

**Structure and Function of Glycoside Hydrolase
Enzymes Involved in Mannose and
N-acetylglucosamine Processing**

Alexandra Males

PhD

University of York

Chemistry

September 2019

Abstract

The reaction mechanisms of enzymes are often imposed by the stereochemical features of their substrates. By virtue of the proximity to the reactive anomeric carbon, the C2 substituent on mannosides and *N*-acetylglucosaminides plays a key role in the reactivity and conformation of these substrates on enzyme. These conformational and mechanistic features can be exploited to produce powerful enzyme inhibitors.

In this thesis, the conformational itineraries and reaction mechanisms of three enzyme classes were studied; bacterial α -mannosidases from CAZY families 47 and 125 and human O-GlcNAcase from family 84. The thesis will be based upon six research publications, two published reviews and unpublished data.

After a general introduction on carbohydrates, Chapter 2 (two papers) explores how family GH125 alleviates steric clashes of the *manno*-O2 through preferential distortion of the Michaelis complex to an ${}^{\circ}S_2$ conformation. Observation of a transition-state-mimicking mannoimidazole in $B_{2,5}$ conformation resulted in a proposed conformational itinerary of ${}^{\circ}S_2 \rightarrow [B_{2,5}]^{\ddagger} \rightarrow {}^1S_5$.

In contrast to GH125, GH47 α -mannosidases operate through a Southern hemisphere ${}^3S_1 \rightarrow [{}^3H_4]^{\ddagger} \rightarrow {}^1C_4$ itinerary. Chapter 3 (two papers) examines, at atomic resolution, the nature of GH47 α -mannosidase inhibitors to bind in a ring-flipped conformation provided that enzyme:inhibitor steric clashes do not form. A new class of mannose phosphorylases/mannosyltransferases, which synthesise the *Leishmania* β -1,2 linked storage polysaccharide mannogen, is described in an Appendix.

The key mechanistic feature of human O-GlcNAc hydrolase is the neighbouring-group participation of the *N*-acetyl group in catalysis. In Chapter 4 (one paper), a series of genetic constructs were produced to improve crystallisation, mimic *in vivo* phosphorylation and improve GlcNAcylated peptide complex formation. As a validated drug target for anti-tauopathy agents, Chapter 5 describes the dissection of a panel of new inhibitor classes with picomolar affinity against OGA due to human-specific active site interactions. The thesis is concluded by discussing the key results and possible approaches for future work.

Table of Contents

Abstract.....	- 3 -
Table of Contents.....	- 4 -
List of Figures	- 11 -
List of Tables.....	- 15 -
Preface.....	- 16 -
Acknowledgements	- 17 -
Declaration.....	- 18 -
Published Papers.....	- 19 -
CHAPTER 1: Introduction	- 20 -
1.1 Defining a carbohydrate.....	- 21 -
1.1.1 The importance of carbohydrates within the body	- 23 -
1.1.2 Carbohydrates can distort into different conformations.....	- 24 -
1.2 Carbohydrate active enzymes (CAZymes).....	- 28 -
1.2.1 Importance of CAZymes	- 28 -
1.2.2 The CAZy database (www.cazy.org).....	- 28 -
1.2.3 Glycoside hydrolases	- 28 -
1.2.3.1 The inverting mechanism	- 31 -
1.2.3.2 The retaining mechanism.....	- 31 -
1.2.3.3 The double displacement mechanism.....	- 32 -
1.3 Overcoming energy barriers	- 34 -
1.3.1 Substrates and enzymes bind in an activated state	- 34 -
1.3.2 The conformational itinerary of a carbohydrate active enzyme.....	- 35 -
1.3.3 The transition state	- 36 -
1.3.3.1 <i>Syn</i> - and <i>anti</i> -protonation	- 38 -

1.3.4	Techniques to visualise the conformational itinerary	- 38 -
1.3.4.1	Identifying the conformational itinerary using nuclear magnetic resonance.....	- 39 -
1.3.4.2	Identifying the conformational itinerary using crystallography	- 39 -
1.3.4.3	Identifying the conformational itinerary using quantum mechanical approaches	- 40 -
1.4	Inhibition of carbohydrate active enzymes.....	- 42 -
1.4.1	Rationale of design.....	- 42 -
1.4.2	Transition state mimicry	- 42 -
1.4.3	The synthesis of inhibitors for glycoside hydrolases.....	- 44 -
1.5	Comparison of the puckering of mannose and <i>N</i> -acetyl- β -D-glucosamine....	- 44 -
1.5.1	Mannose.....	- 45 -
1.5.2	<i>N</i> -acetyl- β -D-glucosamine.....	- 46 -
1.6	Outline of the thesis.....	- 46 -
SECTION 1: N-Glycan Processing		- 49 -
i.	N-linked glycan processing	- 50 -
ii.	N-glycan assembly and protein transfer.....	- 51 -
iii.	N-glycan trimming pathway in the ER and ER-associated degradation for quality control	- 52 -
iv.	N-glycan trimming pathway in the Golgi	- 54 -
CHAPTER 2: Family GH125 Exo- α -1,6-Mannosidases.....		- 56 -
2.1	Abstract.....	- 57 -
2.2	Family GH125 exo- α -1,6-mannosidases introduction.....	- 57 -
2.2.1	Pathogenic bacteria can utilise host glycans as a nutrient source	- 57 -
2.2.2	Discovery of α -1,6-mannosidases	- 59 -
2.2.3	Specificity.....	- 59 -
2.2.3.1	Activity assays	- 62 -

2.2.4	N-glycan processing model.....	- 62 -
2.2.5	Structure.....	- 65 -
2.2.6	Mechanism of catalysis.....	- 67 -
2.2.7	Inhibition of α -1,6-mannosidases.....	- 70 -
2.2.8	A summary of the results in this chapter.....	- 71 -
Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α -Mannosidases.....		- 72 -
Declaration.....		- 73 -
Distortion of Mannoimidazole Supports a $B_{2,5}$ Boat Transition State for the Family GH125 α -1,6-Mannosidase from <i>Clostridium perfringens</i>		- 78 -
Declaration.....		- 79 -
CHAPTER 3: Family GH47 <i>Exo</i> - α -1,2-Mannosidases.....		- 87 -
3.1	Abstract.....	- 88 -
3.2	Family GH47 <i>exo</i> - α -1,2-mannosidases (Golgi ManIA) introduction.....	- 88 -
3.2.1	Discovery of α -1,2-mannosidases.....	- 88 -
3.2.2	Inhibition of α -1,2-mannosidases.....	- 88 -
3.2.2.1	Inhibition by DMJ.....	- 89 -
3.2.2.2	Inhibition by kifunensine.....	- 90 -
3.2.3	Diseases associated with GH47 α -1,2-mannosidases.....	- 91 -
3.2.4	The structure of GH47 enzymes.....	- 92 -
3.2.5	Mechanism of catalysis.....	- 94 -
3.2.5.1	Proposition of catalytic residues.....	- 94 -
3.2.5.2	Catalysis is a calcium dependent mechanism.....	- 96 -
3.2.5.3	The conformational itinerary of GH47 α -1,2-mannosidase.....	- 96 -
3.2.6	Specificity of ER and Golgi α -1,2-mannosidases.....	- 98 -
3.2.7	<i>Caulobacter</i> K31 strain.....	- 100 -

3.2.8	A summary of the results in this chapter	- 100 -
	Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Structural Approach.....	- 101 -
	Declaration.....	- 102 -
	Conformational Behaviour of Mannuronic Acid Based Azasugars	- 109 -
	Declaration.....	- 110 -
	Additional material.....	- 119 -
3.3	Introduction.....	- 119 -
3.4	Materials and Methods.....	- 121 -
3.4.1	Expression and purification of <i>CkGH47_{E427A}</i>	- 121 -
3.4.2	Crystallisation screening of <i>CkGH47_{E427A}</i> and soaking experiments with mannuronic acid derivatives	- 122 -
3.4.3	Observation of the change in thermal denaturation.....	- 122 -
3.5	Results.....	- 123 -
3.5.1	Crystallisation of <i>CkGH47_{E427A}</i> in HEPES buffer as an alternative to Bis-Tris buffer reduced the likelihood of unwanted ligands in the active site	- 123 -
3.5.2	Crystal structure of <i>CkGH47_{E427A}</i> in complex with 3 showed binding in the reducing +1 subsite.....	- 123 -
3.5.3	The binding affinity of mannuronic acid derivatives to <i>CkGH47</i> was too weak to observe by ITC and TSA	- 125 -
3.6	Discussion.....	- 126 -
	SECTION 2: O-Glycan Processing Enzymes.....	- 127 -
i.	Uridine diphosphate <i>N</i> -acetylglucosamine.....	- 128 -
ii.	Hexosamine biosynthetic pathway	- 129 -
iii.	Fluctuation between GlcNAc and glucose levels.....	- 131 -
	CHAPTER 4: Family GH84 O-GlcNAc Hydrolase	- 132 -
4.1	Abstract.....	- 133 -

4.2	Family GH84 structure and function introduction.....	133 -
4.2.1	O-GlcNAc cycling enzymes.....	133 -
4.2.2	Discovery of O-GlcNAc hydrolase	134 -
4.2.2.1	Activity of OGA.....	135 -
4.2.2.2	The structure of OGA.....	136 -
4.2.2.3	Mechanism of catalysis	142 -
4.2.2.4	Substrate recognition.....	143 -
4.2.3	Summary of the results in this chapter	145 -
	O-GlcNAc Hydrolase Binds S-GlcNAcylated Peptides	147 -
4.3	Obtaining a peptide complex structure.....	148 -
4.4	Experimental methodology.....	148 -
4.5	Results.....	148 -
4.5.1	CKII was present in one active site of a homodimer	148 -
4.6	Discussion.....	150 -
	Structural Studies of a Surface Entropy Reduction Mutant of O-GlcNAcase	151 -
	Declaration.....	152 -
	Residue Ser364 of <i>HsOGA</i> is Phosphorylated by Casein Kinase II	162 -
	Declaration.....	163 -
4.7	Introduction to phosphomimetics	164 -
4.8	Experimental methodology.....	165 -
4.8.1	Trypsin digest of phosphorylated <i>HsOGA</i> and LC-MS/MS detection..	165 -
4.8.2	Circular dichroism of <i>HsOGA</i> constructs.....	165 -
4.8.3	Kinetic assay of <i>HsOGA</i> constructs and <i>pNP-GlcNAc</i>	165 -
4.8.4	<i>HsOGA</i> _{S364D} plasmid mutation, expression, purification, crystallisation and data collection.....	166 -
4.9	Results.....	166 -

4.9.1	Casein kinase II phosphorylates OGA at residue Ser364	- 166 -
4.9.2	The phosphomimetic mutation of <i>Hs</i> OGA did not alter the secondary structure of the protein.....	- 169 -
4.9.3	<i>Hs</i> OGA _{S364D} mutant has the same kinetic activity as WT <i>Hs</i> OGA _{split1}	- 170 -
4.10	Discussion.....	- 171 -
CHAPTER 5: Inhibition of O-GlcNAc Hydrolase.....		- 173 -
5.1	Abstract.....	- 174 -
5.2	Family GH84 inhibition introduction.....	- 174 -
5.2.1	The importance of OGA inhibition	- 174 -
5.2.1.1	O-GlcNAcylation and phosphorylation of tau.....	- 174 -
5.2.1.2	Involvement of crosstalk between O-GlcNAcylation and phosphorylation in Alzheimer's disease.....	- 175 -
5.2.2	The first inhibitors of OGA.....	- 177 -
5.2.2.1	Thiamet-G reduces the pathogenicity of tau	- 181 -
5.2.3	Interactions between the residues of <i>Hs</i> OGA and different inhibitors.-	- 182 -
5.2.4	Conformational itinerary of GH84 versus GH20 enzymes	- 183 -
5.2.5	Recent OGA inhibitors.....	- 185 -
5.2.6	Inhibitors in clinical trials	- 186 -
Crystal Structure Guided Human O-GlcNAcase Iminocyclitol and Bicyclic Inhibitors Access Sub-Nanomolar Potency		- 189 -
5.3	Abstract.....	- 190 -
5.4	Introduction.....	- 190 -
5.5	Experimental methodology.....	- 192 -
5.6	Results.....	- 192 -
5.6.1	Modification of the pyrrolidine-based iminocyclitols reveals nanomolar affinity inhibitors.....	- 192 -

5.6.2	1 causes a rearrangement of the loop containing Trp679 due to pi stacking interactions.....	- 193 -
5.6.3	A triazole-based library reveals picomolar affinity inhibitors.....	- 195 -
5.6.4	Pi stacking interactions with active site residues increase the potency of the inhibitors	- 197 -
5.6.5	Bicyclic inhibitors sit in the <i>Hs</i> OGA specific cleft.....	- 200 -
5.7	Discussion.....	- 202 -
CHAPTER 6: Conclusions and Perspectives		- 203 -
Appendix		- 212 -
Abbreviations		- 289 -
References		- 291 -

List of Figures

Figure 1.1. Ring atoms of mono- and disaccharides.....	- 22 -
Figure 1.2. Standardised symbolic representation for 150 monosaccharides.	- 22 -
Figure 1.3. Roles of glycans in the cell.	- 23 -
Figure 1.4. Fischer and Haworth projections of glyceraldehyde and glucose enantiomers.	- 25 -
Figure 1.5. Mutarotation of D-glucose in β and α configurations.	- 26 -
Figure 1.6. Conformations of pyranose rings with nomenclature.....	- 27 -
Figure 1.7. <i>Exo</i> - and <i>endo</i> -acting glycoside hydrolases.	- 30 -
Figure 1.8. Inverting mechanism.....	- 31 -
Figure 1.9. Retaining mechanism.	- 32 -
Figure 1.10. Neighbouring group participation by double displacement mechanism. -	- 33 -
Figure 1.11. Mechanism for GH47 α -1,2-mannosidases.....	- 34 -
Figure 1.12. Projection styles for visualising the conformations of carbohydrates described in Figure 1.6.	- 36 -
Figure 1.13. The oxocarbenium ion transition state conformation.	- 37 -
Figure 1.14. Space quadrants of the -1 subsite around a glycosidic ring.	- 38 -
Figure 1.15. Conformational free energy landscape of α -mannose.....	- 41 -
Figure 1.16. Structures of common glycosidase inhibitors.....	- 44 -
Figure 1.17. Conformational changes for mannose and GlcNAc.	- 45 -
Figure 1.i. N-linked glycan structures.	- 51 -
Figure 1.ii. N-glycan assembly occurs in the cytoplasm and the ER lumen.....	- 52 -

Figure 1.iii. N-glycan trimming and protein folding recognition in the ER.	- 53 -
Figure 1.iv. N-glycan processing in the Golgi apparatus.	- 55 -
Figure 2.1. Schematic of α -1,6-mannobiose.	- 59 -
Figure 2.2. Structure of 2,4-dinitrophenyl α -D-mannopyranoside.	- 60 -
Figure 2.3. Specificity of GH125 was determined by testing the cleavage of disaccharides, trisaccharides and high-mannose glycans.	- 61 -
Figure 2.4. Coupled fluorogenic assay for GH125.	- 62 -
Figure 2.5. Complete N-glycan degradation pathway in <i>Streptococcus pneumoniae</i>	- 64 -
Figure 2.6. Overall $(\alpha/\alpha)_6$ -barrel fold of GH125 and catalytic subsites.	- 66 -
Figure 2.7. Evidence for the catalytic mechanism of GH125.	- 69 -
Figure 2.8. Structure of thiomannobiose.	- 71 -
Figure 3.1. Inhibitors of α -1,2-mannosidases.	- 89 -
Figure 3.2. Secondary structure of the $(\alpha/\alpha)_7$ barrel of GH47 mannosidases.	- 92 -
Figure 3.3. Disulfide bond only stabilises the secondary structure in eukaryotes.	- 93 -
Figure 3.4. Catalytic mechanism of GH47.	- 95 -
Figure 3.5. Capturing the conformational itinerary of GH47 enzymes.	- 97 -
Figure 3.6. Conformational free energy landscape (FEL) of <i>Ck</i> GH47 and the α -mannosyl moiety of different inhibitors in the -1 subsite.	- 98 -
Figure 3.7. The base of the active site determines substrate specificity.	- 99 -
Figure 3.8. Electronegative potential of the neighbouring residues in the active site surrounding 2.	- 120 -
Figure 3.9. Steric hindrance of residues in the active site prevents binding.	- 121 -
Figure 3.10. Hydrogen bonding interactions of 3 in complex with <i>Ck</i> GH47 _{E427A}	- 124 -

Figure 3.11. Fluorescence curves show no shift in melting temperature upon addition of mannuronic acid derivatives.....	125 -
Figure 2.i. O-GlcNAcylation of a serine residue.....	128 -
Figure 2.ii. The hexosamine biosynthetic pathway.....	131 -
Figure 4.1. Reciprocal relationship between phosphorylation and GlcNAcylation....	134 -
Figure 4.2. Domain structure of O-GlcNAc hydrolase long and short isoforms.....	135 -
Figure 4.3. Domain structure of <i>HsOGA</i> and bacterial homologues.....	137 -
Figure 4.4. Construct design for the expression of soluble <i>HsOGA</i>	138 -
Figure 4.5. The variable conformations of the C-terminal domain.....	139 -
Figure 4.6. <i>HsOGA</i> dimer formation.....	140 -
Figure 4.7. Structures of bacterial acetyltransferases.....	142 -
Figure 4.8. Catalytic mechanism of OGA.....	143 -
Figure 4.9. Overlay of <i>HsOGA</i> in complex with various peptides.....	145 -
Figure 4.10. Structure of CKII GlcNAcylated peptide in complex with <i>HsOGA</i>	149 -
Figure 4.11. A phosphorylated serine residue is mimicked by aspartate.....	164 -
Figure 4.12. The hydrolysis of <i>pNP-GlcNAc</i> to 4-nitrophenol and GlcNAc.....	166 -
Figure 4.13. Comparison of the relative peak heights of the phosphorylated and unphosphorylated OGA fragment 3.....	167 -
Figure 4.14. A comparison of the ions present after fragmentation of unphosphorylated and phosphorylated samples.....	168 -
Figure 4.15. The secondary structure of the <i>HsOGA</i> phosphomimetic mutant was similar to the WT.....	170 -
Figure 4.16. Michaelis-Menten curves for <i>HsOGA</i> mutant kinetic assays.....	171 -

Figure 5.1. Tau undergoes phosphorylation and O-GlcNAcylation.	175 -
Figure 5.2. Hyperphosphorylation of tau causes its dissociation from microtubules and aggregation.	176 -
Figure 5.3. The evolution of O-GlcNAc hydrolase inhibitors.	178 -
Figure 5.4. Thiamet-G increases O-GlcNAc levels and reduces phosphorylation levels within the hippocampus.	181 -
Figure 5.5. Ligand complexes in the active site of <i>HsOGA</i>	183 -
Figure 5.6. Inhibitor evidence for the conformational itinerary of GH84 enzymes.	184 -
Figure 5.7. Recent potent and selective OGA inhibitors.	186 -
Figure 5.8. First-in-human OGA inhibitor MK-8719.	187 -
Figure 5.9. Influential inhibitors to the design of a novel triazole series.	191 -
Figure 5.10. An initial test series of triazole-based compounds.	193 -
Figure 5.11. Structure of 1 in complex <i>HsOGA</i>	194 -
Figure 5.12. Starting substrate used to build the compound library.	195 -
Figure 5.13. Triazole compound library.	197 -
Figure 5.14. Structure of <i>HsOGA</i> in complex with compound 3.	199 -
Figure 5.15. Structure of <i>HsOGA</i> in complex with compound 4.	200 -
Figure 5.16. The structures of a bicyclic series of OGA inhibitors.	201 -
Figure 5.17. Inhibitor interactions with the unique active site cleft of <i>HsOGA</i>	202 -
Figure 6.1. Fusion construct design.	208 -
Figure 6.2. Sequence alignment of mammalian HAT constructs.	210 -
Figure 6.3. Production of different mammalian HAT constructs.	211 -

List of Tables

Table 3.1. Data and refinement statistics for <i>CkGH47</i> in complex with 3.	- 125 -
Table 4.1. Trypsin digested fragments of OGA that contain a potential site of phosphorylation.	- 167 -
Table 4.2. Fragment 3 mass/charge table for phosphorylation modifications.	- 167 -
Table 4.3. Calculated B ion type masses.	- 168 -
Table 4.4. Data collection and refinement statistics for <i>HsOGA_{S364D}</i>	- 169 -
Table 4.5. MM kinetic assay results for <i>HsOGA_{split1}</i> and <i>HsOGA_{S364D}</i>	- 171 -
Table 5.1. Data collection and refinement statistics of <i>HsOGA</i> with inhibitors 1 and 3.	- 198 -

Preface

The growth in the knowledge of the glycome vastly increased in the 20th century to advance the field of chemical biology. The variety of combinations that can be created from monosaccharide building blocks lead to the diverse functions that carbohydrates possess in the cell and organism. The importance of understanding the chemical structures of carbohydrates was a catalyst to acquiring the complex geometries and distorted substrate conformation analysis of the enzymes performing the reactions.

Glycoside hydrolases (GHs) cleave the glycosidic bond between a carbohydrate and another carbohydrate or non-carbohydrate molecule. They are classified into different families according to amino-acid sequence. This thesis focuses on three CAZy families, where activity and mechanisms reflect their C2-substituents. GH47 and GH125 enzymes are α -mannosidases involved in the hydrolysis of α -1,2-linked and α -1,6-linked mannosaccharides, respectively. CAZy family GH84 enzymes cleave *N*-acetyl β -D-glucosamine (GlcNAc) from serine or threonine residues of proteins. Although each enzyme family catalyses a different reaction with different mechanisms and substrates, mannose and GlcNAc have a common feature.

Dominating the reactivity of these enzymes, the substituent at the C2 position shows a preference for the alignment of the anomeric oxygen in an axial position and limits the number of accessible distorted conformations. In the case of *manno*-active enzymes, the axial disposition of O2 in the preferred ⁴C₁ conformation hinders nucleophilic substitution and so, on-enzyme, mannose favours skew boat and boat conformations that render O2 equatorial. In contrast, the C2 *N*-acetyl substituent position of GlcNAc allows nucleophilic attack at C1 via a neighbouring group participation mechanism.

In this thesis, studying the conformational itinerary of these enzymes and conducting inhibition and activity studies using X-ray crystallography and biophysical techniques provides information on substrate specificity and will inform the design of mechanism-based inhibitors as probes with the potential for therapeutic applications.

Acknowledgements

Principally, I would like to show my immense gratitude to my supervisor, Prof. Gideon J. Davies, for the immeasurable support and encouragement he has given me throughout my PhD by his kindness, clever ideas, jokes and belief. My thesis wouldn't be what it is today without the "side-projects" that were only supposed to take me a few months. I would also like to thank Dr. Martin Fascione, my co-supervisor, and Dr. Alison Parkin, my internal panel member, for guiding my TAP meetings and providing inspiration.

Without my collaborators, the projects would not have been accomplishable. I would like to thank Prof. David Vocadlo for hosting me in SFU, Matt, Johannes, Hong Yee, Dustin and Roger for sending compounds and showing me around Vancouver. Prof. Spencer Williams, Prof. Carme Rovira, Prof. Hermen Overkleeft, Marta Artola and Dr. Jeroen Codee have all influenced my PhD by providing direction and guidance to the projects.

The PhD office was a friendly and fun place to work, made this way by past PhD students: Drs Claire, Imogen, Adam, Lukasz, Herman and Bryony; current students: Rhianna, Eleni, Paul, Mark, Alex, Bas, Mihaela, Tom, Mateusz, Catherine, Wendy, Ben; and MAF students: Richard, Emily, Harriet and Robin.

I would like to express my gratitude to all the members of the GJD group and YSBL staff. Christian, Liang, Nick and Zach were always available to answer any questions and show me techniques, especially John with the ITC. Thank you to Wendy for sharing a bench with me. Simon, Louise and Jules do an amazing job in keeping the lab running and thank you to Johan and Sam for collecting X-ray data. Kate was always available for a de-stressing chat and Jane was a huge help in finding me a place to live in Lisbon. Lastly, I would like to thank Tim K for improving my badminton skills.

The love and constant emotional support from my parents and family has been incredible. Thank you for understanding my decision to go to York, helping me to settle and coming to visit.

"I used to dream about love of the tall dark mysterious kind." Thank you to Matt who kept me sane, distracted me and encouraged me to work throughout the PhD and write up period!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author, except where it is stated. A declaration, prior to the results section of each chapter, details the contribution from all of the co-authors of the attached journal article. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Published Papers

Chapter 2

Alonso-Gil, S., Males, A., Fernandes, P. Z., Williams, S. J., Davies, G. J. and Rovira, C. (2017). Computational design of experiment unveils the conformational reaction coordinate of GH125 α -mannosidases. *J. Am. Chem. Soc.* *139*, 1085-1088. **Pg 73.**

Males, A., Speciale, G., Williams, S. J. and Davies, G. J. (2019). Distortion of mannoimidazole supports a $B_{2,5}$ boat transition state for the family GH125 α -1,6-mannosidase from *Clostridium perfringens*. *Org. Biomol. Chem.* *17*, 7863-7869. **Pg 79.**

Chapter 3

Males, A., Raich, L., Williams, S. J., Rovira, C. and Davies, G. J. (2017). Conformational analysis of the mannosidase inhibitor kifunensine: a quantum mechanical and structural approach. *ChemBioChem.* *18*, 1496-1501. **Pg 102.**

van Rijssel, E. R., Janssen, A. P. A., Males, A., Davies, G. J., van der Marel, G. A., Overkleeft, H. S., and Codée, J. D. C. (2017). Conformational behaviour of azasugars based on mannuronic acid. *ChemBioChem.* *18*, 1297-1304. **Pg 110.**

Chapter 4

Males, A., and Davies, G. J. (2019). Structural studies of a surface-entropy reduction mutant of O-GlcNAcase. *Acta Crystallogr D Struct Biol.* *75*, 70-78. **Pg 154.**

Appendix

Sernee, M. F., Ralton, J. E., Nero, T. L., Sobala, L. F., Kloehn, J., Vieira-Lara, M. A., Cobbold, S. A., Stanton, L., Pires, D. E. V., Hanssen, E., Males, A., Ward, T., Bastidas, L. M., van der Peet, P. L., Parker, M. W., Ascher, D. B., Williams, S. J., Davies, G. J., McConville, M. J. (2019). A family of dual-activity glycosyltransferase-phosphorylases mediates mannogen turnover and virulence in *Leishmania* parasites. *Cell Host Microbe.* *26*, 385-399.e9. **Pg 241.**

King, D. T., Males, A., Davies, G. J., Vocadlo, D. J. (2019). Molecular mechanisms regulating O-GlcNAc processing enzymes. *Curr Opin Chem Biol.* *53*, 131-144. **Pg 266.**

Males, A., Rovira, C., Davies, G. J., Williams, S. J. (2019). Mannosidase mechanism: At the intersection of conformation and catalysis. *Curr Opin Struct Biol.* (Accepted). **Pg 288.**

CHAPTER 1: Introduction

1.1 Defining a carbohydrate

A carbohydrate is defined as a molecule composed of a base of carbon, hydrogen and oxygen atoms that form multi-hydroxylated ketones and aldehydes. Monosaccharides are defined as $(C \cdot H_2O)_n$, where $n \geq 3$ is the number of atoms in the ring, Figure 1.1; however, not all monosaccharides fit this definition. If a monosaccharide has more than 4 carbons (tetrose), then it exists as a stable cyclic structure in aqueous solution due to the intramolecular reactivity of the aldehyde and alcohol groups. Interconversion of a carbohydrate between its linear and cyclic form has a relaxation rate of $\sim 100 \mu s$, Figure 1.5 (M. Kuttel, 2011). The most stable cyclic alkanes are five membered rings (furanose), which consist of four carbon atoms and one oxygen atom, and six membered rings (pyranose), which consist of five carbon atoms and one oxygen atom.

There are over 150 naturally occurring monosaccharide building blocks. Glycosidic bonds join multiple monosaccharides together to form disaccharides, oligosaccharides (2-10 monosaccharides bonded together) that can be linear or branched in structure and polysaccharides (>10 monosaccharides bonded together); the boundaries for each term are not rigorously defined. The nomenclature for a glycosidic bond is determined by the number of the carbon atom on the ring that donates or receives the hydroxyl, for example, the α -1,4 disaccharide maltose is formed by a glycosidic bond between the C1 hydroxyl of one glucose molecule and the C4 hydroxyl of C4 of another, Figure 1.1. Glycoconjugate is a term used to describe a biomolecule covalently bonded to a carbohydrate. Examples of these forms are glycoproteins/proteoglycans (carbohydrates and proteins) and glycolipids/lipopolysaccharides (carbohydrates and lipids).

The complexity of drawing oligosaccharides and polysaccharides in publications led to the standardisation of their representation. Chemically drawn and IUPAC (International Union of Pure and Applied Chemistry) abbreviations (Chester, 1998; McNaught, 1997; Sharon, 1986) became a series of coloured symbols, initially designed by Kornfield and colleagues to represent vertebrate glycans (polysaccharides) (Kornfeld et al., 1978). Implementation of this system for universal use further extended the number of glycans (Pérez, 2018; Varki, 1999; Varki and Marth, 1995; Varki et al., 2015) Figure 1.2.

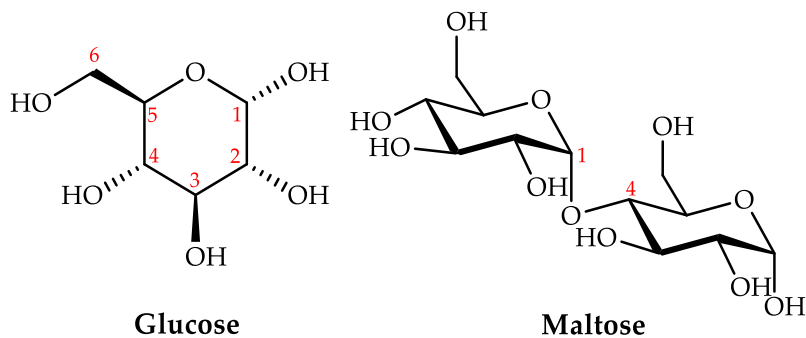


Figure 1.1. Ring atoms of mono- and disaccharides.

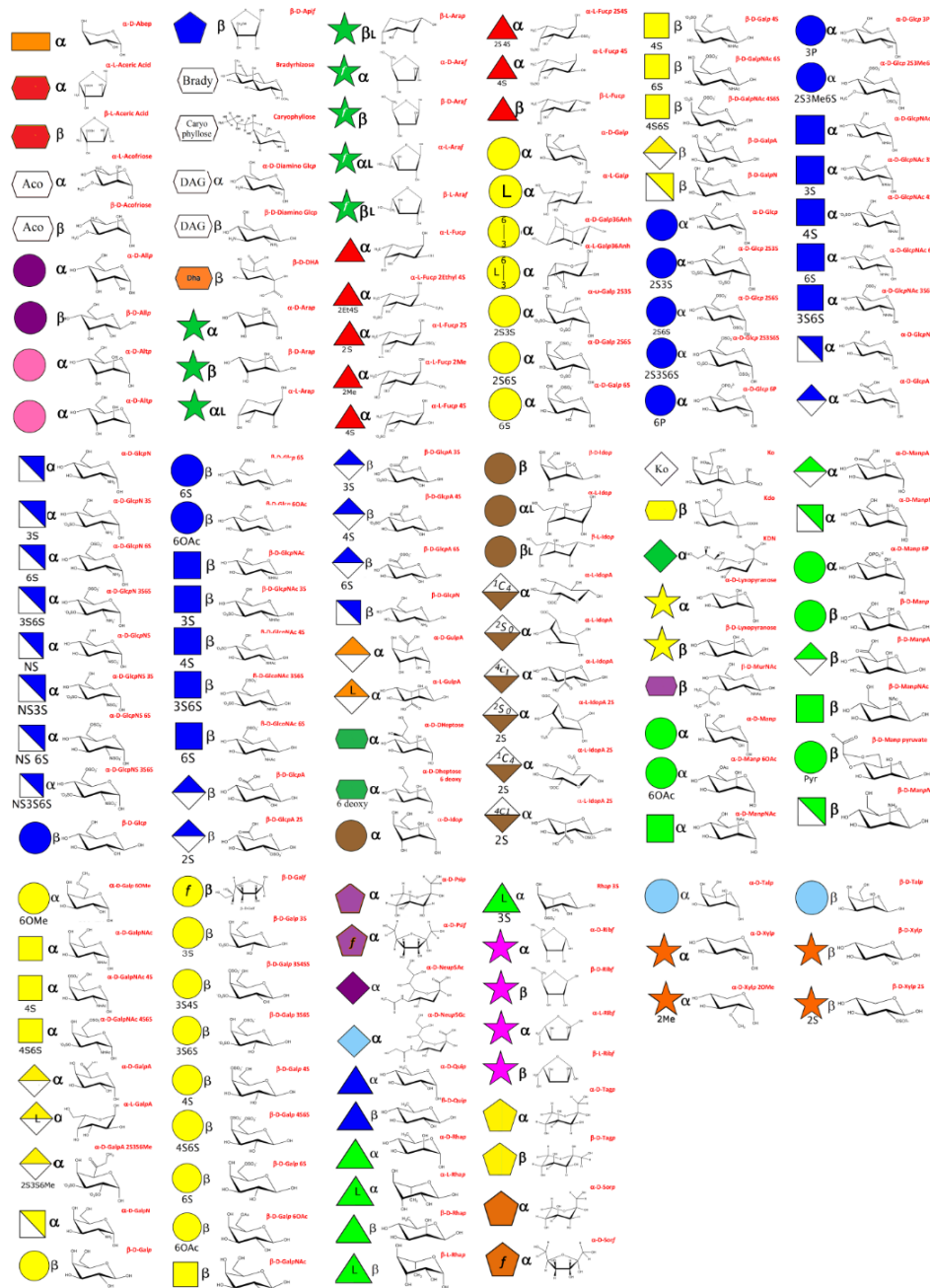


Figure 1.2. Standardised symbolic representation for 150 monosaccharides. Symbolic representation is followed by the structure and name of the monosaccharide. Figure adapted from glycopedia.eu (Pérez, 2018).

1.1.1 The importance of carbohydrates within the body

The glycome is a term used to describe the complete collection of glycans within a cell that cover a multitude of functions both in and on the cell surface (reviewed in Gastaldello et al., 2018; Varki, 2017), Figure 1.3. Glucose is important in glycolysis by providing adenosine triphosphate (ATP) to metabolic reactions. Humans ingest carbohydrates for use as energy in the form of disaccharides, sucrose and lactose, and store excess carbohydrates by synthesising polysaccharides (recently shown in *Leishmania* (Sernee et al., 2019)- see appendix pg 241), glycogen and starch.

In the endoplasmic reticulum, the glycans present on a protein indicate if the protein is correctly folded (Ellgaard et al., 1999; Hammond and Helenius, 1995). Glycosylation modifications can act as degradation signals for example a protein linked with mannose-6-phosphate will be targeted by lysosomal degradation enzymes (Natowicz et al., 1979). Glycans on the cell surface control cell-to-cell signalling and are important in the immune response (Galli-Stampino et al., 1997; Haurum et al., 1994). The migration pattern of the immune cell, the affinity for receptors and the discrimination between “self” and “non-self” are influenced by glycans (Crocker, 2005; Varki and Angata, 2005). Virus virulence and immune evasion are other important mechanisms controlled by glycans (Olden et al., 1985).

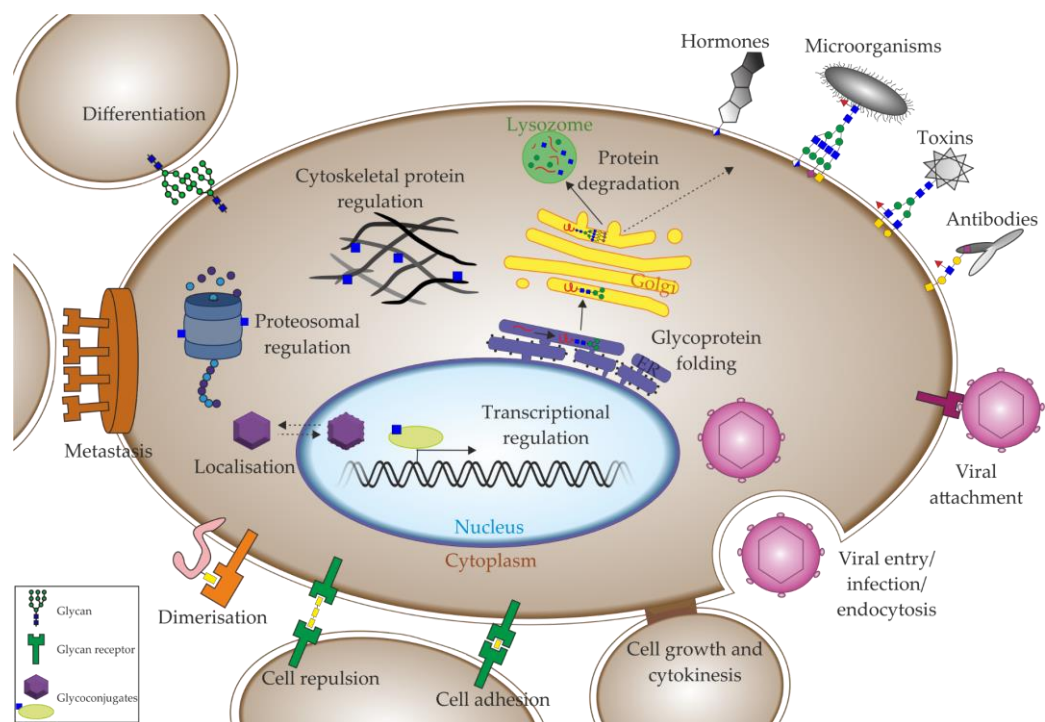


Figure 1.3. Roles of glycans in the cell.

1.1.2 Carbohydrates can distort into different conformations

Glyceraldehyde, the smallest monosaccharide, is present as two enantiomers, D- and L-, due to the presence of a chiral carbon (Rao, 1998), Figure 1.4a. Abundant naturally occurring monosaccharides are mostly present as the D-enantiomer. The Fischer projection was developed to allow unambiguous determination of chirality, since the molecule is drawn as a vertical linear line around the chiral centre (Fischer, 1891), Figure 1.4a and b. Named after Sir Norman Haworth, the three-dimensional Haworth representation details the preferential cyclic form of a carbohydrate, Figure 1.4a and b. The pyranose carbon atoms lie flat along the same plane perpendicular to the plane of the paper (2008).

The most abundant enantiomer of glucose is D-glucose; it is present as 4 different cyclic isomers: α -D-glucopyranose, β -D-glucopyranose, α -D-glucofuranose and β -D-glucofuranose, Figure 1.4b. The furanose isomers are present in solution in negligible amounts compared to the pyranose isomers. The structure of glucose was elucidated by Fischer in 1890, who determined that the cause of the mutarotation (the transition from α - to β -) was governed by hydration (Bijvoet et al., 1951; Fischer, 1890; Lichtenthaler, 1992). The assignment of the diastereomeric conformers α - and β - depend on the orientation of the hydroxyl group on C1 and CH₂OH on C5. For the α - arrangement, the two groups are in a *trans* orientation, opposite sides of the plane, and for the β - isomer, the two groups are in a *cis* orientation. ~36 % of glucose in water at 25 °C is present in the α - diastereomer and ~64% in the β - diastereomer, Figure 1.5. Comparably, ~67% of mannose is present in the α - form and ~33% in the β - form, (Angyal, 1969, 1984; Polacek et al., 2002; Stenger et al., 2002). This ratio was calculated by observing the average optical rotation of a sugar solution as each anomer will rotate plane polarised light by a different degree (reviewed in Fletcher, 1940; Pigman and Isbell, 1968).

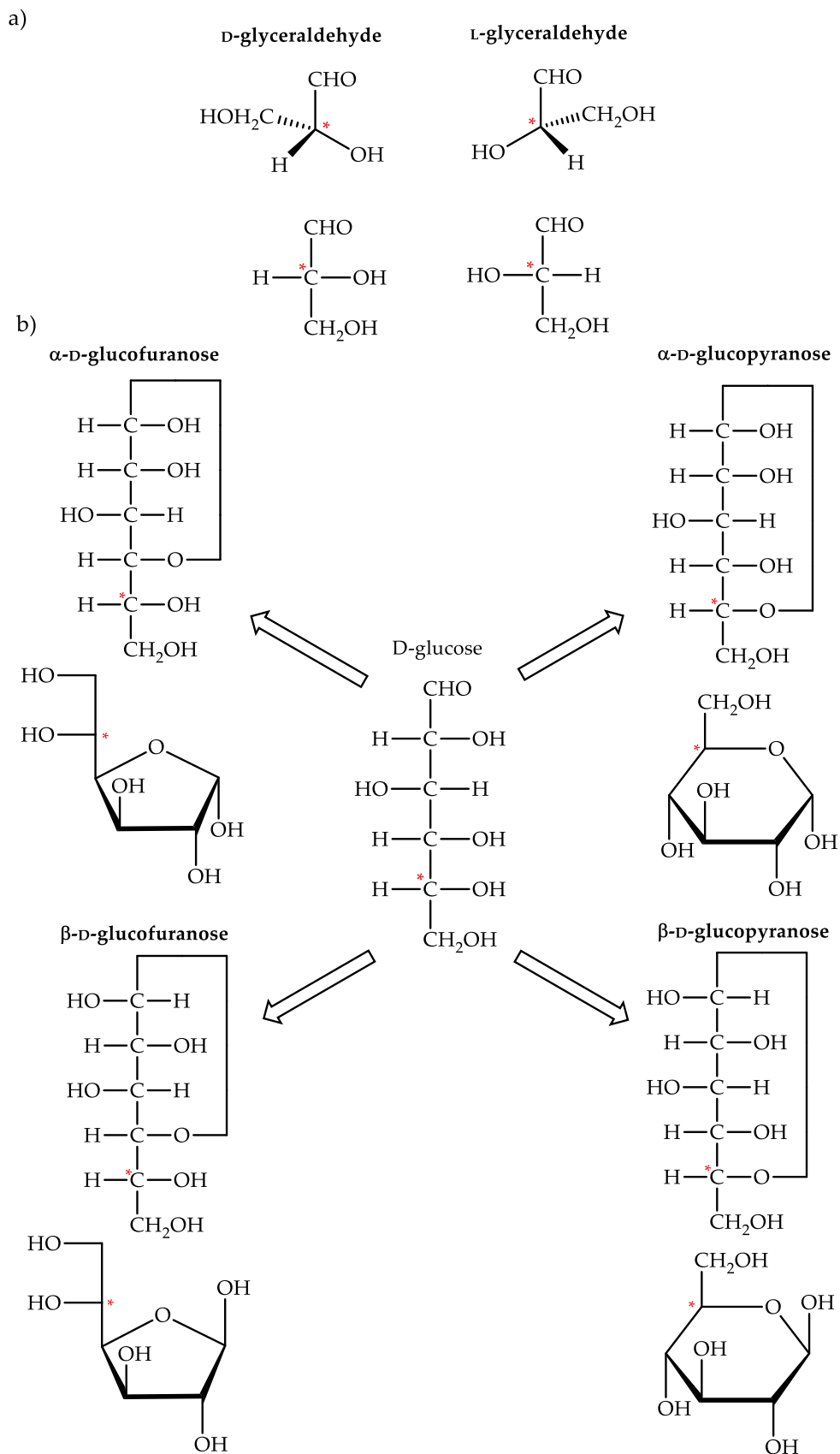


Figure 1.4. Fischer and Haworth projections of glyceraldehyde and glucose enantiomers. The red asterisk indicates the position of the chiral atom. a) Fischer projection of D- and L-glyceraldehyde. b) Fischer and Haworth projections of α- and β-anomers of D-glucopyranose and D-glucofuranose.

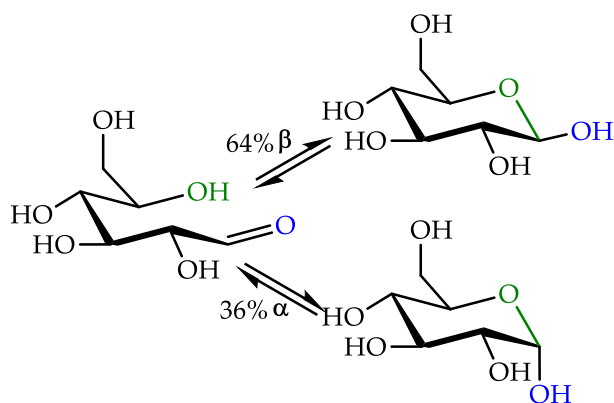


Figure 1.5. Mutarotation of D-glucose in β and α configurations.

Physical forces placed upon the bonding and non-bonding electrons of carbohydrates from molecular interactions with its environment (aqueous or other solvent (Kirschner and Woods, 2001)) leads to the puckering of cyclic carbohydrates into 38 different possible conformations. There are two chair (*C*), six boat (*B*), six skew-boat (*S*), twelve half-chair (*H*) and twelve envelope (*E*) conformations possible (Rao, 1998; Stoddart, 1971), Figure 1.6. The nomenclature for the different conformations is determined by the atoms that lie above and below the plane (composed of 4 or more of the atoms) (1980); the single letter nomenclature was defined by Reeves where *C*₁ has become ⁴*C*₁ (Reeves, 1949). A ⁴*C*₁ conformation states that C₄ is above the plane and C₁ (must be exoplanar) is below the plane; a ¹*C*₄ conformation is a mirror image to ⁴*C*₁. ⁴*C*₁ is the lowest energy conformation and was the first conformation to be observed in solution (Reeves, 1949). In a ring formation, the sp³ carbon atoms prefer torsion angles of ±60°, tetrahedral bonding geometry angles for a stable chair conformation. The preference for ⁴*C*₁ and ¹*C*₄ is due to the anomeric effect (preference for an axial orientation of the hydroxyl on the anomeric carbon), dipoles, intramolecular hydrogen bonding and 1,3-diaxial interactions (between axial substituents). Representation of this phenomenon was devised by Sponslor and Dore modifying the Haworth projection to apply tetrahedral geometry to all of the carbon atoms (Sponslor and Dore, 1926).

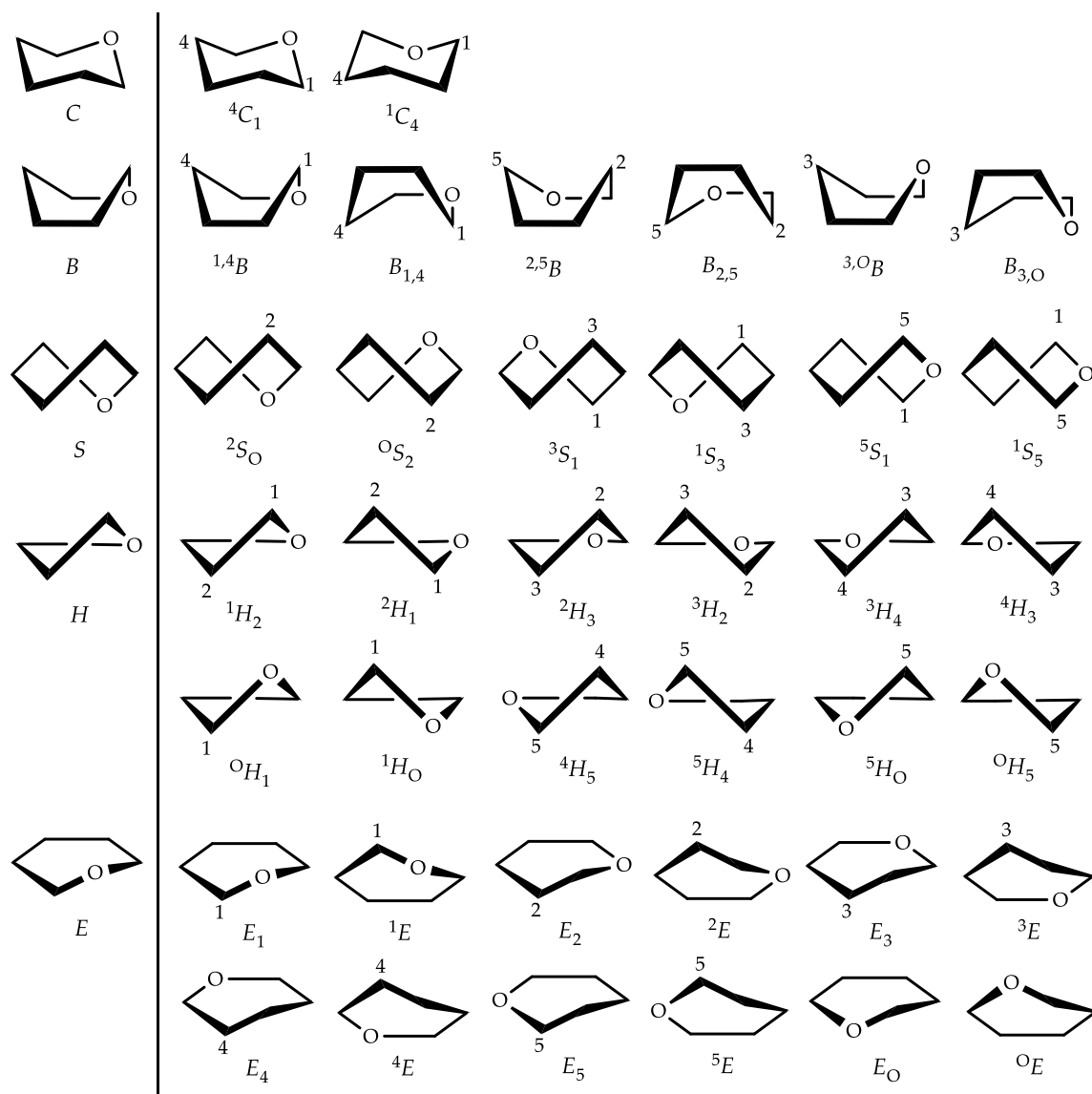


Figure 1.6. Conformations of pyranose rings with nomenclature. Chair conformations (C): two atoms on opposite sides of the ring are out of the plane, one up and one down; Boat conformations (B): two atoms on opposite sides of the ring are out of the plane, either both up or both down; Half-chair conformations (H): four consecutive atoms are coplanar; Envelope conformations (E): one atom is out of the plane; Skew-boat conformations (S): three consecutive atoms are coplanar.

In a solution at room temperature, a carbohydrate is in an equilibrium between different conformations. Notably, ${}^4C_1 \leftrightarrow {}^1C_4$ has a ~ 100 ns interconversion relaxation rate (Polacek et al., 2002; Stenger et al., 2002). The interconverting conformations depend on the solvent used (Kirschner and Woods, 2001), the presence of cations and the energy barriers of the conformations. In the active site of an enzyme, the conformations are limited by interactions with protein residues and by the presence of any ions. Divalent metal ions are

present in some active sites to assist in conformation distortion (Numao et al., 2003; Petersen et al., 2010; Zhong et al., 2008; Zhu et al., 2010).

1.2 Carbohydrate active enzymes (CAZymes)

1.2.1 Importance of CAZymes

~1-5% of the translated human genome encodes Carbohydrate Active enZymes (CAZyme) that are involved in oligosaccharide production, removal and recognition (Varki and Marth, 1995). The numerous functions of carbohydrates in the cell, as described in section 1.1.1, leads to a multitude of applications of CAZymes in health, agriculture and industrial processes. The discovery of novel enzymes is, therefore, of wide interest.

1.2.2 The CAZy database (www.cazy.org)

The Carbohydrate-Active enZymes (CAZy) database classifies enzymes into families based on their amino-acid sequence (Lombard et al., 2014). The enzymes, carbohydrate-binding modules and functional domains recorded on the database catalyse either the degradation, biosynthesis or modification of glycosidic bonds. An enzyme is categorised into a class: glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide lyase, carbohydrate esterase or auxiliary activity; and then into a family. Subdividing enzymes into families facilitates the identification of structural features, association of evolutionary relationships and can reveal mechanistic information (Henrissat, 1991; Henrissat and Bairoch, 1993). An enzyme can be further classified into a clan if the secondary structural fold resembles members of different families and is better conserved than sequence alignment (Henrissat and Bairoch, 1996). Over 350 families have been discovered and deposited on CAZy, with over 100,000 non-redundant entries.

1.2.3 Glycoside hydrolases

Glycoside hydrolases (EC 3.2.1.X) catalyse the hydrolysis of glycosidic bonds between carbohydrate moieties (Davies and Henrissat, 1995; Henrissat and Davies, 1997). As of September 2019, the CAZy database subdivides glycoside hydrolases into 165 different families depending on their sequence (Henrissat, 1991; Lombard et al., 2014). Examined in this report, are enzymes from family GH125 exo- α -1,6-mannosidase, GH47 α -1,2-

mannosidase and GH84 O-GlcNAcase. They work in conjunction with glycosyl transferases to break and create bonds. GHs can hydrolyse C-O glycosidic bonds (O-glycosylation) of a saccharide or between a carbohydrate and a protein residue, C-N glycosidic bonds (N-glycosylation) where a complex branched sugar can be cleaved from an asparagine residue by peptide:N-glycosidase F, an amidase (Norris et al., 1994), and a minority can hydrolyse C-S glycosidic bonds (S-glycosylation) (Macauley et al., 2005a).

Over 50% of the structural folds for a GH are triose phosphate isomerase (TIM) barrels, Figure 1.7a-c (Rigden et al., 2003). This conserved fold consists of eight α -helices and eight β -strands (β/α)₈. Variations on the secondary structure of the barrel fold are observed in GH47, the first to be identified to have an (α/α)₇ fold (Vallée et al., 2000a), and in GH125, which has an (α/α)₆ fold (Gregg et al., 2011). A minority of enzymes have a β -barrel, for example endoglucanase V from *Humicola insolens* (Davies et al., 1993), or a β -sandwich structure, for example, cellobiohydrolase I from *Trichoderma reesei* (Divne et al., 1994).

The differentiation of an active site into subsites demarcates monosaccharide binding (Davies et al., 1997), Figure 1.7a. Cleavage of a polysaccharide can occur in the middle of the chain by an *endo*-acting glycosidase or at the end by an *exo*-acting glycosidase. Usually, the non-reducing sugar binds in the -1 subsite and the reducing sugar binds in the +1 subsite.

The active sites of GHs are mostly located in pockets and craters, grooves and clefts, or tunnels (Davies and Henrissat, 1995), Figure 1.7b-d. A pocket or crater, with a closed end on *exo*-acting enzymes is optimal for monosaccharides or oligosaccharides with multiple exposed chain ends, for example starch, by recognising the non-reducing extremity. A cleft or groove topology on *endo*-acting enzymes is an open surface allowing polymeric substrates to bind. Threading a polysaccharide through a tunnel topology on *endo*- and *exo*-acting enzymes can retain the tight binding of the polysaccharide in the tunnel whilst cleaving the glycosidic bond. This increases an enzymes processivity (Rouvinen et al., 1990).

Utilising a classical S_N2 reaction, cleavage of a glycosidic bond can occur. GHs catalyse reactions by either an inverting or retaining mechanism differing in net inversion or net retention of the anomeric configuration (Koshland, 1953; Vocadlo and Davies, 2008). A common mechanism of hydrolysis is shared within a family.

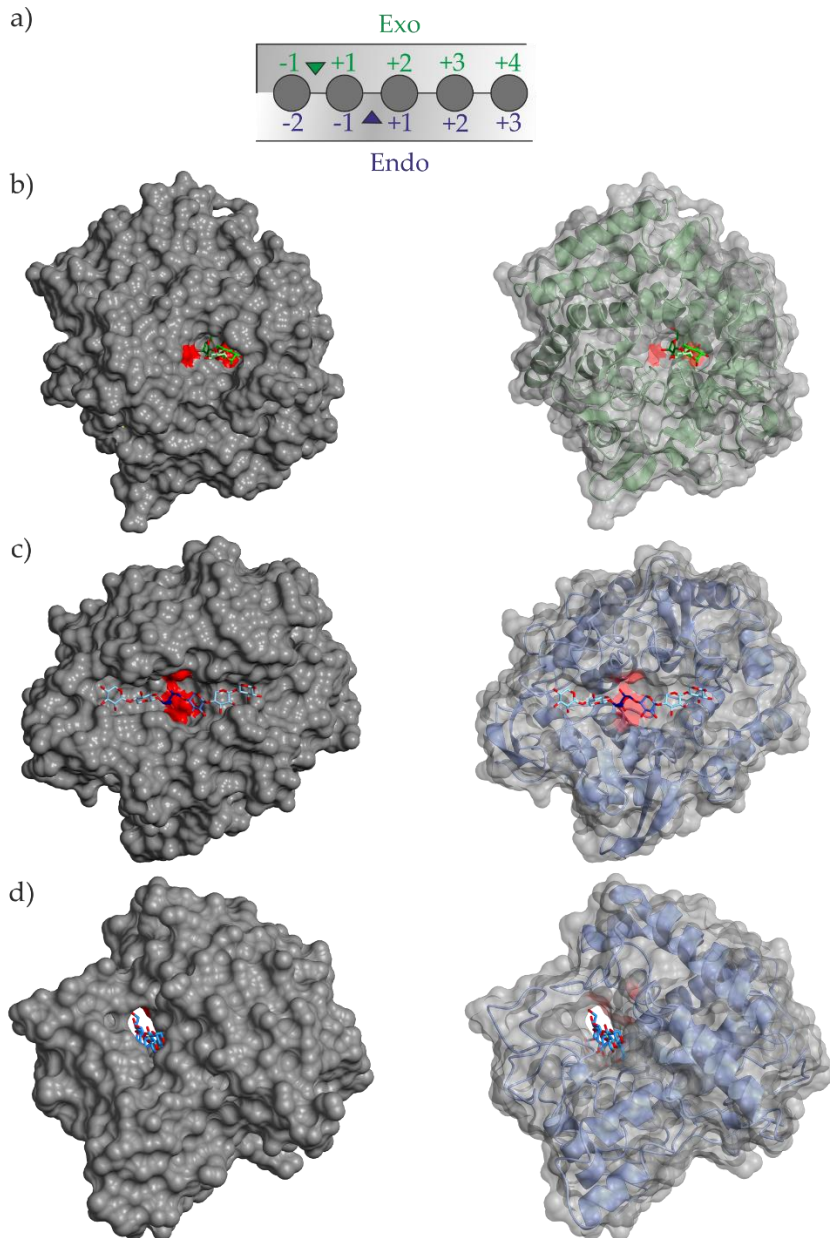


Figure 1.7. Exo- and endo-acting glycoside hydrolases. a) Non-reducing end *exo*-acting enzymes cleave at the point indicated by a green triangle. *Endo*-acting enzymes cleave at the point indicated by a purple triangle. b), c) and d) colour scheme of *exo*- and *endo*-enzymes continued throughout figure. b) Pocket active site of *exo*- α -1,6-mannosidase, *Clostridium perfringens* GH125 (Alonso-Gil et al., 2017). Catalytic residues, Asp220 (mutated to Asn in PDB ID: 5M7Y) and Glu393, highlighted in red. Mannotriose, green, is cleaved between the dark and mid green mannose. c) Cleft active site of *endo*-xylanase, *Teredinibacter turnerae* GH8 (PDB ID: 6G0N) (Fowler et al., 2018), catalytic residues, Asp281 and Glu73 and Xylohexabiose. d) Tunnel active site of *endo*-cellobiohydrolase II, *Trichoderma reesei* GH6 (PDB ID: 1QK2) (Zou et al., 1999), potential catalytic residues, Asp175, Asp401 and Asp221, methyl 4-S- β -cellobiosyl-4-thio- β -cellobioside, blue, is cleaved between the dark and mid blue glucose.

Catalysis is assisted by two residues in the enzyme, most commonly aspartates and glutamates, that act as a proton donor and/ or nucleophile/general base (Koshland, 1953; Sinnott, 1990; White and Rose, 1997).

1.2.3.1 The inverting mechanism

An inverting mechanism involves a single nucleophilic substitution in a one-step mechanism, Figure 1.8. This results in the stereochemistry of the hydroxyl at the anomeric centre being inverted. The general acid facilitates the departure of the aglycon (a compound, carbohydrate or non-carbohydrate, usually a phenol or alcohol group, remaining after glycolysis) leaving group and the general base facilitates the attack of the nucleophilic water at the anomeric centre (Koshland, 1953).

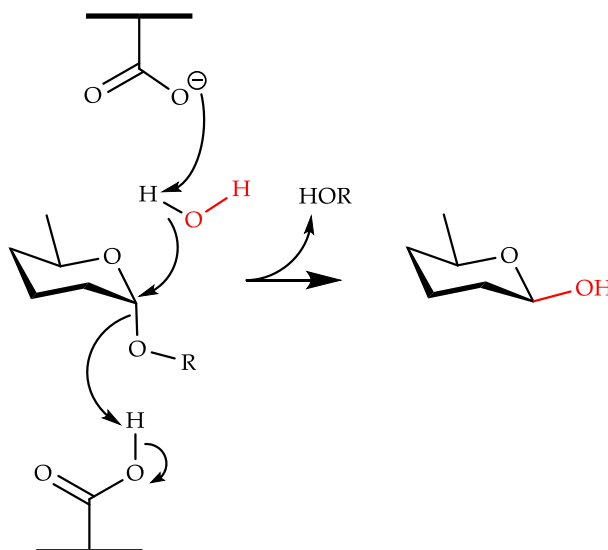


Figure 1.8. Inverting mechanism. α -D- to β -D-saccharide inversion with the donated hydroxyl from the nucleophilic water shown in red.

1.2.3.2 The retaining mechanism

For an enzyme with a retaining mechanism, the catalytic residues act as both an acid and a base in 2-step mechanism that results in net retention of the anomeric configuration (Koshland, 1953), Figure 1.9. In the first “glycosylation” step, the general acid/base facilitates the leaving group departure and the second catalytic residue stabilises the intermediate through either an ion pair or forming a covalent glycosyl bond. In the second “deglycosylation” step, the attack of a water is facilitated by the general acid/base in the

second nucleophilic substitution. The glycosyl-enzyme intermediate is broken down to form the hemiacetal product (reviewed in Vocadlo and Davies, 2008). The orientation of the general acid/base relative to the endocyclic C5 to O5 bond can be *syn* or *anti*, depending on the position of the proton in a half space of the -1 subsite, see section 1.3.3.1 (Heightman and Vasella, 1999; Nerinckx et al., 2005). The pKa value of the general acid/base fluctuates during catalysis to be optimal for the reaction to occur (McIntosh et al., 1996).

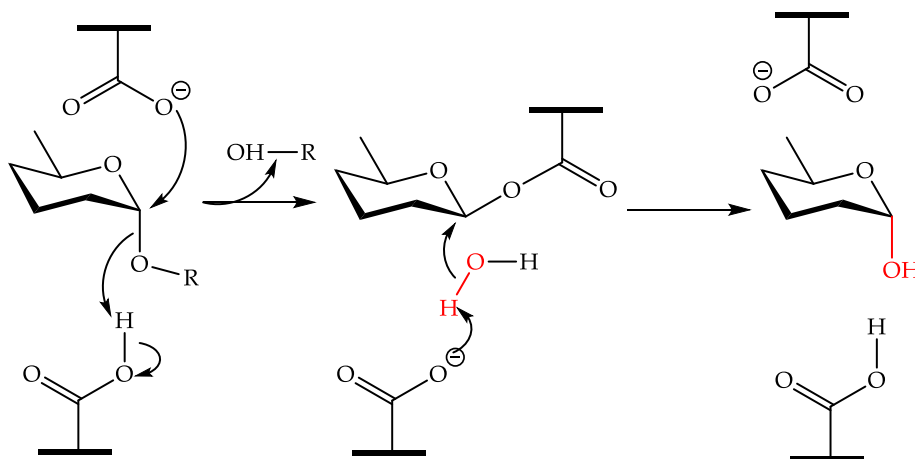


Figure 1.9. Retaining mechanism. Net retention of α configuration with the donated hydroxyl from the nucleophilic water shown in red.

The assignment of an inverting or retaining mechanism can also be inferred by the spatial position of the catalytic base. In both the inverting and retaining mechanisms, the proton donor is in the same position. However, inverting mechanisms utilise a nucleophilic water that is accommodated by increasing the distance between the carboxylates from a consistent ~ 5 Å (retaining) to a variable ~ 9.5 Å (reviewed in McCarter and Stephen Withers, 1994; Zechel and Withers, 2000).

1.2.3.3 The double displacement mechanism

Catalytic residue mutation studies analysed the effect of the C2 acetamido group of *N*-acetyl-chitobioside substrates, which is cleaved by lysozyme and some retaining chitinases. Removal of the C2 group reduced the k_{cat} by 100-fold, a larger reduction than the effect of mutating a potential catalytic base. These enzymes, as well as soluble lytic transglycosylases (Thunnissen et al., 1994), goose lysozyme (Weaver et al., 1995) and

hevamine (Terwisscha van Scheltinga et al., 1994) (and enzymes from families GH18, 20, 25, 56, 84 (see section 1.5.2) and 85) lack a suitably positioned catalytic base for their retaining mechanism; they utilise the C2 acetamido group as an intramolecular nucleophile (Martinez-Fleites et al., 2009). Neighbouring group participation of the *N*-acetyl carbonyl group and the carboxylate pair promotes a double displacement mechanism whereby the intermediate is a covalent oxazoline (deprotonated and stable) or an oxazolinium ion, Figure 1.10 (Tews et al., 1997; reviewed in Vocadlo and Davies, 2008). It is difficult to decipher the nature of the intermediate using experimental methods, therefore, computational QM/MM is used and recently showed that chitinases pass through a neutral oxazoline intermediate before an oxazolinium ion is formed on the way to product (Coines et al., 2018). The positive charge generated in the progression to the transition state is stabilised by a carboxylate in the active site.

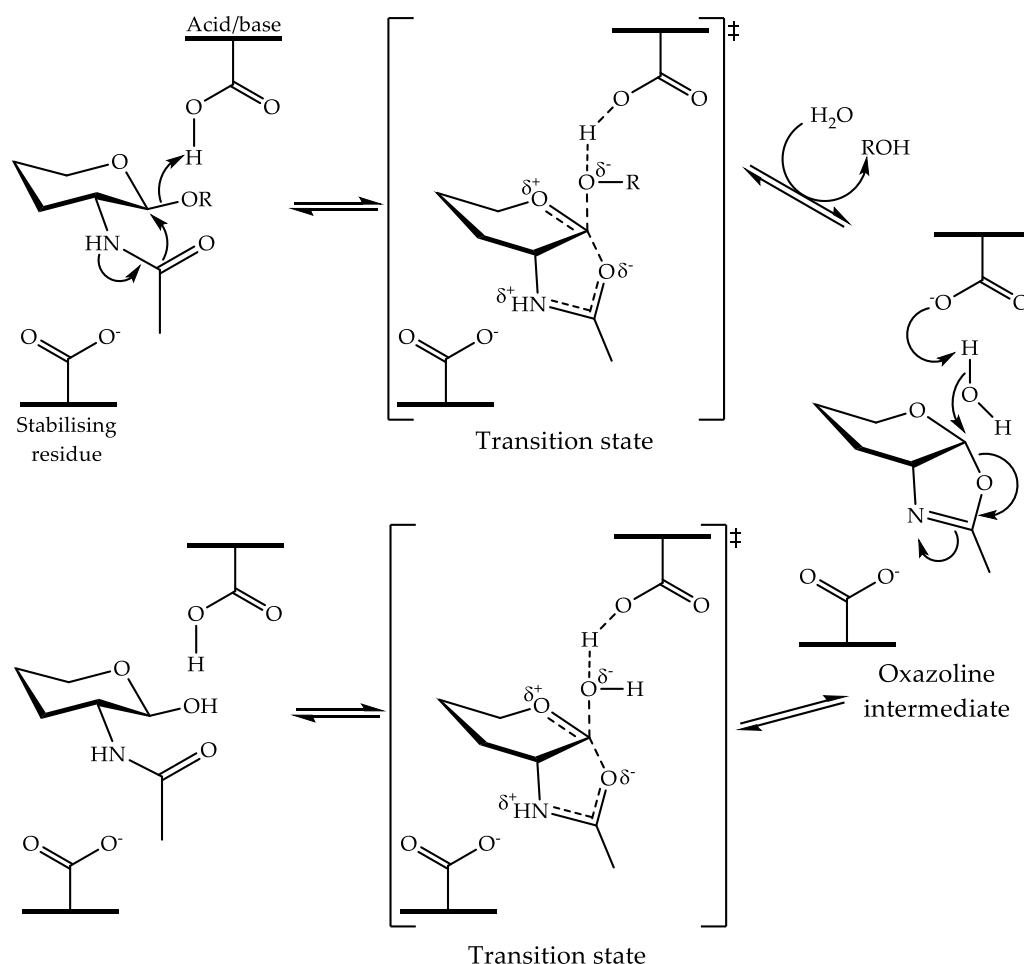


Figure 1.10. Neighbouring group participation by double displacement mechanism. Double displacement mechanism has two transition states either side of an oxazoline intermediate. Figure adapted from Withers and Williams, 2017.

1.3 Overcoming energy barriers

1.3.1 Substrates and enzymes bind in an activated state

A small amount of energy is required to form an enzyme-substrate (ES) complex upon binding of the substrate, Figure 1.11. A larger energy barrier, activation energy, must be overcome to reach the short lived (10^{-14} to 10^{-13} seconds) unstable transition state (Banait and Jencks, 1991). This barrier is lowered by a catalyst (enzyme) in addition to an amount of heat. It is followed by release of the lower energy product (second peak in Figure 1.11).

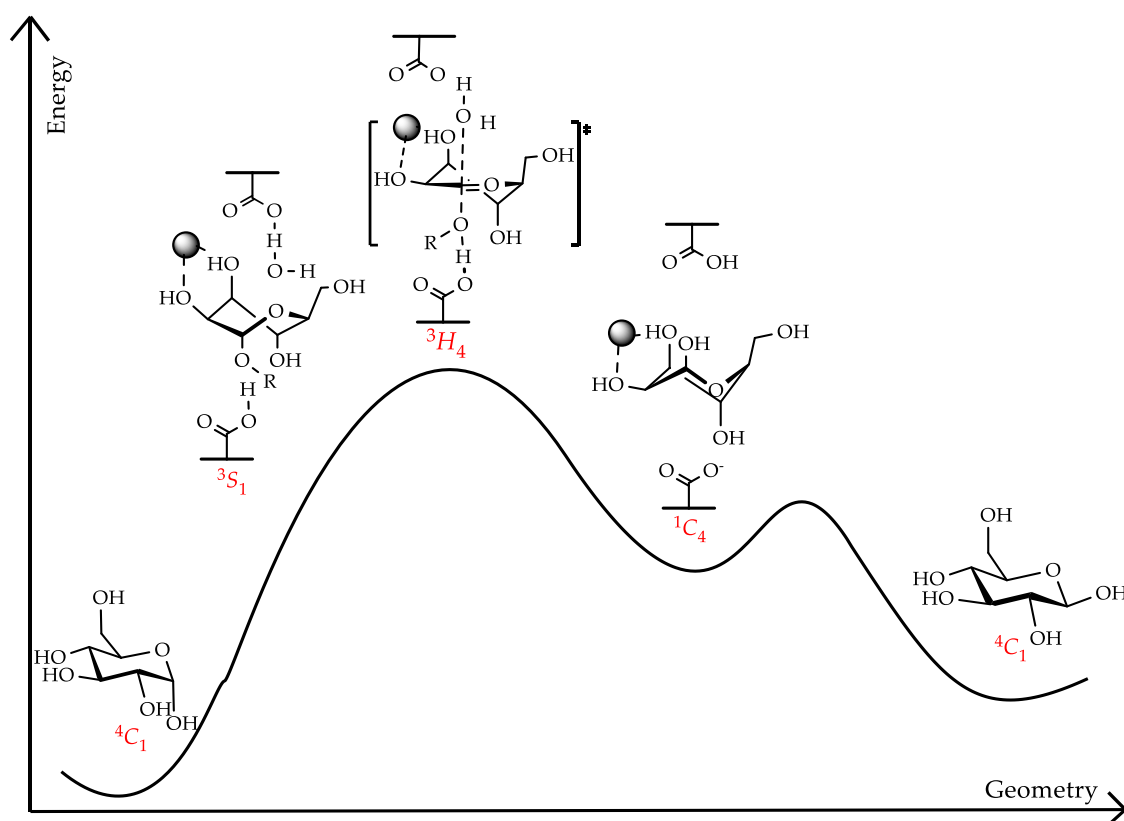


Figure 1.11. Mechanism for GH47 α -1,2-mannosidases. From the ground state 4C_1 conformation, energy is needed to reach the transition state conformation.

The “lock-and-key” model suggests that an enzyme and substrate must fit together like a lock and key (tightly bound) in order for the reaction to progress (Fischer, 1894). This idea advocates that the active site of an enzyme complements the strained conformation of the substrate, therefore, the substrate and enzyme are bound tightly in an activated complex (Pauling, 1946, 1948). An enzyme can reduce the energy barrier, allowing the reaction to proceed at a faster rate, through a different pathway to that of an uncatalysed reaction by

positioning two molecules close together in the correct orientation and strain the bonds of the substrate (reviewed in Závodszky and Hajdú, 2013). For a carbohydrate, the physical forces and strain on the molecule imposed by enzyme interactions leads to different conformations along the reaction coordinate, from Michaelis complex (MC) to transition state (TS/‡) to product conformation, Figure 1.11.

1.3.2 The conformational itinerary of a carbohydrate active enzyme

A “conformational itinerary” or “reaction coordinate” describes the changes that occur when a molecule undergoes a chemical modification. The conformations can be represented on a polar (spherical), Figure 1.12a, or a Mercator projection (grid), Figure 1.12b. Both projections can be compared to a globe map where 4C_1 is at the North Pole ($\theta=0$ or π) and 1C_4 is at the South Pole ($\theta=0$ or π). Each conformation has a specific phi (Φ) and theta (θ) value associated with its position on the projections as defined by the equations in Figure 1.12c. For example, the sequentially placed skew-boat and boat conformations along the equator are in steps of ($\phi = \pi/6$) and ($\theta = \pi/2$).

A molecule can pass from MC to TS to product, sometimes via an intermediate. These states must lie next to one another on the plot to obey the law of least nuclear motion (Sinnott, 1988). The principle of least nuclear motion states that if multiple species (conformations) could form during a chemical reaction, then the species that requires the least change in nuclear atomic positions will most likely occur and, therefore, will have the lowest energy of activation. For example, the conformational itinerary of an α -1,2-mannosidase from family GH47 is ${}^3S_1 \rightarrow [{}^3H_4]^\ddagger \rightarrow {}^1C_4$, Figure 1.12c and Figure 1.11. A conformational itinerary is specific to each family (although some families do have the same conformational itinerary if the enzymes are acting on the same substrate with the same stereochemical mechanism, for example, GH38 and GH76 α -mannosidases both share the same ${}^oS_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^1S_5$ conformational itinerary (for a full list see the review by Ardèvol and Rovira, 2015)). Distortion between different conformers across the equator of a projection is often of a lower energy requirement to overcome the activation barrier than interconversion between the two poles.

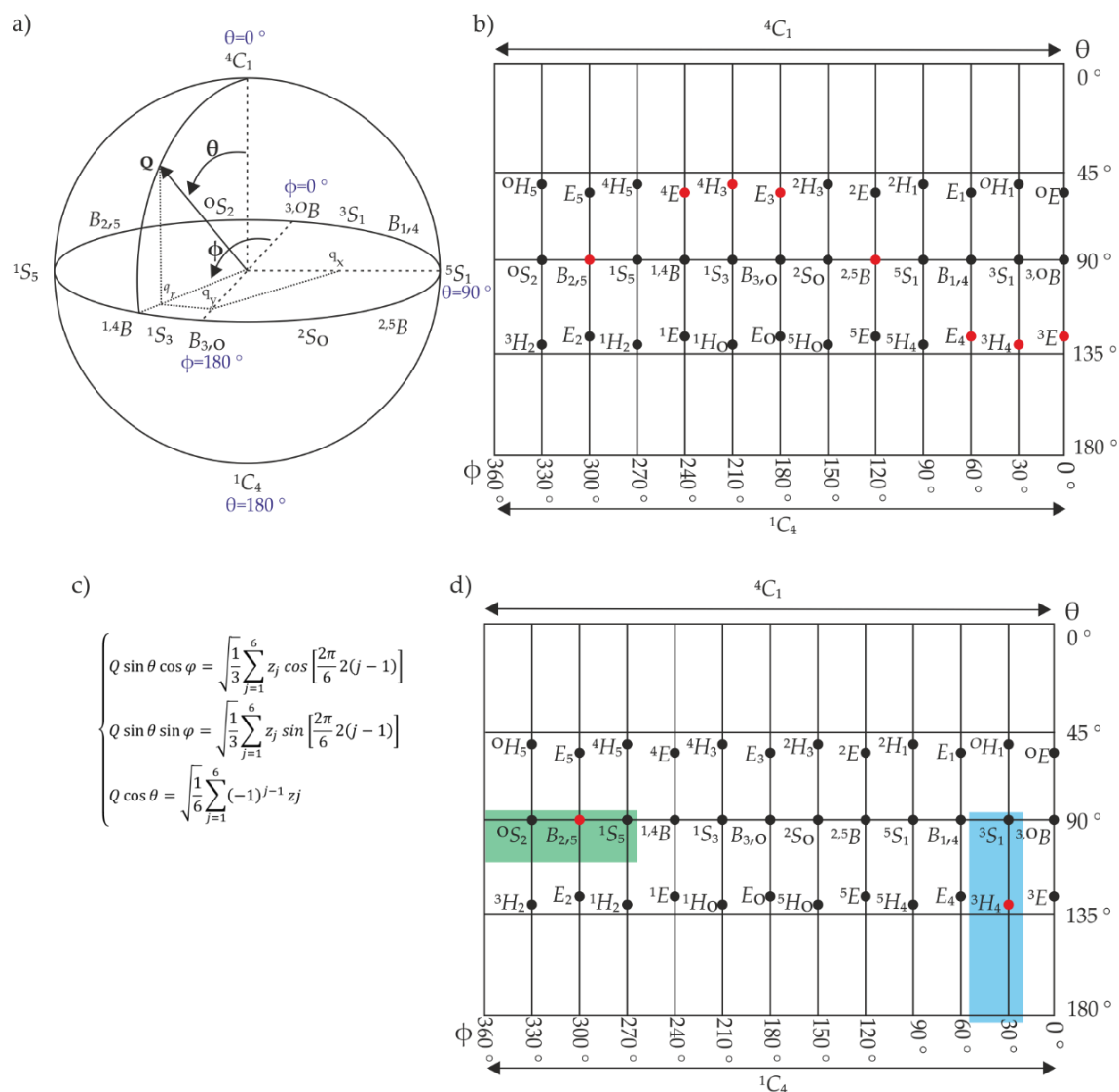


Figure 1.12. Projection styles for visualising the conformations of carbohydrates described in Figure 1.6. a) Polar projection with angles θ and ϕ and the total puckering amplitude, Q . Easy to observe conformational itineraries as there is no discontinuity. b) Mercator plot allows easy visualisation of all conformations that obey the principle of least nuclear motion. Highlighted in red are transition state-like conformations. c) The highlighted conformational pathways are common for mannosidases.

1.3.3 The transition state

The unstable TS has oxocarbenium-ion like character caused by ionisation (formation of a partial positively charged carbocation) of the glycosidic bond, Figure 1.13 (Sinnott, 1990). It is characterised by sp^2 hybridisation of C1 and O5 and a partial double-bond between the anomeric carbon and ring oxygen (borderline SN_1 - SN_2 displacement mechanism) (Koshland, 1953; Sinnott, 1987, 1990). The scissile C1-O bond of the TS species should be

elongated and the bond to the catalytic base should have formed (Vasella et al., 2002). The positively charged carbocation (C1) is stabilised by the overlap of the orthogonal lone pair containing $2p_z$ orbital of O5 (which carries a positive charge) and the orthogonal orbital lobes from the leaving and incoming groups of C1, Figure 1.13a. The enzyme can then make complementary interactions with the two localised areas, which are sterically accessible. In comparison, the ground state has two fully occupied sp^3 lone pair orbitals (Nerinckx et al., 2005).

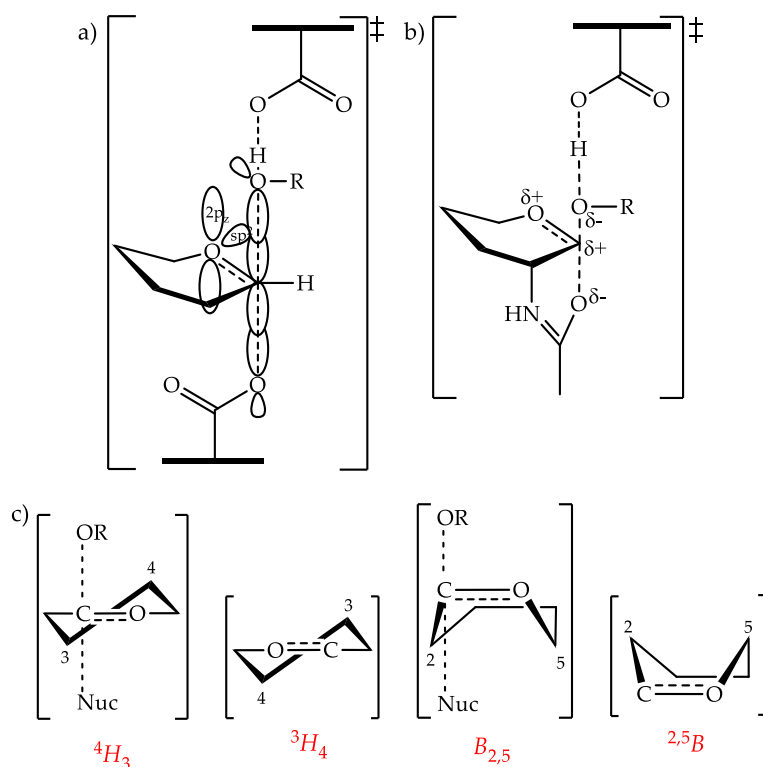


Figure 1.13. The oxocarbenium ion transition state conformation. a) The local positive charge at C1 is stabilised by an overlap of the $2p_z$ ring oxygen orbital and the C1–O1 orbitals of the scissile bond. b) The delta positive and negative charges at each atom of substrate assisted catalysis. c) Common transition states for glycoside hydrolyses. The specific boat, half-chair and envelope conformations are the only possible conformers for the transition state of scissile bond hydrolysis. The partial double bond is shown between C1 and the ring oxygen (Sinnott, 1990).

The only conformations where distortion causes C2, C1, O5, and C5 to become coplanar and obey these rules are $B_{2,5}$, 2B , 4H_3 and 3H_4 , (or the related envelopes 3E , E_3 , 4E and E_4), Figure 1.13. Therefore, the transition state of glycoside hydrolyses can only be close to one of these 8 conformations.

1.3.3.1 *Syn*- and *anti*-protonation

The -1 subsite can be split into sections known as half-spaces. Heightman and Vasella divided the space into *syn* and *anti* half-spaces in regards to the plane of O1, C1 and H1 carbohydrate (Heightman and Vasella, 1999). Vyas also divided the subsite into A and B half-spaces (perpendicular to *syn* and *anti*) dependent on the clockwise (above) or anticlockwise (below) average sugar ring plane and numbering (Vyas, 1991). The positioning of the catalytic acid, with respect to the carbohydrate, is different for *syn*- and *anti*-protonation, Figure 1.14. Consequently, the strategy to stabilise the transition state is different (Nerinckx et al., 2005; Vasella et al., 2002).

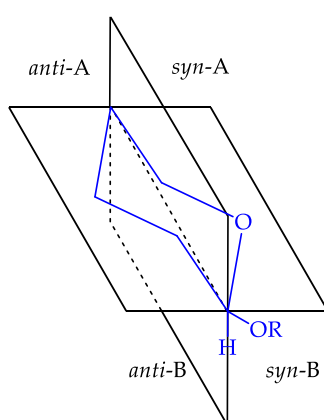


Figure 1.14. Space quadrants of the -1 subsite around a glycosidic ring. Figure adapted from Nerinckx et al., 2005.

1.3.4 Techniques to visualise the conformational itinerary

The techniques employed to identify the conformations along a conformational itinerary include X-ray/neutron crystallography, vibrational spectroscopy, nuclear magnetic resonance spectroscopy (Lipkind et al., 1984, 1985; Shashkov et al., 1986), optical rotation, Boltzmann weighting, computer modelling, quantum mechanics and molecular mechanics (Ardèvol and Rovira, 2015; Davies et al., 2012; Nieto, 2018; Woods, 1998; Wormald et al., 2002). Very few of these techniques can observe the conformational changes that occur on the nanosecond time scale required. Linear to cyclic ring cyclisation, puckering and rotation of the dihedral angles of the glycosidic bond occur at ~10 ns, and rotation of the primary hydroxyls occurs at ~1 ns and of secondary hydroxyls at ~100 ps.

Consequently, a combination of these techniques have been used simultaneously to corroborate the findings (there are many examples some of which include papers discussed in this thesis: Alonso-Gil et al., 2017; Males et al., 2017) (for a general review see Davies et al., 2012; Speciale et al., 2014).

1.3.4.1 Identifying the conformational itinerary using nuclear magnetic resonance

NMR has the advantage of observing the carbohydrate in solution, however, this is complicated by the extreme flexibility of the molecule (decreased upon covalent attachment to a protein (Zierke et al., 2013)). Therefore, nuclear Overhauser enhancements and bulk longitudinal relaxation time, which are both measured using ^1H -NMR, are dynamically averaged to produce the final “virtual” state that only a minority of molecules may possess (Cumming and Carver, 1987). Analysis of changes to the J coupling constants (describing the effect of neighbouring protons), over a range of pHs can identify chemical shifts of ring protons and the preferred orientation of the anomeric hydrogen indicating a change in conformation (Unione et al., 2015). For a recent review of identifying the conformation of a sugar using NMR techniques see Valverde et al., 2019.

1.3.4.2 Identifying the conformational itinerary using crystallography

To identify the conformational itinerary using crystallography, a series of high resolution inhibitor complexes can deduce the pathway (Thompson et al., 2012). “Solid” state crystallography is a snapshot of the reaction and is limited by resolution (Agirre et al., 2015a). For enzymes with multiple subsites, the carbohydrate in the -1 subsite can be forced into higher energy conformations due to interactions with the active site residues, see section 1.2.3, (Davies et al., 1997). The carbohydrate in the +1 binding site is usually in the lowest energy conformation (dependent on the saccharide) and may not undergo a conformational change during the reaction.

Distortion from ideal carbohydrate geometries and coordinates may occur upon fitting the carbohydrate model into the entire density map ($2F_o-F_c$) and the difference/omit density map (F_o-F_c). Incorrect distortion can arise due to weak density, low resolution, wrong linkage distance, or a mistaken anomer, by causing the total puckering amplitude to increase and if the deviation from the correct value is too large, a conformational change

will arise. The crystallographic programme PRIVATEER is used to validate the manual assignment of carbohydrate conformations (Agirre et al., 2015a, 2015b). The software exploits the Cremer-Pople algorithm that provides the angles of phi (Φ) and theta (θ) (only phi for furanose) and a total puckering amplitude (Q) (Cremer and Pople, 1975), Figure 1.12a. The algorithm calculates a mean plane, where the sum of all of the vertical distances from the atoms in the ring to the plane is zero. The sum of the perpendicular distance of each atom on the ring to the ring average plane gives the coordinates, Q. The relationship between these factors describes the conformation.

Limitations of using crystallography to identify the conformational itinerary include the following; it is often challenging and labour intensive to obtain diffracting crystals of the protein of interest; multiple inhibitors must be identified and synthesised to fix the protein in various states and capture each time point (Thompson et al., 2012); an inhibitor may bind in the lowest energy conformation and be misinformative; non-natural substrates, for example, compounds with thiol bonds, may distort the ligand due to natural bond length discrepancies (Gregg et al., 2011).

1.3.4.3 Identifying the conformational itinerary using quantum mechanical approaches

Quantum mechanical (QM) calculations can be an expensive but accurate technique to observe the different conformations that are accessible to a small carbohydrate molecule. Large molecules (~150,000 atoms) are limited by the phase space that can be explored (Canfield et al., 2006), but regions of interest can be isolated into super cells. Typically, experiments with QM are conducted in the gas phase, lacking the dynamic environment of solvents, however, programmes can be utilised that can approximate the solvent content using explicit or bulk waters. Using approximations of the Schrödinger equation, the exact energies of all orbitals of a molecule can be calculated from given *xyz* atomic positions.

Molecular mechanics (MM) can be calculated for all atoms of a large molecule, since some parameters (bond strength of bond lengths, bond angles and dihedrals) are predefined for proteins. The minimisations and energy calculated from “charges on a spring” are not as accurate as QM, but MM is faster.

By integrating QM and MM (QM/MM), the environment around the active site of an enzyme can be calculated using MM and the inhibitor and catalytic residues can be accurately calculated with QM. An initial structure (X-ray or homologous model) of highest resolution is required for QM/MM simulations. *Ab initio* QM/MM can produce free energy landscapes (FELs) to infer the energetic cost for the reaction to occur “on-enzyme”. As previously mentioned in section 1.3.1, small structural and electronic changes in the substrate distort it into an activated state. The shape of the -1 and +1 subsites and substrate interactions with residues in the active site have evolved to accommodate the activated substrate. An FEL is unique to each carbohydrate stereoisomer. It observes the energy, typically Gibbs free energy, required for a carbohydrate to distort into a certain conformation. The lowest energy conformations (most energetically accessible) are often the predisposed distorted conformations; harmonising the link to the activated state (Wolfenden et al., 1998). To calculate an FEL, several density functional theories (DFTs, a first-principle method (Hohenberg and Kohn, 1964)) can be applied with slight variations in the results depending on the DFT (± 0.6 kcal/mol for β -glucose). The most stable minimum of the FEL can then be used in MD simulations.

Comparisons can be made between the FELs of free substrate and substrate in the presence of an enzyme. A different low energy conformation upon enzyme binding (“off-enzyme” versus “on-enzyme” states) can occasionally be observed, Figure 1.15.

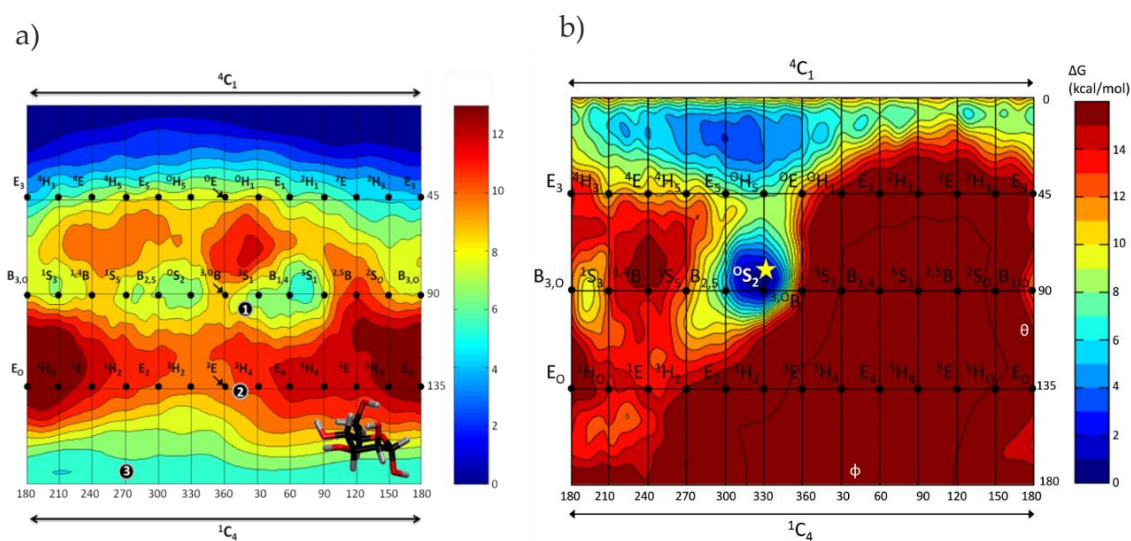


Figure 1.15. Conformational free energy landscape of α -mannose. The lowest accessible energy conformation of isolated α -mannose is 4C_1 . b) α -D-mannose in complex with CpGH125 has a lowest conformation of 0S_2 (Alonso-Gil et al., 2017).

Molecular dynamics uses a set of defined variables (Cremer-Pople Cartesian puckering coordinates) to allow the protein to “wobble” and “relax” into a low energy state. It is commonly used before QM/MM to allow the protein to adjust to the modelling of an inhibitor in the active site or building missing elements and solvation. The calculations are repeated on a fs protein timescale to calculate where each atom will have moved to with a certain speed and a certain force. Longer time scales for MD increase the chances of the substrate binding to the enzyme and conformational changes occurring.

1.4 Inhibition of carbohydrate active enzymes

Inhibition of a carbohydrate active enzyme is important for biochemical research to understand a) how a novel enzyme functions, b) for structural studies, c) the catalytic mechanism and interactions in the active site, and d) therapeutic applications where the inhibitors can be used to clinically treat defective enzyme disorders. Inhibitor design is an iterative process with the aim of enhancing the binding affinity, selectivity and synthesis pathway of the previous generation of inhibitors.

1.4.1 Rationale of design

Designing inhibitors to comply with a variety of different biological necessities, for example, stability in biological samples and crossing the blood-brain barrier, can be difficult. Promiscuous inhibitors are not ideal for clinical use. Ideally an inhibitor should target a specific enzyme; this is problematic since different GH families can have functionally and spatially similar active sites. For clinical use, the inhibitor should be stable in biological fluids and have minimal off-target effects, high potency and low toxicity.

Using high-throughput screening methods, for example, fragment-based drug design, 1000s of inhibitors can be tested using crystallography and binding assays. Prior knowledge of a system can be used to repurpose and modify existing ideas.

1.4.2 Transition state mimicry

Identifying the conformational itinerary of a substrate in an active site is useful for designing inhibitors that are potent to the target enzyme and importantly, mimic the transition state. Inhibitors that have a bias for specific conformations complementary to a

conformational itinerary for an enzyme are likely to have a high affinity (Kuntz et al., 2008; Males et al., 2017). This is due to a lower energy requirement for distortion of the bound inhibitor by the enzyme.

Naturally occurring inhibitors that mimic the TS of target enzymes have been identified and exploited, indicating the necessity to bind the TS even in nature. These inhibitors are generally more potent than the current synthetically designed inhibitors of low (pM) affinity.

The design of TS mimics can fall under four categories: charge mimics, shape mimics, charge and shape mimics and combination glycan-aglycon mimics (detailed reviews by Davies et al., 2012; Ganem, 1996; Gloster and Davies, 2010; Gloster and Vocadlo, 2012; Speciale et al., 2014; Williams et al., 2014). Compounds mimicking an overall positive charge of the TS have a basic nitrogen atom that replaces either the endocyclic oxygen or anomeric carbon on the related natural substrate. To mimic the planar geometry of the TS, a trigonal anomeric centre and/or ring oxygen is incorporated. Mimicking both planarity and positive charge increases the efficacy of the inhibitor. Combining an aglycon (non-carbohydrate) group, such as hydrophobic moieties and aromatic groups, can occupy more subsites and increase the interactions across the active site.

Linear (Gibbs) free energy relationships (LFERs) describe the relationship of enzyme rate and the affinity of the inhibitor for the enzyme upon changing environments, for example, the structure of the inhibitor or non-detrimental mutations/alterations to the enzyme. If the inhibitor is a true TS mimic, then these effects and relationships (K_m/k_{cat} and K_i) should correlate with each other (reviewed in Gloster and Davies, 2010). The most common types are the Brønsted, taft-like and Bartlett relationships (reviewed by Vocadlo and Davies, 2008).

Kinetic isotope effects (KIEs) compare the reaction rates and vibrational frequencies of a substrate and its enriched heavy isotope isoform to inform of the reactional geometry changes undergone (Berti and Tanaka, 2002). A combination of LFERs and KIEs can provide useful complementary insights into charge distribution (LFERs) and geometry (KIEs).

1.4.3 The synthesis of inhibitors for glycoside hydrolases

The first families of glycoside hydrolases inhibitors were monosaccharide-derived δ -aldonolactones, e.g. *D*-gluconolactone, Figure 1.16, (Conchie and Levvy, 1957) and glycosylamines, e.g. 1-amino-1-deoxypyranose *D*-glucosylamine (Lai and Axelrod, 1973). Replacement of the pyranose oxygen of *D*-glucose and *D*-mannose with a nitrogen formed the azasugar analogues termed polyhydroxylated piperidine, pyrrolidine and indolizine alkaloids. At basic pH, the nitrogen atom, which replaces a naturally occurring oxygen atom, is protonated mimicking the incipient partial charge of the transition state. The popular azasugars, Figure 1.16, nojirimycin, 1-deoxynojirimycin, 1-deoxymannojirimycin (DMJ), castanospermine and swainsonine, which closely resemble *D*-glucopyranose and *D*-mannopyranose, have a wide range of biochemical applications (Winchester and Fleet, 1992).

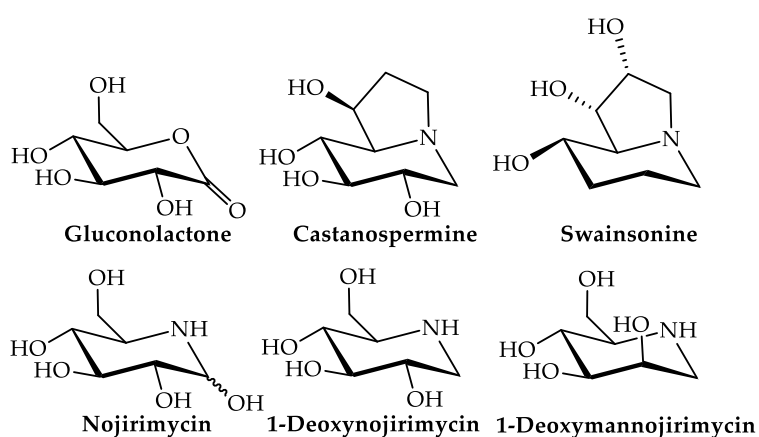


Figure 1.16. Structures of common glycosidase inhibitors.

1.5 Comparison of the puckering of mannose and *N*-acetyl- β -*D*-glucosamine

Both mannose and *N*-acetyl- β -*D*-glucosamine (GlcNAc) have interesting chemistry of the substituents at the 2-position, adjacent to the reactive anomeric centre (C1). The preference for an axial anomeric oxygen limits the number of available activation conformations to skew boat and boat conformations, Figure 1.17. Another factor increasing susceptibility for catalysis is the partial charge on C1, and distance between C1 and O5. As explained in sections 1.2.3.3 and 1.3.3, during glycoside hydrolase catalysis, nucleophilic attack occurs at C1 and a high partial charge and a short C1-O5 bond length is characteristic of the transition state.

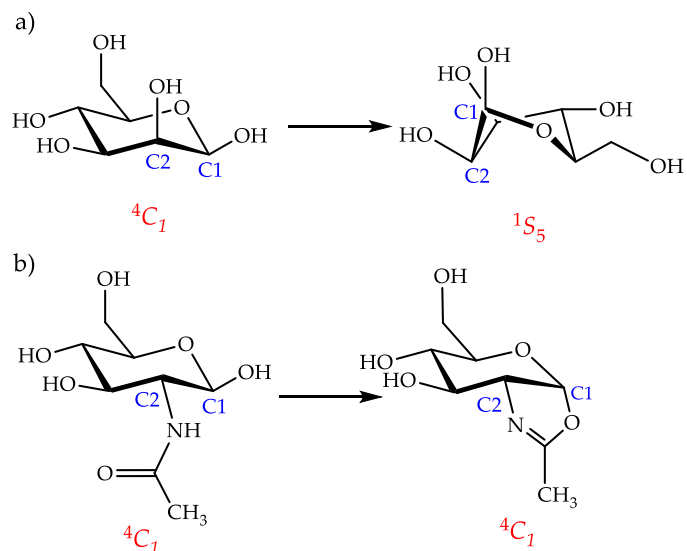


Figure 1.17. Conformational changes for mannose and GlcNAc. From the solution-stable 4C_1 conformation to an activated substrate conformation where the anomeric oxygen is orientated axially. a) The transition from 4C_1 to 1S_5 is catalysed by family GH38 α -1,3-mannosidase enzymes. b) The transition from 4C_1 to a 4C_1 oxazoline intermediate is catalysed by GH84 *N*-acetyl glucosaminidases.

1.5.1 Mannose

The first synthetically manufactured mannose was published by Fischer and Hirschberger in 1888 (Lichtenthaler, 1992). Mannose is a C2 epimer of glucose and functions both as a monomer and as polymeric mannans (a component of hemicelluloses (plant cell walls), fungal cell walls, storage polysaccharides in parasites and N-glycan processing). It is used in the food industry for its bitter (β -) and sweet (α -) taste in food. The arrangement of O1-O2-O3 in β -mannose is *cis-cis* (in comparison, the arrangement for β -glucose is *trans-trans*) which causes a region of high *cis-cis* clashes around 4E - 4H_5 - E_5 , correlating to steric hindrance (Mayes et al., 2014).

The 4E - 4H_5 - E_5 region lies between the ground state, 4C_1 , conformation and the activated substrate, 1S_5 , conformation, which is an energetically unfavourable route for β -mannose to pass through, Figure 1.17a. Remarkably, β -mannose cannot distort into a conformation that is correlated with a puckering geometry that has favourable kinetics, bond lengths and charge. For example, a 1S_3 conformation has the lowest Gibbs free energy and a 0_3B conformation has the highest partial charge on C1, but the largest contribution to conformation choice is enzyme-carbohydrate interactions, since the popular 1S_5 offers a

larger distance between O1 and O2, preventing clashes with the catalytically important O2 (Mayes et al., 2014), see sections 2.2.6 and 3.2.5.

1.5.2 *N*-acetyl- β -D-glucosamine

N-acetyl- β -D-glucosamine (GlcNAc) functions as a regulatory post-translational modification to serine and threonine residues of intracellular protein targets and plays an important role in cell signalling. It is also a monomer of chitin, which functions in the cell walls of fungi, algae and yeast, and of peptidoglycan in the cell wall of bacteria. Atypical of other carbohydrates, GlcNAc comprises a long *N*-acetyl arm attached to C2. The oxygen of the *N*-acetyl group can form an intramolecular hydrogen bond to O3, with a distance of 2.12 Å, causing C2 to be in the same plane as O5, C1, C3 and C4. This condition tolerates a stable low energy ⁵*E* conformation (Mayes et al., 2014).

The positioning of the *N*-acetyl arm is important for enzymes that do not utilise enzyme carboxyl residues for nucleophilic attack but rather utilise the *N*-acetyl group of GlcNAc (van Aalten et al., 2001), Figure 1.13b and Figure 1.17b. The *N*-acetyl group at the C2 position must be in an axial orientation to be below the position of C1 described by the O5-C1-C2-N angle. Therefore, the available conformations where the *N*-acetyl group is poised to execute a nucleophilic attack at C1 are favoured. Family 18, 20, 25, 56, 84 and 85 GHs hydrolyse substrates with an *N*-substituent (acetyl/acetamido/glycolyl) at the 2-position.

1.6 Outline of the thesis

The overarching theme of this thesis is to examine the enzymes that utilise exocyclic group properties. The unusual self-interaction of GlcNAc, due to the intramolecular hydrogen bond formed between the *N*-acetyl arm and O3, and steric clashes of mannose, formed between the substituents of C1 and C2 due to the *cis-cis* orientation, result in the stabilisation of conformations that would otherwise be short lived. As a result of the necessary axial orientation of the 2-position substituents, enzymes handling these carbohydrates display features that can be exploited for inhibitor design, often leading to cellular-active tools with therapeutic implications.

Glycoside hydrolases are key enzymes that regulate a multitude of biological functions. Defects in glycoside hydrolases can result in impaired glycosylation, leading to many different genetic diseases and cancer progression. They catalyse the hydrolysis of oligosaccharides by two mechanisms whereby the substrate in the -1 subsite of the enzyme passes through a series of conformations beginning with the Michaelis/substrate, through the transition state and ending in the product conformation. The activation energy to overcome the transition state conformation, which is commonly a half-chair or boat conformation, is reduced by assisted enzyme catalysis due to their restrictive active site and presence of metal ions. The transition state is short lived with potent binding, making it the perfect target for drugs that have been developed to mimic the transition state and inhibit GHs. The work presented in this thesis aims to analyse the three-dimensional structures and conformational itineraries of mannosidases and β -*N*-acetylglucosaminidase and propose specific and high affinity small molecule inhibitors to be used as therapeutic agents.

The thesis is divided into two sections, the first relating to N-linked glycan processing by α -mannosidase families GH125 and GH47 and the second relating to O-linked glycan processing, specifically by the β -*N*-acetylglucosaminidase, O-GlcNAc hydrolase, family GH84. Each section contains two chapters composed of both published (marked by a preliminary page of declaration with reference) and unpublished results.

The second chapter of this thesis challenges the conformational itinerary of α -1,6-mannosidase GH125 enzymes. Using X-ray crystallography and computational QM/MM metadynamics in tandem, the conformation of the natural substrate (α -1,6-linked mannose) and of a mannosidase inhibitor (mannoimidazole) revealed the conformational pathway for GH125 enzymes (${}^oS_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$).

The third chapter reveals the necessity of developing conformational-restricted inhibitors that are specific to a conformation along the conformational itinerary. This results in inhibitors that are specific to a certain family of enzymes (α -1,2-mannosidases family GH47 in this chapter) and have high affinity binding.

Divergence to GlcNAc occurs in the fourth chapter of this thesis. A variety of different constructs were designed and produced to improve on published structures of the human OGA enzyme since there are regions of high disorder which cannot be modelled and

observed in crystal structures. A detailed structure of human OGA was produced by mutating residues to lower the surface entropy to encourage crystallisation. This resulted in a 5% increase in the number of residues modelled, without compromising its kinetic ability. Analysis of OGA phosphorylation by casein kinase II by mass spectrometry revealed the phosphate binding site, Ser364. Subsequently, a phosphomimetic mutant (S364D) was produced to mimic phosphorylation and potentially stabilise a region of high disorder between Lys341–Thr370. Unfortunately, there was no change in the crystal structure or to the enzyme kinetics.

The final chapter of results, chapter 5, will introduce a new class of inhibitors for GH84 enzymes. Specificity is of particular importance considering that family GH20 enzymes possess a similar enzyme function. The inhibitors are designed to cross the blood-brain barrier. A compound library of C1 triazoles revealed a subset of inhibitors with picomolar inhibition constants. The inhibitors achieve this potency by altering their conformation to increase the number of enzyme stacking interactions.

SECTION 1: N-Glycan Processing

i. N-linked glycan processing

N-linked glycosylation occurs in the endoplasmic reticulum (ER) and Golgi apparatus. It is an important facilitator of protein folding, solubility, recognition and antigenicity. The conserved N-glycan binding protein sequence is **Asn**-X-Ser/Thr where N-glycans are covalently linked by an N-glycosidic bond to the asparagine residue (in bold). The efficiency of the glycosylation is affected by the residue, X, between Asn and Ser/Thr. If the amino acid is acidic, an aspartate or glutamate, the efficiency is reduced. Glycosylation is not viable when the amino acid is a proline. A glycan at the same residue within a population can display heterogeneity by acquiring different branches (Stanley et al., 2009).

As the protein passes through the ER and Golgi, the glycan is modified by a sequential series of enzymes that remove and incorporate certain carbohydrate moieties to the core pentasaccharide, $\text{Man}_3\text{GlcNAc}_2$ (Stanley et al., 2009). During passage through the Golgi, high mannose N-glycans are processed to further complex glycans. High mannose glycans have between 5 and 9 mannose moieties linked to GlcNAc_2 (chitobiose) with 3 mannose sugars at the termini, Figure 1.i. $\text{Man}_5\text{GlcNAc}_2$ is then processed to hybrid and branched glycans. Hybrid and branched glycans contain both terminal mannose residues and another form of carbohydrate at the termini. Lastly, further processing to complex sugars occurs in the *trans*-Golgi. Complex sugars modify the core pentasaccharide with bi-, tri- and tetraantennary GlcNAc -linked antennae saccharides and the most common terminal monosaccharide is sialic acid. The final glycans in the process differ in structure between different kingdoms.

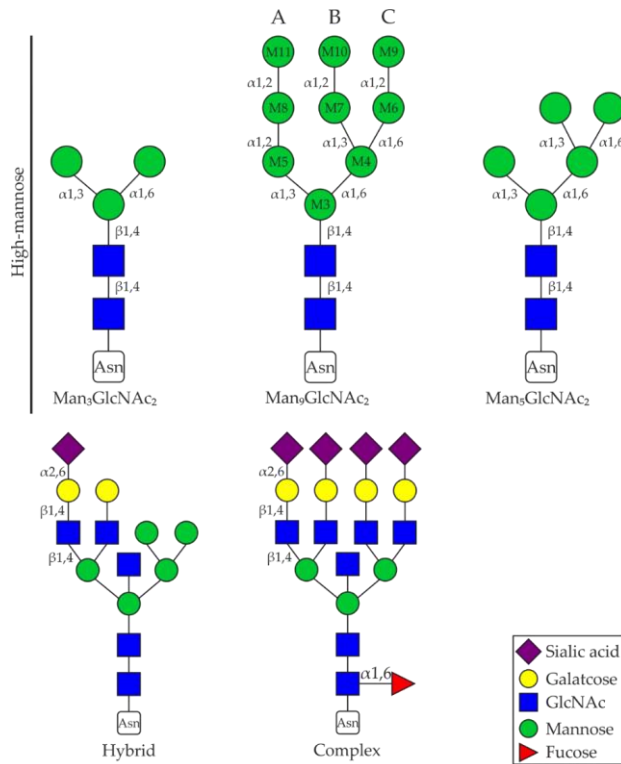


Figure 1.i. N-linked glycan structures. Processing of $\text{Man}_9\text{GlcNAc}_2$ through the ER to $\text{Man}_3\text{GlcNAc}_2$ and through the Golgi to form complex glycans.

ii. N-glycan assembly and protein transfer

Firstly, in the cytoplasm of the cell, the building platform and membrane anchor dolichyl phosphate is formed from dolichol and cytidine triphosphate catalysed by dolichol kinase, Figure 1.ii (Rip et al., 1981). A phosphorylated GlcNAc sugar, from UDP-GlcNAc (uridine diphosphate GlcNAc, see section 2i), is added to dolichyl phosphate by GlcNAc-1-P-transferase (Lehrman, 1991). Subsequently, a second GlcNAc is added with a β -1,4-glycosidic linkage, Figure 1.ii. Next, five mannose sugars are added by a complex of GT33 and GT4 transferases (Huffaker and Robbins, 1982; Kämpf et al., 2009; O'Reilly et al., 2006). The product $\text{Man}_3\text{GlcNAc}_2\text{-PP-Dol}$ is translocated to the lumen of the ER by requiring fifty three 1 (Jelk et al., 2013). A series of seven transmembrane transferases then attach mannose M7, M9, M10 and M6 (Aebi et al., 1996; Takahashi et al., 1996; Taron et al., 2004), then three glucose (Burda and Aebi, 1998; Reiss et al., 1996; Runge and Robbins, 1986). This results in the final product of the assembly $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$. Nascent proteins from ER-bound ribosomes are translocated through the translocon to the lumen. Here the glycan is transferred from dolichol phosphate to the asparagine residue, Figure 1.ii (Słomińska-Wojewódzka and Sandvig, 2015).

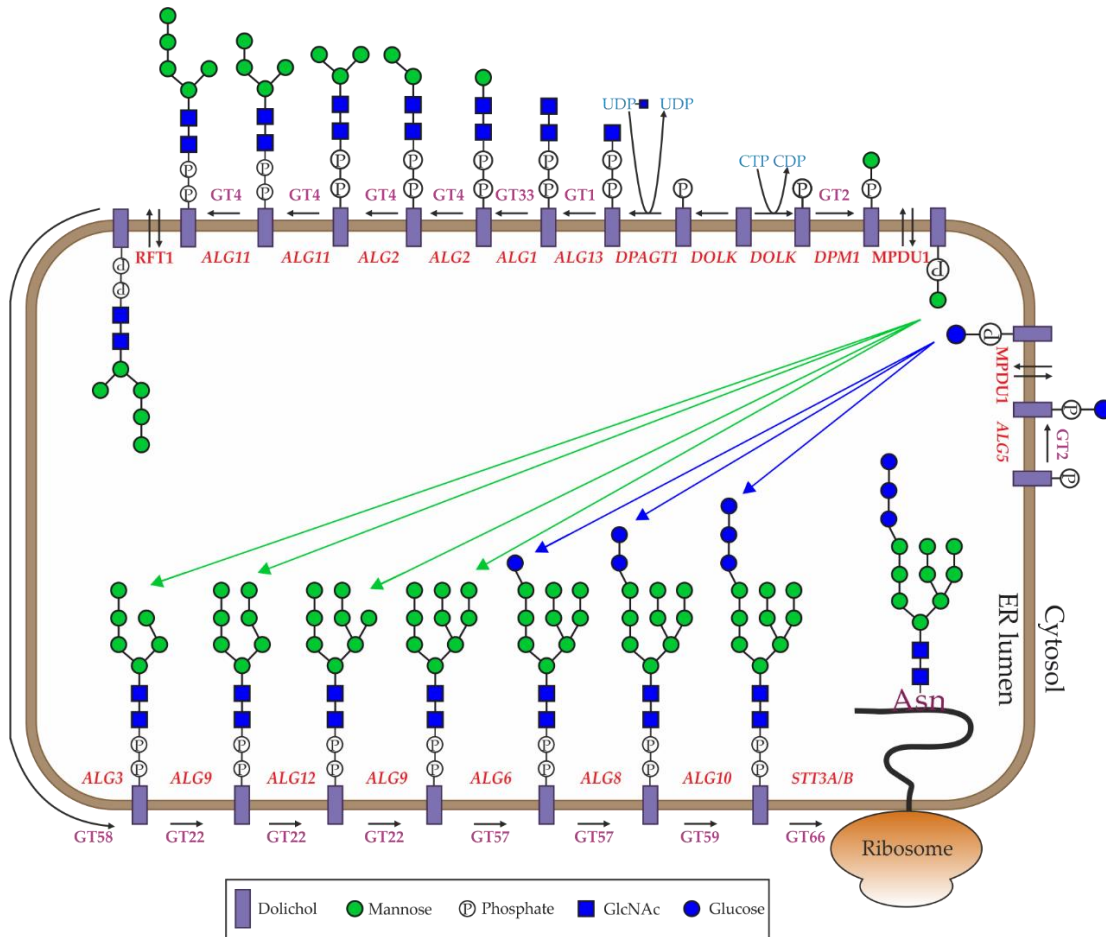


Figure 1.ii. N-glycan assembly occurs in the cytoplasm and the ER lumen. N-glycan assembly pathway from dolichol to $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ and its attachment to an asparagine protein residue. Gene and protein names are in red and glycosyltransferase (GT) families are in pink. Blue and green arrows indicate the movement of GlcNAc and mannose across the ER lumen to the relevant glycan.

iii. N-glycan trimming pathway in the ER and ER-associated degradation for quality control

Immediately after entering the ER lumen, the terminal glucose residue is cleaved by ER α -1,2-exoglucosidase I, GH63 (Elting et al., 1980), and is recognised by the chaperone protein, malectin, Figure 1. (Schallus et al., 2008). Following dissociation from the lectin site, GH31 ER glucosidase II consecutively deglycosylates the second and final glucose sugars from the glycoprotein to form $\text{Man}_9\text{GlcNAc}_2$ (Grinnat and Robbins, 1979). The protein is then folded with the assistance of the chaperones calnexin (Chen et al., 1995)

and calreticulin (Peterson et al., 1995) Figure 1.iii. Once the protein is folded, ERManI removes the central terminal mannose, M10, forming Man₈GlcNAc₂ (Stanley et al., 2009).

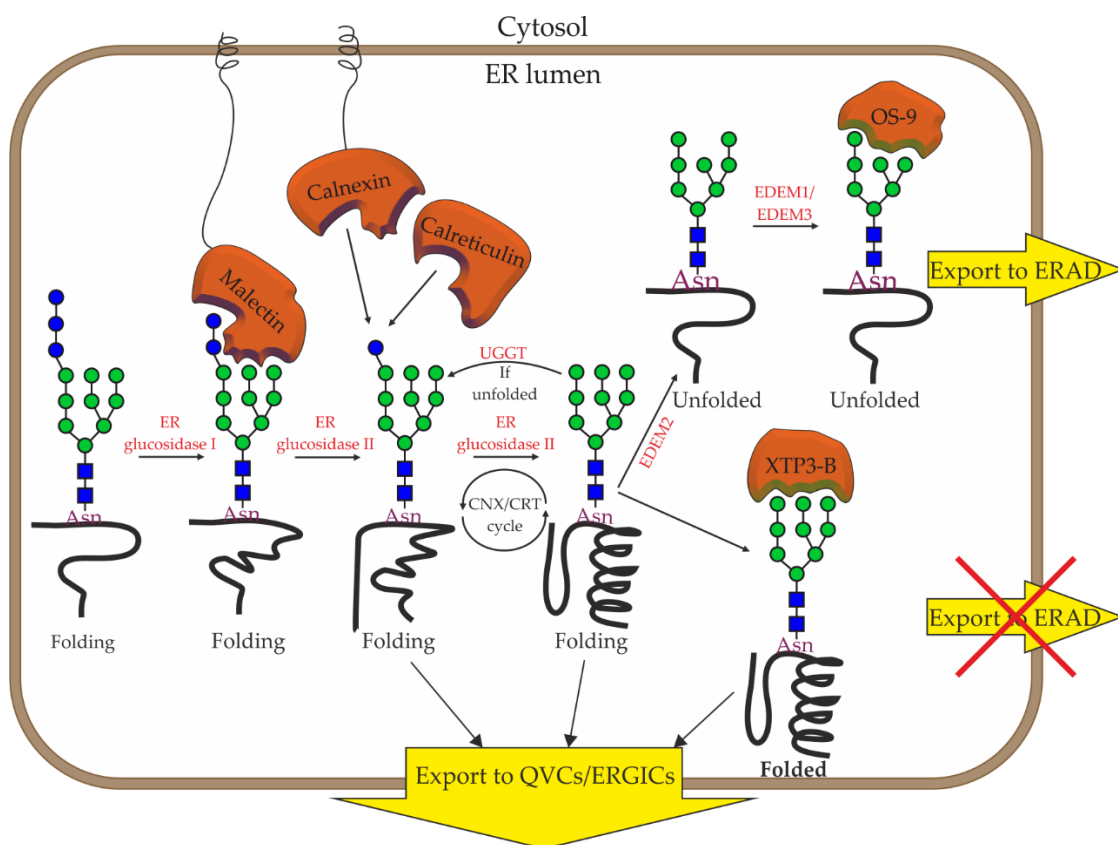


Figure 1.iii. N-glycan trimming and protein folding recognition in the ER. The chaperones, malectin, calnexin and calreticulin help to fold the glycosylated protein. If the protein is folded correctly, it is recognised by the lectin XTP3-B and degradation by the ERAD pathway is inhibited (van der Goot et al., 2018). If the protein is still unfolded after multiple calnexin/calreticulin binding cycles, it is exported to the ERAD pathway.

Before the glycosylated protein leaves the ER, UDP-glucose:glycoprotein glucosyltransferase, GT24 (Parodi et al., 2014), assesses protein folding by recognising hydrophobic patches in partially folded proteins, Figure 1.iii. If the protein is not folded correctly (Sousa et al., 1992) or a protein complex has misassembled (Gardner and Kears, 1999), it is re-glycosylated and the calnexin-calreticulin cycle continues until folding occurs (Parodi et al., 1983). If the protein is still unfolded, it will be targeted to the ER-associated degradation (ERAD) pathway (Vembar and Brodsky, 2008). EDEMs also target proteins to the ERAD. EDEM1 binds to calnexin to form a functional complex to prevent aggregation of misfolded proteins (Olivari et al., 2006); EDEM2 cleaves Man₉GlcNAc₂ to

Man₈GlcNAc₂ (Ninagawa et al., 2014); EDEM3 (and EDEM1) accelerate the removal of mannose from branch A or C forming Man₇GlcNAc₂ from Man₈GlcNAc₂ (Cormier et al., 2009; Hosokawa et al., 2001). The adaptor lectins, osteosarcoma amplified 9 and XTP3-B have an antagonistic relationship, whereby osteosarcoma amplified 9 regulates and promotes HRD1-dependent ERAD for incorrectly folded/unfolded proteins and XTP3-B inhibits ERAD for correctly folded proteins (Fujimori et al., 2013; Hosokawa et al., 2009).

iv. N-glycan trimming pathway in the Golgi

Transportation of the correctly folded proteins through the ER is carried out through QCVs or ER-Golgi intermediate compartments and then repacked into vesicles to be transported to the Golgi (Pimpaneau et al., 1991; Schweizer et al., 1988). Cargos expelled into ER-Golgi intermediate compartments are packaged into vesicles coated by coat protein II (Schekman and Orci, 1996).

Cis-Golgi α -mannosidase IA (ManIA), IB (ManIB) and IC (ManIC) process Man₈GlcNAc₂ to Man₅GlcNAc₂ by the removal of M11, M8 and M9, Figure 1. (Igdoura et al., 1999). Unlike ERManI, ManIA prefers cleaving M11 branch A from Man₉GlcNAc₂ first. If ERManI has cleaved M10 branch B, ManIA preferentially cleaves M9 on branch C (Ogen-Shtern et al., 2016). Removal of M11 prevents reglucosylation and removal of M9 allows binding to the MRH domain of OS-9, a lectin, to target the protein to ERAD (Hosokawa et al., 2009). The amino acid sequence of ManIC is 54% homologous to ManIA and 38% homologous to ManIB. The protein travels to the *medial*-Golgi where branched processing is initiated by 6 *N*-acetylglucoaminyltransferases (GlcNAcT I-VI) (Tanichi and Korekane, 2011). Lastly, in the *trans*-Golgi, the branches can be elongated, capped and decorated to comprise of seven or more component monosaccharides.

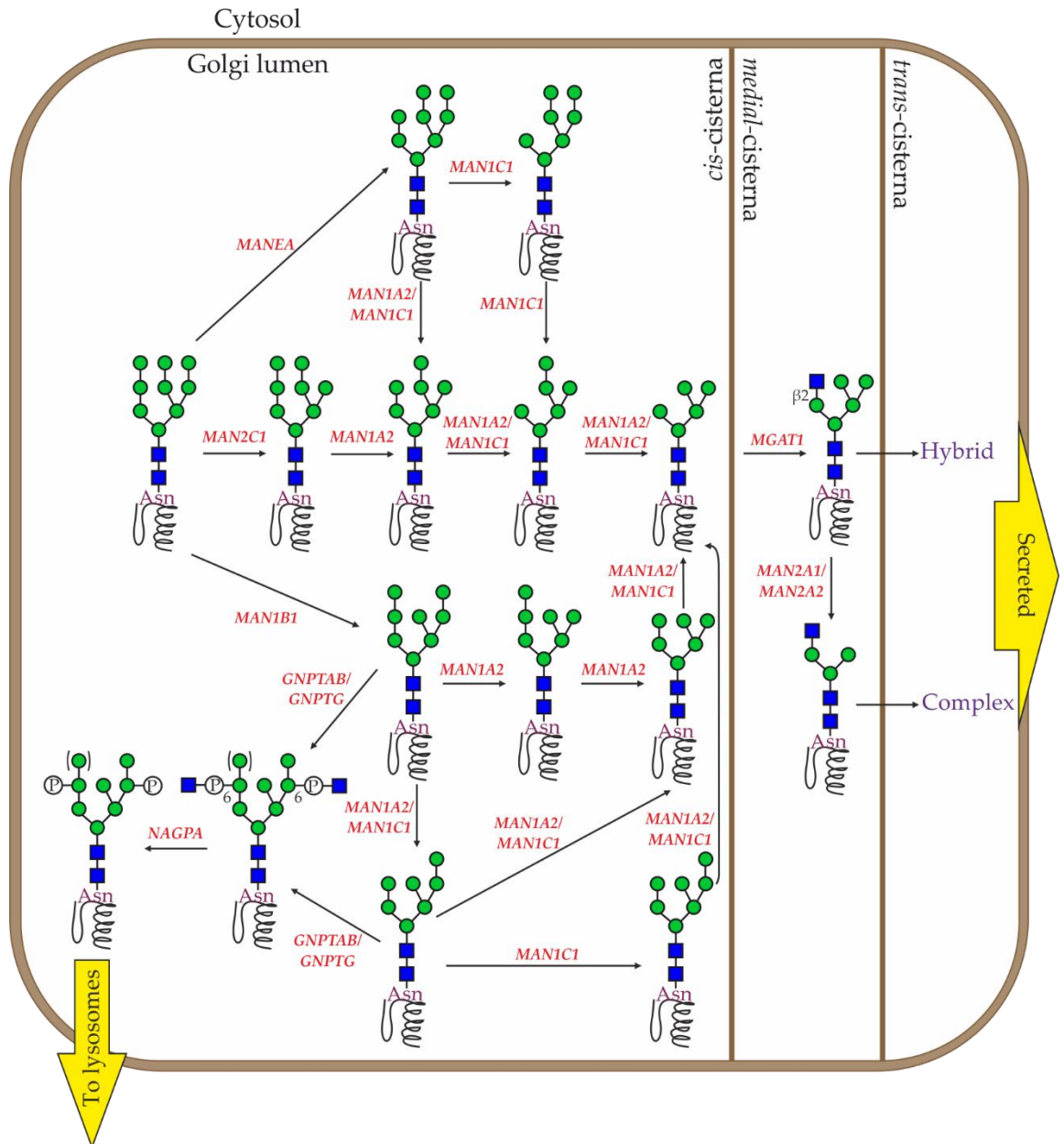


Figure 1.iv. N-glycan processing in the Golgi apparatus. Glossary: *NAGPA* encodes *N*-acetylglucosamine-1-phosphodiester α -*N*-acetylglucosaminidase; *GNPTAB* encodes *N*-acetylglucosamine-1-phosphotransferase; *MANEA* encodes *endo*- α -mannosidase; *MAN1A2* encodes Golgi mannosidase IB; *MAN1C1* encodes Golgi mannosidase IC; *MAN1B1* encodes ER mannosidase I; *MAN1C1* encodes α -mannosidase 2C1; *MAN2A1/MAN2A2* encodes Golgi α -mannosidase II.

CHAPTER 2: Family GH125 Exo- α -1,6- Mannosidases

2.1 Abstract

Pathogenic bacteria contain virulence factors that allow them to colonise within a host organism and potentially cause harm. They have evolved to use host N-glycans either on the surface of cells or on secreted proteins as a source of nutrients for growth and essential cellular processes. Some of the well-documented enzymes involved in the N-glycan processing pathway are present on the cell surface of the bacteria and attack $\text{Man}_9\text{GlcNAc}_2$ glycoproteins. A model for complete N-glycan depolymerisation was proposed after the discovery of a specific cytoplasmic *exo- α -1,6-mannosidase*, CAZy family GH125. GH38 and GH125 reduce $\text{Man}_5\text{GlcNAc}$ and $\text{Man}_3\text{GlcNAc}$ to mannose monosaccharides and ManGlcNAc , respectively, in the cytoplasm. The substrate, α -1,6-mannobiose, was observed to bind in an ${}^{\circ}\text{S}_2$ Michaelis-complex conformation without the requirement for assisted distortion by a metal ion. The powerful mannosidase inhibitor, mannoimidazole, was discovered to mimic the transition state conformation of $B_{2,5}$; both results providing evidence for an ${}^{\circ}\text{S}_2 \rightarrow B_{2,5} \rightarrow {}^1\text{S}_5$ conformational itinerary. Understanding the conformational itinerary of the enzyme is important for discerning catalytic mechanisms and developing conformational mimics to inhibit bacterial colonisation.

2.2 Family GH125 *exo- α -1,6-mannosidases* introduction

2.2.1 Pathogenic bacteria can utilise host glycans as a nutrient source

A diverse community of commensals and quiescent pathogens colonise the tracts of multicellular eukaryotic organisms. *Streptococcus pneumoniae* and *Clostridium perfringens* are (latent) pathogenic gram-positive bacteria found in the human upper respiratory tract and intestinal tract, respectively (reviewed in Henriques-Normark and Tuomanen, 2013). The negative impact of colonisation (shifting the balance towards one microorganism) can result in mild diseases, for example, food poisoning and sinusitis, or severe diseases, such as pneumonia, septicaemia and meningitis (reviewed in Bosch et al., 2013).

Glycans form important structural architecture components for many bacterial species, from cell walls and capsules to pili and flagella. An anionic extracellular capsule prevents clearance by the host's mucus by electrostatic repulsion and their ability to transit to the epithelial surface (Nelson et al., 2007). A capsule is composed of oligosaccharide repeating

units in equimolar ratios interspersed by O-acetyl, phosphoglycerol and pyruvyl acetal substitutions (Kalelkar et al., 1997). There are >90 different serotypes that differ by the composition of the carbohydrate repeats, scissile linkages and modifications (reviewed in Geno et al., 2015).

Virulence factors (cytoplasmic, membrane-associated or secreted) assist the bacteria in colonisation, immunoevasion, immunosuppression and acquirement of nutrients from the host. The major virulence factor is the polysaccharide capsule; there have been >380 genes identified to be important in virulence (Hava and Camilli, 2002; Lau et al., 2001; Polissi et al., 1998). Between 20-40% of isolates are resistant to the drugs currently being administered to combat pneumococcal diseases (including β -lactams, macrolides, lincosamides, fluoroquinolones, tetracyclines and trimethoprim-sulfamethoxazole) (reviewed in Cherazard et al., 2017; Stacevičienė et al., 2016). Penicillin resistance has arisen from heterologous recombination of genes encoding penicillin binding proteins and, due to the nature of the recombination event, has progressed from low level resistance to high level resistance. Therefore, identification of novel virulence factors and the development of drugs and vaccines, is important in the race against antibiotic resistance.

As with many organisms, carbohydrates play a particularly important role in *S. pneumoniae*. Around 30% of its transporters have been shown to, or are predicted to be involved in the translocation of carbohydrates and 10 genes have been identified that encode extracellular glycoside hydrolases (Tettelin et al., 2001). Another strategy that enables *S. pneumoniae* to colonise is its ability to modify and degrade host cell glycans and utilise the monosaccharides as a nutrient source for growth (in the upper respiratory tract free sugars are scarcely found), in biofilm formation, competition with other bacteria and adherence to the epithelium (reviewed in King, 2010). There are many processes that utilise different combinations of GHs to cleave various carbohydrate linkages; in this chapter, I will focus on family GH125 enzymes, which cleave the *exo*- α -1,6-mannosidic bonds of Man₃GlcNAc and Man₂GlcNAc and are the last enzyme family in the N-glycan degradation pathway of *S. pneumoniae*.

2.2.2 Discovery of α -1,6-mannosidases

Before 2011, there was an absence of a known α -mannosidase that could efficiently and exclusively hydrolyse an α -1,6-mannosidic bond, Figure 2.1. Family GH38 enzymes, (*medial*) Golgi α -mannosidase II, primarily hydrolyse the α -1,3-glycosidic bond between M7 and M4 prior to the α -1,6-glycosidic bond between M6 and M4, Figure 1. (Rabouille et al., 1999). Due to the low efficiency of the α -1,6-cleavage catalysed by GH38 in *Streptococcus pyogenes*, it was postulated that the bacteria may express another gene encoding a partner GH (Suits et al., 2010). A previously uncharacterised gene (SP_2144) from *S. pneumoniae*, TIGR4, was suggested to be active on carbohydrates due to its position in a locus containing multiple α -mannosidases, an α -fucosidase and an β -hexosaminidase (Gregg et al., 2011). A second gene from ATCC 13124 from *C. perfringens* was also examined because it shares 40% sequence identity with SP_2144.

Predicted alignment of the $(\alpha/\alpha)_6$ -barrel fold between the GHX enzyme and GH15 enzymes suggested α -glucanase activity (Gregg et al., 2011). Glucoamylases have α -1,4-glucosidic cleavage activity but they can also cleave α -1,6-linkages when the next linkage in the oligosaccharide is α -1,4 (Brown and Brown, 1965). Also within this family are glucodextranases that cleave α -1,6-glucosidic linkages (Walker and Pulkownik, 1973). However, aligning the amino acid sequence of the GHX enzyme and a member of family GH15 prompted a new classification of GH enzymes, the family GH125 hydrolases.

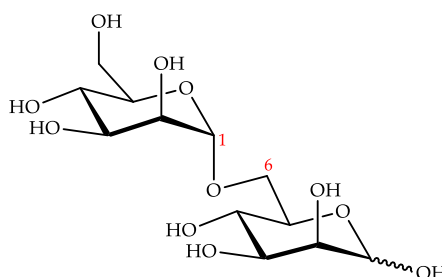


Figure 2.1. Schematic of α -1,6-mannobiose. Linkage between C1 and C6 numbered in red.

2.2.3 Specificity

Boraston and colleagues performed three tests to investigate the specificity of GH125 on different carbohydrates and glycosidic linkages. Firstly, the activity on a variety of different arylglycosides was tested by monitoring the production of 2,4-dinitrophenolate

at 400 nm. Very low activity was observed with 2,4-dinitrophenyl α -D-mannopyranoside (DNP-Man) (*Sp*GH125 $k_{cat}/K_M = 0.13 \text{ min}^{-1} \text{ mM}^{-1}$ and *Cp*GH125 $k_{cat}/K_M = 0.58 \text{ min}^{-1} \text{ mM}^{-1}$), Figure 2.2; suggesting that GH125 can tolerate an aglycon group (in this context means a non-carbohydrate moiety) in the +1/+2 subsite.

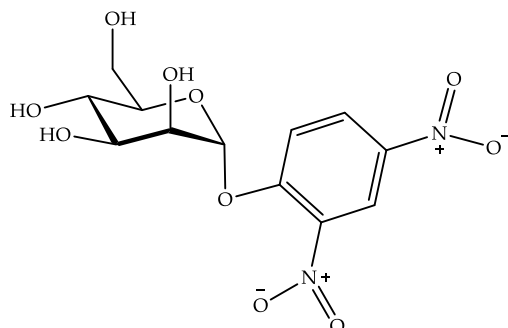


Figure 2.2. Structure of 2,4-dinitrophenyl α -D-mannopyranoside.

Secondly, a variety of disaccharides and trisaccharides with different mannosidase linkages (α -1,2-, α -1,3-, α -1,6-mannobiose and α -1,3- α -1,6-mannotriose) were tested for activity against *Sp*GH125 and *Cp*GH125 by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), Figure 2.3a. *Sp*GH125 and *Cp*GH125 could only accept α -1,6-mannobiose as a substrate.

Lastly, GH125 was tested against high mannose N-glycans by capillary electrophoresis, Figure 2.3b. Man₉, Man₅ and Man₃ were labelled by reductive amination with 8-aminopyrene-1,3,6-trisulfonic acid, which carries three negative charges from sulfonic acid groups. The anionic labelled oligosaccharides are separated by electrophoretic mobility through a capillary from the cathode to the anode passing through a fluorescence detector (reviewed in Ruhaak et al., 2010). Comparison of the products with standards revealed that GH125 was unable to cleave any of the oligosaccharides. When Man₃ was pre-incubated for 3 hours with an α -1,2- α -1,3-mannosidase to cleave M5, GH125 was able to cleave M4, indicating it is a specific *exo*- α -1,6-mannosidase.

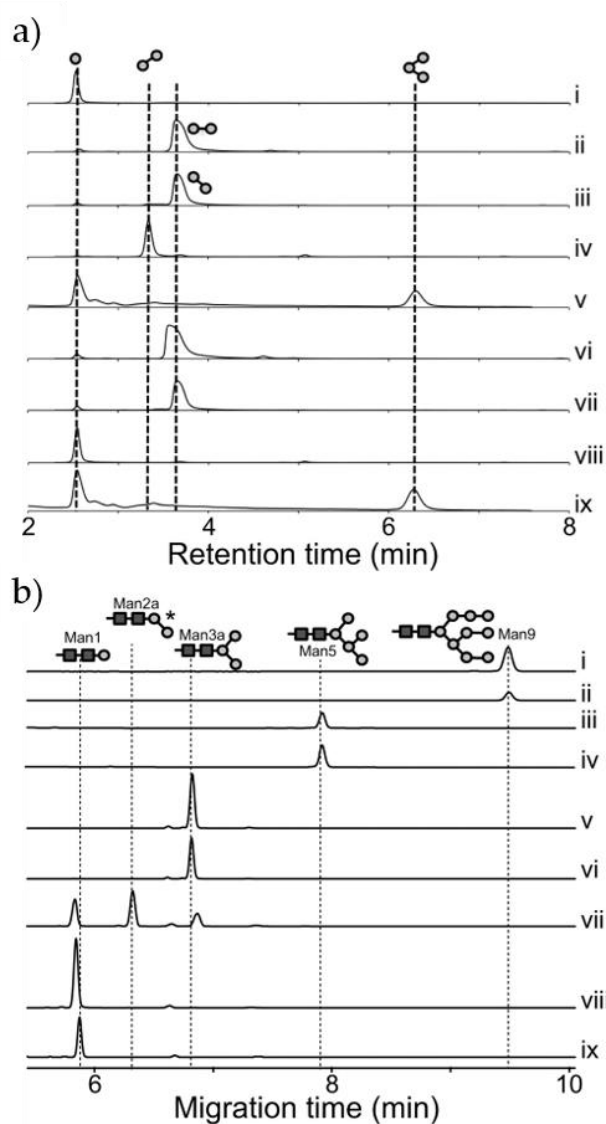


Figure 2.3. Specificity of GH125 was determined by testing the cleavage of disaccharides, trisaccharides and high-mannose glycans. HPAEC-PAD elution profile of the standards: (i) mannose, (ii) α -1,2-mannobiose, (iii) α -1,3-mannobiose, (iv) α -1,6-mannobiose and (v) α -1,3- α -1,6-mannotriose. The products after incubation with *Sp*GH125 and: (vi) α -1,2-mannobiose, (vii) α -1,3-mannobiose, (viii) α -1,6-mannobiose and (ix) α -1,3- α -1,6-mannotriose. b) The capillary electrophoresis traces of different oligosaccharides with *Sp*GH125 and *Cp*GH125. *Man*₉ treated with (i) *Sp*GH125 and (ii) *Cp*GH125; *Man*₅ treated with (iii) *Sp*GH125 and (iv) *Cp*GH125; *Man*_{3a} treated with (v) *Sp*GH125 and (vi) *Cp*GH125; *Man*_{3a} treated with (vii) α -1,2- α -1,3-mannosidase; pre-treatment of *Man*_{3a} with α -1,2- α -1,3-mannosidase and then (viii) *Sp*GH125 and (ix) *Cp*GH125. The asterisk next to *Man*_{2a} indicates that the peak is inferred due to the absence of a standard. Figure taken from (Gregg et al., 2011).

2.2.3.1 Activity assays

As stated in section 2.2.3, GH125 shows low activity towards aryl α -mannopyranosides and binds substrates more efficiently when the substrate has an aglycon that can bind in the +1 subsite. To facilitate effective and robust kinetic assays, a fluorogenic disaccharide substrate was synthesised with a 400-fold higher efficiency of hydrolysis than DNP-Man (Deng et al., 2013). 4-methylumbelliferone α -D-mannopyranosyl-(1 \rightarrow 6)- β -D-mannopyranoside (α -Man- β -Man-4MU) lies across the -1, +1 and +2 subsites with 4MU in the +2 subsite. 4MU emits light at a wavelength of 460 nm when excited by light at 365 nm. The second mannose, which sits in the +1 subsite, mimics the natural substrates, for example host glycans and α -1,6-mannobiose. In a coupled enzyme fluorescent assay, GH125 cleaves α -D-mannopyranosyl (α -Man) followed by hydrolysis of 4-methylumbelliferone β -D-mannopyranoside (β -Man-4MU) by a GH2 β -mannosidase to release 4-methylumbelliferone (4MU), Figure 2.4 (Deng et al., 2013). The β -mannosidase from *Helix pomatia* did not show any activity on α -Man- β -Man-4MU. CpGH125 cleaved (α -Man- β -Man-4MU) with a $k_{cat}/K_m = 0.77 \pm 0.05 \text{ min}^{-1} \mu\text{M}^{-1}$ and *H. pomatia* β -mannosidase cleaved with a $k_{cat}/K_m = 9.6 \pm 0.5 \times 10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$.

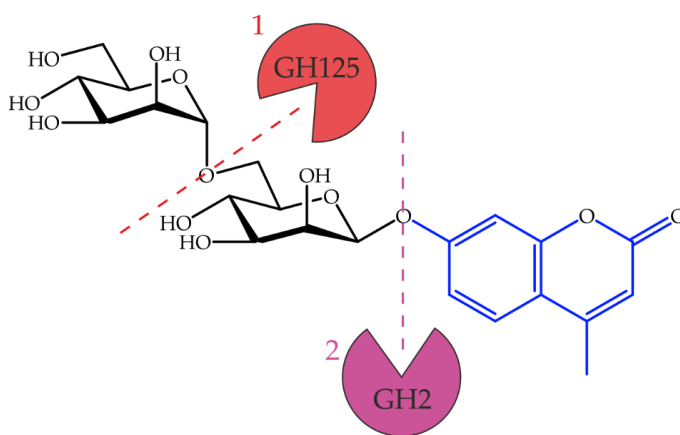


Figure 2.4. Coupled fluorogenic assay for GH125. GH125 cleaves the first α -1,6-linked mannose, only then is GH2 able to cleave the second β -1,4-linked mannose to release 4-methylumbelliferone, which emits light at 460 nm.

2.2.4 N-glycan processing model.

Anchored to the peptidoglycan layer and exposed at the surface of *S. pneumoniae* are several GH families specialising in host N-glycan depolymerisation, Figure 2.5. Complex glycans can be broken down by the following sequence of enzymes (Robb et al., 2017).

NanA is a sialidase that is able to cleave the α -2,3- and α -2,6-linkages between neuraminic acid and galactose (Cámara et al., 1994). BgaA is an *exo*- β -D-galactosidase that cleaves the β -1,4-linkage between galactose and GlcNAc (Zähner and Hakenbeck, 2000). StrH is an *exo*- β -D-N-acetylglucosaminidase that cleaves the β -1,2-linkage between GlcNAc and mannose resulting in Man₃GlcNAc₂ (Clarke et al., 1995). High mannose glycans (Man₉GlcNAc₂) can be cleaved by SpGH92 α -1,2-mannosidase (Robb et al., 2017). The monosaccharides released by these enzymes are transported into the cytoplasm by dedicated ABC and phosphotransferase system transporters (reviewed in Buckwalter and King, 2012).

Both Man₅GlcNAc₂ and Man₃GlcNAc₂ glycans are substrates for EndoD. EndoD cleaves the β -1,4-linkage between the two GlcNAc residues releasing the glycan from the glycoconjugate (Muramatsu et al., 2001). Man₅GlcNAc and Man₃GlcNAc are transported into the cytoplasm by the ABC transporter ABC_{NG}, which is composed of an extracellular solute binding protein NgtS, two membrane spanning permeases, NgtP1 and NgtP2, and two ATPases located in the cytoplasm, Figure 2.5 (Robb et al., 2017).

It has been shown that GH125 requires the removal of M5 and M7 (α -1,3-linked mannose) from high-mannose glycans for efficient hydrolysis. The location of the gene encoding *exo*- α -1,6-mannosidase is always found adjacent to the gene encoding GH38 in Firmicutes (except *C. perfringens*, where the two GH38 genes are elsewhere in the genome) (Gregg et al., 2011). Therefore, it is proposed that GH38 cleaves the N-glycan in the cytoplasm before GH125 for complete N-glycan depolymerisation, Figure 2.5 (Robb et al., 2017).

N-glycan degradation by pathogenic bacteria in plants has been recently demonstrated (Dupoirion et al., 2015). *Xanthomonas campestris* pv. *Campestris* causes black rot disease on *Brassica* plants. It can cleave the most widespread N-glycan motif in plants, Man₃XylGlcNAc₂Fuc. The genes encoding the proteins that cleave the N-glycan are clustered and transcribed polycistronically. The removal of the α -1,3-mannose by NixK (GH38) and β -1,2-xylose by NixI (GH3) is necessary before NixJ (GH125) can cleave the α -1,6-mannose (Dupoirion et al., 2015).

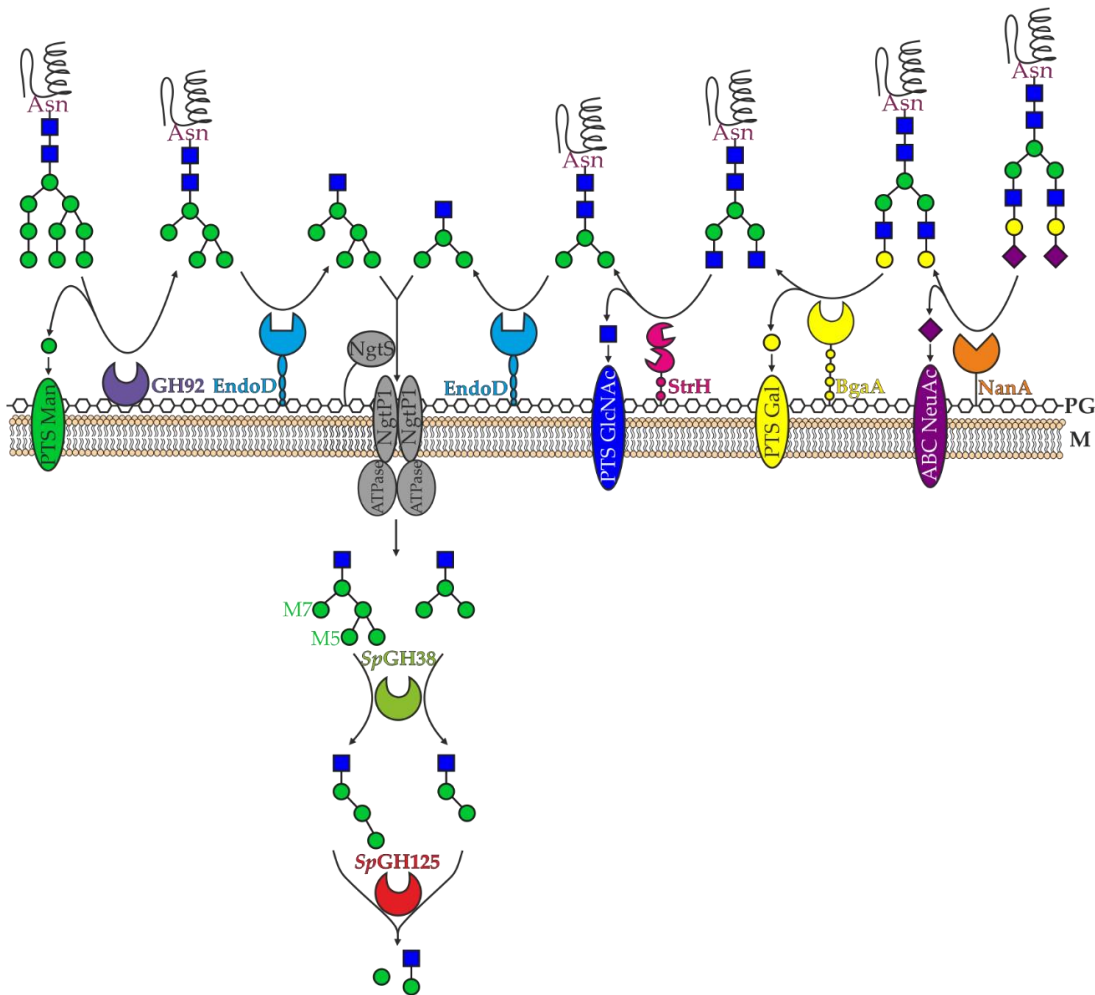


Figure 2.5. Complete N-glycan degradation pathway in *Streptococcus pneumoniae*. Host N-glycan degradation by NanA, BgaA, StrH, EndoD and GH92 on the extracellular cell wall. They are anchored into the peptidoglycan (PG) layer by LPXTG anchoring motifs. The monosaccharides are transported through the cytoplasmic membrane (M) by ABC and phosphotransferase transporters. $\text{Man}_5\text{GlcNAc}_2$ and $\text{Man}_3\text{GlcNAc}_2$ are transported into the cytoplasm by ABC_{NG} .

The cleavage of the human host glycome is also carried out by symbiotic microbiota in the gut, in particular *Bifidobacteria* and *Bacteriodes thetaiotaomicron*, which encode a vast number of mannosidases and other GHs (*B. thetaiotaomicron* has >170 GH proteins) (O'Callaghan and van Sinderen, 2016; Xu et al., 2003). *Bacteriodes thetaiotaomicron* contains 23 different family GH92 α -mannosidases that can cleave α -1,2- α -1,3- α -1,4 and α -1,6-linked glycosidic bonds of host N-glycans in the human gut (Zhu et al., 2010). *Bt3994* can hydrolyse α -1,6-linked mannose from $\text{Man}_5\text{GlcNAc}_2$ to $\text{Man}_4\text{GlcNAc}_2$ and $\text{Man}_3\text{GlcNAc}_2$

to Man₂GlcNAc₂. *B. thetaiotaomicron* also contains a surface *endo*- α -1,6-mannase belonging to family GH76, Bt3792 (Cuskin et al., 2015).

2.2.5 Structure

Crystal structures of both *Sp*GH125 and *Cp*GH125 revealed that both possess an (α/α)₆-fold, Figure 2.6a. The barrel is formed of thirteen antiparallel α -helices, seven of which are on the outside of the barrel (α 1, α 2, α 4, α 6, α 8, α 10 and α 12) and paired to six helices on the inside of the barrel (α 3, α 5, α 7, α 9, α 11, α 13), Figure 2.6b. Between alternate outer and inner α -helices, for example α 2 and α 3, are longer loops with β -sheet secondary structure (most commonly two antiparallel β -strands. Between alternate inner and outer helices, for example α 3 and α 4, is a short loop of 2-7 residues in length.

Located in the middle of the barrel is a tunnel active site with the dimensions $\sim 35 \times 9$ Å. The carbohydrate substrate binds at the top of the barrel on the β -sheet face and is blocked from the lower half of the barrel by Trp410 which lies across it (~ 15 Å into the length of the barrel), Figure 2.6c.

The active site contains a conserved and defined -1 and +1 subsite that are separated by ~ 6.5 Å, Figure 2.6c. The non-reducing end of mannan binds in the -1 subsite and the reducing end binds in the +1 subsite (Gregg et al., 2011). A third subsite, +2, is ~ 6.5 Å from the +1 subsite. It has limited definition due to an absence of residues that are within bonding distance to the hydroxyls of the carbohydrate but was observed in the crystal complexes of *Cp*GH125 with α -1,6-mannotriose and *Sp*GH125 with α -1,6-mannobiose (Gregg et al., 2011). The closest residue to mannose in the +2 subsite is CZ2 of Trp116, 3.2 Å away from O5. Therefore, GH125 should be able to tolerate a number of different carbohydrate groups within the +2 subsite, such as the GlcNAc from Man₂GlcNAc₂.

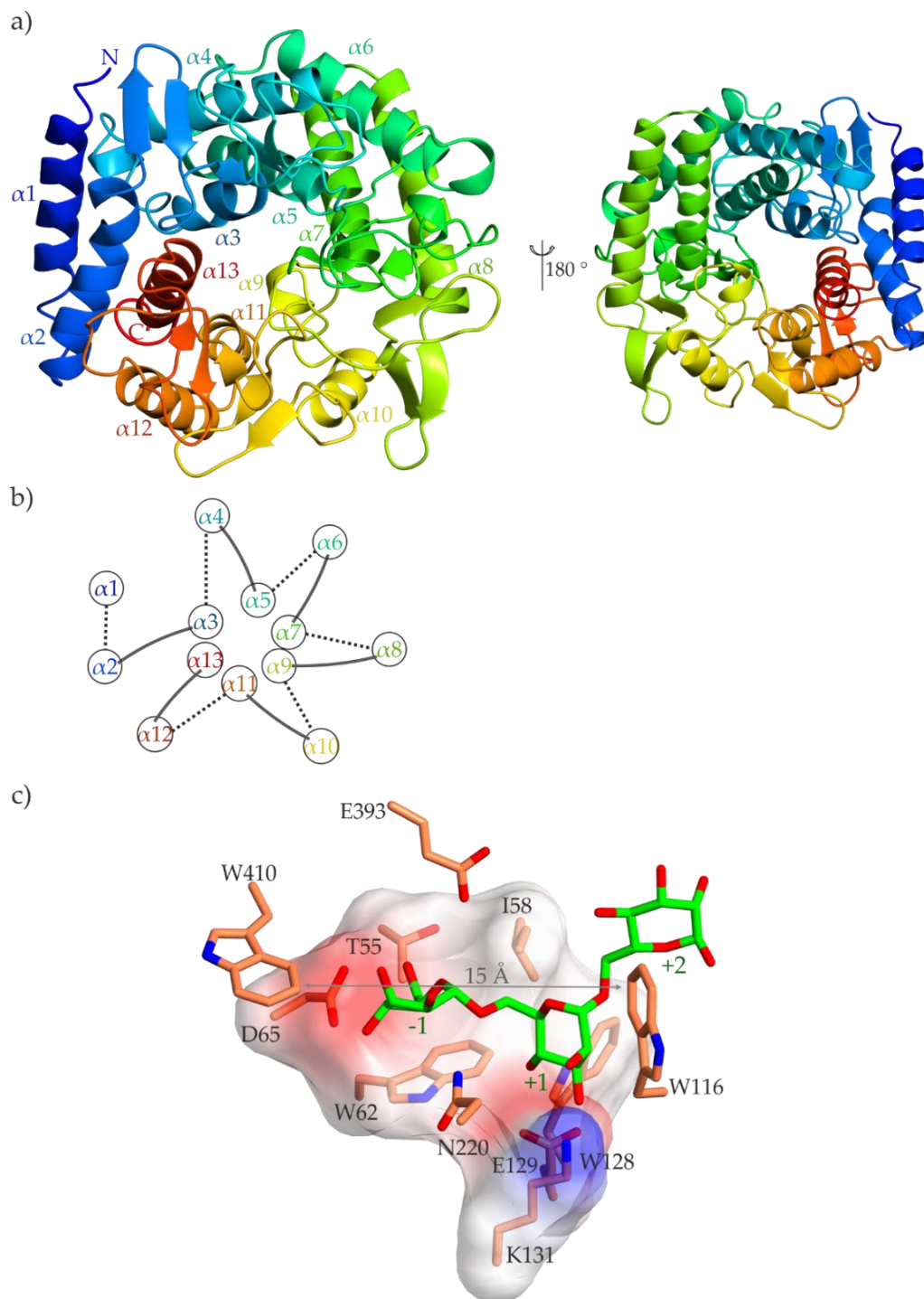


Figure 2.6. Overall $(\alpha/\alpha)_6$ -barrel fold of GH125 and catalytic subsites. a) Top (left) and bottom (right) view of GH125 TIM barrel. The helices are labelled in ascending order in a clockwise direction from the N-terminus (blue) to the C-terminus (red). b) A schematic diagram of the position of each helix showing the connecting short loop (black dotted line) on the bottom face and the long loop with β -strands (grey curved line) on the top face. c) α -1,6-mannotriose in green lies across the three subsites of GH125. The residues coloured in coral line the active site with Trp410 and Trp116 marking the end of the active site and the top of the barrel, respectively.

2.2.6 Mechanism of catalysis

The key catalytic residues in the inverting catalytic mechanism are Asp220 and Glu393 from *C. perfringens* and the equivalent residues Asp218 and Glu393 from *S. pneumoniae*. Asp220 acts as the general acid, *syn*-protonator, donating a proton to the C1 substituent (Heightman and Vasella, 1999). Glu393 acts as the general base by assisting in the attack of the anomeric carbon and displacement of the leaving group by coordinating a water molecule. The catalytic residues were proposed after alignment of the structures of GH125 and GH15 glucodextranase, Figure 2.7a (Gregg et al., 2011). The conservation of residues in the -1 subsite was higher than that of the +1 subsite. Since GH15 catalyses the hydrolysis of oligosaccharides that contain glucose, the equatorial O2 of glucose forms a hydrogen bond to Arg567; in comparison, the axial O2 of mannose forms a hydrogen bond to Asn302 (*Cp*GH125). The catalytic residues were also inferred from the structure of *Cp*GH125 in complex with thiomannobiose. The general acid was 3.0 Å away from the sulfur and the general base was 2.4 Å away from the nucleophilic water poised 3.2 Å away from C1.

Unlike other α -mannosidases, GH125 catalysis does not require any assistance from metal ions. Upon the addition of 10 mM ethylenediaminetetraacetic acid (EDTA) to enzyme kinetic assays, the activity of the enzyme for 1 mM DNP-Man was not altered (Gregg et al., 2011). EDTA is a chelating agent that forms four bonds to a metal ion (two from nitrogen atoms and two from oxygens of different carboxyl groups). Furthermore, there was no indicative density observed in the active site structure. In contrast, GH38 requires a Zn^{2+} ion and, GH47 (Figure 2.7b) and GH92 mannosidases require a Ca^{2+} ion (Numao et al., 2003; Zhu et al., 2010). They function to coordinate the O2 and O3 hydroxyls of the ligand to stabilise the conformational distortion necessary for catalysis, see section 3.2.5.2.

1H -NMR was used to monitor the hydrolysis of methyl 6-*O*-(α -D-mannopyranosyl)- β -D-mannopyranoside (Man α 1-6Man β 1-OMe) catalysed by *Sp*GH125 and *Cp*GH125 (Gregg et al., 2011). The C1 hydroxyl configuration of the product, mannose, was observed over a time course of 660 minutes, Figure 2.7c. The presence of a hemiacetal β -configured mannose product was identified after 14.5 minutes. However, a small peak was also observed for an α -configured mannose product. To decipher which orientation is formed first, the anomeric ratio was plotted against time. There was a delay in the production of

α -mannose suggesting that β -mannose is formed first and that it undergoes spontaneous mutarotation to form α -mannose.

As mentioned in chapter 1, the -1 subsite is catalytically relevant to observe the conformations of different substrates and inhibitors. Complexes of CpGH125 with thiomannobiose and SpGH125 with DMJ revealed a carbohydrate conformation of 4C_1 in the -1 subsite, Figure 2.7d (Gregg et al., 2011). This is a low energy state and suggested that upon enzyme binding the substrate had not distorted. This was an unusual conformation to observe for a mannosidase enzyme.

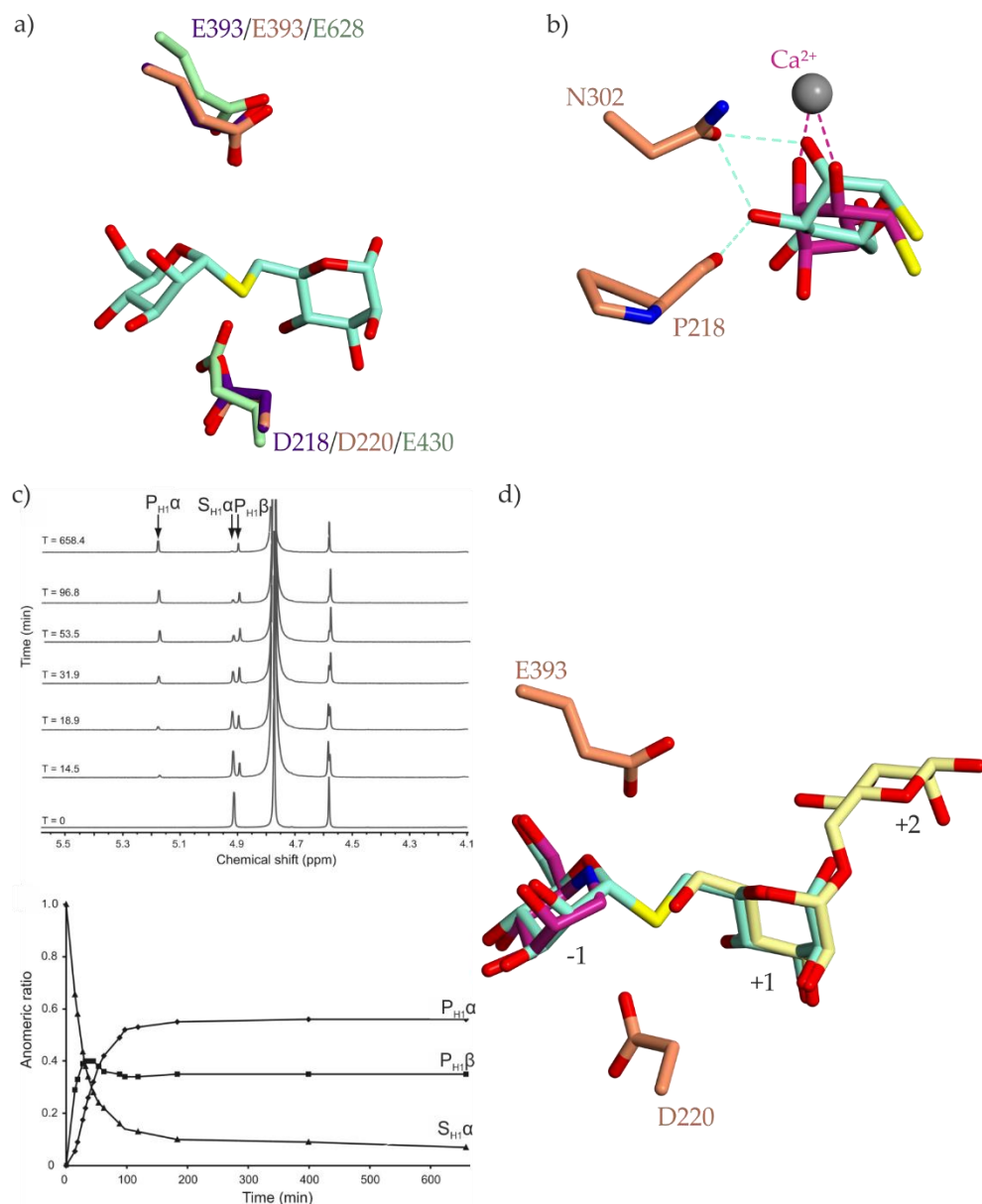


Figure 2.7. Evidence for the catalytic mechanism of GH125. a) Overlay of catalytic residues from Glucodextranase GH15 (PDB ID: 1ULV) in green, *Sp*GH125 (PDB ID: 3QPF) in purple and *Cp*GH125 (PDB ID: 3QT9) in coral with thiomannobiose in blue/green across the -1 and +1 subsites (left to right). b) Overlay of the residues that coordinates the O2 and O3 hydroxyls of thiomannobiose in blue/green (*Cp*GH125 (PDB ID: 3QT9) in coral) and the calcium ion of *Ck*GH47 (PDB ID: 4AYP) that coordinates the O2 and O3 hydroxyls of DMJ in pink. c) Top: spectra of the cleavage of $\text{Man}\alpha 1\text{-6Man}\beta 1\text{-OMe}$ monitoring the configuration of the anomeric proton; bottom: anomeric ratio of α/β -product and substrate. d) Substrate/inhibitor complexes show three binding sites. *Cp*GH125 in coral with thiomannobiose in blue/green (PDB ID: 3QT9) and α -1,6-mannobiose in yellow (PDB ID: 3QSP) and DMJ (*Sp*GH125 PDB ID: 3QRY) in pink.

2.2.7 Inhibition of α -1,6-mannosidases

Thiosaccharides mimic the natural ligand of GHs and are expected to bind to the enzyme in a Michaelis complex conformation. Most glycoside hydrolases (with the possible exception of GH84 (De Leon et al., 2017; Macauley et al., 2005a)) are unable to cleave a C-S bond due to the poor hydrogen bonding ability of thiols, consequently, *O,S*-acetals cannot undergo general acid catalysis (Fife and Przystas, 1980).

The first structure of a GH obtained with a bound thiosaccharide was of endoglucanase I from *Fusarium oxysporum* (Sulzenbacher et al., 1996). Synthesised as five consecutive D -glucopyranosyl carbohydrates linked by β -1,4-linkages (with the C1 position of the fifth mannose containing an alpha configured methoxy group \equiv methyl α - D -glucopyranoside) (reviewed in Schou et al., 1993). The unit in the -1 subsite was deciphered to be distorted to a ${}^1\text{4B}/{}^1\text{S}_3$ conformation after analysis of different inhibitor structures and QM/MM analysis (Knott et al., 2014; Sulzenbacher et al., 1996).

Since then, various natural carbohydrates and inhibitors have been combined as tools for probing substrate binding and catalysis, for example, mixed linkage thiopentasaccharides with only one sulphur between the carbohydrates that binds in the -1 and +1 subsites (Money et al., 2008). Methyl S-(α - D -mannopyranosyl)-(1-6)- α - D -mannopyranose (thiomannobiose) is a non-hydrolysable mannose substrate mimic that has a sulfur atom substituted in the position of the oxygen at the 1,6 position, Figure 2.8 It has been observed in a distorted ${}^3\text{S}_1$ conformation upon binding to GH47 α -1,2-mannosidase (Thompson et al., 2012). However, in a complex with GH92 α -mannosidase the substrate mimic was observed in a ground state ${}^4\text{C}_1$ undistorted conformation (Zhu et al., 2010). GH47, GH92 and GH125 all act by an inverting mechanism with the catalytic acid in a *syn*-trajectory. Therefore, thiosaccharides may not be mechanistically informative but they are informative for observing ligand interactions and deciphering the catalytic apparatus.

1-deoxymannojirimycin (DMJ) is an azasugar and an analogue of mannose. It has a ring nitrogen (piperidine) instead of an oxygen (tetrahydropyran) and it lacks a catalytically essential hydroxyl group from C1. DMJ has been observed to bind in the -1 non-reducing subsite of the GH125 mannosidase active site in an undistorted conformation. For further detail on the effects of mannosidase inhibition see section 3.2.5.1.

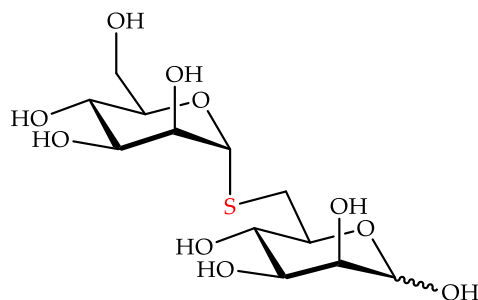


Figure 2.8. Structure of thiomannobiose.

2.2.8 A summary of the results in this chapter

Previous structural studies of α -1,6-mannosidase from *C. perfringens* and *S. pneumoniae* in complex with a non-hydrolysable substrate mimic, α -1,6-thiomannobiose, and with DMJ, portrayed an undistorted and unusual 4C_1 conformation. However, theoretical QM/MM metadynamics (FEL of 1-thio- α -mannosyl vs α -mannosyl) and X-ray crystallography (GH125-D220N in complex with α -1,6-mannobiose and α -1,6-mannotriose) unveiled a preference for an oS_2 Michaelis complex conformation, described in the first paper of this chapter.

Evidence for a $B_{2,5}$ transition state was provided by a structure of CpGH125 in complex with mannoimidazole, a powerful transition state mimic due to its sp^2 -hybridised anomeric centre, and the relative affinities of the three subsites (-1, +1 and +2) are described in the second paper of this chapter. Resulting from these two papers, the proposed conformational itinerary for GH125 enzymes is ${}^oS_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$.

**Computational Design of Experiment Unveils the
Conformational Reaction Coordinate of GH125 α -
Mannosidases**

Declaration

I contributed structural data to this publication from crystallographic studies of CpGH125 in complex with α -1,6-mannobiose and α -1,6-mannotriose. Santiago Alonso-Gil at the University of Barcelona generated the conformational free-energy landscapes of α -mannosyl and 1-thio- α -mannosyl in the -1 subsite of CpGH125. He also conducted QM/MM simulations to observe the reaction itinerary. Pearl Z. Fernandes at the University of Melbourne conducted the synthesis of ligands for GH125. Prof. Spencer J. Williams, Prof. Carme Rovira and Prof. Gideon J. Davies conceived the initial idea for the paper and wrote the final draft of the manuscript taking into consideration my input describing the crystal structure, the methodology and the data statistics.



Signed

Prof. Gideon J. Davies



Signed

Prof. Carme Rovira

Reference

Alonso-Gil, S., Males, A., Fernandes, P. Z., Williams, S. J., Davies, G. J. and Rovira, C. (2017). Computational design of experiment unveils the conformational reaction coordinate of GH125 α -mannosidases. *J. Am. Chem. Soc.* 139, 1085-1088.

Supplementary information – pg 219



Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α -Mannosidases

Santiago Alonso-Gil,[†] Alexandra Males,[‡] Pearl Z. Fernandes,[§] Spencer J. Williams,[§] Gideon J. Davies,^{*,‡} and Carme Rovira^{*,†,||}

[†]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain

[‡]York Structural Biology Laboratory, Department of Chemistry, The University of York, YO10 5DD York, United Kingdom

[§]School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria 3010, Australia

^{||}Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

* Supporting Information

ABSTRACT: Conformational analysis of enzyme-catalyzed mannoside hydrolysis has revealed two predominant conformational itineraries through $B_{2,5}$ or 3H_4 transition-state (TS) conformations. A prominent unassigned catalytic itinerary is that of *exo*-1,6- α -mannosidases belonging to CAZy family 125. A published complex of *Clostridium perfringens* GH125 enzyme with a non-hydrolyzable 1,6- α -thiomannoside substrate mimic bound across the active site revealed an undistorted 4C_1 conformation and provided no insight into the catalytic pathway of this enzyme. We show through a purely computational approach (QM/MM metadynamics) that sulfur-for-oxygen substitution in the glycosidic linkage fundamentally alters the energetically accessible conformational space of a thiomannoside when bound within the GH125 active site. Modeling of the conformational free energy landscape (FEL) of a thioglycoside strongly favors a mechanistically uninformative 4C_1 conformation within the GH125 enzyme active site, but the FEL of corresponding O-glycoside substrate reveals a preference for a Michaelis complex in an 0S_2 conformation (consistent with catalysis through a $B_{2,5}$ TS). This prediction was tested experimentally by determination of the 3D X-ray structure of the pseudo-Michaelis complex of an inactive (D220N) variant of *C. perfringens* GH125 enzyme in complex with 1,6- α -mannobiose. This complex revealed unambiguous distortion of the -1 subsite mannoside to an 0S_2 conformation, matching that predicted by theory and supporting an ${}^0S_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$ conformational itinerary for GH125 α -mannosidases. This work highlights the power of the QM/MM approach and identified shortcomings in the use of nonhydrolyzable substrate analogues for conformational analysis of enzyme-bound species.

The conformational itineraries employed by glycoside hydrolases to perform nucleophilic substitution reactions at the anomeric center of glycosides have been the topic of sustained interest since the mid-1990s.¹ Physical organic studies

have provided compelling evidence that glycosidase-catalyzed glycoside cleavage occurs through oxocarbenium ion-like transition states (TSs) with significant partial double-bond character between the anomeric carbon and the ring oxygen.² Sinnott postulated that glycosidases must react through TSs in one of 4 major conformations: 4H_3 and 3H_4 half chairs (or their related envelopes) or $B_{2,5}$ and ${}^{2,5}B$ boats. The topological relationships of such conformations are conveniently visualized in a Mercator representation (Figure 1).

By the principle of least nuclear motion,³ the conformations of the ground states of the enzymatic Michaelis complex, products, and (if relevant) associated intermediates must flank the TSs. Studies over the last 20 years have identified that all four major TS conformations are co-opted by various enzymes working across the breadth of stereochemically diverse carbohydrate substrates.¹ As TS mimicry provides a practical blueprint for the development of tight binding inhibitors, analysis of these reaction coordinates is invaluable in the design and application of TS mimics as mechanistic probes and therapeutic agents.⁴

The canvas upon which nature's treasure chest of glycosidases is depicted is the carbohydrate-active enzymes (CAZy) classification.⁵ Enzymes are classified into families by amino-acid sequence (and hence 3D structural) similarity. Of particular interest are the diverse α - and β -mannanases and mannosidases that catalyze the sterically challenged reaction at the crowded anomeric carbon of mannose for which mechanistic insights can inform and enlighten key challenges involved in the chemical synthesis of mannosides.⁶ The α - and β -mannanases are involved in glycan processing within important industrial and biological processes. In the latter, assorted α -mannosidases are involved in N-glycan maturation and processing,⁷ fungal cell-wall biosynthesis⁸ and catabolism,⁹ and other cellular reactions of high interest for therapeutic intervention.

By the CAZy classification, α - and β -mannosidases (both *exo*- and *endo*-acting) populate a large number of GH families: α : 38, 47, 76, 92, 99, and 125; β : 2, 5, 26, 113, 130, and 134. Systematic analysis of the conformational itineraries of these enzyme

Received: November 3, 2016

Published: December 27, 2016

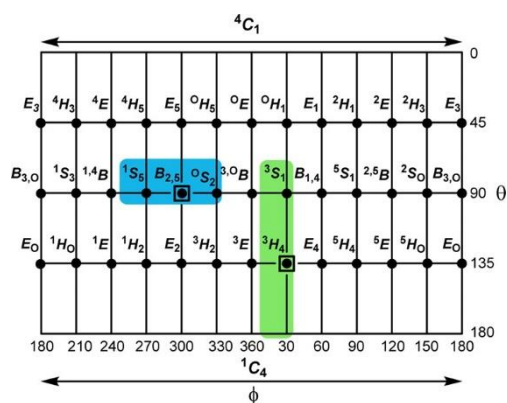


Figure 1. Mercator plot of major canonical conformations of a pyranose ring. The TS conformations (boxed) and associated ground-state conformations of mannosidase conformational itineraries through TSs with $B_{2,5}$ (blue) and 3H_4 (green) conformations.

families, primarily through crystallography of stable species flanking or mimicking the TS(s), has revealed two predominant strategies employed by these catalysts to overcome the challenges of mannoside chemistry (GH99 α -mannosidases are believed to react through an epoxide intermediate¹⁰ and are not discussed further). One group of α - and β -mannosidases belonging to GH families 2,¹¹ 5,¹² 26,¹³ 38,¹⁴ 76,¹⁵ 92,¹⁶ 113,¹⁷ and 130,¹⁸ perform catalysis through a pathway around the 0S_2 - $B_{2,5}$ - 1S_5 region of the conformational space (Figure S2). The other group of GHs include the family GH47 α -mannosidases^{19,20} and the GH134 β -mannanases,²¹ which react in a “ring-flipped” (southern hemisphere) 3S_1 - 3H_4 - 1C_4 conformational arena (Figure S2).

In a seminal work, Gregg et al. reported the creation of GH family 125 based on the discovery of 1,6- α -mannosidase activity for enzymes from *Clostridium perfringens* (CpGH125) and *Streptococcus pneumoniae*.²² This family was shown to operate through an inverting mechanism, and insight into the active site residues was provided through X-ray structures of these enzymes in complex with the nonhydrolyzable substrate mimic 1,6- α -thiomannobiose (Protein Databank (PDB) ID: 3QT9) and deoxymannojirimycin (PDB ID: 3QRY). Despite the 1,6- α -thiomannobiose substrate mimic spanning the active site, the mechanistically informative -1 subsite mannoside residue was observed in an undistorted, ground-state 4C_1 conformation, providing no insight into the conformational itinerary of this family of α -mannosidases. Intrigued by this surprising but uninformative result, we were motivated to investigate further. Although the distortion-free binding of the thiomannoside is surprising, it is not unprecedented. A similar situation was noted in the case of 1,2- α -thiomannobiose bound to a GH92 1,2- α -mannosidase; however, in that case a complex with the TS mimic mannoimidazole provided evidence in support of an ${}^0S_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$ conformational itinerary.¹⁶ In the GH125 case the same approach cannot be applied as the general acid residue is not appropriately situated to allow lateral protonation of the basic mannoimidazole nitrogen, whereas in family GH92 enzymes the orientation of the general acid residue is “anti”²³ to the C1–O5 bond, which enables lateral protonation and binding of this inhibitor. The inability to assign a conformational itinerary to GH family 125 prevents rational application and design of conformationally locked or biased inhibitors selective for this family of biomedically important enzymes. To understand the conformational preferences of thioglycosides

within the active site of CpGH125, we first adopted a computational approach (*ab initio* QM/MM metadynamics)²⁴ to map the conformational free energy landscape (FEL) of the -1 mannoside ring as a function of the Cremer–Pople ring puckering coordinates,²⁵ an approach that has been applied to other GH families.²⁶ We first calculated the conformational FEL for isolated 1-thio- α -mannopyranose (see computational details in the Supporting Information). As previously found for α -mannopyranose,^{19a} the sugar has a preference for 4C_1 conformation, but with other regions of the conformational energy surface energetically accessible most notably the region around the 0S_2 conformation (Figure S1).

When the QM/MM metadynamics approach was applied to the 1-thio- α -mannosyl residue of the complex of CpGH125 with 1,6- α -thiomannobiose the FEL is transformed (Figure 2a) from

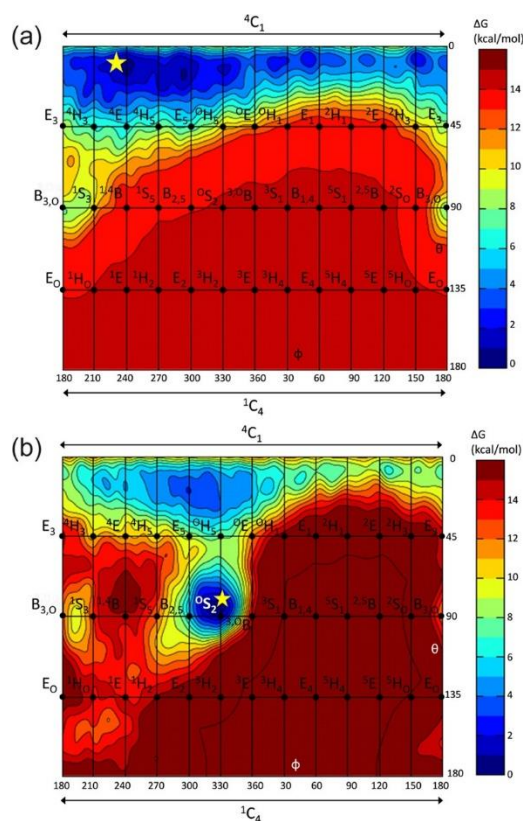


Figure 2. (a) Conformational FEL of the 1-thio- α -mannosyl residue at the -1 subsite of CpGH125 in complex with 1,6- α -thiomannobiose. The star symbol plots the conformation observed experimentally.²² (b) Conformational FEL of the α -mannosyl residue at the -1 subsite of CpGH125 in complex with 1,6- α -mannobiose. The star symbol plots the new conformation subsequently observed experimentally (see below). Contour lines are every 1 kcal mol⁻¹.

that calculated for 1-thio- α -mannopyranose. The use of an S-linked substrate analogue results in a strong bias to a 4C_1 conformation, matching that observed in the original report of Gregg et al.,²² with other more mechanistically relevant conformations not energetically accessible. In this case, while the thiomannoside substrate mimic is informative on the gross details of the catalytic apparatus and the ligand interactions, it is silent in terms of conformational insight.

We next sought to establish whether a solely computational approach could make testable predictions for the catalytic itinerary consistent with that previously observed for α - and β -

mannosidases. Starting with the experimentally determined *CpGH125* 1,6- α -thiomannobiose complex, the glycosidic sulfur was substituted for oxygen in silico to generate a catalytically viable Michaelis complex, which was subjected to minimization to generate a lower energy form, followed by MD equilibration. The full conformational landscape of the -1 sugar ring of this competent substrate containing an *O*-glycosidic linkage was then calculated using the same procedure as in the case of the 1,6- α -thiomannobiose complex. Figure 2b shows that within this Michaelis complex (on-enzyme), an *O*-glycoside strongly favors an 0S_2 conformation, consistent with the α -mannosidase performing catalysis through an $^0S_2 \rightarrow B_{2,5} \rightarrow ^1S_5$ conformational itinerary. Subsequent QM/MM simulations of the reaction mechanism (Figures 3 and S2) starting from the 0S_2

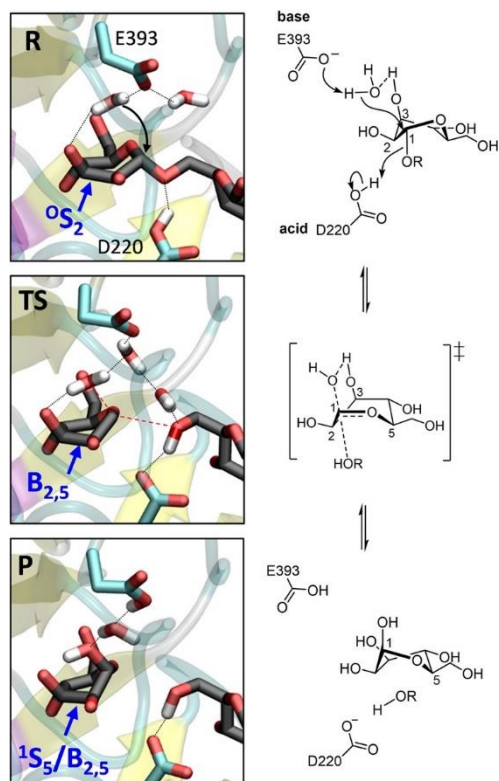


Figure 3. Reaction coordinates for *CpGH125* inverting 1,6- α -mannosidase obtained by QM/MM metadynamics with four collective variables. Most hydrogen atoms have been omitted for clarity.

conformation led to a $B_{2,5}$ TS, in a dissociative reaction pathway generating a β -mannose product bound to *CpGH125* with a $^1S_5/B_{2,5}$ conformation. This computational data, derived from the coordinate of the *CpGH125* 1,6- α -thiomannobiose complex, matches that proposed for GH families 2, 5, 26, 38, 76, 92, 113, and 130.^{1,4}

In order to validate, experimentally, the in silico prediction, an inactive variant in which the general acid (D220) of *CpGH125* was mutated to a nonacidic asparagine residue was engineered. This catalytically inactive variant was crystallized and soaked with the native *O*-glycosides and 1,6- α -mannobiose and -mannotriose to obtain pseudo-Michaelis complexes. Comparison of the structures of the ligand-free *CpGH125* wildtype and ligand-bound D220N enzymes revealed no changes in the position of the amino acid side-chain or other residues, providing confidence that the observed ligand conformation

was not a result of nonisomorphism. The *CpGH125* D220N complexes, solved at resolutions of 2.10 and 1.55 Å (Supporting Information Table 1), unambiguously reveal the -1 subsite mannoside distorted to a 0S_2 conformation (Figure 4a; for 1,6- α -mannotriose complex, see Figure S3), matching that predicted *a priori* by computation.

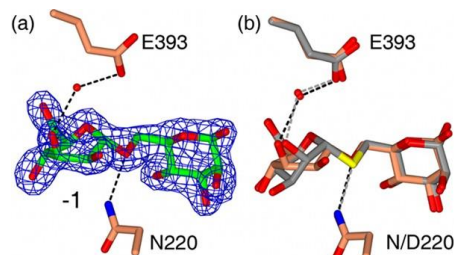


Figure 4. (a) Observed electron density ($2F_{\text{obs}} - F_{\text{calc}}$, σ_A , and maximum likelihood weighted) for D220N 1,6- α -mannobiose complex of *CpGH125*, contoured at 0.31 electrons/Å³. (b) Comparison of *CpGH125* complexes with 1,6- α -mannobiose (this work, brick red) with 1,6- α -thiomannobiose complex (gray, PDB 3QT9, ref 22). D220 is the general acid, E393 is the general base, and the proposed nucleophile water is shown.

The *CpGH125* complexes highlight the molecular basis for catalysis, with a water poised for in-line nucleophilic attack at the anomeric carbon and with E393 positioned to act as the Brønsted base in an inverting mechanism, essentially as proposed previously.²² The nucleophilic water molecule is engaged in a hydrogen-bonding interaction with O3, rather than with O2, as was instead observed in the 1,6- α -thiomannobiose complex. This interaction with O3 is reminiscent of that seen for the nucleophilic residue for a GH family 76 retaining 1,6- α -mannanase from *Bacillus circulans*¹⁵ and is thus a feature of the non-metal-dependent family 76 and 125 α -mannosidases. Overlay of the *CpGH125* D220N 1,6- α -mannobiose complex with the previously determined 1,6- α -thiomannobiose complex (Figure 4b) highlights the structural basis for the conformational differences: The +1 (leaving group) subsite mannoses are essentially identical in terms of conformation and interactions, but the -1 subsite mannoside moieties adopt different conformations and match those predicted by computation. One major contributor to these different conformations is the longer C-S bond (1.89 vs 1.48 Å for C-O); presumably as a result of this key structural difference, the -1 thiomannoside in a 4C_1 conformation with an axial O2 group makes similar interactions to the pseudoaxial O3 of the mannoside in an 0S_2 conformation (Figure S4). Both theory-based calculations and subsequent experimental observation support a conformational itinerary for the inverting GH125 α -mannosidases that proceeds through a (near) $B_{2,5}$ TS conformation. This TS is accessed following binding of the substrate in the ES complex in an 0S_2 conformation, Figure 3.

GH family 125 joins the growing list of mannose-active enzymes that follows a latitudinal pathway around a $B_{2,5}$ TS in which a key “feature” is the near-eclipsed 40° torsional angle between O3 and O2 that positions a *manno*-configured O2 pseudoequatorial and stabilized through H-bonding on-enzyme. There is a remarkable connection to Crich β -mannosylation methodology wherein judicious choice of a 4,6-*O*-benzylidene protecting group favors a similar pathway.⁶ Thiooligosaccharide substrate mimics have been widely used in X-ray crystallographic studies where they have provided mechanistically relevant insight into conformations possible on

enzyme, most notably in the case of distorted thiocellopentaoside bound to *Fusarium oxysporum* cellulase of GH family 7.²⁷ Could other thioglycoside complexes be misleading? In the case of another inverting α -mannosidase, from GH family 47, a 1,2- α -thiomannobioside was ring-flipped and distorted to the 3S_1 conformation suggesting that in this case it is mechanistically relevant.¹⁹ Other GH47 complexes, notably with mannoimidazole in a 3H_4 conformation, kifunensine in a 1C_4 conformation, as well as subsequent QM/MM analysis of FEL of α -mannose “on-enzyme”, collectively support the ${}^3S_1 \rightarrow {}^3H_4 \rightarrow {}^1C_4$ pathway for that enzyme.¹⁹ In contrast, as discussed earlier, the 1,2- α -thiomannobioside complex reported for family GH92, as described here for GH125, was also observed undistorted and again silent to conformational pathways; in that case distortion of enzyme-bound mannoimidazole to a boat conformation allowed assignment of a ${}^0S_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$ pathway for that enzyme.¹⁶

This work highlights the power of computational methods to use preliminary enzyme–ligand complexes to explore conformational space and generate testable predictions that can provide mechanistic insight using X-ray structural methods. Here these approaches predicted ES distortion for CpGH125 and informed an experimental approach that enabled direct observation of a distorted pseudo-Michaelis complex. This combined *in silico*-experimental approach could be applied to identify catalytic itineraries for other GH families that are presently unknown or for which unusual conformations have been proposed, guiding inhibitor design and leading to the development of mechanistic probes, cellular probes, and ultimately therapeutic agents. In this latter context, the ultimate goal is to obtain conformationally selective and thus specific inhibition of just one enzyme family, as has been achieved through inhibition of (${}^3S_1 \rightarrow {}^3H_4 \rightarrow {}^1C_4$ pathway) GH47 α -mannosidases by kifunensine, a 1C_4 chair mimic.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b11247. Coordinates have also been deposited in the RCSB Protein Databank under codes 5M7Y and 5M7I.

Additional methods and experimental data (PDF)

AUTHOR INFORMATION

Corresponding Authors

*gideon.davies@york.ac.uk

*c.rovira@ub.edu

ORCID

Spencer J. Williams: 0000-0001-6341-4364

Carme Rovira: 0000-0003-1477-5010

Author Contributions

C.R., S.J.W., and G.J.D. designed experiments. S.A.-G. performed computational work, A.M. carried out structural work, and P.F. conducted synthesis. G.J.D., S.J.W., and C.R. wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are supported by the Royal Society (Ken Murray Research professorship to G.J.D.), the Biotechnology and Biological

Sciences Research Council (A.M.), the Australian Research Council (FT130100103) (S.J.W.), the Spanish Ministry of Economy and Competitiveness (CTQ2014-55174-P) (C.R. and S.A.-G.) and GENCAT (2014SGR-987) (C.R.). We thank Diamond Light Source for access to beamline I02 and I04 (mx-13587) and BSC-CNS for computer resources and technical support at MareNostrum (RES-QCM-2016-3-0017).

REFERENCES

- (1) (a) Davies, G. J.; Planas, A.; Rovira, C. *Acc. Chem. Res.* **2012**, *45*, 308. (b) Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. *Curr. Opin. Struct. Biol.* **2014**, *28*, 1.
- (2) (a) Koshland, D. E. *Biol. Rev.* **1953**, *28*, 416. (b) Sinnott, M. L. Glycosyl group transfer. In *Enzyme Mechanisms*; Page, M. I., Williams, A., Eds.; Royal Society of Chemistry: London, **1987**; pp 259–297.
- (3) Sinnott, M. L. *Adv. Phys. Org. Chem.* **1988**, *24*, 113.
- (4) Davies, G. J.; Williams, S. J. *Biochem. Soc. Trans.* **2016**, *44*, 79.
- (5) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490.
- (6) Crich, D. *Acc. Chem. Res.* **2010**, *43*, 1144.
- (7) (a) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631. (b) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019.
- (8) Free, S. J. *Adv. Genet.* **2013**, *81*, 33.
- (9) Cuskin, F.; et al. *Nature* **2015**, *517*, 165.
- (10) Thompson, A. J.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 781.
- (11) Tailford, L. E.; Offen, W. A.; Smith, N. L.; Dumon, C.; Morland, C.; Gratien, J.; Heck, M. P.; Stick, R. V.; Bleriot, Y.; Vasella, A.; Gilbert, H. J.; Davies, G. J. *Nat. Chem. Biol.* **2008**, *4*, 306.
- (12) Vincent, F.; Gloster, T. M.; Macdonald, J.; Morland, C.; Stick, R. V.; Dias, F. M. V.; Prates, J. A. M.; Fontes, C. M. G. A.; Gilbert, H. J.; Davies, G. J. *ChemBioChem* **2004**, *5*, 1596.
- (13) Ducros, V.; Zechel, D. L.; Murshudov, G. N.; Gilbert, H. J.; Szabo, L.; Stoll, D.; Withers, S. G.; Davies, G. J. *Angew. Chem., Int. Ed.* **2002**, *41*, 2824.
- (14) Numao, S.; Kuntz, D. A.; Withers, S. G.; Rose, D. R. *J. Biol. Chem.* **2003**, *278*, 48074.
- (15) Thompson, A. J.; et al. *Angew. Chem., Int. Ed.* **2015**, *54*, 5378.
- (16) Zhu, Y.; et al. *Nat. Chem. Biol.* **2010**, *6*, 125.
- (17) Williams, R. J.; et al. *Angew. Chem., Int. Ed.* **2014**, *53*, 1087.
- (18) Cuskin, F.; Baslé, A.; Ladevež, S.; Day, A. M.; Gilbert, H. J.; Davies, G. J.; Potocki-Veroneš, G.; Lowe, E. C. *J. Biol. Chem.* **2015**, *290*, 25023.
- (19) (a) Thompson, A. J.; et al. *Angew. Chem., Int. Ed.* **2012**, *51*, 10997. (b) Karaveg, K.; Siriwardena, A.; Tempel, W.; Liu, Z. J.; Glushka, J.; Wang, B. C.; Moremen, K. W. *J. Biol. Chem.* **2005**, *280*, 16197.
- (20) Vallé, F.; Karaveg, K.; Herscovics, A.; Moremen, K. W.; Howell, P. L. *J. Biol. Chem.* **2000**, *275*, 41287.
- (21) Jin, Y.; Petricevic, M.; John, A.; Raich, L.; Jenkins, H.; Portela De Souza, L.; Cuskin, F.; Gilbert, H. J.; Rovira, C.; Goddard-Borger, E. D.; Williams, S. J.; Davies, G. J. *ACS Cent. Sci.* **2016**, *21*, 896.
- (22) Gregg, K. J.; Zandberg, W. F.; Hehemann, J. H.; Whitworth, G. E.; Deng, L.; Vocadlo, D. J.; Boraston, A. B. *J. Biol. Chem.* **2011**, *286*, 15586.
- (23) Heightman, T. D.; Vasella, A. T. *Angew. Chem., Int. Ed.* **1999**, *38*, 750.
- (24) (a) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826. (b) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
- (25) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.
- (26) Ardevol, A.; Rovira, C. *J. Am. Chem. Soc.* **2015**, *137*, 7528.
- (27) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280

**Distortion of Mannoimidazole Supports a $B_{2,5}$ Boat
Transition State for the Family GH125 α -1,6-
Mannosidase from *Clostridium perfringens***

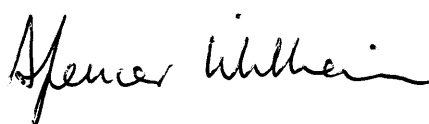
Declaration

Prof. Spencer J. Williams and Prof. Gideon J. Davies conceived the initial idea for the paper. All of the experimental work in this paper was conducted by me. I produced, purified and crystallised CpGH125 before soaking the crystals with mannoimidazole, generously synthesised by supplied by Gaetano Speciale and Prof. Spencer J. Williams. I subsequently solved the structure of CpGH125 in complex with mannoimidazole by X-ray crystallography. Additionally, I conducted the isothermal titration calorimetry experiment with input from Dr. John Darby and the inhibitory concentration analysis with input from Dr. Zach Armstrong. I wrote a complete first draft of the manuscript, which was revised and reworked thoroughly through multiple reiterations involving Prof. Spencer J. Williams, Prof. Gideon J. Davies and myself.



Signed

Prof. Gideon J. Davies



Signed

Prof. Spencer J. Williams

Reference

Males, A., Speciale, G., Williams, S. J. and Davies, G. J. (2019). Distortion of mannoimidazole supports a $B_{2.5}$ boat transition state for the family GH125 α -1,6-mannosidase from *Clostridium perfringens*. *Org. Biomol. Chem.* 17, 7863-7869.

Cite this: *Org. Biomol. Chem.*, 2019, 17, 7863Received 17th May 2019,
Accepted 2nd July 2019

DOI: 10.1039/c9ob01161g

rsc.li/obc

Distortion of mannoimidazole supports a $B_{2,5}$ boat transition state for the family GH125 α -1,6-mannosidase from *Clostridium perfringens*

Alexandra Males,^a Gaetano Speciale,^b Spencer J. Williams^b * and Gideon J. Davies^a

Enzyme transition-state mimics can act as powerful inhibitors and allow structural studies that report on the conformation of the transition-state. Here, mannoimidazole, a mimic of the transition state of mannosidase catalyzed hydrolysis of mannosides, is shown to bind in a $B_{2,5}$ conformation on the *Clostridium perfringens* GH125 α -1,6-mannosidase, providing additional evidence of a 0S_2 - $B_{2,5}$ - 1S_5 conformational itinerary for enzymes of this family.

A conformational itinerary describes how a molecule changes shape along a reaction coordinate. For reactions catalysed by enzymes, the three-dimensional fold of the protein and its electrostatic microenvironment provides constraints such that well-defined conformational itineraries are observed. These conformational itineraries may differ from those observed for chemically-catalyzed reactions in the gas or condensed phase, but nonetheless must still obey stereoelectronic dictates of chemical reactivity. The study of conformational itineraries of enzymes has been assisted by the full spectrum of kinetic, computational and structural methods.^{1,2} In the case of glycoside hydrolases, conformational itineraries are often assigned through determination of the three-dimensional structures of ligands bound to enzymes, most commonly substrate (Michaelis), intermediate, and product complexes, and by using inhibitors that mimic the transition state.³

GHs have been grouped into >150 families on the basis of primary sequence and ability to cleave specific glycoside substrates.^{4,5} As sequence defines the three-dimensional structure of proteins, family members typically adopt similar folds and possess highly conserved amino acids within their active site clefts. The chemical mechanisms through which

family members perform catalysis tend to be conserved across a family and lead to substrate hydrolysis with either retention or inversion of anomeric stereochemistry.⁶ Inversion of stereochemistry occurs in a single step reaction when water acts as a nucleophile. Retention of stereochemistry is typically achieved through a two-step substitution mechanism involving participation by an enzymic nucleophile or by a pendant neighboring group on the substrate.⁶⁻⁸ Both pathways benefit from enzymatic amino acids that provide general acid catalysis to promote the departure of the anomeric groups, and general base catalysis to promote nucleophilic attack by water. In almost all cases the transition states possess oxocarbenium-ion-like character and the stability of this TS is greatest when partial double bond character is maximized between C1 and O5, which occurs when the system C5-O5-C1-C2 is planar.⁹ Such transition-state conformations include two boats: $B_{2,5}$ and ${}^{2,5}B$; two half-chairs: 4H_3 and 3H_4 (Fig. 1a); and the related envelopes: 3E , E_3 , 4E and E_4 . As transition state mimicry is a powerful strategy for inhibitor design, efforts have been made to assign conformational itineraries, and in particular the TS conformation, for a large number of GH families.^{1,2}

Mannosidases and mannanases catalyze the hydrolysis of mannoside linkages and these enzymes are found within 11 GH families. Enzymes within families GH2, 5, 26, 38, 76, 92, 99, 113, 125 and 134 perform catalysis through itineraries along the 0S_2 - $B_{2,5}$ - 1S_5 region of conformational space whereas enzymes of families GH47 and 134 react through the 3S_1 - 3H_4 - 1C_4 conformational axis (Fig. 1b). The use of mannoimidazole and analogues thereof have been instrumental in the assignment of conformational itineraries and transition state conformation (Table 1).

In the case of family GH125, the conformational itinerary remained elusive for many years. Initial X-ray structures of the foundation member *Cp*GH125 *exo*- α -1,6-mannosidase (*Cp*GH125) with the nonhydrolyzable substrate mimic methyl 1,6- α -thiomannobioside (1) (Protein Databank (PDB) ID: 3QT9) and deoxymannojirimycin (2) (PDB ID: 3QRY) revealed undis-

^aYork Structural Biology Laboratory, Department of Chemistry,

The University of York, YO10 5DD York, UK. E-mail: gideon.davies@york.ac.uk ^bSchool of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria 3010, Australia.

E-mail: sjwill@unimelb.edu.au

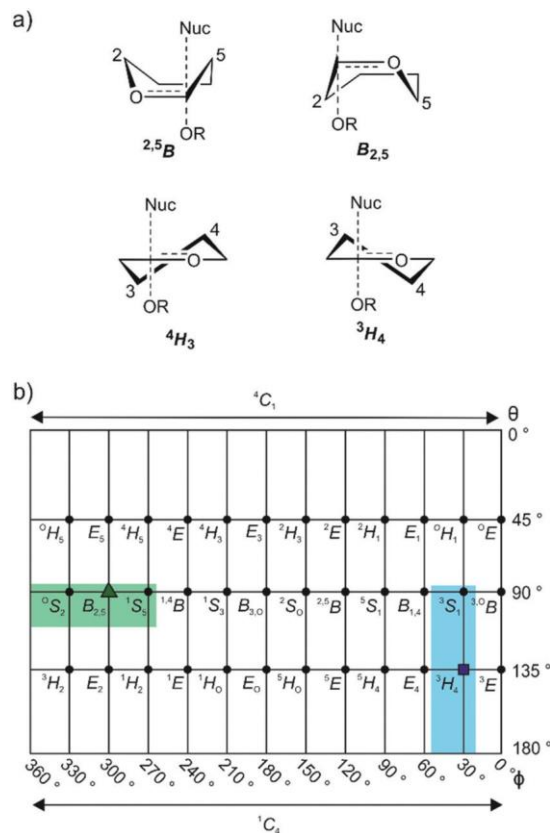


Fig. 1 (a) Common half-chair and boat transition state conformations, illustrated for α -glycosidases. (b) Common mannosidase conformational itineraries shown on a Mercator plot with that for GH125 highlighted in green. Shown in blue is the 3S_1 - 3H_4 - 1C_4 itinerary used by GH47 and 134 enzymes. The transition states are represented by a green triangle and a blue square.

Table 1 Proposed transition state conformation of mannosidases from various glycoside hydrolase families. Families for which the transition state conformation has been assigned on the basis of mannoimidazole complexes are indicated with an asterisk (*)

	Transition state conformation	
	$^3H_4^\ddagger$	$B_{2,5}^\ddagger$
α -Mannosidases	47* (ref. 10)	38*, ³ 76, ¹¹ 92*, ¹² 125* ^a
β -Mannosidases	134 (ref. 13)	2*, ¹⁴ 5*, ¹⁵ 26*, ^{3,16} 113*, ³ 130 ¹⁷

^aThis work.

torted, ground-state 4C_1 conformations (Fig. 2a).¹⁸ Instead, a computational approach involving *ab initio* QM/MM metadynamics was used to model the conformation of substrate, 1,6- α -mannobiose, in the enzyme active site, by *in silico* substitution of the sulfur in the experimentally determined X-ray structure.¹⁹ This calculation predicted that the *O*-glycoside favoured an 0S_2 conformation on-enzyme and hinted at an $^0S_2 \rightarrow B_{2,5}^\ddagger \rightarrow ^1S_2$ conformational itinerary, which was fully supported by QM/MM simulations of the reaction mechanism (Fig. 2b). Direct experimental support for this mechanism was obtained through determination of the Michaelis complex for

the catalytically-disabled acid mutant (D220N) of *CpGH125* in complex with 1,6- α -mannobiose, which revealed an 0S_2 conformation consistent with the computational predictions.¹⁹ As the ultimate determinant of the activation barrier for catalysis, the TS is the cardinal feature of the conformational itinerary. Mannoimidazole (3) (MIm) has emerged as a powerful chemical probe for TS conformation of mannosidases owing to its flattened nature, with an sp^2 -hybridized anomeric centre and energetically feasible access to all of the possible stereo-electronically-compliant conformations expected for an oxocarbenium ion-like transition state (Fig. 2b).³ Indeed, for a single step reaction, determination of the conformation of the transition state allows prediction of the conformational itinerary (from substrate to product *via* the TS) by invoking the principle of least nuclear motion.²¹ This principle states that elementary reactions that involve the least change in atomic position and electronic configuration will be favoured. Given the importance of the TS conformation in a reaction coordinate, and the utility of MIm in assigning conformational itineraries, we chose to revisit family GH125 to gather additional evidence in support of the proposed conformational itinerary. Here we report an X-ray structure of mannoimidazole (3) bound to wild-type *CpGH125* that provides compelling evidence in support of a 'latitudinal' $^0S_2 \rightarrow B_{2,5}^\ddagger \rightarrow ^1S_2$ conformational itinerary, complementing previous computational and structural data.

Results and discussion

Glycoimidazoles have been argued to act as glycosidase inhibitors through 'lateral' protonation of the imidazole nitrogen lone pair by the enzymatic general acid approaching in the plane of the imidazole ring.²² In fact, this interaction served as the basis for an early classification of glycosidases into *syn* and *anti* protonators, based on whether their general acid residues are situated *syn* or *anti* to the substrate C1-O5 bond (Fig. 2c).²³ However, analysis of published complexes with GH125 enzymes shows that the general acid residue is located below the mean plane of the ring,^{18,19} suggesting that there are poor prospects for binding of MIm to *CpGH125*. Yet, there are now several examples of glycoimidazoles binding to glycosidases with good affinity even without achieving a prototypical interaction between the enzymatic general acid and the imidazole nitrogen, such as β -glucosidases of family GH116,²⁴

α -mannanases of family GH99²⁵ and α -mannosidases of family GH47¹⁰ (Fig. 2d). The success of these cases inspired us to investigate binding of MIm to *CpGH125*.

We studied the thermodynamics for binding of MIm to *CpGH125* using isothermal titration calorimetry (Fig. 3a). Analysis of the calorimogram using a one site binding model allowed prediction of binding at 3.2 sites with an overall K_D value of 1.6 mM. Using a sequential binding site model, individual dissociation constants can be calculated at each site yielding values of 40 μ M, 105 μ M, and 3.1 mM. However, whilst this is indicative of multiple binding sites, it is not clear that the individual K_D values are accurately determined by the

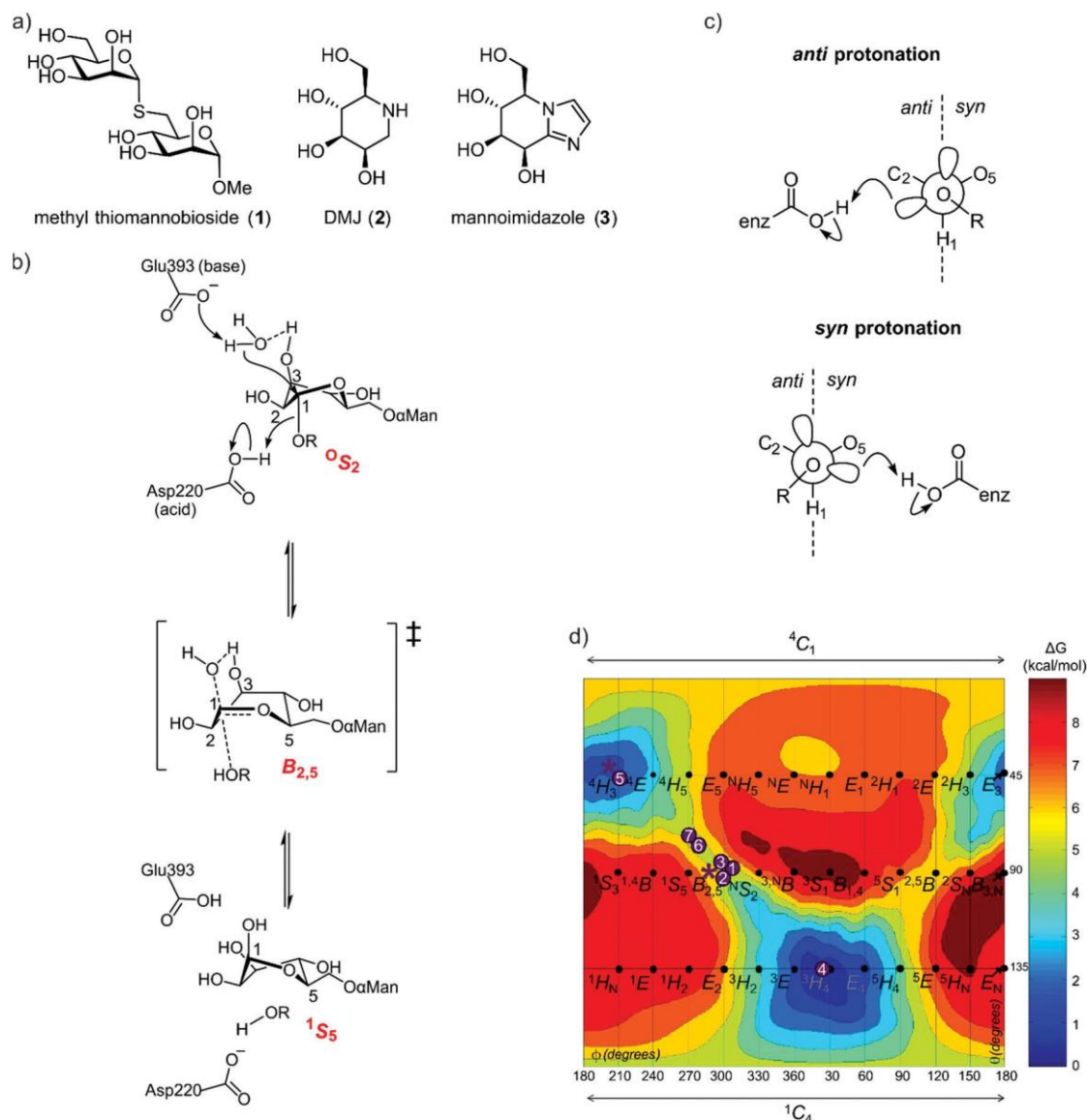


Fig. 2 (a) Structures of methyl α -thiomannobioside, deoxymannojirimycin and mannoimidazole. (b) Mechanism showing distortion of mannose from an 0S_2 conformation to 1S_5 , via a $B_{2,5}$ transition state, a credible conformational itinerary of GH125 α -mannosidases. (c) *syn* and *anti* trajectory of the protonator/general acid residue in respect to the carbohydrate ligand. (d) Conformational free energy landscape of mannoimidazole in complex with a variety of GH family mannosidases. **1 + 2**: β -Man-1,4-MIm bound to GH26 *CjMan26C* and GH113 *AaManA*; **3**: MIm bound to GH2 *BtMan2A*; **4**: MIm bound to GH47 *CkMan47*; **10** **5 + 6**: MIm bound to GH38 *DmGManII*; **3,20** **7**: MIm bound to GH92 *BtMan3990*; **12** * (asterisk) represent MIm bound to *CpGH125* (this work).

data. In order to gain evidence that binding by MIm, as shown by ITC, results in enzyme inhibition, we determined its IC_{50} value. At a concentration of 0.5 mM DNPM, we determined an IC_{50} of 1.7 mM (Fig. 3b), consistent with the overall ITC figure of 1.6 mM; this IC_{50} value cannot be converted to a K_i value since the substrate does not give saturation kinetics and so K_M cannot be determined.

Binding stoichiometry of >1 for basic sugar-shaped heterocycles is unusual and suggests that the ligand may not solely be binding at the -1 subsite. Previous work has identified three subsites, -1 , $+1$ and $+2$, in *CpGH125* through a complex with 1,6- α -mannotriose (PDB ID: 5M7Y),¹⁹ and studies of a di-

saccharide substrate that binds in the $-1/+1$ subsites revealed saturation kinetics, whereas a monosaccharide substrate, 2,4-dinitrophenyl α -mannoside (DNPM), that bound only in the -1 subsite did not.^{18,26} Previous poor electron density in the $+2$ subsite suggests that this may represent the low affinity binding site, a conclusion supported by subsequent X-ray crystallographic studies (*vide infra*). The two tighter affinities therefore should likely reflect binding at the -1 and $+1$ subsites; however, we cannot draw conclusions as to which value is assigned to each of these subsites. Nevertheless, the interesting ITC data encouraged efforts to determine the 3D structure of a complex of *CpGH125* with mannoimidazole.

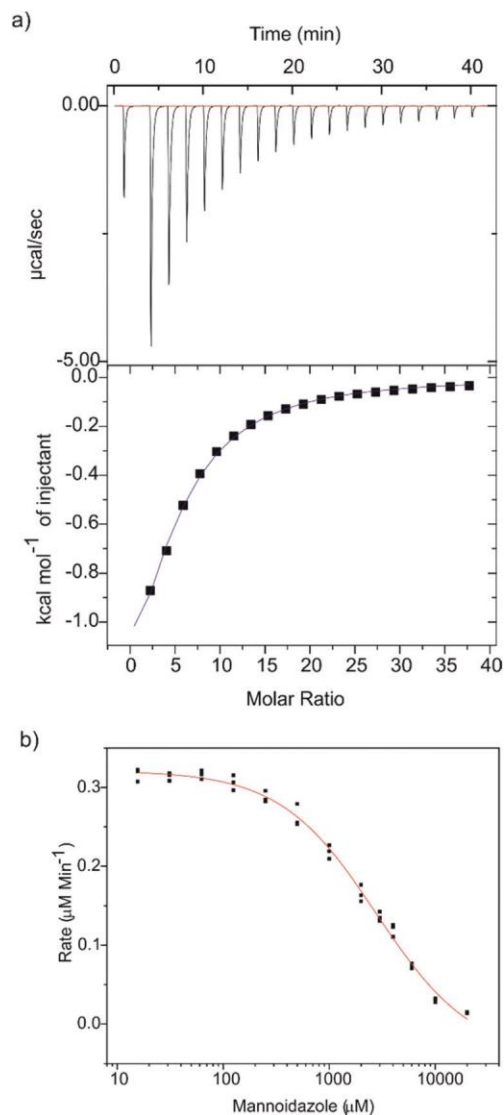


Fig. 3 (a) Raw thermodynamic data obtained by isothermal titration calorimetry. Top: injection profile with the baseline in red; bottom: titration curve with the fitting curve in blue. (b) IC₅₀ curve of MIm against GH125 using the monosaccharide substrate DNPMAN.

Preformed crystals of CpGH125 recombinantly expressed in *E. coli* were soaked with MIm and the structure of the complex was solved at 1.85 Å resolution (Table 2). The crystal form contains 2 monomers in the asymmetric unit. Density for MIm was observed both in the -1 and +1 subsites, but not in the +2 subsite (Fig. 4a). The MIm molecule in the -1 subsite bound in a ^B_{2,5} conformation and the MIm molecule in the +1 subsite bound in a ⁴H₃ conformation. The occupancy of MIm in the -1 subsite was estimated to be 80% and 70% for monomer A and B, respectively; the occupancy of MIm in the +1 subsite was 100% for both monomers. These observations are consistent with tight binding of MIm at the -1/+1 subsites, and weaker binding at the +2 subsite.

MIm molecules in each site engage with a range of active site residues (Fig. 4a). Key interactions with the -1 subsite

Table 2 Data collection and refinement statistics of CpGH125 and MIm

Resolution [Å]	84.8–1.85 (1.89–1.85) ^a
<i>R</i> _{merge}	0.08 (0.51)
CC(1/2)	0.99 (0.54)
Completeness [%]	99.7 (99.9)
<i>R</i> _{work} / <i>R</i> _{free}	0.20/0.24
RMSD _{bonds} [Å]	0.01
PDB code	6RQK

^a Values in parentheses are for highest-resolution shell.

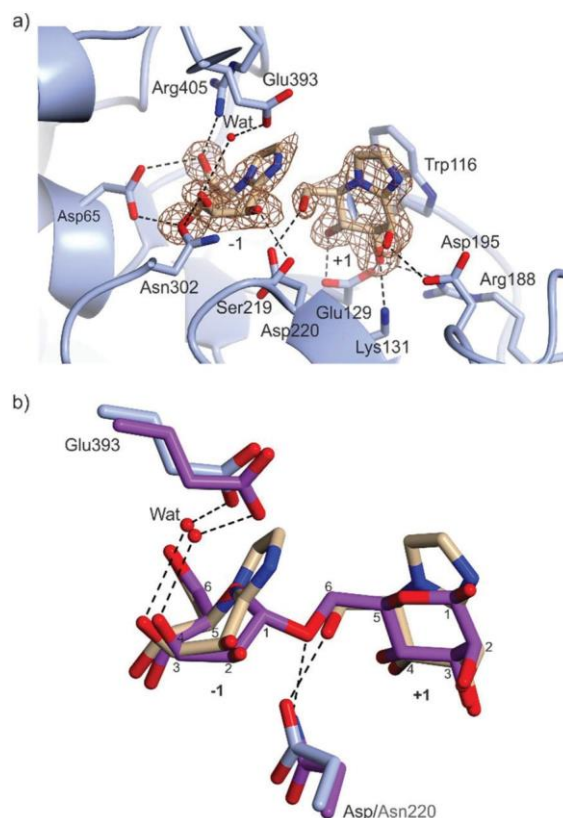


Fig. 4 (a) Complex of CpGH125 with mannoimidazole in the -1 and +1 subsites. The residues of the enzyme that form hydrogen bonding interactions and close contacts with mannoimidazole (<3 Å distance) are shown in blue and the nucleophilic water molecule is depicted by a red sphere. The electron density map is a REFMAC maximum-likelihood/ σ A-weighted $2F_o - F_c$ synthesis contoured at 0.75 electrons Å⁻³ and 1.20 electrons Å⁻³ clipped to MIm in the -1 and +1 subsite, respectively. (b) Superposition of CpGH125 in complex with MIm (monomer A) and 1,6- α -mannobiose (PDB ID: 5M71), shown in purple. Hydrogen bonds shown in black between the catalytic residues of CpGH125 and nucleophilic water

MIm include the carbonyl oxygen of Ser219 with O2, O δ 1 of Asn302 with O3, O δ 1 and O δ 2 of Asp65 with O4 and O6, respectively, and NH₂ of Arg405 with O6. The general base Glu393 is 2.3 Å away from a water poised for inline attack at the 'anomeric carbon' of MIm. Major interactions with MIm in subsite +1 include O δ 1 of Asp195 with O2, Glu129 O ϵ 1 with O3 and O ϵ 2 with O4, and the terminal amine nitrogen, NZ of Lys131 and NH₂ of Arg188 with O3. The catalytic amino acid,

Asp220, interacts with O6 of MIm in the +1 subsite. Trp116 is stacked parallel to the imidazole of MIm, with CZ2 of Trp119 and N2 of MIm spaced 3.79 Å apart.

Overlay of the complex of MIm with wildtype *CpGH125* with that of the D220N mutant of the same enzyme with 1,6- α -mannobiose (PDB ID: 5M7I) is shown in Fig. 4b. This shows that the two MIm molecules bind such that their carbohydrate rings are well-matched with the equivalent mannose rings of 1,6- α -mannobiose. Moreover, the general acid (Asp220; or variant Asn220), and general base (Glu393) residues occupy similar positions in space. This overlay highlights that the interaction of Asp220 with the +1 subsite MIm molecule, in the complex determined here, can be considered equivalent to the interaction of Asn220 with the inter-glycosidic oxygen of 1,6- α -mannobiose seen in the *CpGH125*-D220N complex. Both complexes include a well-defined water molecule hydrogen-bonded to Glu393 and O3 of the -1 subsite MIm that is poised to undertake a nucleophilic attack at the anomeric centre. The interaction of this water molecule with O3 assists in orientating the water molecule to perform an in-line attack on C1 (Fig. 2b).

Nerinx generalized the concept of *syn/anti* protonation to consider divergence from canonical orientation in the mean plane of the ring to include orientations in a *syn* or *anti* hemisphere (Fig. 2c).²⁷ According to this extended concept, *CpGH125* can be considered a *syn* protonator and with the present work *CpGH125* joins a growing list examples of glycosidases that can bind glycoimidazoles even without achieving lateral protonation. While binding is seen in this and related cases, presumably the potency of binding is less than what might be expected for an ideally located general acid that can achieve such an interaction, an expectation that is borne-out by much tighter binding to canonical anti-protonators such as *Bt3130* and *Bt3965* from *Bacteroides thetaiotaomicron* GH92 ($K_i = 1.0$ and $0.4 \mu\text{M}$, respectively) whereas other α -mannosidases from *B. thetaiotaomicron* have a K_i values ranging from 40–6000 μM .¹² Also, imidazole-substituted mannoimidazoles can exhibit even tighter binding, for example, the β -mannosidase *BtMan2A* (from family GH2) binds phenethyl-mannoimidazole with $K_i = 57$ nM compared to binding to mannoimidazole with $K_i = 1400$ nM.¹⁴ In this case since the inhibitor lies across both the -1 and +1 subsite, the number of interactions the inhibitor can make with the protein increases. We highlight one important anomalous case, in which a complex of a disaccharide MIm analogue with *BxGH99* 1,2- α -mannanase revealed an unusual ${}^2H_3/E_3$ conformation possibly due to steric clashes with an active site residue, Tyr252.²⁵ In this case no K_D value could be measured by ITC indicating poor affinity for the enzyme.

Conclusions

The present work builds upon prior computational and structural studies of *CpGH125* focussed on the Michaelis complex that provided evidence for an ${}^0S_2 \rightarrow B_{2,5} \ddagger \rightarrow {}^1S_5$ conformational

itinerary (Fig. 4). The present TS mimicking complex with MIm in a $B_{2,5}$ conformation provides compelling evidence for this conformation in the transition state of this enzyme. Since the free energy landscape³ for MIm reveals that the most stable conformations for this inhibitor in the gas phase are the 4H_3 and 3H_4 conformations, with the $B_{2,5}$ conformation some 6 kcal mol⁻¹ less stable, the observation of a $B_{2,5}$ conformation provides persuasive evidence that this conformation is induced upon binding because of the enzyme's strong preference for stabilizing a transition state with this conformation. Combined with the previously published Michaelis complex in an 0S_2 conformation, we now have crystallographic 'snapshots' of states that together define two of the three critical points along the reaction coordinate of family GH125 enzymes.

Experimental

Expression, purification and crystallisation of *CpGH125*_{WT}

Wild-type *CpGH125* was expressed and purified under the same conditions as the mutant, *CpGH125*-D220N.¹⁹ *CpGH125*, at 12 mg ml⁻¹, was crystallised under 0.2 M MgCl₂, 0.1 M HEPES pH 7.0 and 24% PEG 3350, protein : reservoir ratio 0.3 : 0.5 μL conditions and subsequently soaked with mannoimidazole. Crystals were soaked in 5 mM inhibitor for 2 h at 20 °C and then cryoprotected stepwise in 25% ethylene glycol before snap freezing in liquid nitrogen.

Data were collected on Diamond MX beamline I24 and processed using xia2 3dii.²⁸ Integrated data were processed using the CCP4 software suite;²⁹ the programs used were aimless,^{30,31} MOLREP³² with 5M7I as the reference model and cycles of REFMAC5^{33–38} and COOT³⁹ for refinement of the structure. *CpGH125* structures figures were produced using CCP4 mg.⁴⁰ Coordinate and observed structure-factor amplitudes have been deposited on the PDB with accession code 6RQK.

Measurement of dissociation constant (K_D) by isothermal titration calorimetry

CpGH125 and MIm were buffer matched by dialysis into 20 mM HEPES pH 7.0 and 200 mM NaCl. MIm was added by syringe to a solution of *CpGH125* in the calorimeter cell using a MicroCal ITC₂₀₀ calorimeter at 25 °C with 20 injections. The concentration of *CpGH125* was 200 μM and the ligand concentration was varied between 4 mM and 36 mM to obtain a consistent K_D value. The binding affinity was obtained using one site fitting and a sequential binding fit (3 sites) within the Origin7 software. The association constant (K_A) was converted to a dissociation constant using the equation $K_D = 1/K_A$. An average over 5 experiments was taken to obtain accurate K_D values for the three subsites.

Measurement of inhibition constant (IC_{50}) of MIm

Enzyme assays were performed in 40 μL quartz cuvettes with path lengths of 10 mm. Change in absorbance values at 405 nm were monitored using a Cary UV-Vis spectropho-

meter. Reactions were initiated by addition of CpGH125 at 9.05 μM to a solution of 0.5 mM DNPM in PBS buffer (pH 7.4) at 37 °C. Rates were measured by observing change in absorption for 3 min. Each reaction was performed in triplicate and averaged, and the data fit to a Hill plot.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by Biotechnology and Biological Sciences Research Council (grant BB/M011151/1) for support AM. SJW is funded by the Australian Research Council (DP180101957, DP160100597). GJD is the Royal Society Ken Murray Research Professor. We thank Diamond Light Source for access to beamline I24 (proposal number mx-9948-71) that contributed to the results presented here and Johan Turkenburg and Sam Hart for coordinating data collection. PDB accession code: 6RQK. We would also like to thank Zach Armstrong for support with the IC₅₀ enzyme assay.

Notes and references

- 1 G. J. Davies, A. Planas and C. Rovira, *Acc. Chem. Res.*, 2012, **45**, 308–316.
- 2 G. Speciale, A. J. Thompson, G. J. Davies and S. J. Williams, *Curr. Opin. Struct. Biol.*, 2014, **28**, 1–13.
- 3 R. J. Williams, J. Iglesias-Fernández, J. Stepper, A. Jackson, A. J. Thompson, E. C. Lowe, J. M. White, H. J. Gilbert, C. Rovira, G. J. Davies and S. J. Williams, *Angew. Chem., Int. Ed.*, 2014, **53**, 1087–1091.
- 4 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490–D495.
- 5 B. Henrissat and A. Bairoch, *Biochem. J.*, 1996, **316**(Pt 2), 695–696.
- 6 D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
- 7 M. L. Sinnott, *Chem. Rev.*, 1990, **90**, 1171–1202.
- 8 J. D. McCarter and G. Stephen Withers, *Curr. Opin. Struct. Biol.*, 1994, **4**, 885–892.
- 9 R. J. Woods, C. W. Andrews and J. P. Bowen, *J. Am. Chem. Soc.*, 1992, **114**, 859–864.
- 10 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, *Angew. Chem., Int. Ed.*, 2012, **51**, 10997–11001.
- 11 A. J. Thompson, G. Speciale, J. Iglesias-Fernández, Z. Hakki, T. Belz, A. Cartmell, R. J. Spears, E. Chandler, M. J. Temple, J. Stepper, H. J. Gilbert, C. Rovira, S. J. Williams and G. J. Davies, *Angew. Chem., Int. Ed.*, 2015, **54**, 5378–5382.
- 12 Y. Zhu, M. D. L. Suits, A. J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K. W. Moremen, Y. Xiang, A. Siriwardena, S. J. Williams, H. J. Gilbert and G. J. Davies, *Nat. Chem. Biol.*, 2010, **6**, 125–132.
- 13 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, *ACS Cent. Sci.*, 2016, **2**, 896–903.
- 14 L. E. Tailford, W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M.-P. Heck, R. V. Stick, Y. Blériot, A. Vasella, H. J. Gilbert and G. J. Davies, *Nat. Chem. Biol.*, 2008, **4**, 306–312.
- 15 F. Vincent, T. M. Gloster, J. Macdonald, C. Morland, R. V. Stick, F. M. V. Dias, J. A. M. Prates, C. M. G. A. Fontes, H. J. Gilbert and G. J. Davies, *ChemBioChem*, 2004, **5**, 1596–1599.
- 16 A. Tankrathok, J. Iglesias-Fernandez, R. J. Williams, S. Pengthaisong, S. Baiya, Z. Hakki, R. C. Robinson, M. Hrmova, C. Rovira, S. J. Williams and J. R. Ketudat Cairns, *ACS Catal.*, 2015, **5**, 6041–6051.
- 17 F. Cuskin, A. Baslé, S. Ladevèze, A. M. Day, H. J. Gilbert, G. J. Davies, G. Potocki-Véronèse and E. C. Lowe, *J. Biol. Chem.*, 2015, **290**, 25023–25033.
- 18 K. J. Gregg, W. F. Zandberg, J.-H. Hehemann, G. E. Whitworth, L. Deng, D. J. Voadlo and A. B. Boraston, *J. Biol. Chem.*, 2011, **286**, 15586–15596.
- 19 S. Alonso-Gil, A. Males, P. Z. Fernandes, S. J. Williams, G. J. Davies and C. Rovira, *J. Am. Chem. Soc.*, 2017, **139**, 1085–1088.
- 20 D. A. Kuntz, C. A. Tarling, S. G. Withers and D. R. Rose, *Biochemistry*, 2008, **47**, 10058–10068.
- 21 M. L. Sinnott, *Adv. Phys. Org. Chem.*, 1988, **24**, 113–204.
- 22 A. Varrot, M. Schülein, M. Pipelier, A. Vasella and G. J. Davies, *J. Am. Chem. Soc.*, 1999, **121**, 2621–2622.
- 23 T. D. Heightman and A. T. Vasella, *Angew. Chem., Int. Ed.*, 1999, **38**, 750–770.
- 24 R. Charoenwattanasatien, S. Pengthaisong, I. Breen, R. Mutoh, S. Sansenya, Y. Hua, A. Tankrathok, L. Wu, C. Songsiriritthigul, H. Tanaka, S. J. Williams, G. J. Davies, G. Kurisu and J. R. K. Cairns, *ACS Chem. Biol.*, 2016, **11**, 1891–1900.
- 25 P. Z. Fernandes, M. Petricevic, L. Sobala, G. J. Davies and S. J. Williams, *Chem. – Eur. J.*, 2018, **24**, 7464–7473.
- 26 L. Deng, P. Tsybina, K. J. Gregg, R. Mosi, W. F. Zandberg, A. B. Boraston and D. J. Voadlo, *Bioorg. Med. Chem.*, 2013, **21**, 4839–4845.
- 27 W. Nerinckx, T. Desmet, K. Piens and M. Claeysens, *FEBS Lett.*, 2005, **579**, 302–312.
- 28 G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186–190.
- 29 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 235–242.
- 30 P. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 72–82.

- R. Evans and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **69**, 1204–1214.
- A. A. Vagin and A. Teplyakov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 22–25.
- N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. Vagin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 355–367.
- N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1997, **53**, 240–255.
- S. Pannu, G. N. Murshudov, E. J. Dodson and R. J. Read, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1998, **54**, 1285–1294.
- 36 M. D. Winn, G. N. Murshudov and M. Z. Papiz, *Methods Enzymol.*, 2003, **374**, 300–321.
- 37 A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 2184–2195.
- 38 R. A. Nicholls, F. Long and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 404–417.
- 39 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.
- 40 S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 386–394.

CHAPTER 3: Family GH47 *Exo- α -1,2-Mannosidases*

3.1 Abstract

α -1,2-mannosidase enzymes, CAZy family GH47, cleave the α -1,2-mannosidic bond of Man₉GlcNAc₂ glycans. Kifunensine is the only naturally occurring inhibitor identified (to date) that is truly specific (K_D of 39 nM) to GH47 enzymes. It is commonly utilised in cell biology to prevent heterologous glycosylation. Further investigation into its structure, interactions with residues of the active site of GH47s and free energy “on-enzyme” of the conformationally-locked inhibitor was undertaken to inform development of specific GH47 inhibitors and modify currently available inhibitors. Kifunensine was revealed to be biased towards a ¹C₄ conformation and required 8 kcal mol⁻¹ to bind to GH38 and GH92 mannosidases in the necessary equatorial conformations. The general mannosidase inhibitor, 1-deoxymannojirimycin (DMJ), was shown to bind to GH47 in a dual ³S₁ and ¹C₄ conformation. Mannuronic acid-based derivatives were shown to prefer a ring flipped ¹C₄ conformation at a low pH in solution, however, the enzyme was not able to tolerate the size of the R group substituent.

3.2 Family GH47 *exo*- α -1,2-mannosidases (Golgi ManIA) introduction

3.2.1 Discovery of α -1,2-mannosidases

In 1979, Golgi α -mannosidase (CAZy family GH47) was isolated from rat liver and characterised (Tabas and Kornfeld, 1979). One protein was shown to be responsible for the cleavage of Man₉GlcNAc₂ to Man₅GlcNAc₂ and a variety of intermediate products. An equivalent protein in *Saccharomyces cerevisiae* could be separated to produce a second fraction that was specific to α -1,2-linkages but only cleaves Man₉GlcNAc₂ to Man₈GlcNAc₂ (Jelinek-Kelly et al., 1985).

3.2.2 Inhibition of α -1,2-mannosidases

The subclass of α -mannosidases (either I or II) and the location of class I mannosidases can be distinguished by their inhibition. Class I enzymes are inhibited by DMJ and kifunensine but not by swainsonine and the opposite is true for class II enzymes (Soussillane et al., 2009). Swainsonine is an indolizidine that inhibits class II Golgi mannosidases, Figure 3.1 (reviewed in Wrodnigg et al., 2008).

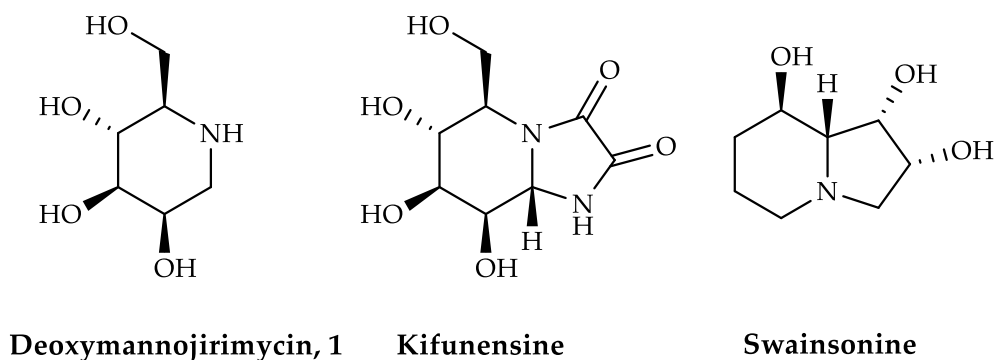


Figure 3.1. Inhibitors of α -1,2-mannosidases.

The dissociation constant, K_D , for DMJ and kifunensine against human ER mannosidase (*HsERManI*) was 1600 nM and 30 nM, respectively (Karaveg and Moremen, 2005) and the K_D against a bacterial GH47 ManIA from *Caulobacter* was 481 nM and 39 nM, respectively (Males et al., 2017; van Rijssel et al., 2017). DMJ is specific to Golgi ManI and has an IC_{50} of 10 μ M. Kifunensine is 50-100 times more effective than DMJ for ERManI (Elbein et al., 1990).

3.2.2.1 Inhibition by DMJ

DMJ is a known specific inhibitor of class I α -mannosidases (Bischoff et al., 1986). It is an analogue of mannose; it is an azasugar meaning it has a ring nitrogen (piperidine) instead of an oxygen (tetrahydropyran) and it lacks a catalytically essential hydroxyl group from C1, Figure 3.1.

DMJ has been shown to possess antiviral activity, including against human immunodeficiency virus (HIV). The HIV virus uses host cell glycan machinery to produce the gp120 envelope, which is heavily glycosylated with 24-28 potential N-glycosylation sites. Carbohydrate-binding agents (CBAs) have been used to inhibit HIV-1 by targeting the gp120 envelope (Balzarini, 2007a). However, they do not target mutant N-glycans that have deletions. This leads to resistance and selection of mutant HIV forms. Inhibition of α -1,2-mannosidases, using low levels of DMJ, does not affect host cell viability but instead suppresses the mutant virus selection. A combination of DMJ and CBA have been shown to both increase viral sensitivity to DMJ and the antiviral efficacy of CBA (reviewed in Balzarini, 2007b).

3.2.2.2 Inhibition by kifunensine

A plethora of antibiotics have been discovered in *Actinobacteria*, in particular the soil-dwelling *Actinomycete* order (reviewed in Barka et al., 2016). Kifunensine was discovered in *Kitasatosporia kifunense* no 9482, and identified as an immunomodulatory compound (Iwami et al., 1987). In the 1990s, the structure of the 1,5-iminopyranose was determined through various spectroscopic techniques (UV, ¹H NMR, ¹³C NMR) and confirmed using X-ray crystallography (Kayakiri et al., 1989). Kifunensine is a cyclic oxamide derivative of 1-amino-substituted mannojirimycin. A bridgehead nitrogen connects a six membered ring with four asymmetric hydroxyl groups and a five membered ring with two carbonyl groups, Figure 3.1 (Elbein et al., 1990).

The glycosylation pattern of a given residue on a glycoprotein shows heterogeneity posing challenges throughout the crystallisation and data interpretation processes (reviewed in Chang et al., 2007). Previously, Chinese hamster ovary cell-based expression systems were utilised with restricted processing of the N-glycan. After purification, the protein of interest was incubated with EndoH to cleave the glycan from the oligosaccharide. This approach was not suitable for high-throughput large scale expression of glycoproteins. By adding kifunensine to the expression cell suspension, the protein can fold correctly but the glycans are prevented from further assembly into complex, highly branched forms (Yu et al., 2011); the EndoH that is produced in the cell is then utilised. Partial glycosylation is important to prevent protein aggregation.

Treatment of human sarcoglycanopathies lysosomal storage disorders Gaucher's and Tay-Sachs disease using kifunensine has been unsuccessful due to unanalysed side effects (Bartoli et al., 2008). A low concentration, 50 nM, was administered to inhibit ERManI and either lead the misfolded sarcoglycans down the ERAD pathway (reviewed in Sandonà and Betto, 2009) or prolong the retention of mutated lysosomal enzymes in the ER, providing them time to be repaired (Wang et al., 2011).

Analogues of kifunensine have been developed to improve the specificity of the inhibitor. The analogues stem from *N*-1 substituted 2-desoxakifunensine, which lack the C2 carbonyl on the furanose ring to make it flexible, and *N*-1 substituted kifunensine, which features different functional groups attached to the nitrogen of the furanose ring. Tethered pseudodisaccharide mimetics (lie across -1 and +1 subsites) were synthesised but they did

not bind favourably. An *N*-bis(hydroxymethyl)methyl kifunensine analogue had a 10-fold greater selectivity for the inhibition of ERManI over Golgi ManI. Thiokifunensine analogues, replacement of nitrogen with sulfur in the five membered ring, were also developed and tested against jack bean α -mannosidase and HIV reverse transcriptase. They had improved stability and inhibitory activity against HIV reverse transcriptase and reduced hydrolysis (Chen et al., 2013). However, none of the analogues developed to date are as potent as kifunensine (Hering et al., 2005).

3.2.3 Diseases associated with GH47 α -1,2-mannosidases

α -1,2-mannosidases are associated with many diseases, either directly or indirectly, due to their involvement in many processes within the ER and Golgi. Many cancers have been associated with defects in the genes encoding α -1,2-mannosidases, for example, elevated levels of mannosidases in larynx cancer (Olszewska et al., 2012). *MAN1A1* (ManIA) is an oncogene associated with breast cancer that, when upregulated, activates the unfolded protein response and initiates metastasis (Milde-Langosch et al., 2014; Tu et al., 2017). Upregulation of *MAN1B1* (ERManI) promotes hepatocellular carcinoma (HCC) formation, since it is a target of miR-125b (a microRNA that is correlatively downregulated in HCC) (Pan et al., 2013). Clear cell renal cell carcinoma (ccRCC) and Wilm's tumour are urologic malignancies that are suppressed by *MAN1C1* (ManIC) (Krepischi et al., 2016; Li et al., 2018a). *MAN1C1* is downregulated (by tumour necrosis factor- α) during stage I of tumour progression, whilst the expression levels of *MAN1A1*, *MAN1A2* and *MAN1B1* were increased (Tu et al., 2017). However, inducing overexpression of *MAN1C1* caused apoptosis, upregulation of *Bax/Bcl-2* and *E-CADHERIN* and downregulation of mesenchymal genes, which inhibited the transition from epithelial to mesenchymal cells (Li et al., 2018a), reducing the migration of cancerous cells and their ability to form colonies (Tu et al., 2017). Lastly, *MAN1A2* (ManIB), along with other genes, is a prognostic indicator in diffuse large B cell lymphoma (Kim et al., 2014). Understanding how inhibition affects the structure and regulation of these enzymes is crucial for combating these diseases.

3.2.4 The structure of GH47 enzymes

A crystallisable construct of the catalytic domain of the *S. cerevisiae* α -1,2-mannosidase, which cleaves the first single mannose from branch B (M10) (Byrd et al., 1982), was the first structure of family GH47 enzymes to be obtained (Dole et al., 1997). However, the 3.5 Å resolution construct was heavily cleaved by proteases at the N-terminus and between residues 368 and 369 situated on a non-conserved loop. A new construct was designed starting at residue 34 and since has missing residues 367-371 (Lipari and Herscovics, 1994) that could be reproducibly crystallised and produced structural data with an improved resolution of 1.54 Å (Vallée et al., 2000a). Within the same year, a structure was published of the *HsERManI* (Vallée et al., 2000b). It was the first structure to reveal the role of calcium in mannosidase catalysis (see section 3.2.5.2) and propose the catalytic amino acids through structures of ERManI in complex with DMJ and kifunensine (two established mannosidase inhibitors - see section 3.2.2) (Vallée et al., 2000b).

GH47 enzymes were the first GH family observed to have an $(\alpha/\alpha)_7$ barrel fold, Figure 3.2 (Vallée et al., 2000a). The 14 α -helices consecutively alternate between the outer (α 1, α 3, α 5, α 7, α 9, α 11 and α 13) and inner barrel (α 2, α 4, α 6, α 8, α 10, α 12 and α 14) with dimensions of 50 x 50 x 50 Å and an internal seven-fold symmetry.

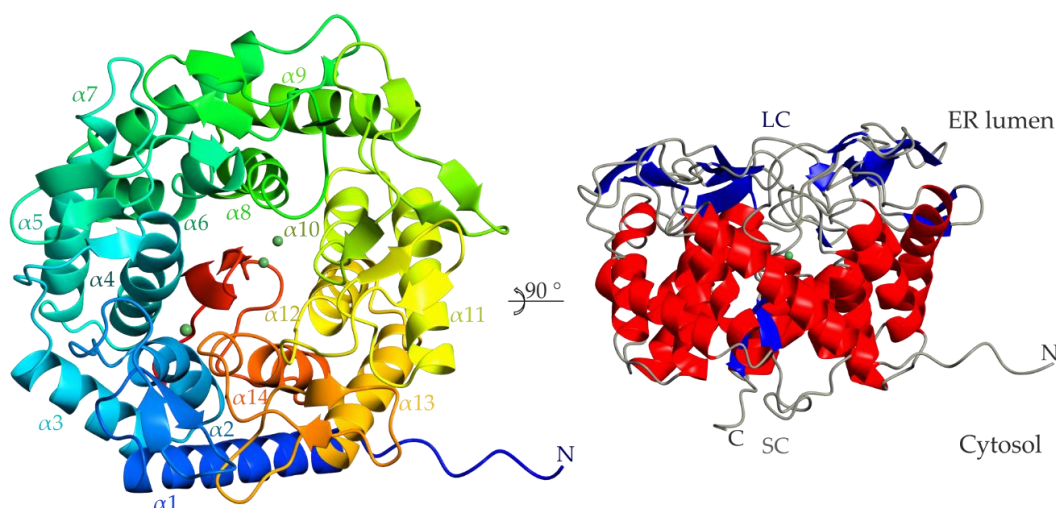


Figure 3.2. Secondary structure of the $(\alpha/\alpha)_7$ barrel of GH47 mannosidases. Left: α -helices are labelled in ascending order from the N-terminus (blue) to the C-terminus (red) according to Vallée et al., 2000b. The divalent metal ions are shown in green. Right: The secondary structure is highlighted, β -sheets (blue) line the long connection face of the α -helix (red) barrel. The β -hairpin plugs the active site from the short connection face. Protein from *CkGH47* PDB ID: 5MEH (van Rijssel et al., 2017).

GH47s are type II membrane proteins where the N-terminus, SC side, of ERManI is located in the cytoplasm linked via a single transmembrane helix to the catalytic LC side that protrudes into the ER lumen. This is a favourable orientation for generating interactions with newly synthesised glycoproteins (Grondin and Herscovics, 1992; Massaad et al., 1999). The two faces, short connection (SC) and long connection (LC), of the barrel are structurally different, Figure 3.2. On the SC side, the loops connecting the helices have no secondary structure and are up to four residues in length. In comparison, the loops on the LC side are structured into antiparallel β -sheets that surround the barrel. There are three sites of N-glycosylation on the SC side, with one site (on *S. cerevisiae*) clearly showing electron density for Man₅GlcNAc₂. The glycan chains point away from the protein and assist in crystallisation by making protein-oligosaccharide contacts.

A conserved disulfide bond (present in eukaryotic species) was observed between Cys340 on the α_{10c} helix at the start of helix 10 and Cys385 on helix 11 (numbered according to *S. cerevisiae* α -1,2-mannosidase and is equivalent to Cys527 and Cys556 of *HsERManI* (Lipari and Herscovics, 1996). It is proposed to be important for enzyme activity through stabilising the secondary structure as the catalytic base Glu435 on helix 12 lies next to this bond. Interestingly, the disulfide bond is not present in the *Caulobacter* K31 strain (this work), instead the equivalent residues are Ala312/Phe334, and the enzyme is active. Instead, a hydrogen bonding network is present between helix 10 and 11 and Glu309 interacts with the catalytic base, Figure 3.3.

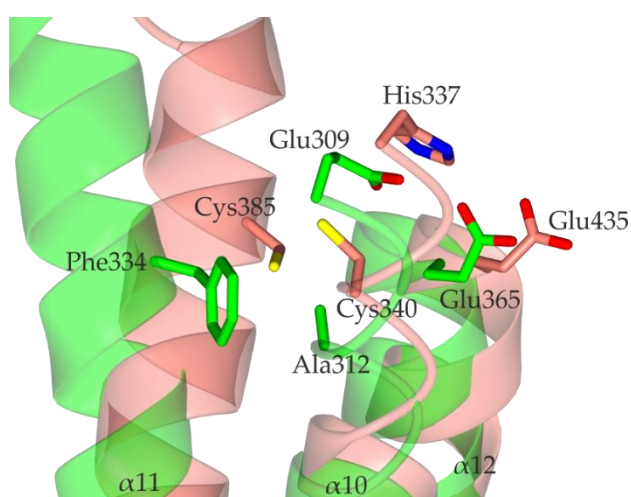


Figure 3.3. Disulfide bond only stabilises the secondary structure in eukaryotes. Superposition of *Caulobacter* GH47 is shown in green (PDB ID:5MEH) and *S. cerevisiae* GH47 is shown in coral (PDB ID:1DL2). The loop between helix 11 and 12 is not shown for clarity.

The active site is located within a tapered tunnel in the middle of the barrel whereby the widest point is at the opening (25 Å). At the C-terminus, residues 512-549, a β -hairpin plugs the barrel on the SC side forming a pocket active site with the dimensions $\sim 15 \times 10$ Å (the narrowest part is at the top of the β -hairpin and base of the active site), Figure 3.2.

3.2.5 Mechanism of catalysis

Catalysis occurs by an *anti*-protonating inverting mechanism whereby the *anti* proton donor is on the opposite side of the anomeric carbon (Lipari et al., 1995). This was first discovered by Lipari et al., in 1995 using $^1\text{H-NMR}$ spectroscopy to analyse the stereochemistry of the reaction and confirmed through crystal structures. Hydrolysis of $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_8\text{GlcNAc}_2$ revealed net inversion of the α -anomeric conformation (Lipari et al., 1995).

3.2.5.1 Proposition of catalytic residues

The residues that are involved in the catalysis of $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_8\text{GlcNAc}_2$ are still a matter of debate. Suggested from crystal structures and enzyme activity assays, there were three possible carboxylate residues in the active site that were identified as candidates for the general base, Glu132, Asp275 and Glu435 (corresponding to Glu330, Asp463 and Glu599 in the human sequence; Glu121, Asp249 and Glu365 from *Caulobacter*), Figure 3.4a (Vallée et al., 2000a).

A water molecule was positioned close to Glu132 to potentially act as a nucleophile. After further investigation, mutation of the residue from glutamate to glutamine decreased the k_{cat} of the reaction but the enzyme retained residual activity (Lipari and Herscovics, 1999; Vallée et al., 2000a). Subsequently, Glu435 was proposed to be the catalytic base due to its interaction with a water molecule, which was hydrogen bonded to a glycerol (glycerol is known to bind in active sites in the place of carbohydrate moieties) and a calcium ion (Vallée et al., 2000a). Therefore, the most generally accepted residue as the general base is Glu435 (Glu599 from human ERMAnI and Glu365 from *Caulobacter*) (Karaveg et al., 2005).

The remaining residues Glu132 and Asp275 (human Glu330 and Asp463; *Caulobacter* Glu121 and Asp249) were suggested as potential catalytic acids. Asp275 was suggested to be the general acid based on docking studies and analysis of the protonation state of a residue by computationally calculating the pKa (Cantú et al., 2008). Despite this, the

consensus in the field is that Glu132 is most likely to fulfil this role due to its positioning across from the catalytic base in the *anti*-quadrant of the active site. Interestingly, O ϵ 1 was 3.9 Å away from O4 of the substrate (or inhibitor) and O ϵ 2 was 4.5 Å away. Consequently, catalysis by the proton acceptor must be indirect through the participation of another water molecule (O ϵ 2 to water and water to O4 were both 2.8 Å), Figure 3.4a (Vallée et al., 2000b). The catalytic residues in the active site are indirectly coordinated via a water to the α -mannosyl substrate.

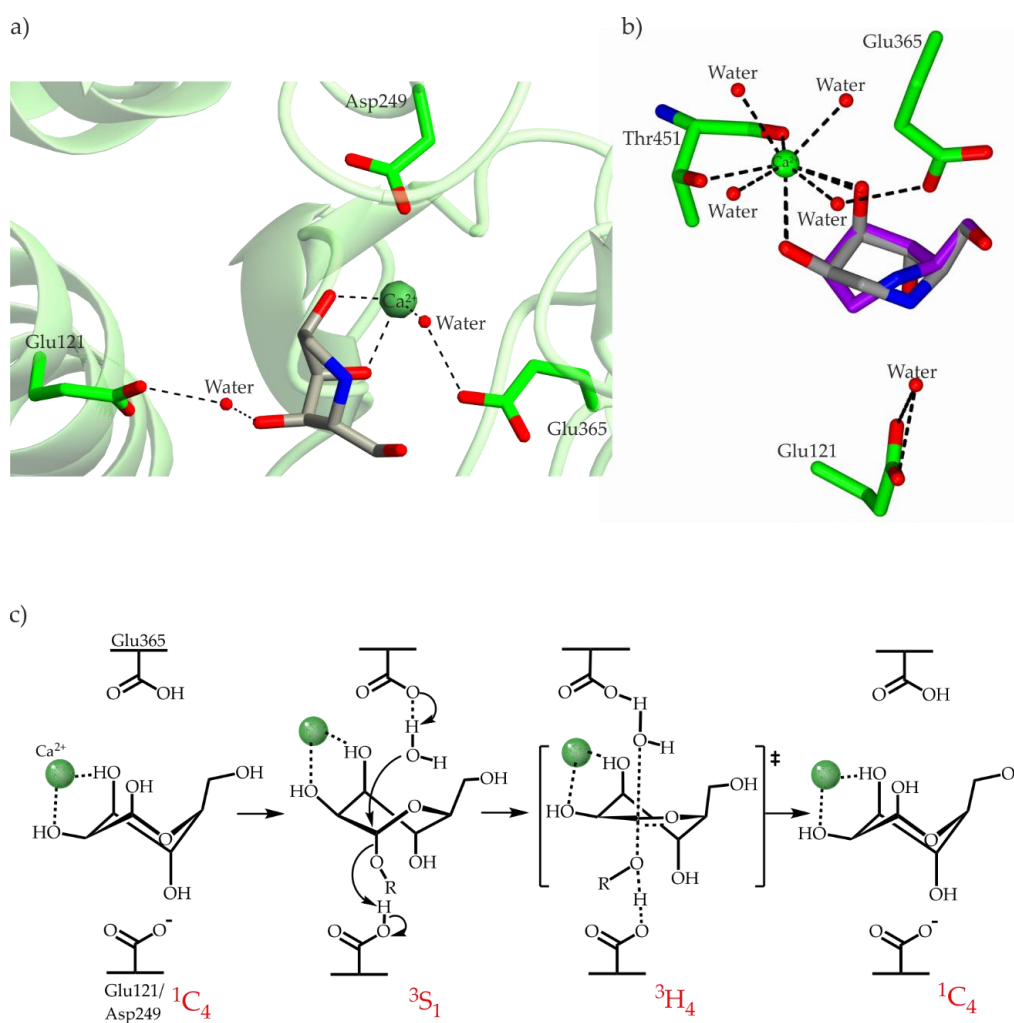


Figure 3.4. Catalytic mechanism of GH47.Numbering is according to the sequence from *Caulobacter* K31 strain GH47. a) and b) *CkGH47*, green, in complex with DMJ PDB ID: 5MEH (van Rijssel et al., 2017). a) Glu365 is the catalytic base and Glu121 is the catalytic acid. b) The calcium ion (Ca^{2+}) is coordinated by 8 components. c) The proposed conformational itinerary of GH47 α -1,2-mannosidase. Mannose ring flips upon binding to the active site of GH47 from a ${}^4\text{C}_1$ (in solution) to ${}^1\text{C}_4$ conformation, the transient intermediate ${}^3\text{S}_1$ distorts to the oxocarbenium ion ${}^3\text{H}_4$ conformation and lastly the transition to the product ${}^1\text{C}_4$ conformation where ${}^4\text{C}_1$ β -mannose is released.

3.2.5.2 Catalysis is a calcium dependent mechanism

Catalysis occurs by a calcium ion dependent mechanism (Vallée et al., 2000b). Located in the -1 subsite of the active site, an important Ca^{2+} ion coordinates the O2 and O3 hydroxyls of the sugar substrate to facilitate ring distortion for catalysis. A second Ca^{2+} in the active site is thought to function as a stabiliser of the β -hairpin plug. The coordination number of the Ca^{2+} ion is 8, hence, the divalent metal ion is surrounded by a conserved network consisting of 4 water molecules and enzyme residues, including the main chain, O, and side chain, O γ 1, of Thr451 and O2 and O3 of the α -mannosyl substrate, Figure 3.4b. One of these water molecules is the pronucleophile, aligned for in-line attack, and bonded to the catalytic base (Thompson et al., 2012).

The importance of the coordination of the GH47 Ca^{2+} metal ion was analysed by mutating Thr688 of ERManI, a residue that coordinates to Ca^{2+} , to alanine. The enzyme was no longer able to catalyse the hydrolysis of Man₉GlcNAc₂ but had increased substrate binding affinity. The same result was achieved by replacement of Ca^{2+} with a lanthanum ion (La^{3+}) allowing the visualisation of high-mannose glycan substrates in complex with inactive GH47 enzymes (Xiang et al., 2016).

GH38 and GH92 α -mannosidase families are also unique in their requirement for a Zn^{2+} or Ca^{2+} metal ion, respectively, in order to perform catalysis (Numao et al., 2003; Zhu et al., 2010). The metal ion required for activity may, therefore, be influenced by the coordination number of the metal, electrostatic interactions with surrounding atoms, chemical and physical properties of the metal and kinetic rates for ligand exchange. Other non-metal dependent enzymes, for example, GH125, require protein residues within the vicinity of the substrate to assist in sugar distortion, see section 2.2.6 (Alonso-Gil et al., 2017; Gregg et al., 2011).

3.2.5.3 The conformational itinerary of GH47 α -1,2-mannosidase

Mannose substrates distort along the unusual conformational itinerary of GH47 enzymes. Firstly, the substrate binds in a ${}^3\text{S}_1$ Michaelis complex conformation, then undergoes a conformational change to ${}^3\text{H}_4$, the transition state, and lastly adopts the product ${}^1\text{C}_4$ conformation, Figure 1.11 and Figure 3.4c (Karaveg et al., 2005; Thompson et al., 2012; Vallée et al., 2000b).

The first insight into the conformational itinerary of GH47s was revealed through a protein:inhibitor complex of *Sc*ERManI and DMJ or kifunensine (Vallée et al., 2000b). Both inhibitors bound in a ring-flipped 1C_4 conformation pertaining to the substrate or product state. A Michaelis complex was obtained when the human orthologue structure was solved in complex with a thiol-linked α -1,2-mannobiose substrate (Karaveg et al., 2005). The conformation of mannose in the -1 subsite was reported as 3S_1 and the conformation of the transition state was assumed to be 3H_4 due to the principle of least nuclear motion.

Computational docking studies (calculating the energies to dock the conformers and the forces on the docked ligand atoms) suggested the pathway proceeded from ${}^1C_4 / {}^0S_2$ (starting conformer) $\rightarrow {}^3H_2 \rightarrow {}^0S_2 \rightarrow {}^3OB \rightarrow {}^3S_1 \rightarrow {}^3E^\ddagger$ (transition state) $\rightarrow {}^1C_4 \rightarrow {}^1H_2 \rightarrow B_{2,5}$ (β -mannose release) (Mulakala et al., 2006, 2007). Another computational study observed the -1 substrate mannobiose in a 3S_1 conformation (Cantú et al., 2008).

Using three different inhibitors, each conformer was identified by X-ray crystallography to confirm a ${}^3S_1 \rightarrow {}^3H_4 \rightarrow {}^1C_4$ reaction pathway. A thiol-linked α -1,2-mannobiose, the same substrate used in the human structure, bound in a Michaelis-complex conformation, mannoimidazole mimicked the transition state and noeuromycin distorted to the product conformation, Figure 3.5 (Thompson et al., 2012). This was the first time a 3H_4 transition state conformation was identified for a glycoside hydrolase (Davies et al., 2003).

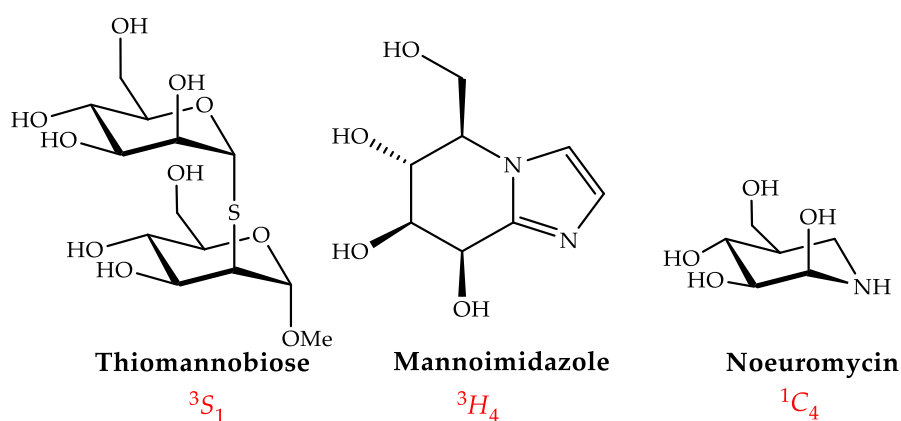


Figure 3.5. Capturing the conformational itinerary of GH47 enzymes.

The FEL of α -D-mannopyranose off-enzyme has a strong preference for a 4C_1 conformation and for the conformers that lie along the equator. The sulfur of the thiol-linked mannobiose bound in the active site of the GH47 structure was computationally substituted for oxygen and used for metadynamic studies. In comparison, the FEL of

mannobiose on-enzyme is strongly biased towards a restricted single conformational itinerary of ${}^3S_1 \rightarrow {}^3H_4 \rightarrow {}^1C_4$, Figure 3.6 (Thompson et al., 2012). The energy difference between the 3S_1 and 1C_4 conformations was 0.5 kcal mol⁻¹ and eluded to a conformational equilibrium between the two conformers. In a structure of DMJ in complex with GH47 from *Caulobacter*, the inhibitor was observed at 50% occupancy of both conformations (3S_1 and 1C_4) (van Rijssel et al., 2017). This could be a flux between Michaelis complex (3S_1) – product (1C_4) or Michaelis complex (1C_4) – transition intermediate (3S_1), Figure 3.4c (Xiang et al., 2016). This ambiguity is still unresolved, although the sugar must distort into a 3S_1 conformation for nucleophilic substitution whereby the C1-H bond is in a pseudo-equatorial orientation.

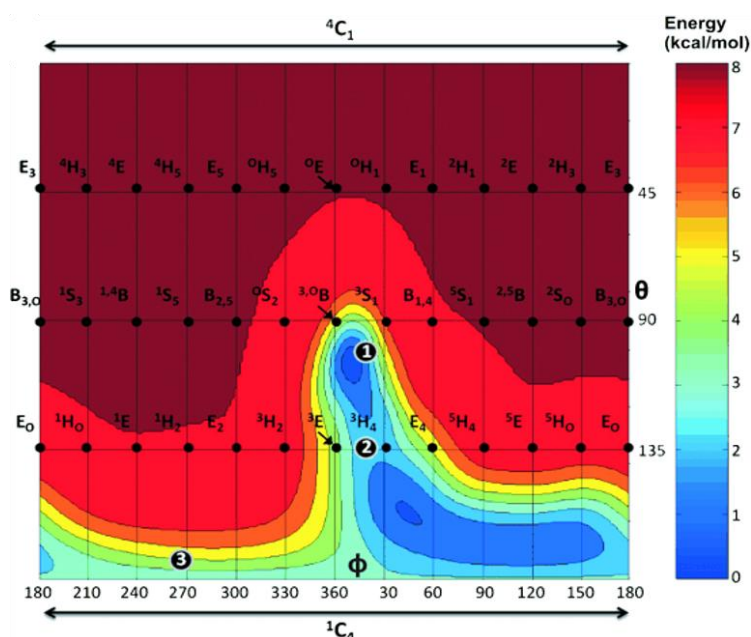


Figure 3.6. Conformational free energy landscape (FEL) of *CkGH47* and the α -mannosyl moiety of different inhibitors in the -1 subsite. The inhibitors are labelled on the FEL: Thiol-linked α -mannobiose (1), mannoimidazole (2) and noeuromycin (3). Figure taken from Thompson et al., 2012.

3.2.6 Specificity of ER and Golgi α -1,2-mannosidases

Bioinformatic sequence and mutant studies of the yeast ER mannosidase revealed the key residue responsible for the specificity of GH47 α -1,2-mannosidases for different high mannose glycan structures. In the *ScERManI* structure, Arg273 (*HsERManI* Arg461) was found at the bottom of the active site and identified as the key residue for specificity due

to the vast number of contacts it makes with an incoming oligosaccharide, specifically, interactions with three mannose sugars (M3, M4 and M7) and one GlcNAc (NAG2), Figure 3.7 (Vallée et al., 2000a). The protein can only hydrolyse $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_8\text{GlcNAc}_2$; Arg273 positions the B branch in the ideal restricted conformation for cleavage between M7 and M10. Golgi α -mannosidases do not have an arginine at the equivalent 273 position, instead, a leucine is present enlargening the size of the active site pocket. The mutant R273L ERManI was capable of cleaving $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_5\text{GlcNAc}_2$ but in a different order since the entropic penalty for binding branches A and C over branch B is lowered (Lobsanov et al., 2002; Romero, 2000). In the fungal orthologue, Arg273 is replaced by a glycine, Gly265, enlarging the cavity further, Figure 3.7. Therefore, the residue at the bottom of the active site controls the orientation of the oligosaccharide and the size of the pocket, hence the difference in specificity (Lobsanov et al., 2002; Romero, 2000; Tempel et al., 2004).

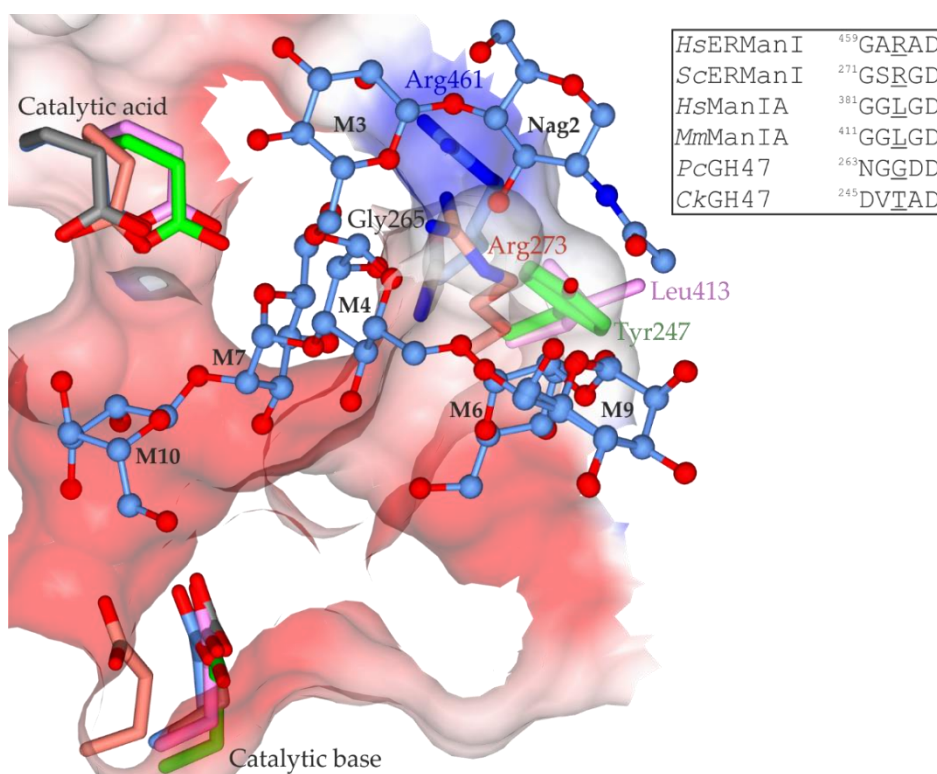


Figure 3.7. The base of the active site determines substrate specificity. Left: Overlay of the important residues for specificity against $\text{Man}_9\text{GlcNAc}_2$ -PA (in blue). *HsERManI* (PDB ID:5KIJ) is shown in blue, *ScERManI* (PDB ID:1DL2) is shown in coral, murine *ManIA* (PDB ID:1NXC) is shown in pink, *PcGH47* (PDB ID:1KKT) is shown in grey and *CkGH47* (PDB ID:5MEH) is shown in green. Right: Partial sequence alignment of the homologues with the first residue numbered.

GH47 from *Caulobacter* K31 strain (*CkGH47*) is thought to be homologous to Golgi mannosidase as it can hydrolyse $\text{Man}_9\text{GlcNAc}_2$ to $\text{Man}_5\text{GlcNAc}_2$. As *CkGH47* is a bacterial protein it does not contain an arginine or a leucine at the equivalent 273 position, alternatively, a tyrosine is present, Figure 3.7. When the structure of *ScERManI* was overlaid with *CkGH47* Tyr247 makes similar interactions with M3, M4 and NAG2.

3.2.7 *Caulobacter* K31 strain

The protein used in this project was cloned from the *Caulobacter* K31 strain (a gram-negative stalked bacteria) and was chosen because it can crystallise to produce atomic resolution X-ray structures (Thompson et al., 2012). This is due to an N-terminal tail that forms interactions with the proteins in the crystal lattice. High resolution structures were necessary to capture the exact conformation of each state along the conformational pathway. Previous attempts to co-crystallise kifunensine with the yeast *ERManI* had failed due to the N-glycan of one symmetry monomer binding into the active site of the other (Vallée et al., 2000b). The addition of kifunensine to the drop blocked this interaction perturbing crystal formation and the N-glycan sterically hindered binding of kifunensine within preformed crystals.

3.2.8 A summary of the results in this chapter

The first paper in this chapter investigates the ability of the potent and naturally occurring inhibitor, kifunensine, to bind to *CkGH47* ($K_D = 39$ nM) in a restricted 1C_4 conformation by QM/MM metadynamics and a 1 Å crystal complex. A comparison is made to other α -mannosidases (GH38 and GH92) for which kifunensine is a poor inhibitor. This is due to a requirement for kifunensine to distort to a higher energy conformation along the conformational pathway of the enzyme.

The second paper explores the design of a conformationally-restricted GH47 inhibitor by exploitation and manipulation of the mannosidase inhibitor DMJ with different C5 substituents. This increased the likelihood of the compound to ring-flip to a 1C_4 conformation upon protonation rather than 4C_1 . Unexpectedly, the sub-Angstrom structure of *CkGH47* in complex with DMJ ($K_D = 481$ nM) revealed two conformations of the inhibitor in the -1 subsite, 3S_1 and 1C_4 , but the mannuronic acid derivatives were unable to bind, possibly due to steric clashes with active site residues.

**Conformational Analysis of the Mannosidase Inhibitor
Kifunensine: A Quantum Mechanical and Structural
Approach**


Declaration

I contributed to the following research publication by providing structural analysis of the conformation of kifunensine bound to α -mannosidase. I conducted crystallisation studies for CkGH47 in complex with kifunensine and isothermal titration calorimetry to compare the binding affinities of different mannosidases for kifunensine. Lluís Raich and Prof. Carme Rovira at the University of Barcelona generated the conformational free-energy landscape providing the quantum mechanical approach to the paper. An equal experimental contribution to the paper was provided by Lluís Raich and myself. Prof. Spencer J. Williams contributed to the initial idea for the paper and provided the figures for scheme 1 and 2. I wrote an initial draft of the paper with guidance from Prof. Gideon J. Davies. The draft was circulated through Lluís Raich, Prof. Carme Rovira and Prof. Spencer J. Williams for modification and proof-reading.



Signed

Prof. Gideon J. Davies



Signed

Prof. Carme Rovira

Reference

Males, A., Raich, L., Williams, S. J., Rovira, C. and Davies, G. J. (2017). Conformational analysis of the mannosidase inhibitor kifunensine: a quantum mechanical and structural approach. *ChemBioChem*. 18, 1496-1501.

Supplementary information - pg 231

Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Structural Approach

Alexandra Males⁺,^[a] Lluís Raich⁺,^[b] Spencer J. Williams,^[c] Carme Rovira,^{*[b, d]} and Gideon J. Davies^{*[a]}

The varied yet family-specific conformational pathways used by individual glycoside hydrolases (GHs) offer a tantalising prospect for the design of tightly binding and specific enzyme inhibitors. A cardinal example of a GH-family-specific inhibitor, and one that finds widespread practical use, is the natural product kifunensine, which is a low-nanomolar inhibitor that is selective for GH family 47 inverting α -mannosidases. Here we show, through quantum-mechanical approaches, that kifunensine is restrained to a “ring-flipped” 1C_4 conformation with another accessible, but higher-energy, region around the 1A_B conformation. The conformations of kifunensine in complex with a range of GH47 enzymes—including an atomic-level resolution (1 Å) structure of kifunensine with *Caulobacter* sp. CkGH47 reported herein and with GH family 38 and 92 α -mannosidases—were mapped onto the kifunensine free-energy landscape. These studies revealed that kifunensine has the ability to mimic the product state of GH47 enzymes but cannot mimic any conformational states relevant to the reaction coordinate of mannosidases from other families.

There is compelling evidence that the enzymatic hydrolysis of glycosides, catalysed by glycoside hydrolases (GHs) or glycosidases, occurs via transition states with significant oxocarbenium ion character. For pyranoside-active enzymes, Sinnott was the first to argue that the allowed canonical conformations of the transition state sugar ring were two half-chairs (4H_3 and 3H_4 ; and their closely related envelope conformations) and two boats (${}^{2,5}B$ and $B_{2,5}$).^[1] Sustained efforts to map the conformational reaction pathways of glycosidases leading from substrate to product via the transition state(s) have revealed that individual glycosidases are optimised to act on substrates and follow a defined conformational itinerary through a specific transition-state conformation.^[2] Glycosidases that are sequence related (for family classification see www.cazy.org and www.cazypedia.org)^[3] and act on sugars with the same configuration are believed to act with identical conformational reaction itineraries. It is also apparent that GHs from different families that act on substrates with the same stereochemical configuration can follow different conformational itineraries during catalysis. Given that mimicry of the enzymatic transition state is a powerful approach to inhibitor design and enzyme inhibition,^[4] the potential exists for the design or discovery of molecules with intrinsically biased conformations that could act as GH-family-specific enzyme inhibitors. However, to achieve specificity, inhibitor design does not have to be limited to transition-state mimicry but could target any distinctive conformation state along the reaction coordinate. Although highly appealing, efforts to design molecular scaffolds that predispose inhibitors to conformations matching that of the transition state of a specific GH have been disappointing. For example, molecular constraints that restrict conformational mobility typically result in steric clashes that prevent efficient binding.^[5] Moreover, a recent study that identified a unique stereoelectronic bias of mannopyranosiduronic iminosugars failed to deliver effective inhibition of a conformationally matched α -mannosidase, presumably because of the inability to accommodate the requisite carboxylate group.^[6] In the context of the failure thus far of rational design methods to achieve conformationally specific inhibition, one compound from nature, kifunensine, stands out as a powerful and family-specific enzyme inhibitor. Kifunensine (**1**; Scheme 1) is an unusual oxalamide-fused iminosugar with high specificity for α -mannosidases of GH47.

Kifunensine is produced by the actinobacterium *Kitasatosporia kifunense* strain no. 9482.^[7] Originally isolated on the basis of immunomodulatory properties, it was soon identified as a potent inhibitor of selected α -mannosidases.^[8] The major

[a] A. Males,⁺ Prof. G. J. Davies

York Structural Biology Laboratory, Department of Chemistry
The University of York
York YO10 5DD (UK)
E-mail: gideon.davies@york.ac.uk

[b] L. Raich,⁺ Prof. C. Rovira

Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) and
Institut de Química Teòrica i Computacional (IQTCUB) Universitat de Barcelona
Martí i Franquès 1, 08028 Barcelona (Spain)
E-mail: c.rovira@ub.edu

[c] Prof. S. J. Williams School
of Chemistry and

Bio21 Molecular Science and Biotechnology Institute University of Melbourne
Parkville, VIC 3010 (Australia)

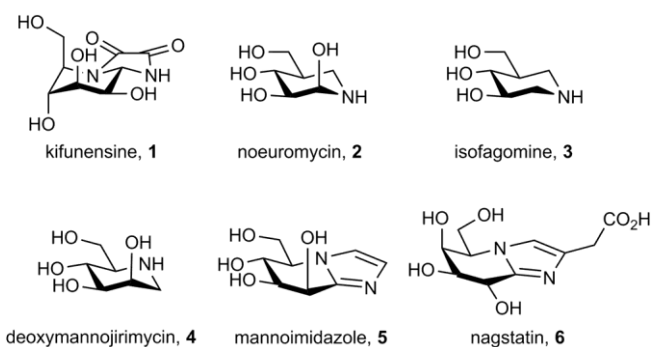
[d] Prof. C. Rovira

Institució Catalana de Recerca i Estudis Avançats (ICREA) Passeig Lluís Companys 23, 08010 Barcelona (Spain)

[*] These authors contributed equally to this work

Supporting information and ORCID identification numbers for the authors of this article can be found under <https://doi.org/10.1002/cbic.201700166>.

© 2017 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

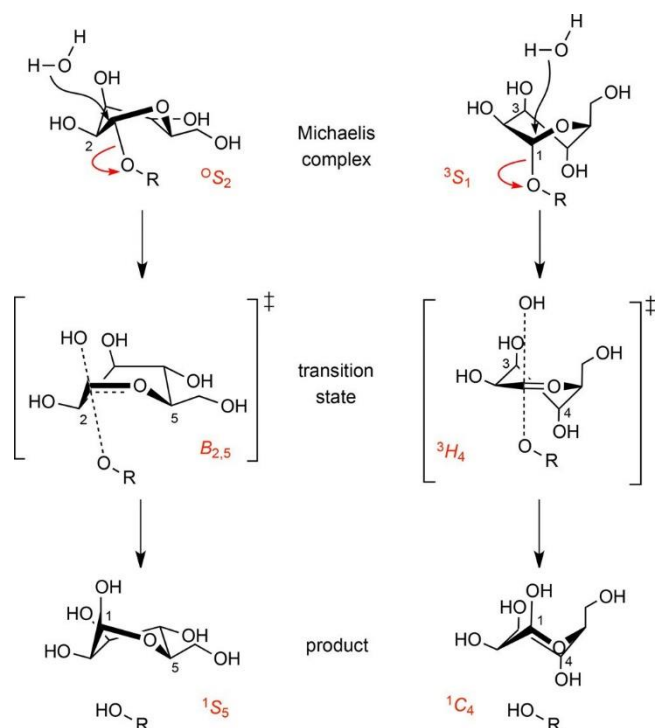


Scheme 1. Assorted glycosidase inhibitors.

contemporary applications of kifunensine stem from its activity as a specific inhibitor of the class I GH47 α -mannosidases involved in glycoprotein biosynthesis within the secretory pathway that is specific for the hydrolysis of α -1,2-glycosidic bonds.^[9] During glycoprotein biosynthesis, glucosylated high-mannose N-glycans are appended to nascent unfolded peptide chains in the endoplasmic reticulum (ER), whereupon they undergo a range of trimming reactions in the ER that assist in folding the peptide, and subsequent trimming reactions in the Golgi apparatus that remove additional mannose residues prior to late-stage glycosylation reactions.^[10] As part of this process, a quality control mechanism termed ER-associated degradation (ERAD) extracts terminally misfolded proteins from the secretory pathway for proteosomal degradation.^[11] Kifunensine-sensitive α -mannosidases are found within both the normal trimming pathway (ER mannosidase I, Golgi mannosidase I), and within the ERAD pathway (ER degradation-enhancing mannosidase-like proteins 1, 2 and 3).^[12]

The powerful and specific inhibition of GH47 α -mannosidases by kifunensine has led to its widespread use for manipulating the N-glycan structure. In the structural-biology context it is used to improve the homogeneity and crystallisation of proteins by arresting glycan remodelling so as to yield high-mannose glycans that are more easily cleaved by endoH.^[13] Kifunensine is also used in the production of therapeutic proteins. The effectiveness of acid β -glucocerebrosidase as a treatment for the lysosomal storage disorder Gaucher's disease depends upon the presence of high-mannose N-glycans of this protein, which enable ligation to mannose receptors and delivery to lysosomes.^[14] The lysosomal-replacement-therapy protein Velaglucerase alfa (acid β -glucocerebrosidase) is produced by Shire Plc by culturing HT1080 fibrosarcoma cells expressing acid β -glucocerebrosidase in the presence of kifunensine—this promotes the biosynthesis of enzyme decorated with immature high-mannose-type N-linked glycan chains.^[15]

Structural studies on GH47 enzymes, firstly the seminal work by the Howell group on the human ER ManB1 α -1,2-mannosidase,^[16] and subsequent work on a *Penicillium citrinum* homologue^[12c] revealed that kifunensine binds in a “ring-flipped” 1C_4 conformation. This conformation is unusual for an iminosugar. Relative to the proposed conformational pathway ${}^3S_1 \rightarrow [{}^3H_4]^\ddagger \rightarrow {}^1C_4$ (Scheme 2)^[2b,16,17] for this family of enzymes, it pro-



Scheme 2. Conformational itineraries for inverting α -mannosidases proceeding through left: ${}^0S_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^1S_5$ (e.g., GH125)^[24] and right: ${}^3S_1 \rightarrow [{}^3H_4]^\ddagger \rightarrow {}^1C_4$ (GH47) pathways. Catalytic residues acting as proton donors/acceptors are omitted for clarity.

vides conformational mimicry of the product. Of relevance to this observation, a complex of noeuromycin **2** bound to a bacterial α -mannosidase of family GH47, *Caulobacter* sp. K31 (*CkGH47*) also adopted a 1C_4 conformation that, on the basis of computational work, was assigned as a product-mimicking species.^[17a] However, in the case of noeuromycin, computational analysis of the inhibitor reveals that it favours a 4C_1 conformation,^[18] and the observed conformation on-enzyme would therefore appear to be a consequence of the enzyme restricting its shape to a higher-energy conformation. Because of the widespread application of kifunensine, and its unique specificity as a GH47-selective α -mannosidase inhibitor, we sought to understand its specificity by developing a quantitative view of the conformational preferences of this compound when isolated, relative to bound states on α -mannosidases from GH47 and other families.

Structures are available for kifunensine bound to *Homo sapiens*^[16] and *P. citrinum*^[12c] family 47 α -mannosidases at medium resolutions of 1.75 and 2.20 Å, respectively. In order to allow comparison to conformations that the α -mannoside substrate follows during catalysis, we sought to obtain a higher-resolution structure of **1** with *CkGH47* α -mannosidase, which we have previously shown to provide atomic resolution diffraction data (Table 1). Kifunensine, shown here using isothermal titration calorimetry, is a tight binding ($K_D = 39$ nM) inhibitor of *CkGH47* (Figure 1 A). Crystals of a complex of *CkGH47* bound to kifunensine diffracted to 1.05 Å resolution (Figure 1 B). Kifunensine binds to *CkGH47* in a 1C_4 conformation, in line with

Table 1. Data and structure quality for 1 in complex with CkGH47 (PDB ID: 5NE5).	
resolution [Å]	72–71.05
R_{merge}	0.090 (0.88)
CC(1/2)	1.00 (0.53)
completeness [%]	99.2 (99.6)
$R_{\text{work}}/R_{\text{free}}$	0.12/0.14
RMSD _{bonds} [Å]	0.020

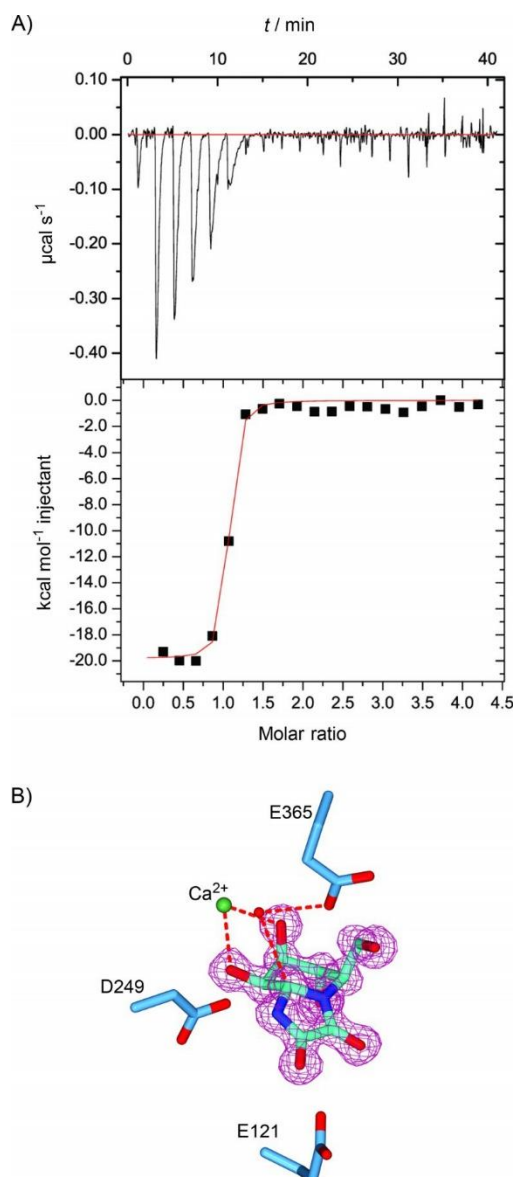


Figure 1. Kifunensine is a potent inhibitor of *Caulobacter* sp. K31 (*CkGH47*) α -mannosidase and binds in a 1C_4 conformation. A) Thermodynamics of binding between *CkGH47* and kifunensine obtained by ITC. The raw data are shown in the injection profile (top) and the titration curve (bottom); $n = 0.98 \pm 0.01$ sites $K_A = (2.6 \pm 1.1) \times 10^7 \text{ M}^{-1}$, $\Delta H = (19.8 \pm 0.3) \text{ kcal mol}^{-1}$.

B) Complex with kifunensine showing the catalytic residues, the Ca^{2+} ion, and the nucleophilic water molecule (red sphere). The electron density map is a REFMAC maximum-likelihood/ σA -weighted $2F_o - F_c$ synthesis contoured at 0.80 electrons \AA^{-3} .

the conformation observed in the structures with the *H. sapiens* and *P. citrinum* GH47 α -mannosidases.

Conformational free-energy landscapes (FELs) are quantitative maps of the energy of the full suite of conformations of a molecule, and provide insight into locally and globally stable conformations as well as the barriers that must be crossed in order to achieve them. In order to understand the conformational landscape of kifunensine, and how this contributes to its GH47 specific inhibition, its conformational FEL was calculated by ab initio metadynamics^[19] according to methods we have developed and applied to a range of glycosidase inhibitors.^[18,20] The kifunensine FEL (Figure 2 A) exhibits a strong preference for the “southern hemisphere” 1C_4 conformation; this is consistent with the conformation observed in a small-molecule, single-crystal X-ray structure.^[8,21] A low energy zone is also seen around the ${}^1S_3/{}^{1,4}B/{}^1S_5$ region of the FEL, which, at approximately 7–8 kcal mol⁻¹ higher in energy and with a minimal barrier, can be considered energetically accessible. Overall, the isolated kifunensine FEL is strongly constrained to a ring-flipped 1C_4 conformation, with a limited ability to adopt a small set of “equatorial” conformations. Plotting the conformations of kifunensine observed in representative GH47 crystal structures (for enzymes from *H. sapiens* and *Bacteroides thetaiotaomicron*) onto the isolated kifunensine FEL reveals that they lie close together in the low energy zone, with similar conformations to that of the small-molecule, single-crystal X-ray structure, thus suggesting that little distortion occurs upon binding to GH47 enzymes—the apparent differences near the poles are a consequence of the deformation of the Mercator projection (for a Polar projection and direct comparison, see Figure S1 in the Supporting Information). In concordance with the FEL for **1**, an FEL calculated for methyl α -mannoside within the active centre environment of the *CkGH47* (Figure 2 B) revealed that this molecule is heavily distorted away from its preferred 4C_1 conformation when bound to the enzyme.^[17a]

The conformational bias of kifunensine towards the southern hemisphere is unusual for a glycosidase inhibitor. For example, FELs of the azasugar isofagomine reveal that this molecule favours a “normal” 4C_1 conformation,^[20b] which is also expected for the iminosugar deoxymannojirimycin **4**. A range of iminosugars containing sp^2 -hybridised atoms, best illustrated by mannoimidazole **5**, favour half-chair (4H_3) and boat ($B_{2,5}$) conformations that lie along the FEL tropic/equator.^[20] One unusual example worth mentioning that is similar what is observed here is a mannoypyranosiduronic iminosugar that, under acidic conditions, favours a 1C_4 chair, as shown by ${}^1\text{H}$ NMR analysis of vicinal coupling constants; however, the presence of the carboxylate group prevents binding to α -mannosidases.^[6]

Although kifunensine typically binds to GH47 α -mannosidases with nanomolar dissociation constants, it is usually a far poorer binder/inhibitor of GH families other than family 47. The available evidence suggests that all other *exo*-mannosidases operate through ${}^0S_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^1S_5$ conformational pathways [GH families 2 (retaining),^[5b] 38 (retaining),^[22] 92 (inverting),^[23] 125 (inverting)^[24] and 130 (inverting)^[25]]. Data are available for kifunensine binding to representatives of families 38 and 92. For the *Drosophila melanogaster* Golgi class II GH38 α -mannosi-

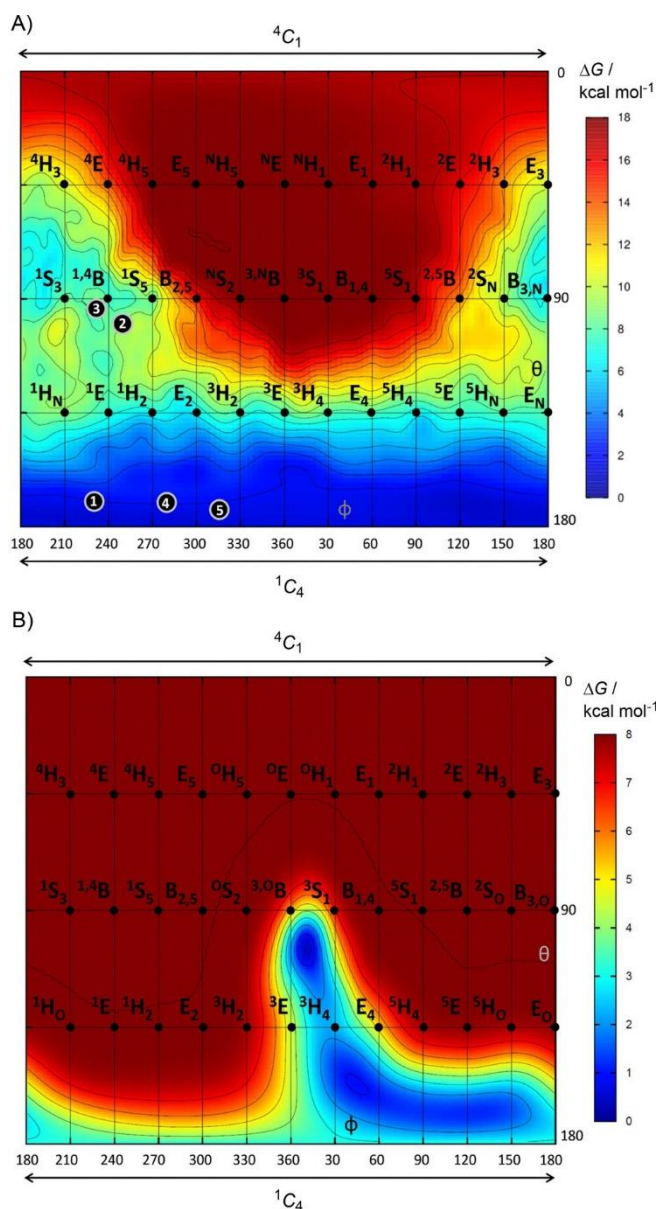


Figure 2. Mercator projections of conformational FELs of kifunensine and methyl α -mannopyranoside obtained by ab initio metadynamics simulations.

A) FEL for kifunensine in vacuo. The indicated observed kifunensine conformations are for the following α -mannosidases: 1) *H. sapiens* GH47 (PDB ID: 1FO3), 2) *B. thalotomycin* GH92 (PDB ID: 2WVZ), 3) *D. melanogaster* GH38 (PDB ID: 1PS3), 4) *Caulobacter* sp. CkGH47 (this work) and 5) the small molecule crystal structures of kifunensine (KASNOH/KASNOH10). (3), (4) and (5) adopt almost identical conformations, the apparent differences reflect distortions nearing the poles of a Mercator projection (see Figure S1 for an alternative Polar projection). B) FEL for methyl α -mannopyranoside bound to CkGH47 from ref. [17a]. Contour lines are separated by 1 kcal mol⁻¹. Angles (x -/ y -axes) are given in degrees.

dase, dGMII, **1** has a K_i value of 5 mM,^[26] and it is reported to be a very poor inhibitor of jack bean GH38 α -mannosidase.^[8,12a] Similarly, for a range of *B. thalotomycin* GH92 α -mannosidases, **1** is reported to have K_i values in the range of 100–200 mM.^[23] Structures of **1** bound to GH38 (PDB ID: 1PS3) and GH92 (PDB ID: 2WVZ)^[23] α -mannosidases in both cases revealed a ^{1,4}B conformation. Consideration of these results in

the context of the kifunensine FEL reveals that, although this conformation lies in the energetically accessible equatorial region (Figure 2 A), it also lies some distance away from the ⁰S₂ → [⁰B_{2,5}][‡] → ¹S₅ conformational pathway proposed for both GH38 and GH92 α -mannosidases; this suggests that the poor inhibition results from an inability of the inhibitor to adopt a conformation that matches species formed on the enzyme during catalysis.

Our conformational analysis of kifunensine **1** highlights the unique ability of this compound to target a region of the FEL that is not accessed by other GH inhibitors. We suggest that the specificity and potency of **1** for GH47 family enzymes is a direct consequence of its strong preference for the southern hemisphere ¹C₄ conformation, which matches that of the product state of the ³S₁ → [³H₄][‡] → ¹C₄ conformational itinerary that this family of enzymes is believed to follow.^[2b,16,17] We highlight that this analysis suggests that potency is achieved in the absence of transition state mimicry. It is surprising that this conformational restraint is achieved by introducing an oxalamide bridge without interfering with binding in the active site, as a previous attempt with a similar goal to synthetically introduce a bridge into an iminosugar failed owing to steric clashes in the active site.^[5] The preference of kifunensine for a ¹C₄ conformation likely arises not merely from the fusion of the bridge to the ring, but also from the sp² hybridisation of the endocyclic nitrogen as part of an amide. Further, we note that kifunensine is a neutral species; this shows that, unlike most aza/iminosugar glycosidase inhibitors, kifunensine achieves its potency through shape, rather than charge. For a discussion of shape versus charge mimicry for the inhibition of a GH99 α -mannosidase see ref. [18]; for an example of a potent class of neutral GH inhibitors see ref. [27].

In conclusion, this work highlights the capacity of natural products to provide inspiration for selectively inhibiting glycosidases based on mechanistic principles. Aligned with this goal, we highlight the inspiration provided by the natural product nagstatin (**6**),^[28] which informed the development of the concept of lateral protonation by GHs with a catalytic acid or acid/base located *anti* to the C1-O5 bond, and the design of the glycoimidazole-class inhibitors that are selective for *anti*-protonating GHs.^[29] The existence of kifunensine as a GH47-selective inhibitor should inspire continuing efforts to develop selective and potent GH inhibitors based on targeting unique conformational features of their catalysis reaction coordinates. We highlight that one other mannosidase family operates through a reversed ¹C₄ → [³H₄][‡] → ³S₁ conformational itinerary, namely GH134 β -mannanases.^[30] Based on the analysis presented herein, this family is likely to be specifically targeted by substrate-mimicking kifunensine-derived oligosaccharides in contrast to the β -mannanases of GH26 and GH113, which instead operate through ⁰S₂ ↔ [⁰B_{2,5}][‡] ↔ ¹S₅ conformational pathways.^[20b]

Experimental Section

CkGH47 protein was cloned, expressed and purified as described previously.^[17a] The crystallisation conditions were the same as discussed in ref. [6]. Mature crystals were soaked in kifunensine

(1 mm) for 15 h. Data were collected on the Diamond I04-1 MX beamline. After data collection, the diffraction images were integrated by using XIA2^[31] and reintegrated by using AIMLESS from the CCP4 software suite.^[32] The Free R flag data set was copied from PDB ID: 4AYO. The model was refined by using multiple rounds of REFMAC and manual model building in COOT.^[33] Water molecules were added by using FIND WATERS in COOT and validated. The conformation of the ligand was validated by using PRIVATEER. Coordinates have been deposited in the PDB with ID 5NE5; details of refinement quality are shown in Table 1 and the PDB header. Figures of the structure were produced by using CCP4mg.^[34]

ITC was performed by using a MicroCal ITC₂₀₀ calorimeter at 25 °C with 20 injections. CkGH47 and kifunensine **1** were transferred into matching buffer by dialysis into HEPES (25 mM, pH 7.0) containing NaCl (50 mM) and CaCl₂ (2 mM). The CkGH47 concentration in the cell was 50 μM, and the ligand concentration was 500 μM. The binding affinity was calculated by using Origin (OriginLab, Northampton, MA).

The free-energy landscape of kifunensine was obtained by density functional theory-based metadynamics^[19] by using the Car–Parrinello (CP) method.^[35] The molecule was enclosed in an isolated cubic box of 14.0 x 14.0 x 14.0 Å. A fictitious electron mass of 700 au and a time step of 0.12 fs ensured a proper conservation of the total energy during the simulation. The Kohn–Sham orbitals were expanded on a plane wave (PW) basis set with a kinetic energy cutoff of 70 Ry. Ab initio pseudopotentials, generated within the Troullier–Martins scheme,^[36] were employed. The Perdew, Burke and Ernzerhoff generalised, gradient-corrected approximation (PBE)^[37] was selected in view of its good performance in our previous work.^[38] Two collective variables of the puckering coordinates of Cremer and Pople (θ and ϕ) were used to explore the conformational space.^[39] Initially, the height/width of these Gaussian terms was set at 0.6 kcal mol⁻¹/0.10 rad, and a new Gaussian-like potential was added every 250 MD steps. Once the whole free-energy space had been explored, the height of the Gaussian terms was reduced to half of its initial value (0.3 kcal mol⁻¹), and a new Gaussian-like potential was added every 500 MD steps. The simulation was stopped when energy differences among wells remained constant. The phase space was fully explored in less than 50 ps, and the simulation was further extended up to 95 ps. The error in the energy difference of the principal minima, taken as a standard deviation (SD) from the last 30 ps, is below 0.6 kcal mol⁻¹.

Acknowledgements

We thank Diamond Light Source for access to beamline I04-1 (proposal number mx-13587), which contributed to the results presented here. G.J.D. is the Royal Society Ken Murray Research Professor. S.J.W. is an Australian Research Council Future Fellow (FT130100103). A.M. and L.R. are supported through BBSRC (BB/M011151/1) and Ajuts de Personal Investigador predoctoral en Formació–Universitat de Barcelona (APIF-UB) PhD studentships, respectively. We thank the Spanish Ministry of Economy and Competitiveness (grant CTQ2014-55174-P to C.R.) and AGAUR (grant SGR2014-987 to C.R.), The authors gratefully acknowledge the computer resources at MareNostrum and the technical support provided by BSC-CNS (RES-QCM-2016-3-00017).

Conflict of interest

The authors declare no conflict of interest.

Keywords: ab initio calculations · carbohydrates · enzymes · hydrolases · iminosugar

- [1] M. L. Sinnott, *Chem. Rev.* **1990**, *90*, 1171–1202.
- [2] a) G. Speciale, A. J. Thompson, G. J. Davies, S. J. Williams, *Curr. Opin. Struct. Biol.* **2014**, *28*, 1–13; b) G. J. Davies, A. Planas, C. Rovira, *Acc. Chem. Res.* **2012**, *45*, 308–316.
- [3] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490–D495.
- [4] a) V. L. Schramm, *ACS Chem. Biol.* **2013**, *8*, 71–81; b) M. M. Mader, P. A. Bartlett, *Chem. Rev.* **1997**, *97*, 1281–1301.
- [5] a) M. Böhm, E. Lorthiois, M. Meyyappan, A. Vasella, *Helv. Chim. Acta* **2003**, *86*, 3818–3835; b) L. E. Tailford, W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M. P. Heck, R. V. Stick, Y. Bleriot, A. Vasella, H. J. Gilbert, G. J. Davies, *Nat. Chem. Biol.* **2008**, *4*, 306–312.
- [6] E. van Rijssel, A. Janssen, A. Males, G. Davies, G. van der Marel, H. S. Overkleeft, J. Codée, *ChemBioChem* **2017**, *18*, 1297–1304.
- [7] M. Iwami, O. Nakayama, H. Terano, M. Kohsaka, H. Aoki, H. Imanaka, *J. Antibiot.* **1987**, *40*, 612–622.
- [8] H. Kayakiri, S. Takase, T. Shibata, M. Okamoto, H. Terano, M. Hashimoto, T. Tada, S. Koda, *J. Org. Chem.* **1989**, *54*, 4015–4016.
- [9] a) S. W. Mast, K. W. Moremen, *Methods Enzymol.* **2006**, *415*, 31–46; b) T. Kuribara, M. Hirano, G. Speciale, S. J. Williams, Y. Ito, K. Totani, *ChemBioChem* **2017**, *18*, 1027–1035.
- [10] R. Kornfeld, S. Kornfeld, *Annu. Rev. Biochem.* **1985**, *54*, 631–664.
- [11] A. A. McCracken, J. L. Brodsky, *J. Cell Biol.* **1996**, *132*, 291–298.
- [12] a) A. D. Elbein, J. E. Tropea, M. Mitchell, G. P. Kauschal, *J. Biol. Chem.* **1990**, *265*, 15599–15605; b) N. Hosokawa, I. Wada, K. Hasegawa, T. Yorihuzi, L. O. Tremblay, A. Herscovics, K. Nagata, *EMBO Rep.* **2001**, *2*, 415–422; c) Y. D. Lobsanov, F. Vallee, A. Imbert, T. Yoshida, P. Yip, A. Herscovics, P. L. Howell, *J. Biol. Chem.* **2002**, *277*, 5620–5630; d) T. Zhou, D. A. Frabutt, K. W. Moremen, Y.-H. Zheng, *J. Biol. Chem.* **2015**, *290*, 22184–22192.
- [13] C. Yu, M. Crispin, A. F. P. Sonnen, D. J. Harvey, V. T. Chang, E. J. Evans, C. N. Scanlan, D. I. Stuart, R. J. C. Gilbert, S. J. Davies, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67*, 785–789.
- [14] B. Brumshtein, P. Salinas, B. Peterson, V. Chan, I. Silman, J. L. Sussman, P. J. Savickas, G. S. Robinson, A. H. Futerman, *Glycobiology* **2010**, *20*, 24–32.
- [15] T. M. Cox, *Am. J. Hematol.* **2013**, *88*, 163–165.
- [16] F. Vall8e, K. Karaveg, A. Herscovics, K. W. Moremen, P. L. Howell, *J. Biol. Chem.* **2000**, *275*, 41287–41298.
- [17] a) 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, G. J. Davies, *Angew. Chem. Int. Ed.* **2012**, *51*, 10997–11001; *Angew. Chem.* **2012**, *124*, 11159–11163; b) K. Karaveg, A. Siriwardena, W. Tempel, Z.-J. Liu, J. Glushka, B.-C. Wang, K. W. Moremen, *J. Biol. Chem.* **2005**, *280*, 16197–16207.
- [18] M. Petricevic, L. F. Sobala, P. Z. Fernandes, L. Raich, A. J. Thompson, G. Bernardo-Seisdedos, O. Millet, S. Zhu, M. Sollogoub, J. Jiménez-Barbero, C. Rovira, G. J. Davies, S. J. Williams, *J. Am. Chem. Soc.* **2017**, *139*, 1089–1097.
- [19] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.
- [20] a) A. Tankrathok, J. Iglesias-Fernández, R. J. Williams, S. Pengthaisong, S. Baiya, Z. Hakkii, R. C. Robinson, M. Hrmova, C. Rovira, S. J. Williams, J. R. Ketudat Cairns, *ACS Catal.* **2015**, *5*, 6041–6051; b) R. J. Williams, J. Iglesias-Fernández, J. Stepper, A. Jackson, A. J. Thompson, E. C. Lowe, J. M. White, H. J. Gilbert, C. Rovira, G. J. Davies, S. J. Williams, *Angew. Chem. Int. Ed.* **2014**, *53*, 1087–1091; *Angew. Chem.* **2014**, *126*, 1105–1109.
- [21] H. Kayakiri, S. Takase, T. Sibata, M. Hashimoto, T. Tada, S. Koda, *Chem. Pharm. Bull.* **1991**, *39*, 1378–1381.
- [22] S. Numao, D. A. Kuntz, S. G. Withers, D. R. Rose, *J. Biol. Chem.* **2003**, *278*, 48074–48083.

- [23] Y. Zhu, M. D. Suits, A. J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K. W. Moremen, Y. Xiang, A. Siriwardena, S. J. Williams, H. J. Gilbert, G. J. Davies, *Nat. Chem. Biol.* **2010**, *6*, 125–132.
- [24] S. Alonso-Gil, A. Males, P. Z. Fernandes, S. J. Williams, G. J. Davies, C. Rovira, *J. Am. Chem. Soc.* **2017**, *139*, 1085–1088.
- [25] F. Cuskin, A. Basl8, S. Ladevèze, A. M. Day, H. J. Gilbert, G. J. Davies, G. Potocki-Véronèse, E. C. Lowe, *J. Biol. Chem.* **2015**, *290*, 25023–25033.
- [26] N. Shah, D. A. Kuntz, D. R. Rose, *Biochemistry* **2003**, *42*, 13812–13816.
- [27] S. J. Williams, V. Notenboom, J. Wicki, D. R. Rose, S. G. Withers, *J. Am. Chem. Soc.* **2000**, *122*, 4229–4230.
- [28] a) T. Aoyagi, H. Suda, K. Uotani, F. Kojima, T. Aoyama, K. Horiguchi, M. Hamada, T. Takeuchi, *J. Antibiot.* **1992**, *45*, 1404–1408; b) K. Tatsuta, S. Miura, S. Ohta, H. Gunji, *J. Antibiot.* **1995**, *48*, 286–288.
- [29] T. D. Heightman, A. T. Vasella, *Angew. Chem. Int. Ed.* **1999**, *38*, 750–770; *Angew. Chem.* **1999**, *111*, 794–815.
- [30] 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, G. J. Davies, *ACS Cent. Sci.* **2016**, *2*, 896–903.
- [31] G. Winter, *J. Appl. Crystallogr.* **2010**, *43*, 186–190.
- [32] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 235–242.
- [33] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.
- [34] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 386–394.
- [35] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- [36] N. Troullier, J. L. Martins, *Phys. Rev. B* **1991**, *43*, 1993–2006.
- [37] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [38] A. Ardèvol, C. Rovira, *J. Am. Chem. Soc.* **2015**, *137*, 7528–7547. [39] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

Manuscript received: March 23, 2017

Accepted manuscript online: May 11, 2017

Version of record online : June 26, 2017

Conformational Behaviour of Mannuronic Acid Based Azasugars

Declaration

The research was instigated by Dr. Jeroen D. C. Codée, Prof. Herman S. Overkleeft and Prof. Gijsbert A. van der Marel. The synthesis of the mannuronic acid derivatives and NMR spectroscopy experiments were conducted by Dr. Erwin R. van Rijssel and Dr. Antonius P. A. Janssen. I conducted X-ray crystallography experiments using the α -1,2-mannosidase, *CkGH47*, to attempt to obtain ligand complexes with DMJ (2) and other mannuronic acid derivatives (3, 4 and 5). I also conducted ITC and mutational studies to obtain a dissociation constant and to investigate the binding of the mannuronic acid derivatives. Together with Professor Gideon J. Davies, the sections in the article describing the structure and conformation of DMJ and *CkGH47*, methods and data statistics tables were written. The final version included input from all authors.



Signed

Professor Gideon J. Davies



Signed

Dr. Jeroen D. C. Codée

Reference:

van Rijssel, E. R., Janssen, A. P. A., Males, A., Davies, G. J., van der Marel, G. A., Overkleeft, H. S., and Codée, J. D. C. (2017). Conformational Behaviour of Azasugars Based on Mannuronic Acid. *ChemBioChem*. 18, 1297-1304.

Supplementary information - pg 233

SPECIAL
ISSUE

Conformational Behaviour of Azasugars Based on Mannuronic Acid

Erwin R. van Rijssel,^[a] Antonius P. A. Janssen,^[a] Alexandra Males,^[b] Gideon J. Davies,^[b] Gijsbert A. van der Marel,^[a] Herman S. Overkleef, ^{*[a]} and Jeroen D. C. Codée ^{*[a]}

Dedicated to the memory of Professor Dr. Werner Reutter.

A set of mannuronic-acid-based iminosugars, consisting of the C-5-carboxylic acid, methyl ester and amide analogues of 1-deoxymannorjirimycin (DMJ), was synthesised and their pH-dependent conformational behaviour was studied. Under acidic conditions the methyl ester and the carboxylic acid adopted an “inverted” 1C_4 chair conformation as opposed to the “normal” 4C_1 chair at basic pH. This conformational change is explained in terms of the stereoelectronic effects of the ring substituents and it parallels the behaviour of the mannuronic

acid ester oxocarbenium ion. Because of this solution-phase behaviour, the mannuronic acid ester azasugar was examined as an inhibitor for a *Caulobacter* GH47 mannosidase that hydrolyses its substrates by way of a reaction itinerary that proceeds through a 3H_4 transition state. No binding was observed for the mannuronic acid ester azasugar, but sub-atomic resolution data were obtained for the DMJ·CkGH47 complex, showing two conformations— 3S_1 and 1C_4 —for the DMJ inhibitor.

Introduction

Stereoelectronic substituent effects have a profound effect on the three-dimensional structures of molecules. Whereas substituents on a cyclic compound generally have a preference for (pseudo)equatorial positions for steric reasons, the electronic spatial preferences depend on different forces such as charge–charge and dipole–dipole interactions.^[1] The conformation and reactivity of carbohydrates are determined to a large extent by the natures and orientations of the substituents. This influence becomes apparent in glycosylation reactions, in which the amount, nature and orientation of the hydroxy groups, protected with electron-withdrawing esters or more electron-neutral ether groups, determine the overall reactivity.^[2]

It has long been known that, in glycosylations, axial substituents are less deactivating or “disarming” than their equatorially

positioned equivalents.^[3] Similarly, the basicity of iminosugars (or “azasugars”), carbohydrates in which the endocyclic oxygen atom is replaced by an amine group, is influenced by the orientation of the ring substituents, with azasugars bearing more axially positioned hydroxy groups being more basic than their stereoisomers bearing equatorially positioned substituents.^[4] These effects can be explained in terms of more favourable interaction of the axially positioned electronegative oxygen substituents with the positive charge present on the azasugar ring in a protonated state and the (partial) positive charge of oxocarbenium ion (-like) intermediates in glycosylation reactions.^[4–7]

In mannuronic acids, mannosides in which the C6-OH group is oxidised to a carboxylic acid functionality, the carboxylic acid has a profound effect on the conformation and reactivity of the pyranoside.^[8] In the context of the construction of bacterial oligosaccharides we have studied the glycosylation behaviour of a variety of mannuronic acid donors in detail and we have found these to be unexpectedly reactive.^[9] In addition, glycosylations involving these donors proceed with an extraordinary selectivity to provide 1,2-*cis* glycosidic linkages. These findings were explained in terms of the conformational preferences of (partially) positively charged mannuronic acid oxocarbenium ion (-like) intermediates that are governed by the ring substituent effects. These species prefer to adopt a “flipped” ring structure and in the 3H_4 (-like) oxocarbenium ion all substituents take up the most stabilising (or least destabilising) orientation: the C2-OR pseudoequatorial and the C3-, C4-OR and C5-COOR groups pseudoaxial. Indeed, DFT calculations indicate that the 3H_4 oxocarbenium ion is significantly more stable than the alternative (“non-ring-flipped”) 4H_3 ion (Scheme 1 A).^[10]

[a] Dr. E. R. van Rijssel, A. P. A. Janssen, Prof. Dr. G. A. van der Marel, Prof.

Dr. H. S. Overkleef, Dr. J. D. C. Codée

Leiden Institute of Chemistry, Leiden University

Einsteinweg 55, 2333 CC Leiden (The Netherlands)

E-mail: jcodee@chem.leidenuniv.nl

h.s.overkleef@chem.leidenuniv.nl

[b] A. Males, Prof. Dr. G. J. Davies York

Structural Biology Laboratory

Department of Chemistry, The University of York

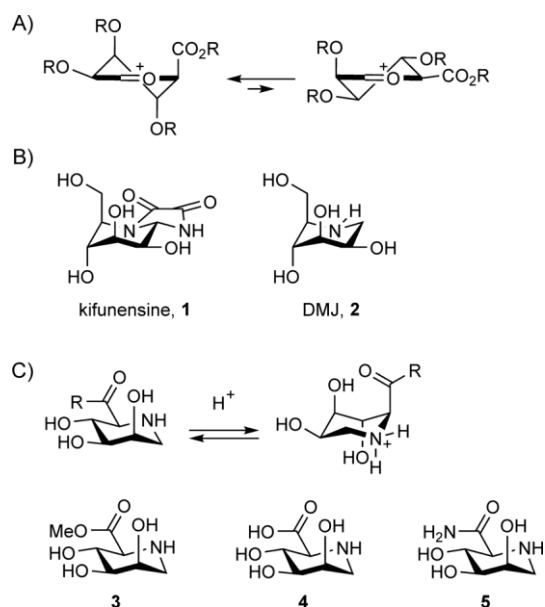
York YO10 5DD (UK)

Supporting information and the ORCID identification numbers for the authors of this article can be found under <http://dx.doi.org/10.1002/cbic.201700080>.

© 2017 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.



This manuscript is part of a Special Issue on Glycobiology, dedicated to the late Werner Reutter.



Scheme 1. A) Conformational equilibrium for the mannuronic acid oxocarbenium ion. B) Structures of the mannosidase inhibitors kifunesine (1) and 1-deoxymannojirimycin (DMJ, 2) in the 1C_4 conformation. C) Mannuronic acid azasugars studied here.

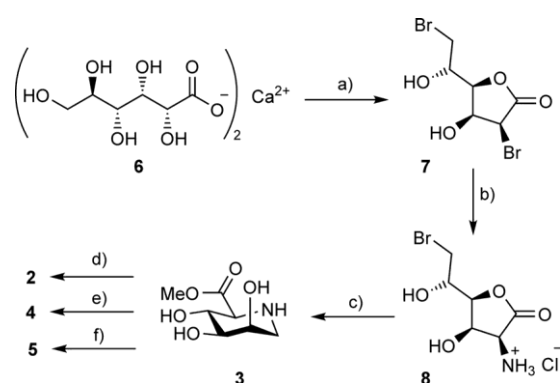
Carbohydrate-processing enzymes, such as glycoside hydrolases, may induce a chemical transformation by forcing the carbohydrate substrate into an unusual conformation.^[11] α -Mannosidases that belong to the CAZY family GH47 are inverting glycoside hydrolases that cleave α -1,2-mannosidic linkages. The mammalian GH47 mannosidases can be found in the Golgi and endoplasmic reticulum (ER), where they cleave mannose residues from N-glycans, thereby playing an important role in protein biosynthesis and quality control. The mechanism by which these hydrolases cleave the 1,2-mannosidic bonds is notable because they employ an unusual catalytic itinerary. The substrate that is to be cleaved binds in a ${}^3,0B/{}^3S_1$ conformation and is hydrolysed in a reaction that proceeds through a transition state in which the mannose ring adopts a 3H_4 conformation.^[12] Kifunesine (1, Scheme 1 B), a potent inhibitor of the mannosidase I enzyme, has been shown to adopt a ring-flipped 1C_4 conformation, and a similar conformation was found for 1-deoxymannojirimycin (DMJ, 2, Scheme 1 B) bound in the active site of *Saccharomyces cerevisiae*.^[12]

We were inspired by the conformations of the inhibitors of the GH47 enzymes to explore the behaviour of azasugars based on mannuronic acid. Here we report on the synthesis of mannuronic-acid-based azasugars 3, 4 and 5 (Scheme 1 C) and their conformational behaviour. We show that the stereoelectronic effects that determine the structures of the mannuronic acid oxocarbenium ions also impact the three-dimensional structures of these azasugars and that protonation of the ring nitrogen can induce a ring flip leading to an axial-rich 1C_4 conformation in solution. We build on this to show how deoxymannojirimycin, the “parent” compound, binds to a bacterial GH47 enzyme from *Caulobacter* sp. K31^[12] but also that - un-

fortunately, despite improved solution behaviour—the mannuronic acid derivatives do not bind to the GH47 enzyme, likely by virtue of their altered C5-substituents.

Results and Discussion

The synthesis of DMJ (2) and its C5 analogues was achieved according to the route devised by Wrodnigg and co-workers.^[13] As depicted in Scheme 2, methyl mannuronic acid ester azasugar 3 was obtained in four steps from the commercially available calcium D-gluconate monohydrate (6).^[14] The gluconate 6



Scheme 2. Synthesis of DMJ and its C5 analogues for this study. Reaction conditions: a) i: HBr, AcOH; ii: MeOH, 26 % over two steps; b) i: NaN_3 , acetone; ii: H_2 , Pd/C, HCl (aq.), MeOH, 55% over two steps; c) i: Et_3N , MeOH, quant.; ii: HMDS, $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ (cat.), CAN; iii: MeOH, AcCl (cat), 70 % over three steps. d) NaBH_4 , EtOH, 29 %; e) NaOH, H_2O , quant.; f) NH_3 , MeOH, quant.

was treated with HBr in acetic acid to form 3,5-di-O-acetyl-2,6-dibromo-2,6-dideoxy-D-manno-1,4-lactone after a series of acid-catalysed transformations (i.e., substitution of the C2 and C6 hydroxy groups, intramolecular ring closure and acetylation of the remaining hydroxy groups). Next, the acetyl groups at O3 and O5 were removed in an acid-catalysed transesterification with methanol to provide the pure dibromolactone 7 after crystallisation from chloroform/water, in 26% yield over the two steps. Regioselective displacement of the C2 bromide with an azide occurred with retention of configuration, as explained by Bock et al.^[15] with epimerisation of the C2-bromide to the more reactive glucose-configured dibromide and subsequent regioselective substitution by the azide. Thereafter, palladium-catalysed reduction of the intermediate azide and subsequent crystallisation from ethanol gave 2-amino-6-bromolactone 8 as its hydrochloric acid salt in 55 % yield. Treatment of this salt with triethylamine in methanol led to ring opening and intramolecular bromide displacement by the C2 amine to give crude azasugar methyl ester 3.

Purification of this compound from the triethylammonium and sodium salts formed in the reaction proved difficult, because of the high polarity of the compound as well as the lability of the methyl ester towards hydrolysis. Attempts to crystallise the compound were to no avail. Therefore, all of the hydroxy groups in 3 were capped with trimethylsilyl groups^[16] to allow for the purification of the compound by chromatogra-

phy. After desilylation, the pure methyl ester **3** was obtained as its hydrochloric acid salt.

DMJ (**2**) was synthesised from **3** by a sodium-borohydride-mediated reduction and was obtained in 29 % yield after column chromatography. D-Mannuronic acid azasugar **4** and amide **5** were obtained from **3** through saponification with sodium hydroxide or aminolysis with methanolic ammonia, respectively.

With the set of azasugars to hand we established their pK_a values by titration and investigated their conformational behaviour at different pH^* (the pH measured in D_2O) values by NMR spectroscopy. Table 1 summarises the results of these studies. For DMJ a pK_a value of 7.4 was measured, in line with the pK_a previously established for this compound (7.5).^[4c] The pK_a values of methyl ester **3**, amino acid **4** and amide **5** were determined to be 5.3, 7.5 and 5.8, respectively. The drop in pK_a value for the ester and the amide is a clear manifestation of the electron-withdrawing effect of the carboxylic acid ester and amide functionalities. At higher pH^* , at which acid **4** is deprotonated, the electron-withdrawing effect of the carboxylate is lowered because of its negative charge.

Table 1. pK_a values for compounds 2–5 and observed and calculated coupling constants and determined ${}^4C_1:{}_1C_4$ conformer ratio.						
Compound	pH^*	$J_{3,4}$ (obs.)	$J_{3,4}$ (calcd)	NMR ratio	pK_a	
			4C_1	${}_1C_4$	${}^4C_1:{}_1C_4$	
2	9	9.5	9.0	3.3	100:0	7.4
	2	9.5	9.4	4.3	100:0	
3	8	9.4	9.4	3.9	99:1	5.3
	2	7.5	9.5	4.9	56:44	
4	11	9.8	9.3	3.1	100:0	7.5
	5	8.8	9.1	3.8	94:6	
5	1	8.3	9.5	4.7	75:25	5.8
	9	9.7	9.0	2.8	100:0	
3	2	9.3	9.1	4.7	100:0	
	CD ₃ OD	9.2	9.3	4.2	99:1	
	CD ₃ OD (+ TFA)	4.8	9.3	4.7	1:99	

Figures 1–4 show the 1H NMR spectra of azasugars **2–5** recorded at varying pH^* values. In Figure 1, 1H NMR spectra of DMJ (**2**) in D_2O at pH^* 1–12 are collected. From pH^* 1 to pH^* 6.5 no changes are observed either in chemical shifts or in coupling constants. The coupling constants are indicative of a “normal” 4C_1 chair conformation for the azasugar ring. On-going from pH^* 6.5 to pH^* 12 significant shifts in chemical shift are observed for all ring protons, with the direct neighbours of the amino group experiencing the largest shifts. No changes in the coupling constants of the ring protons are observed, thus indicating that no major conformation change takes place.

In Figure 2, the 1H NMR spectra of methyl ester **3** at different pH^* values are displayed. Because hydrolysis of the methyl ester was observed above pH^* 8, no spectra were recorded above this pH^* . Large chemical shift changes are seen with increasing pH^* . Especially, H5 shows a large chemical shift change and shifts from $\delta = 4.04$ at pH^* 2 to 3.22 at pH^* 8. In addition, a change in coupling constants is observed for the ring protons. For example, $J_{3,4}$ changes from 9.4 Hz at basic pH^* to 7.5 Hz at acidic pH^* , indicative of a change in conformation of the azasugar ring. At high pH^* the azasugar adopts a single conformation, whereas both the 1C_4 and 4C_1 conformers are present at low pH^* (vide infra).

Mannuronic acid **4** can exist in three different charged states: the fully protonated state, the neutral zwitterionic state and the negatively charged state. Figure 3 shows 1H NMR spectra of **4** from pH^* 1 to pH^* 12. Again, large chemical shift changes are observed with changing pH^* (especially for H5, which shows a shift from $\delta = 3.9$ to $\delta = 2.9$ ppm). A small change in coupling constants is also apparent: $J_{3,4}$ changes from 9.8 Hz at high pH^* to 8.8 Hz at neutral pH^* to 8.3 Hz at acidic pH^* . Thus, in line with the conformational behaviour of methyl ester **3**, mannuronic acid **4** can change its conformation in a pH-dependent manner.

Figure 4 displays a set of 1H NMR spectra for amide **5** at different pH^* values. Smaller changes are observed for the chemi-

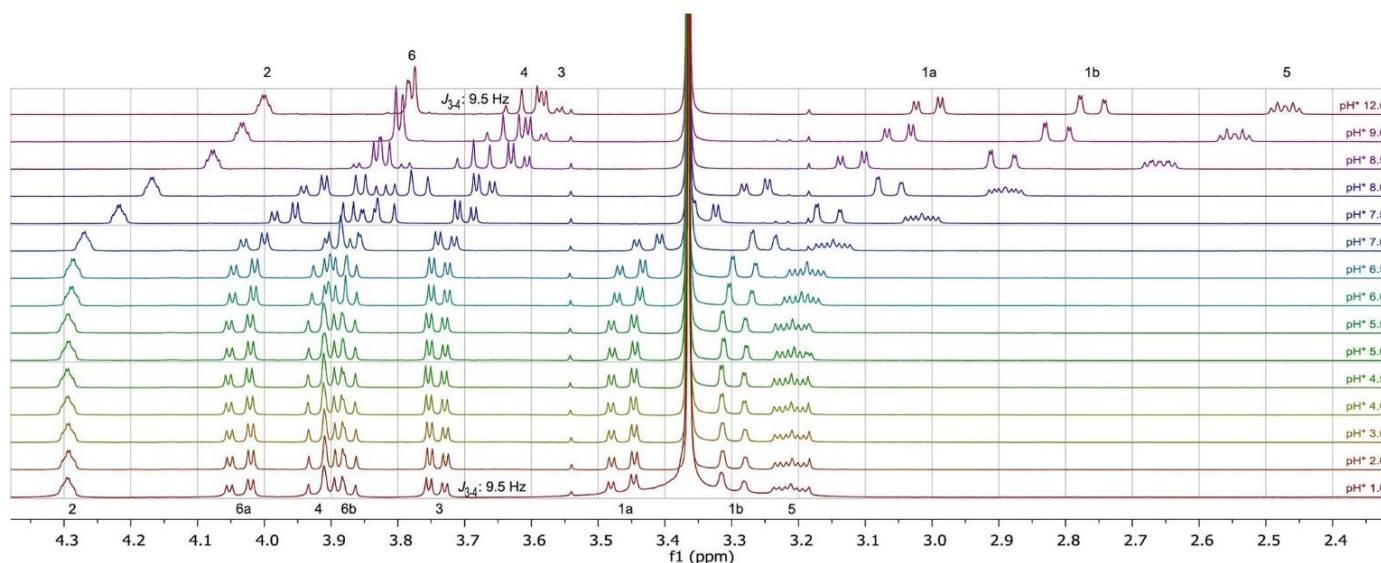


Figure 1. 400 MHz 1H NMR spectra for DMJ (**2**) at different pH^* values; spectra are referenced to residual methanol.

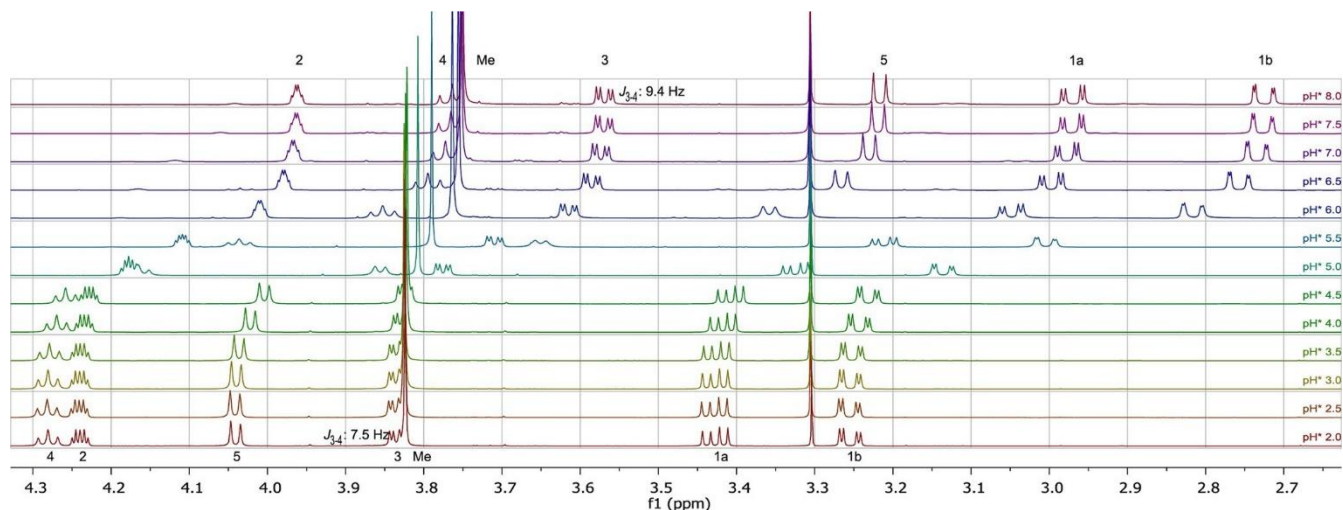


Figure 2. ^1H NMR spectra for 2,6-dideoxy-2,6-iminomannuronic acid methyl ester (**3**) at different pH^* values; spectra are referenced to residual methanol.

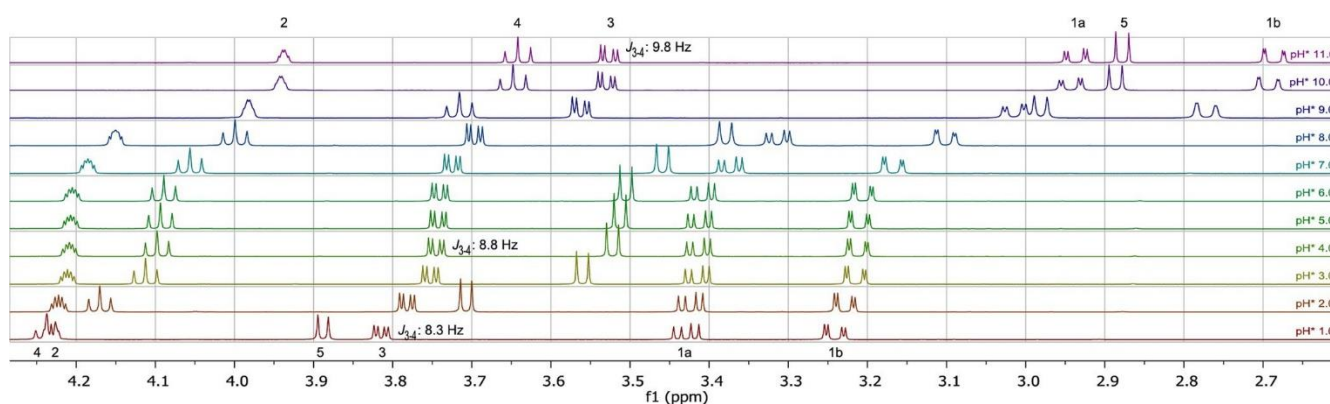


Figure 3. ^1H NMR spectra for 2,6-dideoxy-2,6-iminomannuronic acid (**4**) at different pH^* values; spectra are referenced to water.

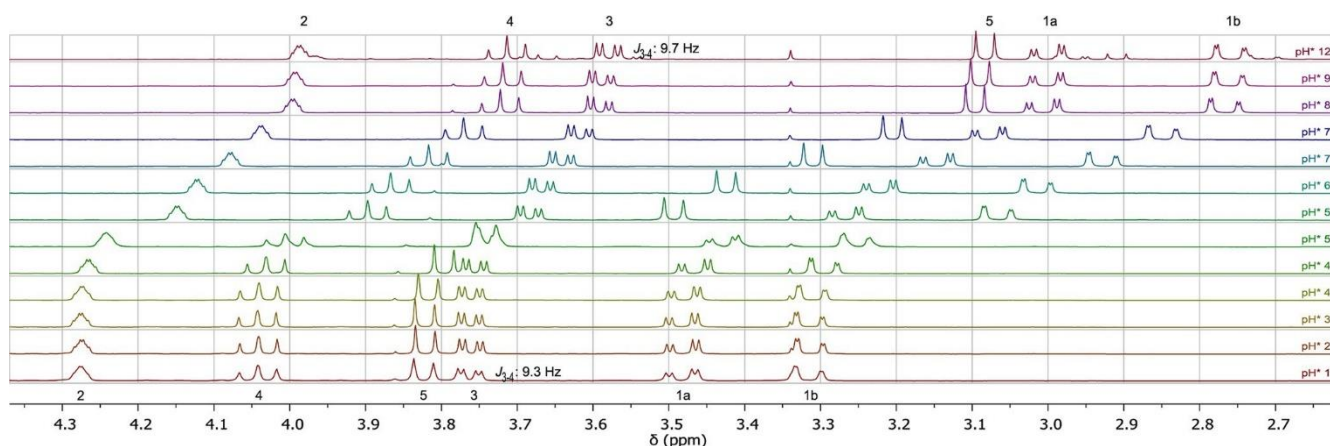


Figure 4. ^1H NMR spectra for 2,6-dideoxy-2,6-iminomannuronic amide (**5**) at different pH^* values; spectra are referenced to water.

cal shift change of H5 and there is no significant change in the coupling constants, thus indicating minimal conformation changes on going from high to low pH^* for this azasugar.

To establish the ratio of $^1\text{C}_4$ and $^4\text{C}_1$ conformers for the different azasugars we used DFT calculations to determine the coupling constants of the two conformers of both the protonated

and the deprotonated azasugars (for details see the Supporting Information).^[17] Table 1 shows the measured coupling constants ($J_{3,4}$) for the four azasugars at low and high pH^* values, the calculated $J_{3,4}$ values for the $^4\text{C}_1$ and $^1\text{C}_4$ conformers and the ratios of the two conformers, established from the measured average coupling constants. As can be seen from

Table 1, there is good agreement between the calculated and the measured coupling constants at high pH* values. With the two values for $J_{3,4}$ the ratios of the 1C_4 and 4C_1 conformers were established. It is clear that DMJ (**2**) exists in a single 4C_1 conformation at both low and high pH values. For the methyl ester **3** the situation is different. With the calculated values for the coupling constants of both conformers ($J_{3,4} = 9.5$ Hz and 4.9 Hz for the 1C_4 and 4C_1 azasugars, respectively) and the measured average coupling constant ($J_{3,4} = 7.5$ Hz) the ratio of the two conformers was established to be 56:44, thus indicating that the two chair conformers are equally stable. In similar vein, the ratio of the two chair conformers of the acid **4** was determined at three different pH values. As can be seen in Table 1, at high pH the anionic azasugar **4** exists as a single conformer, whereas at pH 5 the measured average coupling constant indicates a 94:6 mixture of conformers. At low pH the two conformers are observed in a 75:25 ${}^4C_1/{}^1C_4$ ratio. For the amide **5**, at both high and low pH the 4C_1 chair is almost exclusively present.

To investigate the conformational behaviour in a less polar environment, the azasugar showing the largest conformational change, methyl ester **3**, was investigated in CD_3OD . Figure 5 shows the spectra of the non-protonated and the protonated azasugar. In this medium the $J_{3,4}$ coupling constant changes from 9.2 Hz to 4.8 Hz upon protonation, thus indicating that the non-protonated azasugar exists in the 4C_1 conformation whereas the protonated species is found in the opposite 1C_4 conformation.

The NMR results show that DMJ analogues possessing a methyl ester or carboxylic acid at C5 (as in **3** and **4**, respectively) can change their conformation from the 4C_1 chair form to the opposite 1C_4 chair upon protonation. This conformational change is seen even in a highly polar medium such as water and is significantly enhanced in a more apolar solvent (CD_3OD). The nature of the substituent at C5 of the DMJ analogues is of major importance, because DMJ (**2**) and its C5 amide **5** do not display any conformational change with changing pH. The difference between the ester and amide is notable, because both functional groups—the C5 carboxylic acid ester and the C5 carboxamide—have a similar effect on

the basicity of the azasugars. The electron-withdrawing effect of both groups leads to a significant drop in the pK_a values for **3** and **5**, with the more strongly electron-withdrawing functionality—the ester—having the stronger inductive effect. The conformational flip of ester **3** and acid **4** can be accounted for by considering that electron-withdrawing groups prefer to occupy an axial position on a positively charged pyranose ring to minimise their destabilising effect.^[4–7] The fact that amide **5** does not change its conformation to accommodate this intrinsic preference might be due to internal hydrogen bonds that can be formed between the amide $-NH_2$ and the C4-OH, which provides an extra stabilising factor in the 4C_1 amide.^[18]

Having established that the mannuronic acid azasugars readily undergo ring flip upon protonation we probed the binding of the azasugars in the binding pocket of the α -1,2- mannosidase GH47, from the *Caulobacter* K31 strain. All four compounds were tested for binding through X-ray crystallography and isothermal titration calorimetry. Initially we analysed the binding of the parent compound DMJ (**2**).

DMJ (**2**) binds to CkGH47 with a K_D of 481 nm (determined by isothermal titration calorimetry, Figure 6 A). Although DMJ binding is essentially as observed previously for the mammalian GH47 structures,^[19] the subatomic resolution data (Supporting Information) in this case allow us to observe DMJ bound in the active site of CkGH47 in two different ring conformations (Figure 6 B). In the -1 subsite, the conformation of DMJ is in both 3S_1 and 1C_4 , each with a modelled occupancy of 0.5. Both conformations are consistent with the conformational itinerary of GH47 in which the structure adopts a 3S_1 conformation in the Michaelis complex to react via a 3H_4 transition state to form the product in a 1C_4 conformation.^[12,19,20] This dual conformation observation could perhaps be explained by the proximity of the pH of the crystallisation conditions (6.5) to the pK_a of DMJ (7.5) and the protonation of the species, although one cannot deconvolute which conformer relates to which protonation state).

The structure (PDB ID: 5MEH) confirms proposals made by others, and by us, concerning the catalytic apparatus.^[12,19,20] Briefly, catalytic base E365 is hydrogen-bonded to the O6 of

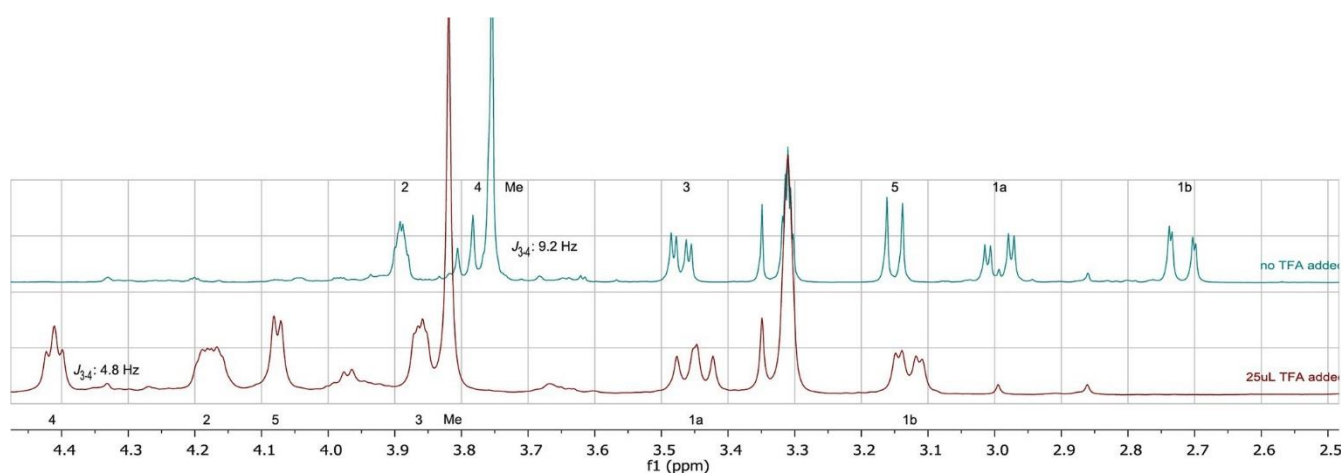


Figure 5. DMJ methyl ester (**3**) in CD_3OD (≈ 0.7 mL). The non-protonated azasugar is shown on top (no TFA added), the protonated azasugar on the bottom (25 μ L TFA added). Spectra are referenced to CD_2HOD .

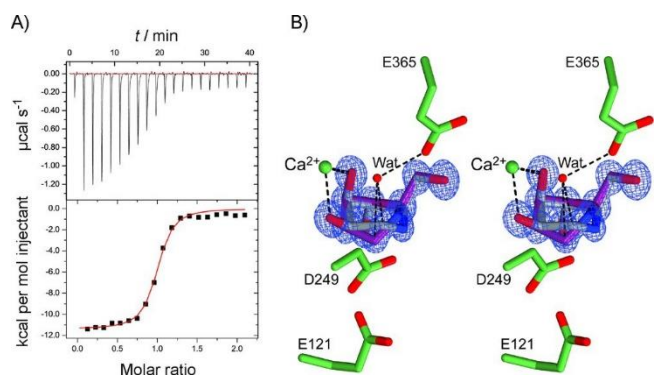


Figure 6. Binding of DMJ (**2**) to CkGH47. A) ITC-derived thermodynamics of binding. The stoichiometry (n) was 0.96 ± 0.01 sites. The association constant (K_A) was $(2.1 \times 10^6 \pm 3.3 \times 10^5) \text{ m}^{-1}$. The enthalpy change (ΔH) was $(1140 \pm 150) \text{ cal mol}^{-1}$. B) Divergent wall-eyed stereo electron density for the structure of CkGH47 in complex with DMJ (**2**), with the two conformations shown with grey/purple bonds). The map shown is a maximum-likelihood/ σ_A weighted $2F_{\text{obs}} - F_{\text{calcd}}$, contoured at $0.87 \text{ electrons } \text{\AA}^{-3}$. The active centre calcium is shown as a green sphere and a water (Wat)—likely equating to the nucleophilic water in catalysis—is shown as a red sphere. Key active centre residues discussed in the text are labelled.

DMJ, at 2.6 \AA . It is held in place by the nucleophilic water, which in turn is coordinated by calcium. The indirect route to protonic assistance is likely given by the O ϵ 2 of E121, although a role for D249 has also been considered. The riding hydrogens of this bond are visible, matching the level of detail of structure achieved by Thompson et al.^[12a] With the assistance of a metal ion, the O2-C2-C3-O3 torsion angle of a 4C_1 conformation is tightened from $\approx 60^\circ$ to $0\text{--}15^\circ$, consistent with the known conformational pathway of GH47 via a 3H_4 transition state.^[12,19,20]

Unfortunately, despite promising solution characteristics (Table 1), we were not able to detect binding of mannuronic acid derivatives **3**, **4** or **5**. Simple modelling of these compounds in the active centre, using the DMJ complex as a template, suggested that the likely reason would be steric clashes with residues in the active site, in particular E427. To test this hypothesis, a mutant containing an E427A mutation to increase the size of the active site was produced; however, attempts to obtain complexes with this variant still did not allow observation of **3-5** in the -1 subsite (data not shown), thus suggesting that further steric clashes might also be contributing to the lack of inhibition.

Conclusion

Azasugars based on mannuronic acid can change their conformation from a “normal” 4C_1 chair to the inverted 1C_4 chair form upon protonation of the endocyclic amine. The molecules thereby position their substituents such that they are optimally positioned to accommodate the positive charge. Although the conformational behaviour of other glycuronic-acid-based azasugars with different substituent configurations has not yet been studied in detail, it is likely that the spatial preferences of the substituents in the mannuronic acid azasugar work in concert to affect the ring flip. This behaviour is in line

with the conformational effects observed for fully protected mannuronic acid glycosyl donors, and therefore the results described here provide an extra indication that the positive charge at the anomeric centre of a mannuronic acid oxocarbenium ion is responsible for the observed unusual ring flip. This intriguing ring-flipping behaviour pointed to the potential use of the mannuronic acid azasugars as inhibitors for mannosidases that hydrolyse their substrates through a ring-flipped conformational itinerary. Unfortunately, the mannuronic acid azasugars did not bind to the studied GH47 mannosidase. Although the concept of chemically flipped inhibitors worked in solution, especially in low-polarity buffers, they sadly highlight the challenges of conformationally specific enzyme inhibition. For, whilst the introduction of favourable chemistry—including, in some cases, locking groups—frequently introduces substituents that prevent binding for steric reasons, in enzyme active centres that have evolved to harness the interactions of and thus distort unsubstituted sugars (e.g., the elegant locking of a mannoside mimic into $B_{2,5}$ conformation with a three-carbon bridge^[21]—in order to target $B_{2,5}$ transition-state mannosidases specifically) it simply resulted in steric clashes with the target β -mannosidase and no inhibition of the wild-type enzyme.^[22] Indeed, although the concept of conformation-specific targeted inhibition is one of the most exciting in glycochemistry, it is only rarely achieved: the use of ring-flipped kifunensine (**1**) to inhibit “southern hemisphere” mannosidases is one of the very few cases in which a conformationally restrained inhibitor works (and has indeed found considerable application in cell biology).^[23] The challenge therefore is still to provide the specific tools and therapeutic compounds required for cellular or patient use, whilst also maintaining binding to the target enzyme.

Experimental section

General methods for organic synthesis: All reagents were of commercial grade and used as received unless stated otherwise. Reactions were performed at room temperature unless stated otherwise. Molecular sieves (4 Å) were flame-dried before use. Flash column chromatography was performed on silica gel (40–63 μm). ${}^1\text{H}$ and ${}^{13}\text{C}$ NMR spectra were recorded with Bruker AV 600, Bruker AV 400 or Bruker DPX 400 spectrometers in D₂O or CD₃OD. Chemical shifts (δ) are given in ppm relative to the solvent residual signals. Coupling constants (J) are given in Hz. All given ${}^{13}\text{C}$ spectra are proton-decoupled. Compound names are given with use of the standard iminosugar nomenclature numbering.

2,6-Dibromo-2,6-dideoxy-d-mannono-1,4-lactone (7): Calcium D-gluconate monohydrate (6, 126 g, 280 mmol) was put under argon before being dissolved in 33 % HBr in acetic acid (500 mL, 3.0 mol). The reaction mixture was stirred for 18 h to give an acetylated form of **6**. MeOH (1 L) was added, and the mixture was heated at reflux for 2 h. It was then concentrated to half its original volume under reduced pressure before addition of more MeOH (500 mL). The reaction mixture was left to stir overnight, after which it was concentrated, resulting in a slightly oily residue. This was co-evaporated with MeOH (100 mL) and three times with H₂O (100 mL). The residue was extracted with diethyl ether (4 v 100 mL), and the organic layers were combined, dried with MgSO₄, filtered and concentrated under vacuum, yielding a yellow oily residue. This was

crystallised from $\text{CHCl}_3/\text{H}_2\text{O}$ to yield a white crystalline solid (44 g, 146 mmol, 26 % yield). M.p. 130 °C; $[\alpha]^{20} = +58.6^\circ$ ($c = 1$, MeOH); ^1H NMR (400 MHz, D_2O): $\delta = 5.20$ (d, $J = 4.5$ Hz, 1H; C-2), 4.64 (m, 2H; C-4, C-3), 4.19 (m, 1H; C-5), 3.77 (dd, $J = 11.4$, 2.4 Hz, 1H; C-6a), 3.65 ppm (dd, $J = 11.4$, 4.9 Hz, 1H; C-6b); ^{13}C NMR (101 MHz, D_2O): $\delta = 174.0$ (C-1), 81.6 (C-4), 69.1 (C-3), 66.2 (C-5), 47.6 (C-2), 36.6 ppm (C-6).

2-Amino-6-bromo-2,6-dideoxy-d-mannono-1,4-lactone hydrochloride (8): 2,6-Dibromo-2,6-dideoxy-d-mannono-1,4-lactone (**7**, 5.0 g, 16.5 mmol) was put under argon and dissolved in dry acetone (MgSO_4 , 100 mL). Sodium azide **Caution: highly toxic** (15.0 g, 231 mmol) was added, and the suspension was heated at reflux for 20 h. The mixture was filtered, and the filtrate was concentrated under reduced pressure. The residue was dissolved in H_2O (50 mL) and extracted with diethyl ether (5 x 100 mL), and the organic layers were combined, dried over MgSO_4 , filtered and concentrated under reduced pressure to give a brown oil that was identified as the 2-azido compound but included some of its diastereoisomer. ^1H NMR (400 MHz, D_2O): $\delta = 4.68$ (dd, $J = 4.5$, 3.3 Hz, 1H; C-3), 4.56 (d, $J = 4.6$ Hz, 1H; C-2), 4.46 (dd, $J = 9.2$, 2.7 Hz, 1H; C-4), 4.09 (m, 1H; C-5), 3.69 (dd, $J = 11.4$, 2.7 Hz, 1H; C-6a), 3.56 ppm (dd, $J = 11.5$, 4.9 Hz, 1H; C-6b); ^{13}C NMR (101 MHz, D_2O): $\delta = 174.1$ (C-1), 81.0 (C-4), 69.6, 65.7 (C-3, C-5), 62.3 (C-2), 36.6 ppm (C-6).

The crude compound (16.5 mmol) was put under argon and dissolved in MeOH (100 mL). Palladium on activated carbon (10 %, 300 mg, 0.3 mmol) and HCl (37 % in H_2O , 10 mL, 121 mmol) were added, and the suspension was put under hydrogen. The reaction mixture was stirred for 22 h, after which the catalyst was filtered off over a Whatman microfilter. The filtrate was concentrated under reduced pressure and co-evaporated once with HCl (37 % in H_2O , 60 mL), thrice with toluene (60 mL) and once with CHCl_3 (50 mL). Crystallisation from EtOH yielded 2-amino-6-bromo-2,6-dideoxy-d-mannono-1,4-lactone hydrochloride (**8**) as white crystals (2.6 g, 9.2 mmol, 55 % over two steps). M.p. 207 8C (decomp.); $[\alpha]^{20} = +41.68$ ($c = 1$, MeOH); ^1H NMR (400 MHz, D_2O): $\delta = 4.83$ (dd, $J = 4.8$, 2.8 Hz, 1H; C-3), 4.63 (dd, $J = 9.2$, 2.7 Hz, 1H; C-4), 4.59 (d, $J = 4.9$ Hz, 1H; C-2), 4.20 (m, 1H; C-5), 3.77 (dd, $J = 11.5$, 2.6 Hz, 1H; C-6a), 3.65 ppm (dd, $J = 11.4$, 5.0 Hz, 1H; C-6b); ^{13}C NMR (101 MHz, D_2O): $\delta = 172.1$ (C-1), 78.9 (C-4), 64.1 (C-3), 62.9 (C-5), 50.2 (C-2), 33.7 ppm (C-6).

1-Deoxymannojirimycin (2): 2-Amino-6-bromo-2,6-dideoxy-d-mannono-1,4-lactone hydrochloride (**8**, 501 mg, 1.8 mmol) was co-evaporated three times with dry toluene, put under argon and suspended in dry MeOH (10 mL). The suspension was cooled to 0 °C before addition of distilled triethylamine (1.0 mL, 7.2 mmol), and the resulting clear solution was stirred overnight. The reaction mixture was concentrated under reduced pressure, yielding the methyl ester as a white semicrystalline solid. The residue was put under argon, dissolved in dry EtOH (molsieves, 10 mL) and cooled to 0 °C. Sodium borohydride (709 mg, 19 mmol) was added, and the suspension was stirred overnight. Dry MeOH (20 mL) was added, after which the mixture was filtered, concentrated under reduced pressure and co-evaporated with HCl in MeOH (1 m, 3 x 10 mL). The residue was purified by column chromatography (EtOAc/EtOH 1:1→100 % EtOH), yielding a pure sample of DMJ (**2**) in 29 % yield (105 mg, 0.50 mmol). $[\alpha]^{20} = +14.08$ ($c = 0.5$, MeOH); ^{13}H NMR (399 MHz, D_2O): $\delta = 3.99$ (dt, $J = 2.9$, 1.6 Hz, 1H; C-3), 3.76 (dd, $J = 12.5$, 3.9 Hz, 1H; C-7), 3.71 (dd, $J = 12.5$, 5.5 Hz, 1H; C-7a), 3.60 (t, $J = 9.7$ Hz, 1H; C-5), 3.53 (dd, $J = 9.6$, 3.1 Hz, 1H; C-4), 3.03 (dd, $J = 14.2$, 2.8 Hz, 1H; C-2a), 2.80 (dd, $J = 14.2$, 1.5 Hz, 1H; C-2b), 2.57 ppm (ddd, $J = 9.7$, 4.9, 3.4 Hz, 1H; C-6); ^{13}C NMR (101 MHz, D_2O): $\delta = 74.4$ (C-4), 67.6 (C-5), 67.5 (C-3), 62.4 (C-6), 59.6 (C-7), 48.9 ppm (C-2).

Methyl 2,6-dideoxy-2,6-imino-d-mannonate hydrochloride (3): 2-Amino-6-bromo-2,6-dideoxy-d-mannono-1,4-lactone hydrochloride (**8**, 0.60 g, 2.2 mmol), was co-evaporated thrice with dry toluene, put under argon and suspended in dry MeOH (12 mL). The suspension was cooled to 0 °C before addition of distilled triethylamine (1.2 mL, 8.7 mmol), and the resulting clear solution was stirred overnight. The reaction mixture was concentrated under reduced pressure before being taken up in acetonitrile (15 mL) and charged with 1,1,1,3,3,3-hexamethyldisilazane (2.5 mL, 12 mmol) and copper sulfate pentahydrate (cat.). After 1 h, the reaction mixture was concentrated and a fraction of 234 mg (0.57 mmol) was purified by column chromatography (1–2.5 % 1,4-dioxane/ CH_2Cl_2) to give the per-TMSylated compound (162 mg, 0.40 mmol). The protected product was put under argon and dissolved in MeOH (8 mL), and acetyl chloride (1 equiv) was added to generate HCl in situ. The mixture was stirred for 0.5 h, after which the compound was concentrated and co-evaporated with MeOH to yield the title compound (98 mg, 0.40 mmol, 70 % over two steps). $[\alpha]^{20} = +31.8$ ($c = 1$, MeOH); ^1H NMR (400 MHz, D_2O): $\delta = 4.40$ (dd, $J = 5.3$, 4.8 Hz, 1H; C-5), 4.17 (ddd, $J = 9.5$, 4.1, 2.8 Hz, 1H; C-3), 4.09 (d, $J = 4.4$ Hz, 1H; C-6), 3.86 (dd, $J = 5.6$, 2.7 Hz, 1H; C-4), 3.82 (s, 3H; OCH₃), 3.44 (dd, $J = 12.2$, 9.6 Hz, 1H; C-2a), 3.13 ppm (dd, $J = 12.2$, 4.2 Hz, 1H; C-2b); ^{13}C NMR (101 MHz, D_2O): $\delta = 168.1$ (C-7), 71.1 (C-4), 69.9 (C-5), 64.2 (C-3), 58.8 (C-6), 53.4 (OCH₃), 43.5 ppm (C-2).

Sodium 2,6-dideoxy-2,6-imino-d-mannonate (4): Methyl 2,6-dideoxy-2,6-imino-d-mannonate hydrochloride (**3**, 24 mg, 0.10 mmol) was dissolved in H_2O (0.5 mL). A sodium hydroxide solution (1 m aq., 170 mL, 0.17 mmol) was added, and the mixture was stirred for 2 h. The mixture was concentrated under reduced pressure to yield the title compound, pure but with added sodium hydroxide. $[\alpha]^{20} = +7.2$ ($c = 1$, MeOH); ^1H NMR (400 MHz, D_2O): $\delta = 4.01$ (m, 1H; C-3), 3.71 (t, $J = 9.7$ Hz, 1H; C-5), 3.60 (dd, $J = 9.6$, 3.2 Hz, 1H; C-4), 3.0 (dd, $J = 14.6$, 2.7 Hz, 1H; C-2a), 2.95 (d, $J = 9.8$ Hz, 1H; C-6), 2.75 ppm (dd, $J = 14.6$, 1.6 Hz, 1H; C-2b); ^{13}C NMR (101 MHz, D_2O): $\delta = 178.4$ (C-7), 74.1 (C-4), 70.6 (C-5), 69.1 (C-3), 65.2 (C-6), 47.9 ppm (C-2).

2,6-Dideoxy-2,6-imino-d-mannonic amide (5): 2-Amino-6-bromo-2,6-dideoxy-d-mannono-1,4-lactone hydrochloride (**8**, 500 mg, 1.8 mmol), was co-evaporated thrice with dry toluene, put under argon and suspended in dry MeOH (10 mL). The suspension was cooled to 0 °C before addition of distilled triethylamine (1.0 mL, 7.2 mmol), and the resulting clear solution was stirred overnight. The reaction mixture was concentrated under reduced pressure, yielding the methyl ester as a white semicrystalline solid. The residue was dissolved in ammonia in MeOH (6 m, 10 mL, 60 mmol) and stirred overnight. The reaction mixture was concentrated under reduced pressure, affording 2,6-dideoxy-2,6-imino-d-mannonic amide (**5**) in quantitative yield. An analytical sample was prepared by crystallisation from pure MeOH (133 mg, 0.76 mmol, 42 %). $[\alpha]^{20} = +31.68$ ($c = 0.5$, H_2O); ^1H NMR (399 MHz, D_2O): $\delta = 3.97$ (m, 1H; C-3), 3.70 (t, $J = 9.7$ Hz, 1H; C-5), 3.57 (dd, $J = 9.6$, 3.1 Hz, 1H; C-4), 3.07 (d, $J = 9.8$ Hz, 1H; C-6), 2.99 (dd, $J = 14.6$, 2.7 Hz, 1H; C-2a), 2.75 ppm (dd, $J = 14.6$, 1.6 Hz, 1H; C-2b); ^{13}C NMR (101 MHz, DMSO): $\delta = 173.5$ (C-7), 74.9 (C-4), 70.0 (C-5), 68.7 (C-3), 63.3 (C-6), 49.1 ppm (C-2); HRMS: m/z calcd for $\text{C}_6\text{H}_{12}\text{O}_4\text{N}_2$: 177.08698 [$M+H$]⁺; found: 177.08683.

Acknowledgments

We thank the European Research Council (ERC-2011-AdG-290836 “Chembiosphing”, to H.S.O.). We thank Diamond Light Source for access to beamline I04-1 (proposal number mx-9948), which contributed to the results presented here. G.J.D. is the Royal Society Ken Murray Research Professor. A.M. is supported by a BBSRC PhD studentship (BB/M011151/1).

Conflict of interest

The authors declare no conflict of interest.

Keywords: azasugars · conformation analysis · inhibitor design · mannosidases · stereoelectronic effects

- [1] D. M. Smith, K. A. Woerpel, *Org. Biomol. Chem.* **2006**, *4*, 1195–1201.
- [2] a) B. Fraser-Reid, Z. F. Wu, U. E. Udodong, H. Ottosson, *J. Org. Chem.* **1990**, *55*, 6068–6070; b) D. R. Mootoo, P. Konradsson, U. Udodong, B. Fraser-Reid, *J. Am. Chem. Soc.* **1988**, *110*, 5583–5584; c) C. M. Pedersen, L. G. Marinescu, M. Bols, *Chem. Commun.* **2008**, 2465–2467.
- [3] a) W. G. Overend, J. S. Sequeira, C. W. Rees, *J. Chem. Soc.* **1962**, 3429–3439; b) H. Paulsen, *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 155–173; *Angew. Chem.* **1982**, *94*, 184–201; c) Z. Y. Zhang, I. R. Ollmann, X. S. Ye, R. Wischnat, T. Baasov, C. H. Wong, *J. Am. Chem. Soc.* **1999**, *121*, 734–753; d) T. K. Ritter, K. K. T. Mong, H. T. Liu, T. Nakatani, C. H. Wong, *Angew. Chem. Int. Ed.* **2003**, *42*, 4657–4660; *Angew. Chem.* **2003**, *115*, 4805–4808; e) C. McDonnell, O. Lopez, J. G. F. Bolanos, R. Hazell, M. Bols, *J. Am. Chem. Soc.* **2004**, *126*, 12374–12385.
- [4] a) H. H. Jensen, L. Lyngbye, M. Bols, *Angew. Chem. Int. Ed.* **2001**, *40*, 3447–3449; *Angew. Chem.* **2001**, *113*, 3555–3557; b) H. H. Jensen, L. Lyngbye, A. Jensen, M. Bols, *Chem. Eur. J.* **2002**, *8*, 1218–1226; c) H. H. Jensen, M. Bols, *Acc. Chem. Res.* **2006**, *39*, 259–265; d) M. Bols, X. F. Liang, H. H. Jensen, *J. Org. Chem.* **2002**, *67*, 8970–8974.
- [5] a) J. A. C. Romero, S. A. Tabacco, K. A. Woerpel, *J. Am. Chem. Soc.* **2000**, *122*, 168–169; b) L. Ayala, C. G. Lucero, J. A. C. Romero, S. A. Tabacco, K. A. Woerpel, *J. Am. Chem. Soc.* **2003**, *125*, 15521–15528; c) S. Chamberland, J. W. Ziller, K. A. Woerpel, *J. Am. Chem. Soc.* **2005**, *127*, 5322–5323.
- [6] a) C. H. Larsen, B. H. Ridgway, J. T. Shaw, K. A. Woerpel, *J. Am. Chem. Soc.* **1999**, *121*, 12208–12209; b) C. H. Larsen, B. H. Ridgway, J. T. Shaw, D. M. Smith, K. A. Woerpel, *J. Am. Chem. Soc.* **2005**, *127*, 10879–10884; c) C. G. Lucero, K. A. Woerpel, *J. Org. Chem.* **2006**, *71*, 2641–2647; d) M. T. Yang, K. A. Woerpel, *J. Org. Chem.* **2009**, *74*, 545–553.
- [7] E. R. van Rijssel, P. van Delft, G. Lodder, H. S. Overkleeft, G. A. van der Marel, D. V. Filippov, J. D. C. Codée, *Angew. Chem. Int. Ed.* **2014**, *53*, 10381–10385; *Angew. Chem.* **2014**, *126*, 10549–10553.
- [8] a) J. D. C. Codée, A. E. Christina, M. T. C. Walvoort, H. S. Overkleeft, G. A. van der Marel, *Top. Curr. Chem.* **2011**, *301*, 253–289; b) L. J. van den Bos, J. D. C. Codée, R. E. J. N. Litjens, J. Dinkelaar, H. S. Overkleeft, G. A. van der Marel, *Eur. J. Org. Chem.* **2007**, 3963–3976.
- [9] a) J. D. C. Codée, M. T. C. Walvoort, A. R. de Jong, G. Lodder, H. S. Overkleeft, G. A. van der Marel, *J. Carbohydr. Chem.* **2011**, *30*, 438–457; b) M. T. C. Walvoort, W. de Witte, J. van Dijk, J. Dinkelaar, G. Lodder, H. S. Overkleeft, J. D. C. Codée, G. A. van der Marel, *Org. Lett.* **2011**, *13*, 4360–4363; c) J. D. C. Codée, L. J. van den Bos, A. R. de Jong, J. Dinkelaar, G. Lodder, H. S. Overkleeft, G. A. van der Marel, *J. Org. Chem.* **2009**, *74*, 38–47.
- [10] a) M. T. C. Walvoort, G. Lodder, J. Mazurek, H. S. Overkleeft, J. D. C. Codée, G. A. van der Marel, *J. Am. Chem. Soc.* **2009**, *131*, 12080–12081; b) M. T. C. Walvoort, J. Dinkelaar, L. J. van den Bos, G. Lodder, H. S. Overkleeft, J. D. C. Codée, G. A. van der Marel, *Carbohydr. Res.* **2010**, *345*, 1252–1263; c) M. T. C. Walvoort, G. Lodder, H. S. Overkleeft, J. D. C. Codée, G. A. van der Marel, *J. Org. Chem.* **2010**, *75*, 7990–8002; d) S. van der Vorm, T. Hansen, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Sci.* **2017**, *8*, 1867–1875.
- [11] For reviews see: a) G. J. Davies, A. Planas, C. Rovira, *Acc. Chem. Res.* **2012**, *45*, 308–316; b) G. Speciale, A. Thompson, G. J. Davies, S. J. Williams, *Curr. Opin. Chem. Biol.* **2014**, *18*, 1–14.
- [12] a) A. J. Thompson, J. Dabin, J. Iglesias-Fernandez, A. Ardevol, Z. Dinev, S. J. Williams, O. Bande, A. Siriwardena, C. Moreland, T. C. Hu, D. K. Smith, H. J. Gilbert, C. Rovira, G. J. Davies, *Angew. Chem. Int. Ed.* **2012**, *51*, 10997–11001; *Angew. Chem.* **2012**, *124*, 11159–11163; b) R. J. Williams, J. Iglesias-Fernandez, J. Stepper, A. Jackson, A. J. Thompson, E. C. Lowe, J. M. White, H. J. Gilbert, C. Rovira, G. J. Davies, S. J. Williams, *Angew. Chem. Int. Ed.* **2014**, *53*, 1087–1091; *Angew. Chem.* **2014**, *126*, 1105–1109.
- [13] B. M. Malle, I. Lundt, T. M. Wrodnigg, *Org. Biomol. Chem.* **2008**, *6*, 1779–1786.
- [14] K. Bock, I. Lundt, C. Pedersen, *Carbohydr. Res.* **1979**, *68*, 313–319.
- [15] K. Bock, I. Lundt, C. Pedersen, *Acta Chem. Scand. Ser. B* **1987**, *41*, 435–441.
- [16] B. Akhlaghinia, S. Tavakoli, *Synthesis* **2005**, 1775–1777.
- [17] Gaussian 03 (Revision E.01), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.
- [18] It has previously been observed that mannuronic acid ester thio donors adopt a ¹C₄ conformation when fully protected, whereas they exist in a ⁴C₁ conformation when the C4-OH is free. This contrasting conformational behaviour can also be accounted for in terms of a stabilising hydrogen bond between the C5-carboxylic acid ester and the C4-OH in the ⁴C₁ chair conformation.
- [19] See, for example: F. Vallee, K. Karaveg, A. Herscovics, K. W. Moremen, P. L. Howell, *J. Biol. Chem.* **2000**, *275*, 41287–41298.
- [20] K. Karaveg, A. Siriwardena, W. Tempel, Z. J. Liu, J. Glushka, B. C. Wang, K. W. Moremen, *J. Biol. Chem.* **2005**, *280*, 16197–16207.
- [21] L. Amorim, D. Díaz, L. P. Calle-Jimenez, J. Jimenez-Barbero, P. Sinaÿ, Y. Blériot, *Tetrahedron Lett.* **2006**, *47*, 8887–8891.
- [22] L. E. Tailford, W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M. P. Heck, R. V. Stick, Y. Blériot, A. Vasella, H. J. Gilbert, G. J. Davies, *Nat. Chem. Biol.* **2008**, *4*, 306–312.
- [23] V. T. Chang, M. Crispin, A. R. Aricescu, D. J. Harvey, J. E. Nettleship, J. A. Fennelly, C. Yu, K. S. Boles, E. J. Evans, D. I. Stuart, R. A. Dwek, E. Y. Jones, R. J. Owens, S. J. Davis, *Structure* **2007**, *15*, 267–273.

Manuscript received: February 20, 2017

Accepted manuscript online: March 3, 2017

Version of record online: April 18, 2017

Additional material

3.3 Introduction

The additional information provided here will further examine the structurally unobservable binding of the mannuronic acid based iminosugars to *CkGH47*.

Hypothesised to be conformationally dependent on the pH to which it is exposed, DMJ (**2**) was observed to bind to the -1 subsite of *CkGH47* in two conformations, 3S_1 and 1C_4 . Experimentally determined in van Rijssel *et al.*, 2017 (table 1) and detailed in Jensen and Bols, 2006, the pKa of **2** (7.4) is close to the pH of the crystallisation condition of 6.5. Conformer ratios of $^4C_1 : ^1C_4$ calculated from NMR studies at different pHs, 2 and 9, showed a ratio of 100 : 0. In solution, it is energetically favourable for **2** and mannose to be in a 4C_1 conformation (Thompson *et al.*, 2012), unlike the ring flipped kifunensine, an α -mannosidase inhibitor (Males *et al.*, 2017). The preference of the mannuronic acid derivatives to be found in the ring flipped 1C_4 conformation occurs upon protonation of the endocyclic amine, experimentally due to an acidic pH. The pH of the endoplasmic reticulum is ~ 7.2 and the Golgi is slightly more acidic at ~ 6.4 (Maeda and Kinoshita, 2010; Schapiro and Grinstein, 2000; Wu *et al.*, 2000). This is a similar pH to that of the crystallisation condition, 6.5, replicating the natural environment. However, this may not be acidic enough to alter the ratio from $^4C_1 : ^1C_4$.

The active site is encased by a negative electrostatic potential, Figure 3.8. This environment may be problematic for a ligand that needs to be protonated due to residues which are most likely predisposed to accept a proton.

The conformational itinerary for GH47 has been shown by inhibitor complex structural determination to be $^3S_1 - ^3H_4 - ^1C_4$ (Karaveg *et al.*, 2005; Thompson *et al.*, 2012; Vallée *et al.*, 2000b). Xiang and colleagues propose a conformational itinerary of $^4C_1 \rightarrow ^1C_4 \rightarrow ^3S_1 \rightarrow [^3H_4]^\ddagger \rightarrow ^1C_4 \rightarrow ^4C_1$ whereby α -1,2-linked mannose and β -mannose are found as 4C_1 in solution before and after catalysis, (Xiang *et al.*, 2016), Figure 3.4c. The dual conformation observed for **2** in complex with *CkGH47* could either be capturing the progression from a ring-flipped solution conformation to Michaelis complex or initial transition from the Michaelis-complex to transient intermediate, which would be short-lived.

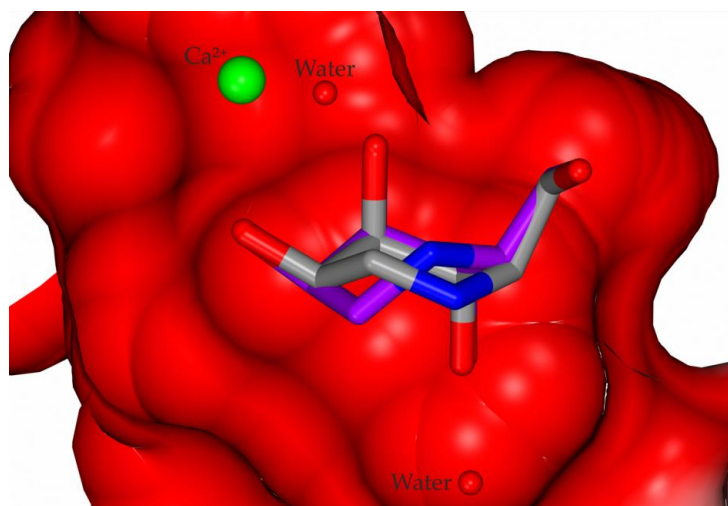


Figure 3.8. Electronegative potential of the neighbouring residues in the active site surrounding 2. Red active site indicates a negative electrostatic potential for 2.

Docking of 3, 4 and 5 into the active site of the structure solved with 2, enabled potential residues that could cause steric hindrance and block the binding of the inhibitor to be identified. One such residue was Glu427 where O ϵ 2 was 2.60 Å away from O3 of 2. Other potential residues that might narrow the active site were Arg363 (NH1 – C7 with a distance of 2.09 Å) and Tyr247 (C δ 2 - O1 with a distance of 2.19 Å), Figure 3.9. The additional work in this chapter will detail the binding of 3 to the active site of *CkGH47*, Glu427 of was mutated to an alanine (*CkGH47*_{E427A}), under different crystallisation conditions. In soaking experiments, Bis-Tris would have to be competitively forced out of the active site by a compound that has a higher affinity 2>Bis-Tris>3/4/5. Therefore, the challenge was to identify a crystallisation condition without Bis-Tris as a buffer and to maintain the sub-Angstrom resolution crystal form. When a structure was obtained, 3 was located in the catalytically irrelevant +1 subsite. Upon further study of the affinity of the mannuronic-acid derivatives for *CkGH47*, binding affinities were undetectable.

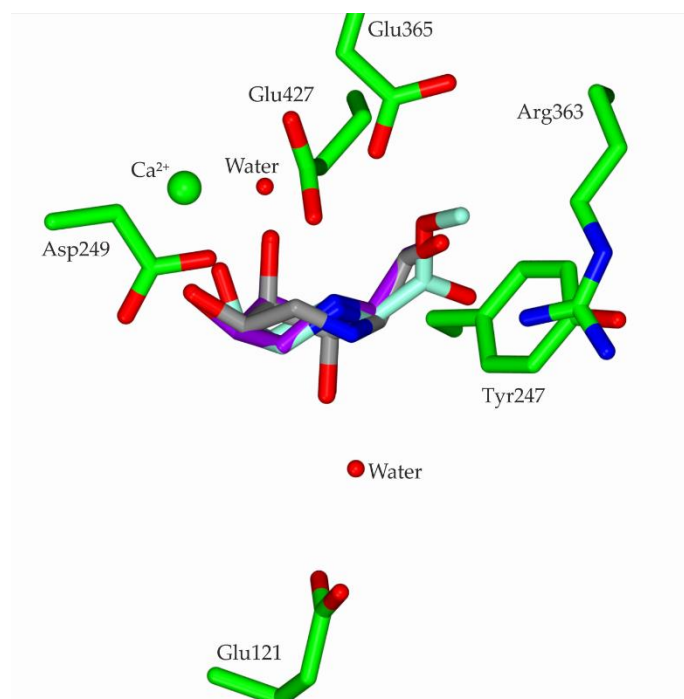


Figure 3.9. Steric hindrance of residues in the active site prevents binding. 2 shown in purple and grey in two conformations. **3** was docked into the active site using superposition of atoms C2, C4, C5, C6, N5, O6 against **2**. Catalytic residues coordinated by two waters and a calcium ion and predicted clashing residues are shown in green.

3.4 Materials and Methods

3.4.1 Expression and purification of *CkGH47*_{E427A}

Using a Q5 site directed mutagenesis kit (New England Biolabs) *CkGH47*, Glu427 of *CkGH47* was mutated to an alanine (*CkGH47*_{E427A}). The primers are described in the appendix Table 1.2. NEB5 α competent *Escherichia coli* (*E. coli*) cells were transformed with the PCR product and colonies containing the plasmid were selected for against 35 $\mu\text{g mL}^{-1}$ ampicillin. The DNA was extracted and sequenced to verify the mutation. Tuner (DE3) *E. coli* cells were transformed with *CkGH47*_{E427A} DNA. The expression and purification of *CkGH47*_{E427A} was conducted under the same conditions as the wild type (WT) protein discussed in Thompson et al., 2012.

3.4.2 Crystallisation screening of *CkGH47*_{E427A} and soaking experiments with mannuronic acid derivatives

The conditions of INDEX well G8 (0.2 M ammonium acetate, 0.1 M HEPES pH 7.5 and 25% PEG 3350) were optimised using the sitting drop vapour diffusion method in 48-well MRC MAXI optimisation plates (Molecular Dimensions). In the reservoir, 100 μ L of crystallisation solution was used and 500 nL of protein and 500 nL reservoir in the well. The conditions were varied by 0.1 M-0.2 M ammonium acetate, 0.1 M HEPES pH 7.0-8.0, 22%-29% PEG 3350. To further enhance protein-ligand interaction, the co-crystals were soaked in the appropriate ligand before flash freezing in liquid nitrogen; co-crystals of **3** from the 48-well optimisation plate were soaked with 10 mM of **3**.

The crystals were tested using the in-house Rigaku MicroMax 007HF x-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using an Oxford Cryosystem 700 Series Cryostream. Crystals that diffracted were sent to Diamond I02 MX beamline. The diffraction images were integrated using XIA2 (Winter, 2010) and reintegrated using Aimless (Evans, 2006, 2011) from the CCP4 software suite (Winn et al., 2011). Multiple rounds of refinement using REFMAC (Murshudov et al., 1997, 2011; Nicholls et al., 2012; Pannu et al., 1998; Vagin et al., 2004; Winn et al., 2003) and model building in COOT (Emsley et al., 2010) were used to refine the model. Figures of the structure were produced using CCP4mg (McNicholas et al., 2011).

3.4.3 Observation of the change in thermal denaturation

19.8 μ M of protein was added to a 96-well PCR plate with 5x SYPRO Orange dye and 500 μ M of inhibitor in a 25 μ L total reaction volume. The plate was incubated at room temperature (RT) for 1 hour prior to data collection. The temperature increased in intervals from 24 $^{\circ}$ C to 94 $^{\circ}$ C over 70 cycles. R(multicomponent view) data was processed (Bond, 2017) using a five parameter sigmoid equation (Schulz et al., 2013) to obtain a melting temperature (T_m) and then the data was processed to derive the fluorescence as a percentage.

3.5 Results

3.5.1 Crystallisation of *CkGH47*_{E427A} in HEPES buffer as an alternative to Bis-Tris buffer reduced the likelihood of unwanted ligands in the active site

It was hypothesised that the reason for the inability of mannuronic acid derivatives to bind to the active site of *CkGH47* was that the dimensions of the active site were not large enough to accommodate the R group. To test this hypothesis, the active site was artificially enlarged by replacement of Glu427 with a shorter, non-polar alanine residue, *CkGH47*_{E427A}. Co-crystallising and soaking techniques were used to produce diffracting crystals. To remove Bis-Tris from the active site, HEPES was found to be a suitable replacement at pH 7.5. The final crystallisation condition was 0.2 M ammonium acetate, 0.1 M HEPES pH 7.5 and 26% PEG 3350.

3.5.2 Crystal structure of *CkGH47*_{E427A} in complex with **3** showed binding in the reducing +1 subsite

A crystal structure was only attainable with compound **3** possibly due to the polarity of the derivatives **4** and **5**, Table 3.1. However, it was not bound in the catalytically relevant -1 subsite and was, instead bound in the +1 subsite, Figure 3.10a. Negative electron density around C7 indicates the last carbon may have been hydrolysed. The C6 moiety was flipped compared to the docking trials, and the occupancy of the atoms C6, C7 and O1 were reduced to 0.5. It was in a ⁴C₁ conformation, following expectations from the +1 subsite. The C2 and C3 hydroxyls were coordinated by O δ 1 and O δ 2 of Asp249, the proposed catalytic acid, with a distance of 2.62 Å and 2.78 Å, respectively. When overlaid with the structure of *CkGH47* in complex with thiomannobiose (PDB ID: 4AYP), rmsd of 0.27 Å, **3** is in a similar position to the mannose group in the +1 site, Figure 3.10b. The hydroxyl groups of C23 and C24 also coordinated to O δ 1 and O δ 2 of Asp249 with a similar distance of 2.63 Å and 2.74 Å, respectively. When overlaid with the structure of DMJ (PDB ID: 5MEH), rmsd of 0.3924, it is in the same position as the glycerol, Figure 3.10c. The methyl group protrudes from the active site. To accommodate for this, Tyr247, has moved 1.4 ° away from the active site.

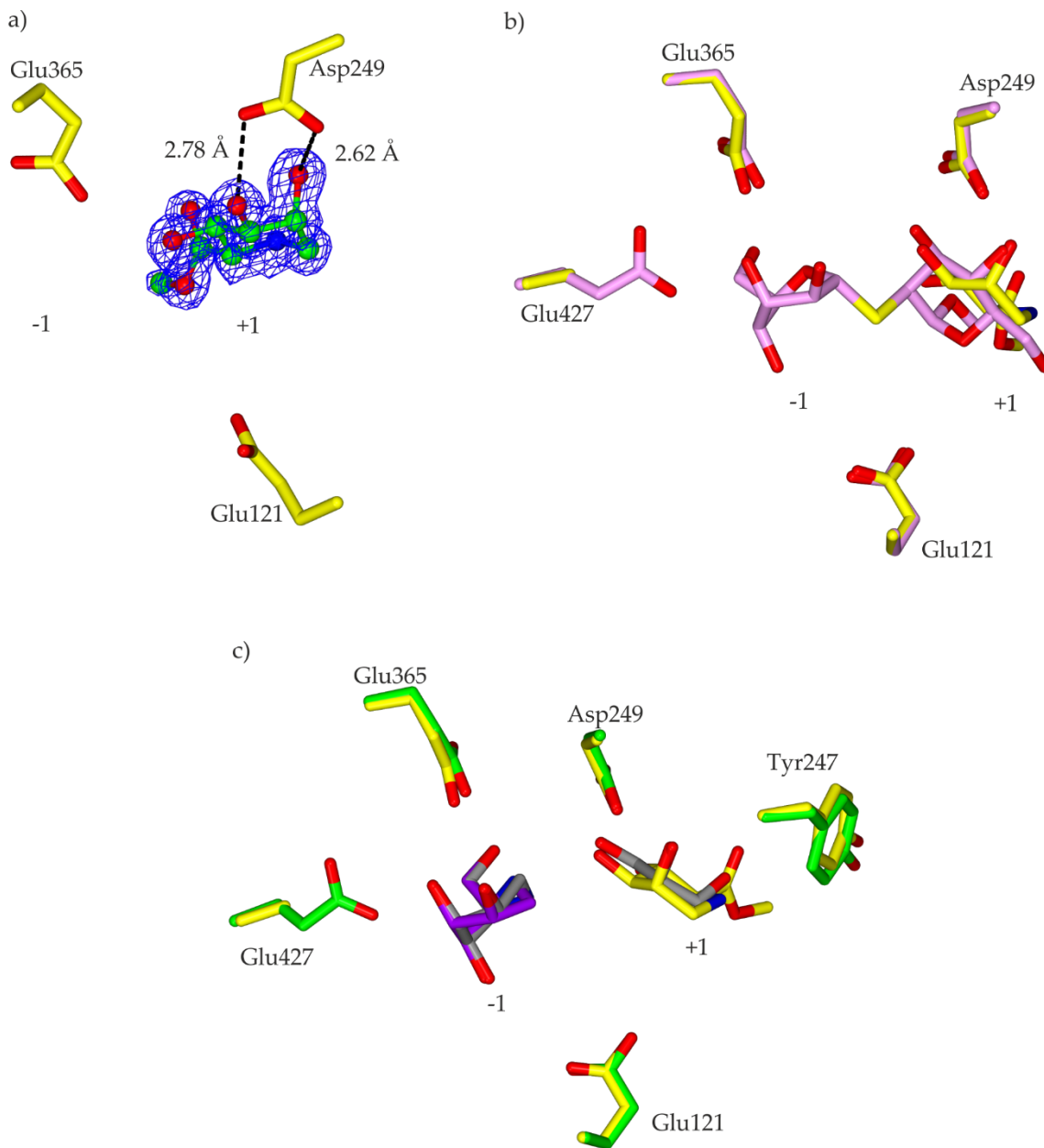


Figure 3.10. Hydrogen bonding interactions of 3 in complex with *CkGH47E427A*. Comparison of +1 subsite bound 3 with published inhibitor structures. Residues and 3 are represented in yellow. a) 3 forms hydrogen bonds with Asp249. The maximum-likelihood/ σ_A -weighted $2F_{\text{obs}}-F_{\text{calc}}$ map, shown in blue, was contoured at 0.9 electrons/ \AA^3 . b) Superposition of main chain protein of 4AYP and *CkGH47E427A*. Residues and thiomannobiose are coloured pink. Overlapping inhibitors show the definition between the -1 and +1 subsite, which may have flexibility. c) Superposition of main chain protein of 5MEH and *CkGH47E427A*. The colour scheme for the residues and DMJ of 5MEH are consistent with Figure 3.8, glycerol in the +1 subsite is coloured in grey.

Table 3.1. Data and refinement statistics for CkGH47 in complex with 3. *Values in parentheses are for high resolution data.

CkGH47_3	
Data collection	
Space group	H3
Cell dimensions:	
<i>a, b, c</i> (Å)	144.3, 144.3, 50.2
γ (°)	120.0
Resolution (Å)	41.65-1.10 (1.12-1.10)*
<i>R</i> _{merge}	0.06 (0.13)
<i>CC</i> (1/2)	1.00 (0.98)
<i>I</i> / σ <i>I</i>	20.1 (11.2)
Refinement	
<i>R</i> _{work} / <i>R</i> _{free}	0.13/0.14
B-factors (Å ²)	
Protein	8
Bond lengths (Å)	0.02
Bond angles (°)	1.97

3.5.3 The binding affinity of mannuronic acid derivatives to CkGH47 was too weak to observe by ITC and TSA

To further investigate the unobtainable crystal structures with the mannuronic acid derivatives in the -1 subsite, the binding affinity was assessed through protein stabilisation and through thermodynamic experiments. Initial thermal shift assays showed a stabilisation of melting temperature when **2** was added to CkGH47, Figure 3.11a.

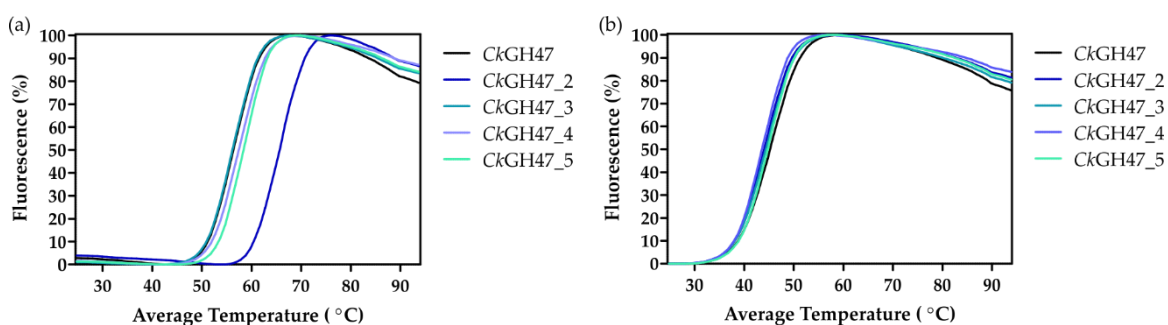


Figure 3.11. Fluorescence curves show no shift in melting temperature upon addition of mannuronic acid derivatives. a) The melting temperature of CkGH47 was tested in the purification buffer, Tris. A significant shift of 9 °C for **2** and weaker ΔT_m of 1.9 °C for **5**. b) In phosphate buffered saline solution, there were no significant shifts for any compound.

The protein was also dialysed into PBS buffer to eliminate the competition of the ligand and Tris for the active site. However, the protein was less stable by -11.3 °C and no significant change in melting temperature was observed with either **2** or **3**, **4**, and **5**, Figure 3.11b.

3.6 Discussion

Conformationally restricted compounds have been shown to have tight binding affinity to a specific protein if the conformation lies along the reaction pathway of the enzyme. Synthesis of compounds that are specific to GH47 was conducted by manipulation of a well characterised inhibitor DMJ. The different substituents at the C6 position shifted the equilibrium in solution from 4C_1 to 1C_4 at different pHs.

Through screening of the buffer used in the crystallisation condition, the buffer was changed from Bis-Tris pH 6.5 to HEPES pH 7.5 as it was speculated that Bis-Tris could have been outcompeting mannuronic acid derivatives and blocking the active site of GH47. pH driven conformers were indicated by NMR and coupling constant analysis, where a single conformer 4C_1 shifted to a ratio of 56:44, ${}^4C_1:1C_4$. Crystals were not obtainable in a buffer system with a low pH.

Residue Glu427, located in the active site, was mutated to widen the pocket allowing for the extended C6 moiety. The crystal complex with **3** did not inform elucidation of the catalytic mechanism since it bound in the +1 subsite. **3** is the least polar derivative, however, it was still not suitable to bind in a negatively charged active site.

In conclusion, the inability to obtain a crystal structure of *Ck*GH47 with a mannuronic-acid based derivative could be because of two reasons i) the conformational dependency on pH, or ii) steric hindrance and polarity of the active site. The challenge to improve the specificity and affinity of existing inhibitors for a target protein is ongoing and in the case of *Ck*GH47, finding a compound that satisfies the criteria for i) and ii).

SECTION 2: O-Glycan Processing Enzymes

i. Uridine diphosphate *N*-acetylglucosamine

Uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) is an acetylated aminosugar nucleotide that acts as a coenzyme in metabolism, Figure 2i. The second carbohydrate of focus in this thesis, GlcNAc, can be attached to serine and threonine residues via a β -linkage resulting in O-linked β -*N*-acetyl-D-glucosamine (Figure 2i). It was discovered as a single O-linked GlcNAc moiety by Torres and Hart in 1984, when the surface of murine thymocytes, splenic B- and T-lymphocytes and macrophages of tumour cells were probed with a galactosyltransferase that attached UDP-[3 H]galactose (radiolabelled substrate) in a β -1,4-linkage to GlcNAc (Torres and Hart, 1984).

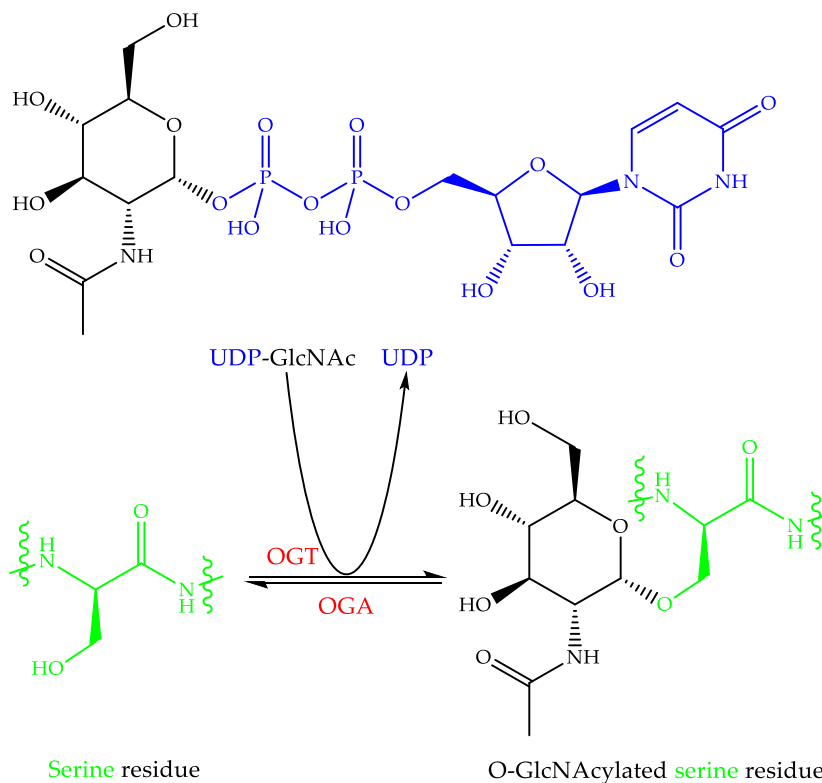


Figure 2.i. O-GlcNAcylation of a serine residue. The addition of O-GlcNAc is conducted by O-GlcNAc transferase (OGT) using the coenzyme UDP-GlcNAc. The removal of O-GlcNAc from a serine residue is catalysed by O-GlcNAc hydrolase (OGA).

There is between 2-30 μ M of UDP-GlcNAc in the cytoplasm, nucleus and mitochondria that forms a “privileged pool” for O-GlcNAc cycling (reviewed by Bond and Hanover, 2015). The permeability of an organelle to UDP-GlcNAc is related to the abundance of O-

GlcNAc or complex glycans (Bond and Hanover, 2015). Permeable organelles, for example the nucleus and mitochondria, have a higher abundance of O-GlcNAc and impermeable organelles, for example the cell membrane, lysosomes, peroxisome and endosomes, have a higher abundance of complex glycans. The uncharged acetylated hexosamine sugar is produced by *de novo* synthesis through the hexosamine biosynthetic pathway.

ii. Hexosamine biosynthetic pathway

Around 2-3% of cellular glucose enters the hexosamine biosynthetic pathway (HBP) as the starting substrate (Marshall et al., 1991). Firstly, glucose is phosphorylated to glucose-6-phosphate by glucokinase and then undergoes isomerisation by glucose-6-phosphate isomerase to fructose-6-phosphate, Figure 2.ii (Aguilera and Zimmermann, 1986; Stachelek et al., 1986). Fructose-6-P is the first compound in the HBP. The reversible formation of glucosamine-6-phosphate from fructose-6-phosphate is catalysed by glutamine:fructose-6-phosphate-amidotransferase (GFAT) (Watzel and Tanner, 1989). Glutamine is transported through glucose transporters across the membrane to be utilised in this step. It is the rate limiting step as it is sensitive to inhibition by the final product UDP-GlcNAc (Marshall et al., 1991). To bypass the rate limiting GFAT catalysis, glucosamine, which also enters the HBP via glucose transporters, is phosphorylated by hexokinase. Next, an acetyl group is added by glucosamine-6-phosphate acetyltransferase to form *N*-acetylglucosamine 6-phosphate (Boehmelt et al., 2000). Penultimately, it is isomerised by phosphoglucomutase 3 to produce GlcNAc-1-phosphate. Lastly, the reaction between UTP and GlcNAc-1-P is catalysed by UDP-GlcNAc pyrophosphorylase resulting in the final product of the hexosamine biosynthetic pathway, UDP-GlcNAc (Mio et al., 1998). This enzyme is a part of UDP-*N*-acetylhexosamine pyrophosphorylases that also catalyse the reaction that generates UDP-GalNAc from GalNAc-1-P. Specificity of UDP-GalNAc pyrophosphorylase can be skewed towards the synthesis of UDP-GlcNAc by the insertion of 17 amino acids (residues 454-496) into the protein sequence (Wang-Gillam et al., 1998).

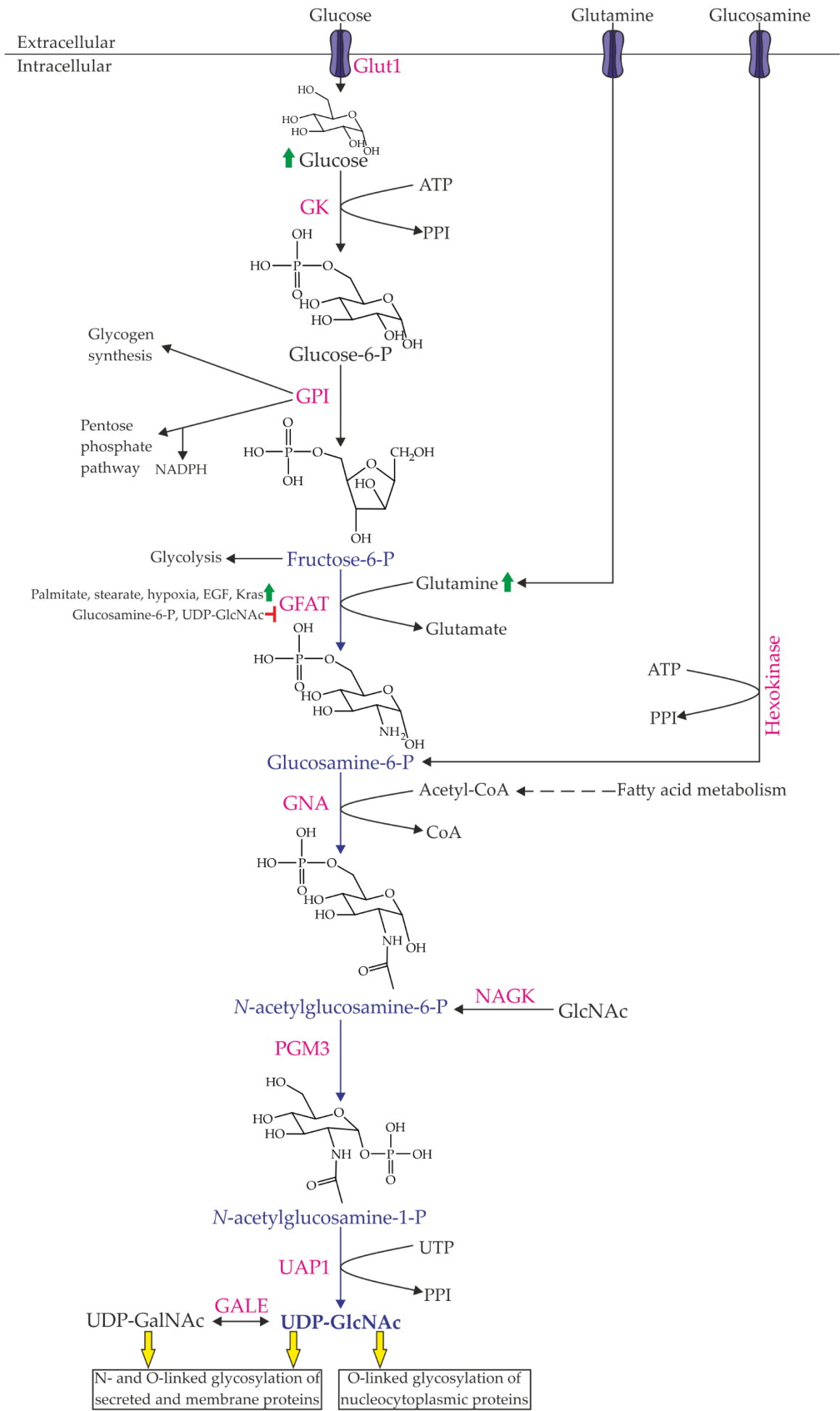


Figure 2.ii. The hexosamine biosynthetic pathway. The enzymes involved in the pathway are coloured in pink. The compounds in blue represent the HBP starting with fructose-6-P. The green arrows indicate components that, when in abundance, enhance the HBP. The red block indicates components that inhibit the HBP.

UDP-GlcNAc can also be produced by two alternative salvage pathways. In the first route, GlcNAc is derived from the degradation of the metabolic components and from lysosomal degradation of glycans (Figure 1.3 and Figure 1.iv) (Hinderlich et al., 1998). The other route uses galactosamine, which is acetylated to generate *N*-acetylgalactosamine 1-phosphate, and reacted with UTP to form UDP-*N*-acetylgalactosamine (Cunneen et al., 2013).

iii. Fluctuation between GlcNAc and glucose levels

Glucose homeostasis and metabolism is regulated by the O-GlcNAc cycling enzymes (Ruan et al., 2013); excess glucose and glutamine increase the levels of UDP-GlcNAc, hence, increased *Ogt* expression and O-GlcNAcylation of proteins (Abdel Rahman et al., 2013). Moreover, linked to increased glucose and glutamine metabolism (biosynthesis of glycoconjugates) is the progression of cancer due to the metabolic contribution of carbon and nitrogen atoms necessary for anabolic pathways (Warburg effect) (Vander Heiden et al., 2009). For example, GLUT1 glucose transporters are upregulated in tumour cells with mutations in the *K-Ras* oncogene (Ying et al., 2012).

CHAPTER 4: Family GH84 O-GlcNAc Hydrolase

4.1 Abstract

The O-GlcNAc post-translational modification (PTM) to serine and threonine residues is controlled by two enzymes, O-GlcNAc transferase and O-GlcNAc hydrolase. Thousands of nucleocytoplasmic proteins are GlcNAcylated by a long isoform, which contains an N-terminal catalytic domain, a stalk domain and a C-terminal domain with putative acetyltransferase activity. To understand how O-GlcNAc hydrolase modifies proteins, past work by others has visualised OGA in complex with glycosylated peptides. Without recognising a strict consensus sequence or the orientation of the protein, the substrate “specificity” of O-GlcNAc hydrolase is unresolved.

In this chapter, the peptide substrate binding mechanism will be explored. Using *HsOGA*:peptide soaking, surface entropy reduction, mutant *HsOGA* crystal structures revealed new details, albeit with mixed results. Although disappointing, this work did constitute a large portion of the PhD. To improve the crystal structure, surface entropy reduction was employed by the mutation of Glu602 and Glu605 to alanine. This resulted in a new crystal form where, when crystallised, only one active site of the dimer was blocked by the C-terminus; ultimately this construct could not be used for ligand binding. Lastly, I show that casein kinase II phosphorylates Ser364 of *HsOGA*. However, a S364D mimic of this phosphorylation moiety did not affect the secondary structure or catalytic efficiency of OGA.

4.2 Family GH84 structure and function introduction

4.2.1 O-GlcNAc cycling enzymes

O-GlcNAc is an unusual and dynamic post-translational modification (PTM) as it is not extended or modified once attached to the protein unlike N-GlcNAc. It requires a cycle of two proteins, O-GlcNAc transferase (OGT), which recognises the conserved sequence [PAT][VAT][ST][VRL][ASPTGY], and O-GlcNAc hydrolase (OGA), to install and remove the modification, respectively, Figure 4.1 (Dong and Hart, 1994; Hart et al., 2007; Holt and Hart, 1986; Kreppel et al., 1997; Lubas et al., 1997; Torres and Hart, 1984). OGT will not be covered in this thesis; for detailed reviews see Joiner et al., 2019; Martinez-Fleites et al., 2010; Nagel and Ball, 2014; Vocadlo, 2012. Over 4000 target proteins, 10-20% of all

functional proteins, have been identified that are localised both to the cytoplasm and the nucleus, see review in appendix pg.266. O-GlcNAcylation sites are often sites targeted for phosphorylation resulting in a reciprocal relationship, Figure 4.1. There are over 430 Ser/Thr kinases that have been identified and ~40 phosphatases encoded in the human genome (Ruan et al., 2013), in contrast to two enzymes for O-GlcNAcylation.

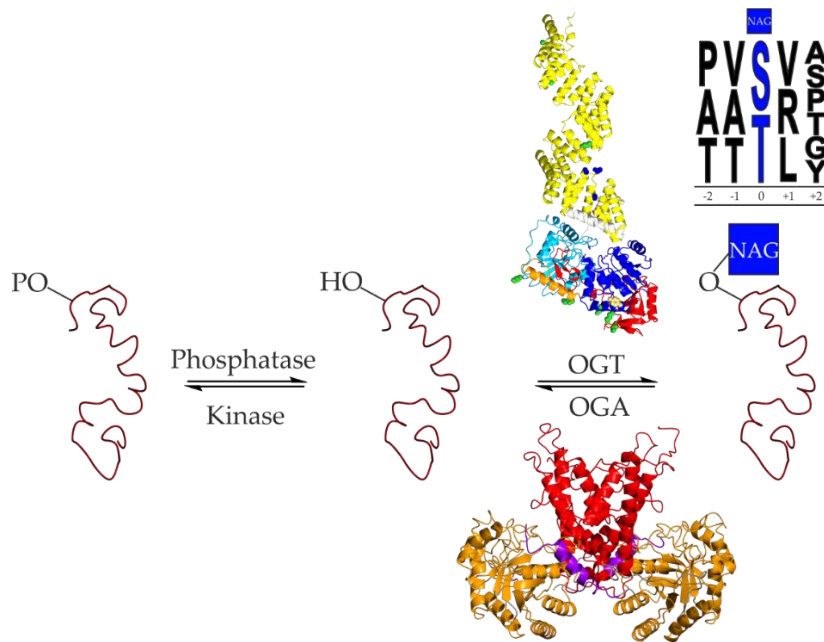


Figure 4.1. Reciprocal relationship between phosphorylation and GlcNAcylation. Target protein Tau (residues 267-312 shown in red PDB ID: 2MZ7) can be phosphorylated and O-GlcNAcylated at Ser400 following the conserved [PAT][VAT][ST][VRL][ASPTGY] glycosylation sequence. O-GlcNAc transferase (OGT) model was built using structures of *HsOGT_{TPR1.5}* (PDB ID: 1W3B) and *HsOGT_{4.5}* (PDB ID: 3PE3) by CCP4mg superposition. O-GlcNAc hydrolase (OGA) (PDB ID: 6HKI) is coloured by domain, orange is the catalytic domain and red is the stalk domain.

4.2.2 Discovery of O-GlcNAc hydrolase

Terminal GlcNAc monosaccharides were first observed attached to the cell surface and cytoplasmic proteins of lymphocytes (Torres and Hart, 1984). They could not be cleaved by peptide:N-glycosidase, which cleaves between GlcNAc and asparagine residues, indicating that GlcNAc attaches to serine or threonine residues. An enzyme with hydrolysing capability was identified later in splenic T cells (Dong and Hart, 1994). After purification, it was shown to be active on O-GlcNAc modified peptides. The gene

annotated as *Meningioma expressed antigen 5 (Mgea5)* encodes the protein O-glycoprotein 2-acetamido-2-deoxy- β -D-glucopyranosidase (OGA) from CAZy database family GH84, that cleaves O-GlcNAc from proteins (Dong and Hart, 1994; Hart et al., 2007; Holt and Hart, 1986; Torres and Hart, 1984). An abundance of tightly regulated OGT and OGA are found in the brain, localised neuronal synapses where they regulate synaptosome proteins by the O-GlcNAc modification (Cole and Hart, 2001; Okuyama and Marshall, 2003; Yang et al., 2012).

It is now known that there are splice variants that produce two isoforms; the short (OGA-S) isoform is localised in lipid droplets and the long (OGA-L) isoform is localised in the cytosol and nucleus, Figure 4.2 (Comtesse et al., 2001; Keembiyehetty et al., 2011). They are identical in their catalytic domains but they vary in the C-terminal region, where OGA-L has a “pseudo-HAT” domain (Butkinaree et al., 2008; Toleman et al., 2004), see section 4.2.2.1, whereas OGA-S terminates at residue 677 (and in which the residues between 662-677 are different to OGA-L). OGA-S has lower activity than OGA-L, a K_M of 810 μ M compared to 170 μ M, possibly because of the absence of cellular cofactors or absence of the C-terminal region (Elsen et al., 2017), see section 4.2.2.1.

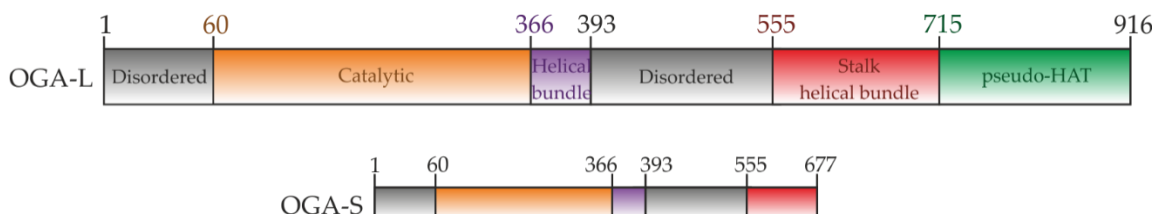


Figure 4.2. Domain structure of O-GlcNAc hydrolase long and short isoforms. OGA-S has the sequence RCTRNNLFSSNILSL from residue 663 to 677.

4.2.2.1 Activity of OGA

Initially, the *N*-acetylglucosaminidase activity of O-GlcNAcase was confused between hyaluronidase and *N*-acetylglucosaminidase activity (Gao et al., 2001; Heckel et al., 1998) Pils and colleagues hypothesised that the N-terminus had hyaluronidase activity and the C-terminal domain (715-916) had *N*-acetylglucosaminidase activity but characterised within the acetyltransferase family. They suggested that O-GlcNAcase is the evolutionary descendent of GNA1, a (non-GH) GCN5-related *N*-acetyltransferase (GNAT), that

catalyses the addition of acetyl to Glc6P (Schultz and Pils, 2002). The ambiguous activity was resolved by showing that Spy1600, a GH84 protein from *Streptococcus pyogenes*, was not active on galactosaminides or hyaluronan (Sheldon et al., 2006). Although OGA-S has much lower activity, it still retains the same kinetic properties, therefore, the N-terminal domain was assigned to have *N*-acetylglucosaminidase activity on protein substrates (Macauley and Vocadlo, 2009).

4.2.2.2 The structure of OGA

4.2.2.2.1 Bacterial homologues of *Homo sapiens* OGA

The first structural insights into the catalytic and stalk domains originated from the crystal structures of *C. perfringens* strain 13 NagJ (*CpNagJ* or *CpOGA*) (Rao et al., 2006) and *Bacteriodes thetaiotaomicron* VPI-5482 hexosaminidase (*BtGH84*) (Dennis et al., 2006), Figure 4.3. *CpNagJ* and *BtGH84* have a sequence identity to *HsOGA* of 27% and 26%, respectively and ~36% sequence identity to each other, Figure 4.3a. Since then, numerous structures of these proteins have been published; additionally, OGAs from *C. perfringens* ATCC 13124, *Oceanicola granulosus* HTCC2516 and *Thermobaculum terrenum* ATCC BAA-798 have also been structurally characterised.

At the N-terminus of *BtGH84* a β -sheet formed from six antiparallel β -strands and two α -helices adjoins the catalytic domain, Figure 4.3b (Dennis et al., 2006). However, this domain is not present in *HsOGA*, Figure 4.3c. An extremely disordered final carbohydrate binding module is only present in the bacterial orthologues, Figure 4.3a and b. Both *CpNagJ* and *BtGH84* crystallised and act in solution as a dimer with the interface at the stalk domain (helix 1 and 3), Figure 4.3b (Dennis et al., 2006).

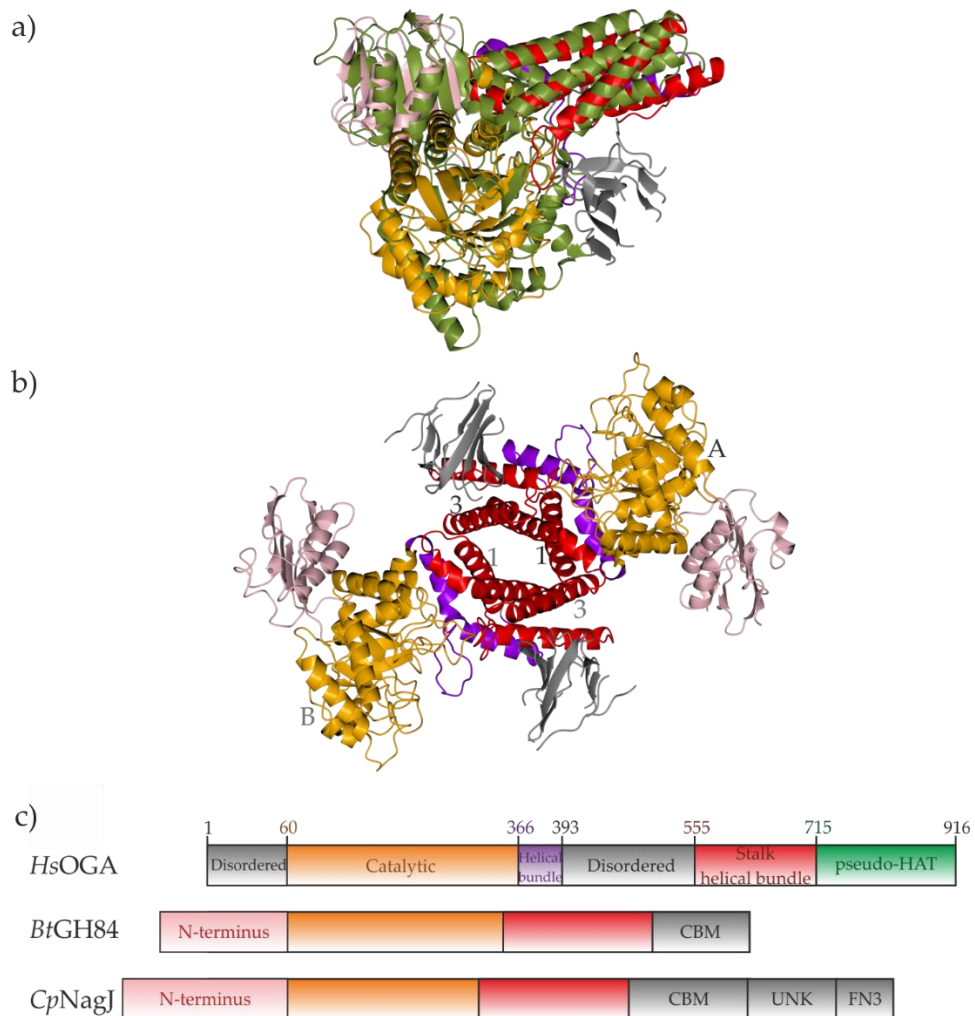


Figure 4.3. Domain structure of *HsOGA* and bacterial homologues. a) Superposition of chain A of *BtGH84* and *CpNagJ*, in green, (PDB ID: 2CBI). b) Structure of *BtGH84* (PDB ID: 2CHO) coloured according to part a. c) CBM = Carbohydrate binding module; UNK = unknown function; FN3 = fibronectin-3 homology module.

4.2.2.2.2 Expression strategies to achieve the structure of *HsOGA*

A crystal structure of OGA from *Homo sapiens* (*HsOGA*) was ultimately achieved by three groups (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). The strategies for the expression of the *HsOGA* DNA were varied but all were designed to remove the “pseudo”-HAT domain and the various sections of the regions predicted to be disordered, Figure 4.4.

Jiang and colleagues designed a fusion construct with a 10 residue glycine/serine amino acid linker (GGGGSGGGGS) (Li et al., 2017a). The linker directly joins the catalytic domain (residue 400) to the stalk domain (residue 553).

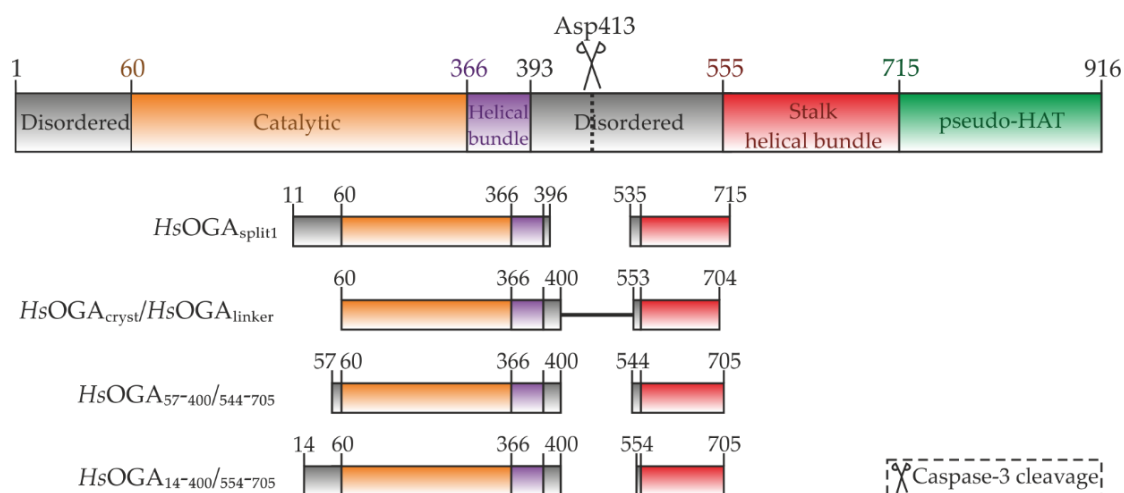


Figure 4.4. Construct design for the expression of soluble *HsOGA*. From top to bottom: the full length construct shown with domain boundaries and the scissors represent the caspase-3 cleavage site; *HsOGA_{split1}* (Roth et al., 2017) residues 11-396+535-715; *HsOGA_{linker}* (Li et al., 2017a); *HsOGA_{57-400/544-705}* and *HsOGA_{14-400/554-705}* (Elsen et al., 2017).

Klien and Davies and colleagues employed a construct screening process based on the apoptotic cleavage of OGA (Elsen et al., 2017; Roth et al., 2017). There is a cleavage site after Asp413, which is induced upon apoptosis, in the disordered region between the catalytic and stalk domains, Figure 4.4 (Butkinaree et al., 2008). The two fragments can recombine when associated and regenerate enzyme activity. Truncated constructs were designed from limited proteolysis, sequence and structural data.

4.2.2.2.3 The structure of *HsOGA*

Apo crystals of *HsOGA*, which produced a relatively low resolution structure between 2.4 Å to 3.0 Å, have been published (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). The catalytic domain has a $(\beta/\alpha)_8$ -barrel formation and the stalk domain comprises of six α -helices as supported by bacterial orthologue structures, Figure 4.5a and c.

An overlay of the human structures revealed very similar structures with slight differences in the positioning of the C-terminal stalk domain with respect to the active site (resulting in a narrower or wider active site), Figure 4.5b. When the active site of *HsOGA* was overlaid with that of the bacterial orthologue, *BtGH84*, the N-terminal $(\beta/\alpha)_8$ -barrel did, again, have a similar rigid structure (RMSD of 0.7 Å of the C α s) with the differences

lying in the position of the C-terminal stalk domain (RMSD of 1.1 Å) demonstrating a high degree of structural flexibility (Li et al., 2017a). This flexibility is thought to be driven by the intersubunit interactions of residues on helix $\alpha 16$ to coin two states, narrow and wide (Elsen et al., 2017).

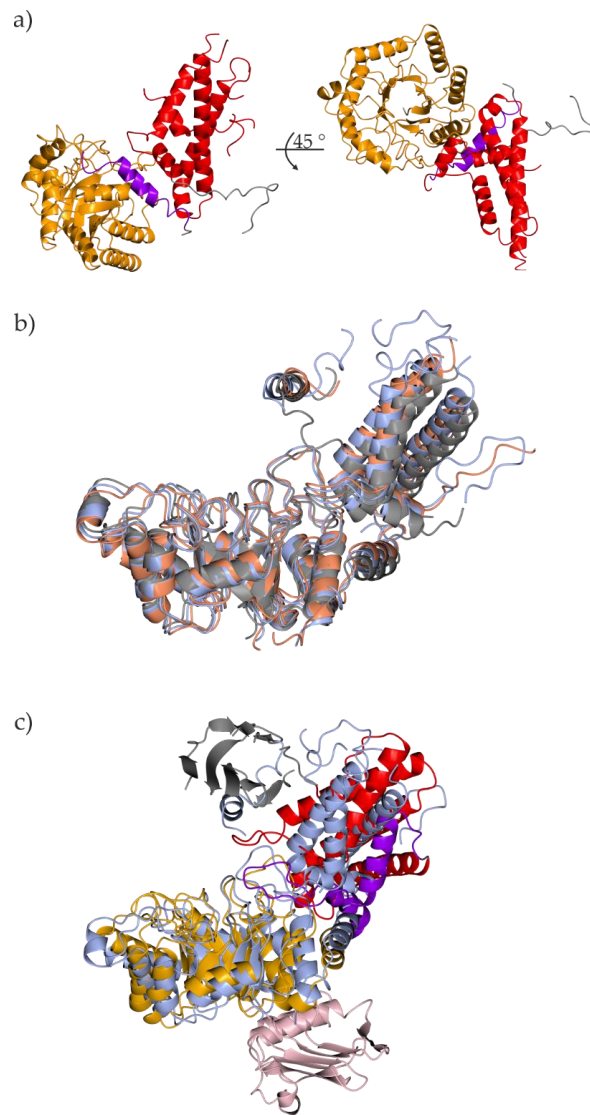


Figure 4.5. The variable conformations of the C-terminal domain. a) Ribbon representation of chain A coloured according to the domain functions. b) Overlay of chain A of *HsOGA*_{split1} (PDB ID: 5M7R) in blue, *HsOGA*_{linker} (PDB ID: 5TKE) in grey and *HsOGA*_{57-400/544-705} (PDB ID: 5UHK) in coral. c) Overlay of *BtGH84* (PDB ID: 2CHO) and *HsOGA*_{split1} using superposition of residues in the $(\beta/\alpha)_8$ barrel.

4.2.2.2.4 Dimer formation is stabilised by a swapped helix

Dimerisation was confirmed by both Size Exclusion Chromatography with Multi-Angle Light Scattering and Analytical Ultracentrifugation of *HsOGA*_{split1}, Figure 4.6a.

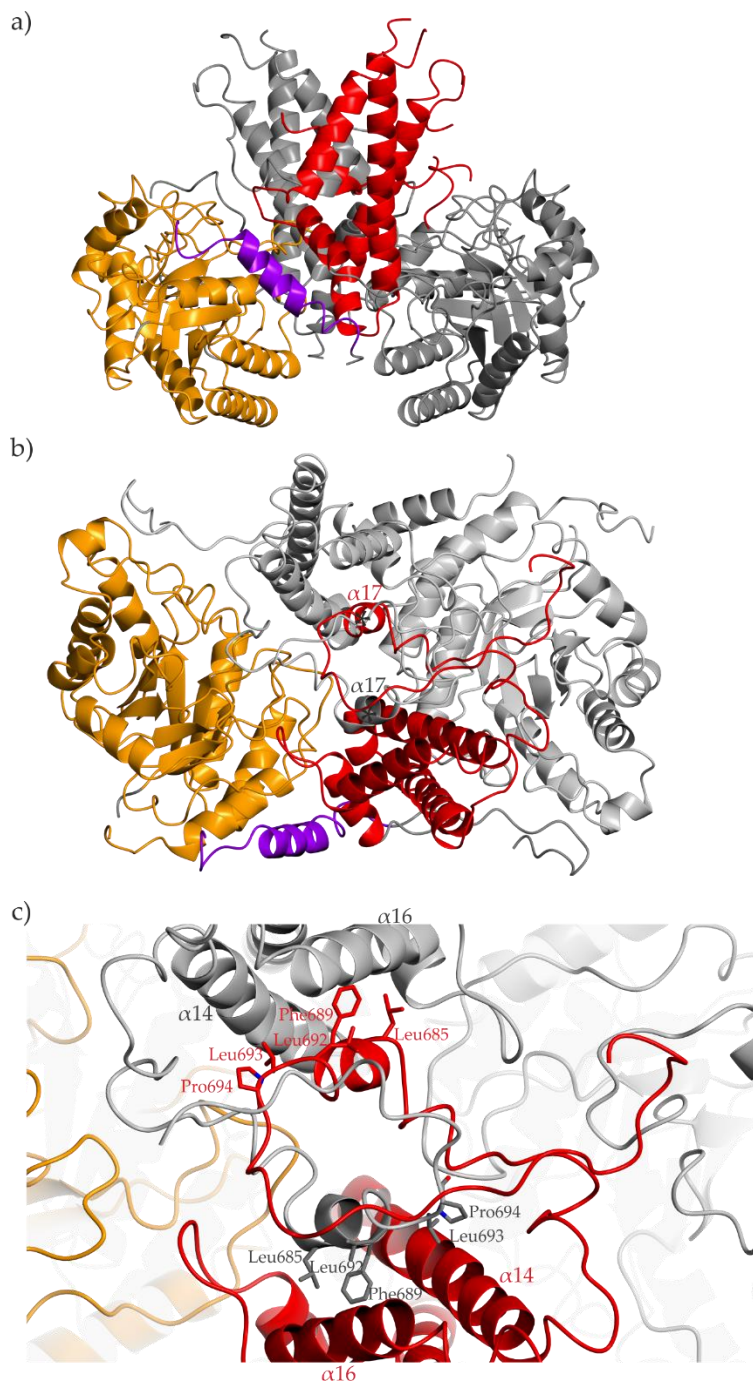


Figure 4.6. *HsOGA* dimer formation. a) Chain B of *HsOGA*_{split1} (PDB ID: 5M7R) is shown in grey. b) and c) PDB ID: 6HKI. b) Monomer A is coloured by domain function and monomer B is coloured in grey. Domain swapped helix, $\alpha17$, helps to stabilise the dimer formation. Two asterisks indicate the domain swapped helix. c) Highlighted residues on the $\alpha17$ helix interact with residues on $\alpha14$ and $\alpha16$.

To stabilise the two domains and the dimer formation, the final C-terminal helix, $\alpha 17$, from the stalk domain (Gln676 to Pro694) undergoes a dimer “swap”, Figure 4.6b. Hydrophobic interactions of the non-polar groove between Leu685, Phe689, Leu692, Leu693, Pro694 of $\alpha 17$ and both residues of helix $\alpha 14$ (helix 1 of the stalk domain) Met578, Leu579 and Phe582, and residues of helix $\alpha 16$ (helix 3 of the stalk domain) Ile647, Ile650, Met651, Met653, Val654 and Phe657, Figure 4.6c. OGA-S, which terminates at residue 667, lacks the swapped helix and acts as a monomer. This is a distinguishing factor of the two isoforms.

4.2.2.2.5 Structure and function of the pseudo-HAT domain

In vitro intrinsic histone acetyltransferase (HAT) activity was proposed for the C-terminal O-GlcNAcase domain and a new bifunctional name was assigned, nuclear cytoplasmic O-GlcNAcase and acetyltransferase (NCOAT) (Toleman et al., 2004). The functional regulation of NCOAT was suggested to reduce the glycosylation level of transcriptional activators and increase the level of acetylation on histones to coordinate gene transcription. However, NCOAT HAT activity on synthetic histone substrate tails and free core histones could not be repeatedly observed (Butkinaree et al., 2008).

Currently, the structure of pHAT from *Homo sapiens* (HspHAT), or any other eukaryotic species, has not been achieved. The full-length protein cannot be crystallised, and the HspHAT domain alone is not soluble. There are two structures from bacterial homologues, *Oceanicola granulos* (OgpAT) (Rao et al., 2013), Figure 4.7a, and *Streptomyces sviveus* (SsAT) (He et al., 2014), Figure 4.7b. They have 26% and 29% sequence identity to HspHAT. The pyrophosphate-interacting amino acids are located on a binding motif, the “P-loop”. The motif has a conserved sequence of Q/R-X-X-G-X-G/A. However, HspHAT has the sequence ⁸⁵²TDPSVA⁸⁵⁷, suggesting that it does not support the negatively charged pyrophosphate of AcCoA, Figure 4.7c. Surface plasmon resonance analysis of an OgpAT mutant, whereby the P-loop was mutated to the HsAT sequence, showed no detectable binding of AcCoA (Rao et al., 2013).

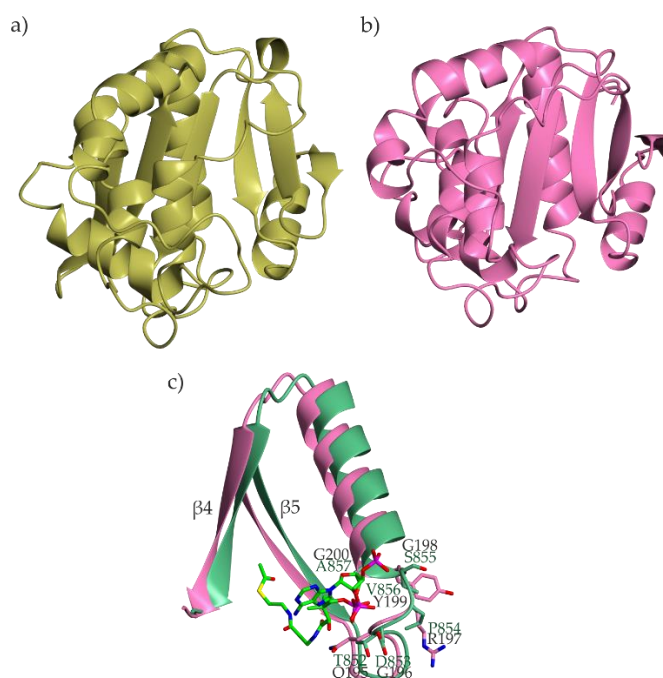


Figure 4.7. Structures of bacterial acetyltransferases. Gold, *OgpAT* (PDB ID: 3ZJ0). b) Pink, *SsAT* (PDB ID: 4BMH). c) β -strand splay and residues of the P-loop interact with AcCoA in green.

4.2.2.3 Mechanism of catalysis

Between the catalytic domain of monomer A and the C-terminal stalk domain of monomer B lies a V-shaped cleft ($\sim 22 \times 25 \text{ \AA}$ with an angle of $\sim 70^\circ$) that forms the active site, Figure 4.8a (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). Bioinformatics sequence alignment and mutation studies compared OGA from *Xanthomonas axonopodis* to the active site of GH20 β -hexosaminidases, which contain a conserved DE catalytic pair (Rigden et al., 2003), and assigned Asp174 as the polarising residue and Asp175 as a general acid/base catalyst (Çetinbaş et al., 2006).

In the first step cyclisation occurs, Figure 4.8b. Asp175 acts as a general acid to promote the departure of the aglycon leaving group. Asp174 acts as a general base to polarise the acetamido group to increase nucleophilicity and stabilise the oxazoline intermediate. Water then replaces the aglycon group. Lastly, the ring opening step occurs. Asp174 acts as a general acid and the acetamido group departs. Asp175 acts as a general base to facilitate attack of the water molecule on the anomeric centre. The catalytic aspartate

residues, Asp174 and Asp175, are located on an inherently flexible loop (residues Leu172 to Ala182) that moves closer to the active site upon ligand binding.

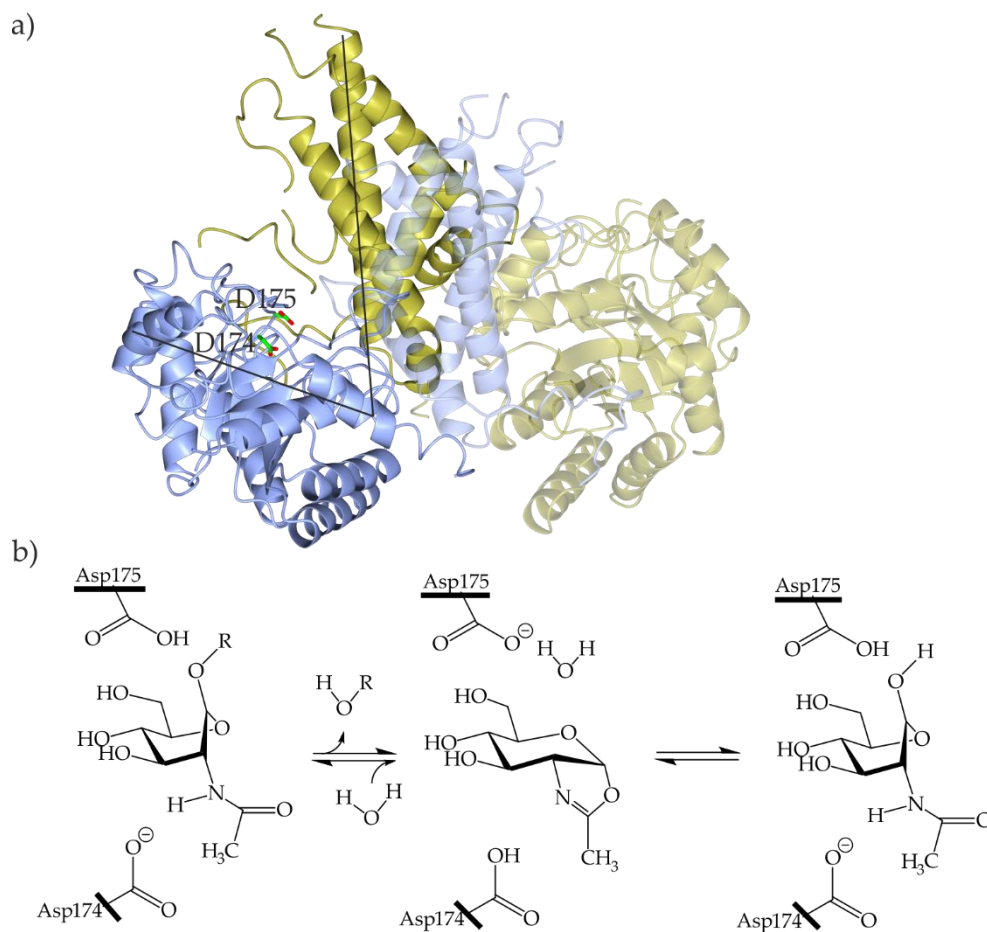


Figure 4.8. Catalytic mechanism of OGA. a) The catalytic aspartate pair is highlighted as green residues. Monomer A is shown in ice blue and monomer B is shown in gold. b) The *N*-acetyl group is involved in nucleophilic participation in neighbouring group catalysis of O-GlcNAc from serine or threonine (R).

4.2.2.4 Substrate recognition

The specific recognition of *Hs*OGA for glycopeptide substrates had been postulated in reports that resulted in two different conclusions. The first surmised that OGA must recognise the sugar moiety due to similar catalytic efficiencies against different glycopeptides and modifications to the C2, C4 and C6 positions of GlcNAc were not tolerated (Li et al., 2018b; Schimpl et al., 2012; Shen et al., 2012). The second revealed that mutations in a number of the residues that line the pocket reduced the affinity to the peptides and molecular docking simulations with glycosylated peptides showed

mainchain hydrogen bonds between the +1, +3 and +4 positions with enzyme residues at the -1 and -2 positions (Martin et al., 2014; Schimpl et al., 2010).

Crystallographic evidence would suggest that OGA makes contacts with both the peptide and the sugar moiety. The recognition mode was initially investigated by obtaining the structure of *Cp*OGA in complex with the tumour suppressor protein, p53 (LWVDSTPP), TAB1 (VPYSS) and *Hs*OGA (VAHSGA) (Schimpl et al., 2012). The different peptides were tethered in similar conformations in the active site.

Recently, a structure of *Hs*OGA in complex with p53 (QLWVDSTPPPG) was obtained (Li et al., 2017a). Essential hydrophobic interactions were created with proline residues in the peptide in the +2 and +3 position. In the apo structure from Roth, et al. the C-terminus of the protein, from residues Pro707 to Tyr715, can be traced to the active site grooves in which they bind (Roth et al., 2017). The position of Tyr715 lies directly on top of the position of the O6 of an O-GlcNAcylated serine or threonine residue. Unexpectedly, it bound in the “opposite” direction (C to N) to the bacterial:peptide complex structures and the *Hs*OGA:p53 structure.

Further peptide complexes were obtained with GlcNAcylated α -crystallin B, TAB1, ELK1 and Lamin B1, Figure 4.9 (Li et al., 2017b). The peptide complexes showed TAB1, lamin B1 and p53 glycosylated peptides bind in a forward direction (N to C) and α -crystallin B, ELK1 and the C-terminus of OGA, Lys713-Tyr715, bind in a reverse direction (C to N) (Li et al., 2017b; Roth et al., 2017). The ability of peptides to bind in a bidirectional yet conserved V-shape conformation in the substrate binding cleft is due to different side-chain interactions with OGA, compared to the conserved GlcNAc binding site (Hu, 2018). The question now stands as to whether *Hs*OGA is able to bind two different substrates concurrently to achieve highly efficient deglycosylation.

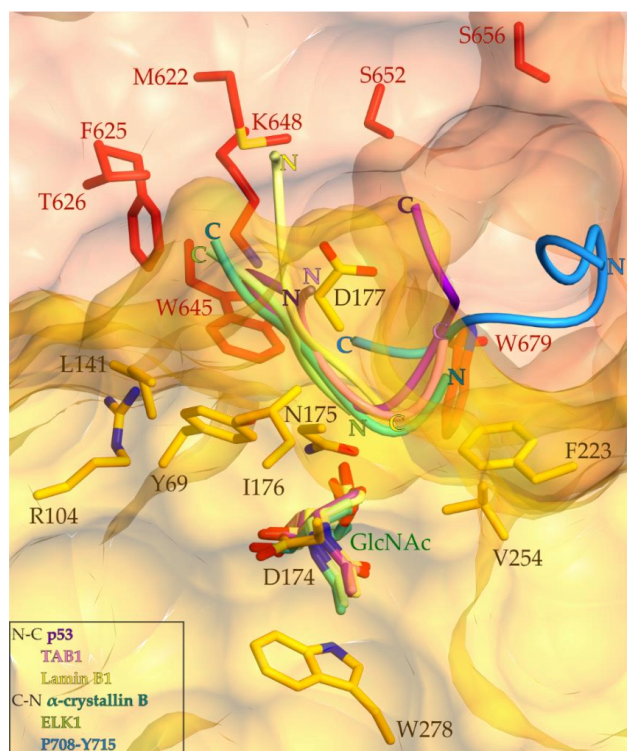


Figure 4.9. Overlay of *HsOGA* in complex with various peptides. α -crystallin B (PDB ID: 5VVV, dark green, residues TSTSL), ELK1 (PDB ID: 5VVT, light green, residues TLSP), TAB1 (PDB ID: 5VVU, pink, residues PYSS), Lamin B1 (PDB ID: 5VVX, yellow, residues SSRVT), and p53 (PDB ID: 5UN8, purple, residues WVDSTPP). The C-terminus of *HsOGA*, residues LTPTSKVY, are shown in blue. The peptides are shown as worms with the N- and C-termini labelled. The catalytic and neighbouring residues are highlighted in the active site in orange.

4.2.3 Summary of the results in this chapter

Further insight is needed to define the action of OGA in terms of substrate specificity. The binding preference for orientation and conformation of the target peptide is still not understood. Furthermore, there has been a deficit of data on the full-length *HsOGA* with full-length protein targets. Efforts to obtain peptide crystal structures, to improve crystals and to study possible effects of phosphorylation, will be described in this chapter.

In the first section, I describe attempts to obtain a peptide complex between *HsOGA*_{split} and an O-glycosylated (serine residue) casein kinase II peptide. To prevent GlcNAc hydrolysis, the protein was mutated to an inactive (D174AD174A) as well as a partially-inactive (D175N) variant. However, the protein could not be crystallised successfully.

Subsequently, the peptide was S-glycosylated at a cysteine residue as OGA is unable to hydrolyse the C-S bond. Unfortunately, there was ambiguity as to the direction of the peptide in the active site and the electron density was poor.

When soaking a target ligand into a crystallisation drop of *HsOGA_{split1}* crystals, the C-terminus has to be expelled. To increase the availability of the active site to ligands, crystals with a different crystal packing were explored, firstly, by using surface entropy reduction. Various mutants, which had couples of mutations from lysine and glutamate protein surface residues to alanine (information on the constructs can be found in the appendix Table 3.), were tested for their ability to crystallise in a new space group. Mutation of Glu602 and Glu605 to alanine resulted in a structure where the C-terminus could be traced into the active site of only symmetry related monomer A. The C-terminus had moved through the active site so that Lys713 was in the position of a superimposed GlcNAc moiety. However, the crystals were fragile and could not be soaked with a ligand.

Several sites of *HsOGA_{split1}* (S268, S364, T370, T709 and S712) are phosphorylated to regulate the enzyme. The second strategy was to exploit the phosphorylated state to obtain a different crystal form. In this section, the kinase, casein kinase II, was shown by liquid chromatography-tandem mass spectrometry (LC-MS/MS) to phosphorylate Ser364, located in a disordered region (Asn337-Asp371) of the crystal structure. A phosphomimetic mutation of Ser364 to alanine, revealed the mutant had the same secondary structure and catalytic efficiency to the WT and the active site was blocked by Tyr715.

O-GlcNAc Hydrolase Binds S-GlcNAcylated Peptides

4.3 Obtaining a peptide complex structure

Five structures of *HsOGA* in complex with various O-GlcNAcylated peptides have been obtained by others (Li et al., 2017b, 2017a). Mutation of Asp175 to asparagine in *HsOGA* rendered the protein catalytically inactive so that the glycosidic bond would not be hydrolysed. In this thesis, a structure of a catalytically inactive *HsOGA*_{split1} mutant in complex with casein kinase II was attempted but crystals could not be obtained. To circumvent this problem, the serine residue of a casein kinase II (CKII) peptide (³³⁹YPGGSTPVSSANMM³⁵² α1-subunit) was mutated to a cysteine and then GlcNAcylated.

4.4 Experimental methodology

Matt Alteen in Professor David Vocadlo's lab enzymatically (OGT) glycosylated the CKII peptide. DNA of *HsOGA*₁₁₋₃₉₆ was mutated using the NEB Q5 Site-Directed Mutagenesis Kit with Q5 Hot Start High-Fidelity DNA Polymerase. The primers are described in table 1.2. of the appendix. The inactive protein was produced, purified and crystallised under the same conditions as the WT, described in Roth et al., 2017. The crystals were soaked with a 5 mM final concentration of peptide for 5 days. Crystals were sent to Diamond I04-1 and processed using the reflections and freeR set from PDB ID:5M7R.

4.5 Results

4.5.1 CKII was present in one active site of a homodimer

The peptide could be partially built into the active site of monomer B only, Figure 4.10a. In the active site of monomer A, the C-terminus of *HsOGA* from Pro707-Tyr715 was still observed to be blocking the active site. Although a 14 residue peptide had been soaked into the crystal, only two of these residues were visible, Cys347 and Ser348.

There are several interactions with the GlcNAc moiety and protein residues, from N2, O7, O3, O4 and O6 of GlcNAc and Asp174, Asn280, Lys98, Gly67, Asn313 and Asp285. There are no interactions between the CKII peptide and the active site residues of *HsOGA*. The weak interactions between the CKII peptide and *HsOGA* allow for flexibility of the peptide (or protein) in the active site. Disorder was also observed in the structures with

TAB1, ELK1, α -crystallin B and Lamin B1 (Li et al., 2017b). Very few of the amino acids for these peptides could be built into the active sites of *HsOGA*.

An overlay of *HsOGA* in complex with the peptide p53 reveals the similar position of GlcNAc in the active site, Figure 4.10b. The p53 peptide binds into the active site in a forward direction. Since Val346 could not be built into the electron density which supports the amino acid Ser348, CKII was concluded to bind in a forward direction.

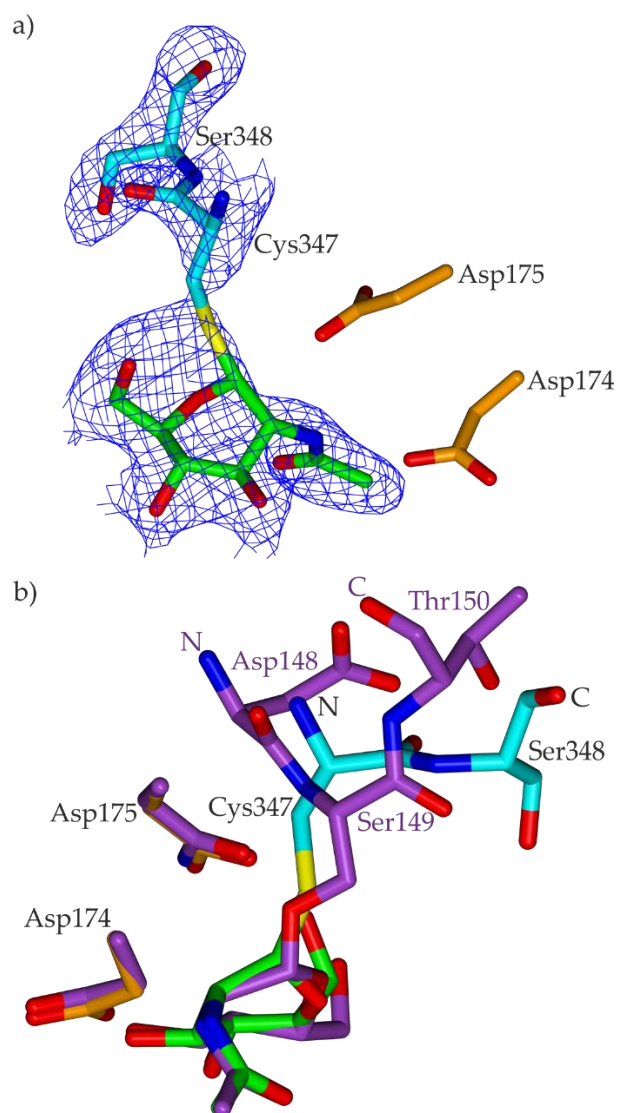


Figure 4.10. Structure of CKII GlcNAcylated peptide in complex with *HsOGA*. a) GlcNAc is shown in green, residues of the peptide CKII are in blue and the catalytic residues of *HsOGA* are in orange. b) Superposition of *HsOGA* in complex with GlcNAcylated p53 in purple (PDB ID: 5UN8).

4.6 Discussion

CKII peptide joins the peptides TAB1, lamin B1 and p53 that traverse the active site in a forward direction from the N-terminus to the C-terminus.

The C-terminus must be expelled before a ligand of a higher affinity can bind. The GlcNAc moiety forms more interactions with residues of the active site and should bind with a higher affinity than the C-terminus. It would be possible to test this theory by designing an inactive construct that is missing the residues from 704-715 and constructing a synthetic peptide of those residues. Association of the two could be measured by ITC or another biophysical method. A second route to test this interaction would be to mutate Tyr715 to a cysteine or serine and glycosylate this residue.

To obtain a structure of a peptide bound in both active sites of *HsOGA* different constructs were designed with the aim of removing the C-terminal peptide bound in the active site. The constructs will be discussed in the next sections.

**Structural Studies of a Surface Entropy Reduction
Mutant of O-GlcNAcase**

Declaration

Under the guidance of Prof. Gideon J. Davies, I designed the surface entropy reduction experiment. I performed the crystallography experiments at the University of York. Prof. David Vocadlo provided advice regarding the kinetics and circular dichroism experiments. I conducted the kinetics experiments in the laboratory of Prof. David Vocadlo, Simon Fraser University, with the assistance of Matt Alteen. The circular dichroism experiment was conducted by Dr. Dustin King, Simon Fraser University. I wrote the complete first draft of the manuscript and with Prof. Gideon J. Davies revised it into its final form.

A handwritten signature in black ink, appearing to read 'Gideon J. Davies', is centered on a light gray rectangular background.

Signed

Prof. Gideon J. Davies

Reference:

Males, A., and Davies, G. J. (2019). Structural studies of a surface-entropy reduction mutant of O-GlcNAcase. *Acta Crystallogr D Struct Biol.* 75, 70-78.

Supplementary information – pg 239



Structural studies of a surface-entropy reduction mutant of O-GlcNAcase

Alexandra Males and Gideon J. Davies*

Department of Chemistry, University of York, York YO10 5DD, England. *Correspondence e-mail: gideon.davies@york.ac.uk

Received 24 September 2018

Accepted 21 November 2018

Edited by M. Czjzek, Station Biologique de Roscoff, France

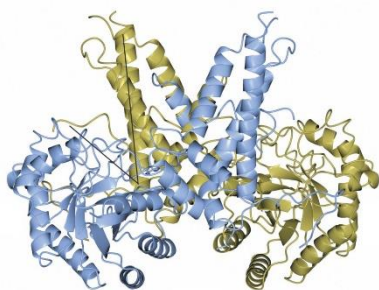
Keywords: O-GlcNAc; O-GlcNAcase; neurodegeneration; surface-entropy reduction; crystallization.**PDB reference:** surface-entropy reduction mutant of O-GlcNAcase, 6hki**Supporting information:** this article has supporting information at journals.iucr.org/d

The enzyme O-GlcNAcase catalyses the removal of the O-GlcNAc co/post-translational modification in multicellular eukaryotes. The enzyme has become of acute interest given the intimate role of O-GlcNAcylation in tau modification and stability; small-molecular inhibitors of human O-GlcNAcase are under clinical assessment for the treatment of tauopathies. Given the importance of structure-based and mechanism-based inhibitor design for O-GlcNAcase, it was sought to test whether different crystal forms of the human enzyme could be achieved by surface mutagenesis. Guided by surface-entropy reduction, a Glu602Ala/Glu605Ala variant [on the Gly11–Gln396/Lys535–Tyr715 construct; Roth *et al.* (2017), *Nature Chem. Biol.* **13**, 610–612] was obtained which led to a new crystal form of the human enzyme. An increase in crystal contacts stabilized disordered regions of the protein, enabling 88% of the structure to be modelled; only 83% was possible for the wild-type construct. Although the binding of the C-terminus was consistent with the wild type, Lys713 in monomer *A* was bound in the -1 subsite of the symmetry-related monomer *A* and the active sites of the *B* monomers were vacant. The new crystal form presents an opportunity for enhanced soaking experiments that are essential to understanding the binding mechanism and substrate specificity of O-GlcNAcase.

1. Introduction

The regulation of O-GlcNAc cycling on thousands of nuclear and cytoplasmic proteins is coordinated by two enzymes. O-GlcNAc transferase (OGT) catalyses the addition of GlcNAc, derived from UDP-GlcNAc, to serine and threonine residues, and O-GlcNAcase (OGA; CAZY database family GH84) cleaves O-GlcNAc (Torres & Hart, 1984; Holt & Hart, 1986; Kreppel *et al.*, 1997; Dong & Hart, 1994; Lubas *et al.*, 1997; Hart *et al.*, 2007). Two isoforms, OGA-L and OGA-S, are localized to the nucleus/cytoplasm (Comtesse *et al.*, 2001) and to the surface of lipid droplets, respectively. The reciprocal relationship between O-phosphorylation and O-glycosylation on the particular protein tau has been keenly studied in the context of neurodegeneration (Arnold *et al.*, 1996; Yuzwa *et al.*, 2008, 2014; Shen *et al.*, 2012; Griffith & Schmitz, 1995; Gao *et al.*, 2001; Liu *et al.*, 2004). In patients with Alzheimer's disease, tau undergoes hyperphosphorylation, causing it to dissociate from microtubules and aggregate into paired helical filaments (PHF) and neurofibrillary tangles (NFTs) (Grundke-Iqbal *et al.*, 1986; Marotta *et al.*, 2015). O-GlcNAc cycling has also been implicated in tumorigenesis owing to its significant role in orchestrating a vast number of cellular processes, for example transcriptional and cytoskeletal regulation, cell signalling and division, and metabolism (Slawson & Hart, 2011).

Structural characterization of the human O-GlcNAcase orthologue (*Hs*OGA-L/*Hs*OGA) showed dimer formation



OPEN ACCESS

both in solution and in the crystal where it has been shown to be promoted by helix-exchange (Roth *et al.*, 2017) in contrast to the non-helix-exchanged bacterial dimers (Dennis *et al.*, 2006; Schimpl *et al.*, 2010; Rao *et al.*, 2006). Composed of two ordered domains, the N-terminal glycoside hydrolase domain forms a $(\beta/\alpha)_8$ barrel (Li *et al.*, 2012), while the C-terminal stalk domain (Toleman *et al.*, 2004; He *et al.*, 2014) is composed of six α -helices. *a17*, consisting of Glu676–Pro694, undergoes a dimer ‘swap’ (Roth *et al.*, 2017), playing a structural role that contributes to dimerization (Li, Li, Lu *et al.*, 2017; Elsen *et al.*, 2017; Roth *et al.*, 2017). Located in a V-shaped cleft formed between the N-terminal domain of monomer *A* and the C-terminal domain of monomer *B* (Fig. 1*a*) are the catalytic residues, Asp174 and Asp175, that initially act as a general acid and a general base through a two-step retaining mechanism (Macauley *et al.*, 2005). The terminal domain, which is not present in OGA-S, has a ‘pseudo’-histone acetyltransferase activity but was not included in the crystallized construct in the three *HsOGA* structures as it has a high degree of disorder.

The recognition mode for glycopeptide substrates has been explored by obtaining structures of the catalytically inactive Asp175Asn mutant of *HsOGA* in complex with a variety of glycosylated peptides (Li, Li, Lu *et al.*, 2017; Li, Li, Hu *et al.*, 2017). The peptides bound in each structure can be segregated into two binding modes with forward or reverse orientations of the peptide (amino to carboxyl or carboxyl to amino, respectively) within the binding site. Initially, crystallographic peptide studies were conducted using an orthologue from the bacterium *Clostridium perfringens* (Schimpl *et al.*, 2012; Mariappa *et al.*, 2015) in complex with TAB1, lamin B1 and p53 glycosylated peptides. These peptides bound in the forward direction. Later, Li, Li, Hu *et al.* (2017) compared the same glycosylated peptides with *HsOGA* and they were found to act in the same way. However, α -crystallin B and ELK1 bound in the reverse direction, strengthening the interest in determining how OGA selects its target (Li, Li, Hu *et al.*, 2017).

In the ‘apo’ structure, *HsOGA*_{11–396+535–715} (thus named owing to the co-expression of two plasmids containing residues 11–396 and 535–715), which was published by Roth *et al.* (2017) and deposited as PDB entry 5m7r, the C-terminus of monomer *A* from residues Pro707 to Tyr715 can be traced in the reverse direction into the active-site groove of the symmetry-related molecule of monomer *B* (Fig. 1*b*). When overlaid with the *HsOGA*–p53 complex (PDB entry 5un8; Li, Li, Lu *et al.*, 2017), the position of the Tyr715 hydroxyl group lies directly on top of the O6 of O-GlcNAc.

To conduct biochemical/biophysical studies and rational drug design, complete and coherent structures are required. However, the published structures of *HsOGA* have stretches of residues that are incomplete (Elsen *et al.*, 2017; Li, Li, Lu *et al.*, 2017; Li, Li, Hu *et al.*, 2017). For example, the structure described by Roth *et al.* (2017) has regions within the N-terminus (Glu11–Arg58 and Lys341–Thr370) and the C-terminus (Asp696–Pro706) that are not modelled in the structure. Expulsion of the C-terminus of *HsOGA*_{11–396+535–715}

from the active site is required before a competing compound can bind, therefore making it challenging to conduct crystal-soaking experiments; this is a problem for weak-binding compounds/inhibitors.

The work in this paper utilized surface-entropy reduction (SER) to enhance the structural characterization of *HsOGA* and to contribute towards the hypothesis for the substrate-recognition mode of OGA, in which either the O-GlcNAc moiety or the peptide sequence is important for recognition. Using the *SERP* online server (Goldschmidt *et al.*, 2007), potential pairs of mutations were identified for *HsOGA*. The rational design of mutating clusters of residues on a protein is a favourable strategy, with the aim of making the protein more susceptible to crystallization (Derewenda, 2004*a,b*; Derewenda & Vekilov, 2006; Cooper *et al.*, 2007). Surface-entropy reduction (SER) is a concept in which flexible, solvent-exposed residues, primarily lysine and glutamate, are mutated to alanine to reduce the entropic loss during the packing of the

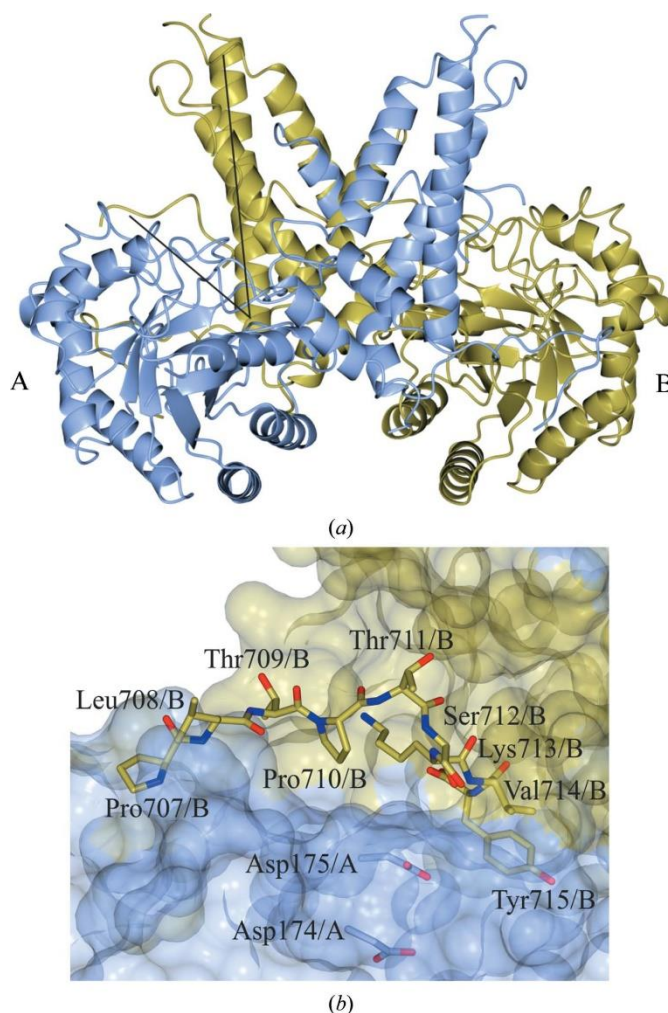


Figure 1
HsOGA forms a dimer in solution and in the crystal structure. Monomer *A* is shown in blue and monomer *B* in gold. (a) One of the two active sites is indicated by intersecting black lines. (b) The C-terminal peptide, Pro707–Tyr715 in chain *B*, bound in the active site of monomer *A* with the catalytic residues, Asp174/*A* and Asp175/*A*, displayed.

protein into a crystalline lattice (Longenecker *et al.*, 2001; Mateja *et al.*, 2002; Vekilov, 2003; Vekilov *et al.*, 2002).

Following the co-expression strategy of Roth *et al.* (2017) using pACYC-Duet-1 Gly11–Gln396 and pET-YSBLIC3C Lys535–Tyr715 plasmids, the successful pair of mutations was Glu602Ala/Glu605Ala (*HsOGA*_{E602AE605A}). Unlike *HsOGA*_{1–396+535–715}, the C-terminus of monomer A can be traced into the active site of the symmetry-related monomer A, with Lys713 binding in the position of O-GlcNAc. Furthermore, previously disordered loops had become ordered and could be built into the final model. Additionally, the activities and secondary-structure profiles of full-length *HsOGA* (*HsOGA*_{FL}), *HsOGA*_{11–396+535–715} and *HsOGA*_{E602AE605A} were examined to verify the loss of the ‘pseudo-HAT’ domain and that the SER mutation did not alter the activity. The results showed that the SER mutant exhibited similar K_m values to the full-length enzyme, since the mutation is distant from the active site, highlighting the potential of SER variants for studying the structural and ligand-binding characteristics of *HsOGA*.

2. Materials and methods

2.1. Macromolecular production

The cloning of the constructs has been described previously (Roth *et al.*, 2017). Homologous DNA for the Lys535–Tyr715 construct, in the vector pET-YSBL-LIC-3C, was mutated using a Q5 Site-Directed Mutagenesis Kit with Q5 Hot Start High-Fidelity DNA Polymerase. The forward primer, A GAT AGC gct AAA ATC gct GAA TGG, was designed to mutate the primary sequence A GAT AGC GAA AAA ATC GAA GAA; the reverse primer was TTA CCC TTG CAG TTA ACC GAA. NEB 5-alpha competent *Escherichia coli* cells were transformed with the Lys535–Tyr715 E602AE605A construct. The DNA was extracted and sequenced to verify the mutation.

The Gly11–Gln396 and Lys535–Tyr715 E602AE605A constructs in the vectors pACYCDUET-1 and pET-YSBL-LIC-3C, respectively, were co-transformed into *E. coli* BL21 (DE3) cells. Luria–Bertani broth was inoculated with a cell suspension and was incubated at 37°C until the OD₆₀₀ reached 1.0. The cells were induced with 1 mM IPTG and incubated at 16°C for 20 h.

The purification of *HsOGA*_{E602AE605A} followed the same purification protocol as that of *HsOGA*_{11–396+535–715} described previously (Roth *et al.*, 2017; Supplementary Figs. S1a and S1b).

2.2. Crystallization

*HsOGA*_{E602AE605A} was initially crystallized by sitting-drop vapour diffusion at 15 mg ml⁻¹ under condition E11 of the PACT *premier* screen from Molecular Dimensions (Newman *et al.*, 2005): 0.2 M sodium citrate tribasic, 20% PEG 3350. Further optimization of the conditions to 0.2 M sodium citrate tribasic pH 7.5, 17% polyethylene glycol 3350 in a 48-well MRC MAXI optimization plate improved the crystal shape.

Table 1

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Beamline I04-1, DLS
Wavelength (Å)	0.9282
Temperature (K)	100
Detector	PILATUS 6M-F
Rotation range per image (°)	0.10
Total rotation range (°)	360
Exposure time per image (s)	0.040
Space group	<i>P</i> 3 ₁ 21
<i>a</i> , <i>b</i> , <i>c</i> (Å)	222.2, 222.2, 72.4
α , β , γ (°)	90.0, 90.0, 120.0
Resolution range (Å)	192.4–3.3 (3.5–3.3)
Total No. of reflections	523842 (75864)
No. of unique reflections	31103 (4440)
Completeness (%)	99.8 (98.6)
Multiplicity	16.8 (17.1)
$\langle I/\sigma(I) \rangle$	7.3 (1.8)†
$R_{p.i.m.}$	0.11 (0.62)
CC _{1/2}	0.99 (0.67)
Overall <i>B</i> factor from Wilson plot (Å ²)	58.53

† The mean $\langle I/\sigma(I) \rangle$ in the outer shell fell below <2.0 at 3.4 Å resolution.

The total volume of the drop was 1 µl and the protein: reservoir solution ratio was 1:1; the total volume in the reservoir was 100 µl.

2.3. Data collection and processing

Diffraction images were collected on beamline I04-1 at Diamond Light Source (DLS). After data collection, the diffraction images were integrated using the *-3dii* option in *xia2* (Winter, 2010) and reintegrated using *AIMLESS* (Evans, 2006; Evans & Murshudov, 2013) from the *CCP4* software suite (Winn *et al.*, 2011). Data-collection and processing statistics are given in Table 1.

2.4. Structure solution and refinement

Molecular replacement against the coordinates of PDB entry 5m7r was conducted using *MOLREP* (Vagin & Teplyakov, 2010). Refinement of the model was conducted using multiple rounds of *REFMAC* (Murshudov *et al.*, 1997, 2011; Pannu *et al.*, 1998; Winn *et al.*, 2003; Vagin *et al.*, 2004; Nicholls *et al.*, 2012) and manual model building in *Coot* (Emsley *et al.*, 2010). Waters were added using Find Waters in *Coot* and validated. The data were processed to a resolution of 3.3 Å (Table 2).

2.5. Michaelis–Menten kinetics

Michaelis–Menten kinetics were assayed using *HsOGA*_{FL} and *HsOGA*_{11–396+535–715} as positive controls against the mutant *HsOGA*_{E602AE605A}. In a 200 ml reaction volume, 50 nM protein solution and a serial dilution of the ligand 4-nitrophenyl *N*-acetyl-β-D-glucosaminide (*p*NP-GlcNAc) from 1500 to 11.7 µM [dissolved in 2.5% DMSO (final concentration)] was added to PBS buffer at pH 7.4. The reaction was monitored at 405 nM continuously using a Molecular Devices SpectraMax M5 spectrophotometer. The

experiment, which was conducted at 25°C, was duplicated and each ligand concentration was repeated in triplicate.

GraphPad Prism v.5 was used to process the data, with nonlinear regression of Michaelian saturation curves. The initial velocities were calculated from the linear range of the reaction-progress curve. A standard curve of 4-nitrophenol was used to extract a molar extinction coefficient.

2.6. Circular-dichroism spectroscopy

The protein samples were dialysed overnight into 25 mM sodium phosphate pH 8.0 and diluted to 0.1 mg ml⁻¹. The spectra were recorded at 21 °C in a QS 248 0.2 mm cuvette with 0.5 s per point and 78 s per read. The wavelength ranged from 195 to 320 nm. The background for each protein was measured immediately before the experiment in the same cuvette and values were taken as averages from triplicate reads.

3. Results and discussion

3.1. Comparison of the mutant crystal structure with that of the wild type

To incorporate protein molecules into a crystal, a thermodynamic cost is endured to bury hydrophobic residues into a constrained conformation and from the immobilization of the prevalent flexible hydrophilic side chains on the surface (Avbelj & Fele, 1998). Reducing the entropic shield can lead to an increase in the variety of conditions, morphologies and crystallographic space groups (Parthasarathy *et al.*, 2008; Kim *et al.*, 2005). Therefore, crystallization conditions were re-screened using the *HsOGA*_{E602AE605A} variant; crystals were obtained in 17% polyethylene glycol 3350, 0.2 M sodium citrate pH 8.0. This is comparable to the wild-type crystallization condition of 0.1–0.2 M triammonium citrate pH 6.5–7.5, 16–24% PEG 3350.

In a different crystal form, flexible loops can become ordered by making backbone crystal contacts or adopting preferential conformations (Derewenda, 2004a). The space group was *P*3₁21, which is a lower symmetry group compared with *P*4₃2₁2 for the *HsOGA*_{11–396+535–715} structure. The data statistics are presented in Tables 1 and 2.

Theoretically, SER should lead to an improvement in the resolution and hence the overall quality of the structure (Parthasarathy *et al.*, 2008). However, the structure was determined to a resolution of 3.3 Å which was a lower resolution compared with the *HsOGA*_{11–396+535–715} and catalytically inactive mutant structures, possibly owing to the thin-rod crystal morphology compared with the large trigonal bipyramid wild-type crystals.

A solvent channel can be observed through the crystal structure (Fig. 2a). The dimers form a trigonal spring such that the C-terminus of monomer *A* binds into the active-site monomer *A* of a symmetry-related dimer, *etc.* (Figs. 2a and 2b). Owing to the different crystal contacts made on the surface of the protein, regions of high disorder could be built into the structure. 88% of the structure was complete, in comparison with 83% of the *HsOGA*_{11–396+535–715} structure.

Table 2

Structure solution and refinement.

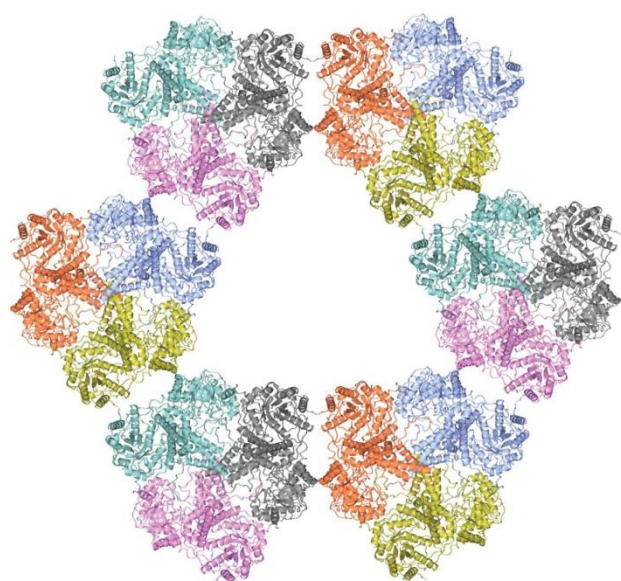
Values in parentheses are for the outer shell.

Resolution range (Å)	192.4–3.3 (3.5–3.3)
No. of reflections	31064
Final <i>R</i> _{cryst}	0.17
Final <i>R</i> _{free}	0.23
Cruickshank DPI	0.40
No. of non-H atoms	
Protein	7877
Water	1
Total	7878
R.m.s. deviations	
Bonds (Å)	0.010
Angles (°)	1.58
Average <i>B</i> factors (Å ²)	
Protein	75
Water	16
Ramachandran plot	
Most favoured (%)	92
Allowed (%)	5

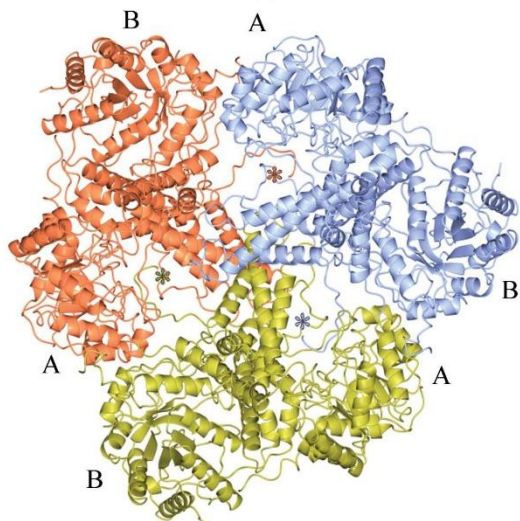
The regions of highest disorder in both monomers were between Lys341 and Thr370, in addition to loops on the protein surface. After multiple rounds of refinement, the confidence for the inclusion of residues Lys341–Asp347 increased, enabling further visualization of the disordered region (Figs. 2c and 2d). The residues that were still disordered in monomers *A* and *B* were Gly11–Gly56, Ser348–Glu369, Lys535 and Pro707–Tyr715 for monomer *B* only. In the protein structures described by Elsen *et al.* (2017) and Li, Li, Lu *et al.* (2017) the residues Lys341–Asp371 in monomer *A* (PDB entry 5uhk) and Asn335–Val372 in monomer *A* (PDB entry 5tke), respectively, were also not observed.

When the N-terminal domains of monomer *A* from *HsOGA*_{E602AE605A} and *HsOGA*_{11–396+535–715} were overlapped, the r.m.s.d of the N-terminal domains was 7.1 Å, indicating a high degree of flexibility in the latter domain (Fig. 2e).

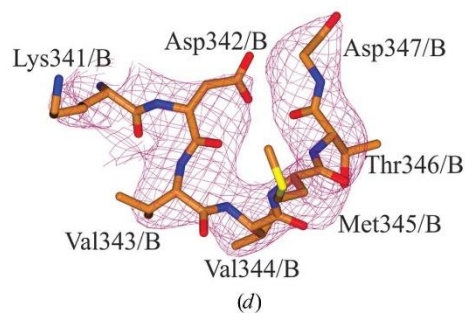
As mentioned, the C-terminus of monomer *A* was found to bind into the active site of a symmetry-related monomer *A*, aiding the formation of the new crystal packing (Figs. 2b and 3a). Initially, the residues of α17 interact with monomer *B* in a domain swap; the residues from Pro694 to Phe703 then bend back towards the residues of the respective monomer, with Gln704–Tyr715 binding into the active site (Figs. 3a and 3b). Electrostatic interactions between the C-terminus of monomer *A* and other residues of monomer *A*, *B* and a symmetry-related molecule *B* stabilize this formation. Pro707 of *HsOGA*_{E602AE605A} has drastically moved position and faces in the opposite direction. There are three consecutive prolines that facilitate the change in direction. Hence, the C-terminus binds to the active site of the symmetry-related monomer *A* rather than monomer *B* (as in the wild type). The C-terminus of monomer *B* could not be built in from Pro707, indicating that it does not bind into an active site because of the crystal packing. In *HsOGA*_{11–396+535–715}, residues Asp696–Pro707 in monomer *A* are missing; therefore, the direction of the peptide is ambiguous. In the natural human sequence, the pseudo-histone acetyltransferase domain is connected to the



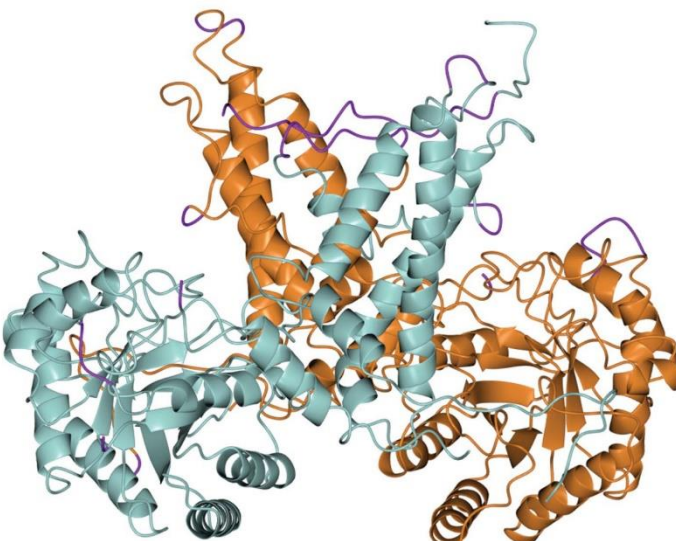
(a)



(b)



(d)



(c)



(e)

stalk domain; therefore, binding of the C-terminus in the active site or alternatively C-termini disorder are possible artefacts of the removal of the HAT domain and of crystal packing.

The C-terminal residues Lys713/A and Tyr715/A hydrogen-bond to and make electrostatic interactions with residues of the active site (Fig. 3*b*). Interestingly, the NZ atom of Lys713 is in the same position as the N2 atom of O-GlcNAc and is 2.97 Å away from the O^δ atom of Asp175/A on the symmetry-related monomer (Fig. 3*d*). This pushes Tyr715 into the +2 subsite, where it interacts with the O^γ atom of Ser652/B of the

Figure 2

(*a, b*) Crystal symmetry of *HsOGA*_{E602AE605A}. (*a*) The connections made between the dimers show the trigonal solvent channel, with 50% solvent as calculated from the Matthews coefficient of 2.51 Å³ Da⁻¹ (Kantardjieff & Rupp, 2003; Matthews, 1968). (*b*) Side view of the repeating trigonal dimers showing the linking C-terminus of monomer A binding into monomer A of the next dimer complex. The monomers are labelled A and B, with asterisks indicating the C-termini. (*c, d*) Disordered loop regions were stabilized in the new crystal form. Monomer A is shown in sea green and monomer B in brown. (*c, d*) The regions in purple were built into the *HsOGA*_{E602AE605A} structure using PDB entry 5m7r as the template model and were missing from the wild-type structure. (*d*) Residues Lys341–Gly347 and the maximum-likelihood/ σ_A -weighted $2F_{\text{obs}} - F_{\text{calc}}$ map shown in red contoured at 0.12 e Å⁻³. (*e*) Overlap of the N-terminal monomers A from chain A for both *HsOGA*_{11–396+535–715} (monomer A, blue; monomer B, gold) and *HsOGA*_{E602AE605A} (monomer A, sea green; monomer B, brown) as calculated by CCP4mg Superpose models. The residue range selected for superposition was Arg59/A–His395/A.

symmetry-related monomer (Fig. 3a). In comparison, the Tyr715/A hydroxyl group of the wild type lies above O6 of O-GlcNAc and hydrogen-bonds to the O^{δ1} atom of Asp285/B (Fig. 3c).

In comparison to the crystal structure of *HsOGA* in complex with α -crystallin B and ELK1 (PDB entries 5vvv and 5vvt, respectively; Li, Li, Lu *et al.*, 2017), the C-terminal residues are in the same reverse direction as the glycosylated

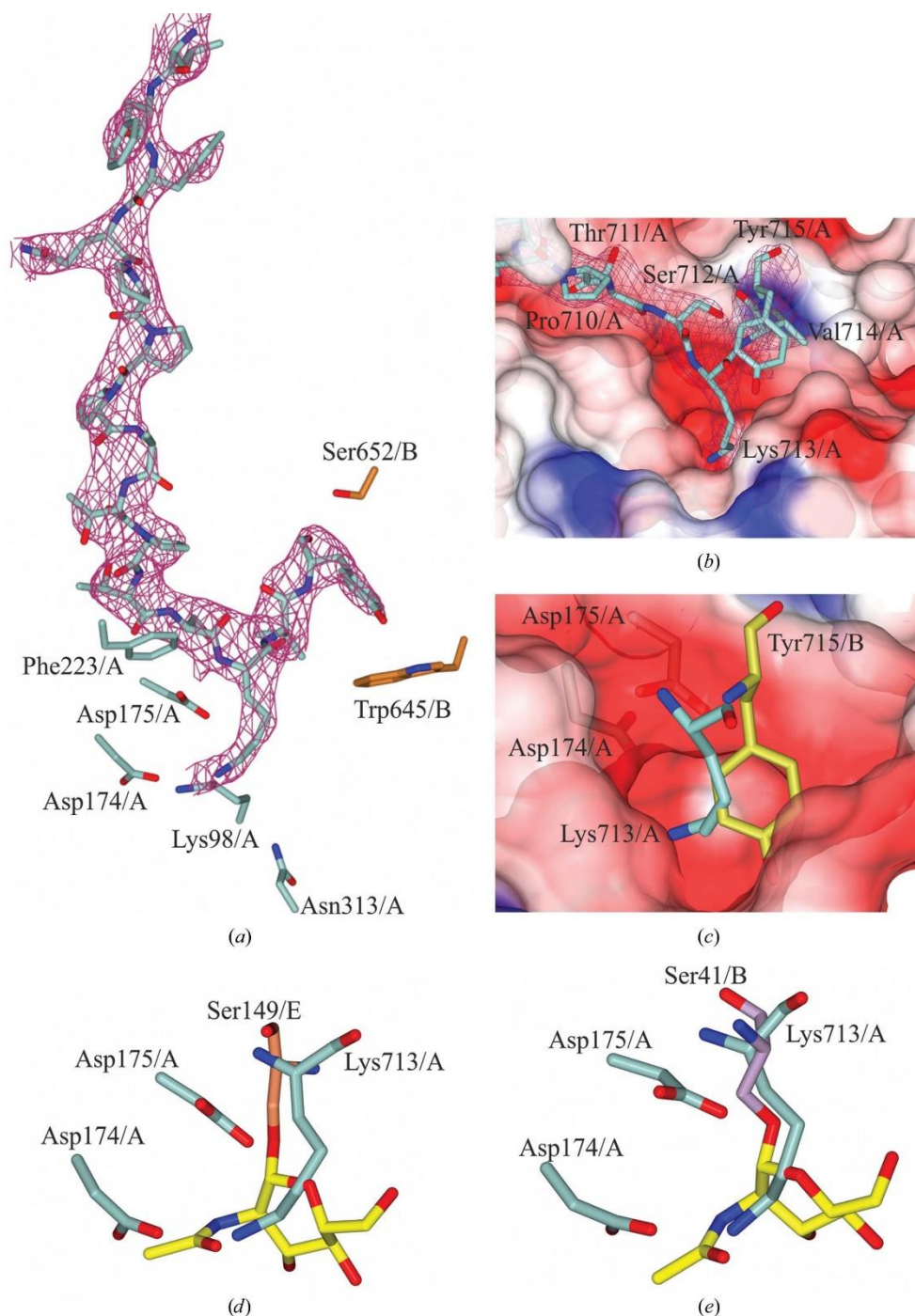


Figure 3

The C-terminus of monomer A binds into the active site of the symmetry-related monomer A. (a) Neighbouring residues in the active site of symmetry-related monomer A of *HsOGA*_{E602AE605A}, with the C-terminal residues bound and the maximum-likelihood/ σ_A -weighted $2F_{obs}-F_{calc}$ map shown in red contoured at $0.17 \text{ e} \text{ \AA}^{-3}$. (b) Surface representation of the active site with the C-terminus bound in a negatively charged pocket. (c) Overlay of Lys713/A from *HsOGA*_{E602AE605A} and Tyr715/B from *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ (in gold) in the binding pocket. (d) Overlay of the C-terminus of chain A of *HsOGA*_{E602AE605A} and that of *HsOGA* in complex with glycosylated p53 peptide (PDB entry 5un8), showing Ser149/E in coral. O-GlcNAc is shown in yellow. (e) Overlay of Lys713/A from *HsOGA*_{E602AE605A} with *HsOGA* in complex with glycosylated α -crystallin B (PDB entry 5vvv), showing Ser41/B in purple.

Table 3
Kinetics analysis comparing different *HsOGA* constructs.

Construct	<i>HsOGA</i> _{FL}	<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₇₁₅	<i>HsOGA</i> _{E602AE605A}
V_{\max} ($\mu\text{M min}^{-1}$)	2.96 ± 0.07	2.50 ± 0.06	4.62 ± 0.15
K_m (μM)	298 ± 15	227 ± 13	230 ± 20
k_{cat} (min^{-1})	59.1 ± 1.4	49.9 ± 1.2	92.5 ± 3.1
k_{cat}/K_m ($\text{min}^{-1} \mu\text{M}^{-1}$)	0.198 ± 0.012	0.219 ± 0.013	0.401 ± 0.038

peptides (Fig. 3e). This is in contrast to the structure containing a glycosylated p53 peptide shown in Fig. 3(d).

The density for all available *HsOGA* peptide-complex structures supports the notion that OGA can bind peptides in both the forward and reverse directions. Comparison between the different peptide structures shows the versatility of the active site for different peptides.

3.2. Comparison of the constructs and mutants

To ensure that the structural stabilization had not occurred owing to a change in the secondary structure of *HsOGA*_{E602AE605A} and that the mutation had not affected the activity of the enzyme, the full-length protein (*HsOGA*_{FL}),

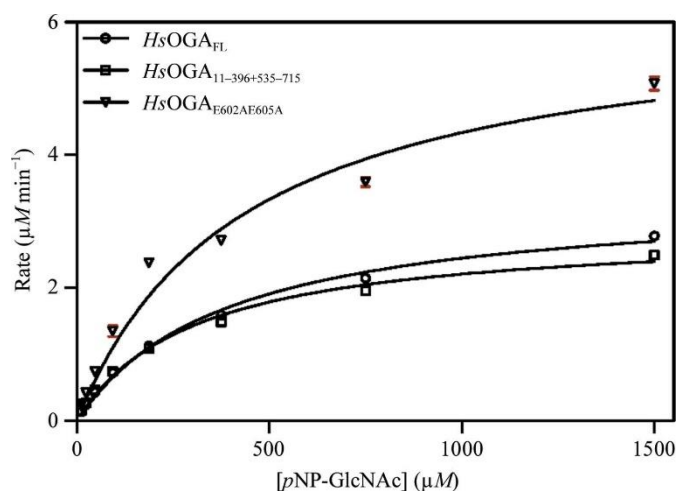


Figure 4
Michaelis–Menten curves for the kinetic assay of the *HsOGA* mutant. pNP-GlcNAc was used as the substrate at concentrations up to seven times higher than the K_m .

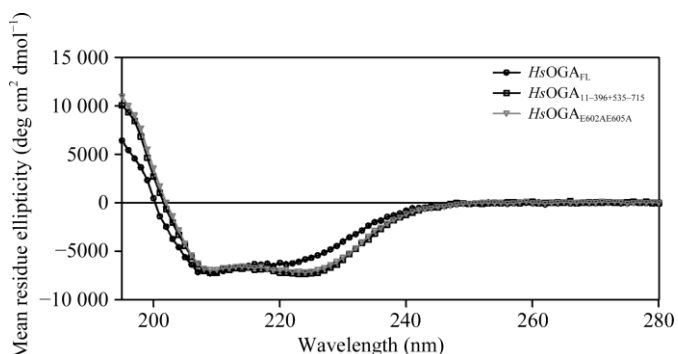


Figure 5
Circular dichroism spectra for the different constructs. The experimental data in millidegree units were converted to mean residue ellipticity values with units of $\text{deg cm}^2 \text{dmol}^{-1}$ using the equation given in Ishtikhari *et al.* (2014).

*HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ and *HsOGA*_{E602AE605A} underwent kinetic and secondary-structure analyses.

Comparison of the results for *HsOGA*_{FL} and *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ shows that the split construct has similar activity to the full-length variant, suggesting that co-expression of the two domains does not affect the activity (Fig. 4). Therefore, the ligand-binding data should be an accurate representation of binding to the full-length protein. When comparing *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ and *HsOGA*_{E602AE605A}, the values are very similar. The V_{\max} is higher, indicating an increase in the reaction rate (Table 3). The increase in V_{\max} may be owing to a discrepancy in the enzyme concentration ($[E]$), as V_0 is directly proportional to $[E]$. The similarity of V_{\max} and K_m for the mutant and the wild type suggest that the mutation did not alter the substrate-binding affinity or the enzyme activity of the protein for its substrate.

The CD spectra show that all of the constructs are fully folded and the spectra for *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ and *HsOGA*_{E602AE605A} are very similar, suggesting that there is no change in the composition of the mutant (Fig. 5). The spectral profile of *HsOGA*_{FL} differs from those of the split constructs in that it has a less pronounced minimum in the 222 nm region and an overall blue-shifted spectrum in the <210 nm region. This suggests that there could be a lower proportion of α -helical structure and/or higher structural disorder. This links to the inclusion of the pseudo-HAT domain that is connected to the C-terminal stalk domain. Since the structure of the human homologue of this domain is unknown owing to the inability of *HsOGA*_{FL} to crystallize, the structure can only be inferred from the structural homologues from *Oceanicola granulosus* and *Streptomyces sviveus* (He *et al.*, 2014; Rao *et al.*, 2013). An estimate of the secondary-structure content of the proteins suggests a decrease of 6.4% in α -helical components and an increase of 3.7% in β -sheet components in *HsOGA*_{FL} compared with *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ (Supplementary Table S1). Homology modelling using crystal structures of the HAT domain from the bacterial homologue *O. granulosus* suggests a similar overall structure minus the acetyltransferase activity (Rao *et al.*, 2013). However, the data could be skewed by the inclusion of the linker region to this domain and the potential difference in the homology model structure. The secondary-structure contents of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ and *HsOGA*_{E602AE605A} are consistent, further confirming that the mutation did not affect the overall structure.

4. Conclusions

In this study, surface-entropy reduction has been utilized to produce further structural information on O-GlcNAcase by the incorporation of residues Ala57–Arg58, Lys341–Asp347, Thr370, Glu536, Cys596–Gly598, Gly674–Asp675 and Asp696–Pro707, an increase in the number of observed residues of 5%. Although the binding of the C-terminus to the active site may be an artefact of crystallization, it reveals further details regarding the substrate specificity of OGA, as peptides have been shown to bind in a bidirectional yet conserved conformation. The results described in this study

present an opportunity for further investigation of the binding orientation of peptides within an SER-modified OGA enzyme. Given the progression of hOGA inhibitors into clinical trials, different surface mutants of the enzyme may afford new routes to drug development.

Acknowledgements

CD data collection was conducted by Dustin King and assistance in conducting the kinetics experiments and analysing the data was provided by Matthew Alteen, both of whom are employed by Professor David Vocadlo at Simon Fraser University. We thank Diamond Light Source for access to beamline I04-1 (proposal No. mx-13587) that contributed to the results presented here, and Johan Turkenburg and Sam Hart for coordinating data collection.

Funding information

This research was funded by the Biotechnology and Biological Sciences Research Council (grant BB/M011151/1) for the support of AM. GJD is the Royal Society Ken Murray Research Professor.

References

- Arnold, C. S., Johnson, G. V. W., Cole, R. N., Dong, D. L.-Y., Lee, M. & Hart, G. W. (1996). *J. Biol. Chem.* 271, 28741–28744.
- Avbelj, F. & Fele, L. (1998). *J. Mol. Biol.* 279, 665–684.
- Comtesse, N., Maldener, E. & Meese, E. (2001). *Biochem. Biophys. Res. Commun.* 283, 634–640.
- Cooper, D. R., Boczek, T., Grelewsk, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Acta Cryst.* D63, 636–645.
- Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocadlo, D. J. & Davies, G. J. (2006). *Nature Struct. Mol. Biol.* 13, 365–371. Derewenda, Z. S. (2004a). *Structure*, 12, 529–535. Derewenda, Z. S. (2004b). *Methods*, 34, 354–363.
- Derewenda, Z. S. & Vekilov, P. G. (2006). *Acta Cryst.* D62, 116–124.
- Dong, D. L.-Y. & Hart, G. W. (1994). *J. Biol. Chem.* 269, 19321–19330.
- Elsen, N. L., Patel, S. B., Ford, R. E., Hall, D. L., Hess, F., Kandula, H., Kornienko, M., Reid, J., Selnick, H., Shipman, J. M., Sharma, S., Lumb, K. J., Soisson, S. M. & Klein, D. J. (2017). *Nature Chem. Biol.* 13, 613–615.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D66, 486–501.
- Evans, P. (2006). *Acta Cryst.* D62, 72–82.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D69, 1204–1214.
- Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. (2001). *J. Biol. Chem.* 276, 9838–9845.
- Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. (2007). *Protein Sci.* 16, 1569–1576.
- Griffith, L. S. & Schmitz, B. (1995). *Biochem. Biophys. Res. Commun.* 213, 424–431.
- Grundke-Iqbal, I., Iqbal, K., Tung, Y.-C., Quinlan, M., Wisniewski, H. M. & Binder, L. I. (1986). *Proc. Natl Acad. Sci. USA*, 83, 4913–4917.
- Hart, G. W., Housley, M. P. & Slawson, C. (2007). *Nature (London)*, 446, 1017–1022.
- He, Y., Roth, C., Turkenburg, J. P. & Davies, G. J. (2014). *Acta Cryst.* D70, 186–195.
- Holt, G. D. & Hart, G. W. (1986). *J. Biol. Chem.* 261, 8049–8057.
- Ishtikhari, M., Khan, S., Badr, G., Osama Mohamed, A. & Hasan Khan, R. (2014). *Mol. Biosyst.* 10, 2954–2964.
- Kantardjiev, K. A. & Rupp, B. (2003). *Protein Sci.* 12, 1865–1871.
- Kim, A.-R., Dobransky, T., Rylett, R. J. & Shilton, B. H. (2005). *Acta Cryst.* D61, 1306–1310.
- Kreppel, L. K., Blomberg, M. A. & Hart, G. W. (1997). *J. Biol. Chem.* 272, 9308–9315.
- Li, B., Li, H., Hu, C.-W. & Jiang, J. (2017). *Nature Commun.* 8, 666.
- Li, B., Li, H., Lu, L. & Jiang, J. (2017). *Nature Struct. Mol. Biol.* 24, 362–369.
- Li, J., Li, Z., Li, T., Lin, L., Zhang, Y., Guo, L., Xu, Y., Zhao, W. & Wang, P. (2012). *Biochemistry (Mosc.)*, 77, 194–200.
- Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C.-X. (2004). *Proc. Natl Acad. Sci. USA*, 101, 10804–10809.
- Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst.* D57, 679–688.
- Lubas, W. A., Frank, D. W., Krause, M. & Hanover, J. A. (1997). *J. Biol. Chem.* 272, 9316–9324.
- Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocadlo, D. J. (2005). *J. Biol. Chem.* 280, 25313–25322.
- Mariappa, D., Selvan, N., Borodkin, V. S., Alonso, J., Ferenbach, A. T., Shepherd, C., Navratilova, I. H. & van Aalten, D. M. F. (2015). *Biochem. J.* 470, 255–262.
- Marotta, N. P., Lin, Y. H., Lewis, Y. E., Ambroso, M. R., Zaro, B. W., Roth, M. T., Arnold, D. B., Langen, R. & Pratt, M. R. (2015). *Nature Chem.* 7, 913–920.
- Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst.* D58, 1983–1991.
- Matthews, B. W. (1968). *J. Mol. Biol.* 33, 491–497.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D67, 355–367.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D53, 240–255.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D61, 1426–1431.
- Nicholls, R. A., Long, F. & Murshudov, G. N. (2012). *Acta Cryst.* D68, 404–417.
- Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D54, 1285–1294.
- Parthasarathy, G., Cummings, R., Becker, J. W. & Soisson, S. M. (2008). *Acta Cryst.* D64, 141–148.
- Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M. & van Aalten, D. M. F. (2006). *EMBO J.* 25, 1569–1578.
- Rao, F. V., Schüttelkopf, A. W., Dorfmueller, H. C., Ferenbach, A. T., Navratilova, I. & van Aalten, D. M. F. (2013). *Open Biol.* 3, 130021.
- Roth, C., Chan, S., Offen, W. A., Hemsworth, G. R., Willems, L. I., King, D. T., Varghese, V., Britton, R., Vocadlo, D. J. & Davies, G. J. (2017). *Nature Chem. Biol.* 13, 610–612.
- Schimpl, M., Borodkin, V. S., Gray, L. J. & van Aalten, D. M. F. (2012). *Chem. Biol.* 19, 173–178.
- Schimpl, M., Schüttelkopf, A. W., Borodkin, V. S. & van Aalten, D. M. F. (2010). *Biochem. J.* 432, 1–7.
- Shen, D. L., Gloster, T. M., Yuzwa, S. A. & Vocadlo, D. J. (2012). *J. Biol. Chem.* 287, 15395–15408.
- Slawson, C. & Hart, G. W. (2011). *Nature Rev. Cancer*, 11, 678–684.
- Toleman, C., Paterson, A. J., Whisenhunt, T. R. & Kudlow, J. E. (2004). *J. Biol. Chem.* 279, 53665–53673.
- Torres, C. R. & Hart, G. W. (1984). *J. Biol. Chem.* 259, 3308–3317.
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* D60, 2184–2195.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D66, 22–25.
- Vekilov, P. G. (2003). *Methods Enzymol.* 368, 84–105.
- Vekilov, P. G., Feeling-Taylor, A., Yau, S.-T. & Petsev, D. (2002). *Acta Cryst.* D58, 1611–1616.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W.,

- McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst. D* 67, 235–242.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* 374, 300–321.
- Winter, G. (2010). *J. Appl. Cryst.* 43, 186–190.
- Yuzwa, S. A., Macauley, M. S., Heinonen, J. E., Shan, X., Dennis, R. J., He, Y., Whitworth, G. E., Stubbs, K. A., McEachern, E. J., Davies, G. J. & Vocadlo, D. J. (2008). *Nature Chem. Biol.* 4, 483–490.
- Yuzwa, S. A., Shan, X., Jones, B. A., Zhao, G., Woodward, M. L., Li, X., Zhu, Y., McEachern, E. J., Silverman, M. A., Watson, N. V., Gong, C.-X. & Vocadlo, D. J. (2014). *Mol. Neurodegener.* 9, 42

**Residue Ser364 of *Hs*OGA is Phosphorylated by Casein
Kinase II**

Declaration

The LC-MS/MS data was collected with the help of Dr. Richard Spears. I conducted the kinetics experiments in the laboratory of Prof. David Vocadlo, Simon Fraser University, with the assistance of Matt Alteen. The circular dichroism experiment was conducted by Dr. Dustin King, Simon Fraser University.

4.7 Introduction to phosphomimetics

11 sites are postulated to be phosphorylated on the full length of *HsOGA* (Hornbeck et al., 2015, 2004). Of these, five are present within the *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ construct (11-396+535-715 as described in Roth): S268, S364, T370, T709 and S712. Phosphorylation of Ser364 was observed in 51 separate HTP experiments including mass phosphoproteomic analysis of human tumour and cancer cells (Carrier et al., 2016; Klammer et al., 2012) and mitotic cells (Dephoure et al., 2008); the high frequency could be indicative of its importance.

Mimicking the phosphorylation of a residue can be achieved through mutation of the residue to an aspartate or glutamate, Figure 4.11 (Thorsness and Koshland, 1987). It is hypothesised that Ser/Thr/Tyr sites of dynamic phosphorylation evolved from aspartate and glutamate residues (Pearlman et al., 2011). Both phosphate (PO_4^{3-}) and aspartate are negatively charged, however, aspartate has a single charge (-1) and PO_4^{3-} has a charge of -1.5. To compensate for this, the environment around the aspartate is expected to contribute to the negative charge. Furthermore, PO_4 has a larger ionic shell than aspartate, since it can form three covalent bonds owing to ten lone pair electrons.

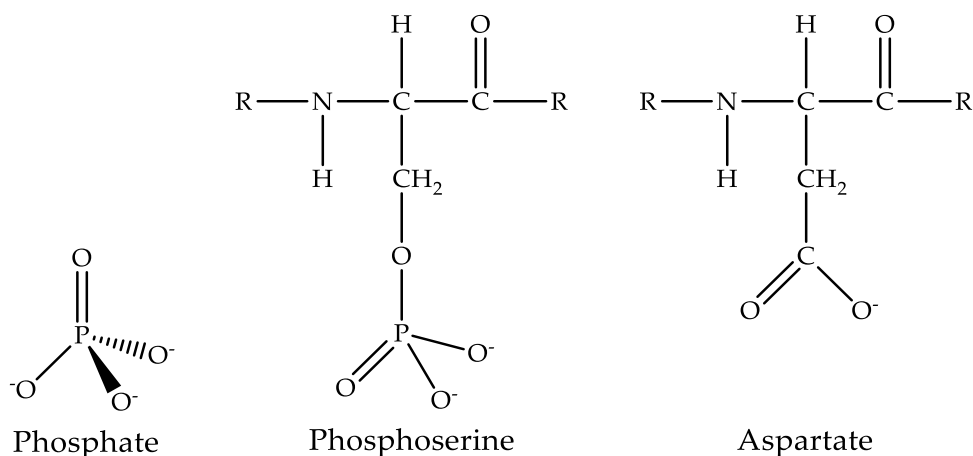


Figure 4.11. A phosphorylated serine residue is mimicked by aspartate.

LC-MS/MS of *HsOGA* highlighted Ser364 as being phosphorylated by CKII by a shift of 79.98 Da (average mass of PO_4^{2-}). The residue is in a disordered region that extends from Asn337 to Asp371 and cannot be modelled in crystal structures, possibly due to its flexibility. To mimic the modification and possibly stabilise the disordered region, Ser364 was mutated to an aspartate. Circular dichroism, kinetic assays and X-ray crystallography revealed no difference between the phosphomimetic and WT protein.

4.8 Experimental methodology

4.8.1 Trypsin digest of phosphorylated *HsOGA* and LC-MS/MS detection

78.1 μM of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ was incubated with CKII in 1x NEBuffer for protein kinases and 200 μM ATP (added from 1x T4 DNA ligase buffer) for 19 hours at 25 °C. The phosphorylated protein was digested with trypsin using the following protocol. 100 μL of buffer 1 (6 M urea, 50 mM Tris-HCl pH 8.0) and 5 μL of buffer 2 (200 mM DTT, 50 mM Tris-HCl pH 8.0) were incubated with the protein fragments for 1 hour at RT. 20 μL of buffer 3 (200 mM iodoacetamide, 50 mM Tris-HCl pH 8.0) was added and the mixture was incubated for 1 hour in the dark. 20 μL of buffer 2 was added and the mixture was incubated for 1 hour. 775 μL of buffer 4 (50 mM Tris-HCl pH 7.6, 1 mM CaCl_2) and 1 vial of Promega trypsin (sequencing grade modified, porcine) were added and the mixture was gently vortexed and incubated at 37 °C for 20 hours. The pH was adjusted to pH 3.0 using formic acid. The protein fragments were buffer exchanged into 50 mM Tris-HCl using a Viva spin protein concentrator spin column. The results analysed using Bruker Daltonics DataAnalysis programme. The fragments of interest were identified using ExPASy peptide mass tool (Gasteiger et al., 2005; Wilkins et al., 1997), Phosphosite (Hornbeck et al., 2015) and a fragment ion calculator.

4.8.2 Circular dichroism of *HsOGA* constructs

The protocol followed for this experiment was the same as the protocol stated in Males and Davies, 2019.

4.8.3 Kinetic assay of *HsOGA* constructs and *pNP*-GlcNAc

Michaelis-Menten kinetics of *HsOGA*_{S364D} was assayed against the positive control, *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅. The protocol followed for this experiment was the same as the protocol stated in Males and Davies, 2019. The reaction of *pNP*-GlcNAc hydrolysis was monitored at 405 nm, the wavelength for the cleavage product 4-nitrophenol, Figure 4.12.

Table 4.1. Trypsin digested OGA fragments containing potential phosphorylation sites.

Fragment	Residue range	Sequence	Site of phosphorylation
1	109-127	EMY S VEEAEQLMTLISAAR	Ser112
2	258-269	EIPVESIEEV S K	Ser268
3	359-381	LENE S DE D IED I TDVLYSPQMALK	Ser364 / Thr370
4	692-713	LLPIDGANDLFFQPPPL T P S K	Thr709 / Ser712

Analysis of the UV spectra when comparing the unphosphorylated and phosphorylated samples showed that there was only one fragment that had a shift in mobility indicating phosphorylation, Figure 4.13, see appendix for UV traces and masses of the other fragments. Fragment 3, LENE**S**DE**D**IED**I**TDVLYSPQMALK, is predicted to contain two sites of phosphorylation, Ser364 and Thr370. There is also a small portion of unphosphorylated OGA in the phosphorylated sample, this may be due to the activity of CKII or the accessibility of the site for phosphorylation.

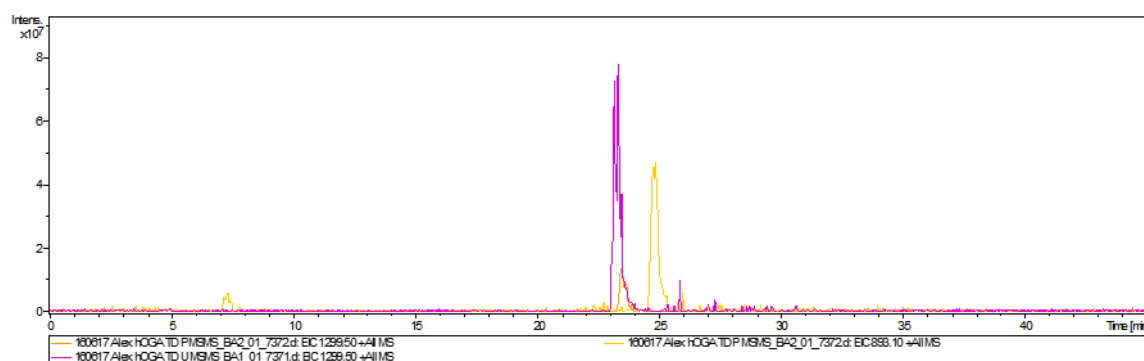


Figure 4.13. Comparison of the relative peak heights of the phosphorylated and unphosphorylated OGA fragment 3. An overlay of three peaks: pink peak is the unphosphorylated fragment; yellow peak is the phosphorylated fragment and the orange peak is the portion of the unphosphorylated fragment in the phosphorylated sample.

The highlighted peak for the unphosphorylated sample was 1299 Da ($[M+H]^{2+}$ a double charge) and the highlighted peak for the phosphorylated sample was 893 Da ($[M+H]^{3+}$ a triple charge), Table 4.2. A second peak was not observed at a mass indicative of having two phosphates attached. To decipher which one of the sites was phosphorylated, the chromatogram of the mass spectrometry experiment was analysed.

Table 4.2. Fragment 3 mass/charge table for phosphorylation modifications.

	$[M+H]^{1+}$	$[M+H]^{2+}$	$[M+H]^{3+}$
Unphosphorylated	2591.2	1299.1	864.7
Phosphorylated +1	2677.2	1339.1	893.3
Phosphorylated +2	2757.2	1379.1	920.1

MS/MS was performed on the peak containing one phosphorylated site (observed at 893 Da). The main peak of 1339.45 Da (corresponding to a double charge) was subjected to fragmentation. A comparison of the peaks from the MS/MS fragmentation showed a vast number of the calculated B ions were present, Figure 4.14. This aided the analysis of site identification. To identify the phosphorylation site, Ser364 or Thr370, the peak corresponding to a phosphorylated Ser364 B ion was searched for. This would correspond to a shift from 630.3 Da to 710.3 Da, Table 4.3. If it was not present at 710.3 Da, then the increase of 80 Da would be seen from 1332.5 to 1412.5 Da.

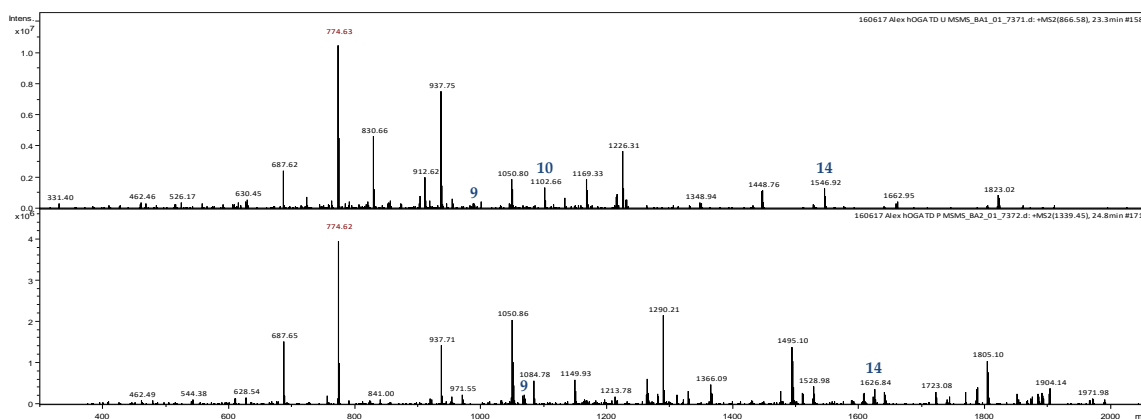


Figure 4.14. A comparison of the ions present after fragmentation of unphosphorylated and phosphorylated samples. MS/MS of the, top: unphosphorylated fragment at 866 Da ($[M+Z]^{3+}$), and bottom: phosphorylated fragment at 1339 Da ($[M+Z]^{2+}$). The blue numbers correspond to the B ion type masses in Table 4.3.

Table 4.3. Calculated B ion type masses. Highlighted in grey are the sites of phosphorylation and mass changes.

Seq	#	Unphosphorylated	S364 Phosphorylated	T370 Phosphorylated
G	5	543.24	543.24	543.24
S	6	630.27	710.27	630.27
D	7	745.30	825.30	745.30
E	8	874.34	954.34	874.34
D	9	989.37	1069.37	989.37
I	10	1102.45	1182.45	1102.45
E	11	1231.50	1311.50	1231.50
T	12	1332.54	1412.54	1412.54
D	13	1447.57	1527.57	1527.57
V	14	1546.64	1626.64	1626.64

The shift to 710.3 Da was observed and the other peaks past this point were also present with an 80 Da increase, Figure 4.14. In conclusion, the observed peaks correspond to the calculated masses for phosphorylation of Ser364 by CKII. To ascertain whether the other predicted phosphorylation sites are phosphorylated by a different kinase, an LC-MS/MS experiment was also conducted on *HsOGA* phosphorylated by CDK1 after digestion with trypsin.

4.9.2 The phosphomimetic mutation of *HsOGA* did not alter the secondary structure of the protein.

*HsOGA*_{S364D} has a similar overall fold to that of *HsOGA*_{split1} as can be seen in the circular dichroism trace, Figure 4.15a. It has two regions of minima at 210 nm and 225 nm indicating α -helical nature, Figure 4.15b. The protein has an overall α -helical composition with the predicted percentage of α -helices being 16% and of β -strands is 35%.

The protein was screened against a variety of different conditions to find a different crystal form that would potentially have higher resolution or where Asn337 to Asp371 would be stabilised and visible. Although initially crystals were obtained in well E8 of the JCSG screen, the crystals diffracted poorly and they could not be reproduced. Better crystals were obtained in the same condition as *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅.

Processed diffraction data revealed no change in the conformation of the disordered region between Asn337 and Asp371, since the residues between Asn337 and Asp371 in chain A and Arg340 and Asp371 in chain B could not be built in, Figure 4.15c and d. The space group was also the same as the WT ($P 4_3 2_1 2$) unlike the different packing seen with the surface entropy reduction mutant *HsOGA*_{E602AE605A}, Table 4.4.

Table 4.4. Data collection and refinement statistics for *HsOGA*_{S364D}. Values in parentheses are for high resolution data.

<i>HsOGA</i> _{S364D}	
Data collection	
Space group	$P 4_3 2_1 2$
Cell dimensions: <i>a, b, c</i> (Å)	101.7, 101.7, 285.3
γ (°)	90
Resolution (Å)	95.8-2.6 (2.6-2.6)

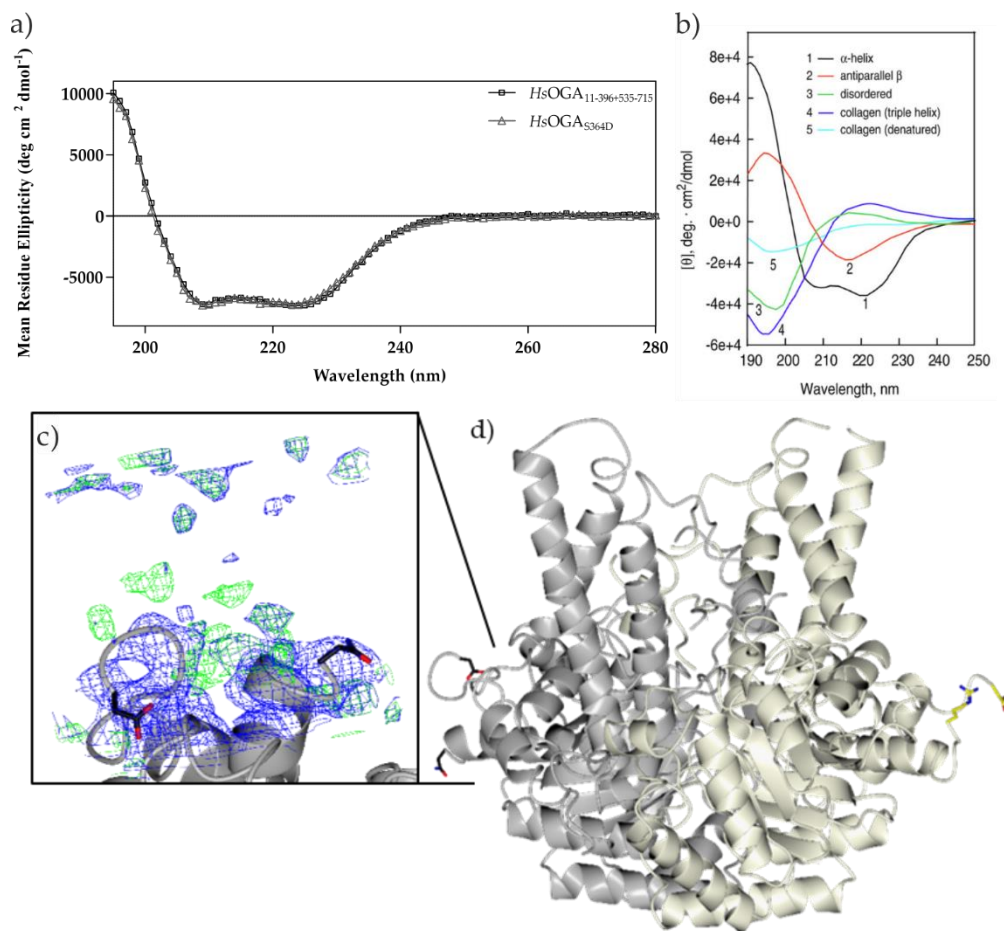


Figure 4.15. The secondary structure of the *HsOGA* phosphomimetic mutant was similar to the WT. a) The trace line of *HsOGA*_{split1} in black and data points are represented by black squares and *HsOGA*_{S364D} is in grey with triangles. b) CD spectra of secondary structural elements. Taken from Greenfield, 2006. c) Density between Asn337 and Asp371 with the maximum-likelihood/ σ_A -weighted $2F_{\text{obs}}-F_{\text{calc}}$ map, shown in blue, contoured at 0.1 electrons/ \AA^3 and the $F_{\text{obs}}-F_{\text{calc}}$ map, shown in green, at 0.12 electrons/ \AA^3 . d) *HsOGA*_{S364D} represented by grey ribbon with Asn337 and Asp371 from chain A and Arg340 and Asp371 from chain B highlighted in black and yellow cylinders, respectively.

4.9.3 *HsOGA*_{S364D} mutant has the same kinetic activity as WT *HsOGA*_{split1}.

Michaelis-Menten (MM) kinetics was used to investigate the effect of the activity of the mutant *HsOGA*_{S364D} on *p*NP-GlcNAc compared to the WT *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅. A change in activity may indicate an alteration of the secondary structure of the protein or detail the importance of the loop to be closer to its natural state.

The comparison of V_{\max} and K_M values for $HsOGA_{\text{split1}}$ versus $HsOGA_{S364D}$ are very similar, Table 4.5 and Figure 4.16. However, there is a 3-fold decrease in the k_{cat} value. This could be interpreted as an increase in the length of time in which the substrate is staying in the active site. However, since it is only a 3-fold decrease it is classed as not significant, therefore, the results suggest that there is no difference between the catalytic activities.

Table 4.5. MM kinetic assay results for $HsOGA_{11-396+535-715}$ and $HsOGA_{S364D}$.

Rates	$HsOGA_{11-396+535-715}$	$HsOGA_{S364D}$
V_{\max} ($\mu\text{M min}^{-1}$)	2.50 ± 0.06	2.28 ± 0.11
K_M (μM)	227.4 ± 12.8	212.9 ± 24.5
k_{cat} (min^{-1})	49.9 ± 1.2	16.4 ± 0.8
k_{cat}/K_M ($\text{min}^{-1} \mu\text{M}^{-1}$)	0.219 ± 0.0134	$0.077 \pm$

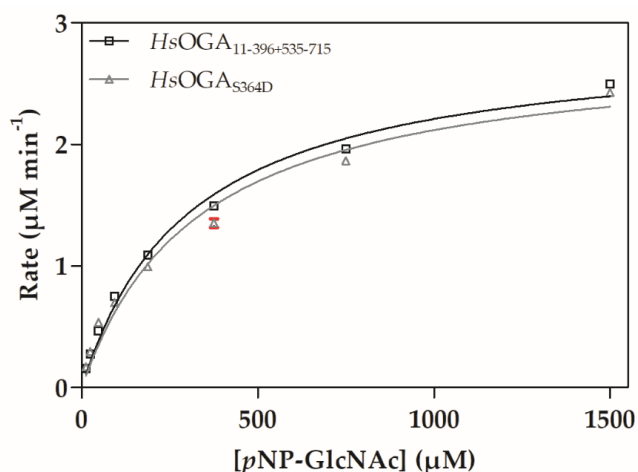


Figure 4.16. Michaelis-Menten curves for HsOGA mutant kinetic assays. $p\text{NP-GlcNAc}$ as the substrate up to 7x the K_M . $HsOGA_{11-396+535-715}$.

4.10 Discussion

Ser364 was identified to be phosphorylated by casein kinase II. However, using mass spectrometry to detect phosphorylation has its limitations. A low signal:noise is due to the lability of the phosphate ion (Jagannadham and Nagaraj, 2008); reviewed in (Dephoure et al., 2013; McLachlin and Chait, 2001; Ubersax and Ferrell, 2007). The mass change upon phosphorylation could not be observed in the fragment ions less than 400 Da (since the peptides are difficult to detect by MS) which was particularly important for MS/MS analysis of fragment 4 from the stalk domain.

MS analysis indicates that the stalk domain of OGA is not phosphorylated by CKII. This work contradicts evidence of phosphorylation of Thr709 when the cell is undergoing mitosis (Dephoure et al., 2008). A cyclin dependent kinase, cell cycle kinase subunit cdc2, was used to phosphorylate proteins in the unfractionated lysate from HeLa cells. However, the kinase used in this experiment was CKII and the consensus sequences for the two kinases are not compatible. Cdc2 kinase recognises X-S-P-X-K and CKII recognises S-X-X-E/D/Yp/Dp (Marin et al., 1992). Residue Ser364 (fragment 3), is a part of the consensus sequence for CKII of S-D-E-D. On the contrary, residue Thr709 (fragment 4), is not part the CKII consensus sequence and is instead a substrate for cdc2 L-T-P-T-S-K. This could be the reason that phosphorylation of Thr709 was not identified.

The activity of *HsOGA*_{S364D} on *pNP-GlcNAc* was not altered in comparison to *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅. Ser364 is not close to the active site and may be phosphorylated for regulation at certain points in the mitotic cell cycle. It may be useful in protein signalling and protein-protein interactions or for homodimeric interactions. Since the charge and ionic shell of a phosphorylation modification is poorly mimicked by aspartate it may have been appropriate to use protein phosphorylated by CKII. A hydrogen bond between S364D and another residue may shield effect of the negative charge. Futhermore, mutation of the 11 potential sites of phosphorylation to aspartates might result in a greater difference in activity. Also observed in the disordered region between Asn337 to Asp371 was the sumoylation of Lys358 (Lumpkin et al., 2017).

Although the results collected in this chapter did not directly progress the knowledge of OGA in a successful manner, it concludes that there is still much to understand in terms of the bidirectional nature of protein-substrate binding and how PTMs affect the regulation of OGA. Not only do PTMs affect the regulation of OGA, but O-GlcNAcylation of proteins by OGA can act as a regulatory mechanism and when this is disrupted diseases can occur. One such disease is the neurodegenerative disorder, Alzheimer's disease. In the next chapter of this thesis, I will discuss the development of inhibitors against OGA and present structures of novel inhibitor classes.

CHAPTER 5: Inhibition of O-GlcNAc Hydrolase

5.1 Abstract

O-GlcNAc hydrolase (OGA) cleaves the Ser-/Thr-linked O-GlcNAc moiety from a variety of different sequences on thousands of different proteins. OGA has a dynamic relationship with O-GlcNAc transferase, which adds the O-GlcNAc moiety to proteins, and inhibition of OGA can shift the equilibrium towards a global increase in GlcNAcylation levels and autophagy. Increased levels of O-GlcNAc have been shown to reduce the levels of pathogenic tau and remove pathogenic tau species. Currently, there are two inhibitors that have passed through the first stage of clinical trials, with both showing promising results against progressive supranuclear palsy. However, it is still important to develop new first-in-class inhibitors. Pyrrolidine-based iminocyclitols, bicyclic inhibitors and NAG-thiazoline derivatives (amongst other inhibitor classes) have been proven to be able to cross the blood-brain barrier, have high potency and increase the cellular levels of O-GlcNAc. A series of pyrrolidine-based triazole compounds have been developed from a library of ~30 molecules that have picomolar inhibition constant to OGA. Pi stacking interactions with Tyr69, Trp645 and Trp679 constrain two conformations of the triazole moiety of the inhibitor in the active site. It is expected that further cell-based assays and selectivity tests will reveal this series of triazole inhibitors to be auspicious in treating a tau-based neurological disease.

5.2 Family GH84 inhibition introduction

5.2.1 The importance of OGA inhibition

Inhibition of OGA can not only help to understand its catalytic mechanism but also be used to treat certain diseases. OGA inhibition has been shown to slow the progression of Alzheimer's disease in patients. One of the many factors contributing to neurodegenerative disorders are tauopathies (the aggregation of tau into pathogenic filaments).

5.2.1.1 O-GlcNAcylation and phosphorylation of tau

Originally discovered in porcine brain extracts, Tau (*MAPT*) is a highly soluble microtubule-associated protein (Weingarten et al., 1975). It interacts with hydrophobic pockets of tubulin heterodimers (Kadavath et al., 2015a) via proline rich regions (highly

conserved PGGG motifs which have a hairpin-like structure (Kadavath et al., 2015b)) in the C-terminal microtubule binding repeats (MTBRs) to promote and stabilise assembly into microtubules (MTs) (Butner, 1991; Gustke et al., 1994).

The multisite phosphorylation of tau was first discovered by the addition of cyclic AMP dependent protein kinases and casein kinase I to rat brains *in vitro* (Pierre and Nunez, 1983). >65 phosphorylation sites have since been identified on Tau441; the phosphorylation of which are catalysed by various kinases and phosphatases, Figure 5.1. O-GlcNAcylation of tau was first detected from bovine brain samples and then confirmed to be present in human brain samples (Arnold et al., 1996; Liu et al., 2004). Both detection by blotting using wheat germ agglutinin and observation of the incorporation of [³H]galactose revealed there to be around four O-GlcNAc molecules per tau molecule. Tau has eight sites of O-GlcNAcylation: Thr123, Ser208, Ser328, S356, Ser400 and Ser409/Ser412/Ser413 detected *in vivo* in mice and rats and *in vitro* by NMR studies, Figure 5.1 (Wang et al., 2010; Yuzwa et al., 2011, 2012).

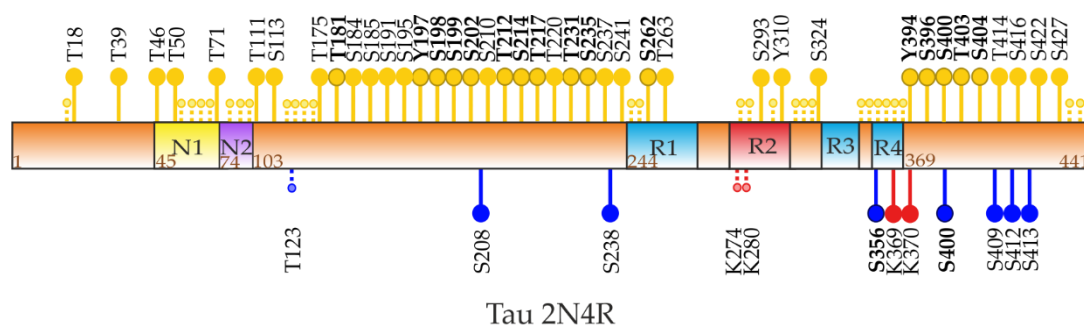


Figure 5.1. Tau undergoes phosphorylation and O-GlcNAcylation. The residues that have been observed to have PTMs are shown by the following colour scheme: yellow for phosphorylation, blue for O-GlcNAcylation and red for acetylation. Larger indicators represent PTM sites that have >5 references or have been identified by low throughput methods and deposited in the database PhosphoSitePlus. PTM sites that have more than 40 corresponding references are shown in bold.

5.2.1.2 Involvement of crosstalk between O-GlcNAcylation and phosphorylation in Alzheimer's disease

Studies have shown a reciprocal relationship between phosphorylation and glycosylation (Griffith and Schmitz, 1995; Liu et al., 2004). When OGA hydrolyses O-GlcNAc, the same

site (or an adjacent site) can subsequently be phosphorylated by a protein kinase. It was shown that when OGA was inhibited the levels of phosphorylation of tau in PC12 cells was decreased, namely to Ser199, Ser202, Thr205, Thr212, Ser214, Ser262 and Ser396 (Liu et al., 2004). When Ser400 is O-GlcNAcylated, it blocks phosphorylation not only of Ser400 but also of Ser396 (Smet-Nocca et al., 2011).

In healthy brain tissue, each molecule of tau has ~1.9 phosphates. However, in patients with AD, tau undergoes hyperphosphorylation and has 3-8 phosphates per molecule of tau (Ksiezak-Reding et al., 1992). Hyperphosphorylation of tau causes it to dissociate, compact and aggregate (Ksiezak-Reding et al., 1992) into paired helical filaments (PHF) and neurofibrillary tangles (NFT), Figure 5.2 (Grundke-Iqbal et al., 1986).

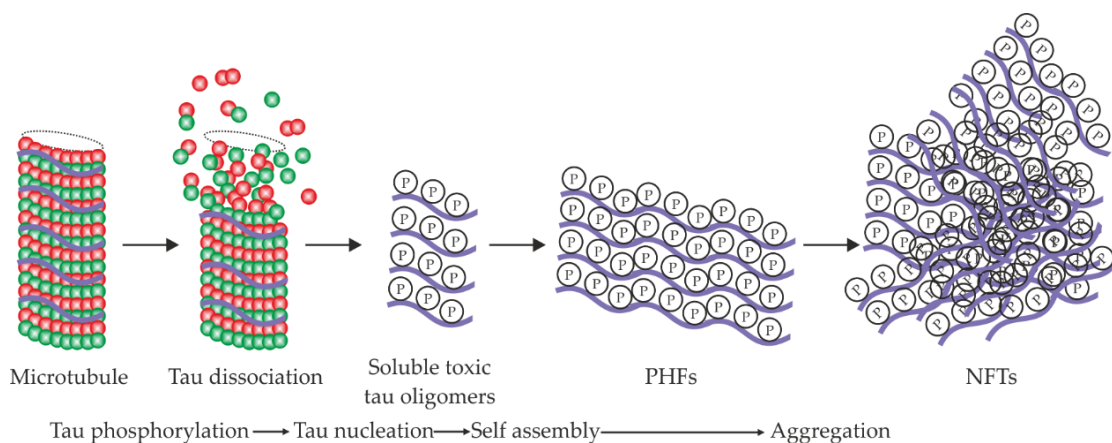


Figure 5.2. Hyperphosphorylation of tau causes its dissociation from microtubules and aggregation. Tau, purple, is attached to microtubules and aggregates into PHFs and NFTs once hyperphosphorylated.

The microtubules then polymerise (Cho and Johnson, 2004; Lindwall and Cole, 1984; Sengupta et al., 1998) and break down, leading to progressive memory loss. Tau pathology then spreads like prions throughout the brain (De Calignon et al., 2012). The toxic species are the small oligomers of tau rather than the large aggregates (Frost et al., 2009; Kfoury et al., 2012; SantaCruz et al., 2005).

A turn-like structure around residues Ser202, Thr205 and Ser208 is destabilised when the residues are phosphorylated (detected using an AT8 antibody (Braak et al., 2006)), encouraging the tau fibres to aggregate (Despres et al., 2017). Since phosphorylation of

different residues varies over the different Braak stages of AD (Braak and Braak, 1991) residues Tyr18 and Thr231 represent new early detection targets (Neddens et al., 2018).

Further to the detection mechanisms for Alzheimer's diseases, OGA is a target enzyme for the development of inhibitors against AD.

5.2.2 The first inhibitors of OGA

The first inhibitors to be used against OGA, streptozotocin (STZ) and *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)-amino *N*-phenyl carbamate (PUGNAc), were not specific to OGA meaning they both have off-target enzyme (GH89 and GH3 (Ficko-Blean et al., 2008; Stubbs et al., 2007)) and cell effects (Gao et al., 2000). PUGNAc was initially observed to inhibit the hexosaminidases, HexA and HexB and other GH enzymes (Horsch et al., 1991).

Streptozotocin is a derivative of *N*-methyl-*N*-nitrosourea *D*-glucosamine, an analogue of GlcNAc with a weak K_i of 64 μ M against OGA, Figure 5.3 (Vavra et al., 1960). It is naturally synthesised and isolated from *Streptomyces achromogenes*. Since pancreatic β -cell islets contain an abundance of glucose transporters, they have a higher uptake of STZ than the brain. The levels of O-GlcNAc and apoptosis were increased upon STZ addition; its cytotoxicity stems from the *N*-methyl-*N*-nitrosourea moiety, which releases nitric oxide and causes DNA alkylation. STZ was proved to be a general toxin to pancreatic cells rather than a specific inhibitor of OGA (Gao et al., 2000; Pathak et al., 2008). STZ is no longer used as a probe for OGA but it is still used to create short term animal models of type I diabetes (Mansford and Opie, 1968).

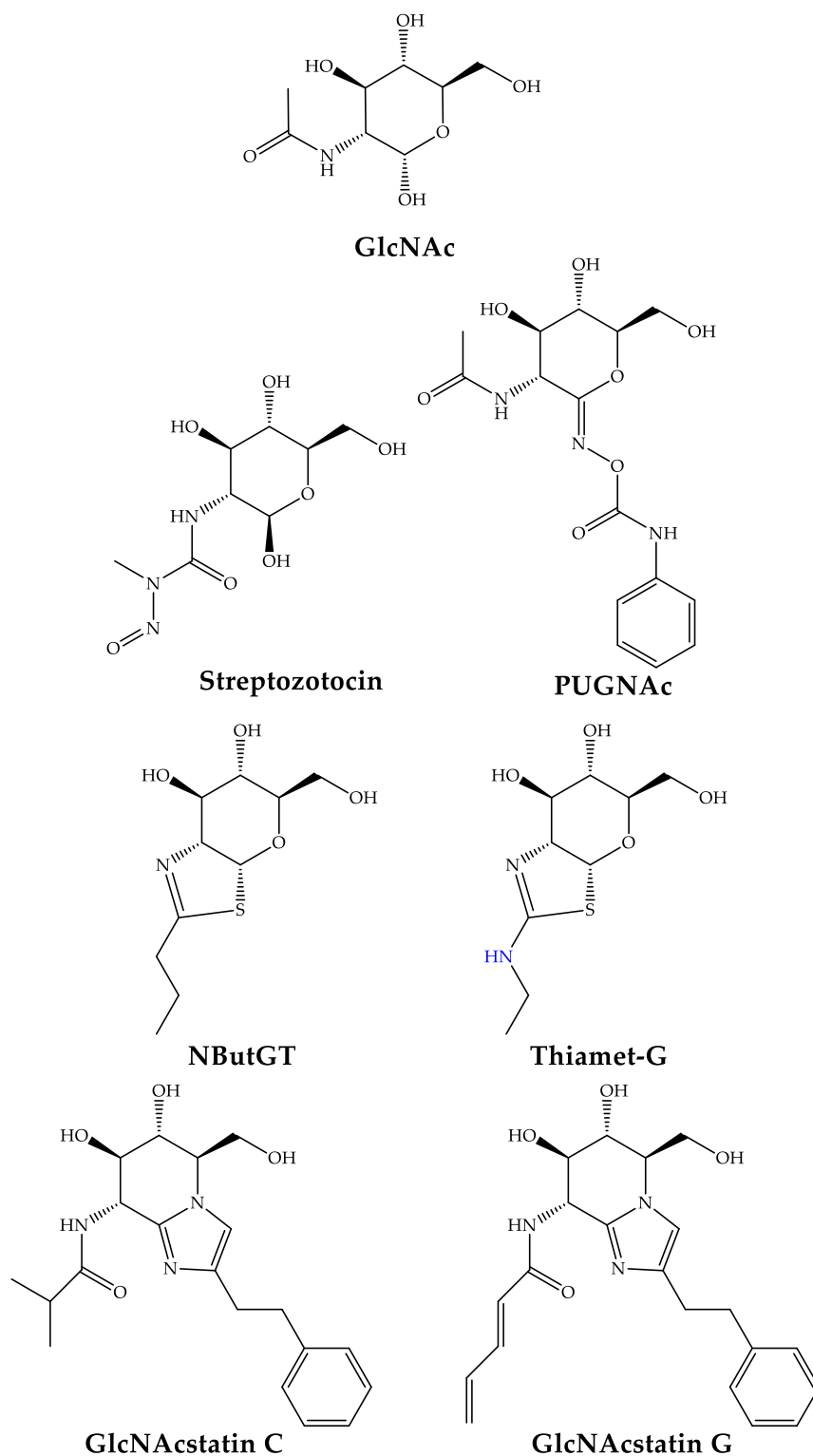


Figure 5.3. The evolution of O-GlcNAc hydrolyase inhibitors. Nitrogen in blue increases the basicity of the imidazole moiety.

The potent inhibitor, PUGNAc (K_i of 50 nM), was used to study the activity of OGA when the enzyme was first discovered, Figure 5.3 (Dong and Hart, 1994; Dorfmüller et al., 2010; Haltiwanger et al., 1998). It was shown before this to be a potent lysosomal

hexosaminidase inhibitor (Beer et al., 1990) inhibiting HexA/B on an equal scale (36 nM K_i) (Macauley et al., 2005b). When Chinese hamster ovary cells were treated with PUGNAc, an accumulation of atypical HexNAc molecules, 2-5 saccharides in length, was observed (Mehdy et al., 2012) linking the physiological effects of PUGNAc association to HexA/B and OGA inhibition. Unfortunately, any modifications to the inhibitor, specifically to the phenylcarbamate moiety, resulted in a decrease (or no change) in the potency of the inhibitor (Hattie et al., 2016; Shanmugasundaram et al., 2006).

After the mechanistic details of OGA were unveiled, by visualisation of the bacterial OGA structures from *C. perfringens* and *B. thetaiotaomicron*, inhibitors were designed against the shape of the active site and to replicate an oxazoline intermediate (Dennis et al., 2006; Rao et al., 2006).

1,2-dideoxy-2'-methyl- α -D-glucopyranoso[2,1-*d*]- Δ 2'-thiazoline (NAG-thiazoline) is a GlcNAc molecule substituted with thiazoline that mimics the oxazoline intermediate of the reaction (Knapp et al., 1996). It is used to inhibit other families that use the same substrate assisted mechanism, such as GH18 and GH20 (Liu et al., 2015; Mark et al., 2001; Whitworth et al., 2007). A derivative of NAG-thiazoline, NButGT, contains two methylene units increasing selectivity by 1500-fold for OGA over β -hexosaminidase and slightly decreasing the K_i to 230 nM, Figure 5.3 (Macauley et al., 2005b). NButGT was utilised in examining the levels of O-GlcNAc within cells injured by ischemia/reperfusion (Champattanachai et al., 2008). A 5-10 fold increase in O-GlcNAc levels was observed compared to OGT overproduction, however, NButGT did not protect against ischemia/reperfusion (Champattanachai et al., 2008).

Structural studies of NButGT in complex with *Bt*GH84 revealed that Asp242 (the general acid) interacts with the amide proton of the *N*-acetyl group (Whitworth et al., 2007). Increasing the basicity of this nitrogen, by addition of a nitrogen in the alkyl chain next to the thiazoline ring (shown in blue in Figure 5.3), was presumed to increase the number of favourable electrostatic interactions in the active site and its stability in solution (Macauley et al., 2005b). Thiamet-G has the highest selectivity (37,000-fold) for OGA over HexA/B due to the formation of a second ionic interaction between Asp242 (Asp174 *Hs*OGA) of 2.8 Å to the nitrogen as seen in the crystal structure of Thiamet-G and *Bt*GH84 (Yuzwa et al., 2008). The extra interaction caused a rearrangement of a cysteine residue at the base of

the active site and a reduction in the overall volume of the active site cleft, a possible reason for the increase in potency. When Thiamet-G was tested against various other GHs (2, 3, 13, 36 and 38) and a human lysosomal β -hexosaminidase, only β -hexosaminidase inhibition was observed (K_i of 750 μ M). It has been used in cell-based assays and in animal studies, since it can penetrate the cell membrane and blood-brain barrier due to the water-solubility and stability properties, see section 5.2.2.1.

Whilst Thiamet-G was being developed by Vocadlo and colleagues, Nagstatin was being modified by Van Aalten and colleagues. Nagstatin is a naturally occurring compound secreted by *Streptomyces amakusaensis* MG846-fF3 (Aoyagi et al., 1992). It inhibits *N*-acetyl- β -D-glucosaminidase with a K_i of 17 nM. A panel of glycoimidazole derivatives of Nagstatin were synthesised and showed activity against a variety of different GHs (Terinek and Vasella, 2005).

A fusion of Nagstatin and PUGNAc resulted in a new family of cell-permeable inhibitors named GlcNAcstatins (Dorfmueller et al., 2006, 2010). The tetrahydroimidazopyridine scaffold of Nagstatin and the phenyl carbamate moiety of PUGNAc, since it stacks against Trp286 in the active site of *BtGH84*, were exploited. To increase the potency, modifications to the C2 substituent by the addition of an aglycon group were shown to be beneficial (Panday et al., 2000; Shanmugasundaram and Vasella, 2005). The K_i of GlcNAcstatin against *CpGH84* was 4.6 pM when an isobutanamido moiety was added (Dorfmueller et al., 2006). A panel of five derivatives (A-E) were synthesised, which had modifications to the *N*-acyl substituent and the imidazole moiety (Dorfmueller et al., 2009). The *N*-acetyl group of GlcNAcstatin C contains an additional $\text{COCH}(\text{CH}_3)_2$, Figure 5.3, which has a K_i of 4.4 nM against *HsOGA* and a 164-fold increase in selectivity against HexA/B (Dorfmueller et al., 2009). Sequence alignment of *CpGH84* and *HsOGA* suggested Val331, which is present next to the *N*-acetyl moiety corresponds to Cys215 (now confirmed using available *HsOGA* structures). Mutation of *CpGH84* Val331 to cysteine resulted in a 25-fold decrease in the K_i of GlcNAcstatin C to 98 pM. Using this knowledge, the cysteine residue was targeted by different *N*-acetyl groups and further inhibitors were added to the panel (F-H) (Dorfmueller et al., 2010). GlcNAcstatin G has a similar K_i to GlcNAcstatin C against *HsOGA* of 4.1 nM, however, it is over 900,000-fold more selective against OGA than HexA/B, Figure 5.3.

5.2.2.1 Thiamet-G reduces the pathogenicity of tau

Tau pathogenicity was shown to be high within the CA1 region of the hippocampus, hence, targeted for observation of immunoreactivity, Figure 5.4 (Yuzwa et al., 2008). Maximal O-GlcNAc levels in nerve growth factor differentiated PC-12 cells, containing mature neurons and large amounts of tau, were detected using 200 nM of Thiamet-G after 12 hours (Yuzwa et al., 2008). Thiamet-G was also shown to reduce the phosphorylation of Thr231 and Ser396 on tau within PC-12 cells.

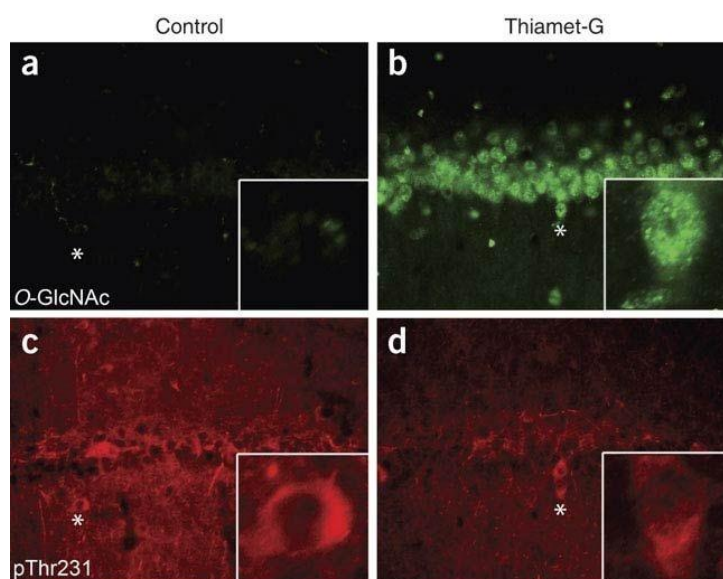


Figure 5.4. Thiamet-G increases O-GlcNAc levels and reduces phosphorylation levels within the hippocampus. Immunoreactivity of O-GlcNAcylated tau in the CA1 region of rat brains. c) Immunoreactivity of pThr231 on tau. b) and d) Results after administration of 200 mg kg⁻¹ d⁻¹ of Thiamet-G for one day by drinking water. Figure taken from Yuzwa et al., 2008.

Using click GalNAz coupled with TAMRA labelling, tau pathology was detected in the brains and cerebrospinal fluid of transgenic mice with a mutated human tau gene (P301L). Treatment with 500 mg/kg Thiamet-G reduced the presence of several species of insoluble pathological (and non-pathological) tau in the brain fraction and elevated the presence of soluble 55 kDa tau (Hastings et al., 2017). ~1.5 O-GlcNAcs were observed to undergo reciprocal phosphorylation and O-GlcNAcylation, which may be critical for the pathological state of tau (Hastings et al., 2017; Yuzwa et al., 2011). A physiological increase in the body weight of transgenic mice (tau P301L) and the number of motor

neurons resulted in a decrease to the nerve damage of the skeletal muscle of the mice after Thiamet-G treatment (Yuzwa et al., 2012).

Also, Thiamet-G inhibition of OGA induces autophagy *in vitro* and *in vivo* (shown in neuroblastoma cells and transgenic mice) (Zhu et al., 2018). Not only does Thiamet-G inhibition increase the levels of O-GlcNAc but it also decreases the levels of Sequestosome-1, a biomarker of autophagy (autophagy causes its degradation). The upregulation of autophagic flux has been shown to slow down the effects of neurodegeneration (Schaeffer et al., 2012). A decrease in pathological tau fragments was also observed, complementing previous studies by Yuzwa et al., 2012. Therefore, an increase in autophagic flux may assist in pathological tau removal (Zhu et al., 2018).

5.2.3 Interactions between the residues of *Hs*OGA and different inhibitors

Ligand binding supported the designation and role of the catalytic residues in the double-displacement mechanism. In the crystal structures of *Hs*OGA, Thiamet-G, PUGNAc (or a PUGNAc-imidazole derivative) and VV347 were found to bind in the -1 subsite, Figure 5.5 (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). VV347 is a potent derivative of pyrrolidine inhibitors with a K_i of 8 nM against *Hs*OGA (Bergeron-Brlek et al., 2015).

The catalytic aspartate residues, Asp174 and Asp175, are located on an inherently flexible loop (residues 172-182) that moves closer to the active site upon ligand binding. The role of Asp174 was confirmed as the polarising residue for the 2-acetamido group by the interaction with the acetamido nitrogen of VV347 and PUGNAc-imidazole (Roth et al., 2017). Lys98 stabilises the deprotonated state of Asp174. The role of Asp175 was also confirmed by its expected positioning and movement towards the anomeric carbon by 2.8 Å. A hydrogen bond was formed between the hydroxyl group of Tyr219 and PUGNAc. The *N*-ethyl group of Thiamet-G was enclosed by Cys215, Tyr219 and Trp278 and also engaged in hydrogen bonding with the main carbonyl of Gly67, Asn280, and Asn313, Figure 5.5 (Li et al., 2017a). The aglycon group of PUGNAc interacts with Phe223 and Val254. The phenyl group was 4.5 Å away from the loop formed by Glu677 to Gly683. Trp645 undergoes reorientation to accommodate the trifluoromethylphenyl group of VV347. Mutation studies of Lys98 and Asp285, which form a bidentate hydrogen bond

between the O4 and O6 hydroxyl of Thiamet-G, showed a loss in catalytic OGA activity upon mutation.

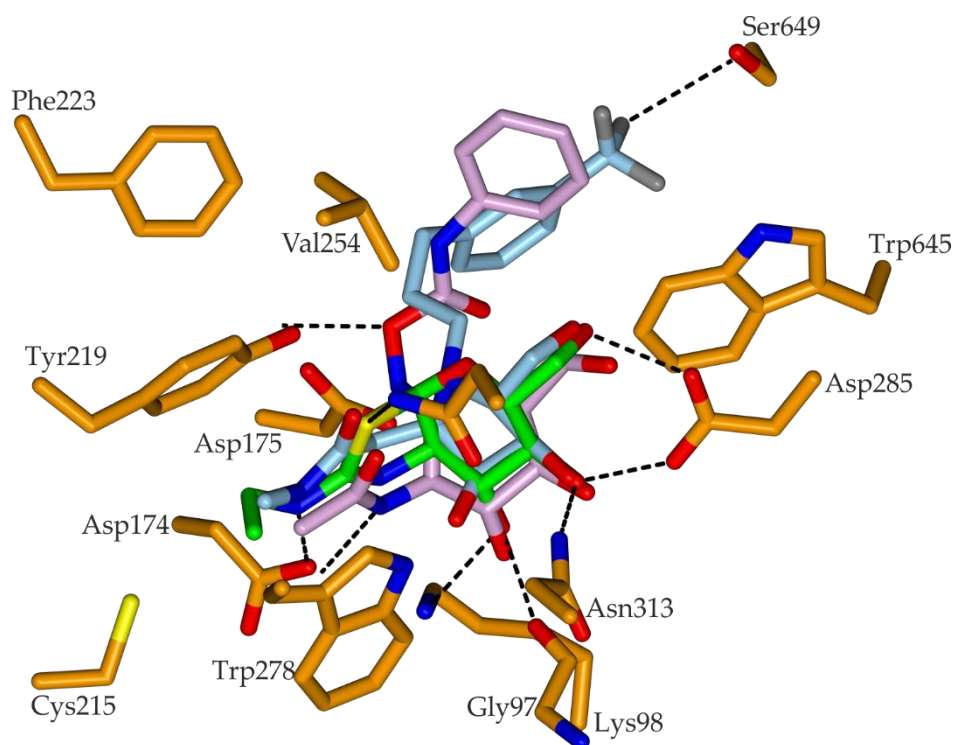


Figure 5.5. Ligand complexes in the active site of *HsOGA*. The ligands are surrounded by residues in orange (PDB ID: 5M7S) that form a hydrophobic pocket. PUGNAc is shown in pink (PDB ID: 5UHO), Thiamet-G in green (PDB ID: 5M7S), VV347 in blue (PDB ID: 5M7U).

5.2.4 Conformational itinerary of GH84 versus GH20 enzymes

A theme of this thesis is the correlation between the ability of an inhibitor to mimic the transition state and potency. The inhibitors described above were analysed by experimental and computational methods to provide a conformational itinerary.

NAG-thiazoline was shown by molecular dynamics to be a poor transition state mimic, since NButGT was observed in a semi-undistorted 4C_1 conformation with slight distortion to an 4E conformation (also observed in the crystal structure) (Greig and Williams, 2007). While the thiazoline derivatives interact with the general acid, the protonation step is assumed to occur before the transition state of the reaction (Lameira et al., 2008).

Computational analysis of Nagstatin analogues showed that they mimic the transition state (Greig and Williams, 2007). The oxocarbenium ion charge distribution is mimicked by protonation of the imidazole ring and the lowest energy conformation of Nagstatin is a half-chair/envelope (Panday et al., 2000; Shanmugasundaram and Vasella, 2005). The crystal structures of GlcNAcstatins confirm that they bind in a 4E conformation in the active site of *CpGH84*, Figure 5.6 (Dorfmueller et al., 2009, 2010). Furthermore, VV347 binds in a 4E conformation in the active site of *HsOGA* (Bergeron-Brlek et al., 2015; Roth et al., 2017).

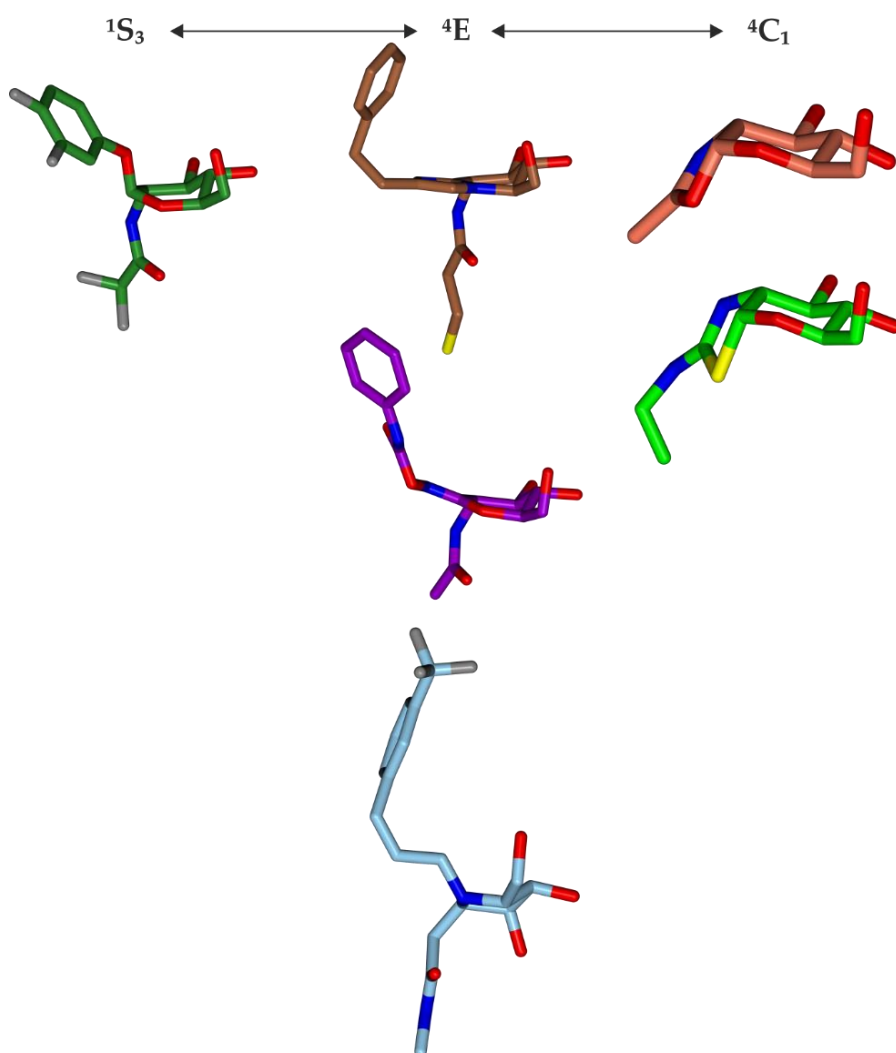


Figure 5.6. Inhibitor evidence for the conformational itinerary of GH84 enzymes. 3,4-difluorophenyl 2-deoxy-2-difluoroacetamido- β -D-glucopyranoside, bound to *BtGH84*, is shown in dark green (PDB ID: 2X0H) in a 3S_1 conformation; GlcNacstatin F in brown (PDB ID: 2XPK) and PUGNAc in purple (PDB ID: 2CBJ), bound to *CpGH84*, and VV347, bound to *HsOGA*, in blue (PDB ID: 5M7U) are in a 4E transition state conformation; Oxazoline, bound to *BtGH84*, in red (PDB ID: 2WZH) and Thiamet-G, bound to *HsOGA*, in green (PDB ID: 5M7S) are in a 4C_1 conformation.

The compound 3,4-difluorophenyl 2-deoxy-2-difluoroacetamido- β -D-glucopyranoside was observed in the active site of *BtGH84* in a 1S_3 conformation and the compounds 2-acetamido-2-deoxy-5-fluoro- β -D-glucopyranosyl fluoride (5F-oxazoline intermediate) and 4-methylumbelliferyl 2-acetamido-2-deoxy- β -D-glucopyranoside (oxazoline intermediate) were observed in a 4C_1 conformation, Figure 5.6 (He et al., 2010). Also, Thiamet-G was visualised in a 4C_1 conformation in the active site of *HsOGA* (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). Together, this data reveals a ${}^1S_3 \leftrightarrow [{}^4H_3/{}^4E]^\ddagger \leftrightarrow {}^4C_1$ conformational itinerary for GH84, comparative to the conformational itinerary of GH20 hexosaminidase (${}^1A/B/{}^1S_3 \leftrightarrow [{}^4H_3/{}^4E]^\ddagger \leftrightarrow {}^4C_1$) (Hattie et al., 2015; Ito et al., 2013).

5.2.5 Recent OGA inhibitors

New and modified inhibitors to OGA are still highly desirable. The design of novel first-in-class inhibitors, modifications to existing inhibitors and multivalency inhibitors will be discussed in this section.

A series of multivalent inhibitors were designed based on acetamido azepane as the enzyme binding moiety, Figure 5.7 (Alvarez-Dorta et al., 2017). The tetra and octavalent compounds have an IC_{50} value of 430 and 320 nM for *HsOGA*, respectively, and 33 and 7 nM for HexA/B, respectively. Although, there is a 10-fold higher selectivity for HexA/B than for *HsOGA*, the enzyme binding moiety has the potential to be modified. This provides an interesting new route towards multivalent inhibitors.

From a library of 50,000 compounds, naphthalimides were shown to be potent inhibitors of *N*-acetyl- β -hexosaminidases (Tropak et al., 2007). Attaching the naphthalimide via a thiol linkage to a GlcNAc moiety resulted in an inhibitor with a K_i of 3.8 μ M against *HsOGA* and 10-fold higher potency compared to *HsHexB* (Shen et al., 2018a). The naphthalimide group in the 1 subsite interacts with Ser649, Arg682 and Trp679. Using rational design to improve the potency and selectivity, a piperidyl moiety was added and the length of the flexible linker was varied to an optimum of ~6 atoms, Figure 5.7 (Chen et al., 2018). This resulted in a compound with a K_i of 0.6 μ M against *HsOGA* and a 167-fold higher potency compared to *HsHexB* (Shen et al., 2018b). Further cell-based analysis is now needed for these compounds.

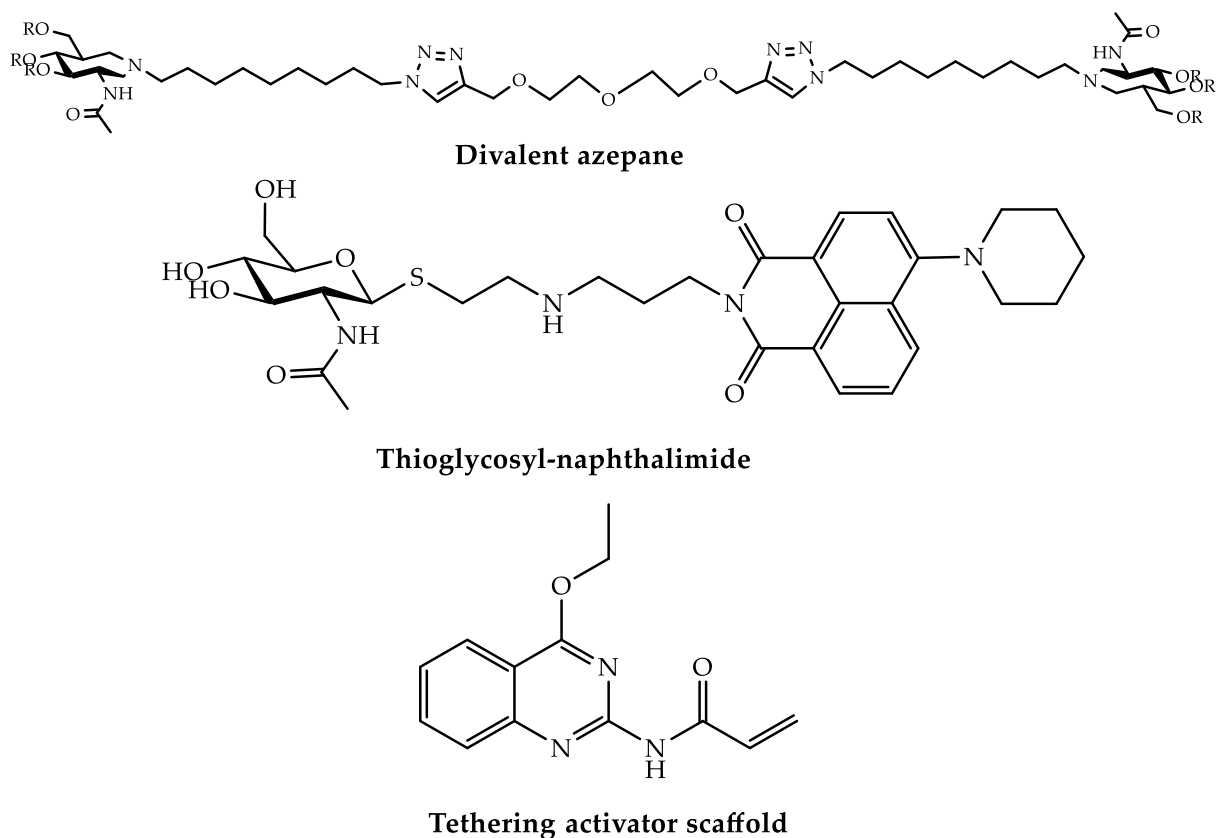


Figure 5.7. Recent potent and selective OGA inhibitors.

Small molecules can act as either an inhibitor or activator. A covalent activator can increase both the catalytic efficiency of the enzyme and its susceptibility to an inhibitor. From a fragment library developed to screen small molecular weight compounds, 4-ethoxyquinazoline was observed to increase the k_{cat}/K_M of *BtGH84* by 2-fold due to the interaction between Arg347 (Gln288 of *HsOGA*) (Darby et al., 2014). Mutation of Tyr550 to a cysteine and the inclusion of an additional moiety to tether the activator to the enzyme resulted in up to a 32-fold increase in k_{cat}/K_M , Figure 5.7 (Darby et al., 2017). A 17-fold difference in the dissociation constant for binding of PUGNAc in the presence and absence of the activator was observed. The affinity of OGA for Thiamet-G was not increased, since Thiamet-G lacks the hydrophobic aglycon group of PUGNAc (which stacks against the activator).

5.2.6 Inhibitors in clinical trials

As well as the promising results of Thiamet-G in *in vitro* and *in vivo* assays, there are two small molecule inhibitors, MK-8719 and ASN120290, that have shown promising *in vivo* data (reviewed in Medina, 2018; Ryan et al., 2018; Sandhu et al., 2016; Smith et al., 2016).

At the time of writing, MK-8719, Alectos Therapeutics, has received Orphan Drug Designation from the FDA and has advanced through phase I clinical trials, Figure 5.8. It is being administered as a treatment for progressive supranuclear palsy, a neurodegenerative disorder developed from nerve damage due to tau abnormalities. After varied doses of MK-8719 (between 5 and 1200 mg), a reduction in the presence of NFTs was observed in 16 healthy volunteers and no adverse effects were seen at the highest concentration (Smith et al., 2016). Peripheral blood mononuclear cells of healthy volunteers were monitored for O-protein levels by positron emission tomography imaging using the PET tracer ^{18}F -MK-8553 after administration of a range of doses of MK-8719 (Sandhu et al., 2016). O-protein levels increased with increasing inhibitor concentration and a pharmacokinetic model was developed as a dosing guide. In an X-ray structure, interactions were observed between the N-ethyl group (selection marker) and a hydrophobic pocket formed by Tyr219 and Trp278 (Selnick et al., 2019).

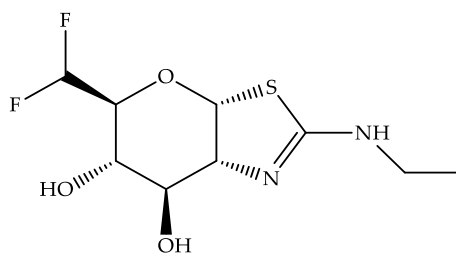


Figure 5.8. First-in-human OGA inhibitor MK-8719.

A second first-in-class OGA inhibitor, ASN120290 from Asceneuron (structure undisclosed), has been successful in the first stage of clinical trials to treat progressive supranuclear palsy and has also received Orphan Drug Designation from the FDA (Ryan et al., 2018). Transgenic mice, containing the P301A tau mutation, responded positively to treatment of ASN120290 over 3.5 months. By taking samples of cerebrospinal fluid, the pharmacokinetic properties were tested and concluded that the inhibitor is highly capable of crossing the blood-brain barrier. The trial was split into three sections categorised by dose type (single 1000 mg dose/single ascending/multiple ascending doses up to 500 mg), food effect (fasted or fed) and volunteer age (young/elderly). Subjects who had fasted or were fed did not show any considerable differences and no volunteer suffered from severe adverse events. Both of the inhibitors, MK-8719 and ASN120290, are planned to enter phase II clinical trials.

A novel positron emission tomography (PET) radioligand is currently in phase I of clinical trials to image OGA in the brain of healthy volunteers. High uptake and slow release of (18F)-OGA-1 (LSN331662) was shown to be reduced in *Oga*^{ΔBR} (mouse brain specific OGA knockout) (by 82% compared to control mice) and in Rhesus monkey brains pre-treated with ThiametG (Paul et al., 2019).

**Crystal Structure Guided Human O-GlcNAcase
Iminocyclitol and Bicyclic Inhibitors Access Sub-
Nanomolar Potency**

5.3 Abstract

In collaboration with Professor Robert Britton, Professor David Vocadlo and colleagues, first-in-class OGA inhibitors have been synthesised and analysed, both in a structural and in an *in vitro/vivo* context. The inhibitors were developed with the long-term outcome of being used therapeutically after passing clinical trials. Through obtaining the crystal structures of the compound in complex with OGA, development into the chain length and side groups could be undertaken to improve the potency and the permeability of the inhibitors to the blood brain barrier. The triazole-based series of pyrrolidine inhibitors have constrained stereochemistry, therefore, it overcomes many of the factors of pharmacokinetics, for example, drug absorption and distribution. The precursor contains an alkyne bond that, through utilising click chemistry, can be reacted with an azide bonded to different R groups. This can be applied to pull down assays, where the compound can be conjugated directly to biotin. Picomolar inhibitors were revealed when the R group was an indole ring or contained fluorine and bromine substituents. The preliminary data for the inhibitors shows that the picomolar affinity may be due to interactions of the aglycon group of the inhibitor in the +1 subsite with aromatic residues of the stalk domain of OGA within a cleft not observed in bacterial OGA structures.

5.4 Introduction

Pyrrolidine-based iminocyclitols have been shown to be extremely potent and it is straightforward to synthesise a large yield of product (Bergeron-Brlek et al., 2015). A K_i of 9 nM against *Hs*OGA was explained by optimal geometries of the ligand and hydrogen bonding interactions to residues in the active site of *Bt*GH84, Figure 5.7. Pyrrolidines with a carbon-3 side chain and substituted aryls were ideal for inhibition, Figure 5.9.

Testing of the pharmacokinetic properties of the inhibitor, revealed it is able to cross the blood-brain barrier and was well distributed, showing good bioavailability (Bergeron-Brlek et al., 2015). O-GlcNAc levels, within the sagittal brain of mice, were quantitatively and qualitatively increased in the cortex and hippocampus 16 hours after administration of the inhibitor.

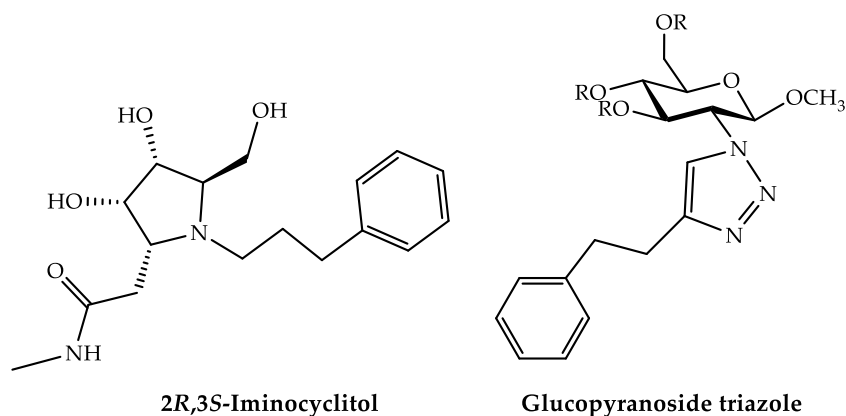


Figure 5.9. Influential inhibitors to the design of a novel triazole series.

Recently, a series of derivatives of glucopyranoside have been synthesised which, instead of modifying the C1 position of the -1 subsite group, they replace the *N*-acetyl group with a 1,4-disubstituted 1,2,3-triazole ring, Figure 5.9 (Igal et al., 2019). A variety of different substituents were attached to this moiety resulting in different potencies. The simplest phenyl ring, Figure 5.9, increased O-GlcNAc levels comparable to that of Thiamet-G and maintained conserved interactions with the phenyl linker and Cys278, Tyr282 and Trp337 at the base of the active site. An IC₅₀ value of 0.5 μ M against OGA and 550 μ M against HexA/B revealed promising selectivity and the compound did not show any cytotoxic effect.

Crystal structures of *Hs*OGA revealed a cleft formed between the two homodimeric subunits. This is a feature solely found in the human isoform (against the current bacterial models). Enhancing interactions in this region by varying the substituents of the side chain attached to C1 (rather than C2) could lead to increased potency. In this section, the design, synthesis and validation of a library of five-membered structures will be discussed along with a first-in-class series of bicyclic inhibitors. It is difficult to assign the precise stereochemistry of five-membered structures as the coupling constants are more complicated than in a six-membered system. Synthesising the inversion of a specific stereocentre has a published yield of ~30%, however with improvements to the synthesis protocol by providing a different route to access the specific molecule, precise stereochemistry was achieved.

5.5 Experimental methodology

The compounds were synthesised by Johannes Lehmann and Hong-Yee Tan (members of Professor Robert Britton and Professor David Vocadlo's groups, respectively). Cameron Proceviat conducted the *in vitro* K_i inhibition assay. Crystals of *Hs*OGA were soaked with a final concentration of 5mM for the inhibitors **1** and **3**. Diffracting crystals containing inhibitors **1** and **3** were sent to Diamond I04 beamline and processed using the Xia2 pipeline. The data was reintegrated using Aimless (Evans, 2006, 2011) from the CCP4 software suite (Winn et al., 2011). Multiple rounds of refinement using REFMAC (Murshudov et al., 1997, 2011; Nicholls et al., 2012; Pannu et al., 1998; Vagin et al., 2004; Winn et al., 2003) and model building in COOT (Emsley et al., 2010) were used to refine the model.

5.6 Results

5.6.1 Modification of the pyrrolidine-based iminocyclitols reveals nanomolar affinity inhibitors

The basic 2*R*-3*S*-iminocyclitol inhibitor was used as a platform to design a series of triazole inhibitors. A second five-membered ring, 1,2,3 triazole, was added to the ring nitrogen via different lengths of carbon chains. One carbon to six carbons were tested for potency, Figure 5.10a. All had nanomolar affinity, with the highest having four carbons and a K_i of 8.3 nM against *Hs*OGA. Based on this result, the orientation of the triazole ring was examined, however, there was very little change in affinity (from 6.5 nM to 5.4 nM), Figure 5.10b.

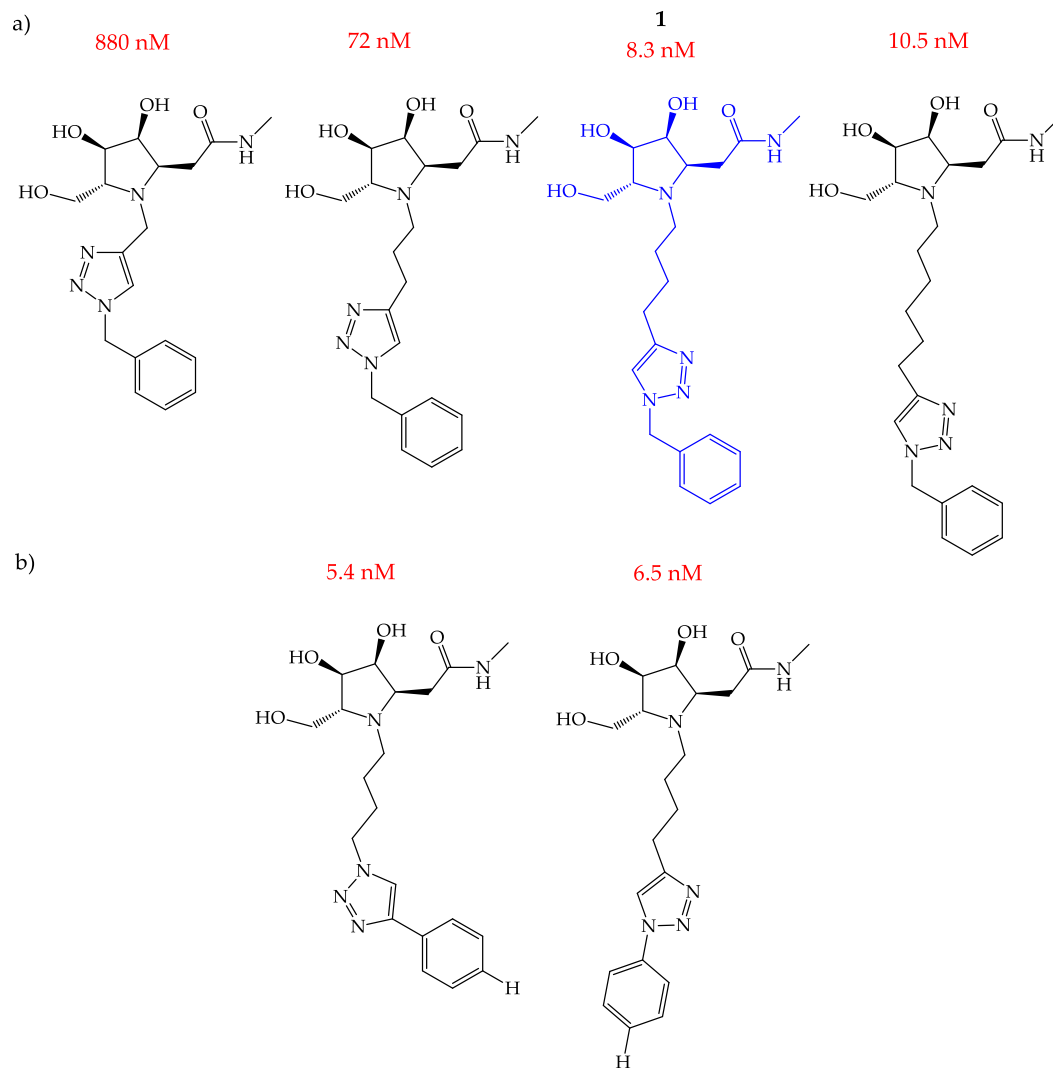


Figure 5.10. An initial test series of triazole-based compounds. a) Compounds with altered lengths of carbon linkers. The compound with the highest potency is shown in blue. b) The orientation of the 1,2,3-triazole moiety was altered.

5.6.2 **1** causes a rearrangement of the loop containing Trp679 due to pi stacking interactions

A crystal structure was obtained for *Hs*OGA in complex with compound **1**. The catalytic residues, Asp174 and Asp175, are poised for attack,

Figure 5.11a. Of the pyrrolidine ring O3 interacts with Gly67, O4 interacts with both Asn313 and Asp285, and the oxygen and nitrogen of the *N*-acetyl group interacts with Asn280 and Asp174, respectively,

Figure 5.11b. Compared to the published structures of different iminocyclitols and *BtGH84*, the bonding interactions in the active site of *HsOGA* are identical except for

His433,

Figure 5.11c.

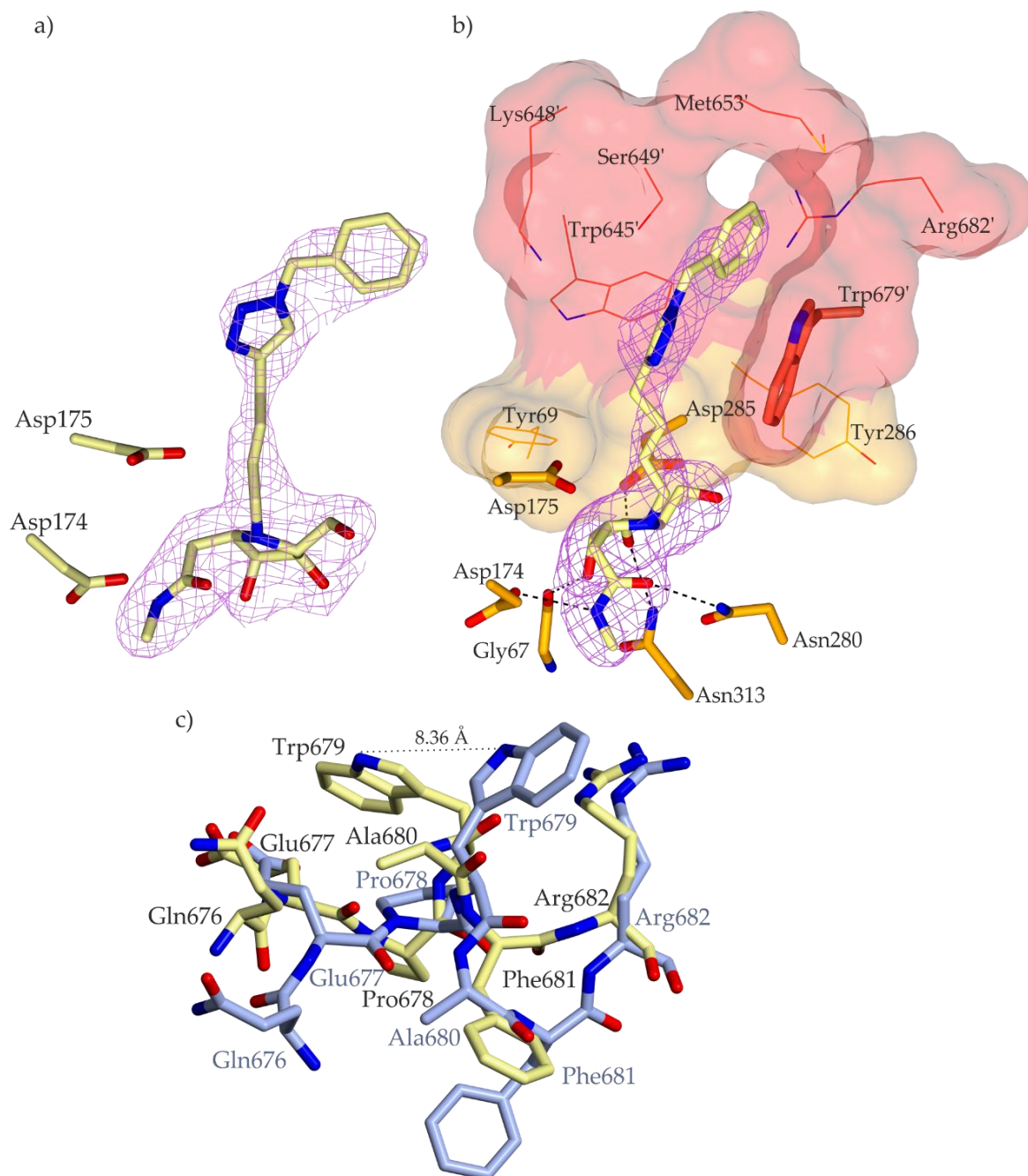


Figure 5.11. Structure of 1 in complex *HsOGA*. a) Compound 1 and residues of *HsOGA* are shown yellow. The maximum-likelihood/ σ_A -weighted $2F_{\text{obs}}-F_{\text{calc}}$ map, shown in pink, was contoured at $0.08 \text{ e}/\text{\AA}^3$. b) Black dotted lines represent hydrogen bonds between the residues of the active site and 1. The residues and surface of chain A are shown in orange

and chain B are shown in red. c) Movement of the loop between residues Gln676 and Arg682 for *Hs*OGA (PDB ID: 5M7R) in blue and *Hs*OGA_1 in yellow.

In the structure of the iminocyclitol **31** and *Bt*GH84, His433 interacts with O5 via a water molecule (Bergeron-Brlek et al., 2015). The histidine residue corresponds to Trp679 of *Hs*OGA, which in comparison, forms pi stacking interactions with the triazole group of **1**. In order for this to happen the residues on the loop between Phe671 and Gly683 have shifted closer to the centre of the active site,

Figure 5.11d. Trp679 has moved 8.36 Å in reference to the WT “apo” structure (Roth et al., 2017). The backbone between residues Gln676 and Phe681 has also shifted to accommodate for this movement.

5.6.3 A triazole-based library reveals picomolar affinity inhibitors

A new chemical scheme was developed based on the highest affinity four carbon linker compound. A starting substrate was developed to allow the building of a library, which using click chemistry, could easily test the type of moiety in the +1 subsite, Figure 5.12. The library contained ~30 molecules or triazoles which were proposed to increase the interaction between the inhibitor and residue Trp679. The improved synthesis allowed different side chain decorations on the nitrogen (allyl and alkyne) favouring them for further derivatisation reactions.

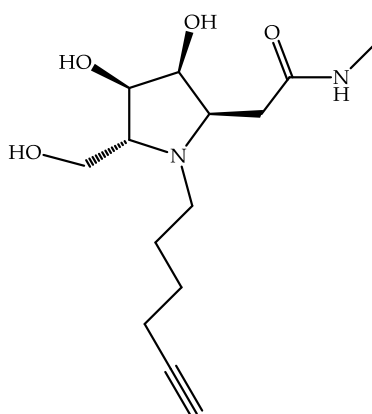


Figure 5.12. Starting substrate used to build the compound library.

The potency of the inhibitors varied from over 100 nM inhibitory concentrations to less than 1 nM. Four moieties had picomolar inhibition; a fluorobenzene, a 2,4,6-tribromobenzene and two indoles differing by orientation and the length of the carbon linker to the triazole moiety, Figure 5.13. This series of novel inhibitors are the most potent inhibitors towards *HsOGA* currently known.

To explore the potency of the inhibitors, crystal structures of *HsOGA* in complex with the inhibitors that have the highest potency were obtained. The active site was examined for enzyme:inhibitor interactions and changes in conformation. Novel conformations of the inhibitors in the +1 subsite was revealed due to stacking pi-interactions of the triazoles. A structure was obtained with compound **2**, **3** and **4** but crystal structures of, **5** and **6** could not be obtained, possibly due to the hydrophobicity of the compounds.

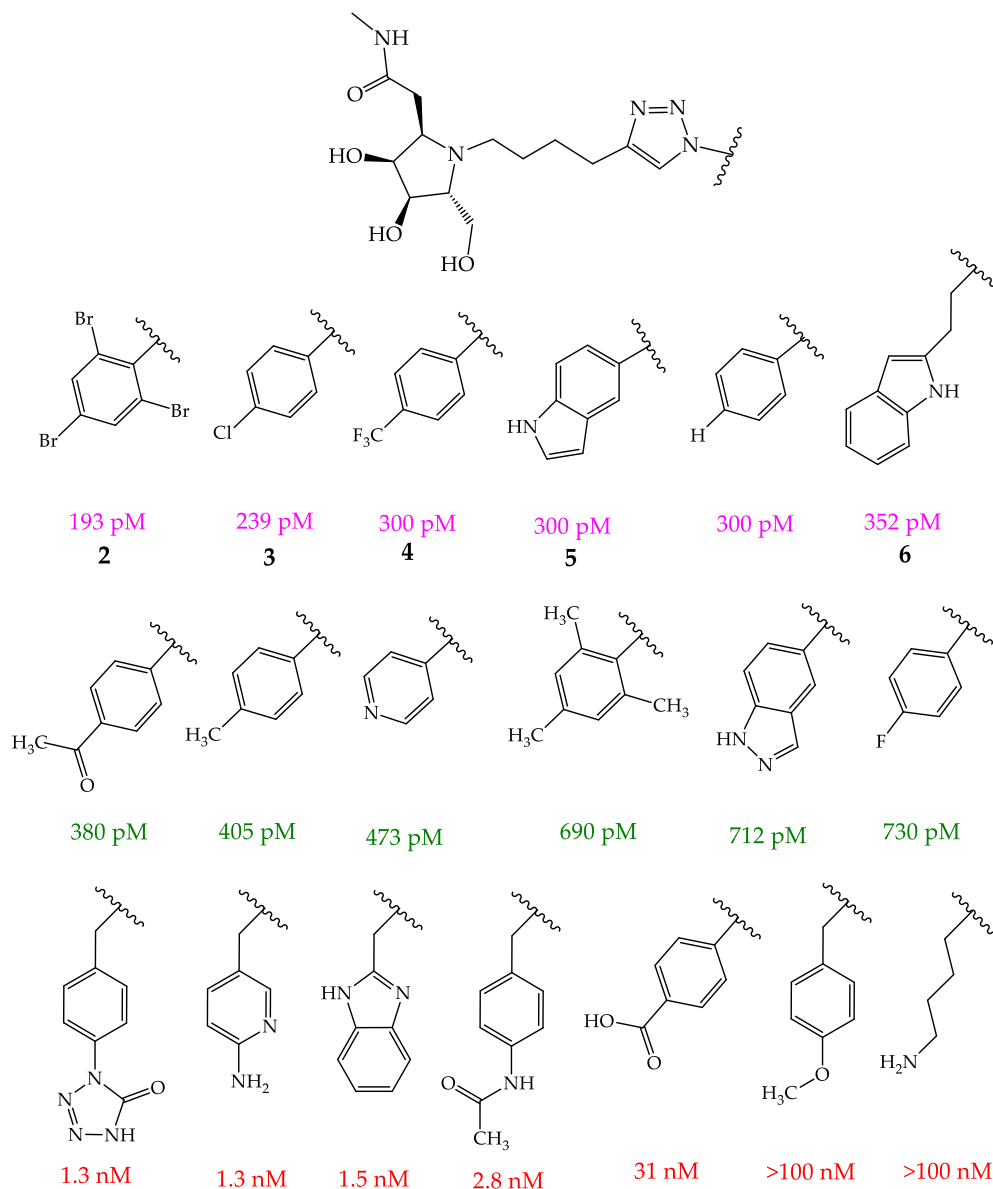


Figure 5.13. Triazole compound library. Top: Triazole precursor. Bottom: 20 compounds proposed to attach to the triazole moiety. The K_i is coloured according to an inhibition range, <360 pM in Pink, 1 nM $> X > 360$ pM in green and >1 nM in red.

5.6.4 π stacking interactions with active site residues increase the potency of the inhibitors

Interestingly, two conformations of compound **2** were observed in 75% of the active sites of *Hs*OGA, Figure 5.14a. The vertical form was present in monomer A and monomer B (*Hs*OGA is a dimer of monomer A and B) and the “L”-shaped form was present with 40% occupancy. The same interactions as compound **1** were observed for the pyrrolidine ring of **2** and the residues of the active site, Figure 5.14b. The two conformations of **2** do not

correspond to the conformation of compound **1** in the active site of *Hs*OGA Figure 5.14c. However, the same movement of the loop between residues Phe671 and Gly683 lead to a similar interaction between the vertical form of compound **2** and Trp679. This second interaction is not possible in the active site of *Bt*GH84 since Arg347 blocks the pocket, Figure 5.14d. The “L”-form of compound **2** was stabilised by the aromatic residues Tyr69 and Trp645, whereby the 2,4,6-tribromobenzene ring is sandwiched between the two residues. In comparison to the *Hs*OGA_1 structure, the Nε1 of Trp645 has moved 1.65 Å compared to the respective position of Trp645 of *Hs*OGA_2, Figure 5.14e. Trp645 does not move position when comparing *Hs*OGA_1 to the apo structure.

Table 5.1. Data collection and refinement statistics of *Hs*OGA with inhibitors 1 and 3.

*Values in parentheses are for high resolution data.

	<i>Hs</i> OGA_1	<i>Hs</i> OGA_2	<i>Hs</i> OGA_4
Data collection			
Beamline	Diamond I04	Diamond I04	Diamond I03
Space group	<i>P</i> 4 ₃ 2 ₁ 2	<i>C</i> 1 2 1	<i>P</i> 4 ₃ 2 ₁ 2
Cell dimensions:			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	101.1, 101.1, 285.1	143.5, 143.1, 276.7	101.4, 101.4, 282.7
β (°)	90.0	92.7	90.0
Resolution (Å)	82.48-2.34 (2.40-2.34)*	71.68-3.10 (3.15-3.10)*	95.41-2.57 (2.65-2.57)
<i>R</i> _{merge}	0.15 (2.42)	0.14 (1.15)	0.09 (3.17)
<i>CC</i> (1/2)	1.00 (0.52)	1.00 (0.65)	1.00 (0.60)
<i>I</i> / σ <i>I</i>	8.4 (1.0)	4.3 (0.9)	16.6 (0.8)
Refinement			
Resolution (Å)	82.48-2.34 (2.40-2.34)*	71.68-3.10 (3.15-3.10)*	95.41-2.57 (2.65-2.57)
R _{work} / R _{free}	0.20/0.26	0.24/0.31	0.21/0.26
B-factors (Å ²)			
Protein	84	76	98
1/2/4	88	70	102
Water	61	0	83
R.m.s. deviations			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	1.68	1.82	1.77

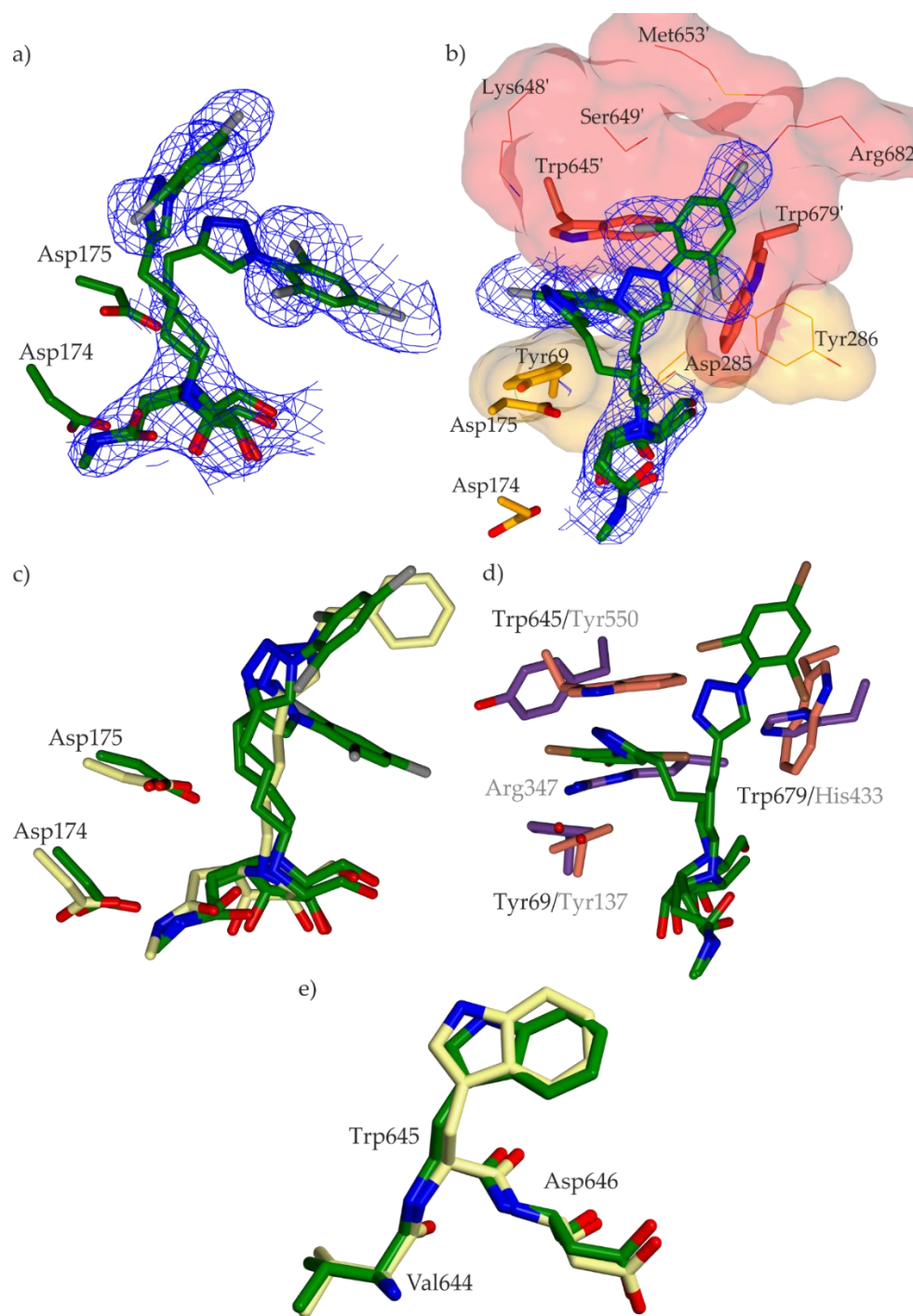


Figure 5.14. Structure of *HsOGA* in complex with compound 3. a) Compound 3 and residues of *HsOGA* are shown green. The maximum-likelihood/ σ_A -weighted $2F_{\text{obs}}-F_{\text{calc}}$ map, shown in blue, was contoured at $0.14 \text{ e}/\text{\AA}^3$. b) Black dotted lines represent hydrogen bonds between the residues of the active site and 3. The residues and surface of chain A are shown in orange and chain B are shown in red. c) Overlay of compound 1 in yellow and compound 3 in green. d) Pi stacking interactions between hydrophobic aromatic residues and the 2,4,6-tribromobenzene and triazole moieties of the inhibitors. Overlay with *BtGH84* in purple. e) Movement of Trp645 compared to the *HsOGA* apo and compound 1 structures.

A third structure of *HsOGA* in complex with compound **4** was achieved. Compound **4** extends further into the pocket of chain B. Hydrogen bond interactions with the side chains of Ser649, Met653 and Arg682 and the main chain of Trp679 may contribute to the high potency of the compound, Figure 5.15.

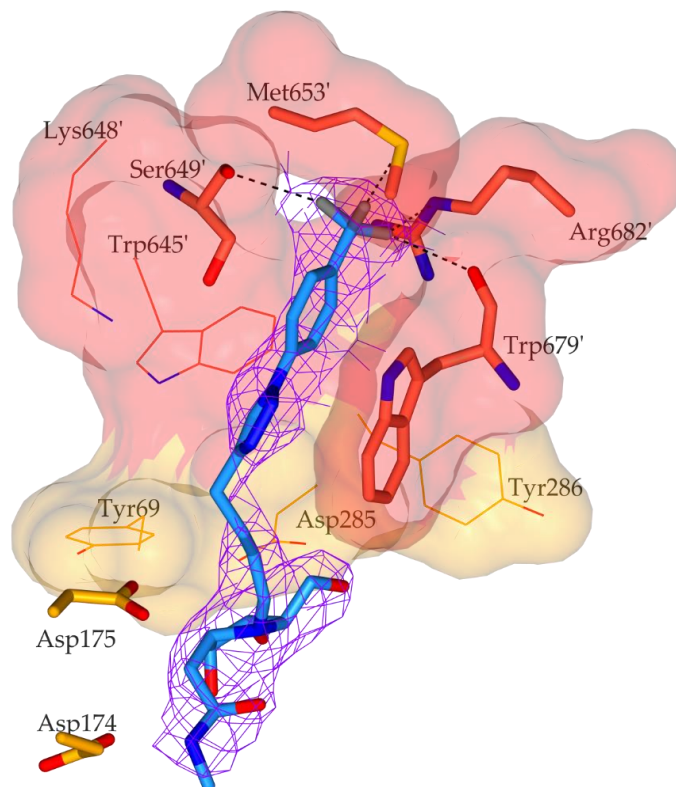


Figure 5.15. Structure of *HsOGA* in complex with compound **4.** Compound **4** in blue is in the active site pocket of *HsOGA* coloured by chain (chain A in orange and chain B in red). The interacting residues are shown with thicker bonds than the residues that compose the active site pocket.

5.6.5 Bicyclic inhibitors sit in the *HsOGA* specific cleft

Assessment of another series of first-in-class inhibitors has been initiated. This class of inhibitors utilise a bicyclic carbohydrate moiety (pyrrolizidine ring). The different compounds vary by the stereochemistry of the 5 and 7 position substituents, Figure 5.16. Crystal structures and *in vivo* data have revealed more information on the potential therapeutic use of these inhibitors. The same loop movement is observed between residues Phe671 and Gly683. The trifluoromethyl groups lie within the cleft formed between residues Tyr69, Asp285 and Tyr286 of chain A and Trp645, Lys648, Ser649 and

Trp679 of chain B, Figure 5.17. A chlorine ion, possibly from the NaCl component of the gel filtration buffer, is thought to be within interacting distance from the trifluoromethyl benzene. Initially, water was modelled into the positive density, however, it did not contain enough electrons and positive density was still present, Figure 5.17.

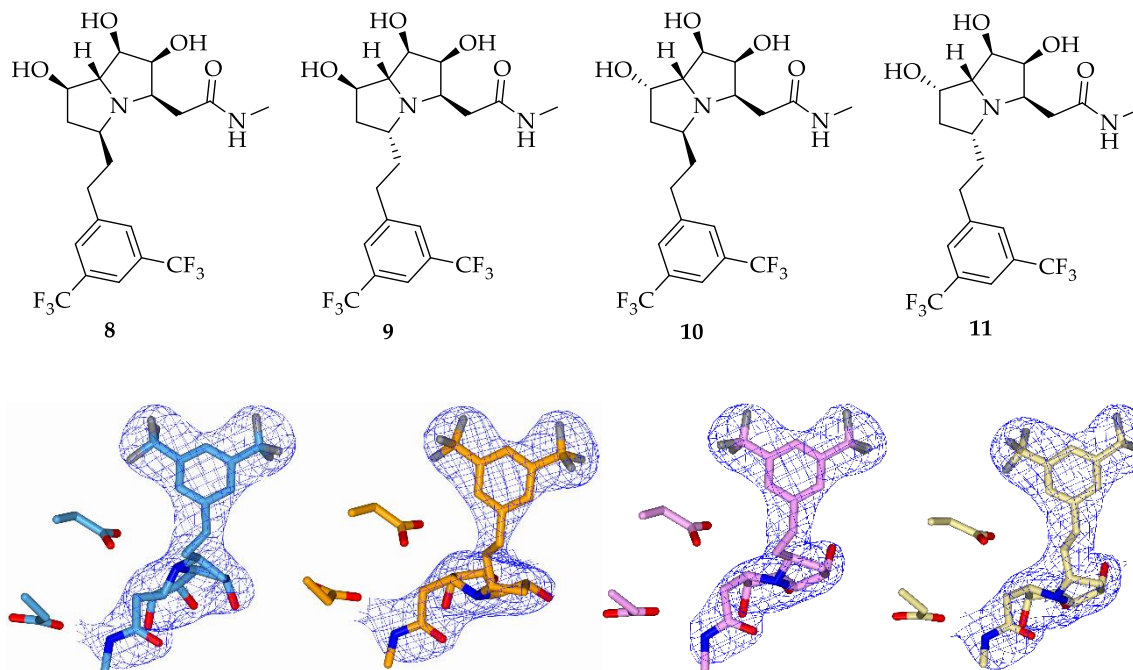


Figure 5.16. The structures of a bicyclic series of OGA inhibitors. Top: Chemical structure of four bicyclic inhibitors. Bottom: *HsOGA* active site interaction; the top residue is Asp175 and the bottom residue is Asp174.

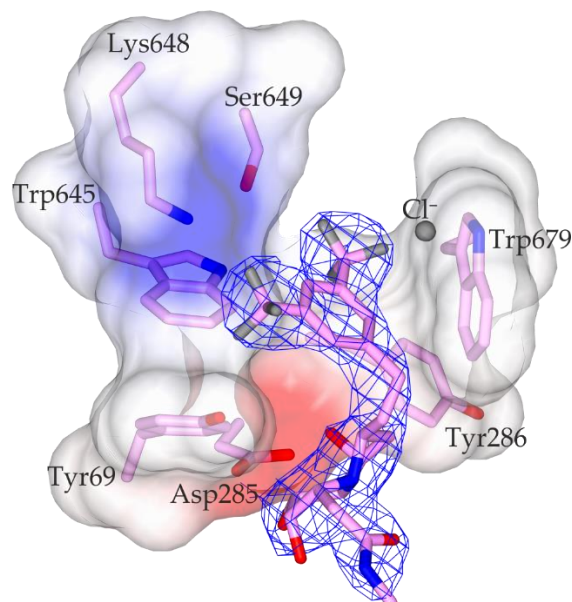


Figure 5.17. Inhibitor interactions with the unique active site cleft of *HsOGA*. Bicyclic inhibitor in blue electron density surrounded by pink residues in a surface representation coloured by electrostatic potential.

5.7 Discussion

The original pyrrolidine-based iminocyclitols were able to cross the blood-brain barrier and were shown to increase the overall glycosylation levels. By conducting cell-based assays these important features of an inhibitor can be assessed for possible therapeutic uses. The O-GlcNAc levels were shown to have significantly increased in mice treated with a pyrrolidine-based iminocyclitol after 16 hours, a target for these triazole inhibitors to surpass. The 2,4,6-tribromobenzene moiety is not suitable for clinical testing due to its toxicity, however, bromine has a high scattering factor, since it is a heavy atom (more electrons than carbon), and can be easily identified in crystal maps. The interaction of **4** with Ser649, Trp679 and Met653 contributes to the high affinity.

CHAPTER 6: Conclusions and Perspectives

Mannose and *N*-acetylglucosamine processing enzymes are characterized by the stereochemistry of their O2 substituents. In the case of mannosidases and mannanases, the steric hindrance of an axial *manno*-O2 (in the -1 subsite) must be alleviated through distortion of the sugar to various boat and skew-boat conformations. For β -*N*-acetylglucosaminidases, catalysis harnesses the internal nucleophile provided by the neighbouring group carbonyl. Although, recently it was shown that *endo*- α -1,2-mannanases, CAZy family GH99, utilise the *manno*-O2 hydroxyl for neighbouring group participation rather than utilising an enzymatic nucleophile and making it equatorial (Sobala et al., 2019) (see appendix for an authored review article on mannose chemistry). In both cases, these unusual mechanistic features can be exploited for the design of enzyme inhibitors, and potentially as drugs for clinical diseases.

Structural analyses along the conformational itinerary of an enzyme has provided considerable insight into the catalytic mechanisms. The transition state, a high energy and short-lived conformation, is characterised by the sp^2 hybridisation of C1 and O5 (formation of a partial positively charged carbocation) and a partial double-bond between the anomeric carbon and ring oxygen. This can be achieved by half-chair and boat conformations. Designing inhibitors that mimic the transition state requires: (i) understanding the conformational itinerary of a GH and (ii) understanding the properties of the inhibitor. Not only can one inhibit an enzyme but also activate it. It was recently shown that aromatic residues of the active site, which form electrostatic and van der Waals interactions, can increase the rate of the reaction through glycosyl cation stabilisation and a decrease in stereoselectivity, leading to enzyme engineering (Montalvillo-Jiménez et al., 2019). The work in this thesis was focussed on providing such information on conformational itinerary mimics for three main classes of enzyme.

Chapter 2 reported on the CAZy family GH125 a grouping of α -1,6-mannosidases that cleave the terminal α -1,6-mannosidic bond of $\text{Man}_3\text{GlcNAc}$ and $\text{Man}_2\text{GlcNAc}$. Prior to this work, a thio-oligosaccharide substrate mimic and the inhibitor 1-deoxymannojirimycin showed no distortion in the active site of GH125 enzymes (although thiosaccharides have accurately provided information on the substrate conformation for many GH enzymes, for example endoglucanase I and *CkGH47* (Sulzenbacher et al., 1996; Thompson et al., 2012)). The lack of distortion was attributed to the bond length and angles of the C-S bond as seen in computational experiments conducted by the Rovira group. The work described

in this thesis showed that by trapping the substrate α -1,6-mannobiose in the -1 and +1 subsites using the inactive enzyme (D220N), the α -mannosyl in the -1 subsite was indeed now observed in an 0S_2 Michaelis complex conformation. This is consistent with the computational work whereby the 0S_2 conformation was predicted to be preferential for α -mannosyl (requires the lowest input of energy) in contrast to the 4C_1 conformation for 1-thio- α -mannopyranose.

Evidence for a boat transition state was provided by visualising the conformation of a powerful transition state inhibitor mannoimidazole. The inhibitor was also observed in the +1 subsite suggesting the requirement for a larger inhibitor. Although mannoimidazole requires lateral protonation, which cannot be provided by GH125 enzymes, it bound and inhibited the enzyme with millimolar affinity. Thus, we were able to propose a conformational itinerary of ${}^0S_2 \rightarrow [B_{2,5}]^\ddagger \rightarrow {}^1S_5$.

In addition to defining the conformational itinerary for GH125, one can imagine the work leading to potential enzyme inhibitors. Since monosaccharides provide nutrition for pathogenic bacteria, a staple for colonisation, GH125 enzymes in combination with GH38 (α ,1,3- α -1,6-) mannosidases are virulence factors for *S. pneumoniae* and *C. perfringens*. The enzymes catalyse the final step in the N-glycan degradation pathway to provide complete depolymerisation. Inhibition of this pathway could prevent the bacterial colonisation initiating or the growth of the colonies to ultimately reduce the risk of infection and severe diseases. That said, with a few exceptions, synthetic approaches to generate conformationally-restricted inhibitors for mannosidases have not been greatly successful.

To complete the visualisation of the conformational pathway via structural biology techniques, the final piece of evidence needed would be to capture the 1S_5 product conformation. Noeuromycin, an azasugar designed against isofagamine and deoxymannojirimycin, was observed in a 1S_5 skew-boat conformation when in a complex with Golgi mannosidase II (GH38) (Kuntz et al., 2008; Liu et al., 2001). Isoquinuclidine was tested for inhibition of *BfMan2a* where it was also observed in a 1S_5 conformation (Tailford et al., 2008). Since GH125 enzymes seem to prefer a disaccharide (or a carbohydrate attached to a non-sugar moiety) isoquinuclidine may be preferred over noeuromycin. Using this information, tight binding inhibitors could be designed that traverse the -1 and +1 subsites in the active site.

One classical, conformationally-constrained, inhibitor that specifically inhibits a mannosidase family, is kifunensine. Chapter 3 examines the binding of kifunensine to family GH47 inverting α -1,2-mannosidases. Kifunensine is a natural product produced by Actinobacteria; they naturally synthesise many compounds that have been utilised in human medicine. In this thesis, kifunensine was shown to inhibit α -1,2-mannosidases by binding in a “ring-flipped” 1C_4 product conformation. Computational experiments showed that this was the lowest energy conformer and one that matches the conformational 3S_1 -[3H_4] \rightarrow 1C_4 itinerary used by these enzymes. Kifunensine is selective and potent because of its shape rather than charge. Given the power of transition-state charge mimics, it would be advantageous to have conformationally-restricted shape mimics.

Further work in Chapter 3 studied binding of 1-deoxymannojirimycin to GH47 enzymes, at atomic resolution; the resolution being essential to deconvolute the two conformations in the active site, 3S_1 and 1C_4 . Since it is not conformationally restricted, DMJ binds 12-fold less tightly than kifunensine. Derivatives of DMJ were synthesised to be constrained to a 1C_4 conformation in solution at low pH. However, they contained bulky groups causing steric hinderance with residues of the active site. The active site of the enzyme should be taken into consideration when designing future inhibitors, for example, by using molecular modelling. Prior to synthesis of the molecule, a model could be built into the active site of an enzyme to see whether any groups would clash with enzyme residues. Clearly there is much work to be done to obtain clinically relevant α -mannosidase inhibitors, in contrast, one glycobiological target now has inhibitors in advanced clinical trials - the human O-GlcNAcase, *HsOGA*.

O-GlcNAc is a dynamic post-translational modification to nucleocytoplasmic proteins attached by O-GlcNAc transferase and cleaved by O-GlcNAc hydrolase (the target for study in Chapters 4 and 5). Chapter 4 focussed on different constructs of OGA in order to provide better crystals and some insight into the role of phosphorylation and substrate binding. Mimicking phosphorylation of Ser364 by casein kinase II was shown that it does not affect the structure or catalytic activity of the protein, suggesting that the modification must play another role, perhaps in controlling mitotic expression levels during the cell cycle (Dephoure et al., 2008).

Prior to this thesis, four structures of *HsOGA* were obtained, each utilising a different construct design and expression strategy, to understand substrate recognition, regulation and the clinical value of inhibitor classes (Elsen et al., 2017; Li et al., 2017a; Roth et al., 2017). The *HsOGA* construct designed by Roth et al., 2017 (*HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅) separates the catalytic domain and the stalk domain on two different plasmids (11-396 and 535-715). Crystallisation of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ requires the C-terminus (residues 707-715) to bind into the active site of a symmetry related monomer. In my work, the surface entropy reduction construct, *HsOGA*_{E602AE605A}, was used to obtain better diffracting crystals, without great success. Although, due to different crystal packing, the path of the C-terminus into the active site of the symmetry related monomer was revealed. Tyr715 had moved across and out of the binding site so that the nitrogen of Lys713 was situated in the same position as the nitrogen of the C2 *N*-acetyl group.

In contrast to our work, the single construct designed by Li et al., 2017 (*HsOGA*_{linker}) ends at residue 704, therefore, the C-terminus does not bind into the active site (60-401/(GGGGS)₂/552-704). They were able to obtain five peptide structures that displayed the ability of OGA to bind bidirectional peptide substrates. In Chapter 4, I further sought to obtain complexes with an S-GlcNAcylated peptide of casein kinase II, whereby serine was mutated to a cysteine. The structure corroborates and joins the family of peptides: TAB1, lamin B1 and p53, which bind in a “forward” direction (N to C).

Superposition of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ and *HsOGA*_{linker} and *HsOGA*_{E602AE605A} showed an overlap of residues between Phe391 and Pro540, Figure 6.1a. The residues Gln537 to Pro540 of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ loosely overlay with Ser398 to Arg400 of *HsOGA*_{linker} in terms of backbone positioning. Also, Phe391 to Gln396 of *HsOGA*_{linker} and *HsOGA*_{E602AE605A} overlay well. Utilising this information, a new “fusion” construct was designed from 60-398/538-704 (*HsOGA*_{fusion}) to avoid the C-terminus blocking the active site. It does not contain a linker between the two domains, instead it fuses the N-terminal catalytic domain directly to the C-terminal stalk domain, Figure 6.1b. This construct will be used in the future to obtain more peptide substrate structures providing further knowledge into substrate recognition.

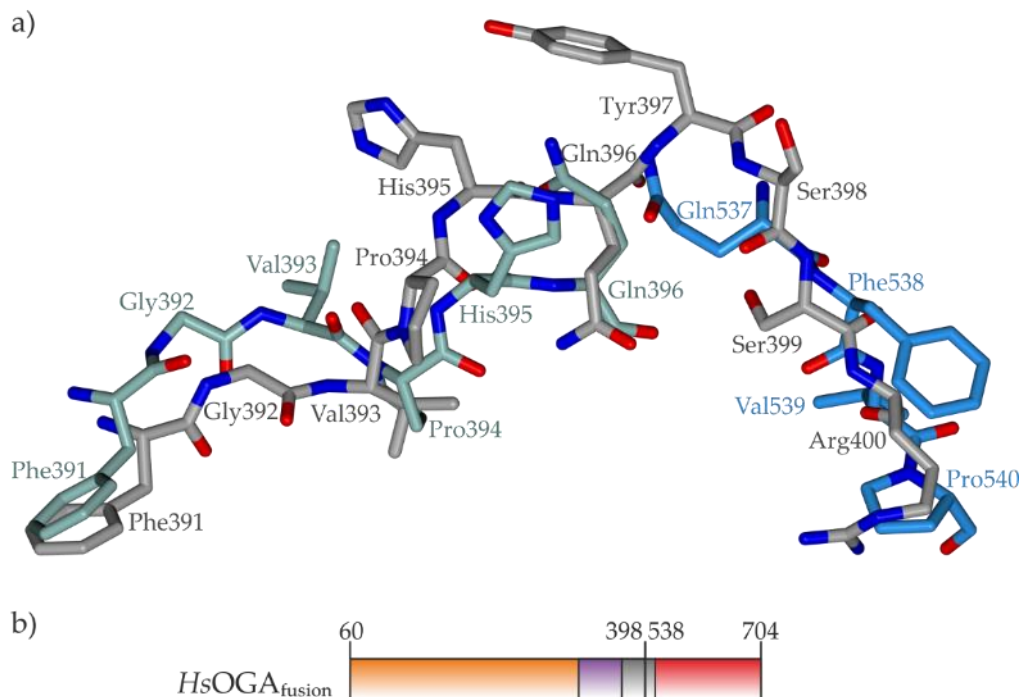


Figure 6.1. Fusion construct design. Overlay of *HsOGA*₁₁₋₃₉₆₊₅₃₅₋₇₁₅ (PDB ID:5M7R) in blue, *HsOGA*_{E602AE605A} (PDB ID: 6HKI) in green and *HsOGA*_{linker} (PDB ID: 5VVO). in grey.

Inhibition of OGA has been shown to reduce the prevalence of pathogenic tau aggregates by reducing the availability of phosphorylation sites. This occurs at sites that are both O-GlcNAcylated and phosphorylated and at phosphorylation sites that are adjacent to O-GlcNAcylation sites. Thiamet-G, PUGNAc, NButGT, STZ, NAG-thiazoline and GlcNAcstatin C are potent inhibitors, however, some show selectivity towards GH20 hexosaminidases as well. Analogues of these inhibitors have been synthesised and examined for inhibition *in vitro*, but the clinical *in vivo* data for these analogues have been lacking. The development of novel OGA inhibitor classes, described in chapter 5 has provided evidence, not only, for the potent and selective inhibition of OGA, but also the application of these inhibitors in diseases other than neurological diseases. The data provided here is clear evidence that these triazole library of iminocyclitols and bicyclic pyrrolizidine inhibitors are suitable for clinical trials, like MK-8719, after collecting *in vivo* data within single and multi-cellular organisms. As well as producing a potent inhibitor, it also needs to be selective. The compounds should be tested against other GH20 enzymes, HexA/B, for inhibition. The aim would be to have greater than 100-fold specificity to OGA (the specificity of the original pyrrolidine-based iminocyclitols was not described by Bergeron-Brlek and colleagues). The triazole inhibitors could then be

compared to Thiamet-G activity, currently the most frequently used OGA inhibitor. The initial data for the triazole pyrrolidine inhibitors is encouraging and the pi stacking interactions with aromatic active site residues reveals new interactions to exploit and explore.

Beyond the search for new inhibitors is the examination of how and why OGA interacts with partner proteins, for example fatty acid synthase, filamin A, heat shock cognate 70 kDa and OGT. When fatty acid synthase is overexpressed during oxidative stress, the levels of O-GlcNAcylation in the cell is increased due to an ~85% decrease in activity of OGA (Groves et al., 2017). Also, OGT and OGA act in an equilibrium, whereby in patients with X-linked intellectual disability, protein levels of mutated OGT (L254F-OGT) are reduced and a repressive complex (OGT-mSin3A-HDAC1) binds at the promoter region of *Oga*. This results in a decrease in *Oga* mRNA levels and no change to global O-GlcNAc levels (Vaidyanathan et al., 2017). For a review highlighting the recent accomplishments on the regulatory mechanisms of OGT and OGA see the appendix page 266.

Common to all published OGA structures, is the removal of the C-terminal “pseudo” histone acetyltransferase domain, pHAT. pHAT is speculated to play a role in transcription activation, to check the correct fold of an O-GlcNAcylated substrate and to guide it to the active site of the hydrolase domain. It acts by catalysing the acetylation of either serotonin or another small molecule, rather than histones. Of the two published bacterial orthologues, from *Oceanicola granulosus* and *Streptomyces sviveus*, less than 30% of the sequence is identical to that of the human sequence. Therefore, the structure and role of the bacterial homologues may not be comparable. One major structural and functional direction would be to study the animal HAT domains, especially in light of recent work. The pHAT domain of OGA was shown to acetylate pyruvate kinase M2 making it a target for O-GlcNAcylation by OGT. This modification inhibits pyruvate kinase M2 activity and, under high glucose conditions, promotes cancer through aerobic glycolysis and tumour growth (Singh et al., 2019). Hence, I initiated studies to obtain a soluble HAT domain from a functional OGA.

The production of *HspHAT* is hindered by protein precipitation once purified. Therefore, three eukaryotic species of pHAT were chosen and tested for protein production. *Mus musculus* has 99.5% sequence identity (a mutation of residue 742 from leucine (*HspHAT*))

to phenylalanine is found on the surface away from the active site (inferred from the HspHAT model). *Danio rerio* and *Anopheles gambiae* HATs have 76% and 33% sequence identity to HspHAT, Figure 6.2.

Although the same level of expression of the gene encoding *MmpHAT* from *E. coli* was observed, the same problem of protein precipitation occurred, Figure 6.3. The level of protein production for the *D. rerio* and *A. gambiae* was lower but the protein did not precipitate. In future work, these constructs should be investigated further by structural biology and by substrate analysis assays.



Figure 6.2. Sequence alignment of mammalian HAT constructs. Residues are coloured by their property: small in red, acidic in blue, basic in blue, hydroxyl, sulfhydryl, amine and glycine in green.

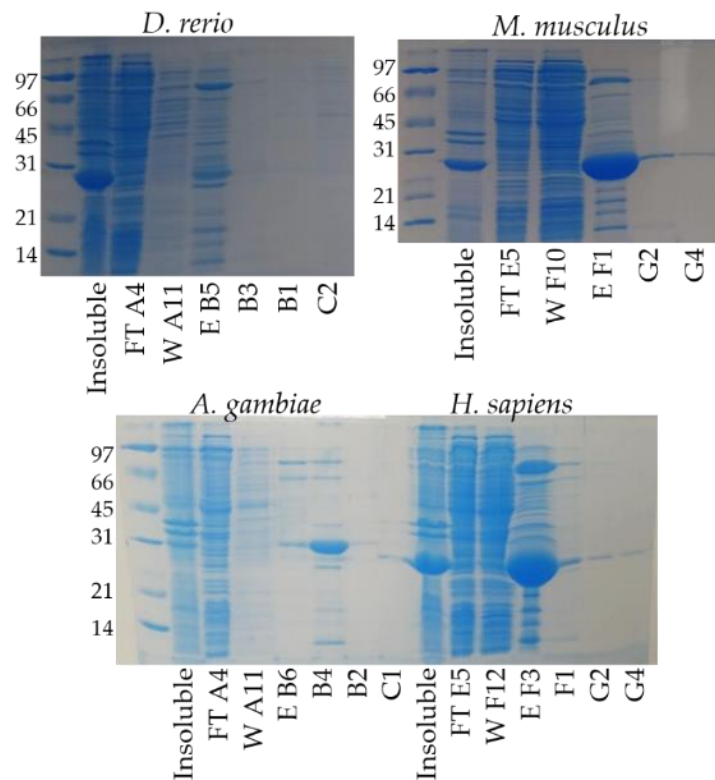


Figure 6.3. Production of different mammalian HAT constructs. SDS-PAGE gels of fractions from the purification of *D. rerio*, *M. musculus*, *A. gambiae* and *H. sapiens* HAT from *E. coli*. The insoluble fraction from the pellet, flow-through (FT) from a nickel affinity column, wash (W) with 20 mM imidazole and elution (E) by 500 mM imidazole over several fractions.

In conclusion, structural studies of the inhibition of glycoside hydrolase enzymes (linked in this case by their unusual features at C2) provides insight into the catalytic mechanism, conformational itinerary of the enzyme and, importantly, informs future drug design. The importance of developing and researching potent, conformationally-restricted, selective and pharmacologically relevant inhibitors will one day lead to the application of these inhibitors for disease treatment.

Appendix

Materials and methods

Strains

Table 3.i. Strains of bacteria used in the thesis for cloning and expression.

Strain	Genotype
NEB 5-alpha competent <i>E. coli</i>	Efficient transformation of unmethylated DNA (<i>hsdR</i>). Activity of nonspecific endonuclease I (<i>endA1</i>) eliminated. Resistance to phage T1 (<i>fhuA2</i>). α -complementation of the β -galactosidase gene. Reduced recombination of cloned DNA (<i>recA1</i>). DH5 α TM derivative.
<i>E. coli</i> BL21 (DE3)	(F ⁻ <i>ompT</i> r _B ⁻ m _B ⁻) + bacteriophage DE3
<i>E. coli</i> Tuner (DE3)	F ⁻ <i>ompT</i> <i>hsdS_B</i> (r _B ⁻ m _B ⁻) <i>gal dcm lacY1</i> (DE3)

Plasmids

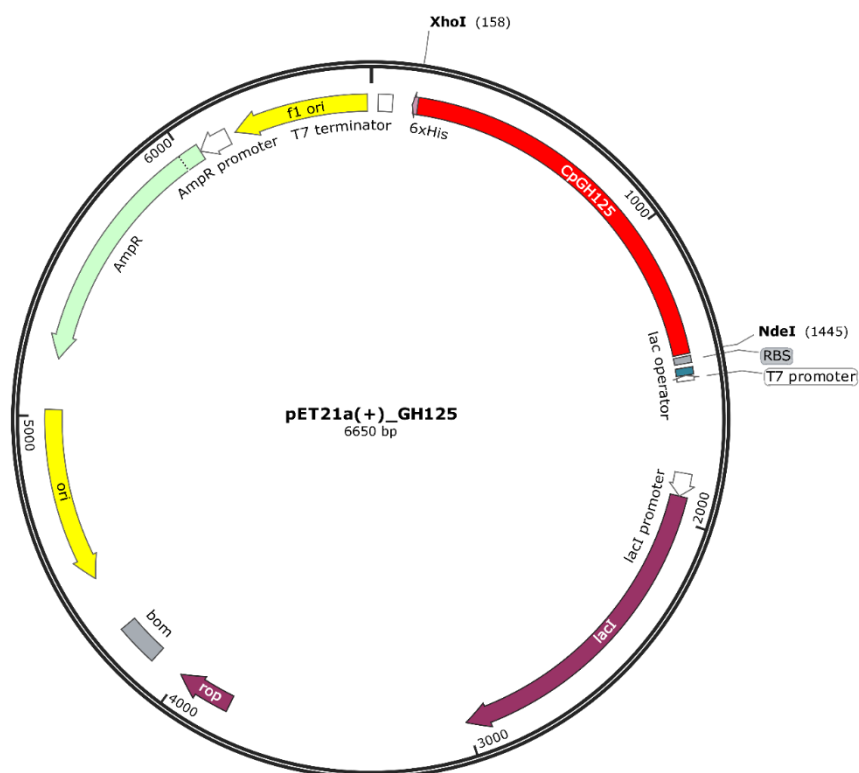


Figure 3.i. *CpGH125* cloned into pET21a vector. The pET21a(+) vector controls expression by the *lac* operon and is selected for against ampicillin (AmpR). The *CpGH125* gene is highlighted in red and was inserted between the NdeI and XhoI restriction sites (5'-3').

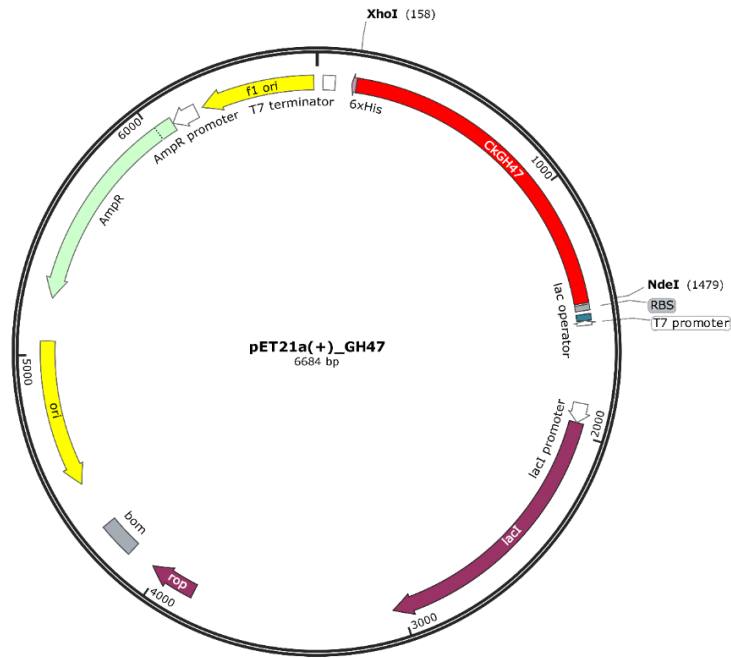


Figure 3.ii. *CkGH47* cloned into pET21a vector. The pET21a(+) vector controls expression by the lac operon and is selected for against ampicillin (AmpR). The *CkGH47* gene is highlighted in red and was inserted between the NdeI and XhoI restriction sites (5'-3').

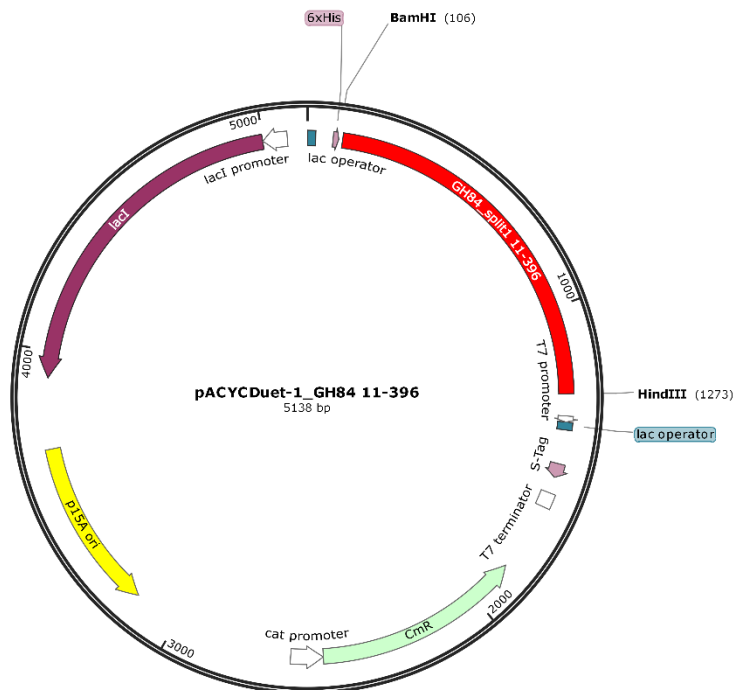


Figure 3.iii. *HsOGA*₁₁₋₃₉₆ cloned into pACYCDuet-1 vector. The pACYCDuet-1 vector controls expression by the lac operon and is selected for against chloramphenicol (CmR). The gene is highlighted in red and was inserted between the BamHI and HindIII restriction sites (5'-3').

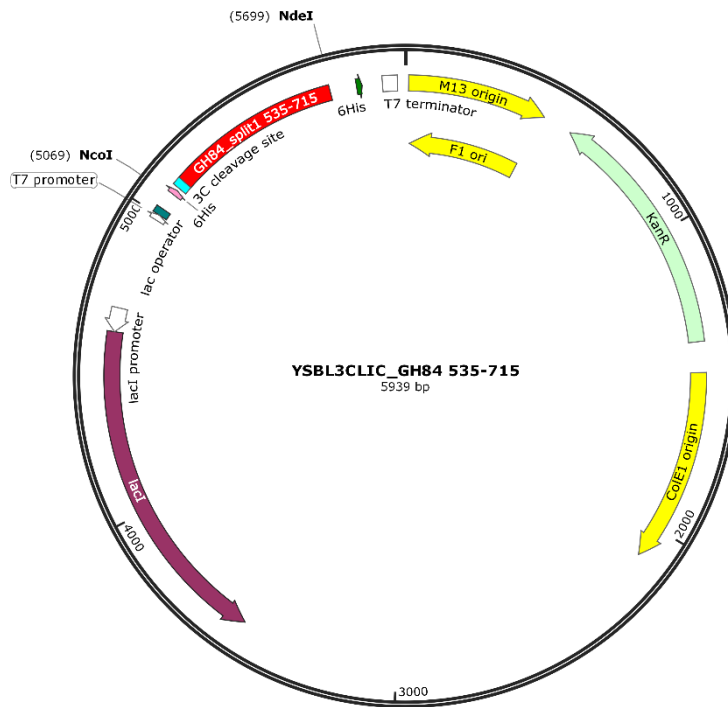


Figure 3.iv. *HsOGA*₅₃₅₋₇₁₅ cloned into YSBL3CLIC vector. The YSBL3CLIC vector controls expression by the lac operon and is selected for against kanamycin (KanR). The gene is highlighted in red and was inserted between the NcoI and NdeI restriction sites (5'-3').

Protein sequence

CpGH125

atgtctctgtctaccaacgaactgaaagaaatcgttcgtaaaatcggtaaagacctgtctggtaaaatc
M S L S T N E L K E I V R K I G K D L S G K I
gaagacaaaaaactgcaggaactgttctacaactgcttcatcaacaccatggacaccaccgttgaagtt
E D K K L Q E L F Y N C F I N T M D T T V E V
tctgaaggtagcgcggttcggttatcaccggtgacatccccggcgatgtggctgctgactctacctctcag
S E G D A F V I T G D I P A M W L R D S T S Q
gttgaacactacctgcccgttcggttaaagaatacccggaactgaaagcgatcttcaccggtctgatcaac
V E H Y L P F V K E Y P E L K A I F T G L I N
cgtcagggttaaatagcattctcatcgaccgtagcgaacgcggttcaacaagaaccgaacggtcagaaa
R Q V K C I F I D P Y A N A F N K E P N G Q K
tgggacaacgacatcaccaaagactctccgtgggttgggaacgtaaatcgaaatcgactctctgtgc
W D N D I T K D S P W V W E R K Y E I D S L C
taccggttcgctctgatccacaaatactggaaagaatctggtgacgaaaccttcttcaacgacgacatc
Y P V R L I H K Y W K E S G D E T F F N D D I
aaaaagcgttcaacatgatcatcgacctgtggcggttgaacagtaccaccgtgaaaaatctgactac
K K A F N M I I D L W R V E Q Y H R E K S D Y
tctttccagcgtctgaaactgctctgttaccgacaccctgtctcaggaaggtctgggtaccccggttacc
S F Q R L N C S V T D T L S H E G L G T P V T
tacaccggtatgacctggtctggttccgctccgctctgacgacgctgcaatacggttacctgatcccg
Y T G M T W S G F R P S D D A C E Y G Y L I P
gcaacatggttcgcggttgttgcgctgcggttacatctctgaaatcgcggaaaaagtttacaagacgaa
A N M F A V V A L R Y I S E I A E K V Y K D E
gaactgaaagaaaaagcggactctctgcggtgaagaaatcgacaacgcgatcgaaaaacacggtaaagtt

E L K E K A D S L R E E I D N A I E K H G K V
 tacaaagaaggttttcggtgaagtttacgcgtacgaaaccgacggtatgggtaactacaacttcatggac
 Y K E G F G E V Y A Y E T D G M G N Y N F M D
 gacgcgaacgtttccgtctctgctgtctatcccgtacctggaatacaaaggtatcgaagacgaagtttac
 D A N V P S L L S I P Y L E Y K G I E D E V Y
 cagaacacccgtaaattcatcctgtctaaaaacaaccgtttcttcttcgaaggtaaagcggcgaaaggt
 Q N T R K F I L S K N N R F F F E G K A A K G
 atcggttctccgcacaccccgaccagtagatctggcacatcgcgctgtctatgcagggtctgaccacc
 I G S P H T P D Q Y I W H I A L S M Q G L T T
 aacaaccaggaagaatcgaccagctgatcaactgctgaaagaaaccgacgcgggtaccgggttacatg
 N N Q E E I D Q L I K L L K E T D A G T G Y M
 cagaaggtttccacggttgacgacccgaccaaattcaccctgactggttcgcgctgactaactctg
 H E G F H V D D P T K F T R D W F A W S N S L
 ttctctcacttcatctacgaaaaagttatcaacaaaaaactcgagcaccaccaccaccactga
 F S H F I Y E K V I N K K L E H H H H H H -

*Green highlight indicates nucleotides/amino acids added by the vector.

Pink highlight indicates location of mutation CpGH125_{D220N}.

CkGH47 ManIA

atgaccgcttgcacctctcccgcgccacgcccctggccctggctcgccctcggccgcgccgctccggcc
 M T A S H L S R R H A L A L V A S A A A A P A
 ttccgcccattggctagcgaaccacccccggaagattggaaagcgtggcggccgatgtgcttctgaa
 F A A M A S E T T P E D W K A L A A D V R S E
 ttccagtgggcctggcagggttatgttgcaaaagcctggggcaaaagatgaaattaacctgtgagtgcc
 F Q W A W Q G Y V A K A W G K D E I N P V S G
 accagccgtagctttttcatcgaaggtcatgatctgggcctgagcctggtggaagcgtggatacgtg
 T S R S F F I E G H D L G L S L V E A L D T L
 tggattatgggctggatgcggaatttcaggccggtggtgattgggttaaagcaaactctgagctttgat
 W I M G L D A E F Q A G V D W V K A N L S F D
 gtggatggcaacgcgcaggttttcgaaaccaatccgctctggttggcggctctgctgagcgcacatctg
 V D G N A Q V F E T N I R L V G G L L S A H L
 gcgtctggatccggtgctgctggcaaaagcgcgtgatctggccgatcgccctggccaaagcattcgaa
 A S G D P V L L A K A R D L A D R L A K A F E
 gcatctccgcacggctctgcccgtggcgttatgtgaacctgcgacccggtgccggttagtgatccggaaacc
 A S P H G L P W R Y V N L R T G A V S D P E T
 aatctggccgaaattggcagctacctgagtgaaatttggcgttctgagccagctgaccggtgaaacgcaa
 N L A E I G T Y L S E F G V L S Q L T C G E R K
 tatttcgatatggccaaacgtgcaatgcgccatacgtggatcgctgcagtaaaaattggcctgatggca
 Y F D M A K R A M R H T L D R R S K I G L M A
 gcgaacatccacgcaatgaccggtgctgtttacgtctcgtaatgcgagtatcgatgtgtacgccgatagc
 A N I H A M T G A F T S R N A S I D V Y A D S
 ttttacgaatatctgtgggatgctggccctgttcggcgatgaagattgcaaacgctgggcggtggaa
 F Y E Y L W D A W A L F G D E D C K R W A V E
 tgtgttgatgctgagctggcccaccaggcaaaacggttatgatggctgcctgtggtttccgatggtggat
 C V D A Q L A H Q A K R Y D G R L W F P M V D
 ttgaaaccgggtgctgggttaccggtaccgcacagagcgaactggccgcatattatgcccgtctgctgggt
 F E T G A V T G T A Q S E L A A Y Y A G L L G
 caagtgggtcgtaaagcccagggcgatgattacctggcgagctttacctatctgcaggcgcaggttcggt
 Q V G R K A Q G D D Y L A S F T Y L Q A T F G
 gtgattccggaatctatcgatgttaccacgggccagccgctcgcaaacataaccggtctgcccggaa
 V I P E S I D V T T G Q P R R K H T G L R P E
 taccggatgctgctgacctgtggctgattgatcgctgacccgctaccgctgcctggccgcaatc
 Y P D A C L N L W L I D R D P R Y R L A A A I
 cactatcgtgaaatgaaagccacctctcgcgcagcgtttggctacacggcactgaaagatattaccag
 H Y R E M K A T S R A A F G Y T A L K D I T T
 cgtccgatgaccaggtatgataactgtccgggttattgggtgctgaaacagatgaaataactactacctg
 R P M T Q D D N C P G Y W W S E Q M K Y Y Y L
 ctgttttagtgataccccgcgatcgattatggccagctgcagctgagcaggaagcgaatgtgctgctg
 L F S D T P R I D Y G Q L Q L S T E A N V L R
 ggtttccgcaaagttctcgagcaccaccaccaccactga
 G F R K V L E H H H H H H -

*Yellow highlight indicates 26 amino acid signal sequence that was removed.

Green highlight indicates nucleotides/amino acids added by the vector.

Pink highlight indicates location of mutation *CkGH47*_{E427A}.

HsOGA

```
atgggcgtgcaaaaagaatctcaagctaccctggaagaacgtgaaagcgaactgtcgtcaaatccggcg
M G V Q K E S Q A T L E E R E S E L S S N P A
gcgtcagcgggtgcctctctggaaccgccagccgaccggcaccgggtgaagataatccggccgggtgcc
A S A G A S L E P P A A P A P G E D N P A G A
ggcgggtgcagctgtggcaggcgcggccggcggtgctcgtcgtcttctgtgcggtgtggttgaaggcttc
G G A A V A G A A G G A R R F L C G V V E G F
tatggctcgtccgtgggttatggaacagcgcgaaagaactgtttcgtcgcctgcaaaaatgggaactgaat
Y G R P W V M E Q R K E L F R R L Q K W E L N
acctatctgtacgcaccgaaagatgactataaacatcgtatggttctggcgcgaaatgtacagtggtgaa
T Y L Y A P K D D Y K H R M F W R E M Y S V E
gaagcgaacagctgatgaccctgatctccgcagctcgtgaatacgaatcgaattcatctacgccatc
E A E Q L M T L I S A A R E Y E I E F I Y A I
tctccgggctggatattaccttctccaaccgaaagaagtttcaacgctgaaacgtaaactggaccag
S P G L D I T F S N P K E V S T L K R K L D Q
gtctcgcgaatgtggtgcccgcagcttggcgtgctggttcgatgacatcgatcataacatgtgtgccc
V S Q F G C R S F A L L F D D I D H N M C A A
gacaaaagaagtttttagctcttctcgcacacgctcaagtcagtatcaccatgaaatattatcaatacctg
D K E V F S S F A H A Q V S I T N E I Y Q Y L
ggcgaaccgaaacgtttctgttctgcccgcaccgaatattgcccgcagttttgttaccggaatgtgtca
G E P E T F L F C P T E Y C G T F C Y P N V S
cagtcgccgatctgcccaccgttggcgaaaaactgctgccgggtattgaagtgtgtggacggggcccg
Q S P Y L R T V G E K L L P G I E V L W T G P
aaagtcgtgagcaaaagaatcccggctcgaatcgattgaagaagtgagcaaaattatcaaactgcccgg
K V V S K E I P V E S I E E V S K I I K R A P
ggtatctgggataaacattcatgccaatgattatgaccagaaacgctgtttctggcccgtacaaaggt
V I W D N I H A N D Y D Q K R L F L G P Y K G
cgttcaaccgaaactgatcccgcgctgaaaggtgctcctgacgaaccggaattgtgaattcgaagcgaac
R S T E L I P R L K G V L T N P N C E F E A N
tatgtggccattcacaccctggcagctggtacaaaacatgaatggcgtgctgtaagatggttctc
Y V A I H T L A T W Y K S N M N G V R K D V V
atgaccgattcagaagactcgacggttagcattcagatcaaactggaaaatgaaggtagccgatgaagac
M T D S E D S T V S I Q I K L E N E G S D E D
atcgaaacccgatgtgctgtattctccgcagatggcactgaaactggctctgacggaatggctgcaagaa
I E T D V L Y S P Q M A L K L A L T E W L Q E
tttggcgtcccgcacatcagctacagttcccgcgaagtggcccactctggcgcgaaagccagtggttgat
F G V P H Q Y S S R Q V A H S G A K A S V V D
ggtaaccgctggtcgcagctccgtctctgaacgcaaccacggctcgtgaccacgggtgatcaggaaccg
G T P L V A A P S L N A T T V V T T V Y Q E P
attatgtctcaaggtgccgcactgagtggtgaaccgaccacctgaccaagaagaagagaaaaaacag
I M S Q G A A L S G E P T T L T K E E E K K Q
ccggatgaagaaccgatggacatggttgcgaaaaacaggaagaaccgatcataaaaacgacaatcaa
P D E E P M D M V V E K Q E E T D H K N D N Q
atcctgtcagaaaattgtggaagcaaaaatggctgaagaactgaaaccgatggataccgacaagaatcg
I L S E I V E A K M A E E L K P M D T D K E S
attgcagaagaacaaatctccggaaatgtccatgcaggaattgcattctcagacattgctccgatgcag
I A E S K S P E M S M Q E A D C I S D I A P M Q
accgatgaacaaacgaacaaagaacagtttggttccgggcccgaatgaaaaaccgctgtataccgcccga
T D E Q T N K E Q F V P G P N E K P L Y T A E
ccggtcacgctggaagatctgcagctgctggccgacctgttttatctgccgtacgaacacggcccga
P V T L E D L Q L L A D L F Y L P Y E H G P K
ggtgacagatgctgctggaattccaatggctgcccgcctaactcatcgggtggttccggttaactgcaag
G A Q M L R E F Q W L R A N S S V V S V N C K
ggtaagatagcgaacaaatcgaagaatggcgttctcgcgcagctaaatgtgagaatgtgcccgtc
G K D S E K I E E W R S R A A K F E E M C G L
```

gtgatgggcatgttcacccgctctgtcaaaactgtgccaatcgacgattctgtatgatatgtattcgtac
 V M G M F T R L S N C A N R T I L Y D M Y S Y
 gtttgggacatcaaatctatcatgagatggttaaatacattcgtccagtggtggctgggctgctcgttcgcac
 V W D I K S I M S M V K S F V Q W L G C R S H
 agctctgcgagttcctgatcggatcaagaaccgtggcctttcgtggcggctctggcgggtgaattc
 Q F L I G D Q E P W S S A A F R G G L A G E F
 cagcgtctgctgccgattgatggtgccaacgacctgttttccagccgccgctgacccccgaccagc
 Q R L L P I D G A N D L F F Q P P P L T P T S
 aaagtgtataccatccgctccgtactttccgaaagatgaagcgtctgtgtataaaaatttgccgcgaaatg
 K V Y T I R P Y F P K D E A S V Y K I C R E M
 tacgatgacggcggttggtctgcccgttccagagccaaccggatctgattggcgacaaaactggttggcgggt
 Y D D G V G L P F Q S Q P D L I G D K L V G G
 ctgctgagctgtccctggattattgctttgtcctgggaagatgaagacggcatctgtggttacgcgctg
 L L S L S L D Y C F V L E D E D G I C G Y A L
 ggcaccgtcgatgtgacgcccgttcatcaaaaaatgtaaaatcagttggatcccgttcatgcaggaaaaa
 G T V D V T P F I K K C K I S W I P F M Q E K
 tataccaaaccgaacggtgataaagaactgagtgaaagcagaaaaaatcatgctgtcctttcatgaagaa
 Y T K P N G D K E L S E A E K I M L S F H E E
 caggaagtgtgcccggaaacctttctggcgaatttcccagacctgatcaaaatggatatccacaaaaaa
 Q E V L P E T F L A N F P S L I K M D I H K K
 gtgacggaccgctccggttgcgaaatcaatgatggcctgctgctgagttccctgaaagcaaatggcagt
 V T D P S V A K S M M A C L L S S L K A N G S
 cgcgggtgctttctgtgaagttcgtccggatgacaaacgcattctggaattttactccaaactgggtgtg
 R G A F C E V R P D D K R I L E F Y S K L G C
 ttcgaaatcgctaaaatggaaggcttcccgaaagatgtcgttatcctgggtcgtagtctgtaa
 F E I A K M E G F P K D V V I L G R S L -

*Yellow highlight indicates the N-terminal catalytic domain.

Green highlight indicates the C-terminal stalk domain.

Pink highlight indicates pseudo-histone acetyl transferase domain.

Primers

Table 3.ii. Primers used for mutating and subcloning the DNA of enzymes in this thesis. Lower case letters represent codons that were changed for the mutation. Codons in bold represent bases from the vector.

Primer name	Sequence (5'-3')	Chapter
CpGH125_{D220N} Forward	C CGT CCG TCT aat GAC GCG TGC G	2
CpGH125_{D220N} Reverse	AAA CCA GAC CAG GTC ATA CCG	2
CkGH47_{E427A} Forward	T TGG TGG TCT gcg CAG ATG AAA TAC	3
CkGH47_{E427A} Reverse	TAA CCC GGA CAG TTA TCA TC	3
HsOGA_{D175N} Forward	G CTG CTG TTC GAT aac ATC GAT CAT AA	
HsOGA_{D175N} Reverse	GCA AAG CTG CGG CAA CCA	
HsOGA_{E602A} Forward	T AAA GAT AGC gcg AAA ATC GAA GAA TGG	4
HsOGA_{E602A} Reverse	CCC TTG CAG TTA ACC GAA	4
HsOGA_{E602AE605A} Forward	A GAT AGC gct AAA ATC gct GAA TGG	4
HsOGA_{E602AE605A} Reverse	TTA CCC TTG CAG TTA ACC GAA	4
HsOGA_{S364D} Forward	A AAT GAA GGT gac GAT GAA GAC ATC	4
HsOGA_{S364D} Reverse	TCC AGT TTG ATC TGA ATG	4

Published structures of HsOGA

Table 3.iii. List of published constructs of OGA.

Abbreviation	Comment	Construct	Expression Plasmid	Mutation	PDB codes
HsOGA _{FL}	Full length protein	1-916	pET30a(+)		
HsOGA _{split}	Split expression construct Roth, C., et al (Roth et al., 2017)	11-396 + 535-715	pACYC-Duet (11-396) and pET-YSBLIC3C (535-715) (transformation into <i>E. coli</i>)		5M7R Apo 5M7S Thiamet G 5M7T PUGNAc 5M7U iminocyclitol
HsOGA _{E602AE605A}	Surface Entropy reduction mutant	11-396 + 535-715	pACYC-Duet (11-396) and pET-YSBLIC3C (535-715) (transformation into <i>E. coli</i>)	E602A + E605A	6HKI
HsOGA _{Linker}	Li, B., et al (Li et al., 2017a)	60-401 -15 residue Gly/Ser linker- 552-704	pETSUMO (transformation into <i>E. coli</i>)		5TKE Apo 5UN8 p53 peptide 5UN9 Thiamet G
HsOGA _{57-400/544-705}	Elsen, N. L., et al (Elsen et al., 2017)	14-400 + 554-705 57-400 + 544-705	pBAC-1 (Co-infection into <i>Sf21</i> cells)		5UHK Apo 5UHL Thiamet G 5UHO PUGNAc 5UHP Cat. domain

Table 3.iv. Data collection and refinement statistics for the published *HsOGA* crystal structures. ¹(Roth et al., 2017); ²(Li et al., 2017a); ³(Li et al., 2017b); ⁴(Elsen et al., 2017); ⁵(Males and Davies, 2019).

PDB code	Ligand	Space Group	Resolution (Å)	R _{merge}	R _{work} / R _{free}
5M7R ¹	Apo	<i>P</i> 4 ₃ 2 ₁ 2	96.3-2.4	0.10 (2.19)	0.22 / 0.25
5M7S ¹	Thiamet-G	<i>P</i> 4 ₃ 2 ₁ 2	72.1-2.4	0.12 (2.86)	0.20 / 0.23
5M7T ¹	PUGNAc- Imidazole	<i>P</i> 4 ₃ 2 ₁ 2	82.3-2.6	0.09 (2.28)	0.20 / 0.24
5M7U ¹	Iminocyclitol VV347	<i>P</i> 4 ₃ 2 ₁ 2	95.4-2.3	0.08 (2.70)	0.20 / 0.23
5TKE ²	Apo	<i>P</i> 1 2 ₁ 1	50.0-2.5	0.09 (0.72)	0.18 / 0.24
5UN8 ²	Glycosylated p53 peptide	<i>P</i> 1 2 ₁ 1	45.0-2.1	0.08 (0.72)	0.18 / 0.23
5UN9 ²	Thiamet-G	<i>P</i> 1 2 ₁ 1	30.0-2.5	0.06 (0.58)	0.21 / 0.26
5VVO ³	Apo	<i>P</i> 1 2 ₁ 1	50.0-2.6	0.08 (0.63)	0.21 / 0.26
5VVV ³	Crystallin B	<i>P</i> 1 2 ₁ 1	50.0-2.7	0.07 (0.94)	0.20 / 0.26
5VVT ³	ELK1	<i>P</i> 1 2 ₁ 1	50.0-2.8	0.08 (0.72)	0.21 / 0.29
5VVX ³	Lamin B1	<i>P</i> 1 2 ₁ 1	50.0-2.9	0.09 (0.69)	0.20 / 0.28
5VVU ³	TAB1	<i>P</i> 1 2 ₁ 1	50.0-2.7	0.08 (0.68)	0.21 / 0.28
5UHK ⁴	Apo (57-400 + 544-705)	<i>P</i> 4 ₁ 2 ₁ 2	34.0-3.0	0.08 (1.06)	0.19 / 0.25
5UHL ⁴	Thiamet-G (57- 400 + 544-705)	<i>P</i> 4 ₁ 2 ₁ 2	92.0-3.1	0.12 (1.42)	0.19 / 0.26
5UHO ⁴	PUGNAc (57- 400 + 544-705)	<i>P</i> 4 ₁ 2 ₁ 2	34.0-3.2	0.09 (1.07)	0.20 / 0.26
5UHP ⁴	Apo (14-400 + 554-705)	<i>P</i> 1	41.0-2.8	0.14 (0.51)	0.24 / 0.29
6HKI ⁵	Apo (E602AE605A)	<i>P</i> 3 ₁ 2 1	192.4-3.3	0.42 (2.51)	0.17 / 0.23

Table 3.v. Construct progress for *HsOGA* mutants. Green indicates a positive result and red indicates a negative result.

Construct	Mutation	Cloning	Expression	Purification	Crystallisation
11-396 + 535-715	S364D	Green	Green	Green	Green
11-396 + 535-715	E128A E130A	Green	Red	Green	Green
11-396 + 535-715	E239A K240A	Green	Green	Green	Green
11-396 + 535-694	E602A E605A	Green	Green	Green	Green
11-396 + 535-712	E602A E605A	Red	Green	Green	Green
11-396 + 535-715	E602A E605A	Green	Green	Green	Green
11-396 + 535-715	D174A D175A	Green	Green	Green	Green
11-396 + 535-715	D174A D175A E602A E605A	Green	Green	Green	Green

Phosphorylation of *HsOGA*

Table 3.vi. The masses of fragments predicted to contain a site of phosphorylation. Site of phosphorylation is highlighted in red.

Fragment	Sequence	[M+H] ¹⁺	[M+H] ²⁺	[M+H] ³⁺
1	EMY S VEEAEQLMTLISAAR			
	Unphosphorylated	2172.5	1086.7	724.8
	Phosphorylated 1 +80	2252.5	1126.7	751.5
2	EIPVESIEEV S K			
	Unphosphorylated	1359.5	680.3	453.8
	Phosphorylated 1 +80	1439.5	720.75	480.8
4	LLPIDGANDLFFQPPPL T P T S K			
	Unphosphorylated	2382.8	1191.9	794.9
	Phosphorylated 1 +80	2462.8	1232.4	821.9
	Phosphorylated 2 +160	2542.8	1271.9	848.3

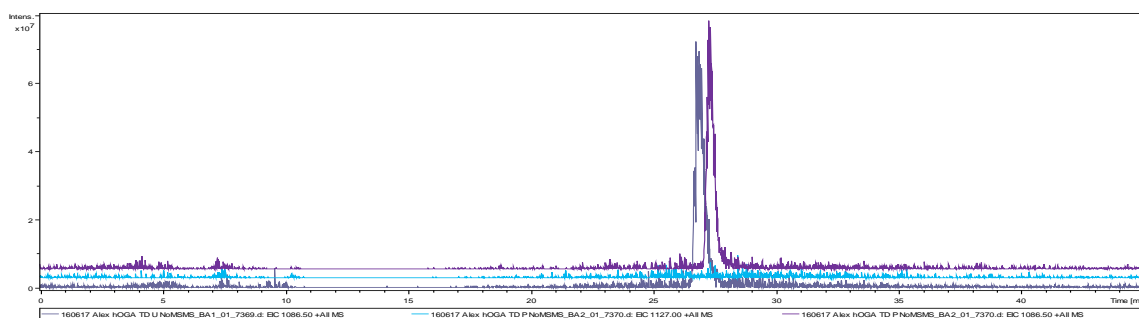


Figure 3.iv. Overlay of fragment 2 from phosphorylated and unphosphorylated samples. The grey chromatogram shows the unphosphorylated sample, the blue and purple chromatograms show the phosphorylated and unphosphorylated proportions of the sample incubated with CKII, respectively.

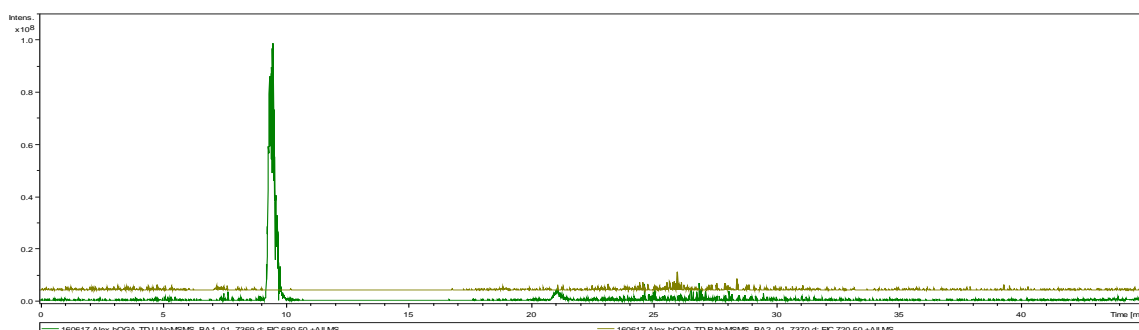


Figure 3.v. Overlay of fragment 3 from phosphorylated and unphosphorylated samples. The green chromatogram shows the unphosphorylated sample, the brown chromatogram shows the phosphorylated proportion of the sample incubated with CKII.

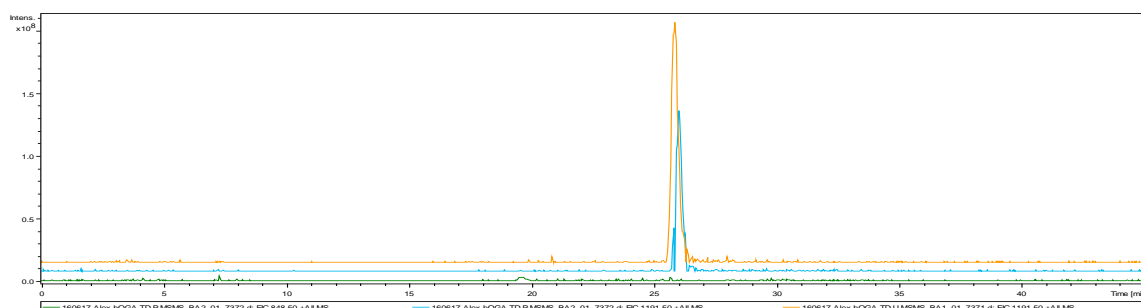


Figure 3.vi. Overlay of fragment 3 from phosphorylated and unphosphorylated samples. The yellow chromatogram shows the unphosphorylated sample, the green and blue chromatograms show the phosphorylated and unphosphorylated proportions of the sample incubated with CKII, respectively.

Michaelis-Menten kinetics

The MM model defines the reaction of enzyme and substrate as the reversible formation of to the enzyme-substrate, Michaelis complex, then irreversible formation of the free enzyme and product; $E + S \rightleftharpoons ES \rightarrow E + P$, where E is enzyme, S is substrate and P is product. The reaction rate is dependent on the substrate concentration and is quantified by the rate constants k_{on} (forward rate), k_{off} (reverse rate) and k_{cat} (catalytic rate, decomposition of the ES). Steady-state approximations assume that the concentration of ES is constant when the rate of formation and breakdown of the ES are equal. By plotting V_0 (initial velocity experimentally measured after a short time period assumes ES has formed but the substrate concentration is constant) against varying substrate concentrations, $V_0 = \frac{V_{max} \cdot [S]}{K_M + [S]}$. V_{max} is the maximum rate when the enzyme is saturated with substrate. K_M is concentration of substrate at half the V_{max} , where a small K_M indicates a high affinity. K_{cat}/K_M is the measure of catalytic efficiency of conversion of substrate into product.

Mutation studies

Table 3.vii. Kinetics data for different HsOGA constructs from published data.
Michaelis-Menten results for the K_M when either using *p*NP-GlcNAc as the substrate or p53 glycopeptide.

Construct	pNP-GlcNAc K_M (μ M)	k_{cat} (s^{-1})	p53 glycopeptide K_M (μ M)
<i>HsOGA</i> ₁₋₉₁₆	170 ± 10	1.80 ± 0.10	
<i>HsOGA</i> ₁₋₉₁₆	88 ± 7	1.04 ± 0.02	
<i>HsOGA</i> ₁₋₉₁₆	92 ± 5	0.37 ± 0.01	
<i>HsOGA</i> ₁₋₆₆₂	810 ± 30	0.10 ± 0.01	
<i>HsOGA</i> ₁₄₋₃₄₁	4000 ± 1000	0.13 ± 0.04	
<i>HsOGA</i> ₁₄₋₄₀₀	2100 ± 300	0.08 ± 0.01	
<i>HsOGA</i> ₁₄₋₄₀₀₊₅₅₄₋₇₀₅	70 ± 30	0.20 ± 0.02	
<i>HsOGA</i> ₅₇₋₄₀₀₊₅₄₄₋₇₀₅	40 ± 10	1.50 ± 0.10	
<i>HsOGA</i> ₆₀₋₄₀₁₊₅₅₂₋₇₀₄	30 ± 3	0.73 ± 0.02	796 ± 57
<i>HsOGA</i> ₆₀₋₄₀₁₊₅₅₂₋₇₀₄ F223A			3840 ± 86
<i>HsOGA</i> ₆₀₋₄₀₁₊₅₅₂₋₇₀₄ W679A			2639 ± 116
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₇₁₅	41 ± 3	0.37 ± 0.01	
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₇₁₅ E677A	65 ± 4	0.27 ± 0.01	
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₇₁₅ S649A	29 ± 2	0.40 ± 0.01	
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₇₁₅ W678A	113 ± 7	0.33 ± 0.01	
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₆₆₃	N/A	N/A	
<i>HsOGA</i> ₁₁₋₃₉₆₊₅₃₅₋₆₉₄	195 ± 20		
<i>HsOGA</i> _{10-(410 535)-694*}	308 ± 24		
<i>HsOGA</i> _{Isoform3}	N/A	N/A	

Circular dichroism equation for mean residue ellipticity

$$MRE = \frac{\theta_{obs}}{10 \times n \times C \times l}$$

n = number of amino acids

C = Concentration (molar)

l = length of light path

Supplemental information

Chapter 2

Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α -Mannosidases. DOI: 10.1021/jacs.6b11247.

https://pubs.acs.org/doi/suppl/10.1021/jacs.6b11247/suppl_file/ja6b11247_si_001.pdf

Chapter 3

Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Structural Approach. DOI: 10.1002/cbic.201700166.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fcbic.201700166&file=cbic201700166-sup-0001-misc_information.pdf

Conformational Behaviour of Azasugars Based on Mannuronic Acid. DOI: 10.1002/cbic.201700080.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fcbic.201700080&file=cbic201700080-sup-0001-misc_information.pdf

Chapter 4

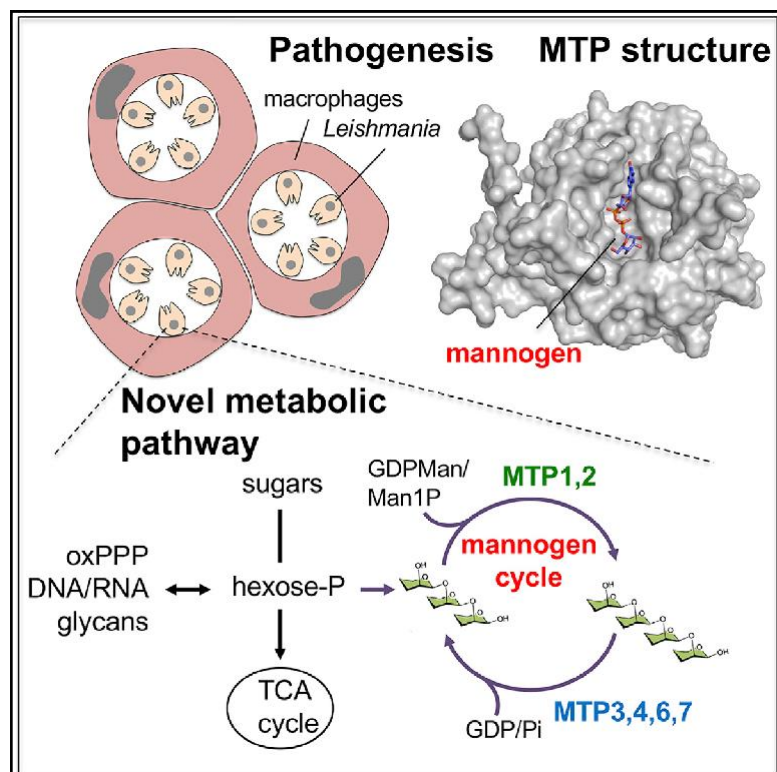
Structural studies of a surface-entropy reduction mutant of O-GlcNAcase. DOI: 10.1107/S2059798318016595.

<http://journals.iucr.org/d/issues/2019/01/00/jc5018/jc5018sup1.pdf>

Cell Host & Microbe

A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in *Leishmania* Parasites

Graphical Abstract



Authors

M. Fleur Sernee, Julie E. Ralton, Tracy L. Nero, ..., Spencer J. Williams, Gideon J. Davies, Malcolm J. McConville

Correspondence

malcolmm@unimelb.edu.au

In Brief

Sernee et al. show that metabolism of the *Leishmania* carbohydrate reserve, mannogen, is mediated by a single family of enzymes acquired by horizontal gene transfer. The evolution of this pathway may have allowed these parasites to expand their host range and colonize intracellular niches in their vertebrate hosts.

Highlights

- *Leishmania* have replaced canonical carbohydrate reserves with mannogen
- Mannogen cycling is regulated by dual-activity sugar transferase/phosphorylases
- Enzyme evolution required horizontal gene transfer and enzyme substrate diversification
- Mannogen cycling protects against glucose toxicity and is essential for virulence

Sernee et al., 2019, Cell Host & Microbe 26, 385–399



September 11, 2019 © 2019 The Authors. Published by Elsevier Inc.

<https://doi.org/10.1016/j.chom.2019.08.009>

CellPress

A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in *Leishmania* Parasites

M. Fleur Sernee,^{1,6} Julie E. Ralton,^{1,6} Tracy L. Nero,^{1,2} Lukasz F. Sobala,³ Joachim Kloehn,¹ Marcel A. Vieira-Lara,¹ Simon A. Cobbold,¹ Lauren Stanton,¹ Douglas E.V. Pires,¹ Eric Hanssen,^{1,4} Alexandra Males,³ Tom Ward,³ Laurence M. Bastidas,³ Phillip L. van der Peet,⁵ Michael W. Parker,^{1,2} David B. Ascher,¹ Spencer J. Williams,⁵ Gideon J. Davies,³ and Malcolm J. McConville^{1,7,*}

¹Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia

²ACRF Rational Drug Discovery Centre, St. Vincent's Institute of Medical Research, Fitzroy, VIC 3065, Australia

³Department of Chemistry, University of York, York YO10 5DD, UK

⁴Advanced Microscopy Facility, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia

⁵School of Chemistry, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: malcolmm@unimelb.edu.au <https://doi.org/10.1016/j.chom.2019.08.009>

SUMMARY

Parasitic protists belonging to the genus *Leishmania* synthesize the non-canonical carbohydrate reserve, mannogen, which is composed of β -1,2-mannan oligosaccharides. Here, we identify a class of dual-activity mannosyltransferase/phosphorylases (MTPs) that catalyze both the sugar nucleotide-dependent biosynthesis and phosphorolytic turnover of mannogen. Structural and phylogenetic analysis shows that while the MTPs are structurally related to bacterial mannan phosphorylases, they constitute a distinct family of glycosyltransferases (GT108) that have likely been acquired by horizontal gene transfer from gram-positive bacteria. The seven MTPs catalyze the constitutive synthesis and turnover of mannogen. This metabolic rheostat protects obligate intracellular parasite stages from nutrient excess, and is essential for thermotolerance and parasite infectivity in the mammalian host. Our results suggest that the acquisition and expansion of the MTP family in *Leishmania* increased the metabolic flexibility of these protists and contributed to their capacity to colonize new host niches.

INTRODUCTION

The Trypanosomatidae are parasitic protists that cause diseases such as human African trypanosomiasis (*Trypanosoma brucei*), Chagas disease (*T. cruzi*), and human leishmaniasis (*Leishmania spp*) that are estimated to chronically or acutely infect more than 100 million people worldwide (Stuart et al., 2008; Singh et al., 2014; Burza et al., 2018; Ban˘uls et al., 2011). These pathogens often have complex life cycles, occu-

pying extracellular and intracellular niches in different insect, plant, and vertebrate hosts. Comparative analyses of the genomes of these parasites and related free-living protists indicate that the evolution of parasitic lifestyles has been associated with loss of metabolic pathways and genome reduction, as well as the acquisition of lineage-specific innovations, that have contributed to parasite survival in different host niches (Jackson et al., 2016; Opperdoes et al., 2016). The absence of canonical reserve poly- or oligosaccharides (e.g., glycogen, starch, sucrose, trehalose) in the trypanosomatids constitutes a particularly dramatic example of metabolic pathway loss. The dynamic synthesis and turnover of carbohydrate reserves in other eukaryotes plays a key role in regulating central carbon metabolism and cellular stress responses (Franc, ois and Parrou, 2001; Shi et al., 2010; MacNeill et al., 2017), suggesting that early trypanosomatids may have developed compensatory mechanisms for dealing with variable nutrient levels and other environmental stresses (Jackson et al., 2016). Intriguingly, *Leishmania* and closely related trypanosomatids (including the plant pathogen *Phytophthora* and insect parasites *Leptomonas*, *Crithidia*, and *Herpetomonas*) have subsequently acquired a new pathway of carbohydrate reserve biosynthesis (Ralton et al., 2003; Sernee et al., 2006). These parasites synthesize mannogen, a family of linear oligosaccharides made up of 2-60 β -1,2-linked mannose residues (Previato et al., 1984; Keegan and Blum, 1992; Ralton et al., 2003). The absence of mannogen in the *Trypanosoma* suggests that this pathway evolved following the separation of these different trypanosomatid lineages. How the acquisition of mannogen synthesis contributes to the parasitic lifestyle of *Leishmania* remains unclear.

Mannogen accumulates to very high levels (>10 mM) in pathogenic stages of *Leishmania* (Ralton et al., 2003). These include the non-dividing promastigote stages that develop in the foregut of the sandfly vector and initiate infection in the mammalian host and the obligate intracellular amastigote stages that colonize the phagolysosome of macrophages and other phagocytic cells in the mammalian host and perpetuate both acute and long-term

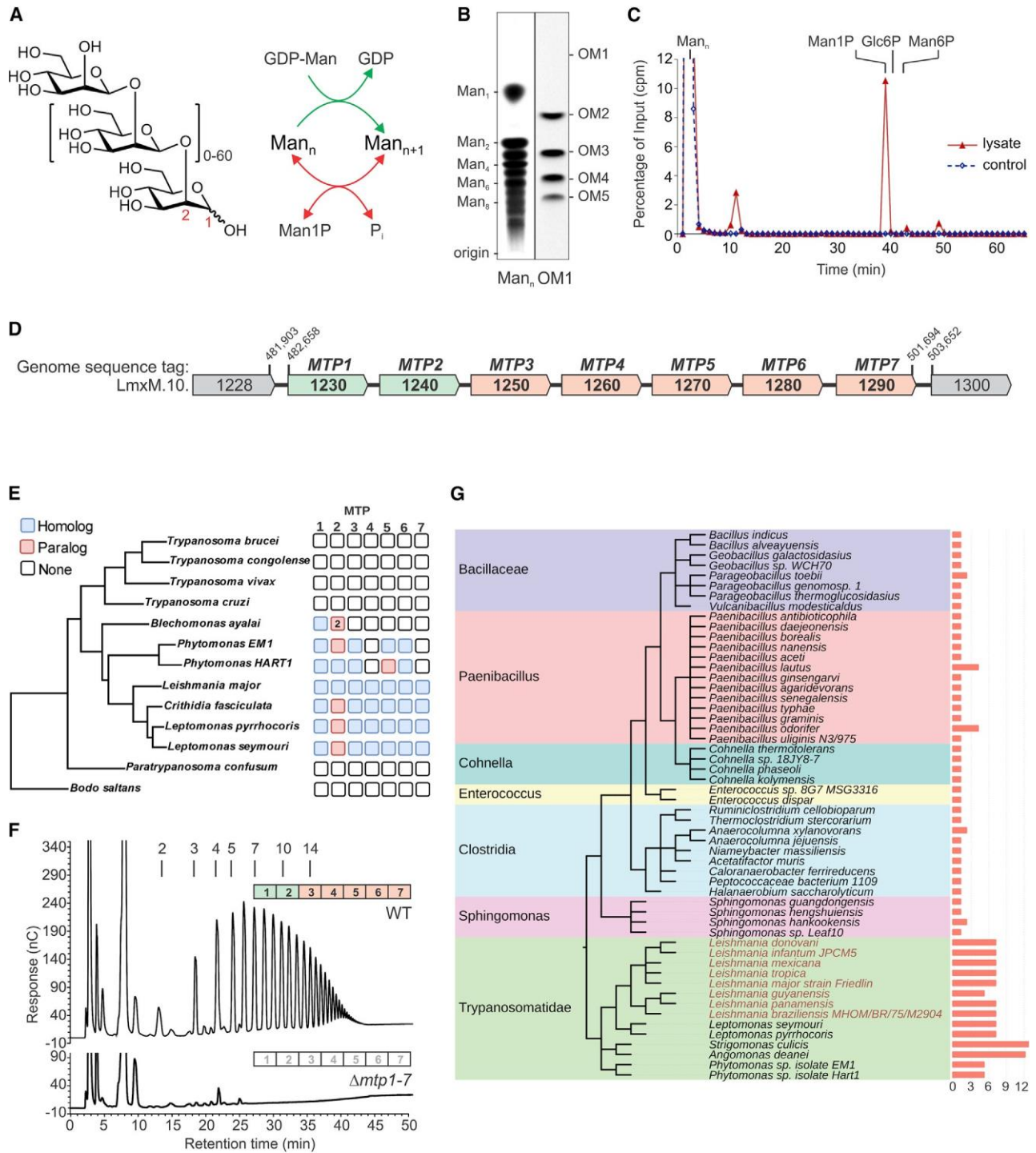


Figure 1. Identification of a *Leishmania* Gene Family Required for Mannogen Synthesis

(A) Putative role of GDP-Man-dependent transferase(s) and glycan phosphorylase(s) in mannogen cycling.

(B) HPTLC separation of ^3H -mannogen (DP 2–10) or octyl- ^3H -mannogen (DP 1–5) oligomers generated by incubation of *L. mexicana* Dgmp cell-free lysates with GDP- ^3H -Man and purified mannogen (Man_n) or the synthetic mannose acceptor OM1.

(C) HPAEC detection of ^3H -Man1P released after incubation of *L. mexicana* Dpmm cell-free lysates (and boiled control) with ^3H -mannogen (Man_n).

(D) The *L. mexicana* MTP1-7 gene array on chromosome 10.

(legend continued on next page)

chronic infections in humans. While the enzymes involved in mannogen synthesis and turnover have yet to be identified, biochemical studies suggest that mannogen oligosaccharides are initially assembled on a mannose phosphate primer and subsequently extended by guanosine diphosphate (GDP)-Man-dependent mannosyltransferases (Sernee et al., 2006; van der Peet et al., 2006). Conversely, the rapid degradation of mannogen under glucose-limiting conditions (Ralton et al., 2003) implies the presence of putative mannosidases (which remove neutral sugars or disaccharides from the non-reducing termini of glycans) or glycan phosphorylases (which remove sugar-1-phosphates that can feed directly into central carbon metabolism without consumption of ATP) (Figure 1A). Significantly, *Leishmania* mutants lacking key enzymes in GDP-Man synthesis have global defects in mannogen synthesis, as well as protein and lipid mannosylation (Ralton et al., 2003), and are avirulent in animal models (Garami and Iig, 2001), suggesting that one or more of these pathways is important for virulence.

Here, we show that mannogen synthesis and turnover is regulated by a single family of glycosyltransferases that, uniquely, exhibit dual sugar-nucleotide-dependent mannosyltransferase and glycan phosphorylase activities. Analysis of the structure and function of these enzymes suggests that the mannogen pathway evolved through a process of horizontal gene transfer (HGT), gene duplication, and expanded substrate promiscuity. We show that the synthesis and turnover of mannogen regulates *Leishmania* central carbon metabolism and cellular stress responses under both nutrient replete, as well as nutrient-limiting conditions, and is essential for parasite infectivity in the mammalian host. The evolution of this pathway may therefore have played a pivotal role in enabling these parasites to colonize new insect and mammalian host niches.

RESULTS

Detection of Mannogen Synthase and Phosphorylase Activities in *Leishmania* Extracts

In support of earlier studies (Ralton et al., 2003; van der Peet et al., 2006, 2012), mannogen synthesis was shown to be mediated by GDP-Man-dependent mannosyltransferase(s), as incubation of *L. mexicana* cell-free extracts with GDP-³H-Man and either native mannogen or the synthetic substrate, octyl- α -D-mannopyranoside (OM1), resulted in the synthesis of labeled oligosaccharides (Figure 1B) comprising linear chains of β -1,2-linked mannose (Figures S1A and S1B). Competition experiments showed that OM1 was elongated by the same mannosyltransferases that elongate native mannogen (Figure S1C), while short pulse-labeling of live parasites with ³H-mannose, followed by subcellular fractionation, showed that mannogen synthesis occurs primarily in the cytoplasm (Figure S1D).

To assess whether mannogen turnover is regulated by mannosidases (exo or endo) or a glycan phosphorylase, cell lysates were prepared from *L. mexicana* Dpmm promastigotes and

incubated with ³H-mannogen. This parasite line lacks the enzyme phosphomannomutase, preventing the catabolism of any mannose-1-phosphate (Man1P) generated in these assays. A single labeled product was generated that co-eluted with authentic Man1P on high-pH anion exchange chromatography (HPAEC) (Figure 1C). Free [³H]-mannose or manno-oligosaccharides were not generated in these assays, indicating that mannogen turnover is primarily mediated by a glycan phosphorylase.

Identification of a Gene Family in *Leishmania* encoding Putative Mannoside Phosphorylases

Bioinformatic searches of the *Leishmania* genomes revealed a tandem array of seven genes on chromosome 10 (Figure 1D) that shared 45%–68% sequence identity with each other and distant sequence homology (5%–12% identity) to β -1,2-, β -1,3-, and β -1,4-mannosidases or phosphorylases belonging to the carbohydrate active enzyme (CAZy) family 130 glycoside hydrolases (Cuskin et al., 2015; Senoura et al., 2011; Lombard et al., 2014; Ye et al., 2016; Saburi, 2016) (Figures S2A and S3). Related gene arrays, containing between 2 and 12 genes, were also present in closely related trypanosomatids (e.g., *Phytomonas* spp, *C. fasciculata*, *Leptomonas* spp) that synthesize mannogen, but not in more distantly related trypanosomatids (e.g., *T. cruzi*, *T. brucei*) that lack any storage carbohydrates (Figure 1E). Interestingly, the crystal structure of one of the proteins encoded by the *L. major* gene array (hypothetical *L. major* protein encoded by LmjF.10.1260; PDB: 2B4W) has been deposited and shown to have the same five-bladed β -propeller fold as the bacterial GH130 proteins. To assess the function of this gene family and its role in mannogen catabolism, a *L. mexicana* null mutant lacking the entire seven gene array was generated by replacement of the diploid chromosomal loci with bleomycin drug-resistance cassettes. Unexpectedly, the mutant promastigotes, which grew normally in rich medium, were found to be completely deficient in mannogen (Figure 1F), indicating that the encoded proteins may have a role in mannogen biosynthesis in addition to, or instead of, their predicted catabolic role. Based on the genetic and biochemical studies described below, the encoded *Leishmania* proteins have been termed mannosyltransferase/phosphorylases (MTPs) and assigned to a new CAZy family of glycosyltransferases, GT108.

Phylogenetic Analysis of Family GT108

Phylogenetic analysis indicated that the *Leishmania* MTP gene array was acquired after the split that gave rise to the Trypanosomatidae (*Typanosoma* spp) and the subfamily Leishmaniinae (comprising the genera *Leishmania*, *Phytomonas*, *Leptomonas*, *Crithidia*, and *Blechnomonas*) (Figures 1E and S2). The absence of GT108 homologs in free-living *Bodo saltans* and *Paratrypanosoma confusum* indicates that the lack of GT108 homologs in *Typanosoma* spp is not due to recent loss in these lineages. The earliest branches of the Leishmaniinae, *B. ayalai* and

(E) Distribution of MTP homologs and paralogs in different kinetoplastid lineages. Homologs and paralogs are defined based upon their distribution within the phylogenetic tree. *B. ayalai* contains two paralogs of MTP2.

(F) HPAEC analysis of mannogen levels in *L. mexicana* wild-type (WT) and Dmtp1-7 promastigotes. The DP of mannogen oligomers are indicated.

(G) Distribution of GT108 members in different gram-positive bacterial and trypanosomatid taxa (GT108 gene copy number is indicated in the side bar). See also Figures S1–S3 and Tables S4 and S5.

Phytomonas spp, contain two and five GT108 homologs or paralogs, respectively, suggesting that the primordial mannogen pathway may have required just two GT108 genes and that these loci have subsequently undergone several rounds of duplication in other members of this subfamily (Figure 1E).

tBLASTN searches of the non-redundant protein sequence and TriTrypDB databases using the *Leishmania* GT108 sequences as a query led to the identification of >100 sequences of bacterial origin filtered based on e-value ($<1e^{-50}$) (Figures 1G, S2A, and S2B). No eukaryotic sequences were identified outside the trypanosomatids (Figures 1G and S2B). Most of the bacterial GT108 family members were associated with sequenced and annotated genomes of gram-positive bacteria belonging to the phylum Firmicutes and are annotated as hypothetical proteins with unknown function (Figure 1G). Several of these genera are present in the gut microbiota (e.g., *Paenibacillus*) of trypanosomatid insect vectors (Dey et al., 2018), suggesting that the GT108 genes may have been acquired by horizontal gene transfer in this host niche.

Several *Leishmania* MTPs Exhibit β -1,2-Mannoside Phosphorylase Activity

The seven *L. mexicana* proteins were expressed in *E. coli* and recombinant proteins incubated with the β -1,2-mannan acceptor, OM4, together with either 1 mM phosphate (Pi) or Man1P, to assess their phosphorylytic and reverse phosphorylytic (biosynthetic) activities, respectively (Figure 2A) (Kitaoka, 2015). MTP3, 4, 6, and 7 exhibited strong phosphorylytic activity, catalyzing the degradation of OM4 to OM1–3 in the presence of Pi (Figure 2B) with production of Man1P (Figure 2C). No activity was observed in the absence of Pi, indicating that the MTPs lack β -mannosidase activity. Conversely, incubation of MTP3, 4, 6, and 7 with Man1P instead of Pi resulted in elongation of OM4 (Figure 2B) and release of Pi (Figure S1E). The forward and reverse phosphorylytic activity of these enzymes did not require the divalent cations, Mn^{2+} or Mg^{2+} (Figure S1F). MTP3 and MTP6 could generate long mannogen chains (degree of polymerization [DP] >16) by reverse phosphorylysis, while MTP4 and 7 generated shorter mannoooligosaccharides (DP 1–12) at equilibrium (Figure 2B). In contrast, recombinant MTP1, 2, and 5 proteins lacked detectable forward or reverse phosphorylytic activity in these assays (Figure 2B).

The *Leishmania* MTPs Exhibit Dual Phosphorylytic and Mannosyltransferase Activities

To investigate whether the MTPs are also responsible for the GDP-Man-dependent mannosyltransferase activity detected in cell lysates (Figure 1B), recombinant MTP proteins were incubated with OM4 (or OM1) and GDP-Man (Figures 2D, 2E, S1G, and S1H). Strikingly, both MTP1 and MTP2 extended these primers in the presence of GDP-Man, generating either highly extended octyl-mannogen chains (DP 5 to >30) or a family of shorter octyl-mannogen oligomers (DP 2–8) (Figures 2D and S1H). The GDP-Man-dependent mannosyltransferase activities of MTP1 and 2 are reversible, as indicated by the trimming of the OM4 substrate in the presence of GDP-Man or GDP (Figures 2D and 2F) and detection of GDP-Man synthesis in the presence of GDP (Figure S1I).

MTP3, 4, 6, and 7 could also elongate the synthetic acceptor, OM1, with a single β -1,2-mannose residue in the presence of

GDP-Man (Figure S4A). However, they could not elongate OM4 but instead generated new products that co-migrated with mannose (Man_1), $Man\beta$ 1-2Man (Man_2), and $Man\beta$ 1-2Man β 1-2Man (Man_3) (Figure 2D). These observations suggest that GDP-Man is cleaved by these MTPs in the absence of suitable acceptor to generate free mannose that acts as an alternative primer of β -1,2-mannose oligosaccharide synthesis. Interestingly, MTP4 and 6 (but not MTP3, 5, and 7) generated a mixed series of octyl-mannogen oligomers when incubated with OM4 and GDP (Figure 2D), consistent with the finding that they can catalyze synthesis of GDP-Man (Figure S1I). The fact that GDP was more effective than GDP-Man in promoting MTP4- and MTP6-dependent elongation of OM4 indicates that the newly synthesized GDP-Man remains bound to the catalytic –1 subsite, while an acceptor glycan enters the +1 subsite, allowing net elongation of the oligomannoside pool.

Co-incubation of MTP3, 4, and 6 with both GDP-Man and Man1P enhanced OM1 elongation, indicating that both donors are accommodated within the same active site. Consistent with this conclusion, incubation of MTP4 with OM1 (1 mM), GDP- 3H -Man (50 mM) and increasing concentrations of α -Man1P (50 mM to 5 mM) competitively inhibited GDP- 3H -Man incorporation (Figure 2G). In contrast, β -Man1P only competed with GDP- 3H -Man when added at 100-fold excess (reflecting the fact that this sugar can also act as an acceptor). α -Man1P and GDP-Man may therefore compete for the same mannose donor (–1) subsite in these enzymes.

These data show that MTP1 and 2 function primarily as GDP-Man-dependent β -1,2-mannosyltransferases, while MTP3, 4, 6, and 7 act as β -1,2-mannoooligosaccharide phosphorylases with dual transferase activity. MTP5 lacks detectable activity in all of these assays, and the function of this protein remains unknown (Figures 2B and 2D).

Identification of Catalytic Site Residues Responsible for Phosphorylytic and Mannosyltransferase Activities

To better understand the dual transferase and phosphorylase activities of this enzyme family, the 3D structures of MTP1, MTP2, and MTP4 containing different substrates were determined (Figures 3A–3C, S5, and S6; Tables S1 and S2). Like the deposited structure of the *L. major* MTP4 protein (PDB: 2B4W), all three proteins adopt a five-bladed β -propeller fold, with each blade composed of a β sheet with three or four antiparallel strands surrounding an extended central pocket comprising the active site (Figures 3A–3C). The catalytic pockets of these proteins have similar architecture to those of the sequence-divergent bacterial CAZy family GH130 proteins, such as the *Listeria innocua* β -1,2-mannobiose phosphorylase (PDB: 5B0R) (Tsuda et al., 2015) (Figure 3E). In particular, overlay of the catalytic pocket of MTP4 and the *L. innocua* protein revealed equivalent Pi coordinating residues (Arg150, His208, Asp134, and Tyr224; numbering for *L. mexicana* MTP4) arranged identically in space (Figure 3E; Table S2). Similarly, the identity, spatial arrangement, and interactions of residues in the –1 mannosyl transfer site are identical. These include the Phe265 aromatic platform and residues that interact with different hydroxyls of the –1 mannose, such as Lys133 with mannose C(O)2, Asp83 with mannose C(O)3 (proposed catalytic acid/base in the reaction mechanism; see below), Asn32 with mannose C(O)4, and Asp285 with mannose C(O)6

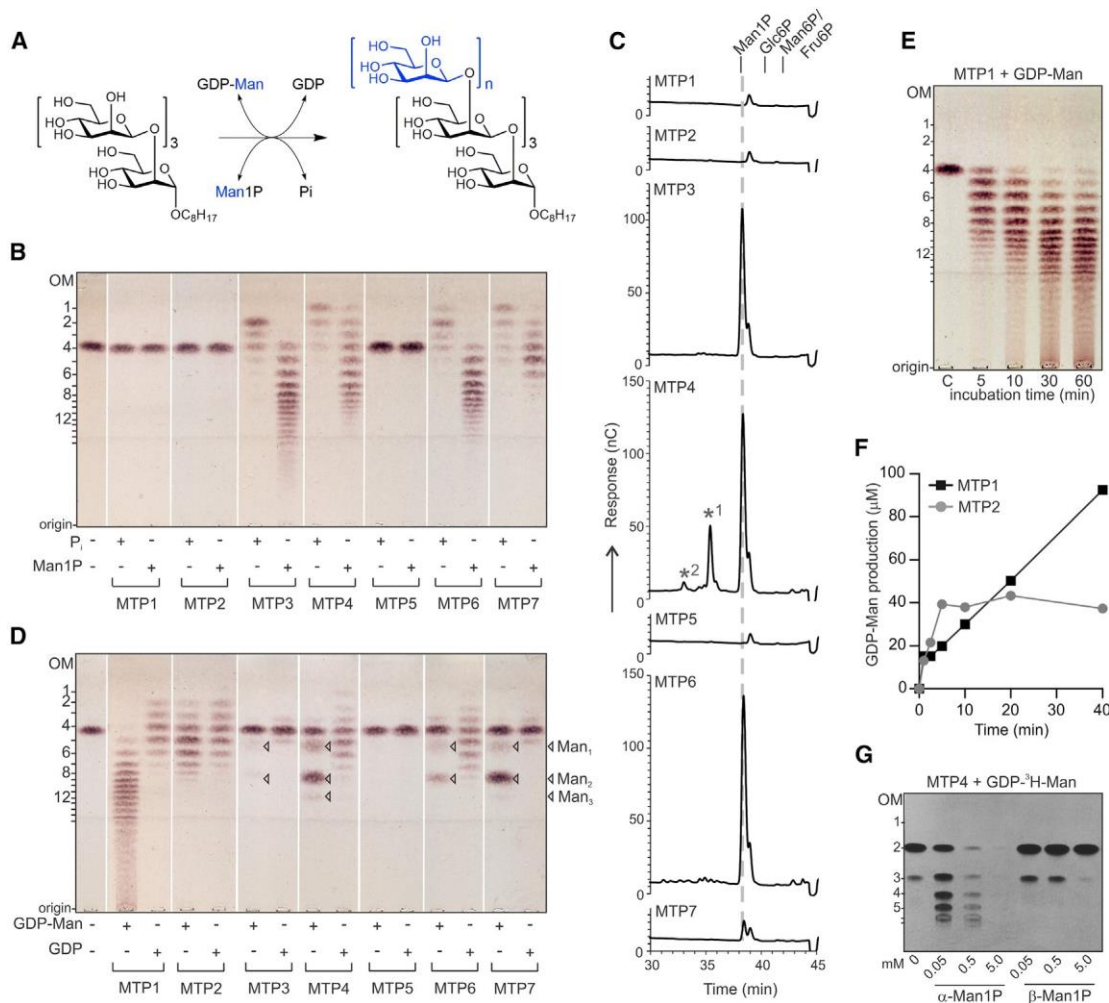


Figure 2. The *Leishmania* MTPs Exhibit Dual Mannosyltransferase and Phosphorylase Activities

(A) Enzyme assays used to measure MTP mannosyltransferase and reverse phosphorylase activity.
 (B) HPTLC separation of octyl-mannogen oligosaccharides (DP 1–16) generated by incubation of MTP proteins with OM4 and either Pi or Man1P.
 (C) HPAEC separation of Man1P and phosphooligosaccharides (*) synthesized by MTP proteins incubated with mannogen and Pi. MTP4 synthesized Man β -1,2-Man α -P (*) and [Man β -1,2]₂-Man α -P (*2), in addition to Man1P.
 (D) High-performance thin-layer chromatography (HPTLC) separation of octyl-mannogen oligosaccharides generated by incubation of MTP proteins with OM4 and either GDP-Man or GDP. Products indicated with open triangle correspond to mannose, Man β -1,2-Man α -P (Man₂), and [Man β -1,2]₂-Man (Man₃).
 (E) Time-dependent synthesis of high-DP octyl-mannogen following incubation of MTP1 with OM4 and GDP-Man.
 (F) Liquid chromatography-mass spectrometry (LC-MS) detection of GDP-Man synthesized by MTP1 (black squares) and MTP2 (gray dots) in the presence of octyl-mannogen (DP 1–16) and GDP.
 (G) HPTLC separation of octyl-³H-mannogen species synthesized by MTP4 in the presence of OM1 (1 mM) and GDP-³H-Man (50 mM) and increasing concentrations of α -Man1P and β -Man1P.
 See also Figures S1 and S4.

(Figures 3E and S5E). The catalytic pockets of MTP4 and the bacterial GH130 proteins also have a very similar mannose acceptor +1 subsite, where the mannose C(O)3 interacts with Arg48, C(O)4/(O)6 interacts with Glu82, and C(O)5/(O)6 interacts with Lys133 (Figures 3E, 3H, and 4A; Table S2).

The MTP1 and 2 proteins have 3D structures that are similar to MTP4 but form dimers rather than monomers in the crystal lattice and in solution (data and structure statistics in Table S1; Figures S4D, S5A, and S5B). The structure of the catalytically inactive MTP2 mutant lacking the proposed aspartate acid/base (D94N), containing β -1,2-mannobiose in the -1/+1 subsites

(Figure 3D), showed that amino acid side chains in the MTPs are spatially conserved and partake in identical interactions as the equivalent residues in the bacterial GH130 enzymes (Table S2). Neither MTP1 nor MTP2 contained canonical nucleotide binding domains despite utilizing GDP-Man as primary donor, confirming that the MTP enzymes represent a new paradigm in nucleotidyl sugar glycosyltransferases.

We were unable to obtain complexes of MTP1 and 2 with GDP or GDP-Man through co-crystallization and crystal-soaking experiments, suggesting that GDP-Man binding only occurs after binding of the mannan acceptor and/or that structural changes

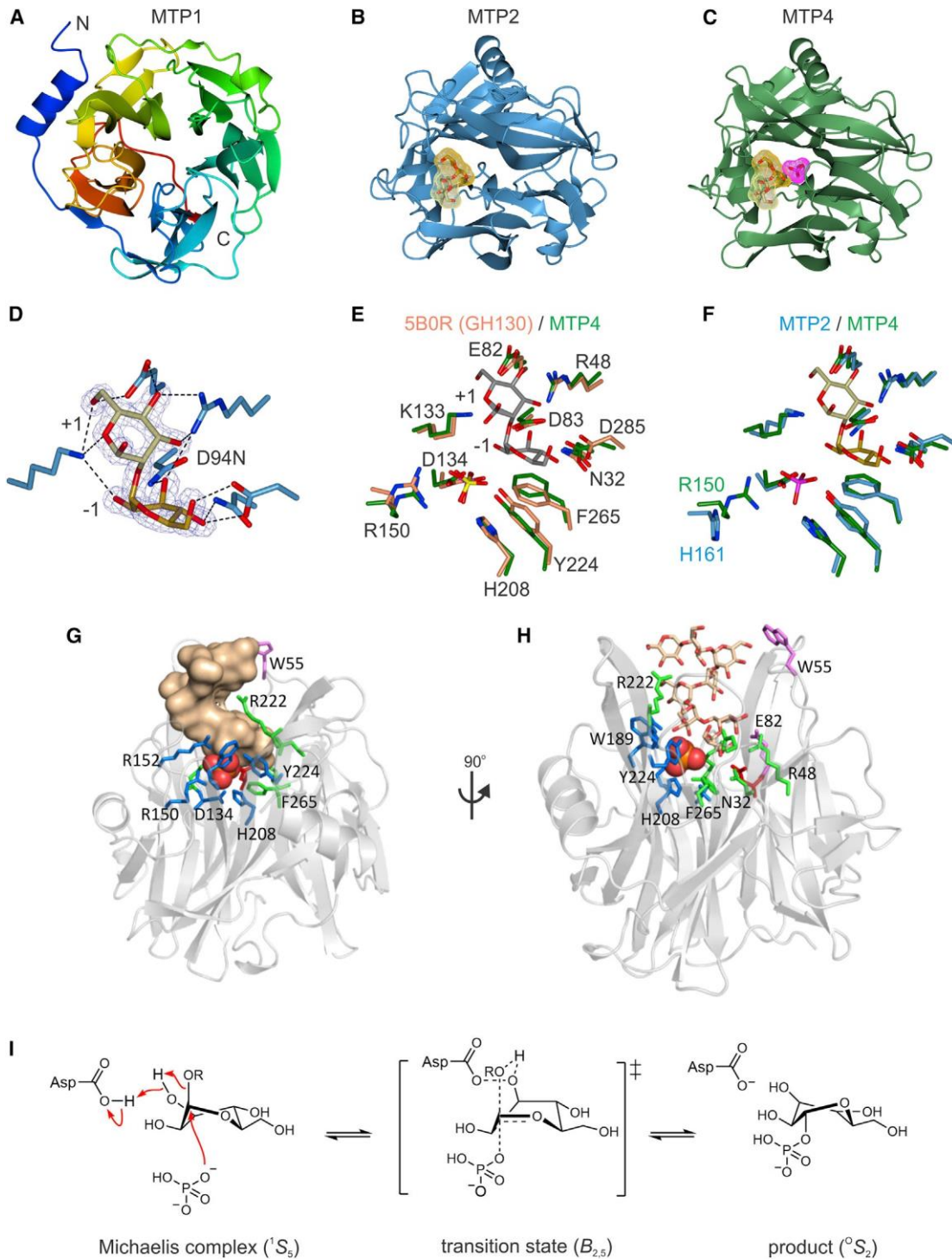


Figure 3. MTP Crystal Structures and Modeled Interactions

- (A) The five-bladed β -propeller fold of MTP1 (top view). The β sheets containing three to four antiparallel strands are color coded.
- (B) MTP2 (side view) with β -1,2-mannose (yellow stick and molecular surface representation) in the +1/-1 catalytic position. The complex was captured using the MTP2 D94N mutant and the crystal structure overlaid on native MTP2.
- (C) MTP4 (side view) with the observed Pi shown in magenta. The β -1,2-mannose of the overlaid D94N MTP2 variant crystal complex is included for reference (D94N MTP2 protein not shown).
- (D) The β -1,2-mannose ligand in complex with the MTP2 D94N protein.

(legend continued on next page)

may be required for nucleotide binding. Nonetheless, we could identify structural differences in the catalytic pocket of MTP1 and 2 and MTP4 that likely contribute to the different donor and acceptor specificities of these enzymes (Figures 3F, 4A, and 4C). In particular, MTP4 (and other bacterial GH130 phosphorylases) contain an Arg residue (Arg150 in MTP4) in the Pi-binding pocket, while MTP1 and 2 contain a His residue (His168 and His161, respectively) in the same position (Figures 3F, 4A, 4C, S2C, S3, and S5C–S5F; Table S2). Sequence analysis confirmed that the Arg/His switch is diagnostic for enzymes that primarily exhibit phosphorylytic versus mannosyltransferase activities, supporting the conclusion that these residues mediate critical interactions with the phosphate or GDP leaving groups, respectively (Figures S2C and S3).

Our data suggest that the MTPs have the same reaction mechanism as the bacterial GH130 phosphorylases (Nakae et al., 2013). During phosphorylysis or the reverse transferase reactions, the conserved Asp (Asp83 in MTP4, Asp94 in MTP2, Asp101 in MTP1) acts as a general acid, protonating the glycosidic leaving group via a proton relay through C(O)3, allowing phosphate to displace the anomeric glycoside leaving group (Figure 3I). Conversely, during reverse-phosphorylysis or forward-mannosyltransferase reactions, Asp83 acts as general base, deprotonating C(O)2 of a sugar nucleophile and promoting glycosidic bond formation with displacement of phosphate (or GDP) leaving group. Consistent with this mechanism, the –1 subsite mannoside is distorted into an unusual 1S_5 skew conformation (Figure 3I), facilitating unhindered phosphate attack and departure without steric clashes with C(O)2 (Cuskin et al., 2015). Mutagenesis studies confirmed the importance of the active center residues in the dual phosphorylytic and mannosyltransferase activities of MTP4 (Figures 4A and 4B). Mutation of the catalytic Asp83 acid/base (D83N) resulted in complete loss of both phosphorylase and transferase activities (Figures 4A and 4B). Mutation of Lys133 (K133A, K133R) predicted to bind the OM1 acceptor in the +1 subsite of the pocket also resulted in loss of both activities. In contrast, mutations of Asp285 (D285N), which forms hydrogen bonds with C(O)4/(O)6 of the –1 subsite mannose; Asp134 (D134N), which binds the Man1P phosphate group; and Glu82 (E82A), which hydrogen bonds with C(O)4/(O)6 of the +1 sugar, led to loss of phosphorylase activity but only had minimal effect on the transferase activity. Mutation of MTP4 His208 (H208A) also had a modest effect on phosphorylysis but no effect on transferase activity. Selective retention of transferase activity indicates that the amino acid residues that interact with the GMP moiety of this donor likely assist leaving group departure.

Mutation of the conserved active center aspartate at position 94 in MTP2 (D94N) resulted in complete loss of transferase activ-

ity (Figures 4C and 4D). The active site pocket of the MTP1 and MTP2 enzymes is larger than in MTP4 and contained a His residue (His168 in MTP1 and His161 in MTP2) in place of the Arg residue in MTP4 (Arg150), which establishes hydrogen bonds with the guanine moiety of GDP-Man (Figures 4C, 4D, and S5F). The importance of the Arg/His switch was confirmed by mutating MTP2 His161 to Arg, which resulted in reduced transferase activity (Figure 4D). In contrast, mutation of MTP2 Asp145 (D145N) had little effect on transferase activity, consistent with this residue having a primary role in coordination to Man1P phosphate during phosphorylysis (Figures 4B and 4D). These data suggest that all of the MTP proteins contain a single catalytic site that preferentially accommodates either Man1P or GDP-Man as donors, together with a β -1,2-mannooligosaccharide acceptor.

Modeling of the bacterial GT108 proteins provided further support for the hypothesis that the *Leishmania* MTPs were acquired by HGT. Structural predictions of eleven bacterial GT108 proteins (Table S3) indicated that they all share the same five-bladed β -propeller fold (Figures S6A and S6B) and retain key residues in the catalytic pocket. Notably, all of the modeled bacterial proteins retained an Arg residue at the position of the Arg/His switch, indicating that they are primarily glycan phosphorylases/hydrolases (Figures S2C and S6D). *B. ayalai*, the most divergent member of the Leishmaniinae, contains three MTP genes: a homolog of MTP1 with a His residue in the Arg/His switch position and two paralogs of MTP2, which both contain Arg in the same position (Figures 1E and S2C). A functional mannogen cycle may therefore have arisen in ancestral members of the Leishmaniinae following HGT of a bacterial phosphorylytic GT108 gene and subsequent gene duplication and mutation around the Arg/His switch position to allow use of GDP-Man as donor.

Leishmania MTPs Have Non-redundant Functions *In Vivo*

To investigate the function of the MTPs *in vivo*, the *L. mexicana* $\Delta mtp1-7$ mutant was complemented with individual MTP genes using episome or integration vectors. Expression of MTP1 alone resulted in production of hyper-elongated mannogen oligomers (DP >60), while expression of MTP2 resulted in synthesis of a restricted pool of short mannogen oligomers (DP 2–10), consistent with the *in vitro* activities and products of each of these enzymes (Figure 5A). Co-expression of both MTP1 and 2 in the complemented *L. mexicana* $\Delta mtp1-7$ mutant, or targeted deletion of MTP3–7, resulted in an intermediate mannogen profile containing both short and long mannogen oligomers (DP 4 to >60) (Figure 5A). MTP1 and MTP2 thus appear to act cooperatively *in vivo* to regulate the synthesis of high- and low-DP pools of mannogen, respectively.

(E) Overlay of the active sites of the Pi complex of MTP4 (green) and the β -1,2-mannobiose-sulfate complex of *L. innocua* GH130 β -1,2-mannobiose phosphorylase (pink) (PDB: 5B0R) (Tsuda et al., 2015). Sulfate (yellow and red sticks) and Pi (magenta and red sticks) are co-localized; β -1,2-mannobiose in gray and red sticks.

(F) Overlay of D94N MTP2 (blue) and MTP4 (green) reveals minor changes (His or Arg residues) in the Pi site of these MTPs (Pi in magenta and red sticks and β -1,2-mannobiose in yellow and red sticks).

(G) Man₈ (beige molecular surface) docked into the catalytic site of MTP4 (semi-transparent cartoon). Residues in the Pi site and +1/–1 acceptor and donor subsites shown as sticks (colored according to their location as defined in Figure S3 and Table S2); Pi depicted as spheres.

(H) View is a 90° rotation about the y axis to that shown in (G). Man₈ depicted as thin sticks (beige and red).

(I) Canonical catalytic mechanism for GH130 and MTP enzymes.

See also Figures S2–S6 and Tables S1–S3.

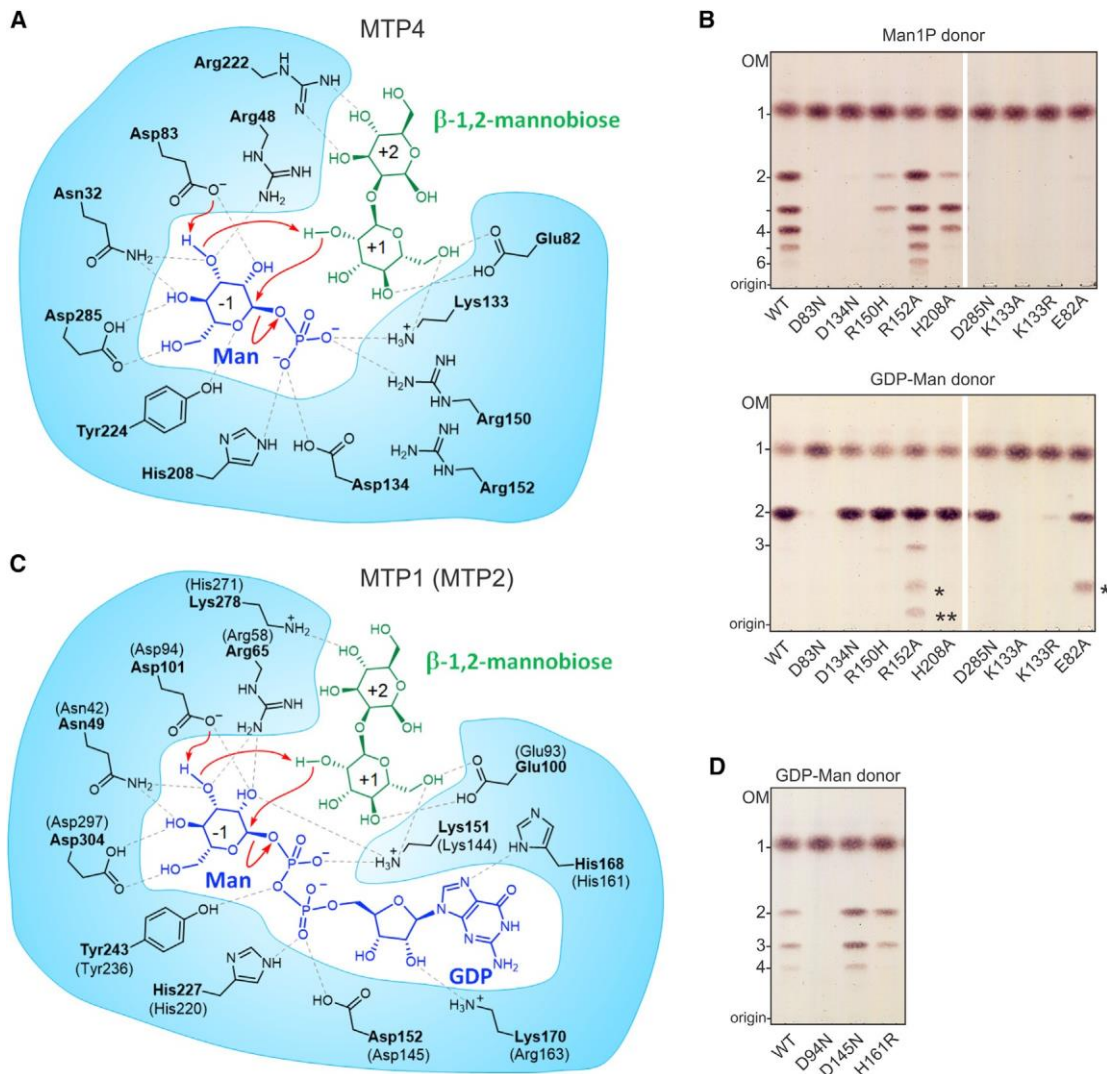


Figure 4. Identification of Key Residues in the *L. mexicana* MTP Catalytic Pocket Required for Phosphorolytic or Mannosyltransferase Activity

- (A) 2D interaction map showing hydrogen bonds and electrostatic interactions (black dashed lines) between MTP4 amino acid residues and the Man1P donor and mannogen acceptor combination.
- (B) Mutant MTP4 proteins were incubated with OM1 and either Man1P (upper panel) or GDP-Man (lower panel) and the products analyzed by HPTLC.
- (C) 2D interaction map of MTP1 active site showing side-chain interactions with GDP-Man and mannogen acceptors. Equivalent MTP2 residues are shown in brackets.
- (D) MTP2 mutant proteins were incubated with OM1 and GDP-Man and products analyzed by HPTLC.
- See also Figures S3, S5, and S6.

Mannogen levels were elevated in parasite lines lacking the phosphorolytic MTPs, indicating that they are primarily involved in mannogen turnover (Figure 5B). In support of this notion, expression of MTP4 in the $\Delta mtp3-7$ mutant resulted in a modest decrease in mannogen levels (Figure 5B). Complementation of the *L. mexicana* $\Delta mtp1-7$ null mutant with MTP3, 4, 6, or 7, individually or in combination, did not restore mannogen synthesis (Figure S4B), even though the expressed proteins were active in cell lysates (Figure S4C). These data suggest that the phosphorolytic MTPs cannot prime mannogen synthesis *in vivo* either because of limited availability of suitable mannose primers and/or because of

high intracellular Pi levels that drive the reaction in the phosphorolytic direction.

Regulated Mannogen Synthesis Is Required for Infection in the Mammalian Host

Despite growing normally in rich medium at 27°C, $\Delta mtp1-7$ parasites did not induce lesions or proliferate in highly susceptible BALB/c mice (Figures 6A and 6C). Strikingly, expression of MTP1 alone or MTP1/MTP2 together in the absence of phosphorolytic MTPs ($\Delta mtp1-7::pX\text{-MTP1}$ and $\Delta mtp3-7$ lines, respectively) exacerbated the loss of virulence phenotype, with no recovery of parasites from tissue biopsies (Figures 6A and 6C).

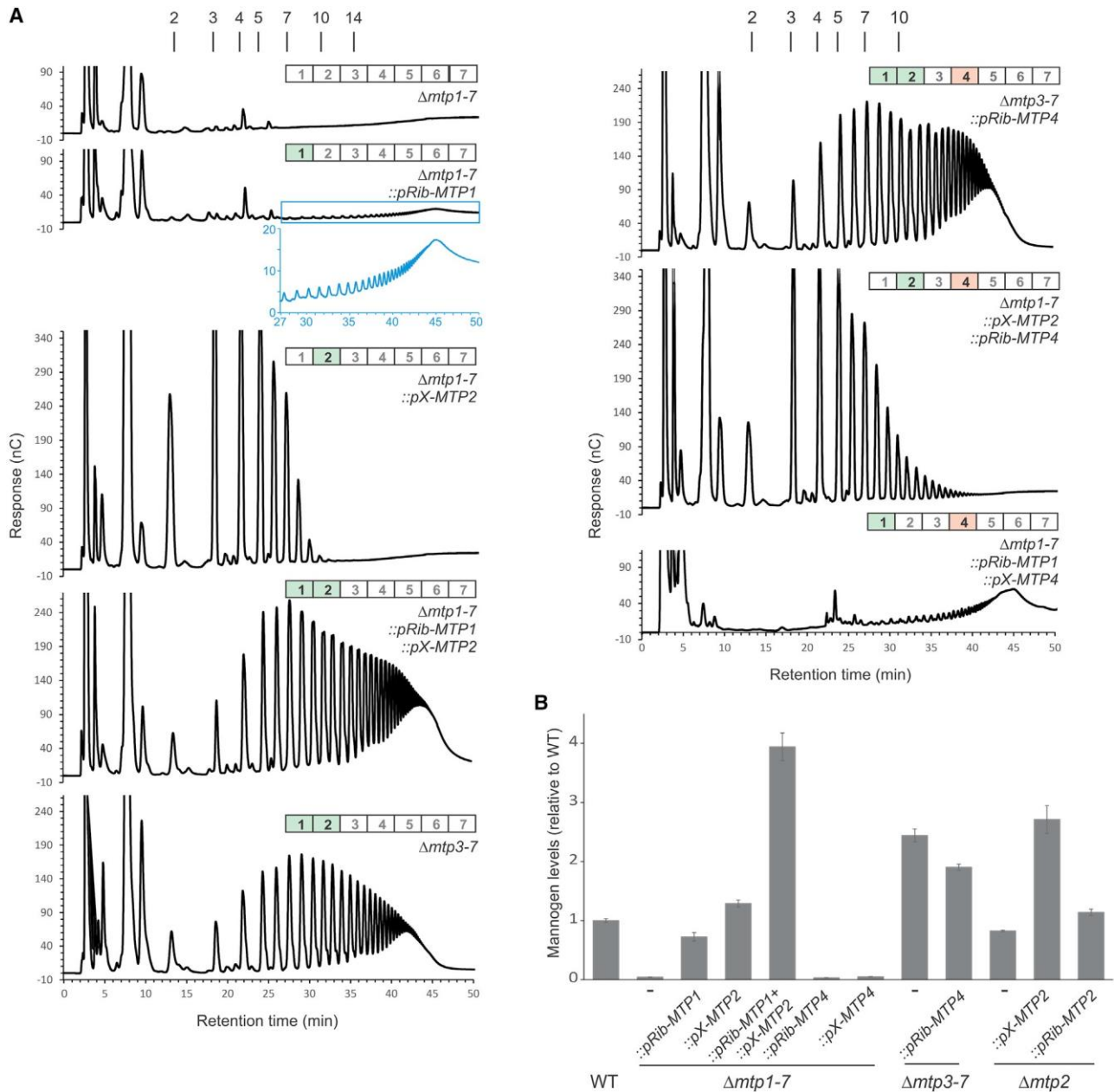


Figure 5. *Leishmania* MTPs Have Non-redundant Functions In Vivo

(A) HPAEC-PAD analysis of the mannogen profiles of *L. mexicana* $\Delta mtp1-7$ and $\Delta mtp3-7$ and complemented lines expressing *MTP* genes on an episomal plasmid (*pX*) or inserted into the ribosomal loci (*pRib*). Expression (green and orange) or absence (no color) of individual *MTP* genes in each line is indicated. The insert shows the hyperextended mannogen chains in $\Delta mtp1-7::pRib-MTP1$ parasites.

(B) Mannogen levels in the different transgenic parasite lines ($n = 3$ biological replicates). Data presented are mean \pm SEM.

See also Figures S4 and S7.

Complementation of $\Delta mtp1-7$ with *MTP2* and *MTP4* also failed to restore virulence (Figure 6A), suggesting that *Leishmania* virulence in animals is highly sensitive to loss of one or more MTPs. As *MTP2* plays a key role in regulating mannogen DP *in vivo*, we generated a knockout line lacking only *MTP2*. The $\Delta mtp2$ mutant did not induce lesions in BALB/c mice even when a high inoculum (10^7 parasites) was used (Figures 6B and S7B). Genetic complementation of this mutant ($\Delta mtp2::pX-MTP2$ or

$\Delta mtp2::pRib-MTP2$), restored synthesis of low DP mannogen (Figure S7A) as well as virulence in mice (Figures 6B and S7B), suggesting that mannogen synthesis and/or cycling is critical for survival of mammalian infective stages.

Defects in mannogen synthesis or cycling were also associated with attenuated growth in macrophages (Figure S7C) which, in the case of $\Delta mtp2$, could be restored by complementation with *MTP2* (Figure S7C). Mannogen synthesis and

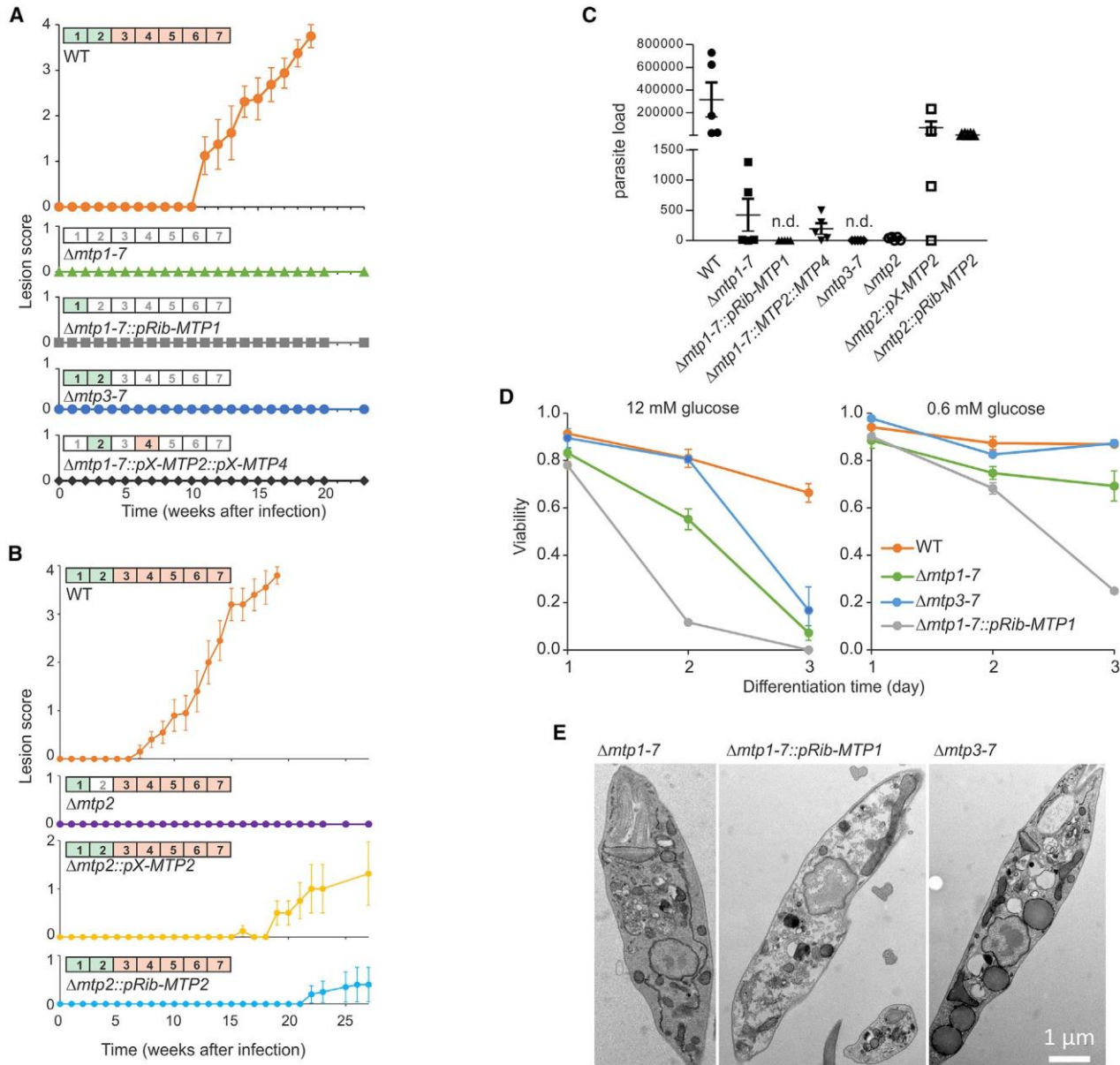


Figure 6. Loss of Single or Multiple *Leishmania* MTPs Leads to Loss of Virulence

(A) Induction of lesions by *L. mexicana* WT, $\Delta mtp1-7$, $\Delta mtp3-7$ and selected complemented parasite lines in BALB/c mice (n = 5 animals per treatment).

(B) Induction of lesions by $\Delta mtp2$ and corresponding pX- or pRib-MTP2 complemented lines (n = 5 animals per treatment).

(C) Parasite burden in proximal lymph nodes 23 weeks post infection (n = 5 tissue samples). n.d., no parasites were detected.

(D) WT and Δmtp mutant promastigotes were cultivated in medium, pH 5.5, containing either 12 mM or 0.6 mM glucose (33°C for 3 days) and parasite viability determined by propidium iodide staining (n = 3 biological replicates).

(E) TEM ultrastructure of $\Delta mtp1-7$, $\Delta mtp1-7::pRib-MTP1$ and $\Delta mtp3-7$ promastigotes.

Data presented for (A)–(D) are mean \pm SEM.

See also Figure S7.

turnover also appeared to be required for amastigote differentiation as $\Delta mtp1-7$ and $\Delta mtp3-7$ promastigotes lost viability when incubated at elevated temperature (33°C) and low pH (5.5), conditions that normally induce amastigote differentiation (Figures 6D and S7D). Viability was partly rescued if parasites were induced to differentiate in low-glucose medium (Figure 6D), suggesting that mannogen cycling may protect parasites from

glucose toxicity. $\Delta mtp1-7::pRibMTP1$ promastigotes were also acutely sensitive to glucose toxicity (Figure 6D), which likely reflects the hyper-accumulation of high-DP mannogen in these parasites and disruption of the cytoplasm (Figure 6E).

To further assess the role of mannogen synthesis and cycling in regulating glucose homeostasis, *L. mexicana* promastigotes and amastigotes were labeled with $^{13}C_6$ -glucose and mannogen

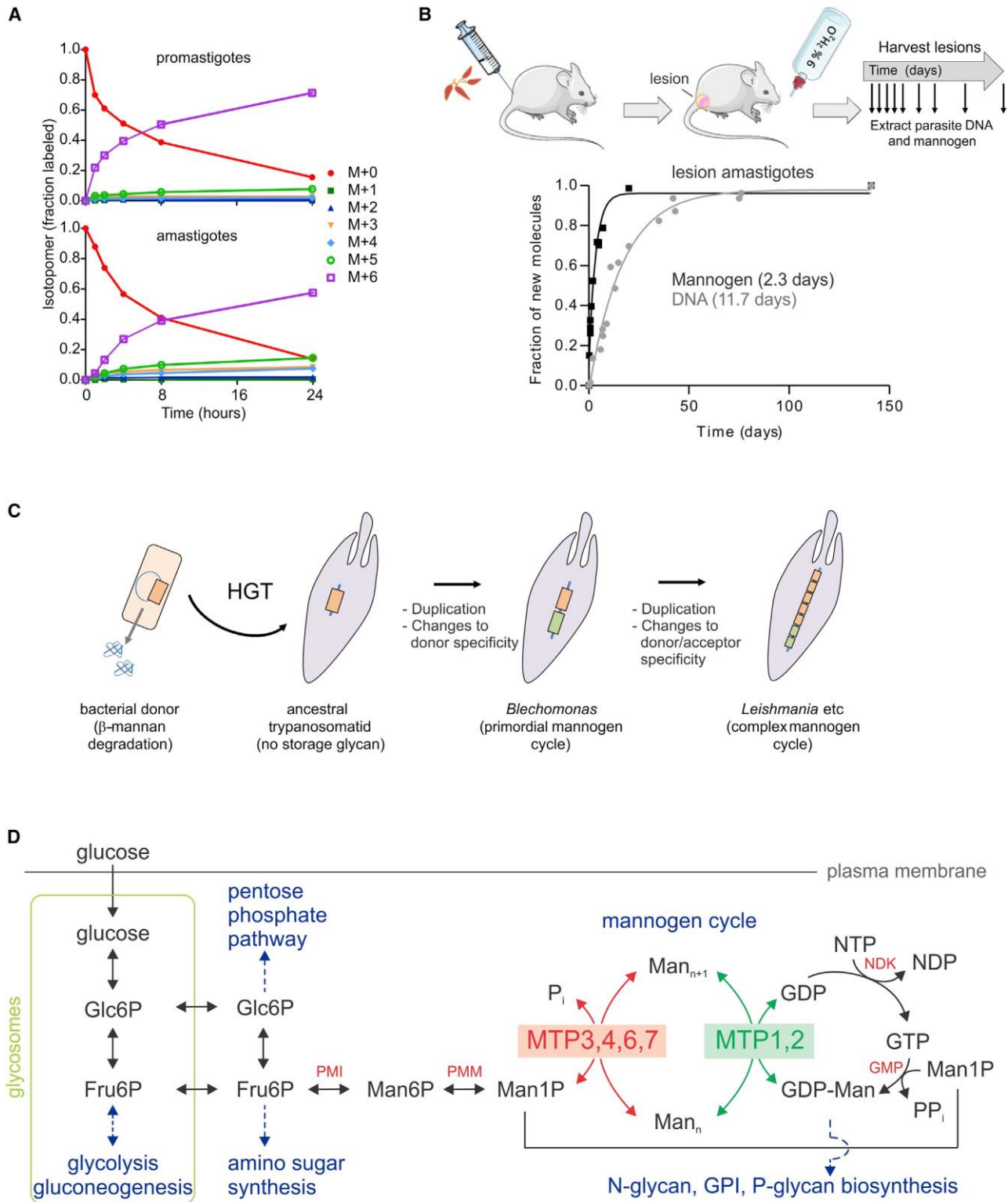


Figure 7. Mannogen is Constitutively Cycled under Nutrient-Replete Growth Conditions *In Vitro* and in Murine Lesions

(A) Incorporation of $^{13}\text{C}_6$ -glucose into mannogen (M+0 to M+6) in *L. mexicana* promastigotes and axenic amastigotes.

(B) *L. mexicana* infected BALB/c mice were labeled with 9% $^2\text{H}_2\text{O}$ and the level of deuterium incorporation into parasite DNA (deoxyribose moiety) and mannogen determined by gas chromatography-mass spectrometry (GC-MS). The $t_{1/2}$ for DNA and mannogen turnover was 11.7 and 2.3 days, respectively.

(C) Proposed evolution of the mannogen cycle in Trypanosomatids by HGT, gene duplication, and changes in MTP donor and/or acceptor specificity.

(legend continued on next page)

turnover under glucose-replete conditions measured. Constitutive mannogen turnover was observed in both promastigote ($t_{1/2}$ 4 h) and amastigote stages ($t_{1/2}$ 5 h) (Figure 7A). The high turnover rate of mannogen in amastigotes is notable, as these stages have a 10-fold lower rate of glucose uptake compared to promastigotes (Saunders et al., 2014). Interestingly, most of the mannose in the mannogen pool was fully labeled (M+6), indicating that a significant proportion of exogenous glucose is channeled directly into mannogen synthesis. The accumulation of additional mannose isotopomers (M+3, M+4, and M+5) reflects the recycling of hexose-phosphates (and derived triose-, tetraose-, and pentose-phosphates) through the oxidative and non-oxidative arms of the pentose phosphate pathway (Figure 7A). The complexity of these labeling patterns suggests that mannogen cycling regulates glucose homeostasis and carbon fluxes under glucose-replete conditions.

To investigate whether constitutive mannogen cycling also occurs in metabolically quiescent amastigote stages in tissue granulomas, infected BALB/c mice were labeled with 9% deuterium oxide ($^2\text{H}_2\text{O}$) and turnover of parasite DNA and mannogen determined from ^2H -incorporation into the deoxyribose and mannose constituents of these molecules, respectively (Figure 7B). Mannogen turnover was approximately five times faster than DNA ($t_{1/2}$ of 2.3 days versus 11.7 days, respectively) (Kloehn et al., 2015). As mannogen comprises a significant fraction of amastigote biomass (Ralton et al., 2003), mannogen cycling represents a major metabolic flux in these stages.

DISCUSSION

The absence of canonical carbohydrate reserves in all of the Trypanosomatidae suggest that these pathways were lost early in the evolution of the parasitic lifestyles of these protists (Oppenoes et al., 2016). Here, we show that an important subgroup of the Trypanosomatidae, the Leishmaniinae, have subsequently evolved a non-canonical pathway of carbohydrate reserve biosynthesis through the acquisition and functional repurposing of bacterial genes involved in carbohydrate degradation. As a result of gene duplication and donor and/or acceptor diversification, the MTP enzymes catalyze a cycle of mannogen synthesis and turnover that contributes to the regulation of key pathways in central carbon metabolism and is essential for virulence. We propose that the evolution of this pathway has facilitated the colonization of both extracellular and intracellular niches in the insect vectors and mammalian hosts of these parasites.

In most organisms, the biosynthesis and turnover of intracellular carbohydrate reserves is regulated by distinct families of sugar-nucleotide glycosyltransferases and glycan phosphorylases or glycosidases. In contrast, the biosynthesis and cycling of *Leishmania* mannogen is regulated by a single family of glycosyltransferases that contain both GDP-Man-dependent mannosyltransferase and β -1,2-mannogen phosphorylase activities

(Figure 7D). The *Leishmania* MTPs are founding members of the CAZy GT108 family, which includes a large number of bacterial proteins of unknown function. GT108 family members share limited sequence identity but a high degree of structural similarity to the CAZy GH130 family of bacterial β -mannan phosphorylase or mannosidases. The six enzymatically active *Leishmania* MTP proteins share overlapping but non-redundant roles in mannogen biosynthesis and cycling. In particular, MTP1 and MTP2 are responsible for priming mannogen synthesis *in vitro* and *in vivo* and for generating distinct pools of mannogen with high DP (>12) and low DP (2–10), respectively. In contrast, MTP3, 4, 6, and 7 primarily act as phosphorylases *in vitro* and *in vivo* and only exhibit transferase activity on small acceptors (Figure 7D). While these enzymes catalyze the reverse phosphorolytic elongation of mannogen *in vitro*, they cannot prime mannogen synthesis *in vivo* under normal growth conditions. Finally, MTP5 contains key catalytic residues representative of both the transferase and phosphorylase MTPs but lacks detectable enzymatic activity. The function of this protein is unknown.

Phylogenetic analysis showed that the *Leishmania* MTPs are sequence related to a large family of proteins from gram-positive bacteria. Intriguingly, genes encoding proteins in the GT108 family are found in bacteria that are commonly present in the midgut of the modern insect vectors colonized by *Leishmania* and other members of the Leishmaniinae subfamily (Dey et al., 2018). The trypanosomatid MTP genes may thus have been acquired by HGT within a similar insect niche, most likely after the split of this group from the trypanosomes (*T. brucei*, *T. cruzi*) (Figure 7C). While the bacterial GT108 genes have yet to be characterized, they contain signature residues of glycan phosphorylases, suggesting a role in the degradation of complex plant glycans in the insect gut or other niches (Nakae et al., 2013). Phylogenetic analysis indicates that acquisition of a GT108 gene in the common Leishmaniinae ancestor was followed by gene duplication and changes in the donor specificity of the primordial MTP proteins. In particular, *B. ayalai*, the most divergent member of the Leishmaniinae, contains three MTP homologs or paralogs, one of which is predicted to be a transferase. The acquisition of MTPs with efficient mannosyltransferase activity is likely to have been the key step in the establishment of a functional mannogen cycle, while the further amplification of this gene family in other members of the Leishmaniinae may have been critical in regulating the overall capacity and flux through this cycle and its responsiveness to different nutrient or growth conditions.

The *Leishmania* MTPs have the same five-bladed β -propeller fold and similar reaction mechanism as bacterial GH130 mannan phosphorylases or hydrolases, although they differ at the sequence level. The MTPs have acquired their dual mannosyltransferase and phosphorylase activities through the acquisition of amino acid changes that expanded the Man1P binding pocket to enhance binding of sugar nucleotides. In particular, the introduction of specific amino acid side chains into the

(D) Schematic of the *Leishmania* mannogen cycle. The synthesis of low- and high-DP mannogen is mediated by MTP2 and MTP1, respectively, while turnover is regulated by MTP3, 4, 6, and 7. Constitutive mannogen cycling may regulate intracellular levels of sugar phosphates ATP and Pi/PPi. PMI, phosphomannose isomerase; PMM, phosphomannomutase; GMP, GDP-Man pyrophosphorylase; NDK, nucleotide diphosphate kinase. See also Figure S7.

phosphate-binding pocket (such as His168/161 in MTP1/MTP2) to bind the guanine moiety demonstrates that GDP is not simply a surrogate for Pi. Interestingly, a number of fungal pathogens, including *Candida albicans*, also synthesize linear β -1,2-mannans, which are incorporated into cell wall and surface glycoconjugates. These glycans are synthesized by family GT91 GDP-Man-dependent β -1,2-mannosyltransferases expressed in the secretory pathway (Sfihi-Loualia et al., 2016). The *C. albicans* mannosyltransferases also share low sequence identity to bacterial GH130 glycan hydrolases and phosphorylases and appear to have independently evolved a similar active site geometry to the *Leishmania* MTPs (Figures S6C and S6E; Table S3). Whether other glycan phosphorylases have also evolved sugar-nucleotide-dependent transferase activities is worthy of further investigation.

Leishmania promastigotes characteristically accumulate high-DP mannogen as they enter stationary phase growth and can rapidly utilize these oligomers under glucose-limiting conditions, suggesting that they primarily function as a canonical energy and carbon reserve in these stages (Ralton et al., 2003). In contrast, the obligate intracellular amastigotes constitutively accumulate high levels (~10 mM) of low-DP mannogen, suggesting a distinct role in mammalian infective stages (Ralton et al., 2003). We have shown that amastigotes enter a slow growing, metabolically quiescent state associated with reduced rates of utilization of glucose and other carbon sources, as well as repression of energy-intensive processes such as transcription and protein translation (Saunders et al., 2014; Kloehn et al., 2015; McConville et al., 2015). In contrast, we show here that low-DP mannogen is constitutively cycled in both cultured and lesion amastigotes. Mannogen cycling may therefore function as a metabolic rheostat, buffering changes in sugar uptake and down-stream fluxes into glycolysis, the pentose phosphate pathway and nucleotide/sugar nucleotide biosynthesis (Figure 7). Mannogen cycling may also have a direct role in modulating intracellular levels of ATP and Pi, which are utilized in the cycle (Figure 7). Importantly, we show that mannogen cycling confers resistance to glucose toxicity and elevated temperatures and is essential for virulence in the mammalian host. *Leishmania* are also exposed to transient and potentially toxic pulses of sucrose (up to 30%) when the sandfly periodically feeds on plant saps and honeydews (Louradour et al., 2017; Dey et al., 2018). Under these conditions, mannogen cycling may prevent catastrophic imbalance in fluxes in upper and lower glycolysis and cell death, as has been proposed for the Pi and nucleotide-triphosphate-consuming trehalose cycle in yeast (van Heerden et al., 2014). The evolution of the mannogen cycle in monogenic trypanosomatid ancestors may thus have allowed these parasites to colonize new insect hosts, as well as to survive and proliferate within the phagolysosomes of macrophages. The central role of mannogen cycling in *Leishmania* central carbon metabolism in the mammalian host suggests that the *Leishmania* MTPs, and in particular MTP2, are potential targets for new therapeutics.

STAR+METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Parasites
 - Bone Marrow Derived Macrophages
 - Mice
 - Bacteria
- METHOD DETAILS
 - Generation of *L. mexicana* Mutants and Complemented Strains
 - Bacterial Gene Expression and Protein Purification
 - GDP-Man-Dependent Mannosyltransferase Assays
 - Mannogen Phosphorylase Assays
 - HPTLC Analysis of Enzyme Products
 - Mannogen Profiling in Wild Type and Mutant Parasite Lines
 - GC-MS and LC-MS Analysis of Unlabeled and ^{13}C -/ ^2H -Labeled Mannogen
 - Chemical and Enzymatic Analyses
 - Macrophage Infection
 - Mouse Infections
 - Microscopy
 - Protein Production for Crystallization and SECIMALS
 - Crystallization of MTP1, MTP2 and MTP4
 - X-ray Data Collection and Crystal Structure Solution
 - SECIMALS
 - Phylogenetic Analysis
 - Homology Modeling and Molecular Docking
 - Multiple Sequence Alignments and Phylogenetic Tree Building for the GT108 MTPs, and GH130 Proteins
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.chom.2019.08.009>.

ACKNOWLEDGMENTS

We thank Professor Steve Beverley for the pXG plasmids; Dr. Thomas Ilg for the *L. mexicana* *Dgmp*, *Dpmi*, and *Dpmm* mutants; and Dr Eleanor Saunders for assistance with GC-MS analyses. MJM and MWP are NHMRC Principal Research and Senior Principal Research Fellows, respectively. GJD is a Royal Society Ken Murray Research Professor. DBA and DEVP are NHMRC (APP1072476) and Endeavour Fellows, respectively. We acknowledge grant support from the NHMRC (APP1100000 to MJM), the Australian Research Council (DP160100597 and DP180101957 to SJW), the BBSRC (BB/M011151/1 to AM) and ERC (ERC-2012-AdG-322942 to LS), the MRC (RCUK-CONFAP Grant to DRP), and the Jack Brockhoff Foundation (JBF 4186 to DBA). We thank Diamond Light Source for access to beamlines I03, I04 and I24 (proposals mx13587 and mx18598) and support from the Victorian Government Operational Infrastructure Support Scheme.

AUTHOR CONTRIBUTIONS

Conceptualization: M.F.S., J.E.R., M.J.M., S.J.W., and G.J.D.; Investigation: M.F.S., J.E.R., T.L.N., L.F.S., J.K., M.A.V.-L., S.A.C., L.S., D.E.V.P., E.H., A.M., T.W., and L.M.B.; Resources: P.L.v.d.P.; Writing: M.F.S., J.E.R., T.L.N., S.J.W., D.B.A., G.J.D., and M.J.M.; Supervision and Project Management: M.J.M., S.J.W., G.J.D., D.B.A., and M.W.P.; Funding Acquisition: M.J.M., S.J.W., G.J.D., D.B.A., and M.W.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 23, 2019

Revised: July 24, 2019

Accepted: August 15, 2019

Published: September 11, 2019

REFERENCES

- Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G.J., Wilson, K.S., and Cowtan, K.D. (2015). Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834.
- Banuls, A.L., Bastien, P., Pomares, C., Arevalo, J., Fisa, R., and Hide, M. (2011). Clinical pleiomorphism in human leishmaniasis, with special mention of asymptomatic infection. *Clin. Microbiol. Infect.* **17**, 1451–1461.
- Benzel, I., Weise, F., and Wiese, M. (2000). Deletion of the gene for the membrane-bound acid phosphatase of *Leishmania mexicana*. *Mol. Biochem. Parasitol.* **111**, 77–86.
- Burza, S., Croft, S.L., and Boelaert, M. (2018). Leishmaniasis. *Lancet* **392**, 951–970.
- Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920.
- Cobbold, S.A., Chua, H.H., Nijagal, B., Creek, D.J., Ralph, S.A., and McConville, M.J. (2016). Metabolic dysregulation induced in *Plasmodium falciparum* by dihydroartemisinin and other front-line antimalarial drugs. *J. Infect. Dis.* **213**, 276–286.
- Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
- Cuskin, F., Basle, A., Ladeveze, S., Day, A.M., Gilbert, H.J., Davies, G.J., Potocki-Ve´rone, G., and Lowe, E.C. (2015). The GH130 Family of Mannosidic Phosphorylases Contains Glycoside Hydrolases That Target β -1,2-Mannosidic Linkages in Candida Mannan. *J. Biol. Chem.* **290**, 25023–25033.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469.
- Dey, R., Joshi, A.B., Oliveira, F., Pereira, L., Guimaraes-Costa, A.B., Serafim, T.D., de Castro, W., Coutinho-Abreu, I.V., Bhattacharya, P., Townsend, S., et al. (2018). Gut microbes egested during bites of infected sand flies augment severity of leishmaniasis via inflammasome-Derived IL-1b. *Cell Host Microbe* **23**, 134–143.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Francois, J., and Parrou, J.L. (2001). Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* **25**, 125–145.
- Garami, A., and Ilg, T. (2001). Disruption of mannose activation in *Leishmania mexicana*: GDP-mannose pyrophosphorylase is required for virulence, but not for viability. *EMBO J.* **20**, 3657–3666.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.
- Ha, D.S., Schwarz, J.K., Turco, S.J., and Beverley, S.M. (1996). Use of the green fluorescent protein as a marker in transfected *Leishmania*. *Mol. Biochem. Parasitol.* **77**, 57–64.
- Hanssen, E., Dekiwadia, C., Riglar, D.T., Rug, M., Lemgruber, L., Cowman, A.F., Cyrklaff, M., Kudryashev, M., Frischknecht, F., Baum, J., and Ralph, S.A. (2013). Electron tomography of *Plasmodium falciparum* merozoites reveals core cellular events that underpin erythrocyte invasion. *Cell. Microbiol.* **15**, 1457–1472.
- He, Z., Zhang, H., Gao, S., Lercher, M.J., Chen, W.H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44** (W1), W236–W241.
- Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638.
- Jackson, A.P., Otto, T.D., Aslett, M., Armstrong, S.D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J.M., Dacks, J.B., et al. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr. Biol.* **26**, 161–172.
- Jubb, H.C., Higuero, A.P., Ochoa-Montaño, B., Pitt, W.R., Ascher, D.B., and Blundell, T.L. (2017). Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **429**, 365–371.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
- Keegan, F.P., and Blum, J.J. (1992). Utilization of a carbohydrate reserve comprised primarily of mannose by *Leishmania donovani*. *Mol. Biochem. Parasitol.* **53**, 193–200.
- Kitaoka, M. (2015). Diversity of phosphorylases in glycoside hydrolase families. *Appl. Microbiol. Biotechnol.* **99**, 8377–8390.
- Kloehn, J., Saunders, E.C., O’Callaghan, S., Dagley, M.J., and McConville, M.J. (2015). Characterization of metabolically quiescent *Leishmania parasites* in murine lesions using heavy water labeling. *PLoS Pathog.* **11**, e1004683.
- Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J., and Lesk, A.M. (2006). MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559–574.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486.
- LeBowitz, J.H., Coburn, C.M., McMahon-Pratt, D., and Beverley, S.M. (1990). Development of a stable *Leishmania* expression vector and application to the study of parasite surface antigen genes. *Proc. Natl. Acad. Sci. USA* **87**, 9736–9740.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495.
- Louradour, I., Monteiro, C.C., Inbar, E., Ghosh, K., Merkhofer, R., Lawyer, P., Paun, A., Smelkinson, M., Secundino, N., Lewis, M., et al. (2017). The midgut microbiota plays an essential role in sand fly vector competence for *Leishmania major*. *Cell. Microbiol.* **19**, e12755.
- MacNeill, G.J., Mehrpouyan, S., Minow, M.A.A., Patterson, J.A., Tetlow, I.J., and Emes, M.J. (2017). Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *J. Exp. Bot.* **68**, 4433–4453.
- McConville, M.J., Homans, S.W., Thomas-Oates, J.E., Dell, A., and Bacic, A. (1990). Structures of the glycoinositolphospholipids from *Leishmania major*. A family of novel galactofuranose-containing glycolipids. *J. Biol. Chem.* **265**, 7385–7394.
- McConville, M.J., Saunders, E.C., Kloehn, J., and Dagley, M.J. (2015). *Leishmania* carbon metabolism in the macrophage phagolysosome: feast or famine? *F1000Res.* **4** (F1000 Faculty Rev), 938.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674.
- McNicholas, S., Potterton, E., Wilson, K.S., and Noble, M.E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 386–394.
- Melamud, E., Vastag, L., and Rabinowitz, J.D. (2010). Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**, 9818–9826.
- Misslitz, A., Mottram, J.C., Overath, P., and Aebischer, T. (2000). Targeted integration into a rRNA locus results in uniform and high level expression of

- transgenes in *Leishmania* amastigotes. *Mol. Biochem. Parasitol.* *107*, 251–261.
- Murshudov, G.N., Skuba'k, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 355–367.
- Naderer, T., Wee, E., and McConville, M.J. (2008). Role of hexosamine biosynthesis in *Leishmania* growth and virulence. *Mol. Microbiol.* *69*, 858–869.
- Naderer, T., Heng, J., and McConville, M.J. (2010). Evidence that intracellular stages of *Leishmania major* utilize amino sugars as a major carbon source. *PLoS Pathog.* *6*, e1001245.
- Nakae, S., Ito, S., Higa, M., Senoura, T., Wasaki, J., Hijikata, A., Shionyu, M., Ito, S., and Shirai, T. (2013). Structure of novel enzyme in mannan biodegradation process 4-O- β -D-mannosyl-D-glucose phosphorylase MGP. *J. Mol. Biol.* *425*, 4468–4478.
- Oliveira, D.M., Costa, M.A.F., Chavez-Fumagalli, M.A., Valadares, D.G., Duarte, M.C., Costa, L.E., Martins, V.T., Gomes, R.F., Melo, M.N., Soto, M., et al. (2012). Evaluation of parasitological and immunological parameters of *Leishmania chagasi* infection in BALB/c mice using different doses and routes of inoculation of parasites. *Parasitol. Res.* *110*, 1277–1285.
- Opperdoes, F.R., Butenko, A., Flegontov, P., Yurchenko, V., and Lukeš, J. (2016). Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J. Eukaryot. Microbiol.* *63*, 657–678.
- Peng, J., and Xu, J. (2011). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins* *79* (Suppl 10), 161–171.
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., et al. (2018). CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Struct. Biol.* *74*, 68–84.
- Previato, J.O., Xavier, M.T., Brazil, R.P., Gorin, P.A., and Mendonc, a-Previato, L. (1984). Formation of (1—2)-linked beta-D-mannopyranan by *Leishmania mexicana amazonensis*: relationship with certain Crithidia and Herpetomonas species. *J. Parasitol.* *70*, 449–450.
- Ralton, J.E., Naderer, T., Piraino, H.L., Bashtannyk, T.A., Callaghan, J.M., and McConville, M.J. (2003). Evidence that intracellular beta1-2 mannan is a virulence factor in *Leishmania* parasites. *J. Biol. Chem.* *278*, 40757–40763.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* *42*, W320–W324.
- Saburi, W. (2016). Functions, structures, and applications of cellobiose 2-epimerase and glycoside hydrolase family 130 mannoside phosphorylases. *Biosci. Biotechnol. Biochem.* *80*, 1294–1305.
- Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins* *23*, 318–326.
- Saunders, E.C., Ng, W.W., Kloehn, J., Chambers, J.M., Ng, M., and McConville, M.J. (2014). Induction of a stringent metabolic response in intracellular stages of *Leishmania mexicana* leads to increased dependence on mitochondrial metabolism. *PLoS Pathog.* *10*, e1003888.
- Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., Wasaki, J., and Ito, S. (2011). New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. *Biochem. Biophys. Res. Commun.* *408*, 701–706.
- Sernee, M.F., Ralton, J.E., Dinev, Z., Khairallah, G.N., O'Hair, R.A., Williams, S.J., and McConville, M.J. (2006). *Leishmania* beta-1,2-mannan is assembled on a mannose-cyclic phosphate primer. *Proc. Natl. Acad. Sci. USA* *103*, 9458–9463.
- Sfihi-Loualia, G., Hurtaux, T., Fabre, E., Fradin, C., Me' e, A., Pourcelot, M., Maes, E., Bouckaert, J., Mallet, J.-M., Poulain, D., et al. (2016). *Candida albicans* β -1,2-mannosyltransferase Bmt3 prompts the elongation of the cell-wall phosphopeptidomannan. *Glycobiology* *26*, 203–214.
- Shi, L., Sutter, B.M., Ye, X., and Tu, B.P. (2010). Trehalose is a key determinant of the quiescent metabolic state that fuels cell cycle progression upon return to growth. *Mol. Biol. Cell* *21*, 1982–1990.
- Simonetti, F.L., Teppa, E., Chernomoretz, A., Nielsen, M., and Marino Buslje, C. (2013). MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Res.* *41*, W8–W14.
- Singh, O.P., Hasker, E., Sacks, D., Boelaert, M., and Sundar, S. (2014). Asymptomatic *Leishmania* infection: a new challenge for *Leishmania* control. *Clin. Infect. Dis.* *58*, 1424–1429.
- Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* *41*, 641–643.
- Stuart, K., Brun, R., Croft, S., Fairlamb, A., Grtler, R.E., McKerrow, J., Reed, S., and Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. *J. Clin. Invest.* *118*, 1301–1310.
- Tsuda, T., Nihira, T., Chiku, K., Suzuki, E., Arakawa, T., Nishimoto, M., Kitaoka, M., Nakai, H., and Fushinobu, S. (2015). Characterization and crystal structure determination of β -1,2-mannobiose phosphorylase from *Listeria innocua*. *FEBS Lett.* *589* (24 Pt B), 3816–3821.
- van der Peet, P., Gannon, C.T., Walker, I., Dinev, Z., Angeli, M., Tam, S., Ralton, J.E., McConville, M.J., and Williams, S.J. (2006). Use of click chemistry to define the substrate specificity of *Leishmania* beta-1,2-mannosyltransferases. *ChemBioChem* *7*, 1384–1391.
- van der Peet, P., Ralton, J.E., McConville, M.J., and Williams, S.J. (2012). Discovery of inhibitors of *Leishmania* β -1,2-mannosyltransferases using a click-chemistry-derived guanosine monophosphate library. *PLoS One* *7*, e32642.
- van Heerden, J.H., Wortel, M.T., Bruggeman, F.J., Heijnen, J.J., Bollen, Y.J.M., Planque' , R., Hulshof, J., O'Toole, T.G., Wahl, S.A., and Teusink, B. (2014). Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* *343*, 1245114.
- Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* *43*, 186–190.
- Winter, G., Lobley, C.M.C., and Prince, S.M. (2013). Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* *69*, 1260–1273.
- Ye, Y., Saburi, W., Odaka, R., Kato, K., Sakurai, N., Komoda, K., Nishimoto, M., Kitaoka, M., Mori, H., and Yao, M. (2016). Structural insights into the difference in substrate recognition of two mannoside phosphorylases from two GH130 subfamilies. *FEBS Lett.* *590*, 828–837.
- Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kbler, J., Lozajic, M., Gabler, F., Sding, J., Lupas, A.N., and Alva, V. (2018). A completely reimplemented MPI bioinformatics Toolkit with a new HHpred server at its core. *J. Mol. Biol.* *430*, 2237–2243.

STAR+METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
XL10-Gold	Agilent	Cat#210515
XL-1-Blue	Agilent	Cat#200249
BL21(DE3)pLysS	Agilent	Cat#200132
Biological Samples		
<i>Canavalia ensiformis</i> (jack bean) α -mannosidase	Sigma	Cat#M7257
<i>Helix pomatia</i> (snail) β -mannosidase	Sigma	Cat#M9400
Chemicals, Peptides, and Recombinant Proteins		
OM1	Williams laboratory	van der Peet et al., 2012
OM4	This study	N/A
GDP-[2- ³ H]-Man	McConville laboratory	N/A
GDP-Man (guanosine 5'-diphospho-D-mannose sodium salt)	Sigma	Cat#G5131-50 mg
GDP (Guanosine 5'-diphosphate sodium salt)	Sigma	Cat#G7127-100 mg
α -Man1P	Sapphire Bioscience	Cat#M185010
β -Man1P	Williams laboratory	N/A
D-Glucose-U- ¹³ C	Sigma	Cat#297046
² H ₂ O	Cambridge isotope laboratories	Cat#DLM-4-99-1000
Critical Commercial Assays		
PiColorLock Gold phosphate assay	Innova Biosciences	Cat#303-0030
Deposited Data		
<i>L. mexicana</i> LmxM.10.1230 no ligand (MTP1)	This study	PDB: 6Q4W
<i>L. mexicana</i> LmxM.10.1240 no ligand (MTP2)	This study	PDB: 6Q4X
<i>L. mexicana</i> LmxM.10.1240 + Man (MTP2 + Man)	This study	PDB: 6Q4Y
<i>L. mexicana</i> LmxM.10.1240 D94N + β -1,2-mannobiose (MTP2 D94N + β -1,2-mannobiose)	This study	PDB: 6Q4Z
<i>L. mexicana</i> LmxM.10.1260 + Pi (MTP4 + Pi)	This study	PDB: 6Q50
Experimental Models: Cell Lines		
Δ gmp	T. Ilg	Garami and Ilg, 2001
Δ pmi	T. Ilg	Garami and Ilg, 2001
Δ pmm	T. Ilg	Garami and Ilg, 2001
Δ mtp1-7	This study	N/A
Δ mtp3-7	This study	N/A
Δ mtp2	This study	N/A
pRib-MTP1	This study	N/A
Δ mtp1-7::pRib-MTP1	This study	N/A
Δ mtp1-7::pRib-MTP2	This study	N/A
Δ mtp1-7::pX-MTP2	This study	N/A
Δ mtp1-7::pRib-MTP-1/pX-MTP2	This study	N/A
Δ mtp1-7::pRib-MTP1/pX-MTP4	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Δmtp1-7::pRib-MTP4</i>	This study	N/A
<i>Δmtp1-7::pX-MTP4</i>	This study	N/A
<i>Δmtp1-7::pX-MTP2 + pRib-MTP4</i>	This study	N/A
<i>Δmtp1-7::pX-MTP3+4</i>	This study	N/A
<i>Δmtp1-7::pX-MTP3+4/ pRib-MTP5+6</i>	This study	N/A
<i>Δmtp1-7::pX-MTP4+7</i>	This study	N/A
<i>Δmtp1-7::pX-MTP4+7/ pRib-MTP5+6</i>	This study	N/A
<i>Δmtp3-7::pRib-MTP4</i>	This study	N/A
<i>Δmtp3-7::pX-MTP2/ pRib-MTP4</i>	This study	N/A
<i>Δmtp2::pX-MTP2</i>	This study	N/A
<i>Δmtp2::pRib-MTP2</i>	This study	N/A
Experimental Models: Organisms/Strains		
<i>L. mexicana</i> (MNYC/ BZ/ 62/M379)	ATCC	Cat#50156
BALB/c mice	Animal Resources Centre, Canning Vale, WA, Australia	Cat#BC
Oligonucleotides		
Primers for MTP knockout constructs and confirmation, see Table S4	This study	N/A
Primers for MTP expression, see Table S4	This study	N/A
Primers for deletion mutants, see Table S4	This study	N/A
Recombinant DNA		
pET-28a(+)	Novagen	Cat#69864-3
pGEX-6p-3	GE Healthcare	Cat#28-9546-51
pGEX-4T-2	GE Healthcare	Cat#28-9545-50
pRIBII	JC Mottram laboratory	Misslitz et al., 2000
pX-Neo	SM Beverley laboratory	LeBowitz et al., 1990
pX-Hyg	This study	N/A
pBluescript II SK+	Statagene	GenBank Acc#X52328
pXG-BLEO (PHLEO)	SM Beverley laboratory http://beverleylab.wustl.edu/plasmids_vectors.html	B3324
pXG-SAT	SM Beverley laboratory http://beverleylab.wustl.edu/plasmids_vectors.html	B2352
pXG-GFP+	SM Beverley laboratory http://beverleylab.wustl.edu/plasmids_vectors.html	B2863
Knockout constructs (pBluescript-SKII(+)) made, Table S5	This study	N/A
Expression (including deletion) constructs (pGEX-4T-2, pGEX-6p-3 and pET28a(+)), Table S5	This study	N/A
Software and Algorithms		
Prism 5	GraphPad	https://www.graphpad.com/
AXIOVISION 4.8	Zeiss	N/A
Modeler v9.20	Sali et al., 1995	https://salilab.org/modeller/9.20/release.html
SYBYL-X 2.1.1	Certara	https://www.certara.com/
PROCHECK	Laskowski et al., 1996	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
PyMOL Molecular graphics system, version 1.8.2.2	Schrodinger	https://pymol.org/2/
MUSTANG v3.2.3	Konagurthu et al., 2006	http://lcb.infotech.monash.edu.au/mustang/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ETE3 v3.1.1	Huerta-Cepas et al., 2016	http://etetoolkit.org/
PhyML v3.12	Guindon et al., 2010	http://www.atgc-montpellier.fr/phyml/versions.php/
TreeDyn v198.3	Dereeper et al., 2008	http://www.phylogeny.fr/one_task.cgi? task_type=treedyn
ESPript3.0	Robert and Gouet, 2014	http://esript.ibcp.fr/ESPript/ESPript/ index.php
ClustalW	N/A	http://www.mybiosoftware.com/bioedit-7-0-9- biological-sequence-alignment- editor.html
ClustalU	Zimmermann et al., 2018	https://toolkit.tuebingen.mpg.de/
RaptorX	Peng and Xu, 2011	http://raptorX.uchicago.edu/ StructurePrediction/predict/
MAFFT v7.0	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/server/
Evolview	He et al., 2016	https://www.evolgenius.info/evolview/
MISTIC	Simonetti et al., 2013	N/A
Arpeggio	Jubb et al., 2017	http://biosig.unimelb.edu.au/arpeggioweb/
WebLogo 3	Crooks et al., 2004	N/A
DIALS		https://dials.diamond.ac.uk/
Xia2	Winter et al., 2013	https://xia2.github.io/
Aimless	Potterton et al., 2018	http://www.ccp4.ac.uk/html/aimless.html
Phaser	McCoy et al., 2007	https://www.phaser.cimr.cam.ac.uk/index. php/Phaser_Crystallographic_Software; http://www.ccp4.ac.uk/
CCP4mg	McNicholas et al., 2011	http://www.ccp4.ac.uk/MG/
Chainsaw	Stein, 2008	http://www.ccp4.ac.uk/html/ chainsaw.html
Refmac5	Murshudov et al., 2011	https://www2.mrc-lmb.cam.ac.uk/groups/ murshudov/content/refmac/refmac.html
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/ personal/pemsley/cool/
Ccp4i2 GUI	Potterton et al., 2018	https://www.ccp4.ac.uk/ccp4i_main.php
Privateer	Agirre et al., 2015	http://www.ccp4.ac.uk/html/privateer.html
ASTRA software		https://www.wyatt.com/products/ software/astra.html
MSD-ChemStation	Agilent Technologies	N/A
MassHunter	Agilent Technologies	N/A
MS convert	Chambers et al., 2012	http://proteowizard.sourceforge.net/ tools.shtml
Maven	Melamud et al., 2010	http://genomics-pubs.princeton.edu/ mzroll/index.php

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Malcolm J. McConville (malcolmm@unimelb.edu.au).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Parasites

L. mexicana (MNYC/BZ/62/M379) promastigotes were grown in Roswell Park Memorial Institute (RPMI)-1640 medium (Invitrogen) containing 10% heat inactivated fetal bovine serum (FBS, Invitrogen), pH 7.4 at 27°C. Mutant and complemented parasite lines were cultivated in the presence of appropriate antibiotics: Geneticin (100 mg/mL, GIBCO), Puromycin (20 mg/mL, Sigma), Hygromycin

B (50 mg/mL, InvivoGen), Bleomycin (10 mg/mL, Calbiochem), Nourseothricin (100 mg/mL, JenaBioscience). Stationary phase promastigotes were induced to differentiate to amastigotes by cultivation in RPMI, 20% FBS, pH 5.5 at 33°C (Saunders et al., 2014). For ¹³C-glucose labeling experiments, *L. mexicana* promastigotes or amastigotes were suspended in RPMI-1640 medium containing ¹³C-U-glucose (6 mM) and harvested with rapid chilling to quench metabolism at indicated time points (Saunders et al., 2014). For ³H-mannose labeling experiments, *L. mexicana* wild type promastigotes (3 × 10⁸) were pulse labeled in glucose-free RPMI-1%BSA medium containing D-[2-³H]-mannose (50 mCi/mL) for 5 min, harvested by rapid centrifugation (1,000 × g, 30 s) and washed in ice cold phosphate buffered saline (× 3), prior to hypotonic lysis.

Bone Marrow Derived Macrophages

Bone marrow macrophages (BMDM) were isolated from BALB/c mice. Cells were flushed from tibia of hind-legs of mice seeded in TC Petri dishes at 1 × 10⁷ cells/plate in 10 mL RPMI containing 15% FBS, 20% L929-cell medium (containing macrophage colony stimulating factor; MCSF) and incubated at 37°C, 5% CO₂. After 24 h non-adherent progenitor cells from each dish were transferred to two non-TC Petri dishes. Progenitor cells converted to macrophages within 3-4 days. At day 3 after differentiation, 1 mL extra L929-cell medium was added to the plate. At day 5, cells were removed from the plates and plated onto coverslips in 24 well plates at 1 × 10⁵ cells/well in RPMI containing 15% FBS, 10% L929-cell medium.

Mice

Mice usage was approved by the Institutional Animal Care and Use Committee of the University of Melbourne (ethics number 1212647.1). All animal experiments were performed in accordance with the Australian National Health Medical Research Council (NHMRC) (Australian code of practice for the care and use of animals for scientific purposes, 8th edition, 2013, ISBN: 1864965975). Mice used in this study were maintained in the biological research facility of the Bio21 Molecular Science and Biotechnology Institute under specific pathogen free (SPF) conditions according to institutional guidelines. BALB/c mice were gender (female) and age matched (6-8 weeks) within individual experiments. Mice were bred in house or purchased from Animal Resources Centre, Canning Vale, WA, Australia (product code: BC; SPF status in health report). Animals were not subject to water or food restrictions and monitored daily by facility staff.

Bacteria

The strain and source of all the bacteria used in this study are detailed in Key Resources Table.

METHOD DETAILS

Generation of *L. mexicana* Mutants and Complemented Strains

The *L. mexicana* $\Delta mtp1-7$ null mutant was generated by homologous replacement of the entire LmxM.10.1230-LmxM.10.1290 gene locus with bleomycin and nourseothricin resistant cassettes. The bleomycin and nourseothricin resistance cassettes were digested from pXG-BLEO and pXG-SAT, respectively, using the EcoRI and BamHI restriction sites, and cloned into the pBluescript II SK+ vector. The 745 bp 5'UTR region of LmxM.10.1230 was amplified using forward and reverse primers (Table S4) and cloned into pBluescriptII SK+ vector already containing the resistance cassettes using the *HindIII* and *BstBI* restriction sites. The 952 bp 3'UTR region of LmxM.10.1290 was cloned into these plasmids after amplification and digest using the BamHI and *XbaI* restriction sites. The resulting bleomycin and nourseothricin knock-out constructs were verified by sequencing. *L. mexicana* was transfected with 2-5 mg *HindIII* and *XbaI* digested and gel-purified linearized DNA fragments. Drug resistant clones were screened by PCR for the presence of the resistance cassettes using a forward primer in the 5'UTR outside the cloned region and a reverse primer for the resistance bleomycin and nourseothricin cassettes and the absence of the genes using 5' and 3' primers of each gene (Table S4).

The *L. mexicana* $\Delta mtp3-7$ null mutant was generated using the above constructs, but with the 5'UTR region of LmxM.10.1230 replaced with the 5'UTR 896 bp region of LmxM.10.1250 amplified from wild type cDNA using forward and reverse primers (Table S4) and was cloned using the *HindIII* and *BstBI* restriction sites. *L. mexicana* was transfected as described below and null mutants were screened as described above, with its specific primers (Table S4)

The *L. mexicana* $\Delta mtp2$ null mutant was generated by sequential deletion of both LmxM.10.1240 genes using a linearized constructs of plasmids that were constructed as follows: The 740 bp 5' UTR region of LmxM.10.1240 was amplified using primers indicated in Table S4 and cloned using *HindIII* and *BstBI* restriction sites into a pBluescript II SK+ vector already containing the either a bleomycin or nourseothricin resistance gene and the 896 bp 3' UTR region of LmxM.10.1240 was amplified using primers indicated in Table S4 and cloned using BamHI and *XbaI* restriction sites. Positive clones were verified as described above (specific primers indicated in Table S4)

Complementation of the $\Delta mtp1-7$, $\Delta mtp3-7$ and $\Delta mtp2$ mutants was achieved by expression of individual genes from the pXG episome or by stable insertion into the ribosomal RNA loci as described below (Misslitz et al., 2000; Benzel et al., 2000). LmxM.10.1260 was expressed in the $\Delta mtp1-7$ mutant as a C-terminal GFP fusion protein using pXG-GFP+ (B2863) (Ha et al., 1996), using the primers from Table S4 with the *XmaI* and *EcoRV* restriction sites. For stable integration into the ribosomal RNA locus, LmxM.10.1260 and LmxM.10.1230 genes were amplified using *BglII* and *NotI* containing primers and both were cloned into the pRIBII plasmid using *BglII* and *NotI* restriction sites (Benzel et al., 2000; Misslitz et al., 2000).

LmxM.10.1240 was expressed using an episomal expression construct without tag under the selection of hygromycin. LmxM.10.1240 was amplified using primers from Table S4 and cloned using *Bam*HI and *Not*I restriction sites in pX-Hyg derived from pX-Neo vector by replacing the neo/DHFR between *Spe*I sites of pX with the hygromycin gene using splice overlap PCR. LmxM.10.1270 and LmxM.10.1280 were amplified with the intergenic region and cloned in pRIBII plasmid (Benzel et al., 2000; Mislitz et al., 2000) for stable integration into the ribosomal RNA locus for constitutive expression. The LmxM.10.1270-80 region was amplified with the forward primers of LmxM.10.1270 and reverse primer of LmxM.10.1280 (see Table S4).

pX-Hyg constructs with LmxM.10.1250 and LmxM.10.1290 were made by inserting *Bgl*III and *Not*I amplified material in *Bam*HI and *Not*I digested plasmid using primers listed in Table S4.

For transfection, promastigotes (mid-log phase, 4×10^7 cells/transfection) were suspended in chilled electroporation buffer (EPB; 21 mM HEPES, 137 mM NaCl, 5 mM KCl, 6 mM glucose, 0.7 mM Na_2PO_4 pH 7.4), in a 4 mm cuvette and electroporated with 3-5 mg of linearized DNA or 5-10 mg of circular (uncut) plasmid DNA and pulsed twice at 1,700 V and 25 mF with a 10 s interval in a BioRad Gene Pulser. Cells were transferred to SDM-79 media containing 10% FBS and 100 units/mL penicillin/streptomycin. After 24 h the medium was supplemented with selection drugs.

Bacterial Gene Expression and Protein Purification

L. mexicana MTPs were expressed in *E. coli* using pGEX-6p-3, pGEX-4T-2 or pET-28a(+) plasmids resulting in N-terminal GST-tag for pGEX or His-tag for pET plasmid (Table S5). The MTP encoding gene was amplified from cDNA with either a *Bgl*III or *Bam*HI restriction site and *Not*I and was cloned into pGEX vectors using *Bam*HI and *Not*I sites or into pET-28a(+) using *Nde*I and *Not*I restriction sites. MTP constructs in the pGEX-6p-3 expression vector were used as templates for mutagenesis. Mutations were introduced with QuickChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's instructions. Once sequences were confirmed, the DNA was subsequently amplified with primers containing *Nde*I and *Not*I restriction sites for cloning into pET-28(a)+.

Recombinant proteins were expressed in BL21(DE3)-pLysS at 37°C at 220 rpm until OD_{600} reached approximately 0.8 (Table S5). The culture was allowed to adapt to 16°C for 2 h (240 rpm), before being induced with 0.1 mM IPTG for > 16 h at 16°C at 240 rpm. Protein was extracted from cell pellets in 4 mL Bugbuster (Merck Millipore) containing protease inhibitors without EDTA (Roche) per 50 mL culture pellet for 10 min.

GST-tagged protein was purified from 25,000 x g supernatants using Glutathione Sepharose 4B (GE-healthcare) washed with 10 volumes of PBS before being eluted with 10 mM glutathione in 50 mM Tris pH 8.0. His-tagged protein was purified from 25,000 x g supernatants using Complete His-Tag Purification resin (Roche) (1 ml) equilibrated with 5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7.9. The resin was washed with increased imidazole concentrations from 5, 20, 45 to 60 mM in 500 mM NaCl, 20 mM Tris-HCl, pH 7.9 and protein was eluted with 250 mM imidazole in 500 mM NaCl, 20 mM Tris-HCl, pH 7.9. Amicon Ultra-15 (10K) centrifugal filters (Merck Millipore) were used to exchange buffers to low ionic strength assay buffer (5 mM NaHEPES-NaOH buffer, pH 7.4 with 5 mM MgCl_2 , 1 mM MnCl_2 , 2 mM EGTA and 2 mM DTT) or with 20 mM MES, 2 mM EGTA, pH 5.5.

GDP-Man-Dependent Mannosyltransferase Assays

Mannosyltransferase activities in parasite lysates or of recombinant proteins were measured using octyl- α -mannosides (OM1) or octyl-mannogen (DP 4 (OM4), or DP 1-16 (OM1-16)) as acceptors and either unlabeled GDP-Man or GDP-[2- ^3H]-Man as donors. The use of OM_n acceptors facilitated the recovery of products by solvent phase partitioning and analysis by HPTLC and were used with similar affinity as corresponding native mannogen oligomers. In brief, cell lysates were prepared by hypotonic lysis of *L. mexicana* wild type or *Dgmp* promastigotes harvested at late log growth phase (van der Peet et al., 2012). Promastigotes (4×10^7 /assay) were washed in cold PBS then suspended in 50 ml of cold hypotonic lysis buffer (1 mM HEPES-NaOH buffer, pH 7.4, containing 2 mM EGTA, 2 mM DTT and protease inhibitors without EDTA (Roche)) and incubated for 10 min on ice. Cell lysis was monitored by microscopy and facilitated by brief bath sonication (3-5 s) if required. Lysed cells were centrifuged (2000 x g, 4°C, 4 min) and the supernatant removed. The pellet was washed with 10 vol of low ionic strength assay buffer (5 mM NaHEPES-NaOH buffer, pH 7.4 with 2 mM EGTA, 2 mM DTT and protease inhibitors without EDTA (Roche)). The supernatant was adjusted to 5 mM HEPES buffer, pH 7.4 with 2 mM EGTA, 2 mM DTT and protease inhibitors without EDTA (Roche) and the pellet fractions suspended in the same buffer containing 0.1% TX-100, to achieve the same cell equivalents. Assays were initiated by addition of acceptor (native mannogen or OM1), and GDP-[2- ^3H]-Man (50 mM) /unlabeled GDP-Man (0.05-10 mM), and incubated at 27°C for indicated times. Reactions were stopped by addition of chloroform/methanol (1:2 v/v) to give a final ratio of chloroform/ methanol/ aqueous 1:2:0.8 (v/v), centrifuged (15,000 rpm, 5 min), and the supernatant dried under nitrogen. OM_n products were recovered by phase partitioning between water-saturated 1-butanol (200 ml) and water (150 ml), the 1-butanol phase dried and analyzed by high performance TLC (HPTLC). For analysis of longer OM products, extracts were desalted by passage down a small column (400 ml) of Ag 50-X12(H⁺) over Ag 4-X4(OH⁻) (Bio-Rad) and freeze-dried prior to HPTLC analysis. For assay of recombinant proteins, final reaction mixtures contained 50 ml of 20 mM MES-NaOH, pH 5.5, 5-10 mg purified recombinant protein, 0.05-5 mM OM1, OM4 or OM1-16 as acceptor, and 0.05-10 mM GDP-Man or GDP as donor. Reactions were initiated by addition of protein and reactions (27°C, 5-90 min) stopped and processed as described above. OM4 was generated by incubating *L. mexicana* MTP2 with OM1 (5 mM) and GDP-Man (10 mM) at 27°C for 90 min, and purified by preparative HPTLC on Silica Gel 60 coated aluminum backed HPTLC sheets (Merck), developed

twice in 1-butanol: ethanol: water (4:3:3 v/v). Silica bands containing individual octyl-mannogen species were scraped from the HPTLC plates and extracted in water-saturated 1-butanol (2 x 200 ml).

Mannogen Phosphorylase Assays

L. mexicana promastigote lysates prepared from wild type or the *Dpmm* mutant line (lacking the enzyme phosphomannose mutase) as described above were incubated with ³H-mannogen. Hypotonic lysates were adjusted to contain 5 mM HEPES buffer, pH 7.4, protease inhibitors without EDTA (Roche), EDTA, and 10⁶ cell equivalents/mL. ³H-mannogen was prepared by metabolically labeling stationary phase promastigotes of *L. mexicana Dpmi* in glucose-free RPMI-1%BSA medium containing D-[2-³H]-mannose for 4 h at 27°C (Sernee et al., 2006). The absence of phosphomannose isomerase in this mutant line maximizes efficiency of labeling. Labeled parasites were extracted in hot water (100°C, 10 min) and mannogen oligomers desalted over a small ion exchange column of AG 50W-X12(H⁺) over AG 4-X4(OH⁻) (Bio-Rad) (100 ml each) and freeze-dried. Phosphorylase assays were initiated by incubating 100 ml lysate and ³H-mannogen (10,000 dpm) at 27°C for 2 h. Assays were stopped by addition of chloroform/methanol (1:2 v/v) as described above and neutral and phosphorylated sugar products analyzed by HPAEC. For analysis of recombinant MTP proteins, assays contained 50 ml 20 mM MES-NaOH, pH 5.5, 5-10 mg recombinant protein and either NaHPO₄ (0.5-5 mM) or Man1P (1 mM) in the case of reverse phosphorolysis, and OM4 (200 mM mannose equivalent) or unlabeled mannogen (2 mM mannose equivalent) as substrate. After incubation at 27°C for 90 min, reactions were stopped by heating at 100°C for 2 min and either the C8Man_x products were analyzed by HPTLC after desalting as described above, or release of phosphate was measured using the PiColorLock Gold phosphate assay (Innova Biosciences).

HPTLC Analysis of Enzyme Products

Desalted reaction products were analyzed by HPTLC on Silica Gel 60 aluminum-backed HPTLC sheets (Merck), developed twice in 1-butanol: ethanol: water (4:3:3 v/v) for neutral glycans or polar samples, or once in chloroform: methanol: 13 M ammonia: 1 M ammonium acetate: water (180:140:9:9:23 v/v) for apolar samples (Ralton et al., 2003). Unlabeled glycans were visualized by staining with 0.2% orcinol (Sigma-Aldrich) in 10% H₂SO₄, 80% ethanol and developed at 100°C. Labeled species were detected by fluorography after spraying the HPTLC sheets with EN²HANCE spray (PerkinElmer Life Sciences) and exposing them to Biomax MR film (Kodak) at -80°C.

Mannogen Profiling in Wild Type and Mutant Parasite Lines

Parasites were metabolically quenched by transferring aliquots of culture suspension into 15 mL plastic tube and immersion into a dry ice-ethanol bath for 20 s (until suspension reached 5°C). The chilled suspension (8 x 10⁷ parasites) was centrifuged (4,000 rpm, 0°C, 10 min), washed three times with ice-cold PBS, then suspended in 200 ml ice-cold water and immersed in a boiling water bath for 20 min. The extracts were centrifuged (16,000 x g, 15 min) and the supernatant containing mannogen desalted over a small ion exchange column of AG 50W-X12(H⁺) over AG 4-X4(OH⁻) (Bio-Rad). In some cases, extracts were hydrolyzed in 80 mM TFA (100°C, 10 min) to cleave acid labile phosphate mono-esters, prior to desalting. Sugar-phosphates, neutral sugars and mannogen oligosaccharides were analyzed by HPAEC, using a Dionex GP-50 gradient pump, a CarboPac PA-1 column (4 x 250 mm), with a PA-1 guard column and an ED₅₀ integrated pulsed amperometric detector (PAD). The system was controlled and data analyzed by Chromeleon version 6.50 software. Sugar phosphates were resolved on a gradient of 1 mM sodium hydroxide (E1) and 1 mM sodium hydroxide containing 1 M sodium acetate (E2) at a flow rate of 1 mL/min: T₀ = 98% (v/v) E1; T₁₅ = 98% (v/v) E1; T₃₀ = 90% (v/v) E1; T₄₀ = 80% (v/v) E1; T₅₅ = 100% (v/v) E2; T₆₅ = 100% (v/v) E2. Neutral sugars/oligosaccharides were resolved on a gradient of 75 mM NaOH (E1) and 75 mM NaOH containing 250 mM sodium acetate (E2) at a flow rate of 0.6 mL/min: T₀ = 100% (v/v) E1; T₅ = 100% (v/v) E1; T₄₀ = 100% (v/v) E2, T₆₀ = 100% (v/v) E2.

GC-MS and LC-MS Analysis of Unlabeled and ¹³C-/²H-Labeled Mannogen

Mannogen was extracted from promastigotes in hot water as described above, desalted and resuspended in 20% 1-propanol. For analysis of mannose isotopomers in the mannogen pool, an aliquot of the mannogen extract was subjected to solvolysis in 0.5 M methanolic HCl (50 ml, Supelco) in sealed glass capillaries under vacuum or in deactivated GC-MS glass vial inserts (Agilent part no: 5181-8872) (80°C, 16 h). Samples were neutralized by addition of pyridine (10 ml), dried under nitrogen, and derivitized with N,O-Bis(trimethylsilyl) trifluoroacetamide containing 1% TMCS (40 ml; Thermo Scientific) prior to being analyzed on either an Agilent 6890/5973 GC-MS system in electron impact mode or an Agilent 7890/5975 system in chemical ionization mode. Separation was achieved on a 30 m DB-5ms GC capillary column (J&W scientific; 250 mm i.d., 0.25 mm film thickness), incorporating a 10 m inert duraguard section. The levels of mannogen-derived mannose were quantified by GC-MS in EI mode by determining the total area under the curve for its specific peak (an authentic standard was used to confirm the retention time and ion fragmentation of mannose following methanolysis and silylation). Equal cell equivalents (1x10⁷ cells) were analyzed for each genotype and the mannose levels were normalized to an internal standard added during the hot water extraction. Selected ion monitoring was used to quantify the labeling in mannose and its mass isotopologs (M₀-M₆; 467-473 m/z; a fragment of the mannose derivative associated with loss of CH₄ [M+1-16]⁺). The ²H-labeling in mannogen-derived mannose was determined as the excess molar enrichment in the M₁ mass isotopolog (468 m/z) over the molecular ion M₀ (467 m/z). GC-MS data was analyzed using MSD ChemStation and MassHunter (Agilent technologies). For LC-MS analysis detection of GDP and GDP-Man, reaction mixtures were separated on a Merck ZIC-pHILIC column (5 mM, 150 x 4.6 mm, Millipore) (Cobbold et al., 2016). A gradient of water with 20 mM ammonium carbonate (solvent A) and

acetonitrile (solvent B) ran from 80%–20% solvent B from 0.5 to 15 min, then 20%–5% between 15 and 20 min, before returning to 80% at 20.5 min and held until 29.5 min. MS detection was performed on an Agilent QTOF mass spectrometer 6545 operating in negative ESI mode and data analyzed using the MAVEN software package (Melamud et al., 2010). Following alignment, metabolites were assigned using exact mass (< 10 ppm) and retention time (compared to standards of GDP and GDP-Man). The area top for each positively assigned metabolite was integrated and the concentration of each metabolite in each sample was determined from standard curves that were acquired for both GDP and GDP-Man.

Chemical and Enzymatic Analyses

Glycan products of recombinant enzyme reactions as well as native mannogen were desalted and subjected to methylation linkage analysis as previously described (McConville et al., 1990). Neutral glycans were digested with *Canavalia ensiformis* (jack bean) α -mannosidase or *Helix pomatia* (snail) β -mannosidase (both Sigma-Aldrich) in 0.1 M sodium acetate, pH 5.0 (2 Units/20 ml) or in 0.1 M citric acid phosphate buffer, pH 4.5 (0.5 Units/20 ml), respectively. Reactions were stopped by incubation at 100°C for 2 min, then desalted over a small column of AG 50W-X12(H⁺) over AG 4-X4(OH⁻) (Bio-Rad) (as described above) prior to HPTLC analysis.

Macrophage Infection

BALB/c bone marrow derived macrophages were seeded on 10 mm glass coverslips (1 \times 10⁵ macrophages/well in 24-well plates) and incubated in RPMI containing 15% FBS and 10% L929-cell conditioned medium at 33°C, 5% CO₂ for 24 h prior to infection with *L. mexicana* wild type and mutant promastigotes (1 \times 10⁶ cells/well) to give a final MOI of 1:10. Non-internalized parasites were removed after 4 h by washing the monolayer three times with PBS, and infected macrophages incubated as described above. Coverslips were washed with PBS (x3) at different time points after infection, fixed in 4% paraformaldehyde in PBS and stained with Hoechst 33342 (20 mg/mL, Molecular Probes). Slides were mounted in MoWiol and images acquired using a Zeiss Axioplan2 imaging microscope, equipped with AxioCam MRm camera and the AXIOVISION 4.8 software (Zeiss) (Naderer et al., 2010)

Mouse Infections

BALB/c mice (female, 6-8 weeks old) were injected subcutaneously in the hind-rump and lesion size scored weekly as previously described (Naderer et al., 2008). Parasite burden was also monitored at time of death by removal of the right draining inguinal lymph-nodes and limiting dilution cloning of homogenized tissues (Oliveira et al., 2012). For *in vivo* labeling studies, infected mice were injected intra-peritoneal with ²H₂O (99%, 35 ml/g body weight), then provided with 9% ²H₂O in the drinking water to maintain a level of labeling of 5% in the body (Kloehn et al., 2015). Levels of ²H₂O in the body water was regularly checked by GC-MS analysis of blood samples (Kloehn et al., 2015). Mice were culled at indicated time after initiation of labeling and lesion amastigotes isolated from granulomatous lesions as previously described (Kloehn et al., 2015). Purified parasites were washed three times with ice cold PBS, then extracted in hot water (100°C, 20 min). Extracted mannogen was desalted on a small ion exchange column and ²H-labeling in incorporated mannose residues measured by GC-MS.

Microscopy

For fluorescence microscopy, live *Leishmania* promastigotes were incubated with propidium iodide (final 20 mg/mL) for 5 min before being immobilized on L-lysine (Sigma) coated coverslips. Images were acquired by using a Zeiss Axioplan2 imaging microscope, equipped with AxioCam MRm camera and the AXIOVISION 4.8 software (Zeiss). For transmission microscopy, parasites were fixed in 0.1 M sodium cacodylate containing 2.5% glutaraldehyde overnight at 4°C and further processed using the ROTO technique (Hanssen et al., 2013). Sections (70 nm) were visualized on a Tecnai F30 (FEI, Eindhoven) and images captured on a Ceta CMOS 4k x 4k camera (FEI, Eindhoven).

Protein Production for Crystallization and SECMLS

Genes encoding N-terminally His-tagged variants of MTP1, 2, and 4 were expressed in *E. coli* BL21 (DE3) cells. An overnight starter culture was used to inoculate (at 1% v/v) liter-scale TB expression media containing 35 mg/l of kanamycin. The cultures were shaken in 2 L plastic flasks at 37°C at 220 rpm until the OD₆₀₀ was above 1.5. Gene expression was induced by adding a sterile-filtered solution of IPTG to the culture up to a concentration of 1 mM. The temperature was subsequently reduced to 16°C and the cultures were shaken overnight at 220 rpm. The cells were harvested by centrifugation at > 5000 x g for 20 min, after which a pellet was formed. The pellet was stored at –80°C until further use. Pellet thawing was done by adding 10 x w/v of 50 mM sodium phosphate pH 6.2, 200 mM NaCl, 1 mM DTT, 30 mM imidazole buffer and gently mixing at RT until fully thawed. Subsequently, the volume was adjusted to 100 mL with the same buffer and the cells were lysed by sonication. The resulting solution was centrifuged at > 30,000 x g for 30 min; only the supernatant was collected for further use and loaded onto a 5 mL nickel affinity column (HisTrap FF (GE)). After washing with the abovementioned buffer, the protein was eluted in an imidazole gradient (up to 400 mM). Any precipitated protein was removed by centrifugation and the solution was concentrated to 1-2 ml. The concentrated sample was loaded onto a size-exclusion column (Superdex S75 or S200 (GE)) pre-equilibrated in 50 mM sodium phosphate pH 6.2, 200 mM NaCl, 1 mM DTT buffer. Protein-bearing fractions were collected and the protein was concentrated up to 12 mg/mL (MTP1), 60 mg/mL (MTP2) and 18 mg/mL (MTP4).

Crystallization of MTP1, MTP2 and MTP4

All proteins were crystallized using either sitting drop or hanging drop methods. In the former, the droplet sits on a plastic plate surface and the plates sealed using adhesive film. In the latter, the droplet hangs above the reservoir solution on a siliconized coverslip and the well is sealed with vacuum grease. MTP1 crystals were obtained in 12% w/v PEG 20K, 100 mM MES pH 6.5. The droplet contained 50 nL crystal seed stock solution obtained from previous experiments, 250 nL 5.5 mg/mL protein solution and 200 nL reservoir solution. MTP2 and MTP2 D94N variant crystals were obtained in either 200 mM trisodium citrate, 18%–22% w/v PEG 3350 or in 200 mM triammonium citrate pH 7.0, 18%–22% PEG 3350. The droplet contained a 1:1 ratio of protein (12 mg/mL) to reservoir solution (a total volume of 1 ml for sitting drop and 2 ml for hanging drop). MTP4 was crystallized in 100 mM PCTP pH 7.0, 25% w/v PEG 1500. The droplet contained 150 nL 7 mg/mL protein solution and 150 nL reservoir solution.

X-ray Data Collection and Crystal Structure Solution

Diffraction data were collected at Diamond Light Source synchrotron at beamlines I03, I04 and I24. Crystallographic data were indexed and integrated using Diamond Light Source autoprocessing pipelines that incorporate DIALS software into Xia2 (Winter, 2010). Data truncation, merging and scaling was performed in Aimless. The structures of MTP2 and 4 were solved using molecular replacement (MR) in Phaser (McCoy et al., 2007) with the initial model being 2B4W (PDB code) truncated in Chainsaw (Stein, 2008). The structure of MTP1 was solved using MR in Phaser as well, and the starting model was the polypeptide chain of MTP2. Subsequently, the structures were refined in Refmac5 (Murshudov et al., 2011) and real-space refinement was done using Coot (Emsley et al., 2010). In case of isomorphous crystals in complex with different ligands, the HKL index of the complex was matched to the first obtained solution, the R_{free} set was copied and a model containing the protein only was directly refined to the observed data using Refmac5. Waters were added after the refinement of the polypeptide chain was complete and the ligand molecules were added in the last steps of the refinement. The model geometry and the correspondence of the model to experimental data were validated using Coot validation tools and the PDB validation pipeline. Sugar geometry was validated using Privateer. The process of model building, refinement and validation was performed using the CCP4i2 GUI (Potterton et al., 2018). Data and refinement statistics are shown in Table S1.

SECMALS

Size Exclusion Chromatography Multiangle Light Scattering (SECMALS) was used to investigate the protein assembly molecular weight in solution. The pump used was a Shimadzu HPLC system (CBM-20A Controller, LC-20AD Pump with degasser, SIL-20A Autosampler and SPD-20A detector) and Wyatt detectors: HELEOS-II (light scattering) and Optilab rEx (refractive index). The proteins were kept in 50 mM sodium phosphate pH 6.2, 200 mM NaCl, 1 mM DTT buffer at 3 mg/mL. A standard solution of bovine serum albumin (BSA) was used to calibrate the instrument, as its exact molecular weight and amount used was known. The flow through a Superdex S75 column was kept at a constant 0.5 mL/min. Data were analyzed using ASTRA software (Wyatt Technology).

Phylogenetic Analysis

Sequences of the seven MTP proteins were queried against the NCBI non-redundant sequence database (NR) and TritypDB release 40 using *blastp* and hits filtered based on e-value ($< 1e^{-50}$). Multiple sequence alignments for 21 NR and 148 TritypDB sequences were generated using MAFFT v7.310 (Katoh and Standley, 2013). Phylogenetic trees were generated based on the multiple sequence alignments with PhyML, with initial tree generation done with BIONJ and the LG substitution model employed with branch support with bootstrap analyses (100). Tree diagrams were generated with Evolview (He et al., 2016) and annotated based on taxonomy mapped using the ETE3 Toolkit (Huerta-Cepas et al., 2016).

Co-evolution analysis was performed using the MISTIC platform (Simonetti et al., 2013), which uses Mutual Information (MI) to infer coevolution between residue pairs, using a MI Z-score threshold of > 6.5 , leading to a specificity of 95%. Residue pairings with strong coevolution signals were assessed, and their interactions analyzed using Arpeggio (Jubb et al., 2017). Sequence conservation of protein active site and co-evolving residues was visualized using WebLogo 3 (Crooks et al., 2004).

Homology Modeling and Molecular Docking

Models for *L. mexicana* MTP3, MTP5, MTP6 and MTP7 were constructed using the crystal structure of *L. mexicana* MTP4 as the template. The amino acid sequence identity of these MTPs with *L. mexicana* MTP4 is: MTP3 = 57%, MTP5 = 53%, MTP6 = 60% and MTP7 = 68% (ClustalW as implemented in BioEdit v7.2.5). The four *L. mexicana* MTP homology models were generated using Modeler v9.20 (Sali et al., 1995). Steric clashes between amino acid side chains that may have arisen during model construction were removed by replacing the clashing side chains with an alternative low energy side chain conformation from a conformer library within the Biopolymer module of the modeling program SYBYL-X 2.1.1 (Certara; <https://www.certara.com/>). Each model was then geometry optimized for 1000 iterations (or until the gradient of successive iterations was < 0.05 kcal/mol, Å) using the molecular mechanics MMFF94s force field and partial atomic charges and conjugate gradient minimization method (all other parameters were at default values) within SYBYL-X 2.1.1. The resulting four *L. mexicana* MTP models were deemed to be good quality models using PROCHECK, with $> 99.6\%$ of residues in the allowed regions of the Ramachandran plot (Laskowski et al., 1996).

The amino acid sequence identity between the seven *L. mexicana* MTPs and GH130 family members is very low (5%–12%, ClustalW). To compare the 3-D structures, the four *L. mexicana* MTP models, along with the crystal structures of *L. mexicana* MTP1, MTP2 and MTP4, *L. major* MTP4 and members of the GH130 family, were aligned via the C α atoms of the five-bladed

β -propeller structural motif using the structural alignment program MUSTANG v3.2.3 (Konagurthu et al., 2006). The structural alignment was used to compare the amino acids located in the Pi site, the -1 subsite and the +1 subsite (Table S2).

Man1P was manually docked into its putative binding site in the *L. mexicana* MTP4 crystal structure, using its location in the aligned GH130 family crystal complexes as a guide. GDP-Man and Man₈ were manually docked into their putative binding sites in the *L. mexicana* MTP1, MTP2 and MTP4 crystal structures so that the α -mannose portion of the molecules were in a similar location to that observed in the β -1,2-mannobiose:MTP2 (D94N mutant) crystal complex. Rotatable ligand torsion angles were adjusted manually as required to optimize the ligand interaction with nearby amino acid side chains. All manual manipulations were conducted using SYBYL-X 2.1.1. Each ligand, and any residues within a 6 Å radius of the ligand in the protein complex, was then geometrically optimized within SYBYL-X 2.1.1 using the same optimization protocol described above.

Models of eleven bacterial GT108 proteins identified in the evolutionary phylogenetic analysis of the MTPs (Figure S2; Table S3) and the nine characterized GT91 family members listed in the CAZy database (accessed on the 20th December 2018, <http://www.cazy.org>; Figure S5; Table S3) were constructed in an unbiased manner using the RaptorX structure prediction web server (<http://raptorx.uchicago.edu/StructurePrediction/predict/>) (Peng and Xu, 2011). For the bacterial GT108 proteins (Table S3), the optimal structural template selected by the RaptorX algorithm was *L. major* MTP4 (PDB: 2B4W). The GT91 proteins are predicted to be single pass type II membrane proteins comprising a short cytoplasmic domain, a helical transmembrane domain and a large extracellular domain. For eight of the characterized GT91 family members (BMT1_CANAL, BMT2_CANAL, BMT3_CANAL, BMT4_CANAL, BMT5_CANAL, BMT6_CANAL, BMT2_KOMPG and BMT3_KOMPG; Table S3), the largest region within the extracellular domain was predicted to be a five-bladed β -propeller. GT91 protein models were constructed using single, and also multiple, structural templates. The crystal structure of the GH130_1 family member MGP (PDB: 3WAS) was consistently selected as the single structural template for the highest ranked GT91 protein models, whereas lower ranked models used the crystal structures of Lin0857, RaMP1, RaMP2, BT3780 and *L. major* MTP4 (PDBs: 5B0P, 5YA9, 5AYD, 5A7V and 2B4W, respectively; Figures S2 and S5; Tables S2 and S3). The GT91 protein models constructed using multiple templates (all five-bladed β -propeller GH130 proteins) often had one slightly distorted or displaced blade, therefore structural comparisons with the *L. mexicana* MTPs were carried out using the GT91 protein models constructed from only the MGP template. Interestingly, one of the characterized GT91 proteins, BMT1_KOMPG, was predicted to be a six-bladed β -propeller and was excluded from the structural comparison with the *L. mexicana* MTPs. The PyMOL Molecular Graphics System, Version 2.1.0 (Schrodinger; <https://pymol.org/2/>), was used to visualize and analyze the protein structures and generate figure images.

Multiple Sequence Alignments and Phylogenetic Tree Building for the GT108 MTPs, and GH130 Proteins

The amino acid sequences of the seven *L. mexicana* MTPs and the homolog *L. major* MTP4 were aligned using ClustalU (as implemented in the MPI Bioinformatics Toolkit, <https://toolkit.tuebingen.mpg.de/>) (Zimmermann et al., 2018). The resulting multiple sequence alignment was then used as input to ESPript3.0 (<http://esript.ibcp.fr/ESPript/ESPript/index.php>) (Robert and Gouet, 2014; Edgar, 2004) to generate an annotated alignment (Figure S3).

For the phylogenetic analysis of the seven *L. mexicana* MTPs, ten bacterial GT108 proteins and eighteen characterized GH130 proteins, the amino acid sequences were aligned using MAFFT v7.0 (Katoh and Standley, 2013). The resulting multiple sequence alignment was used as input to ETE3 v3.0.0b32 (as implemented at GenomeNet, <https://www.genome.jp/tools-bin/ete>) (Huerta-Cepas et al., 2016) to construct the phylogenetic trees. The ETE3 methodology used in the tree construction was PhyML v20160115 with model JTT (parameters: -f m-pinv e -o tlr-nclasses 4-bootstrap 100-alpha e) (Guindon et al., 2010). The branch supports were computed out of 100 bootstrapped trees. TreeDyn v198.3 (<http://www.treedyn.org>) as implemented on the Méthodes et Algorithmes pour la Bio-informatique LIRMM web server (http://www.phylogeny.fr/one_task.cgi?task_type=treedyn) (Dereeper et al., 2008) was used to draw the unrooted tree (Figure S2A).

QUANTIFICATION AND STATISTICAL ANALYSIS

All values are the mean \pm standard error using Excel and GraphPad Prism. All experiments were repeated at least twice and numbers of replicates, including the value of n in each experiment are indicated in the figure legends.

DATA AND CODE AVAILABILITY

The PDB accession codes for unliganded MTP1, unliganded MTP2, MTP2 in complex with mannose, MTP2 in complex with β -1,2-mannobiose and MTP4 in complex with phosphate anion, are PDB: 6Q4W, PDB: 6Q4X, PDB: 6Q4Y, PDB: 6Q4Z and PDB: 6Q50, respectively.

Molecular mechanisms regulating O-GlcNAc processing enzymes

Dustin T. King^{1,2}, Alexandra Males³, Gideon J. Davies³, David J. Vocadlo^{1,2}

Addresses

¹ Department of Chemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

² Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

³ Department of Chemistry, University of York, York YO10 5DD,

England. Corresponding author: David J. Vocadlo (dvocadlo@sfu.ca)

Abstract

The post-translational modification of proteins by O-linked N-acetylglucosamine (O-GlcNAc) dynamically programs cellular physiology to maintain homeostasis and tailor biochemical pathways to meet context-dependant cellular needs. Despite diverse roles for O-GlcNAc, only two enzymes act antagonistically to govern its cycling; O-GlcNAc transferase (OGT) installs the monosaccharide on target proteins and O-GlcNAc hydrolase (OGA) removes it. Recent literature has exposed a network of mechanisms regulating these two enzymes to choreograph global, and target-specific, O-GlcNAc cycling in response to cellular stress and nutrient availability. Herein, we amalgamate these emerging mechanisms from a structural and molecular perspective to explore how the cell exerts fine control to regulate O-GlcNAcylation of diverse proteins in a selective fashion.

Introduction

The covalent attachment of β -linked *N*-acetylglucosamine to the hydroxyl group of protein serine and threonine residues has been identified on hundreds of nucleocytoplasmic and mitochondrial proteins, and is abundant throughout all multicellular eukaryotes [1]. The donor substrate for OGT catalyzed O-GlcNAcylation is uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc), the biosynthetic product of the nutrient responsive hexosamine biosynthetic pathway (HBP). Protein O-GlcNAc levels are thus tied to nutrient status through HBP flux, facilitated by the reversible nature of O-GlcNAc that can be installed and removed multiple times during the lifespan of a protein [1]. O-GlcNAc is being found to modulate proteins in an increasing variety of ways including manipulation of: protein function, protein-protein interactions, protein localization, and protein stability (reviewed in [1]). Additionally, O-GlcNAc has the potential to compete directly and indirectly with phosphorylation sites and may agonize or antagonize other post-translational modifications (PTMs), including ubiquitinylation, sumoylation, and acetylation (reviewed in [1]), although relatively few such interactions have been demonstrated. Nevertheless, taken together, dynamic cycling and interplay with other PTMs make O-GlcNAc ideally suited as a nutrient and stress responsive molecular switch.

Various noxious stimuli including heat shock, nutrient depletion, ER dysfunction, and redox imbalance cause fluctuations in global O-GlcNAc levels to help combat these stressors. Aberrant O-GlcNAc cycling has been linked to chronic human diseases, including neurodegeneration, cancer, diabetes mellitus, and X-linked intellectual disability (XLID) [2,3]. The diverse physiological roles of O-GlcNAc and its emerging importance in human health underscore the importance of uncovering the complex molecular mechanisms orchestrating its installation and removal on protein targets, both in normal physiology and pathological states.

Herein, we summarize the biochemical underpinnings of the O-GlcNAc modification and integrate salient concepts pertinent to its regulation. Notably, the O-GlcNAc enzymes have evolved elaborate multi-domain structures that are ideally suited to enable the integrated fine-tuning of their function by diverse regulatory mechanisms. Accordingly, we start with an overview of their structures and then consolidate select recent literature illuminating key regulatory mechanisms that define the abundance, localization, and specificity of the O-GlcNAc cycling enzymes. We focus on emerging interconnected regulatory processes including: subcellular localization, PTMs, protein-protein interactions, metabolic control, and transcriptional and post-transcriptional regulation (Figure 1). Throughout, we highlight current paradigms and raise key unanswered questions. Due to space constraints we regrettably make various omissions to focus on the selected literature most relevant to this topic.

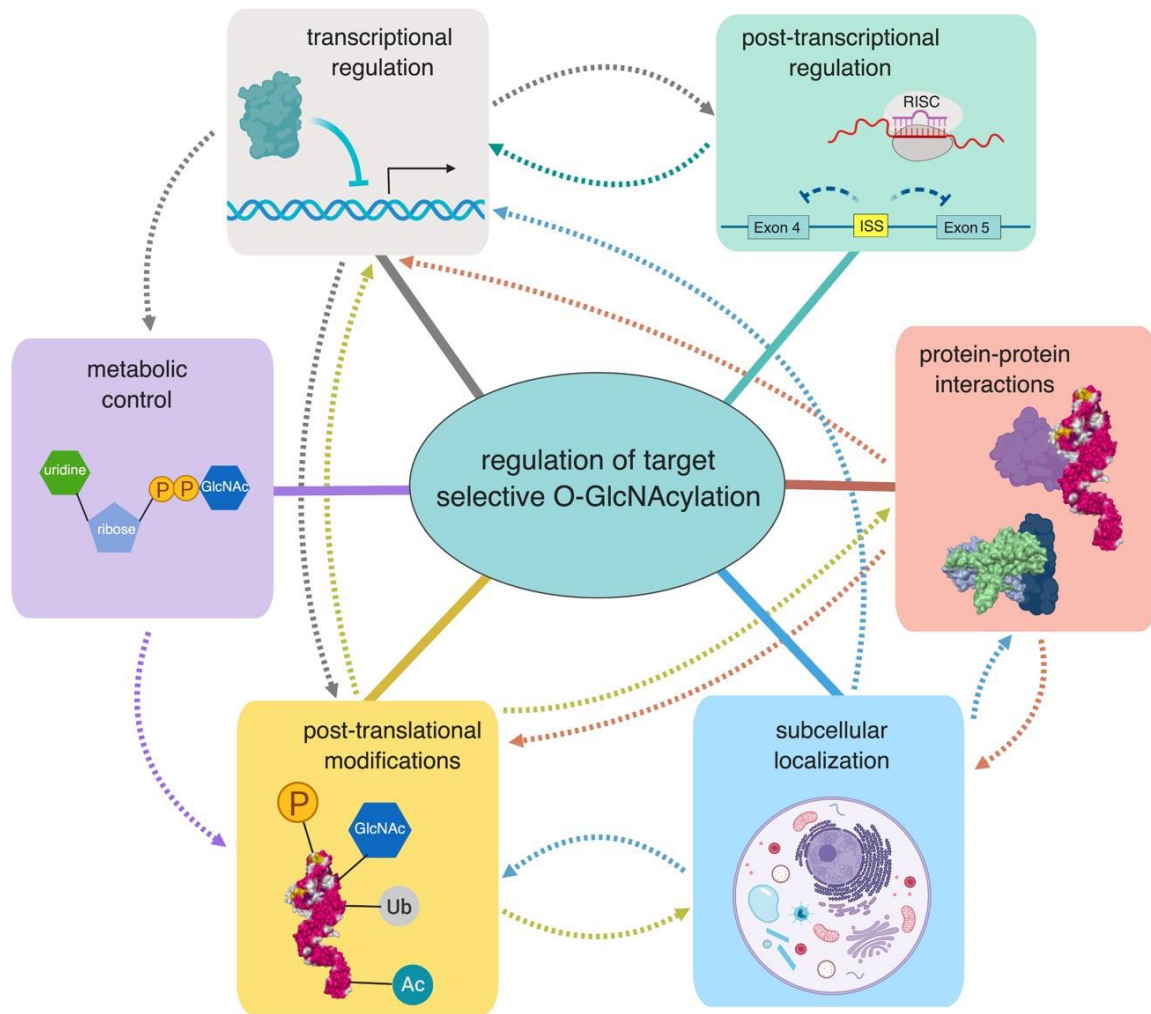


Figure 1. Concept map depicting the interconnected mechanisms regulating target selective O-GlcNAcylation by OGA and OGT. Each dotted arrow represents higher-level crosstalk between the various regulatory mechanisms.

OGT structure and polypeptide acceptor specificity

OGT encodes a metal-independent CAZY (www.cazy.org) family GT41 β -*N*-acetylglucosaminyltransferase belonging to the GT-B superfamily of glycosyltransferases. This essential gene is located on the X-chromosome and encodes a modular protein composed of five distinct regions: i) an N-terminal region consisting of an isoform specific number of contiguous helix-turn-helix tetratricopeptide repeats (TPRs), ii) a transitional (H3) helix that sits between the N- and C-terminal halves, iii) a C-terminal region containing the GT41 catalytic domain with two catalytic lobes (Cat1 and Cat2) that each have a Rossmann fold, iv) an intervening domain with a topologically novel fold comprised of a seven stranded β -sheet core stabilized by flanking α - helices (IntD), and v) a putative C-

terminal phosphatidylinositol-binding region (PPO) (Figure 2A). There are three different isoforms of OGT expressed in human cells that vary in the number of N-terminal TPRs. Nucleocytoplasmic OGT (ncOGT) has 13.5 TPR repeats, mitochondrial OGT (mOGT) has 9 repeats, and short OGT (sOGT) has 2.5 repeats (Figure 2A). These isoforms have distinct subcellular localization patterns and activities (see localization section below).

In 2011, the seminal crystallographic structure of human OGT was determined illuminating its unique atomic architecture. The structure comprises 4.5 N-terminal TPR repeats and the full C-terminal catalytic region in a ternary complex with UDP and a CKII peptide substrate (hereafter called hOGT_{4.5}) [4]. The TPRs in this structure (aa 313-400) overlap with an earlier OGT crystal structure consisting of 11.5 TPR repeats, and alignment of the two structures provides a putative model of the full-length protein (Figure 2B) [4]. The TPR superhelix consists of two layers of helices, an inner concave face formed by helix-A and an outer convex face formed by helix-B. Notably, helix-A has a highly conserved asparagine ladder that runs the length of the TPR domains that is involved in binding to peptide acceptors (see below) [4]. The C-terminal end of the TPR superhelix packs against the catalytic domain carving out a narrow acceptor peptide binding groove that feeds into the catalytic centre found at the interface of Cat1 and Cat2 [4]. The Int-D domain packs against the outer face of the Cat2 lobe oriented away from the active site core. The function of the Int-D domain remains unknown but in conjunction with a Cat2 helix, and the C-terminal PPO, it forms a large electropositive surface comprising ten lysine residues that may be involved in ligand binding (see below) [4].

As an inverting glycosyltransferase, OGT catalyzes the formation of a β -glycosidic linkage between GlcNAc and the acceptor peptide serine. The peptide acceptor binds OGT predominantly through amide backbone hydrogen bonds with the β -hydroxyl group of its glycosylated serine projecting directly into the nucleotide sugar binding site. The peptide acceptor covers the solvent exposed face of the bound UDP-GlcNAc, effectively trapping it within the active site (Figure 3A). This observed binding mode in combination with kinetic data support the broadly accepted proposal that OGT catalyzes glycosyltransfer via an ordered bi-bi mechanism whereby UDP-GlcNAc must bind the active site first followed by the acceptor peptide [4]. Intriguingly, fundamental mechanistic details remain unknown, including the identity of general base responsible for deprotonating the catalytic serine [4]. Briefly as reviewed [4], three alternative general bases were proposed including His498, Asp554, or the α -phosphate of UDP-GlcNAc. Although mutation of His498 results in inactive enzyme, the side chain interactions suggest it is protonated at neutral pH making it poorly suited to serve as a general base. Direct transfer of a proton to α -phosphate of UDP-GlcNAc appears energetically unfavorable. Structural studies, however, revealed a series of ordered water molecules in the active site acting as a potential proton shuttle adjacent to Asp554 and functional studies reveal defective sugar transfer when this residue was mutated to asparagine, offering the best support for the identity of the general base to

date.

OGT lacks a strict consensus sequence governing its activity. However, subtle sequence preferences help govern substrate selectivity. The specificity of OGT was assessed using a saturation library of 3,480 unique acceptor peptides displayed in a monolayer. As predicted by earlier work [5], valine and alanine were preferred in the -3 and +2 subsites. Additionally, a preference for aromatic residues in the -4 position was uncovered [6]. Notably, proline was previously shown to be preferred at the -2 subsite [5]. In contrast, proline at the +1 position has been found to prevent O-GlcNAcylation within certain peptides [7]. OGT predominantly glycosylates serine and threonine residues located within disordered regions on protein targets but also appears able to glycosylate residues positioned in a variety of different secondary structures [8]. However, these secondary structures may need to be unfolded to be accommodated within the OGT active site. Although not yet characterized, such local unfolding may be triggered in response to various stimuli, or potentially occur during the co-translational O-GlcNAcylation of some nascent chains [9], prior to protein folding. The presence of other PTMs can also influence protein targeting as exemplified by the observation that acceptor peptide phosphorylation can reduce OGT catalyzed O-GlcNAcylation *in vitro* [7]. However, elucidating the physiological significance of this proposed PTM code awaits quantitative analysis of levels on endogenous proteins.

Emerging evidence shows that peptide binding occurs not just in the GT41 active site but also within the TPR domains. A protein microarray in conjunction with UDP-GlcNAz was used to establish OGT targeting of proteins using downstream click-chemistry to incorporate biotin [10]. This method was used to assess the catalytic contribution of a poly-asparagine ladder located within the TPR domains, known to engage in numerous bidentate hydrogen bonds with the peptide backbone of Tab1 [11] (Figure 3A). An OGT mutant construct consisting of alanine in place of these 5 asparagine residues (5N5A) retained activity toward short peptides but impaired modification of proteins [10], supporting the importance of these residues in indiscriminate recognition of the polypeptide backbone of proteins distant from the site of glycosylation. In a different approach, the Jiang group developed a GlcNAc electrophilic probe (GEP) that consists of a UDP-GlcNAc analog containing an allyl chloride electrophile extending as an N-acyl group from the 2-amino position [12]. This GEP was used to discern whether mutations in OGT affected sugar binding or protein acceptor binding and glycosylation by examining its relative partitioning to OGT (C917) or the acceptor protein. This method yielded results further supporting an important role for the asparagine ladder in acceptor protein binding [12]. Recently, the XLID OGT mutation, L254F, within the seventh TPR was found to distort the TPR superhelix resulting in reduced thermal stability and catalytic activity [13]. Furthermore, five additional mutations mapping to the OGT TPR domains have recently been correlated with XLID [13-15]. Very recently, two aspartate residues (D386 and D420) located in the OGT TPR lumen distal from the active site were shown to be important for substrate selection. Interestingly, mutation of these residues to alanine altered substrate specificity and increased rates of

protein glycosylation *in vitro*. In particular, a preference was found for substrates with a lysine or arginine positioned 7-11 residues away from the site of glycosylation [16]. Taken together, the OGT TPR domains are emerging as essential mediators of substrate binding, enzyme activity, and stability.

OGA structure and target specificity

OGA encodes two distinct splice variants referred to as long (OGA-L) and short (OGA-S) isoforms. Human OGA-L (hOGA) consists of three domains connected by disordered regions: i) an N-terminal *N*-acetyl- β -D-glucosaminidase catalytic domain (CAZy GH84) with a $[(\beta/\alpha)_8]$ triose-phosphate isomerase (TIM) barrel structure, ii) a helical bundle referred to as the stalk domain composed of four α -helices, and iii) a C-terminal pseudo histone acetyl-transferase (HAT) domain with a GCN5 acetyltransferase-like fold (Figure 2C). The pseudo-HAT domain lacks the conserved P-loop motif required for binding of the cofactor, acetyl-CoA, implying that it likely has a non-enzymatic function. OGA-S lacks the pseudo-HAT domain and part of the stalk domain, leading to greatly reduced enzyme activity.

Recently, three groups independently solved the long-anticipated crystal structure of hOGA. This was achieved by the systematic design of various constructs lacking disordered regions [17-19]. The homodimer is stabilized by a domain swapped helix (α 17: aa 676-694) [18,19]. Notably, a construct lacking this α 17 helix (residues 1-662) does not dimerize and displays 100-fold lower catalytic efficiency as compared to WT OGA in processing *p*NP-GlcNAc suggesting that dimerization is required for maximal activity [19].

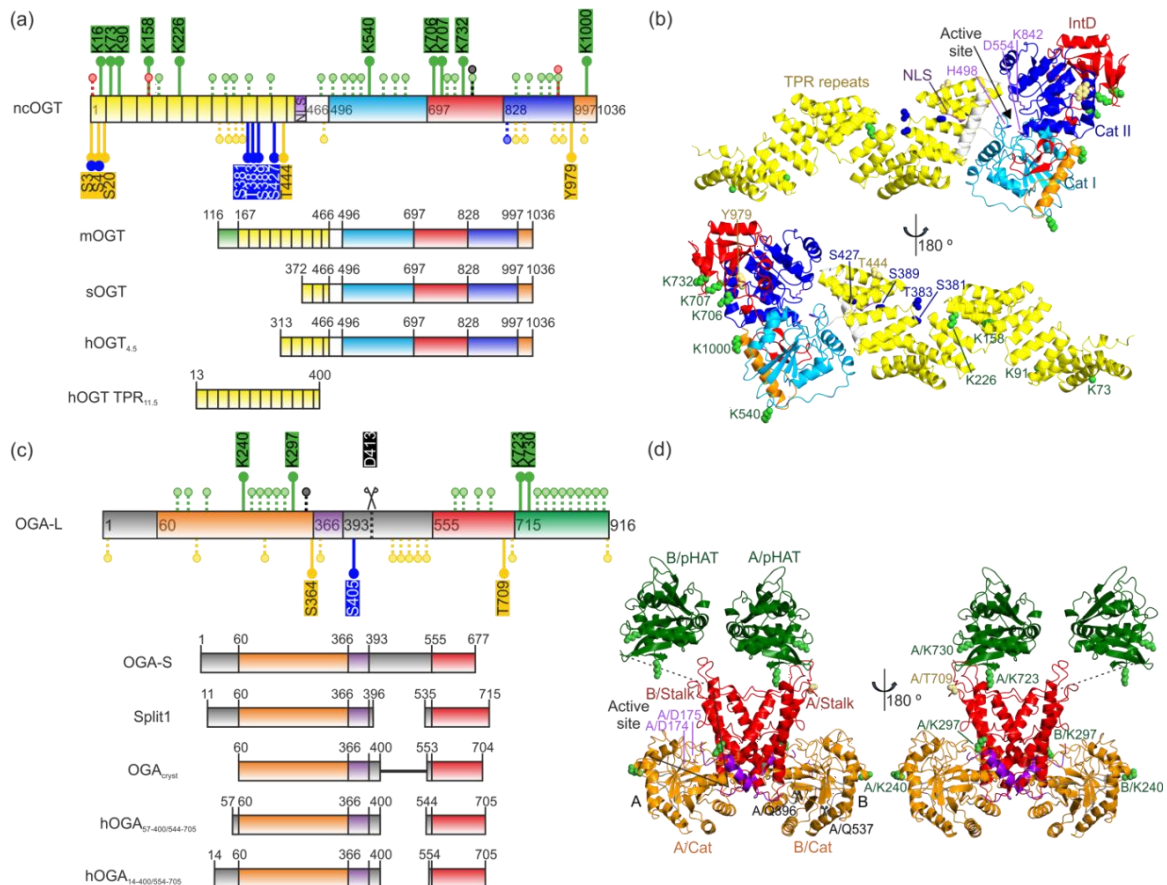


Figure 2. Domain structure and isoforms of OGT and OGA. A) Schematic of hOGT showing splice variants and constructs used for crystallography. OGT is found as three isoforms distinguished by the number of TPRs. The TPRs are shown in yellow, the NLS (D451-P453) is coloured purple, transitional helix (H3) is white, the glycosyltransferase catalytic domains are illustrated in light blue and dark blue, the Int-D (intervening domain) is red, and the phosphoinositide binding domain of OGT (PPO) is coloured orange with sequence numbered according to UniProtKB/SwissProt accession (P56558.1). In panels A and C, putative PTMs compiled using PhosphoSitePlus [20] are represented on ncOGT and OGA-L as follows: phosphorylation (yellow), O-GlcNAcylation (blue), ubiquitinylation (green), acetylation (red), and sumoylation (black). High confidence PTM sites with 5 or more references or experimentally determined by low-throughput mass spectrometry mapping methods are shown with solid lines and bold colours. Lower confidence sites with fewer than 5 references and determined by high-throughput methods are shown with dotted lines and light colours. B) Overall crystallographic model of ncOGT. The model was built by superposing hOGT_{4.5} (PDB ID: 3PE3) and hOGT TPR_{11.5} (1W3B). High confidence PTMs are labelled, as well as domains and key active site catalytic residues. C) Schematic of human OGA splice variants as well as constructs used for crystallography. Disordered regions are shown in black, the N-terminal GH84 domain is illustrated in orange, the helical bundle domain is purple, the C-terminal stalk domain is shown in red and the “pseudo HAT” domain is represented as a green cartoon. The caspase-3 cleavage site is indicated by scissors. D) Overall crystallographic model of hOGA. The hOGA Split1 crystal structure (PDB ID: 6HKI) is shown attached at the C-terminal end (Y715) to a homology model of the OGA pseudo-HAT domain constructed using Phyre2 from the *Oceanicola granulosus* homolog (PDB ID: 3ZJ0, 29% sequence identity). Structures in B and D are colour coded as in A and C with high confidence PTMs labelled on the structures.

A V-shaped active site cleft (~22 x 25 Å with an angle of ~70 °) is formed at the interface of the catalytic domain of monomer A and the stalk domain of monomer B [18] (Figure 2D). Multiple hOGA-glycopeptide bound structures reveal peptides bind within the cleft in either orientation [18,21,22] (Figure 3B-C). Two conserved aspartate residues flank the glycosidic bond, acting as general acid/base catalysts to facilitate a two-step catalytic mechanism involving a transient oxazoline intermediate [4]. Kinetic analysis of OGA mediated O-GlcNAc hydrolysis on intact proteins revealed that OGA indiscriminately cleaves diverse O-GlcNAc sites across different targets with roughly equivalent catalytic efficiency [23], an observation that is supported by the enzyme cleaving O-GlcNAc residues incorporated at “unnatural” sites [24]. This promiscuous substrate selectivity is likely in-part due to the predominant interaction of OGA with amide backbone atoms within the bound glycopeptide. As seen elsewhere [23], the rate of OGA catalyzed hydrolysis of an O-GlcNAcylated serine was shown to be reduced by half in the presence of an adjacent phosphotyrosine on a ZO3-derived model peptide (RESSYDIY_(p)RVPSS_(g)QS) [25]. The ability for PTMs to influence OGA catalyzed removal merits further investigation. In summary, recent research has provided a vital structural foundation for understanding OGA and OGT. The challenge for the future is to build upon this knowledge to uncover how regulatory factors modulate their structure and function to choreograph target specific O-GlcNAcylation.

Localization

The O-GlcNAc cycling enzymes adopt distinct subcellular loci during different phases of the cell cycle, in response to stimuli, and across cell types. The OGT and OGA isoforms (Figure 2) have long been known to display different subcellular localization patterns; ncOGT is found within the nucleus and cytoplasm, sOGT is predominantly localized to the cytosol, and mOGT is thought to be largely mitochondrial – though this is a topic of some debate [26,27]. OGA has two dominant splice variants with OGA-L being nucleocytoplasmic with preferential localization in the cytoplasm and the largely inactive OGA-S residing in the nucleus and on the surface of nascent lipid droplets, where it is thought to regulate proteasome-mediated droplet remodelling (Figure 4A) [1].

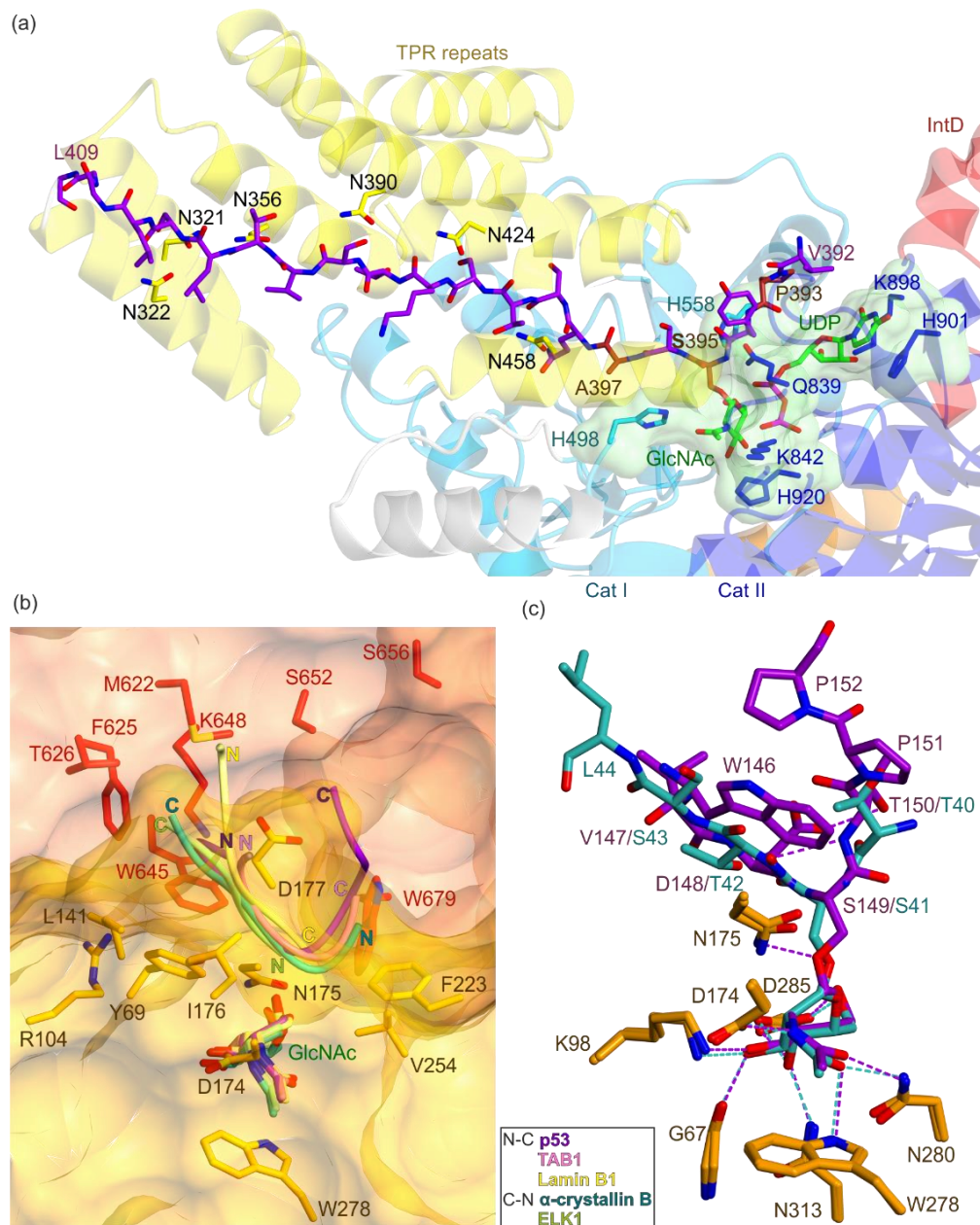


Figure 3. Active-sites of OGT and OGA. A) OGT in complex with a TAB1 peptide. The bound peptide is fused to the N-terminus of OGT and is shown in purple (PDB ID: 5LVV). Asparagine ladder residues shown in yellow are highlighted against the yellow TPRs. The peptide covers the solvent exposed face of UDP-

GlcNAc, shown in green, with the active site highlighted. B) Overlay of OGA in complex with various peptides. The following peptides are displayed: α -crystallin B (PDB ID: 5VVV, dark green, residues TTSLSL), ELK1 (PDB ID: 5VVT, light green, residues TLSP), TAB1 (PDB ID: 5VVU, pink, residues PYSS), Lamin B1 (PDB ID: 5VVX, yellow, residues SSRVT), and p53 (PDB ID: 5UN8, purple, residues WVDSTPP). The peptides are shown as worms with the N- and C-termini labelled. The catalytic and neighbouring residues are highlighted in the active site in orange. C) Superposition of peptide bound OGA complexes. The bound peptides α -crystallin B (PDB ID: 5VVV) and p53 are shown in cyan and purple. Key hydrogen bond interactions are coloured according to peptide.

Excitingly, the molecular mechanisms regulating these isoform specific localization patterns are beginning to emerge. Recently, the ncOGT nuclear localization signal (DFP; residues 451-453), which is required for interaction with importin α 5 and subsequent translocation through the nuclear pore complex (NPC) (Figure 4A) was characterized [28]. mOGT is proposed, based on sequence analysis, to contain a putative N-terminal mitochondrial targeting sequence. Consistent with this hypothesis, OGT appears highly concentrated in the mitochondrial inner membrane where it interacts with electron transport chain complex IV and is suspected of O- GlcNAcylating various mitochondrial proteins [29]. Curiously, the pyrimidine nucleotide carrier 1 transporter protein (SLC25A33) translocates UDP-GlcNAc from the cytosol to the mitochondrial matrix, potentially facilitating mOGT catalyzed mito-specific glycosyltransfer (Figure 4A) [29]. Isoform specific siRNA knockdown of mOGT leads to Drp1 dependant mitochondrial fragmentation, reduced membrane potential, and loss of mitochondrial content, supporting an important role for OGT in maintaining mitochondrial integrity and function [27].

In addition to these isoform specific localization patterns, overexpression of tagged O- GlcNAc cycling enzymes shows their partitioning to distinct subcellular locations during cell division and in response to signalling events. During mitosis, OGT is concentrated at the central spindle assembly during anaphase, and the midbody during cytokinesis. This is consistent with the formation of transient complex between OGT, OGA, mitotic Aurora kinase B (AURKB), and protein phosphatase 1 (PP1), which regulates vimentin filament glycosylation to control intermediate bridge formation and severing during cytokinesis (Figure 4B) [30]. Recently localization of endogenously expressed O-GlcNAc cycling enzymes as well as global O-GlcNAcylation were investigated in bovine and human oocytes during meiosis by immunocytochemistry and confocal microscopy [31]. O-GlcNAcylated proteins were distributed throughout the nucleus and cytoplasm with enrichment at the nuclear envelope during prophase-I, whereas during metaphase O-GlcNAcylated substrates localized throughout the cytoplasm with slight enrichment at the cortex. OGA was localized at the cortex throughout meiosis, whereas OGT was localized to the meiotic spindles [31]. These results support the idea that localization of the O-GlcNAc cycling enzymes is regulated during cell division. Despite having cell division-dependent localization patterns the mitotic and meiotic roles of the O-GlcNAc cycling enzymes remain poorly understood, though altering levels of these enzymes clearly affects cell division. Upon prolonged insulin stimulation of 3T3-L1 adipocytes, a subset of ncOGT is proposed to be recruited to the plasma membrane by

phosphatidylinositol (3,4,5)-triphosphate (PIP3), to modulate the insulin signalling pathway by O-GlcNAcylation (Figure 4A) [32-34]. OGT was initially proposed to have a PIP3 binding site located in the C-terminal catalytic and Int-D domains, however, this topic remains controversial [4]. We speculate that this type of triggered localization of the O-GlcNAc cycling enzymes may be relatively common to enable temporal control of O-GlcNAc levels at distinct subcellular locations. Yet, the molecular bases for these changes need more focused biochemical studies combined with cell-based experiments.

Finally, OGT displays cell type specific localization patterns. For example, OGT is enriched in the postsynaptic density of excitatory synapses perhaps to enhance the O-GlcNAcylation of multiple proteins as seen to occur upon neuronal stimulation. Postsynaptic KO of OGT decreases synaptic expression of ionotropic transmembrane glutamate receptor AMPA subunits (GluA2 and GluA3), reduces presynaptic terminals, and stunts the development of mature dendritic spines [35]. Future discovery-based investigations to systematically catalogue OGA and OGT localization patterns across diverse cell types and in response to varied stimuli would likely be illuminating, as has been the case for kinases and phosphatases.

Post-translational modifications

The O-GlcNAc enzymes are decorated with numerous PTMs. In fact, a survey of the PhosphoSitePlus database reveals that proteomic discovery-based mass spectrometry efforts have uncovered a total of 55 and 42 distinct PTMs on human OGT and OGA (Figure 2) [20]. For OGT, a handful of these PTMs have been ascribed functions including dynamic fine-tuning of its localization, turnover, and substrate specificity (Table 1). O-GlcNAcylation of OGT Ser389 within TPR12 enables its nuclear localization [28]. Similarly, AMPK mediated phosphorylation of Thr444 contributes to OGT nuclear import in myotubes [36]. We note that the T444 and S389 phosphosites are both located within TPR12 in proximity to the recently characterized NLS (451- 453), and thus may regulate recruitment of the nuclear import machinery. This NLS is present in all OGT isoforms, however, its contribution to nuclear import has only been investigated in ncOGT. Phosphorylation of OGT T444 changes its substrate selectivity *in vitro* and in proliferating HEK293T kidney cells, and alters its binding partners. T444 phosphorylation thus appears to regulate OGT target specificity, localization, and protein-protein interactions – phenomena that are likely not mutually exclusive. Upon glucagon-induced calcium signaling, calcium/calmodulin- dependant kinase II (CamKII) phosphorylates OGT S20, which alters the specificity of OGT to promote O-GlcNAcylation and activation of the Ulk proteins involved in autophagic induction [37]. Similarly, Checkpoint kinase 1 (Chk1) also phosphorylates Ser-20, and this event was separately found to stabilize OGT against proteasomal degradation and be required for its previously mentioned localization to the midbody upon cytokinesis (Figure 4B) [38]. Understanding the roles of OGT phosphorylation may yield insight into its stimulus-induced spatiotemporal regulation. Even less explored are the multiple putative

ubiquitylation sites discovered on OGT in proteome-wide mass spectrometry experiments (PhosphoSitePlus database [20]). For example, the LSD2 histone demethylase acts as an E3 ligase and targets OGT for polyubiquitylation at an unknown site, thereby promoting its proteasomal degradation [39], whereas the BAP1 deubiquitinase protects OGT against UPS mediated degradation [40]. However, the identities of the E1 and E2 enzymes regulating these UPS mediated degradation pathways remain unknown.

PTMs can also directly modulate OGT catalyzed glycosyltransfer. OGT can be phosphorylated at S3 and S4 by glycogen synthase kinase 3 β (GSK3 β) [41]. Interestingly, GlcNAc has also been identified on OGT S3 and S4 suggesting the possibility that O-GlcNAcylation and phosphorylation may antagonize each other at this site [41]. Although an appealing hypothesis, O-GlcNAc and phosphorylation are both most often substoichiometric and, accordingly, such proposed direct reciprocal crosstalk has been difficult to clearly demonstrate as this would require quantitative assessment of PTMs on endogenous proteins. Taken together, the roles for PTMs in controlling OGT localization, turnover, and activity are slowly emerging. The challenge for the future is to confirm these effects, decode the combinatorial influence of these diverse PTMs on OGT function in response to subcellular signalling, and delve deeper into the structural and biochemical basis behind these observations.

OGA is also heavily decorated with PTMs, however very little is understood regarding their functional ramifications. During apoptosis, OGA is cleaved by caspase-3 within a non-canonical S₄₁₀VVD₄₁₃ recognition site [42]. Co-expression of the two caspase-3 cleavage products recapitulates OGA activity in cells, suggesting the two halves remain associated within cells. However, a functional role for this cleavage has not been defined. Studies have reproducibly identified GlcNAc at S405, yet it is not known whether this modification has a functional role [20]. Bearing in mind that individual sites of modification do not necessarily have physiological roles, identifying those which do and exploring how some of these post-translational events influence OGA function will bridge a key gap in our understanding of the regulation of O-GlcNAc.

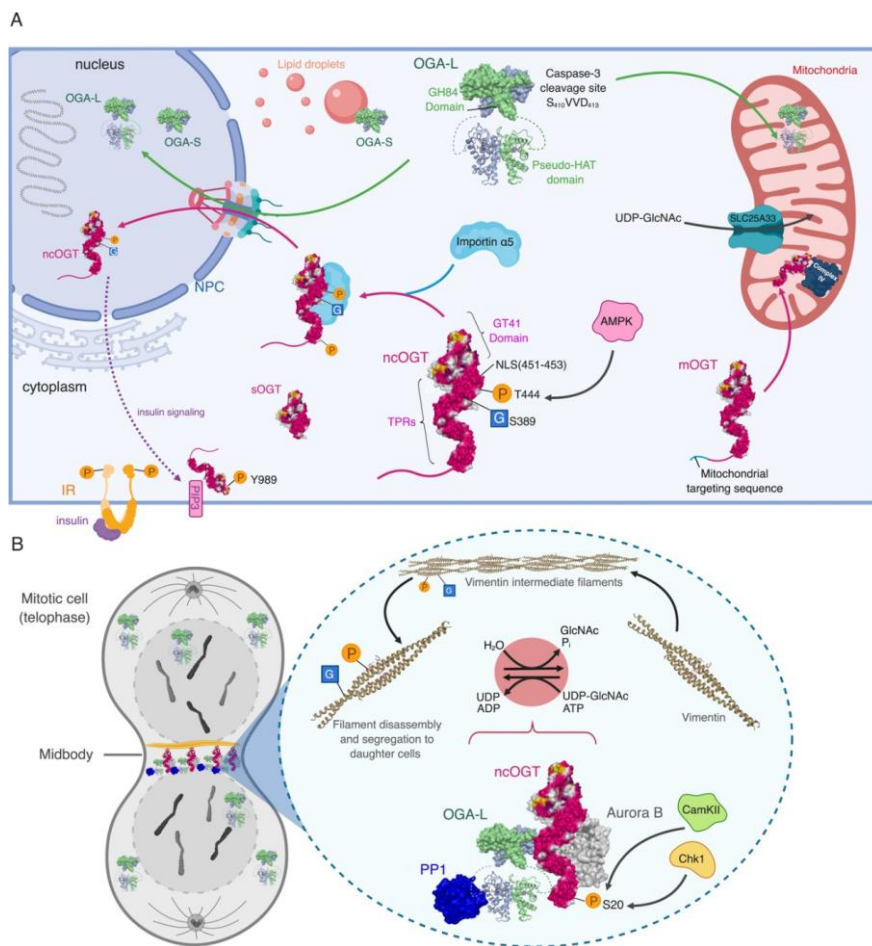


Figure 4: Modulation of OGA and OGT localization by PTMs and protein binding partners. A) Subcellular localization of OGA and OGT. Importin $\alpha 5$ facilitates the nuclear import of ncOGT and this process is dependent upon S389 O-GlcNAcylation and AMPK mediated T444 phosphorylation. The OGT-S isoform is predominantly cytosolic. OGA-L localizes to the cytosol and nucleus, and OGA-S exists in the nucleus and on the surface of lipid droplets. OGA-L and mOGT localize to the mitochondria, and mOGT associates with complex IV. **B)** OGT and OGA localization and complexation during mitosis. OGT is localized almost exclusively to the midbody during mitosis. In contrast, OGA localizes diffusely throughout the cell. OGT, OGA, AURKB, and PP1 are thought to form a complex at the midbody during mitosis. This complex modulates vimentin intermediate filament formation and disassembly, thereby regulating segregation of daughter cells during cytokinesis. We note that structural details regarding the (OGT, OGA, AURKB, PP1) complex remain unresolved, therefore, our representation of this complex in panel B is for illustrative purposes only.

Interaction partners

Interaction partners play an integral role in modulating the activity and localization of the O-GlcNAc cycling enzymes. Here, we focus on select recent examples germane to the regulation of O-GlcNAc. The nuclear receptor REV-ERB α integrates the circadian clock with both glucose and lipid metabolism and has recently been shown to interact with and modulate OGT [43]. REV-ERB α shields cytoplasmic OGT from proteasomal degradation and appears to enhance OGT catalytic activity in the nucleus, thereby increasing O-GlcNAcylation of proteins within both compartments. For example, REV-ERB α is proposed to promote OGT-mediated regulation of insulin signalling by boosting O-GlcNAcylation of AKT in the

cytoplasm [43]. Thus, REV-ERB α exerts pleiotropic circadian activities through direct interaction with OGT.

UDP-N-acetylglucosamine pyrophosphorylase-1 (UAP1) catalyzes the last step in the synthesis of UDP-GlcNAc. A UAP1 paralog, UAP1-like-1 (UAP1L1), is required for OGT mediated protein O-GlcNAcylation in HepG2 liver hepatocellular carcinoma cells [44]. Surprisingly, UAP1L1 does not efficiently catalyze UDP-GlcNAc synthesis, but instead interacts directly with the C-terminal catalytic domain of OGT. Knockdown of UAP1L1 markedly reduced global O-GlcNAcylation in HepG2 and TONG cells, yet UAP1L1 alone is not sufficient to bolster OGT activity *in vitro* [44]. Thus, UAP1L1 likely maintains OGT activity within cells through direct binding, though more detailed characterization is necessary.

5-methylcytosine (5mC) in DNA plays important roles in transcriptional regulation, genomic imprinting, and suppression of transposable elements. The TET1-3 (Ten Eleven Translocation) proteins oxidize 5mC and thereby facilitate reversal of locus specific DNA methylation particularly within promoter regions [45]. Though sometimes contradictory, a range of studies have shown OGT binds to TET proteins and regulates their activity (reviewed in, [46]). Sadly, aside from the work on REV-ERB α [43], this promising area has largely languished since 2015 and a challenge for the future is to better determine the context-dependant effects of these interactions on chromatin structure and gene regulation.

To identify putative OGA binding partners, BioID proximity biotinylation in combination with stable isotope labeling of amino acids in cell culture (SILAC) was used [47]. This analysis revealed 90 proteins displaying stress-dependant interactions with OGA. OGA interactions with fatty acid synthase (FAS), filamin-A, heat shock cognate 70 kDA, and OGT were verified by co-immunoprecipitation. *In vitro* analysis of the pool of OGA bound to FAS revealed it had only ~15% of the activity of free enzyme, and FAS overexpression appeared to boost oxidative stress-induced global O-GlcNAc levels in cells [47]. Overall, the structural and biochemical features governing these protein-protein interactions remain unexplored and limit our understanding of how binding partners functionally regulate OGT and OGA.

Table 1

Key post-translational modifications and processing events on OGT and OGA

Protein (residue)	Modification	Functional role	Enzyme(s)	Ref.
OGT (S3)	Phosphorylation, O-GlcNAc	Increases activity	GSK3 β	[41]
OGT (S4)	Phosphorylation, O-GlcNAc	Increases activity	GSK3 β	[41]
OGT (S20)	Phosphorylation	OGT activity / mitotic localization	Chk1, CamKII	[37,38]
OGT (S389)	O-GlcNAc	Required for nuclear localization	OGT	[28]
OGT (T444)	Phosphorylation	Nuclear uptake and activity	AMPK	[36]
OGT (D451-P453)	Nuclear localization signal	Nuclear uptake	*N.A.	[28]
³² P-OGT (Y989)	Phosphorylation	Increased activity	IR, Src	[32]
OGA (S364)	Phosphorylation	Unknown	Unknown	[20]

OGA (S405)	O-GlcNAc	Unknown	Unknown	[20]
OGA (S410-D413)	Proteolysis site	Unknown	Caspase-3	[42]
*N.A. means “non-applicable”				
‡ Speculative modification proposed based on blots and sequence analysis on rat OGT Y979 (hOGT Y989)				

Metabolic Regulation

The concentration of the UDP-GlcNAc donor substrate used for OGT catalyzed glycosyltransfer is responsive to nutrient flux due to the direct dependency of HBSP on major cellular metabolites. In particular, intracellular concentrations of glucose, glutamine, fatty acids, and GlcNAc modulate HBSP flux and UDP-GlcNAc abundance (for an early perspective see [48]). Notably, HBSP enzyme expression and allosteric modulation are under tight control and respond to various changes in nutrient and metabolite abundance and intracellular signalling (reviewed in [49]). For example, the unfolded protein response (UPR) signal transducer, spliced X box- binding protein 1 (XBP1), acts as a transcriptional activator to induce expression of rate-limiting HBSP enzymes boosting cellular O-GlcNAcylation in response to ER stress [50]. While nutrient supply clearly alters UDP-GlcNAc levels, nutrient flux may also induce intracellular signalling pathways that ultimately modulate the abundance and function of the O-GlcNAc cycling enzymes. Accordingly, while generally accepted in the field, the effects of nutrients on regulation of O-GlcNAc have surprisingly not been systematically investigated to simultaneously monitor UDP-GlcNAc levels, levels of HBSP enzymes, OGA and OGT, and downstream levels of O-GlcNAc. Such studies would help provide a solid foundation to systematically assess the functional and mechanistic basis for the effects of nutrients.

Transcriptional activation and repression

New findings are emerging to support expression of OGA and OGT being coordinately regulated by control of transcriptional programs that work to maintain balanced O-GlcNAc levels. For example, increases in OGT lead to concomitant increases in OGA expression, and vice versa, in what appears to be an effort to buffer the cell from rapid changes in O-GlcNAcylation. Furthermore, treatment of cells with the selective OGA inhibitor ThiametG, boosts O-GlcNAc levels and induces a compensatory increase in OGA gene expression and a correlated decrease in OGT abundance [51], and the converse is also true for OGT inhibitors [52].

We are just beginning to uncover transcriptional programs governing these coordinated shifts in expression of *OGT* and *OGA*. For example, *OGA* serves as a co-activator in conjunction with the histone acetyltransferase p300 to positively regulate transcription factor CCAAT/enhancer-binding protein beta (C/EBP β) – among the most common activators occupying the *OGT* promoter (Figure 5) [53]. This

transcriptional regulatory mechanism enables OGT expression to be tuned to OGA levels to balance O-GlcNAcylation. Additionally, the *OGA* promoter is sensitive to OGT expression as exemplified by a recent analysis of O-GlcNAcylation in XLID patients. An XLID variant L254F-OGT, results in reduced L254F-OGT protein levels in isolated patient lymphoblastoid cells [3]. However, global O-GlcNAc abundance appeared unaltered, presumably due to a compensatory decrease in *OGA* mRNA and downstream *OGA* levels. OGT appears to be involved in a balancing mechanism that regulates global O-GlcNAc levels in XLID through formation of an OGT-mSin3A-HDAC1 repressive complex at the *OGA* proximal promoter region (Figure 5) [3]. Curiously, E2F transcription factor 1 (E2F1) concurrently represses transcription at both *ogt* and *oga* promoters in a retinoblastoma (Rb1)-dependant manner in mice to maintain a balanced mRNA pool (Figure 5) [54]. OGT transcription also appears to be regulated by hepatocyte nuclear factor 1 homologue A (HNF1A), which is itself O-GlcNAc modified. Loss of O-GlcNAc appeared to increase transcription of OGT, suggesting a mechanism for autoregulation of O-GlcNAc homeostasis [55]. Overall, a higher order network of multiple transcriptional mechanisms likely act in concert to maintain O-GlcNAc homeostasis but decoding this will require both more detailed biochemical and system wide studies.

Transcription of O-GlcNAc cycling enzymes is also dysregulated in response to other specific stimuli. For example, OGT levels are markedly increased in human papillomavirus (HPV) induced cervical neoplasms, and transfection of HPV oncogenes into mouse embryonic fibroblasts induces a pronounced increase in OGT mRNA and protein levels [56]. HPV infection alters the expression of several transcription factors with predicted binding sites within the *OGT* promoter. Overexpression of transcription factors AP-1, SP-1, NF- κ B, p65, and c-MYC enhanced OGT promoter activity on a reporter construct [56]. However, further *in vivo* analysis of the relative contribution of these transcription factors to enhancement of endogenous *OGT* expression is required. As an additional example, upon lipopolysaccharide stimulation, E2-related factor – 2 (Nrf2) binds the *ogt* promoter and enhances its transcription in bone-marrow derived mouse macrophages, resulting in amplified O-GlcNAcylation and inhibition of pro-inflammatory transcription factor STAT3. Interestingly, Nrf2 is not required for baseline OGT expression but rather only boosts its abundance in response to stimuli. Furthermore, the E3 ubiquitin ligase CUL3 negatively regulates OGT expression through its targeted degradation of Nrf2 (Figure 5) [57]. Future efforts should aim to uncover additional *cis*- and *trans*-acting factors controlling transcription of *OGT* and *OGA* and illuminate how they are integrated to tailor transcriptional programs in response to various stimuli.

Post-transcriptional regulation

MicroRNAs (miRNAs) are short non-coding RNAs that typically hybridize to the 3'UTR of targeted transcripts. This binding guides Argonaute proteins to direct targeted mRNAs to the miRNA- induced

silencing complex (miRISC), leading to translational repression and/or nucleolytic degradation. An early report showed MicroRNA-539 destabilizes *OGA* mRNA transcripts [58]. Recently, miRNA-15b has been shown to target *OGT* mRNA in Th17 cells [59], and the oxidative

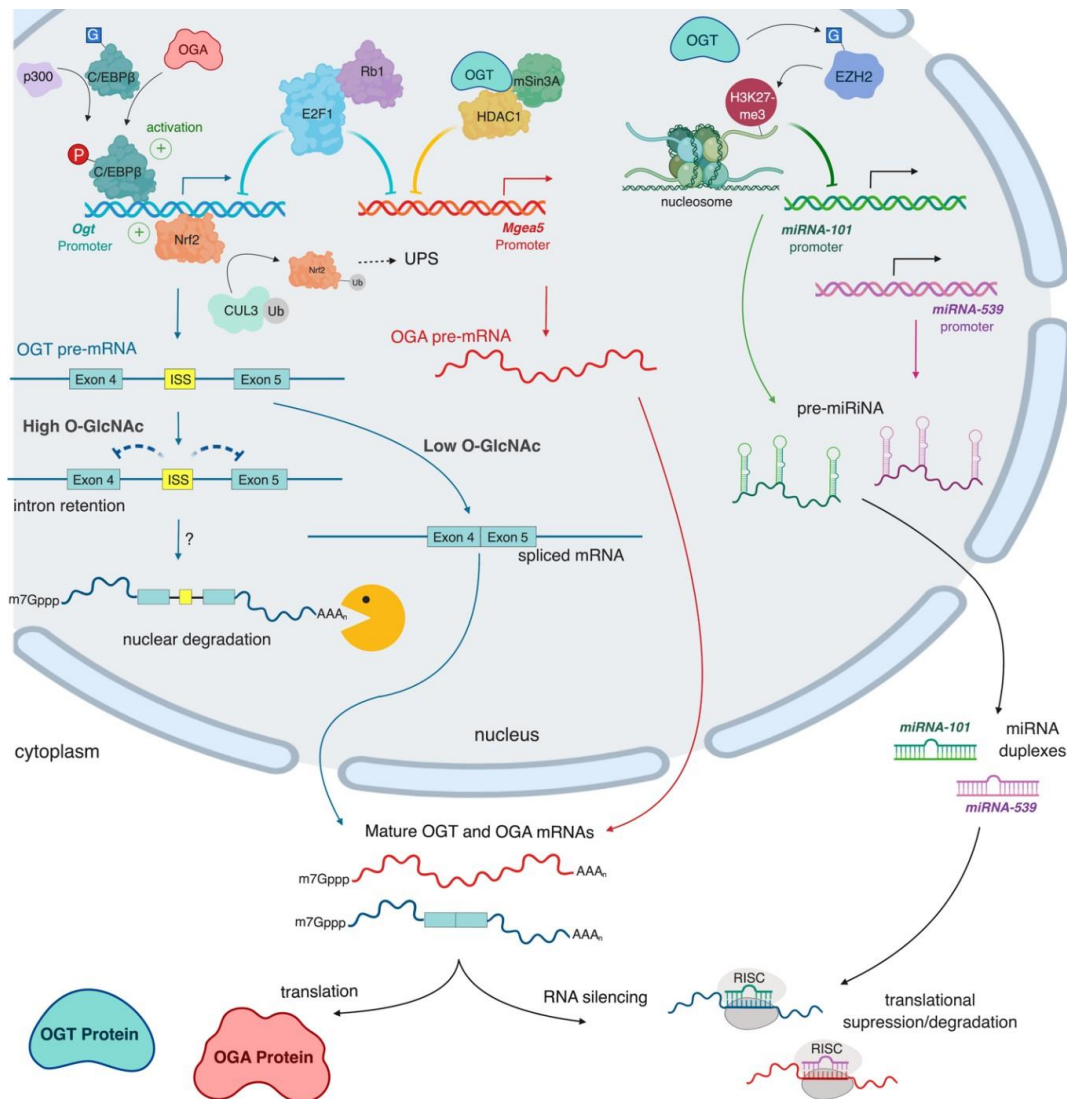


Figure 5. Transcriptional and posttranscriptional regulation of *OGA* and *OGT* expression. OGT is regulated at the transcriptional level by activators (C/EBP β , Nrf2) that are themselves modulated by secondary regulatory mechanisms. Both *OGA* and *OGT* promoters are negatively regulated by E2F1 in a Rb1 dependant fashion. Furthermore, *OGA* is thought to be transcriptionally repressed by the OGT-mSin3A-HDAC1 complex. The *OGT* mRNA transcript contains a retained fourth intron with an O-GlcNAc sensitive internal splicing silencer (ISS). In conditions of high O-GlcNAc the ISS prevents splicing of the detained intron leading to nuclear degradation of the transcript, whereas low O-GlcNAc results in splicing of the intron and subsequent nuclear export. The cytoplasmically-localized mature *OGT* and *OGA* mRNA transcripts are either translated into protein product or can be targeted for degradation via hybridization to miRNAs and subsequent RNA-induced silencing. *OGT* and *EZH2* take part in a negative feedback mechanism to repress the miRNA-101 promoter.

stress responsive miRNA-200a and miRNA-200b antagonize the *OGT* mRNA transcript in conditions of high glucose in human aortic endothelial cells [59,60]. Recently, miRNA-101 has been found to target the transcripts of both *OGT* and *EZH2* for degradation. Interestingly, *OGT* mediated O-GlcNAcylation has

been proposed to stabilize EZH2 thereby promoting EZH2 catalyzed histone H3K27 trimethylation within the miRNA-101 promoter. Because the H3K27me3 mark transcriptionally silences miRNA-101, a negative feedback mechanism that promotes metastatic colorectal cancer may operate (Figure 5) [61]. Finally, in hepatocytes, new data shows HCV infection induces miRNA-501-61p binding to the 3'UTR of OGT transcripts, blocking translation and leading to decreased OGT levels [62]. Moreover, this last observation is recapitulated bi-directionally using OGA and OGT inhibitors. On balance, miRNA mediated degradation is emerging as a key mechanism regulating context dependant fine control of expression of O-GlcNAc cycling enzymes.

Another post-transcriptional mechanism controlling OGT expression is the regulated splicing of its fourth intron, through the recently recognized phenomenon of retained introns (RI) [63]. High cellular O-GlcNAc levels are correlated with increased abundance of the RI spliceoform of the *OGT* transcript (OGT-RI) within the nucleus. Notably, pharmacological inhibition of OGT reduced OGT-RI levels, which correlated with enhanced OGT expression [63]. Furthermore, a 526-nts O-GlcNAc sensitive intronic splicing silencer (ISS) located within the fourth intron regulates abundance of this non-translated spliceoform (Figure 5) [63]. Currently, the cellular fate of the OGT-RI, and the trans-acting factors that bind and regulate the ISS in response to fluctuating O-GlcNAc levels represent major unknowns of high interest.

Conclusion

In summary, recent crystallographic investigations of OGT and OGA have provided a structural foundation for understanding the molecular mechanisms regulating O-GlcNAc cycling. Upon exploring the current state of knowledge about these mechanisms, we emerge with a view that an interconnected network of regulatory factors dynamically tailors the abundance, localization, and activity of the O-GlcNAc cycling enzymes. For example, PTMs and protein interaction partners can exert pleiotropic effects on the localization and activity of OGT and OGA (Figure 5). Pre-translational regulatory mechanisms are emerging as major factors orchestrating global O-GlcNAc homeostasis. Additionally, nutrient-based regulation plays defined roles as well, both through modulation of these processes and by directly regulating levels of donor sugar substrates. While many new regulatory processes are emerging, understanding of events at the atomic level remains largely unknown. A major challenge is to integrate cell and tissue-based discovery studies with quantitative biochemical analyses to uncover the detailed molecular and atomic basis for regulation of O-GlcNAc. Such findings will illuminate how these events are antagonistic or synergistic in coordinating both global and target specific O-GlcNAcylation to meet cellular needs. Finally, given that O-GlcNAc is emerging as a druggable pathway, delving deeper to uncover these mysteries will likely enable targeted therapeutic approaches that modulate O-GlcNAc for human benefit.

Conflict of interest statement

DJV is cofounder of and holds equity in the company Alectos Therapeutics. DJV serves as CSO and Chair of the Scientific Advisory Board (SAB) of Alectos Therapeutics. GJD serves on the SAB of Alectos Therapeutics. DJV may receive royalties from SFU for commercialization of technology relating to OGA inhibitors.

Acknowledgements

This work was supported by grants from the Canadian Institutes for Health Research (CIHR) (D.J.V., PJT-148732, PJT-156202) and the Natural Sciences and Engineering Council of Canada (Discovery-RGPIN298406). D.J.V. is a Tier I Canada Research Chair in Chemical Biology. D.T.K. is supported by a postdoctoral fellowship from CIHR and a Trainee Award from the Michael Smith Foundation for Health Research (MSFHR). G.J.D is a Royal Society Ken Murray Research Professor. A.M. is supported by the Biotechnology and Biological Sciences Research Council (BBSRC BB/M011151/1).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

*of special interest

** of outstanding interest

**[3] An important paper in the field. The authors provide unambiguous evidence for mutations in *ogt* implicated in XLID influencing OGT function. Despite largely normal global O- GlcNAc levels, significant changes in gene transcription are observed. Furthermore, a compensatory mechanism to maintain global O-GlcNAc levels involving OGT-mSin3A- HDAC1 dependant transcriptional repression at the OGA promoter is proposed.

*[6] The authors develop a systematic platform for glycosylation sequence characterization and optimization and use this method to characterize OGT sequence preference. This analysis revealed a surprising preference for an aromatic residue at the -4 subsite.

*[10] An asparagine ladder in the TPR domains mediates OGT interaction with protein substrates.

*[12] New chemical biology tools to identify features of OGT responsible for substrate targeting. This method uncovered a patch of residues lining the inner surface of the N-terminal domain that contribute to interaction with multiple substrates.

**[13] A nice structural study defines the effect of XLID-associated OGT mutation L254F as distorting the TPR helix.

*[27] mOGT was shown to be catalytically active *in vivo* and plays an important role in supporting mitochondrial structure and function.

**[28] A fundamental advance in discovery of the OGT NLS and the effects of S389 O- GlcNAcylation and interaction with importin α 5 for nuclear import.

*[35] The authors demonstrate that OGT is enriched in the post-synaptic density of excitatory neuronal synapses whereby it regulates synapse maturation.

*[37] Upon glucagon-induced calcium signalling CamKII phosphorylates OGT at S20, which induces O-GlcNAc modification of Ulk proteins through potentiation of AMPK- dependant phosphorylation.

**[38] Chk1 phosphorylates OGT S20, and this modification regulates OGT localization to the mid-body and regulation of vimentin bridge formation and severing during mitosis.

*[39] LSD2 was shown to act as an E3 ligase to ubiquitinylate OGT and thereby facilitate its UPS

mediated degradation.

**[43] Circadian nuclear receptor REV-ERB α was found to form a stable complex with OGT and regulate OGT-controlled processes in both the cytosol and nucleus.

*[44] The authors characterize UAP1L1 as a novel OGT binding partner that is necessary for O-GlcNAcylation in cells.

**[47] This study leverages BioID proximity biotinylation to identify 90 OGA-interaction partners including FAS, which when bound causes a substantial reduction in OGA catalytic activity.

** [52] The authors report the structure based evolution of OGT inhibitors with nanomolar affinity. Moving forward, these inhibitors may provide a key tool to facilitate discovery and targeted studies of O-GlcNAc in cells.

*[53] This study demonstrates that OGA acts in conjunction with histone acetyltransferase p300 to positively regulate transcription factor C/EBP β , which acts as a transcriptional activator at the *ogt* promoter.

* [57] Nrf2 was demonstrated to bind the *ogt* promoter and enhance transcription. Furthermore, E3 ubiquitin ligase CUL3 negatively regulates *ogt* transcript through targeted degradation of Nrf2.

* [61] The authors demonstrate that miRNA-101 targets degradation of both OGT and EZH2. Furthermore, OGT and EZH2 catalyze O-GlcNAcylation and H3K27 trimethylation within the miR-101 promoter providing a negative feedback mechanism that transcriptionally silences miRNA-101.

**[63] An exceptional finding. Retention of the OGT fourth intron is sensitive to global O-GlcNAc levels. Furthermore, this retained intron contains an ISS that regulates its splicing – by what mechanism remains a tantalizing mystery.

References

1. Yang X, Qian K: **Protein O-GlcNAcylation: emerging mechanisms and functions.** *Nat Rev Mol Cell Biol* 2017, **18**:452-465.
2. Banerjee PS, Lagerlöf O, Hart GW: **Roles of O-GlcNAc in chronic diseases of aging.** *Mol Aspects Med* 2016, **51**:1-15.
3. Vaidyanathan K, Niranjana T, Selvan N, Teo CF, May M, Patel S, Weatherly B, Skinner C, Opitz J, Carey J, et al.: **Identification and characterization of a missense mutation in the O-linked β -N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability.** *J Biol Chem* 2017, **292**:8948-8963.
4. Joiner CM, Li H, Jiang J, Walker S: **Structural characterization of the O-GlcNAc cycling enzymes: insights into substrate recognition and catalytic mechanisms.** *Curr Opin Struct Biol* 2019, **56**:97-106.
5. Pathak S, Alonso J, Schimpl M, Rafie K, Blair DE, Borodkin VS, Albarbarawi O, van Aalten DMF: **The active site of O-GlcNAc transferase imposes constraints on substrate sequence.** *Nat Struct Mol Biol* 2015, **22**:744-750.
6. Kightlinger W, Lin L, Rosztochy M, Li W, DeLisa MP, Mrksich M, Jewett MC: **Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases.** *Nat Chem Biol* 2018, **14**:627-635.
7. Loney AC, El Atmioui D, Wu W, Ovaa H, Heck AJR: **Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation.** *Proc Natl Acad Sci U S A* 2017, **114**:E7255-E7261.
8. Britto-Borges T, Barton GJ: **A study of the structural properties of sites modified by the O-linked β -N-acetylglucosamine transferase.** *PLoS One* 2017, **12**:e0184405.
9. Zhu Y, Liu TW, Cecioni S, Eskandari R, Zandberg WF, Vocadlo DJ: **O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains.** *Nat Chem Biol* 2015, **11**:319-325.
10. Levine ZG, Fan C, Melicher MS, Orman M, Benjamin T, Walker S: **O-GlcNAc Transferase Recognizes Protein Substrates Using an Asparagine Ladder in the Tetratricopeptide Repeat (TPR) Superhelix.** *J Am Chem Soc* 2018, **140**:3510-3513.
11. Rafie K, Raimi O, Ferenbach AT, Borodkin VS, Kapuria V, van Aalten DMF: **Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats.** *Open Biol* 2017, **7**.
12. Hu CW, Worth M, Fan D, Li B, Li H, Lu L, Zhong X, Lin Z, Wei L, Ge Y, et al.: **Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase.** *Nat Chem Biol* 2017, **13**:1267-1273.
13. Gundogdu M, Llabrés S, Gorelik A, Ferenbach AT, Zachariae U, van Aalten DMF: **The O-GlcNAc Transferase Intellectual Disability Mutation L254F Distorts the TPR Helix.** *Cell Chem Biol* 2018, **25**:513-518.e514.

14. Selvan N, George S, Serajee FJ, Shaw M, Hobson L, Kalscheuer V, Prasad N, Levy SE, Taylor J, Aftimos S, et al.: **O-GlcNAc transferase missense mutations linked to X-linked intellectual disability deregulate genes involved in cell fate determination and signaling.** *J Biol Chem* 2018, **293**:10810-10824.
15. Willems AP, Gundogdu M, Kempers MJE, Giltay JC, Pfundt R, Elferink M, Loza BF, Fuijkschot J, Ferenbach AT, van Gassen KLI, et al.: **Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability.** *J Biol Chem* 2017, **292**:12621- 12631.
16. Joiner CM, Levine ZG, Aonbangkhen C, Woo CM, Walker S: **Aspartate Residues Far from the Active Site Drive O-GlcNAc Transferase Substrate Selection.** *J Am Chem Soc* 2019.
17. Elsen NL, Patel SB, Ford RE, Hall DL, Hess F, Kandula H, Kornienko M, Reid J, Selnick H, Shipman JM, et al.: **Insights into activity and inhibition from the crystal structure of human O- GlcNAcase.** *Nat Chem Biol* 2017, **13**:613-615.
18. Roth C, Chan S, Offen WA, Hemsworth GR, Willems LI, King DT, Varghese V, Britton R, Vocadlo DJ, Davies GJ: **Structural and functional insight into human O-GlcNAcase.** *Nat Chem Biol* 2017, **13**:610-612.
19. Li B, Li H, Lu L, Jiang J: **Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode.** *Nat Struct Mol Biol* 2017, **24**:362-369.
20. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.** *Nucleic Acids Res* 2015, **43**:D512-520.
21. Li B, Li H, Hu CW, Jiang J: **Structural insights into the substrate binding adaptability and specificity of human O-GlcNAcase.** *Nat Commun* 2017, **8**:666.
22. Males A, Davies GJ: **Structural studies of a surface-entropy reduction mutant of O- GlcNAcase.** *Acta Crystallogr D Struct Biol* 2019, **75**:70-78.
23. Shen DL, Gloster TM, Yuzwa SA, Vocadlo DJ: **Insights into O-linked N-acetylglucosamine (O- GlcNAc) processing and dynamics through kinetic analysis of O-GlcNAc transferase and O-GlcNAcase activity on protein substrates.** *J Biol Chem* 2012, **287**:15395-15408.
24. Wright TH, Bower BJ, Chalker JM, Bernardes GJ, Wiewiora R, Ng WL, Raj R, Faulkner S, Vallée MR, Phanumartwiwath A, et al.: **Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity.** *Science* 2016, **354**:597.
25. Sharif S, Shi J, Ruijtenbeek R, Pieters RJ: **Study of cross talk between phosphatases and OGA on a ZO-3-derived peptide.** *Amino Acids* 2019, **51**:739-743.
26. Trapannone R, Mariappa D, Ferenbach AT, van Aalten DM: **Nucleocytoplasmic human O- GlcNAc transferase is sufficient for O-GlcNAcylation of mitochondrial proteins.** *Biochem J* 2016, **473**:1693-1702.
27. Sacoman JL, Dagda RY, Burnham-Marusich AR, Dagda RK, Berninsone PM: **Mitochondrial O- GlcNAc transferase (mOGT) regulates mitochondrial structure, function, and survival in HeLa cells.** *J Biol Chem* 2017, **292**:4499-4518.
28. Seo HG, Kim HB, Kang MJ, Ryum JH, Yi EC, Cho JW: **Identification of the nuclear localisation signal of O-GlcNAc transferase and its nuclear import regulation.** *Sci Rep* 2016, **6**:34614.
29. Banerjee PS, Ma J, Hart GW: **Diabetes-associated dysregulation of O-GlcNAcylation in rat cardiac mitochondria.** *Proc Natl Acad Sci U S A* 2015, **112**:6050-6055.
30. Slawson C, Lakshmanan T, Knapp S, Hart GW: **A mitotic GlcNAcylation/phosphorylation signaling complex alters the posttranslational state of the cytoskeletal protein vimentin.** *Mol Biol Cell* 2008, **19**:4130-4140.
31. Zhou LT, Romar R, Pavone ME, Soriano-Úbeda C, Zhang J, Slawson C, Duncan FE: **Disruption of O-GlcNAc homeostasis during mammalian oocyte meiotic maturation impacts fertilization.** *Mol Reprod Dev* 2019, **86**:543-557.
32. Whelan SA, Lane MD, Hart GW: **Regulation of the O-linked beta-N-acetylglucosamine transferase by insulin signaling.** *J Biol Chem* 2008, **283**:21411-21417.
33. Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S: **Structure of human O-GlcNAc transferase and its complex with a peptide substrate.** *Nature* 2011, **469**:564-567.
34. Yang X, Ongusaha PP, Miles PD, Havstad JC, Zhang F, So WV, Kudlow JE, Michell RH, Olefsky JM, Field SJ, et al.: **Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance.** *Nature* 2008, **451**:964-969.
35. Lagerlöf O, Hart GW, Haganir RL: **O-GlcNAc transferase regulates excitatory synapse maturity.** *Proc Natl Acad Sci U S A* 2017, **114**:1684-1689.
36. Bullen JW, Balsbaugh JL, Chanda D, Shabanowitz J, Hunt DF, Neumann D, Hart GW: **Cross-talk between**

- two essential nutrient-sensitive enzymes: O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *J Biol Chem* 2014, **289**:10592-10606.
37. Ruan HB, Ma Y, Torres S, Zhang B, Feriod C, Heck RM, Qian K, Fu M, Li X, Nathanson MH, et al.: **Calcium-dependent O-GlcNAc signaling drives liver autophagy in adaptation to starvation.** *Genes Dev* 2017, **31**:1655-1665.
 38. Li Z, Li X, Nai S, Geng Q, Liao J, Xu X, Li J: **Checkpoint kinase 1–induced phosphorylation of O- linked β -N-acetylglucosamine transferase regulates the intermediate filament network during cytokinesis.** *J Biol Chem* 2017, **292**:19548-19555.
 39. Yang Y, Yin X, Yang H, Xu Y: **Histone demethylase LSD2 acts as an E3 ubiquitin ligase and inhibits cancer cell growth through promoting proteasomal degradation of OGT.** *Mol Cell* 2015, **58**:47-59.
 40. Dey A, Seshasayee D, Noubade R, French DM, Liu J, Chaurushiya MS, Kirkpatrick DS, Pham VC, Lill JR, Bakalarski CE, et al.: **Loss of the tumor suppressor BAP1 causes myeloid transformation.** *Science* 2012, **337**:1541-1546.
 41. Kaasik K, Kivimäe S, Allen JJ, Chalkley RJ, Huang Y, Baer K, Kissel H, Burlingame AL, Shokat KM, Ptáček LJ, et al.: **Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock.** *Cell Metab* 2013, **17**:291-302.
 42. Butkinaree C, Cheung WD, Park S, Park K, Barber M, Hart GW: **Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis.** *J Biol Chem* 2008, **283**:23557-23566.
 43. Berthier A, Vinod M, Porez G, Steenackers A, Alexandre J, Yamakawa N, Gheeraert C, Ploton M, Maréchal X, Dubois-Chevalier J, et al.: **Combinatorial regulation of hepatic cytoplasmic signaling and nuclear transcriptional events by the OGT/REV-ERB α complex.** *Proc Natl Acad Sci U S A* 2018, **115**:E11033-E11042.
 44. Lai CY, Liu H, Tin KX, Huang Y, Yeh KH, Peng HW, Chen HD, He JY, Chiang YJ, Liu CS, et al.: **Identification of UAP1L1 as a critical factor for protein O-GlcNAcylation and cell proliferation in human hepatoma cells.** *Oncogene* 2019, **38**:317-331.
 45. Gambetta MC, Müller J: **A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin.** *Chromosoma* 2015, **124**:429-442.
 46. Wu D, Cai Y, Jin J: **Potential coordination role between O-GlcNAcylation and epigenetics.** *Protein Cell* 2017, **8**:713-723.
 47. Groves JA, Maduka AO, O'Meally RN, Cole RN, Zachara NE: **Fatty acid synthase inhibits the O-GlcNAcase during oxidative stress.** *J Biol Chem* 2017, **292**:6493-6511.
 48. Marshall S, Nadeau O, Yamasaki K: **Dynamic actions of glucose and glucosamine on hexosamine biosynthesis in isolated adipocytes: differential effects on glucosamine 6- phosphate, UDP-N-acetylglucosamine, and ATP levels.** *J Biol Chem* 2004, **279**:35313- 35319.
 49. Chiaradonna F, Ricciardiello F, Palorini R: **The Nutrient-Sensing Hexosamine Biosynthetic Pathway as the Hub of Cancer Metabolic Rewiring.** *Cells* 2018, **7**.
 50. Wang ZV, Deng Y, Gao N, Pedrozo Z, Li DL, Morales CR, Criollo A, Luo X, Tan W, Jiang N, et al.: **Spliced X-box binding protein 1 couples the unfolded protein response to hexosamine biosynthetic pathway.** *Cell* 2014, **156**:1179-1192.
 51. Zhang Z, Tan EP, VandenHull NJ, Peterson KR, Slawson C: **O-GlcNAcase Expression is Sensitive to Changes in O-GlcNAc Homeostasis.** *Front Endocrinol (Lausanne)* 2014, **5**:206.
 52. Martin SES, Tan ZW, Itkonen HM, Dubeau DY, Paulo JA, Janetzko J, Boutz PL, Törk L, Moss FA, Thomas CJ, et al.: **Structure-Based Evolution of Low Nanomolar O-GlcNAc Transferase Inhibitors.** *J Am Chem Soc* 2018, **140**:13542-13545.
 53. Qian K, Wang S, Fu M, Zhou J, Singh JP, Li MD, Yang Y, Zhang K, Wu J, Nie Y, et al.: **Transcriptional regulation of O -GlcNAc homeostasis is disrupted in pancreatic cancer.** *J Biol Chem* 2018, **293**:13989-14000.
 54. Muthusamy S, Hong KU, Dassanayaka S, Hamid T, Jones SP: **E2F1 Transcription Factor Regulates O-linked N-acetylglucosamine (O-GlcNAc) Transferase and O-GlcNAcase Expression.** *J Biol Chem* 2015, **290**:31013-31024.
 55. Zhang C, Xie F, Li L, Zhang Y, Ying W, Liu L, Yan X, Yin F, Zhang L: **Hepatocyte nuclear factor 1 alpha (HNF1A) regulates transcription of O-GlcNAc transferase in a negative feedback mechanism.** *FEBS Lett* 2019, **593**:1050-1060.
 56. Zeng Q, Zhao RX, Chen J, Li Y, Li XD, Liu XL, Zhang WM, Quan CS, Wang YS, Zhai YX, et al.: **O- linked GlcNAcylation elevated by HPV E6 mediates viral oncogenesis.** *Proc Natl Acad Sci U S A* 2016, **113**:9333-

9338.

57. Li X, Zhang Z, Li L, Gong W, Lazenby AJ, Swanson BJ, Herring LE, Asara JM, Singer JD, Wen H: **Myeloid-derived cullin 3 promotes STAT3 phosphorylation by inhibiting OGT expression and protects against intestinal inflammation.** *J Exp Med* 2017, **214**:1093-1109.
58. Muthusamy S, DeMartino AM, Watson LJ, Brittian KR, Zafir A, Dassanayaka S, Hong KU, Jones SP: **MicroRNA-539 is up-regulated in failing heart, and suppresses O-GlcNAcase expression.** *J Biol Chem* 2014, **289**:29665-29676.
59. Liu R, Ma X, Chen L, Yang Y, Zeng Y, Gao J, Jiang W, Zhang F, Li D, Han B, et al.: **MicroRNA-15b Suppresses Th17 Differentiation and Is Associated with Pathogenesis of Multiple Sclerosis by Targeting O-GlcNAc Transferase.** *J Immunol* 2017, **198**:2626-2639.
60. Lo WY, Yang WK, Peng CT, Pai WY, Wang HJ: **MicroRNA-200a/200b Modulate High Glucose- Induced Endothelial Inflammation by Targeting.** *Front Physiol* 2018, **9**:355.
61. Jiang M, Xu B, Li X, Shang Y, Chu Y, Wang W, Chen D, Wu N, Hu S, Zhang S, et al.: **O- GlcNAcylation promotes colorectal cancer metastasis via the miR-101-O-GlcNAc/EZH2 regulatory feedback circuit.** *Oncogene* 2019, **38**:301-316.
62. Herzog K, Bandiera S, Pernot S, Fauvelle CJh, FrankWeiss, Amélie Bull, AnneDurand, Sarah C Chane-Woon-Ming, Béatrice Pfeffer, Sébastien et. al.: **Functional microRNA screen uncovers O-linked N-acetylglucosamine transferase as a host factor modulating hepatitis C virus morphogenesis and infectivity.** *Gut* 2019, **0**:1-13.
63. Park SK, Zhou X, Pendleton KE, Hunter OV, Kohler JJ, O'Donnell KA, Conrad NK: **A Conserved Splicing Silencer Dynamically Regulates O-GlcNAc Transferase Intron Retention and O- GlcNAc Homeostasis.** *Cell Rep* 2017, **20**:1088-10

Mannosidase mechanism: At the intersection of conformation and catalysis

Carme Rovira,^{1,2,*} Alexandra Males,³ Gideon J. Davies,³ Spencer J. Williams^{4,*}

1 Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

2 Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain

3 Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

4 School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

sjwill@unimelb.edu.au

Abstract

Mannosidases are a diverse group of enzymes that are important in the biological processing of mannose-containing polysaccharides and complex glycoconjugates. They are found in 12 of the >160 sequence-based glycosidase families. We discuss evidence that nature has evolved a small set of common mechanisms that unite almost all of these mannosidase families. Broadly, mannosidases (and the closely related rhamnosidases) perform catalysis through just two conformations of the oxocarbenium ion-like transition state: a $B_{2,5}$ (or enantiomeric $^{2,5}B$) boat and a 3H_4 half-chair. This extends to a new family (GT108) of GDPMan-dependent β -1,2-mannosyltransferases/phosphorylases that perform mannosyl transfer through a boat conformation as well as some mannosidases that are metalloenzymes and require divalent cations for catalysis. Yet, among this commonality lies diversity. New evidence shows that one unique family (GH99) of mannosidases use an unusual mechanism involving anchimeric assistance via a 1,2-anhydro sugar (epoxide) intermediate.

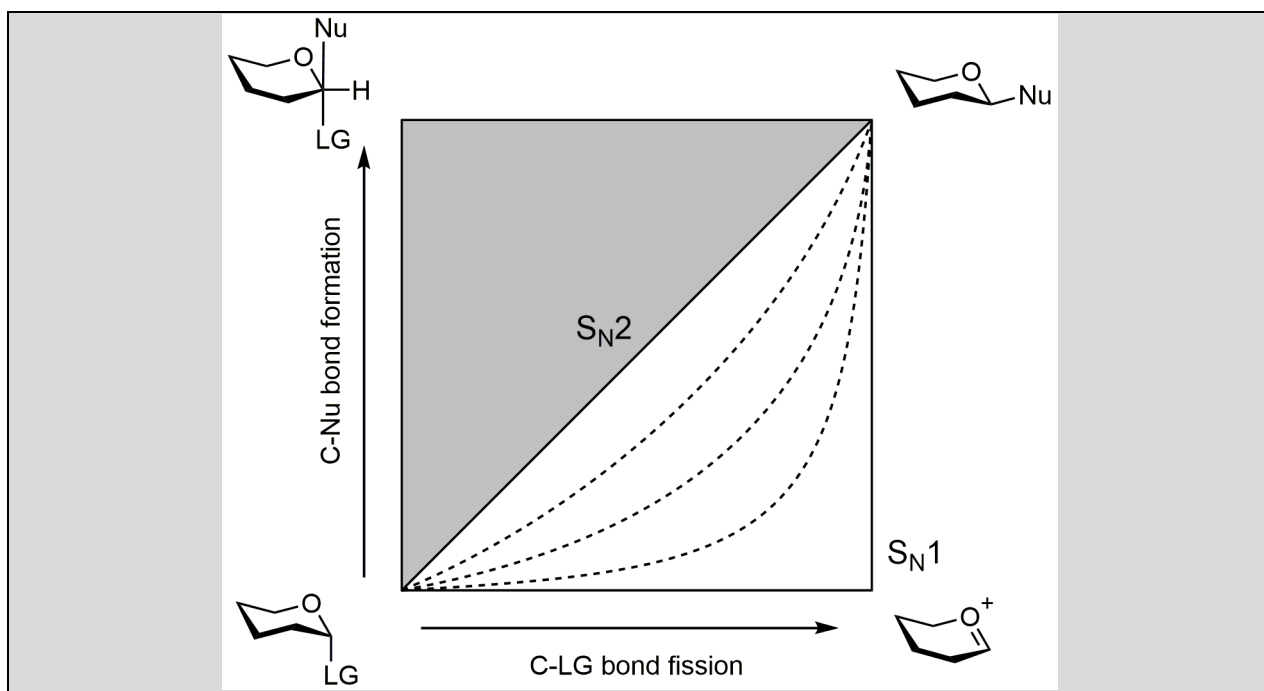
Introduction

Mannose-containing polysaccharides and glycoconjugates are widespread in nature. For example, β -1,4-mannans are found in a variety of plant sources and include the galactomannans and glucomannans [1]. The cell wall of fungi contain complex heteromannans with an α -1,6-mannan backbone [2] and related structures are present within the lipomannan and lipoarabinomannan structures of mycobacteria and corynebacterial [3]. β -1,2-mannans are produced as a soluble cytosolic carbohydrate reserve in leishmania [4]. And as part of N-linked glycoprotein synthesis, mammals and other eukaryotes produce complex N-glycans that contain β -1,4-, α -1,2-, α -1,3 and α -1,6-mannose linkages. The degradation of these glycosides is achieved by diverse mannosidases (here we use this term to include mannanases) that cleave α - or β -mannoside linkages in a wide variety of substrates. Mannosidases are found within 12 of the glycoside hydrolase families of the Carbohydrate Active enzyme classification (www.cazy.org [5]; www.cazypedia.org [6]). **Table 1** summarizes the various sequence-based families containing mannosidases. Three of these families host α -mannosidase metalloenzymes, with dependency on divalent metals for activity, typically Ca^{2+} or Zn^{2+} .

The chemical transformations used by glycosidases to cleave the glycosidic linkage primarily involve substitution reactions at the anomeric centre of the sugar. Sugars possess a ring oxygen that can provide varying degrees of electronic stabilization to developing charge at C1 and under appropriate conditions the anomeric centre can engage in a continuum of nucleophilic substitution mechanisms that span the spectrum from $\text{S}_{\text{N}}1$ to $\text{S}_{\text{N}}2$ reactions [7]. The case of a "quintessential" concerted $\text{S}_{\text{N}}2$ reaction at the anomeric position represents one mechanistic extreme, and has only been observed in non-biological catalysis using exceptional nucleophiles (e.g. a glycosyl fluoride with azide) [8] (see **Side panel A**). The other mechanistic extreme involves a "quintessential" stepwise $\text{S}_{\text{N}}1$ reaction with the formation of a glycosyl cation. Glycosyl cations have recently been synthesized in the condensed phase [9], and their involvement as intermediates in non-enzymatic solution-phase glycosylation reactions firmly established [10].

Side panel A. Illuminating the spectrum of reactivity with More O'Ferrall-Jencks plots

More O'Ferrall [11] and Jencks [12] introduced a 2-D plot (which may be contoured to represent energy in the third dimension) that is a useful tool to illustrate the continuity between concerted and stepwise processes. Each axis represents the formation or breakage of a given bond, and a trajectory between substrate at the origin with product at the diagonal illustrates a particular reaction coordinate as a function of changes in bond order for two reactive bonds. More O'Ferrall-Jencks plots are useful to demonstrate the continuity between concerted and stepwise processes such as substitution and elimination reactions. A More O'Ferrall-Jencks plot for an anomeric substitution is shown below. The diagonal represents a perfect $\text{S}_{\text{N}}2$ process, the pathway along the x -axis a completely stepwise $\text{S}_{\text{N}}1$ process, and the dotted lines are pathways highlighting the continuum between the two (with dissociative character increasing as the pathway approaches the x -axis). The shaded area above the diagonal is excluded for substitution at carbon, but may be involved in substitution at phosphorus or sulfur.



Glycosidases typically possess active sites that stabilize oxocarbenium ion character by favouring certain sugar shapes or conformations [13]. However, in biological catalysis the presence of nucleophilic species within active sites means that glycosyl cations are generally believed not to possess a discrete existence in enzymatic mechanisms [14]. Instead, within enzyme active sites, nucleophilic participation at the anomeric centre is coupled to leaving group departure (dotted lines in **Side panel A**). 'Tight' and 'loose' transition states can be distinguished defined by the sum of the nucleophile...C1 and leaving group...C1 distances [15]. Nonetheless, given the oxocarbenium ion character that develops during glycosidic bond cleavage, the structure and reactivity of glycosyl cations are central to considerations of reactivity in biological catalysis, and provide key insights into the nature of oxocarbenium ion-like transition states. Electronic stabilization of charge development at the transition state results in partial double bond development between O5 and C1 meaning that conformations in which C2-C1-O5-C5 are planar (or close to) are preferred [14]. This includes the idealized half-chair (3H_4 , 4H_3), boat ($B_{2,5}$, ${}^{2,5}B$) and envelope (4E , E_4 and 3E , E_3) conformations [16] (see **Side panel B**).

Side panel B. Conformational nomenclature of pyranoses

IUPAC have developed a useful nomenclature to denote the conformations adopted by a pyranose ring [17]. This system starts by assigning a capital letter (in italics) to indicate the overall shape, *C* = chair, *B* = boat, *H* = half chair, *S* = skew boat, *E* = envelope. The first four of these conformations has four atoms in a plane; the *E* conformation has five. The specific conformation is derived by determining the four (or five) atoms that lie in a plane; assigning the top and bottom face of the ring by using a left-hand rule counting in order of increasing ring carbon number; and then indicating the relative positions of the remaining two (or one) atoms that lie out of the plane using top-face = superscript prefix; bottom-face = subscript suffix. This system needs to be considered carefully when applied to enantiomers of the common sugars. Enantiomeric conformers have the same reference plane, but application of the left-hand rule results in a swap of the top and bottom faces. Thus, the enantiomer of D-mannose in a 4C_1 conformation is L-mannose in a 1C_4 conformation. This situation applies especially in the context of this review to identifying conformers of L-rhamnose.

Glycosidases nominally follow two main nucleophilic substitution mechanisms, distinguished by the stereochemical outcome of the anomeric substitution reaction [14]. Inverting glycosidases operate through one-step mechanisms involving a general acid that assists leaving group departure, and a water nucleophile, assisted by a general base (**Figure 1c,d**). Retaining glycosidases use two-step mechanisms

involving a covalent intermediate assisted by an enzymic general acid/base residue [18]. Most retaining glycosidases use an enzyme-borne nucleophile (typically aspartate, glutamate or tyrosine). In the first step the enzyme nucleophile approaches the anomeric centre; coupled to its approach is the departure of the anomeric leaving group, assisted by the enzymic acid/base acting as general acid, resulting in the formation of a covalent glycosyl enzyme (**Figure 1a,b**). In the second step the acid/base acts as general base to deprotonate a water molecule, which substitutes the anomeric centre and cleaves the glycosyl enzyme to restore the enzymatic nucleophile. An important variant of the retaining mechanism applies for substrates bearing a pendant acetamido group situated trans to the glycoside, such as N-acetyl- β -hexosaminides. Certain N-acetyl- β -hexosaminidases utilise a neighboring group participation mechanism in which the substrate acetamido group acts as a nucleophile to displace the leaving group, again with the involvement of an enzymatic general acid/base to promote leaving group departure [19]. These enzymes form a cyclic intermediate that may be either an oxazoline or oxazolinium ion [20,21], and possess a general acid/base that both enhances the nucleophilicity of the acetamido group, and stabilizes the oxazolinium ion-like transition state [20,22,23].

In and out of trouble: Nucleophilic substitution of mannosides

Among the range of carbohydrates that nature's catalysts must contend with, polymers and glycoconjugates formed from D-mannose (and its pseudo-enantiomer L-rhamnose) comprise a special case. In mannopyranosides in the normal 4C_1 chair conformation, the 2-hydroxyl group occupies an axial orientation. This orientation presents a challenge for nucleophilic substitution reactions at the anomeric position from the β -face. To understand why, one may consider the neopentyl effect (**Figure 2a**): rates of bimolecular S_N2 substitution by ethoxide in neopentyl bromide is around 2,000,000-fold slower than for methyl bromide, more than 150,000-fold slower than for ethyl bromide, and around 4,500-fold slower than the equivalent substitution in isobutyl bromide [24]. The resistance of neopentyl bromide to substitution is credited to blocking of the trajectory required for nucleophilic attack.

Pyranoses can take advantage of their flexibility to relieve the nucleophile blocking effect of the axial 2-hydroxyl and adopt more reactive conformations. The full set of pyranose conformations can be represented using the elegant method of Cremer and Pople [25] using θ/ϕ puckering coordinates on a pseudo-sphere or in 2D in a Mercator projection. In this representation, the 'normal' 4C_1 conformation lies at the north pole; the inverted 1C_4 conformation at the south pole; the equator consists of various boat and skew conformations; and at the two tropics lie the envelope and half-chair conformations. Varying the ring conformation results in changes to the orientation of O2, which can alleviate the steric clash that occurs as a nucleophile approaches C1, for example by adopting 0S_2 or 3S_1 conformations (**Figure 2a**). Which conformations are most beneficial for substitutions on mannose C1? Fig 1b shows the degree of 'axiality' of the 2-OH across the conformational orientations, which reveals that the most advantageous are the 3H_4 and $B_{2,5}$ conformations; these conformations achieve planarity across C2-C1-O5-C5, and maximize the 'equatoriality' of O2, opening up a nucleophilic trajectory for substitution at C1. Almost all mannosidase Michaelis complexes that have been structurally characterized exhibit pyranose ring conformations in which the 2-hydroxyl is not axial (blue and white regions in **Figure 2b**); family GH99 represents an important exception that will be discussed later.

It has been known since the 1960s that several α -mannosidase (and α -rhamnosidase) families are metalloenzymes [26] that require divalent metal cations (typically Zn^{2+} or Ca^{2+}) for activity (**Table 1**). Cation and substrate bind in that order through an ordered sequential mechanism [27], and X-ray structures reveal that the metal coordinates the 2- and 3-hydroxyls of the substrate. It has been suggested that the role of the metal ion is to assist in distortion of α -mannosides/rhamnosides by effects on the torsion angle of O2-C2-C3-O3 (Ω_{O2-O3}) [28]. Possibly, for inverting enzymes, the nucleophilic water required in the reaction mechanism is delivered from within the coordination sphere of the metal ion, which may enhance the nucleophilicity of the water by increasing its acidity [28]. A computational study of a GH38 α -mannosidase (a retaining enzyme) argued that Zn^{2+} coordination to the sugar stabilizes charge development

on O2 at the transition state [29], but the effect of the cation on sugar ring conformation was not investigated.

Based on observations that can be traced back to the work of Tanret in 1894 [30] it is known that α -mannosides and β -glucosides can be cleaved under basic conditions through a mechanism involving neighboring group participation [31]. These reactions involve the formation of a C2-oxyanion that performs an intramolecular nucleophilic attack on C1 to form a 1,2-epoxide (more formally termed a 1,2-anhydro sugar), which is short-lived under basic conditions, and cleaves through exclusive ring-opening at C1 to form a product with overall stereochemical retention. A recent KIE study of the basic solvolysis of PNPMann revealed a distinctive nucleophile KIE value for O2-¹⁸O of 1.04 and anomeric KIEs for anomeric-²H and anomeric 1-¹³C supporting an oxocarbenium ion-like transition state [32]. As nucleophilic substitution will be preferred when O2 and the anomeric leaving group adopt an antiperiplanar arrangement, this reaction occurs with little requirement for conformational distortion from a ⁴C₁ conformation with an axial O2. Until very recently (vide infra, see GH99 enzymes below) proposals for a biological equivalent of this neighboring group participation mechanism by Wallenfels for LacZ β -galactosidase [33,34] and Oppenheimer for NAD⁺ glycohydrolase [35-37] have not withstood mechanistic scrutiny.

Mannoimidazoles: Privileged probes of transition state conformation

The transition state is the cardinal feature of the catalysis reaction coordinate, but as the least stable species is also the most fleeting. How then can we gain experimental insights into the structure and conformation of the transition state? One approach is to make molecules with the ability to mimic key features of the transition state and use X-ray crystallography to obtain high resolution complexes with enzymes and thereby study the structure and interactions that occur at the transition state. As glycosidases possess a transition state with partial double-bond character and positive charge development at C1 and O5, various sugar-shaped imino/azasugars have proven very useful, especially when validated as transition state mimics through kinetic analysis using Linear Free Energy Relationships on Bartlett plots [38,39]. Prime among the most useful compounds are glycoimidazoles and related species that were developed through the work of Vasella [40] and others. Mannoimidazole especially has emerged as a powerful probe of transition state conformation through a detailed understanding of its intrinsic 'off-enzyme' conformational preferences. The conformational free energy landscape of mannoimidazole determined by quantum mechanical methods revealed that this compound provides good transition-state shape mimicry, with the mechanistically relevant half-chair (⁴H₃ and ³H₄), envelope (³E, E₃, ⁴E, and E₄), and boat (^{2,5}B or B_{2,5}) conformations all energetically accessible (**Figure 3**) [41]. The ⁴H₃ conformation is the global minimum, with the ³H₄ conformation just 1 kcal mol⁻¹ higher in energy. The B_{2,5} conformation is 5 kcal mol⁻¹ higher in energy than the ³H₄ conformation and it is in an accessible region of the FEL. Thus, the observation of mannoimidazole bound to enzymes in conformations other than the ⁴H₃ ground state can be considered mechanistically significant, with the enzyme tuning its preferred conformation.

Mannosidases that operate through a B_{2,5} boat conformation

A B_{2,5} transition state conformation for mannosidase catalysis was first proposed based on X-ray structures of a GH26 β -mannanase showing a Michaelis complex in a ¹S₅ conformation [42], and a glycosyl enzyme formed using a 2-deoxy-2-fluoro-mannoside in a ⁰S₂ conformation [42], suggestive of a boat transition state through application of the principle of least nuclear motion (**Figure 1b**) [43,44]. This interpolative approach was bolstered by use of a mannoibiose-derived imidazole, which bound in a B_{2,5} conformation [41]. B_{2,5} conformations have been observed for mannoimidazole complexes of family GH1, 2, 26, 38, 92, 113, and 125 α - and β -mannosidases (**Table 1**). In both retaining family GH38 and inverting GH92 α -mannosidases, Ca²⁺ coordinates O2 and O3 in the mannoimidazole complexes [28,41].

The conformational preference of an enzyme has been shown to overcome even greater conformational ligand preferences. A complex of a retaining GH76 enzyme from *Bacillus circulans* with a 1,6- α -mannobiose-derived isofagomine adopted a B_{2,5} conformation [45]. This was surprising as the

computed FEL revealed this precise conformer to be approx. 8 kcal mol⁻¹ greater than the preferred ⁴C₁ conformation of this inhibitor [41]. QM/MM metadynamics simulation of isofagomine binding to the enzyme revealed that the enzyme dramatically reshapes the FEL to favour the B_{2,5} and ⁰S₂ conformations of this inhibitor on-enzyme, most likely due to a direct hydrogen bonding interactions between the catalytic nucleophile and inhibitor 3-OH and NH²⁺ groups, allowing the enzyme to qualitatively recapitulate the major hydrogen-bonding interactions predicted for the transition state. Collectively, structural and computational studies support an ⁰S₂ → B_{2,5}[‡] → ¹S₅ conformational itinerary for the GH76 glycosylation half-reaction.

Assignment of the conformational itinerary to the metal-independent family GH125 α-mannosidases has proved challenging. Initial complexes with a non-hydrolysable thiosugar and the iminosugar deoxymannojirimycin were undistorted [46]. Ab initio QM/MM metadynamics of the thiosugar complex but with an S→O replacement performed in silico predicted an ⁰S₂ conformation on-enzyme [47]. This prompted a crystallographic experiment using the acid/base mutant D220N and led to direct observation of an ⁰S₂ conformation for the Michaelis complex [47]. This was subsequently extended to the observation of a B_{2,5} conformation of mannoimidazole with this enzyme [48], data that collectively supports the assignment of an ⁰S₂ → B_{2,5}[‡] → ¹S₅ conformational itinerary for GH125 enzymes.

Family GH130 contains β-mannosidases/phosphorylases, both of which act with inversion of stereochemistry. This interesting circumstance arises from two subsets of enzymes that contain either a triad of positively charged (His/Lys/Arg) residues that promote binding of phosphate that can act as the catalytic nucleophile, or instead a pair of glutamates that can activate a water molecule for nucleophilic attack [49]. These enzymes have proven resistant to the obtention of inhibitor complexes. However, various product complexes contain mannose bound in distorted B_{2,5} or B_{2,5}/⁰S₂ conformations, and most informatively, Michaelis complexes with disaccharides in which the -1 subsite sugar was bound in a ¹S₅ conformation [50,51]. These data are again collectively consistent with an ⁰S₂ → B_{2,5}[‡] → ¹S₅ itinerary (**Figure 4a**) [49].

Family GH1 retaining glycosidases are typically β-glucosidases or β-galactosidases. Cairns and co-workers reported the unusual case of rice Os7BGlu26 β-glycosidase which hydrolyzes 4-nitrophenyl β-glucoside and β-mannoside with similar efficiencies [52]. This is an especially unusual combination as β-glucosidases typically utilize ⁴H₃ transition states and β-mannosidases use either B_{2,5} or in rare cases (*vide infra*) ³H₄ transition states. X-ray structures of glucoimidazole (K₁ 2.7 nM) and mannoimidazole (K₁ 10.4 μM) revealed that the former binds in a ⁴E conformation, while the latter binds in a B_{2,5} conformation, suggestive of different transition state conformations for the two substrates. QM/MM simulations supported a ¹S₃ → ⁴E/⁴H₃[‡] → ⁴C₁ itinerary for β-glucosides and a ¹S₅ → B_{2,5}[‡] → ⁰S₂ itinerary for β-mannosides (**Figure 1b**). Thus, Os7BGlu26 hydrolyzes glucosides and mannosides through distinct pyranoside transition state conformations.

Mannosidases that operate through a ³H₄ half-chair conformation

α-Mannosidases of family GH47 are inverting and Ca²⁺-dependent metalloenzymes. The most conspicuous members of this family are the mammalian class I GH47 α-mannosidases involved in N-linked glycoprotein trimming in the secretory pathway in both normal biosynthetic processing and in endoplasmic reticulum associated degradation. A series of X-ray 'snapshots' of an S-linked substrate analogue, mannoimidazole, and the product mannose, combined with FEL analysis gave support for a ³S₁ → ³H₄[‡] → ¹C₄ conformational itinerary (**Figure 1c**) [53,54].

Family 47 α -mannosidases are uniquely sensitive to inhibition to the bicyclic and neutral inhibitor kifunensine, an observation that is exploited in cell biology [55] and in the industrial production of therapeutic proteins [56]. Early crystallographic work applied to mammalian ER α -mannosidase 1 showed that kifunensine adopts an inverted 1C_4 conformation bound to this enzyme [57]. In fact, QM/MM calculations show that kifunensine prefers the ring-flipped 1C_4 conformation over that of the transition state mimicking 3H_4 conformation, which is ~ 10 kcal mol $^{-1}$ higher in energy and does not correspond to an energy minimum. In the context of the conformational itinerary, this argues that kifunensine achieves its high potency by mimicking the product (rather than the transition state) of the reaction, suggesting that kifunensine achieves selectivity for GH47 mannosidases by its bias to a conformational state that is not accessed by α -mannosidases belonging to other families [58].

β -Mannanases of family GH134 are inverting glycoside hydrolases that have the capacity to degrade various β -mannans, including unprecedented activity on a challenging substrate, the regular unadorned β -1,4-mannan from the ivory nut [59]. An X-ray structure of mannopentaose complexed with an inactive mutant bacterial GH134 enzyme revealed that the sugar at the -1 subsite adopts a 1C_4 conformation [60]. QM/MM analysis of the reaction coordinate provided evidence for a ${}^1C_4 \rightarrow {}^3H_4^\ddagger \rightarrow {}^3S_1$ conformational itinerary (**Figure 1d**). Thus, the GH47 α -mannosidases and the GH134 β -mannanases use the same transition state in equivalent but reversed conformational itineraries for catalysis.

Leishmania β -1,2-mannosyltransferases/phosphorylases: a missing link between GHs and GTs?

Leishmania parasites accumulate a soluble intracellular β -1,2-mannan (mannogen) that functions as a carbohydrate reserve material and confers stability to stress. McConville and co-workers identified a family of enzymes termed β -1,2-mannosyltransferases/phosphorylases (MPTs) that can catalyze the synthesis and degradation of mannogen [61]. Five enzymes possess reversible phosphorylic activity, catalysing the stereochemically inverting release of Man-1-phosphate from mannogen (and elongation of mannogen from Man-1-P); two other homologous enzymes catalyzed the reversible extension of mannogen from GDPMan, with the difference in activities associated with an His/Arg switch at a key site within the extended active site cleft that appears to promote phosphate or GDP binding, respectively. The sugar nucleotide transfer activity of this group of enzymes defines them as glycosyltransferases, leading to their assignment to a new family, GT108. However, their ability to act as phosphorylases hints at a connection with the GHs, which in most cases are classified within GH families. Strikingly, X-ray structures of several representatives revealed that they do not adopt one of the typical GT folds, but rather a β -propeller (as predicted for GT91 enzymes), with remarkable similarity of active site with GH130 enzymes that possess β -1,2-mannosidase/phosphorylase activity, including a similar arrangement of catalytic residues. In fact, like enzymes of GH130, a substrate complex with β -1,2-mannobiose was observed with the -1 sugar residue bound in a 1S_5 conformation, and a conformational itinerary ${}^1S_5 \rightarrow B_{2,5}^\ddagger \rightarrow {}^0S_2$ was proposed, matching that of GH130 β -mannosidases (**Figure 4**).

GH family 99 endomannosidase/endomannanases: a new mechanism of glycosidic bond cleavage

Glycoside hydrolase family 99 contains retaining enzymes with two closely related activities: *endo*- α -1,2-mannosidases and *endo*- α -1,2-mannanases, which act on corresponding structures present within mammalian glucosylated high mannose N-glycans [62] and yeast mannans [63,64], respectively. An early set of ligand-bound 3-D structures of α -glucosyl-1,3-deoxymannojirimycin and α -glucosyl-1,3-isofagomine complexed to a *Bacteroides* sp. GH family 99 enzyme failed to identify a candidate enzymatic nucleophile, and also did not display any distortion away from a 'normal' 4C_1 conformation [65]. Instead, a conserved carboxylate residue was observed involved in a hydrogen bond with the 2-OH of GlcDMJ and a related inhibitor α -mannosyl-1,3-neoeuromycin [66]. These observations gave indirect support to the suggestion that this family may use a neighboring group participation mechanism involving the 2-OH group via a 1,2-anhydro sugar intermediate [65].

A comprehensive study of a *Bacteroides* sp. *endo*- α -1,2-mannanase yielded kinetic isotope effects for C1- ^2H (1.123 ± 0.012), anomeric- ^{13}C (1.030 ± 0.005) and C2- ^{18}O (1.052 ± 0.006) (**Figure 5**) [67]. These values are strikingly similar to those measured for the base-promoted solvolysis of PNPMan, and directly support neighboring group participation by the 2-hydroxyl group. A series of X-ray structures including a tetrasaccharide bound in a Michaelis complex, with cyclohexane-1,2-aziridine and cyclohexane-1,2-epoxide (as mimics of the 1,2-anhydro sugar intermediate) and with mannobiose in a product complex provided insight into the conformations of species along the reaction pathway. These structures were used in QM/MM calculations to model the reaction coordinate that also predicted reaction through a 1,2-anhydro sugar intermediate and a $^2E/{}^2H_3 \rightarrow [E_3]^\ddagger \rightarrow ^4E/{}^4H_5$ conformational itinerary. Interestingly, this transition state conformation is very close to the $^2H_3/E_3$ conformation observed upon binding of α -mannosyl-1,3-mannoimidazole (see Fig. 3) [68].

The unique features of this mechanism has inspired the design of activity based probes that seek to capitalize on the conserved base that deprotonates O2. Schroder synthesized spiro-epoxides that were designed to react with this residue to form a covalent link to the enzyme [69].

Through the looking glass: L-rhamnosidases

Rhamnosidases act on L-rhamnosides, which are the pseudo-enantiomer of D-mannosides. α -L-rhamnoside linkages occur widely in various pectins and the associated degradative enzymes occur within plants and in bacteria in the gut microbiota. All rhamnosidases studied to date process α -L-rhamnosides. Like α -mannosidases, α -rhamnosidases follow inverting and retaining mechanisms, and the rhamnosidases of GH106 are Ca^{2+} -dependent metalloenzymes (Table 1). A complex of a rhamnogalacturonan II degrading bacteroides GH106 α -L-rhamnosidase with L-rhamnotetrazole showed coordination of Ca^{2+} with the 2- and 3-hydroxyl groups and adopted a $^{2,5}B$ conformation [70], suggesting that enzymes of this family follow a $^2S_0 \rightarrow ^{2,5}B^\ddagger \rightarrow ^5S_1$ conformational itinerary, which is enantiomeric to the $B_{2,5}^\ddagger$ pathway followed by most α -mannosidases. Much less is known about the conformational itinerary of all other α -rhamnosidases. Complexes of a GH90 bacteriophage P22 tailspike protein *endo*-rhamnosidase from family GH90 (PDB 1TYU) and *Streptomyces avermitilis* α -rhamnosidase from family GH78 (PDB 3W5N) with rhamnose are found in $^{2,5}B$ conformations hint at a similar conformational itinerary; however, without further study the implications of these observations are unclear. On the basis of a failure to identify a candidate nucleophile for the retaining α -rhamnosidases of GH145, a neighboring group participation mechanism, similar to that of GH99 *endo*- α -1,2-mannanases, was proposed for enzymes of this family [71].

Conclusions

Extensive efforts over the last 20 years combining X-ray crystallography of ligands, kinetic studies and increasingly high-level *ab initio* simulations to the problem of understanding glycoside hydrolase catalysis across the bounty of nature's treasure chest. There is now good evidence that almost all mannosidases follow just two main conformational itineraries involving either a $B_{2,5}^\ddagger$ or $^3H_4^\ddagger$ conformation of the transition state. Thus certain α -mannosidases (GH families 38, 76, 125) use an $^0S_2 \rightarrow B_{2,5}^\ddagger \rightarrow ^1S_5$ itinerary and certain β -mannosidases (GH families 1, 2, 26, 92, 113) and the GH-like GT108 MTPs use a reversed $^1S_5 \rightarrow B_{2,5}^\ddagger \rightarrow ^0S_2$ itinerary. The first itinerary, $^2S_0 \rightarrow ^{2,5}B^\ddagger \rightarrow ^5S_1$, has also been assigned to GH106 α -rhamnosidases. Separately, GH family 47 α -mannosidases use a $^3S_1 \rightarrow ^3H_4^\ddagger \rightarrow ^1C_4$ itinerary and GH134 β -mannosidases use the reversed $^1C_4 \rightarrow ^3H_4^\ddagger \rightarrow ^3S_1$ itinerary. This situation is reminiscent of the use of symmetry related $^1S_3 \leftrightarrow ^4H_3^\ddagger \leftrightarrow ^4C_1$ itineraries of α - and β -glucosidases [72].

Nature has finally revealed its closely held secret that the GH99 α -mannosidases operate through a fundamentally distinct mechanism involving neighboring group participation [67]. This work demonstrates that nature can exploit long-established solution phase chemistry to cleave α -mannosides.

Unlike classical retaining and inverting mechanisms, neighboring group participation by the 2-hydroxy group can be considered a specialized rather than a general mechanism that is intrinsically limited to 1,2-trans related substrates in which the two groups can readily adopt an antiperiplanar arrangement. Thus, it is not applicable for cleavage of β -mannosides or 2-substituted α -mannosides, and is unlikely to be used for β -glucosides or β -galactosides, which require an energetically costly ring flip to the all (or mostly) axial 1C_4 conformation. Possibly, the mechanism may be found to be utilized by α -rhamnosides, which can readily achieve the necessary stereoelectronic arrangement.

The enzymatic control over conformational pathways identified for mannosidases used by nature shares analogy with discoveries made in synthetic chemistry on the effects of bridging protecting groups on controlling access to the flattened oxocarbenium ions invoked in solution phase glycosylation chemistry. The Crich laboratory in particular has utilized 4,6-benzylidene acetals to reduce the stability of mannosyl cations, thereby favouring covalent α -triflates and allowing access to prized β -mannosides [73]. On the other hand, bridging 2,3-*O*-carbonate are argued to (pre)distort the ring and provide easier access to a mannosyl cation and thereby deliver α -mannosides [73]. This latter example presents a clear analogy to one proposed role of O2-O3 coordination by divalent metals in α -mannosidase/ α -rhamnosidase metalloenzymes. It is to be hoped that continued investigations into nature's solutions to the problems of biological glycosyl transfer will inspire new solutions to the age-old problem of stereochemical control in chemical glycosylation reactions.

Conflicts of interest

None declared.

Acknowledgements

The Australian Research Council (ARC), the Spanish Ministry of Science, Innovation and Universities (MICINN) and the Agency for Management of University and Research Grants (AGAUR) are thanked for financial support. GJD is supported by the Royal Society through the Ken Murray Research Professorship. We thank Alba Nin-Hill for technical assistance with one of the figures of the manuscript.

References

Papers of particular interest, published within the period of review, have been highlighted as:

● of special interest

●● of outstanding interest

1. Yamabhai M, Sak-Ubol S, Srila W, Haltrich D: Mannan biotechnology: from biofuels to health. *Crit. Rev. Biotechnol.* 2016, 36:32-42.
2. Free SJ: Fungal cell wall organization and biosynthesis. *Adv. Gen.* 2013, 81:33-82.
3. Cao B, Williams SJ: Chemical approaches for the study of the mycobacterial glycolipids phosphatidylinositol mannosides, lipomannan and lipoarabinomannan. *Nat. Prod. Rep.* 2010, 27:919-947.
4. Ralton JE, Naderer T, Piraino HL, Bashtannyk TA, Callaghan JM, McConville MJ: Evidence that Intracellular β -1,2-Mannan is a Virulence Factor in Leishmania Parasites. *J. Biol. Chem.* 2003, 278:40757-40763.
5. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014, 42:D490-495.
6. Ten years of CAZyPedia: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology* 2018, 28:3-8.
7. Adero PO, Amarasekara H, Wen P, Bohé L, Crich D: The Experimental Evidence in Support of Glycosylation Mechanisms at the SN1–SN2 Interface. *Chem. Rev.* 2018, 118:8242-8284.
- A major review discussing the evidence for oxocarbenium ions in glycosylation reactions, the factors that influence their formation, and the types of reactive intermediates formed in chemical glycosylation reactions.
8. Chan J, Sannikova N, Tang A, Bennet AJ: Transition-State Structure for the Quintessential SN2 Reaction of a Carbohydrate: Reaction of α -Glucopyranosyl Fluoride with Azide Ion in Water. *J. Am. Chem. Soc.* 2014, 136:12225-12228.

9. Martin A, Arda A, Désiré J, Martin-Mingot A, Probst N, Sinaÿ P, Jiménez-Barbero J, Thibaudeau S, Blériot Y: Catching elusive glycosyl cations in a condensed phase with HF/SbF₅ superacid. *Nat. Chem.* 2015, 8:186.
10. Hansen T, Lebedel L, Remmerswaal WA, van der Vorm S, Wander DPA, Somers M, Overkleeft HS, Filippov DV, Desire J, Mingot A, et al.: Defining the S_N1 Side of Glycosylation Reactions: Stereoselectivity of Glycopyranosyl Cations. *ACS Cent. Sci.* 2019, 5:781-788.
11. More O'Ferrall RA: Relationships between E₂ and E₁cB mechanisms of β-elimination. *J. Chem. Soc. B* 1970:274-277.
12. Jencks WP: General acid-base catalysis of complex reactions in water. *Chem. Rev.* 1972, 72:705-718.
13. Biarnés X, Nieto J, Planas A, Rovira C: Substrate Distortion in the Michaelis Complex of Bacillus 1,3–1,4-β-Glucanase: Insight from first principles molecular dynamics simulations. *J. Biol. Chem.* 2006, 281:1432-1441.
14. Sinnott ML: Catalytic mechanisms of enzymatic glycosyl transfer. *Chem. Rev.* 1990, 90:1171-1202.
15. Schramm VL, Shi W: Atomic motion in enzymatic reaction coordinates. *Curr. Opin. Struct. Biol.* 2001, 11:657-665.
16. Speciale G, Thompson AJ, Davies GJ, Williams SJ: Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* 2014, 28:1-13.
17. Conformational nomenclature for five- and six-membered ring forms of monosaccharides and their derivatives. *Pure Appl. Chem.* 1981, 53:1901.
18. Zechel DL, Withers SG: Glycosidase mechanisms: Anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 2000, 33:11-18.
19. Voadlo DJ, Davies GJ: Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* 2008, 12:539-555.
20. Coines J, Alfonso-Prieto M, Biarnés X, Planas A, Rovira C: Oxazoline or Oxazolinium Ion? The Protonation State and Conformation of the Reaction Intermediate of Chitinase Enzymes Revisited. *Chem. Eur. J.* 2018, 24:19258-19265.
- This insightful computational study provides clear evidence that neighboring group participation of a GH18 chitinase proceeds via a neutral oxazoline intermediate.
21. Cekic N, Heinonen JE, Stubbs KA, Roth C, He Y, Bennet AJ, McEachern EJ, Davies GJ, Voadlo DJ: Analysis of transition state mimicry by tight binding aminothiazoline inhibitors provides insight into catalysis by human O-GlcNAcase. *Chem. Sci.* 2016, 7:3742-3750.
22. Williams SJ, Mark B, Voadlo DJ, James MN, Withers SG: Aspartate 313 in the *Streptomyces plicatus* hexosaminidase plays a critical role in substrate assisted catalysis by orienting the 2-acetamido group and stabilizing the transition state. *J. Biol. Chem.* 2002, 277:40055-40065.
23. Roth C, Petricevic M, John A, Goddard-Borger ED, Davies GJ, Williams SJ: Structural and mechanistic insights into a *Bacteroides vulgatus* retaining N-acetyl-β-galactosaminidase that uses neighbouring group participation. *Chem. Commun.* 2016, 52:11096-11099.
24. Hughes ED: Mechanism and kinetics of substitution at a saturated carbon atom. *Trans. Faraday Soc.* 1941, 37:603-631.
25. Cremer D, Pople JA: General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 1975, 97:1354-1358.
26. Snaith SM, Levvy GA: α-Mannosidase as a Zinc-dependent Enzyme. *Nature* 1968, 218:91-92.
27. Nielsen JW, Poulsen NR, Johnsson A, Winther JR, Stipp SL, Willemoës M: Metal-ion dependent catalytic properties of *Sulfolobus solfataricus* class II α-mannosidase. *Biochemistry* 2012, 51:8039-8046.
28. Zhu Y, Suits MDL, Thompson AJ, Chavan S, Dinev Z, Dumon C, Smith N, Moremen K, Xiang Y, Siriwardena A, et al.: Mechanistic insights into a calcium-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* 2010, 6:125-132.
29. Petersen L, Ardevol A, Rovira C, Reilly PJ: Molecular mechanism of the glycosylation step catalyzed by Golgi α-mannosidase II: a QM/MM metadynamics investigation. *J. Am. Chem. Soc.* 2010, 132:8291-8300.
30. Tanret M: Sur la piceine, glucoside des feuilles du sapin epicea. *Bull. Soc. Chim. Fr.* 1894, 211:944.
31. Gasman RC, Johnson DC: C-2 Oxyanion Participation in the Base-Catalyzed Cleavage of p-Nitrophenyl β-D-Galactopyranoside and p-Nitrophenyl α-D-Mannopyranoside. *J. Org. Chem.* 1966, 31:1830-1838.
32. Speciale G, Farren-Dai M, Shidmoosavee FS, Williams SJ, Bennet AJ: C2-Oxyanion neighboring group participation: Transition state structure for the hydroxide-promoted hydrolysis of 4-nitrophenyl α-D-mannopyranoside. *J. Am. Chem. Soc.* 2016, 138:14012-14019.
33. Brockhaus M, Dettinger H-M, Kurz G, Lehmann J, Wallenfels K: Participation of HO-2 in the cleavage of β-D-galactosides by the β-D-galactosidase from *E. coli*. *Carbohydr. Res.* 1979, 69:264-268.
34. Wallenfels K, Weil R: In *The Enzymes*, edn 3rd Edition. Edited by Boxer PD: Academic Press; 1972:617-663. vol 7.]

35. Egea PF, Muller-Steffner H, Kuhn I, Cakir-Kiefer C, Oppenheimer NJ, Stroud RM, Kellenberger E, Schuber F: Insights into the Mechanism of Bovine CD38/NAD⁺ Glycohydrolase from the X-Ray Structures of Its Michaelis Complex and Covalently-Trapped Intermediates. *PLOS One* 2012, 7:e34918.
36. Handlon AL, Xu C, Muller-Steffner HM, Schuber F, Oppenheimer NJ: 2'-Ribose Substituent Effects on the Chemical and Enzymic Hydrolysis of NAD⁺. *J. Am. Chem. Soc.* 1994, 116:12087-12088.
37. Johnson RW, Marschner TM, Oppenheimer NJ: Pyridine nucleotide chemistry. A new mechanism for the hydroxide-catalyzed hydrolysis of the nicotinamide-glycosyl bond. *J. Am. Chem. Soc.* 1988, 110:2257-2263.
38. Wicki J, Williams SJ, Withers SG: Transition-state mimicry by glycosidase inhibitors: a critical kinetic analysis. *J. Am. Chem. Soc.* 2007, 129:4530-4531.
39. Tailford LE, Offen WA, Smith NL, Dumon C, Morland C, Gratien J, Heck MP, Stick RV, Bleriot Y, Vasella A, et al.: Structural and biochemical evidence for a boat-like transition state in β -mannosidases. *Nat. Chem. Biol.* 2008, 4:306-312.
40. Heightman TD, Vasella AT: Recent insights into inhibition, structure and mechanism of configuration retaining glycosidases. *Angew. Chem. Int. Ed.* 1999, 38:750-770.
41. Williams RJ, Iglesias-Fernandez J, Stepper J, Jackson A, Thompson AJ, Lowe EC, White JM, Gilbert HJ, Rovira C, Davies GJ, et al.: Combined inhibitor free-energy landscape and structural analysis reports on the mannosidase conformational coordinate. *Angew. Chem. Int. Ed.* 2014, 53:1087-1091.
42. Ducros VM, Zechel DL, Murshudov GN, Gilbert HJ, Szabo L, Stoll D, Withers SG, Davies GJ: Substrate distortion by a β -mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state. *Angew. Chem. Int. Ed.* 2002, 41:2824-2827.
43. Hine J: The Principle of Least Nuclear Motion. In *Adv. Phys. Org. Chem.* Edited by Gold V, Bethel D: Academic Press; 1978:1-61. vol 15.
44. Sinnott ML: The principle of least nuclear motion and the theory of stereoelectronic control. *Adv. Phys. Org. Chem.* 1988, 24:113-204.
45. Thompson AJ, Speciale G, Iglesias-Fernandez J, Hakki Z, Belz T, Cartmell A, Spears RJ, Chandler E, Temple MJ, Stepper J, et al.: Evidence for a boat conformation at the transition state of GH76 α -1,6-mannanases-key enzymes in bacterial and fungal mannoprotein metabolism. *Angew. Chem. Int. Ed.* 2015, 54:5378-5382.
46. Gregg KJ, Zandberg WF, Hehemann JH, Whitworth GE, Deng L, Vocadlo DJ, Boraston AB: Analysis of a new family of widely distributed metal-independent α -mannosidases provides unique insight into the processing of N-linked glycans. *J. Biol. Chem.* 2011, 286:15586-15596.
47. Alonso-Gil S, Males A, Fernandes PZ, Williams SJ, Davies GJ, Rovira C: Computational design of experiment unveils the conformational reaction coordinate of GH125 α -mannosidases. *J. Am. Chem. Soc.* 2017, 139:1085-1088.
- Computational QM/MM metadynamics showed that sulfur-for-oxygen substitution in the substrate unfavourably perturbs the energetically-accessible conformational space on-enzyme, allowing design of a crystallographic experiment to access a mechanistically-relevant Michaelis complex.
48. Males A, Speciale G, Williams SJ, Davies GJ: Distortion of mannoimidazole supports a B2,5 boat transition state for the family GH125 α -1,6-mannosidase from *Clostridium perfringens*. *Org. Biomol. Chem.* 2019, 17:7863-7869.
49. Cuskin F, Basle A, Ladeveze S, Day AM, Gilbert HJ, Davies GJ, Potocki-Veronese G, Lowe EC: The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β -1,2-Mannosidic Linkages in *Candida Mannan*. *J. Biol. Chem.* 2015, 290:25023-25033.
50. Nakae S, Ito S, Higa M, Senoura T, Wasaki J, Hijikata A, Shionyu M, Ito S, Shirai T: Structure of novel enzyme in mannan biodegradation process 4-O- β -D-mannosyl-D-glucose phosphorylase MGP. *J. Mol. Biol.* 2013, 425:4468-4478.
51. Tsuda T, Nihira T, Chiku K, Suzuki E, Arakawa T, Nishimoto M, Kitaoka M, Nakai H, Fushinobu S: Characterization and crystal structure determination of β -1,2-mannobiose phosphorylase from *Listeria innocua*. *FEBS Lett.* 2015, 589:3816-3821.
52. Tankrathok A, Iglesias-Fernández J, Williams RJ, Pengthaisong S, Baiya S, Hakki Z, Robinson RC, Hrmova M, Rovira C, Williams SJ, et al.: A single glycosidase harnesses different pyranoside ring transition state conformations for hydrolysis of mannosides and glucosides. *ACS Catalysis* 2015, 5:6041-6051.
53. Karaveg K, Siriwardena A, Tempel W, Liu Z-J, Glushka J, Wang B-C, Moremen KW: Mechanism of Class 1 (Glycosylhydrolase Family 47) α -Mannosidases Involved in N-Glycan Processing and Endoplasmic Reticulum Quality Control. *J. Biol. Chem.* 2005, 280:16197-16207.
54. Thompson AJ, Dabin J, Iglesias-Fernandez J, Ardevol A, Dinev Z, Williams SJ, Bande O, Siriwardena A, Moreland C, Hu TC, et al.: The reaction coordinate of a bacterial GH47 α -mannosidase: A combined quantum mechanical and structural approach. *Angew. Chem. Int. Ed.* 2012, 51:10997-11001.
55. Elbein AD, Tropea JE, Mitchell M, Kaushal GP: Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* 1990, 265:15599-15605.

56. Cox TM: Competing for the treasure in exceptions. *Am. J. Hematol.* 2013, 88:163-165.
57. Vallée F, Karaveg K, Herscovics A, Moremen KW, Howell PL: Structural basis for catalysis and inhibition of N-glycan processing class I α -1,2-mannosidases. *J. Biol. Chem.* 2000, 275:41287-41298.
58. Males A, Raich L, Williams SJ, Rovira C, Davies GJ: Conformational analysis of the mannosidase inhibitor kifunensine: a quantum mechanical and structural approach. *ChemBiochem* 2017, 18:1496-1501.
- This study uses computational methods to define the free energy landscape and show that kifunensine achieves selectivity for GH47 mannosidases by mimicry of the product conformation.
59. Shimizu M, Kaneko Y, Ishihara S, Mochizuki M, Sakai K, Yamada M, Murata S, Itoh E, Yamamoto T, Sugimura Y, et al.: Novel β -1,4-Mannanase Belonging to a New Glycoside Hydrolase Family in *Aspergillus nidulans*. *J. Biol. Chem.* 2015, 290:27914-27927.
60. Jin Y, Petricevic M, John A, Raich L, Jenkins H, Portela De Souza L, Cuskin F, Gilbert HJ, Rovira C, Goddard-Borger ED, et al.: A β -mannanase with a lysozyme-like fold and a novel molecular catalytic mechanism. *ACS Cent. Sci.* 2016, 2:896-903.
61. Sernee MF, Ralton JE, Nero TL, Sobala LF, Kloehn J, Vieira-Lara MA, Cobbold SA, Stanton L, Pires DEV, Hanssen E, et al.: A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in *Leishmania* Parasites. *Cell Host Microbe* 2019, 26:385-399.e389.
- Discovery of a new family of glycosyltransferases (GT108) that shares a fold and mechanism with GH130 β -mannoside phosphorylases.
62. Lubas WA, Spiro RG: Golgi endo- α -D-mannosidase from rat liver, a novel N-linked carbohydrate unit processing enzyme. *J. Biol. Chem.* 1987, 262:3775-3781.
63. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, et al.: Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 2015, 517:165-169.
64. Hakki Z, Thompson AJ, Bellmaine S, Speciale G, Davies GJ, Williams SJ: Structural and kinetic dissection of the endo- α -1,2-mannanase activity of bacterial GH99 glycoside hydrolases from *Bacteroides* spp. *Chem. Eur. J.* 2015, 21:1966-1977.
65. Thompson AJ, Williams RJ, Hakki Z, Alonzi DS, Wennekes T, Gloster TM, Songsrirote K, Thomas-Oates JE, Wrodnigg TM, Spreitz J, et al.: Structural and mechanistic insight into N-glycan processing by endo- α -mannosidase. *Proc. Natl. Acad. Sci. USA* 2012, 109:781-786.
66. Petricevic M, Sobala LF, Fernandes P, Raich L, Thompson AJ, Bernardo-Seisdedos G, Millet O, Zhu S, Sollogoub M, Jimenez-Barbero J, et al.: Contribution of shape and charge to the inhibition of a family GH99 endo- α -1,2-mannanase. *J. Am. Chem. Soc.* 2017, 139:1089-1097.
67. Sobala LF, Speciale G, Zhu S, Raich L, Sannikova N, Thompson AJ, Hakki Z, Lu D, Shamsi Kazem Abadi S, Lewis AR, et al.: An Epoxide Intermediate in Glycosidase Catalysis. *ChemRxiv* 2019, 10.26434/chemrxiv.9745388.v1.
- This pre-print is a tour de force of structural, mechanistic and computational chemistry and reveals a new mechanism in glycosidase catalysis involving neighboring group participation by the substrate 2-hydroxyl group.
68. Fernandes PZ, Petricevic M, Sobala L, Davies GJ, Williams SJ: Exploration of Strategies for Mechanism-Based Inhibitor Design for Family GH99 endo- α -1,2-Mannanases. *Chem. Eur. J.* 2018, 24:7464-7473.
69. Schroder SP, Kallemeijn WW, Debets MF, Hansen T, Sobala LF, Hakki Z, Williams SJ, Beenakker TJM, Aerts J, van der Marel GA, et al.: Spiro-epoxyglycosides as Activity-Based Probes for Glycoside Hydrolase Family 99 Endomannosidase/Endomannanase. *Chem. Eur. J.* 2018, 24:9983-9992.
- The work details the rational design of 2-spiro epoxides as mechanism-based inactivators and affinity-based activity probes of family GH99 enzymes.
70. Ndeh D, Rogowski A, Cartmell A, Luis AS, Basle A, Gray J, Venditto I, Briggs J, Zhang X, Labourel A, et al.: Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 2017, 544:65-70.
- This in-depth study of rhamnogalacturonan metabolism by a gut bacterium led to discovery of 7 new glycoside hydrolase families, and showed the first definitive transition state conformation for an α -rhamnosidase.
71. Munoz-Munoz J, Cartmell A, Terrapon N, Henrissat B, Gilbert HJ: Unusual active site location and catalytic apparatus in a glycoside hydrolase family. *Proc. Natl. Acad. Sci. USA* 2017, 114:4936-4941.
72. Davies GJ, Planas A, Rovira C: Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* 2012, 45:308-316.
73. Crich D: Mechanism of a Chemical Glycosylation Reaction. *Acc. Chem. Res.* 2010, 43:1144-1153.
74. Vincent F, Gloster TM, Macdonald J, Morland C, Stick RV, Dias FMV, Prates JAM, Fontes CMGA, Gilbert HJ, Davies GJ: Common Inhibition of Both β -Glucosidases and β -Mannosidases by Isofagomine Lactam Reflects Different Conformational Itineraries for Pyranoside Hydrolysis. *ChemBioChem* 2004, 5:1596-1599.
75. Helbert W, Poulet L, Drouillard S, Mathieu S, Loiodice M, Couturier M, Lombard V, Terrapon N, Turchetto J, Vincetelli R, et al.: Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proc. Natl. Acad. Sci. USA* 2019, 116:6063-6068.

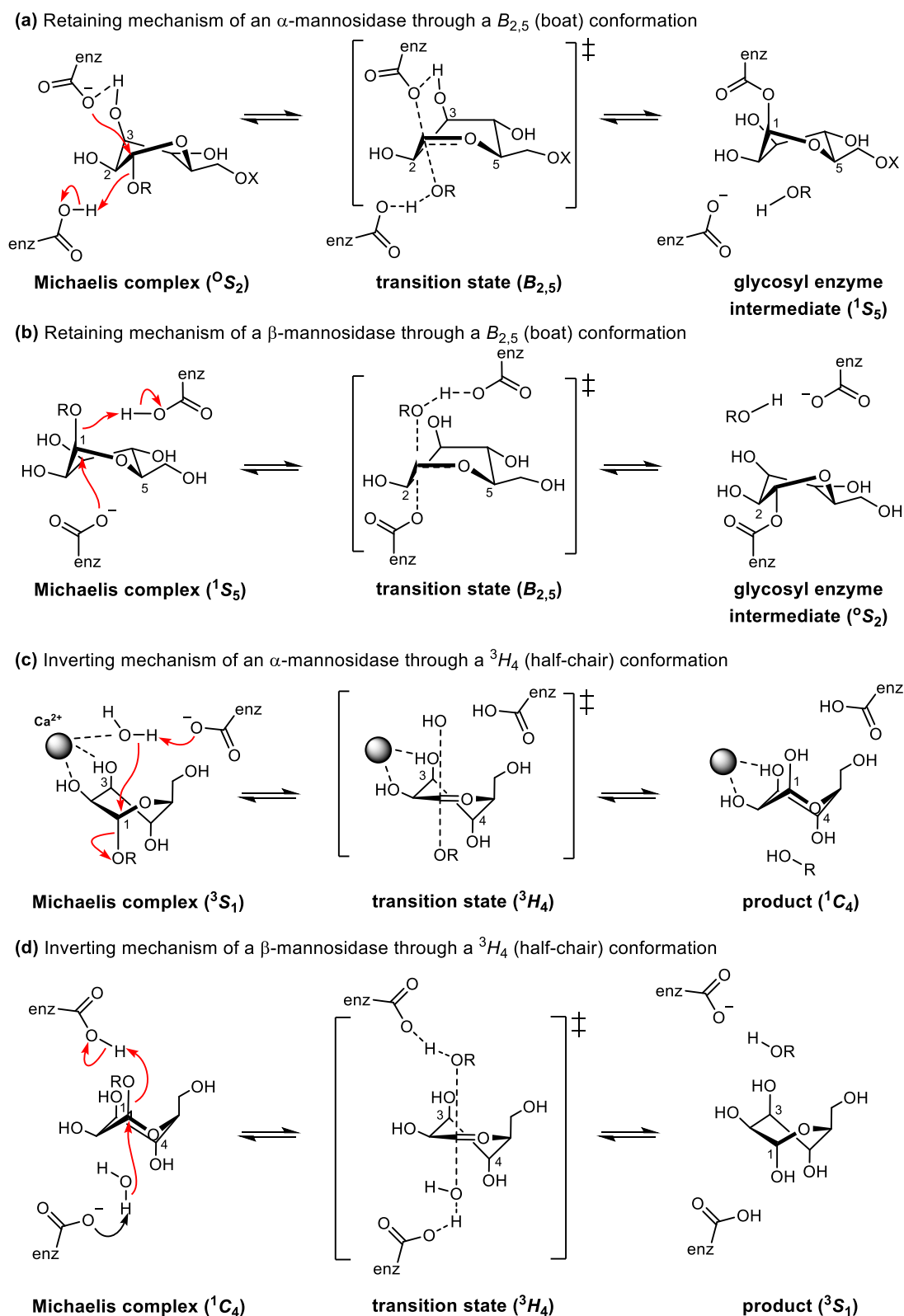
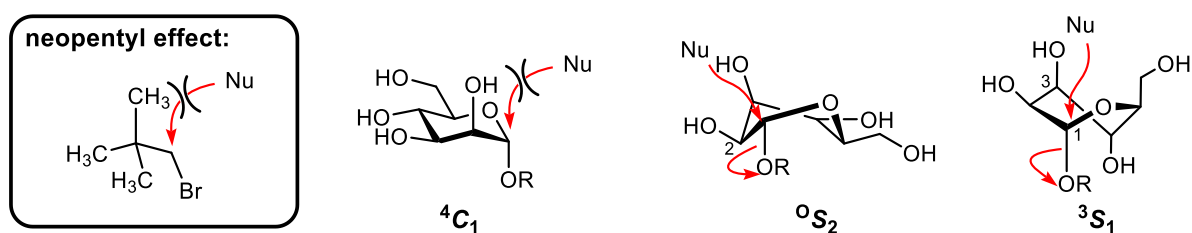
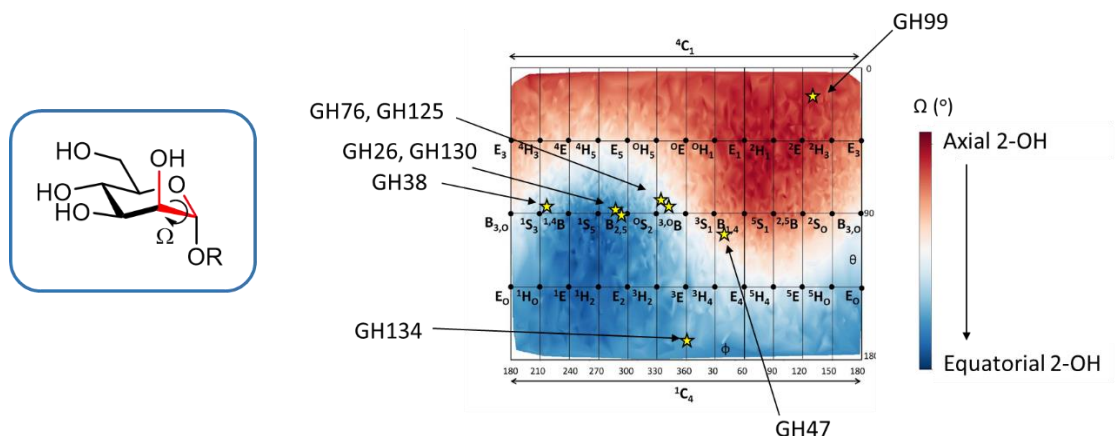


Figure 1. Mechanisms for (a,b) glycosylation half-reaction of retaining α -mannosidases (GH38, 76) and β -mannosidases (GH1, 2, 5, 26, 113) that proceed through a $B_{2,5}$ transition state conformation (GH38 retaining α -mannosidases and GH92 inverting α -mannosidases follow similar mechanisms but are metal dependent), and (c,d) inverting α -mannosidases (GH47) and β -mannosidases (GH134) that proceed through a 3H_4 transition state conformation.

(a) Blocked nucleophilic trajectories in mannose can be alleviated by conformational distortion



(b) O2-Axiality plot for mannose across the Cremer-Pople coordinates



(c) Ligands

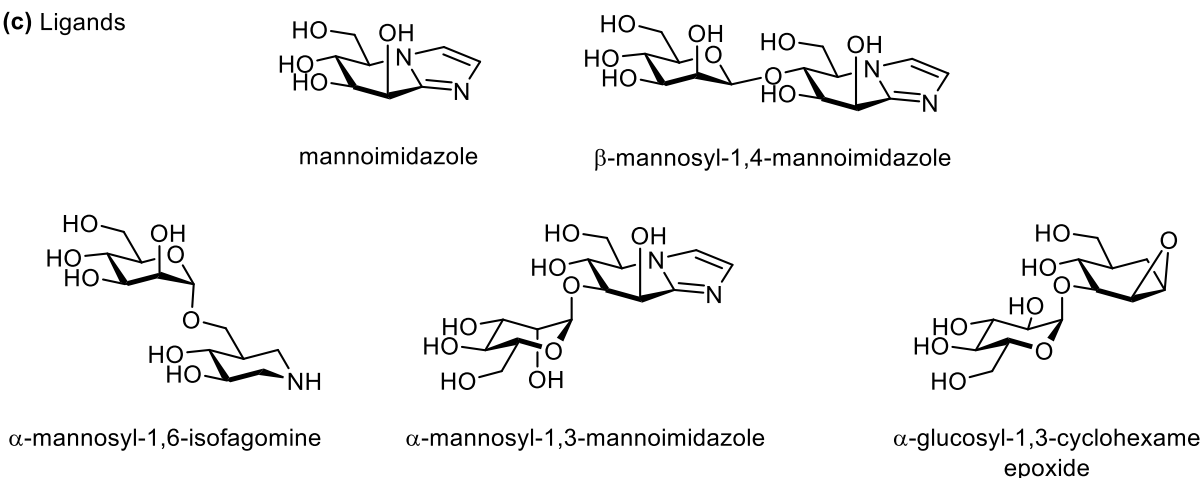


Figure 2. (a) Nucleophilic attack on an α -mannoside in a 4C_1 conformation suffers from a 1,2-diaxial interaction that interrupts the nucleophilic trajectory, analogous to the neopentyl effect. This effect can be relieved by conformational distortion to an 0S_2 or 4H_3 conformation. (b) Plot showing O2 axial/equatorial percentage across the Cremer-Pople conformational landscape, and with the conformation of the individual family representatives from PDB files of X-ray structures of Michaelis or intermediate complexes indicated with stars (GH26 1GVY; GH38 1QX1; GH47 4AYP; GH76 5AGD; GH125 5M7Y; GH130 5B0R; GH134 5JUG). (c) Ligands discussed in this article.

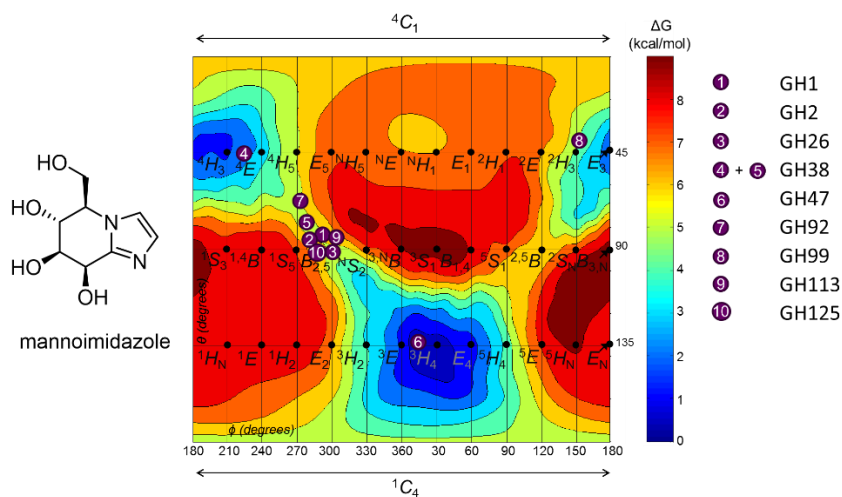
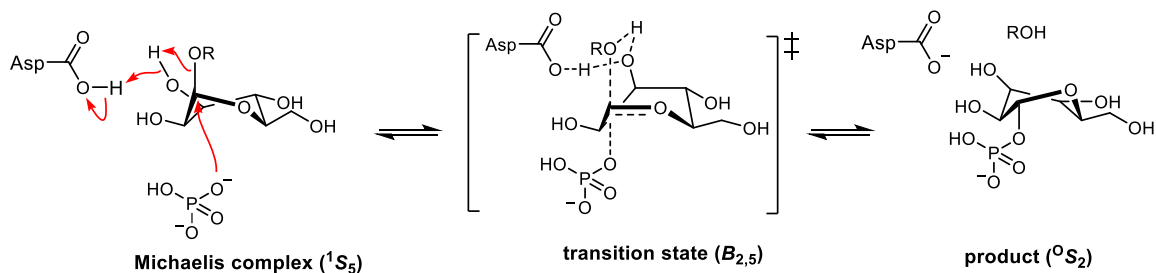
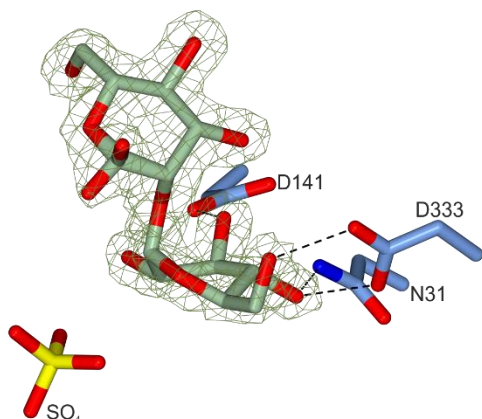


Figure 3. Free energy landscape for mannoimidazole, with coordinates for the observed conformation of mannoimidazole-type ligands in the -1 subsite for enzymes from families (1) GH1 (PDB 4RE2), (2) GH2 (PDB 2VMF), (3) GH26 (PDB 4CD5), (4)+(5) GH38 (PDB 3D4Y), (6) GH47 (PDB 4AYQ), (7) GH92 (PDB 2WZS), (8) GH99 (PDB 6FAR), (9) GH113 (PDB 4CD8), (10) GH125 (PDB 6RQK).

(a) Proposed inverting mechanism of GH130 β -mannoside phosphorylase and GT108 β -mannosyltransferase/phosphorylase



(b) 3-D structure of GH130 bound to β -Man-1,4-Glc



(c) 3-D structure of GT108 bound to β -Man-1,2-Man

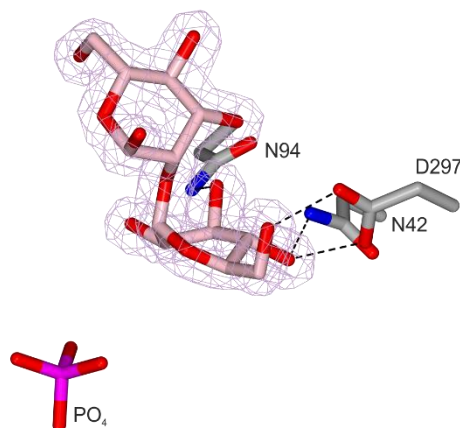
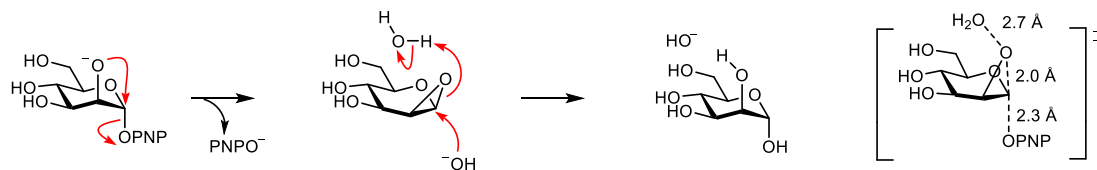
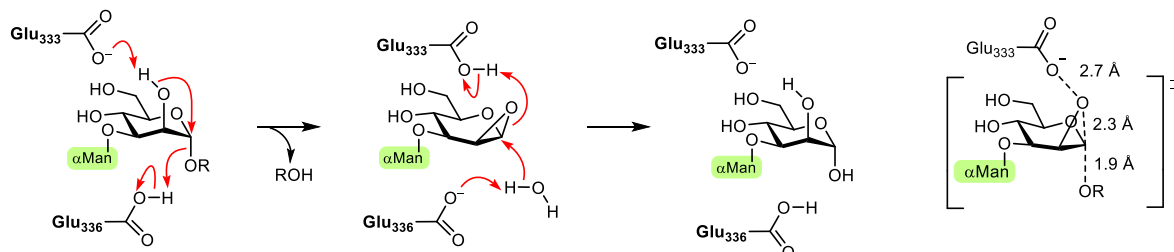


Figure 4. (a) Inverting mechanism proposed for GH130 β -mannoside phosphorylases and GT108 β -mannosyltransferase/phosphorylases, showing participation of the 3-hydroxyl. (b) 3-D structure of β -Man-1,2-Man (green) complex of *Listeria innocua* β -1,2-mannobiose phosphorylase (blue) showing substrate in a 1S_5 conformation and sulfate location (yellow). $2mF_o - F_c$ weighted electron density map (dark green) contoured at $1.0 e^-/\text{\AA}^3$. D141 is the putative general acid catalyst. (c) 3-D structure of β -Man-1,2-Man (pink) complex of *Leishmania mexicana* β -mannosyltransferase/phosphorylase MPT2 D94N mutant (grey) showing substrate in a 1S_5 conformation, superposed with the phosphate location (magenta) from the homologous MTP4 structure. $2mF_o - F_c$ weighted electron density map (light pink) contoured at $0.5 e^-/\text{\AA}^3$.

(a) C2-oxyanion participation and calculated transition state geometry



(b) Family GH99 C2-hydroxyl participation and calculated transition state geometry



(c) Experimental kinetic isotope effects

position	base promoted hydrolysis of PNPMan	
	<i>Br</i> GH99 catalyzed	
1- ² H	1.123 ± 0.012	1.112 ± 0.004
1- ¹³ C	1.052 ± 0.006	1.026 ± 0.006
2- ¹⁸ O	1.030 ± 0.005	1.044 ± 0.006

(d) X-ray structures along the reaction coordinate

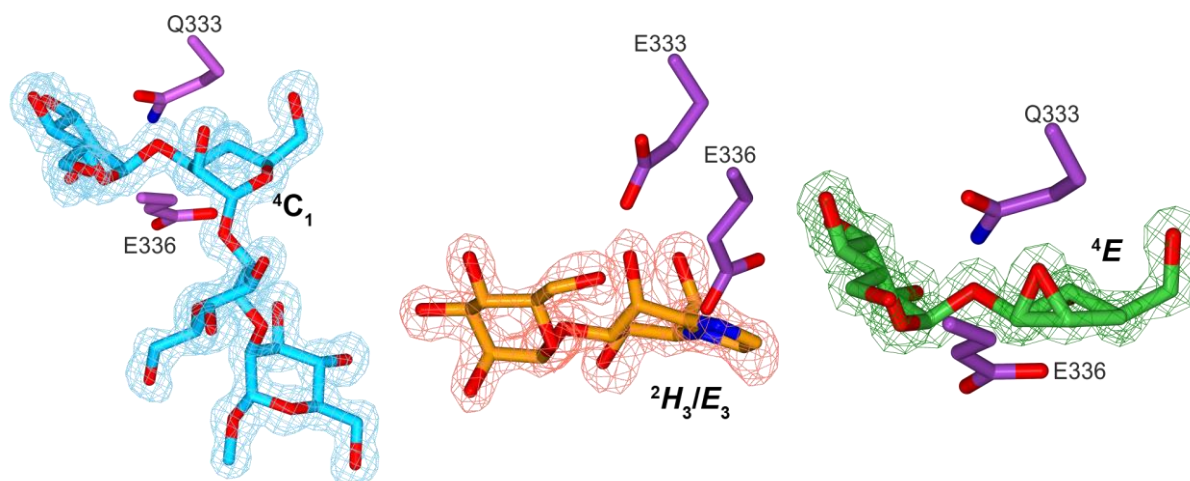


Figure 5. Neighboring group participation by the 2-hydroxyl of α -mannosides. (a) Mechanism for neighboring group participation by a C2-oxyanion in the base-mediated solvolysis of 4-nitrophenyl α -D-mannopyranoside (PNPMan). Right: ab initio calculated interatomic distances for the critical interactions at the solvolytic transition state [32]. (b) Mechanism for neighboring group participation by family GH99 *endo*- α -1,2-mannanase. Right: QM/MM calculation of interatomic distances for the critical interactions at the enzymatic transition state [67]. (c) Experimental kinetic isotope effects for the solvolytic [32] and enzymatic [67] processes indicate an oxocarbenium ion-like transition state with O2 participation. (d) Snapshots along the reaction coordinate: Michaelis complex with α -Man-1,3- α -Man-1,2- α -Man-1,2- α -Man-OMe [67], α -mannosyl-1,3-mannoimidazole [68] and α -glucosyl-1,3-cyclohexane- α -1,2-epoxide [67].

Abbreviations

α -Man	α -D-mannopyranosyl
4MU	4-methylumbelliferone
AcCoA	Acetyl coenzyme A
AD	Alzheimer's disease
<i>B. thetaiotaomicron</i>	<i>Bacteriodes thetaiotaomicron</i>
<i>C. perfringens</i>	<i>Clostridium perfringens</i>
CAZy/CAZyme	Carbohydrate active enzyme
CBA	Carbohydrate-binding agents
<i>Ck</i>	<i>Caulobacter</i> sp. K31 strain
CKII	Casein kinase II
DFT	Density Functional Theory
DMJ	1-Deoxymannojirimycin
DNP-Man	2,4-Dinitrophenyl α -D-mannopyranoside
<i>E. coli</i>	<i>Escherichia coli</i>
EDEM	ER-degradation enhancing α -mannosidase-like
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER	Endoplasmic reticulum
ERAD	Endoplasmic-reticulum-associated protein degradation
ERGIC	ER-Golgi intermediate compartment
ERManI	Endoplasmic reticulum α -mannosidase I
FEL	Free energy landscape
GalNAc	<i>N</i> -acetyl-D-galactosamine
GH#X	Glycoside hydrolase family #X
GlcNAc	<i>N</i> -acetyl- β -D-glucosamine
GNAT	GCN5-related <i>N</i> -acetyltransferase
ManIA/B/C	Golgi α -mannosidase IA/B/C
GT#X	Glycosyltransferase family #X
<i>H. sapiens</i>	<i>Homo sapiens</i>
(p)HAT	(pseudo) Histone acetyltransferase
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid
HIV	Human immunodeficiency virus
HPAEC-PAD	High-performance Anion Exchange Chromatography with pulsed amperometric detection
ITC	Isothermal titration calorimetry
IUPAC	International union of pure and applied Chemistry
K_A	Equilibrium association constant
k_{cat}	Turnover number
K_D	Equilibrium dissociation constant
K_i	Inhibitor constant
KIE	Kinetic isotope effect
K_M	Michaelis constant
LC	Long connection
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LFER	Linear (Gibbs) free energy relationship

M	Cytoplasmic membrane
Man	D-Mannnose
Man ₉ GlcNAc ₂	9 Mannose 2 <i>N</i> -acetyl glucosamine
MC	Michaelis complex
MD	Molecular dynamics
MIm	Mannoimidazole
MT	Microtubule
MTBR	Microtubule binding repeats
NAG-thiazoline	1,2-dideoxy-2'-methyl- α -D-glucofuranose[2,1- <i>d</i>]- Δ 2'-thiazoline
NButGT	1,2-Dideoxy-2'-propyl- α -d-glucofuranose-[2,1- <i>d</i>]- Δ 2'-thiazoline
NCOAT	Nuclear cytoplasmic O-GlcNAcase and acetyltransferase
NeuAc	Neuraminic acid
NFT	Neurofibrillary tangles
NMR	Nuclear magnetic resonance
OGA	O-GlcNAc hydrolase
OGT	O-GlcNAc transferase
PDB ID	Protein data bank identification
PEG	Polyethylene glycol
PG	Peptidoglycan
PHF	Paired helical filaments
PISA	Protein Interfaces, Surfaces and Assemblies
<i>p</i> NP-GlcNAc	4-Nitrophenyl <i>N</i> -acetyl- β -D-glucosaminide
PTM	Post-translational modification
PUGNAc	O-(2-acetamido-2-deoxy-d-glucofuranosylidene)amino- <i>N</i> -phenylcarbamate
QM/MM	Quantum mechanics / Molecular mechanics
RMSD	Root mean square deviation
RT	Room temperature
<i>S. cerevisiae/Sc</i>	<i>Saccharomyces cerevisiae</i>
<i>S. pneumoniae/Sp</i>	<i>Streptococcus pneumoniae</i>
SC	Short connection
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SER	Surface entropy reduction
STZ	Streptozotocin (STZ)
TIM	Triose phosphate isomerase
T _m	Melting temperature
TS/‡	Transition state
TSA	Thermal shift assay
UDP	Uridine diphosphate
UV	Ultra violet
WT	Wild type

References

- van Aalten, D.M., Komander, D., Synstad, B., Gåseidnes, S., Peter, M.G., and Eijssink, V.G. (2001). Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 8979–8984.
- Abdel Rahman, A.M., Ryczko, M., Pawling, J., and Dennis, J.W. (2013). Probing the Hexosamine Biosynthetic Pathway in Human Tumor Cells by Multitargeted Tandem Mass Spectrometry. *ACS Chem. Biol.* *8*, 2053–2062.
- Aebi, M., Gassenhuber, J., Domdey, H., and te Heesen, S. (1996). Cloning and characterization of the ALG3 gene of *Saccharomyces cerevisiae*. *Glycobiology* *6*, 439–444.
- Agirre, J., Davies, G., Wilson, K., and Cowtan, K. (2015a). Carbohydrate anomalies in the PDB. *Nat. Chem. Biol.* *11*, 303.
- Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G.J., Wilson, K.S., and Cowtan, K.D. (2015b). Privateer: Software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* *22*, 833–834.
- Aguilera, A., and Zimmermann, F.K. (1986). Isolation and molecular analysis of the phosphoglucose isomerase structural gene of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* *202*, 83–89.
- Alonso-Gil, S., Males, A., Fernandes, P.Z., Williams, S.J., Davies, G.J., and Rovira, C. (2017). Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α -Mannosidases. *J. Am. Chem. Soc.* *139*, 1085–1088.
- Alvarez-Dorta, D., King, D.T., Legigan, T., Ide, D., Adachi, I., Deniaud, D., Désiré, J., Kato, A., Vocadlo, D., Gouin, S.G., et al. (2017). Multivalency To Inhibit and Discriminate Hexosaminidases. *Chem. - A Eur. J.* *23*, 9022–9025.
- Angyal, S.J. (1969). Zusammensetzung und Konformation von Zuckern in Lösung. *Angew. Chemie* *81*, 172–182.
- Angyal, S.J. (1984). The Composition of Reducing Sugars in Solution. *Adv. Carbohydr. Chem. Biochem.* *42*, 15–68.
- Aoyagi, T., Suda, H., Uotani, K., Kojima, F., Aoyama, T., Horiguchi, K., Hamada, M., and Takeuchi, T. (1992). Nagstatin, a new inhibitor of N-acetyl-beta-D-glucosaminidase, produced by *Streptomyces amakusaensis* MG846-ff3. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J. Antibiot. (Tokyo)*. *45*, 1404–1408.
- Ardèvol, A., and Rovira, C. (2015). Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* *137*, 7528–7547.
- Arnold, C.S., Johnson, G.V.W., Cole, R.N., Dong, D.L.-Y., Lee, M., and Hart, G.W. (1996). The Microtubule-associated Protein Tau Is Extensively Modified with O-linked N-acetylglucosamine. *J. Biol. Chem.* *271*, 28741–28744.
- Balzarini, J. (2007a). Carbohydrate-binding agents: A potential future cornerstone for the chemotherapy of enveloped viruses? *Antivir. Chem. Chemother.* *18*, 1–11.
- Balzarini, J. (2007b). The $\alpha(1,2)$ -mannosidase I inhibitor 1-deoxymannojirimycin potentiates the antiviral activity of carbohydrate-binding agents against wild-type and mutant HIV-1 strains containing glycan deletions in gp120. *FEBS Lett.* *581*, 2060–2064.
- Banait, N.S., and Jencks, W.P. (1991). Reactions of anionic nucleophiles with α -D-glucopyranosyl fluoride in aqueous solution through a concerted, ANDN (S_N2) mechanism. *J. Am. Chem. Soc.* *113*, 7951–7958.
- Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y.,

- van Wezel, G.P., and van Wezel, G.P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* *80*, 1–43.
- Bartoli, M., Gicquel, E., Barrault, L., Soheili, T., Malissen, M., Malissen, B., Vincent-Lacaze, N., Perez, N., Udd, B., Danos, O., et al. (2008). Mannosidase I inhibition rescues the human α -sarcoglycan R77C recurrent mutation. *Hum. Mol. Genet.* *17*, 1214–1221.
- Beer, D., Maloisel, J.-L., Rast, D.M., and Vasella, A. (1990). Synthesis of 2-Acetamido-2-deoxy-D-gluconhydroximolactone- and Chitobionhydroximolactone-Derived N-Phenylcarbamates, Potential Inhibitors of β -N-Acetylglucosaminidase. *Helv. Chim. Acta* *73*, 1918–1922.
- Bergeron-Brlek, M., Goodwin-Tindall, J., Cekic, N., Roth, C., Zandberg, W.F., Shan, X., Varghese, V., Chan, S., Davies, G.J., Vocadlo, D.J., et al. (2015). A Convenient Approach to Stereoisomeric Iminocyclitols: Generation of Potent Brain-Permeable OGA Inhibitors. *Angew. Chemie - Int. Ed.* *54*, 15429–15433.
- Berti, P.J., and Tanaka, K.S.E. (2002). Transition State Analysis Using Multiple Kinetic Isotope Effects: Mechanisms of Enzymatic and Non-Enzymatic Glycoside Hydrolysis and Transfer. *Adv. Phys. Org. Chem.* *37*, 239–314.
- Bijvoet, J.M., Peerdeman, A.F., and van Bommel, A.J. (1951). Determination of the absolute configuration of optically active compounds by means of X-rays. *Nature* *168*, 271–272.
- Bischoff, J., Liscum, L., and Kornfeld, R. (1986). The use of 1-deoxymannojirimycin to evaluate the role of various alpha-mannosidases in oligosaccharide processing in intact cells. *J. Biol. Chem.* *261*, 4766–4774.
- Boehmelt, G., Fialka, I., Brothers, G., McGinley, M.D., Patterson, S.D., Mo, R., Hui, C.C., Chung, S., Huber, L.A., Mak, T.W., et al. (2000). Cloning and Characterization of the Murine Glucosamine-6-phosphate Acetyltransferase EMeg32: Differential expression and intracellular membrane association. *J. Biol. Chem.* *275*, 12821–12832.
- Bond, P. (2017). No Title.
- Bond, M.R., and Hanover, J.A. (2015). A little sugar goes a long way: The cell biology of O-GlcNAc. *J. Cell Biol.* *208*, 869–880.
- Bosch, A.A.T.M., Biesbroek, G., Trzcinski, K., Sanders, E.A.M., and Bogaert, D. (2013). Viral and Bacterial Interactions in the Upper Respiratory Tract. *PLoS Pathog.* *9*, e1003057.
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* *82*, 239–259.
- Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* *112*, 389–404.
- Brown, B.I., and Brown, D.H. (1965). The subcellular distribution of enzymes in type II glycogenosis and the occurrence of an oligo- α -1,4-glucan glucohydrolase in human tissues. *Biochim. Biophys. Acta - Enzymol. Biol. Oxid.* *110*, 124–133.
- Buckwalter, C.M., and King, S.J. (2012). Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* *20*, 517–522.
- Burda, P., and Aebi, M. (1998). The ALG10 locus of *Saccharomyces cerevisiae* encodes the alpha-1,2 glucosyltransferase of the endoplasmic reticulum: the terminal glucose of the lipid-linked oligosaccharide is required for efficient N-linked glycosylation. *Glycobiology* *8*, 455–462.
- Butkinaree, C., Cheung, W.D., Park, S., Park, K., Barber, M., and Hart, G.W. (2008). Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *J. Biol. Chem.* *283*, 23557–23566.
- Butner, K.A. (1991). Tau protein binds to microtubules through a flexible array of distributed weak sites. *J.*

Byrd, J.C., Tarentino, A.L., Maley, F., Atkinson, P.H., and Trimble, R.B. (1982). Glycoprotein synthesis in yeast. Identification of Man8GlcNAc2 as an essential intermediate in oligosaccharide processing. *J. Biol. Chem.* 257, 14657–14666.

De Calignon, A., Polydoro, M., Suárez-Calvet, M., William, C., Adamowicz, D.H., Kopeikina, K.J., Pitstick, R., Sahara, N., Ashe, K.H., Carlson, G.A., et al. (2012). Propagation of Tau Pathology in a Model of Early Alzheimer's Disease. *Neuron* 73, 685–697.

Cámara, M., Boulnois, G.J., Andrew, P.W., and Mitchell, T.J. (1994). A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infect. Immun.* 62, 3688–3695.

Canfield, P., Dahlbom, M.G., Hush, N.S., and Reimers, J.R. (2006). Density-functional geometry optimization of the 150 000-atom photosystem-I trimer. *J. Chem. Phys.* 124, 024301.

Cantú, D., Nerinckx, W., and Reilly, P.J. (2008). Theory and computation show that Asp463 is the catalytic proton donor in human endoplasmic reticulum alpha-(1→2)-mannosidase I. *Carbohydr. Res.* 343, 2235–2242.

Carrier, M., Joint, M., Luttinger, R., Page, A., and Rochette-Egly, C. (2016). Phosphoproteome and transcriptome of RA-Responsive and RA-Resistant breast cancer cell lines. *PLoS One* 11.

Çetinbaş, N., Macauley, M.S., Stubbs, K.A., Drapala, R., and Vocadlo, D.J. (2006). Identification of Asp174 and Asp175 as the key catalytic residues of human O-GlcNAcase by functional analysis of site-directed mutants. *Biochemistry* 45, 3835–3844.

Champattanachai, V., Marchase, R.B., and Chatham, J.C. (2008). Glucosamine protects neonatal cardiomyocytes from ischemia-reperfusion injury via increased protein O-GlcNAc and increased mitochondrial Bcl-2. *Am. J. Physiol. Cell Physiol.* 294, C1509-20.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., et al. (2007). Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* 15, 267–273.

Chen, H., Li, R., Liu, Z., Wei, S., Zhang, H., and Li, X. (2013). Synthesis of kifunensine thioanalogs and their inhibitory activities against HIV-RT and α -mannosidase. *Carbohydr. Res.* 365, 1–8.

Chen, W., Helenius, J., Braakman, I., and Helenius, A. (1995). Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc. Natl. Acad. Sci.* 92, 6229–6233.

Chen, W., Shen, S., Dong, L., Zhang, J., and Yang, Q. (2018). Selective inhibition of β -N-acetylhexosaminidases by thioglycosyl-naphthalimide hybrid molecules. *Bioorganic Med. Chem.* 26, 394–400.

Cherazard, R., Epstein, M., Doan, T.L., Salim, T., Bharti, S., and Smith, M.A. (2017). Antimicrobial Resistant *Streptococcus pneumoniae*. *Am. J. Ther.* 24, e361–e369.

Chester, M.A. (1998). IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature of glycolipids--recommendations 1997. *Eur. J. Biochem.* 257, 293–298.

Cho, J.H., and Johnson, G.V.W. (2004). Primed phosphorylation of tau at Thr231 by glycogen synthase kinase 3 β (GSK3 β) plays a critical role in regulating tau's ability to bind and stabilize microtubules. *J. Neurochem.* 88, 349–358.

Clarke, V.A., Platt, N., and Butters, T.D. (1995). Cloning and Expression of the β -N-Acetylglucosaminidase Gene from *Streptococcus pneumoniae*. *J. Biol. Chem.* 270, 8805–8814.

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

- Cole, R.N., and Hart, G.W. (2001). Cytosolic O-glycosylation is abundant in nerve terminals. *J. Neurochem.* 79, 1080–1089.
- Comtesse, N., Maldener, E., and Meese, E. (2001). Identification of a nuclear variant of MGEA5, a cytoplasmic hyaluronidase and a β -N-acetylglucosaminidase. *Biochem. Biophys. Res. Commun.* 283, 634–640.
- Conchie, J., and Levvy, G.A. (1957). Inhibition of glycosidases by aldonolactones of corresponding configuration. *Biochem. J.* 65, 389–395.
- Cormier, J.H., Tamura, T., Sunryd, J.C., and Hebert, D.N. (2009). EDEM1 Recognition and Delivery of Misfolded Proteins to the SEL1L-Containing ERAD Complex. *Mol. Cell* 34, 627–633.
- Cremer, D., and Pople, J.A. (1975). A General definition of ring puckering coordinates. *J. Am. Chem. Soc.* 97, 1354–1358.
- Crocker, P.R. (2005). Siglecs in innate immunity. *Curr. Opin. Pharmacol.* 5, 431–437.
- Cumming, D.A., and Carver, J.P. (1987). Virtual and solution conformations of oligosaccharides. *Biochemistry* 26, 6664–6676.
- Cunneen, M.M., Liu, B., Wang, L., and Reeves, P.R. (2013). Biosynthesis of UDP-GlcNAc, UndPP-GlcNAc and UDP-GlcNAcA Involves Three Easily Distinguished 4-Epimerase Enzymes, Gne, Gnu and GnaB. *PLoS One* 8, e67646.
- Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., et al. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517, 165–169.
- Darby, J.F., Landström, J., Roth, C., He, Y., Davies, G.J., and Hubbard, R.E. (2014). Discovery of selective small-molecule activators of a bacterial glycoside hydrolase. *Angew. Chem. Int. Ed. Engl.* 53, 13419–13423.
- Darby, J.F., Atobe, M., Firth, J.D., Bond, P., Davies, G.J., O'Brien, P., and Hubbard, R.E. (2017). Increase of enzyme activity through specific covalent modification with fragments. *Chem. Sci.* 8, 7772–7779.
- Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859.
- Davies, G.J., Dodson, G.G., Hubbard, R.E., Tolley, S.P., Dauter, Z., Wilson, K.S., Hjort, C., Mikkelsen, J.M., Rasmussen, G., and Schüle, M. (1993). Structure and function of endoglucanase V. *Nature* 365, 362–364.
- Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* 321 (Pt 2), 557–559.
- Davies, G.J., Ducros, V.M.A., Varrot, A., and Zechel, D.L. (2003). Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. *Biochem. Soc. Trans.* 31, 523–527.
- Davies, G.J., Planas, A., and Rovira, C. (2012). Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* 45, 308–316.
- Deng, L., Tsybina, P., Gregg, K.J., Mosi, R., Zandberg, W.F., Boraston, A.B., and Vocadlo, D.J. (2013). Synthesis of 4-methylumbelliferyl α -D-mannopyranosyl-(1 \rightarrow 6)- β -D-mannopyranoside and development of a coupled fluorescent assay for GH125 exo- α -1,6-mannosidases. *Bioorg. Med. Chem.* 21, 4839–4845.
- Dennis, R.J., Taylor, E.J., Macauley, M.S., Stubbs, K.A., Turkenburg, J.P., Hart, S.J., Black, G.N., Vocadlo, D.J., and Davies, G.J. (2006). Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity. *Nat. Struct. Mol. Biol.* 13, 365–371.
- Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J., and Gygi, S.P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci.* 105, 10762–10767.
- Dephoure, N., Gould, K.L., Gygi, S.P., and Kellogg, D.R. (2013). Mapping and analysis of phosphorylation

sites: a quick guide for cell biologists. *Mol. Biol. Cell* 24, 535–542.

Despres, C., Byrne, C., Qi, H., Cantrelle, F.X., Huvent, I., Chambraud, B., Baulieu, E.E., Jacquot, Y., Landrieu, I., Lippens, G., et al. (2017). Identification of the Tau phosphorylation pattern that drives its aggregation. *Proc. Natl. Acad. Sci. U. S. A.* 114, 9080–9085.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K., Teeri, T.T., and Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265, 524–528.

Dole, K., Lipari, F., Herscovics, A., and Howell, P.L. (1997). Crystallization and Preliminary X-Ray Analysis of the Class 1 α 1,2-Mannosidase from *Saccharomyces cerevisiae*. *J. Struct. Biol.* 120, 69–72.

Dong, D.L., and Hart, G.W. (1994). Purification and characterization of an O-GlcNAc selective N-acetyl- β -D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* 269, 19321–19330.

Dorfmueller, H.C., Borodkin, V.S., Schimpl, M., Shepherd, S.M., Shpiro, N.A., and Van Aalten, D.M.F. (2006). GlcNAcstatin: A picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-GlcNAcylation levels. *J. Am. Chem. Soc.* 128, 16484–16485.

Dorfmueller, H.C., Borodkin, V.S., Schimpl, M., and van Aalten, D.M.F. (2009). GlcNAcstatins are nanomolar inhibitors of human O-GlcNAcase inducing cellular hyper-O-GlcNAcylation. *Biochem. J.* 420, 221–227.

Dorfmueller, H.C., Borodkin, V.S., Schimpl, M., Zheng, X., Kime, R., Read, K.D., and van Aalten, D.M.F. (2010). Cell-Penetrant, Nanomolar O-GlcNAcase Inhibitors Selective against Lysosomal Hexosaminidases. *Chem. Biol.* 17, 1250–1255.

Dupoiron, S., Zischek, C., Ligat, L., Carbonne, J., Boulanger, A., Dugé de Bernonville, T., Lautier, M., Rival, P., Arlat, M., Jamet, E., et al. (2015). The N-Glycan cluster from *Xanthomonas campestris* pv. *campestris*: a toolbox for sequential plant N-glycan processing. *J. Biol. Chem.* 290, 6022–6036.

Elbein, A.D., Tropea, J.E., Mitchell, M., and Kaushal, G.P. (1990). Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* 265, 15599–15605.

Ellgaard, L., Molinari, M., and Helenius, A. (1999). Setting the Standards: Quality Control in the Secretory Pathway. *Science* (80-.). 286, 1882–1888.

Elsen, N.L., Patel, S.B., Ford, R.E., Hall, D.L., Hess, F., Kandula, H., Kornienko, M., Reid, J., Selnick, H., Shipman, J.M., et al. (2017). Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* 13, 613–615.

Elting, J.J., Chen, W.W., and Lennarz, W.J. (1980). Characterization of a glucosidase involved in an initial step in the processing of oligosaccharide chains. *J. Biol. Chem.* 255, 2325–2331.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 486–501.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 62, 72–82.

Evans, P.R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 282–292.

Ficko-Blean, E., Stubbs, K.A., Nemirovsky, O., Vocadlo, D.J., and Boraston, A.B. (2008). Structural and mechanistic insight into the basis of mucopolysaccharidosis IIIB. *Proc. Natl. Acad. Sci.* 105, 6560–6565.

Fife, T.H., and Przytas, T.J. (1980). Intramolecular electrostatic and general acid catalysis in the hydrolysis of O,S-thioacetals. *J. Am. Chem. Soc.* 102, 292–299.

Fischer, E. (1890). Synthese des Traubenzuckers. *Berichte Der Dtsch. Chem. Gesellschaft* 23, 799–805.

- Fischer, E. (1891). No Title. *Berichte Der Dtsch. Chem. Gesellschaft* 24, 2683–2687.
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Der Dtsch. Chem. Gesellschaft* 27, 2985–2993.
- Fischer, E., and Hirschberger, J. (1888). Über mannose. *Berichte Der Dtsch. Chem. Gesellschaft* 21, 1805–1809.
- Fletcher, H.G. (1940). Augustin-Pierre Dubrunfaut—An early sugar chemist. *J. Chem. Educ.* 17, 153–156.
- Fowler, C.A., Hemsworth, G.R., Cuskin, F., Hart, S., Turkenburg, J., Gilbert, H.J., Walton, P.H., and Davies, G.J. (2018). Structure and function of a glycoside hydrolase family 8 endoxylanase from *Teredinibacter turnerae*. *Acta Crystallogr. Sect. D Struct. Biol.* 74, 946–955.
- Frost, B., Jacks, R.L., and Diamond, M.I. (2009). Propagation of Tau Misfolding from the Outside to the Inside of a Cell. *J. Biol. Chem.* 284, 12845–12852.
- Fujimori, T., Kamiya, Y., Nagata, K., Kato, K., and Hosokawa, N. (2013). Endoplasmic reticulum lectin XTP3-B inhibits endoplasmic reticulum-associated degradation of a misfolded α 1-antitrypsin variant. *FEBS J.* 280, 1563–1575.
- Galli-Stampino, L., Meinjohanns, E., Frische, K., Meldal, M., Jensen, T., Werdelin, O., and Mouritsen, S. (1997). T-cell recognition of tumor-associated carbohydrates: the nature of the glycan moiety plays a decisive role in determining glycopeptide immunogenicity. *Cancer Res.* 57, 3214–3222.
- Ganem, B. (1996). Inhibitors of Carbohydrate-Processing Enzymes: Design and Synthesis of Sugar-Shaped Heterocycles.
- Gao, Y., Parker, G.J., and Hart, G.W. (2000). Streptozotocin-Induced β -Cell Death Is Independent of Its Inhibition of O-GlcNAcase in Pancreatic Min6 Cells. *Arch. Biochem. Biophys.* 383, 296–302.
- Gao, Y., Wells, L., Comer, F.I., Parker, G.J., and Hart, G.W. (2001). Dynamic O-Glycosylation of Nuclear and Cytosolic Proteins: Cloning and Characterization of a Neural, Cytosolic β -N-Acetylglucosamine from Human Brain. *J. Biol. Chem.* 276, 9838–9845.
- Gardner, T.G., and Kears, K.P. (1999). Modification of the T cell antigen receptor (TCR) complex by UDP-glucose:glycoprotein glucosyltransferase. TCR folding is finalized convergent with formation of $\alpha\beta\delta\epsilon\gamma\epsilon$ complexes. *J. Biol. Chem.* 274, 14094–14099.
- Gastaldello, A., Alocci, D., Barletta, E., Mariethoz, J., Lisacek, F., and Ghraichy, M. (2018). Understanding the glycome: an interactive view of glycosylation from glyco compositions to glycoepitopes. *Glycobiology* 28, 349–362.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, (Totowa, NJ: Humana Press), pp. 571–607.
- Geno, K.A., Gilbert, G.L., Song, J.Y., Skovsted, I.C., Klugman, K.P., Jones, C., Konradsen, H.B., and Nahm, M.H. (2015). Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin. Microbiol. Rev.* 28, 871–899.
- Gloster, T.M., and Davies, G.J. (2010). Glycosidase inhibition: assessing mimicry of the transition state. *Org. Biomol. Chem.* 8, 305–320.
- Gloster, T.M., and Vocadlo, D.J. (2012). Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nat. Chem. Biol.* 8, 683–694.
- van der Goot, A.T., Pearce, M.M.P., Leto, D.E., Shaler, T.A., and Kopito, R.R. (2018). Redundant and Antagonistic Roles of XTP3B and OS9 in Decoding Glycan and Non-glycan Degrons in ER-Associated Degradation. *Mol. Cell* 70, 516–530.

- Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890.
- Gregg, K.J., Zandberg, W.F., Hehemann, J.-H., Whitworth, G.E., Deng, L., Vocadlo, D.J., and Boraston, A.B. (2011). Analysis of a new family of widely distributed metal-independent α -mannosidases provides unique insight into the processing of N-linked glycans. *J. Biol. Chem.* 286, 15586–15596.
- Greig, I.R., and Williams, I.H. (2007). Glycosidase inhibitors as conformational transition state analogues. *Chem. Commun.* 3747.
- Griffith, L.S., and Schmitz, B. (1995). O-linked N-acetylglucosamine is upregulated in Alzheimer brains. *Biochem. Biophys. Res. Commun.* 213, 424–431.
- Grinnat, L.S., and Robbins, P.W. (1979). Glycoprotein biosynthesis: rat liver microsomal glucosidases which process oligosaccharides. *J. Biol. Chem.* 254, 8814–8818.
- Grondin, B., and Herscovics, A. (1992). Topology of ER processing α -mannosidase of *Saccharomyces cerevisiae*. *Glycobiology* 2, 369–372.
- Groves, J.A., Maduka, A.O., O'Meally, R.N., Cole, R.N., and Zachara, N.E. (2017). Fatty acid synthase inhibits the O-GlcNAcase during oxidative stress. *J. Biol. Chem.* 292, 6493–6511.
- Grundke-Iqbal, I., Iqbal, K., Tung, Y.C., Quinlan, M., Wisniewski, H.M., and Binder, L.I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc. Natl. Acad. Sci. U. S. A.* 83, 4913–4917.
- Gustke, N., Trinczek, B., Biernat, J., Mandelkow, E.-M., and Mandelkow, E. (1994). Domains of tau Protein and Interactions with Microtubules. *Biochemistry* 33, 9511–9522.
- Haltiwanger, R.S., Grove, K., and Philipsberg, G.A. (1998). Modulation of O-Linked N-Acetylglucosamine Levels on Nuclear and Cytoplasmic Proteins in Vivo Using the Peptide O-GlcNAc- β -N-acetylglucosaminidase Inhibitor O-(2-Acetamido-2-deoxy-dglucopyranosylidene)amino-N-phenylcarbamate. *J. Biol. Chem.* 273, 3611–3617.
- Hammond, C., and Helenius, A. (1995). Quality control in the secretory pathway. *Curr. Opin. Cell Biol.* 7, 523–529.
- Hart, G.W., Housley, M.P., and Slawson, C. (2007). Cycling of O-linked β -N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446, 1017–1022.
- Hastings, N.B., Wang, X., Song, L., Butts, B.D., Grotz, D., Hargreaves, R., Fred Hess, J., Hong, K.-L.K., Huang, C.R.-R., Hyde, L., et al. (2017). Inhibition of O-GlcNAcase leads to elevation of O-GlcNAc tau and reduction of tauopathy and cerebrospinal fluid tau in rTg4510 mice. *Mol. Neurodegener.* 12, 39.
- Hattie, M., Ito, T., Debowski, A.W., Arakawa, T., Katayama, T., Yamamoto, K., Fushinobu, S., and Stubbs, K.A. (2015). Gaining insight into the catalysis by GH20 lacto-N-biosidase using small molecule inhibitors and structural analysis. *Chem. Commun.* 51, 15008–15011.
- Hattie, M., Cekic, N., Debowski, A.W., Vocadlo, D.J., and Stubbs, K.A. (2016). Modifying the phenyl group of PUGNAc: Reactivity tuning to deliver selective inhibitors for N-acetyl-d-glucosaminidases. *Org. Biomol. Chem.* 14, 3193–3197.
- Haurum, J.S., Arsequell, G., Lellouch, A.C., Wong, S.Y., Dwek, R.A., McMichael, A.J., and Elliott, T. (1994). Recognition of carbohydrate by major histocompatibility complex class I-restricted, glycopeptide-specific cytotoxic T lymphocytes. *J. Exp. Med.* 180, 739–744.
- Hava, D.L., and Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol. Microbiol.* 45, 1389–1406.
- He, Y., Macauley, M.S., Stubbs, K.A., Vocadlo, D.J., and Davies, G.J. (2010). Visualizing the Reaction

Coordinate of an O-GlcNAc Hydrolase. *J. Am. Chem. Soc.* *132*, 1807–1809.

He, Y., Roth, C., Turkenburg, J.P., and Davies, G.J. (2014). Three-dimensional structure of a *Streptomyces sviveus* GNAT acetyltransferase with similarity to the C-terminal domain of the human GH84 O-GlcNAcase. *Acta Crystallogr. Sect. D* *70*, 186–195.

Heckel, D., Comtesse, N., Brass, N., Blin, N., Zang, K.D., and Meese, E. (1998). Novel immunogenic antigen homologous to hyaluronidase in meningioma. *Hum. Mol. Genet.* *7*, 1859–1872.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* (80-.). *324*, 1029–1033.

Heightman, T.D., and Vasella, A.T. (1999). Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases. *Angew. Chemie Int. Ed.* *38*, 750–770.

Henriques-Normark, B., and Tuomanen, E.I. (2013). The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb. Perspect. Med.* *3*.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* *280*, 309–316.

Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* *293*, 781–788.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* *316*, 695–696.

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* *7*, 637–644.

Hering, K.W., Karaveg, K., Moremen, K.W., and Pearson, W.H. (2005). A Practical Synthesis of Kifunensine Analogues as Inhibitors of Endoplasmic Reticulum α -Mannosidase I. *J. Org. Chem.* *70*, 9892–9904.

Hinderlich, S., Nohring, S., Weise, C., Franke, P., Stasche, R., and Reutter, W. (1998). Purification and characterization of N-acetylglucosamine kinase from rat liver. Comparison with UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase. *Eur. J. Biochem.* *252*, 133–139.

Hohenberg, P., and Kohn, W. (1964). Inhomogeneous Electron Gas. *Phys. Rev.* *136*, B864–B871.

Holt, G.D., and Hart, G.W. (1986). The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. *J. Biol. Chem.* *261*, 8049–8057.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* *43*, D512–D520.

Hornbeck, P. V., Chabra, I., Kornhauser, J.M., Skrzypek, E., and Zhang, B. (2004). PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* *4*, 1551–1561.

Horsch, M., Hoesch, L., Vasella, A., and Rast, D.M. (1991). N-acetylglucosaminono-1,5-lactone oxime and the corresponding (phenylcarbamoyl)oxime. Novel and potent inhibitors of beta-N-acetylglucosaminidase. *Eur. J. Biochem.* *197*, 815–818.

Hosokawa, N., Wada, I., Hasegawa, K., Yorihuzi, T., Tremblay, L.O., Herscovics, A., and Nagata, K. (2001). A novel ER alpha-mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep.* *2*, 415–422.

Hosokawa, N., Kamiya, Y., Kamiya, D., Kato, K., and Nagata, K. (2009). Human OS-9, a lectin required for glycoprotein endoplasmic reticulum-associated degradation, recognizes mannose-trimmed N-glycans. *J. Biol. Chem.* *284*, 17061–17068.

Huffaker, T.C., and Robbins, P.W. (1982). Temperature-sensitive yeast mutants deficient in asparagine-linked

glycosylation. *J. Biol. Chem.* 257, 3203–3210.

Igdoura, S.A., Herscovics, A., Lal, A., Moremen, K.W., Morales, C.R., and Hermo, L. (1999). Alpha-mannosidases involved in N-glycan processing show cell specificity and distinct subcompartmentalization within the Golgi apparatus of cells in the testis and epididymis. *Eur J Cell Biol* 78, 441–452.

Igual, M.O., Nunes, P.S.G., da Costa, R.M., Mantoani, S.P., Tostes, R.C., and Carvalho, I. (2019). Novel glucopyranoside C2-derived 1,2,3-triazoles displaying selective inhibition of O-GlcNAcase (OGA). *Carbohydr. Res.* 471, 43–55.

Ito, T., Katayama, T., Hattie, M., Sakurama, H., Wada, J., Suzuki, R., Ashida, H., Wakagi, T., Yamamoto, K., Stubbs, K.A., et al. (2013). Crystal Structures of a Glycoside Hydrolase Family 20 Lacto- N -biosidase from *Bifidobacterium bifidum*. *J. Biol. Chem.* 288, 11795–11806.

Iwami, M., Nakayama, O., Terano, H., Kohsaka, M., Aoki, H., and Imanaka, H. (1987). A new immunomodulator, FR-900494: Taxonomy, fermentation, isolation, and physico-chemical and biological characteristics. *J. Antibiot. (Tokyo)*. 40, 612–622.

Jagannadham, M. V, and Nagaraj, R. (2008). Detecting the site of phosphorylation in phosphopeptides without loss of phosphate group using MALDI TOF mass spectrometry. *Anal. Chem. Insights* 3, 21–29.

Jelinek-Kelly, S., Akiyama, T., Saunier, B., Tkacz, J.S., and Herscovics, A. (1985). Characterization of a specific α -mannosidase involved in oligosaccharide processing in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 260, 2253–2257.

Jelk, J., Gao, N., Serricchio, M., Signorell, A., Schmidt, R.S., Bangs, J.D., Acosta-Serrano, A., Lehrman, M.A., Bütikofer, P., and Menon, A.K. (2013). Glycoprotein biosynthesis in a eukaryote lacking the membrane protein Rft1. *J. Biol. Chem.* 288, 20616–20623.

Jensen, H.H., and Bols, M. (2006). Stereoelectronic Substituent Effects. *Acc. Chem. Res* 39, 259–265.

Joiner, C.M., Li, H., Jiang, J., and Walker, S. (2019). Structural characterization of the O-GlcNAc cycling enzymes: insights into substrate recognition and catalytic mechanisms. *Curr. Opin. Struct. Biol.* 56, 97–106.

Kadavath, H., Hofele, R. V, Biernat, J., Kumar, S., Tepper, K., Urlaub, H., Mandelkow, E., and Zweckstetter, M. (2015a). Tau stabilizes microtubules by binding at the interface between tubulin heterodimers. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7501–7506.

Kadavath, H., Jaremko, M., Jaremko, Ł., Biernat, J., Mandelkow, E., and Zweckstetter, M. (2015b). Folding of the Tau Protein on Microtubules. *Angew. Chemie Int. Ed.* 54, 10347–10351.

Kalelkar, S., Glushka, J., van Halbeek, H., Morris, L.C., and Cherniak, R. (1997). Structure of the capsular polysaccharide of *Clostridium perfringens* Hobbs 5 as determined by NMR spectroscopy. *Carbohydr. Res.* 299, 119–128.

Kämpf, M., Absmanner, B., Schwarz, M., and Lehle, L. (2009). Biochemical characterization and membrane topology of Alg2 from *Saccharomyces cerevisiae* as a bifunctional α 1,3- and 1,6-mannosyltransferase involved in lipid-linked oligosaccharide biosynthesis. *J. Biol. Chem.* 284, 11900–11912.

Karaveg, K., and Moremen, K.W. (2005). Energetics of substrate binding and catalysis by class 1 (glycosylhydrolase family 47) α -mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *J. Biol. Chem.* 280, 29837–29848.

Karaveg, K., Siriwardena, A., Tempel, W., Liu, Z.-J., Glushka, J., Wang, B.-C., and Moremen, K.W. (2005). Mechanism of class 1 (glycosylhydrolase family 47) α -mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *J. Biol. Chem.* 280, 16197–16207.

Kayakiri, H., Takase, S., Shibata, T., Okamoto, M., Terano, H., Hashimoto, M., Tada, T., and Koda, S. (1989). Structure of kifunensine, a new immunomodulator isolated from an actinomycete. *J. Org. Chem.* 54, 4015–4016.

- Keembiyehetty, C.N., Krzeslak, A., Love, D.C., and Hanover, J.A. (2011). A lipid-droplet-targeted O-GlcNAcase isoform is a key regulator of the proteasome. *J. Cell Sci.* *124*, 2851–2860.
- Kfoury, N., Holmes, B.B., Jiang, H., Holtzman, D.M., and Diamond, M.I. (2012). Trans-cellular Propagation of Tau Aggregation by Fibrillar Species. *J. Biol. Chem.* *287*, 19440–19451.
- Kim, S.J., Sohn, I., Do, I.-G., Jung, S.H., Ko, Y.H., Yoo, H.Y., Paik, S., and Kim, W.S. (2014). Gene expression profiles for the prediction of progression-free survival in diffuse large B cell lymphoma: results of a DASL assay. *Ann. Hematol.* *93*, 437–447.
- King, S.J. (2010). Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Mol. Oral Microbiol.* *25*, 15–24.
- Kirschner, K.N., and Woods, R.J. (2001). Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci.* *98*, 10541–10545.
- Klammer, M., Kaminski, M., Zedler, A., Oppermann, F., Blencke, S., Marx, S., Muller, S., Tebbe, A., Godl, K., Schaab, C., et al. (2012). Phosphosignature Predicts Dasatinib Response in Non-small Cell Lung Cancer. *Mol. Cell. Proteomics* *11*, 651–668.
- Knapp, S., Vocadlo, D., Gao, Z., Kirk, B., Lou, J., and Withers, S.G. (1996). NAG-thiazoline, an N-acetyl- β -hexosaminidase inhibitor that implicates acetamido participation. *J. Am. Chem. Soc.* *118*, 6804–6805.
- Knott, B.C., Haddad Momeni, M., Crowley, M.F., Mackenzie, L.F., Götz, A.W., Sandgren, M., Withers, S.G., Ståhlberg, J., and Beckham, G.T. (2014). The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* *136*, 321–329.
- Kornfeld, S., Li, E., and Tabas, I. (1978). The synthesis of complex-type oligosaccharides. II. Characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein. *J. Biol. Chem.* *253*, 7771–7778.
- Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* *28*, 416–436.
- Krepischi, A.C. V., Maschietto, M., Ferreira, E.N., Silva, A.G., Costa, S.S., da Cunha, I.W., Barros, B.D.F., Grundy, P.E., Rosenberg, C., and Carraro, D.M. (2016). Genomic imbalances pinpoint potential oncogenes and tumor suppressors in Wilms tumors. *Mol. Cytogenet.* *9*, 20.
- Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic Glycosylation of Nuclear and Cytosolic Proteins: Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* *272*, 9308–9315.
- Ksiazek-Reding, H., Liu, W.-K., and Yen, S.-H. (1992). Phosphate analysis and dephosphorylation of modified tau associated with paired helical filaments. *Brain Res.* *597*, 209–219.
- Kuntz, D.A., Tarling, C.A., Withers, S.G., and Rose, D.R. (2008). Structural analysis of Golgi α -mannosidase II inhibitors identified from a focused glycosidase inhibitor screen. *Biochemistry* *47*, 10058–10068.
- Lai, H.-Y.L., and Axelrod, B. (1973). 1-Aminoglycosides, a new class of specific inhibitors of glycosidases. *Biochem. Biophys. Res. Commun.* *54*, 463–468.
- Lameira, J., Alves, C.N., Moliner, V., Martí, S., Kanaan, N., and Tuñón, I. (2008). A Quantum Mechanics/Molecular Mechanics Study of the Protein–Ligand Interaction of Two Potent Inhibitors of Human O-GlcNAcase: PUGNAc and NAG-Thiazoline. *J. Phys. Chem. B* *112*, 14260–14266.
- Lau, G.W., Haataja, S., Lonetto, M., Kensit, S.E., Marra, A., Bryant, A.P., McDevitt, D., Morrison, D.A., and Holden, D.W. (2001). A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. *Mol. Microbiol.* *40*, 555–571.
- Lehrman, M.A. (1991). Biosynthesis of N-acetylglucosamine-P-P-dolichol, the committed step of asparagine-

linked oligosaccharide assembly. *Glycobiology* 1, 553–562.

De Leon, C.A., Levine, P.M., Craven, T.W., and Pratt, M.R. (2017). The Sulfur-Linked Analogue of O-GlcNAc (S-GlcNAc) Is an Enzymatically Stable and Reasonable Structural Surrogate for O-GlcNAc at the Peptide and Protein Levels. *Biochemistry* 56, 3507–3517.

Li, B., Li, H., Lu, L., and Jiang, J. (2017a). Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat. Struct. Mol. Biol.* 24, 362–369.

Li, B., Li, H., Hu, C.-W., and Jiang, J. (2017b). Structural insights into the substrate binding adaptability and specificity of human O-GlcNAcase. *Nat. Commun.* 8, 666.

Li, H., Wang, G., Yu, Y., Jian, W., Zhang, D., Wang, Y., Wang, T., Meng, Y., Yuan, C., and Zhang, C. (2018a). α -1,2-Mannosidase MAN1C1 Inhibits Proliferation and Invasion of Clear Cell Renal Cell Carcinoma. *J. Cancer* 9, 4618–4626.

Li, S., Wang, J., Zang, L., Zhu, H., Guo, J., Zhang, J., Wen, L., Chen, Y., Li, Y., Chen, X., et al. (2018b). Production of Glycopeptide Derivatives for Exploring Substrate Specificity of Human OGA Toward Sugar Moiety. *Front. Chem.* 6, 646.

Lichtenthaler, F.W. (1992). Emil Fischer's Proof of the Configuration of Sugars: A Centennial Tribute.

Lindwall, G., and Cole, R.D. (1984). Phosphorylation affects the ability of tau protein to promote microtubule assembly. *J. Biol. Chem.* 259, 5301–5305.

Lipari, F., and Herscovics, A. (1994). Production, purification and characterization of recombinant yeast processing α 1,2-mannosidase. *Glycobiology* 4, 697–702.

Lipari, F., and Herscovics, A. (1996). Role of the Cysteine Residues in the α 1,2-Mannosidase Involved in N-Glycan Biosynthesis in *Saccharomyces cerevisiae*: The conserved Cys340 and Cys385 residues form an essential disulfide bond. *J. Biol. Chem.* 271, 27615–27622.

Lipari, F., and Herscovics, A. (1999). Calcium binding to the class I α -1,2-mannosidase from *Saccharomyces cerevisiae* occurs outside the EF hand motif. *Biochemistry* 38, 1111–1118.

Lipari, F., Goursalin, B.J., and Herscovics, A. (1995). The *Saccharomyces cerevisiae* Processing α 1,2-Mannosidase Is an Inverting Glycosidase. *Biochem. Biophys. Res. Commun.* 209, 322–326.

Lipkind, G.M., Verovsky, V.E., and Kochetkov, N.K. (1984). Conformational states of cellobiose and maltose in solution: A comparison of calculated and experimental data. *Carbohydr. Res.* 133, 1–13.

Lipkind, G.M., Shashkov, A.S., and Kochetkov, N.K. (1985). Nuclear overhauser effect and conformational states of cellobiose in aqueous solution. *Carbohydr. Res.* 141, 191–197.

Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W., and Gong, C.-X.C.-X. (2004). O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10804–10809.

Liu, H., Liang, X., Søhoel, H., Bülow, A., and Bols, M. (2001). Noeuromycin, a glycosyl cation mimic that strongly inhibits glycosidases. *J. Am. Chem. Soc.* 123, 5116–5117.

Liu, T., Xia, M., Zhang, H., Zhou, H., Wang, J., Shen, X., and Yang, Q. (2015). Exploring NAG-thiazoline and its derivatives as inhibitors of chitinolytic β -acetylglucosaminidases. *FEBS Lett.* 589, 110–116.

Lobsanov, Y.D., Vallée, F., Imbert, A., Yoshida, T., Yip, P., Herscovics, A., and Howell, P.L. (2002). Structure of *Penicillium citrinum* alpha 1,2-mannosidase reveals the basis for differences in specificity of the endoplasmic reticulum and Golgi class I enzymes. *J. Biol. Chem.* 277, 5620–5630.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–495.

- Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats. *J. Biol. Chem.* 272, 9316–9324.
- Lumpkin, R.J., Gu, H., Zhu, Y., Leonard, M., Ahmad, A.S., Clauser, K.R., Meyer, J.G., Bennett, E.J., and Komives, E.A. (2017). Site-specific identification and quantitation of endogenous SUMO modifications under native conditions. *Nat. Commun.* 8, 1171–1181.
- M. Kuttel, M. (2011). The Conformational Free Energy of Carbohydrates. *Mini. Rev. Org. Chem.* 8, 256–262.
- Macauley, M.S., and Vocadlo, D.J. (2009). Enzymatic characterization and inhibition of the nuclear variant of human O-GlcNAcase. *Carbohydr. Res.* 344, 1079–1084.
- Macauley, M.S., Stubbs, K.A., and Vocadlo, D.J. (2005a). O-GlcNAcase Catalyzes Cleavage of Thioglycosides without General Acid Catalysis. *J. Am. Chem. Soc.* 127, 17202–17203.
- Macauley, M.S., Whitworth, G.E., Debowski, A.W., Chin, D., and Vocadlo, D.J. (2005b). O-GlcNAcase uses substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors. *J. Biol. Chem.* 280, 25313–25322.
- Maeda, Y., and Kinoshita, T. (2010). The Acidic Environment of the Golgi Is Critical for Glycosylation and Transport. In *Methods in Enzymology*, pp. 495–510.
- Males, A., and Davies, G.J. (2019). Structural studies of a surface-entropy reduction mutant of O-GlcNAcase. *Acta Crystallogr. Sect. D, Struct. Biol.* 75, 70–78.
- Males, A., Raich, L., Williams, S.J., Rovira, C., and Davies, G.J. (2017). Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Structural Approach. *ChemBioChem* 18, 1496–1501.
- Mansford, K.R., and Opie, L. (1968). Comparison of metabolic abnormalities in diabetes mellitus induced by streptozotocin or by alloxan. *Lancet (London, England)* 1, 670–671.
- Marin, O., Meggio, F., Draetta, G., and Pinna, L.A. (1992). The consensus sequences for cdc2 kinase and for casein kinase-2 are mutually incompatible A study with peptides derived from the β -subunit of casein kinase-2. *FEBS Lett.* 301, 111–114.
- Mark, B.L., Vocadlo, D.J., Knapp, S., Triggs-Raine, B.L., Withers, S.G., and James, M.N. (2001). Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase. *J. Biol. Chem.* 276, 10330–10337.
- Marshall, S., Bacote, V., and Traxinger, R.R. (1991). Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* 266, 4706–4712.
- Martin, J.C., Fadda, E., Ito, K., and Woods, R.J. (2014). Defining the structural origin of the substrate sequence independence of O-GlcNAcase using a combination of molecular docking and dynamics simulation. *Glycobiology* 24, 85–96.
- Martinez-Fleites, C., Korczynska, J.E., Davies, G.J., Cope, M.J., Turkenburg, J.P., and Taylor, E.J. (2009). The crystal structure of a family GH25 lysozyme from *Bacillus anthracis* implies a neighboring-group catalytic mechanism with retention of anomeric configuration. *Carbohydr. Res.* 344, 1753–1757.
- Martinez-Fleites, C., He, Y., and Davies, G.J. (2010). Structural analyses of enzymes involved in the O-GlcNAc modification. *Biochim. Biophys. Acta* 1800, 122–133.
- Massaad, M.J., Franzusoff, A., and Herscovics, A. (1999). The processing α 1,2-mannosidase of *Saccharomyces cerevisiae* depends on Rer1p for its localization in the endoplasmic reticulum. *Eur. J. Cell Biol.* 78, 435–440.
- Mayes, H.B., Broadbelt, L.J., and Beckham, G.T. (2014). How sugars pucker: Electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for

enzymatic catalysis. *J. Am. Chem. Soc.* *136*, 1008–1022.

McCarter, J.D., and Stephen Withers, G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* *4*, 885–892.

McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Körner, M., Plesniak, L.A., Ziser, L., Wakarchuk, W.W., and Withers, S.G. (1996). The pKa of the General Acid/Base Carboxyl Group of a Glycosidase Cycles during Catalysis: A ¹³C-NMR Study of *Bacillus circulans* Xylanase. *Biochemistry* *35*, 9958–9966.

McLachlin, D.T., and Chait, B.T. (2001). Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* *5*, 591–602.

McNaught, A.D. (1997). International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology. Joint Commission on Biochemical Nomenclature. Nomenclature of carbohydrates. *Carbohydr. Res.* *297*, 1–92.

McNicholas, S., Potterton, E., Wilson, K.S., and Noble, M.E.M. (2011). Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *67*, 386–394.

Medina, M. (2018). An Overview on the Clinical Development of Tau-Based Therapeutics. *Int. J. Mol. Sci.* *19*, 1160.

Mehdy, A., Morelle, W., Rosnoblet, C., Legrand, D., Lefebvre, T., Duvet, S., and Foulquier, F. (2012). PUGNAC treatment leads to an unusual accumulation of free oligosaccharides in CHO cells. *J. Biochem.* *151*, 439–446.

Milde-Langosch, K., Karn, T., Schmidt, M., zu Eulenburg, C., Oliveira-Ferrer, L., Wirtz, R.M., Schumacher, U., Witzel, I., Schütze, D., and Müller, V. (2014). Prognostic relevance of glycosylation-associated genes in breast cancer. *Breast Cancer Res. Treat.* *145*, 295–305.

Mio, T., Yabe, T., Arisawa, M., and Yamada-Okabe, H. (1998). The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression, and catalytic mechanism. *J. Biol. Chem.* *273*, 14392–14397.

Money, V.A., Cartmell, A., Guerreiro, C.I.P.D., Ducros, V.M.-A., Fontes, C.M.G.A., Gilbert, H.J., and Davies, G.J. (2008). Probing the β -1,3:1,4 glucanase, CtLic26A, with a thio-oligosaccharide and enzyme variants. *Org. Biomol. Chem.* *6*, 851.

Montalvillo-Jiménez, L., Santana, A.G., Corzana, F., Jiménez-Osés, G., Jiménez-Barbero, J., Gómez, A.M., and Asensio, J.L. (2019). Impact of Aromatic Stacking on Glycoside Reactivity: Balancing CH/ π and Cation/ π Interactions for the Stabilization of Glycosyl-Oxocarbenium Ions. *J. Am. Chem. Soc.* *141*, 13372–13384.

Mulakala, C., Nerinckx, W., and Reilly, P.J. (2006). Docking studies on glycoside hydrolase Family 47 endoplasmic reticulum α -(1 \rightarrow 2)-mannosidase I to elucidate the pathway to the substrate transition state. *Carbohydr. Res.* *341*, 2233–2245.

Mulakala, C., Nerinckx, W., and Reilly, P.J. (2007). The fate of β -d-mannopyranose after its formation by endoplasmic reticulum α -(1 \rightarrow 2)-mannosidase I catalysis. *Carbohydr. Res.* *342*, 163–169.

Muramatsu, H., Tachikui, H., Ushida, H., Song, X., Qiu, Y., Yamamoto, S., and Muramatsu, T. (2001). Molecular Cloning and Expression of Endo- β -N-Acetylglucosaminidase D, Which Acts on the Core Structure of Complex Type Asparagine-Linked Oligosaccharides. *J. Biochem.* *129*, 923–928.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *53*, 240–255.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* *67*, 355–367.

Nagel, A.K., and Ball, L.E. (2014). O-GlcNAc transferase and O-GlcNAcase: Achieving target substrate

specificity. *Amino Acids* 46, 2305–2316.

Natowicz, M.R., Chi, M.M., Lowry, O.H., and Sly, W.S. (1979). Enzymatic identification of mannose 6-phosphate on the recognition marker for receptor-mediated pinocytosis of beta-glucuronidase by human fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4322–4326.

Neddens, J., Temmel, M., Flunkert, S., Kerschbaumer, B., Hoeller, C., Loeffler, T., Niederkofler, V., Daum, G., Attems, J., and Hutter-Paier, B. (2018). Phosphorylation of different tau sites during progression of Alzheimer's disease. *Acta Neuropathol. Commun.* 6, 52.

Nelson, A.L., Roche, A.M., Gould, J.M., Chim, K., Ratner, A.J., and Weiser, J.N. (2007). Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infect. Immun.* 75, 83–90.

Nerinckx, W., Desmet, T., Piens, K., and Claeysens, M. (2005). An elaboration on the syn-anti proton donor concept of glycoside hydrolases: Electrostatic stabilisation of the transition state as a general strategy. *FEBS Lett.* 579, 302–312.

Nicholls, R.A., Long, F., and Murshudov, G.N. (2012). Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 404–417.

Nieto, P.M. (2018). The Use of NMR to Study Transient Carbohydrate–Protein Interactions. *Front. Mol. Biosci.* 5, 33.

Ninagawa, S., Okada, T., Sumitomo, Y., Kamiya, Y., Kato, K., Horimoto, S., Ishikawa, T., Takeda, S., Sakuma, T., Yamamoto, T., et al. (2014). EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step. *J. Cell Biol.* 206, 347–356.

Norris, G.E., Stillman, T.J., Anderson, B.F., and Baker, E.N. (1994). The three-dimensional structure of PNGase F, a glycosylasparaginase from *Flavobacterium meningosepticum*. *Structure* 2, 1049–1059.

Numao, S., Kuntz, D.A., Withers, S.G., and Rose, D.R. (2003). Insights into the mechanism of *Drosophila melanogaster* Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates. *J. Biol. Chem.* 278, 48074–48083.

O'Callaghan, A., and van Sinderen, D. (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Front. Microbiol.* 7, 925.

O'Reilly, M.K., Zhang, G., and Imperiali, B. (2006). In vitro evidence for the dual function of Alg2 and Alg11: Essential mannosyltransferases in N-linked glycoprotein biosynthesis. *Biochemistry* 45, 9593–9603.

Ogen-Shtern, N., Avezov, E., Shenkman, M., Benyair, R., and Lederkremer, G.Z. (2016). Mannosidase IA is in Quality Control Vesicles and Participates in Glycoprotein Targeting to ERAD. *J. Mol. Biol.* 428, 3194–3205.

Okuyama, R., and Marshall, S. (2003). UDP-N-acetylglucosaminyl transferase (OGT) in brain tissue: temperature sensitivity and subcellular distribution of cytosolic and nuclear enzyme. *J. Neurochem.* 86, 1271–1280.

Olden, K., Bernard, B.A., Humphries, M.J., Yeo, T.K., Yeo, K.T., White, S.L., Newton, S.A., Bauer, H.C., and Parent, J.B. (1985). Function of glycoprotein glycans. *Trends Biochem. Sci.* 10, 78–82.

Olivari, S., Cali, T., Salo, K.E.H., Paganetti, P., Ruddock, L.W., and Molinari, M. (2006). EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation. *Biochem. Biophys. Res. Commun.* 349, 1278–1284.

Olszewska, E., Borzym-Kluczyk, M., Rzewnicki, I., Wojtowicz, J., Rogowski, M., Pietruski, J.K., Czajkowska, A., and Sieskiewicz, A. (2012). Possible role of α -mannosidase and β -galactosidase in larynx cancer. *Współczesna Onkol.* 2, 154–158.

Pan, S., Cheng, X., Chen, H., Castro, P.D., Ittmann, M.M., Hutson, A.W., Zapata, S.K., and Sifers, R.N. (2013). ERMAnI Is a Target of miR-125b and Promotes Transformation Phenotypes in Hepatocellular Carcinoma

(HCC). PLoS One 8, e72829.

Panday, N., Canac, Y., and Vasella, A. (2000). Very Strong Inhibition of Glucosidases by C(2)-Substituted Tetrahydroimidazopyridines. *Helv. Chim. Acta* 83, 58–79.

Pannu, N.S., Murshudov, G.N., Dodson, E.J., and Read, R.J. (1998). Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54, 1285–1294.

Parodi, A.J., Mendelzon, D.H., and Lederkremer, G.Z. (1983). Transient glucosylation of protein-bound Man9GlcNAc2, Man8GlcNAc2, and Man7GlcNAc2 in calf thyroid cells. A possible recognition signal in the processing of glycoproteins. *J. Biol. Chem.* 258, 8260–8265.

Parodi, A.J., Caramelo, J.J., and D'Alessio, C. (2014). UDP-Glucose: Glycoprotein Glucosyltransferase 1,2 (UGGT1,2). In *Handbook of Glycosyltransferases and Related Genes*, (Tokyo: Springer Japan), pp. 15–30.

Pathak, S., Dorfmüller, H.C., Borodkin, V.S., and van Aalten, D.M.F. (2008). Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death. *Chem. Biol.* 15, 799–807.

Paul, S., Haskali, M.B., Liow, J.-S., Zoghbi, S.S., Barth, V.N., Kolodrubetz, M.C., Bond, M.R., Morse, C.L., Gladding, R.L., Frankland, M.P., et al. (2019). Evaluation of a PET Radioligand to Image O-GlcNAcase in Brain and Periphery of Rhesus Monkey and Knock-Out Mouse. *J. Nucl. Med.* 60, 129–134.

Pauling, L. (1946). Molecular Architecture and Biological Reactions. *Chem. Eng. News* 24, 1375–1377.

Pauling, L. (1948). Nature of Forces between Large Molecules of Biological Interest. *Nature* 161, 707–709.

Pearlman, S.M., Serber, Z., and Ferrell, J.E. (2011). A Mechanism for the Evolution of Phosphorylation Sites. *Cell* 147, 934–946.

Pérez, S. (2018). Symbolic Representation of Monosaccharides in the Age of Symbolic Representation of Monosaccharides in the Age of Glycobiology. 1–16.

Petersen, L., Ardèvol, A., Rovira, C., and Reilly, P.J. (2010). Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi α -Mannosidase II: A QM/MM Metadynamics Investigation. *J. Am. Chem. Soc.* 132, 8291–8300.

Peterson, J.R., Ora, A., Van, P.N., and Helenius, A. (1995). Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Mol. Biol. Cell* 6, 1173–1184.

Pierre, M., and Nunez, J. (1983). Multisite phosphorylation of tau proteins from rat brain. *Biochem. Biophys. Res. Commun.* 115, 212–219.

Pigman, W., and Isbell, H.S. (1968). Mutarotation of Sugars in Solution: Part I: History, Basic Kinetics, and Composition of Sugar Solutions. *Adv. Carbohydr. Chem.* 23, 11–57.

Pimpaneau, V., Midoux, P., Monsigny, M., and Roche, A.C. (1991). Characterization and isolation of an intracellular D-mannose-specific receptor from human promyelocytic HL60 cells. *Carbohydr. Res.* 213, 95–108.

Polacek, R., Behrends, R., and Kaatze, U. (2002). Chair–Chair Conformational Flexibility of Monosaccharides Linked to the Anomer Equilibrium. *J. Phys. Chem. B* 105, 2894–2896.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998). Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infect. Immun.* 66, 5620–5629.

Rabouille, C., Kuntz, D.A., Lockyer, A., Watson, R., Signorelli, T., Rose, D.R., van den Heuvel, M., and Roberts, D.B. (1999). The *Drosophila* GMII gene encodes a Golgi α -mannosidase II. *J. Cell Sci.* 112 (Pt 19), 3319–3330.

- Rao, V.S.R. (Vallurupalli S.R. (1998). Conformation of carbohydrates (Harwood Academic Publishers).
- Rao, F. V, Dorfmüller, H.C., Villa, F., Allwood, M., Eggleston, I.M., and van Aalten, D.M.F. (2006). Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J.* 25, 1569–1578.
- Rao, F. V, Schüttelkopf, A.W., Dorfmüller, H.C., Ferenbach, A.T., Navratilova, I., and van Aalten, D.M.F. (2013). Structure of a bacterial putative acetyltransferase defines the fold of the human O-GlcNAcase C-terminal domain. *Open Biol.* 3, 130021.
- Reeves, R.E. (1949). Cuprammonium-glycoside complexes. III. The conformation of the D-glycopyranoside ring in solution. *J. Am. Chem. Soc.* 71, 215–217.
- Reiss, G., te Heesen, S., Zimmerman, J., Robbins, P.W., and Aebi, M. (1996). Isolation of the ALG6 locus of *Saccharomyces cerevisiae* required for glucosylation in the N-linked glycosylation pathway. *Glycobiology* 6, 493–498.
- Rigden, D.J., Jedrzejewski, M.J., and de Mello, L. V (2003). Identification and analysis of catalytic TIM barrel domains in seven further glycoside hydrolase families. *FEBS Lett.* 544, 103–111.
- van Rijssel, E.R., Janssen, A.P.A., Males, A., Davies, G.J., van der Marel, G.A., Overkleeft, H.S., and Codée, J.D.C. (2017). Conformational Behaviour of Azasugars Based on Mannuronic Acid. *ChemBioChem* 18, 1297–1304.
- Rip, J.W., Rugar, C.A., Chaudhary, N., and Carroll, K.K. (1981). Localization of a dolichyl phosphate phosphatase in plasma membranes of rat liver. *J. Biol. Chem.* 256, 1929–1934.
- Robb, M., Hobbs, J.K., Woodiga, S.A., Shapiro-Ward, S., Suits, M.D.L., McGregor, N., Brumer, H., Yesilkaya, H., King, S.J., and Boraston, A.B. (2017). Molecular Characterization of N-glycan Degradation and Transport in *Streptococcus pneumoniae* and Its Contribution to Virulence. *PLoS Pathog.* 13, e1006090.
- Romero, P.A. (2000). Mutation of Arg273 to Leu Alters the Specificity of the Yeast N-Glycan Processing Class I alpha 1,2-Mannosidase. *J. Biol. Chem.* 275, 11071–11074.
- Roth, C., Chan, S., Offen, W.A., Hemsworth, G.R., Willems, L.I., King, D.T., Varghese, V., Britton, R., Vocadlo, D.J., and Davies, G.J. (2017). Structural and functional insight into human O-GlcNAcase. *Nat. Chem. Biol.* 13, 610–612.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K., and Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 249, 380–386.
- Ruan, H.-B., Singh, J.P., Li, M.-D., Wu, J., and Yang, X. (2013). Cracking the O-GlcNAc Code in Metabolism. *Trends Endocrinol. Metab.* 24, 301.
- Ruhaak, L.R., Zauner, G., Huhn, C., Bruggink, C., Deelder, A.M., and Wuhrer, M. (2010). Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* 397, 3457–3481.
- Runge, K.W., and Robbins, P.W. (1986). A new yeast mutation in the glucosylation steps of the asparagine-linked glycosylation pathway. Formation of a novel asparagine-linked oligosaccharide containing two glucose residues. *J. Biol. Chem.* 261, 15582–15590.
- Ryan, J.M., Quattropani, A., Abd-Elaziz, K., Daas, I. den, Schneider, M., Ousson, S., Neny, M., Sand, A., Hantson, J., Permanne, B., et al. (2018). PHASE 1 STUDY IN HEALTHY VOLUNTEERS OF THE O-GLCNAcase INHIBITOR ASN120290 AS A NOVEL THERAPY FOR PROGRESSIVE SUPRANUCLEAR PALSY AND RELATED TAUOPATHIES. *Alzheimer's Dement.* 14, P251.
- Sandhu, P., Lee, J., Ballard, J., Walker, B., Ellis, J., Marcus, J., Toolan, D., Dreyer, D., McAvoy, T., Duffy, J., et al. (2016). PHARMACOKINETICS AND PHARMACODYNAMICS TO SUPPORT CLINICAL STUDIES OF MK-8719: AN O-GLCNAcase INHIBITOR FOR PROGRESSIVE SUPRANUCLEAR PALSY. *Alzheimer's Dement.* 12, P1028.

- Sandonà, D., and Betto, R. (2009). Sarcoglycanopathies: molecular pathogenesis and therapeutic prospects. *Expert Rev. Mol. Med.* 11, e28.
- SantaCruz, K., Lewis, J., Spires, T., Paulson, J., Kotilinek, L., Ingelsson, M., Guimaraes, A., DeTure, M., Ramsden, M., McGowan, E., et al. (2005). Tau Suppression in a Neurodegenerative Mouse Model Improves Memory Function. *Science* (80-.). 309, 476–481.
- Schaeffer, V., Lavenir, I., Ozcelik, S., Tolnay, M., Winkler, D.T., and Goedert, M. (2012). Stimulation of autophagy reduces neurodegeneration in a mouse model of human tauopathy. *Brain* 135, 2169–2177.
- Schallus, T., Jaechk, C., Fehér, K., Palma, A.S., Liu, Y., Simpson, J.C., Mackeen, M., Stier, G., Gibson, T.J., Feizi, T., et al. (2008). Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein *N*-Glycosylation. *Mol. Biol. Cell* 19, 3404–3414.
- Schapiro, F.B., and Grinstein, S. (2000). Determinants of the pH of the Golgi complex. *J. Biol. Chem.* 275, 21025–21032.
- Schekman, R., and Orci, L. (1996). Coat proteins and vesicle budding. *Science* 271, 1526–1533.
- Schimpl, M., Schüttelkopf, A.W., Borodkin, V.S., and van Aalten, D.M.F. (2010). Human OGA binds substrates in a conserved peptide recognition groove. *Biochem. J.* 432, 1–7.
- Schimpl, M., Borodkin, V.S., Gray, L.J., and Van Aalten, D.M.F. (2012). Synergy of peptide and sugar in O-GlcNAcase substrate recognition. *Chem. Biol.* 19, 173–178.
- Schou, C., Rasmussen, G., Schulein, M., Henrissar, B., and Driguez, H. (1993). 4-Thiocellooligosaccharides : Their Synthesis and use as Inhibitors of Cellulases. *J. Carbohydr. Chem.* 12, 743–752.
- Schultz, J., and Pils, B. (2002). Prediction of structure and functional residues for O-GlcNAcase, a divergent homologue of acetyltransferases. *FEBS Lett.* 529, 179–182.
- Schulz, M.N., Landström, J., and Hubbard, R.E. (2013). MTSA—A Matlab program to fit thermal shift data. *Anal. Biochem.* 433, 43–47.
- Schweizer, A., Fransen, J.A., Bächli, T., Ginsel, L., and Hauri, H.P. (1988). Identification, by a monoclonal antibody, of a 53-kD protein associated with a tubulo-vesicular compartment at the cis-side of the Golgi apparatus. *J. Cell Biol.* 107, 1643–1653.
- Selnick, H., Hess, J.F., Tang, C., Liu, K., Schachter, J., Ballard, J.E., Marcus, J.N., Klein, D.J., Wang, X., Pearson, M., et al. (2019). Discovery of MK-8719, A Potent O-GlcNAcase Inhibitor as a Potential Treatment for Tauopathies. *J. Med. Chem.* [acs.jmedchem.9b01090](https://doi.org/10.1021/acs.jmedchem.9b01090).
- Sengupta, A., Kabat, J., Novak, M., Wu, Q., Grundke-Iqbal, I., and Iqbal, K. (1998). Phosphorylation of tau at both Thr 231 and Ser 262 is required for maximal inhibition of its binding to microtubules. *Arch. Biochem. Biophys.* 357, 299–309.
- Sernee, M.F., Ralton, J.E., Nero, T.L., Sobala, L.F., Kloehn, J., Vieira-Lara, M.A., Cobbold, S.A., Stanton, L., Pires, D.E.V., Hanssen, E., et al. (2019). A Family of Dual-Activity Glycosyltransferase-Phosphorylases Mediates Mannogen Turnover and Virulence in Leishmania Parasites. *Cell Host Microbe* 26, 385-399.e9.
- Shanmugasundaram, B., and Vasella, A. (2005). Synthesis of new C(2)-substituted gluco-configured tetrahydroimidazopyridines and their evaluation as glucosidase inhibitors. *Helv. Chim. Acta* 88, 2593–2602.
- Shanmugasundaram, B., Debowski, A.W., Dennis, R.J., Davies, G.J., Vocadlo, D.J., and Vasella, A. (2006). Inhibition of O-GlcNAcase by a gluco-configured nagstatin and a PUGNAc-imidazole hybrid inhibitor. *Chem. Commun.* 4372–4374.
- Sharon, N. (1986). IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature of glycoproteins, glycopeptides and peptidoglycans. Recommendations 1985. *Eur. J. Biochem.* 159, 1–6.

- Shashkov, A.S., Lipkind, G.M., and Kochetkov, N.K. (1986). Nuclear overhauser effects for methyl β -maltoside and the conformational states of maltose in aqueous solution. *Carbohydr. Res.* *147*, 175–182.
- Sheldon, W.L., Macauley, M.S., Taylor, E.J., Robinson, C.E., Charnock, S.J., Davies, G.J., Vocadlo, D.J., and Black, G.W. (2006). Functional analysis of a group A streptococcal glycoside hydrolase Spy1600 from family 84 reveals it is a beta-N-acetylglucosaminidase and not a hyaluronidase. *Biochem. J.* *399*, 241–247.
- Shen, D.L., Gloster, T.M., Yuzwa, S.A., and Vocadlo, D.J. (2012). Insights into O-linked N-acetylglucosamine (O-GlcNAc) processing and dynamics through kinetic analysis of O-GlcNAc transferase and O-GlcNAcase activity on protein substrates. *J. Biol. Chem.* *287*, 15395–15408.
- Shen, S., Chen, W., Dong, L., Yang, Q., Lu, H., and Zhang, J. (2018a). Design and synthesis of naphthalimide group-bearing thioglycosides as novel β -N-acetylhexosaminidases inhibitors. *J. Enzyme Inhib. Med. Chem.* *33*, 445–452.
- Shen, S., Dong, L., Chen, W., Zeng, X., Lu, H., Yang, Q., and Zhang, J. (2018b). Modification of the Thioglycosyl-Naphthalimides as Potent and Selective Human O-GlcNAcase Inhibitors. *ACS Med. Chem. Lett.* *9*, 1241–1246.
- Singh, J.P., Qian, K., Lee, J.-S., Zhou, J., Han, X., Zhang, B., Ong, Q., Ni, W., Jiang, M., Ruan, H.-B., et al. (2019). O-GlcNAcase targets pyruvate kinase M2 to regulate tumor growth. *Oncogene*.
- Sinnott, M.L. (1987). Glycosyl group transfer. In *Enzyme Mechanisms*, M.I. Page, and A. Williams, eds. (London: Royal Society of Chemistry), pp. 259–297.
- Sinnott, M.L. (1988). The Principle of Least Nuclear Motion and the Theory of Stereoelectronic Control. *Adv. Phys. Org. Chem.* *24*, 113–204.
- Sinnott, M.L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* *90*, 1171–1202.
- Śłomińska-Wojewódzka, M., and Sandvig, K. (2015). The Role of Lectin-Carbohydrate Interactions in the Regulation of ER-Associated Protein Degradation. *Molecules* *20*, 9816–9846.
- Smet-Nocca, C., Broncel, M., Wieruszkeski, J.-M., Tokarski, C., Hanouille, X., Leroy, A., Landrieu, I., Rolando, C., Lippens, G., and Hackenberger, C.P.R. (2011). Identification of O-GlcNAc sites within peptides of the Tau protein and their impact on phosphorylation. *Mol. Biosyst.* *7*, 1420.
- Smith, S.M., Struyk, A., Jonathan, D., Declercq, R., Marcus, J., Toolan, D., Wang, X., Schachter, J.B., Cosden, M., Pearson, M., et al. (2016). EARLY CLINICAL RESULTS AND PRECLINICAL VALIDATION OF THE O-GLCNACASE (OGA) INHIBITOR MK-8719 AS A NOVEL THERAPEUTIC FOR THE TREATMENT OF TAUOPATHIES. *Alzheimer's Dement.* *12*, P261.
- Sobala, L.F., Speciale, G., Zhu, S., Raich, L., Sannikova, N., Thompson, A.J., HAKKI, Z., Lu, D., Abadi, S.S.K., Lewis, A.R., et al. (2019). An Epoxide Intermediate in Glycosidase Catalysis. *ChemRxiv*.
- Sousa, M.C., Ferrero-Garcia, M.A., and Parodi, A.J. (1992). Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* *31*, 97–105.
- Soussillane, P., D'Alessio, C., Paccalet, T., Fitchette, A.C., Parodi, A.J., Williamson, R., Plasson, C., Faye, L., and Gomord, V. (2009). N-glycan trimming by glucosidase II is essential for Arabidopsis development. *Glycoconj. J.* *26*, 597–607.
- Speciale, G., Thompson, A.J., Davies, G.J., and Williams, S.J. (2014). Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* *28*, 1–13.
- Sponslor, O.L., and Dore, W.H. (1926). The Structure of Ramie Cellulose as Derived from X-Ray Data. In *Colloid Symposium Monograph*, H.B. Weiser, ed. (New York), pp. 174–202.
- Stacevičienė, I., Petraitienė, S., Vaičiūnienė, D., Alasevičius, T., Kiršlienė, J., and Usonis, V. (2016). Antibiotic resistance of *Streptococcus pneumoniae*, isolated from nasopharynx of preschool children with acute

respiratory tract infection in Lithuania. *BMC Infect. Dis.* 16, 216.

Stachelek, C., Stachelek, J., Swan, J., Botstein, D., and Konigsberg, W. (1986). Identification, cloning and sequence determination of the genes specifying hexokinase A and B from yeast. *Nucleic Acids Res.* 14, 945–963.

Stanley, P., Schachter, H., and Taniguchi, N. (2009). Chapter 8. N-Glycans, *Essentials of Glycobiology*, 2nd Edition.

Stenger, J., Cowman, M., Eggers, F., Eyring, E.M., Petrucci, S., and Kaatze, U. (2002). Molecular Dynamics and Kinetics of Monosaccharides in Solution. A Broadband Ultrasonic Relaxation Study. *J. Phys. Chem. B* 104, 4782–4790.

Stoddart, J.F. (1971). *Stereochemistry of Carbohydrates* (wiley).

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

Suits, M.D.L., Zhu, Y., Taylor, E.J., Walton, J., Zechel, D.L., Gilbert, H.J., and Davies, G.J. (2010). Structure and Kinetic Investigation of *Streptococcus pyogenes* Family GH38 α -Mannosidase. *PLoS One* 5, e9006.

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

Tabas, I., and Kornfeld, S. (1979). Purification and Characterization of a Rat Liver Golgi α -Mannosidase Capable of Processing Asparagine-linked Oligosaccharides*. *J. Biol. Chem.* 254, 11655–11663.

Tailford, L.E., Offen, W. a, Smith, N.L., Dumon, C., Morland, C., Gratien, J., Heck, M.-P., Stick, R. V, Blériot, Y., Vasella, A., et al. (2008). Structural and biochemical evidence for a boat-like transition state in beta-mannosidases. *Nat. Chem. Biol.* 4, 306–312.

Takahashi, M., Inoue, N., Ohishi, K., Maeda, Y., Nakamura, N., Endo, Y., Fujita, T., Takeda, J., and Kinoshita, T. (1996). PIG-B, a membrane protein of the endoplasmic reticulum with a large luminal domain, is involved in transferring the third mannose of the GPI anchor. *EMBO J.* 15, 4254–4261.

Tanichi, N., and Korekane, H. (2011). Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Rep.* 44, 772–781.

Taron, B.W., Colussi, P.A., Wiedman, J.M., Orlean, P., and Taron, C.H. (2004). Human Smp3p Adds a Fourth Mannose to Yeast and Human Glycosylphosphatidylinositol Precursors *in Vivo*. *J. Biol. Chem.* 279, 36083–36092.

Tempel, W., Karaveg, K., Liu, Z.-J., Rose, J., Wang, B.-C., and Moremen, K.W. (2004). Structure of mouse Golgi α -mannosidase IA reveals the molecular basis for substrate specificity among class 1 (family 47 glycosylhydrolase) α 1,2-mannosidases. *J. Biol. Chem.* 279, 29774–29786.

Terinek, M., and Vasella, A. (2005). Synthesis of N-acetylglucosamine-derived nagstatin analogues and their evaluation as glycosidase inhibitors. *Helv. Chim. Acta* 88, 10–22.

Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J., and Dijkstra, B.W. (1994). Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. *Structure* 2, 1181–1189.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 293, 498–506.

Tews, I., Terwisscha Van Scheltinga, A.C., Perrakis, A., Wilson, K.S., and Dijkstra, B.W. (1997). Substrate-assisted catalysis unifies two families of chitinolytic enzymes. *J. Am. Chem. Soc.* 119, 7954–7959.

- Thompson, A.J., Dabin, J., Iglesias-Fernández, J., Ardèvol, A., Dinev, Z., Williams, S.J., Bande, O., Siriwardena, A., Moreland, C., Hu, T.-C., et al. (2012). The Reaction Coordinate of a Bacterial GH47 α -Mannosidase: A Combined Quantum Mechanical and Structural Approach. *Angew. Chemie Int. Ed.* *51*, 10997–11001.
- Thorsness, P.E., and Koshland, D.E. (1987). Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. *J. Biol. Chem.* *262*, 10422–10425.
- Thunnissen, A.-M.W.H., Dijkstra, A.J., Kalk, K.H., Rozeboom, H.J., Engel, H., Keck, W., and Dijkstra, B.W. (1994). Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature* *367*, 750–753.
- Toleman, C., Paterson, A.J., Whisenhunt, T.R., and Kudlow, J.E. (2004). Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *J. Biol. Chem.* *279*, 53665–53673.
- Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* *259*, 3308–3317.
- Tropak, M.B., Blanchard, J.E., Withers, S.G., Brown, E.D., and Mahuran, D. (2007). High-throughput screening for human lysosomal beta-N-Acetyl hexosaminidase inhibitors acting as pharmacological chaperones. *Chem. Biol.* *14*, 153–164.
- Tu, H.-C., Hsiao, Y.-C., Yang, W.-Y., Tsai, S.-L., Lin, H.-K., Liao, C.-Y., Lu, J.-W., Chou, Y.-T., Wang, H.-D., and Yuh, C.-H. (2017). Up-regulation of golgi α -mannosidase IA and down-regulation of golgi α -mannosidase IC activates unfolded protein response during hepatocarcinogenesis. *Hepatol. Commun.* *1*, 230–247.
- Ubersax, J.A., and Ferrell, J.E. (2007). Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* *8*, 530–541.
- Unione, L., Xu, B., Díaz, D., Martín-Santamaría, S., Poveda, A., Sardinha, J., Rauter, A.P., Blériot, Y., Zhang, Y., Cañada, F.J., et al. (2015). Conformational Plasticity in Glycomimetics: Fluorocarbamethyl- L -idopyranosides Mimic the Intrinsic Dynamic Behaviour of Natural Idose Rings. *Chem. - A Eur. J.* *21*, 10513–10521.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *60*, 2184–2195.
- Vaidyanathan, K., Niranjana, T., Selvan, N., Teo, C.F., May, M., Patel, S., Weatherly, B., Skinner, C., Opitz, J., Carey, J., et al. (2017). Identification and characterization of a missense mutation in the O-linked β -N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* *292*, 8948–8963.
- Vallée, F., Lipari, F., Yip, P., Sleno, B., Herscovics, A., and Howell, P.L. (2000a). Crystal structure of a class I alpha1,2-mannosidase involved in N-glycan processing and endoplasmic reticulum quality control. *EMBO J.* *19*, 581–588.
- Vallée, F., Karaveg, K., Herscovics, A., Moremen, K.W., and Howell, P.L. (2000b). Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases. *J. Biol. Chem.* *275*, 41287–41298.
- Valverde, P., Quintana, J.I., Santos, J.I., Ardá, A., and Jiménez-Barbero, J. (2019). Novel NMR Avenues to Explore the Conformation and Interactions of Glycans. *ACS Omega* *4*, 13618–13630.
- Varki, A. (1999). *Essentials of glycobiology* (Cold Spring Harbor Laboratory Press).
- Varki, A. (2017). Biological roles of glycans. *Glycobiology* *27*, 3–49.
- Varki, A., and Angata, T. (2005). Siglecs—the major subfamily of I-type lectins. *Glycobiology* *16*, 1R-27R.

- Varki, A., and Marth, J. (1995). Oligosaccharides in vertebrate development. *Semin. Dev. Biol.* 6, 127–138.
- Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., et al. (2015). Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25, 1323–1324.
- Vasella, A., Davies, G.J., and Böhm, M. (2002). Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* 6, 619–629.
- Vavra, J.J., Deboer, C., Dietz, A., Hanka, L.J., and Sokolski, W.T. (1960). Streptozotocin, a new antibacterial antibiotic. *Antibiot. Annu.* 7, 230–235.
- Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.* 9, 944–957.
- Vocadlo, D.J. (2012). O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and enzyme regulation. *Curr. Opin. Chem. Biol.* 16, 488–497.
- Vocadlo, D.J., and Davies, G.J. (2008). Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* 12, 539–555.
- Vyas, N.K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* 1, 732–740.
- Walker, G.J., and Pulkownik, A. (1973). Degradation of dextrans by an α -1,6-glucan glucohydrolase from *Streptococcus mitis*. *Carbohydr. Res.* 29, 1–14.
- Wang-Gillam, A., Pastuszak, I., and Elbein, A.D. (1998). A 17-amino acid insert changes UDP-N-acetylhexosamine pyrophosphorylase specificity from UDP-GalNAc to UDP-GlcNAc. *J. Biol. Chem.* 273, 27055–27057.
- Wang, F., Song, W., Brancati, G., and Segatori, L. (2011). Inhibition of Endoplasmic Reticulum-associated Degradation Rescues Native Folding in Loss of Function Protein Misfolding Diseases. *J. Biol. Chem.* 286, 43454–43464.
- Wang, Z., Udeshi, N.D., O'Malley, M., Shabanowitz, J., Hunt, D.F., and Hart, G.W. (2010). Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* 9, 153–160.
- Watzel, G., and Tanner, W. (1989). Cloning of the glutamine:fructose-6-phosphate amidotransferase gene from yeast. Pheromonal regulation of its transcription. *J. Biol. Chem.* 264, 8753–8758.
- Weaver, L.H., Grütter, M.G., and Matthews, B.W. (1995). The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the “goose-type” lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* 245, 54–68.
- Weingarten, M.D., Lockwood, A.H., Hwo, S.Y., and Kirschner, M.W. (1975). A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci.* 72, 1858–1862.
- White, A., and Rose, D.R. (1997). Mechanism of catalysis by retaining β -glycosyl hydrolases. *Curr. Opin. Struct. Biol.* 7, 645–651.
- Whitworth, G.E., Macauley, M.S., Stubbs, K.A., Dennis, R.J., Taylor, E.J., Davies, G.J., Greig, I.R., and Vocadlo, D.J. (2007). Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: mechanistic and structural insights into inhibitor selectivity and transition state poise. *J. Am. Chem. Soc.* 129, 635–644.
- Wilkins, M.R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F., and Appel, R.D. (1997). Detailed peptide characterization using PEPTIDEMASS - A World-Wide-Web-accessible tool. *Electrophoresis* 18, 403–408.

- Williams, R.J., Iglesias-Fernández, J., Stepper, J., Jackson, A., Thompson, A.J., Lowe, E.C., White, J.M., Gilbert, H.J., Rovira, C., Davies, G.J., et al. (2014). Combined inhibitor free-energy landscape and structural analysis reports on the mannosidase conformational coordinate. *Angew. Chemie - Int. Ed.* *53*, 1087–1091.
- Winchester, B., and Fleet, G.W.J. (1992). Amino-sugar glycosidase inhibitors: Versatile tools for glycobiologists. *Glycobiology*.
- Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS Refinement in REFMAC at Moderate Resolutions. *Methods Enzymol.* *374*, 300–321.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *67*, 235–242.
- Winter, G. (2010). Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* *43*, 186–190.
- Withers, S.G., and Williams, S.J. (2017). Glycoside hydrolases.
- Wolfenden, R., Lu, X., and Young, G. (1998). Spontaneous hydrolysis of glycosides. *J. Am. Chem. Soc.* *120*, 6814–6815.
- Woods, R.J. (1998). Computational carbohydrate chemistry: what theoretical methods can tell us. *Glycoconj. J.* *15*, 209–216.
- Wormald, M.R., Petrescu, A.J., Pao, Y.L., Glithero, A., Elliott, T., and Dwek, R.A. (2002). Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem. Rev.* *102*, 371–386.
- Wrodnigg, T.M., Steiner, A.J., and Ueberbacher, B.J. (2008). Natural and synthetic iminosugars as carbohydrate processing enzyme inhibitors for cancer therapy. *Anticancer. Agents Med. Chem.* *8*, 77–85.
- Wu, M.M., Llopis, J., Adams, S., McCaffery, J.M., Kulomaa, M.S., Machen, T.E., Moore, H.P., and Tsien, R.Y. (2000). Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* *7*, 197–209.
- Xiang, Y., Karaveg, K., and Moremen, K.W. (2016). Substrate recognition and catalysis by GH47 α -mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway. *Proc. Natl. Acad. Sci.* *113*, E7890–E7899.
- Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L. V., and Gordon, J.I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* *299*, 2074–2076.
- Yang, Y.R., Song, M., Lee, H., Jeon, Y., Choi, E.-J., Jang, H.-J., Moon, H.Y., Byun, H.-Y., Kim, E.-K., Kim, D.H., et al. (2012). O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell* *11*, 439–448.
- Ying, H., Kimmelman, A.C., Lyssiotis, C.A., Hua, S., Chu, G.C., Fletcher-Sananikone, E., Locasale, J.W., Son, J., Zhang, H., Coloff, J.L., et al. (2012). Oncogenic Kras Maintains Pancreatic Tumors through Regulation of Anabolic Glucose Metabolism. *Cell* *149*, 656–670.
- Yu, C., Crispin, M., Sonnen, A.F.-P., Harvey, D.J., Chang, V.T., Evans, E.J., Scanlan, C.N., Stuart, D.I., Gilbert, R.J.C., and Davis, S.J. (2011). Use of the α -mannosidase I inhibitor kifunensine allows the crystallization of apo CTLA-4 homodimer produced in long-term cultures of Chinese hamster ovary cells. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* *67*, 785–789.
- Yuzwa, S.A., Macauley, M.S., Heinonen, J.E., Shan, X., Dennis, R.J., He, Y., Whitworth, G.E., Stubbs, K.A., McEachern, E.J., Davies, G.J., et al. (2008). A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nat. Chem. Biol.* *4*, 483–490.
- Yuzwa, S.A., Yadav, A.K., Skorobogatko, Y., Clark, T., Vosseller, K., and Vocadlo, D.J. (2011). Mapping O-

GlcNAc modification sites on tau and generation of a site-specific O-GlcNAc tau antibody. *Amino Acids* 40, 857–868.

Yuzwa, S.A., Shan, X., Macauley, M.S., Clark, T., Skorobogatko, Y., Vosseller, K., and Vocadlo, D.J. (2012). Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* 8, 393–399.

Zähner, D., and Hakenbeck, R. (2000). The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J. Bacteriol.* 182, 5919–5921.

Závodszky, P., and Hajdú, I. (2013). Evolution of the concept of conformational dynamics of enzyme functions over half of a century: A personal view. *Biopolymers* 99, 263–269.

Zechel, D.L., and Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 33, 11–18.

Zhong, W., Kuntz, D.A., Ember, B., Singh, H., Moremen, K.W., Rose, D.R., and Boons, G.-J. (2008). Probing the Substrate Specificity of Golgi α -Mannosidase II by Use of Synthetic Oligosaccharides and a Catalytic Nucleophile Mutant. *J. Am. Chem. Soc.* 130, 8975–8983.

Zhu, Y., Suits, M.D.L., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A., et al. (2010). Mechanistic insights into a calcium-dependent family of alpha-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* 6, 125–132.

Zhu, Y., Shan, X., Safarpour, F., Erro Go, N., Li, N., Shan, A., Huang, M.C., Deen, M., Holicek, V., Ashmus, R., et al. (2018). Pharmacological Inhibition of O-GlcNAcase Enhances Autophagy in Brain through an mTOR-Independent Pathway. *ACS Chem. Neurosci.* 9, 1366–1379.

Zierke, M., Smieško, M., Rabbani, S., Aeschbacher, T., Cutting, B., Allain, F.H.-T., Schubert, M., and Ernst, B. (2013). Stabilization of Branched Oligosaccharides: Lewis \times Benefits from a Nonconventional C–H \cdots O Hydrogen Bond. *J. Am. Chem. Soc.* 135, 13464–13472.

Zou, J.Y., Kleywegt, G.J., Ståhlberg, J., Driguez, H., Nerinckx, W., Claeysens, M., Koivula, A., Teeri, T.T., and Jones, T.A. (1999). Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure* 7, 1035–1045.

(1980). Conformational Nomenclature for Five and Six-Membered Ring Forms of Monosaccharides and Their Derivatives: Recommendations 1980. *Eur. J. Biochem.* 111, 295–298.

(2008). Haworth representation. In *IUPAC Compendium of Chemical Terminology*, (Research Triangle Park, NC: IUPAC), p.