residue Glu144. This conformation is consistent with all the theoretical and experimental information available. The initial structure of *Sm*ChiB in the Michaelis complex was taken from the crystal structure of a $(GlcNAc)_5$ bound at *-3*, *-2*, *-1*, *+1* and *+2* subsites. The initial structure of Hya was taken from the $(GlcA\text{-}GlcNAc)_2$ trapped in its *-4*, *-3*, *-2* and *-1* subsites, in the products of the reaction state (PDB entry 1FCV)[61]. The equatorial configuration of O1 atom in the GlcNAc at *-1* subsite was reverted to axial, in order to model a β-1,3 linked GlcNAc in $^{1,4}B$ conformation.

In *Sm*ChiB, all acidic residues were considered deprotonated, except Asp142 and Glu144. In the Michaelis complex, both residues were protonated (Figure 3.2). In the case of the reaction intermediate, two scenarios were considered: either the substrate featuring a Glc-ox$^+$ and Asp142 protonated at the O atom closest to Glu144; or both Asp142 and Glu144 protonated and the substrate described as a Glc-ox (Figure 3.3.). In both cases, a proton is shared between the catalytic residues, as inferred from the crystallographic studies in which these residues are oriented and located to interact. Histidine protonation was performed according their chemical environment. Three sodium cations were added to neutralize the total charge of the *Sm*ChiB system. In Hya, both assisting and acid/base residues (Asp111 and Glu113) were protonated to maintain the expected hydrogen bond network for the mechanism and according to the crystal structure. All the other acidic residues were considered deprotonated. Histidine residues were protonated according their chemical environment. Six chlorine ions were added to neutralize the total charge of Hya system. All crystallographic water molecules were kept, and extra ones were added to form a 15 Å water box around the protein surface in each system.

This led to a total number of 86579, 92422 and 64821 atoms for the reaction intermediates in *Sm*ChiB, its Michalis complex and Hya, respectively.

Classical molecular dynamics

AMBER 11 software[131] was used to run classical MD calculations. Protein residues were described employing the ff99SB force field.[132] The TIP3P force field[80] was used to model water molecules. While GlcNAc and GlcA carbohydrates were described with GLYCAM06,[78] the reaction intermediate in *Sm*ChiB required a parametrization of the Glc-ox and Glc-ox$^+$ linked by a β-1,4 bond. This was performed using antechamber module,[133] considering GLYCAM06 and GAFF[134] parameters. RESP atomic charges were obtained from first principle calculations using Gaussian09[135] at HF/6-31G* level of theory. All MD simulations were run under periodic boundary conditions employing the following protocol. First, all ions and water molecules were relaxed by 20,000 cycles of energy minimization, while the protein and substrate were fixed. Afterwards, the entire system was relaxed with 20,000 minimization cycles more. Subsequently, the system was gradually heated in the NVT ensemble to 100 K, 200 K and 300 K in intervals of 50 ps, except the last step that took 100 ps. Restraints were imposed to the protein and ligand during the first heating step, allowing only water and ions to diffuse. Afterwards, the density was converged up to water density at 300 K with 100 ps in the NPT ensemble. At this point, the time step was incremented from 1 fs to 2 fs via the SHAKE algorithm.[136] Production runs were performed in the NPT ensemble until the RMSD of the protein backbone reached a plateau and all the protein-ligand interactions were stable. In *Sm*ChiB, 60 ns and 45 ns were run for the Michaelis complex and the reaction intermediate (Glc-ox and Glc-ox$^+$ scenarios), respectively. In the case of Hya, the boat conformation of the GlcNAc at the *-1* subsite was maintained imposing a soft restrain of 50 kcal/mol/rad^2 to the C1-O5-C5-C4 dihedral angle in the interval of -25 and 25 degrees. 75 ns of classical MD production were run for the Hya system. Analysis of the obtained trajectories was carried out employing standard tools of AMBER and VMD.[137]

QM/MM MD calculations

After classical MD, quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) calculations were run for each system. We used the method developed by Laio et al.,[95] which couples classical MD with Car–Parrinello MD.[90] The QM region for the reaction intermediate in *Sm*ChiB included the disaccharide GlcNAc-Glc-ox/ox$^+$, the catalytic

water and the side chains of the catalytic residues Asp142 and Glu144 (74 atoms, 17.05 × 22.59 × 21.61 Å^3). In the case of the Michaelis complex of *Sm*ChiB, Asp142 and Glu144 and carbohydrate moieties at *-2* and *+1* subsites were capped, and thus the QM region included the side chain of Asp142 and Glu144, the GlcNAc at *-1* subsite, and half of each adjacent monosaccharide (61 atoms, 17.44 × 17.64 × 16.99 Å^3). The calculation of Hya included the side chain of Asp111 and Glu113, the GlcNAc at *-1* subsite and half of the *-2* subsite sugar (54 atoms, 16.89 × 15.95 × 19.80 Å^3). All side chains were capped by saturating the C_α of each protein residue with capping hydrogens. Sugar residues were capped in a similar manner but saturating the parallel C1···C2 and C4···C5 bonds of each sugar ring. The NN, MIX and ESP radii were set at 9.52 Å, 14.82 Å and 26.46 Å, respectively, in the reaction intermediate in *Sm*ChiB, at 7.94 Å, 10.58 Å and 15.87, respectively, in the Michaelis complex in *Sm*ChiB, and at 5.29 Å, 10.58 Å and 15.87 Å, respectively, in Hya. The electronic structure was described within the DFT, using the PBE,[138] which has provided a reliable description of carbohydrate systems.[88,89] Kohn–Sham orbitals were expanded in a plane wave basis set with a kinetic energy cut-off of 70 Ry. Norm-conserving *ab initio* pseudopotentials were employed, generated within the Troullier–Martins Scheme.[139] In all QM/MM calculations, a time step of 0.12 fs was used. A fictitious electron mass of 850 a.u., 800 a.u. and 900 a.u. were used for the reaction intermediates in *Sm*ChiB, the Michalis complex in *Sm*ChiB and Hya, respectively.

The following protocol for the QM/MM MD simulations was followed. First, the structure was optimized by MD with annealing of nuclear velocities, until the maximal component of the nuclear gradient was lower than 5×10^4 a.u. Afterwards, the system was equilibrated with several ps of QM/MM MD at 300K employing the Nosé–Hoover thermostat[140,141]. In particular, 15.0 ps, 3.5 ps and 2.5 ps were run for the reaction intermediate in *Sm*ChiB, the Michalis complex in *Sm*ChiB and Hya, respectively. A snapshot of these simulation was used as starting point for the QM/MM metadynamics calculations for each system. In the case of the reaction intermediate in *Sm*ChiB, QM/MM MD simulations using three different exchange-correlation functionals (PBE,[138] HCTH[142] and BLYP[86,87]) were tested starting from the Glc-ox$^+$ model. The evolution of relevant active site distances for each simulation is shown in Figure S2. In all cases, the system was found to evolve towards the oxazoline configuration.

QM/MM metadynamics simulations

All metadynamics simulations were calculated using the PLUMED 2 plugin.[103] The CVs chosen to explore the conformational FES of the Glc-ox pyranose ring within the *Sm*ChiB active site were the Cartesian puckering coordinates q_x and q_y. The height and width of the Gaussian terms was set at 0.3 kcal/mol and 0.1 Å, respectively, for the two CVs. The deposition time was set at 200 MD steps. The conformational FES was reconstructed after depositing 1100 Gaussian functions and the energetic difference between the observed minima remained stable. Hydrolysis of the GlcNAc-Glc-ox by *Sm*ChiB was driven using the following CVs. The first CV was defined as the distance between the oxygen of the nucleophilic water and the oxazoline anomeric carbon (CV_1=C1···O_w), and thus accounts for the nucleophilic attack of the water molecule. The second CV was defined as the distance between the oxazoline nitrogen atom and the hydrogen atom of the Asp142 assisting residue (CV_2=N_x···H_x). The height of the Gaussian terms was set at 1 kcal/mol. The width of the Gaussian potentials was set at 0.15 Å for both distances, according to their oscillations observed in free dynamics. The deposition time was set at 300 MD steps. Tables S2-S4 show the main distances, charges, and dihedral angles at each characteristic state of the FES. The first reaction step of *Sm*ChiB reaction mechanism was modelled using two CVs. The first CV (CV_1) was defined as the difference between C1···O_x and C1···O1 distances. Thus, CV_1 accounts for the intramolecular nucleophilic attack of the carbonyl oxygen (O_x) and the cleavage of the glycosidic bond. The second CV (CV_2) was defined as the distance difference between H_{Glu144}···O1 and O_{Glu144}···H_{Glu144}. Therefore, this CV accounts for the proton transfer between Glu144 and the glycosidic oxygen (O1). The height of the Gaussian terms was set at 1 kcal/mol and their width at 0.1 Å for both distance differences. The deposition time was set at 300 MD steps. The FES was reconstructed after depositing 580 Gaussian terms. Obtention of the reaction intermediate in Hya was performed using the distance between C1 and the glycosidic oxygen (O1) as CV (CV_1). The height of the Gaussian terms was set at 1 kcal/mol and its width at 0.1 Å. The deposition time was set at 300 MD steps. The FES was reconstructed after depositing 450 Gaussian terms. In all metadynamics calculations of enzymatic reactions the Gaussian term height was reduced to 0.5 kcal/mol just before crossing the TS to obtain a smoother and more converged FES. The simulations were stopped once a TS recrossing events was observed.

3.5. Supplementary material

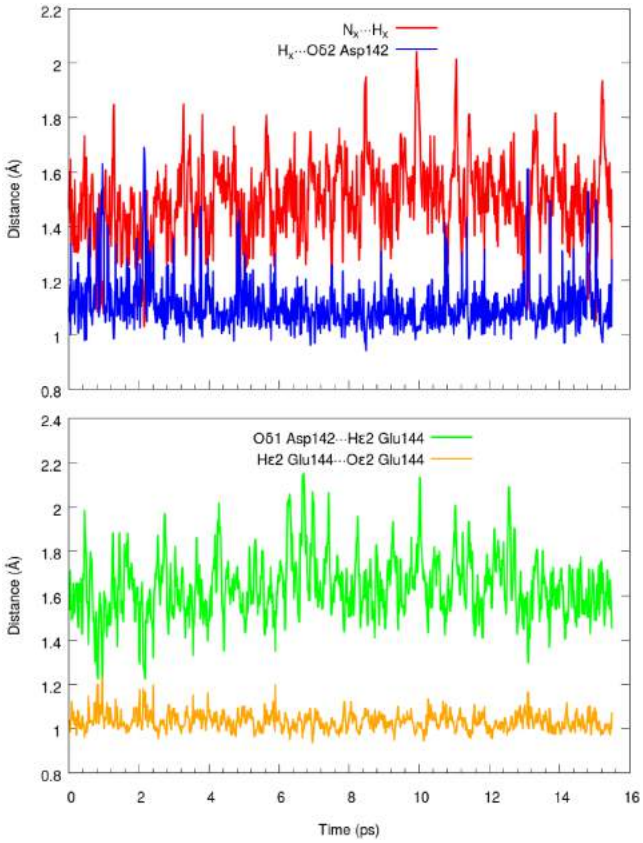

Figure S3.1. Distances involved in the *Sm*ChiB active site hydrogen bond network during the QM/MM MD simulation in the reaction intermediate. The substrate remains deprotonated 93.8% of the time.

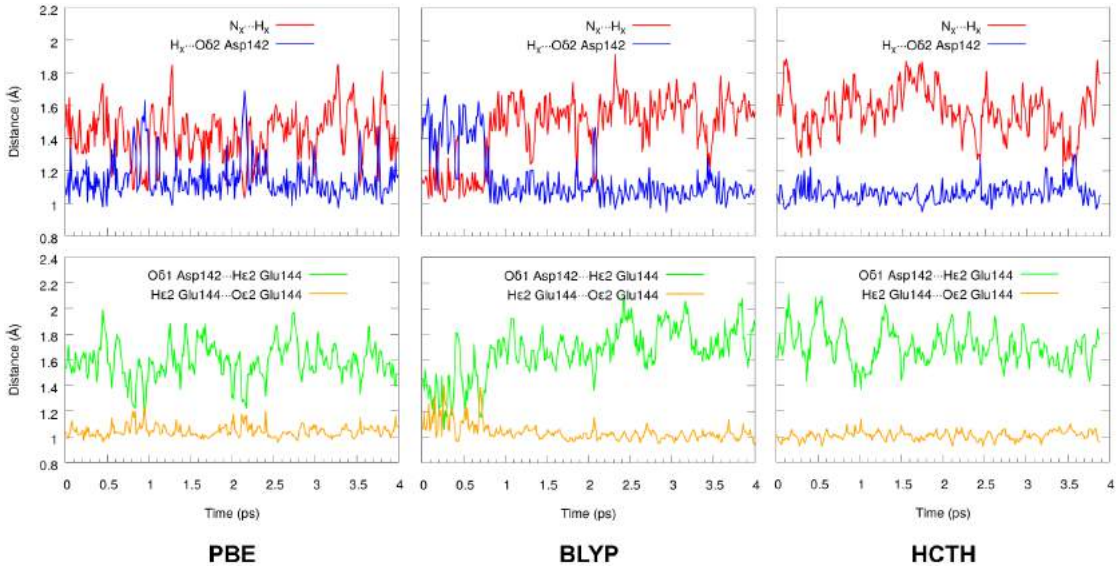

Figure S3.2. Main distances defining the proton transfer events occurring during QM/MM MD simulations with different DFT functionals in the reaction intermediate of *Sm*ChiB.

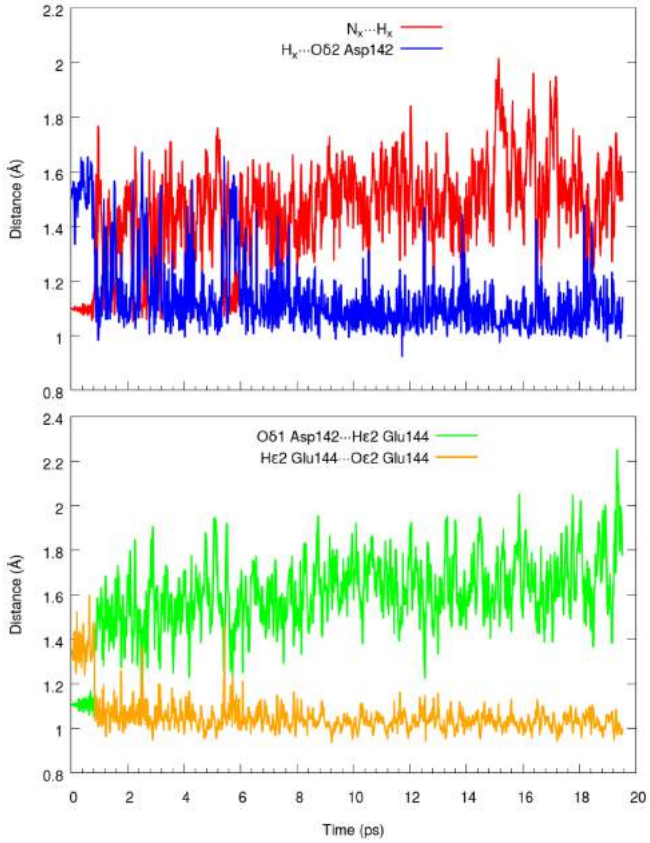

Figure S3.3. Evolution of the distances involved in the active site hydrogen bond network during the restrained QM/MM MD simulation (a restraint on the $N_x \cdots H_x$ distance was applied during the first ps).

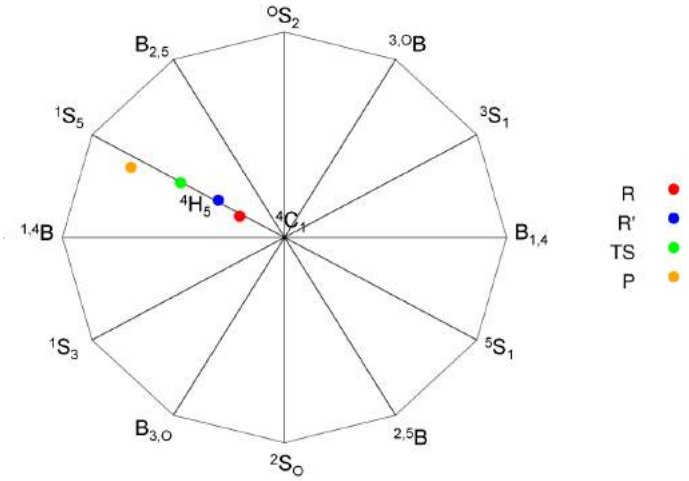

Figure S3.4. Puckering coordinates of the Glc-ox at the *-1* subsite for relevant states along the second reaction step in *Sm*ChiB.

Table S3.1. X-ray structures available for GH18, GH20, GH84 and GH85 enzymes in complex with Glc-oxazoline and its derivatives.

GH family	PDB entry	Year	Resolution (Å)	Ligand	Enzyme
GH18	1E6Z	2001	1.99	Glc-oxazoline	Chitinase
GH18	4HMD	2012	2.26	Glc-oxazoline	Chitinase
GH20	4C7G	2013	1.80	Glc-oxazoline	β-*N*-acetylhexosaminidase
GH84	2WZH	2009	2.20	Glc-oxazoline	O-GlcNAcase
GH84	2WZI	2009	1.90	Glc-5F-oxazoline	O-GlcNAcase
GH18	2WK2	2009	2.05	Glc-thiazoline	Chitinase
GH18	2WM0	2009	1.90	Glc-thiazoline	Chitinase
GH20	1HP5	2001	2.10	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	1NP0	2003	2.50	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	2EPN	2008	1.61	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	3OZO	2011	2.00	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	3SUS	2011	1.80	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	4JAW	2013	1.80	Glc-thiazoline	Lacto-N-biosidase
GH20	5A6A	2015	2.69	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH20	5OAR	2017	2.30	Glc-thiazoline	β-*N*-acetylhexosaminidase
GH84	2CHN	2006	1.95	Glc-thiazoline	O-GlcNAcase
GH85	3FHQ	2009	2.45	Glc-thiazoline	Endo-β-*N*--acetylglucosaminidase
GH85	2W92	2009	1.65	Glc-thiazoline	Endo-β-*N*--acetylglucosaminidase

Table S3.2. Relevant distances (mean and standard deviation) of each representative structure along the reaction pathway.

Distance (Å)	I	I'	TS	P
C1···O_w	3.79 ± 0.04	3.08 ± 0.05	2.11 ± 0.02	1.47 ± 0.02
C1···O5	1.38 ± 0.03	1.36 ± 0.02	1.32 ± 0.02	1.40 ± 0.04
C1···O_x	1.51 ± 0.04	1.55 ± 0.05	2.23 ± 0.21	3.03 ± 0.15
N_x···H_x	1.61 ± 0.04	1.15 ± 0.03	1.06 ± 0.01	1.04 ± 0.01
H_x··· O_δ2 Asp142	1.07 ± 0.04	1.38 ± 0.07	1.63 ± 0.13	1.94 ± 0.08
O_δ1 Asp142···H_ε2 Glu144	1.67 ± 0.13	1.49 ± 0.13	1.18 ± 0.12	1.01 ± 0.03

$H_{\varepsilon}2$ Glu144···$O_{\varepsilon}2$ Glu144	1.03 ± 0.03	1.06 ± 0.04	1.35 ± 0.12	1.86 ± 0.12
$O_{\varepsilon}1$ Glu144···H_w	1.80 ± 0.10	2.19 ± 0.52	1.74 ± 0.15	1.01 ± 0.04
H_w···O_w	0.99 ± 0.03	1.00 ± 0.03	1.05 ± 0.02	1.88 ± 0.13
H_{η} Tyr214···O_x	2.00 ± 0.15	1.97 ± 0.11	1.92 ± 0.14	1.88 ± 0.07

Table S3.3. Dihedral angles defining the conformation of the 5-membered ring for each relevant state along the reaction pathway.

Dihedral angle	**I**	**I'**	**TS**	**P**
C2-C1-O5-C5	-34.5 ± 10.6	-31.6 ± 8.5	-13.5 ± 9.6	8.4 ± 12.9
C3-C2-C1-O5	25.7 ± 6.6	16.7 ± 6.4	-5.3 ± 11.6	-32.3 ± 8.8

Table S3.4. ESP charges (e) of each representative structure along the reaction pathway. See Figure S2 for atom labeling. δ^{an} corresponds to the sum of the charges of C1 and its neighboring atoms O5, C2, O_x and H).

Atom	**I**	**I'**	**TS**	**P**
C2	-0.01 ± 0.10	0.05 ± 0.14	0.09 ± 0.09	0.19 ± 0.07
N_x	-0.42 ± 0.08	-0.45 ± 0.12	-0.57 ± 0.08	-0.59 ± 0.04
C_x	0.42 ± 0.04	0.46 ± 0.10	0.59 ± 0.07	0.61 ± 0.06
O_x	-0.52 ± 0.04	-0.55 ± 0.04	-0.66 ± 0.03	-0.65 ± 0.02
C1	0.63 ± 0.11	0.75 ± 0.14	0.71 ± 0.14	0.63 ± 0.05
H1	0.01 ± 0.02	0.02 ± 0.03	0.03 ± 0.02	0.00 ± 0.01
O5	-0.53 ± 0.06	-0.48 ± 0.07	-0.43 ± 0.08	-0.52 ± 0.03
H_x	0.33 ± 0.04	0.40 ± 0.04	0.38 ± 0.02	0.32 ± 0.01
$O_{\delta}2$ Asp142	-0.55 ± 0.04	-0.67 ± 0.05	-0.64 ± 0.04	-0.58 ± 0.02
$O_{\delta}1$ Asp142	-0.59 ± 0.05	-0.67 ± 0.03	-0.60 ± 0.05	-0.49 ± 0.06
$H_{\varepsilon}2$ Glu144	0.33 ± 0.03	0.38 ± 0.03	0.40 ± 0.05	0.34 ± 0.02
$O_{\varepsilon}2$ Glu144	-0.49 ± 0.05	-0.59 ± 0.05	-0.62 ± 0.07	-0.52 ± 0.03
$O_{\varepsilon}1$ Glu144	-0.63 ± 0.06	-0.64 ± 0.04	-0.66 ± 0.05	-0.47 ± 0.02
H_w	0.42 ± 0.06	0.37 ± 0.07	0.45 ± 0.06	0.37 ± 0.02
O_w	-0.95 ± 0.11	-0.92 ± 0.12	-0.96 ± 0.09	-0.81 ± 0.05
δ^{an}	-0.41 ± 0.10	-0.21 ± 0.11	-0.25 ± 0.09	-0.36 ± 0.05

4. Conversion of a GH84 glycosidase into a phosphorylase.

Abstract: *O*-GlcNAcases belong to family GH84 and catalyse the cleavage of the GlcNAc attached to Ser/Thr residues on proteins (*O*-GlcNAc) through a substrate-assisted mechanism. This post-translational modification is crucial in biological processes such as protein phosphorylation or cellular signalling. Besides their role in biology, *O*-GlcNAcases — as other GHs — are being engineered into novel and efficient chemoenzymatic tools. In this chapter, a *O*-GlcNAcase is converted into a glycoside phosphorylase by site-directed mutagenesis, and we rationalize this novel conversion by molecular dynamics on the *O*-GlcNAcase from *Thermobaculum terrenum* (*Tt*OGA). We show that the Asp to Asn mutation of the catalytic acid/base residue modifies the active site electrostatic potential, enhancing the phosphate ion population in the engineered variant (*Tt*OGA-Asp120Asn). Afterwards, we explore the hydrolysis/phosphorylation mechanisms of these enzymes by means of QM/MM metadynamics calculations, revealing that both native and mutant enzymes feature a Glc-ox^{+} as a reaction intermediate, differing from the Glc-ox observed in GH18 chitinases (chapter 3). This remarkable difference is due to the arrangement of the catalytic residues in the protein sequence and their environment. Elucidating the *O*-GlcNAcase catalytic mechanisms can assist the design of new inhibitors for disease-related GH enzymes that participate in protein post-translational.

Publication: D. Teze[‡], J. Coines[‡], L. Raich, V. Kalichuk, C. Solleux, C. Tellier, C. Andre-Miral, B. Svensson and C. Rovira. A Single Point Mutation Converts GH84 *O*-GlcNAc Hydrolases into Phosphorylases: Experimental and Theoretical Evidence., *J. Am. Chem. Soc.*, 2020, 142, 2120–2124 ([‡] equal contribution).

4.1. Introduction

O-GlcNAcylation is a post-translational modification that consists in attaching a GlcNAc at Ser/Thr residues on a myriad of nuclear, cytosolic, and mitochondrial proteins. This modification is found across species and controls a wide range of cellular processes, such as transcription, nutrient sensing and signalling.[143] The dysregulation of *O*-GlcNAcylation is related to diabetes, cancer, and neurodegeneration.[144-146] In addition, there is a tight crosstalk between *O*-GlcNAcylation and phosphorylation, another major post-translational modification crucial in biology as well.[147] Contrary to phosphorylation, which requires over 600 kinases and phosphatases in humans, only two enzymes catalyse *O*-GlcNAc cycling. While *O*-GlcNAc transferase (OGT) transfers a single *O*-GlcNAc moiety from UDP-GlcNAc to a Ser/Ter residue in proteins, *O*-GlcNAcase (OGA) catalyses the removal of the attached *O*-GlcNAc. (Figure 4.1). As a consequence, the enzymes involved in such modification have been extensively investigated.[148] OGA enzymes are glycosidases that belong to family GH84 and catalyse the hydrolysis of *O*-GlcNAc operating via a substrate-assisted reaction mechanism (Figure 4.2).

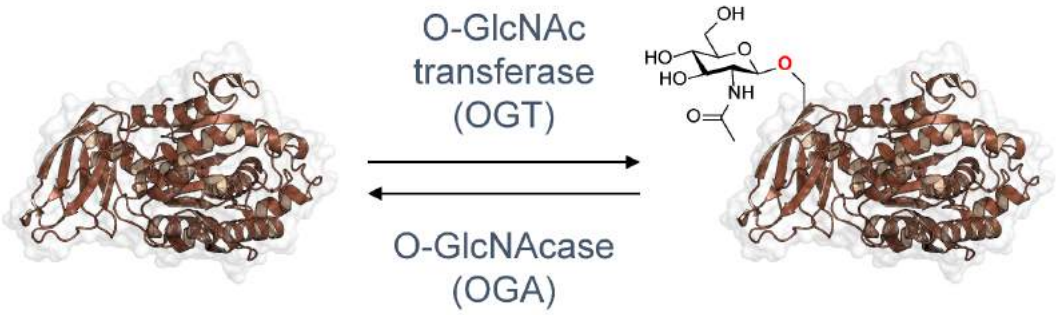

Figure 4.1. Dynamical interchange of *O*-GlcNAcylation catalysed by *O*-GlcNAc transferase and *O*-GlcNAcase. The structure of the human carboxypeptidase N (PDB entry 2NSM) is used as illustrative example.

Our collaborators Dr. David Teze and Prof. Birte Svensson (Technical University of Denmark, Denmark) were recently evaluating mutant variants of OGA enzymes in order to explore novel enzymatic activities and stumbled upon an oddity. Instead of obtaining an inactive enzyme by mutating the acid/base residue from Asp to Asn, as seen in other glycosidases, the mutant cleaved *p*-nitrophenyl-*N*-acetyl-β-D-glucosaminide (*p*NP-GlcNAc) in phosphate buffer. While the native GH84 enzyme produced exclusively GlcNAc, the Asn mutant released β-GlcNAc-1-phosphate (GlcNAc-P). Catalysis took place with retention of

the anomeric configuration. Importantly, when Glc-ox (reaction intermediate) was employed as a substrate, the reaction rate was the same as using *p*NP-GlcNAc. This indicates that the second step of the observed retaining mechanism is rate-limiting.

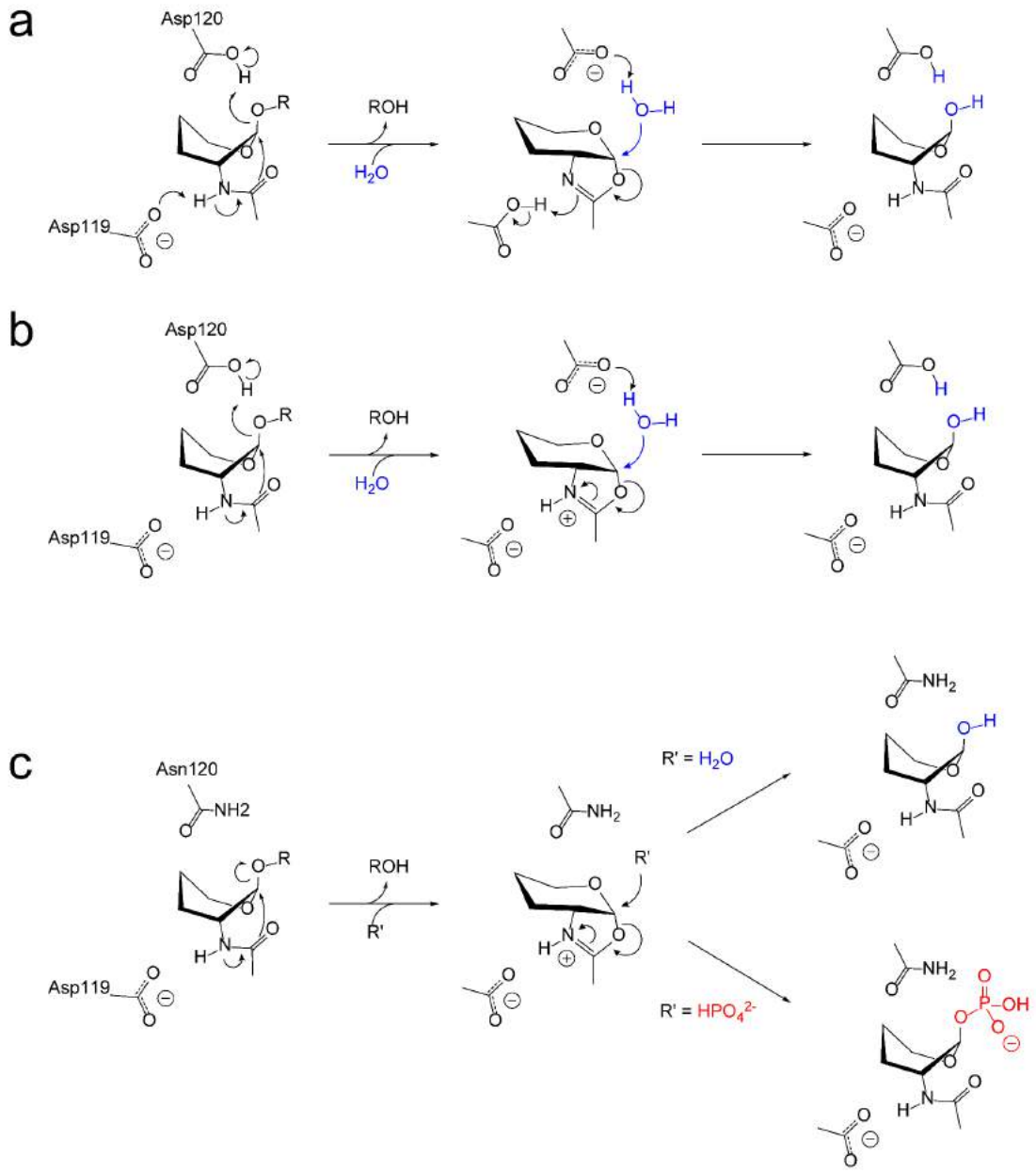

Figure 4.2. The two-steps substrate-assisted reaction mechanism of *Tt*OGA considering a (a) Glc-ox (glucose oxazoline) reaction intermediate; or a (b) Glc-ox$^+$ (glucose oxazolinium ion). (c) Hydrolysis (blue) and phosphorylation (red) reactions catalysed *Tt*OGA-Asp120Asn mutant. From left to right: Michaelis complex, reaction intermediate and products.

Depending on the experimental conditions and the specific OGA enzyme, hydrolysis was abolished at different degrees. For instance, 1 μg/mL of the OGA from *Streptococcus pyogenes* (*Sp*OGA-Asp245Asn) with 7.5 mM pNP-GlcNAc in 0.5 M phosphate buffer at pH 7.8 produced GlcNAc-P exclusively, whereas GlcNAc was not detected. Instead, only 30% of phosphorolysis was observed using the OGA from *Thermobaculum terrenum* (*Tt*OGA-Asp120Asn) as a catalyst in 20 mM phosphate concentration at pH 7.8. Despite the transformation of hydrolytic GHs into synthetic enzymes is well established, this is the very first example of a genuine GH converted into a glycoside phosphorylase (GP).

GPs catalyse the transfer of an inorganic phosphate to a glycoside acceptor. From mechanistic and structural points of view, most GPs are similar to hydrolytic GHs.[149] However, rather than being classified in separated families, they are found within several GHs (GH3, 13, 65, 94, 112, 130, 149) and GTs families (GT4 and 35).[150,151] The retaining β-GPs reported to date belong to family GH3, acting on either β-GlcNAc or β-Glc, via a glycosyl-enzyme intermediate.[152,153] Interestingly, GHs and GPs from GH3 exhibit similar active sites but for their acid/base residue (Glu in GHs, His in some GPs),[152-154] suggesting that this position may determine enzyme specificity. Nevertheless, neither phosphorolytic nor hydrolytic activities were observed by *Sp*OGA-Asp245His mutant in 0.1 M phosphate buffer pH 7.8 (k_{cat}= < 10^{-3} s^{-1} for pNP-GlcNAc cleavage), highlighting possible differences between the classical retaining mechanism of GH3 and the substrate-assisted mechanism followed by GH84. Hence, further analyses at molecular level were required to comprehend the conversion of GH84 glycosidases into GPs (Figure 4.2c).

To understand such novel transformation, we turned out attention to *Tt*OGA, for which a crystal structure was available[155] and shares 33% sequence identity with *Sp*OGA. The *Tt*OGA structure (Figure 4.3) comprises a catalytic domain that shapes a $(\beta/\alpha)_8$ TIM barrel (residues 1-280) and a C-terminal helical bundle domain (residues 281-474), resembling those found in other bacterial OGA homologues.[156] *Tt*OGA is active on both synthetic substrates (*p*NP-GlcNAc and 4-methylumbelliferyl GlcNAc) and *O*-linked GlcNAc.[155] The adjacent catalytic dyad Asp119 and Asp120 in *Tt*OGA corresponds to the assisting and the acid/base residues, respectively, which perform catalysis via a substrate-assisted mechanism. The nature of the reaction intermediate in family GH84 has not yet been characterized, although a neutral Glc-ox has been frequently depicted, thus assumed, in the literature (Figure 4.2a).[64,157,158] Notably, there is no insertion between the catalytic residues (Asp-Asp) in GH84 enzymes, contrary to *Sm*ChiB and *Bv*Hya (Asp-X-Glu in both GH18 and GH56 families), analyzed in Chapter 3. However, the occurrence of a Glc-ox reaction intermediate in *Sm*ChiB and *Bv*Hya

was found to be due to the existence of a proton shuttle that links the catalytic dyad, thus one might expect a different situation in family GH84 (Figure 4.2b), since both catalytic aspartate residues do not interact. This will be one of the aspects particularly analyzed in this chapter, in which we investigate the conversion from a GH to a GP in *Tt*OGA.

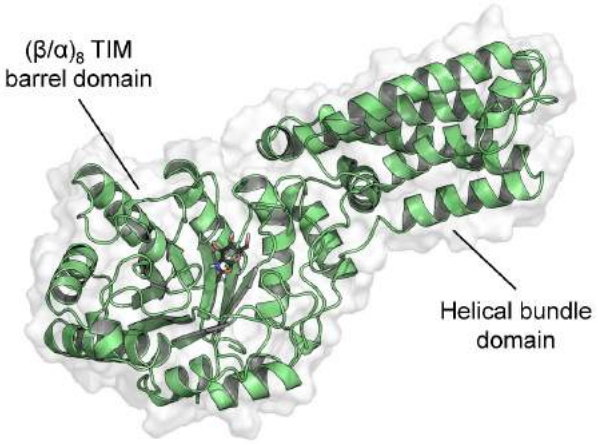

Figure 4.3. *Tt*OGA structural overview (PDB entry 5DIY)[155]. The catalytic domain resembles a $(\beta/\alpha)_8$ TIM barrel, while the C-terminal shapes a helix bundle domain. A Glc-ox$^+$ ligand is present (shown as black sticks) inside the pocket active site of *Tt*OGA.

4.2. Results and discussion

4.2.1. Phosphate ions near the active site of *Tt*OGA and *Tt*OGA-Asp120Asn?

As a first step in our investigation, we investigated the stability of phosphate ions in the vicinity of the active site of *Tt*OGA and *Tt*OGA-Asp120Asn, a requirement to undergo the phosphorylation reaction. Several MD simulations were designed to determine the identity of the phosphate ion (mono or dianion, i.e., $H_2PO_4^-$ or HPO_4^{2-}) that can perform a nucleophilic attack on the active site sugar, as well as the most likely protonation state of the reaction intermediate (Glc-ox or Glc-ox$^+$).

The initial structure for all calculations was taken from the crystallographic complex of the glycopeptide hTab1-*O*-GlcNAc bound to *Tt*OGA-Asp120Asn mutant (PDB entry 5DIY).[155] While the GlcNAc residue located at the *-1* subsite served as a scaffold to build the Glc-ox/Glc-ox$^+$ moiety, the rest of the peptide was removed. The simulations were started from the reaction intermediate for two main reasons: (1) because we would like to elucidate its protonation state (Glc-ox or Glc-ox$^+$) in the context of substrate-assisted catalysis and (2) because deglycosylation is the rate-limiting step of the observed phosphorylation reaction. The initial protonation state of the assisting residue and the substrate in the reaction intermediate cannot be directly inferred from structural data, as occurred in our previous study of GH18 *Sm*ChiB.[128] In addition, both HPO_4^{2-} and $H_2PO_4^-$ phosphate ion species are present in solution at pH 7.8 ($pK_{a2} \approx 7.21$). Therefore, all possible combinations were set up depending on the protonation state of the substrate in the reaction intermediate (Glc-ox or Glc-ox$^+$), the phosphate ion form (HPO_4^{2-} or $H_2PO_4^-$) and the enzyme variant (wild type or Asp120Asn). Figure 4.4 illustrates all models considered.

A first set of unrestrained MD simulations were performed starting with a phosphate ion located at the active site (see computational details), in order to elucidate which anion species is preferred — and how it behaves — in both *Tt*OGA and *Tt*OGA-Asp120Asn. The resulting eight models were submitted to short classical MD calculations of 10 ns each, which was enough to see significant differences.

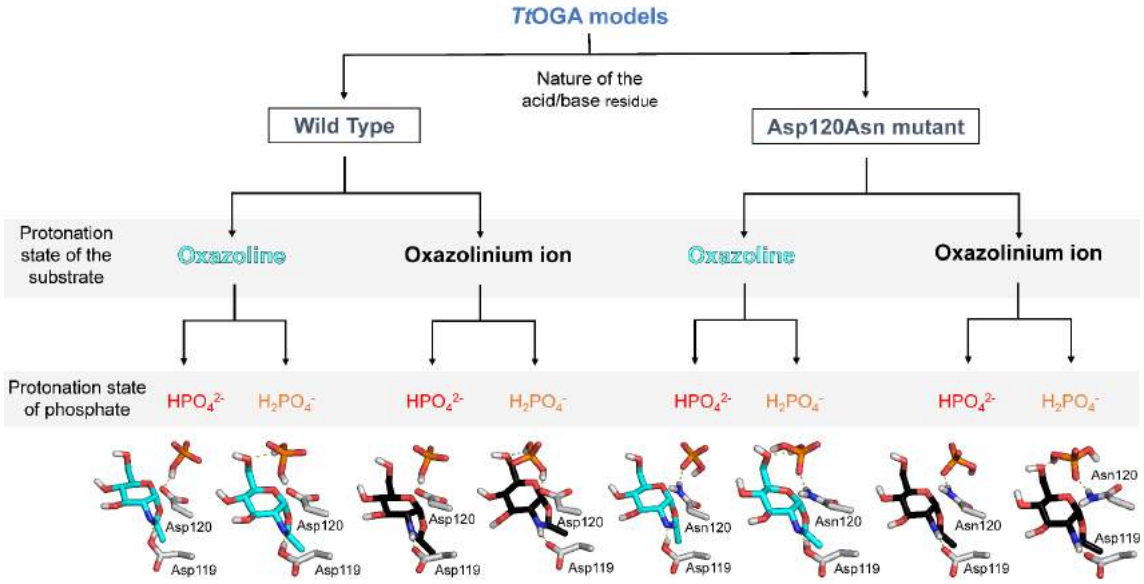

Figure 4.4. Schematic representation of the models for the *Tt*OGA enzyme considered here, depending on the nature of the acid/base residue (native or Asp120Asn variant), the protonation state of the substrate (Glc-ox or Glc-ox$^+$) and the phosphate form (HPO_4^{2-} or $H_2PO_4^-$).

The results obtained show that all Glc-ox models can be excluded because none of them resulted in a reactive conformation (Figure S4.1). In particular, the protonated Asp119 side chain oriented towards the carboxyl group of Asp120 in *Tt*OGA (or the carbonyl group of Asn120 in *Tt*OGA-Asp120Asn), instead of interacting with the N_x atom of the substrate through a hydrogen bond. This did not happen in the Glc-ox$^+$ models, which were the ones further considered. However, phosphate ions diffused away from the active site in all cases.

To elucidate which phosphate form most likely reacts with Glc-ox$^+$ in native and mutant *Tt*OGA enzymes, either HPO_4^{2-} or $H_2PO_4^-$ anions were maintained near the anomeric carbon (C1) employing a restraint between C1 and the phosphate oxygen (O_p) during the first 10 ns of a new set of MD simulations. Subsequently, the constant force was modified with values ranging from 0.0 to 3.0 kcal/mol in 5 independent replicas of 1 ns each, considering that phosphate departure can already take place at this timescale (Figure 4.5).

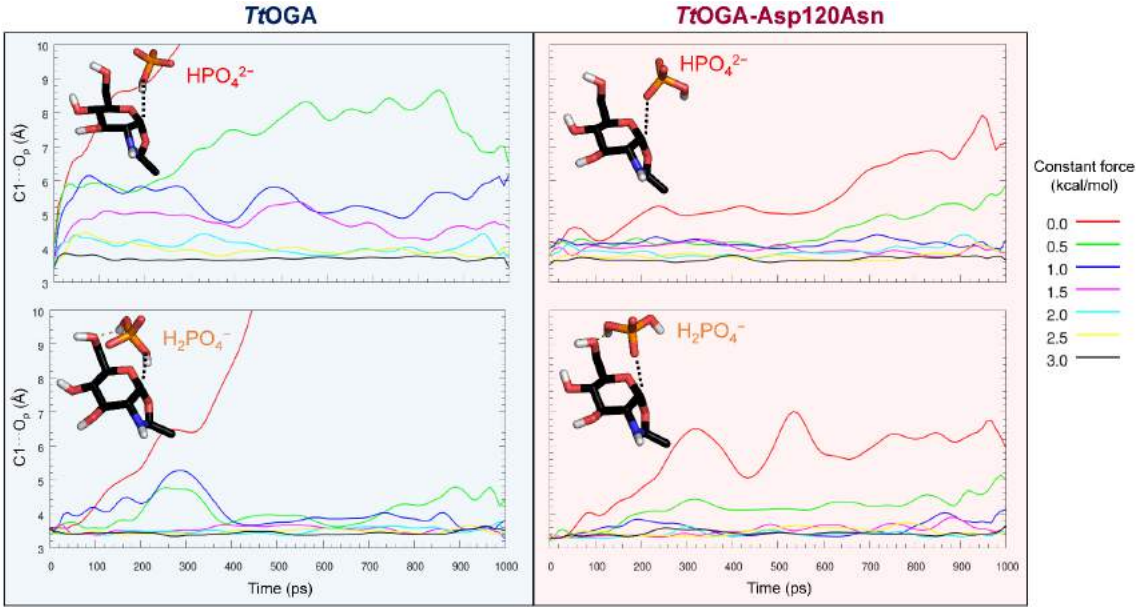

Figure 4.5. Evolution of the distance between the anomeric carbon (C1) and the phosphate oxygen (O_p) during the 1 ns restrained MD simulations of *Tt*OGA and *Tt*OGA-Asp120Asn models with Glc-ox$^+$. Different values for the constant force of the restraint applied in this distance were evaluated. For the sake of clarity, results are averaged over five independent MD replicas that were performed for each applied restraint.

Phosphate ions were not stable near the active site in all cases after releasing the restraints (read line in Figure 4.5, i.e. no constant force applied), although the dianion remains not too far from the active site in the case of the mutant enzyme. Migration of the phosphates occurs rapidly in the case of native *Tt*OGA, suggesting that they are electrostatically repelled by the acid/base residue Asp120, since both are negatively charged. In the same vein, greater force constant values are required to maintain the more charged HPO_4^{2-}, compared to $H_2PO_4^-$. In the case of *Tt*OGA-Asp120Asn, the behaviour of both anions is similar, emphasizing that the electrostatic repulsion observed in wild-type *Tt*OGA is now faded by the mutation to Asn. However, even in the case of the Asn mutant, phosphate anions do not remain close than ≈ 6 Å from the anomeric carbon if a restraint is no applied. This migration of phosphate ions into the solvent is probably due to the well-known tendency of force fields to oversolvate phosphate ions.[159] This is also obvious from our simulations, as clusters of phosphate ions were observed in the solution. Indeed, Jiří Vymětal et al. have pointed out that further development of phosphate parameters is still required for accurate atomistic simulations.[82] As a consequence, and in view of the results from Figure 4.5, it is reasonable to assume that phosphate ions would remain in the active site in a parameter-free dynamics. Thus, $H_2PO_4^-$, for which electrostatic repulsion with Asp120 is lower than HPO_4^{2-}, can be considered to

model phosphorylation by *Tt*OGA. In the case of the *Tt*OGA-Asp120Asn mutant, the electrostatic repulsion with Asp120 vanishes and the more charged and nucleophilic HPO_4^{2-} can be accommodated in the active site employing just a weak restraint to counterbalance the limitations of the phosphate force field. All restraints were released during the following QM/MM simulations, in which the phosphate ion is described at QM level.

A second set of classical MD simulations were performed to quantify the phosphate distribution around *Tt*OGA and *Tt*OGA-Asp120Asn. This is another way to check whether phosphate ions are able to visit the active site and whether the mono or dianion is preferred. Experimental evaluation of phosphate concentrations in *Tt*OGA-Asp120Asn showed that a maximal phosphorylation enhancement was obtained with 20 mM phosphate concentration at pH 7.8. To model such conditions, 9 HPO_4^{2-} and 2 $H_2PO_4^-$ ions were randomly placed in a simulation cell of 910 nm^3 for both *Tt*OGA and *Tt*OGA-Asp120Asn systems. Note that no phosphate ion was initially placed close to Glc-ox^+, as it was done in all previous simulations.

After equilibration of each system, a production run of 50 ns was performed. The relative phosphate density was obtained by computing the radial distribution function of both HPO_4^{2-} and $H_2PO_4^-$ with respect to the C_γ of the residue at position 120 (Asp in the wild-type enzyme and Asn in the mutant enzyme, Figure 4.6).

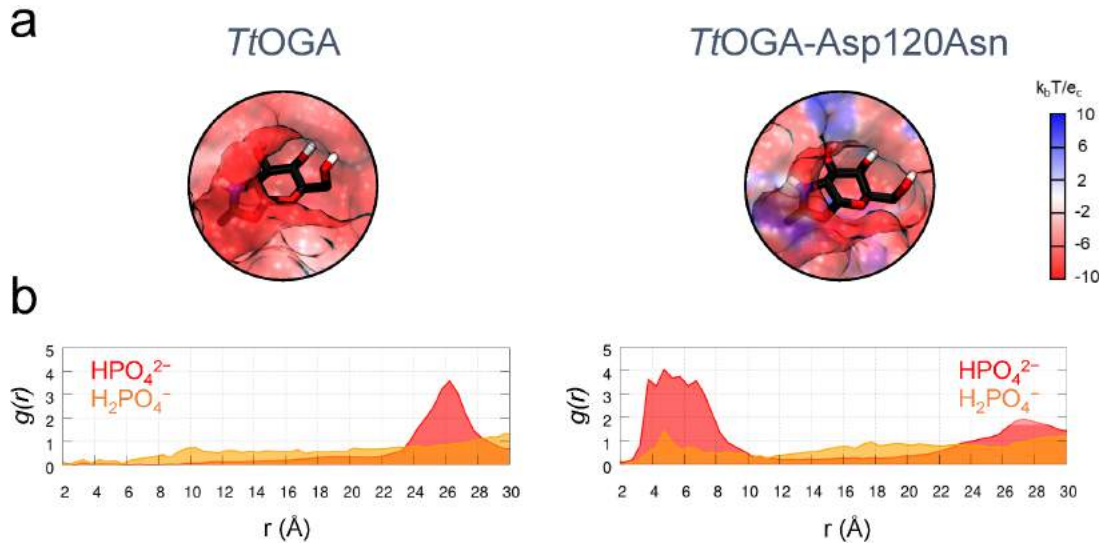

Figure 4.6. Phosphate distribution in GH84 enzymes. (a) Computed electrostatic potential of *Tt*OGA (left) and *Tt*OGA-Asp120Asn (right) at the active site, where the Glc-ox^+ substrate is depicted as black sticks. (b) radial distribution function *g(r)* of the distance between the C_γ of the Asp120/Asn120 residue and the closest oxygen of each phosphate ion (r in Å).

In addition, the electrostatic potential was calculated for both enzymes (Figure 4.6). The results show that phosphate anions do not reach the active site of *Tt*OGA, due to electrostatic repulsion from the Asp120 carboxylate, as reflected in the negative electrostatic potential at the active site pocket. However, the more positive active site of *Tt*OGA-Asp120Asn displays a higher phosphate density at ≈ 4-7 Å from the residue at position 120 compared with the wild-type enzyme, with the density being the highest for HPO_4^{2-}. These results are consistent with the tendencies observed in our previous simulations using a restraint for the C1···O_p distance.

In summary, our classical MD simulations indicate that the best model to analyse phosphorylation should contain a HPO_4^{2-} anion in the active site of *Tt*OGA-Asp120Asn. In contrast, no phosphate anions, regardless of their charge, are able to come close to the active site of the native enzyme and therefore its hydrolytic activity should be considered (Figure 4.6). Yet, we explored the hypothetical situation in which the wild-type *Tt*OGA uses the less charged, thus less repelled, $H_2PO_4^-$ monoanion to perform phosphorylation.

4.2.2. Reactivity of *Tt*OGA and *Tt*OGA-Asp120Asn mutant

The mechanisms of hydrolysis/phosphorylation in *Tt*OGA and *Tt*OGA-Asp120Asn were modelled by QM/MM metadynamics (Figure 4.7, left and S4.3-S4.4). The QM region for all systems included the Glc-ox$^+$ substrate, the side chain of Asp119 and Asp120 and the molecule that performs the nucleophilic attack on C1 (either water or $H_2PO_4^-$). It is worth mentioning that during both geometry optimization and QM/MM MD, the proton from Glc-ox$^+$ was not transferred to Asp119, suggesting that the protonated form of the substrate is the most stable, as predicted by the classical MD simulations. Further free energy calculations verified this assumption and allowed us to quantify the energetic difference between Glc-ox and Glc-ox$^+$ (next section). The reactivity of the Glc-ox$^+$ intermediate was evaluated using two CVs (Figure S4.2). CV_1 takes into account the nucleophilic attack on C1 from either water or $H_2PO_4^-$ (O_w or O_p atoms, respectively), whereas CV_2 assesses the proton transfer from the protonated nucleophile (water or $H_2PO_4^-$) to the acid/base catalytic residue Asp120. In Figure 4.7a, the free energy was projected into CV_1 to obtain the free energy profile, while the two-dimensional FESs are illustrated in Figure S4.3. Hydrolysis was calculated starting from a structure in which a water molecule is properly oriented for nucleophilic attack, which was often observed in the classical MD simulations since the pocket active site is exposed to bulk water. The hydrolysis reaction is found to be exergonic by 3.8 kcal/mol, with a free

energy barrier of 13.2 kcal/mol, in good agreement with kinetic data (k_{cat}=106 s^{-1}; $\Delta G^{\ddagger}$=14.7 kcal/mol).

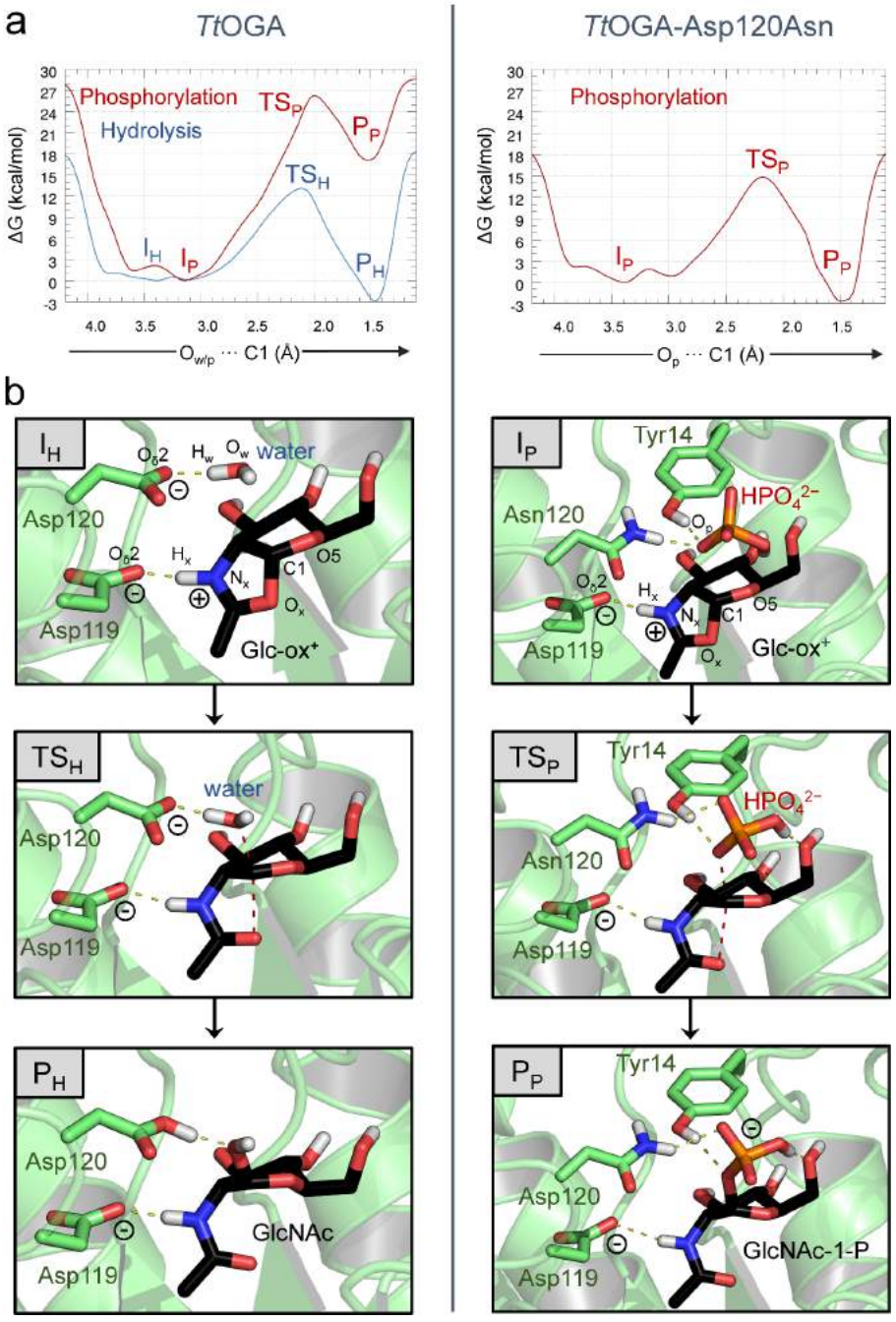

Figure 4.7. Hydrolysis/phosphorylation of Glc-ox^{+} by *Tt*OGA (left) and *Tt*OGA-Asp120Asn (right) enzymes. (a) Free energy profile reconstructed from the metadynamics simulation. (b) Average structures of the main states along the reaction coordinate for *Tt*OGA hydrolysis and *Tt*OGA-Asp120Asn phosphorolysis (I=Glc-ox^{+}, TS=transition state, P=reaction products). H and P subscripts correspond to hydrolysis and phosphorylation, respectively. Hydrogen atoms are omitted for clarity, except those attached to heteroatoms. Bonds being formed/broken are indicated with red dashed lines.

The reaction coordinate evolves as follows. From the reaction intermediate (I_H), the process starts with the approach of the catalytic water to C1 of Glc-ox$^+$ (C1$\cdots$O$_w$ = 3.39±0.05 Å, Table S1), until the TS (TS_H) is reached (C1$\cdots$O$_w$ = 2.10±0.05 Å). At this point, the C1$\cdots$O$_x$ bond (1.60±0.07 Å at I_H) is already broken (2.40±0.23 Å at TS_H), revealing that the reaction mechanism is dissociative. At TS_H, the proton from water is not yet transferred to Asp120 (O$_w\cdots$H$_w$ = 1.07±0.02 Å) and the interaction between the oxazolinium ion oxygen (O$_x$) and Asn223 is strengthened (O$_x\cdots$Hδ2$_{Asn223}$ distance from 2.78±0.28 Å at I_H to 2.11±0.11 Å at TS_H), suggesting a stabilizing role for this residue (Figure S4.5). The C1$\cdots$O5 distance is slightly shortened from 1.35±0.03 Å at I_H to 1.32±0.03 Å at TS_H, in accordance with an oxocarbenium ion-like TS.

Finally, the reaction products (P_H, GlcNAc) are obtained when the oxygen from the catalytic water collapses with C1 (C1$\cdots$O$_w$ and C1$\cdots$O$_x$ distances of 1.46±0.02 Å and 3.05±0.26 Å, respectively, at P_H) and its hydrogen is already transferred to Asp120 (H$_w\cdots$Oδ2$_{Asp120}$ = 1.03±0.02 Å at P_H). The conformational itinerary that the pyranose ring of Glc-ox$^+$ follows during hydrolysis is $^4H_5/^4E \rightarrow [^4H_5]^{\ddagger} \rightarrow {}^4H_5$ (Figure S4.6a).

To model phosphorylation reaction by *Tt*OGA and, given the absence of phosphate ions close to Glc-ox$^+$, we manually placed a $H_2PO_4^-$ anion in the active site and maintained it with a soft restraint, before performing QM/MM metadynamics simulations. However, phosphorylation results in a dissociative mechanism (C1$\cdots$O$_x$ = 2.59±0.08 Å at TS_P, Table S4.2) with a high energy barrier (26.3 kcal/mol, Figures 4.7 and S.4.3-S4.4) and the reaction turns out to be endergonic by 16.3 kcal/mol; thus the reaction appears to be unfavourable both kinetically and thermodynamically. This indicates that would a phosphate ion visit the active site, which does not happen according to our classical MD calculations, it would not react. This is consistent with the lack of phosphorolytic activity for the native *Tt*OGA. The obtained phosphorylation mechanism by *Tt*OGA shares some features with the native hydrolytic mechanism, such as shrinkage of the C1$\cdots$O5 bond at the TS (from 1.34±0.03 Å at I_P to 1.29±0.03 Å at TS_P), approach of the nucleophilic oxygen to C1 (O$_p$ in this case) and stabilization of the carbonyl group from Glc-ox$^+$ by Asn223 (shorter distance of O$_x\cdots$Hδ2$_{Asn223}$ interaction at TS_P and P_P, compared to I_P). Nevertheless, it differs in several aspects with its native activity. For instance, the proton from $H_2PO_4^-$ that interacts with Asp120 (H$_p$, equivalent to H$_w$ in hydrolysis) is transferred to this residue before reaching the TS. The conformational pathway of the pyranose ring of Glc-ox$^+$ followed during phosphorolysis is $^4C_1/^4E \rightarrow [^4E]^{\ddagger} \rightarrow {}^{1,4}B/^4E$ (Figure S4.6b), highlighting small conformational

differences during catalysis in comparison with the obtained hydrolytic conformational pathway.

The same approach was employed to understand phosphorolysis catalysed by *Tt*OGA-Asp120Asn (Figure 4.7, right). In this case, the considered phosphate form was HPO_4^{2-}, according to our classical MD calculations. Only one CV was used to drive the reaction (CV_1=C1···O_p), since HPO_4^{2-} does not need deprotonation (Figure S4.2). The obtained free energy profile reveals that, in contrast to the simulations for the wild-type *Tt*OGA, phosphorylation is exergonic by 2.7 kcal/mol, with a low reaction free energy barrier (14.8 kcal/mol) that is in good agreement with kinetic data (k_{cat}=4.4 s^{-1}; $\Delta G^{\ddagger}$=16.3 kcal/mol). Hence, both experiments and simulations evidence that *Tt*OGA-Asp120Asn is an efficient β-GlcNAc phosphorylase, with an activity comparable to the hydrolytic one from *Tt*OGA. The enzymatic reaction starts at the reaction intermediate (I_P) with the approach of the nucleophilic oxygen from HPO_4^{2-} until the TS (TS_P) is reached (C1···O_p distance from 3.40±0.08 Å at I_P to 2.17±0.01 Å at TS_P, Table S4.3). Once the phosphate finishes the nucleophilic attack on the C1 of Glc-ox^+ (C1···O_p = 1.47±0.03 Å at P_P), the reaction products (P_P, GlcNAc-P) are obtained. Common mechanistic features with *Tt*OGA catalysis are the dissociative character of the reaction (C1···O_X = 2.55±0.12 Å at TS_P), the shrinkage of the C1···O5 bond at the TS (from 1.34±0.03 Å at I_P to 1.29±0.02 Å at TS_P), and the strengthening of the O_x···$H\delta2_{Asn223}$ interaction along the reaction coordinate. The conformational itinerary followed by the pyranose ring of Glc-ox^+ is similar as the one obtained for *Tt*OGA phosphorylation, $^4C_1/^4E \rightarrow [^4E]^{\ddagger} \rightarrow {}^{1,4}B/^4E$ (Figure S4.6c), indicating that hydrolysis and phosphorylation show small differences in their conformational pathways, independently on the *Tt*OGA enzyme variant.

Analysis of the atomic rearrangements during phosphorylation reveals distinct changes in active site interactions upon the Asp120Asn substitution (Figures S4.5). In *Tt*OGA, the carboxylate of the acid/base catalyst (Asp120) interacts with the hydroxyls of both Tyr14 and Tyr168. However, the corresponding residue in the mutated enzyme (Asn120) does not interact with Tyr14 but with Tyr168 and the nucleophilic oxygen atom (O_p) of the HPO_4^{2-}, which in turn interacts with Tyr14 (Figure 4.7, right). Thus, the Asn mutation ultimately provides two hydrogen bond donors (Asn120 and Tyr14), properly oriented for catalysis. These small structural changes, together with the electrostatic potential modification of the active site upon the mutation, keep the incoming phosphate ion above the substrate anomeric carbon in a reactive configuration and further stabilize the GlcNAc-P products, hence enabling phosphorylation.

4.2.3. The nature of the reaction intermediate in *Tt*OGA GH84 enzyme

A debated issue in enzymes that follow the substrate-assisted mechanism is the nature of the reaction intermediate, whether it is protonated (Glc-ox$^+$) or deprotonated (Glc-ox). Our classical MD simulations of *Tt*OGA were consistent with a Glc-ox$^+$ species. In addition, we analysed this aspect by QM/MM simulations, starting from a structure in which a Glc-ox$^+$ is equilibrated in *Tt*OGA active site and a water molecule is placed for hydrolysis, mimicking the reactive hydrolysis configuration of the enzyme. In the light of previous classical MD calculations, the Glc-ox$^+$ and Asp119 deprotonated (assisting residue) configuration was considered.

All our QM/MM simulations for native or mutated *Tt*OGA (Figure 4.7 and S4.4) show that the Glc-ox$^+$ intermediate is stable (i.e. the proton remains with the sugar substrate). To quantify the relative free energy difference between both protonation states, we performed a QM/MM metadynamics calculation using as CV the distance between the hydrogen form the oxazolinium ring (H_x) and the carboxylic oxygen from Asp119 (Figure S4.2).

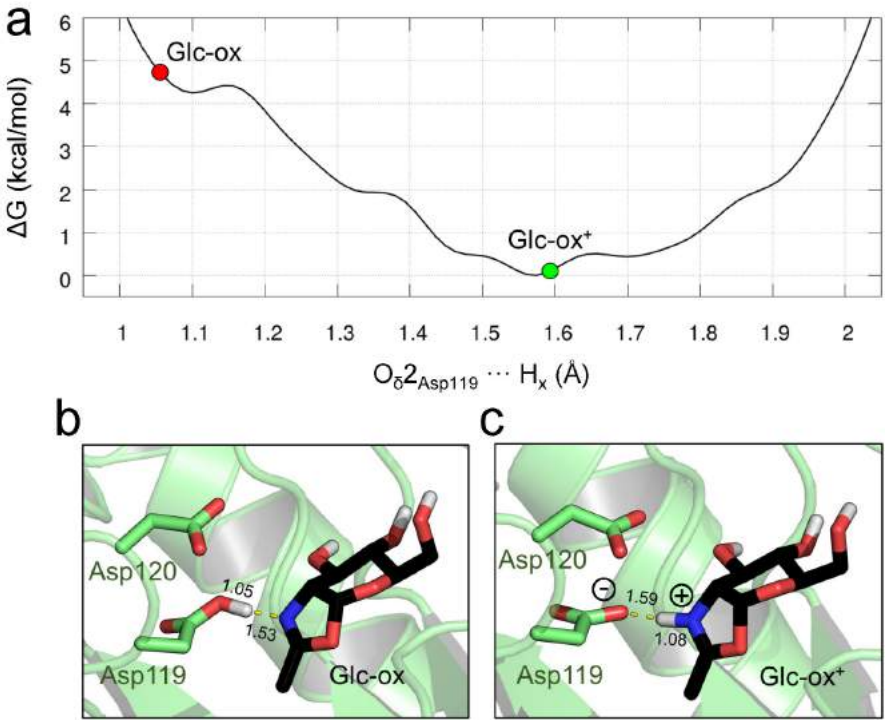

Figure 4.8. Analysis of the protonation state of the reaction intermediate in *Tt*OGA. (a) Free energy profile reconstructed from metadynamics. (b) Initial structure for geometry optimization: the proton is bound to Asp119, the oxazoline is neutral (red dot in the free energy profile). (c) The optimized structure, in which the proton has transferred to the N_x atom of the substrate (green dot).

The resulting free energy profile (Figure 4.8) displays a minimum that corresponds to the protonated intermediate (Glc-ox$^+$) in combination with deprotonated Asp119 (CV_1 = 1.6 Å). The alternative state (Glc-ox) is located at CV_1=1.1 Å, lying on a shoulder 4.2 kcal/mol higher in energy than the global minimum, thus it is less preferred. In addition, Glc-ox$^+$ was obtained even after a simple geometry optimization starting from Glc-ox/Asp-OH. Thus, Glc-ox$^+$ is the most stable species in the reaction intermediate of *Tt*OGA active site, consistent with our previous equilibrium QM/MM simulations and classical MD simulations.

These results obtained for *Tt*OGA (this chapter), *Sm*ChiB and *Bv*Hya (previous chapter), suggest that there are at least two distinct substrate-assisted mechanisms among GHs: one in which the assisting residue acts as an acid/base, protonating and deprotonating the NAc group, leading to a Glc-ox intermediate (families GH18[128] and GH56) and another one where the assisting residue plays a stabilizing role, forming an ionic bond with the Glc-ox$^+$ (GH84). The basis of this observation lies in the pK_a of the assisting residue, which depends on its spatial organization and environment. For GH families in which there is an insertion between the catalytic residues (Asp-X-Glu), the protein backbone bends allowing a hydrogen bond interaction between both catalytic residues that increases the pK_a of the assisting residue. This generates a deprotonated Glc-ox reaction intermediate, as observed in *Sm*ChiB and *Bv*Hya from families GH18 and GH56, respectively (chapter 3). In contrast, GH families in which the catalytic residues are adjacent in the sequence — as occurs with *Tt*OGA from family GH84 (Asp119-Asp120) — these residues cannot interact, thus the instrumental hydrogen bond network that facilitates deprotonation of the intermediate in GH18 *Sm*ChiB is not formed. Moreover, the assisting Asp119 in GH84 is involved in an ionic interaction with Lys43, which lowers its pK_a and precludes Asp119 from deprotonating the Glc-ox$^+$ species. The presence of a Lys residue similarly positioned in the active site is also observed in other X-ray structures from GH84 enzymes such as *h*OGA or *Og*OGA (PDB entries 5VV0 and 2XSB, respectively). Hence, the architecture of the active site in GHs that perform catalysis via a substrate-assisted mechanism modulates the pK_a of the assisting residue, which ultimately determines the nature of the reaction intermediate (Figure 4.9).

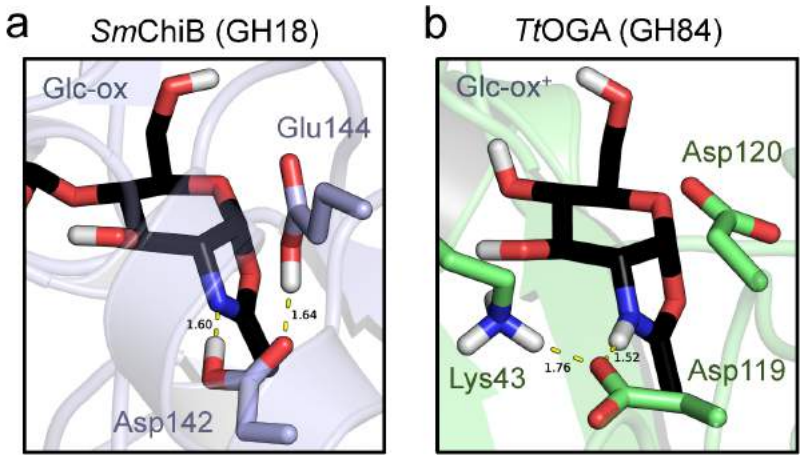

Figure 4.9. Comparison between (a) *Sm*ChiB from family GH18 and (b) *Tt*OGA from family GH84 active site architectures after optimization at QM/MM level.

4.3. Conclusions

- The single point mutation of the acid/base residue from Asp to Asn in OGA enzymes (*Tt*OGA and *Sp*OGA) converts these hydrolytic enzymes (GHs) in synthetic phosphorylases (GPs) that can be used to produce valuable phosphorylated glycosides.[160]

- Multiscale calculations were able to rationalize this novel case of enzyme engineering, showing that the aforementioned mutation modifies both the electrostatic potential of the OGA active site and its architecture, allowing phosphate ions to approach and react. This expands the understanding of what distinguishes GHs from GPs, providing a basis for the generation of novel enzymatic functions via mutagenesis.

- The nature of the reaction intermediate in GH84 enzymes was revealed by means of QM/MM metadynamics, showing that the protonated state of the substrate (Glc-ox^{+}) is the most stable. Thus, two possible substrate-assisted mechanisms take place in glycosidases, which can be explained by the impact of the environment of the assisting residue.

4.4. Computational details

Systems preparation

The initial structure was taken from the X-ray structure of the *Tt*OGA-Asp120Asn variant in complex with the glycopeptide hTab1-O-GlcNAc (PDB entry 5DIY).[155] The Asp120Asn mutation was reverted *in silico* to regenerate the native *Tt*OGA enzyme. We removed the peptide, except the GlcNAc residue, which served as a scaffold to build the glucose oxazoline/oxazolinium moiety (Glc-ox/Glc-ox$^+$). The GlcNAc to Glc-ox/Glc-ox$^+$ conversion was driven by minimizing the energy in the equilibration step (classical MD calculations). In the simulations with a phosphate ion in the active site, its initial position was generated by substituting the O and C atoms of the bound side chain Ser residue of the hTab1-*O*-GlcNAc glycopeptide by the O and P phosphate atoms, respectively.

The protonation state of all residues was assigned considering the experimental pH of 7.8. The protonation of His residues was further assessed according their chemical environment. Specifically, His residues 20, 48, 112, 127, 200, 206, 242, 248, 250 were considered neutral with their proton located at N_δ; while His236 were protonated at N_ε. Crystallographic water molecules were retained. To compute the distribution of phosphate ions around the active site, we tested different models in which an appropriate number of Na^+ and phosphate ions were added to the system to achieve a concentration of 20 mM phosphate and reach neutrality. Additional water molecules were added to solvate the protein and generate a simulation box of 86.4 × 97.6 × 132.7 Å^3. The composition of each system is specified in Table S4.4.

Classical MD calculations

The AMBER14 software[161] was used to run all classical MD calculations. The Amber ff14SB[75] and TIP3P[80] force fields were employed to describe the protein and water molecules, respectively. Phosphate ions (HPO_4^{2-} and $H_2PO_4^-$) were described with parameters adapted from phosphorylated serine parameters.[162] Glc-ox and Glc-ox$^+$ were parametrized by using the antechamber module,[133] in conjunction with GAFF[134] parameters and the RESP atomic charges obtained from first principles calculations at the HF/6-31G* level of theory employing Gaussian09.[135]

All systems were equilibrated in periodic boundary conditions, using the following protocol. First, the energy was minimized with 20,000 minimization cycles, relaxing Na^+ ions and water molecules. Subsequently, the entire system was relaxed with 20,000 minimization

cycles. Afterwards, the system was heated gradually to 100, 200, 250 and 300 K in the NVT ensemble at intervals of 50 ps. Spatial restraints were applied to the protein and ligands during the first heating interval, while all restraints were released after reaching 100K. Subsequently, the density was converged up to water density at 300 K during 100 ps in the NPT ensemble. A short equilibration run of 150 ps was performed to equilibrate the RMSD in the NVT ensemble with a time step of 1 fs. By employing the SHAKE algorithm,[136] the time step was subsequently increased to 2 fs. The simulations were further continued for 10 ns or 50 ns, depending on the calculation (reactivity studies or phosphate density analyses, respectively). Analysis of the trajectories was carried out using standard tools of AMBER and VMD.[137]

QM/MM MD calculations

The method developed by Laio et al.[95] was employed to perform all QM/MM MD simulations. Three different systems were set up: wild-type *Tt*OGA with either $H_2PO_4^-$ or water molecule located above the Glc-ox$^+$ substrate (in order to study phosphorylation and hydrolysis reactions, respectively); and *Tt*OGA-Asp120Asn in complex with the HPO_4^{2-} ion in a reactive configuration. Snapshots from the classical MD simulations were taken as starting configurations. The QM region for all systems included the Glc-ox$^+$ substrate, the side chain of Asp119 and Asp120/Asn120 (saturating the C_α atom of the $C_\alpha \cdots C_\beta$ bond) and the molecule that performs the nucleophilic attack on the anomeric carbon (either water, HPO_4^{2-} or $H_2PO_4^-$). The MM region included the rest of the protein and the solvent. The parameters for each system are shown in Table S4.5. Density Functional Theory (DFT) was used to describe the electronic structure, employing the Perdew, Burke, and Ernzerhoff generalized gradient-corrected approximation (PBE).[138] A plane wave basis set was employed to expand the Kohn-Sham orbitals, with a kinetic energy cut-off of 70 Ry. Norm-conserving *ab initio* pseudopotentials were used, generated within the Troullier-Martins Scheme.[139] The following protocol was followed in our QM/MM MD calculations. The structure was optimized with annealing of the nuclei until reaching a maximal component of the nuclear gradient of $5 \cdot 10^{-4}$ a.u. Afterwards, 3 ps of MD were performed to re-equilibrate the system at 300 K using the Nosé–Hoover thermostat.[140,141] A snapshot of these simulations was used as a starting point for further metadynamics calculations.

Metadynamics simulation of enzymatic reactions by *Tt*OGA and *Tt*OGA-Asp120Asn

All metadynamics calculations were performed employing the PLUMED plugin.[103] All reactions were activated using atomic distances as CVs, which are schematically represented

in Figure S4.2. In the case of the reactions catalysed by native *Tt*OGA, two distances were used as CVs. The first CV (CV_1) accounts for the nucleophilic attack of the oxygen atom of either water or $H_2PO_4^-$ (hydrolysis and phosphorylation, respectively) on the anomeric carbon of Glc-ox$^+$ (C1). The second CV (CV_2) accounts for the proton transfer from water or $H_2PO_4^-$ to the carboxylic oxygen of the acid/base residue (Asp120). In the case of the phosphorylation reaction catalysed by *Tt*OGA-Asp120Asn, only CV_1 was used, since no proton transfer event is necessary (moreover, Asn120 cannot accept a proton from the phosphate). The parameters used in the metadynamics simulations are collected in Table S4.6. Relevant distances and charges of the species along the enzymatic reactions are listed in Tables S4.1-S4.3. The conformations of pyranose ring of the substrate for the different reactions studied in this chapter are shown in the Mercator representation in Figure S4.6, considering average structures of the relevant states along each enzymatic reaction.

Analysis of substrate protonation state

Starting from the equilibrated structure of *Tt*OGA in complex with Glc-ox$^+$, we computed the free energy change with respect to transferring the proton from the nitrogen of the Glc-ox$^+$ to the assisting residue Asp119. We used metadynamics with one CV that accounts for the aforementioned proton transfer (Figure S4.2). The parameters used in the metadynamics simulations are specified in Table S4.6. The resulting free energy profile (Figure 4.8) shows that the pair Asp119-COOH/Glc-ox is 4.2 kcal/mol higher in energy than Asp119-COO$^-$/Glc-ox$^+$ (Figure 4.8). Therefore the proton prefers to stay in the substrate rather than being transferred to Asp119. In the same vein, the substrate remained protonated >99.7% of the time in all the simulations performed (three enzymatic reactions). All these results disqualify the possibility that a neutral oxazoline as the intermediate of the enzymatic reaction, hence supporting the proposed oxazolinium ion intermediate (Figure 4.2b).

4.5. Supplementary material

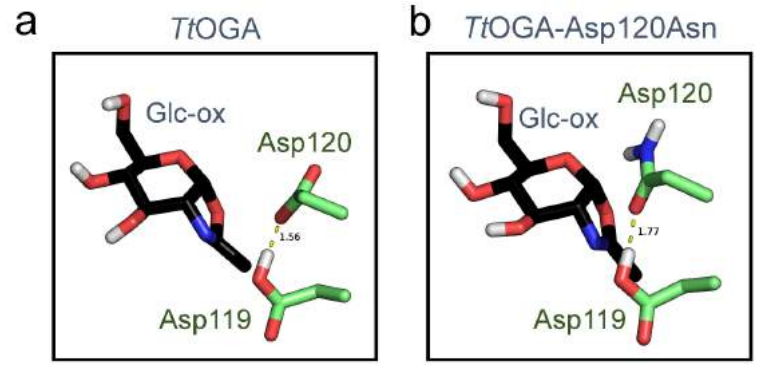

Figure S4.1. Resulting structures for *Tt*OGA and *Tt*OGA-Asp120Asn active sites obtained by classical MD considering a Glc-ox protonation state. The interaction between Asp119 and the nitrogen from Glc-ox is lost.

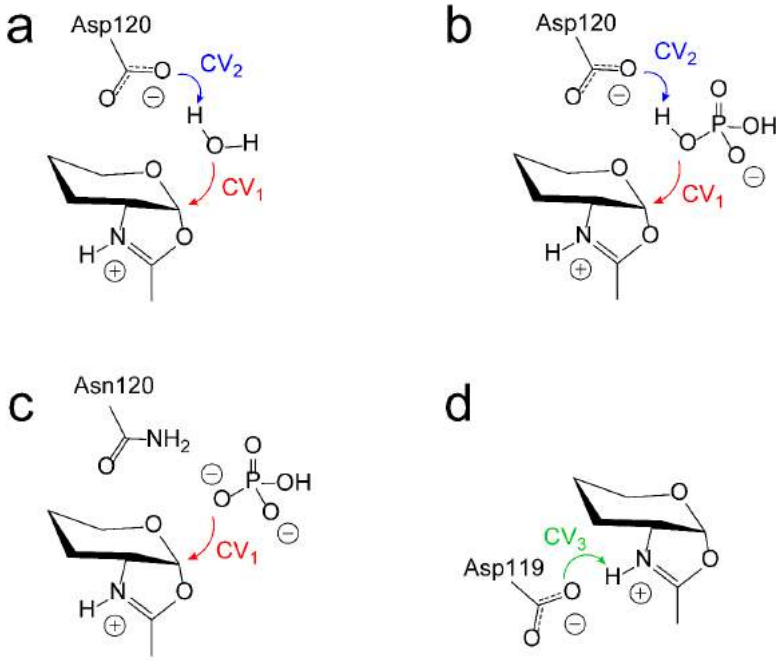

Figure S4.2. CVs used in this work to study: (a) hydrolysis, (b) phosphorylation by wild-type *Tt*OGA, (c) phosphorylation by *Tt*OGA-Asp120Asn, and (d) the protonation state of the bicyclic reaction intermediate in wild-type *Tt*OGA.

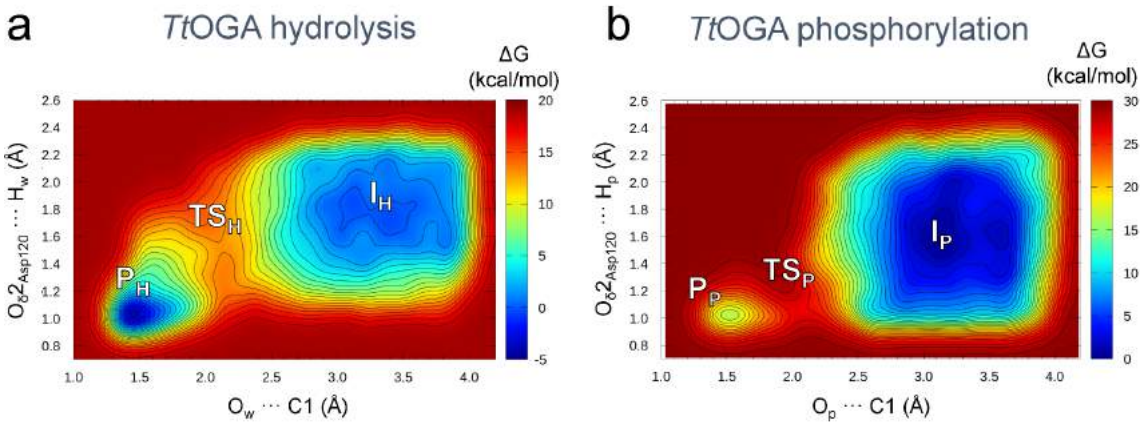

Figure S4.3. FES reconstructed from the metadynamics simulations of (a) hydrolysis and (b) phosphorylation reactions of Glx-ox$^+$ by *Tt*OGA. Isolines are at 1 kcal/mol. The two employed CVs account for the deprotonation of either water or phosphate by Asp120, and the nucleophilic attack of either the water or phosphate oxygens on C1, respectively.

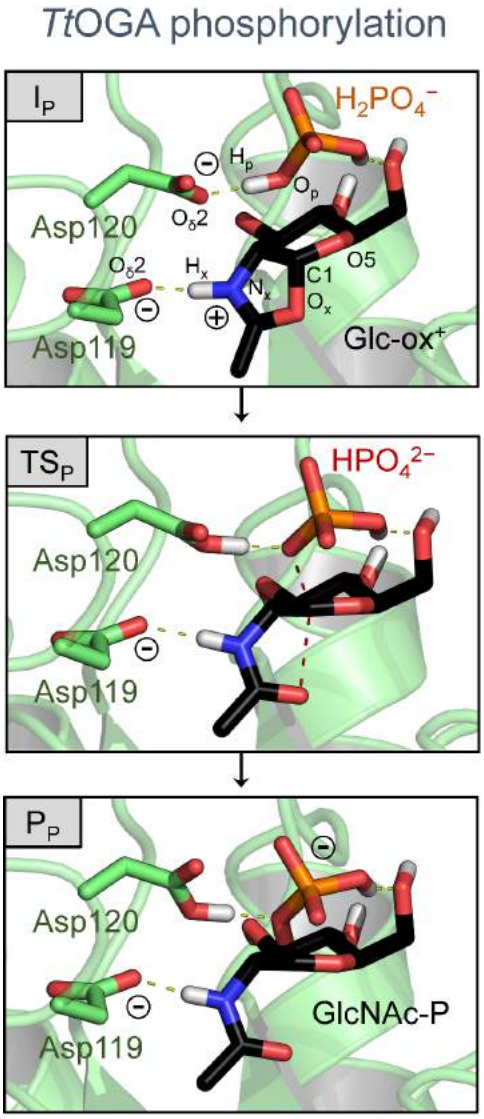

Figure S4.4. Average structures of the main states along the reaction coordinate for *Tt*OGA phosphorolysis (I=Glc-ox$^+$, TS=transition state, P=reaction products). P subscript corresponds to phosphorylation. Hydrogen atoms are omitted for clarity, except those attached to heteroatoms. Bonds being formed/broken are indicated with red dashed lines. The FES of this metadynamics calculation is found in Figure S.4.3.

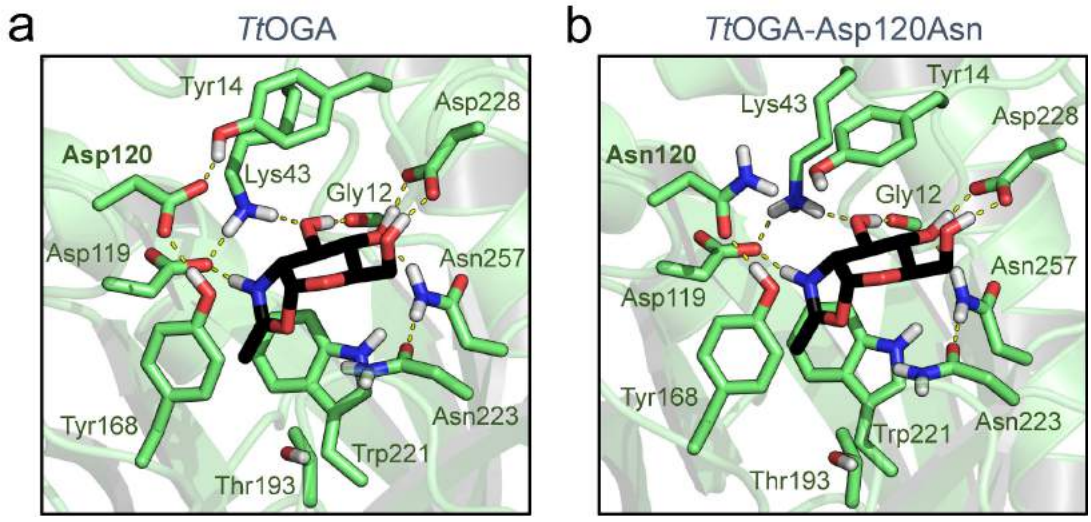

Figure S4.5. Active site architecture of (a) *Tt*OGA wild-type and (b) TtOGA-Asp120Asn enzymes. The main interactions between the enzyme and the Glc-ox^{+} substrate are shown. Side chains of Tyr168, Thr193 and Trp221 provide a hydrophobic cavity for the methyl group of the substrate. The OH3 of Glc-ox^{+} interacts with Lys43 and the backbone carbonyl of Gly12, while OH4 and OH6 interacts by hydrogen bonding with both oxygens of the carboxylate group of Asp228. The assisting residue Asp119 establishes ionic interactions with the nitrogen of the oxazolinium ion ring and the side chain of Lys43. The main difference between *Tt*OGA and *Tt*OGA-Asp120Asn lies in the environment of the acid/base residue (Asp120 and Asn120, respectively). In *Tt*OGA, Asp120 is oriented for catalysis and stabilized by the phenolic hydroxyls of both the Tyr14 and Tyr168 residues, while in *Tt*OGA-Asp120Asn, Asn120 interacts only with Tyr168. The NH group of the amide Asn120 can form a hydrogen bond with the incoming HPO_4^{2-}.

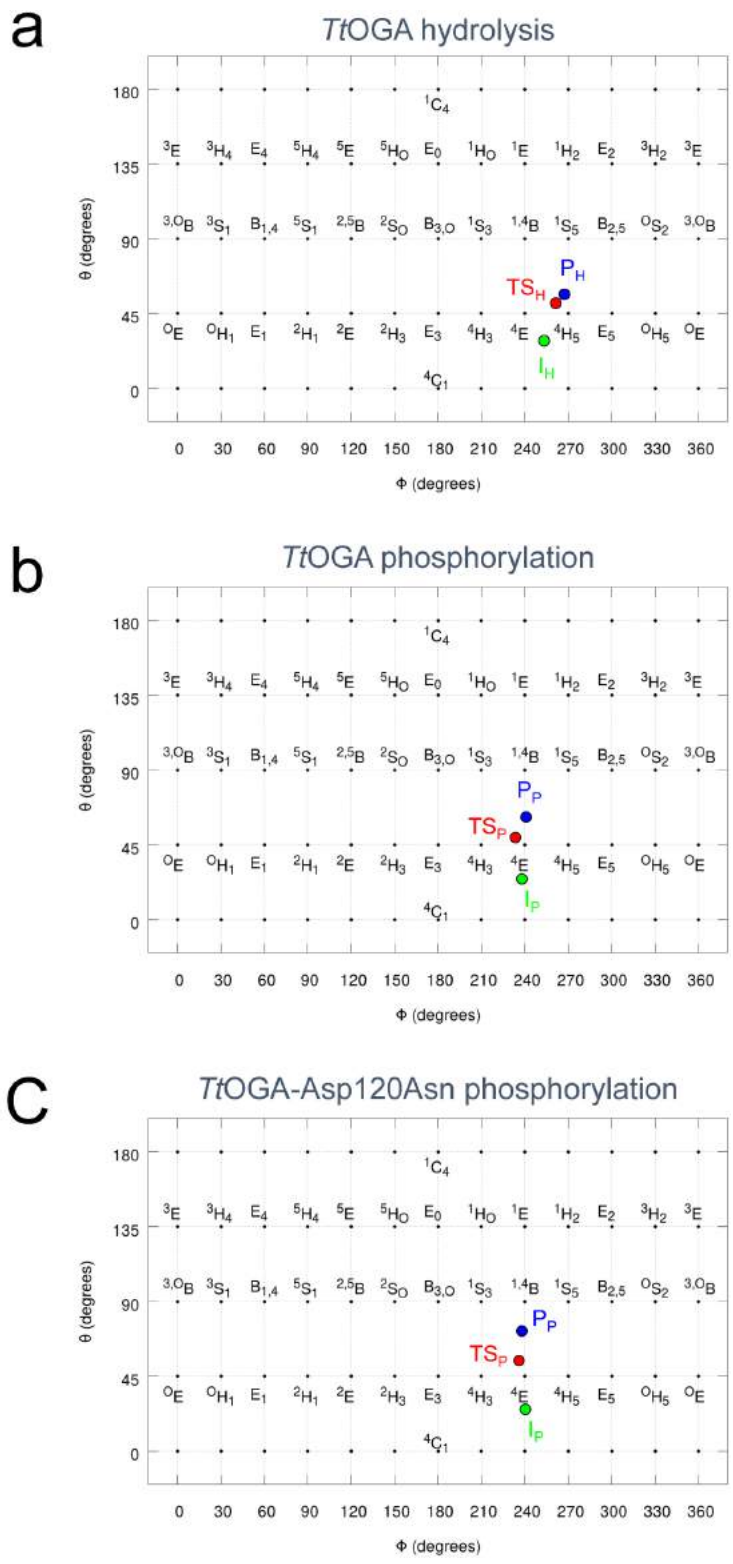

Figure S4.6. Mercator representation of the puckering coordinates calculated for the averaged structures of the main states along the enzymatic reactions: (a) hydrolysis, (b) phosphorylation by wild-type *Tt*OGA, and (c) phosphorylation by *Tt*OGA-Asp120Asn. I = Glc-ox^{+} intermediate, TS = transition state, P = products. H and P subscripts correspond to hydrolysis and phosphorylation, respectively.

Table S4.1. Relevant distances (mean and standard deviation) of each representative structure along the reaction pathway of hydrolysis of Glc-ox$^+$ by *Tt*OGA.

Distance (Å)	**I_H**	**TS_H**	**P_H**
C1···O_w	3.39 ± 0.05	2.10 ± 0.05	1.46 ± 0.02
C1···O_x	1.60 ± 0.07	2.40 ± 0.23	3.05 ± 0.26
C1···O5	1.35 ± 0.03	1.32 ± 0.03	1.43 ± 0.05
O_w···H_w	1.01 ± 0.03	1.07 ± 0.02	1.60 ± 0.06
H_w···$O_\delta2_{Asp120}$	1.75 ± 0.06	1.53 ± 0.03	1.03 ± 0.02
H_x···$O_\delta2_{Asp119}$	1.58 ± 0.10	1.90 ± 0.28	1.80 ± 0.03
H_x···N_x	1.08 ± 0.05	1.03 ± 0.06	1.06 ± 0.03
O_x···$H_\delta2_{Asn223}$	2.78 ± 0.28	2.11 ± 0.11	2.30 ± 0.20
$H_{\eta Tyr14}$···$O_\delta2_{Asp120}$	1.84 ± 0.12	1.84 ± 0.08	1.93 ± 0.10
$H_{\eta Tyr168}$···$O_\delta1_{Asp120}$	1.88 ± 0.16	1.84 ± 0.12	1.91 ± 0.12

Table S4.2. Relevant distances (mean and standard deviation) of each representative structure along the reaction pathway of phosphorylation of Glc-ox$^+$ by *Tt*OGA.

Distance (Å)	**I_P**	**TS_P**	**P_P**
C1···O_p	3.15 ± 0.02	2.10 ± 0.04	1.51 ± 0.02
C1···O_x	1.63 ± 0.07	2.59 ± 0.08	2.85 ± 0.07
C1···O5	1.34 ± 0.03	1.29 ± 0.03	1.40 ± 0.03
O_p···H_p	1.05 ± 0.04	1.53 ± 0.12	1.96 ± 0.09
H_p···$O_\delta2_{Asp120}$	1.56 ± 0.04	1.05 ± 0.01	1.03 ± 0.02
H_x···$O_\delta2_{Asp119}$	1.60 ± 0.14	1.94 ± 0.13	1.79 ± 0.18
H_x···N_x	1.09 ± 0.03	1.05 ± 0.04	1.03 ± 0.02
O_x···$H_\delta2_{Asn223}$	2.34 ± 0.22	1.91 ± 0.07	1.87 ± 0.10
$H_{\eta Tyr14}$···$O_\delta2_{Asp120}$	2.61 ± 0.35	2.64 ± 0.12	2.75 ± 0.13
$H_{\eta Tyr168}$···$O_\delta1_{Asp120}$	2.98 ± 0.69	3.08 ± 0.47	3.22 ± 0.12

Table S4.3. Relevant distances (mean and standard deviation) of each representative structure along the reaction pathway of phosphorylation of Glc-ox$^+$ by *Tt*OGA-Asp120Asn.

Distance (Å)	I_P	TS_P	P_P
C1···O_p	3.40 ± 0.08	2.17 ± 0.01	1.47 ± 0.03
C1···O_x	1.66 ± 0.09	2.55 ± 0.12	3.07 ± 0.13
C1···O5	1.34 ± 0.03	1.29 ± 0.02	1.40 ± 0.04
O_p···$H_{\eta Tyr14}$	1.94 ± 0.11	2.03 ± 0.21	2.12 ± 0.16
O_p···$H\delta2_{Asn120}$	1.95 ± 0.33	2.95 ± 0.19	3.11 ± 0.23
H_x···$O\delta2_{Asp119}$	1.68 ± 0.15	2.19 ± 0.13	2.38 ± 0.27
H_x···N_x	1.07 ± 0.04	1.04 ± 0.03	1.03 ± 0.03
O_x···$H\delta2_{Asn223}$	2.59 ± 0.38	2.11 ± 0.14	2.01 ± 0.14
$H_{\eta Tyr168}$···$O\delta2_{Asn120}$	1.92 ± 0.15	1.81 ± 0.09	1.89 ± 0.11
O_p2···$H\delta2_{Asn120}$	3.63 ± 0.27	2.30 ± 0.71	2.37 ± 0.49

Table S4.4. Computational set up for all systems analyzed in this chapter.

Simulation	Enzyme	Reaction intermediate	HPO_4^{2-}/ $H_2PO_4^-$	Na^+	nº atoms	Production run (ns)
Equilibration of the reaction intermediate complex with classical MD	*Tt*OGA	Glc-ox	1/0	24	93553	20
	*Tt*OGA	Glc-ox	0/1	23	93559	20
	*Tt*OGA	Glc-ox$^+$	1/0	24	93553	20
	*Tt*OGA	Glc-ox$^+$	0/1	23	93559	50
	*Tt*OGA-Asp120Asn	Glc-ox	1/0	23	93557	20
	*Tt*OGA-Asp120Asn	Glc-ox	0/1	22	93560	20
	*Tt*OGA-Asp120Asn	Glc-ox$^+$	1/0	23	93557	50
	*Tt*OGA-Asp120Asn	Glc-ox$^+$	0/1	22	93560	20
Phosphate density	*Tt*OGA	Glc-ox$^+$	9/2	35	93555	50
	*Tt*OGA-Asp120Asn	Glc-ox$^+$	9/2	34	93559	50

Table S4.5. QM/MM MD parameters for the three systems studied in this chapter.

System	Total nº atoms	QM nº atoms	QM box (Å)	NN/ MIX/ESP radii (Å)	Timestep (fs)	Electronic mass (a.u.)
*Tt*OGA + H_2O	93561	45	14.26×14.17× 18.45	9.00/13.23/15.88	0.12	600
*Tt*OGA + $H_2PO_4^-$	93561	49	15.45×15.56× 18.33	7.94/10.58/15.88	0.12	600
*Tt*OGA-Asp120Asn + HPO_4^{2-}	93559	50	13.26×17.08× 18.22	6.35/10.58/15.88	0.12	600

Table S4.6. Metadynamics parameters for hydrolysis/phosphorylation by *Tt*OGA and phosphorylation by *Tt*OGA-Asp120Asn mutant.

System	Collective variable (CVs)	Gaussian height (kcal/mol)	Gaussian width (Å)	Deposition time (MD steps)
*Tt*OGA + H_2O	$CV_1 = O_w\cdots C1$; $CV_2 = H_w\cdots O_{Asp120}$	0.6	0.10; 0.10	300
*Tt*OGA + $H_2PO_4^-$	$CV_1 = O_p\cdots C1$; $CV_2 = H_p\cdots O_{Asp120}$	1.0	0.10; 0.10	400
*Tt*OGA-Asp120Asp + HPO_4^{2-}	$CV_1 = O_p\cdots C1$	0.6	0.10	400

5. Asparagine tautomerization in glycosidase catalysis?

Abstract: Endo-β-*N*-acetylglucosaminidases (ENGases) are family GH18 or GH85 glycosidases that cleave the chiotobiose core of *N*-glycans. This post-translational modification affects critical protein properties such as stability, conformation, or circulatory time. Whereas GH18 enzymes have been extensively characterized, little is known about GH85 ENGases from a mechanistic point of view. All the available X-ray structures came out last decade, showing important discrepancies such as the conformation of the catalytic dyad. These enzymes are known to operate via substrate-assisted catalysis. However, the assisting residue in GH85 enzymes is an Asn instead of an Asp, in contrast with other *N*-acetyl-glycosidases that follow substrate-assisted catalysis. Thus, not only the nature of the reaction intermediate (Glc-ox or Glc-ox$^+$) remains unknown for ENGases, but also the behaviour of the catalytic Asn. Despite it is rarely considered in proteins, tautomerization of the assisting Asn has been invoked. Here we accurately built the Michaelis complex of *Sp*GH85 with the *N*-glycan core in order to study the first step of the reaction mechanism by means of QM/MM metadynamics. Our results show that both Glc-ox and Glc-ox$^+$ are stable in the active site and that amide/imidic acid Asn335-mediated mechanisms may coexist.

5.1. Introduction

N-glycosylation is crucial in many relevant biological processes such as protein folding, cell-cell interaction or the immune response.[163] *N*-glycans can be classified in three groups depending on their structure: high mannose, hybrid and complex. Despite their structural differences, all of them share the same $(Man)_3(GlcNAc)_2$ core attached to an Asn residue (Figure 5.1). The hydrolysis of the β-1,4 linkage between both GlcNAc units of the core of these oligosaccharides (red arrow in Figure 5.1) is catalysed by endo-β-*N*-acetylglucosaminidases or ENGases (EC 3.2.1.96).[164] Once the glycosidic bond is cleaved, the innermost GlcNAc remains linked to the side chain of the Asn residue of the protein (GlcNAc-Asn); while the rest of the oligosaccharide is released. ENGases are classified within GH18 and GH85 families, both belonging to the GH-K clan and exhibiting similar $(\beta/\alpha)_8$-TIM barrel folds. GH85 enzymes are widely distributed in animals,[165] plants,[166] fungi[167] and bacteria.[168-170]

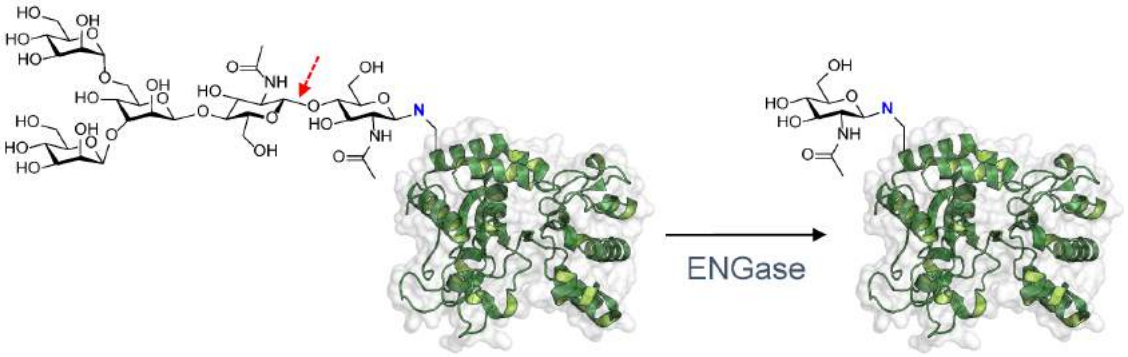

Figure 5.1. Schematic representation of ENGase enzymatic cleavage of the β-1,4 linkage between GlcNAc units in the *N*-glycan core $(Man)_3(GlcNAc)_2$ (red arrow). The structure of the human sialyltransferase is shown as example (PDB entry 5BO6).

A remarkable feature of ENGases is their ability to catalyse transglycosylation reactions, coexisting with their expected hydrolytic activity. For this reason, these enzymes have gained great interest to serve as potential chemoenzymatic tools.[14,171] Despite Nature has already provided synthetic enzymes to form glycosidic linkages, GTs, their use in industry is limited. The reason behind it lies in the fact that GTs frequently are membrane-associated enzymes

that require nucleotide-phosphate substrates and, consequently, they can be challenging to express and expensive to employ on a large scale. Instead, GHs can be used for this purpose, either by using wild-type enzymes together with activated donors[172] or by introducing specific mutations in the active site to generate glycosynthase variants.[54] Examples of GH85 ENGases that have been engineered into efficient glycosynthases include those named Endo-A,[173-175] Endo-D[176] and Endo-M.[54,177,178]

Understanding ENGases is critical in the fight against pathogenic organisms such as *Streptococcus pneumoniae*, a Gram-positive bacterium that causes major human infections. This pathogen employs a battery of GHs to degrade glycans from host proteins, such as glycosylated antibodies, in order to acquire sugar nutrients and deactivate the immunological response.[179,180] Thus, the glycan degrading enzymes from *Streptococcus pneumoniae*, such GH85 ENGase, [62,181] are considered virulence factors and targets for drug discovery.[182]

Because of their potential use as competent chemoenzymatic tools and their relevance in human health, a mechanistic description of ENGases from family GH85 is of utmost importance. In 2009, X-ray crystal structures of these enzymes shed light on the understanding of their active site architecture and mechanism of action.[62,175,183] GH85 ENGases perform catalysis via a substrate-assisted mechanism with retention of the anomeric configuration (Figure 5.2), where a Glu is the acid/base residue. A remarkable exception compared with mechanistically related enzymes, such as GH18 chitinases and GH84 *O*-GlcNAcases analysed in previous chapters, is the chemical nature of the assisting residue, an Asn, instead of an Asp. This Asn residue is conserved both in sequence and structures within enzymes that follow the substrate-assisted mechanism. Thus, alternatively to an acidic residue, in this case presumably a non-charged polar residue assists catalysis by interacting with the *N*-acetamido group of the GlcNAc substrate. The two residues (the assisting Asn and the acid/base Glu) are linked by another residue, forming a conserved Asn-X-Glu motif, which interacts via a well-defined hydrogen bond network with the substrate. As discussed in previous chapters (see Introduction) this feature is observed within families of GHs that use substrate-assisted catalysis and have an insertion in the sequence between their catalytic residues (chapter 3).[128]

The molecular details of the reaction mechanism of GH85 ENGases remain unknown, but two pathways have been proposed (Paths A and B in Figure 5.2) that differ in two main aspects. First, they differ on the protonation of the substrate in the reaction intermediate, whether its oxazoline ring is protonated or not (Glc-ox$^+$ or Glc-ox, respectively), an aspect that has been also investigated in other GHs that use a substrate-assisted mechanism.[128,184]

Secondly, the proposed mechanisms differ on the initial orientation of the amide group of the Asn assisting residue.

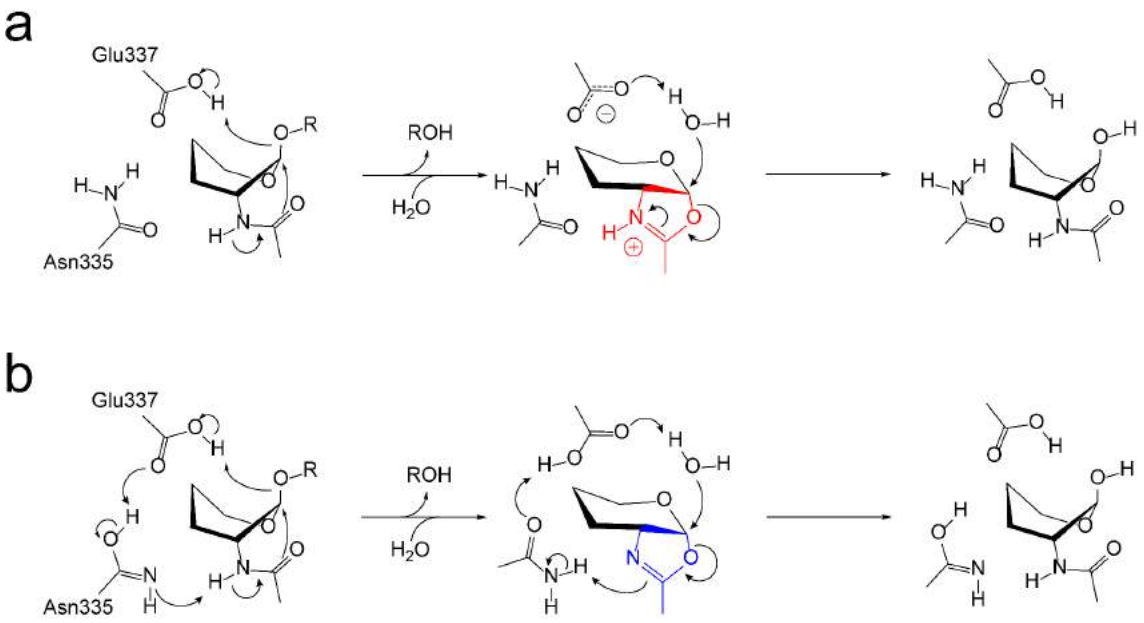

Figure 5.2. The two-steps substrate-assisted reaction mechanism of Endo-D (GH85 enzyme from *Streptococcus pneumoniae*) considering Path A: (a) a neutral Asn and a Glc-ox$^+$ reaction intermediate; or Path B: (b) an imidic acid tautomer of Asn and a Glc-ox intermediate. R corresponds to GlcNAc-Asn. From left to right: Michaelis complex, reaction intermediate and products.

The available structures of Endo-A and Endo-D enzymes correspond to mimics of the reaction intermediate, in the form of the enzyme in complex with a Glc-thiazoline inhibitor complexes.[refs] These structures show that the side chain of the Asn assisting residue interacts with both the carboxyl group of the acid/base residue and the N_x of the Glc-thiazoline inhibitor, if present. However, depending on the X-ray structure, the carboxyl group of the Glu acid/base residue interacts with either the nitrogen of the Asn sidechain (PDB entries 3FHQ and 3GDB) or its oxygen (PDB entries 3FHA, 2W91 and 2W92). Thus, any of the two mechanism of Figure 5.2 can be excluded on the basis of structural information. In fact, even within structures of the same enzyme (Endo-A free and complexed with Glc-thiazoline) show that both rotamers are possible,[183] which suggests that X-ray crystallography is unable to discern the rotameric state of the Asn.

It should be noted that reaction path shown in Figure 5.2b (Path B) would require tautomerization of the Asn assisting residue along the enzymatic reaction. In particular, the reaction involves an imidic acid tautomer in the Michaelis complex and the products of the reaction, actively participating in proton relays during catalysis. In proteins, Asn/Gln

tautomerization from the amide to its imidic acid tautomer is rarely considered under biological conditions since amides are more stable in water than their imidic acid counterparts. However, in a hydrophobic and negatively charged environment, polarization and resonance might trigger the formation of the imidic acid tautomer, which may be more reactive and thus convenient for catalysis. Indeed, tautomerization of Asn/Gln has already been proposed for diverse enzymatic reactions, mainly on the basis of QM/MM calculations but also structural analyses.[185-191] A noteworthy example is the recent investigation of *Pc*Cel45A,[192] a GH that hydrolyses cellulose through several proton transfers involving an Asn in its imidic acid form, as revealed by combining neutron diffraction and high-resolution X-ray crystallography, which allows hydrogen atoms visualization. In summary, the reaction mechanism of GH85 ENGases still hinders catalytic questions that remain to be answered.

Here we analysed Endo-D (1646 residues), an ENGase from *Streptococcus pneumoniae*, as a representative GH85 enzyme system, for which the X-ray structure with the highest resolution among ENGases (1.40–1.65 Å) and a bound substrate is available (PDB entry 2W92).[62] ENGases exhibit different selectivity depending on the composition of the *N*-glycan substrate. In particular, Endo-D hydrolyses oligosaccharides that are elongated from the α-1,6 linkage of the last mannosyl group attached to $(GlcNAc)_2$ core, but not from its α-1,3 bond (Figure 5.1), suggesting that Endo-D acts in synergy with other GHs that first cleave branching carbohydrates in other to fully degrade more complex *N*-glycans.[193,194] The more recent biochemical and structural investigations on Endo-D were performed on the truncated recombinant form of Endo-D (residues 159–807), hereafter referred to as *Sp*GH85.[62] The truncated enzyme is active towards 3-fluoro-4-nitrophenyl-GlcNAc (3F4NP-GlcNAc) and is inhibited by GlcNAc-thiazoline, consistent with the substrate-assisted mechanism. The *Sp*GH85 three-dimensional structure is composed of three domains: *N*-terminal $(\beta/\alpha)_8$ TIM barrel (catalytic domain) where the active site pocket is found, followed by two β-sandwich domains (Figure 5.3a). The protein surface surrounding the active site has several grooves that might accommodate the glycan branches.

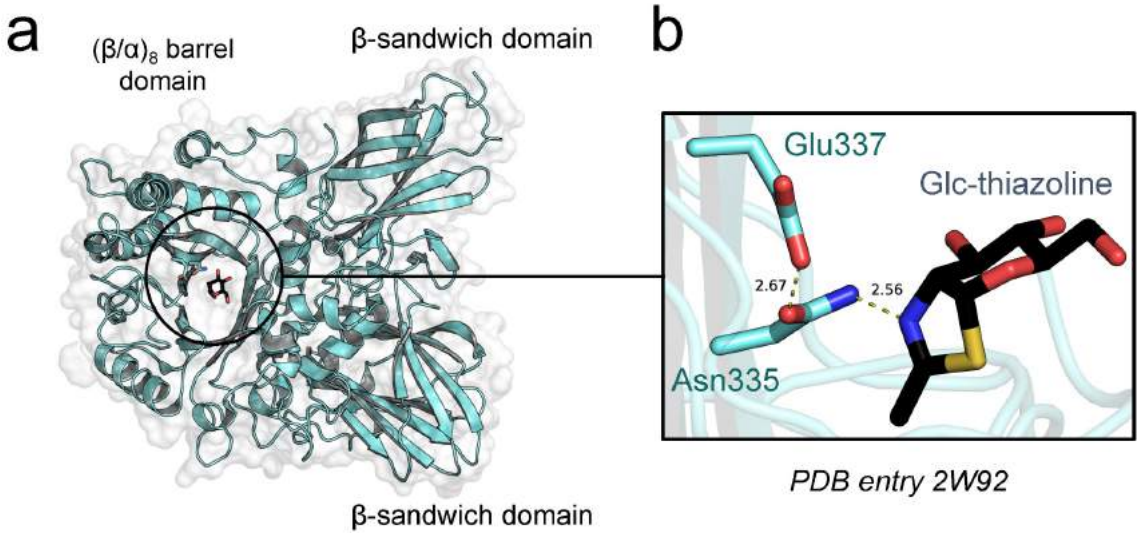

Figure 5.3. X-ray structure of *Sp*GH85 trapped with the inhibitor Glc-thiazoline at the *-1* subsite (PDB entry 2W92). (a) Enzyme three-dimensional structure. (b) Zoom of the active site showing both the inhibitor Glc-thiazoline (black) and the catalytic residues (cyan, Asn335 and Glu337 as assisting and acid/base residues, respectively).

From the above structural data of *Sp*GH85[62] it has been proposed that the substrate-assisted mechanism involves tautomerization of the assisting residue (Asn335) along the reaction and thus the formation of a neutral Glc-ox reaction intermediate (Path B; Figure 5.2b). However, experiments could not discard the alternative mechanism which considers a Glc-ox$^+$ intermediate and the normal amide form of Asn335 (Path A; Figure 5.2a). Path B was proposed based on the following observations: (i) the electron density of Asn335 is more compatible with the orientation involving an imidic acid tautomer than the amide one; (ii) Path A can be excluded because the amide form would not be able to properly stabilize the positively charged Glc-ox$^+$ reaction intermediate with a salt bridge. Instead, a cyclic series of proton transfers between the GlcNAc substrate, Asn335 and Glu337 (acid/base residue) occurring in a concerted manner — together with the consequent changes in pK_a of the groups involved —would be favoured (Figure 5.2b), similarly to *Sm*ChiB from family GH18.[128]

Here we investigated the reaction mechanism of *Sp*GH85 by means of QM/MM metadynamics. Starting from the reaction intermediate mimic obtained experimentally, we reconstructed the Michaelis complex, considering the minimal substrate model (i.e., $(Man)_3(GlcNAc)_2$ *N*-glycan core). Afterwards, we examined the first reaction step (glycosylation) catalysed by *Sp*GH85, considering the two possible mechanisms (Paths A and B).

5.2. Results and discussion

5.2.1. *In-silico* generation of the reaction intermediate in *Sp*GH85

Since no structure of the Michaelis complex (MC) of GH85 enzymes was available, we used a multi-step approach to produce a reliable MC structure. In a first step, we reconstructed the intermediate of the substrate-assisted reaction. The calculations were started from the X-ray structure of the inhibitor Glc-thiazoline bound to *Sp*GH85 (truncated form of Endo-D) at the *-1* subsite, a mimic of the reaction intermediate. In this structure, the $(Man)_3$ trisaccharide of the *N*-glycan core (*-2* and *-3* subsites; Figure 5.3) was missing, as well as the innermost GlcNAc (*+1* subsite, i.e., the one that would be attached to an Asn protein residue). The structure of the $(Man)_3$ was taken from the X-ray crystal of $(Man)_3$-Glc-thiazoline tetrasaccharide bound to Endo-A (PDB entry 3FHQ, Figure S5.1).[183] The GlcNAc unit at *+1* subsite was introduced in the active site of *Sp*GH85 by means of molecular docking (see computational details). Finally, the sulfur atom from the Glc-thiazoline at *-1* subsite was substituted by an oxygen, converting it to Glc-ox/ox^+ (the protonation state of the oxazoline will be analysed later on). In this way, we reconstructed the complete substrate (*N*-glycan core) just after the first step of the enzymatic reaction (Figure 5.2). Concerning active site residues, two cases were considered, as depicted in Figure 5.4. In the first case, a positively charged Glc-ox^+ intermediate interacts with rotamer 1 of Asn335 (Figure 5.4a). In the second case, Asn335 shows the other rotameric form (rotamer 2), which interacts with the neutral Glc-ox substrate (Figure 5.4b).

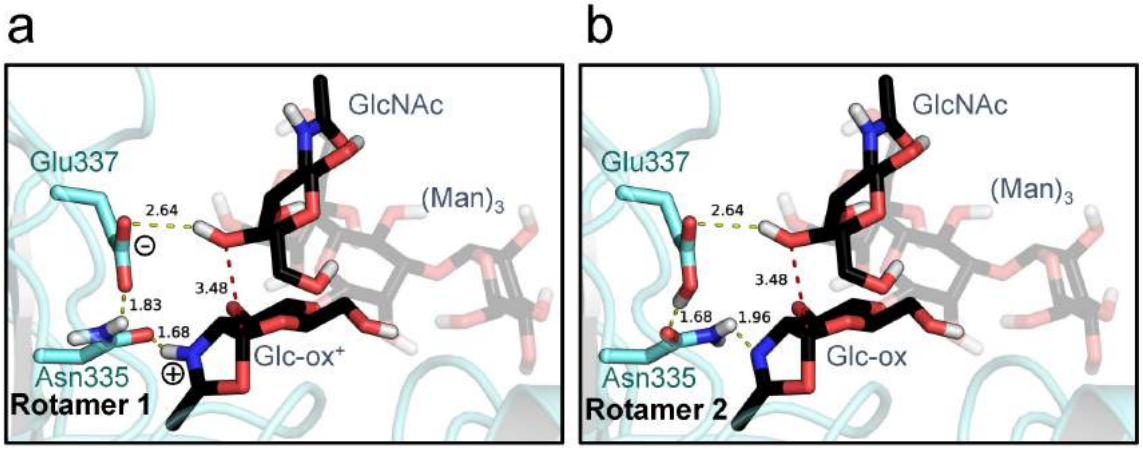

Figure 5.4. Initial structure of *Sp*GH85 in the reaction intermediate considering the two possible Asn rotameric states discussed in the text: (a) rotamer 1 / Glc-ox^+ or (b) rotamer 2 / Glc-ox. Only polar hydrogens are shown for clarity. The $(Man)_3$ fragment located at *-2* and *-3* subsites is shown as transparent. Yellow dashed lines indicate hydrogen bond interactions. See Figure 5.2 for details of both paths.

Two classical MD simulations of 30 ns of *Sp*GH85 were performed considering both Asn rotameric states (Figure 5.4). During the calculations, a restraint was employed to maintain the innermost GlcNAc residue at *+1* subsite, mimicking the real and far more complex scenario in which this residue is linked to another protein through an Asn, thus displaying additional interactions that keeps it in place. Specifically, a restraint was used to ensure that the 4-OH of the GlcNAc at the *+1* subsite interacts with the carboxylate oxygen from Glu337 and also remains at nucleophilic attack distance from the C1 of Glc-ox/ox$^+$, allowing the rest of the GlcNAc to accommodate and explore the available conformational space. The two rotameric states of Asn shown in Figure 5.4 were found to be stable under MD simulations. Noteworthy, a Trp residue (Trp292), was found to be properly oriented to provide stacking interactions to the GlcNAc at *+1* subsite. In particular, the axial C···H bonds from the pyranose ring and the methyl group of the *N*-acetamido group establish CH/π contacts with the Trp292 side chain (Figure 5.5, left). Surprisingly, this Trp residue is neither structurally conserved among the available X-ray structures of GH85 enzymes (*Sp*GH85 and Endo-A, Figure S5.2) nor in their sequences,[195] suggesting that there are different mechanisms to orientate and accommodate the incoming GlcNAc at subsite *+1*. Indeed, it has been demonstrated that despite Trp216 in Endo-A does not correspond in the sequence/structure to Trp292 in *Sp*GH85, it is essential for transglycosylation in Endo-A.[196]

5.2.2. Asn335 in the reaction intermediate: amide or imidic acid?

To discern among the two Asn335 rotameric states (Figure 5.4), as well as investigating the possible occurrence of an imidic acid tautomer in the reaction intermediate, QM/MM calculations were performed. The simulations were initiated from snapshots of the previous MD simulations, considering both Asn335 rotameric states. The QM region included the side chain of both Asn335 and Glu337, part of the Man-Glc-ox/ox$^+$ disaccharide (negative subsites) and the GlcNAc (subsite *+1*). After geometry optimization and 6 ps of QM/MM MD, both systems are stable (i.e. all active site interactions maintained and no proton transfers). This differs from the investigations of *Sm*ChiB and *Tt*OGA (chapters 3 and 4, respectively) in which only one of the two possible reaction intermediates, Glc-ox or Glc-ox$^+$, was found to be stable. Furthermore, we additionally considered the occurrence of an imidic acid tautomer of Asn335 (Figure S5.3) by manually transferring the protons along the hydrogen bond network (i.e., between N_x oxazoline atom, Asn335 and Glu337, Figure S5.4). After QM/MM geometry optimization of both imidic acid Asn355 reaction intermediates, all protons involved in the aforementioned hydrogen bond network transfer to their original

heteroatom and the amide form of Asn335 is recovered. This demonstrates that, independently from the rotameric state, Asn335 only exists in its neutral amide tautomeric form in the reaction intermediate (Figure S5.4).

5.2.3. Generation of the Michaelis complex in *Sp*GH85

We employed QM/MM metadynamics to model the transglycosylation reaction in *Sp*GH85, which leads to the $(Man)_3(GlcNAc)_2$ *N*-glycan core. We included in the QM region the same atoms as in the previous section (5.2.2). The corresponding structure should be equivalent to the MC of the glycosylation reaction (Figure 5.2). Because our objective is to obtain a *bona fide* MC structure rather than investigating the transglycosylation reaction pathway in detail, a "rough" QM/MM metadynamics simulation was performed in which the reaction was driven by employing only one collective variable (CV). This CV was taken as the $C1_{Glc\text{-}ox/ox+}\cdots O4_{GlcNAc}$ distance, which measures the formation of the glycosidic bond. In the following, $O4_{GlcNAc}$ atom will be referred as the glycosidic oxygen (O_g). The QM/MM metadynamics simulation successfully lead to the formation of a stable enzyme in complex with the $(Man)_3(GlcNAc)_2$ *N*-glycan core (Figure 5.5).

The structure of the $(Man)_3(GlcNAc)_2$ *N*-glycan core in the resulting MC is very similar to the one observed in the very recent X-ray structure of Endo-BT (PDB entry 6TCV), Figure S5.6a),[197] a GH18 ENGase enzyme from *Bacteroides thetaiotaomicron*. This structure contains a larger *N*-glycan with the same $(Man)_3(GlcNAc)_2$ core in the active site, which is linked to an Asn and elongated from the terminal Man residues (i.e., R-$(Man)_9(GlcNAc)_2$-Asn). At the time when our simulations were performed, this structure was not yet available. Despite the fact that the GlcNAc located at subsite *-1* exhibits a $^1S_5$ distortion in the structure by Trastoy et al.,[197] similar to the obtained in our QM/MM metadynamics simulation (Figure S5.7), the orientation of the GlcNAc at the *+1* subsite slightly differs from the one we modelled for *Sp*GH85 (Figure S5.6b). This might be due to structural differences in their active sites. In fact, structural alignment shows that the larger R-$(Man)_9(GlcNAc)_2$-Asn acceptor would lead to a steric clash with the positive subsites of *Sp*GH85. In addition, the two catalytic residues were mutated in order to capture the Michaelis complex of Endo-BT (Asp312Ala and Glu314Leu, respectively), which could have altered the architecture of the Endo-BT active site. Indeed, the nitrogen atom of the carbonyl group of the GlcNAc at *-1* subsite is not oriented towards the catalytic residues (Figure S5.6a). In a prior investigation, the same team tried this double mutation in Endo-S, another GH18 ENGase, but obtained the

reaction products containing the *N*-glycan core with only one GlcNAc residue (i.e., the innermost GlcNAc bound to an Asn was not present, Figure S5.6c), which was undistorted.[198]

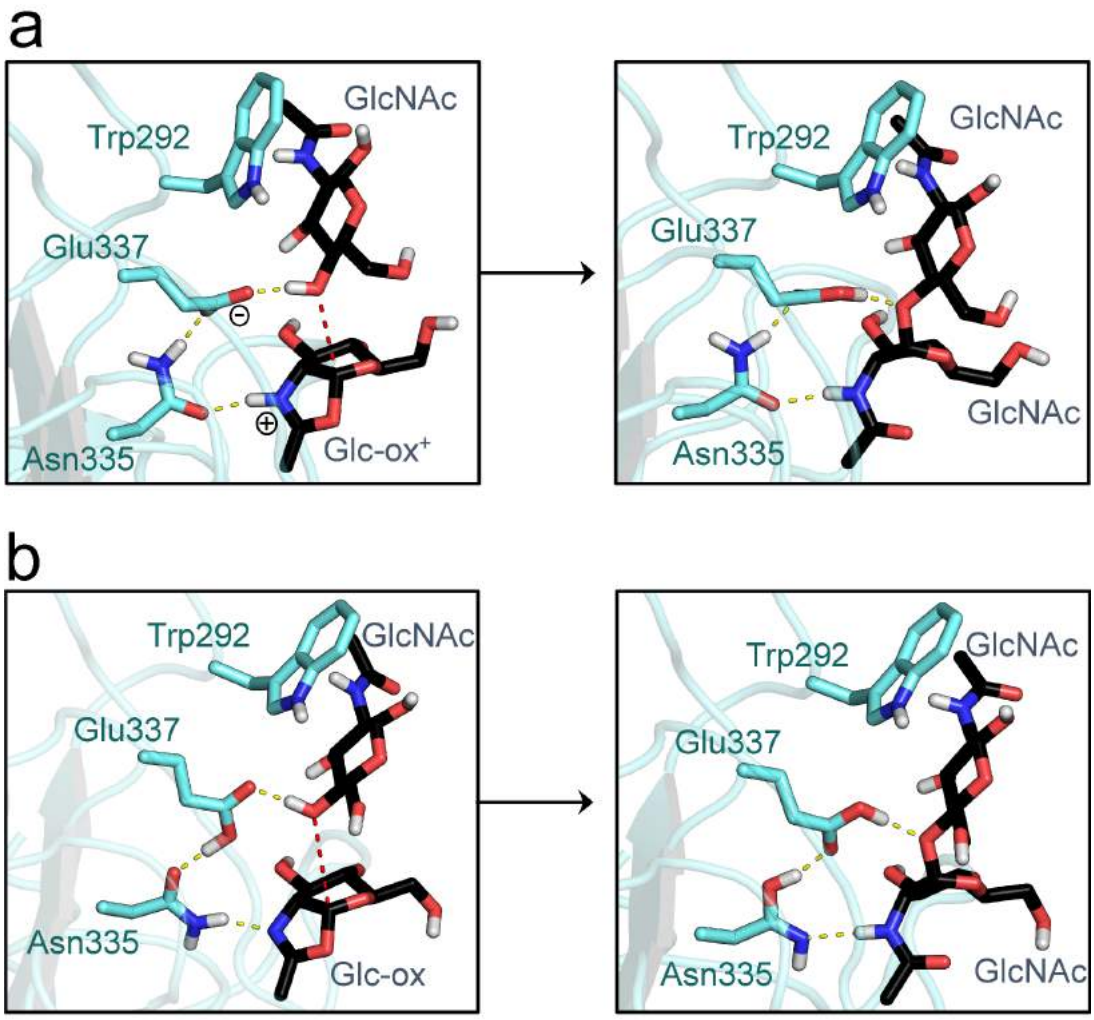

Figure 5.5. Modeling of the Michaelis complex of *Sp*GH85 from the reaction intermediate by using one-dimensional QM/MM metadynamics. The employed CV is shown as a red dashed line. The two possible combinations of the rotameric state of Asn335 and the protonation of the oxazoline were considered. (a) rotamer 1 / Glc-ox$^+$; (b) rotamer 2 / Glc-ox. In the second case, the reaction results in the imidic acid form of Asn335. For the sake of clarity only hydrogens attached to heteroatoms are displayed. The $(Man)_3$ moiety located at *-2* and *-3* subsites is not represented for clarity. Hydrogen bond interactions are shown as yellow dashed lines.

Regarding GH85 enzymes, molecular docking calculations were also employed by others to place the $(Man)_3(GlcNAc)_2$-Asn moiety into an homology model of Endo-Om.[195] In the reported model, the *-1* GlcNAc of the oligosaccharide was not distorted. Altogether highlights the difficulties to experimentally trap Michaelis complexes in ENGase systems (both GH18 and GH85 families) and the limitations of classical methods, such as molecular docking, to properly describe the binding mode of distorted carbohydrates.

5.2.4. The glycosylation reaction catalysed by *Sp*GH85

From the equilibrated Michaelis complex of *Sp*GH85 we examined the first reaction step (glycosylation) considering both reaction paths previously proposed (Path A and Path B, Figure 5.2). The enzymatic reactions were driven by means of the metadynamics approach, using two CVs that take into account the main covalent bonds being broken or formed during the reaction. CV_1 was defined as the difference between the C1···O_x and C1···O_g distances. This CV accounts for the intramolecular nucleophilic attack of the carbonyl oxygen (O_x) and the disruption of the glycosidic bond. CV_2 was defined as the difference between the $H_{\varepsilon}2_{Glu337}$···$O_g$ and $O_{\varepsilon}2_{Glu337}$···$H_{\varepsilon}2_{Glu337}$ distances. This CV accounts for the proton transfer between Glu337 and the glycosidic oxygen (O_g). These variables are equivalent to the ones used to investigate the first step of the reaction catalysed by *Sm*ChiB (chapter 3). The FESs reconstructed from the metadynamics simulation for each path are shown in Figure 5.6a and 5.6b. The position of the stationary points on both FESs in terms of the CV space is similar in both cases, which show a single concerted step (only one **TS**). The reaction mechanism can be analysed by taking snapshots along the minimum free energy pathway connecting reactants and product complexes (**MC** and **I** in Figure 5.6). Surprisingly, the two FESs feature an exergonic reaction path with a low free energy barrier. Path A, which involves Asn335 in its amide form, involves a free energy barrier of just 7.0 kcal/mol, whereas Path B, which starts with Asn335 in its amidic acid form, involves a free energy barrier of just 9.7 kcal/mol. Before analysing the energetic results, let us describe the atomic details of both reaction paths.

***Path A**:* The reaction starts from the Michaelis complex ($\mathbf{MC_A}$) where the hydrogen bond network between Gluc337, Asn335 and GlcNAc at *-1* subsite is observed, the acid/base Glu337 carboxylic acid group points towards O_g and the carbonyl oxygen from the GlcNAc at *-1* subsite is properly positioned to perform the intramolecular nucleophilic attack (Figure 5.7a). To reach the $\mathbf{TS_A}$, the NAc oxygen atom (O_x) approaches the anomeric carbon (the C1···O_x distance changes from 3.06±0.09 Å at $\mathbf{MC_A}$ to 2.73±0.01 Å at TS, Table S5.1) and the glycosidic bond breaks (C1···O_g distance from 1.53±0.05 Å at MC to 2.23±0.11 Å at TS), while, simultaneously, the acid/base residue starts to transfer its proton to O_g ($H_{\varepsilon}2$···O_g distance from 1.61±0.11 Å at MC to). Indeed, at the $\mathbf{TS_A}$ of the reaction the proton is located between Glu337 and O_g (the $H_{\varepsilon}2_{Glu337}$···$O_g$ and $H_{\varepsilon}2_{Glu337}$···$O_{\varepsilon}2_{Glu337}$ distances amount to 1.20±0.05 Å and 1.26±0.05 Å, respectively). At the TS, the glycosidic bond is already broken and the O_x has not collapsed with C1 yet, thus the reaction can be described as dissociative, as found in previous studies of GH mechanisms.[199] The oxocarbenium-like character of the *-1* sugar at the **TS** of the reaction is evidenced by the decrease of the internal C1···O5 distance,

which changes from 1.39±0.05 Å at $\mathbf{MC_A}$ to 1.30±0.05 Å at the $\mathbf{TS_A}$. The pyranose ring is almost planar around the O5-O-C1-C2 atoms (Table S5.3) and the anomeric charge increases by 0.98 e (Table S5.5) at the $\mathbf{TS_A}$, indicating strong oxocarbenium-ion character. Finally, the reaction intermediate (**I**), 14.7 kcal/mol lower in energy than $\mathbf{MC_A}$, forms when the bicyclic sugar is formed (C1···O_x distance of 1.60±0.06 Å) and the proton from Glu337 completely transfers to O_g ($H_\varepsilon 2_{Glu337}$···O_g = 1.06±0.06 Å at $\mathbf{I_A}$).

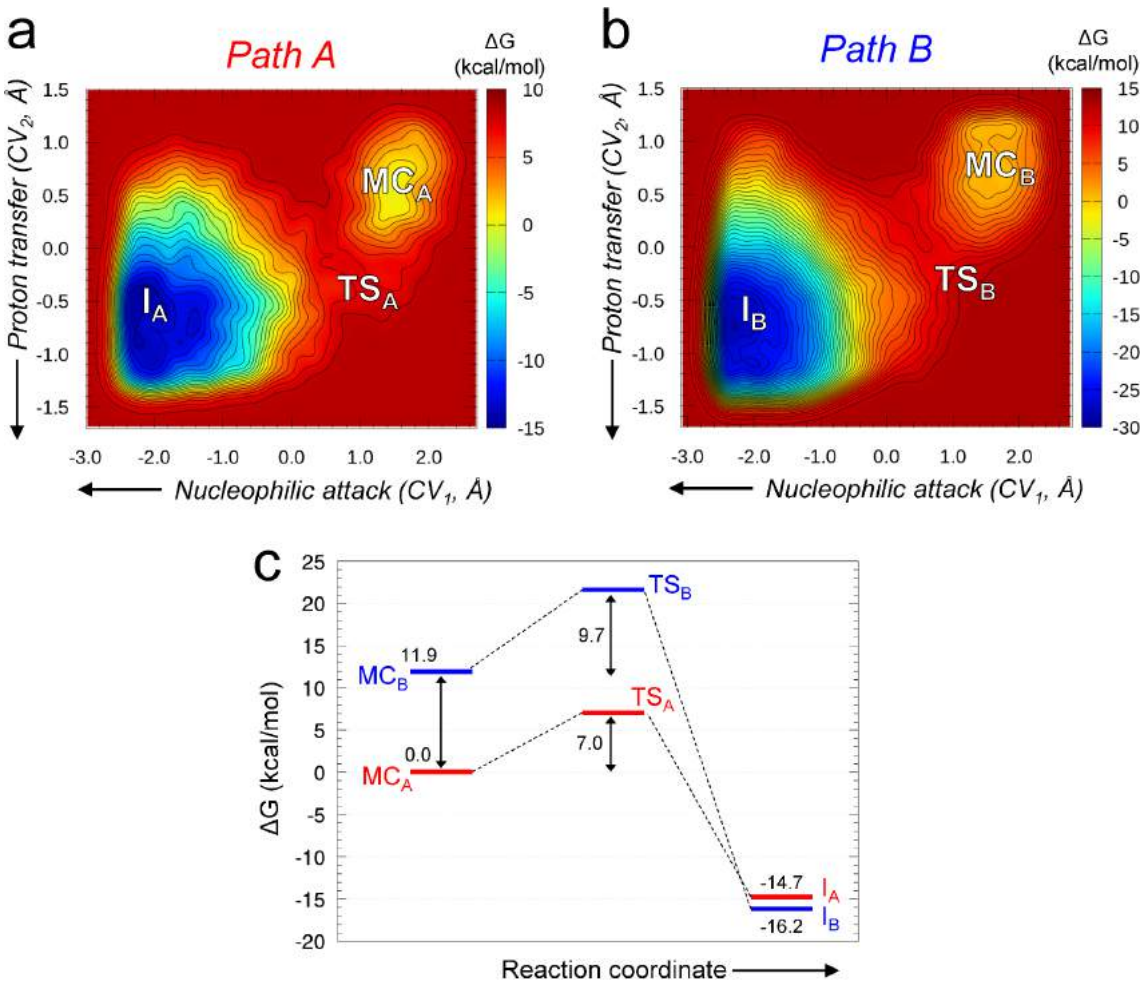

Figure 5.6. Energetics for the first reaction step of *Sp*GH85. (a) FES reconstructed from the metadynamics calculation of path A (amide Asn335 and Glc-ox$^+$) and (b) for path B (acid imidic Asn335 and Glc-ox intermediate). Isolines at 1 kcal/mol. (c) Free energy profile scheme comparing both paths examined in this study and taking into account the initial energetic difference between amide/imidic acid tautomeric states.[200] (MC= Michaelis complex, TS=transition state, I= Glc-ox/ox$^+$). A and B subscripts indicate its corresponding reaction path.

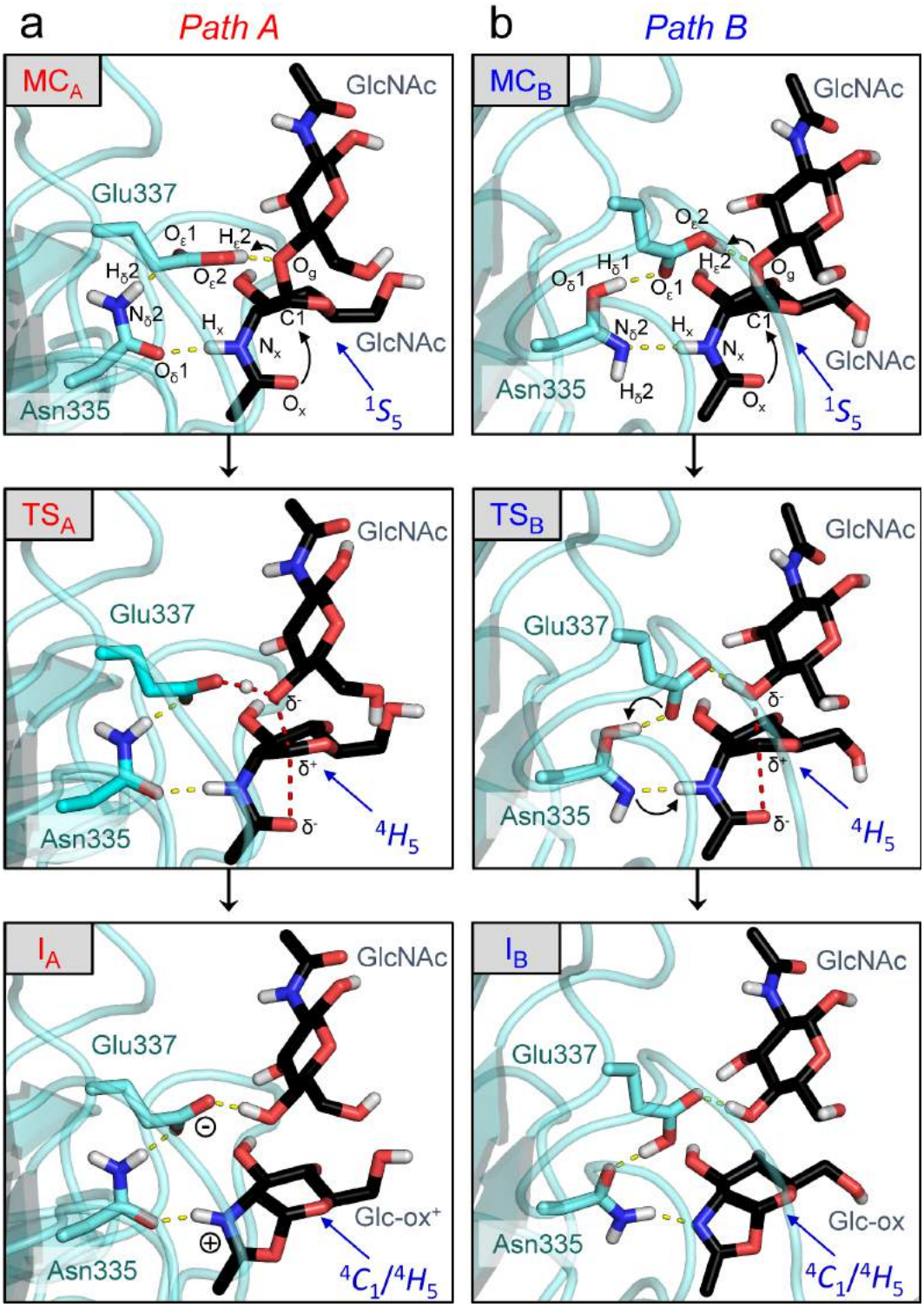

Figure 5.7. Average structures extracted from the main states along the reaction coordinate of the first reaction step by *Sp*GH85, considering paths A (red) and B (blue). (MC= Michaelis complex, TS=transition state, I=Glc-ox/ox+). Hydrogen atoms have been omitted for clarity, except those attached to heteroatoms. Bonds being formed/broken are indicated with red dashed lines. The $(Man)_3$ moiety at the negative subsites has been omitted for the sake of clarity.

***Path B**:* Contrary to the previous path, in this case several proton transfers between the catalytic residues and the ligand are necessary to obtain the Glc-ox intermediate with Asn335 in its amide form (Figure 5.7b). This proton shuttle reassembles what we previously found for *Sm*ChiB (Figure 3.6, chapter 3). The assisting residue Asn335 in Path B is initially in its imidic acid form; thus it adopts the alternative rotameric conformation (rotamer 2) in contrast to the active site configuration of Path A. The reaction starts with the O_x approach to C1 and the protonation of O_g by Glu337, as occurred in Path A. This is evidenced by the change from $\mathbf{MC_B}$ to $\mathbf{TS_B}$ in both C1···O_x and $H_{\varepsilon}2_{Glu337}$···$O_g$ distances (from 3.08±0.08 Å to 2.67±0.17 Å and 1.81±0.05 Å to 1.15±0.02 Å, respectively, Table S5.2). The proton from Glu337 is completely transferred at $\mathbf{TS_B}$, and the glycosidic bond broken (C1···O_g distance of 1.99±0.18 Å at $\mathbf{TS_B}$). The free energy barrier for Path B is 9.7 kcal/mol, just 2.7 kcal/mol higher than Path A (Figure 5.6). The shrinkage of the C1···O5 distance from 1.38±0.03 Å at $\mathbf{MC_B}$ to 1.34±0.04 Å at $\mathbf{TS_B}$, in conjunction with pyranose ring planarity (Table S5.4) and positive anomeric charge development (0.83 e, Table S5.6), manifest the oxocarbenium ion-like behaviour of the $\mathbf{TS_B}$. Once the $\mathbf{TS_B}$ is crossed, Asn335 tautomerizes from its imidic acid to the amide one via two proton transfers (Figure 5.7b and Table S5.2): one proton transfers from the Asn335 oxygen ($O_{\delta}1_{Asn335}$) to the Glu337 carboxylic oxygen ($O_{\varepsilon}1_{Glu337}$); and the one transfers from the *N*-acetamido group of the reactive GlcNAc to the nitrogen atom from Asn335 ($N_{\delta}2_{Asn335}$). With that, Asn335 recovers its amide form and the reaction intermediate Glc-ox **I** is formed, being 28.1 kcal/mol lower in energy than $\mathbf{MC_B}$ (Figure 5.6). Hence, the proton shuttle observed in Path B resembles the one obtained for GH18 *Sm*ChiB.[128]

Noteworthily, we assumed that the $\mathbf{MC_B}$ was 11.9 kcal/mol higher in energy than $\mathbf{MC_A}$, since there is an intrinsic energetic difference between both Asn tautomeric states. This value was computed by quantum chemical calculations of an Asn considering that the tautomerization process was mediated by a water molecule,[200] which may be the case in the solvent exposed active site of *Sp*GH85. However, from our simulations and the structural data available we cannot determine the mechanism in which *Sp*GH85 would generate such Michaelis complex ($\mathbf{MC_B}$). First, in apo and holo X-ray structures of GH85 enzymes both Asn rotamers may exist, but the specific tautomeric form cannot be inferred. Secondly, Asn335 tautomerization to imidic acid required in path B could be influenced by the binding of the substrate (i.e. it can occur during or after the binding process). Moreover, the imidic acid tautomer of Asn335 may be coupled with the transition from rotamer 1 to rotamer 2, which in turn depends on ligand binding as well. Hence, the generation of $\mathbf{MC_B}$ in *Sp*GH85 needs further investigation. Here, we will assume that both $\mathbf{MC_A}$ and $\mathbf{MC_B}$ can take place and focus on how they evolve during the first reaction step.

Considering that both paths have a similar and low free energy barrier (7.0 and 9.7 kcal/mol for Paths A and B, respectively), one would think that none of them is favoured, and thus they may coexist. Direct comparison of the calculated reaction with experimental data (k_{cat}) is not possible since the reaction volume and the enzyme concentration used in the kinetics assays of 3F4NP-GlcNAc (comparable with the first reaction step) have not been reported.[62] The obtained low energy barriers for the glycosylation reaction suggest that the first reaction step might not be rate limiting. Instead, the experimental k_{cat} for the hydrolysis of $(Man)_2$-Glc-ox substrate (the reaction intermediate) by *Sp*GH85 is 3.7 s^{-1},[170] corresponding to a free energy barrier of 16.8 kcal/mol. This value is similar to the value for Glc-ox hydrolysis by *Tt*OGA, 14.7 kcal/mol, investigated in Chapter 4.[184] In the case of *Tt*OGA, the second reaction step (deglycosylation) turns out to be rate-limiting. In view of the theoretical values obtained here and the experimental reaction rates, we can expect that the second reaction step (hydrolysis of the reaction intermediate) is also rate-limiting in *Sp*GH85. The calculation of the second step in Paths A and B in *Sp*GH85 would further confirm this hypothesis, and these calculations are under way in our laboratory. Another explanation for the relatively low free energy barriers obtained for the first reaction step of *Sp*GH85 might be that other processes than the chemical reaction limit catalysis. These processes include the binding of the reactant substrate or the release of the reaction products. Further experiments and simulations considering larger substrates (e.g. the *+2* Asn residue and the entire linked protein, which were not present in our present model) are necessary to identify the rate-limiting step in *Sp*GH85.

From our present calculations, and together with the available structural and kinetic data, we cannot discard any of the two pathways. Despite path A exhibits a ~ 3 kcal/mol lower free energy barrier than path B, the corresponding Glc-ox^+/ neutral Asn335-rotamer 1 reaction intermediate would exhibit a charge separation configuration. In path A, a positively charged Glc-ox^+ would preferably be stabilized by a negatively charged residue through salt bridge, as occurs in family GH84 *Tt*OGA studied in Chapter 4 (Glc-ox^+ with Asp119),[184] instead of a neutral Asn. Indeed, the total charge of Asn335 side chain is 0.14 e more negative at $\mathbf{I_A}$ compared to the $\mathbf{MC_A}$, indicating that it requires polarization in order to stabilize the charged substrate. In particular, the $O\delta1_{Asn335}$ partial charge changes from -0.53 ± 0.04 e at $\mathbf{MC_A}$ to -0.61 ± 0.06 at $\mathbf{I_A}$ (Table S5.5). In path B, despite that the free energy barrier is 2.7 higher than path A, there is no charge separation. Even though several proton transfers take place, the charge of the Asn335 side chain practically does not change during the reaction (0.04 e more positive at $\mathbf{I_B}$ than $\mathbf{MC_B}$). The feasibility of these proton transfers depends on the pK_a shift of the active site residues along the reaction. In solution, the pK_a for *N*-deprotonation of an amide

is 18, while it drastically decreases to 7.5 when the amide oxygen is protonated.[201] Thus, the occurrence of the rare Asn tautomer might be caused by its strong interaction with the acid/base Glu337 and its own pK_a changes during catalysis, which stabilizes the protonation of the Asn335 oxygen. Another reason that might argue in favour of path B is that the reaction intermediate is slightly lower in energy than the one in path A (-16.2 kcal/mol vs. -14.7 kcal/mol, respectively). This would be consistent with the high tendency of *Sp*GH85 to catalyse transglycosylation. Indeed, it has been shown that the stabilization of reaction intermediates in glycosidase catalysis correlates with their ability to perform transglycosylation reactions.[202,203]

In conclusion, the available data indicates that both paths A and B are feasible and may coexist. However, the structural similarities of GHs that operate via a substrate-assisted mechanism and have an insertion between the catalytic residues (*Sm*ChiB, Chapter 3) suggest that the *Sp*GH85 reaction mechanism could resemble the one observed in *Sm*ChiB. This would include its associated proton shuttle machinery and thus a neutral Glc-ox intermediate and tautomerization of Asn335. If confirmed, this would be the first evidence in which an imidic acid Asn residue participates in glycosidase catalysis. Our results provide new mechanistic information of GH85 enzymes and will boost further investigations to fully comprehend their entire reaction mechanism.

5.3. Conclusions

- Both Glc-ox and Glc-ox$^+$ states of the oxazoline intermediate are stable in the *Sp*GH85 active site, but the assisting Asn335 exists only in its amide tautomeric form.

- Asn335 can tautomerize to its acid imidic form inside the active site of *Sp*GH85.

- Our simulations cannot exclude neither Path A or Path B and further investigations are required to elucidate the entire *Sp*GH85 reaction mechanism. Still, sequence and structural similarities with other GHs that follow the substrate-assisted mechanism suggest that Path B could be the most likely scenario for *Sp*GH85 catalysis and probably GH85 ENGases in general. A proton shuttle machinery, similar to the one observed in GH18 chitinases, is expected.

5.4. Computational details

Systems preparation

The initial structure for our calculations was taken from the X-ray structure of the *Sp*GH85 in complex with the inhibitor Glc-thiazoline at *-1* subsite (PDB entry 2W92)[62]. The active site residue Tyr373 exhibits two conformations in this structure, but one shows a steric clash with the methyl group of the inhibitor and thus only the alternative conformation was modelled. Glc-ox/ox$^+$ intermediates were obtained after substituting the S atom from Glc-thiazoline by O. The $(Man)_3$ trisaccharide of the *N*-glycan core (*-2* and *-3* subsites) was introduced in *Sp*GH85 by aligning the Glc-thiazoline moieties from aforementioned structure and the X-ray structure of $(Man)_3$-Glc-thiazoline bound to Endo-A[183] (PDB entry 3FHQ, Figure S5.1). The innermost GlcNAc that should be located at *+1* subsite was placed using molecular docking with Autodock Vina.[105] The selected binding pose for the next calculations fulfilled the following criteria: the O4 was at nucleophilic attack distance from the C1 of the Glc-ox/ox$^+$, the hydroxyl at position 4 of the GlcNAc was at hydrogen bond distance from the carboxylate group of the acid/base residue Glu337, and the *N*-acetamido group was facing the alternative side compared to the *-1* sugar, as generally observed in *N*-glycans. Figure 5.4 shows the resulting system considering both paths studied here (Figure 5.2).

All acidic residues were modelled deprotonated, except Glu337 in the case of the reaction intermediate considering Path A. Two different scenarios were considered: either the substrate featuring a neutral Glc-ox or a protonated Glc-ox$^+$. Each reaction path involved a different rotameric form of Asn335 (assisting residue) in order to maintain the hydrogen bond network that links the catalytic residues and the ligand, as expected from the mechanistic evidence and structural data. His protonation was assessed according their chemical environment. Specifically, His residues 214, 384, and 921 were considered neutral with their proton located at N_δ; while His residues 366, 487, 625, 684 and 704 were protonated at N_ε. His239 and His278 were double protonated. Nine sodium cations were added to neutralize the total charge of the system. Crystallographic water molecules were retained, and extra ones were added to form a 15 Å water box around the protein surface. This led to a total number of 108686 atoms and a simulation box of $114.6 \times 105.9 \times 104.4$ Å^3.

Classical MD

All classical MD simulations were performed employing the AMBER 11.[131] The ff99SB force field[132] was used to describe protein residues. Water molecules were modelled employing the TIP3P force field.[80] While GlcNAc and Man saccharides were described with GLYCAM06,[78] the reaction intermediate Glc-ox/ox$^+$ was parametrized using antechamber module,[133] considering GLYCAM06 and GAFF[134] parameters. RESP atomic charges were obtained from first principle calculations using Gaussian09[135] at HF/6-31G* level of theory.

MD calculations were run under periodic boundary conditions employing the following protocol. First, while the protein and substrate were fixed, all ions and water molecules were relaxed by 20,000 cycles of energy minimization. Afterwards, the entire system was relaxed with 20,000 additional minimization cycles. The whole system was gradually heated in the NVT ensemble to 100 K, 200 K and 300 K in intervals of 50 ps, except the last step (100 ps). The protein and ligand were restrained during the first heating step and thus only water and ions were able to diffuse. Subsequently, the density was converged up to water density at 300 K with 100 ps in the NPT ensemble. In all previous steps the time step was set up at 1 fs, but for the following production runs was incremented up to 2 fs via the SHAKE algorithm.[136] Production runs of 30 ns for each system were performed in the NPT ensemble until the RMSD of the protein backbone reached a plateau. The GlcNAc at the *+1* subsite was maintained imposing two soft restrains of 50 kcal/mol/rad^2 to the $O4_{GlcNAc}\cdots C1_{Glc\text{-}ox}$ and $H4O_{GlcNAc}\cdots O\varepsilon 2_{Glu337}$ distances at nucleophilic attack and hydrogen bond distances, respectively. Analysis of the obtained trajectories was carried out employing standard tools of AMBER and VMD.[137]

QM/MM MD

After classical MD equilibration, QM/MM MD calculations were performed using the method developed by Laio et al.[95] The QM region (Figure S5.8) included the side chain of both Asn335 and Glu337, part of the Man at *-2* subsite, Glc-ox/ox$^+$ at *-1* subsite and GlcNAc (subsite *+1*), leading to a total number of 88 atoms and cell of 17.61 × 18.27 × 21.56 Å^3. The Man residue was capped in three different C···C bonds. Asn335 and Glu337 side chains were capped by saturating the C_α of each protein residue with capping hydrogens. The NN, MIX and ESP radii were set at 6.35 Å, 13.23 Å and 16.94 Å, respectively. DFT level of theory with PBE functional[138] was used to describe the electronic structure. Kohn–Sham orbitals were expanded in a plane wave basis set with a kinetic energy cut-off of 70 Ry. Norm-

conserving *ab initio* pseudopotentials were employed, generated within the Troullier–Martins Scheme.[139] In all QM/MM calculations, a time step of 0.12 fs was used and a fictitious electron mass of 800 a.u.. The following protocol for the QM/MM MD simulations was followed. Firstly, the structure was optimized by MD with annealing of nuclear velocities, until the maximal component of the nuclear gradient was lower than 5×10^4 a.u. Subsequently, each system was equilibrated by running 6 ps of QM/MM MD at 300K employing the Nosé–Hoover thermostat[140,141]. A snapshot of these simulation was used as starting point for the QM/MM metadynamics calculations for each system.

QM/MM metadynamics simulations

All metadynamics calculations were run using the PLUMED 2 plugin.[103] All reactions were activated using atomic distances — or distance difference — as CVs. In the case of the preparation of the Michaelis complex of *Sp*GH85 considering both Path A and B, the $C1_{Glc\text{-}ox/ox+}\cdots O4_{GlcNAc}$ distance was used as CV in order to form the glycosidic bond. The height and width of the Gaussian terms was set at 1.0 kcal/mol and 0.15 Å, respectively. The deposition time was set at 300 MD steps. The obtained Michaelis complexes were further equilibrated with 15 ps of QM/MM MD. From these equilibrated structures, the first reaction step was evaluated employing two distance differences as CVs. CV_1 was defined as the distance difference between $C1\cdots O_x$ and $C1\cdots O_g$ distances and accounts for the intramolecular nucleophilic attack of O_x and the disruption of the glycosidic bond. CV_2 was defined as the distance difference between $H_{\varepsilon}2_{Glu337}\cdots O_g$ and $O_{\varepsilon}2_{Glu337}\cdots H_{\varepsilon}2_{Glu337}$ distances and accounts for the proton transfer between Glu337 and O_g. The width of the Gaussian terms was set at 0.15 and 0.10 Å for CV_1 and CV_2, respectively. The height of the Gaussian terms was set at 1.0 kcal/mol but was reduced to 0.5 kcal/mol before crossing the TS. The deposition time was set at 300 MD steps. Both simulations were stopped once the reaction products were explored and a recrossing event occurred (i.e., returning to the Michaelis complex state). All analyses such as distances, charges, puckering conformations were performed on the relevant states of the enzymatic reactions, selected via capturing frames on the region from each FES.

5.5. Supplementary material

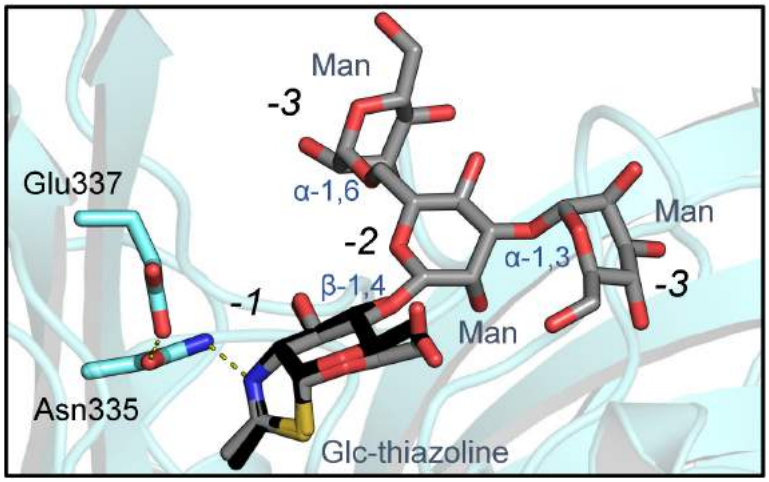

Figure S5.1. Construction of the $(Man)_3$-Glc-ox/ox^+ moiety in *Sp*GH85. Structural alignment between the X-ray structure of Glc-thiazoline (carbon atoms coloured in black) bound to *Sp*GH85 (cyan, PDB entry 2W92) and the substrate $(Man)_3$-Glc-thiazoline (grey carbon atoms) trapped in Endo-A (PDB entry 3FHQ). In particular, the alignment was conducted on the Glc-thiazoline residue since this part is common in both mono- and tetrasaccharide inhibitors.

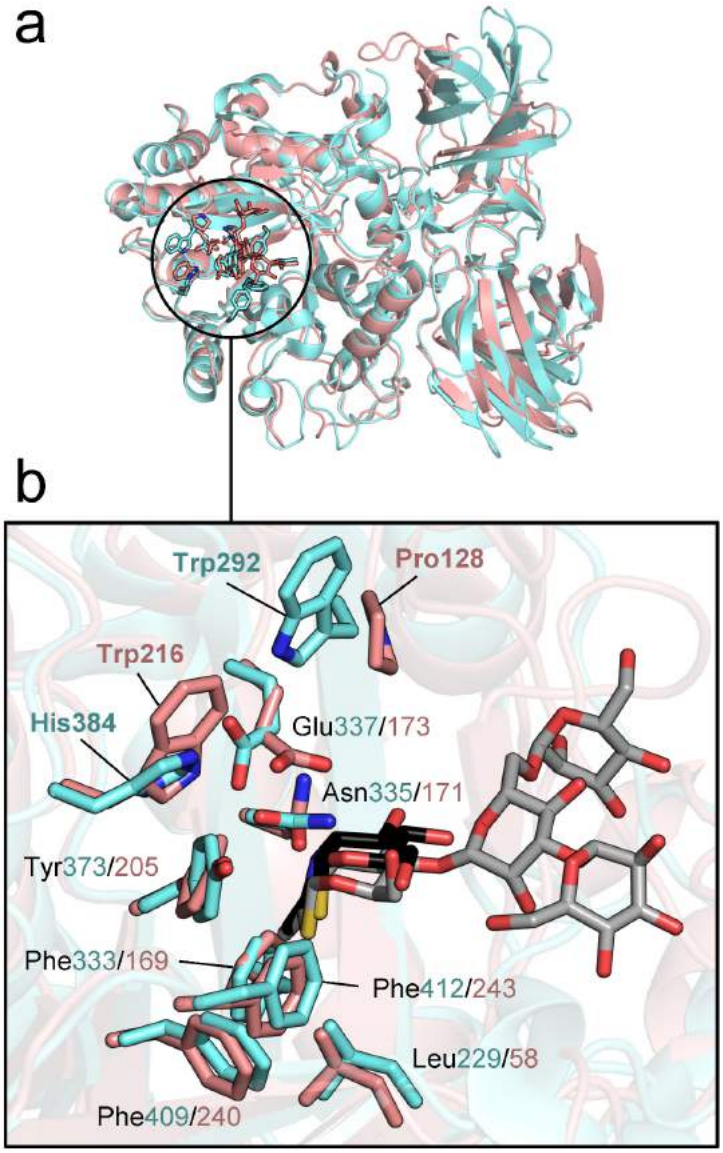

Figure S5.2. Active site comparison between Endo-D and Endo-A (cyan and salmon, respectively). (a) Structural alignment between the X-ray structures of *Sp*GH85 (Endo-D) and Endo-A (PDB entries 2W92 and 3FHQ, respectively). (b) Phe, Leu and Tyr residues at the *-1* subsite (shown as sticks) provide an aromatic cavity to accommodate the methyl group of the Glc-ox/ox^{+} residue. Noteworthy, the amino acids at *+1* subsite (sticks with bold labels) differ between both enzymes.

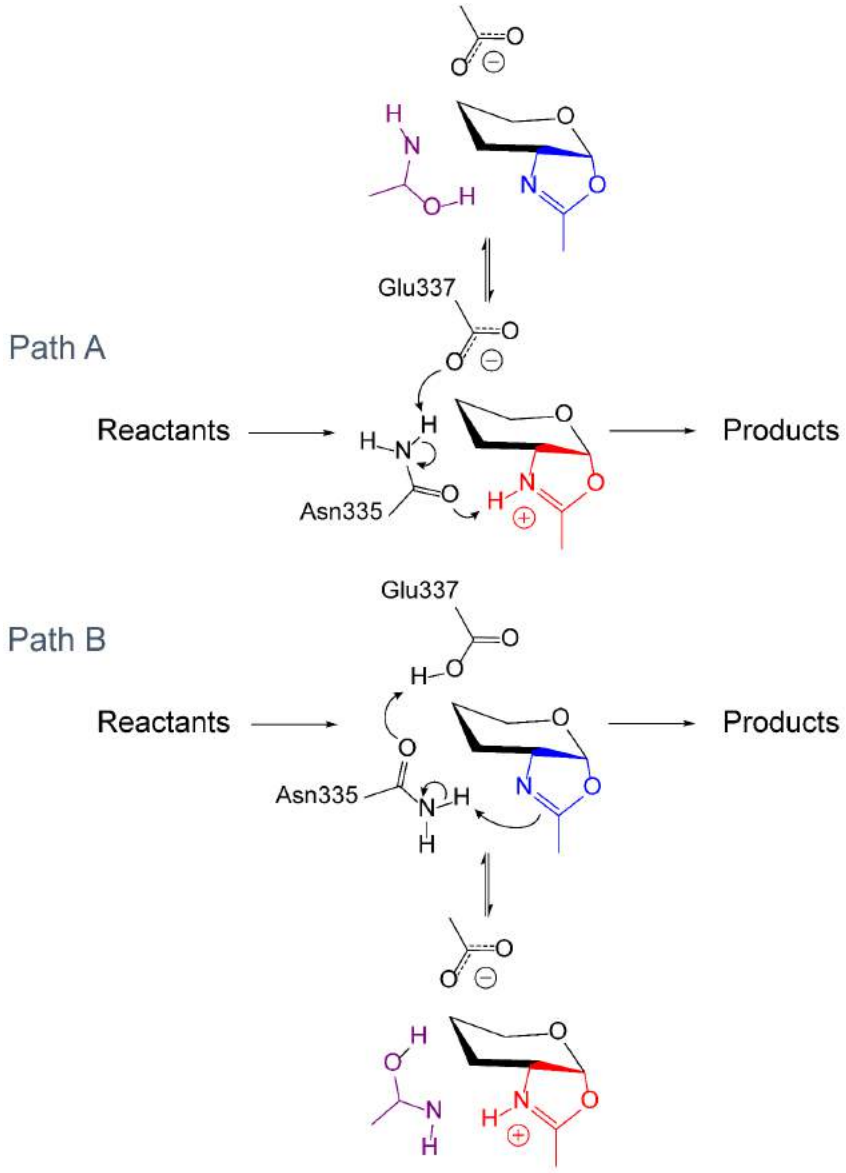

Figure S5.3. Additional scheme of the two reaction paths from *Sp*GH85 studied in this work but including the occurrence of an acid imidic tautomeric form of Asn335 (purple) in the intermediate of the reaction. These structures have been submitted to geometry optimization (Figure S5.4) and are found unstable. Glc-ox and Glc-ox^{+} are shown in red and blue colours, respectively.

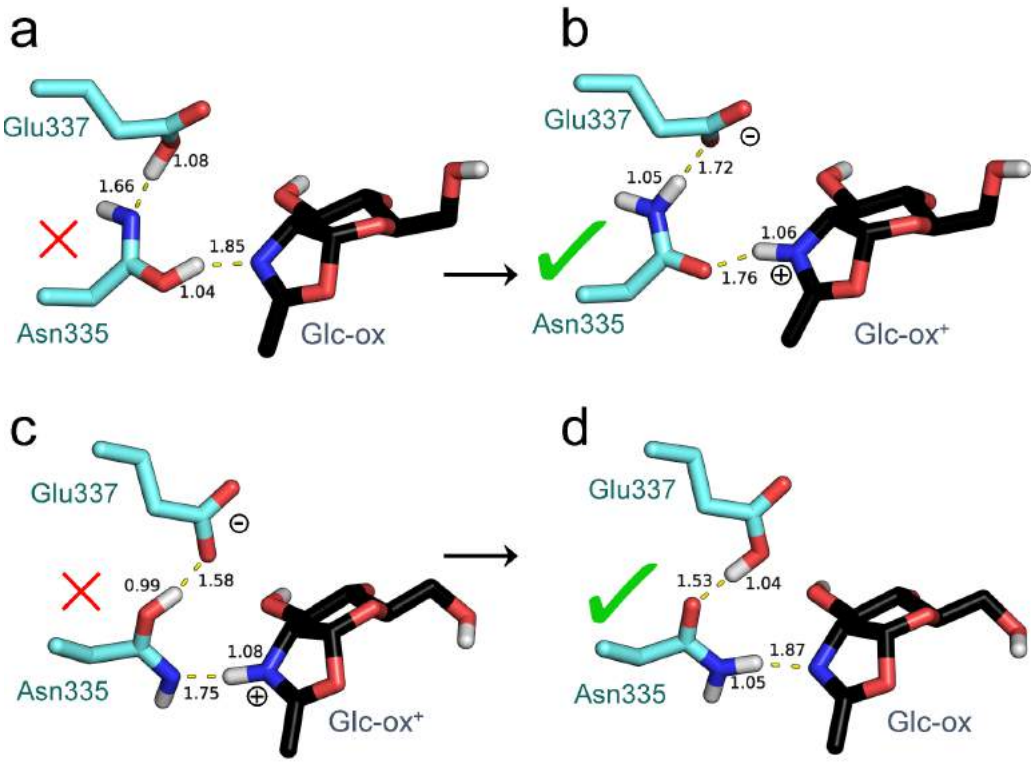

Figure S5.4. QM/MM geometry optimizations considering Asn335 as an imidic acid in the reaction intermediate of both analysed paths. (a) Initial imidic acid tautomer in path A and (b) its corresponding optimized structure where the amide form of Asn335 is recovered. (c) Initial imidic acid tautomeric form of Asn335 in path B and (d) its corresponding optimized structure. A red cross and a green stick represent an unstable and stable state, respectively. The recovery of the amide form of Asn335 in the reaction intermediate involves two proton transfers: one between Glu337 and Asn335 and another between the oxazoline ring and Asn335.

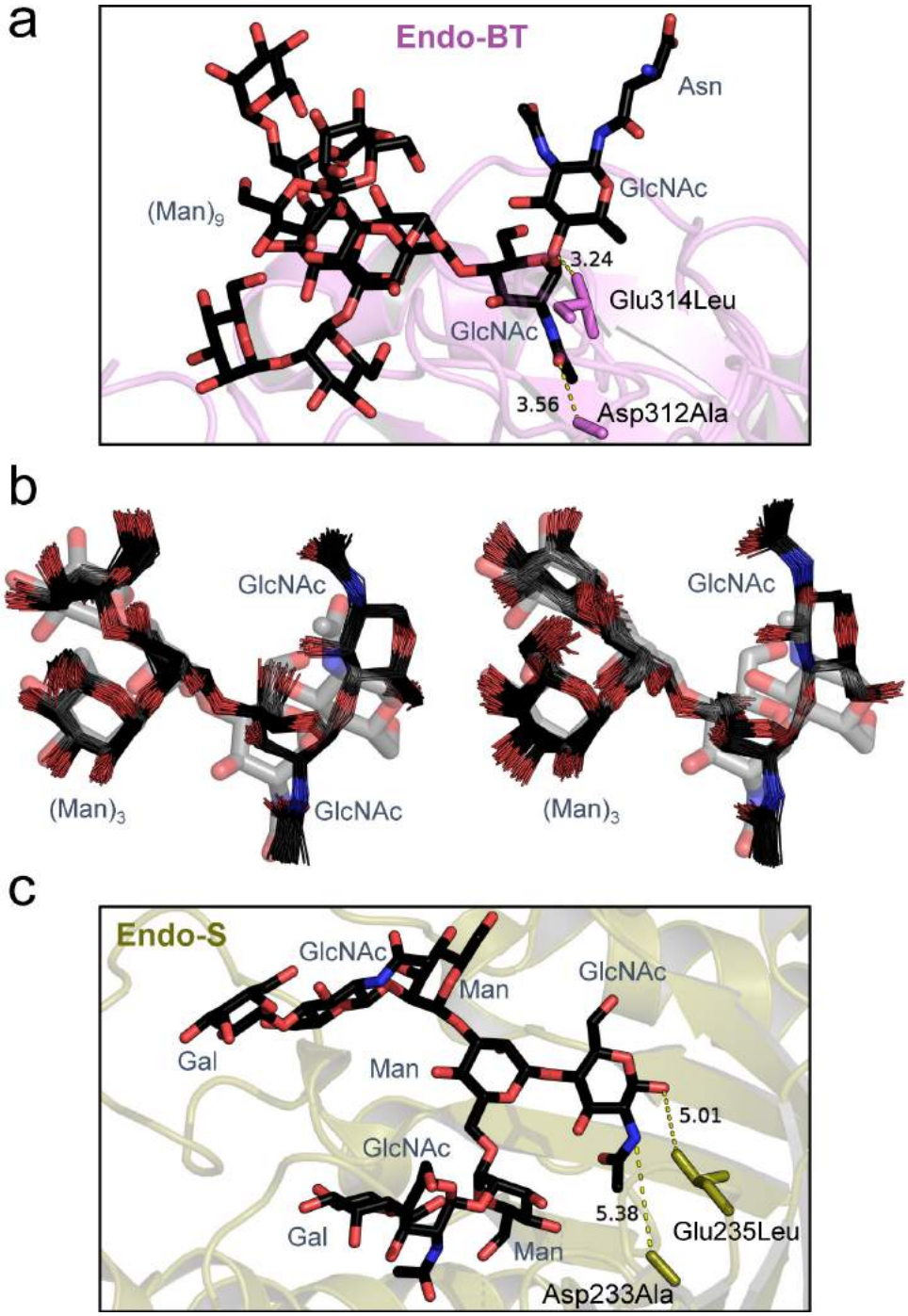

Figure S5.6. Comparison of our *Sp*GH85 results with structural studies of GH18 ENGases. (a) X-ray structure of Endo-BT double mutant in complex with (Man)9(GlcNAc)Asn (Michaelis complex). (b) The *N*- glycan core from Endo-BT structure (a, grey sticks) compared with the corresponding ensemble obtained by QM/MM MD in *Sp*GH85 (thin black sticks). (c) X-ray structure of Endo-S double mutant with a complex glycan where the innermost GlcNAc and the Asn are missing.

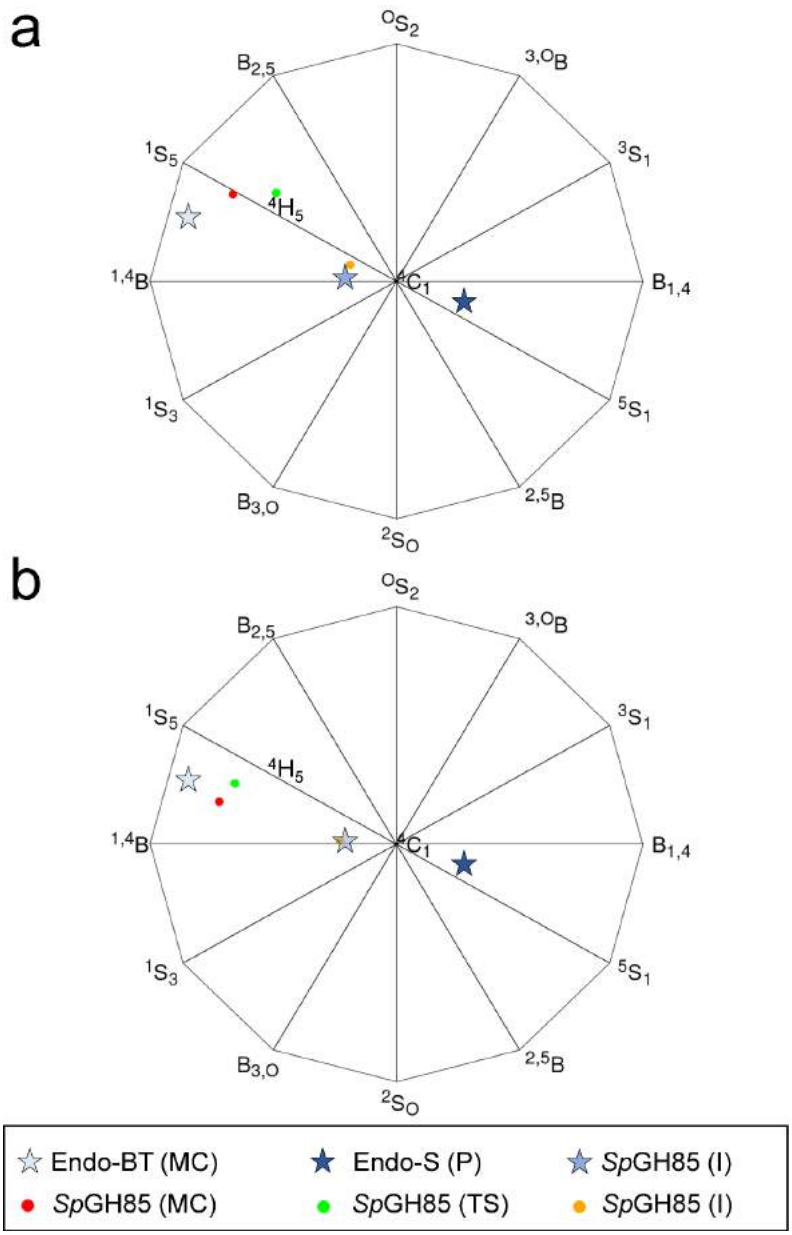

Figure S5.7. Puckering coordinates of the GlcNAc at the *-1* subsite for the relevant states computed along (a) Path A and (b) Path B in *Sp*GH85 (dots); and for X-ray structures of additional ENGase systems (blue stars). PDB entries 6TCV, 6EN3 and 2W92 for Endo-BT, Endo-S and *Sp*GH85, respectively.

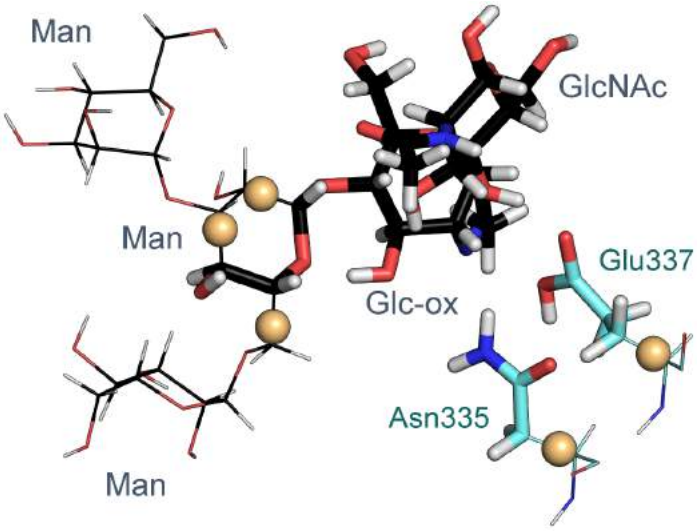

Figure S5.8. Active site atoms described at QM level in QM/MM calculations of *Sp*GH85 (thicker sticks). Capping hydrogens are shown as light orange spheres.

Table S5.1. Relevant distances (mean and standard deviation) of each representative structure along Path A. See atom labels in Figure 5.7a.

Distance (Å)	**MC**	**TS**	**I**
$C1\cdots O_x$	3.06 ± 0.09	2.73 ± 0.01	1.60 ± 0.06
$C1\cdots O_g$	1.53 ± 0.05	2.23 ± 0.11	3.76 ± 0.07
C1···O5	1.39 ± 0.04	1.30 ± 0.05	1.36 ± 0.03
$H_{\varepsilon}2_{Glu337}\cdots O_g$	1.61 ± 0.11	1.20 ± 0.05	1.06 ± 0.03
$O_{\varepsilon}2_{Glu337}\cdots H_{\varepsilon}2_{Glu337}$	1.04 ± 0.03	1.26 ± 0.05	1.56 ± 0.04
$O_{\varepsilon}1_{Glu337}\cdots H_{\delta}2_{Asn335}$	2.11 ± 0.16	1.81 ± 0.13	1.83 ± 0.21
$O_{\delta}1_{Asn335}\cdots H_x$	1.95 ± 0.16	2.00 ± 0.19	1.71 ± 0.13

Table S5.2. Relevant distances (mean and standard deviation) of each representative structure along Path B. See atom labels in Figure 5.7b.

Distance (Å)	**I_P**	**TS_P**	**P_P**
$C1\cdots O_x$	3.08 ± 0.08	2.67 ± 0.17	1.47 ± 0.04
$C1\cdots O_g$	1.55 ± 0.03	1.99 ± 0.18	3.74 ± 0.05
C1···O5	1.38 ± 0.03	1.34 ± 0.04	1.39 ± 0.03
$H_{\varepsilon}2_{Glu337}\cdots O_g$	1.81 ± 0.05	1.15 ± 0.02	1.00 ± 0.03
$O_{\varepsilon}2_{Glu337}\cdots H_{\varepsilon}2_{Glu337}$	1.02 ± 0.02	1.31 ± 0.04	1.73 ± 0.05
$O_{\varepsilon}1_{Glu337}\cdots H_{\delta}1_{Asn335}$	1.82 ± 0.17	1.54 ± 010	1.05 ± 0.04
$H_{\delta}1_{Asn335}\cdots O_{\delta}1_{Asn335}$	1.02 ± 0.04	1.03 ± 0.03	1.53 ± 0.11
$N_{\delta}2_{Asn335}\cdots H_{\delta}2_{Asn335}$	2.25 ± 0.13	2.03 ± 0.15	1.03 ± 0.02
$H_{\delta}2_{Asn335}\cdots H_x$	1.04 ± 0.03	1.04 ± 0.02	2.03 ± 0.20

Table S5.3. Dihedral angles defining the conformation of the pyranose ring of GlcNAc at *-1* subsite for each relevant state along Path A.

Dihedral angle	MC	TS	I
C2-C1-O5-C5	2.7 ± 9.1	-18.8 ± 12.3	-41.1 ± 6.6
C3-C2-C1-O5	-27.4 ± 5.9	-5.9 ± 11.6	33.2 ± 5.2

Table S5.4. Dihedral angles defining the conformation of the pyranose ring of GlcNAc at *-1* subsite for each relevant state along Path B.

Dihedral angle	MC	TS	I
C2-C1-O5-C5	19.8 ± 12.9	11.2 ± 3.8	-36.3 ± 9.5
C3-C2-C1-O5	-30.4 ± 6.9	-27.5 ± 2.9	31.8 ± 7.5

Table S5.5. ESP charges (e) of each representative structure along Path A. See Figure 5.7a for atom labeling. δ^{an} corresponds to the sum of the charges of C1 and its neighboring linked atoms O5, C2, H1 and/or O_x/O_g).

Atom	MC	TS	I
C1	0.36 ± 0.19	0.15 ± 0.10	0.53 ± 0.10
C2	0.46 ± 0.13	0.46 ± 0.02	-0.06 ± 0.09
C_x	0.65 ± 0.06	0.51 ± 0.06	0.29 ± 0.06
O_x	-0.60 ± 0.04	-0.52 ± 0.01	-0.30 ± 0.04
N_x	-0.68 ± 0.06	-0.55 ± 0.06	-0.21 ± 0.07
H_x	0.30 ± 0.02	0.27 ± 0.04	0.30 ± 0.02
O5	-0.53 ± 0.13	-0.12 ± 0.12	-0.59 ± 0.06
H1	0.04 ± 0.03	0.12 ± 0.02	0.06 ± 0.03
O_g	-0.62 ± 0.10	-0.72 ± 0.05	-0.77 ± 0.07
$O\delta1_{Asn335}$	-0.53 ± 0.04	-0.51 ± 0.03	-0.61 ± 0.06
$N\delta2_{Asn335}$	-0.52 ± 0.07	-0.55 ± 0.05	-0.55 ± 0.07
$H\delta2_{Asn335}$	0.31 ± 0.04	0.34 ± 0.01	0.34 ± 0.04
$O\varepsilon1_{Glu337}$	-0.68 ± 0.03	-0.76 ± 0.02	-0.77 ± 0.07
$O\varepsilon2_{Glu337}$	-0.73 ± 0.07	-0.78 ± 0.04	-0.84 ± 0.05
$H\varepsilon2_{Glu337}$	0.30 ± 0.02	0.27 ± 0.04	0.30 ± 0.02
δ^{an}	-0.28 ± 0.12	0.61 ± 0.07	-0.37 ± 0.06

Table S5.6. ESP charges (e) of each representative structure along Path B. See Figure 5.7b for atom labeling. δ^{an} corresponds to the sum of the charges of C1 and its neighboring linked atoms O5, C2, H1 and/or O_x/O_g).

Atom	MC	TS	I
C1	0.25 ± 0.09	0.20 ± 0.15	0.59 ± 0.13
C2	0.35 ± 0.08	0.45 ± 0.10	0.09 ± 0.14
C_x	0.48 ± 0.06	0.47 ± 0.03	0.30 ± 0.06

O_x	-0.55 ± 0.04	-0.58 ± 0.01	-0.35 ± 0.05
N_x	-0.48 ± 0.09	-0.48 ± 0.04	-0.51 ± 0.07
H_x	0.24 ± 0.02	0.26 ± 0.03	0.37 ± 0.04
O5	-0.47 ± 0.08	-0.30 ± 0.13	-0.71 ± 0.08
H1	0.06 ± 0.03	0.12 ± 0.02	0.02 ± 0.03
O_g	-0.49 ± 0.07	-0.66 ± 0.11	-0.66 ± 0.08
$O_{\delta}1_{Asn335}$	-0.59 ± 0.05	-0.71 ± 0.07	-0.62 ± 0.05
$H_{\delta}1_{Asn335}$	0.43 ± 0.03	0.54 ± 0.08	0.37 ± 0.05
$N_{\delta}2_{Asn335}$	-0.72 ± 0.05	-0.72 ± 0.03	-0.77 ± 0.08
$H_{\delta}2_{Asn335}$	0.23 ± 0.02	0.24 ± 0.02	0.31 ± 0.04
$O_{\varepsilon}1_{Glu337}$	-0.56 ± 0.06	-0.81 ± 0.04	-0.54 ± 0.06
$O_{\varepsilon}2_{Glu337}$	-0.57 ± 0.05	-0.67 ± 0.03	-0.56 ± 0.05
$H_{\varepsilon}2_{Glu337}$	0.39 ± 0.04	0.51 ± 0.06	0.40 ± 0.05
δ^{an}	-0.30 ± 0.07	0.47 ± 0.10	-0.36 ± 0.09

6. Conformational analysis of pyranose rings in substrate-assisted catalysis.

Abstract: The puckering coordinates developed by Cremer and Pople allow to describe unequivocally and quantitatively the conformation for a six-membered ring. Ring conformation is a crucial aspect in the reaction mechanism of GHs, being key when designing GH inhibitors. It has been proven that, by analyzing the intrinsic conformations for a given sugar ring, one is able to predict and rationalize its conformational behavior in the context of GH catalysis. Following this approach, we here inspected the species involved in substrate-assisted catalysis by means of metadynamics simulations, using the puckering coordinates as collective variables. Furthermore, we investigated the conformational landscape of sugar-based inhibitors that are specific for GH families that follow a substrate-assisted reaction mechanism.

6.1. Introduction

The computational analysis of carbohydrate conformations has served to deeply comprehend catalysis by GHs.[35,88] As introduced in section 1.2.3, the evaluation of isolated monosaccharides by means of *ab initio* MD calculations allows to understand their intrinsic features such as the relative stability of the different ring conformations and their respective electronic and structural properties. The very first example was β-D-glucose computed in 2007 by Biarnés et al (Figure 6.1, left).[38] The authors found that despite the most stable conformation for an isolated Glc is the $^4C_1$ chair, as expected for a Glc in solution, its conformational FES exhibits several local minima, the most stable ones along $^1S_5$-$^{1,4}B$-$^1S_3$-$B_{3,O}$-^2S_O line on the Stoddart diagram. Interestingly, the conformation of the sugar at the *-1* subsite β-D-glucosidase MCs obtained experimentally fall into those local minima.

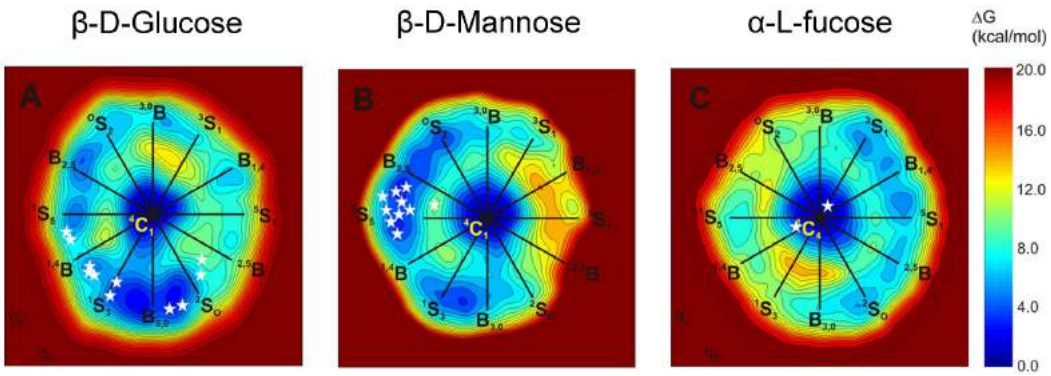

Figure 6.1. Conformational FES projected on the Stoddart diagram for β-D-glucose, β-D-Mannose and α-L-fucose. Contour lines are at 0.5 kcal/mol. Star symbols indicate the observed conformations for Michaelis complexes of β-D-glucosidases, β-D-mannosidases and α-L-fucosidases. Figure adapted from reference.[88]

It was latter demonstrated that GHs initiate the enzymatic reaction from a distorted sugar conformation in the MC, as evidenced for a 1,3-1,4-β-glucanase (catalytic itinerary: $^{1,4}B/^1S_3 \rightarrow {}^4E/^4H_3 \rightarrow {}^4C_1$).[204] This was a clear indication that GHs take advantage of the intrinsic electronic and structural properties of each carbohydrate via accommodating distorted conformations to efficiently perform hydrolysis. Some of these intrinsic characteristics are the elongation/shortening of the C1···O_g/C1···O5 bonds, C1···O_g orientation, and charge development at the anomeric carbon. These relevant sugar properties were collected and integrated into a pre-activation index when Man was under investigation (Figure 6.1, middle).[39] The pre-activation index allows to identify the best conformation for GH catalysis.

This approach has been successfully applied to other isolated sugars such as β-D-xylose[40] or α-L-fucose[205] (Figure 6.1, right). Besides providing an in-depth understanding of the GH catalysed hydrolysis of their natural substrates, the conformational analysis of carbohydrate-based inhibitors allowed to comprehend GH inhibition and their specificities as well. Examples of such investigations include the inhibitors glucoimidazol,[206] mannoimidazole,[207] isofagomine[207,208] kifunensine[209] or carba-cyclophellitol.[210] These inhibitors bind tightly the active site of GHs because they reassemble the substrate at the TS in terms of ring conformation and/or sp^2 character at the anomeric carbon.

In order to obtain the conformational free energy landscape of an isolated sugar, the metadynamics scheme is applied together with CPMD, using the Cremer and Pople puckering coordinates as CVs. These CVs allow to accurately describe the conformation of a six-membered ring and have been subsequently employed by others.[211-213] The Cremer-Pople polar coordinates consist in a radius Q and two-phase angles θ and ϕ (Figure 6.2). The Q coordinate is the sum of the perpendicular distance of each ring atom (j) to the ring average plane ($Q = \sum_j^6 z_j$). θ and ϕ coordinates are obtained by solving the following system of equations.

$$Q \sin\theta \cos\phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \cos[\frac{2\pi}{6} 2(j-1)] \qquad \text{Equation 25}$$

$$Q \sin\theta \sin\phi = -\sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \sin[\frac{2\pi}{6} 2(j-1)] \qquad \text{Equation 26}$$

$$Q \cos\theta = \sqrt{\frac{1}{6}} \sum_{j=1}^{6} (-1)^{j-1} z_j \qquad \text{Equation 27}$$

All 38 canonical conformations for a six-member ring are located on the external shell of the puckering volume (Figure 1.5b), which is not spherical but ellipsoidal, with a slow increase of the puckering amplitude Q from the poles to the equator.[40] The Stoddart diagram is represented by the cartesian coordinates (q_x and q_y) and can be calculated in terms of the polar coordinates as

$$q_x = Q \cdot \sin\theta \cdot \cos\phi \qquad \text{Equation 28}$$
$$q_y = Q \cdot \sin\theta \cdot \sin\phi \qquad \text{Equation 29}$$
$$q_z = Q \cdot \cos\theta \qquad \text{Equation 30}$$

This two-dimensional projection of q_z (i.e. The Stoddart diagram, Figure 1.5b) facilitates the visualization of the computed free energy surface in a continuous space and, therefore, the interpretation of the GH catalytic itineraries.

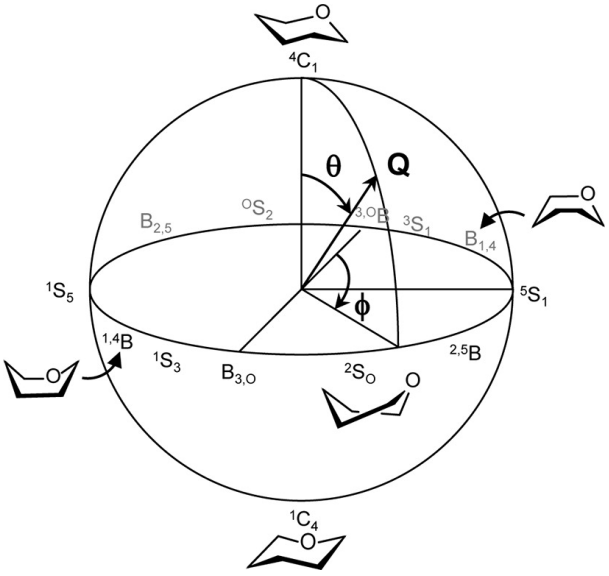

Figure 6.2. Cremer-Pople puckering coordinates. Figure adapted from reference.[35]

In the context of substrate-assisted catalysis, we investigated both the isolated species involved in the reaction mechanism and their derived inhibitors (Figure 6.3). While the conformational space of GlcNAc has been already analysed by others,[213,214] the conformational energy landscape of the reaction intermediate Glc-ox/ox$^+$ and the GlcNAc oxocarbenium ion transition state are yet unexplored. Furthermore, a remarkable difference of the substrate-assisted catalysis in comparison with the classical retaining Koshland mechanism of hydrolysis is that the substrate does not covalently bound to the enzyme in the reaction intermediate. This complicates the understanding of substrate-assisted GH inhibition, since it is not straightforward whether a certain inhibitor reassembles the substrate bound in the MC, the TS or reaction intermediate.[215] Traditionally, when researchers design TS-like inhibitors for GHs they try to incorporate the following features: sp^2-hybridized centre installed at the C1 to mimic the TS geometry, a nitrogen atom at the C1/O5 position to mimic the partial positive charge developed at C1 and, if possible, conformational constrains that force the pyranose ring to adopt a TS-like conformation. Additionally, slight mechanistic differences between the GH families that use substrate-assisted catalysis should be considered, as it has been shown in this Thesis (families GH18/GH56, GH84 and GH85 in chapters 3, 4 and 5, respectively).

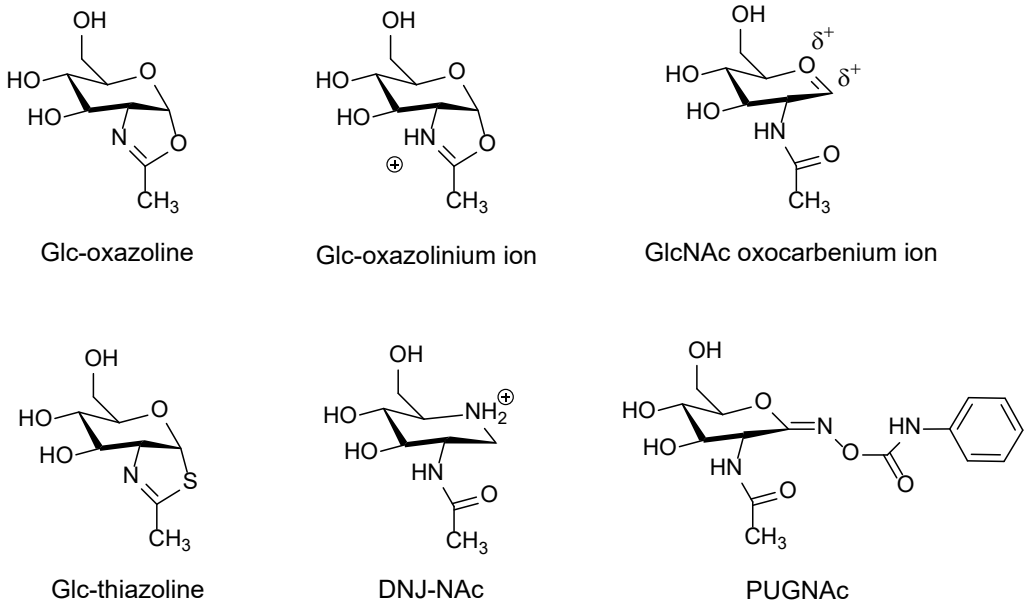

Figure 6.3. Carbohydrate compounds related with the substrate-assisted mechanism evaluated in this work. Top: species corresponding to the reaction intermediate and the transition state (Glc-ox/ox$^+$ and GlcNAc oxocarbenium ion, respectively). Bottom: inhibitors of the families that follow the substrate assisted-mechanism, from left to right: 1,2-dideoxy-2′-methyl-α-$_D$-glucopyranoso-[2,1-*d*]-Δ2′-thiazoline (Glc-thiazoline),[48] 2-acetamido-1,2-dideoxynojirimycin (DNJ-NAc)[216] and *O*-(2-acetamido-2-deoxy-$_D$-glucopyranosylidene)amino-*N*-phenyl carbamate (PUGNAc).[217]

Here we analysed the conformation and properties of the isolated carbohydrates that correspond to the reaction intermediate and the TS involved in the enzymatic reactions catalysed by GHs that follow the substrate assisted-mechanism (Glc-ox/ox$^+$ and GlcNac oxocarbenium ion, respectively, Figure 6.3 top). Moreover, we selected three carbohydrate-based compounds that inhibit these enzymes and display different degrees of selectivity. The conformational map of an isolated compound and related inhibitors allows us to rationalize their similarities, thus predict their inhibition ability, in the context of the substrate-assisted mechanism.

6.2. Results and discussion

6.2.1. Conformational analysis of the species involved in the enzymatic reaction

Ab initio MD simulations of isolated Glc-ox and Glc-ox$^+$ compounds were performed with CPMD method at DFT level of theory. From the equilibrated molecule at 300 K, the metadynamics approach was used to explore the conformational space of the pyranose ring by using the Cremer and Pople puckering coordinates as CVs (see computational details).

The obtained FESs for Glc-ox and its protonated form Glc-ox$^+$ are shown in Figures 6.4 and 6.5, respectively. They show clear similarities such as the position and number of free energy minima (two in the equator of the Cremer and Pople sphere and the global minimum around the half-chair /envelope region).

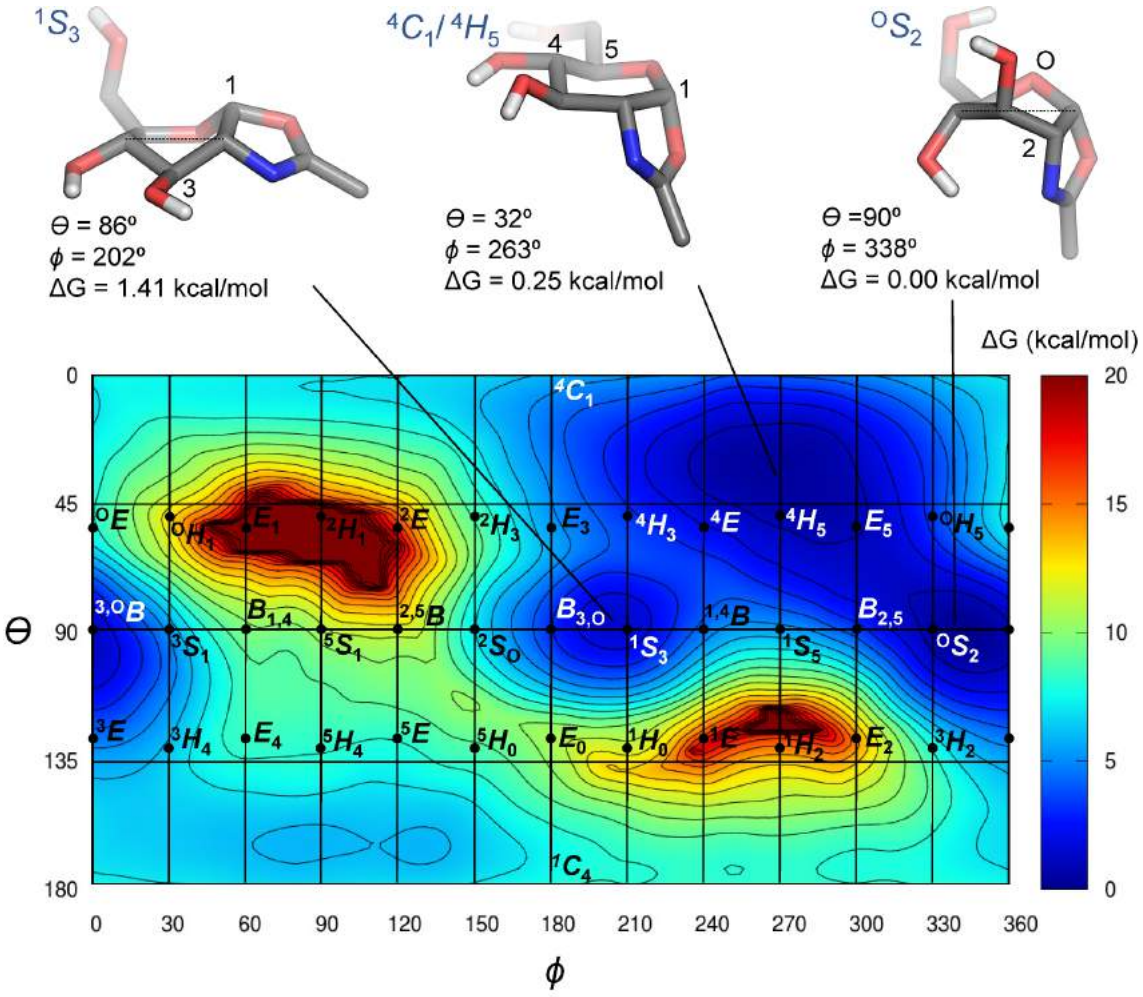

Figure 6.4. Metadynamics calculation of the conformational landscape of Glc-ox. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

The conformational FES of Glc-ox has a global minimum cantered at the $^{O}S_2$ conformation, and two local minima that correspond to $^{4}C_1/^{4}H_5$ and $^{1}S_3$ conformations, 0.25 and 1.41 kcal/mol higher in energy than the global minimum, respectively. The $^{4}C_1/^{4}H_5$ conformation is also the one that was observed in the *Sm*ChiB active site (chapter 3), showing that the enzyme stabilizes this specific conformation and constricts the rest of the conformational space (the global minimum of the isolated Glc-ox is not present in the conformational FES inside the enzyme, Figure 3.5). In the case of Glc-ox^{+} (Figure 6.5), the global minimum of the conformational FES lies at $^{3,O}B$ conformation, in close proximity to the global minimum obtained for the unprotonated molecule ($^{O}S_2$ conformation). Similarly to Glc-ox, two secondary minima are located at $^{4}C_1/^{4}H_5$ and $B_{3,O}/^{1}S_3$ regions, being 0.43 and 3.20 kcal/mol higher in energy than the global minimum.

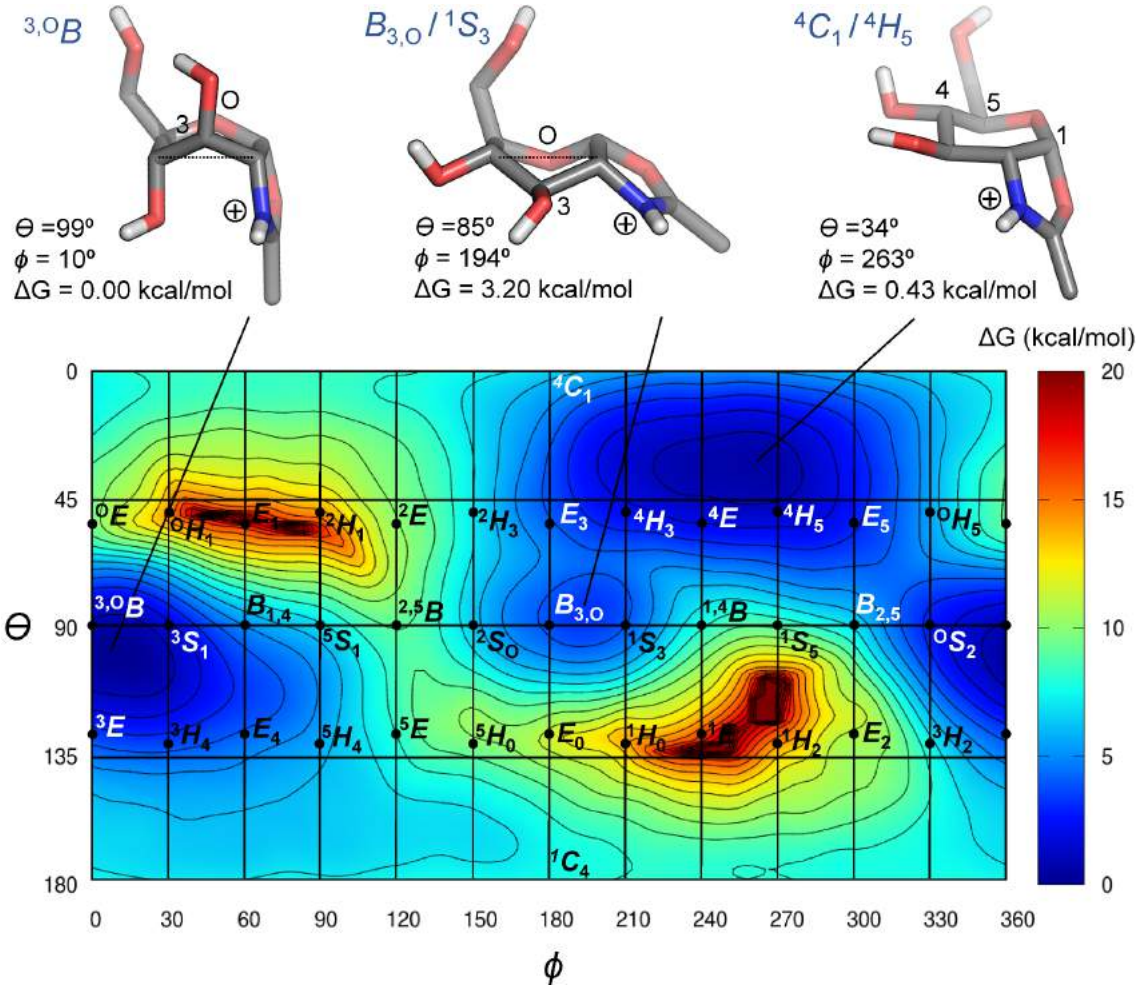

Figure 6.5. Metadynamics calculation of the conformational landscape of Glc-ox^{+}. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

The above results indicate that, despite Glc-ox and Glc-ox$^+$ do not significantly differ in terms of pyranose conformation, they exhibit various structural differences important for catalysis. For instance, when comparing $^4C_1/^4H_5$ conformers of Glc-ox and Glc-ox$^+$ (conformation observed inside GHs), the C1···O$_x$ distance, which corresponds to the bond that is going to break in the second reaction step, is greater in the protonated form (1.66 ± 0.09 Å and 1.50 ± 0.04 Å, for Glc-ox$^+$ and Glc-ox, respectively). Additionally, the C1···O5 distance is shorter in Glc-ox$^+$ than Glc-ox: 1.33 ± 0.03 Å and 1.39 ± 0.03 Å, respectively. This suggests that Glc-ox$^+$ is a species closer to the TS of the reaction in comparison with its unprotonated form, Glc-ox, thereby the latter being more stable. However, we described in this Thesis that depending on the GH family, the enzyme stabilizes either Glc-ox or its protonated form Glc-ox$^+$, showing that ultimately the protein environment dictates the nature of the intermediate. The reason why GHs that operate via substrate-assisted catalysis have found different solutions to degrade GlcNAc substrates is still a mystery.

To further understand the features of the TS in substrate-assisted catalysis, we investigated the conformation of the isolated GlcNAc oxocarbenium ion. Considering that the acetamido group would definitely collapse on the positively charged sp^2 anomeric carbon, the C1···O$_x$ distance was restrained at 2.35 Å along the entire simulation. The value for this restrain was chosen considering the C1···O$_x$ distance observed at the TS in the enzymatic reactions studied in this work. The obtained conformational FES displays two minima (Figure 6.6).

The global minimum corresponds to the ^4E, one of the canonical TS-like conformations. A secondary minimum located at the $^3S_1$ / $^3H_4$ region is 4.11 kcal/mol higher than the global minimum. This conformation is stabilized by a hydrogen bond between the hydroxyl at position 3 and the lone pair of the oxygen of the CH_2OH group. In the enzymatic reaction catalyzed by GHs studied in this work (chapters 3-5), the conformation is a $^4H_5$, indicating that the enzyme environment slightly shifts the conformational preferences of this species (^4E and $^4H_5$ are adjacent conformations in the Cremer-Pople space). GlcNAc oxocarbenium ion has an anomeric center with a strong sp^2 character: a positive charge, a short C1···O5 distance of 1.28 ± 0.02 Å, almost coplanarity of the pyranose ring (C3-C2-C1-O5 and C2-C1-O5-C5 dihedral angles of 18.9 ± 3.6º and -17.4 ± 4.0º, respectively). Coplanarity is not completely reached probably because of the presence of the *N*-acetyl group that might affect the position of the positively charged anomeric carbon.

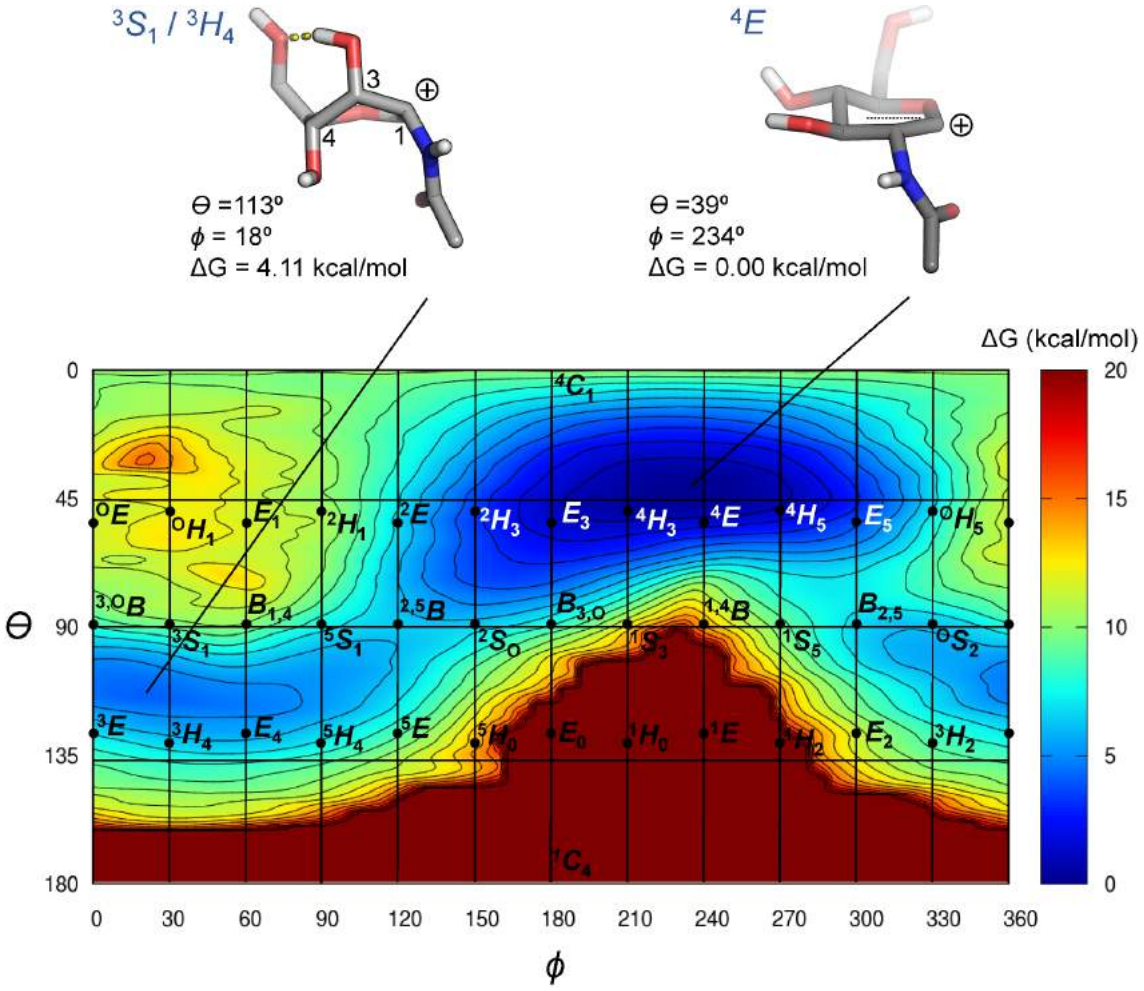

Figure 6.6. Metadynamics calculation of the conformational landscape of GlcNAc oxocarbenium ion. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

6.2.2. Conformational analysis of inhibitor compounds

We investigated the conformational features of three carbohydrate-based compounds that inhibit GHs that operate via substrate-assisted catalysis: Glc-thiazoline, DNJ-NAc and PUGNAc (Figure 6.3, bottom). The main objective is to compare these results with the conformational preferences obtained for Glc-ox/ox$^+$ and GlcNAc oxocarbenium ion as representation of the reaction intermediate and TS, respectively.

First, we analysed Glc-thiazoline (Figure 6.3, bottom left, a potent inhibitor of enzymes that perform substrate-assisted catalysis, achieving nM inhibition in GHs from families GH18, GH20, GH84[48,157,218] and μM in GH85 enzymes.[62] It has been extensively used to identify and prove substrate-assisted catalysis in several GHs. The main difference of this molecule in comparison with the reaction intermediate (Glc-ox/ox$^+$) is the presence of a sulfur atom at the position of the oxygen of the oxazolinic ring. The conformational FES is shown in Figure 6.7, exhibiting the both the same location and energetic tendencies as Glc-ox (Figure 6.4).

The global minimum corresponds also to $^{O}S_2$ conformation, while the secondary minima are $^4C_1/^4H_5$ and $^1S_3$ conformations, 0.81 and 1.71 kcal/mol higher in energy than the global minimum, respectively (almost equienergetic). Despite that the $^4C_1/^4H_5$ conformation is located at the second most stable minimum in the FES of the isolated compound, Glc-thiazoline in GHs X-ray structures shapes this conformation (coloured symbols in Figure 6.7). Considering that Glc-ox and Glc-ox$^+$ feature a $^4C_1/^4H_5$ conformation as well in our QM/MM simulations, we can argue that Glc-thiazoline mimics the conformation of the reaction intermediate in GHs that follow the substrate-assisted mechanism. Furthermore, the GH-substrate interactions both in the X-ray structures (Glc-thiazoline) and the computed Glc-ox/ox$^+$ are conserved. However, it has been experimentally proven by linear free energy relationships of log(K_M/k_{cat}) versus log(K_I) that Glc-thiazoline is a TS-like inhibitor for a GH84 *O*-GlcNAcase, not a ground state analogue.[215] This suggests that not only the conformation of a certain carbohydrate-based inhibitor needs to be under consideration, but also other features. For instance, the C1$\cdots$S$_x$ distance of the $^4C_1/^4H_5$ conformer ranges a values of 1.89 ± 0.06 Å, reassembling the dissociative nature of the TS in GHs studied here (i.e. longer C1$\cdots$O$_x$ distance than the reaction intermediate). Nevertheless, the sp^3 hybridization of the anomeric carbon and the C1$\cdots$O5 distance (1.40 ± 0.04 Å) are equivalent to the reaction intermediate.

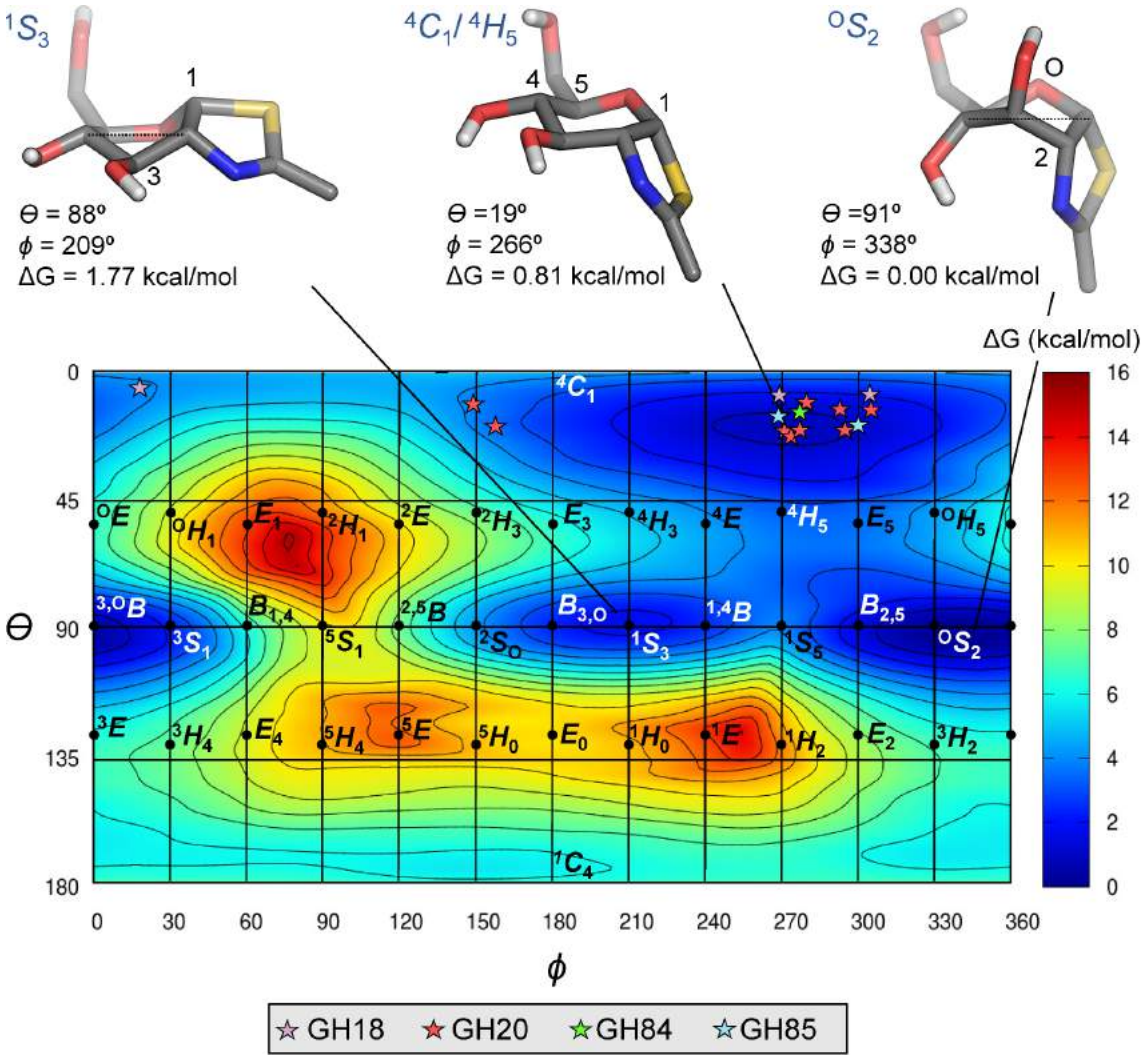

Figure 6.7. Metadynamics calculation of the conformational landscape of Glc-thiazoline. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

The PUGNAc inhibitor (Figure 6.3, bottom right) displays a sp^2 hybridized anomeric carbon, since it has a double bond involving the *N*-phenyl carbamate group. This inhibitor is a nanomolar inhibitor of GH3, GH18, GH20 and GH84 enzymes.[53,59,219,220] The resulting conformational FES is shown in Figure 6.8, displaying two free energy minima that are both stabilized by an intramolecular hydrogen bond between O_x and the hydroxyl at position 3. The global minimum corresponds to $^4C_1/\,^4E$ conformation and the second local minimum to $^5S_1$ conformation, being 3.19 kcal/mol higher than the global minimum. Remarkably, the conformation of PUGNAc in the active site of GHs obtained by X-ray crystallography is closer to ^4E than $^4C_1$, which suggests that each binding site shifts the conformation preference of PUGNAc. This effect is most probably caused by the interactions established between the

N-phenyl carbamate group from PUGNAc and the positive subsites of the inhibited GH. Consequently, the orientation of the aglycon part of the inhibitor affects the conformation of the pyranose ring. In terms of pyranose conformation, PUGNAc reassembles more the conformational tendencies observed in GlcNAc oxocarbenium ion (Figure 6.6) than Glc-ox/ox^{+} (Figures 6.4 and 6.5). Nevertheless, as occurred with Glc-thiazoline, free energy relationships experiments indicates that PUGNAc behaves as a poor TS mimic.[215] Regarding structural features, the $^{4}E/^{4}C_{1}$ conformation of PUGNAc shows a C1···O5 distance of 1.35 ± 0.03 Å, closer to Glc-ox^{+} (1.33 ± 0.03 Å) than GlcNAc oxocarbenium ion (1.28 ± 0.02 Å). Moreover, C3-C2-C1-O5 and C2-C1-O5-C5 dihedral angles are far from planarity (34.4 ± 4.4º and -32.5 ± 4.6º, respectively).

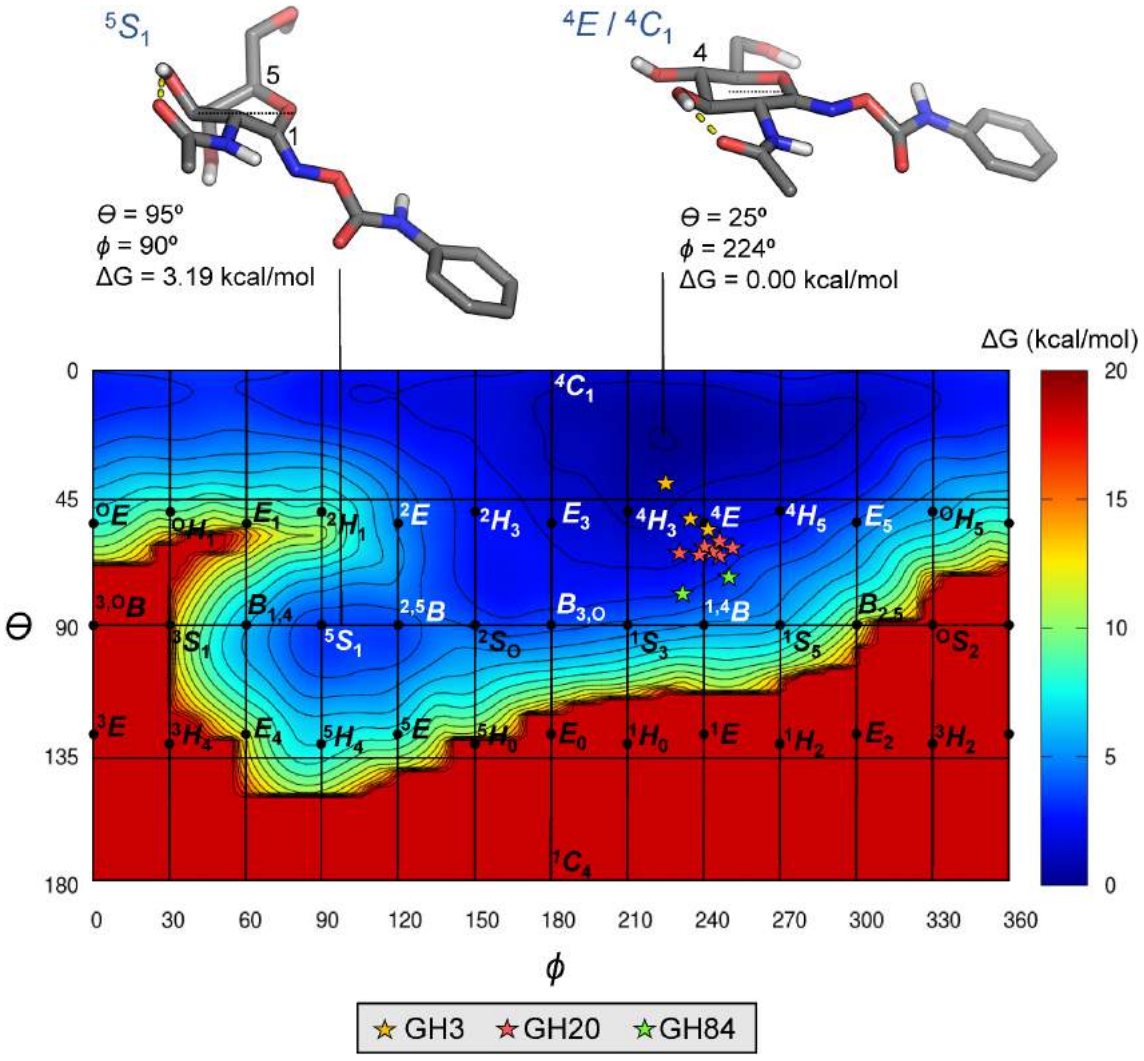

Figure 6.8. Metadynamics calculation of the conformational landscape of PUGNAc. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

Therefore, despite the sp^2 hybridization of C1 in PUGNAc, this compound should be classified as ground state analogue inhibitor (i.e., an inhibitor that mimics the substrate at the MC) that establish serendipitous interactions between the aglycon part of the inhibitor with hydrophobic residues from GHs. Indeed, *N*-acetylglucosamino-1,5-lactone oxime (LOGNAc), that lacks the phenyl carbamate group of PUGNAc, binds 30-fold less tightly than PUGNAc to a GH84 *O*-GlcNAcase,[221] highlighting the importance of these interactions.

Finally, we studied DNJ-NAc (Figure 6.3, bottom centre), an imino sugar GH inhibitor that has a positively charged nitrogen at the position of the pyranose oxygen and a 2-acetamido group, thus it mimics the charge of the reaction TS. DNJ-NAc is a nanomolar inhibitor of enzymes from families GH3, GH20, while GH84 is inhibited at µM level.[222,223]

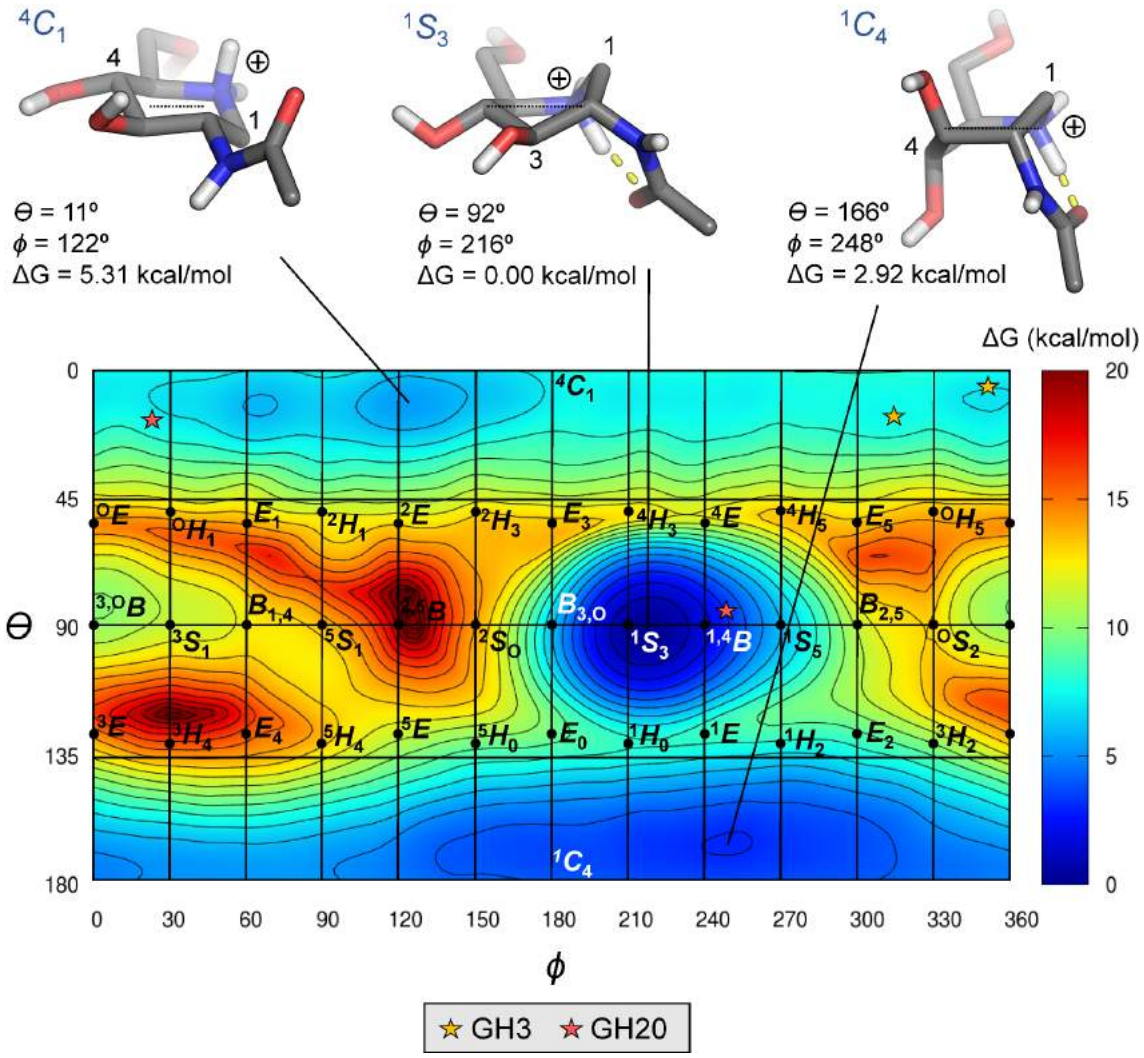

Figure 6.9. Metadynamics calculation of the conformational landscape of DNJ-NAc. Top, structure, conformation, energy, Θ and ϕ for each minimum in the conformational FES (Mercator representation, bottom). Contour lines are at 1 kcal/mol.

The obtained conformational FES for DNJ-NAc displays three different minima (Figure 6.9). The global minimum corresponds to $^{1}S_{3}$ conformation and the other to both chairs conformations ($^{4}C_{1}$ and $^{1}C_{4}$), which are 5.31 and 2.92 kcal/mol higher in energy, respectively. Both conformations that exhibit the anomeric carbon above the plain ($^{1}S_{3}$ and $^{1}C_{4}$) are stabilized by the intramolecular hydrogen bond between the O_x from the acetamido group and the positively charged nitrogen. However, DNJ-NAc in almost all X-ray crystal structures where it is complexed shapes a $^{4}C_{1}$ conformation, despite being the less stable conformer. Since GH3 *N*-acetyl-β-D-glucosaminidases operate via the classical retaining mechanism, it is expected that they show different interactions as in enzymes that perform substrate-assisted catalysis. This is the reason why the *N*-acetyl group of DNJ-NAc shows a different rotameric form in GH3 active sites (PDB codes 5G5K and 5LY7),[222] compared with GH20 enzymes. In GH3 enzymes, the nitrogen atom from *N*-acetyl group points towards below the plain (as the $^{4}C_{1}$ chair structure shown in Figure 6.9), interacting with the nucleophile residue. In contrast, the conformation of the acetamido group in GH20 structures (PDB codes 5BXR and 3SUV) is equivalent to the typically observed in GHs that use the substrate-assisted mechanism, where the nitrogen interacts with the assisting residue and O_X atom is closer to the anomeric carbon. Overall, the conformational properties of DNJ-NAc do not reassemble those of neither Glc-ox/ox^{+} nor the GlcNAc oxocarbenium ion species. This suggests that DNJ-NAc is not a good shape inhibitor for GHs that perform substrate-assisted catalysis.

In conclusion, we characterized the conformational landscape of the species involved in substrate-assisted catalysis, showing that not only conformation but also anomeric charge and intramolecular bond distances are important when assessing the TS mimicry of a GH inhibitor.

6.3. Conclusions

- We computed the conformational landscape of the species involved in the substrate-assisted mechanism (Glc-ox/ox$^+$ and GlcNAc oxocarbenium ion), which are in good agreement with the conformations observed in X-ray structures of GHs and our QM/MM simulations in previous chapters. The protonation of the Glc-ox reaction intermediate does not significantly modify the conformational map of the sugar.

- The pyranose ring conformation of Glc-thiazoline, PUGNAc and DNJ-NAc carbohydrate-based inhibitors was investigated. Our simulations reveal that their specificity for GHs that operate via substrate-assisted mechanism can be assessed by analysing their conformational properties, in particular how their FES resembles to that of the Glc-ox/ox$^+$ and GlcNAc oxocarbenium ions. However, for PUGNAc and Glc-thiazoline, this is not enough to predict whether the inhibitor is a TS-mimic or ground state binder, thereby further investigations are required.

6.4. Computational details

Systems preparation

The initial structures for Glc-ox, Glc-ox$^+$, Glc-thiazoline and GlcNAc oxocarbenium ion molecules were built by using as scaffold the Glc-ox molecule found inside the *Sm*ChiB X-ray structure (PDB entry 1E6Z). The initial coordinates for PUGNAc and DNJ-NAc were taken from the following X-ray structures: PDB entries 2CBJ and 3SUV, respectively. Further modifications of each compound such as adding hydrogens or atom replacement/elimination were carried out by means of Molefacture tool implemented in VMD software.[137] *Ab initio* MD simulations of all isolated molecules were performed within the Car-Parrinello (CP) approach.[90] The electronic structure was described within DFT, using the PBE generalized gradient-corrected approximation.[138] Kohn-Sham orbitals were expanded in a plane wave basis set with a kinetic energy cut-off of 70 Ry. Norm-conserving ab initio pseudopotentials were employed, generated within the Troullier-Martins scheme.[139] Each system was enclosed in an orthorhombic isolated box. The dimensions for Glc-ox, Glc-ox$^+$ and Glc-thiazoline boxes were 14.5 × 14.5 × 14.5 Å^3, while for GlcNAc oxocarbenium ion, DNJ-NAc and PUGNAc were 16.0 × 16.0 × 16.0 Å^3, 15.0 × 15.0 × 15.0 Å^3 and 22.0 × 22.0 × 22.0 Å^3, respectively. The fictitious electron mass for each system was tested. A fictitious electron mass of 600 a.u. was used for Glc-ox, Glc-ox$^+$, Glc-thiazoline and GlcNAc oxocarbenium ion. For DNJ-NAc and PUGNAc systems, the fictitious electron mass was set at 700 a.u. A time step of 0.12 fs was used. The following protocol for the QM MD simulations was used. First, the molecular structure was minimized until the maximal component of the nuclear is lower than $1 \cdot 10^{-5}$ a.u. Afterwards, each molecule was equilibrated during 10 ps at 300 K employing the Nosé-Hoover thermostat.[140,141]

Metadynamics of sugar puckering

The conformational FES of the molecules under investigation was explored by means of metadynamics[97] taking as CVs the Cremer-Pople puckering coordinates.[9] In particular, q_x, q_y and q_z divided by the Q amplitude (CV_1, CV_2 and CV_3). The metadynamics algorithm was provided by Plumed 2 plugin.[103]

The initial height of the gaussian-like potential for all systems was set to 0.6 kcal/mol and added every 250 MD time steps. Once the whole free energy space was explored, the height of the Gaussian terms was reduced to 0.2 kcal/mol and the deposition time duplicated to 500 MD time steps to facilitate convergence of the FES. The width of the Gaussian terms was set according to their oscillations in the free dynamics, which corresponded to 0.05, 0.05, 0.02 rad for CV_1, CV_2 and CV_3, respectively. The simulations were stopped when energy differences among the energy wells remain constant. This corresponded to 8350 deposited Gaussians in the case of Glc-ox, Glc-ox$^+$ and Glc-thiazoline compounds; while 5800 for the GlcNAc oxocarbenium ion. For DNJ and PUGNAc inhibitors, 8500 and 2500 Gaussians were deposited, respectively.

The FES was projected against the θ and ϕ puckering coordinates in order to obtain the Mercator representation using the reweighting method described by Branduardi et. al.[224] Structures from each metadynamics simulation were selected for structural analysis considering a 0.10 rad interval at the θ and ϕ space, centred at each free energy minimum.

7. Conclusions

The substrate-assisted mechanism followed by several retaining GHs has been investigated by means of computational techniques based on MD techniques and combined with metadynamics as an enhanced sampling method. We selected enzymes among several GH families in order to catch their subtitle differences in terms of catalytic machinery and mechanism, thereby obtaining a big picture of this specific reaction mechanism. We inspected enzymes that possesses a catalytic dyad with an insertion in their sequence (Cat-X-Cat) in Chapters 3 and 5; and a GH with their catalytic residues adjacent in the sequence (Cat-Cat) in Chapter 4. We have found that this feature impacts directly on their catalytic mechanism. Finally, we aimed to map the conformational preferences of isolated carbohydrates involved in substrate-assisted catalysis, and their related inhibitors, to comprehend their nature (Chapter 6). The investigation of GH reaction mechanisms provides an in-depth understanding of the catalytic factors that play a role in catalysis, which ultimately might guide further drug design and enzyme engineering.

The conclusions obtained in this doctoral Thesis are the following:

- The reaction mechanism of *Sm*ChiB, a GH18 chitinase that operates via substrate-assisted, has been fully elucidated. The reaction intermediate, whose nature was still controversial, features a neutral Glc-ox with a $^{4}C_{1}/^{4}H_{5}$ conformation. The deprotonation of the initial GlcNAc is performed by the catalytic dyad Asp142-X-Glu144, that define a well-defined hydrogen bond network with the substrate as well.

- The observed features in *Sm*ChiB are very likely transferable to other GHs that exhibit a similar active site: Asp-X-Glu catalytic residues and a Tyr residue to stabilize the *N*-acetyl oxygen (O_x). We prove this in the GH56 hyaluronidase (*Bv*Hya), which also produces a neutral Glc-ox reaction intermediate by means a hydrogen-bond network.

- The single point mutation of the acid/base residue from Asp to Asn in GH84 *O*-GlcNAcases converts hydrolytic enzymes (GHs) in synthetic phosphorylases (GPs). We computed the phosphate ion distribution and the reaction mechanism of hydrolysis and phosphorylation, finding that this mutation alters both the electrostatic potential and architecture of their active sites, allowing the phosphate ions to visit and react in a more positively charged environment.

- The reaction intermediate in *Tt*OGA (GH84 *O*-GlcNAcase) is protonated (Glc-ox$^+$), since the catalytic dyad is adjacent in the sequence (Asp119-Asp120) and thus cannot establish a hydrogen bond network that ultimately deprotonates the substrate. Lys43 interacts with the assisting residue Asp119 by an ionic interaction that lowers its pK_a, thereby not allowing the Glc-ox$^+$ deprotonation.

- *Sp*GH85, a GH85 ENGase, possesses an Asn assisting residue, instead of the more commonly observed Asp (Asn335-X-Glu337). This enzyme is able to stabilize a both Glc-ox and Glc-ox$^+$ reaction intermediates as long as the Asn is in its amide tautomeric form.

- We demonstrated that the enzymatic environment of *Sp*GH85 allows Asn335 to tautomerize in the obtained Michaelis complex with a *N*-glycan core substrate. This feature is very unusual in GH catalysis, only being reported in one experimental study. Further investigation is required to determine which reaction pathways actually occurs in this enzymes, since the computed first reaction step for both paths are almost isoenergetic. Thus, both Glc-ox and Glc-ox$^+$ reaction intermediates may take place, involving or not imidic acid tautomerization of Asn335, respectively.

- The conformational analysis of the isolated species involved in the substrate-assisted mechanism (Glc-ox/ox$^+$ and GlcNAc oxocarbenium ion) allows to identify and map the conformational preferences of these molecules. Glc-ox and Glc-ox$^+$ do not differ in terms of sugar conformation. By computing the conformational free energy landscape of the carbohydrate-based inhibitors Glc-thiazoline, PUGNAc and DNJ-NAc we comprehended their ability to inhibit GHs that operate via substrate assisted catalysis.

- Two substrate-assisted mechanisms have been unravelled by means of QM/MM metadynamics calculations. While one involves several proton transfer events and a neutral Glc-ox intermediate, in the other there are no proton transfers between the catalytic dyad and the substrate, producing a Glc-ox$^+$ intermediate. The active site environment of the assisting residue and the spacial disposition of the catalytic dyad most probably determine which mechanism is employed by a certain GH.

List of publications

This Thesis has resulted in the following publications (‡ equal contribution):

J. Coines, M. Alfonso-Prieto, X. Biarnés, A. Planas and C. Rovira. Oxazoline or oxazolinium ion? the protonation state and conformation of the reaction intermediate of chitinase enzymes revisited. *Chem. Eur. J.* 2018, 24, 19258–19265.

J. Coines, L. Raich and C. Rovira. Modeling catalytic reaction mechanisms in glycoside hydrolases. *Curr. Op. Chem. Bio*., 2019, 53, 183–191.

D. Teze‡, **J. Coines**‡, L. Raich, V. Kalichuk, C. Solleux, C. Tellier, C. Andre-Miral, B. Svensson and C. Rovira. A single point mutation converts a GH84 O-GlcNAc hydrolase into phosphorylases. Experimental and theoretical evidence. *J. Am. Chem. Soc*. 2020, 142, 2120–2124.

Publications not included in this Thesis and unrelated to its main topic:

T. Belz, Y. Jin, **J. Coines**, C. Rovira, G. J. Davies, S. J. Williams. An atypical interaction explains the high-affinity of a non-hydrolyzable S-linked 1,6-α-mannanase inhibitor. *Chem. Commun*. 2017, 53, 9238–9241.

J. Coines, S. Acosta-Gutierrez, I. Bodrenko, C. Rovira and M. Ceccarelli. Glucose transport via the pseudomonad porin OprB: implications for the design of Trojan Horse anti-infectives. *Phys. Chem. Chem. Phys*. 2019, 21, 8457–8463.

S. Alonso-Gil, **J. Coines**, I. André and C. Rovira. Conformational Itinerary of Sucrose During Hydrolysis by Retaining Amylosucrase. *Front. Chem*. 2019, 7, 269.

R. Pequerul, J. Vera, J. Giménez-Dejoz, I. Crespo, **J. Coines**, S. Porté, C. Rovira, X. Parés and J. Farrés. Structural and kinetic features of aldehyde dehydrogenase 1A (ALDH1A) subfamily members, cancer stem cell markers active in retinoic acid biosynthesis. *Arch. Biochem. Biophys.* 2020, 681, 108256.

M. A. Morais, **J. Coines**, M. N. Domingues, R. A. Pirolla, C. C. C. Tonoli, C. R. Santos, V. P. M. Martins, C. Rovira, M. T. Murakami. Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations (*accepted*, *Nat. Commun*.).

N. G. S. McGregor‡, **J. Coines**‡, V. Borlandelli, S. Amaki, M. Artola, A. Nin-Hill, D. Linzel, C. Yamada, T. Arakawa, G. A. van der Marel, J. D. C. Codée, S. Fushinobu, H. S. Overkleeft, C. Rovira, and G. J. Davies. Cysteine nucleophiles in glycosidase catalysis: application of a covalent β-L-arabinofuranosidase inhibitor (*accepted, Angew. Chem. Int. Ed.*).

D. Teze‡, **J. Coines**‡, F. Fredslund, K. D. Dubey, G. Bidart, P. D. Adams, J. E. Dueber, C. Rovira, B. Svensson and D. H. Welner. O-/N-/S-specificity in glycosyltransferases: from mechanistic understanding to engineering (*under revision, ACS Catal.*).

Oral and poster presentations

The work presented in this Thesis has been presented as oral and poster communications in workshops and congresses:

The reaction intermediate of GH18 chitinases: substrate conformation and protonation state. Poster. 19th European Carbohydrate Symposium (EUROCARB), Barcelona (Spain), 2-6 July, 2017.

Oxazoline or oxazolinium ion? The protonation state and conformation of the reaction intermediate of chitinase enzymes revisited. Invited talk. Advanced Materials Workshop. Duni (Bulgary), 11-14 Sept, 2018.

Oxazoline or oxazolinium ion? The reaction mechanism of GH18 chitinases. Carbohydrate Bioengineering Meeting. Flash talk and poster. Toulouse (France), May 19-22, 2019.

Mechanistic insights into substrate-assisted glycoside hydrolases. Invited talk. X IQTCUB Symposium. Barcelona (Spain), 31 May, 2019.

Oxazoline Or Oxazolinium Ion? The Protonation State and Conformation of the Reaction Intermediate of Chitinase Enzymes Revisited. EurocarbXX. Flash talk and poster. Leiden (Netherlands). 30 June-4 July, 2019.

The Sweet Gateway of Pseudomonads: Sugar Transport Across the OprB Porin. EurocarbXX. Poster. Leiden (Netherlands). 30 June-4 July, 2019.

Collaborations

The investigation of the GH18 chitinase *Sm*ChiB (Chapter 3), was conducted in collaboration with:

> **Professors Antoni Planas and Xevi Biarnés**. Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Via Augusta, 390, 08017 Barcelona, Spain.

The experimental work of *Tt*OGA *O*-GlcNAcase in Chapter 4 was mostly performed by **Dr. David Teze** in the group led by:

> **Professor Birte Svensson.** Department of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark.

Additionally, I visited for three months and a half a computational biophysics group in Italy, thanks to the support of a PhD scholarship from MINECO (FPI-BES-2015-072055). The research performed resulted in the following publication: Coines et al. *Phys. Chem. Chem. Phys*., 2019, 21, 8457–8463. The host research group is led by:

> **Professor Matteo Ceccarelli.** Department of Physics, University of Cagliari, Cittadella Universitaria, S.P.8-km 0.700, 09042 Monserrato, Cagliari, Italy.

Acknowledgments

Que em trobi escrivint aquestes línies no seria possible sense les persones que he tingut el plaer de creuar-me al llarg del camí. Comencem doncs:

A la Carme, per l'oportunitat que m'has brindat podent fer la tesi sota la teva direcció i el suport amb el que he comptat al llarg d'aquests anys. Gràcies per haver-me introduït en el món del *modeling* d'enzims, per tots els cafès plegats — que seguim gaudint a dia d'avui — i tot el que m'has ensenyat. Has sigut una mentora exemplar.

A la Mertxe, perquè tot i que els meus inicis queden ben lluny, la teva ajuda en els primers passos va marcar la diferència. Em vas ensenyar molt, com aprendre per mi mateix però sense deixar de sentir que podia comptar amb tu. Tot i que em diguis pilota (in italià sona pitjor), t'ho seguiré agraint sempre.

Al Lluís, perquè no tinc més que paraules d'admiració per la teva tenacitat, rigor i, sobretot, la teva humilitat. Ens has donat un cop de mà sempre que t'hem necessitat. He trobat un amic molt particular en tu. Les nostres *rutas de la fruta*, les pràctiques de QO i els dinars plegats me'ls enduc per sempre.

A l'Alba, per ser companys de viatge durant tots aquests anys. Encara recordo quan compartíem taula tot fent el tutorial de CPMD. Han estat molts copets a l'esquena i suport mutu. No podria haver tingut més sort en coincidir amb tu. Gràcies per fer-me el dia a dia millor.

A la Bea, Irene i Oriol, les noves generacions, la nova fornada. Tant l'Alba i jo estem contents i tranquils en deixar el despatx en bones mans. Gràcies per totes les estones plegats i aguantar-me en les meves crisis de final de tesi. Us auguro un bon futur.

Devo ringraziare a tutte le persone che ho incontrato in Sardegna per avere reso la mia visita una bellissima esperienza. Grazie a i colleghi dell Dipartimento di Fisica, in particolare all Professore Ceccarelli, Matteo e'stato un vero piacere lavorare con te. Grazie per insegnarmi non solo tutte le tue conoscenze sulle porine ma anche per avermi accolto e per portarmi a fare surf con la tua famiglia. A Sílvia, mi canaria perdida, por haberme ayudado tantísimo des del día que te conocí. Tengo muchas ganas de verte, así que miraré vuelo a Baleares en cuanto pueda. A Francesco, perche ho trovato un bel amico romano in Sardegna. Ci vediamo a Genova presto!

To Mariana, because it has been a great pleasure working along your side. You are an amazing example of resilience and hard work. I have learnt a lot from you. Hope to see you again soon, somewhere fancier than our hotel in Toulouse!

Als amics, tot i que no sé ni per on començar. Seguiré un criteri cronològic. De l'institut, Sara i Ari, que passen els anys i seguim aquí, quina sort. Maria Núñez, gràcies per mai deixar de ser-hi, també. De l'Armentera, a la Mireia Boher, no sols per tots els anys que portem a les esquenes sinó perquè ets de les millors persones que conec. Comptar amb tu és una sort immensa. Marina, Blanca, Cervelló i Villalba també us tinc presents, gràcies per fer-me gaudir de l'Empordà en bona companyia. De bioquímica, el grup amb qui he gaudit una de les millors etapes de la meva vida, *BQ shore*, us porto a tots dins i que no faltin els nostres retrobaments. Destacar, merescudament, Dani i Fer (tio!), que sou un pilar fonamental en la meva vida i doneu tot el sentit del món a la paraula *bromance*. Del CSIC-IQAC, Broto i Sanchís sou com les meves germanes grans però sense el com. No me olvido de ti, Diego, que soportaste un sabotaje a tu propuesta de matrimonio en las colinas escocesas. A los lunchicompis, Héctor, Jordi y Dani, no sólo por los mediodías de risas y desconexión, también por los cines y los viernes de cervezas. A Lorena, por todas las *machines* y el apoyo brindado. Al final se te coge cariño, quién lo hubiese dicho en Bulgaria. A Nuria Senar, por un matrimonio disfuncional que nunca funcionó pero que ha dejado un pozo infinito de estima. A Irene, ladrona de olas con alevosía, que, aunque estemos siempre lejos, no te cambiaría ni por un gofre. Muchas ganas de ver mundo y coger olas a tu lado. Y no me digas que reme, porque no, no va a pasar.

Y, para acabar por todo lo alto, a mis padres, por haber creído en mí más que yo mismo, desde el principio. Sin vosotros absolutamente nada de esto sería posible. Me faltan vidas para agradecéroslo.

Bibliography

(1) Stick, R.; Williams, S. J. *Carbohydrates: The Essential Molecules of Life*; 2009.

(2) Suhas; Gupta, V. K.; Carrott, P. J. M.; Singh, R.; Chaudhary, M.; Kushwaha, S. Cellulose: A Review as Natural, Modified and Activated Carbon Adsorbent. *Bioresour. Technol.* 2016, *216*, 1066–1076.

(3) Pauly, M.; Keegstra, K. Cell-Wall Carbohydrates and Their Modification as a Resource for Biofuels. *Plant J.* 2008, *54* (4), 559–568.

(4) Deng, J.; Li, M.; Wang, Y. Biomass-Derived Carbon: Synthesis and Applications in Energy Storage and Conversion. *Green Chem.* 2016, *18* (18), 4824–4854.

(5) Cummings, R. D. "Stuck on Sugars – How Carbohydrates Regulate Cell Adhesion, Recognition, and Signaling." *Glycoconj. J.* 2019, *36* (4), 241–257.

(6) Lemieux, R.U. , Stevens, J. D. The Proton Magnetic Resonance Spectra and Tautomeric Equilibria of Aldoses in Deuterium Oxide. *Can. J. Chem.* 1966, *44* (3), 249–262.

(7) Cramer, C. J.; Truhlar, D. G.; French, A. D. Exo-Anomeric Effects on Energies and Geometries of Different Conformations of Glucose and Related Systems in the Gas Phase and Aqueous Solution. *Carbohydr. Res.* 1997, *298* (1–2), 1–14.

(8) Karlson, P. Conformational Nomenclature for Five and Six-Membered Ring Forms of Monosaccharides and Their Derivatives. Recommendations 1980. *Eur. J. Biochem.* 1980, *111* (Copyright (C) 2013 American Chemical Society (ACS). All Rights Reserved.), 295–298.

(9) Cremer, D.; Pople, J. a. A General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* 1975, *97* (6), 1354–1358.

(10) André, I.; Potocki-Véronèse, G.; Barbe, S.; Moulis, C.; Remaud-Siméon, M. CAZyme Discovery and Design for Sweet Dreams. *Curr. Opin. Chem. Biol.* 2014, *19* (1), 17–24.

(11) Vocadlo, D. J.; Davies, G. J. Mechanistic Insights into Glycosidase Chemistry. *Curr. Opin. Chem. Biol.* 2008, *12* (5), 539–555.

(12) Lopez-Casado, G.; Urbanowicz, B. R.; Damasceno, C. M.; Rose, J. K. Plant Glycosyl Hydrolases and Biofuels: A Natural Marriage. *Curr. Opin. Plant Biol.* 2008, *11* (3), 329–337.

(13) Sathya, T. A.; Khan, M. Diversity of Glycosyl Hydrolase Enzymes from Metagenome and Their Application in Food Industry. *J. Food Sci.* 2014, *79* (11), R2149–R2156.

(14) Danby, P. M.; Withers, S. G. Advances in Enzymatic Glycoside Synthesis. *ACS Chem. Biol.* 2016, *11* (7), 1784–1794.

(15) Dvir, H.; Harel, M.; McCarthy, A. a; Toker, L.; Silman, I.; Futerman, A. H.; Sussman, J. L. X-Ray Structure of Human Acid-Beta-Glucosidase, the Defective Enzyme in Gaucher Disease. *EMBO Rep.* 2003, *4* (7), 704–709.

(16) Hers, H. G. Alpha-Glucosidase Deficiency in Generalized Glycogenstorage Disease (Pompe's Disease). *Biochem. J.* 1963, *86* (1959), 11–16.

(17) Wood, S.; Nadler, H. L. Fabry's Disease : Absence of an α-Galactosidase Isozyme. 1972, 250–255.

(18) Suzuki, K.; Suzuki, Y. Globoid Cell Leucodystrophy (Krabbe's Disease): Deficiency of Galactocerebroside Beta-Galactosidase. *Proc. Natl. Acad. Sci. U. S. A.* 1970, *66* (2), 302–309.

(19) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* 2014, *42* (D1), 490–495.

(20) Garron, M.; Henrissat, B. The Continuing Expansion of CAZymes and Their Families. *Curr. Opin. Chem. Biol.* 2019, *53*, 82–87.

(21) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes.

(22) Davies, G.; Henrissat, B. Structures and Mechanisms of Glycosyl Hydrolases. *Structure* 1995, *3* (9), 853–859.

(23) Knott, B. C.; Haddad Momeni, M.; Crowley, M. F.; MacKenzie, L. F.; Götz, A. W.; Sandgren, M.; Withers, S. G.; Staìšhlberg, J.; Beckham, G. T. The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* 2014, *136* (1), 321–329.

(24) Gaiser, O. J.; Piotukh, K.; Ponnuswamy, M. N.; Planas, A.; Borriss, R.; Heinemann, U. Structural Basis for the Substrate Specificity of a Bacillus 1,3-1,4-β-Glucanase. *J. Mol. Biol.* 2006, *357* (4), 1211–1225.

(25) van Bueren, A. L.; Ghinet, M. G.; Gregg, K.; Fleury, A.; Brzezinski, R.; Boraston, A. B. The Structural Basis of Substrate Recognition in an Exo-β-d-Glucosaminidase Involved in Chitosan Hydrolysis. *J. Mol. Biol.* 2009, *385* (1), 131–139.

(26) Sinnott, M. L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem. Rev.* 1990, *90* (7), 1171–1202.

(27) Koshland, D. E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biol. Rev.* 1953, *28* (4), 416–436.

(28) Jongkees, S. a K.; Withers, S. G. Unusual Enzymatic Glycoside Cleavage Mechanisms. *Acc. Chem. Res.* 2014, *47* (1), 226–235.

(29) Teze, D.; Shuoker, B.; Chaberski, E. K.; Kunstmann, S.; Fredslund, F.; Nielsen, T. S.; Stender, E. G. P.; Peters, G. H. J.; Karlsson, E. N.; Welner, D. H.; et al. The Catalytic Acid-Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable α-Retaining and β-Inverting Activity in an *N*-Acetylgalactosaminidase from *Akkermansia Muciniphila*. *ACS Catal.* 2020, *10* (6), 3809–3819.

(30) Sobala, L. F.; Speciale, G.; Zhu, S.; Raich, L.; Sannikova, N.; Thompson, A. J.; Hakki, Z.; Lu, D.; Shamsi Kazem Abadi, S.; Lewis, A. R.; et al. An Epoxide Intermediate in Glycosidase Catalysis. *ACS Cent. Sci.* 2020, *6* (5), 760–770.

(31) Phillips, D. C. The Hen Egg-White Lysozyme Molecule. *Proc. Natl. Acad. Sci. U. S. A.* 1967, *57* (3), 483–495.

(32) Blake, C. C.; Koenig, D. F.; Mair, G. a; North, a C.; Phillips, D. C.; Sarma, V. R. Structure of Hen Egg-White Lysozyme. A Three-Dimensional Fourier Synthesis at 2 Angstrom Resolution. *Nature* 1965, *206* (986), 757–761.

(33) Davies, G. J.; Ducros, V. M.; Varrot, A.; Zechel, D. L. Mapping the Conformational Itinerary of Beta-Glycosidases by X-Ray Crystallography. *Biochem Soc Trans* 2003, *31* (Pt 3), 523–527.

(34) Coines, J.; Raich, L.; Rovira, C. Modeling Catalytic Reaction Mechanisms in Glycoside Hydrolases. *Curr. Opin. Chem. Biol.* 2019, *53*, 183–191.

(35) Davies, G. J.; Planas, A.; Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. *Acc. Chem. Res.* 2012, *45* (2), 308–316.

(36) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J. Structure of the Fusarium Oxysporum Endoglucanase I with a Nonhydrolyzable Substrate Analogue: Substrate Distortion Gives Rise to the Preferred Axial Orientation for the Leaving Group. *Biochemistry* 1996, *35* (48), 15280–15287.

(37) Biarnés, X.; Nieto, J.; Planas, A.; Rovira, C. Substrate Distortion in the Michaelis Complex of Bacillus 1,3-1,4-β-Glucanase: Insight from First Principles Molecular Dynamics Simulations. *J. Biol. Chem.* 2006, *281* (3), 1432–1441.

(38) Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. The Conformational Free Energy Landscape of β-D-Glucopyranose. Implications for Substrate Preactivation in β-Glucoside Hydrolases. 2007, 10686–10693.

(39) Ardèvol, A.; Biarnés, X.; Planas, A.; Rovira, C. The Conformational Free-Energy Landscape of β-d-Mannopyranose: Evidence for a 1S5 → B2,5 → OS2 Catalytic Itinerary in β-Mannosidases. *J. Am. Chem. Soc.* 2010, *132* (45), 16058–16065.

(40) Iglesias-Fernández, J.; Raich, L.; Ardèvol, A.; Rovira, C. The Complete Conformational Free Energy Landscape of β-Xylose Reveals a Two-Fold Catalytic Itinerary for β-Xylanases. *Chem. Sci.* 2015, *6* (2), 1167–1177.

(41) Gloster, T. M.; Davies, G. J. Glycosidase Inhibition: Assessing Mimicry of the Transition State. *Org. Biomol. Chem.* 2010, *8* (2), 305–320.

(42) Gloster, T. M.; Vocadlo, D. J. Developing Inhibitors of Glycan Processing Enzymes as Tools for Enabling Glycobiology. *Nat. Chem. Biol.* 2012, *8* (8), 683–694.

(43) Wu, L.; Armstrong, Z.; Schröder, S. P.; de Boer, C.; Artola, M.; Aerts, J. M.; Overkleeft, H. S.; Davies, G. J. An Overview of Activity-Based Probes for Glycosidases. *Curr. Opin. Chem. Biol.* 2019, *53*, 25–36.

(44) J. F. Stoddart. *Stereochemistry of Carbohydrates*; John Wiley & Sons Inc, 1971.

(45) Nin-Hill, A.; Rovira, C. The Catalytic Reaction Mechanism of β-Galactocerebrosidase, the Enzyme Deficient in Krabbe Disease. *ACS Catal.* 2020, *10* (29), 12091−12097.

(46) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. Catalysis by Hen Egg-White Lysozyme Proceeds via a Covalent Intermediate. *Nature* 2001, *412* (6849), 835–838.

(47) Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W. Stereochemistry of Chitin Hydrolysis by a Plant Chitinase/Lysozyme and X-Ray Structure of a Complex with Allosamidin: Evidence for Substrate Assisted Catalysis. *Biochemistry* 1995, *34* (48), 15619–15623.

(48) Knapp, S.; Vocadlo, D.; Gao, Z.; Kirk, B.; Lou, J.; Withers, S. G. NAG-Thiazoline, An *N*-Acetyl-β-Hexosaminidase Inhibitor That Implicates Acetamido Participation. *J. Am. Chem. Soc.* 1996, *118*, 6804–6805.

(49) Piszkiewicz, D.; Bruice, T. C. Glycoside Hydrolysis. II. Intramolecular Carboxyl and Acetamido Group Catalysis in β-Glycoside Hydrolysis. *J. Am. Chem. Soc.* 1968, *90* (8), 2156–2163.

(50) Paulsen, H. Advances in Selective Chemical Syntheses of Complex Oligosaccharides. *Angew. Chemie Int. Ed. English* 1982, *21* (3), 155–173.

(51) Synstad, B.; Gåseidnes, S.; Van Aalten, D. M. F.; Vriend, G.; Nielsen, J. E.; Eijsink, V. G. H. Mutational and Computational Analysis of the Role of Conserved Residues in the Active Site of a Family 18 Chitinase. *Eur. J. Biochem.* 2004, *271* (2), 253–262.

(52) Vaaje-Kolstad, G.; Houston, D. R.; Rao, F. V.; Peter, M. G.; Synstad, B.; Van Aalten, D. M. F.; Eijsink, V. G. H. Structure of the D142N Mutant of the Family 18 Chitinase ChiB from *Serratia Marcescen*s and Its Complex with Allosamidin. *Biochim. Biophys. Acta - Proteins Proteomics* 2004, *1696* (1), 103–111.

(53) Dennis, R. J.; Taylor, E. J.; Macauley, M. S.; Stubbs, K. A.; Turkenburg, J. P.; Hart, S. J.; Black, G. N.; Vocadlo, D. J.; Davies, G. J. Structure and Mechanism of a Bacterial β-Glucosaminidase Having *O*-GlcNAcase Activity. *Nat. Struct. Mol. Biol.* 2006, *13* (4), 365–371.

(54) Umekawa, M.; Huang, W.; Li, B.; Fujita, K.; Ashida, H.; Wang, L.; Yamamoto, K. Mutants of Mucor Hiemalis Endo-β-*N*-Acetylglucosaminidase Show Enhanced Transglycosylation and Glycosynthase-like Activities. *J. Biol. Chem.* 2008, *283* (8), 4469–4479.

(55) Frump, J. A. Oxazolines. Their Preparation, Reactions, and Applications. *Chem. Rev.* 1971, *71* (1884), 483.

(56) Greig, I. R.; Zahariev, F.; Withers, S. G. Elucidating the Nature of the *Streptomyces Plicatus* β-Hexosaminidase-Bound Intermediate Using *Ab Initio* Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 2008, *130* (51), 17620–17628.

(57) Pitha, J.; Jonas, J.; Kovář, J.; Bláha, K. Zur Konfiguration Stickstoffhaltiger Verbindungen XIII. Über Die Darstellung Und Tautomerie Der Aminooxazoline. *Collect. Czechoslov. Chem. Commun.* 1961, *26* (3), 834–846.

(58) van Aalten, D. M.; Komander, D.; Synstad, B.; Gåseidnes, S.; Peter, M. G.; Eijsink, V. G. Structural Insights into the Catalytic Mechanism of a Family 18 Exo-Chitinase. *Proc. Natl. Acad. Sci. U. S. A.* 2001, *98* (16), 8979–8984.

(59) Robb, M.; Robb, C. S.; Higgins, M. A.; Hobbs, J. K.; Paton, J. C.; Boraston, A. B. A Second β-Hexosaminidase Encoded in the *Streptococcus Pneumoniae* Genome Provides an Expanded Biochemical Ability to Degrade Host Glycans. *J. Biol. Chem.* 2015, *290* (52), 30888–30900.

(60) Rau, A.; Hogg, T.; Marquardt, R.; Hilgenfeld, R. A New Lysozyme Fold. Crystal Structure of the Muramidase from *Streptomyces Coelicolor* at 1.65 Å Resolution. *J. Biol. Chem.* 2001, *276* (34), 31994–31999.

(61) Markovíc-Housley, Z.; Miglierini, G.; Soldatova, L.; Rizkallah, P. J.; Müller, U.; Schirmer, T. Crystal Structure of Hyaluronidase, a Major Allergen of Bee Venom. *Structure* 2000, *8* (10), 1025–1035.

(62) Abbott, D. W.; Macauley, M. S.; Vocadlo, D. J.; Boraston, A. B. *Streptococcus Pneumoniae* Endohexosaminidase D, Structural and Mechanistic Insight into Substrate-Assisted Catalysis in Family 85 Glycoside Hydrolases. *J. Biol. Chem.* 2009, *284* (17), 11676–11689.

(63) Roth, C.; Petricevic, M.; John, A.; Goddard-Borger, E. D.; Davies, G.; Williams, S. J. Structural and Mechanistic Insights into a *Bacteroides Vulgatus* Retaining *N*-Acetyl-β-Galactosaminidase That Uses Neighboring Group Participation. *Chem. Commun.* 2016, *52*, 11096–11099.

(64) He, Y.; Macauley, M. S.; Stubbs, K. A.; Vocadlo, D. J.; Davies, G. J. Visualizing the

Reaction Coordinate of an *O*-GlcNAc Hydrolase. *J. Am. Chem. Soc.* 2010, *132* (6), 1807–1809.

(65) Jitonnom, J.; Limb, M. A. L.; Mulholland, A. J. QM/MM Free-Energy Simulations of Reaction in *Serratia Marcescens* Chitinase b Reveal the Protonation State of Asp142 and the Critical Role of Tyr214. *J. Phys. Chem. B* 2014, *118* (18), 4771–4783.

(66) Roth, C.; Petricevic, M.; John, A.; Goddard-Borger, E. D.; Davies, G. J.; Williams, S. J. Structural and Mechanistic Insights into a *Bacteroides Vulgatus* Retaining *N*-Acetyl-β-Galactosaminidase That Uses Neighbouring Group Participation. *Chem. Commun.* 2016, *52* (74), 11096–11099.

(67) Tamai, E.; Yoshida, H.; Sekiya, H.; Nariya, H.; Miyata, S.; Okabe, A.; Kuwahara, T.; Maki, J.; Kamitori, S. X-Ray Structure of a Novel Endolysin Encoded by Episomal Phage PhiSM101 of *Clostridium Perfringens*. *Mol. Microbiol.* 2014, *92* (2), 326–337.

(68) Alteen, M. G.; Oehler, V.; Nemčovičová, I.; Wilson, I. B. H.; Vocadlo, D. J.; Gloster, T. M. Mechanism of Human Nucleocytoplasmic Hexosaminidase D. *Biochemistry* 2016, acs.biochem.5b01285.

(69) Vocadlo, D. J.; Withers, S. G. Detailed Comparative Analysis of the Catalytic Mechanisms of *N*-Acetylglucosaminidases from Families 3 and 20 of Glycoside Hydrolases. *Biochemistry* 2005, *44*, 12809–12818.

(70) Bottaro, S.; Lindorff-Larsen, K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? *Science (80-.).* 2018, *361*, 355–360.

(71) Zinovjev, K.; Tuñón, I. Quantifying the Limits of Transition State Theory in Enzymatic Catalysis. *Proc. Natl. Acad. Sci.* 2017, No. 30, 201710820.

(72) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Jacob, I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations : V1 . 0. 2018, 1–27.

(73) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* 1976, *159* (1), 98–103.

(74) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley, 2004.

(75) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* 2015, *11* (8), 3696–3713.

(76) Brooks, B. R. et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 2009, *30*, 1545–1614.

(77) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 1988, *110* (6), 1657–1666.

(78) Kirschner, K. N.; Yongye, A.; Tschampel M., S.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* 2008, *29* (4), 622–655.

(79) Raman, E. P.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *J. Phys. Chem. B* 2010, *114* (40), 12981–12994.

(80) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983, *79* (2), 926.

(81) Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E. Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent. Sci.* 2020, *6* (2), 189–196.

(82) Vymětal, J.; Jurásková, V.; Vondrasek, J. AMBER and CHARMM Force Fields Inconsistently Portray the Microscopic Details of Phosphorylation. *J. Chem. Theory Comput.* 2018, *15* (1), 665–679.

(83) Mongan, J.; Case, D. A.; McCammon, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* 2004, *25* (16), 2038–2048.

(84) D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E.Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. M, X. W. and P. A. K. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 1995, *117* (19), 5179–5197.

(85) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. 1965, *385* (1951).

(86) A. D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* 1988, *38* (6), 3098–3100.

(87) Chengteh Lee, Weitao Yang, and R. G. P. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* 1988, *37* (2), 785–789.

(88) Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from First Principles QM/MM Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 2015, *137*, 7528–7547.

(89) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and α-Maltose. *J. Chem. Theory Comput.* 2016, *12* (12), 6157–6168.

(90) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* 1985, *55* (22), 2471–2474.

(91) Raich, L.; Nin-Hill, A.; Ardèvol, A.; Rovira, C. Enzymatic Cleavage of Glycosidic Bonds: Strategies on How to Set Up and Control a QM/MM Metadynamics Simulation, 1st ed.; Elsevier Inc., 2016; Vol. 577.

(92) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* 1976, *103* (2), 227–249.

(93) Singh, U. C.; Kollman, P. A. A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH_3Cl + Cl^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* 1986, *7* (6), 718.

(94) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comp. Chem.* 1990, *11* (6), 700–733.

(95) Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car-Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* 2002, *116* (16), 6941–6947.

(96) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* 2016, *67* (1), 159–184.

(97) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* 2002, *99* (20), 12562–12566.

(98) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011, *1* (5), 826–843.

(99) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. *J. Phys. Chem. B* 2005, *109*, 6676–6687.

(100) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* 2008, *100* (2), 1–4.

(101) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* 2007, *111* (17), 4553–4559.

(102) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* 2006, *110* (8), 3533–3539.

(103) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* 2014, *185* (2), 604–613.

(104) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* 2009, *31* (16), 2785–2791.

(105) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 2010, *31* (2), 455–461.

(106) Adrangi, S.; Faramarzi, M. A. From Bacteria to Human: A Journey into the World of Chitinases. *Biotechnol. Adv.* 2013, *31* (8), 1786–1785.

(107) Arbia, W.; Arbia, L.; Adour, L.; Amrane, A. Chitin Extraction from Crustacean Shells Using Biological Methods – A Review. *Food Technol. Biotechnol.* 2013, *51* (1), 12–25.

(108) Kaur, S.; Dhillon, G. S. Recent Trends in Biological Extraction of Chitin from Marine Shell Wastes: A Review. *Crit. Rev. Biotechnol.* 2013, *8551* (1), 44–61.

(109) Muanprasat, C.; Chatsudthipong, V. Chitosan Oligosaccharide: Biological Activities and Potential Therapeutic Applications. *Pharmacol. Ther.* 2017, *170*, 80–97.

(110) Hamed, I.; Özogul, F.; Regenstein, J. M. Industrial Applications of Crustacean By-Products (Chitin, Chitosan, and Chitooligosaccharides): A Review. *Trends Food Sci. Technol.* 2016, *48*, 40–50.

(111) Park, B. K.; Kim, M. M. Applications of Chitin and Its Derivatives in Biological Medicine. *Int. J. Mol. Sci.* 2010, *11* (12), 5152–5164.

(112) Di Rosa, M.; Distefano, G.; Zorena, K.; Malaguarnera, L. Chitinases and Immunity: Ancestral Molecules with New Functions. *Immunobiology* 2015, *221* (3), 399–411.

(113) Koch, B. E. V; Stougaard, J.; Spaink, H. P. Keeping Track of the Growing Number of Biological Functions of Chitin and Its Interaction Partners in Biomedical Research. *Glycobiology* 2015, *25* (5), 469–482.

(114) Vinetz, J. M.; Dave, S. K.; Specht, C. A.; Brameld, K. A.; Xu, B.; Hayward, R.; Fidock, D. A. The Chitinase PfCHT1 from the Human Malaria Parasite *Plasmodium Falciparum* Lacks Proenzyme and Chitin-Binding Domains and Displays Unique Substrate Preferences. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96* (24), 14061–14066.

(115) Vinetz, J. M.; Valenzuela, J. G.; Specht, C. A.; Aravind, L.; Langer, R. C.; Ribeiro, J. M. C.; Kaslow, D. C. Chitinases of the Avian Malaria Parasite *Plasmodium Gallinaceum*, a Class of Enzymes Necessary for Parasite Invasion of the Mosquito Midgut. *J. Biol. Chem.* 2000, *275* (14), 10331–10341.

(116) Lockhart, D. E. A.; Schuettelkopf, A.; Blair, D. E.; Van Aalten, D. M. F. Screening-Based Discovery of *Aspergillus Fumigatus* Plant-Type Chitinase Inhibitors. *FEBS Lett.* 2014, *588* (17), 3282–3290.

(117) Vaaje-Kolstad, G.; Horn, S. J.; Sørlie, M.; Eijsink, V. G. H. The Chitinolytic Machinery of *Serratia Marcescens* - A Model System for Enzymatic Degradation of Recalcitrant Polysaccharides. *FEBS J.* 2013, *280* (13), 3028–3049.

(118) Horn, S. J.; Sørlie, M.; Vaaje-Kolstad, G.; Norberg, a. L.; Synstad, B.; Vårum, K. M.; Eijsink, V. G. H. Comparative Studies of Chitinases A, B and C from *Serratia Marcescens*. *Biocatal. Biotransformation* 2006, *24* (1–2), 39–53.

(119) van Aalten, D. M.; Synstad, B.; Brurberg, M. B.; Hough, E.; Riise, B. W.; Eijsink, V. G.; Wierenga, R. K. Structure of a Two-Domain Chitotriosidase from *Serratia Marcescens* at 1.9-Å Resolution. *Proc. Natl. Acad. Sci. U. S. A.* 2000, *97* (11), 5842–5847.

(120) Tsuji, H.; Nishimura, S.; Inui, T.; Kado, Y.; Ishikawa, K.; Nakamura, T.; Uegaki, K. Kinetic and Crystallographic Analyses of the Catalytic Domain of Chitinase from *Pyrococcus Furiosus*- the Role of Conserved Residues in the Active Site. *FEBS J.* 2010, *277* (12), 2683–2695.

(121) Ivo Tews; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. Bacterial Chitobiase Structure Provides Insight into Catalytic Mechanism and the Basis of Tay-Sachs Disease. *Nat. Struct. Mol. Biol.* 1996, *3* (7), 638–648.

(122) Fadel, F.; Zhao, Y.; Cousido-Siah, A.; Ruiz, F. X.; Mitschler, A.; Podjarny, A. X-Ray Crystal Structure of the Full Length Human Chitotriosidase (CHIT1) Reveals Features of Its Chitin Binding Domain. *PLoS One* 2016, *11* (4), e0154190.

(123) Fadel, F.; Zhao, Y. G.; Cachau, R.; Cousido-Siah, A.; Ruiz, F. X.; Harlos, K.; Howard, E.; Mitschler, A.; Podjarny, A. New Insights into the Enzymatic Mechanism of Human Chitotriosidase (CHIT1) Catalytic Domain by Atomic Resolution X-Ray Diffraction and Hybrid QM/MM. *Acta Crystallogr. Sect. D-Biological Crystallogr.* 2015, *71*, 1455–1470.

(124) Greig, I. R.; Zahariev, F.; Withers, S. G. Elucidating the Nature of the *Streptomyces Plicatus* β-Hexosaminidase- Bound Intermediate Using *Ab Initio* Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 2008, *130* (51), 17620–17628.

(125) Malecki, P. H.; Raczynska, J. E.; Vorgias, C. E.; Rypniewski, W. Structure of a Complete Four-Domain Chitinase from *Moritella Marina*, a Marine Psychrophilic Bacterium. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2013, *69* (5), 821–829.

(126) Jitonnom, J.; Lee, V. S.; Nimmanpipug, P.; Rowlands, H. a.; Mulholland, A. J. Quantum Mechanics/Molecular Mechanics Modeling of Substrate-Assisted Catalysis in Family 18 Chitinases: Conformational Changes and the Role of Asp142 in Catalysis in ChiB. *Biochemistry* 2011, *50* (21), 4697–4711.

(127) Iino, T.; Sakurai, M.; Furuta, T. A Novel Ring-Shaped Reaction Pathway with

Interconvertible Intermediates in Chitinase A as Revealed by QM/MM Simulation Combined with a One-Dimensional Projection Technique. *Phys. Chem. Chem. Phys.* 2019, *21* (45), 24956–24966.

(128) Coines, J.; Alfonso-Prieto, M.; Biarnés, X.; Planas, A.; Rovira, C. Oxazoline or Oxazolinium Ion? The Protonation State and Conformation of the Reaction Intermediate of Chitinase Enzymes Revisited. *Chem. – A Eur. J.* 2018, *24* (72), 19258–19265.

(129) Meekrathok, P.; Kukic, P.; Nielsen, J. E.; Suginta, W. Investigation of Ionization Pattern of the Adjacent Acidic Residues in the DXDXE Motif of GH-18 Chitinases Using Theoretical PKaCalculations. *J. Chem. Inf. Model.* 2017, *57* (3), 572–583.

(130) Krokeide, I. M.; Synstad, B.; Göseidnes, S.; Horn, S. J.; Eijsink, V. G. H.; Sörlie, M. Natural Substrate Assay for Chitinases Using High-Performance Liquid Chromatography: A Comparison with Existing Assays. *Anal. Biochem.* 2007, *363* (1), 128–134.

(131) 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, P. A. K. AMBER 11. 2010, University of California, San Francisco.

(132) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Genet.* 2006, *65* (3), 712–725.

(133) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* 2006, *25* (2), 247–260.

(134) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a; Case, D. a. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 2004, *25* (9), 1157–1174.

(135) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had, J. C. and D. J. F. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT. 2013.

(136) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* 1977, *23* (3), 327–341.

(137) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 1996, *14* (1), 33–38.

(138) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* 1996, *77* (18), 3865–3868.

(139) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. II. Operators for Fast Iterative Diagonalization. *Phys. Rev. B* 1991, *43* (11), 8861–8869.

(140) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* 1984, *52* (2), 255–268.

(141) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* 1985, *31* (3), 1695–1697.

(142) Hamprecht, F.; Cohen, A.; Tozer, D. J.; Handy, N. C. Development and Assessment of New Exchange-Correlation Fnctionals. *J. Chem. Phys.* 1998, *109* (15), 6264–6272.

(143) Hardivillé, S.; Hart, G. W. Nutrient Regulation of Signaling, Transcription, and Cell Physiology by *O*-GlcNAcylation. *Cell Metab.* 2014, *20* (2), 208–213.

(144) Yuzwa, S. A.; Shan, X.; MacAuley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J. Increasing *O*-GlcNAc Slows Neurodegeneration and Stabilizes Tau against Aggregation. *Nat. Chem. Biol.* 2012, *8* (4), 393–399.

(145) Ferrer, C. M.; Lynch, T. P.; Sodi, V. L.; Falcone, J. N.; Schwab, L. P.; Peacock, D. L.; Vocadlo, D. J.; Seagroves, T. N.; Reginato, M. J. *O*-GlcNAcylation Regulates Cancer Metabolism and Survival Stress Signaling via Regulation of the HIF-1 Pathway. *Mol. Cell* 2014, *54* (5), 820–831.

(146) Tavassoly, O.; Yue, J.; Vocadlo, D.; Fraser, S. Pharmacological Inhibition and Knockdown of *O*-GlcNAcase Reduces Cellular Internalization of α -Synuclein Pre-Formed-Fibrils. *FEBS J.*

(147) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. Cross Talk Between *O*-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* 2011, *80* (1), 825–858.

(148) King, D. T.; Males, A.; Davies, G. J.; Vocadlo, D. J. Molecular Mechanisms Regulating *O*-Linked *N*-Acetylglucosamine (*O*-GlcNAc)–Processing Enzymes. *Curr. Opin. Chem. Biol.* 2019, *53*, 131–144.

(149) Puchart, V. Glycoside Phosphorylases: Structure, Catalytic Properties and Biotechnological Potential. *Biotechnol. Adv.* 2015, *33* (2), 261–276.

(150) Kitaoka, M. Diversity of Phosphorylases in Glycoside Hydrolase Families. *Appl. Microbiol. Biotechnol.* 2015, *99* (20), 8377–8390.

(151) Macdonald, S. S.; Armstrong, Z.; Morgan-Lang, C.; Osowiecka, M.; Robinson, K.; Hallam, S. J.; Withers, S. G. Development and Application of a High-Throughput Functional Metagenomic Screen for Glycoside Phosphorylases. *Cell Chem. Biol.* 2019, *26* (7), 1001–1012.

(152) Macdonald, S. S.; Blaukopf, M.; Withers, S. G. *N*-Acetylglucosaminidases from CAZy Family GH3 Are Really Glycoside Phosphorylases, Thereby Explaining Their Use of Histidine as an Acid/Base Catalyst in Place of Glutamic Acid. *J. Biol. Chem.* 2015, *290* (8), 4887–4895.

(153) Macdonald, S. S.; Patel, A.; Larmour, V. L. C.; Morgan-Lang, C.; Hallam, S. J.; Mark, B. L.; Withers, S. G. Structural and Mechanistic Analysis of a β-Glycoside Phosphorylase Identified by Screening a Metagenomic Library. *J. Biol. Chem.* 2018, *293* (9), 3451–3467.

(154) Ducatti, D. R. B.; Carroll, M. A.; Jakeman, D. L. On the Phosphorylase Activity of GH3 Enzymes: A β-*N*-Acetylglucosaminidase from *Herbaspirillum Seropedicae* SmR1 and a Glucosidase from *Saccharopolyspora Erythraea*. *Carbohydr. Res.* 2016, *435*, 106–112.

(155) Ostrowski, A.; Gundogdu, M.; Ferenbach, A. T.; Lebedev, A. A.; Van Aalten, D. M. F. Evidence for a Functional *O*-Linked *N*-Acetylglucosamine (O-GlcNAc) System in the Thermophilic Bacterium *Thermobaculum Terrenum*. *J. Biol. Chem.* 2015, *290* (51), 30291–30305.

(156) Ficko-Blean, E.; Gregg, K. J.; Adams, J. J.; Hehemann, J. H.; Czjzek, M.; Smith, S. P.; Boraston, A. B. Portrait of an Enzyme, a Complete Structural Analysis of a Multimodular β-N-Acetylglucosaminidase from *Clostridium Perfringens*. *J. Biol. Chem.* 2009, *284* (15), 9876–9884.

(157) Macauley, M. S.; Whitworth, G. E.; Debowski, A. W.; Chin, D.; Vocadlo, D. J. *O*-GlcNAcase Uses Substrate-Assisted Catalysis: Kinetic Analysis and Development of Highly Selective Mechanism-Inspired Inhibitors. *J. Biol. Chem.* 2005, *280* (27), 25313–25322.

(158) Cekic, N.; Heinonen, J. E.; Stubbs, K. a.; Roth, C.; He, Y.; Bennet, a. J.; McEachern, E. J.; Davies, G. J.; Vocadlo, D. J. Analysis of Transition State Mimicry by Tight Binding Aminothiazoline Inhibitors Provides Insight into Catalysis by Human *O*-GlcNAcase. *Chem. Sci.* 2016, *84*, 25–27.

(159) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* 2012, *8* (11), 4405–4412.

(160) Zhao, G.; Guan, W.; Cai, L.; Wang, P. G. Enzymatic Route to Preparative-Scale Synthesis of UDP-GlcNAc/GalNAc, Their Analogues and GDP-Fucose. *Nat. Protoc.* 2010, *5* (4), 636–646.

(161) D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E.Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. M, X. W. and P. A. K. AMBER 14. *University of California,*. 2014, p San Francisco.

(162) Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. AMBER Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J. Mol. Model.* 2006, *12* (3), 281–289.

(163) Unverzagt, C.; Kajihara, Y. Chemical Assembly of *N*-Glycoproteins: A Refined Toolbox to Address a Ubiquitous Posttranslational Modification. *Chem. Soc. Rev.* 2013, *42* (10), 4408.

(164) Fairbanks, A. J. The ENGases: Versatile Biocatalysts for the Production of Homogeneous *N*-Linked Glycopeptides and Glycoproteins. *Chem. Soc. Rev.* 2017, *46* (16), 5128–5146.

(165) Ito, K.; Okada, Y.; Ishida, K.; Minamiura, N. Human Salivary Endo-β-*N*-Acetylglucosaminidase HS Specific for Complex Type Sugar Chains of Glycoproteins. *J. Biol. Chem.* 1993, *268* (21), 16074–16081.

(166) Li, S.-C.; Asakawa, M.; Hirabayashi, Y.; Li, Y.-T. Isolation of Two Endo-β-*N*-Acetylglucosaminidases from Fig Latex. *Biochim. Biophys. Acta - Gen. Subj.* 1981, *660*, 278–283.

(167) Fujita, K.; Kobayashi, K.; Iwamatsu, A.; Takeuchi, M.; Kumagai, H.; Yamamoto, K. Molecular Cloning of Mucor Hiemalis Endo-β-*N*-Acetylglucosaminidase and Some Properties of the Recombinant Enzyme. *Arch. Biochem. Biophys.* 2004, *432* (1), 41–49.

(168) Barreaud, J. P.; Bourgerie, S.; Julien, R.; Guespin-Michel, J. F.; Karamanos, Y. An Endo-*N*-Acetyl-Beta-D-Glucosaminidase, Acting on the Di-*N*-Acetylchitobiosyl Part of *N*-Linked Glycans, Is Secreted during Sporulation of *Myxococcus Xanthus*. *J. Bacteriol.* 1995, *177* (4), 916–920.

(169) Karamanos, Y.; Bourgerie, S.; Barreaud, J. P.; Julien, R. Are There Biological Functions for Bacterial Endo-*N*-Acetyl-Beta-D-Glucosaminidases? *Res. Microbiol.* 1995, *146* (6), 437–443.

(170) Parsons, T. B.; Patel, M. K.; Boraston, A. B.; Vocadlo, D. J.; Fairbanks, A. J. Streptococcus Pneumoniae Endohexosaminidase D; Feasibility of Using *N*-Glycan Oxazoline Donors for Synthetic Glycosylation of a GlcNAc-Asparagine Acceptor. *Org. Biomol. Chem.* 2010, *8* (8), 1861–1869.

(171) Fairbanks, A. J. Chemoenzymatic Synthesis of Glycoproteins. *Curr. Opin. Chem. Biol.* 2019, *53*, 9–15.

(172) Fujita, M.; Shoda, S. ichiro; Haneda, K.; Inazu, T.; Takegawa, K.; Yamamoto, K. A Novel Disaccharide Substrate Having 1,2-Oxazoline Moiety for Detection of Transglycosylating Activity of Endoglycosidases. *Biochim. Biophys. Acta - Gen. Subj.* 2001, *1528* (1), 9–14.

(173) Heidecke, C. D.; Ling, Z.; Bruce, N. C.; Moir, J. W. B.; Parsons, T. B.; Fairbanks, A. J. Enhanced Glycosylation with Mutants of Endohexosaminidase A (Endo A). *ChemBioChem* 2008, *9* (13), 2045–2051.

(174) Huang, W.; Li, C.; Li, B.; Umekawa, M.; Yamamoto, K.; Zhang, X.; Wang, L. Glycosynthases Enable a Highly Efficient Chemoenzymatic Synthesis of *N*-Glycoproteins Carrying Intact Natural *N*-Glycans. *J. Am. Chem. Soc.* 2009, No. 131, 2214–2223.

(175) Ling, Z.; Suits, M. D. L.; Bingham, R. J.; Bruce, N. C.; Davies, G. J.; Fairbanks, A. J.; Moir, J. W. B.; Taylor, E. J. The X-Ray Crystal Structure of an Arthrobacter Protophormiae Endo-β-*N*-Acetylglucosaminidase Reveals a $(\beta/\alpha)_8$ Catalytic Domain, Two Ancillary Domains and Active Site Residues Key for Transglycosylation Activity. *J. Mol. Biol.* 2009, *389* (1), 1–9.

(176) Fan, S. Q.; Huang, W.; Wang, L. X. Remarkable Transglycosylation Activity of Glycosynthase Mutants of Endo-D, an Endo-β-*N*-Acetylglucosaminidase from *Streptococcus Pneumoniae*. *J. Biol. Chem.* 2012, *287* (14), 11272–11281.

(177) Umekawa, M.; Li, C.; Higashiyama, T.; Huang, W.; Ashida, H.; Yamamoto, K.; Wang, L. X. Efficient Glycosynthase Mutant Derived from *Mucor Hiemalis* Endo-β-*N*-Acetylglucosaminidase Capable of Transferring Oligosaccharide from Both Sugar Oxazoline and Natural *N*-Glycan. *J. Biol. Chem.* 2010, *285* (1), 511–521.

(178) Umekawa, M.; Higashiyama, T.; Koga, Y.; Tanaka, T.; Noguchi, M.; Kobayashi, A.; Shoda, S. ichiro; Huang, W.; Wang, L. X.; Ashida, H.; et al. Efficient Transfer of Sialo-Oligosaccharide onto Proteins by Combined Use of a Glycosynthase-like Mutant of *Mucor Hiemalis* Endoglycosidase and Synthetic Sialo-Complex-Type Sugar Oxazoline. *Biochim. Biophys. Acta - Gen. Subj.* 2010, *1800* (11), 1203–1209.

(179) Dalia, A. B.; Standish, A. J.; Weiser, J. N. Three Surface Exoglycosidases from Streptococcus Pneumoniae , NanA , BgaA , and StrH , Promote Resistance to Opsonophagocytic Killing by Human Neutrophils. 2010, *78* (5), 2108–2116.

(180) Shelburne, S. A.; Davenport, M. T.; Keith, D. B.; Musser, J. M. The Role of Complex Carbohydrate Catabolism in the Pathogenesis of Invasive Streptococci. 2008, No. May.

(181) Koide, N.; Nose, M.; Muramatsu, T. Recognition of IgG by Fc Receptor and Complement: Efects of Glycosidase Digestion. *Biochem. Biophys. Res. Commun.* 1977, *75* (4), 838–844.

(182) Pluvinage, B.; Higgins, M. A.; Abbott, D. W.; Robb, C.; Dalia, A. B.; Deng, L.; Weiser, J. N.; Parsons, T. B.; Fairbanks, A. J.; Vocadlo, D. J.; et al. Inhibition of the Pneumococcal Virulence Factor StrH and Molecular Insights into *N*-Glycan Recognition and Hydrolysis. *Structure* 2011, *19* (11), 1603–1614.

(183) Yin, J.; Li, L.; Shaw, N.; Li., Y.; Song, J. K.; Zhang, W.; Xia, C.; Zhang, R.; Joachimiak, A.; Zhang, H. C.; et al. Structural Basis and Catalytic Mechanism for the Dual Functional Endo-β-*N*-Acetylglucosaminidase A. *PLoS One* 2009, *4* (3).

(184) Teze, D.; Coines, J.; Raich, L.; Kalichuk, V.; Solleux, C.; Tellier, C.; André-Miral, C.; Svensson, B.; Rovira, C. A Single Point Mutation Converts GH84 *O*-GlcNAc Hydrolases into Phosphorylases: Experimental and Theoretical Evidence. *J. Am. Chem. Soc.* 2020, *142* (5), 2120–2124.

(185) Sadeghian, K.; Bocola, M.; Schütz, M. A Conclusive Mechanism of the Photoinduced Reaction Cascade in Blue Light Using Flavin Photoreceptors. *J. Am. Chem. Soc.* 2008, *130* (37), 12501–12513.

(186) Domratcheva, T.; Grigorenko, B. L.; Schlichting, I.; Nemukhin, A. V. Molecular Models Predict Light-Induced Glutamine Tautomerization in BLUF Photoreceptors. *Biophys. J.* 2008, *94* (10), 3872–3879.

(187) Grigorenko, B. L.; Khrenova, M. G.; Nemukhin, A. V. Amide-Imide Tautomerization in the Glutamine Side Chain in Enzymatic and Photochemical Reactions in Proteins. *Phys. Chem. Chem. Phys.* 2018, *20* (37), 23827–23836.

(188) Liu, H.; Wang, C.; Lee, S.; Ning, F.; Wang, Y.; Zhang, Q.; Chen, Z.; Zang, J.; Nix, J.; Dai, S.; et al. Specific Recognition of Arginine Methylated Histone Tails by JMJD5 and JMJD7. *Sci. Rep.* 2018, *8* (1), 1–11.

(189) Chrysina, M.; De Mendonça Silva, J. C.; Zahariou, G.; Pantazis, D. A.; Ioannidis, N. Proton Translocation via Tautomerization of Asn298 during the S_2-S_3 State Transition in the Oxygen-Evolving Complex of Photosystem II. *J. Phys. Chem. B* 2019, *123* (14), 3068–3078.

(190) Goings, J. J.; Hammes-Schiffer, S. Early Photocycle of Slr1694 Blue-Light Using Flavin Photoreceptor Unraveled through Adiabatic Excited-State Quantum Mechanical/Molecular Mechanical Dynamics. *J. Am. Chem. Soc.* 2019, *141* (51), 20470–20479.

(191) Grigorenko, B. L.; Kots, E. D.; Nemukhin, A. V. Diversity of Mechanisms in Ras-GAP Catalysis of Guanosine Triphosphate Hydrolysis Revealed by Molecular Modeling. *Org. Biomol. Chem.* 2019, *17* (19), 4879–4891.

(192) Nakamura, A.; Ishida, T.; Kusaka, K.; Yamada, T.; Fushinobu, S.; Tanaka, I.; Kaneko, S.; Ohta, K.; Tanaka, H.; Inaka, K.; et al. "Newton's Cradle" Proton Relay with Amide-Imidic Acid Tautomerization in Inverting Cellulase Visualized by Neutron Crystallography. *Sci. Adv.* 2015, *1* (7), e1500263.

(193) Muramatsu, T. Endo-β-N-Acetylglucosaminidase D from *Diplococcus Pneurnoniae*. *Methods Enzymol.* 1978, *50*, 555–559.

(194) Tai, T.; Yamashita, K.; Ogata-Arakawa, M.; Koide, N.; Muramatsu, T.; Iwashita, S.; Inoue, Y.; Kobata, A. Structural Studies of Two Ovalbumin Glycopeptides in Relation to the Endo-β-*N*-Acetylglucosaminidase Specificity. *J. Biol. Chem.* 1975, *250* (21), 8569–8575.

(195) Kitajima, T.; Jia, Y.; Komatsuzaki, A.; Cui, J.; Matsuzawa, F.; Aikawa, S. I.; Gao, X. D.; Chiba, Y. Structural Modeling and Mutagenesis of Endo-β-N-Acetylglucosaminidase from *Ogataea Minuta* Identifies the Importance of Trp295 for Hydrolytic Activity. *J. Biosci. Bioeng.* 2018, *125* (2), 168–174.

(196) Fujita, K.; Takegawa, K. Tryptophan-216 Is Essential for the Transglycosylation Activity of Endo-β-*N*-Acetylglucosaminidase A. *Biochem. Biophys. Res. Commun.* 2001, *283* (3), 680–686.

(197) Trastoy, B.; Du, J. J.; Klontz, E. H.; Li, C.; Cifuente, J. O.; Wang, L. X.; Sundberg, E. J.; Guerin, M. E. Structural Basis of Mammalian High-Mannose *N*-Glycan Processing by Human Gut Bacteroides. *Nat. Commun.* 2020, *11* (1), 1–11.

(198) Trastoy, B.; Klontz, E.; Orwenyo, J.; Marina, A.; Wang, L. X.; Sundberg, E. J.; Guerin, M. E. Structural Basis for the Recognition of Complex-Type *N*-Glycans by Endoglycosidase S. *Nat. Commun.* 2018, *9* (1), 1–11.

(199) Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *Ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* 2015, *137* (24), 7528–7547.

(200) Constantino, E.; Solans-Monfort, X.; Sodupe, M.; Bertran, J. Basic and Acidic Bifunctional Catalysis: Application to the Tautomeric Equilibrium of Formamide. *Chem. Phys.* 2003, *295* (2), 151–158.

(201) Eriksson, M. A. L.; Hard, T.; Nilsson, L. On the PH Dependence of Amide Proton Exchange Rates in Proteins. *Biophys. J.* 1995, *69*, 329–339.

(202) Abdul Manas, N. H.; Md. Illias, R.; Mahadi, N. M. Strategy in Manipulating Transglycosylation Activity of Glycosyl Hydrolase for Oligosaccharide Production. *Crit. Rev. Biotechnol.* 2018, *38* (2), 272–293.

(203) Bissaro, B.; Monsan, P.; Fauré, R.; O'Donohue, M. J. Glycosynthesis in a Waterworld: New Insight into the Molecular Basis of Transglycosylation in Retaining Glycoside Hydrolases. *Biochem. J.* 2015, *467* (1), 17–35.

(204) Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. Catalytic Itinerary in 1,3-1,4-β-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not yet Fully Developed at the Oxocarbenium Ion-like Transition State. *J. Am. Chem. Soc.* 2011, *133* (50), 20301–20309.

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

(206) Tankrathok, A.; Iglesias-Fernández, J.; Williams, R. J.; Pengthaisong, S.; Baiya, S.; Hakki, Z.; Robinson, R. C.; Hrmova, M.; Rovira, C.; Williams, S. J.; et al. A Single Glycosidase Harnesses Different Pyranoside Ring Transition State Conformations for Hydrolysis of Mannosides and Glucosides. *ACS Catal.* 2015, *5* (10), 6041–6051.

(207) Williams, R. J.; Iglesias-Fernández, J.; Stepper, J.; Jackson, A.; Thompson, A. J.; Lowe, E. C.; White, J. M.; Gilbert, H. J.; Rovira, C.; Davies, G. J.; et al. Combined Inhibitor Free-Energy Landscape and Structural Analysis Reports on the Mannosidase Conformational Coordinate. *Angew. Chemie Int. Ed.* 2014, *53* (4), 1087–1091.

(208) Belz, T.; Jin, Y.; Coines, J.; Rovira, C.; Davies, G. J. An Atypical Interaction Explains the High-Affinity of a Non-Hydrolyzable S-Linked 1,6-α-Mannanase Inhibitor. *Chem. Commun.* 2017, No. 53, 9238–9241.

(209) Males, A.; Raich, L.; Williams, S. J.; Rovira, C.; Davies, G. J. Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Structural Approach. *ChemBioChem* 2017, *18* (15), 1496–1501.

(210) Beenakker, T. J. M.; Wander, D. P. A.; Offen, W. A.; Artola, M.; Raich, L.; Ferraz, M. J.; Li, K. Y.; Houben, J. H. P. M.; Van Rijssel, E. R.; Hansen, T.; et al. Carba-Cyclophellitols Are Neutral Retaining-Glucosidase Inhibitors. *J. Am. Chem. Soc.* 2017, *139* (19), 6534–6537.

(211) Sega, M.; Autieri, E.; Pederiva, F. On the Calculation of Puckering Free Energy Surfaces. *J. Chem. Phys.* 2009, *130* (22).

(212) Spiwok, V.; Králová, B.; Tvaroska, I. Modelling of β-D-Glucopyranose Ring Distortion in Different Force Fields: A Metadynamics Study. *Carbohydr. Res.* 2010, *345* (4), 530–537.

(213) Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Fivebiologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* 2014, *136* (3), 1008–1022.

(214) Sattelle, B. M.; Almond, A. Is *N*-Acetyl-D-Glucosamine a Rigid $^4C_1$ Chair? *Glycobiology* 2011, *21* (12), 1651–1662.

(215) Whitworth, G. E.; Macauley, M. S.; Stubbs, K. A.; Dennis, R. J.; Taylor, E. J.; Davies, G. J.; Greig, I. R.; Vocadlo, D. J. Analysis of PUGNAc and NAG-Thiazoline as Transition State Analogues for Human *O*-GlcNAcase: Mechanistic and Structural Insights into Inhibitor Selectivity and Transition State Poise. *J. Am. Chem. Soc.* 2007, *129* (3), 635–644.

(216) Böshagen, H.; Heiker, F.-R.; Schüller, A. M. The Chemistry of the 1-Deoxynojirimycin System. Synthesis of 2-Acetamido-1,2-Dideoxynojirimycin from 1-Deoxynojirimycin. *Carbohydr. Res.* 1987, *164*, 141–148.

(217) Beer, D.; Maloisel, J.; Rast, D. M.; Vasella, A. Synthesis of 2-Acetamido-2-Deoxy-D-Gluconhydroximolactone- and Chitobionhydroximolactone-Derived *N*-Phenylcarbamates, Potential Inhibitors of β-*N*-Acetylglucosaminidase. *Helv. Chim. Acta* 1990, *73*, 1918–1922.

(218) Macdonald, J. M.; Tarling, C. a; Taylor, E. J.; Dennis, R. J.; Myers, D. S.; Knapp, S.; Davies, G. J.; Withers, S. G. Chitinase Inhibition by Chitobiose and Chitotriose Thiazolines. *Angew. Chem. Int. Ed. Engl.* 2010, *49* (14), 2599–2602.

(219) Stubbs, K. A.; Balcewich, M.; Mark, B. L.; Vocadlo, D. J. Small Molecule Inhibitors of a Glycoside Hydrolase Attenuate Inducible AmpC-Mediated β-Lactam Resistance. *J. Biol. Chem.* 2007, *282* (29), 21382–21391.

(220) Liu, T.; Zhang, H.; Liu, F.; Chen, L.; Shen, X.; Yang, Q. Active-Pocket Size Differentiating Insectile from Bacterial Chitinolytic β-*N*-Acetyl-D-Hexosaminidases. *Biochem. J.* 2011, *438* (3), 467–474.

(221) Dong, D. L. Y.; Hart, G. W. Purification and Characterization of an *O*-GlcNAc Selective *N*-Acetyl-β-D- Glucosaminidase from Rat Spleen Cytosol. *J. Biol. Chem.* 1994, *269* (30), 19321–19330.

(222) Acebrón, I.; Mahasenan, K. V.; De Benedetti, S.; Lee, M.; Artola-Recolons, C.; Hesek, D.; Wang, H.; Hermoso, J. A.; Mobashery, S. Catalytic Cycle of the *N*-Acetylglucosaminidase NagZ from Pseudomonas Aeruginosa. *J. Am. Chem. Soc.* 2017, *139* (20), 6795–6798.

(223) Ho, C. W.; Popat, S. D.; Liu, T. W.; Tsai, K. C.; Ho, M. J.; Chen, W. H.; Yang, A. S.; Lin, C. H. Development of GlcNAc-Inspired Iminocyclitiols as Potent and Selective N -Acetyl-β-Hexosaminidase Inhibitors. *ACS Chem. Biol.* 2010, *5* (5), 489–497.

(224) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* 2012, *8* (7), 2247–2254.