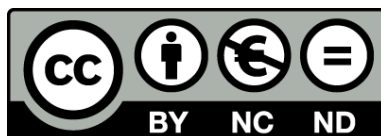




UNIVERSITAT^{DE}
BARCELONA

Simulations of glycoside hydrolase and phosphorylase reaction mechanism: families GH20, GH29, GH129 and GH130

Irene Cuxart Sánchez



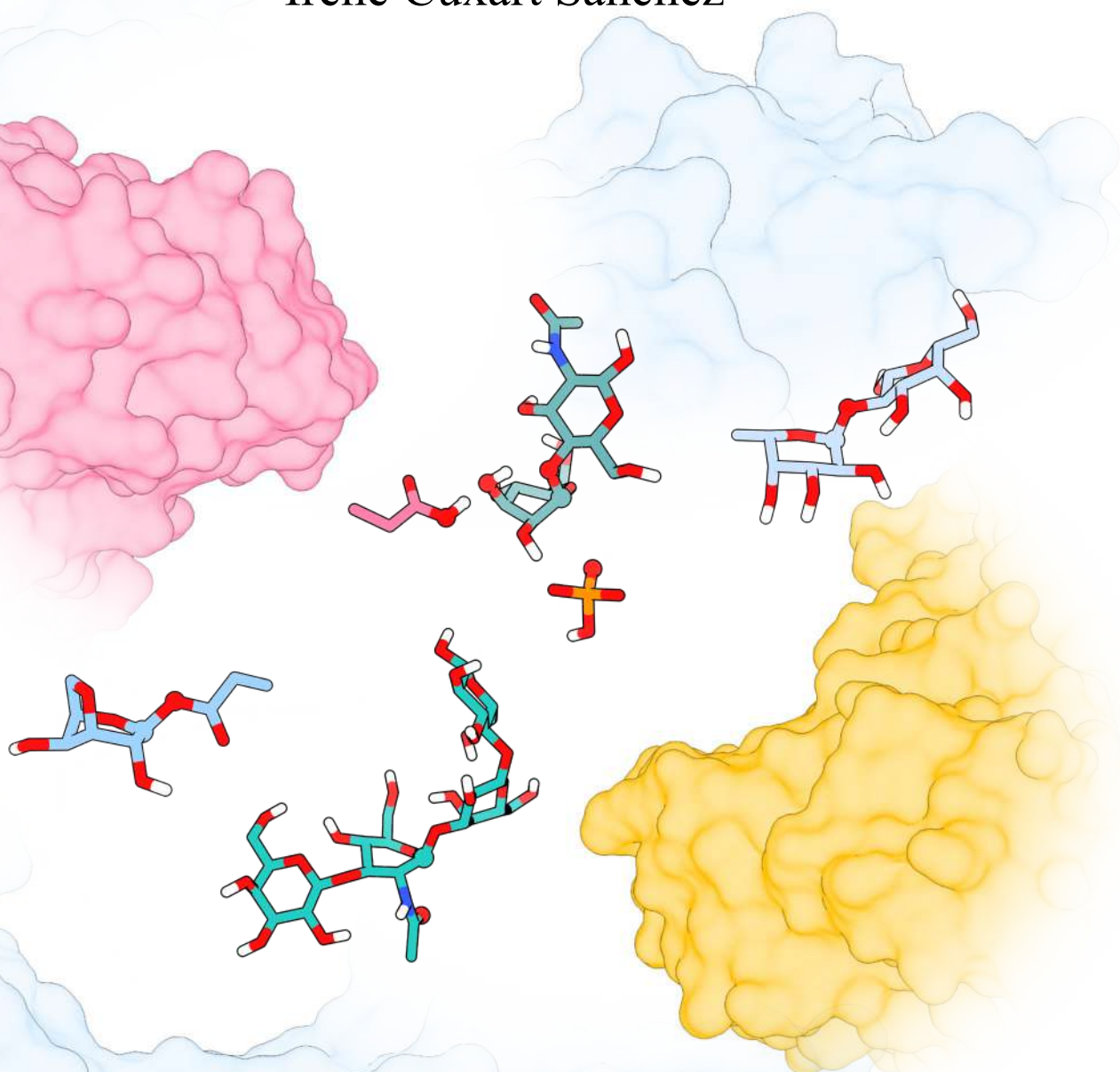
Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 4.0. Spain License.**

Simulations of glycoside hydrolase and phosphorylase reaction mechanisms: families GH20, GH29, GH129 and GH130

Irene Cuxart Sánchez



UNIVERSITAT DE
BARCELONA



UNIVERSITAT DE
BARCELONA

UNIVERSITAT DE BARCELONA

FACULTAT DE QUÍMICA

Departament de Química Inorgànica i Orgànica

Secció de Química Orgànica

Programa de Doctorat en Química Orgànica

Línia de Recerca en Química Teòrica i Computacional

Simulations of glycoside hydrolase and phosphorylase reaction mechanisms: families GH20, GH29, GH129 and GH130

Memòria de Recerca presentada per optar al títol de Doctora per la Universitat de
Barcelona, Setembre 2024

Nom, cognoms i signatura de la doctoranda:

Irene Cuxart Sánchez

Nom, cognoms i signatura de la directora i tutora:

Dra. Carme Rovira Virgili

Departament de Química Inorgànica i Orgànica, Secció de Química Orgànica i

Institut de Química Teòrica i Computacional, Universitat de Barcelona

*Amb els llibres per amics
no et faltarà companyia.
Cada pàgina pot ser
un estel que et fa de guia.
— Joana Raspall*

Acknowledgments

There are so many people I am deeply grateful for the work of all these years leading to the writing of the present dissertation, and I will try to go step by step to thank all. Not only the work of all these years has done with the encouragement of many people, but also your company and support has been fundamental in all life moments.

Primer de tot vull donar les gràcies a la meva directora, la Dra. Carme Rovira. Gràcies, Carme, per oferir-me l'oportunitat de quedar-me al grup per fer el doctorat, ara ja fa més de quatre anys. Va ser un salt que em vas animar molt a fer i per això estic molt agraïda. També vull donar gràcies per tota la formació que he rebut, com a investigadora, científica i persona. Per l'oportunitat de fer-me conèixer una branca de la química que uneix el món biològic a l'escala molecular. Gràcies també pels encoratjaments a l'hora de fer xerrades, els ànims per ser més segura, les crítiques i els consells. Dels anys aquí m'enduc grans aprenentatges que espero poder aplicar en tantes altres etapes de la vida.

Vull donar les gràcies als meus companys de viatge Bea i Oriol. El vostre suport ferm m'ha ajudat a empènyer, les converses, els riures, els esmorzars de cafè amb gel i croissant de quilo, i sopars de sopa de formatge. Els viatges que hem fet a Amsterdam, Irlanda i París, mai oblidarem el fred que vam passar a Hill of Tara. Quina millor manera de començar un matí amb -*Com estan les machines?* I acabar amb -*Que sigueu feliços*, i últimament -*Seguim*, [sic]. El club de lectura ha estat un pilar especialment mentre escriviem. Bea, no hauria pogut imaginar un millor final de Tesi, amb coordinació vestimental inclosa. Ha estat una cursa de resistència i m'has fet costat en moments àrids del final. Ha estat un privilegi compartir aquest temps amb vosaltres.

To the members of the SQPBio Research Group, all present and past, thank you. Alba, Qinghua, José Pablo, Alessandro, Òscar, Mert, Marina, Samanta, Laura, you are exceptional colleagues and also a beautiful group of people. I have learned so much from you, and also had tons of fun with the lunches together, discussions and outings. Thank you to the mentors of my early times in the group: Joan, Myriam, Alba especially. Thank you also to other people whose presence in the group has left

a profound mark: Mariana, Martín, Mireia. Thank you to the all visitors, former Postdocs Bachelor and MSc students that have been with us these years.

To the scientists which whom I have worked, thank you as well. Thank you Mertxe for the fantastic opportunity of working with you, your patience, care and detail you put in answering my questions. Thank you Elisa for hosting me in your group in Maynooth. The weekly group meetings were incredibly enlightening and I will keep the memories of those three months in Ireland dearly. Thank you for your insight in our work together: I am always happy to read your feedback and suggestions. Thank you also to the ELab members in Maynooth: Callum, Carl, Ojas, Akash, Silvia, Bea and Marium. I also would like to extend these acknowledgments to the rest of researchers and staff of the Hamilton Institute, who made my visit there memorable.

I also want to thank my collaborators for the different projects. To the Biochemistry group of IQS: Antoni Planas, Magda Faijes and Xevi Biarnés. From the days I was an undergrad in your lectures of the Chemistry degree to have been worked with you during this Ph.D. It is a pleasure for me to come back to IQS for all the beautiful seminars you organize. Thank you Xevi, because he suggested me to contact Carme in the first place. I would also like to thank the Roscoff and Australia teams, especially Elizabeth Ficko-Blean and Keith Stubbs, but also all the other members with whom we have worked hard especially this last year. It has been great to learn from you all and give shape to an interdisciplinary project together. Thank you for introducing me to marine glycobiology.

I would also like to extend my thanks to my Ph.D. Committee of the Organic Chemistry Doctorate Program: Profs. Pere Romea, Jordi Poater and Joaquim Crusats. Thank you for the encouragement during the yearly meetings, the feedback and the suggestions. I have always enjoyed talking with you all, especially this last year. I would also like to thank our institute, the IQTC-UB, and the other staff of the department and Facultat de Química that made this possible. These years I have really enjoyed coming to work to the Chemistry department.

I would like to warmly thank all other people that hosted me to work both in Leiden and Dresden. Thanks to Franco Buda, Agnieszka Kuc and Thomas Heine for welcoming me. I would like to thank as well the other friends I have made there: Rico, Bas, Titus, Irene, Joan, Maria and Silvia. Also I'd like to thank Augusto, Reza, Ali, Shubha, Puja, Preeti, Rico. Also thank you to the friends made in the different conferences attended during these years: Maria, Antonio, Andrea, and the rest of the team in Erice MolSim 2022 and CECAM QM/MM School 2022.

I also want to thank all my professors and mentors. My years at the Institut Químic de Sarrià are unforgettable. I am very grateful care and patience they put in our studies. They are a crucial group of people I think of as I go step by step in life. Also, to my teachers at Els Pins and IES Joan Miró, thank you. Thank you especially to Josep S, Ana T F, Lali M, Conchita M and Maria del Prado D. Your love for teaching was so contagious it was actually difficult to choose a University degree. Those were wonderful learning years.

Vull agrair aquesta tesi a les meves estimades Lluïcia i Helena. El vostre suport ha estat capdal, i sempre m'he sentit encoratjada al vostre costat. Els dies estressants esdevenen més lleugers amb les vostres bromes i berenars, vídeos d'ossos rentadors i tacos al pastor. Gràcies també Isaac, Anna, i a la resta del Cor Bruckner Barcelona per l'acompanyament musical i d'amistat.

També vull dedicar aquesta tesi als meus amics de la universitat: Anna D, Meritxell, Lúdia, Marina, Alícia, Marc M, Marc F, Juan, Laura C, Anna A, Matías, Adrià. Gràcies Meritxell per convidar-me a fer una xerrada a la UdG. Gràcies també a la Mireia i el Bernat. Gracias Marta, Raúl y Cristina.

Este trabajo también está dedicado a mis abuelos, Pepe y Rafaela. Vuestro cuidado y carácter me acompaña, y de vosotros he aprendido tanto. Gracias también a mis tíos y primos. Paco, Loli, Jose, Loli, Rafi y Joan, Carlos y Ari, Ana, Albert y Anna, Marc y Ari, Elena, Aleix. Gràcies padrins, Alfons i Anna, Teresa, Martí, Eulàlia, Clara, Iris i Rita. Els vostres ànims i comprensió han estat una gran companyia especialment els últims mesos d'enclaustrament. Grazie alla famiglia in Toscana: Marialuisa, Ennio, Martina, Ilaria, Ivana, Ernesto e tutti gli altri. Da voi mi sento a casa e sono sempre felice di tornare all'altra riva del mare per rivedervi.

Gràcies als meus pares. La vostra il·lusió per aprendre no us deixa mai, i en conseqüència alguna cosa d'això espero que se m'hagi quedat a mi també. Aquests anys he pogut compartir amb vosaltres força moments d'aquest recorregut. Em quedo amb un comentari d'un dia estressant, al juliol passat: *estàs la mar de bé!*. Que pugui aprendre una mica de la vostra calma, empeny i generositat, tot compartint unes cloïsses banyades en salsa espinaler.

Finalment, gràcies al Dario Calvani. Conoscerti è stata la sorpresa più inaspettata di questo viaggio dottorale. Da te imparo ogni giorno passione, costanza, entusiasmo, concisione e generosità. Il tuo senso dell'umorismo è un faro. Come nella mezza maratona di questa primavera, un passo alla volta, come i provinciali dell'Orsa Minore alla conquista degli spazi interstellari. One by one. Accomplir aquesta fita al teu costat m'omple d'alegria, i he sentit els teus ànims i alegria en

tot moment, amb menció especial al final de tot. Gràcies pel teu suport constant i absolutament entusiasta. *E quannu t'ancontru 'nda strata.*

List of publications

This thesis has resulted in the following publications:

- Cuxart, I., Coines, J., Esquivias, O., Faijes, M., Planas, A., Biarnés, X., Rovira, C. **Enzymatic hydrolysis of human milk oligosaccharides. The catalytic molecular mechanism of *Bifidobacterium bifidum* lacto-*N*-biosidase**, *ACS Catalysis* **2022**, 12, 8, 4737-4743
- Alfonso-Prieto, M.[#], Cuxart, I.[#], Potocki-Véronèse, G., André, I., Rovira, C. **Substrate-Assisted Mechanism for the Degradation of *N*-Glycans by a Gut Bacterial Mannoside Phosphorylase**, *ACS Catalysis* **2023**, 13, 4283-4289
- Wallace, M. D.[#], Cuxart, I.[#], Roret, T., Guée, L., Debowski, A. W., Czjzek, M., Rovira, C., Stubbs, K. A., Ficko-Blean, E. **Constrained catalytic itinerary of a retaining 3,6-anhydro-D-galactosidase, a key enzyme in red algal cell wall degradation**, *Angewandte Chemie International Edition*, **2024**, e202411171
- A manuscript including the results of the last chapter (human GH29 fucosidase, FucA1) is in preparation

[#] Equal contribution

Publications not included in this thesis:

- Coines, J., Cuxart, I., Teze, D., Rovira, C. **Computer Simulation to Rationalize "Rational" Engineering of Glycoside Hydrolases and Glycosyltransferases**, *The Journal of Physical Chemistry B* **2022**, 126, 4, 802-812

List of collaborations

The chapters of this thesis have been conducted in collaboration with the following research groups:

- **Prof. Antoni Planas (group leader), Prof. Magda Faijes and Prof. Xevi Biarnés**, Laboratory of Biochemistry, Grup de Química Biològica i Biotecnologia (GQBB), Institut Químic de Sarrià, Via Augusta 390, 08017 Barcelona, Spain. (Chapter 3)
- **Prof. Mercedes Alfonso-Prieto**, Institute of Neurosciences and Medicine, Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Wilhelm-Johnen-Straße, Gebäude 16.15, 52425 Jülich, Germany. (Chapter 4)
- **Prof. Isabelle André (PixCELLS Enzymes for Cell Factories and Plastics team leader) and Prof. Gabrielle Potocki-Véronèse (Catalysis and Enzyme Molecular Engineering team leader)**, Toulouse Biotechnology Institute, Bio and Chemical Engineering, 135 Avenue de Rangueil, 31077 Toulouse, France. (Chapter 4)
- **Prof. Elizabeth Ficko-Blean (group leader)**, Laboratory of Integrative Biology of Marine Models, Station Biologique de Roscoff, CNRS / Sorbonne Université, Place Georges Teissier, 29680 Roscoff, France. (Chapter 5)
- **Prof. Keith A. Stubbs (group leader)**, School of Molecular Sciences, The University of Western Australia (M310), 35 Stirling Highway, 6009 Perth, Australia. (Chapter 5)

Finally, the research included in Chapter 6, corresponding to GH29 FucA1, of this thesis was done in collaboration with:

- **Prof. Elisa Fadda (group leader)**, School of Biological Sciences, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom (present address).

As part of a three-month research visit at her group in her previous address at the Department of Chemistry Hamilton Institute, Maynooth University, Maynooth, County Kildare, Ireland.

List of abbreviations

Abbreviation	Description
ADG	Anhydro-D-galactose
<i>Bb</i> GH129	<i>Bifidobacterium bifidum</i> GH129 <i>N</i> -acetylgalactosaminidase
<i>Bf</i> MGP	<i>Bacteroides fragilis</i> 4- <i>O</i> - β -D-mannosyl-D-glucose phosphorylase
BOMD	Born-Oppenheimer molecular dynamics
CarPUL	Carrageenan polysaccharide utilization locus/loci
CAZY	Carbohydrate active enzyme
CAZYme	Carbohydrate active enzyme
CPMD	Car-Parrinello molecular dynamics
Cryo-EM	Cryogenic electron microscopy
CV	Collective variable
DFJ	Deoxyfuconojirimycin
DFT	Density functional theory
DP4	neo- β / κ -carrageenan tetrasaccharide
FEL	Free energy landscape
Fuc	Fucose
FucA1	<i>Homo sapiens</i> fucosidase 1
Gal	Galactose
GalNAc	<i>N</i> -acetylgalactosamine

GEI	Glycosyl-enzyme intermediate
Glc	Glucose
GlcNAc	<i>N</i> -acetylglucosamine
GH	Glycoside hydrolase, glycosidase
GP	Glycoside phosphorylase
GT	Glycoside transferase
<i>H</i>	Half-chair
HMO	Human milk oligosaccharide
IBD	Inflammable bowel disease
Int	Intermediate
KS	Kohn-Sham
Lac	Lactose
LNB	Lacto- <i>N</i> -biose
LnbB	<i>Bifidobacterium bifidum</i> lacto- <i>N</i> -biosidase
LNT	Lacto- <i>N</i> -tetraose
Man	Mannose
MC	Michaelis complex
MD	Molecular dynamics
MEP	Minimum energy pathway
MMPBSA	Molecular mechanics Poisson-Boltzmann surface area
NAD	Nicotinamide adenine dinucleotide
NeuAc	<i>N</i> -acetylneuraminic acid
P	Product
PBE	Perdew, Burke and Ernzerhof
PDB	Protein data bank

Pi	Inorganic phosphate group
<i>p</i> NP	<i>para</i> -nitrophenyl
POFUT1	Protein <i>O</i> -fucosyltransferase 1
PUL	Polysaccharide utilization locus/loci
QM/MM	Quantum mechanics / molecular mechanics
R	Reactant
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
S	Skew-boat
SAXS	Small-angle X-ray scattering
SNFG	Symbol nomenclature for glycans
TS	Transition state
UDP	Uridine diphosphate
UhgbMP	Unknown human gut bacterium mannoside phosphorylase
WT	Wild type
<i>Zg</i> GH129	<i>Zobellia galactanivorans</i> 3,6-anhydro-D-galactosidase

Summary

The most abundant family of biomolecules on earth are the carbohydrates. Carbohydrates, also known as glycans, sugars or saccharides, are based on a polyhydroxy-carbonyl backbone with a vast array of chemical and structural modifications, which confer them different physico-chemical properties. They participate in many biological processes related to structural, energy storage and molecular signaling functions, and the many possibilities regarding structure and function of carbohydrates open questions in the attractive field of glycobiology.

The simplest form of carbohydrates, the monosaccharides, can assemble via the glycosidic bond to form larger structures (oligo- or polysaccharides) or glycoconjugates in combination with other families of biomolecules (for example, glycoproteins). Carbohydrates are very flexible: from the possible atomic positions of the atoms that form a monosaccharide ring, to the rotation of the linkages between sugar units, it is a property that confers them the possibility to arrange their atoms in different positions resulting in different conformations, which is a crucial aspect in processes where they take part in.

Carbohydrates undergo chemical changes in the form of degradation, synthesis and modifications in the several biological processes where they participate. Operating these reactions we find the Carbohydrate Active Enzymes (also known as CAZymes), whose functions are increasingly being discovered and characterized. The study of CAZymes has attracted research attention, as it opens windows in how complex metabolic pathways protagonized by carbohydrates work and paves new paths for medical, biotechnological and environmental applications. In this dissertation we uncover the catalytic mechanisms of four enzymes involved in the breakdown of saccharide chains: three via hydrolysis of glycosidic bonds, known as glycoside hydrolases (GHs), and one via sugar phosphorylation (glycoside phosphorylases or GPs), respectively. With computer simulations based in all-atoms models, combining QM/MM approaches with enhanced sampling methods, we simulate the chemical transformations that diverse saccharides undergo in the active site of the enzymes, including their conformational itineraries.

The study of enzyme catalysis is an interdisciplinary field, where different techniques from structural biology to biochemistry give insights on their structure and function. Computational techniques such as molecular dynamics give us the opportunity to characterize at the atomic and molecular level states that would be otherwise difficult to capture with experimental methods, such as the short-lived transition states or the sometimes elusive native enzyme-substrate complexes.

Following the experience of our group in elucidating the catalytic mechanisms of CAZYmes, in this Thesis we aim to uncover the catalytic mechanisms of enzymes relevant for medical and biotechnological applications that have some particularity that sets them apart from other well-established GH mechanisms. This Thesis dissertation contains the following chapters:

- **Chapter 1:** Introduction.

Contextualization of the general topics covered in this dissertation, description of the scope of this Thesis and listing of the specific objectives.

- **Chapter 2:** Methods.

Section outlining the main theoretical basis of the computational techniques applied and a description of the general workflow used in this work.

- **Chapter 3:** The catalytic mechanism of GH20 lacto-*N*-biosidase.

We uncover the catalytic mechanism of the enzyme LnbB, involved in the degradation of human milk oligosaccharides, which uses the substrate for the nucleophilic role of catalysis.

- **Chapter 4:** The catalytic mechanism of GH130 *N*-glycan mannoside phosphorylase.

We elucidate the catalytic mechanism of the enzyme UhgbMP, which incorporates phosphate to mannosyl substrates via a mechanism involving the substrate in a double proton transfer.

- **Chapter 5:** The catalytic mechanism of GH129 3,6-anhydro-D-galactosidase.

We characterize the catalytic mechanism of bacterial *Zg*GH129, which degrades the unique algal 3,6-anhydro-D-galactosidase-based carrageenan oligosaccharides via an unusual conformational itinerary.

- **Chapter 6:** Substrate recognition and catalytic mechanism of human GH29 fucosidase.

We elucidate the first reaction step of the reaction mechanism of the enzyme FucA1, recognizes and degrades fucosylated glycans, and unveil structural and dynamic details of its recognition of fucosylated motifs.

- **Chapter 7:** Conclusions.

General conclusions extracted from this work.

Contents

Acknowledgments	iii
List of publications	vii
List of collaborations	ix
List of abbreviations	xi
Summary	xv
1 Introduction	1
1.1 Carbohydrates	1
1.2 Carbohydrate ring conformations	3
1.3 Glycoside hydrolases	6
1.3.1 Glycoside phosphorylases	8
1.3.2 Reaction mechanisms	9
1.3.3 Non-classical mechanisms: substrate-assisted catalysis	11
1.3.4 Conformational itineraries in GH catalysis	14
1.4 Scope of this Thesis	16
Objectives	18
2 Methods	19
2.1 Introduction to molecular dynamics	20
2.2 Energy calculation: Forcefield methods	21
2.3 Quantum mechanics methods	23
2.3.1 Energy calculation: DFT	24
2.3.2 Quantum molecular dynamics	26
2.4 QM/MM molecular dynamics	28
2.5 The timescale challenge	32
2.5.1 Enhanced sampling with metadynamics	32
2.6 General workflow in simulations of enzyme catalysis	34

3	The catalytic mechanism of GH20 lacto-<i>N</i>-biosidase	37
3.1	Introduction	38
3.2	Results and discussion	44
3.2.1	Modeling the Michaelis complex of LnbB with lacto- <i>N</i> -biose	44
3.2.2	First reaction step	45
3.2.3	Modeling the reaction intermediate	52
3.2.4	Second reaction step	53
3.3	Summary and conclusions	56
3.4	Computational details	58
3.4.1	Model of the Michaelis complex and first reaction step	58
3.4.1.1	System preparation	58
3.4.1.2	Classical molecular dynamics	58
3.4.1.3	QM/MM molecular dynamics	59
3.4.1.4	QM/MM metadynamics	60
3.4.2	Model of the reaction intermediate and second reaction step	61
3.4.2.1	System preparation	61
3.4.2.2	Classical molecular dynamics	61
3.4.2.3	QM/MM molecular dynamics	62
3.4.2.4	QM/MM metadynamics	62
3.5	Supplementary information	63
4	The catalytic mechanism of GH130	
	<i>N</i>-glycan mannoside phosphorylase	71
4.1	Introduction	72
4.1.1	Project history	78
4.2	Results and discussion	79
4.2.1	Modeling the Michaelis complex of UhgbMP with Man- β -1,4-GlcNAc	79
4.2.1.1	The oligomerization order of UhgbMP	80
4.2.1.2	The substrate recognition by UhgbMP	83
4.2.1.3	Distal acid catalyst: a non-canonical active site for a non-canonical mechanism	84
4.2.2	The phosphorylase reaction	85
4.2.3	Study of protonation states	89
4.3	Summary and conclusions	91
4.4	Computational details	93
4.4.1	System preparation	93

4.4.2	Classical molecular dynamics of hexameric UhgbMP-Man- β -1,4-GlcNAc	94
4.4.3	Oligomerization state of UhgbMP	94
4.4.4	Substrate promiscuity of UhgbMP	95
4.4.5	QM/MM molecular dynamics	95
4.4.6	QM/MM metadynamics	96
4.4.7	Additional QM/MM simulations	97
4.4.8	Additional QM/MM metadynamics simulations	97
4.4.9	Protonation state of Asp104	98
4.4.9.1	Molecular dynamics simulations of different scenarios	98
4.4.9.2	Computed estimations of pK_a values	98
4.5	Supplementary information	99
5	The catalytic mechanism of GH129	
	3,6-anhydro-D-galactosidase	117
5.1	Introduction	118
5.1.1	An interdisciplinary study. Experimental results	120
5.2	Results and discussion	121
5.2.1	Modeling the Michaelis complex of <i>Zg</i> GH129 with a neo- β/κ -carrageenan oligosaccharide	121
5.2.1.1	The conformational preferences of isolated 3,6-ADG .	125
5.2.1.2	The conformational preferences of 3,6-ADG in the active site of <i>Zg</i> GH129	127
5.2.2	The glycosylation reaction	128
5.2.3	Modeling the glycosyl intermediate of <i>Zg</i> GH129 with 3,6-ADG	133
5.2.4	The deglycosylation reaction	134
5.3	Summary and conclusions	137
5.4	Computational details	139
5.4.1	Model of the <i>Zg</i> GH129-substrate complex and glycosylation reaction	139
5.4.1.1	System preparation	139
5.4.1.2	Classical Molecular Dynamics	140
5.4.1.3	QM metadynamics of isolated 3,6-anhydro-D-galactose	140
5.4.1.4	QM/MM metadynamics of 3,6-anhydro-D-galactose in <i>Zg</i> GH129 active site	141
5.4.1.5	QM/MM molecular dynamics of the <i>Zg</i> GH129-substrate complex	142
5.4.1.6	QM/MM metadynamics of the glycosylation step . .	143

5.4.2	Model of the <i>Zg</i> GH129 glycosyl-enzyme intermediate and deglycosylation reaction	143
5.4.2.1	System preparation	143
5.4.2.2	Classical Molecular Dynamics	144
5.4.2.3	QM/MM molecular dynamics	144
5.4.2.4	QM/MM metadynamics of the deglycosylation step .	145
5.4.2.5	QM/MM metadynamics of the <i>Zg</i> GH129-ADG product complex	145
5.5	Supplementary information	147
6	Substrate recognition and catalytic mechanism of human GH29 fucosidase	163
6.1	Introduction	164
6.2	Results and discussion	168
6.2.1	Unliganded FucA1	168
6.2.2	FucA1-glycan complexes	171
6.2.2.1	FucA1 with Fuc- α -1,2-Gal (complex 1)	172
6.2.2.2	Fuc- α -1,2-Gal trisaccharides: FucA1 with substrates 2, 3, 4 and 5	175
6.2.2.3	Other linkage types: α -1,3-GlcNAc, α -1,4-GlcNAc and α -1,6-GlcNAc (complexes 6, 7 and 8)	179
6.2.3	The glycosylation reaction of FucA1 with Fuc- α -1,2-Gal	182
6.3	Summary and conclusions	185
6.4	Computational details	187
6.4.1	Model of the unliganded FucA1	187
6.4.1.1	System preparation	187
6.4.1.2	Classical Molecular Dynamics	187
6.4.2	Building the FucA1-glycan complexes	188
6.4.2.1	System preparation	188
6.4.2.2	Classical Molecular Dynamics	190
6.4.3	The glycosylation reaction of FucA1 with Fuc- α -1,2-Gal	191
6.4.3.1	QM/MM molecular dynamics simulations of FucA1-Fuc- α -1,2-Gal complex	191
6.4.3.2	QM/MM metadynamics simulations of the glycosylation step	191
6.5	Supplementary information	193
7	Conclusions	207

Bibliography	209
---------------------	------------

Chapter 1

Introduction

1.1 Carbohydrates

Carbohydrates are the most abundant family of biomolecules. From the trees cellulose trunk to the sea-life chitin exoskeletons, to the starch in a seed, or many of the foods we eat, to the glycogen that gives us energy to walk to the office and work in our research, carbohydrates participate in many biological functions. Although building structures and storing energy are their most widely known functions, they are also involved in communication between cells, even between organisms, and so they participate in many biological processes [1].

Carbohydrates, or glycans, exist in their free forms or linked to other families of biomolecules forming glycoconjugates such as glycoproteins or glycolipids. The chemical definition of a carbohydrate is a polyhydroxycarbonyl, $(\text{CH}_2\text{O})_n$, or a larger compound that can be hydrolyzed into such [1,2]. The position of the carbonyl group in carbohydrates can be terminal (aldehyde, named aldoses) or in the inner carbons of the chain (ketone, ketoses). They have vast structural possibilities thanks to variations in the nature of their substituent groups, regio- and stereochemistry, and the diversity of linkages between them. Monosaccharides are the simplest building block of carbohydrates, namely, those that cannot be hydrolyzed any further, and are usually found in their ring forms, forming 5- or 6-member rings (furanoses and pyranoses, respectively) [1].

The simplest variation among monosaccharides is the arrangement of the hydroxyl group of the chiral center farthest from the most oxidized carbon (C5 in hexoses). The direction of this -OH group in the Fischer projection of the monosaccharide, right (D-) or left (L-), distinguishes enantiomers of the same sugar. Most natural monosaccharides are D- isomers, although some L-sugars are major players in biology like L-fucose [3]. Two monosaccharides that share stereochemistry of all but one chiral center are known as epimers (for example, Glc and Gal at C4, or Glc and

Man at C2) (Figure 1.1.a, top). Further variation is introduced in the cyclic forms with the creation of a new chiral center, the anomeric carbon, which corresponds to the carbonyl carbon of the linear saccharide after forming the cyclic hemiacetal: the α and β forms, depending on the position of the anomeric substituent respect to the $-\text{CH}_2\text{OH}$ substituent of the ring (Figure 1.1.a, bottom). Some of the most common hexoses in Nature will be a matter of study of this Thesis, such as D-galactose (D-Gal), *N*-acetyl D-glucosamine (GlcNAc), D-mannose (D-Man) and L-fucose (L-Fuc), which are important building blocks for many larger glycans (Figure 1.1.b). The rich stereochemistry of carbohydrates is complemented with other chemical modifications, such methylations, esterifications, deoxygenations, *N*-acetylations, sulfations and formation of anhydro-bridges [2, 4].

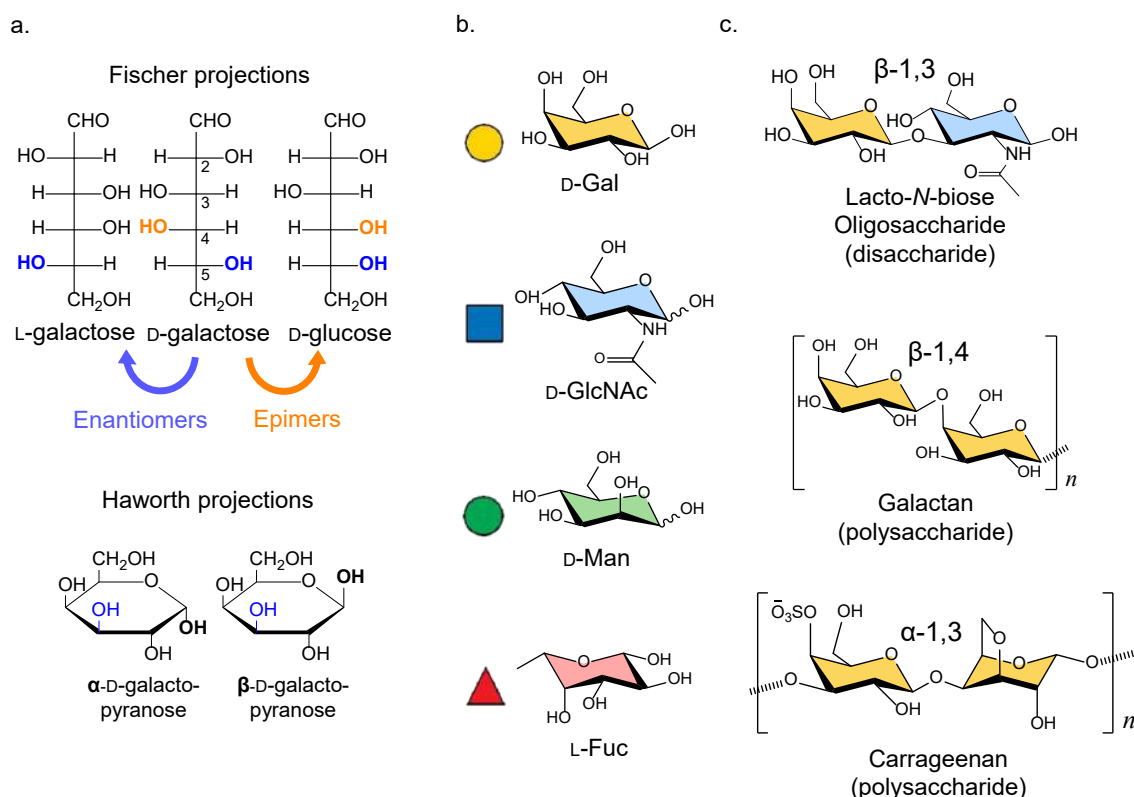


Figure 1.1: a. (Top) Isomeric variations of carbohydrates (6-member depicted): L/D and epimerization of C4, which is the difference between Gal and Glc. (Bottom) Haworth projections of pyranoses showing the stereoisomers of the anomeric C: α and β . b. Some of the most commonly found monosaccharides, which will appear in this Thesis Chapters and their associated symbols. Small differences in stereochemistry and exocyclic groups confer them very different chemical properties and biological roles. c. The glycosidic bond links monosaccharides to form oligo- and polysaccharides. Sugars can be further modified with sulfo- groups and anhydro bridges, like in carrageenans. Adapted from [1], SNFG symbols drawn with DrawGlycan (www.virtualglycome.org/DrawGlycan/) [5, 6].

Monosaccharides link with each other via glycosidic bonds to form larger struc-

tures: oligosaccharides and polysaccharides. The regio- and stereochemistry of these linkages, therefore, the different possibilities of monosaccharide assembly (for example, linear or branched, α or β linkages, position of the binding carbon) introduce further diversity in glycan structures (Figure 1.1.c), which results in different functions. For example, starch and glycogen are glucose polysaccharides with energy storage functions, and galactans are structural polysaccharides in macroalgal cell walls (Figure 1.1.c, bottom) [7]. A relevant example of unconjugated carbohydrates are the human milk oligosaccharides, diverse and complex structures with a common lactose (Gal- β -1,4-Glc) and lacto-*N*-biose (Gal- β -1,3-GlcNAc) [8].

Although carbohydrates, or glycans, are found in their free, native forms, mostly as oligo- and polysaccharides, they are often linked to other types of biomolecules to form glycoconjugates. Glycoconjugates are found across all families of biomolecules (lipids, proteins and nucleic acids) featuring different glycan types. Some examples are glycosphingolipids [9], glycoproteins (with *N*-glycans [10] and *O*-glycans [11]) and proteoglycans (containing for example heparan sulphate) [1]. The glycosylation of proteins is an important post-translational modification in eukaryotic organisms [12] and a key aspect in cell-to-host recognition for pathogenic agents [13,14]. Sugars are present also in nucleic acids, composed by deoxyribose (DNA) and ribose (RNA) [15].

The rich structural diversity, together with the fact that their sequences are not directly encoded in the genome like proteins [1], challenged the study of carbohydrates but opened the door to a fascinating field.

1.2 Carbohydrate ring conformations

Carbohydrates are flexible entities. Their different structural possibilities confer them many degrees of freedom, therefore with a vast conformational and chemical space to explore [2]. These variations include the changes in the relative ring atom positions of the cyclic forms, resulting in the so-called ring puckering conformations, which depend on the regio- and stereochemistry of each monosaccharide [16]. Pyranose ring conformations have been described in 38 canonical conformations which are classified in five groups: chair (*C*), boat (*B*), half-chair (*H*), envelope (*E*) and skew boat (*S*), each distinguishing one or a group of atoms out-of-plane respect to the rest of atoms. The ring atoms are indexed as O, 1, 2, 3, 4 and 5 [17], and to indicate which of them are exoplanar, they are super- or subscripted to the name of the conformation (Figure 1.2).

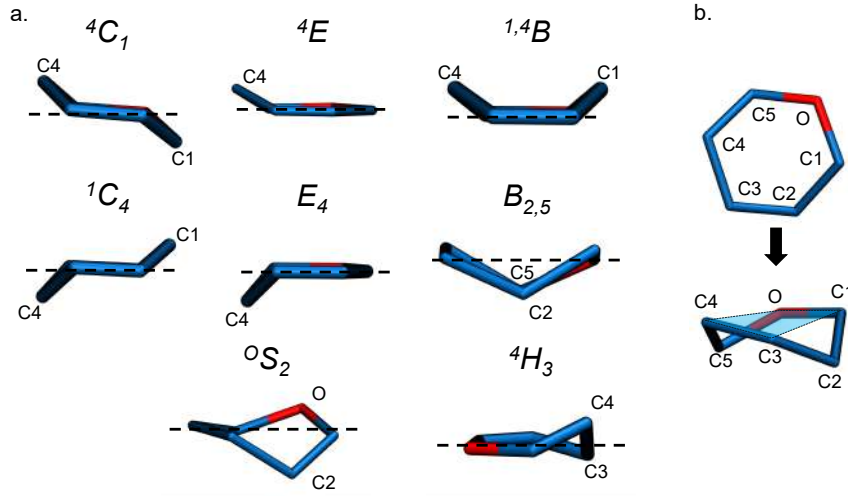


Figure 1.2: a. Examples out of the 38 canonical pyranose puckering conformations. Dashed lines indicate the position of the plane formed by the endoplanar atoms. Exoplanar atoms are indicated with their indices (C1-C5, O). Included the two conformers of the poles (4C_1 and 1C_4) and some representatives of the other classes. b. View from the top of a $B_{2,5}$ conformer showing the convention for index nomenclature for pyranoses, and lateral vision showing the plane formed by atoms C1, C3 C4 and O), and the two exoplanar atoms.

For quantifying pyranose ring puckering mathematically and unequivocally assign conformations, Cremer and Pople presented in 1975 a coordinate system to uniquely assign a conformation to a set of coordinates depending on the position of each plane respect to the mean plane of the ring [18]. In the polar coordinate system, these conformations describe the Cremer-Pople sphere, which shows the inter-relations between all possible conformations. Although this coordinate system is applicable to any ring of N atoms, the discussion here will be centered on six-atom rings. The spherical coordinates of the Cremer-Pople system are θ , ϕ and Q , the amplitude. To calculate which atoms are exoplanar, they first defined the mean ring plane as a function of all six perpendicular distances (j) to the ring average plane (Equation 1.1) [16, 19]

$$Q = \sum_{j=1}^6 z_j \quad (1.1)$$

Which is centered at $Z = 0$. Then, the mean plane is fixed by [18]

$$\begin{aligned} \sum_{j=1}^6 z_j \cos [2\pi (j-1)/6] &= 0 \\ \sum_{j=1}^6 z_j \sin [2\pi (j-1)/6] &= 0 \end{aligned} \quad (1.2)$$

The angles θ and ϕ are computed by solving the expressions (Equation 1.3):

$$\begin{cases} Q \sin \theta \cos \phi = \sqrt{\frac{1}{3}} \sum_{j=1}^6 z_j \cos \left[\frac{2\pi}{6} 2(j-1) \right] \\ Q \sin \theta \sin \phi = \sqrt{\frac{1}{3}} \sum_{j=1}^6 z_j \sin \left[\frac{2\pi}{6} 2(j-1) \right] \\ Q \cos \theta = \sqrt{\frac{1}{6}} \sum_{j=1}^6 (-1)^{j-1} z_j \end{cases} \quad (1.3)$$

Ring conformations can also be quantified in a Cartesian coordinate system by transforming the polar coordinates to the components q_x , q_y and q_z (Equation 1.4).

$$\begin{aligned} q_x &= Q \sin \theta \sin \phi \\ q_y &= Q \sin \theta \cos \phi \\ q_z &= Q \cos \theta \end{aligned} \quad (1.4)$$

Q is the amplitude of the puckering sphere and accounts for the ring planarity, and can also be expressed with the Cartesian components (Equation 1.5):

$$Q = \sqrt{q_x^2 + q_y^2 + q_z^2} \quad (1.5)$$

In this coordinate system, the chair conformers are at the poles of the sphere (4C_1 at $\theta = 0$, and 1C_4 at $\theta = \pi$). While the different ring conformations entail large variations of θ and ϕ (between 0 and π), the value of Q varies a little among conformers, which can be represented in the surface of the Cremer-Pople spheroid of radius Q [19].

There are different ways to represent puckering coordinates in two dimensions. The Stoddart projection represents either the Northern or Southern hemisphere of the Cremer-Pople sphere, and each conformation is assigned based on the Cartesian components q_x and q_y . The spherical coordinates θ and ϕ can be projected into the Mercator representation [16] (Figure 1.3).

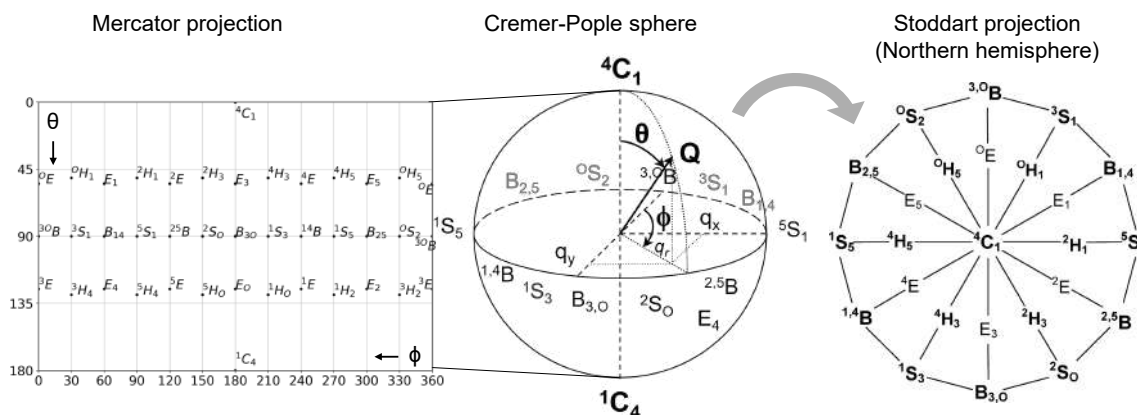


Figure 1.3: Used projections of the Cremer-Pople sphere to represent puckering conformations in 2 dimensions. Mercator representation projecting the spherical ϕ and θ coordinates and Stoddart projection of the Northern hemisphere of Cremer-Pople, projecting the Cartesian components q_x and q_y .

It must be noted that these mathematical systems only assign conformations based on ring atomic positions and their inter-relationships, but they do not contain information on the energies of the ring conformations [16]. Computer simulations of the conformational free energy landscapes give information of the relative stability of conformers, thus, of the conformational preferences of each saccharide [19]. Carbohydrate ring conformations are a key aspect of their properties including their reactivity, and the conformational changes they adopt when undergoing degradation, synthesis and modification processes are known as their catalytic itinerary and will be further discussed in this general introduction.

1.3 Glycoside hydrolases

Carbohydrates and glycoconjugates are major players in many biological processes, in which are involved enzymes responsible of their degradation, synthesis and modification. The enzymes that process glycans are known as carbohydrate Active Enzymes (CAZYmes). The CAZY database (www.cazy.org) [20] is the cornerstone for systematic reporting, classification and collection of structural, genomic and biochemical data on carbohydrate active enzymes and carbohydrate binding modules. CAZYmes are distinguished in different classes depending on how they modify the glycan substrates. Glycoside hydrolases or glycosidases (GHs) cleave the glycosidic bond that link monosaccharide units in an oligo- or polysaccharide [21]. Glycoside transferases (GTs) form new glycosidic bonds between monosaccharides using

sugar donors activated with nucleotides like UDP [22]. Glycoside phosphorylases (GPs) are a class of CAZymes similar to both GHs and GTs that cleave glycosidic bonds adding an inorganic phosphate group to the substrates [23]. Other classes include non-hydrolytic degradative enzymes and other operating with redox mechanisms [24, 25].

Glycoside hydrolases are the class of CAZymes that hydrolyze glycosidic bonds between linked monosaccharides to degrade oligo- or polysaccharides into shorter forms, therefore, involving a water molecule as a catalytic agent at some point during the mechanism [26]. The degradation of glycosidic bonds is present across many biological processes such as cell waste management [27], onset and development of diseases [28], plant biomass degradation [29] and microorganism-host interactions [30]. The myriad of applications of glycans in different production sectors has raised interest in their systematic and efficient production, which is challenging with traditional chemical synthesis methods due to their structural complexity and isomeric variability. Glycoside hydrolases are a potential tool to overcome these challenges, as they are able to synthesize glycosidic bonds providing the appropriate conditions [31]. Some CAZymes are potential therapeutic targets due to their involvement in many processes both in health and disease, as they are involved in metabolic diseases like fucosidosis [32], Krabbe's disease [9], inflammatory bowel diseases [33] and virulence processes by pathogenic bacteria [34, 35] and other diseases like cancer and Alzheimer's [36].

At the present time glycoside hydrolases (GHs) are classified in 189 families based on sequence homology [21]. This is useful for comparing structures, reaction mechanisms and other characteristics [37]. GHs are distinguished according to the cleaving position in a glycan chain, which relates to the shape of their active site cavities: endo-GHs degrade glycosidic bonds in middle positions of the oligo- or polysaccharide chain, whereas exo-GHs recognize and cleave terminal saccharides, thus, at the end of the chain (which is generally the non-reducing one [38]) (Figure 1.4.a). The active site cavity topologies are prepared to host the substrates according to the cleavage site: endo-GHs tend to have more closed, tunnel-like active sites, whereas exo-GHs have more open, cleft-like active sites. In this Thesis the focus is in exo-GHs. However, we will show examples of an exo-acting GH that cleaves two sugars at the non-reducing end of an oligosaccharide (Chapter 3), with a fairly-solvent exposed active site, an exo-GH operating on algal oligosaccharides with a very narrow, crevasse-like active site (Chapter 5) and two other enzymes that have flexible elements or openings around the active site cavity, allowing the entrance of different substrates (Chapters 4 and 6).

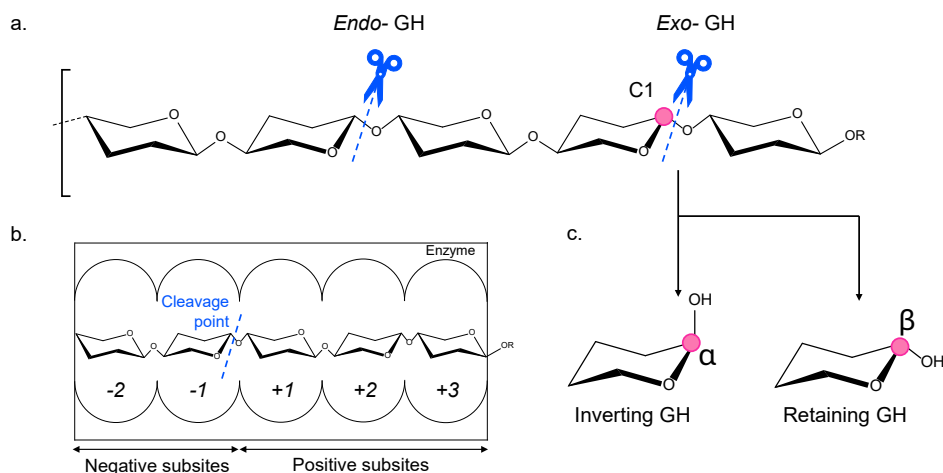


Figure 1.4: a. Classification of GH types according to the cleavage site along the substrate. Endo-acting GHs cleave bonds within the saccharide chain, whereas exo-acting GHs cleave the terminal positions, or the two last in some cases [30]. b. Schematic representation of the saccharide substrate binding sites in a GH. Each subsite accommodates one saccharide unit, and the linkage to degrade between the -1 and $+1$ subsites. This scheme represents an endo-GH. The nomenclature system for an exo-GH is equivalent. Adapted from [16] and [38]. c. Classification of GHs according to the stereochemistry of the products anomeric carbon (in pink circle) respect to the reactants.

Another differential feature of GHs that is key for their classification is the stereochemistry of the reaction products versus the substrate. When a GH cleaves a glycosidic linkage between two sugar units, the moiety that is at the nonreducing end of the cleaved bond is named the -1 sugar (it is the sugar bound to the -1 subsite of the enzyme active site), and the one at the other side, the $+1$ (Figure 1.4.b). The configuration of the anomeric carbon of the sugar at subsite -1 is inverted in inverting GHs (for example, from the β anomer to the α or vice versa), while it is maintained the same before and after the catalytic cycle in retaining GHs (therefore, a β glycosidic bond cleavage yields a β -sugar moiety at subsite -1 [16] (Figure 1.4.c).

1.3.1 Glycoside phosphorylases

Glycoside phosphorylases (GPs) are a class of CAZYmes that degrade glycosidic linkages and form a new bond between the substrate and a phosphate group, thus incorporating it to the non reducing end of non-phosphorylated saccharides [23] (Figure 1.5). GPs can be classified among either GH and GT families due to sequence and fold similarity and share aspects of the reaction mechanism with both of them, with the main difference being that one of the catalytic agents is a phosphate group in GPs. They can also be classified as retaining or inverting depending on the configuration of the resulting product with respect to the substrate. For more details on GPs, their similarities and differences with GHs and GTs, as well as examples

from CAZY families, see Chapter 4 of this Thesis. There we present in detail the mechanism of an inverting glycoside phosphorylase of family GH130 with a prominent role in microbe-host interactions in the human gut [13,33].

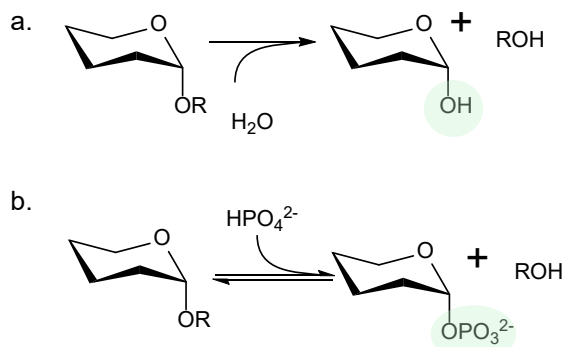


Figure 1.5: Scheme representation of enzymatic action of a. glycoside hydrolase (GH) and b. glycoside phosphorylase (GP). Although both result in the cleavage of a glycosidic bond, GHs yield a product with a hydroxyl group in the anomeric position of the -1 sugar, whereas GPs incorporate a phosphate group in that position, resulting in a phosphorylated sugar. For more details on GPs, see the introductory section, 4.1 of chapter 4 of this thesis. Adapted from [23,39].

1.3.2 Reaction mechanisms

In July 2024, the CAZY database contains data for glycoside hydrolases classified across 189 families (www.cazy.org), and this number rises every year as new functions are discovered, for example in metagenomics studies [40]. However, the heart of most GH mechanisms shares a common basis: the S_N2 substitution and general acid/base catalysis, to cleave the glycosidic bond and assist the departure of the leaving group of the reaction, and inversion or retention of the stereochemistry of the anomeric carbon. The catalytic roles are generally performed by a pair of essential amino acid residues of the enzyme, which are usually aspartate or glutamate (general acid/base), and aspartate (nucleophile or base) [41]. From this common element stem differences across GHs and, as will be presented in this work, include notable deviations from the shared basic mechanism.

In 1953, D. E. Koshland described what we call nowadays the "classical GH mechanisms" [42], and its difference lies in the stereochemistry of the reaction product versus the initial substrate. The inverting mechanism takes place in a single reaction step, or single displacement, where a water molecule performs the nucleophilic attack on the anomeric carbon (Figure 1.6.a), assisted by a base residue, and a general acid residue donates a proton to the oxygen of the scissile bond. At the end of the reaction, the product has an inverted anomeric configuration respect to the initial state. Some inverting GH families include GH8 β -endoglucanases [43],

GH43 xylanases [29], GH47 α -mannosidases [44], and, some inverting GPs, like family GH65 and GH94 enzymes [13], and GH130 β -mannosidases [33] where the nucleophile agent is a phosphate group, discussed in chapter 4 of this Thesis.

Retaining GHs operate via a two-step, also named as double displacement, mechanism where two consecutive S_N2 substitution take place with the formation of a covalent glycosyl-enzyme intermediate (Figure 1.6.b). During the first reaction step, an enzyme residue acting as nucleophile attacks the anomeric carbon, which results in the cleavage of the glycosidic bond and the formation of the covalent intermediate, while a residue that acts as general acid/base protonates the oxygen of the scissile bond to facilitate the departure of the leaving group. The second reaction step proceeds subsequently as a catalytic water molecule, activated by the residue now acting as a base, attacks the anomeric carbon, thus cleaving the glycosyl-enzyme bond and releasing the reaction product. After this, the catalytic cycle is complete. It is during this second step that, if using an appropriate acceptor instead of water, the enzyme can synthesize a new glycosidic bond and yield an oligosaccharide [31]. The net anomeric configuration of the sugar product is maintained at the end of the catalytic cycle (see reaction products in Figure 1.6). The reaction mechanisms of several retaining GHs have been studied both with experimental and computational approaches, such as enzymes in GH59 galactosidases [9] or GH16 glucanases [45].

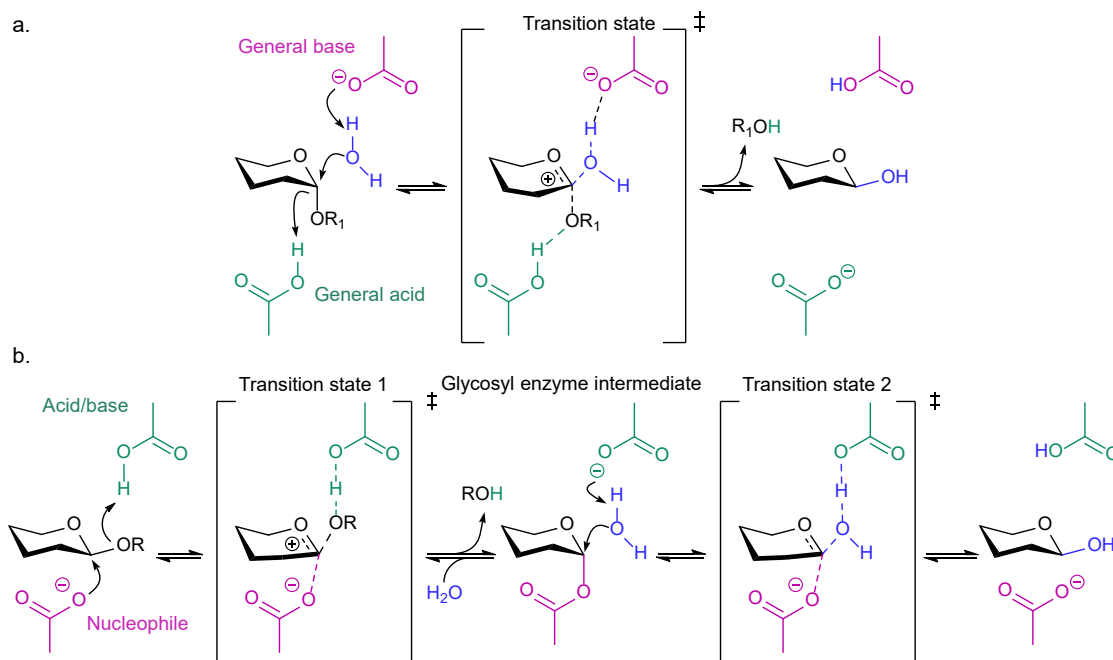


Figure 1.6: Classical GH mechanisms. a. Mechanism by inverting GHs. b. Mechanism by retaining GHs. Adapted from [26] and [16].

GH mechanisms, either via single- or double-displacement have in common the

oxocarbenium ion-like species at the transition state [16, 38, 46]. The experimental observation of this species is challenging due to its short-lived nature. The oxocarbenium ion-like TS is formed when a positive charge develops at the anomeric carbon (in retaining GHs it happens in two different scenarios: when the enzyme nucleophilic attack takes place and the glycosidic linkage breaks, in the first reaction step, or when the catalytic water molecule attacks the glycosyl-enzyme intermediate, in the second reaction step). The charge separation between the nucleophile agent (enzyme residue or water molecule), with negative charge density, and the anomeric center, with positive charge density, results in the shrinkage of the endocyclic O-C1 distance, as the bond between these two atoms acquires partial double bond character. The conformation of the reactive sugar at subsite -1 is a crucial aspect of the oxocarbenium ion-like species, and will be discussed in section 1.3.4.

1.3.3 Non-classical mechanisms: substrate-assisted catalysis

Notable variations from the classical Koshland mechanisms have long been described for GHs, constituting a diverse group of alternative mechanisms. Some variations of the classical GHs mechanism consist in residues other than the typical Asp and Glu playing the role of nucleophile, for example a Cys residue [47], or general acid/base [38]. Alternative mechanisms also include those of processive enzymes like β -glucosidases from family GH3, in which the reaction product has a role in preparing the active site for the next catalytic cycle (known as product-assisted) [48].

Substrate-assisted mechanisms (also known as neighboring group participation mechanisms) are a particular and diverse alternative to the classical GH catalytic mechanisms where the substrate itself plays a role in the catalytic action, for example as nucleophile or general acid of the reaction, via one of its functional groups assisted by an enzyme residue. This can take place via an aminoacetyl substituent at C2, such as the 2-NHAc group in GlcNAc substrates targeted by hexosaminidases (GH18 chitinases [49], GH20 hexosaminidases [30], GH84 O-GlcNAcases [36], or a hydroxyl substituent of the substrate. A remarkable example of this is the mechanism of GH99 endo- β -mannanase, in which the Man substrate 2-OH participation results in the formation of an epoxide intermediate [26, 50]. An example of 2-NHAc substrate-assisted mechanism will be presented in chapter 3 of this Thesis.

Another case of hydroxyl substrate-assisted catalysis is the one present in inverting β -mannosyl phosphorylases of family GH130 [33]. Enzymes of this family have been proposed to operate via a substrate-assisted mechanism where the 3-OH substituent of the mannosyl substrate participates in the reaction but, unlike to the

GH99 mannanase, the hydroxyl plays the role of last proton donor of the leaving group, instead of an enzyme residue. This catalytic mechanism will be studied in detail in chapter 4 of this Thesis.

Catalysis with 2-NHAc substrate assistance proceeds as shown in Figure 1.7.a. A negatively charged residue of the enzyme activates the 2-aminoacetyl exocyclic group of the -1 sugar by interacting with its NH- group. Then, the 2-NHAc group attacks the adjacent anomeric carbon via an intramolecular nucleophilic substitution, instead of an enzyme residue taking this role, while a general acid/base residue transfers a proton to the scissile bond oxygen. In this type of mechanisms, since no enzyme residue acts as nucleophile of the reaction, the covalent glycosyl-enzyme intermediate does not form. Therefore, after the formation of an oxocarbenium ion-like transition state, an oxazolinium ion or oxazoline intermediate is formed [49]. Then, the second displacement step starts as a water molecule enters the active site cavity and acts as the nucleophile, activated by the now general base catalyst, thus promoting the cleavage of the five member oxazolinium ring and yielding the reaction product [30, 49].

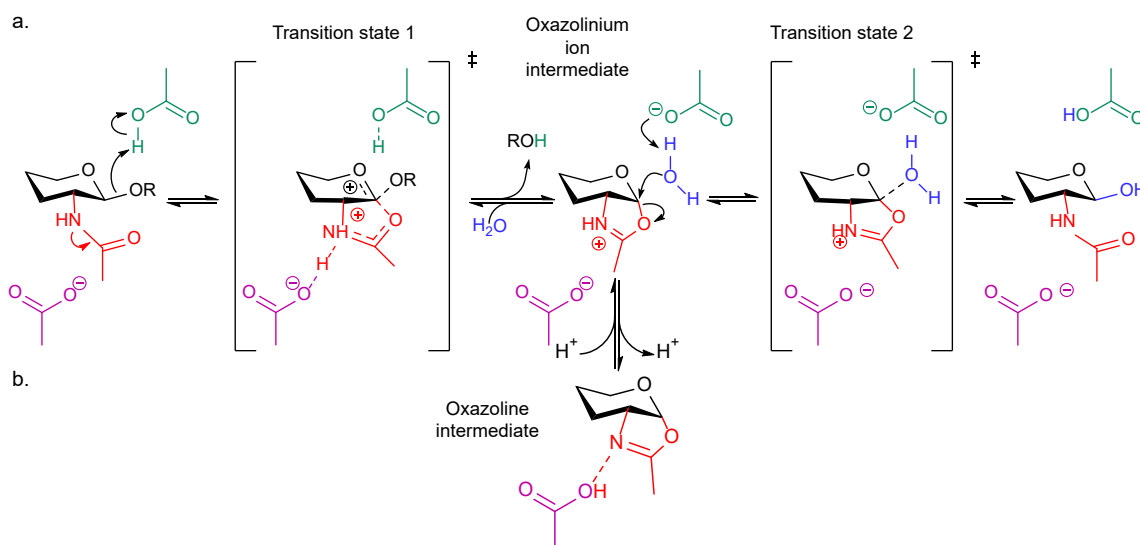


Figure 1.7: a. Scheme of 2-NHAc substrate assisted mechanism that will be studied in this Thesis. Adapted from [26]. b. Scheme of the equilibrium between oxazolinium ion and oxazoline intermediates.

The nature of reaction intermediates of enzymes operating with 2-NHAc assisted catalysis has been debated in previous works, especially in the PhD. Thesis Dissertation of Dr. Joan Coines, former member of our group [51]. After the first reaction step a five member ring is formed, constituting a bicyclic intermediate Figure 1.7.a. It has two possible forms: oxazoline (neutral), where the proton of the former 2-NHAc group has completely transferred to the assistant residue, and oxazolinium

ion (positively charged), where the proton remains with the endocyclic nitrogen. In a 2018 computational study by Coines *et al.* the authors reported a substrate-assisted GH18 chitinase featuring a neutral oxazoline intermediate (Figure 1.7.b). They hypothesized that the direct interaction between the acid/base, separated in the amino acid sequence by one position (while other substrate-assisted enzymes have adjacent catalytic residues), facilitated the appropriate conditions to stabilize this form via a proton relay, which adopted oxazolinium ion character towards the second reaction step. Other substrate assisted GHs with adjacent catalytic residues have been reported to favor the charged form [52].

Hydroxyl-assisted mechanisms have been reported in family GH130 inverting β -mannosyl phosphorylases. These enzymes feature an active site architecture where the acid catalyst is further from the oxygen of the scissile bond, thus challenging a direct proton transfer. Crystal structures of a GH130 inverting phosphorylase solved in 2013 by Nakae *et al.* showed that the 3-OH substituent of the Man substrate was suitably oriented to form a hydrogen bond network connecting the putative acid catalyst and the glycosidic bond [53]. The authors proposed an alternative mechanism to the direct protonation where this enzyme residue transfers a proton to the 3-OH of Man, which in its turn transfers its proton to the scissile bond. The substrate engages in a proton relay to assist the leaving group protonation, while the nucleophilic attack is performed by the inorganic phosphate group in the active site [33] (Figure 1.8).

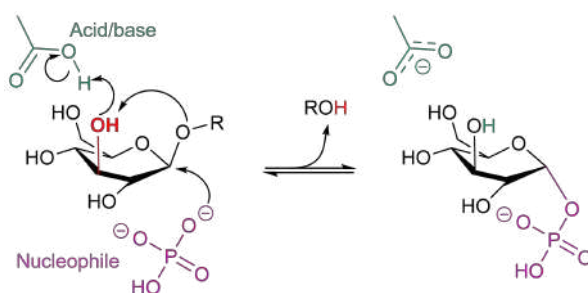


Figure 1.8: Schematic representation of the substrate-assisted mechanism of some inverting phosphorylases. In red, H atom from the substrate 3-OH that is transferred to the departing group.

The details of this double hydrogen bond reaction pathway will be analyzed in detail in chapter 4 of this Thesis.

1.3.4 Conformational itineraries in GH catalysis

Glycosidase catalysis has yet another trait that distinguishes them, the conformational itinerary. The conformational itinerary is described as the changes in conformation that substrates undergo during catalysis [54] (Figure 1.9.a).

The conformation adopted by the substrates in the Michaelis complex is a topic that has long been discussed in GH catalysis [55]. Substrates are observed to bind to enzymes in conformations that depart from the relaxed chair 4C_1 or 1C_4 , which are ground states of the unbound saccharides conformational free energy landscapes [9, 19, 41]. The reason behind this is that several of this so-called "distorted" conformations, meet stereochemical criteria favorable for catalysis [56]. When unbound to enzymes, cyclic carbohydrates adopt different conformations. Which ones are more stable is an intrinsic characteristic of each sugar, and depends on the nature of its exocyclic substituents and the stereochemistry of their chiral centers (as described in section 1.2). The conformational preferences of isolated sugars have long been studied in our group with QM/MM metadynamics approaches to compute the conformational free energy landscape of saccharides [16, 19], and it has been suggested that the enzymes have evolved to take advantage of these conformations to make catalysis more efficient [16].

The oxocarbenium ion-like species at the TS of GH catalyzed reactions has the lone pairs of the endocyclic oxygen delocalized. The pyranose ring geometry most suitable for this requires the ring atoms C5, O, C1 and C2 (Figure 1.2.b) in the same plane [38], which is possible only in some of the canonical pyranose ring conformations (3E , E_3 , 4E , E_4 , $B_{2,5}$, ${}^{2,5}B$, 4H_3 and 3H_4) [16] (Figure 1.9.b). Therefore, the TS of any GH-catalyzed reaction is expected to adopt one of these eight conformations.

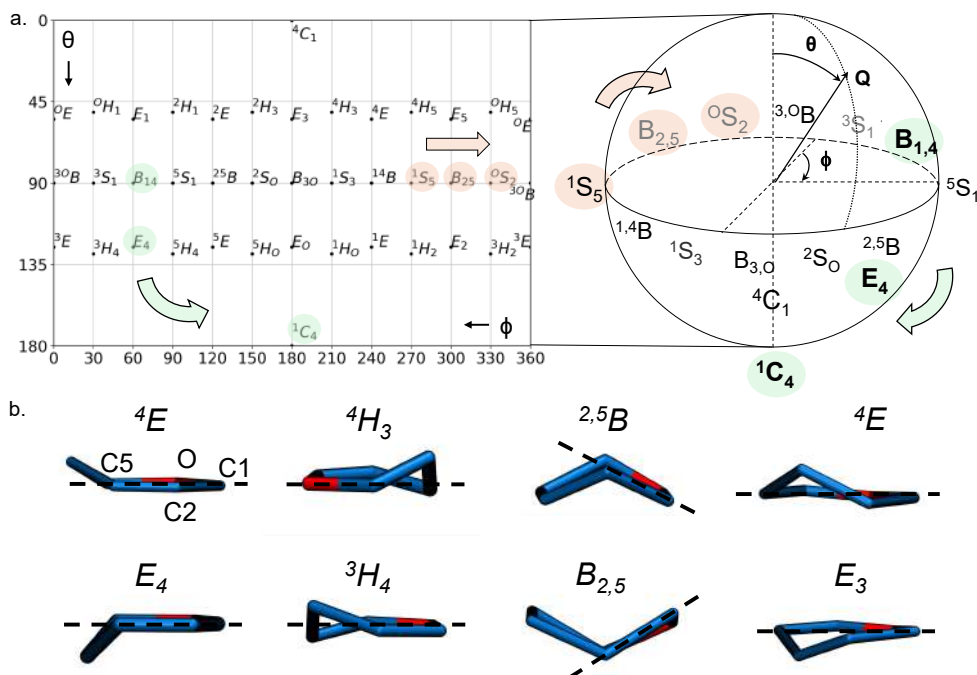


Figure 1.9: a. Possible pyranose conformations of the reaction TS. In E_4 , indicated the four atoms C5-O-C1-C2 that are coplanar in these conformations. b. Representation of two GH catalytic itineraries on the Cremer-Pople sphere (right) and its Mercator projection (left). In green, a catalytic itinerary that takes place at the Southern hemisphere of the sphere (GH129). In orange, an equatorial itinerary (GH130).

Substrate distortion at the Michaelis complex has been observed in X-ray structures [57, 58] and confirmed in first-principles computational studies [16]. Substrate distortions provide certain geometric and electronic characteristics that facilitate catalysis: they are pre-activated conformations, and allow for reaching the TS with the least possible nuclear motion. Substrate conformations also provide the suitable geometry for an in-line nucleophilic attack (electrons of the nucleophile will populate the σ^* orbital of the glycosidic bond), and place the glycosidic bond in axial orientation [59]. This relieves steric hindrance at the anomeric carbon for the nucleophile attack (as the anomeric hydrogen is in equatorial orientation, whereas in chair conformers is axial, hampering nucleophilic attack) (Figure 1.10). Also, they facilitate developing partial double bond character between the endocyclic oxygen and anomeric carbon [54]. Therefore, substrates adopt distortions that are on the path to become one of the eight possible conformers at the TS (see that conformational itineraries always follow adjacent conformations in Figure 1.9.a) [38]. Some exo-GHs have been reported to bind and process undistorted substrates, in cases where the leaving group is already axially oriented without the distortion or where leaving group departure is strongly favored, or other geometric traits [9, 27, 28, 38].

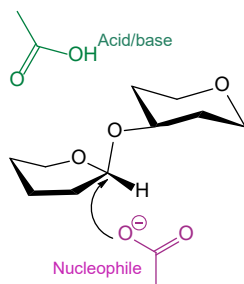


Figure 1.10: Schematic representation of a distorted substrate at the MC. Conformations departing from the ground states 4C_1 and 1C_4 position the glycosidic bond axially, while leaving the hydrogen of the anomeric carbon in equatorial orientation, ready for nucleophilic attack (black arrow).

Unveiling conformational itineraries is useful for designing specific inhibitors and activity probes, as organic chemists can have insights on what conformations the molecules they design should adopt in complex with an enzyme, and predict the conformational preferences within a GH family. Especially relevant are the conformations at the transition states, as their short-lived nature challenges their observation. Atomistic modeling of reaction mechanisms can report these at atomic and molecular level, thus providing valuable information.

1.4 Scope of this Thesis

In this Thesis we aim to expand the knowledge on different types of glycosidases that have some particularity setting them apart from the classical inverting or retaining ones. We have seen that, while many glycosidases and glycoside phosphorylases share a common mechanism, others introduce variations that deal with specific substrate types and unusual active site geometries. Enzymes are plastic entities that have evolved to adapt to the vast structural diversity of glycans, and even small differences in substrate substituents or the stereochemistry of one carbon requires a whole different enzyme to process them. We have also studied the conformations cyclic form of pyranoses, and how their conformational itineraries impact GH catalysis.

Unveiling the molecular details of enzyme reaction mechanism opens doors to enzyme engineering strategies. This has applications in enhancing certain enzyme activities or fine-tuning variants to selectively degrade heterogeneous substrates. For example, knowing the role of active site residues of a GH can guide the design of a mutant that performs better the alternative synthetic mechanism towards the production of high added-value oligosaccharide products, like in the case of human milk oligosaccharide-processing enzymes, or to selectively degrade algal polysaccharides to obtain industrially relevant algal-derived gellifying materials.

Disease-related CAZYmes are also important therapeutic targets, with the development of specific inhibitors and activity-based probes to design new diagnostic tools. Understanding the conformational itineraries of a substrate, as well as the reaction mechanism, can assist in rationalizing the design of molecules. For example, as the intestinal degradation of *N*-glycans is related to the onset and development of pathogenic and disease processes, developing specific inhibitors and probes can be useful to treat them, mediate microorganism-host relationships and to design diagnosis tools. Other enzymes related to the management of cell waste can be a target for developing efficient sequencing tools, as well as therapeutic targets for metabolic diseases. The enzymes we have focused on in this Thesis are the following:

- In chapter 3 we study the reaction mechanism of *Bifidobacterium bifidum* lacto-*N*-biosidase, LnbB, an enzyme that catalyzes the degradation of the main type of human milk oligosaccharides through a substrate assisted mechanism, where the enzyme environment details that regulated it had not yet been elucidated.
- In chapter 4 we study the enzyme Unknown Human Gut Mannoside Phosphorylase (UhgbMP), an enzyme secreted by an intestinal bacterium of the *Bacteroides* genus that incorporates phosphate to β -linked mannose motifs of *N*-glycans. We study its substrate assisted mechanism and reveal a protonation pathway alternative to other inverting glycoside phosphorylases.
- In chapter 5 we study the reaction mechanism of *Zobellia galactanivorans* 3,6-anhydro-D-galactosidase, *ZgGH129*, an enzyme involved in the selective processing of complex algal oligosaccharides. We also study the conformational preferences of its unique substrate 3,6-anhydro-D-galactose and compare it to other galactose-processing enzymes.
- In chapter 6 we study human fucosidase FucA1, an enzyme that cleaves fucose from different glycans with a retaining mechanism and obtain insights on its ability to recognize different fucosylated substrates.

Objectives

In this Thesis we have unveiled the catalytic mechanism of two glycosidases operating with substrate-assisted catalysis, and two glycosidases that operate by the classical Koshland mechanism. First, we studied a retaining GH where the substrate plays the role of nucleophile and one inverting GP where the substrate is the proton donor of the scissile glycosidic bond. We have also studied the catalytic mechanism of a retaining 3,6-anhydro-D-galactosidase and the conformational preferences of its unique substrate that delineates a Southern hemisphere itinerary. Finally, we have studied the catalytic mechanism of a fucosidase, also operating in the Southern hemisphere, and unlocked its structural plasticity to accept different substrates. For these studies we used molecular dynamics techniques, both classical and the hybrid QM/MM approach, together with enhancing sampling methods (metadynamics).

The specific objectives for each study are as follows:

- Describe the full catalytic mechanism of LnbB with the natural substrate LNT and elucidate the protonation state of its reaction intermediate. Elucidate the catalytic itinerary of the substrate. (Chapter 3)
- Unveil the molecular features of the LnbB active site that make possible the intramolecular nucleophilic substitution for substrate-assisted mechanism. (Chapter 3)
- Elucidate the rare substrate-assisted mechanism of UhgbMP with an *N*-glycan motif substrate and the strategies of the enzyme to deal with a distal catalytic acid residue. (Chapter 4)
- Characterize the reaction mechanism of ZgGH129 with a natural algal oligosaccharide substrate from the first crystal structure of this enzyme-substrate. (Chapter 5)
- Determine the conformational preferences of the constrained 3,6-anhydro-D-galactose *in vacuo* and in the enzymatic environment and identify its reactive conformations. (Chapter 5)
- Study the structural determinants used by FucA1 to recognize different fucosylated oligosaccharides to unveil the key behind its substrate plasticity and validate its reactivity with one of the natural substrates. (Chapter 5)

Chapter 2

Methods

Enzyme catalysis is an interdisciplinary field, where areas of scientific research as diverse as molecular biology, structural biology, or spectroscopy techniques provide structural and kinetic information. Enzymes are dynamic entities whose structures serve specific functions. From the large secondary structure elements, such as helices, sheets and barrels, to the small details of what residues make up their active sites, the interplay between structure and function is a long-standing research question. If we put the focus in the heart of an enzyme, we will study the chemical reactions that make possible the transformations of substrates. Computational chemistry techniques are widely used to understand, rationalize and predict enzyme catalytic mechanisms.

One of the cornerstones in the study of enzyme and, in general, protein systems, are the different crystallographic methods, which have a long-standing role for structural determination. To this is added the recent development of predictive tools. In all, both historic and well-established methods and newer techniques aim to aiding researchers in approaching the fundamental question of sequence to structure relationship, and thanks to them we can elucidate structures of many large biomolecules at the atomic detail.

However, when we aim to study the dynamics of a system and also, the chemical transformations that it may undertake, we need to go beyond a fix set of atomic coordinates. While classic biochemistry techniques, such as spectrophotometric assays, chromatography or kinetic isotope assays give invaluable insights about the nature of the atoms undergoing transformations, the type of mechanisms taking place and the nature of the reaction adducts and products, some shortly-lived states are elusive to these techniques, therefore challenging their characterization. In other situations, we might want to access different states of our system that are separated by high energy barriers and therefore not likely to be sampled.

Computational Chemistry has developed techniques that, through models based

on physical laws, estimate properties of the models they represent. Researchers can build models at different levels of description (for example, at the quantum level) depending on the phenomena they study. The concept *molecular simulations* refers to how we generate representative sampling of a system [60]. With the different molecular simulations techniques, researchers propagate the models, whichever the level of description chosen. The way a system is propagated to generate an ensemble of states is one of the main divisions in molecular simulation techniques. First, Monte Carlo (MC) techniques use probabilistic laws to propagate the system from an initial configuration to other states. At the end, an ensemble of states is obtained, although its states do not have a temporal relationship between them. Monte Carlo methods will not be described further in this dissertation.

The other great family of methods in molecular simulations, and the one that has been used for the work in this Thesis, uses the equations of motion to propagate, over time, a system from an initial state. The final obtained ensemble is a trajectory comprising time-evolving states, which allows the dynamic interpretation of system properties [60]. Molecular dynamics will be described in more detail in the upcoming sections.

2.1 Introduction to molecular dynamics

Molecular dynamics is the molecular simulation method for propagating a model, comprised by a set of initial coordinates and velocities, through Newton’s equation of motion [60] (Equation 2.1).

$$\vec{F}_i(t) = m_i \vec{a}_i(t) = m_i \frac{\partial^2 \vec{r}_i(t)}{\partial t^2} = -\frac{\partial V(\vec{r}_i)}{\partial \vec{r}_i} \quad (2.1)$$

What results from the propagation is a *trajectory*, whose definition is a series of time-correlated points in phase space [60]. This stems from the approximation that atoms, nuclei, are heavy particles that behave classically, thus the deterministic propagation of the Newton motion laws holds true. In Equation 2.1, $\vec{r}_i$ is a vector containing the positions of all the particles of the system. In other words, the coordinates of all the atoms. Therefore, when propagating the equations of motion, Molecular Dynamics algorithms calculate new positions after a discrete interval of time, the *timestep* (Δt). Since the analytical integration of the differential equations of motion is not possible for more than three interacting bodies with no constraints [61], the role of the numerical *integrator* comes at play. An important example of MD integrators is the Verlet algorithm, which is based in the following Taylor expansion (Equation 2.2):

$$\vec{r}_{i+1} = \vec{r}_i + \frac{\partial \vec{r}}{\partial t}(\Delta t) + \frac{1}{2} \frac{\partial^2 \vec{r}}{\partial t^2}(\Delta t)^2 + \frac{1}{6} \frac{\partial^3 \vec{r}}{\partial t^3}(\Delta t)^3 + \dots \quad (2.2)$$

Each partial derivative of the positions (r) respect to time t corresponds to one of the vector quantities of the system (velocities, v_i , accelerations, a_i) of all of the system particles. By making appropriate arrangements, the new positions of the system after a timestep unit (Δt) are deterministically calculated as (Equation 2.3):

$$\vec{r}_{i+1} = (2\vec{r}_i - \vec{r}_{i-1}) + \vec{a}_i(\Delta t)^2 \quad (2.3)$$

Being r_{i+1} the new atomic positions. This iterative process is repeated for every timestep unit, thus generating the correlated trajectory points. The choice of timestep remains a tradeoff between accuracy and cost: the smaller Δt , the more precise, although then more steps are necessary for the same amount of simulated time, which is more computationally demanding.

In expression 2.3 the acceleration, a_i is calculated using the Newton equation of motion, which are also as the change in potential energy of the system as a function of position (Equation 2.1). How this potential energy is calculated for all particles of the system depends on the level of theory used for each model, and will be described shortly.

2.2 Energy calculation: Forcefield methods

We have seen in the last section that what molecular dynamics algorithms do is take a set of initial conditions from a system made of particles, and numerically and iteratively solve the Newton equations of motion to propagate over time and generate a trajectory. The energy component of this expression can be obtained from the system using fundamentally different approaches, which depend on what laws of Physics we respect when designing our model. If to the initial assumption that atoms in a model behave like classical particles, therefore, with a mass m_i , that follow the deterministic laws of motion, we decide that our model is composed by *nuclei* and do not account for the electrons and thus, Quantum Mechanics to any effect, we are in front of a Molecular Mechanics (MM) method to calculate the potential energy of a system. Although molecular dynamics algorithms may use molecular mechanics methods to obtain the potential energy, thus, the system's forces in equation 2.1, to discretely advance to the next iteration and approximate the continuum system real dynamics, other computational chemistry approaches use energy calculations without propagating system dynamics.

A *forcefield* is a parametric expression to calculate the energy of a system only as a function of its *nuclei* positions. The parameters of a forcefield do not come from solving the Schrödinger equation, but are provided from either experimental data or higher level quantum calculations [60]. A typical expression of a forcefield is equation 2.4 a sum of contributions of mechanical and electrostatic phenomena [62]:

$$\begin{aligned}
 E_{total} &= E_{bonded} + E_{non-bonded} = \\
 &= \sum_{bonds} k_d(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi(1 + \cos(n\phi - \phi_0)) \quad (2.4) \\
 &+ \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{ij} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]
 \end{aligned}$$

The forcefield expression contains terms for different types of contributions to the potential energy E_{total} : E_{bonded} describes atomic interactions that stem from atoms that share a bond and $E_{non-bonded}$, those atoms that are not bonded. In forcefield energy calculation schemes, to the classical description of particles we add the assumption that bonds between atoms behave like springs. Thus, bonded terms of equation 2.4 include the contributions of bond stretching (2 atoms involved), angle bending (3 atoms involved) and dihedral rotations (4 atoms involved) (see Figure 2.1), for each pair, group of 3 and 4 directly bonded atoms of the system. Non-bonded contributions to the potential energy include the electrostatic (assuming point charges q_i and q_j for any pair of atoms at a r_{ij} distance) and Van der Waals interactions.

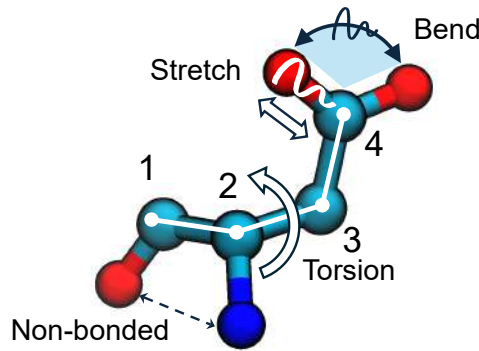


Figure 2.1: Representation of the bonded terms included in a molecular mechanics forcefield. Bond stretching, angle bending and dihedral angle rotation, represented on an aspartate molecule. Adapted from [61]

All terms of the forcefield equation that do not depend on the particles positions

in space are the *parameters*. Forcefield parameters are obtained through different methods (for example, experimental methods, or computational at a higher level of theory) [60], and they aim to emulate the real behavior of the atomic types or molecules they represent (for example, a sp^3 carbon, or an oxygen belonging in a carboxylate group) [63]. Forcefield parametrization is an entire field within computational chemistry, and in this dissertation examples of different forcefields used for modeling different types of molecules (GLYCAM for carbohydrates, FF19SB for proteins) [64, 65] will be shown. The main strengths of forcefield methods are their transferrability and relatively low computational cost [60]. Therefore they are very attractive in biomolecular simulations, where researchers can model systems with thousands of atoms. However, as molecular mechanics methods do not account for quantum effects, their main limitation lies in their inability to account for any phenomena involving electron rearrangements. Thus, photochemical phenomena, or chemical reactions in enzymes cannot be modeled with forcefields.

2.3 Quantum mechanics methods

In a classical system, *state* means that the position and velocity can be determined at any instant (Equation 2.4). However, things are very different at the subatomic scale: the Heisenberg uncertainty principle reveals that at that scale an observer cannot determine the position and velocity of a particle (its linear momentum), therefore it is not possible to predict its future behavior, like we did deterministically with classical systems (Equation 2.5) [66].

$$\Delta x \Delta p > h \quad (2.5)$$

At the subatomic level, we must find another way [67]. Quantum mechanics has been formulated through *postulates*, of which the central one states that all properties of microscopic systems are described by wavefunctions. Through quantum operators applied to wavefunctions, one can *predict* the probability of the system acquiring a certain value for a property [63]. The central equation of quantum mechanics is the Schrödinger equation (Equation 2.6).

$$\hat{H}(r, t)\psi(r, t) = i \frac{\partial \psi(r, t)}{\partial t} \quad (2.6)$$

The Schrödinger equation is a differential equation with solution $\psi(x)$ which describes a particle of mass m moving in a potential field, $V(x)$. We can imagine the wavefunction $\psi(x)$ as the amplitude of the wave associated to a particle [66], which can be written as (Equation 2.7):

$$-\frac{\hbar}{2m} \frac{d^2\psi}{dr^2} + V(x)\psi(r) = E\psi(r) \quad (2.7)$$

The left part of the expression corresponds to the Hamiltonian operator. Linear operators have a central role in quantum mechanics formalism and they are related to classical quantities. Thus, the Hamiltonian operator represents the energy [66], and the eigenvalue formulation of the Schrödinger equation is expressed in equation 2.8.

$$\hat{H}\psi(r) = E\psi(r) \quad (2.8)$$

Thus, it all comes down to a matter of how to obtain the energy of a system, only that for reaching the quantum level of description, the Schrödinger equation should be solved for all particles in a model. Several methods have been developed over the years to deal with the electronic structure of atomic and molecular systems, each proposing different approximations for solving the Hamiltonian, thus resulting in different levels of accuracy (which, as always, entail a tradeoff with the cost of the calculation) [60, 63].

2.3.1 Energy calculation: DFT

The Density Functional Theory (DFT) is one of the pillars of computational chemistry for retrieving the energy of a atomic or molecular system and is the method for energy calculation used in all reaction and carbohydrate conformation simulations in this Thesis. In section 2.3, we have seen that at the quantum level, the energy of the system state is associated to a Hamiltonian operator, which rather than the position of the system particles, depends on its wavefunction, ψ , which could be obtained by solving the Schrödinger equation.

However, working with wavefunctions is not trivial, starting by the high dimensionality (a set of three spatial coordinates and one spin for each electron), which becomes an unsurmountable task for polyelectronic (and thus, polyatomic) systems [63]. For addressing this, several approximations have been developed in order to interpret the energy of a quantum system like any other physical observable. Since the energy of a system is closely tied to the Hamiltonian operator, which depends on the electronic structure of the system, developing an observable depending on the electron number seemed a good way to obtain the energy, namely the *electron density* [63] (Equation 2.9).

$$E = F[\rho(r)] \quad (2.9)$$

In this expression, $\rho(r)$ is the electronic density. In 1964 Hohenberg and Kohn [68] established the necessary bases that eventually solidified DFT as a robust method for modelling quantum systems in Chemistry [63] by proposing that the ground state energy of a system of interacting electrons can be calculated as a function of the electronic density. Later, Kohn and Sham developed the Kohn-Sham orbitals to approximate the ground state energy of the polyelectronic system from single-electron orbitals represented as ϕ_i (Equation 2.10.1) [69].

$$\rho(r) = 2 \sum_{occ.}^i |\phi_i(r)|^2 \quad (2.10.1)$$

$$E^{DFT} = \min_{\phi_i} E^{KS}[\{\phi_i(r)\}, \{R_N\}] \quad (2.10.2)$$

The energy of Kohn-Sham is expressed as Equation 2.11 [70]:

$$E^{KS}[\{\phi_i\}] = T_s[\{\phi_i\}] + \int V_{ext}(r)n(r)dr + \frac{1}{2} \int V_H(r)n(r)dr + E_{XC}[n] \quad (2.11)$$

An expression defining the Kohn-Sham energy as a sum of contributions corresponding to the kinetic energy, two potential energy terms corresponding to the attraction between nuclei and electrons and the interelectronic repulsion, respectively [60]. The Kohn-Sham energy E_{KS} is then equal to the DFT energy, although it leaves the last term, the exchange and correlation energy, E_{XC} , unknown. There are different ways to approximate the exchange correlation energy contribution, from which one of the most widespread is the Generalized Gradient Approximation [70].

When applying DFT to solve the electron structure of a system, it is fundamental to choose carefully the functional that will ultimately describe the interactions and behavior of our system's electrons. In the present work, the gradient-corrected Perdew, Burke and Ernzerhof functional (PBE) [71] has been used, as its performance has been assessed for carbohydrate ring conformations and carbohydrate-processing enzymes [72, 73].

The basis set is the term to refer to the functions used to represent the Kohn-Sham orbitals to calculate the electronic density [70]. Different approaches have been historically proposed. For example, Slater or Gaussian basis sets, which are based in the positions of the nuclei. Another type of functions to periodically expand KS orbitals are the so-called plane waves, which are defined as a set of functions (Equation 2.12):

$$f^{PW}_G(r) = N \exp[iGr] \quad (2.12)$$

Where N depends on the volume of the periodic cell. G is a wave vector quantity that can be tuned with an appropriate energy cutoff. Plane waves perform well in periodic systems and systems with delocalized electrons. Core electrons, namely, those that are not valence electrons, are strongly localized. Since their description with plane waves would be computationally very expensive, they are described with pseudopotentials [60, 74].

2.3.2 Quantum molecular dynamics

In section 2.3.1 we have provided an overview of how to calculate the energy of a system taking into account the contribution of its electrons. In contrast, we have also summarized how to calculate the energy of a classical system and how it is applied in solving the Newton equations of motion to propagate it with molecular dynamics. However, in sections 2.1 and 2.2 we have only discussed systems where its particles behave deterministically, with a mass m and point charges. Again, if we wish to descend to the subatomic level to explain phenomena that involve electron rearrangements, or where quantum effects cannot be overlooked, we need a formalism that explicitly includes the electrons.

Again, the expression that governs the quantum world is the time-dependent Schrödinger equation (Equation 2.13):

$$i\hbar \frac{\partial}{\partial t} \psi(\{r_i\}, \{R_I\}; t) = \hat{H} \psi(\{r_i\}, \{R_I\}; t) \quad (2.13)$$

In the quantum molecular dynamics formalism, we express the Schrödinger equation with terms accounting for the electronic ($\{r_i\}$) and nuclear ($\{R_I\}$) degrees of freedom separately, due to the fundamental different behaviors of these particles. The Hamiltonian operator, $\hat{H}$ is a function of both the nuclei and atomic degrees of freedom, and contains contributions accounting for the interactions between all them [70]. How the degrees of freedom of each fundamental particle, therefore, how to they move over time, is addressed distinguishes the main methods in quantum molecular dynamics.

Analogous to what we do for classical molecular dynamics, in quantum (or first principles) molecular dynamics, a succession of calculations of the system energy is performed at discrete intervals of time and the positions of the system particles are updated, thus reconstructing a trajectory. In first principles or *ab initio* MD, the electronic structure is solved in order to compute the energy, for example with

DFT [70].

Born-Oppenheimer molecular dynamics (BOMD) is the approach that solves the electronic structure considering static, fix, nuclear positions at each discrete time interval. Thus, the time-independent Schrödinger equation (Equation 2.14.2) is used, instead of the time-dependent expression of equation 2.7:

$$M_I \ddot{R}_I(t) = -\nabla_I \min \left\{ \left\langle \psi_0 \left| \hat{H} \right| \psi_0 \right\rangle \right\} \quad (2.14.1)$$

$$E_0 \psi_0 = \hat{H} \psi_0 \quad (2.14.2)$$

The main characteristic of BOMD is that, at each step of the propagation (namely, the discrete time interval or timestep), $\hat{H}$ has to be minimized. As expressed in Equation 2.14.1, in BOMD the positions of the *nuclei*, $\ddot{R}_I(t)$, are taken into consideration, and their motions follow the rules of classical mechanics, while the electronic structure is addressed by solving the time-independent Schrödinger equation at each time interval (known as self consistently) [70]. The impact of this rationale in calculations using BOMD is that the chosen value for the discrete time intervals to iterate over can be of the order of magnitude of the movement of nuclei (about 0.5 fs).

Car-Parrinello molecular dynamics (CPMD) entails a different conceptualization of the dynamics of nuclei and electrons. Proposed in 1985 by Roberto Car and Michele Parrinello, it takes into account both the degrees of freedom of nuclei and electrons, separately [75]. Contrary to BOMD, in CPMD the equations of motion cannot be integrated in the timescale of nuclear motion, as electrons move orders of magnitude faster. Thus, the time-independent Schrödinger equation cannot be used to treat electrons here [70]. For incorporating both nuclear and electronic motion in the molecular dynamics formalis, Car and Parrinello introduced a new expression, in the form of a Lagrangian (Equation 2.15.1):

$$\mathcal{L}_{CP} = \underbrace{\sum_I \frac{1}{2} M_I \dot{R}_I^2}_{\text{nuclei } E_k} + \underbrace{\sum_i \mu \left\langle \dot{\phi}_i \left| \dot{\phi}_i \right\rangle}_{\text{electrons } E_k} - \underbrace{\left\langle \psi_0 \left| H_e \right| \psi_0 \right\rangle}_{E^{KS}, \text{ DFT}} + \text{constraints} \quad (2.15.1)$$

$$E_{el}^k = \sum_i \mu \int d\vec{r} |\dot{\Psi}_i(\vec{r})|^2 \quad (2.15.2)$$

Where each term accounts for the different motion scales [70]. To propagate the system, eventually this expression is used to obtain the Newton equations of motion,

adapted to this formalism. Therefore, eventually it all comes down to iterate over discrete time intervals, now closer to the faster motion of electrons, compute the energy of the state (for example, with an electronic structure method like DFT) and output the new state of the system, for nuclei and orbitals. Due to the simultaneous evolution over time of these fundamentally different particles, the concept of *adiabatic separation* is introduced with the electronic fictitious mass, μ , which slows down electrons, thus enabling larger timestep values without compromising the accuracy of the simulation (Equation 2.15.2). Still, an usual timestep of a CPMD simulation is of the order of 0.12 fs, while in BOMD is around 5 fs.

2.4 QM/MM molecular dynamics

Ab initio methods are very appealing for modeling natural phenomena accurately at the quantum level. However, these methods are demanding in terms of computational cost, as the bigger the system the more complex the calculations become. After some hundreds of atoms, they are unsurmountable with today's technologies. And yet, more often than not we cannot reduce our research question to just a few atoms. For example, the systems studied in this dissertation are large biomolecules, with thousands of atoms, solvated by some more thousands of water molecules. In 1976 a new approach bridged the two worlds: the small-scale quantum description, with the larger, classical scale: QM/MM, or quantum mechanics/molecular mechanics is a hybrid method that combines two different levels of theory for one system [76, 77]. Its authors A. Warshel and M. Levitt, and M. Karplus [77], were awarded the Nobel Prize for Chemistry in 2013. If we go back to the protein example, we might be interested in a chemical transformation involving a small region of a few atoms (for example, the active site), while we might not be interested in what is going on at the electronic level with all the other atoms (which today is still almost unattainable anyway). Thus, this method combines a description at the quantum level for a region of chemical interest with mechanical description for the rest of the system (see Figure 2.3).

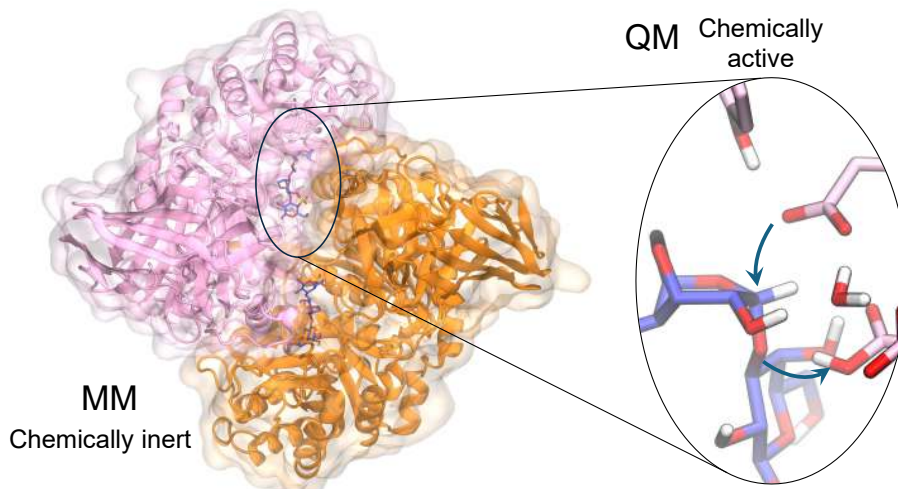


Figure 2.2: (Left) Overview of a system consisting in a dimeric protein (*ZgGH129*, which will be discussed in chapter 5, represented in orange and pink cartoons, with a bound substrate, represented in violet sticks. In a QM/MM approach for studying how the enzyme processes this substrate, atoms in the active site cavity would be treated at the quantum level (QM), whereas the rest of atoms would be described classically. (Right) Detail of the enzyme active site, of chemical interest if one wishes to study its catalytic mechanism.

How is this concept incorporated to the calculation of the state of the system? We have seen that in molecular mechanics methods the energy of all system particles is obtained with the parametric equation known as forcefield. At the quantum level, DFT among others calculates the energy as a function of the electron density. In the QM/MM hybrid approach, the energy is again a sum of the contributions accounting for each type of particle (Equation 2.16):

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.16)$$

The E_{QM} and E_{MM} terms are obtained with energy calculation methods appropriate for the level of theory, like DFT or another electronic structure method for the QM region and a suitable forcefield for the MM region. How the $E_{QM/MM}$ term is approached deserves further discussion. One of the caveats of QM/MM is that, even though the system is divided quantum and classical level of description, both subsystems are *interacting*, in the sense that they are not independent from each other. This interface between the quantum-described and classical regions is accounted for in the $E_{QM/MM}$ term, and there are different strategies to approach it. The QM-MM boundary energy contribution distinguishes the two great types of interatomic interactions in an all-atoms system: bonded and non bonded interactions (Equation 2.17) [78].

$$E_{QM/MM} = E_{QM/MM}^{bonded} + E_{QM/MM}^{non-bonded} \quad (2.17)$$

Bonded QM and MM interactions take place when, due to the choice of the QM region for that system, the boundary cuts through a linkage between two atoms. Several methods address this, by either saturating the valence that has been left in the MM region with a *dummy* hydrogen atom or pseudopotential. The treatment of non bonded QM-MM interactions, namely, van der Waals and electrostatic interactions (see expression in Equation 2.18), differs according the type of non-bonded interaction. While repulsion and dispersion interactions are computed with the classical definition of van der Waals interaction, electrostatic effect are computationally demanding and has been addressed in different ways. Here we will discuss the strategies used by the two QM/MM schemes used in this work, CPMD and CP2K.

$$E_{non-bonded}^{QM/MM} = \sum_{MM} q_i \int dr \frac{\rho(r)}{|r - r_i|} \sum_{\substack{QM \\ MM}} v_{vdw}(r_{i,j}) \quad (2.18)$$

The QM/MM scheme implemented in Car-Parrinello molecular dynamics formalism calculates electrostatic contributions of the QM/MM boundary region, aiming to address the interaction between two fundamentally different observables: the classical point charges in the MM region (represented as q_i in expression 2.18) and the electron density in the QM region represented as the total charge $\rho(r)$ in 2.18. Laio, VandeVondele and Röthlisberger developed in 2002 a multilayer electrostatic coupling scheme [78]. The idea of this scheme is that, computing electrostatic QM/MM interactions is very expensive because the number of calculations is of several folds the order of the MM defined atoms. Thus, the multilayer approach lies in distinguishing these interactions in short- and long-range interactions, and applying different schemes to calculate them, according to how far they are from the QM region.

The shortest range electrostatic interactions are calculated with the following expression (Equation 2.19).

$$H_{el} = \sum_{j \in MM} q_j \int dr \rho(r) v(|r - r_j|) \quad (2.19)$$

Where a potential, $v(r)$ is introduced to calculate the interactions of short ranges, which derives from the Coulomb potential [78]. This is the most precise approximation for calculating the QM/MM interaction forces. For longer range interactions, Laio and coauthors proposed a classical Coulomb interaction (Equation 2.20):

$$H_{el} = \sum_{\substack{j \in ESP \\ j \in QM}} \frac{q_i q_j}{r_{ij}} \quad (2.20)$$

Finally, the remaining interactions, at very long range, are described by multipole expansion schemes [78]. The areas of the system under each of these approaches are defined by the so-called electrostatic coupling radii (NN for definition of Equation 2.19, ESP for Equation 2.20). The assessment of what atoms of the system are best described with each approach is system dependent and is done by scanning different values for the radii. For compromising between accuracy and computational cost, the rule of thumb is to select small values yielding errors under $1.0 \text{ kcal mol}^{-1}$ [79]. A third layer, known as MIX can be added combining NN and ESP methods.

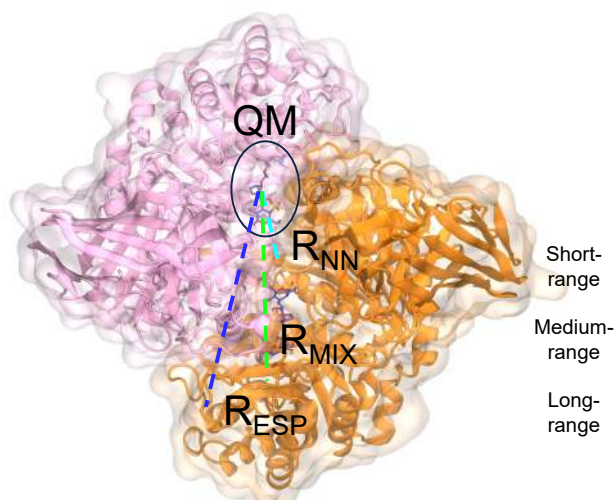


Figure 2.3: Schematic representation of the electrostatic radii defining which parts of the MM subsystem will be treated with the different methods for dealing with QM/MM electrostatic interactions near the boundary region between the two subsystems.

In other QM/MM MD schemes, other methodologies address the contribution to the Hamiltonian by QM/MM electrostatic interactions near the boundary region. The program CP2K, instead of the multilayer electrostatic embedding developed by Laio and coworkers [78] uses a method known as GEEP: Gaussian Expansion of the QM/MM Electrostatic Potential, developed by Parrinello *et al.* [80] in 2005. In this method the electrostatic potential is treated as Gaussian functions like Equation 2.21:

$$n(|r - R_a|) = \left(\frac{1}{\sqrt{\pi} r_{c,a}} \right) \exp \left(-\frac{|r - R_a|^2}{r_{c,a}^2} \right) \Rightarrow v_a(r, R_a) = \frac{\text{erf} \frac{|r - R_a|}{r_{c,a}}}{|r - R_a|} \quad (2.21)$$

This method maps the MM point charges-as-Gaussian-like functions onto grids whose mesh size depends on the Gaussian shape. A threshold value discriminates whether a particular MM atom is considered for the electrostatic interaction calculation or not based on distance, thus, reducing the number of necessary operations and thus the computational cost [80].

In summary, the multiscale QM/MM approach is a method that has allowed to model natural phenomena involving chemical transformations in systems with a high number of atoms that would otherwise be unsurmountable with first principles methods alone. Although the cost bottleneck is still the quantum description of the chemically active area, the electrostatic interactions, especially the effects between QM and MM atoms at the boundary between the two subsystems, are also expensive and thus QM/MM formalisms include different ways to ease this. Although QM/MM methods allow the accurate description of chemical systems taking into account the role of electrons, propagating the system with QM/MM MD schemes soon reaches a different type of challenge: the timescale.

2.5 The timescale challenge

When we propagate a system over time to reconstruct its dynamic behavior, every discrete time step forward we take is often confined to a local energy minimum of a potential energy surface. The phase space of most systems is vast, and often states of interest, or system configurations, are located in several of these local minima, separated by energy barriers. It would take an enormous, unrealistic amount of time to effectively sample all of them, unless the system was governed by a diffusion regime. Plus, this is aggravated by the fact that the timestep of integration of molecular simulations at QM/MM level are orders of magnitude smaller than the timescale in which many chemical phenomena occur [81]. This difficulty for effectively accessing different system configurations called for schemes to accelerate the sampling of a conventional molecular dynamics simulation, which oscillate approximately at the k_bT order. There are many different enhanced sampling algorithms, among which metadynamics has been used throughout this thesis, coupled to QM/MM MD simulations.

2.5.1 Enhanced sampling with metadynamics

Metadynamics is a method for effectively sampling the free energy surface associated to a complex, multidimensional phenomenon and accessing *rare events*, namely, states of the phase space that would otherwise be very unlikely to be ever sampled,

as a function of a few degrees of freedom named collective variables (CVs) [82]. Introduced in 2002 by Laio and Parrinello [83], it is based in adding a history-dependent bias potential term to the system that discourages it from re-visiting regions, thus pushes it towards other regions of the phase space (Figure 2.4). An advantage of this method respect to other enhanced sampling schemes like umbrella sampling [84] is that no previous information about the other regions is needed, as the system will evolve first to low energy regions [85].

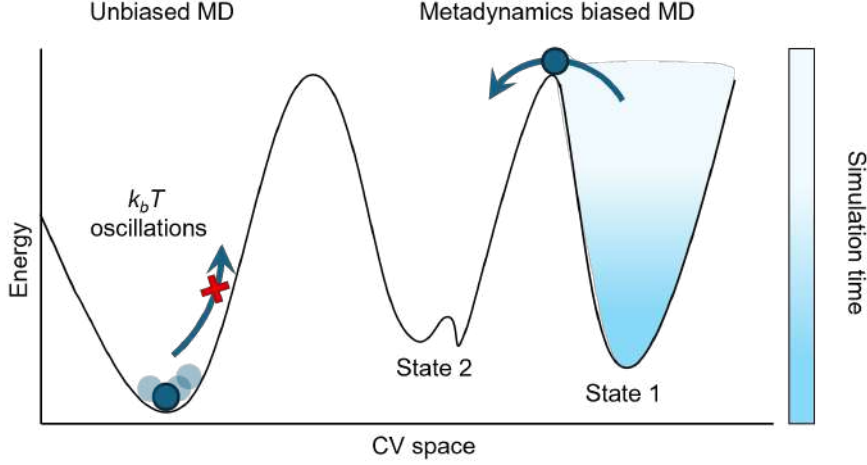


Figure 2.4: Schematic representation of the main difference between unbiased MD (left) and applying the metadynamics method for enhanced sampling (right).

Collective variables are descriptors of the system, a small set of functions $S_i(r)$ of the system coordinates [85] that describe the exploration space. In metadynamics, a Gaussian-like bias potential is constructed along the phase space created by the CVs, which is then integrated over time. The estimation of this method is that, at a long simulation time, the applied potential function corresponds to the underlying free energy surface, which in its turn is a function of the collective variables (Equation 2.22) [85]:

$$V_G(S, t) = \frac{W}{\tau_G} \int_t^0 \exp \left(- \sum_{i=1}^d \frac{(S_i(r) - S_i(r(t')))^2}{2\sigma_i^2} \right) dt \quad (2.22)$$

Where W is the height of the Gaussian kernels, in energy units, and τ_G is their deposition stride, in time units. Note that the historical potential that adds to the system Hamiltonian is a function of the CVs $S_i(r)$ and thus, of the system coordinates [82]. σ_i is the width of the Gaussian kernels for each CV, in CV units.

One of the main challenges of metadynamics is a dimensionality reduction problem: choosing a small number of CVs that is able to describe univocally the phase

space of a certain event. In the context of enzyme catalysis, CVs should be able to assign the states of the system in the reaction coordinate (reactants, reaction intermediates, products, etc.). The other major drawback of metadynamics simulations is the convergence. One of the cornerstones of the method is that, at a long simulation time:

$$V_G(S, t \rightarrow \infty) = G(S) \quad (2.23)$$

However, it is not trivial to know when a simulation is long enough, and the bias potential does not converge smoothly in standard metadynamics scheme. This impacts in both the estimation of the error and poses a risk in overfilling the underlying free energy estimation. For computing the error of a reconstructed free energy landscape, in the context of simulations of enzyme catalysis, a common strategy is to compare independent simulation [85,86].

2.6 General workflow in simulations of enzyme catalysis

Throughout the present PhD dissertation thesis are presented modeling studies of enzyme catalysis. For choosing which methods are most suitable for addressing a research question, we can take a look to what physical and chemical phenomena are at the center of the system and what level of description they might require. Enzyme catalysis consists in the chemical transformations that a molecule, the substrate, undergoes in the active site of an enzyme, a large biomolecule of hundreds of atoms. Since we will need to describe electron rearrangements to characterize chemical transformations, we need an electron structure method, like DFT. On the other hand, the sheer size of the system (not only the enzyme, but the solvation cell), requires a hybrid approach as the quantum description of atoms that are not involved in the reaction, or in the close vicinity, is not needed. Finally, since we aim to drive the system through the different states of a reaction coordinate, QM/MM MD should be coupled to an enhanced sampling method like metadynamics. The usual starting point for all models is a set of (fix) coordinates of the enzyme-substrate complex, or a variant thereof. This structures might have been obtained with crystallographic techniques, or more recently, with protein structure prediction algorithms. For propagating the system from its initial state and obtaining its dynamical behavior, we first model the system with forcefield-based, classical MD (ns timescale). With this approach we can not only study some dynamic characteristics of the system, but also have an idea of its most populated configurations. Then, after formulating an

hypothesis of the catalytic mechanism, we turn to hybrid QM/MM methods. We start from a trajectory frame of classical MD that is representative. After choosing the region to be described at the quantum level, we choose appropriate parameters for the implementation of the QM/MM formalism, either in the CPMD or CP2K programs. We then run QM/MM MD (ps timescale) and study the behavior of the system slow modes to assist in the choice of collective variables that might be best for the hypothesized reaction mechanism and the parameters for the Gaussian-like kernels. After this QM/MM MD step we can start the QM/MM metadynamics simulation.

6

Chapter 3

The catalytic mechanism of GH20 lacto-*N*-biosidase

Publication:

Cuxart, I., Coines, J., Esquivias, O., Faijes, M., Planas, A., Biarnés, X., Rovira, C., "Enzymatic hydrolysis of human milk oligosaccharides. The catalytic molecular mechanism of *Bifidobacterium bifidum* lacto-*N*-biosidase." *ACS Catalysis*, 12, 8, 4737-4743 (2022)

Abstract: The intestinal tract of breast-fed infants is rich in bacteria equipped with enzymatic machinery for breaking down complex carbohydrates present in human milk (human milk oligosaccharides or HMOs), which constitute their carbon source. *Bifidobacterium bifidum* segregates Lacto-*N*-biosidase (LnbB), an enzyme that releases lacto-*N*-biose (Gal- β -1,3-GlcNAc) from most HMOs. Crystal structures show that the enzyme likely works via a two-step, substrate-assisted mechanism in which the substrate itself is the nucleophile of the first reaction step. However, the structure of a LnbB-substrate complex remains unknown, as well as the details of the catalytic mechanism and the conformational itinerary of the substrate. Here we compute the free energy landscape of the two-step hydrolytic mechanism of LnbB on a natural substrate using QM/MM and metadynamics. Our results uncover the molecular determinants of the mechanism, showing that the substrate undergoes the conformational itinerary ${}^1S_3/{}^1,4B \rightarrow [{}^4E]^{\ddagger} \rightarrow {}^4C_1/{}^4H_5 \rightarrow [{}^4E/{}^4H_5]^{\ddagger} \rightarrow {}^1,4B$. We also identified a key histidine residue (His263) that modulates the pK_a of the acid/base residue (Glu321) by changing orientation, thus favoring catalysis. The reaction intermediate of the reaction is mostly in its ionic form (oxazolinium ion), with a minor population of the neutral oxazoline form. This study brings light to the degradative processes by intestinal tract microbiota in the first stages of life and opens doors to enzyme engineering strategies to produce high added value products.

3.1 Introduction

Human milk is a complex mixture containing a multitude of biomolecules that provide nutrients and other bioactive elements needed for the short- and long-term growth and development of infants. It contains lactose, fatty acids, proteins, vitamins and minerals, nucleotides, immunoglobulins, lysozyme and carbohydrates, but its exact composition is variable and dependent on multiple factors [8, 87].

The carbohydrate fraction of human milk consists in glycans either conjugated with other biomolecules forming glycoproteins and glycolipids or in free form, the most abundant of which is lactose (Lac, Gal- β -1,4-Glc). While lactose is also the main glycan component of other mammals' milk, there is a fraction of free, highly diverse and complex saccharides uniquely reported in human milk [8]. Carbohydrates in this fraction are known as human milk oligosaccharides (HMOs), and they have a vast structural diversity as well as of composition and concentration between individuals and during the lactation period. The highest concentration of HMOs in milk, however, is reported both in colostrum and milk for pre-term born infants [88]. Five monosaccharides are the fundamental building blocks of their structures, β -D-glucose (Glc), β -D-*N*-acetylglucosamine (GlcNAc), β -D-galactose (Gal), α -L-fucose (Fuc) and α -D-*N*-acetylneuraminic acid (NeuAc) [87, 89] (Figure 3.1.a). The diverse O-glycosidic linkage types between them, like β -1,(3,4,6), confer them linear or branched structures of different lengths up to 15 saccharide units [90] (Figure 3.1.b).

HMOs play a fundamental role in defining the intestinal microbiome of infants, as they are the carbon source for different types of bacteria that proliferate in the colon. The gut microbiome constitutes a defensive barrier against other pathogens and helps regulating the immunologic system at the beginning of human life. Since limited access to human milk affects the correct growth and development of the intestinal microbioma at the early stages of life, there is interest to explore the enzymatic machinery responsible for the degradation, synthesis and modification of HMOs to improve formula milks [31].

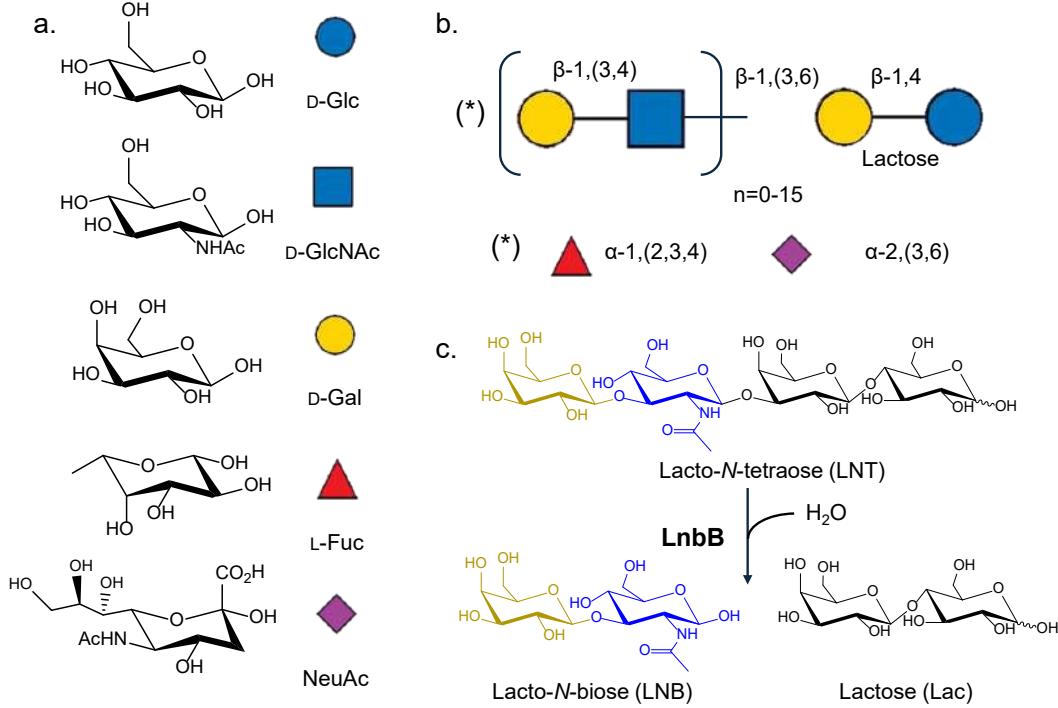


Figure 3.1: a. Monosaccharide building blocks of HMOs, in molecular (left) and Symbol Nomenclature for Glycans (SNFG, right) representations. b. SNFG representations of possible HMO backbone (right) with different linkages and complexity that are linked to lactose (left) or at the termini with Fuc or NeuAc. c. Hydrolysis reaction of LNT by LnbB yields the disaccharides LNB and Lac, which can be further metabolized by gut bacteria. Adapted from [87] and [30].

Lacto-*N*-biosidase from *Bifidobacterium bifidum* (LnbB) is a key enzyme that degrades HMOs in breastfed infants' gut. It was isolated for the first time in 2008 [88]. LnbB hydrolyzes HMOs by releasing lacto-*N*-biose (LNB, Gal- β -1,3-GlcNAc) from the non-reducing end of the lacto-*N*-tetraose core (LNT, Gal- β -1,3-GlcNAc- β -1,3-Gal- β -1,4-Glc) present in the most prevalent HMO types [91] (Figure 3.1.c). LnbB belongs in family 20 of glycoside hydrolases (GH20) [92], like other β -hexosaminidases, β -1,6-*N*-acetylglucosaminidases and β -6-SO₃-*N*-acetylglucosaminidases. GH20 enzymes have diverse domain organizations that may include catalytic and non-catalytic domains. Val-Cid *et al.* determined in 2015 that the GH20 catalytic domain by itself is inactive and requires a remote element to shape the active site and restore the activity, which in LnbB in particular corresponds to a lectin-like domain (residues 570-578) at the C-terminus of the catalytic domain extends towards the active site [93]. Tridimensional structures of LnbB were first solved in 2013 by Fushinobu *et al.* [94] corresponding to a complex with either the product LNB or a LNB-thiazoline inhibitor (Protein Data Bank, PDB, entries 4H04

and 4JAW at www.rcsb.org, respectively, Figure S3.1). The bicyclic structure of the LNB-thiazoline in this last complex led the crystallographers to hypothesize that LnbB works by a substrate-assisted mechanism for its catalytic action (Figure 3.2). Later, in 2015, Hattie *et al.* solved the crystal structure of LnbB in complex with other inhibitors which gave more insight about the pyranose conformations at the enzyme active site and also supported that the enzyme operates via substrate-assisted catalysis [95].

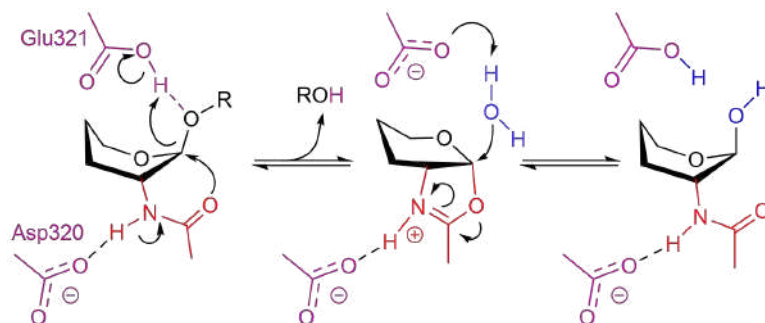


Figure 3.2: Scheme of the double-displacement, neighboring group participation mechanism hypothesized for LnbB, where the 2-NHAc substituent adjacent to the scissile bond attacks the anomeric carbon, assisted by Asp320. R=Gal- β -1,4-Glc. Adapted from [30].

In this type of substrate-assisted catalysis, also known as neighboring group participation, the enzyme uses a substituent group of the reactive sugar (located at the -1 subsite of the active site cavity, adjacent to the scissile glycosidic bond) to perform a part of the catalytic reaction [96]. LnbB uses the NHAc substituent at the C2 of the GlcNAc at subsite -1 to perform a nucleophilic substitution on its anomeric carbon leading to the cleavage of the glycosidic bond. A glutamate residue acts as general acid/base catalyst to protonate the leaving group, and the intramolecular nucleophilic attack is facilitated by an assisting aspartate that interacts with the NHAc substituent inducing its polarization. The identity of the catalytic residues of LnbB was confirmed with mutagenesis studies: Residues Asp320 and Glu321 were identified as the assistant and acid/base respectively, as they both are necessary for catalysis. The mutation Glu321Ala abolishes all enzymatic activity, whereas the Asp320Ala variant retains 1% of the wild-type enzyme [94, 97]. After the glycosidic bond cleavage the reaction proceeds through the formation of an oxazoline or oxazolinium ion intermediate depending on the location of the 2-NHAc proton, which is a priori not known. Other enzymes reported operating via substrate assisted catalysis are GH18 chitinases [98, 99], GH20 hexosaminidases [96, 100–102] and 6S-GlcNAcases [103], GH84 *O*-GlcNAcases [104–106], GH56 hyaluronidases [107–109], GH85 endo- β -*N*-glucosaminidases [110–112] and GH123 endo- β -*N*-galactosaminidases [113, 114].

Besides the identity of the catalytic residues, another key aspect in GH catalysis is the conformational changes that the substrate experiences during the reaction. Fushinobu *et al.* proposed a substrate conformational itinerary inferred from the conformations of the LnbB crystal complexes and similar enzymes [101, 115–117]. According to their hypothesis, the substrate follows an itinerary at the Northern hemisphere of the Cremer-Pople sphere [18], adopting ${}^{1,4}B \rightarrow [{}^4E]^{\ddagger} \rightarrow {}^4C_1$ (first reaction step) and ${}^4C_1 \rightarrow [{}^4E]^{\ddagger} \rightarrow {}^{1,4}B$ (second reaction step) itineraries along the double displacement mechanism. The proposed itinerary starts from a distorted boat ${}^{1,4}B$ conformation and planar 4E at the transition states of both displacement reactions. At ${}^{1,4}B$ the substrate is pre-activated for catalysis, as the scissile glycosidic bond is in axial orientation and exposes the least hindered face of the anomeric carbon for the NHAc nucleophilic substitution [16, 94].

Since LnbB is the only reported member of the GH20 family with two subsites (-2 and -1) in the active site cavity to bind the disaccharide Gal- β -1,3-GlcNAc (LNB) [94], it remains unclear whether the molecular details of the catalytic mechanism and the conformational changes of the substrate differ from the other GH20s and substrate-assisted GHs in general. The residues involved in shaping the active site cavity contrast from other studied GH20 enzymes, such as the bacterial β -hexosaminidase from *Streptomyces plicatus* SpHex [101], thus affecting the architecture of the active site. Residues Gln190 and Glu216 are key for Gal recognition at the LnbB -2 subsite (Figure 3.3.a), while in the other GH20 SpHex, Arg162 interacts with the monosaccharide at the -1 subsite and blocks the space for a -2 subsite. Other examples are the *N*-glycan degrading virulence factor StrH from *Streptococcus pneumoniae* and the 6S-GlcNAcases from *Bacteroides thetaiotaomicron* and *Bifidobacterium bifidum*, where this space is occupied by Arg196 [34], Arg171 and Arg358 [103], respectively (Figure 3.3.b).

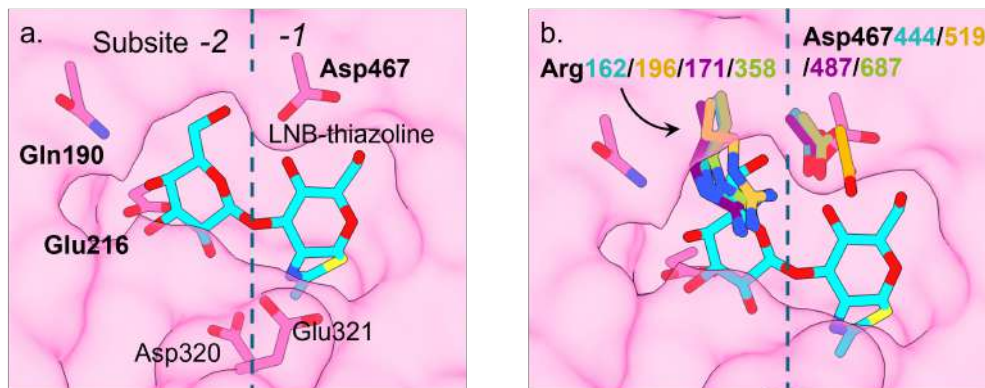


Figure 3.3: a. Active site of LnbB complex with LNB-thiazoline (PDB ID: 4JAW). Bold: residues Gln190 and Glu216 shape the active site cavity with a -2 subsite to host the Gal moiety of LNB. b. Superimposition of other GH20 enzymes whose active site Arg hinders the space for a -2 subsite. *SpHex* Arg162, teal (PDB 1M01), *StrH* Arg196, orange (PDB 2YL8), *Bt4394* Arg171, purple (PDB 7DVA), *BbhII* Arg358, green (PDB 7DTW).

Several residues are involved in substrate recognition of LnbB. Asn259 and Asp320 prevent binding of fucosylated sugars at the galactose unit, a function of the promiscuous FucA1, which will be studied in Chapter 6 of this Thesis, a characteristic shared by other HMO building blocks like 2'-fucosyllactose and iso-lacto-*N*-decaose [88]. The residue Asp467, a position that is conserved across the GH20 family (Figure 3.3), has a dual role as it is involved in substrate recognition as well as contributing to the conformation of the GlcNAc ring. This residue is conserved in other GH20s like *SpHex* [101], *SmChB* [118] or human lysosomal hexosaminidases [119], as shown in the sequence alignments done in the work of Fushinobu *et al.* [94]. By forming a hydrogen bond with the substrate 4-OH, in one hand it disfavors the binding of galactose-based saccharides like galacto-*N*-biose (GNB, also present in the intestine as part of mucin glycoproteins and utilized by gut bacteria [120]), as it is a C4 epimer of glucose, and in the other binding the substrate 4-OH assists in maintaining a distorted substrate conformation. Other conserved positions at the active site cavity are the aromatic nest around the substrate NHAc group (Trp373, Trp394 and Trp 465, Figure 3.4), Tyr419 and His263. This last residue raised questions in previous studies: When mutated to phenylalanine by Fushinobu *et al.* the enzyme activity decreased by two orders of magnitude, although this residue is not directly interacting with the substrate. The WT enzyme k_{cat} with *pNP*-LNB is $15 \pm 1 \text{ s}^{-1}$, whereas for the His263Phe variant it is $0.15 \pm 0.01 \text{ s}^{-1}$ [94]. Therefore, while this residue is key for catalytic activity, whether its exact role is to contribute to the integrity of the active site cavity, assist in substrate binding or to play a part in the catalytic mechanism remained unclear.

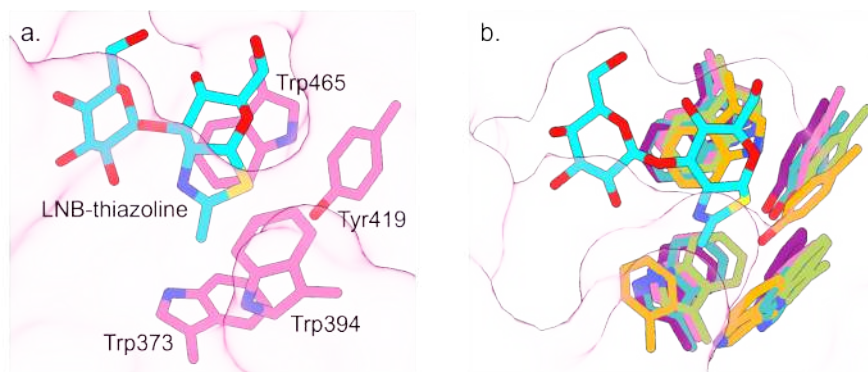


Figure 3.4: a. Active site of LnbB highlighting conserved residues around the 2-HNAc group of LNB-thiazoline. b. Superposition of other GH20 enzymes with conserved residues at the 2-HNAc position. *SpHex* Arg162 in teal (PDB 1M01), *StrH* Arg196 in orange (PDB 2YL8), *Bt4394* Arg171 in purple (PDB 7DVA), *BbhII* Arg358 in green (PDB 7DTW).

Other enzymes acting with neighboring group participation have been studied with quantum mechanics/molecular mechanics (QM/MM) methods, like a family 20 β -hexosaminidase [121], family 18 chitinases [49] and chitobiases [116,117], and family 84 *O*-GlcNAcases [36,52]. Studies of the substrate-assisted mechanism via NHAc substitution suggest that the nature of the reaction intermediate, so, whether it is a neutral oxazoline or oxazolinium ion, depends on the active site configuration. Coines *et al.* found that the oxazoline form is the preferred one in substrate assisted hexosaminidases where the catalytic assistant and acid/base are not adjacent in the amino acid sequence (therefore, with an insertion), with the example of GH18 *SmChiB* [49]. The active site architecture makes possible the direct interaction between the catalytic residues, which establish a proton relay that connects with the substrate. In consequence, at the reaction intermediates the 2-NHAc proton transfers to the assistant residue. However, this seems not to be the case for enzymes with catalytic residues adjacent in the sequence, as shown for GH84 OGA [52]. The molecular mechanism of LnbB has not been studied with computational methods yet and, to the best of our knowledge, neither other enzymes involved in the degradation of HMOs. Concerning the experimental structures, they correspond to a non-hydrolysable reaction intermediate analogs (LNB-thiazoline) and the product complex, but the reactive substrate-enzyme complex, the Michaelis complex, with a natural substrate has not yet been characterized.

Glycoside hydrolases can have a secondary synthetic activity that competes with the main degradative one [31,97]. If an appropriate acceptor is used as nucleophile during the second displacement reaction instead of water to react with the intermediate, complex oligosaccharides can be synthesized with the desired regio- and stereoselectivity. Tuning the synthetic activity, however, is challenging as the intrin-

sic hydrolytic activity is favored. Knowing the intricacies of the enzymatic mechanism, as well as characterizing the reaction intermediate can give ideas to develop engineering strategies for improving the synthetic activity of GHs and using them as biocatalysts.

In this chapter we unveil the full two-step catalytic mechanism of hydrolysis by LnbB. Our results confirm the predicted neighboring group participation by the 2-NHAc group of the GlcNAc at the -1 subsite, a trait shared with other enzymes across GH families. Our simulations also reveal that His263, at the active site, interacts with both catalytic agents, Asp320 and Glu321, modulating their pK_a . By means of QM/MM MD simulations we also characterized the relevant states along the reaction pathway, including the elusive oxocarbenium ion-like transition states and the oxazolinium-ion intermediate, and we obtained the full conformational itinerary of the substrate. The results of this study add to the knowledge of GHs operating with substrate assistance and give clues to develop enzyme engineering strategies for the synthesis of complex saccharides.

3.2 Results and discussion

3.2.1 Modeling the Michaelis complex of LnbB with lacto-*N*-biose

To gain insight on the enzyme-substrate interactions at the active site of LnbB and the main determinants for its catalytic action we built the Michaelis complex with the natural substrate LNT (Lacto-*N*-tetraose, Gal- β -1,3-GlcNAc- β -1,3-Gal- β -1,4-Glc). The enzyme initial atomic coordinates were obtained from the tridimensional structure (available at PDB, www.rcsb.org ID: 4JAW) of LnbB in complex with the inhibitor LNB-thiazoline, a nonreactive analog of the reaction intermediate [94]. Since no complex with a natural substrate is available we built the complex with LNT using docking and forcefield molecular dynamics (MD) simulations (Figures 3.5 and S3.2.a and .b), guided by the inhibitor and product (PDB ID: 4H04 [94]) complexes. The reactive GlcNAc at subsite -1 , was in 4C_1 conformation in the initial docking pose, however it evolved to a distorted conformation between ${}^{1,4}B$ and 1S_3 during classical MD simulations [30] (Figure S3.2.c). These conformations are stabilized by the substrate-enzyme interactions at the -2 and -1 subsites observed in the crystal structures described in section 3.1, and they position the scissile glycosidic bond and leaving group an axial orientation. This provides space at the GlcNAc anomeric carbon for the hypothesized intramolecular nucleophilic substitution, a key feature in many GH enzymes [16, 38, 54]. During classical MD simulations, the two

catalytic residues (assistant Asp320 and acid/base Glu321) were in an orientation suitable for the neighboring group participation mechanism (Figure 3.5).

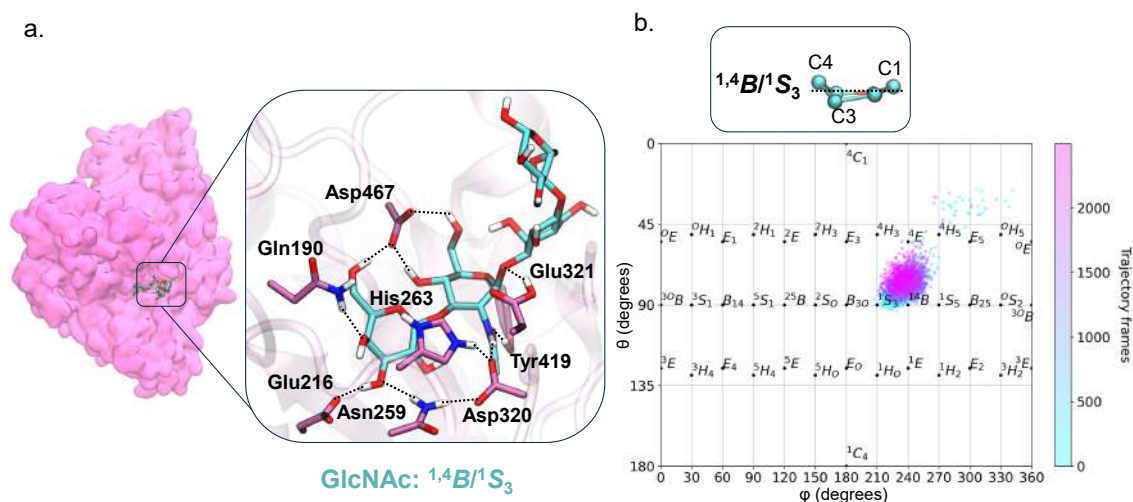


Figure 3.5: a. Active site of the Michaelis complex of LnbB-LNT after classical MD simulations. b. The conformation of the reactive GlcNAc at subsite -1 evolves toward a distorted $1S_3/1,4B$ during MD simulations.

3.2.2 First reaction step

After obtaining a suitable snapshot from classical MD simulations, we ran 2 ps of unbiased QM/MM MD. During this simulation, the saccharide moieties at subsites -1 and -2 established interactions with active site residues, including the catalytic assistant Asp320 with the 2-NHAc group H_{NAc} . The general acid/base Glu321 is close to the solvent and forms a weaker hydrogen bond interaction with the glycosidic bond oxygen. Thus, two conformations were sampled for Glu321: interacting with the scissile bond oxygen ($O_{Glu321}-O1$) and with the 2-OH substituent of Gal at subsite +1 (Figure S3.5).

After the QM/MM MD simulation, we modeled the first of the two-step substrate-assisted mechanism, where the bicyclic oxazoline or oxazolinium ion intermediate is formed. Two collective variables (CVs) were used to drive the system from the reactant state (the Michaelis complex) to the reaction intermediate. The CVs combine the interaction distances of the atoms involved in the 2-NHAc intramolecular substitution and the proton donation by Glu321 to the leaving group. CV1 includes the distance for the GlcNAc anomeric carbon to the 2-NHAc carbonyl oxygen (O_{NAc}) and the scissile glycosidic bond distance, thus accounting for the glycosidic bond cleavage and covalent bond formation. CV2 describes the protonation of the leaving group, as it includes the OH linkage distance of Glu321 and the hydrogen bond

distance with the oxygen atom of the scissile glycosidic bond (O1). These CVs were able to discriminate between the reactants and reaction intermediate states (Figure S3.6.a), as required in metadynamics.

The reconstructed reaction free energy landscape (FEL, Figure 3.6.b) reveals that the system evolves concertedly from the Michaelis complex towards the reaction intermediate with a free energy barrier of 15.3 kcal mol⁻¹. This value is in agreement with the value estimated from the experimental reaction rate on LNT ($k_{cat}=42.1 \pm 0.8$ s⁻¹, $\Delta G \approx 15$ kcal mol⁻¹ applying the Transition State Theory [122]). The 3 FEL minima correspond to the Michaelis complex (MC in Figure 3.6.a), a non-reactive configuration of the MC (indicated as MC') and the reaction intermediate, I. In MC' the acid/base Glu321 is forming a hydrogen bond with the 2-OH exocyclic group of Gal at the +1 subsite instead of with the oxygen of the glycosidic bond (Figure 3.6.a), thus preventing its protonation. This interaction pattern is consistent with the observations during the unbiased QM/MM MD simulations of the Michaelis complex Figure S3.5).

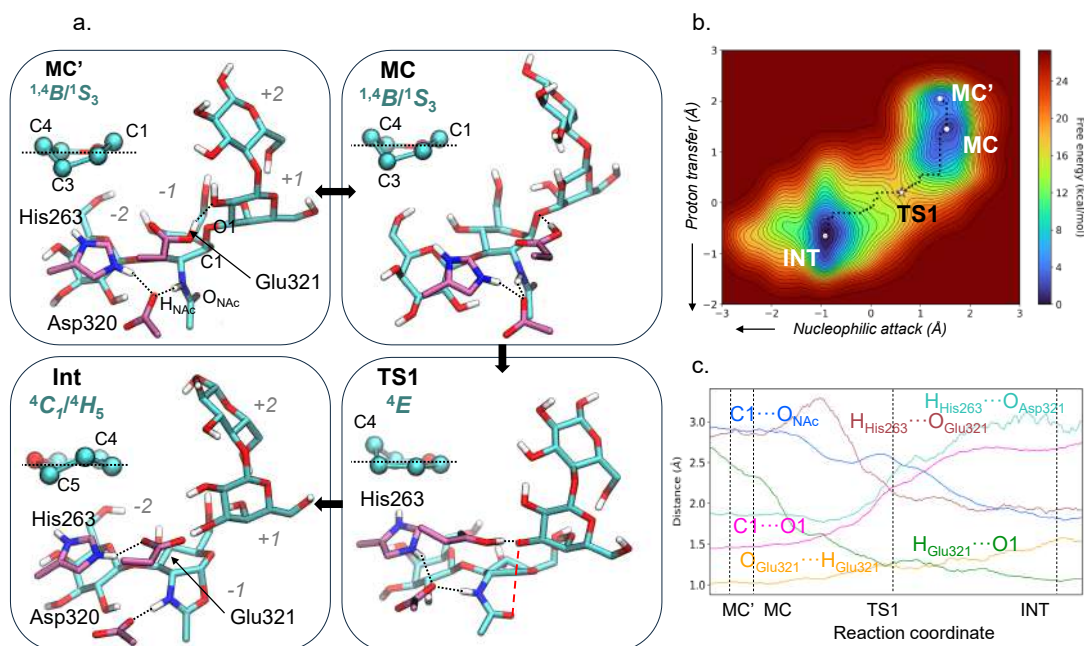


Figure 3.6: a. Representative structures of the explored states along the minimum energy pathway of the reaction FEL. b. Reconstructed FEL of the first reaction step, driven by two CVs corresponding to combinations of the distances involved in bond cleavage and formation in the two main events of the reaction (Nucleophile attack, CV1: $d[C1 - O_{NAC}] - d[C1 - O1]$, proton transfer, CV2: $d[(O - H)_{Glu321}] - d[H_{Glu321} - O1]$. Isolines are drawn every 1 kcal mol⁻¹. c. Evolution of relevant interatomic distances along the reaction coordinate, which corresponds to the MEP.

To obtain an atomistic view of the enzyme active site dynamics during the reaction coordinate, we analyzed the enzyme-saccharide interactions along the minimum

free energy pathway (MEP) (Figure 3.6.c). The reaction initiates with the elongation of the scissile glycosidic bond and the approach of the carbonyl oxygen (O_{NAc}) of the -1 GlcNAc 2-NHAc towards the anomeric carbon (C1), while Glu321 transfers a proton (H_{Glu321}) to the oxygen of the glycosidic bond (O1). The first reaction step takes place through a single transition state, TS_1 . The oxocarbenium ion-like nature of TS_1 is evidenced in the planarity of the reactive GlcNAc pyranose ring, which adopts a 4E conformation. At TS_1 the glycosidic bond is already significantly elongated (C1-O1: 1.99 ± 0.15 Å, Table 3.1), while the covalent bond between O_{NAc} and C1 is not yet formed (C1- O_{NAc} : 2.59 ± 0.13 Å), thus showing the dissociative character of the reaction. The C1-O5 distance shrinks, which is coherent with the development of a partial double bond at TS_1 (Figure 3.6.a, Table 3.1, MC: 1.43 ± 0.04 Å vs. TS_1 : 1.33 ± 0.04 Å). Regarding the protonation of the leaving group, it takes place concertedly with the glycosidic bond cleavage (Figure 3.6.c), as at TS_1 the proton of Glu321 is partially transferred (H_{Glu321} -O1: 1.39 ± 0.05 Å; H_{Glu321} - O_{Glu321} : 1.14 ± 0.04 Å, Table 3.1).

Distance (Å)	R (${}^1S_3/{}^1,4B$)	State	
		TS_1 (4E)	INT (${}^4C_1/{}^4H_5$)
C1-O5	1.43 ± 0.04	1.33 ± 0.04	1.32 ± 0.03
C1- O_{NAc}	2.89 ± 0.12	2.59 ± 0.13	1.85 ± 0.09
C1-O1	1.47 ± 0.05	1.99 ± 0.15	2.70 ± 0.09
H_{Glu321} - O_{Glu321}	1.01 ± 0.03	1.14 ± 0.04	1.69 ± 0.06
H_{Glu321} -O1	2.33 ± 0.18	1.39 ± 0.05	1.02 ± 0.02
H_{His263} - O_{Asp320}	1.88 ± 0.11	2.04 ± 0.17	3.19 ± 0.67
H_{His263} - O_{Glu321}	2.84 ± 0.23	2.41 ± 0.27	1.93 ± 0.16
N_{NAc} - H_{NAc}	1.04 ± 0.02	1.02 ± 0.03	1.13 ± 0.08
H_{NAc} - O_{Asp320}	1.90 ± 0.12	1.83 ± 0.08	1.45 ± 0.14

Table 3.1: Main interatomic distances of the active site at the main states of the first displacement reaction (Michaelis complex MC, TS_1 and reaction intermediate Int)

At what point of the reaction the leaving group is protonated depends on its nature and microenvironment. Catalysis of substrates with *p*-nitrophenyl (*p*NP) or methylumbelliferyl (MeU) groups have been reported to feature late protonation of the leaving group, both after the TS or upon formation of the reaction intermediate [9, 45, 123, 124]. Leaving groups with lower pK_a than the carbohydrate moiety at the negative subsites (MeU $pK_a \approx 7$ [45, 123], *p*NP $pK_a = 7.50$ [123]) can undergo catalysis where the protonation might not even be necessary. In the case of *p*NP, as reported by Nin-Hill and Rovira in 2020 [9], an ion-pair interaction with a neighboring Arg stabilizes the *p*NP anion, thus rendering it a better leaving group.

In contrast, saccharide leaving groups have high pK_a values of the order of $pK_a \approx 12$ [125] and need the protonation by the acid/base catalyst in order to continue the reaction.

The characteristics described above make the first displacement step by LnbB a proton-assisted S_N2 reaction. Based on the relevant atomic motions during the reaction coordinate (see Figure 3.6.c), the reaction is described as D_NA_N according to the IUPAC nomenclature due to its dissociative character [30, 124, 126]. The conformational itinerary of the reactive sugar computed along the reaction coordinate is ${}^1S_3/{}^1,4B \rightarrow [{}^4E]^\ddagger \rightarrow {}^4C_1/{}^4H_5$ (Figure 3.6.a, insets and Figure S3.7.a), which confirms the hypotheses from structural studies [94, 95]. The reaction starts with the substrate in a distorted conformation that favors catalysis, proceeds through a planar transition state and finishes near the Northern pole of the Cremer-Pople sphere at the reaction intermediate.

The interaction patterns during the reactive process vary across the subsites of LnbB. The two negative subsites of LnbB, a feature so far unique of this enzyme in family GH20, provide several binding points for hydrogen bond interactions during cleavage of the glycosidic bond (Table 3.2). The active site residues forming interactions with the LNB moiety are Asp467, Glu216 and Asp320, Asn259 and Tyr419, therefore of both negatively charged and neutral nature (Figure 3.7.a). Glu190 is in a position occupied by larger residues in other GH20 enzymes [34, 103, 121] (Figure 3.3), so their active sites can only host one saccharide at the negative subsite. Additionally, the reactive 2-NHAc group is interacting with aromatic residues Trp373, Trp394 and Trp465, in a similar way as in other enzymes operating via substrate-assisted NHAc catalysis [94, 99, 113] (Figure 3.7.b). The lactose moiety, the leaving group at the end of the first reaction step, is solvent exposed (Figure S3.8). Therefore, subsites -2 and -1 provide several binding points evidencing the specificity LnbB for LNB (Gal- β -1,3-GlcNAc) motifs, while being more promiscuous for sugars bound to the GlcNAc moiety. This allows *Bifidobacterium bifidum* to degrade many different HMOs.

Distance (Å)	State		
	R ($^1S_3/^{1,4}B$)	TS ₁ (4E)	INT ($^4C_1/{}^4H_5$)
H _{Gln190} -O _{4Gal}	2.44 ± 0.20	2.68 ± 0.16	2.45 ± 0.36
O _{Glu216} -O _{3Gal}	1.68 ± 0.09	1.70 ± 0.09	1.74 ± 0.14
H _{Asn259} -O _{Asp320}	3.12 ± 0.20	2.97 ± 0.18	2.88 ± 0.20
O _{Asp467} -H _{4O_{GlcNAc}}	1.60 ± 0.07	1.63 ± 0.18	1.58 ± 0.11
O _{Asp467} -H _{6O_{Gal}}	3.90 ± 0.33	3.62 ± 0.59	3.48 ± 0.36
O' _{Asp467} -H _{6O_{GlcNAc}}	1.62 ± 0.09	1.58 ± 0.12	1.60 ± 0.09
H _{Tyr419} -O _{N_{Ac}}	1.85 ± 0.14	1.76 ± 0.15	1.99 ± 0.18

Table 3.2: Interaction distances between active site residues and the LNB moiety at the main states of the first displacement step (Michaelis complex MC, TS₁ and reaction intermediate INT)

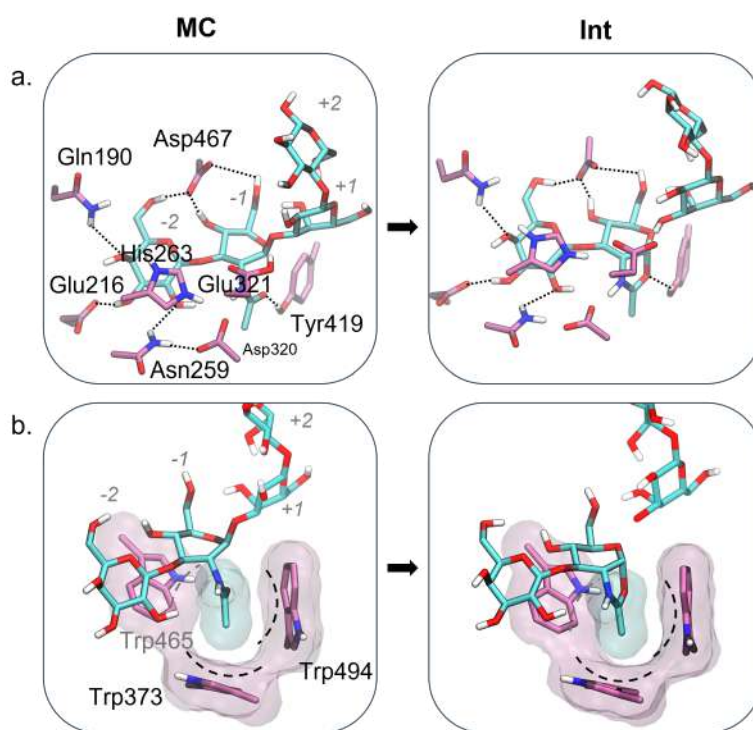


Figure 3.7: a. Active site residues that bind the LNB moiety at the -1 and -2 subsites of the enzyme, both at the beginning of the reaction (MC) and once the reaction intermediate is formed. b. Aromatic nest interactions around 2-NHAc at subsite -1 conserved around the bicyclic oxazolinium ion intermediate as well.

Residue His263 already raised questions when the crystal structure of LnbB was resolved, as mutating this position to Phe and Ala resulted in an activity loss of two orders of magnitude (see the Introduction of this Chapter) [94,95,97]. This residue however is not engaging in hydrogen bonds with the substrate, but rather interacting with the acid/base catalyst Glu321 in the structure. During the QM/MM metady-

namics simulation of the first reaction step, however, His263 changes its interaction pattern: it starts the reaction by forming a hydrogen bond with the assisting residue Asp320, then switches to interact with Glu321 upon formation of the reaction intermediates. To understand why this interaction switch may be favoring catalysis, we examined the interaction network formed between Asp320, His263, Glu321 and the substrate throughout the reaction. At the reactants state (MC) His263 interacts with Asp320, which is also forming a hydrogen bond with $H_{N_{Ac}}$ (Figure 3.6.a). By interacting with His263, the pK_a of Asp320 is lowered [127], which enhances the polarization of the 2-NHAc substituent by Asp320. The 2-NHAc in turn improves its nucleophilicity to perform the intramolecular substitution. When the cleavage of the scissile glycosidic bond starts, a proton is partially transferred from Glu321 to O1 and the residue eventually becomes negatively charged. This negative charge is stabilized by the newly formed interaction with His263, and the complex is ready for the second reaction step, in which Glu321 takes the role of general base for the hydrolysis [30].

The protonation state of the bicyclic reaction intermediate is an interesting question for substrate-assisted hexosaminidase catalysis [49, 128]. In these GHs, there are two possible forms for the reaction intermediate: neutral oxazoline or oxazolinium ion. Examples for both species have been reported in previous studies applying QM/MM metadynamics methods, like GH18 *SmChiB* (oxazoline) and GH84 *O*-GlcNAcases both bacterial and human (oxazolinium ion) [36, 52]. In our QM/MM metadynamics simulations, the most populated state is the oxazolinium ion, with a minor population of the oxazoline form [30]. We would expect the pK_a of Asp320 to rise after the first reaction step as His263 starts engaging with Glu321, the catalytic base for the second reaction step [129]. This pK_a rise would increase the population of the oxazoline form, which is indeed sampled during the metadynamics simulations of the first reaction step. However, the oxazolinium ion form is still the most populated at the reaction intermediate in our simulations. This is consistent with an additional QM/MM metadynamics simulation where we described the proton transfer from the 2-NHAc group to Asp320, where both states are almost isoenergetic (< 1 kcal mol⁻¹ difference) with a slight preference for the oxazolinium ion state [30] (Figure 3.8).

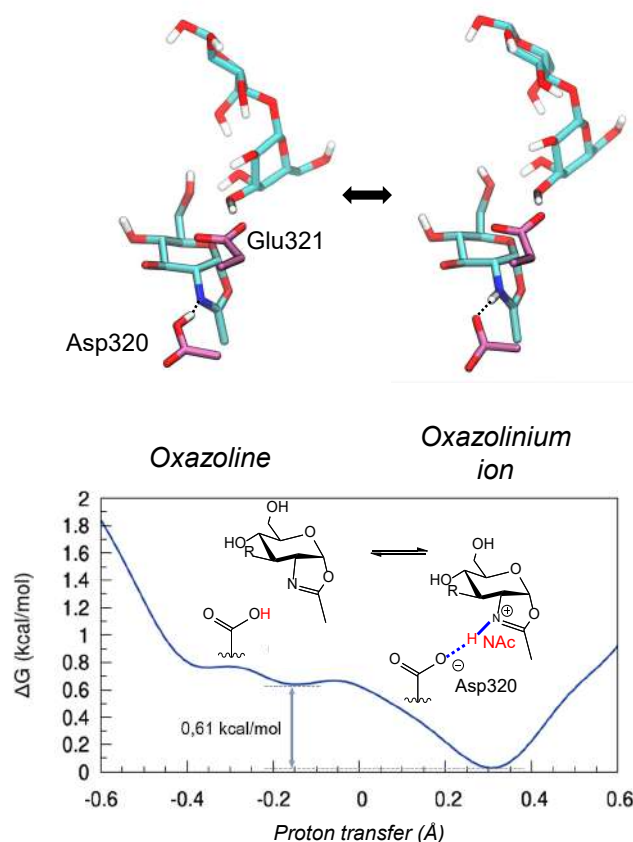


Figure 3.8: Free energy profile of the simulated proton transfer between Asp320 and the reaction intermediate to assess the relative stability of the charged oxazolinium ion and neutral oxazoline forms.

It is worth comparing the active site interactions that may be at play when favoring one intermediate type or the other. In GH18 *SmChiB*, the catalytic residues Asp142 and Glu144 interact directly and form a hydrogen network with the substrate, which favors the oxazoline intermediate (although the reaction progresses towards the second step via a transient oxazolinium ion, as its C-O bond is more labile [49]), whereas in GH84 *TtOGA* and GH20 *LnbB* the catalytic residues (Asp119 and Glu120, and Asp320 and Glu321, respectively) are not interacting directly, but rather the interaction network is facilitated by another residue (His263 in *LnbB* and Lys83 in *TtOGA*). Additionally, in 2023 a QM/MM metadynamics study of human GH84 *O*-GlcNAcase by Calvelo *et al.* showed similar results as the GH84 *TtOGA* and GH20 *LnbB*, with also a lysine (Lys98) interacting with the assisting residue Asp174 [36].

3.2.3 Modeling the reaction intermediate

To elucidate the molecular mechanism of the second reaction step, in which a water molecule attacks the reaction intermediate and the product is released, we first modeled the reaction intermediate complex. Our QM/MM simulations yielded structures of the reaction intermediate that are in agreement with the available structure of LnbB in complex with LNB-thiazoline, an analog for the reaction intermediate (PDB accession code 4JAW, Figure 3.3.a, Figures S3.1.b and S3.9) [94]. First, we performed classical MD simulations to study the solvent behavior around the active site. The second reaction step proceeds when a water molecule, activated by Glu321 acting as base, attacks the anomeric carbon of the bicyclic intermediate, the oxazolinium ring bond formed in the first step is broken and a proton is transferred back to the acid/base catalyst. Therefore, the presence of a water molecule near the active site and suitably oriented for catalysis (with one of its oxygen lone electron pairs oriented towards the anomeric carbon), is key in this reaction step. This occurrence was consistently observed in the MD simulations (Figure 3.9), a structure from which was chosen to continue to with QM/MM and metadynamics to eventually simulate the hydrolysis reaction. Two collective variables (CVs) defined as the difference of two interatomic distances were chosen to drive the bias. CV1 accounts for the oxazolinium covalent bond cleavage, as it was defined as the subtraction of the distance between its anomeric carbon to water oxygen and the anomeric carbon and five-member ring oxygen. The second represents the proton transfer from the catalytic water molecule to Glu321 (Figure S3.6.b).

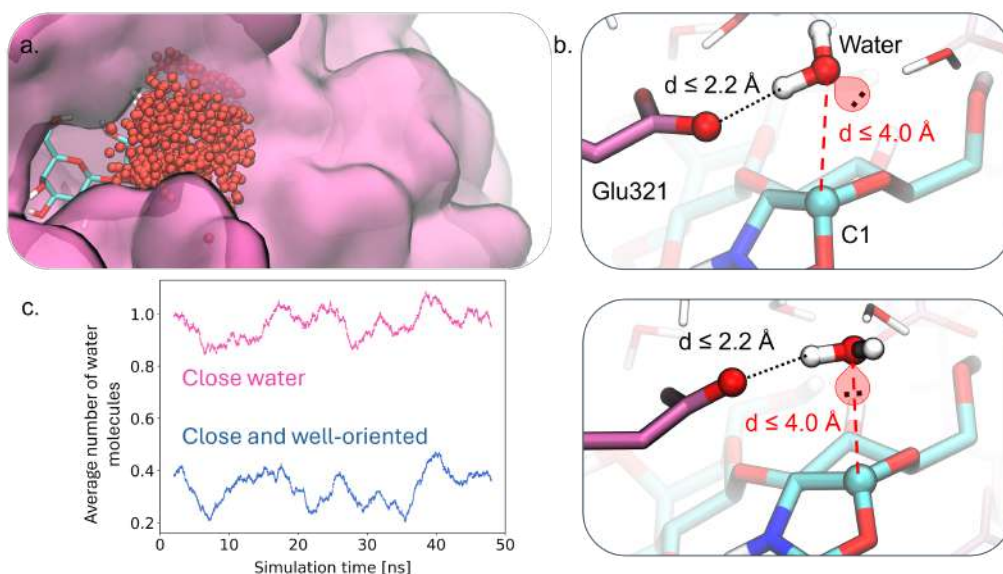


Figure 3.9: a. Representation of the active site of LnbB at the reaction intermediate. As red balls, superimposition water molecules within $4.5 \text{ \AA}$ of the anomeric center of the oxazolinium moiety of all frames of the 50 ns of classical MD trajectory. Note that the -1 subsite is fairly exposed, so water molecules can approach. b. Trajectory snapshots showing water molecules close to C1 ($\leq 4.0 \text{ \AA}$) and forming a H-bond with Glu321 ($\leq 2.2 \text{ \AA}$). In the bottom snapshot, one of the lone electron pairs of O_w is in line with C1. c. Running average of close and well-oriented water molecules along 50 ns of classical MD simulation.

3.2.4 Second reaction step

The metadynamics simulation drove the system from the state of reaction intermediate (without the leaving group, not present in the starting crystal structure) to the products, which are LnbB in complex with the LNB moiety, as represented by the reconstructed FEL (Figure 3.10, S3.6.b). The computed free energy barrier is $11.7 \text{ kcal mol}^{-1}$, which is $3.3 \text{ kcal mol}^{-1}$ lower than the barrier of the first reaction step ($15.3 \text{ kcal mol}^{-1}$). This indicates that this second reaction step is not rate-limiting, but rather the first, that leads to the formation of the oxazolinium ion intermediate. Our simulations allow characterizing the second step as a concerted, dissociative S_N2 reaction, as the oxazolinium ion opens before the water molecule forms the new covalent bond.

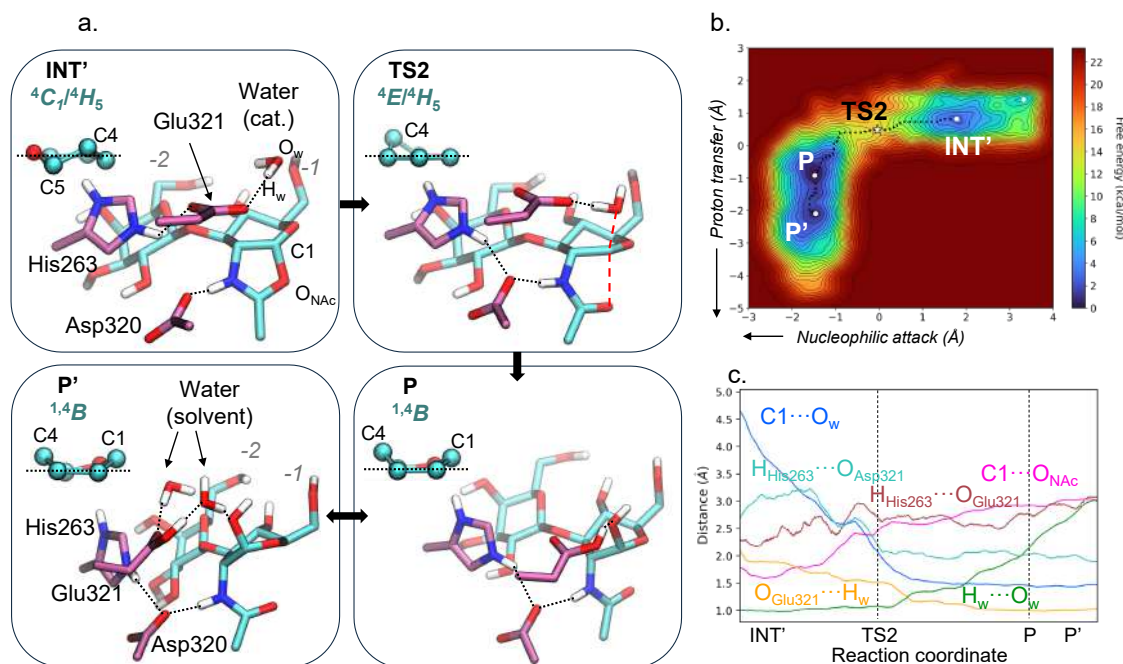


Figure 3.10: a. Representative structures of the main states of the minimum energy pathway of the second reaction step FEL. b. Reconstructed FEL of the second reaction step, simulation driven by two CVs corresponding to combinations of the distances involved in bond cleavage and formation in the two main events of the reaction (nucleophile attack, CV1: $d[\text{C1} - \text{O}_w] - d[\text{C1} - \text{O}_{\text{NAc}}]$, proton transfer, CV2: $d[\text{O}_{\text{Glu321}} - \text{H}_w] d[(\text{O} - \text{H})_w]$. Isolines are drawn every 1 kcal mol⁻¹. c. Evolution of relevant interatomic distances along the reaction coordinate, which corresponds to the minimum energy pathway of the reaction.

The hydrolysis step starts with the elongation of the C1-O_{NAc} bond, while a water molecule interacting with Glu321 approaches the anomeric carbon. The transition state, TS₂, exhibits oxocarbenium ion-like character. The conformational itinerary for this reaction step, ${}^4C_1/{}^4H_5 \rightarrow [{}^4E/{}^4H_5]^{\ddagger} \rightarrow {}^{1,4}B$ (Figure 3.10.a, insets and Figure S3.7.b), mirrors the itinerary from the first displacement and proceeds through a planar conformation, in agreement with the initial hypotheses and experimental data [94, 95]. This is supported by a detailed view on the FEL minima and minimum free energy pathway as well (Figure 3.10.b and c). At TS₂, the covalent bond C1-O_{NAc} is partially broken (see Figure 3.10.c, Table 3.3, C1-O_{NAc}: 2.35 ± 0.16 Å) whereas the bond with the catalytic water molecule is not yet formed (Figure 3.10.c, Table 3.3, C1-O_w: 2.30 ± 0.15 Å). The O5-C1 distance shrinks at TS₂ (INT': 1.34 ± 0.03 Å vs. TS₂: 1.29 ± 0.03 Å), indicating its partial double bond character. The water molecule H_w is transferred to Glu321 after TS₂ (3.10.a, Table 3.3, O_w-H_w: 1.07 ± 0.03 Å, O_{Glu321}-H_w: 1.53 ± 0.06 Å).

Distance (Å)	State		
	INT' (${}^4C_1/{}^4H_5$)	TS ₂ (${}^4E/{}^4H_5$)	P (1A_B)
C1-O _w	3.45 ± 0.20	2.30 ± 0.15	1.46 ± 0.04
C1-O _{N_{Ac}}	1.70 ± 0.11	2.35 ± 0.16	2.95 ± 0.10
O _{Glu321} -H _w	1.79 ± 0.14	1.53 ± 0.06	1.01 ± 0.03
H _w -O _w	1.01 ± 0.03	1.07 ± 0.03	1.85 ± 0.14
O _{Asp320} -H _{N_{Ac}}	1.65 ± 0.11	1.91 ± 0.13	2.14 ± 0.22
N _{N_{Ac}} -H _{N_{Ac}}	1.07 ± 0.04	1.05 ± 0.03	1.03 ± 0.03
C1-O5	1.34 ± 0.03	1.29 ± 0.03	1.41 ± 0.03
H _{Tyr419} -O _{N_{Ac}}	2.05 ± 0.19	1.96 ± 0.14	1.86 ± 0.14
H _{His263} -O _{Asp320}	3.18 ± 0.47	2.15 ± 0.34	2.08 ± 0.24
H _{His263} -O _{Glu321}	2.49 ± 0.84	2.69 ± 0.38	2.66 ± 0.40

Table 3.3: Main interatomic distances of the active site at the main states of the hydrolysis reaction (reaction intermediate without the leaving group INT', TS₂ and products P state)

Following the molecular motions along the reaction coordinate gives us a detailed picture on the enzyme-substrate interactions at subsite -1. During the second reaction step, the saccharide moiety is tightly bound to the enzyme negative subsites by engaging with the active site residues (Asp467, Glu216, Asp320, Asn259, Tyr417 and the three aromatic Trp373, Trp394 and Trp465 around the oxazolinium ion ring) (see Supplementary Table S3.1). Regarding residue His263, it is interesting to note that it changes partners again during the second reaction step, and the pre-catalytic interaction network is recovered. While at the reaction intermediate His263 is forming a hydrogen bond with the acid/base Glu321, it moves and interacts again with the assisting residue Asp320 when a proton is transferred from the catalytic water molecule to Glu321 (Figure 3.10.a). The motion of His263, together with the proton transfer to Glu321, recovers the active site configuration of the enzyme-substrate complex prior to catalysis. After product release the turnover is then complete.

Our simulations of both first and second reaction steps of the LnbB catalytic mechanism revealed the molecular details that could explain the role of His263 in LnbB catalysis [30, 94]. By engaging in hydrogen bond interactions with the assistant residue then the acid/base as the reaction progresses, His263 modulates the pK_a cycle of the catalytic pair. Therefore, it is apparent from the experimental data and the simulations that His263 has a role beyond substrate recognition and plays a crucial part in the catalytic mechanism. The pK_a cycle of the catalytic pair has long been described in retaining GHs [36, 49, 129]. In their active sites, close to the catalytic pair there is a residue whose function is to modulate the pK_a of the acid/base catalyst, which decreases when the reaction intermediate is formed and in-

creases again upon product formation. This phenomenon is shared by non-classical retaining GHs, like those operating via substrate-assisted mechanism commented in this chapter and the identity of this residue varies throughout GH families. As suggested in previous work by Dr. Joan Coines, the position in the sequence of the catalytic pair [49] seems to be a determining factor in the active site architecture, which then determines the interaction network of the catalytic residues. For example, and as mentioned above, *Sm*ChiB has non-consecutive catalytic residues in the sequence that interact directly throughout the catalytic process, whereas GH84 OGA and GH20 LnbB catalytic pairs are consecutive and never interact directly, therefore an additional residue is the pK_a modulator (Lys98, His263). Since the His263 position is conserved across family GH20, the results of our simulations can be extended to other GH20 members and could be a position to consider in enzyme engineering strategies.

3.3 Summary and conclusions

In this chapter, we have studied the two-step catalytic mechanism of the GH20 enzyme lacto-*N*-biosidase, LnbB, with the natural substrate LNT. Starting from the enzyme 3D structure in complex with the mechanism inhibitor LNB-thiazoline, we first built an enzyme-substrate complex model with LNT, then a model of the oxazolinium ion reaction intermediate. We started from the hypothesis that the enzyme operates with a substrate-assisted mechanism via a 2-NHAc substituent as the nucleophile of the first reaction step. The experimental structures show the bicyclic intermediate analog characteristic of this mechanism. However, to the best of our knowledge, our model provided the first all-atom structure of the LnbB Michaelis complex with a natural substrate. Our simulations unveiled the details of the enzyme catalytic mechanism at an atomistic level, in good agreement with experiments, which revealed that a histidine residue at the active site, previously identified as crucial for activity in mutagenesis and kinetic experiments, moves during catalysis and induces a pK_a modulation of the catalytic residue pair, facilitating catalysis. Also, our simulations described the full conformational itinerary of the pyranose ring, revealing that it is in a distorted 1,4B conformation at the Michaelis complex, which was previously not known. Additionally, we characterized the reaction intermediate as oxazolinium ion, with a minor population of the neutral oxazoline.

This study, then, reveals with QM/MM MD and metadynamics the mechanism of degradation of one of the main building blocks of HMOs, therefore shedding light on how a central piece of the human gut microbiome works at the beginning of life.

These insights can open new doors to enzyme engineering strategies for the enzymatic synthesis of HMOs for improving formula milk as well as add understanding other members of the GH20 family and substrate-assisted operating enzymes.

The conclusions of this Chapter are as follows:

- LnbB operates via a two-ste substrate-assisted mechanism to degrade its natural substrate LNT via an intramolecular substitution by the 2-NHAc exocyclic group of GlcNAc, where the first is the rate-limiting step of the reaction.
- The residue His263 is directly involved in the reaction mechanism as pK_a modulator of the Asp320/Glu321 catalytic pair by alternating interactions with them along the reaction. This can explain the impact of this residue in enzyme activity.
- The Michaelis complex of LnbB features a distorted conformation of the reactive saccharide, delineating a ${}^1S_3/{}^{1,4}B \rightarrow [{}^4E]^\ddagger \rightarrow {}^4C_1/{}^4H_5$ (first step) and ${}^4C_1/{}^4H_5 \rightarrow [{}^4E/{}^4H_5]^\ddagger \rightarrow {}^{1,4}B$ (second step) itinerary at the Northern Hemisphere of the Cremer-Pople sphere.
- The reaction intermediate is best described as a charged oxazolinium ion, with a minor population of neutral oxazoline.

3.4 Computational details

The systems modeled in this study were LnbB in complex with the natural substrate LNT (Michaelis complex), as well as and LnbB in complex with the disaccharide LNB-oxazolinium ion (the bicyclic reaction intermediate). Each of the models were employed to eventually simulate the first and second displacement reactions of the LnbB substrate-assisted mechanism, respectively. For all simulations, trajectory analyses were performed with VMD version 1.9.3 (<http://www.ks.uiuc.edu/Research/vmd/>) [130], the CPPTRAJ [131] module of AMBER18 [132, 133].

3.4.1 Model of the Michaelis complex and first reaction step

3.4.1.1 System preparation

To build the model we used as initial coordinates the crystal structure of *Bifidobacterium bifidum* LnbB in complex with LNB-thiazoline (PDB ID 4JAW at www.rcsb.org [94,134], resolution 1.8 Å), considering the first protomer only, residues 40 to 662. The first residues of the construct (30 to 39, whose function is to assist the structure packing and are added as part of the cloning vector in the protein expression step) were excluded from the model, as well as the LNB-thiazoline ligand and sulfate counter ions. The protonation state of protein residues was assigned with the assistance of the programs PropKa3 and H++ [135–137] and inspecting the environment of each residue at the optimal pH for LnbB activity, 4.5 [94]. Regarding the catalytic residues, Asp320 was modeled in its carboxylate form and the acid/base Glu321 was modeled as protonated. LNT was docked in the active site using the program AutoDock VINA [138,139] (Figure S3.2). The docked poses were provided by our collaborators at the Laboratory of Biochemistry from Institut Químic de Sarrià, Dr. Xevi Biarnés, Dr. Magda Faijes and Prof. Dr. Antoni Planas. In the best pose, the substrate was engaged in interactions with the residues of the -2 and -1 subsites. The system was solvated in a cell of dimensions $105.15 \times 98.89 \times 107.20$ Å including 33868 water molecules and 10 Cl⁻ anions to achieve neutrality.

3.4.1.2 Classical molecular dynamics

Classical molecular dynamics (MD) simulations of the system LnbB-LNT were performed using the AMBER package, version 18 [133]. The employed forcefields were FF99SB (protein) [140], GLYCAM06 (LNT) [64] and TIP3P (solvent) [141]. The protocol for the forcefield simulations was carried out as follows: first an energy minimization of the solvent box was performed (first with steepest descent then conjugated gradients schemes), followed by a global minimization. Next, the system

was heated to a temperature of 300 K by increasing 100 K over consecutive runs of 50 ps with a timestep of 1 fs (solvent only to 100 K first, then entire system to 100 K, 200 K and 300 K). Then a 2 ps run in the NPT ensemble was performed to equilibrate the system density to 1 g cm⁻¹ at 1 bar with the Berendsen barostat [142]. The simulation was then continued in the NVT ensemble for 50 ns (Figure S3.3.a). Runs in both NPT and NVT ensemble used a timestep of 2 fs, applying the SHAKE algorithm [143] to constrain the bonds with hydrogen atoms. Potential wall restraints were placed at active site interaction distances to assist in the stabilization of LNT at the active site cavity, and they were smoothly released by decreasing the force constant during the production run.

3.4.1.3 QM/MM molecular dynamics

Hybrid QM/MM MD simulations were computed using the CPMD code, version 3.15.1 (<http://www.cpmc.org>), which combines Density Functional Theory-based Car-Parrinello molecular dynamics [75] (for the chemically active QM part) with the AMBER potential energy function (for the MM part) developed by Laio, Vandevondele and Roethlisberger [78,144], starting from a trajectory snapshot extracted from the classical MD of the LnbB-LNT complex. 67 atoms were selected for the QM region and were enclosed in an orthorhombic cell of dimensions 15.09 × 19.84 × 18.52 Å. The atoms in the QM region include the reactive GlcNAc at subsite -1, the sidechains of the catalytic residues (Asp320 and Glu321), part of the Gal units at subsites -2 and +1 and 6 monovalent atoms included to saturate the C-C bonds separated by the QM-MM boundary (Figure S3.4.a). The MM region includes the rest of the system (111345 atoms). To assess whether His263 could transfer one of its protons to Asp320 at the Michaelis complex, a QM region including the sidechain of His263 was also built (box dimensions: 19.37 × 16.62 × 20.39 Å). The method combining Car-Parrinello MD with DFT developed by Laio *et al.* [78] was applied using the Perdew-Burke Ernzerhof (PBE) generalized gradient-corrected approximation [71]. This functional is appropriate, as tested by Marianski *et al.* [73], for carbohydrate conformations [16,145] and glycoside hydrolase systems [9,29,36,52], as well as other carbohydrate-processing enzymes [146,147]. For the charges near the QM/MM boundary the fully Hamiltonian multilayer electrostatic coupling scheme developed by Laio *et al.* was used [78], adjusting the radii at 9.52 Å (NN), 11.11 Å (MIX) and 15.87 Å (ESP), after scanning values ranging from 5.29 Å to 105.84 Å to assess which values yield an accuracy of the order of 1 kcal mol⁻¹. The kinetic energy cutoff for the expansion of Kohn-Sham orbital in the plane-wave basis set was set at 70 Ry. Norm-conserving *ab initio* Trouiller-Martins pseudopotentials were used for all elements [74]. The fictitious electron mass was set at 600 au. After

optimization of the QM/MM with a simulated annealing of the atomic velocities, the system was heated to 300 K with the Nosé-Hoover thermostat with a frequency of 2500 cm^{-1} [148] during a 5 ps unbiased QM/MM MD run with a 0.12 fs timestep, after which the metadynamics simulations started.

3.4.1.4 QM/MM metadynamics

The first displacement reaction, from the Michaelis complex of LnbB-LNT to the formation of LNB-oxazolinium was modeled with of QM/MM metadynamics. To apply the metadynamics scheme to QM/MM MD, the open-source, community-developed PLUMED library [149], version 2.3.3 [150] was coupled to CPMD. A history-dependent bias potential built to enhance the exploration of the reaction free energy landscape was driven by two collective variables (CVs) able to discriminate the different states along the reaction coordinate. The two CVs account for the linkages broken and formed during the reaction and were designed as differences of distances as in previous simulations of GH mechanisms [38,49,79]. The first CV, $d(\text{C1-O}_{N\text{Ac}}) - d(\text{C1-O1})$, represents the 2-NHAc nucleophilic attack and glycosidic bond cleavage and the second CV, $d(\text{O1-H}_{\text{Glu321}}) - d(\text{H}_{\text{Glu321}}-\text{O}_{\text{Glu321}})$, accounts for the proton transfer by the acid/base residue to the departing group (Figure S3.6.a, top). The CV space exploration was carried out by adding Gaussian-like biasing potential functions every 40.8 fs (340 MD simulation steps, in-tune with the motion of the slowest CV). Gaussian height was set at 1.0 kcal mol^{-1} and the width at 0.2 Å for both CVs. The simulation was considered as finished when a recrossing event from the intermediates back to the initial state was observed. This is the recommended approach for QM/MM MD metadynamics simulations of enzyme reactions [151]. At that point, 528 Gaussian functions were accumulated in approximately 22 ps of simulation (S3.6.a, bottom). The minimum free energy pathway (MEP) of the FEL was computed with the program MEPSAnd [152], and the program to order the trajectory frames along the MEP was created by the former group member Dr. Oriol Esquivias [30].

An additional QM/MM metadynamics simulation was performed to explore the proton transfer between the 2-NHAc of the reaction intermediate and the assisting Asp320. For this, one CV was employed, $d(\text{O}_{\text{Asp320}}-\text{H}_{N\text{Ac}}) - d(\text{H}_{N\text{Ac}}-\text{N}_{N\text{Ac}})$. Gaussian-like functions were deposited every 15.6 fs (130 MD simulation steps), with a height of 0.1 kcal mol^{-1} and 0.1 Å , respectively.

3.4.2 Model of the reaction intermediate and second reaction step

3.4.2.1 System preparation

The starting point of the LnbB-reaction intermediate complex model was the crystal structure with PDB ID 4JAW (LnbB in complex with the inhibitor LNB-thiazoline) [94]. The same residues from the crystallographic construct were used as in the model of the enzyme-substrate complex (see Section 3.4.1.1). In this model the sulfur atom at LNB-thiazoline five-membered ring was changed to oxygen to obtain LNB-oxazolinium ion, which was the only difference between the two molecules. The protonation states for this model were also assigned with the assistance of pK_a prediction programs [135–137] and visual inspection of the enzymatic environment at pH 4.5. Since at the state of reaction intermediates the acid/base catalyst is hypothesized to extract a proton from a water molecule to complete the hydrolysis reaction, Glu321 was modeled in its carboxylate form. Regarding the protonation state of the reaction intermediate, the form LNB-oxazolinium ion with carboxylate Asp320 was selected. For the forcefield MD simulations of the LnbB-intermediate complex the system was solvated in a cubic cell (dimensions $111.01 \times 104.40 \times 113.17$ Å) with 33883 water molecules and 9 Cl^- ions to achieve neutrality.

3.4.2.2 Classical molecular dynamics

The AMBER package [133] was used for all simulations with the forcefields FF14SB (enzyme) [153], GLYCAM06 (saccharide) and TIP3P (water) [64, 141]. Since the parameters for Glc-oxazolinium are not included in GLYCAM, they were obtained by calculating RESP charges with Gaussian version 09 [154] at the HF/6-31G* level of theory, then obtaining parameters with the Antechamber [155] suite of AMBER using GLYCAM_06j and GAFF [156]. The protocol for the classical MD simulations was performed as described in Section 3.4.1.2 (Figure S3.3.b). To increase the sampling of potential catalytic molecules near the active site, we performed 6 independent replicas of 50 ns each. The behavior of waters at the active site during MD simulations was thoroughly analyzed using an in-house Python program designed by Dr. Oriol Esquivias [30]. To shortlist suitably positioned water molecules, cutoffs of 4.0 Å for the $\text{O}_w\text{-C1}$ distance and 2.2 Å for $\text{H}_w\text{-O}_{\text{Glu321}}$ were applied. With this procedure a trajectory snapshot was selected to continue with the QM/MM MD protocol.

3.4.2.3 QM/MM molecular dynamics

A QM region of 56 atoms were enclosed in an orthorhombic cell with dimensions $16.50 \times 19.97 \times 15.39$ Å. The QM region includes the saccharide at subsite -1 and part of the Gal unit at subsite -2, the sidechains of the catalytic residues (Asp320/Glu321), a water molecule positioned at a suitable orientation for catalysis and 4 monovalent atoms used to cap the valences of the C-C bonds divided by the QM-MM interface (Figure S3.4.b). The MM region contains the remaining 111354 atoms of the system. The method used for treating the charges near the QM-MM boundary region is the same as explained in section 3.4.1.3, with radii for the electrostatic coupling areas of 8.46 Å (NN), 11.64 Å (MIX) and 15.34 Å (ESP). The fictitious electronic mass was set at 600 au. A 4 ps unbiased QM/MM MD run was performed with a timestep of 0.12 fs and keeping the temperature at 300 K with a Nosé-Hoover thermostat [148].

3.4.2.4 QM/MM metadynamics

The QM/MM metadynamics simulation of the second displacement step was driven using two CVs representing the main bond cleavage and formation events taking place. Both CVs were defined as the subtraction of interaction distances: The first CV includes the attack distance between the catalytic water molecule and the anomeric carbon and the opening of the oxazolinium ring, $d(O_w-C1) - d(C1-O_{NAc})$, and the second CV accounts for the proton transfer from the reactive water molecule to the acid/base catalyst, now acting as general base, $d(O_{Glu321}-H_w) - d(H_w-O_w)$. Gaussian-like kernels were deposited every 45.6 fs of the MD simulation (380 MD steps) with an initial height of $1.0 \text{ kcal mol}^{-1}$ (lowered to $0.5 \text{ kcal mol}^{-1}$ near the crossings over the TS) and a width of 0.2 Å for each variable. The first recrossing criterion was applied to discern the end of the simulation.

3.5 Supplementary information

Distance (Å)	State		
	INT' (${}^4C_1/{}^4H_5$)	TS ₂ (${}^4E/{}^4H_5$)	P (${}^1,{}^4B$)
H _{Gln190} -O4 _{Gal}	2.94 ± 0.23	2.89 ± 0.15	2.85 ± 0.23
O _{Glu216} -O3 _{Gal}	1.69 ± 0.12	1.68 ± 0.09	1.70 ± 0.09
H _{Asn259} -O2 _{Gal}	2.68 ± 0.33	2.68 ± 0.22	2.67 ± 0.19
O _{Asp467} -H6O _{Gal}	1.66 ± 0.11	1.77 ± 0.12	1.68 ± 0.14
O _{Asp467} -H4O _{ox}	1.58 ± 0.09	1.57 ± 0.08	1.62 ± 0.10
O' _{Asp467} -H6O _{ox}	1.59 ± 0.09	1.55 ± 0.07	1.59 ± 0.09
H _{Tyr419} -O _{NAc}	2.05 ± 0.19	1.96 ± 0.14	1.86 ± 0.14

Table S3.1: Interaction distances between atoms of active site residues at the -1 and -2 subsites and selected atoms of the substrate at the main states of the second displacement step (reaction intermediate Int', TS₂ and products)

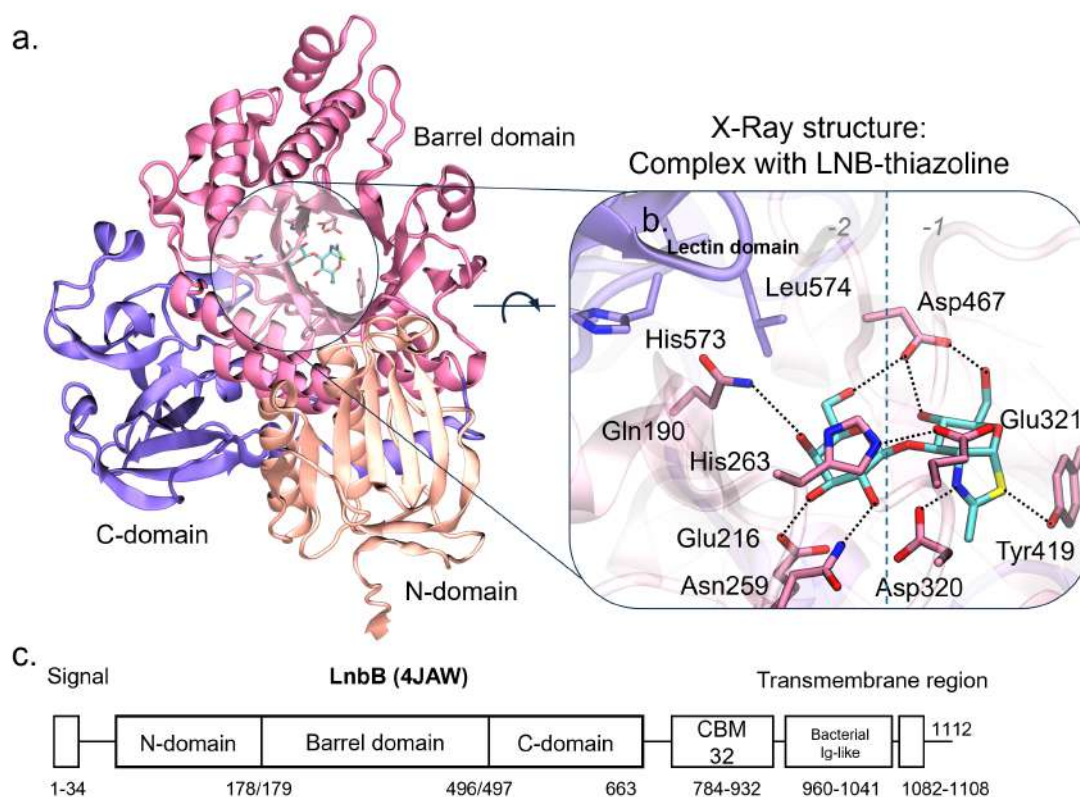


Figure S3.1: a. Tridimensional structure of LnbB in complex with LNB-thiazoline (PDB 4JAW) showing the three domains in purple (C-domain), mauve (barrel domain) and orange (N-domain). b. Active site of the crystal structure of LnbB. In purple, lectin domain close to the active site reported to be crucial in [93]. This structure is the starting point for both the models of the Michaelis complex (with LNT) and for the model of the reaction intermediate. c. Scheme representation of the domains of LnbB. The structure construct includes until residue 663. Adapted from [94].

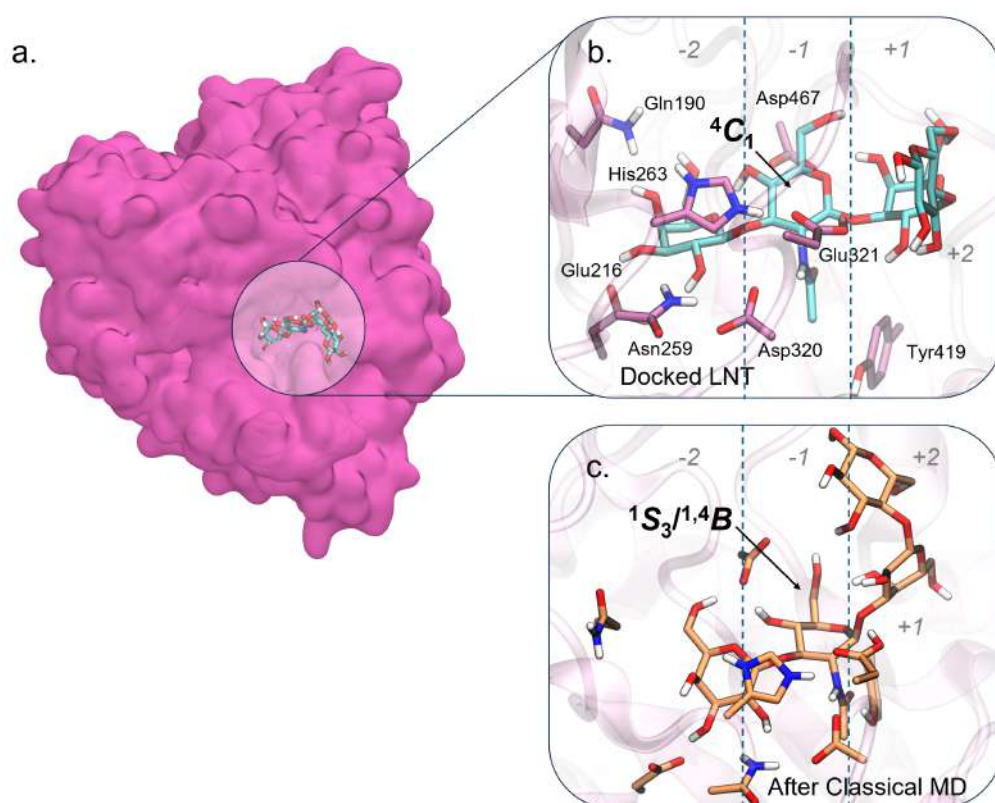


Figure S3.2: a. LnbB enzyme-substrate complex with LNT (cyan stick representation). b. Active site of the model before classical MD simulations (cyan) and c. After MD (orange). The substrate accommodates in the active site after classical MD and forms the corresponding interactions with the active site residues. The conformation of GlcNAc at subsite -1 evolves from 4C_1 to ${}^1S_3/{}^1,4B$.

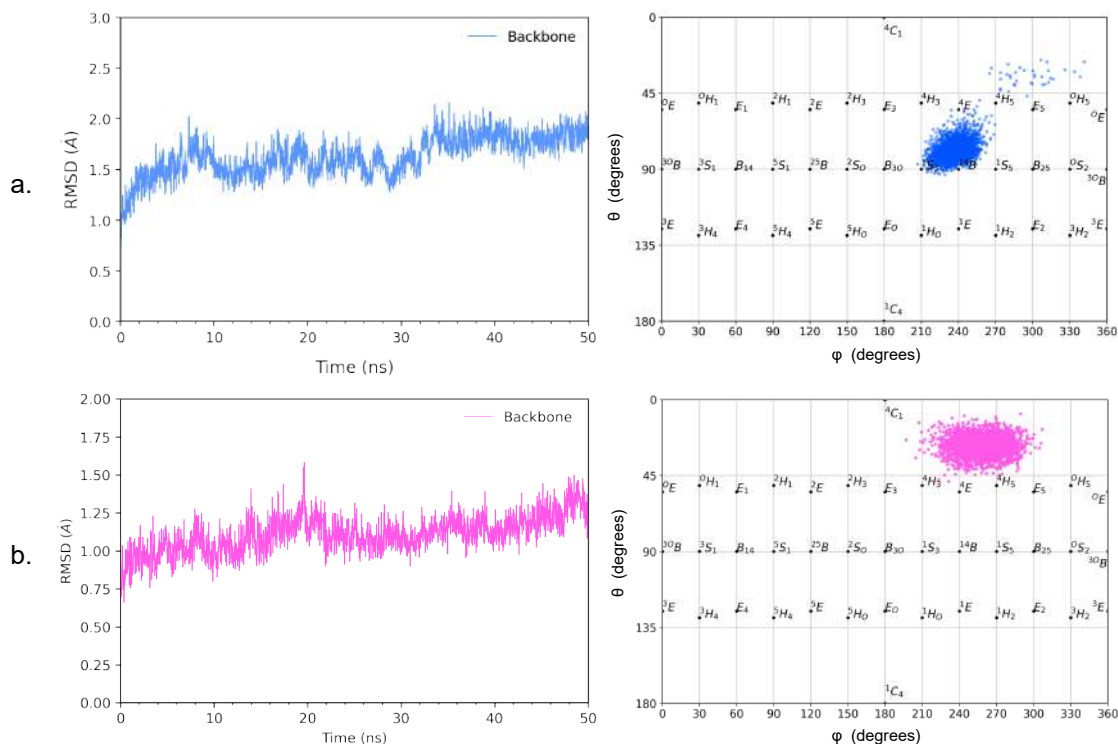


Figure S3.3: RMSD evolution and Mercator representations of the sugar ring at subsite -1 during the production run of classical MD simulations of a. the Michaelis complex and b. the reaction intermediate.

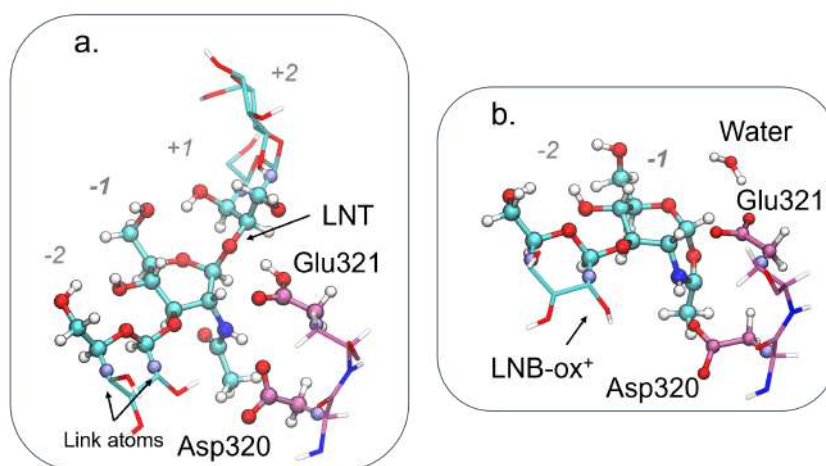


Figure S3.4: Atoms included in the QM regions for the QM/MM MD simulations of the a. Michaelis complex and b. Reaction intermediate. In cyan, the corresponding saccharides, mauve, the catalytic residues Asp320 and Glu321, violet, the link *dummy* atoms used to cap the valences of the C-C bonds crossed by the QM-MM region interface. Ball representations, QM atoms. Stick representation, MM atoms.

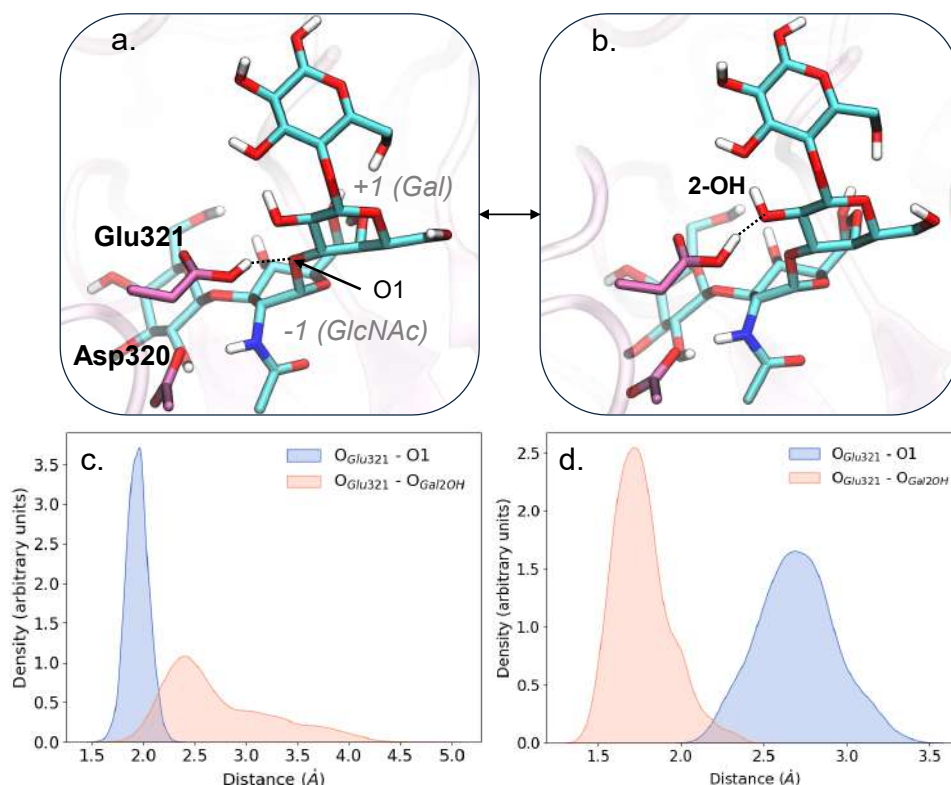


Figure S3.5: Two possible conformations of the acid/base Glu321 sidechain sampled in QM/MM MD simulations. a. Forming a H-bond interaction with O1 and b. Interacting with the nearby Gal moiety at the +1 subsite 2-OH. c. Kernel density estimation (KDE) plots of the classical MD trajectory showing that, although Glu321 forms a H-bond with O1, it also interacts with the Gal 2-OH substituent. d. Same representation during the QM/MM MD trajectory showing that the H bond with Gal 2-OH is stronger. This is coherent with the MC' minimum visited along the MEP during the QM/MM metadynamics simulation (Figure 3.6) and corresponds to a pre-catalytic configuration of the active site.

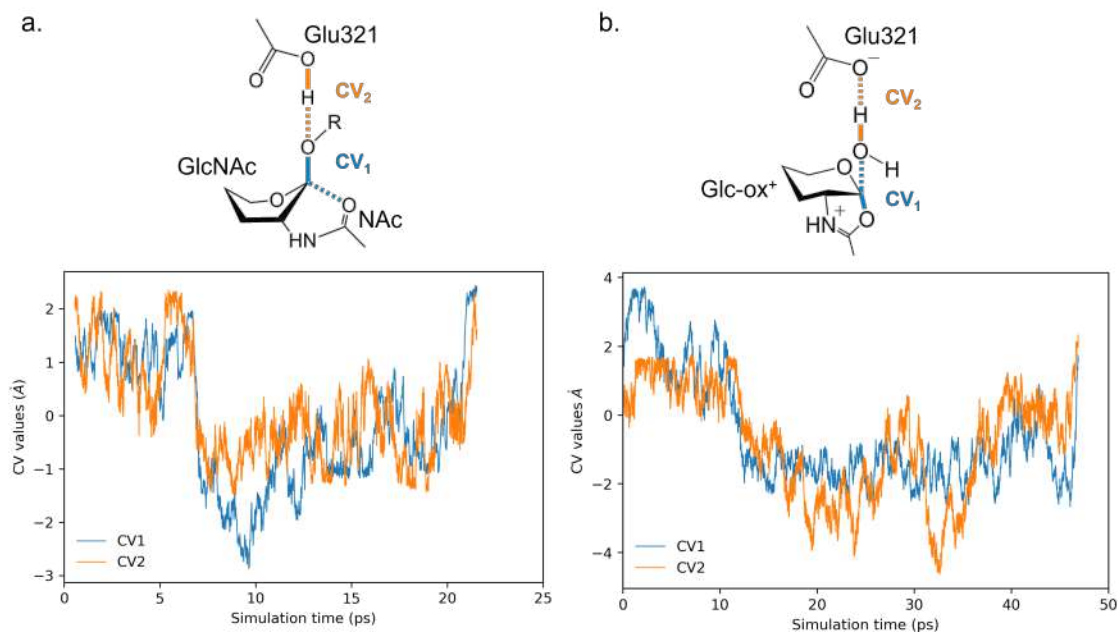


Figure S3.6: a. Top: Collective variables chosen to drive the bias in the first reaction step. CV1 (blue) corresponds to the intramolecular nucleophilic substitution by the 2-NHAc group GlcNAc and CV2 (orange) to the proton transfer by the general acid catalyst Glu321. Bottom: Evolution of CV1 and CV2 during the QM/MM metadynamics simulation until the first recrossing criterion was fulfilled. b. Top: Collective variables for the QM/MM metadynamics simulation of the second reaction step. In this reaction, a water molecule is included in the biased CVs. Bottom: CVs evolution during the simulation of the second reaction step.

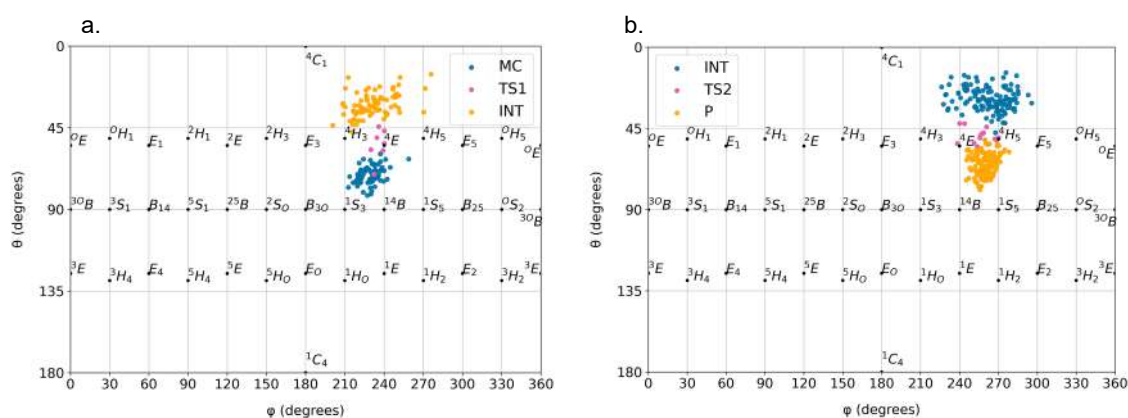


Figure S3.7: Mercator representations of the -1 sugar ring of ensembles of structures taken from the main FEL minima of a. The first reaction step (MC, TS₁, INT) and b. The second reaction step (INT', TS₂, P).

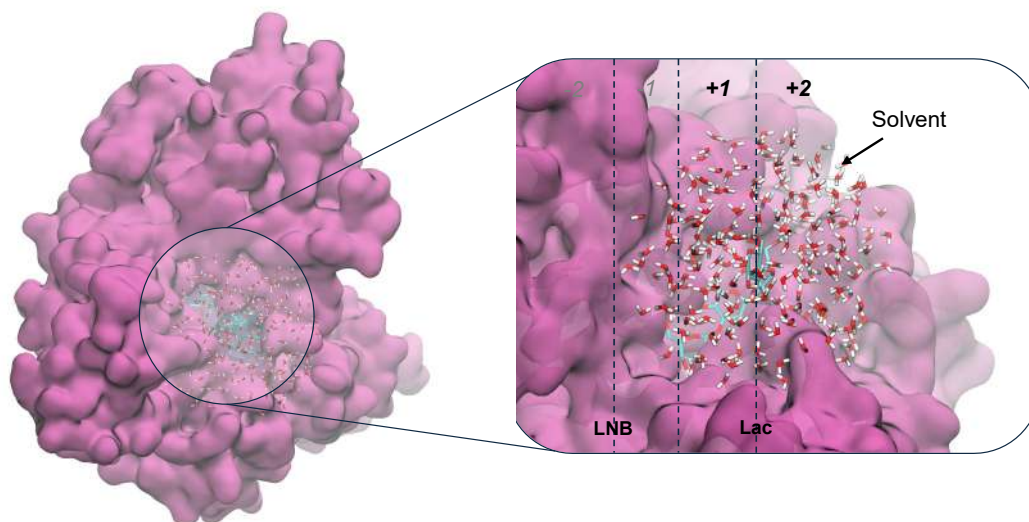


Figure S3.8: Detail of the water molecules that can access near the active site of LnbB and solvate the Lac moiety of the substrate. The fact that the active site of LnbB is relatively open at the part of the leaving group explains the promiscuity of this enzyme, that can cleave HMOs with LNB at their non-reducing end.

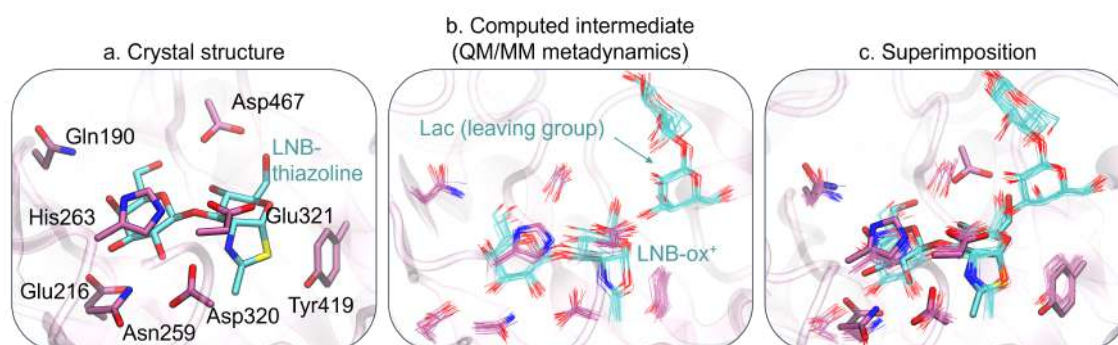


Figure S3.9: a. Active site of the LnbB crystal structure [94]. b. Superimposition of a selection of trajectory frames corresponding to the reaction intermediate state obtained with QM/MM metadynamics. c. Superimposition of a. (in stick representation) and b. (in line representation) shows that the computed reaction intermediate structure is in good agreement with the crystal structure, thus the latter is a suitable starting point for modeling the reaction intermediate of LnbB (Section 3.4.2.1)

Chapter 4

The catalytic mechanism of GH130 *N*-glycan mannoside phosphorylase

Publication:

Alfonso-Prieto, M.[#], Cuxart, I.[#], Potocki-Véronèse, G., André, I., Rovira, C.
"Substrate-Assisted Mechanism for the Degradation of *N*-Glycans by a Gut Bacterial Mannoside Phosphorylase." *ACS Catalysis*, 13, 4283-4289 (2023) [#] Equal contribution

Abstract: The bacterial degradation of *N*-glycans lining the human intestinal epithelium is associated with conditions like inflammatory bowel diseases. A participant in this process, the enzyme unknown human gut bacterium mannoside phosphorylase (UhgbMP), releases mannose units from *N*-glycans by phosphorylation. Crystal structures show that, unlike most glycoside phosphorylases operating with the classical Koshland mechanism, the putative catalytic acid of UhgbMP (Asp104) is far from the scissile glycosidic bond. By means of QM/MM simulations combined with metadynamics we showed the reason for this anomaly. We computed the catalytic mechanism with a natural substrate and demonstrate that UhgbMP catalysis is assisted by the mannose substrate, with its 3-OH as the last proton donor to the scissile bond via a double proton shuttle involving Asp104. During this process, the mannose substrate undergoes the conformational itinerary $^1S_5/B_{2,5} \rightarrow [B_{2,5}]^\ddagger \rightarrow ^OS_2$. The active site architecture enabling the hydrogen network between the catalytic acid and the substrate is conserved across CAZY family GH130, therefore this mechanism can be extended to other GH130s and other mannoside phosphorylases displaying similar active site features. Understanding the molecular determinants of the catalytic mechanisms of these enzymes can benefit the development of specific inhibitors to mediate gut bacteria-host interactions and enzyme engineering strategies for synthesizing high added-value products.

4.1 Introduction

The intestinal tract is a complex environment where microorganisms thrive and secrete enzymes capable of degrading complex carbohydrates like *N*-glycans [157] coming from the host diet [40], other intestinal microorganisms [158, 159], and the lining of the epithelial tissues themselves [160]. The relationship between microbes and host is complex and crucial for many biological processes, both in health and disease [161]. For example, gut bacteria secrete specific enzymes for degrading HMOs [88] which is helpful for the development of the infant human gut microbiome, as described in Chapter 3 of this Thesis, while others can degrade glycans from the host gut epithelial tissues themselves. Alterations in their prevalence is linked to diseases such as inflammatory bowel diseases (IBDs) and colorectal cancer [162–165].

Present in epithelial tissues, *N*-glycans are a type of carbohydrates linked to glycoproteins through a β -GlcNAc linked to an Asn-X-Ser/Thr motif (in eukaryotic organisms) involved in molecular recognition functions like modulation of the antibody-mediated immune response [166, 167], regulation of protein folding and cell-matrix and cell-cell recognition and constitute the most common post-translational modification of proteins [164, 168]. They have a common core structure consisting in Man- α -1,3-(Man- α -1,6)-Man- β -1,4-GlcNAc- β -1,4-GlcNAc, which is then extended, branched and modified (Figure 4.1.a) [1, 58]. In the human digestive tract, they are attached to glycoproteins of the epithelial tissues thus constituting a protective barrier and carbon source for microorganisms [13].

The degradation of intestinal *N*-glycans has long attracted researchers' attention due to their role in molecular recognition and host-microbe relationships, and for some time only hydrolytic degradation by GHs had been reported [58, 169]. In 2013, however, a mannoside phosphorylase secreted by a human gut bacterium was identified for the first time after studying and comparing sequences obtained in a 2010 functional metagenomics study of the human gut [13, 40]. This newly discovered enzyme was key for degradation of *N*-glycans in the intestinal tract, which can be of human origin (from the intestinal epithelial lining) or from other microorganisms (like the opportunistic pathogen *Candida albicans*) [159], thus in both cases of eukaryotic origin. Its secretor organism has not been fully identified yet besides belonging to the *Bacteroides* genus, therefore it has been coined as "unknown" human gut bacterium mannoside phosphorylase (UhgbMP) and classified in the family GH130 of glycoside hydrolases.

UhgbMP is an *exo*-acting enzyme that cleaves the β -1,4 glycosidic bond between of the human *N*-glycan motif Man- β -1,4-chitobiose using inorganic phosphate (Pi)

and releases as α -Man-1-P [13] (Figure 4.1.b). It has been experimentally reported that UhgbMP acts on different substrates: besides the Man-(GlcNAc)₂ motif, it recognizes and phosphorylyzes substrates with Man- β -Glc and mannan oligosaccharides, Man- β -1,4-(Man)_n ($n=1-15$), of dietary origin [13], and can perform the inverse reaction, the reverse phosphorylysis, to synthesize β -1,4-mannosyl linkages with different acceptors as well [13]. However, it is specific for β -mannosyl units at the non-reducing end and does not process Glc- or GlcNAc-linked glycans [13].

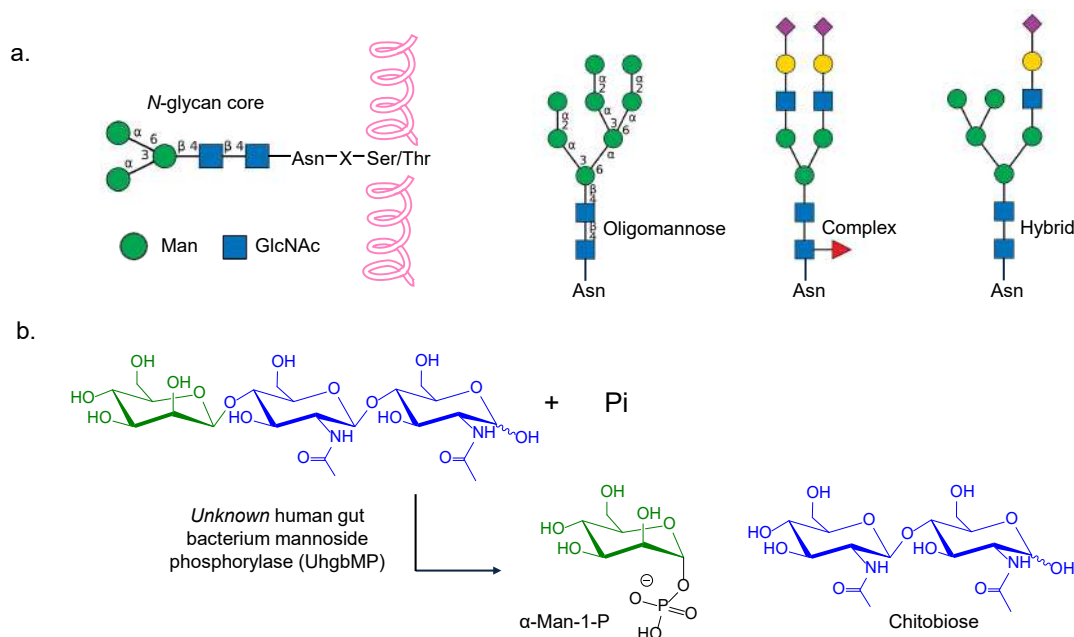


Figure 4.1: a. From left to right: Core structure of N-glycans. "X" in the Asn-X-Ser/Thr motif can be any of the 20 proteogenic amino acids except proline. From the common core, structures are extended to form the different types of N-glycans, which are transferred "en bloc" to the correspondent glycoproteins during biosynthesis. Adapted from [1]. b. UhgbMP catalyzes the phosphorylysis reaction on N-glycan fragments with Man- β -1,4-(GlcNAc)₂ moiety at the non-reducing end to release Man-1-P.

The CAZY family GH130 includes both GHs and GTs. Glycoside phosphorylases, or GPs, are carbohydrate-processing enzymes that incorporate a phosphate group to the non reducing end of their substrates resulting in the cleavage of the glycosidic bond with the subsequent saccharide unit (Figure 4.2). Their reaction mechanism shares aspects with both GHs and GTs: they either invert or retain the substrate anomeric configuration [23] and is based in nucleophilic and acid/base catalysis. However, the nucleophile is an enzyme residue in classical GHs, whereas in GPs is an inorganic phosphate group (Pi) bound to the enzyme active site via hydrogen bond interactions [13]. Also, the Pi in GPs is not covalently bound to

a sugar-nucleotide donor, characteristic in GTs [41]. The phosphorolysis reaction by GPs is reversible and is known as reverse phosphorolysis, which has applications for producing high added value products without nucleotide-activated sugar donors. GPs are classified in both GH and GT families (for example GH65, GH94 and GH130, GT4 and GT108 [13,170]). It took some time to distinguish them in metagenomics studies such as the ones trying to unravel new enzymatic functions of gut bacteria [40], as their sequence similarities with GHs and GTs challenged their identification, while other types of CAZymes like polysaccharide lyases (PLs) are more distinct. Therefore, as sequence and structural data confirmed these similarities, GPs established their place among GH and GT families [13].

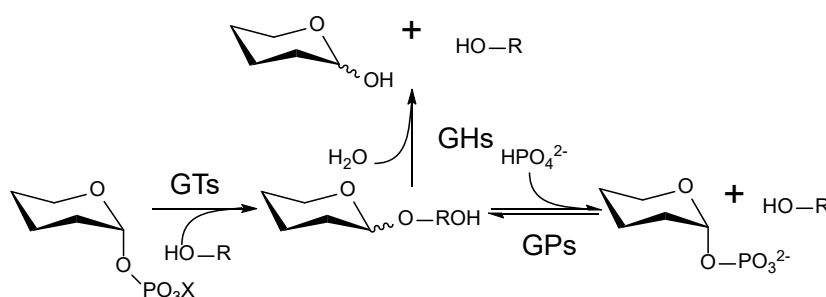


Figure 4.2: Comparison of the reactions by glycoside hydrolases (GHs), glycosyltransferases (GTs) and glycoside phosphorylases (GPs). R=acceptor group, X=activating group of donor saccharide, such as a nucleotide. Figure adapted from [23,39].

Other GH130 members are mannosidases and phosphorylases degrading β -mannosides [53,171,172]. GH130s are further classified in subfamilies GH103.1 and GH130.2 based on substrate specificity. While GH130.1 members act specifically on Man- β -1,4-Glc [53], GH130.2 phosphorylases, where UhgbMP belongs, accept diverse mannosides. UhgbMP belongs to subfamily GH130.2 along with another identified gut bacterium phosphorylase, *Bt1053*, also degrading *N*-glycan core oligosaccharides [173].

The tridimensional structure of UhgbMP was solved in 2015 by X-Ray crystallography [58] and consists in a homohexamer with tightly bound protomers each having the same five-bladed β -propeller fold characteristic in family GH130. The obtention of the crystal structure of UhgbMP opened the door to exploring its catalytic mechanism, proposed for the first time in 2013 for a *Bacteroides fragilis* GH130.1 phosphorylase [53] and helped identifying the structural factors at play for substrate specificity. Regarding the oligomerization state of the enzyme, although the structural data, together with other techniques (SAXS analysis, multiangle laser light scattering and size-exclusion chromatography), indicate that the active form of UhgbMP is a homohexamer [58], it was not yet clear why this configuration is the one

favoring catalysis. The active site was identified hosting a phosphate ion (Pi), Man and GlcNAc in the four different structures available at the PDB (unliganded, 4UDI, and complexes with Man+Pi, 4UDJ, GlcNAc+Pi, 4UDG and Man+GlcNAc+Pi, 4UDK) [58]. The solved structures of UhgbMP also shed some light on the relaxed substrate specificity of UhgbMP with respect to similar enzymes: Its short loop at positions Gly121-Pro125 where other GPs like *Bf*MGP have a larger loop that would preclude the entrance of larger manno-oligosaccharides (Figure 4.3). However, it was not yet clear what interactions governed the specificity towards different saccharides at substrate $+1$. Regarding the oligomerization state of the enzyme, although the structural data, together with other techniques (SAXS analysis, multiangle laser light scattering and size-exclusion chromatography), indicate that the active form of UhgbMP is a homohexamer [58], it was not yet known why this configuration is the one favoring catalysis.

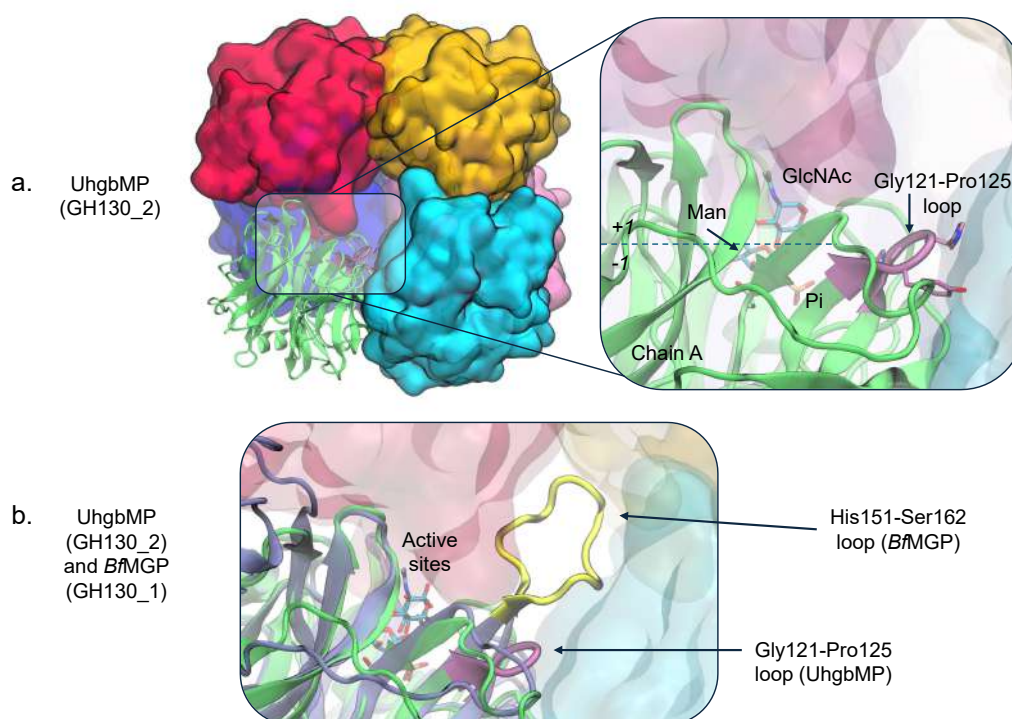


Figure 4.3: a. Crystal structure of UhgbMP (PDB 4UDJ and 4UDK) with the Gly121-Pro125 loop near the active site. The position of this loop in the hexameric form is probably the structural explanation of UhgbMP ability to bind manno-oligosaccharides beyond subsite -1 . b. Superposition of structures of UhgbMP and *Bf*MGP (PDB 3WAT) shows that the equivalent position of UhgbMP Gly121-Pro125 loop is occupied by a larger motif in *Bf*MGP, also reported in other subfamily GH130.1 members [58].

Concerning the reaction mechanism, previous hypothesis suggested that UhgbMP operated with the general mechanism for inverting glycoside phosphorylases (GP), a single displacement mechanism like the classical inverting GHs (Figure 4.4.a and 4.4.b), reported in families GH65, GH94 and GH112 [13,23,174]. In classical inverting GPs, an amino acid residue of the enzyme protonates the leaving group, while the Pi group at the active site is the nucleophile (Figure 4.4.b). A bioinformatics study identified in 2013 potential candidates for the general acid: Asp104, Asp304 and Glu273, which are conserved in family GH130 and proved to be essential as the enzyme activity was severely impaired for its respective variants to asparagine and glutamine [13].

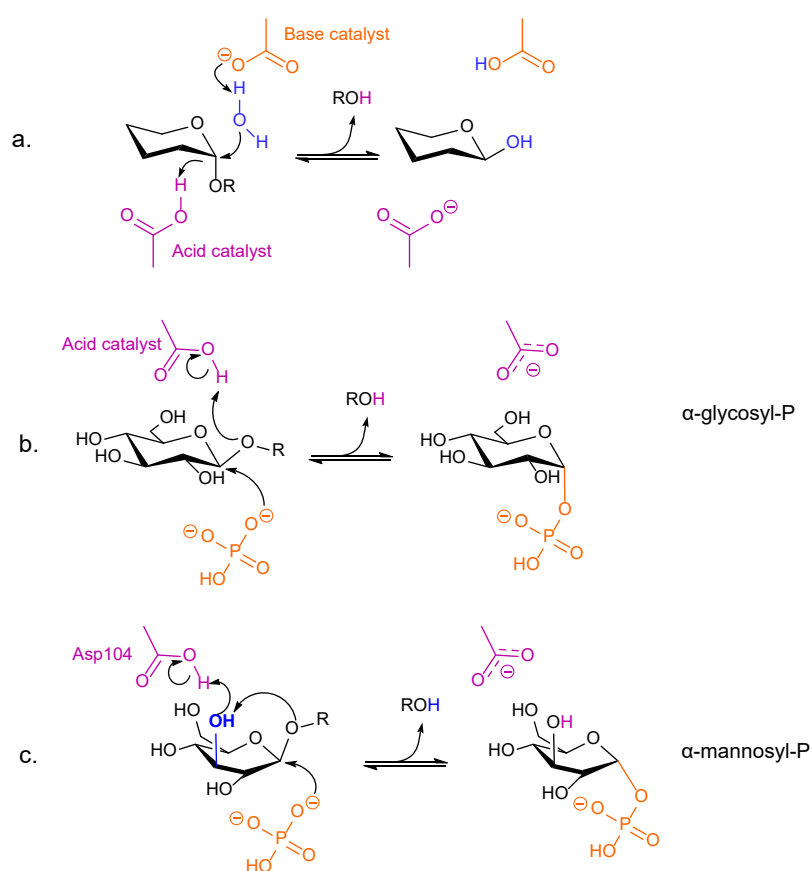


Figure 4.4: a. Classical inverting glycoside hydrolase mechanism. Single displacement like in GPs, but the base catalyst of the reaction is an enzyme residue activating a water molecule for the nucleophilic attack on the anomeric carbon. b. Classical mechanism of inverting GPs, where the nucleophile of the reaction is a Pi group and the enzyme acidic catalyst is the direct proton donor of the scissile bond. c. Inverting, substrate-assisted hypothesized mechanism for GH130 mannoside phosphorylases where a hydroxyl substituent of β -mannosyl is involved in a proton relay and acts as the last proton donor. Adapted from [26,33].

The newly solved crystal structures of UhgbMP finally identified Asp104 as the

putative acid base [58], while Asp304 interacts with Man 4-OH and 6-OH, thus rather involved in Man binding and stabilizing its $B_{2,5}$ conformation. Regarding Glu273, it is not in the active site (although is part of a short α -helix, so its role may be crucial to contribute to the enzyme integrity) [13, 58]. The position of Asp104, however, remained intriguing: Although close to β -mannosyl in the active site, the structures revealed that it was farther from the scissile glycosidic bond than other GPs (Figure 4.5) ($d[\text{O}_{\text{Asp104}}-\text{O1}] = 3.59 \pm 0.12 \text{ \AA}$ [33, 58]). This challenged the classical inverting GP mechanism, and consequently a novel mechanism was proposed (Figure 4.4.c): Asp104 is close enough to the 3-OH substituent of mannose to establish a hydrogen-bond interaction, then the Man 3-OH in turn can interact with the oxygen of the scissile glycosidic bond. It was hypothesized that UhgbMP follows a mechanism involving a double proton transfer from the acid catalyst to the oxygen at the scissile bond via the substrate 3-OH exocyclic group [58]. This mechanism can be described as substrate-assisted catalysis. Although the enzyme acid catalyst, Asp104, plays a key role establishing the double proton transfer shuttle, ultimately the leaving group is protonated by the substrate itself.

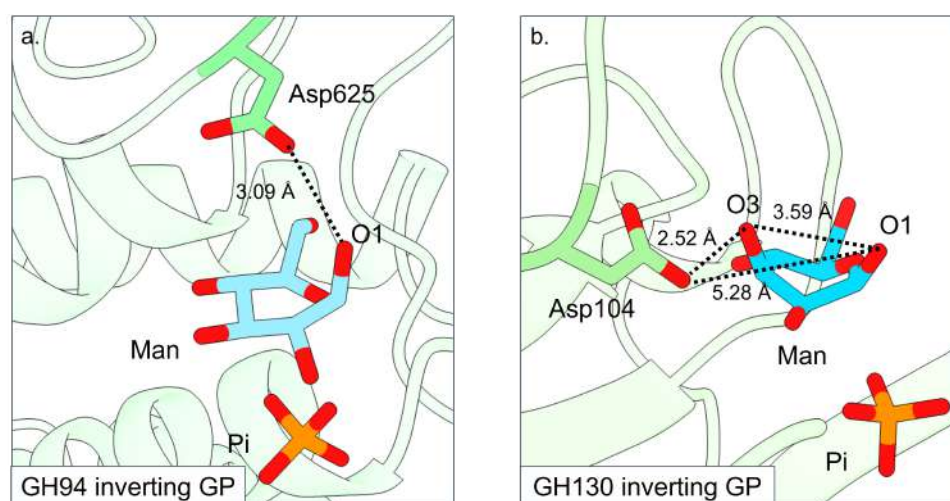


Figure 4.5: a. Active site of GH94 bacterial inverting GP in complex with Man and Pi (Cdp PDB ID 8H2V [175]) with a canonically positioned acid catalyst at 3.09 Å of the scissile bond O. b. Active site of UhgbMP in complex with Man and Pi (PDB ID 4UDJ). Note that the acid catalyst is far from O1, but at H-bond distance of Man-O3. The distances in the figure correspond to the active site in protomer A of the structure, but this configuration is observed in the active sites of all subunits.

The reaction mechanism was first proposed for 4-*O*- β -1,4-mannosylglucose phosphorylase (MGP) [53], also with a distal catalytic acid residue and no alternative for substrate protonation: Neither in UhgbMP nor MGP have water molecules at the active site that could protonate the leaving group [53, 58]. This active site architecture has been reported in other GPs of family GH130 (Tables S4.2 and S4.3) [53, 176, 177] and recently in family GT108 GPs [170]. These latter dual-activity enzymes share

the fold with GH130 GPs; therefore, it is likely that their phosphorolytic activity takes place with a common mechanism [33]. Therefore, although other enzymes similar to UhgbMP have been identified, to the best of our knowledge none have been studied with computational methods yet. By modeling the reaction mechanism of UhgbMP, it is possible to describe at an atomistic level the double proton transfer from the distal catalytic acid to the scissile bond and characterize the molecular determinants, such as the substrate conformational catalytic itinerary.

Characterizing the molecular details of the catalytic mechanism of UhgbMP can give insight for the design of specific inhibitors targeting its activity. Besides the interest as therapeutic target, UhgbMP has attracted attention for biosynthetic applications. Alternatively to its main phosphorolytic activity UhgbMP can perform the reverse reaction, the so-called reverse phosphorolysis, where a saccharide acceptor is incorporated to the mannosyl-1-P donor to form a new glycosidic linkage and an extended oligosaccharide is released [23,178]. Since UhgbMP can recognize different types of mannose-oligosaccharides (even plant mannans) with β -1,4 linkage [13], enhancing this activity *in vitro* opens the door to engineering biosynthetic tools to produce complex oligosaccharides [13] with regio- and stereoselectivity control. This process that can be aided by knowing the main determinants for substrate recognition and processing during the phosphorolysis reaction that we computed in this study.

In this Thesis Chapter, we modeled the catalytic mechanism of UhgbMP. Our results confirm that the predicted phosphorolysis reaction takes place via the predicted substrate-assisted mechanism where the 3-OH of the reactive mannosyl ring is the last proton donor of the scissile bond, involved in a proton relay with the enzyme Asp104 residue. With QM/MM MD simulations we also revealed the conformational itinerary of the reactive mannose as it undergoes catalysis: The reaction starts from a distorted $^1S_5/B_{2,5}$ and proceeds via a $[B_{2,5}]^\ddagger$ transition state, trait that shares with several retaining β -mannosidases (families GH1, 2, 5 26, 113, 130, and for GH164, $^OS_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow ^1S_5$) [41,56,179]. Additionally, we characterized the enzyme-substrate complex by demonstrating that the hexameric organization of the enzyme is key for substrate recognition and integrity of the active site cavity, while the main residues for substrate recognition were identified.

4.1.1 Project history

This project was conceived and designed by Dr. Mercedes Alfonso-Prieto, as part of her Postdoctoral research in the group of Prof. Carme Rovira, in collaboration with Dr. Gabrielle Potocki-Véronèse and Dr. Isabelle André from the Toulouse Biotechnology Institute (TBI), who resolved the UhgbMP crystal structures in 2015. In

the first place M.A.P. built the UhgbMP-Man- β -1,4-GlcNAc Michaelis complex, as well as the complexes in the low-order oligomeric forms (monomer and dimer) and with alternative acceptors, and performed classical MD simulations. Afterwards, she built the QM/MM system and performed QM/MM metadynamics simulations to build the first free energy landscape of the phosphorolysis reaction with CPMD, after assessing the suitable parameters and CVs. Then, she conducted the committor analysis to characterize the reaction transition state structure. My contribution in the project started by collaborating with the analyses of the previous results and assisted in the elaboration of a manuscript of the article. When we decided to expand the project, I performed independent replicas of the classical MD simulations and conducted the analyses. Subsequently I built and conducted QM/MM MD simulations starting from suitable structures alternative active sites and replicas and performed additional metadynamics simulations using CPMD+PLUMED, all yielding the same reaction mechanism with a free energy barrier in agreement with the experimental rate constants. Additionally, I performed pK_a estimation calculations to assess the choice of active site protonation states. After, together with M.A.P we designed classical MD simulations for the alternative protonation states (MC, MC', P and P' in Section 4.2.3). More details about this in the Computational Methods of this Chapter (see Section 4.4).

4.2 Results and discussion

4.2.1 Modeling the Michaelis complex of UhgbMP with Man- β -1,4-GlcNAc

The Michaelis complex of UhgbMP with the natural substrate Man- β -1,4-GlcNAc and inorganic phosphate (Pi) was built by superimposing two of the available crystal structures of the enzyme (See Section 4.4.1 and Figure S4.1). On one hand the structure corresponding to the enzyme in complex with β -Man, GlcNAc and Pi (PDB accession code 4UDK) and on the other, the structure of UhgbMP in complex with Man and Pi (PDB accession code 4UDJ) [58]. The last structure has a slightly higher resolution (1.60 Å vs. 1.76 Å in 4UDK) and displayed the Man moiety better positioned for catalysis, in-line for the predicted Pi nucleophile attack and its O substituent at C1 in a geometry compatible with an axial glycosidic bond with GlcNAc after superimposition. The upcoming sections address different aspects regarding UhgbMP function we investigated in this study. Specifically, why the hexameric oligomerization order favors binding and phosphorolysis of substrate? What characteristics of the active site allow the relative relaxed specificity for dif-

ferent carbohydrates β -1,4-linked to mannose? And finally, what are the transfer pathway for relaying a proton to the scissile bond for the posphorolysis reaction and the conformational itinerary of β -mannosyl during the reaction?

4.2.1.1 The oligomerization order of UhgbMP

The homo-hexameric organization of UhgbMP has been reported with experimental techniques such as SAXS analyses, size-exclusion chromatography and multi-angle laser light scattering [58]. However, to gain molecular insight on why this organization is the one carrying out the biological function and characterize the enzyme residues that might be important for keeping the integrity of the enzyme-substrate complex, MD simulations of the UhgbMP in complex with Pi and Man- β -1,4-GlcNAc for three different oligomeric orders of the enzyme (see section 4.4.3) were designed. The simulations with the monomer, dimer and hexameric forms revealed that the latter keeps the substrate tightly bound in the active site cavity (Figures 4.6, S4.2, S4.3), whereas in the monomeric and dimeric forms the binding is less stable. They are also less prone to sample conformations competent with the hypothesized mechanism, as the main distances involved in the phosphorolysis reaction fluctuate more in these simulations than in those of the hexamer (Figures S4.8 and S4.3).

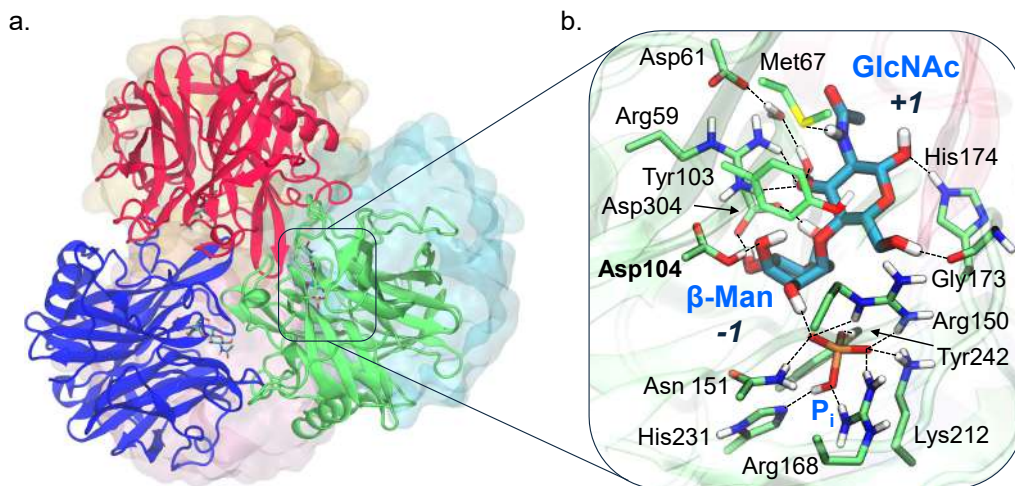


Figure 4.6: a. Hexameric model of the UhgbMP Michaelis complex with Man- β -1,4-GlcNAc and Pi, built from PDBs 4UDK and 4UDJ. Each subunit in a different color. Cartoon representation (front) and transparent surface (back). b. View of active site A with bound substrate and Pi interacting with enzyme residues after MD simulations. Although the direct interactions are with enzyme residues in the same subunit, close to the subsite +1 another subunit (red cartoon) assists in the packing of the active site and prevents the entrance of bulk water molecules.

The differences between oligomers can be also found in the residues at the other side of the active site cavity: MD simulations of monomer, dimer and hexamer give answers to what residues are important for substrate recognition at the enzyme subsite $+1$, where GlcNAc is bound. The hexamer shows that residues Arg59, Tyr103, His174 and Met67, from the same protomer as the binding site, engage in interactions with GlcNAc (Figures 4.6.b, 4.8). However, although the main residues interacting with the GlcNAc at $+1$ are all in the same subunit, the presence of other subunits assists in the overall packing of the complex. The MD simulations of the monomer and dimeric forms show a GlcNAc with increased overall flexibility (Figures 4.7 and S4.9), and solvent exposure (Figure 4.8). This, besides being deterrent for the binding of GlcNAc, could lead to substrate hydrolysis instead of phosphorolysis as water molecules can enter the active site cavity and hydrolyze the scissile glycosidic bond instead of the analogous reaction with Pi. With these results, we concluded that the hexameric construct of the crystal structure was necessary for modeling the enzyme-substrate complex and subsequently the reaction mechanism.

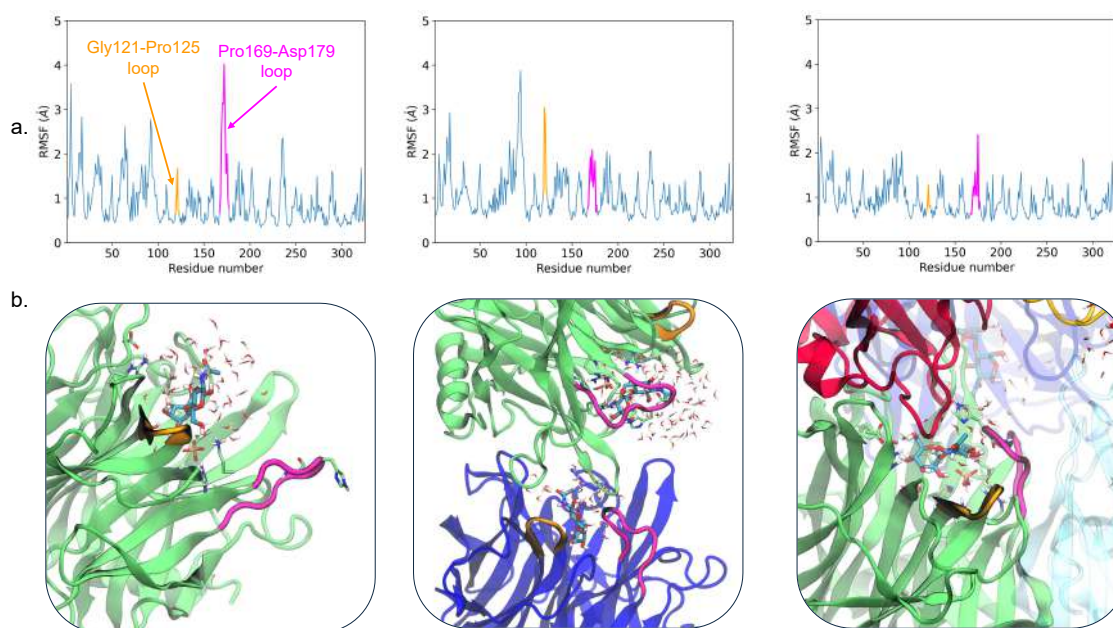


Figure 4.7: a. (From left to right) Root mean square fluctuation over protein residues (one subunit represented for clarity) of the monomer, dimer and hexamer MD simulations. Highlighted, different flexible regions of the protein. b. Cartoon representations of the simulated oligomeric orders. The lower the order, the higher the flexibility of the overall structure. Note that the highlighted flexible loops are close to the active site and may play a part in packing the active site for substrate recognition and binding.

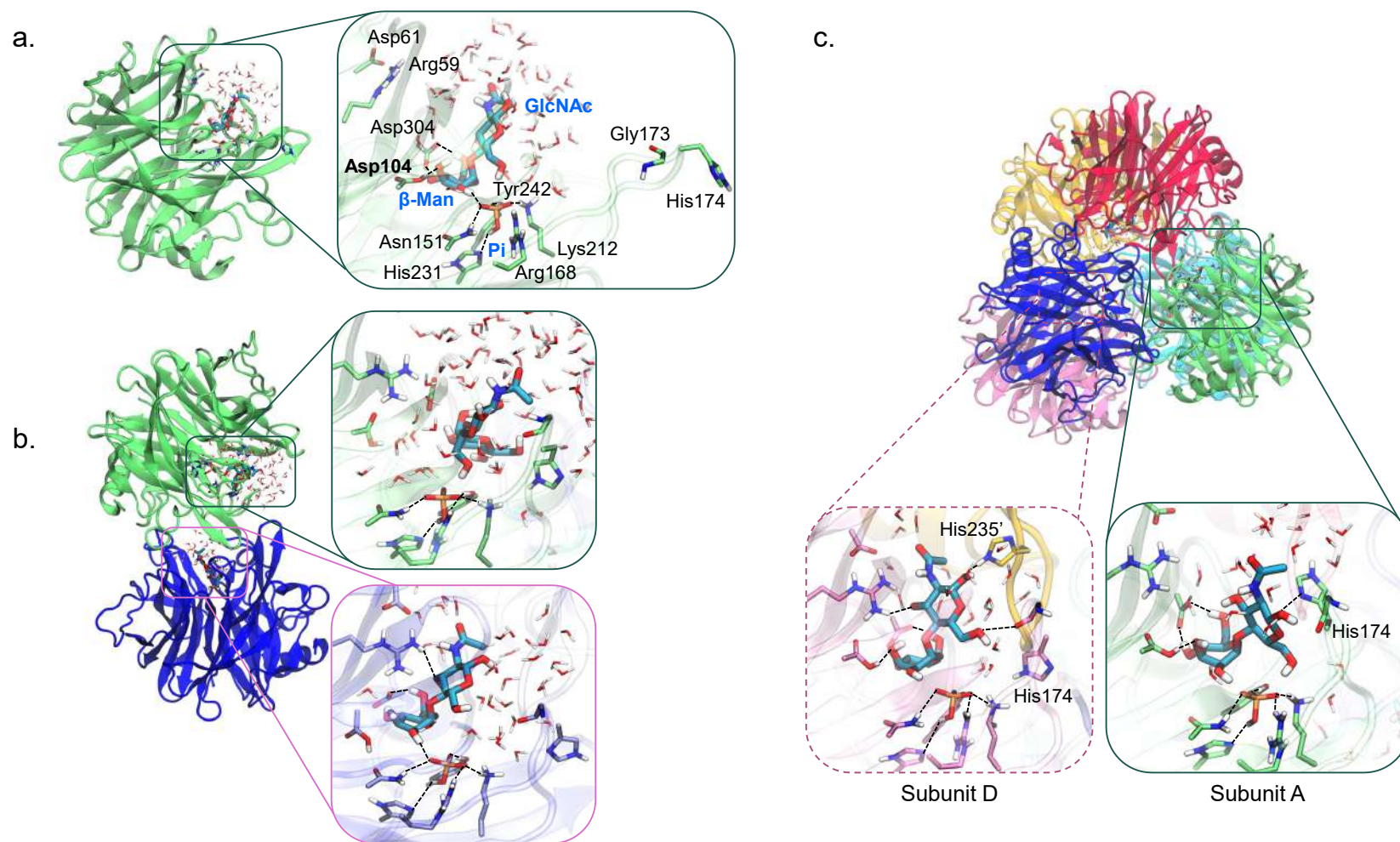


Figure 4.8: Representative trajectory snapshots of the classical MD simulations of the different oligomers of UhgbMP. a. Monomer, b. dimer and c. hexamer. In cartoon representation, each protomer and in sticks, detailed view of the active sites.

4.2.1.2 The substrate recognition by UhgbMP

The flexibility of UhgbMP for different substrates has been previously assessed with experimental techniques [58], showing that the enzyme can accommodate β -1,4-mannosyl moieties linked to different carbohydrates and even large *N*-glycans [13]. To gain insight on the role of the enzyme residues in the acceptor recognition, we performed classical MD simulations of UhgbMP in complex with substrates considering other acceptor sugars at subsite +1 that can also be found in *N*-glycan motifs [1,58]: Man- β -1,4-Glc and Man- β -1,4-Man, and compared the interactions they form with the enzyme with the previously simulated UhgbMP-Man- β -1,4-GlcNAc complex. These simulations show that UhgbMP complexes with Glc and Man acceptors are stabilized by forming interactions with residues at subsite +1. Two interactions seem to be the key for substrate recognition: A hydrogen bond with Arg59 and a CH- π interaction with Tyr103 are stable during the simulations and observed for the Man- β -1,4-GlcNAc complex as well (Figures 4.9 and S4.10 to S4.12). Other residues at subsite +1, Met67 and His174, interact with the acceptors as well. The differences between interaction patterns with the different acceptors (Figures S4.10 to S4.12) may be related to the differences in K_M values of the Man- β -1,4-Man and Man- β -1,4-Glc and Man- β -1,4-GlcNAc complexes [13]. Overall, these results provide a molecular perspective on the known promiscuity of UhgbMP for β -1,4-linked mannositides.

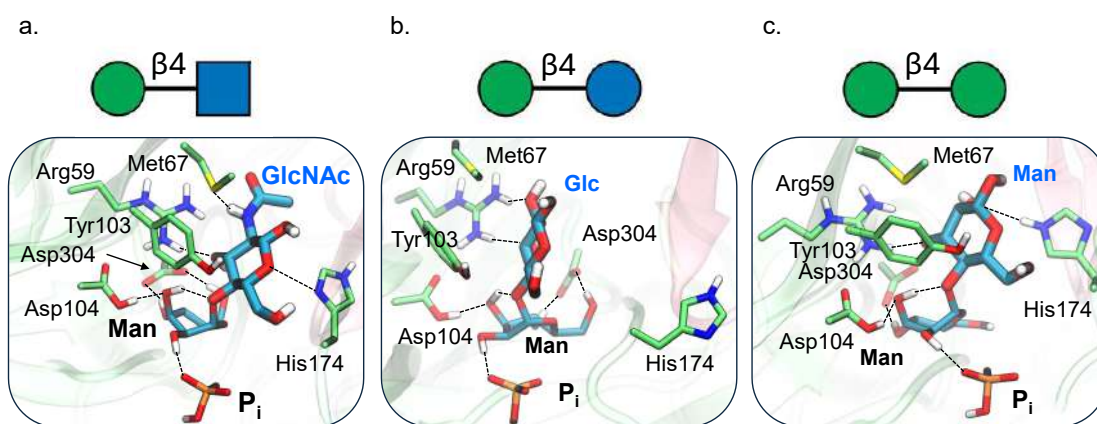


Figure 4.9: Representative classical MD snapshots showing the main UhgbMP-substrate interactions at the active site. Note that UhgbMP can accommodate different acceptor sugars that constitute part of *N*-glycans: a: Man- β -1,4-GlcNAc. b: Man- β -1,4-Glc. c: Man- β -1,4-Man. Adapted from [33]

4.2.1.3 Distal acid catalyst: a non-canonical active site for a non-canonical mechanism

We analyzed the atomistic details of the active site architecture during classical MD simulations. The β -mannosyl ring at subsite -1 forms persistent hydrogen bond interactions with residues Asp304 and 104 (Figure 4.6.b and S4.3 to S4.7). The inorganic phosphate group is conserved close to the anomeric center of the mannosyl moiety. Although in some of the independent MD replicas the Pi group rotates (Figures S4.3 to S4.7, central columns), one of its oxygen atoms is always in-line for the predicted nucleophile attack according to the classical GP mechanism [23]. Asp104 forms a hydrogen bond with the Man 3-OH exocyclic group, which in turn forms another hydrogen bond with the oxygen atom of the scissile glycosidic bond (Figures 4.6.b and S4.3 to S4.7, left columns). This interaction network, supported by structural data [58], is conserved during the forcefield MD simulations and seems suited for relaying protons, with the β -Man being the last proton donor instead of, for example, a water molecule.

The conformational preferences of the substrate are utilized by enzymes to favor catalysis [16, 54, 56]. In the Michaelis complex of UhgbMP the Man ring at subsite -1 is distorted between 1S_5 and $B_{2,5}$ (Figures 4.6.b and S4.2.c), which is typical in other β -mannosidases [16, 179]. These conformations ease the steric hindrance at the face of the anomeric carbon in line with the Pi for nucleophilic attack, as well as the Man 2-OH, and the scissile bond and the GlcNAc leaving group are axially oriented, which favors its departure. Some crucial active site interactions for stabilizing this conformation are the following: Asp304 forms a bidentate hydrogen bond with both 4- and 6-OH substituents (Figure 4.6.b), placing the carbons 4 and 5 in- and out-of-plane, respectively. Additionally, the Pi group forms a hydrogen bond interaction with the 2-OH substituent. Asp304 is a conserved position across family GH130 and had been found part of a putative catalytic groove identified via sequence alignments and structural modeling in a study by Potocki-Véronèse *et al.* in 2013. Also, kinetic essays of the Asp304Asn variant reported a notable decrease in activity which remained a 4% of the WT enzyme [13]. The role of Asp304 the mannose ring conformation in the active site of UhgbMP is then coherent with the conservation of this position in GH130 mannoside phosphorylases and its importance for UhgbMP activity. Therefore, our results explain previous bioinformatic and experimental results [13].

Although classical MD simulations of the UhgbMP Michaelis complex gave valuable insight into its active site characteristics, the reaction coordinate involving the substrate-assisted proton relay cannot be confirmed yet, therefore it was subse-

quently studied by means of QM/MM metadynamics simulations.

4.2.2 The phosphorolysis reaction

The catalytic mechanism of UhgbMP in complex with Man- β -1,4-GlcNAc and Pi was modeled with QM(DFT)/MM MD simulations, according to the approach developed by Laio *et. al* implemented in the CPMD code [75, 78], starting from a trajectory snapshot of the classical MD simulations of the Michaelis complex. The chemically active area was included in the QM region: The Man- β -1,4-GlcNAc substrate, the Pi, the sidechain of Asp and His231, which interacts directly with Pi. This residue was included in the QM description due to its direct interaction with a catalytic agent. The remaining atoms of the system were included in the MM region, described with the AMBER forcefield. This approach for modeling reaction mechanisms is reported for several carbohydrate processing enzymes [16, 26, 30, 46, 147, 180, 181], as well as for an engineered GH84-turned GP [52].

Two collective variables (CVs) were used to drive the reaction from the reactants state (Michaelis complex of UhgbMP with Man- β -1,4-GlcNAc and Pi) to the α -Man-1-P product (Figure 4.10.a). The first CV is the interaction distance between the inorganic phosphate O atom closest to the Man anomeric carbon during QM/MM MD and the Man C1. CV2 was defined as the glycosidic bond distance. Both CVs account for covalent bonds broken and formed during the Pi nucleophilic attack, thus the protonation pathway of the glycosidic bond oxygen was not biased in our simulations.

The chosen CVs successfully biased the system from UhgbMP in complex with Man- β -1,4-GlcNAc and Pi towards the reaction products, α -Man-1-P and leaving group GlcNAc. The reconstructed FEL shows two minima corresponding to the aforementioned states connected through one transition state (Figure 4.10.b), therefore the reaction proceeds in a single step. The reaction trajectory confirms the hypothesized substrate-assisted mechanism in which the mannosyl 3-OH substituent transfers a proton to the leaving group and engages in a proton relay with the enzyme residue Asp104 which acts as general acid catalyst. This FEL shows that the phosphorolysis reaction on mannosyl substrates by UhgbMP is best described as S_N2 reaction, which is consistent with its nature as inverting GP (although a substrate-assisted one) [23, 174]. The computed energy barrier is 21.6 kcal mol⁻¹, which is close to the value estimated from the experimental reaction rate on a similar substrate [122] (Man- β -1,4-(GlcNAc)₂, $k_{cat} = 0.072 \pm 0.002$ s⁻¹, $\Delta G = 19.8$ kcal mol⁻¹) [13]. Additional simulations of the phosphorolysis reaction starting from an alternative structure of the classical MD simulations yielded a barrier of 21.36 $\pm$

2.02 kcal mol⁻¹ (Figure S4.16).

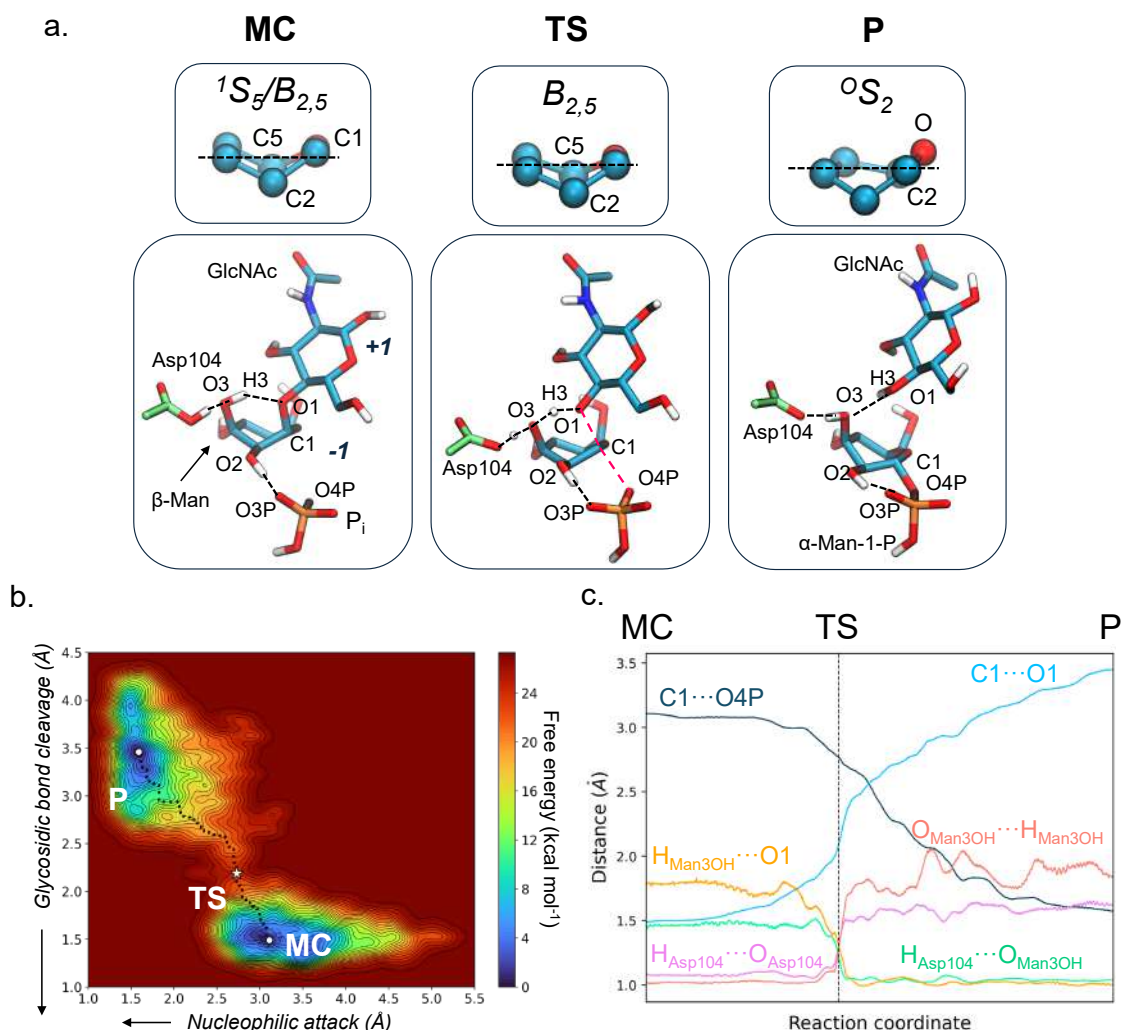


Figure 4.10: a. Trajectory snapshots of the main states of the reaction (reactants, MC, TS and P). MC and P structures are representative from an ensemble of configurations at the FEL minima MC and P, respectively, whereas the TS structure results from committor analysis. On top, ring conformations of the β -Man at MC, TS and P, indicating in each case the atoms out-of-plane with respect to the average ring plane. b. FEL of the simulated phosphorolysis reaction by UhgBMP on Man- β -1,4-GlcNAc. Isolines every 1 kcal mol⁻¹. c. Evolution of the main interatomic distances (bond formation and cleavage, and double proton transfer) along the reaction coordinate, defined as the MEP.

The reaction was simulated by Dr. Mercedes-Alfornso Prieto who started and designed this project as part of her Postdoctoral research in our group. The transition state was assessed by committor analysis, performed by her as well [33]. The results further reinforced the computed reaction profile (Figure S4.15).

By analyzing the structures of the relevant states of the reaction pathway we characterized the molecular details of the reaction mechanism. Initially the glycosidic bond elongates ($C1-O1 = 1.50 \pm 0.09$ Å, Table 4.1) and Pi approaches the

β -Man ($\text{C1-O}_{4P} = 3.41 \pm 0.51 \text{ \AA}$). At the same time, the hydrogen bond distance between Asp104 and the β -Man O3 decreases, and, in turn, the 3-OH bond lengthens ($\text{O3-H}_{\text{Asp104}} = 1.48 \pm 0.13 \text{ \AA}$, $\text{O3-H3} = 1.01 \pm 0.04 \text{ \AA}$). Simultaneously, a double proton transfer takes place (Figure 4.10.c). The Asp104 proton is transferred to the β -Man 3-OH and the latter’s proton is ultimately transferred to O1, resulting in the protonation of the GlcNAc moiety. The dissociative character of the TS (Figure 4.10.c) is demonstrated by the distances of the already broken glycosidic bond ($\text{C1-O1} = 2.19 \text{ \AA}$) and the not yet formed covalent bond between the β -Man and Pi ($\text{C1-O}_{4P} = 2.57 \text{ \AA}$), also reported with QM/MM MD methods in a GH84 engineered to become a GP [52]. The endocyclic oxygen distance with the anomeric carbon shrinks at the TS ($\text{C1-O5} = 1.28 \text{ \AA}$ versus $\text{C1-O5} = 1.41 \pm 0.04 \text{ \AA}$ (MC) and $\text{C1-O5} = 1.36 \pm 0.06 \text{ \AA}$ (P)), which is coherent with the development of a partial double bond in the pyranose ring, indicate its oxocarbenium-ion like nature [54]. This is further evidenced in the $B_{2,5}$ conformation of the β -Man at the TS (Figure 4.10.a, top panels).

Distance ($\text{\AA}$)	State		
	R ($^1S_5/B_{2,5}$)	TS ($B_{2,5}$)	P (0S_2)
C1-O_{4P}	3.41 ± 0.51	2.57	1.81 ± 0.42
C1-O1	1.50 ± 0.09	2.19	3.51 ± 0.44
C1-O5	1.41 ± 0.04	1.28	1.36 ± 0.06
$\text{O3-H}_{\text{Asp104}}$	1.48 ± 0.13	1.22	1.04 ± 0.03
O3-H3	1.01 ± 0.04	1.21	1.95 ± 0.19
H3-O1	1.79 ± 0.15	1.27	1.00 ± 0.03
H2-O_{3P}	1.73 ± 0.31	1.50	1.73 ± 0.19

Table 4.1: Main interatomic distances of the active site at the relevant states of the phosphorolysis reaction (Michaelis complex MC, TS and reaction product P). TS data from committor analysis (Figure S4.15).

After the TS Pi forms a covalent bond with the β -Man anomeric carbon ($\text{C1-O}_{4P} = 1.81 \pm 0.42 \text{ \AA}$), thus forming the α -Man-1-P product, in 0S_2 conformation, and releasing GlcNAc. The $B_{2,5}$ conformation of the Man ring at the TS, with an almost coplanar configuration of the C5-O-C1-C2 atoms, evidences the formation of an oxocarbenium ion-like species. The interaction between the 2-OH and the O_{3P} atom of Pi (stronger in the TS, $\text{H2-O}_{3P} = 1.50 \text{ \AA}$, compared to the MC, $\text{H2-O}_{3P} = 1.73 \pm 0.31 \text{ \AA}$, see Table 4.1) also contributes to TS stabilization by the inductive effect on the anomeric carbon [182], which has been reported to be in the order of 10 kcal mol^{-1} in GHs [183].

The conformational determinants of the Man substrate ring were also charac-

terized. The computed conformational itinerary is ${}^1S_5/B_{2,5} \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^0S_2$ which lays at the equator of the Cremer and Pople sphere (Figure 4.11), as reported in other mannose-degrading enzymes [41,179]. The ring is distorted at MC, in a conformation best described between 1S_5 and $B_{2,5}$. These conformations meet the stereoelectronic properties that favor catalysis [54]. First, the scissile glycosidic bond is axially oriented and lengthened [41]. Then, the 2-OH exocyclic group of Man is equatorially oriented, which facilitates hydrogen-bond formation with one of the Pi O atoms. This interaction helps placing the Pi in a geometry favorable for the subsequent nucleophilic attack (Table 4.1, $C1-O_{4P} = 3.41 \pm 0.51$ Å) with an angle between $O_{4P}-C1-O1$ of 140 ± 8.27 degrees. Important interactions that stabilize the Man conformational itinerary are the bidentate hydrogen bond between Asp304 and the Man 4-OH and 6-OH groups, maintained throughout the reaction, and the aforementioned Man 2-OH with O_{3P} , stabilizing the TS with $C5-O1-C1-C2$ coplanarity.

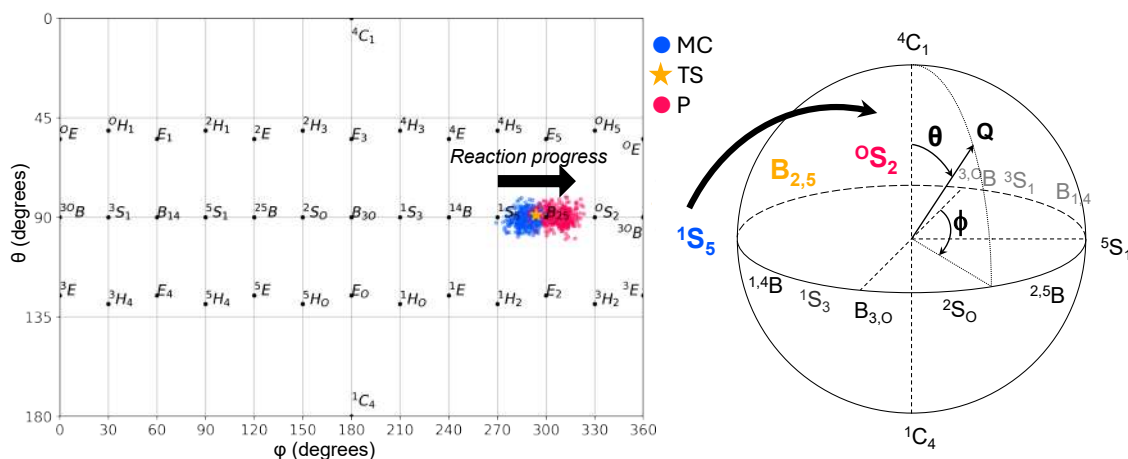


Figure 4.11: Right, Mercator projection of the conformational itinerary of the β -Man at subsite -1. Data points correspond to ensembles of conformers at the main minima of the reaction FEL (reactants, MC, blue and products, P, in magenta. The TS, star, was obtained from the committor analysis). Left, representation of the conformational itinerary over the Cremer-Pople sphere. The itinerary ${}^1S_5/B_{2,5} \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^0S_2$ is along the equator of the sphere and is followed by other β -mannosidases.

In summary, our QM/MM metadynamics simulations confirm a novel mechanism for UhgbMP in which, unlike other inverting GHs and GPs, the substrate is involved in catalysis by acting as the last proton donor of the scissile glycosidic linkage via the 3-OH substituent, which in its turn establishes a proton relay with the enzyme acid catalyst Asp104, instead of this residue transferring the proton directly [16,46]. This result is further reinforced by the fact that, not being included in any of the CVs of the metadynamics simulation, the proton transfer pathway did occur. Compared with a direct transfer from Asp104 to the scissile bond oxygen (O1), the double

proton transfer mechanism involves less nuclear motion [184,185]: to transfer the proton directly, the Asp104 residue should approach the scissile bond from around 3.6 Å, which would require significant atomic rearrangements of the active site. The buried position of Asp104 complicates this motion, whereas the network OH_{Asp104}-O3-H-O1 is in the suitable geometry for the synchronous double proton transfer, a phenomenon that has been reported in other types of enzymes [186–189].

4.2.3 Study of protonation states

How this substrate-assisted proton transfer occurs via the hydroxyl substituent of a saccharide deserves extended discussion. Sugar OH groups in solution have high pK_a values, of the order of 13 or higher [190], which makes them poor proton donors. Thus, the role of the hydrogen bond network connecting the putative catalyst with the glycosidic bond also improves the acidity of β -Man 3-OH. This is achieved through consecutive pK_a shifts, dependent on the electrostatic conditions provided by the specific enzyme environment: in GHs, both retaining and inverting, the acid catalyst is often an aspartate or glutamate residue with an increased pK_a respect to that in aqueous solution [127,129,191].

In the predicted UhgbMP mechanism Asp104 needs protonated to transfer the first proton to the Man 3-OH. Additional MD simulations of the enzyme-substrate complex with deprotonated Asp104 (MC') showed the disruption of the hydrogen bond network necessary for protonating the leaving group (Figure 4.12). The Man 3-OH forms a hydrogen bond with Asp104, thus losing its interaction with O1 (Figure 4.12.c, MC'). After some simulation time, Asp104 flips downwards and interacts with Arg106 via a bidentate hydrogen bond (Figure 4.12.c, MC''). Since no other residue took up the role of Asp104 to reestablish the proton relay [58], we concluded that Asp104 must be in its carboxylic acid form for the reaction to proceed. In the alternative protonation scenario for the reaction products (protonated Asp104 in P', Figure 4.12.c) the proton relay is also disrupted, which would preclude the reverse phosphorolysis reaction (synthesis of Man- β -1,4-GlcNAc).

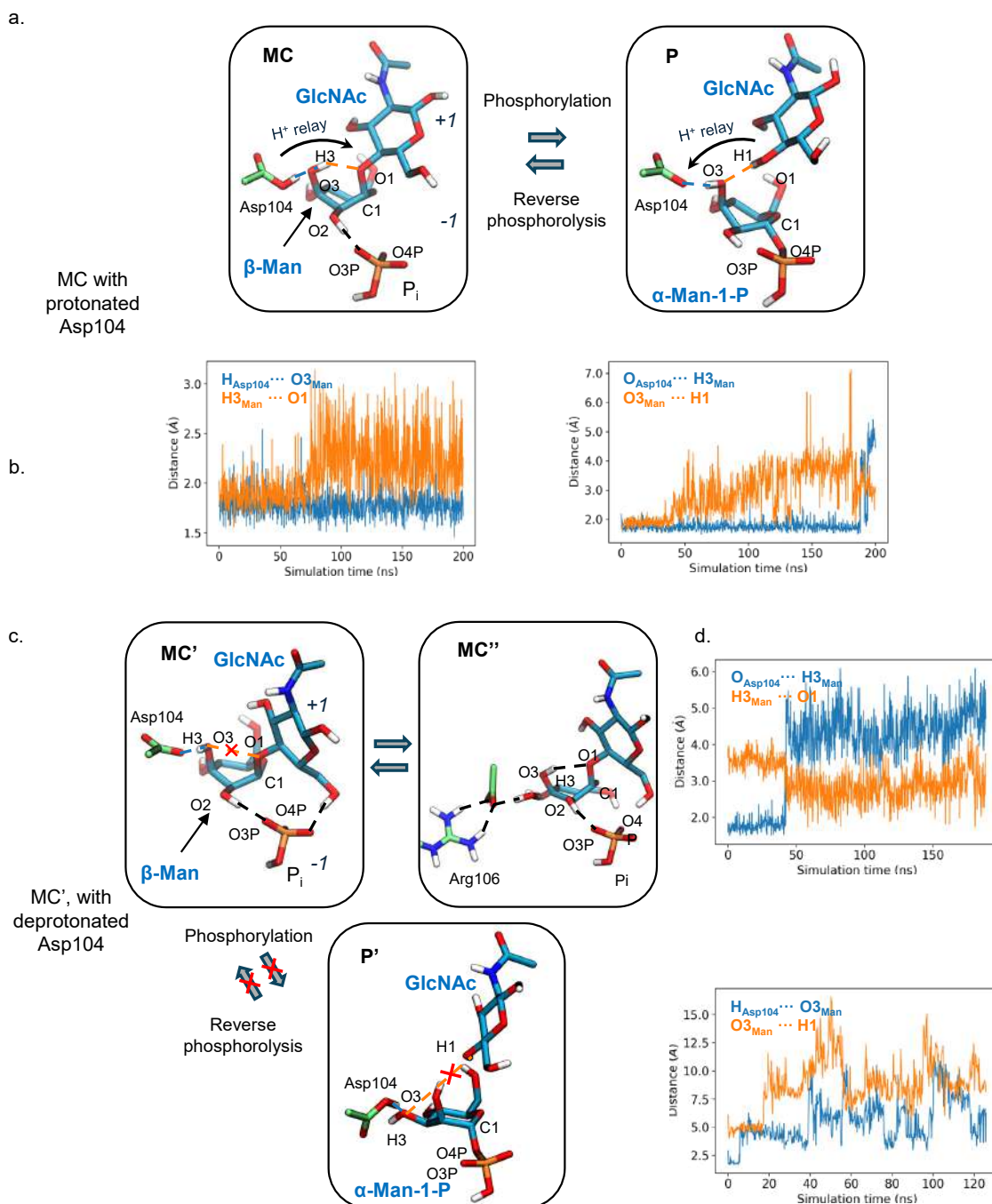


Figure 4.12: Representative snapshots of classical MD simulations of reactants and products states considering alternative protonation states for Asp104. a. (Left) Michaelis complex with protonated Asp104 can progress towards phosphorylation to form α -Man-1-P product (P). (Left) Products state with the deprotonated Asp104 obtained in this work. b. This network is stable during classical MD simulations in both the MC and P states. c. Michaelis complex with carboxylate Asp104 (MC') where the double H-bond network is disrupted (MC'') and Asp104 interacts with Arg106 (MC''). This active site configuration cannot progress towards the phosphorylation reaction. (Bottom) Products complex with protonated Asp104 (P'), where the proton relay to perform reverse phosphorolysis cannot be formed. d. The H-bond network is disrupted in the latter scenarios, precluding the proton relay.

The participation of hydroxyl groups in enzyme catalysis is unusual but not unprecedented: The ribosomal peptidyl transferase uses ribose hydroxyl groups from the mRNA substrate to catalyze peptide bond formation in the ribosome [192,193], and in a NAD⁺-dependent protein deacetylase, the nicotinamide ribose hydroxyl groups of the substrate are also reactive [194]. Substrate hydroxyl group participation in catalysis has been also reported in a GH99 endomannosidase, in which the Man moiety 2-OH group forms an epoxide intermediate [50]. At a molecular level, the pK_a rise of Asp104 with respect to the value in solution is assisted mostly by two factors: The water-shielded active site and negative charge of the Pi group nearby (at ~ 7 Å from Asp104) in a similar fashion of inverting GHs, whose catalytic acid pK_a shift is modulated by the negatively charged nucleophile at 9-10 Å [129]. Additional pK_a estimations using web server-hosted H++ program supported the need for Asp104 to be protonated in the Michaelis complex of UhgbMP (Table S4.1). Therefore, these simulations contributed in assessing the choice for the protonation state of Asp104 in UhgbMP.

4.3 Summary and conclusions

In this Chapter we have investigated the catalytic mechanism of the GH130 inverting mannoside phosphorylase "Unknown human gut bacterium mannoside phosphorylase", UhgbMP, with a disaccharide substrate Man- β -1,4-GlcNAc, a recurring element in human epithelium lining *N*-glycans. The initial hypothesis was that UhgbMP follows a substrate-assisted mechanism where the last proton donor of the glycosidic bond oxygen atom is the 3-OH exocyclic group of the β -mannosyl moiety of the substrate. What motivated this hypothesis is that structural data showed the putative acid catalyst, Asp104, far from the scissile bond, an unusual orientation compared to other inverting glycoside phosphorylases which made unlikely the direct proton transfer. However, this residue seemed to be able to interact with the Man 3-OH, which in turn could interact with the scissile bond O, in a double hydrogen bond network. The simulations confirmed that this network is indeed stable at the active site of UhgbMP and provided an atomistic view of the catalytic mechanism, confirming the experimental interpretation as well as providing the complete conformational itinerary of the mannosyl substrate.

The QM/MM metadynamics simulations of this study unveiled an important step of the degradative mechanism of *N*-glycan fragments by intestinal bacteria, involved in onset and evolution of inflammatory processes that compromise the integrity of the intestinal epithelium lining. The characterized mechanism and conformational itinerary of the substrate can give clues for the development of specific inhibitors

targeting these processes. The obtained results can be extended to other phosphorylases with shared fold, such as *Bf*MGP [53], the β -1,2-mannobiose phosphorylases *Lin*0857 [195] and *Bt*3780, the β -1,4-mannooligosaccharide phosphorylase *Ra*MP2 [172] and the GT108 mannose phosphorylases [170].

The main conclusions of this Chapter are the following:

- UhgbMP phosphorylates its natural substrates through a substrate-assisted, inverting mechanism, where the distal catalytic acid Asp104 participates in a proton relay network with the β -mannosyl moiety, thus being the last proton donor to the leaving group its 3-OH substituent.
- The precise positions of the hydrogens at the UhgbMP-substrate complex define the proton network that results in the protonation of the leaving group.
- The hexameric organization of UhgbMP likely maintains the tight packing of the active site and prevents hydrolysis by entrance of bulk solvent, whereas in the lower order oligomeric states the active site is more flexible and solvent-exposed.
- Although UhgbMP is specific for Man- β -1,4 linkages, it accepts different acceptors such as Glc, Man and Glc due to their ability to sustain interactions with the enzyme subsite *+1* residues Arg59, Tyr103, Met67 and His174, which is coherent with the state-of-the art.
- The β -mannosyl substrate undergoes a conformational itinerary on the equator of the Cremer-Pople sphere, $^1S_5/B_{2,5} \rightarrow [B_{2,5}]^\ddagger \rightarrow ^OS_2$, typical of inverting β -mannosidases.
- The atomistic detail of the mechanism obtained for UhgbMP can be used for guiding the development of specific inhibitors targeting gut inflammatory processes, as well as enzyme engineering strategies for mannose oligosaccharide recognition. The results can most likely be extended to GH130 and GT108 members that share the active site organization.

4.4 Computational details

The systems modeled in the present Chapter correspond to UhgbMP in complex with several substrates. All trajectory analyses were carried out using VMD version 1.9.3 (<https://www.ks.uiuc.edu/Research/vmd/>) [130], the CPPTRAJ module [131] of AMBER [133] and the open-source, community developed PLUMED library [149], version 2.7 [150]. The setup for classical MD systems (Sections 4.4.1 to 4.4.4) was designed by Dr. Mercedes Alfonso-Prieto, who started this project as a former Postdoc in the group of Prof. C. Rovira, as well as the first reaction (Sections 4.4.5 and 4.4.6).

4.4.1 System preparation

Designed by Dr. M. Alfonso-Prieto

The Michaelis complex of UhgbMP in complex with inorganic phosphate (Pi), β -mannose- β -1,4-*N*-acetylglucosamine was built using two of the available X-ray structures of the enzyme. The structure with PDB accession code 4UDK (resolution 1.76 Å) was used for modeling the enzyme, Pi and GlcNAc moiety, and the structure with PDB 4UDJ (resolution 1.94 Å), for the β -mannosyl moiety at subsite -1 [58]. The β -Man moiety at the 4UDK structure had low occupancy and was in a position not suitable for the predicted mechanism. By superimposing both structures and adding the β -1,4 glycosidic bond between the Man and GlcNAc moieties with the LeAP program of AMBER [133], a *bona fide* Michaelis complex was built. Crystallographic water molecules were kept in the model. Regarding the protonation states of titratable residues, the propKa [136] and H++ [137] programs were used together with MolProbity [196, 197] and visual inspection of their close environment. Special attention was given to the Asp residues in the active site (see Section 4.4.9). The predicted acid catalyst, Asp104, was assigned in its protonated carboxylic acid form to establish the double hydrogen bond network with the Man 3-OH and also considering the computed pK_a shifts (Table S4.1). Regarding the protonation state of Pi, it was set as monohydrogenphosphate (HPO_4^{2-}). The second pK_a of inorganic phosphate is 7.2 [190] and the group is coordinated by Arg150, Arg168 and Lys212, all residues with positive charge (Figure S4.1) interacting with the non-protonated Pi O atoms. His321, located in the active site, interacts with the Pi H atom so it was assigned as δ -tautomer. The system was solvated with 164589 water molecules in a box of $121.60 \times 127.25 \times 127.01$ Å, and 248 Na^+ and 182 Cl^- ions were added to achieve neutrality to a concentration of 150 mM.

4.4.2 Classical molecular dynamics of hexameric UhgbMP-Man- β -1,4-GlcNAc

R0 performed by Dr. M. Alfonso-Prieto. R1-3 performed by I. Cuxart

Classical molecular dynamics of the Michaelis complex of UhgbMP with Man- β -1,4-GlcNAc were carried out using NAMD version 2.9 [198]. Parameters from parm99SB for the protein [140], GLYCAM06 [64], adapted parameters from phosphorylated Ser for the Pi [199]. The TIP3P model was used for water [141] and Åqvist parameters for Na⁺ and Cl⁻ [200]. The general protocol for the forcefield simulations was the following: First, an energy minimization, followed by heating the system to room temperature (300 K). Restraints were applied on the protein backbone and saccharide rings to allow the movement of the amino acid side chains and saccharide substituents while keeping the backbone fixed and avoid large displacements due to the initial dynamics steps. Then, restraints were progressively released during a 15 ns NPT run keeping the temperature at 300 K with a Nosé-Hoover thermostat [201] and the pressure at 1 atm with a Langevin piston Nosé-Hoover barostat [202]. Subsequently, 25 ns of MD were performed applying harmonic restraints on the main protein-substrate interaction distances, progressively decreasing the force constant. Then, the production run continued without any restraints until 200 ns. MD was carried out with multiple timestep integration using r-RESPA [203]. A base timestep of 2 fs was used applying the SHAKE/RATTLE [143] algorithm for constraining bonds with hydrogen atoms, then a secondary timestep of 4 fs was applied for long-range interactions. A cutoff of 9 Å was applied for computing non-bonded interactions. For increasing sampling, we ran 3 additional replicas randomizing the initial velocities from an already equilibrated frame of R0, thus obtaining 800 ns in 4 uncorrelated replicas named R0 to R3 hereafter. Analyses of classical MD trajectories of all 6 active sites located at each subunit across replicas R0-3 showed stable and similar active site configurations (Figures S4.2 to S4.7), with RMSD values below 2 Å (Figures S4.2.b and S4.4) and the Man ring in ¹S₅/B_{2,5} conformation (Figures S4.2.c and S4.4).

4.4.3 Oligomerization state of UhgbMP

Performed by Dr. M. Alfonso-Prieto

The NAMD package was used for all the subsequent classical MD simulations. The enzyme-substrate complex of monomeric and dimeric forms of UhgbMP with Man- β -1,4-GlcNAc was prepared using the same structural data as with the hex-

americ construct (See Section 4.4.2, above). The monomeric system was solvated in a box of dimensions $94.60 \times 99.09 \times 92.60 \times \text{\AA}$ with 80046 water molecules. 86 Na^+ and 75 Cl^- ions were added to achieve neutrality at 150 mM concentration, whereas the dimeric system cell was $105.07 \times 132.57 \times 91.71 \text{\AA}$, with 116217 water molecules, 134 Na^+ and 112 Cl^- ions. The same protocol as described in Section 4.4.2 was applied, except for the simulations time applying restraints was of 80 ns (monomer) and 50 ns (dimer). Unrestrained production runs were carried out up to 100 ns for both systems (Figures S4.8 and S4.9).

4.4.4 Substrate promiscuity of UhgbMP

Performed by Dr. M. Alfonso-Prieto

Two additional complexes with Man- β -1,4-Glc and Man- β -1,4-Man were built starting from the structure of UhgbMP in complex with Man- β -1,4-GlcNAc. While the mannosyl unit at subsite -1 was kept the same, the Glc/Man at subsite $+1$ were modeled by replacing the GlcNAc and changing the 2-OH configuration for Man. The rest of the setup proceeded as described in Section 4.4.1, with GLYCAM06 parameters for all saccharide substrates. Classical MD simulations ran with NAMD, applying the same equilibration protocol as in Section 4.4.2, until 70 ns for each complex (Figures S4.10, S4.11 and S4.12).

4.4.5 QM/MM molecular dynamics

Performed by Dr. M. Alfonso-Prieto

QM/MM MD simulations were performed starting from a suitable classical MD trajectory snapshot. The snapshot was taken from the equilibrated part of replica R0 and the region chosen to model at the QM level corresponds to the active site of monomer A. Hybrid QM/MM MD calculations were performed with the CPMD code (<http://www.cpmc.org>), applying the method which combines Density Functional Theory-based [75] Car-Parrinello MD (QM) with forcefield MD developed by Laio, VandeVondele and Roethlisberger (MM) [78,144]. The QM region includes 77 atoms (net charge -2) corresponding to the full Man- β -1,4-GlcNAc substrate, the inorganic phosphate (HPO_4^{2-}) and the sidechains of Asp104 and His231, interacting with Pi, as well as two monovalent atoms saturating the C-C bonds at the QM-MM boundary region (Figure S4.13). The remaining 195436 atoms remained in the MM region. The QM region was enclosed in an orthorhombic cell of dimensions $24.34 \times 17.46 \times 15.88 \text{\AA}$. Electrostatic interactions near the QM-MM interface were handled with a fully Hamiltonian multilayer electrostatic coupling scheme developed by Laio *et al.* [78] the corresponding radii of 7.41 (NN), 8.47 (MIX) and 10.58 $\text{\AA}$ (ESP).

Density Functional Theory was applied using the Perdew-Burke-Ernzerhof (PBE) exchange correlation functional [71], which was previously reported to perform well in the description of hydrogen bonds [204], carbohydrate ring conformations and carbohydrate active enzyme reactions [16, 73], with a plane-wave basis set with a kinetic energy cutoff of 70 Ry and norm conserving Martins-Trouillier pseudopotentials [74]. After a geometry optimization with a simulated annealing of atomic velocities, a QM/MM MD run of 5 ps with a timestep of 0.12 fs was performed, setting the electronic fictitious mass at 700 au and the temperature at 300 K with a Nosé Hoover thermostat [148] with a frequency of 3500 cm⁻¹. Metadynamics simulations started from a trajectory snapshot of this simulation.

4.4.6 QM/MM metadynamics

Performed by Dr. M. Alfonso-Prieto

The phosphorysis reaction of UhgbMP with Pi and Man- β -1,4-GlcNAc was modeled with QM/MM metadynamics [81, 83] starting from a trajectory frame of the unbiased QM/MM MD. Two CVs corresponding to interatomic distances were defined to drive the history-dependent bias over the reaction FEL and discriminate its relevant states (Figure S4.14.a). CV1 is the distance between O_{4P} (the Pi oxygen atom closest to Man) , and the β -Man C1. CV2 is the glycosidic linkage distance (C1-O1). Thus, CV1 drives the Pi nucleophilic attack and CV2, the cleavage of the scissile bond. To let the system evolve freely through any possible proton transfer pathways, this process was not included in any of the CVs. For building up the bias Gaussian-like functions were deposited every 18 fs (150 MD steps) setting the width at 0.10 Å (CV1) and 0.05 Å (CV2). The deposition pace and Gaussian with based on the oscillations of the CVs during the previous QM/MM MD run (Section 4.4.5). The initial Gaussian height was set at 1.0 kcal mol⁻¹. The simulation was stopped after the deposition of 1684 Gaussians, which amounts to approximately 30 ps of simulation (Figure S4.14.b and .c). The reaction free energy barrier was validated with the experimental rate value, $k_{cat}(\text{Man-}\beta\text{-1,4-chitobiose}) = 0.072 \text{ s}^{-1} = 19.8 \text{ kcal mol}^{-1}$ [13]. For refining the estimation of the computed free energy barrier, the QM/MM metadynamics simulation was restarted from a point where the reactants well of the FEL was half full changing the Gaussian height to both 0.5 and 0.25 kcal mol⁻¹, as in the procedure described by Nair *et al.* [86].

The reaction TS structure was further validated with a committor analysis [205] performed by Dr. M. Alfonso-Prieto. An ensemble of data points was extracted close to the saddle point of the reaction FEL (Figure 4.10.b). QM/MM MD simulations were performed applying random initial velocities assessing the probability of each

structure of falling back to the reactants state or evolving to products (Figure S4.15). The parameters for the QM/MM simulations follow the setup of Section 4.4.5.

4.4.7 Additional QM/MM simulations

Performed by I. Cuxart

Additional QM/MM MD and metadynamics simulations were performed starting from alternative snapshots of the classical MD simulations. First, a starting snapshot was chosen from the end of replica R0, although the active site D of the complex was selected for modelling the QM region. This subunit was chosen because, although active site RMSD and overall phosphorolysis reaction distances are stable in all classical MD simulations (Figures S4.2 and S4.3), subunit D was especially stable regarding subsite +1 interactions, which transiently can involve an adjacent subunit (His174 from the same subunit vs. His235' from a different one) (Figure 4.8.c, bottom left). Second, another snapshot was selected from the end of the classical MD replica R1, choosing subunit F for the QM region. The CPMD code was used to perform QM/MM MD simulations following the protocol in Section 4.4.5.

4.4.8 Additional QM/MM metadynamics simulations

Performed by I. Cuxart

After QM/MM MD equilibration, the phosphorolysis reaction was simulated coupling the CPMD QM/MM MD to the open-source PLUMED library, version 2.3.3 [149, 150], biasing the same 2 CVs described in Section 4.4.6, with the same parameters except for the Gaussian deposition time which was increased to 36 fs (300 MD steps). The initial Gaussian height was set at 1.0 kcal mol⁻¹ and was then decreased to 0.25 kcal mol⁻¹ and 0.5 kcal mol⁻¹ restarting the simulation from a half-full reactants well (Section 4.4.6). The simulation was stopped after observing one recrossing event from products back to reactants (first recrossing criterion), after approximately 140 ps and 3707 deposited Gaussians. The computed free energy barrier amounts to 21.36 ± 2.02 kcal mol⁻¹, with the same reaction mechanism as in the previous simulation observed (Figure S4.16). For the simulation starting from the R1 frame, it was also stopped following the first recrossing criterion at approximately 115 ps of simulation and after depositing 3071 Gaussians. The computer free energy barrier amounts to 21.51 kcal mol⁻¹ (Figure S4.17).

4.4.9 Protonation state of Asp104

4.4.9.1 Molecular dynamics simulations of different scenarios

Performed by both Dr. M. Alfonso-Prieto and I. Cuxart

In all previous classical MD simulations Asp104 was set in its carboxylic acid form. To assess this choice, we prepared additional systems considering different protonation scenarios and subjected them to classical MD. We built three additional complexes contemplating different possibilities: On one hand, the Michaelis complex of UhgbMP with Man- β -1,4-GlcNAc with Asp104 in the carboxylate form (not protonated, MC'). On the other hand, two product complexes also considering the different possibilities. First, the product complex as obtained in the QM/MM metadynamics simulations (P): A metadynamics trajectory snapshot with UhgbMP with α -Man-1-P and Asp104 as carboxylate was selected and the appropriate topology to run classical MD simulations with NAMD was built with forcefields parm99SB [140], GLYCAM06 (saccharide) [64]. For the alternative state, the neutral Asp104 was assigned to this structure (P'). After an equilibration following the protocol in Section 4.4.2, 200 ns of unrestrained MD were carried out for each system.

4.4.9.2 Computed estimations of pK_a values

Performed by I. Cuxart

The web server-hosted H++ program (newbiophysics.vt.edu/H++) was used to perform pK_a calculations of the acid catalyst Asp104. We computed pK_a values for all titratable residues of the protein. We discussed residues Asp104, Asp304, Asp76, Asp187 and Asp112 in Table S4.1. The two first residues are in the active site of UhgbMP, whereas the rest are at the surface to validate the pK_a tendencies. Each pK_a value of Table S4.1 is averaged from the results of three structures corresponding to QM/MM MD trajectory snapshots. For the unliganded UhgbMP, Pi and Man- β -1,4-GlcNAc were removed from the structure. For the Michaelis complex, charges and radii of the ligand atoms were added to the PQR file along to atomic coordinates following the H++ protocol. Charge values were obtained from the parameter file used in the classical MD simulations described in Section 4.4.1. The product complex structure is a snapshot from the metadynamics simulation of Section 4.4.6, and the non-protein part of the system was included in the calculation likewise. For simplicity, only one enzyme subunit was considered after testing that the tendencies were consistent in a dimer and tetramer. Calculations were performed at pH=7.0 and saline concentration 150 mM NaCl, external dielectric was set at 80 and the internal dielectric at 4 after screening values between 3 and 5.

4.5 Supplementary information

Residue	States		
	Unliganded	MC	P
Asp104 (catalytic)	0.0	7.4	0.0
Asp304 (active site)	2.5	0.0	5.2
Asp76 (solvent-exposed)	3.2	3.2	3.2
Asp187 (solvent-exposed)	2.8	2.8	2.7
Asp112 (solvent-exposed)	0.8	0.8	1.1

Table S4.1: Estimated pK_a values computed with H++ [137] of Asp residues in different regions of the protein. Besides the catalytic Asp104, Asp304 is in the active site close to Asp104 therefore likely to act as a pK_a modulator, and interacting with the Man 4-OH and 6-OH groups. Asp76, Asp112 and Asp187 are at the surface of UhgbMP and interact with polar residues (both Asp76 and Asp187 with neutral and Asp112 with an arginine).

CHAPTER 4. THE CATALYTIC MECHANISM OF GH130
N-GLYCAN MANNOSIDE PHOSPHORYLASE

Enzyme	GH130 subfamily	Oligomeric state	Ligands	PDB code (resolution, Å)	Ref.
<i>Bf</i> MGP	1	Hexamer	β -Man-1,4- β -Glc and Pi	3WAS (1.50)	[53]
			β -Man, β -Glc and Pi	3WAT (1.60)	
			α -Man-1-P	3WAU (1.70)	
<i>Ra</i> MP1	1	Trimer	β -Man-1,4- β -Glc and sulphate	5AYC (1.90)	[172]
UhgbMP	2	Hexamer	β -Man, α -GlcNAc and Pi	4UDK (1.76)	[58]
			β -Man and Pi	4UDJ (1.60)	
			α -GlcNAc and Pi	4UDG (1.60)	
<i>Ra</i> MP2	2	Hexamer	β -Man-1,4- β -Man and Pi	5AYE (2.20)	[172]
<i>Lin</i> 0857	NC	Dimer	β -Man-1,2- β -Man and sulphate	5B0R (1.80)	[195]
			β -Man-1,2- β -Man-1,2- α -Man and sulphate	5B0S (2.10)	
			β -Man and sulphate	5B0Q (2.30)	
<i>Bt</i> 3780	NC	Monomer	α -Man and β -Man	5A7V (1.35)	[176]
<i>Teth</i> 514_1789	NC	Monomer	α -Man and Zn ²⁺	7FIQ (2.22)	[177]
			α -L-Gul-1,4- α -L-All and Zn ²⁺	7FIR (2.20)	
			α -Man-1-P and Zn ²⁺	7FIS (2.19)	

Table S4.2: Available crystal structures of GH130 members in complex with carbohydrate ligands. Data extraction: CAZY, (http://www.cazy.org/GH130_structure.html), last update on 12.09.2022, Table adapted from [33]. Standard carbohydrate abbreviations are employed for carbohydrates. Ligands are D- isomers unless the contrary is specified.

Enzyme	PDB ID	O _{Asp} -O3 _{Man} (Å)	O3 _{Man} -O1 Å
<i>Bf</i> MGP	3WAS	2.62 ± 0.01	2.73 ± 0.02
	3WAT	2.67 ± 0.00	2.83 ± 0.01
	3WAU	2.75 ± 0.04	-
<i>Ra</i> MP1	5AYC	2.58 ± 0.00	2.79 ± 0.00
UhgbMP	4UDK	-	-
	4UDJ	2.48 ± 0.09	3.59 ± 0.12
<i>Ra</i> MP2	5AYE	2.86 ± 0.06	2.68 ± 0.12
<i>Lin</i> 0857	5B0R	2.61 ± 0.05	2.86 ± 0.02
	5B0S	2.57 ± 0.05	2.67 ± 0.01
<i>Bt</i> 3780	5A7V	2.73 ± 0.01	2.82 ± 0.07
<i>Teth</i> 514_1789	7FIS	2.81 ± 0.14	-

Table S4.3: Distance (in Å) between the atoms that form the conserved H-bond network connecting the putative acid base (O_{Asp}), the Man O3 (O_{Asp102}-O3_{Man}) and the scissile bond O (O3_{Man}-O1). Values are averaged across the enzyme subunits of each crystal structure. Structures correspond either to complexes with Man and saccharide at subsite *+1* bonded (reactants state, 3WAS, 5AYC, 5AYE, 5B0R, 5B0S) or unbonded (products state, 4UDK 4UDJ 5A7V). Not included GH130 structures without Man at subsite *-1* (4UDG and 5B0Q at Table S4.2), or with low occupancy (4UDK), and complexes with Zn²⁺ and inhibitors (see 7FIQ and 7FIR at Table S4.2). Adapted from [33].

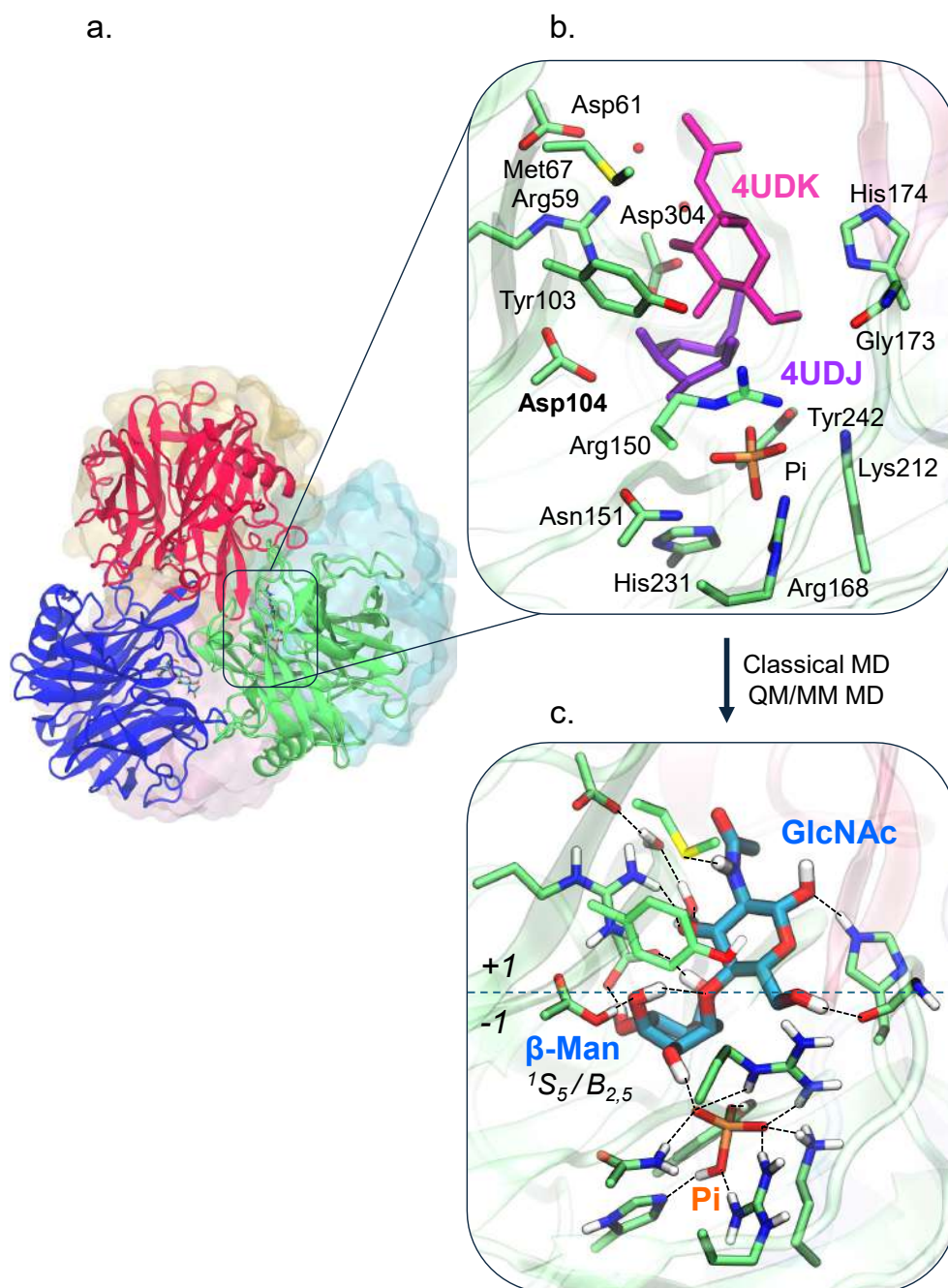


Figure S4.1: a. UhgMP crystal structure. Each protomer is represented in different color, the protomers at the back as transparent surfaces for clarity. b. Detail of the active site corresponding to protomer A after superimposing structures 4UDJ (GlcNAc, magenta sticks) and 4UDK (Man, purple sticks) for building the UhgMP-Man- β -1,4-GlcNAc complex. Since the occupancy of Man was <1 in the 4UDK structure, the 4UDJ coordinates for Man moiety were taken instead. c. Detail of the active site interactions formed in the modeled UhgMP Michaelis complex and conformation of the β -Man after classical and QM/MM MD simulations.

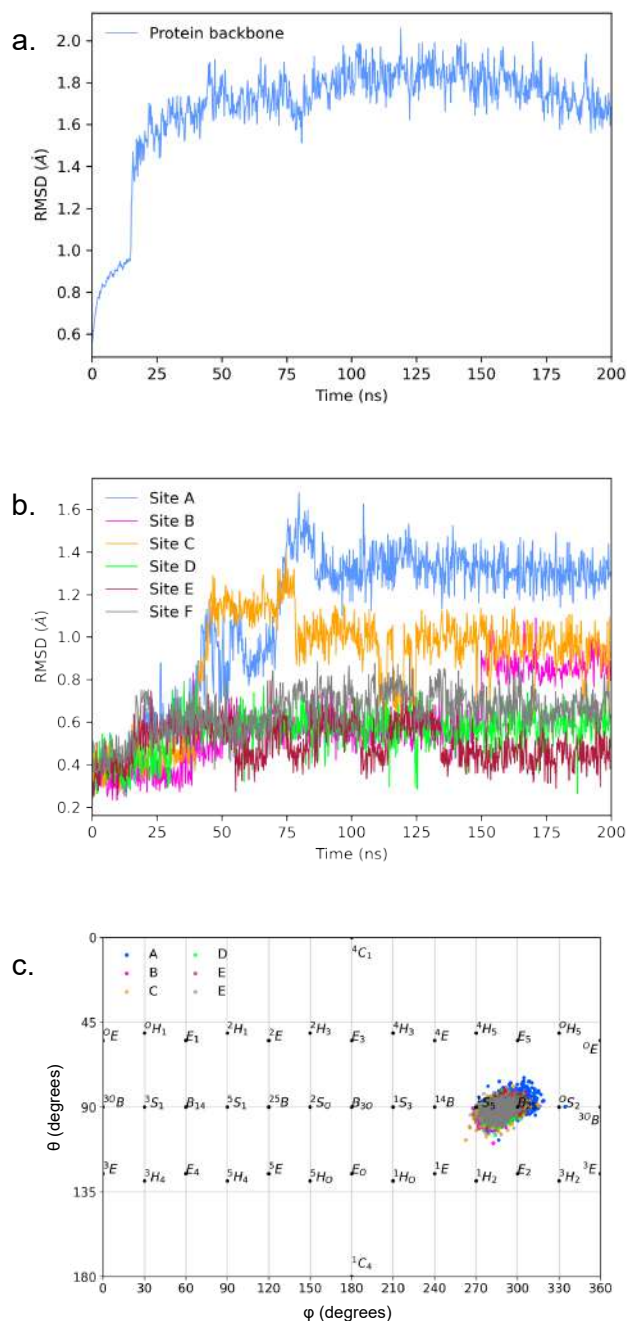


Figure S4.2: Classical MD simulation of the hexameric form of UhgbMP in complex with Man- β -1,4-GlcNAc (first simulation, named R0). a. Evolution of the protein backbone RMSD of protomer A along simulation time, stable around 1.8 Å after 200 ns of simulation. The initial jump in RMSD values is caused by the removal of distance-based restraints. b. Time evolution of the RMSD of the active site of each protomer of the homohexamer. Each computed RMSD included Man- β -1,4-GlcNAc, Asp104 and Pi. c. Mercator projection of the pyranose conformations explored by β -Man ring during classical MD, stable between 1S_5 and $B_{2,5}$ in all subunits.

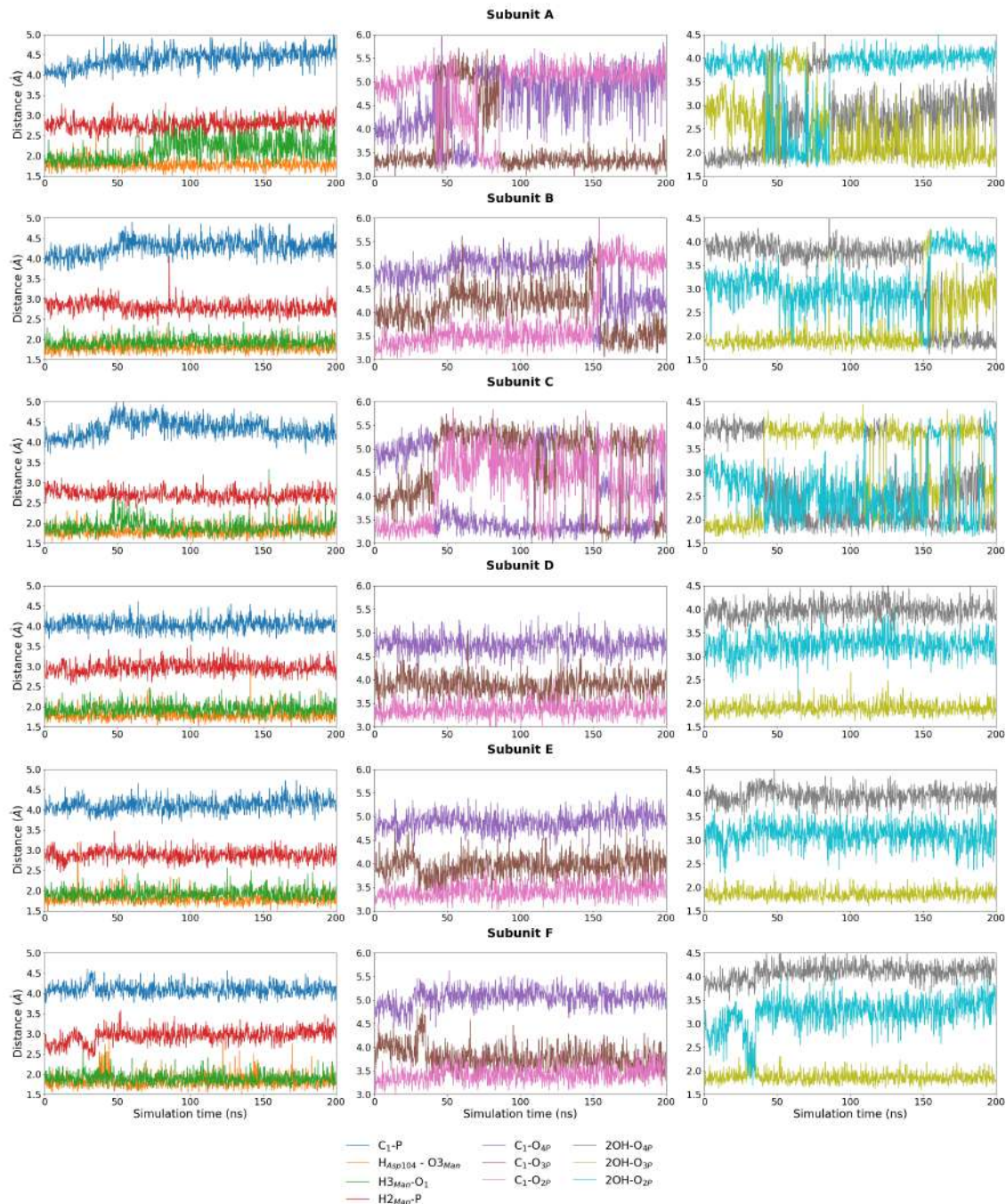


Figure S4.3: Evolution of the main active site distances during classical MD simulation of the UhgbMP hexamer in complex with Man- β -1,4-GlcNAc (first simulation, named R0). Each row corresponds to the active site at each subunit (A-F). First column: Interaction distances involved in the hypothesized substrate-assisted phosphorylisis mechanism. Second column: C1- O_{xP} , $x=2-4$, showing that although Pi can rotate, one O_{xP} atoms is always close to C1. Third column: interactions between the β -Man and the Pi oxygens 2 to 4.

CHAPTER 4. THE CATALYTIC MECHANISM OF GH130 N-GLYCAN MANNOSIDE PHOSPHORYLASE

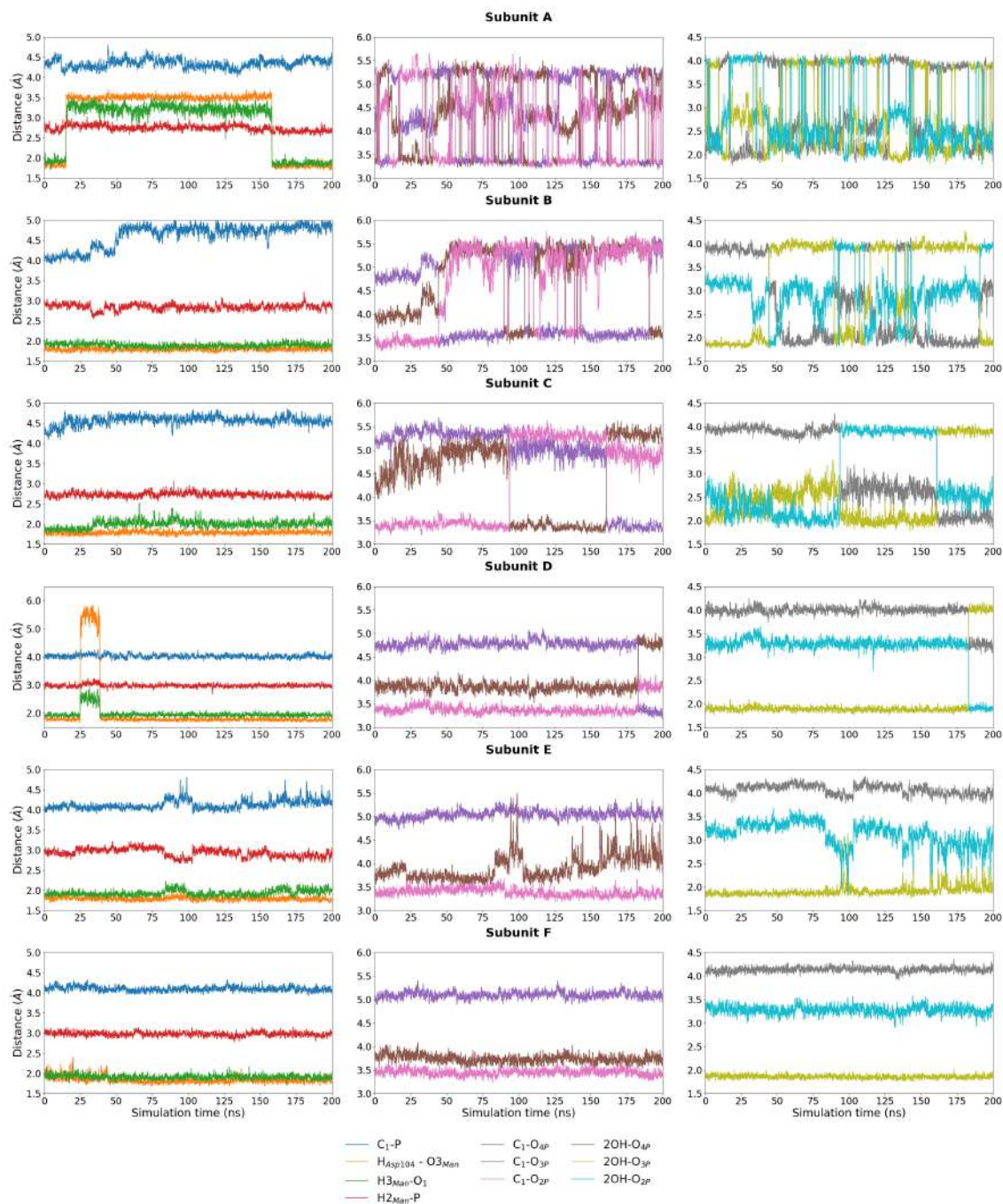


Figure S4.5: Time evolution of main substrate-enzyme interactions during classical MD of replica R1. For more details about each column, see Figure S4.3.

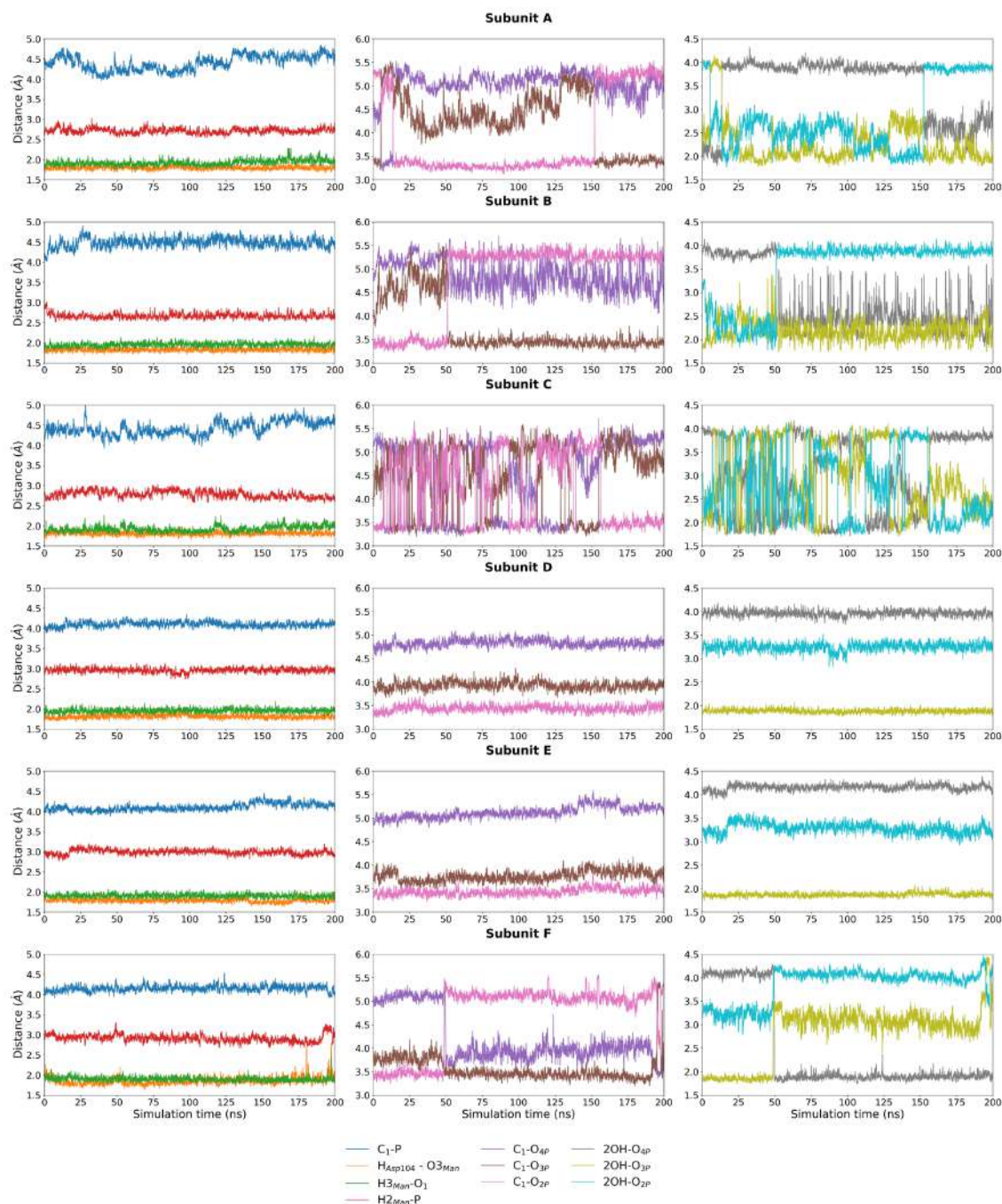


Figure S4.6: Time evolution of main substrate-enzyme interactions during classical MD of replica R2. For more details about each column, see Figure S4.3.

CHAPTER 4. THE CATALYTIC MECHANISM OF GH130 N-GLYCAN MANNOSIDE PHOSPHORYLASE

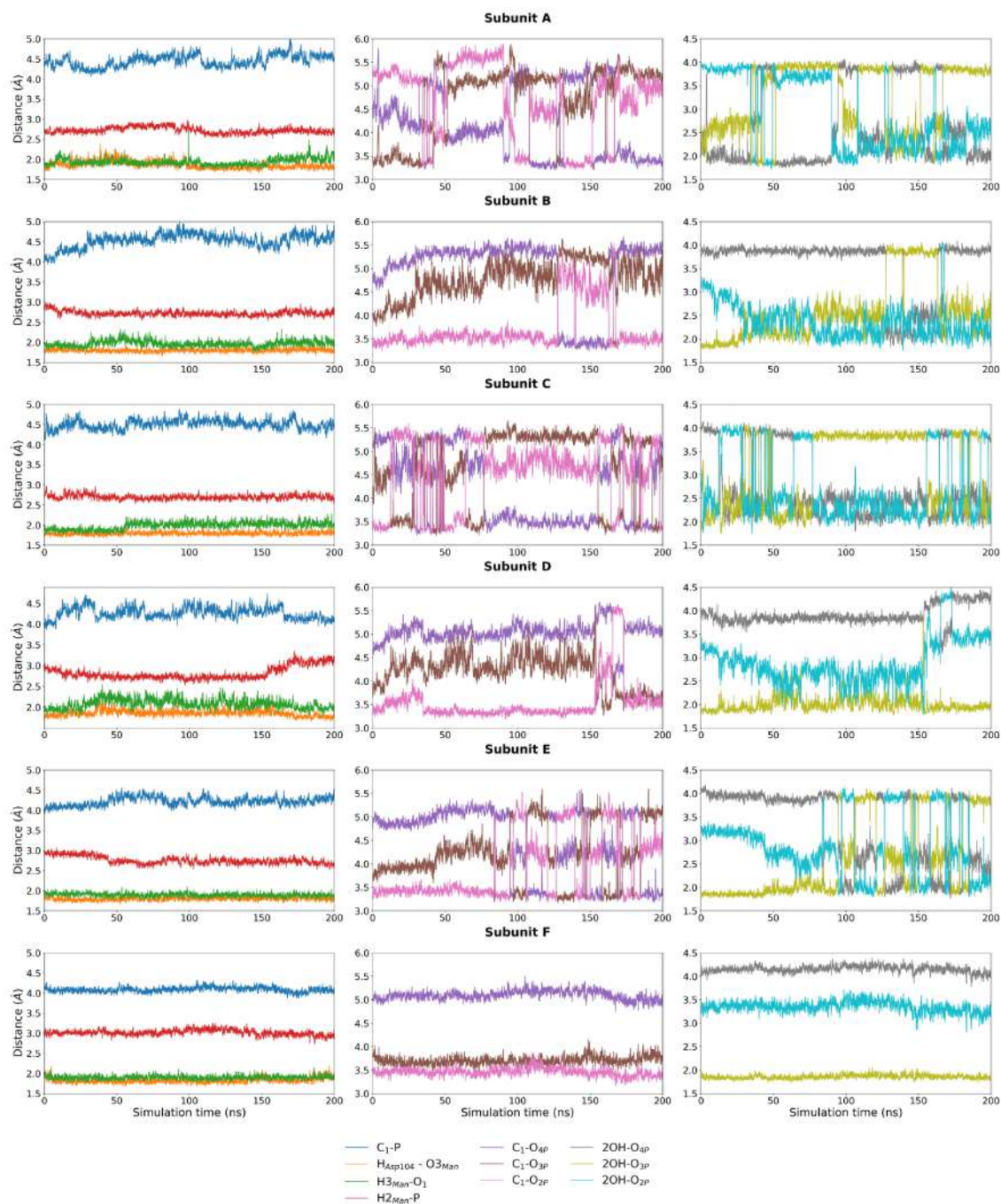


Figure S4.7: Time evolution of main substrate-enzyme interactions during classical MD of replica R3. For more details about each column, see Figure S4.3.

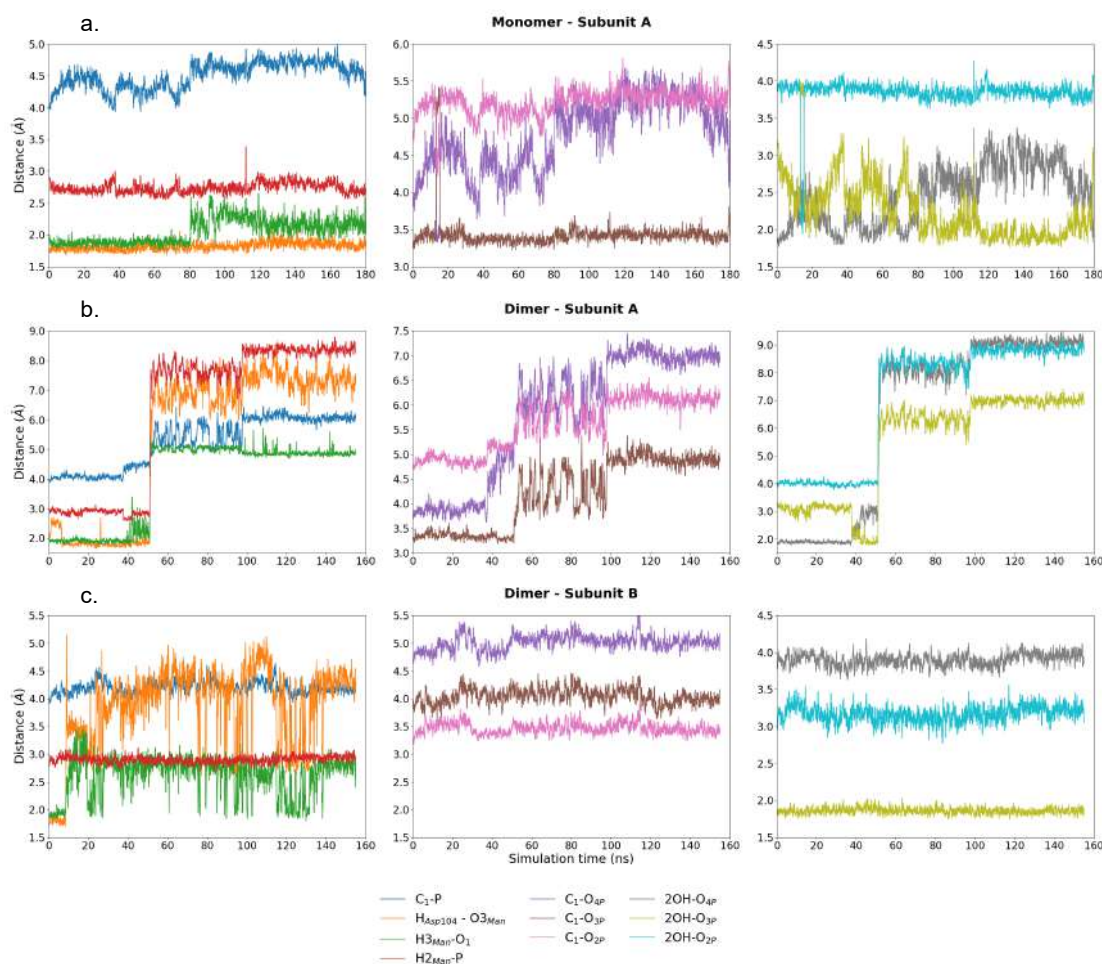


Figure S4.8: Classical MD simulations of the a. monomeric and b,c. dimeric forms of UhgbMP in complex with Man- β -1,4-GlcNAc. As in Figure S4.3, first column represents the main interaction distances involved in the 3-OH substrate-assisted mechanism, second column Pi- $C1_{Man}$ distances and third column Pi-2- OH_{Man} . Note that restraints were removed for the last 100 ns of each simulation, reflected in larger fluctuations of the represented distances than in the hexameric form.

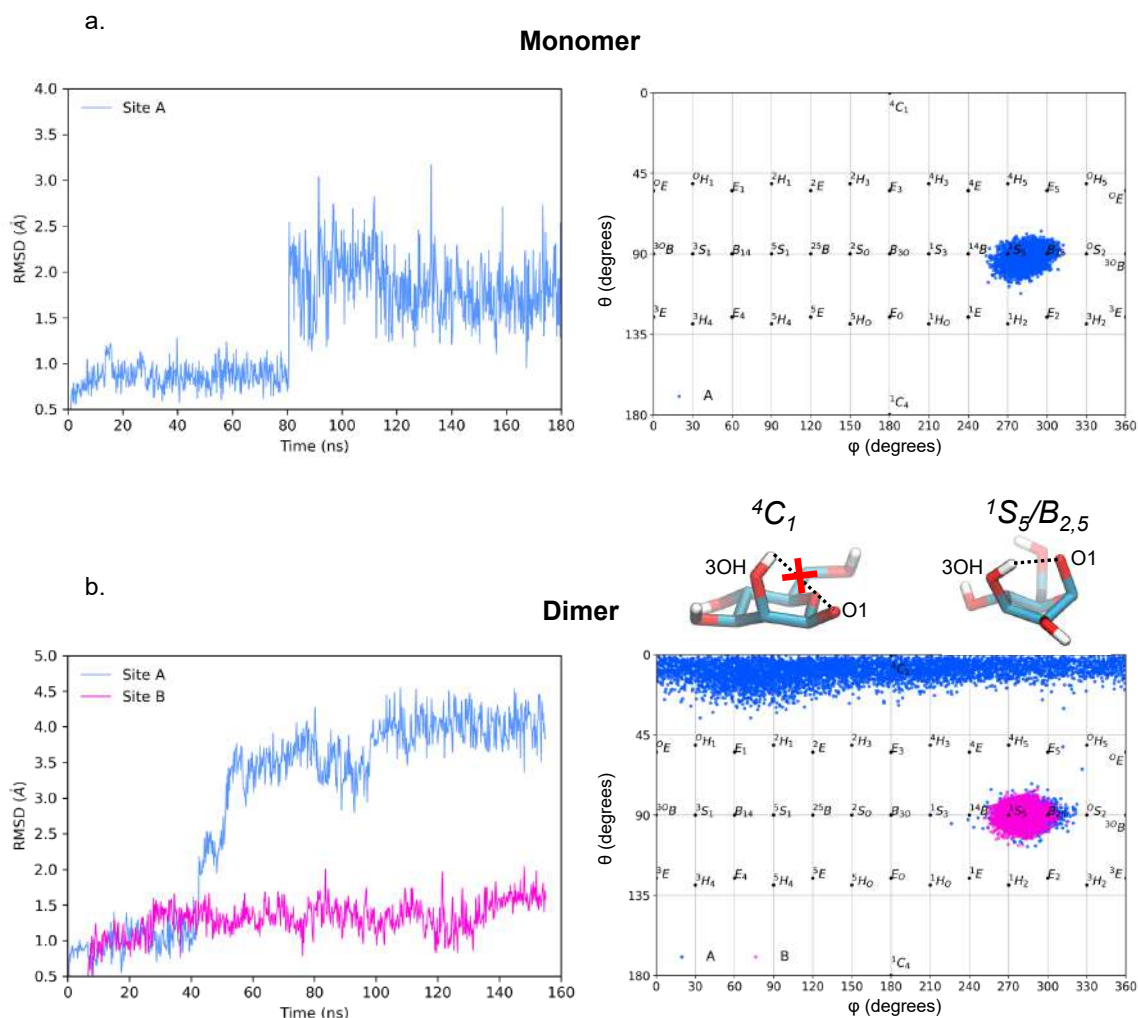


Figure S4.9: Left, RMSD evolution (substrate, Asp and Pi group) and right, Mercator projections of the conformations adopted by the β -Man moiety at subsite -1 of classical MD simulations of monomeric and dimeric forms of UhgMP in complex with Man- β -1,4-GlcNAc. a. Monomer and b. Dimer, active sites A and B. Note that in the dimeric forms in active site A the distortion is lost, and the β -Man adopts the 4C_1 conformation, which disrupts the H-bond network from Man 3-OH to O1 (stick representations).

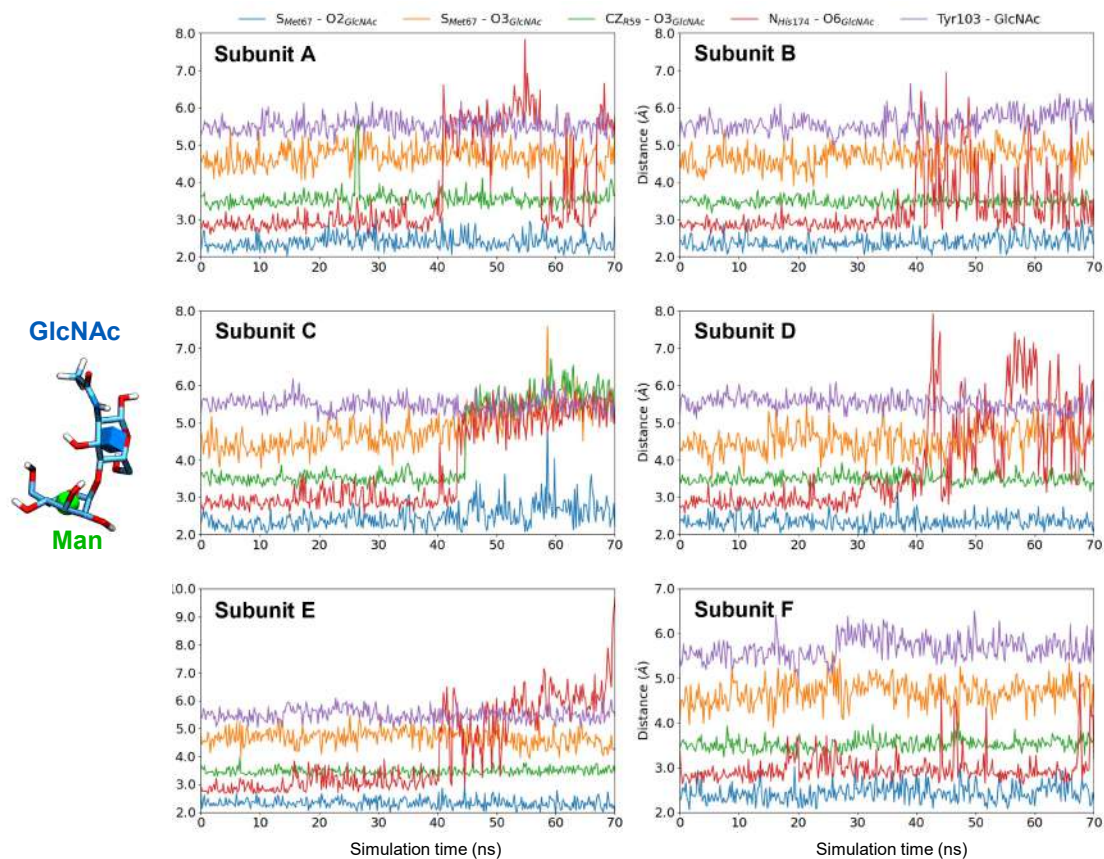


Figure S4.10: Time evolution of selected interaction distances between the acceptor sugar and residues at subsite +1. In this case the acceptor is GlcNAc, and the trajectory corresponds to the first 70 ns of the initial classical MD replica, R0, for the active site at each protomer of the hexameric form of UhgbMP.

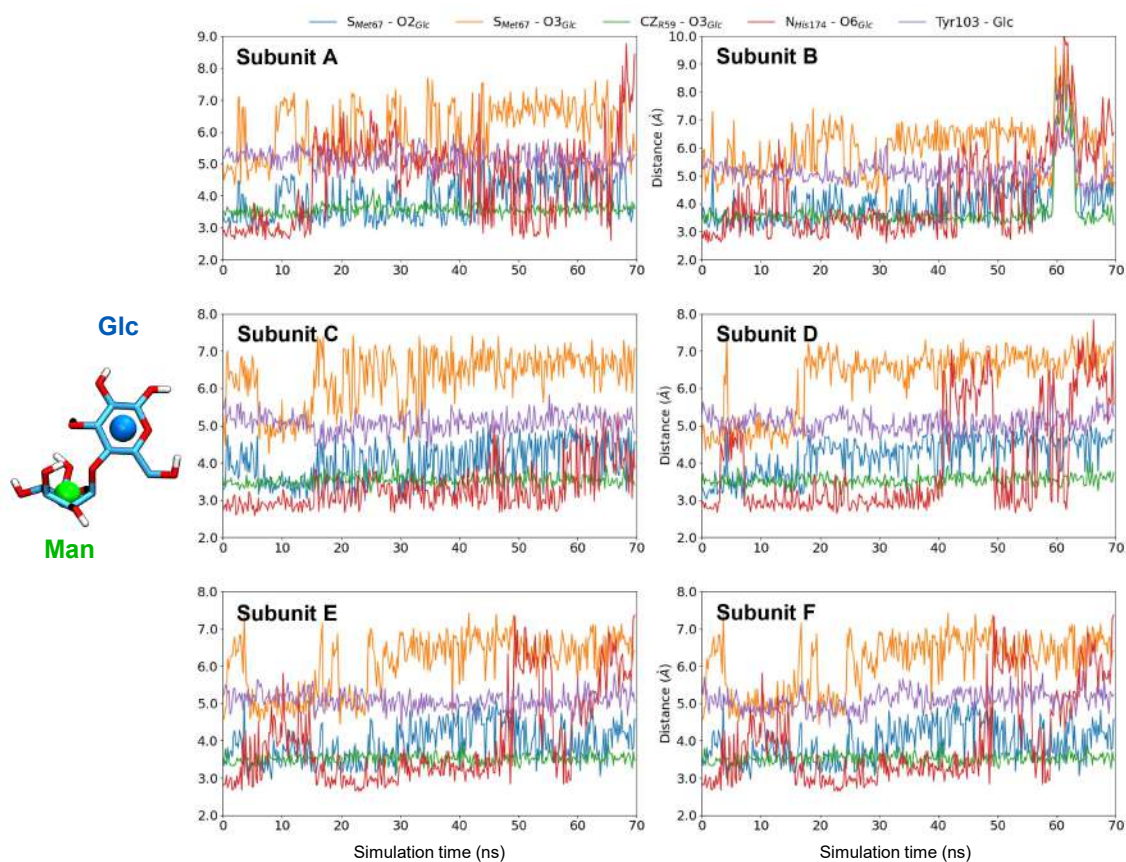


Figure S4.11: Time evolution for acceptor-enzyme interaction distances at subsite +1 during 70 ns of classical MD. In this case, the acceptor is β -1,4-Glc, with one panel for each active site.

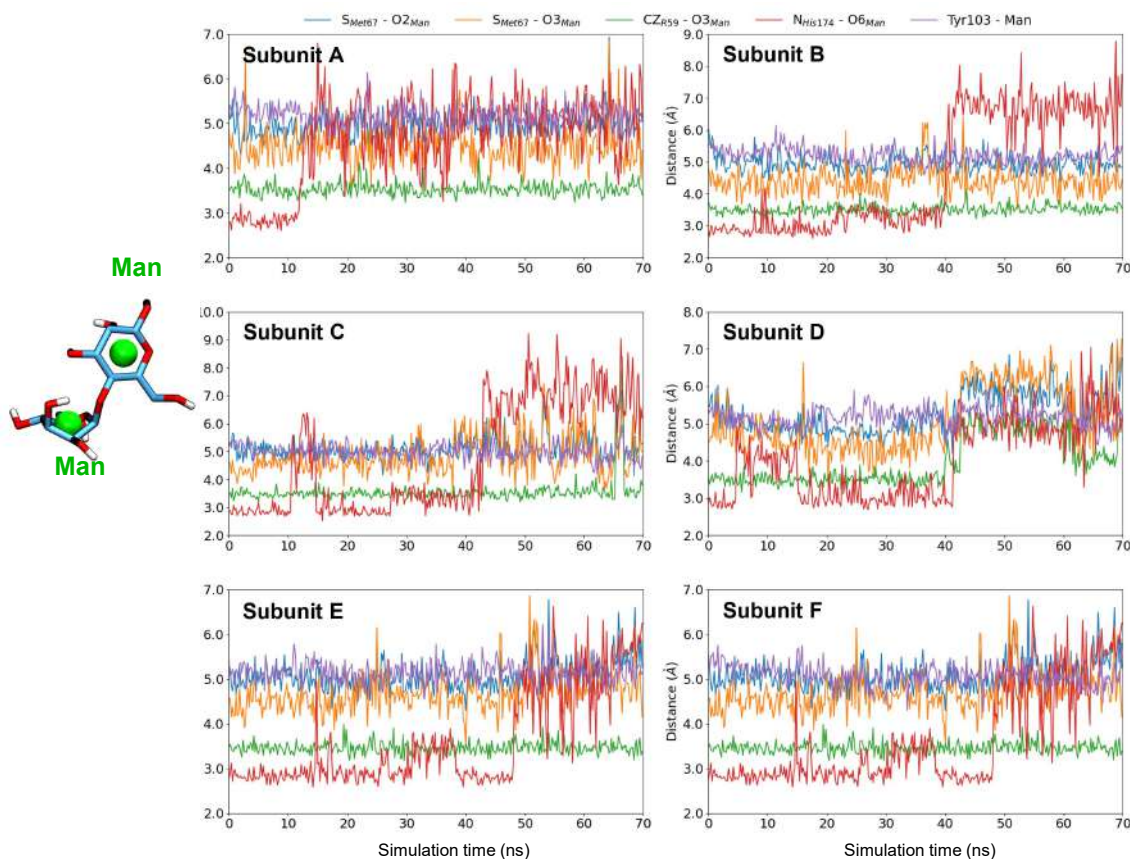


Figure S4.12: Time evolution for interaction distances between acceptor and enzyme $+1$ subsite residues during 70 ns of classical MD, for the active site at each protomer of UhbMP. In this case, the acceptor is β -1,4-Man.

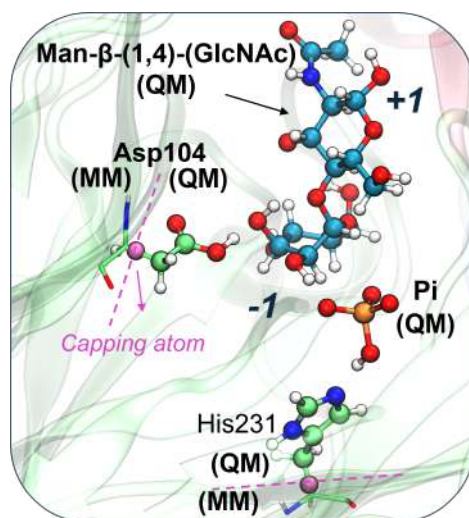


Figure S4.13: Scheme of the QM region selected for QM/MM MD simulations of UhbMP Man- β -1,4-GlcNAc complex. The QM region consists in 77 atoms, as ball-and-stick representation, including the full substrate, the side chains of Asp 104 and His231, the Pi group and H capping atoms placed to saturate the valences of the C-C bonds separated by the QM/MM boundary in Asp104 and His231 (mauve balls).

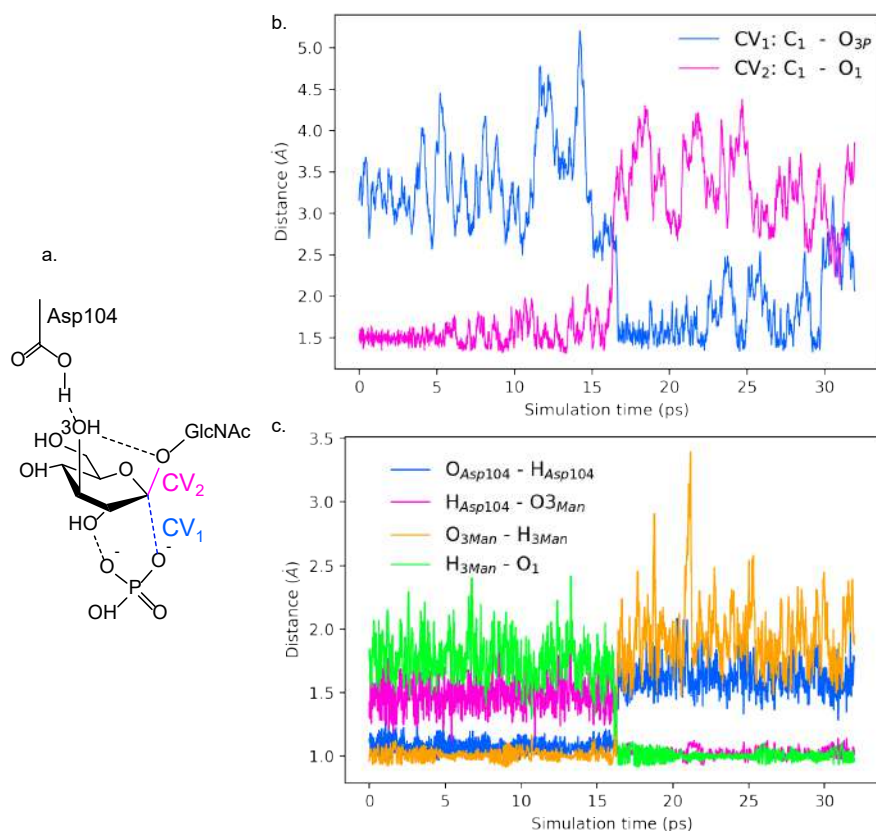


Figure S4.14: QM/MM metadynamics simulation of the UhgbMP phosphorolysis as described in section 4.2.2. a. Scheme of metadynamics CVs, defined as the distances involved in covalent bond formation ($O_{4P}-C1$) and cleavage ($C1-O1$). b. Evolution of CV1 and CV2 over the simulation. c. Evolution of the distances involved in the proton relay Asp104-3OH-O1.

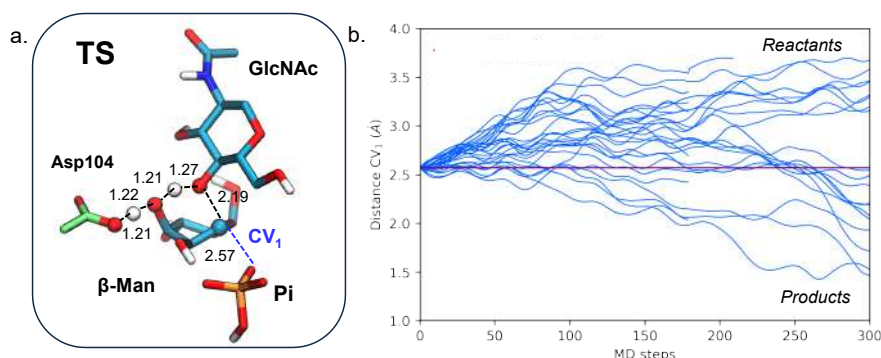


Figure S4.15: Results of the committer analysis. a. Structure extracted from the reaction FEL for running independent QM/MM MD simulations that yielded approximately a 1:1 ratio of collapsing to reactants or products state. This structure, with $O_{4P}-C1 = 2.57$ Å and $C1-O1 = 2.19$ Å, is close to the MEP computed TS (Figure 4.10.b). b. Blue: distance evolution of the evaluated collective variable (CV1: $O_{4P}-C1$) along the QM/MM MD committer simulations. Violet: value of CV1 of the initial TS structure.

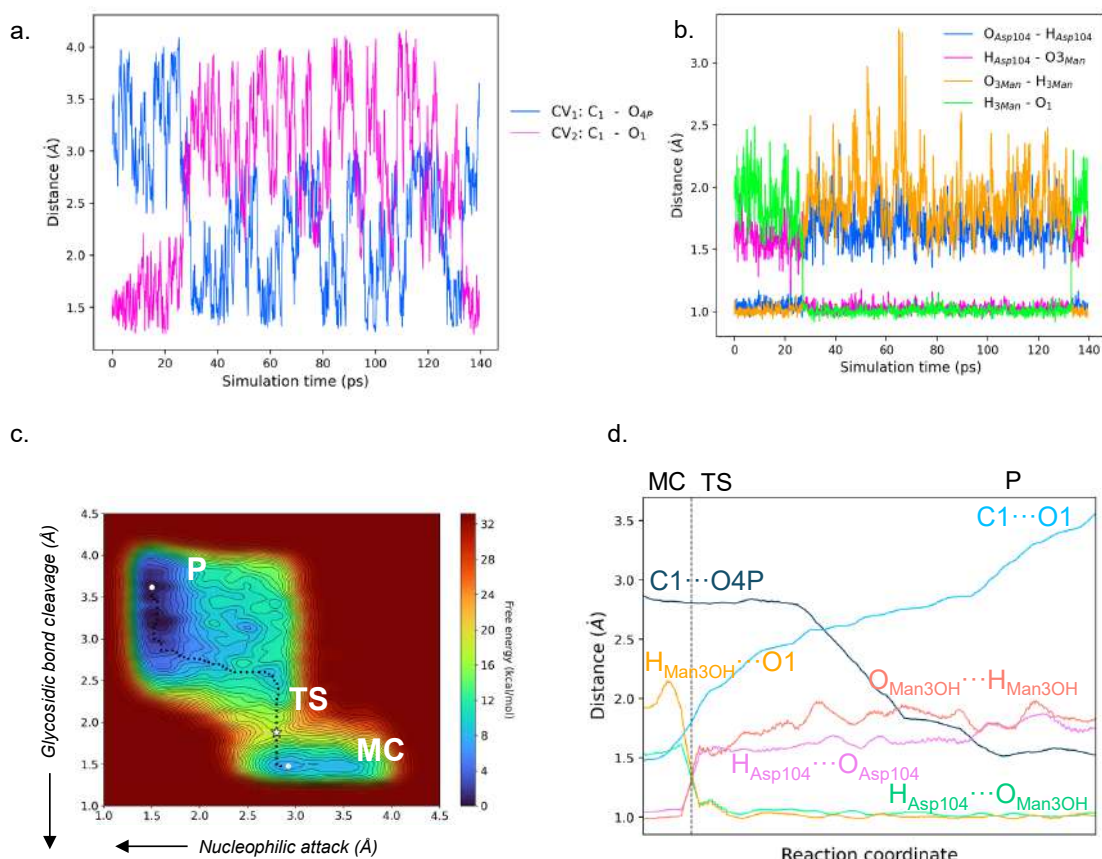


Figure S4.16: QM/MM metadynamics simulation of the phosphorolysis reaction by UhgbMP starting from an alternative snapshot of R0 and selecting the active site in subunit D for the QM region. More details of setup in Section 4.4.7. a. Evolution along the simulation of a. both CVs, b. all distances involved in the double H-bond between Asp104 and the substrate, c. Reconstructed FEL, with a 21.36 ± 2.02 kcal mol⁻¹ barrier (average of simulations performing the crossing with Gaussian heights of 1.0, 0.5 and 0.25 kcal mol⁻¹ and d. evolution of the main interatomic distances involved in the reaction along its MEP.

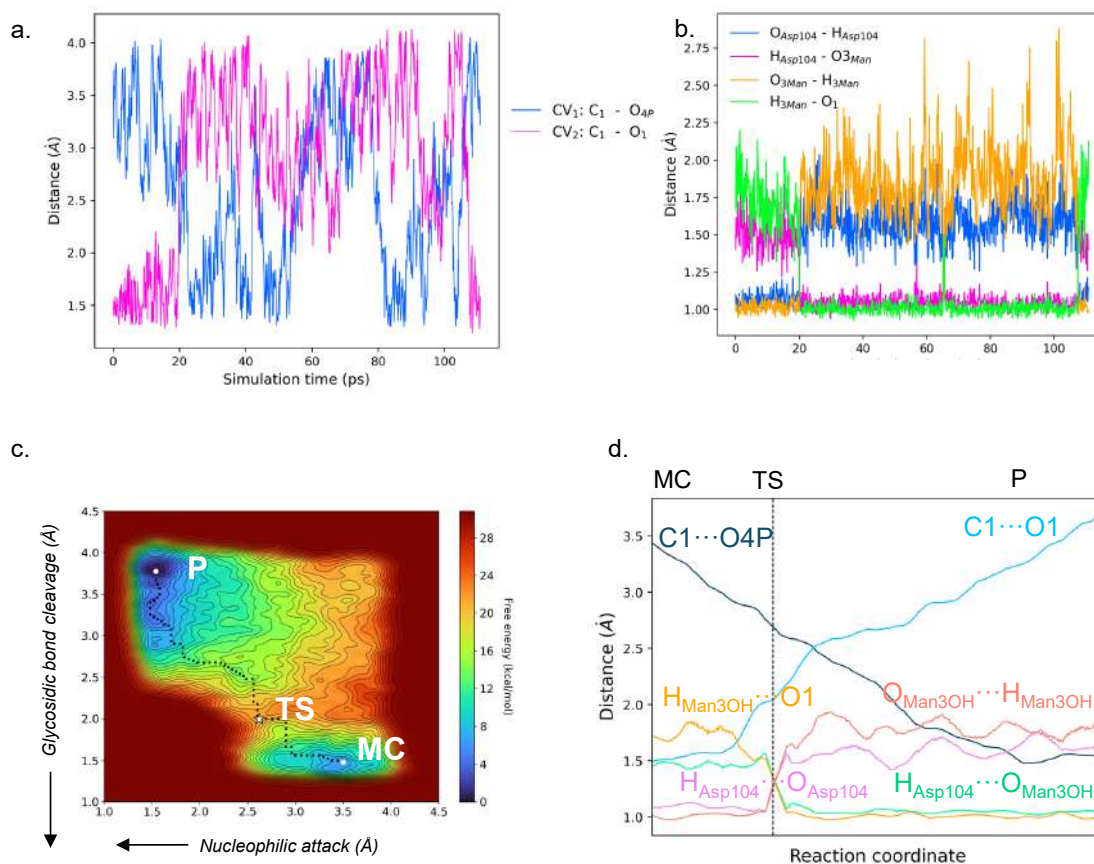


Figure S4.17: QM/MM metadynamics simulation of the phosphorylase reaction by UhgBMP starting from trajectory snapshot of the classical MD R1 and selecting the active site in subunit A for the QM region. More details of setup in Section 4.4.7. a. Evolution along the simulation of a. both CVs, b. all distances involved in the double H-bond between Asp104 and the substrate, c. Reconstructed FEL, with a 21.51 kcal mol⁻¹ barrier. d. MEP evolution of the main distances involved in the reaction.

Chapter 5

The catalytic mechanism of GH129 3,6-anhydro-D-galactosidase

Publication:

Wallace, M. D.[#], Cuxart, I.[#], Roret, T., Guée, L., Debowski, A. W., Czjzek, M., Rovira, C., Stubbs, K. A., Ficko-Blean, E. "Constrained Catalytic Itinerary of a Retaining 3,6-Anhydro-D-Galactosidase, a Key Enzyme in Red Algal Cell Wall Degradation" *Angewandte Chemie International Edition*, e202411171 (2024) [#] Equal contribution

Abstract: The marine bacterium *Zobellia galactanivorans* has the appropriate machinery for degrading red algal cell wall galactan carrageenans. Its 3,6-anhydro-D-galactosidase ZgGH129 releases the rare α -1,3-linked 3,6-anhydro-D-galactose monosaccharide (ADG) using specific molecular interactions to deal with the steric restrictions imposed by the bicyclic ADG, which is ring-flipped relative to D-Gal. Using crystallographic structures of relevant states with natural substrate and novel chemical probes mimicking species along the reaction coordinate, together with QM/MM metadynamics methods and kinetic studies, we unveil a retaining mechanism where the second step is rate limiting. The conformational landscape of the 3,6-anhydro-D-galactopyranose ring proceeds through enzyme glycosylation $B_{1,4} \rightarrow [E_4]^{\ddagger} \rightarrow E_4/{}^1C_4$ and deglycosylation $E_4/{}^1C_4 \rightarrow [E_4]^{\ddagger} \rightarrow B_{1,4}$ itineraries in the Southern hemisphere of the Cremer-Pople sphere. The results demonstrate the conformational changes throughout catalysis in a non-standard, sterically restrained and bicyclic monosaccharide, and provide insights for the design of specific inhibitors for anhydro-type carbohydrate-processing enzymes and for applications involving finely-tuned carrageenan degradation.

5.1 Introduction

Red algae are photosynthetic, ancient eukaryotic organisms that live mostly in the sea and coastal environments [206,207]. The highly diverse types of their cell wall polysaccharides, which represent the main constituents of their biomass [7], make them adapted for the oceanic environment, where they face high osmotic stress and currents [208,209]. The main cell wall polysaccharides of red macroalgae are agars and carrageenans, both linear, sulfated galactans [210]. Carrageenans are a type of galactans alternating D-galactose units linked with α -1,3- and β -1,4-glycosidic bonds. This D-galactose base is further diversified by incorporating the modified sugars D-galactose-6-sulfate and 3,6-anhydro-D-galactose (named ADG hereafter), unique of red algae, in the β -linked positions of some carrageenans, as well as other modifications such as sulfations, methylations and pyruvations [4,7]. In fact, the main distinction between agars and carrageenans is the isomeric form of this unique bicyclic saccharide: while carrageenans incorporate the D-form, in agars the 3,6-anhydro-L-galactose is found [211] (Figure 5.1). ADG is formed in carrageenans by cyclization catalyzed by sulfurylase enzymes [212] and presents a special feature: the 3,6-anhydro bridge imposing structural constraints and flipping the pyranose ring of the α -galactosyl from 4C_1 to the inverted 1C_4 conformation. Thus, its 4-OH exocyclic group is equatorial within the carrageenan chains [213].

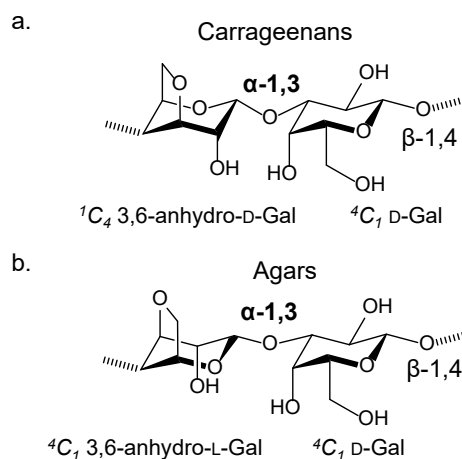


Figure 5.1: a. Repeating motif of κ -type of carrageenans. b. Repeating motif of agarobiose. Besides the decoration patterns, the main difference between carrageenans and agars is that carrageenans contain the D-form of 3,6-anhydro-galactose, whereas agars have the L form, so the pyranose ring conformation of ADG in carrageenans is flipped respect to agars. R=H, CH₃. Adapted from reference [213].

Algal cell wall polysaccharides like carrageenans are a major carbon source for marine organisms, including heterotrophic bacteria. Some of them, like those be-

longing to the *Bacteroidetes* phylum, are equipped with complex machinery for processing these algal polysaccharides, especially carrageenans [7], carried out by proteins encoded by adjacent and co-regulated genes in the so-called Polysaccharide Utilization Loci (PUL) [7]. The importance of organic matter degradation in marine environments brought research attention to the machinery of the marine *Bacteroidetes* bacterium *Zobellia galactanivorans* Dsij^T, and the regulon responsible for carrageenan utilization was characterized in a 2017 study by E. Ficko-Blean *et al.* [7]. The authors identified for the first time an enzyme able to cleave ADG units at the non-reducing end of carrageenan neo-oligosaccharides (those already degraded at the β -1,4 linkages with *endo*-carrageenases [210]) within the CarPUL regulon and provided a crystal structure of the unliganded form of the enzyme (PDB ID 5OPQ), named *ZgGH129*. Later, in 2020, Ficko-Blean *et al.* characterized for the first time the substrate specificity of *ZgGH129*, developed novel probes for determining its activity and found that the minimum substrate unit that the enzyme can process is neo- β -carrabiose (ADG- α -1,3-Gal) [214, 215].

Algal galactans, like carrageenans, are used in different industrial sectors, mainly for their attractive rheological properties [7, 216, 217]. In addition, they are attracting interest for medical applications for the development of carrageenan oligosaccharide-based anti-tumor and antiviral drugs [210, 218]. Therefore, the fine-tuned degradation of heterogeneous, large carrageenan polysaccharides into specific oligosaccharides has promising applications.

Exo-(α -1,3)-3,6-anhydro-D-galactosidase is a homodimer with a 40 Å deep crevasse hosting the putative active site. With the structural data and mutagenesis studies, the authors identified four potential catalytic residues (Asp486, Glu517, Glu531 and Asp202), although the lack of saccharide substrate or analogs precluded further assignation at the time [7]. The enzyme was assigned to the family 129 of glycoside hydrolases. This family contains another characterized member, *Bifidobacterium bifidum* α -*N*-acetylgalactoseaminidase [219] (NagBb or *BbGH129*). The hydrolysis by *ZgGH129* follows a retaining mechanism [214], with the formation of a glycosyl-enzyme intermediate (Figure 5.2) [16]. For identifying the catalytic residues, phylogenetic and structural analyses conducted by Wallace, Stubbs and Ficko-Blean comparing *ZgGH129* and *BbGH129* proposed that the catalytic pair is Asp486/Glu517 [214]. Although in 2020 Wallace *et al.* determined the minimum substrate that *ZgGH129* can recognize [214, 215], the recognition of larger, possibly heterogeneous carrageenan substrates by *ZgGH129* remained still unknown due to the lack of structural data with natural substrates or mimics thereof. Also, *BbGH129* binds an α -linked GalNAc, which in the crystal structure adopts a ⁴C₁

conformation whereas for ADG, due to the 3,6-anhydro cycle, it would be expected to find a flipped conformation of the pyranose ring. Due to the lack of structural information, it was not yet known how this feature of ADG would impact ligand recognition and catalysis by *Zg*GH129.

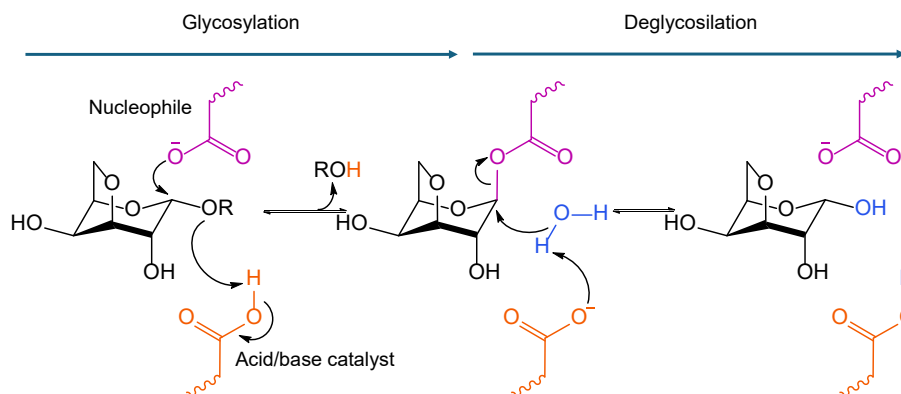


Figure 5.2: Scheme of the double-displacement catalytic mechanism predicted for *Zg*GH129, where the putative catalytic residues are Asp486 and Glu517 for the roles of nucleophile and general acid/base, respectively. $R = \text{Gal-}\beta\text{-1,4-ADG-}\alpha\text{-ADG-SO}_3^-$.

Another open question of *Zg*GH129 catalysis is the conformational itinerary of the ADG moiety at the non-reducing end, the one adjacent to the scissile bond. The 3,6-anhydro bridge introduces a constraint in the D-Gal ring as it places the C4 below the average ring plane. This entails a whole inversion of the ring to the 1C_4 conformation, which lies at the South pole of the Cremer-Pople sphere. Therefore, the conformational itinerary may take place at its Southern hemisphere. This would distinguish *Zg*GH129 from all other studied galactosidases, as well as *Bb*GH129, which, either processing α - or β -linked Gal, always work through the Northern hemisphere, although the inverted 1C_4 chair is one of the accessible conformations for an isolated galactosyl [9]. Southern hemisphere itineraries are reported for GH47 inverting α -mannosidases [44], GH134 β -mannanases [38, 220] or L-fucosidases [221]. Therefore, these itineraries are more scarce than their Northern hemisphere counterparts, which is attributed to the energetic cost to flip the ring conformation and to the fact that L-sugars are less common in nature [1, 56].

5.1.1 An interdisciplinary study. Experimental results

This project was performed in collaboration with the Laboratory of Integrative Biology of Marine Models, of the CNRS and Sorbonne Université, located in Roscoff (France), led by Prof. Dr. Elizabeth Ficko-Blean, and the research group of Prof. Dr. Keith A. Stubbs, of the School of Molecular Sciences of the University of

Western Australia. The first authorship of the publication of this work is shared with Dr. Michael Wallace (University of Western Australia). The team of M.W. and K.A.S designed, synthesized and characterized several putative substrates, inhibitors and probes, aiming to the creation of a chemical toolbox for the study of carrageenans and the enzymes processing them. With some of the designed probes, the team of E.F.B prepared crystallographic structures of *ZgGH129* (the apo form structure, PDB 5OPQ, was obtained earlier [7]). ADG-based chemical probes (3,6-anhydro-D-1-deoxygalactonojirimycin, 3,6-anhydro-D-galacto-isofagomine) and one mechanism-based inhibitor (2-deoxy-2-fluoro-3,6-anhydro-D-galactose) [57] (Figure S5.1) bound to the enzyme provided important mechanistic information regarding the conformation of the putative substrate along the reaction coordinate, at the Southern hemisphere of the Cremer-Pople sphere. This, and the fact that there were divergences in inhibition power in some of them with the other characterized member of the GH129 family, *BbGH129*, differentiated *ZgGH129* in a distinct clade within the GH129 family. The crystallographic structures of the enzyme in complex with putative substrate neo- β/κ -carrageenan (DP4) (Figure S5.1), reaction product ADG and mechanism-based inhibitors (2-deoxy-2-fluoro-3,6-anhydro-D-galactose) (Figure S5.1) gave further insight in the identity of the catalytic residues as well as the oligomeric order of the active enzyme, with room for questions in the identity of the catalytic acid/base.

In this Thesis Chapter we present our part in the project in which we modeled the catalytic mechanism of *ZgGH129* and uncovered the conformational preferences of the highly constrained 3,6-anhydro-D-galactosyl substrate both in isolation and along catalysis.

5.2 Results and discussion

5.2.1 Modeling the Michaelis complex of *ZgGH129* with a neo- β/κ -carrageenan oligosaccharide

Guided by the structures of *ZgGH129* solved and characterized by our collaborators, corresponding to different stages of its catalytic mechanism, we modeled the full hydrolytic mechanism of an oligo-carrageenan by *ZgGH129* and described the catalytic itinerary and conformational preferences of the rigid substrate 3,6-anhydro-D-galactose, using QM/MM and metadynamics methods.

We built the Michaelis complex of *ZgGH129* with the natural substrate neo- β/κ -DP4 (ADG- α -1,3-Gal- β -1,4-ADG- α -Gal-SO₃) using the catalytically inactive

Asp486Asn variant structure, reverting the mutation. All resolved structures are homodimers with two equivalent active sites at the ends of a 40 Å deep cavity (Figure 5.3), with four subsites for the neo- β - κ -DP4 tetrasaccharide (Figure S5.2). Since the ZgGH129 active site contains residues of both subunits, we used both of them for building the enzyme models of this work. During the 200 ns of classical MD simulation (Figure S5.3, 4 independent replicas of 50 ns each), we observed that residues Asp486 and Glu517 are well-positioned for the hypothesized retaining mechanism. Residue Glu531, although near the scissile glycosidic bond, is on average at 6.80 ± 0.30 Å of the substrate O1 ($O_{Glu531}-O_1$ in Figure 5.4), further than the putative acid catalyst Glu517, which is at hydrogen bond distance during the simulation ($O_{Glu517}-O_1 = 2.77 \pm 0.10$ Å, $H_{Glu517}-O_1 = 1.90 \pm 0.12$ Å). Glu531 is coordinating a water molecule that in its turn interacts ADG at -1, a persistent interaction that spans most simulation time (Figure 5.4). Therefore, our classical MD simulations in combination with the structural data, helped identifying acid/base and confirmed a water-mediated interaction important for substrate binding at subsite -1. Our simulations show the role of Glu531 in substrate binding, providing a structural and molecular insights on why Glu531 is a crucial residue, as its mutated variant abolishes enzymatic activity [7].

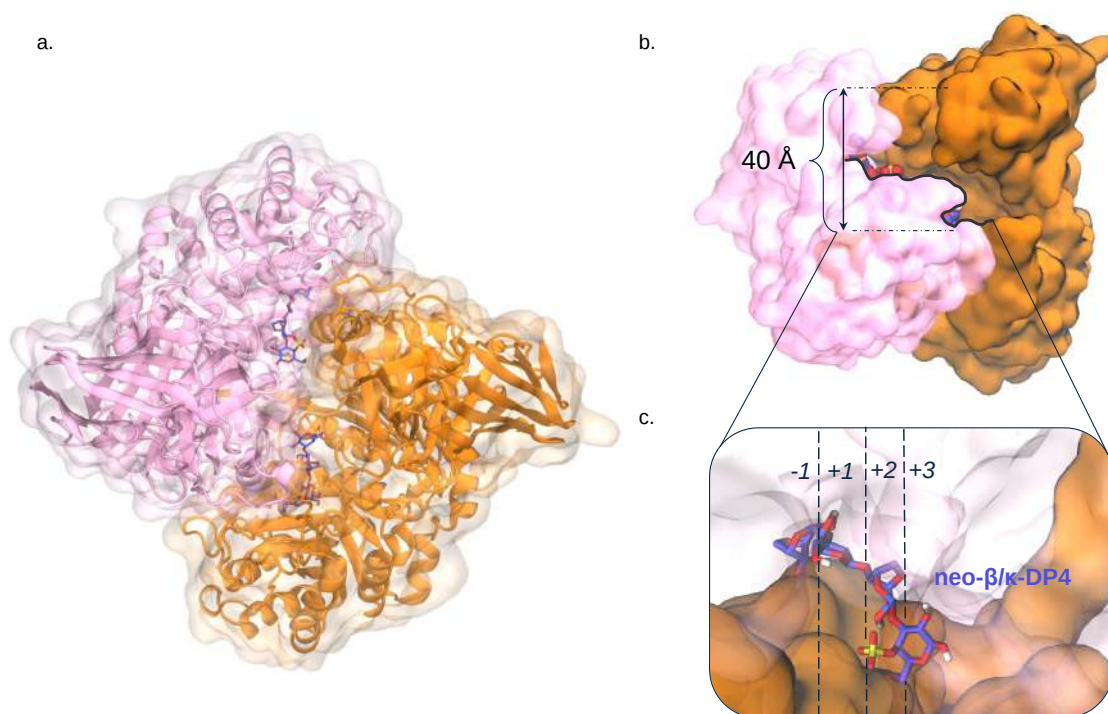


Figure 5.3: a. *ZgGH129* crystal structure in complex with neo- β - κ -carrageenan (purple sticks) solved by the group of Dr. Elizabeth Ficko-Blean. b. Each substrate oligosaccharide lies at the bottom of the 40 Å cavity formed by the two protein subunits c. Purple sticks, detail of the substrate in the active site cavity of *ZgGH129*.

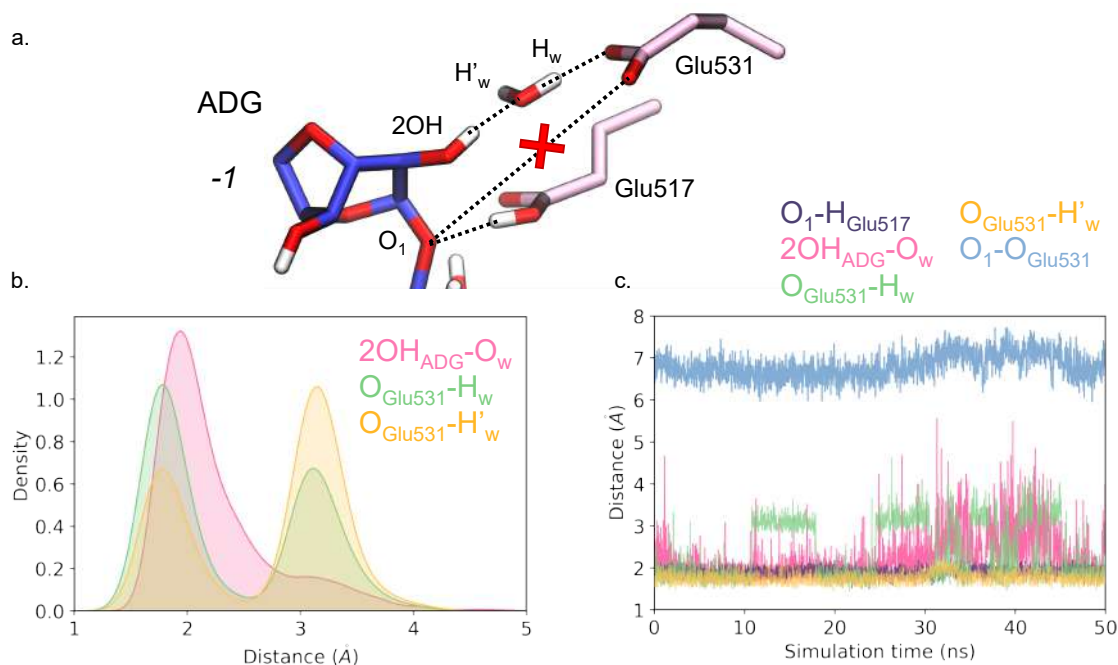


Figure 5.4: a. Detail of the H-bond interactions of residues Glu517 and Glu531 with ADG directly and via a water molecule during classical MD simulations. Glu531 is far from the scissile glycosidic bond. The water molecule interacts with Glu531 through either of its H (H_w and H'_w). c. Distance evolution of the interaction distances between ADG, Glu517 and Glu531, during 50 ns of classical MD simulation (only replica 3 represented for clarity, but the other replicas share this behavior).

The structural information points towards a conformational itinerary where the reactive ADG moiety undergoes conformational changes across the Southern hemisphere of the Cremer-Pople sphere (Figure 5.5). A Southern itinerary, to the best of our knowledge has not yet been reported for any α - or β - D-galactosidase (Table S5.1), with the enzyme-substrate complex in a $B_{1,4}$ conformation. We hypothesize that the enzyme takes advantage of how the unusual 3,6-anhydro bridge of ADG impacts its conformational possibilities, which we proceeded to study with QM metadynamics.

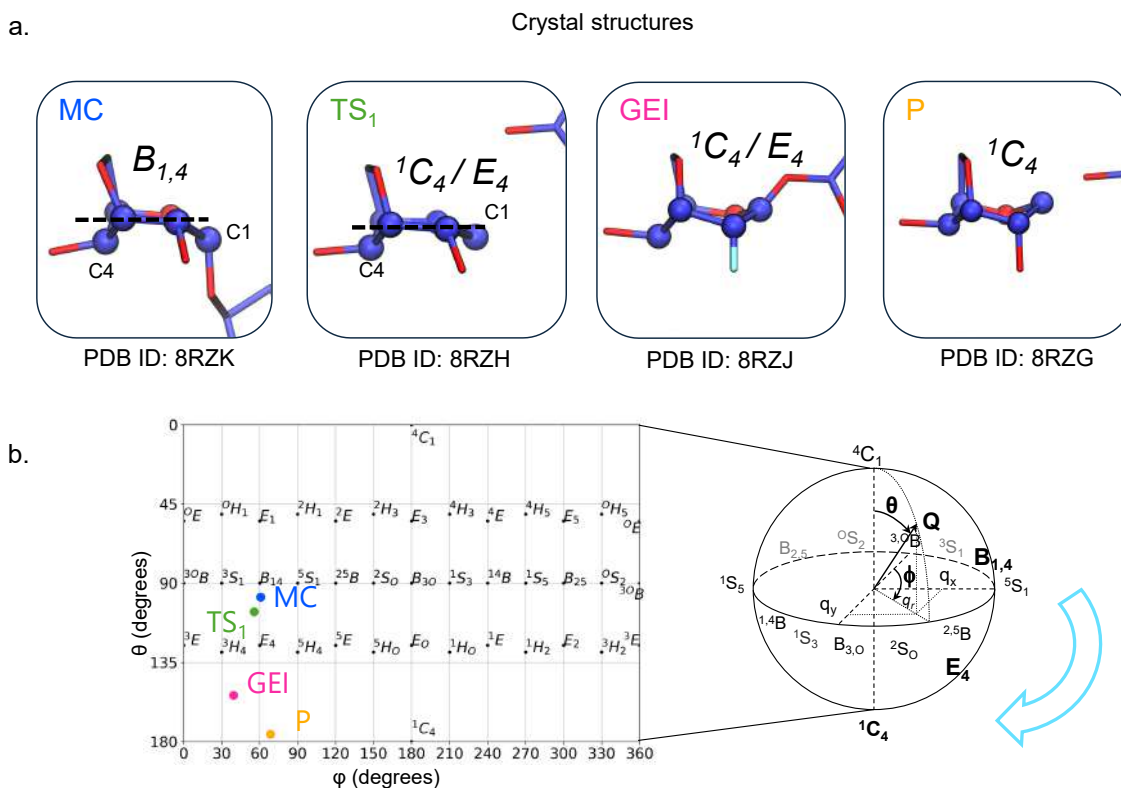


Figure 5.5: a. Crystal structures of putative substrate and chemical probes designed by our collaborators in complex with *ZgGH129*, indicating their ID numbers in PDB. From left to right: neo- β/κ -carrageenan (DP4), AD-DGJ, 1,2-diF-ADG and 3,6-anhydro-D-Gal. For more details about the structure of the ligand and probes, check supplementary Figure S5.1. b. Mercator projection (left) of the Cremer-Pople sphere (right) shows the different conformations of the ADG or ADG-like bicyclic ring, all of them in the Southern hemisphere of the sphere, which gives structural insight on their conformational itinerary during *ZgGH129* catalysis.

5.2.1.1 The conformational preferences of isolated 3,6-ADG

Starting from the Michaelis complex model with neo- β/κ -DP4 we first studied the conformational preferences of the isolated, α -3,6-anhydro-D-galactosyl ring, therefore not part of any oligosaccharide nor in the enzyme active site cavity, which gives information about the intrinsic conformational preferences of saccharides [16]. For this purpose we computed the conformational FEL of the α -3,6-ADG with *ab initio* QM(DFT) coupled to metadynamics to drive the bias across the coordinate space of the Cremer-Pople pyranose ring coordinates (Figure S5.4).

The reconstructed conformational FEL of ADG shows that the saccharide can mainly access two conformations, corresponding to the two FEL minima, which are $B_{1,4}$ and 1C_4 (Figure 5.6). The $B_{1,4}$ and 1C_4 are almost equienergetic (1C_4 is 2.10 kcal mol $^{-1}$ more stable than $B_{1,4}$) and are separated by a free energy barrier of 3.0 kcal mol $^{-1}$. This results contrast with the conformational FEL of galactose, which shows several other accessible conformations in both hemispheres of the Cremer-

Pople sphere [9].

We next analyzed the intramolecular hydrogen bond patterns of both $B_{1,4}$ and 1C_4 conformers (Figure 5.6). The strongest hydrogen bonds in both conformations are between the H atom of ADG 2-OH group and O1, followed by the H of 4-OH with O3. The 1C_4 conformers display an additional strong interaction between the H of 1-OH and the O2, which is not present in the $B_{1,4}$ conformers. This is possibly due to the axial position of the 1-OH exocyclic group when C1 is below the average ring plane in $B_{1,4}$, which precludes this group from interacting with O2. The difference in hydrogen bonding patterns can explain the energy difference between both 1C_4 and $B_{1,4}$ conformers.

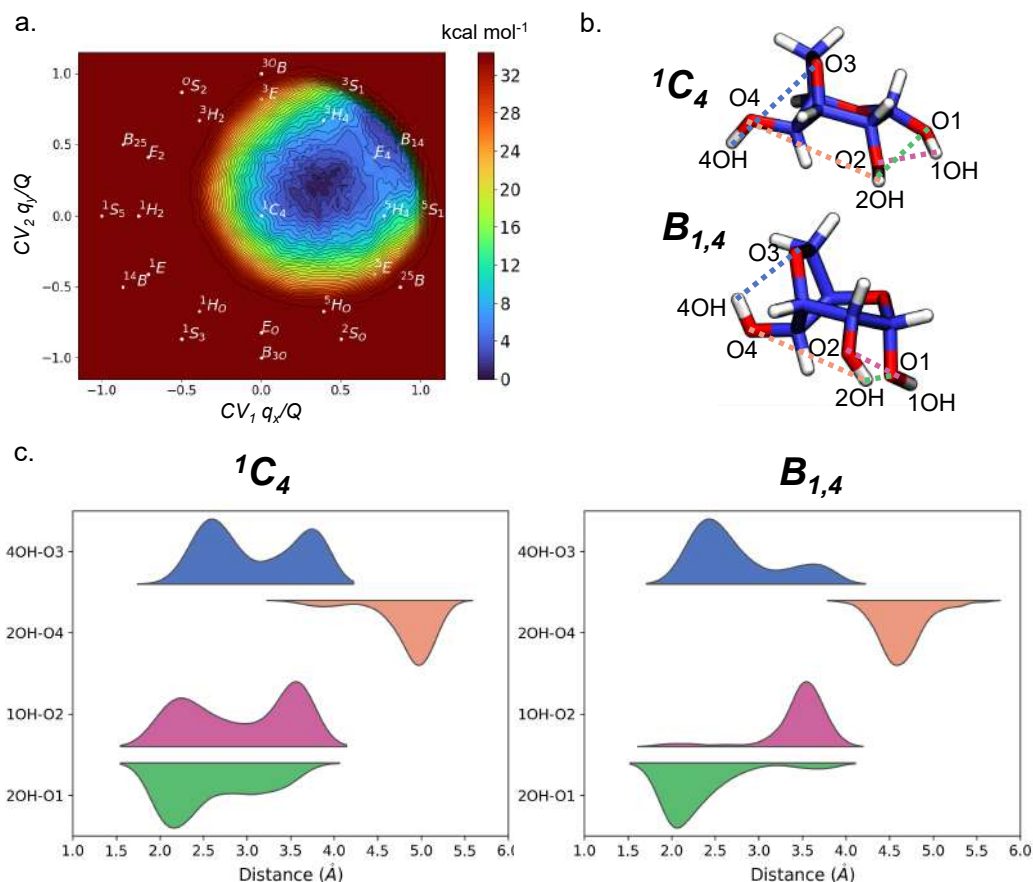


Figure 5.6: a. Stoddart projection in CV1 and CV2 of the computed conformational FEL of isolated α -3,6-ADG. Isolines drawn every 1.0 kcal mol⁻¹. b. Conformations of the free energy minima. In dashed lines, all possible intramolecular H-bonds. c. Distribution of the values adopted by the intramolecular H-bonds between both hydroxyl exocyclic groups and the anhydro O3. Values correspond to an ensemble of structures at the 1C_4 and $B_{1,4}$ minima, respectively.

A most notable aspect of the conformational FEL of ADG is the fact that the

two available conformations are in the Southern hemisphere of the Cremer and Pople sphere (Figure S5.5), as no transitions across Northern and Southern hemispheres were sampled. This contrasts with the available conformational FEL of D-galactose, in which conformations of both hemispheres are accessible [9], including 1C_4 , although is less populated than 4C_1 . Although the North to South pole transitions (so, flipping from the 4C_1 to 1C_4 conformers) can have an associated energetic cost in some pyranoses due to steric effects between exocyclic groups [19,56], in ADG the double cycle structure constraints the pyranose ring, positioning the C4 far below the average ring plane and rendering all Northern hemisphere conformations high in energy. To gain insight into how this translates to the enzyme active site, and which of these conformations are favored for catalysis, we moved on to study the conformational preferences of ADG in the ZgGH1129-substrate complex.

5.2.1.2 The conformational preferences of 3,6-ADG in the active site of ZgGH129

To investigate how the enzyme exploits the accessible conformations of ADG, we computed the conformational FEL of the native substrate neo- β/κ -DP4 in the active site of ZgGH129. For this purpose we applied QM/MM methods coupled to metadynamics, using the Cartesian components of the Cremer-Pople ring puckering coordinates to drive the bias. QM/MM MD simulations started from a structure of the classical MD of the Michaelis complex (section 5.4.1.2). In this case, in view that only the Southern hemisphere was explored in the simulations of isolated ADG, we selected two collective variables: q_x/Q (CV1) and q_y/Q (CV2) (Figure S5.6). Since q_z/Q accounts for transitions across the hemispheres of the Cremer-Pople sphere we decided to not use this variable in the QM/MM metadynamics simulation, as it would increase the computing time.

The reconstructed conformational FEL (Figure 5.7) shows that the enzyme environment favors the $B_{1,4}$ conformation, where the ADG moiety engages in several hydrogen bond interactions with enzyme residues, in agreement with the crystallographic structures of the ZgGH129 neo- β/κ -DP4 complex.

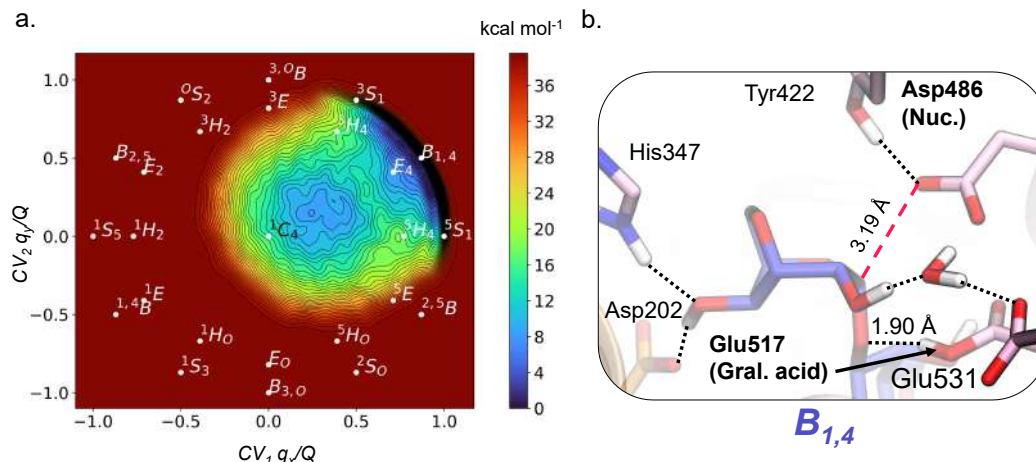


Figure 5.7: a. Stoddart projection of CV1 and CV2 of the computed conformational FEL of α -3,6-ADG as part of the neo- β/κ -DP4 substrate in *ZgGH129*. Isolines every 1.0 kcal mol⁻¹. b. Main conformations of the Cremer-Pople sphere, note that the minima of the conformational FEL correspond to conformations at the Southern hemisphere. c. Subsite -1 of *ZgGH129*. The $B_{1,4}$ conformer is the one found in the newly solved crystal structures, with the axial scissile bond ready for cleavage by Asp486 and Glu517.

The $B_{1,4}$ conformer of 3,6-ADG features the leaving group of the reaction (Gal- β -1,4-ADG- α -4-sulfate-D-galactosyl at subsites +1 to +3) in axial orientation, thus the nucleophile Asp486 points to the least hindered face of the ADG anomeric carbon and is at an appropriate distance for nucleophilic attack ($C1-O_{Asp486} = 3.19$ Å). The acid/base, Glu517, is at hydrogen bond distance with the scissile bond oxygen (1.90 Å), so the active site and the substrate are prepared for catalysis. This conformation is also observed in the crystal structure of *ZgGH129* with the natural substrate (Figure S5.2.b). Therefore our results indicate that the starting enzyme-substrate complex features a $B_{1,4}$ conformation of ADG. In this conformation, the substrate is preactivated for catalysis.

5.2.2 The glycosylation reaction

For modeling the glycosylation reaction by *ZgGH129* (the first reaction step) we selected a trajectory snapshot of the conformational FEL of the substrate in the enzyme active site (see Section 5.2.1.2) where the ADG moiety at subsite -1 is in a $B_{1,4}$ conformation (Figure 5.7.a,b). After applying QM/MM MD methods to model the *ZgGH129*-substrate complex, expanding the QM region to include the sidechains of the catalytic residues Asp486 and Glu517 (Figure S5.7), we selected two collective variables (CVs) to model the glycosylation reaction. The first CV (CV1) was defined as a difference of two distances, those being $d(C1-O_{Asp486})$ and $d(C1-O1)$, which accounts for the nucleophilic attack and covalent bond formation

to form the glycosyl-enzyme intermediate. The second CV (CV2) corresponds to the hydrogen bond interaction distance from the general acid to the scissile bond, $d(\text{H}_{\text{Glu517}}-\text{O1})$, and represents the protonation of the leaving group (Figure S5.8.a).

The reaction FEL shows two principal basins connected by one transition state, in concordance with a concerted reaction (Figure 5.8.a). The system evolves from the Michaelis complex, to the covalent glycosyl-enzyme reaction intermediate with a free energy barrier of $14.3 \pm 0.1 \text{ kcal mol}^{-1}$, in agreement with the one estimated from the experimental rate constant applying Transition State Theory, $15.6 \text{ kcal mol}^{-1}$. The additional minimum near the reactants state basin is a pre-catalytic configuration of the MC, which we name MC', connected through a small free energy barrier of $1.74 \text{ kcal mol}^{-1}$. In this configuration, the acid catalyst Glu517 interacts with the substrate ADG 2-OH group instead of the scissile bond O1 (Figure S5.9).

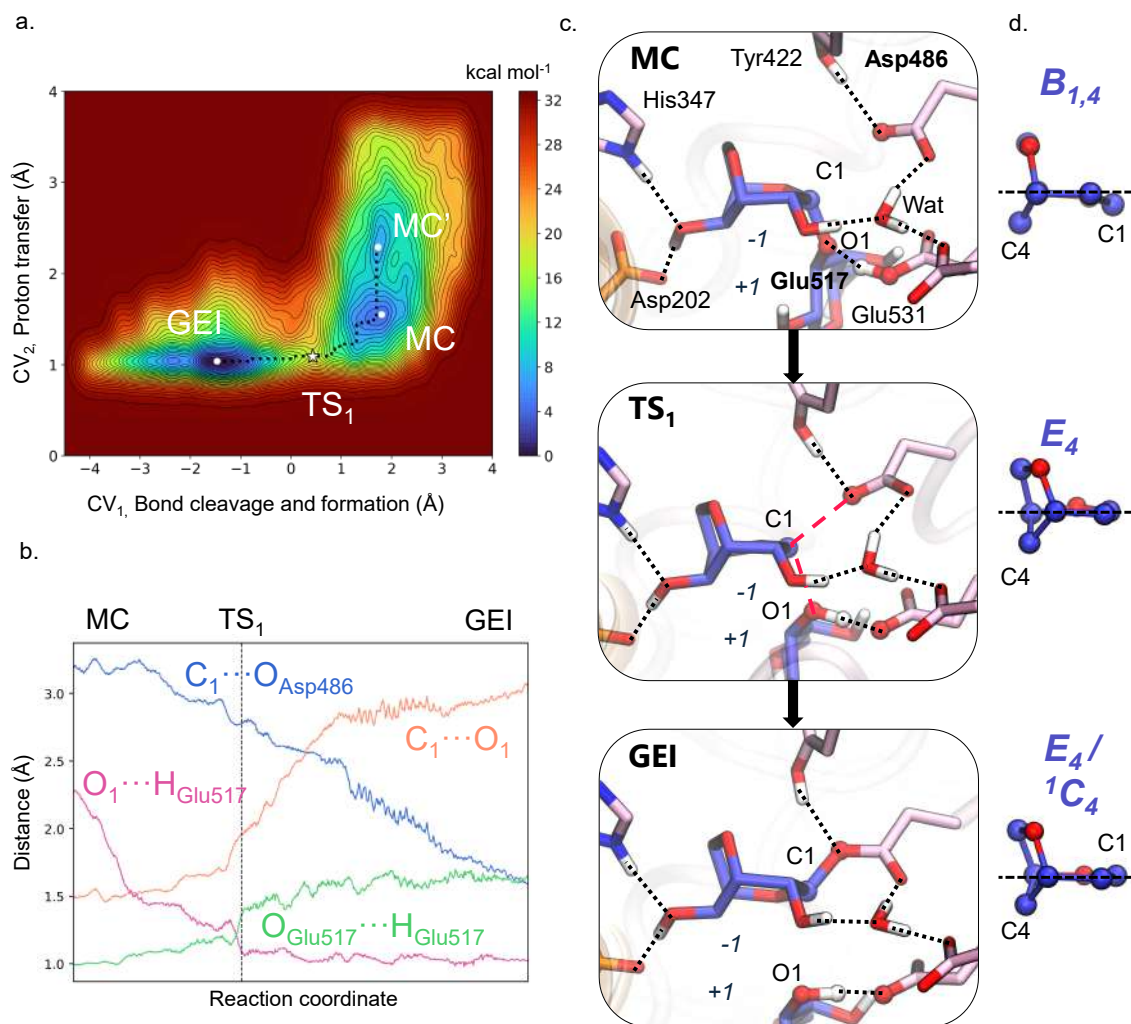


Figure 5.8: a. FEL of the glycosylation reaction *ZgGH129*. Isolines every 1.0 kcal mol⁻¹. b. Minimum energy pathway (MEP) evolution of the distances involved in the catalytic step. c. Representative structures of the main reaction states (MC, TS₁ and GEI) extracted from each FEL region. Black, dashed lines, H-bond interactions. Red, dashed lines at TS₁, covalent bonds broken and formed. Ball representation, atoms involved in the reactive steps. d. Ball-and-stick representation of the catalytic itinerary of ADG during the reaction pathway.

We analyzed in detail the reaction coordinate to obtain an atomistic point of view of the glycosylation reaction step (Figure 5.8.b and .c, and Table 5.1). For this we by calculated the evolution of relevant distances along the minimum free energy pathway of the reaction (MEP). At the MC, when Glu517 forms a hydrogen bond with O1 ($O1-H_{Glu517} = 1.51 \pm 0.09$ Å), the glycosidic bond distance increases ($C1-O1 = 1.54 \pm 0.06$ Å). After, the nucleophile approaches the anomeric carbon ($C1-O_{Asp486} = 3.20 \pm 0.17$ Å) and the proton of Glu517 starts transferring to the scissile bond O1 ($H_{Glu517}-O1 = 1.51 \pm 0.09$ Å). At the TS (TS₁, Figure 5.8.c) the ADG adopts an *E*₄ conformation, in which the C-O5-C1-C2 atoms of the pyranose ring are almost coplanar, typical of an oxocarbenium ion-like of the TS. This is

further reinforced by the shrinking of the C1-O5 bond distance (C1-O5 at MC is 1.39 ± 0.03 Å, whereas at TS₁ is 1.29 ± 0.01 Å), which indicates the partial double bond character developed between the endocyclic oxygen and the anomeric carbon. The covalent bond cleavage and formation distances indicate that the mechanism is dissociative, as at TS₁ the glycosidic bond is almost broken (C1-O1 = 2.19 ± 0.04 Å) whereas the glycosyl-enzyme bond is not yet formed (C1-O_{Asp486} = 2.62 ± 0.05 Å).

Distance (Å)	State			
	MC' ($B_{1,4}$)	MC ($B_{1,4}$)	TS ₁ (E_4)	GEI ($E_4/{}^1C_4$)
C1-O _{Asp486}	3.24 ± 0.09	3.20 ± 0.17	2.62 ± 0.05	1.65 ± 0.12
C1-O1	1.49 ± 0.05	1.54 ± 0.06	2.19 ± 0.04	3.08 ± 0.15
O1-H _{Glu517}	2.29 ± 0.18	1.51 ± 0.09	1.07 ± 0.04	1.02 ± 0.04
O _{Glu511} -H _{Glu517}	1.00 ± 0.03	1.06 ± 0.04	1.39 ± 0.12	1.66 ± 0.10
2OH _{ADG} -O _w	1.70 ± 0.13	1.67 ± 0.09	1.59 ± 0.05	1.63 ± 0.07
H _w -O _{Glu531}	2.27 ± 0.72	1.78 ± 0.14	1.72 ± 0.09	1.64 ± 0.10
H' _w -O' _{Asp486}	1.74 ± 0.10	1.84 ± 0.10	1.81 ± 0.26	1.87 ± 0.12
4OH _{ADG} -O _{Asp202}	1.55 ± 0.07	1.59 ± 0.10	1.58 ± 0.10	1.59 ± 0.08
4OH _{ADG} -H _{His347}	1.95 ± 0.13	2.06 ± 0.15	2.01 ± 0.09	2.42 ± 0.25
C1-O5	1.41 ± 0.04	1.39 ± 0.03	1.29 ± 0.01	1.37 ± 0.04
O _{Asp486} -H _{Y422}	1.93 ± 0.17	1.88 ± 0.14	1.87 ± 0.10	2.02 ± 0.20
O6 _{ADG} -H _{Y422}	3.85 ± 0.26	4.04 ± 0.28	3.75 ± 0.11	3.75 ± 0.11

Table 5.1: Main interatomic distances of the active site at the main states of the first reaction step of ZgGH129 hydrolysis (MC, TS₁ and GEI). Structures of all states were extracted from the computed reaction FEL. The water molecule referred (H_w, H'_w and O_w) is the one coordinating Glu531 with the ADG 2-OH group.

During the reaction, ZgGH129 binds the substrate in the active site through hydrogen bond interactions like the ones involving the 4-OH group of ADG at subsite -1 with both Asp202 and His347, the water-mediated interaction between its 2-OH and Glu531 (Figure 5.8.c) and the hydrophobic interaction between Trp455 and the ADG C5-C6 ridge of ADG (Figure 5.9). The active site is also maintained interactions between residues, like Tyr422-Asp486, and Trp455-Asp202 hydrogen bond. It is noteworthy that the water molecule mediating the interaction between the 2-OH group and Glu531 also coordinates this residue with Asp486, via the O atom that is not involved in the nucleophile attack (H'_w-O'_{Asp486} at Table 5.1).

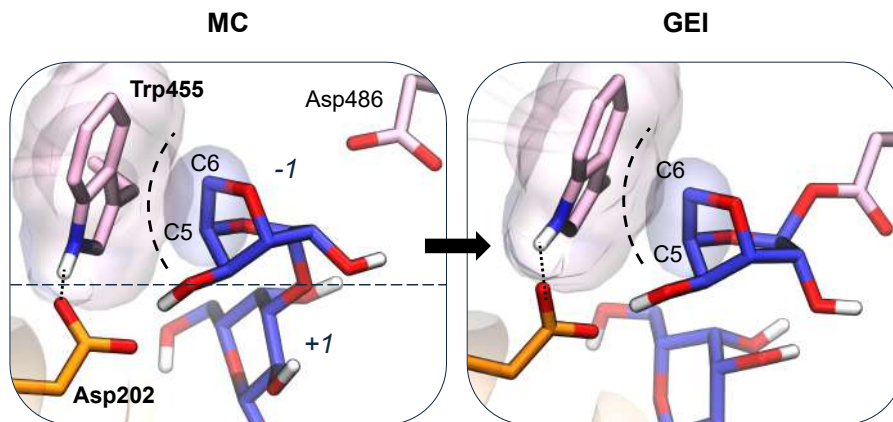


Figure 5.9: Detailed view of the *ZgGH129* active site at MC and GEI states. Trp455 provides hydrophobic support for the C5-C6 ridge of ADG at subsite -1, and also contributes to the active site integrity by interacting with Asp202.

When the glycosidic bond starts elongating, the ADG at subsite -1 evolves from the initial boat $B_{1,4}$ conformation to an envelope E_4 at the transition state TS_1 . At the end of the glycosylation reaction, ADG is between E_4 and 1C_4 . The conformation at the transition state TS_1 , E_4 is one of the possible pyranose TS structures [16] which render the atoms C5-O5-C1-O2 in the same plane. Therefore, the conformational itinerary is $B_{1,4} \rightarrow [E_4]^\ddagger \rightarrow E_4/{}^1C_4$, at the Southern hemisphere of the Cremer-Pople sphere (Figure S5.8.d and S5.10), in agreement with the crystallographic structures of the inactive variant in complex with the carrageenan substrate and the 2-fluorinated GEI analog (Figure 5.5). Our simulations provided the details of how the enzyme has adapted to the geometry of the constrained ADG substrate and assists with several interactions the conformational changes necessary to undergo catalysis. Additional QM/MM MD simulations starting with the enzyme-substrate complex featuring 1C_4 ADG (the alternative minimum of the conformational FEL, Figure 5.7) showed that during the unbiased QM/MM MD equilibration the conformation of the saccharide evolved towards $B_{1,4}$, which supports that the reactive Michaelis complex features a distorted $B_{1,4}$ ADG (Figure S5.11).

5.2.3 Modeling the glycosyl intermediate of *ZgGH129* with 3,6-ADG

For building the covalent glycosyl-enzyme intermediate of *ZgGH129* with ADG, we started from the structure solved by our collaborators, corresponding to the WT enzyme in complex with the mechanism-based inhibitor 1,2-diF-ADG. This molecule covalently bonds with Asp486 via its C1, but it is non-hydrolysable, therefore it can be trapped when forming the GEI without proceeding to the reaction product. We reconstructed the natural GEI *ZgGH129*-ADG from this structure by replacing the F atom at C2 with a hydroxyl group (section 5.4.2.1 and Figure S5.12).

The modeled glycosyl intermediate complex corresponds to the state where ADG is covalently bound to the enzyme, and after the leaving group of the first reaction step has already left the enzyme active site. Therefore, this state (named GEI' hereafter) differs from the QM/MM metadynamics obtained intermediate (GEI, see section 5.2.2) in the absence of the leaving group. The modeled GEI' was stable during 300 ns of classical MD, performed in 6 independent replicas of 50 ns (Figure S5.13.a). This structure is in agreement with the GEI obtained in our QM/MM metadynamics simulations of the glycosylation step (Section 5.2.2). During classical MD simulations, water molecules that enter the active site cavity accommodate under the ADG anomeric carbon and interact with Glu517, the putative general base for the *ZgGH129* deglycosylation step, which in this stage of the mechanism is in its carboxylate form (Figure S5.13). The water-mediated interaction between the substrate 2-OH and Glu531 is maintained as in the MC, and is also conserved in the crystal structures (Figure S5.12.a). The position of this water, far from the ADG C1 (approximately between 4 and 5 Å) (see Figure S5.12.b), together with the fact that catalytically competent positioned waters are sampled in our simulations, supports the hypothesis that this water molecule is involved in ADG binding and not catalysis. It also supports and the catalytic role of Glu517 as acid/base, rather than Glu531.

5.2.4 The deglycosylation reaction

We selected a structure with a water molecule suitably positioned for the deglycosylation step of catalysis to proceed with QM/MM MD and metadynamics approach, including this water in the QM region. Two collective variables (CVs) representing the water nucleophilic attack and ADG-enzyme bond cleavage successfully drove the system from the initial state (GEI') to the products state (P) of the deglycosylation reaction. The first CV (CV1) is the distance from the catalytic water oxygen to the ADG anomeric carbon, $d(O_{wat}-C1)$, and the second (CV2) is the ADG-*Zg*GH129 covalent bond distance, $d(C1-O_{Asp486})$ (Figure S5.15.a).

The computed FEL shows that the mechanism takes place concertedly, with one minimum corresponding to the initial state (reaction intermediates, GEI') connected to the reaction products (P) via a single transition state (TS₂) (Figure 5.10.a). The computed free energy barrier is 17.2 ± 0.7 kcal mol⁻¹, higher than that of the glycosylation step of the reaction (14.3 ± 0.1 kcal mol⁻¹), which, combined with the kinetic assays performed by our collaborators, predict that the glycosylation reaction is not the rate-limiting step in *Zg*GH129 catalysis. The reaction FEL is slightly endergonic, with a free energy difference of 3.47 kcal mol⁻¹ between GEI' and P. Our collaborators obtained the Brønsted plots of $\log(k_{cat})$ vs. pK_a probing the glycosilation reaction of *Zg*GH129 with different aryl leaving groups (see Supplementary section Figure S5.1 **1** to **6**, and Figure S5.16), which showed no dependence of the reaction rate with departing group ability [57]. This lead to hypothesize that the glycosylation step is not rate determining, but rather the deglycosylation step or a non-chemical process such as a conformational change, which is supported by the higher computed free energy barrier of deglycosylation (17.2 ± 0.7 kcal mol⁻¹, versus 14.3 ± 0.1 kcal mol⁻¹ of glycosylation).

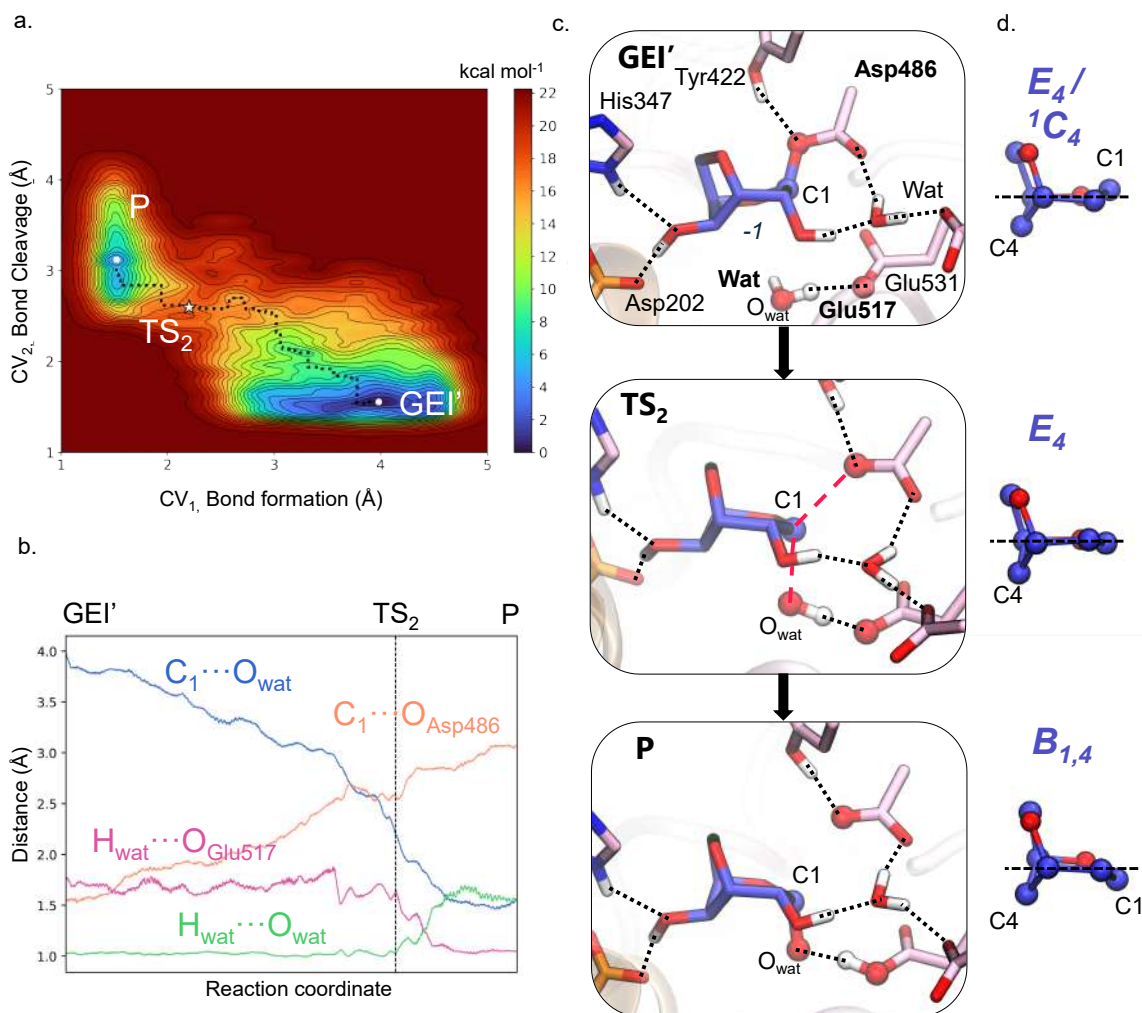


Figure 5.10: a. FEL of the simulated deglycosylation reaction step by *ZgGH129*. Isolines every $1.0 \text{ kcal mol}^{-1}$. b. MEP evolution of the main interatomic distances. c. Representative structures of the main states of the reaction (GEI' , TS_2 and P) extracted from each FEL basin. Black lines, H-bond interactions at the active site. Red lines at TS_2 , covalent bonds being broken or formed. Ball representation, atoms involved in the reaction. d. Ball-and-stick representation of the catalytic itinerary of ADG at the main states of the reaction pathway, mirroring that of the glycosylation step of the mechanism.

Analysis of the relevant states of the FEL (GEI' , TS_2 and P) uncovered further mechanistic details of the deglycosylation reaction step (Figure 5.10.b,c, Table 5.2). At the beginning of the reaction the water molecule forms a hydrogen bond interaction with the catalytic base (GEI' , $\text{H}_{\text{wat}}\cdots\text{O}_{\text{Glu517}} = 1.71 \pm 0.13 \text{ \AA}$) and starts approaching the ADG anomeric carbon ($\text{C1}\cdots\text{O}_{\text{wat}} = 3.86 \pm 0.25 \text{ \AA}$), in-line for the nucleophilic attack. The covalent bond between ADG and the enzyme starts elongating as the reaction progresses (Figure 5.10.b, $\text{C1}\cdots\text{O}_{\text{Asp486}}$) and at TS_2 is already broken, whereas the new bond with the water molecule O_{wat} , although close to C1 , is not yet formed, which indicates that the transition state of the deglycosylation

step is dissociative like the first step (TS_2 , $\text{C1-O}_{\text{wat}} = 2.17 \pm 0.05 \text{ \AA}$ and $\text{C1-O}_{\text{Asp486}} = 2.61 \pm 0.06 \text{ \AA}$). At TS_2 the conformation is E_4 , and the C1-O5 bond decreases from $1.39 \pm 0.03 \text{ \AA}$ (GEI') to $1.31 \pm 0.01 \text{ \AA}$ (TS_2), acquiring partial double bond character. Unlike in the glycosylation step, the proton is not yet transferred to the general base when the reaction reaches the transition state. At TS_2 , H_{wat} is still forming a bond with the catalytic water oxygen ($\text{H}_{\text{wat}}\text{-O}_{\text{wat}} = 1.06 \pm 0.04 \text{ \AA}$) while at hydrogen bond distance with the catalytic base ($\text{H}_{\text{wat}}\text{-O}_{\text{Glu517}} = 1.53 \pm 0.06 \text{ \AA}$). At P, the water transfers a proton to Glu517 (P, $\text{H}_{\text{wat}}\text{-O}_{\text{Glu517}} = 1.05 \pm 0.03 \text{ \AA}$), thus restoring the enzyme original protonation state at the beginning of catalysis and completing the reactive catalytic cycle.

Distance ($\text{\AA}$)	State		
	$\text{GEI}' (E_4/{}^1C_4)$	$\text{TS}_2 (E_4)$	P ($B_{1,4}$)
C1-O_{wat}	3.86 ± 0.25	2.17 ± 0.05	1.51 ± 0.04
$\text{C1-O}_{\text{Asp486}}$	1.58 ± 0.04	2.61 ± 0.06	3.12 ± 0.03
$\text{H}_{\text{wat}}\text{-O}_{\text{Glu517}}$	1.71 ± 0.13	1.53 ± 0.06	1.05 ± 0.03
$\text{H}_{\text{wat}}\text{-O}_{\text{wat}}$	1.02 ± 0.03	1.06 ± 0.04	1.53 ± 0.08
$2\text{OH}_{\text{ADG}}\text{-O}_w$	1.66 ± 0.11	1.66 ± 0.08	1.65 ± 0.10
$\text{H}_w\text{-O}_{\text{Glu531}}$	1.68 ± 0.18	1.63 ± 0.13	1.88 ± 0.21
$\text{H}'_w\text{-O}'_{\text{Asp486}}$	1.89 ± 0.22	1.84 ± 0.17	1.83 ± 0.12
$4\text{OH}_{\text{ADG}}\text{-O}_{\text{Asp202}}$	1.60 ± 0.11	1.54 ± 0.07	1.67 ± 0.14
$4\text{OH}_{\text{ADG}}\text{-H}_{\text{His347}}$	2.28 ± 0.29	2.22 ± 0.27	2.23 ± 0.23
C1-O5	1.39 ± 0.03	1.31 ± 0.01	1.41 ± 0.03
$\text{O}_{\text{Asp486}}\text{-H}_{\text{Y422}}$	2.16 ± 0.27	1.81 ± 0.11	1.90 ± 0.13
$\text{O6}_{\text{ADG}}\text{-H}_{\text{Y422}}$	2.80 ± 0.42	3.88 ± 0.28	3.98 ± 0.30

Table 5.2: Main interatomic distances of the active site at the main states of the second reaction step of *Zg*GH129 hydrolysis (GEI' , TS_2 and P). Structures of all states were extracted from the computed reaction FEL. The water molecules are the catalytic water (H_{wat} and O_{wat}) and the one coordinating Glu531 with the ADG 2-OH group (H_w , H'_w and O_w)

The conformational itinerary of ADG during the glycosylation reaction, ${}^1C_4/E_4 \rightarrow [E_4]^\ddagger \rightarrow B_{1,4}$ (Figure 5.10.d and S5.15.c), mirrors the glycosylation step and takes place in the Southern hemisphere of the Cremer-Pople sphere.

To further explore the endoergonic character of the deglycosylation reaction observed in the FEL (Figure 5.10.a), and study if there might be an additional step needed to complete the reaction and attaining a more stable, final state. To explore a possible conformational change after reaching P (Figure 5.10.a) we performed a QM/MM metadynamics simulation the reaction products. We followed the approach described in section 5.4.1.4 and obtained a conformational FEL where 1C_4 is the most stable conformation. This conformer is stabilized via an interaction of the 1-OH with Asp486 (Figure S5.17), and the conformational change taking place from $B_{1,4}$ (P) to 1C_4 has an estimated free energy barrier of 1.6 kcal mol⁻¹, although more sampling is needed to converge the simulation and refine this value. However, the 1C_4 conformation after reaching the products state is observed experimentally in the crystal structure of *ZgGH129* in complex with ADG, which the crystallographers described as an average representing the covalently bound and unbound (product) states in equilibrium, rather than one of them univocally [57].

In summary, our QM/MM metadynamics simulations support a double displacement, retaining mechanism for the hydrolysis of a carrageenan oligosaccharide including the α -bonded 3,6-anhydro-D-galactosyl moiety by *ZgGH129*, with residues Asp486 and Glu517 acting as nucleophile and general acid/base, respectively. The conformations of ADG are limited to the Southern hemisphere of the Cremer-Pople sphere by the 3,6-anhydro bridge, and the enzyme active site further restricts them, so the $B_{1,4}$ conformation is the one adopted by the —textit-1 sugar at the MC. This is consistent with the newly resolved structures. The reaction FELs, together with kinetic experiments, point towards the deglycosylation step being rate limiting.

5.3 Summary and conclusions

In this Chapter we have investigated the two-step hydrolytic mechanism of neo- β/κ -carrageenans, containing the α -linked 3,6-anhydro-D-galactosyl moiety, by *ZgGH129*. The hypothesized retaining, double displacement with rate-limiting deglycosylation step was confirmed, as well as the role of the residues Asp486 and Glu517 of nucleophile and general acid/base, respectively. This was corroborated with chemical, biochemical and structural data combined in inter-disciplinary effort between our group and the groups of Dr. Elizabeth Ficko-Blean and Dr. Keith Stubbs. To the best of our knowledge, *ZgGH129* is the first reported D-galactosidase operat-

ing through the Southern hemisphere of the Cremer-Pople sphere. This pinpoints the fundamental difference between D-galactose, which adopts conformations in the Northern hemisphere, and 3,6-anhydro-D-galactose, which is the 3,6-anhydro bridge that constricts the available conformations of this sugar at the Southern hemisphere. *ZgGH129* has evolved to take advantage of this, thus further restricting the conformational landscape to $B_{1,4}$ in the enzyme-substrate complex, which evolves through $B_{1,4} \rightarrow [E_4]^\ddagger \rightarrow E_4/{}^1C_4$ and $E_4/{}^1C_4 \rightarrow [E_4]^\ddagger \rightarrow B_{1,4}$ itineraries.

This study describes the conformational changes of a bicyclic monosaccharide and reveals, combining different techniques, how marine heterotrophic bacteria degrade prominent algal polysaccharides. Overall, this study provides new insights for developing specific probes for enzymes processing anhydro-carbohydrates and enzyme engineering approaches for the use of algal polysaccharides.

The main conclusions of this Chapter are the following:

- *ZgGH129* hydrolyzes α -1,3-linked, 3,6-anhydro-D-galactose of algal carrageenans via a two-step retaining mechanism where the catalytic nucleophile and acid/base are Asp486 and Glu517 respectively. The second reaction step, the deglycosylation, is rate-limiting.
- Both the 3,6-anhydro bridge and the enzyme scaffold restrict the conformational preferences of ADG to the Southern hemisphere of the Cremer-Pople sphere.
- The ADG ring adopts a distorted $B_{1,4}$ conformation in the Michaelis complex of *ZgGH129*, and delineates a $B_{1,4} \rightarrow [E_4]^\ddagger \rightarrow E_4/{}^1C_4$ (glycosylation) and $E_4/{}^1C_4 \rightarrow [E_4]^\ddagger \rightarrow B_{1,4}$ (deglycosylation) conformational itineraries. This is the first reported Southern hemisphere itinerary for a D-galactosidase.
- Once the reaction ends, a conformational change from $B_{1,4}$ to 1C_4 further stabilizes the reaction products.

5.4 Computational details

In this study we modeled *ZgGH129* in complex with the natural substrate neo- β/κ -DP4 and the covalent intermediate with ADG to simulate the glycosylation and deglycosylation steps of the enzyme reaction mechanism. We also modeled the 3,6-anhydro-D-galactosyl moiety *in vacuo* and in the enzyme active site to study its conformational preferences. All simulations were analyzed using VMD version 1.9.3 (<http://www.ks.uiuc.edu/Research/vmd/>) [130] the CPPTRAJ [131] module of AMBER [133] and the open-source PLUMED library [150].

5.4.1 Model of the *ZgGH129*-substrate complex and glycosylation reaction

5.4.1.1 System preparation

For building the Michaelis complex (MC) model, we used as initial coordinates the crystal structure of *ZgGH129*_Asp486Asn with neo- β/κ -DP4 (PDB 8RZK resolution 2.29 Å), a nucleophile variant corresponding to the reactant state of the glycosylation step. This structure was solved by the group of Dr. Elizabeth Ficko-Blean at the Station Biologique de Roscoff, CNRS Sorbonne Université. We reverted the Asp486Asn mutation using the atomic coordinates of the Asp486 sidechain from the *ZgGH129*_ADG crystal structure (PDB 8RZG, resolution 1.81 Å), also solved by our collaborators. As all *ZgGH129* structures are homodimers with equivalent active sites formed by residues belonging to both subunits, we used the full dimer to build all subsequent protein-saccharide models (chain A from residues 1 to 665 and chain B, from 666 to 1330, see Figure S5.2). We used PropKa3.0 [136] and H++ [137] to assign the protonation states of the titratable residues at pH = 5.0, as well as visual analysis of their environment. We assigned the carboxylate form for all Asp and Glu residues except for the acid/base, Glu517, in its protonated form to model the glycosylation reaction. Close to Glu517 there is Glu531, although Glu517 is closer to the scissile bond oxygen (chain A; $O_{Glu517}-O1 = 3.02$ Å, chain B; 3.05 Å) than Glu531 (chain A; $O_{Glu531}-O1 = 7.18$ Å, chain B; 7.33 Å), therefore is a more likely candidate for direct protonation of the scissile bond O. Glu531 coordinates a crystallographic water which in turn interacts with ADG at subsite -1. The system was solvated with 50204 water molecules in a cell of $132.73 \times 23.02 \times 122.06$ Å, spanning 15.0 Å from the enzyme surface, and 17 Cl⁻ ions to achieve neutrality. Crystallographic waters were included in the model.

5.4.1.2 Classical Molecular Dynamics

Classical molecular dynamics simulations of the system $ZgGH129\text{-nep}\beta/\kappa\text{-DP4}$ were performed with the AMBER package [133]. We employed the forcefields FF14SB (enzyme) [153], GLYCAM06j (saccharide and sulfate) [64] and TIP3P (solvent) [141], respectively. The non-reducing end ADG and the β -1,4-linked ADG units, not included in GLYCAM, were parametrized with the Antechamber suite [155] of AMBER with GLYCAM06j parameters. Atomic RESP charges were computed at the HF/6-31G* level of theory using Gaussian version 09.D01 [154]. The protocol for classical MD simulations was the following: first, an energy minimization of the solvent for 20000 cycles (first with the steepest descent scheme, then switching to conjugate gradients), then a subsequent minimization of the whole system. Next we performed a heating protocol up to 300 K consisting in 500 ps runs in the NVT ensemble, with a 1 fs timestep. First, only the solvent was heated at 100 K, then all atoms to 100, 200 and 300 K consecutively. Subsequently the density was converged to 1.0 g cm^3 in a 500 ps run in the NPT ensemble at 300 K and 1 bar with the Berendsen barostat [142]. The simulation continued in the NVT ensemble with a 2 fs timestep in four replicas of 50 ns (total of 200 ns). We applied the SHAKE algorithm for constraining bonds involving H in the NPT and NVT runs [143]. Harmonic restraints were applied to ensure sampling of the crystallographic $B_{1,4}$ ADG conformation at subsite -1. A representative snapshot was chosen for continuing the conformational studies of ADG.

5.4.1.3 QM metadynamics of isolated 3,6-anhydro-d-galactose

The conformational FEL of the isolated α -3,6-anhydro-D-galactose pyranose ring was computed with QM metadynamics. We selected the starting structure from a classical MD trajectory snapshot of the simulations described above (section 5.4.1.2 featuring the ring in a $B_{1,4}$ conformation. For this model the enzyme, solvent cell and most of the substrate atoms were stripped, leaving only the terminal ADG at subsite -1. The anomeric carbon was capped with a α -OH exocyclic group. We defined a cell measuring $11.92 \times 13.44 \times 11.79\text{ \AA}$ to enclose the 21-atom system, which was treated with a QM approach. We used the Car-Parrinello method as implemented in the CPMD code (<http://www.cpmc.org>) [75], which applies the Density Functional Theory (DFT) for calculating the electronic structure with the Perdew, Burke and Ernzerhoff (PBE) generalized gradient-corrected approximation [71]. PBE performance has been assessed for carbohydrate ring conformations [73] and has been applied in previous studies of GH catalysis [29,36]. A plane waves basis set with a 70 Ry kinetic energy cutoff was used to expand the Kohn-Sham orbitals

and we used norm-conserving, *ab initio* Troullier-Martins pseudopotentials [74]. The system was heated to 300 K with a Nosé Hoover thermostat [148] in a 4 ps MD run, with fictitious electronic mass of 700 a.u. and 0.12 fs timestep.

For sampling the conformational FEL the standard metadynamics scheme of the open-source, community-developed PLUMED [149,150] was coupled to CPMD. To drive the metadynamics bias we used 3 collective variables corresponding to the Cartesian projections of the six-member ring puckering coordinates divided by the amplitude: q_x/Q (CV1), q_y/Q (CV2) and q_z/Q (CV3) (Figure S5.4.a), which have been used to compute the conformational landscapes of other carbohydrates [16,18,19,41,222]. Gaussian-like kernels were deposited every 36 fs with a width set at 0.045, 0.040 and 0.045 Å for each CV respectively, considering their amplitude oscillations along the QM/MM MD simulation. The height of the terms was set at 0.4 kcal mol⁻¹. The simulation was stopped after deposition of 11998 Gaussian terms, when the FEL did not display any more significant changes (Figure S5.4.b).

5.4.1.4 QM/MM metadynamics of 3,6-anhydro-d-galactose in ZgGH129 active site

We computed the conformational FEL of ADG as part of the substrate in the active site of ZgGH129 using hybrid QM/MM MD coupled with metadynamics. The starting coordinates correspond to the same trajectory snapshot of DP4-ZgGH129 classical MD simulations (section 5.4.1.2) we used to start the isolated ADG puckering simulations (section 5.4.1.3), where ADG is in $B_{1,4}$ conformation. The system corresponding to the ZgGH129-substrate complex and the solvent cell was treated with QM/MM. The QM region amounting to 30 atoms included the ADG ring at subsite -1 and some atoms of the adjacent α -1,3-D-gal unit at subsite +1, which were enclosed in a QM cell of $13.80 \times 13.64 \times 14.03$ Å (net charge of 0). To saturate the C-C bonds separated by the QM/MM boundary at the α -1,3-D-gal we applied 2 monovalent pseudopotentials. The MM region contained the remaining 173308 atoms of the system. The hybrid QM/MM approach was performed as implemented in the CPMD code, combining DFT-based Car-Parrinello MD (QM) [75] with force-field MD (MM), developed by Laio, VandeVondele and Roethlisberger [78,144]. The charges treatment near the interface of the QM and MM regions were treated with the Hamiltonian electrostatic coupling scheme developed by Laio *et al.* [78], selecting radii at 13.22 (NN) and 18.52 (ESP) Å according to the approach advised for QM/MM simulations of GHs [79]. The PBE generalized gradient-corrected approximation was applied [71], and a plane-waves basis set with a kinetic energy cutoff of 70 Ry with Troullier Martins *ab initio* pseudopotentials [74]. The system was equi-

librated up to 300 K with a Nosé-Hoover thermostat [148], a fictitious electronic mass of 600 au and a 0.12 fs timestep for 2 ps.

For enhancing the sampling of conformational transitions we applied the standard metadynamics scheme by PLUMED [149,150] coupled to CPMD (see Section 5.4.1.3). In this case, in view of the obtained results for the isolated ADG FEL, only 2 CVs were used to drive the metadynamics bias in this simulations, q_x/Q (CV1) and q_y/Q (CV2) were selected (Figure S5.6.a). q_z/Q accounts for the transitions across the equator to change hemispheres, and since the Northern conformations are high in energy, this change is not expected to take place [41]. Gaussian-like kernels were added every 38 fs with a 0.045 Å (CV1) and 0.035 Å (CV2) width and a 1.0 kcal mol⁻¹ initial height, which was later decreased to 0.4 kcal mol⁻¹ to refine the energy accuracy. The simulation was stopped once the FEL displayed no additional significant changes, after depositing 4640 Gaussians.

5.4.1.5 QM/MM molecular dynamics of the *Zg*GH129-substrate complex

QM/MM MD simulations of the Michaelis complex of *Zg*GH129 were performed starting from a structure in where the ADG unit at subsite -1 was in $B_{1,4}$ conformation, extracted from the puckering QM/MM metadynamics trajectory of section 5.4.1.4 and corresponds to the most stable region of the conformational FEL. The QM/MM MD approach implemented in CPMD was applied as explained in section 5.4.1.4. The QM region included 47 atoms (ADG at subsite -1, some atoms of α -1,3-D-gal at subsite +1 and the sidechains of Asp486 and Glu517), including 4 monovalent atoms to saturate the C-C bonds divided by the QM/MM boundary (see Figure S5.7) [223]. The dimensions of the QM cell were $17.50 \times 16.40 \times 14.10$ Å with a net charge of -1. The remaining 173296 atoms of the system were part of the MM region. The radii for the QM/MM layer interface were set at 9.53 (NN), 13.76 (MIX) and 15.88 (ESP) Å. 6 ps QM/MM MD simulation was performed with a 0.12 fs timestep and a fictitious electronic mass of 600 au. A trajectory snapshot of this simulation was used as starting structure for the subsequent QM/MM metadynamics simulation of the glycosylation step.

We performed additional simulations of the *Zg*GH129 model starting from a 1C_4 conformer of ADG, starting from a structure of the 1C_4 basin of the conformational FEL of section 5.4.1.4 and applied the same protocol for QM/MM setup as described in the paragraph above. An equivalent QM region was enclosed in a $17.05 \times 16.79 \times 15.48$ Å cell (net charge -1). The electrostatic embedding radii were set at 9.53 Å (NN), 13.76 Å (MIX) and 15.88 Å (ESP). A 10 ps QM/MM MD simulation was

performed at 300 K with a 0.12 fs timestep and a 600 au fictitious electronic mass.

5.4.1.6 QM/MM metadynamics of the glycosylation step

The glycosylation reaction of *ZgGH129* with neo- β/κ -DP4 was simulated applying the metadynamics scheme of PLUMED coupled to CPMD QM/MM MD (see section 5.4.1.4). To drive the system from the reactants state (MC) to the covalently bound reaction intermediate (GEI) we selected two collective variables corresponding to the main reactive events taking place in the predicted mechanism. CV1 is the difference between the C1-O_{Asp486} and the C1-O1 distances, representing the glycosidic bond cleavage and enzyme-saccharide covalent bond formation. CV2 is the O1-H_{Glu517} distance, accounting for the proton transfer from the acid catalyst to the scissile bond oxygen (Figure S5.8.a). Gaussian-like functions were added every 40.8 fs considering the slowest motion of the CVs monitored during the QM/MM MD of section 5.4.1.5. Gaussian width was set at 0.15 Å for both CVs after evaluating their variation during the aforementioned QM/MM unbiased run, and the height was set initially at 1.2 kcal mol⁻¹. The simulation was stopped when all relevant reaction FEL states (MC, TS and GEI) were sampled at least once, following the recommended first recrossing criterion [151], after depositing 1132 Gaussian terms, which amounts to approximately 50 ps of simulation time (Figure S5.8.b). For improving the accuracy of the computed reaction free energy barrier the simulation was restarted from a point where the reactants basin was half filled lowering the Gaussian height to 0.5 kcal mol⁻¹. The free energy barrier averages to 14.3 ± 0.1 kcal mol⁻¹ and the obtained mechanism was the same (with 1.2 kcal mol⁻¹, 14.2 kcal mol⁻¹, with 0.5 kcal mol⁻¹, 14.3 kcal mol⁻¹). This approach by Nair *et al.* [86] has been used in other GH catalysis studies [9]. The experimental free energy barrier was estimated from the experimental average rate constants for different substrates applying the Eyring-Polanyi expression to approximate Transition State Theory [122].

5.4.2 Model of the *ZgGH129* glycosyl-enzyme intermediate and deglycosylation reaction

5.4.2.1 System preparation

To build the glycosyl-enzyme intermediate model of *ZgGH129* we obtained the initial atomic coordinates from the crystallographic structure in complex with the inhibitor 3,6-anhydro-2-deoxy-2-fluoro- α -D-galactopyranosyl (1,2-diF-ADG) (PDB ID 8RZI, resolution 1.96 Å), solved by our collaborators. To obtain the corresponding ADG structure, we substituted the F atom at C2 of the inhibitor for an O (FigureS5.12),

which is the covalently bound reaction intermediate obtained after the first reaction step. Our model corresponds to the state where the leaving group of the first reaction step has already left the enzyme active site cavity, thus this structure will be referred as GEI' thereafter. For this model, as well as for the MC (section 5.4.1.1), both subunits of the enzyme were included in the model, as both contribute to the integrity of the active site. The same protonation states for the titratable residues were assigned as in the enzyme-substrate complex mode (section 5.4.1.1), with the only difference at Glu517 being in the carboxylate form, consistent with its predicted role of base catalyst of the second reaction step. The GEI' system was solvated with 46045 water molecules in a cell with dimensions $121.74 \pm 121.07 \pm 127.14$ Å, accounting from 15.0 Å from the enzyme surface in each direction. 18 Cl⁻ ions were added for neutralizing the system net charge. All crystallographic waters were included in this model.

5.4.2.2 Classical Molecular Dynamics

The AMBER package [133] was used to perform all subsequent classical MD simulations with the forcefields FF14SB (enzyme) [153], GLYCAM06j (saccharide) [64] and TIP3P (solvent) [141]. We calculated the parameters for the enzyme Asp486 covalently bound to the ADG unit at subsite -1 using the Antechamber [155] suite of programs from AMBER applying GLYCAM06j and GAFF parameters [156]. Atomic RESP charges for the ADG-Asp were computed with Gaussian version 09.D01 [154] at the HF/6-31G* level of theory. The protocol for classical MD simulations was applied as detailed in section 5.4.1.2. After the density equilibration in NPT converged to 1.0 g cm⁻³, we extended the simulation in the NVT ensemble for 300 ns in 6 independent replicas of 50 ns to increase the sampling of water molecules near the active site that may take up the catalytic role in the subsequent reaction step. We analyzed the dynamics of water molecules near the active site during all classical MD replicas with an in-house Python script developed by Dr. Oriol Esquivias during his PhD. research in our group [30]. We found a trajectory snapshot with a water molecule suitably oriented for catalysis with distances $O_{wat}-C1 = 3.22$ Å and $H_{wat}-O_{Glu517} = 1.72$ Å. This water was included in the QM region of the subsequent QM/MM MD simulations as the putative nucleophile of the deglycosylation reaction.

5.4.2.3 QM/MM molecular dynamics

The QM/MM MD simulations of ZgGH129 in complex with the ADG covalent intermediate were performed applying the methodology as for the enzyme-substrate

complex described in section 5.4.1.5, as well as the same parameters for plane-waves expansion and DFT functional. For the present system, the QM region contained 36 atoms including the side chains of the covalently bound Asp486 and Glu517, the ADG linked to the enzyme, a water molecule at 3.22 Å of the ADG C1 and 1.72 Å of O_{Glu517} and two monovalent atoms to saturate the C-C bonds separated by the QM/MM boundary (see Figure S5.14). The QM region, with net charge of -1, was enclosed in a cell with dimensions $14.89 \pm 14.32 \pm 16.04$ Å. The MM region included the remaining 162552 atoms of the system. The radii for the multilayer scheme for charges treatment was set at 9.53 (NN), 13.76 (MIX) and 15.88 (ESP) Å. The system was heated to 300 K with a Nosé-Hoover thermostat [148] in a 6 ps MD run, with a fictitious electronic mass of 600 au and a 0.12 fs timestep.

5.4.2.4 QM/MM metadynamics of the deglycosylation step

The QM/MM metadynamics simulation of the de deglycosylation reaction was modeled starting from a QM/MM MD trajectory snapshot of the GEI' (section 5.4.2.3). We selected 2 CVs to drive the metadynamics bias, as implemented by the PLUMED plugin coupled to CPMD (also see section 6.4.3.2). The first collective variable (CV1) accounts to the formation of the new covalent bond, $d(C1-O_{wat})$, and the second (CV2), to the covalent bond degradation of ADG with Asp486, $d(C1-O_{Asp486})$ (Figure S5.15.a). The Gaussian-like kernels were added every 36 fs with an initial height of 1.0 kcal mol⁻¹ and a width of 0.15 and 0.05 Å for each CV, respectively, based on their variation along QM/MM MD. The simulation was stopped when all relevant states of the reaction coordinate were sampled at least once (initial intermediates state GEI', transition state TS₂ and reaction product P), therefore after one re-crossing event to the starting state, approximately after 40 ps of simulation time and deposition of 1068 Gaussian functions (Figure S5.15.b). To refine the free energy reaction barrier, we restarted the simulation from the point in which the reaction intermediate basin (GEI') was half-full and changing the Gaussian height to 0.5 and 0.25 kcal mol⁻¹. The obtained mechanism was the same, with an average free energy barrier of 17.2 ± 0.7 kcal mol⁻¹.

5.4.2.5 QM/MM metadynamics of the ZgGH129-ADG product complex

We computed the conformational FEL of the hydrolysis reaction product ADG in the active site of ZgGH129 with QM/MM MD metadynamics. We started from a snapshot of the reaction products (P), where ADG is in $B_{1,4}$. The product system included 162547 water molecules in a cell of $116.57 \times 115.93 \times 121.74$ Å with 18

Cl^- ions. For QM/MM MD simulations of the product complex we enclosed the QM system corresponding to the full ADG molecule in a cell of $11.74 \times 12.71 \times 13.2$ Å (21 atoms, net charge 0). The MM region included 162565 atoms. QM/MM MD was performed with the CPMD code applying the protocol described in section 5.4.1.5 [75,78]. The radii for the electrostatic embedding regions were 10.05 Å, 14.76 Å (NN) and 13.35 (MIX) Å (ESP). We equilibrated the system at 300 with 2 ps of QM/MM MD, using a fictitious electron mass of 600 au and a 0.12 fs timestep.

We applied the metadynamics scheme as implemented in PLUMED with CPMD [150] for sampling the conformational FEL of ADG. We used two CVs to drive the bias corresponding to q_x/Q (CV1) and q_y/Q (CV2). The deposition time of Gaussian functions was set at 36.3 fs, and the width at 0.03 and 0.04 Å, respectively. The initial Gaussian height was 0.6 kcal mol⁻¹. The simulation ran approximately for 50 ps, after depositing 1263 Gaussians.

5.5 Supplementary information

CAZY family	Conformational itinerary	Enzyme function	Reference
GH1	${}^{1,4}B \rightarrow [{}^4H_3/{}^4E]^{\ddagger} \rightarrow {}^4C_1$	β -D-galactosidase	[224]
GH2	${}^4H_3 \rightarrow [{}^4E]^{\ddagger} \rightarrow {}^4C_1$	β -D-galactosidase	[224, 225]
GH27	${}^4C_1 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow {}^1S_3$	α -D-galactosidase	[226, 227]
GH35	${}^{1,4}B \rightarrow [{}^4H_3/{}^4E]^{\ddagger} \rightarrow {}^4C_1$	β -D-galactosidase	[224]
GH36	${}^2E/{}^2H_3 \rightarrow [E_3/{}^4H_3]^{\ddagger} \rightarrow {}^2H_3/E_3$	α -D-galactosidase	[228]
GH42	${}^1S_5 \rightarrow [{}^4E/{}^4H_5]^{\ddagger} \rightarrow {}^4C_1$	β -D-galactosidase	[224]
GH59	${}^1S_3 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow {}^4C_1$ ${}^4C_1 \rightarrow [{}^4H_3]^{\ddagger} \rightarrow {}^4C_1$	β -D-galacto- cerebrosidase	[9]
GH129	$B_{1,4} \rightarrow [E_4]^{\ddagger} \rightarrow E_4/{}^1C_4$ $E_4/{}^1C_4 \rightarrow [E_4]^{\ddagger} \rightarrow B_{1,4}$	α -3,6-anhydro- -D-galactosidase (<i>Zg</i> GH129)	This study
	Unknown, 4C_1 in crystal complexes	α - <i>N</i> -acetyl- galactosaminidase (<i>Bb</i> GH129)	[229]

Table S5.1: Conformational catalytic itineraries of D-galactosidases, both α - and β - acting, elucidated across different GH families.

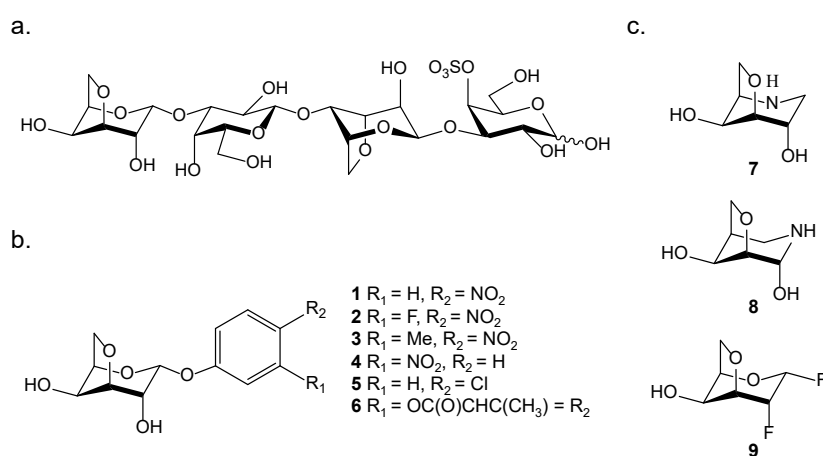


Figure S5.1: a. Structure of the neo- β - κ -oligo-carrageenan tetrasaccharide (DP4). b. Structures of aryl-based compounds designed to probe the activity of *Zg*GH129: **1** *p*NP-ADG, **2** 3-F-*p*NP-ADG, **3** 3-MepNP-ADG, **4** *m*NP-ADG, **5** CIP-ADG, **6** MeUMB-ADG. c. Structures of designed ADG mimics to probe the catalytic mechanism of *Zg*GH129: **7** AD-DGJ, **8** ADG-IF and **9** 1,2-diF-ADG. Our collaborators solved crystal structures of *Zg*GH129 in complex with the natural substrate DP4, the reaction product ADG and the probes **7**, **8** and **9** (PDB IDS 8RZK, 8RZG, 8RZH, 8RZJ and 8RZI, respectively).

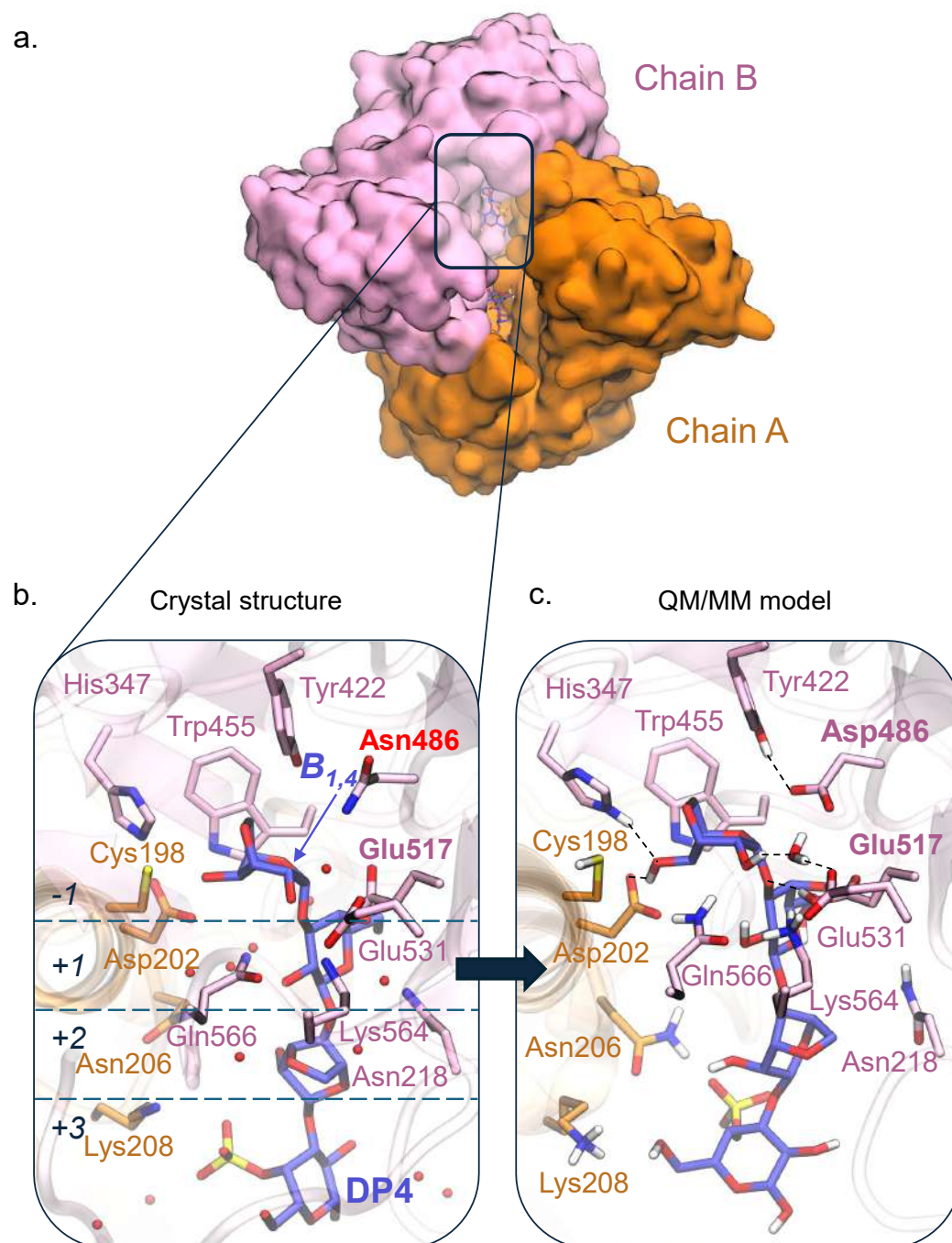
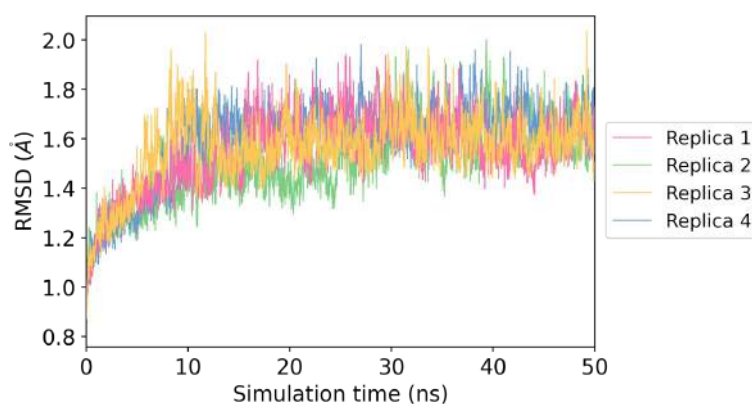


Figure S5.2: a. Crystal structure used to build the model of *ZgGH129* in complex with the tetrasaccharide neo- β/κ DP4 considering both subunits of the structure. b. Detail of one of the active sites. We built the WT reverting the nucleophile Asp486Asn mutation (coordinates for Asp486, from structure with PDB ID 8RZG). c. Enzyme-substrate interactions after classical and QM/MM MD.

a.



b.

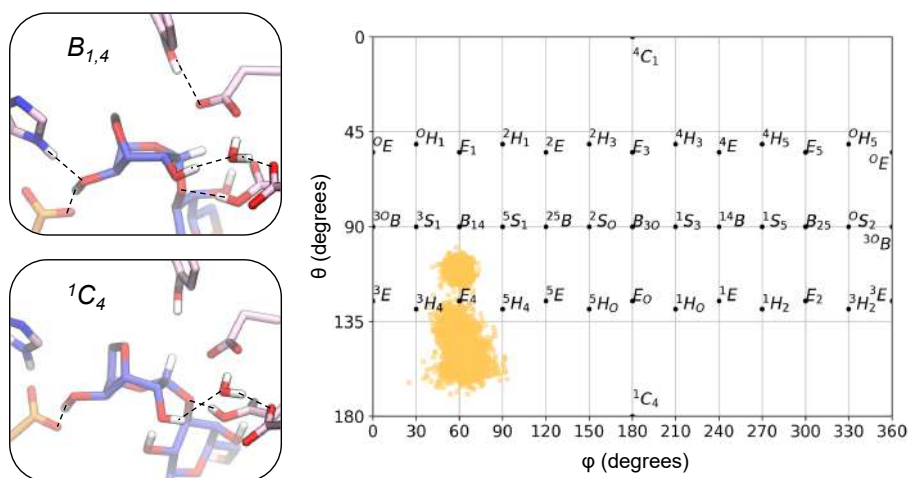
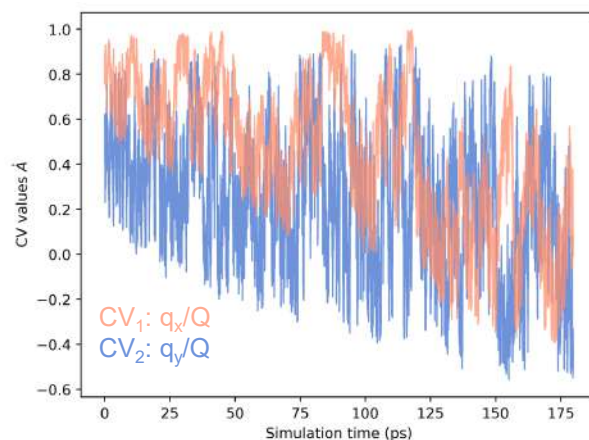


Figure S5.3: a. Protein backbone RMSD evolution of the classical MD simulations of *Zg*GH129 in complex with neo- β - κ -DP4 carrageenan. b. (Left) Active site detail of ADG in $B_{1,4}$ and 1C_4 . (Right) Mercator projection of the ADG ring conformations in classical MD. Since force-field methods do not accurately describe puckering ring conformations, we selected a snapshot in $B_{1,4}$ and analyzed the *Zg*GH129-ADG conformational landscape with QM/MM metadynamics.

a.



b.

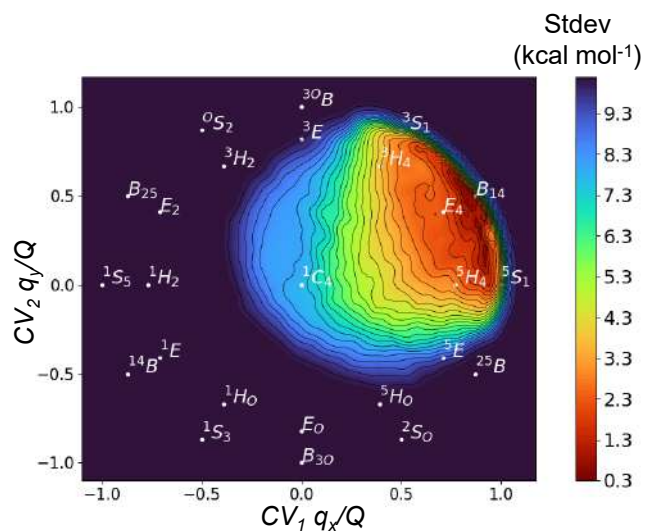


Figure S5.6: a. Simulation time evolution of the collective variables used to reconstruct the conformational FEL of ADG as part of neo- β - κ -DP4 in the ZgGH129 active site. b. Representation of the standard deviation of the conformational FEL of Figure 5.7.a represented with the free energy estimator developed by Tiwary and Parrinello [230] for assessing the convergence of the simulation.

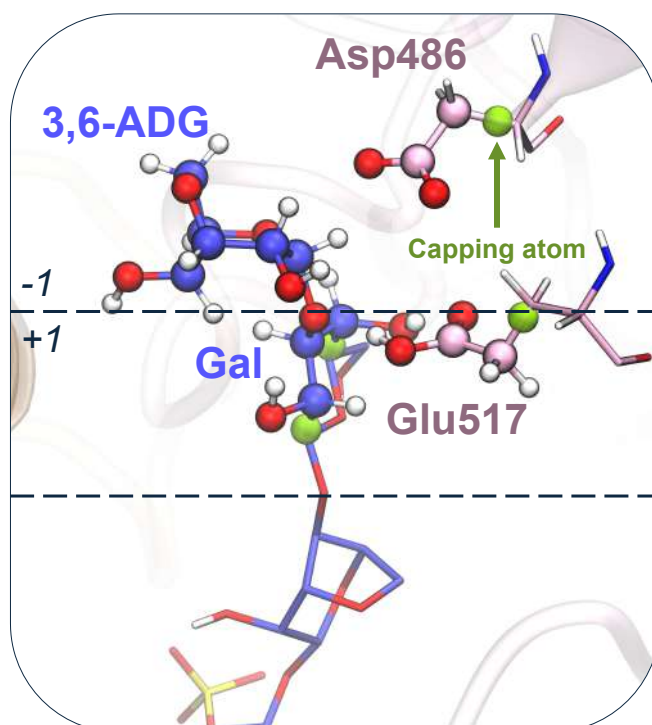
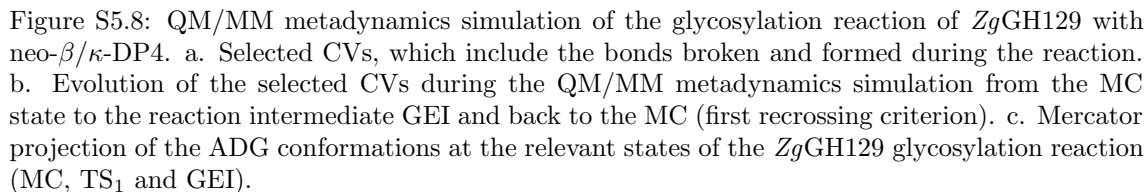


Figure S5.7: Representation of the QM atoms selected for the QM/MM simulations of the *ZgGH129-neo-β/κ-DP4* MC. In ball-and-stick representation, 47 QM atoms corresponding to the reactive ADG at subsite -1, a part of the adjacent Gal, the sidechains of Asp486 and Glu517 and the capping monovalent atoms (green balls). As sticks, the remaining substrate and backbone atoms in the MM region.



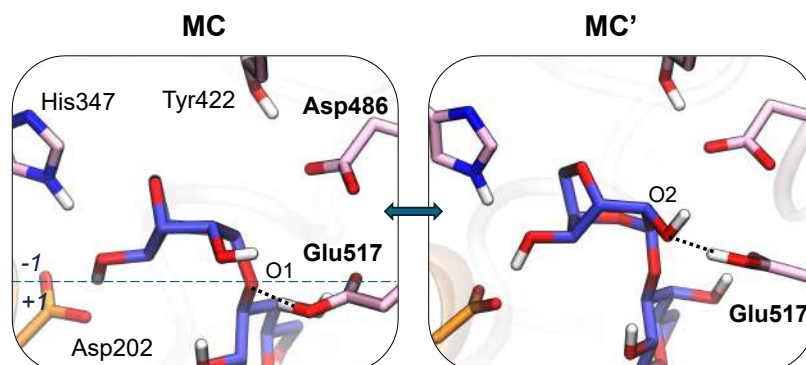


Figure S5.9: Representative structures of the two states in the reactants basin during QM/MM metadynamics of the glycosylation step. MC is the catalytic Michaelis complex, where the H_{Glu517} forms a H-bond with O1. Right: MC' is a pre-catalytic configuration of the enzyme-substrate complex where Glu517 forms a H-bond with the 2-OH exocyclic group of ADG.

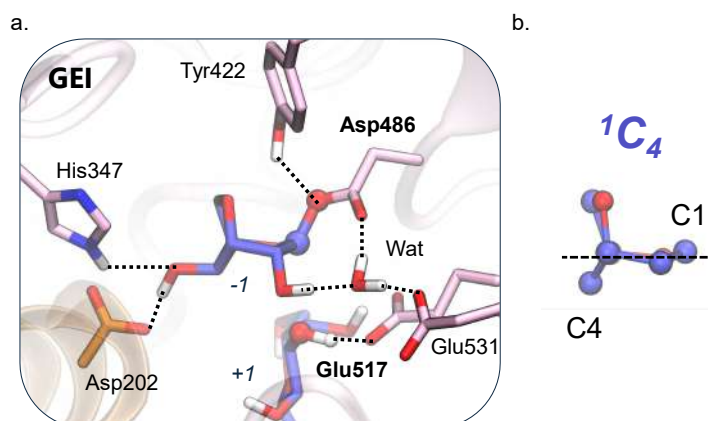


Figure S5.10: a. Active site detail of a QM/MM metadynamics snapshot of the GEI complex where the ADG bound to Asp486 is in a conformation close to the inverted chair 1C_4 . b. Ball-and-stick representation of the ADG ring shows that, while this conformation can be described as 1C_4 , the difference with the E_4 is subtle: the geometrical constraint imposed by the 3,6-anhydro bridge positions the C4 far from the ring plane, more than the upwards position of C1.

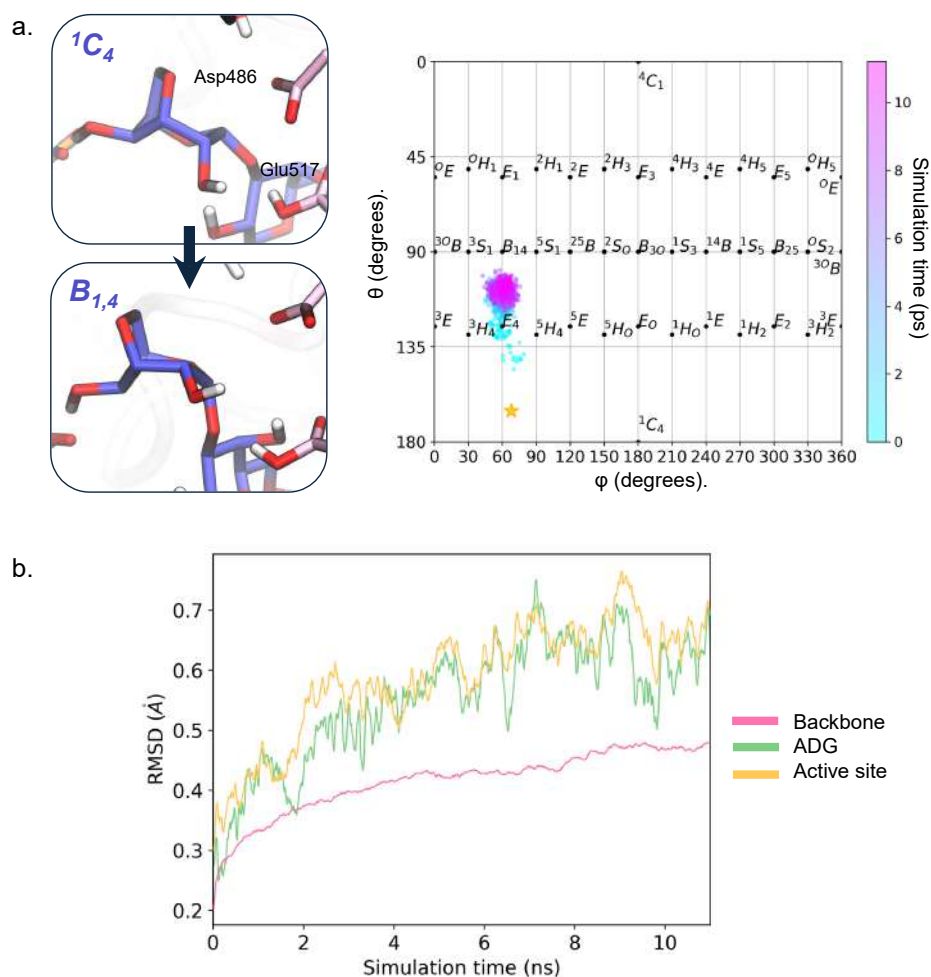


Figure S5.11: a. Left: Top, starting trajectory snapshot from the QM/MM metadynamics simulation from section 5.4.1.4 with ADG in 1C_4 , bottom, final snapshot of the QM/MM MD (unbiased), where ADG adopts $B_{1,4}$ conformation. Right: Mercator projection of the puckering ring conformation of ADG in polar coordinates along the QM/MM MD simulation. Yellow star: conformation of the chosen frame to start QM/MM. b. RMSD evolution along the QM/MM MD simulation.

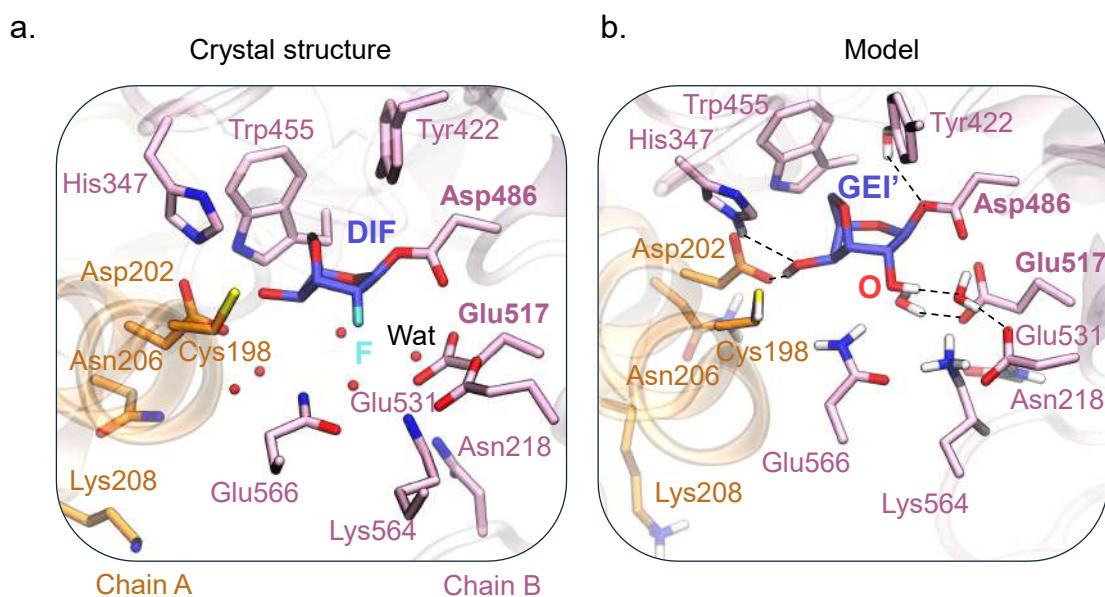


Figure S5.12: a. Detail of the active site of the *ZgGH129* crystal structure in complex with the DIF inhibitor used to build the covalent glycosyl-enzyme intermediate (GEI'). Represented as a cyan stick, the F atom present in the covalently bound inhibitor that was changed to O. The presence of F possibly precludes water molecules coordinating the saccharide moiety with Glu531, otherwise present near the active site, including the one coordinating Glu517. b. Detail of the modeled complex (after classical MD and QM/MM MD), where the saccharide engages in several interactions with the enzyme, including two water-mediated interactions with both Glu517 and Glu531.

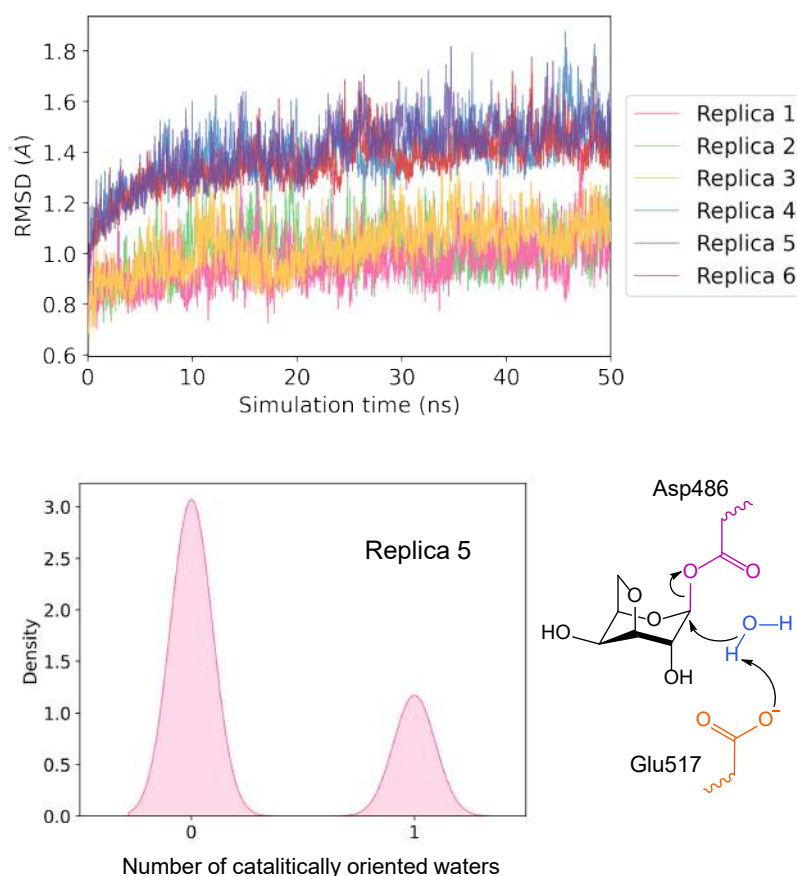


Figure S5.13: a. Protein backbone RMSD evolution during classical MD simulations of the reaction intermediate covalent complex of *ZgGH129* with ADG.

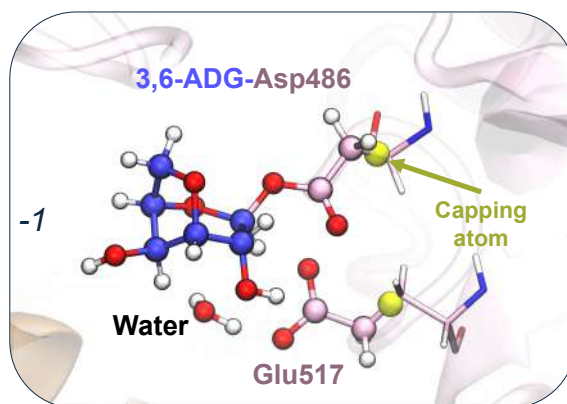
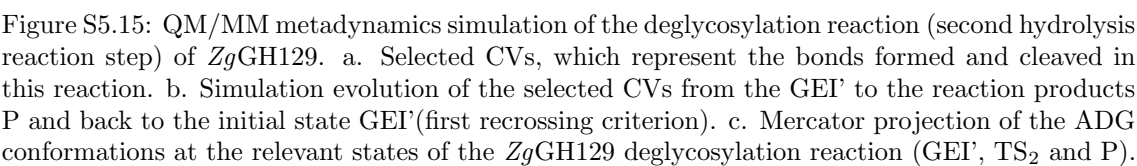


Figure S5.14: Selected QM atoms for the *ZgGH129*-ADG QM/MM simulations (GEI'). Ball-and-stick representation: the 36 QM atoms (ADG covalently bound to Asp486 at subsite -1, the sidechains of the catalytic Asp486 and Glu517 and the capping atoms as yellow balls). Thin stick representation: backbone MM atoms.



a.

Substrate	pK_a	λ (nm)	ϵ ($M^{-1}cm^{-1}$) ^a	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}mM^{-1}$)
3-F- <i>p</i> NP-ADG 2	6.42 ^[83]	385	18450	0.35 ± 0.03	25.4 ± 0.5	72.6 ± 4.4
<i>p</i> NP-ADG 1	7.18 ^[83]	405	18500	0.58 ± 0.04	28.4 ± 0.6	49.0 ± 3.5
3-MepNP-ADG 3	7.4 ^[84]	395	16200	0.64 ± 0.04	29.3 ± 0.5	45.8 ± 2.3
MeUMB-ADG 6 ^b	7.5 ^[83]	360	8600	—	—	60.1 ± 1.5
<i>m</i> NP-ADG 4	8.39 ^[83]	390	1400	0.84 ± 0.07	23.9 ± 0.8	28.5 ± 2.5
4-CIP-ADG 5	9.47 ^[83]	290	1800	1.01 ± 0.10	24.7 ± 1.0	24.5 ± 2.6

^aDetermined experimentally. ^bFor **6**, only initial velocities for low concentrations (0.05-0.8 mM) could be obtained due to low solubility. Nonetheless, the second order rate constant could be calculated from the slope of the initial linear region. ZgGH129 was found to be stable over the timeframe of the experiment.

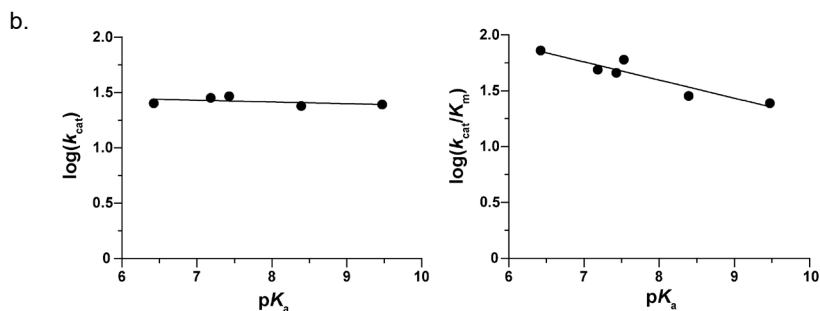


Figure S5.16: a. Kinetic parameters for ZgGH129 with different aryl-substituted 3,6-anhydro-D-galactose probes. For structures of the probes, check Figure S5.1. b. Brønsted plots with different pK_a corresponding to the different aryl probes. Adapted from [57]

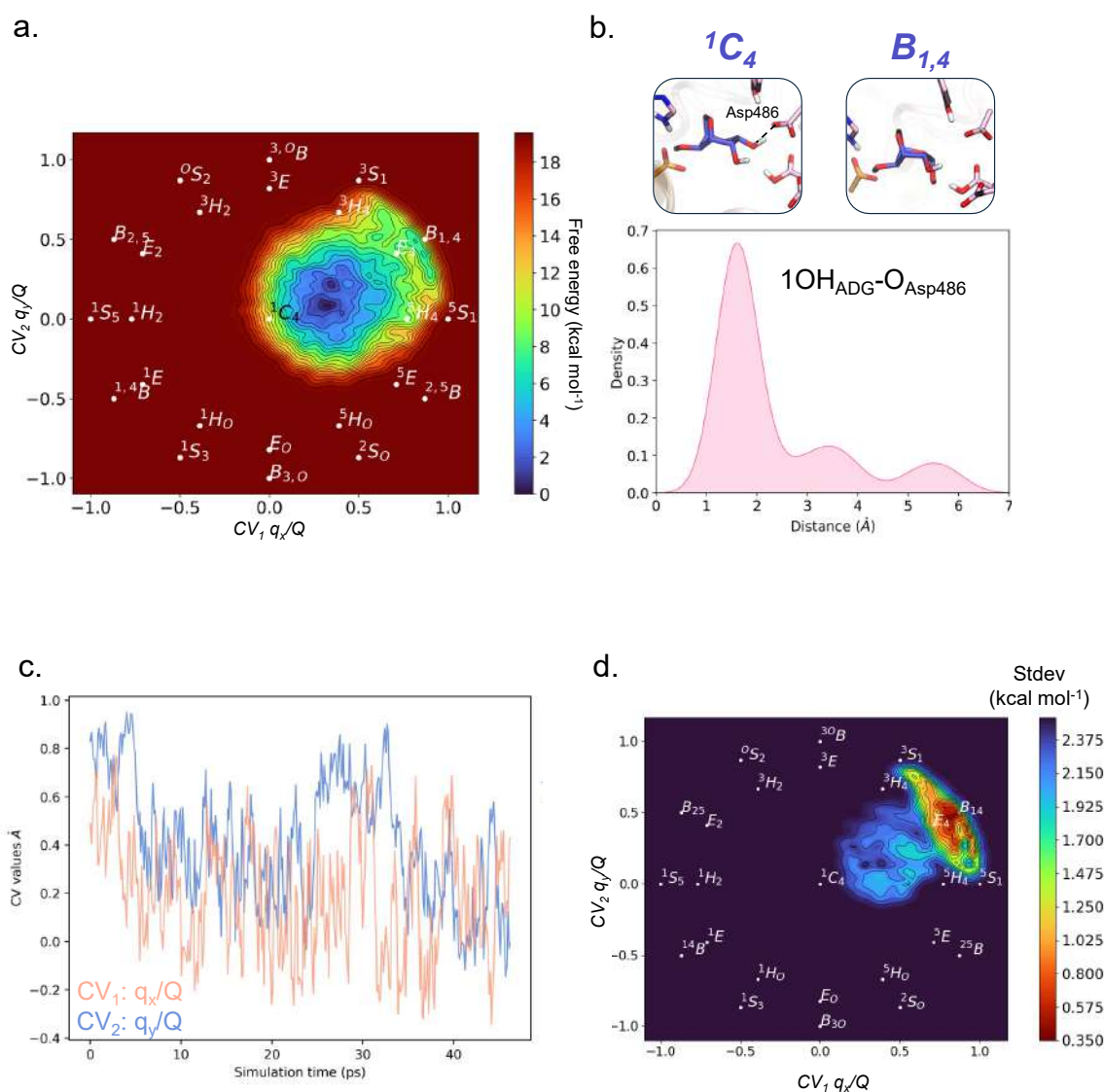


Figure S5.17: a. Reconstructed conformational FEL of the reaction product. b. Distribution of the interaction distance between the nucleophile Asp486 and the 1-OH of ADG. c. Simulation time evolution of the CVs used to bias the QM/MM metadynamics. d. Representation of the standard deviation of the conformational FEL represented with the free energy estimator of the FEL standard deviation developed by Tiwary and Parrinello [230] for assessing the convergence of the simulation. This simulation is ongoing in order to improve the sampling in the 1C_4 region.

Chapter 6

Substrate recognition and catalytic mechanism of human GH29 fucosidase

Abstract:

The degradation of L-fucosylated glycans is paramount for many biological functions such as those related to cell waste management and cell-to-cell signaling. Fucosidases are a type of glycoside hydrolases cleaving terminal α -linked fucose from fucosylated motifs. Their diversity and the lack of structural information for animal fucosidases has challenged the understanding of fucosylated glycans recognition and catalysis. Recent structures of human fucosidase FucA1 show that the enzyme, which was known to operate with a retaining, two step classical GH mechanism, used the catalytic pair Asp225 and Asp276. FucA1 is reported to have a relatively relaxed substrate specificity compared to other bacterial fucosidases. However, it is not clear how the enzyme can accommodate different substrates in its active site, as no structures of the enzyme in complex with a natural substrate are available yet. Here we present models of the enzyme in complex with different natural substrates, corresponding to the different possible fucose linkage types (Fuc- α -1,2-Gal, Fuc- α -1,3/4/6-GlcNAc) and show that the main structural motif responsible for enzyme plasticity is a loop region from Asp258 to Lys287. This region displays high mobility to adapt to the different substrates and also contains the catalytic Asp276. We also uncover the mechanism of the first step of the hydrolytic mechanism of FucA1 with the natural substrate Fuc- α -1,2-Gal using QM/MM and metadynamics. Our study sheds light on the substrate recognition strategies by FucA1 and the degradation of fucosylated glycans, and opens doors to the design of specific inhibitors and enzyme engineering strategies.

6.1 Introduction

Fucose is a 6-deoxy hexose found across many prokaryotic and eukaryotic organisms, including mammals, where it functionalises complex carbohydrates like *N*-glycans, *O*-glycans and other glycoconjugates like glycolipids [3]. Fucosylation plays a crucial role in many biological processes involving cell-to-cell communication, including tissue development, inflammatory and immune responses, angiogenesis and cancer development [3, 231]. While less prevalent in prokaryotic organisms, fucosylation is key for bacterial cell signaling processes such as those related to cell adhesion and host immune response interaction [231]. It also constitutes an important nutrient for bacteria. For example, *Streptococcus pneumoniae* scavenges fucose from human tissues [35] and gut bacteria harvest human milk oligosaccharides (HMOs), of which fucose is a prevalent component [232], with an important role in the development of the gut microbiome, as discussed in chapter 3 of this Thesis. Therefore, the study of fucosylated glycans and glycoconjugates, as well as the enzymes involved in their synthesis, degradation, transport and modification, has potential applications in developing diagnosis tools, drug design and the synthesis of high added-value oligosaccharides.

Fucose, also known as 6-deoxy-L-galactose [3], is distinct from other vertebrate hexoses, both in its isomeric configuration and exocyclic substituent groups. While other monosaccharides, like galactose, are present in Nature in their D- forms, L-fucose is the form constituting vertebrate glycans [1]. Although Fuc and Gal share the configuration of C4, Fuc is a 6-deoxy sugar (Figure 6.1), with a methyl substituent in its carbon 5.

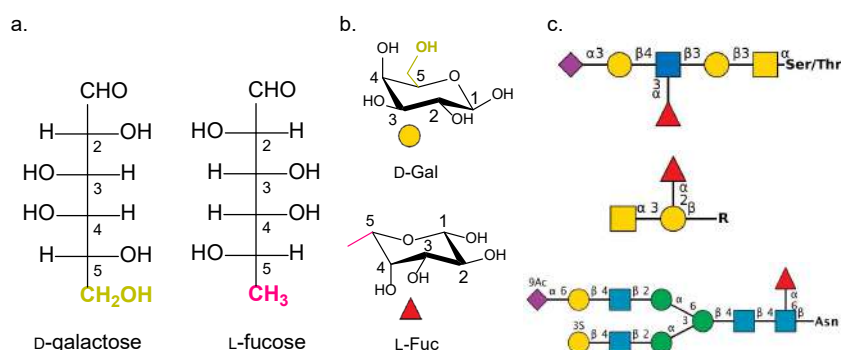


Figure 6.1: a. Fischer projections of D-Gal and L-Fuc. b. Cyclic forms and SNFG symbols (Symbol Nomenclature for Glycans). Fuc and Gal share C4 stereochemistry but differ in the exocyclic group at C6. c. (Top to bottom) human glycans with terminal Fuc: extended core 1 of *O*-GalNAc glycans, antigen from blood group A determinant, also in *O*-GalNAc glycans (R=Ser/Thr-linked oligosaccharide), and *N*-glycans. Adapted from [3] and [1], SNFG symbols with DrawGlycan (www.virtualglycome.org/DrawGlycan/) [5, 6].

Enzymes that process fucosylated glycans are essential in many biological functions, whose up- or downregulation is related to many diseases such as immune deficiencies and cancer. Therefore, research on fucose-containing saccharides degradation and synthesis can provide tools for both diagnosis and treatment of diseases [3]. Fucose is incorporated to glycans and glycoconjugates as part of post-translational modification processes by fucosyltransferases like protein *O*-fucosyltransferase 1 (POFUT1), which incorporates fucose to threonine or serine residues within specific sequences [12]. In humans, degradation of fucosylated glycans is related to immune response, cell waste management and many pathological processes [27]. FucA1 is part of essential degradation pathways of several glycans and glycoconjugates in the lysosome, and its impaired functioning leads to an abnormal accumulation of partially undegraded oligosaccharides, glycoproteins and glycolipids in tissue cells [233], causing the neurodegenerative disease fucosidosis [32].

FucA1 is classified in family GH29 of glycoside hydrolases. It is a promiscuous enzyme, with relaxed substrate specificity [27, 234–236], being able to cleave fucose from many different motifs like Fuc- α -1,2-Gal, Fuc- α -1,3-GlcNAc, Fuc- α -1,4-GlcNAc and Fuc- α -1,6-GlcNAc-linked substrates. Members of CAZY family GH29 are both bacterial and eukaryotic fucosidases, including *Streptococcus pneumoniae* SpGH29 [35] and *Bacteroides thetaiotaomicron* [221]. Fucosidases are known to follow a classical Koshland retaining mechanism [42]. The fucosyl-enzyme covalent intermediate has been obtained experimentally [221, 237, 238], using a nucleophile and general acid/base catalytic pair. Some fucosidases can also synthesize glycosidic linkages and produce fucosylated oligosaccharides [236, 239, 240]. Although the nucleophile has been identified as an aspartate in GH29 fucosidases with mass spectrometry analyses [241] and X-ray crystallography [237], the identity of the general base is not conserved within the family. Phylogenetic relationships and differences in substrate preferences led to the distinction of two GH29 subfamilies [238, 242]. While fucosidases from family A, among which human FucA1 is found, exhibit broad substrate specificity, members of subfamily B are specific for α -1,3 and α -1,4 linkages. GH29 subfamilies also diverge in the identity of the catalytic acid base, and in many bacterial GH29 fucosidases it has been identified as a glutamate residue [238]. However, the identity of the acid/base in human fucosidase remained a matter of debate until recently [243].

The tridimensional structure of human fucosidase was recently resolved by cryo-electron microscopy (cryo-EM) in 2022 by the group of Prof. Gideon Davies at the University of York, UK, both in the native unliganded form and in complex with the inhibitor deoxyfuconojirimycin (DFJ) [27] (Figure 6.2). Both structures

have 4 identical subunits with three glycosylation sites each, only one of which was partially resolved (GlcNAc-Asn236). This work provides structural evidence for the catalytic mechanism of this enzyme, which was suggested to be retaining [237] (Figure 6.3). The pair of catalytic residues was identified to be two aspartate residues, Asp225 and Asp276. This was an important breakthrough especially for the identity of the general acid/base catalyst, suggested to be Glu289 in previous work [238,243]. Thanks to the structural data it was possible to see that this residue lies far away from the active site, and the role of acid/base catalyst was finally assigned to Asp276. The identity of the acid/base catalyst is not conserved throughout the GH29 family, as in many bacterial fucosidases this role is played by a glutamate residue [238], whereas recent bioinformatics studies by Zeuner *et al.* suggest that all animal fucosidases feature a catalytic aspartate for the role of acid/base [236]. From the conformation of the DFJ molecule, we can also hypothesize that the catalytic itinerary of the fucose ring will start from the inverted chair, 1C_4 .

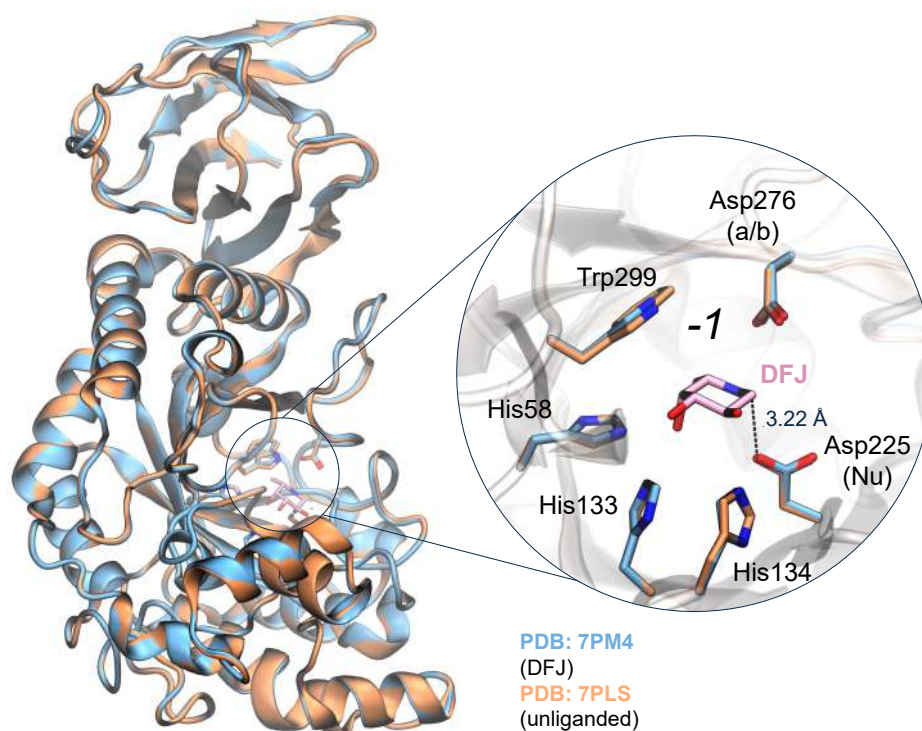


Figure 6.2: Crystal structures of human fucosidase FucA1 solved by Armstrong *et al* [27]. Left, superimposition of the native, unbound enzyme (orange cartoon, PDB ID 7PLS) and the complex with DFJ inhibitor (blue cartoon, PDB ID 7PM4). Right, detail of the active site of FucA1 with the bound DFJ, indicating the main residues responsible for ligand binding at subsite -1.

The hydrolysis of fucosylated glycans by FucA1 has been suggested to follow

a classical, double displacement mechanism with retention of anomeric configuration (Figure 6.3) [27]. A suitably positioned residue, Asp225, takes the role of the nucleophile of the reaction and attacks the anomeric carbon of the fucose moiety, leading to the formation of the glycosyl-enzyme intermediate, and another aspartate, Asp276, protonates the leaving group. In the second catalytic step, a water molecule attacks the anomeric carbon activated by Asp276, now as general base, and the fucose product is released. With an appropriate sugar acceptor, during the deglycosylation step, fucosidases can be engineered to synthesize fucosylated oligosaccharides [239].

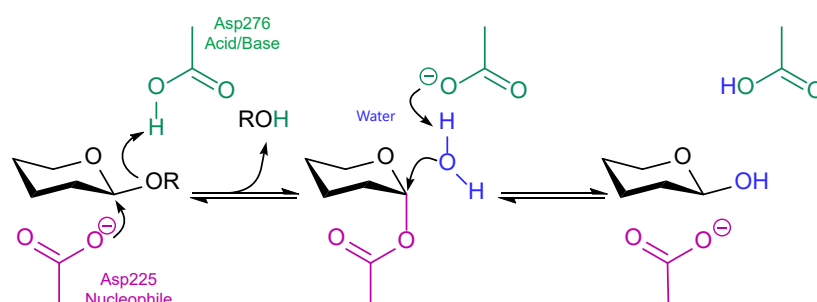


Figure 6.3: Scheme of the classical, double displacement retaining mechanism of GHs, where Asp225 is the nucleophile and Asp276, the general acid/base.

Although the recent structural work provides kinetic data for a natural substrate, namely the disaccharide Fuc- α -1,2-Gal ($k_{cat} = 0.74 \pm 0.02s^{-1}$), the structure of an enzyme-substrate complex was not solved. Additionally, it is known that FucA1 can hydrolyze Fuc- α -1,2-Gal, Fuc- α -1,3-GlcNAc, Fuc- α -1,4-GlcNAc and Fuc- α -1,6-GlcNAc linkages [236], that are present in many structurally different fucosylated glycans. What mechanisms FucA1 uses to discriminate the different fucosylated substrates, however, remains unclear. An early study by Fadda *et al.* addressed the question of enzyme-substrate interactions of FucA1 [244]. Using a homology model for bovine kidney fucosidase (BKF) based in the fucosidase of *Thermotoga maritima* (PDB ID 1ODU) [237] and binding several aminoquinoline-based fluorophore molecules linked to fucose, enzyme-probe models showed that the tight active site cavity cannot accommodate the fluorophore moieties of the substrates, which hampered the enzymatic activity [244]. These molecules however are more rigid than oligosaccharides, whose linkages can rotate. Therefore, we hypothesize that the enzyme has mechanisms to adapt its active site cavity and bind different types of fucosylated glycans.

The aim of this project is to gain insight on human fucosidase FucA1 substrate

specificity and reactivity. Our initial hypothesis is that the enzyme is plastic and has structural mechanisms for adapting to different fucosylated substrates despite having a small active site cavity, with one negative subsite for the fucose moiety, and accommodate them. We also aim to provide a first model for the enzyme-substrate complex of FucA1 with a natural substrate, and to uncover its catalytic mechanism.

In this Thesis chapter we present a molecular dynamics study of the human fucosidase 1 and simulate the glycosylation step of its reaction mechanism with the natural substrate Fuc- α -1,2-Gal. We identify a structural element as responsible for the relaxed substrate specificity of FucA1. We also support the proposal of a classical retaining mechanism with the formation of a covalent glycosyl-enzyme intermediate delineating a conformational itinerary in the Southern hemisphere of the Cremer-Pople sphere. This work is performed in collaboration with Prof. Elisa Fadda (Maynooth University, Ireland, currently at University of Southampton, United Kingdom), as part of a PhD. mobility visit to her group in Maynooth University.

6.2 Results and discussion

6.2.1 Unliganded FucA1

Based on the recent structures of human FucA1 [27] we first built a model for the unbound form of FucA1 and described its dynamics and flexibility with classical MD simulations. The FucA1 structures were solved as homotetramers, each carrying a binding site. Inspection of the 3D structure gave some initial insights on what structural elements might be important for adapting the structure of FucA1 to fucosylated glycans. The enzyme has a narrow active site cavity with one defined subsite to host a Fuc unit, with the catalytic pair Asp225/Asp276 at the entrance. The acid/base, Asp276, is located at 6.98 Å vertically from the nucleophile Asp225, and is part of a large loop spanning from residue Asp250 to Lys287 (named as "catalytic loop" hereafter). The β -factors of this loop are higher than those of surrounding motifs, including active site residues, therefore we hypothesized that the flexibility of this structure could be behind the substrate promiscuity of FucA1 (Figure 6.4, regions closer to blue represent residues with low B-values, therefore less flexible, and closer to red, more flexible regions).

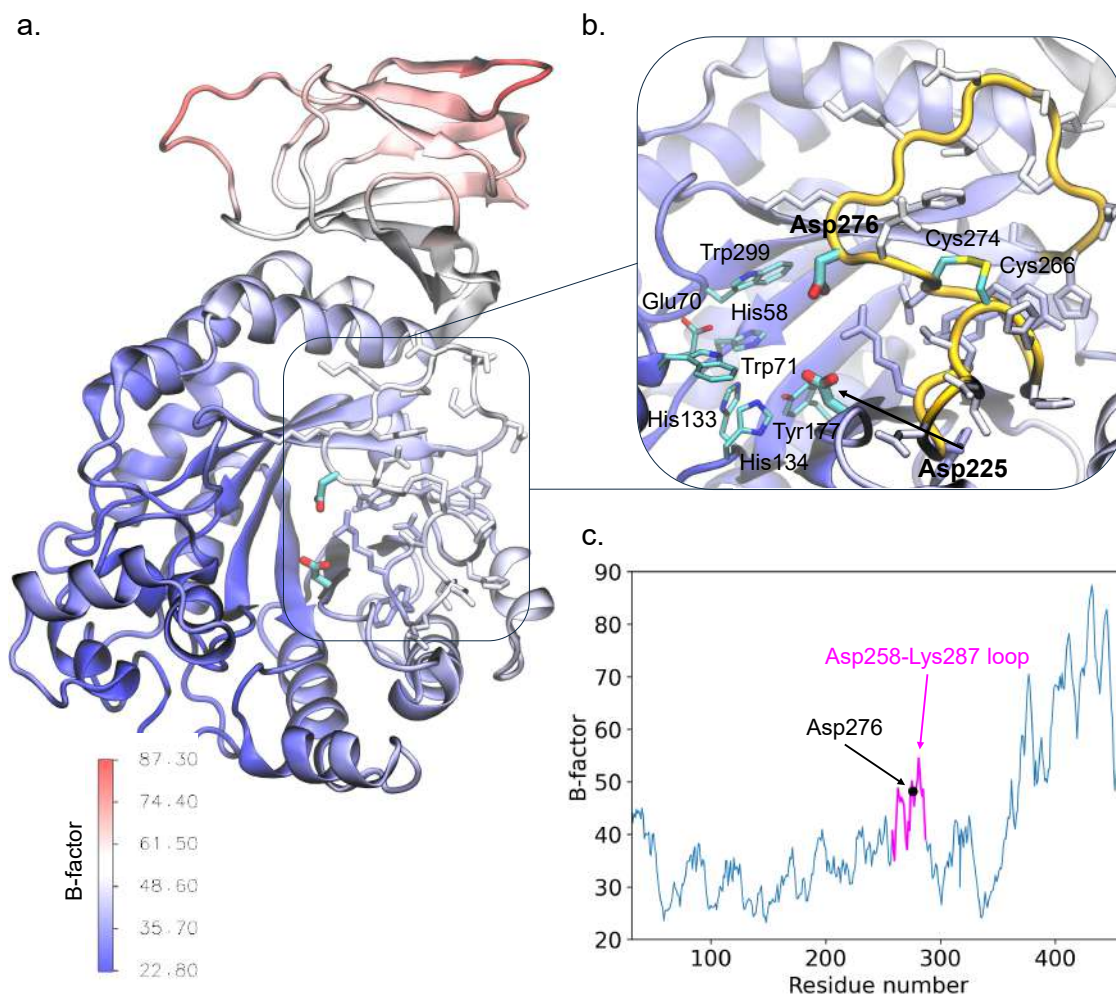


Figure 6.4: a. Cryo-EM structure of unliganded FucA1 colored by B-factor (blue for lower values, red for higher values, see color bar), stick representation: loop residues covering the active site. b. Detailed view of FucA1 active site highlighting the loop that hosts the catalytic Asp276 in yellow. c. B-factor of FucA1 subunit A residues, including the disulfide bond between Cys266-Cys274 within the loop. Highlighted in pink, the loop containing the catalytic acid/base Asp276, with higher B-factor values than the rest of structure, except for the flexible C-terminal β -sandwich domain.

Guided by these structural insights we performed classical MD simulations starting from one subunit of the unliganded structure of FucA1 (PDB ID 7PLS [27]). Since we were interested in analyzing the enzyme dynamics, we performed long timescale MD, specifically 4 μ s of MD simulations (in 4 independent replicas). In two of the replicas (replica 1 and 2), we observed the Asp258-Lys287 loop opening and closing again, recovering the initial state (Figures 6.5 and S6.1), thus, different conformations of this loop are accessible through low free energy barriers, of the order of thermal energy fluctuations $k_b T$ (see Figure S6.4).

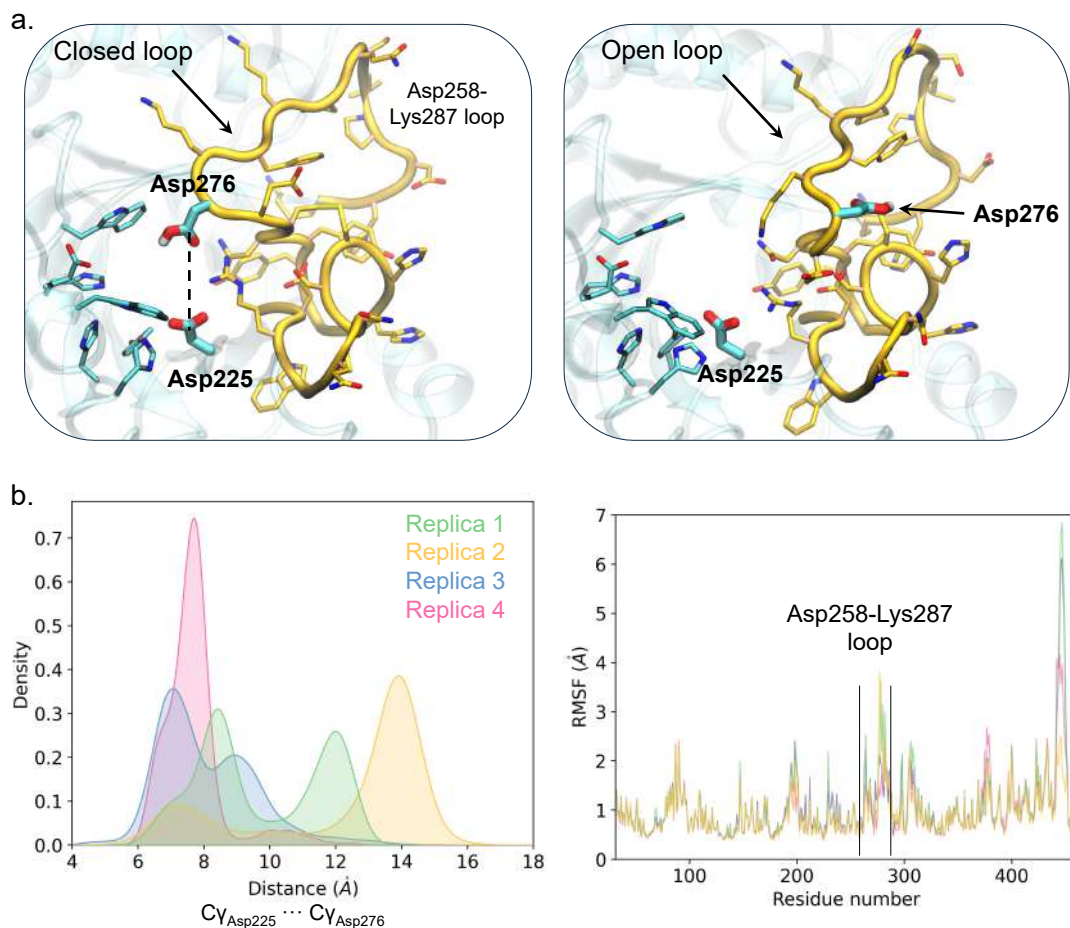


Figure 6.5: a. Trajectory snapshots corresponding to replica 2 showing the two possible conformations of the loop hosting Asp276. As yellow sticks, loop residues. H atoms have been omitted for clarity except the proton on the Asp276 carboxylic group. b. (Left) Distribution of values of the distance between the catalytic Asp225 and Asp276 C γ considering the entire simulation. (Right) RMSF of all residues. Highlighted, the Asp258-Lys287 loop.

With these results in hand we hypothesized that the flexibility of the Asp258-Lys287 loop, together with the fact that it hosts the catalytic acid/base Asp276 may be the key feature in allowing FucA1 to operate on a wide range of structurally different substrates. To this end, we looked at the loop residues to see which ones are more involved in stabilizing the closed conformation. Crucial interactions are the hydrogen bond between Arg259, in the active site, and Asn273 (Figures S6.2 and S6.3). Our simulations support the hypothesis from Armstrong *et al.* [27] that this large flexible region near the active site and containing the catalytic acid/base may be governing substrate recognition.

FucA1 has a narrow active site cavity, with a single subsite (-1), thus only being able to host a fucose unit, yet it has a relaxed substrate specificity for fucosylated glycans [236]. How does the enzyme adapt to different substrates α -linked to fucose?

Are there residues with a key role in substrate recognition and binding or does this mechanism vary from substrate to substrate?

6.2.2 FucA1-glycan complexes

For studying how fucosylated substrates fit in the narrow active site cavity of FucA1 we first built a library of fucosylated glycan epitopes, accounting for all possible fucosyl linkages found in humans, and of length of di- and trisaccharides (Figure 6.6, names of all glycans in Table 6.1) [1,3].

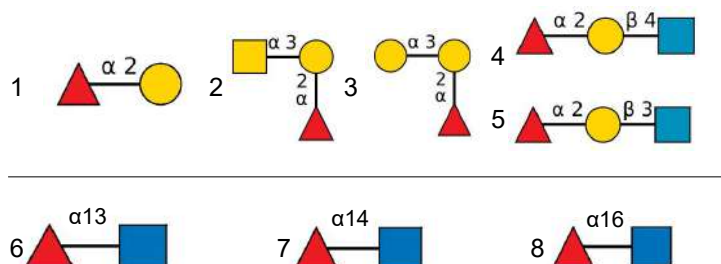


Figure 6.6: Fucosyl-based substrates considered in this work, in the SNFG representation. SNFG symbols drawn with DrawGlycan (www.virtualglycome.org/DrawGlycan/) [5,6].

The names of these substrates can also be found in Table 6.1. For the sake of simplicity, the models we build of FucA1 in complex with the glycans in this table will be referred as **complex 1** to **8** hereafter.

	Substrate	Epitope name
1	Fuc- α -1,2-Gal	H-antigen fragment
2	GalNAc- α -1,3-(Fuc- α -1,2)-Gal	Blood group A
3	Gal- α -1,3-(Fuc- α -1,2)-Gal	Blood group B
4	Fuc- α -1,2-Gal- β -1,4-GlcNAc	H-antigen, type 2
5	Fuc- α -1,2-Gal- β -1,3-GlcNAc	H-antigen, type 1
6	Fuc- α -1,3-GlcNAc	
7	Fuc- α -1,4-GlcNAc	
8	Fuc- α -1,6-GlcNAc	Core fucose of <i>N</i> -glycans

Table 6.1: Structures of the fucosylated motifs found in human glycans simulated in this study. [1]

6.2.2.1 FucA1 with Fuc- α -1,2-Gal (complex 1)

Guided by the available 3D structures of FucA1 in the apo form (PDB ID 7PLS) and in complex with the DFJ inhibitor (PDB ID 7PM4) we first built the enzyme-substrate complex of the H-antigen fragment Fuc- α -1,2-Gal (glycan **1** in Figure 6.6, complex **1** hereafter). To obtain an initial guess of the starting structure we docked the substrate to the enzyme cryo-EM structure corresponding to the enzyme-inhibitor complex (PDB ID 4PM4) (Figure S6.5, for more details on how this complex was built check section 6.4.2.1). We ran 1,5 μ s of classical MD in 3 independent replicas of 500 ns each (Figure S6.6.a). Our simulations show that, during this time, the substrate strongly binds to the enzyme (Figure 6.7.a and S6.7) and that residues Asp225 and Asp276 remain in the same orientation as in the experimental structure, which we suspect is the catalytically competent (Figure 6.7.b). During these trajectories the catalytic loop remained closed, with Asp276 almost always interacting with the glycosidic bond oxygen (Figure S6.6.b), although still retained some degree of mobility (Replica 1, Figure S6.6.b). This is reflected in the RMSF of the loop, which display lower values in these simulations relative to the apo form (Figure 6.7.c).

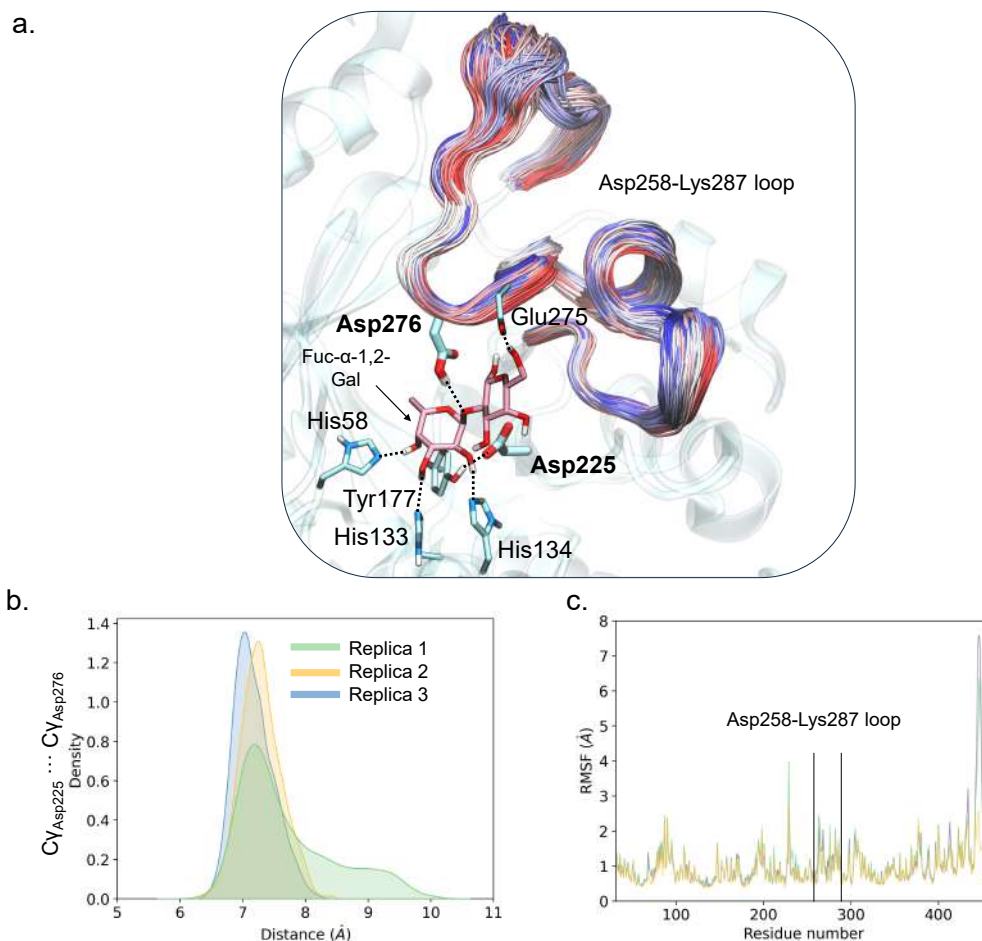


Figure 6.7: a. Classical MD trajectory snapshot (replica 2) showing the ligand **1** (Fuc- α -1,2-Gal) in the active site of FucA1. The Asp258-Lys287 loop is depicted every 0.2 ns over the trajectory. Coloring indicates simulation time, from blue (0 ns) to red (500 ns). b. Density distribution of the distance between the catalytic residues over the classical MD replicas. c. RMSF of all residues, highlighting the loop.

To get insight about important residues for substrate recognition, we analyzed substrate-enzyme and intra-enzyme hydrogen bonding network along the trajectories. The interactions between the fucose moiety and residues at the -1 subsite His58, His133 and His134 (Figure S6.7) are the most persistent during the simulations, together with Asp276 and the oxygen of the glycosidic bond. Besides these interactions, the substrate engaged in hydrogen bonds with other residues. Glu275, in the catalytic loop, establishes hydrogen bonds with either of its 2 oxygen atoms with the Gal moiety 1-OH and 6-OH groups (Figure 6.8.a). Another relevant loop residue is Arg259, which, although it forms short-lived hydrogen bonds with the Gal moiety, it interacts with both catalytic residues. These results suggest that Arg259 has a role in assisting the catalytic pair for catalysis by modulating its microenvironment. The bulky presence of the Arg sidechain also blocks the space for oligosaccharides branched in that direction. This position is conserved in other

members of subfamily GH29A [236]. Asn273, also in the catalytic loop, interacts with Arg259. These positions govern the loop dynamics and thus, its plasticity. We suggest that modifying these positions might result in altered loop dynamics, which might impact substrate recognition. Finally Tyr177 interacts with the nucleophile, but also with the Fuc 4-OH substituent.

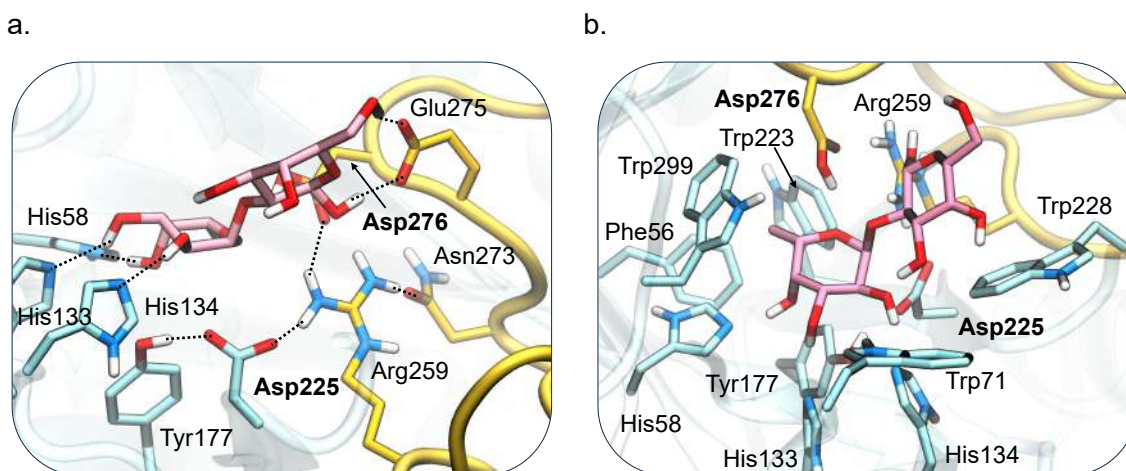


Figure 6.8: a. Detail of the main hydrogen bond interactions of complex **1**. In orange, residues belonging to the catalytic loop. In blue sticks, other enzyme residues. This snapshot corresponds to R2. The other replicas behave similarly. b. Residues that contribute the most in ligand-enzyme binding free energy according to MMPBSA analysis.

To obtain more information on the main residues involved in the substrate-enzyme binding energy we performed MMPBSA calculations to identify the key binding residues (Figure 6.8.b and S6.8). Some of the residues with the highest contribution to the binding free energy form hydrogen bonds with the substrate (His58, His133, His134) or establish hydrophobic interactions with the pyranose rings and the methyl group of fucose (Trp71, Trp299, Trp223, Phe56). Tyr177 can interact with the substrate, but also with the catalytic base, suggesting a fundamental role in stabilizing the substrate (Figure S6.7.b). Residue Trp228 interacts with the Gal moiety and thus, this residue may contribute to the recognition of the saccharide at the $+1$. Some of the residues most relevant for substrate binding are part of the catalytic loop (Arg259, Glu275), which reinforce the role of this flexible region in conferring the enzyme a unique ability to recognize structurally different substrates. Guided by these findings, we proceeded to build alternative FucA1-substrate complexes considering different oligosaccharide lengths and linkage types (Figure 6.6).

6.2.2.2 Fuc- α -1,2-Gal trisaccharides: FucA1 with substrates **2**, **3**, **4** and **5**

Complexes **2** and **3**: Blood groups A and B

We first built FucA1 in complex with the blood groups A (GalNAc- α -1,3-[Fuc- α -1,2]-Gal, complex **2**) and B (Gal- α -1,3-[Fuc- α -1,2]-Gal, complex **3**) (Table 6.1). With these complexes we explored how the enzyme would respond to the addition of a third monosaccharide to the existing α -1,2-linked Gal. For this purpose, we selected one suitable trajectory snapshot of the classical MD simulations of complex **1** (section 6.2.2.1) where the complex was stable, and added the α -1,3-linked GalNAc or Gal moiety, respectively. Then, after running 500 ns of constrained classical MD where we fixed the atomic positions of the saccharide to let the enzyme adapt around it, we performed 500 ns more of unbiased classical MD.

Our simulations showed that both trisaccharides are able to bind in the tight active site cavity of FucA1 (Figure S6.9). The catalytic loop is flexible enough to move and accommodate the larger substrate (Figure 6.9.a,c), while staying closed during the span of the simulation, reflected in the interatomic distance between C $_{\gamma}$ of the catalytic pair sidechains (Figure 6.9.b). However, we detected some differences in the enzyme-substrates hydrogen bond patterns with respect to the smaller disaccharide substrate of section 6.2.2.1.

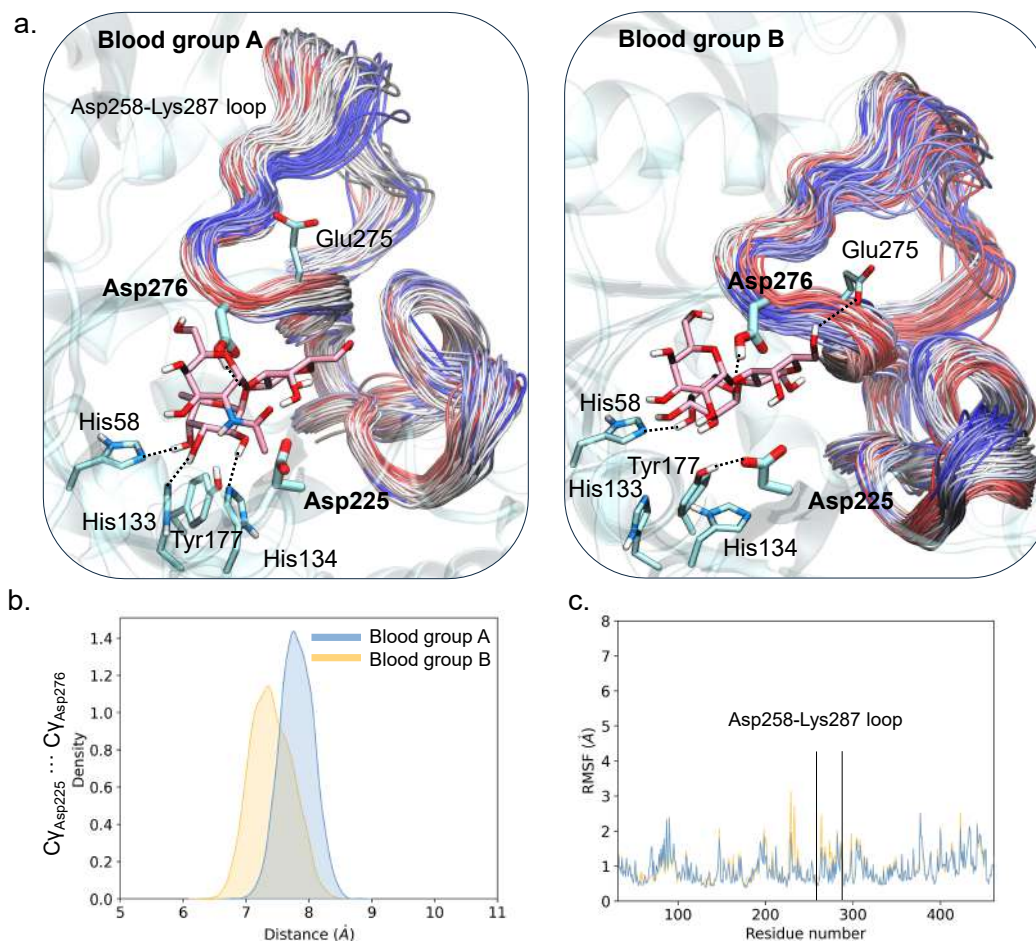


Figure 6.9: a. Classical MD trajectory snapshots of complex **2** (left) and **3** (right). Asp258-Lys287 loop is depicted every 0.4 ns. Coloring indicates simulation time, from blue (0 ns) to red (1 μ s). b. Density distribution of the distance between the catalytic residues during the 500 ns of free MD. c. RMSF of all residues, highlighting the loop.

By analyzing the hydrogen bond interactions during the unrestrained part of the simulation we observed that some enzyme residues hold persistent interactions with the substrate *+2* sugar (complex **2**), like Glu70, Trp299 and Glu227 (Figure 6.10). Residues Trp299 and Glu227 also belong to flexible loop elements (Ile294 to Ala306 and Gly226 to Pro 231, shorter in number of residues than the Asp258-Lys287 loop) close to the active site cavity, which shows the plasticity of the enzyme to adapt to a longer substrate. In this case, however, interactions with the loop Glu275 were much less sampled. As in complex **1**, catalytic loop residues Arg259 and Asn273 are key for loop plasticity and the integrity of the active site. These hydrogen bonds are also present in the complex with the blood group B antigen (complex **3**, Figure 6.10.b).

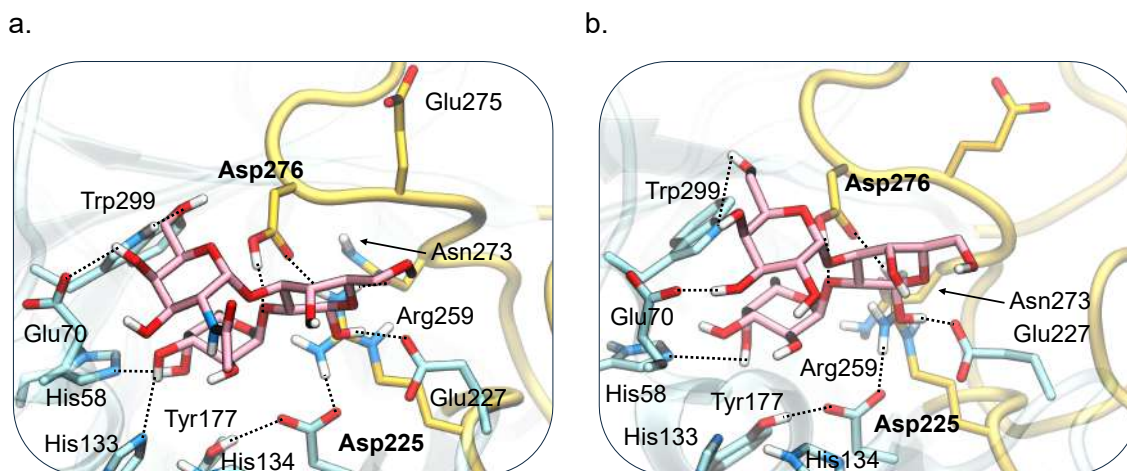


Figure 6.10: Detail of the main hydrogen bonding interactions of a. complex **2** and b. complex **3**. In orange, residues belonging to the catalytic loop. In blue sticks, other enzyme residues.

Complexes 4 and 5: H-type antigens 2 and 1

To account for different types of Fuc- α -1,2-Gal trisaccharides, we also built models of FucA1 in complex with two other common α -1,2 fucosylated trisaccharide epitopes, the H-antigens type 2 (Fuc- α -1,2-Gal- β -1,4-GlcNAc, complex **4**) and type 1 (Fuc- α -1,2-Gal- β -1,3-GlcNAc, complex **5**) (Table 6.1) [3]. After 800 ns of classical MD, we observed that the enzyme can also adapt to these substrates by fine tuning the motions of its flexible loop around the active site (Figure 6.11). H-antigen 1, was stable on the active site for approximately 500 ns, therefore the analyzed results focus on that part of the simulation (Figure S6.10).

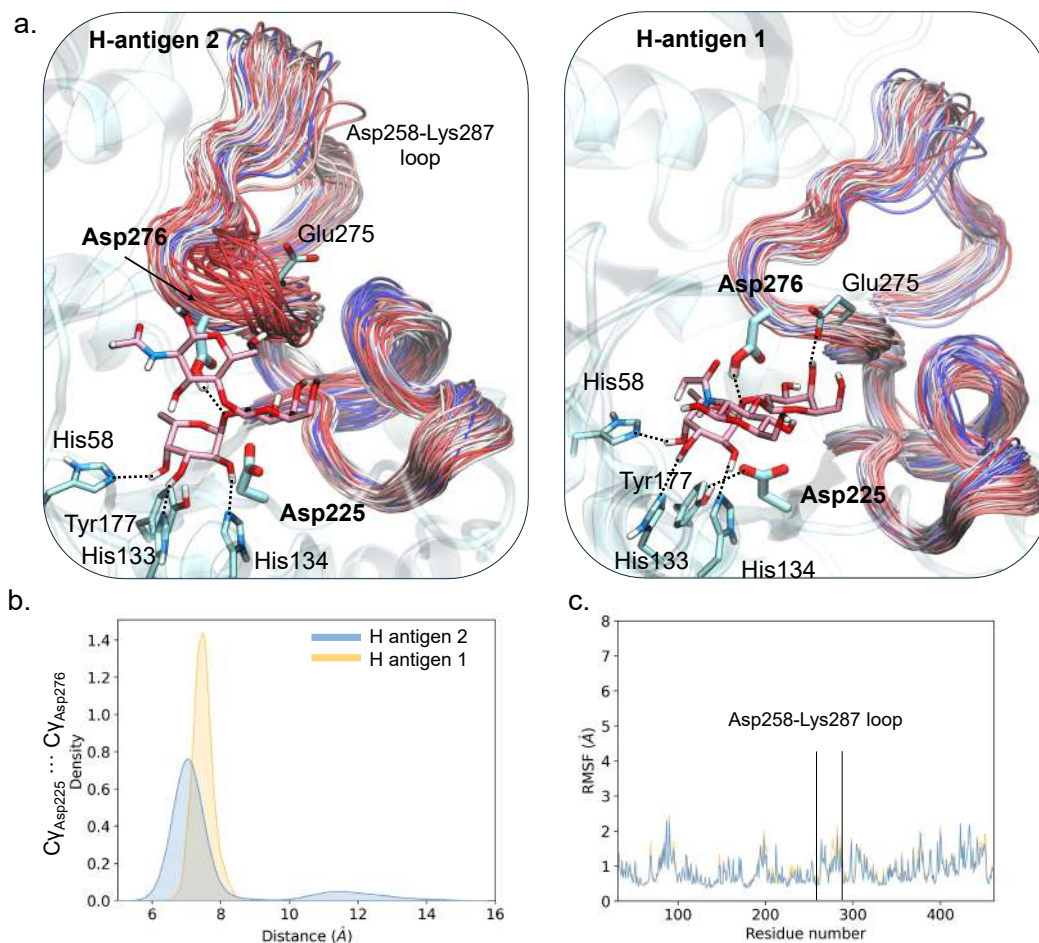


Figure 6.11: a. Classical MD trajectory snapshots of complex 4 (left) and 5 (right). Asp258-Lys287 loop is depicted every 2 ns. Coloring indicates simulation time, from blue (0 ns) to red (end of the simulation). b. Density distribution of the distance between the catalytic residues during the 500 ns of free MD. c. RMSF of all residues, highlighting the loop.

FucA1 is also able to accommodate the two GlcNAc-containing trisaccharides by establishing hydrogen bond interactions with the β -1,4 and β -1,3-linked saccharides (Figure 6.11). Although fairly solvent-exposed, some enzyme residues provide additional binding points. Two tryptophan residues interact with GlcNAc: Trp71 forms stacking interactions with the β -1,3-GlcNAc moiety in the complex with H-antigen 1 (Figure 6.12, right), whereas in complex with the H-antigen 2 this residue forms a hydrogen bond with fucose at subsite -1. Trp299 interacts with the exocyclic hydroxyl substituents of GlcNAc. Altogether, this residue is part of a loop from Asp295 to Glu312. This shows how the enzyme resorts to flexible loops near the active site (Asp25-Lys287, Asp295-Glu312) to adapt to different substrates.

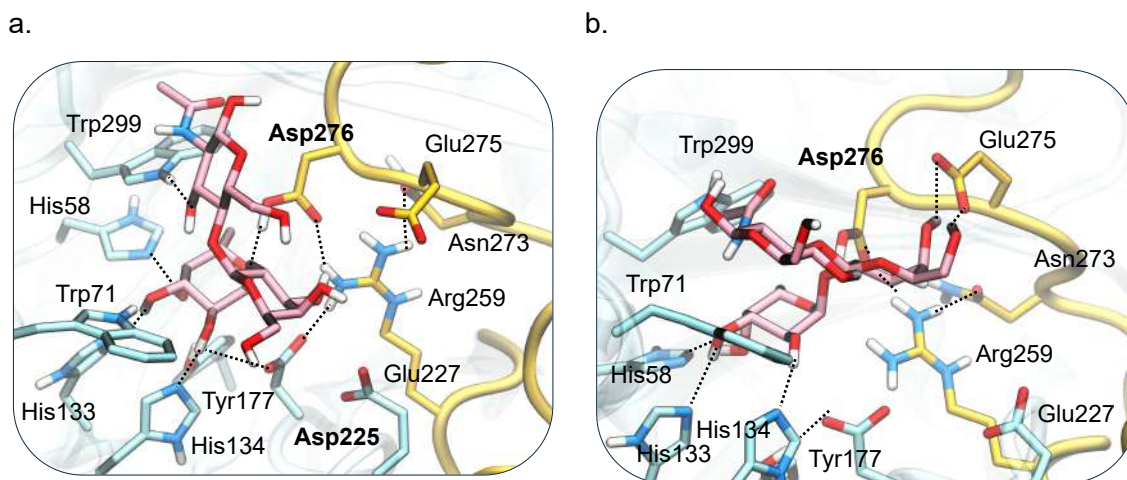


Figure 6.12: Detail of the main H-bond interactions of a. complex **4** and b. complex **5**. In orange, residues belonging to the catalytic loop. In blue sticks, other enzyme residues.

Our simulations gave insight on the mechanisms that govern the plasticity of the FucA1 binding site, and show examples of how the enzyme can accommodate α -1,2-linked fucosylated substrates which are characterized experimentally [236]. From tight active site interactions, to the motion of large structural elements, the active site of FucA1 provides several binding points that can adapt to structurally different substrates. Also, the loop spanning from residue Asp258 to Lys287 is mobile enough to adapt to trisaccharides containing the Fuc- α -1,2 moiety. These simulations provide an atomistic point of view for the reported relaxed substrate specificity of this enzyme, towards Fuc- α -1,2-Gal linked glycans [27, 236].

6.2.2.3 Other linkage types: α -1,3-GlcNAc, α -1,4-GlcNAc and α -1,6-GlcNAc (complexes **6**, **7** and **8**)

After analyzing thoroughly the plasticity of FucA1 with Fuc- α -1,2-Gal linked substrates, and observing how it combines the motions of a large structure region (the catalytic loop) with small changes in the hydrogen bond partners in the active site to accommodate different glycans, we moved to test how the enzyme would respond to the other types of fucosyl linkages commonly found in mammalian glycans, namely Fuc- α -1,3-GlcNAc, Fuc- α -1,4-GlcNAc and Fuc- α -GlcNAc [3]. The first two disaccharides are part of important antigenic epitopes, including the Lewis^X and Lewis^A sugars, respectively [5]. α -1,6-linked fucose is the core fucosylation motif of *N*-glycans, also known as "core" fucose [3]. For these complexes we applied the approach described in section 6.2.2.2. Applying a similar procedure, after running 1 μ s, 800 ns (the time the substrate remained stable in the active site) and 1 μ s of classical MD (Figures S6.12 and S6.13), respectively we observed that, the enzyme

adapts to the saccharide linked to fucose (3,4 and 6-linked GlcNAc). This agrees with experimental data showing fucosidase activity for these substrates [234, 236] (Figure 6.13).

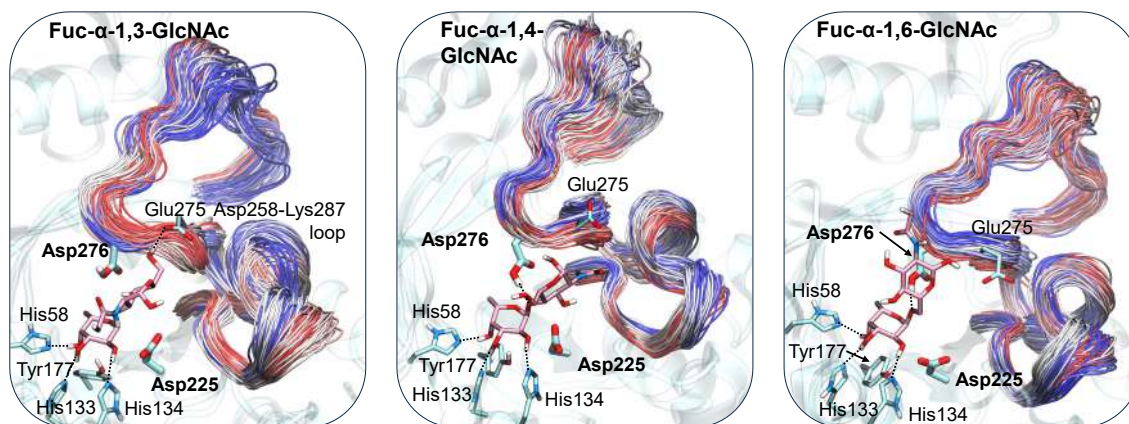


Figure 6.13: Classical MD trajectory snapshots of complexes **6** (left), **7** (center) and **8** (right). The loop containing the catalytic acid/base is depicted every 2 ns, colored from blue (0 ns) to red (end of the simulation).

Our simulations showed that FucA1 is able to accommodate fucose linked to GlcNAc also by fine motions of the catalytic loop around GlcNAc, which is bulkier than Gal (Figure 6.13). During the simulations, the catalytic loop stayed mostly closed, as evidenced by the distance between the catalytic pair (Figure 6.14) (except for some population of slightly open loop during MD simulations of complex **6** (Figure 6.14.b).

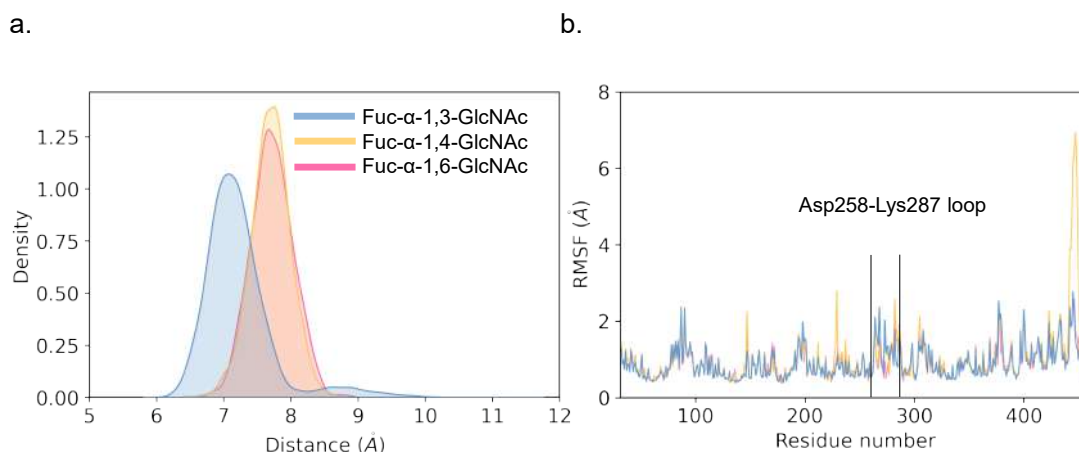


Figure 6.14: a. Density distribution of the distance between the $C\gamma$ of the catalytic Asp225 and Asp276. c. Computed RMSF of all protein residues, highlighting the loop region.

A detailed view on hydrogen bond interaction patterns shows that the residues that were involved in stabilizing the different Fuc- α -1,2-Gal complexes also contribute stabilizing the linked GlcNAc (Figure 6.15). Some residues, however, have a different role in substrate binding than in the complexes commented previously: Trp299, rather than forming a hydrogen bond with the substrate (see Figure 6.12.a), provides a hydrophobic cavity with the α -1,3 and 4 linked GlcNAc moieties. In the case of Fuc- α -GlcNAc, the α -1,6 glycosidic bond is more flexible than the corresponding α -1,2/3/4 linkages due to the rotation of the GlcNAc CH₂ group. The GlcNAc moiety interacts mostly with Glu275, and does not form hydrophobic interactions with Trp299.

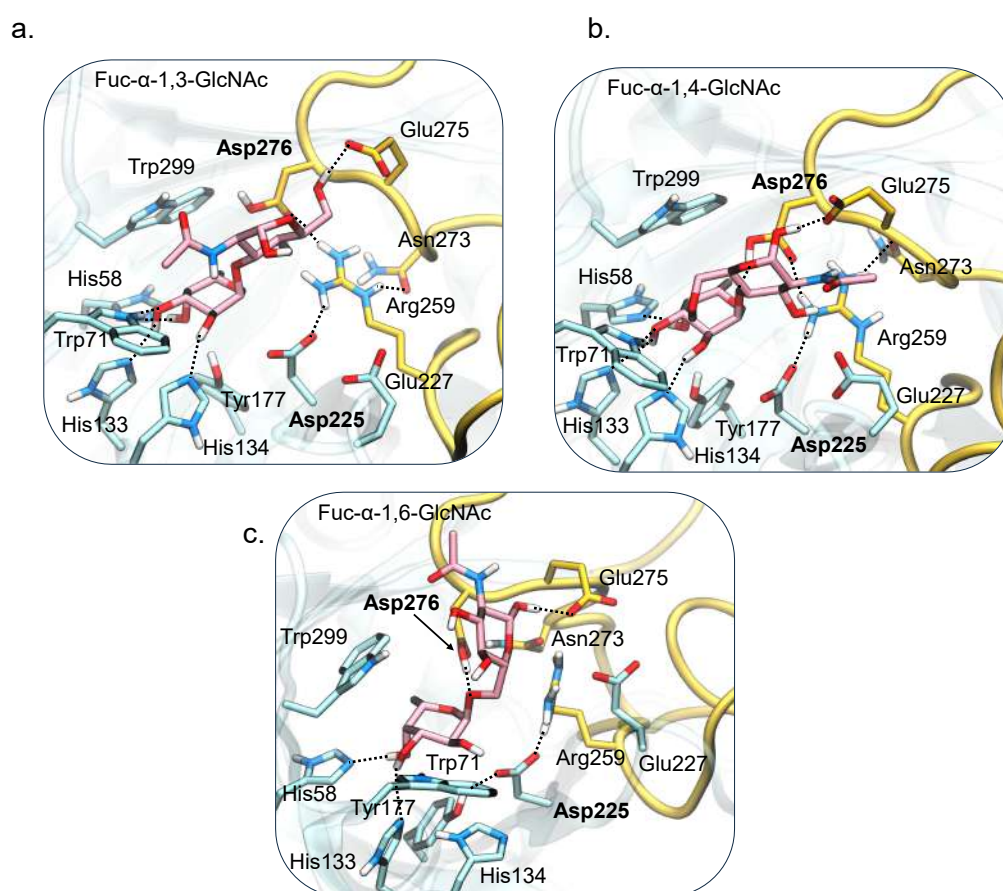


Figure 6.15: Detail of the main H-bond interactions of a. complex **6**, b. complex **7** and c. complex **8**. In orange, residues belonging to the catalytic loop. In blue sticks, other enzyme residues.

In summary, our classical MD simulations provided models of the human fucosidase FucA1 in complex with different naturally occurring fucosylated substrates, whose structure is not yet available experimentally. The results could explain the relaxed substrate specificity FucA1 [236] in terms of specific substrate-enzyme in-

teractions. Our results expand previous work from Fadda *et al.*, in which it was observed that different aminoquinoline-based fucosylated probes were difficult to accommodate in the tight active site of FucA1, which was attributed to the low flexibility of the aminoquinoline-based aglycone moiety of the probes. Here, we report bound structures with both disaccharides and trisaccharides, which are intrinsically more flexible, and provide a molecular point of view on the structural mechanisms of the enzyme to be able to accommodate different substrates. First, we report a flexible large loop ranging from Asp258 to Lys287 that is, not only near the active site, but also hosts the general acid/base residue Asp276. Although having one of the catalytic residues might seem counter productive for catalysis, as loops have increased flexibility with respect to the rest of the protein, reflected in the experimental B-factors [27], it is one of these occasions where it can be "a feature rather than a bug". The mobility of this loop, which is accessible at room temperature (300 K) because of the low free energy barriers, appears to be the main mechanism responsible of the fucosylated substrates promiscuity of FucA1. Besides being the structural element that hosts the catalytic acid/base, the Asp 258-Lys 287 loop also contains residues that are important for both catalysis and binding. Arg259 interacts with the catalytic Asp225 and Asp276, and is stabilized via interactions with Asn273, also in the loop. Furthermore, the hydrogen bond interaction network adapts to the structure of the substrate that binds the active site cavity, through residues from the catalytic loop (Glu275, Asp276), as well as from other flexible regions near the active site (Glu70, Trp71, Glu227, Trp228, Trp299) (see Figures S6.11 and S6.14). These results expand current knowledge about substrate recognition in subfamily A fucosidases, and give new directions into what positions could be mutated to develop enzyme engineering strategies.

6.2.3 The glycosylation reaction of FucA1 with Fuc- α -1,2-Gal

In order to study the reactivity of FucA1 with a natural substrate we turned to QM/MM MD and metadynamics. We built a QM/MM model of FucA1 in complex with the blood group H-antigen fragment (Fuc- α -1,2-Gal, Figure 6.6). To note, there is no structural information on the FucA1 Michaelis complex, while the experimental catalytic rate constant is available in the work of Armstrong *et al.* for pNP- α -L-Fuc ($k_{cat}=9.96 \pm 0.09 \text{ s}^{-1}$) and the H-antigen fragment disaccharide Fuc- α -1,2-Gal ($k_{cat}=0.74 \pm 0.02 \text{ s}^{-1}$) [27]. Starting from a trajectory snapshot of a classical MD simulation of complex **1** we applied QM/MM MD methods defining a QM region that included the sidechains of the catalytic pair Asp225/Asp276 and the disaccharide

(Figure S6.15). We selected two collective variables (CVs) to model the glycosylation reaction of FucA1, which combine the interatomic distances between the atoms involved in the reactive process. CV1 is the difference between the $C1-C_{\gamma_{Asp225}}$, and $C1-O1$ distances, accounting for the nucleophilic attack. By including the Asp225 C_{γ} we did not specify which of the carboxylate oxygens the reaction pathway should evolve through. CV2 was defined as the difference of $O1-H_{Asp276}$ and $H-O_{Asp276}$, accounting for the proton donation to the scissile bond oxygen O1 (Figure S6.16.a).

The reconstructed reaction FEL has two basins corresponding to the reactants state (Michaelis complex, MC) and glycosyl-enzyme covalent intermediate (GEI), connected via one transition state. The computed free energy barrier is $14.07 \text{ kcal mol}^{-1}$, which is in reasonable agreement with the free energy barrier estimated from the experimental rate constant ($k_{cat} = 0.74 \pm 0.02 \text{ s}^{-1}$, which amounts to $17.74 \text{ kcal mol}^{-1}$) (Figure 6.16.a).

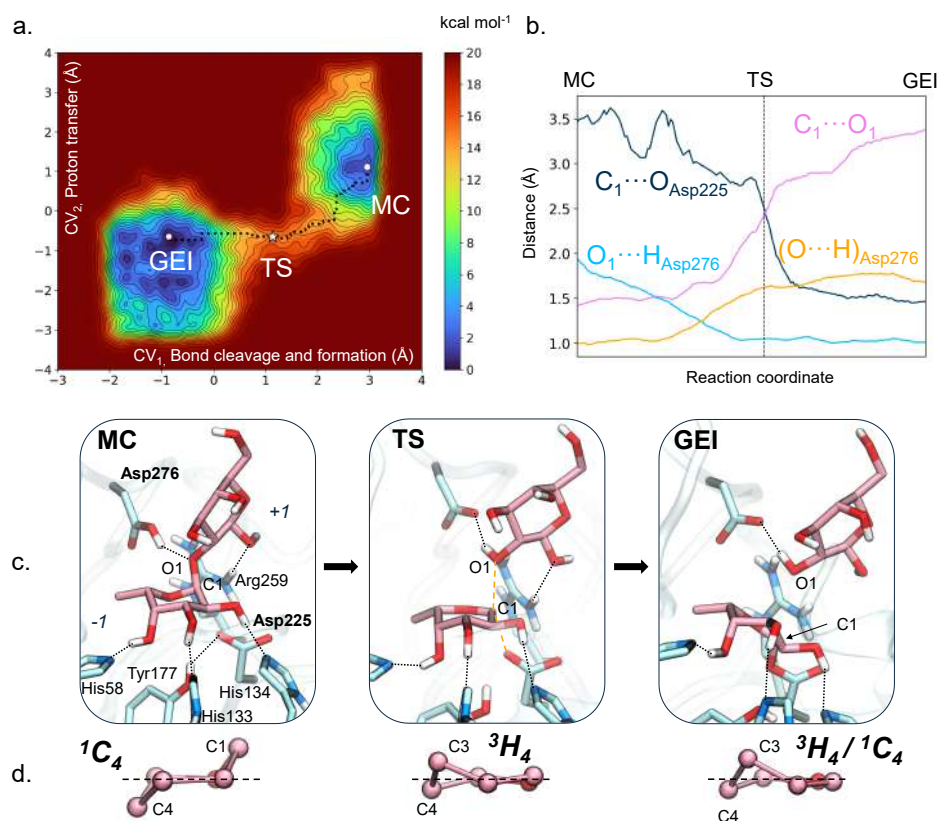


Figure 6.16: a. Reaction FEL of glycosylation by FucA1. Isolines every $1.0 \text{ kcal mol}^{-1}$. b. Distance evolution between atoms involved in the reaction along the MEP. c. Detailed representation of the FucA1 active site at the main reaction states (MC, TS and GEI). Black, dashed lines correspond to H-bonds. Orange dashed lines at TS indicate bond cleavage and formation. d. Ball-and-stick representation of the L-Fuc ring conformational itinerary, indicating the exoplanar atoms at every state.

The analysis at atomic detail of the main interatomic distances along the reaction coordinate (Figure 6.16.b and Table 6.2) reveals that the reaction starts as Asp225 approaches the anomeric carbon of Fuc (C1) and Asp276 tightens its hydrogen bond interaction with the scissile bond O1 (at MC, $C1-O_{Asp225} = 3.43 \pm 0.45$ Å and $O1-H_{Asp276} = 2.12 \pm 0.13$ Å) (Figure 6.16.c). At the same time, the glycosidic bond distance starts increasing (C1-O1: 1.44 ± 0.05 Å). At the transition state (TS) the glycosidic bond is already broken (C1-O1 at TS is 2.38 ± 0.16 Å) and the new covalent bond between Asp225 and the anomeric carbon of Fuc is not yet formed ($C1-O_{Asp225} = 2.60 \pm 0.06$), indicating the dissociative nature of the reaction mechanism. The fucose moiety adopts a 3H_4 conformation at the transition state. This is one of the possible conformations of an oxocarbenium ion-like TS [41], in which the pyranose ring atoms O5, C1, C2 and C5 lay on the same plane. The shortening of the C1-O5 distance reflects the partial double bond character between these atoms at the TS (at MC is 1.42 ± 0.03 Å, whereas at TS is 1.29 ± 0.02). When the reaction intermediate is formed (GEI), the anomeric bond distance with the endocyclic O increases again (C1-O5 1.40 ± 0.03 Å). The protonation of the departing Gal moiety takes place before reaching the TS (Figure 6.16.b), where the proton is completely transferred from Asp276 to O1 (TS panel in Figure 6.16.c) (at TS, $O1-H_{Asp276} = 1.04 \pm 0.07$ Å, whereas $H_{Asp276}-O_{Asp276} = 1.76 \pm 0.23$ Å). Other interactions keep the Fuc moiety tightly bound at subsite -1 while the reaction takes place, such as three hydrogen bonds with active site His58, His133 and His134 (Table 6.2 and Figure S6.17.a).

Distance (Å)	State		
	MC (1C_4)	TS (3H_4)	GEI (${}^3H_4/{}^1C_4$)
$C1-O_{Asp225}$	3.43 ± 0.45	2.59 ± 0.08	1.48 ± 0.06
C1-O1	1.44 ± 0.05	2.42 ± 0.18	3.46 ± 0.21
$O1-H_{Asp276}$	2.12 ± 0.13	1.06 ± 0.07	1.01 ± 0.05
C1-O5	1.42 ± 0.03	1.29 ± 0.02	1.40 ± 0.03
$H_{Asp276}-O_{Asp276}$	1.00 ± 0.05	1.84 ± 0.26	1.86 ± 0.17
$H_{2-OHFuc}-N_{\epsilon His58}$	2.86 ± 0.81	2.09 ± 0.20	1.95 ± 0.29
$H_{3-OHFuc}-N_{\epsilon His133}$	2.45 ± 0.53	1.93 ± 0.04	2.16 ± 0.31
$H_{4-OHFuc}-N_{\epsilon His134}$	1.81 ± 0.12	1.78 ± 0.05	1.81 ± 0.15

Table 6.2: Main FucA1 active site distances at the relevant states of the glycosylation reaction (MC, TS and GEI). Distances averaged from an ensemble of trajectory frames of each of the corresponding areas in the FEL.

The computed conformational itinerary of the L-Fuc moiety is ${}^1C_4 \rightarrow [{}^3H_4]^\ddagger \rightarrow {}^3H_4/{}^1C_4$ (Figures 6.16.d and S6.16.b). The conformational preferences of fucose

have been elucidated by means of QM/MM MD methods, locating it across the Southern hemisphere of the Cremer-Pople sphere [221]. In contrast with the 3,6-anhydro-D-galactosyl (Chapter 5 of this Thesis) where the conformational landscape was restricted by the 3,6-anhydro bridge in D-Gal. L-fucose itineraries take place in the Southern hemisphere as expected for L-sugars [221]. The reaction starts from the inverted chair conformation 1C_4 , therefore without the typical distortion widely reported in many GH families, such as the GH20, GH129 and GH130 described in this Thesis. The substrate binds FucA1 in 1C_4 conformation at the Michaelis complex, as this conformation provides the necessary geometry for in-line nucleophilic attack [221].

In summary, our QM/MM metadynamics simulations confirm the catalytic roles assigned to Asp225 and Asp276, and confirm the classical glycosylation step for FucA1, and propose a first model for an enzyme-substrate complex of the human fucosidase 1. They also shed light on how the mobile loop of FucA1 is positioned for catalysis, thus opening the door to engineering strategies for developing variants for synthesizing fucosylated oligosaccharides and the design of mechanism-based inhibitors and probes. Further work is ongoing to model the deglycosylation step and elucidate which is the rate-limiting step in the FucA1 catalytic mechanism.

6.3 Summary and conclusions

In this Thesis chapter we presented and discussed the results of a combined classical MD and hybrid QM/MM study of the recognition, binding and catalytic mechanism of human FucA1. Starting from the native, unliganded 3D structure of FucA1, we modeled the enzyme with classical MD. We started with the hypothesis that, although the active site of FucA1 is small, with one subsite -1 for hosting the fucose moiety of the substrates, a flexible structural element near the active site could be responsible for the known relaxed substrate specificity. Our simulations of the unliganded form of the enzyme confirmed that the loop region ranging from residue Asp258 to Lys287, including part of the catalytic machinery (acid/base Asp276), is able to open and close, thus probably able to adapt around different substrates. Based on this insight, we used again classical MD to build and analyze models of FucA1 in complex with natural substrates, which have not yet been resolved experimentally to date, to our knowledge. These models showed what binding site regions and residues are involved in substrate recognition, both through direct hydrophobic and hydrogen bond interactions between residues near the active site cavity and the substrate, and by the motion of large structural elements (the Asp258-Lys287 loop). Our simulations provide a molecular point of view to rationalize the previ-

ous experimental results reporting the promiscuity of FucA1 for naturally occurring fucosylated substrates. Our QM/MM metadynamics simulations give insight on the catalytically competent configuration in the active site of FucA1 with the natural substrate Fuc- α -1,2-Gal, confirming that the mobile Asp258-Lys287 loop stays closed for catalysis, providing the right orientation of the catalytic acid/base, pointing towards the scissile bond. Our simulations provided atomistic details of the glycosylation step of the catalytic mechanism, and provide a first putative structure of human FucA1 glycosyl-enzyme intermediate. Further work is ongoing towards simulating the full reaction mechanism, which will give insight on the rate limiting step, and to build complexes with Fuc- α -1,3/4/6-GlcNAc trisaccharides, which will expand the knowledge in the plasticity of this enzyme to accommodate different substrates.

This study reveals, with a combination of classical MD and QM/MM MD metadynamics techniques, the structural features of the human fucosidase 1, FucA1, for binding diverse substrates, and the first reaction step of the degradative mechanism of fucosylated glycans. The results give insights on important enzyme residues for substrate recognition, which adds to the understanding of GH29 fucosidases and opens doors for engineering strategies for the enzymatic synthesis of fucosylated oligosaccharides and the development of specific inhibitors and probes for fucosidases.

The conclusions of this Chapter are as follows:

- The region of FucA1 ranging from Asp258 to Lys287, containing the catalytic acid/base Asp276, is highly flexible and is able to adopt open and closed conformations at room temperature. This dynamic feature is responsible for the relaxed substrate recognition by FucA1.
- FucA1 can bind Fuc- α -1,2-Gal, Fuc- α -1,3/4/6-GlcNAc linked substrates by adapting the flexible regions of its active site. Residues Glu275, Trp299, Trp71 and Arg259 are some of the key positions for proper substrate binding.
- FucA1 performs a classical, retaining mechanism with the formation of a fucosyl-enzyme intermediate to degrade the naturally occurring substrate Fuc- α -1,2-Gal. The Fuc moiety delineates a ${}^1C_4 \rightarrow [{}^3H_4]^{\ddagger} \rightarrow {}^3H_4/{}^1C_4$ catalytic itinerary, featuring a dissociative TS. Further work is on the way to refine the transition step, elucidate the molecular details of the deglycosylation reaction and determine the rate limiting step.

6.4 Computational details

In this study we built several models involving FucA1 in the apo form and in complex with different fucosylated glycans, among which the H-antigen fragment Fuc- α -1,2-Gal complex was used to model the first displacement step of the hydrolysis reaction mechanism. All trajectory analyses were performed with VMD version 1.9.3 (<http://www.ks.uiuc.edu/Research/vmd/>) [130], the CPPTRAJ [131] module of AMBER21 [133] and the open-source, community-developed PLUMED library [149, 150].

6.4.1 Model of the unliganded FucA1

6.4.1.1 System preparation

We built the model starting from the cryo-EM structure of native FucA1 in the apo form (PDB ID 7PLS at www.rcsb.org, resolution 2.49 Å [27]), considering the first protomer, from residue 31 to 461. Crystallographic waters were included in the model. Protonation states of protein were assigned with the assistance of PropKa3 [135] and H++ [136], as well as visual inspection of residues environment considering pH 5.0. The protonation state of the catalytic residue was selected as follows: we assigned the carboxylate form to Asp225 (nucleophile) and carboxylic acid to Asp276 (general acid/base). The system including one monomer of FucA1 and the crystallographic waters was solvated in a cell of $117.21 \times 92.77 \times 84.51$ Å with 23391 water molecules, at a distance of 15.0 Å of the protein surface at each direction, and 69 Na⁺ cations and 61 Cl⁻ anions to achieve net neutrality of the system and a concentration of 150 mM. For estimating the number of ions, we used the SLTCAP (Screening Layer Tally by Container Average Potential) tool (web server: <https://www.phys.ksu.edu/personal/schmit/SLTCAP/SLTCAP.html>) [245].

6.4.1.2 Classical Molecular Dynamics

We performed classical molecular dynamics (MD) simulations of the native FucA1 using the AMBER package, version 20 [133]. We used the forcefields FF19SB (protein) [65] and TIP3P (solvent) [141]. Classical MD simulations were performed according to the following protocol: first an energy minimization of the solvent box was performed (first with steepest descent then conjugated gradients schemes), followed by a global minimization, both with 20000 maximum number of cycles. Next, the system was heated to a temperature of 300 K by increasing 100 K over consecutive runs of 500 ps with a timestep of 1 fs (solvent only to 100 K first, then entire system to 100 K, 200 K and 300 K). Then a 200 ps run in the NPT ensemble was

performed to equilibrate the system density to 1 g cm^{-3} at 1 bar with the Berendsen barostat [142]. The simulation was then continued in the NVT ensemble for 200 ps. After this equilibration we performed 4 independent replicas of $1 \mu\text{s}$ each in the NPT ensemble (Figure S6.1). Runs in both NPT and NVT ensemble used a timestep of 2 fs, applying the SHAKE algorithm [143] to constrain the bonds with hydrogen atoms.

6.4.2 Building the FucA1-glycan complexes

The subsequent sections detail how the FucA1-glycan complexes were built. In all cases we started from the disaccharides covering all possible fucosyl linkage types in humans: Fuc- α -1,2-Gal, Fuc- α -1,3-GlcNAc, Fuc- α -1,4-GlcNAc and Fuc- α -1,6-GlcNAc, then elongated the structures adding different monosaccharides to build the most common fucosylated trisaccharide epitopes. There is ongoing work to cover the linkage types that have not been included in this Thesis dissertation. All the subsequent MD simulations were performed with AMBER [133]. All protocols were maintained as similar as possible, although the atomic positional constraint schemes were system depending.

6.4.2.1 System preparation

Complex 1 The disaccharide Fuc- α -1,2-Gal was docked to one monomer of the crystal structure of FucA1 in complex with DNJ (PDB 7PM4). The initial structure of the disaccharide was built with Glycam (<http://glycam.org>) [246] To obtain an initial guess of the substrate glycosidic bond torsion, we used AutoDock Vina version 1.1.2 [138]. This approach was chosen as the pairwise alignment of the most abundant rotamer in water to the DFJ inhibitor at the active site did not yield stable results (not shown). We centered the search space at the enzyme active site, including the catalytic loop, with a cell of $30 \times 24 \times 30 \text{ \AA}$. We used the flexible docking approach, including the all the substrate dihedrals and residues of the catalytic loop surrounding Asp276 (Tyr272, Asn273, Cys274, Glu275, Asp276, Lys277, Phe278, Lys279, Gln281) (Figure S6.5). The flexibility of these region is observed in the cryo-EM structures [27] and in the classical MD simulations of section 6.4.1.2. The protocol was replicated with 10 different random seeds, and the maximum number of modes was set to 20. The position of the Gal moiety had been previously assessed to be relevant in previous MD simulations (not shown). Some poses corresponded to rotameric states of the glycosidic bond that alleviated the steric hindrance with the catalytic acid, and one of these was selected to continue the simulations. The Fuc moiety was positioned in subsite -1 like the DFJ inhibitor (Figure S6.5). We

assigned protonation states using H++ [137] at pH 5.0. The putative acid/base Asp276 was assigned in its carboxylic acid form, as necessary for protonating the scissile bond O according to the predicted mechanism. A rise in the predicted pK_a of this residue in the MC with Fuc- α -1,2-Gal with respect to the unliganded form supported this [137]. The system was solvated in a cell of $117.21 \times 92.77 \times 84.14$ Å with 23106 water molecules. 69 Na⁺ cations and 61 Cl⁻ were added to reach system neutrality and a salt concentration of 150 mM, with the SLTCAP webserver [245]. The same protocol was applied to the other complexes (from **2** to **8**).

Complexes with substrates 2 and 3 We built the models of FucA1 in complex with GalNAc- α -1,3-[Fuc- α -1,2]-Gal and Gal- α -1,3-[Fuc- α -1,2]-Gal starting from a trajectory snapshot of the classical MD simulation of complex **1** (see section 6.4.2.2) and we attached the α -1,3-linked GalNAc or Gal units, respectively, using the LEaP program from the AMBER suite of programs [133]. The dimensions of the solvation box were: complex **2**, $97.87 \times 116.99 \times 95.59$ Å, including 28767 water molecules, and complex **3**, $97.88 \times 116.99 \times 95.59$ Å.

Complexes with substrates 4 and 5 We built the models of FucA1 in complex with Fuc- α -1,2-Gal- β -1,4-GlcNAc (H-antigen 2, complex **4**) and Fuc- α -1,2-Gal- β -1,3-GlcNAc (H-antigen 1, complex **5**) starting from a trajectory snapshot of the classical MD simulation of complex **1** and added the β -1,3 and β -1,4 linked GlcNAc units, respectively. The dimensions of solvation boxes were $81.70 \times 96.74 \times 119.42$ Å, with 24028 solvent molecules (complex **4**) and $81.70 \times 96.74 \times 119.42$ Å, with 24029 solvent molecules (complex **5**). The coordinates for the saccharides attached to the initial substrates were retrieved from Glycam (<http://glycam.org>) [246].

Complexes with substrates 6 to 8 We built the models of FucA1 in complex with Fuc- α -1,3-GlcNAc (complex **6**), Fuc- α -1,4-GlcNAc (complex **7**) and Fuc- α -1,6-GlcNAc (complex **8**) by pairwise alignment of the Fuc moieties of each substrate to the Fuc- α -1,2-Gal substrate of the classical MD trajectory snapshot. The pairwise alignment was performed with the PyMOL Molecular Graphics System, version 3.0 Schrödinger, LLC [247]. We solvated each model with cells of dimensions $97.87 \times 116.99 \times 95.99$ Å with 28768 solvent molecules (complex **6**), $97.86 \times 116.99 \times 95.59$ Å with 28772 solvent molecules (complex **7**) and $90.53 \times 89.87 \times 114.37$ Å, with 23746 solvent molecules (complex **8**). Coordinates for the initial structures of the saccharides were kindly provided by Dr. Callum Ives, Postdoctoral Researcher in the group of Prof. Elisa Fadda in Maynooth University.

6.4.2.2 Classical Molecular Dynamics

Complex 1 Classical molecular dynamics simulations of FucA1 in complex with Fuc- α -1,2-Gal were performed with the AMBER package [133], applying the forcefields FF19SB (protein) [65], GLYCAM06j (disaccharide) [64] and TIP3P (solvent) [141]. For running the simulations we applied the following protocol: first we performed a system equilibration with the same steps and parameters detailed in section 6.4.1.2. After the 200 ps run in the NVT ensemble we performed 3 independent replicas of 500 ns each in the NPT ensemble (Figure S6.6), with a 2 fs timestep. The same forcefields and classical MD protocol was applied for the other complexes (from **2** to **8**).

We analyzed hydrogen bonds with the CPPTRAJ module [131] of AMBER [133]. For calculating the binding free energy of the enzyme-substrate complex we used the MM/PB(GB)SA method [248] (https://valdes-tresanco-ms.github.io/gmx_MMPBSA/dev/) as implemented in the Python suite of programs gmx_MMPBSA [249], using the ACPYPE program [250] to obtain the inputs in GROMACS format. We used an ensemble of trajectory frames corresponding (where the RMSD evolution stabilized) for each replica, respectively. To alleviate correlation effects we selected frames with a stride of 2 (0.4 ns). For the calculation we used the default Generalized Born method [251] and a salt concentration of 150 mM.

Complexes 2 and 3 We employed the same forcefields as detailed in section 6.4.2.2, complex **1**. For the classical MD protocol, we applied the same one detailed in complex **11**. After 500 ns of production run, we removed the restraints for another 500 ns.

Complexes 4 and 5 We employed the same forcefields as detailed in section 6.4.2.2, complex **1**. For the classical MD protocol, we applied the same one detailed in complex **11**. For complex **4**, we ran 400 ns of classical MD applying positional constraints to the substrate, then extended the simulation without constraints until 800 ns. For complex **5**, we ran 500 of classical MD applying positional constraints on the substrate, and progressively lowering the force constants.

Complexes 6 to 8 We employed the same forcefields as detailed in section 6.4.2.2, complex **1**. For the classical MD protocol, we applied the same one detailed in complex **11**. For complexes **6** and **8** we ran 1 μ s of classical MD, 500 applying positional constraints on the substrate and 500 without. For complex **7**, we applied constraints for 400 ns, then released them until 800 ns.

6.4.3 The glycosylation reaction of FucA1 with Fuc- α -1,2-Gal

6.4.3.1 QM/MM molecular dynamics simulations of FucA1-Fuc- α -1,2-Gal complex

QM/MM MD simulations of the Michaelis complex of FucA1 started from a representative trajectory snapshot from the classical MD simulations of complex 1 (Fuc- α -1,2-Gal) where the nucleophile Asp225 and acid/base Asp276 were suitably positioned for catalysis. The QM/MM approach was applied as implemented in CP2K 9.1 [252], combining Born-Oppenheimer MD at the DFT level of theory for the defined QM atoms (via the program QUICKSTEP [253]). The QM region included 57 atoms, corresponding to the disaccharide substrate and the sidechains of residues Asp225 and Asp276, and was enclosed in a cell of $15.98 \times 15.03 \times 19.58$ Å (net charge -1.0) (Figure S6.15). The rest of the system atoms (75996) were treated with molecular mechanics (with the program FIST, also implemented in CP2K). The Integrated Molecular Orbital Molecular Mechanics (IMOMM) method [254] was used to treat the QM-MM boundary between atoms C_δ and C_β of Asp225 and Asp276. For calculating the electronic structure with DFT we used the Perdew, Burke and Ernzerhoff (PBE) functional [71], applying DFTD3 dispersion corrections [255] in the Gaussian and plane-waves (GPW) scheme. We used GTH pseudopotentials [256], as well as Gaussian triple- ζ valence polarized (TZV2P) basis set. We applied a plane-wave cutoff of 300 Ry and a relative cutoff of 60 Ry. The system was heated to 300 K with a Nosé-Hoover thermostat 6 ps MD run, using a 0.5 fs timestep.

6.4.3.2 QM/MM metadynamics simulations of the glycosylation step

We simulated the glycosylation step of the reaction mechanism of FucA1 with the natural substrate Fuc- α -1,2-Gal applying QM/MM MD to the metadynamics algorithm implemented in the open-source, community-developed PLUMED library [149,150]. We started from a trajectory snapshot of the QM/MM MD simulation described in section 6.4.3.1, and to drive the bias we described two collective variables (CVs) accounting for the reactive events taking place, both corresponding to the difference between two interatomic distances (Figure S6.15). CV1 was chosen as the difference between the $C1-C_{\gamma_{Asp225}}$ and the glycosidic bond distance $C1-O1$, which accounts for bond formation and cleavage. We chose $C_{\gamma_{Asp225}}$ avoid biasing the glycosylation from any of the O atoms of Asp225. CV2 was defined as the difference between the

$H_{Asp276-O1}$ and $H_{Asp276-O_{Asp276}}$ distances to represent the proton donation from the general acid/base. The deposition step of Gaussian kernels was set at 80 MD steps (40 fs), while the width was set at 0.11 and 0.16 for each CV, based on their amplitude during the unbiased QM/MM MD of section 6.4.3.1. The free energy barrier was set initially at 1.2 kcal mol⁻¹ to accelerate the initial sampling, then lowered to 1.0 kcal mol⁻¹ and 0.5 kcal mol⁻¹ for the crossing to the GEI, for improving the accuracy of the computed free energy barrier. The experimental free energy barrier was estimated from the rate constant k_{cat} obtained by Armstrong *et al.* applying Transition State Theory with the Eyring-Polanyi expression [122].

6.5 Supplementary information

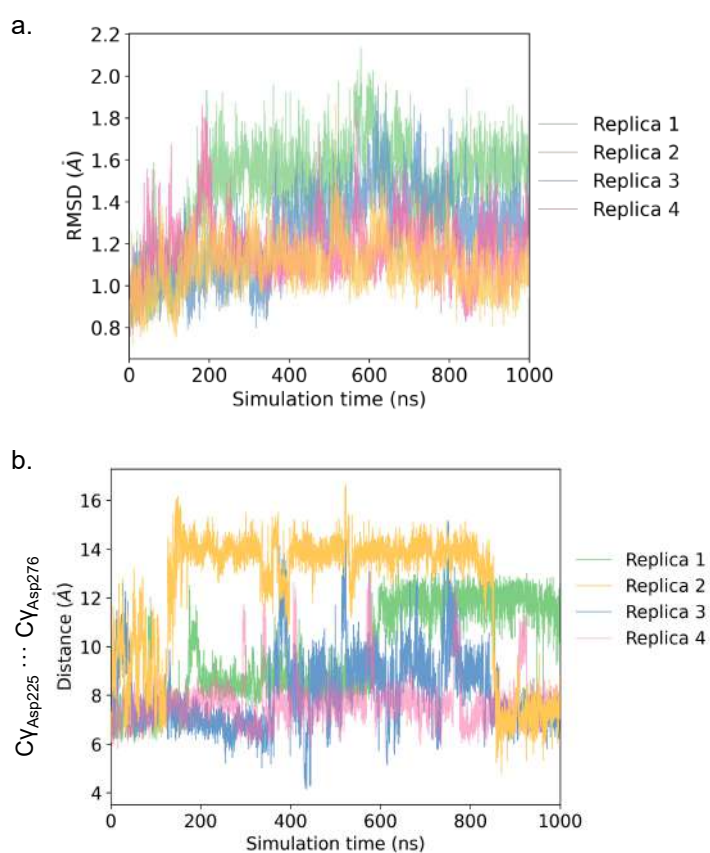


Figure S6.1: a. RMSD evolution of the 4 1 μ s replicas of classical MD simulations of the unliganded form of FucA1. b. Distance evolution between the C γ of the catalytic residues Asp225 and Asp275 during classical MD simulations of the unliganded FucA1.

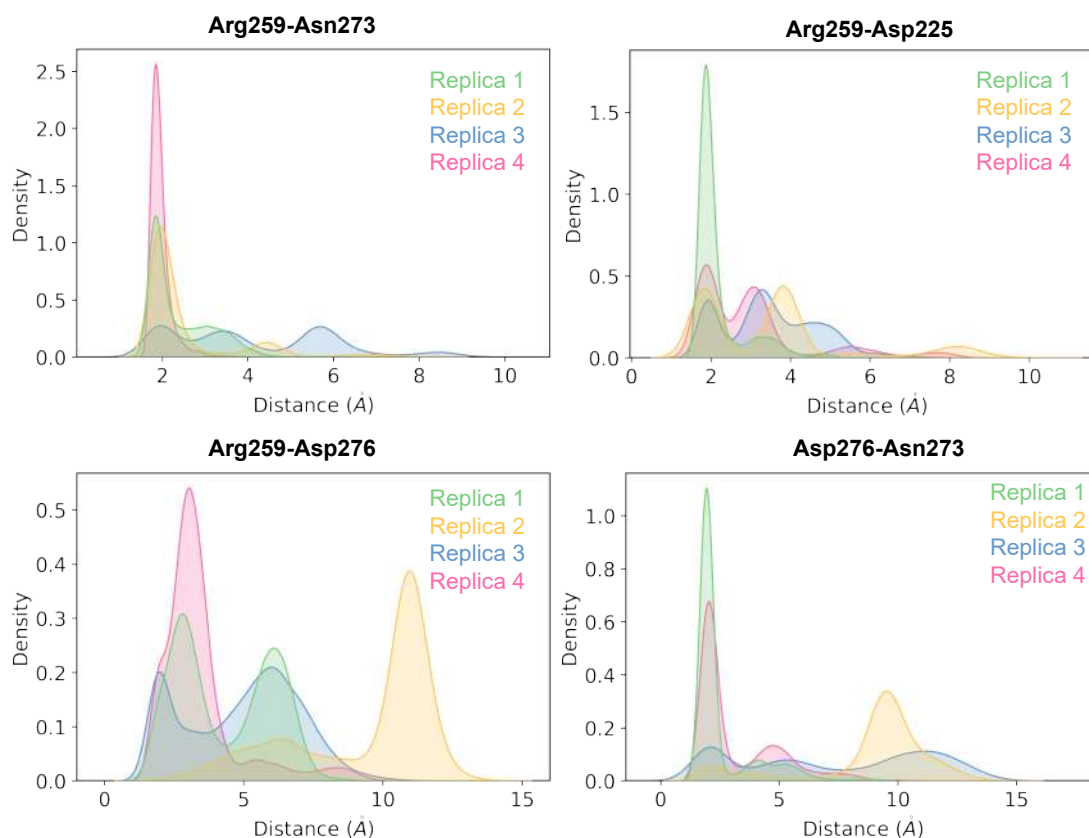


Figure S6.2: Distribution of important interaction distances involving residues of the Asp258-Lys287 loop across the four 1 μ s classical MD replicates. The high values of the hydrogen bond interactions of Asp276 with Arg259 and Asn273 respectively indicate the full loop opening in replica 2.

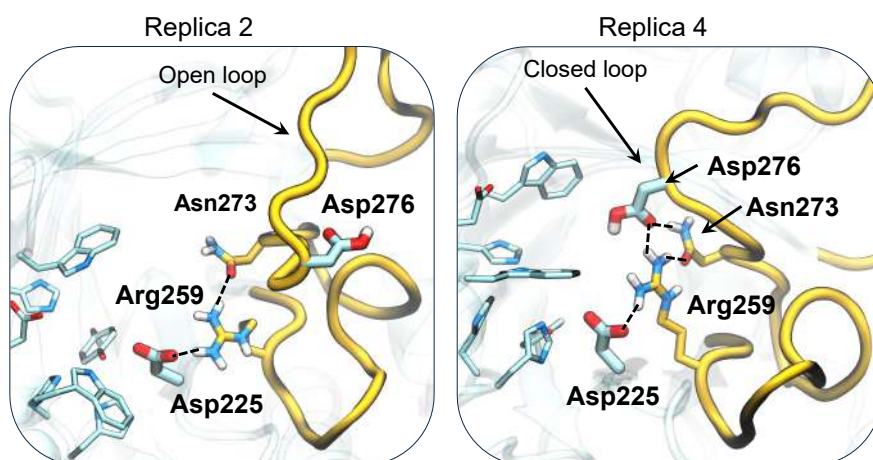


Figure S6.3: Trajectory snapshots corresponding to replica 2 (left) and replica 4 (right), showing the hydrogen bond interaction network that takes place between the catalytic residues (Asp225 and Asp276) and residues that are part of the Asp258-Lys287 recognition loop (Arg259, Asn273).

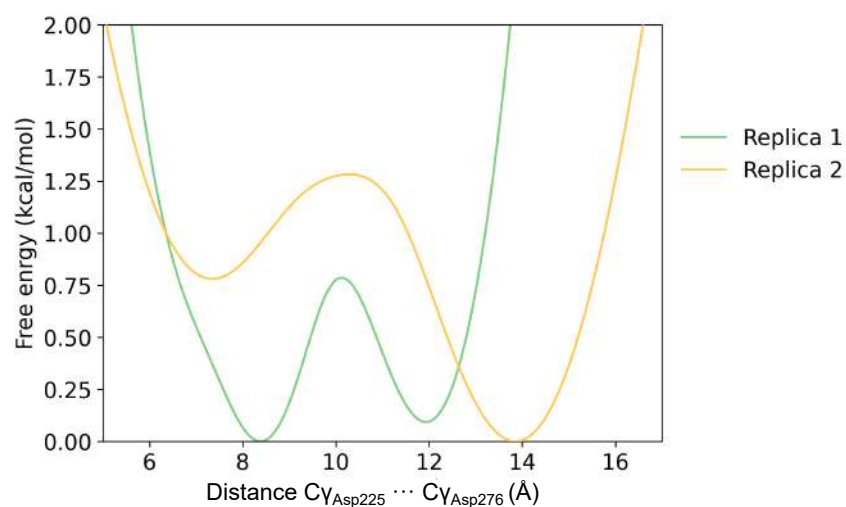


Figure S6.4: Free energy profile of the interatomic distance between the C γ of Asp225 and Asp276. Only replicas where both opened and closed states were sampled (replicas 1 and 2)

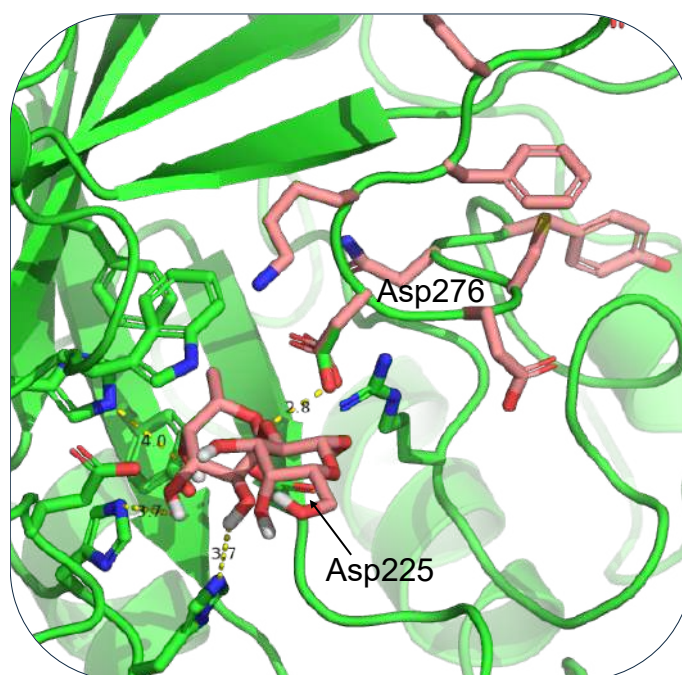


Figure S6.5: Docked Fuc- α -1,2-Gal used to build the complex with substrate **1**. In green, protein residues fixed in the docking calculations. In pink, ligand and residues whose dihedrals were allowed to rotate.

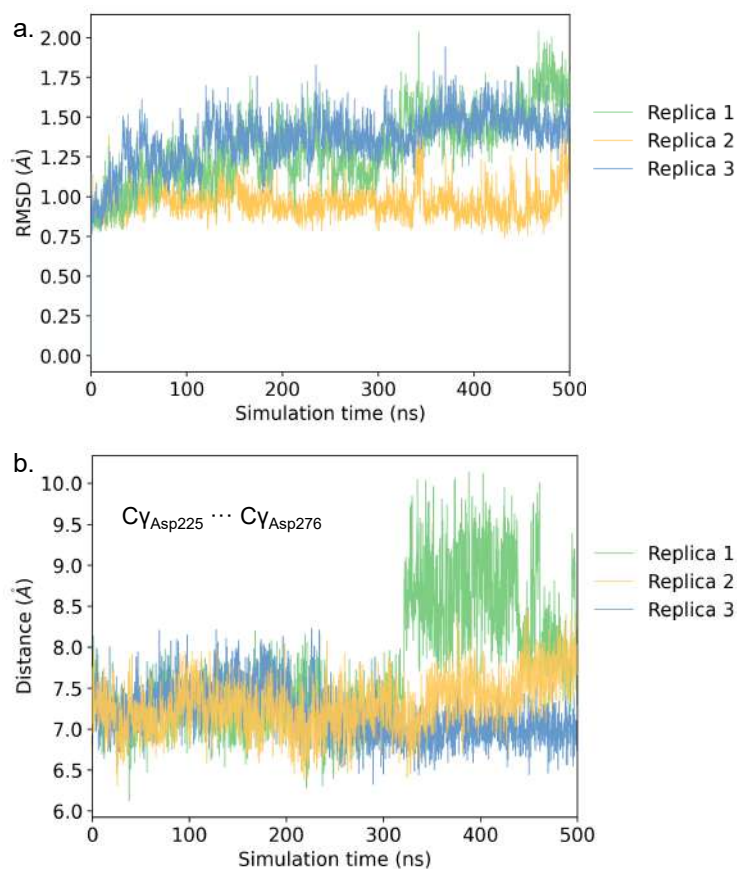


Figure S6.6: a. RMSD evolution of the three 500 ns replicas of classical MD of FucA1 in complex with Fuc- α -1,2-Gal. b. Time evolution of the interatomic distance between the sidechains of the catalytic pair Asp225 and Asp276, measured from each C _{γ} for the three replicas.

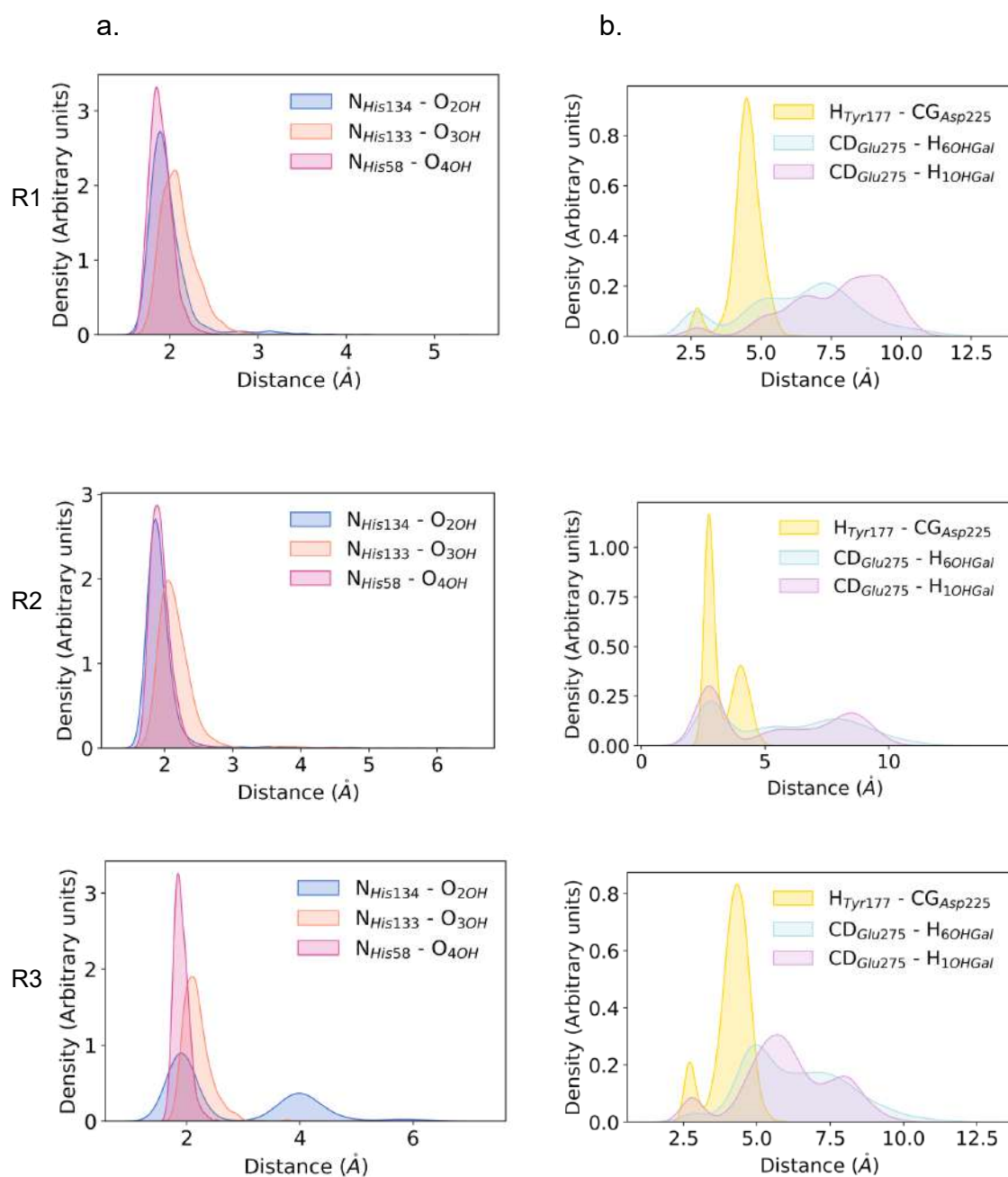


Figure S6.7: Density distribution of important interaction distances over classical MD simulations of FucA1 in complex with Fuc- α -1,2-Gal. Each row represents one of the 500 ns replicas. a. Interactions between active site residues at subsite -1 and the fucose moiety. b. Interactions between other relevant enzyme residues.

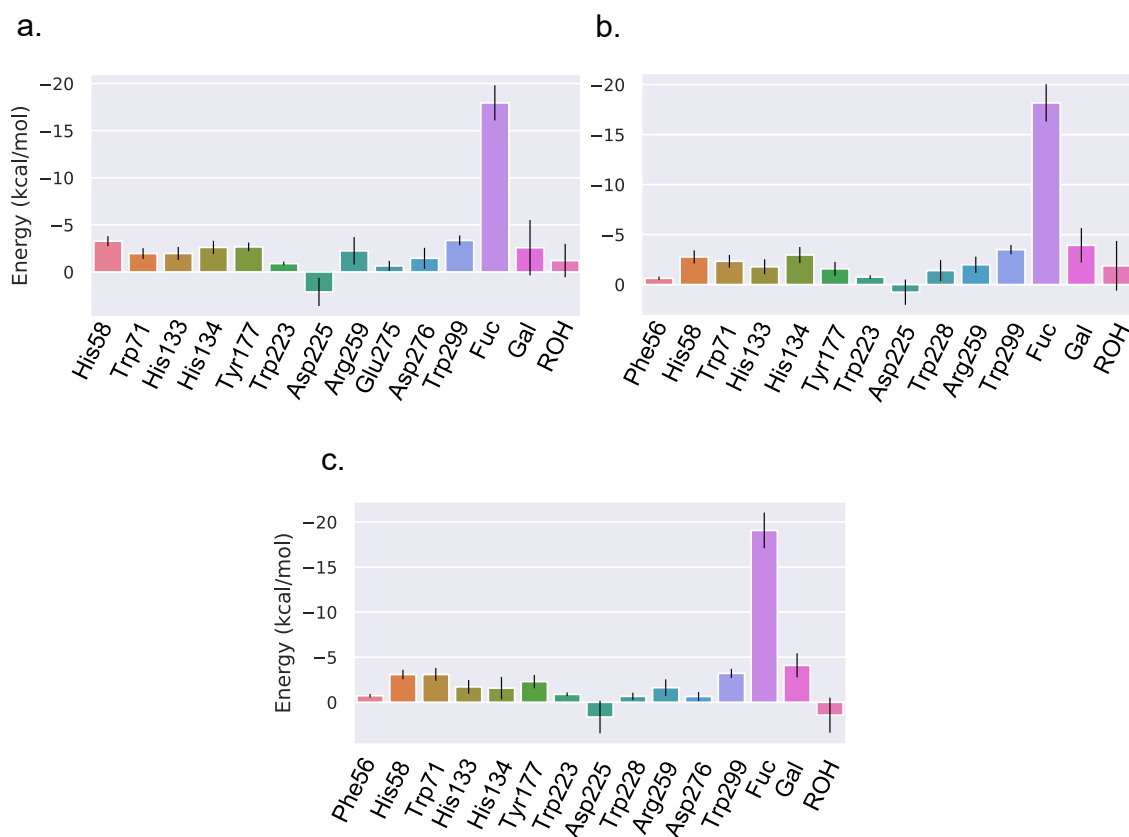


Figure S6.8: Binding free energy contribution by residue sidechain corresponding to the three 500 ns replicas of complex **1** simulations (Fuc- α -1,2-Gal). Plots generated with gmx_MMPBSA [248]

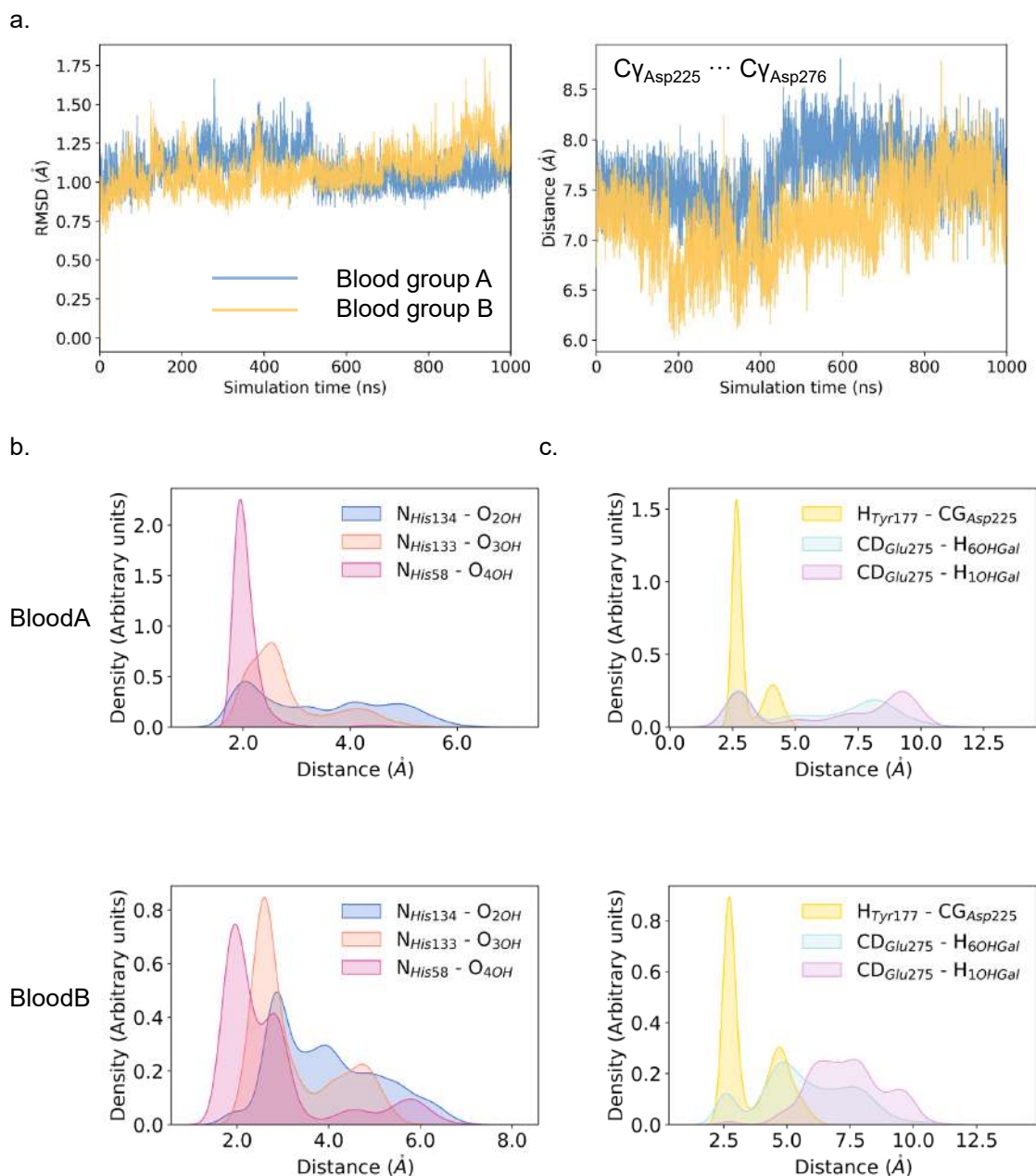


Figure S6.9: a. (Left) RMSD evolution of the classical MD of FucA1 in complex with the Fuc- α -1,2-linked blood groups A and B trisaccharides. (Right) Time evolution of the interatomic distance between the sidechains of the catalytic residues Asp225 and Asp276. b. Density distribution of interactions between residues at subsite -1 and the bound Fuc. c. Other interatomic distances with relevant residues: Tyr177 and Glu275.

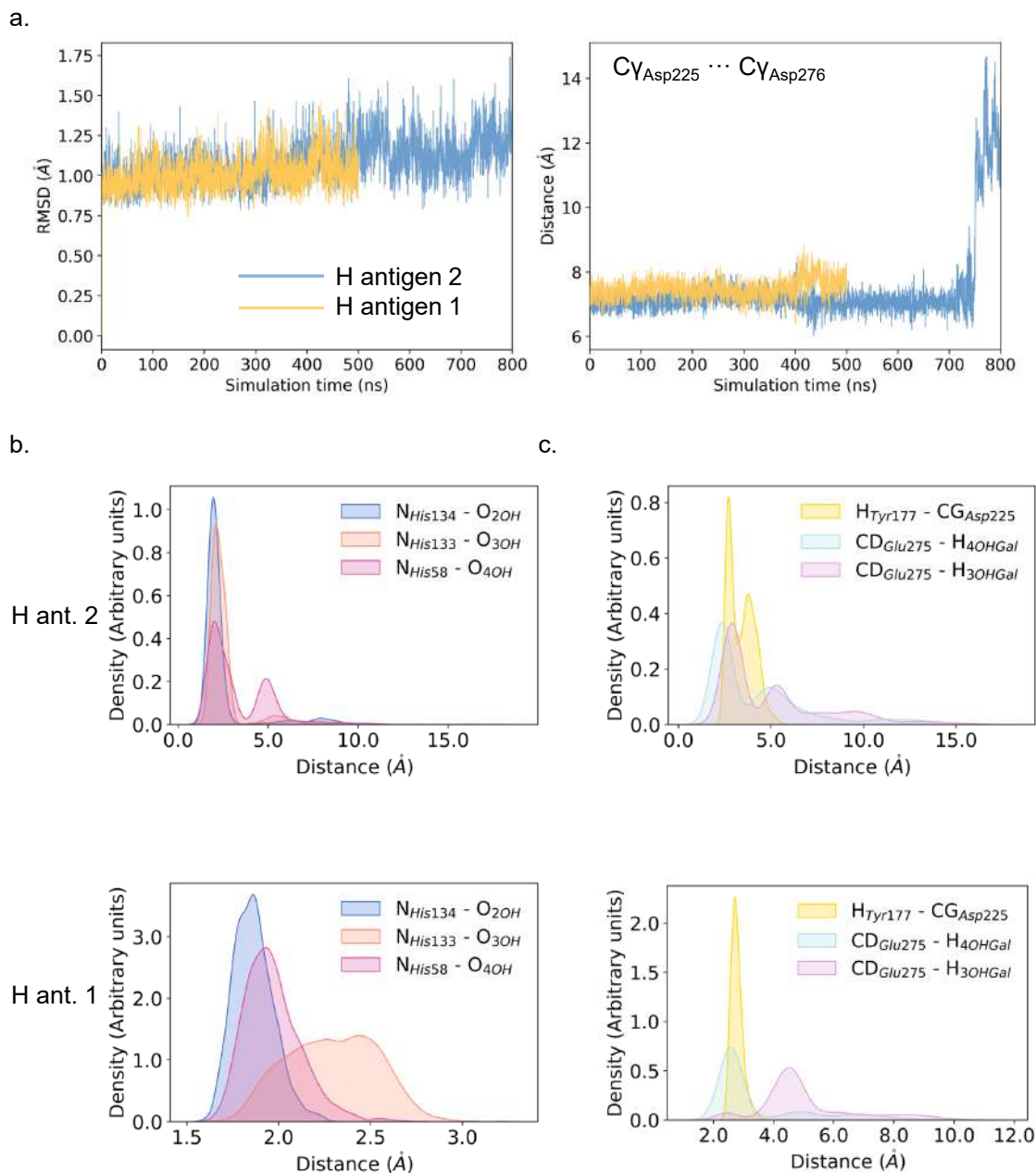


Figure S6.10: a. (Left) RMSD evolution of the classical MD of FucA1 in complex with the Fuc- α -1,2-linked H-antigens 2 and 1 trisaccharides. (Right) Time evolution of the interatomic distance between the sidechains of the catalytic residues Asp225 and Asp276. b. Density distribution of interactions between residues at subsite -1 and the bound Fuc. c. Other interatomic distances with relevant residues: Tyr177 and Glu275.

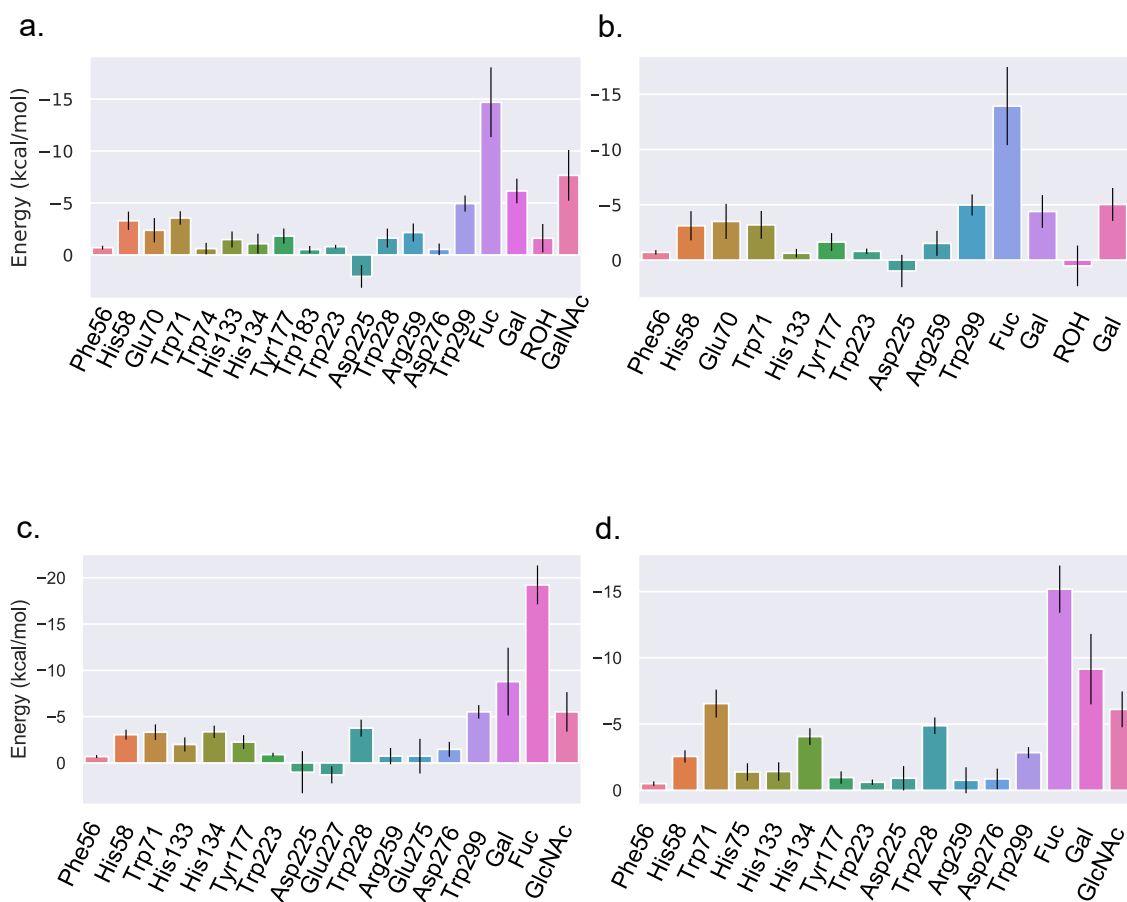


Figure S6.11: Binding free energy contribution by residue sidechain corresponding to the α -1,2-linked trisaccharides. a. Complex **2**, blood group A. b. Complex **3**, blood group B. c. Complex **4**, H antigen type 2. d. Complex **5**, H antigen type 1. Plots generated with gmx_MMPBSA [248]

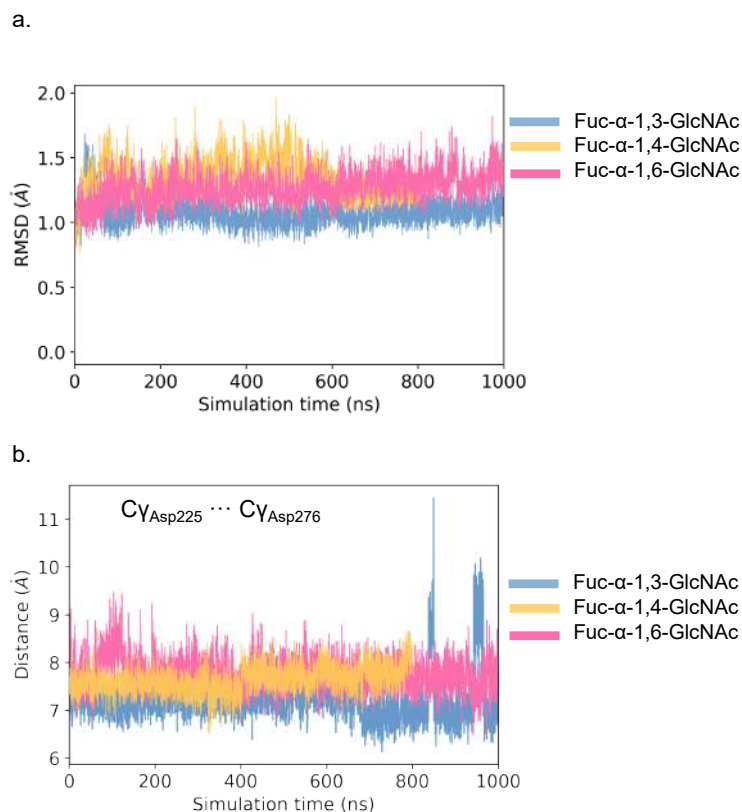


Figure S6.12: a. RMSD evolution of the classical MD of FucA1 in complex with the α -1,3, α -1,4 and α -1,6-GlcNAc disaccharides. b. Time evolution of the interatomic distance between the sidechains of the catalytic residues Asp225 and Asp276 during classical MD simulations of the GlcNAc-linked substrates.

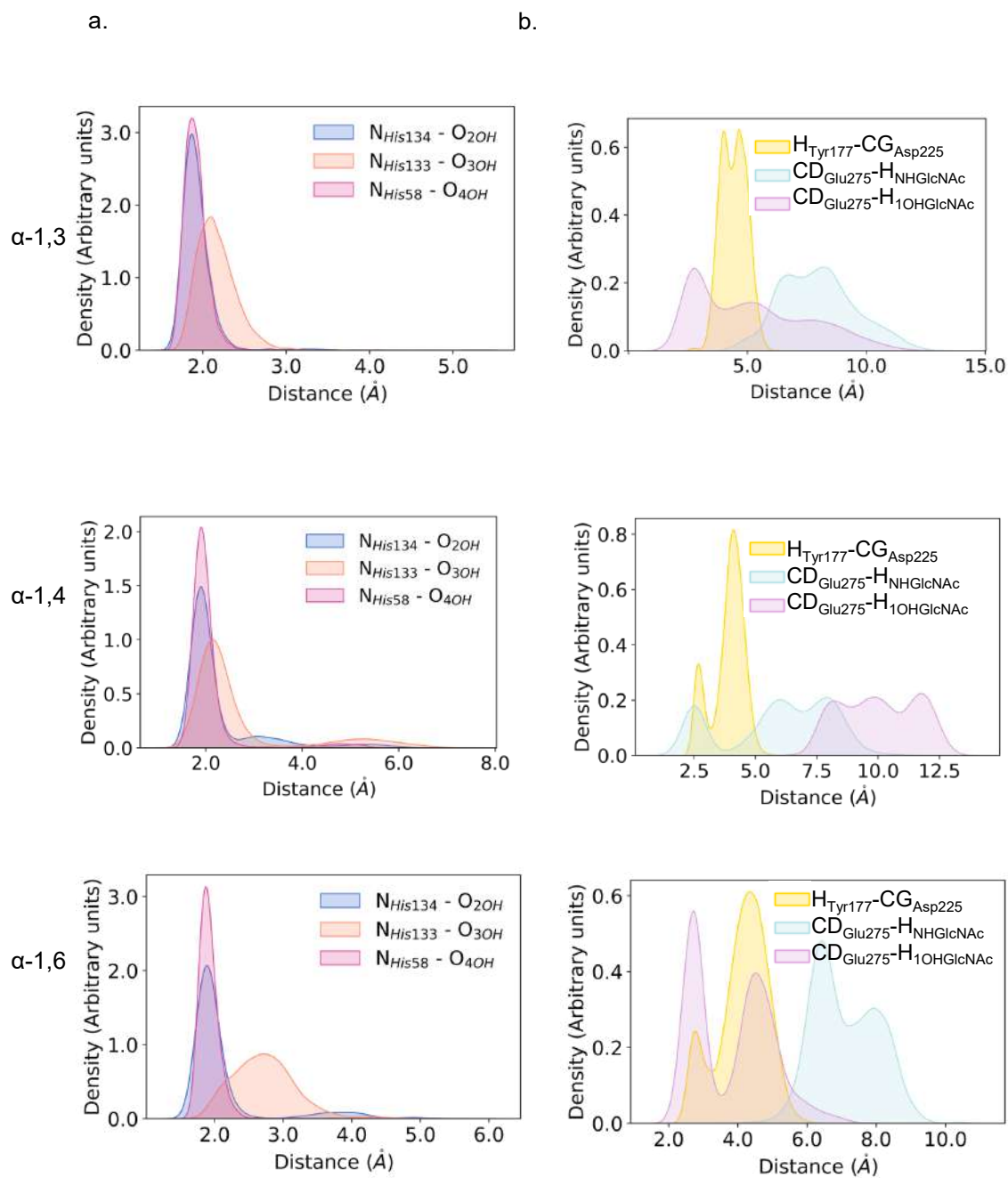


Figure S6.13: a. Density distribution of interactions between residues at subsite -1 and the bound Fuc. b. Other interatomic distances with relevant residues: Tyr177 and Glu275.

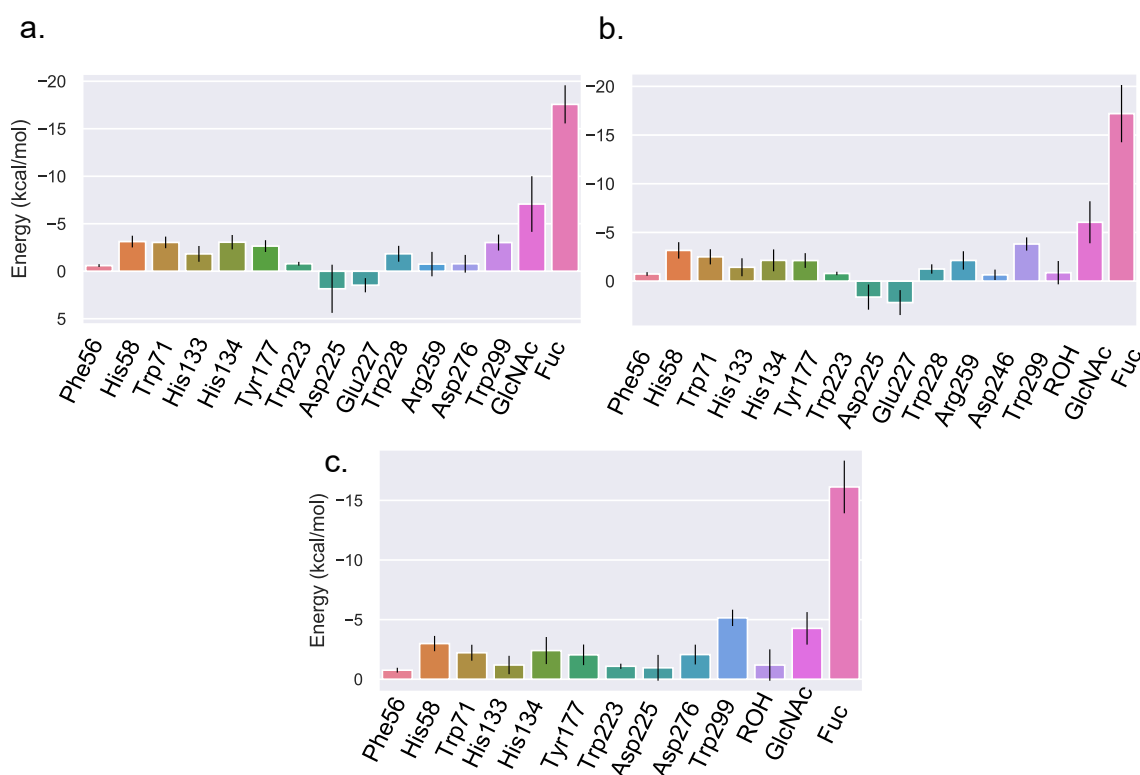


Figure S6.14: Binding free energy contribution by residue sidechain corresponding to the α -1,3/4/6-linked epitopes. a. Complex **6**, Fuc- α -1,3-GlcNAc. b. Complex **7**, Fuc- α -1,4-GlcNAc. c. Complex **8**, Fuc- α -1,6-GlcNAc. Plots generated with gmx.MMPBSA [248]

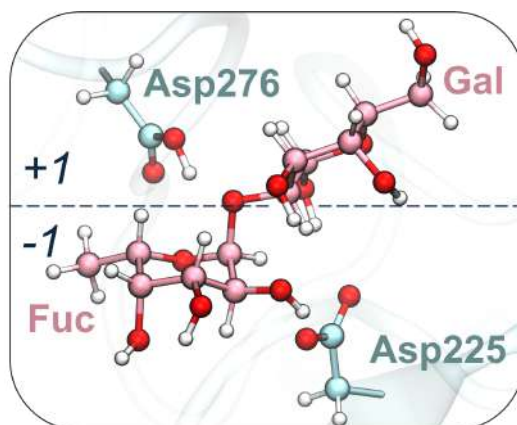


Figure S6.15: The 57 atoms chosen for the QM region in the QM/MM MD simulations of FucA1 in complex with naturally-occurring substrate Fuc- α -1,2-Gal.

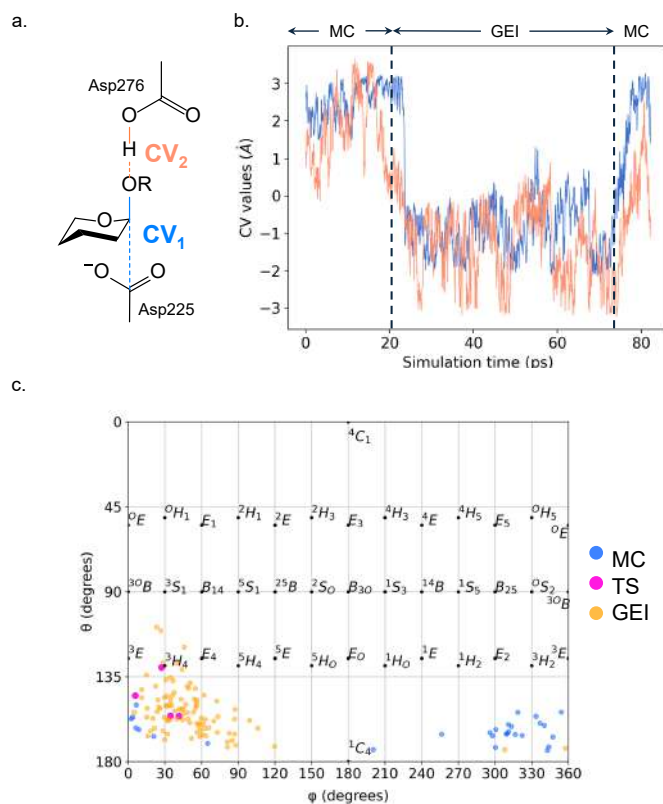


Figure S6.16: a. Scheme of the selected collective variables (CVs) for modelling the glycosylation step of the FucA1 catalytic mechanism. b. Evolution of the CVs during the QM/MM metadynamics simulation. Indicated on top, the sampled states (MC at 80 ps corresponds to the first recrossing). c. Mercator projection of the puckering coordinates of the fucose ring at the relevant states of the reaction mechanism.

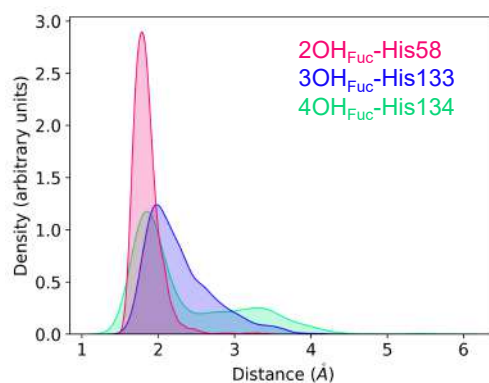


Figure S6.17: Distribution of the interaction distance between the fucosyl moiety at the enzyme -1 subsite and active site residues during the QM/MM metadynamics simulation of the glycosylation reaction step of FucA1.

Chapter 7

Conclusions

In this Thesis dissertation, we have presented four reaction mechanisms by glycoside hydrolases with naturally occurring substrates, studied with computational techniques based in MD methods coupled with the hybrid QM/MM approach, for describing the electronic rearrangements. We focused in four enzymes of biotechnological and biomedical interest. The first two enzymes investigated (LnbB and UhgbMP) display non-classical substrate-assisted mechanisms, where the substrate takes a part in the reaction as a catalytic agent, either as nucleophile (which is the case in LnbB) or as the last proton donor of the scissile glycosidic bond (in UhgbMP). Next, we studied the reaction mechanism of an enzyme that, while it follows a classically described mechanism, it operates on a nonstandard substrate (3,6-anhydro-D-galactose), becoming the first reported galactosidase with a conformational itinerary in the Southern hemisphere of the Cremer and Pople sphere (*ZgGH129*). All the enzymes above feature either non-classical catalytic residues or act on unusual substrates. Finally, we focus on an enzyme with a classical Koshland mechanism (the retaining fucosidase *FucA1*) that recognizes several fucose-based substrates.

The general conclusions extracted in this Thesis are the following:

- The GH20 enzyme LnbB degrades the natural substrate LNT, an important building block of human milk oligosaccharides, operating via substrate assisted mechanism where the substrate 2-NHAc group performs an intramolecular nucleophilic substitution. The role of Asp320 is to polarize the substrate, whose conformer suited for catalysis is ${}^{1,4}B$.
- Residue His263 of LnbB regulates the electrostatic micro-environment of the catalytic pair Asp320/Glu321, thus contributing to the necessary pK_a shifts. The prominent role of this residue in the catalytic reaction can explain the drop in activity of variants of this residue.

- The catalytic mechanism of UghbMP, an inverting bacterial glycoside phosphorylase that takes place with substrate assistance, where an exocyclic group of the β -mannosyl unit transfers a proton to the leaving group. This process occurs with a proton relay from the catalytic acidic residue, Asp104, situated far from the scissile bond. This mechanism can be extended to other glycoside phosphorylases with similar fold, such as GH130 *Ra*MP2, *Lin*0857 and GT108 MTPs.
- The hexameric organization of UghbMP contributes to the packing of the active site and prevents the entrance of water molecules, thus preventing hydrolysis instead of phosphorylation. Its structure also allows different acceptor sugars, a characteristic that sets it apart from other previously known bacterial phosphorylases and can be exploited to perform reverse phosphorolysis aiming to the synthesis of oligosaccharides.
- The catalytic mechanism of *Zg*GH129 3,6-anhydro-D-galactosidase for degrading the algal oligosaccharide β/κ -carrageenan takes place in the Southern hemisphere of the Cremer-Pople sphere. The presence of the 3,6-anhydro bridge in the substrate results in a very constrained geometry, which is exploited by the enzyme.
- The conformational free energy landscapes of 3,6-anhydro-D-galactose show few available conformations due to its constrained nature. Although the isolated molecule adopts a 1C_4 conformation, in the active site of the enzyme the $B_{1,4}$ conformation is the one suitable for catalysis.
- The catalytic mechanism of 3,6-anhydro-D-galactosidase *Zg*GH129 takes place via a two-step retaining mechanism, where the hydrolysis is the rate limiting step.
- GH29 human fucosidase FucA1 displays a recognition loop ranging from Asp258-Lys287 over its active site. It contains the catalytic acid/base Asp276 and its flexibility allows accommodating different fucosyl linkage types. This structural feature of FucA1 explains its promiscuity for different fucosylated substrates.
- FucA1 degrades fucosylated substrates via a retaining mechanism where residues Asp225 and Asp276 acts as nucleophile and general acid/base, respectively. A fucosyl-enzyme intermediate is formed via one transition state featuring a 3H_4 conformation.

-
- Enzymes take advantage of the saccharides inherent properties for enhancing catalysis and substrate binding. This can be in the form of small changes in their active sites, like the movement of a residue, or the motions of large structural elements, like loops. This work provides an example of this in the active sites of LnbB, where the motion of a residue assists the catalytic action, UhgbMP, where the scissile bond is far from the catalytic machinery, ZgGH129, which has adapted to cleave a substrate with a very constrained structure, and in FucA1, where the motion of a loop containing a catalytic residue is the responsible of substrate recognition.
 - Some enzymes utilize their substrates for direct intervention in catalysis, either as nucleophiles or proton donors of the departing group, thus elegantly capitalizing on the suitably oriented functional groups of the substrates. In this Thesis dissertation, we have presented two enzymes where the substrate is in charge of different catalytic functions, either as the nucleophile (LnbB) or the leaving group proton donor (UhgbMP).

Bibliography

- [1] J. Prestegard, *Essentials of Glycobiology*. United States: Cold Spring Harbor Laboratory Press, 4th ed., 2022.
- [2] E. Fadda, “Molecular simulations of complex carbohydrates and glycoconjugates,” *Current Opinion in Chemical Biology*, vol. 69, p. 102175, 2022.
- [3] M. Schneider, E. Al-Shareffi, and R. S. Haltiwanger, “Biological functions of fucose in mammals,” *Glycobiology*, vol. 27, no. 7, pp. 601–618, 2017.
- [4] N. Anderson, T. Dolan, and D. Rees, “Evidence for a common structural pattern in the polysaccharide sulphates of the *Rhodophyceae*,” *Nature*, vol. 205, no. 4976, pp. 1060–1062, 1965.
- [5] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, *et al.*, “Symbol nomenclature for graphical representations of glycans,” *Glycobiology*, vol. 25, no. 12, pp. 1323–1324, 2015.
- [6] K. Cheng, Y. Zhou, and S. Neelamegham, “DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information,” *Glycobiology*, vol. 27, no. 3, pp. 200–205, 2017.
- [7] E. Ficko-Blean, A. Préchoux, F. Thomas, T. Rochat, R. Larocque, Y. Zhu, M. Stam, S. Génicot, M. Jam, A. Calteau, B. Viart, D. Ropartz, D. Pérez-Pascual, G. Correc, M. Matard-Mann, K. A. Stubbs, H. Rogniaux, A. Jeudy, T. Barbeyron, C. Médigue, M. Czjzek, D. Vallenet, J. McBride, Mark, E. Duchaud, and M. Gurvan, “Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria,” *Nature Communications*, vol. 8, no. 1, p. 1685, 2017.
- [8] L. Bode, “Human milk oligosaccharides: Every baby needs a sugar mama,” *Glycobiology*, vol. 22, no. 9, pp. 1147–1162, 2012.
- [9] A. Nin-Hill and C. Rovira, “The catalytic reaction mechanism of the β -galactocerebrosidase enzyme deficient in Krabbe disease,” *ACS Catalysis*, vol. 10, no. 20, pp. 12091–12097, 2020.

- [10] C. A. Fogarty and E. Fadda, “Oligomannose *N*-Glycans 3D architecture and its response to the fc γ RIIIa structural landscape,” *The Journal of Physical Chemistry B*, vol. 125, no. 10, pp. 2607–2616, 2021.
- [11] V. Taleb, Q. Liao, Y. Narimatsu, A. García-García, I. Compañón, R. J. Borges, A. M. González-Ramírez, F. Corzana, H. Clausen, C. Rovira, and R. Hurtado-Guerrero, “Structural and mechanistic insights into the cleavage of clustered *O*-glycan patches-containing glycoproteins by mucinases of the human gut,” *Nature Communications*, vol. 13, no. 1, p. 4324, 2022.
- [12] B. Piniello, E. Lira-Navarrete, H. Takeuchi, M. Takeuchi, R. S. Haltiwanger, R. Hurtado-Guerrero, and C. Rovira, “Asparagine tautomerization in glycosyltransferase catalysis. The molecular mechanism of protein *O*-fucosyltransferase 1,” *ACS catalysis*, vol. 11, no. 15, pp. 9926–9932, 2021.
- [13] S. Ladevèze, L. Tarquis, D. A. Cecchini, J. Bercovici, I. André, C. M. Topham, S. Morel, E. Laville, P. Monsan, V. Lombard, B. Henrissat, and G. Potocki-Véronèse, “Role of glycoside phosphorylases in mannose foraging by human gut bacteria,” *Journal of Biological Chemistry*, vol. 288, no. 45, pp. 32370–32383, 2013.
- [14] L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, and R. E. Amaro, “Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein,” *ACS Central Science*, vol. 6, no. 10, pp. 1722–1734, 2020.
- [15] S. Minchin and J. Lodge, “Understanding biochemistry: structure and function of nucleic acids,” *Essays in Biochemistry*, vol. 63, no. 4, pp. 433–456, 2019.
- [16] A. Ardèvol and C. Rovira, “Reaction mechanisms in carbohydrate-active enzymes: Glycoside hydrolases and glycosyltransferases. Insights from *ab initio* quantum mechanics/molecular mechanics dynamic simulations,” *Journal of the American Chemical Society*, vol. 137, no. 24, pp. 7528–7547, 2015.
- [17] A.-A. Nomenclature, “Nomenclature Committee of IUB (NC-IUB) IUB-IUPAC Joint Commission on Biochemical Nomenclature (JCBN),” *Eur. J. Biochem*, vol. 104, pp. 321–322, 1980.
- [18] D. Cremer and J. Pople, “General definition of ring puckering coordinates,” *Journal of the American Chemical Society*, vol. 97, no. 6, pp. 1354–1358, 1975.
- [19] X. Biarnés, A. Ardèvol, A. Planas, C. Rovira, A. Laio, and M. Parrinello, “The conformational free energy landscape of β -D-glucopyranose. Implications for substrate preactivation in β -glucoside hydrolases,” *Journal of the American Chemical Society*, vol. 129, no. 35, pp. 10686–10693, 2007.

- [20] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, and N. Terrapon, “The carbohydrate-active enzyme database: functions and literature,” *Nucleic Acids Research*, vol. 50, no. D1, pp. D571–D577, 2022.
- [21] B. Henrissat and G. Davies, “Structural and sequence-based classification of glycoside hydrolases,” *Current Opinion in Structural Biology*, vol. 7, no. 5, pp. 637–644, 1997.
- [22] P. M. Coutinho, E. Deleury, G. J. Davies, and B. Henrissat, “An evolving hierarchical family classification for glycosyltransferases,” *Journal of Molecular Biology*, vol. 328, no. 2, pp. 307–317, 2003.
- [23] V. Puchart, “Glycoside phosphorylases: Structure, catalytic properties and biotechnological potential,” *Biotechnology Advances*, vol. 33, no. 2, pp. 261–276, 2015.
- [24] V. Lombard, T. Bernard, C. Rancurel, H. Brumer, P. M. Coutinho, and B. Henrissat, “A hierarchical classification of polysaccharide lyases for glycogenomics,” *Biochemical Journal*, vol. 432, no. 3, pp. 437–444, 2010.
- [25] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, and B. Henrissat, “Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes,” *Biotechnology for Biofuels*, vol. 6, pp. 1–14, 2013.
- [26] J. Coines, I. Cuxart, D. Teze, and C. Rovira, “Computer simulation to rationalize “rational” engineering of glycoside hydrolases and glycosyltransferases,” *The Journal of Physical Chemistry B*, vol. 126, no. 4, pp. 802–812, 2022.
- [27] Z. Armstrong, R. W. Meek, L. Wu, J. N. Blaza, and G. J. Davies, “Cryo-EM structures of human fucosidase FucA1 reveal insight into substrate recognition and catalysis,” *Structure*, vol. 30, no. 10, pp. 1443–1451, 2022.
- [28] S. Pengthaisong, B. Piniello, G. J. Davies, C. Rovira, and J. R. Ketudat Cairns, “Reaction mechanism of glycoside hydrolase family 116 utilizes perpendicular protonation,” *ACS Catalysis*, vol. 13, no. 9, pp. 5850–5863, 2023.
- [29] M. A. Morais, J. Coines, M. N. Domingues, R. A. Pirolla, C. C. Tonoli, C. R. Santos, J. B. Correa, F. C. Gozzo, C. Rovira, and M. T. Murakami, “Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations,” *Nature Communications*, vol. 12, no. 1, p. 367, 2021.
- [30] I. Cuxart, J. Coines, O. Esquivias, M. Faijes, A. Planas, X. Biarnés, and C. Rovira, “Enzymatic hydrolysis of human milk oligosaccharides. The molecular mechanism of *Bifidobacterium bifidum* lacto-*N*-biosidase,” *ACS Catalysis*, vol. 12, no. 8, pp. 4737–4743, 2022.

- [31] M. Faijes, M. Castejón-Vilatersana, C. Val-Cid, and A. Planas, “Enzymatic and cell factory approaches to the production of human milk oligosaccharides,” *Biotechnology Advances*, vol. 37, no. 5, pp. 667–697, 2019.
- [32] K. M. Stepien, E. Ciara, and A. Jezela-Stanek, “Fucosidosis—Clinical manifestation, long-term outcomes, and genetic profile—review and case series,” *Genes*, vol. 11, no. 11, p. 1383, 2020.
- [33] M. Alfonso-Prieto, I. Cuxart, G. Potocki-Véronèse, I. André, and C. Rovira, “Substrate-assisted mechanism for the degradation of *N*-glycans by a gut bacterial mannoside phosphorylase,” *ACS Catalysis*, vol. 13, no. 7, pp. 4283–4289, 2023.
- [34] B. Pluvinae, M. A. Higgins, D. W. Abbott, C. Robb, A. B. Dalia, L. Deng, J. N. Weiser, T. B. Parsons, A. J. Fairbanks, D. J. Vocadlo, and A. B. Boraston, “Inhibition of the pneumococcal virulence factor StrH and molecular insights into *N*-glycan recognition and hydrolysis,” *Structure*, vol. 19, no. 11, pp. 1603–1614, 2011.
- [35] J. K. Hobbs, B. Pluvinae, M. Robb, S. P. Smith, and A. B. Boraston, “Two complementary α -fucosidases from *Streptococcus pneumoniae* promote complete degradation of host-derived carbohydrate antigens,” *Journal of Biological Chemistry*, vol. 294, no. 34, pp. 12670–12682, 2019.
- [36] M. Calvelo, A. Males, M. G. Alteen, L. I. Willems, D. J. Vocadlo, G. J. Davies, and C. Rovira, “Human *O*-GlcNAcase uses a preactivated boat-skew substrate conformation for catalysis. Evidence from X-ray crystallography and QM/MM metadynamics,” *ACS Catalysis*, vol. 13, no. 20, pp. 13672–13678, 2023.
- [37] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J.-P. Mornon, and G. Davies, “Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases,” *Proceedings of the National Academy of Sciences*, vol. 92, no. 15, pp. 7090–7094, 1995.
- [38] M. A. B. Morais, A. Nin-Hill, and C. Rovira, “Glycosidase mechanisms: Sugar conformations and reactivity in endo-and exo-acting enzymes,” *Current Opinion in Chemical Biology*, vol. 74, p. 102282, 2023.
- [39] E. C. O’Neill and R. A. Field, “Enzymatic synthesis using glycoside phosphorylases,” *Carbohydrate Research*, vol. 403, pp. 23–37, 2015.
- [40] L. Tasse, J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, B. L. Cantarel, P. M. Coutinho, B. Henrissat, M. Leclerc, J. Doré, P. Monsan, M. Remaud-Simeon, and G. Potocki-Véronèse, “Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes,” *Genome Research*, vol. 20, no. 11, pp. 1605–1612, 2010.

- [41] A. Ardèvol, X. Biarnés, A. Planas, and C. Rovira, "The conformational free-energy landscape of β -D-mannopyranose: evidence for a ${}^1S_5 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^oS_2$ catalytic itinerary in β -mannosidases," *Journal of the American Chemical Society*, vol. 132, no. 45, pp. 16058–16065, 2010.
- [42] D. Koshland Jr, "Stereochemistry and the mechanism of enzymatic reactions," *Biological Reviews*, vol. 28, no. 4, pp. 416–436, 1953.
- [43] L. Petersen, A. Ardèvol, C. Rovira, and P. J. Reilly, "Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: a QM/MM metadynamics study," *The Journal of Physical Chemistry B*, vol. 113, no. 20, pp. 7331–7339, 2009.
- [44] A. J. Thompson, J. Dabin, J. Iglesias-Fernández, A. Ardèvol, Z. Dinev, S. J. Williams, O. Bande, A. Siriwardena, C. Moreland, T.-C. Hu, D. K. Smith, H. J. Gilbert, C. Rovira, and G. J. Davies, "The reaction coordinate of a bacterial GH47 α -mannosidase: a combined quantum mechanical and structural approach," *Angewandte Chemie International Edition*, vol. 124, no. 44, pp. 11159–11163, 2012.
- [45] X. Biarnés, A. Ardèvol, J. Iglesias-Fernández, A. Planas, and C. Rovira, "Catalytic itinerary in 1,3-1,4- β -glucanase unraveled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state," *Journal of the American Chemical Society*, vol. 133, no. 50, pp. 20301–20309, 2011.
- [46] J. Coines, L. Raich, and C. Rovira, "Modeling catalytic reaction mechanisms in glycoside hydrolases," *Current Opinion in Chemical Biology*, vol. 53, pp. 183–191, 2019.
- [47] N. G. McGregor, J. Coines, V. Borlandelli, S. Amaki, M. Artola, A. Nin-Hill, D. Linzel, C. Yamada, T. Arakawa, A. Ishiwata, Y. Ito, G. A. van der Marel, J. D. C. Codee, S. Fushinobu, H. S. Overkleeft, C. Rovira, and G. J. Davies, "Cysteine nucleophiles in glycosidase catalysis: Application of a covalent β -l-arabinofuranosidase inhibitor," *Angewandte Chemie International Edition*, vol. 133, no. 11, pp. 5818–5822, 2021.
- [48] S. Luang, X. Fernández-Luengo, A. Nin-Hill, V. A. Streltsov, J. G. Schwerdt, S. Alonso-Gil, J. R. Ketudat Cairns, S. Pradeau, S. Fort, J.-D. Maréchal, L. Masgrau, C. Rovira, and M. Hrmova, "The evolutionary advantage of an aromatic clamp in plant family 3 glycoside exo-hydrolases," *Nature Communications*, vol. 13, no. 1, p. 5577, 2022.
- [49] J. Coines, M. Alfonso-Prieto, X. Biarnés, A. Planas, and C. Rovira, "Oxazoline or oxazolinium ion? The protonation state and conformation of the reaction intermediate of chitinase enzymes revisited," *Chemistry—A European Journal*, vol. 24, no. 72, pp. 19258–19265, 2018.

- [50] L. F. Sobala, G. Speciale, S. Zhu, L. Raich, N. Sannikova, A. J. Thompson, Z. Hakki, D. Lu, S. Shamsi Kazem Abadi, A. R. Lewis, V. Rojas-Cervellera, A. J. Bennet, M. Sollogoub, C. Rovira, G. J. Davies, and S. J. Williams, "An epoxide intermediate in glycosidase catalysis," *ACS Central Science*, vol. 6, no. 5, pp. 760–770, 2020.
- [51] J. Coines Lopez-Nieto, *Mechanistic insights into substrate-assisted catalysis in glycosidases by means of QM/MM molecular dynamics*. Ph.D. dissertation, Universitat de Barcelona, Barcelona, 2021.
- [52] D. Teze, J. Coines, L. Raich, V. Kalichuk, C. Solleux, C. Tellier, C. André-Miral, B. Svensson, and C. Rovira, "A single point mutation converts GH84 *O*-glcnac hydrolases into phosphorylases: Experimental and theoretical evidence," *Journal of the American Chemical Society*, vol. 142, no. 5, pp. 2120–2124, 2020.
- [53] S. Nakae, S. Ito, M. Higa, T. Senoura, J. Wasaki, A. Hijikata, M. Shionyu, S. Ito, and T. Shirai, "Structure of novel enzyme in mannan biodegradation process 4-*O*- β -D-mannosyl-D-glucose phosphorylase MGP," *Journal of Molecular Biology*, vol. 425, no. 22, pp. 4468–4478, 2013.
- [54] G. Speciale, A. J. Thompson, G. J. Davies, and S. J. Williams, "Dissecting conformational contributions to glycosidase catalysis and inhibition," *Current Opinion in Structural Biology*, vol. 28, pp. 1–13, 2014.
- [55] X. Biarnés, J. Nieto, A. Planas, and C. Rovira, "Substrate distortion in the michaelis complex of bacillus 1, 3–1, 4- β -glucanase: Insight from first principles molecular dynamics simulations," *Journal of Biological Chemistry*, vol. 281, no. 3, pp. 1432–1441, 2006.
- [56] G. J. Davies, A. Planas, and C. Rovira, "Conformational analyses of the reaction coordinate of glycosidases," *Accounts of Chemical Research*, vol. 45, no. 2, pp. 308–316, 2012.
- [57] M. D. Wallace, I. Cuxart, T. Roret, L. Guée, A. W. Debowski, M. Czjzek, C. Rovira, K. A. Stubbs, and E. Ficko-Blean, "Constrained catalytic itinerary of a retaining 3, 6-anhydro-D-galactosidase, a key enzyme in red algal cell wall degradation," *Angewandte Chemie International Edition*, p. e202411171.
- [58] S. Ladevèze, G. Cioci, P. Roblin, L. Mourey, S. Tranier, and G. Potocki-Véronèse, "Structural bases for *N*-glycan processing by mannoside phosphorylase," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 6, pp. 1335–1346, 2015.
- [59] A. Vasella, G. J. Davies, and M. Böhm, "Glycosidase mechanisms," *Current Opinion in Chemical Biology*, vol. 6, no. 5, pp. 619–629, 2002.

- [60] F. Jensen, *Introduction to Computational Chemistry*. Hoboken, NJ, USA: John Wiley & Sons, Inc., second ed., 2017.
- [61] S. Prasad, D. Mobley, E. Braun, H. Mayes, J. Monroe, D. Zuckerman, *et al.*, “Best practices for foundations in molecular simulations [article v1. 0],” *Living Journal of Computational Molecular Science*, vol. 1, pp. 1–28, 2018.
- [62] L. Monticelli and D. P. Tieleman, *Force Fields for Classical Molecular Dynamics*, pp. 197–213. Totowa, NJ: Humana Press, 2013.
- [63] C. J. Cramer, *Essentials of Computational Chemistry – Theories and Methods*. Hoboken, NJ, USA: John Wiley & Sons, Inc., second ed., 2004.
- [64] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods, “GLYCAM06: A generalizable biomolecular force field. Carbohydrates,” *Journal of Computational Chemistry*, vol. 29, no. 4, pp. 622–655, 2008.
- [65] C. Tian, K. Kasavajhala, K. A. Belfon, L. Raguette, H. Huang, A. N. Migués, J. Bickel, Y. Wang, J. Pincay, Q. Wu, *et al.*, “ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution,” *Journal of Chemical Theory and Computation*, vol. 16, no. 1, pp. 528–552, 2019.
- [66] D. A. McQuarrie and J. D. Simon, *Physical Chemistry – A molecular approach*. Melville, NY, USA: University Science Books, first ed., 1997.
- [67] I. N. Levine, *Quantum Chemistry*. Upper Saddle River, NJ, USA: Pearson Education, Inc., seventh ed., 2014.
- [68] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas,” *Physical Review*, vol. 136, no. 3B, p. B864, 1964.
- [69] W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects,” *Physical Review*, vol. 140, no. 4A, p. A1133, 1965.
- [70] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics*. Cambridge, UK: Cambridge University Press, first ed., 2011.
- [71] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized gradient approximation made simple,” *Physical Review Letters*, vol. 77, no. 18, p. 3865, 1996.
- [72] G. I. Csonka, A. D. French, G. P. Johnson, and C. A. Stortz, “Evaluation of density functionals and basis sets for carbohydrates,” *Journal of Chemical Theory and Computation*, vol. 5, no. 4, pp. 679–692, 2009.

- [73] M. Marianski, A. Supady, T. Ingram, M. Schneider, and C. Baldauf, "Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies for the examples of glucose and α -maltose," *Journal of Chemical Theory and Computation*, vol. 12, no. 12, pp. 6157–6168, 2016.
- [74] N. Troullier and J. L. Martins, "Efficient pseudopotentials for plane-wave calculations," *Physical Review B*, vol. 43, no. 3, p. 1993, 1991.
- [75] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical Review Letters*, vol. 55, no. 22, p. 2471, 1985.
- [76] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, no. 2, pp. 227–249, 1976.
- [77] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 11, no. 6, pp. 700–733, 1990.
- [78] A. Laio, J. VandeVondele, and U. Rothlisberger, "A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 116, no. 16, pp. 6941–6947, 2002.
- [79] L. Raich, A. Nin-Hill, A. Ardèvol, and C. Rovira, "Enzymatic cleavage of glycosidic bonds: Strategies on how to set up and control a QM/MM metadynamics simulation," *Methods in Enzymology*, vol. 577, pp. 159–183, 2016.
- [80] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, "An efficient real space multigrid QM/MM electrostatic coupling," *Journal of Chemical Theory and Computation*, vol. 1, no. 6, pp. 1176–1184, 2005.
- [81] M. Iannuzzi, A. Laio, and M. Parrinello, "Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics," *Physical Review Letters*, vol. 90, no. 23, p. 238302, 2003.
- [82] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics*, vol. 71, no. 12, p. 126601, 2008.
- [83] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12562–12566, 2002.
- [84] J. Kästner, "Umbrella sampling," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 6, pp. 932–942, 2011.
- [85] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 5, pp. 826–843, 2011.

- [86] N. N. Nair, E. Schreiner, and D. Marx, "Peptide synthesis in aqueous environments: The role of extreme conditions on amino acid activation," *Journal of the American Chemical Society*, vol. 130, no. 43, pp. 14148–14160, 2008.
- [87] L. Bode and E. Jantscher-Krenn, "Structure-function relationships of human milk oligosaccharides," *Advances in Nutrition*, vol. 3, no. 3, pp. 383S–391S, 2012.
- [88] J. Wada, T. Ando, M. Kiyohara, H. Ashida, M. Kitaoka, M. Yamaguchi, H. Kumagai, T. Katayama, and K. Yamamoto, "*Bifidobacterium bifidum* lacto-*N*-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure," *Applied and Environmental Microbiology*, vol. 74, no. 13, pp. 3996–4004, 2008.
- [89] B. Zeuner, D. Teze, J. Muschiol, and A. S. Meyer, "Synthesis of human milk oligosaccharides: Protein engineering strategies for improved enzymatic transglycosylation," *Molecules*, vol. 24, no. 11, p. 2033, 2019.
- [90] D. Garrido, D. Barile, and D. A. Mills, "A molecular basis for bifidobacterial enrichment in the infant gastrointestinal tract," *Advances in Nutrition*, vol. 3, no. 3, pp. 415S–421S, 2012.
- [91] G. N. Bidart, J. Rodríguez-Díaz, M. Palomino-Schätzlein, V. Monedero, and M. J. Yebra, "Human milk and mucosal lacto-and galacto-*N*-biose synthesis by transgalactosylation and their prebiotic potential in *Lactobacillus* species," *Applied Microbiology and Biotechnology*, vol. 101, pp. 205–215, 2017.
- [92] E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, and N. Terapon, "The carbohydrate-active enzyme database: Functions and literature," *Nucleic Acids Research*, vol. 50, no. D1, pp. D571–D577, 2022.
- [93] M. Faijes, C. Val Cid, X. Biarnés Fontal, and A. P. S. Planas, "Structural-functional analysis reveals a specific domain organization in family GH20 hexosaminidases," *PLoS ONE*, vol. 10, no. 5, p. e0128075, 2015.
- [94] T. Ito, T. Katayama, M. Hattie, H. Sakurama, J. Wada, R. Suzuki, H. Ashida, T. Wakagi, K. Yamamoto, K. A. Stubbs, and S. Fushinobu, "Crystal structures of a glycoside hydrolase family 20 lacto-*N*-biosidase from *Bifidobacterium bifidum*," *Journal of Biological Chemistry*, vol. 288, no. 17, pp. 11795–11806, 2013.
- [95] M. Hattie, T. Ito, A. W. Debowski, T. Arakawa, T. Katayama, K. Yamamoto, S. Fushinobu, and K. A. Stubbs, "Gaining insight into the catalysis by GH20 lacto-*N*-biosidase using small molecule inhibitors and structural analysis," *Chemical Communications*, vol. 51, no. 81, pp. 15008–15011, 2015.

- [96] S. Knapp, D. Vocadlo, Z. Gao, B. Kirk, J. Lou, and S. G. Withers, "NAG-thiazoline, an *N*-acetyl- β -hexosaminidase inhibitor that implicates acetamido participation," *Journal of the American Chemical Society*, vol. 118, no. 28, pp. 6804–6805, 1996.
- [97] M. Castejón-Vilatersana, M. Faijes, and A. Planas, "Transglycosylation activity of engineered *Bifidobacterium* lacto-*N*-biosidase mutants at donor subsites for lacto-*N*-tetraose synthesis," *International Journal of Molecular Sciences*, vol. 22, no. 6, p. 3230, 2021.
- [98] D. Van Aalten, B. Synstad, M. Brurberg, E. Hough, B. Riise, V. Eijsink, and R. Wierenga, "Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9 Å resolution," *Proceedings of the National Academy of Sciences*, vol. 97, no. 11, pp. 5842–5847, 2000.
- [99] D. M. van Aalten, D. Komander, B. Synstad, S. Gåseidnes, M. G. Peter, and V. G. Eijsink, "Structural insights into the catalytic mechanism of a family 18 exo-chitinase," *Proceedings of the National Academy of Sciences*, vol. 98, no. 16, pp. 8979–8984, 2001.
- [100] S. Drouillard, S. Armand, J. G. Davies, E. C. Vorgias, and B. Henrissat, "*Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation," *Biochemical Journal*, vol. 328, no. 3, pp. 945–949, 1997.
- [101] B. L. Mark, D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers, and M. N. James, "Crystallographic evidence for substrate-assisted catalysis in a bacterial β -hexosaminidase," *Journal of Biological Chemistry*, vol. 276, no. 13, pp. 10330–10337, 2001.
- [102] Y.-L. Jiang, W.-L. Yu, J.-W. Zhang, C. Frolet, A.-M. Di Guilmi, C.-Z. Zhou, T. Vernet, and Y. Chen, "Structural basis for the substrate specificity of a novel β -*N*-acetylhexosaminidase StrH protein from *Streptococcus pneumoniae* R6," *Journal of Biological Chemistry*, vol. 286, no. 50, pp. 43004–43012, 2011.
- [103] Z. Zhang, M. Dong, R. Zallot, G. M. Blackburn, N. Wang, C. Wang, L. Chen, P. Baumann, Z. Wu, Z. Wang, H. Fan, C. Roth, Y. Jin, and Y. He, "Mechanistic and structural insights into the specificity and biological functions of bacterial sulfoglycosidases," *ACS Catalysis*, vol. 13, no. 1, pp. 824–836, 2022.
- [104] M. S. Macauley, G. E. Whitworth, A. W. Debowski, D. Chin, and D. J. Vocadlo, "*O*-glcNAcase uses substrate-assisted catalysis: Kinetic analysis and development of highly selective mechanism-inspired inhibitors," *Journal of Biological Chemistry*, vol. 280, no. 27, pp. 25313–25322, 2005.
- [105] Y. He, M. S. Macauley, K. A. Stubbs, D. J. Vocadlo, and G. J. Davies, "Visualizing the reaction coordinate of an *O*-GlcNAc hydrolase," *Journal of the American Chemical Society*, vol. 132, no. 6, pp. 1807–1809, 2010.

- [106] C. Roth, S. Chan, W. A. Offen, G. R. Hemsworth, L. I. Willems, D. T. King, V. Varghese, R. Britton, D. J. Vocadlo, and G. J. Davies, “Structural and functional insight into human *O*-GlcNAcase,” *Nature Chemical Biology*, vol. 13, no. 6, pp. 610–612, 2017.
- [107] Z. Marković-Housley, G. Miglierini, L. Soldatova, P. J. Rizkallah, U. Müller, and T. Schirmer, “Crystal structure of hyaluronidase, a major allergen of bee venom,” *Structure*, vol. 8, no. 10, pp. 1025–1035, 2000.
- [108] K. L. Chao, L. Muthukumar, and O. Herzberg, “Structure of human hyaluronidase-1, a hyaluronan hydrolyzing enzyme involved in tumor growth and angiogenesis,” *Biochemistry*, vol. 46, no. 23, pp. 6911–6920, 2007.
- [109] L. Zhang, A. G. Bharadwaj, A. Casper, J. Barkley, J. J. Barycki, and M. A. Simpson, “Hyaluronidase activity of human Hyal1 requires active site acidic and tyrosine residues,” *Journal of Biological Chemistry*, vol. 284, no. 14, pp. 9433–9442, 2009.
- [110] D. W. Abbott, M. S. Macauley, D. J. Vocadlo, and A. B. Boraston, “*Streptococcus pneumoniae* endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases,” *Journal of Biological Chemistry*, vol. 284, no. 17, pp. 11676–11689, 2009.
- [111] Z. Ling, M. D. Suits, R. J. Bingham, N. C. Bruce, G. J. Davies, A. J. Fairbanks, J. W. Moir, and E. J. Taylor, “The X-ray crystal structure of an *Arthrobacter protophormiae* endo- β -*N*-acetylglucosaminidase reveals a $(\beta/\alpha)_8$ catalytic domain, two ancillary domains and active site residues key for transglycosylation activity,” *Journal of Molecular Biology*, vol. 389, no. 1, pp. 1–9, 2009.
- [112] J. Yin, L. Li, N. Shaw, Y. Li, J. K. Song, W. Zhang, C. Xia, R. Zhang, A. Joachimiak, H.-C. Zhang, L.-X. Wang, Z.-J. Liu, and P. Wang, “Structural basis and catalytic mechanism for the dual functional endo- β -*N*-acetylglucosaminidase A,” *PLoS One*, vol. 4, no. 3, p. e4658, 2009.
- [113] I. Noach, B. Pluinage, C. Laurie, K. T. Abe, M. G. Alteen, D. J. Vocadlo, and A. B. Boraston, “The details of glycolipid glycan hydrolysis by the structural analysis of a family 123 glycoside hydrolase from *Clostridium perfringens*,” *Journal of Molecular Biology*, vol. 428, no. 16, pp. 3253–3265, 2016.
- [114] C. Roth, M. Petricevic, A. John, E. D. Goddard-Borger, G. J. Davies, and S. J. Williams, “Structural and mechanistic insights into a *Bacteroides vulgatus* retaining *N*-acetyl- β -galactosaminidase that uses neighbouring group participation,” *Chemical Communications*, vol. 52, no. 74, pp. 11096–11099, 2016.
- [115] D. J. Vocadlo and S. G. Withers, “Detailed comparative analysis of the catalytic mechanisms of β -*N*-acetylglucosaminidases from families 3 and 20 of glycoside hydrolases,” *Biochemistry*, vol. 44, no. 38, pp. 12809–12818, 2005.

- [116] I. Tews, A. Perrakis, A. Oppenheim, Z. Dauter, K. S. Wilson, and C. E. Vorgias, "Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay–Sachs disease," *Nature Structural Biology*, vol. 3, no. 7, pp. 638–648, 1996.
- [117] I. Tews, A. C. Terwisscha van Scheltinga, A. Perrakis, K. S. Wilson, and B. W. Dijkstra, "Substrate-assisted catalysis unifies two families of chitinolytic enzymes," *Journal of the American Chemical Society*, vol. 119, no. 34, pp. 7954–7959, 1997.
- [118] G. Prag, Y. Papanikolau, G. Tavlas, C. E. Vorgias, K. Petratos, and A. B. Oppenheim, "Structures of chitobiase mutants complexed with the substrate di-*N*-acetyl-D-glucosamine: The catalytic role of the conserved acidic pair, aspartate 539 and glutamate 540," *Journal of Molecular Biology*, vol. 300, no. 3, pp. 611–617, 2000.
- [119] B. L. Mark, D. J. Mahuran, M. M. Cherney, D. Zhao, S. Knapp, and M. N. James, "Crystal structure of human β -hexosaminidase B: Understanding the molecular basis of Sandhoff and Tay–Sachs disease," *Journal of Molecular Biology*, vol. 327, no. 5, pp. 1093–1109, 2003.
- [120] S. Fushinobu, "Unique sugar metabolic pathways of bifidobacteria," *Bio-science, Biotechnology, and Biochemistry*, vol. 74, no. 12, pp. 2374–2384, 2010.
- [121] I. R. Greig, F. Zahariev, and S. G. Withers, "Elucidating the nature of the *Streptomyces plicatus* β -hexosaminidase-bound intermediate using *ab initio* molecular dynamics simulations," *Journal of the American Chemical Society*, vol. 130, no. 51, pp. 17620–17628, 2008.
- [122] M. Garcia-Viloca, J. Gao, M. Karplus, and D. G. Truhlar, "How enzymes work: Analysis by modern rate theory and computer simulations," *Science*, vol. 303, no. 5655, pp. 186–195, 2004.
- [123] A. Planas, M. Abel, Ó. Millet, J. Palasi, C. Pallarés, and J.-L. Viladot, "Synthesis of aryl 3-O- β -cellobiosyl- β -d-glucopyranosides for reactivity studies of 1, 3-1, 4- β -glucanases," *Carbohydrate Research*, vol. 310, no. 1-2, pp. 53–64, 1998.
- [124] V. L. Schramm and W. Shi, "Atomic motion in enzymatic reaction coordinates," *Current Opinion in Structural Biology*, vol. 11, no. 6, pp. 657–665, 2001.
- [125] S. Feng, C. Bagia, and G. Mpourmpakis, "Determination of proton affinities and acidity constants of sugars," *The Journal of Physical Chemistry A*, vol. 117, no. 24, pp. 5211–5219, 2013.

- [126] R. D. Guthrie and W. P. Jencks, "IUPAC recommendations for the representation of reaction mechanisms," *Accounts of Chemical Research*, vol. 22, no. 10, pp. 343–349, 1989.
- [127] D. G. Isom, C. A. Castañeda, B. R. Cannon, and B. García-Moreno E, "Large shifts in pK_a values of lysine residues buried inside a protein," *Proceedings of the National Academy of Sciences*, vol. 108, no. 13, pp. 5260–5265, 2011.
- [128] A. C. Terwisscha van Scheltinga, S. Armand, K. H. Kalk, A. Isogai, B. Henrisat, and B. W. Dijkstra, "Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin evidence for substrate assisted catalysis," *Biochemistry*, vol. 34, no. 48, pp. 15619–15623, 1995.
- [129] L. P. McIntosh, G. Hand, P. E. Johnson, M. D. Joshi, M. Körner, L. A. Plesniak, L. Ziser, W. W. Wakarchuk, and S. G. Withers, "The pK_a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a ^{13}C -NMR study of *Bacillus circulans* xylanase," *Biochemistry*, vol. 35, no. 31, pp. 9958–9966, 1996.
- [130] W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.
- [131] D. R. Roe and T. E. Cheatham III, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, pp. 3084–3095, 2013.
- [132] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1–41, 1995.
- [133] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, H. Gohlke, A. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon Ferrer, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, L. Xiao, D. York, and P. Kollman, "AMBER 2018, University of California, San Francisco," *University of California, San Francisco*, 2018.

- [134] wwPDB Consortium, "Protein Data Bank: The single global archive for 3D macromolecular structure data," *Nucleic Acids Research*, vol. 47, no. D1, pp. D520–D528, 2019.
- [135] C. R. S ndergaard, M. H. Olsson, M. Rostkowski, and J. H. Jensen, "Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values," *Journal of Chemical Theory and Computation*, vol. 7, no. 7, pp. 2284–2295, 2011.
- [136] M. H. Olsson, C. R. S ndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions," *Journal of Chemical Theory and Computation*, vol. 7, no. 2, pp. 525–537, 2011.
- [137] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev, "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations," *Nucleic Acids Research*, vol. 40, no. W1, pp. W537–W541, 2012.
- [138] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, pp. 455–461, 2010.
- [139] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, "AutoDock Vina 1.2. 0: New docking methods, expanded force field, and Python bindings," *Journal of Chemical Information and Modeling*, vol. 61, no. 8, pp. 3891–3898, 2021.
- [140] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 3, pp. 712–725, 2006.
- [141] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.
- [142] H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [143] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n -alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

- [144] A. Laio, J. VandeVondele, and U. Roethlisberger, “D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations,” *The Journal of Physical Chemistry B*, vol. 106, no. 29, pp. 7300–7307, 2002.
- [145] V. Borlandelli, W. Offen, O. Moroz, A. Nin-Hill, N. McGregor, L. Binkhorst, A. Ishiwata, Z. Armstrong, M. Artola, C. Rovira, G. J. Davies, and H. S. Overkleeft, “ β -L-Arabinofurano-cyclitol aziridines are covalent broad-spectrum inhibitors and activity-based probes for retaining β -L-arabinofuranosidases,” *ACS Chemical Biology*, vol. 18, no. 12, pp. 2564–2573, 2023.
- [146] B. Piniello, J. Macías-León, S. Miyazaki, A. García-García, I. Compañón, M. Ghirardello, V. Taleb, B. Veloz, F. Corzana, A. Miyagawa, C. Rovira, and R. Hurtado-Guerrero, “Molecular basis for bacterial *N*-glycosylation by a soluble HMW1C-like *N*-glycosyltransferase,” *Nature Communications*, vol. 14, no. 1, p. 5785, 2023.
- [147] A. J. Borg, O. Esquivias, J. Coines, C. Rovira, and B. Nidetzky, “Enzymatic C4-epimerization of UDP-glucuronic acid: Precisely steered rotation of a transient 4-keto intermediate for an inverted reaction without decarboxylation,” *Angewandte Chemie International Edition*, vol. 62, no. 4, p. e202211937, 2023.
- [148] S. Nosé, “A unified formulation of the constant temperature molecular dynamics methods,” *The Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.
- [149] M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. d. L. Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhner, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provati, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, A. White, and T. P. Consortium, “Promoting transparency and reproducibility in enhanced molecular simulations,” *Nature Methods*, vol. 16, no. 8, pp. 670–673, 2019.
- [150] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, “PLUMED 2: New feathers for an old bird,” *Computer Physics Communications*, vol. 185, no. 2, pp. 604–613, 2014.

- [151] B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, "A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions," *The Journal of Physical Chemistry B*, vol. 109, no. 14, pp. 6676–6687, 2005.
- [152] I. Marcos-Alcalde, E. López-Viñas, and P. Gómez-Puertas, "MEPSAnd: Minimum energy path surface analysis over n -dimensional surfaces," *Bioinformatics*, vol. 36, no. 3, pp. 956–958, 2020.
- [153] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015.
- [154] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. Hratchian, J. Ortiz, A. Izmaylov, J. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. Peralta, F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin, V. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. Burant, S. Iyengar, J. Tomasi, M. Cossi, J. Millam, M. Klene, C. Adamo, R. Cammi, J. Ochterski, R. Martin, K. Morokuma, O. Farkas, J. Foresman, and D. Fox, "Gaussian 09 Revision D.01." Gaussian Inc. Wallingford CT 2009.
- [155] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *Journal of Molecular Graphics and Modelling*, vol. 25, no. 2, pp. 247–260, 2006.
- [156] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general Amber force field," *Journal of Computational Chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.
- [157] A. E. Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, and B. Henrissat, "The abundance and variety of carbohydrate-active enzymes in the human gut microbiota," *Nature Reviews Microbiology*, vol. 11, no. 7, pp. 497–504, 2013.
- [158] F. Cuskin, E. C. Lowe, M. J. Temple, Y. Zhu, E. A. Cameron, N. A. Pudlo, N. T. Porter, K. Urs, A. J. Thompson, A. Cartmell, A. Rogowski, B. S. Hamilton, R. Chen, T. J. Tolbert, K. Piens, D. Bracke, W. Vervecken, Z. Hakki, G. Speciale, J. L. Muñoz-Muñoz, A. Day, M. J. Peña, R. McLean, M. D. Suits,

- A. B. Boraston, T. Atherly, C. J. Ziemer, S. J. Williams, G. J. Davies, D. W. Abbott, E. C. Martens, and H. J. Gilbert, "Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism," *Nature*, vol. 517, no. 7533, pp. 165–169, 2015.
- [159] A. Li, E. Laville, L. Tarquis, V. Lombard, D. Ropartz, N. Terrapon, B. Henrissat, D. Guieysse, J. Esque, J. Durand, D. P. Morgavi, and G. Potocki-Véronèse, "Analysis of the diversity of the glycoside hydrolase family 130 in mammal gut microbiomes reveals a novel mannoside-phosphorylase function," *Microbial Genomics*, vol. 6, no. 10, p. e000404, 2020.
- [160] L. I. Crouch, M. V. Liberato, P. A. Urbanowicz, A. Baslé, C. A. Lamb, C. J. Stewart, K. Cooke, M. Doona, S. Needham, R. R. Brady, J. E. Berrington, K. Madunic, M. Wuhrer, P. Chater, J. P. Pearson, R. Glowacki, E. C. Martens, F. Zhang, R. J. Linhardt, D. I. R. Spencer, and D. N. Bolam, "Prominent members of the human gut microbiota express endo-acting *O*-glycanases to initiate mucin breakdown," *Nature Communications*, vol. 11, no. 1, p. 4017, 2020.
- [161] A. L. Kau, P. P. Ahern, N. W. Griffin, A. L. Goodman, and J. I. Gordon, "Human nutrition, the gut microbiome and the immune system," *Nature*, vol. 474, no. 7351, pp. 327–336, 2011.
- [162] C. Huttenhower, A. D. Kostic, and R. J. Xavier, "Inflammatory bowel disease as a model for translating the microbiome," *Immunity*, vol. 40, no. 6, pp. 843–854, 2014.
- [163] Y. H. Sheng, S. Z. Hasnain, T. H. Florin, and M. A. McGuckin, "Mucins in inflammatory bowel diseases and colorectal cancer," *Journal of Gastroenterology and Hepatology*, vol. 27, no. 1, pp. 28–38, 2012.
- [164] C. Reily, T. J. Stewart, M. B. Renfrow, and J. Novak, "Glycosylation in health and disease," *Nature Reviews Nephrology*, vol. 15, no. 6, pp. 346–366, 2019.
- [165] A. S. Tauzin, M. R. Pereira, L. D. Van Vliet, P.-Y. Colin, E. Laville, J. Esque, S. Laguerre, B. Henrissat, N. Terrapon, V. Lombard, M. Leclerc, J. Doré, F. Hollfelder, and G. Potocki-Véronèse, "Investigating host-microbiome interactions by droplet based microfluidics," *Microbiome*, vol. 8, pp. 1–20, 2020.
- [166] A. M. Harbison, L. P. Brosnan, K. Fenlon, and E. Fadda, "Sequence-to-structure dependence of isolated IgG Fc complex biantennary *N*-glycans: A molecular dynamics study," *Glycobiology*, vol. 29, no. 1, pp. 94–103, 2019.
- [167] A. Harbison and E. Fadda, "An atomistic perspective on antibody-dependent cellular cytotoxicity quenching by core-fucosylation of IgG1 Fc *N*-glycans from enhanced sampling molecular dynamics," *Glycobiology*, vol. 30, no. 6, pp. 407–414, 2020.

- [168] I. L. Grothaus, G. Bussi, and L. Colombi Ciacchi, “Exploration, representation, and rationalization of the conformational phase space of *N*-glycans,” *Journal of Chemical Information and Modeling*, vol. 62, no. 20, pp. 4992–5008, 2022.
- [169] S. R. Gill, M. Pop, R. T. DeBoy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, and K. E. Nelson, “Metagenomic analysis of the human distal gut microbiome,” *Science*, vol. 312, no. 5778, pp. 1355–1359, 2006.
- [170] M. F. Sernee, J. E. Ralton, T. L. Nero, L. F. Sobala, J. Kloehn, M. A. Vieira-Lara, S. A. Cobbold, L. Stanton, D. E. Pires, E. Hanssen, A. Males, T. Ward, L. M. Bastidas, P. K. van der Peet, M. W. Parker, D. B. Ascher, S. J. Williams, G. J. Davies, and M. J. McConville, “A family of dual-activity glycosyltransferase-phosphorylases mediates mannogen turnover and virulence in *Leishmania* parasites,” *Cell Host & Microbe*, vol. 26, no. 3, pp. 385–399, 2019.
- [171] A. Li, M. Benkoulouche, S. Ladeveze, J. Durand, G. Cioci, E. Laville, and G. Potocki-Véronèse, “Discovery and biotechnological exploitation of glycoside-phosphorylases,” *International Journal of Molecular Sciences*, vol. 23, no. 6, p. 3043, 2022.
- [172] Y. Ye, W. Saburi, R. Odaka, K. Kato, N. Sakurai, K. Komoda, M. Nishimoto, M. Kitaoka, H. Mori, and M. Yao, “Structural insights into the difference in substrate recognition of two mannoside phosphorylases from two GH130 subfamilies,” *FEBS Letters*, vol. 590, no. 6, pp. 828–837, 2016.
- [173] T. Nihira, E. Suzuki, M. Kitaoka, M. Nishimoto, K. Ohtsubo, and H. Nakai, “Discovery of β -1, 4-D-mannosyl-*N*-acetyl-D-glucosamine phosphorylase involved in the metabolism of *N*-glycans,” *Journal of Biological Chemistry*, vol. 288, no. 38, pp. 27366–27374, 2013.
- [174] C. Luley-Goedl and B. Nidetzky, “Carbohydrate synthesis by disaccharide phosphorylases: Reactions, catalytic mechanisms and application in the glycosciences,” *Biotechnology Journal*, vol. 5, no. 12, pp. 1324–1338, 2010.
- [175] T. Kuga, N. Sunagawa, and K. Igarashi, “Effect of free cysteine residues to serine mutation on cellodextrin phosphorylase,” *Journal of Applied Glycoscience*, vol. 71, no. 2, pp. 37–46, 2024.
- [176] F. Cuskin, A. Baslé, S. Ladevèze, A. M. Day, H. J. Gilbert, G. J. Davies, G. Potocki-Véronèse, and E. C. Lowe, “The GH130 family of mannoside phosphorylases contains glycoside hydrolases that target β -1,2-mannosidic linkages in *Candida mannan*,” *Journal of Biological Chemistry*, vol. 290, no. 41, pp. 25023–25033, 2015.

- [177] L. Dai, Z. Chang, J. Yang, W. Liu, Y. Yang, C.-C. Chen, L. Zhang, J.-W. Huang, Y. Sun, and R.-T. Guo, "Structural investigation of a thermostable 1,2- β -mannobiose phosphorylase from *Thermoanaerobacter* sp. X-514," *Biochemical and Biophysical Research Communications*, vol. 579, pp. 54–61, 2021.
- [178] M. Kitaoka, "Diversity of phosphorylases in glycoside hydrolase families," *Applied Microbiology and Biotechnology*, vol. 99, pp. 8377–8390, 2015.
- [179] C. Rovira, A. Males, G. J. Davies, and S. J. Williams, "Mannosidase mechanism: At the intersection of conformation and catalysis," *Current Opinion in Structural Biology*, vol. 62, pp. 79–92, 2020.
- [180] M. K. Bilyard, H. J. Bailey, L. Raich, M. A. Gafitescu, T. Machida, J. Iglésias-Fernández, S. S. Lee, C. D. Spicer, C. Rovira, W. W. Yue, and B. G. Davis, "Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis," *Nature*, vol. 563, no. 7730, pp. 235–240, 2018.
- [181] J. F. Darby, A. K. Gilio, B. Piniello, C. Roth, E. Blagova, R. E. Hubbard, C. Rovira, G. J. Davies, and L. Wu, "Substrate engagement and catalytic mechanisms of *N*-acetylglucosaminyltransferase V," *ACS Catalysis*, vol. 10, no. 15, pp. 8590–8596, 2020.
- [182] A. White, D. Tull, K. Johns, S. G. Withers, and D. R. Rose, "Crystallographic observation of a covalent catalytic intermediate in a β -glycosidase," *Nature Structural Biology*, vol. 3, no. 2, pp. 149–154, 1996.
- [183] L. Raich, V. Borodkin, W. Fang, J. Castro-Lopez, D. M. van Aalten, R. Hurtado-Guerrero, and C. Rovira, "A trapped covalent intermediate of a glycoside hydrolase on the pathway to transglycosylation. Insights from experiments and quantum mechanics/molecular mechanics simulations," *Journal of the American Chemical Society*, vol. 138, no. 10, pp. 3325–3332, 2016.
- [184] J. Hine, "The principle of least nuclear motion," in *Advances in Physical Organic Chemistry*, vol. 15, pp. 1–61, Elsevier, 1977.
- [185] M. L. Sinnott, "The principle of least nuclear motion and the theory of stereoelectronic control," in *Advances in Physical Organic Chemistry*, vol. 24, pp. 113–204, Elsevier, 1988.
- [186] A. Nakamura, T. Ishida, K. Kusaka, T. Yamada, S. Fushinobu, I. Tanaka, S. Kaneko, K. Ohta, H. Tanaka, K. Inaka, Y. Higuchi, N. Nimura, M. Samejima, and K. Igarashi, "'Newton's cradle' proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography," *Science Advances*, vol. 1, no. 7, p. e1500263, 2015.
- [187] T. P. Silverstein, "How enzymes harness highly unfavorable proton transfer reactions," *Protein Science*, vol. 30, no. 4, pp. 735–744, 2021.

- [188] P. Kumar, P. K. Agarwal, and M. J. Cuneo, "On the case of the misplaced hydrogens," *ChemBioChem*, vol. 22, no. 2, pp. 288–297, 2021.
- [189] V. Borshchevskiy, K. Kovalev, E. Round, R. Efremov, R. Astashkin, G. Bourenkov, D. Bratanov, T. Balandin, I. Chizhov, C. Baeken, I. Gushchin, A. Kuzmin, A. Alekseev, A. Rogachev, D. Willbold, M. Engelhard, E. Bamberg, G. Büldt, and V. Gordeliy, "True-atomic-resolution insights into the structure and functional role of linear chains and low-barrier hydrogen bonds in proteins," *Nature Structural & Molecular Biology*, vol. 29, no. 5, pp. 440–450, 2022.
- [190] G. D. Fasaman, *Handbook of Biochemistry and Molecular Biology: Physical and chemical data*. CRC Press, 1976.
- [191] T. K. Harris and G. J. Turner, "Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites," *IUBMB Life*, vol. 53, no. 2, pp. 85–98, 2002.
- [192] T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, and T. A. Steitz, "Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction," *Molecular Cell*, vol. 20, no. 3, pp. 437–448, 2005.
- [193] M. Beringer and M. V. Rodnina, "The ribosomal peptidyl transferase," *Molecular Cell*, vol. 26, no. 3, pp. 311–321, 2007.
- [194] K. G. Hoff, J. L. Avalos, K. Sens, and C. Wolberger, "Insights into the sirtuin mechanism from ternary complexes containing NAD^+ and acetylated peptide," *Structure*, vol. 14, no. 8, pp. 1231–1240, 2006.
- [195] T. Tsuda, T. Nihira, K. Chiku, E. Suzuki, T. Arakawa, M. Nishimoto, M. Kitaoka, H. Nakai, and S. Fushinobu, "Characterization and crystal structure determination of β -1,2-mannobiose phosphorylase from *Listeria innocua*," *FEBS Letters*, vol. 589, no. 24PartB, pp. 3816–3821, 2015.
- [196] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 1, pp. 12–21, 2010.
- [197] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids," *Nucleic acids research*, vol. 35, no. suppl_2, pp. W375–W383, 2007.

- [198] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, “Scalable molecular dynamics with NAMD,” *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005.
- [199] N. Homeyer, A. H. Horn, H. Lanig, and H. Sticht, “AMBER force-field parameters for phosphorylated amino acids in different protonation states: Phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine,” *Journal of Molecular Modeling*, vol. 12, pp. 281–289, 2006.
- [200] J. Aqvist, “Ion-water interaction potentials derived from free energy perturbation simulations,” *The Journal of Physical Chemistry*, vol. 94, no. 21, pp. 8021–8024, 1990.
- [201] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Physical Review A*, vol. 31, no. 3, p. 1695, 1985.
- [202] G. J. Martyna, D. J. Tobias, and M. L. Klein, “Constant pressure molecular dynamics algorithms,” *The Journal of Chemical Physics*, vol. 101, no. 5, pp. 4177–4189, 1994.
- [203] J. A. Izaguirre, S. Reich, and R. D. Skeel, “Longer time steps for molecular dynamics,” *The Journal of Chemical Physics*, vol. 110, no. 20, pp. 9853–9864, 1999.
- [204] J. Ireta, J. Neugebauer, and M. Scheffler, “On the accuracy of DFT for describing hydrogen bonds: dependence on the bond directionality,” *The Journal of Physical Chemistry A*, vol. 108, no. 26, pp. 5692–5698, 2004.
- [205] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, “Transition path sampling and the calculation of rate constants,” *The Journal of Chemical Physics*, vol. 108, no. 5, pp. 1964–1977, 1998.
- [206] S. H. Brawley, N. A. Blouin, E. Ficko-Blean, G. L. Wheeler, M. Lohr, H. V. Goodson, J. W. Jenkins, C. E. Blaby-Haas, K. E. Helliwell, C. X. Chan, T. N. Marriage, D. Bhattacharya, A. S. Klein, Y. Badis, J. Brodie, Y. Cao, J. Collén, S. M. Dittami, C. M. M. Gachon, B. R. Green, S. K. Karpowicz, J. W. Kim, U. J. Kudahl, S. Lin, M. Gurvan, M. Mittag, B. J. S. C. Olson, J. L. Pangilinan, Y. Peng, H. Qiu, S. Shu, J. T. Singer, A. G. Smith, B. N. Sprecher, V. Wagner, W. Wang, Z.-Y. Wang, J. Yan, C. Yarish, S. Zäuner-Riek, Y. Zhuang, Y. Zou, E. A. Lindquist, J. Grimwood, K. W. Barry, D. S. Rokhsar, J. Schmutz, J. W. Stiller, A. R. Grossmann, and S. E. Prochnik, “Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta),” *Proceedings of the National Academy of Sciences*, vol. 114, no. 31, pp. E6361–E6370, 2017.

- [207] S. Bengtson, T. Sallstedt, V. Belivanova, and M. Whitehouse, “Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae,” *PLoS Biology*, vol. 15, no. 3, p. e2000735, 2017.
- [208] G. Manat, M. Fanuel, D. Jouanneau, M. Jam, J. Mac-Bear, H. Rogniaux, T. Mora, R. Larocque, A. Lipinska, M. Czjzek, D. Ropartz, and E. Ficko-Blean, “Specificity of a β -porphyranase produced by the carrageenophyte red alga *Chondrus crispus* and implications of this unexpected activity on red algal biology,” *Journal of Biological Chemistry*, vol. 298, no. 12, 2022.
- [209] A. Chevenier, D. Jouanneau, and E. Ficko-Blean, “Carrageenan biosynthesis in red algae: A review,” *The Cell Surface*, vol. 9, p. 100097, 2023.
- [210] E. Sokolova, D. Jouanneau, A. Chevenier, M. Jam, N. Desban, P. Colas, E. Ficko-Blean, and G. Michel, “Enzymatically-derived oligo-carrageenans interact with α -Gal antibodies and Galectin-3,” *Carbohydrate Polymers*, vol. 324, p. 121563, 2024.
- [211] E. Ficko-Blean, D. Duffieux, É. Rebuffet, R. Larocque, A. Groisillier, G. Michel, and M. Czjzek, “Biochemical and structural investigation of two paralogous glycoside hydrolases from *Zobellia galactanivorans*: novel insights into the evolution, dimerization plasticity and catalytic mechanism of the GH117 family,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 2, pp. 209–223, 2015.
- [212] S. Genicot-Joncour, A. Poinas, O. Richard, P. Potin, B. Rudolph, B. Kloareg, and W. Helbert, “The cyclization of the 3,6-anhydro-galactose ring of ι -carrageenan is catalyzed by two D-galactose-2, 6-sulfurylases in the red alga *Chondrus crispus*,” *Plant Physiology*, vol. 151, no. 3, pp. 1609–1616, 2009.
- [213] E. Ficko-Blean, C. Hervé, and G. Michel, “Sweet and sour sugars from the sea: the biosynthesis and remodeling of sulfated cell wall polysaccharides from marine macroalgae,” *Perspectives in Phycology*, vol. 2, pp. 51–64, 05 2015.
- [214] M. D. Wallace, L. Guée, D. Ropartz, M. Fanuel, G. Lannuzel, G. Correc, K. A. Stubbs, and E. Ficko-Blean, “Characterisation of an exo-(α -1, 3)-3, 6-anhydro-D-galactosidase produced by the marine bacterium *Zobellia galactanivorans* dsijt: insight into enzyme preference for natural carrageenan oligosaccharides and kinetic characterisation on a novel chromogenic substrate,” *International Journal of Biological Macromolecules*, vol. 163, pp. 1471–1479, 2020.
- [215] M. D. Wallace, E. Ficko-Blean, and K. A. Stubbs, “Red algal molecules—Synthesis of methyl neo- β -carrabioside and its *S*-linked variant via two synthetic routes: a late stage ring closure and using a 3, 6-anhydro-D-galactosyl donor,” *The Journal of Organic Chemistry*, vol. 85, no. 24, pp. 16182–16195, 2020.

- [216] M. Ciancia, M. C. Matulewicz, and R. Tuvikene, “Structural diversity in galactans from red seaweeds and its influence on rheological properties,” *Frontiers in Plant Science*, vol. 11, p. 559986, 2020.
- [217] A. P. Lipinska, J. Collén, S. A. Krueger-Hadfield, T. Mora, and E. Ficko-Blean, “To gel or not to gel: differential expression of carrageenan-related genes between the gametophyte and tetrasporophyte life cycle stages of the red alga *Chondrus crispus*,” *Scientific Reports*, vol. 10, no. 1, p. 11498, 2020.
- [218] M. Khotimchenko, V. Tiasto, A. Kalitnik, M. Begun, R. Khotimchenko, E. Leonteva, I. Bryukhovetskiy, and Y. Khotimchenko, “Antitumor potential of carrageenans from marine red algae,” *Carbohydrate Polymers*, vol. 246, p. 116568, 2020.
- [219] M. Kiyohara, T. Nakatomi, S. Kurihara, S. Fushinobu, H. Suzuki, T. Tanaka, S.-i. Shoda, M. Kitaoka, T. Katayama, K. Yamamoto, and H. Ashida, “ α -N-acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway,” *Journal of Biological Chemistry*, vol. 287, no. 1, pp. 693–700, 2012.
- [220] Y. Jin, M. Petricevic, A. John, L. Raich, H. Jenkins, L. Portela De Souza, F. Cuskin, H. J. Gilbert, C. Rovira, E. D. Goddard-Borger, S. J. Williams, and G. J. Davies, “A β -mannanase with a lysozyme-like fold and a novel molecular catalytic mechanism,” *ACS Central Science*, vol. 2, no. 12, pp. 896–903, 2016.
- [221] A. Lammerts van Bueren, J. Fayers-Kerr, B. Luo, Y. Zhang, M. Sollogoub, Y. Bleériot, C. Rovira, and G. J. Davies, “Analysis of the reaction coordinate of α -l-fucosidases: combined structural and quantum mechanical approach,” *Journal of the American Chemical Society*, vol. 132, no. 6, pp. 1804–1806, 2010.
- [222] J. Iglesias-Fernández, L. Raich, A. Ardèvol, and C. Rovira, “The complete conformational free energy landscape of β -xylose reveals a two-fold catalytic itinerary for β -xylanases,” *Chemical Science*, vol. 6, no. 2, pp. 1167–1177, 2015.
- [223] Y. Zhang, T.-S. Lee, and W. Yang, “A pseudobond approach to combining quantum mechanical and molecular mechanical methods,” *The Journal of Chemical Physics*, vol. 110, no. 1, pp. 46–54, 1999.
- [224] R. Kumar, B. Henrissat, and P. M. Coutinho, “Intrinsic dynamic behavior of enzyme: substrate complexes govern the catalytic action of β -galactosidases across clan GH-A,” *Scientific Reports*, vol. 9, no. 1, p. 10346, 2019.

- [225] N. F. Brás, P. A. Fernandes, and M. J. Ramos, “QM/MM studies on the β -galactosidase catalytic mechanism: hydrolysis and transglycosylation reactions,” *Journal of Chemical Theory and Computation*, vol. 6, no. 2, pp. 421–433, 2010.
- [226] X.-L. Pan, W. Liu, and J.-Y. Liu, “Mechanism of the glycosylation step catalyzed by human α -galactosidase: A QM/MM metadynamics study,” *The Journal of Physical Chemistry B*, vol. 117, no. 2, pp. 484–489, 2013.
- [227] M. Artola, C. Hedberg, R. J. Rowland, L. Raich, K. Kytidou, L. Wu, A. Schaaf, M. J. Ferraz, G. A. van der Marel, J. D. Codée, C. Rovira, J. M. Aerts, G. J. Davies, and H. S. Overkleeft, “ α -D-Gal-cyclophellitol cyclosulfamidate is a Michaelis complex analog that stabilizes therapeutic lysosomal α -galactosidase A in Fabry disease,” *Chemical Science*, vol. 10, no. 40, pp. 9233–9243, 2019.
- [228] W. Ren, R. Pengelly, M. Farren-Dai, S. Shamsi Kazem Abadi, V. Oehler, O. Akintola, J. Draper, M. Meanwell, S. Chakladar, K. Świderek, V. Moliner, R. Britton, T. M. Gloster, and A. J. Bennet, “Revealing the mechanism for covalent inhibition of glycoside hydrolases by carbasugars at an atomic level,” *Nature Communications*, vol. 9, no. 1, p. 3243, 2018.
- [229] M. Sato, D. Liebschner, Y. Yamada, N. Matsugaki, T. Arakawa, S. S. Wills, M. Hattie, K. A. Stubbs, T. Ito, T. Senda, H. Ashida, and S. Fushinobu, “The first crystal structure of a family 129 glycoside hydrolase from a probiotic bacterium reveals critical residues and metal cofactors,” *Journal of Biological Chemistry*, vol. 292, no. 29, pp. 12126–12138, 2017.
- [230] P. Tiwary and M. Parrinello, “A time-independent free energy estimator for metadynamics,” *The Journal of Physical Chemistry B*, vol. 119, no. 3, pp. 736–742, 2015.
- [231] B. Ma, J. L. Simala-Grant, and D. E. Taylor, “Fucosylation in prokaryotes and eukaryotes,” *Glycobiology*, vol. 16, no. 12, pp. 158R–184R, 2006.
- [232] B. Zeuner and A. S. Meyer, “Enzymatic transfucosylation for synthesis of human milk oligosaccharides,” *Carbohydrate Research*, vol. 493, p. 108029, 2020.
- [233] J.-C. Michalski and A. Klein, “Glycoprotein lysosomal storage disorders: α - and β -mannosidosis, fucosidosis and α -N-acetylgalactosaminidase deficiency,” *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1455, no. 2-3, pp. 69–84, 1999.
- [234] G. Dawson and G. Tsay, “Substrate specificity of human α -L-fucosidase,” *Archives of Biochemistry and Biophysics*, vol. 184, no. 1, pp. 12–23, 1977.

- [235] H. Grootaert, L. Van Landuyt, P. Hulpiau, and N. Callewaert, “Functional exploration of the GH29 fucosidase family,” *Glycobiology*, vol. 30, no. 9, pp. 735–745, 2020.
- [236] V. N. Perna, K. Barrett, A. S. Meyer, and B. Zeuner, “Substrate specificity and transglycosylation capacity of α -L-fucosidases across GH29 assessed by bioinformatics-assisted selection of functional diversity,” *Glycobiology*, vol. 33, no. 5, pp. 396–410, 2023.
- [237] G. Sulzenbacher, C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers, B. Henrissat, and Y. Bourne, “Crystal structure of *Thermotoga maritima* α -L-fucosidase: Insights into the catalytic mechanism and the molecular basis for fucosidosis,” *Journal of Biological Chemistry*, vol. 279, no. 13, pp. 13119–13128, 2004.
- [238] F. A. Shaikh, A. Lammerts van Bueren, G. J. Davies, and S. G. Withers, “Identifying the catalytic acid/base in GH29 α -L-fucosidase subfamilies,” *Biochemistry*, vol. 52, no. 34, pp. 5857–5864, 2013.
- [239] T. Burgin and H. B. Mayes, “Mechanism of oligosaccharide synthesis via a mutant GH29 fucosidase,” *Reaction Chemistry & Engineering*, vol. 4, no. 2, pp. 402–409, 2019.
- [240] E. H. Klontz, C. Li, K. Kihn, J. K. Fields, D. Beckett, G. A. Snyder, P. L. Wintrode, D. Deredge, L.-X. Wang, and E. J. Sundberg, “Structure and dynamics of an α -fucosidase reveal a mechanism for highly efficient IgG transfucosylation,” *Nature Communications*, vol. 11, no. 1, p. 6204, 2020.
- [241] C. A. Tarling, S. He, G. Sulzenbacher, C. Bignon, Y. Bourne, B. Henrissat, and S. G. Withers, “Identification of the catalytic nucleophile of the family 29 α -L-fucosidase from *Thermotoga maritima* through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis,” *Journal of Biological Chemistry*, vol. 278, no. 48, pp. 47394–47399, 2003.
- [242] H. Sakurama, E. Tsutsumi, H. Ashida, T. Katayama, K. Yamamoto, and H. Kumagai, “Differences in the substrate specificities and active-site structures of two α -L-fucosidases (glycoside hydrolase family 29) from *Bacteroides thetaiotaomicron*,” *Bioscience, Biotechnology, and Biochemistry*, vol. 76, no. 5, pp. 1022–1024, 2012.
- [243] S.-W. Liu, C.-S. Chen, S.-S. Chang, K.-K. T. Mong, C.-H. Lin, C.-W. Chang, C. Y. Tang, and Y.-K. Li, “Identification of essential residues of human α -L-fucosidase and tests of its mechanism,” *Biochemistry*, vol. 48, no. 1, pp. 110–120, 2009.
- [244] R. O’Flaherty, A. M. Harbison, P. J. Hanley, C. H. Taron, E. Fadda, and P. M. Rudd, “Aminoquinoline fluorescent labels obstruct efficient removal of

- N*-glycan core $\alpha(1-6)$ fucose by bovine kidney α -L-fucosidase (BKF),” *Journal of Proteome Research*, vol. 16, no. 11, pp. 4237–4243, 2017.
- [245] J. D. Schmit, N. L. Kariyawasam, V. Needham, and P. E. Smith, “SLTCAP: A simple method for calculating the number of ions needed for MD simulation,” *Journal of Chemical Theory and Computation*, vol. 14, no. 4, pp. 1823–1827, 2018.
- [246] R. Woods, “Woods group, gLYCAM Web; Complex Carbohydrate Research Center,” *Athens, GA: University of Georgia*, 2005.
- [247] Schrödinger, LLC, “The PyMOL molecular graphics system, version 1.8.” November 2015.
- [248] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno, “gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS,” *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6281–6291, 2021.
- [249] B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, “MMPBSA.py: an efficient program for end-state free energy calculations,” *Journal of Chemical Theory and Computation*, vol. 8, no. 9, pp. 3314–3321, 2012.
- [250] A. W. Sousa da Silva and W. F. Vranken, “ACPYPE-Antechamber python parser interface,” *BMC Research Notes*, vol. 5, pp. 1–8, 2012.
- [251] A. Onufriev, D. Bashford, and D. A. Case, “Exploring protein native states and large-scale conformational changes with a modified generalized born model,” *Proteins: Structure, Function, and Bioinformatics*, vol. 55, no. 2, pp. 383–394, 2004.
- [252] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessi, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, “CP2K: An electronic structure and molecular dynamics software package—Quickstep: Efficient and accurate electronic structure calculations,” *The Journal of Chemical Physics*, vol. 152, no. 19, 2020.
- [253] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, “Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach,” *Computer Physics Communications*, vol. 167, no. 2, pp. 103–128, 2005.

- [254] F. Maseras and K. Morokuma, “IMOMM: A new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states,” *Journal of Computational Chemistry*, vol. 16, no. 9, pp. 1170–1179, 1995.
- [255] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu,” *The Journal of Chemical Physics*, vol. 132, no. 15, 2010.
- [256] C. Hartwigsen, S. Goedecker, and J. Hutter, “Relativistic separable dual-space Gaussian pseudopotentials from H to Rn,” *Physical Review B*, vol. 58, no. 7, p. 3641, 1998.

