Theoretical Study of the Catalytic Mechanism of Retaining Glycosyltransferases

Hansel Gómez Martínez

Ph.D. Thesis

Ph.D. program in Chemistry

Supervisors:

Laura Masgrau i Fontanet

Josep M. Lluch i López

Departament de Química

Facultat de Ciències

2013

UNIVERSITAT AUTÒNOMA DE BARCELONA
DEPARTAMENT DE QUÍMICA
UNITAT DE QUÍMICA FÍSICA

LAURA MASGRAU i FONTANET, investigadora Ramón y Cajal a l'Institut de Biotecnologia i de Biomedicina i JOSEP M. LLUCH i LÓPEZ, Professor Catedràtic del Departament de Química i membre de l'Institut de Biotecnologia i de Biomedicina de de la Universitat Autònoma de Barcelona, certifiquem que

Hansel Gómez Martínez, llicenciat en Bioquímica per la Universidad de La Habana (Cuba) ha realizat sota la nostra direcció, en l'Institut de Biotecnologia i de Biomedicina i el Departament de Química de la Universitat Autònoma de Barcelona, el treball d'investigació titulat:

Theoretical Study of the Catalytic Mechanism of Retaining Glycosyltransferases

que es presenta en aquesta memòria per optar al grau de Doctor en Química (amb menció de doctorat europeu).

I perquè consti als efectes escaients, signem aquest certificat a Bellaterra al 17 de juliol de 2013:

Dra. Laura Masgrau i Fontanet

Dr. Josep M. Lluch i Lóp

PREFACE

The present thesis is focused on the catalytic mechanism of retaining glycosyltransferases from the computational point of view. Different proposed mechanisms for these enzymes are studied by using hybrid quantum mechanics/ molecular mechanics (QM/MM) calculations. Moreover, the factors responsible for the catalytic efficiency of Lipopolysaccharyl-α-1,4-galactosyltransferase C from *Neisseria meningitides* (LgtC), bovine α-1,3-galactosyltransferase (α1,3-GalT) and human UDP-N-acetylgalactosamine:polypeptide N-acetyl-α-galactosaminyltransferase 2 (ppGalNAcT-2) are discussed.

The results for LgtC, α1,3-GalT and ppGalNAcT-2 are exposed in Chapters 4, 5, and 7, respectively. For convenience, a comparative study between LgtC and α1,3-GalT is presented in Chapter 6. A general discussion providing a global overview of the catalytic mechanism of retaining glycosyltransferases is presented in Chapter 8 while the general conclusions of this work are outlined in Chapter 9.

For clarity and to make the reading easier, some tables and figures are included in the Appendix section, even though every relevant data are mentioned in the main text.

This doctoral thesis was performed thanks to the financial support from the Spanish "Ministerio de Ciencia e Innovación / Ministerio de Economía y Competitividad" through the pre-doctoral fellowship BES-2009-019796 and projects CTQ2008-02403 and CTQ2011-24292; as well as from the "Generalitat de Catalunya", project 2009SGR409.

Most of the results presented in this document have already been published in the following papers:

- Gómez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L. *J. Am. Chem. Soc.* **2012**, *134*, 4743-4752.
- Gómez, H.; Lluch, J. M.; Masgrau, L. *Carbohydr. Res.* **2012**, *356*, 204-208.
- Gómez, H.; L. luch, J. M.; Masgrau, L. *J. Am. Chem. Soc.* **2013**, *135*, 7053-7063.

CONTENTS

Page

I GENERAL INTRODUCTION

1. General Introduction

II OBJETIVES

III THEORETICAL FRAMEWORK

3. Theoretical Framework

IV RESULTS AND DISCUSSION

4. Lipopolysaccharyl-α-1,4-galactosyltransferase C (LgtC)

5. α-1,3-galactosyltransferase (α1,3-GalT)

6. Further comparison of α1,3-GalT and LgtC catalytic mechanisms

V GENERAL CONCLUSIONS

PART I

General Introduction

1

GENERAL INTRODUCTION

"Life … is a relationship between molecules"
Linus Pauling

1.1.ENZIMATIC REACTIONS

Enzymes are the catalysts in most biological processes and are therefore the main responsible for the chemical interconversions that sustain life.[1] Not surprisingly, alteration of the enzymatic activity is commonly related to pathological conditions. Moreover, they are highly efficient when compared with chemical catalysts. Putting all together it seems logic why there is such a major practical and fundamental interest in finding out what makes enzymes so efficient and to characterize the detailed mechanism that controls each reaction step and provides the desired regio- and stereospecificity.

L. Pauling stated long ago the fact that enzymes reduce the activation barrier of the catalyzed reaction.[4] Later on, R. Wolfenden illustrated the catalytic power of enzymes by comparing the rate constant of one catalyzed reaction with that of the same reaction in aqueous solution and in the absence of the enzyme (Figure 1.1).[6,7]

After binding of the substrates takes place, one or more mechanisms of catalysis lower the free energy of the reaction's transition state. In that sense, studies of enzymatic reactions show that natural selection has developed different mechanisms, and enzymes may operate through electrostatic stabilization, bond strain, proximity and orientation, active-site proton donors or acceptors, covalent catalysis and quantum tunneling.[9]

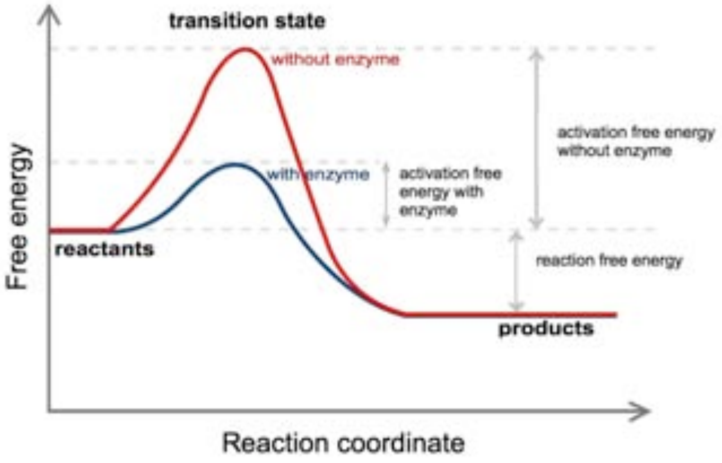

Figure 1.1. Schematic representation of Gibbs free energy as a function of the reaction coordinate in the same reaction catalyzed or not by an enzyme.

To understand enzyme catalysis and mechanism it is necessary, and often challenging, to elucidate the unique ways in which each enzyme exerts electrostatic and other forces on the substrate and the transition state. Each enzyme has its unique characteristics, and enzymes use all possible means to achieve the ultimate objective of reducing the free energy of activation. A proper description of the enzyme active site and catalytic mechanism is always desirable since understanding the changes in electronic structure along the reaction path is needed to design inhibitors and novel catalysts.

Enzyme kinetics itself cannot prove which modes of catalysis an enzyme uses although some kinetic data can suggest possibilities to be examined by other techniques. On the other hand, in the past 10 years computational enzymology has become (coupled with advances in protein structure determination, site-directed mutagenesis, and fast computers and algorithms) an effective way of determining the catalytic characteristics of enzymes. Therefore, theoretical chemistry as a more general approach, has become a widely used tool that has changed our vision of enzymes and their powerful machinery while also providing a whole new framework to the rational design of inhibitors and enzyme engineering.

1.2.GLYCOSYLTRANSFERASES OVERVIEW

1.2.1. Function and activity. Gycosyltransferases (GTs) (EC 2.4.x.y) altogether with glycoside hydrolases (GHs), glycan phosphorylases and polysaccharide lyases are responsible for the remarkable complexity and diversity of the oligosaccharides, polysaccharides and glycoconjugates found in nature.[10] Such complexity would not be possible without a proper interrelationship between the glycosidic bond breakdown and formation processes.

Much of the glycoproteins and glycolipids that are processed by these enzymes are distributed on cell surfaces and within the extracellular matrices and play key roles in cell functions including cell growth and differentiation, recognition by the immune system, as well as cell-cell interactions.[11-13] That is why, and not surprisingly, changes in the composition of these glycoconjugates are often associated with pathological states, including the metastasis of tumoral cells and autoimmune responses.[14-17] Moreover, some of these glycoconjugates modulate interactions with viral and bacterial pathogens leading to infection and are involved in mechanisms to evade the host immune responses.[18-20] Therefore, GTs not only represent an attractive class of therapeutic targets but also are considered important tools for the enzymatic synthesis of synthetically challenging therapeutic agents (e.g. antibiotics). A detailed understanding of the mechanisms by which this class of enzyme catalyzes glycosyl group transfer is therefore of central importance to rationally inhibit or manipulate their enzymatic activity as convenient.

GTs are specifically defined as enzymes that catalyze the biosynthesis of glycosidic linkages by transferring monosaccharides from a donor substrate to an acceptor molecule. It is estimated that from the gene products of an organism, whether archaeal, bacterial, or eukaryotic, about 1 % to 2% correspond to GTs.[10] Logically, those organisms with large genomes like plants that synthesize a complex cell wall or use the glycosylation of small molecules to tune bioactivity have many GTs (e.g., *Arabidopsis* encodes approximately 450 GTs and *Populus* more than 800). The number of genes expressing GTs is much smaller in mammals (e.g., humans have ~230). On the other hand, organisms that have undergone massive gene loss during evolution to become obligate symbionts or obligate parasites appear to have very few or no genes at all coding for GTs (e.g., several *Mycoplasma* species).

In eukaryotes, most of the glycosylation reactions occur in the Golgi apparatus.[21] These GTs are type-II transmembrane proteins with a large C-terminal globular catalytic domain facing the luminal side.

Regarding the donor substrates, GTs commonly use molecules activated in the form of nucleotide diphospho-sugars (e.g., UDP-Gal, UDP-Glc, GDP-Man).[10,21] These GTs are also known as Leloir enzymes in honor of Luis F. Leloir, who discovered the first sugar nucleotide and was awarded the Nobel Prize in Chemistry in 1970 for his contributions to the understanding of glycoside biosynthesis and sugar metabolism. Some GTs can also use nucleotide monophospho-sugars as well as sugar phosphates. On the other hand, all GTs using non-nucleotide donors are termed non-Leloir GTs. This group includes the GTs using lipid-linked glycosyl donors wherein the lipid is frequently a terpenoid such as dolichol or polyprenol (e.g, polyprenol pyrophosphates, polyprenol phosphates).[10] Sugar-1-phosphates or sugar-1-pyrophosphates are also found as donor substrates in GTs better known as phosphorylases and pyrophosphorylases respectively.

As for the acceptor substrate, GTs are able to glycosylate a broad set of molecules. These include sugars, lipids, peptides or proteins itself, nucleic acids and antibiotics.[10] Moreover, the glycosyl transfer most frequently occurs to the nucleophilic oxygen of a hydroxyl group of the acceptor substrate, but it can also occur to a nitrogen (e.g., N-linked glycoproteins), sulfur (e.g., thioglycosides in plants), or even carbon atoms (e.g., C-glycoside antibiotics).

It should by pointed out that the functional characterization of GTs remains a great challenge and that over all the open reading frames known that encode for this type of enzymes yet the donor and acceptor specificity for the vast majority (>95%) is unknown.[10]

1.2.2. Sequence similarities and folding. On a sequence base and according to the Carbohydrate-Active enZyme database (CAZy),[22] GTs using nucleotide diphospho-sugar, nucleotide monophospho-sugars and sugar phosphates (EC 2.4.1.x) have been classified into families as first described by Campbell et al.[23] As of January 2013, CAZy database (**http://www.cazy.org**) collected 107874 sequences of GTs organized in 94 families (GT1 to GT94), as well as 2368 sequences that were not classified into a particular family.

In striking contrast to GHs, which exhibit a wide variety of folds, GTs exhibit a much narrower subset. More specifically, only two general folds, called GT-A and GT-B as proposed by Bourne and Henrissat,[24] have been identified for all the structures of nucleotide-sugar-dependent GTs solved to date.[25-27] The foregoing suggests a convergence process of catalytic mechanisms during the evolution in GHs, while GTs would have evolved from a small number of progenitor sequences. More specifically, most GTs might be evolved from primitive domain Archae.[25]

The GT-A fold is typified by the first member to have its X-ray structure determined, the SpsA from *Bacillus subtilis*, for which both the apo and the UDP-bound 3-D X-ray crystal structures were obtained.[2] This fold consists of two different domains, involved in the nucleotide and acceptor binding respectively.[26] However these are tightly associated β/α/β domains that lead to the formation of a continuous central β-sheet giving the appearance of a single domain (Figure 1.2A). This is a seven-stranded β-sheet with a topology in which strand 6 is antiparallel to the rest.[21] Moreover, the structural architecture of the GT-A fold can be considered as a reminiscent of two abutting Rossmann-like folds, typical of nucleotide-binding proteins.[10]

There was probably a divergence process during the evolution of the catalytic mechanisms, since not all the enzymes exhibiting this fold are GTs, that is, the GT-A fold is not exclusively related to glycosyl transfer reactions.[28] This is the case of the sugar-1-phosphate pyrophosphorylase/nucleotidyl transferase superfamily of enzymes, responsible for the synthesis of nucleoside diphosphate sugars.

GT-A enzymes commonly possess an Asp-X-Asp (referred to as DXD) signature in which the carboxylates of the aspartates coordinate a divalent cation and may also be involved in the binding of the donor substrate.[29,30] GT-A GTs are commonly dependent on this divalent ion for their activity.[30,31] This DXD motif is localized at the end of the Rossmann-type nucleotide-binding domain, encompassing the first 100–120 residues.[21] Other key amino acids that interact with UDP are mainly found at the C-term of strands β1 and β4, although in some crystal structures residues in the C-terminus of the catalytic domain were shown to interact with the nucleotide. It is to be cautioned that no part of the DXD motif is invariant among GTs. On the other hand, there are enzymes from this folding family that do not possess the DXD signature and, thus, it can not be considered a reliable distinction.[32] Finally, more than 50% of all protein sequences possess a DXD signature but, definitely, they are not all GTs, as well as many glycosyltransferases are metal-ion independent and do not require a metal-binding motif.

In addition to the carboxylate side chains of a DXD motif in the active site, the divalent cation is also coordinated by the diphosphate from the donor substrate leaving group. Thus, the role of the divalent metal ion cofactors (tipically Mn^{2+} or Mg^{2+}) in the metal-ion dependent GTs is commonly assumed to be as a Lewis acid to facilitate leaving group departure by electrostatically stabilizing the developing negative charge associated with the bond breakage process (See Figures 1.5, 1.6).[10]

In the case of the structural region involved in the recognition of the acceptor substrate in GTs with the GT-A fold this is much more variable and comprises the C-terminal portion of the enzyme.[21] Thus, the small variety of folds observed in GTs is compensated by a large structural variability in the acceptor-binding domain that confers some functional plasticity to these enzymes.

The case of the GT-B fold is exemplified by its first member: the DNA-modifying β-glycosyltransferase from bacteriophage T4.[33] This folding can also be described as two similar β/α/β Rossmann-like domains, even though GT-A and GT-B enzymes appear to be unrelated.[21] The GT-B fold clearly differentiates from the GT-A one since in the former the two domains are less tightly associated and face each other with the active-site lying within the resulting cleft (Figure 1.2B).[10] Once again, these two domains can be separately associated with the donor and acceptor substrate binding sites, respectively.

There are several examples of proteins adopting the GT-B fold without being a GT (e.g., UDP GlcNAc 2-epimerase[34]) and, thus, this structural pattern is also not exclusive to this kind of enzymes. The GT-B GTs do not exhibit the DXD signature and are commonly assumed to follow a metal ion-independent mechanism to facilitate the

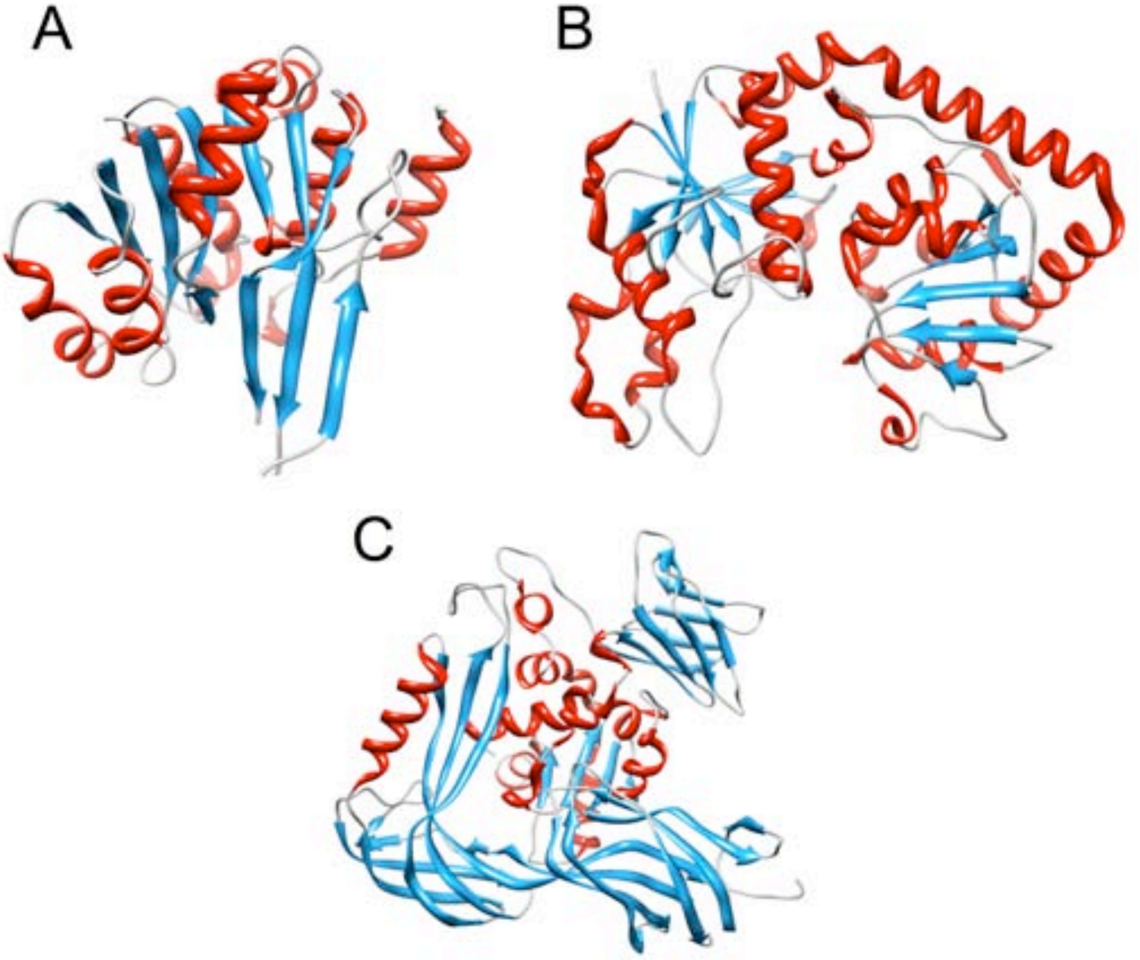

Figure 1.2. Overall folds observed for glycosyltransferase enzymes. (***A***) The GT-A fold is represented by the inverting enzyme SpsA from *Bacillus subtilus* (PDB Code 1QGQ[2]), (***B***) the GT-B fold, by bacteriophage T4 β-glucosyltransferase (PDB Code 1JG7[5]) and (***C***) first structure of an enzyme assigned to the GT-C fold, corresponding to the soluble C-terminal domain of the *Pyrococcus furiosius* oligosaccharyltransferase STT3 (PDB Code 2ZAI[8]).

leaving group departure, although some of them may require divalent cations for full activity.[21]

The ratio of loops to secondary structural elements is high in GTs.[21] Consequently, many crystal structures do not describe the entire catalytic domain because the polypeptide extremities and/or several loops are too flexible and do not present clear electron density. The latest constitutes an additional problem while performing theoretical studies of this type of enzymes.

More recently, Liu & Mushegian[35] suggested a third structural pattern on the basis of an iterative primary sequence analysis. From this initial study several GTs were predicted to adopt what is known as the GT-C fold. With the exception of the members of families GT48 and GT53 that use UDP-Glc and UDP-L-arabinose, respectively, the rest of GTs predicted to have the GT-C fold use lipid phosphate-activated donor sugar substrates. All these enzymes are large hydrophobic integral membrane protein located in the endoplasmic reticulum or on the plasma membrane, having between 8 and 13 transmembrane helices and an active site located on a long-loop region.[36-38] The first of the 3-D structures from a GT expected to adopt the GT-C fold was obtained by Igura et al. in 2008[8] and corresponds to the soluble C-terminal domain of the *Pyrococcus furiosius* oligosaccharyltransferase STT3 (Figure 1.2C). Curiously, it adopts a novel architecture with a central, mainly α-helical domain surrounded by three β-sheet-rich domains. It was then noticed that the previous predictions of the GT-C fold were mostly based on the trans-membrane region, which does not include the loop bearing the active site of these GTs. Consequently, the major structural similarity of these enzymes could be in their trans-membrane region rather than in their catalytic domain. It is actually possible to find different patterns for this part of the protein since these non-Leloir GTs do not require Rossmann-like domains to achieve their function.[10] In other words, it is not clear whether the predicted GT-C fold have any predictive relevance beyond indicating the presence of a large trans-membrane component.

Finally, other sequence families are currently *"orphan"* families that are predicted to adopt neither the GT-A, nor the GT-B nor the proposed GT-C fold and, therefore, more experimental data is needed in order to assign them a folding pattern.

1.2.3. Glycosyltransferases reaction mechanism. Starting with the formation of the ternary complex (GT + donor substrate + acceptor substrate), many, probably all, GTs display conformational changes upon ligands binding. These conformational changes have proven to have direct and important implications for GTs function.[21,39,40] However, for many GTs the nature and extent of these changes have not been defined.

Kinetic and structural data support that most GTs follow an ordered bi bi reaction. The sugar donor binds first and results in an induced conformational change prior to the binding of the acceptor molecule.[41-43] More specifically, a disordered loop (or C-term extremity) in the apoenzyme becomes ordered upon nucleotide sugar binding and creates a lid over the donor substrate where additional residues establish direct contacts with the diphosphate moiety.[21] This new conformation, frequently known as the *'closed'* active conformation, creates a pocket that will serve as binding site for the acceptor molecule. Affinity studies performed on LgtC (GT-A fold) with the use of titration microcalorimetry confirmed that the *"open"* state (free enzyme) has no or little affinity for the oligosaccharidic acceptor.[42] These loops have been proven to play a crucial role during substrate binding and catalysis as certified in several other studies with GT-A GTs.[42,44-50] Moreover, these conformational changes in GTs have also been the subject of pioneering Molecular Dynamic (MD) simulation studies that demonstrate the correlated motions of several loops as well as the importance of contacts between loops in the catalytic mechanism.[51-53]

This mechanism involving protein rearrangement upon ligand binding has historically been suggested to restrict water access to the active site and, thus, to limit the hydrolysis of the energetically precious nucleotide sugar. Moreover, another significant feature is the observed distorted conformation of the bound nucleotide sugar in the active site, which was suggested to be important in the catalytic mechanism.[54] In GTs, the nucleotide sugar is forced to adopt a folded shape that brings the sugar over the pyrophosphate. This special conformation facilitates the sugar transfer by several means: the anomeric carbon atom is more spatially accessible for the reaction, the anomeric bond is elongated and weakened as calculated by *ab initio* methods,[55] and a hydrogen bond can be established between the O2' atom of the sugar ring and the phosphate leaving group, lowering the energy barrier. This phenomenon has also been certified in the case of a glycosidase in a theoretical study by Ardèvol et al.[56] where they found that a *"preactivated"* conformation of the sugar ring in the Michaelis complex strongly determined the catalytic itinerary followed by the enzyme.

Once the sugar transfer has taken place the glycosylated acceptor first and then the sugar-activating group (e.g., nucleotide) are released. In the case of the divalent cation depending enzymes, it has been suggested that the cation can bind to the apoenzyme and that it does not dissociate after each catalytic cycle.[41-43,45,57-59]

GTs can catalyze the glycosyl transfer in two well-defined ways according to the stereochemistry of the substrates and reaction products, being classified as either retaining (ret-GTs) or inverting (inv-GTs) GTs.[60] Stated differently, the reaction takes place with net retention or inversion of the configuration at the anomeric center, which in case of the Leloir enzymes is the carbon atom linked to the nucleotide leaving group (Figure 1.3).

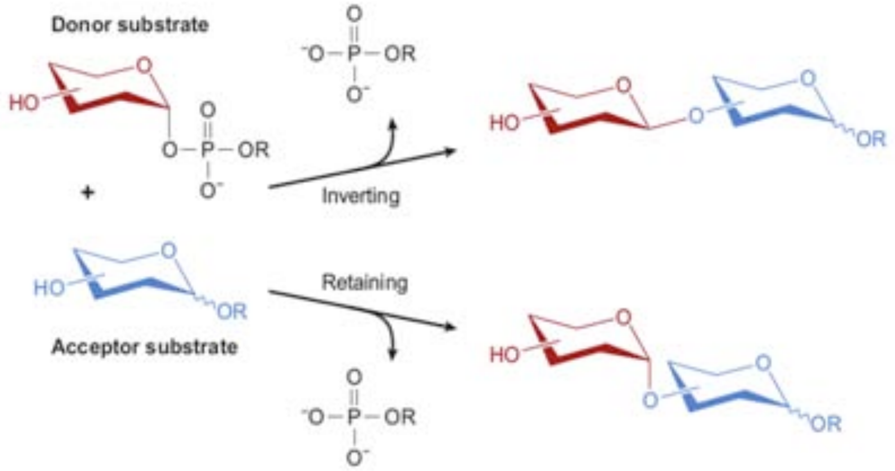

Figure 1.3. Classification of glycosyltransferases as inverting (inv-GT) or retaining (ret-GT) depending on the reaction stereochemical outcome at the anomeric center. Figure 2 in Ref. [10].

Even if it cannot always be predicted with reliability from sequence comparison alone,[21] the catalytic mechanism (i.e. retaining or inverting) is conserved within a GT family. However, the overall fold of the enzyme does not dictate the stereochemical outcome of the reaction that it catalyzes, since there are examples of both ret-GT and inv-GT that exhibit both the GT-A or GT-B fold.[25] The foregoing indicates that common structural elements are necessary for the glycosyl transfer reaction, irrespective of the stereochemistry of the reaction.[21] The exception would be the enzymes adopting the GT-C fold, as they have all been described as inv-GTs.

Differences are observed in the function of the residues of the DXD motif in retaining and inverting GTs with the GT-A fold. In ret-GTs, the two aspartate residues can interact with the divalent ion, whereas in inverting enzymes only the last aspartate coordinates the metal cation.[61,62] However, in both cases, the variable amino acid (i.e.

X) of the DXD motif (usually a polar or an aliphatic residue of moderate size) interacts with the ribose ring of the nucleotide leaving group.[21]

Finally, GTs families can be also classified into clans depending on their fold and on mechanistic features, as proposed by Coutinho and coworkers[25] which is assumed to be the most inclusive classification criteria (Figure 1.4). Notice also from Figure 1.4 that there are some GTs families for which no folding pattern/reaction mechanism can be assigned at present. Finally, it has to be mentioned that the classification of GTs is under continuous revision since it is commonly based upon thermodynamics, often upon sequence, and occasionally upon structural homologies, or even influenced by historical perspective.

1.2.3.1. Catalytic mechanism of inverting glycosyltransferases (inv-GTs). Structural and kinetic data for inv-GTs strongly supports a catalytic mechanism that proceeds through a single nucleophilic substitution step facilitated by an amino acid residue in the active site of the enzyme (Figure 1.5).[2,32,63-70] This residue is commonly an aspartate or glutamate which side chain serves as a base catalyst that deprotonates the incoming nucleophile of the acceptor molecule, facilitating a direct displacement of the activated (substituted) phosphate leaving group. This S_N2-like reaction is the same mechanistic strategy described for inverting glycoside hydrolases (inv-GHs) and involves the formation of a transition state (TS) with a substantial oxocarbenium ion character.

1.2.3.2. Catalytic mechanism of retaining glycosyltransferases (ret-GTs). It is logical to believe that a different reaction outcome, that is, inversion or retention of stereochemistry at the anomeric center must result from the utilization of different mechanisms by inv-GTs and ret-GTs.

It was initially proposed that the catalysis by ret-GTs proceeded through a double-displacement mechanism.[71] The foregoing implies the formation and subsequent cleavage of a covalent enzyme-glycosyl intermediate (CGE) and the presence of a strong and appropriately positioned nucleophile within the active site (Figure 1.6A).

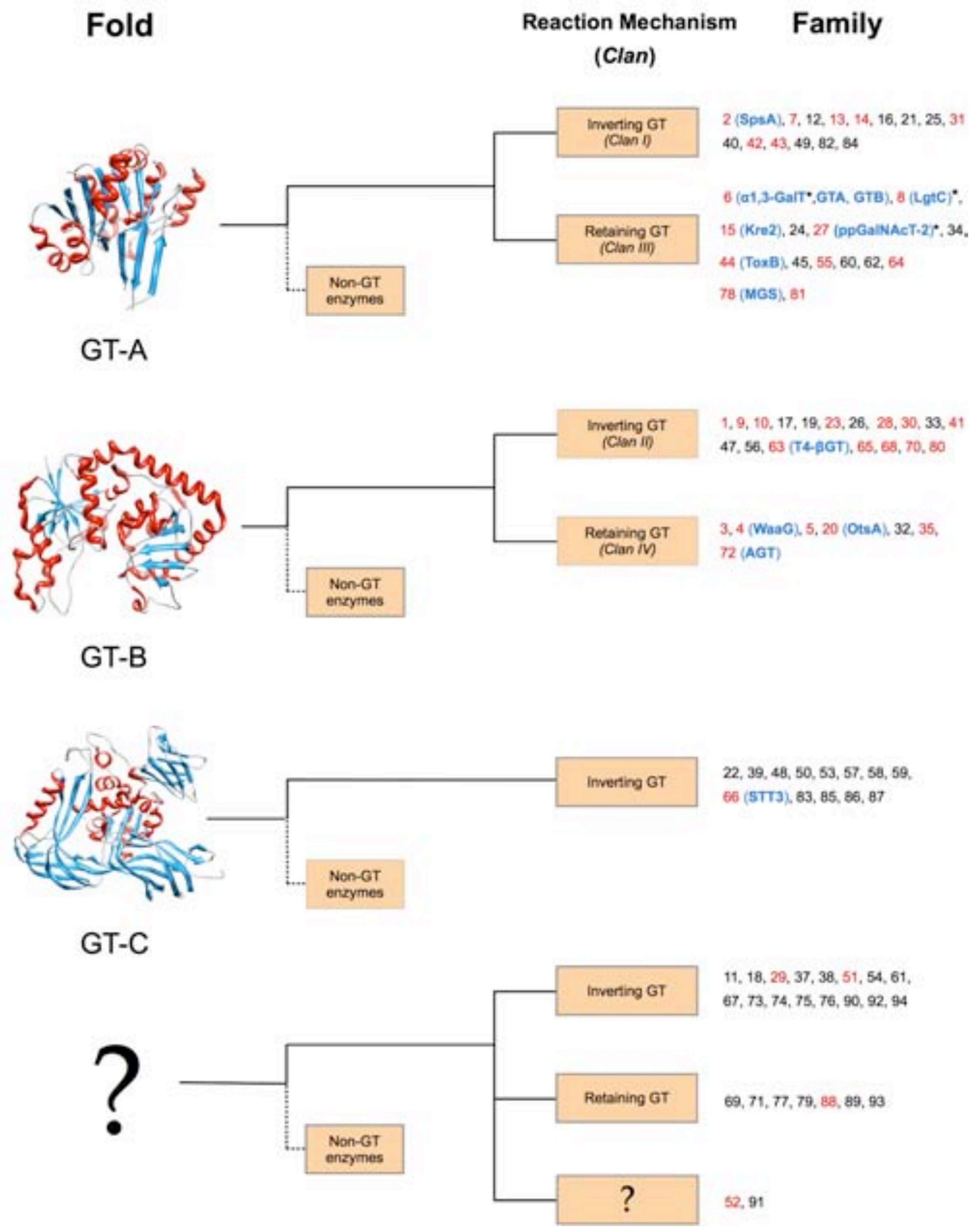

Figure 1.4. Glycosyltransferase (GT) classification system. In the case of GT-A and GT-B like folding families of GTs, they are classified into clans on the basis of their fold and activity as suggested by Coutinho et al.[25] Families exhibiting a predicted (by Liu & Mushegian[35] and/or the CAZY Web site) or confirmed (with solved 3-D structures) folding type are indicated in black or red, respectively. Notice that there are some X-ray structures available for some members of

the *"orphan"* families (those where no folding GT-A, GT-B or GT-C can be assigned). In these cases they adopt novel folding patters resembling GT-A or GT-B ones (i.e., GT29/GT88, or GT52 respectively). GT51 is better defined as adopting a lysozyme-like fold.[10] Names of the GTs mentioned throughout this manuscript are highlighted in blue. Those enzymes that were studied in the present work are additionally labelled with an asterisk. SpsA, *Bacillus subtilis* glycosyltransferase;[2] α1,3-GalT, bovine α -1,3-galactosyltransferase;[72] GTA, human α-1,3-N-acetylgalactosaminyltransferase;[73] GTB, human α-1,3-galactosyltransferase;[73] LgtC, *Neisseria meningitides* α-galactosyltransferase;[61] Kre2, *Saccharomyces cerevisiae* α-1,2-mannosyltransferase;[74] ppGalNAcT-2, human UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase-2;[75] ToxB, *Clostridium difficile* Toxin B;[76] MGS, *Rhodothermus marinus* mannosylglycerate synthase;[48] T4-βGT, bacteriophage T4 DNA β-glucosyltransferase;[33] Waag, *Escherichia coli* α-1,3-glucosyltransferase I;[77] OtsA, *Escherichia coli* trehalose-6-phosphate synthase;[78] AGT, Bacteriophage T4 α-glucosyltransferase;[79] STT3, *Pyrococcus furiosius* oligosaccharyltransferase.[8]

Figure 1.5. Catalytic mechanism followed by the inverting glycosyltransferases (inv-GTs). Abbreviations: R, a substituent (e.g., nucleoside, a nucleoside monophosphate, a lipid phosphate, or phosphate) and R'OH, an acceptor group (e.g., another sugar, a protein, or an antibiotic). A sphere is used to represent a divalent cation, although GTs with GT-B fold are known to use ion-independent mechanisms to stabilize the negative charge and facilitate the leaving group departure. Adapted from Figure 4 in Ref. [10].

The double-displacement mechanism was initially suggested by Koshland[80] for the retaining glycoside hydrolases (ret-GHs), for which it was strongly supported by experimental data.[81-85] In this way, ret-GTs were initially assumed to follow the same catalytic mechanism as ret-GHs despite an evident lack of evolutionary relatedness.[71]

On the other hand, and even if there are some experiments pointing in the direction of such a double-displacement, they are mainly inconclusive and direct evidence for the formation of CGE intermediates in ret-GTs has remained elusive along the years.[10,21,86,87]. In one of these experiments, Lairson et al.[86] replaced the side chain amide of the putative nucleophile glutamine 189 in LgtC with the more nucleophilic carboxylate-containing side chain of glutamate in the hope of accumulating a CGE intermediate. Liquid chromatographic/mass spectrometric analysis of fragmented proteolytic digests in the Q189E mutant identified the site of labeling not as Glu189 but, surprisingly, as the adjacent Asp190. However, the side chain carboxylate of Asp190 is located 8.9 Å away from the donor substrate analogue in the available wild-type crystal structure (PDB Code 1GA8[61]) and that is why it was initially discounted. The foregoing implies that if Asp190 were to serve as the catalytic nucleophile in a double-displacement mechanism, a conformational change would be required during catalysis allowing for an appropriate positioning relative to the donor sugar substrate. Interestingly, the authors discarded an altered active site conformation or mode of donor sugar binding in the Q189E mutant by obtaining the corresponding crystal structure (PDB Code 1SS9). Nevertheless, and even though this study represents the first report of the direct observation of a CGE intermediate covalently bound in a ret-GT, the site of glycosylation is possibly an artifact arising from the creation of an active site mutant and that is why is considered by no means conclusive.

Later on, Monegal & Planas[87] applied a *"chemical rescue"* methodology in bovine α -1,3-galactosyltransferase (α1,3-GalT) with the aim of providing new mechanistic information. This approximation was used before in a ret-GH[88] and is based on reactivation of inactive mutants of an enzyme by exogenous small molecules.

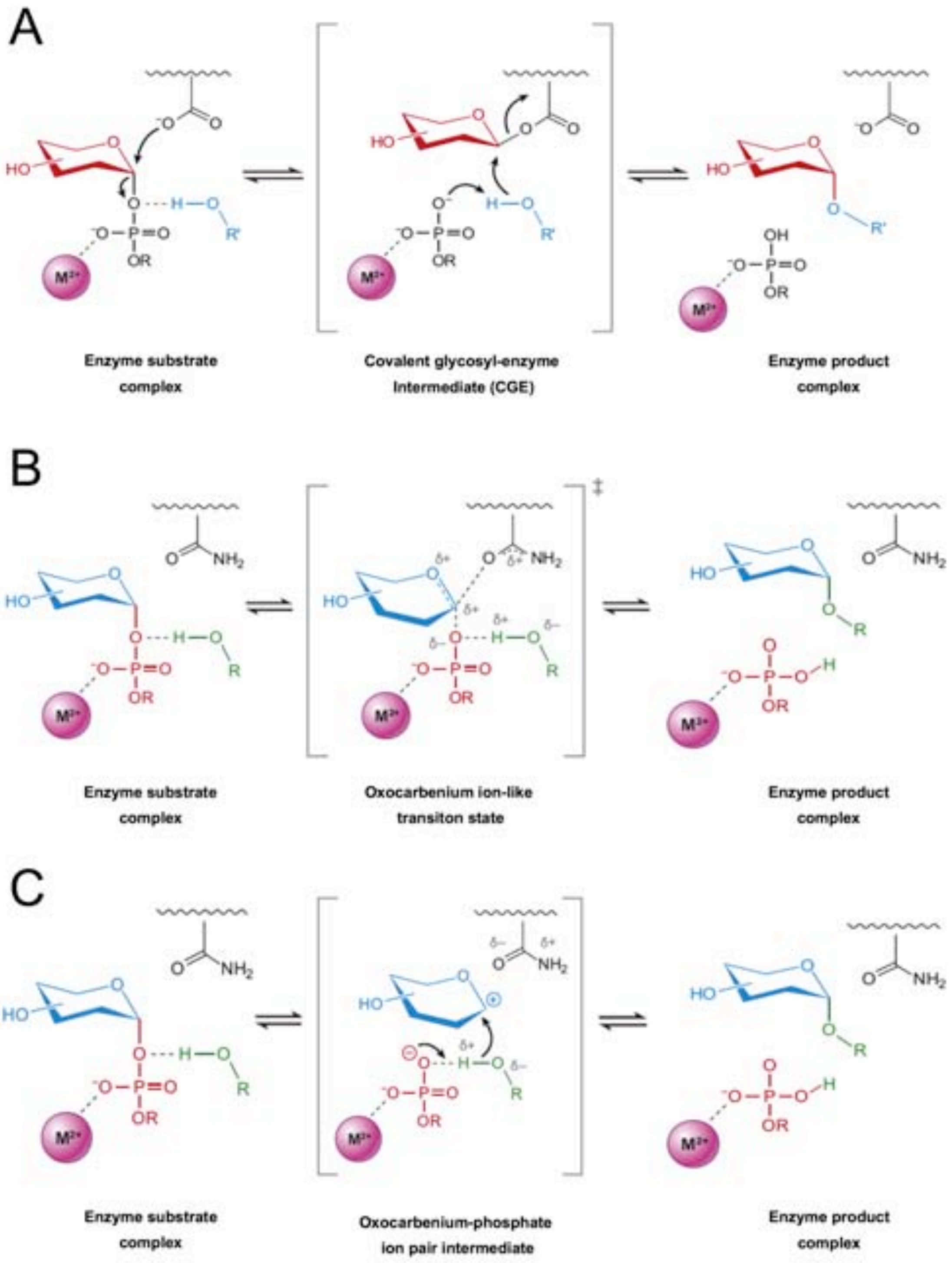

Figure 1.6. Proposed catalytic mechanisms for the retaining glycosyltransferases (ret-GTs). (***A***) Double-displacement mechanism with formation of a covalently bound glycosyl-enzyme intermediate (CGE). (***B***) Concerted front-side single displacement mechanism (S_Ni) with an oxocarbenium ion-like transition state. (***C***) S_Ni-like mechanism with formation of a short-lived oxocarbenium-phosphate ion pair intermediate. Adapted from Figures 4 and 10 in Ref. [10].

The addition of an exogenous nucleophile such as azide to mutants in which the catalytic nucleophile or the general acid/base has been replaced by alanine reactivates the enzyme leading to the corresponding α - or β -glycosyl azide adduct, the stereochemistry of which correlates with the putative function of the mutated residue. In this study the authors mutated the residue Glu317 (tentative nucleophile) and recovered part of the enzymatic activity when the side chain of Glu317 was removed and a cavity was generated (i.e. E317A mutant) but not in the case of the E317Q and E317I variants. The foregoing would support a double displacement mechanism where Glu317 acts as the catalytic nucleophile, since its role can be replaced by the exogenous nucleophile upon mutation to Ala, the same behavior obtained with ret-GHs. However, this experiment does not fully discard an alternative mechanism for ret-GTs and can only be considered as an evidence of the suitable position that this glutamate residue occupies in the active site of the enzyme for playing the role of a catalytic nucleophile.

More recently, direct detection of CGE intermediates in two ret-GTs, the human blood group synthesizing α -1,3-N-acetylgalactosaminyltransferase (GTA) and α -1,3-galactosyltransferase (GTB), was reported by Soya et al.[89] More specifically, mutants of GTA or GTB in which the putative catalytic nucleophile Glu303 was replaced with Cys (i.e. $\text{GTA}_{\text{E303C}}$ and $\text{GTB}_{\text{E303C}}$), were incubated with their respective donor substrate resulting in a covalent intermediate. Tandem mass spectrometry analysis confirmed Cys303 as the site of glycosylation. Finally, the exposure of these CGE intermediates to a disaccharide acceptor resulted in the formation of the corresponding enzymatic trisaccharide products, thus proving that these mutants could actually operate by a double-displacement mechanism.

Interestingly, the experimental approaches that have been successful applied to support the double-displacement mechanism in ret-GHs (e.g. fluorinated substrate analogues and mutagenesis studies) are commonly useless in ret-GTs.[10] In first place, the introduction of an electronegative fluorine at either the C2' or the C5' position of the pyranose sugar ring destabilizes the oxocarbenium ion-like transition states and would also removes key hydrogen-bonding interactions, resulting in a significant decrease in the rate of the overall reaction catalyzed by the ret-GT. Secondly, and since the leaving group has been postulated to play the role of base catalyst[10] in ret-GTs (See

Figure 1.6), the deglycosylation step cannot be slowed down in these enzymes by simple mutagenesis of a base catalyst as it is done with ret-GHs. Thus, this inability to alter the relative rates of glycosylation versus deglycosylation steps has rendered the available experimental approaches ineffective in the trapping of covalent intermediates in ret-GTs.[90] However, it should be pointed out that the lack of systematic evidence for the formation of covalent intermediates in ret-GTs does not discount the double displacement mechanism itself and it should still be taken into account when performing reactivity studies on ret-GTs.

The problem is even more challenging for those ret-GTs where no good candidate for the catalytic nucleophile seems to exist within the active site. In most cases a relatively poor nucleophilic agent (i.e., side chain or main chain amide) is the most suitably positioned and at a reasonable distance from the anomeric center to play the role of the catalytic nucleophile in a putative double-displacement mechanism.[10] An exception to this observation is found in ret-GTs from family GT6 that have a side chain carboxylate at this position. However, if all ret-GTs utilized the double-displacement mechanism, it would seem likely that during the course of divergent evolution they had conserved this feature in the active site, since it would be the most important component of their catalytic machinery. On the contrary, there is an apparent lack of conserved active-site architecture among these enzymes, in strong contrast with ret-GHs where at least a highly conserved carboxylate group is clearly positioned within the active site to play the role of the catalytic nucleophile that leads to the formation of the covalent intermediate.[10]

Consequently, alternative mechanisms involving retention of configuration at the anomeric center have been proposed. The most relevant is the so-called S_Ni mechanism, which occurs through a front-side attack of the acceptor nucleophilic group at the same side from which the leaving group departs and was first described for the decomposition of alkyl chlorosulfites.[91-93] Strictly speaking, the S_Ni mechanism is a form of a S_N1 reaction in which the nucleophile is derived by the decomposition of the leaving group and attacks the anomeric atom from the same face. Decomposition of the initial intermediate species and attack by the formed nucleophile occur at a rate that exceeds that of solvent attack and the ion pair reorganization that would be required for nucleophilic attack from the back face. Therefore the *"i"* indicates an *"**i**nternal return"*. Person et al. were the first ones to propose a S_Ni type mechanism for a ret-GT; the bacterial LgtC.[61] Notice, however, that in the sugar transfer reactions catalyzed by GTs the acceptor group is external and not internal, that it, it is not produced after the leaving group undergoes decomposition. The terminology S_Ni has been conserved, though, for historical reasons.

We could make a distinction regarding this front-side attack mechanism as it could take place in a single-step displacement process forming an oxocarbenium ion-like transition state (Figure 1.6B) or via a short-lived ion-pair if the oxocarbenium has a life time long enough to be considered an intermediate of the reaction (Figure 1.6C). According to IUPAC nomenclature they should be called as A_ND_N and $D_N{*}A_N$. For clarity, we will be calling these alternatives as S_Ni and S_Ni-like mechanisms respectively throughout this document.

In summary, there is no definitive evidence for one or another mechanism for ret-GTs and the debate remains open. As stated before, even for ret-GT from the GT-6 family, where it is possible to identify a putative and strong catalytic nucleophile (i.e., Glu or Asp residue), the identification of a covalent intermediate has been elusive. In any case, a double displacement mechanism has historically been assumed for the ret-GTs. On the other hand, the alternative front-side attack mechanism has been primarily suggested for some ret-GTs with available 3D structure (e.g., LgtC,[61] Kre2,[74] ToxB,[76] Extl2,[94] Mgs,[48] WaaG,[77] OtsA,[78] and AGT[79]) only because of the lack of an appropriately positioned nucleophile within their active sites but, once again, there is a notorious lack of direct supporting evidence.

1.2.4. Previous computational studies of catalytic mechanisms of glycosyltransferases. When this doctoral project started, by the end of 2009, there were just a few computational works devoted to the study of reaction mechanisms of GTs and, most of them, focused on inv-GTs. These precedent studies are summarized in Table 1.1, where some geometrical parameters and energy barriers associated with the transition states described, as well as the methods employed, are given.

Table 1.1. Geometrical parameters (Å) and energy barriers (kcal/mol) corresponding to the transition states found in previous theoretical studies of catalytic mechanisms of GTs. Opt., geometry optimization; SP, Single Point energy calculation; N_{QM}, number of QM atoms considered; $R_{(C1'-OA)}$, distance between the anomeric carbon (C1') and the oxygen from the incoming hydroxyl group of the acceptor molecule; $R_{(C1'-OB)}$, distance between C1' and the initially bonded oxygen atom from the leaving group; $R_{(H-OA)}$, distance between the hydrogen of the acceptor hydroxyl group (H) and the donor oxygen; $R_{(H-OB)}$, distance between H and its final acceptor oxygen; $V^{\ddagger}$, potential energy barrier; Ref., Reference.

System	RC/ Reopt. / SP [a]	N_{QM}	$R_{(C1'-OA)}$	$R_{(C1'-OB)}$	$R_{(H-OA)}$	$R_{(H-OB)}$	$V^{\ddagger}$	Ref.
GlcNAc-T (inv)	DFT(B3LYP/6-31G(d)) DFT(B3LYP/6-31++G(d,p))	86	2.16	2.54	1.00	1.81	13.4	95
			1.50	3.01	1.36	1.80	14.7	
GlcNAc-T (inv)	DFT(B3LYP/6-31G(d,p)) DFT(B3LYP/6-31++G(d,p))	127	1.88	2.38	1.23	1.18	42.2 ~ (20-25)	96
GlcNAc-T (inv)	DFT(BP/TZP)/MM(AMBER)	88	1.91	2.54	1.31	1.12	18.74	97
β4Gal-T1 (inv)	DFT(BP/DZP)/MM(AMBER) DFT(BP/TZZP)/MM(AMBER)	253	2.70	2.09	1.11	1.35	15	98
Model (ret)	DFT(B3LYP/6-31G(d)) DFT(B3LYP/6-311++G(d,p))	74	2.20 [b]	1.64	---	---	16.5	99
			2.19	2.28 [b]	1.02	---	8.45	
LgtC (ret)	DFT(B3LYP/6-31G(d)) DFT(B3LYP/6-311++G(d,p))	136	2.34	2.66	1.09	1.34	31.34 ~ (10-24)	3

[a] Methodologies used in the geometry optimizations and energy (SP) calculations, respectively. The presence of only one method indicates that a single level of calculation was considered in the entire study. For a better understanding of the theoretical approaches mentioned, check Chapter 3 Section 3.2.

[b] Distance to oxygen of the catalytic residue forming the covalent intermediate in the proposed model.

Concerning the inv-GTs, the available theoretical studies also support the concerted S_N2-like displacement mechanism as provided by both pure Quantum Mechanical (QM)[95,96] or hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) studies.[97,98] Therefore, and as we have stated before, there is a general agreement about the catalytic mechanism of inv-GTs since it has been extensively and systematically supported by experimental and theoretical approaches.

On the other hand, there were only two studies dealing with the retention of the configuration at the anomeric center. Moreover, these were purely QM studies of small clusters that did not consider the entire enzyme environment, the latest being an additional motivation at the time of starting the present study of the ret-GTs catalytic mechanisms. In the first of these studies, André et al.[99] built a general and reduced model based on what they considered a typical ret-GT active site. A double displacement mechanism is presumed, and therefore no alternative mechanisms where tested. More specifically, the attention was focused on the activation process of the incoming hydroxyl group from the acceptor molecule. They considered three possibilities for the proton subtraction: *a*) by the catalytic nucleophile forming the covalent intermediate, *b*) an additional base catalyst in the active site or *c*) the UDP leaving group. The latest option resulted in more consistent results when considering the available kinetic data for the ret-GTs and, thus, this work supports from a theoretical perspective the role of the UDP leaving group as a base catalyst in ret-GTs.

The second and final work considered a cluster of 136 atoms based on the available crystallographic coordinates of the bacterial ret-GT LgtC (PDB Code 1GA8[61]). This time, a front-side mechanism where the UDP activates the incoming hydroxyl group was shown as the most favorable.[3] The initial calculated energy barrier was slightly high (~31 kcal/mol) although it was finally estimated in ~10-24 kcal/mol after simulating a continuous dielectric field and performing energy corrections. The latest can be considered as a limitation of these pure QM approximations where only a reduced portion of the active site is considered and, thus, a more realistic effect of the protein environment is commonly requested to achieve proper or more consistent results.

PART II

Objetives

2

OBJETIVES

After Proteomics and Genomics have done it, Glycobiology/Glycomics is now starting its revolution and, at the centre of it, are glycosyltransferases (GTs). These are the biocatalyst involved in the synthesis of the wide variety of glycan molecules present in nature. Glycans participate in a variety of biological functions, including cellular and molecular recognition, energy storage and structural stability. It has been estimated that > 50% of the proteins are glycosylated[100] and several GTs have also been identified as drug targets to fight against different diseases and infection. Moreover, a detailed understanding of the mechanism by which glycans are synthesized will open new horizons in many research areas, with an ultimate benefit in industry, health and energy/environment. While the catalytic mechanism of other enzymes involved in carbohydrate metabolism is in general understood (i.e. for glycosyl hydrolases, glycan phosphorylases, polysaccharide lyases and inv-GTs), it has remained less clear in the case of ret-GTs; basically because there is no conclusive evidence directly supporting neither a double-displacement nor a front-side attack mechanism.

In that context, the results presented in the present thesis are aimed to provide a systematic study of the catalytic mechanism of ret-GTs, which we hope will be useful to the scientific community. More specifically, theoretical hybrid quantum-mechanics/molecular mechanics (QM/MM) studies were performed for three ret-GTs: Lipopolysaccharyl-α-1,4-galactosyltransferase C from *Neisseria meningitides* (LgtC), bovine α-1,3-galactosyltransferase (α1,3-GalT) and human UDP-N-acetylgalactosamine:polypeptide N-acetyl-α-galactosaminyltransferase 2 (ppGalNAcT-2).

Three specific objectives were fulfilled in each case:

I. To model the Michaelis complex of the enzyme in presence of the donor and acceptor substrates.
II. To simulate the main catalytic mechanisms proposed in the literature, that is, a double-displacement or a front-side attack mechanism and characterize the corresponding stationary points (i.e. minima and transition states).
III. To analyse the factors responsible for the catalytic efficiency of the enzyme.

By complementing these specific objectives for every ret-GT under study we intent to obtain useful information that can be ultimately integrated to fulfil our primary goal: to provide a general overview of the catalytic mechanism of ret-GTs.

Finally and taking into account our theoretical outcomes we want to answer some intriguing questions like:

a) Is there a common mechanism for retaining glycosyltransferases?
b) Why just some of these enzymes (i.e. from family GT6) might require a strong nucleophile in their active site?
c) Why the isolation of a covalent intermediate (even in the case of ret-GTs with such a strong nucleophile) has been so elusive for experimentalists?

PART III

Theoretical Framework

3

THEORETICAL FRAMEWORK

"*Everything should be made as simple as possible, but not simpler.*"
Albert Einstein

3.1. PRELUDE

Computational Chemistry is a relatively new field of knowledge that appeared in the late 1960's. From that moment on Chemistry has not been uniquely an experimental science and theoretical calculations are now widely used for a big number of relevant topics, including enzymatic catalysis.

Most of the theoretical frame used nowadays was already developed in the early stage of the 20th century, but applications in chemistry were long in coming. More specifically, it was practically impossible to handle the complicated mathematical relations of quantum mechanics for such complex systems as biomolecules.

One of the founders of quantum physics, Paul Dirac, expressed the problem in 1929 as follows: *"The fundamental laws necessary for the mathematical treatment of large parts of physics and the whole of chemistry are thus fully known, and the difficulty* *lies only in the fact that application of these laws leads to equations that are too complex to be solved."*[101]

Fortunately, things began to change at the beginning of the 1960s when computers came into use for solving these equations. As well as producing quantitative information on molecules and their interactions, the theory also provides deeper understanding of molecular processes that cannot be obtained from experiments alone. Thus, theory and experimentation combine today in the search for understanding the inner structure of matter. Moreover, computational chemistry has become a useful tool for getting the reasons behind a chemical phenomenon, possibly because it is easier (and surely cheaper) than performing the relevant experiment at the laboratory. Finally, and since the computational capabilities are in a continuously improvement process, more complicated systems are now affordable.

The goal of this chapter is to give a general overview of the computational methodologies used throughout the thesis. Most relevant references are provided, in case the reader would like to go deeper into details. More specifically, basic concepts of Molecular Mechanics (MM), Quantum Mechanics (QM), hybrid QM/MM and Molecular Dynamics (MD) approaches will be addressed. Specific details about the methods and their utilization are included in the following chapters whenever it was considered necessary. Finally, a general vision about the theoretical study of enzymatic catalysis will be provided.

3.2. MOLECULAR MECHANICS (MM)

Molecular mechanics (MM) uses classical mechanics to model molecular systems. It is based on the assumption that the electrons find an optimal distribution about the nuclei without being explicitly treated. Therefore, it has the advantage that it can be used to study small molecules as well as large biological systems or material assemblies from just a few to many thousands of atoms.

In this approximation the potential energy can be represented through molecular mechanics as a parametric function: a force field (FF) depending on all the nuclear coordinates. The latest implies that the MM approach cannot be used to predict any property depending on the electronic distribution, including reactive processes.

Force fields treat a molecule as a collection of particles held together by simple forces. These various types of forces are described in terms of covalent and noncovalent interactions, and the sum of these interactions constitutes the overall molecular potential energy of the molecule, described as follows:

$$E_{total} = E_{bonded} + E_{non-bonded} \tag{3.1a}$$

$$E_{bonded} = E_{stretching} + E_{bending} + E_{torsion} + E_{improper} \tag{3.1b}$$

$$E_{non-bonded} = E_{vanderWaals} + E_{electrostatic} \tag{3.1c}$$

The exact analytical expression for each of these energetic terms depends on the particular FF being used. Generally the bond and angle terms are modeled as harmonic potentials centered on equilibrium values (eq) derived from experiment or theoretical calculations.

$$E_{stretching} = \sum_{i}^{bonds} k_i \left(r_i - r_i^{eq} \right)^2 \tag{3.2a}$$

$$E_{bending} = \sum_{j}^{angles} k_j \left(\theta_j - \theta_j^{eq} \right)^2 \tag{3.2b}$$

$$E_{improper} = \sum_{h}^{improp} k_h \left(\gamma_h - \gamma_h^{eq} \right)^2 \tag{3.2c}$$

There are other terms having multiple minima that cannot be modeled as harmonic oscillators (e.g, torsion term; Eq. 3.3). These terms are treated by regular analytical functions that may vary with the implementation.

$$E_{torsion} = \sum_{k}^{dihed} \frac{V_k}{2} \left(1 + \cos\left(n\phi_k - \gamma_k \right) \right) \tag{3.3}$$

The non-bonded terms are commonly defined as follows:

$$E_{vanderWaals} = \sum_{l}^{atoms} \sum_{m>l}^{atoms} 4\varepsilon_{lm} \left[\left(\frac{\sigma_{lm}}{r_{lm}} \right)^{12} - \left(\frac{\sigma_{lm}}{r_{lm}} \right)^{6} \right] \tag{3.4a}$$

$$E_{electrostatic} = \sum_{l}^{atoms} \sum_{m>l}^{atoms} \frac{q_l q_m}{r_{lm}} \tag{3.4b}$$

where ε and σ are parameters defining the depth of the potential well and the finite distance at which the inter-particle potential is zero, respectively. Both non-bonded terms are normally modulated by a shifting or switching function.

The non-bonded terms are much more computationally costly to calculate, since a typical atom is bonded to only a few of its neighbors, but interacts with every other atom in the molecule. The van der Waals term is typically modeled using a *"6-12"* or "*Lennard-Jones potential"*[102] that falls off rapidly (i.e. attractive forces fall off with distance as r^{-6} and repulsive forces as r^{-12}) as shown in expression 3.4a.

On the other hand, the electrostatic terms are notoriously difficult to calculate well because they do not fall off rapidly with distance, and long-range electrostatic interactions are often important features of the system under study (especially for proteins). The basic functional form is the *Coulomb potential*, which only falls off as r^{-1}, as shown in equation 3.4b. A variety of methods are used to address this problem, the simplest is to consider a cutoff radius similar to that used for the van der Waals terms. Other more sophisticated but computationally intensive methods are the *particle mesh Ewald*[103] (PME) and the *multipole algorithm.*[104]

The set of parameters from equations 3.2-3.4 identify a particular FF and parameterization is typically done through agreement with experimental values and theoretical calculations. Even so, such a fast way to get a molecular total potential energy needs some approximations to make it computationally affordable. Molecules consist of atoms, and atoms can usually have a similar bond environment in very different molecules. The development of parameter sets is a very laborious task, requiring extensive optimization. On the other hand, if each atom would have its own set, one should optimize an unaffordable quantity of parameters. For this reason, fixed sets of atom types are employed when determining the parameters for a FF. Moreover, the sum terms in equations 3.4a,b implies to consider all the possible atom pairs increasing the volume of operations in an unfeasible way. Finally, if the non-bonded interactions are roughly set to zero for interatomic distances greater than the cutoff distance, the method can lead to large fluctuations in the energy. In this way, some improvements to terminate the interaction between two atoms have been developed over the years: the *shift* and *switch* approximations.[105]

The first use of a FF is attributed to M. Levitt and S. Lifson, when in 1969 refined the conformation of a protein from its X-ray structure.[106] One year later, A. Warshel and Lifson published the first consistent FF[107] and N. L. Allinger and J. T. Sprague implemented in 1973 the FF MM1 in the first MM software package.[108] Up to now, several FFs have been developed, published and incorporated into various molecular modeling packages. For simulations on biomolecular systems, like enzymes (the particular case of this thesis), the most noteworthy FFs are AMBER,[109,110] OPLS,[111] GROMOS,[112] and CHARMM.[113,114] Each of them has different parameters or even different atom types, and these variations cause the information transfer between them to be arduous. However, some comparative studies of different FFs reveal that, in spite of some discrepancies, the global result is usually similar for all of them.[115]

3.3. QUANTUM MECHANICS (QM)

There are two general approximations to describe the properties of the molecular systems using quantum mechanics; based on the *wavefunction* or the *electronic density* as primary source of information, respectively. A description of both approaches will be made here, with emphasis on the second one since it is the most used QM methodology in the present thesis.

3.3.1. Wavefunction approach. The final goal of most of the methods based on the wavefunction is to solve the non-relativistic and time independent Schrödinger equation

$$\hat{H}\Psi(\vec{r},\vec{R}) = E\Psi(\vec{r},\vec{R}) \tag{3.5}$$

where $\Psi(\vec{r},\vec{R})$ is the wavefunction as a function of the electron and nuclear coordinates, respectively, E is the total energy of the state described by Ψ and $\hat{H}$ is the Hamiltonian operator for the entire system including *M* nucleus and *N* electrons. The latest equation becomes mathematical unfeasible and that is why some approximations are commonly assumed. The first and most important of such approximations in quantum chemistry separates the nuclear and the electronic motions based on the different masses of the particles involved. Neutrons and protons are around 1800 times heavier than electrons; thus, the electron motion around the nucleus is much faster than the nuclear movement itself. Consequently, electrons can be considered to move in the field generated by *"static"* nucleus. The latest is also known as the *Born-Oppenheimer approximation*[116] and has a very useful implication by splitting the wavefunction into electronic and nuclear parts. In practice, and since the nuclear movement is voided, their kinetic energy is null and the potential energy associated to the nucleus-nucleus repulsion ($\hat{V}_{Nuc-Nuc}$) is constant for a given nuclear configuration. By making such a consideration, the Hamiltonian in equation 3.5 is reduced to the electronic Hamiltonian as

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{r_{ij}} = \hat{T}_e + \hat{V}_{Nuc-e} + \hat{V}_{ee} \tag{3.6}$$

The three terms of this electronic Hamiltonian ($\hat{H}_{elec}$) describe the kinetic energy for the electrons, nucleus-electron attractions and electron-electron repulsions, respectively. Then, the electronic Schrödinger equation is expressed like

$$\hat{H}_{elec}\Psi_{elec}(\vec{r};\vec{R}) = E_{elec}(\vec{R})\Psi_{elec}(\vec{r};\vec{R}) \tag{3.7}$$

and the potential energy surface is obtained as

$$U(\vec{R}) = E_{elec}(\vec{R}) + V_{Nuc-Nuc}(\vec{R}) \tag{3.8}$$

where $U(\vec{R})$ is the total energy of the system assuming that the nuclei are frozen at the nuclear configuration $\vec{R}$.

However, the equation 3.7 does not have an exact solution in systems with more than one electron and new approximations have to be considered.

3.3.1.1. Hartree-Fock method. The Hartree-Fock[117] (HF) is the most simple *ab initio* methodology dealing with polielectronic systems. The term *ab initio* (Latin for *"from scratch"*) is referred to that group of methods in which molecular structures can be calculated using nothing but the Schrödinger equation, the values of the fundamental constants and the atomic numbers of the atoms present.

This is a variational method that basically represents the *N*-electronic wavefunction by an antisymmetrized product of *N* monoelectronic functions χ_i. Such a product is also known as the *Slater determinant*. The monoelectronic functions are called spin-orbitals and are the product of a spatial function φ_i and one function of the spin, $\alpha(s)$ or $\beta(s)$. The latest form an orthonormal set, that is, $\langle\alpha|\alpha\rangle = \langle\beta|\beta\rangle = 1$ and $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle = 0$. Varying the spin-orbitals the Slater determinant of lower energy is found, subject to the constraint that the spin-orbitals remain orthonormal. The HF equation is obtained like

$$\hat{f}\chi_i = \varepsilon_i\chi_i \tag{3.9}$$

where $\hat{f}$ is the Fock operator which includes the mean potential created over an electron by the nucleus and the rest of electrons; χ_i are the canonical spin orbitals and ε_i is the energy of the ith spin-orbital. In equation 3.9 the $\hat{f}$ operator depends on the spin-orbitals to be determined and that is why an iterative approach has to be considered in order to solve the equations. The latest is known as the *Self-Consistent Field* (SCF) technique.

3.3.1.1.1. Basis set and Roothan-Hall equations. To solve the HF equation Roothann[118] and Hall[119] proposed in 1951 to express every molecular orbital as a linear combination of atomic orbitals (ϕ_α), which constitute what is called the *basis set*:

$$\varphi_i = \sum_{\alpha}^{K} c_{\alpha i}\phi_\alpha \qquad (3.10)$$

In this equation the coefficients $c_{\alpha i}$ are determined in a variational way and ϕ_α represents the basis functions. The expression in Eq. 3.10 is also known as the *Linear Combination of Atomic Orbitals* (LCAO) approximation. It's implementation to the HF equation leads to the so-called *Roothan-Hall equations*

$$FC = SC\varepsilon \qquad (3.11)$$

which once expressed in matrix form, *S* corresponds to the overlap matrix, *F* to the Fock matrix and ε to the diagonal matrix of the orbital energies. The elements of the F matrix can be written as

$$F_{\alpha\beta} = h_{\alpha\beta} + \sum_{\gamma\delta} P_{\gamma\delta} G_{\alpha\beta\gamma\delta} \qquad (3.12)$$

The $h_{\alpha\beta}$ elements are the *core* integrals, $G_{\alpha\beta\gamma\delta}$ contains the two-electron integrals, and $P_{\gamma\delta}$ are the element of the so-called density matrix, which depends on the molecular orbital coefficients $C_{\gamma i}$ and $C_{\delta i}$. Once the equation 3.12 is solved the molecular orbital coefficients can be obtained.

The HF limit, that is, the lowest possible energy when using the HF approximation is obtained when considering an infinite number of atomic basis functions. Since this is obviously impossible, a finite subset of truncated basis functions has to be chosen. In a *minimal basis set* there is only a basis function per atomic orbital. Further extensions have to be considered to improve the results. These larger basis sets are called double-ζ, triple-ζ, etc. depending on the number of basis functions used to describe the atomic orbitals. Aditionally, *polarization functions* can be added to take into consideration the distortion of the atomic orbitals when bonds are formed. Finally, other chemical aspects like those concerning anions or radicals require the use of *diffuse functions* to properly describe the behavior of the most distant electrons from the nucleus.

As mentioned before, is it desirable to use an infinite basis set to solve the HF equations. This way the HF methodology is able to provide up to 99% of the total electronic energy. Although it might look negligible, the remaining 1% can result important in certain events. This small fraction represents the correlation energy of the system and is calculated as the difference between the exact non-relativistic energy and the HF limit energy, $E_{corr}= E_0-E_{HF}$.[120] The correlation energy is a direct consequence of considering the inter-electron interactions as a mean within the HF approximation. Many efforts have been made to include this energy in the calculations in the so-called post-Hartree-Fock methodologies (e.g., *Møller-Plesset*[121,122] and *Coupled Cluster*[123,124] methods). Moreover, methods based on the density-functional theory (section 3.3.2.) also take into account the correlation energy.

3.3.1.2. Semiempirical methods. The cost of performing a HF calculation scales formally as the fourth power of the number of basis functions. This arises from the number of two-electron integrals necessary for constructing the Fock matrix (Eq. 3.12). Semiempirical methods reduce the number of these integrals to be calculated, thus decreasing the scalability.

The first step to reduce the computational problem is to consider explicitly only the valence electrons. The core electrons are accounted by reducing the nuclear charge or introducing functions to model the combined repulsion due to the nuclei and core electrons. Furthermore, a minimum basis set (1s and 3p closed-shell orbitals) is commonly used for the valence electrons. The original assumption of semiempirical methods was the *Zero Differential Overlap* (ZDO) approximation, which neglects all products of two different basis functions that depend on the same electron coordinates; that is, two-center one-electron as well as the numerous three and four-center two-electron integrals are null. The orbital coefficients are treated variationally, the Roothaan-Hall equations (Eq. 3.11) are used and the corresponding orbital energies are solutions of the following secular equation:

$$\left| F_{\mu\nu} - \varepsilon S_{\mu\nu} \right| = 0 \qquad (3.13)$$

It is imposed that the orbitals form an orthonormal set, so $S_{\mu\nu} = 0$ unless $\mu = \nu$ and the overlap matrix S is reduced to the identity matrix. To make up for these approximations, the remaining integrals are converted into parameters, whose values come from both rigorous theoretical calculations and experimental results. Stated differently, the parameterization is performed by adjusting the constants involved in the different semiempirical methods so that the results fit experimental data as closely as possible. This is in a sense wrong since as it was mentioned before, the HF method cannot give the correct result, even in the limit of an infinite basis set. Moreover, semiempirical methods share the advantages and disadvantages of force field methods: they perform best for systems where much experimental information is already available but they are unable to predict totally unknown compound types.

J. A. Pople[*] formulated the semiempirical approximations CNDO and NDDO in 1965,[125,126] from which most of the current semiempirical methods are derived. Many others have been developed since (e.g., popular semiempirical methods MNDO,[127] AM1,[128] and PM3[129]). All these methods apply the ZDO approximation only for basis functions centered on different atoms. For further details about semiempirical approximations, weak points and perspectives look at Ref. [130].

In the present thesis the *"self-consistent-charge density-functional tight-*

---

[*] John A. Pople was co-awarded the 1998 Nobel Prize in Chemistry *"for his development of computational methods in quantum chemistry."* He is considered one of the founders of Theoretical Chemistry as a discipline.

binding"[131] (SCC-DFTB) method was used. This is a semiempirical approach inspired on the *density-functional theory* (DFT) (described in the next section) that is increasingly being applied in theoretical studies on biomolecules. Strictly speaking, SCC-DFTB is an extension of the standard *tight-binding* (TB)[132] theory to operate in a self-consistent charge mode based on a second-order expansion of the Kohn-Sham total-energy functional as calculated within DFT.

3.3.2. Density-Functional Theory (DFT). Conventional calculation of the properties of molecules (i.e. using wavefunction-based methods) is based on the description of the motion of individual electrons. For this reason, such approaches are mathematically complicated. On the other hand, Walter Kohn[*] showed that the electronic energy for a system described by the laws of quantum mechanics could be theoretically calculated if the spatial distribution of the electrons (i.e., electron density (ρ)) is known.

$$\rho(\vec{r}\,) = N\int\cdots\int\left|\Psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_N)\right|^2 ds_1 d\vec{x}_2 ... d\vec{x}_N \qquad (3.14)$$

Thus, for a system with N electrons $\rho(\vec{r})d\vec{r}$ represents the probability of finding any of the N electrons in the infinitesimal volume space $d\vec{r}$, while the remaining N-1 electrons have arbitrary positions in the rest of the space. $\rho(\vec{r})$ has limit zero when $\vec{r}$ tends to infinite and integrates the total number of electrons.

$$\rho(\vec{r} \to \infty) = 0\,, \qquad \int\rho(\vec{r})d\vec{r} = N \qquad (3.15)$$

Therefore, it is not necessary to consider the motion of each individual electron but it suffices to know the average number of electrons located at any one point in space. This has led to a computationally simpler approach, the *density-functional theory*

---

[*] Walter Kohn (1923-) is an American physicist who was co-awarded the 1998 Nobel Prize in Chemistry *"for his development of the density-functional theory."*

(DFT). The advantage of using this approximation over the wavefunction-based methods is clear; while the wavefunction of a system with N electrons depends on $4N$ coordinates (3 spatial and an additional one if the spin is considered), $\rho(\vec{r})$ only depends on three spatial coordinates, regardless of the number of electrons N. Thus, the complexity of the wavefunction grows exponentially with N but the density function always depends on the same number of variables, no matter the system's size. Moreover, the wavefunction is physically meaningless while the electron density is an observable that can be measured experimentally (e.g. by X-ray diffraction). Consequently, DFT methods have emerged as a powerful tool and are widely used nowadays in theoretical calculations. For further details on the theoretical and methodological aspects, as well as practical applications in specific problems, several reviews are available in the literature.[133-143]

3.3.2.1. Hohenberg-Kohn Theorems. The cornerstones of all the methodologies relying on the DFT theory are the *Hohenberg-Kohn theorems*[144] from 1964.

The first of these theorems demonstrates that the ground state properties of a many-electron system are uniquely determined by the electron density $\rho(\vec{r})$. In this context, the electronic energy E_0 can be expressed as

$$E_0\left[\rho_0\right] = \int \rho_0(\vec{r})V_{ne}(\vec{r})d\vec{r} + F[\rho_0(\vec{r})] \tag{3.16}$$

where $V_{ne}(\vec{r})$ is the potential energy due to the electron-nucleus interactions and $F[\rho_0(\vec{r})]$ is the universal functional defined as

$$F[\rho_0(\vec{r})] = T[\rho_0(\vec{r})] + J[\rho_0(\vec{r})] + E_{ncl}[\rho_0(\vec{r})] \tag{3.17}$$

Here, $T[\rho_0(\vec{r})]$ includes the individual contributions of the kinetic energy, $J[\rho_0(\vec{r})]$ includes the classical Coulomb interaction and the functional $E_{ncl}[\rho_0(\vec{r})]$ includes the *non-classical* contribution to the electron-electron interaction (i.e. self-interaction correction, exchange and Coulomb correlation). $J[\rho_0(\vec{r})]$ is the only known term in expression 3.15 and to find explicit forms of functionals $T[\rho_0(\vec{r})]$ and $E_{ncl}[\rho_0(\vec{r})]$ is the major challenge in DFT.

The second Hohenberg-Kohn theorem allows putting the theory into practice. It states that the functional $F[\rho_0(\vec{r})]$ delivers the lowest energy of the system if and only if the input or trial density ρ_{pr} is the true ground state density. This is nothing but the variational principle:

$$E_0\left[\rho_0(\vec{r})\right] \le \int \rho_{pr} V_{ne}(\vec{r})dr + T\left[\rho_{pr}(\vec{r})\right] + J\left[\rho_{pr}(\vec{r})\right] + E_{ncl}\left[\rho_{pr}(\vec{r})\right] \qquad (3.18)$$

It should be pointed out that although the second Hohenberg-Kohn theorem indicates how to choose the best electronic density, the main problem of the DFT methodology is still that you do not know a priori the form of the functionals for the kinetic energy and the non-classical interactions inside of the universal functional $F[\rho_0(\vec{r})]$. Consequently, some approximations are needed in order to build these unknown functionals. One of the most common approaches is to use atomic orbitals as proposed by Kohn and Sham[145] in 1965.

3.3.2.2. The Kohn-Sham method. In this approach, a *fictitious reference system* of *N non-interacting* electrons is assumed. All the electrons are under the influence of the same potential energy function $v_s(\vec{r}_i)$, and thus, the probability density function for the ground state $\rho_s(r)$ of the reference system is equal to the exact probability density function for the ground state of the system of interest $\rho_0(\vec{r})$ (i.e., $\rho_s(\vec{r}) = \rho_0(\vec{r})$). Accordingly, the Hamiltonian of the reference system can be defined as a sum of monoelectronic Hamiltonians of *non-interacting* electrons.

$$\hat{H}_s = \sum_{i=1}^{n}\left[-\frac{1}{2}\nabla_i^2 + v_s(\vec{r}_i)\right] \equiv \sum_{i=1}^{n}\hat{h}^{KS}(i) \tag{3.19}$$

where $\hat{h}^{KS}$ is the Kohn-Sham one-electron Hamiltonian. The exact wavefunction for this system is a Slater determinant, and the orbitals φ_i^{KS} are determined by the eigenvalue equation

$$\hat{h}^{KS}\varphi_i^{KS} = \varepsilon_i\varphi_i^{KS} \tag{3.20}$$

where φ_i^{KS} and ε_i are the Kohn-Sham orbitals and monoelectronic energies, respectively. These monoelectronic equations are solved in an iterative way and since the operator h_i^{KS} depends on the electronic density, a trial electronic density is needed as starting point. The spatial part of φ_i^{KS} can be expanded as a linear combination of basis functions and that is why to find the solution of the Kohn-Sham equations consists in finding the corresponding coefficients, analogously to what is done in the Hartree-Fock approximation. Since we are dealing with a system where the electrons do not interact, the kinetic energy can be exactly determined as

$$T_s = -\frac{1}{2}\sum_{i=1}^{N}\left\langle\varphi_i^{KS}\middle|\nabla^2\middle|\varphi_i^{KS}\right\rangle \tag{3.21}$$

Obvioulsy and since the kinetic energy of the reference system is different to the real kinetic energy the functional $F[\rho(\vec{r})]$ was redefined by Kohn and Sham as

$$F[\rho(\vec{r})] = T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})] \tag{3.22}$$

where $E_{XC}[\rho(\vec{r})]$ is the so called exchange-correlation energy defined like

$$E_{XC} = (T[\rho(\vec{r})] - T_s[\rho(\vec{r})]) + E_{ncl}[\rho(\vec{r})] \tag{3.23}$$

This expression includes the correlation kinetic energy within the first parenthesis as well as the rest of the non-classical contributions ($E_{ncl}[\rho(\vec{r})]$). Since the exact form of the exchange-correlation energy functionals $E_{XC}[\rho(\vec{r})]$ is unknown, the main difficulty of DFT methods rely on their construction. In that sense, some of the most popular approaches will be briefly reviewed.

3.3.2.3. Local Density Approximation (LDA). The Local Density Approximation[146-148] (LDA) is the basis of all approximate exchange-correlation functionals. At the center of this model is the idea of an uniform electron gas. This is a model system in which electrons move on a positive background charge distribution such that the total ensemble is neutral. In practice what is assumed is that $\rho(\vec{r})$ varies slowly with position and thus $E_{XC}[\rho(\vec{r})]$ can be expressed in the following form:

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r})\varepsilon_{XC}(\rho)d\vec{r} \tag{3.24}$$

The quantity $E_{XC}[\rho(\vec{r})]$ can be further split into exchange and correlation contributions; that is, $E_{XC}(\rho(\vec{r})) = E_X(\rho(\vec{r})) + E_C(\rho(\vec{r}))$. No explicit expression is known for the correlation part $E_C(\rho(\vec{r}))$ while $E_X(\rho(\vec{r}))$ was derived by Bloch and Dirac in the late 1920's and $E_X^{LDA}(\rho)$ can finally be expressed as

$$E_X^{LDA}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int [\rho(\vec{r})]^{4/3} d\vec{r} \qquad (3.25)$$

In a more general sense, when the spin densities α and β are different the *Local Spin Density Approximation*[149] (LSDA) is considered and $\rho(\vec{r})$ is split into $\rho^{\alpha}(\vec{r})$ and $\rho^{\beta}(\vec{r})$.

Regardless of its simplicity, LDA provides with reasonable results in a vast number of systems. The main reason is a systematic cancelation of errors since LDA underestimates $E_X(\rho)$ and overestimates $E_C(\rho)$ so that it commonly generates a correct $E_{XC}(\rho)$ value.[143] However, LDA outcomes are insufficient in many systems like heavy atoms, so dominated by electron-electron interaction effects.

3.3.2.4. Generalized Gradient Approximation (GGA). The first logical step to go beyond LDA is to use not only the information about the density $\rho(\vec{r})$ at a particular point $\vec{r}$, but to supplement it with information about the gradient of the charge density $\nabla\rho(\vec{r})$ in order to account for the non-homogeneity of the real electron density. Thus, in the context of the *Generalized Gradient Approximation* (GGA), the exchange-correlation energy can be written as

$$E_{XC}^{GGA}[\rho^{\alpha},\rho^{\beta}] = \int f(\rho^{\alpha}(\vec{r}),\rho^{\beta}(\vec{r}),\nabla\rho^{\alpha}(\vec{r}),\nabla\rho^{\beta}(\vec{r}))d\vec{r} \qquad (3.26)$$

Two main philosophies have been used while developing successful GGA's. The first has an empirical nature as initially proposed by Becke[150-155] and is based on numeric fitting processes considering a big set of test molecules. Some examples are the exchange functionals Becke88 (B88),[156] modified Perdew-Wang functional (mPW),[157,158] OptX (O),[159] and X;[160] and the exchange-correlation Perdew-Wang (PW).[157] The second group of GGA's methods was proposed by Perdew,[157,161,162] and considers that the development of exchange-correlation functionals should be based on principles coming strictly from quantum mechanics. Some of the functionals belonging to this group are the exchange Becke86 (B86);[150] the correlation Perdew86 (P86);[163] and the exchange-correlation Perdew-Burke-Ernzerhof (PBE),[164] and the modified Perdew-Burke-Ernzerhof (mPBE)[164,165] functionals.

More recently, a new subclass of GGA methods has been developed including additional information as provided by the Laplacian of the electron density as well as the kinetic energy density. They are know as meta-GGA (M-GGA) methods and some examples are the exchange functional VSXC;[166] and the correlation functionals KCIS[167] and TPSS.[168]

3.3.2.5. Hiper-GGA or hybrid methods. The hybrid or hiper-GGA methods combine the exchange-correlation effects from the GGA methods with some percentage of exact exchange (Hartree-Fock). The simpler combination considered the same proportion of each term as conceived by Becke[169] in 1993, and that is why this is known as the *half-and-half* functional. Thus, the exchange-correlation energy of this functional has the form

$$E_{XC}^{BH} = \frac{1}{2} E_{XC}^{LSDA} + \frac{1}{2} E_{X}^{exact} \tag{3.27}$$

On the other hand, B3LYP[153,170,171] is one of the most used functionals, mainly because it is able to describe many chemical properties with special robustness. Its exchange-correlation energy expression contains three parameters (i.e. a, b and c with values 0.20, 0.72 and 0.81, respectively). These parameters were optimized to reproduce experimental molecular atomization energies and they modulate the exchange-correlation contributions from functionals B88 and LYP as

$$E_{XC}^{B3LYP} = (1-a)E_{X}^{LSDA} + aE_{X}^{exact} + b\Delta E_{X}^{B88} + (1-c)E_{C}^{LSDA} + cE_{C}^{LYP} \tag{3.28}$$

Another functional that was considered in the present thesis is M05-2X.[172] The latest belongs to a bigger family of functionals developed in the group of Prof. Truhlar at the University of Minnesota. M05-2X is a hybrid meta exchange-correlation functional (hybrid meta-GGA) that incorporates electron spin density (ρ), density gradient ($\nabla\rho$), kinetic energy density (τ), and Hartree-Fock (HF) exchange so that the associated exchange-correlation energy $E_{XC}[\rho]$ can be expressed as

$$E_{XC}^{M05-2X}[\rho^{\alpha},\rho^{\beta}] = \int \rho\varepsilon_{X}^{UEG}(\rho)F_{XC}(\rho^{\alpha},\rho^{\beta},\vec{\nabla}\rho^{\alpha},\vec{\nabla}\rho^{\beta},\tau^{\alpha},\tau^{\beta})d\vec{r} \qquad (3.29)$$

where $\rho = \rho^{\alpha} + \rho^{\beta}$ is the total density; $\varepsilon_{X}^{UEG}(\rho) = -(3/4\pi)(3\pi^{2}\rho)$ is the exchange energy per electron of a spin-unpolarized ($\rho^{\alpha} = \rho^{\beta}$) uniform electron gas; and the enhancement factor F_{XC} shows the effects of correlation and inhomogeneity.[173]

Moreover, M05-2X satisfies the *Uniform Electron Gas* limit (UEG). It also exhibits a very good performance for thermochemical kinetics and noncovalent interactions; (i.e., especially weak interaction, hydrogen bonding, $\pi \cdots \pi$ stacking, and interaction energies of nucleobases).[172] Finally, and even more important in our particular case, it describes carbohydrate chemistry well, including the dissociation of the glycosidic bond in sugar phosphates[174,175] and that is why it has been considered as the reference method throughout this thesis.

There are many other functionals available in the literature and they are under constant evolution. However, in some cases they are not universal and provide proper results for only a subset of molecules or a particular chemical property. Therefore, choosing the proper functional is a crucial step when performing DFT calculations.

3.4. QM/MM METHODOLOGY

The size and conformational complexity of systems as biopolymers demands for methods capable of treating up to several 100 000 atoms and allowing for simulations over time scales of tens of nanoseconds. This is only achieved in practice by highly efficient, force-field-based molecular mechanics (MM) methods. On the other hand and as has been mentioned before, Quantum-Mechanics (QM) methods are required for describing chemical reactions and other electronic processes, such as charge transfer or electronic excitation. However, QM methods are restricted nowadays to systems of up to a few hundred atoms. Therefore, combined quantum-mechanics/molecular-mechanics (QM/MM) approaches have become the method of choice for modeling reactions in biomolecular systems[9,176,177] and it is the approach that was used throughout this thesis.

The basic idea of the QM/MM methodology as conceived by Warshel and Levitt[178] is to use a QM method for the chemically active region (e.g., catalytic residues, substrates and co-factors in an enzymatic reaction) and a MM treatment for the surroundings (e.g., rest of the protein and solvent). QM/MM methods enable the modeling of reactive biomolecular systems at a reasonable computational effort while providing the necessary accuracy.

In practice, many current biomolecular QM/MM applications use DFT theory as the QM method owing to its favorable computational-effort/accuracy ratio although semiempirical QM methods remain important, mainly for QM/MM molecular dynamics simulations (Section 3.5.).

3.4.1. Boundary treatment. The treatment of the boundary between the QM and MM partitions is a critical part within the QM/MM methodology and different approaches have been proposed.[179]

The first of these is termed the *"link atom"*[180] approach and was the one considered in the present thesis. In this case a dummy atom is introduced into the quantum system at the location of the boundary between the QM and MM regions, thus satisfying the truncation of the electron density in the QM part. This link atom is usually taken to be a hydrogen atom,[180,181] or a parameterized atom like a one-free valence atom (e.g., *connection atom*,[182] *pseudobond*,[183] and *quantum capping potential*[184] schemes) which involve a parameterized semiempirical Hamiltonian[182] or a parameterized *effective core potential*[183,184] (ECP) adjusted to mimic the properties of the original bond being cut.

The link atom method is straightforward and widely used. However, by definition, a link atom is neither a QM nor an MM atom. It is introduced artificially and, as a consequence, causes ambiguity and makes the definition of the QM/MM energy more complicated (see below).

Moreover, in the original link atom approach the polarization of the bond between the QM frontier atom and the link atom is unphysical due to the nearby point charge on the MM boundary atom. The simplest way to avoid overpolarization is to ignore the first MM charge by setting it to zero. Some improvements have been done to solve this problem by means of eliminating more MM charges (*eliminated-charge* schemes) or adding scaling factors (*scaled-charge* schemes). Another problem of the link atom approach is the introduction of its own coordinates, thus adding extra degrees of freedom. One way to avoid this is to make the coordinates of the link atom dependent on the coordinates of the QM and MM frontier atoms.[185,186]

An alternative to the link atom method started with the frozen orbital approach developed by V. Thery and coworkers[187] (the *"local self-consistent field"* (LSCF)). These methods using local orbitals are theoretically more fundamental than those using link atoms, since they provide a quantum mechanical description of the charge distribution around the QM/MM boundary. A more generalized form of this implementation was presented by Gao et al.,[188] in which a set of hybrid orbitals, called GHO (from *"generalized hybrid orbitals"*), are used at the boundary region. In this approach, a set of four *sp3* hybrid orbitals is assigned to each MM boundary atom. The hybridization scheme is determined by the local geometry of the three MM atoms to which the boundary atom is bonded, and the parametrization is assumed to be transferable. The hybrid orbital that is directed toward the frontier QM atom is called the active orbital whereas the other three are the auxiliary ones. The four hybrid orbitals are included in the QM calculations, but only the active hybrid orbital participates in the SCF procedure.

QM-MM partitioning is commonly fixed, i.e., the boundary between QM and MM regions is defined once-and-for-all at the start of the calculation. However, there are QM/MM schemes that allow the boundary to change during the course of a simulation (*adaptive partitioning*), suitable for processes with shifting active region(s), such as defect propagation in materials or solvated ions. For further information about these approaches see Ref. [189].

Choosing an appropriate boundary is really important and there are some general rules to take into account.[176] The more important ones are that the QM-MM frontier should be as distant from the chemically active region as feasible in terms of computational effort (which is controlled by the size of the QM part) and that the bond being cut should be unpolar and not involved in conjugative interactions (e.g. aliphatic C-C bond).

3.4.2. QM/MM energy expression. In the QM/MM approach a system (*S*) is partitioned into the inner region (*I*) that is treated quantum-mechanically (QM region) and the outer region (*O*) that is described by a force field (MM region). Owing to the strong QM-MM interactions, the total energy of the system cannot simply be written as the sum of the energies of the subsystems and coupling terms have to be considered.[176] There are basically two ways to deal with this; known as *Subtractive* and *Additive* schemes.

Subtractive QM/MM schemes require *1*) an MM calculation on the entire system; *2*) a QM calculation on the inner subsystem; and *3*) an MM calculation on the inner subsystem. The QM/MM energy of the entire system is then obtained by summing (*1*) and (*2*) and subtracting (*3*) to avoid double counting.

$$E_{QM/MM}^{sub} = E_{MM}(S) + E_{QM}(I+L) - E_{MM}(I+L) \tag{3.30}$$

Here the link atom (*L*) approach was considered and thus calculations are performed on a capped inner subsystem (*I* + *L*).

Conceptually, the subtractive QM/MM scheme can be seen as an MM approach where a certain region of space has been cut out and is treated at the QM level. Therefore, its main advantage is simplicity. No explicit QM-MM coupling terms are needed; the standard QM and MM procedures can be used without any modifications and the implementation is thus straightforward. However, a complete set of MM parameters for the inner subsystem is needed and more severely, the coupling between the subsystems is handled entirely at the MM level. This is particularly problematic for the electrostatic interaction, which will typically be represented by the Coulomb interaction between fixed atomic charges in the QM and MM regions.

One of the most popular QM/MM methods based on this approach is ONIOM[190-192] (from *"our n-layered integrated molecular orbital and molecular mechanics"*). However, recent improvements of the ONIOM approach enable the inclusion of MM charges into the QM Hamiltonian (*electrostatic embedding*, see below) and thus take it beyond a strictly subtractive scheme.

On the other hand, the basic energy expression for an additive QM/MM scheme is given in Equation (3.31).

$$E_{QM/MM}^{add} = E_{MM}(O) + E_{QM}(I+L) + E_{QM-MM}(I,O) \qquad (3.31)$$

In contrast to Equation (3.30) the MM calculation is now performed on the outer subsystem only. In addition, it appears an explicit coupling term $E_{QM-MM}(I,O)$, which collects the interaction terms between the two subsystems. The capped inner subsystem (I + L) is treated at the QM level as in the subtractive scheme. The majority of QM/MM schemes presently in use are of the additive type,[176] including the one considered in the present work.

The exact form of the QM-MM coupling term $E_{QM-MM}(I,O)$ defines a particular QM/MM method. It includes bonded, van der Waals and electrostatic interactions, being the last ones the most important and also the most technically difficult ones to deal with.

3.4.2.1. Electrostatic coupling schemes. The electrostatic coupling between the QM charge density and the charge model used in the MM region can be handled at different levels of sophistication, characterized essentially by the extent of mutual polarization and classified accordingly as *mechanical*, *electrostatic*, or *polarized embeddings*.[176]

In the *mechanical embedding* scheme the charge model of the MM method (typically rigid atomic point charges) is simply applied to the QM region as well. The main limitation is that charges in the outer region do not interact with the QM density, which is thus not directly influenced (polarized) by the electrostatic environment.

The major shortcomings of mechanical embedding can be avoided by performing the QM calculation in presence of the MM charge model. For instance, by incorporating the MM point charges as one-electron terms in the QM Hamiltonian, which is then augmented by an additional term that has the form

$$\hat{H}^{el}_{QM-MM} = -\sum_{i}^{N}\sum_{J\in O}^{P}\frac{q_j}{\left|\vec{r}_i - \vec{R}_J\right|} + \sum_{\alpha\in I+L}^{M}\sum_{J\in O}^{P}\frac{q_j Q_\alpha}{\left|\vec{R}_\alpha - \vec{R}_J\right|} \tag{3.32}$$

where q_j are the MM point charges located at $\vec{R}_J$; Q_α are the nuclear charges of the QM atoms at $\vec{R}_\alpha$; and $\vec{r}_i$ designate electron positions. The indices i, J, and α run over the N electrons, P point charges, and M QM nuclei, respectively.

In this approach, namely *electrostatic* or *electronic scheme* the electronic structure of the inner region can adapt to changes in the charge distribution of the environment and is automatically polarized by it. Moreover, no charge model needs to be derived for the inner region, and the QM-MM electrostatic interaction is treated at the QM level, which provides a more advanced and more accurate description than a mechanical embedding. Even if increasing the computational demands, the electrostatic embedding is the most popular embedding scheme in use today, certainly for biomolecular applications.[176]

Even a more sophisticated approach would introduce a flexible MM charge model that is polarized by the QM charge distribution. These *polarized embedding* schemes can be further divided into approaches where the polarizable charge model in the MM region is polarized by the QM electric field but does not itself act back on the QM density; and an ideal fully self-consistent formulation that includes the polarizable MM model into the QM Hamiltonian and therefore allows for mutual polarization. Although some QM/MM implementations have already used polarized-embedding schemes, biomolecular applications have remained scarce and there are yet no generally established polarizable biomolecular force fields.[176]

3.4.2.2. Other non-bonded and bonded QM-MM interactions. Although electrostatic interactions chiefly control the Hamiltonian $\hat{H}_{QM-MM}$, a good choice of van der Waals parameters can be crucial for a correct description of the QM-MM interaction. The van der Waals interaction is typically described by a Lennard-Jones potential (Eq. 3.4a) meaning that suitable parameters are needed for each quantum atom. The most common way to get dispersion parameters for the QM atoms is using those already derived for the MM force field. However, in some cases and in order to improve results, a costly option is to derive some specific Lennard-Jones parameters for the quantum atoms.[193]

The same approach of using basically standard MM parameters applies also to the bonded QM-MM interactions (bond stretching, angle bending, torsional, etc). Once again these parameters can be complemented as necessary with additional bonded terms not covered by the default assignment rules of the force field.

3.5. MOLECULAR DYNAMICS (MD)

Molecular dynamics (MD) as originally conceived within theoretical physics in the late 1950s is applied today in chemical physics, materials science and the modeling of biomolecules. Moreover, MD simulations provide the methodology for performing time-evolution studies of systems at the atomic scale. In the most common version, the trajectories of atoms are determined by solving the Newton's equations of motion for a system of interacting particles over MM or QM/MM potential energy surfaces.

3.5.1. Computational algorithms. In MD simulations the classical dynamics equations (Eq. 3.33) over all the nuclear Cartesian coordinates (*q*) are propagated.

$$-\frac{\delta V}{\delta q_i} = m_i \frac{d^2 q_i}{dt^2} \qquad (3.33)$$

Solving Newton's equation of motion requires a numerical procedure for integrating the differential equations. The most common integration algorithm used in the study of biomolecules has been the *Verlet integrator*,[194] based on two Taylor expansions. However, the Verlet algorithm suffers from several numerical problems. Moreover, the fact that the velocity is not explicitly considered is a problem when generating ensembles at constant temperature. There are several improvements to the Verlet algorithm such as the *leap-frog*[195] or higher order algorithms that already include the velocity in their equations.

3.5.2. Experimental conditions and thermodynamic ensembles. The total energy is a constant when using the equations of motion as specified in Eq. 3.33. In this case, time averages obtained in the MD simulation are equivalent to ensemble averages in a microcanonical ensemble (NVE).

Experiments are usually performed at constant temperature and volume (i.e., the canonical NVT ensemble) or constant pressure and temperature (i.e., the isobaric-isothermal NPT ensemble). In order to run MD simulations in other ensembles different than NVE, thermostats and/or barostats must be introduced. Some of the most commonly used are the Nosé-Hoover thermostat,[196,197] and the Berendsen[198] thermostat/ barostats.

3.5.3. Time step and constraints. Characteristic time scales for protein motions range from the fast atomic fluctuations (10^{-15} s) to some protein transitions (10^{-1} s). Moreover, all these motions tend to be interconnected. Such interdependence implies that it is not possible to study a long-time scale motion without taking into account the short-time ones.

An appropriate time step (Δt) should be small by comparison to the period of the fastest motion in the system being simulated. In the case of biomolecular systems, the highest frequency motions are the bond stretching vibrations but these are generally of minimal interest in the study of their structure and function. Hence, improvements in efficiency are often obtained through freezing the fastest modes of vibration by constraining the bonds involving hydrogen atoms and using algorithms such as SHAKE,[199,200] RATTLE,[201] and LINCS.[202] In addition, constraints in MD simulations can be applied to other interesting areas. It may be used, for example in *Potential of Mean Force* (PMF) calculations to estimate free energy barriers in enzymatic reactions (see below).

3.5.4. Simulated environment. MD can simulate a range of experimental conditions. The earliest protein simulations[203,204] considered the molecules as isolated entities, that is, in vacuum. Later simulations started including explicit water and neighboring protein molecules as in a crystal environment to be more realistic. The so-called *periodic boundary conditions* (PBC) ensure that the system is periodical in all directions to represent an essentially infinite system.

Other times it is not completely necessary to model a periodic system. An example would be the study of localized phenomena such as the reaction mechanism in the active site of an enzyme. In the latter case (of interest for this thesis), the reactant part of the system is enclosed within a shell, usually spherical, and centered in the active site. The atoms included in such a shell are able to move while the outer region forms a barrier that maintains the overall structure of the system.

3.5.5. Potential of Mean Force (PMF). MD simulations have become a powerful and useful tool that can be used to determine macroscopic thermodynamic properties of the system based on the *ergodic hypothesis.*[*] Thus, statistical mechanics provides a rigorous framework of mathematical expressions that relate the distributions and motions of atoms and molecules as obtained in MD simulations to macroscopic observables such as pressure, heat capacity and free energies. The latest is of high interest when studying reaction mechanisms as allows the estimation of the entropic effects that otherwise would be neglected. In these applications, as chemical bonds are being broken/formed, a QM or QM/MM description of the potential energy will be necessary.

In enzymatic reactivity, once the reaction coordinate has been defined, one can determine the associated free energy of activation by computing the *Potential of Mean*

---

[*] Originally due to L. Boltzmann it states that the statistical ensemble averages are equal to time averages of the system.

Force[205] (PMF) along that reaction coordinate through sampling techniques. As well known, the Boltzmann distribution expression indicates that given a potential function it is possible to calculate the probability for all the states of the system. The PMF procedure works in the reverse direction: given an observed distribution of values (from the trajectory), the corresponding effective potential function can be derived.

In order to obtain satisfactory kinetic results, the free energy profile mapped along a reaction coordinate z is required; so, this means computing the PMF as

$$W(T,z) = -RT \ln \rho(T,z) \tag{3.34}$$

where $\rho(T,z)$ is the classical mechanical probability density as a function of the reaction coordinate.

Two methods are generally used in the calculation of PMFs for enzymatic reactions; namely, the *Umbrella Sampling*[205] (US) and the *Free Energy Perturbation*[206] (FEP) techniques. We will focus on US, since it was the one used in the present thesis.

In conventional MD simulations, a straightforward sampling of the reaction path is not possible since the dynamics at ordinary temperatures only very rarely visit the high-energy region near the TS (unless the corresponding free energy barrier is close to zero). In order to achieve a sampling of a specific region of the energy surface with MD methods, the sampling must be biased towards a specific volume of phase space. This is exactly the idea behind the US technique; that is, augmenting the energy surface $V(r)$ with a biasing potential (restraining potential) $U(r)$ as

$$V'(r) = V(r) + U(r) \tag{3.35a}$$

$$U(r) = k_U (r - r_0)^2 \tag{3.35b}$$

where k_U determines the extension of the accessible sampling around r_0.

The ensemble calculated with the augmented potential is non-Boltzmann but the probability distribution is corrected after the simulation to remove the artificial biases. In practice what is done is to perform a series of simulations with biasing potentials located at different positions along the reaction path, and to use the unbiased distribution to determine the corresponding free energy along such reaction path (i.e., PMF).

The main disadvantage of the US technique is that the location of the biasing potential must be selected manually and an a priori knowledge of an approximate reaction coordinate is therefore required.

Finally, in has to be mentioned that when the PMF calculations are carried out using purely classical dynamic equations (its usual implementation) some effects like the quantization of the vibrational motions or quantum nuclear tunneling are neglected. In some systems this omission can lead to significant errors in computed free energy barriers so that some corrections are then desirable.[9]

3.6. ENZYME KINETICS AND THEORETICAL STUDY OF ENZYMATIC REACTIVITY

3.6.1. Enzyme kinetics. Like other catalysts, enzymes do not alter the position of equilibrium between substrates and products. However, unlike other catalyzed chemical reactions, the enzymes display saturation kinetics, that is, their rate of catalysis does not show a linear response and at relatively high substrate concentrations the reaction rate asymptotically approaches the theoretical maximum ($v_{\max}$).

In order to give an explanation to this experimental observation, in 1913 L. Michaelis and M. Menten[207] proposed the following simplified mechanism:

$$\mathrm{E + S} \overset{K_M}{\rightleftarrows} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E + P} \tag{3.36}$$

In this mechanism the enzyme (*E*) and substrate (*S*) reversibly associate to form a non-covalent complex (i.e. enzyme-substrate complex (*ES*)). The latest, also known as the *"Michaelis complex"*, is said to be in thermodynamic equilibrium with the free enzyme and substrate. Accordingly, $K_M = K_S$ is the dissociation constant of the Michaelis complex. K_M is experimentally defined as the substrate concentration at which the reaction rate is at half-maximum ($v_{\max}/2$) and it depends on both the enzyme and the substrate, as well as conditions such as temperature and pH. Moreover, in the context of the Michaelis-Menten mechanism, K_M is an inverse measure of the substrate's affinity for the enzyme (i.e., a small K_M indicates high affinity, meaning that the reaction rate will approach $v_{\max}$ more quickly).

Once the Michaelis complex is formed, the chemical reaction takes place with a first order rate constant k_{cat}, resulting in the product (P) and the free enzyme. k_{cat} is also known as the *"turnover number"* and represents the maximum number of substrate molecules converted to product per enzyme molecule per second.

Based on this mechanism, the Michaelis-Menten kinetic equation is obtained as

$$v = \frac{v_{\max}[S]}{K_M + [S]} \tag{3.37}$$

In this equation, v is the reaction rate at the beginning of the catalytic process and $v_{\max}$ is the maximum theoretical velocity and equal to $k_{cat}[E_0]$, where $[E_0]$ is the total enzyme concentration.

It should be noted that the model proposed by Michaelis and Menten is not valid in many cases, but still equation 3.37 reproduces reasonably well the experimental kinetic results. However, most of the time, k_{cat} and K_M can no longer be considered as the rate or the *ES* dissociation constants as defined before, but are actually a combination of the rate and the equilibrium constants of the different steps of the real mechanism.

If the substrate concentration ($[S]$) is low enough so that the enzyme is considered to be free in solution (i.e. $[E_0] = [E]$), the velocity can be obtained as

$$v = \frac{k_{cat}}{K_M}[E_0][S] \tag{3.38}$$

In this case, the kinetics corresponds to a bimolecular reaction with a rate constant k_{cat}/K_M. This parameter is a measure of how efficiently an enzyme converts a substrate into product. It has a theoretical upper limit of 10^8-10^{10} $(M*s)^{-1}$ and enzymes working close to this, such as fumarase, are termed superefficient.[208]

Moreover, and most commonly, the substrate concentration is much higher than the enzyme concentration, so that all the enzyme is forming the Michaelis complex (i.e. $[E_0] = [ES]$). Accordingly, equation 3.37 can be then expressed as

$$v = v_{\max} = k_{cat}[E_0] \tag{3.39}$$

and we are in the presence of a unimolecular reaction with rate constant k_{cat}.

3.6.2. Theoretical study of enzymatic reactivity. Most of the methodological framework considered in the present thesis has already been addressed. In the next few pages, the application of these approaches in theoretical studies of enzymatic reactivity will be briefly explained.

3.6.2.1. Michaelis complex modeling. Modeling the Michaelis complex (i.e., enzyme + substrate/s + cofactor/s) is usually the first step in theoretical studies on chemical reactions catalyzed by enzymes. The construction of this model is typically based on X-ray crystal structures. Unfortunately, and with the exception of a few rare cases, the native substrates cannot be crystallized all together, as they would undergo rapid chemical transformations. Very often, a substrate analogue or inhibitor is thus present and accordingly, real substrate/s have to be derived from the existing coordinates or even docked in the active site. Another typical problem arises when the protein structure has not been completely solved (e.g. flexible loops) and thereby the missing parts have to been modeled, for example, by homology modeling techniques.

Once these issues have been tackled, MD simulations are commonly performed to relax this initial model and equilibrate the system to its *"hypothetical"* active catalytic form. Thus, the obtained computational outcomes are directly coupled to the structure of the Michaelis complex used as starting point in the reactivity study. The latest implies that the modeling process of the Michaelis complex is an important part of the work that cannot be underestimated.

3.6.2.2. Exploration of the Potential Energy Surface (PES). Once a reliable Michaelis complex is obtained, the next step would be to explore the *potential energy surface* (PES) searching for structures that correspond to the reactants, the products, the reactions intermediates and the saddle points that interconnect them.

In the simplest description, a chemical reaction takes place along the lowest potential energy path connecting the reactants and products and passing over the transition structure as the highest point. This is the *Minimum Energy Path* (MEP), which in mass-weighted coordinates is called the *Intrinsic Reaction Coordinate* (IRC).

The definition of the reaction coordinate (RC) is a critical step in order to build a potential energy profile or a potential energy surface that represents the reaction under study and provides a first guess of the transition state nature. Most works use a simple function of valence coordinates (e.g., geometrical parameters such as a dihedral angle or the difference between the forming and breaking bond lengths). For systems for which little is known about the reaction path, a number of methods have been developed for finding the best description of the reaction coordinate (e.g. *conjugate peak refinement*[209] and *nudged elastic band*[210]).

Once a representative PES has been obtained, that is, potential energy as a function of atomic positions properly describing the reaction under study, several points of interest can be determined. These are global and local minima (which correspond to the most stable nuclear configuration and reactive intermediates, respectively) and the *"saddle points"* or local maxima along the RC. The latest connect two minima (e.g. reactants and products in a chemical reaction) and will also be referred to as transition states in this manuscript. It is expected that the minimum energy structures will be representative of stable chemical species, as well as the potential energy and the structure of a transition state will describe the mechanism and the kinetics of the considered reaction.

There are different methods to optimize the potential energy as a function of the nuclear coordinates. These methods are classified depending on the need of first or second derivatives of the potential energy, which is related with its efficiency and computational cost.

Methods that require up to first derivatives of the potential energy with respect to the nuclear coordinates are mainly the *steepest descent* and the *conjugate gradient* family methods,[211] which are widely used to minimize molecular structures of thousands of atoms. Moreover, when looking for transition state structures, the information of the PES curvature provided by the Hessian matrix (i.e. the matrix of second derivatives of the potential energy) is essential for the success of the search.

In the case of big systems (e.g. those studied in the present thesis), energy and gradient evaluations are expensive and, accordingly, efficient methods are required. An option would be to apply the so-called *micro-iterative method*, which was first used by Maseras and Morokuma.[190] This method splits the system in two parts, a *core* zone where an accurate second order search is done, and an environment that is kept minimized with a cheap first order method. Both processes are carried out until consistency. This is maybe the only strategy that can locate real saddle points in big systems with the direct usage of second derivatives information. Obviously, the control and information given by a Hessian calculated like this will only be referred to the core zone, where the main relevant movements related to the reaction are expected to occur.

Moreover, if the Hessian matrix is mass-weighted and diagonalized, eigenvalues ε_i are obtained and vibrational frequencies can be calculated as

$$\overline{\upsilon}_i = \frac{1}{2\pi c}\sqrt{\varepsilon_i} \tag{3.40}$$

In a minimum all frequency values are real, whereas there is one (an only one) imaginary frequency if the structure corresponds to a transition state.

3.6.2.3. Connecting experimental with theoretical data. The rate constant (k_{cat}) of the second step of the mechanism shown in Eq. 3.36 can be determined experimentally. Within the framework of *"Transition State Theory"* (TST), this rate constant can be used to derive a phenomenological (or effective) Gibbs free energy barrier (for convenience, here we will call it $\Delta G^{\ddagger}$) according to:

$$k_{cat}(T) = \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\ddagger}}{RT} \right) \tag{3.41}$$

where k_B is the Boltzmann constant, T is the temperature and h is the Planck constant.

Strictly speaking, this phenomenological free energy barrier would include differences in potential energies, thermal and entropic contributions, quantum nuclear effects, recrossing and even the contribution to the rate from different starting protein:substrate complexes (local and global conformational changes).

The theoretical estimation of a comparable, accurate free energy magnitude is very time consuming and not always required. In practice, free energy barriers are usually determined by using molecular-dynamics based approaches (e.g. umbrella sampling) combined with QM/MM methods for the description of the potential. When accurate electronic structure methods (e.g. DFT methods) are used for the QM part, the simulations are very expensive and do not allow for much conformational sampling. A reasonable alternative would be considering semiempirical methods to treat the QM part, but sometimes there are no parameters for all the atoms types involved or they are not very accurate and fail to reproduce the reaction under study. Correcting schemes (namely dual-level approaches) can be applied in these cases to improve the energetics. It is important to notice that, even with semiempirical methods, the typical simulation times used in most applications only allow for local conformational sampling.

On the other hand, in systems where the enthalpic and entropic contributions to the free energy barrier are not significant, and/or when an accurate description of the electronic structure is more advantageous, a generally used approximation consists in performing energy optimizations and thus estimate potential energy barriers ($\Delta V^{\ddagger}$) instead of free energy barriers ($\Delta G^{\ddagger}$). A qualitative comparison with the experimentally-derived phenomenological free energy barrier can then still be done.

PART IV

Results and Discussion

4

Lipopolysaccharyl-α-1,4-galactosyltransferase C (LgtC)

4.1. INTRODUCTION

Lipopolysaccharyl-α-1,4-galactosyltransferase C (EC 2.4.1.44, LgtC) from *Neisseria meningitides* (a member of family 8 of GTs; see Figure 1.4)[212] plays a key role in the biosynthesis of the oligosaccharide part of lipooligosaccharides (LPS) structures in this microorganism. *Neisseria meningitidis* is a bacterium that lives within the nasopharyngeal tract of humans and can become a human-specific pathogen causing invasive, life-threatening infections, such as meningitis and septicemia.[213] Its great virulence is related to the presence of LPSs on its cell surface that can mimic those of human glycolipids, thus avoiding the recognition by the human immune system. Therefore, LgtC represents a very attractive therapeutic target for new antibiotics.[214]

LgtC transfers an α-galactose from uridine 5'-diphospho-α-galactose (UDP-Gal) to a galactose of the terminal lactose (LAT) on the bacterial LPS to yield an elongated oligosaccharide with an overall retention of stereochemistry at the anomeric carbon atom (See Figure 4.1) and with an α-1,4 specificity of the new glycosidic bond. This

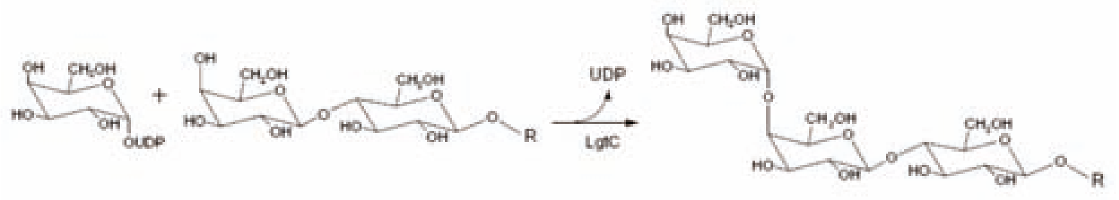

Figure 4.1. Schematic representation of the galactosyl transfer from UDP-Gal to the LPS core oligosaccharide of *N. meningitidis* catalyzed by the retaining GT LgtC. Scheme 2 from Ref. [3]

retaining glycosyltransferase presents a GT-A fold and the typical structural motif Asp-X-Asp (DXD) implicated in the binding of the donor sugar substrate and the coordination of a divalent cation. LgtC was the first retaining nucleotide sugar dependent transferase with solved structure, and the only one so far of any type with structural data for both the donor and acceptor sugars (substrate analogs) in the same X-ray structure (ternary complex).[61]

The structure shows a glutamine residue (Gln189) that could act as a putative nucleophile in a double-displacement mechanism, which its oxygen at 3.5 Å of the anomeric carbon. However, the nucleophilic character of the amide of Gln189 should be rather poor and experimentally, both the Q189A and Q189E LgtC mutants display 3% residual transferase activity, indicating the limited relevance of Gln189 as catalytic nucleophile.[61,86] Therefore, an alternative mechanism would seems more likely.

We present in this chapter the first full-enzyme hybrid quantum mechanical/ molecular mechanical (QM/MM) study of the catalytic mechanism of a retaining glycosyltransferase with the GT-A fold, using LgtC as a model. The different mechanistic alternatives outlined before (See Chapter 1 Section 1.2.3.2) are explored and the role of the most relevant active-site residues supporting the catalytic mechanism of LgtC is analysed. In addition, the effect of different mutations in the enzyme and of modifications in the substrates is considered.

4.2. MODELS AND METHODS

Initial coordinates for the wild type LgtC enzyme were taken from the X-ray structure (PDB Code: 1GA8,[61] resolution 2.0 Å). The protonation states of the titratable residues (His, Glu, Asp, Arg, Lys) were chosen based on the pKa values given by the empirical PROPKA procedure[215] and verified through visual inspection. The substrate analogues present in the crystal structure (UDP 2-deoxy-2'-fluoro-galactose and 4-deoxylactose) were transformed manually into the original substrates, UDP-galactose (UDP-Gal) and lactose (LAT), respectively. Finally, the original Mn^{2+} ion in PDB structure was modeled by Mg^{2+} (See below for further discussion).

A partial solvation scheme was used to solvate the region of 24 Å around the anomeric center by overlaying a water sphere on the enzyme. The solvated system was first relaxed by performing energy minimizations at the MM level using the CHARMM22 force field[114,216] as implemented in the CHARMM program.[113,217] For the sugar moieties, the topology and parameters from the CHARMM force field for carbohydrates were used, including the recently released ones for glycosidic linkages between hexopyranoses.[218] The solute atoms were initially frozen for 10000 conjugate gradient optimization steps. In the subsequent minimizations, the restraints on the protein and ligand atoms were gradually released. The prepared system (see Figure 4.2A) contained 6728 atoms, including 755 TIP3P water molecules, and served as starting point for the QM/MM calculations.

The QM/MM calculations were done with the modular program package ChemShell[219] using TURBOMOLE,[220] Gaussian03[221] or MNDO2005[222] to obtain the QM energies and gradients at the DFT (BP86,[149,223-225] B3LYP[153,170,171] and M05-2X[172] functionals) or SCC-DFTB[131] levels, respectively. MM energies and gradients were evaluated by DL_POLY,[226] which was accessed through the ChemShell package, using the CHARMM force field. An electronic embedding scheme[227] was adopted in the QM/MM calculations with the MM point charges being incorporated into the one-electron Hamiltonian during the QM calculation. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. Seven hydrogen link atoms were employed to treat the QM/MM boundary with the charge shift model.[228,229]

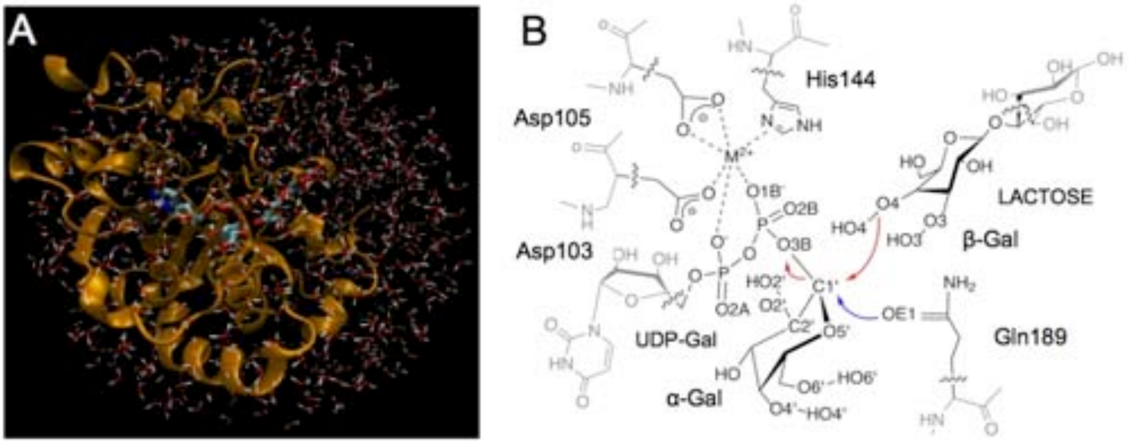

Figure 4.2. System overview (A) and (B) active site representation showing the QM/MM partition considered in the present work. In (B), QM (MM) atoms are depicted in black (grey). The boundary between the QM and MM regions is indicated by wavy lines. Distances considered to model the first step of the double displacement mechanism and the front-side single displacement mechanism are shown with blue and red arrows, respectively. Atoms mentioned in the text are labeled.

All residues and water molecules within 12 Å of the anomeric center were included in the optimization process (1225 atoms) as the active region while the remaining atoms were kept fixed. The QM region incorporated 101 atoms: those from the α and β galactose rings (from UDP-Gal and LAT, respectively), Mg^{2+} and its first coordination sphere (phosphate groups from UDP and the side chains of residues Asp103, Asp105 and His244), as well as the side chain of Gln189 (see Figure 4.2B). The total charge of the QM region was -2. Reaction paths were scanned by performing constrained optimizations along properly defined reaction coordinates in steps of 0.2 Å. This provided starting structures for subsequent full optimization of all relevant stationary points, employing the low-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS)[230,231] algorithm in the case of minimizations and the microiterative optimizer combining both the partitioned rational function optimizer (P-RFO)[232,233] and L-BFGS during the transition state search. All these algorithms are implemented in the HDLCopt[234] module of ChemShell. Frequency calculations on the QM region confirmed that all reported transition states are characterized by a single imaginary frequency and a suitable transition vector that corresponds to the investigated reaction.

We ensured by intrinsic reaction coordinate (IRC) calculations and visual inspection of the optimized structures that the computed stationary points are connected by continuous pathways.

Geometry optimizations were generally carried out at the QM(BP86/SVP[235])/ CHARMM level in combination with the resolution of the identity (RI) approximation.[236,237] For more accurate energy evaluations, we performed single-point energy calculations with other functionals and larger basis sets: B3LYP, M05-2X, and DFT-D56 methods; TZVP[238] and def2-TZVPP(d)[239] basis sets. Natural population analysis (NPA)[240] charges were determined from QM/MM calculations with QM = B3LYP/TZVP and BP86/TZVP.

Umbrella sampling at the SCC-DFTB/CHARMM22 level was performed to compute the potential of mean force (PMF) and the free energy profile for the front-side mechanism using the dynamics module within ChemShell. The reaction coordinate was scanned in steps of 0.1 Å, with a force constant of 237 kcal mol^{-1} $Å^{-2}$. The same active region as in QM/MM optimizations (see above) was subjected to NVT dynamics with the Nosé-Hoover thermostat.[196,197] The SHAKE procedure[199] was applied at every step for the O-H bonds in the water molecules. A 20 ps molecular dynamics (MD) run for equilibration was followed by 5 ps of data collection for every sampling window. Both the umbrella integration analysis[241,242] and the weighted histogram analysis method[243,244] (WHAM) were used to compute the free energy profile.

The contribution of different residues to the QM/MM energy in the front-side attack mechanism was examined by setting their point charges to zero in additional energy calculations along the QM(BP86/SVP)/CHARMM22 reaction path. In the case of Gln189, extra QM/MM calculations were performed in which Gln189 was assigned the charges of a glutamate.

The program VMD[245] was used to generate the *in silico* mutants of the enzyme and the modified substrates as well as the drawings showing molecular structures.

4.3. RESULTS AND DISCUSSION

4.3.1. Catalytic mechanism. We started by simulating the mechanistic alternatives depicted in Figure 1.6; that is, a double displacement with Gln189 being the nucleophile and a front-side attack mechanism.

4.3.1.1. Front-side attack mechanism. Defining a reaction coordinate (RC) to properly describe a catalytic mechanism is a critical step in reactivity studies. Some information of the system and chemical intuition is needed but even so, many tests should be frequently performed before finding the most appropriate RC.

The simplest RC to consider for modeling a front side attack mechanism would be $d(\mathrm{O4}_{\beta\text{-Gal}}\text{–C1'}_{\alpha\text{-Gal}})$, which describes the nucleophilic attack of the incoming hydroxyl group from LAT on the anomeric carbon (See Figure 4.2B). However, an energy profile with a very high barrier is obtained (See Appendix (Appx.) Figure A4.1A) and more complex RCs need to be considered.

On the other hand, the use of the RC = $d(\mathrm{O3B}_{\mathrm{UDP}}\text{–C1'}_{\alpha\text{-Gal}})$ - $d(\mathrm{O4}_{\beta\text{-Gal}}\text{–C1'}_{\alpha\text{-Gal}})$ (i.e. to combine bond breakage and bond forming processes in the RC) to model the front-side attack of LAT on UDP-Gal results in a smooth energy profile with a single energy barrier of ~12 kcal/ mol at the QM(B3LYP/TZVP//BP86/SVP)/CHARMM22 level (see Figure 4.3). The changes in the distances between the reactive atoms along this reaction path indicate that the computed $\mathrm{S_Ni}$ mechanism has a highly dissociative character: the $\mathrm{O3B}_{\mathrm{UDP}}\text{–C1'}_{\alpha\text{-Gal}}$ bond breaks early in the reaction, the attacking O4 atom from LAT slowly approaches C1′, and both the $\mathrm{HO4}_{\beta\text{-Gal}}\text{–O3B}_{\mathrm{UDP}}$ and $\mathrm{O4}_{\beta\text{-Gal}}\text{–C1'}_{\alpha\text{-Gal}}$ bonds form simultaneously and concomitant with the rapid drop in energy that is observed right after the maximum in Figure 4.3. It should be noted that, in the reactant complex, the O3B oxygen of the phosphate leaving group is already well oriented to act as the base to deprotonates the acceptor in the $\mathrm{S_Ni}$ mechanism ($d(\mathrm{HO4}_{\beta\text{-Gal}}\text{–O3B}_{\mathrm{UDP}})$ = 1.67 Å). This possible role of the leaving group as the base catalyst has initially been invoked by Sinnott and Jecks[246] for the solvolysis of glycosyl fluoride by trifluoroethanol and has recently also been suggested for OtsA.[247] In the cluster model by Tvaroška, UDP was also found to accept the $\mathrm{HO4}_{\beta\text{-Gal}}$ proton, but in the reactant

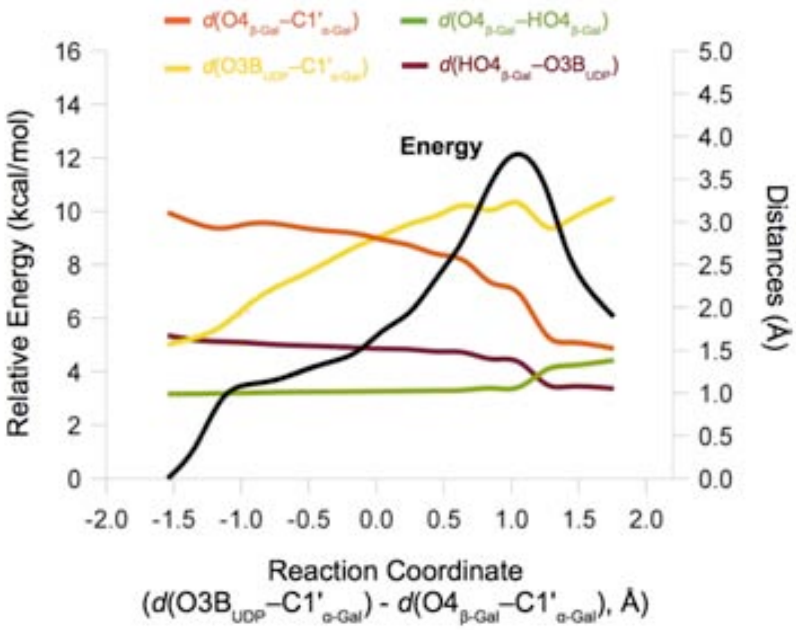

Figure 4.3. QM(B3LYP/TZVP//BP86/SVP)/CHARMM energy profile for the S_Ni mechanism. The variation of several interatomic distances involved in the reaction is also depicted.

state, the $HO4_{β\text{-Gal}}–O3B_{UDP}$ distance was significantly longer than in the present work ($d(HO4_{β\text{-Gal}}–O3B_{UDP}) = 2.33$ Å).[3] The distances at the energy maximum of Figure 4.3 are $d(O3B_{UDP}–C1'_{α\text{-Gal}}) = 3.22$ Å, $d(O4_{β\text{-Gal}}–C1'_{α\text{-Gal}}) = 2.16$ Å, $d(HO4_{β\text{-Gal}}–O4_{β\text{-Gal}}) = 1.07$ Å, and $d(HO4_{β\text{-Gal}}–O3B_{UDP}) = 1.36$ Å, clearly indicating a highly dissociative character. This rationalizes the finding that the use of a single distance $d(O4_{β\text{-Gal}}–C1'_{α\text{-Gal}})$ as reaction coordinate for this chemical event did not work satisfactorily (Appx., Figure A4.1A).

The structure of the energy maximum in Figure 4.3 was used as the starting point for a transition state (TS) search. The computed energy barriers and reaction energies are given in Table 4.1. Key bond distances and NPA charges of reactants, TS, and products are listed in Table 4.2. The energy barriers obtained range between 8.1 and 16.6 kcal/mol (BP86/TZVP and M05-2X/SVP single-point energies, respectively). The computed reaction energies vary less (from 2.8 kcal/mol at the B3LYP-D/TZVP level to 6.1 kcal/mol at the B3LYP/TZVP and B3LYP/def2-TZVPP(d) levels). Similar geometries for R, TS and P are obtained at the BP86/SVP and B3LYP/SVP levels, with some small differences for the TS: at the B3LYP/SVP level the breaking $C1'_{α\text{-Gal}}–$

$O3B_{UDP}$ bond is slightly longer, and the forming $O4_{\beta\text{-Gal}}$–$C1'_{\alpha\text{-Gal}}$ bond is slightly shorter. Interestingly, the distances obtained for the optimized TS structure [$d(O3B_{UDP}$–$C1'_{\alpha\text{-Gal}})$ = 3.12 Å, $d(O4_{\beta\text{-Gal}}$–$C1'_{\alpha\text{-Gal}})$ = 2.18 Å, $d(HO4_{\beta\text{-Gal}}$–$O4_{\beta\text{-Gal}})$ = 1.07 Å, and $d(HO4_{\beta\text{-Gal}}$–$O3B_{UDP})$ = 1.36 Å] are very similar to those reported for the energy maximum in Figure 4.3, confirming that the reaction coordinate used to describe the reactive process is adequate. Stronger deviations are found when the QM region is treated at the SCC-DFTB level, which gives a high energy barrier of 32.7 kcal/mol and a reaction energy of 9.7 kcal/mol. The transition state at this level of theory is still of dissociative character but with shorter $d(C1'_{\alpha\text{-Gal}}$–$O3B_{UDP})$ = 2.63 Å and longer $d(O4_{\beta\text{-Gal}}$–$C1'_{\alpha\text{-Gal}})$ = 2.44 Å distances than the other two methods used for optimization in this work (See Appx., Table A4.1). These discrepancies are most likely related to the difficulty of the SCC-DFTB method to correctly describe chemical processes involving phosphorous. When comparing the different DFT methods for a given basis set, the barrier chemical heights increase in the order BP86 < B3LYP < M05-2X as has been observed before because of the Hartree-Fock exchange that is present in B3LYP and M05-2X but not in BP86. The M05-2X relative energies are higher at the TS and also along the reaction pathway (See Appx., Figure A4.2A), in agreement with the tendency observed between different DFT methods when describing the breakage of the glycosidic bond in sugar phosphates.[248] For a given method, the extension from the SVP to the TZVP basis set reduces the energy barrier, whereas the inclusion of more polarization and of diffuse functions in the def2-TVZPP(d) basis set increases the barrier height again.

As stated before in Chapter 3 Section 3.3.2, M05-2X is a functional that describes particularly well the dissociation of the glycosidic bond in sugar phosphates. However, and for efficiency, we decided to perform M05-2X/TZVP single-point QM/CHARMM energies calculations at geometries optimized with QM = BP86/SVP. The latest is expected to be the most reliable energy data presented in the present thesis.

The energy barrier obtained with single point energy calculations using M05-2X/TZVP is of 14.6 kcal/mol, which slightly underestimate the phenomenological free energy barrier of ~16 kcal/mol (derived from the experimental k_{cat} values of 14-34 s^{-1} at 303 K),[61,86,90] but it shows that a front-side attack or S_Ni mechanism is plausible and consistent with the experimental data.

Table 4.1. QM/MM potential energy barrier ($V^{\ddagger}$) and reaction energy (ΔV_R) (in kcal/mol) for the proposed S_Ni mechanism at different levels of theory.

QM treatment	**$V^{\ddagger}$**	**ΔV_R**
BP86/SVP	9.2	3.3
BP86/TZVP//BP86/SVP	8.1	4.6
BP86/def2-TZVPP(d)//BP86/SVP	8.8	4.6
B3LYP/SVP// BP86/SVP	12.9	4.6
B3LYP/SVP	11.1	4.6
B3LYP/TZVP//BP86/SVP	11.8	6.1
B3LYP-D/TZVP//BP86/SVP	9.5	2.8
B3LYP/def2-TZVPP(d)//BP86/SVP	12.5	6.1
M05-2X/SVP//BP86/SVP	16.6	3.6
M05-2X/TZVP//BP86/SVP	14.6	4.6
SCC-DFTB	32.7	9.7
B3LYP/TZVP//SCC-DFTB	15.4	10.3
M05-2X/TZVP//SCC-DFTB	17.2	10.5

Table 4.2. Selected QM/CHARMM22 bond distances *d* (Å) and atomic charges *q* (a.u.) in the optimized reactants, transition state, and products for the front-side single displacement mechanism, with QM = BP86/SVP(B3LYP/SVP) for the distances and QM = B3LYP/TZVP// BP86/SVP for the charges.

	Reactants	**TS**	**Products**
$d(\mathrm{O3B_{UDP}–C1'_{\alpha\text{-}Gal}})$	1.57(1.53)	3.12(3.19)	3.28(3.30)
$d(\mathrm{O4_{\beta\text{-}Gal}–C1'_{\alpha\text{-}Gal}})$	3.11(3.14)	2.18(2.06)	1.52(1.49)
$d(\mathrm{HO4_{\beta\text{-}Gal}–O4_{\beta\text{-}Gal}})$	0.99(0.97)	1.07(1.06)	1.38(1.44)
$d(\mathrm{HO4_{\beta\text{-}Gal}–O3B_{UDP}})$	1.67(1.71)	1.36(1.36)	1.06(1.02)
$d(\mathrm{OE1_{Q189}–C1'_{\alpha\text{-}Gal}})$	3.38(3.40)	2.87(2.89)	3.05(3.09)
$d(\mathrm{C1'_{\alpha\text{-}Gal}–O5'_{\alpha\text{-}Gal}})$	1.36(1.36)	1.29(1.29)	1.38(1.38)
$q(\mathrm{C1'_{\alpha\text{-}Gal}})$	0.38	0.49	0.34
$q(\mathrm{O5'_{\alpha\text{-}Gal}})$	-0.47	-0.41	-0.50

In Figure 4.4, the TS structure calculated at QM(BP86/SVP)/MM(CHARMM22) is depicted. The sugar ring goes from the initial distorted $^4C_1$ chair (φ = 236°, θ = 9°) in the reactant complex to a conformation between an ^4E envelope and a $^4H_5$ half-chair at the TS (φ = 250°, θ = 41°). The oxocarbenium nature of the TS is also reflected in a $C1'_{\alpha\text{-Gal}}$–$O5'_{\alpha\text{-Gal}}$ distance that is slightly shorter than in the reactant (by 0.07 Å) and in the charge development during the reaction (Table 4.2). When going from reactant to TS, the charge of the α-Gal moiety increases by $\Delta q(\alpha\text{-Gal}) = 0.30$ a.u., with the main contributions coming from the ring atoms $C1'_{\alpha\text{-Gal}}$ and $O5'_{\alpha\text{-Gal}}$, $\Delta q(C1') = 0.11$ a.u. and $\Delta q(O5') = 0.06$ a.u.. This is accompanied by an increase in the negative charge of the UDP moiety ($\Delta q(\text{UDP}) = -0.36$ a.u.) dominated by the large change at the leaving oxygen atom ($\Delta q(O3B_{\text{UDP}}) = -0.24$ a.u.), and by a smaller change in the lactose, mainly at the attacking oxygen atom ($\Delta q(\beta\text{-Gal}) = 0.09$ a.u., $\Delta q(O4_{\beta\text{-Gal}}) = 0.05$ a.u.).

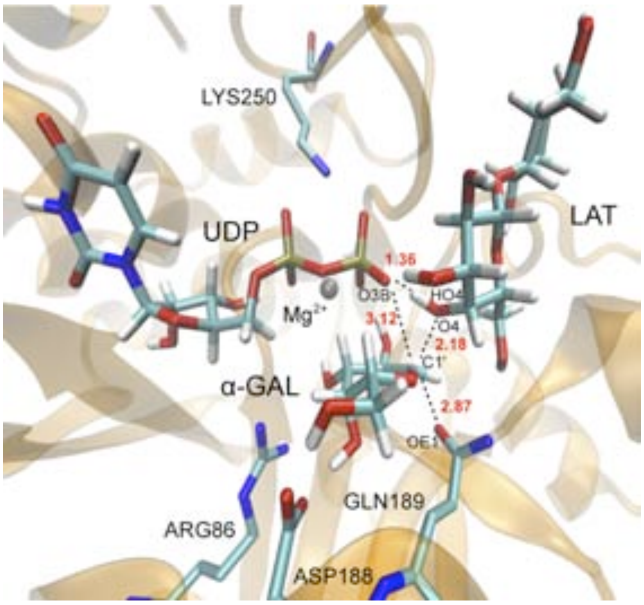

Figure 4.4. QM(BP86/SVP)/CHARMM22 optimized transition state for the S_Ni mechanism. The donor and acceptor substrates, together with some relevant residues, are represented as sticks. Selected distances (in Å) are indicated in red.

Our results for LgtC thus indicate that there is a direct one-step path connecting reactant, TS, and products, which corresponds to a highly dissociative S_Ni mechanism (Figure 1.6B). In the computational work of Tvaroška[3] on a cluster model of LgtC, an

S_Ni mechanism was also suggested although some structural differences with the present study should be highlighted: the gas-phase cluster model predicts an earlier TS with shorter $d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}}) = 2.66$ Å and longer $d(\mathrm{O4_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}}) = 2.34$ Å distances, and in the reactant complex, the $d(\mathrm{O4_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}}) = 3.50$ Å and $d(\mathrm{HO4_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}}) = 2.33$ Å distances are much longer than those obtained here, probably due to a different donor-acceptor orientation because of the missing enzyme environment.

Notice from Table 4.2 that during the front-side attack of LAT on UDP-Gal, the anomeric carbon and the OE1 atom of Gln189 (the putative nucleophile in a double displacement mechanism) get closer by ~0.5 Å. This is mainly caused by the change in the ring puckering on the way from the reactants to the TS (see above). The decrease in the $d(\mathrm{C1'_{\alpha\text{-}Gal}}\text{–}\mathrm{OE1_{Q189}})$ distance, from 3.38 to 2.87 Å, may help to stabilize the increasing positive charge at the anomeric center. Like in the cluster model by Tvaroška,[3] Gln189 (via Nε) forms a hydrogen bond with the O6 atom of β-Gal of LAT, with $d(\mathrm{H}\text{–}\mathrm{O6})$ distances of 2.0 and 1.9 Å in the reactants and the TS, respectively. This interaction is thus involved both in the binding of LAT and in keeping the proper orientation of the substrate during the reaction. The previously reported[3] interaction between $\mathrm{N\varepsilon_{Q189}}$ and $\mathrm{O5'_{\alpha\text{-}Gal}}$ is not seen in our QM/MM calculations where $\mathrm{N\varepsilon_{Q189}}$ is permanently hydrogen bonded to the hydroxyl group of Tyr151. Again, this difference may arise from the limitations of the cluster model that does not include Tyr151 and therefore we are more confident in our results.

4.3.1.1.1. Free energy calculations. Umbrella sampling molecular dynamics simulations at the SCC-DFTB/CHARMM22 level were used to estimate the free energy barrier of the front-side mechanism. Although this electronic structure method severely overestimates the energy barrier (Table 4.1), it is still expected to provide a reasonable estimate of the differences between the potential energy and free energy profiles. These and the corresponding barriers are found to be practically identical, indicating that entropic effects are minor in the reaction under study (Figure 4.5). This is consistent with the highly dissociative character of the TS for the reaction catalysed by ret-GTs. Based on the previous results and for effciency, we decided hereafter to avoid performing systematic free energy calculations.

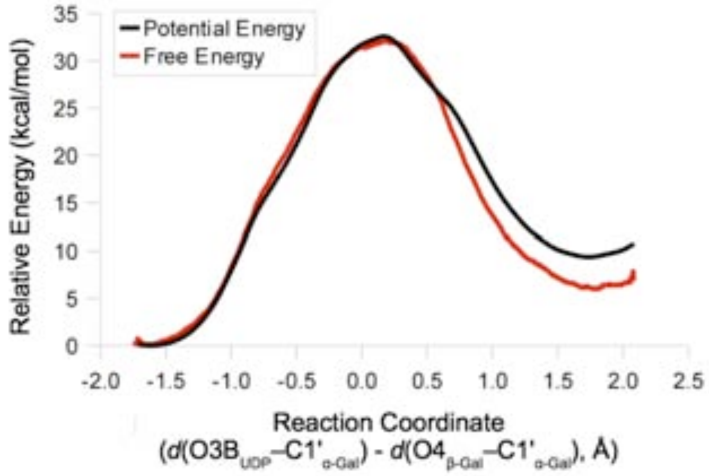

Figure 4.5. SCC-DFB/CHARMM22 potential energy profile and potential of mean force (PMF) for the front-side attack mechanism.

4.3.1.2. Double-displacement mechanism. All attempts to locate the covalent glycosyl-enzyme complex (CGE) for the wild-type enzyme failed. The $d(\mathrm{OE1_{Q189}–C1'_{\alpha\text{-}Gal}})$ distance would seem to be a natural reaction coordinate for driving the reactants toward the CGE intermediate. A corresponding reaction path calculation indeed yields a CGE-type structure, with $d(\mathrm{OE1_{Q189}–C1'_{\alpha\text{-}Gal}}) = 1.58$ Å and $d(\mathrm{O3B_{UDP}–C1'_{\alpha\text{-}Gal}}) = 3.27$ Å. However, the energy profile is monotonously increasing (Appx., Figure A4.1B), and an unrestrained minimization of the last point leads back to the reactants.

To illustrate the different mechanistic proposals for LgtC we built a two-dimensional potential energy surface (PES) (Figure 4.6). At fixed values of the $d(\mathrm{OE1_{Q189}–C1'_{\alpha\text{-}Gal}})$ distance, the QM/MM energy was computed by constrained optimizations along the RC = $d(\mathrm{O3B_{UDP}–C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{O4_{\beta\text{-}Gal}–C1'_{\alpha\text{-}Gal}})$; note that the x-axis in Figure 4.6 is just $d(\mathrm{O4_{\beta\text{-}Gal}–C1'_{\alpha\text{-}Gal}})$. Obviously, the only energy minima in Figure 4.6 are those for the reactants (R, top-right corner) and for the products (P, top-left corner). There is no minimum in the region of the putative intermediate (CGE, bottom-right corner), and hence no evidence for the proposed double-displacement mechanism in the wild-type enzyme that would require CGE formation by nucleophilic attack of Gln189. By contrast, one can easily identify a path connecting reactants and products via an oxocarbenium ion-like transition state. The corresponding TS is of dissociative character and belongs to a one-step S_Ni reaction.

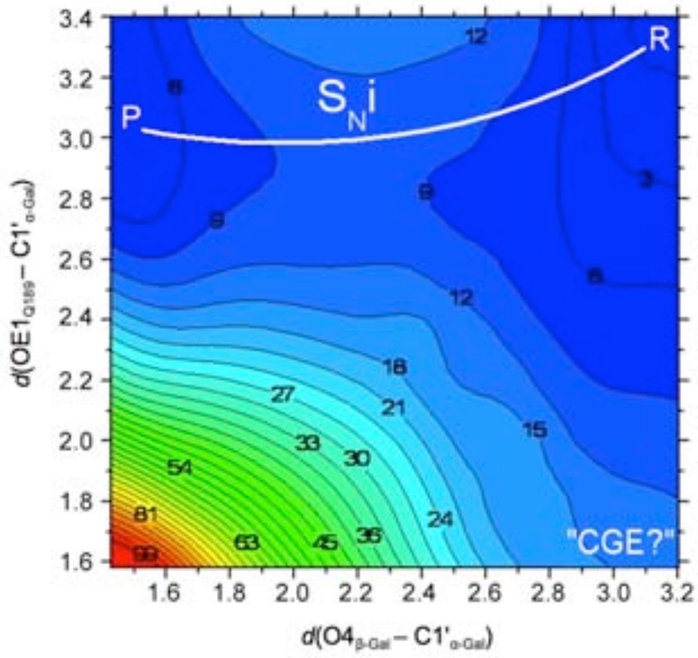

Figure 4.6. Two-dimensional QM(BP86/SVP)/CHARMM22 potential energy surface. Energies are given in kcal/mol and distances in Å. Contour lines are drawn in intervals of 3 kcal/mol.

4.3.2. Modeling Mn^{2+} as Mg^{2+}. As was indicated in the methods section we use Mg^{2+} to model the manganese ion present in the PDB structure (PDB Code 1GA8). Considering the original Mn^{2+} is less convenient from the computational point of view not only because of the increased number of electrons (i.e. 23 *vs* 10 electrons for Mg^{2+}) but also because it would imply open-shell calculations where different spin states should be considered (e.g. triplet, quintuplet).

In order to ensure that the replacement used in this work is valid and does not compromise the mechanistic conclusions previously drawn, we performed test calculations with Mn^{2+} in place of Mg^{2+}. As can be seen in Figure 4.7 the potential energy profiles for both mechanisms remain essentially the same.

The latest outcomes seem logical since in ret-GTs with the GT-A fold, the divalent cation is known to act as a Lewis acid that facilitates the departure of the leaving group,[10] and in that sense, is known that both Mn^{2+} and Mg^{2+} can adopt this role in a similar manner.[249] Fortunately, and according to previous theoretical works, both cations behave remarkably similar; they exhibit the same coordination preferences.[250] Moreover, both divalent ions have given almost the same results in a thorough

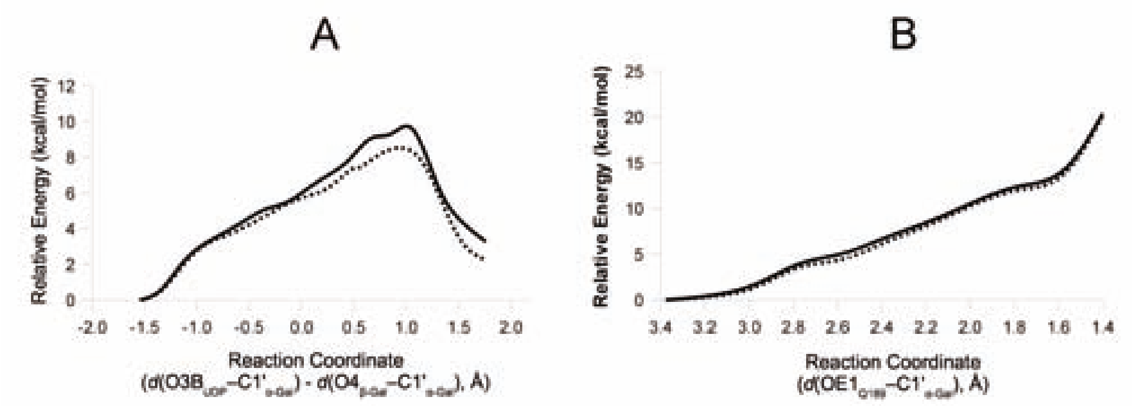

Figure 4.7. QM(BP86/SVP)/CHARMM potential energy profiles for (**A**) the front-side attack (S_Ni) mechanism and (**B**) the first step of the double displacement mechanism with Mg^{2+} and Mn^{2+} as divalent cation (solid and dotted lines, respectively).

validation study[248] of α-glycosidic bond dissociation in sugar phosphates derived from the crystal structure of LgtC; for example, the M05-2X energies (bond lengths) differ between the studied Mg^{2+} and Mn^{2+} complexes by less than 0.2 kcal/mol (0.01 Å).[248]

Finally, this assumption was also made in a previous work[3] on ret-GT and as our own results show, this can be considered as a well-justified choice that we have hereafter made while studying the catalytic mechanism of ret-GTs.

4.3.3. Analysis of factors contributing to catalysis. The QM/MM results presented so far indicate that the front-side attack of LAT at UDP-Gal (S_Ni mechanism) is preferred and can proceed at a reasonable energetic cost in the LgtC active site. The next goal will be to identify which are the factors stabilizing the TS and thus responsible for the catalytic efficiency exhibited by the enzyme.

4.3.3.1. Enzyme-substrates interactions; key enzyme residues. The TS structure in Figure 4.4 indicates several interactions between the enzyme and the substrates that may contribute to TS and/or product stabilization and thus facilitate the reaction via this mechanism. The contribution of individual residues to the (electrostatic) stabilization/ destabilization of the system was estimated by charge deletion analysis (see Section 4.2). The results for the most relevant residues are shown in Figure 4.8.

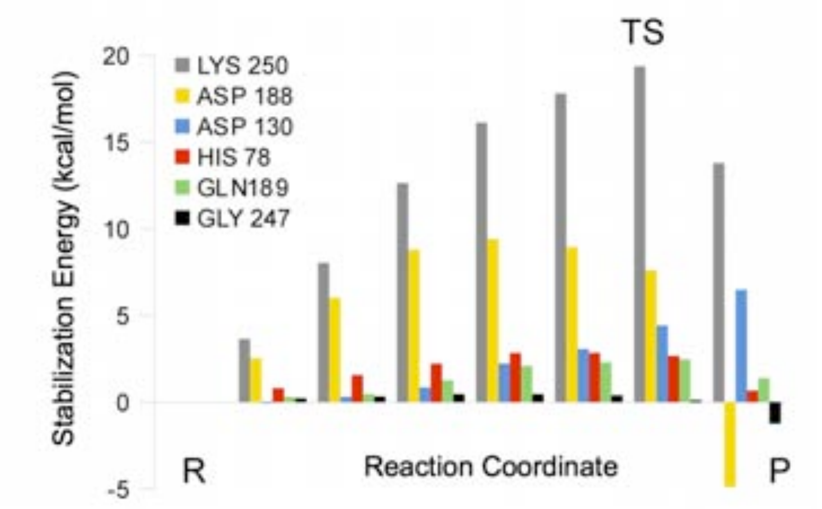

Figure 4.8. Electrostatic contribution to the stabilization of the QM region by selected residues of LgtC surrounding the substrates. The reaction coordinate corresponds to the front-side attack mechanism (RC = $d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{O4_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$) [QM(BP86/SVP)/CHARMM22 data].

As can be seen, the most prominent contribution comes from Lys250, with ~19 kcal/mol of TS stabilization compared to the reactants. This strong effect is caused by the interaction of the charged Lys side chain with the α and β phosphates of the leaving UDP: in fact, a new H-bond involving $\mathrm{O2B_{UDP}}$ appears at the TS (H···O2B distance: 2.98 Å in the reactants and 2.18 Å in the TS). On the other hand, the H-bond between Lys250 and $\mathrm{O2A_{UDP}}$ lengthens from 1.67 Å in the reactants to 1.80 Å in the TS. Lys250 is highly conserved in the GT8 family and belongs to one of the two loops that fold over the donor substrate and are thought to be disordered in its absence;[61] hence, Lys250 is also crucial for substrate binding. The H-bond of the β phosphate (O2B) with Gly247 is very weak, while that with His78 provides a stabilization of 2.5 kcal/mol around the TS (and up to 3 kcal/mol earlier). The electrostatic interactions involving Asp188 (H-bonded to O4′ and O6′ of α-Gal) are most important (up to 9 kcal/mol) at an early stage of the reaction, while those involving Asp130 (H-bonded to O6 of β-Gal of lactose) become more prominent as the reaction proceeds (6 kcal/mol around the TS). Although some residues like Asp188 and Lys250 exhibit strong stabilization effects, their inclusion in the QM region results in negligible differences when comparing the potential energy profiles (Appx., Figure A4.3). These results confirm that the QM/MM partition used in this work satisfactorily describes such enzyme-substrates interactions and that the effect of key residues like Asp188 and Lys250 is properly described despite not being included in the QM partition.

Also notice from Figure 4.8 that Gln189, the putative nucleophile in a double displacement mechanism, provides very little stabilization in the reactants, which increases up to 2.5 kcal/mol on the route to the TS, where it amounts to ~2 kcal/mol before it drops off again. This stabilization mostly comes from the interaction of the anomeric center C1′ and the OE1 atom of Gln189 and is correlated with the changes in the $OE1_{Q189}$–$C1'_{\alpha\text{-Gal}}$ distance (see Table 4.2). The transient decrease of this distance helps to stabilize the charge development at $C1'_{\alpha\text{-Gal}}$ during the reaction: this charge shows a quick initial rise and then levels off before decreasing again around the TS (Figure 4.9A).

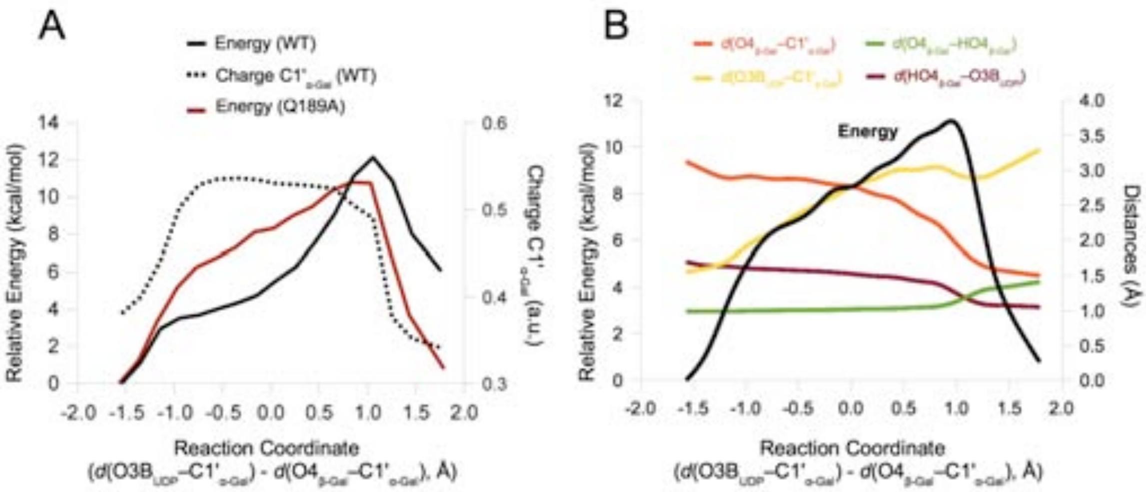

Figure 4.9. QM(B3LYP/TZVP//BP86/SVP)/CHARMM22 energy profile along the $d(O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}) - d(O4_{\beta\text{-Gal}}$–$C1'_{\alpha\text{-Gal}})$ reaction coordinate, for (**A**) the wild type enzyme and (**A**,**B**) the Q189A mutant. The charges at the anomeric center of the wild type enzyme were computed at the same level of theory and depicted in (**A**). The variation of selected interatomic distances involved in the reactive process is also shown in (**B**) for mutant Q189A.

4.3.3.1.1. Q189A mutant. Experimentally, the Q189A mutant exhibits a residual activity of 3% (i.e. the k_{cat} value reduced to 3% of that of the wild-type enzyme)[61] which translates into an increase of about 2 kcal/mol in the phenomenological free energy barrier. At first sight, this agrees well with the computed electrostatic TS stabilization due to Gln189 (see Figure 4.8) which will be absent in the Q189A mutant. Of course, the Q189A replacement causes other changes too, and therefore, we decided to build the Q189A mutant *in silico* starting from the coordinates of the wild-type enzyme and making the required substitutions for residue 189. The energy profile for the front-side attack reaction was then recalculated (see Figures 4.9A,B) and the corresponding reactants, TS, and products were reoptimized.

Contrary to expectation, the computed QM/CHARMM22 energy barrier for the Q189A mutant (See Figure 4.9A and Appx., Table A4.2) is slightly smaller than that for the wild-type enzyme, by 1.8 kcal/mol for QM = B3LYP/TZVP//BP86/SVP and by 0.4 kcal/mol for QM = M05-2X/TZVP//BP86/SVP (See Figure 4.12). The effect of the mutation on the potential energy barrier is thus relatively small, especially for M05-2X, but still in the wrong direction.

Inspection of the computed energy profiles shows that the initial breaking of the $O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$ bond and approach of the attacking $O4_{\beta\text{-Gal}}$ atom is more difficult for the Q189A mutant, making its energy profile much wider and higher initially (Figure 4.9A). However, the TS occurs somewhat earlier in the mutant compared with the wild type, and the energy thus has to rise more in the latter before the TS is reached (Figure 4.9A). In the optimized QM(BP86/SVP)/CHARMM22 TS structure, the key distances are $d(O3B_{UDP}–C1'_{\alpha\text{-Gal}})$ = 3.03 Å, $d(O4_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}})$ = 2.40 Å, $d(HO4_{\beta\text{-Gal}}–O4_{\beta\text{-Gal}})$ = 1.04 Å, and $d(HO4_{\beta\text{-Gal}}–O3B_{UDP})$ = 1.66 Å, that is, the breaking bond is shorter and the forming bond is longer than in the case of the wild-type enzyme. The Q189A mutant thus has an earlier TS with less dissociative character. One may speculate that this could enhance entropic effects (a tighter TS would make $\Delta S^{\ddagger}$ more negative and $\Delta G^{\ddagger}$ more positive, because of the larger $-T\Delta S^{\ddagger}$ term). We have not quantified such effects, however, and therefore refrain from speculating whether this may explain the wrong trend in the computed barriers.

Still, an interesting feature of the PES of this system comes out of this *in silico* mutagenesis experiment. According to the calculated potential energy profile, the Q189A mutant follows a S_Ni mechanism although with a somewhat less dissociative TS. Notice that whether the reaction has a single dissociative oxocarbenium ion-like TS (S_Ni) or a short-lived ion-pair intermediate (S_Ni-like) obviously depends on the actual shape of the potential (or free) energy surface along the reaction coordinate (from RC ~ -1.5 to 2.0 Å, see Figure 4.9A), or in other words, on the extent and timing of the making and breaking of bonds.[251] The Q189A mutation with the replacement of a weak nucleophile already significantly affects the energy profile and the location of the TS (see Figure 4.9A). It is conceivable, thus, that other mutations lead to larger changes, up to a point where another local minimum for an ion-pair intermediate appears. Therefore, the tuning by the environment may determine the exact mechanism followed by the enzyme (S_Ni *vs* S_Ni-like) and this could introduce differences between ret-GTs.

4.3.3.1.2. Q189E mutant. If Gln189 assists the reaction by stabilizing the charge development at the anomeric center, the Q189E mutation would be expected to provide a much better stabilization and, thus, favor the S_Ni mechanism even more. This is exactly what we find if we transform *in silico* Gln189 to a pseudo-glutamate (Q189E*) by simply including this residue in the MM part of the system and assigning it the charges of a glutamate, without changing the geometries obtained for the WT enzyme (Appx., Figure A4.4). However, when we actually build the Q189E mutant *in silico* and perform reoptimizations (in analogy to the procedure outlined above for the Q189A mutant), a quite different scenario emerges. Energy minimization of the Q189E mutant takes the system straight to the CGE complex, indicating that in this case CGE formation is barrierless. At the QM(BP86/SVP)/MM(CHARMM22) level, the $OE1_{Q189}$–$C1'_{\alpha\text{-Gal}}$ distance in the formed CGE is 1.49 Å. The interaction of Glu189 with the Ala154 amide is maintained, but not the hydrogen bond with the O6 atom of LAT (which is oriented to make a hydrogen bond with Asp130, Figure 4.10A).

The S_N2 attack of LAT at the CGE, in what would be the second step of the double displacement mechanism, was studied by scanning the $d(O4_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}})$ reaction coordinate at different levels of theory (Figure 4.11). In all DFT-based energy profiles, the computed barrier is higher than 30 kcal/mol. TS optimization has only been successful for SCC-DFTB/CHARMM22, yielding a barrier height of 28.8 kcal/mol that increased to 33.9 and 39.5 kcal/mol in single-point energy calculations with QM = B3LYP/TZVP and M05-2X/TZVP, respectively (See Appx., Table A4.3). According to these QM/MM results, the overall reaction for the Q189E mutant would thus be very

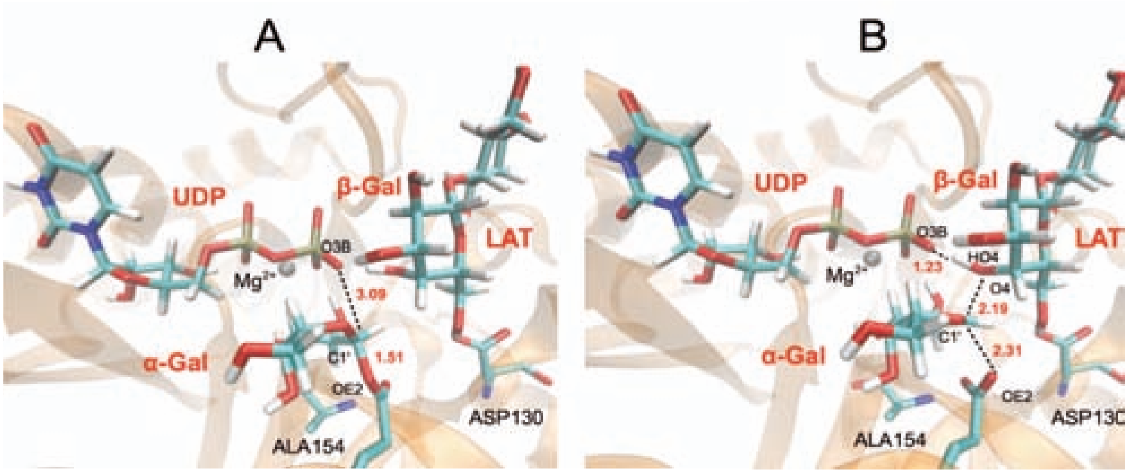

Figure 4.10. Optimized (**A**) covalent glycosyl-enzyme complex and (**B**) transition state for the second step of a double-displacement mechanism in the LgtC Q189E mutant optimized at the SCC-DFTB/CHARMM level. Selected distances (in Å) are indicated in red.

slow, and moreover, the reaction energy is also computed to be very high (26-30 kcal/mol). Key distances in the SCC-DFTB/CHARMM22 TS structure are $d(\mathrm{OE1_{Q189}{-}C1'_{\alpha\text{-}Gal}})$ = 2.31 Å, $d(\mathrm{O4_{\beta\text{-}Gal}{-}C1'_{\alpha\text{-}Gal}})$ = 2.19 Å, $d(\mathrm{HO4_{\beta\text{-}Gal}{-}O4_{\beta\text{-}Gal}})$ = 1.18 Å, and $d(\mathrm{HO4_{\beta\text{-}Gal}{-}O3B_{UDP}})$ =1.23 Å (Figure 4.10B).

The case of the Q189E mutant would be an example of a change in mechanism introduced by a mutation. Recently, Goedl and Nidetzky[252] have remodeled sucrose phosphorylase to change its kinetics and chemical mechanism from a double-displacement to a direct front-side nucleophilic displacement reaction. In our calculations, we observe CGE formation in the Q189E mutant, but galactosyl transfer via a double displacement mechanism involving Glu189 has a high barrier and is thus too slow to be feasible, at least in the presently studied conformation of the enzyme. We note that $\mathrm{O4_{\beta\text{-}Gal}}$ access to $\mathrm{C1'_{\alpha\text{-}Gal}}$ is hampered by $\mathrm{H1'_{\alpha\text{-}Gal}}$ and also limited by the need to keep the $\mathrm{HO4_{\beta\text{-}Gal}{-}O3B_{UDP}}$ interaction. This enforces a TS structure with a dissociated $\mathrm{OE1_{Q189}{-}C1'_{\alpha\text{-}Gal}}$ bond, but in the resulting species, Glu189 does not seem to be stabilized enough by the environment. In fact, kinetic experiments on the Q189E mutant of LgtC found a reduction in k_{cat} to 3% of that of the wild-type enzyme and a reduction in its (low) hydrolytic activity by a factor of 10.[86] Formation of a CGE intermediate was indeed detected but involved Asp190, whose side chain carboxylate is located as far as 8.9 Å away from the anomeric carbon $\mathrm{C1'_{\alpha\text{-}Gal}}$ in the Q189E:UDP-

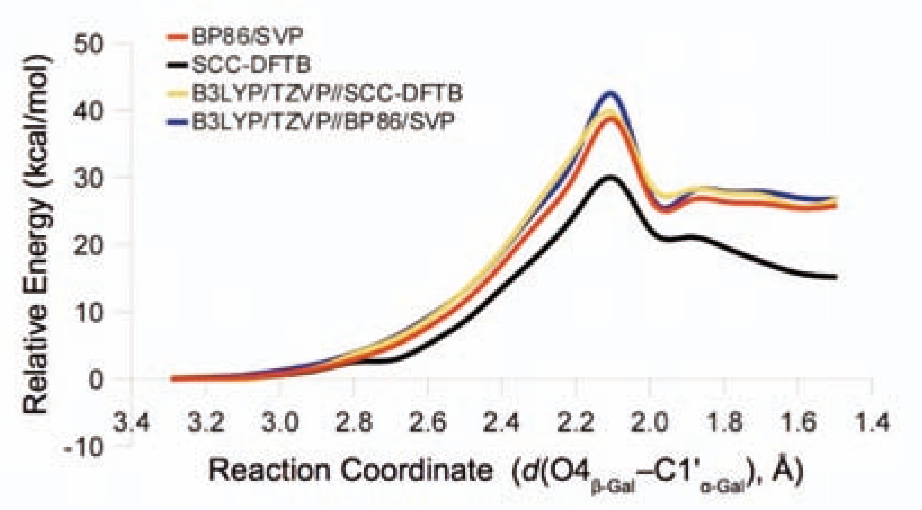

Figure 4.11. QM/CHARMM potential energy profiles computed at different QM levels of theory, for the second step of the double displacement mechanism in the Q189E mutant.

2′FGal crystal structure. These experimental results suggest the possibility of a double-displacement mechanism in the Q189E mutant, in which the remote residue Asp190 acts as the catalytic nucleophile. This would obviously require significant conformational changes (relative to the wild-type crystal structure) to correctly position Asp190, which might occur as a consequence of the mutation or upon acceptor binding. A detailed understanding of the reaction in the Q189E mutant would clearly require further experimental and computational studies that are beyond the scope of the present work.

4.3.3.2. Inter- and intra- substrate interactions. Clearly, and as despicted in Figure 4.4 some inter- or intra-substrate interactions migh be involved in the TS stabilization. More specifically, three substrate-substrate hydrogen bonds can be seen in Figure 4.4: one involving O4$_{\beta\text{-Gal}}$ and O3B$_{\text{UDP}}$; another one between O2′$_{\alpha\text{-Gal}}$ and O1B$_{\text{UDP}}$ (i.e., the oxygen atom of the β phosphate coordinated to the metal cation); and a third one between O3$_{\beta\text{-Gal}}$ and O3B$_{\text{UDP}}$. All of them are present in the reactants (with O−H···O distances of 1.67, 1.68, and 1.94 Å, respectively, at the QM(BP86/SVP)/MM(CHARMM22) level. These hydrogen bonds get shorter at the TS (1.36, 1.64, and 1.65 Å, respectively) and thereby stabilize the increasing negative charge in UDP.

Another interaction that may help stabilizing the highly charged phosphate at the TS is the hydrogen bond between $O2A_{UDP}$ of the α phosphate and O3 of the UDP ribose.

To estimate the energy contribution provided by these hydrogen bonds, we produced *in silico* variants of the substrates (UDP-2′-deoxygalactose and 3-deoxylactose) and calculated the front-side attack pathway for each of them at the QM(BP86/SVP)/CHARMM22 level using the $d(O3B_{UDP}–C1'_{\alpha\text{-Gal}})$ - $d(O4_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}})$ reaction coordinate. This was followed by optimization and characterization of the corresponding transition states. The results are presented in Figure 4.12 (See also Appx., Tables A4.4 and A4.5 for other levels of theory). The computed barrier heights confirm that substituting one of these OH-groups by a hydrogen atom impedes the reaction significantly: it increases the barrier by 2.3-3.3 kcal/mol compared with the unmodified substrates. Assuming that these differences are maintained in the free energy barriers, this would reduce the k_{cat} value of the mutants to 0.6-3% of that of the wild type at room temperature. The use of modified substrates leads to somewhat earlier TSs, as indicated by the optimized distances $d(O3B_{UDP}–C1'_{\alpha\text{-Gal}})$ = 3.09/3.06 Å, $d(O4_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}})$ = 2.39/2.35 Å, $HO4_{\beta\text{-Gal}}–O4_{\beta\text{-Gal}}$ = 1.11/1.08 Å, and $HO4_{\beta\text{-Gal}}–O3B_{UDP}$ = 1.29/1.35 Å for the substrates UDP-2’-deoxyGal/3-deoxyLAT, respectively.

The changes in the computed TS geometries are similar for both modifications, although the replacement of the 3-OH group by H in lactose tends to increase the computed barrier slightly more. By contrast, the reaction energy is much more affected when the 2’-OH group of the α-galactose of UDP-Gal is replaced by H, leading to a rather high endoergicity of ~9 kcal/mol. If we would substitute the 2’-OH group of UDP-Gal by fluorine instead of hydrogen, the oxocarbenium ion-like transition state should be further destabilized inductively, and the result can easily be an inert UDP-2’FGal (which acts as a competitive inhibitor with respect to UDP-Gal), as has been observed experimentally.[61]

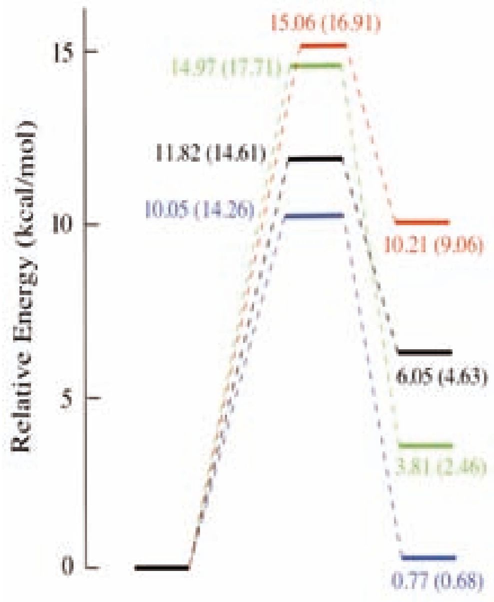

Figure 4.12. QM/CHARMM22 potential energy barriers and reaction energies for the front-side attack mechanism (S_Ni), with QM = B3LYP/TZVP//BP86/SVP (M05-2X/TZVP//BP86/SVP). Color code: *black*, unmodified wild type enzyme and substrates; *red*, wild type with modified substrates UDP-2'-deoxygalactose and lactose; *green*: wild type with modified substrates UDP-Gal and 3-deoxylactose; *blue*, Q189A mutant with unmodified substrates.

4.4. CONCLUSIONS

In the present study we have used QM(DFT)/MM calculations on the full enzyme to study the reaction catalyzed by LgtC as well as the corresponding reaction with alternative substrates and with LgtC mutants. This provides us with a detailed description of the reaction catalyzed by this enzyme. The different mechanisms proposed in the literature (S_Ni, S_Ni-like and double displacement mechanism *via* the formation of a CGE intermediate) have been investigated and compared. We find a dissociative S_Ni mechanism for the wild type enzyme with the most reliable QM/CHARMM22 barriers ranging between 11.8 kcal/mol (B3LYP/TZVP) and 14.6 kcal/mol (M05-2X/TZVP), in reasonable agreement with the experimental kinetic data. We have identified several factors that help the front-side mechanism, in particular enzyme-substrate and substrate-substrate interactions. Among them, the largest effects come from Lys250, which is also involved in binding. Gln189, the putative nucleophile in a double displacement mechanism, is found to favor the charge development at the anomeric center during the reaction by about 2 kcal/mol. Moreover, we predict that 3-deoxylactose as acceptor will increase the barrier height by 2-3 kcal/mol (reduction of k_{cat} to 0.6-3% of that for the unmodified substrates).

LgtC mutants Q189A and Q189E were also analysed. When Gln189 is substituted by Ala (Q189A), an S_Ni mechanism is still predicted but with a less pronounced maximum and a wider and flatter barrier top. For the LgtC Q189E mutant, an even more drastic change in mechanism is computed, from a front-side attack to the formation of a CGE with Glu189 that cannot evolve to the products, at least not with the present enzyme conformation. These two examples ilustrate how mutation of enzyme residues can alter the potential energy surface of the system with changes that can even modify the catalytic mechanism.

5

α-1,3-galactosyltransferase (α1,3-GalT)

5.1. INTRODUCTION

Bovine α-1,3-galactosyltransferase (EC 2.4.1.87, α1,3-GalT) is a ret-GT that catalyzes the formation of an α-1,3 glycosidic linkage and has been the focus of several structural and mechanistic investigations.[41,42,49,253,254] Just like LgtC, α1,3-GalT catalyzes the transfer of an α-galactose from UDP-Gal to another saccharide to yield an elongated oligosaccharide. The enzyme presents a GT-A fold and the typical Asp-X-Asp (DXD) signature implicated in the binding of the donor sugar substrate and the coordination of a divalent cation (M^{2+}).[72] α1,3-GalT, though, has a glutamate at position 317, equivalent to Gln189 in LgtC.

Based on an initial X-ray crystal structure, Glu317 was supposed to play the role of the catalytic nucleophile in a double-displacement mechanism.[72] More recently, experiments by different groups[87,255,256] seem to agree on the importance of Glu317 in catalysis. E317D, E317C, E317H[256] or E317Q[255] mutants present residual activities of 0.04-0.8 %. For the E317A mutant Molina et al.[256] reported a residual activity of 0.1 %

while Monegal and Planas[87] obtained an even higher decrease in the catalytic activity (k_{cat} of 1.3 s^{-1} and $<10^{-4}$ s^{-1} for WT and E317A, respectively). Additionally, the authors achieved a 100-fold increase in the E317A k_{cat} via a *"chemical rescue"* by adding azides (See Chapter I Section 1.2.3.2 for further details). However, and even though this data was initially interpreted as consistent with the double-displacement mechanism, the covalent intermediate (CGE) has never been isolated and, therefore, alternative mechanisms cannot be completely ruled out.

On the other hand, the availability of an increasing number of crystal structures has shown that in most retaining GTs there is not a well-positioned nucleophile in the active site to act as the nucleophile.[10] Therefore, alternative mechanisms involving retention of the configuration are pushed forward (e.g. S_Ni or S_Ni-like), at least for those retaining GTs where no good nucleophile is suitably positioned to form the CGE. The exception would precisely be for members of family GT6 (like α1,3-GalT) where such a residue has always been identified. More interestingly, in these alternative front-attack mechanisms, the presence of a nucleophile on the β-face of the sugar ring could facilitate catalysis by *"pushing"* the leaving group[10] without implying the formation of a CGE and thus, the exact role of such nucleophile remains unclear.

At the present stage of investigation there are several pending questions to be clarified. First, it is necessary to know whether GT6 family members follow a double displacement mechanism or not. Secondly, to reveal the exact role of this glutamate/aspartate residue in the vicinity of the anomeric center, and finally, to rationalize the difficulties found to isolate the covalent intermediate (if formed) experimentally.

In this chapter we present the first full-enzyme hybrid quantum mechanical/molecular mechanical (QM/MM) study on a ret-GT where a *bona fide* nucleophile is present in the active site. More specifically, the GT6 member α1,3-GalT will be used as a model to give answer to the above questions. In brief, the different mechanistic alternatives outlined before are studied; and the substrates implication and the role of the most relevant active-site residues of α1,3-GalT are analyzed, giving special attention to the putative nucleophile Glu317.

5.2. MODELS AND METHODS

Coordinates from the X-ray structure (PDB Code 1O7O,[257] resolution 1.97 Å) were considered as starting point to model the Michaelis complex of the enzyme with its ligands (α1,3-GalT + M^{2+} + UDP-Gal + LAT). Coordinates for the enzyme, the acceptor substrate (LAT), the divalent cation and the crystallographic water molecules were taken from this X-ray structure. The protonation states of the titratable residues were assigned by the PROPKA procedure.[215] The coordinates for the donor substrate were derived from the PDB Code 1GA8,[2] from where the substrate analogue (UDP-2'FGal) was manually transformed into the original UDP-Gal substrate and superimposed with the UDP moiety present in the 1O7O crystal structure. The Mn^{2+} ion present in the original X-ray structure was modeled by the computationally more convenient Mg^{2+}. Finally, the system was fully solvated with a cubic box of TIP3P water molecules (83 x 77 x 73 $Å^3$) and one Cl^- ion was added to neutralize the system using the program VMD.[245]

Starting from this model of the ternary complex we performed classical Molecular Dynamics (MD) simulations at 300 K using the CHARMM22 force field[114,216] and periodic boundary conditions, as implemented in the NAMD software.[258] Specific topology and parameters from the CHARMM force field for carbohydrates were considered.[218] During the MD simulations, the Langevin piston and the Nosé-Hoover method[196,197] were used to keep the system at 1 bar and 300 K respectively. The integration time step was 2 fs. During the initial equilibration process only water molecules were free to move and the volume was kept constant for 100 ps of MD simulation. Afterwards, the protein structure was minimized for 2000 conjugate gradient optimization steps and released for 100 ps of constant pressure and temperature MD (with the backbone atoms fixed for the initial 50 ps and the acceptor and donor substrates fixed the whole time). Next, the UDP-Gal was also minimized and 20 ps steps of MD simulation were performed applying gradually decreasing restraints to this substrate (5.0, 2.0, 1.0, 0.5, 0.2, 0.1, 0.05 and 0.00 kcal mol^{-1} $Å^{-2}$) while keeping the acceptor substrate (LAT) fixed. A similar approach was then applied for the LAT, with decreasing force constants of 5.0, 2.0, 1.0 and 0.5 kcal mol^{-1} $Å^{-2}$. Within these conditions, a final run of 5 ns of MD simulation was carried out.

A random snapshot from this latest MD simulation was taken as starting point for QM/MM MD simulations. All non-protein atoms being more than 24 Å away from the anomeric center ($C1'_{\alpha\text{-Gal}}$) were deleted. This procedure resulted in a system with 12694 atoms, including 2614 TIP3P water molecules. All residues and water molecules within 15 Å of the anomeric center (2089 atoms) were included in the active region (see Figure 5.1A). Six hydrogen link atoms were employed to treat the QM/MM boundary. The charge of the QM region was -3 and included 84 atoms: those from the α- and β-galactose rings from UDP-Gal and LAT, respectively, Mg^{2+} and its first coordination sphere (phosphate groups from UDP and the side chains of residues Asp225, Asp227 and a crystallographic water), as well as the side chain of Glu317 (see Figure 5.1B).

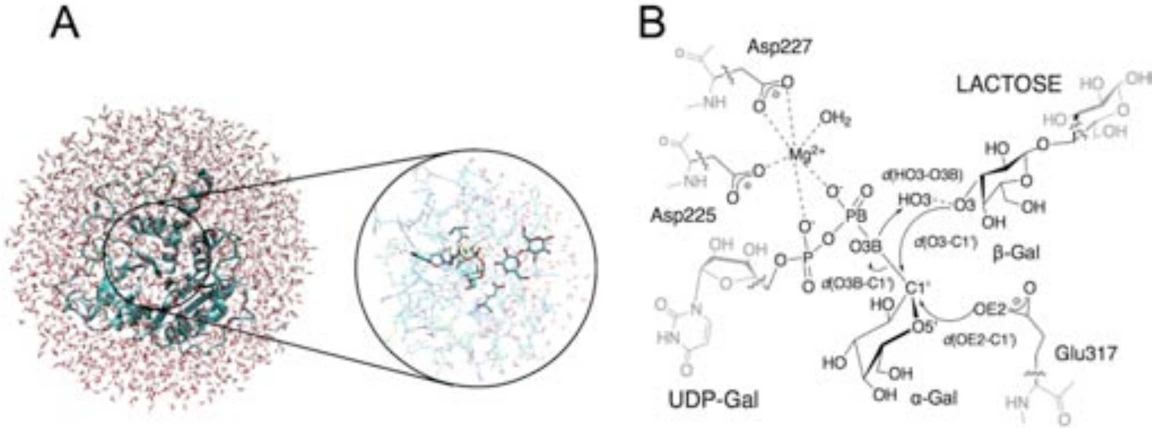

Figure 5.1. Model system used in the QM/MM calculations (**A**) and (**B**) QM/MM partition considered in the present work. In (**A**) the active region is enlarged and the QM atoms represented in licorice. In (**B**), QM (MM) atoms are depicted in black (grey). Wavy lines indicate the boundary between the QM and MM regions. The arrows indicate the distances considered in the reaction coordinates and the atoms involved are labeled.

Later on, a NVT QM(SCC-DFTB)[131]/MM(CHARMM22) MD was performed using the dynamics module within ChemShell.[219] The SHAKE procedure[199] was applied at every step for the O-H bonds in the water molecules. A 10 ps MD equilibration run was followed by 80 ps of production MD. One selected snapshot from this simulation was used in a preliminary study at the QM = SCC-DFTB level for primary optimizations. A re-optimization of the stationary points obtained was carried out at the QM(BP86[149,223-225]/SVP[235])/CHARMM22 level of theory. In a later study four snapshots were used in QM/MM calculations with QM = BP86/SVP.

The rest of the QM/MM calculations were performed as exposed in Chapter IV and no new procedures need to be commented here.

5.3. RESULTS AND DISCUSSION

5.3.1. Preliminary study. The QM(SCC-DFTB)/MM(CHARMM22) level of calculation was initially considered together with complex reaction coordinates (i.e. involving bond-breaking (*d*(O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$)), bond-forming (*d*(OE2$_{E317}$–C1'$_{\alpha\text{-Gal}}$)/*d*(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$)) and proton transfer (*d*(HO3$_{\beta\text{-Gal}}$–O3B$_{UDP}$)) processes) to model the different mechanistic alternatives for one frame. Once the R, TS and P had been characterized at this level of theory (Tables A5.1 and S5.2), reoptimization of these stationary points at the QM = BP86/SVP level was carried out (Table A5.3). Single point energy calculations followed (Table 5.1).

Table 5.1. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and S$_N$i mechanisms at different levels of theory. The calculations were carried out on the re-optimized QM(BP86/SVP)/MM(CHARMM22) geometries of the stationary points.

Double-displacement mechanism				
	BP86	**B3LYP//BP86/SVP**	**M05-2X//BP86/SVP**	
	SVP	**SVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00
TS$^d_1$	12.11	12.45	15.61	14.64
CGE	11.39	11.69	10.69	13.73
TS$^d_2$	18.66	20.84	21.64	21.15
P	2.17	1.68	-2.49	2.62
S$_N$i mechanism				
TS^i	14.32	15.02	17.98	15.42
P	0.74	-0.16	-4.05	2.15

For the double-displacement mechanism the potential energy barriers at the QM=M05-2X/TZVP//BP86/SVP level for the TS^d_1, CGE and TS^d_2 are 14.64, 13.73 and 21.15 kcal/mol respectively and so, the second step would be the rate-limiting one.

With regard to the other mechanistic alternatives (Figure 1.6B-C), the calculations performed at the QM=M05-2X/TZVP//BP86/SVP level also lead to the characterization of a S_Ni mechanism with an energy barrier of 15.42 kcal/mol. This implies that no high energy (short-lived) oxocarbenium intermediate (IP) but a dissociative TS (i.e. TS^i) was localized (see Tables 5.1 and S5.1 for energies and relevant geometrical parameters, respectively).

An especial distinction should be made about the important effect revealed for Glu317 in the case of a front-side attack mechanism. Very interestingly, during the front-side attack of LAT on UDP-Gal, the anomeric carbon and the OE2 atom of Glu317 get closer by ~1.5 Å (Appx., Table A5.1). Such significant decrease in the $d(\mathrm{OE2_{E317}–C1'_{\alpha\text{-}Gal}})$ distance, from 4.26 at reactants to 2.80 Å at the TS^i, is most likely assisting the UDP departure. Our results suggest that the incoming acceptor substrate (LAT) would hardly get closer to the anomeric center unless the leaving group is previously *"pushed"* (without formation of a covalent intermediate) by the back-side nucleophile Glu317.

In order to further analyze the role of Glu317 in the galactosyl transfer catalyzed by α1,3-GalT, we calculated a two-dimensional potential energy surface (PES) (Figure 5.2). At fixed values of the distance $d(\mathrm{OE2_{E317}–C1'_{\alpha\text{-}Gal}})$, the energy was computed by constrained QM(SCC-DFTB)/CHARMM22 optimizations along the reaction coordinate RC = $d(\mathrm{O3B_{UDP}–C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{O3_{\beta\text{-}Gal}–C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{HO3_{\beta\text{-}Gal}–O3B_{UDP}})$. For clarity, only $d(\mathrm{O3_{\beta\text{-}Gal}–C1'_{\alpha\text{-}Gal}})$ is represented in the *x*-axis. Notice that QM(SCC-DFTB)/MM(CHARMM22) method provides higher energy barriers and tighter TS geometries than the QM(BP86/SVP)/MM(CHARMM22) method (see Tables A5.2 and S5.3), but it is much faster and served our purpose here of having a general overview of the catalytic mechanism followed by α1,3-GalT.

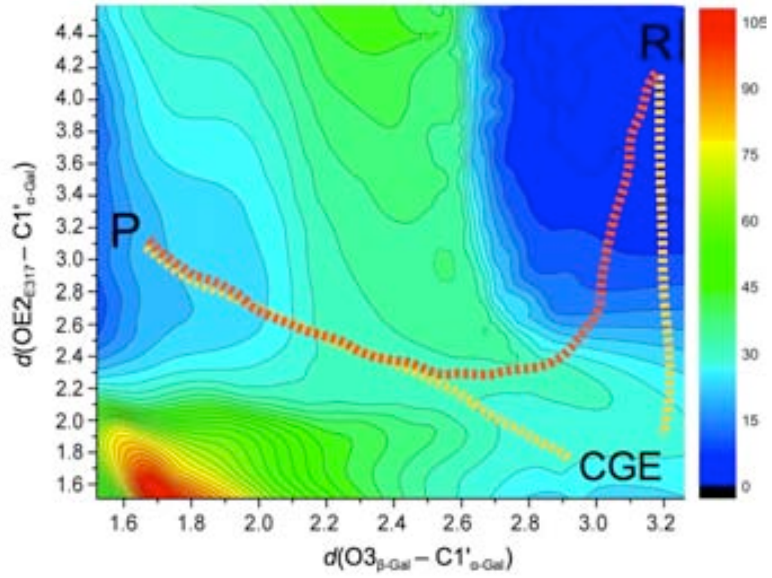

Figure 5.2. Two-dimensional QM(SCC-DFTB)/CHARMM22 potential energy surface. Energies are given in kcal/mol and distances in Å. Contour lines are drawn in intervals of 3 kcal/mol.

The PES depicted in Figure 5.2 shows two prominent minima, at the top-right corner and at the middle-left border, corresponding to the reactants (R) and the products (P), respectively. There is a third minimum in the region of the covalent intermediate (CGE, bottom-right corner), although it is much less pronounced. We can easily identify two possible pathways connecting reactants and products. One of them would proceed in two steps and the formation of a CGE intermediate, while the other one would imply a concerted mechanism *via* an oxocarbenium ion-like transition state (S_Ni mechanism; See Figure 1.6B).

Moreover, the PES depicted in Figure 5.2 shows very clearly that the S_Ni mechanism is not feasible unless the d(OE2$_{E317}$–C1'$_{\alpha\text{-Gal}}$) distance gets shorter (~2.4 Å). Accordingly, one might think of a catalytic mechanism where Glu317 plays a key role as a nucleophile, not leading to the formation of a CGE intermediate but facilitating the leaving group departure in a *"S_Ni-assisted"* mechanism.

5.3.2. Catalytic mechanism. We started considering the QM(SCC-DFTB)/ MM(CHARMM22) level of theory for geometry optimizations along the reaction paths with the intention of having a first overview of the different proposed mechanistic scenarios in α1,3-GalT. The potential energy barriers obtained for the S_Ni and the double-displacement mechanisms are 15.42 and 21.15 kcal/mol (QM = M05-2X/TZVP), respectively, being both values in qualitative agreement with the phenomenological free energy barrier of ~17 kcal/mol derived from the experimental k_{cat} values of 6.4 s^{-1} at 310 K.[259] Even if a double-displacement mechanism has always been assumed for α1,3-GalT, our results so far suggest that the front-side attack may also be feasible. At this point, two remarks should be made. First, considering the relatively close barrier heights reported in Table 5.1 for both mechanisms, it would be dangerous to conclude about the dominant mechanism used by the enzyme; a quantitative discrimination would require ensemble-averaged energies and, probably, a more extended testing of electronic structure methods. Secondly, it is important to notice that the difference between a S_Ni and a S_Ni-like mechanism (that is between having an oxocarbenium TS or intermediate, respectively; see Figures 1.6B-C) may be quite subtle and difficult to assess computationally. For example, in the work of Ardèvol et al. the oxocarbenium ion corresponds to a shallow minimum which represents an extremely short-lived ion pair.[260] In fact, a very flat free energy surface (± 2 kcal/mol) was obtained for the zone corresponding to the oxocarbenium intermediate and its surroundings. In such type of energy landscapes, the performance of the electronic structure methods used to calculate the potential energy might be critical again, as it can be the effect of the dynamics of the environment.[261] Nonetheless, it should be pointed out that the calculations reported so far allowed us to have a qualitative view of the different proposed mechanistic scenarios and revealed some interesting features like the relevance of Glu317 in the case of a front-side attack mechanism.

In any case, and taking all the above into consideration we decided to perform further calculations at a higher level of theory as well as considering different starting structures. More specifically, four random frames of the QM(SCC-DFTB)/MM(CHARMM22) MD were considered for QM(DFT)/MM geometry optimizations along the reaction paths with QM = BP86/SVP.

5.3.2.1. Double-displacement mechanism. We considered the same reaction coordinates as before to model the CGE formation and following attack of LAT on the anomeric center (i.e. RC = [$d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{OE2_{E317}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}})$] and RC = [$d(\mathrm{OE2_{E317}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{O3_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ - $d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}})$], respectively). The QM/MM potential energy barriers and reaction energies calculated at the QM = (M05-2X/TZVP//BP86/SVP) level for the four frames considered are summarized in Figure 5.3A. A more detailed comparison of the energies calculated at different levels is available in the Supporting Information (Tables A5.4-S5.7).

The average potential energy barriers for formation and subsequent cleavage to products of the CGE intermediate are quite similar (15.1 ± 3.5 kcal/mol and 15.9 ± 3.1 kcal/mol, respectively, both measured with respect to the reactants), and are also quite similar within a given frame. These results are in qualitative agreement with the corresponding experimentally derived phenomenological free energy barrier of ~17 kcal/mol (k_{cat} = 6.4 s^{-1} at 310 K),[259] and would give theoretical support to the double displacement mechanism hypothesis for this enzyme-substrates system. As seen in Figure 5.3A, CGE formation is quite endoergic (CGE lays 11.4 ± 4.2 kcal/mol over the reactants) and the barrier for going back to reactants is relatively small (3.7 ± 1.3 kcal/mol). The latest suggests that, even in the presence of the acceptor substrate, the covalent intermediate could just move back instead of going directly to the final reaction products, which would in part explain why the isolation of glycosyl-enzyme intermediates for wild type ret-GTs has been so elusive for experimentalists.

At this point it might have been noticed that two groups of potential energy barriers can be distinguished in Figure 5.3: those from frames 1-2 and 3-4, respectively. In order to explain this issue, we measured the RMSD values in the reactants for all the residues in the active space between the four structures considered in the present study. The only important difference in the vicinity of the reaction center was found for a water molecule (crystallographic water ID 2283 in original PDB). Taking frame 1 as a reference, the corresponding RMSD values are 0.32, 1.51 and 1.91 Å for frames 2-4 respectively, showing that the position of this residue is clearly different between frames 1-2 and 3-4 (See Figure 5.4). This is a crystallographic water molecule that was conserved when the Michaelis complex was built using the coordinates from PDB

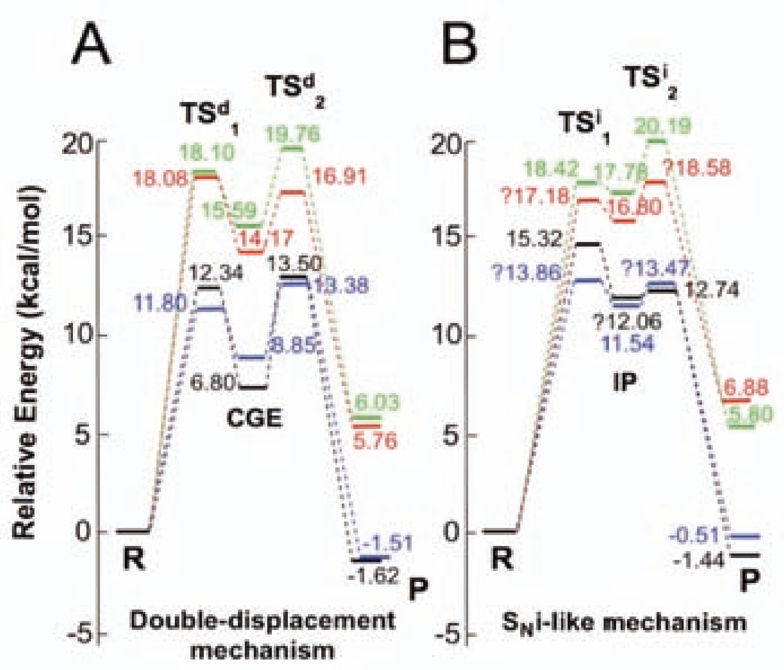

Figure 5.3. QM/CHARMM potential energy barriers and reaction energies for α1,3-GalT. QM=M05-2X/TZVP//BP86/SVP. (**A**) Double-displacement mechanism. (**B**) S_Ni-like mecha-nism. ?: The TS or IP nature of this stationary point could not be confirmed by the frequency calculation. Blue, black, red and green lines correspond to the frames considered in the present study as starting points for the QM/MM calculations (frames 1 to 4 respectively).

1O7O as a template. This water residue was 4.82 Å away from the O3B$_{UDP}$ atom in this structure but got closer when we performed MM and QM/MM MDs on the ternary complex. In particular, in frames 1 and 2 this water molecule is hydrogen-bonded to O3B$_{UDP}$ in the reactants, whereas in frames 3 and 4 is not. In order to confirm the stabilizing role of such water molecule along the paths we switched off its charge and made SP calculations at M05-2X/TZVP//BP86/SVP level for frames 1 to 4. The results are presented in the Appx. section, Table A5.8.

As shown in Table A5.8, there is a notorious electrostatic stabilization by water 2283 in frames 1-2, that is, there is an increment of the (QM + QM/MM$_{interactions}$) potential energy barriers and reaction energies of ~3-8 kcal/mol after switching off the charge of the water molecule. In contrast, for frame 4 no significant effect is observed. Unexpectedly, for frame 3 a significant effect has been obtained. A more detailed analysis of this result revealed that, although water molecule 2283 is not hydrogen-bonded to O3B$_{UDP}$ in the reactants in frames 3-4, in frame 3 it gets reoriented along the

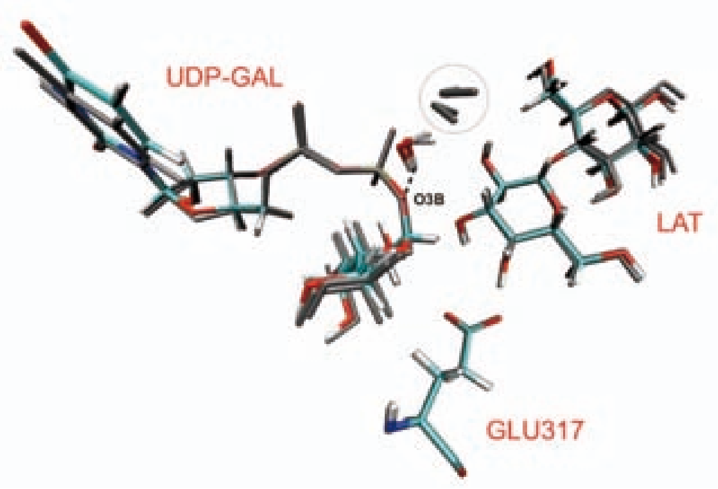

Figure 5.4. Structural superposition of the four structures of α1,3-GalT considered in the present study as starting point for the QM/MM calculations. For clarity, only the ligands, Glu317 and the crystallographic water 2283 are depicted. Structures from frames 3-4 appear in gray and the water molecule responsible for the energy barrier differences between the frames is circulated.

reaction path (e.g. $d(\mathrm{O3B}_{\mathrm{UDP}}\text{-}\mathrm{O}_{\mathrm{WAT2283}})$ = 4.53 (4.35) Å and 3.06 (4.08) Å, for the R and TS^{i}_{1} in frame 3 (4)). Interestingly, in frame 3 this change in the QM/MM electrostatic interaction is accompanied by an increment of the MM energy term thus resulting in the high total QM/MM energy barriers reported in Figure 5.3. The foregoing indicates that the system is highly sensitive to any stabilizing interaction supporting the UDP leaving process. This discussion will resume shortly (especially in the following chapter), as we have found that is quite relevant to understand the catalytic mechanism followed by ret-GTs.

The evolution of key distances along this reaction path is similar for the four frames considered. The values corresponding to the selected frame are listed in Table 5.2 (from now on, given the similar results obtained for the four frames and in order to facilitate the discussion, only the results of frame 2 will be reported and discussed in the main text (See Appx. Tables A5.9-A5.11 for the other frames). Essentially, in the transition state of the first step (TS^{d}_{1}) the bond between the UDP leaving group and the α-Gal is already broken ($d(\mathrm{O3B}_{\mathrm{UDP}}\text{–}\mathrm{C1'}_{\alpha\text{-Gal}}) > 2.5$ Å) while the Glu317 OE2 atom is still at 2.6 Å from the anomeric center. These distances are 3.95 Å and 1.56 Å at the CGE, respectively. In the transition state of the second step (TS^{d}_{2}), $d(\mathrm{OE2}_{\mathrm{E317}}\text{–}\mathrm{C1'}_{\alpha\text{-Gal}})$

Table 5.2. Selected QM/MM bond distances *d* (Å) and atomic charges *q* (a.u.) in the optimized reactants (R), transition states (TS), covalent glycosyl-enzyme intermediate (CGE), ion-pair intermediate (IP) and products (P) for the double-displacement and S_Ni-like mechanisms for frame 2. QM=BP86/SVP and M05-2X/TZVP//BP86/SVP for the distances and charges respectively. ?: The IP intermediate was not supported by the frequency calculation.

		Double-displacement mechanism				**S_Ni-like mechanism**			
	R	**TS^d_1**	**CGE**	**TS^d_2**	**P**	**TS^i_1**	**?IP**	**TS^i_2**	**P**
$d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$	1.51	3.18	3.95	3.45	3.45	2.48	3.40	3.59	3.40
$d(\mathrm{O3_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$	3.04	2.85	3.09	2.49	1.48	2.87	2.79	2.63	1.48
$d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3_{\beta\text{-}Gal}})$	1.00	1.02	1.02	1.03	1.51	1.01	1.02	1.03	1.51
$d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}})$	4.06	1.58	1.59	1.51	1.04	1.63	1.57	1.52	1.04
$d(\mathrm{OE2_{E317}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$	4.25	2.60	1.56	2.77	3.18	3.16	2.51	2.61	3.19
$d(\mathrm{C1'_{\alpha\text{-}Gal}}\text{–}\mathrm{O5'_{\alpha\text{-}Gal}})$	1.38	1.28	1.37	1.28	1.40	1.28	1.28	1.28	1.40
$q(\mathrm{C1'_{\alpha\text{-}Gal}})$	0.36	0.58	0.35	0.58	0.32	0.57	0.58	0.58	0.32
$q(\mathrm{O5'_{\alpha\text{-}Gal}})$	-0.51	-0.40	-0.52	-0.41	-0.53	-0.41	-0.41	-0.41	-0.54

= 2.77 Å and $d(\mathrm{O3_{\beta\text{-}Gal}}\text{–}\mathrm{C1'_{\alpha\text{-}Gal}})$ = 2.49 Å, which indicates that the CGE has already dissociated while the incoming hydroxyl group from LAT is still approaching the anomeric center. Thus, both TSs present a highly dissociative character. The TSs and CGE structures are shown in Figure 5.5.

It should be noticed that the $d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}})$ distance was explicity considered to model the CGE formation and following attack of LAT on the anomeric center. Indeed, reorientation of $\mathrm{HO3_{\beta\text{-}Gal}}$ to form a hydrogen bond with $\mathrm{O3B_{UDP}}$ is needed in the first step of the double-displacement mechanism and this could only be accomplished by using the condensed RC reaction coordinate as described. Moreover, CGE formation is not occurring when the $d(\mathrm{HO3_{\beta\text{-}Gal}}\text{–}\mathrm{O3B_{UDP}})$ distance is not included in the definition of the RC; but interestingly, it is not until the second step of the double-displacement mechanism when the proton transfer takes place (See Table 5.2). The foregoing suggests that interactions between the substrates might be a critical factor to take into account in these enzymes and therefore it will be deeply analyzed in further sections.

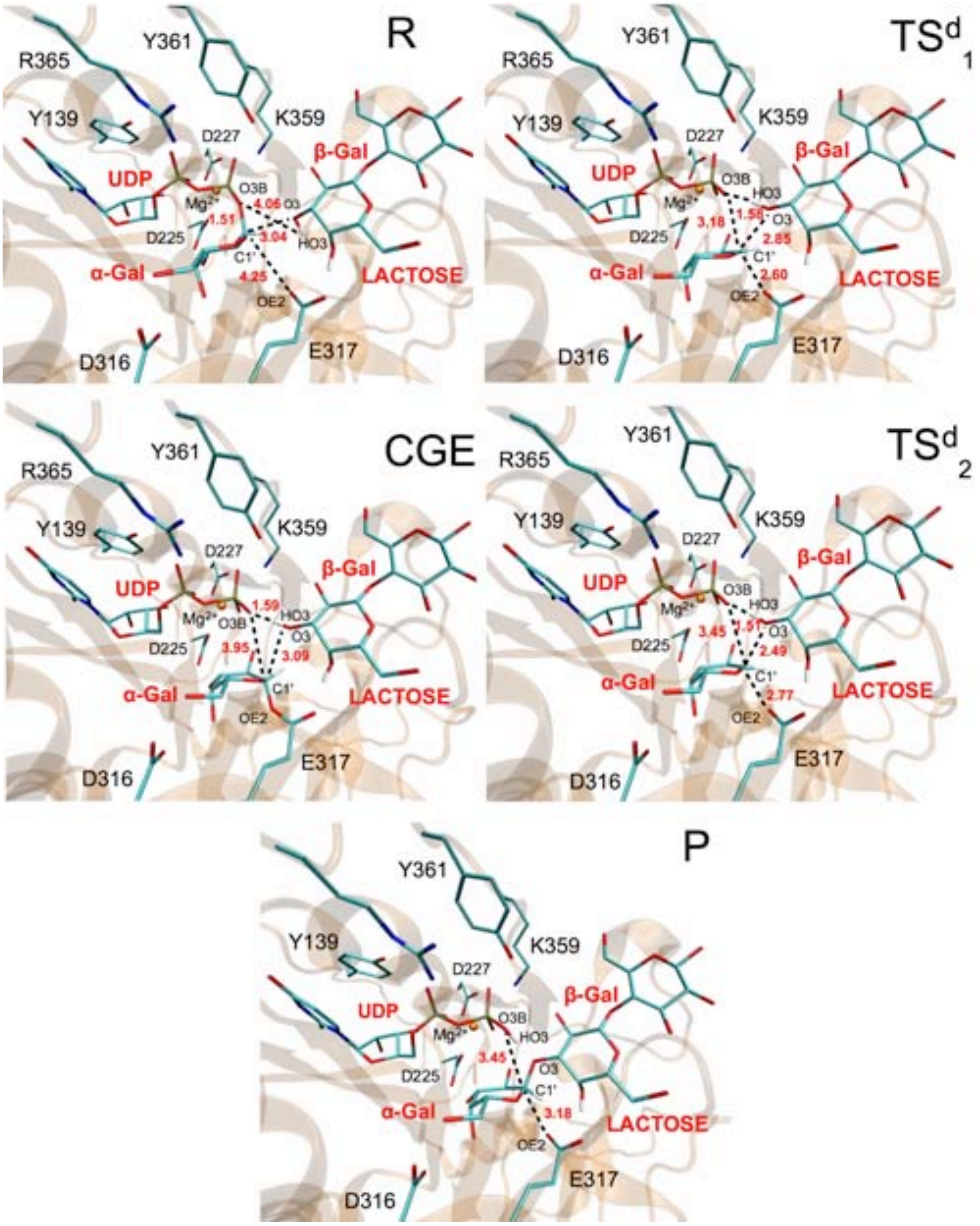

Figure 5.5. QM(BP86/SVP)/MM(CHARMM22) optimized reactants (R), transition states (TS^{d}_{1-2}), covalent intermediate (CGE) and products (P) for the double-displacement mechanism in frame 2. The donor and acceptor substrates, together with some relevant residues in the active site, are represented as sticks. Selected distances (in Å) are indicated in red.

5.3.2.2. Front-side attack mechanism. The front-side attack mechanism was considered next. For that, the RC = [d(O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-Gal}}$) – d(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$) – d(HO3$_{\beta\text{-Gal}}$–O3B$_{\mathrm{UDP}}$)] was used to obtain the starting points for TSs and intermediate search and characterization. The energy barriers and reaction energies calculated are shown in Figure 5.3B. We were able to find an (short-lived) ion pair (IP) intermediate for this mechanism, corresponding to the S$_{\mathrm{N}}$i-like depicted in Figure 1.6C. The TSs and IP structures are shown in Figure 5.6. Again, the transition states for the first and second steps (TS$^{\mathrm{i}}_1$ and TS$^{\mathrm{i}}_2$, respectively) are very similar in average energy (16.2 ± 2.0 and 16.3 ± 3.7 kcal/mol, respectively) and are in the range of the experimentally derived ones. The IP lays 14.6 ± 3.2 kcal/mol over the reactants and between 0.6 and 3.2 kcal/mol below the two TSs (depending on the frame), indicating its short-lived nature. Inspection of Table 5.2 reveals that the two TS$^{\mathrm{i}}$s are again very dissociative and that TS$^{\mathrm{i}}_1$, IP and TS$^{\mathrm{i}}_2$ exhibit similar bond distances. In particular, changes > 0.1 Å are only seen for d(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$) (change of 0.24 Å), d(OE2$_{\mathrm{E317}}$–C1'$_{\alpha\text{-Gal}}$) (0.65 Å) and d(O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-Gal}}$) (1.1 Å), which is the leaving group-sugar bond.

The evolution of key distances along the front-side attack mechanism shows some interesting traits that is worth mentioning. As depicted in Figure 5.7, different events seem to take place in an ordered way. At first (before TS$^{\mathrm{i}}_1$), the main process taking place is the reorientation of HO3$_{\beta\text{-Gal}}$ to form a hydrogen-bond with O3B$_{\mathrm{UDP}}$. It should be pointed out that the d(HO3$_{\beta\text{-Gal}}$–O3B$_{\mathrm{UDP}}$) distance had to be explicitly considered in order to have potential energy barriers consistent with the experimental free energy values (derived from the reported k_{cat} values). It is not until this hydrogen-bond between HO3$_{\beta\text{-Gal}}$ and O3B$_{\mathrm{UDP}}$ is established, that the O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-Gal}}$ bond starts to break. Moreover, lengthening and shortening of the d(O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-Gal}}$) and d(OE2$_{\mathrm{E317}}$–C1'$_{\alpha\text{-Gal}}$) distances appear to be correlated. Also notice from Figure 5.7 that the attacking hydroxyl group from LAT is not considerably approaching to C1'$_{\alpha\text{-Gal}}$ until the latest steps of the catalytic mechanism (i.e. after IP), when the UDP is strongly dissociated, the d(OE2$_{\mathrm{E317}}$–C1'$_{\alpha\text{-Gal}}$) distance increases again and the proton transfer process takes place.

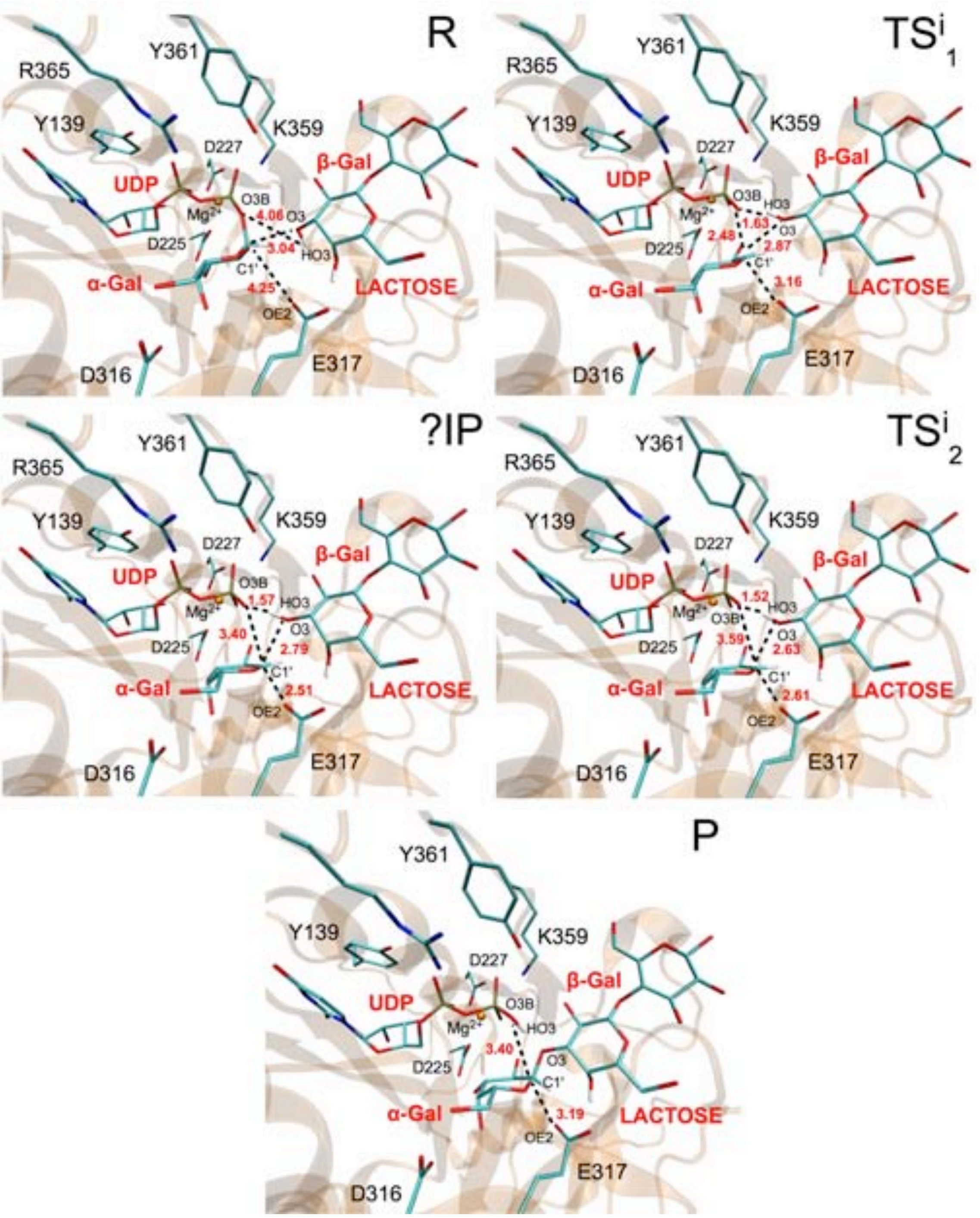

Figure 5.6. QM(BP86/SVP)/MM(CHARMM22) optimized reactants (R), transition states (TS^{i}_{1-2}), ion-pair (IP) intermediate and products (P) for the front-side attack mechanism in frame 2. The donor and acceptor substrates, together with some relevant residues in the active site, are represented as sticks. Selected distances (in Å) are indicated in red. ?: The IP was not supported by the frequency calculation.

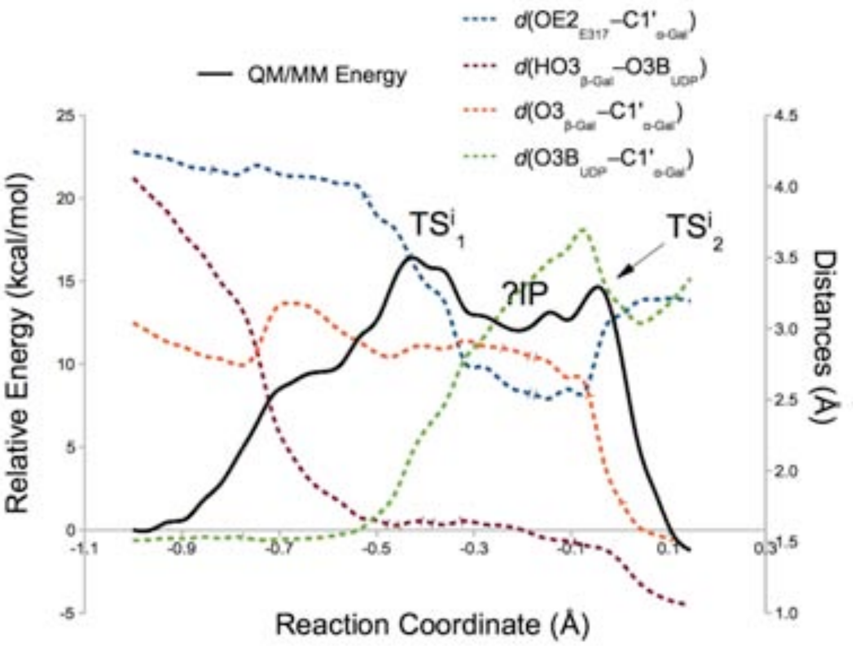

Figure 5.7. QM(BP86/SVP)/MM(CHARMM22) energy profile for the front-side attack mechanism in frame 2. Reaction coordinate (RC) = $d(\mathrm{O3B_{UDP}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{O3_{\beta\text{-}Gal}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{HO3_{\beta\text{-}Gal}}–\mathrm{O3B_{UDP}})$. The variation of several interatomic distances involved in the reaction is also depicted.

We decided to build a two-dimensional QM(BP86/SVP)/CHARMM22 potential energy surface (PES) to better appreciate the flat character of the zone describing all the oxocarbenium-like species described so far (i.e. $\mathrm{TS^d}_{1\text{-}2}$, $\mathrm{TS^i}_{1\text{-}2}$, IP) as depicted in Figure 5.8 To build this PES, from every structure coming from the reaction path used to model the CGE formation (RC = $[d(\mathrm{O3B_{UDP}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{OE2_{E317}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{HO3_{\beta\text{-}Gal}}–\mathrm{O3B_{UDP}})]$) we fixed $d(\mathrm{OE2_{E317}}–\mathrm{C1'_{\alpha\text{-}Gal}})$ and simulated a front-side attack mechanism (RC = $[d(\mathrm{O3B_{UDP}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{O3_{\beta\text{-}Gal}}–\mathrm{C1'_{\alpha\text{-}Gal}}) - d(\mathrm{HO3_{\beta\text{-}Gal}}–\mathrm{O3B_{UDP}})]$). For clarity, only $d(\mathrm{OE2_{E317}}–\mathrm{C1'_{\alpha\text{-}Gal}})$ and $d(\mathrm{O3_{\beta\text{-}Gal}}–\mathrm{C1'_{\alpha\text{-}Gal}})$ are represented in the *y* and *x*-axis, respectively. The reactants and products minima are located at the top-right corner and the middle left border, respectively. Two possible pathways connect them: one going through the CGE, the other one corresponding to the $\mathrm{S_N}$i-like mechanism. Notice the well-defined flat zone (within a range of 2-3 kcal/mol) surrounding the IP, which evidences the idea that the different oxocarbenium-like species have very similar potential energies. In addition, notice that $\mathrm{TS^i}_2$ and $\mathrm{TS^d}_2$ are geometrically very similar

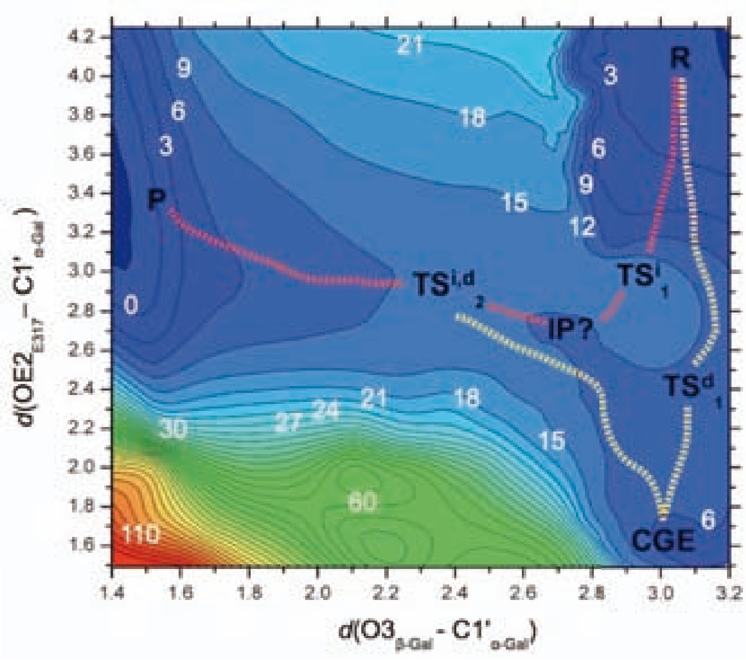

Figure 5.8. Two-dimensional QM(BP86/SVP)/CHARMM potential energy surface for frame 2. Energies are given in kcal/mol and distances in Å. Contour lines are drawn in intervals of 3 kcal/mol.

(See Table 5.2). This flatness of the PES is probably the reason why a systematic validation of all the IP and TSs by frequency calculations has not always been possible.

Therefore, galactosyl transfer catalyzed by α1,3-GalT seems to have access to an ensemble of oxocarbenium-like species in which the oxocarbenium (and C1'$_{\alpha\text{-Gal}}$) is critically poised between the three nucleophiles (O3B$_{UDP}$, O3$_{\beta\text{-Gal}}$ and OE2$_{E317}$).

In our preliminar study, a single TS corresponding to the S$_N$i mechanism (Figure 1.6B) was characterized and no IP could be found; these differences with the present results are probably because at that time the initial exploration of the PES was carried out at the SCC-DFTB/CHARMM22 level of calculation. In principle, we are more confident with the latest results as we have applied a higher level of theory and we have sampled over different enzyme:substrates configurations. Nevertheless, in systems presenting a PES with these characteristics the difference between a S$_N$i and a S$_N$i-like mechanism may be quite subtle and, thus, it is difficult to assess both from theoretical and experimental methods. Fortunately, the energetic and structural similarities between the oxocarbenium species found for this system suggests that such strict differentiation between these two mechanisms may actually be irrelevant from a more general and practical point of view.

On the other hand, the comparison between the double displacement and a front-side attack mechanisms is of much more interest. Averaging over the different frames the highest energy barrier calculated in each case, we obtain barrier heights of 16.2 ± 3.2 and 16.9 ± 2.9 kcal/mol for the double displacement and the front-side attack, respectively. Both values are very similar and in agreement with the experimentally derived phenomenological free energy of activation (~17 kcal/mol). In conclusion, our new results suggest that for α1,3-GalT both mechanisms could be acting at the same time in a competitive or complementary manner.

5.3.2.3. Analysis of factors contributing to catalysis. A detailed analysis of the factors modulating catalysis in α1,3-GalT is presented in the next subsections. At first, the role of Glu317 along the front-side attack mechanism was reassessed.

5.3.2.3.1. Nucleophilically assisted catalysis in α1,3-GalT. Looking at the changes in $d(\mathrm{OE2_{E317}–C1'_{\alpha\text{-}Gal}})$ distance for the front-side attack (Table 5.1 and Figure 5.8) it becomes apparent that Glu317 also participates in this mechanism (i.e. $d(\mathrm{OE2_{E317}–C1'_{\alpha\text{-}Gal}})$ = 4.25, 3.16, 2.51 and 2.61 Å in R, $\mathrm{TS^i_1}$, IP and $\mathrm{TS^i_2}$, respectively). This trend was already observed in our previous results where we proposed that Glu317 would also be a crucial residue in a front-side attack mechanism (See Table A5.1 and Figure 5.2).

5.3.2.3.1.1. E317A and E317Q mutants. To shed more light on the role of Glu317, we decided to perform a more quantitative analysis by modeling the front-side attack mechanism for the *in silico* E317A and E317Q mutants. As can be seen in Figure 5.9, much sharper energy profiles are obtained for the mutants when compared to the WT enzyme, and just a maximum is identified ($\mathrm{S_Ni}$ mechanism, Figure 1.6B). The energy barriers corresponding to the TSs are of 30.9 and 26.0 kcal/mol for the E317A and E317Q mutants, respectively. This destabilization of the TSs correlates with a worst stabilization of the positive charge on the α-Gal ring, which makes UDP departure to

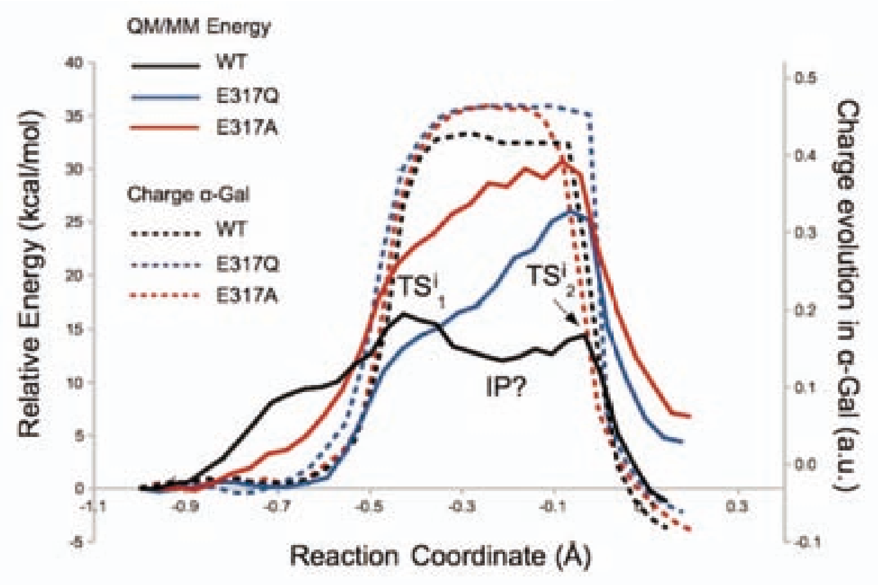

Figure 5.9. QM(M05-2X/TZVP//BP86/SVP)/CHARMM energy profiles along the $d(\mathrm{O3B_{UDP}{-}C1'_{\alpha\text{-}Gal}}) - d(\mathrm{O3_{\beta\text{-}Gal}{-}C1'_{\alpha\text{-}Gal}}) - d(\mathrm{HO3_{\beta\text{-}Gal}{-}O3B_{UDP}})$ reaction coordinate for the wild-type enzyme and the E317A and E317Q mutants. The charge evolution at the α-Gal ring along the reaction path is depicted as dashed lines.

become less favorable. Therefore, it is confirmed that even in the case of a front-side attack mechanism, the role of Glu317 is very noticeable and could also explain the mutagenesis experimental results. As anticipated by Lairson et al.,[10] a nucleophile in the back side could *"push"* the UDP leaving group to form a ion-pair intermediate. In that sense, our findings suggest that some retaining GTs may actually require a strong nucleophile on the β-face of the donor sugar substrate without necessarily involving the formation of a CGE but in a *"nucleophilically assisted"* front-side attack mechanism. This fact would also explain why the CGE has not been isolated for α1,3-GalT. Again, such mechanism will only be possible for those retaining GTs presenting a well-positioned nucleophile in the active site.

5.3.2.3.2. Enzyme-substrates interactions; key enzyme residues. We calculated the electrostatic contribution of all the residues in the active space to the stabilization of the QM region for the four TSs obtained (i.e., $\mathrm{TS^d_{1\text{-}2}}$, $\mathrm{TS^i_{1\text{-}2}}$). In addition to the putative nucleophile (Glu317), other residues of the enzyme contribute to the stabilization of the oxocarbenium-like species described before (TSs and IP) (see Table 5.3).

Table 5.3. Electrostatic contribution (kcal/mol) to the stabilization of the QM region by residues surrounding the reaction center in the double-displacement and in the front-side attack mechanisms. Only residues with some contribution higher than 1.5 kcal/mol are specified. QM=(M05-2X/TZVP//BP86/SVP).

	Double-displacement mechanism		**S_Ni-like mechanism**		
Residue	**TS^d_1**	**TS^d_2**	**TS^i_1**	**IP**	**TS^i_2**
Lys359	14.43	14.03	10.68	15.04	14.33
Asp316	10.37	7.19	11.42	8.24	7.77
Arg365	7.55	8.02	4.00	8.58	8.53
Glu317	4.35	7.40	-4.68	7.10	8.10
Tyr361	3.24	3.47	2.75	3.39	3.32
Tyr139	1.40	1.47	0.90	1.50	1.57

Two general observations can be made regarding the results shown in Table 5.3. First, most of the residues result to stabilize all the species under consideration. This is not surprising since they are an ensemble of geometrically similar (generally speaking) oxocarbenium-like species. Stated differently, key residues in the active site of α1,3-GalT would support both a double-displacement and a front-side attack mechanism since their corresponding TSs and/or intermediates share common features, that is, a highly dissociative character. Secondly, four over the six residues are interacting with UDP. Thus, stabilization of the negative charge on the UDP leaving group provided by the enzyme's residues might be an important feature for catalysis and we will certainly retake this discussion in next chapters.

However, the result obtained for Glu317 in TS^i_1 may look contradictory. It implies that in this case Glu317 would have a moderate destabilizing effect, even if we have stated before that this residue would play a key role in a *"nucleophilically assisted"* front-side attack mechanism. Moreover, our results suggest that Glu317 could facilitate catalysis by *"pushing"* the leaving group since a significant shortening of the $d(\mathrm{OE2}_{\mathrm{E317}}–\mathrm{C1'}_{\alpha\text{-Gal}})$ distance is observed while modeling a front-side attack mechanism (even if this distance is not considered explicitly in the corresponding reaction coordinates) (See Tables 5.2, S5.1 and Figures 5.2, 5.8). It is worth mentioning, though, that Glu317 is basically unmoving during the catalytic mechanism, due to hydrogen bonds with LAT and Tyr278. Thus, is the anomeric carbon ($C1'_{\alpha\text{-Gal}}$) that is getting closer to $OE2_{E317}$ because of the changes in the ring puckering conformation while the UDP-Gal bond breaks (i.e. from a distorted $^4C_1$ chair ($\varphi = 298°$, $\theta = 6°$) in the reactants to a conformation close to an ^4E envelope at the TS^i_1 ($\varphi = 247°$, $\theta = 32°$)) (See Figure 5.10).

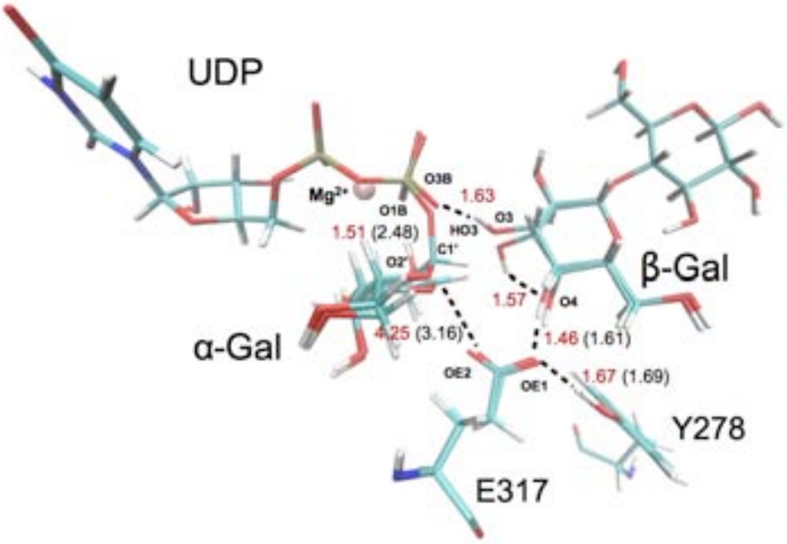

Figure 5.10. Structural superposition of reactants (R) and the first transition state in the S_Ni-like mechanism (TS^i_1) for frame 2. The donor and acceptor substrates, together with Glu317 and Tyr278, are represented as sticks. Selected distances (in Å) are indicated in red and black for R and TS^i_1, respectively.

On the other hand, TS^i_1 is an early transition state, in the sense that the leaving group is still relatively close ($d(O3_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}}) = 2.48$ Å) and therefore the anomeric center is critically poised between three nucleophiles (UDP, Glu317 and LAT). Thus, the Glu317 may actually have a destabilizing role at the beginning of the catalytic mechanism, that would ultimately drive the leaving group departure supported by the $HO3_{\beta\text{-Gal}}–O3B_{UDP}$ hydrogen-bond and interactions with residues of the enzyme. Put differently, at the very beginning the Glu317 is not facilitating the catalytic process by stabilizing the emerging positive charge on the anomeric center, but by contributing to a destabilizing scenario that promotes the leaving group departure. On the contrary, when the UDP moves farther and a pronounced positive charge develops on $C1'_{\alpha\text{-Gal}}$, the situation changes and Glu317 plays a key stabilizing role. The latest would also explain why in the case of mutants E317A/E317Q the beginning the reaction seems to be favored despite the fact that the glutamate is a stronger nucleophile; but later on, it turns out that the foregoing positive charge on the α-Gal ring is better stabilized when the glutamate is present (See Figure 5.9) All the above support the idea that Glu317 plays a significant role in the front-side attack mechanism.

5.3.2.3.3. Inter- and intra- substrate interactions. We pointed out in the previous chapter the importance of inter- and intra- substrate interactions for catalysis in LgtC. More specifically, several substrate-substrate interactions were identified to promote reaction, some of them involving LAT (See Figures 4.8 and 4.10): a hydrogen bond between $O2'_{\alpha\text{-Gal}}$ and the UDP β-phosphate ($O1B_{UDP}$), a hydrogen bond between $O3B_{UDP}$ and the $O4_{\beta\text{-Gal}}$ attacking hydroxyl of LAT, and a hydrogen bond between $O3B_{UDP}$ and the $O3_{\beta\text{-Gal}}$ hydroxyl (adjacent to the attacking $O4_{\beta\text{-Gal}}$). In α1,3-GalT only the $O2'_{\alpha\text{-Gal}}$–$O1B_{UDP}$ hydrogen bond and the one of $O3B_{UDP}$ with the attacking hydroxyl (now $O3_{\beta\text{-Gal}}$) are possible (See Figure 5.10).

Interestingly, we have found that the $O3B_{UDP}$-$HO4_{\beta\text{-Gal}}$ (LgtC) interaction is present both at the TS and at the reactants, whereas the equivalent $O3B_{UDP}$–$O3_{\beta\text{-Gal}}$ (α1,3-GalT) hydrogen bond is formed along the galactosyl transfer (e.g. *d*($HO3_{\beta\text{-Gal}}$–$O3B_{UDP}$) = 4.06 Å in R and ~1.6 Å at the TSs, see Table 5.2, Figures 5.7 and 5.10) so that it is not present at the beginning of any catalytic mechanism considered in the present study.

In the α1,3-GalT reactants, $O3_{\beta\text{-Gal}}$ is facing the neighboring $O4_{\beta\text{-Gal}}$ group, which in turn is hydrogen-bonded to $OE1_{E317}$ (See Figure 5.10). This network of hydrogen bonds was observed along all the MM and QM/MM MDs performed during the present study, suggesting that it is actually present in the Michaelis complex. It is thus a consequence of the substrates' binding orientation and interactions which ultimately defines the enzyme's specificity (i.e. α-1,3).

The latest would explain why in the case of α1,3-GalT, the *d*($HO3_{\beta\text{-Gal}}$–$O3B_{UDP}$) distance had to be explicitly included in the reaction coordinates for both the front-side attack and the double displacement mechanisms (which was not necessary the case for LgtC), even though it is not until the second step when the proton transfer takes place (see Table 5.2, Figures 5.5-5.7).

5.3.2.4. Trapping a covalent intermediate (CGE)? Our results so far suggest that in the absence of the acceptor molecule (LAT), formation of the covalent intermediate would not be possible (or extremely slow), which could explain why it is being so difficult to isolate such CGE. In fact, it is important to remember that even in the presence of LAT the CGE formation would be quite endoergic (11.4 ± 4.2 kcal/mol over the reactants) and that the CGE could easily go back to reactants (reverse barrier of 3.7 ± 1.3 kcal/mol).

We have further studied the importance of the acceptor in CGE formation by calculating the QM/MM energy profile leading to the CGE in the absence of LAT. For that, we removed LAT from the Michaelis complex, resolvated the resulting binary complex and reequilibrated the system (water molecules first, also the protein side chains afterwards) by a 2 ns MM(CHARMM22) MD. Three snapshots were randomly taken as starting points in the following QM/MM calculations. The corresponding energy profiles (See Figure 5.11) have high energy barriers (~30-38 kcal/mol) and are much more endoergic than when LAT is present (> 10 kcal/mol). Dissociation of the CGE back to reactants has a barrier of only ~5 kcal/mol. All together, these results confirm our hypothesis.

5.4. CONCLUSIONS

The results presented so far suggest that both the front-side attack and the double displacement mechanism could be operating at the same time and with similar rates in galactosyl transfer catalyzed by α1,3-GalT. Moreover, both mechanisms would require the presence of Glu317 and in that sense, our findings suggest that some retaining GTs may actually require a strong nucleophile on the β-face of the donor sugar substrate

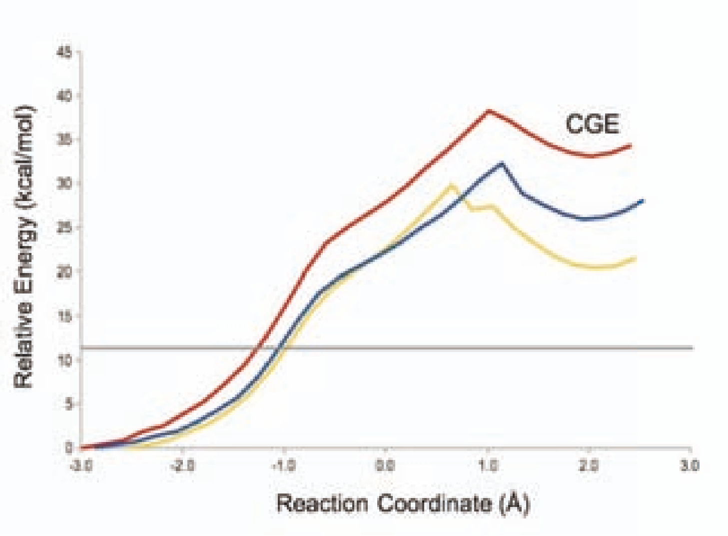

Figure 5.11. QM(M05-2X/TZVP//BP86/SVP)/CHARMM energy profiles for CGE formation when the acceptor LAT is not in the active site of α1,3-GalT. Reaction Coordinate (RC) = *d*(O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$) - *d*(OE2$_{E317}$–C1'$_{\alpha\text{-Gal}}$). Three different snapshots from a MD simulation were considered. The line at 11.4 kcal/mol indicates the average energy of the CGE formed in the original ternary complex (with LAT).

without necessarily involving the formation of a CGE in what we have defined as a *"nucleophilically assisted"* front-side attack mechanism. Our results for the *in silico* E317Q and E317A mutants support this statement, since the corresponding energy barriers for the front-side attack mechanism increase over 10 kcal/mol. In conclusion, the essential role of Glu317 in the galactosyl transfer catalyzed by bovine α1,3-GalT has now been confirmed computationally. Unfortunately, and since its presence seems to be crucial for all the main possible mechanisms proposed in the literature (double-displacement and front-side attack), the interpretation of the available mutagenesis data is hampered.

On the other hand, most of the stabilizing residues in the active site of α1,3-GalT (i.e. Tyr139, Lys359, Tyr361 and Arg365) are interacting with the UDP, but more importantly, interactions between the acceptor molecule and the leaving group are crucial in order to support the UDP departure and thus achieve the reported k_{cat} values. Finally, our results show that the CGE formation is not affordable unless the ternary complex is formed (i.e. in the presence of the aceptor substrate), which may partly explain why it is being so difficult to isolate the covalent intermediate by the experimentalists.

6

Further comparison of α1,3-GalT and LgtC catalytic mechanisms

6.1. INTRODUCTION

The catalytic mechanisms of α1,3-GalT and LgtC have been described in the previous chapters. These enzymes can be considered as models of retaining GTs where a strong nucleophile is present or not in the β-face of the donor sugar substrate (i.e. Glu317 and Gln189 for α1,3-GalT and LgtC, respectively). In case of LgtC, a S_Ni mechanism was described and we have not found any evidence of a double-displacement mechanism for the wild-type enzyme. On the other hand, for α1,3-GalT we were able to characterize a double-displacement and a S_Ni-like mechanism, which could be acting in a competitive or complementary manner. More interestingly, we have found that Glu317 plays an important role even in a front-side attack mechanism to facilitate the leaving group (UDP) departure in α1,3-GalT.

Why would some retaining GTs require a nucleophile in the β-face of the donor sugar substrate to facilitate the leaving group departure? In other words, is UDP-sugar bond cleavage more difficult in some retaining GTs that in others? Notice that in all the proposed mechanisms the beginning of the reaction is dominated by UDP-Gal bond cleavage. Thus, the main goal of the present chapter will be to get an answer to the foregoing question by comparing the leaving group departure process in the ternary complex for these two retaining GTs.

6.2. MODELS AND METHODS

Models considered for LgtC and α1,3-GalT, were already described in chapters 4 and 5, respectively.

In the same way, all the QM/MM calculations were performed as before. We only need to specify that a Natural Bond Orbital (NBO) analysis[240,262-264] was also performed for some structures, using the NBO program v3.1[265] included in Gaussian09.[266]

6.3. RESULTS AND DISCUSSION

The QM/MM energy profiles for the UDP-Gal bond cleavage process are depicted in Figure 6.1A. Surprisingly, and despite the presence of Glu317, bond cleavage following the $d(\mathrm{O3B}_{\mathrm{UDP}}\text{–}\mathrm{C1'}_{\alpha\text{-Gal}})$ coordinate is much more difficult in α1,3-GalT than in LgtC. In the results exposed before for LgtC (chapter IV), only a moderate electrostatic stabilization of the TS by Gln189 was estimated (~2 kcal/mol). In α1,3-GalT, Glu317 was expected to provide a better stabilization of the positive charge developed in α-Gal but, apparently, its contribution is important but not *"enough"* to make the bond breakage sufficiently easy (See also Figure 6.1B). It is important to remind here that, despite this difference in bond cleavage, the two enzymes exhibit similar experimentally and theoretically derived energy barriers (~17 kcal/mol for α1,3-GalT[259] and ~16 kcal/mol for LgtC[61,86,90]). One of the main goals of the present chapter is

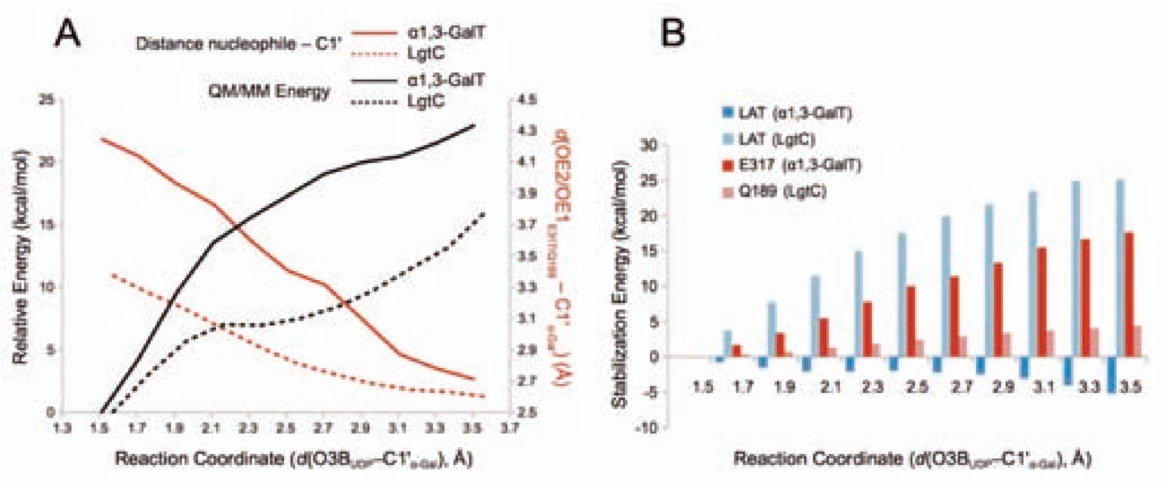

Figure 6.1. Comparative analysis of the bond-breakage process in α1,3-GalT and LgtC at the QM(M05–2X/TZVP//BP86/SVP) level of calculation. (**A**) Energy profile for the bond-breakage process and variation of the *d*(OE2/OE1$_{E317/Q189}$–C1'$_{\alpha\text{-Gal}}$) distance. (**B**) Electrostatic contribution to the stabilization of the QM region by the acceptor substrate (LAT) and the active site nucleophile (Glu317 or Gln189 in α1,3-GalT or LgtC, respectively).

then to clarify this alleged inconsistency that we have found in the bond-breakage readiness when comparing α1,3-GalT and LgtC.

We started by analyzing the charge evolution in the donor substrate as the O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$ bond is broken (Figure 6.2). The positive charge development on α-Gal is very similar for both enzymes, whereas α1,3-GalT develops a higher negative charge on UDP (-0.51 a.u. *vs*. -0.41 a.u.). The data in Figure 6.1A, thus, indicate that although it has always been suggested that stabilization of the anomeric carbon positive charge is a key factor in glycosyl transfer, stabilization of the charge on the UDP leaving group is also crucial and it seems to be making an important difference between these two enzymes.

6.3.1. Contribution from the enzyme's residues. An analysis of the active site residues capable of stabilizing this charge development was performed. We calculated, for all the residues in the active space, their electrostatic contribution to the stabilization of the QM region as the O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$ bond is broken. The contribution from the putative nucleophile (Glu317 or Gln189) is shown in Figure 6.1B. Other residues that

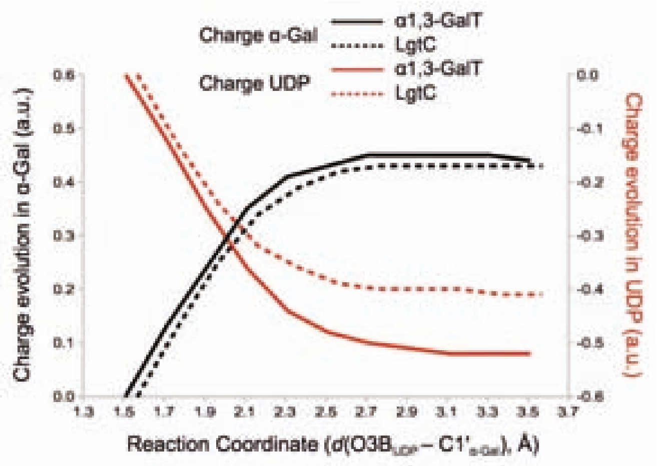

Figure 6.2. Comparative analysis of the bond-breakage process in LgtC and α1,3-GalT. QM=(M05-2X/TZVP//BP86/SVP). The charge evolution of α-Gal and UDP along the *d*($O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$) reaction coordinate is depicted.

contribute significantly are depicted in Figure 6.3. As it can be seen, Glu317 contribution in α1,3-GalT is much more significant (four times higher) than that of Gln189 in LgtC. Curiously, we found more stabilizing residues in the active site of α1,3-GalT than in LgtC, most of them localized in the upper face of the UDP-Gal. Thus, according to the different stabilization of the negative charge on the UDP provided by the enzyme's residues, bond cleavage in α1,3-GalT would be expected to be easier than in LgtC.

6.3.2. Inter-substrates interactions. If stabilization by the amino acidic residues in the active site is not responsible for the easier leaving group departure calculated for LgtC when compared to α1,3-GalT, other explanations need to be found. An analysis of the stabilization provided by the substrates themselves (inter-substrates interactions) during the cleavage of the $O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$ bond was then performed. In particular, the contribution of the acceptor substrate (LAT) was calculated.

As Figure 6.1B shows, the electrostatic contribution of LAT to the stabilization of the QM region along the *d*($O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$) bond-cleavage coordinate is completely different between the two enzymes: in LgtC, the inter-substrates interaction of UDP-Gal with LAT clearly facilitates the cleavage of the glycosidic bond, whereas in α1,3-GalT,

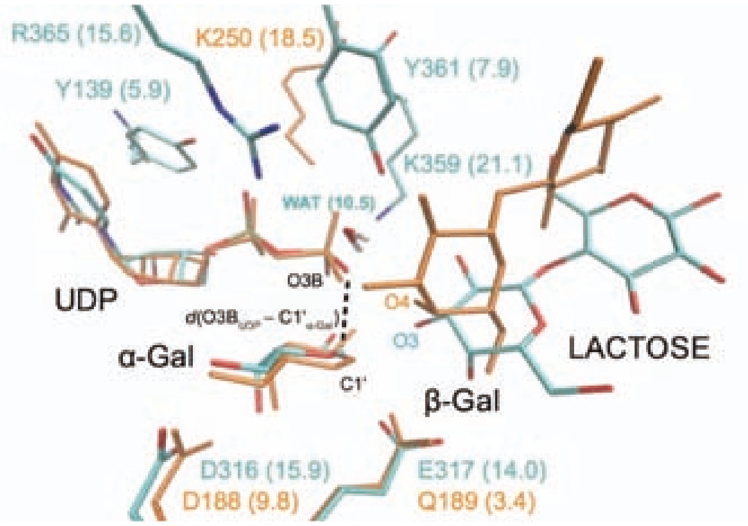

Figure 6.3. Main electrostatic contributions (kcal/mol) of residues in the active site of α1,3-GalT (cyan) and LgtC (orange) to the stabilization of the QM region in the $O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$ bond–breakage process. The structures with $d(O3B_{UDP}$–$C1'_{\alpha\text{-Gal}})$ ~3 Å were considered to compute the energy differences with the reactants.

it slightly destabilizes it. Thus, the different orientation of the acceptor substrate in the active site of these enzymes is conditioning not only their specificity (α1-3 and α1-4 in α1,3-GalT and LgtC, respectively) but also their ability to assist $O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$ bond breakage. A detailed analysis of the stabilizing role of LAT follows.

6.3.2.1. Substrate-assisted catalysis in α1,3-GalT and LgtC. According to the above analysis, α1,3-GalT needs a strong nucleophile because, contrary to what is found in LgtC, UDP is a bad leaving group as LAT is not able to support the initial UDP-Gal bond cleavage in α1,3-GalT. However, this nucleophilic assistance is not enough to achieve the reported reaction rate (which is of the same order as that of LgtC) and new interactions must be formed during galactosyl transfer. In our work on LgtC several substrate-substrate interactions were identified to promote reaction, some of them involving LAT (See Figure 4.3): a hydrogen bond between $O2'_{\alpha\text{-Gal}}$ and the UDP β-phosphate ($O1B_{UDP}$), a hydrogen bond between $O3B_{UDP}$ and the $O4_{\beta\text{-Gal}}$ attacking hydroxyl of LAT, and a hydrogen bond between $O3B_{UDP}$ and the $O3_{\beta\text{-Gal}}$ hydroxyl (adjacent to the attacking $O4_{\beta\text{-Gal}}$). In α1,3-GalT only the $O2'_{\alpha\text{-Gal}}$–O_{UDP} hydrogen bond and the one of $O3B_{UDP}$ with the attacking hydroxyl (now $O3_{\beta\text{-Gal}}$) are possible. Nonetheless, the stabilization by $O3_{\beta\text{-Gal}}$ (LgtC) was estimated to be of ~ 2 kcal/mol, which is not enough to explain the differences observed in Figure 6.1A. Interestingly, we have found that the $O3B_{UDP}$–$HO4_{\beta\text{-Gal}}$ (LgtC) interaction is present both at the TS and at the reactants, whereas the equivalent $O3B_{UDP}$–$O3_{\beta\text{-Gal}}$ (α1,3-GalT) hydrogen bond is formed along the galactosyl transfer so that it is not present to stabilize the beginning of the bond breakage process (e.g. $d(HO3_{\beta\text{-Gal}}–O3B_{UDP})$= 4.06 Å in R and ~ 1.6 Å at the TSs, see Table 5.2). Therefore, the initial orientation of $HO3_{\beta\text{-Gal}}$ can be at the origin of the differences observed in UDP-Gal bond cleavage between α1,3-GalT and LgtC. To test this hypothesis we forced a hydrogen bond between $O3_{\beta\text{-Gal}}$ and $O3B_{UDP}$ in the reactants of α1,3-GalT and calculated again the energy for breaking the UDP-Gal bond (Figure 6.4A).

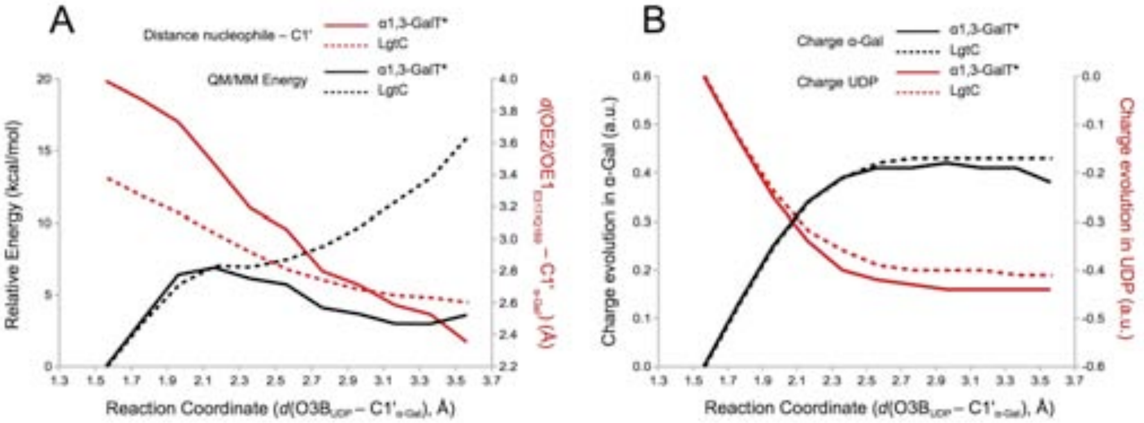

Figure 6.4. Comparative analysis of the bond-breaking process in LgtC and α1,3-GalT after imposing the O3 atom from LAT to be initially hydrogen-bonded to O3B in the reactants (α1,3-GalT*). QM=(M05-2X/TZVP//BP86/SVP). (**A**) Energy profile for bond-breakage process and variation of the distance between the putative nucleophile and the anomeric center. (**B**) Charge evolution of α-Gal and UDP along the $d(O3B_{UDP}–C1'_{\alpha\text{-Gal}})$ reaction coordinate.

Comparing Figures 6.1A and 6.4A, it can be seen that when the $HO3_{\beta\text{-Gal}}$ is initially facing its final acceptor the leaving group is more easily released as the developing negative charge on UDP gets stabilized from the beginning (See Figure 6.4B and compare to Figure 6.2), thus confirming our hypothesis. Moreover, when we model the double-displacement and front-side attack mechanisms starting from these *"artificial"* ternary complex where an initial $O3_{\beta\text{-Gal}}$–$O3B_{UDP}$ hydrogen bond is forced, the potential energy barriers are ~ 10 kcal/mol less than the originally reported ones (~ 16 kcal/mol). The new energy barriers are inconsistent with the available experimental data and therefore it can be assumed that $O3_{\beta\text{-Gal}}$ is actually not hydrogen-bonded to $O3B_{UDP}$ in the Michaelis complex, as a consequence of the initial orientation of the substrates in the active site of α1,3-GalT, which ultimately defines its specificity (α1-3).

A detailed study of the magnitude and nature of these inter-substrates interactions in galactosyl transfer by α1,3-GalT and LgtC was then carried out by a full donor-acceptor Natural Bond Orbitals (NBO) analysis where the putative nucleophiles were also included (Table 6.1). The previously characterized TS for LgtC and the equivalent one for α1,3-GalT, more specifically the TS^{i}_{2} obtained for the front-side attack mechanism, have been compared. Some catalytically relevant interacting molecular orbitals are depicted in Figures 6.5 and in Appx., Figures A6.1 and A6.2.

Three main conclusions can be outlined from these results. First, the largest contributions to TS stabilization involve interactions of the incoming hydroxyl group from LAT ($O4_{\beta\text{-Gal}}$ and $O3_{\beta\text{-Gal}}$ in LgtC and α1,3-GalT, respectively) with the final acceptor of the proton, $O3B_{UDP}$ (Table 6.1; Figures 6.5B, E). Interestingly, a interaction between $O4_{\beta\text{-Gal}}$(LgtC)/$O3_{\beta\text{-Gal}}$(α1,3–GalT) and a molecular orbital involving $(C1'–O5')_{\alpha\text{-Gal}}$ is also seen to participate in TS stabilization (Table 6.1; Appx., Figures A6.1A, B). In the case of LgtC, the interaction of the neighboring hydroxyl group with $O3B_{UDP}$ mentioned above is also highlighted in the analysis (Table 6.1, Figures 6.5C and S6.2A). Secondly, it is confirmed that a major difference between the two enzymes is that this hydrogen bond between $O3B_{UDP}$ and $O4_{\beta\text{-Gal}}$(LgtC)/$O3_{\beta\text{-Gal}}$(α1,3-GalT) is already present in the reactants only for LgtC. In the α1,3-GalT reactants, $O3_{\beta\text{-Gal}}$ is facing the neighboring $O4_{\beta\text{-Gal}}$ group, which in turn is hydrogen-bonded to $OE1_{E317}$ (Figure A6.2B). This network of hydrogen bonds was observed along all the MM and QM/MM MDs performed, suggesting that it is present in the Michaelis complex, as was mentioned in the previous chapter. This explains why $O3B_{UDP}$–$C1'_{\alpha\text{-Gal}}$ bond cleavage,

Table 6.1. Donor-Acceptor Natural Bond Orbitals (NBO) analysis at the QM(M05-2X/TZVP//BP86/SVP)/CHARMM22 level. Previously characterized TS for LgtC in chapter 4 and the equivalent one for α1,3-GalT (TS$^{i}_{2}$) were considered. LP: lone pair; BD: bonding molecular orbital; BD*: antibonding molecular orbital. Only the main interacting pairs involving the nucleophile and/or the ligands are given. ΔE (in kcal/mol) corresponds to the energy difference considering the same interacting MO in the TS and the reactants. The reference to the figure in which some of them are depicted is given within parentheses. $^{\&}$Interactions that were already present in the reactants.

LgtC			**α1,3-GalT**		
Donor NBO	Acceptor NBO	ΔE (kcal/mol)	Donor NBO	Acceptor NBO	ΔE (kcal/mol)
Putative Nucleophile (Gln189/Glu317)					
LP$_2$ (OE1)$_{\text{Q189}}$	BD*$_1$ (C1'-H1')$_{\alpha\text{-Gal}}$	1.42 (A6.1C)	BD$_2$ (CD-OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	2.87 (A6.1D)
BD$_2$ (CD-OE1)$_{\text{Q189}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	1.19	LP$_1$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	2.74
			LP$_2$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	1.53
β-Gal (LAT)					
LP$_3$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	96.05 (6.5B)	LP$_3$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	46.05 (6.5E)
LP$_2$ (O4)$_{\beta\text{-GAL}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	44.56 (A6.1A)	LP$_1$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	11.00
LP$_2$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	15.49 (6.5C)	LP$_1$ (O3)$_{\beta\text{-Gal}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	5.99 (A6.1B)
LP$_1$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	6.98$^{\&}$	BD*$_1$ (PB-O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	1.94
LP$_1$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	5.48$^{\&}$ (A6.2A)	LP$_2$ (O4)$_{\beta\text{-Gal}}$	BD* (C1'-H1')$_{\alpha\text{-Gal}}$	1.92
BD*$_1$ (PB-O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	5.99			
BD$_1$ (O4-HO4)$_{\text{BGAL}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	5.70			
LP$_1$ (O4)$_{\beta\text{-GAL}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	2.80			
CR$_1$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	2.47			
BD*$_1$ (PB-O3B)$_{\text{UDP}}$	BD*$_1$ (O3-HO3)$_{\beta\text{-Gal}}$	2.41			
BD$_1$ (PB-O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	1.55			
LP$_2$ (O3B)$_{\text{UDP}}$	BD*$_1$ (O4-HO4)$_{\beta\text{-Gal}}$	-17.45$^{\&}$ (6.5A)			

described only by the *d(*O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-Gal}}$) coordinate, is initially more difficult in α1,3-GalT than in LgtC (despite the assistance by Glu317 in the former). It would also explain why in α1,3-GalT the *d*(HO3$_{\beta\text{–Gal}}$–O3B$_{\text{UDP}}$) distance had to be explicitly included in the reaction coordinate for both the front-side attack and the double displacement mechanisms, even though it is not until the second step when the transfer of HO3$_{\beta\text{–Gal}}$ to O3B$_{\text{UDP}}$ takes place (see Table 5.2 and Figure 5.6). During the sugar transfer reaction, reorientation of the incoming hydroxyl to form the O3$_{\beta\text{-Gal}}$–O3B$_{\text{UDP}}$

hydrogen bond will facilitate UDP departure and lead to energy barriers of the same order for both enzymes. Finally, contributions from the nucleophiles to TSs stabilization have also been identified and are quantified to be comparatively much less important than those from the inter-substrate interactions, especially in LgtC. In the case of α1,3-GalT, the stabilizing effect by Glu317 is more substantial (~7 kcal/mol, see Table 6.1).

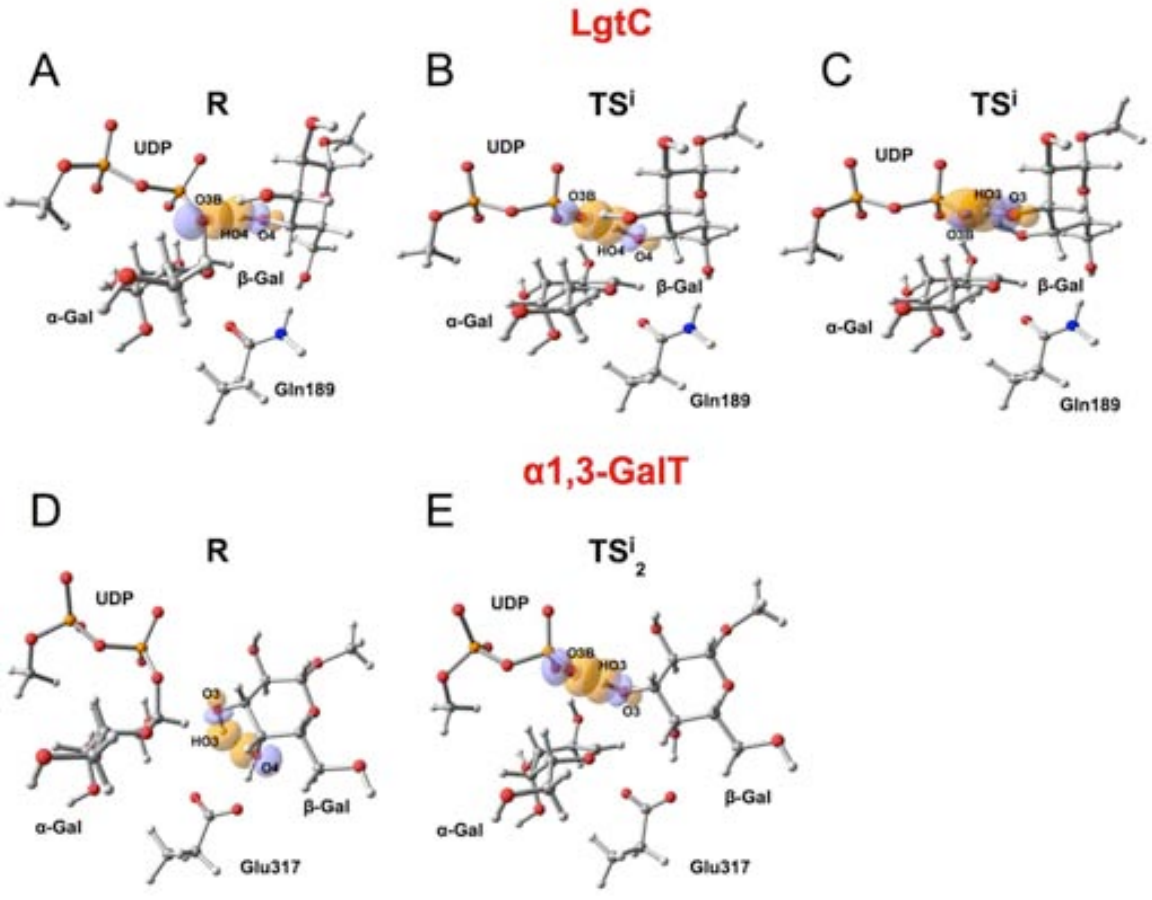

Figure 6.5. Interactions between molecular orbitals of the substrates relevant in galactosyl transfer according to a NBO analysis. These interactions involve: for LgtC, $O3B_{UDP}$ and the incoming $O4_{\beta\text{-Gal}}$ in the reactants (**A**) and in the S_Ni transition state (**B**), or the neighboring hydroxyl ($O3_{\beta\text{-Gal}}$), here only shown in the transition state (**C**); for α1,3-GalT, $O3_{\beta\text{-Gal}}$ interacts with $O4_{\beta\text{-Gal}}$ in the reactants (**D**) and with $O3B_{UDP}$ in the TS^{i}_{2} transition state (**E**). For clarity, just a fraction of the QM atoms is shown.

6.3.3. Nucleophilic strength of Glu317. Results so far suggest that Glu317 is not such a strong nucleophile as primarily expected. Moreover, as described in the previous chapter, the formation of the covalent intermediate (CGE) would be unfeasible unless the acceptor molecule is present in the active site of α1,3-GalT. More especifically, when we removed LAT from the Michaelis complex, resolvated the resulting binary complex and reequilibrated the system, the corresponding energy profiles for the CGE formation had high energy barriers (~ 30-38 kcal/mol) and were much more endoergic than when LAT is present (> 10 kcal/mol).

Interestingly, if the system is not further resolvated after LAT removal, so that the nucleophile only interacts with the α-Gal, CGE formation is *"artificially"* observed when the UDP-Gal bond is broken following the RC = $d(\mathrm{O3B_{UDP}}–\mathrm{C1'_{\alpha\text{-}Gal}})$ (Figure 6.6A). Notice that in the case of LgtC the effect is quite different; that is, the bond-breakage process becomes much less favorable (Figure 6.6B). Therefore, interaction of Glu317 with LAT or with water molecules seems to be reducing its nucleophilic character.

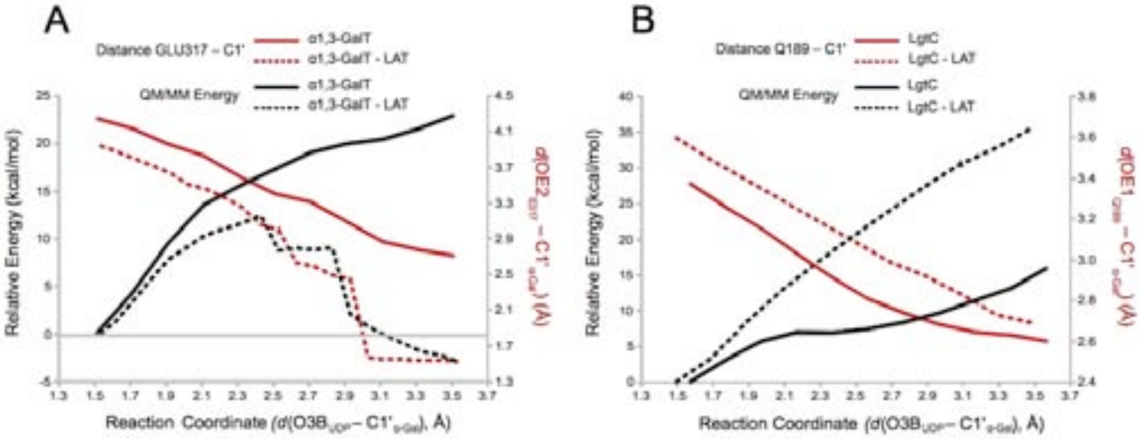

Figure 6.6. QM(M05-2X/TZVP//BP86/SVP)/CHARMM22 energy profiles along the $d(\mathrm{O3B_{UDP}}–\mathrm{C1'_{\alpha\text{-}Gal}})$ reaction coordinate for (A) α1,3-GalT and (B) LgtC with and without (- LAT) the acceptor substrate in the active site. The variation of the distance between the nucleophiles Glu317/Gln189 and the anomeric center is also depicted for α1,3-GalT and LgtC, respectively.

Table 6.2. Donor-Acceptor Natural Bond Orbitals (NBO) analysis for the bond-breaking process in the wild-type (WT) α1,3-GalT as well as after removing the LAT from the initial complex and resolvating (WT - LAT + H_2O) or not (WT - LAT). QM = M05-2X/TZVP//BP86/SVP. Structures where *d*(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-Gal}}$) ~2.8 Å were considered. LP: lone pair; BD: bonding molecular orbital; BD*: antibonding molecular orbital; CR: core pair. Only the main interacting pairs involving the nucleophile and/or the ligands are given. ΔE corresponds to the energy difference between the interacting molecular orbitals in these structures and the reactants.

Donor NBO	**Acceptor NBO**	**ΔE (kcal/mol)**
WT (α1,3-GalT + UDP-Gal + LAT)		
BD$_2$ (CD-OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.53
LP$_1$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.29
BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	BD*$_2$ (CD-OE2)$_{\text{E317}}$	0.18
LP$_2$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.12
WT – LAT + H_2O		
BD$_2$ (CD-OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.22
LP$_2$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.07
WT – LAT		
LP$_3$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	8.20
LP$_1$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	4.13
LP$_2$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	3.40
BD$_1$ (CD-OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.53
CR$_1$ (OE2)$_{\text{E317}}$	BD*$_2$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.19
LP$_3$ (OE2)$_{\text{E317}}$	BD*$_1$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.19
LP$_2$ (OE2)$_{\text{E317}}$	BD*$_1$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.12
LP$_1$ (OE2)$_{\text{E317}}$	BD*$_1$ (C1'-O5')$_{\alpha\text{-Gal}}$	0.10

Again, a NBO analysis of the different scenarios provides an explanation for the latest outcomes (See Table 6.2). As mentioned before, when LAT is present in the active site there is a hydrogen bond between O4$_{\beta\text{-Gal}}$ and OE1$_{\text{E317}}$; as a result of that, the (CD–OE2)$_{\text{E317}}$ bond acquires a double bond character and the negative charge on OE2$_{\text{E317}}$ is relatively moderate (Figure 6.7A). In this context, the antibonding (BD*) (C1'–O5')$_{\alpha\text{-Gal}}$ molecular orbital interacts with molecular orbitals of the Glu317 with a total interaction energy of 1.12 kcal/mol in favor of UDP departure (for *d*(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-}}$

$_{\text{Gal}}$) = 2.81 Å). When LAT is removed and the system resolvated, some water molecules come to interact with Glu317 and a similar scenario is found (Figure 6.7B), but with a less significant interaction energy (0.29 kcal/mol, for *d*(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-Gal}}$) = 2.89 Å). However, in the absence of the acceptor substrate and water molecules, the (CD–OE2)$_{\text{E317}}$ bond would no longer have a double–bound character (the (CD–OE1)$_{\text{E317}}$ bond has) and the OE2$_{\text{E317}}$ atom that is facing the anomeric center would exhibit a higher negative charge. In this case, stronger interactions would be established between Glu317 and α-Gal (16.9 kcal/mol, at *d*(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-Gal}}$) = 2.89 Å, Figure 6.7C). Thus, in this context, Glu317 is assuming a more efficient role in the leaving-group departure so that, as mentioned above, the CGE would be readily formed (Figure 6.6A).

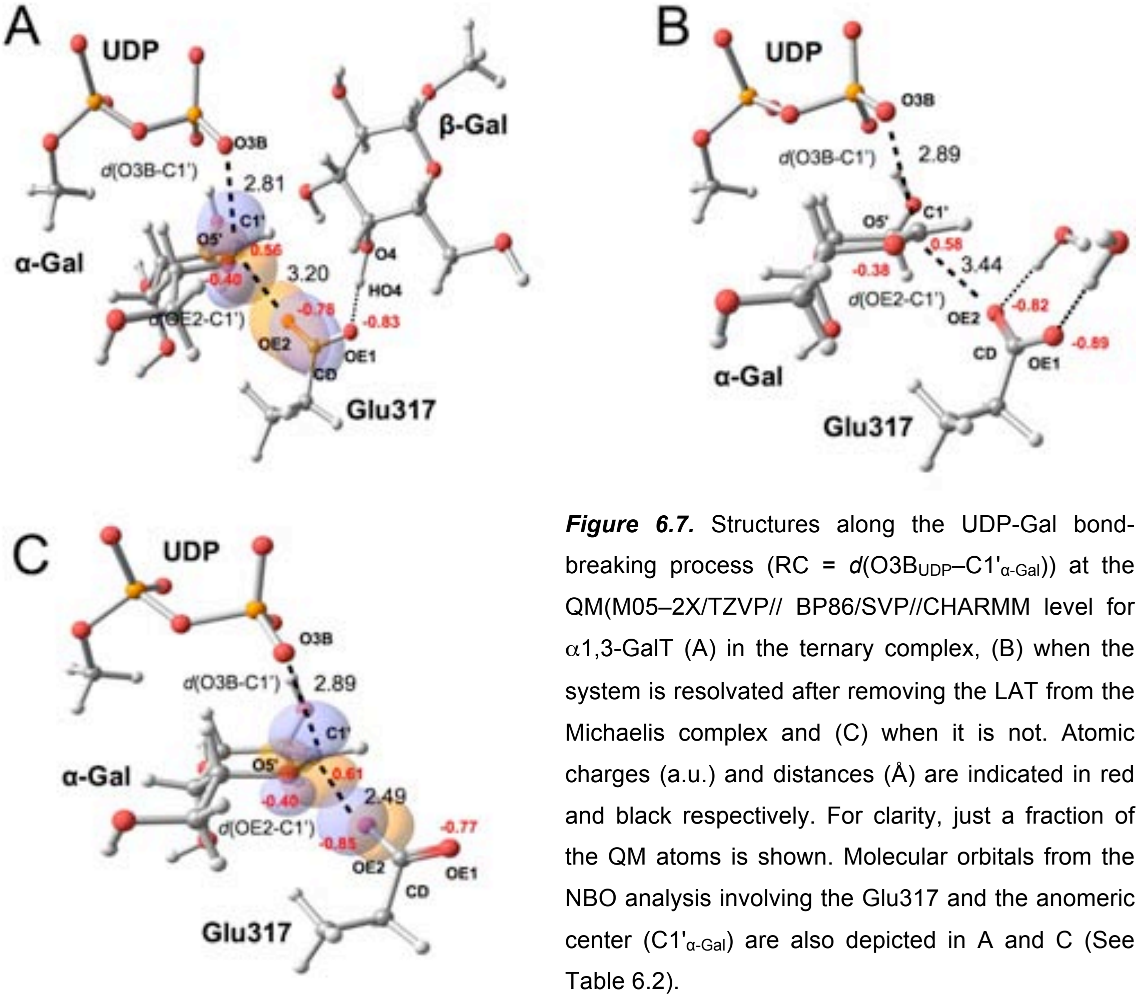

Figure 6.7. Structures along the UDP-Gal bond-breaking process (RC = *d*(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-Gal}}$)) at the QM(M05–2X/TZVP// BP86/SVP//CHARMM level for α1,3-GalT (A) in the ternary complex, (B) when the system is resolvated after removing the LAT from the Michaelis complex and (C) when it is not. Atomic charges (a.u.) and distances (Å) are indicated in red and black respectively. For clarity, just a fraction of the QM atoms is shown. Molecular orbitals from the NBO analysis involving the Glu317 and the anomeric center (C1'$_{\alpha\text{-Gal}}$) are also depicted in A and C (See Table 6.2).

The previous analysis shows that the nucleophilic strength of Glu317 in α1,3-GalT is reduced by its interaction with LAT so that Glu317 alone cannot promote the formation of a covalent intermediate and substrate-assisted catalysis (curiously by the same LAT) is required. Moreover, the interaction of Glu317 with water molecules also reduces its nucleophilic character; which is also supported by the fact that in this case no *"spontaneous"* CGE formation is obtained while simulating the leaving group departure (See Figure 6.8A). Interestingly, in the case of LgtC, to resolvate or not the binary complex (i.e. LgtC + UDP-Gal) does not make a real difference, which confirms the minor role of Gln189 in the leaving group departure in LgtC (See Figure 6.8B).

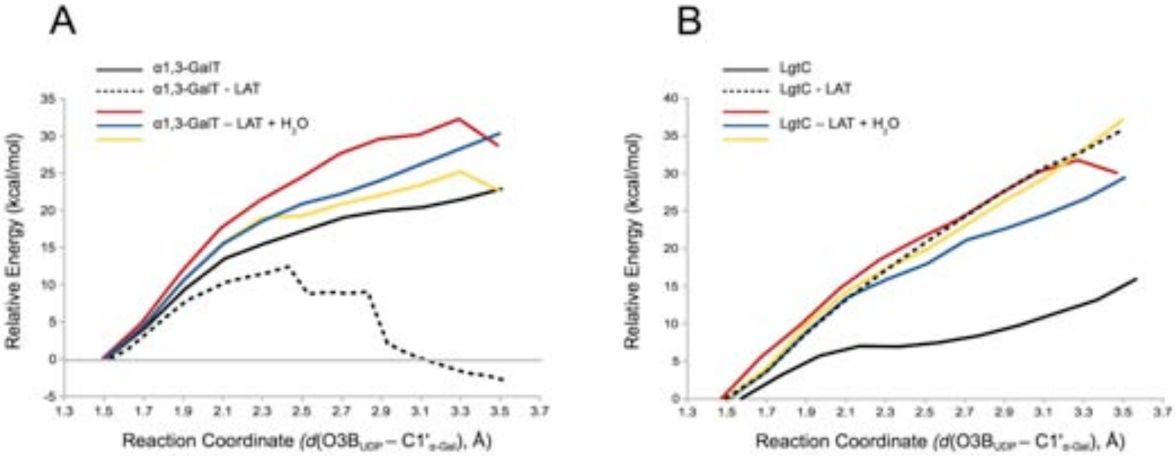

Figure 6.8. QM(M05-2X/TZVP//BP86/SVP)/CHARMM22 energy profiles along the *d*(O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$) reaction coordinate for (**A**) α1,3-GalT and (**B**) LgtC with and without (- LAT) the acceptor substrate as well as after resolvating the initial complex once LAT was removed (LgtC - LAT + H_2O). For the latter, the results for three random frames of a MM(CHARMM22) MD equilibration are depicted as red, yellow and blue lines.

We then suggest that, altogether with the endoergicity for CGE formation and its low stability, the interactions of Glu317 with water molecules and the acceptor ligand in the binary and Michaelis complexes, respectively, are the reasons why a covalent intermediate is not experimentally detected for the α1,3-GalT WT enzyme.

6.4. CONCLUSIONS

By studying in detail galactosyl transfer catalyzed by α1,3-GalT and LgtC, (i.e. taking into account the results exposed in chapters 4-6), and comparing them, we have shown that LgtC binds the substrates in a relative orientation very convenient for catalysis as substrate-substrate interactions can be readily established that efficiently participate in the stabilization of the β-phosphate negative charge (*substrate-assisted catalysis*). On the contrary, the binding orientation and interactions that donor and acceptor must adopt in α1,3-GalT in order to achieve the desired reaction specificity (α1-3 linkage), reduces the number of interactions that facilitate initial UDP-Gal bond cleavage, requiring then a stronger nucleophile (Glu317) in the β-face of UDP-Gal to assist initial leaving group departure (*nucleophilically-assisted catalysis*) (See Figure 6.9). In particular, a hydrogen bond between the β-phosphate of the leaving group and the attacking hydroxyl of the acceptor molecule is not present in the α1,3-GalT Michaelis complex. However, the pushing effect by Glu317 is not enough, but the attacking hydroxyl will reorient during the reaction to form the hydrogen bond with the β-phosphate in the TS (TS stabilization), which finally result in similar energy barriers for both enzymes.

In both, the double-displacement and the front-side attack mechanistic alternatives, substrate-assisted catalysis is required to proceed at reliable rates due to the reduced nucleophilic strength of Glu317 in α1,3-GalT, as a result of its interactions with the acceptor substrate in the Michaelis complex (or with water molecules in the binary complex). The foregoing could be a strategy to avoid undesired hydrolysis of the donor substrate and, together with its limited stability, would also explain why it is being so difficult to isolate a glycosyl-enzyme covalent intermediate in α1,3-GalT.

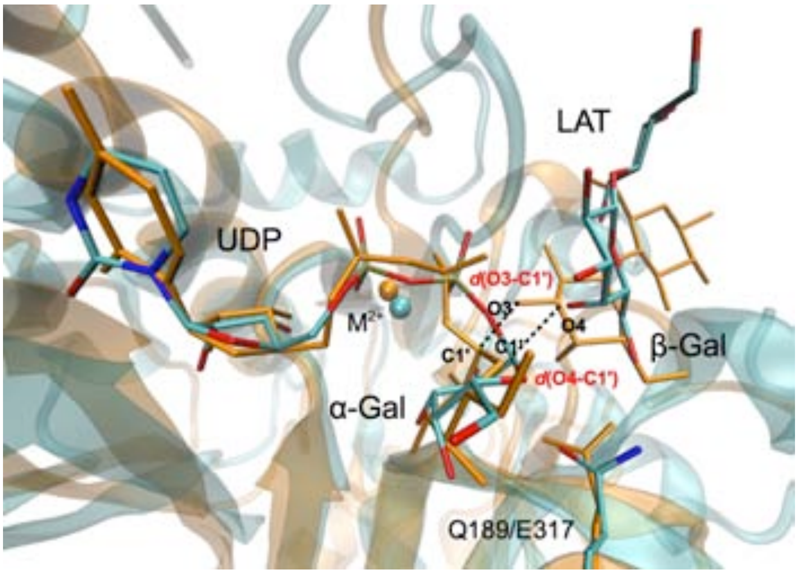

Figure 6.9. Different orientation of the substrates in the active sites of LgtC (cyan; PDB code: 1GA8) and α1,3-GalT (orange; PDB codes: 1O7O (protein, LAT); 2VS5 (UDP-Gal)) (using the crystallographic data).

7

UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase 2 (ppGaNAcT-2)

7.1. INTRODUCTION

The enzymes UDP-Nacetylgalactosamine:polypeptide N-acetylgalactosaminyl-transferases (ppGaNAcTs, EC 2.4.1.41) transfer GalNAc from the sugar donor UDP-GalNAc to serine and threonine residues to form the Tn antigen[267] (GalNAcα1-O-Ser/Thr) and belong to family 27 in the CAZy database.[22] This is a large family, with up to 20 members in humans, and evolutionarily conserved, but the polypeptide substrate preferences of individual isoforms have not been elucidated even though some structural information is available for several isoforms.[268]

The O-linked glycosylation proceeds step-wise[269] and the addition of GalNAc to serine or threonine by ppGaNAcTs represents the first committed step in mucin biosynthesis. Several experiments indicate that there is a hierarchical addition of core GalNAc residues to apomucins, implying that the complete glycosylation of certain substrates is dependent on the coordinated action of multiple ppGaNAcTs.[267]

O-glycans provide unique structural features to mucin glycoproteins and numerous membrane receptors,[270-272] as well as they impart resistance to thermal change and proteolytic attack in a number of diverse proteins.[273,274] Moreover, O-linked carbohydrate side chains function as ligands for receptors (e.g. in host-microbial interactions,[275] lymphocyte and leukocyte homing[276,277]) and as signals for protein sorting.[278-282] Accordingly, a detailed understanding of the catalytic mechanism of ppGaNAcTs would have important practical implications.

The available PDB structures of ppGaNAcTs indicate the presence of two domains connected by a flexible linker that allows significant changes in their relative orientation.[268] The catalytic domain belongs to the GT-A fold, whereas the carbohydrate-binding lectin domain adopts a β-trefoil fold, as classified in the Conserved Domain Database.[283] The enzymatic activity of the catalytic domain depends on the presence of a Mn^{2+} ion coordinated by His and Asp residues, which form the so-called DXH motif of the active site, and by an additional His residue, which are all conserved throughout the ppGalNAcT family.[267]

Regarding the catalytic mechanism, the available crystallographic structures of ppGaNAcTs show that the nearest acidic residues that might function as nucleophiles in a double-displacement mechanism (i.e. Asp224 of the DXH motif binding Mn^{2+} and Glu334) are ~7 Å away from the β-phosphate oxygen.

Consequently, a double-displacement looks unlikely or would require a large conformational change. The latest is not observed on the timescale of the MD simulations performed on the Michaelis complex of human ppGaNAcT-2 by Milac et al.[268] On the other hand, their results are more consistent with a front-side mechanism, since the distance between the glycosidic oxygen and the nucleophilic hydroxyl group is about 3 Å and is maintained nearly constant during the simulation, which would at least structurally be consistent with a nucleophilic role of the acceptor.[268] These results, together with the available X-ray structures and site-directed mutagenesis data, point to a single-displacement mechanism as the most likely.

We present in this chapter a full-enzyme hybrid quantum mechanical/molecular mechanical (QM/MM) study of the catalytic mechanism of human ppGaNAcT-2. Moreover, and as usual, the key factors supporting the catalytic mechanism are analyzed.

This study is result of collaboration with the experimental group leaded by Dr. L. A. Tabak from the National Institute of Dental and Craniofacial Research (NIDCR) (National Institutes of Health; NIH) in Bethesda, MD, USA.

7.2. MODELS AND METHODS

An initial ternary complex modeled by Milac et al.[268] was used as starting point in the reactivity study. They built this ternary complex as follows: coordinates of catalytic domain of ppGaNAcT-2 and acceptor peptide (EA2; sequence PTTDSTTPAPTTK) were taken from PDB Code 2FFU.[284] The donor substrate (UDP-GalNAc) was modeled in the active site using as a template the human ppGaNAcT isoform 10 structure (PDB Code 2D7I[285]), which contains hydrolyzed UDP-GalNAc.

All water atoms in the solvated ternary complex being more than 30 Å away from the anomeric center ($C1'_{GalNAc}$) in this model were deleted. This procedure resulted in a system with ~12630 atoms, including ~2170 TIP3P water molecules (see Figure 7.1A). All residues and water molecules within 15 Å of the anomeric center (~2080 atoms) were included in the active region.

The Mn^{2+} ion present in the original X-ray structure was modeled by the computationally more convenient Mg^{2+}.

Five hydrogen link atoms were added to treat the QM/MM boundary. The charge of the QM region was -1 and included 80 atoms: those from the GalNAc ring, side chain of Thr7 from the acceptor substrate (peptide EA2), Mg^{2+} and its first coordination sphere (phosphate groups from UDP and the side chains of residues Asp224, His226, His359 and one crystallographic water) (see Figure 7.1B).

Starting from the Michaelis complex, a NVT QM(SCC-DFTB)[131]/ MM(CHARMM22) MD was performed using the dynamics module within ChemShell.[219] The SHAKE procedure[199] was applied at every step for the O-H bonds in the water molecules. A 10 ps MD equilibration run was followed by 80 ps of production MD. Two randomly selected snapshot from this simulation were used in QM/MM calculations with QM = BP86/SVP for geometry optimizations.

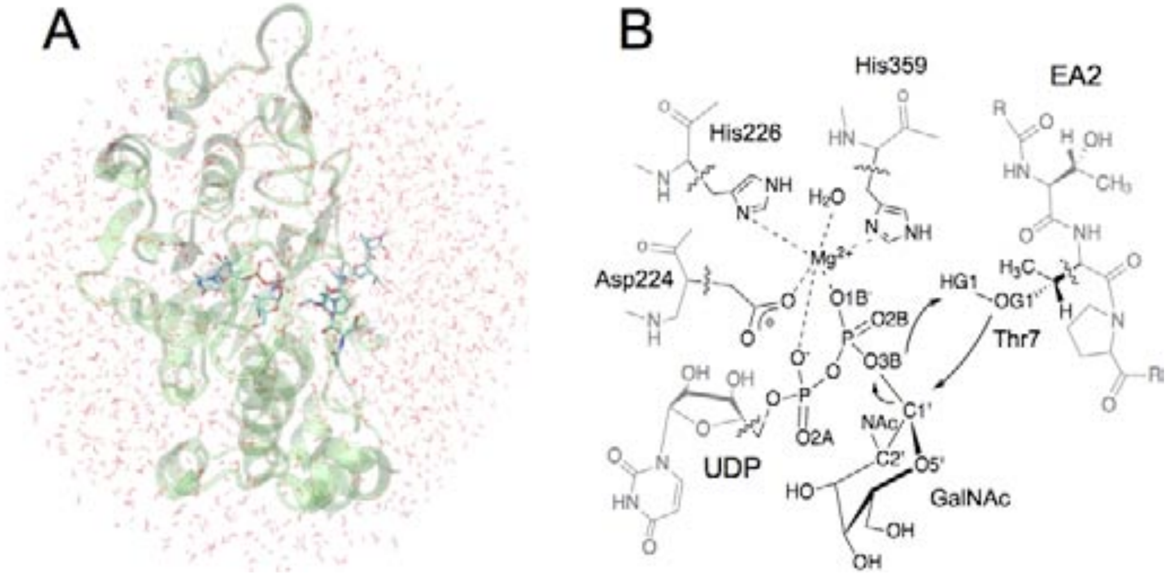

Figure 7.1. Model system used in the QM/MM calculations (**A**) and (**B**) QM/MM partition considered in the present work. In (**B**), QM (MM) atoms are depicted in black (grey). Wavy lines indicate the boundary between the QM and MM regions. The arrows indicate the distances considered in the reaction coordinates and the atoms involved are labeled.

The rest of the QM/MM calculations were performed as exposed in previous chapters.

7.3. RESULTS AND DISCUSSION

7.3.1. Catalytic mechanism. As mentioned in Section 7.1, there is no strong nucleophile in the vicinity of the anomeric center in ppGaNAcT-2 and, therefore, the present reactivity study focuses on analyzing the reliability of a front-side attack mechanism and the factors contributing to it.

Such a mechanism was modeled by using both double (RC = [$d(\mathrm{O3B}_{\mathrm{UDP}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$ - $d(\mathrm{OG1}_{\mathrm{T7}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$]) or triple (RC = [$d(\mathrm{O3B}_{\mathrm{UDP}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$ - $d(\mathrm{OG1}_{\mathrm{T7}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$ - $d(\mathrm{HG1}_{\mathrm{T7}}\text{–}\mathrm{O3B}_{\mathrm{UDP}})$]) reaction coordinates. In both cases, the energy profiles were very similar given that in the reactants the attacking hydroxil group from the acceptor peptide was already hydrogen bonded to $\mathrm{O3B}_{\mathrm{UDP}}$, just like we described before for LgtC. Moreover, very similar results were obtained for the two frames considered and for simplicity most of the results presented in the present chapter will only refer to the first of them.

The potential energy profile is depicted in Figure 7.2, altogether with the evolution of key distances along the front-side attack mechanism.

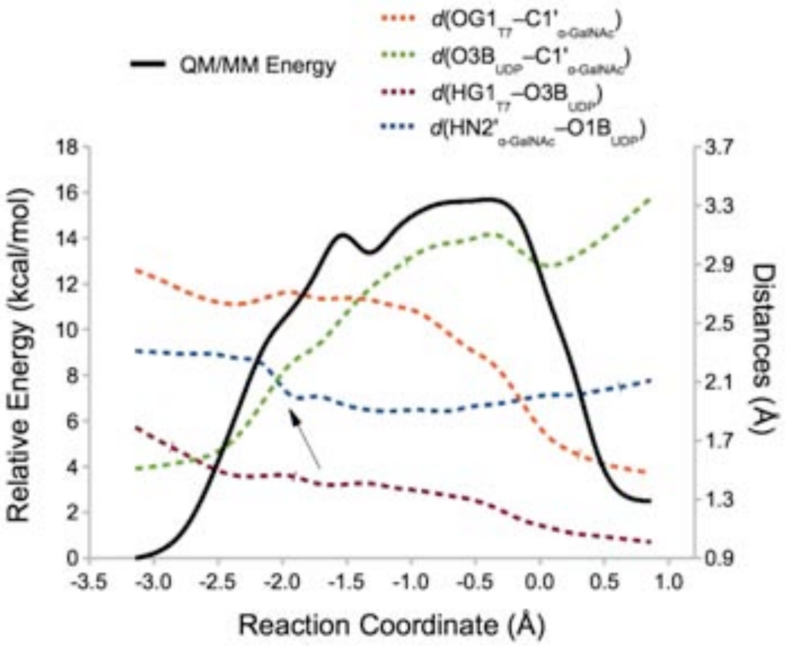

Figure 7.2. QM(BP86/SVP)/MM(CHARMM22) energy profile for the front-side attack mechanism in ppGaNAcT-2. Reaction coordinate (RC) = [*d*(O3B$_{UDP}$–C1'$_{\alpha\text{-GalNAc}}$) - *d*(OG1$_{T7}$–C1'$_{\alpha\text{-GalNAc}}$) - *d*(HG1$_{T7}$–O3B$_{UDP}$)]. The variation of several interatomic distances involved in the reaction is also depicted. The arrow indicates the moment when the NAc group from the α-GalNAc gets properly oriented to favor the O3B$_{UDP}$–C1'$_{\alpha\text{-GalNAc}}$ bond-breaking process.

As can be seen, the surface is quite planar at the region corresponding to the maximum potential energy values. At first sight it would seem to correspond to a S$_{N}$i-like mechanism but no ion-pair intermediate (IP) was characterized so that the S$_{N}$i term would be more appropriate. However, and as stated in case of our previous study of α1,3-GalT, the difference between these two alternatives of front-side attack mechanism are subtle so that the succeeding outcomes would be essentially the same in one case or the other. Moreover, the topology of this surface conditioned that we were also unable to find the corresponding transition state, and we will be considering the TS guess (i.e. ?TS^{i}; structure corresponding to the maximum potential energy value) through out this study.

Notice that the bond-breakage process starts at the very beginning of the reaction, being consistent with the fact that the oxygen from the attacking hydroxyl group was hydrogen bonded with the UDP leaving group in the reactants. Such interaction has been proven to be essential to assist the leaving group departure, and was not present in the Michaelis complex of α1,3-GalT, so that the bond-breakage was not observed at the beginning of the catalytic process (See Figure 5.7).

The energy barriers and reaction energies calculated at different levels of theory are shown in Table 7.1. The potential energy barriers obtained for the S_Ni mechanism are equal to 19.84 kcal/mol and 20.20 kcal/mol (QM = M05-2X/TZVP) for the two frames considered, respectively (See Table A7.1 for values corresponding to the other levels of theory for the second frame).

Table 7.1. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed front-side attack mechanism at different levels of theory. The calculations were carried out on the corresponding QM(BP86/SVP)/MM(CHARMM22) geometries of reactants (R), transition state guess (?TS^i) and products (P).

	BP86		**B3LYP**		**M05-2X**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
?TS^i	15.66	10.74	19.27	13.97	26.34	19.84
P	-2.26	-0.46	1.75	-1.12	3.56	0.27

The values of the energy barriers are in qualitative agreement with the phenomenological free energy barrier of 17.30 kcal/mol derived from the experimental k_{cat} values of 3.7 s^{-1} at 310 K[284] and therefore our results so far suggest that the front-side attack may be feasible.

Key distances corresponding to the stationary points are listed in Table 7.2 (See Appx. Table A7.2 for the other frame) while the structures of the stationary points are depicted in Figure 7.3. Some common trends are again observed when comparing to the results obtained for LgtC and α1,3-GalT; that is, the ?TS^i is highly dissociative (d(O3B$_{UDP}$–C1'$_{\alpha\text{-GalNAc}}$) = 3.10 Å) and the proton transfer process takes place late in the

Table 7.2. Selected QM/MM bond distances *d* (Å), dihedral angle (Degrees) and atomic charges *q* (a.u.) in the optimized reactants (R), transition state guess (?TS^i), and products (P) for the front-side attack mechanism. QM=BP86/SVP and M05-2X/TZVP//BP86/SVP for the geometrical parameters and charges, respectively.

	Reactants	**?TS^i**	**Products**
$d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	1.51	3.10	3.30
$d(\mathrm{OG1_{T7}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	2.86	2.20	1.49
$d(\mathrm{HG1_{T7}}\text{–}\mathrm{OG1_{T7}})$	0.99	1.15	1.47
$d(\mathrm{HG1_{T7}}\text{–}\mathrm{O3B_{UDP}})$	1.79	1.24	1.02
$d(\mathrm{O_{A307}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	4.09	3.12	3.25
$d(\mathrm{C1'_{\alpha\text{-}GalNAc}}\text{–}\mathrm{O5'_{\alpha\text{-}GalNAc}})$	1.38	1.29	1.39
$d(\mathrm{HN2'_{\alpha\text{-}GalNAc}}\text{–}\mathrm{O1B_{UDP}})$	2.31	1.91	2.09
$(\mathrm{H2'}\text{–}\mathrm{C2'}\text{–}\mathrm{N2'}\text{–}\mathrm{HN2'})_{\alpha\text{-}\mathrm{GalNAc}}$	164.5	168.0	158.8
$q(\mathrm{C1'_{\alpha\text{-}GalNAc}})$	0.39	0.57	0.37
$q(\mathrm{O3B_{UDP}})$	-0.91	-1.16	-1.07
$q(\mathrm{O5'_{\alpha\text{-}GalNAc}})$	-0.51	-0.44	-0.56
$q(\mathrm{O1B_{UDP}})$	-1.18	-1.23	-1.20
$q(\mathrm{HN2'_{\alpha\text{-}GalNAc}})$	0.47	0.49	0.46

reaction. The closest residue on the β-face of the donor sugar substrate is Ala307, which carbonyl group is 4.09 Å from the anomeric center at the reactants and therefore the evolution of such a distance was also followed.

Since we were unable to characterize a TS at the DFT level, we decided to consider the SCC-DFTB level of theory. The corresponding PES is shown in Figure 7.4 and, as can be seen, a much more pronounced maximum is obtained. A TS was easily identified with an imaginary frequency consistent with the reaction under study.

The potential energy values and distances corresponding to the stationary points at the QM(SCC-DFTB)/MM(CHARMM22) level of calculation are giving in the Appendix section (Tables A7.3 and A7.4, respectively). Once again, SCC-DFTB tends to considerably overestimate the potential energy values (energy barrier of 38.95 kcal/mol), although single point calculations at the reference method (QM = M05-2X/TZVP) yielded an energy barrier of 22.72 kcal/mol. In the TS^i $d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}}}$

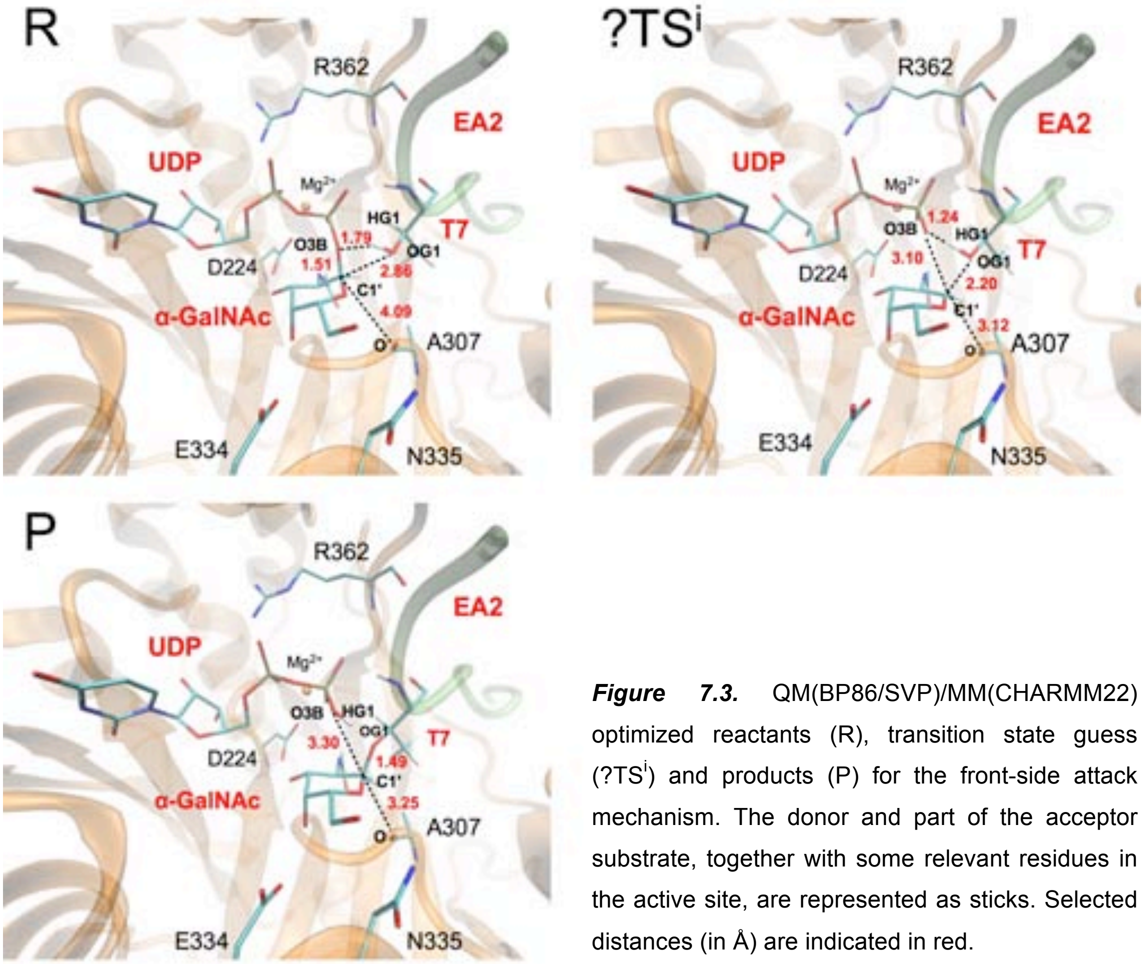

Figure 7.3. QM(BP86/SVP)/MM(CHARMM22) optimized reactants (R), transition state guess (?TS^i) and products (P) for the front-side attack mechanism. The donor and part of the acceptor substrate, together with some relevant residues in the active site, are represented as sticks. Selected distances (in Å) are indicated in red.

$_{\text{GalNAc}}$) = 2.59 Å so that it corresponds to a less dissociative reaction intermediate. The reaction energy was 13.22 (13.32) kcal/mol at QM = SCC-DFTB (M05-2X/TZVP) levels of theory. The foregoing is in contrast with the results obtained when performing geometry optimizations at the QM(BP86/SVP)/MM(CHARMM22) level, where the reaction turned to be almost isoergic. This is consistent with the higher barrier obtained at the SCC-DFTB level. Of course we are more confident on the results obtained at the DFT level, which presumably provides a better description of the catalytic mechanism. On the other hand, the use of the SCC-DFTB method enabled us to perform an umbrella sampling calculation, which would be much more expensive at higher levels of theory. Both, the potential energy and free energy profiles are depicted in Figure 7.4. Qualitatively speaking, the entropic effects might be relatively small when comparing the energy barriers, just like we mentioned in the case of LgtC. Regarding the reaction energy, a bigger difference is obtained between the potential and free energy profiles. In the last case the reaction is less endoergic, and in better agreement with the results obtained at the DFT level. Presumably, performing structural sampling allowed an improved description of the end of the catalytic mechanism when considering the SCC-DFTB level of calculation.

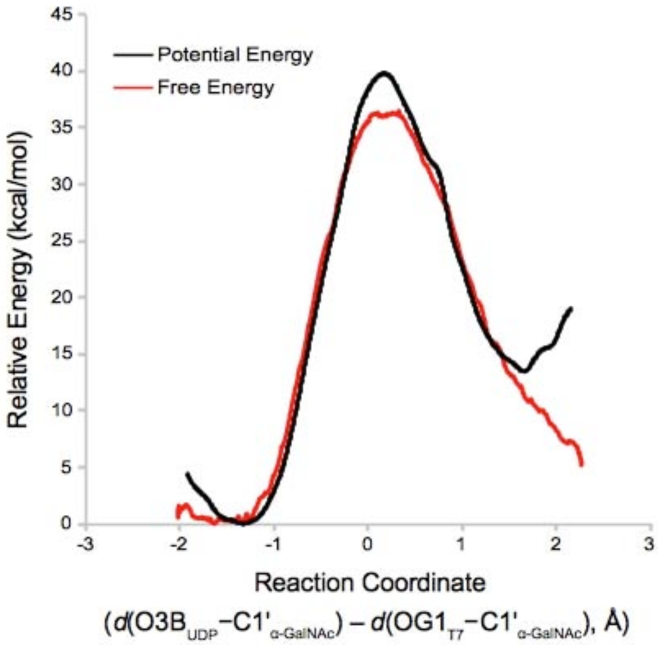

Figure 7.4. SCC-DFB/CHARMM22 potential energy profile and potential of mean force (PMF) for the front-side attack mechanism.

7.3.2. Analysis of factors contributing to catalysis. Once the front-side attack mechanism was modeled, a detailed analysis of the interactions supporting the catalytic process was performed.

7.3.2.1 Enzyme-substrates interactions; key enzyme residues. Since the intermediates of the reaction catalyzed by ret-GTs are quite charged we focused once again on the electrostatic stabilization of the QM region provided by all the residues in the active space. As explained in the Methods section, this is performed by switching off the charge of the residue and recalculating the QM/MM interaction energy. There are basically 4 residues displaying such effect; Arg362, Asp334, Ala307 and Trp331, for which stabilizations energies (QM = M05-2X/TZVP) of 18.62, 11.53, 2.83 and 2.30 kcal/mol were estimated for the ?TS^{i} as compared to the reactants. The most stabilizing residue (i.e. Arg362) is interacting with UDP (See Figure 7.3), just like we described in our previous studies with LgtC and α1,3-GalT, supporting the idea that the stabilization of the negative charge on the leaving group is a key feature that strongly contributes to the catalytic efficiency of ret-GTs. Tyr331 is also interacting with the leaving group through a hydrogen bond with O3B$_{\mathrm{UDP}}$. Moreover, Glu334 is hydrogen-bonded to the donor substrate and therefore it is also a key residue for the binding of this substrate (See Figure 7.3) Moreover, it is located on the β-face of the sugar ring and it is a negatively charged residue so that it is expected to be important in catalysis.

Also notice that Asn335, a residue lying on the β-face of the donor sugar substrate (See Figure 7.3) and that has been suggested to be the putative nucleophile in a double-displacement mechanism in ppGaNAcT-2, is not expected to have a very significant effect on the stabilization of the oxocarbenium species according to this analysis. This is not surprising as it located at $d(\mathrm{OD1}_{\mathrm{N335}}–\mathrm{C1'}_{\alpha\text{-GalNAc}})$ = 7.05 Å in the Michaelis complex and at a distance of 6.96 ± 0.43 Å along the simulations performed by Dr. Milac. Moreover, the carbonyl side chain of Asn335 is pointing away from the anomeric carbon, whereas the amide nitrogen is hydrogen bonded to the Ala307 backbone carbonyl. This carbonyl group of Ala307 is better positioned to stabilize the positive charge development in C1'$_{\alpha\text{-GalNAc}}$, and this is reflected in the 2.83 kcal/mol of stabilization provided by ths residue.

7.3.2.1.1. E334Q, R362K, N335A and N335D mutants. Once the most important residues of the enzyme (from the catalytic point of view) were identified, a list of mutants was made to be test experimentally. Conservative mutations, which are presumed to preserve the structural role of the residue while targeting the chemistry in question were applied in most cases. In particular, mutants E334Q, N335A, N335D, N335H, N335S and R362K have been produced and are being tested in the laboratory of Dr. Tabak (Divya, Rojas and Tabak) following a previously defined protocol.[286,287] The preliminary experimental results obtained so far indicate that ppGaNAcT-2 is highly sensitive to mutations in such positions, although in most cases the mutant still shows some activity. While this chapter is being written, the conditions to be used in the kinetics assays are being set up so that the direct effect on catalysis can be measured.

Some of these mutants have also been modeled *in silico* (i.e. R362K, E334Q, N335A, N335D) and the energy profile for the front-side attack has been calculated to qualitatively assess the effect that such mutations would produce in catalysis. An important distinction should be made, though; we built the models of the mutants by just replacing the side chain of the original residue, meaning that we intended to evaluate the effect of the mutations in the catalytic mechanism itself if there were not significant structural perturbations of the enzyme and without considering the effect on the binding of the substrates. The latest may not be a very good assumption for some of the mutants (even if conservative mutations have been done), since we are considering residues that are directly implicated in the binding and/or could also have a structural role in the active site. However, this is probably the most reliable way to estimate if the catalytic performance of the mutants can be explained by only considering the role of the specific residues in the reaction, while important inconsistencies with the experimental kinetic results might indicate that the overall structure of the active site and/or the binding of the substrates is also affected.

The potential energy barrier and reaction energies associated to these mutants are shown in Figure 7.5. The corresponding potential energy profiles were equally planar at their maximum (See Figure 7.7) and, therefore, we did not perform a proper TS search but used the maximum of the potential energy profile as a TS guess.

As can be seen in Figure 7.5, the energy barrier would increase in ~7 kcal/mol in the case of mutant E334Q. The carboxylic group of Glu334 is initially at 6 Å from the anomeric center but this negatively charged residue on the β-face of α-GalNAc would still contribute to the stabilization of the developing positive charge on the sugar ring. Suppressing this negative charge by mutating Glu by Asn significantly reduces the electrostatic stabilization role of position 334 in catalysis (See Figure 7.6A). Moreover, this is a key residue in the binding of the donor substrate, which is hydrogen-bonded to OH4 of UDP-GalNAc in the Michaelis complex. Therefore it is expected to affect both the K_M for the donor substrate and the k_{cat} values if this residue is mutated. Since ret-GTs bind their substrates in an ordered and interdependent way, which have also been

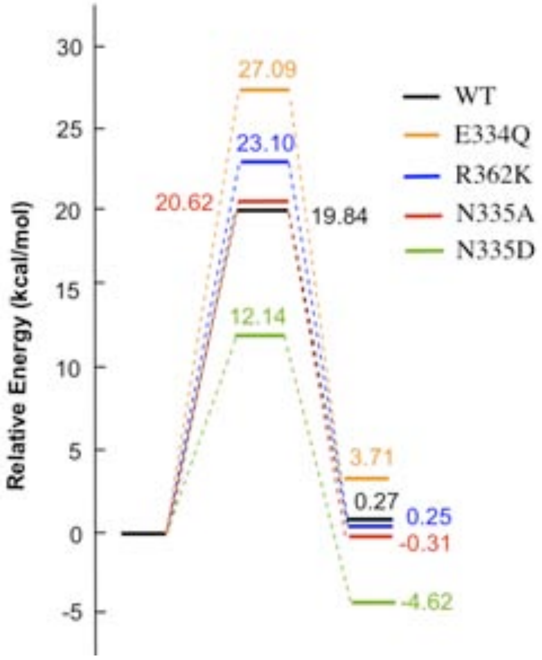

Figure 7.5. QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM22) potential energy barriers and reaction energies for wild-type (WT) ppGaNAcT-2 and considered mutants. For the energy barriers a transition state guess (?TS^{i}) was considered.

certified in the case of ppGaNAcTs,[288] it could also be expected an increase in the K_M value for the acceptor substrate. Stated differently, Glu334 is an essential residue and to mutate such position is expected to affect the binding of the ligands and the catalytic process itself, given the importance of the negative charge of the aspartate carboxyl group for the catalytic activity.

All the latest would be consistent with the available experimental mutagenesis data for murine ppGaNAcT-1, where mutant E319Q (being Glu319 the equivalent of Glu334 in human ppGaNAcT-2) exhibits a residual activity of 0.04 %.[289]

In the case of mutant R362K the energy barrier is increased by ~3 kcal/mol, even if a lysine in such position would still stabilize the developing negative charge on the leaving group. Despite this, notice from Figure 7.6C that Arg362 shows a higher stabilizing effect than lysine. The side chain of Arg362 is closer to the UDP leaving group thus explaining the differences found (i.e.; distances $d(\text{NH2}_{\text{R362}}–\text{O2A}_{\text{UDP}})/(d(\text{NZ}_{\text{K362}}–\text{O2A}_{\text{UDP}})$ = 3.23/4.97 Å and 2.72/4.62 Å in R and ?TS^{i}, respectively).

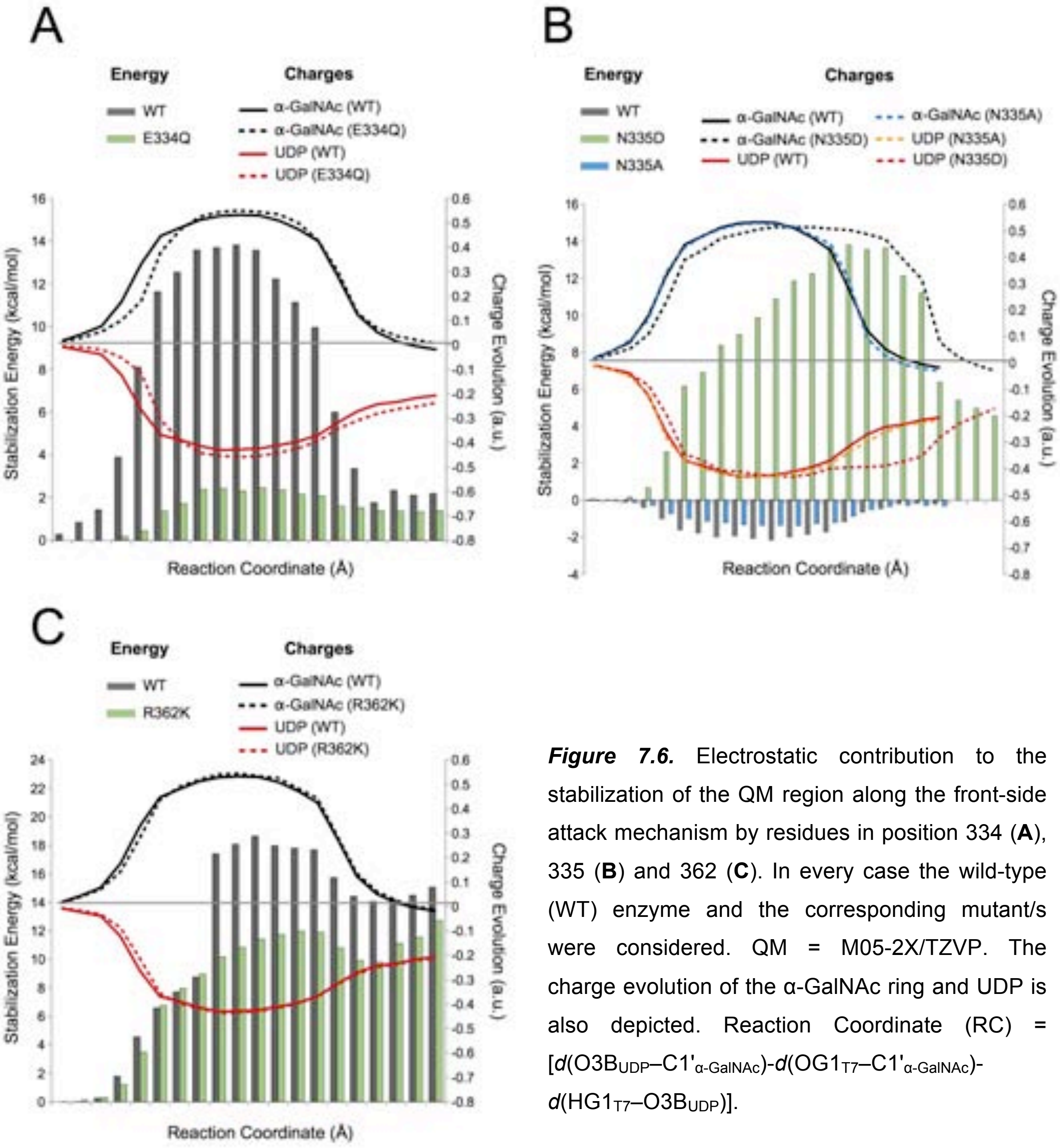

Figure 7.6. Electrostatic contribution to the stabilization of the QM region along the front-side attack mechanism by residues in position 334 (**A**), 335 (**B**) and 362 (**C**). In every case the wild-type (WT) enzyme and the corresponding mutant/s were considered. QM = M05-2X/TZVP. The charge evolution of the α-GalNAc ring and UDP is also depicted. Reaction Coordinate (RC) = [d(O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-GalNAc}}$)-d(OG1$_{\mathrm{T7}}$–C1'$_{\alpha\text{-GalNAc}}$)-d(HG1$_{\mathrm{T7}}$–O3B$_{\mathrm{UDP}}$)].

Increments of the energy barriers of ~3 or ~7 kcal/mol, as observed in mutants R362K or E334Q, would imply less than 0.04 % of residual activity. It should be remembered that we built the *in silico* mutant by just changing the side chain of these residues and therefore our results indicate that ppGaNAcT-2 is highly sensitive to any change in these key positions.

For the mutant N335A the energy barrier remains practically unaffected, which is consistent with our expectations since Asn335 was not found to be an important residue in terms of electrostatic stabilization (See Figure 7.6B). However, an important change is observed in the case of mutant N335D, where a drop in the energy barrier of ~7 kcal/mol is observed. This result predicts that to place a negatively charged residue on the β-face of the donor sugar substrate would turn 335 into a key position (even if it is too far away of the anomeric center to participate in a double-displacement mechanism ($d(\mathrm{OD1}_{\mathrm{N335}}–\mathrm{C1'}_{\alpha\text{-GalNAc}})$ = 7.05 Å in reactants)); and would definitely alter the charge evolution along the reaction. Interestingly, the presence of a strong nucleophile in this position facilitates the leaving group departure and delays the nucleophilic attack of the incoming hydroxyl group, basically because the positive charge on the α-GalNAc ring is better stabilized (See Figure 7.6B). In that sense, notice from Figure 7.7 that the evolution of the leaving group departure and nucleophilic attack is actually different when comparing to the wild-type enzyme or the rest of the mutants. Stated differently, the acceptor substrate is less *"attracted"* by the α-GalNAc ring resulting in a very flat potential energy surface.

Mutations to alanine in the equivalent asparagine residue in the murine isoform (i.e. ppGaNAcT-1) just had a little effect on catalysis[289] and our calculations suggest that the role of Asn335 in human ppGaNAcT-2 might be even slightly destabilizing (See Figure 7.6B). The latest seems logical since the NH2 of the carboxamide group is the closest part of this residue to the anomeric center (See Figure 7.3), where a positive charge is developed along the reaction.

As it was done for the other two enzymes studied in this thesis, the leaving group departure (or UDP-GalNAc bond breakage) was also modeled by using $d(\mathrm{O3B}_{\mathrm{UDP}}–\mathrm{C1'}_{\alpha\text{-GalNAc}})$ as reaction coordinate. The results are shown in Figure 7.8, where can be seen that the bond-breaking process looks harder in case of mutant E334Q and much easier for N335D, while in case of mutants R362K and N335A the differences are minor compared to the wild-type enzyme. The latest is supporting an idea that we have been emphasizing throughout this thesis; that is, the leaving group departure process is the driving force of the sugar transfer by ret-GTs and in consequence, any residue *"pushing"* away the UDP and/or providing electrostatic

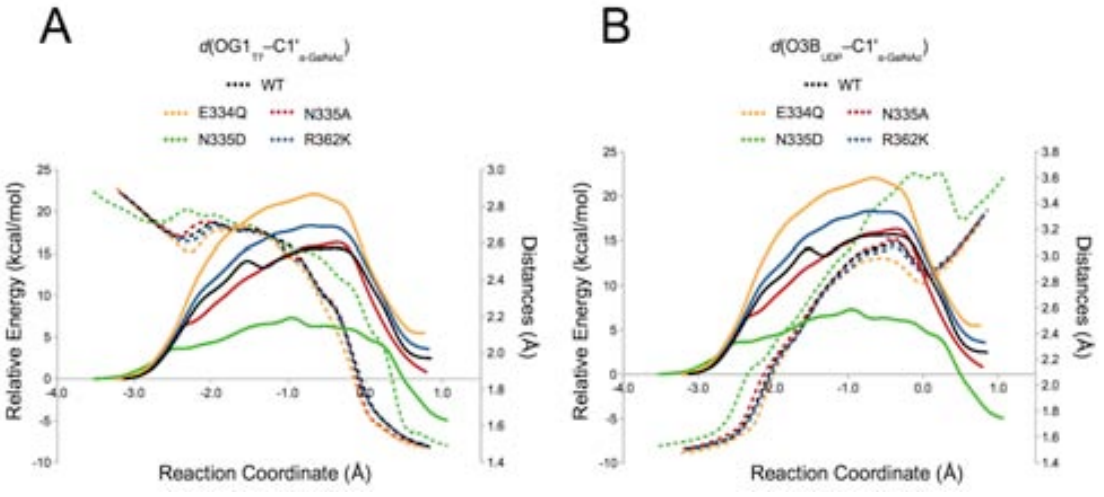

Figure 7.7. QM(BP86/SVP)/MM(CHARMM22) energy profile for the front-side attack mechanism in wild-type (WT) ppGaNAcT-2 and mutants E334Q, N335A, N335D and R362K. Reaction coordinate (RC) = [d(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-GalNAc}}$) - d(OG1$_{\text{T7}}$–C1'$_{\alpha\text{-GalNAc}}$) - d(HG1$_{\text{T7}}$–O3B$_{\text{UDP}}$)]. The variation of d(OG1$_{\text{T7}}$–C1'$_{\alpha\text{-GalNAc}}$) (**A**) and d(O3B$_{\text{UDP}}$–C1'$_{\alpha\text{-GalNAc}}$) (**B**) is also shown. Solid lines correspond to the QM/MM potential energy profiles and share the same color code used for the distances.

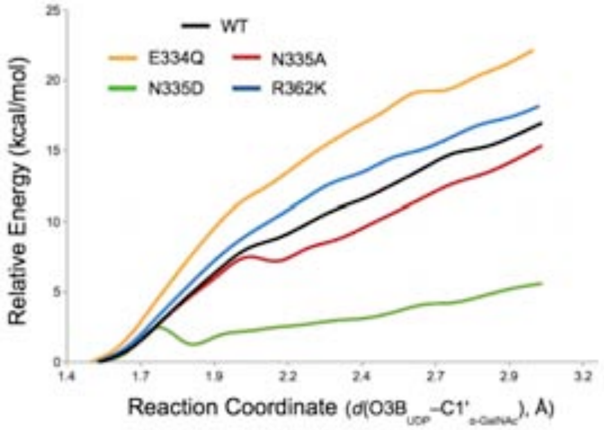

Figure 7.8. QM(BP86/SVP)/MM(CHARMM22) energy profiles for the bond-breaking process in wild-type (WT) ppGaNAcT-2 and mutants E334Q, N335A, N335D and R362K.

stabilization to the resulting oxocarbenium-like species will become essential in catalysis.

7.3.2.2. Inter- and intra- substrate interactions. We have extensively shown how important the inter- and intra-substrate interactions are to explain the catalytic efficientcy of LgtC and α1,3-GalT. In order to identify equivalent interactions in the case of human ppGaNAcT-2, a NBO analysis was performed by considering the reactants and the transition state guess for the wild-type enzyme. The results are summarized in Table 7.3.

Table 7.3. Donor-Acceptor Natural Bond Orbitals (NBO) analysis for the front-side attack mechanism in wild-type ppGaNAcT-2. QM=(M05-2X/TZVP//BP86/SVP). LP: lone pair; BD: bonding molecular orbital; BD*: antibonding molecular orbital; CR: core pair. Only the main interacting pairs are given. ΔE corresponds to the energy difference between the interacting molecular orbitals in the transition state guess (?TS^i) and the reactants (R). [&, §] Interactions only present in R and ?TS^i, respectively. Reference to the figure where the interaction is depicted is within parenthesis.

Donor NBO	Acceptor NBO	ΔE (kcal/mol)
UDP-GalNAc – Thr 7 (EA2)		
BD$_2$ (PB-O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	124.98[§] (7.9B)
LP$_1$ (O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	12.20
BD*$_1$ (PB-O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	14.06[§]
BD$_1$ (PB-O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	5.03[§]
CR$_1$ (O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	3.75[§]
BD*$_1$ (OG1-HG1)$_{T7}$	BD*$_2$ (PB-O3B)$_{UDP}$	3.55[§]
LP$_2$ (O3B)$_{UDP}$	BD*$_1$ (OG1-HG1)$_{T7}$	-8.70 (7.9A)
LP$_3$ (O3B)$_{UDP}$	BD*$_1$ (N-HN)$_{T7}$	2.13 (7.9D)
LP$_2$ (O3B)$_{UDP}$	BD*$_1$ (N-HN)$_{T7}$	1.37
LP$_1$ (O3B)$_{UDP}$	BD*$_1$ (N-HN)$_{T7}$	-2.05 (7.9C)
NAc – Asp224/UDP		
LP$_1$ (OD2)$_{D224}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	-7.34[&] (7.9E)
LP$_3$ (OD2)$_{D224}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	-2.11[&]
LP$_1$ (O1B)$_{UDP}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	6.86[§] (7.9F)
LP$_2$ (O1B)$_{UDP}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	2.06[§]
BD*$_1$ (PB-O1B)$_{UDP}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	2.45[§]
LP$_3$ (O1B)$_{UDP}$	BD*$_1$ (N2'-HN2')$_{\alpha\text{-GalNAc}}$	1.18[§]

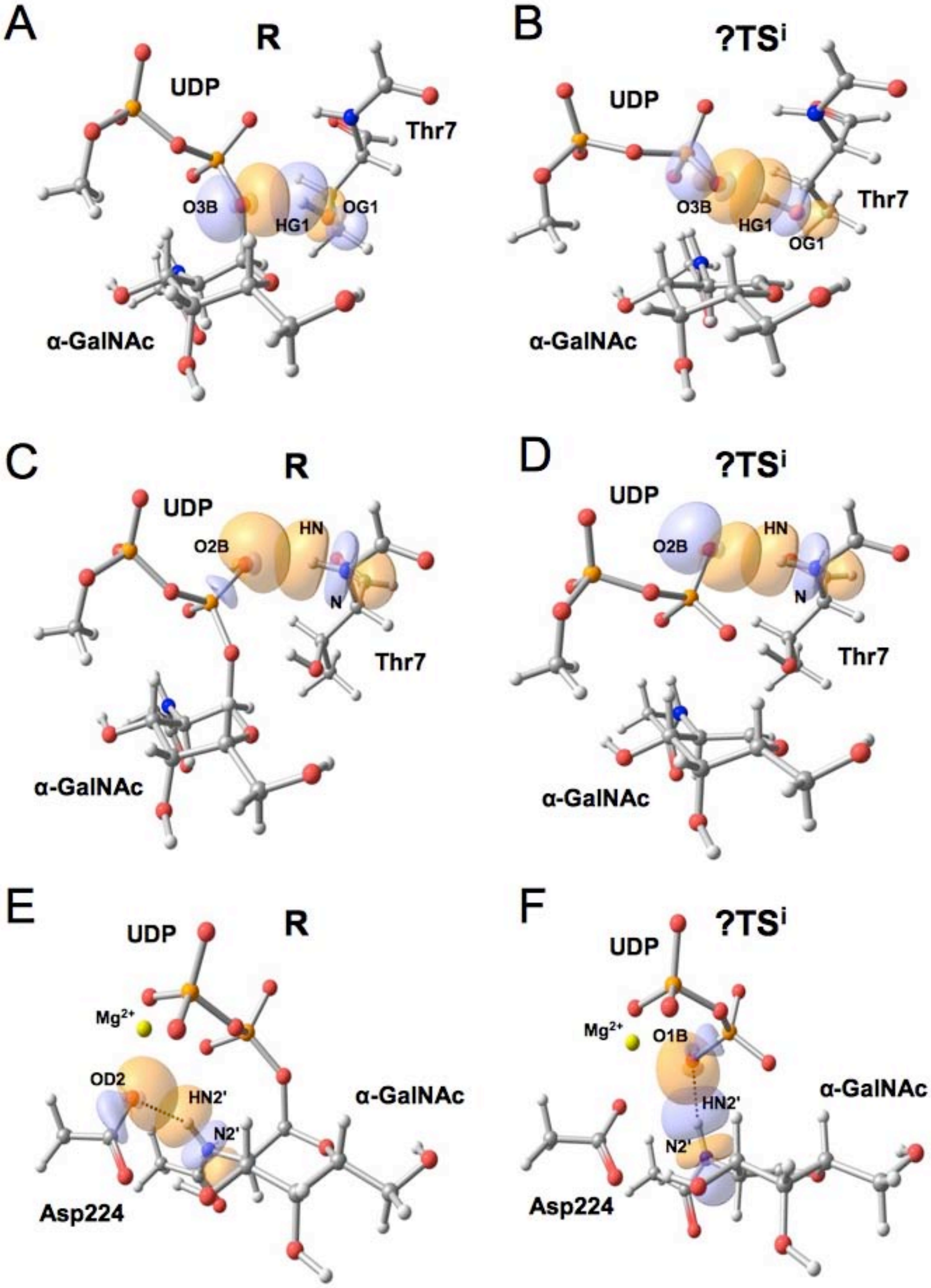

Figure 7.9. Interactions between molecular orbitals of the substrates relevant in N-acetylGal transfer by ppGaNAcT-2 according to a NBO analysis. These interactions involve: $O3B_{UDP}$ and the incoming $OG1_{T7}$ in the reactants (R) (**A**) and in the front-side attack transition state guess (?TS^i) (**B**); the backbone amide group of Thr7 in R (**C**) and ?TS^i (**D**); NAc group of the donor substrate with Asp224 in R (**E**) and UDP in ?TS^i (**F**). For clarity, just a fraction of the QM atoms is shown.

Just as described for LgtC and α1,3-GalT, the interactions between molecular orbitals involving the leaving group oxygen bonded to the monosaccharide to be transferred ($O3B_{UDP}$) and the attacking nucleophilic group from the acceptor substrate are the most significant ones. Also notice from Table 7.3 and Figures 7.9 A-B that the attacking hydroxyl group from Thr7 was already facing the UDP leaving group in the reactants, in a similar way as described before for LgtC.

There is a second inter-substrate interaction involving the leaving group and the backbone amide from the same residue to which the α-GalNAc is getting transferred (i.e. Thr7 from peptide EA2) (See Table 7.3 and Figures 7.9 C-D). Both inter-substrates interactions were already present in the Michaelis complex so that they are primarily determined by the orientation between the substrates in the active site and the specificity of the enzyme.

On the other hand, there is a distinctive trait in the case of the intra-substrate interactions in ppGaNAcT-2 when comparing to the other ret-GTs that we have studied before. We refer to the interaction between the 2'-N-acetyl (NAc) group from the donor substrate and the UDP leaving group, which is only present in the TS guess. In the reactants, NAc is interacting with Asp224, one of the residues coordinating the metallic cofactor (i.e. Mg^{2+}) (See Figure 7.9E). The latest was also observed throughout the 40 ns of MD simulation that Milac et al. performed on the Michalis complex.[268] However, and due to the change in the sugar ring puckering along the reaction (i.e.; from a distorted $^{4}C_{1}$ (puckering parameters $\phi = 137.3°, \theta = 15.0°$) to a ^{4}E-like ring conformation ($\phi = 242.9°, \theta = 31.3°$) in R and $?TS^{i}$, respectively), NAc gets reoriented and establishes a stabilizing interaction with the UDP leaving group (See Table 7.2 and Figures 7.2 and 7.9F). The distance $d(HN2'_{\alpha\text{-GalNAc}}–O1B_{UDP})$ gets shorter, from 2.31 Å in the Michaelis complex to 1.91 Å in the $?TS^{i}$, and as result $N2'_{\alpha\text{-GalNAc}}$ gets hydrogen-bonded with $O1B_{UDP}$.

It is known that some hydrolases like OGA (a GH involved in O-GlcNAcylation cycling) employ the NAc group from GlcNAc itself as a nucleophile to cleave the monosaccharide from serine/threonine.[290] However, for a GT a catalytic role by a NAc group from the donor substrate has just recently being described. In that work, Tvaroška et al.[291] found a substrate-assisted mechanism by this functional group while performing a QM/MM study on one inv-GT O-GlcNAc transferase (i.e.; uridine diphospho-N-acetylglucosamine:poly-peptide β-N-acetylaminyltransferase, OGT) that belongs to the GT-B superfamily. More specifically, they described a rotation of the C2'-N2' bond that approaches the HN2' proton to the oxygen of the breaking glycosidic linkage, thus stabilizing the leaving group negative charge and assisting its departure. Finally, the authors hypothesize that OGT would require a mechanism like this to deal with the fact that there are no extra inter-substrates interactions and this is a metal-ion-independent enzyme that otherwise would not efficiently catalyze the reaction. The latest is not the case for ppGaNAcT-2, but still the NAc group resulted important in catalysis.

On the other hand, and according to our results, in ppGaNAcT-2 the NAc group has a stabilizing effect through interactions with $O1B_{UDP}$ and not with the oxygen atom of UDP linked to the monosaccharide ($O3B_{UDP)}$ (See Figure 7.9F). There is another important difference when comparing our results with that obtained by Tvaroška et al.;[291] that is, in the case of OGA there is a clear rotation of the NAc group along the reaction pathway (dihedral angle $\chi_{(H2'-C2'-N2'-HN2')}$ = 145° and 175° in the Michaelis complex and the TS, respectively), while in ppGaNAcT-2 is the change in the sugar ring puckering which promotes the referred NAc-UDP interaction, whereas the variation of $\chi_{(H2'-C2'-N2'-HN2')}$ is just of 3.5° between the reactants and the transition state guess.

To shed more light on the relevance of the NAc group in catalysis, alternative donor substrates were considered and tested. More specifically, we substituted this NAc group by OH, H or OCH_3 in the original Michaelis complex, which corresponds to consider UDP-Gal, 2'-deoxy-Gal and 2'-oxymethyl-Gal as donor substrates. Once again, and as was mentioned for the study of the mutants, we did not intended to reproduce any effect that this change could have in the binding of the donor substrate, which would require performing long MD simulations. The point here was to evaluate the direct effect that these substitutions would have in the catalytic process itself. The effects of this functional group substitution on the energy and reaction barriers are summarized in Figure 7.10.

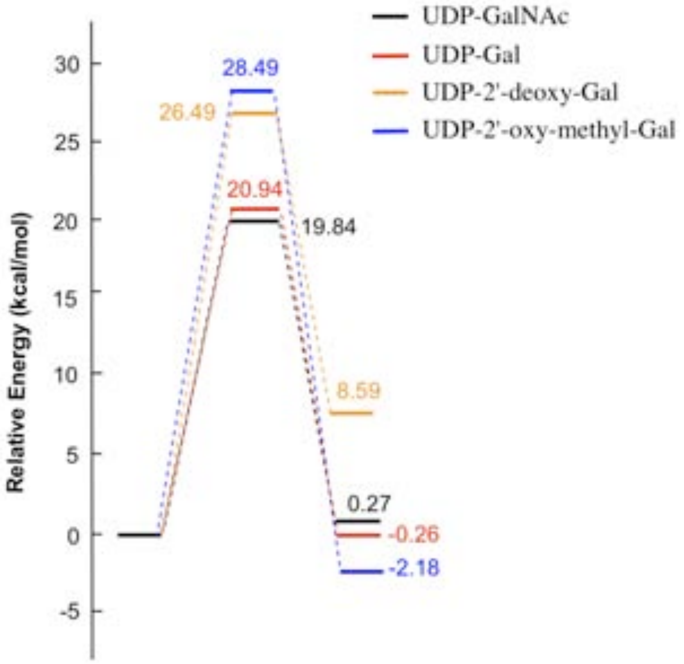

Figure 7.10. QM(M05-2X/TZVP//BP86/SVP)/MM(CHARMM22) potential energy barriers and reaction energies for wild-type ppGaNAcT-2 and different donor substrates. For the energy barriers a transition state guess was considered (?TS^{i}).

When UDP-Gal was considered, the 2’-hydroxyl group is predominantly interacting with Asp224 in the Michaelis complex, and gets reoriented along the reaction, thus behaving similarly to the NAc group in UDP-GalNAc (See Figures 7.11 A-B). The latest would explain why the energy barrier is so similar when considering UDP-Gal instead of UDP-GalNAc. Interestingly, in the case of LgtC and α1,3-GalT, which use UDP-Gal as donor substrate, the interaction between the 2’-hydroxyl group and UDP was already present in the reactants (See Appx., Figure A7.2). We performed a QM(SCC-DFTB)/ MM(CHARMM22) MD simulation of 100 ps for ppGaNAcT-2 in order to estimate if this hydroxyl group could establish an initial interaction with UDP just like it is observed in LgtC and α1,3-GalT; but still the prevailing interaction was with Asp224. Altogether, our results suggest that human ppGaNAcT-2 would be able to transfer Gal to the peptide EA2, being consistent with the results obtained for another acceptor peptide (i.e. Muc2; sequence PTTTPISTTTMVTPTPTPTC).[292] In this work, the V_{max} values corresponding to the transfer of Gal-NAc and Gal were estimated in 46.1 and 79.9 pmol/min, respectively, which implies a decrease of less than 1 kcal/mol

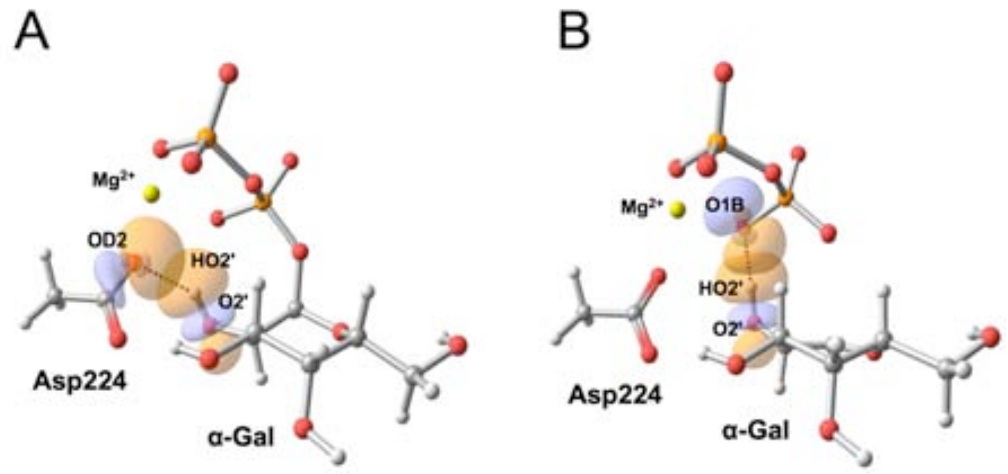

Figure 7.11. Interactions between molecular orbitals of the 2' hydroxyl group of Gal with Asp224 (**A**) and UDP (**B**) in R and ?TS^{i}, respectively, and according to a NBO analysis in ppGaNAcT-2. For clarity, just a fraction of the QM atoms is shown.

in the energy barrier when using UDP-Gal as a donor substrate. Moreover, the authors concluded that giving the relatively small difference between the K_M values for UDP-GalNAc and UDP-Gal (10 and 27 μM, respectively), UDP-Gal might actually be a naturally relevant substrate of ppGaNAcT-2.

Even if our results predict a slightly higher difference (energy barrier is ~1 kcal/mol higher for UDP-Gal), it falls within the order of error that could be expected for the methods. Moreover, it should be pointed out that we did not look for the corresponding TSs. In any case, we propose that the energy barrier corresponding to UDP-Gal would be quite similar to the observed for the GalNac transfer reaction.

A most obvious effect is observed when the other two alternative donor substrates are considered. As can be seen in Figure 7.10, the energy barrier increases by ~7 and ~9 kcal/mol for 2'-deoxy-Gal and 2'-oxymethyl-Gal, respectively. Also, the reaction energies are more affected. These results would support the idea that the interaction of 2'-NAc (OH) in UDP-GalNAc (UDP-Gal) is a key factor, and to suppress it would significantly disrupt catalysis in ppGaNAcT-2. No surprisingly, and as observed in Figure 7.12, the identity of the functional group in position 2' also conditions the facility of the leaving group departure, which would be easier in the case of the donor substrates UDP-GalNAc and UDP-Gal. Finally, and according to our

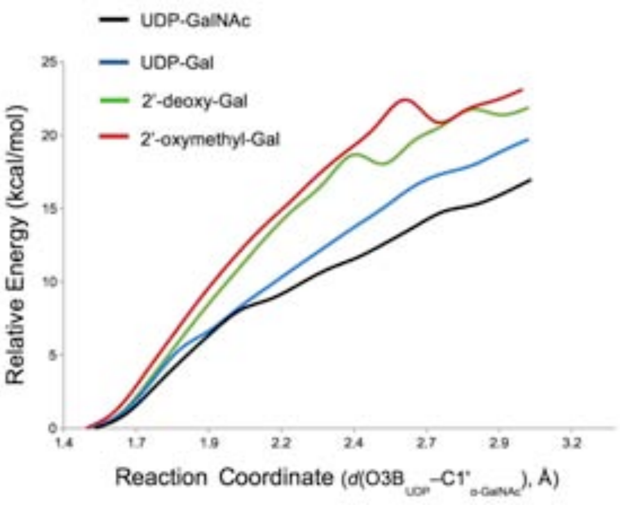

Figure 7.12. QM(BP86/SVP)/MM(CHARMM22) energy profiles for the bond-breaking process in wild-type ppGaNAcT-2 and different donor substrates.

theoretical results, negligible or no detectable residual activity would be observed when 2'-deoxy-Gal or 2'-oxymethyl-Gal are used as donor substrates.

7.4. CONCLUSIONS

We performed full QM(DFT//SCC-DFTB)/MM calculations to study the catalytic mechanism of human ppGaNAcT-2. A front-side attack mechanism was described (probably S_Ni) but the flatness of the potential energy surfaces impeded the exact localization of the corresponding TS. We estimated the energy barrier in ~20 kcal/mol (QM = M05-2X/TZVP), which would be in reasonable agreement with the experimental kinetic data.

A different scenario is obtained when considering the SCC-DFTB level of theory for the QM atoms and the TS is then easily found, with an energy barrier of ~23 kcal/mol (QM = M05-2X/TZVP). In this case, the reactions results much more endoergic (reaction energy of ~13 kcal/mol) although such tendency seems to be corrected by performing a free energy calculation. For the energy barriers, the entropic effects appear to be relatively small when comparing potential and free energy profiles.

The analysis of factors contributing to catalysis revealed that there are two key amino acids in the active site of the enzyme: Arg362 and Glu334. Tyr331 and the backbone of Ala307 were also found to stabilize the transition state but to a less extend. Accordingly, and as part of a collaboration with the experimental group leaded by Dr. L. A. Tabak, we proposed to mutate Arg362 and Glu334. Additionally, position 335 was also considered, since the original Asn is situated on the β-face of the GalNAc ring and the identity of the residue may have an important impact in catalysis. At the time of finishing the writing of this thesis, the experimental kinetics of those mutants are being studied.

On the other hand, we have studied *in silico* the reactivity of some of these mutants. In case of E334Q and R362K an energy barrier increase of ~3 and ~7 kcal/mol is estimated, thus supporting the idea of their relevance in catalysis. In case of mutants N335A there is almost no effect, but for N335D a decrease of ~7 kcal/mol in the energy barrier is obtained. The latest is consistent with the fact that any negatively charged residue on the β-face of the sugar ring would stabilize the developing positive charge on the anomeric carbon and would assist catalysis. It should be pointed out that we intended to estimate the effect of the mutations on the catalytic mechanism itself and not on the structure of the enzyme or the binding of the substrates. The latest could also be affected by the mutation thus contributing to adecay in the catalytic activity, but the consideration of these aspects was beyond the scope of the present work. We expect to provide a much complete discussion of these results once the experimental kinetic parameters are available.

Just like in LgtC and α1,3-GT, a NBO analysis showed that the interactions between the leaving group and the incoming hydroxyl group from the acceptor substrate are essential in catalysis. Moreover, there is an extra interaction between the UDP and the amide group from the Thr7.

Finally, an intra-substrate interaction involving the NAc group of α-GalNAc has been described. Essentially, the change in the sugar ring puckering along the catalytic mechanism promotes N-acetamide to interact with the leaving group. Moreover, reactivity studies with alternative donor substrates (i.e. UDP-Gal, 2’-deoxy-Gal and 2’-oxymethyl-Gal) support the importance that the functional group in position 2’ has in catalysis. In that sense, our results suggest that human ppGaNAcT-2 would be inactive if 2’-deoxy-Gal or 2’-oxymethyl-Gal are used, while UDP-Gal is confirmed as a valid substrate.

8

GENERAL DISCUSSION

The goal of the present chapter is to provide a general overview of the catalytic mechanism of ret-GTs by taking into account the results previously discussed and in the context of the corresponding mechanistic implications for these enzymes.

It is worth mentioning that the detailed understanding of the reaction mechanism and the substrate-enzyme interactions in GHs has led to important developments, such as new drugs[293] or engineered GHs for synthetic or biotechnological applications.[294-296] On the contrary, the reaction mechanism used by GTs has been a subject of debate for more than a decade and still remains an open question.[10,71] By analogy with retaining GHs, and despite an evident lack of evolutionary relatedness,[71] retaining GTs were first assumed to follow a double displacement mechanism with formation of a covalent glycosyl-enzyme intermediate (CGE). However, structural data has shown that only very few enzymes, namely family GT6, have a suitably positioned nucleophilic residue.[10] Therefore, it is most likely that most retaining GTs follow an alternative mechanism. Both experimental and theoretical studies are now showing that the proposed front-side attack mechanism is perfectly feasible,[3,247,251,260] and we have successfully modeled such catalytic alternative as the only possible one in the case of LgtC and ppGaNAcT-2. Still, it is unclear if GT6 (e.g. α1,3-GalT) family members use the double displacement strategy. We have shown that for α1,3-GalT both mechanisms could be taken place at the same time, as similar barrier heights have been calculated for them. Moreover, both mechanisms would require substrate-assisted catalysis to stabilize the negative charge on the leaving group to proceed at reliable rates. In fact, inter- or intra-substrate interactions that facilitate reaction are being described for both retaining and inverting GTs.[10,61,259,260] Apparently, different glycosyltransferases could be using different substrate-substrate interactions to promote reaction depending on the nature and the relative orientation of their ligands. In that sense, by studying in detail galactosyl transfer catalyzed by α1,3-GalT, LgtC and ppGaNAcT-2; we have shown that the last two ret-GTs bind the substrates in a relative orientation very convenient for catalysis, as substrate-substrate interactions can be readily established that efficiently participate in the stabilization of the β-phosphate negative charge (*"substrate-assisted catalysis"*). In contrast, the binding orientation and interactions that donor and acceptor must adopt in α1,3-GalT in order to achieve the desired reaction specificity (α1-3 linkage) reduce the number of interactions that facilitate initial UDP-Gal bond cleavage.

What is emerging as a common inter-substrates interaction in all ret-GTs is an hydrogen bond between the β-phosphate and the attacking hydroxyl of the acceptor molecule ($O3_{\beta\text{-Gal}}$, $O4_{\beta\text{-Gal}}$ and $OG1_{T7}$ in α1,3-GalT, LgtC and ppGaNAcT-2, respectively). The latest theoretical and experimental works on ret- GTs[3,61,99,247,251,260] have confirmed that the final accepting base for the proton is the same phosphate group and support the existence of such a hydrogen bond at the transition state. For LgtC and ppGaNAcT-2 (GT-A fold) this hydrogen bond is already present in the reactants. For OtsA (GT-B), the interaction in the reactants is with another oxygen of the β-phosphate and its reorientation towards the glycosidic oxygen has been suggested to be a driving force for the reaction.[260] In α1,3-GalT, this inter-substrate interaction is missing in the reactants but will be formed along the galactosyl transfer and will significantly contribute to the TS stabilization. The reason is that in the reactants $O3_{\beta\text{-Gal}}$ is making a hydrogen bond with $O4_{\beta\text{-Gal}}$, which is in turn hydrogen bonded to $OE1_{E317}$. We propose that, as a consequence of this, α1,3-GalT requires the presence of a nucleophile (Glu317) in the β-face of UDP-Gal to assist initial leaving group departure (*"nucleophilically assisted catalysis"*). However, pushing of UDP by Glu317 is not enough to achieve the catalytic efficiency and proper inter-substrates interactions are still needed.

Once again, the presence of Glu317 in α1,3-GalT, which is also involved in proper binding of LAT, opens the door to the possibility of a double displacement mechanism. However, it should be notice that even the double displacement requires the $O3B_{UDP}$–$HO3_{\beta\text{-Gal}}$ interaction as the nucleophilic strength of Glu317 is compromised by its interaction with $O4_{\beta\text{-Gal}}$. On the other hand, we have described important stabilization effects by charged residues in the active of these enzymes but contrary to what has recently been proposed based on model compounds,[297] our results suggest that the tight coordination of the pyrophosphate by positively charged residues is not enough and that substrate-assisted catalysis by the sugar acceptor hydroxyl group is required to achieve the observed reaction rates for ret-GTs. Moreover, and according to our calculations, proton transfer will not occur until the end of the reaction, but a tight hydrogen bond with the glycosidic oxygen needs to be formed earlier to allow leaving group departure, even if a nucleophile is present.

Another intra-substrate interaction involving the functional group in position 2' (i.e. next to the anomeric carbon) with the UDP in the donor substrates has also been found to contribute to catalysis in the three enzymes under study. The only difference is that this interaction is already present in the Michaelis complex of LgtC and α1,3-GalT, while in ppGaNAcT-2 it is the change in the ring puckering along the reaction, which leads to the interaction between the 2'-NAc of the α-GalNAc and the UDP.

Finally, in the case of LgtC and ppGaNAcT-2, there is an additional inter-substrates interaction that contributes to the stabilization of the oxocarbenium-like species and involves the hydroxyl $OH3_{(\beta\text{-Gal})}$ or the backbone amide from Thr7, respectively.

From the enzyme's point of view, the use of the acceptor substrate to promote donor-glycosidic bond cleavage imposes that reaction will more easily be initiated (at suitable rates) once the ternary complex will be formed. This could be a way to slow down undesired hydrolysis in retaining GTs. In α1,3-GalT, hydrolysis of the donor substrate is slower than the transfer reaction[87] and for human ABO(H) blood group related retaining GTs, which also have a glutamate as a putative nucleophile, slow hydrolysis of the donor with retention of the configuration has been reported in the absence of the acceptor substrate.[298] On the contrary, in retaining glycosidases (GHs), for which the double displacement mechanism first proposed by Koshland[80] is generally accepted, hydrolysis is the target reaction. The difficulty of breaking the glycosidic bond between two sugars and the limited potential stabilization of the leaving group departure by the acceptor substrate (a water molecule), are overcome by the implication of a protein residue acting as an acid catalyst that supplies the required proton to the leaving group. By doing this, CGE formation is made energetically accessible.[10] In transglycosidases,[299] the transfer of the monosaccharide follows a ping pong mechanism; thus, the acceptor substrate is not present in the active site to cleave the glycosidic bond of the donor substrate and a mechanism equivalent to the one used by retaining GHs has prevailed.

Finally, and according to our results, we propose that it is the identity of the donor and acceptor substrates (i.e. specificity of the enzyme), as well as their relative orientation in the active site of ret-GTs, which ultimately determine the putative inter- and intra-substrate interactions. These interactions are absolutely necessary (regardless of the catalytic mechanism) since the stabilizing residues, or even the putative nucleophiles in the active site, cannot ensure by themselves the catalytic efficiency achieved by ret-GTs. These inter- and intra-substrate interactions will, in turn, modulate the energy landscape of the system and, ultimately, the exact reaction mechanism followed by the enzyme, being the front-side attack the central one for ret-GTs.

PART V

General Conclusions

9

GENERAL CONCLUSIONS

Getting a clear picture of the reaction mechanism used by retaining glycosyltransferases (ret-GTs) is very difficult experimentally and remains one of the fundamental challenges in glycosciences. In the present thesis we have used QM(DFT)/MM calculations on fully solvated ret-GTs to shed light on this topic. More specifically, we have studied the reactions catalyzed by LgtC from *N. meningitides* (family GT8), bovine α1,3-GalT (family GT6) and human ppGaNAcT-2 (family 27). The different mechanisms proposed in the literature (S_Ni, S_Ni-like, and double-displacement mechanism via the formation of a CGE intermediate) have been investigated and compared. Alternative substrates and mutants were also considered, which provides us with a detailed description of the reaction catalyzed by these enzymes that we expect to be valuable to the experimental groups working in the field.

In that sense, the main conclusions drawn from our theoretical outcomes are:

1) A front-side attach mechanism (i.e. S_Ni or S_Ni-like) seems to be the general mechanism for ret-GTs. However, enzymes of the family 6 (with a strong nucleophile on the β-face of the donor substrate), could also operate through a double-displacement mechanism. At least in the case of α1,3-GalT these

mechanisms are taking place with similar rates in a competitive o complementary manner.

2) Stabilization of the developing negative charge on the leaving group (UDP) is very important to catalysis in ret-GTs. Such stabilization is performed by protein residues as well as by the same substrates.

3) The intra- and inter-substrates interactions are essential in the catalytic process to assist the UDP-sugar bond cleavage and thus to obtain reliable kinetic rates, in both the double-displacement or front-side attack mechanisms. We have used the term *"substrate-assisted"* catalysis to refer to them.

4) The most important interaction between the substrates is the hydrogen bond between the β-phosphate and the attacking hydroxyl group of the acceptor substrate, although proton transfer does not occur until the last step of the reaction. The latest feature is found independently of the catalytic mechanism followed by the enzyme.

5) The requirement of the inter-substrates interactions in the catalysis could be a way to slow down undesired hydrolysis of the donor substrate.

6) Some ret-GT like α1,3-GalT may require a strong nucleophile on the β-face of the donor sugar substrate for catalysis, without necessarily involving the formation of a covalent enzyme-glycosyl intermediate (i.e. in a double-displacement mechanism). This nucleophile would be a key residue to assist the leaving group departure in an *"assisted"* S_Ni or S_Ni-like mechanism (*"nucleophilically assisted"* catalysis). Moreover, the foregoing would also be consistent with the experimental mutagenesis data.

7) The nucleophilic strength of the putative nucleophile is limited by its interactions with the acceptor substrate or water molecules in the Michaelis or the binary complex, respectively.

8) The covalent intermediate (CGE) formation is quite endoergic and it could easily go back to reactants in both the ternary and binary complexes. This low stability of the CGE, altogether with the reduced nucleophilic strength of the nucleophile (i.e. Glu317 in α1,3-GalT) might explain why the experimentalists have not been able to isolate such a covalent intermediate so far.

In an attempt to put everything together and to provide a more general overview of the catalytic mechanism of ret-GTs we conclude that the catalytic strategy followed by each enzyme-substrate complex is influenced by several factors, including the nature of the substrates and the specificity of the glycosidic linkage to be formed. The foregoing strongly conditions the orientation of the bound substrates in the active site and, accordingly, the interactions between the substrates, which have shown to be a key factor in catalysis. In cases where the leaving group departure is hampered, ret-GTs like α1,3-GalT may need a strong nucleophile to assist this process, but curiously, it does not necessarily imply the formation of a covalent glysosyl-enzyme intermediate, as has always been assumed by the scientific community.

REFERENCES

(1) Grisham, C. M.; Reginald, H. G. *Biochemistry*; Saunders College Pub.: Philadelphia, 1999.

(2) Charnock, S.; Davies, G. *Biochemistry* **1999**, *38*, 6380-6385.

(3) Tvaroška, I. *Carbohydr. Res.* **2004**, *339*, 1007-1014.

(4) Pauling, L. *Chem. Eng. News* **1946**, *34*, 1375-1377.

(5) Moréra, S.; Larivière, L.; Kurzeck, J.; Aschke-Sonnenborn, U.; Freemont, P.; Janin, J.; Rüger, W. *J. Mol. Biol.* **2001**, *311*, 569-577.

(6) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90-93.

(7) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938-945.

(8) Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D. *EMBO J.* **2008**, *27*, 234-243.

(9) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Sience* **2004**, *303*, 186-195.

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

(11) Varki, A. *Glycobiology* **1993**, *3*, 97-130.

(12) Akira, K. *Acc. Chem. Res.* **1993**, *26*.

(13) Dwek, R. *Chem. Rev.* **1996**, *96*, 683-720.

(14) Dennis, J.; Granovsky, M.; Warren, C. *BioEssays : news and reviews in molecular, cellular and developmental biology* **1999**, *21*, 412-421.

(15) Dennis, J.; Granovsky, M.; Warren, C. *Biochim. Biophys. Acta* **1999**, *1473*, 21-34.

(16) Brockhausen, I. *Biochim. Biophys. Acta* **1999**, *1473*, 67-95.

(17) Fukuda, M. *Cancer Res.* **1996**, *56*, 2237-2244.

(18) Jack, D.; Dodds, A.; Anwar, N.; Ison, C.; Law, A.; Frosch, M.; Turner, M.; Klein, N. *Journal of immunology (Baltimore, Md. : 1950)* **1998**, *160*, 1346-1353.

(19) Ljungh, A.; Moran, A.; Wadström, T. *FEMS Immunol. Med. Microbiol.* **1996**, *16*, 117-126.

(20) Moran, A.; Prendergast, M.; Appelmelk, B. *FEMS Immunol. Med. Microbiol.* **1996**, *16*, 105-115.

(21) Breton, C.; Snajdrová, L.; Jeanneau, C.; Koca, J.; Imberty, A. *Glycobiology* **2006**, *16*.

(22) Cantarel, B.; Coutinho, P.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, 8.

(23) Campbell, J.; Davies, G.; Bulone, V.; Henrissat, B. *Biochem. J.* **1997**, *326 (Pt 3)*, 929-939.

(24) Bourne, Y.; Henrissat, B. *Curr. Opin. Struct. Biol.* **2001**, *11*, 593-600.

(25) Coutinho, P.; Deleury, E.; Davies, G.; Henrissat, B. *J. Mol. Biol.* **2003**, *328*, 307-317.

(26) Unligil, U.; Rini, J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 510-517.

(27) Hu, Y.; Walker, S. *Chem. Biol.* **2002**, *9*, 1287-1296.

(28) Brown, K.; Pompeo, F.; Dixon, S.; Mengin-Lecreulx, D.; Cambillau, C.; Bourne, Y. *EMBO J.* **1999**, *18*, 4096-4107.

(29) Wiggins, C.; Munro, S. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 7945-7950.

(30) Breton, C.; Bettler, E.; Joziasse, D.; Geremia, R.; Imberty, A. *J. Biochem. (Tokyo).* **1998**, *123*, 1000-1009.

(31) Breton, C.; Imberty, A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 563-571.

(32) Pak, J.; Arnoux, P.; Zhou, S.; Sivarajah, P.; Satkunarajah, M.; Xing, X.; Rini, J. *J. Biol. Chem.* **2006**, *281*, 26693-26701.

(33) Vrielink, A.; Rüger, W.; Driessen, H.; Freemont, P. *EMBO J.* **1994**, *13*, 3413-3422.

(34) Campbell, R.; Mosimann, S.; Tanner, M.; Strynadka, N. *Biochemistry* **2000**, *39*, 14993-15001.

(35) Liu, J.; Mushegian, A. *Protein Sci.* **2003**, *12*, 1418-1431.

(36) Strahl-Bolsinger, S.; Immervoll, T.; Deutzmann, R.; Tanner, W. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 8164-8168.

(37) Takahashi, M.; Inoue, N.; Ohishi, K.; Maeda, Y.; Nakamura, N.; Endo, Y.; Fujita, T.; Takeda, J.; Kinoshita, T. *EMBO J.* **1996**, *15*, 4254-4261.

(38) Maeda, Y.; Watanabe, R.; Harris, C.; Hong, Y.; Ohishi, K.; Kinoshita, K.; Kinoshita, T. *EMBO J.* **2001**, *20*, 250-261.

(39) Guerin, M.; Kordulakova, J.; Schaeffer, F.; Svetlikova, Z.; Buschiazzo, A.; Giganti, D.; Gicquel, B.; Mikusova, K.; Jackson, M.; Alzari, P. *J. Biol. Chem.* **2007**, *282*, 20705-20714.

(40) Guerin, M.; Schaeffer, F.; Chaffotte, A.; Gest, P.; Giganti, D.; Korduláková, J.; van der Woerd, M.; Jackson, M.; Alzari, P. *J. Biol. Chem.* **2009**, *284*, 21613-21625.

(41) Boix, E.; Swaminathan, G.; Zhang, Y.; Natesh, R.; Brew, K.; Acharya, K. *J. Biol. Chem.* **2001**, *276*, 48608-48614.

(42) Boix, E.; Zhang, Y.; Swaminathan, G.; Brew, K.; Acharya, K. *J. Biol. Chem.* **2002**, *277*, 28310-28318.

(43) Ramakrishnan, B.; Balaji, P.; Qasba, P. *J. Mol. Biol.* **2002**, *318*, 491-502.

(44) Unligil, U.; Zhou, S.; Yuwaraj, S.; Sarkar, M.; Schachter, H.; Rini, J. *EMBO J.* **2000**, *19*, 5269-5280.

(45) Ramakrishnan, B.; Qasba, P. *J. Mol. Biol.* **2001**, *310*, 205-218.

(46) Pedersen, L.; Tsuchida, K.; Kitagawa, H.; Sugahara, K.; Darden, T.; Negishi, M. *J. Biol. Chem.* **2000**, *275*, 34580-34585.

(47) Chiu, C.; Watts, A.; Lairson, L.; Gilbert, M.; Lim, D.; Wakarchuk, W.; Withers, S.; Strynadka, N. *Nat. Struct. Mol. Biol.* **2004**, *11*, 163-170.

(48) Flint, J.; Taylor, E.; Yang, M.; Bolam, D.; Tailford, L.; Martinez-Fleites, C.; Dodson, E.; Davis, B.; Gilbert, H.; Davies, G. *Nat. Struct. Mol. Biol.* **2005**, *12*, 608-614.

(49) Jamaluddin, H.; Tumbale, P.; Withers, S.; Acharya, K.; Brew, K. *J. Mol. Biol.* **2007**, *369*, 1270-1281.

(50) Alfaro, J.; Zheng, R.; Persson, M.; Letts, J.; Polakowski, R.; Bai, Y.; Borisova, S.; Seto, N.; Lowary, T.; Palcic, M.; Evans, S. *J. Biol. Chem.* **2008**, *283*, 10097-10108.

(51) Gunasekaran, K.; Nussinov, R. *Chembiochem : a European journal of chemical biology* **2004**, *5*, 224-230.

(52) Snajdrová, L.; Kulhánek, P.; Imberty, A.; Koča, J. *Carbohydr. Res.* **2004**, *339*, 995-1006.

(53) Milac, A.; Buchete, N.; Fritz, T.; Hummer, G.; Tabak, L. *J. Mol. Biol.* **2007**, *373*, 439-451.

(54) Qasba, P.; Ramakrishnan, B.; Boeggeman, E. *Trends Biochem. Sci.* **2005**, *30*, 53-62.

(55) Petrova, P.; Koca, J.; Imberty, A. *J. Am. Chem. Soc.* **1999**, 5535-5547.

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

(57) Morrison, J.; Ebner, K. *J. Biol. Chem.* **1971**, *246*, 3992-3998.

(58) Nishikawa, Y.; Pegg, W.; Paulsen, H.; Schachter, H. *J. Biol. Chem.* **1988**, *263*, 8270-8281.

(59) Ramakrishnan, B.; Boeggeman, E.; Ramasamy, V.; Qasba, P. *Curr. Opin. Struct. Biol.* **2004**, *14*, 593-600.

(60) Michael, L. S. *Chem. Rev.* **1990**, *90*.

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

(62) Tarbouriech, N.; Charnock, S.; Davies, G. *J. Mol. Biol.* **2001**, *314*, 655-661.

(63) Jinek, M.; Chen, Y.-W.; Clausen, H.; Cohen, S.; Conti, E. *Nat. Struct. Mol. Biol.* **2006**, *13*, 945-946.

(64) Ohtsubo, K.; Imajo, S.; Ishiguro, M.; Nakatani, T.; Oka, S.; Kawasaki, T. *J. Biochem. (Tokyo).* **2000**, *128*, 283-291.

(65) Kakuda, S.; Shiba, T.; Ishiguro, M.; Tagawa, H.; Oka, S.; Kajihara, Y.; Kawasaki, T.; Wakatsuki, S.; Kato, R. *J. Biol. Chem.* **2004**, *279*, 22693-22703.

(66) Garinot-Schneider, C.; Lellouch, A.; Geremia, R. *J. Biol. Chem.* **2000**, *275*, 31407-31413.

(67) Larivière, L.; Gueguen-Chaignon, V.; Moréra, S. *J. Mol. Biol.* **2003**, *330*, 1077-1086.

(68) Grizot, S.; Salem, M. l.; Vongsouthi, V.; Durand, L.; Moreau, F. o.; Dohi, H.; Vincent, S. p.; Escaich, S.; Ducruix, A. *J. Mol. Biol.* **2006**, *363*, 383-394.

(69) Sun, H.-Y.; Lin, S.-W.; Ko, T.-P.; Pan, J.-F.; Liu, C.-L.; Lin, C.-N.; Wang, A.; Lin, C.-H. *J. Biol. Chem.* **2007**, *282*, 9973-9982.

(70) Ni, L.; Chokhawala, H.; Cao, H.; Henning, R.; Ng, L.; Huang, S.; Yu, H.; Chen, X.; Fisher, A. *Biochemistry* **2007**, *46*, 6288-6298.

(71) Davies, G. S.; Withers, S. G. In *Comprehensive Biological Catalysis*; Sinnott, M. L., Ed. London:Academic, 1998, p 119–208.

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

(73) Patenaude, S.; Seto, N.; Borisova, S.; Szpacenko, A.; Marcus, S.; Palcic, M.; Evans, S. *Nat. Struct. Biol.* **2002**, *9*, 685-690.

(74) Lobsanov, Y.; Romero, P.; Sleno, B.; Yu, B.; Yip, P.; Herscovics, A.; Howell, P. *J. Biol. Chem.* **2004**, *279*, 17921-17931.

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

(76) Reinert, D.; Jank, T.; Aktories, K.; Schulz, G. *J. Mol. Biol.* **2005**, *351*, 973-981.

(77) Martinez-Fleites, C.; Proctor, M.; Roberts, S.; Bolam, D.; Gilbert, H.; Davies, G. *Chem. Biol.* **2006**, *13*, 1143-1152.

(78) Gibson, R.; Turkenburg, J.; Charnock, S.; Lloyd, R.; Davies, G. *Chem. Biol.* **2002**, *9*, 1337-1346.

(79) Larivière, L.; Sommer, N.; Moréra, S. *J. Mol. Biol.* **2005**, *352*, 139-150.

(80) Koshland, D. E. *Biol. Rev. Camb. Philos. Soc.* **1953**, *28*, 416–36.

(81) Sinnott, M.; Souchard, I. *Biochem. J.* **1973**, *133*, 89-98.

(82) Kempton, J.; Withers, S. *Biochemistry* **1992**, *31*, 9961-9969.

(83) Vocadlo, D.; Wicki, J.; Rupitz, K.; Withers, S. *Biochemistry* **2002**, *41*, 9727-9735.

(84) Xicai, H.; Kelly, S. E. T.; Andrew, J. B. *J. Am. Chem. Soc.* **1997**, *119*.

(85) Vocadlo, D.; Davies, G.; Laine, R.; Withers, S. *Nature* **2001**, *412*, 835-838.

(86) Lairson, L. L.; Chiu, C. P.; Ly, H. D.; He, S.; Wakarchuk, W. W.; Strynadka, N. C.; Withers, S. G. *J. Biol. Chem.* **2004**, *279*, 28339-28344.

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

(88) Viladot, J.; de Ramon, E.; Durany, O.; Planas, A. *Biochemistry* **1998**, *37*, 11332-11342.

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

(90) Ly, H.; Lougheed, B.; Wakarchuk, W.; Withers, S. *Biochemistry* **2002**, *41*, 5075-5085.

(91) Cowdrey, W.; Hughes, E. D.; Ingold, C. K.; Masterman, S.; Scott, A. *J. Chem. Soc.* **1937**, 1252-1271.

(92) Hughes, E. D.; Ingold, C. K.; Whitfield, I. C. *Nature* **1941**, *147*, 206-207.

(93) Edward, S. L.; Charles, E. B. *J. Am. Chem. Soc.* **1952**, *74*, 308-311.

(94) Pedersen, L.; Dong, J.; Taniguchi, F.; Kitagawa, H.; Krahn, J.; Pedersen, L.; Sugahara, K.; Negishi, M. *J. Biol. Chem.* **2003**, *278*, 14420-14428.

(95) Tvaroška, I.; André, I.; Carver, J. P. *J. Am. Chem. Soc.* **2000**, *122*, 8762-8776.

(96) Tvaroška, I.; André, I.; Carver, J. P. *Glycobiology* **2003**, *13*, 559-566.

(97) Kozmon, S.; Tvaroška, I. *J. Am. Chem. Soc.* **2006**, *128*, 16921-16927.

(98) Krupicka, M.; Tvaroška, I. *J. Phys. Chem. B* **2009**, *113*, 11314-11319.

(99) André, I.; Tvaroška, I.; Carver, J. P. *Carbohydr. Res.* **2003**, *338*, 865-877.

(100) Hart, G. W.; Copeland, R. J. *Cell* **2010**, *143*, 672-676.

(101) In *"Press Release: The 1998 Nobel Prize in Chemistry". Nobelprize.org.* ; http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/press.html: 2013.

(102) Lennard-Jones, J. E. *Proc. R. Soc. Lond. A* **1924**, *106*, 463–477.

(103) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253–287.

(104) Rokhlin, V. *J. Comput. Phys.* **1985**, *60*, 187-207.

(105) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667-683.

(106) Levitt, M.; Lifson, S. *J. Mol. Biol.* **1969**, *46*, 269-279.

(107) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582-594.

(108) Allinger, N. L.; Sprague, J. T. *J. Am. Chem. Soc.* **1973**, *95*, 3893-3907.

(109) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G. S.; Profeta, J.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

(110) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1-41.

(111) Jorgensen, W., L. ; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

(112) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596-3607.

(113) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

(114) MacKerell, A. D. J.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(115) Price, D. J.; Brooks, C. L. I. *J. Comput. Chem.* **2002**, *23*, 1045-1057.

(116) Born, M.; Oppenheimer, J. R. *Ann. Physik* **1927**, *84*, 457.

(117) Hartree, D. R. *Proc. Cambridge Philos. Soc.* **1928**, *24*, 89.

(118) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69-89.

(119) Hall, G. G.; Lennardjones, J. *Proc. Roy. Soc. Lond. Math. Phys. Sci.* **1951**, *205*, 357-374.

(120) Jensen, F. *Introduction to Computational Chemistry*; 2 ed.; John Wiley & Sons Ltd: Chichester, 2007.

(121) Freed, K. F. *Annu. Rev. Phys. Chem.* **1971**, *22*, 313-&.

(122) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359-401.

(123) Scuseria, G. E.; Schaefer, H. F. *J. Chem. Phys.* **1989**, *90*, 3700-3703.

(124) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Headgordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479-483.

(125) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, 129-135.

(126) Pople, J. A.; Segal, G. A. *J. Chem. Phys* **1965**, *43*, 136-149.

(127) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899-4907.

(128) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.

(129) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209-220.

(130) Winget, P.; Selcuki, C.; Horn, A. H. C.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110*, 254-266.

(131) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260-7268.

(132) Slater, P. C.; Koster, G. F. *Phys. Rev.* **1954**, *94*, 1498-1524.

(133) Parr, R. G. *Annu. Rev. Phys. Chem.* **1983**, *34*, 631-656.

(134) Parr, R. G.; Yang, W. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701-728.

(135) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651-667.

(136) Gonze, X. *Phys. Rev. A: At., Mol., Opt. Phys* **1995**, *52*, 1096-1114.

(137) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem. A* **1996**, *100*, 12974-12980.

(138) Baerends, E. J.; Gritsenko, O. V. *J. Phys. Chem. A* **1997**, *101*, 5383-5403.

(139) Chermette, H. *Coord. Chem. Rev.* **1998**, *178-180*, 699-721.

(140) Siegbahn, P. E. M.; Blomberg, M. R. A. *Annu. Rev. Phys. Chem.* **1999**, *50*, 221-249.

(141) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793-1873.

(142) Burke, K.; Werschnik, J.; Gross, E. K. U. *J. Chem. Phys.* **2005**, *123*, 62206.

(143) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439-10452.

(144) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, B864.

(145) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.

(146) Thomas, L. H. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542-548.

(147) Fermi, E. *Rend. Accad. Naz. Lincei* **1927**, *6*, 602-607.

(148) Dirac, P. A. M. *Proc. Cambridge Philos. Soc.* **1930**, *26*, 376-385.

(149) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385-390.

(150) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524-4529.

(151) Becke, A. D. *J. Chem. Phys.* **1992**, *96*, 2155-2160.

(152) Becke, A. D. *J. Chem. Phys.* **1992**, *97*, 9173-9177.

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

(154) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040-1046.

(155) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554-8560.

(156) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(157) Levy, M.; Perdew, J. P. *J. Chem. Phys.* **1986**, *84*, 4519-4523.

(158) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664-675.

(159) Handy, N. C.; Cohen, A. J. *J. Mol. Phys.* **2001**, *99*, 403-412.

(160) Xu, X.; Goddard, W. A., III *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673-2677.

(161) Kurth, S.; Perdew, J. P.; Blaha, P. *Int. J. Quantum Chem.* **1999**, *75*, 889-909.

(162) Wang, Y.; Perdew, J. P. *Phys. Rev. B* **1991**, *43*, 8911-8916.

(163) Perdew, J. P.; Levy, M. *Phys. Rev. B: Condens. Matter* **1985**, *31*, 6264-6272.

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

(165) Adamo, C.; Barone, V. *J. Chem. Phys.* **2002**, *116*, 5933-5940.

(166) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400-410.

(167) Krieger, J. B.; Chen, J.; Iafrate, G. J.; Savin, A. *Electron Correl. Mater. Prop.* **1999**, 463-477.

(168) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401/1-146401/4.

(169) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

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

(171) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.

(172) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364-382.

(173) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. J. *J. Chem. Phys.* **2004**, *120*, 6898.

(174) Kóňa, J.; Tvaroška, I. *Chem. Pap.* **2009**, *63*, 598−607.

(175) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. *J. Chem. Theory Comp.* **2009**, *5*, 679-692.

(176) Senn, H.; Thiel, W. *Angew. Chem.* **2009**, *48*, 1198-1229.

(177) Lodola, A.; Woods, C. J.; Mulholland, A. J. *Annu. Rep. Comput. Chem.* **2008**, *4*, 155-169.

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

(179) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720-1735.

(180) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

(181) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.

(182) Antes, I.; Thiel, W. *J. Phys. Chem. B.* **1999**, *103*, 9290-9295.

(183) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.

(184) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578-9584.

(185) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959-1967.

(186) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452-10467.

(187) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269-282.

(188) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *12*, 4714-4721.

(189) Heyden, A.; Lin, H.; Truhlar, D. *J. Phys. Chem. B* **2007**, *111*, 2231-2241.

(190) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.

(191) Svensson, M.; Humbel, S. p.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

(192) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419-1432.

(193) Riccardi, D.; Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467-6478.

(194) Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.

(195) Hockney, R. W. *Meth. Comput. Phys.* **1970**, *9*, 135-211.

(196) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511-519.

(197) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.

(198) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(199) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327-341.

(200) Krautler, V.; van Gunsteren, W. F.; Hunenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501-508.

(201) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24-34.

(202) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463-1472.

(203) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585-590.

(204) van Gunsteren, W. F.; Berendsen, H. J. *Mol. Phys.* **1977**, *34*, 1311-1327.

(205) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187-199.

(206) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

(207) Michaelis, L.; Menten, M. L. *Biochem. Z.* **1913**, *49*, 333-369.

(208) Stroppolo, M. E.; Falconi, M.; Caccuri, A. M.; Desideri, A. *Cell. Mol. Life Sci.* **2001**, *58*, 1451–1460.

(209) Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252-261.

(210) Crehuet, R.; Field, M. J. *J. Chem. Phys.,* **2003**, *118*, 9563-9571.

(211) Fletcher, R. *Practical methods of optimization.* ; 2nd Ed. ed.; John Wiley & Sons: Essex, United Kingdom, 1987.

(212) Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. *Biochem. J.* **1997**, *326*, 929-942.

(213) Tzeng, Y. L.; Stephens, D. S. *Microbes Intect.* **2000**, *2*, 687-700.

(214) Takayama, S.; Chung, S. G.; Igarashi, Y.; Ichikawa, Y.; Sepp, A.; Lechler, R. I.; Wu, J.; Hayashi, T.; Siuzdak, G.; Wong, C. H. *Bioorg. Med. Chem.* **1999**, *7*, 401-409.

(215) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704-721.

(216) MacKerell, A. D. J.; Feig, M.; Brooks, C. L. I. *J. Am. Chem. Soc.* **2004**, *126*, 698-699.

(217) Brooks, B. R.; Brooks, C. L. I.; MacKerell, A. D. J.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545-1614.

(218) Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. J. *J. Chem. Theory Comput.* **2009**, *5*, 2353-2370.

(219) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (Theochem.)* **2003**, *632*, 1-28.

(220) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165-169.

(221) Frisch, M. T. G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J.; Vreven, T.; Kudin, K.; Burant, J.; Millam, J.; Iyengar, S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima,

T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.; Hratchian, H.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P.; Morokuma, K.; Voth, G.; Salvador, P.; Dannenberg, J.; Zakrzewski, V.; Dapprich, S.; Daniels, A.; Strain, M.; Farkas, O.; Malick, D.; Rabuck, A.; Raghavachari, K.; Foresman, J.; Ortiz, J.; Cui, Q.; Baboul, A.; Clifford, S.; Cioslowski, J.; Stefanov, B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.; Fox, D.; Keith, T.; Al-Laham, M.; Peng, C.; Nanayakkara, A.; Challacombe, M.; Gill, P.; Johnson, B.; Chen, W.; Wong, M.; Gonzalez, C.; Pople, J. *Gaussian 03, revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.*

(222) Thiel, W.; Program MNDO2005, version 7.0 ed. Max-Planck-Institut für Kohlenforschung: Mülheim, 2005.

(223) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200-1211.

(224) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100.

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

(226) Smith, W.; Forester, T. R. *J. Mol. Graph.* **1996**, *14*, 136-141.

(227) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580-10594.

(228) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133-6141.

(229) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79-92.

(230) Nocedal, J. *Math. Comp.* **1980**, *35*, 773-782.

(231) Liu, D. C.; Nocedal, J. *Math. Programming* **1989**, *45*, 503-528.

(232) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52-57.

(233) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385-395.

(234) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177-2186.

(235) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571-2577.

(236) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-289.

(237) Eichkorn, K.; Welgend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119-124.

(238) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(239) Rappoport, D.; Furche, F. *J. Chem. Phys.* **2010**, *133*, 134105.

(240) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735-746.

(241) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.

(242) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2006**, *123*, 234106.

(243) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011-1021

(244) Souaille, M. *Comput. Phys. Commun.* **2001**, *135*, 40-57.

(245) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33-38.

(246) Sinnott, M. L.; Jencks, W. P. *J. Am. Chem. Soc.* **1980**, *102*, 2026– 2032.

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

(248) Kóňa, J.; Tvaroška, I. *Chem. Pap.* **2009**, *63*, 598-607.

(249) Fraústo da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of Elements. The Inorganic Chemistry of Life*; Clarendon Press: Oxford, England, 1991.

(250) Bock, C. W.; Katz, A. K.; Markham, G. D.; Glusker, J. P. *J. Am. Chem. Soc.* **1999**, *121*, 7360-7372.

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

(252) Goedl, C.; Nidetzky, B. *ChemBioChem* **2009**, *10*, 2333-2337.

(253) Zhang, Y.; Deshpande, A.; Xie, Z.; R., N.; Acharya, K. R.; Brew, K. *Glycobiology* **2004**, *14*, 1295-1302.

(254) Tumbale, P.; Jamaluddin, H.; Thiyagarajan, N.; Brew, K.; Acharya, K. R. *Biochemistry* **2008**, *47*, 8711-8718.

(255) Zhang, Y.; Swaminathan, G. J.; Deshpande, A.; Boix, E.; Natesh, R.; Xie, Z.; Acharya, K. R.; Brew, K. *Biochemistry* **2003**, *42*, 13512-13521.

(256) Molina, P.; Knegtel, R. M.; Macher, B. A. *Biochim. Biophys. Acta* **2007**, *1770*, 1266-1273.

(257) Zhang, Y.; Swaminathan, G.; Deshpande, A.; Boix, E.; Natesh, R.; Xie, Z.; Acharya, K.; Brew, K. *Biochemistry* **2003**, *42*, 13512-13521.

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

(259) Zhang, Y.; Wang, P. G.; Brew, K. *J. Biol. Chem.* **2001**, *276*, 11567-11574.

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

(261) Ruiz, J.; Tuñón, I.; Williams, I. *J. Phys. Chem. B* **2010**, *114*, 5769-5774.

(262) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211-7218.

(263) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066-4073.

(264) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899-926.

(265) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F.; NBO Version 3.1 ed.

(266) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09, revision A.1; Gaussian, Inc., Wallingford CT, 2009.*

(267) Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A. *Glycobiology* **2003**, *13*, 1R-16R.

(268) Milac, A. L.; Buchete, N. V.; Fritz, T. A.; Hummer, G.; Tabak, L. A. *J. Mol. Biol.* **2007**, *373*, 439-451.

(269) Strous, G. J. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 2694-2698.

(270) Jentoft, N. *Trends Biochem. Sci.* **1990**, *15*, 291-294.

(271) Moody, A. M.; Chui, D.; Reche, P. A.; Priatel, J. J.; Marth, J. D.; Reinherz, E. L. *Cell* **2001**, *107*, 501-512.

(272) Xu, Z.; Weiss, A. *Nat. Immunol.* **2002**, *3*, 764-771.

(273) Sauer, J.; Sigurskjold, B. W.; Christensen, U.; Frandsen, T. P.; Mirgorodskaya, E.; Harrison, M.; Roepstorff, P.; Svensson, B. *Biochim. Biophys. Acta* **2000**, *1543*, 275-293.

(274) Garner, B.; Merry, A. H.; Royle, L.; Harvey, D. J.; Rudd, P. M.; Thillet, J. *J. Biol. Chem.* **2001**, *276*, 22200-22208.

(275) Hooper, L. V.; Gordon, J. I. *Glycobiology* **2001**, *11*, 1R-10R.

(276) Yeh, J. C.; Hiraoka, N.; Petryniak, B.; Nakayama, J.; Ellies, L. G.; Rabuka, D.; Hindsgaul, O.; Marth, J. D.; Lowe, J. B.; Fukuda, M. *Cell* **2001**, *105*, 957-969.

(277) Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T. *Cell* **2000**, *103*, 467-479.

(278) Alfalah, M.; Jacob, R.; Preuss, U.; Zimmer, K. P.; Naim, H.; Naim, H. Y. *Curr. Biol.* **1999**, *9*, 593-596.

(279) Altschuler, Y.; Kinlough, C. L.; Poland, P. A.; Bruns, J. B.; Apodaca, G.; Weisz, O. A.; Hughey, R. P. *Mol. Biol. Cell* **2000**, *11*, 819-831.

(280) Breuza, L.; Garcia, M.; Delgrossi, M.-H. H.; Le Bivic, A. *Exp. Cell Res.* **2002**, *273*, 178-186.

(281) Naim, H. Y.; Joberty, G.; Alfalah, M.; Jacob, R. *J. Biol. Chem.* **1999**, *25*, 17961-17967.

(282) Zheng, X.; Sadler, J. E. *J. Biol. Chem.* **2002**, *277*, 6858-6863.

(283) Marchler-Bauer, A.; Bryant, S. H. *Nucleic Acids Res.* **2004**, *32*, 327-331.

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

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

(286) Hagen, F. K.; Ten Hagen, K. G.; Beres, T. M.; Balys, M. M.; VanWuyckhuyse, B. C.; Tabak, L. A. *J. Biol. Chem.* **1997**, *272*, 13843-13848.

(287) Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A. *Glycobiology* **2003**, *13*, 1R-16R.

(288) Wragg, S.; Hagen, F. K.; Tabak, L. A. *J. Biol. Chem.* **1995**, *270*, 16947-16954.

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

(290) Vocadlo, D. J.; Whiters, S. G. *Biochemistry* **2005**, *44*, 12809-12818.

(291) Tvaroška, I.; Kozmon, S.; Wimmerová, M.; Koča, J. *J. Am. Chem. Soc.* **2012**, *134*, 15563-15571.

(292) Wandall, H. H.; Hassan, H.; Mirgorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. *J. Biol. Chem.* **1997**, *272*, 23503-23514.

(293) von Itzstein, M.; Wu, W.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. *Nature* **1993**, *363*, 418-423.

(294) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 5583-5584.

(295) Malet, C.; Planas, A. *FEBS Lett.* **1998**, *440*, 208.

(296) Moracci, M.; Tricone, A.; Perugino, G.; Ciaramella, M.; Rossi, M. *Biochemistry* **1998**, 17262.

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

(298) Soya, N.; Shoemaker, G. K.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2009**, *19*, 1224-1234.

(299) Amaya, M. F.; Watts, A. G.; Damager, I.; Wehenkel, A.; Nguyen, T.; Buschiazzo, A.; Paris, G.; Frasch, A. C.; Withers, S. G.; Alzari, P. M. *Structure* **2004**, *12*, 775-784.

APPENDIX

A1. APPENDIX TO CHAPTER 4

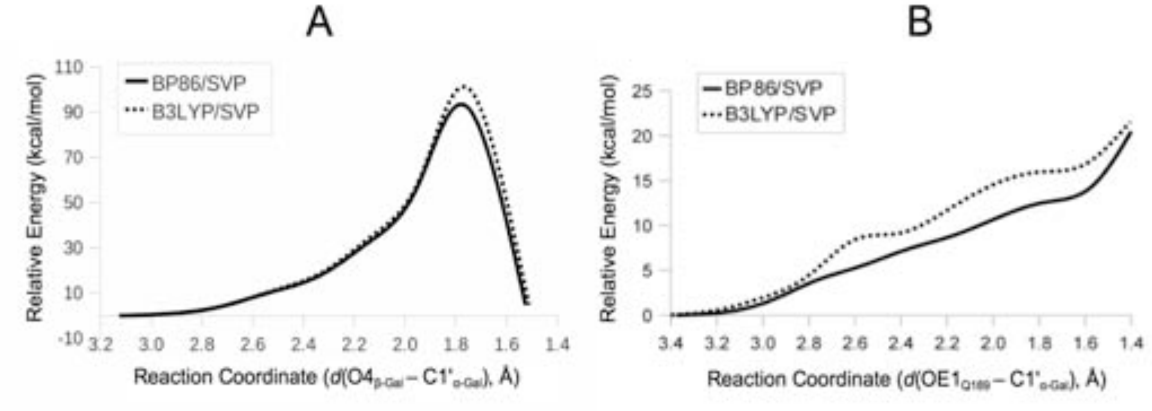

Figure A4.1. QM/CHARMM22 with QM= BP86/SVP and B3LYP/SVP energy profiles for (A) the front-side attack and (B) the first step of the double-displacement mechanism using simple distances as reaction coordinates.

Table A4.1. Bond distances (in Å) in the optimized reactants, transition state, and products for the proposed S_Ni mechanism at the SCC-DFTB level of theory.

	Reactant	**TS**	**Product**
$d(C1'_{\alpha\text{-Gal}}–O3B_{UDP})$	1.51	2.63	3.23
$d(O4_{\beta\text{-Gal}}–C1'_{\alpha\text{-Gal}})$	3.13	2.44	1.51
$d(HO4_{\beta\text{-Gal}}–O4_{\beta\text{-Gal}})$	0.99	1.15	1.69
$d(HO4_{\beta\text{-Gal}}–O3B_{UDP})$	1.78	1.27	0.99
$d(OE1_{Q189}–C1'_{\alpha\text{-Gal}})$	3.38	2.78	3.08
$d(C1'_{\alpha\text{-Gal}}–O5'_{\alpha\text{-Gal}})$	1.44	1.28	1.44

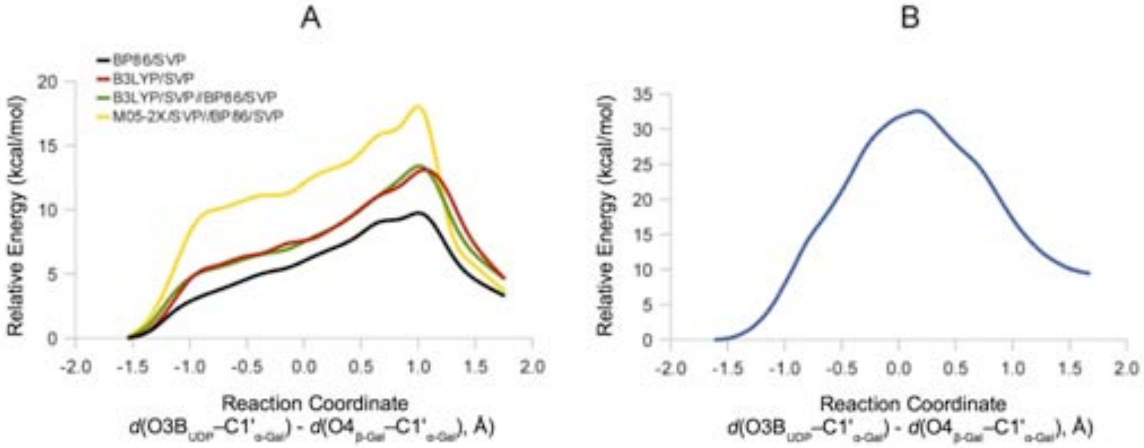

Figure A4.2. QM/CHARMM22 potential energy profiles for the front-side attack (S_Ni) mechanism at different QM levels of theory: (**A**) BP86/SVP, B3LYP/SVP, B3LYP/SVP//BP86/SVP, M05-2X/SVP//BP86/SVP; and (**B**) SCC-DFTB.

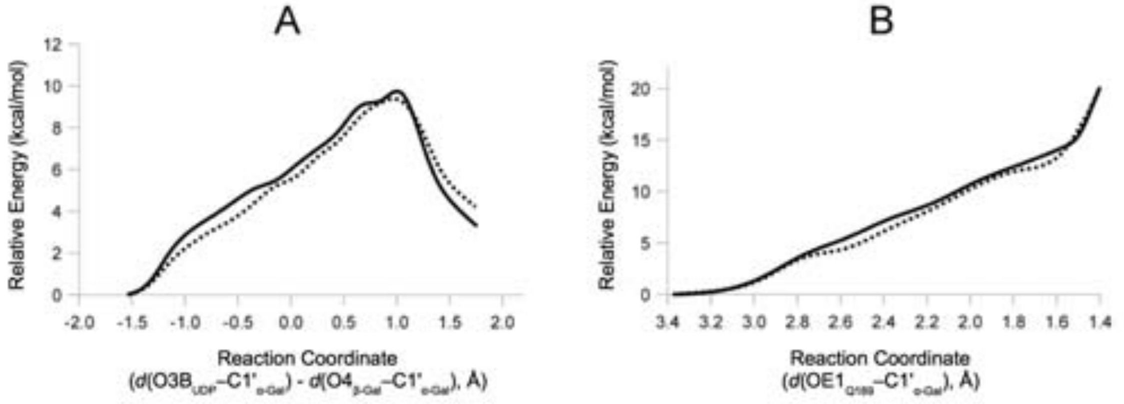

Figure A4.3. QM(BP86/SVP)/CHARMM22 potential energy profiles for (**A**) the front-side attack (S_Ni) mechanism and (**B**) the first step of the double displacement mechanism in the wild type enzyme, including or not the residues Asp188 and Lys250 in the QM region (dashed and solid lines, respectively).

Table A4.2. QM/MM potential energy barriers and reaction energies for the proposed S_Ni mechanism in the Q189A mutant enzyme calculated at stationary points optimized at the QM(BP86/SVP)/CHARMM22 level. Energies are given in kcal/mol.

	BP86			**B3LYP**			**B3LYP-D**	**M05-2X**	
	SVP	TZVP	def2-TZVPP(d)	SVP	TZVP	def2-TZVPP(d)	TZVP	SVP	TZVP
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TS	9.26	7.40	8.06	12.05	10.05	10.68	9.58	17.22	14.26
P	-0.96	0.12	0.11	-0.42	0.77	0.85	-0.94	0.02	0.68

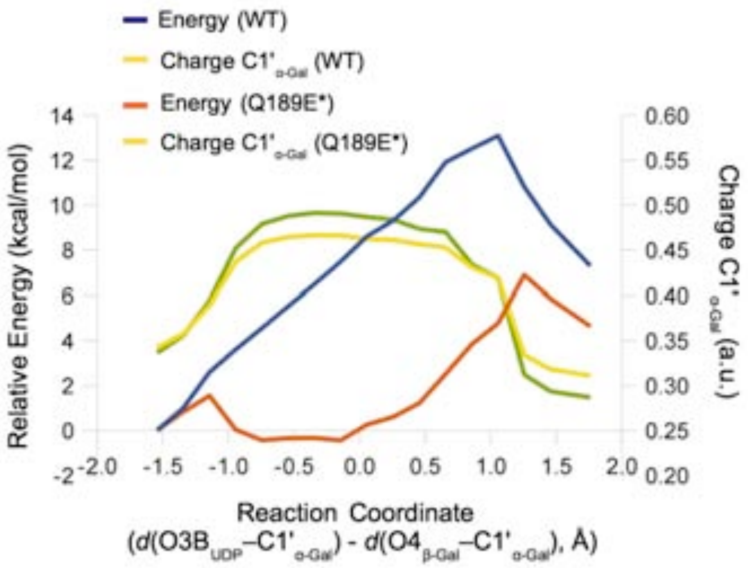

Figure A4.4. QM(BP86/SVP)/CHARMM22 potential energy profile for the proposed S_Ni mechanism in the wild type and the *pseudo* Q189E* mutant enzymes. In both cases, residue 189 is described by the MM force field. The *pseudo* Q189E* mutant is built by simply assigning glutamate charges to Gln189. Its energy profile is calculated at the optimized geometries of the wild type enzyme. QM(BP86/TZVP)/CHARMM22 charge evolution at $C1'_{\alpha\text{-Gal}}$ calculated at QM(BP86/SVP)/CHARMM22 optimized geometries is also depicted.

Table A4.3. QM/MM potential energy barriers and reaction energies for the second step of a double displacement mechanism in the Q189E mutant enzyme. The stationary points were optimized at the SCC-DFTB/CHARMM22 level. Energies are given in kcal/mol.

	SCC-DFTB	**B3LYP**			**B3LYP-D**	**M05-2X**	
		SVP	TZVP	def2-TZVPP(d)	TZVP	SVP	TZVP
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TS	28.57	36.16	33.86	34.25	32.13	41.74	39.47
P	14.87	27.71	26.38	25.12	24.48	30.65	30.18

Table A4.4. QM/MM potential energy barriers and reaction energies for the proposed S_Ni mechanism in the wild type enzyme with 2'-deoxygalactose. The stationary points were optimized at the QM(BP86/SVP)/CHARMM22 level. Energies are given in kcal/mol.

	BP86			**B3LYP**			**B3LYP-D**	**M05-2X**	
	SVP	TZVP	def2-TZVPP(d)	SVP	TZVP	def2-TZVPP(d)	TZVP	SVP	TZVP
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TS	12.71	10.77	11.19	17.02	15.06	15.49	12.41	19.69	16.91
P	7.83	8.46	7.98	9.40	10.21	9.83	7.19	8.44	9.06

Table A4.5. QM/MM potential energy barriers and reaction energies for the proposed S_Ni mechanism in the WT enzyme with 3-deoxylactose calculated at stationary points optimized at the QM(BP86/SVP)/CHARMM22 level. Energies are given in kcal/mol.

	BP86			**B3LYP**			**B3LYP-D**	**M05-2X**	
	SVP	TZVP	def2-TZVPP(d)	SVP	TZVP	def2-TZVPP(d)	TZVP	SVP	TZVP
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TS	13.77	11.20	12.13	17.51	14.97	15.82	12.84	21.15	17.71
P	2.46	2.60	2.51	3.45	3.81	3.78	0.28	2.58	2.46

A2. APPENDIX TO CHAPTER 5

Table A5.1. Selected QM/CHARMM22 bond distances *d* (Å) in the optimized reactants (R), transition state (TS), and products (P) for the double-displacement and S_Ni mechanisms. QM = SCC-DFTB.

		Double-displacement mechanism				**S_Ni mechanism**	
	R	**TS^d_1**	**CGE**	**TS^d_2**	**P**	**TS^i**	**P**
$d(\text{C1'}_{\alpha\text{-Gal}}\text{–O3B}_{\text{UDP}})$	1.49	2.52	3.57	3.40	3.38	2.87	3.31
$d(\text{O3}_{\beta\text{-Gal}}\text{–C1'}_{\alpha\text{-Gal}})$	3.09	2.99	3.15	2.33	1.49	2.60	1.48
$d(\text{HO3}_{\beta\text{-Gal}}\text{–O3}_{\beta\text{-Gal}})$	1.00	1.08	1.22	1.23	1.73	1.14	1.74
$d(\text{HO3}_{\beta\text{-Gal}}\text{–O3B}_{\text{UDP}})$	3.95	1.46	1.24	1.21	1.00	1.31	1.00
$d(\text{OE2}_{\text{E317}}\text{–C1'}_{\alpha\text{-Gal}})$	4.59	2.12	1.51	2.22	3.11	2.81	3.16
$d(\text{C1'}_{\alpha\text{-Gal}}\text{–O5'}_{\alpha\text{-Gal}})$	1.45	1.30	1.45	1.30	1.46	1.27	1.47

Table A5.2. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and S_Ni mechanisms at different levels of theory. The calculations were carried out on the optimized SCC-DFTB/CHARMM22 geometries.

Double-displacement mechanism					
	SCC-DFTB	**BP86//SCC-DFTB**	**B3LYP//SCC-DFTB**	**M05-2X//SCC-DFTB**	
		SVP	**SVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00
TS^d_1	31.19	17.61	19.60	20.33	20.35
CGE	20.83	20.23	21.46	17.36	18.04
TS^d_2	31.41	24.50	27.81	27.17	27.37
S_Ni mechanism					
TS^i	35.16	15.74	17.66	18.38	16.25
P	1.16	5.12	3.77	-0.19	2.90

Table A5.3. Selected QM/CHARMM22 bond distances *d* (Å) and atomic charges *q* (a.u.) in the optimized reactants (R), transition state (TS), and products (P) for the double-displacement and the S_Ni mechanisms. QM = BP86/SVP or M05-2X/TZVP//BP86/SVP for the distances and charges, respectively.

		Double-displacement mechanism				**S_Ni mechanism**	
	R	**TS^d_1**	**CGE**	**TS^d_2**	**P**	**TS^i**	**P**
$d(\mathrm{C1'}_{\alpha\text{-Gal}}–\mathrm{O3B}_{\mathrm{UDP}})$	1.51	2.89	3.63	3.43	3.41	3.04	3.43
$d(\mathrm{O3}_{\beta\text{-Gal}}–\mathrm{C1'}_{\alpha\text{-Gal}})$	3.08	2.94	3.16	2.41	1.47	2.72	1.47
$d(\mathrm{HO3}_{\beta\text{-Gal}}–\mathrm{O3}_{\beta\text{-Gal}})$	1.00	1.02	1.03	1.09	1.49	1.03	1.49
$d(\mathrm{HO3}_{\beta\text{-Gal}}–\mathrm{O3B}_{\mathrm{UDP}})$	3.95	1.59	1.56	1.40	1.04	1.52	1.04
$d(\mathrm{OE2}_{\mathrm{E317}}–\mathrm{C1'}_{\alpha\text{-Gal}})$	4.26	2.56	1.58	2.41	3.23	2.80	3.22
$d(\mathrm{C1'}_{\alpha\text{-Gal}}–\mathrm{O5'}_{\alpha\text{-Gal}})$	1.38	1.28	1.37	1.29	1.40	1.28	1.41
$q(\mathrm{C1'}_{\alpha\text{-Gal}})$	0.36	0.59	0.35	0.59	0.33	0.57	0.33
$q(\mathrm{O5'}_{\alpha\text{-Gal}})$	-0.51	-0.41	-0.53	-0.45	-0.53	-0.41	-0.54

Table A5.4. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and S_Ni-like mechanisms at different levels of theory in frame 1. Geometry optimizations were performed at the QM(BP86/SVP)/CHARMM22 level of calculation. ?: The TS was not supported by the frequency calculation.

	Double-displacement mechanism					
	BP86		**B3LYP//BP86/SVP**		**M05-2X//BP86/SVP**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
TS^d_1	10.42	7.18	11.28	8.09	16.27	11.80
CGE	9.06	10.49	9.25	9.56	9.46	8.85
TS^d_2	11.08	9.26	12.89	11.37	16.11	13.38
P	-1.12	3.1	-1.3	3.55	-5.43	-1.51
	S_Ni-like mechanism					
?TS^i_1	11.27	9.08	12.46	9.28	18.12	13.86
IP	10.27	7.57	11.13	7.96	16.00	11.54
?TS^i_2	11.51	9.89	12.91	11.11	16.47	13.47
P	0.93	3.91	0.80	4.47	-3.46	-0.51

Table A5.5. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and SNi-like mechanisms at different levels of theory in frame 2. Geometry optimizations were performed at the QM(BP86/SVP)/CHARMM22 level of calculation. ?: The IP intermediate was not supported by the frequency calculation.

	Double-displacement mechanism					
	BP86		**B3LYP//BP86/SVP**		**M05-2X//BP86/SVP**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
TS^{d}_{1}	10.22	8.93	11.18	8.18	16.62	12.34
CGE	6.22	6.61	6.72	7.31	7.20	6.80
TS^{d}_{2}	10.38	8.62	12.31	10.81	16.26	13.50
P	-1.69	2.02	-1.56	2.81	-5.11	-1.62
	S_Ni-like mechanism					
TS^{i}_{1}	12.68	9.01	13.74	10.34	19.73	15.32
?IP	9.51	8.57	10.84	8.32	15.92	12.06
TS^{i}_{2}	10.15	8.60	11.64	9.68	15.98	12.74
P	-1.50	2.14	-1.29	2.99	-4.87	-1.44

Table A5.6. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and S_Ni-like mechanisms at different levels of theory in frame 3. Geometry optimizations were performed at the QM(BP86/SVP)/CHARMM22 level of calculation. ?: The TS was not supported by the frequency calculation.

	Double-displacement mechanism					
	BP86		**B3LYP//BP86/SVP**		**M05-2X//BP86/SVP**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
TS^{d}_{1}	16.88	15.06	16.45	14.72	21.12	18.08
CGE	13.18	14.38	13.65	14.92	13.85	14.17
TS^{d}_{2}	14.22	13.35	15.80	14.35	19.48	16.91
P	4.11	9.12	5.06	10.66	1.06	5.76
	S_Ni-like mechanism					
?TS^{i}_{1}	14.81	14.49	15.97	13.79	20.56	17.18
IP	16.35	14.03	15.80	13.49	20.35	16.80
?TS^{i}_{2}	14.35	14.92	16.99	17.26	19.52	18.58
P	5.20	10.23	6.19	11.82	2.15	6.88

Table A5.7. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed double-displacement and S_Ni-like mechanisms at different levels of theory in frame 4. Geometry optimizations were performed at the QM(BP86/SVP)/CHARMM22 level of calculation.

	Double-displacement mechanism					
	BP86		**B3LYP//BP86/SVP**		**M05-2X//BP86/SVP**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
TS^{d}_{1}	16.78	14.57	16.35	14.35	21.13	18.10
CGE	13.94	15.2	14.74	16.15	14.95	15.59
TS^{d}_{2}	15.89	16.81	18.14	17.56	21.43	19.76
P	5.41	9.51	5.71	10.44	2.06	6.03
	S_Ni-like mechanism					
TS^{i}_{1}	16.47	13.84	16.41	14.09	21.67	18.42
IP	14.82	14.86	16.16	14.20	20.80	17.78
TS^{i}_{2}	16.49	15.16	18.44	17.65	22.08	20.19
P	5.31	9.42	5.55	10.29	1.82	5.80

Table A5.8. Relative energies ((QM + $QM/MM_{interactions}$), in kcal/mol) of all the stationary points of the double-displacement and the S_Ni-like mechanisms for frames 1-4 in α1,3-GalT. Values correspond to the relative energies when the charge of the water molecule 2283 set to zero or to the original force field value (within parentheses). QM=(M05-2X/TZVP//BP86/SVP).

	Frame 1	**Frame 2**	**Frame 3**	**Frame 4**
	Double-displacement mechanism			
R*	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
TS^{d}_{1}	16.02 (8.52)	13.12 (8.40)	22.22 (17.24)	21.38 (21.11)
CGE	9.90 (2.77)	4.23 (-1.44)	15.18 (5.97)	16.71 (16.60)
TS^{d}_{2}	15.90 (8.92)	13.08 (8.49)	19.64 (11.81)	20.49 (20.23)
P	-0.08 (-2.04)	-2.53 (-4.16)	4.5 (0.13)	4.49 (4.63)
	S_Ni-like mechanism			
TS^{i}_{1}	18.89 (12.74)	16.36 (13.29)	20.56 (13.96)	22.48 (22.24)
IP	15.60 (8.19)	11.99 (7.21)	18.72 (10.83)	20.97 (20.73)
TS^{i}_{2}	18.30 (10.47)	13.35 (7.93)	23.99 (15.90)	22.21 (21.87)
P	-2.81 (-4.9)	-2.37 (-4.09)	3.81 (-0.60)	3.26 (3.40)

Table A5.9. Selected QM/MM bond distances *d*(Å) and atomic charges *q*(a.u.) in the optimized reactants (R), transition states (TS), ion-pair intermediate (IP) and products (P) for the double-displacement and S$_N$i-like mechanisms in frame 1. QM=BP86/SVP and M05-2X/TZVP// BP86/SVP for the distances and charges respectively. ?: The TS was not supported by the frequency calculation.

		Double-displacement mechanism				**S$_N$i-like mechanism**			
	R	**TS^{d}_{1}**	**CGE**	**TS^{d}_{2}**	**P**	**?TS^{i}_{1}**	**IP**	**?TS^{i}_{2}**	**P**
d(O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$)	1.50	3.13	3.79	3.32	3.43	2.52	3.17	3.47	3.36
d(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$)	3.03	2.87	3.10	2.43	1.47	2.90	2.83	2.51	1.47
d(HO3$_{\beta\text{-Gal}}$–O3$_{\beta\text{-Gal}}$)	1.00	1.02	1.03	1.04	1.48	1.01	1.02	1.04	1.48
d(HO3$_{\beta\text{-Gal}}$–O3B$_{UDP}$)	4.00	1.56	1.55	1.47	1.04	1.60	1.55	1.47	1.04
d(OE2$_{E317}$–C1'$_{\alpha\text{-Gal}}$)	4.36	2.64	1.57	2.88	3.26	3.10	2.69	2.82	3.27
d(C1'$_{\alpha\text{-Gal}}$–O5'$_{\alpha\text{-Gal}}$)	1.38	1.28	1.36	1.28	1.41	1.28	1.28	1.28	1.40
q(C1'$_{\alpha\text{-Gal}}$)	0.36	0.58	0.35	0.57	0.33	0.57	0.57	0.57	0.33
q(O5'$_{\alpha\text{-Gal}}$)	-0,51	-0.40	-0.52	-0.42	-0.53	-0.41	-0.40	-0.41	-0.53

Table A5.10. Selected QM/MM bond distances d(Å) and atomic charges *q*(a.u.) in the optimized reactants (R), transition states (TS), ion-pair intermediate (IP) and products (P) for the double-displacement and S$_N$i-like mechanisms in frame 3. QM=BP86/SVP and M05-2X/TZVP// BP86/SVP for the distances and charges respectively. ?: The TS was not supported by the frequency calculation.

		Double-displacement mechanism				**S$_N$i-like mechanism**			
	R	**TS^{d}_{1}**	**CGE**	**TS^{d}_{2}**	**P**	**?TS^{i}_{1}**	**IP**	**?TS^{i}_{2}**	**P**
d(O3B$_{UDP}$–C1'$_{\alpha\text{-Gal}}$)	1.50	3.25	3.81	3.32	3.30	3.00	3.19	3.48	3.32
d(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$)	3.12	2.87	3.18	2.65	1.48	2.89	2.87	2.40	1.48
d(HO3$_{\beta\text{-Gal}}$–O3$_{\beta\text{-Gal}}$)	1.00	1.02	1.02	1.03	1.50	1.02	1.02	1.05	1.50
d(HO3$_{\beta\text{-Gal}}$–O3B$_{UDP}$)	4.08	1.62	1.62	1.57	1.04	1.62	1.63	1.48	1.04
d(OE2$_{E317}$–C1'$_{\alpha\text{-Gal}}$)	4.27	2.67	1.57	2.85	3.34	2.77	2.72	2.91	3.34
d(C1'$_{\alpha\text{-Gal}}$–O5'$_{\alpha\text{-Gal}}$)	1.38	1.28	1.36	1.28	1.41	1.28	1.28	1.29	1.41
q(C1'$_{\alpha\text{-Gal}}$)	0.36	0.58	0.35	0.57	0.33	0.57	0.57	0.57	0.33
q(O5'$_{\alpha\text{-Gal}}$)	-0,51	-0.41	-0.52	-0.40	-0.54	-0.40	-0.40	-0.43	-0.54

Table A5.11. Selected QM/MM bond distances d(Å) and atomic charges q(a.u.) in the optimized reactants (R), transition states (TS), ion-pair intermediate (IP) and products (P) for the double-displacement and S$_\mathrm{N}$i-like mechanisms in frame 4. QM=BP86/SVP and M05-2X/TZVP//BP86/SVP for the distances and charges respectively.

		Double-displacement mechanism				**S$_\mathrm{N}$i-like mechanism**			
	R	**TS$^\mathrm{d}_1$**	**CGE**	**TS$^\mathrm{d}_2$**	**P**	**TS$^\mathrm{i}_1$**	**IP**	**TS$^\mathrm{i}_2$**	**P**
d(O3B$_{\mathrm{UDP}}$–C1'$_{\alpha\text{-Gal}}$)	1.50	2.90	3.76	3.29	3.35	2.69	3.07	3.48	3.36
d(O3$_{\beta\text{-Gal}}$–C1'$_{\alpha\text{-Gal}}$)	3.07	2.89	3.11	2.47	1.48	2.87	2.83	2.53	1.47
d(HO3$_{\beta\text{-Gal}}$–O3 $_{\beta\text{-Gal}}$)	1.00	1.02	1.03	1.05	1.51	1.02	1.03	1.04	1.51
d(HO3$_{\beta\text{-Gal}}$–O3B$_{\mathrm{UDP}}$)	4.07	1.57	1.56	1.47	1.04	1.58	1.55	1.48	1.04
d(OE2$_{\mathrm{E317}}$–C1'$_{\alpha\text{-Gal}}$)	4.28	2.69	1.58	2.84	3.23	2.98	2.70	2.78	3.23
d(C1'$_{\alpha\text{-Gal}}$–O5'$_{\alpha\text{-Gal}}$)	1.38	1.28	1.36	1.28	1.41	1.28	1.28	1.28	1.41
q(C1'$_{\alpha\text{-Gal}}$)	0.36	0.58	0.35	0.57	0.33	0.57	0.58	0.58	0.33
q(O5'$_{\alpha\text{-Gal}}$)	-0,51	-0.41	-0.52	-0.42	-0.54	-0.41	-0.40	-0.41	-0.54

A3. APPENDIX TO CHAPTER 6

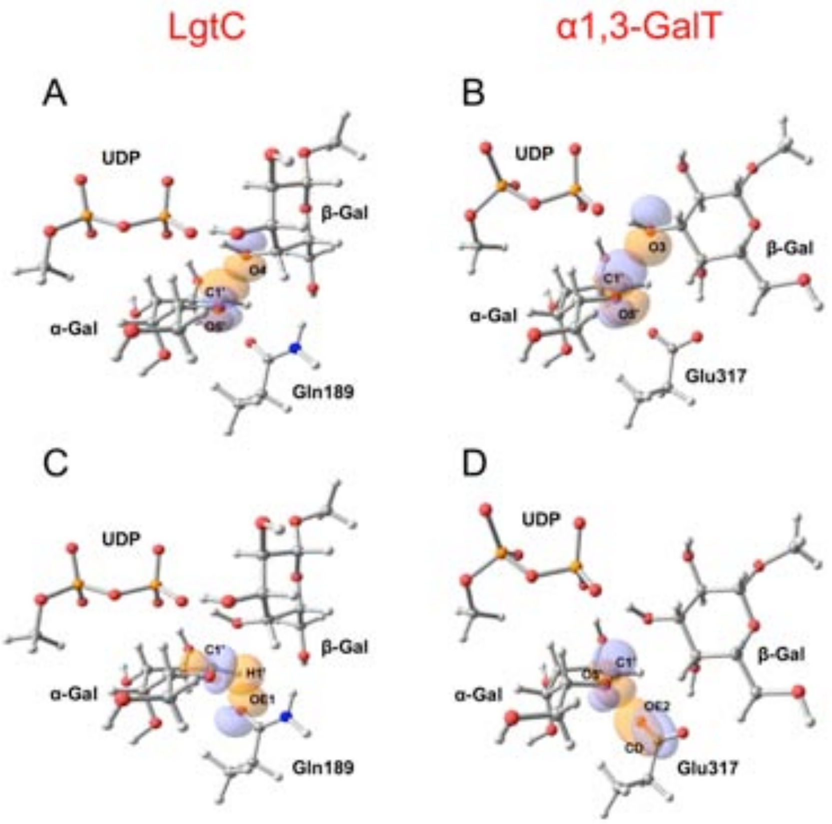

Figure A6.1. Some of the main molecular orbital interactions contributing to TS stabilization according to a NBO analysis in the TS and TS^{i}_{2} for the front-side attack mechanism in LgtC and α1,3-GalT, respectively. It complements Figure 6.5. For clarity, just a fraction of the QM atoms is shown.

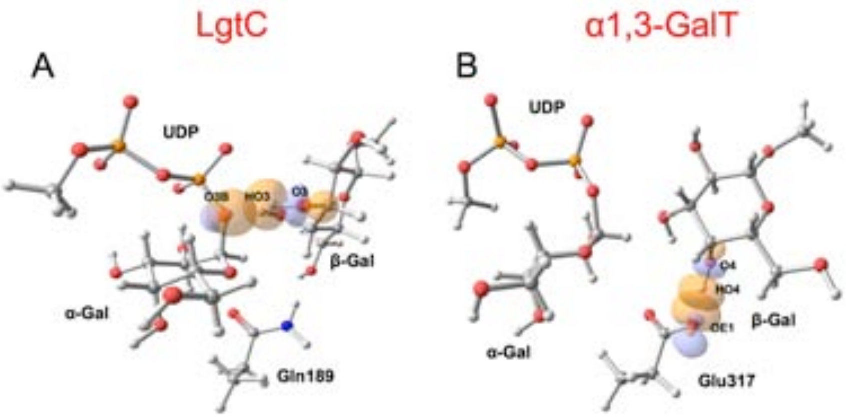

Figure A6.2. Some of the main molecular orbital interactions present in the reactants according to a NBO analysis in LgtC and α1,3-GalT, respectively. It complements Figure 6.5. For clarity, just a fraction of the QM atoms is shown.

A4. APPENDIX TO CHAPTER 7

Table A7.1. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed front-side attack mechanism at different levels of theory in frame 2. The calculations were carried out on the corresponding QM(BP86/SVP)/MM(CHARMM22) geometries of reactants (R), transition state guess (?TS^{i}) and products (P).

	BP86		**B3LYP**		**M05-2X**	
	SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00
?TS^{i}	15.67	10.73	19.28	13.89	26.80	20.20
P	1.49	-0.88	0.92	-1.57	2.79	-0.14

Table A7.2. Selected QM/MM bond distances *d* (Å), dihedral angle (Degrees) and atomic charges *q* (a.u.) in the optimized reactants (R), transition state guess (?TS^{i}), and products (P) for the front-side attack mechanism in frame 2. QM=BP86/SVP and M05-2X/TZVP//BP86/SVP for the geometrical parameters and charges respectively.

	Reactants	**?TS^{i}**	**Products**
$d(\mathrm{O3B}_{\mathrm{UDP}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$	1.51	3.13	3.36
$d(\mathrm{OG1}_{\mathrm{T7}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$	2.86	2.24	1.49
$d(\mathrm{HG1}_{\mathrm{T7}}\text{–}\mathrm{OG1}_{\mathrm{T7}})$	0.99	1.16	1.48
$d(\mathrm{HG1}_{\mathrm{T7}}\text{–}\mathrm{O3B}_{\mathrm{UDP}})$	1.79	1.23	1.02
$d(\mathrm{O}_{\mathrm{A307}}\text{–}\mathrm{C1'}_{\alpha\text{-GalNAc}})$	4.11	3.10	3.25
$d(\mathrm{C1'}_{\alpha\text{-GalNAc}}\text{–}\mathrm{O5'}_{\alpha\text{-GalNAc}})$	1.38	1.29	1.39
$d(\mathrm{HN2'}_{\alpha\text{-GalNAc}}\text{–}\mathrm{O1B}_{\mathrm{UDP}})$	2.32	1.90	2.04
$(\mathrm{H2'}\text{–}\mathrm{C2'}\text{–}\mathrm{N2'}\text{–}\mathrm{HN2'})_{\alpha\text{-GalNAc}}$	163.73	167.53	158.68
$q(\mathrm{C1'}_{\alpha\text{-GalNAc}})$	0.40	0.57	0.37
$q(\mathrm{O3B}_{\mathrm{UDP}})$	-0.91	-1.16	-1.07
$q(\mathrm{O5'}_{\alpha\text{-GalNAc}})$	-0.51	-0.43	-0.56
$q(\mathrm{O1B}_{\mathrm{UDP}})$	-1.18	-1.23	-1.20
$q(\mathrm{HN2'}_{\alpha\text{-GalNAc}})$	0.47	0.49	0.47

Table A7.3. QM/MM potential energy barriers and reaction energies (in kcal/mol) for the proposed front-side attack mechanism at different levels of theory in frame 1. The calculations were carried out on the corresponding QM(SCC-DFTB)/MM(CHARMM22) geometries of reactants (R), transition state (TS^{i}) and products (P).

	SCC-DFTB	**BP86**		**B3LYP**		**M05-2X**	
		SVP	**TZVP**	**SVP**	**TZVP**	**SVP**	**TZVP**
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TS^{i}	38.95	18.82	14.26	22.24	17.39	28.65	22.72
P	13.22	16.48	13.06	15.08	11.61	16.75	13.32

Table A7.4. Selected QM/MM bond distances *d* (Å), dihedral angle (Degrees) and atomic charges *q* (a.u.) in the optimized reactants (R), transition state (TS^{i}), and products (P) for the front-side attack mechanism in frame 2. QM=SCC-DFTB and M05-2X/TZVP//BP86/SVP for the distances and charges respectively.

	Reactants	**TS^{i}**	**Products**
$d(\mathrm{O3B_{UDP}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	1.48	2.59	3.20
$d(\mathrm{OG1_{T7}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	2.82	2.40	1.49
$d(\mathrm{HG1_{T7}}\text{–}\mathrm{OG1_{T7}})$	0.99	1.24	2.55
$d(\mathrm{HG1_{T7}}\text{–}\mathrm{O3B_{UDP}})$	1.85	1.17	0.98
$d(\mathrm{O_{A307}}\text{–}\mathrm{C1'_{\alpha\text{-}GalNAc}})$	4.14	3.14	3.24
$d(\mathrm{C1'_{\alpha\text{-}GalNAc}}\text{–}\mathrm{O5'_{\alpha\text{-}GalNAc}})$	1.45	1.28	1.45
$d(\mathrm{HN2'_{\alpha\text{-}GalNAc}}\text{–}\mathrm{O1B_{UDP}})$	2.03	1.75	1.85
$(\mathrm{H2'}\text{–}\mathrm{C2'}\text{–}\mathrm{N2'}\text{–}\mathrm{HN2'})_{\alpha\text{-}\mathrm{GalNAc}}$	167.19	164.19	152.28
$q(\mathrm{C1'_{\alpha\text{-}GalNAc}})$	0.37	0.59	0.37
$q(\mathrm{O3B_{UDP}})$	-0.87	-1.12	-0.99
$q(\mathrm{O5'_{\alpha\text{-}GalNAc}})$	-0.52	-0.40	-0.54
$q(\mathrm{O1B_{UDP}})$	-1.16	-1.20	-1.16
$q(\mathrm{HN2'_{\alpha\text{-}GalNAc}})$	0.48	0.49	0.46

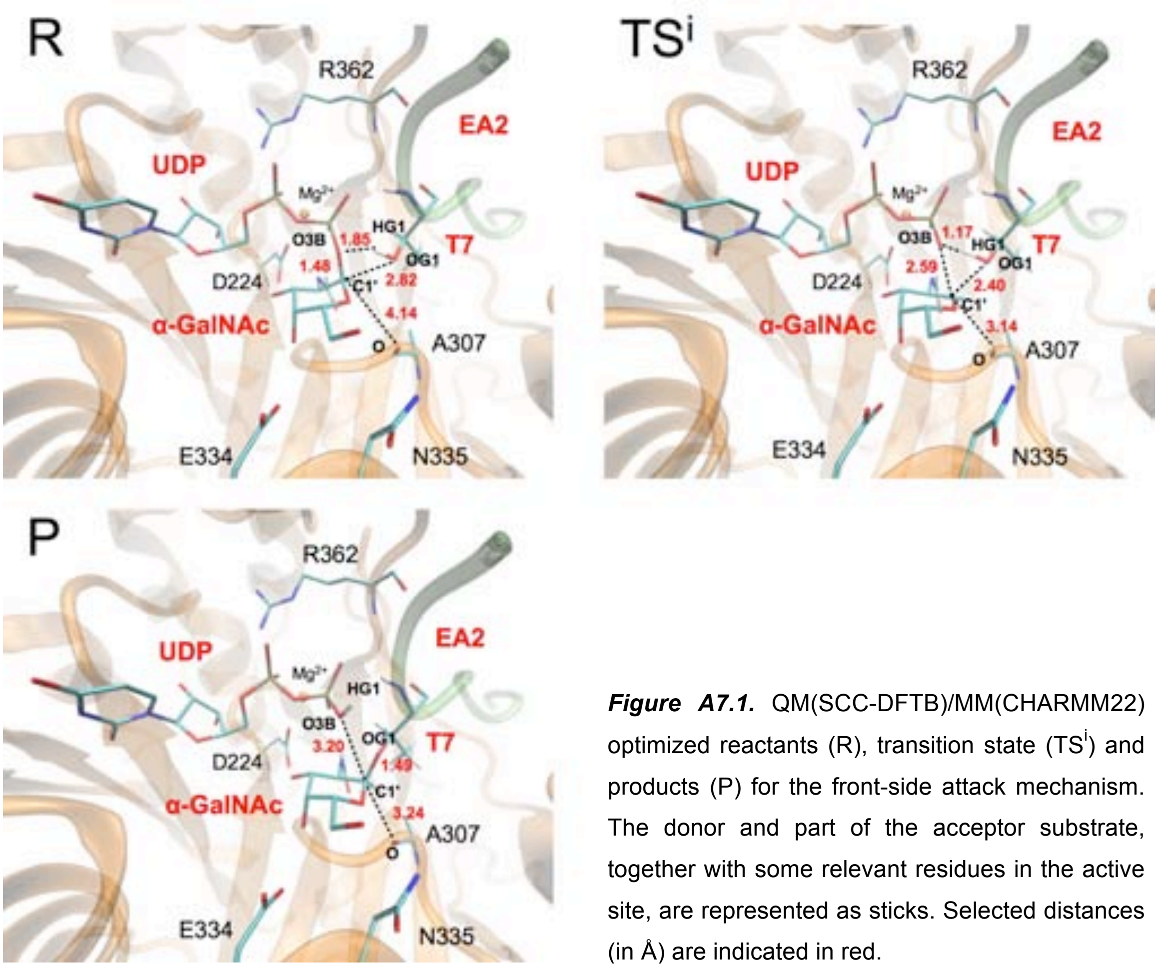

Figure A7.1. QM(SCC-DFTB)/MM(CHARMM22) optimized reactants (R), transition state (TS^{i}) and products (P) for the front-side attack mechanism. The donor and part of the acceptor substrate, together with some relevant residues in the active site, are represented as sticks. Selected distances (in Å) are indicated in red.

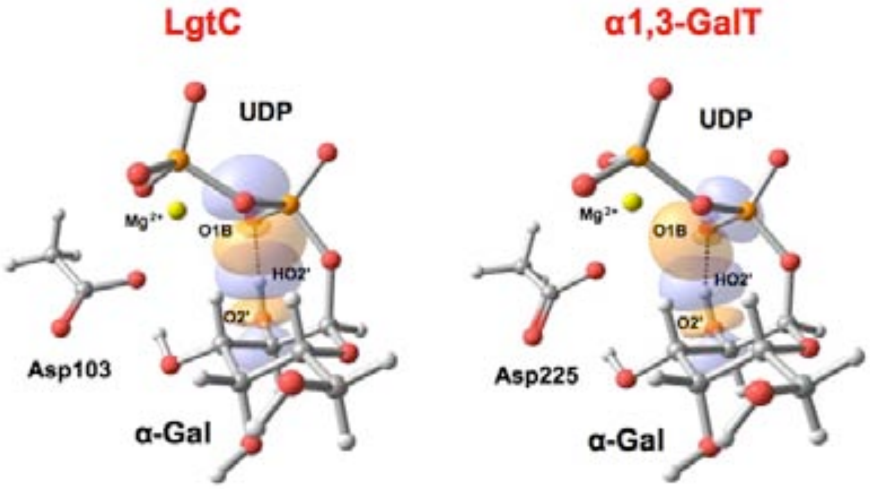

Figure A7.2. Interactions between molecular orbitals of the 2'-hydroxyl group of Gal and UDP in LgtC and α1,3-GalT according to a NBO analysis. The interactions are only depicted in the reactants. For clarity, just a fraction of the QM atoms is shown.