Structural, mechanistic and functional characterization of glycoside hydrolases of family GH99

Łukasz Filip Sobala

Doctor of Philosophy

University of York

Chemistry

March 2018

Abstract

Glycosylation is a very common post-translational modification and the glycans can be attached to oxygen (*O*-linked), nitrogen (*N*-linked) or carbon (*C*-linked). *N*-linked glycosylation has implications for protein folding and is also essential in viral infectivity and cell-cell signalling. *Endo*-α-1,2-mannosidase from family GH99 is a unique enzyme within the *N*-glycosylation pathway as it is the only one which does not cleave the terminal sugar from the reducing end of the glycan, but instead releases an α-Glc-1,3-Man disaccharide, with overall retention of stereochemistry at the anomeric carbon. Previously it was proposed that GH99 *endo*-acting mannosidases and mannanases proceed through a neighbouring group participation mechanism with a 1,2-anhydrosugar as a reaction intermediate. This Thesis contains evidence supporting this hypothesis. Chapter 2 presents structures of the bacterial GH99 with its substrate, with mimics of the reaction intermediate and with the products of the reaction. Kinetic and structural data on various intermediate mimics show that the compound whose structure is the closest to the intermediate is turned over by the enzyme. In Chapter 3, analysis of different designs of GH99 inhibitors and their conformation on-enzyme is presented. Chapter 4 presents purification and solution of the crystal structure of the catalytic domain of the human endomannosidase (MANEA). Multiple crystal forms were obtained, which made it possible to look at the conformation of a feature present in the eukaryotic but not bacterial GH99: a loop spanning residues 191–201. This loop was disordered when no ligand was present in the –2/–1 sites, and ordered when these sites were occupied. Chapter 5 explores attempts at producing MANEAL, a paralog of MANEA which is found in bony vertebrates. The Thesis concludes with an analysis of the phylogeny of endomannosidase genes and perspectives for future research: studies of endomannosidase in mammalian model organisms are needed to understand its significance.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I would like to thank my supervisor, Gideon Davies, for trusting me with this project. I was tasked with two ambitious strands of a large grant, which was, at first, very humbling. He was very understanding of my periodic frustrations and helped me find the right path when I strayed, which also happened occasionally. The time spent here in York was filled with surprises and changes in perception, and I am sure I would never grow as a person as much if I had not been allowed so much time to listen and think.

The project would have never gone very far if it had not been for the kind, patient supervision and help of Dr Andrew Thompson, who worked with this system before I started. In the first 6 months of the project he taught me techniques that were useful on almost every day after he left. The collection of reagents he left was indispensable to me during the whole period after his departure, and they all were labelled clearly. Thank you for being so considerate.

I was always welcomed with great hospitality by the collaborators I have visited, and a big part of this Thesis would not exist without these collaborations. Meeting you at your labs or conferences was always a wonderful experience. I am grateful to all the intrepid synthetic chemists who sent me compounds embodying months of their work. Again, this Thesis would never be what it is without their work.

I would like to express my gratitude to the current and previous members of the Davies group who helped me and kept life interesting along the years: Dr Imogen Breen, Dr Marc Cabry, Claire Fowler, Dr Rebecca Gregory, Dr Glyn Hemsworth, Dr Yi Jin, Dr Esther Johnston, Alexandra Males, Wendy Offen, Dr Christian Roth, Rhianna Rowland, Dr Daniel Wright, Dr Liang Wu and Dr Saioa Urresti.

I am grateful to all the people in YSBL, especially the technical staff who kept the laboratory going, the X-ray technicians whose experience in running the facility is invaluable, Dr Jon Agirre and Professor Eleanor Dodson who helped me with many questions about crystallography, Dr Stuart McNicholas for his support with ccp4i2, Dr Tim Ganderton who assisted me with my work on mammalian cells and Professor Andrzej Brzozowski for having an excellent critical eye as an Independent Panel Member.

Finally I thank my family who was always there for me.

Author's declaration

I declare that I am the sole author of this Thesis and all research detailed throughout this Thesis was carried out by me, with the exception of collaborative works which are listed below and stated in the text. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

- All X-ray datasets were collected remotely by Sam Hart and Dr Johan Turkenburg.
- The first structure of *Bx*GH99 inactive mutant with Man_4OMe was produced by Dr Andrew Thompson. The structure reported in Chapter 2 is the work of the author and was solved at a higher resolution.
- Compound Man_4OMe used in Chapter 2 was synthesized at Spencer Williams laboratory by Dr Zalihe Hakki and Dr Gaetano Speciale.
- QM/MM calculations reported in Chapter 2 were done by Lluis Raich and Professor Carme Rovira.
- Cloning of genes encoding *Bt*GH99 protein used in Chapter 2 and Chapter 3 was carried out by Dr Tracey Gloster before the commencement of this Thesis work.
- Crystal fishing for structures of *Bx*GH99 with compounds Glc-β-aziridine and Glc-cyclohexene reported in Chapter 2, Chapter 3 was done by Dr Andrew Thompson.
- Cloning and mutagenesis of genes encoding *Bx*GH99 used in Chapter 2 and Chapter 3 were performed by Dr Andrew Thompson before the commencement of this Thesis work.
- Compounds Glc-β-epoxide, Glc-α-epoxide, Glc-β-aziridine, Glc-α-aziridine and Glc-cyclohexene reported in Chapter 2 and Chapter 3 were synthesized by Sha Zhu and Dan Lu at Matthieu Sollogoub laboratory.
- Compounds Man-D-glucal, ManddMan, and $GlcMan_3OMe$ used in Chapter 2,

Chapter 3 and Chapter 5 were synthesized by Pearl Fernandes.

- Compounds ManNOE, ManADMJ and ManManIm used in Chapter 2, Chapter 3 and Chapter 5 were synthesized by Marija Petricevic.
- Analysis of the results of NMR titrations and addition of a number of data points to the graphs reported in Chapter 3 were done by Dr Ganeko Bernardo-Seisdedos.
- Subcloning of the *Hs*MANEA gene optimized for bacterial expression that was used as a template for truncations reported in Chapter 4 was done by Dr Andrew Thompson before the commencement of this Thesis work.
- The restraints dictionary used in the refinement of structures containing the $[TeW_6O_{24}]^{6-}$ molecule presented in Chapter 4 was made by Dr Jon Agirre.
- The GlcIFG, ManIFG and BGlcIFG inhibitors used in Chapter 4 and Chapter 5 were synthesized at Spencer Williams laboratory by Dr Zalihe Hakki.
- Preparation of sugars for MS analysis, running of the samples and data presentation reported in Chapter 5 was done by Rachel Bates.
- The experiments involving BVDV reported in Chapter 5 were performed by Dr Dominic Alonzi.
- The BT3990 enzyme used in the linked assay in Chapter 5 was purified by Dr Michael Suits before the commencement of this Thesis work.
- Biohazard level III work and immunostaining of Huh7 cells in the experiments reported in Chapter 5 were performed by Scott Davies and Dr Zania Stamataki.

Chapter 1

Introduction

1.1 Why are sugars and glycans interesting? Carbohydrates, their roles and diversity

Carbohydrates are a ubiquitous group of biomolecules, found in all of the known life forms. They share this status with nucleic acids, proteins and lipids[1] (nucleic acids themselves incorporate carbohydrates, ribose or deoxyribose, as a part of their backbone structure). The specific roles of carbohydrates are too numerous to quote here, but four general categories have been proposed[2] and are depicted in Figure 1.1. Below I will list the categories and explain them briefly. The field of science that is concerned with biological roles of sugars is glycobiology. The rate of progress in the field of glycobiology means that new roles are discovered frequently, but they are likely to belong these broad categories. The detailed definition of a carbohydrate will be presented in Section 1.2. For the purpose of this Section, a working definition of "organic compounds with a formula $C_nH_{2n}O_n$ and the oligo- and polymers of these compounds" will suffice.

- **Structural and modulatory roles** – these encompass functions relating to the physical structure of a life form and the consequences of this structure, such as its resistance to shear stress, viscosity, diffusion, protection from degradation by enzymes, acquisition and storage of nutrients, promotion of protein folding,[3] developmental gradient generation,[4] histone modifications and others.
- **Interspecies recognition roles** – glycans in a host organism might be recognized by other organisms. A host can in turn protect itself by taking advantage of the

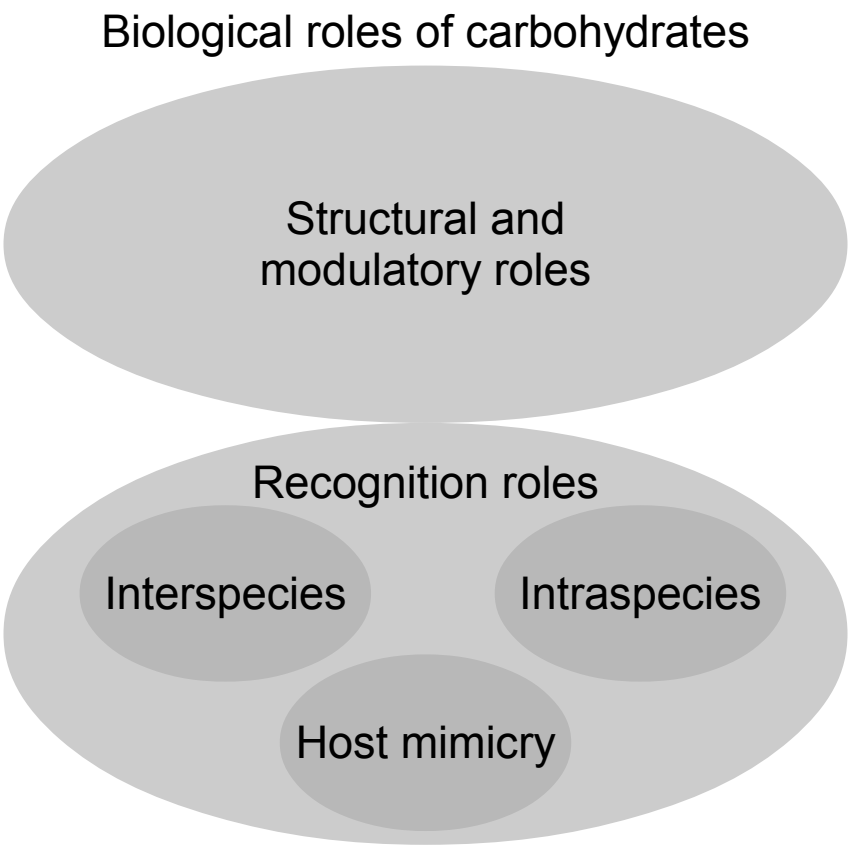


Figure 1.1: Biological roles of carbohydrates – a diagram.

information about its parasite and producing carbohydrate-based "decoys" (for example, the surface of erythrocytes, which do not have the machinery to replicate viruses).[5,6] Milk oligosaccharides are another example of host-protective carbohydrates.[7] Glycans can also be recognized as antigens.[8]

- **Intraspecies recognition roles** – processes like intra- and intercellular signaling, interactions with the extracellular matrix and fertilization are often mediated by glycans. The ABO blood group system is an effect of genetic polymorphisms leading to various glycans present on erythrocytes[9].
- **Molecular mimicry of the host** – bacteria in the gut or pathogens can evolve to synthesize host glycans in order to influence the host immune system.[10] Viruses, as they reproduce through selfish utilization of the host glycan synthesis machinery, commonly contain glycoproteins with glycans similar to these of a host.

In addition to the multiplicity of their roles, carbohydrates are a structurally diverse group of biomolecules (see the inset in Figure 1.2). Some monosugars, such as D-glucose, are more common than others. But the structures of monosaccharides from

various organisms are not as evolutionarily conserved as those of nucleotides and amino acids. DNA is a polymer of four nucleotides and only two more are encountered ubiquitously (uracil in RNA, and nicotinamide in NAD (nicotinamide adenine dinucleotide)). As few as 20 amino acids make up most proteins. In contrast, the number of monosaccharides encountered is higher and constantly growing. The primary sequence of nucleic acids and polypeptides is encoded genetically, but the sequence of polysaccharides is an effect of the concerted action of enzymes specific towards different substrates. This adds to glycan diversity.

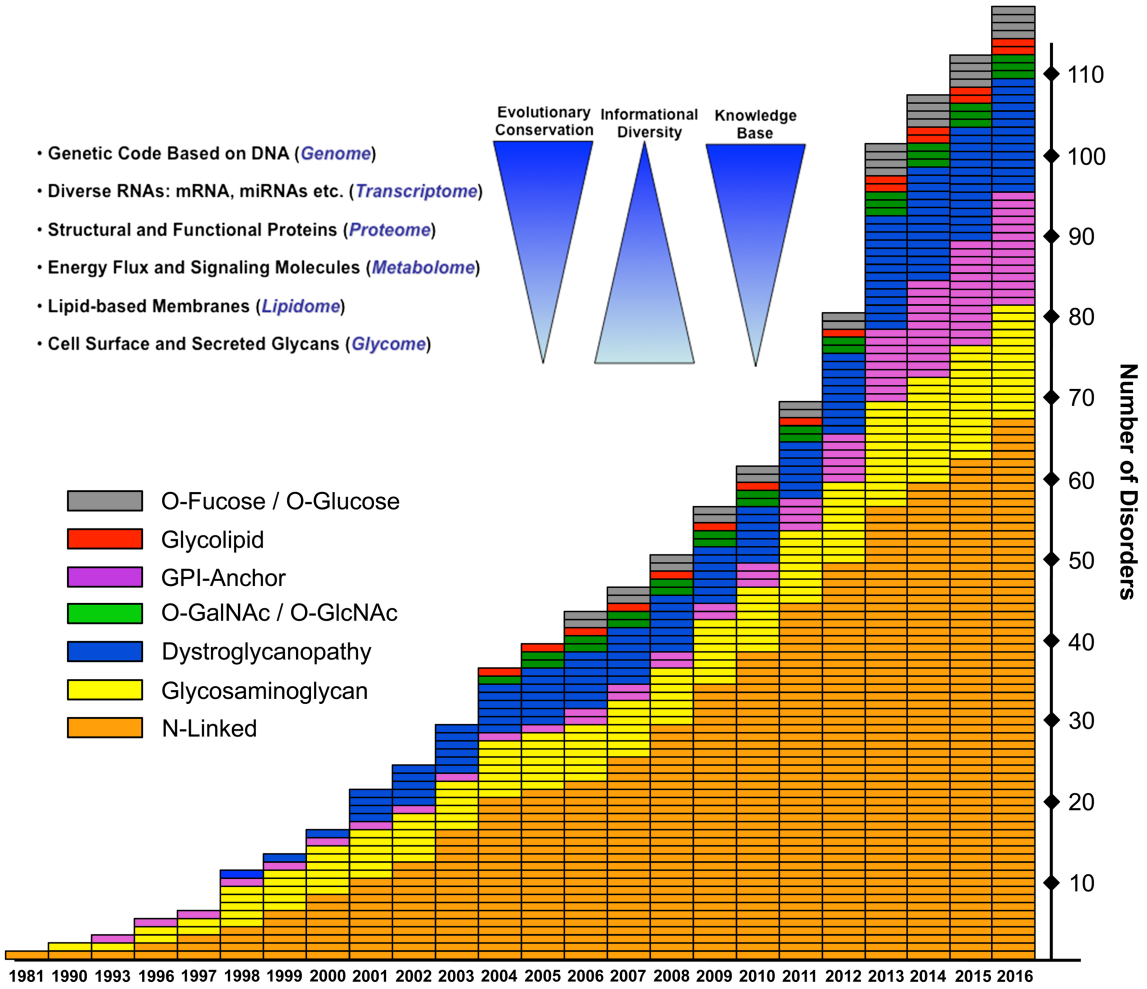


Figure 1.2: The number of human glycosylation disorders known in the years stated and the processes they relate to. Inset: A representation of the diversity, conservation and the knowledge base of different biomolecules. Glossary: GPI – glycosylphosphatidylinositol, GalNAc – *N*-acetylgalactosamine, GlcNAc – *N*-acetylglucosamine. Figure adapted from Ref. 2.

In living organisms, it is proteins that play the roles of recognition sensors of glycans and catalyze the reactions involving glycans. Interspecies recognition of glycans is mediated by proteins such as adhesins, viral agglutinins, toxic lectins[11] or pathogen glycosidases. Host killer proteins recognize the xenosugars that are present on a

parasite.[12] In the process of *N*-glycosylation (which will be discussed in detail in Section 1.4), chaperone lectins, glycosidases and glycosyltransferases recognize specific glycan structures. This is an example of intraspecies recognition. The proteins that recognize sugars have various functions: lectins bind to glycan "epitopes" but do not catalyze the cleavage of any bonds, glycoside hydrolases hydrolyze bonds between sugars (glycosidic bonds), and glycosyltransferases elongate the sugar chain using sugar donor molecules. These and other categories will be explained in more detail in Section 1.3. Many of the drugs that are produced using biotechnology are glycoproteins, which reflects the importance of glycan recognition.

Each of the multiple copies of a protein molecule in a single organism can have different *N*-glycans attached after full processing inside the cell. Apart from the variation in their sequence, oligo- and polymers of saccharides are often branched or functionalized at their OH groups. In 2002, 41 different types of linkages between a glycoside moiety and a polypeptide chain were known.[13] Over a hundred human disorders have been linked to aberrations in glycosylation (Figure 1.2).[2] It is evident that glycobiology is a broad field of study, in which it is only feasible to track progress in the narrower areas of one's interest. Carbohydrates have been called "the last frontier of cell and molecular biology" which is "rapidly being conquered".[14]

The following Introduction will focus on the area of glycobiology relevant to work presented in this Thesis. First, carbohydrates and the ways to analyze their conformation will be introduced (Section 1.2), followed by the classification system of carbohydrate active enzymes into families and other categories (Section 1.3). These concepts are crucial for understanding why a structural, mechanistic and functional characterization of glycoside hydrolases of family GH99 was conducted, and the significance of the obtained results. An in-depth look at the process on *N*-glycosylation in the ER (endoplasmic reticulum), the Golgi apparatus and related organelles will follow (Section 1.4). Family GH99 glycoside hydrolases participate in this pathway, and the current state of knowledge about GH99 enzymes will be presented in Section 1.4.5. The Introduction will close with a discussion of the potential relevance of GH99-targeting therapies on human health (Section 1.5), and a summary of project aims (Section 1.6).

1.2 Chemistry and conformation of carbohydrates

Carbohydrates as biomolecules have unique properties that stem from their chemical structure. Some of these properties are specific to and stem from a sugar being in a ring form. The sugar ring can assume different shapes, called conformations. Carbohydrate active enzymes take advantage of the changes in the chemical properties of sugars that stem from changes in their conformations. This is often the basis of their mechanism of catalysis. Here, the various effects that give rise to certain conformations, and general theory of the sugar ring conformation.

Carbohydrates were originally defined as *hydrates of carbon*, i.e. compounds with the formula $C_nH_{2n}O_n$.[15] The definition has evolved, now encompassing structures with a slightly different general formula, such as deoxy sugars, unsaturated sugars, sugar acids and derivatives such as *N*-acetylglucosamine, with at least three carbon atoms. In their chemical structure, sugars are polychiral molecules: more than one carbon in their structure has four different substituents. A convention for numbering the carbon atoms in carbohydrates exists. The numbering of atoms in the sugar chain is such that the carbon with an attached aldehydo or keto group has the lowest number (1 or ≥2, respectively). The absolute configuration at the chiral carbon assigned the highest number determines whether the sugar is D- or L-. If the OH group is on the right when the structure is drawn in the Fischer projection it is a D-sugar and if on the left, an L-sugar (see Figure 1.3). The majority of sugars in biology are D-sugars.

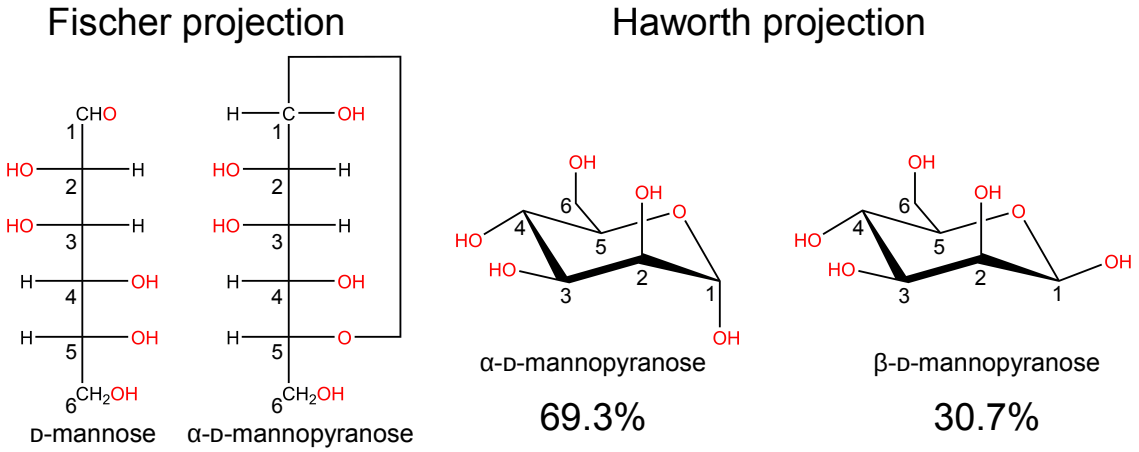


Figure 1.3: Structure of mannose, an example of a monosaccharide, in acyclic and cyclic forms. The equivalent Fischer and Haworth projections of α-D-mannopyranose are shown. The percentages refer to anomeric equilibrium in water at 20 °C.[16]

Most sugars can exist in a linear form or a cyclic form. When an OH group from

another carbon in the chain performs a nucleophilic attack at the carbonyl carbon, the product of this reaction is formation of a hemiacetal/hemiketal and a new chiral centre at the former carbonyl carbon. Biologically, the most important sugar rings are five-membered (furanoses) and six-membered (pyranoses). The new chiral centre determines the anomer of the sugar, which can be α or β. When drawn in the Fischer projection the α-anomer has the OH1 on the same side as the reference OH describing the sugar absolute configuration, and the β-anomer on the opposite side. Cyclization is reversible – the anomeric configuration can flip in a process called mutarotation. A temperature-dependent equilibrium between anomers is achieved after the cyclic sugar spends enough time in solution.[17]

Other effects of cyclization relate to the characteristics of the five- and six-membered rings. As with cyclohexane, the conformation of the ring is constrained by preference for ideal bond angles. Another contributor to sugar ring conformation is the bulkiness of substituents. Large substituents (OH, CH_2OH groups) usually prefer the *equatorial* (close to the mean sugar plane) position, and small substituents (hydrogens) prefer the *axial* position. For example, in β-D-glucopyranose all the OH groups are in equatorial positions when the sugar assumes a $^4\boldsymbol{C}_1$ chair conformation. When α-D-mannopyranose is in the same $^4\boldsymbol{C}_1$ chair conformation, only OH1 and OH2 are in axial positions. This preference of equatorial over axial position is partly a consequence of varying energetic minima when substituents of different size are bound to two adjacent carbons (they can be eclipsed or staggered).[18] An unusual property of pyranoses and furanoses is that, counterintuitively, the OH group connected to the anomeric carbon often prefers the axial position. This is called the anomeric effect. Traditionally, the anomeric effect was explained by dipole moment repulsion.[19] Recent computational studies point out that most of the anomeric effect can be explained by steric hindrance and classical electrostatic interactions.[20,21] As shown in Figure 1.3, in standard conditions D-mannopyranose prefers the axial α-anomer. Anomers can differ in surprising ways, for example the taste of α-D-mannose is perceived as sweet by humans and the taste of β-D-mannose as bitter.[22] The anomeric carbon, bound to two oxygen atoms, is the most reactive part of a cyclic sugar. Nature takes advantage of this by linking sugar moieties together through the OH1 group and thus creating disaccharides, trisaccharides and higher order oligosaccharides, as well as polysaccharides. The catalysis of these reactions is performed by glycosyltransferases. In

some cases, glycoside hydrolases can be engineered to catalyze formation, not hydrolysis, of glycosidic bonds. These enzymes are referred to as glycosynthases.[23] Examples of the mechanisms of reactions catalyzed by these enzymes are presented in Figure 1.4.

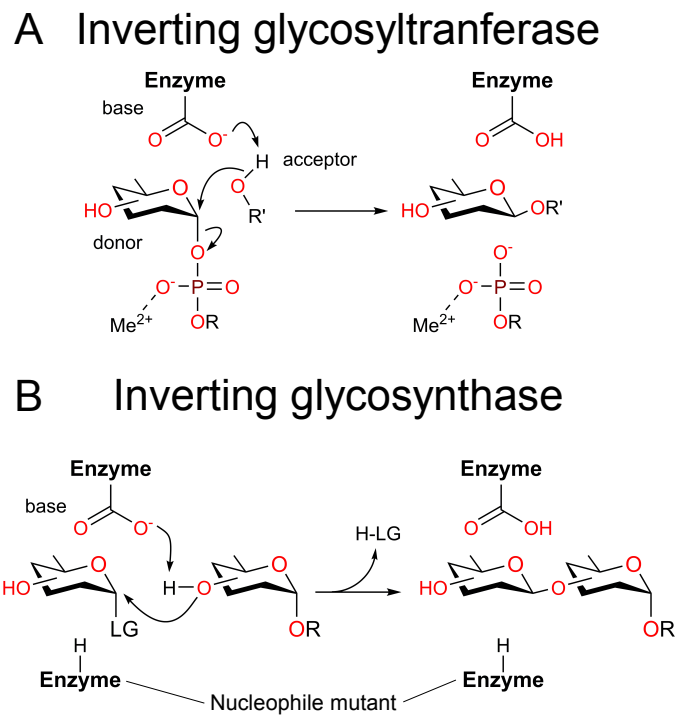


Figure 1.4: Reaction mechanisms of inverting, metal-dependent glycosyltranferases (**A**) and engineered, inverting glycosynthases (**B**). Glossary: LG – leaving group, R – radical. Figures adapted from references 24 and 25.

Conformation of the sugar ring is of interest to enzymologists, because CAZymes often rely on making conformations other than those which are favoured in solution more energetically favourable when the sugar is bound to the enzyme. This lowers the activation energy of certain reactions and makes catalysis possible. The pyranoses of perhaps the greatest scientific interest due to their abundance, D-glucose and D-mannose, both prefer the $^{4}\boldsymbol{C}_{1}$ conformation in aqueous solutions. The alternative chair conformation, $^{1}\boldsymbol{C}_{4}$, puts most bulky substituents in these sugars in the axial positions, making them energetically unfavourable. Certain intermediate pyranose conformations are also possible, at least theoretically: boats ***B***, skew-boats ***S***, half-chairs ***H*** and envelopes ***E*** (Figure 1.5**A**). For a more detailed explanation of the so-called Cremer–Pople sphere and a guide to sugar conformations, see Appendix B. The conformation of a pyranose ring can be described using only three parameters, called puckering coordinates: the angles ϕ (phi) and θ (theta) and the total pucker-

ing amplitude Q.[26] In this system, $^{4}\boldsymbol{C}_{1}$ is the "North pole" (θ=0°), boat conformations lie on the "equator" (θ=90°) and $^{1}\boldsymbol{C}_{4}$ is the "South pole" (θ=180°). Angle ϕ points to the direction of ring puckering. This simple description makes it possible to visualize the conformational landscape on projections similar to that used for the globe: θ is the longitude and ϕ the latitude. The Stoddart projection (Figure 1.5**B**) is usually used to visualize the poles, and the Mercator (Figure 1.5**B**) projection for the whole "globe". Depending on the chemical environment of a sugar, different conformations will have different free energies, with those with lower free energies being more commonly observed: when bound to the enzyme, the free energy of a $^{4}\boldsymbol{H}_{5}$ conformation may not be the same as when in solution. Computationally derived free energy landscape can be plotted on the projection, which constitutes a powerful tool for studying enzyme mechanisms, especially when combined with structural data obtained from crystallography.[27,28]

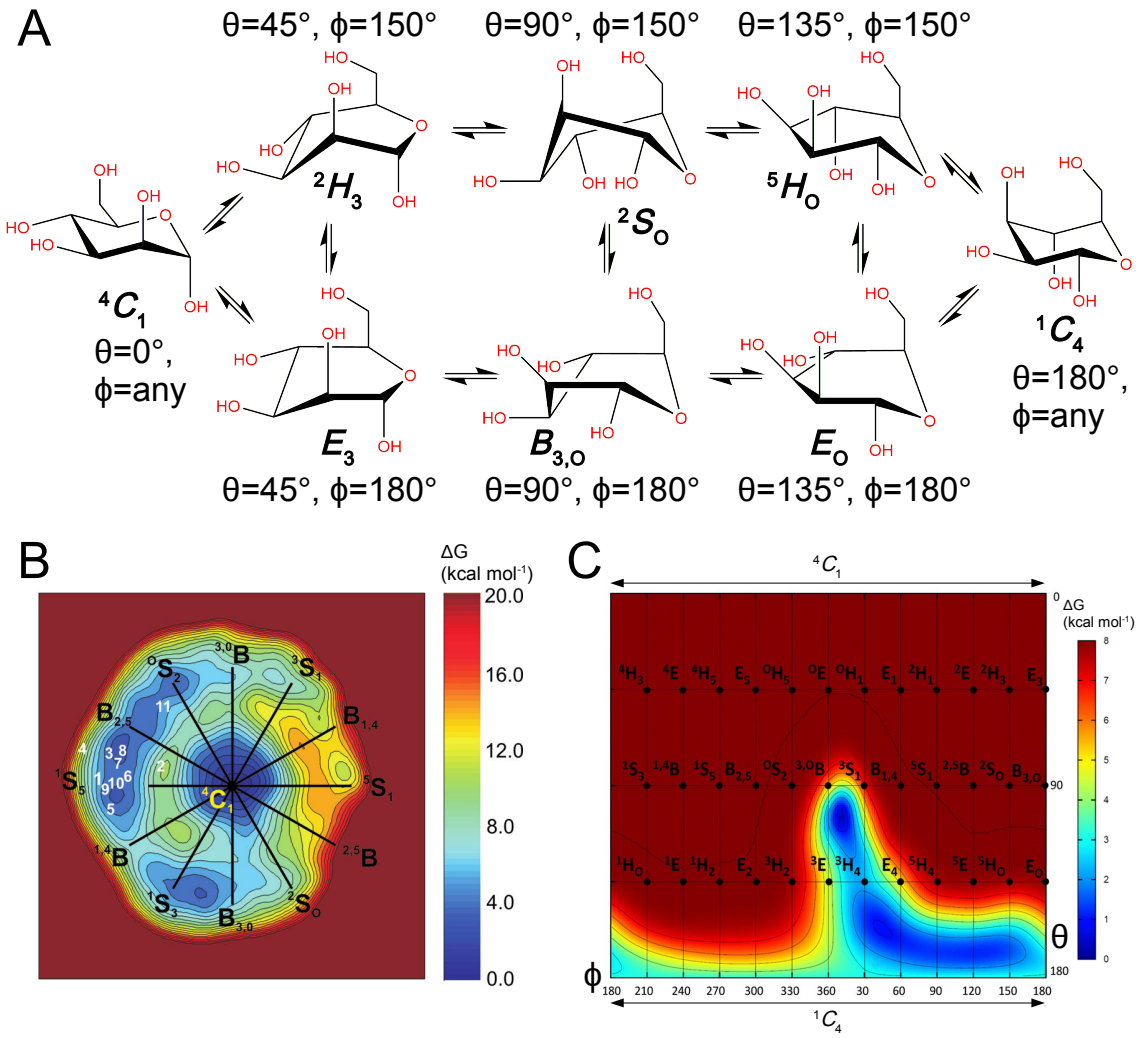


Figure 1.5: Sugar conformations, projections and their free energy. (**A**) Examples of theoretical α-D-mannopyranose conformations and their transitions. Note the inverted equatorial and axial positions in $^4C_1$ and $^1C_4$. (**B**) An example of conformational free energy landscape (FEL) plotted on the Stoddart projection (β-D-mannose) – the northern hemisphere. Figure from Ref. 28. The numbers in white correspond to the conformations of various saccharides at the –1 subsite of various β-mannosidases. (**C**) An example of conformational FEL plotted on the Mercator projection (α-D-mannose bound to GH47 from *Caulobacter* strain K31). Figure adapted from Ref. 29, originally published in 30. ΔG: Gibbs free energy of a conformation.

1.3 CAZy – a database classifying carbohydrate active enzymes

Classification of proteins according to their fold can be useful for establishing deep evolutionary links between them, but it is not very informative with respect to their biochemistry. For this, special interest databases are more practical. One of such databases, probably most relevant to this Thesis is the CAZy (**C**arbohydrate **A**ctive En**Zy**mes) database. The principle behind it was elaborated in a 1989 publication[31]. Initially,[32] 31 families were described and the database was accessible on the Internet under a different name. CAZy was set up in 1999 and described in the *Nucleic Acids Research* journal in 2009.[33,34] The curators categorize proteins into *Families*, *Clans* and sometimes *Subfamilies*. In addition to enzymes, carbohydrate-binding modules (CBMs) with no enzymatic activity have their own category (see Table 1.1). A CAZy family is defined whenever a carbohydrate active enzyme or a CBM is biochemically described, its sequence cannot be categorized into other families and a number of uncategorizable homologues exist. The impetus to set up the database was that the EC classification based on the activity was not specific enough for CAZymes: it is based on the substrate, thus cannot detect divergence or convergent evolution, and is inadequate for enzymes with broad specificity. The number of sequences and activities is increasing rapidly, and tools for sequence analysis coevolve, enabling sophisticated alignment and prediction. At the beginning, hydrophobic cluster analysis was used in addition to sequence alignment.[31] The tools used by CAZy today also include HMMer (which uses hidden Markov models) and BLAST similarity searches.[33] 1-5% of the predicted coding sequences from any organism codes for CAZymes.[34] With such an abundance of sequencing information, the number of sequences to categorize is considerable but the database seems to remain manageable. In this Thesis, both the conventional single-letter and three-letter amino acid name abbreviations will be used, according to the IUPAC-IUB (International Union of Pure and Applied Chemistry-International Union of Biochemistry) Joint Commission on Biochemical Nomenclature guidelines.[35]

Table 1.1: CAZy classification system at the time of writing (excluding deleted families).

Family	Count	Meaning	Description
GH	143	Glycoside Hydrolases	Enzymes hydrolyzing the glycosidic bond.
GT	104	Glycosyltransferases	Enzymes transferring sugar moieties from activated donors to acceptors, forming glycosidic bonds.
PL	26	Polysaccharide Lyases	Enzymes acting on polysaccharide chain using an elimination mechanism and produce an unsaturated product and a new reducing end.
CE	16	Carbohydrate Esterases	Enzymes catalyzing removal of ester sugar decorations (de-*O*-acetylation or de-*N*-acetylation).
AA	14	Auxilliary Activities	Redox enzymes which degrade saccharides.
CBM	81	Carbohydrate-Binding Modules	Polypeptides that bind carbohydrates but do not alter their covalent bonds.

Polypeptides active towards carbohydrates can have broad activity, and single proteins can possess multiple modules that are categorized into different CAZy families. In some cases, subfamilies are defined within a CAZy family, in order to cluster together polypeptides with closer sequence identity. The overall approach showed good predictive value in identifying active residues.[36] This makes CAZy an extremely useful resource for research into these proteins.

1.4 *N*-glycosylation and its implications

N-linked glycosylation is an abundant post-translational modification. The *N* refers both to the nitrogen atom to which sugars are linked, and the amino acid asparagine, which is the site of the modification. Other types of protein-linked glycosylation exist: *O*-linked glycosylation, which is a modification of the serine and threonine side chains, and *C*-linked mannosylation,[37,38] where tryptophan residues are modified. It will be presented below in order to contextualize the research presented in later chapters. *N*-glycosylation is relevant to family GH99 enzymes, and plays a central role in eukaryotes as an intricate mechanism of protein folding quality control.

N-glycosylation itself promotes protein folding,[3,39] is essential in quality control,[40] protein stability, immune responses and other cellular processes.

In eukaryotes, the process of *N*-glycosylation starts in the ER lumen with the transfer of the glycan $Glc_3Man_9GlcNAc_2$ from $Glc_3Man_9GlcNAc_2$-PP-dolichol to an asparagine residue, catalyzed by an enzyme complex oligosaccharyltransferase (OST).[41] The glycan is then trimmed in the ER while the protein folds. When folded, the protein is transported to the Golgi, where the smallest structure observed is $GlcNAcMan_3GlcNAc_2$ and it is later readorned with various saccharides and non-saccharide groups. The pathway is well-studied and has been reviewed many times, but unanswered questions still exist.[13,42–48]

1.4.1 *N*-glycan assembly and transfer

The association of proteins with saccharides has been recognized in late 19th century.[49] Initially an obscure phenomenon, the discovery opened many questions about the structure of proteins: why are the sugars and proteins linked? How are the sugars attached to the protein? What may their function be? Many of these questions, especially about the function of specific glycan structures, are still awaiting an answer, but much progress has been made. Initial studies of the chemistry of glycosylated peptides obtained from digesting ovalbumin led a conclusion that the aspartate side chain must be the site of glycan attachment,[50] and later the Asn-glycan linkage was proposed, although first without direct evidence.[51] The structure of the Asn-linked precursor glycan (shown is Figure 1.6) was elucidated in 1970s and shown to be conserved among eukaryotes from yeast to mammals[52–54] as $Glc_3Man_9GlcNAc_2$. This conservation reflects the functional importance of the pathway.

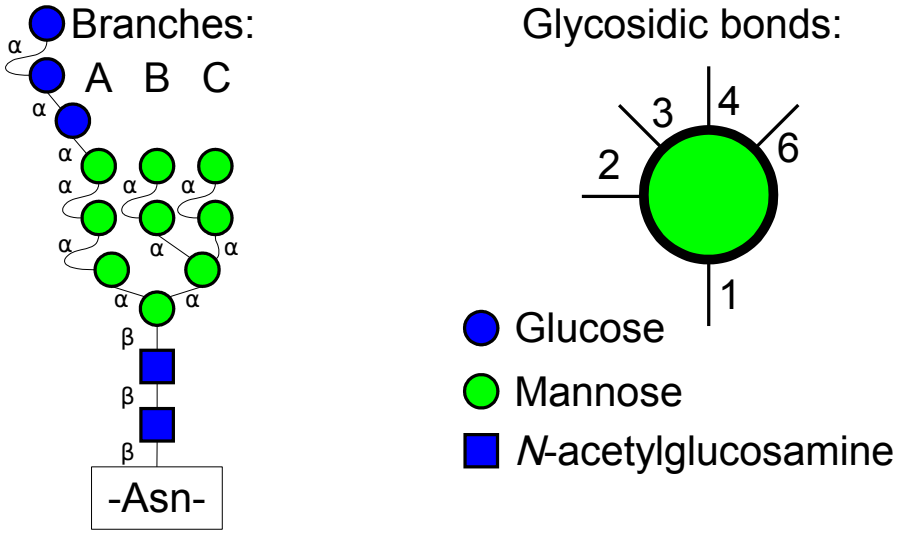


Figure 1.6: Structure of the precursor *N*-glycan bound to the protein, its link nomenclature and branch naming conventions. The lines representing bonds will convey the linkage type as pictured only for sugar moieties present in the precursor glycan.

Synthesis of the glycan is achieved by a tandem of monomeric enzymes and enzyme complexes, both in the cytoplasm and the ER lumen (see Figure 1.7 for a scheme). In the cytosol, UDP-Glc (uridine diphosphate glucose) and GDP-Man (guanosine diphosphate mannose) derived from primary metabolism serve as building blocks for specific transferases. Dolichol kinase DOLK produces dolichyl phosphate (Dol-P) at the ER membrane. The first step of the oligosaccharide assembly is DPAGT1- (dolichyl-phosphate *N*-acetylglucosaminephosphotransferase 1-) catalyzed conversion of Dol-P to α-GlcNAc-PP-Dol. DPAGT1 is a multipass transmembrane protein formerly categorized in family GT4, now uncategorized. Subsequently a GT1 transferase adds a GlcNAc molecule, and a complex of three GT33 and GT4 transferases adds five mannose residues, yielding $Man_5GlcNAc_2$-PP-Dol. The glycan is then flipped from the cytoplasmic side of the ER to the lumenal side. The protein responsible for this is thought to be RFT1 (RFT1 (requiring fifty three 1) homolog), although the data are inconsistent.[55] From then on, the building of the oligosaccharide is achieved by the action of six inverting transferases from families GT58, GT22, GT57 and GT59. They all are multi-pass transmembrane proteins with catalytic machinery on the side of the ER lumen, and their donor substrates are Man-P-Dol and Glc-P-Dol. These donors are synthesized on the cytosolic side of the ER[56] from GDP-Man, UDP-Glc and P-Dol by, respectively, DPM (dolichol-P-mannose synthase) complex whose catalytic subunit is DPM1 (dolichyl-phosphate mannosyltransferase subunit 1, catalytic) and ALG5 (asparagine-linked glycosylation 5 homolog) from

family GT2. The products are then transported to the lumenal side by MPDU1 (mannose-P-dolichol utilization defect 1) flippase belonging to the lysosomal cystine transporter family.[57,58] Mutations in many of the genes involved in the precursor glycan synthesis pathway result in congenital disorders of glycosylation (CDG).[59]

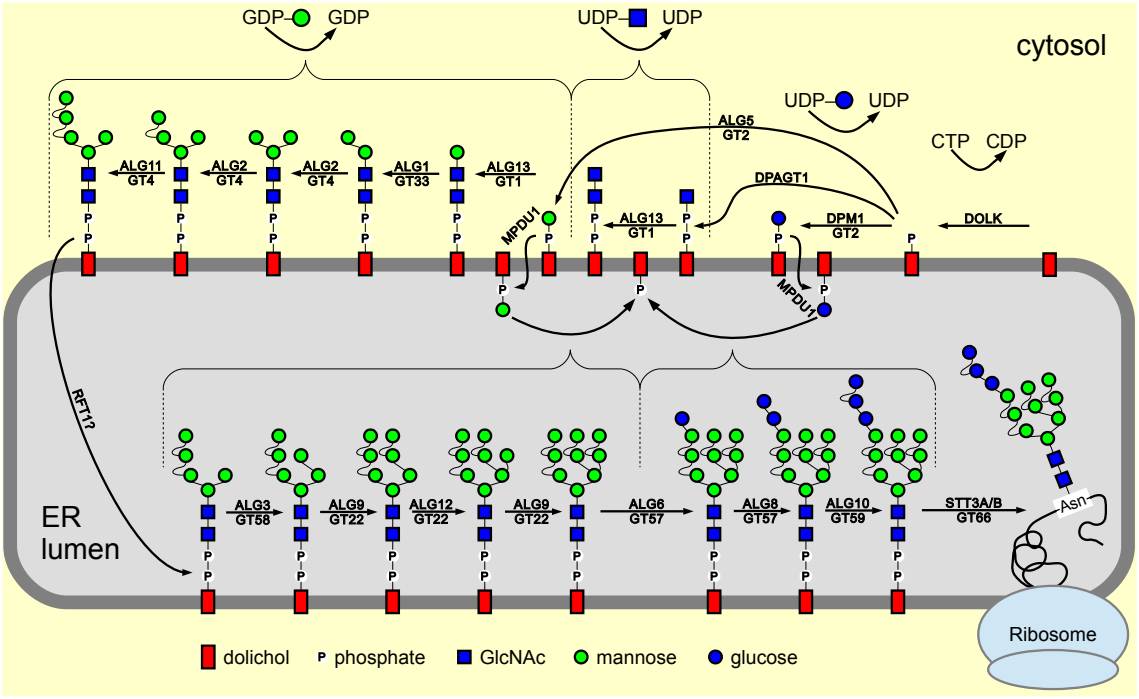


Figure 1.7: Synthesis and transfer of the lipid-linked precursor *N*-glycan. The initial steps of this process occur in the cytosol, and the later steps take place in the ER. The names of enzymes catalyzing each step are above the reaction arrows. If known, the CAZy family of each protein is shown. Linkages as in Figure 1.6.

The transfer of the oligosaccharide to Asn residues is performed by OST. OST in humans and yeast is a protein complex, while in bacteria it is monomeric. Structures of exemplary OSTs have been solved and confirmed initial hypotheses about the reaction mechanism.[60–64] The recognition site of OST in eukaryotes is a sequon N-X-T/S/(C)-X, where X is any amino acid other than proline. Bioinformatics and mapping studies in mice indicate that two thirds of all possible sites are glycosylated.[65,66] The NXTX motifs are the best substrates for OST, followed by NXSX and NXCX. A motif with cysteine can be thought of as non-canonical for the reaction, as it constitutes only 1% of all glycosylated loci. 75% of the glycosylation occurred in predicted loop regions, 15% in β-strands and only 5% in α-helices.[66] Proteins STT3A and STT3B (from: staurosporin and temperature sensitive 3 A/B, current name: catalytic subunit of the oligosaccharyltransferase complex) were recognized as the catalytic subunits in mammalian OST and assigned to family GT66. In *Saccharomyces cerevisiae* only one isoform, Stt3p, exists and also is catalytic.[61] The roles of STT3A an STT3B are

partially non-overlapping: complexes with STT3A are more efficient in glycosylating the N-terminal recognition sites, while STT3B works on C-terminal sites. OST with STT3A works co-translationally close to the growing polypeptide chain, while OST with STT3B glycosylates its substrates after translation. The OST complex in yeast was found to associate with the outlet of the ribosome.[67] These catalytic subunits arose from a gene duplication, and then developed different specificities. The boundary between STT3A and STT3B OST specificity is 50–55 C-terminal residues.[68,69] The α-1,2-linked terminal glucose residue in the precursor glycan is the most important moiety for recognition of an oligosaccharide as a substrate by OST, but it was found that the minimal substrate that can be used is as small as GlcNAc-PP-Dol.[70]

The STT3B complex contains proteins MagT1 (magnesium transporter 1) and Tusc3 (tumor suppressor candidate 3), which are oxidoreductases. The environment of the ER lumen is oxidizing which makes disulfide bond formation energetically favourable. It has been proposed that when the STT3B-contaning OST associates with the protein, MagT1 and Tusc3 may form transient disulfide bridges with the protein or reduce those already formed, facilitating a situation in which the polypeptide can fold first and form correct disulfide bridges later.[71,72] When the protein is folded, the recognition sequons are inaccessible to OST, as shown early on using ovalbumin as a model.[73]

1.4.2 *N*-glycan trimming in the ER and quality control vesicles

Glycoside hydrolases can be broadly classified as *exo* and *endo*-acting, with some proteins presenting semiprocessive* activity.[74] *Exo*glycosidases release carbohydrates one by one acting at the terminal end of an oligo/polysaccharide, and *endo*glycosidases cleave non-terminal glycosidic bonds. In the ER, exomannosidases, exoglucosidases and a glucosyltransferase is responsible for glycan processing. The ER *N*-glycan processing pathway is instrumental for assuring correct protein folding in higher eukaryotes. It is interesting to note that the optimal pH for ER glycosidases is neutral. The first enzyme to process the *N*-glycan is ER glucosidase I, an inverting α-1,2-

*Semiprocessive activity refers to enzymes that attack an internal bond and proceed by hydrolyzing a number of consecutive moieties before initiating an attack at another internal position.

glycosidase from family GH63. It is a type II[†] transmembrane protein[75] encoded in humans by the gene *MOGS* (mannosyl-oligosaccharide glucosidase). The structure of yeast ER glucosidase I, which is kinetically similar to the mammalian enzyme,[76,77] has been solved.[78] The minimal substrate for the enzyme is a glucotriose with linkages which are configured as in the native substrate, $Glc_3Man_9GlcNAc_2$-Asn(…) (see Figure 1.6). Due to crystal packing issues, it was impossible to investigate ligand binding experimentally, but docking efforts supported the view that Asp568 and Glu771 are the catalytic acid/base residues. The glucotriose molecule recognized by ER glucosidase I has a unique shape: the +1 Glc has glycosidic bonds at adjacent groups, OH1 and OH2. This type of linked glycan structure occurs nowhere else in biology. This is probably relevant to the function of *N*-glycosylation, which depends on enzymes in the pathway catalyzing only specific reactions. The structure of GH63 from *Thermus thermophilus* and enzymologic NMR studies were used to elucidate its reaction mechanism (shown in Figure 1.8**A**).[79,80] Cleavage by ER glucosidase I is a rapid reaction and the enzyme does not recognize the folding state of the protein.[81]

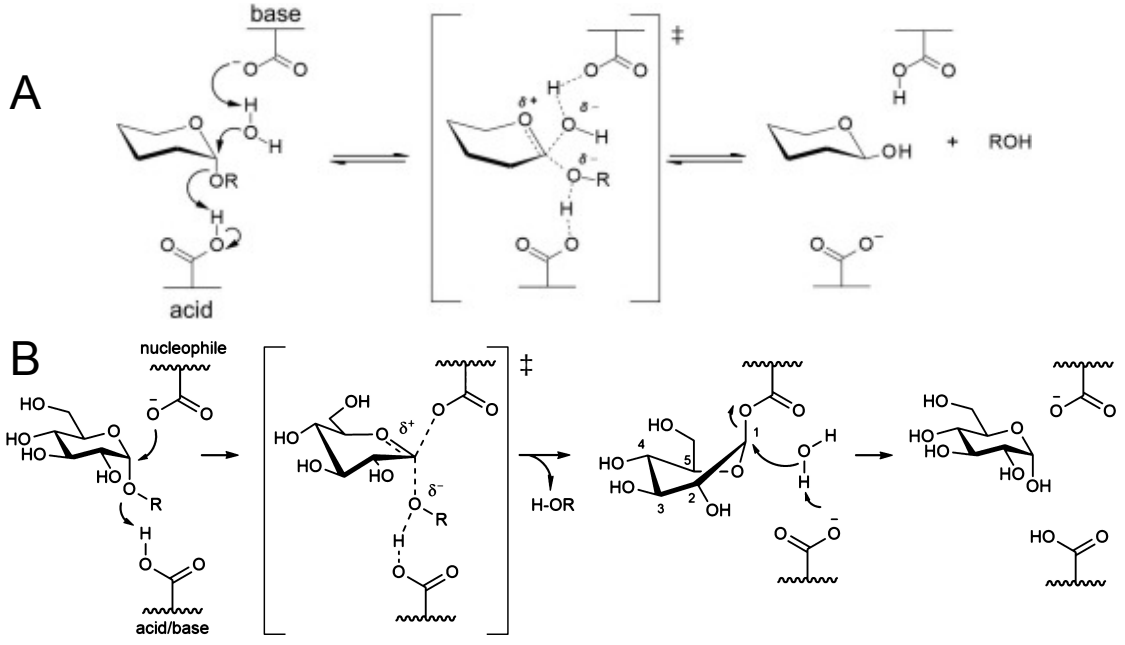


Figure 1.8: Two common reaction mechanisms of glycosidases. (**A**) The single displacement, inverting mechanism of GH63 enzymes which act on the α-1,2-glucosidic. Figure from Ref. 80. (**B**) The double displacement mechanism of GH31-catalyzed hydrolysis of α-1,3-glucosidic bond. The anomeric configuration is retained. Figure adapted from Ref. 82.

[†]Type II membrane proteins are single-pass transmembrane proteins with their C-terminus facing the ER lumen.

After the first glucose moiety is trimmed, the glycan becomes a substrate for ER glucosidase II and a ligand for malectin, a chaperone protein (see Section 1.4.3 for discussion of the latter and Figure 1.9 for an overview of glycan trimming and recognition in the ER). ER glucosidase II is a GH31 retaining α-1,3-glucosidase that is able to cleave both α-1,3-linked glucoses from the *N*-glycan. In mammals and yeast it is a heterodimer composed of a catalytic subunit encoded by *GANAB* (glucosidase II alpha subunit) gene and a regulatory subunit encoded by *PRKCSH* (protein kinase C substrate 80K-H).[83] Mechanistically, GH31 proteins follow a classical Koshland[84] double displacement reaction coordinate with one residue acting as an acid/base and one as a nucleophile, as presented in Figure 1.8**B**.[82,85] The enzyme is key to the first level of ER quality control (ERQC). ER glucosidase II recognizes its substrates by interacting with the C branch (see Figure 1.6 on page 13 for reference) of the glycan, as removing the C branch-linked mannose decreases the rate of the deglucosylation reactions *in vitro*.[86] Cleavage of the first glucose moiety results in oligosaccharide $Glc_1Man_9GlcNAc_2$ (G1M9Gn2), which is recognized by calnexin and calreticulin, molecular chaperone lectins. The lectins can bind cotranslationally when the protein chain is as short as 50 aa[87] (amino acids) and two *N*-glycans are necessary for the lectins to associate with the glycoprotein.[81] The second cleavage of glucose yields glycoproteins with $Man_9GlcNAc_2$ (M9), a structure for which the lectins have low affinity and dissociate. However, another key enzyme, UDP-glucose:glycoprotein glucosyltransferase (UGGT), which has a folding sensing domain[88] can recognize $Man_9GlcNAc_2$ and reglucosylate it, letting the chaperones bind again. UGGT belongs to the GT24 CAZy family and there are two genes with associated protein products of similar length in the human genome (*UGGT1* and *UGGT2*). Structural analysis of full-length eukaryotic UGGT[89] revealed considerable conformational flexibility. The fold of the protein is complex: its seven domains act like a clamp around the unfolded protein and the GT24 domain is located inside the clamp, facing the folding sensing machinery. Expression of *UGGT* is higher when cells undergo ER stress and the protein plays an important role in response to an increase in unfolded proteins.[90] UGGT1 was found to be essential for embryonic development: *UGGT1*$^{-/-}$ mouse mutant embryos do not develop.[91] This deglucosylation-reglucosylation cycle continues until the protein is folded and no longer a UGGT substrate, or misfolded, which is sensed by the EDEM (ER degradation enhancing alpha-mannosidase like) proteins.

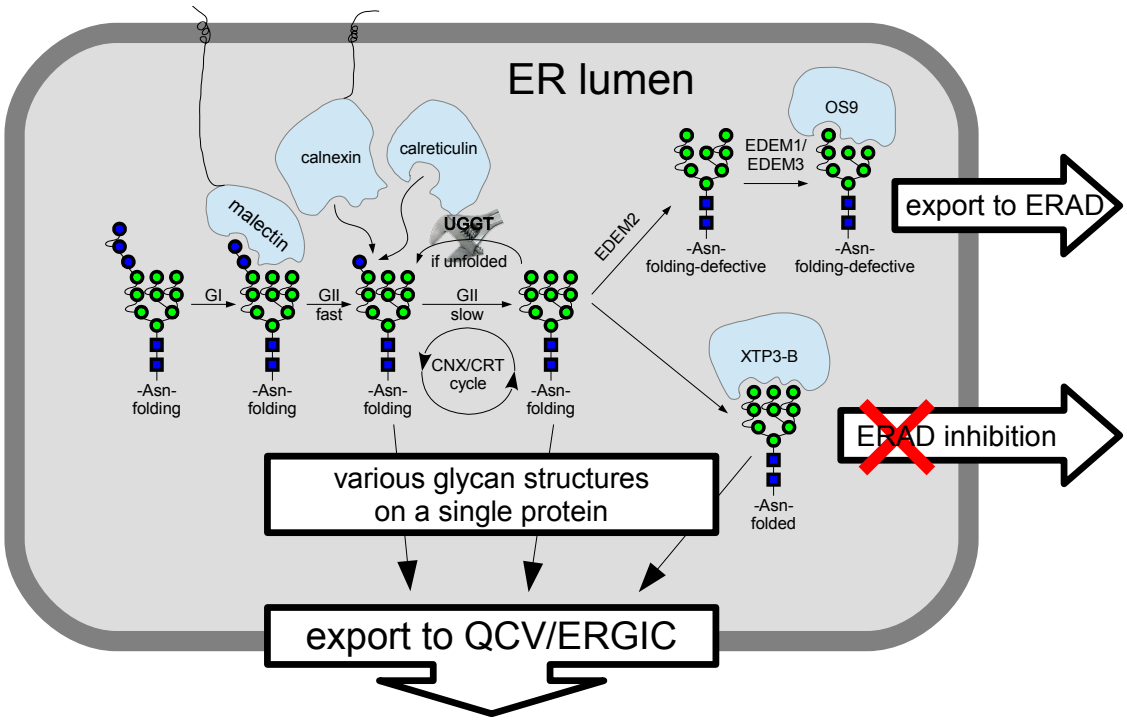


Figure 1.9: *N*-glycan recognition and processing in the ER. Folded proteins are directed to the QCVs/ERGIC, and a large part of folding-defective proteins is degraded in ERAD. Glossary: GI – ER glucosidase I, GII – ER glucosidase II.

At this point the glycan has all mannose termini exposed and mannosidases step onto the stage. The yeast and mammalian pathways are substantially different and below the human pathway will be described, however there is still controversy about particular steps. There are several genes in the human genome known to encode *N*-glycosylation-associated α-1,2-mannosidases: *EDEM1*, *EDEM2*, *EDEM3*, *MAN1A1* (mannosidase alpha class 1A member 1), *MAN1A2* (mannosidase alpha class 1A member 2), *MAN1B1* (etc.) and *MAN1C1*. They all belong to the GH47 family and their catalysis results in overall inversion of stereochemistry at the anomeric carbon,[92] but their specificities are different. ER mannosidase I (*MAN1B1* product) contains a short N-terminal cytoplasmic tail and a transmembrane domain while EDEMs are resident in the ER lumen. In addition to a GH47 domain, EDEM3 also contains an N-terminal protease-associated domain.[93] The conventional names of the *MAN1A1*, *MAN1A2* and *MAN1C1* gene products, accordingly, Golgi mannosidase IA, IB and IC and all are type II transmembrane proteins. A landmark study[93] revealed that it is EDEM2 which is responsible for most of the conversion of M9Gn2 to $Man_8GlcNAc_2$ with the α-1,2-mannose removed from the B branch of the precursor glycan (M8B). ER mannosidase I, contrary to its accepted name, is resident in the quality control vesicles (QCVs).[94] Due to a particular experimental and imaging procedure, previously it was

thought the enzyme is localized to the ERQC because QCVs congregate in ERQC-like clusters when cells are under stress. Golgi mannosidase IA was also found to localize to QCVs and not the Golgi,[95] after many years of controversy and conflicting data.[96] To avoid confusion, these will be referred to as QCVManI (ER mannosidase I) and QCV-ManIA (Golgi mannosidase IA). Substrate specificity studies revealed that QCVManI, like EDEM2, prefers to cleave the mannose on the B branch of the M9Gn2 glycan, but the preference is not strict.[97,98] QCVManIA prefers the A branch of M9Gn2 while mannosidase IB prefers the C branch.[99] This is substantiated by a later comparison of mammalian structures of these two enzymes.[100] Mannosidase IC is the only enzyme that can remove the last α-1,2-linked mannose from M8AGn2 producing M7AAGn2, and then proceed to cleave the C branch making M6AACGn2.[101] Hypotheses about their localization are discussed in Section 1.4.4. The picture is complicated by EDEM1 and EDEM3 proteins, which are specific towards the A and C arms of M8BGn2 and act downstream of EDEM2. EDEM3 activity is more pronounced than EDEM1 activity and mammalian cells with *EDEM3* knocked out (ko) accumulate 10–20% more M8BGn2 (~5% more for *EDEM1* ko).[93] A 2008 study[102] reported that most EDEM1 is localized to EDEMosomes coated with LC3-I‡, and high turnover of this protein ($t_{1/2}$=1 h) was observed. No structural data on EDEM proteins exists, but structures of QCVManI and QCVManIA were solved. Curiously, in addition to M9Gn2, cells with QCVManI knocked down (kd) also accumulated proteins with G1M9Gn2 glycans in a pulse-chase experiment.[105] This happens in cells treated with kifunensine (KIF), a selective[29] GH47 inhibitor, as well. It can be explained by accumulation of M9Gn2 and its subsequent reglucosylation to G1M9Gn2 by UGGT. Cultured EDEM2-ko cells also show an increased amount of G1M9Gn2.[93] It is important to note that QCVManI can process glucosylated *N*-glycans at the same rate as non-glucosylated glycans.[106] It preferentially removes mannose moieties from unfolded proteins and very slowly from native proteins.[106,107] Coupled with its localization in mammalian cells outside the Golgi, it seems to play a role of a post-translational quality check enzyme. An overview of the current understanding of glycan processing in QCVs and the ERGIC is presented in Figure 1.10.

‡LC3 is a canonical name for the protein products of the gene MAP1LC3A and MAP1LC3B (microtubule associated protein 1 light chain 3 alpha/beta). LC3-I is nonlipidated and LC3-II is lipidated.[103,104] LC3-I is cytosolic and also associated with EDEMosomes, while LC3-II is associated with authophagosomes.

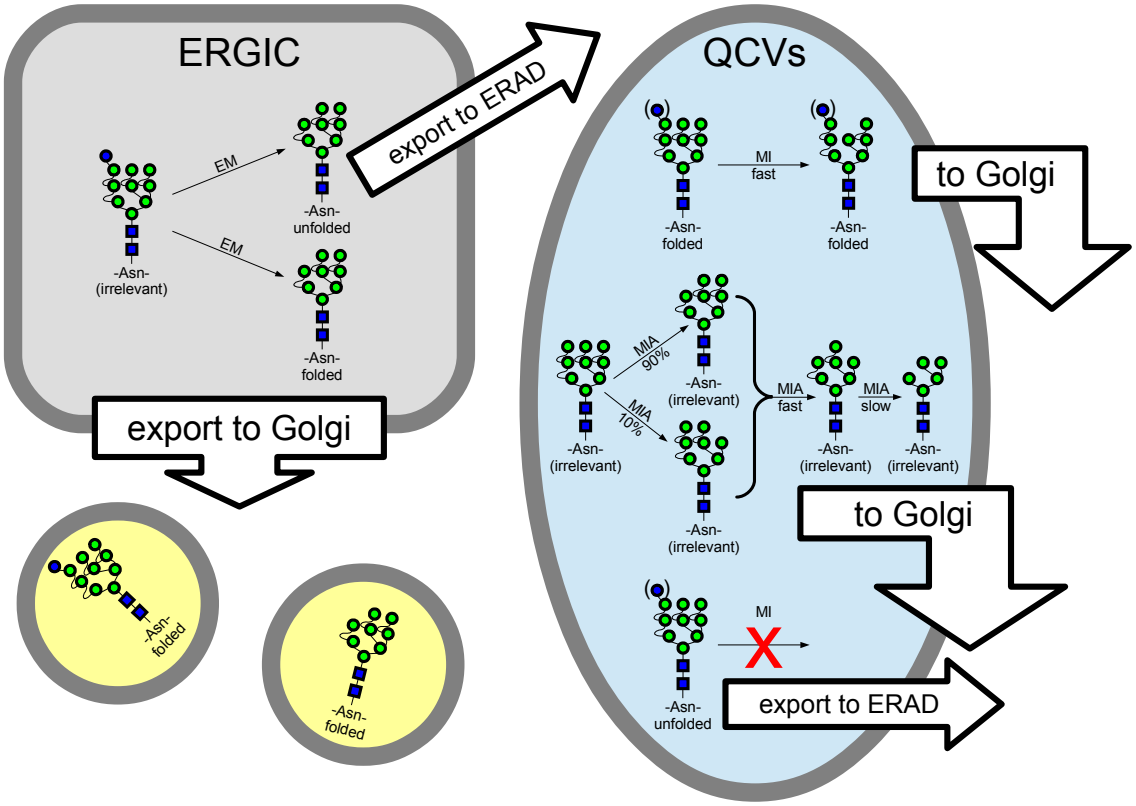


Figure 1.10: *N*-glycan processing in the ERGIC and the QCVs. Glossary: EM – *endo*-α-1,2-mannosidase, MI – QCV mannosidase I, MIA – QCV mannosidase IA. Residues in parentheses make no difference for enzyme specificity.

In summary, the only initial mannosyl residue cleavage reaction in the ER can occur at the B branch of the *N*-glycan (catalyzed by EDEM2), and this makes the glycan vulnerable to further cleavage of the C branch by EDEM3 and EDEM1. This process often leads to the commencement of ERAD (ER-associated degradation). It is possible that in rapidly and correctly folding proteins, no mannose residues are cleaved, and an α-1,3-linked glucose can be still present on the glycan A branch. The most likely situation is that GH47 mannosidases resident in the QCVs cleave off the mannoses from the B and C branch of (G)M7-9, leaving proteins with (G)M5-7 glycans as they enter the Golgi. Glycans are then processed further by unique Golgi mannosidases, described in Section 1.4.4. Golgi processing involves subsequent readorning of the *N*-glycan with other oligosaccharides, performed by various glycosyltransferases.

1.4.3 ER quality control and ER-associated degradation

The quality control mechanism in the ER depends on lectins that recognize different high-mannose glycans. The first one in the pathway is malectin, which recognizes the G1M9Gn2 glycan and is upregulated under conditions of cellular stress.[108,109] It is a type I transmembrane protein and acts upstream of calnexin, calreticulin and BiP (binding immunoglobulin protein) as it does not compete with them for substrates. Overexpression of the gene encoding for malectin causes a reduction in folding-incompetent protein secretion and less deglucosylation by ER glucosidase II.[109] Two lectins have been identified as major contributors to the regulation of this pathway. Lectin OS9 (OS9, endoplasmic reticulum lectin, abbreviation from: osteosarcoma amplified 9), the structure of whose sugar-binding domain is solved (PDB (Protein Data Bank) code: **3AIH**),[110] has been shown to promote ERAD. OS9 binds with a high affinity to M7AGn2 structures (where two mannose molecules from the B and C branches were removed, see Figure 1.9 on page 18). Crucially, the OS9 must bind after the EDEM3/1 cleavage, which produces this glycan. Two isoforms of OS9 are produced in humans owing to alternative splicing and both are often overexpressed in tumours. OS9 isoform 2 with an inactivated carbohydrate-binding module was still competent at binding to a folding-defective protein, but when the glycosylation site was removed, the interaction did not target it to ERAD.[111] Therefore, the exact function of this domain remains unclear. The other of the two is XTP3-B (XTP3-transactivated gene B protein;), which has an opposite regulatory activity to OS9. XTP3-B is a lectin which specifically binds to M9Gn2 oligosaccharides and makes their degradation through ERAD less likely.[112] As such, its binding activity would compete with that of UGGT and EDEM2. However, conflicting reports exist that assign the role of XTP3-B as essentially redundant with that of OS9.[113]

The function of EDEM proteins is important in ERAD. In *Drosophila melanogaster*, dEDEM1 is a homolog of human EDEM2, and dEDEM2 is homologous to EDEM1 and EDEM3. Under the model described in Ref. 93, dEDEM1 would function upstream of dEDEM2 in the fly. A recent study found that overexpression of either of the *EDEM* genes protects against age-related proteinopathy and the mannosidase activity is not essential for this effect.[114] However, the same study found that the mannosidase activity is essential to degrade the culprit protein, $\alpha\beta42$, in neurons. This suggests that

EDEMs may possess chaperone-like qualities independent of their mannosidase activity. Earlier[115] it was proposed that EDEM proteins release folding-incompetent proteins from the calnexin/calreticulin cycle, although only EDEM1 was known at that time and its role is less important than that of EDEM2 or EDEM3, as described above.

HSPA5 (heat shock protein family A member 5) encoding BiP was found as one of the genes upregulated in conditions of stress. It belongs to the HSP70 family[116] and is constitutively expressed at a moderately high level. It has multiple roles in the ER, and one of them is to keep polypeptides in a folding-competent state directly after translation. As a member of the HSP70 (70-kDa (kilodalton) heat shock protein) family it has an ATPase domain and a substrate binding domain. It may have a foldase activity as upon binding the substrate, the rate of ATP hydrolysis increases. BiP is thought to be stored in an inactive oligomeric form that is post-translationally modified and released in a monomeric form when needed, providing a reservoir for rapid response to conditions of stress.[117,118] It is possible that in QCVs, where QCVManI and QCVManIA reside, there are lectins that recognize the oligosaccharides produced by these enzymes. Such colocalization would provide a basis for another quality control mechanism. Overexpression of *MAN1A1* has been shown to enhance ERAD, whereas its knockdown suppresses the degradation.[95]

The overall logic of the ERAD pathway, to which OS9-tagged proteins are directed, is well understood. Terminally misfolded proteins are transported to the cytosol and polyubiquitinylated. There is controversy about the exact mechanism, but in certainty the ubiquitinylation machinery is located in the cytosol. N-linked glycans are cleaved by PNGase (peptide:N-glycosidase) in the cytosol and then degraded by lysosomal glycosidases. Polyubiquitinylated proteins are then degraded inside the proteasome complex into 7-9 aa long peptides, and ubiquitin molecules are recycled for further use.[119–121] Free oligosaccharides (FOS) are generated during ERAD: overexpression of *EDEM1* in CHO-K1 cells increases the total amount of FOS resident in the cytosol but not in the ER lumen.[122] Due to the presence of PNGase, fully functional proteins resident in the cytosol are in general deglucosylated (although at various rates).[123,124]

Folded glycoproteins are transported from the ER to the Golgi through QCVs or ER-Golgi intermediate compartment (ERGIC). The cargo from the ER to ERGIC is

packaged into vesicles coated by COPII (coat protein II).[125] A marker for ERGIC is a D-mannose-specific lectin ERGIC-53,[126,127] whose another characteristic is a long half-life. The identity of ERGIC-53 cargo is not clear, but it probably interacts with the mannose "epitopes" on the high-mannose *N*-glycans. There is dispute about the role of the ERGIC, but the prevailing understanding is that it plays a sorting role.[128] The glycoproteins are then packed in vesicles again and transported to the Golgi.

1.4.4 *N*-glycan processing in the Golgi

In the Golgi, glycoproteins encounter at least two enzymes with mannosidase activity: the *endo*-α-1,2-mannosidase (ERGIC/*cis*-Golgi)[129] and Golgi mannosidase II (*medial*/*trans*-Golgi).[96] The endomannosidase from family GH99 is the main focus of this work and will reviewed separately in Section 1.4.5. It catalyzes the removal of $Glc_{3\text{-}1}Man$ from the A branch of $Glc_{3\text{-}1}Man_{9\text{-}7}GlcNAc_2$ glycans, leaving $Man_{8\text{-}6}GlcNAc_2$ glycosylation.[130] Golgi mannosidase II cleaves α-1,3- and α-1,6-linked mannoses, converting GnM5Gn2 ($GlcNAcMan_5GlcNAc_2$) to GnM3Gn2 glycans.

The report about QCVManIA not being localized to the Golgi[95] casts doubt on the subcellular localization of related GH47 enzymes: "Golgi" mannosidases IB and IC. Below, the possibility of both Golgi and QCV localization of them will be discussed. In living cells, Golgi mannosidase II acts only after GlcNAc transferase I (MGAT1, mannosyl (alpha-1,3-)-glycoprotein beta-1,2-*N*-acetylglucosaminyltransferase) adds a GlcNAc residue to the A branch of the glycan, which first must be free of α-1,2-linked mannoses. For an overview of processing of *N*-glycan in the Golgi, refer to Figure 1.11.

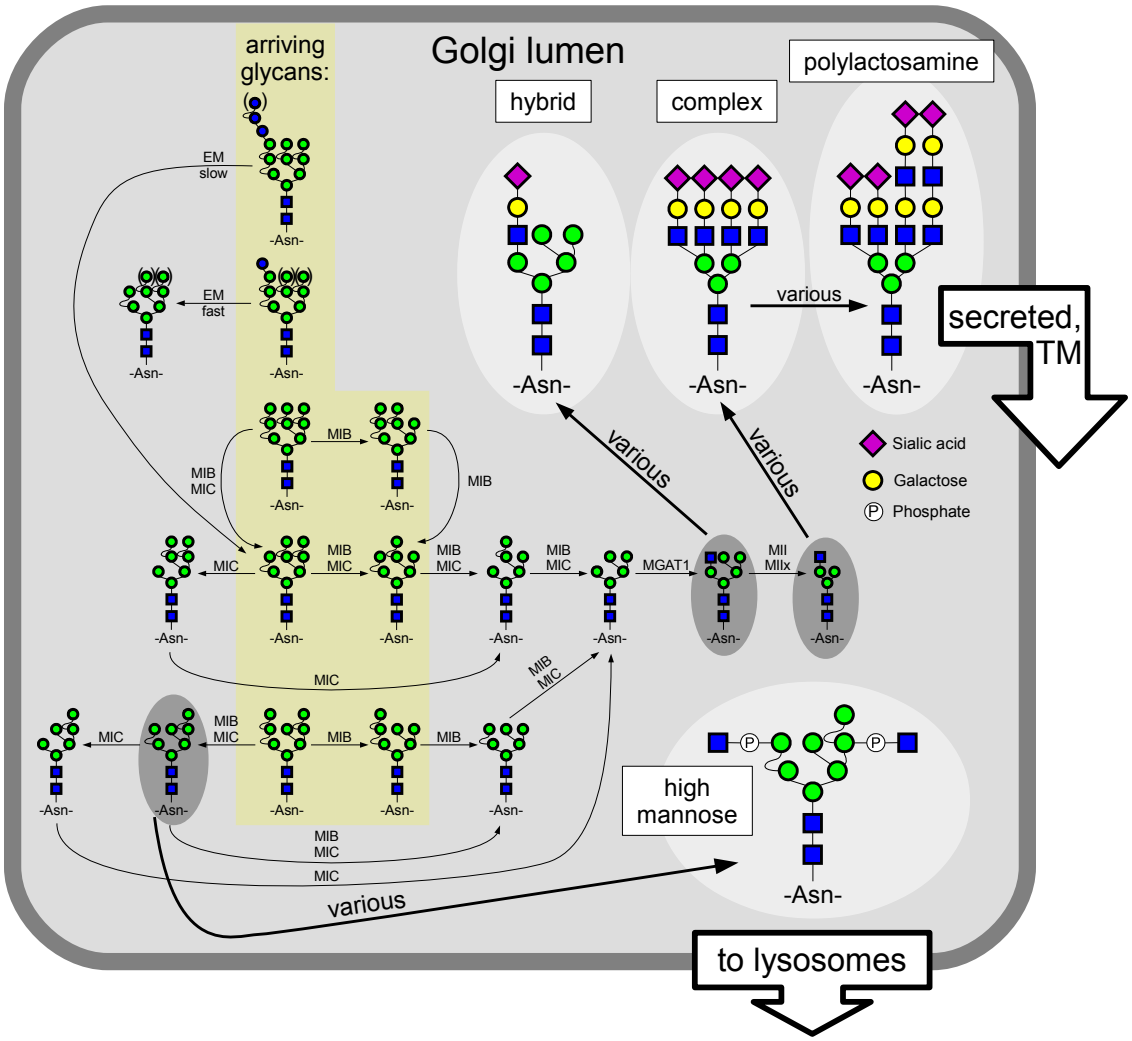


Figure 1.11: *N*-glycan processing in the Golgi apparatus. Glossary: EM – *endo*-α-1,2-mannosidase, MIB – Golgi mannosidase IB, MIC – Golgi mannosidase IC, MII – Golgi mannosidase II, MIIx – Golgi mannosidase IIx. Structures most likely to arrive at the Golgi shaded yellow. Glycans used as unique starting material for modified structures shaded grey. Mature glycans shaded white. Residues in parentheses make no difference for enzyme specificity.

Under the assumption that the glycan is still glucosylated and there are only two QCV mannosidases, they would remove moieties from the B and C branch and mostly GM7AGn2 glycans enter the Golgi. After the cleavage by the endomannosidase the glycan would still possess one α-1,2-linked mannose (on the A branch), which implies that at least one α-1,2-mannosidase resides in the Golgi. The best candidate would be Golgi mannosidase IC, which is capable of catalyzing the removal of the α-1,2-linked mannose from the A branch irrespective of the structure of other branches.[101] Golgi mannosidase IB would serve as a backup Golgi mannosidase, trimming the B and C branches. If Golgi mannosidases IB and IC were also localized to the QCVs, the endomannosidase would need to be in these vesicles as well because there would be no backup mechanism for cleaving the A branch in the Golgi. The data on endomannosidase localization is, however, entirely consistent with Golgi/ERGIC (see Section 1.4.5.3 on page 31).

Golgi mannosidase II belongs to the CAZy family GH38, was first isolated from rat liver Golgi and is a type II membrane protein.[131,132] A crystal structure of *Drosophila melanogaster* Golgi mannosidase II in complex with its substrate has been solved.[133] The model suggests that the enzyme cleaves two mannose residues sequentially without dissociating from the oligosaccharide, and that considerable movement in the active site is needed to achieve this: only one subsite is catalytic. The first mannose to be removed is the α-1,3-linked moiety (branch B), followed by the α-1,6-linked Man (branch C). In humans and other mammals, there also exists an isozyme, Golgi mannosidase IIx,[134] whose role is largely complementary but it is essential for some processes, such as spermatogenesis.[135] Golgi mannosidase II/IIx double null mutant mice completely lack complex-type *N*-glycans and fail to thrive.[136] Single null mutant mice are viable, demonstrating that the presence of one of these enzymes can compensate for the lack of the other.

Subsequent processing of *N*-glycans occurs in *cis*, *medial*, *trans*-Golgi and *trans*-Golgi network (TGN) and involves action of a vast array of enzymes. It is outside the scope of this introduction to characterize them all. Much non-homogeneity exists in mature glycoproteins, even within a single glycosylation site, although some sites are essential for the function of a particular protein.[137] Three main classes of *N*-glycans observed on mature proteins can be distinguished: high-mannose (in which the A, B and C mannose branches of the precursor glycan are found), hybrid (only branch

A removed and subsequently modified) and complex (all three branches removed and modified, see Figure 1.11).[138] One common modification is branching (up to 6 times) via the action of various GlcNAc transferases. Their specificities are tuned so that they act in tandem, i.e. one branch must be created before another Golgi transferase recognizes the glycan as its substrate and creates another branch. The activities of some branching enzymes prevent others from recognizing the glycan as a substrate.[139,140] Other common saccharide-based modifications include fucosylation, sialylation and galactosylation.[45] Glycans can also be phosphorylated (for example, phosphorylated high-mannose structures are tags for delivery of enzymes to lysosomes, see Figure 1.11),[141] and sulfated. When the processing completes, proteins are packaged into secretory vesicles and exported.[142]

1.4.5 *Endo*-α-1,2-mannosidase: a review

1.4.5.1 The enzymatic activity of *endo*-α-1,2-mannosidase

The *endo*-α-1,2-mannosidase, or endomannosidase for short, is the only *endo*-acting mannosidase in the *N*-glycan trimming pathway. It was first isolated from rat Golgi membranes by Lubas and Spiro.[130] Before its discovery it was postulated that more complex processing of mannoses may take place in the Golgi.[143] The first report established that the endomannosidase is active against GM9Gn2 glycans, has a neutral (7.0–7.2) pH optimum and its activity is the highest in the Golgi fractions. It also provided the explanation for an incomplete blockage of synthesis of complex-type *N*-glycans in conditions of ER glucosidase II inhibition.[144,145] The enzyme was tested in presence of EDTA and had full activity in presence, leading to a conclusion that it does not depend on any metal ions as cofactors. The enzyme also retained full activity when tested with and 1-deoxynojirimycin (DNJ, a glucosidase II inhibitor) in solution. The novel activity was 69 times higher in the Golgi extract than in the ER. The second report about the enzyme by the same authors[146] has shown no activity against M9Gn2 glycans, high activity against monoglucosylated glycans and low (~7-8%) activity with di- and triglucosylated substrates. Interestingly, the fewer mannose molecules the glycans possessed, the higher the endomannosidase activity (2.3 times higher for $Glc_1Man_4GlcNAc_2$ than

for $Glc_1Man_9GlcNAc_2$). Despite the lower activity of this enzyme with di- and triglucosylated free oligosaccharides, the predominant products from its processing of lipid-linked precursor glycan were Glc_3Man and Glc_2Man. A suggestion was made that the additional glucose moieties may sterically inhibit the enzyme. Buffers HEPES (2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid), HEPPS (3-[4-(2-Hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid) and Tris (2-amino-2-(hydroxymethyl)propane-1,3-diol) were inhibitory towards the enzyme, but HEPES and HEPPS were uniquely inhibiting the endomannosidase while Tris inhibited *exo* glucosidase and *exo*mannosidase activity as well. Other exoglucosidase inhibitors, such as castanospermine (CST), bromoconduritol (BCD), 1-deoxymannojirimycin (DMJ) or swainsonine (SW) had no effect on the endomannosidase (for structures of some of these compounds, see Appendix A). The minimal substrate for endomannosidase is a tetrasaccharide mimicking the linkages on the A-branch of the *N*-glycan: α-Glc-1,3-α-Man-1,2-α-Man-1,2-Man[147] (see Figure 1.12). Bacterial GH99, a model for endomannosidase studies, was shown to retain the anomeric configuration of its substrates. The reaction mechanism was proposed to be a unique neighbouring group participation mechanism with a 1,2-anhydrosugar intermediate (see Figure 2.1 on page 45 for details).[148] Evidence of OH-catalyzed hydrolysis of 4-*para*-nitrophenyl α-D-mannopyranoside through such an intermediate has been made available recently, but no enzymology-related proof has been published.[149] The activity-based EC classification of *endo*-α-1,2-mannosidase is 3.2.1.130.

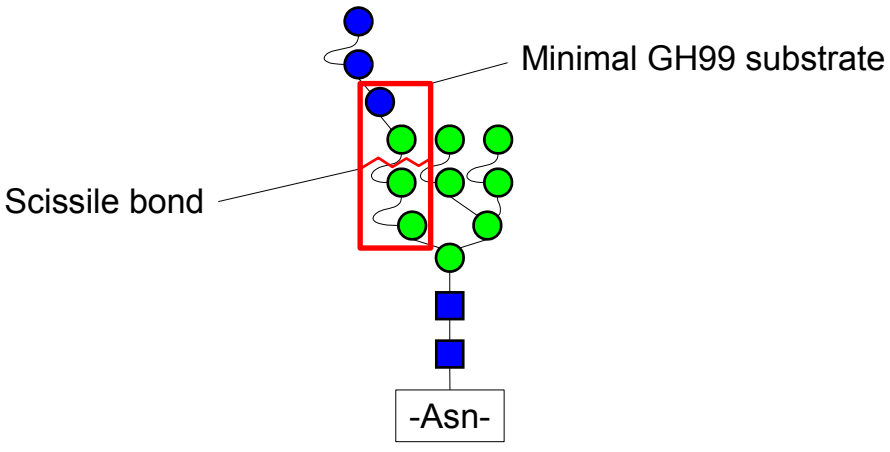
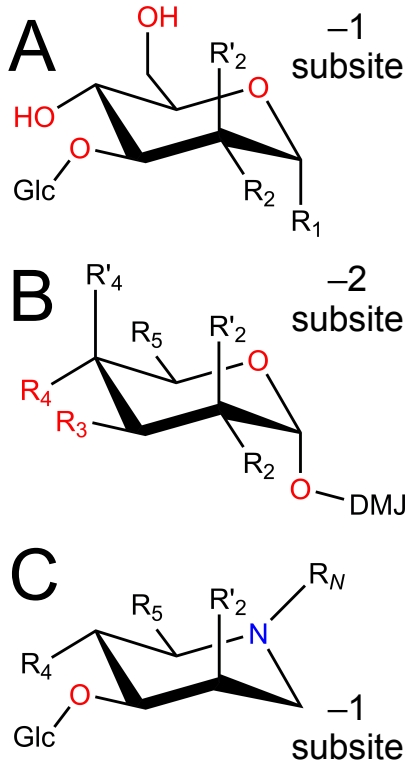


Figure 1.12: Substrate specificity of the GH99 endomannosidase.

Inhibition of mannosidases through the use of DMJ and KIF in a cell line lacking ER glucosidase II results in production of glycoproteins with a specific high-mannose oligosaccharide M8AGn2 attached. This result is exactly what one would expect if the endomannosidase was the only enzyme at work downstream of ER glucosidase I. Without this inhibition a product characteristic of the endomannosidase, G2M, was released into the media. When CST, an ER glucosidase I inhibitor, was added to the cocktail, the sugar released was G3M, again exactly as expected.[150] These findings helped define the specificity of the endomannosidase activity. Usage of the enzyme as a tool for assessing glycosylation state of proteins was proposed[151] and applied.[152]

1.4.5.2 Inhibition of GH99 through use of small molecules

The first inhibitors of GH99 were reported by Hiraizumi and Spiro,[153] and shortly after that more inhibitors were synthesized by Ardron *et al.*[154] The backbone of endomannosidase inhibitors is an α-1,3-linked disaccharide which mimics the natural substrate bound in the –2/–1 subsites. Various –1 "warheads" were surveyed and DMJ without modifications was the most potent (see Figure 1.13**A** and **C**). The –2 sugars tried included Glc, 2-, 3-, 4- and 6- (methyl- or deoxy-) glucose, Gal, Man and Xyl (see Figure 1.13**B**). Of interest, the –2/–1 endomannosidase cleavage product, α-Glc-1,3-Man, was 130 times less potent than GlcDMJ, but two inhibitors were approaching GlcDMJ affinity: Glc-D-glucal (1.3 × less potent) and Glc-ddMan (Glc-dideoxymannose, 2.2 × less potent). Modifications and different substituents around the DMJ ring were tried, but none toppled GlcDMJ affinity.



R_1	R_2	R'_2	–1 compound	IC_{50} (μM)
OH	OH	H	Glc	290
	H	OH	Man	220
H			dMan	150
		H	ddMan	3.8
H/π bonds			glucal	2.3
F	H	OH	Man-F	48
p NP			*p* NP-Man	460

R_2	R'_2	R_3	R_4	R'_4	R_5	–2 compound	IC_{50} (μM)
OH	H	OH	OH	H	CH_2OH	Glc	1.7 (5.6)
H/OCH_3						2d-/2-*O* Me-Man	28/4.4
OH		H/OCH_3				3d-/3-*O* Me-Man	390/350
		OH	H/OCH_3			4d-/4-*O* Me-Man	ND/ND
			OH		CH_3/CH_2OCH_3	6d-/6-*O* Me-Man	10/400
H	OH				CH_2OH	Man	(25.1)
OH	H		H	OH		Gal	10
			OH	H	H	Xyl	ND

R'_2	R_4	R_5	R_N	–1 compound	IC_{50} (μM)
OH	OH	OH	H	DMJ	1.7 (5.6)
H/OCH_3				2d/2-O Me-DMJ	ND/300
OH	H/OCH_3			4d/4-O Me-DMJ	ND/ND
	OH	H/OCH_3		6d/6-O Me-DMJ	ND/ND
		OH	CH_3	*N*-Me-DMJ	ND
			C_3H_7	*N*-Pr-DMJ	ND
			C_4H_6-COOH	*N*-5-carboxypentyl-DMJ	ND

Figure 1.13: The pioneering work on GH99 inhibition. (**A**) Various Glc-linked –1 subsite compounds and their IC_{50} values. IC_{50} is the inhibitor concentration at which an enzyme catalyzes the reaction at 50% of the rate of the uninhibited reaction. (**B**) Various –2 subsite compounds linked to DMJ and their IC_{50} values. (**C**) Modifications on Glc-linked DMJ and their effects on the IC_{50} values. Abbreviations used: d – deoxy, dd – dideoxy, Me – methyl, ND – not determined. The magnitude of IC_{50} depend on the substrate concentration,[155]. As they cannot be directly compared, the values from Ref. 153 are given without parentheses, and from Ref. 154 with parentheses.

GlcDMJ was inhibiting 92-100% of the endomannosidase activity when tested at 80 μM (micromolar) with cultured cell membrane extract in vitro.[153] GlcDMJ was cleaved by the enzymes of mouse membrane extract, but remained intact when CST was added to the extract.[153] A report of the synthesis of Glc- and Man-linked α- and β-homomannojirimycin was published, but there are no reports of testing their potency.[156] An alternative approach to GlcDMJ synthesis was published in 2004.[157] It was not until 2012 that reports of new GH99 inhibitors were published: Fleetamine (GlcSW) was a potentially powerful inhibitor, originally proposed in 1993,[154] but its synthesis was not attempted. When produced, however, the compound did not bind to bacterial GH99, which by then had become a more convenient testing platform.[158] By contrast, Glc-isofagomine (GlcIFG) was found to be a better GH99 inhibitor than GlcDMJ by a factor of 38.[148] After that, a 4.5-fold improvement in binding to the bacterial GH99 was made by synthesizing and testing ManIFG.[159]. Structures of these inhibitors are shown in Figure 1.14.

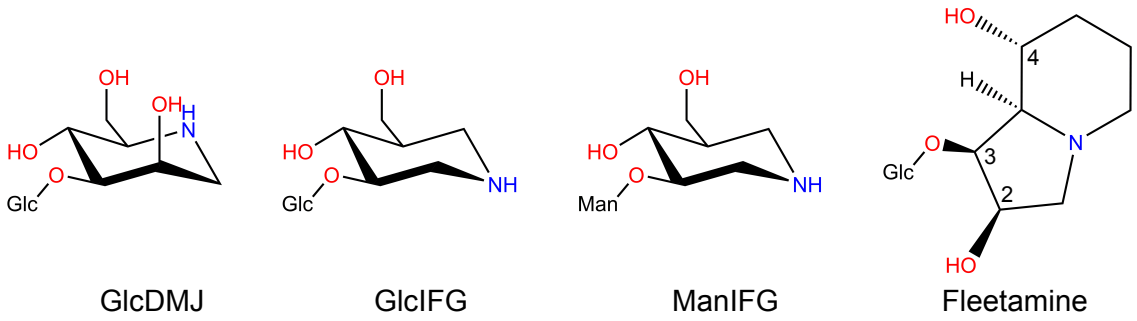


Figure 1.14: Different designs of actual and potential inhibitors of the endomannosidase activity. The numbering on Fleetamine represents the –1 subsite pyranose atom mimicked by the fused five- and six-membered ring.

1.4.5.3 An "alternative" pathway for complex glycans processing?

The third pioneering study of the endomannosidase concerned its activity in liver (hepatocellular carcinoma) cell line HepG2. Inhibition of glucosidases by DNJ and CST resulted in endomannosidase activity markers G1-3M being produced in the growth media, although in a different pattern[160] consistent with CST being more effective than DNJ at inhibiting ER glucosidase I. About 50% of the *N*-glycans produced by the cells were complex-type, confirming previous observations and providing evidence that the endomannosidase was responsible for this effect. Endomannosidase activity was then detected in the brain and purification of a protein that was its source was reported.[161] The isolated enzyme had a pH optimum of 6.2–6.8. It is now known that mammals have two genes encoding endomannosidase enzymes: *MANEA* (mannosidase endo-alpha) and *MANEAL* (mannosidase endo-alpha-like). *MANEAL* is expressed predominantly in the central nervous system (see Figure 4.1 in Section 4.2) and *MANEA* in other tissues. Tulsiani *et al.* might have, unknowingly, described an an isozyme of the one discovered by Lubas and Spiro.

In ER glucosidase II-deficient cells treated with CST, the predominant (91% molar) *N*-glycans were complex-type, showing that endomannosidase can process glycans on most proteins that are made in cells.[150] Subsequent results showed that endomannosidase does not discriminate between folded and unfolded proteins.[162–164]

Some results led the scientists to believe that a proportion proteins do not reach the compartment in which endomannosidase resides.[165] Rabouille and Spiro[162] found that the susceptibility of glycoproteins to endomannosidase cleavage is probably time-constrained, as fewer glycans were processed with increasing number of glycosylation sites in CST+DMJ-treated cells. As the GH99 activity towards triglucosylated glycans is low, in heavily *N*-glycosylated proteins many glycans remained uncleaved. The same study[162] is the first to describe the technique of investigating endomannosidase activity by comparing the electrophoretic mobility of glycoproteins treated or untreated with *endo*-β-*N*-acetylglucosaminidase H (endo H).

Once endomannosidase inhibitors became available, it became apparent that Chinese hamster ovary (CHO) cells do not have active mannosidase in their enzyme repertoir,

and they can be used as a convenient control to investigate the effectiveness of GH99 inhibition. Without GlcDMJ, the endo H-cleavable glycans produced by BW5147 cells (mouse lymphoma) were mostly M9Gn2 and M8Gn2. When *exo*mannosidases and glucosidases were inhibited, G3M9-7Gn2 and M8Gn2 became the predominant species, and M8Gn2 corresponded to endomannosidase product. With GlcDMJ, only G3M9-8Gn2 were detected, suggesting that some demannosylation (consistent with Golgi mannosidase IB activity) still took place. In CHO cells, G3M9Gn2 was detected in conditions of exoglycosidase inhibition.[153] The lack of endomannosidase in CHO cells was a finding that added another layer of consideration for researchers when choosing a cell line in which to conduct glycobiology research.

The products of endomannosidase activity are rapidly processed by Golgi mannosidases (IB and IC under the current understanding) and cannot be observed without the use of exomannosidase inhibitor. It was confirmed that the endomannosidase pathway functions without a glucosidase blockade.[166] In a seminal publication, Karaivanova *et al.* determined that the G protein[§] from vesicular stomatitis virus (VSV) and influenza haemagglutinin is processed by endomannosidase, the extent of which was host-cell dependent.[168] HepG2, PtK1 (*Potorous tridactylus* kidney) and BHK (baby hamster kidney) cells were found to utilize the endomannosidase pathway heavily, while MDBK (Madin-Darby bovine kidney) and MDCK (Madin-Darby canine kidney) not at all. The utilization level did not correspond to enzymatic activity in postnuclear membrane extracts. Rather, it seemed to depend on which cellular components proteins are trafficked through while their glycans are assembled (MDCK cells were shown earlier[153] to make use of the endomannosidase pathway). Differential mobility of uncleaved, endo-H-treated and PNGase F-treated proteins was used to establish differences in the level of utilization of the endomannosidase pathway by each cell line. Incubation of HepG2 cells with GlcDMJ caused all VSV glycans to be susceptible to endo-H cleavage, showing that this cell line uses the endomannosidase pathway heavily.

Subsequent research established that the low activity of endomannosidase in MDBK and MDCK cells was determined genetically and not physiologically. When bovine and canince *MANEA* were overexpressed, the protein products had an altered speci-

[§]G protein (glycoprotein) from VSV and related viruses are essential for the processes of cell attachment and membrane fusion. These viruses (enveloped viruses) enter the by fusion of the viral membrane with the host cell membrane.[167]

ficity and lower activity than the human version. The bovine enzyme strictly processed monoglucosylated and not di- or triglucosylated glycans; this result pointed to the physiological significance of endomannosidase pathway.[169] The canine enzyme showed a preference towards monoglucosylated glycans more pronounced than the human enzyme, but – unlike the bovine enzyme – it was not restricted in its processing. Shortly after, a study in B16 mice melanoma cells showed that the endomannosidase pathway is preferentially utilized by the cell when it produces tyrosinase-related protein-1, but less so with tyrosinase.[170]

FOS that are by-products of ERAD (and molecular markers of ERAD having taken place) are transported to the Golgi and processed by MANEA there. When cells are treated with brefeldin A (BFA), which causes fusion of parts of the ER and Golgi, endomannosidase activity can also be detected in the ER.[171] The trafficking of the enzyme to the Golgi is COPII-mediated.[172] Endomannosidase knockdown increases the total amount of FOS produced and ERAD can occur downstream of the cleavage of glycans by this enzyme. When *MANEA* was overexpressed, small FOS, which implied endomannosidase cleavage before ERAD, were observed during glucosidase inhibition.[173] In a later study, Alonzi *et al.*[122] showed that in addition to total FOS, glucosylated FOS accumulating in the ER lumen in cells lacking the endomannosidase activity can serve as a marker for ERAD. Only non-glucosylated FOS are substrates for a transporter that trafficks them across the membrane (from the ER to the cytosol),[174] and they accumulate in the ER in cells lacking an active form of endomannosidase. This model, however, necessitates the existence of ER-resident PNGase, which has not been confirmed yet, although some data suggest it is ER membrane-associated.[175]

Expression of endomannosidase genes is tissue-specific and cell-specific. The highest level of the enzyme mRNA was found in the lung and the liver, moderate in the brain and kidney and low in spleen, muscles and kidneys. This did not directly translate to the level of enzymatic activity, which was the highest in lungs and the liver, and moderate in all other cells.[176] Immunofluorescence showed the protein colocalizes with Golgi mannosidase II, but is not present in endothelial liver cells, unlike the other, ubiquitous enzyme. It was also not detected in other epithelia and adrenal cortex and medullar cells.[177] The general but not definitive pattern was that endomannosidase is expressed in cells that need to secrete high amounts of glycoproteins. A later study confirmed the Golgi localization using both immunofluorescence and im-

munogold labelling.[129] In particular, 84% of gold particles were found over *cis*/*medial* Golgi, 15% over ERGIC and 1% over the rough ER. *Trans*-Golgi and TGN were not labelled. Presence of DTT (1,4-dithiothreitol) arrests trafficking of proteins from the ER.[178] It was shown that when this process is halted, glycans do not undergo cleavage by endomannosidase, again providing evidence of Golgi localization.[179]

In some cases, endomannosidase activity was directly observed in humans. In an individual with an incredibly rare CDG-IIb[¶] (ER glucosidase I deficiency), Glc_3Man tetrasaccharide, the product of endomannosidase activity, accumulated in urine. This unusual FOS could not be detected at all in control urine.[181] Tissue samples showed no increase in the activity of the enzyme. The patient died at 74 days of age. A follow-up report which used cultured fibroblasts from the infant showed a dramatically altered pattern of endo H-released oligosaccharides which was normal in the father and only slightly altered in the mother. ~80% of total glycans were correctly processed and only ~16% were G3M9Gn2, showing the extent of rescuing capability of endomannosidase in humans. As in cell culture studies during glucosidase blockade,[160] about 50% of all *N*-glycans were complex-type. The procedure for measuring MANEA activity was refined and showed that in the child cell extract it was 2.5 times more active than in controls (1.8 × in parents' cells). This led the authors to speculate that ER glucosidase I and *endo*-α-1,2-mannosidase pathways do not operate independently, but their capacity may be tuned depending on the situation.[182] A pair of siblings with CDG-IIb were alive at ages of 6 and 11.[183] There are no reports of CDG caused by endomannosidase deficiency.

No evidence of endomannosidase activity was detected in the protozoan *Trypanosoma cruzi* (incidentally, one of the earliest reports of observation of UGGT activity).[184] Literature produced before the invention of Next-Generation Sequencing (NGS) suggests this pathway of *N*-glycan processing is limited to chordates with one exception: it was also detected in mollusks. Interestingly, the most active endomannosidase seemed to come from a species of frog, *Rana catesbeiana*.[185] Currently available data are reanalyzed in Section 6.1.

A number of studies found *MANEA* genetic variants associated with with various

[¶]Congenital disorders of glycosylation are separated into two types: type I – "referring to defects in the initial steps of *N*-linked protein glycosylation", and type II – "defects in the processing of protein-bound glycans or the addition or [*sic*] other glycans to the protein".[180]

ailments. The allele rs1133503 C was a top-ranked SNP (single-nucleotide polymorphism) associated with cocaine-induced paranoia (CIP).[186] It is located in the 3′ UTR (untranslated region) of the *MANEA* gene. The discovered allele had an overall frequency of 41% in the surveyed population. A study that was prompted by this finding detected another SNP, whose association with CIP was even higher: rs9387522 A, also in the 3′-UTR. A number of other CIP risk-associated SNPs were also found in the coding and promoter regions, as well as an association of eight markers with cocaine dependence.[187] In another study that was sparked by the discovery of rs1133503, association of the C allele with panic disorder and generalized social anxiety disorder was found to be statistically significant. In cells with both C and T alleles present (heterozygous), 45% less T-allele mRNA was present, and the overall level of mRNA in TT homozygotes was 67% of that in CC and CT cells.[188] In the Han population in mainland China, a SNP close to *MANEA* was found to be associated with sporadic Parkinson's disease.[189] Using a novel approach of integrative genome-wide association (iGWAS, integrative genome-wide association study), *MANEA* was identified as a susceptibility gene for asthma.[190] It was also discovered that *MANEA* expression was decreased in response to androgens.[191] Androgen-responsive genes are potential targets for anticancer medication, as high level of androgens is associated with a higher incidence of prostate cancer.

Two general conclusions can be drawn from the body of knowledge about the enzyme. First, endomannosidase cleavage is an event that ultimately prevents proteins from re-entering the calnexin/calreticulin cycle. Second, the enzyme has probably evolved in order to cleave monoglucosylated substrates irrespective of the folding state and not act upon non-glucosylated *N*-glycans. What seems to be often missed in literature is the consideration that when proteins have multiple *N*-glycosylation sites, it is likely that not all *N*-glycans will be processed in the same way before the protein exits from the ER. Only some glycosylation sites may serve as binding sites for chaperone lectins. The presence of the endomannosidase pathway ensures that all *N*-glycans will be available for processing by the Golgi machinery: its localization within ERGIC and *cis/medial* Golgi strongly points to such a function *in vivo*. As such, the pathway does not look as an alternative, but an integrated part of *N*-glycosylation machinery in animals, which ensures glycans can be processed further.

1.4.5.4 Sequence, domain organization and 3D structure of endomannosidase

An affinity-based method of purifying endomannosidase was described in 1994.[192] The group used a non-cleavable derivative of the –2/–1 subsite ligand, GlcMan, which was previously found to be weakly inhibitory.[153] The purified rat liver endomannosidase migrated as two bands and had an apparent molecular weight (Mw) of 56 and 60 kDa. This was followed up and the 60 kDa turned out to be the chaperone calreticulin.[193] Another study isolated the enzyme from pig liver and again it was observed as two bands of mobility corresponding to an Mw of 48 and 50 kDa and not glycosylated (which was also confirmed for rat enzyme in Ref. 172). Enzymes from both bands were active, leading the authors to a conclusion that the enzyme undergoes nonspecific proteolysis.[194]

The gene encoding the rat endomannosidase was cloned and expressed in *E. coli* in 1997 and the sequence of the protein product was not homologous to any other protein known at the time.[176] The enzyme was found to be 451 aa long and have an Mw of 52 kDa. It was claimed that a 8 kDa C-terminal peptide cleaved by *E. coli* proteases was not a part of the active site, which is dubious as the proposed active residues would be within this fragment. Most likely the sequence was cleaved but did not dissociate from the protein, which would destabilize the fold. That first rat endomannosidase sequence was shown to be an artifact of cloning seven years later. The N-terminal sequence preceding aa 55 came from a different chromosome.[195] Full-length, C-terminally GFP-tagged (green fluorescent protein-tagged) human and rat MANEA proteins were localized to the Golgi when produced in CHO-K1 cells, and an N-terminally truncated human protein was mislocalized. When a protein comprising residues 1–59 with a C-terminal GFP tag was produced, fluorescence was seen in the Golgi. This established that it is the N-terminal signal sequence that targets the protein to the Golgi, and that the enzyme is a typical type II membrane protein.[195] When N-terminally tagged with GFP before the signal sequence, the enzyme also localized to the Golgi.[196]

MANEA from CHO cells is inactive.[153] Cloning, sequencing and translation of CHO *MANEA* mRNA revealed two cysteines found in no other endomannosidases known at the time: C177 and C188. In the human gene, these residues are, respectively, R and W. The authors of a study published in 2007[197] looked at "restoring" both of

these residues from the CHO sequence to the human sequence. Whether residue 177 was C or R had no effect, rescuing or detrimental, to the activity of the enzyme (the enzyme used for these substitution studies was CHO MANEA). Indeed, it was later found that residue 177 in the wild-type chinese hamster MANEA protein sequence is a cysteine when its genome was sequenced, annotated and the coding sequences translated (see Figure 1.15).[198] The C188 variants had no activity and localized to the ER, while the C188W variants were active and localized to the Golgi.[197] The ER mislocalization was caused by formation of an intracellular disulfide bridge between the residues. MANEA does not have any *N*-glycosylation sites and therefore its cysteines might be less likely to be shielded by the lectin-based ER chaperoning machinery. Another sequence feature that was investigated was a positively charged patch found in the cytoplasmic tail of endomannosidases from many species. Only the length of the tail was found to influence the Golgi localization of the enzyme, not the identity of amino acids: proteins with shortened tails were retained in the ER.[199]

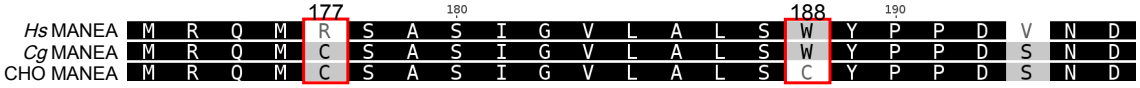


Figure 1.15: An excerpt of multiple sequence alignment of *Homo sapiens* (*Hs*), Chinese hamster (*Cricetulus griseus*, *Cg*) and CHO-K1 cells MANEA proteins. Amino acid residues are coloured by similarity (BLOSUM62 score matrix with a similarity threshold of 1, black background: 100% similar, light grey: 60–80% similar, white: less than 60% similar; for details of the BLOSUM score matrix see Ref.[200]). Residues 177 and 188, talked about in the text, are in red boxes. Figure made in Geneious[201]; sequences aligned using the MUSCLE algorithm.[202]

Endomannosidase is post-translationally modified. Phosphorylation of the luminal domain of MANEA in the Golgi was detected after the protein found its way into the Golgi.[172] The PhosphoSitePlus database[203] lists experimentally confirmed post-translational modifications (PTMs). In the human MANEA, 12 phosphorylation sites (1 in the cytoplasmic tail), 4 *O*-glycosylation sites and 1 ubiquitinylation site were observed. Torossi *et al.*[172] could not, however, detect any *O*-glycosylation on their variant of the protein. They speculated that phosphorylation might be used in order to regulate its activity or its half-life, but also that it can be inconsequential. All potential *O*-glycosylation sites are situated at the C-terminal part of the stem domain of endomannosidase, which is a linker between the transmembrane helix and the catalytic domain.

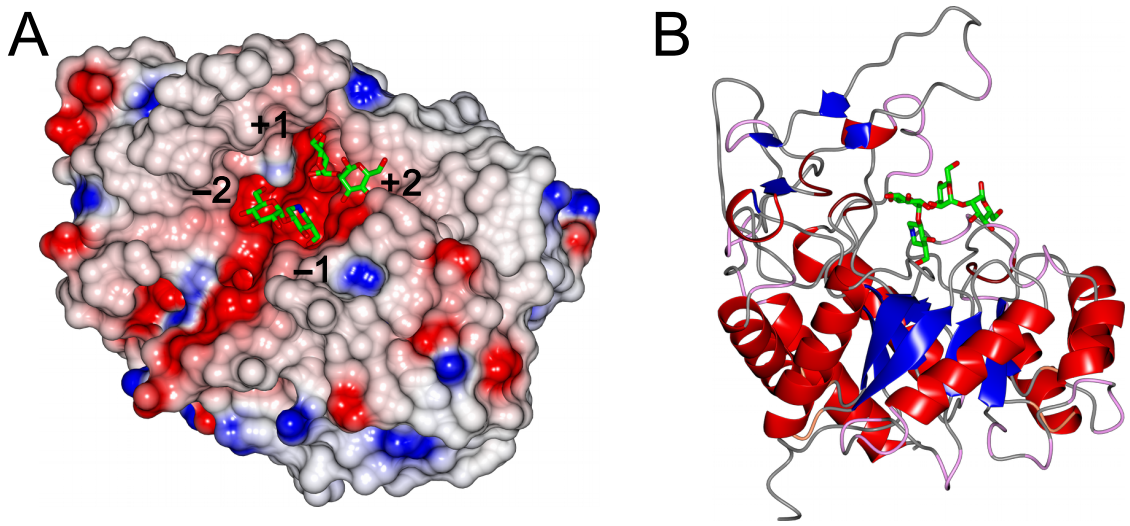


Figure 1.16: Structure of *Bacteroides xylanisolvens* GH99 *endo*-α-1,2-mannanase in complex with GlcIFG and α-1,2-mannobiose (PDB code **4AD4**, Ref. 148). (**A**) Electrostatic surface and ligands. Subsites annotated next to the sugar moieties. (**B**) Ribbon depiction showing the $(\beta/\alpha)_8$ barrel fold of the protein. Assembled in ccp4mg.[204]

Expression of a *Shewanella amazonensis* (prokaryote) gene encoding a GH99 protein and the finding that it had endomannosidase activity was reported in 2011.[205] 3-dimensional structures of two bacterial GH99 endomannosidases whose natural substrate is yeast mannan (and were therefore dubbed *endo*-α-1,2-mannanases) were published by Thompson *et al.* in 2012.[148] *Bacteroides thetaiotaomicron* (*Bt*) and *Bacteroides xylanisolvens* (*Bx*) that live inside the human gastrointestinal tract were source organisms, and the genes were expressed in *E. coli*. These proteins differ from the endomannosidase by not having a type II membrane protein topology, but instead being fully soluble. The structure revealed a new fold: a modified $(\beta/\alpha)_8$ barrel (Figure 1.16**B**) and a groove-like binding site typical of endo-acting enzymes.[206] Ligand soaking experiments helped define four subsites, from –2 to +2, within the groove (Figure 1.16**A**). Most amino acids that directly coordinated the ligand were conserved between the bacterial and the mammalian enzymes, implying similar modes of action. The only exception was a tryptophan residue close to the –2 sugar moiety, which in yeast mannan would be Man, and in the *N*-linked glycan, Glu. In mammalian endomannosidases this is a tyrosine residue. The bacterial endomannanases take part in "selfish" utilization of the yeast glycan by gut endosymbionts.[207]

Of particular importance, the catalytic amino acid residues were discovered. These are two Glu residues 333 and 336 (*Bx*GH99 numbering) that correspond to E329 and

E332 in *Bt*GH99, E404 and E407 in the human MANEA and E406 and E409 in the human MANEAL. The *Bt*GH99 E329A had no activity, and the E332A mutant was 50-fold slower than the WT (wild-type) enzyme. The structure of a Michaelis complex of the inactive *Bx*GH99 mutant with ManMan-4-methylumbelliferone (ManManMUF) was published in 2015, but the ligand was bound in an unorthodox position, which prevented the authors from drawing sweeping conclusions about the real structure of substrate-bound GH99. Its catalyzed hydrolysis with the WT enzyme was remarkably slow, as would be expected from the unorthodox positioning of ManManMUF (the hydrogen bonding of bound ManManMUF was unlike any other *Bx*GH99 structures in complex with ligands; this will be elaborated further in Section 2.2).[159]

A structure of a GH99 family protein from the marine flavobacterium *Ochrovirga pacifica* without the catalytic residues has been published recently, but it bears resemblance to *Bt*/*Bx*GH99 proteins only in its overall fold. The residues in the marine GH99 corresponding to the catalytic residues in GH99 endomannosidases are not likely to catalyze the hydrolysis of a glycosidic bond, and its binding cleft is deeper, wider and longer. The authors of the study speculate that this protein might bind, but not degrade, a mannose-rich glycan.[208]

1.5 Implications of glycosylation in health and disease

1.5.1 Glycosylation and viral diseases

Many viruses are known to depend on their glycans for retaining infectivity. A well-known example is HIV (human immunodeficiency virus), which can lose its ability to infect cells if some of its glycosylation sites are erased. The envelope protein gp120 is glycosylated at approximately 25 sites.[209] The removal of any one of six specific sites within its V1/V2 and C1/C2 domains causes a complete loss of infectivity of HIV, and when others are mutated it can become more infective.[210] The glycosylation was found to be dispensable for proper folding of this protein, but important for the virus to evade antibody-mediated neutralization.[209] Surprisingly, gp120 viral coat contains a large number of hardly processed, high-mannose *N*-glycans, mainly $Man_{9-8}GlcNAc_2$ (a "mannose patch",). This is caused by a high local density of glycosylation of this

protein (see Figure 1.17), which sterically prevents mannosidases from accessing the sites.[211–213]

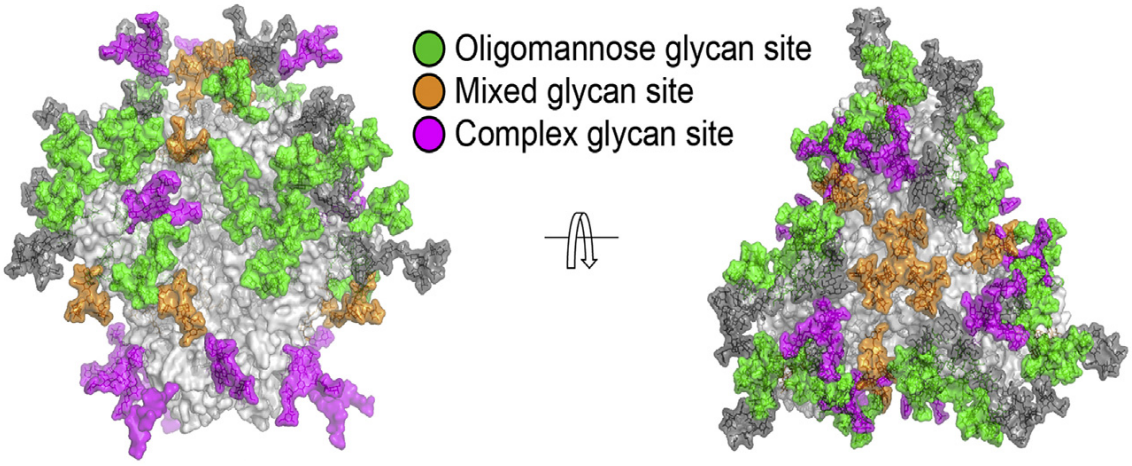


Figure 1.17: Identity of *N*-glycans at various sites on HIV Env (envelope) trimer. Figure adapted from Ref. 212.

Inhibition of glucosidases using *N*-butyl-deoxynojirimycin (*N*B-DNJ) caused G3M9Gn2 glycans to be present on SARS-CoV (severe acute respiratory syndrome coronavirus) spike glycoprotein,[‖] and changed its overall glycosylation pattern. As would be expected, some glycans on the spike glycoprotein were processed due to the action of endomannosidase.[215] Viral proteins are dependent upon ER chaperones for their correct folding and it was speculated that misfolding might have been responsible for the previously observed observed reduction in infectivity.[216] *N*B-DNJ inhibits HIV entry into the immune cells by disrupting the shedding of its envelope after it binds to the CD4 (cluster of differentiation 4) glycoprotein at the cell surface, as well as lowering the amount of viruses released from cells. The inhibitor also increases binding of antibodies to a loop region V3 (variable 3).[216,217] The drug was tested in HIV-infected people. It showed some antiviral activity, especially in patients with previous exposure to a second tested drug zinovudine, but *N*B-DNJ had unpleasant gastrointestinal side effects (osmotic diarrhoea, weight loss).[218] The drug, at lower concentrations (100 mg 3 times per day vs 1500–3000 mg per day in the antiviral therapy study) is approved to treat Gaucher disease type I, a lysosomal storage disease.[219] Due to a potential for broad-range antiviral activity but unpleasant side effects, it was proposed that glucosidase inhibitors are better suited to be used for acute, rather than chronic viral infections.[220] Cells from the

[‖] The SARS-CoV spike glycoprotein, similarly to the G protein, mediates attachment and membrane fusion. SARS-CoV is an enveloped virus.[214]

two siblings with CDG-IIb, mentioned in context of endomannosidase activity being observed in humans, were less susceptible to infections by viruses that depend on glycosylation for cell entry.[183] These cells mimic the effects of glucosidase inhibitors.

1.5.2 Glycosylation in cancer and other diseases

Glycans present on the cell surface are important in signaling to other cells. For example, tumour cells show a decrease in high-mannose glycans and a corresponding increase in branched, highly processed glycans. Their surface glycoprotein glycosylation patterns are reminiscent of those of cells undergoing mitosis – in fact, they are stuck in this part of the cell cycle.[221] In particular, the presence of tetraantennary, branched glycans with the core structure of a GlcNAc α-1,6-linked to the α-1,6-linked mannose was observed as the cause.[222,223] This particular branch can be extended in order to form poly-*N*-lactosamine-type glycans (Figure 1.11, page 24), which are recognized by galectin 3, a protein intimately involved in cell signalling, mediating inflammation and transition to chronic inflammation.[224] The branch can form only when enzyme MGAT5 attaches the α-1,6-linked GlcNAc to a previously triantennary glycan (previously acted upon by a sequence of enzymes: MGAT1, Golgi mannosidase II, MGAT2 and MGAT4) and is associated with tumours. The level of MGAT5 is elevated in neoplastic cells, and lower levels of MGAT5 reduce metastatic potential.[225–227] Remarkably, *Mgat5*$^{-/-}$ knockout mice show less metastases and live longer when tumours are induced in them while appearing otherwise healthy.[228] Later these mice were shown to be more susceptible to a number of automimmune disorders.[229] Galectin signalling activates a cascade of intracellular reactions, and blocking MGAT5 prevents this process.[230] Inhibitors of MGAT5 were developed,[231,232] but have not yet been tested clinically and better binding compounds are needed (the K_i of the best one was about 10 μM). The metastatic potential of a tumour is dependent on the length of the glycopolymers (mimics of the heavily glycosylated protein mucin) present on its surface, strongly suggesting that overproduction of mucin has a similar effect.[233]

Inhibiting key branching transferases shows anti-cancer activity,[234] but so do inhibitors of earlier steps of *N*-glycosylation. Tunicamycin, which inhibits DPAGT1[235] (see Figure 1.7 on page 14) has an antiproliferative effect[236]. CST appeared to

reduce tumour colonization potential, as well as increase the sensitivity of the neoplasm to other agents.[237] A combination of glucosidase and endomannosidase inhibition could offer an alternative approach to this. Due to their unique – in human cells – structure of an α-1,3-linked disaccharide (such as GlcDMJ shown in Figure 1.18), endomannosidase inhibitors can potentially be selectively inhibiting the endomannosidase activity only.

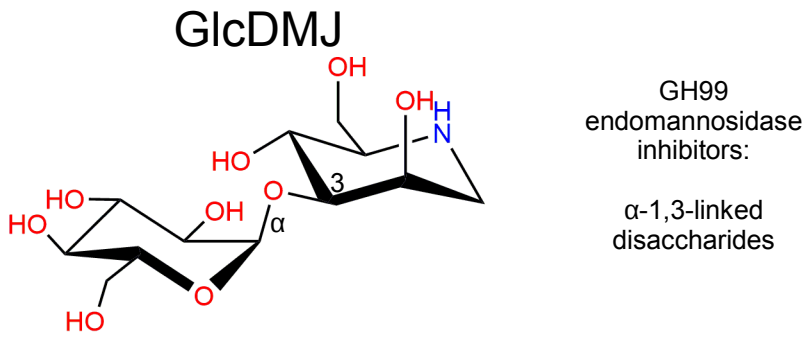


Figure 1.18: The general structure of an endomannosidase inhibitor is an α-1,3-linked disaccharide. Here, GlcDMJ is shown as an example.

19 types of CDG have been identified but none involving mutations in *MANEA* or *MANEAL*.[238] Nothing is known about the effect of the knockdown or knockout of these genes in living and developing organisms and the only data available comes from the previously mentioned GWAS (genome-wide association study).

1.6 Project aims and Thesis outline

Inhibition of glucosidases is a potentially universal antiviral therapy, as viruses depend on the cellular machinery to assume the correct fold. They need to hijack the host protein production systems in order to reproduce. When the system is disrupted, it is almost certain the host will be negatively affected, but also possible that the viral reproduction cycle will be severely disturbed. The potential for disruption of viral proteins is even higher if more than one element of the protein folding machinery is blocked. As endomannosidase can give some viral glycoproteins a way to mature in the Golgi even under conditions of glucosidase inhibition, applying inhibitors of this enzyme could enhance the effectiveness of the intervention. Specific glycan tags were also detected in cancer cells, and endomannosidase inhibitors may prevent those structures from forming, if combined with mannosidase or glucosidase inhibition

regimes. The main aims of the project described in this Thesis are listed below.

- Test the endomannosidase reaction mechanism hypothesis posed by Thompson *et al.*[148] using synthetic reaction intermediate mimics (Chapter 2).
- Use bacterial GH99s as a platform for testing potential novel inhibitors and measure their thermodynamic binding parameters (Chapter 3).
- Purify a recombinant version of human endomannosidase and characterize its Michaelis-Menten kinetics. Chapters 4 and 5 explore results of these efforts for two human GH99 proteins, MANEA and MANEAL.
- Determine the crystal structure of the human endomannosidase. Produce a crystal form able to accommodate ligands in its active site; investigate human endomannosidase in complex with its inhibitors in order to learn about the interactions in the active site and test the conclusions drawn by Thompson *et al.*[148] from the structure of the bacterial GH99. This part is presented in Chapter 4.
- Test the antiviral activity of endomannosidase inhibitors alone and in combination with glucosidase inhibitors using cells infected by a virus with a glycosylated structural protein (Chapter 5).

A compendium of small molecules that are most relevant to this Thesis can be found in Appendix A. Appendix B contains an explanation of Cremer–Pople parameters describing the conformations of furanoses and pyranoses. Published work that was contributed to but not the focus of this Thesis is presented in Appendix C: Bibliography. It includes determination of the oligomeric state of *Talaromyces pinophilus* GH62 arabinofuranosidase and its dissociation constant with AraDNJ (L-arabinodeoxynojirimycin) and crystallization experiments on *Bacteroides ovatus* GH43B protein. The appendices are followed by List of Abbreviations that were used in the Thesis. The list of references is in the last section, entitled References.

Chapter 2

The *endo*-α-mannosidase and *endo*-α-mannanase mechanism

2.1 Abstract

The knowledge of a reaction mechanism that an enzyme uses is key to understanding its biochemistry. It can guide rational design of inhibitors, as well as reveal unique chemical transitions. The mechanism of GH99 *endo*-α-1,2-mannosidase has been a puzzle since the first structures with inhibitors bound in the –2/–1 subsites were published. The Koshland double displacement mechanism that many retaining mannosidases follow was not likely: none of the residues around the –1 sugar were good candidates for a nucleophile. In this chapter, data supporting hypothesis of an alternative, neighbouring group participation reaction mechanism are presented. Mimics of the proposed reaction intermediate and related compounds were synthesized by collaborators and tested with the enzyme and its mutants. Only one such compound, which resembled the α-1,2-anhydrosugar intermediate most closely, was a substrate for the bacterial endomannosidase (endomannanase). 3-dimensional structures of the wildtype enzyme and its mutants with the mimic were obtained and supported the kinetic data, as well as structures with the minimal substrate and the product. QM/MM calculations suggest the overall conformational itinerary is: ${}^{4}\boldsymbol{C}_{1}\rightarrow{}^{2}\boldsymbol{H}_{3}\rightarrow[\boldsymbol{E}_{3}]^{\ddagger}\rightarrow{}^{4}\boldsymbol{E}\rightarrow[\boldsymbol{E}_{3}]^{\ddagger}\rightarrow{}^{4}\boldsymbol{C}_{1}$. This mechanism is unique among mannosidases and enzymes in general. These insights are a crucial addition to the understanding of the process of *N*-glycosylation.

2.2 Introduction

Bacterial *endo*-α-1,2-mannanases have been used as a model for mammalian *endo*-α-1,2-mannosidase due to the ease at which they are expressed in *E. coli* and because they offer a high likelihood of crystallization. Other factors include belonging to the same CAZy family (GH99) and therefore an increased likelihood of a similar mode of action. In a 2012 paper,[148] Andrew Thompson *et al.* suggested that the reaction mechanism of this enzyme is not a canonical retaining mannosidase double-displacement mechanism proposed by Koshland in 1953[84] (Figure 2.1**A**) and true for many previously characterized retaining α-mannosidases. The reaction intermediate in the proposed reaction mechanism for GH99 endomannosidases would be a disaccharide with a 1,2-β-anhydrosugar in the –1 subsite (Figure 2.1**B**).

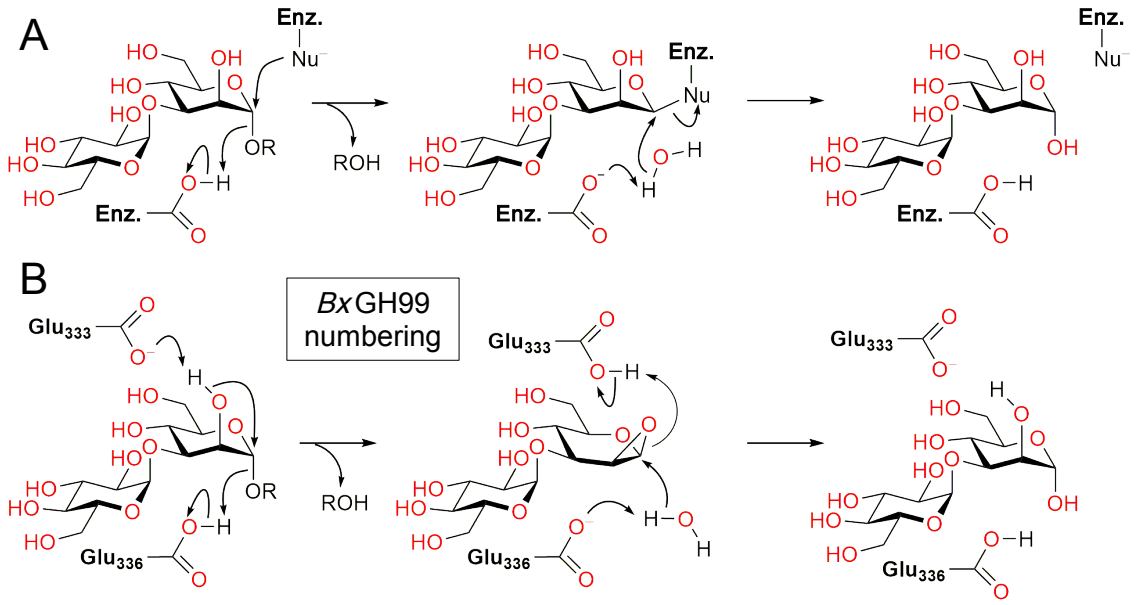


Figure 2.1: Possible mechanisms for *endo*-α-1,2-mannosidase and *endo*-α-1,2-mannanase. (**A**) General double displacement, retaining mechanism proposed by Koshland[84] for glycoside hydrolases (GH). (**B**) Mechanism involving a 1,2-β-anhydrosugar reaction intermediate proposed by Thompson *et al*[148] (*Bx*GH99 numbering of the catalytic residues).

The reasoning behind this suggestion came from an observation that in a complex with a classical GH99 inhibitor,[153] GlcDMJ, none of the hydroxyl groups of the catalytic glutamate residues, E333 and E336, were within 2.6–2.8 Å of the pseudoanomeric atom (PDB code **4AD3**). The same observation was made for a GH99 inhibitor whose synthesis was first reported in the 2012 paper, α-Glc-1,3-isofagomine (GlcIFG, PDB code **4AD2**). Additionally, the inhibitor was not distorted from the

relaxed $^4C_1$ chair conformation when bound to the enzyme. Through soaking of the crystals with both the inhibitor disaccharide and α-1,2-mannobiose, four subsites, from –2 to +2, were defined for the enzyme (PDB codes: **4AD4**, **4AD5**, see Figure 2.2**A**).

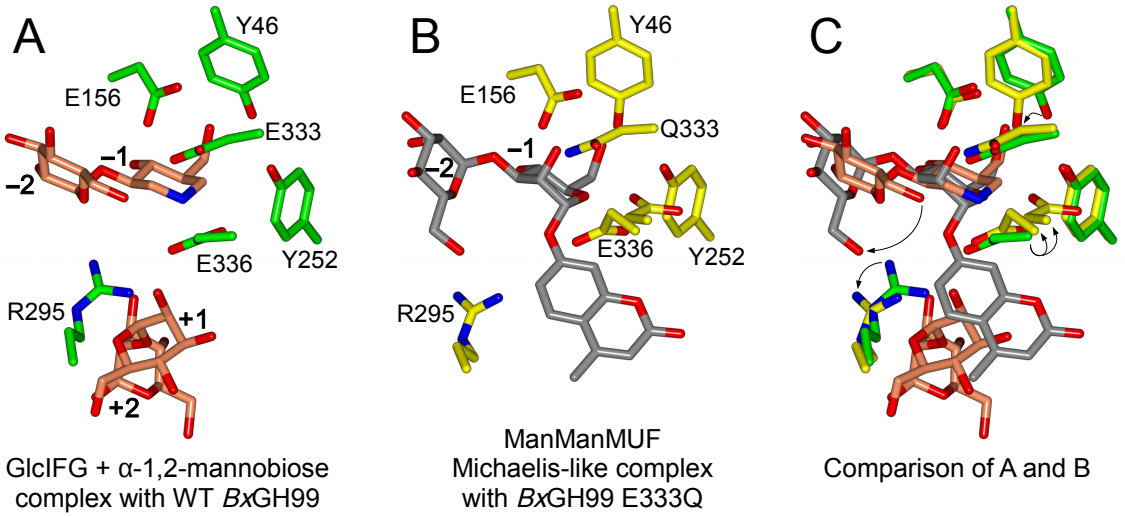


Figure 2.2: Position and conformation of ligands in the binding site of *Bx*GH99 and amino acid residues around the –1 subsite. (**A**) Structure of WT *Bx*GH99 in complex with GlcIFG and α-1,2-mannobiose (PDB code **4AD4**). Coral – carbons of the small molecules, green – polypeptide carbons. Subsites and aa residues are annotated. (**B**) Structure of E333Q *Bx*GH99 in complex with ManManMUF (PDB code **4V28**). Note E336 exists in two conformations. Grey – carbons of the small molecules, yellow – polypeptide carbons. Subsites and aa residues are annotated. (**C**) Comparison of **A** and **B** with major atomic position diffrences highlighted using arrows. Coloured as previously. Subsite and residue annotations omitted for clarity. Assembled in ccp4mg.[204]

*Bt*GH99 mutants E154A and E329A had zero activity, and mutant E332A was 50 × slower than the WT. These residues are equivalent to E156, E333 and E336 in *Bx*GH99. A Michaelis-like complex of the E333Q *Bx*GH99 mutant with ManManMUF was published in 2015.[159] However, ManManMUF is a remarkably poor substrate for the endomannanase and the structure (PDB code **4V28**) provides some clues why. The –2 sugar of the ligand appears bound in an unorthodox position (Figure 2.2**BC**). Notably, the OH6 of the –2 sugar does not form a hydrogen bond with E336 OH – an interaction which was observed in all other structures of *Bx*GH99 with ligands occupying the –2/–1 subsites (interatomic distance 6.8 Å instead of 2.6–2.8 Å for others). MUF is also situated in the active site less ideally than the natural +1/+2 leaving group, α-1,2-mannobiose. In the same article, an endomannanase inhibitor with the best affinity to the enzyme (ManIFG, PDB code **4V27**) was reported. The 4.5-fold increase in affinity over GlcIFG was achieved by changing the –2 sugar moiety from Glc to

Man, supporting the claim that the enzyme has been evolutionarily tuned for yeast mannan, rather than mammalian *N*-glycans.

A neighbouring group participation reaction mechanism was observed for hexosaminidases from families GH20,[239–241] and GH18,[242] as well as human *O*-GlcNAcase from family GH84,[243–246] GH56[247] and GH85.[248] In 2016, Spencer Williams' group[149] presented kinetic isotope effect (KIE) evidence that the OH-catalyzed hydrolysis of 1-*p*NP-α-D-mannopyranoside (1-*para*nitrophenyl-α-D-mannopyranoside) consistent with the transition state being a 1,2-anhydrosaccharide (see Figure 2.3). The reaction intermediate for NaOH-catalyzed solvolysis of α-D-mannosyl fluoride is known to be the same compound.[249]

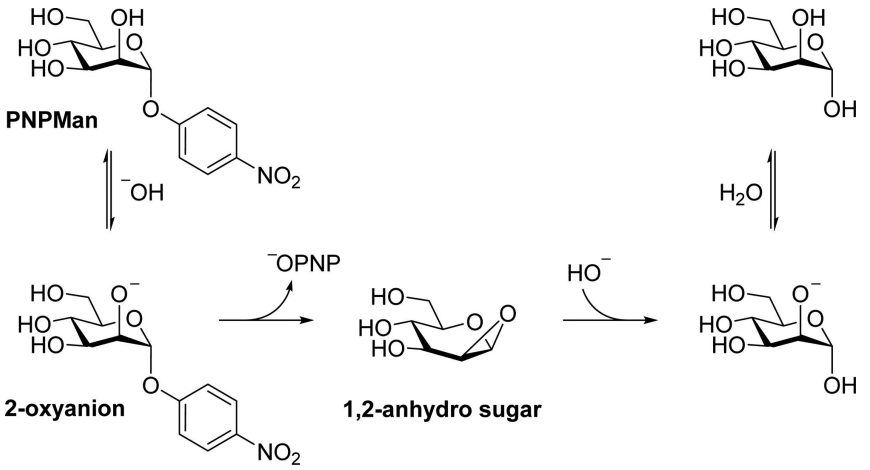


Figure 2.3: Alkaline hydrolysis of 1-*p*NP-α-D-mannopyranoside. Figure from Ref. 149.

The KIE is defined as the ratio of the reaction rate on a compound with the light isotope to that on the compound with the heavy isotope.[250] The change in reaction rate is caused, in theory, by vibrational differences between the atoms due to their slightly different mass.[251] In general, two types of kinetic isotope effects are measured: primary and secondary. Primary isotope effects are measured for atoms that directly participate in a reaction by forming or breaking bonds and secondary isotope effects for atoms that do not.[250] As these KIEs stem from very minimal changes to the natural substrate, they offer insights into the identity of the generally unobservable TS (transition state) that are not available through other methods. For example, the magnitude of a KIE can provide information about the bonding environment around the labelled atom: if it is more than 1 (normal) it is less restricted in the TS and if it is less than 1 (an inverse KIE) it is more restricted and the compound containing the heavy

isotope reacts more quickly than the light one.[252]

That none of the –1 sugars are distorted when bound to the enzyme is significant. In α-D-mannose, the OH2 group is in the axial position when mannose is in a $^{4}\boldsymbol{C_1}$ conformation. This causes an effect destabilizing the sugar conformation when a nucleophilic attack at the anomeric carbon is attempted – collectively called the Δ2 effect.[253] α-Mannosidases need to overcome this effect in order to hydrolyze the glycosidic bond, which is usually accomplished by distorting the conformation of the –1 sugar. For inverting α-mannosidases, the observed conformational itineraries are: $^{O}\boldsymbol{S_2}\rightarrow[\boldsymbol{B_{2,5}}]^{\ddagger}\rightarrow{}^{1}\boldsymbol{S_5}$ (GH125)[254] and $^{3}\boldsymbol{S_1}\rightarrow[^{3}\boldsymbol{H_4}]^{\ddagger}\rightarrow{}^{1}\boldsymbol{C_4}$ (GH47)[29,255]. These enzymes are *exo*-acting and the inhibitors used to elucidate their reaction mechanism are pseudo-monosaccharides like kifunensine or nojirimycin. Swainsonine[256] is a powerful inhibitor of glycoprotein processing, acting on Golgi α-mannosidase II and lysosomal α-mannosidase.[257] Fleetamine (α-Glc-1,3-swainsonine) was proposed in 1993[154] as a potentially tightly-binding inhibitor of *endo*-α-mannosidase, and its synthesis was published almost two decades later[158]. However, neither ITC, nor soaking experiments on *Bx*GH99 showed any binding of Fleetamine to the enzyme. Taken together, these observations about GH99 show this enzyme has unusual features: its active site, and possibly the mechanism of action is unlike that of other mannosidases. But they do not provide direct evidence that would prove or disprove the hypothesized GH99 mechanism. Notably, the lack of a "true" Michaelis complex has been problematic in ascertaining how GH99 works. In this chapter I will present snapshots of the GH99 reaction mechanism at the point of substrate binding, reaction intermediate and the product. Additionally, data on the activity of four variants of the enzyme on various intermediate mimics with correct and incorrect stereochemistry at C1 and C2 will be shown and interpreted.

2.3 Materials and methods

2.3.1 Gene expression

Genes encoding GH99 endomannosidase from *Bacteroides thetaiotaomicron* (UniProt ID: **Q8A109**) and *Bacteroides xylanisolvens XB1A* (UniProt ID: **D6D1V7**) were expressed in *E. coli* BL21(DE3) cells (Agilent). The gene encoding *Bt*GH99 was cloned into pET-YSBL3C vector, which is a pET28a(+) vector modified by the addition of a 3C protease cleavage site. Cloning of *Bt*GH99 was done by Dr Tracey Gloster before the commencement of this Thesis work. The gene encoding *Bx*GH99 was cloned into a canonical pET28a(+) vector (between *NdeI* and *XhoI* restriction endonuclease sites). Cloning of *Bx*GH99, as well as mutagenesis of *Bx*GH99 in order to produce genes encoding *Bx*GH99 E333Q and E336Q was done by Dr Andrew Thompson before this Thesis work was commenced. Competent cells were made using a modified one-step protocol[258] where LB (lysogeny broth) media in TSS (transformation and storage solution) was substituted with MQ (ultrapure) H_2O. The competent cells were transformed with the DNA plasmids using a traditional 45 s at 42 °C heat-shock method with subsequent 1 h recovery in SOC (**s**uper-**o**ptimal broth with **c**atabolite repression) media[259] and plating on LB agar containing 50 mg ml^{-1} kanamycin (Kan). After overnight incubation at 37 °C, single colonies were selected from the plate and shaken at 180 RPM (revolutions per minute) in 4 × 8 ml Kan^+ LB media overnight at 37 °C in order to grow starter cultures. The following morning, the 8 ml starter culture was added to 750 ml of Kan^+ (with a working concentration of kanamycin added) LB media in 2 l flasks. A total of 3 l (4 × 750 ml) media was used at this stage. The cultures were shaken at 180 RPM and 37 °C until the OD_{600} (optical density at 600 nm) was 0.8–2, at which point expression was induced by adding sterile IPTG (isopropyl β-D-1-thiogalactopyranoside) solution up to a concentration of 0.2 mM. After this step, the cultures were shaken at 180 RPM at 16 °C overnight. Cells were harvested by centrifugation for 20–30 min at >5000 × g (multiples of gravitational acceleration on the Earth's surface) and frozen at -80 °C. *Bx*GH99 mutants: E333Q and E336Q were produced using the same protocol.

The cell pellet was resuspended in buffer A (25 mM HEPES pH 7.0, 300 mM NaCl, 20 mM imidazole, 1 mM DTT) until homogeneity and then cells were disrupted by

sonication on ice. The solution for the next step was darker and more viscous. It was centrifuged at ~3900 × g for 10 min and loaded onto a 5 ml HisTrap FF Crude column (GE) preequilibrated with buffer A. The column was washed with 30 ml of buffer A, and then the bound protein was eluted using an imidazole gradient through mixing in buffer B (25 mM HEPES pH 7.0, 300 mM NaCl, 500 mM imidazole, 1 mM DTT). Fractions containing the protein of interest were collected and concentrated using an AMICON 30 kDa cut-off ultrafiltration device. The resulting solution was loaded onto a 120 ml Superdex S75 16/60 gel filtration column, preequilibrated with protein storage buffer (25 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT). The fractions containing pure protein were combined and concentrated using an identical ultrafiltration device. The final concentration was 30–50 mg ml^{-1}. The protein solution was divided into 50 µl aliquots in 1.5 ml eppendorf tubes, flash frozen in liquid nitrogen and stored at -80 °C.

2.3.2 Chemicals, assays and mass spectrometry

The compounds used in the study were synthesized by Dan Lu and Sha Zhu in Matthieu Sollogoub's laboratory at UPMC, Paris, France and Pearl Fernandes, Marija Petricevic, Dr Zalihe Hakki and Dr Gaetano Speciale in Spencer J. Williams' laboratory at University of Melbourne, Parkville, Australia. Spencer Williams' laboratory provided α-Man-1,3-noeuromycin (ManNOE), α-Man-1,3-D-glucal (Man-D-glucal), α-Man-1,3-dideoxymannose (ManddMan), α-Man-1,3-2-aminodeoxynojimycin (ManADMJ), α-Man-1,3-mannoimidazole (ManManIm) and α-Man-1,3-α-Man-1,2-α-Man-1,2-α-1-methyl-Man (Man_4OMe). For structures, see Figure 2.4**A**. Matthieu Sollogoub's laboratory provided α-Glc-1,3-α/β-epoxide (Glc-α-epoxide/Glc-β-epoxide), α-Glc-1,3-α/β-aziridine (Glc-α-aziridine/Glc-β-aziridine) and α-Glc-1,3-cyclohexene (Glc-cyclohexene), depicted in Figure 2.4**B**. α-1,2-Mannobiose and α-1,3-mannobiose were purchased from Sigma.

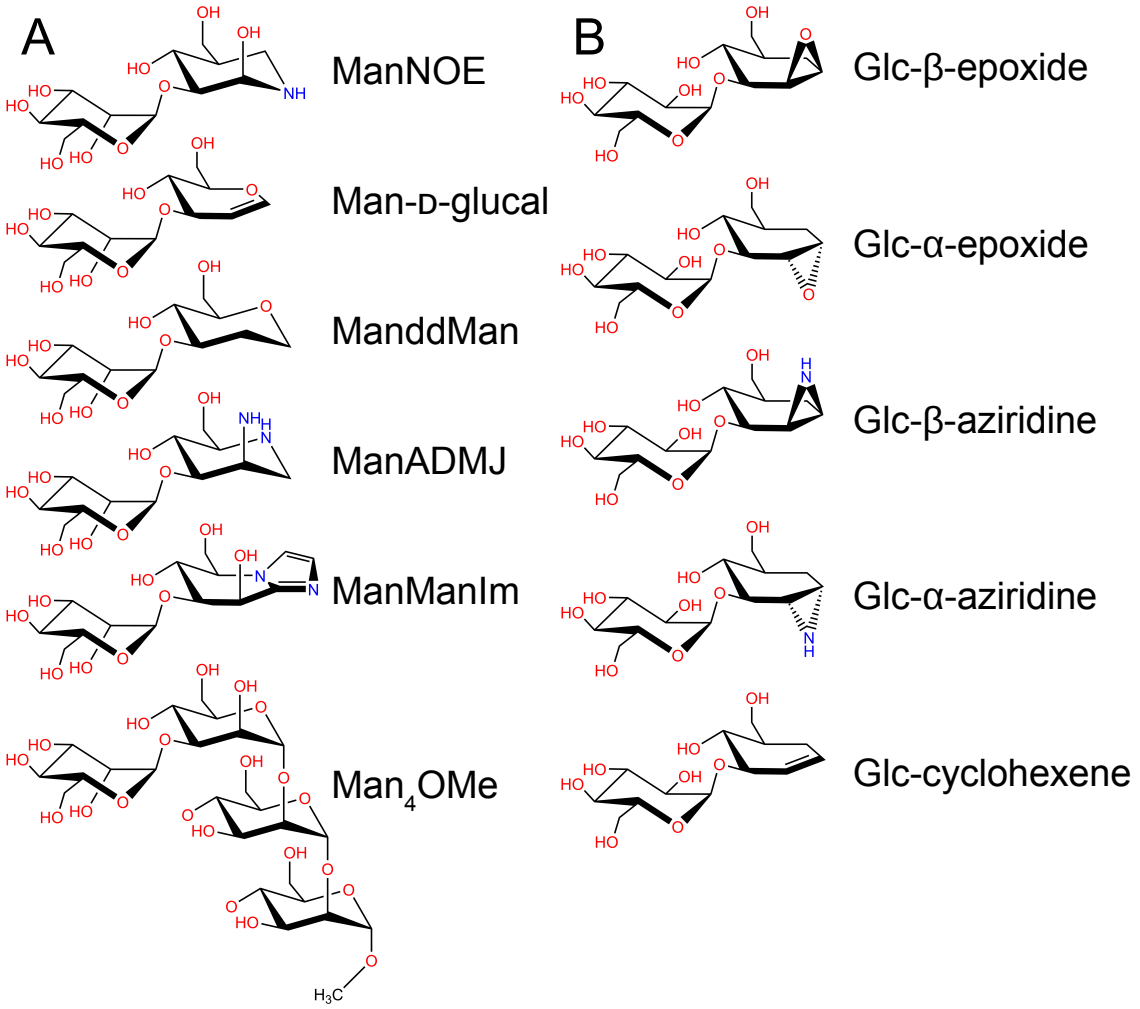


Figure 2.4: Compounds used in the study. (**A**) Chemicals synthesized in Spencer J. Williams' lab: ManNOE, Man-D-Glucal, Man-ddMan, ManADMJ, ManManIm, Man_4OMe. (**B**) Chemicals synthesized in Matthieu Sollogoub's lab: Glc-β-epoxide, Glc-α-epoxide, Glc-β-aziridine, Glc-α-aziridine, Glc-cyclohexene.

Experiments which required quantification of small molecules using mass spectrometry, which involved the use of the five compounds from Matthieu Sollogoub's laboratory, were conducted by me on a visit to UPMC, Paris, France. The conditions for each reaction were: 37 °C in 1.25 mM HEPES pH 7.0, 5 mM NaCl and the total volume varied between 20 and 40 µM. The compounds were incubated in separate reaction tubes with 1 µM of WT *Bt*GH99, WT *Bx*GH99, *Bx*GH99 E333Q, *Bx*GH99 E336Q; 20 µM of WT *Bt*GH99; or buffer only. All five compounds were tested. Each reaction was followed by taking samples after 20 min, 60 min and overnight (after 17–22 h). Each reaction was quenched by mixing 10 µl reaction mixture with 100 µl methanol. Mass spectra were obtained using a Bruker microTOF spectrometer using electrospray ionization (ESI). All compounds were used at a starting concentration of 1 mM with correction factors from calculations applied later. The peaks corresponding to the $[M+Na]^+$ ions

were integrated in Origin software[260]. The concentration of the compounds was estimated as follows: the areas of the peaks corresponding to unreacted epoxide or aziridine and their ring-opened forms in each mass spectrum, were added and assumed to sum up to 1 mM. The proportion of the area of each peak was divided by the total area and this was converted into a concentration. In the case of the Glc-β-epoxide, extended digestion with enzyme revealed that a portion of the substrate peak was not processed by the enzyme, and were most likely traces of Glc-α-epoxide. This amount was subtracted from the total calculated Glc-β-epoxide concentration. Glc-β-epoxide and Glc-β-aziridine were not stable in the buffer used. This can be seen in the reaction graph the case of Glc-β-aziridine, which was the least stable and about 15% of the aziridine ring was hydrolyzed in 1 h.

NMR of open Glc-β-epoxide was performed by Sha Zhu in D_2O using a 600 MHz Bruker Avance magnet.

2.3.3 Crystallization, crystal handling and structure solution

*Bx*GH99 crystals were grown in droplets containing a 1:1 ratio (1 µl each) of protein (34 mg ml^{-1} in 25 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT) to reservoir solution (3 M sodium acetate pH 6.4–7.4). The droplets were hanging on a siliconized cover slip above 500 µl of the precipitant at 19 °C in darkness. The 24-well plates used for crystallization were obtained from Greiner Bio-one. The contact surface between the cover slip and the plate was sealed using high vacuum grease. Crystals were fished from the droplets using nylon cryoloops. Crystals were soaked with the ligand at various concentrations, ranging from 1 to ~40 mM. Cryoprotection briefly submerging the crystal in the reservoir solution with ethylene glycol added up to 20% was applied when fishing the initial crystals soaked with ManddMan and Man-D-glucal, but was omitted for subsequent crystals due to no obvious benefits. Data were collected remotely at beamlines I02, I04 and I04-1 at Diamond Light Source using a wavelength of 0.979 Å.

Diffraction was integrated using Xia2[261] or DIALS[262] and scaled in Aimless.[263] The resolution of the data was cut according to guidelines from Ref. 263, with *CC(1/2)* being the most important factor (the cut-off was 0.3). The model used for direct refine-

ment with each dataset listed in Table 2.2 was *Bx*GH99 with ManddMan with water molecules and the ligand removed from the pdb file (PDB code **5M17**). The details creating the initial model are in Section 3.3.2, where the *Bx*GH99 + ManddMan structure is discussed in context of other inhibitors. Refinement was done in Refmac5[264] and model modifications were made in Coot.[265] Waters were added to the models when the protein chain was refined as well as possible in real space. The model and its agreement with electron density was validated using Coot tools and the wwPDB validation software. CCP4i2 was used for the majority of the process.[266]

2.4 Results and discussion

2.4.1 Structural snapshots of the endomannanase mechanism and the conformation of sugars in the –1 subsite

In order to adequately characterize the GH99 mechanism, a number of *Bx*GH99 complexes with different ligands were solved. An attempt at placing the structures on the conformational itinerary was then made. A $^{4}\boldsymbol{C_1}\rightarrow[^{4}\boldsymbol{E}]^{\ddagger}\rightarrow{}^{4}\boldsymbol{H_5}$ itinerary was proposed for the enzyme in a 2014 review[267] based on the structural analysis of the available complex structures, in all of which the –1 subsite ligand was seen in a $^{4}\boldsymbol{C_1}$ conformation, including the Michaelis-like complex with ManManMUF. In the current work, the structure of a more biologically relevant Michaelis complex (*Bx*GH99 E333Q with Man_4OMe, Figure 2.5**A**) is reported, as well as the product complex with both α-1,3- and α-1,2-mannobiose (Figure 2.5**C**). Structures with reaction intermediate mimics were solved as well: *Bx*GH99 E333Q with Glc-β-epoxide (as shown in Figure 2.9) and WT with Glc-β-aziridine (Figure 2.5**B**). A summary of the conformations of various compounds in complex with *Bx*GH99 is presented in Table 2.1. The significance of these results will be presented in the discussion of each structure.

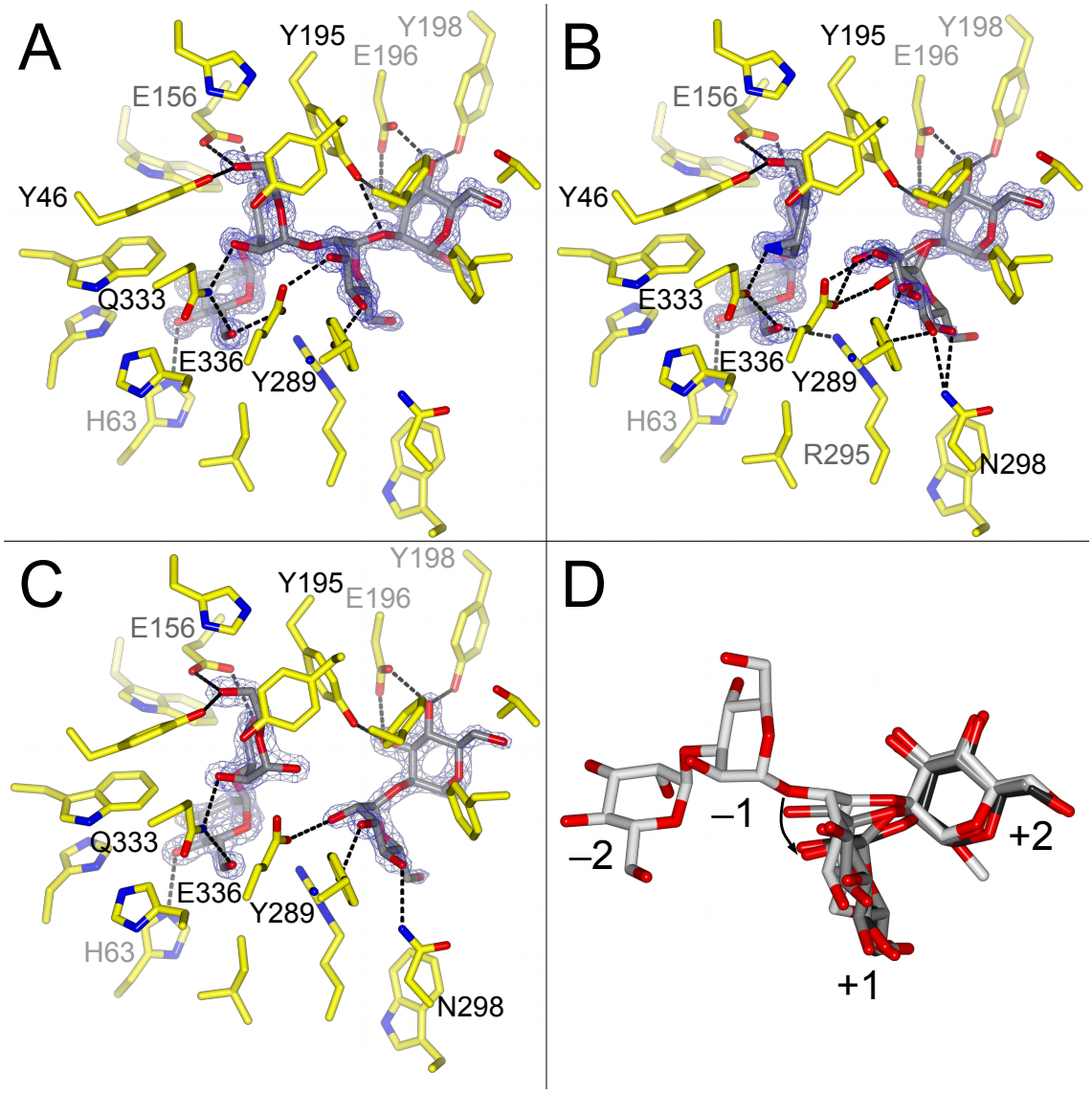


Figure 2.5: Structural snapshots along the endomannanase reaction coordinate with residues surrounding the ligand shown. All electron density maps are 2mF_o-DF_c syntheses contoured at 0.8 $e^-/Å^3$ (**A**) The *Bx*GH99 E333Q Michaelis complex with Man_4OMe. (**B**) Complex of WT *Bx*GH99 with a mimic of the reaction intermediate Glc-β-aziridine. Note the double conformation of the +1 sugar leading to multiple possible hydrogen bonds with residues 289, 298 and 336. (**C**) E333Q product complex with α-1,2-mannobiose and α-1,3-mannobiose. (**D**) Movement of the +1 mannose moiety as the reaction progresses (models – white: **A**, grey: **B**, dark grey: **C**). Assembled in ccp4mg.[204]

Table 2.1: Conformations of different ligands as determined from refinement to electron density inside the *Bx*GH99 active site. For views of ManNOE, Man-D-glucal, ManddMan, ManADMJ, ManManIm and Glc-α-aziridine, see Section 3.4.2.

–1 sugar conformation	Sugar or sugar analog
$^4\boldsymbol{C_1}$	ManNOE, ManddMan, ManADMJ, Man_4OMe, open Glc-β-epoxide, GlcDMJ, ManDMJ, ManManMUF, α-1,3-mannobiose
$^4\boldsymbol{H_5}$	Man-D-glucal, Glc-α-aziridine, Glc-cyclohexene
$^4\boldsymbol{E}$	Glc-β-epoxide, Glc-β-aziridine
$^2\boldsymbol{H_3}$ / $\boldsymbol{E_3}$	ManManIm

The $^4\boldsymbol{H_5}$ conformation for the anhydrosugar intermediate was proposed due to this half-chair being the apparent lowest energy conformation of the –1 anhydrosugar (as calculated using Marvin).[268] $^4\boldsymbol{E}$ envelope was deduced as the conformation of the TS from the principle of least nuclear motion. Transitioning from a $^4\boldsymbol{C_1}$ to $^4\boldsymbol{H_5}$ requires a larger conformational change than $^4\boldsymbol{C_1}$ to $^4\boldsymbol{E}$. These crystallographic data suggest a more simple itinerary of $^4\boldsymbol{C_1}\rightarrow{}^4\boldsymbol{E}$. Both Glc-β-epoxide and Glc-β-aziridine were found to bind to the enzyme in a $^4\boldsymbol{E}$ conformation, and the theoretical lowest energy conformation for them is also $^4\boldsymbol{H_5}$. Another possible explanation is that the compounds act as mimics of the proposed TS, and the reaction intermediate is a $^4\boldsymbol{H_5}$ half-chair. This seems less likely, as both compounds closely resemble the reaction intermediate. A third possibility is that the difference between $^4\boldsymbol{E}$ and $^4\boldsymbol{H_5}$ is not large in terms of the necessary conformational change, and it may be superfluous to explain reaction dynamics by it. In the QM/MM calculations which are presented later in this chapter, conformation $^4\boldsymbol{H_5}$ does not appear. It was observed it in *Bx*GH99 structures in complex with Man-D-glucal, Glc-α-aziridine and Glc-cyclohexene. The calculated lowest energy conformation for these compounds is also $^4\boldsymbol{H_5}$, but $^4\boldsymbol{E}$ is also energetically accessible.[269]

The Michaelis complex (with Man_4OMe, Figure 2.5**A**, Figure 2.6) shows that *Bx*GH99 is an *anti* protonating enzyme. The $^{+1}$C2-$^{-1}$O1-C1-O5 dihedral angle is 63°, which serves to position the +1 sugar on the opposite side to the acid/base E336. The $^{-1}$O5-C1-O1-^{E336}Oε angle is 172°, close to 180°, which is optimal for *anti* protonating enzymes. The $^{4}\boldsymbol{C}_{1}$ chair of the –1 sugar is slightly distorted: its θ and ϕ angles are, respectively, 19.7° and 120.1°. If the θ angle was larger than 22.5° its conformation would be closer to $^{2}\boldsymbol{E}$ than $^{4}\boldsymbol{C}_{1}$. Such a conformation is caused by strong binding of the +1/+2 sugars which may serve to kink the mean sugar plane "up" with respect to E336 Oε, positioning the glutamate oxygen slightly below the plane. This makes the endomannanase an *anti*-B (below) enzyme according to nomenclature by Nerinckx.[270] The *anti* position of the acid/base was known since the first structural study of GH99,[148] but the Michaelis complex was essential to determine whether it acts from above or below the plane. Carbohydrates in other subsites are all relaxed $^{4}\boldsymbol{C}_{1}$ chairs. The hydroxyl group of Tyr252 is poised to interact with the –1 sugar ring oxygen, but in the Michaelis complex the atom is pulled away from it, making the interatomic distance 3.35 Å. For complexes with Man-D-glucal, ManddMan and α-1,3-mannobiose this distance is 2.9–3 Å, suggesting the interaction might be important in the later stages of the reaction.

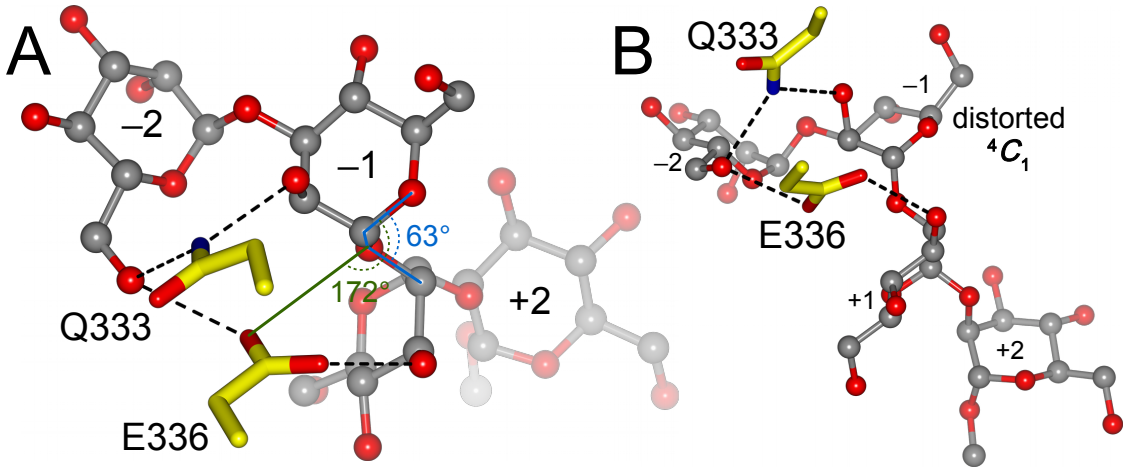


Figure 2.6: Geometry of the *Bx*GH99 E333Q complex with Man_4OMe with respect to the –1 sugar and with the catalytic residues shown. (**A**) View from the top with dihedral angles of interest marked. (**B**) View from the front, showing the distortion of the $^{4}\boldsymbol{C}_{1}$ conformation. Assembled in ccp4mg.[204]

Complexes mimicking the reaction intermediate can offer an intimate look into reaction dynamics. Perhaps the most relevant for the current analysis is the tertiary complex of *Bx*GH99 with Glc-β-aziridine and α-1,2-mannobiose, which was solved at a near-atomic resolution of 1.04 Å (Figure 2.5**B**, Figure 2.7**A** for the –2/–1 subsites). E333 clearly stands out as an important residue, making a strong 2.58 Å hydrogen bond with the aziridine nitrogen but leaving the three-membered ring intact (Figure 2.7**A**). The Cδ…Oε1 linkage (see Figure 2.7**A** for nomenclature) is 0.04 Å shorter than Cδ…Oε2, and the electron density suggests more of a double-bond like character for it. Oε1 is an acceptor of a hydrogen from the H335 backbone nitrogen (distance 2.84 Å) and, more weakly, from the H336 backbone nitrogen (3.02 Å) (Figure 2.7**A**). Such a hydrogen bonding pattern causes the E333 side chain to be an uncommon rotamer: Coot suggests the probability of such a rotamer is 0, but this is the conformation is observed in all *Bx*GH99 structures. The E333 Oε2…C1 distance is 3.49 Å, much longer than required for a nucleophilic attack (2.6–2.8 Å). When mutated to an asparagine, Q333 Oε1 assumes a position opposite the –1 ligand (Figure 2.5) and Q333 Nε1 faces the ligand. Thus, there is strong preference for the mutant to replicate the hydrogen bonding pattern with the polypeptide backbone and this requires the the oxygen atom does not face the solvent or the ligand. Taken together, this supports the hypothesis that the role of the residue in the reaction is to be a crucial acid/base.

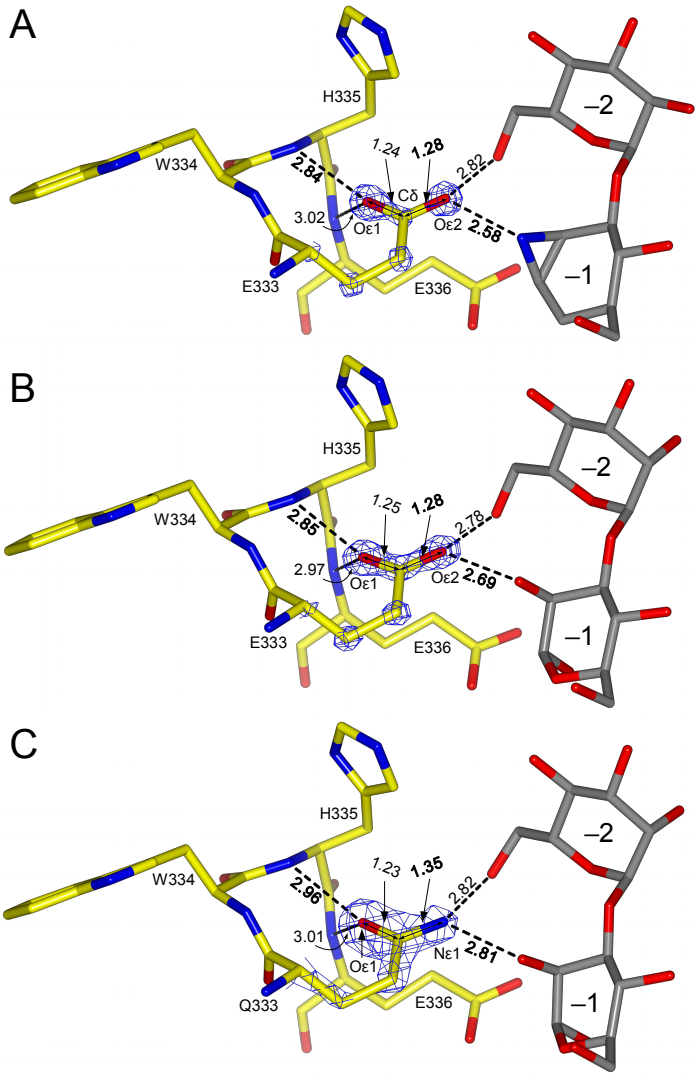


Figure 2.7: The environment around the side chain of the *Bx*GH99 catalytic residue 333. Interatomic distances that vary significantly between different structures are in bold (all distances in Ångstroms). α-1,2-Mannobiose in the +1/+2 subsites hidden for clarity. (**A**) The complex of WT *Bx*GH99 with Glc-β-aziridine and α-1,2-mannobiose. 2mF_o-DF_c synthesis contoured at 2 e$^-$/Å^3. (**B**) WT *Bx*GH99 product complex (with α-1,2-mannobiose and α-1,3-mannobiose). 2mF_o-DF_c synthesis contoured at 2 e$^-$/Å^3. (**C**) *Bx*GH99 E333Q product complex (with α-1,2-mannobiose and α-1,3-mannobiose). 2mF_o-DF_c synthesis contoured at 1 e$^-$/Å^3. Assembled in ccp4mg.[204]

The product complexes, *Bx*GH99 with α-1,3-mannobiose and with or without α-1,2-mannobiose, were solved for the WT and E333Q mutant. A minor, but consistent difference between the binding of the –1 sugar by these variants exists: the WT E333 Oε2…OH2 distance is 2.66–2.69 Å and the mutant Nε1…OH2 distance is 2.79–2.81 Å (Figure 2.7**BC**). A similar bonding pattern is replicated in the structure of the Michaelis complex, where the Nε1…OH2 distance is 2.79 Å. For the complexes with intermediate mimics, it also seems to hold true: the hydrogen bond is 2.58 Å long in the WT structure containing Glc-β-aziridine and 2.78 Å long for E333Q with Glc-β-epoxide. In WT and E336Q structures with open Glc-β-epoxide ("product mimic complexes") this distance is, accordingly, 2.66 and 2.58 Å. These measurements suggest this distance in the WT enzyme-intermediate complex is 2.6–2.7 Å. For consistency, in Figure 2.5 the Michaelis complex was compared directly with the E333Q product complex. The crucial difference between the complex with the substrate and with the products is the positioning of the +1 sugar with respect to the –1 sugar. After the bond is cleaved, the +1 sugar moves away from the –1 sugar O1 while still being transiently bound to the enzyme. That movement can be seen when saccharide structures along the reaction coordinate are directly compared, as in Figure 2.5**D**.

One residue, Y198, was consistently observed in structures with α-1,3-mannobiose to assume a single conformation and making a hydrogen bond with the OH4 group of the +2 mannose, and in two conformations when this subsite is not occupied. Perhaps Y198 is responsible for dragging the sugar out of the binding pocket after the reaction occurs. The OH1 group of the +2 sugar projects outside of the binding pocket into the solvent, as this is where the rest of the yeast mannan or N-linked glycan would be located. Due to a higher resolution (1.08 Å) of data, the two conformations of the +1 mannose can be seen for the WT product complex, and only one in the lower-resolution structure with E333Q (1.65 Å). The data collection and model refinement statistics for the structures presented in this chapter can be found in Table 2.2.

Table 2.2: Data collection and model refinement statistics of *Bx*GH99 structures discussed in this chapter. *Values in parentheses represent the highest resolution shell. RMS – root mean square.

	E333Q Michaelis complex	WT + β-1,2-aziridine	WT + β-1,2-aziridine + α-1,2-mannobiose	E333Q + β-1,2-epoxide	WT + open β-1,2-epoxide	E336Q + open β-1,2-epoxide	WT product complex	E333Q product complex
Data collection								
Beamline	I02	I04-1	I02	I04	I04	I04	I04-1	I04-1
Space group	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4
Cell dimensions								
a = *b*, *c* (Å)	107.9, 67.6	108.5, 68.0	108.6, 67.6	108.2, 67.4	108.3, 67.5	107.6, 67.5	108.9, 68.5	108.7, 68.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	57.30-1.07 (1.09-1.07)*	39.49-1.25 (1.27-1.25)	57.42-1.04 (1.06-1.04)	76.53-1.12 (1.14-1.12)	57.25-1.28 (1.30-1.28)	57.17-1.34 (1.36-1.34)	58.00-1.08 (1.10-1.08)	76.83-1.65 (1.68-1.65)
R_{merge}	0.061 (1.531)	0.059 (1.664)	0.075 (1.346)	0.063 (1.387)	0.066 (1.445)	0.069 (1.278)	0.059 (1.455)	0.084 (1.014)
R_{pim}	0.026 (0.705)	0.022 (0.643)	0.030 (0.601)	0.027 (0.634)	0.028 (0.650)	0.029 (0.544)	0.025 (0.710)	0.047 (0.578)
CC(1/2)	0.997 (0.398)	0.998 (0.462)	0.993 (0.455)	0.999 (0.450)	0.999 (0.439)	0.999 (0.563)	0.999 (0.407)	0.995 (0.406)
$\langle I / \sigma I \rangle$	10.6 (1.0)	17.4 (1.3)	9.5 (0.8)	9.8 (1.1)	10.7 (1.2)	10.6 (1.3)	10.1 (1.0)	8.2 (1.1)
Completeness (%)	100 (99.3)	99.4 (100)	98.8 (80.4)	100 (99.2)	100 (99.9)	99.9 (99.5)	99.9 (99.2)	99.6 (99.8)
Redundancy	6.3 (5.5)	8.2 (7.6)	7.5 (5.5)	6.4 (5.7)	6.4 (5.9)	6.5 (6.3)	6.3 (5.1)	4.0 (4.1)
Refinement								
Resolution (Å)	57.30-1.07	39.49-1.25	57.42-1.04	76.53-1.12	57.25-1.28	57.17-1.34	58.00-1.08	76.83-1.65
No. reflections all / free	170019 / 8420	108087 / 5400	185423 / 9152	149020 / 7398	100095 / 4977	86250 / 4416	170757 / 8460	47947 / 2376
R_{work} / R_{free}	0.12 / 0.14	0.13 / 0.15	0.13 / 0.15	0.13 / 0.15	0.14 / 0.16	0.13 / 0.16	0.13 / 0.14	0.15 / 0.19
No. atoms								
Protein	3067	3045	3150	3165	3129	3034	3080	2980
Ligand/ion	54	38	60	30	27	31	81	47
Water	344	270	382	416	381	278	277	161
B-factors (Å^2)								
Protein	16	17	16	17	18	21	16	28
Ligand/ion	17	22	16	17	23	25	18	26
Water	35	33	33	33	33	36	33	40
RMS. deviations								
Bond lengths (Å)	0.010	0.009	0.013	0.011	0.009	0.011	0.013	0.014
Bond angles (°)	1.5	1.4	1.8	1.6	1.4	1.5	1.4	1.4
PDB code	**6FWG**	**6FWI**	**6FWJ**	**6FWL**	**6FWM**	**6FWO**	**6FWP**	**6FWQ**

2.4.2 Activity of GH99 towards novel substrates

The activity of *Bt*GH99 and *Bx*GH99 WT, E333Q and E336Q against compounds shown in Figure 2.4**B** was tested. During the course of the experiments, compounds Glc-β-epoxide and Glc-β-aziridine decomposed in the aqueous buffer solution used, which was controlled for by performing simultaneous incubations without the enzyme. Activity was observed only for WT *Bt* and *Bx* enzyme, and only for Glc-β-epoxide (Figure 2.8**D**). The buffer-induced decomposition of Glc-β-aziridine proceeded at a faster rate than that of Glc-β-epoxide (compare the controls in Figure 2.8 **C** and **D**).

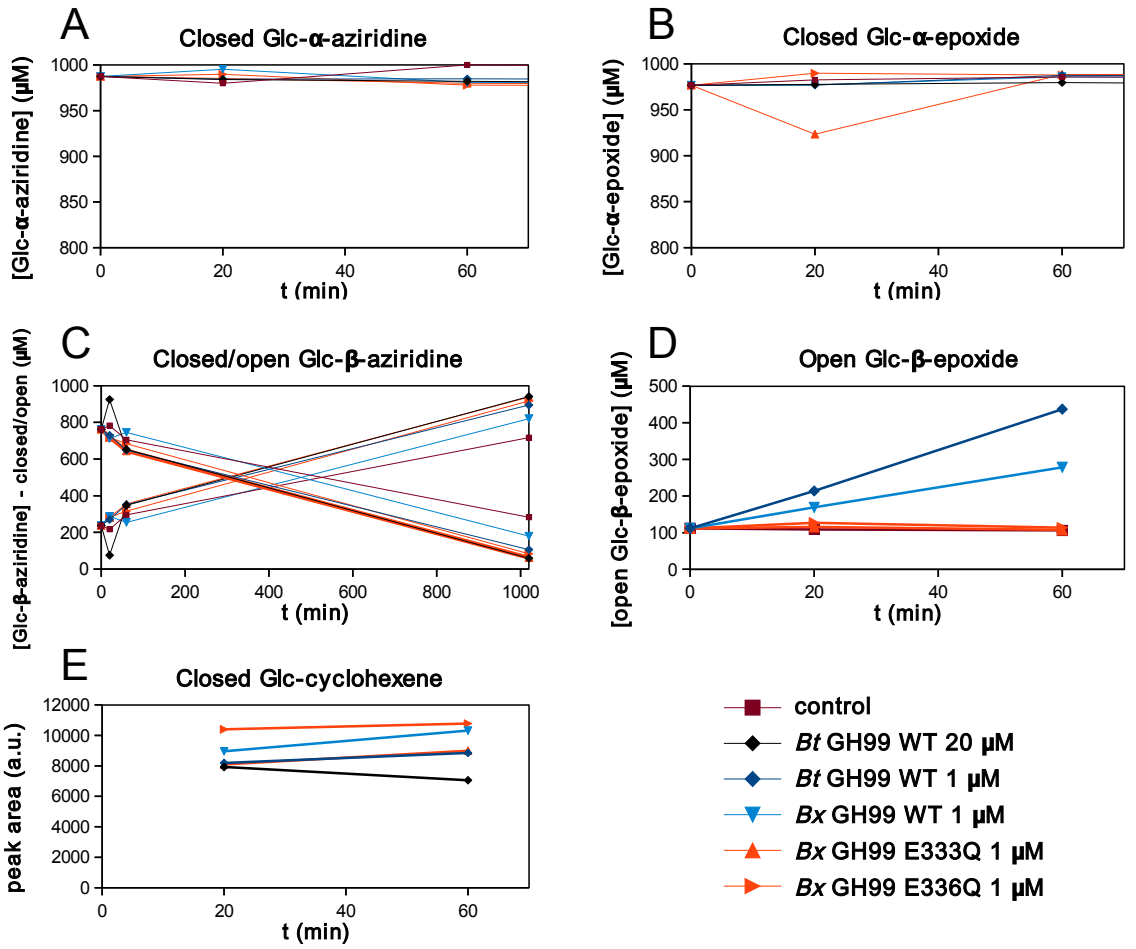


Figure 2.8: MS (Mass spectrometry) quantification of the amount of closed or open potential substrates of GH99 present after incubation with or without the enzyme. Overnight measurements omitted in all but **C** in order to preserve graph clarity. (**A**) Concentration of closed Glc-α-aziridine over time. Note the Y scale starts at 800 μM. (**B**) Concentration of closed Glc-α-epoxide over time. Note the Y scale starts at 800 μM. (**C**) Concentration of closed (descending lines) and open (ascending lines) Glc-β-aziridine over time. (**D**) Concentration of open Glc-β-epoxide over time. (**E**) Area of the MS peak corresponding to closed Glc-cyclohexene over time.

The results shown in Figure 2.8**D** were corroborated by the crystal structures obtained by soaking crystals of *Bx*GH99 WT, E333Q and E336Q with Glc-β-epoxide. In the structures of WT and the E336Q mutant the epoxide ring is *open* and the pseudo-sugar occupying the –1 subsite resembles α-D-mannopyranose. When bound to the E333Q mutant, the epoxide ring remains *closed*, signifying that E333 is crucial to the activity of the enzyme. The ligand in the WT structure was best explained by setting the atom occupancy at 75%, which is likely due to its active turnover by the enzyme. The β-epoxide ring is also seen as open in the structure of E336Q mutant, apparently contradicting the MS kinetic data. *Bx*GH99 E336Q is roughly equivalent to *Bt*GH99 E332A which was shown previously[148] to have residual activity that was about 50-fold lower than the WT. The MS experiment might have not been sensitive enough to detect the small amounts of the open epoxide generated by the *Bx*GH99 E336Q variant. Glc-β-aziridine was observed bound in the active site of *Bx*GH99 WT, as shown earlier in Figure 2.5. The β-aziridine ring remained closed, suggesting the enzyme is not active towards compounds with this ring, however, this cannot be asserted with certainty when looking at the results presented in Figure 2.8**C** due to the fast decomposition of the Glc-β-aziridine. This means that the endomannanase is not tuned for opening the β-aziridine ring. What is also clear is that the compounds with an α-configured three-membered ring are not substrates for the enzyme, which provides indirect evidence supporting the hypothesis about the GH99 mechanism shown in Figure 2.1. It might be argued that the α-configured compounds are unstable and likely to open on their own, but these results suggest otherwise.

The *open* Glc-β-epoxide (α-Glc-1,3-α-*carba*mannose) cannot undergo mutarotation due to possessing a cyclohexane ring, rather than the pyranose ring: mutarotation necessitates that the ring can be transiently opened. After incubation of a larger amount of the compound with *Bt*GH99, an NMR experiment confirmed that the product is indeed α-configured. Figure 2.9 shows the reaction catalyzed by GH99. This result is a strong indication that GH99 can open the epoxide ring only when it resembles the proposed reaction intermediate.

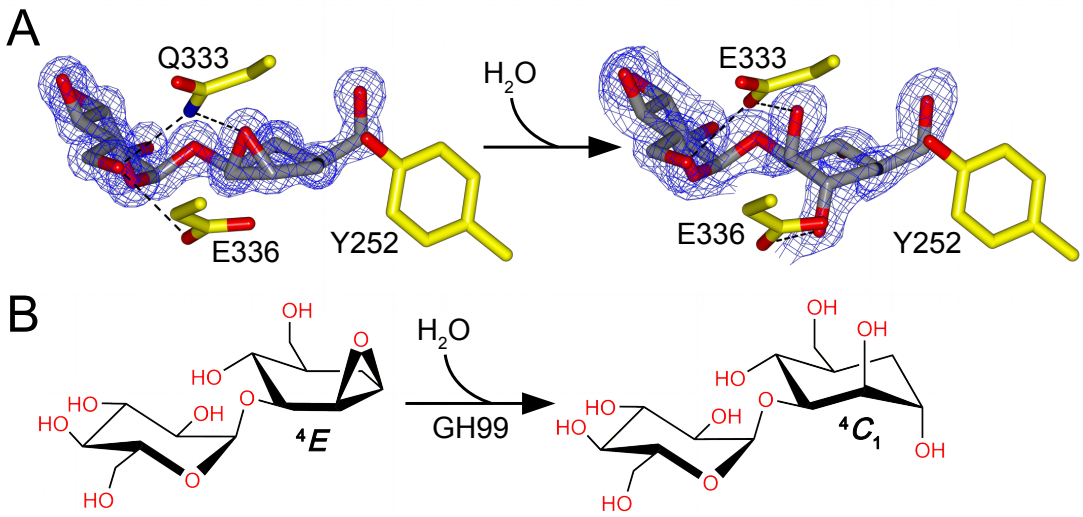


Figure 2.9: Catalysis of ring opening of Glc-β-epoxide by *Bx*GH99. (**A**) Structures of the *closed* Glc-β-epoxide bound in the active site of *Bx*GH99 E333Q variant (left, $2mF_o$-DF_c synthesis contoured at 0.8 $e^-/Å^3$) and *open* Glc-β-epoxide in the active site of WT *Bx*GH99 (right, $2mF_o$-DF_c synthesis contoured at 0.5 $e^-/Å^3$). Hydrogen bonds between the catalytic residues and the sugars are shown. (**B**) Chemical reaction equation.

2.4.3 Quantum mechanics/molecular mechanics

Important corroborating data were generated by our collaborator Professor Carme Rovira and Dr Lluis Raich. QM/MM metadynamics simulations on the Michaelis and intermediate complexes that aimed at replicating the reaction as it happens revealed the likely conformational itinerary is ${}^{2}\boldsymbol{H}_3 \rightarrow [\boldsymbol{E}_3]^{\ddagger} \rightarrow {}^{4}\boldsymbol{E} \rightarrow [\boldsymbol{E}_3]^{\ddagger} \rightarrow {}^{4}\boldsymbol{C}_1$ (see Figure 2.10). The mannose residue in the –1 subsite of the Michaelis complex (*Bx*GH9 E333Q with Man_4OMe) was a ${}^{2}\boldsymbol{E}$-slanted ${}^{4}\boldsymbol{C}_1$, not the ${}^{2}\boldsymbol{H}_3$ conformer. It might be that due to the mutation at residue 333, the enzyme was incapable of fully deforming the substrate in order to catalyze the reaction, which would require that the substrate assumes a ${}^{2}\boldsymbol{H}_3$ conformation. Possibly the difference between the length of the E333-substrate O2 hydrogen bond (discussed earlier in Section 2.4.1) is responsible for this change. A fortuitous finding of a weakly binding inhibitor whose –1 pseudosugar (ManIm) bound as ${}^{2}\boldsymbol{H}_3$ / $\boldsymbol{E}_3$ signifies the enzyme can accommodate such conformers, however, the binding of ManManIm was undetectable (see Figure 3.7 on page 81).

With the data presented so far at hand, it is possible to form a more informed guess about the position of the catalytic water. The analysis is presented in Figure 2.11. Comparison of structures with the intermediate mimic Glc-β-epoxide narrows the search down to two water molecules that are close to the anomeric carbon and below

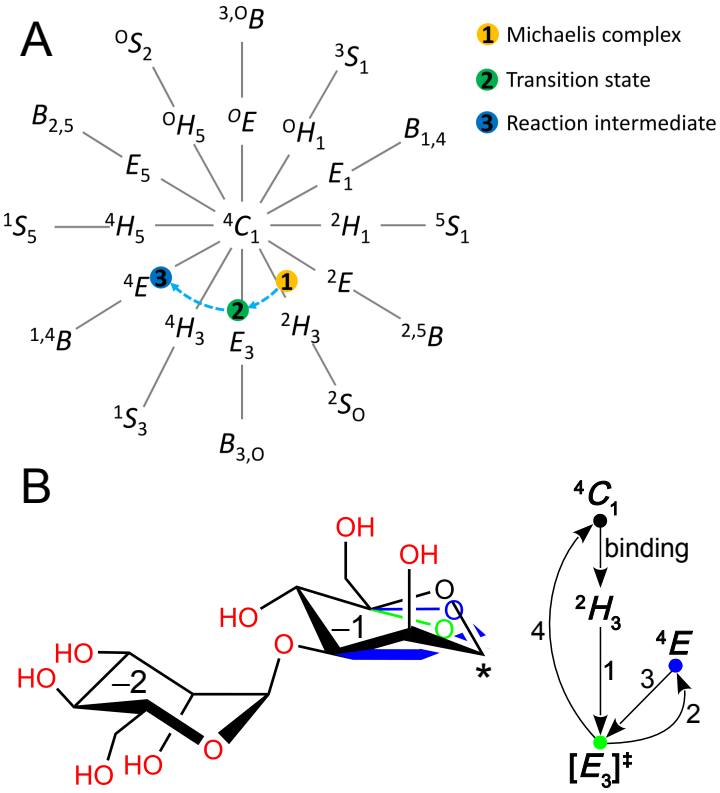


Figure 2.10: Schematic representations of the conformational changes during catalysis of reaction by GH99 *endo*-α-mannanase. (**A**) Stoddart projection of the first part of the reaction. (**B**) Structural changes during the whole reaction. The asterisk points to the anomeric carbon.

the –1 sugar plane (Figure 2.11**A**). In the structure of WT enzyme with Glc-β-epoxide, the electron density for the O1 was found to be elongated (Figure 2.11**B**). This was accurately explained by adding a water molecule (water 1) in an alternative (B) conformation at an occupancy of 0.25 (open Glc-β-epoxide occupancy 0.75, conformation A). The distance between the refined water 1 position and the carbasugar anomeric carbon is only 2.9 Å and the water is positioned below the projected plane of the $\boldsymbol{E}_3$ transition state, ideal for a nucleophilic S_N2 attack (Figure 2.11**D**). The equivalent water molecule in the structure of Glc-β-epoxide with *Bx*GH99 E333Q mutant is 3.9 Å away from the anomeric carbon of the modified anhydrosugar ring of Glc-β-epoxide (Figure 2.11**A**). However, there is no indication of either of these water molecules in the structure of Glc-β-epoxide with the slow E336Q mutant (Figure 2.11**C**). Instead, even after modelling in the opened intermediate, residual positive difference electron density remains in the space around O1 (Figure 2.11**C**). Thus, this result might mean that both are important in hydrolyzing the reaction intermediate.

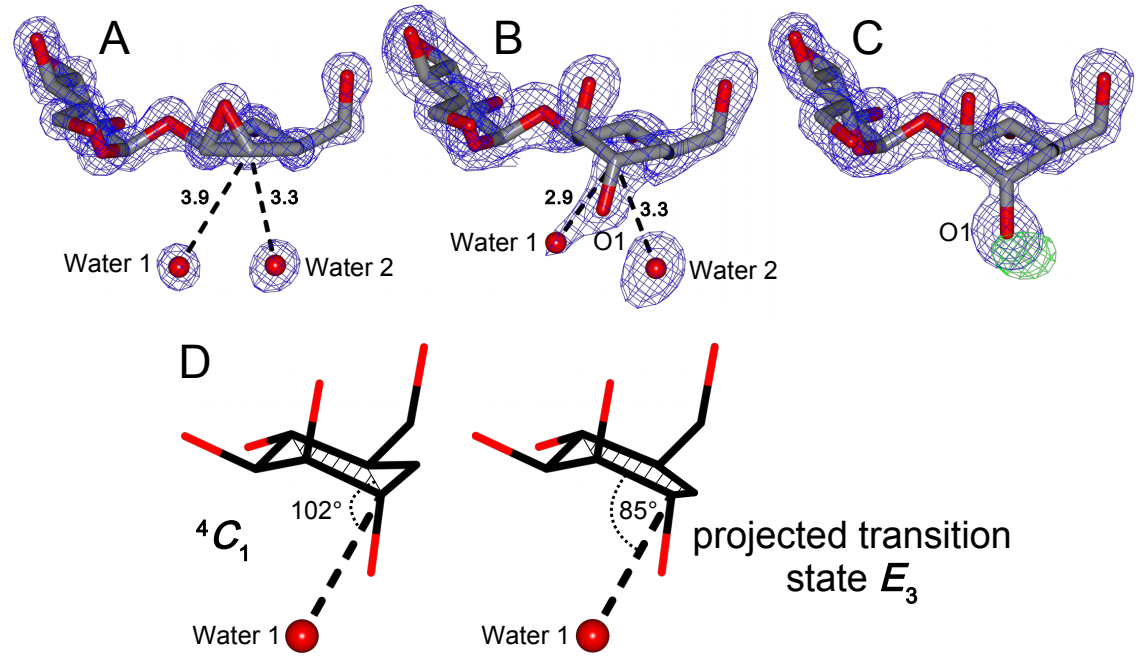


Figure 2.11: Differences between the water molecules around the scissile bond position in intermediate mimic structures. Distances to water molecules nearest to the anomeric carbon are shown as dashed lines and labelled with length in Å. (**A**) Glc-β-epoxide (occupancy of 1) with the inactive E333Q mutant. 2mF_o-DF_c synthesis contoured at 1 $e^-/Å^3$. (**B**) Glc-β-epoxide (occupancy of 0.75) with WT *Bx*GH99. 2mF_o-DF_c synthesis contoured at 0.3 $e^-/Å^3$. (**C**) Glc-β-epoxide (occupancy of 1) with the slow *Bx*GH99 mutant E336Q. 2mF_o-DF_c synthesis contoured at 0.5 $e^-/Å^3$ and positive mF_o-DF_c around the O1 atom at 0.25 $e^-/Å^3$. (**D**) Left: a view of the carbammannose from the structure of Glc-β-epoxide with WT *Bx*GH99 (the chair conformation plane highlighted). Right: the transition state E_3 conformation as projected from the carbamannose. Note that both of the angles shown are close to 90°.

2.5 Conclusions

The structural and kinetic data presented above support the hypothesized neighbouring-group mechanism for bacterial *endo*-α-1,2-mannanases. To date, no other enzymes have been shown to use a 1,2-anhydrosugar "epoxide" intermediate during their catalysis. The synthesis of unprotected 1,2-anhydrosugars was reported recently[271] and the compounds were unstable: 29% of 1,2-anhydroglucose decomposed at -10 °C in 1 hour in 1:1 D_2O/acetonitrile mixture. At room temperature, these compounds are likely to decompose extremely fast. Our result showing that GH99 is active only towards the β- and not α- form of the 1,2-anhydrosugar suggests the enzyme is unique in using this structure to circumvent the steric barriers imposed by the internal α-mannosidic bond of its substrate. Furthermore, the enzyme did not catalyze the hydrolysis of the aziridine ring, but only the epoxide ring.

GH99 endomannanases overcome the unfavourable steric interactions at the axial OH2 group by taking advantage of the presence of these potential clashes. It forms an easily hydrolyzable reaction intermediate, which would decompose on its own in solution, but the enzyme evolved to catalyze this process as well, probably because if it was not broken down, the anhydrosugar would stay bound to the enzyme and inhibit it. It is the first instance of data being consistent with the creation of an epoxide ring during hydrolysis. What is needed as an ultimate proof of the mechanism are measurements of KIEs of substrates synthesized with heavy carbon, hydrogen and oxygen isotopes around the scissile bond. KIEs have recently been used to elucidate the mechanism of hydrolysis of a compound highly structurally related to the GH99 reaction intermediate.[149] Measurements of KIEs for the GH99-catalyzed reaction have been undertaken by the Andrew Bennett's group (see Figure 2.12) but the results are not available yet.

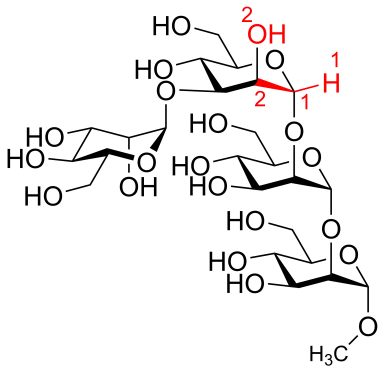


Figure 2.12: Sites of GH99 KIE measurements. Primary isotope effect measurement sites are at C1 and H1, and secondary at C2 and O2.

Knowing the mechanism by which a reaction proceeds is important in designing enzyme inhibitors. Conformationally restricted compounds that mimic the transition state often bind with excellent affinity, outcompeting the substrate in the active site and inhibiting catalysis. However, the electrostatic interactions and hydrogen bonds between GH99 and an inhibitor can only form if a compound has appropriate charge. In the next chapter, data about shape and charge mimicry in inhibiting GH99 will be presented and interpreted.

Chapter 3

Shape and charge of GH99 inhibitors*

3.1 Abstract

The design of enzyme inhibitors is a process which benefits from the knowledge of ligand-protein interactions. One of the most direct methods for investigating these interactions is X-ray crystallography, which enables the observation of interatomic distances and sometimes also the protonation states. This chapter outlines studies on novel GH99 inhibitor designs. Man- or Glc-linked, conformationally restricted and unrestricted –1 subsite warheads were surveyed. Of all of these, the tightest-binding *Bx*GH99 inhibitor was ManNOE. Its affinity to *Bt*GH99 expressed as the dissociation constant K_D was 30 nM, and 13 nM for *Bx*GH99. This was 5–20 × tighter (depending on the protein) than the previously known tightest binder, ManIFG. K_D values of other inhibitors were above 1 µM. 3-dimensional structures of *Bx*GH99 with all of the tested ligands were obtained and helped to explain some surprising findings, namely the lack of measurable binding of Man-mannoimidazole and the poor affinity of Man-aminoDNJ. The results presented in this chapter inform future directions in GH99 inhibition, as well as provide additional data pertaining to the mechanism of this enzyme. Inhibition of endomannosidase is a potential antiviral therapy and the large arsenal of its inhibitors will be useful in possible future medical applications.

*Some of the work described in this chapter has been published in: Petricevic *et al.*, Contribution of Shape and Charge to the Inhibition of a Family GH99 *endo*-α-1,2-Mannanase. *J. Am. Chem. Soc.* **139**, 1089–1097 (2017) and in Fernandes *et al.*, Exploration of Strategies for Mechanism-Based Inhibitor Design for Family GH99 *endo*-α-1,2-mannanases. *Chem. Eur. J.* **24**, 7464–7473 (2018).

3.2 Introduction

Inhibition of endomannosidase can potentially enhance the antiviral activity of glucosidase inhibitors such as *N*B-DNJ. This therapy can also be antiviral on its own. At the time when this project was commenced, the most potent known inhibitor of this enzyme was GlcIFG.[148] ManIFG was subsequently synthesized and determined to be a better inhibitor of the GH99 endomannanase than GlcIFG. As the natural substrate of the bacterial GH99 is yeast mannan, that was an expected result.[159] The obvious approach for creating a GH99 inhibitor is linking Man or Glc with a previously known exomannosidase inhibitor in order to obtain the analog of the –2/–1 ligand. The effects of these compounds on GH99 vary from excellent inhibition (as with GlcDMJ[153] and GlcIFG)[148] to poor or no binding (as with GlcSW, dubbed Fleetamine).[158] Structures of proposed GH99 inhibitors are presented in Figure 3.1**A** and inhibitors whose structure in complex with *Bx*GH99 was solved are presented in Figure 3.1**B**.

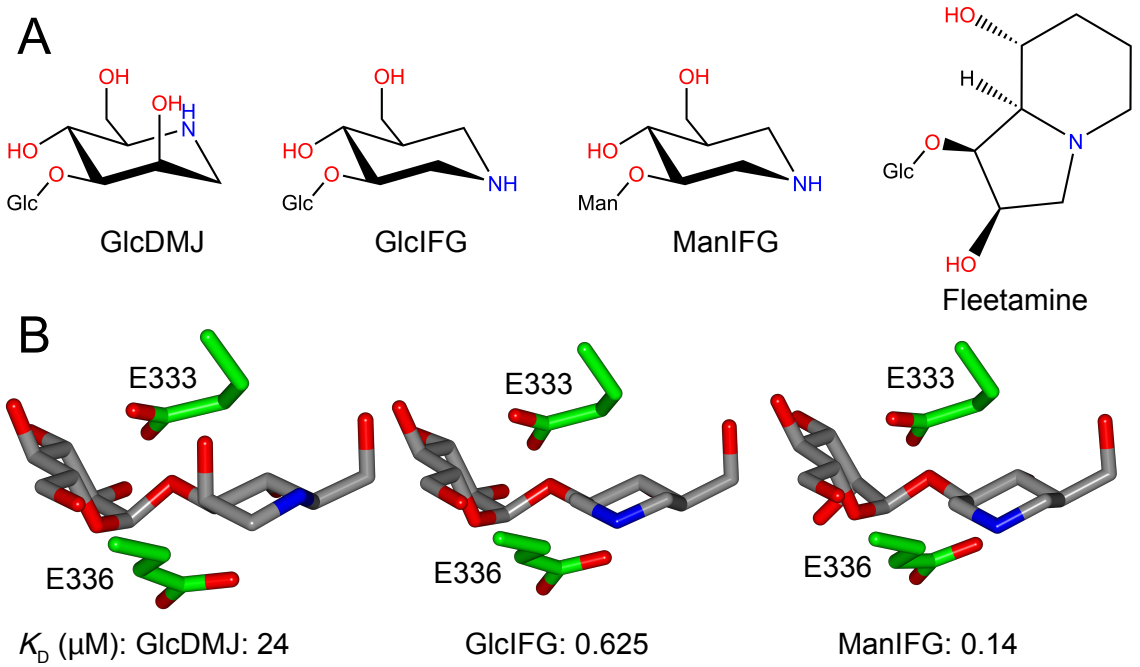


Figure 3.1: Proposed GH99 inhibitors with a history of attempts at characterizing their structure when bound to *Bx*GH99. (**A**) Chemical structures of the compounds. (**B**) Views of the structures of inhibitors bound to *Bx*GH99 and their dissociation constants K_D. Extracted from PDBs (left to right): **4AD3**, **4AD4**, **4V27**. The models were aligned using main chain atoms of residues 333 to 336 in ccp4mg.[204]

Conformationally restricted inhibitors are of interest, as many enzymes depend on deforming the –1 sugar in order to catalyze hydrolysis.[27] This is especially significant for mannosidases, which have to overcome steric clashes due to the presence of the axial OH group at C2.[272] Mannoimidazole and noeuromycin are excellent inhibitors of GH47 α-mannosidases[30] (K_Ds on the GH47 from *Caulobacter* strain K31, accordingly: 47 nM and 99 nM) and moderate inhibitors of GH2 β-mannosidase[273] (K_Ds on *Bt*Man2A: 1.4 μM and approximately 1 μM). Of these two good inhibitors, mannoimidazole mimics the conformation of the transition state: $^{3}\boldsymbol{H}_{4}$ for GH47 and $\boldsymbol{B}_{2,5}$ on GH2, and noeuromycin mimics the carbocation. Kifunensine, a specific GH47 inhibitor is observed with these enzymes in a $^{1}\boldsymbol{C}_{4}$ conformation, which is reminiscent of the product of the reaction but not the transition states.[29] KIF is structurally similar to mannoimidazole but its five-membered ring is not aromatic (see Figure 3.2), which makes the ring pucker – this property allows the pseudopyranose ring sample conformational space different to that of ManIm.

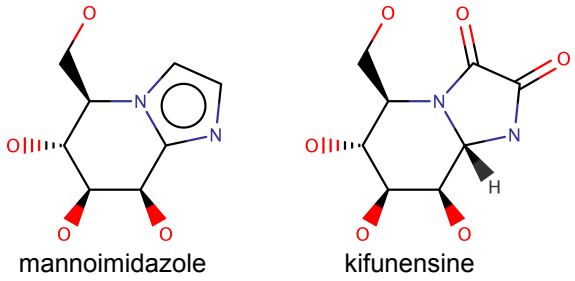


Figure 3.2: Examples of pseudopyranose mannosidase inhibitors with a fused five-membered ring.

In this chapter, novel designs of GH99 inhibitors will be discussed and their binding constants measured. Structures of compounds whose –1 conformation is restrained will be shown in complex with *Bx*GH99. The relative affinity of various types of GH99 inhibitors will be assessed. The influence of inhibitor binding on the solvent molecules observed around the –1 binding site will be examined. Finally, suggestions about possible future directions will be made.

3.3 Materials and methods

3.3.1 Gene expression and protein production

The same stocks of unlabelled *Bx* and *Bt*GH99 as described in Section 2.3.1 were used. ^{15}N-labelled *Bt*GH99 WT and *Bx*GH99 WT, E333Q and E336Q were produced in M9 media with 1 g l^{-1} of $^{15}NH_4Cl$ as the sole source of nitrogen. A procedure for expression and purification identical as with unlabelled proteins expressed in LB media was used. For each purification, a freshly stripped and recharged or brand new 5 ml HisTrap FF column was used. This was done in order to prevent cross-contamination by traces of any unlabelled protein, as well as contamination of each protein variant by a different one. Gel filtration chromatography was performed in 25 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT. In order to perform NMR experiments, the proteins were buffer-exchanged into 50 mM potassium phosphate pH 7.0, 50 mM KCl buffer with no loss of material.

3.3.2 Crystallization and structure solution of *Bx*GH99 complexes with inhibitors

Crystallization conditions were previously described in Section 2.3.3.

Diffraction data reduction were integrated using Xia2[261] and DIALS,[262] and scaled in Aimless.[263] The R_{free}-flagged reflections were kept identical between the isomorphous datasets in order to prevent model bias. For the first, high-resolution structure (*Bx*GH99 with ManddMan) the direct refinement starting model was **4AD4** (PDB code). An improved model obtained after building and refinement of this structure (polypeptide chain only) was used as the starting model in all subsequent datasets from crystals soaked with inhibitors. Reciprocal space refinement was done in Refmac5[264] and model modifications were made in Coot.[265] Dictionary files needed for refinement were made in JLigand.[274] Waters were added to the models when the protein chain was refined as well as possible in real space. The model and its agreement with electron density was validated using Coot tools and wwPDB validation software. CCP4i2 was used for the majority of the process.[266] The

compounds used were presented in Figure 2.4 in Section 2.3.2 (page 51).

3.3.3 NMR experiments with unlabelled and ^{15}N-labelled *Bt* and *Bx*GH99

Initial 1D NMR experiments, which did not yield usable data, were performed in-house with the help of John Darby. The presented NMR was performed by me during a visit to CIC bioGUNE in Derio, Basque Country, Spain with the kind help of Oscar Millet, Sivanadam Veeramuthu Natarajan, Ganeko Bernardo-Seisdedos and Jesús Jiménez-Barbero. 2D NMR spectra: ^{1}H-^{15}N SOFAST-HMQC were recorded at 298 K for 1 h each on a Bruker AVANCE III 800 MHz spectrometer with cryoprobe. Spectra of the unliganded protein and protein mixed with ligands at various concentrations were prepared in order to measure the binding constants. The signal corresponding to Nε-Hε correlation of residue R295 served as a reporter for establishing the fraction of protein that was bound with the ligand: the chemical shift was changing in slow exchange regime. Peak intensities were used to establish the bound and unbound ligand fractions. The exact protein concentration was measured using UV_{280} absorbance of stocks used to prepare the samples for titration, which were treated with 6 M guanidinium chloride in order to expose all absorbing residues. The exact ligand concentration values were determined by recording their spectra and comparing ^{1}H peaks to an internal standard, 3-(trimethylsilyl)propionic-2,2,3,3-D_4 (Sigma-Aldrich). For detailed calculation procedures refer to Ref. 269.

3.3.4 ITC experiments with *Bt* and *Bx*GH99 and their inhibitors

Isothermal titration calorimetry was performed using a MicroCal Auto-ITC200 (GE/Malvern Instruments). The protein was present in the cell and inhibitors in the syringe. Inhibitor and protein concentration were varied according to the requirements of each experiment: c-values were taken into account. Care was taken to perform experiments at a c-value of 5–1000.[275] (c=[Protein]/K_D (dimensionless)). For example, for ManNOE, the concentration of *Bt* and *Bx*GH99 was 5 µM and the ligand concentration was 50 µM. In case of poorly binding ligands such as Man-aminoDMJ, 117 µM protein and 6 mM inhibitor was used and stoichiometry

was fixed at a known value of 1 in order to improve the fit.[276] The K_D values were calculated using a modified version of Origin program with ITC200 adjustments and MicroCal PEAQ-ITC Analysis Software (Malvern Instruments).

3.4 Results and discussion

3.4.1 Inhibitor binding studies

A study of binding constants of several potential GH99 inhibitors was conducted. α-1,3-Man-substituted versions of ddMan and D-glucal, whose Glc-versions were first tested in 1993,[153] were revisited as the authors of these studies reported only IC_{50} values and K_Ds are more useful. The binding of these compounds could not be detected using ITC for reasons that were not uncovered. An alternative approach was pursued, in which ^{15}N-labelled protein was produced and a search for peaks that change in presence of a ligand in 2D NMR spectra was performed. Such peaks were found, and they corresponded to Nε-Hε correlation in residue Arg295 (*Bx*GH99, visible in Figure 2.2 on page 46 and Figure 2.5 on page 54) or Arg292 (*Bt*GH99). The bound and unbound fraction could be directly calculated from the peak intensity, as the chemical shift was perturbed in slow exchange regime. The alternative peak position signifying binding was almost the same for Man-D-glucal and ManddMan (see Figure 3.3). Upon binding of these compounds, the reporting hydrogen atoms were more shielded (lower ^{1}H ppm), and when binding the Glc-α- and β-aziridine, they were more deshielded (higher ^{1}H ppm). As chemical shift changes depend on a myriad of contributions, it is difficult to say what factors caused these particular changes. The curves obtained from the NMR experiments are presented in Figure 3.4, and the results of the ITC experiments in Figure 3.5. The K_D values are summarized in Table 3.1.

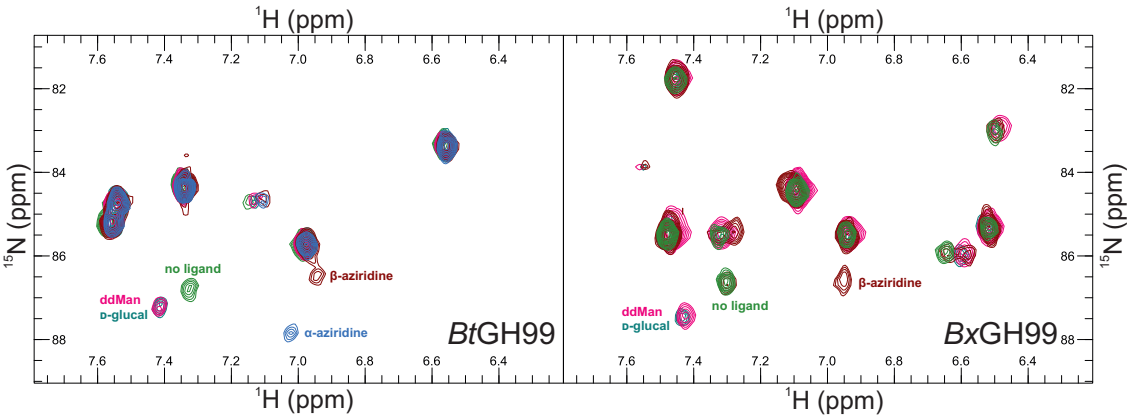


Figure 3.3: ^{1}H-^{15}N SOFAST-HMQC spectra of *Bt*GH99 (left) and *Bx*GH99 (right) recorded in absence or presence of large excesses of different ligands. Refer to the picture for the colour legend.

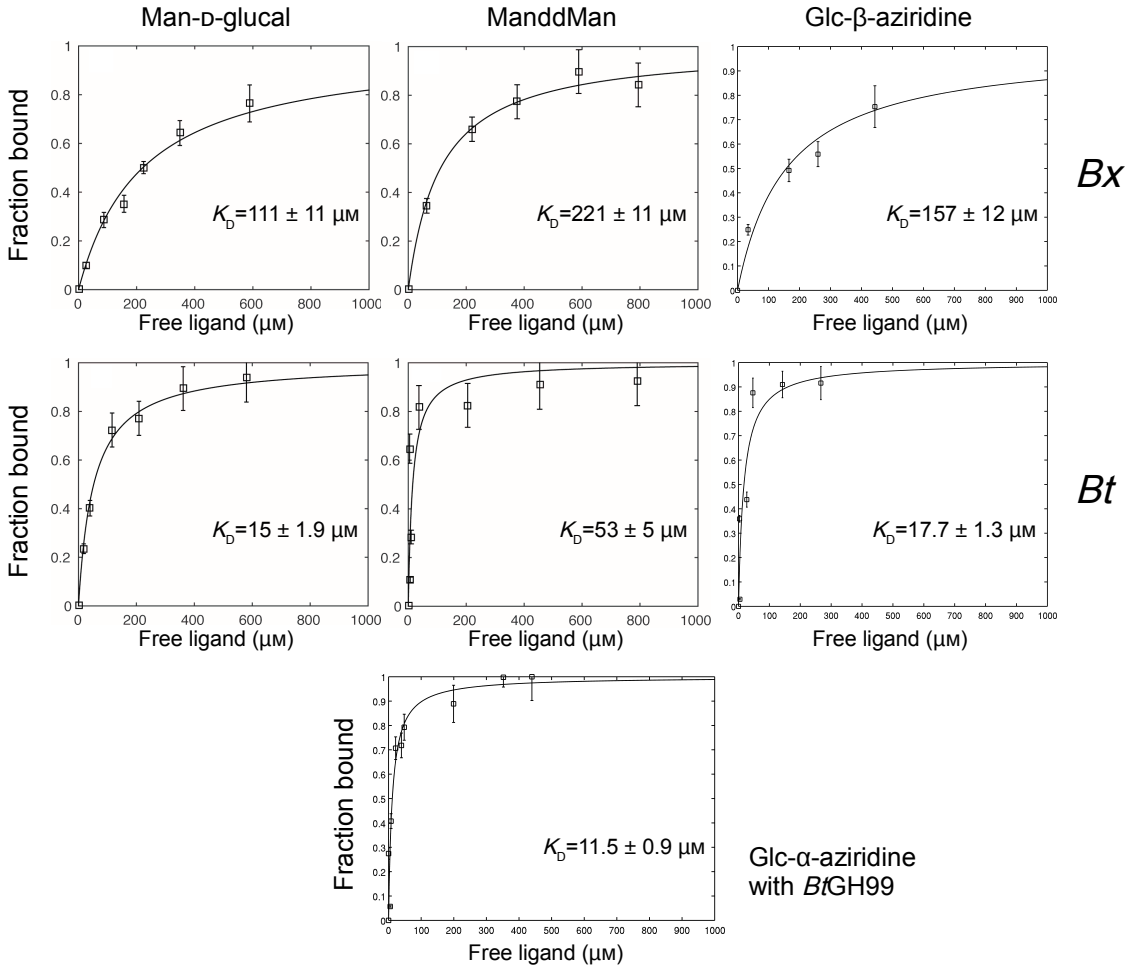
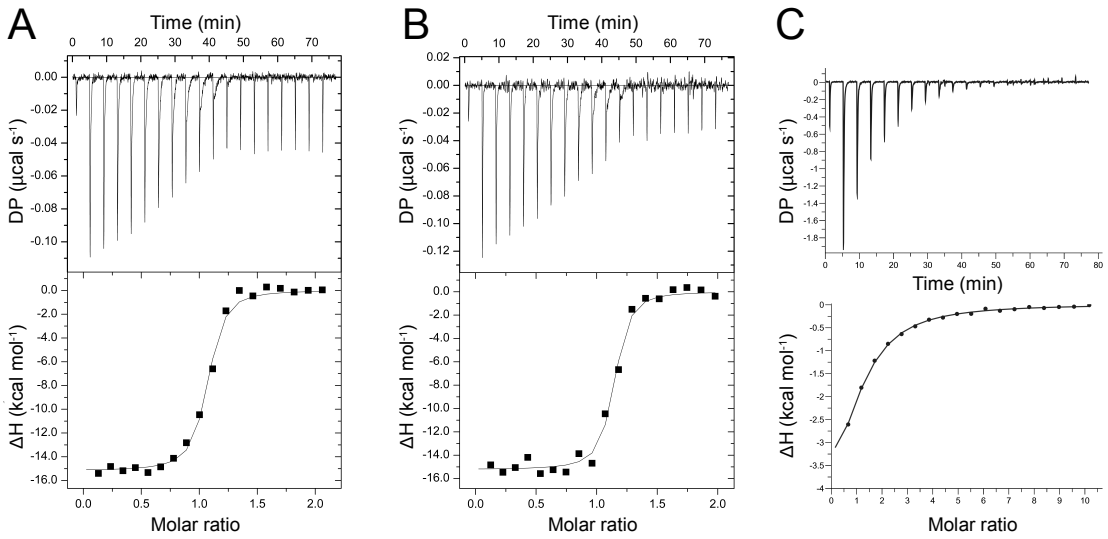


Figure 3.4: NMR titration curves for all measurements. The protein, the inhibitor and the value of the dissociation constant noted next to or on each graph. Figures were prepared using MATLAB by Ganeko Bernardo-Seisdedos.



Protein	*Bt*GH99	*Bx*GH99	*Bt*GH99
Inhibitor	ManNOE	ManNOE	ManADMJ
N	1.02 ± 0.01	1.08 ± 0.02	1 (fixed)
K_D (µM)	0.03 ± 0.01	0.013 ± 0.002	97.7 ± 4.9
ΔH (kcal mol^{-1})	-15.6 ± 0.3	-14.5 ± 0.1	-5.9 ± 0.1
-TΔS (kcal mol^{-1})	5.3	3.7	0.5

Figure 3.5: Results of isothermal titration calorimetry for ligands whose K_Ds were determined using this technique. (**A**) ManNOE binding to *Bt*GH99. (**B**) ManNOE binding to *Bx*GH99. (**C**) ManADMJ binding to *Bt*GH99. DP: differential power N: number of sites, ΔH: change in enthalpy, –TΔS: the additive inverse of the change in entropy multiplied by the absolute temperature. Graph in **C** has a different layout because data were analyzed using upgraded software.

Table 3.1: Dissociation constants of various inhibitors, sorted from weakest to best-binding. *Asterisks denote publications that use data also reported in this Thesis.

–2	–1	K_D *Bt*GH99 (µM)	K_D *Bx*GH99 (µM)	Method	Ref.
Man	ManIm	undetectable	no data	ITC	277*
Glc	cyclohexene	undetectable	no data	ITC/NMR	269*
Man	aminoDMJ	97.7 ± 4.9	no data	ITC	277*
Man	ddMan	53 ± 5	221 ± 11	NMR	269*
Glc	DMJ	24	no data	ITC	148
Glc	β-aziridine	17.7 ± 1.3	157 ± 12	NMR	–
Man	D-glucal	15 ± 1.9	111 ± 11	NMR	269*
Glc	α-aziridine	11.5 ± 0.9	no data	NMR	–
Glc	IFG	0.625	no data	ITC	148
Man	IFG	0.14 ± 0.016	0.27	ITC	159
Man	NOE	0.03 ± 0.01	0.013 ± 0.002	ITC	269*

D-Glucal, which was restricted to a $^4\boldsymbol{H}_5$ conformation, turned out to be 2–3.5 times better binding than ddMan. This result is concordant with the IC_{50} data from Hiraizumi *et al.*,[153] who found that GlcddMan was 1.6 times less effective than Glc-D-glucal at inhibiting rat endomannosidase (see Figure 1.13). ManDMJ was never synthesized, but if the K_D ratio between GlcIFG and ManIFG was taken as a benchmark, its K_D on *Bt*GH99 would be about 5 µM, which suggests tighter binding than that of Man-D-glucal. It is important to note that the presence of the endocyclic oxygen atom is crucial for this inhibitor to bind; Glc-cyclohexene binding could not be detected by ITC or NMR. At the time of the synthesis of these compounds, the $^4\boldsymbol{H}_5$ conformation was thought to be present at the transition state; in fact, it is $\boldsymbol{E}_3$, as elaborated in Section 2.4.3 (page 63). Therefore, this compound likely does not mimic the conformation of the TS.

The reaction intermediate mimic described in Chapter 2, Glc-β-aziridine, binds to both *Bt* and *Bx*GH99 with a K_D close to that of Man-D-glucal. This means that Man-substituted β-aziridine would bind about 3 × more tightly than Man-D-glucal, (K_D on *Bt*GH99 of about 4 µM). Interestingly, Glc-α-aziridine is a marginally better ligand than Glc-β-aziridine. Neither of these compounds, however, is particularly effective as an inhibitor, as the dissociation constant of the ligand that was previously known to be the tightest binder, ManIFG, was an order of magnitude lower. ManNOE was found to be an even more potent inhibitor than ManIFG, with a K_D in the range of tens of nM. The structure of Man-configured noeuromycin combines both the pseu-

doanomeric nitrogen of IFG and the axial OH2 group of DMJ. This result signifies the importance of hydrogen bonds and polar contacts in designing GH99 ligands: the aliphatic –1 subsite warheads with only hydrogen atoms at their C1 and C2 could not form them with the catalytic residues, and were not binding tightly (see Section 3.4.2).

Another unsuccessful avenue of GH99 inhibitor development were compounds Man-aminoDMJ and Man-mannoimidazole. ManADMJ binds with a modest K_D of 97.7 μM and binding of ManManIm is undetectable. The result for ManADMJ was initially puzzling, as it is structurally very similar to GlcDMJ and the hope was to achieve better affinity by exchanging OH2 for NH2, which could possibly strengthen the interaction between E333 Oε3 and the NH2 group in the inhibitor, relative to the Oε2…OH2 interaction with DMJ. This did not happen. The likely explanation for this phenomenon is that E333 protonates the NH2 and the inhibitor fails to mimic the oxocarbenium ion. Such a loss of affinity after introduction of an additional amino group to an inhibitor has been observed previously when a second amino group was introduced to apramycin, resulting in a 100-fold loss of affinity for the ribosome.[278] As for ManManIm, its poor binding can only be explained after looking at its structure (Figure 3.6**D**), which was obtained through soaking of crystals, in spite of the undetectable binding by ITC.

After correcting for the effect of the –2 sugar, the GH99 –1 subsite "warheads", listed from worst-binding to best-binding, are as follows: ADMJ, ddMan, D-glucal, DMJ, β-aziridine, α-aziridine, IFG, NOE.

3.4.2 Crystal structures of *Bx*GH99 complexes

Crystal structures of *Bx*GH99 complexed with all the inhibitors mentioned in Table 3.1 were determined, and ternary complexes with the inhibitor and α-1,2-mannobiose were also solved for ManNOE, Man-D-glucal, ManddMan, Glc-β-aziridine and Glc-α-aziridine. The data collection and refinement statistics are presented in Table 3.2 The binary complexes are shown in Figure 3.6. The occupancy of the ligands was 1, with one exception of ManManIm, for which the highest obtained occupancy was 0.8. Only crystals of the inactive E333Q mutant yielded a dataset in which Glc-cyclohexene was seen. Glc-cyclohexene could not be observed in the active site of WT *Bx*GH99, despite

prolonged soaking and repeated data collections. In contrast, Glc-β-aziridine was observed only when complexed with the WT enzyme. This is not surprising, as the hydrogen bond between Q333 Nε and aziridine NH2 is extremely unlikely to form.

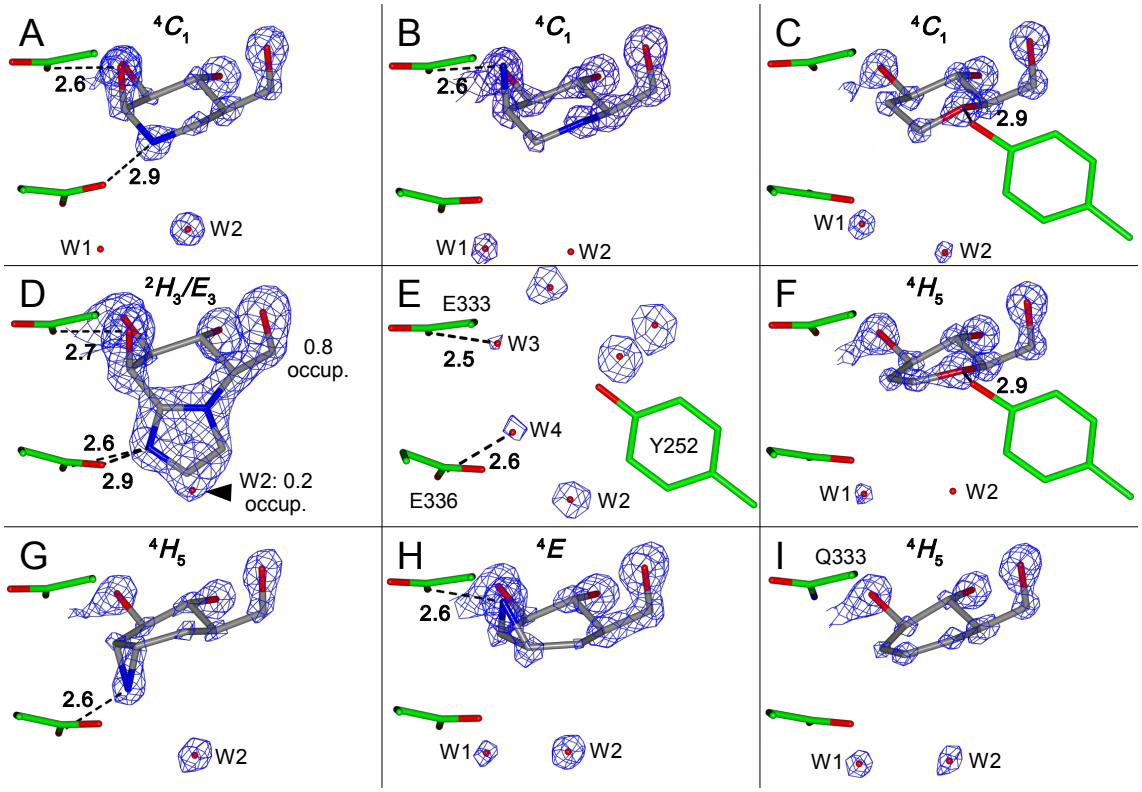


Figure 3.6: Views of the –1 subsite of *Bx*GH99 when complexed with novel inhibitors. Catalytic residues E333 and E336 (and an active site residue Y252 shown when interacting with the endocyclic oxygen) are coloured green, the ligand is coloured grey. Hydrogen bonds/close polar contacts are marked with dashed lines. **ABCDF** compounds are Man-linked and **GHI** are Glc-linked. The contour level of 2mF_o-DF_c syntheses adjusted in order to show size of the electron density on individual atoms (the contour level details are below). One or two water molecules between the –1 and +1 subsites can be seen at the bottom of each panel (numbering as in Figure 2.11, W: water). (**A**) Noeuromycin, 1.2 e$^-$/Å^3; (**B**) 2-amino-DMJ, 1.2 e$^-$/Å^3; (**C**) dideoxymannose, 1.7 e$^-$/Å^3; (**D**) mannoimidazole, 0.5 e$^-$/Å^3; (**E**) Unliganded –1 subsite from PDB code **4AD1**, 1.0 e$^-$/Å^3; (**F**) D-glucal, 1.5 e$^-$/Å^3; (**G**) α-aziridine, 1.5 e$^-$/Å^3; (**H**) β-aziridine, 0.9 e$^-$/Å^3; (**I**) 1,2-cyclohexene, 1.4 e$^-$/Å^3.

In the active site of the unliganded *Bx*GH99, residues E333 and E336 form hydrogen bonds with water molecules positioned roughly where the –1 Man OH2 group and the anomeric carbon would otherwise be (Figure 3.6**E**). The structure with NOE replicates these interactions while not introducing any steric clashes, which makes it a very good inhibitor. DMJ, β-aziridine and ADMJ both mimic the E333-OH2 interaction only. Their K_D values are comparable, but β-aziridine binds slightly more tightly. The binding of aminoDMJ is poorer probably due to from the amino group protonation problem, explained earlier. Cyclohexene, ddMan and D-glucal do not mimic any

Table 3.2: Data collection and model refinement statistics of *Bx*GH99 structures discussed in this chapter. *Values in parentheses represent the highest resolution shell.

	E333Q + Glc-cyclohexene	WT + ManddMan	WT + ManddMan + α-1,2-mannobiose	WT + Man-D-glucal	WT + Man-D-glucal + -1,2-mannobiose	WT + ManNOE	WT + ManNOE + α-1,2-mannobiose	WT + Glc-α-aziridine	WT + Glc-α-aziridine + α-1,2-mannobiose
Data collection									
Beamline	I04	I02	I02	I02	I02	I24	I04	I04	I02
Space group	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4	*I*4
Cell dimensions									
a = *b*, *c* (Å)	108.4, 67.8	108.6, 67.7	108.4, 67.6	108.7, 67.6	108.6, 67.7	108.5, 67.6	108.0, 67.5	108.8, 67.7	108.2, 67.6
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)*	39.54-1.2 (1.22-1.2)*	76.77-1.03 (1.05-1.03)	76.65-1.04 (1.06-1.04)	76.85-1.07 (1.09-1.07)	57.45-1.1 (1.12-1.1)	76.73-1.14 (1.16-1.14)	57.21-1.05 (1.07-1.05)	39.51-1.27 (1.29-1.27)	57.31-1.03 (1.05-1.03)
R_{merge}	0.059 (0.955)	0.052 (0.989)	0.052 (1.063)	0.052 (1.748)	0.061 (1.962)	0.051 (1.158)	0.054 (1.314)	0.058 (1.295)	0.058 (1.238)
R_{pim}	0.022 (0.446)	0.024 (0.566)	0.023 (0.634)	0.023 (0.874)	0.031 (1.002)	0.022 (0.855)	0.023 (0.915)	0.024 (0.550)	0.028 (0.679)
CC(1/2)	0.999 (0.543)	0.994 (0.479)	0.994 (0.427)	0.997 (0.33)	0.998 (0.414)	0.999 (0.387)	0.999 (0.308)	0.999 (0.515)	0.988 (0.409)
$\langle I / \sigma I \rangle$	15.3 (1.7)	14.4 (1.2)	13 (1)	10.9 (0.6)	6.9 (0.4)	10.6 (0.7)	10.8 (0.7)	13.1 (1.4)	12.4 (1.0)
Completeness (%)	98.7 (82.7)	97.5 (73.2)	99 (87.3)	99.8 (96.5)	99.7 (95.2)	94.6 (61.5)	99.3 (89.1)	100 (100)	98.0 (75.8)
Redundancy	7.9 (5.3)	6 (3.6)	6 (3.5)	6.3 (4.9)	4.8 (4.7)	5.6 (2.3)	5.9 (2.7)	6.6 (6.4)	6.1 (3.8)
Refinement									
Resolution (Å)	39.54-1.20	76.77-1.03	76.65-1.04	76.85-1.07	57.45-1.10	76.73-1.14	57.21-1.05	39.51-1.27	57.31-1.03
No. reflections all / free	121915 / 6060	188468 / 9325	184685 / 9151	172107 / 8530	158574 / 7870	134703 / 6708	178944 / 8872	103773 / 5172	187827 / 9303
R_{work} / R_{free}	0.12 / 0.14	0.12 / 0.13	0.12 / 0.13	0.13 / 0.14	0.14 / 0.16	0.12 / 0.14	0.11 / 0.13	0.13 / 0.15	0.12 / 0.14
No. atoms									
Protein	3168	3211	3231	3161	3163	3234	3166	3132	3080
Ligand/ion	21	33	67	21	52	22	60	26	49
Water	355	423	374	394	358	427	480	343	383
B-factors (Å^2)									
Protein	17	15	15	17	17	17	16	19	16
Ligand/ion	15	14	17	14	19	17	14	18	16
Water	32	33	33	36	34	35	36	32	33
RMS. deviations									
Bond lengths (Å)	0.012	0.010	0.011	0.011	0.010	0.010	0.013	0.009	0.012
Bond angles (°)	1.6	1.5	1.6	1.6	1.5	1.5	1.7	1.4	1.6
PDB code	**5MEL**	**5M17**	**5M3W**	**5M5D**	**5MC8**	**5LYR**	**5M03**	–	–

of these hydrogen bonds and their binding is quite poor, although D-glucal is the best of the three. This is likely because of the combined effect of less steric clashes with aliphatic atoms owing to its $^{4}\boldsymbol{H}_{5}$ conformation, and the presence of the endocyclic oxygen. The aziridine NH group of α-aziridine mimics the interation with E336 in a unique way: by replicating an interaction between E336 Oε and water 4, observed in the unliganded structure. Mannoimidazole seems to mimic interactions with both catalytic residues just like ManNOE, but the imidazole nitrogen atom being at a close distance to both E336 Oε atoms is unfavourable (as shown in Figure 3.6**D**). Its conformation is $^{2}\boldsymbol{H}_{3}$ / $\boldsymbol{E}_{3}$, which brings about costs in terms of conformational free energy.

The relative size of the electron density peak corresponding to water 1 and water 2 found between the –1 and +1 subsites seems to depend on the inhibitor bound. Water 2 peak is larger in the structure with the carbocation mimic NOE and in the structure with β-aziridine (Figure 3.6**AH**). The peak is larger for water 1 in complexes with oxocarbenium ion mimics: ADMJ, ddMan and D-glucal (Figure 3.6**BCF**). Fittingly, the sizes are the same in the structure with the cyclohexene (Figure 3.6**I**). In the structure of WT with open Glc-β-epoxide (Figure 2.11**B**, page 65) the electron density peak if water 2 was more pronounced, implying that the molecule that is being turned over has a carbocation-like charge. Both of the water molecules around the epoxide in Figure 2.11**A** have similar electron density peaks. This might mean that the E333Q mutation prevents the formation of a carbocation.

Figure 3.7 offers a closer look at the binding pose of mannoimidazole with *Bx*GH99. The C1-C2 bond in the bound mannoimidazole and the –1 mannoside residue in the Michaelis complex are exactly parallel (Figure 3.7**AC**), which is not the case for any other GH99 inhibitor structures. Comparison of the Michaelis complex and the structure with α-1,3-mannobiose shows the C1-C2 bond in the disaccharide is not parallel to the one in the tetrasaccharide (Figure 3.7**BD**). Mannobiose sits deep in the –2/–1 binding sites and its position is more reminiscent of that of ManNOE. The rat endomannosidase was inhibited by GlcMan disaccharide with a very modest IC_{50} of 220 μM. ITC experiments with α-1,3-mannobiose and *Bt*GH99 indicated binding but the K_D was unmeasurable by this method (over 2 mM). These values suggest that GH99 endomannosidases are not significantly affected by product inhibition.

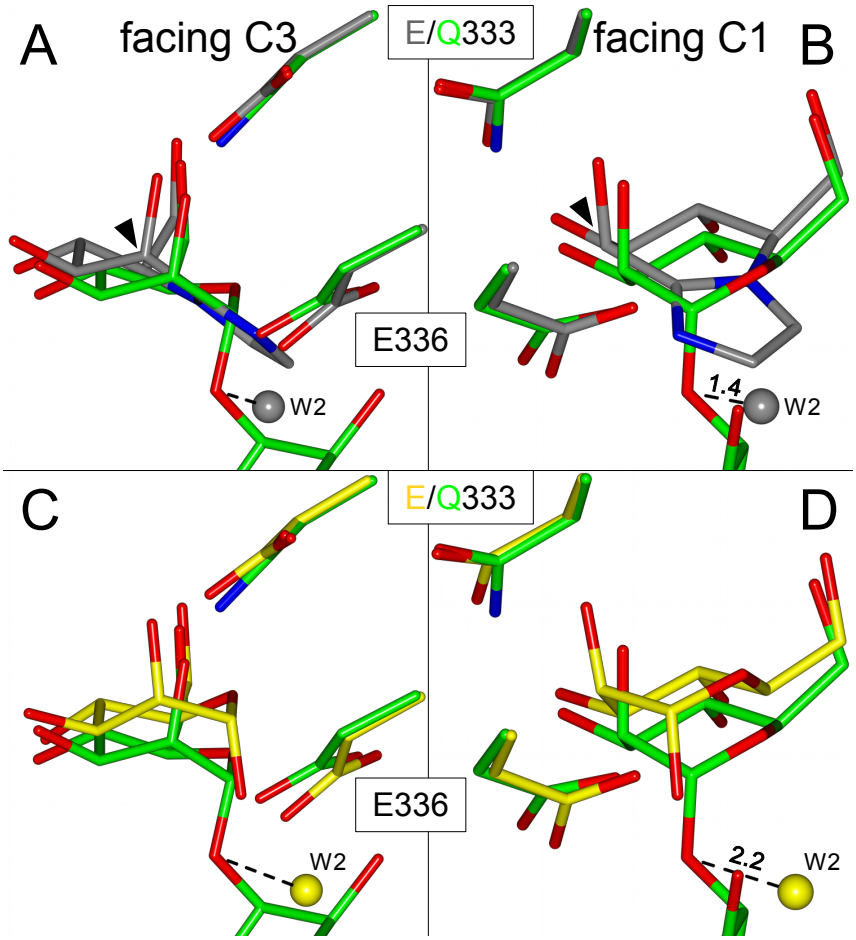


Figure 3.7: Mannoimidazole bound in the –1 subsite mimics the TS. Projected distance between water 2 molecule and the glycosidic oxygen in the aligned model of the Michaelis complex is shown. (**AB**) Comparison between the Michaelis complex (green) and ManManIm complex (grey). Arrowhead points to the C1-C2 bond. (**CD**) Comparison between the Michaelis complex (green) and the complex with α-1,3-mannobiose only (yellow). **AC** show the species bound in the –1 subsite facing the C3 and **BD** facing the C1. Models were aligned using the main chain atoms of residues 333 to 336 in ccp4mg.[204]

The best affinity for GH99 is achieved by azasugars with a pseudoanomeric nitrogen. These are thought to mimic the carbocation, as opposed to ligands with the endocyclic oxygen swapped to a nitrogen which mimic the oxocarbenium ion. Despite its poor affinity, ManADMJ binds in exactly the same position as GlcDMJ (Figure 3.8**B**) and the crystal structure does not make it obvious why the affinity is lost. With ManNOE vs. ManIFG the reason for a 4–20 × (depending on the protein) better affinity is clear: the addition of the OH2 group allows for another hydrogen bond with the enzyme to form in comparison to IFG. However, ManNOE, due to its hemiaminal functional group at C2, is predicted to not be a particularly stable molecule in the sometimes oxidizing conditions inside living cells. Indeed, in solution NOE exists in a 1:2 resonance of *manno*- and *gluco*-forms.[279] ITC experiments with ManNOE with *Bt* and *Bx*GH99 needed an unusually long time period (240 s) between injections in order for the DP to return to baseline (see Figure 3.5**AB**), which might have been caused by part of the compound "mutarotating" from the *gluco*-form to the *manno*-form and binding to the enzyme only then. This is in contrast to IFG which is a stable compound, and unlikely to be hydrolyzed. GlcIFG seems more useful in clinical applications, even if the binding of GlcNOE to the eukaryotic endomannosidase is probably tighter.

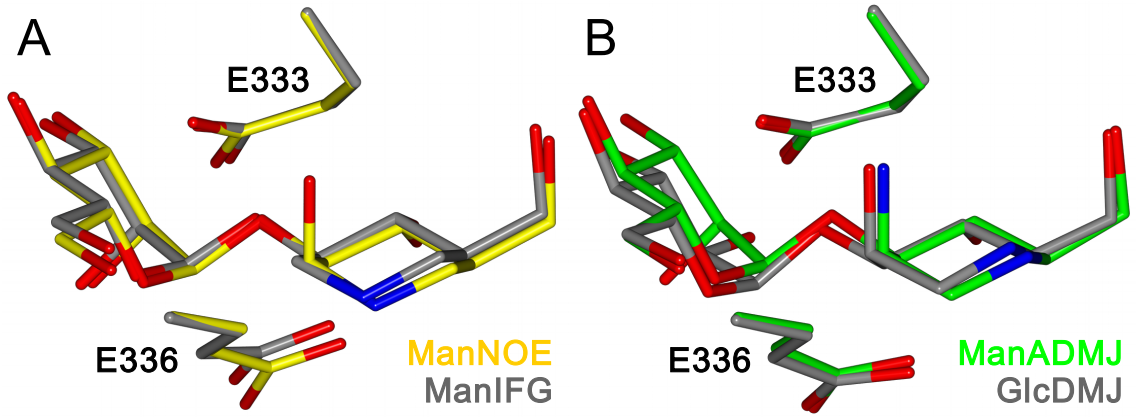


Figure 3.8: Comparison of the binding poses of structurally similar GH99 inhibitors that were published previously (with *Bx*GH99). (**A**) ManNOE (yellow) vs. ManIFG (grey). (**B**) ManADMJ (green) vs. GlcDMJ (grey). Models were aligned using main chain atoms of residues 333 to 336 in ccp4mg.[204]

3.5 Conclusions

The data presented in this chapter in context with the literature paint a picture of a unique enzyme. Two molecules that are excellent inhibitors of many mannosidases, swainsonine and mannoimidazole, do not bind when adapted to act as GH99 inhibitors. Carbocation-mimicking agents like ManNOE and ManIFG are the tightest GH99 inhibitors and it is difficult to think of ways to improve on this rather minimalistic design. It is possible that swapping the pseudoanomeric nitrogen for an atom that carries an even larger positive charge could improve affinity. At this point, however, ManIFG and GlcIFG seem like the best compromise between stability and affinity, and they are also amenable to be modified at other sites for increased stability. Man-2-amino-NOE could be one possible solution to the instability problem, but it is likely to lose its affinity due to the vicinal amino group problem.

Interestingly, the "wrong" (α-configured) aziridine is a marginally stronger-binding molecule than the β-configured aziridine. Only the latter closely resembles the structure of the reaction intermediate. The shape of the molecule is not likely to be the factor that makes a good GH99 ligand, and seems to be secondary to the charge-charge interactions and hydrogen bonds inside the active site. As long as the active site can accommodate the inhibitor with few steric clashes, which is not the case for compunds such as ManManIm, a compound is worth trying with the enzyme.

The research presented in Chapters 2 and 3 was done on the bacterial forms of GH99: *Bt*GH99 and *Bx*GH99. The later part of this thesis will concentrate on a protein which is clinically more relevant, the human endo-α-mannosidase. Their structural similarity will validate that the results obtained for the bacterial enzyme are meaningful, but differences between these proteins will also be discussed, as well as the significance of these differences.

Chapter 4

Gene cloning, protein production and structure determination of human *endo*-α-mannosidase: MANEA

4.1 Abstract

Structural research into *endo*-α-mannosidase had originally concentrated on bacterial GH99 endo-α-mannanases, which are homologs of the human enzyme. The relative difficulty of the studies on the human *endo*-α-mannosidase stems from it being a transmembrane protein that needs the presence of a detergent to be solubilized,[176] which usually means crystallization of such proteins is challenging. Folding of proteins expressed recombinantly in *E. coli* can also be impaired. In this chapter I will describe how these obstacles were overcome, culminating in obtaining the crystal structure of the catalytic domain of the human protein MANEA. The overall fold of the protein is similar to that of the endomannanases whose structures had already been solved. The main differences between the active site of the endomannanase and the endomannosidase are: a tyrosine residue coordinating a water that hydrogen bonds with the –2 sugar in place of a tryptophan in the endomannanase, and the presence of a loop making contacts with the –2 sugar. The active site environment around the –1 and +1 subsites was conserved, which validates the use of *endo*-α-1,2-mannanases as models for studying the reaction mechanism of the *endo*-α-1,2-mannosidase. The results presented below fill in a gap in the knowledge of *N*-glycan-processing enzymes and provide leads for further structural and functional studies on this enzyme.

4.2 Introduction

The *Homo sapiens* genome harbors two genes that encode proteins categorized in CAZy glycoside hydrolase family 99: *MANEA* and *MANEAL*.[280] *MANEA* is located on chromosome 6 and *MANEAL* on chromosome 1.[281] The *MANEA* gene has one isoform whose protein product is 462 aa long. *MANEAL* has three splice forms, the protein product of one of which (isoform 3) is of approximately the same length as MANEA protein (457 aa). The two others – isoforms 1 and 2 – give rise to complementarily truncated N-terminal and C-terminal fragments of isoform 3. The GH99 domain is fully present in isoform 3 and split in half in the two shorter isoforms. The short isoforms of *MANEAL* are expressed at a level that is about 1/2 of the long isoform expression level.[282] Comparison of RNA expression levels on all currently listed human tissues using Metabolic gEne RApid Visualizer (MERAV)[283] suggests they are weakly anticorrelated (*MANEA*/*MANEAL* CC=–0.27). Interestingly, data from the same source indicate that their expression levels in primary tumours (CC=0.01) and cancer cell lines (CC=0.09) are not correlated. A close look at the data reveals that *MANEAL* is expressed predominantly in the brain, while *MANEA* in all other tissues, including non-CNS (non-central nervous system) nerves (Figure 4.1). Provocatively, a 1989 study by Tulsiani *et al*.[161] looked at endomannosidase activity isolated from the rat brain tissue before *MANEAL* was understood to be a separate gene. This might have been the first, inadvertent look at MANEAL, the activity of which has not been confirmed.

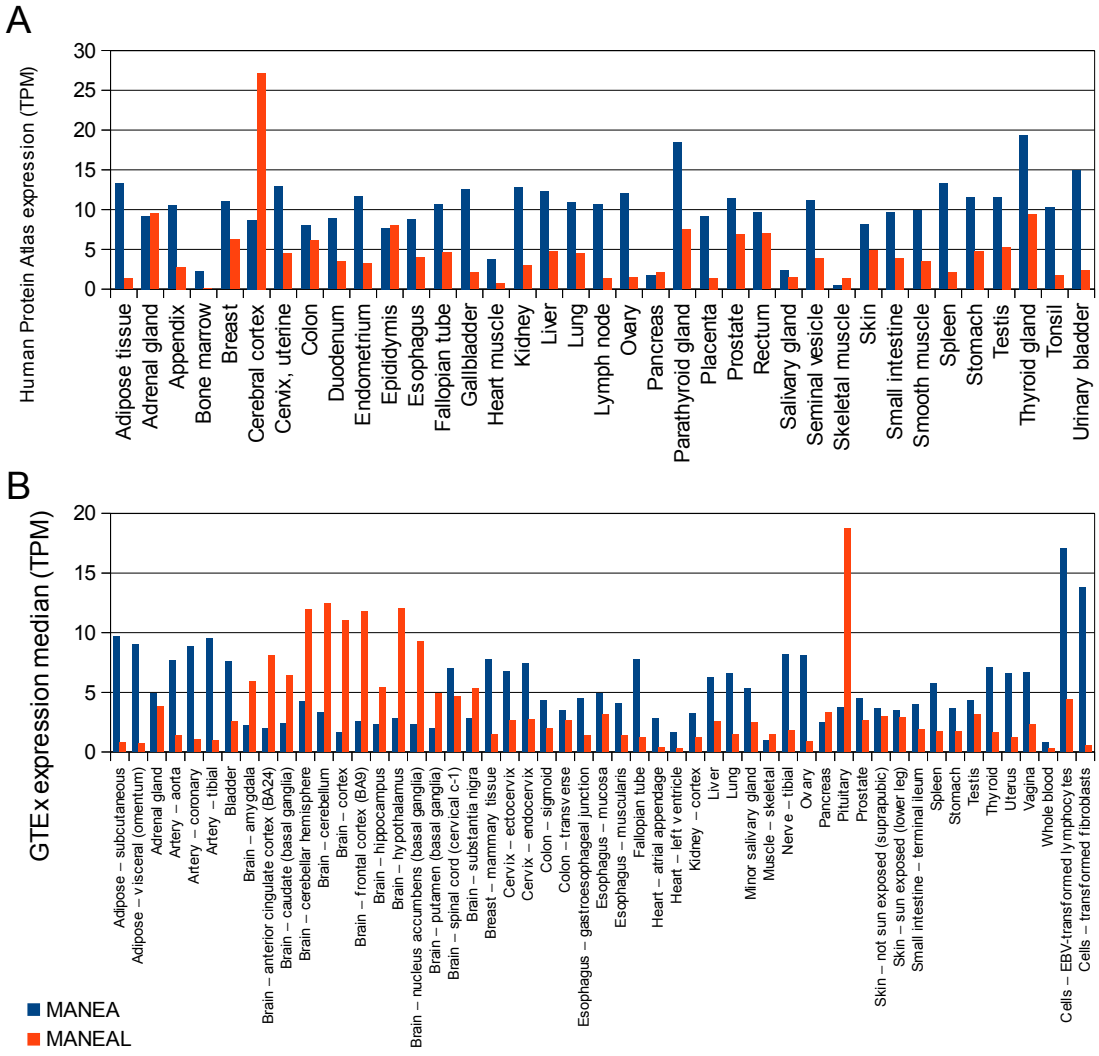


Figure 4.1: Comparison of RNA-seq data for *MANEA* and *MANEAL* from two sources: (**A**) the Human Protein Atlas[284,285] and (**B**) the Genotype-Tissue Expression database (GTEx).[282] The data cover expression in various human tissues, and the GTEx dataset is extended by transformed cell lines. Note the enhanced expression of *MANEAL* in the CNS. TPM: transcripts per million.

Functional studies have shown MANEA is a type II membrane protein.[195] It is highly likely that MANEAL has the same overall fold as MANEA. MANEA was known to be localized to the Golgi membrane since its discovery in 1987[130] – it was first isolated from rat Golgi membranes – but there were no studies targeting MANEAL. Human Protein Atlas[286] provides a recent (2017), direct piece of evidence showing the Golgi localization of MANEAL: a micrograph of immunostained cells (Figure 4.2). It is also likely that MANEAL has the same enzymatic activity as MANEA, but it is localized to the CNS only, and the expression of *MANEAL* is likely cell-specific, just like *MANEA* expression.[177] MANEA from rat liver was observed to be specifically inhibited by HEPES and HEPPS[146] but not MES or MOPSO buffers. HEPES and HEPPS were uniquely affecting the endomannosidase and not α-glucosidase or *exo*-α-mannosidase. Tris abolished the activity of all of these enzymes.

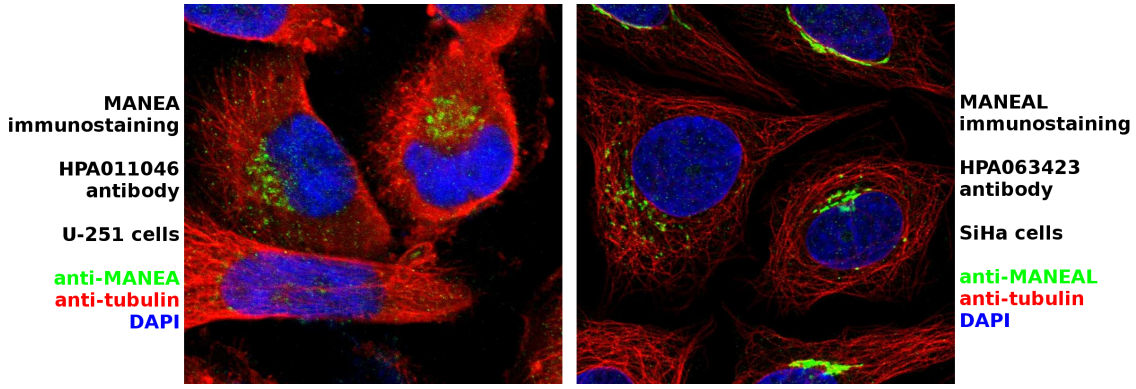


Figure 4.2: Immunofluorescence images showing that both MANEA (left panel) and MANEAL (right panel) are localized to the Golgi apparatus. U-251 cells are derived from human malignant glioblastoma and SiHa from human cervix. Data from the Human Protein Atlas.[284,285]

Until the 2012 study by Thompson *et al.*[148] there have been no attempts at characterizing the fold of GH99 proteins. In the article they show that the *Bacteroides thetaiotaomicron* and *Bacteroides xylanisolvens* GH99 fold is a modified $(\beta/\alpha)_8$ barrel and find that it can catalyze the same reaction as the human GH99 *endo*-α-1,2-mannosidase (EC 3.2.1.130) with overall retention of anomeric conformation. In a later article, it was found that the primary activity of these GH99 proteins is a selfish mechanism by which human gut bacteria utilize yeast mannan.[207] Since the 2015 publication by Hakki *et al.*[159] the bacterial GH99s have been referred to as *endo*-α-1,2-mannanases. *Endo*-α-1,2-mannanase activity does not have an EC number assigned at the time of writing and therefore is referred to as EC 3.2.1.- (although it still correct to refer to these enzymes as EC 3.2.1.130, because they can catalyze the same reaction as the

human enzyme).

The primary amino acid sequence of MANEA shares 39% identity to both *Bt* and *Bx*GH99s proteins, with 45% identity when only the catalytic domain is considered in the alignment. Figure 4.3 shows a global sequence alignment of the GH99 proteins considered in this passage. The main difference between the proteins is that a tyrosine residue conserved for all *endo*-α-1,2-mannosidases, Y189 (*Hs*MANEA numbering), is a tryptophan residue in all *endo*-α-1,2-mannanases (W126 according to *Bx*GH99 numbering). The residues that interact directly with the bound substrate are also conserved in the MANEAL protein, with a minor but consistent difference: *Hs*MANEA E224 is a Q226 in *Hs*MANEAL. The significance of this difference is not known. The analysis of the MANEAL primary sequence, aligned with *Hs*MANEAL, *Bt*GH99 and *Bx*GH99 in Figure 4.3, corroborates the view that MANEAL has the same overall transmembrane topology as MANEA.

The efforts in this part of the project were a continuation of Dr Andrew Thompson's attempts at producing a folded version of MANEA. He established that the Δ34, His-tagged version of the protein, when produced in *E. coli* BL21(DE3) cells, does not fold properly and is instead deposited in inclusion bodies. The protein was amenable to isolation from the insoluble fraction, but did not refold. The same truncation but with an additional N-terminal His-tag and a C-terminal EGFP-tag (enhanced GFP), was determined by Dr Andrew Thompson to be soluble and active. However, with this large tag present, the kinetic data were not reliable, and removing it proved to be difficult. Here I will describe how these problems were overcome and present the results that were obtained from various crystallization efforts of the human enzyme.

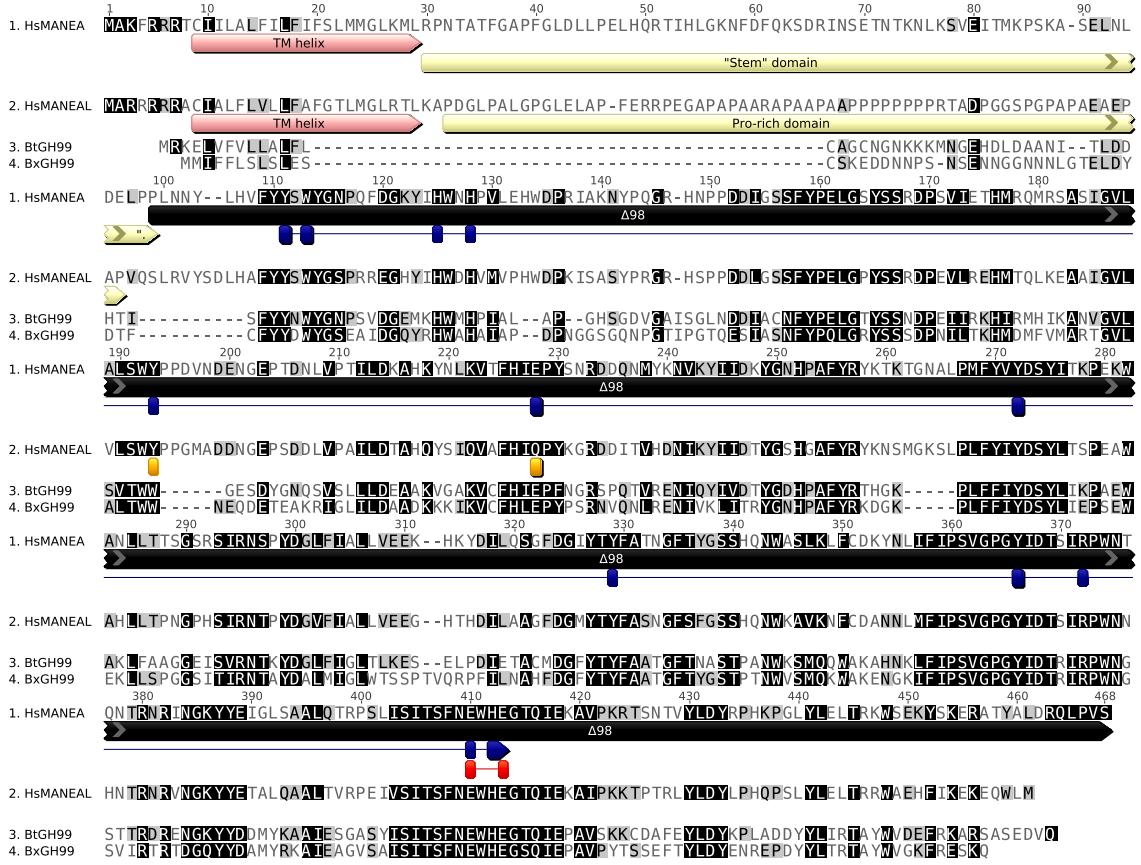


Figure 4.3: Alignment of the primary sequences of human and *Bacteroides* GH99s. The differences between residues conserved within a specific protein family are highlighted with the orange rectangle. Residues close to the substrate are marked in blue, the catalytic residues in red. The endomannosidase catalytic domain corresponds to the MANEA-Δ98 truncation less 8 C-terminal aa. Amino acid symbols are shaded according to conservation (BLOSUM62 score matrix[200] with a similarity threshold of 1, black background: 100% similar, dark grey: 80–100% similar, light grey: 60–80% similar, white: less than 60% similar). Figure made in Geneious[201]; sequences aligned using the MUSCLE algorithm.[202]

4.3 Materials and methods

4.3.1 Methods overview

The following experimental section will start with a commentary on the methods used and the methods will be presented subsequently. The first attempts at *MANEA* expression were made using pGEn1- and pGEn2-MANEA constructs obtained from the GlycoEnzymes repository.[287] The gene in pGEn1-MANEA DNA encodes for MANEA protein which is N-terminally tagged with a signal sequence for extracellular secretion, an octahistidine tag, and a StrepII tag (for affinity purification using Strep-tactin). The gene in pGEn2-MANEA DNA encodes for MANEA protein which is N-terminally tagged with the same signal sequence, the octahistidine tag, an Avi tag (for affinity purification) and a superfolder GFP (GFP whose fold is more stable than the fold of wild-type GFP). The strains used for mammalian expression were CHO-K1 (ATCC) and HEK-293T (ATCC). This approach was chosen due to the lack of success by Dr Andrew Thompson in purification of soluble human *endo*-α-mannosidase from *E. coli*. After numerous attempts (transient transfection of the cells or stable transfection and selection of best-expressing clones), that spanned about 18 months, no protein was obtained. This prompted a reconsideration of bacterial expression methods.

Gene expression was performed in BL21(DE3) cells (Agilent) that were made competent using a modified version of a one-step protocol[258] where LB media in TSS (transformation and storage solution) was substituted with MQ H_2O. Constructs containing four N-terminal truncations of the *MANEA* gene (optimized for expression in *Escherichia coli*) at positions 60, 71, 83 and 98 fused to a C-terminal EGFP tag were cloned using restriction enzymes into the pET-28a(+) vector. The *MANEA* gene was truncated in the stem domain as shown in Figure 4.4. After expression the genes encoding the EGFP fusion proteins, the amount of proteins was checked and the results indicated presence of protein in the soluble fraction, with a larger amount of soluble protein in the longer truncations. This was likely only due to the larger protein being stained by more dye and not a difference in molar concentration or its mRNA level (expression level). Cell-free expression systems have been used to produce correctly folded, full-length MANEA[195] but the yield of such techniques is usually not

high enough to enable structural studies. The publication of a study describing a relatively simple method of producing soluble MANEA in *E. coli*[164] led me to adopt an approach of co-overexpression of *MANEA* with the genes encoding the GroEL molecular chaperone. This became the method of choice for producing large amounts of pure MANEA.

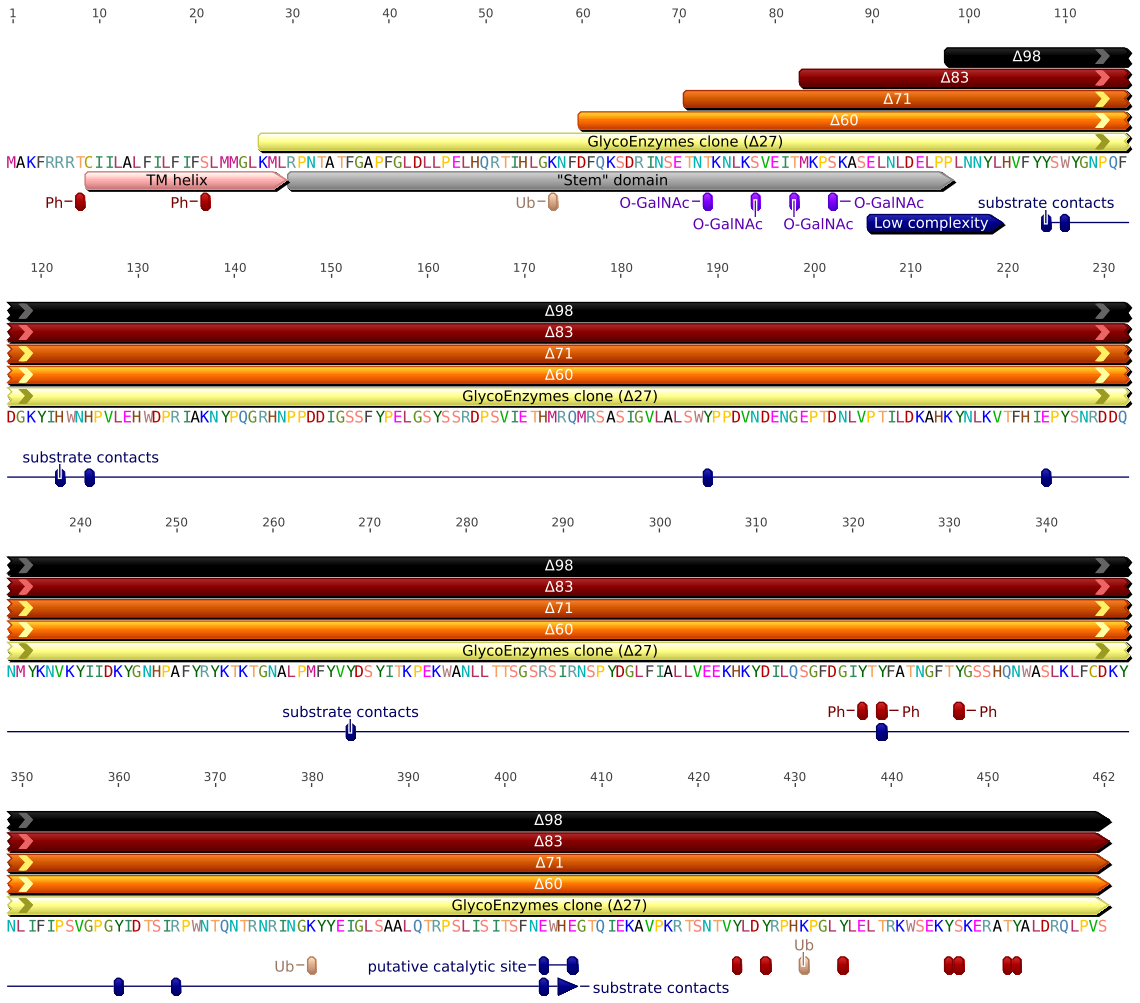


Figure 4.4: Full, annotated sequence of MANEA protein with truncations shown. The GlycoEnzymes clone refers to the sequence present in pGEn1 and pGEn2-MANEA constructs. Major structural motifs are highlighted. Residues making close contacts to the substrate and the active site residues are annotated based on alignment with *Bx*GH99 structures. Putative phosphorylation (Ph), *O*-glycosylation and ubiquitinylation (Ub) sites were found using PhosphoSitePlus[203], additional ubiquitinylation from UbiSite[288] and ESA-UbiSite.[289] The amino acid residues are coloured by type. Figure made in Geneious.[201]

Truncation MANEA-Δ98 was selected, based on the sequence alignment with *Bt* and *Bx*GH99, as the most likely to contain the folded domain only and therefore most crystallizable, and the pCold-I construct was used for expression. After initial screening, it became apparent that GroEL was produced in large amounts, but was not soluble

and therefore not helpful as a molecular chaperone. An experiment by Claire Fowler revealed that adding 20 mM $MgCl_2$ to the growth media stabilized GroEL. After this was included in the protocol, another problem with initial large-scale expression experiments was encountered: the protein yield was poor, 0.35 $mg_{protein}$ per litre of culture. This was rectified by using TB (Terrific Broth) media (modified from 4 to 8 g glycerol per litre of media), which increased the yield 7–8-fold.

4.3.2 Truncated *MANEA* gene cloning and subcloning

Q5 High-Fidelity DNA Polymerase (New England Biolabs, Inc.) was used for all PCR (polymerase chain reaction) experiments. Eight PCR experiments were set up, each with a total volume of 20 µl and the following ingredients: 500 pg template DNA (*Hs*MANEA optimized for *E. coli* cloned into pET28a(+), generated before the start of this Thesis work by Dr Andrew Thompson), 10 mM dNTPs (deoxiribonucleotide triphosphates), 10 µM forward primer, 10 µM reverse primer and 0.02 units μl^{-1} Q5 High-Fidelity DNA Polymerase (0.2 µl in each reaction mixture). The components were mixed in 0.2 ml PCR tubes. The PCR cycler (BioER LifeTouch) was preheated to 98 °C before putting the tubes inside and the hot lid was used in order to prevent condensation on the tube lid. The tubes were put inside the cycler, incubated at 98 °C for 30 s and thermocycled 35 times. The denaturation step was 10 s at 98 °C, the annealing step was 30 s at 60 °C and the extension step was 40 s at 72 °C. After 35 cycles were complete, the mixtures were incubated for 2 min at 72 °C for a final extension step and the temperature was lowered to 10 °C for storage until later use. The primers used in the PCR reactions are presented in Table 4.1. The forward primers used for the amplification of each truncated version of *MANEA* were combined with a reverse primer that added an *EcoRI* cleavage site (for C-terminal EGFP fusions) or a reverse primer that added an *XhoI* cleavage site (for cloning into the MCS (multiple cloning site) of empty pET28a(+) vector).

Table 4.1: Primers used for PCR amplification of the genes encoding truncated MANEA. The *italicized* sequence adds a 5′ overhang improving the efficiency of digestion by restriction enzymes. The underlined sequence is the palindromic restriction site. The region in **bold** is complementary to the template DNA.

Primer sequence	Primer name
*GATCAG*CATATG**GACTTTCAGAAATCAGATCGTA**	*Hs*MANEA Δ60 forward
*GATCAG*CATATG**ACCAACACGAAAAATCTGAA**	*Hs*MANEA Δ71 forward
*GATCAG*CAT**ATG**AAACCGTCGAAAGCAAG	*Hs*MANEA Δ83 forward
*GATCAG*CATAT**G**CCGCTGAACAATTACCTGCA	*Hs*MANEA Δ98 forward
*GATCAG***GAATTC**GGAGACCGGCAGTTGAC	*Hs*MANEA *EcoRI* reverse
*GATCAG***CTCGAG**TCAGGAGACCGG	*Hs*MANEA *XhoI* reverse

Agarose gel (0.7% w/v) electrophoresis was performed on all of the thermocycled PCR mixtures. The electrophoresis grade agarose was obtained from Melford Laboratories, Ltd. When the DNA was resolved well enough to confirm that the size of the PCR product was correct, the electrophoresis was stopped and the band corresponding to each product was cut out from the agarose gel using a scalpel. The agent that visualized the DNA under near-UV light was SYBR Safe Stain (Invitrogen). Subsequently, the DNA was extracted from the bands using NucleoSpin Gel and PCR Cleanup kit (Macherey-Nagel) and dissolved in Buffer NE (5 mM Tris/HCl, pH 8.5) supplied with the kit.

The extracted PCR products (20 µl of each) were double digested with *NdeI* and *EcoRI* (in case of DNA used for C-terminal EGFP fusions) in *EcoRI* buffer + BSA (bovine serum albumin) or *NdeI* and *XhoI* in NEBuffer 4 + BSA. All enzymes and buffers were obtained from New England Biolabs, Inc. 1 µg of each of the vectors: pET28a(+) vector containing an EGFP tag (generated by Dr Andrew Thompson before the start of this Thesis work) and empty pET28a(+) were double digested using the same protocol. The digestion mixture was incubated at 37 °C for 75 min and subsequently ran on 0.7% agarose gel electrophoresis. When resolved, the bands were extracted from the gel as explained above. The insert and the vector DNA were mixed in a 3:1 molar ratio and ligated using T4 DNA ligase for 15 min at RT (room temperature) (20 µl total ligation mixture volume, 1 µl enzyme used). After the ligation, 2 µl of each ligation mixture was used for transformation of DH5α competent cells (New England Biolabs, Inc.). Transformation of the competent cells with the ligated DNA was achieved by incubating the cells (~10 µl) for 15–30 min on ice in 1.5 ml Eppendorf tube with the

ligation mixture added, followed by a heat shock at 42 °C for 45 s and 2 min recovery on ice. After this, SOC media was added to each tube up to a total volume of 500 µl and the cells were shaken at 37 °C and 180 RPM for 1 h. 1–10% of the total cell culture volume was used to plate on LB agar plates with 50 mg ml^{-1} kanamycin added. The plates were incubated at 37 °C overnight and stored at 4 °C. At the end of the same day, cells from single colony forming units (CFU, 2 from each transformation plate) were used to start a 5 ml liquid Kan^+ LB culture, which was shaken overnight at 37 °C. 2 ml of each culture was used for plasmid DNA extraction, which was performed using QIAprep Spin Miniprep Kit (QIAGEN). The coding sequence region in each plasmid (8 in total) was sequenced using the following primers: T7 (sequence: TAATACGACTCACTATAGGG) and pET-RP (sequence: CTAGTTATTGCTCAGCGG). Sequencing was done by GATC Biotech (**SUPREME**RUN service). In case of ambiguous or wrong sequencing result, DNA extracted from the second CFU was resequenced. This was sufficient to confirm that the sequences were correct.

In order to subclone *MANEA* truncations into pCold-I vector (obtained from TaKaRa Bio, Inc.), 1 µg of pET28a(+) containing truncated, optimized *MANEA* truncations were double digested with *NdeI* and *XhoI* as above, and the empty pCold-I vector was linearized in the same way. The digestion products were resolved using 0.7% agarose gel and extracted as above. Each of the four DNA fragments encoding truncated *MANEA* was ligated in 3:1 molar ratio with the linearized pCold-I vector as described earlier. The selective agent used for culturing cells harbouring the pCold-I vector was ampicillin (Amp) at 100 µg ml^{-1}. Sequencing of truncated genes encoding *MANEA* subcloned into the pCold-I vector was not performed. The details of the design of pCold-I vector can be found in Ref. 290.

Chaperone competent cells were made from BL21(DE3) cells that were transformed with each of the five vectors from Chaperone Plasmid Set (TaKaRa Bio, Inc.). The same protocol as referred to previously (published in Ref. 258) was used, with one difference: the cells were cultured in liquid media in presence of chloramphenicol (Cam) at 20 µg ml^{-1}. Each of the plasmids from the Chaperone Plasmid Set carries a chloramphenicol resistance gene. The plasmids are: pG-KJE8, pGro7, pKJE7, pGTf2 and pTf16. Transformation of these cells with pColdI-*Hs*MANEA-Δ98 was achieved by using the same protocol as used for the transformation of DH5α cells. Maps of pCold-I-*Hs*MANEA-Δ98 and the pGro7 plasmid are shown in Figure 4.5.

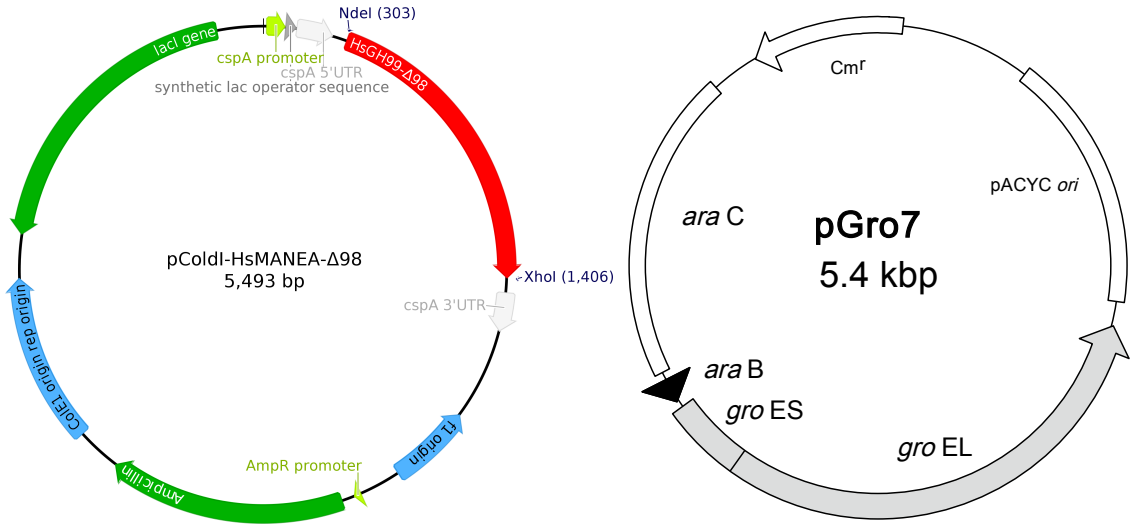


Figure 4.5: The expression system used to produce *Hs*MANEA-Δ98. Left: pCold-I-*Hs*MANEA-Δ98, right: pGro7 vector harbouring the *gro*EL and *gro*ES genes. *Gro*EL and *gro*ES are inducible by L-arabinose. Cm^r: chloramphenicol resistance gene, cspA: cold shock protein CspA gene.

In some cases, glycerol stocks of BL21(DE3) cells transformed with plasmids containing the genes to be expressed were made in order to facilitate subsequent expression experiments. To make a glycerol stock, cells from a single CFU were cultured overnight in 2–5 ml LB media with selective agents (Amp, Cam or Kan) added. 250 µl of the culture was pipetted into a 2 ml cryotube, followed by adding 250 ml of 50% v/v glycerol solution. The components were mixed with a pipette, the tubes were flash-frozen in liquid nitrogen and stored at -80 °C. In order to use the stock, it was removed from the -80 °C storage, the surface of the cells was scraped with a pipette tip, and the tip was then briefly dipped in selective media to inoculate the media.

4.3.3 Production and purification of soluble MANEA and its E404Q variant

Final protocol: starter cultures were grown overnight in LB supplemented with Amp at 100 µg ml^{-1} and Cam at 20 µg ml^{-1}. Modified Amp^+/Cam^+ TB media with 20 mM $MgCl_2$ and 0.5 g l^{-1} L-arabinose was inoculated with 1% v/v of the starter culture. Cells were shaken in 2 l plastic flasks at 220 RPM until the OD_{600} was at least 0.8. At this point, cells were cold-shocked in an ice-water slurry for 5 min and stood at 15 °C for the next 30 min. Subsequently, IPTG was added to a final concentration of 0.2 mM. The cultures were then shaken overnight at 15 °C. The cell pellet was harvested by

centrifugation of the cell culture and then frozen at -80 °C.

The protein was purified in two steps: nickel affinity chromatography and cation exchange chromatography. Firstly, the cell pellet was thawed and 50-150 ml of Buffer A1 (50 mM potassium phosphate pH 7.0, 260 mM KCl, 20 mM $MgCl_2$, 20 mM imidazole) or buffer A2 (25 mM HEPES pH 7.0, 260 mM NaCl, 20 mM $MgCl_2$) was added to the pellet. The cells were resuspended and transferred to a suitable container, then sonicated on ice until the solution became darker and more viscous. Cell debris was separated by centrifugation in 30 ml spin tubes at 41600×g for 30 min at 4 °C. The supernatant was loaded onto a 5 ml Ni-NTA HisTrap FF column (GE) preloaded with Co^{2+} ions instead of Ni^{2+} and preequilibrated with buffer A1 or A2. The column was washed with 5 volumes of buffer A1 or A2 and the protein was eluted in gradient of, respectively, buffer B1 (50 mM potassium phosphate pH 7.0, 300 mM KCl, 500 mM imidazole) or B2 (25 mM HEPES pH 7.0, 300 mM NaCl, 500 mM imidazole). Two peaks were seen in the chromatogram (see Figure 4.7 in the Results and discussion Section) and fractions collected from the peak that eluted at a higher concentration of imidazole was collected and buffer-exchanged into buffer SP A (50 mM potassium phosphate pH 7.0, 50 mM KCl). The solution was loaded onto a 5 ml HiTrap SP FF column (GE). The SP column is packed with a strong cation exchange resin - Sepharose functionalized with $-CH_2CH_2CH_2SO_3^-$ groups. FF relates to the flowrate the column can withstand with no damage (Fast Flow). The difference between the isoelectric point of MANEA (9.2) and GroEL (4.5) was used to separate the chaperone molecule. A 50→500 mM KCl gradient was applied. *Hs*MANEA eluted around 185-200 mM salt, separate from GroEL. Finally, the fractions of interest were concentrated using an Amicon 30,000 Da ultrafiltration device, buffer-exchanged on the filter into storage buffer 1 (50 mM potassium phosphate pH 7.0, 185 mM KCl) or storage buffer 2 (25 mM HEPES pH 7.0, 200 mM NaCl), aliquoted and flash-frozen in liquid nitrogen. The maximal yield was ~3 mg protein per litre of culture and the protein was stable up to a concentration of 15-18 mg ml^{-1}.

The pCold-I-*Hs*MANEA-Δ98 construct was mutated at two sites, encoding residues E404 and E407. These are equivalent to the *Bx*GH99 catalytic residues E333 (base/acid) and E336 (acid/base), respectively. The mutagenesis was performed using of Q5 Site-Directed Mutagenesis Kit (NEB). Mutagenic primers, containing two base changes per codon, were designed using the NEBaseChanger tool (see

Table 4.2). Three mutants: two single and one double (all E→Q), were generated. MANEA E404Q mutant was produced and purified in exactly the same way as the WT, with an additional gel filtration step using a HiLoad Superdex 16/600 S75 column in storage buffer 2 in an attempt to remove an insignificant amount of high-molecular weight impurities (three bands with apparent MWs of 70, 75 and 82 kDa). The same impurities were also present in the WT enzyme (best visible in Figure 4.7 on page 102 as the bands above GroEL in the collected fractions from the cobalt column). In case of the WT protein they were not removed, because they were observed during a purification procedure performed after a successful crystallization attempt and it was assumed they do not interfere with the process of crystallization. The gel filtration chromatography was performed after cation exchange purification. The amount of the impurity appeared to be reduced only slightly, but it constituted less than 1% of the total protein mass in the sample and was unlikely to affect any subsequent experiments.

Table 4.2: Primers used in the mutagenesis of the pCold-I-*Hs*MANEA-Δ98 construct. The bases shown in **bold** are the mutation site.

Primer sequence	Primer name
CCATTCGTTAAAGGAGGTGATACTAATC	*Hs*MANEA 404 WT
CCA**C**T**G**GTTAAAGGAGGTGATACTAATC	*Hs*MANEA E404Q
CATGAAGGCACGCAAATTGAAAAAG	*Hs*MANEA 407 WT
CAT**C**A**G**GGCACGCAAATTGAAAAAG	*Hs*MANEA E407Q

4.3.4 Crystallization and crystal handling

A number of crystal forms of *Hs*MANEA were obtained. The conditions and protocols common to all of them, unless indicated otherwise when describing each crystal form, are presented in this paragraph. The crystals were grown at 19 °C. The robot used to set up crystallization screens was Mosquito LCP (TTP Labtech). 54 µl of the reservoir solution was transferred in parallel from a crystal screen aliquoted in a 96-well deep well block to all the wells of standard 96-well MRC sitting drop crystallization plates using Hydra 96 robot (Robbins Scientific). This was the final volume of the reservoir solution. The final volume of the droplets in 96-well crystallization screens was 300 nl. Crystal fishing was performed using a nylon cryoloop adapted to the crystal size and the diffraction quality was checked using an in-house X-ray

source (Rigaku MicroMax 007HF with a RAXIS IV++ detector). No cryoprotection was used.

Crystal form 1: The first crystal form of *Hs*MANEA WT was obtained in PACT *premier*™ Screen (Molecular Dimensions Ltd) well B3. The crystallization condition contained 100 mM MIB (malonate/imidazole/boric acid) buffer, pH 6.0 (crystals of different quality were observed at pHs between 4 and 7). The protein at a concentration of 5.5 mg ml^{-1} was kept in 50 mM potassium phosphate, 50 mM KCl (storage buffer 1). Crystals suitable for fishing appeared after about one month. The dataset used to solve the MANEA structure was collected using Diamond Light Source (DLS) beamline I03. The space group was $P2_12_12$. The crystal diffracted to a resolution of 2.25 Å.

Crystal form 2: *Hs*MANEA crystallized in 100 mM sodium citrate pH 5.5 and 20-32.5% w/v PEG 3350 (protein solution:reservoir ratio 1:1 or 1:2, *Hs*MANEA concentration 10 mg ml^{-1}). The protein was in this case kept in storage buffer 2 (25 mM HEPES pH 7.0, 200 mM NaCl). The crystals appeared after 2 days but their morphology was not adequate for fishing and X-ray crystallography.

Crystal form 3: *Hs*MANEA crystallized in 1 M sodium succinate pH 7.0, 1% w/v PEG-MME 2000 (JCSG+ Screen, well F11). The protein solution:reservoir ratio 1:1 or 1:2, *Hs*MANEA concentration 10 mg ml^{-1} (in storage buffer 2). The crystal was fished and yielded a dataset collected on DLS beamline I04. The space group was $P4_32_12$. The crystal diffracted to a resolution of 2.25 Å.

Crystal form 4: An attempt at crystallizing the protein with an additive Anderson–Evans polyoxotungstate $[TeW_6O_{24}]^{6-}$ (TEW) was made in order to find a crystal form that would allow the structure solution from high resolution data and would allow ligands to bind in the active site. TEW (shown in Figure 4.6) is a compound that can act as a crosslinker between protein molecules, enabling the otherwise often disordered residues (e.g. Lys or Glu) that hamper the crystallization process to participate in crystal contacts[291]. Only four crystal structures with this agent are currently deposited in the PDB and they comprise three proteins: hen egg white lysozyme[292], a plant aurone synthase[293] and a mushroom tyrosinase.[294] The crystal form of tyrosinase is unusual, as it is an example of heterologous crystallization of the latent and and active forms of the enzyme.

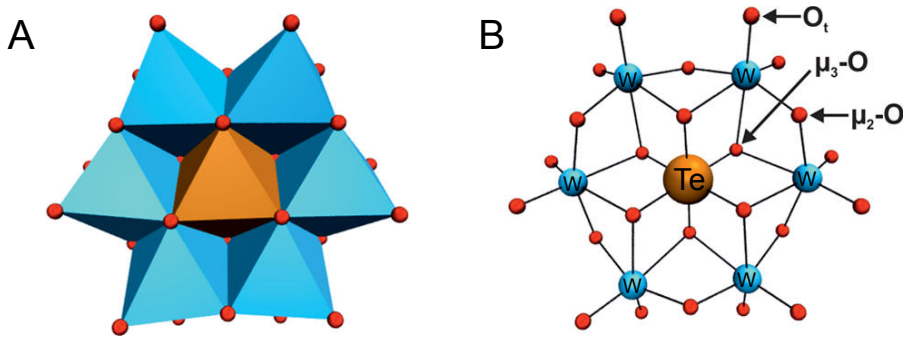


Figure 4.6: The structure of Anderson–Evans polyoxotungstate (TEW). (**A**) Polyhedral representation of the molecule (**B**) Ball and stick representation with coordination modes of oxygen atoms assigned. The molecule carries a –6 charge. Figure adapted from Ref. 291.

XP Screen (Jena Bioscience), which contains 1 mM TEW in each condition, was used in screening for *Hs*MANEA co-crystallization with TEW and the experiment resulted in finding a novel crystal form in well B5 (condition: 100 mM HEPES pH 7.5, 200 mM $MgCl_2$, 30% w/v PEG 400, 1 mM TEW). The protein concentration was 9.2 mg ml^{-1} (in storage buffer 2). The crystals appeared after 8 days in wells with protein solution:reservoir ratio of 1:1 and 1:2 but the crystals were larger in the droplet with a 1:1 ratio. Some of the crystals were fished and a dataset was collected on DLS beamline I03. After this, the crystals from **crystal form 4** were soaked overnight with ~20 mM GlcIFG and subsequently ~40 mM α-1,2-mannobiose to obtain complexes with these compounds. Fishing was performed after each soaking experiment. Structures of a binary complex with of *Hs*MANEA with GlcIFG and a ternary complex with GlcIFG and α-1,2-mannobiose were obtained from data collected on DLS beamline I03. The space group was $P6_2$. These crystals diffracted to a resolution of 1.8–2.0 Å.

Co-crystallization with compounds known to form a complex with MANEA: Co-crystallization experiments of *Hs*MANEA WT and E404Q mutants were performed. For the WT protein, the first co-crystallization was attempted by screening 4 mg ml^{-1} *Hs*MANEA containing a 20 × molar excess of GlcIFG (kept in storage buffer 1). This returned no droplets that contained protein crystals. A second attempt was made with 10 mg ml^{-1} *Hs*MANEA containing a 10 × molar excess of both GlcIFG and α-1,2-mannobiose (in storage buffer 2). The aim of the experiment was to obtain a ternary complex with –2 to +2 sites occupied, but this was unsuccessful as well.

Crystal form 5: The E404Q inactive mutant was co-crystallized with $GlcMan_4OMe$ synthesized by Pearl Fernandes in Spencer Williams' laboratory. Numerous hits were

obtained with the protein. Crystals in Crystal Screen HT (Hampton Research) well F1 were observed after 1 day. The initial conditions were: 100 mM sodium acetate pH 4.6, 200 mM ammonium sulfate, 30% w/v PEG-MME 2000. These condition was later optimized and the crystal used for structure solution was fished from a well with a precipitant containing 100 mM sodium acetate pH 4.6, 200 mM ammonium sulfate and 12.8% w/v PEG-MME 2000. In this case, the plate was kept at 6 °C, the protein solution:reservoir ratio was 1:1 and 500 nl of each was used. The experiment was done on a 48-well MRC Maxi sitting drop crystallization plate. The data were collected on DLS beamline I03, and the space group was $P2_1$. The crystal diffracted to a resolution of 1.1 Å.

4.3.5 Structure determination

The images containing diffraction data were indexed, integrated and merged using DIALS software[262] integrated into Xia2 pipeline for automatic data processing.[261] Data reduction was done in Aimless.[263] 5% of reflections were not used in refinement in order to cross-validate the refinement statistics through the use of R_{free}. A new random R_{free} set of reflections was generated for each crystal form. The initial molecular replacement (MR) model was created from a model with PDB code **5M17** (wild-type *Bx*GH99 in complex with 1,2-dideoxymannose[269]) using CHAINSAW[295] (the polypeptide chain only). The amino acid identity between *Bx*GH99 and *Hs*MANEA was 42%. MR was done using PHASER.[296] After refinement of the initial model to the first *Hs*MANEA dataset the R_{work}/R_{free} values were 0.35/0.38. Subsequently, the model was rebuilt manually using Coot[265] and refined in Refmac5.[264] Hydrogens were added in riding positions during refinement for all structures, except the **crystal form 3**, for which they did not improve statistics. Water molecules were added to the model after the protein chain was as complete as possible. After the first round of refinement with waters present, difference map peaks were looked at in order to find more water molecules. Agreement between the model and the electron density was periodically checked using edstats.[297] Dictionary files needed for refinement were made in JLigand,[274] with the exception of the dictionary file for TEW, which was made by Dr Jon Agirre in Libcheck. NCS restraints were used in the refinement of **crystal form 3** form, which was the only one with non-crystallographic symmetry

(3 molecules in the asymmetric unit). Anisotropic B factors (Debye-Waller factors) were used in the refinement of **crystal form 5** because of a relatively high resolution of the data (1.1 Å).

Chain A from **crystal form 3** was used as the search model for subsequent molecular replacement procedures. Isomorphous structures (complexed with saccharides) of the $P6_2$ crystal form with ligands were solved by matching the HKL index of the solution to the first solution (with HEPES bound in the active site) in Aimless and subsequent direct refinement (in Refmac5) of a pdb file containing only the polypeptide chain of the WT + HEPES structure.

4.3.6 Circular dichroism

Circular dichroism experiments were performed at York Technology Facility using a Jasco J810 CD Spectrophotometer with a prism monochromator. The samples were kept in a 1 mm quartz cuvette. *Hs*MANEA at 0.2 mg ml^{-1} was stable in low salt conditions (2.2 mM KCl). Ellipticity data were collected every 0.5 nm for wavelengths from 190 to 260 nm. Before data analysis, the buffer signal was subtracted from the sample signals. Data were analyzed using the CAPITO web server.[298]

4.3.7 Homology modelling of *Hs*MANEAL

The refined model of **crystal form 3** of MANEA (chain A, Figure 4.10**B**) was used as a template for I-TASSER[299–301] to model MANEAL. All the previously published, bacterial GH99 models were excluded from the modelling job. Only the part of MANEAL that was homologous to MANEA was modelled (aa 95–457).

4.4 Results and discussion

4.4.1 Overview of results

*Hs*MANEA was purified to a high degree of homogeneity. The results of the purification of *Hs*MANEA WT are presented in Figure 4.7.

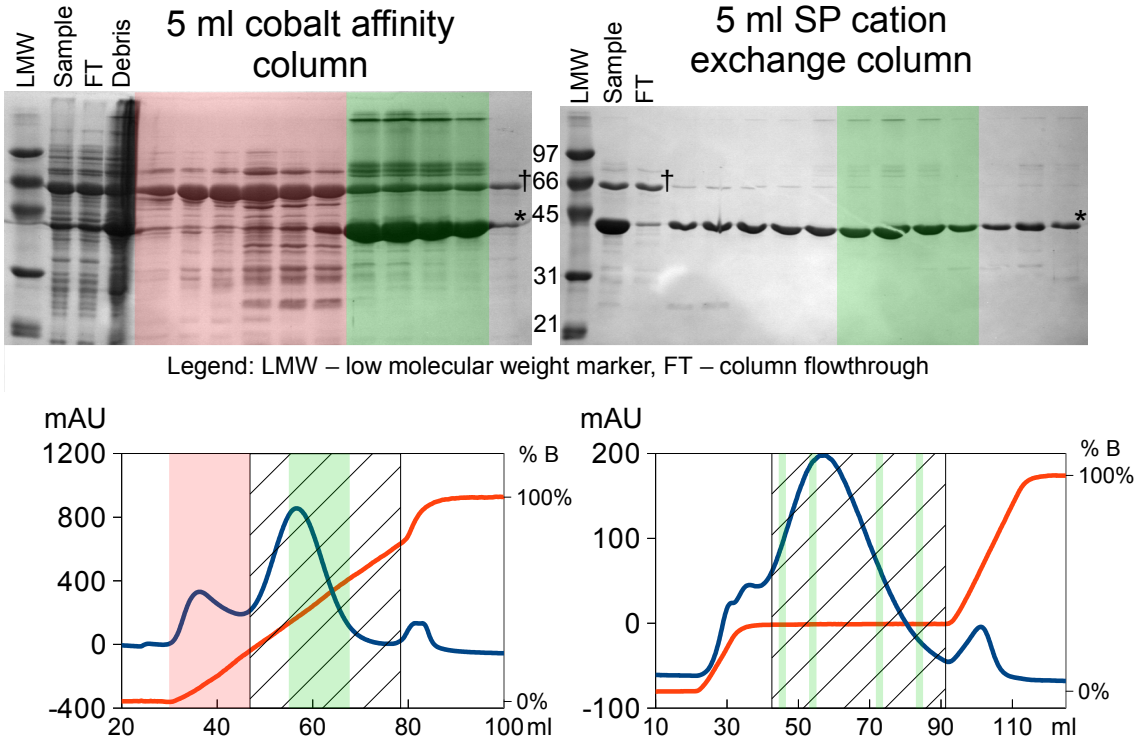


Figure 4.7: Purification of *Hs*MANEA WT from BL21(DE3) cells. Cobalt affinity column (FF crude) was used in the first step, followed by SP cation exchange chromatography. The blue line represents the Abs_{280} (absorbance at 280 nm)) signal (mAU), and the orange line the concentration of buffer B (arbitrary 0-100% scale). The collected fractions are tinted green on the 12% reducing SDS-PAGE gel image and the corresponding chromatogram, the discarded fractions are tinted red. The molecular weight of ladder bands are in kDa. The hatched box represents all fractions that were collected for later use. †GroEL chaperone, *N-terminally His-tagged *Hs*MANEA.

*Hs*MANEA crystallized in 5 crystal forms, 4 of which were solved. **Crystal form 1** was the first one to be solved. Knowing the crystal structures of bacterial GH99s would prove indispensable for phasing the data, a trimmed and mutated model of *Bx*GH99 was used as a search model in MR. An unambiguous solution was found, with one molecule in the asymmetric unit (ASU) and space group $P2_12_12$ (unit cell dimensions (Å): a=38.5, b=86.5, c=135.9, $\alpha=\beta=\gamma=90°$). The diffraction data were usable up to a resolution of 2.25 Å. After model building and refinement it became apparent that the structure, while informative about the protein fold, would not permit ligand binding

analysis. N-terminal His residues from the His-tag and His residues from the active site participated in coordinating a metal ion (of unknown identity, possibly Ni^{2+} that leached from the nickel column used for purification: cobalt affinity columns were used only in later purification runs). Moreover, the integrity of the crystal seemed to be provided by the ion: the crystal contacts between proteins were mediated through it (see Figure 4.8).

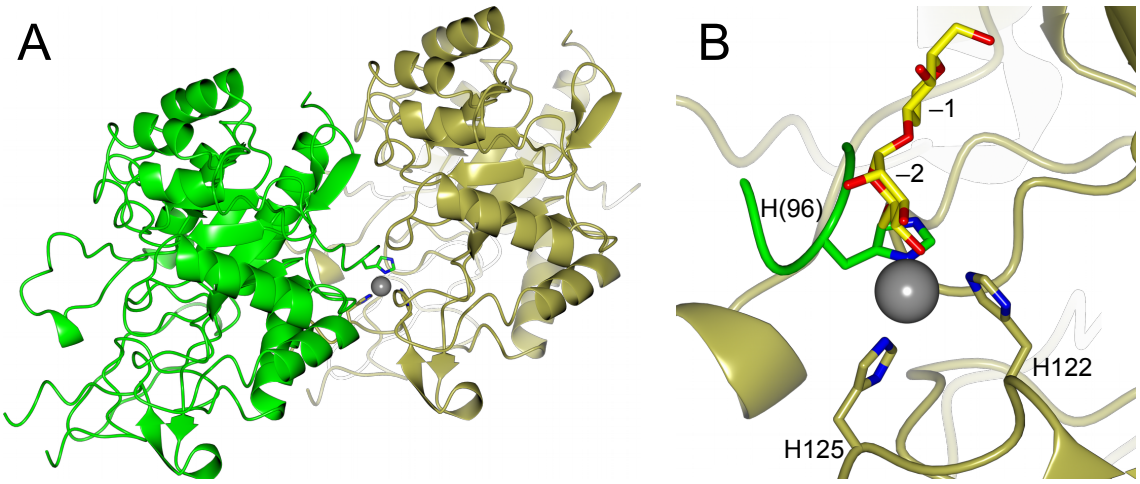


Figure 4.8: Views of the packing of *Hs*MANEA **crystal form 1** around the nickel ion. The molecule which coordinates the ion with an N-terminal H residue is shown in green, and the molecule coordinating the metal ion with the active site H residues is shown in gold. (**A**) Packing of adjacent molecules in the crystal. (**B**) The residues around the metal ion. H(96) is present in the sequence because of the way it was cloned. It is encoded by the first three bases of the *NdeI* restriction site. The saccharide shown in yellow comes from structural alignment (in ccp4mg)[204] of *Hs*GH99 shown in gold with PDB code **5M17**. H(96) and the metal ion block the –2 subsite.

A search for a different crystal form was commenced. Attempts at cleaving off the N-terminal His-tag with the aim of reducing the strength of the chelation did not produce usable crystals. Performing crystallization screening again with protein in 25 mM HEPES buffer instead of 50 mM potassium phosphate yielded conditions that permitted growth of a different crystal form (**crystal form 3**). Its space group was $P4_32_12$ with cubic tendencies (c almost equal to a and b: a=144.2, b=144.2, c=139.7). The collected data were of lower resolution (3 Å), and it was possible to solve the structure using MR with the initial model from the first solution. *Hs*MANEA in **crystal form 3** appeared to be in a more ordered state. In the first solution, a particular loop region (residues 131–141), was not visible in the electron density. The **crystal form 3** allowed for this to be built. More importantly, the metal ion was not present in the active site, but electron density not belonging to the protein was observed there and it was accurately explained by modelling in HEPES, the buffering agent that was found to inhibit the rat endomannosidase enzyme three decades before.[146] **Crystal form 4** was found by using the polyoxometalate TEW (see Figure 4.6) as an additive. The structure from this crystal, which diffracted to a resolution of 2.0 Å was solved with MR using a truncated model from **crystal form 3** (protein chain A only). The data collection and refinement statistics for *Hs*MANEA structures not in complex with saccharides can be found in Table 4.3. The same initial model was used in MR to solve **crystal form 5**, *Hs*GH99 E404Q mutant structure obtained from its co-crystallization with substrate $GlcMan_4OMe$ (the crystal diffracted to a resolution of 1.1 Å). No metal ions likely to be functionally important were found in any *Hs*MANEA models, as was the case for *Bx* and *Bt*GH99.[148]

Table 4.3: Data collection and model refinement statistics of *Hs*MANEA crystal structures without saccharides bound. *Values in parentheses represent the highest resolution shell.

Crystal form	**1**	**3**	**4**
*Hs*MANEA variant	WT	WT	WT
In complex with	Ni^{2+}	HEPES	HEPES + TEW
Data collection			
Space group	$P2_12_12$	$P4_32_12$	$P6_2$
Cell dimensions			
a, *b*, *c* (Å)	38.5 86.5 135.9	144.2, 144.2, 139.7	129.0, 129.0, 48.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 120
Resolution (Å)	135.92–2.25 (2.33–2.25)*	102.00–3.00 (3.16–3.00)	64.48–2.00 (2.05–2.00)
R_{merge}	0.200 (1.872)	0.596 (1.175)	0.126 (0.739)
R_{pim}	0.083 (1.002)	0.135 (0.265)	0.078 (0.450)
CC(1/2)	0.994 (0.345)	0.974 (0.253)	0.992 (0.411)
$<I / \sigma I>$	7.0 (0.9)	5.0 (2.9)	6.2 (1.5)
Completeness (%)	98.1 (84.5)	100 (99.8)	99.6 (98.3)
Redundancy	11.4 (6.9)	20.3 (20.5)	3.5 (3.5)
Refinement			
Resolution (Å)	73.10–2.25	102.2–3.00	111.69–2.00
No. reflections			
all / free	21817 / 1106	30099 / 1429	31481 / 1554
R_{work} / R_{free}	0.18 / 0.22	0.21 / 0.27	0.20 / 0.24
No. atoms			
Protein	2906	8853	2998
Ligand/ion	1	45	140
Water	185	140	165
B-factors (Å^2)			
Protein	51	49	34
Ligand / ion	88	53	76
Water	51	36	36
RMSD			
Bond lengths (Å)	0.011	0.010	0.011
Bond angles (°)	1.4	1.5	1.6
PDB code	–	–	–

The results of the circular dichroism experiment were of limited usefulness. The estimates of the secondary structure content in *Hs*MANEA at 20 °C were: 7–10% α-helix, 37–44% β-sheet and 46-56% irregular residues (curve most similar to that of β-crystallin B2). The actual values for the solved, truncated structure with uncleaved polyhistidine tags are: 27.5% α-helix, 16.5% β-sheet and 56% irregular. A thermal unfolding experiment (Figure 4.9) reveals that the point at which MANEA loses 50% of its secondary structure lies at about 50 °C. Interestingly, when MANEA is unfolded, its position on the $[\theta]_{222}$ vs $[\theta]_{200}$ graph* is closer to that of pre-molten globules, rather than the fully unfolded proteins.

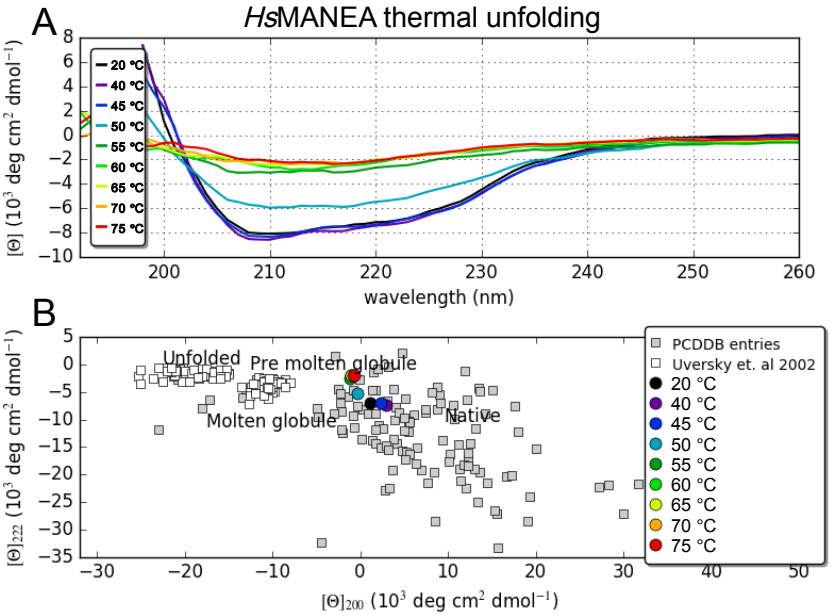


Figure 4.9: Analysis of circular dichroism experiments of MANEA. (**A**) A graph of molar ellipticity [θ] vs wavelength at different temperatures. (**B**) Comparison of $[\theta]_{222}$ vs $[\theta]_{200}$ of MANEA at different temperatures and the values for other proteins.[302,303] Figure made using CAPITO web server[298].

4.4.2 Comparison with other GH99 structures

The structure of *Bx*GH99 is by far the best studied GH99 structure to date. The models of *Bx*GH99 and *Hs*MANEA are very similar (see Figure 4.10). The overall fold is the same, with even short β-strands that form the core replicated in both proteins.

*$[\theta]_{222}$ vs $[\theta]_{200}$ graph is useful for assessing the folding state of a protein. Unfolded proteins have a comparatively low ellipticity at 222 nm (usually below 0) and around 0 at 200 nm.[302] Folded and partially folded proteins usually differ from unfolded proteins in ellipticity at these wavelengths.

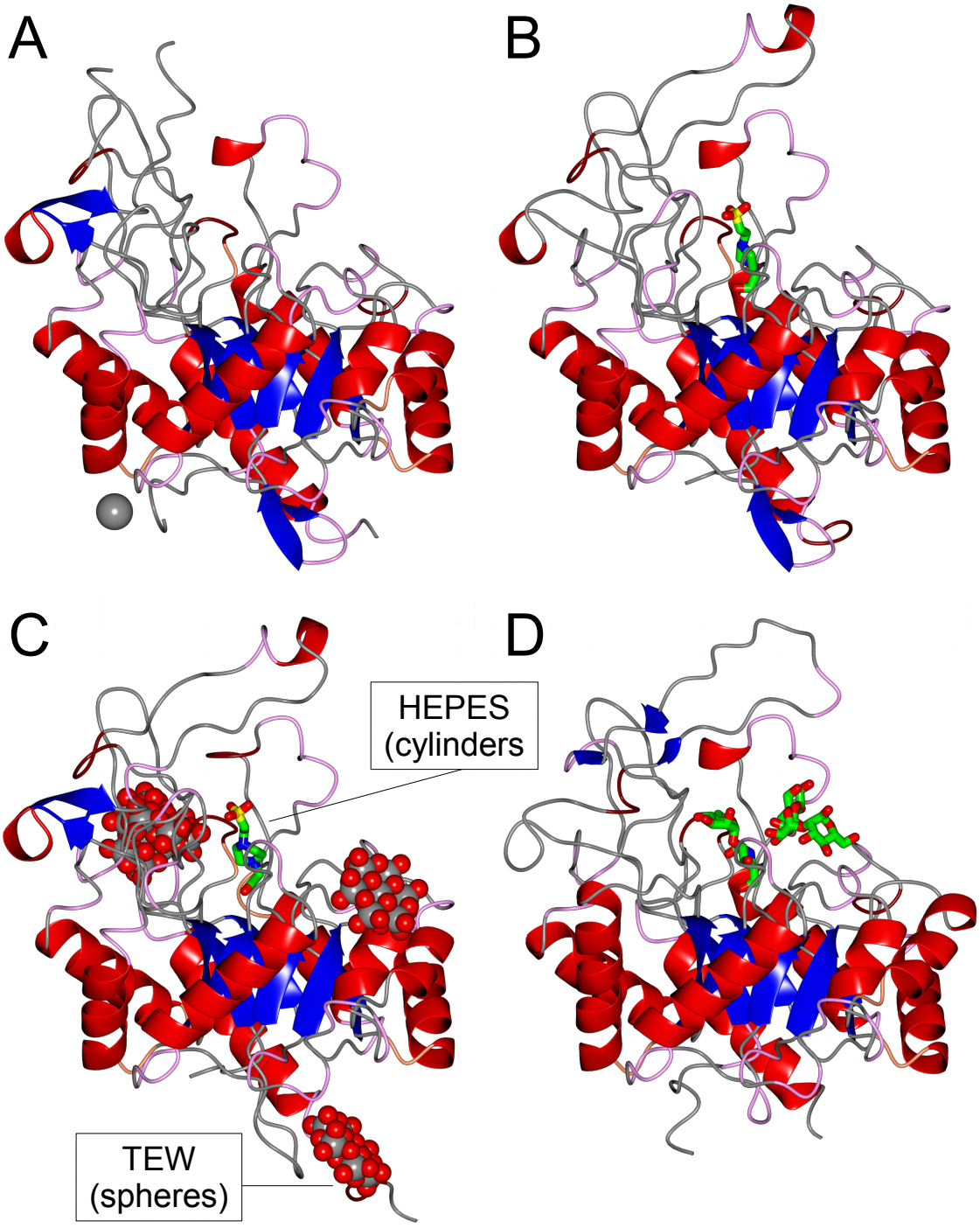


Figure 4.10: Comparison of the structures of GH99 proteins. (**A**) **Crystal form 1**. The Ni^{2+} ion is shown as a grey sphere. (**B**) Chain A of **crystal form 3** with HEPES shown as cylinders. (**C**) **Crystal form 4** with HEPES and TEW shown. (**D**) *Bx*GH99 with Man-NOE and α-1,2-mannobiose (PDB code **5M03**). Note the extended loop, not modelled in **crystal form 1** (**A**) because of disorder, appears to be ordered in **crystal form 3** (**B**), despite the lower resolution of the data collected from the crystal used to solve the structure of the protein forming **crystal form 3**. Assembled using ccp4mg.[204]

Only in loop areas significant differences can be found. The extended loop comprising residues 125 to 143 (*Bx*GH99 63 to 80, see Figure 4.11) differs in conformation between the two enzymes, but does not appear important in protein-substrate interactions. Another region that does not participate in interactions with the ligand, but has a different structure in these proteins is the short β-structure at residues 252–262 (Figure 4.11). In *Hs*MANEA it is a pair of antiparallel β-strands (distorted in **crystal form 4** due to the proximity of the additive TEW), while in *Bx*GH99 it is a 4-residue turn (aa 186–189). In **crystal form 5** the region could not be modelled. It does not form a part of the core fold, and can be seen at the bottom of the proteins in Figure 4.10**AB** and in Figure 4.11. The region has a highly basic charge and forms a part of the MANEA basic patch, which will be discussed in Section 4.4.5.

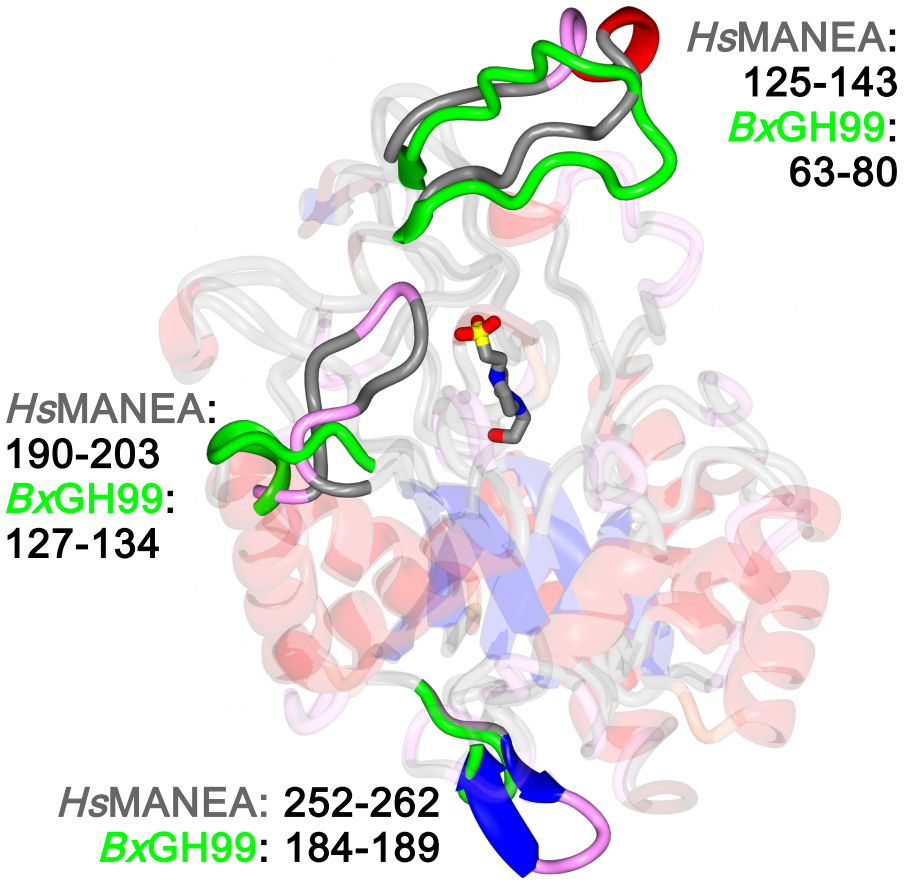


Figure 4.11: Structure of *Hs*MANEA **crystal form 3** chain A overlapped with the structure of *Bx*GH99 with ManNOE and α-1,2-mannobiose (PDB code **5M03**). The polypeptide chains are shown as well as HEPES which is in complex with *Hs*MANEA in **crystal form 3**. The parts of the protein which do not differ significantly are shown as transparent (and coloured by the secondary structure). The parts that differ are non-transparent. The mainchain trace in *Hs*MANEA is coloured according to its secondary structure (grey: none, red: helices, blue: strands, pink/tan/coral: turns) and the parts of *Bx*GH99 with different three dimensional structure are coloured green. Assembled in ccp4mg.[204]

Perhaps the most interesting feature is the conformation of *Hs*MANEA residues 191–201. This loop is absent from *Bx*GH99, where in its place there is a short sequence linking a core barrel β-strand to an α-helix (see Figure 4.11). In **crystal form 1**, this loop does not appear close to the ligand binding pocket. In both crystals forms with HEPES bound (**2** and **4**), the sulfo group seems to stabilize the loop in a position closer to the projected space where the –2 sugar would bind. It is probably the case that the protein co-crystallized with HEPES because it helped stabilize the conformation of this loop. D195 is poised to interact with the –2 sugar OH3 group. This residue is conserved in species that possess *endo*-α-1,2-mannosidases (see Figure 4.12), which also points to its probable importance in enzyme function. However, the presence of HEPES is inconvenient experimentally due to the –2 and –1 subsites being occupied by the buffer molecule. These structures explain why, at high concentrations, the buffer acts as an inhibitor of the enzyme.

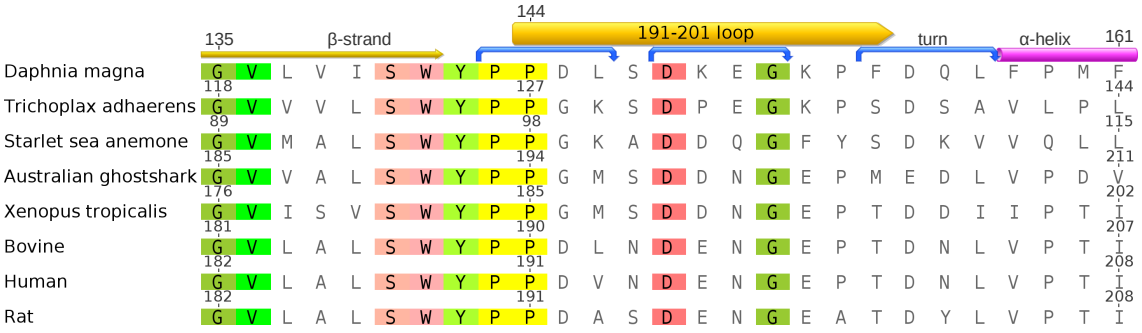


Figure 4.12: An excerpt of an alignment of endomannosidase proteins from various eukaryotes showing the 191–201 loop and neighbouring residues. Numbering in the original sequences is given above each sequence. Secondary structure features from the human enzyme are annotated on top of the sequences. Only the residues that are 100% conserved are highlighted, and the colour of the highlighting reflects the residue type.

The discussion above related to the overall 3-dimensional structure of *Hs*MANEA. However, of particular interest is the binding pose of *Hs*MANEA inhibitors and natural substrates/products. In the next Section, this topic will be discussed.

4.4.3 MANEA ligand binding

A structure of *Hs*MANEA E404Q in complex with α-1,2-Man-O1-α-methylmannose (**crystal form 5**) was fortuitously obtained from a co-crystallization experiment in which the protein was crystallized in the presence of a tenfold molar excess of its substrate: $GlcMan_3OMe$. The initial aim of the experiment was to see the mode of binding of the whole substrate – to obtain the so-called Michaelis complex. In the experimentally determined structure, the ligand is seen only in the +1/+2 sites with an occupancy of 0.75. Owing to the high quality of the data, the structure of the alternative conformation, in which water molecules were present in the active site at an occupancy of 0.25, could also be resolved. The ligand sits exactly in the position that was predicted from structural alignment with the *Bx*GH99 product complex (PDB code **6FWP**). The structural similarity goes beyond that: Y271, a residue homologous to *Bx*GH99 Y198, exists in two conformations. In the bacterial protein, this residue is disordered in structures where the +2 subsite is not fully occupied, and is in a stable conformation when it is occupied. As in the human protein structure the ligand is at 75% occupancy, what is observed in *Hs*MANEA seems to be the same effect. It is tempting to explain this remarkable similarity by convergent evolution, but it is not clear whether this is the reason for it.

Table 4.4: Data collection and model refinement statistics of *Hs*MANEA crystal structures with saccharides bound. *Values in parentheses represent the highest resolution shell. RMSD: root mean square difference.

Crystal form	**5**	**4**	**4**
*Hs*MANEA variant	E404Q	WT	WT
In complex with	α-1,2-Man-ManOMe	GlcIFG	GlcIFG + α-1,2-mannobiose
Data collection			
Space group	$P2_1$	$P6_2$	$P6_2$
Cell dimensions			
a, b, c (Å)	42.7, 81.7, 53.0	127.9 127.9 48.8	127.9 127.9 48.4
α, β, γ (°)	90, 92.93, 90	90, 90, 120	90, 90, 120
Resolution (Å)	52.92–1.10 (1.12–1.10)*	31.97–1.90 (1.94–1.90)	110.76–1.80 (1.84–1.80)
R_{merge}	0.054 (0.223)	0.203 (3.430)	0.148 (0.939)
R_{pim}	0.029 (0.150)	0.050 (0.843)	0.038 (0.237)
CC(1/2)	0.998 (0.944)	0.998 (0.546)	0.999 (0.782)
$<I / \sigma I>$	11.9 (3.3)	10.5 (1.0)	13.4 (2.9)
Completeness (%)	97.6 (71.3)	100.0 (99.8)	100.0 (100.0)
Redundancy	3.7 (2.6)	17.6 (17.4)	16.5 (16.6)
Refinement			
Resolution (Å)	52.98–1.10	110.73–1.90	110.76–1.80
No. reflections			
all / free	143263 / 7225	36184 / 1779	42104 / 2094
R_{work} / R_{free}	0.11 / 0.13	0.19 / 0.22	0.18 / 0.22
No. atoms			
Protein	3220	2961	2974
Ligand/ion	24/5	130	164
Water	491	113	137
B-factors (Å^2)			
Protein	15	32	26
Ligand / ion	19 / 21	66	47
Water	30	32	30
RMSD			
Bond lengths (Å)	0.014	0.013	0.014
Bond angles (°)	1.7	1.74	1.9
PDB code	–	–	–

+1/+2 subsites: A comparison with aligned structures of E404Q + α-1,2-Man-ManOMe and the product complex of *Bx*GH99 reveals an essentially identical arrangement of the conserved residues and the hydrogen bonds around the +1/+2 subsites (Figure 4.13**C**).

–1/–2 subsites: The aa 191–201 loop seems to be stabilized through the hydrogen bond that HEPES sulfo group forms with a backbone nitrogen of E196 (Figure 4.13**D**). The

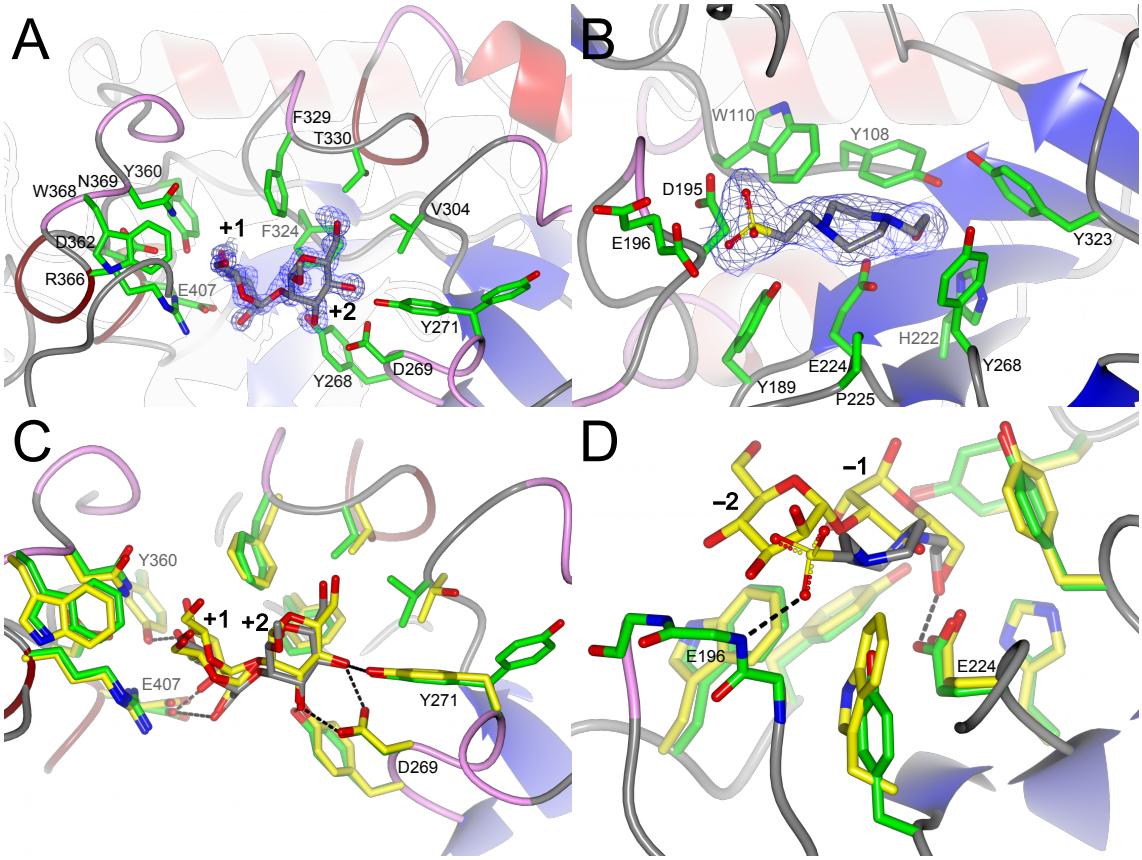


Figure 4.13: *Hs*MANEA ligand binding pocket. (**A**) α-1,2-Man-ManOMe in the +1/+2 subsites of the E404Q mutant (**crystal form 5**). 2mF_o-DF_c electron density map around the ligand shown at 0.5 e$^-$/Å^3. (**B**) HEPES in the active site of the WT co-crystallized with TEW (**crystal form 4**). 2mF_o-DF_c electron density map around the molecule is contoured at 0.8 e$^-$/Å^3. Residues making close contacts to the bound molecules are shown and labelled. (**C**) Comparison of the binding pose of α-1,2-Man-ManOMe (grey) in complex with *Hs*MANEA E404Q (green) with the *Bx*GH99 product complex structure (with α-1,2-mannobiose and α-1,3-mannobiose, only α-1,2-mannobiose is shown (in yellow)). (**D**) Comparison of the binding pose of HEPES (grey) in complex with *Hs*MANEA WT (green) with the *Bx*GH99 product complex structure (with α-1,2-mannobiose and α-1,3-mannobiose, only α-1,2-mannobiose is shown (in yellow)). In (**CD**), hydrogen bonds for MANEA are shown, and only the residues forming them with the compounds bound in the pocket are labelled. Subsite nomenclature as in Ref. 304. Assembled in ccp4mg.[204]

hydroxyethyl group of HEPES mimics the C5-C6-OH6 atoms of the –1 sugar, forming a hydrogen bond with E224 and a hydrophobic interaction with Y195.

Soaking the crystals of **crystal form 4** with GlcIFG and subsequently α-1,2-mannobiose revealed the binding poses of ligands in subsites –2 to +2 (HEPES was outcompeted by GlcIFG in the –2/–1 subsites). The binding position and conformation of both compounds almost exactly mimics their position when bound to *Bx*GH99 (see Figure 4.14). Like in the structure of *Bx*GH99 with GlcIFG and α-1,2-mannobiose (PDB code **4AD4**), isofagomine binds in a $^{4}\boldsymbol{C}_{\mathbf{1}}$ conformation (see Figure 4.14**AC**). The crystallization condition of this crystal form contained 100 mM HEPES at pH 7.5. The protein most likely cocrystallized with HEPES, which primed the 191–200 loop for reception of the inhibitor into the –2/–1 subsites. The loop envelops the inhibitor: residue E196 participates in stabilizing the –2 glucose residue by making close (~3.05 Å) polar contacts with its OH2. Side chains of D195 and N197 form hydrogen bonds with the OH3 group. The conserved D195 Oδ…OH3 interaction is particularly strong (2.55 Å). Y189, by homology to *Bx*GH99 W126 thought to be forming a hydrogen bond with the –2 Glc OH2, is actually coordinating one of the two water molecules that form hydrogen bonds with the OH2 group (Figure 4.14**A**). In *Bx*GH99, this water molecule cannot be seen in the electron density map – it would cause a steric clash with the tryptophan side chain. A tantalizing proposition is that the protein is employing the flexible loop close to the –2 sugar OH3 in order to allow interactions with precursor glycans with two or three terminal glucose molecules still attached, the first of which would be bound to the –2 Glc via an α-1,3-glycosidic bond. It has been shown that rat MANEA can cleave $Glc_{3\text{-}1}Man$ from $Glc_{3\text{-}1}Man_9GlcNAc_2$, although the activity towards the longer glycans is only 7–8% of that towards $Glc_1Man_9GlcNAc_2$.[146] A different situation, however, is more likely. The loop structure is not present in the bacterial enzymes, whose primary substrate is the yeast mannan.[207] The GH99 endomannanase takes care of cleaving the terminal α-1,3-mannobiose from the polysaccharide, which exposes the backbone. It is possible mannans with these branches extended exist. In such a case, selection pressure in the bacterial enzyme would then act to minimize any steric clashes at the –2 subsite, so that it can be cleaved efficiently with no regard to additional sugars being present. Conversely, the endomannosidase might have evolved specifically to *prevent* the longer glycans from binding, and the 7–8% activity

might represent "leakage". In this scenario, the function of the enzyme would be to discriminate against glycans that did not undergo the cleavage by α-glucosidase I and the first cleavage by α-glucosidase II, which usually are fast processes.[81,86] Cell culture studies demonstrated that the predominant product of the enzyme in glucosidase II-deficient mouse cells is Glc_2Man.[150] In normal conditions the two first glucoses are not decorating the glycan, but the last glucose is much more likely to be carried over from the ER due to the presence of UGGT. Therefore, it would make sense for the enzyme to evolve a high affinity towards the terminal GlcMan disaccharide only, and not the longer structures. Kinetic studies are in support of this: bovine MANEA does not cleave glycans that have more than one glucose on the A branch. Figure 4.14**C** demonstrates the different observed conformations of the loop in question. A "closed" conformation (green) is visible in structures with HEPES or GlcIFG in the –2/–1 subsites (crystal forms **3** and **4**). Two conformations of the 191–201 loop can be seen in the high-resolution structure of the E404Q mutant (**crystal form 5**, yellow) and one in the structure with Ni^{2+} ion (**crystal form 1**, red). Exactly how the loop behaves in the presence of a –3 ligand remains unknown.

The +1 mannose in the structure of the tertiary complex is visible in two conformations at 50% occupancy each. This is analogous to how it appears in high resolution structures of *Bx*GH99, as shown in Chapter 2. This speaks to an extremely high similarity between the binding site topography and the mode of action of bacterial and mammalian GH99 endomannosidases.

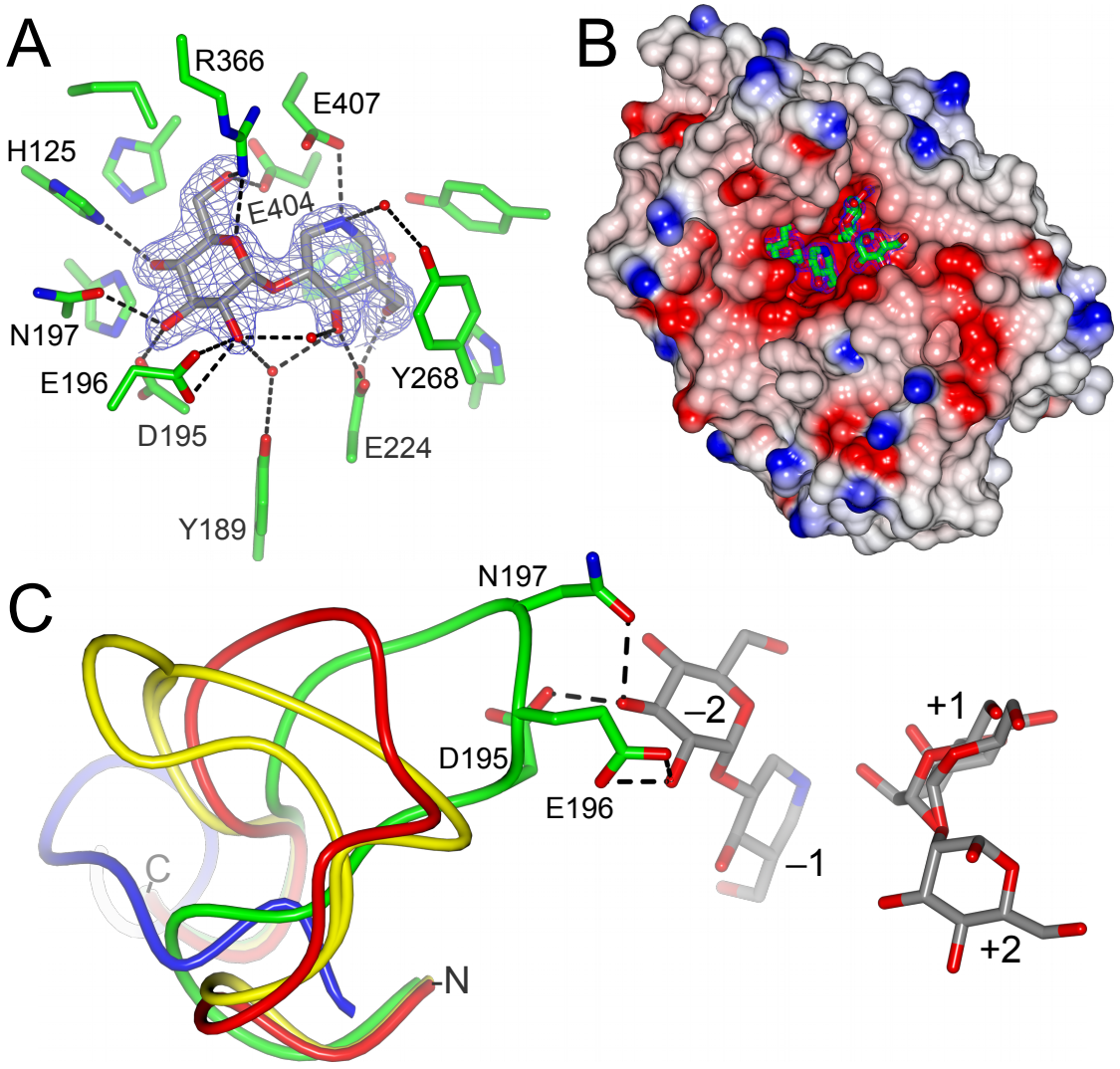


Figure 4.14: *Hs*MANEA binding of GlcIFG, α-1,2-mannobiose and conformational changes of the 191–201 loop. (**A**) Side chains of residues around the –2/–1 subsites and GlcIFG electron density ($2mF_o$-DF_c electron density map around the inhibitor shown at 0.5 $e^-/Å^3$). Close polar contacts and hydrogen bonds are shown as dashed lines. (**B**) Electrostatic surface of *Hs*MANEA with GlcIFG and α-1,2-mannobiose binding pose shown (**crystal form 4**). (**C**) Worm representation of residues 189–203 in different *Hs*MANEA models and 127–135 in *Bx*GH99 PDB code **4AD4**. Green: WT in complex with GlcIFG and α-1,2-mannobiose (**crystal form 4**); yellow: E404Q in complex with α-1,2-Man-ManOMe (**crystal form 5**); red: WT with Ni^{2+} (**crystal form 1**); blue: PDB code **4AD4**. Subsite nomenclature as above. Assembled in ccp4mg.[204]

4.4.4 Analysis of MANEA mutations in context of the structure

The models of the human MANEA protein permit answering long-standing questions about the structure of GH99 enzymes. For example, CHO cells have been known not to possess any detectable endomannosidase activity[153,168] and this is due to a mutation in their MANEA protein, described in a 2007 article by Torossi *et al*.[197] This was discussed in Section 1.4.5.4: see Figure 1.15 on page 37 and the associated text. Two residues were selected for studying the cause of this lack of activity: residue 177 and 188. In the human MANEA and MANEA proteins encoded in the genomes of many other mammals, these residues are R177 and W188. The wild-type Chinese hamster MANEA features an unusual cysteine residue in position 177, and MANEA in CHO cells contains a W188C mutation. While R177 is not fully conserved – being a cysteine in WT Chinese hamster and a histidine in *Otolemur garnettii* (northern greater galago) – W188 is conserved across the animal kingdom, as well as in the bacterial GH99 enzymes. It was found[197] that the W188C mutation causes the protein to be mislocalized away from the Golgi and to the ER. This is mediated by the formation of an intramolecular disulfide bridge. The authors propose two hypotheses as to why this mutation has such an effect: either R188 forms hydrogen bonds with the substrate, or it participates in the stabilization of tertiary structure. The MANEA crystal structure points to the second explanation – with some truth in the first. R188 is a residue that plays many roles: it is a part of a core β-strand and stabilizes the structure against the outside α-helix. It also directly precedes Y189, which coordinates a water molecule that forms a hydrogen bond with the –2 sugar OH2 (see Figure 4.14**A**). With R188 mutated to C, both the protein tertiary structure and its ligand binding propensity would be disturbed, and this effect would probably be exacerbated by the formation of the disulfide bridge. The appearance of this enzyme variant raises questions about the utility of not having endomannosidase activity for CHO cell survival: the point mutation at this locus seems to be ideal for compromising the endomannosidase, and might have been selected for. An interesting experiment would be to see whether CHO cells with restored MANEA activity are indeed outcompeted by those lacking it in culture. An answer to this question would have implications for research into immortalized cell lines, which translates to insights into biology of cancer.

Kukushkin *et al.* speculated that an S to K substitution at position 227 may be respon-

sible for the more narrow specificity of bovine MANEA towards monoglucosylated *N*-glycans, as opposed to a more relaxed specificity of human MANEA. They thought it might occlude access of longer structures to the active site.[169] It is indeed quite possible that the lysine makes an H-bond with the spatially close 191–201 loop E196 side chain or N194 main chain oxygen, which stabilizes the loop. This question can only be resolved by further structural studies.

A number of somatic point mutations were discovered in the endomannosidase gene and deposited to publicly available databases. These are summarized in Table 4.5 and with the structural information at hand, their functional impact on the protein can be assessed. Of the 89 considered mutations (82 excluding the metastases), there were five found in more than one sample (at residues 57, 111, 248, 319 and 366). Only two of these – K57E and G319E – were found in multiple samples that were not clones of the same tumour. R→C and Y→C variants seem to be prevalent, making up 11% of non-metastatic missense mutations. This is notable because MANEA proteins, in general, have low cysteine content (usually 1, rarely 2 in the lumen-resident region). As discussed earlier, a cysteine in the wrong place can lead to formation of a spurious disulfide bridge, compromising the activity of the enzyme. However this might not be significant since the total content of R and Y residues in the MANEA sequence is 12% and the background mutation rate for the protein is not known (11% and 12% are numbers that are not far apart). The catalytic domain makes up 357 of the 462 MANEA residues (77% of the sequence length; residues 98-454) and 81% (54/67) of the mutations burden affects the catalytic domain. As such, there seems to be no preference for mutations to occur within this region. 50% of the mutations within the catalytic domain affected residues that are conserved between MANEA, MANEAL and bacterial GH99 endomannanases.

Table 4.5: Somatic mutations found in the human MANEA. Only the protein variants are shown. Silent mutations were omitted. Variants of particular interest written in bold, the mutations resulting in an additional cysteine residue emphasized in italics. If multiple synonymous missense mutations were observed, the number in parentheses is the number of patients in whom they occurred. Data taken from the COSMIC database.[305,306]

Mutation	Count	Mutation	Count	Mutation	Count
F4V	1	V219F	1	**D318H**	**1**
M23V	1	H222Y	1	**D318Y**	**1**
G37E	1	I223R	1	**G319E**	**2 (2)**
H49R	1	R229*	1	Y323*	1
R51Q	1	K244N	1	N327S	1
K57E	**3 (3)**	H248Y	4 (1)	*Y331C*	*1*
N58I	1	A250V	1	S334L	1
D60E	1	Y252*	1	C345Ffs*6	1
D65V	1	A261V	1	*Y348C*	*1*
E70*	1	L262V	1	G359A	1
K74N	1	S270F	1	T363I	1
K77E	1	T273A	1	***R366C***	***3 (1)***
S78I	1	P275H	1	**R366H**	**1**
M83V	1	A279T	1	*W368C*	*1*
H104N	1	L281M	1	R376*	1
F106S	1	T283I	1	R376Q	1
Y111C	*3 (1)*	S287F	1	*Y381C*	*1*
N113K	1	R288Q	1	R393H	1
P114L	1	N292Y	1	**E407A**	**1**
K137R	1	P294L	1	A414Lfs*12	1
G142E	1	L302P	1	V415L	1
Y155C	*1*	E306*	1	P416S	1
P156S	1	K307N	1	*R428C*	*1*
M173V	1	H308R	1	L434F	1
M176I	1	K309T	1	R440H	1
R177H	1	S315I	1	A454V	1
K218N	1				

Only one MANEA missense mutation, K57E, was found in tumour samples from three separate patients. This lysine residue is a part of the stem domain of MANEA, which is a spacer between the short N-terminal transmembrane helix and the catalytic domain. It is not predicted to be acetylated[307] but is predicted to be ubiquitinylated by UbiSite[288] and ESA-UbiSite[289] servers. As it is not within the sequence of the solved catalytic domain structure, the model cannot inform the assessment of the mutation. As K57 is a predicted ubiquitinylation site, it suggests there might be some selection in tumours towards survival of the protein – ubiquitinylation is a marker for degradation by the proteasome. The G319E variant is found in the MANEA sequence from

two patients. G319 is a residue conserved across vertebrate and bacterial GH99s and a part of the core β-structure: it can be said with a high degree of confidence that this change would prevent the protein from folding properly. D318H and D318Y, each found in one patient, also affects a universally conserved, neighbouring residue and would have a similar effect. Mutations R366C, R366H and E407A change residues that form hydrogen bonds with the substrate/product. E407 is the acid/base residue and an E→A mutation of a homologous glutamate in *Bt*GH99 (E332A) resulted in activity being reduced by approximately a factor of 50.[148] Variant Y111C affects a tyrosine residue that is conserved and neighbours Y110 which forms hydrophobic interactions with the pyranose rings of the –2 and –1 sugars – it is likely to disrupt these interactions.

No definite conclusions can be drawn from the mutational analysis. One one hand, the MANEA variant present in CHO cells, and the relative multiplicity of X→C amino acid changes suggest the immortalized cells that lack the endomannosidase function might have a selective advantage over other phenotypes. On the other hand, the most commonly found mutation was K57E. If the ubiquitinylation prediction software is correct, it might be a site that – if mutated – could decrease the propensity of MANEA to be degraded by the proteasome. Perhaps the individual characteristics of each tumour or cell line favour different levels of MANEA activity. It should be also kept in mind that somatic mammalian cells are diploid and the other allele could compensate for the mutation.

4.4.5 MANEA and MANEAL structural features

As MANEA is the first eukaryotic GH99 to have been solved, the possibility of looking at related proteins was opened. MANEAL is the obvious candidate for such investigation. The MANEAL catalytic domain modelled on the MANEA template revealed a highly similar protein, with structural alignment containing no gaps, as in the primary sequence alignment (Figure 4.3). Only one model was generated and it had a high confidence score (1.76 on a scale from -5 to 2) and a low RMSD (3.1 ± 2.2 Å). By visual examination, the modelling of the chains is likely not correct, especially the residues in the active site, which are likely to assume a low-probability rotamer to which the modelling software is blind. The backbone of the protein, however, is

identical, and the differences in the primary sequence seem to stem from a series of functionally neutral changes, with one exception: Q226 (mentioned in Section 4.2), which is aligned with MANEA E224 (as expected). The interaction that E224 makes with GlcIFG can be seen in Figure 4.14**A** on page 115.

A positively charged patch was found upon examination of the electrostatic surface of MANEA. Such patches have been found in proteins that interact with DNA, but also in proteins that do not form such interactions.[308] The presence of positively charged patches is correlated with difficulty in production of soluble proteins,[309] which might explain the need for chaperone in order to produce active MANEA in bacteria. The patch can be seen for both the structure of the WT (Figure 4.15**A**) and the E404Q mutant (Figure 4.15**B**). It is more pronounced in the WT structure due to a loop containing basic residues being present in the model. Interestingly, the MANEAL I-TASSER homology model does not seem to have such a patch (Figure 4.15**C**), nor does *Bx*GH99 (**D**).

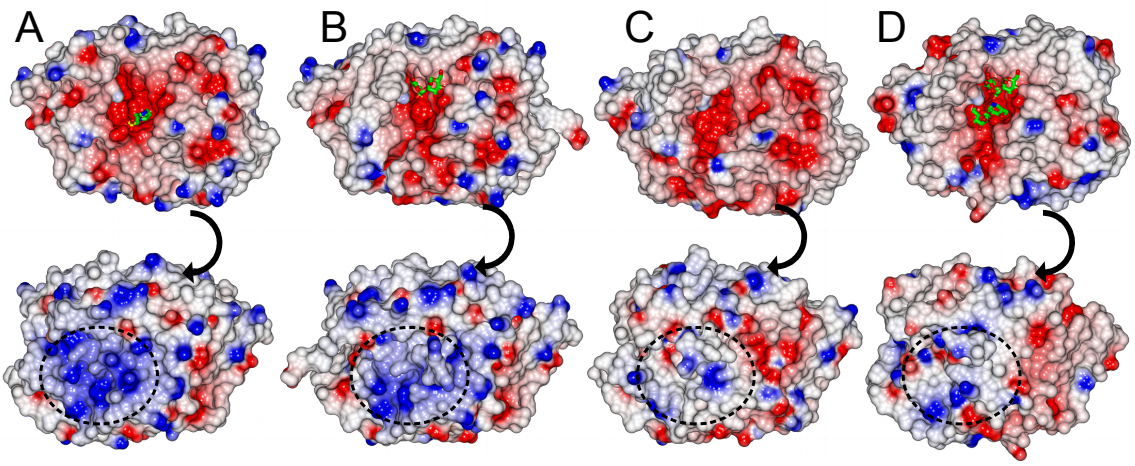


Figure 4.15: Electrostatic surface of the models of various GH99 proteins. Top: view from the active site side, bottom: view from the back of the active site (180° rotation about the Y axis). Red signifies negative charge, and blue positive charge. The positively charge patch is circled. (**A**) **Crystal form 3** form with HEPES visible; (**B**) **crystal form 5** with α-1,2-Man-ManOMe visible; (**C**) *Hs*MANEAL homology model (modelled after *Hs*MANEA **crystal form 2** chain A); (**D**) *Bx*GH99 with all four subsites occupied (by ManNOE and α-1,2-mannobiose, PDB code **5M03**). Surface charges calculated using ccp4mg[204]. Charge scale: ≤-0.5: red, >-0.5 and <0: shades of red, 0: white, >0 and <0.5: shades of blue, ≥0.5: blue

4.5 Conclusions

The solution of the crystal structure of human *endo*-α-mannosidase represents a major step in understanding of this enzyme. The remarkable similarity in the binding of the sugars around the scissile bond validates the mechanistic conclusions drawn from studying the bacterial homolog of the enzyme. Description of subsites –2 to +2 opened new areas of investigation of the enzyme. Three residues surrounding the ligand around the –2 subsite were discovered, and they form a part of a loop that is not present in bacteria. Two of these residues are highly conserved in eukaryotic endomannosidases. In the process, a method for production of high amounts of human MANEA in bacteria was refined, and the boundaries of the catalytic domain were established. Analysis of somatic mutations revealed that the role of MANEA in cancer is not easy to predict, and probably depends on the individual characteristics of each condition. As some mutations that were found are likely to make cells produce more of the enzyme than in healthy cells, there is therapeutic potential for targeted inhibition of the enzyme. Answering the questions of when, how and if to intervene with endomannosidase activity would require more research, but the protocols developed here may serve to make the process easier. It is also a step forward towards a structural picture of the whole *N*-glycosylation pathway. Structures of numerous eukaryotic ER and Golgi glucosidases and mannosidases from various eukaryotic sources are available, and the endomannosidase represented a gap in our knowledge which is now filled.

The human GH99 fold is reasonably stable. The 191–201 loop becoming ordered when the –2/–1 disaccharide or HEPES (and probably HEPPS, as its structure and effect on the enzyme are similar) binds is the largest conformational change that happens in the protein. An interesting avenue of research might be studying the changes in the loop using NMR. This would require assigning the protein, which with an enzyme of this length is likely to be cumbersome (^{2}H, ^{13}C and ^{15}N triply labelled protein is necessary). It would, however, be a good method for uncovering the observed specificity differences that occur between endomannosidases from different mammals (bovine MANEA is specific towards monoglucosylated glycans only, as discussed above).[169] A conformational change might be observed in rat/human MANEA upon binding of the longer substrates, but not in bovine MANEA, or it might require

a higher concentration of the ligand. Understanding of these processes in a larger sample of endomannosidases will uncover the evolutionary identity of this enzyme.

The path to the solution of *Hs*MANEA structure with saccharides occupying four subsites reveals that the use of a weakly binding ligand may be a good strategy for obtaining multiple crystal forms. HEPES was an ideal ligand for this: it binds to the protein with low affinity and stabilizes an otherwise disordered loop, priming it to receive the inhibitor. It is well known that using a molecule which binds to the protein often stabilizes its conformation. Researchers take advantage of this effect by performing thermal shift assays with different molecules present as a way to screen for potentially binding compounds. A higher melting temperature in comparison to that of the protein alone often means that the compound is a ligand. If a compound of interest is in limited supply, preventing co-crystallization, and no suitable crystal form was obtained with the protein alone, trying different buffer molecules can be helpful. Glycoside hydrolases, glycosyltransferases and lectins are especially suitable for this, as their binding pocket is tuned to capture ligands with multiple hydroxyl groups, and biologically useful aqueous buffers often contain those. Additionally, only by the use of a polyoxometalate TEW as an additive, a crystal form suitable for studies of the four subsites, from –2 to +2, was obtained. There are only four structures with this compound in the PDB, even though it has been available for some time already (the first structure with it was published in 2014).[292]

Production of soluble, stable protein enabled the characterization of MANEA ligand binding thermodynamics and studies on its reaction kinetics. In the following chapter, which will close this thesis, data pertaining to these topics, as well as cell culture experiments, will be presented. In addition, efforts with an aim to produce the MANEA sister protein – MANEAL – will be described.

Chapter 5

Kinetic and biochemical characterization of human GH99s

5.1 Abstract

The human genome contains genes encoding two proteins predicted to have a GH99 domain: *MANEA* and *MANEAL*. Previously reported research concentrated on the *MANEA* gene and its protein product. Here, an attempt at expression of a truncated version of the *MANEAL* gene is presented. The catalytic domains of MANEA and MANEAL were produced, but only MANEA showed activity on the native substrate $GlcMan_9GlcNAc_2$. The yield of MANEAL was extremely poor, the protein was not stable in the buffer used for MANEA and likely was misfolded. The binding constants of isofagomine-type GH99 inhibitors were measured. GlcIFG binds to MANEA with a K_D of 19.6 ± 5.6 nM and binding of ManIFG is roughly 10 times weaker. A novel inhibitor, "blocked" GlcIFG, binds with a K_D of 929 ± 52 nM. The compound is predicted to be more stable than GlcIFG *in vivo*. TheK_D values obtained for the inhibitors support previous conclusions about the human enzyme derived from the data obtained from experiments with the bacterial enzyme, namely the preference of the human enzyme towards Glc-substituted ligands. Michaelis-Menten kinetics of MANEA activity on a tetrasaccharide substrate analogue $GlcMan_3OMe$ were established. Alternative approaches to MANEAL production have to be developed in order to characterize this enzyme. It is possible that by omission of this enzyme, crucial aspects in the story of the endomannosidase pathway are not understood.

5.2 Introduction

Production of the other GH99 protein encoded by the human genome, MANEAL, has never been reported. The success in MANEA purification prompted trying a similar approach for MANEAL. The activity of this protein is unknown, but as shown in the previous chapter in Figure 4.2, it is localized to the Golgi and therefore is likely to play the same role as MANEA but in different tissues or cell types. The expression of *MANEAL* mRNA is higher in the tissue depleted in *MANEA* – the brain (Figure 4.1). The protein sequence identity of MANEA and MANEAL is 57% and there is one consistent difference between these proteins: MANEA E224 is Q226 in MANEAL. This chapter will include a discussion of the results of the protein prodution efforts, as well as characterization of the activity of both enzymes.

As discussed in Chapter 1, inhibition of endomannosidase is an attractive, potentially general antiviral therapy. The enzyme offers a pathway of *N*-glycan maturation in regimes of glucosidase inhibition. Results from the Terry Butters group (from Dr Dominic Alonzi) suggested that the GH99 inhibitor GlcIFG was antiviral against bovine viral diarrhea virus (BVDV). They are summarized in Figure 5.1. With increasing GlcIFG concentration, the infectivity of BVDV was decreasing. The inhibitor had an effect on the glycosylation of the viral envelope protein that was distinctly different from that of an ER glucosidase II inhibitor NAP-DNJ. When used together, these compounds had an additive effect on decreasing infectivity. These data prompted an investigation into the effect on the infectivity of other viruses that affect humans, whose results will be presented in this chapter.

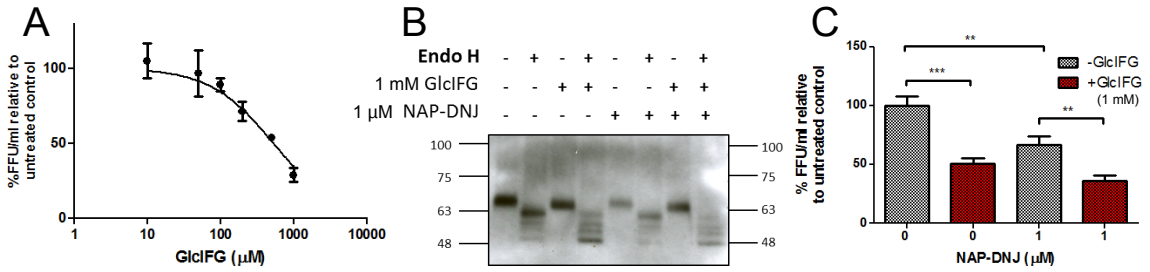


Figure 5.1: Results of BVDV assay in cell line MDBK. (**A**) Percentage of FFU (focus-forming units) relative to untreated cells at different concentrations of GlcIFG at a MOI (multiplicity of infection) of 1. (**B**) Effect of endomannosidase inhibition (GlcIFG) and ER glucosidase II inhibition (NAP-DNJ) on susceptibility of glycans on the BVDV E1/E2 protein to cleavage by endoH. (**C**) Effects of NAP-DNJ and GlcIFG on BVDV infectivity are additive. Figures from the Terry Butters' group.

5.3 Materials and methods

5.3.1 Gene subcloning and protein production

A CDS (coding DNA sequence) encoding MANEAL which lacked the 95 N-terminal amino acids (the stem domain) was optimized for expression in *Escherichia coli* and ordered from Genscript Biotech. The gene was subcloned into pET28a(+) and pCold-I vectors using restriction sites *NdeI* and *XhoI*, similarly to *MANEA*. The gene was co-expressed with *gro*EL/*gro*ES from the pCold-I construct and the protein was purified from BL21(DE3) cells using the same protocol as *Hs*MANEA (Section 4.3.3). As the isoelectric point of MANEAL is considerably lower (6.9 vs. 9.2 for MANEA), the pH of the buffer used for cation exchange was modified from 7.0 to 6.0. An additional gel filtration step using an S75 16/600 HiLoad Superdex column was performed in 50 mM potassium phosphate pH 7.0, 185 mM KCl in order to remove residual chaperone. The protein migrated on the gel as a single band corresponding to a Mw of 42–43 kDa (the actual Mw of the purified protein was 43963 Da). The yield was extremely low (0.15–0.2 mg protein per litre of culture). The buffer used to store *Hs*MANEAL was 50 mM potassium phosphate pH 6.0, 250 mM KCl.

5.3.2 Chemicals and enzymology

MANEA activity on its natural substrate was confirmed by performing an extended assay. This method was chosen in order to make the results of the assay comparable with the results of a similar assay performed by Dr Andrew Thompson with *Bt*GH99 and described in his PhD Thesis.[310] The protein, at a concentration of 100 nM, was incubated with 24.5 µM $GlcMan_8GlcNAc_2$ oligosaccharide (DEXTRA) (total substrate amount: 1 µg) for 19 h at 4 °C, 14 h at RT, then 12 h 30 min at 37 °C. The buffer used was 50 mM potassium phosphate pH 7.0, 185 mM KCl. A negative control containing the oligosaccharide but no enzyme was subjected to the same incubation process. The total volume of each sample was 20 µl. For MS analysis the enzyme was removed using an ion exchange column and the resultant carbohydrate samples were dried. Permethylation of the carbohydrates was achieved using a procedure described in Ref. 311. The samples were analysed using MALDI-MS. Enzyme removal, permethylation, MS data collection and analysis were performed by York Technology Facility staff.

Comparison of the activity of *Hs*MANEA and MANEAL was done using a similar incubation procedure. In this case, all reactions were performed in MANEAL storage buffer: 50 mM potassium phosphate pH 6.0, 250 mM KCl. 4 µl of reaction mixture containing 500 nM enzyme or no enzyme control and 1 µg $GlcMan_8GlcNAc_2$ were incubated for 67 h at 5 °C, 12 h at RT and 12 h 30 min at 37 °C. The contents of the resulting mixture were analyzed using thin layer chromatography (TLC). The samples were transferred in 1 µl droplets onto SUPELCO silica gel on TLC Al foils (SIGMA-ALDRICH). Each droplet was allowed to dry before adding the next droplet. β-1,4-Xylobiose was used as a standard indicator of Mw (282.24 g mol^{-1}). TLC was performed in a preequilibrated chamber using 50% v/v *N*-butanol, 25% v/v acetic acid, 25% v/v H_2O as the running buffer. When the buffer reached 75% of the plate height, the plate was taken out of the chamber, dried and put back again. The buffer was allowed to reach 75% of the plate height again, at which point the migration was terminated. After drying, the samples were visualized by applying a staining solution (3% v/v sulfuric acid, 75% v/v ethanol, 0.1% orcinol monohydrate). The spots appeared after the plate was dried and subsequently heated on a hot plate (70–100 °C).

Blocked GlcIFG (BGlcIFG) was synthesized in Spencer J. Williams' laboratory at the University of Melbourne, Parkville, Australia by Dr Zalihe Hakki and $GlcMan_3OMe$ by Pearl Fernandes from the same group. The structures of these compunds are shown in Figure 5.2. Endomannosidase activity assay was performed by a linked assay method, published earlier.[159,207] The assay involves the use of BT3990, a GH92 α-mannosidase prepared by Dr Michael Suits, which specifically cleaves the α-1,2-mannosidic bond in the reaction product α-1,2-Man-ManOMe. The non-methylated mannose residue was then detected using D-Mannose/D-Fructose/D-Glucose Assay kit (Megazyme, Inc.). The samples were pre-heated to 37 °C before adding the enzymes (GH99 to a final concentration of 1 µM and GH92 to a final concentration of 100 nM). The initial reaction rate was measured at 7–9 min after addition of the enzymes as the rate of increase of absorbance at 340 nm. All samples were kept in quartz cuvettes. Abs_{340} (absorbance at 340 nm) was directly proportional to the amount of mannose present in the solution: 160.8 µM mannose per 1 unit of absorbance. The reactions were performed in 50 mM potassium phosphate pH 7.0, 50 mM KCl. K_M and k_{cat} values were calculated using a non-linear curve fit (Michaelis-Menten fit) in Origin 9. Additional reactions with mannose standards were performed in order to standarize the substrate concentration.

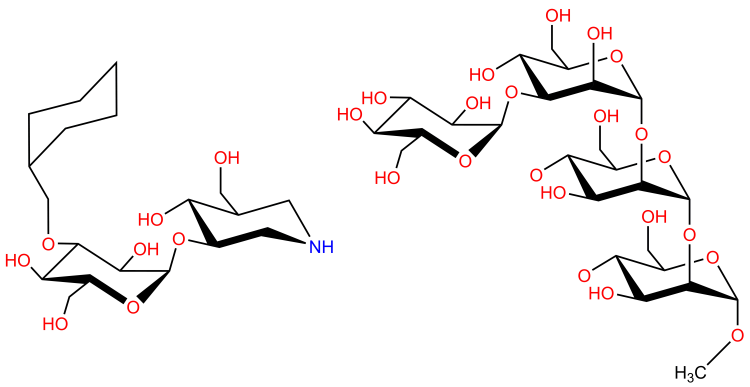


Figure 5.2: Compounds used for studies presented in this chapter. Left: methylcyclohexanyl-3-α-Glc-1,3-isofagomine (BGlcIFG), right: α-Glc-1,3-α-Man-1,2-α-Man-1,2-α-1-methyl-Man ($GlcMan_3OMe$).

5.3.3 ITC

Isothermal titration calorimetry was done in 50 mM potassium phosphate pH 7.0, 50 mM KCl in the case of *Hs*MANEA and 25 mM HEPES pH 7.0, 100 mM NaCl for *Bt*GH99. Protein and ligand concentration was varied in order to optimize the c-value. Details of the ITC hardware and software are explained in Section 3.3.4. The ITC experiment in the presence of α-1,2-mannobiose were performed with 44 µM of *Hs*MANEA in cell, 608 µM of GlcIFG in syringe and 1 mM of α-1,2-mannobiose in both. The details of protein concentration in the calorimeter cell and the inhibitor concentration in the syringe are explained in the discussion. The protein concentration in the sample was double-checked by unfolding the protein in 6 M guanidinium hydrochloride to expose all aromatic residues and measuring Abs_{280} using an Eppendorf BioPhotometer. The extinction coefficient of *Hs*MANEAL used for calculations of protein concentration was 95230 M^{-1} cm^{-1}.

5.3.4 Protein thermal unfolding

Thermal unfolding was done using an Agilent Technologies Mx3005P qPCR system. The proteins were kept in their respective phosphate storage buffers. The final concentration of *Hs*MANEA was 1 mg ml^{-1} and of *Hs*MANEAL was 0.83 mg ml^{-1}. Temperature was increased from 25 to 95 °C in 1 degree increments of 30 s each. The reporter molecule was SYPRO Orange (fluorescence measured at 570 nm). The re-

porter was used at a final concentration of 2.5 × (1/2000 of the 5000 × concentrated stock – this is the concentration at which it is sold). The measurements were done in triplicate with single buffer controls.

5.4 Viral infectivity assay

The Level III procedures in the hepatitis C (HepC) activity assay and the immunostaining were performed at the University of Birmingham by Scott Davies and Dr Zania Stamataki. Huh7 cells (human hepatocyte cellular carcinoma cell line) were grown in 96-well plates for 2 d. After this, the media was exchanged into a 2 × solution of the tested GH99 inhibitor (GlcIFG, GlcDMJ, BGlcIFG or Glc-β-aziridine) with or without a 2 × solution of *N*B-DNJ, in which the cells were incubated for 1 h. The final concentration of the GH99 inhibitor was varied (using serial dilutions) from 2 mM to 3.9 μM. The final concentration of *N*B-DNJ was varied from 8 mM to 500 μM depending on the experiment (it is discussed and presented in the Results and discussion subsection 5.5.4). Then, 50 μl media containing HepC was added to each well, diluting the GH99 inhibitor and *N*B-DNJ to the desired concentration. Experiments were done in triplicate for samples with *N*B-DNJ and untreated controls, and in single samples for GH99 inhibitors only. The infected cells were then grown for 2 d and killed using methanol. This inactivated the virus and made the cells safe to handle outside the Level III facility. The plates were then stained with an anti-NS5A* antibody and a secondary antibody conjugated to a red-fluorescent probe. They were then transported to York, where the cells were stained with the nuclear stain DAPI (4 ,6-diamidino-2-phenylindole). After staining, all of the wells containing relevant samples were imaged in the red and blue channel using a Zeiss LSM 780 microscope with a fully open pinhole. The image data were processed using TissueQuest software (TissueGnostics imaging solutions). About 60% of each image was used, and areas with damaged cells were masked from the analysis. Cells were counted using the DAPI staining and the infected cells by anti-NS5A staining.

*NS5A (non-structural protein 5A) is a zinc-binding protein essential for HCV (hepatitis C virus) RNA replication and virion morphogenesis. The protein exists as a number of phosphorylated species. See Ref. 312 for a more detailed characterization of the protein.

5.5 Results and discussion

5.5.1 *Hs*MANEAL gene expression, protein purification and unfolding

The truncated *Hs*MANEAL was produced in and purified from *E. coli*, but the yield was very low and most of the protein was in the insoluble fraction, even when co-produced with the GroEL chaperone. The protein seemed to purify in a similar way to *Hs*MANEA: two peaks were observed when eluting from the nickel affinity column with an imidazole gradient with a higher concentration of GroEL than *Hs*MANEAL in the first peak, and three peaks during cation exchange chromatography on an SP column (see Figure 5.3). The protein was difficult to solubilize at pH 7.0 and 185 mM salt, the *Hs*MANEA buffer. By changing the pH to 6.0 it was possible to concentrate the protein up to 1.7 mg ml^{-1}.

Thermal unfolding experiments showed that both proteins are folded – the initial fluorescence is stable at 0 at lower temperatures (25–35 °C), which also means that the fluorescence change (differential) is around 0 when the protein is heated at these temperatures (the fluorescence change is the Y axis in Figure 5.4). MANEA unfolds with a T_m of 50.3 °C (concordant with the value obtained from CD spectra in Figure 4.9), while MANEAL exhibits two melting points: 48.7 °C and 60.0 °C (Figure 5.4). All three replicates gave similar results.

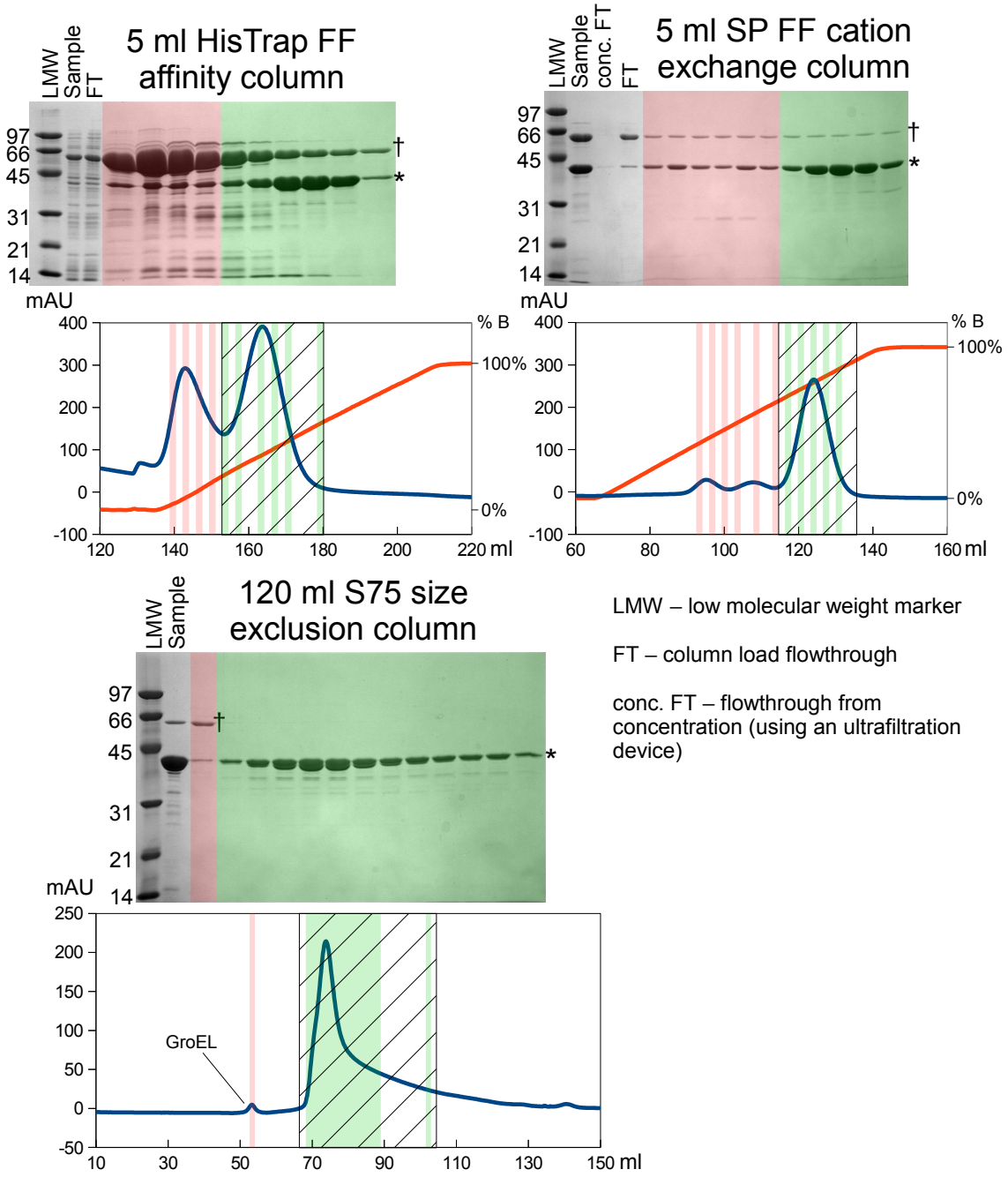


Figure 5.3: Purification of *Hs*MANEAL WT from BL21(DE3) cells. Nickel affinity column (FF crude) was used in the first step, followed by SP cation exchange chromatography and size exclusion chromatography. The blue line represents the Abs_{280} signal (mAU), and the orange line the concentration of buffer B (arbitrary 0-100% scale). The collected fractions are tinted green on the 12% reducing SDS-PAGE gel image and the corresponding chromatogram, the discarded fractions are tinted red. The molecular weight of ladder bands are in kDa. The hatched box represents all fractions that were collected for later use. †GroEL chaperone, *N-terminally His-tagged *Hs*MANEAL.

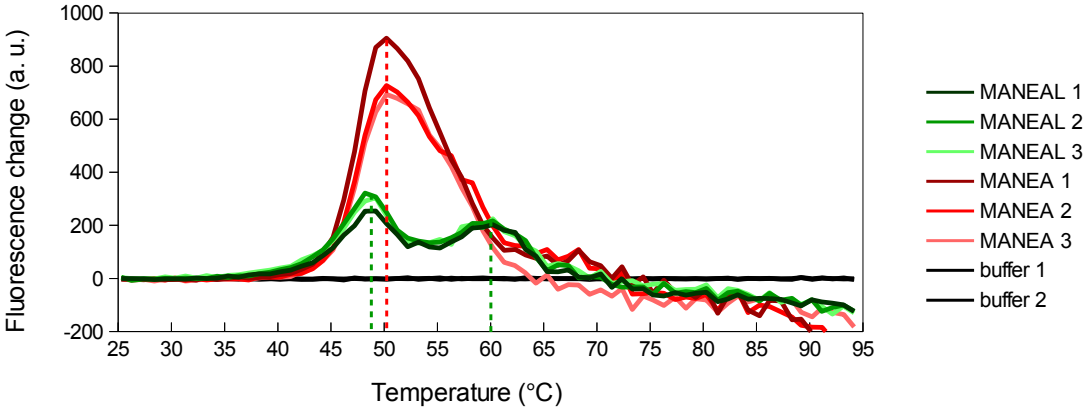


Figure 5.4: First derivative melting plot of *Hs*MANEAL (green lines) and *Hs*MANEA (red lines). The peaks correspond to inflection points of the melting curves, which establish the T_m.

5.5.2 Enzymology of MANEA and MANEAL

The truncated variant of *Hs*MANEA is catalytically active towards its natural substrate $GlcMan_9GlcNAc_2$. It releases α-Glc-1,3-Man and $Man_8GlcNAc_2$: the presence of the latter was confirmed by mass spectrometry. Some of the initial substrate was uncleaved even after prolonged incubation with the enzyme (see Figure 5.5). This result confirms that the 97 N-terminal amino acids in MANEA are not essential for its activity.

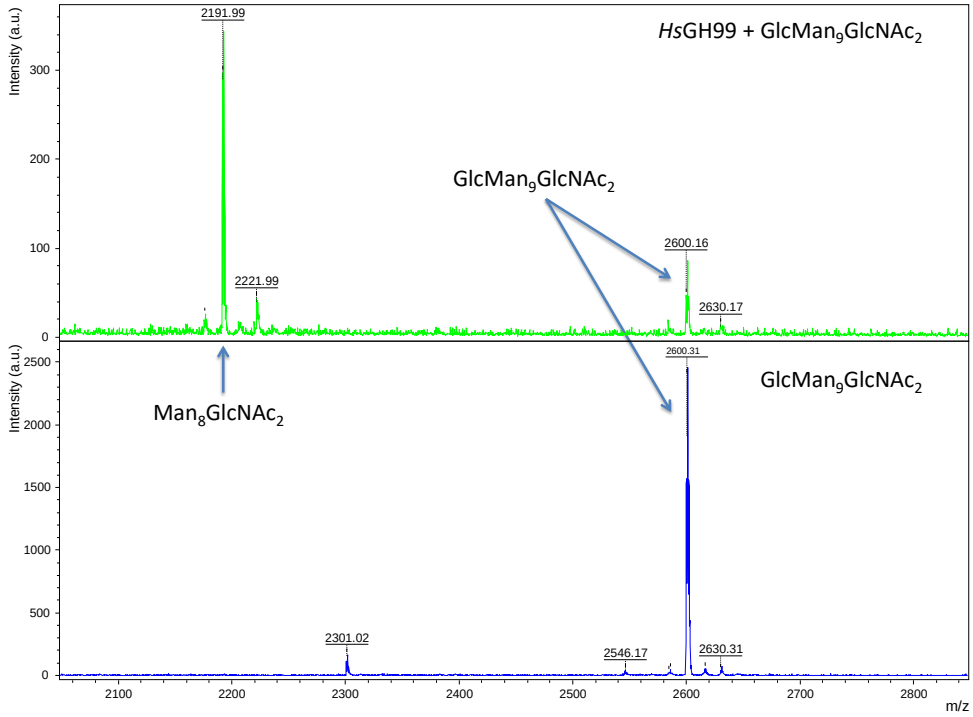


Figure 5.5: Mass spectra of carbohydrates present in solution of $GlcMan_9GlcNAc_2$ after incubation with (top) or without (bottom) *Hs*MANEA. The peaks correspond to m/z of $[M+Na]^+$ permethylated derivatives of the oligosaccharides.

MANEA exhibited considerable activity towards $GlcMan_3OMe$. The K_M value with this substrate was 426 ± 33 μM and k_{cat} 27.7 ± 1.0 min^{-1} (k_{cat}/K_M of 65 mM^{-1} min^{-1}). The Michaelis-Menten curve fit is presented in Figure 5.6. MANEA catalytic efficiency is almost identical to that of *Bt*GH99 on the Man_4OMe substrate:[207] (K_M = 2.6 mM, k_{cat} = 180 min^{-1}, k_{cat}/K_M = 69 mM^{-1} min^{-1}). The 6 × lower K_M of the human enzyme also comes with about 6.5 × lower k_{cat}. The similar values of k_{cat}/K_M for these enzymes, which likely emerged and evolved independently, suggest there is little room for improvement in catalytic efficiency. It is possible that the trade-off for the improved substrate affinity expressed by the lower K_M value is a slower process of dissociation of the products, which reduces the k_{cat}. These values also reflect the environments

in which the enzymes operate: it is more important that no *N*-glycans are left unprocessed in the ERGIC/Golgi, even if their concentration is quite low. In contrast, yeast mannan may reach high concentrations around the cell of a *Bacteroides* species, and in this situation what matters is the velocity at which the enzyme can process the sugars. There was no measurable activity when the substrate was incubated with *Hs*MANEA E404Q mutant.

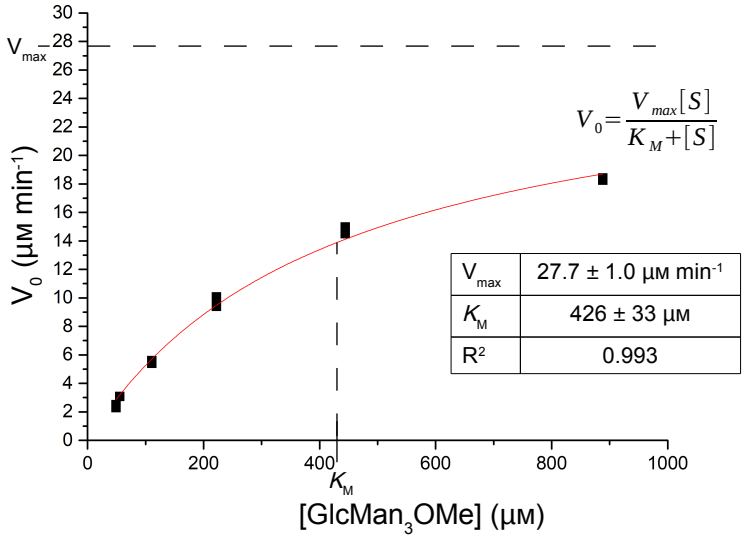


Figure 5.6: Michaelis-Menten kinetics of the catalytic domain of *Hs*MANEA with $GlcMan_3OMe$ as its substrate. The data points at around 49 μM and 111 μM [$GlcMan_3OMe$] done in triplicate and for higher concentrations in duplicate. A single measurement was also takes at 56 μM [$GlcMan_3OMe$]. The error values shown are standard errors as reported by the fitting software. Due to the precious nature of the substrate, experiments at higher [$GlcMan_3OMe$] concentrations were not attempted at this point – the value of V_{max} might be corrected in the future. Figure created using Origin software.

The purified version of *Hs*MANEAL was not active when assessed by thin-layer chromatography (see Figure 5.7). While *Hs*MANEA left no detectable $GlcMan_9GlcNAc_2$ substrate, releasing GlcMan and $Man_8GlcNAc_2$, the contents of the mixture containing *Hs*MANEAL and the oligosaccharide were not different from those of the negative control. Together with the thermal unfolding data, this suggests that the protein might be misfolded. However, the limited amount of the protein prevented analysis by CD. Attempts at crystallization with various screens produced no hits.

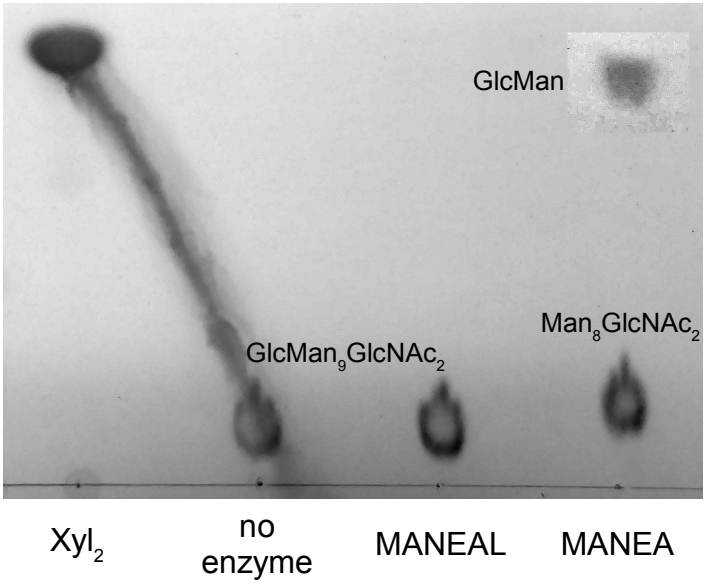


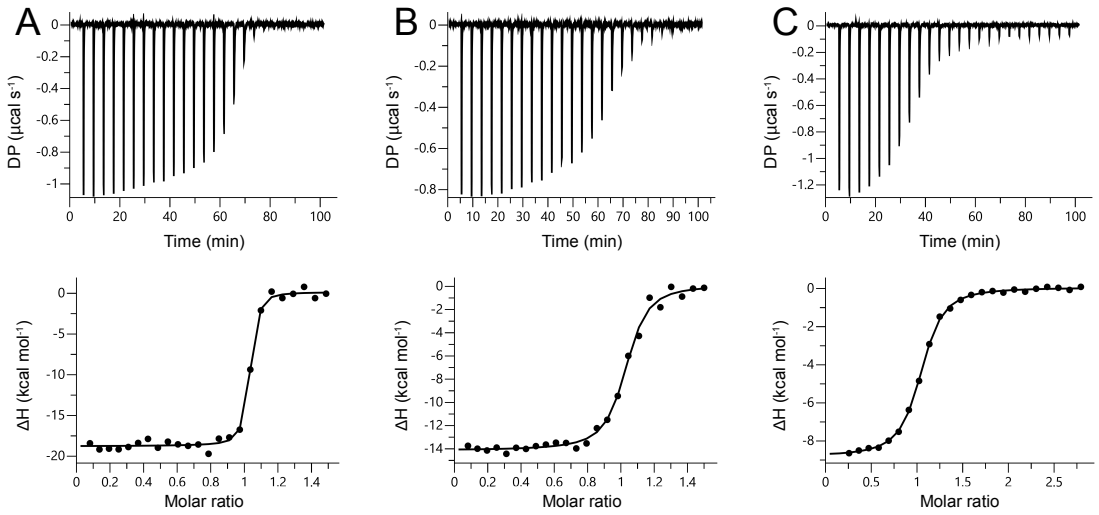
Figure 5.7: Results of MANEAL activity assay. The tailing at the Xyl_2 lane is an artefact of overloading and does not affect the migration of the neighbouring sample. Contrast around the GlcMan spot was enhanced due to a low amount of the species.

5.5.3 MANEA and MANEAL inhibitor binding

Isothermal titration calorimetry experiments with the human GH99 were problematic. In case of GlcIFG and ManIFG, the number of sites as measured by the inflection point of the isotherm was larger than 1 (2.3–2.4), even though the shape of the curve was entirely consistent with a single binding mode. Parallel experiments with *Bt*GH99 were run in order to measure the ligand concentration using ITC, as it was known that *Bt*GH99 has one binding site for ManIFG and GlcIFG. These experiments reduced the discrepancy to 1.7–1.8, which was unsatisfactory. In order to exclude the possibility of a secondary binding site and saturate the +1/+2 subsites, titration experiments with a large excess (1 mM, 23 × protein concentration) of α-1,2-mannobiose in the cell and the syringe were run. α-1,2-Mannobiose is the disaccharide that binds in the +1/+2 subsites, and they were considered as possible culprits changing the binding mode. In this experiment, the apparent number of sites was was 1.1.

The possibility that the protein concentration having not been measured correctly, which would skew the apparent N, was considered. The protein concentration was tested with folded and unfolded *Hs*MANEA and the results were identical. If a part of the protein was unfolded or inactive, the apparent number of binding sites would be lower than 1, not higher than 1, as the inhibitor would saturate the protein at a lower concentration.

When data from the ITC experiments done in presence or without α-1,2-mannobiose were analyzed (assuming that N=1 in both cases) and the syringe (inhibitor) concentration, instead of the N, was left to refine freely, the obtained ΔH and K_D values were similar. This led to a tentative conclusion that the the apparent stoichiometry of 1.7–1.8 was an experimental artefact. If the additional binding site had been real, there would have been a difference between the ΔH and K_D values obtained from experiments with or without α-1,2-mannobiose. In the ITC curves presented in Figure 5.8**AB**, data processing was done with an assumption of a single binding site, and the syringe concentration was left to refine freely. The experimental N for BGlcIFG (Figure 5.8**C**) was 1.



Inhibitor	GlcIFG	ManIFG	BGlcIFG
[Protein] in cell (µM)	42.4	42.4	64
[Inhibitor] in syringe (µM)	317 ± 1	320 ± 2	900 (fixed)
N	1 (fixed)	1 (fixed)	0.998 ± 0.003
K_D (nM)	**19.6 ± 5.6**	**170 ± 32**	**929 ± 52**
ΔH (kcal mol^{-1})	-18.9 ± 0.3	-14.2 ± 0.4	-8.85 ± 0.06
-TΔS (kcal mol^{-1})	8.35	4.96	0.624

Figure 5.8: ITC curves for various ligands titrated into cell containing a solution of *Hs*MANEA. (**A**) GlcIFG, (**B**) ManIFG, (**C**) BGlcIFG.

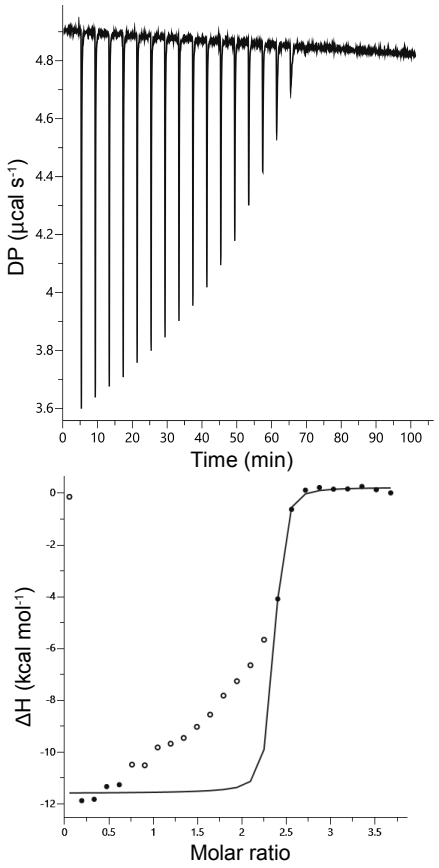
The results of ITC experiments are in good agreement with the differences between the bacterial and the human GH99 enzymes revealed by the kinetic data (see Table 5.1 for a comparison). The overall affinity of the human enzyme to the inhibitors is higher, which is concordant with its lower K_M value on its preferred substrate. Comparison of ΔH and -TΔS values obtained from the ITC experiments on *Hs*MANEA with GlcIFG, ManIFG and BGlcIFG reveals that with increasing inhibitor affinity, the enthalpy change (negative – release of heat) has a larger magnitude, as well as the entropy change (a decrease in entropy). The decrease in entropy for BGlcIFG is 13 × smaller than for GlcIFG – the most extreme change. This may reflect that the 191–201 loop in *Hs*MANEA remains unstructured upon the binding of BGlcIFG. Structural studies are needed to resolve this question.

Table 5.1: Comparison of dissociation constants for isofagomine-type ligands with *Hs*MANEA and *Bt*GH99. *The error is not known, as the value was taken from literature (as reported in Ref. 148). The K_D value of ManIFG with *Bt*GH99 is taken from Ref. 159.

Compound	K_D *Hs*MANEA (nM)	K_D *Bt*GH99 (nM)
BGlcIFG	929 ± 52	–
ManIFG	170 ± 32	140 ± 16
GlcIFG	19.6 ± 5.6	625*

ITC experiments with ManNOE were also attempted. In this case, it was not possible to interpret the curve using standard fitting software. The reported number of sites was 2.2–2.3. The curve had an unusual shape: initial injection points indicated weak binding, and when the number of sites reached about 2, the binding became tight and the protein saturated almost instantly. The unusual behaviour observed might indicate cooperative binding or that both the *gluco*- and *manno*- forms of ManNOE bind to the enzyme. One indicator of this is that the time interval for the DP to return to baseline for *Hs*MANEA is not long, as it was in the case of *Bt* and *Bx*GH99 (compare to Figure 3.5 on page 75). In Figure 5.9 the result of the titration is presented. If the binding was weak, one would expect the curve to not show signs of saturation near the end of the titration. If the binding was strong, all of the injected ligand would initially bind after each injection, producing a curve similar to those in Figure 5.8**AB**. The ManNOE binding isotherm has characteristics of both. Alternative models were tried (two set of sites, two sequential binding sites), but due to no knowledge of the initial parameters, no reasonable results were obtained. Figure 5.9 represents only an

attempt at fitting a curve that would partly explain the ITC result – a more satisfactory answer needs to be arrived at, which requires ITC experiments in presence of α-1,2-mannobiose or displacement titrations, where ManNOE would outcompete a weaker inhibitor in the –2/–1 subsites. Further structural might shed light on the question of what process gives rise to this behaviour.



Inhibitor	ManNOE
[Protein] in cell (µM)	38.5
[Inhibitor] in syringe (µM)	714
N	2.290 ± 0.007
K_D (nM)	**38 ± 17**
ΔH (kcal mol^{-1})	-11.8 ± 0.1
-TΔS (kcal mol^{-1})	1.67

Figure 5.9: ITC traces of ManNOE titrated into a solution of *Hs*MANEA at 38.5 µM (ligand concentration 714 µM). Filled dots represent points used to fit the curve and empty dots have been excluded from the fit. The isotherm was fitted using the model for a single set of sites. The reported K_D, ΔH and -TΔS values are likely incorrect, as the model assumed does not explain the shape of the whole curve.

5.5.4 Hepatitis C infectivity assay

An assay of the effect of incubation of cells infected with HepC with four GH99 ligands and *N*B-DNJ was performed. The GH99 inhibitors (at 10 concentrations from 4 µM to 2 mM) were used either with or without the glucosidase inhibitor. The 2 mM concentration of *N*B-DNJ was chosen so as to achieve full glucosidase inhibition: 0.5 mM is nearly fully inhibiting[313] and it was increased four-fold. Multiple images similar to the one in Figure 5.10 were collected and analyzed *in silico*.

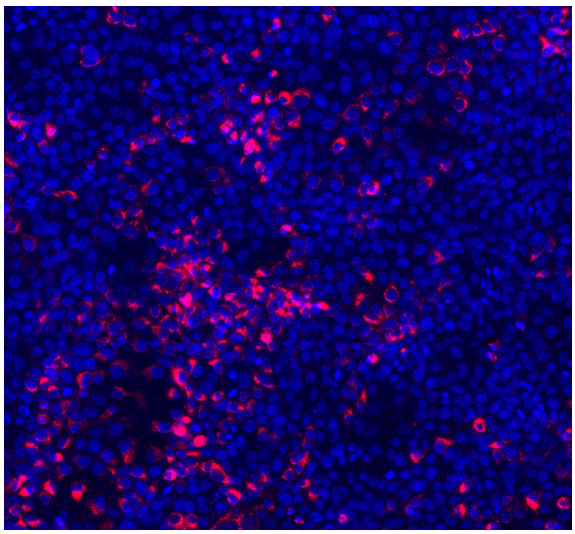


Figure 5.10: An example of DAPI (blue)/anti-NS5A (red) staining of HepC-infected cells. This is a small fragment of the whole image, hundreds of which were used in the numerical analysis.

The results of this experiment, presented in Figure 5.11 were surprising: only *N*B-DNJ reduced the percentage of infected cells. None of the GH99 inhibitors had such an effect, with an exception of Glc-β-aziridine, which was cytotoxic when added to the media (this could also be caused by a change in the media pH it induced). The initial goal of the experiment was to probe HepC infectivity. This would be measured more directly by a change in the average number of cells in the foci, but it would also have an effect on the percentage of cells that were infected. The latter parameter was used as a proxy measure of this effect, and if it had changed, the average number of cells in the foci would have been counted. As no proxy effect was observed, calculations of the direct measure of infectivity were not performed.

It is possible that this underwhelming result was produced by the experimental design. The concentration of *N*B-DNJ used might have been too high: at this concentration, both ER glucosidase I and II are likely to be inhibited. In such conditions, the relevance of GH99 activity might be lower (although this would be more of an issue for bovine MANEA). The experiment would benefit from repeating in a different cell line (for example HepG2) and at different concentrations of *N*B-DNJ. This was not done because of time constraints and difficulty in procuring access to Level III laboratories.

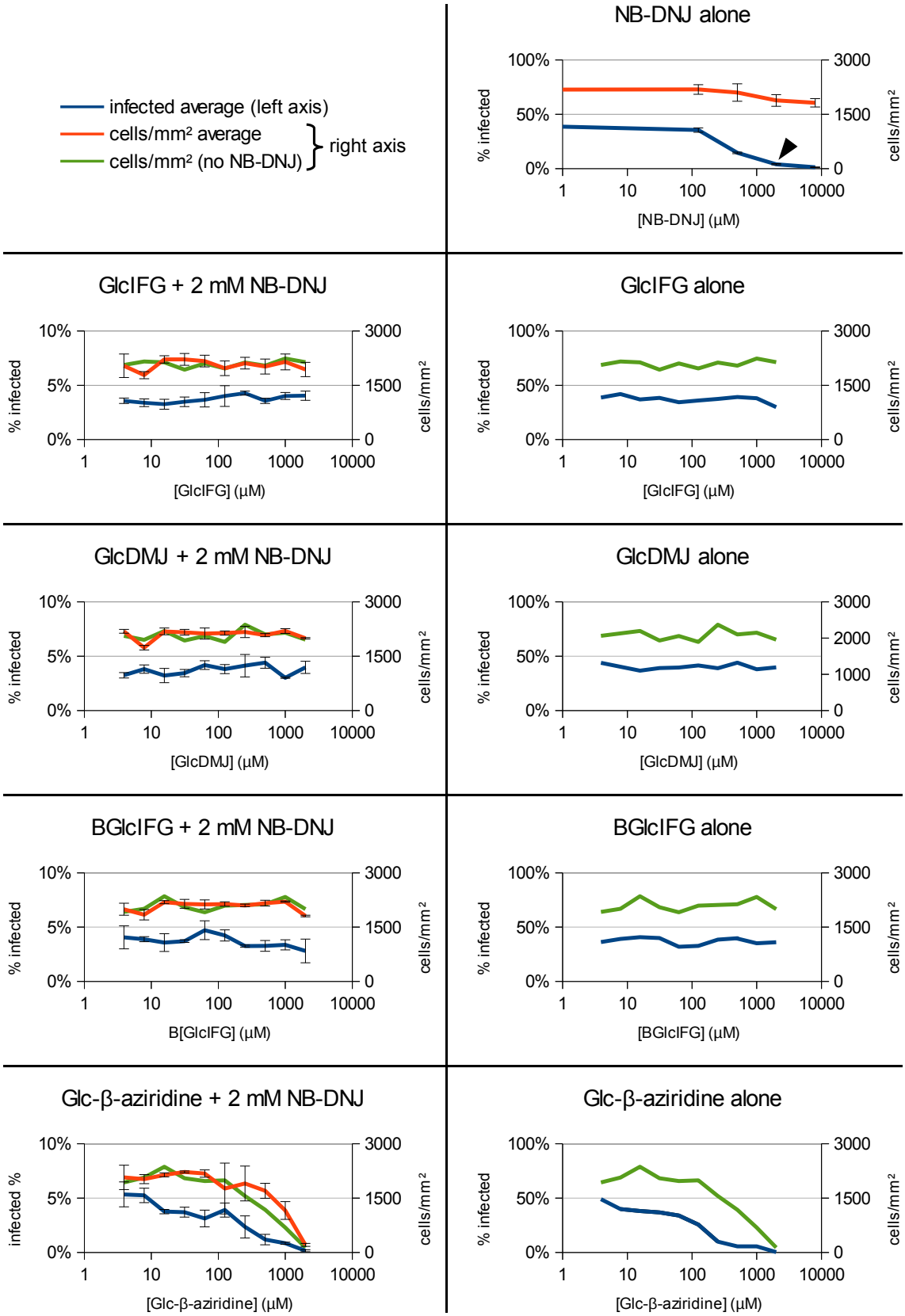


Figure 5.11: Results of the experiment investigating antiviral properties of endomannosidase inhibitors. The *N*B-DNJ-containing samples are triplicates (error bars: standard deviation) and samples with endomannosidase inhibitors only are not replicated. The arrowhead points to the concentration of *N*B-DNJ used in conjuntion with the GH99 inhibitors. Cell density and the number of infected cells were calculated using TissueQuest software.

5.6 Conclusions

GlcIFG is a potent inhibitor of endomannosidase. The affinity of the novel BGlcIFG is 50 × lower. Blocked-GlcIFG was synthesized in order to prevent ER glucosidase II-mediated breakdown of GlcIFG (the process was first reported in 1993[153] for GlcDMJ). In order to achieve this, ManIFG is a better candidate: it will be resistant to breakdown by the glucosidase and it has intermediate affinity to *Hs*MANEA (9 × lower than GlcIFG). The only non-lysosomal enzyme able to cleave α-1,3-mannosidic bonds is Golgi mannosidase II, but it is unlikely to break down this structure, as its recognition site needs a larger molecule with a GlcNAc branch. Therefore, the results presented here add to the knowledge about potential therapeutics acting on the endomannosidase. Results of the experiment that probed the antiviral properties of GH99 inhibitors against HepC were negative, but they do not conclusively rule out this potential therapy, as the effect might be obvious only at doses of glucosidase inhibitor that are lower than the one used.

The expressed version of *Hs*MANEAL was not active. This may mean that its natural substrate is not identical to that of MANEA (which would be a surprising result), or that it was not folded properly. Incorrect folding seems to be a more parsimonious explanation. A substantiating result was that all attempts at its crystallization failed, and the thermal unfolding showed two peaks. More research is needed to check the activity of this protein and characterize its structure, including its differences to MANEA.

Chapter 6

Conclusions and future perspectives

Knowledge of key cellular processes is key to understanding health and disease. Analysis of three-dimensional structures of proteins involved in these processes offers an in-depth look into their mechanism and opens avenues of research that were not considered before. *Endo*-α-1,2-mannosidase is a component of a process encountered widely in various life-forms: *N*-glycosylation. In particular, it is conserved in the *Eumetazoa* clade of animals, and in essentially all vertebrates. As such, its function is not dispensable for humans and organisms most closely related to humans. Glycoside hydrolases have been of interest for their involvement in viral replication, infectivity, cancer biology and congenital disorders of glycosylation (CDGs), which occur because of their abnormal function. In the following, short section, the origin of the genes encoding GH99 endomannosidases in view of the 3-dimensional structures that were obtained in this Thesis will be discussed.

6.1 The evolutionary origins of MANEA and MANEAL

In 1997, Dairaku and Spiro conducted a phylogenetic analysis in which they argued that the *endo*-α-1,2-mannosidase is present in chordates and not in insects.[185] The recent explosion of available genome sequences made it possible to re-evaluate their conclusions. While it is true that some insects seem to have lost the *MANEA* gene with the canonical active site sequence EWHE, they ofter harbor a GH99-like protein: for example CG14015 in *Drosophila melanogaster*. The canonical *MANEA* gene is found in the genome of *Cryptotermes secundus*, a species of termite. One of the results dis-

cussed in 1997 was a chance occurrence: the insect species surveyed just happened to lack the endomannosidase activity. The only species of sea cucumber that was studied was *Cucumaria frondosa*. Today, a protein predicted to be an endomannosidase was found using BLAST in a different sea cucumber species, *Apostichopus japonicus*. This is the only species of sea cucumber whose genome has been sequenced so far.[314] No endomannosidase was detected in protozoans *Tetrahymena*, *Trypanosoma* or *Leishmania*, confirming previous findings.[184,185] NCBI resource HomoloGene[315] claims that *MANEA* is conserved in *Bilateria* and *MANEAL* in *Euteleostomi*. This was also evaluated against new data and is presented in Figure 6.1**A**. No GH99 could be found in *Metazoa* (animals) sister group, *Choanoflagellata*. It also was not present in *Porifera* (sponges), a sister group to *Eumetazoa*. It was, however, present in *Placozoa* and *Cnidaria*, which are sister groups to *Bilateria* and daughter to *Eumetazoa*.[316,317] In *Bilateria*, the gene was found in all subgroups except *Ctenophora* (comb jellies). The last universal common ancestor (LUCA) of all animals lived 650 million years ago.[318] Sponges and *Trichoplax* diverged 600-620 My (megayears) ago. LUCA of comb jellies (*Ctenophora*) lived 550 My ago and no endomannosidase was detected in them. The endomannosidase, therefore, arose between 600–650 My ago in a population of metazoans that did not evolve into sponges or comb jellies.

A curious finding is the presence of an endomannosidase-like protein in *Blastocystis*, which is a heterokont (*Stramenopiles*) but not a metazoan. The common ancestor of metazoans and heterokonts lived about 1.5 Gy (gigayears) ago, which suggests that the origin of the endomannosidase could be in the more distant past. The primary protein sequence of the heterokont GH99 does feature the 191–201 loop region that all metazoan endomannosidase sequences have, but within it, a highly conserved glycine (*Hs*MANEA G198) is an alanine. Such a change would probably not alter the enzyme activity. Prokaryotic enzymes of similar activity were also surveyed. Essential differences between bacterial *endo*-α-1,2-mannanases and the eukaryotic *endo*-α-1,2-mannosidases are that in prokaryotes, a tryptophan is close to the –2 sugar instead of a tyrosine, and the 191–201 loop is absent. All of the surveyed bacterial *endo*-α-1,2-mannanases featured the tryptophan, suggesting their core substrate was yeast mannan (or their endogenous mannans). The heterokont GH99 has a tyrosine in that position, suggesting it has a better affinity towards Glu, rather than Man, in the –2 subsite, similar to metazoan endomannosidases.

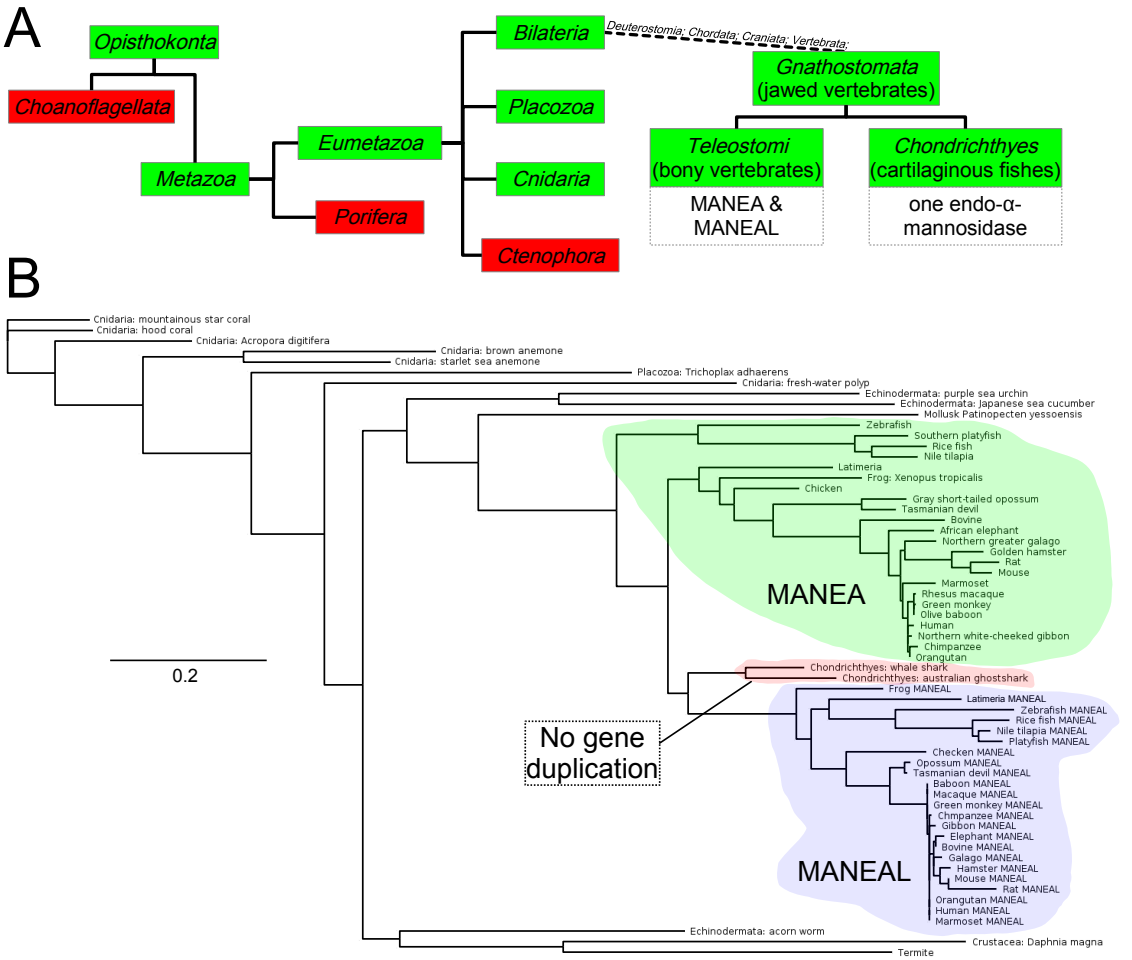


Figure 6.1: Phylogenetic analysis of *endo*-α-1,2-mannosidases. (**A**) A phylogenetic tree of endomannosidase evolution. In green taxa genes encoding endomannosidase were detected, in red no proteins predicted to have this activity exist. (**B**) PhyML[319] maximum-likelihood phylogenetic tree of a collection of endomannosidase proteins (catalytic domain only). The scale bar refers to calculated patristic distances. The unique placing of proteins from cartilaginous fishes (*Chondrichthyes*) is highlighted, as well as the sequences that are unambiguously MANEA and MANEAL.

What has never been investigated is at which point the gene duplication that gave rise to *MANEA* and *MANEAL* had occured. Using the publically available data it was straightforward to answer. *Gnathostomata* is the clade to which bony vertebrates and cartilaginous fishes like sharks belong. The genomes of sharks and rays (*Chondrichthyes*) do not encode any MANEAL proteins, but the *Teleostomi* clade, bony vertebrates, usually have *MANEA* and *MANEAL* genes. Both daughter groups in *Teleostomi*, *Actinopterygii* and *Sarcopterygii*, have these two genes. Therefore, the gene duplication must have occured after cartilaginous fishes diverged (460 million years ago) and before the LUCA of *Teleostomi* lived, which is 430 million years ago. This is supported by the phylogenetic tree analysis (see Figure 6.1**B**) and sequence identity analysis: MANEAL has the highest identity to the endomannosidase from cartilaginous fishes. Early evolution of eukaryotes is the subject of academic debate and all dates given here should be treated as very provisional. Differences in residue conservation in MANEA vs. the "ancestral" endo-α-mannosidases are visualized in Figure 6.2.

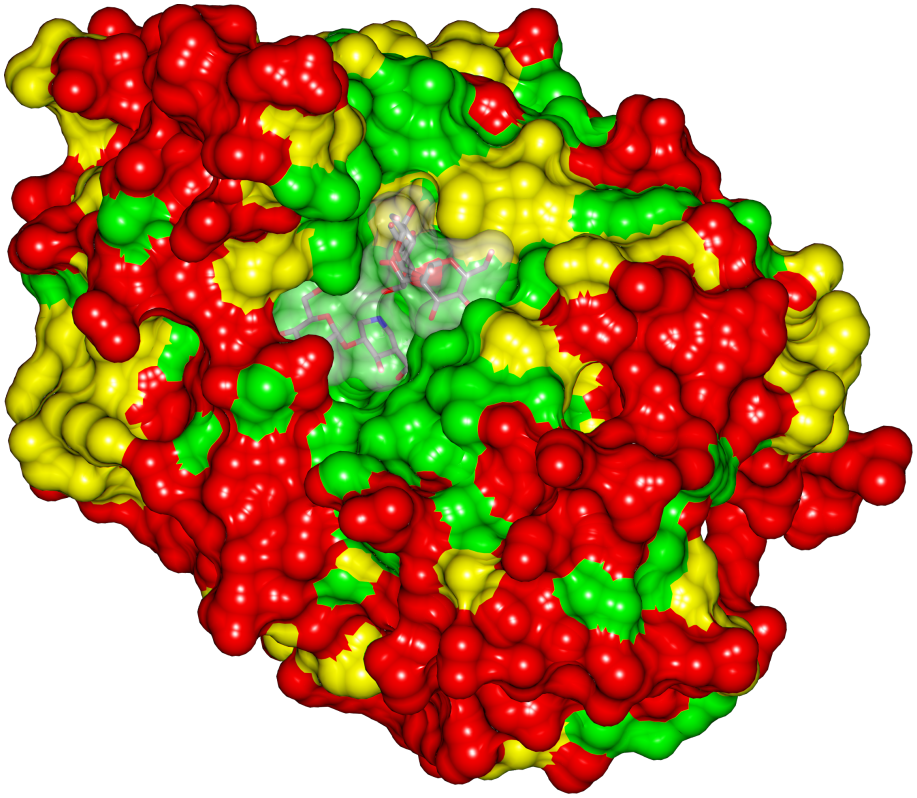


Figure 6.2: Conservation of the endo-α-mannosidase visualized using the structure of human MANEA with GlcIFG and α-1,2-mannobiose. Identity colour-coded on the calculated protein surface. Green: residues conserved in all endo-α-mannosidase genes including *Cnidaria* and *Placozoa*; yellow: residues conserved in MANEA proteins (Asp/Asn and Leu/Ile variations treated as conserved); red: non-conserved residues.

6.2 Synopsis

The research presented in this thesis contributed to the molecular understanding of the mechanism of GH99 endomannosidases/endomannanases. In Chapter 2, mechanistic and structural data entirely consistent with the enzyme using a 1,2-anhydrosugar disaccharide as the reaction intermediate were presented. The α-mannosidic bond is particularly inaccessible and α-mannosidases have to deform the conformation of the –1 sugar in order to catalyze its hydrolysis. This is achieved through various conformational itineraries, many of which involve half-chair and boat conformations, but none seem to use neighbouring-group participation mechanisms.[27,267] Interestingly, all of the *exo*-α-mannosidases involved in processing of *N*-glycans use divalent cations as cofactors. The endomannosidase does not need any metals in solution to work. It is therefore maybe not surprising that its mechanism is different. Through a combination of structural and computational approaches, the full itinerary for GH99 was assigned: $^{2}\boldsymbol{H}_{3}\rightarrow[\boldsymbol{E}_{3}]^{\ddagger}\rightarrow{}^{4}\boldsymbol{E}\rightarrow[\boldsymbol{E}_{3}]^{\ddagger}\rightarrow{}^{4}\boldsymbol{C}_{1}$. The mechanistic data support the hypothesis, but the ultimate proof will come from the kinetic isotope effects studies that are underway.

In Chapter 3, various approaches to GH99 inhibitor design were investigated and evaluated. As usual with this enzyme, there were surprises: Man-mannoimidazole, a potentially excellent inhibitor, was seen to bind only when forced into the –2/–1 subsite after prolonged soaking at a high concentration. On the other hand, Man-NOE was found to be the tightest endomannanase inhibitor ever studied. Classical isofagomine-type inhibitors were tested with human GH99 and found to be binding moderately tightly (even ManIFG, which has the "wrong" –2 sugar). These compounds, which mimic the carbocation, had in all cases better affinities than the reaction intermediate mimic Glc-β-aziridine or the oxocarbenium ion mimics.

The linkages in the *N*-glycan likely evolved under pressure that kept them unique and difficult to cleave. Similarly, yeast mannan is notoriously difficult to break down, but some bacteria manage to live on it.[207] These independent evolutionary pressures led to development of enzymes with a remarkably similar active site and virtually identical fold, which was shown in Chapter 4. The mechanistic data from bacterial endomannanase are directly transferable to human endomannosidase and their substrate binding is the same down to minute details, such as the double conformation of the mannose in the +1 position. Crucial, structural differences were also found.

In the eukaryotic GH99s a loop, disordered when the ligand is not bound, becomes ordered and envelops the –2 subsite from the "rear end" of the sugar. This loop is not present in the bacterial enzymes, making the binding cleft more open. To arrive at this point of understanding, difficulties with the crystal packing of *Hs*MANEA had to be overcome. Only co-crystallization with HEPES buffer and a polyoxotungstate molecule TEW yielded a crystal form suitable for inhibition studies. One structure, although not useful for characterization of inhibitors and saccharides occupying the –2/–1 subsites, was solved at a resolution of 1.1 Å. Such a near-atomic resolution offers a detailed structural view of the enzyme. The multiplicity of crystal forms helped establish that this loop becomes ordered upon binding of the –2/–1 ligand.

Crystal form 4 of *Hs*MANEA which was reproducible and able to accommodate the –2/–1 inhibitor was obtained rather late in the project. If this had been achieved sooner, mutagenesis of the 191–201 loop would have been performed in order to investigate its contribution to reaction kinetics and substrate specificity. For example, excision of the loop (or changing it to the sequence present in bacterial GH99s) may restore full endomannosidase activity in the presence of di- or triglucosylated forms of *N*-glycan. These studies would provide the structural basis for restricted processing of *N*-glycans by this enzyme, observed earlier.[169] Mutating Asp195 which is conserved in all eukaryotic GH99s and makes close contacts with the substrate would show the contribution of the loop to ligand affinity. The knowledge of this contact might also inform the design of novel inhibitors: it may be the case that changing the OH3 or OH4 group of the –2 sugar to an amino group may improve binding.

In Chapter 5 the kinetics of the endomannosidase reaction were discussed, as well as attempts at production of its paralog, MANEAL. The protein had the correct size on the gel and it was possible to purify it using a similar protocol to the one used for recombinantly expressed MANEA. However, it was not active, possibly due to misfolding (but an unfolding experiment suggested it was folded when in solution). The protein requires further investigation as it seems to have been overlooked in the research looking at *N*-glycosylation. The dissociation constants of isofagomine-type inhibitors were measured, including the novel BGlcIFG, which was found to bind somewhat less tightly than others. Finally, experiments with GH99 inhibitors in HepC-infected cells showed no effect, a result that is surprising given that initial data from another laboratory found that GlcIFG was antiviral. What might be responsible for this are the differences in methodology. Future work should concentrate on, first, replicating

the result obtained by Dr Dominic Alonzi and, second, investigating the properties of Man-substituted compounds such as ManIFG, which have the advantage of not being substrates of ER glucosidase II.

6.3 Perspectives for the field

In recent years, the field of glycoscience has immensely benefited from advances in analytical techniques. Mass spectrometry has become the tool of choice for describing the proteome of whole organisms and performing in-depth analyses of glycans on single proteins. Glycan MS can be supplemented by chemical modifications, which label certain features of the glycan.[320] Over 95% of the sites on PhosphoSitePlus, a database of PTMs, came from MS.[203] The glycan shield of the HIV envelope was described by mass spectrometry, and these insights may contribute to the development of one of the holy grails of medicine, a vaccine against HIV.[212] The question of how many proteoforms exist in humans is, however, still difficult to answer, as there is a multitude of PTMs which one protein can be subjected to. The current estimates are in the range of 6 million, but these are based on rough calculations.[321] Knowledge of glycosylation states in health and disease can be crucial in reversing an undesired process: for example, in heart failure, the glycosylation state of LDL protein is directly implicated in disease progression.[322] This also makes glycan profiles useful as biomarkers aiding diagnoses, especially in various types of cancer.[323,324]

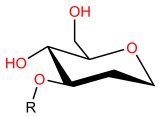
The relevance of the endomannosidase activity to human (and, more generally, animal) health is still largely unclear. Given the substantial amount of knowledge about its structure and mechanism, part of which was contributed by the research presented in this Thesis, this topic warrants investigation. The glycan profiles from organisms with altered endomannosidase function are obscure. It would be of interest to knock down or knock out mouse or rat *Manea* or *Maneal* and observe the differences from WT in the resulting phenotype and patterns of glycosylation. It is possible that endomannosidase-related CDGs exist in humans but have not been diagnosed because of their ambiguous presentation. If the research into antiviral properties of GH99 inhibitors proves they are of value, further studies on animal models, and then clinical trials, will be needed to establish the safety of these interventions.

Appendix A

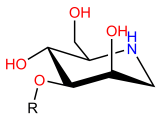
Compendium of compounds discussed at most length in the Thesis

GH99 INHIBITORS

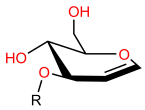
R=αMan or αGlc:



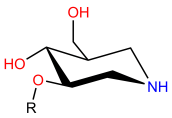
R-dideoxymannose (**R**ddMan)



R-1-deoxymannojirimycin (**R**DMJ)

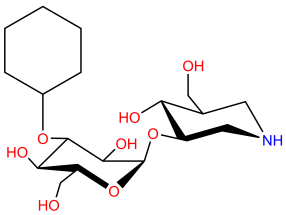


R-D-glucal (**R**glucal)

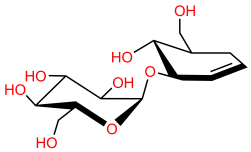


R-isofagomine (**R**IFG)

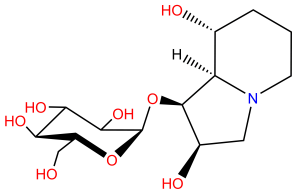
Glc in the –2 GH99 subsite:



blocked Glc-isofagomine (BGlcIFG)



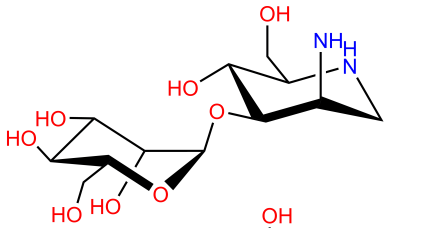
Glc-cyclohexene



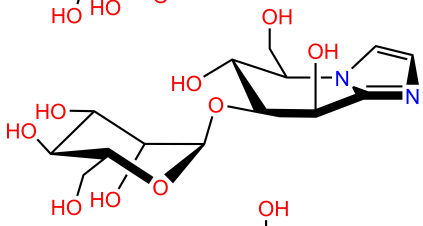
Glc-swainsonine (GlcSW, Fleetamine)

GH99 INHIBITORS

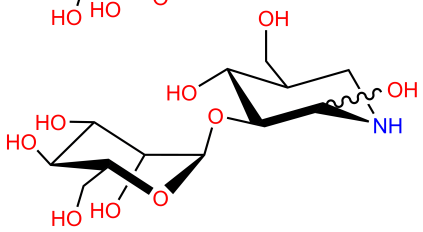
Man in the –2 GH99 subsite:



Man-aminoDNJ (ManADMJ)



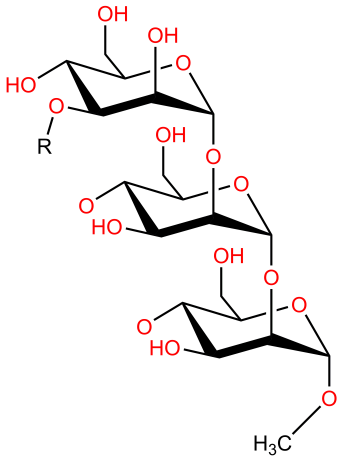
Man-mannoimidazole (ManManIm)



Man-noeuromycin (ManNOE)

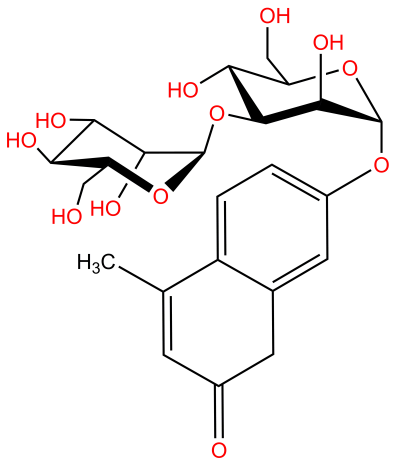
GH99 SUBSTRATES

R=αMan or αGlc:



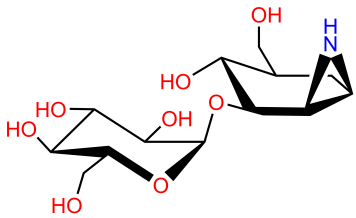
R-Man_3OMe

(Man_4OMe or $GlcMan_3OMe$)

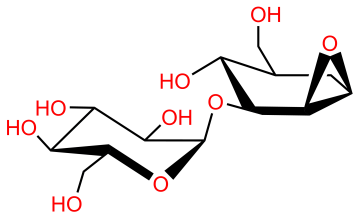


ManMan-4-methylumbelliferone (ManManMUF)

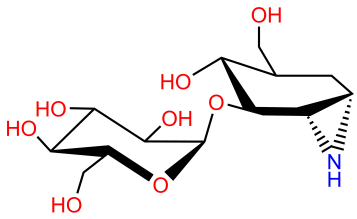
GH99 reaction intermediate mimics:



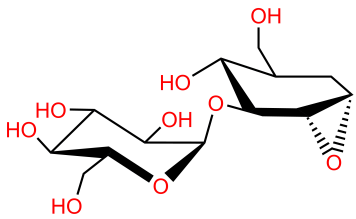
Glc-α-aziridine



Glc-α-epoxide



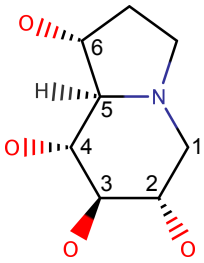
Glc-β-aziridine



Glc-β-epoxide

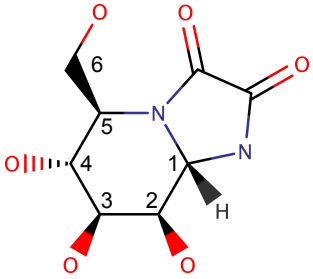
Exoglycosidase inhibitors:

pseudopyranose numbering
on fused ring inhibitors



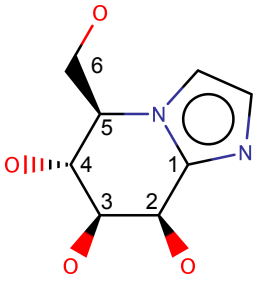
Castanospermine (CST)

Inhibits ER glucosidases I and II



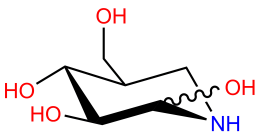
Kifunensine (KIF)

Inhibits GH47 mannosidases



Mannoiminazole (ManIm)

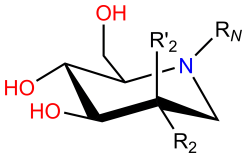
Inhibits GH47 and other
mannosidases



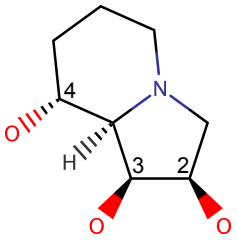
Noeuromycin (NOE)

Inhibits various glycosidases

Nojirimycin-type



R_2	R'_2	R_N	Compound	Abbreviation	Inhibits
OH	H	H	1-deoxynojirimycin	DNJ	glucosidases
OH	H	C_4H_9	*N*-butyl-deoxynojirimycin	*N*B-DNJ	glucosidases
H	OH	H	1-deoxymannojirimycin	DMJ	GH47 mannosidases



Swainsonine (SW)

Inhibits Golgi mannosidase II and
lysosomal α-mannosidase

Appendix B

Explanation of Cremer–Pople parameters describing the conformations of furanoses and pyranoses

The IUPAC nomenclature for sugar conformations ("the (approximate) arrangement of the rind atoms of a monosaccharide in the cyclic form")[325] recommends using a single, capital, italic letter. These letters are: *B* for the boat conformation, *C* for the chair conformation, *E* for the envelope conformation, *H* for the half-chair conformation, *S* for the skew-boat conformation and *T* for the twist conformation. The defined conformations of furanoses (sugars with a five-membered ring) are *E* and *T*, and of pyranoses (sugars with a six-membered ring) are *B*, *C*, *E*, *H* and *S*. These conformations are shown in Figure B.1**CD**. To decide on what the conformation of a particular ring is, a plane of reference is needed. In case of pyranoses, either three (in case of the twist conformation) or four (in case of the envelope conformation) atoms define the plane. In case of furanoses, it is either four atoms (conformations *B*, *C*, *H* and *S*) or five (the envelope conformation) roughly coplanar atoms define the plane. In 1975, Cremer and Pople[26] defined three parameters that are sufficient to describe the conformation of atoms in a closed ring. This concise approach has become the *de facto* standard applied when describing the conformation of furanoses and pyranoses. For furanoses, two parameters are needed: the angle ϕ and the scalar Q, also called the total puckering amplitude. For pyranoses, the parameters are two angles ϕ and θ, as well as the scalar Q. The meaning of these angles and the Q parameter is graphically presented in Figure B.1**AB** and explained below.

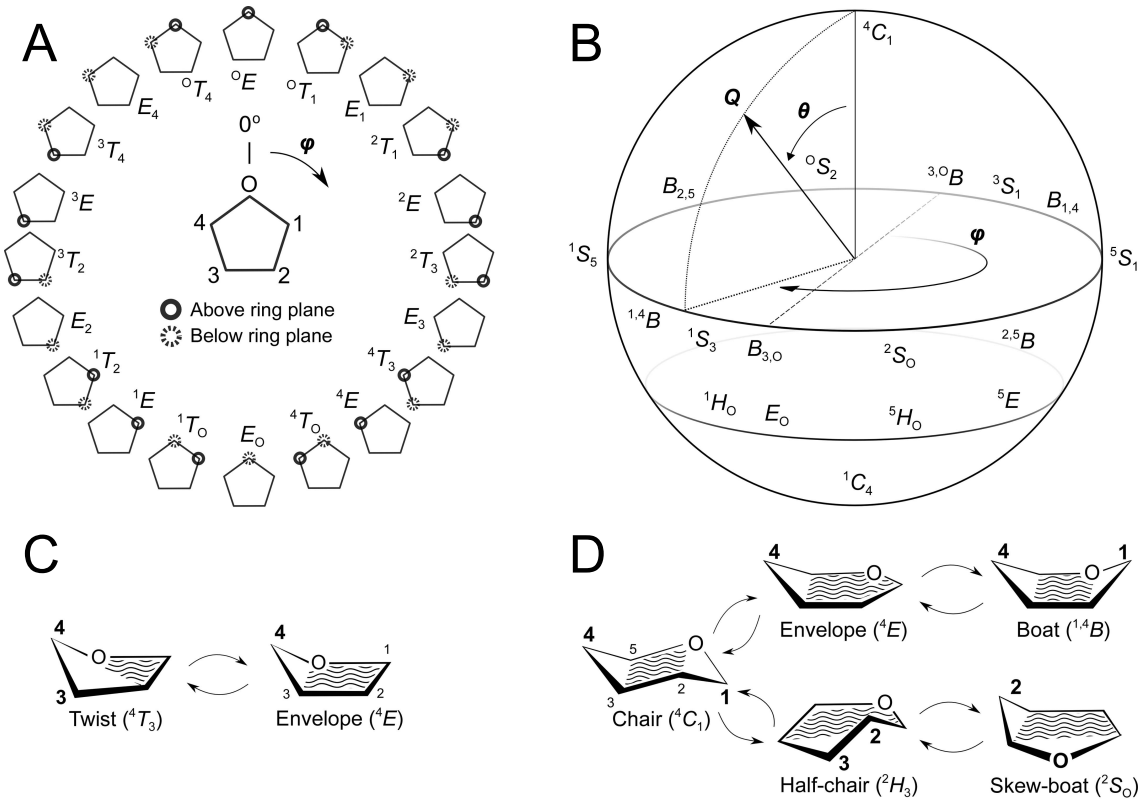


Figure B.1: Conformations of pyranoses and furanoses, their interconversions and the Cremer–Pople sphere. (**A**) The conformation of a furanose depending on the value of ϕ. (**B**) The Cremer–Pople sphere showing the meaning of angles ϕ, θ and the Q coordinate, as well as some conformations (ϕ=0° for $^{3,O}B$ and 180° for $B_{3,O}$). The value of ϕ increases clockwise. (**C**) The sugar planes and the interconversions between furanose conformations $^4T_3$ and ^4E. (**D**) Interconversions between various pyranose conformations and their mean planes. Figure adapted from Ref. 326.

In case of furanoses, there are 10 possible twist conformations and 10 possible envelope conformations (see Figure B.1**A**). In pyranoses, 2 *C* conformations, 6 *B* and *S* conformations and 12 *E* and *H* distinct conformations are possible. The Cremer–Pople parameters are sufficient to define each one. By convention, in furanoses the conformation assigned a ϕ angle of 0° is ${}^{O}E$, an envelope in which the oxygen atom is above the sugar plane (formed by C1, C2, C3 and C4) when the ring is looked at so that the carbon atom numbering grows clockwise (Figure B.1**A**). As the ϕ angle rises to 18°, the conformation changes to ${}^{O}T_1$, in which the oxygen atom is above the sugar plane and C1 is below the plane (formed by C2, C3 and C4). At ϕ=36°, the conformation of the furanose is E_1, and the plane is formed by atoms C2, C3, C4 and O. Finding the envelope conformation by looking at a structural model is straightforward, as four atoms always form a distinct plane. In case of the twist conformation, it requires assessing the structure from all sides coplanar with the possible planes (because any three atoms can form a plane). The mean sugar plane is the one that bisects a bond which is directly opposite the central plane atom. For example, when looked from a direction coplanar with the sugar plane of the conformation ${}^{O}T_1$ so that C3 is the closest to the observer, C2 and C4 are equidistant from the observer and O and C1 are behind them, C3 will be exactly in the middle of the bond between O and C1. The atom symbol in the superscript is always written before the conformation letter and refers to the atom above the plane. The subscript is always written after the conformation letter, and points to the atom below the plane.

In case of six-membered rings such as pyranoses, two angles are needed to describe the conformation, as the atoms may pucker (wrinkle) around the plane from both sides of the plane. The angle ϕ still refers to the side of the deformation in reference to conformations ${}^{3,O}B$, ${}^{O}E$ and ${}^{3}E$ which have a ϕ of 0° (the ϕ angle of the chair conformations, ${}^{4}C_1$ and ${}^{1}C_4$ does not affect their conformation, and indeed cannot be determined, when the θ angle is 0° or 180°). The θ angle helps explain what happens on both sides of the plane in reference to the position about the ring defined by the ϕ angle.

An intuitive understanding of the θ angle is quite difficult to attain. The following example might be helpful. If a sugar conformation changes from ${}^{4}C_1$ to ${}^{1}C_4$ (the "poles" of the sphere, see Figure B.1**B** and Figure B.2) and ϕ is kept at 0°, it has to go through ${}^{O}E$, ${}^{3,O}B$ and ${}^{3}E$ and all the intermediate conformations (see Figure B.2, Mercator pro-

jection). The first change, from ${}^4C_1$ to ${}^{O}E$, requires that atom C3 forms a plane together with atoms C1, C2, C4 and C5. This involves its upward movement if the sugar is looked at so that the atom numbering grows clockwise. ${}^{O}E$ has a θ angle of 45°. The change from ${}^{O}E$ to ${}^{3,O}B$ involves further upward movement of O. A θ angle of 90° at a ϕ of 0° means that atoms O and C3 are both above the sugar plane, and at an equal distance from it. The changes from ${}^{3,O}B$ (θ=90°) to ${}^{3}E$ (θ=135°) and then from ${}^{3}E$ to ${}^{1}C_4$ (θ=180°) require a downward movement of atom O with the position of C3 being kept above the plane. In order to find the mean planes of different conformations, refer to Figure B.1**D**.

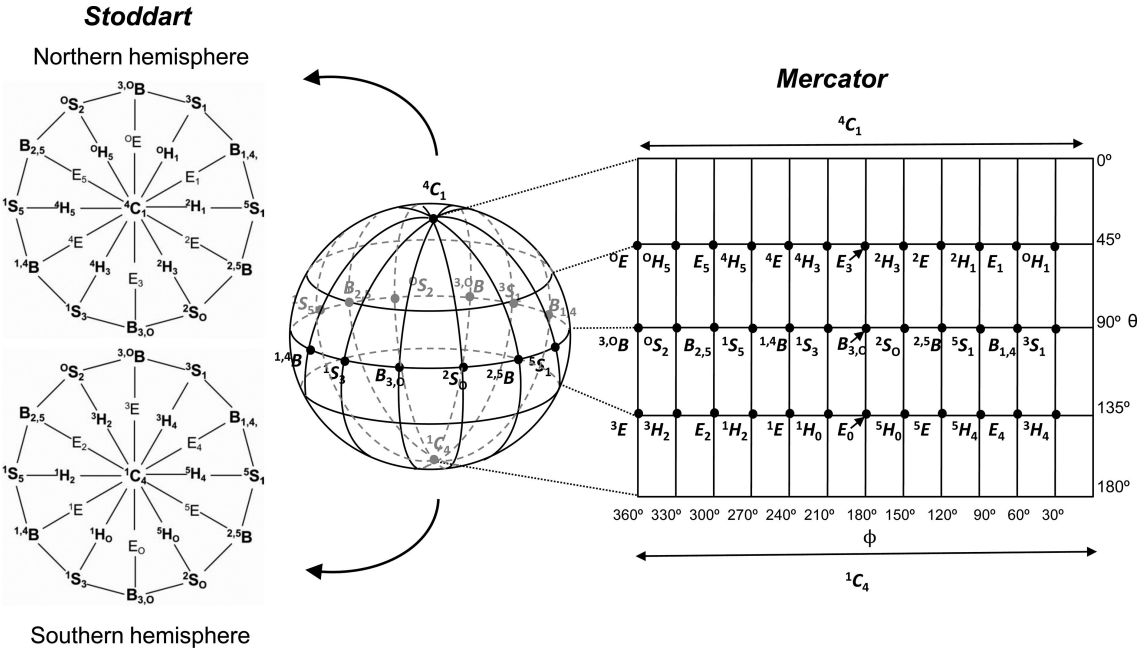


Figure B.2: The Cremer–Pople sphere projections and the conformations of pyranoses defined on the projections. Left: the polar (Stoddart) projections, northern hemisphere (${}^4C_1$ is the pole) and the southern hemisphere (${}^1C_4$ is the pole). Centre: the Cremer–Pople sphere. Right: The Mercator projection with all the defined conformations marked, as well as their corresponding ϕ and θ angles. Figure from Ref. 27.

The Q coordinate describes the total puckering of the sugar: it is a sum of the perpendicular distances of the ring atoms from the mean sugar plane. For a completely planar ring its value would be 0.

Appendix C

Bibliography

One manuscript is in preparation as a result of the research conducted for this Thesis and one is planned. The following papers were published during the course of this thesis work and are listed below in reverse order of publication date:

- Moroz, O. V., Sobala, L. F. Blagova, E., Coyle, T., Peng, W., Mørkeberg Krogh, K. B. R., Stubbs, K. A. Wilson, K. S. & Davies, G. J. Structure of a *Talaromyces pinophilus* GH62 arabinofuranosidase in complex with AraDNJ at 1.25 Å resolution. *Acta Cryst F* **74**, (2018).

- Schröder, S. P., Kallemeijn, W. W., Debets, M. F., Hansen, T., Sobala, L. F., Hakki, Z., Williams, S. J., Beenakker, T. J. M., Aerts, J. M. F. G., van der Marel, G. A., Codée, J. D. C., Davies, G. J., & Overkleeft, H. S. Spiro-epoxyglycosides as Activity-Based Probes for Glycoside Hydrolase Family 99 Endomannosidase/Endomannanase. *Chem. Eur. J.* **24**, 9983–9992 (2018).

- Fernandes, P. Z., Petricevic, M., Sobala, L., Davies, G. J. & Williams, S. J. Exploration of strategies for mechanism-based inhibitor design for family GH99 *endo*-α-1,2-mannanases. *Chem. Eur. J.* **24**, 7464–7473 (2018)

- Petricevic, M. Sobala, L.F., Fernandes, P.Z., Raich, L., Thompson, A.J., Bernardo-Seisdedos, G., Millet, O., Zhu, S., Sollogoub, M., Jiménez-Barbero, J., Rovira, C., Davies, G.J., Williams, S.J. Contribution of Shape and Charge to the Inhibition of a Family GH99 *endo*-α-1,2-Mannanase. *J. Am. Chem. Soc.* **139**, 1089–1097 (2017).

- Hemsworth, G. R., Thompson, A.J., Stepper, J., Sobala, Ł.F., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E.D., Stubbs, K.A., Brumer, H., Davies, G.J. Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut. *Open Biology* **6**, 160142 (2016).

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 23 November 2017
Accepted 4 January 2018

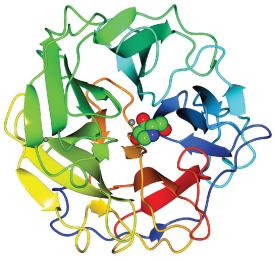
Edited by M. J. van Raaij, Centro Nacional de Biotecnología – CSIC, Spain

‡ Current address: School of Chemistry, University College Dublin, Stillorgan Road, Belfield, Dublin 4, Ireland.

Keywords: biofuels; glycosidases; enzymes; enzyme inhibitors; *Talaromyces pinophilus*; arabinofuranosidase.

PDB reference: *Talaromyces pinophilus* arabinofuranosidase, complex with AraDNJ, 6f1j

Supporting information: this article has supporting information at journals.iucr.org/f



Structure of a *Talaromyces pinophilus* GH62 arabinofuranosidase in complex with AraDNJ at 1.25 Å resolution

Olga V. Moroz,[a] Lukasz F. Sobala,[a] Elena Blagova,[a] Travis Coyle,[b]‡ Wei Peng,[c] Kristian B. R. Mørkeberg Krogh,[d] Keith A. Stubbs,[b] Keith S. Wilson[a] and Gideon J. Davies[a]*

[a]York Structural Biology Laboratory, Department of Chemistry, The University of York, York YO10 5DD, England, [b]School of Molecular Sciences, The University of Western Australia (M313), 35 Stirling Highway, Crawley, WA 6009, Australia, [c]Fungal Diversity, Novozymes A/S, China Headquarters, 14 Xinxi Road, Shangdi Zone, Haidian District, Beijing 100085, People's Republic of China, and [d]Protein Biochemistry and Stability, Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark. *Correspondence e-mail: gideon.davies@york.ac.uk

The enzymatic hydrolysis of complex plant biomass is a major societal goal of the 21st century in order to deliver renewable energy from nonpetroleum and nonfood sources. One of the major problems in many industrial processes, including the production of second-generation biofuels from lignocellulose, is the presence of 'hemicelluloses' such as xylans which block access to the cellulosic biomass. Xylans, with a polymeric β-1,4-xylose backbone, are frequently decorated with acetyl, glucuronyl and arabinofuranosyl 'side-chain' substituents, all of which need to be removed for complete degradation of the xylan. As such, there is interest in side-chain-cleaving enzymes and their action on polymeric substrates. Here, the 1.25 Å resolution structure of the *Talaromyces pinophilus* arabinofuranosidase in complex with the inhibitor AraDNJ, which binds with a K_d of 24 ± 0.4 μM, is reported. Positively charged iminosugars are generally considered to be potent inhibitors of retaining glycosidases by virtue of their ability to interact with both acid/base and nucleophilic carboxylates. Here, AraDNJ shows good inhibition of an inverting enzyme, allowing further insight into the structural basis for arabinoxylan recognition and degradation.

1. Introduction

The production of 'second-generation' biofuels, *i.e.* from nonfood plants, is a major societal goal as we move away from petroleum-based energy towards secure and renewable energy. Although the majority of polysaccharide biomass in plants is cellulose, the cellulose fibres are coated with hemicelluloses such as xylan, which render access to the cellulose more difficult. Enzymatic degradation of xylan is therefore necessary for the action of cellulase on higher plants, but it is also an important substrate in itself in that glucose and xylan, with small quantities of other sugars, are the major substrates for biofuel generation (discussed in Somerville, 2007). The enzymatic degradation of hemicelluloses such as xylan is of major importance in the biofuel industry (reviewed in Pauly & Keegstra, 2008) and also in diverse industries such as bread manufacture, animal feed and the pulp and paper industry (for pulp bleaching). Xylan, which is a major component of the plant cell wall, consists of a backbone β-1,4-linked D-xylosyl chain, which is decorated with diverse substituents including 2- and 3-linked arabinofuranosyl moieties (typically in cereal arabinoxylans) and glucuronic acid (notably in cereal and hardwood glucuronoxylans). Xylan complexity is further

segmented through ester-linked species such as acetyl and ferulate species, with the latter potentially linking the xylan to lignin (Fig. 1*a*). Degradation of xylan both in natural environments and in the industrial milieu therefore requires a plethora of enzymes, with some of the main players including β-xylanases, β-xylosidases, α-glucuronidases, acetyl and ferulate esterases and arabinofuranosidases, all of which are subject to keen academic and industrial study (recently comprehensively reviewed by Biely *et al.*, 2016).

Arabinoxylans, by virtue of being found in many of the plants now favoured for biofuel production, are considered to be a major 'feedstock' if we are to attain these societal goals in terms of renewable and secure energy (for reviews, see, for example, Lagaert *et al.*, 2014; Pauly & Keegstra, 2008). Given that arabinoxylan degradation requires a consortium of enzymes acting in partial synergy, most elegantly emphasized through Gilbert's recent work on xylan degradation by the microbiota (Rogowski *et al.*, 2015), there is much interest in the structure, mechanism and specificity of xylan-active enzymes, with a special focus on side-chain-cleaving enzymes and their potential synergy with backbone-cleaving xylanases. This potential synergy is further complicated by the differing capacities of the endoxylanases themselves to accommodate side chains. Of particular interest are the arabinofuranosidases, which are capable of removing the arabinofuranosyl (Araf) substituents from the 2- and 3-positions of the xylan backbone, thus opening up the xylan backbone for attack by classical endoxylanases. Arabinofuranosidases are found in families GH2, GH3, GH43, GH51, GH54 and GH62 of the CAZy sequence-based classification (http://www.cazy.org; Lombard *et al.*, 2014).

CAZY family GH62 contains many enzymes that act as arabinoxylan-active arabinofuranosidases (extensively reviewed in Wilkens *et al.*, 2017). The first three-dimensional structures of GH62 enzymes appeared in 2014, with structures reported from the bacteria *Streptomyces coelicolor* (Maehara *et al.*, 2014) and *S. thermoviolaceus* (Wang *et al.*, 2014) and of two fungal enzymes from *Ustilago maydis* and *Podospora anserina* (Siguier *et al.*, 2014). The three-dimensional structures share a common five-bladed β-propeller fold with an active centre consistent with hydrolysis with inversion of anomeric configuration, with conserved Glu and Asp residues acting as the catalytic acid and catalytic base, respectively, in the single-displacement mechanism (Fig. 1*b*). GH62 enzymes have been reviewed in CAZYpedia (for a review, see The CAZypedia Consortium, 2018).

Here, we present the three-dimensional structure of a fungal GH62 arabinofuranosidase from *Talaromyces pinophilus* refined at 1.25 Å resolution in complex with the bespoke iminosugar arabinofuranosidase inhibitor 1,4-dideoxy-l,4-imino-L-arabinitol (AraDNJ). The complex sheds light on the active site and, in light of previously published data, allows analysis of how the enzyme interacts with arabinoxylan substrates, serving to remove these side chains from the xylan backbone.

2. Materials and methods

2.1. Macromolecule production and small-molecule synthesis

The enzyme (a single-module GH62 arabinofuranosidase with no predicted N-glycosylation sites; GenBank MG656406)

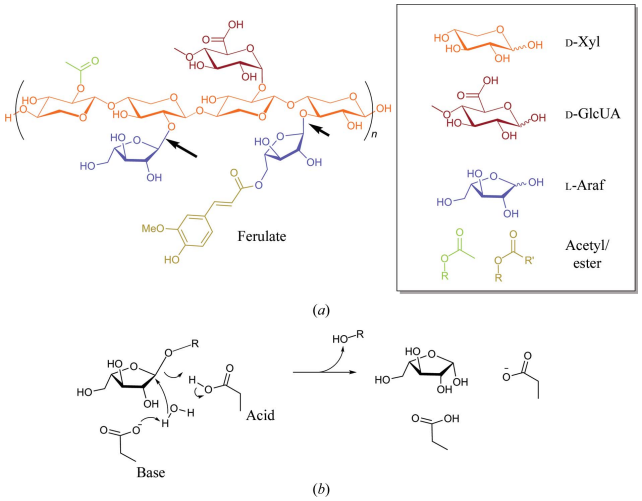


Figure 1
Xylans and their degradation. (*a*) The structure of a generic xylan, colour-coded by chemical group. Arrows indicate the positions of cleavage by arabinoxylan-active arabinofuranosidases. (*b*) The reaction scheme for an inverting arabinofuranosidase, which requires the presence of both Brønsted acid and base residues.

Table 1
Crystallization.

Method	Vapour diffusion, sitting drop; MMS
Plate type	MRC 2-well crystallization microplate, Swissci, Switzerland
Temperature (K)	293
Protein concentration (mg ml^{-1})	25
Buffer composition of protein solution	20 m*M* Tris–HCl pH 8.5, 150 m*M* NaCl
Composition of reservoir solution	30% PEG 2K MME, 0.2 *M* KBr
Volume and ratio of drop	300 nl total, 1:1 ratio
Volume of reservoir (µl)	54

was cloned and expressed by standard heterologous expression at Novozymes A/S using *Aspergillus oryzae* as the expression host, essentially as discussed in Biely *et al.* (2014). A novel band of about 35 kDa was observed in cultures of transformants that was not observed in cultures of the untransformed production strain. The expression level was investigated using SDS–PAGE for several transformants that appeared to express the recombinant arabinofuranosidase. After expression of the transformant with the highest expression level in a 1 l bioreactor, the culture broth was sterile-filtered to remove the mycelia. The filtrated broth was brought to 1.8 *M* ammonium sulfate, and after filtration (0.22 µm PES filter; Nalge Nunc International, Nalgene labware catalogue No. 595-4520) the filtrate was loaded onto a Phenyl Sepharose 6 Fast Flow column (high sub; GE Healthcare, Piscataway, New Jersey, USA) equilibrated with 25 m*M* HEPES pH 7.0 with 1.8 *M* ammonium sulfate; the column was washed with three column volumes of 25 m*M* HEPES pH 7.0, 1.0 *M* ammonium sulfate and bound proteins were eluted with 25 m*M* HEPES pH 7.0. The fractions were pooled and applied onto a Sephadex G-25 column (GE Healthcare) equilibrated with 25 m*M* HEPES pH 7.5. The fractions were applied onto a SOURCE 15Q column (GE Healthcare) equilibrated with 25 m*M* HEPES pH 7.5 and the bound proteins were eluted with a linear gradient from 0 to 1000 m*M* sodium chloride over ten column volumes. Fractions were analyzed by SDS–PAGE and those containing the arabinofuranosidase were combined.

The synthesis of AraDNJ was carried out using literature procedures (Jones *et al.*, 1985; Naleway *et al.*, 1988).

2.2. Crystallization

Crystallization screening was carried out by sitting-drop vapour diffusion with drops set up using a Mosquito Crystal liquid-handling robot (TTP Labtech, England) with 150 nl protein solution plus 150 nl reservoir solution in 96-well format plates (MRC 2-well crystallization microplates, Swissci, Switzerland) equilibrated against 54 µl reservoir solution. Experiments were carried out at room temperature using several commercial screens.

Extensive screening was carried out with no promising hits. As a final resort, the sample was subjected to shallow-gradient ion exchange in Tris–HCl pH 8.5. The resultant peak was asymmetric and the conditions of the run were adjusted to optimize the separation of different regions of the peak (whole gradient 0–1 *M* NaCl, peak separation at 10–20% of elution buffer). Fractions for these regions were pooled separately and concentrated. Crystallization was set up with protein fractions from the beginning of the peak. Crystallizations were performed both with and without the inhibitor AraDNJ which, when used, was mixed with the protein to give a final concentration of 5 m*M*. The best hit was obtained for protein in complex with the inhibitor from Crystal Screen HT condition G3 (0.01 *M* zinc sulfate, 0.1 *M* MES pH 6.5, 25% PEG 550 MME); this was chosen to make a seeding stock for further optimizations.

The seeding stock was prepared and microseed matrix screening (MMS; for a recent review, see D'Arcy *et al.*, 2014) was carried out using an Oryx robot (Douglas Instruments) according to the published protocols (Shaw Stewart *et al.*, 2011; Shah *et al.*, 2005) with two screens, Crystal Screen HT and JCSG, as well as a number of optimizations of the hit conditions. Diffraction-quality crystals were obtained from JCSG screen conditions B2, G7 and G10. That used for data collection was obtained from condition G10, *i.e.* 30% PEG 2K MME, 0.2 *M* KBr. The crystals were cryoprotected by adding PEG 3350 to the mother liquor in a 1:2 ratio (3 µl PEG + 6 µl mother liquor), which corresponded to 16.6% PEG 3350 and 20% PEG 2K in the final cryoprotectant solution. Crystallization conditions are shown in Table 1.

Table 2
Data-collection statistics.

Values in parentheses are for the outer shell.

Diffraction source	I04-1, DLS
Wavelength (Å)	0.93
Temperature (K)	100
Detector	PILATUS 6M-F
Crystal-to-detector distance (mm)	254.2
Rotation range per image (°)	0.1
Total rotation range (°)	180
Exposure time per image (s)	0.0375
Space group	$P2_1$
a, *b*, *c* (Å)	43.83, 88.97, 72.66
α, β, γ (°)	90, 95.22, 90
Mosaicity (°)	0.11
Resolution range (Å)	33.52–1.25 (1.27–1.25)
Total No. of reflections	457639 (14559)
No. of unique reflections	149344 (6813)
Completeness (%)	98 (91)
$CC_{1/2}$†	0.998 (0.79)
Multiplicity	3.1 (2.1)
$\langle I/\sigma(I)\rangle$	13.1 (2.9)
$R_{\rm merge}$	0.044 (0.28)
$R_{\rm r.i.m.}$‡	0.052 (0.34)
Overall *B* factor from Wilson plot (Å^2)	5.1

† $CC_{1/2}$ values for $I_{\rm mean}$ are calculated by splitting the data randomly into two half data sets. ‡ Estimated $R_{\rm r.i.m.} = R_{\rm merge}[N/(N-1)]^{1/2}$, where N is the data multiplicity, and $R_{\rm merge}$ is defined as $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of the reflection.

2.3. Data collection and processing

All computations were carried out using programs from the *CCP*4 suite (Winn *et al.*, 2011) unless otherwise stated. The data were collected on beamline I04-1 at Diamond Light Source (DLS) to 1.2 Å resolution and were processed with

*xia*2 (Winter *et al.*, 2013). Data-collection and processing statistics are given in Table 2.

2.4. Structure solution and refinement

The structure was solved by *MOLREP* (Vagin & Teplyakov, 2010) using *S. coelicolor* α-L-arabinofuranosidase (PDB entry 3wmy; Maehara *et al.*, 2014) as the search model. Chain tracing used *Buccaneer*, and the structure was refined with *REFMAC* (Murshudov *et al.*, 2011) iterated with manual model correction using *Coot* (Emsley *et al.*, 2010). The quality of the final model was validated using *MolProbity* (Chen *et al.*, 2010) as part of the *PHENIX* package (Adams *et al.*, 2011). The final refinement statistics are given in Table 3. The structure has been deposited in the PDB as entry 6f1j.

2.5. Isothermal titration calorimetry

Ligand affinity was measured using isothermal titration calorimetry (ITC). ITC was performed at 25°C in 25 m*M* HEPES pH 7.0, 100 m*M* NaCl using a Malvern MacroCal Auto-iTC200 calorimeter. The ligand in the syringe was at 1.8 m*M* and was titrated into a cell containing a 112 µ*M* solution of the enzyme. Assays were performed in duplicate. The dissociation constant was calculated using the *PEAQ-ITC Analysis* software (Malvern).

3. Results and discussion

The structure (PDB entry 6f1j) was solved and refined at 1.25 Å resolution (Table 3). The protein chain can be traced from residues 25 through to 325 and contains both structural calcium and zinc ions. The five-bladed β-propeller structure (Fig. 2*a*) bears a strong similarity to those of previously published GH62 enzymes, notably those from *S. coelicolor* (Maehara *et al.*, 2014) and *S. thermoviolaceus* (Wang *et al.*, 2014); 300 residues align with 72 and 69% sequence identity and r.m.s. C^{α} deviations of 0.58 and 0.68 Å, respectively, as reflected by high *PDBeFold* (Krissinel & Henrick, 2004) *Q* scores of 0.95 and 0.94, respectively. There are two subunits in the asymmetric unit with high structural similarity (r.m.s.d. of 0.22 Å), with some conformational differences on the outer surfaces, in particular in the region of crystal contacts.

Of the two metal ions, the Ca^{2+} ion is located essentially as reported previously, for example in the *S. coelicolor* enzyme (Maehara *et al.*, 2014). However, this structural Ca^{2+} ion (which is close to, but does not impinge on, the active centre) is coordinated by six water molecules and a carboxylate O atom from Glu215. This is different to previous structures, in which the Ca^{2+} ion was coordinated by a His and Gln pair, which are replaced here by a water molecule hydrogen-bonded to Ser278 (in place of the His) and directly to Glu215 (in place of the Gln observed previously). In the *T. pinophilus* enzyme there are additional Zn^{2+} ions derived from the 'seeding stock' (see above) element of the crystallization conditions. One of those bridges the *A* and *B* molecules in the lattice, presumably aiding lattice formation, with coordination from His180 from molecule *A* and the amino-terminal NH_2 and carbonyl groups of Ser24 and the side chain of Glu220 from molecule *B*. Another Zn^{2+} ion is coordinated by Glu88 from molecule *B*, His180 from the symmetry-related molecule *B* and three waters.

Table 3
Structure solution and refinement.

Resolution range (Å)	33.52–1.25
Completeness (%)	97.8
No. of reflections	
Working set	141792
Test set	7088
Final R_{cryst}	0.120
Final R_{free}	0.136
Cruickshank DPI	0.037
No. of subunits in the asymmetric unit	2
R.m.s. C^{α} deviation between subunits (Å)	0.221
No. of non-H atoms	
Protein	4698
Ion	4
Ligand	18
Water	658
Total	5378
R.m.s. deviations	
Bonds (Å)	0.014 (0.020)
Angles (°)	1.5 (1.9)
Average *B* factors (Å^2)	
Protein	
Chain *A*	7.3
Chain *B*	7.7
Ions	
Ca^{2+}	3.3
Zn^{2+} (1st)	8.8
Zn^{2+} (2nd)	8.4
Ligand	6.7
Water	18.8
Ramachandran plot†	
Favoured (%)	96.4
Outliers (%)	0.33
MolProbity score	0.85

† Ramachandran plot analysis was carried out by *MolProbity* (Chen *et al.*, 2010).

The structure of the *T. pinophilus* GH62 arabinofuranosidase was determined in the presence of the putative arabinofuranosidase inhibitor AraDNJ (Fig. 2*b*), which allows further confirmation of the catalytic apparatus. This compound has found use in studies of other arabinofuranosidases (Axamawaty *et al.*, 1990; Hemsworth *et al.*, 2016) as well as as a scaffold for developing inhibitors of other glycosidases (Siguier *et al.*, 2014; Mena-Barragán *et al.*, 2016). Azasugars and iminosugars are generally considered to be good inhibitors of retaining glycoside hydrolases by virtue of their endocyclic N atom, which can be protonated, thus mimicking the putative positive charge that is thought to exist in the transition state(s) during glycoside hydrolysis. In addition, the N atom provides adventitious interactions with both the acid/base and the nucleophile in the active sites of these enzymes (see, for example, Gloster *et al.*, 2007). GH62 enzymes are inverting and thus do not have a suitably positioned nucleophile. It was therefore surprising to us that AraDNJ acted as an inhibitor with well resolved density. The binding constant for AraDNJ was therefore determined by isothermal titration calorimetry (Fig. 2*c*), revealing a surprisingly tight K_d of 24 ± 0.4 µ*M*. It is rare in glycosidases that iminosugars bind so well to the glycosidase active site without

a close enzyme-derived nucleophilic interaction, but other examples include CAZY family GH6, where cellobio-derived isofagomines have been used to good effect, even reporting on the substrate distortions involved in catalysis (Gloster *et al.*, 2007). Here, AraDNJ binds in a potentially transition-state-mimicking ^{4}E conformation. As might be expected, AraDNJ binds in the same location as observed for Araf itself (see, for example, PDB entry 4o8o; Wang *et al.*, 2014), making similar hydrogen bonds from O2 and O3 to Asp160, from O3 to Gln120 and from O56 to Asp52. There is also a potential hydrophobic contact with the side chain of Ile159. There is no direct interaction of the positively charged N atom (here replacing the endocyclic O atom of arabinose), but the structure reveals a water molecule poised 3.1 Å 'below' the furanose ring, where it hydrogen-bonds to Asp52, the putative catalytic base, consistent with previous studies (Maehara *et al.*, 2014; Wang *et al.*, 2014) and the inverting mechanism (Fig. 1*b*). Glu212, the putative acid, is placed for lateral *anti* protonation of any departing group (Fig. 2*d*). Notably, the positively charged N atom lies exactly where the positively charged N atom of published Tris complexes of homologues sits (see, for example, PDB entry 3wn2, the *S. coelicolor* GH62 enzyme; Maehara *et al.*, 2014), highlighting that these enzymes have evolved to stabilize the positively charged transition state, even without the aid of the direct charge–charge interactions available to retaining enzymes.

The *T. pinophilus* GH62 enzyme in complex with AraDNJ, viewed in light of past work on xylooligosaccharide complexes of GH62 enzymes, provides further insight into the mechanisms by which GH62 enzymes remove the arabinofuranoside

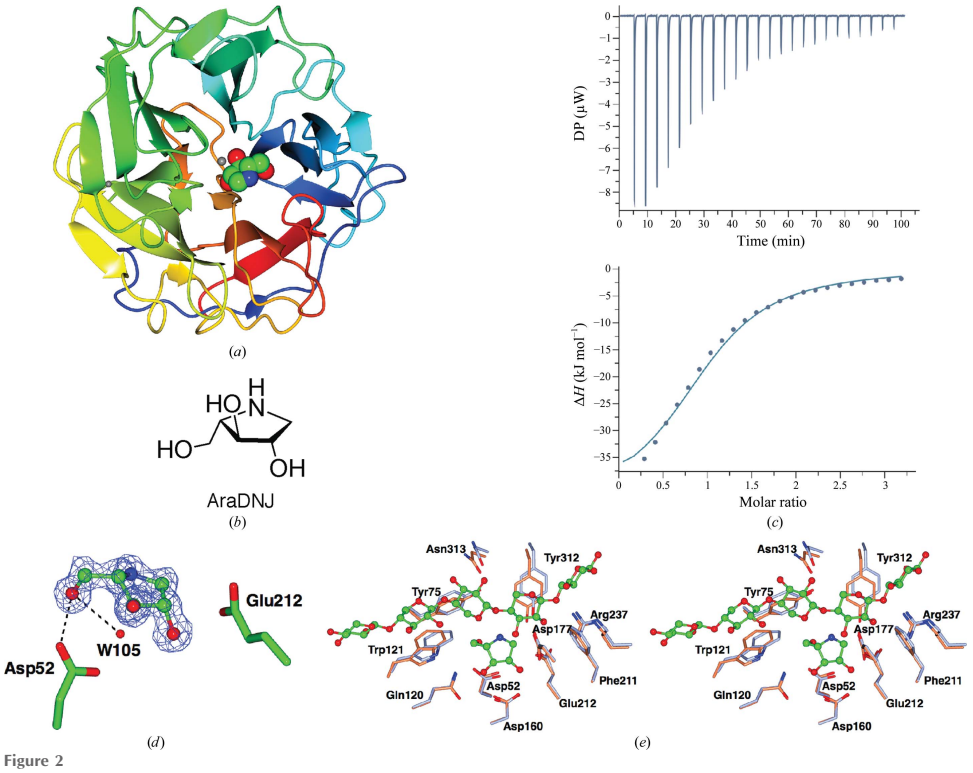


Figure 2
Three-dimensional structure and ligand binding of the *T. pinophilus* GH62 arabinofuranosidase in complex with the inhibitor AraDNJ. (*a*) Three-dimensional structure colour-ramped from the N-terminus (blue) to the C-terminus (red). Metal ions are shown as shaded spheres and AraDNJ as a CPK model. (*b*) The chemical structure of AraDNJ. (*c*) ITC data for AraDNJ binding (K_d of 24 ± 0.4 µ*M*). (*d*) Observed electron density for AraDNJ bound to GH62, $2F_o - F_c$ (maximum-likelihood/σ_A-weighted) at 1.25 Å contoured at 1σ. The catalytic acid Glu212 and base Asp52 are shown, along with a water molecule poised for nucleophilic attack. (*e*) Partial overlay of the *T. pinophilus* GH62 arabinofuranosidase (brown with AraDNJ in green) with the *S. coelicolor* GH62 arabinofuranosidase (PDB entry 3wn2; pale blue with xylopentaose in green), highlighting the highly conserved binding centre and the recognition apparatus for the arabinoxylan chain. Structural figures were drawn with *CCP4mg* (McNicholas *et al.*, 2011).

decorations from arabinoxylan. An overlay with the xylopentaose complex (PDB entry 3wn2) of the *S. coelicolor* GH62 enzyme (Maehara *et al.*, 2014; Fig. 2*e*) shows how the interacting surface for the xylan chain is highly conserved between the two enzymes, with both aromatic platforms (Phe211, Tyr312 and Trp121) and some hydrogen-bonding interactions (Arg237, Asn313 and Asp177) being invariant, suggesting that ligand recognition is similar. Indeed, C1 of the AraDNJ complex lies 1.9 Å from the O3 atom of the 'second' (from the reducing end) xylose moiety in PDB entry 3wn2, highlighting how the *T. pinophilus* GH62 enzyme could act as an arabinofuranosidase that is active on O3-substituted xylans, as was proposed originally for the *S. coelicolor* GH62 enzyme (Maehara *et al.*, 2014), although it is possible to also consider action at the O2 position should the xylan chain occasionally be reversed through the active site (which is possible with xylans given their internal pseudo-symmetry).

The *T. pinophilus* GH62 enzyme thus adds to the growing literature surrounding these key players in natural and industrial arabinoxylan degradation. It demonstrates how arabinofuranoside mimics lie in the active site of the enzyme and how the enzyme recognizes and cleaves arabinoxylan. Furthermore, the nonclassical application of an iminosugar-based glycosidase inhibitor to study inverting-enzyme structure and function should encourage the further non-intuitive application of such compounds in the future.

Acknowledgements

The authors would like to thank Novozymes A/S for partially funding this work. GJD is a Royal Society Ken Murray Research Fellow. KAS thanks the Australian Research Council and TC thanks the Australian Government, the University of Western Australia (UWA) and the Centre for Microscopy, Characterization and Analysis at UWA. The authors would also like to thank Diamond Light Source for beamtime (proposal mx13587) and the staff of beamline I04 for assistance with crystal testing and data collection.

References

Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
Axamawaty, M. T., Fleet, G. W., Hannah, K. A., Namgoong, S. K. & Sinnott, M. L. (1990). *Biochem. J.* **266**, 245–249.
Biely, P., Puchart, V., Stringer, M. A. & Mørkeberg Krogh, K. B. R. (2014). *FEBS J.* **281**, 3894–3903.
Biely, P., Singh, S. & Puchart, V. (2016). *Biotechnol. Adv.* **34**, 1260–1274.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. (2014). *Acta Cryst.* F**70**, 1117–1126.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Gloster, T. M., Meloncelli, P., Stick, R., Zechel, D., Vasella, A. & Davies, G. J. (2007). *J. Am. Chem. Soc.* **129**, 2345–2354.
Jones, D. W. C., Nash, R. J., Bell, E. A. & Williams, J. M. (1985). *Tetrahedron Lett.* **26**, 3125–3126.
Hemsworth, G. R., Thompson, A. J., Stepper, J., Sobala, Ł. F., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E. D., Stubbs, K. A., Brumer, H. & Davies, G. J. (2016). *Open Biol.* **6**, 160142.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Lagaert, S., Pollet, A., Courtin, C. M. & Volckaert, G. (2014). *Biotechnol. Adv.* **32**, 316–332.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
Maehara, T., Fujimoto, Z., Ichinose, H., Michikawa, M., Harazono, K. & Kaneko, S. (2014). *J. Biol. Chem.* **289**, 7962–7972.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* D**67**, 386–394.
Mena-Barragán, T., García-Moreno, M. I., Nanba, E., Higaki, K., Concia, A. L., Clapés, P., García Fernández, J. M. & Ortiz Mellet, C. (2016). *Eur. J. Med. Chem.* **121**, 880–891.
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.* D**67**, 355–367.
Naleway, J. J., Raetz, C. R. & Anderson, L. (1988). *Carbohydr. Res.* **179**, 199–209.
Pauly, M. & Keegstra, K. (2008). *Plant J.* **54**, 559–568.
Rogowski, A. *et al.* (2015). *Nature Commun.* **6**, 7481.
Shah, A. K., Liu, Z.-J., Stewart, P. D., Schubot, F. D., Rose, J. P., Newton, M. G. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 123–129.
Shaw Stewart, P. D., Kolek, S. A., Briggs, A. R., Chayen, N. E. & Baldock, P. F. M. (2011). *Cryst. Growth Des.* **11**, 3432–3441.
Siguier, B., Haon, M., Nahoum, V., Marcellin, M., Burlet-Schiltz, O., Coutinho, P. M., Henrissat, B., Mourey, L., O'Donohue, M. J., Berrin, J.-G., Tranier, S. & Dumon, C. (2014). *J. Biol. Chem.* **289**, 5261–5273.
Somerville, C. (2007). *Curr. Biol.* **17**, R115–R119.
The CAZypedia Consortium (2018). *Glycobiology*, **28**, 3–8.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, W., Mai-Gisondi, G., Stogios, P. J., Kaur, A., Xu, X., Cui, H., Turunen, O., Savchenko, A. & Master, E. R. (2014). *Appl. Environ. Microbiol.* **80**, 5317–5329.
Wilkens, C., Andersen, S., Dumon, C., Berrin, J.-G. & Svensson, B. (2017). *Biotechnol. Adv.* **35**, 792–804.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* D**69**, 1260–1273.

ChemPubSoc Europe

DOI: 10.1002/chem.201801902

CHEMISTRY A European Journal Full Paper

Biochemistry

Spiro-epoxyglycosides as Activity-Based Probes for Glycoside Hydrolase Family 99 Endomannosidase/Endomannanase

Sybrin P. Schröder,[a] Wouter W. Kallemeijn,[b] Marjoke F. Debets,[a] Thomas Hansen,[a] Lukasz F. Sobala,[c] Zalihe Hakki,[d] Spencer J. Williams,[d] Thomas J. M. Beenakker,[a] Johannes M. F. G. Aerts,[b] Gijsbert A. van der Marel,[a] Jeroen D. C. Codée,[a] Gideon J. Davies,*[c] and Herman S. Overkleeft*[a]

Abstract: N-Glycans direct protein function, stability, folding and targeting, and influence immunogenicity. While most glycosidases that process N-glycans cleave a single sugar residue at a time, enzymes from glycoside hydrolase family 99 are *endo*-acting enzymes that cleave within complex N-glycans. Eukaryotic Golgi *endo*-1,2-α-mannosidase cleaves glucose-substituted mannose within immature glucosylated high-mannose N-glycans in the secretory pathway. Certain bacteria within the human gut microbiota produce *endo*-1,2-α-mannanase, which cleaves related structures within fungal mannan, as part of nutrient acquisition. An unconventional mechanism of catalysis was proposed for enzymes of this family, hinted at by crystal structures of imino/azasugars complexed within the active site. Based on this mechanism, we developed the synthesis of two glycosides bearing a spiro-epoxide at C-2 as electrophilic trap, to covalently bind a mechanistically important, conserved GH99 catalytic residue. The spiro-epoxyglycosides are equipped with a fluorescent tag, and following incubation with recombinant enzyme, allow concentration, time and pH dependent visualization of the bound enzyme using gel electrophoresis.

Introduction

N-Linked glycans are complex oligosaccharides linked to asparagine (Asn) residues in eukaryotic proteins.[1] They play important roles in protein function, stability, folding and targeting and are essential for a range of cellular functions.[2] Erroneous N-glycan composition is associated with various diseases including viral infections, Alzheimer's disease and metastatic cancer.[3–5] Assembly of the N-glycan commences in the endoplasmic reticulum (ER) where the 14-mer polysaccharide $Glc_3Man_9GlcNAc_2$-diphosphodolichol is coupled to the Asn residue of the target protein by the enzyme oligosaccharyl transferase. The glycan undergoes stepwise "trimming" of the non-reducing end glucoside residues by α-glucosidase I and II, after which α-mannosidase I truncates the resulting oligomannoside.[6] The resulting $Man_5GlcNAc_2$ structure is redecorated to yield complex N-glycans. Because α-glucosidases I and II play important roles in the early stages of glycan maturation, these enzymes were investigated as therapeutic targets to control diseases involving incorrect N-glycosylation.[7–10] However, inhibition of these enzymes did not block N-glycosylation: mouse lymphoma cells inhibited with the α-glucosidase inhibitor castanospermine as well as mutant cell lines lacking α-glucosidase II retained up to 80% of normal N-glycan maturation.[11–13] Spiro and co-workers identified *endo*-1,2-α-mannosidase,[14,15] (later classified as a member of glucoside hydrolase family 99 (GH99); see http://cazypedia.org),[16] residing in the Golgi apparatus, which circumvents inhibition of ER α-glucosidase I and II. The enzyme cleaves glucose-substituted mannose from the A-branch of ER escaped immature N-glycoproteins bearing $Glc_{1-3}Man_9GlcNAc_2$, releasing $Glc_{1-3}Man$. The resulting $Man_8GlcNAc_2$ glycoprotein subsequently re-enters the normal processing route in the Golgi apparatus.

Bacterial GH99 orthologs including *Bacteroides thetaiotaomicron* (*Bt*) and *Bacteroides xylanisolvens* (*Bx*) enzymes possess *endo*-1,2-α-mannosidase activity, but are more appropriately described as *endo*-1,2-α-mannanases, as they act on yeast

[a] Dr. S. P. Schröder, Dr. M. F. Debets, T. Hansen, Dr. T. J. M. Beenakker, Prof. Dr. G. A. van der Marel, Dr. J. D. C. Codée, Prof. Dr. H. S. Overkleeft
Department of Bioorganic Chemistry, Leiden Institute of Chemistry
Einsteinweg 55, 2333 CC Leiden (The Netherlands)
E-mail: h.s.overkleeft@chem.leidenuniv.nl

[b] Dr. W. W. Kallemeijn, Prof. Dr. J. M. F. G. Aerts
Department of Medical Biochemistry, Leiden Institute of Chemistry
Einsteinweg 55, 2333 CC Leiden (The Netherlands)

[c] L. F. Sobala, Prof. Dr. G. J. Davies
Department of Chemistry, York Structural Biology Laboratory
University of York, Heslington, York, YO10 5DD (UK)
E-mail: gideon.davies@york.ac.uk

[d] Dr. Z. Hakki, Prof. Dr. S. J. Williams
School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne
Parkville, Victoria (Australia)

Supporting information and the ORCID identification numbers for the authors of this article can be found under: https://doi.org/10.1002/chem.201801902.

mannan[17] and exhibit a tenfold preference for mannan-based substrates versus the equivalent glucose-substituted mannans.[18] Several imino/azasugar inhibitors for GH99 endomannosidase have been developed, including α-glucopyranosyl-1,3-isofagomine (GlcIFG, **1**, Figure 1) and α-mannopyranosyl-1,3-isofagomine (ManIFG, **2**). Due to a preference for a mannopyranosyl residue in subsite −2 GH99 *endo*-1,2-α-mannanases show a greater affinity for **2** than for **1**.[18] Recently, mannoeuromycin (ManNOE, **3**), which features a 2-hydroxyl allowing interaction with the proposed general base/acid residue, has been reported as the most potent *endo*-1,2-α-mannanase inhibitor for bacterial GH99 enzymes with K_D values in the low nanomolar range.[19] Additionally, fluorescent[20] and fluorogenic[18,21] substrates have been developed for monitoring *endo*-1,2-α-mannosidase/mannanase activity.

Family GH99 *endo*-1,2-α-mannosidases/mannanases cleave their substrate glycosides with retention of anomeric stereochemistry; however, instead of the classical Koshland double-displacement mechanism for retaining enzymes,[22] an unusual neighboring group participation hydrolytic mechanism was proposed in which a glutamate residue (Glu_{333} in *Bx*GH99) acts as a general base assisting OH-2 to displace the aglycon via a 1,2-anhydro sugar that is subsequently hydrolyzed by water (Figure 1 B).[23] In order to study enzyme function in biological settings, screen for inhibitors, as well as to further illuminate the catalytic reaction mechanism, the development of a mechanism-based irreversible inhibitor would be of interest. Here, the synthesis is described of two putative covalent inhibitors **4** and **5**, designed to, respectively inhibit eukaryotic GH99 *endo*-1,2-α-mannosidases and bacterial *endo*-1,2-α-mannanases and which vary in the nature of the pyranoside at the non-reducing end (Figure 1 A, right). Both compounds contain a spiro-epoxide at position C-2 to serve as an electrophile to trap the general base residue. Inspection of the crystal structures of *Bx*GH99 suggests that the general base will be situated close to the methylene group of the spiro-epoxide, where it may open the ring via nucleophilic attack resulting in a covalent intermediate (Figure 1 C).[23] The compounds are also equipped with a reporter tag, allowing active enzyme labeling by activity-based protein profiling (ABPP)[24] protocols, the efficiency of which is reported as well.

Results

Acceptor **7** was synthesized by 4,6-silylidene protection of compound **6**,[25] followed by formation of the 2,3-orthobenzoate and final treatment with acid (Scheme 1).[26] Glucopyranoside donor **9** was synthesized from thiophenyl β-glucopyranoside **8**.[27] While 4,6-silylidene protection proceeded smoothly, elevated temperatures were required to install the TBS-

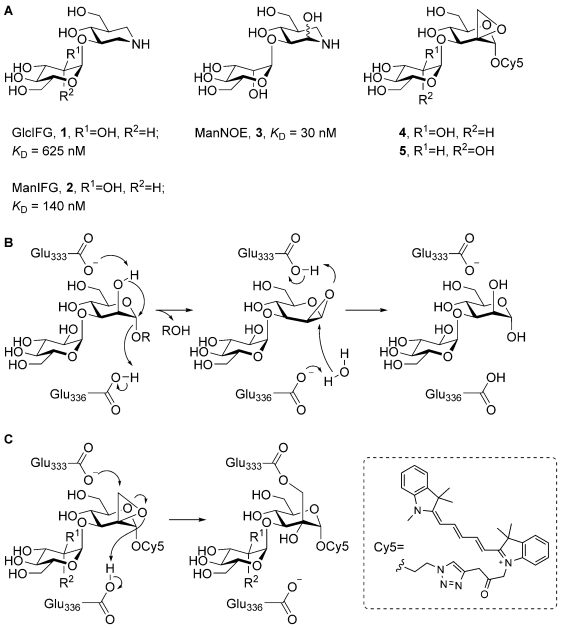
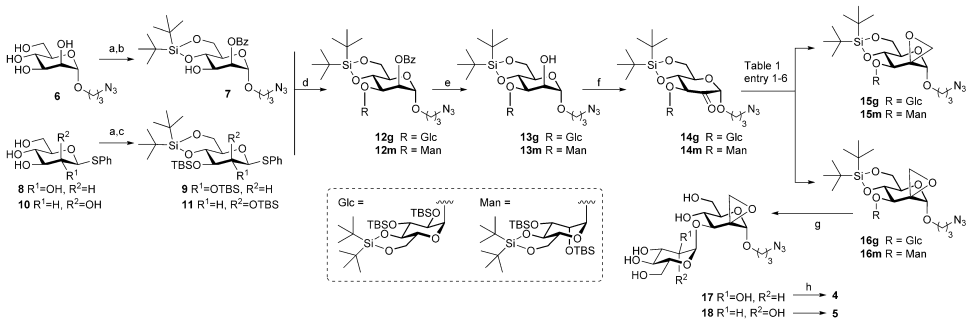


Figure 1. (A) Known GH99 *endo*-1,2-α-mannosidase inhibitors (**1**–**3**) and fluorescent spiro-epoxyglycosides **4** and **5** subject of this study. K_D values are for *B. thetaiotaomicron endo*-1,2-α-mannosidase (*Bt*GH99). (B) The proposed catalytic mechanism for GH99 enzyme (amino acid numbering for *B. xylanisolvens endo*-1,2-α-mannosidase (*Bx*GH99)). (C) Anticipated covalent inhibition mechanism of GH99 enzymes.



Scheme 1. Synthesis of fluorescent spiro-epoxyglycosides **4** and **5**. Reagents and conditions: a) tBu$_2$Si(OTf)$_2$, 2,6-lutidine, DMF, −50 °C; b) PhC(OMe)$_3$, CSA, 2 h, then AcOH, H$_2$O, 16 h, 57% over 2 steps; c) TBSOTf, DMAP, pyridine, 60 °C, 16 h, yield **9**: 85% over 2 steps; yield **11**: 81% over 2 steps; d) donor **9** or **11**, NIS, TMSOTf, DCM, 4 Å MS, −40 °C, 1 h, yield **12g**: 92%; yield **12m**: 88%; e) NaOMe, MeOH, DCM, yield **13g**: 95%; yield **13m**: 86%; f) DMP, DCM, yield **14g**: 98%; yield **14m**: 96%; g) TBAF, THF, 5 days, yield **17**: 97%; yield **18**: 74%; h) Cy5-alkyne,[47] CuSO$_4$·5H$_2$O, sodium ascorbate, DMF, rt, 16 h, yield **4**: 32%; yield **5**: 34%.

groups onto both secondary hydroxyl groups, presumably due to steric hindrance. Using a similar approach, thiophenyl α-mannopyranoside **10**[28] was converted to protected thiomannoside donor **11**.

Glycosylation of acceptor **7** by **9** or **11** was achieved in an *N*-iodosuccinimide (NIS)/trimethylsilyl triflate (TMSOTf) mediated coupling at low temperature, affording **12g** or **12m**, respectively. Both glycosylations proceeded in excellent yield and stereoselectivity. Pedersen, Bols and co-workers[29] recently reported that silylidene-protected mannosyl donors can be used for stereoselective β-mannosylation. The contrasting selectivity obtained here is likely the result of the steric buttressing effect of the large silyl ether protecting groups at the C-2- and C-3-hydroxyls of **7**, consistent with the steric effects that large protecting groups and functionalities have in glycosylations of otherwise β-selective benzylidene-protected mannosyl donors.[30] Thus, the β-face of mannosyl donor **11** is shielded from attack by the incoming nucleophile. The glycosylation stereoselectivity of glucosyl donor **9** can be rationalized by its high reactivity. The "arming" silyl protecting groups allow this donor to readily form an oxocarbenium ion, which will likely take up a $^4H_3$-like conformation, which is preferentially attacked from the α-face to provide the 1,2-*cis*-linked product.[31,32] Next, the benzoyl groups were deprotected under Zemplén conditions affording compounds **13g** and **13m**. The alcohols were then oxidized with Dess–Martin periodinane (DMP) to ketones **14g** and **14m**, which appeared to be in equilibrium with the corresponding hydrates.

Transformation of ketones **14** into their corresponding spiro-epoxides was explored next (Table 1). Reaction of **14g** with diazomethane as methylenating agent[33] resulted in the formation of the equatorial (**15g**) and axial (**16g**) methylenes in a 1:1 ratio and in good yields (entry 1). Their absolute configuration was determined by 1D-NOE difference experiments (see Supplementary Information). Reaction of **14m** with diazomethane also resulted in a mixture of **15m** and **16m**, in a 3:1 ratio, in favor of the equatorial methylene group in almost quantitative yield (entry 2). We anticipated that a Corey–Chaykovsky epoxidation[34] using stabilized dimethylsulfoxonium methylide would favor the formation of the equatorial methylenes **15g** and **15m**. Indeed, also in these cases both isomers were obtained, however the formation of axial methylenes was still favored in both cases (entries 3 and 4). Finally, using the more reactive dimethylsulfonium methylide, only the kinetically favored axial methylenes **16g** and **16m** were formed, albeit in moderate yields (entries 5 and 6). With spiro-epoxides **16g** and **16m** in hand, global deprotection was accomplished by reaction with tetrabutylammonium fluoride (TBAF). Finally, a fluorescent Cy5 tag was installed at the azide handle using copper(I) catalyzed click chemistry, which after HPLC purification afforded spiro-epoxyglycosides **4** and **5**.

Table 1. Transformation of ketones **14g** and **14m** into their corresponding spiro-epoxides.

Entry	s.m.	Conditions	15:16	Yield [%][a]
1	**14g**	CH_2N_2, EtOH, 0 °C	1:1	78
2	**14m**		3:1	97
3	**14g**	$SOMe_3I$, nBuLi, THF, 60 °C	1:5	83
4	**14m**		1:2	85
5	**14g**	SMe_3I, NaH, DMSO, THF, −10 °C	0:1	50
6	**14m**		0:1	53

[a] Combined yield after column chromatography. s.m. = starting material.

The ability of **4** and **5** to label recombinant *Bt*- and *Bx*GH99 *endo*-1,2-α-mannanase was evaluated (Figure 2A). The compounds label both enzymes in a concentration-dependent manner, at concentrations as low as 100 nM. Previous studies indicated a preference for a mannosyl residue at the −2 subsite of both enzymes.[18] However, no difference in potency of labeling was observed. Studies on the effect of the pH dependence on labeling revealed that both spiro-epoxyglycosides label the enzymes maximally at pH 6–8, corresponding to

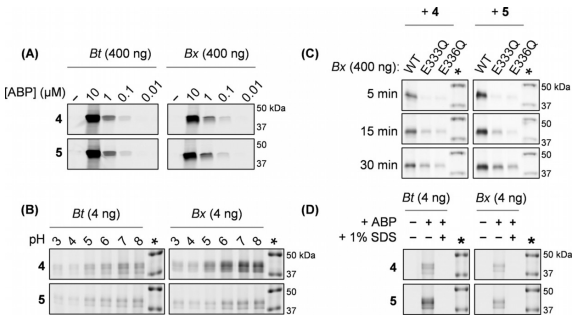


Figure 2. Fluorescent labeling of GH99 endomannanases. (A) Detection limit of *Bt* and *Bx* GH99 endomannanases (left and right, respectively), labeled with various concentrations of fluorescent spiro-epoxyglycosides **4** or **5**. (B) Effect of pH on labeling of *Bt* and *Bx* GH99 enzymes with **4** or **5**. (C) Labeling of wild-type and mutant *Bx*GH99 with **4** or **5** (left or right, respectively) for 5, 15 or 30 minutes. (D) Effect of denaturation with 1% (w/v) SDS and boiling on labeling of *Bt* and *Bx* GH99 enzymes (left and right, respectively) with **4** or **5**. The marker is annotated with an asterisk (*).

the pH optimum of GH99 enzymatic activity (Figure 2B).[23] Notably, more than one band is evident, suggesting enzyme degradation under reducing SDS-PAGE conditions or alternatively that multiple labelling events may be occurring. Next, labeling of wild-type (WT) *Bx*GH99 was compared to analogous active-site mutants (Figure 2C). While WT enzyme is labeled by spiro-epoxyglycosides **4** and **5** within 5 minutes, the general base mutant E333Q and the catalytic acid mutant E336Q were not labeled in the same time period with these compounds, suggesting that labeling is indeed activity-based, and is consistent with reaction occurring in a mechanism-based manner. However, incubation for longer times resulted in labeling of the mutant enzymes, indicating that either the spiro-epoxide is susceptible to ring opening by the mutant catalytic residues, or that other residues may also be involved in covalent labelling. Denaturation of *Bt*GH99 and *Bx*GH99 completely abrogated labeling by spiro-epoxyglycosides **4** and **5**, indicating that labeling requires the natively folded enzyme (Figure 2D).

To further evaluate whether covalent inhibition of *Bt*GH99 and *Bx*GH99 is activity-based, the processing of human α-galactosidase A (GLA) by these enzymes was investigated (Figure 3A). GLA contains three N-glycosylation sites, of which two are decorated with oligo-mannose structures, and one contains complex oligosaccharides low in mannose content.[35,36] We have previously demonstrated that fluorescent α-*galacto*-cyclophellitol aziridines such as TB340 covalently label GLA in activity-based manner.[37] Here, GLA was pre-labeled with TB340 to enable fluorescent detection on gel. Without additives, GLA gives a distinct major band at ≈50 kDa (Figure 3B, lane 1). Incubation of GLA with *Bt*GH99 results in demannosylation of the two high-mannose N-glycans of GLA, resulting in a shift of the GLA band into lower bands at ≈42 kDa (lane 2). This shift in molecular weight is similar to the shift observed when GLA is incubated with Endo-H (lane 4), which causes demannosylation of high-mannose N-glycans by cleaving within the chitobiosyl core leaving a residual GlcNAc on Asn. Treatment of GLA with PNGase-F

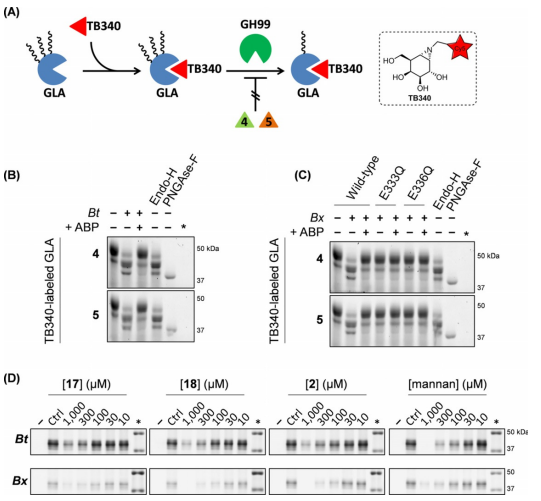


Figure 3. (A) Schematic representation of processing of human α-galactosidase GLA by GH99 endomannosidase. GLA is pre-labeled by fluorescent TB340, and contains high-mannose N-glycans which can be truncated by endomannosidase, resulting in a decrease in GLA molecular weight. Activity-based labeling of endomannosidase by spiro-epoxyglycosides **4** or **5** (prior to incubation with GLA) blocks its activity, and is therefore unable to process GLA. (B) *Bt*GH99 wild-type demannosylates GLA, causing a shift in molecular weight for the protein bands. Pre-labeling *Bt*GH99 wild-type with **4** or **5** abrogates GLA demannosylation. Endo-H cleaves high-mannose structures, PNGase-F cleaves full N-linked glycan (leaving Asp-GlcNAc). (C) *Bx*GH99 wild-type demannosylates GLA, while *Bx*GH99 pre-labeled with **4** or **5** is unable to do so. *Bx*GH99 active-site mutants E333Q and E3336Q are unable to process GLA. (D) Fluorescent labeling of *Bt*GH99 (top) and *Bx*GH99 (bottom) by **4** or **5** competed by different concentrations of **17**, **18**, ManIFG (**2**) and yeast mannan. The marker is annotated with an asterisk (*).

(lane 5), which cleaves most N-glycans leaving Asn, results in a band of a lower molecular weight, most likely as a result of complete deglycosylation of all three N-glycans. When *Bt*GH99 was pre-incubated with **4** or **5**, demannosylation of TB340-labeled GLA was mostly inhibited (lane 3), indicating that binding of **4** and **5** occurs in the *Bt*GH99 active site. An identical experiment was conducted with *Bx*GH99 wildtype and the E333Q and E336Q mutant enzymes (Figure 3C). Similar to *Bt*GH99, wildtype *Bx*GH99 is able to process the N-glycans decorating the surface of the enzyme, giving rise to a shift in molecular weight (lane 2), similar to processing by Endo-H (lane 8). Pre-incubation of WT *Bx*GH99 by **4** or **5**, prior to consecutive incubation with TB340-labeled GLA resulted in an observed absence of glycan processing (lane 3), indicating that binding of **4** and **5** abrogates enzymatic activity. Interestingly, while mutants E333Q and E336Q are labeled by spiro-epoxyglycosides **4** and **5** after prolonged reaction times, they are evidently unable to process TB340-labeled GLA (lanes 4–7).

Finally, the inhibitory potencies of **2**, **17**, **18** and yeast mannan from *S. cerevisiae* (an α-1,6-linked mannose backbone branched with α-1,2 and α-1,3 mannoses)[38] towards *Bt*GH99 were investigated using spiro-epoxyglycoside **5** as fluorescent read-out (Figure 3D). The enzyme was first pre-incubated with the competitor for 30 min at 37 °C, followed by labeling with 1 μM **5**. Compounds **17** and **18** both show a concentration-dependent competition of fluorescent labeling in the range of 10-1,000 μM, although full competition with labeling could not be achieved under these conditions. Similarly, the azasugar ManIFG (**2**) gave concentration-dependent competition but again full competition was not achieved. However, pre-incubation by yeast mannan achieved full competition with labeling, suggesting that processing of spiro-epoxyglycoside **5** by *Bt*GH99 endomannanase is specific and activity based. A similar competition experiment was performed for *Bx*GH99, and it was shown that while pre-incubation with **17** did not fully abrogate labeling, pre-incubation with 1000 μM of **18** provided full competition, possibly hinting at a slight preference for a mannosyl residue in subsite −2. Additionally, yeast mannan showed concentration dependent (albeit incomplete) competition, and azasugar **2** fully competed with labeling at 1,000 μM, suggesting that processing of spiro-epoxyglycoside **5** by *Bx*GH99 is specific and activity-based.

Discussion

Epoxide-based probes have been investigated as mechanism-based inhibitors of a range of glycosidases. Early work led to the development of epoxyalkyl glycosides,[39] which were initially proposed as reagents that could specifically label the nucleophile of retaining glycosidases, however X-ray crystallography later revealed labelling of both acid/base and nucleophile residues.[40] In one classic study, conduritol C epoxide, which was originally believed to label the nucleophile of *E. coli LacZ* β-galactosidase, was subsequently shown to covalently label the acid/base catalyst.[41] Work from our laboratory has investigated related pyranose-mimicking cyclophellitol epoxides and aziridines and shown that these typically exhibit excellent selectivity for labelling the nucleophile of assorted α- and β-glycosidases.[42] We have shown that introduction of a reporter tag (e.g. biotin or a fluorescent dye) onto these small molecule inhibitors affords chemical probes that enable quantification of activity,[43,44] and have distinct advantages over techniques such as transcription analysis and antibody-based detection. We report here the first activity-based probes for detection of GH99 enzymes, which were designed based on the proposed mechanism of this enzyme. In this proposed mechanism a 1,2-anhydro-epoxide intermediate is formed by general base assisted deprotonation of O2 by a carboxylate residue.[23] Our design strategy includes a reactive C2 spiro-epoxide that can potentially covalently label the general base (acting as a nucleophile), and includes a fluorescent label for visualization. Gel-based analysis of labelled bacterial GH99 *endo*-1,2-α-mannanases demonstrated concentration dependent labelling which occurs in a pH dependent manner consistent with the pH optimum of enzyme activity. Labelling could be competed by various substrates and inhibitors, providing evidence that it is active-site directed. While mutation of the key general base and general acid residues inactivated the enzyme towards processing of natural substrate N-glycans in GLA, the mutants could be labelled with the spiro-epoxyglycosides, albeit with reduced potency. Collectively, our data suggests that these spiro-epoxides do result in labelling at the active site, presumeably through the catalytic general base. However, the high reactivity of the primary epoxide means that labelling is most likely not exclusive at a single residue. While *endo*-1,2-α-mannanase has a preference for mannosyl residues at the −2 binding subsite, there was minimal differences in the efficiency of labelling for spiro-epoxides bearing either a mannosyl or glucosyl residue. We believe these compounds represent an important first step in devising probes that take advantage of the unique mechanism proposed for this family. Future studies will seek to better understand the mode of labelling by identifying the covalently labelled residue(s) by X-ray crystallography or MS based techniques. By analogy to previously described irreversible cyclophellitol activity-based probes,[43] we propose these fluorescent spiro-epoxyglycosides could ultimately lead to chemical tools for functional investigation of GH99 *endo*-1,2-α-mannosidase/mannanases, both as isolated species and in tissue extracts.

Experimental Section

Chemicals were purchased from Acros, Sigma Aldrich, Biosolve, VWR, Fluka, Merck and Fisher Scientific and used as received unless stated otherwise. Tetrahydrofuran (THF), *N*,*N*-dimethylformamide (DMF) and toluene were stored over molecular sieves before use. All reactions were performed under an argon atmosphere unless stated otherwise. TLC analysis was conducted using Merck aluminum sheets (Silica gel 60 F_{254}) with detection by UV absorption (254 nm), by spraying with a solution of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ (25 g L^{-1}) and $(NH_4)_4Ce(SO_4)_4{\cdot}2H_2O$ (10 g L^{-1}) in 10% sulfuric acid, followed by charring at ≈150 °C. Column chromatography was performed using Screening Device b.v. silica gel (particle size of 40–63 μm, pore diameter of 60 Å) with the indicated eluents. For reversed-phase HPLC purifications an Agilent Tech-

nologies 1200 series instrument equipped with a semi-preparative column (Gemini C18, 250×10 mm, 5 µm particle size, Phenomenex) was used. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Brüker AV-400 (400 and 101 MHz, respectively) or a Brüker DMX-600 (600 and 151 MHz, respectively) spectrometer in the given solvent. Chemical shifts are given in ppm (δ) relative to the residual solvent peak or tetramethylsilane (0 ppm) as internal standard. High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electronspray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL min^{-1}, capillary temperature 250 °C) with resolution $R=60\,000$ at m/z 400 (mass range m/z 150–2000) and dioctyl phthalate (m/z 391.28428) as lock mass. The high-resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan). ManIFG was prepared as previously reported.[18] Recombinant expression of *B. thetaiotaomicron* (*Bt*) and *B. xylanisolvens* (*Bx*) GH99 was achieved as previously described.[23] Recombinant α-galactosidase (GLA) was purchased from Genzyme (Cambridge, MA, USA). The α-galactosidase ABP TB340 was synthesized as described earlier.[37] Yeast mannan from *S. cerevisiae* was purchased from Sigma.

Synthesis and characterization

(4*aR*,6*S*,7*S*,8*R*,8*aS*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-hydroxyhexahydropyrano[3,2-*d*][1,3,2]dioxasilin-7-yl benzoate (7): Compound **6**[45] (1.00 g, 3.80 mmol) was co-evaporated with dry toluene and dissolved in dry DMF (38 mL). The resulting solution was cooled to −50 °C and Si*t*Bu$_2$(OTf)$_2$ (1.11 mL, 3.42 mmol, 0.9 EQ) and 2,6-lutidine (0.44 mL, 3.80 mmol) were added. The reaction was stirred at −50 °C for 30 minutes and subsequently quenched with brine (400 mL). The aqueous layer was extracted with Et$_2$O (4×100 mL). The combined organic layers were washed with 1 M aqueous HCl (2×100 mL), H$_2$O (100 mL), and brine and dried over Na$_2$SO$_4$. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (EtOAc/pentane, 1:4 to 1:2). The 4,6-silydene product was obtained as white solid (970 mg, 70%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta=$ 4.81 (d, $J=1.4$ Hz, 1H), 4.11 (dd, $J=10.0$, 5.0 Hz, 1H), 4.07–4.00 (m, 2H), 3.96 (t, $J=10.2$ Hz, 1H), 3.86–3.76 (m, 2H), 3.69 (td, $J=10.0$, 5.0 Hz, 1H), 3.50 (ddd, $J=10.0$, 6.3, 5.2 Hz, 1H), 3.40 (td, $J=6.6$, 2.0 Hz, 2H), 1.94–1.82 (m, 2H), 1.06 (s, 9H), 1.00 ppm (s, 9H). ^{13}C NMR (101 MHz, CDCl$_3$): $\delta=166.0$, 133.4, 129.9, 129.7, 128.5, 98.1, 75.2, 72.0, 70.2, 67.4, 66.6, 64.6, 48.2, 28.8, 27.4, 27.0, 22.8, 20.0 ppm. IR (neat): $\tilde{\nu}=$3524, 2934, 2886, 2097, 1732, 1717, 1558, 1472, 1267, 1095, 1072, 1026, 885, 826, 710, 654 cm^{-1}. $[\alpha]_D^{20}$ (c 0.1, DCM): −16. HRMS (ESI) m/z: $[M+Na]^+$ calcd for $C_{24}H_{37}N_3O_7SiNa$ 530.22930, found 530.22907. The 4,-6-silydene compound (889 mg, 2.20 mmol) was dissolved in trimethyl orthobenzoate (5.7 mL) and CSA (102 mg, 0.44 mmol) was added. The reaction was stirred for 2 hours at room temperature and cooled to 0 °C. Aqueous AcOH (50%, 20 mL) was added and the mixture was stirred overnight while the cooling bath was allowed to reach room temperature. The solution was poured into saturated aqueous NaHCO$_3$ (50 mL) and the water layer was extracted with CH$_2$Cl$_2$ (3×50 mL). The combined organic layers were washed with NaHCO$_3$ (50 mL) and dried over MgSO$_4$. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (EtOAc/pentane, 1:99 to 1:10). The title product was obtained as colorless oil (922 mg, 82%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta=$8.15–7.98 (m, 2H), 7.63–7.54 (m, 1H), 7.51–7.41 (m, 2H), 5.42 (dd, $J=3.4$, 1.6 Hz, 1H), 4.88 (d, $J=1.4$ Hz, 1H), 4.23–4.06 (m, 3H), 3.99 (t, $J=10.2$ Hz, 1H), 3.86–3.76 (m, 2H), 3.59–3.49 (m, 1H), 3.43 (t, $J=6.6$ Hz, 2H), 1.99–1.81 (m, 2H), 1.09 (s, 9H), 1.02 ppm (s, 9H). ^{13}C NMR (101 MHz, CDCl$_3$): $\delta=166.0$, 133.4, 129.9, 129.7, 128.5, 98.1, 75.2, 72.0, 70.2, 67.4, 66.6, 64.6, 48.2, 28.8, 27.4, 27.0, 22.8, 20.0 ppm. IR (neat): $\tilde{\nu}=$3524, 2934, 2886, 2097, 1732, 1717, 1558, 1472, 1267, 1095, 1072, 1026, 885, 826, 710, 654 cm^{-1}. $[\alpha]_D^{20}$ (c 0.1, DCM): −16. HRMS (ESI) m/z: $[M+Na]^+$ calcd for $C_{24}H_{37}N_3O_7SiNa$ 530.22930, found 530.22907.

(4*aR*,6*S*,7*R*,8*S*,8*aR*)-2,2-Di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)-6-(phenylthio)hexahydropyrano[3,2-*d*][1,3,2]dioxasiline (9): Compound **8**[27] (2.6 g, 9.5 mmol) was dissolved in dry DMF (100 mL) under Ar-atmosphere. The mixture was cooled to −50 °C and 2,6-lutidine (3.3 mL, 28.5 mmol) and Si*t*Bu$_2$(OTf)$_2$ (3.4 mL, 10.5 mmol) was added. The reaction was stirred for 2 hours at −50 °C and subsequently quenched with H$_2$O (100 mL). The water layer was extracted with EtOAc (3×100 mL). The organic layers were combined and washed with H$_2$O (2×200 mL) and brine (200 mL) and dried over MgSO$_4$. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (EtOAc/pentane, 1:4 to 1:2). The 4,6-silydene product was obtained as a white solid (3.58 g, 91%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta=$7.57–7.46 (m, 2H), 7.38–7.28 (m, 3H), 4.60 (d, $J=9.7$ Hz, 1H), 4.21 (dd, $J=10.2$, 5.1 Hz, 1H), 3.90 (t, $J=10.2$ Hz, 1H), 3.68 (t, $J=9.0$ Hz, 1H), 3.60 (t, $J=8.7$ Hz, 1H), 3.51–3.37 (m, 2H), 2.92 (s, 1H), 2.77 (s, 1H), 1.04 (s, 9H), 0.98 ppm (s, 9H). ^{13}C NMR (101 MHz, CDCl$_3$): $\delta=132.9$, 131.7, 129.1, 128.3, 88.6, 77.8, 76.4, 74.5, 71.8, 66.1, 27.4, 27.0, 22.7, 19.9. IR (neat): $\tilde{\nu}=$3241, 2932, 2858, 1695, 1471, 1058 cm^{-1}. $[\alpha]_D^{20}$ (c 0.06, DCM): −57.0. HRMS (ESI) m/z: $[M+Na]^+$ calcd for $C_{20}H_{32}O_5SSiNa$ 435.16319, found 435.16315. The 4,6-silydene product (1.0 g, 2.42 mmol) was co-evaporated with toluene (3×), dissolved in dry pyridine (5 mL) and cooled to 0 °C. DMAP (30 mg, 0.24 mmol) and TBSOTf (3.33 mL, 14.5 mmol) were added and the mixture was heated to 60 °C and stirred overnight. The mixture was carefully diluted with water (25 mL) and extracted with DCM (3× 50 mL). The combined organic layers were washed with aq. 1 M HCl (3×25 mL) and brine, dried over Na$_2$SO$_4$, filtrated and concentrated. The crude product was purified by gradient column chromatography (pentane/EtOAc, 400:1 to 200:1), affording the title product as a white solid (1.44 g, 93%). Analytical data were in accordance with those reported in literature.[46]

(4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-Di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)-6-(phenylthio)hexahydropyrano[3,2-*d*][1,3,2]dioxasiline (11): The 4,6-silydene compound was prepared from **10**[28] (4.9 g, 18 mmol) as described for the preparation of **9** to afford the product (6.6 g, 89%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$): $\delta=$7.52 −7.20 (m, 5H), 5.53 (s, 1H), 4.30 (d, $J=3.1$ Hz, 1H), 4.24 (td, $J=$ 10.0, 5.0 Hz, 1H), 4.15–4.08 (t, $J=9.4$ Hz, 1H), 4.05 (dd, $J=10.0$, 5.0 Hz, 1H), 3.96 (t, $J=10.1$ Hz, 1H), 3.87 (dd, $J=9.1$, 3.3 Hz, 1H), 2.67 (brs, 2xOH), 1.05 (s, 9H), 1.03 ppm (s, 9H). ^{13}C NMR (101 MHz, CDCl$_3$): $\delta=133.9$, 131.5, 129.3, 127.7, 87.8, 75.0, 72.4, 72.1, 67.9, 66.2, 27.6, 27.2, 22.8, 20.2 ppm. IR (neat): $\tilde{\nu}=$3384, 2932, 2858, 1474, 1064 cm^{-1}. $[\alpha]_D^{20}$ (c 0.4, DCM): +227. HRMS (ESI) m/z: $[M+$ Na$]^+$ calcd for $C_{20}H_{32}O_5SSiNa$ 435.16319, found 435.16301. The title product was prepared from the 4,6-silydene compound (6.6 g, 16 mmol) as described for the preparation of 9 to afford the product (9.3 g, 91%) as a pale yellow oil which crystallized at −20 °C. ^{1}H NMR (400 MHz, CDCl$_3$): $\delta=$7.48–7.25 (m, 5H), 5.29 (d, $J=1.5$ Hz, 1H), 4.28 (t, $J=9.0$ Hz, 1H), 4.19 (m, 1H), 4.17–4.11 (m, 1H), 4.11–4.08 (m, 1H), 3.96 (t, $J=9.7$ Hz, 1H), 3.87 (dd, $J=8.9$, 2.5 Hz, 1H), 1.09 (s, 9H), 1.07 (s, 9H), 0.99 (s, 9H), 0.92 (s, 9H), 0.21 (s, 3H), 0.18 (s, 3H), 0.14 (s, 3H), 0.07 ppm (s, 3H). ^{13}C NMR (101 MHz, CDCl$_3$): $\delta=134.9$, 131.3, 129.3, 127.4, 89.9, 75.0, 74.6, 73.0, 69.6, 67.1, 27.8, 27.3, 26.3, 25.8, 22.9, 20.2, 18.5, 18.2, −3.9, −4.1, −4.4, −4.4 ppm. IR (neat): $\tilde{\nu}=$2931, 2857, 1471, 1250, 1096 cm^{-1}. $[\alpha]_D^{20}$ (c 1.0, DCM):

+91. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{32}H_{61}O_5SSi_3$ 641.35420, found 641.36460.

(4*aR*,6*S*,7*S*,8*R*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*R*,8*S*,8a*R*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-7-yl benzoate (12 g): Compound **9** (2.00 g, 3.12 mmol) and compound 7 (1.58 g, 3.12 mmol) were combined and co-evaporated with toluene (3×). The mixture was dissolved in dry CH_2Cl_2 (20 mL) and stirred with activated 4A MS for 30 minutes at room temperature. The reaction was cooled to −50 °C and NIS (842 mg, 3.74 mmol) and TMSOTf (68 μL, 0.37 mmol) were added. The reaction mixture was warmed to −40 °C, stirred for 1 hour and subsequently neutralized with NEt_3 (2 mL). The mixture was diluted with CH_2Cl_2 (200 mL) and washed with saturated aqueous Na_2SO_3 (2×100 mL), H_2O (100 mL) and subsequently dried over $MgSO_4$. The solvents were removed under reduced pressure and the crude product was purified by gradient column chromatography (EtOAc/pentane, 1:50 to 1:40). The title product was obtained as a white foam (2.98 g, 92%). ^{1}H NMR (400 MHz, $CDCl_3$): δ = 8.08 (d, *J* = 7.3 Hz, 2H), 7.59 (t, *J* = 7.4 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 2H), 5.49 (s, 1H), 5.16 (d, *J* = 2.9 Hz, 1H, H-1 "donor"), 4.86 (s, 1H, H-1 "acceptor"), 4.44 (t, *J* = 9.4 Hz, 1H), 4.22–4.15 (m, 2H, H-3), 4.04 (t, *J* = 10.2 Hz, 1H), 3.95–3.79 (m, 3H), 3.75–3.66 (m, 2H), 3.61 (t, *J* = 8.4 Hz, 1H), 3.54 (dt, *J* = 10.2, 5.7 Hz, 1H), 3.49–3.41 (m, 2H), 1.91 (dq, *J* = 13.5, 6.9 Hz, 1H), 1.13 (s, 9H), 1.04 (s, 9H), 1.03 (s, 9H), 0.97 (s, 9H), 0.91 (s, 9H), 0.79 (s, 9H), 0.17 (s, 3H), 0.05 (s, 3H), 0.03 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 165.4, 133.2, 129.9, 129.6, 128.5, 98.1, 97.9, 78.6, 75.2, 74.3, 73.4, 72.6, 71.1, 67.8, 67.8, 66.9, 66.6, 64.4, 48.1, 28.9, 27.5, 27.1, 27.0, 26.4, 26.2, 22.7, 22.7, 20.0, 20.0, 18.1, 18.0, −3.2, −3.5, −3.6, −4.4 ppm. ^{13}C-HMBC-GATED NMR (101 MHz, $CDCl_3$): δ = 98.1 ($J_{C1,H1}$ = 170.6 Hz, C1 "donor"), 97.9 ($J_{C1,H1}$ = 172.1 Hz, C1 "acceptor"). IR (neat): $\tilde{\nu}$ = 2966, 2859, 2093, 1732, 1472, 1260, 1096, 1069, 1045, 827 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +20. HRMS (ESI) *m/z*: $[M+Na]^+$ calcd for $C_{50}H_{91}N_3O_{12}Si_4$ + Na 1060.55720, found 1060.55694.

(4*aR*,6*S*,7*S*,8*R*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-7-yl benzoate (12 m): This compound was prepared from **11** (378 mg, 0.59 mmol) and 7 (299 mg, 0.59 mmol) as described for the preparation of **12 g**, to afford the title product (538 mg, 88%) as a pale yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 8.02 (m, 2H), 7.62–7.53 (m, 1H), 7.45 (t, *J* = 7.7 Hz, 2H), 5.34 (dd, *J* = 3.5, 1.6 Hz, 1H), 4.96 (d, *J* = 1.9 Hz, 1H), 4.82 (d, *J* = 1.4 Hz, 1H), 4.28 (t, *J* = 9.5 Hz, 1H), 4.15 (m, 3H), 4.08 (dd, *J* = 9.4, 3.6 Hz, 1H), 4.02–3.93 (t, *J* = 10.3 Hz, 1H), 3.90–3.77 (m, 4H), 3.77–3.67 (m, 2H), 3.57–3.37 (m, 3H), 2.02–1.77 (m, 2H), 1.09 (s, 9H), 1.04–1.00 (s, 9H), 1.00 (s, 9H), 0.95–0.89 (m, 9H), 0.87–0.82 (m, 9H), 0.82–0.76 (m, 9H), 0.07 (s, 3H), 0.00 (s, 3H), −0.13 (s, 3H), −0.17 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 165.5, 133.4, 130.0, 129.7, 128.6, 103.4, 98.2, 75.2, 75.1, 74.0, 73.6, 72.4, 71.8, 69.5, 67.7, 67.5, 67.1, 64.8, 48.4, 29.0, 27.9, 27.7, 27.2, 27.2, 26.2, 25.8, 22.9, 22.9, 20.1, 19.9, 18.4, 18.2, −4.3, −4.4, −4.4, −4.7 ppm. ^{13}C-HMBC-GATED NMR (101 MHz, $CDCl_3$): δ = 103.4 ($J_{C1,H1}$ = 172.1 Hz, C1 "donor"), 98.2 ppm ($J_{C1,H1}$ = 172.5 Hz, C1 "acceptor"). IR (neat): $\tilde{\nu}$ = 2931, 2858, 20998, 1729, 1472, 1226, 1096, 1068 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.4, DCM): +1. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{50}H_{92}N_3O_{12}Si_4$ 1038.57526, found 1038.57587.

(4*aR*,6*S*,7*S*,8*R*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*R*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-7-ol (13 g): Compound **12 g** (610 mg, 0.59 mmol) was co-evaporated with toluene (3×) and dissolved in a mixture of DCM/MeOH (9 mL, 1:1). NaOMe (30 wt%, 560 μL) was added and the reaction mixture was stirred for 24 h. The reaction was neutralized with AcOH and the solvents were removed under reduced pressure. The crude product was purified by gradient column chromatography (EtOAc/pentane, 1:11 to 1:8). The title product was obtained as a white foam (519 mg, 95%). ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.34 (d, *J* = 3.1 Hz, 1H), 4.81 (d, *J* = 0.7 Hz, 1H), 4.29 (t, *J* = 9.3 Hz, 1H), 4.10–4.02 (m, 2H), 3.98 (t, *J* = 10.3 Hz, 1H), 3.95 (s, 1H), 3.88 (dd, *J* = 9.2, 3.3 Hz, 1H), 3.86–3.76 (m, 3H), 3.76–3.66 (m, 3H), 3.58 (dd, *J* = 8.2, 3.1 Hz, 1H), 3.54–3.47 (m, 1H), 3.38 (td, *J* = 6.5, 1.7 Hz, 2H), 3.00 (s, 1H, OH), 1.94–1.78 (m, 2H), 1.05 (s, 9H), 1.04 (s, 9H), 1.00 (s, 9H), 0.98 (s, 9H), 0.93 (s, 9H), 0.92 (s, 9H), 0.14 (s, 3H), 0.13 (s, 3H), 0.11 (s, 3H), 0.09 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 99.7, 97.4, 78.7, 75.1, 74.6, 74.5, 74.3, 71.1, 67.6, 67.4, 67.0, 66.4, 64.4, 48.4, 29.0, 27.6 (3×), 27.5 (3×), 27.2 (3×), 27.1 (3×), 26.4 (3×), 26.4 (3×), 22.9, 22.7, 20.1, 20.1, 18.3, 18.3, −3.1, −3.3, −3.4, −3.9 ppm. IR (neat): $\tilde{\nu}$ = 2931, 2856, 2099, 1472, 1252, 1132, 1095, 1069, 1043, 868, 827, 772, 654 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +44. HRMS (ESI) *m/z*: $[M+Na]^+$ calcd for $C_{43}H_{87}N_3O_{11}Si_4$ + Na 956.53099, found 956.53097.

(4*aR*,6*S*,7*S*,8*R*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-7-ol (13 m): This compound was prepared from **12 m** (501 mg, 0.48 mmol) as described for the preparation of **13 g** to afford the title product (386 mg, 86%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.00 (d, *J* = 1.9 Hz, 1H), 4.79 (d, *J* = 1.1 Hz, 1H), 4.17 (t, *J* = 9.2 Hz, 1H), 4.11 (m, 3H), 3.99–3.88 (m, 4H), 3.88–3.82 (m, 2H), 3.82–3.77 (m, 1H), 3.77–3.62 (m, 2H), 3.49 (m, 1H), 3.40 (td, *J* = 6.5, 3.1 Hz, 1H), 2.37–2.03 (brs, OH), 1.97–1.77 (m, 2H), 1.04 (m, 18H), 0.99 (s, 9H), 0.97 (s, 9H), 0.93 (s, 9H), 0.86 (s, 9H), 0.12 (s, 3H), 0.11 (s, 3H), 0.10 (s, 3H), 0.02 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 103.2, 99.6, 77.7, 74.5, 74.1, 73.3, 72.4, 71.5, 69.6, 67.4, 67.33, 66.8, 64.5, 48.4, 28.9, 27.8, 27.6, 27.2, 27.1, 26.3, 25.8, 22.9, 22.7, 20.1, 18.6, 18.2, −3.9, −4.1, −4.3, −4.6. ppm IR (neat): $\tilde{\nu}$ = 2930, 2858, 2098, 1472, 1250, 1096, 1031 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.4, DCM): +32. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{43}H_{88}N_3O_{11}Si_4$ 934.54904, found 934.54959.

(4*aR*,6*S*,8*S*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*R*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydropyrano[3,2-*d*][1,3,2]dioxasilin-7(6*H*)-one (14 g): Compound **13 g** (2.20 g, 2.36 mmol) was co-evaporated with dry toluene (3×) and dissolved in dry CH_2Cl_2 (65 mL). Dess–Martin periodinane (2.00 g, 4.71 mmol) was added and the mixture was stirred overnight. Celite was added and the solvents were removed under reduced pressure. The product was purified by gradient column chromatography (EtOAc/pentane, 1:70 to 1:4). The title product was obtained as a white foam (2.15 g, 98%). ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.18 (d, *J* = 2.8 Hz, 1H), 4.73–4.72 (m, 2H), 4.24–4.02 (m, 5H), 4.02–3.90 (m, 1H), 3.90–3.73 (m, 3H), 3.67 (t, *J* = 8.6 Hz, 1H), 3.62–3.54 (m, 2H), 3.40 (t, *J* = 6.6 Hz, 2H), 1.94–1.81 (m, 2H), 1.06 (s, 9H), 1.04 (s, 9H), 1.02 (s, 9H), 0.98 (s, 9H), 0.93 (s, 9H), 0.92 (s, 9H), 0.17 (s, 3H), 0.12 (s, 3H), 0.11 (s, 3H), 0.08 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 196.2, 100.2, 98.2, 80.0, 78.9, 78.7, 74.9, 73.9, 67.49, 67.45, 67.0, 66.0, 65.3, 48.0, 27.5, 27.4 (3×), 27.1 (3×), 27.0 (3×), 26.5 (3×), 26.41 (3×), 22.7, 22.6, 20.0, 20.0, 18.3, 18.1, −3.0, −3.5, −3.7, −3.9 ppm. IR (neat): $\tilde{\nu}$ = 2932, 2859, 2099, 1757, 1474, 1387, 1362, 1252, 1161, 1093, 1070, 1043, 866, 827, 775, 652 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +50. HRMS (ESI) *m/z*: $[M+Na]^+$ calcd for $C_{43}H_{85}N_3O_{11}Si_4$ + Na 954.51534, found 954.51535.

**(4*aR*,6*S*,8*S*,8*aR*)-6-(3-Azidopropoxy)-2,2-di-*tert*-butyl-8-(((4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethyl-

silyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydropyrano[3,2-*d*][1,3,2]dioxasilin-7(6*H*)-one (14m): This compound was prepared from **13m** (355 mg, 0.38 mmol) as described for the preparation of **14g** to afford the title product (340 mg, 96%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 4.81 (d, J = 1.9 Hz, 1H), 4.75 (s, 1H), 4.45 (d, J = 9.1 Hz, 1H), 4.31–4.19 (m, 2H), 4.18–4.12 (t, J = 9.2 Hz, 1H), 4.12–3.98 (m, 4H), 3.97–3.88 (m, 2H), 3.88–3.78 (m, 2H), 3.61–3.52 (dt, J = 9.9, 5.4 Hz, 1H), 3.41 (t, J = 6.5 Hz, 2H), 2.00–1.76 (m, 2H), 1.05 (s, 9H), 1.05 (s, 9H), 1.04 (s, 9H), 1.00 (s, 9H), 0.93 (s, 9H), 0.86 (s, 9H), 0.15 (s, 3H), 0.12 (s, 3H), 0.09 (s, 3H), 0.01 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 195.8, 103.4, 100.3, 82.2, 79.5, 74.4, 73.3, 72.1, 68.9, 67.3, 67.1, 66.6, 65.3, 48.0, 28.8, 27.6, 27.4, 27.1, 27.0, 26.2, 25.7, 22.8, 22.7, 20.1, 20.0, 18.4, 18.1, −4.0, −4.2, −4.5, −4.7 ppm. IR (neat): $\tilde{\nu}$ = 2933, 2858, 2087, 1755, 1471, 1254, 1155 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.4, DCM): +47. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{43}H_{86}N_3O_{11}Si_4$ 932.53339, found 932.53363.

(2*S*,4*a'R*,6'*S*,8'*S*,8*a'R*)-6'-(3-Azidopropoxy)-2',2'-di-*tert*-butyl-8'-(((4*aR*,6*R*,7*R*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydro-6'*H*-spiro[oxirane-2,7'-pyrano[3,2-*d*][1,3,2]dioxasiline] (15g) and (2*R*,4*a'R*,6'*S*,8'*S*,8*a'R*)-6'-(3-azidopropoxy)-2',2'-di-*tert*-butyl-8'-(((4*aR*,6*R*,7*R*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydro-6'*H*-spiro[oxirane-2,7'-pyrano[3,2-*d*][1,3,2]dioxasiline] (16g): Method A: *dimethyl sulfonium methylide:* A 1 M solution of dimsyl-sodium was prepared from sodium hydride (60 wt%, 200 mg, 5 mmol) in dry DMSO (2.5 mL) and heating this mixture to 70 °C for 1 h. The olive-green solution was cooled to room temperature and diluted with dry THF (2.5 mL). A fraction (0.12 mL, 0.19 mmol) of this mixture was added to a dried flask and cooled on an ice-salt bath. Then, a solution of trimethylsulfonium iodide (26.3 mg, 0.129 mmol) in dry DMSO (0.43 mL) and dry THF (0.4 mL) was added drop wise and the mixture was stirred for 5 minutes. Then, compound 14g (100 mg, 0.11 mmol, co-evaporated with toluene (3×) beforehand) in dry THF (0.64 mL) was added and the mixture was stirred for 30 minutes. The mixture was diluted with water (20 mL) and extracted with Et_2O/pentane (2:1, 4x 15 mL). The combined organic layers were washed with water (20 mL), dried over Na_2SO_4, filtrated and concentrated. The crude product was purified by gradient column chromatography (pentane/EtOAc, 60:1 to 50:1) to afford solely product 15g (51 mg, 50%) as an oil.

Method B: *dimethyl sulfoxonium methylide:* Trimethylsulfoxonium iodide (37.8 mg, 0.172 mmol) was suspended in dry THF (2 mL) and cooled to 0 °C. *n*Butyllithium (2 M in pentane, 80 μL, 0.16 mmol) was added and the mixture was heated to 60 °C. Compound **14g** (100 mg, 0.11 mmol) was co-evaporated with toluene (3×), dissolved in dry THF (1 mL) and added drop wise to the ylide solution. After 10 minutes, the mixture was cooled to room temperature and quenched with MeOH (0.5 mL). The mixture was evaporated and the crude product was purified by gradient column chromatography (pentane/EtOAc, 60:1) to give a mixture of compounds **15g** and **16g** (90 mg, ratio **15g**:**16g** 1:5, total yield 88%) as a colorless oil.

Method C: *diazomethane:* To a glass tube were added aq. KOH (40%, 5 mL) and Et_2O (20 mL) and this mixture was cooled to 0 °C. Then, 1-methyl-3-nitro-1-nitrosoguanidine (2.9 g, 10 mmol) was added in portions with swirling. A fraction (2 mL) of the bright yellow ether layer was added drop-wise to a solution of compound **14g** (100 mg, 0.11 mmol) in EtOH (3 mL) at 0 °C. After stirring for 10 minutes, acetic acid (glacial) was added drop wise until the yellow mixture turned colorless. The mixture was concentrated and co-evaporated with toluene (3×). The crude products were purified by column chromatography (pentane/acetone, 150:1), affording compound **15g** and **16g** (79 mg, ratio **15g**:**16g** 1:1, total yield 78%). Data for compound **15g** (equatorial methylene): ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.33 (d, J = 2.9 Hz, 1H), 4.27–4.15 (m, 3H), 4.18 (s, 1H), 4.10–4.01 (m, 2H), 4.00–3.95 (t, J = 10.1 Hz, 1H), 3.92–3.76 (m, 3H), 3.75–3.59 (m, 4H), 3.50–3.35 (m, 3H), 3.15 (d, J = 5.1 Hz, 1H), 2.61 (d, J = 5.1 Hz, 1H), 1.86 (quintet, J = 6.3 Hz, 2H), 1.04 (s, 9H), 1.04 (s, 9H), 0.98 (s, 18H), 0.91 (s, 9H), 0.90 (s, 9H), 0.13 (s, 3H), 0.10 (s, 3H), 0.09 (s, 3H), 0.08 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 102.5, 97.4, 79.3, 78.5, 76.0, 74.8, 68.7, 67.6, 67.1, 66.3, 64.2, 59.4, 48.3, 46.4, 29.9, 29.1, 27.6, 27.5, 27.2, 27.1, 26.3, 22.9, 22.7, 20.1, 18.4, 18.3, −3.6, −3.6, −4.0 ppm. IR (neat): $\tilde{\nu}$ = 2930, 2858, 2099, 1472, 1252, 1091, 1043 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +54. HRMS (ESI) *m/z*: $[M+Na]^+$ calcd for $C_{16}H_{27}N_3O_{11}Na$ 968.53099, found 968.53089.Data for compound **16g** (axial methylene) ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.34 (d, J = 3.0 Hz, 1H), 4.40–4.22 (m, 2H), 4.09 (dd, J = 10.4, 5.1 Hz, 2H), 4.04–3.93 (m, 3H), 3.87–3.76 (m, 4H), 3.66 (t, J = 8.7 Hz, 1H), 3.56–3.39 (m, 4H), 3.23 (d, J = 5.6 Hz, 1H), 2.63 (d, J = 5.6 Hz, 1H), 1.89 (q, J = 5.7 Hz, 2H), 1.07 (s, 18H), 1.05 (s, 9H), 1.02 (s, 9H), 0.96 (s, 9H), 0.95 (s, 9H), 0.16 (s, 3H), 0.13 (s, 3H), 0.11 (s, 3H), 0.10 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 101.2, 96.9, 79.7, 78.2, 74.7, 73.8, 67.7, 67.4, 66.9, 66.8, 66.1, 64.4, 58.3, 48.1, 48.0, 29.0, 27.6 (3×), 27.3 (3×), 27.2 (3×), 27.0 (3×), 26.5 (3×), 26.4 (3×), 22.8, 22.5, 20.1, 20.0, 18.2, 18.1, −3.08, −3.38 (2×), −4.03 ppm. IR (neat): $\tilde{\nu}$ = 2934, 2858, 2320, 2094, 1095, 1043, 827, 773 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +98 (c 0.1, DCM). HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{44}H_{88}N_3O_{11}Si_4$ 946.54904, found 946.54953.

(2*S*,4*a'R*,6'*S*,8'*S*,8*a'R*)-6'-(3-Azidopropoxy)-2',2'-di-*tert*-butyl-8'-(((4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydro-6'*H*-spiro[oxirane-2,7'-pyrano[3,2-*d*][1,3,2]dioxasiline] (15m) and (2*R*,4*a'R*,6'*S*,8'*S*,8*a'R*)-6'-(3-azidopropoxy)-2',2'-di-*tert*-butyl-8'-(((4*aR*,6*R*,7*S*,8*S*,8*aR*)-2,2-di-*tert*-butyl-7,8-bis((*tert*-butyldimethylsilyl)oxy)hexahydropyrano[3,2-*d*][1,3,2]dioxasilin-6-yl)oxy)tetrahydro-6'*H*-spiro[oxirane-2,7'-pyrano[3,2-*d*][1,3,2]dioxasiline] (16m): Compounds **15m** and **16m** were prepared from **14m** as described for the preparation of 15g and 16g, and could be separated by careful column chromatography. *Method A*: Starting from compound 14m (100 mg, 0.11 mmol), the product **16m** (54 mg, 53%) was obtained as the single product. *Method B*: Starting from compound **14m** (150 mg, 0.16 mmol), product **15m** and **16m** were obtained as a mixture (129 mg, 85%, ratio **15m**:**16m** = 1:2). *Method C:* Starting from compound **14m** (265 mg, 0.28 mmol), product **15m** and **16m** were obtained as a mixture (263 mg, 98%, ratio **15m**:**16m** = 3:1). Data for **15m** (equatorial methylene): ^{1}H NMR (400 MHz, $CDCl_3$): δ = 5.12 (d, J = 2.0 Hz, 1H), 4.28 (d, J = 9.4 Hz, 1H), 4.22–4.14 (m, 3H), 4.14–4.05 (m, 2H), 4.00–3.93 (t, J = 10.3 Hz, 1H), 3.93–3.83 (m, 3H), 3.83–3.77 (m, 2H), 3.55–3.43 (m, 2H), 3.40 (m, 2H), 3.08 (d, J = 4.7 Hz, 1H), 2.67 (d, J = 4.7 Hz, 1H), 1.94–1.82 (m, 2H), 1.05 (s, 9H), 1.04 (s, 9H), 1.00 (s, 9H), 0.99 (s, 9H), 0.94–0.90 (m, 9H), 0.84 (s 9H), 0.14 (s, 3H), 0.11 (s, 3H), 0.10 (s, 3H), 0.00 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 102.2, 101.9, 78.2, 74.4, 73.4, 72.1, 70.8, 69.7, 68.2, 67.3, 66.5, 64.2, 58.9, 48.2, 46.6, 28.8, 27.7, 27.5, 27.01, 26.2, 25.6, 22.7, 22.6, 20.0, 19.9, 18.4, 18.1, −4.0, −4.2, −4.2, −4.8 ppm. IR (neat): $\tilde{\nu}$ = 2929, 2098, 1741, 1251, 1161, 1099 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.05, DCM): +38. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{44}H_{88}N_3O_{11}Si_4$ 946.54904, found 946.54940. Data for compound 16m (axial methylene): ^{1}H NMR (400 MHz, $CDCl_3$): δ = 4.94 (d, J = 2.0 Hz, 1H), 4.28 (s, 1H), 4.25–4.11 (m, 5H), 3.98–3.85 (m, 6H), 3.85–3.75 (m, 3H), 3.48 (m, 3H), 2.98 (d, J = 5.3 Hz, 1H), 2.64 (d, J = 5.3 Hz, 1H), 2.02–1.80 (m, 2H), 1.08 (s,

9H), 1.06 (s, 9H), 1.03 (s, 9H), 1.01 (s, 9H), 0.96 (s, 9H), 0.87 (s, 9H), 0.15 (s, 3H), 0.14 (s, 3H), 0.11 (s, 3H), 0.03 ppm (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ = 102.1, 101.3, 79.0, 74.4, 73.5, 72.6, 68.7, 67.4, 67.2, 66.9, 64.5, 58.3, 48.2, 47.6, 28.9, 27.7, 27.7, 27.2, 27.1, 26.3, 25.7, 22.8, 22.8, 20.2, 20.1, 18.7, 18.2, −4.0, −4.1, −4.3, −4.6 ppm. IR (neat): $\tilde{\nu}$ = 2931, 2098, 1741, 1251, 1159, 1097 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.05, DCM): +44. HRMS (ESI) *m/z*: $[M+H]^+$ calcd for $C_{44}H_{88}N_3O_{11}Si_4$ 946.54904, found 946.54933.

(2*R*,3*R*,4*S*,5*S*,6*R*)-2-(((3*R*,4*S*,6*R*,7*R*,8*S*)-4-(3-Azidopropoxy)-7-hydroxy-6-(hydroxymethyl)-1,5-dioxaspiro[2.5]octan-8-yl)oxy)-6-(hydroxymethyl)tetrahydro-2*H*-pyran-3,4,5-triol (17): Compound **16g** (145 mg, 0.153 mmol) was co-evaporated with toluene (3×) and dissolved in dry THF (14.5 mL). TBAF (1 M in THF, 2.3 mL, 2.3 mmol) was added and the mixture was stirred 5 days at room temperature. The solution was eluted with THF over a small Dowex-50WX4-200-Na^+ packed column, concentrated and purified by gradient column chromatography (EtOAc/MeOH, 19:1 to 9:1). The product was dissolved in water and lyophilized to afford the title compound as a white solid (64.8 mg, 97%). ^{1}H NMR (400 MHz, D_2O): δ = 5.23 (d, *J* = 3.8 Hz, 1H), 4.50 (s, 1H), 4.25 (d, *J* = 9.0 Hz, 1H), 3.95–3.68 (m, 8H), 3.63–3.53 (m, 2H), 3.53–3.43 (m, 3H), 3.41–3.34 (t, *J* = 8 Hz, 1H), 3.17 (d, *J* = 4.5 Hz, 1H), 2.87 (d, *J* = 4.6 Hz, 1H), 1.97–1.85 ppm (m, 2H). ^{13}C NMR (101 MHz, D_2O): δ = 100.0, 99.2, 73.0, 72.9, 72.5, 71.7, 71.5, 70.91, 69.1, 64.8, 60.3, 60.2, 58.7, 48.4, 48.1, 27.8 ppm. IR (neat): $\tilde{\nu}$ = 3369, 2927, 2108, 1521, 1026 cm^{-1}. $[\alpha]_D^{20}$ (*c* 0.1, DCM): +174. HRMS (ESI) *m/z*: $[M+NH_4]^+$ calcd for $C_{16}H_{31}N_4O_{11}$ 455.19838, found 455.19849.

(2*R*,3*S*,4*S*,5*S*,6*R*)-2-(((3*R*,4*S*,6*R*,7*R*,8*S*)-4-(3-azidopropoxy)-7-hydroxy-6-(hydroxymethyl)-1,5-dioxaspiro[2.5]octan-8-yl)oxy)-6-(hydroxymethyl)tetrahydro-2*H*-pyran-3,4,5-triol (18): This compound was prepared from **16m** (59 mg, 0.623 mmol) as described for the preparation of **17** to afford the title product (20 mg, 74%) as a white solid. ^{1}H NMR (400 MHz, D_2O): δ = 5.05 (d, *J* = 1.6 Hz, 1H), 4.45 (s, 1H), 4.18 (d, *J* = 9.3 Hz, 1H), 3.95 (dd, *J* = 3.2, 1.8 Hz, 1H), 3.77 (m, 6H), 3.64 (m, 3H), 3.57–3.48 (m, 1H), 3.41 (t, *J* = 6.5 Hz, 2H), 3.10 (d, *J* = 4.5 Hz, 1H), 2.81 (d, *J* = 4.5 Hz, 1H), 1.95–1.78 ppm (m, 2H). ^{13}C NMR (101 MHz, D_2O): δ = 101.0, 100.0, 73.3, 72.8, 72.5, 70.9, 70.4, 69.9, 66.3, 64.7, 60.8, 60.2, 58.7, 48.3, 48.0, 27.8 ppm. HRMS (ESI) *m/z*: $[M+Na]^+$ calcd for $C_{16}H_{27}N_3O_{11}$ 460.1538, found 460.1544.

1-(6-(((1-(3-(((3*R*,4*S*,6*R*,7*R*,8*S*)-7-Hydroxy-6-(hydroxymethyl)-8-(((2*R*,3*R*,4*S*,5*S*,6*R*)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2*H*-pyran-2-yl)oxy)-1,5-dioxaspiro[2.5]octan-4-yl)oxy)propyl)-1*H*-1,2,3-triazol-4-yl)methyl)amino)-6-oxohexyl)-3,3-dimethyl-2-((1*E*,3*E*)-5-((*Z*)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3*H*-indol-1-ium (4): Compound 17 (4.83 mg, 11.0 μmol) was dissolved in DMF (0.5 mL) and placed under Argon. Then the Cy5-alkyne[47] (6.1 mg, 11.0 μmol), aq. $CuSO_4$ (0.1 M, 44 μL, 4.4 μmol) and aq. sodium ascorbate (0.1 M, 44 μL, 4.4 μmol) were added and the mixture was stirred overnight at room temperature. The product was purified by HPLC (50 mM NH_4CO_3) to afford the title compound as a blue solid (3.54 mg, 32%). ^{1}H NMR (400 MHz, MeOD): δ = 8.24 (t, *J* = 13.0 Hz, 2H), 7.89 (s, 1H), 7.49 (d, *J* = 7.4 Hz, 2H), 7.44–7.38 (m, 2H), 7.32–7.23 (m, 4H), 6.62 (t, *J* = 12.4 Hz, 1H), 6.28 (d, *J* = 13.7 Hz, 2H), 5.14 (d, *J* = 3.8 Hz, 1H), 4.85 (s, 1H), 4.53 (t, *J* = 6.8 Hz, 2H), 4.42 (s, 2H), 4.32 (s, 1H), 4.16 (d, *J* = 9.1 Hz, 1H), 4.10 (t, *J* = 7.4 Hz, 2H), 3.84–3.64 (m, 9H), 3.63 (s, 3H), 3.56 (t, *J* = 9.3 Hz, 1H), 3.40 (dd, *J* = 9.7, 3.8 Hz, 1H), 3.37–3.32 (1, 9H), 3.07 (d, *J* = 5.3 Hz, 1H), 2.70 (d, *J* = 5.4 Hz, 1H), 2.25 (t, *J* = 7.3 Hz, 2H), 2.23–2.15 (m, 2H), 1.88–1.76 (m, 2H), 1.75–1.67 (m, 17H), 1.51–1.44 ppm (m, 2H). ^{13}C NMR (101 MHz, MeOD): δ = 180.3, 175.7, 175.4, 174.7, 155.5, 155.5, 146.1, 144.3, 143.6, 142.6, 142.5, 129.8, 129.7, 126.6, 126.3, 126.2, 124.7, 123.4, 123.3, 112.0, 111.9, 104.4, 104.3, 102.4, 101.8, 76.8, 75.1, 74.4, 74.0, 73.9, 73.0, 71.2, 65.3, 62.4, 62.4, 59.8, 50.6, 50.5, 48.4, 44.8, 36.5, 35.6, 31.5, 31.0, 28.1, 27.9, 27.8, 27.3, 26.4 ppm. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{51}H_{69}N_6O_{12}$ 957.4968, found 957.5005.

1-(6-(((1-(3-(((3*R*,4*S*,6*R*,7*R*,8*S*)-7-Hydroxy-6-(hydroxymethyl)-8-(((2*R*,3*S*,4*S*,5*S*,6*R*)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2*H*-pyran-2-yl)oxy)-1,5-dioxaspiro[2.5]octan-4-yl)oxy)propyl)-1*H*-1,2,3-triazol-4-yl)methyl)amino)-6-oxohexyl)-3,3-dimethyl-2-((1*E*,3*E*)-5-((*Z*)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3*H*-indol-1-ium (5): This compound was prepared from 18 (3.72 mg, 8.5 μmol) as described for the preparation of 4 to afford the product (2.9 mg, 34%) as a blue solid. ^{1}H NMR (600 MHz, MeOD): δ = 8.24 (t, *J* = 13.0 Hz, 2H), 7.90 (s, 1H), 7.49 (d, *J* = 7.4 Hz, 2H), 7.44–7.39 (m, 2H), 7.28 (dt, *J* = 16.4, 7.6 Hz, 4H), 6.62 (t, *J* = 12.4 Hz, 1H), 6.28 (d, *J* = 13.7 Hz, 2H), 5.19 (d, *J* = 1.3 Hz, 1H), 4.85 (s, 1H), 4.54 (t, *J* = 6.8 Hz, 2H), 4.42 (s, 2H), 4.29 (s, 1H), 4.23 (d, *J* = 9.2 Hz, 1H), 4.10 (t, *J* = 7.4 Hz, 3H), 3.90 (dd, *J* = 3.2, 1.7 Hz, 1H), 3.85–3.68 (m, 9H), 3.66 (d, *J* = 9.5 Hz, 1H), 3.63 (s, 3H), 3.60 (dd, *J* = 9.5, 3.3 Hz, 1H), 3.57–3.52 (m, 1H), 3.36–3.32 (m, 1H), 3.01 (d, *J* = 5.3 Hz, 1H), 2.68 (d, *J* = 5.4 Hz, 1H), 2.26 (t, *J* = 7.3 Hz, 2H), 2.20 (dq, *J* = 13.1, 6.7 Hz, 2H), 1.83 (m, 2H), 1.73 (s, 17H), 1.47 ppm (m, 2H). ^{13}C NMR (150 MHz, MeOD): δ = 180.3, 175.7, 175.4, 174.7, 155.5, 155.5, 146.1, 144.3, 143.6, 142.6, 142.5, 129.8, 129.7, 126.6, 126.3, 126.2, 124.8, 123.4, 123.3, 112.1, 111.8, 104.4, 104.3, 102.6, 102.4, 74.7, 74.3, 74.1, 73.3, 72.7, 72.1, 68.2, 65.2, 62.7, 62.3, 59.9, 50.6, 50.5, 48.3, 44.8, 36.5, 35.7, 31.5, 31.1, 31.0, 28.1, 28.0, 27.8, 27.3, 26.4 ppm. HRMS (ESI) *m/z*: $[M]^+$ calcd for $C_{51}H_{69}N_6O_{12}$ 957.4968, found 957.4995.

Labeling of *Bt*GH99 and *Bx*GH99 enzymes

To determine the detection limit, 400 ng recombinant *B. thetaiotaomicron* (*Bt*) and *B. xylanisolvens* (*Bx*) GH99 enzymes were labeled in 150 mM McIlvaine buffer, pH 7.0 (citric acid-Na_2HPO_4) with 0.0001–10 μM spiro-epoxyglycoside **4** or **5** for 1 h at 37 °C. The samples were then denatured with 5× Laemmli buffer (50% (v/v) 1 M Tris-HCl, pH 6.8, 50% (v/v) 100% (v/v) glycerol, 10% (w/v) DTT, 10% (w/v) SDS, 0.01% (w/v) bromophenol blue), boiled for 4 min at 100 °C, and separated by electrophoresis on 10% (w/v) SDS-PAGE gel running continuously at 90 V.[43] Wet slab-gels were scanned on fluorescence using a Typhoon FLA 9500 (GE Healthcare at λ_{EX} 532 nm and λ_{EM} 575 nm for ABP TB340; and at λ_{EX} 635 nm and λ_{EM} 665 nm for **4** and **5**. The pH optimum was analyzed using 4 ng enzyme incubated with 1 μM **4** or **5** dissolved in McIlvaine buffer, pH 3–8, for 30 min at 37 °C. Time-dependent labeling of *Bx*GH99 wild-type, E333Q and E336Q enzymes was assessed by incubating 400 ng for 5, 15 or 30 min with 1 μM **4** or **5** dissolved in McIlvaine buffer, pH 7.0. The effect of denaturation was assessed on 4 ng wild-type *Bt*GH99 and *Bx*GH99 by boiling for 4 min at 100 °C prior to incubating with 1 μM **4** and **5** for 30 min at 37 °C. Competitive ABPP assays utilized 4 ng *Bt*GH99 and *Bx*GH99 enzyme that was pre-incubated with 10–1000 μM **17**, **18** or ManIFG, or 0.3–30 μg μL^{-1} yeast mannan (*S. cerevisiae*), at pH 7.0 for 30 min at 37 °C, followed by labeling with 1 μM **4** and **5** for 30 min at 37 °C.

Functional GLA assay

Recombinant α-galactosidase GLA was diluted 1:2 in 50 mM McIlvaine buffer, pH 4.6, and pre-labeled with 2 μM TB340 for 1 h at 37 °C. Subsequently, the mixture was diluted to 1:500 in 150 mM McIlvaine buffer, pH 7.0. In parallel, 400 ng *Bx*GH99 wild-type, E333Q and E336Q were incubated in the presence or absence of 10 μM **4** or **5**, dissolved in 150 mM McIlvaine buffer, pH 7.0, for 1 h at 37 °C. Subsequently, the *Bx*GH99 mixture (10 μL) was incubated

with 10 μL TB340-labeled GLA for 8 h at 37 °C. Hereafter, samples were denatured, separated on SDS-PAGE gel and visualized by fluorescence scanning, as described above (vide supra). As control, 10 μL TB340-labeled GLA was treated by either Endo-H or PNGase-F, following the manufacturer's instructions (New England Biolabs).

Acknowledgements

We thank the Netherlands Organization for Scientific Research (NWO-CW ChemThem grant to JMFGA and HSO), and the European Research Council (ERC-2011-AdG-290836 "Chembiosphing" to HSO, and ERC-2012-AdG-32294 "Glycopoise" to GJD). GJD thanks the Royal Society for the Ken Murray Research Professorship.

Conflict of interest

The authors declare no conflict of interest.

Keywords: activity-based probes · endomannosidase · GH99 · glycosidase · inhibitors

[1] A. Helenius, M. Aebi, *Science* **2001**, *291*, 2364.
[2] M. Molinari, *Nat. Chem. Biol.* **2007**, *3*, 313.
[3] T. Feizi, M. Larkin, *Glycobiology* **1990**, *1*, 17.
[4] Y.-Y. Zhao, M. Takahashi, J.-G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume, N. Taniguchi, *Cancer Sci.* **2008**, *99*, 1304.
[5] K. Akasaka-Manya, H. Manya, Y. Sakurai, B. S. Wojczyk, Y. Kozutsumi, Y. Saito, N. Taniguchi, S. Murayama, S. L. Spitalnik, T. Endo, *Glycobiology* **2010**, *20*, 99.
[6] A. Herscovics, *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 96.
[7] Y. T. Pan, H. Hori, R. Saul, B. A. Sanford, R. J. Molyneux, A. D. Elbein, *Biochemistry* **1983**, *22*, 3975.
[8] V. W. Sasak, J. M. Ordovas, A. D. Elbein, R. W. Berninger, *Biochem. J.* **1985**, *232*, 759.
[9] L. Foddy, R. C. Hughes, *Eur. J. Biochem.* **1988**, *175*, 291.
[10] R. G. Spiro, *J. Biol. Chem.* **2000**, *275*, 35657.
[11] S. E. H. Moore, R. G. Spiro, *J. Biol. Chem.* **1990**, *265*, 13104.
[12] K. Fujimoto, R. Kornfeld, *J. Biol. Chem.* **1991**, *266*, 3571.
[13] C. Völker, C. M. De Praeter, B. Hardt, W. Breuer, B. Kalz-Füller, R. N. Van Coster, E. Bause, *Glycobiology* **2002**, *12*, 473.
[14] W. A. Lubas, R. G. Spiro, *J. Biol. Chem.* **1987**, *262*, 3775.
[15] W. A. Lubas, R. G. Spiro, *J. Biol. Chem.* **1988**, *263*, 3990.
[16] a) V. Lombard, H. G. Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490 – D495; b) The Cazypedia Consortium, *Glycobiology* **2017**, *28*, 3.
[17] F. Cuskin, E. C. Lowe, M. J. Temple, Y. Zhu, E. A. Cameron, N. A. Pudlo, N. T. Porter, K. Urs, A. J. Thompson, A. Cartmell, A. Rogowski, B. S. Hamilton, R. Chen, T. J. Tolbert, K. Piens, D. Bracke, W. Vervecken, Z. Hakki, G. Speciale, J. L. Munõz-Munõz, A. Day, M. J. Peña, R. McLean, M. D. Suits, A. B. Boraston, T. Atherly, C. J. Ziemer, S. J. Williams, G. J. Davies, D. W. Abbott, E. C. Martens, H. J. Gilbert, *Nature* **2015**, *517*, 165.
[18] Z. Hakki, A. J. Thompson, S. Bellmaine, G. Speciale, G. J. Davies, S. J. Williams, *Chem. Eur. J.* **2015**, *21*, 1966.
[19] 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.
[20] S. Iwamoto, Y. Kasahara, K. I. Kamei, A. Seko, Y. Takeda, Y. Ito, I. Matsuo, *Biosci. Biotechnol. Biochem.* **2014**, *78*, 927.
[21] C. Vogel, G. Pohlentz, *J. Carbohydr. Chem.* **2000**, *19*, 1247.
[22] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416.
[23] A. J. Thompson, R. J. Williams, Z. Hakki, D. S. Alonzi, T. Wennekes, T. M. Gloster, K. Songsrirote, J. E. Thomas-Oates, T. M. Wrodnigg, J. Spreitz, A. E. Stutz, T. D. Butters, S. J. Williams, G. J. Davies, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 781.
[24] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383.
[25] Y. Zhang, C. Chen, L. Jin, H. Tan, F. Wang, H. Cao, *Carbohydr. Res.* **2015**, *401*, 109.
[26] G. Despras, R. Robert, B. Sendid, E. MacHez, D. Poulain, J. M. Mallet, *Bioorg. Med. Chem.* **2012**, *20*, 1817.
[27] M. S. Motawia, C. E. Olsen, K. Enevoldsen, J. Marcussen, B. L. Møller, *Carbohydr. Res.* **1995**, *277*, 109.
[28] M. Martín-Lomas, N. Khiar, S. García, J. L. Koessler, P. M. Nieto, T. W. Rademacher, *Chem. Eur. J.* **2000**, *6*, 3608.
[29] M. Heuckendorff, J. Bendix, C. M. Pedersen, M. Bols, *Org. Lett.* **2014**, *16*, 1116.
[30] D. Crich, V. Dudkin, *Tetrahedron Lett.* **2000**, *41*, 5643.
[31] S. van der Vorm, T. Hansen, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Sci.* **2017**, *8*, 1867.
[32] S. Van der Vorm, H. S. Overkleeft, G. A. Van der Marel, J. D. C. Codée, *J. Org. Chem.* **2017**, *82*, 4793.
[33] K. Sato, J. Yoshimura, *Carbohydr. Res.* **1979**, *73*, 75.
[34] E. J. Corey, M. Chaykovsky, *J. Am. Chem. Soc.* **1965**, *87*, 1353.
[35] S. C. Garman, D. N. Garboczi, *J. Mol. Biol.* **2004**, *337*, 319.
[36] Y. Sohn, J. M. Lee, H. R. Park, S. C. Jung, T. H. Park, D. B. Oh, *BMB Rep.* **2013**, *46*, 157.
[37] J. Jiang, T. J. M. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codée, J. M. F. G. Aerts, H. S. Overkleeft, *Chem. Eur. J.* **2015**, *21*, 10861.
[38] T. Nakajima, C. E. Ballou, *J. Biol. Chem.* **1974**, *249*, 7685.
[39] G. Legler, E. Bause, *Carbohydr. Res.* **1973**, *28*, 45.
[40] R. Havukainen, A. Törrönen, T. Laitinen, J. Rouvinen, *Biochemistry* **1996**, *35*, 9617.
[41] J. C. Gebler, R. Aebersold, S. G. Withers, *J. Biol. Chem.* **1992**, *267*, 11126.
[42] L. I. Willems, J. Jiang, K. Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. N. Beenakker, S. P. Schröder, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée, H. S. Overkleeft, *Chem. Eur. J.* **2014**, *20*, 10864.
[43] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft, J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907.
[44] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. A. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. A. A. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. C. Walvoort, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts, H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529; *Angew. Chem.* **2012**, *124*, 12697.
[45] V. Ladmiral, G. Mantovani, G. J. Clarkson, S. Cauet, J. L. Irwin, D. M. Haddleton, *J. Am. Chem. Soc.* **2006**, *128*, 4823.
[46] C. M. Pedersen, L. U. Nordstrøm, M. Bols, *J. Am. Chem. Soc.* **2007**, *129*, 9222.
[47] O. Kaczmarek, H. A. Scheidt, A. Bunge, D. Föse, S. Karsten, A. Arbuzova, D. Huster, J. Liebscher, *Eur. J. Org. Chem.* **2010**, 1579.

Manuscript received: April 17, 2018
Revised manuscript received: May 23, 2018
Accepted manuscript online: May 24, 2018
Version of record online: June 21, 2018

ChemPubSoc Europe

DOI: 10.1002/chem.201800435

CHEMISTRY A European Journal Full Paper

Enzymes

Exploration of Strategies for Mechanism-Based Inhibitor Design for Family GH99 *endo*-α-1,2-Mannanases

Pearl Z. Fernandes,[a] Marija Petricevic,[a] Lukasz Sobala,[b] Gideon J. Davies,*[b] and Spencer J. Williams*[a]

Abstract: *endo*-α-1,2-Mannosidases and -mannanases, members of glycoside hydrolase family 99 (GH99), cleave α-Glc/Man-1,3-α-Man-OR structures within mammalian *N*-linked glycans and fungal α-mannan, respectively. They are proposed to act through a two-step mechanism involving a 1,2-anhydrosugar "epoxide" intermediate incorporating two conserved catalytic carboxylates. In the first step, one carboxylate acts as a general base to deprotonate the 2-hydroxy group adjacent to the fissile glycosidic bond, and the other provides general acid assistance to the departure of the aglycon. We report herein the synthesis of two inhibitors designed to interact with either the general base (α-mannosyl-1,3-(2-aminodeoxymannojirimycin), Man2NH$_2$DMJ) or the general acid (α-mannosyl-1,3-mannoimidazole, ManManIm). Modest affinities were observed for an *endo*-α-1,2-mannanase from *Bacteroides thetaiotaomicron*. Structural studies revealed that Man2NH$_2$DMJ binds like other iminosugar inhibitors, which suggests that the poor inhibition shown by this compound is not a result of a failure to achieve the expected interaction with the general base, but rather the reduction in basicity of the endocyclic nitrogen caused by introduction of a vicinal, protonated amine at C2. ManManIm binds with the imidazole headgroup distorted downwards, a result of an unfavourable interaction with a conserved active site tyrosine. This study has identified important limitations associated with mechanism-inspired inhibitor design for GH99 enzymes.

Introduction

Glycoside hydrolases of the carbohydrate-active enzyme (see www.cazy.org; www.cazypedia.org)[1,2] family GH99 are *endo*-acting mannosidases that cleave α-mannoside linkages within mammalian high mannose *N*-glycans (*endo*-α-1,2-mannosidases)[3–7] and fungal α-mannans (*endo*-α-1,2-mannanases, Figure 1 A).[8,9] Inhibitor design for these enzymes is driven by their potential use to understand glycoprotein biosynthesis and maturation in the secretory pathway, and to manipulate fungal mannan degradation processes in the human gut microbiota. Structural and mechanistic studies of family GH99 enzymes suggest that they utilise an unusual mechanism involving neighbouring group participation by the substrate 2-hydroxy to form a 1,2-anhydrosugar intermediate.[10] In this proposed mechanism, a conserved active site residue acts as a general base to deprotonate the 2-OH group, thereby facilitating its nucleophilic attack on C1 (Figure 1 A). This process has little biological precedent (for a related proposal see Ref. [11]), but occurs in the base-promoted solvolysis of α-mannosides.[12]

Efforts to develop inhibitors of GH99 enzymes have relied upon appending 1,3-linked α-glucosyl (to target mammalian *endo*-α-1,2-mannosidases) or 1,3-linked α-mannosyl (to target bacterial *endo*-α-1,2-mannanases) groups to various sugar-shaped heterocycles. Spiro and co-workers reported the discovery of α-glucosyl-1,3-deoxymannojirimycin (GlcDMJ) as an effective inhibitor of the mammalian enzyme,[13,14] and follow-on studies by Fleet and co-workers revealed α-mannosyl-1,3-deoxymannojirimycin (ManDMJ) to be a slightly weaker inhibitor for this enzyme (Figure 1 B).[15] The potency of GlcDMJ was subsequently exceeded by α-glucosyl-1,3-isofagomine (GlcIFG).[10,16] Equivalent results have been noted for bacterial GH99 enzymes, which led to the development of α-mannosyl-1,3-isofagomine (ManIFG; dissociation constant, $K_D = 0.14$ μM for *Bacteroides thetaiotaomicron* GH99).[8] Furthermore, reintroduction of the "missing" 2-OH of 1,3-isofagomine (IFG) into ManIFG gave α-mannosyl-1,3-noeuromycin (ManNOE), which was shown to be five-fold more potent towards the *B. thetaiotaomicron* GH99 enzyme ($K_D = 0.03$ μM).[17] These compounds bind in a ground-state $^4C_1$ conformation, as seen in complexes of inactive enzyme with substrate and thus proposed for the

[a] P. Z. Fernandes, M. Petricevic, Prof. S. J. Williams
School of Chemistry
Bio21 Molecular Science and Biotechnology Institute
University of Melbourne, Parkville, Vic 3010 (Australia)
E-mail: sjwill@unimelb.edu.au

[b] L. Sobala, Prof. G. J. Davies
York Structural Biology Laboratory, Department of Chemistry
University of York, Heslington, YO10 5DD (UK)
E-mail: gideon.davies@york.ac.uk

Supporting information and the ORCID number(s) for the author(s) of this article can be found under https://doi.org/10.1002/chem.201800435.

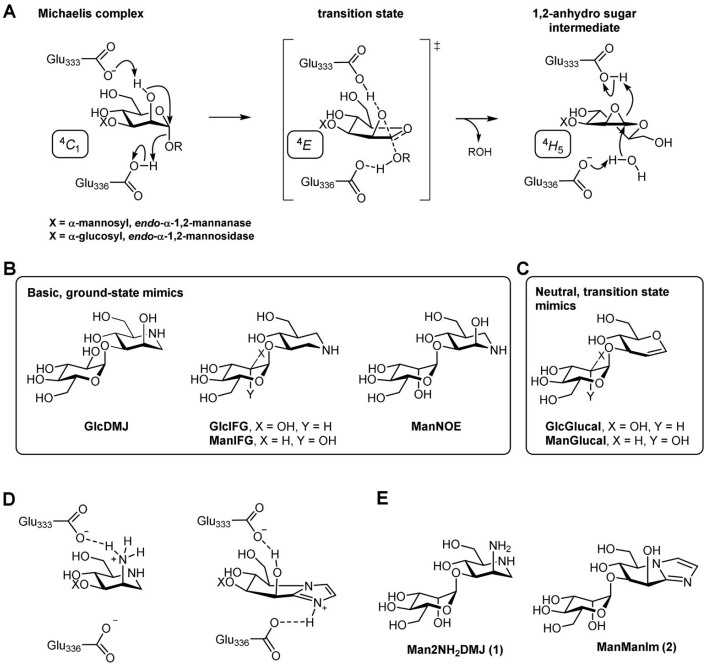


Figure 1. (A) Proposed mechanism for family GH99 enzymes retaining endomannosidases/endomannanases. Only the first half of the catalytic cycle is shown. (B) Saturated basic heterocyclic inhibitors for GH99 enzymes mimicking the ground state conformation. (C) Neutral glycal inhibitors for GH99 enzymes mimicking the transition state. (D) Two inhibitor design concepts explored herein. (E) Structures of Man2NH$_2$DMJ (**1**) and ManManIm (**2**).

conformation of substrate within the Michaelis complex (Figure 1 A), which suggests that potent inhibition of GH99 enzymes can be achieved simply by mimicry of the charge in the transition state.[17]

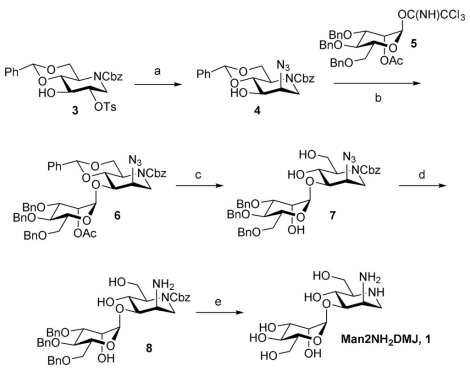
Separately, Spiro and co-workers showed that the neutral compound GlcGlucal (Figure 1 C) was a modest inhibitor of mammalian GH99 (rat Golgi preparation, $IC_{50}=2.3$ μM; for GlcDMJ $IC_{50}=1.7$ μM);[14,18] the equivalent molecule targeting bacterial GH99, ManGlucal, was also a ligand with mildly potent affinity ($K_D=15$ μM for *Bt*GH99).[17] Computational free-energy landscape analysis of the preferred conformation of D-glucal suggested that the inhibition of the glucal-based inhibitors arises from mimicry of the proposed ^4E conformation of the transition state or the proposed $^4H_5$ conformation of the 1,2-anhydro sugar intermediate, but with no contribution from charge mimicry owing to the neutral nature of this compound.[17]

We report here our efforts to explore two new inhibitor design strategies for the inhibition of GH99 enzymes. Considering the role of the basic residue implicated in the 1,2-anhydrosugar mechanism of GH99 enzymes, we speculated that introduction of an amino group into the structure of ManDMJ to give Man-2NH$_2$DMJ (**1**; Figure 1 E) could promote the formation of a favourable ionic interaction upon inhibitor binding (Figure 1 D). Separately, the glycoimidazole class of inhibitors were developed following the discovery of the natural product nagstatin,[19] and are believed to derive their potency from their ability to mimic the shape of the oxocarbenium-like transition state as well as from the ability of the imidazole glycosidic nitrogen to engage in a hydrogen bond with an appropriately situated carboxylate residue in the active site (Figure 1 D).[20] For the present work, this would require the synthesis of ManManIm (**2**; Figure 1 E). Thus, we report herein on the synthesis of these two target inhibitors, the structural characterisation of their binding modes and measurement of their binding constants.

Results and Discussion

Synthesis of Man2NH$_2$DMJ and ManManIm

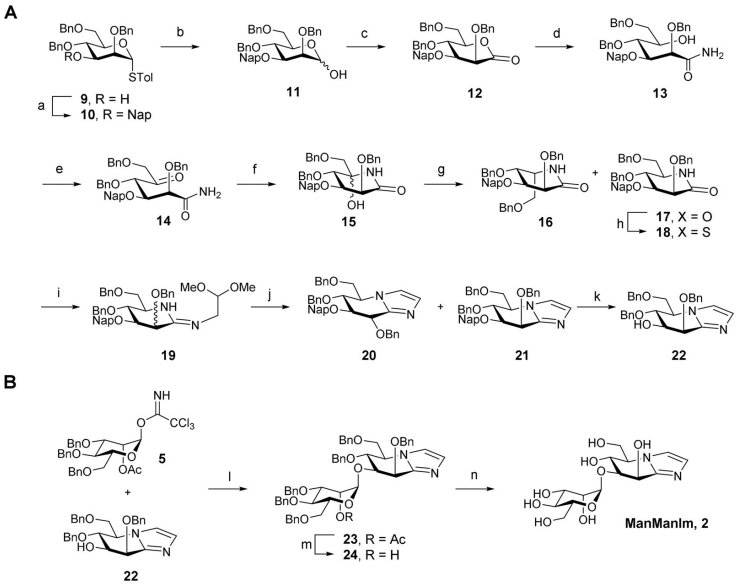
Man2NH$_2$DMJ (**1**) was prepared by substitution of known tosylate **3**[21] with sodium azide in DMF to afford azide **4** (Scheme 1). Coupling of azide **4** with trichloroacetimidate **5**[22] under the agency of TfOH afforded the disaccharide **6** in a yield of 83%. The deprotection of **6** was achieved in a stepwise



Scheme 1. Reagents and conditions: a) NaN_3, DMF, reflux, 74%; b) TfOH, CH_2Cl_2, −30 to 0 °C, 87%; c) i. NaOMe, MeOH, ii. 9:1 TFA/H_2O, 83%; d) DTT, pyr, pH 9.2 $NaHCO_3/Na_2CO_3$, 80%; e) H_2, $Pd(OH)_2$/C, aq. HCl, 2:2:1 EtOAc/MeOH/H_2O, 70%.

manner, as attempts to perform a global deprotection that involved simultaneous removal of benzyloxycarbonyl (Cbz), benzylidene and benzyl ethers as well as the reduction of the azide was unsuccessful. Deacetylation of **6** (NaOMe/MeOH) and then hydrolysis of the benzaldehyde acetal (TFA/H_2O) afforded triol **7**. The azide group was reduced with dithiothreitol (DTT)/pyridine buffer to afford amine **8**. Removal of the Cbz and benzyl groups then proceeded smoothly by using H_2 and Pearlman's catalyst to afford **1**.

ManManIm (**2**) was synthesised through a sequence involving the preparation of the protected mannoimidazole alcohol **22**, followed by elaboration to the disaccharide (Scheme 2). The known alcohol **9**[23] was treated with 2-naphthylmethyl bromide (NapBr)/NaH in DMF to afford **10**. Hydrolysis of the thioglycoside with *N*-iodosuccinimide (NIS) in H_2O/acetone gave the hemiacetal **11**, which was oxidised to the lactone **12** under Albright–Goldman conditions.[24] For the conversion of the lactone **12** to the lactam **17** we followed the protocol developed by Overkleeft et al.,[25] which involved aminolysis to the acyclic amide **13**, Albright–Goldman oxidation (→**14**) and ring closure promoted by ammonia/MeOH (→**15**). Reduction of the hemiaminals **15** with $NaCNBH_3$ afforded a 2:1 mixture of the D-*manno* and L-*gulo* lactams, from which the D-*manno* lactam **17** was isolated in a yield of 38%. Conversion of the lactam to the thionolactam **18** was achieved by using Lawesson's reagent and pyridine in toluene. Annulation of the imidazole ring was achieved by following the general approach of Vasella and co-workers.[26] Reaction of the thionolactam **18** with aminoacetaldehyde dimethyl acetal afforded the amidine **19**, and imidazole ring formation was achieved by catalysis with TsOH to provide a mixture of D-*gluco* and D-*manno* imida-



Scheme 2. A) Preparation of imidazole alcohol **22**. Reagents and conditions: a) NapBr, NaH, DMF, 86%; b) NIS, H_2O, acetone, 0 °C, 99%; c) DMSO, Ac_2O; d) NH_3, THF, reflux; e) DMSO, Ac_2O; f) NH_3, MeOH, 88% over steps c–f; g) HCO_2H, $NaBH_3(CN)$, 38% D-*manno*, 33% L-*gulo*; h) Lawesson's reagent, pyridine, 4 Å molecular sieves, toluene, 93%; i) $H_2NCH_2CH(OMe)_2$; j) TsOH·H_2O, toluene, 60 °C, yields over steps i and j: 42% D-*gluco*, 32% D-*manno*; k) DDQ, CH_2Cl_2/H_2O, 67%. B) Synthesis of ManManIm (**2**). Reagents and conditions: l) TfOH, 4 Å molecular sieves, toluene, −20 °C, 47%; m) K_2CO_3/MeOH, 46%; n) H_2 (34 bar), $Pd(OH)_2$/C, AcOH, EtOAc, MeOH, H_2O, 48%.

zoles in a 2:1 ratio, from which the D-*manno* imidazole **21** was isolated in a yield of 32% over two steps. The naphthylmethyl group was removed under the agency of 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) and CH_2Cl_2/H_2O to afford the alcohol **22**.

Coupling of **22** with trichloroacetimidate **5**[22] catalysed by TfOH afforded the disaccharide **23** in a yield of 47%. Deprotection was achieved in two steps under conditions chosen to avoid epimerisation at C2. Treatment of **23** with K_2CO_3/MeOH afforded the alcohol **24**, and hydrogenation with Pearlman's catalyst afforded **2**.

Binding affinities and 3D structures

Isothermal titration calorimetry (ITC) was used to assess the binding of **1** and **2** to a bacterial endomannosidase. Titration of *Bt*GH99 with $Man2NH_2DMJ$ (**1**) revealed binding with $K_D = 97.7 \pm 4.9$ μM (Figure 2), whereas no binding with ManManIm (**2**) was evident by ITC. Placed in context, **1** has a poorer binding affinity towards *Bt*GH99 than GlcDMJ ($K_D = 24$ μM);[10] the equivalent data is not available for ManDMJ, but as this enzyme prefers to bind Man-configured substrates, the difference would be expected to be even greater.

Three-dimensional structures were obtained for **1** and **2** bound to the *endo*-α-1,2-mannanase *Bx*GH99 from *Bacteroides xylanisolvens*, which is closely related to *Bt*GH99 but more amenable to complex formation. These complexes diffracted to a resolution of 1.1 and 1.3 Å, respectively (Table 1). Occupancy of the active site for the complex with **1** was essentially complete, whereas that with **2**, with prolonged soaking, was estimated to be 80%, likely a consequence of the poor affinity of the compound for the enzyme. As predicted, both compounds bound in the −2/−1 subsites of the enzyme (sub-site nomenclature from Ref. [27]) and will be discussed in turn.

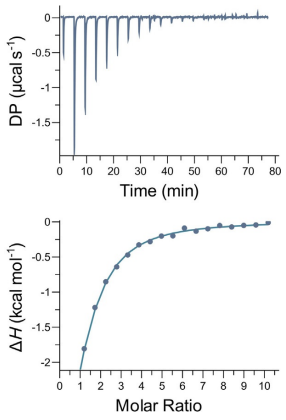


Figure 2. Isothermal titration calorimetry thermogram showing the binding of $Man2NH_2DMJ$ (**1**) to *Bacteroides thetaiotaomicron endo*-α-1,2-mannanase (*Bt*GH99). DP = differential power. Binding parameters $K_D = 97.7 \pm 4.9$ μM, $N = 1$ (fixed) and $\Delta H = -5.9 \pm 0.1$ kcal mol^{-1}.

Table 1. Data collection and refinement statistics for the complexes of *Bx*GH99 with **1** and **2**.

	*Bx*GH99 complexed with aminoDMJ (**1**)	*Bx*GH99 complexed with ManManIm (**2**)
	Data collection	
Space group	*I*4	*I*4
Cell dimensions		
a [Å]	108.1	108.6
b [Å]	108.1	108.6
c [Å]	67.5	67.8
α [°]	90	90
β [°]	90	90
γ [°]	90	90
resolution [Å]	76.44–1.13 (1.15–1.13)[a]	76.81–1.30 (1.32–1.30)[a]
R_{merge}	0.069 (1.501)	0.054 (1.224)
R_{pim}	0.026 (0.735)	0.020 (0.610)
CC(1/2)	0.999 (0.400)	(0.999) 0.486
I/σ*I*	10.2 (1.0)	14.0 (0.9)
completeness [%]	99.1 (86.0)	99.5 (92.7)
redundancy	7.5 (4.8)	7.5 (4.6)
	Refinement	
resolution [Å]	76.44–1.13	76.81–1.30
no. reflections all/free	143544/7133	96144/4810
R_{work}/R_{free}	0.122/0.144	0.134/0.162
no. atoms		
protein	3188	3146
ligand/ion	22	25
water	467	427
B factors [Å^2]		
protein	17.2	20.5
ligand/ion	20.3	22.4
water	35.1	36.7
r.m.s. deviations		
bond lengths [Å]	0.0101	0.011
bond angles [°]	1.495	1.497
PDB ID	6FAM	6FAR

[a] Values in parentheses are for the highest-resolution shell.

Structural analysis of the *Bx*GH99–**1** complex (Figure 3A) revealed the piperidine ring in a $^4C_1$ conformation, which matches that seen for complexes of the wild-type enzyme with GlcDMJ and isofagomine-based inhibitors[8,10,17] as well as that of a disabled mutant with substrate.[8] The 2-amino group is situated appropriately to interact with the E333 residue, that which is proposed to act as a general base/acid through deprotonation of the 2-hydroxy group. Overlay of this complex with that of *Bx*GH99–GlcDMJ reported previously[10] revealed that the positioning and conformations of the rings in the −1 and −2 sub-sites are essentially identical, and that no amino acid residues undergo significant movement (Figure 3C). In particular, the E333···O2 and E333···N2 distances are 2.54 and 2.59 Å, respectively. The poor binding affinity of **1** compared with GlcDMJ therefore does not result from incorrect binding of the inhibitor, and must instead reflect a failure to fully capitalise on the proposed interactions. It is widely acknowledged that iminosugars such as DMJ (and thus GlcDMJ) achieve inhibition through binding to glycosidases in their protonated

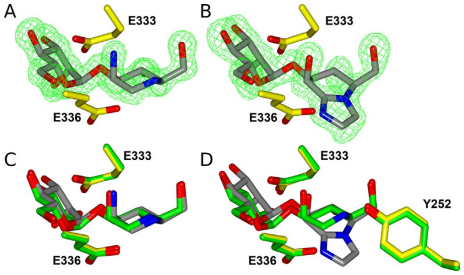


Figure 3. Three-dimensional structures of *Bt*GH99 complexed with A) Man2NH$_2$DMJ (**1**) and B) ManManIm (**2**). Electron density maps are maximum likelihood/σ_A weight F_o–F_c difference syntheses contoured at 0.5 and 0.3 e Å^{-3} for panels A and B, respectively, visible before refining the structure model with the ligand added. (C) Overlay of Man2NH$_2$DMJ (**1**) with GlcDMJ (PDB code 4FAM). (D) Overlay of ManManIm (**2**) with GlcDMJ (PDB code 4FAR).

form;[28] this is supported by first-principles consideration of the basicity of these inhibitors and the relevant pK_a values of the catalytic residues, as well as by studies of the pH dependence of inhibition. In the case of **1**, there are two basic nitrogen residues. However, for vicinal diamines, protonation at one nitrogen has a profound effect on the pK_a value at the second nitrogen; in acyclic systems this effect has been estimated to be $\Delta pK_a = 3.6$ units for NH_3^+ and NR_3^+.[29] Moreover, in cyclic systems there are stereoelectronic and conformational contributions, notable examples for various diamines (pK_{a1}, pK_{a2}) include piperazine (9.8, 5.7),[29] *cis*-1,3-diaminocyclohexane (10.3, 8.3)[30] and *trans*-1,3-diaminocyclohexane (10.4, 8.5).[30] Finally, vicinal hydroxy groups can also perturb amine pK_a values; in Man2NH$_2$DMJ, O4 is antiperiplanar with respect to the endocyclic nitrogen and would be expected to reduce its basicity by around 1.3 pK_a units.[30] Collectively, this analysis would suggest that N2 is protonated by the general acid E333, and that it is unlikely that the dication is formed, and therefore Man2NH$_2$DMJ fails to appropriately mimic an oxocarbenium-like transition state. A related example of this phenomenon was reported in which introduction of a second amine vicinal to a pre-existing one in apramycin resulted in a dramatic loss of binding to a bacterial ribosome of approximately 100-fold.[31] Additionally, the proposed binding mode of **1** shown in Figure 1D highlights the fact that the 2-amino group has additional hydrogen substituents that may cause an energy penalty upon binding of the inhibitor.

Structural analysis of the *Bx*GH99–**2** complex revealed the piperidine ring of the mannoimidazole moiety to be in an unusual $^2H_3/E_3$ conformation (Figure 3B).[32] Overlay of the complex with that of *Bx*GH99–GlcDMJ[10] revealed that although the −2 sugar residues occupy similar positions, the mannoimidazole headgroup is atypically positioned such that the heterocycle projects downward into the active site, below the plane of the piperidine ring of the GlcDMJ complex (Figure 3D). In this case the E336···N (imidazole ring) distance is 2.65 Å, similar to that seen in related glycoimidazole complexes.[33] In the original formulation by Heightman and Vasella, β-equatorial glycosidases were proposed to perform protonation from the side, in what was termed "lateral protonation", with the acid either on the same side as the endocyclic oxygen (*syn*) or opposed to it (*anti*).[20] In a subsequent publication Nerinckx et al. formalised this concept by dividing the space around the −1 sugar into *anti* and *syn* hemispheres through a plane defined by the glycosidic oxygen, C1 and H1 of the sugar residue.[34] Analysis of complexes of various *anti*-protonating glycosidases revealed that the acid/base or acid residues responsible for protonating the leaving group are in fact not universally located lateral to the mean plane of the sugar, but are more commonly positioned above or below it, so as to better protonate the leaving group oxygen. However, this does not prevent glycoimidazoles binding in normal orientations and engaging in hydrogen-bonding interactions with the imidazole nitrogen. For example, in the case of the retaining GH116 β-glucosidase from *Thermoanaerobacterium xylanolyticum*, the acid/base is positioned above the mean plane of the sugar, but a normal orientation and conformation of glucoimidazole was observed.[35] Mannoimidazole also binds in the normal fashion to an inverting GH47 α-mannosidase from *Caulibacter* sp. in which the acid is below the mean plane of the inhibitor, but instead the inhibitor establishes an interaction with another conserved active site carboxylic acid that lies lateral to the imidazole.[36] *Bx*GH99 is an *anti*-protonating enzyme with its general acid/base Glu336 positioned below the plane of the ring to facilitate classical *anti* protonation of the axial glycosidic oxygen (O5-C1-O1 angle is approximately 60°). The distorted mode of binding of the mannoimidazole moiety of **2** seems to be a consequence of the imidazole binding to maximise this interaction with the acid/base. Close examination of the active site of *Bx*GH99 revealed that if the ManIm moiety were to be shifted up to the same position as that of the piperidine of GlcDMJ, a steric interaction would result with Tyr252, a conserved residue. In fact, the distance between the imidazole C=C bond and Tyr252 Cε is only 3.2 Å, which causes the wwPDB validation software[37] to report H/H steric clashes in this region. In fact, a ternary complex of GlcDMJ and α-1,2-mannobiose highlighted the fact that the active site of the enzyme involves a sharp bend in the −1 and +1 sub-sites. The failure of **2** to bind in a typical position in the −1 sub-site is thus likely a result of a failure to accommodate the imidazole ring owing to the location of Tyr252.

Conclusions

We have reported here the design and synthesis of two "mechanism-based" inhibitors of family GH99 endomannanases. Although Man2NH$_2$DMJ (**1**) bound to the bacterial endomannanase *Bx*GH99 in the expected manner, its affinity for *Bt*GH99 did not exceed that seen for GlcDMJ. This appears to be a result of the perturbing effect of the 2-amino substituent, which reduces the basicity of the endocyclic nitrogen and its ability to be protonated in the active site and thereby resemble the oxocarbenium-like transition state. On the other hand, the binding of ManManIm (**2**) to *Bt*GH99 could not be detect-

ed by ITC and, consistent with this, the X-ray structure of **2** complexed with *Bx*GH99 displayed incomplete occupancy. The poor binding of this inhibitor appears to be a consequence of an inability of the active site of *Bx*GH99 to accommodate the annulated imidazole ring because of an interaction with a conserved Tyr active-site residue. This study provides important insights that will inform future strategies for the development of mechanism-inspired and transition-state mimicking inhibitors of GH99 enzymes.

Experimental Section

General: ^{1}H and ^{13}C NMR spectra were recorded by using 400, 500 or 600 MHz Varian INOVA spectrometers. All signals were referenced to TMS ($\delta=0.00$ ppm) or solvent peaks ($CDCl_3$: $\delta=7.26$ ppm for ^{1}H and 77.16 ppm for ^{13}C; D_2O: $\delta=4.80$ ppm for ^{1}H and TMS: $\delta=0.00$ ppm for ^{13}C; [D_4]MeOH: $\delta=3.49$ ppm for ^{1}H and $\delta=49.0$ ppm for ^{13}C). Melting points were obtained by using a Reichert-Jung hot-stage apparatus. TLC analysis was performed with aluminium-backed Merck Silica Gel 60 F254 sheets, detection was achieved by using UV light, 5% H_2SO_4 in MeOH or ceric ammonium molybdate ("Hanessian's stain") with charring as necessary. Flash chromatography was performed by using Geduran silica gel according to the method of Still et al.[38] Dry CH_2Cl_2, THF and Et_2O were obtained from a dry solvent apparatus (Glass Contour of SG Water, Nashua).[39] DMF and DMSO were dried over 4 Å molecular sieves.

2-Azido-4,6-*O*-benzylidene-*N*-benzyloxycarbonyl-1,2,5-trideoxy-1,5-imino-D-mannitol (4): Sodium azide (57.8 mg, 0.890 mmol) was added to a solution of 4,6-*O*-[(*R*)-benzylidene]-*N*-benzyloxycarbonyl-1,5-dideoxy-2-*O*-(*p*-toluenesulfony1)-D-glucitol[21] (**3**; 120 mg, 0.222 mmol) in DMF (1 mL). The suspension was heated at reflux for 18 h, poured into ice, extracted into EtOAc (3×20 mL), washed with brine (2×20 mL), dried over anhydrous $MgSO_4$ and evaporated to dryness. Column chromatography (AcOEt/pet. ether 40-60, 1:5) gave the azide **4** (67.7 mg, 74%) as a white solid. $[\alpha]_D^{24}=-21.9$ ($c=1.12$ in $CHCl_3$); ^{1}H NMR ($CDCl_3$, 500 MHz): $\delta=2.74$ (s, 1H; NH), 2.82 (dd, $J=1.6$, 14.5 Hz, 1H; 1-H_a), 3.06 (td, $J=4.6$, 10.2 Hz, 1H; 5-H), 3.74 (dd, $J=3.8$, 9.2 Hz, 1H; 3-H), 3.79–3.93 (m, 2H; 2,4-H), 4.31 (dd, $J=3.0$, 14.5 Hz, 1H; 1-H_b) 4.46 (t, $J=11$ Hz, 1H; 6-H_a), 4.66 (dd, $J=4.6$, 11.6 Hz, 1H; 6-H_b), 5.01 (d, $J=3.1$ Hz, 2H; CH_2), 5.48 ppm (s, 1H; CH); ^{13}C NMR ($CDCl_3$, 125 MHz): $\delta=48.1$, 55.8, 60.1, 67.8, 69.2, 73.6, 78.2 (7C; C1–C6, CH_2), 101.8 (1C; CH), 126.3, 128.3, 128.4, 128.5, 128.7, 129.4, 136.0, 137.3 (12C; Ph), 155.0 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{21}H_{22}N_4O_5$: 411.1663 [*M*+H]$^+$; found: 411.1664.

2-*O*-Acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2-azido-4,6-*O*-benzylidene-*N*-benzyloxycarbonyl-1,2,5-trideoxy-1,5-imino-D-mannitol (6): TfOH (0.043 μL, 0.0049 mmol) was added to a mixture of acceptor **4** (20 mg, 0.049 mmol) and 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl trichloroacetimidate (**5**;[22] 37 mg, 0.058) in CH_2Cl_2 over 4 Å sieves at −30 °C, The mixture was stirred for 30 min, warmed to 0 °C and quenched with Et_3N (7 μL, 0.05 mmol) and then concentrated under reduced pressure. Flash chromatography (EtOAc/pet. ether, 25:75) gave the disaccharide **6** (37.4 mg, 87%) as a colourless oil. $[\alpha]_D^{24}=-4.2$ ($c=0.89$ in $CHCl_3$); ^{1}H NMR ($CDCl_3$, 500 MHz): $\delta=2.80$ (dd, $J_{1,1}=14.4$, $J_{1,2}=0.9$ Hz, 1H; 1-H_a), 3.15 (dt, $J=10.1$, 4.6 Hz, 1H; 5-H), 3.70–4.00 (m, 6H; 3,4,4′,5′-H, 6″-H_a, 6′-H_b), 4.03 (dd, $J=9.3$, 3.4 Hz, 1H; 3′-H), 4.17–4.20 (m, 1H; 2-H), 4.28 (dd, $J=14.5$, 2.2 Hz, 1H; 1-H_b), 4.47–4.52 (m, 3H; 3×CH_2Ph), 4.60–4.64 (m, 2H; 6-H_a, CH_2Ph), 4.69 (d, $J=11$ Hz, 1H; CH_2Ph), 4.76 (dd, $J=11.6$, 4.5 Hz, 1H; 6-H_b), 4.86 (d, $J=11$ Hz, 1H; CH_2Ph), 5.12 (d, $J=3.6$ Hz, 2H; CH_2), 5.28 (d, $J=1.6$ Hz, 1H; 1′-H), 5.59 (dd, $J=3.3$, 1.8 Hz, 1H; 2′-H), 5.64 (s, 1H; CH), 7.17–7.46 ppm (m, 25H; Ph); ^{13}C NMR ($CDCl_3$, 125 MHz): $\delta=48.3$ (1C; C-1), 56.3 (1C; C-5), 60.0, 72.7, 74.4, 77.8 (4C; C-3,4,4′,5), 67.7 (1C; CH_2), 68.5 (1C; C-2′), 69.1 (1C; C-6), 69.3 (1C; C-6′), 72.2, 73.6, 75.1 (3C; CH_2Ph), 78.1 (1C; C-2), 78.2 (1C; C-3′), 99.5 (1C; C-1′), 100.90 (1C; CH), 100.92, 126.0, 127.77, 127.79, 127.83, 127.9, 128.0, 128.2, 128.28, 128.29, 128.41, 128.44, 128.5, 128.7, 128.9 ppm (30C; Ph); HRMS (ESI, +ve): *m/z* calcd for $C_{50}H_{52}N_4O_{11}$: 907.3525 [*M*+Na]$^+$; found: 907.3544.

3,4,6-Tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2-azido-*N*-benzyloxycarbonyl-1,2,5-trideoxy-1,5-imino-D-mannitol (7): A solution of sodium methoxide in methanol (0.1 M, 10 μL, 1 μmol) was added to **6** (60 mg, 0.068 mmol) in methanol (0.5 mL) and the mixture was stirred for 1 h and then concentrated under reduced pressure to give an alcohol, which was used without purification. TFA/H_2O (9:1, 100 μL) was added to the crude alcohol and the mixture was stirred for 30 min, concentrated and azeotroped with toluene (3×10 mL). Flash chromatography (EtOAc/pet. ether, 9:1) gave the triol **7** (42.5 mg, 83%,). $[\alpha]_D^{25}=44.6$ ($c=1.03$ in MeOH); ^{1}H NMR (500 MHz, CD_3OD): $\delta=2.67$–4.20 (13H; 1-H_a–6-H_b, 2′-H–6′-H_b), 4.43–4.46 (m, 2H; 2×CH_2Ph), 4.52 (d, $J=12.0$ Hz, 1H; CH_2Ph), 4.70 (d, $J=12.7$ Hz, 1H; CH_2Ph), 4.72 (d, $J=11.2$ Hz, 1H; CH_2Ph), 4.89 (d, $J=2.1$ Hz, 1H; 1′-H), 5.12 (s, 2H; CH_2), 5.15 (app. s, 1H; 1′-H), 7.03–7.42 ppm (m, 20H; 4×Ph); ^{13}C NMR ($CDCl_3$, 125 MHz): $\delta=59.5$, 68.0, 68.9, 69.0, 71.9, 72.5, 73.5, 74.2, 74.9, 79.5 (13C; C-1,2,3,4,5,6,1′,2′,3′,4′,5′,6′, CH_2) 127.8, 127.9, 128.0, 128.1, 128.16, 128.19, 128.4, 128.5, 128.6, 128.7, 137.9, 138.0, 138.3 (24C; Ph), 156.5 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{41}H_{46}N_4O_{10}$: 755.3287 [*M*+H]$^+$; found: 755.3300.

3,4,6-Tri-*O*-benzyl-α-D-mannopyranosyl-(1→3)-2-amino-*N*-benzyloxycarbonyl-1,2,5-trideoxy-1,5-imino-D-mannitol (8): DTT (51 mg, 0.331 mmol) was added to a solution of azide **7** (25 mg, 0.0331 mmol) in pyridine (1 mL) and $NaHCO_3/H_2CO_3$ buffer (0.625 mL, pH 9.16). The mixture was stirred at room temperature for 4 h, concentrated and azeotroped with toluene (5×10 mL). Flash chromatography (EtOAc/MeOH/H_2O, 94:4:2) gave the amine **8** (80%, 19.2 mg). ^{1}H NMR (500 MHz, CD_3OD): $\delta=2.89$ (t, $J=12.4$ Hz, 1H; 2-H), 3.21–4.13 (13C; m, 1-H_a, 1-H_b, 3,5-H, 6-H_a, 6-H_b, 1′–6_b′-H), 4.36 (t, $J=7.8$ Hz, 1H; 4-H),4.46–4.54 (m, 2H; 2×CH_2Ph), 4.58 (d, $J=12.0$ Hz, 1H; CH_2Ph), 4.66 (d, $J=11.8$ Hz, 1H; CH_2Ph), 4.77–4.81 (m, 2H; 2×CH_2Ph), 4.98 (d, $J=2.5$ Hz, 1H; 1′-H), 5.15 (s, 2H; CH_2), 7.16–7.47 ppm (m, 20H; Ph); ^{13}C NMR ($CDCl_3$, 125 MHz): $\delta=46.8$, 59.9, 65.6, 68.5, 69.4, 70.4, 72.6, 73.7, 74.4, 75.4, 75.7, 78.1, 80.1, 100.8 (16C; C-1–6, C1′–6′, 4×CH_2), 128.81, 128.84, 129.2, 129.28, 128.30, 129.3, 129.4, 129.5, 138.0, 139.3, 139.5, 139.6 ppm (24C; Ph); HRMS (ESI, +ve): *m/z* calcd for $C_{41}H_{48}N_2O_{10}$: 729.3385 [*M*+H]$^+$; found: 729.3398.

α-D-Mannopyranosyl-(1→3)-2-amino-1,2,5-trideoxy-1,5-imino-D-mannitol (1): The triol **8** (19.2 mg, 0.0264 mmol) in EtOAc/MeOH/H_2O (2:2:1, 3 mL) and 10% HCl in methanol (0.3 mL) was treated with $Pd(OH)_2$/C (50 mg) and H_2 (20 atm, 18 h). The suspension was filtered, concentrated and purified with cation and anion resin (eluted with aqueous NH_3) to give $ManNH_2DMJ$ (**1**; 70%, 6.02 mg) as a colourless oil. $[\alpha]_D^{25}=17.2$ ($c=0.08$ in H_2O); ^{1}H NMR (500 MHz, D_2O): $\delta=2.78$–2.84 (m, 1H; 5-H), 3.09 (dd, $J_{1a,1b}=14.0$, $J_{1a,2}=2.1$ Hz, 1H; 1-H_a), 3.25 (dd, $J_{1a,1b}=14.0$, $J_{1a,2}=3.2$ Hz, 1H; 1-H_b), 3.62–3.95 (m, 9H; 2,3,4,4′,5′-H, 6-H_a, 6′-H_a, 6-H_b, 6′-H_b), 3.98 (dd, $J_{3',4'}=9.2$, $J_{2',3'}=4.3$ Hz, 1H; 3′-H), 4.09 (dd, $J_{2',3'}=3.3$, $J_{1',2'}=1.8$ Hz, 1H; 2′-H), 5.24 ppm (d, $J_{1',2'}=1.6$ Hz, 1H; 1′-H); ^{13}C NMR (125 MHz, D_2O): $\delta=44.5$, 50.4, 60.0, 60.8, 61.0, 66.6, 67.3, 69.7, 70.1, 73.7, 77.3, 101.6 ppm; HRMS (ESI, +ve): *m/z* calcd for $C_{12}H_{24}N_2O_8$: 325.1605 [*M*+H]$^+$; found: 325.1606.

4-Methylphenyl 2,4,6-tri-*O*-benzyl-3-*O*-(2-naphthylmethyl)-1-thio-α-D-mannopyranoside (10): A dry solution of the alcohol **9**[23] (167 mg, 0.30 mmol) in DMF (5 mL) was cooled to 0 °C. The solution was charged with NaH (60% dispersion in mineral oil, 36 mg, 0.9 mmol) and the mixture stirred for 30 min. 2-Bromomethylnaphthalene (79.6 mg, 0.36 mmol) was added and the mixture stirred overnight. The mixture was diluted with Et_2O (20 mL), poured into ice/water and washed with water (3×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$), the solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether, 15:85) to give the protected thioglycoside **10** (179.3 mg, 86%) as a colourless oil. $[\alpha]_D^{24} = +65$ ($c = 0.69$ in $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 2.28$ (s, 3H; TolMe), 3.78 (dd, $J_{5,6a} = 1.8$, $J_{6a,6b} = 10.9$ Hz, 1H; 6-H_a), 3.87 (dd, $J_{5,6b} = 5.2$, $J_{6a,6b} = 10.9$ Hz, 1H; 6-H_b), 3.97 (dd, $J_{2,3} = 3.0$, $J_{3,4} = 9.3$ Hz, 1H; 3-H), 4.04 (dd, $J_{1,2} = 3.0$, $J_{2,3} = 1.8$ Hz, 1H; 2-H), 4.11 (m, 1H; 4-H), 4.33 (ddd, $J_{4,5} = 9.8$, $J_{5,6a} = 5.1$, $J_{5,6b} = 1.6$ Hz, 1H; 5-H), 4.49 (d, $J = 11.9$ Hz, 1H; CH_2Ph), 4.57–4.67 (m, 3H; 3×CH_2Ph), 4.74 (m, 3H; CH_2Ph, 2×CH_2Nap), 4.96 (d, $J = 10.9$ Hz, 1H; CH_2Ph), 5.58 (d, $J_{1,2} = 1.5$ Hz, 1H; 1-H), 7.02 (app. d, $J = 7.9$ Hz, 2H; Tol), 7.21–7.37 (m, 17H; 3×Ph, Tol), 7.44–7.47 (m, 3H; Nap), 7.74–7.83 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 21.2$ (1C; TolMe), 69.3 (1C; C-6), 71.9 (1C; CH_2Ph), 72.2 (1C; CH_2Nap), 72.8 (1C; C-5), 73.3 (1C; CH_2Ph), 75.1 (1C; C-4), 75.2 (1C; CH_2Ph), 76.3 (1C; C-2), 80.3 (1C; C-3), 86.1 (1C; C-1), 125.9–126.5 (4C; Nap), 127.5–128.4 (18C; 3×Ph, Nap), 129.8 (2C; Tol), 132.3 (2C; Tol), 133.4, 135.8, 137.6, 138.0, 138.5, 138.6 ppm (6C; C_q); HRMS (ESI, +ve): *m/z* calcd for $C_{45}H_{44}O_5S$: 719.2802 $[M+Na]^+$; found: 719.2809.

2,4,6-Tri-*O*-benzyl-3-*O*-(2-naphthylmethyl)-α-D-mannopyranose (11): *N*-Iodosuccinimide (216 mg, 0.961 mmol) was added to a solution of the thioglycoside **10** (447 mg, 0.641 mmol) in acetone (1% aq., 10 mL) at 0 °C and left to stir for 2.5 h. The solution was quenched with aq. $Na_2S_2O_3$ (0.5 M, 10 mL), diluted with EtOAc (20 mL) and washed with aq. $Na_2S_2O_3$ (0.5 M, 3×20 mL), $NaHCO_3$ (2×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$), the solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether/Et_3N, 30:69.5:0.5) to afford the hemiacetals **11** (344 mg, 91%; α/β 3.3:1) as a white powder. α anomer: ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 3.69$ (dd, $J_{5,6a} = 6.6$, $J_{6a,6b} = 10.5$ Hz, 1H; 6-H_a), 3.74 (dd, $J_{5,6b} = 2.0$, $J_{6a,6b} = 10.4$ Hz, 1H; 6-H_b), 3.83 (dd, $J_{1,2} = 2.0$, $J_{2,3} = 2.8$ Hz, 1H; 2-H), 3.91 (t, $J_{3,4} = J_{4,5} = 9.6$ Hz, 1H; 4-H), 4.05 (dd, $J_{2,3} = 3.0$, $J_{3,4} = 9.4$ Hz, 1H; 3-H), 4.10 (ddd, $J_{4,5} = 8.7$, $J_{5,6a} = 5.8$, $J_{5,6b} = 1.9$ Hz, 1H; 5-H), 4.51–4.59 (m, 3H; 3×CH_2Ph), 4.74–4.76 (m, 4H; 2×CH_2Ph, 2×CH_2Nap), 4.94 (d, $J = 11.0$ Hz, 1H; CH_2Ph), 5.27 (d, $J_{1,2} = 1.8$ Hz, 1H; 1-H), 7.18–7.41 (m, 17H; 3×Ph), 7.45–7.47 (m, 3H; Nap), 7.72–7.83 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 69.7$ (1C; C-6), 71.4 (1C; C-5), 72.2 (1C; CH_2Nap), 72.7 (1C; CH_2Ph), 73.3 (1C; CH_2Ph), 75.1 (1C; CH_2Ph), 75.1 (1C; C-2), 75.3 (1C; C-4), 79.8 (1C; C-3), 92.6 (1C; C-1), 125.8–126.3 (4C; Nap), 127.6–128.5 (18C; 3×Ph, Nap), 133.0, 133.4, 136.1, 138.0, 138.5 ppm (6C; C_q); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{38}O_6$: 608.3007 $[M+NH_4]^+$; found: 608.3007.

2,4,6-Tri-*O*-benzyl-3-*O*-(2-naphthylmethyl)-D-mannonolactone (12): A solution of the hemiacetal **11** (742 mg, 1.26 mmol) in acetic anhydride (6.1 mL) and dry DMSO (6.6 mL) was stirred under N_2 for 22 h. The mixture was diluted with EtOAc (20 mL), quenched with ice and washed with water (3×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$) and the solvent was evaporated. Azeotropic toluene was used to remove any residual AcOH to afford the crude lactone **12** (823 mg), which was used directly in the next step. A portion of **12** obtained from a separate experiment was purified by flash chromatography (EtOAc/pet. ether, 1:9) to yield analytically pure **12** as a colourless oil. $[\alpha]_D^{25} = +4.05$ ($c = 0.44$ in $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 3.61$ (m, 2H; 6-H_a, 6-H_b), 3.80 (dd, $J_{2,3} = 1.5$, $J_{3,4} = 7.2$ Hz, 1H; 3-H), 4.09 (dd, $J_{1,2} = 2.6$, $J_{2,3} = 1.6$ Hz, 1H; 2-H), 4.23 (m, 2H; 5-H, 4-H), 4.38 (d, $J = 2.6$ Hz, 1H; CH_2Ph), 4.48 (app. d, 2H; 2×CH_2Ph), 4.56 (d, $J = 11.8$ Hz, 1H; CH_2Ph), 4.77 (d, $J = 12.5$ Hz, 1H; CH_2Ph), 4.94 (d, $J = 12.5$ Hz, 1H; CH_2Ph), 5.06 (m, 2H; 2×CH_2Nap), 6.96–7.45 (m, 18H; 3×Ph, Nap), 7.69–7.78 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 69.0$ (1C; C-6), 71.6 (1C; C-4), 72.8 (1C; CH_2Ph), 72.9 (1C; CH_2Nap), 73.3 (1C; CH_2Ph), 75.5 (1C; CH_2Ph), 75.8 (1C; C-3), 76.5 (1C; C-2), 78.4 (1C; C-5), 125.9–126.1 (3C; Nap), 126.9 (1C; Nap), 127.6–128.9 (18C; 3×Ph, Nap), 132.9, 133.0, 135.0, 136.7, 137.3, 137.6 (6C; C_q), 169.3 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{36}O_6$: 606.2850 $[M+NH_4]^+$; found: 606.2853.

2,4,6-Tri-*O*-benzyl-3-*O*-(2-naphthylmethyl)-D-mannonamide (13): A dry-ice/acetone cold finger cooling trap was used to condense ammonia (50 mL) into a solution of the crude lactone **12** (823 mg) in dry THF (30 mL) at −78 °C. The solution was allowed to reflux at 0 °C for 4 h. The mixture was then evaporated to dryness to afford the crude amide **13** (771 mg), which was used directly in the next step. A portion obtained from an independent experiment was purified by flash chromatography (EtOAc/pet. ether, 3:2) to yield analytically pure **13** as a yellow solid. M.p. 120 °C; $[\alpha]_D^{25} = +7.21$ ($c = 0.41$ in $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 3.20$ (d, $J_{5,OH} = 6.2$ Hz, 1H; OH), 3.61 (m, 2H; 6-H_a, 6-H_b), 3.87 (dd, $J_{3,4} = 5.9$, $J_{4,5} = 7.3$ Hz, 1H; 4-H), 3.98 (m, 1H; 5-H), 4.13 (dd, $J_{2,3} = 3.5$, $J_{3,4} = 5.8$ Hz, 1H; 3-H), 4.33 (d, $J_{2,3} = 3.5$ Hz, 1H; 2-H), 4.43–4.60 (m, 6H; 6×CH_2Ph), 4.82 (s, 2H; 2×CH_2Nap), 5.50 (brs, 1H; NH), 6.54 (brs, 1H; NH), 7.11–7.27 (m, 15H; 3×Ph), 7.38–7.43 (m, 3H; Nap), 7.68–7.76 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 71.1$ (1C; C-5), 71.4 (1C; C-6), 72.9 (1C; CH_2Ph), 73.6 (1C; CH_2Ph), 74.6 (1C; CH_2Ph), 75.0 (1C; CH_2Nap), 79.1 (1C; C-4), 80.2 (1C; C-2), 81.6 (1C; C-3), 126.0–126.3 (3C; Nap), 126.9 (1C; Nap), 127.8–128.7 (18C; 3×Ph, Nap), 133.1, 133.4, 135.7, 137.2, 138.2, 138.4 (6C; C_q), 173.4 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{39}NO_6$: 606.2844 $[M+H]^+$; found: 606.2850 ppm.

(3*S*,4*S*,5*S*,6*R*/*S*)-3,5-Bis(benzyloxy)-6-(benzyloxymethyl)-6-hydroxy-4-(2-naphthylmethoxy)piperidin-2-one (15): A solution of the crude amide **13** (771 mg) in acetic anhydride (6.1 mL) and dry DMSO (6.6 mL) was stirred under N_2 for 21 h. The reaction mixture was diluted with EtOAc (20 mL), quenched with ice and washed with water (3×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$) and the solvent was evaporated to afford the keto-amide **14** as a white solid. A dry-ice/acetone cold finger was used to condense ammonia (20 mL) into a solution of the crude keto-amide in dry methanol (30 mL) at 0 °C. The solution was allowed to warm to room temperature and stirred under N_2 for 16 h. The solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether, 1:1) to give a separable mixture of the hydroxy-lactams **15** (669 mg, 88% over four steps; D-*manno*/L-*gulo* 2.2:1). ^{1}H NMR (500 MHz, $CDCl_3$), partial spectrum of the mixture of diastereomers: $\delta = 3.38$ (d, $J = 9.8$ Hz, 1H; CH_2(C6) D-*manno*), 3.43 (d, $J = 9.6$ Hz, 1H; CH_2(C6) L-*gulo*), 3.47 (d, $J = 9.8$ Hz, 1H; CH_2(C6) D-*manno*), 3.57 (d, $J = 9.6$ Hz, 1H; CH_2(C6) L-*gulo*), 3.72 (brs, 1H; OH), 4.22 (d, $J_{3,4} = 3.0$ Hz, 1H; 3-H D-*manno*), 4.26 (d, $J_{3,4} = 3.1$ Hz, 1H; 3-H L-*gulo*), 4.98 (d, $J = 12.5$ Hz, 1H; CH_2Ph D-*manno*), 5.10 (d, $J = 12.3$ Hz, 1H; CH_2Ph L-*gulo*), 6.33 (brs, 1H; NH L-*gulo*), 6.22 ppm (brs, 1H; NH D-*manno*); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 74.0$ (1C; CH_2(C6) D-*manno*), 74.5 (1C; C-3 D-*manno*), 169.6 (1C; C=O D-*manno*), 170.2 ppm (1C; C=O L-*gulo*); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{37}NO_6$: 604.2694 $[M+H]^+$; found: 606.2698 ppm.

(3*S*,4*S*,5*S*,6*R*)-3,5-Bis(benzyloxy)-6-(benzyloxymethyl)-4-(2-naphthylmethoxy)piperidin-2-one (16) and (3*S*,4*S*,5*S*,6*S*)-3,5-bis(benzyloxy)-6-(benzyloxymethyl)-4-(2-naphthylmethoxy)piperidin-2-one (17): Sodium cyanoborohydride (90.4 mg, 1.44 mmol) was added to a solution of the hydroxy-lactams **15** (86.9 mg, 0.144 mmol) and formic acid (0.52 mL) in dry acetonitrile (3 mL) and the mixture stirred under N_2 for 20 h. Sodium cyanoborohydride (90.4 mg, 1.44 mmol) was added and the reaction mixture was stirred for a further 24 h when TLC analysis (EtOAc/pet. ether, 1:3) indicated complete consumption of the starting material. The mixture was diluted with EtOAc (20 mL) and washed with aq. sat. $NaHCO_3$ (3×20 mL) and brine (1×20 mL). The aqueous extracts were treated with sodium hypochlorite prior to disposal. The organic extracts were dried ($MgSO_4$), the solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether, 1:1) to afford the L-*gulo* lactam **16** (28.2 mg, 33%) and the D-*manno* lactam **17** (32.5 mg, 38%), both as colourless oils.

Characterisation for **16**: $[\alpha]_D^{23} = -57$ ($c = 0.535$ in $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 3.36$ (dd, $J_{6,6a} = 4.27$, $J_{6a,6b} = 9.11$ Hz, 1H; CH_2(C6)), 3.46 (m, 2H; 6-H, CH_2(C6)), 3.57 (m, 1H; 3-H), 3.91 (dd, $J_{3,4} = 3.1$, $J_{4,5} = 4.4$ Hz, 1H; 4-H), 3.95 (m, 1H; 6-H), 4.08–4.19 (m, 3H; 2×CH_2Ph, 5-H), 4.40 (m, 2H; 2×CH_2Ph), 4.66 (d, $J = 12.4$ Hz, 1H; CH_2Ph), 4.71 (d, $J = 12.3$ Hz, 1H; CH_2Nap), 4.93 (d, $J = 12.3$ Hz, 1H; CH_2Nap), 5.10 (d, $J = 12.4$ Hz, 1H; CH_2Ph), 5.83 (br s, 1H; NH), 6.84 (app. d, $J = 7.05$ Hz, 2H; Ph), 7.07–7.45 (m, 16H; Ph, Nap), 7.62 (s, 1H; Nap), 7.72–7.79 ppm (m, 3H; Nap); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 52.8$ (1C; C-6), 70.3 (1C; CH_2(C6)), 72.5 (1C; CH_2Nap), 73.6 (1C; CH_2Ph), 73.6 (1C; CH_2Ph), 73.7 (1C; CH_2Ph), 74.2 (1C; C-5), 74.3 (1C; C-3), 74.8 (1C; C-4), 126.0–126.3 (3C; Nap), 126.8 (1C; Nap), 127.8–128.6 (18C; 3×Ph, Nap), 133.2, 133.3, 135.6, 137.0, 137.6, 138.4 (6C; C_q), 171.3 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{37}NO_5$: 588.2749 $[M+H]^+$; found: 588.2747.

Characterisation for **17**: $[\alpha]_D^{25} = -9.49$ ($c = 0.715$ in $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 3.41$ (m, 1H; CH_2(C6)), 3.54 (m, 2H; 6-H, CH_2(C6)), 3.66 (t, $J_{4,5} = J_{5,6} = 5.2$ Hz, 1H; 5-H), 3.98 (dd, $J_{3,4} = 2.9$, $J_{4,5} = 5.0$ Hz, 1H; 4-H), 4.18 (d, $J_{3,4} = 2.9$ Hz, 1H; 3-H), 4.38 (d, $J = 11.6$ Hz, 1H; CH_2Ph), 4.42–4.49 (m, 2H; 2×CH_2Ph), 4.55 (d, $J = 11.6$ Hz, 1H; CH_2Ph), 4.69 (d, $J = 12.1$ Hz, 1H; CH_2Ph), 4.74 (d, $J = 12.2$ Hz, 1H; CH_2Nap), 4.88 (d, $J = 12.2$ Hz, 1H; CH_2Nap), 5.06 (d, $J = 12.2$ Hz, 1H; CH_2Ph), 5.91 (br s, 1H; NH), 7.08–7.49 (m, 18H; 3×Ph, Nap), 7.72–7.84 ppm (m, 4H; Nap); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 55.5$ (1C; C-6), 71.5 (1C; CH_2(C6)), 72.9 (1C; CH_2Nap), 72.9 (1C; CH_2Ph), 73.4 (1C; CH_2Ph), 73.5 (1C; CH_2Ph), 75.0 (1C; C-5), 75.2 (1C; C-3), 77.8 (1C; C-4), 126.1–126.3 (3C; Nap), 127.0 (1C; Nap), 127.8–128.6 (18C; 3×Ph, Nap), 133.2, 133.3, 135.5, 137.5, 138.1 (6C; C_q), 169.6 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{37}NO_5$: 588.2744 $[M+H]^+$; found: 588.2747.

(3*S*,4*S*,5*S*,6*S*)-3,5-Bis(benzyloxy)-6-(benzyloxymethyl)-4-(2-naphthylmethoxy)piperidin-2-thione (18): Lawesson's reagent (202 mg, 0.50 mmol) was added to a mixture containing the mannonolactam **17** (98 mg, 0.167 mmol), pyridine (6.7 μL, 0.083 mmol), freshly activated 4 Å molecular sieves and distilled toluene (6 mL) and the mixture was stirred for 20 h. The mixture was then filtered, stirred with MeOH (1.68 mL) for 2 h and the solvent removed under reduced pressure. The residue obtained was subjected to flash chromatography (EtOAc/pet. ether, 20:80) to afford the thionolactam **18** (94 mg, 93%) as a white solid. M.p. 147 °C; $[\alpha]_D^{23} = -52$ ($c = 0.215$ in $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 3.43$ (m, 1H; CH_2(C6)), 3.56 (m, 2H; 6-H, CH_2(C6)), 3.83 (apt. t, 1H; 5-H), 3.91 (dd, $J_{3,4} = 2.6$, $J_{4,5} = 7.2$ Hz, 1H; 4-H), 4.42 (d, $J_{3,4} = 2.5$ Hz, 1H; 3-H), 4.44–4.52 (m, 3H; 3×CH_2Ph), 4.68–4.73 (m, 2H; CH_2Nap, CH_2Ph), 4.79 (d, $J = 12.1$ Hz, 1H; CH_2Nap), 4.83 (d, $J = 12.0$ Hz, 1H; CH_2Ph), 5.08 (d, $J = 12.1$ Hz, 1H; CH_2Ph), 7.14–7.52 (m, 18H; 3×Ph, Nap), 7.73–7.85 (m, 4H; Nap), 8.13 ppm (br s, 1H; NH); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 59.8$ (1C; C-6), 70.6 (1C; CH_2(C6)), 72.5 (1C; CH_2Nap), 73.2 (1C; CH_2Ph), 73.5 (1C; CH_2Ph), 73.7 (1C; CH_2Ph), 74.2 (1C; C-5), 78.3 (1C; C-4), 79.8 (1C; C-3), 125.9–126.3 (3C; Nap), 126.8 (1C; Nap), 127.8–128.7 (18C; 3×Ph, Nap), 133.1, 133.3, 135.4, 137.3, 137.6, 138.0 (6C; C_q), 200.0 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{38}H_{37}NO_4S$: 604.2516 $[M+H]^+$; found: 604.2524 [].

(5*R*,6*R*,7*S*,8*S*)-7-(2-Naphthylmethoxy)-6,8-bis(benzyloxy)-5-(benzyloxymethyl)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridine (20) and (5*R*,6*R*,7*S*,8*R*)-7-(2-naphthylmethoxy)-6,8-bis(benzyloxy)-5-(benzyloxymethyl)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridine (21): Thionolactam **18** (256 mg, 0.424 mmol) was dissolved in aminoacetaldehyde dimethyl acetal (0.69 mL, 6.33 mmol) and the mixture stirred under N_2 for 18 h. The mixture was diluted with Et_2O (20 mL) and washed with H_2O (2×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$) and the solvent removed under reduced pressure to afford the amidines **19** as a colourless residue. *p*-Toluenesulfonic acid monohydrate (0.14 g, 0.74 mmol) was added to a solution of the crude amidines in toluene (9.5 mL) and the mixture was stirred at 60 °C overnight. The mixture was then diluted with DCM (20 mL) and washed with $NaHCO_3$ (2×20 mL) and brine (1×20 mL). The organic extracts were dried ($MgSO_4$), the solvent was removed under reduced pressure and the residue was subjected to flash chromatography (EtOAc/pet. ether, 1:1) to afford the glucoimidazole **20** (110 mg, 42% over two steps) as a colourless oil and the mannoimidazole **21** (83.3 mg, 32% over two steps) as a yellow oil.

Characterisation for **20**: $[\alpha]_D^{25} = +52$ ($c = 0.315$ in $CHCl_3$; lit.:[39] +52 (in $CHCl_3$)); ^{1}H NMR (600 MHz, $CDCl_3$): $\delta = 3.75$ (dd, $J_{5,5a} = 5.0$, $J_{5a,5b} = 10.3$ Hz, 1H; CH_2(C5)), 3.87 (m, 2H; 6-H, CH_2(C5)), 4.13 (dd, $J_{6,7} = 7.5$, $J_{7,8} = 5.8$ Hz, 1H; 7-H), 4.18 (m, 1H; 5-H), 4.45 (app. d, 2H; 2×CH_2Ph), 4.51 (d, $J = 11.2$ Hz, 1H; CH_2Ph), 4.78 (d, $J_{7,8} = 5.8$ Hz, 1H; 8-H), 4.84 (d, $J = 11.6$ Hz, 1H; CH_2Ph), 4.86 (d, $J = 11.2$ Hz, 1H; CH_2Ph), 4.89 (d, $J = 11.5$ Hz, 1H; CH_2Nap), 4.97 (d, $J = 11.5$ Hz, 1H; CH_2Ph), 5.19 (d, $J = 11.5$ Hz, 1H; CH_2Nap), 7.04 (s, 1H; 2-H), 7.12 (s, 1H; 3-H), 7.14–7.48 (m, 18H; 3×Ph, Nap), 7.68–7.83 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 58.3$ (1C; C-5), 68.5 (1C; CH_2(C5)), 72.9 (1C; CH_2Nap), 73.4 (1C; CH_2Ph), 74.3 (1C; CH_2Ph), 74.4 (1C; CH_2Ph), 74.5 (1C; C-8), 76.2 (1C; C-6), 82.2 (1C; C-7), 117.4 (1C; C-2), 126.1–126.9 (3C; Nap), 127.7 (1C; Nap), 127.8–128.6 (18C; 3×Ph, Nap), 129.5 (1C; C-3), 133.2, 133.4, 135.5, 137.4, 137.7, 138.4 (6C; C_q), 144.2 ppm (C_q, imidazole).

Characterisation for **21**: $[\alpha]_D^{25} = -24$ ($c = 0.24$ in $CHCl_3$; lit.:[39] −20 (in $CHCl_3$)); ^{1}H NMR (600 MHz, $CDCl_3$): $\delta = 3.57$ (dd, $J_{5,5a} = 7.1$, $J_{5a,5b} = 10.1$ Hz, 1H; CH_2(C5)), 3.71 (dd, $J_{5,5a} = 3.4$, $J_{5a,5b} = 10.1$ Hz, 1H; CH_2(C5)), 3.84 (dd, $J_{6,7} = 9.3$, $J_{7,8} = 3.1$ Hz, 1H; 7-H), 4.06 (m, 1H; 5-H), 4.25 (dd, $J_{5,6} = 9.3$, $J_{6,7} = 7.2$ Hz, 1H; 6-H), 4.39 (m, 2H; 2×CH_2Ph), 4.56–4.66 (m, 3H; 2×CH_2Ph, CH_2Nap), 4.69 (d, $J = 12.2$ Hz, 1H; CH_2Nap), 4.74 (d, $J = 12.0$ Hz, 1H; CH_2Ph), 4.78 (d, $J_{7,8} = 3.0$ Hz, 1H; 8-H), 4.96 (d, $J = 11.2$ Hz, 1H; CH_2Ph), 6.98 (s, 1H; 3-H), 7.09 (s, 1H; 2-H), 7.17–7.39 (m, 18H; 3×Ph, Nap), 7.62–7.74 ppm (m, 4H; Nap); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 60.0$ (1C; C-5), 68.3 (1C; C8), 70.6 (1C; CH_2Nap), 71.2 (1C; CH_2(C5)), 71.8 (1C; CH_2Ph), 73.3 (1C; CH_2Ph), 74.3 (1C; C-6), 75.0 (1C; CH_2Ph), 80.2 (1C; C-3), 119.5 (1C; C-2), 125.2–126.9 (3C; Nap), 126.7 (1C; Nap), 128.6–127.7 (18C; 3×Ph, Nap), 129.4 (1C; C-3), 133.2, 133.3, 135.4, 137.6, 138.2, 138.3 (6C; C_q), 143.0 ppm (C_q, imidazole).

(5*R*,6*R*,7*S*,8*R*)-6,8-Bis(benzyloxy)-5-(benzyloxymethyl)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridin-7-ol (22): DDQ (25.2 mg, 0.111 mmol) was added to a solution of the mannoimidazole **21** (22.6 mg, 0.037 mmol) in DCM/H_2O (9:1, 1 mL) and the reaction mixture was stirred at room temperature overnight. DDQ (25 mg,

0.11 mmol) was again added and the mixture stirred for 3 days when TLC analysis (EtOAc/pet. ether, 8:2) indicated complete consumption of the starting material. The mixture was then diluted with DCM (20 mL), washed with water (3×20 mL) and aq. sat. $NaHCO_3$ (3×20 mL), dried ($MgSO_4$), filtered and concentrated. The crude product was purified by flash chromatography (EtOAc/pet. ether, 80:20 to 100:0) to afford the alcohol **22** (11.7 mg, 67%) as a yellow oil. $[\alpha]_D^{24} = -35$ ($c = 0.585$ in $CHCl_3$; lit.:[40] -6 (in $CHCl_3$)); ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 3.64$ (dd, $J_{5,5a} = 5.9$, $J_{5a,5b} = 10.2$ Hz, 1H; CH_2(C5)), 3.78 (dd, $J_{5,5a} = 2.5$, $J_{5a,5b} = 10.2$ Hz, 1H; CH_2(C5)), 4.03 (m, 3H; 7-H, 6-H, 5-H), 4.42 (app. s, 2H; 2×CH_2Ph), 4.54 (d, $J =$ 11.2 Hz, 1H; CH_2Ph), 4.65 (d, $J = 11.6$ Hz, 1H; CH_2Ph), 4.70 (d, $J_{7,8} =$ 3.3 Hz, 1H; 8-H), 4.85 (d, $J = 11.6$ Hz, 1H; CH_2Ph), 4.90 (d, $J =$ 11.2 Hz, 1H; CH_2Ph), 7.05 (s, 1H; 3-H), 7.13 (s, 1H; 2-H), 7.19–7.28 ppm (m, 15H; 3×Ph); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 59.1$ (1C; C-5), 70.2 (1C; CH_2(C5)), 71.2 (2C; C-8, CH_2Ph), 72.4 (1C; C-6), 73.2 (1C; CH_2Ph), 74.6 (1C; CH_2Ph), 75.3 (1C; C-7), 118.9 (1C; C-2), 127.7–128.5 (15C; 3×Ph), 129.6 (1C; C-3), 137.5, 137.7, 137.8 (3C; C_q), 142.3 ppm (C_q, imidazole).

(5*R*,6*R*,7*S*,8*R*)-7-(2-*O*-Acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyloxy)-6,8-bis(benzyloxy)-5-(benzyloxymethyl)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridine (23): A mixture of the alcohol **22** (13.8 mg, 0.029 mmol), 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl trichloroacetimidate (**5**;[22] 32.5 mg, 0.051 mmol) and freshly activated 4 Å molecular sieves in toluene (1.5 mL) was stirred at room temperature for 30 min. Triflic acid (1 μL, 0.011 mmol) was added to the mixture at −20 °C and the mixture was stirred for 1 h, then at 0 °C for 20 min, and at room temperature for another 20 min, quenched with pyridine (1 drop) and filtered through a pad of Celite. The solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether/ Et_3N 80:19:1) to recover alcohol **26** (6.4 mg) and afford the disaccharide **23** (12.9 mg, 47%) as a colourless oil. $[\alpha]_D^{23} = +7.2$ ($c = 0.175$ in $CHCl_3$); ^{1}H NMR (600 MHz, $CDCl_3$): $\delta = 2.11$ (s, 3H; Ac), 3.49 (dd, $J_{5',5a'} = 1.7$, $J_{5a'',5b'} = 10.9$ Hz, 1H; CH_2(C5′)), 3.55 (dd, $J_{5,5a} = 6.7$, $J_{5a,5b} = 10.2$ Hz, 1H; CH_2(C5)), 3.63 (dd, $J_{5',5b'} = 3.5$, $J_{5a'',5b'} = 10.8$ Hz, 1H; CH_2(C5′)), 3.67 (dd, $J_{5,5b} = 3.2$, $J_{5a,5b} = 10.2$ Hz, 1H; CH_2(C5)), 3.87 (m, 1H; 5′-H), 3.93 (t, $J_{3',4'} = J_{4',5'} = 9.5$ Hz, 1H; 4′-H), 4.01 (dd, $J_{2',3'} = 3.3$, $J_{3',4'} = 9.5$ Hz, 1H; 3′-H), 4.07 (dd, $J_{6,7} = 9.5$, $J_{7,8} = 3.1$ Hz, 1H; 7-H), 4.13 (1H, m, 5-H), 4.29 (dd, $J_{5,6} = 7.1$, $J_{6,7} =$ 9.5 Hz, 1H; 6-H), 4.41 (m, 2H; 2×CH_2Ph), 4.46 (d, $J = 10.9$ Hz, 1H; CH_2Ph), 4.51 (d, $J = 11.3$ Hz, 1H; CH_2Ph), 4.54 (d, $J = 12.0$ Hz, 1H; CH_2Ph), 4.57 (d, $J = 11.3$ Hz, 1H; CH_2Ph), 4.64 (app. d, 3H, 3× CH_2Ph), 4.81 (d, $J_{2,3} = 3.1$ Hz, 1H; 2-H), 4.84 (m, 2H; 2×CH_2Ph), 5.19 (d, $J_{1',2'} = 1.6$ Hz, 1H; 1′-H), 5.48 (dd, $J_{1',2'} = 1.6$, $J_{2',3'} = 3.3$ Hz, 1H; 2′-H), 7.07 (s, 1H; 3-H), 7.14 (s, 1H; 2-H), 7.08–7.34 ppm (m, 30H; 6× Ph); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 21.2$ (1C; Me), 60.0 (1C; C-5), 68.5 (1C; C-6′), 69.1 (1C; C-2′), 70.3 (1C; CH_2Ph), 70.8 (1C; CH_2(C5)), 70.9 (1C; C-8), 72.1 (1C; CH_2Ph), 72.4 (1C; C-5′), 73.4 (1C; CH_2Ph), 73.7 (1C; CH_2Ph), 74.2 (1C; C-4′), 74.4 (1C; C-6), 75.1 (2C; CH_2Ph), 78.2 (1C; C-3′), 80.3 (1C; C-7), 100.1 (1C; C-1′), 119.4 (1C; C-2), 127.6–128.7 (30C; 6×Ph), 129.5 (1C; C-3), 137.6, 137.7, 137.9, 138.1, 138.2, 138.8 (6C; C_q), 142.6 (C_q, imidazole), 170.4 ppm (1C; C=O); HRMS (ESI, +ve): *m/z* calcd for $C_{58}H_{60}N_2O_{10}$: 945.4321 $[M+H]^+$; found: 945.4322.

(5*R*,6*R*,7*S*,8*R*)-7-(3,4,6-Tri-*O*-benzyl-α-D-mannopyranosyloxy)-6,8-bis(benzyloxy)-5-(benzyloxymethyl)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridine (24): K_2CO_3 (1 mg, 0.007 mmol) was added to a solution of the acetate **23** (13.1 mg, 0.014 mmol) in dry methanol (0.3 mL) and the resulting suspension was stirred at room temperature for 6.5 h. The reaction mixture was quenched with acetic acid (5 μL, 0.087 mmol), the solvent was removed under reduced pressure and the resulting residue was subjected to flash chromatography (EtOAc/pet. ether/Et_3N 50:49.5:0.5) to afford the alcohol **24** (5.8 mg, 46%) as a colourless oil. $[\alpha]_D^{24} = +13$ ($c = 0.305$ in $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$): $\delta = 2.40$ (d, $J_{2',OH} = 2.5$ Hz, 1H; OH), 3.49 (dd, $J_{5',6a'} = 1.8$, $J_{6a'',6b'} = 10.8$ Hz, 1H; 6′-H_a), 3.58 (m, 2H; CH_2(C5), 6′-H_b), 3.70 (dd, $J_{5,5a} = 3.2$, $J_{5a,5b} = 10.1$ Hz, 1H; CH_2(C5)), 3.87 (m, 1H; 5′-H), 3.91 (m, 2H; 4′,3′-H), 4.03 (m, 1H; 2′-H), 4.08 (dd, $J_{6,7} = 9.6$, $J_{7,8} = 3.1$ Hz, 1H; 7-H), 4.13 (1H, m, 5-H), 4.28 (dd, $J_{5,6} = 7.3$, $J_{6,7} =$ 9.6 Hz, 1H; 6-H), 4.40–4.53 (m, 5H; 5×CH_2Ph), 4.57–4.68 (m, 5H; 5×CH_2Ph), 4.79 (m, 2H; 2×CH_2Ph), 4.85 (d, $J_{7,8} = 3.1$ Hz, 1H; 8-H), 5.23 (d, $J_{1',2'} = 1.5$ Hz, 1H; 1′-H), 7.08 (s, 1H; 3-H), 7.14 (s, 1H; 2-H), 7.11–7.35 ppm (m, 30H; 6×Ph); ^{13}C NMR (125 MHz, $CDCl_3$): $\delta = 60.0$ (1C; C-5), 68.6 (1C; C-6′), 69.0 (1C; C-2′), 70.3 (1C; CH_2Ph), 70.7 (1C; C-8), 71.1 (1C; CH_2(C5)), 72.0 (1C; C-5′), 72.4 (1C; CH_2Ph), 73.4 (1C; CH_2Ph), 73.7 (1C; CH_2Ph), 74.3 (2C; C-6,3′), 75.1 (2C; CH_2Ph), 80.1 (1C; C-4′), 80.4 (1C; C-7), 101.8 (1C; C-1′), 119.3 (1C; C-2), 127.6–128.7 (30C; 6×Ph), 129.6 (1C; C-3), 137.6, 137.8, 138.1, 138.3, 138.7 (6C; C_q), 142.7 ppm (C_q, imidazole); HRMS (ESI, +ve): *m/z* calcd for $C_{56}H_{58}N_2O_9$: 903.4215 $[M+H]^+$; found: 903.4214.

(5*R*,6*R*,7*S*,8*R*)-6,8-Dihydroxy-5-(hydroxymethyl)-7-(α-D-mannopyranosyloxy)-5,6,7,8-tetrahydroimidazo[1,2-*a*]pyridine (2): Pd(OH)2/C (20%, 24 mg) was added to a solution of the deacetylated disaccharide **24** (12.6 mg, 0.014 mol) in EtOAc/MeOH/H_2O (5:17:3, 1.50 mL) and AcOH (0.34 mL). The reaction vessel was filled with H_2 (34 bar) and agitated for 4 days. At this point TLC analysis (EtOAc/MeOH/H_2O, 7:3:2) indicated complete conversion to a single species along with baseline by-products. The suspension was filtered through a pad of Celite, the solvent was evaporated and the resulting residue was subjected to flash chromatography (EtOAc/MeOH/H_2O, 5:2:1) to afford ManManIm (**2**; 2.4 mg, 48%) as a colourless residue. $[\alpha]_D^{27} = +13$ ($c = 0.12$ in H_2O); ^{1}H NMR (500 MHz, D_2O): $\delta = 3.57$ (t, $J_{3',4'} = J_{4',5'} = 9.8$ Hz, 1H; 4′-H), 3.66 (dd, $J_{5',6a'} = 6.3$, $J_{6a'',6b'} = 12.1$ Hz, 1H; 6′-H_a), 3.77 (m, 1H; 5′-H), 3.83 (m, 2H; 3′-H, 6′-H_b), 3.91 (m, 1H; 5-H), 3.95 (dd, $J_{5,5a} = 3.3$, $J_{5a,5b} =$ 12.7 Hz, 1H; CH_2(C5)), 3.99 (dd, $J_{6,7} = 10.2$, $J_{7,8} = 3.7$ Hz, 1H; 7-H), 4.02 (dd, $J_{1',2'} = 3.4$, $J_{2',3'} = 1.7$ Hz, 1H; 2′-H), 4.13 (dd, $J_{5,5b} = 2.6$, $J_{5a,5b} = 12.7$ Hz, 1H; CH_2(C5)), 4.27 (dd, $J_{5,6} = 8.6$, $J_{6,7} = 10.2$ Hz, 1H; 6-H), 4.97 (d, $J_{7,8} = 3.7$ Hz, 1H; 8-H), 5.23 (d, $J_{1',2'} = 1.6$ Hz, 1H; 1′-H), 7.01 (s, 1H; 3-H), 7.20 ppm (s, 1H; 2-H); ^{13}C NMR (125 MHz, D_2O): $\delta = 59.3$ (1C; CH_2(C5)), 60.9 (1C; C-5,6′), 63.5 (1C; C-8), 63.9 (1C; C-6), 66.7 (1C; C-4′), 69.9 (1C; C-2′), 70.3 (2C; C-4,3′), 73.5 (1C; C-5′), 78.1 (1C; C-7), 102.1 (1C; C-1′), 118.3 (1C; C-2), 128.7 (1C; C-3), 144.7 ppm (C_q, imidazole); HRMS (ESI, +ve): *m/z* calcd for $C_{14}H_{22}N_2O_9$: 363.1398 $[M+H]^+$; found: 363.1398.

Isothermal titration calorimetry (ITC): The binding affinity of Man2NH$_2$DMJ (**1**) to *Bt*GH99 was determined by using a Microcal iTC200 calorimeter (GE Healthcare/Malvern Instruments). The assay was carried out at 25 °C with 18×2 μL injections of the inhibitor (6 mM) titrated into the ITC cell containing 117 μM *Bt*GH99. Owing to the low affinity of the ligand, which prevented the observation of a sigmoidal binding isotherm, *N* was fixed at 1.[41] An initial ITC experiment was conducted by using 1 M inhibitor in the syringe and 52 μM protein with 24×1.5 μL injections. The dissociation constant (K_D), change in enthalpy (ΔH) and measurement uncertainty were calculated by using the MicroCal PEAQ-ITC Analysis Software (Malvern Instruments).

Crystallisation and data collection: *Bx*GH99 protein[10] was crystallised by using the vapour diffusion hanging drop method in 3 M sodium acetate at pH 7.4. Crystals were grown at 19 °C in a 24-well plate with 500 μL of reservoir solution in each well and sealed with vacuum grease. The droplet was created by mixing 1 μL of *Bx*GH99 solution (34 mg mL^{-1} in 25 mM HEPES buffer, pH 7.0, 100 mM NaCl) with 1 μL of the crystallant solution. Crystals were fished from the droplet by using a nylon cryoloop, without cryoprotection. Data

were collected at Diamond Light Source beamline i04 using X-rays with a wavelength of 0.979 Å.

Structure solution and refinement: Images containing diffraction patterns were indexed and integrated by using DIALS[42] through xia2.[43] The *hkl* index of each data set was then matched to a previous solution in Aimless.[44] Refinement was performed by using Refmac5[45] and real-space model building in Coot.[46] Model geometry and agreement with electron density were validated in Coot and Edstats.[47] The quality of the carbohydrates and nitrogen heterocycles were verified by using Privateer.[32] The modelling and refinement processes were aided by using ccp4i2 interface.[48]

Acknowledgements

The Australian Research Council is thanked for financial support (DP120101396, FT130100103). We thank Diamond Light Source for access to beamline i04 (proposal mx13587) that contributed to the results presented here. G.J.D. and L.F.S. were supported by the European Research Council (ERC-2012-AdG-32294 'Glycopoise'). G.J.D. thanks the Royal Society for the Ken Murray Research Professorship.

Conflict of interest

The authors declare no conflict of interest.

Keywords: enzymes · glycosidase · imidazole rings · inhibitors · X-ray crystallography

[1] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490 – 495.
[2] *Glycobiology* **2018**, *28*, 3 – 8.
[3] C. Rabouille, R. G. Spiro, *J. Biol. Chem.* **1992**, *267*, 11573 – 11578.
[4] S. E. Moore, R. G. Spiro, *J. Biol. Chem.* **1992**, *267*, 8443 – 8451.
[5] S. E. Moore, R. G. Spiro, *J. Biol. Chem.* **1990**, *265*, 13104 – 13112.
[6] W. A. Lubas, R. G. Spiro, *J. Biol. Chem.* **1988**, *263*, 3990 – 3998.
[7] W. A. Lubas, R. G. Spiro, *J. Biol. Chem.* **1987**, *262*, 3775 – 3781.
[8] Z. Hakki, A. J. Thompson, S. Bellmaine, G. Speciale, G. J. Davies, S. J. Williams, *Chem. Eur. J.* **2015**, *21*, 1966 – 1977.
[9] F. Cuskin, E. C. Lowe, M. J. Temple, Y. Zhu, E. A. Cameron, N. A. Pudlo, N. T. Porter, K. Urs, A. J. Thompson, A. Cartmell, A. Rogowski, B. S. Hamilton, R. Chen, T. J. Tolbert, K. Piens, D. Bracke, W. Vervecken, Z. Hakki, G. Speciale, J. L. Munoz-Munoz, A. Day, M. J. Pena, R. McLean, M. D. Suits, A. B. Boraston, T. Atherly, C. J. Ziemer, S. J. Williams, G. J. Davies, D. W. Abbott, E. C. Martens, H. J. Gilbert, *Nature* **2015**, *517*, 165 – 169.
[10] A. J. Thompson, R. J. Williams, Z. Hakki, D. S. Alonzi, T. Wennekes, T. M. Gloster, K. Songsrirote, J. E. Thomas-Oates, T. M. Wrodnigg, J. Spreitz, A. E. Stutz, T. D. Butters, S. J. Williams, G. J. Davies, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 781 – 786.
[11] J. Munoz-Munoz, A. Cartmell, N. Terrapon, B. Henrissat, H. J. Gilbert, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4936 – 4941.
[12] G. Speciale, M. Farren-Dai, F. S. Shidmoossavee, S. J. Williams, A. J. Bennet, *J. Am. Chem. Soc.* **2016**, *138*, 14012 – 14019.
[13] U. Spohr, M. Bach, R. G. Spiro, *Can. J. Chem.* **1993**, *71*, 1928 – 1942.
[14] S. Hiraizumi, U. Spohr, R. G. Spiro, *J. Biol. Chem.* **1993**, *268*, 9927 – 9935.
[15] H. Ardron, T. D. Butters, F. M. Platt, M. R. Wormald, R. A. Dwek, G. W. J. Fleet, G. S. Jacob, *Tetrahedron: Asymmetry* **1993**, *4*, 2011 – 2024.
[16] D. S. Alonzi, N. V. Kukushkin, S. A. Allman, Z. Hakki, S. J. Williams, L. Pierce, R. A. Dwek, T. D. Butters, *Cell. Mol. Life Sci.* **2013**, *70*, 2799 – 2814.
[17] M. Petricevic, L. F. Sobala, P. Fernandes, L. Raich, A. J. Thompson, G. Bernardo-Seisdedos, O. Millet, S. Zhu, M. Sollogoub, J. Jimenez-Barbero, C. Rovira, G. J. Davies, S. J. Williams, *J. Am. Chem. Soc.* **2017**, *139*, 1089 – 1097.
[18] U. Spohr, M. Bach, R. G. Spiro, *Can. J. Chem.* **1993**, *71*, 1919 – 1927.
[19] T. Aoyagi, H. Suda, K. Uotani, F. Kojima, T. Aoyama, K. Horiguchi, M. Hamada, T. Takeuchi, *J. Antibiot.* **1992**, *45*, 1404 – 1408.
[20] T. D. Heightman, A. T. Vasella, *Angew. Chem. Int. Ed.* **1999**, *38*, 750 – 770; *Angew. Chem.* **1999**, *111*, 794 – 815.
[21] I. K. Khanna, F. J. Koszyk, M. A. Stealey, R. M. Weier, J. Julien, R. A. Mueller, S. N. Rao, L. Swenton, D. P. Getman, G. A. DeCrescenzo, R. M. Heintz, *J. Carbohydr. Chem.* **1995**, *14*, 843 – 878.
[22] M. Hoch, E. Heinz, R. R. Schmidt, *Carbohydr. Res.* **1989**, *191*, 21 – 28.
[23] T. Oshitari, M. Shibasaki, T. Yoshizawa, M. Tomita, K.-i. Takao, S. Kobayashi, *Tetrahedron* **1997**, *53*, 10993 – 11006.
[24] J. D. Albright, L. Goldman, *J. Am. Chem. Soc.* **1967**, *89*, 2416 – 2423.
[25] H. S. Overkleeft, J. van Wiltenburg, U. K. Pandit, *Tetrahedron* **1994**, *50*, 4215 – 4224.
[26] T. Granier, N. Panday, A. Vasella, *Helv. Chim. Acta* **1997**, *80*, 979 – 987.
[27] G. J. Davies, K. S. Wilson, B. Henrissat, *Biochem. J.* **1997**, *321*, 557 – 559.
[28] D. L. Zechel, A. B. Boraston, T. Gloster, C. M. Boraston, J. M. Macdonald, D. M. G. Tilbrook, R. V. Stick, G. J. Davies, *J. Am. Chem. Soc.* **2003**, *47*, 14313 – 14323.
[29] J. Clark, D. D. Perrin, *Q. Rev.* **1964**, *18*, 295 – 320.
[30] S. Inouye, *Chem. Pharm. Bull.* **1968**, *16*, 1134 – 1137.
[31] A. R. Mandhapati, D. Shcherbakov, S. Duscha, A. Vasella, E. C. Böttger, D. Crich, *ChemMedChem* **2014**, *9*, 2074 – 2083.
[32] J. Agirre, J. Iglesias-Fernandez, C. Rovira, G. J. Davies, K. S. Wilson, K. D. Cowtan, *Nat. Struct. Mol. Biol.* **2015**, *22*, 833 – 834.
[33] A. Varrot, M. Schülein, M. Pipelier, A. Vasella, G. J. Davies, *J. Am. Chem. Soc.* **1999**, *121*, 2621 – 2622.
[34] W. Nerinckx, T. Desmet, K. Piens, M. Claeyssens, *FEBS Lett.* **2005**, *579*, 302 – 312.
[35] 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, J. R. Cairns, *ACS Chem. Biol.* **2016**, *11*, 1891 – 1900.
[36] 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.
[37] S. Gore, E. Sanz Garcia, P. M. S. Hendrickx, A. Gutmanas, J. D. Westbrook, H. Yang, Z. Feng, K. Baskaran, J. M. Berrisford, B. P. Hudson, Y. Ikegawa, N. Kobayashi, C. L. Lawson, S. Mading, L. Mak, A. Mukhopadhyay, T. J. Oldfield, A. Patwardhan, E. Peisach, G. Sahni, M. R. Sekharan, S. Sen, C. Shao, O. S. Smart, E. L. Ulrich, R. Yamashita, M. Quesada, J. Y. Young, H. Nakamura, J. L. Markley, H. M. Berman, S. K. Burley, S. Velankar, G. J. Kleywegt, *Structure* 2017, 25, 1916-1927. doi: 10.1016/j.str.2017.10.009.
[38] W. C. Still, M. Kahn, A. M. Mitra, *J. Org. Chem.* **1978**, *43*, 2923 – 2925.
[39] A. B. Pangborn, M. A. Giardello, R. H. Grubbs, R. K. Rosen, F. J. Timmers, *Organometallics* **1996**, *15*, 1518 – 1520.
[40] C. Ouairy, T. Cresteil, B. Delpech, D. Crich, *Carbohydr. Res.* **2013**, *377*, 35 – 43.
[41] W. B. Turnbull, A. H. Daranas, *J. Am. Chem. Soc.* **2003**, *125*, 14859 – 14866.
[42] D. G. Waterman, G. Winter, R. J. Gildea, J. M. Parkhurst, A. S. Brewster, N. K. Sauter, G. Evans, *Acta Crystallogr. Sect. D* **2016**, *72*, 558 – 575.
[43] G. Winter, *J. Appl. Crystallogr.* **2010**, *43*, 186 – 190.
[44] P. R. Evans, G. N. Murshudov, *Acta Crystallogr. Sect. D* **2013**, *69*, 1204 – 1214.
[45] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. Sect. D* **2011**, *67*, 355 – 367.
[46] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486 – 501.
[47] I. Tickle, *Acta Crystallogr. Sect. D* **2012**, *68*, 454 – 467.
[48] L. Potterton, J. Agirre, C. Ballard, K. Cowtan, E. Dodson, P. R. Evans, H. T. Jenkins, R. Keegan, E. Krissinel, K. Stevenson, A. Lebedev, S. J. McNicholas, R. A. Nicholls, M. Noble, N. S. Pannu, C. Roth, G. Sheldrick, P. Skubak, V. Uski, F. von Delft, D. Waterman, K. Wilson, M. Winn, M. Wojdyr, *Acta Crystallographica Section D: Structural Biology* **2018**, *74*, 68-84.

Manuscript received: January 28, 2018
Accepted manuscript online: March 5, 2018
Version of record online: April 30, 2018

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Contribution of Shape and Charge to the Inhibition of a Family GH99 *endo*-α-1,2-Mannanase

Marija Petricevic,[†] Lukasz F. Sobala,[‡] Pearl Z. Fernandes,[†] Lluís Raich,[⊥] Andrew J. Thompson,[‡] Ganeko Bernardo-Seisdedos,[∥] Oscar Millet,[∥] Sha Zhu,[§] Matthieu Sollogoub,[§] Jesús Jiménez-Barbero,[∥,▽] Carme Rovira,[*,⊥,#] Gideon J. Davies,[*,‡] and Spencer J. Williams[*,†]

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

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

[§]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Institut Parisien de Chimie Moléculaire (IPCM), UMR 8232, 4, place Jussieu, 75005 Paris, France

[∥]Molecular Recognition and Host−Pathogen Interactions, CIC bioGUNE, Bizkaia Technology Park, Building 800, 48160 Derio, Spain

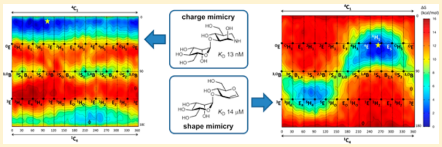
[⊥]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

[#]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain

[▽]Ikerbasque, Basque Foundation for Science, María Díaz de Haro 3, 48013 Bilbao, Spain

Supporting Information

ABSTRACT: Inhibitor design incorporating features of the reaction coordinate and transition-state structure has emerged as a powerful approach for the development of enzyme inhibitors. Such inhibitors find use as mechanistic probes, chemical biology tools, and therapeutics. *Endo*-α-1,2-mannosidases and *endo*-α-1,2-mannanases, members of glycoside hydrolase family 99 (GH99), are interesting targets for inhibitor development as they play key roles in N-glycan maturation and microbiotal yeast mannan degradation, respectively. These enzymes are proposed to act via a 1,2-anhydrosugar "epoxide" mechanism that proceeds through an unusual conformational itinerary. Here, we explore how shape and charge contribute to binding of diverse inhibitors of these enzymes. We report the synthesis of neutral dideoxy, glucal and cyclohexenyl disaccharide inhibitors, their binding to GH99 *endo*-α-1,2-mannanases, and their structural analysis by X-ray crystallography. Quantum mechanical calculations of the free energy landscapes reveal how the neutral inhibitors provide shape but not charge mimicry of the proposed intermediate and transition state structures. Building upon the knowledge of shape and charge contributions to inhibition of family GH99 enzymes, we design and synthesize α-Man-1,3-noeuromycin, which is revealed to be the most potent inhibitor (K_D 13 nM for *Bacteroides xylanisolvens* GH99 enzyme) of these enzymes yet reported. This work reveals how shape and charge mimicry of transition state features can enable the rational design of potent inhibitors.



INTRODUCTION

Over 500 000 gene sequences have been discovered encoding glycoside hydrolases that are grouped into more than 130 families according to the Carbohydrate Active enZyme classification (CAZy: www.cazy.org).[1] Glycoside hydrolases of family 99 possess two distinct activities: *endo*-α-1,2-mannosidase and *endo*-α-1,2-mannanase. *endo*-α-1,2-Mannosidases are eukaryotic proteins involved in N-linked glycan maturation, folding, and quality control[2−5] and are of clinical significance as they provide a means for viruses and cancer to evade the effect of *exo*-glucosidase inhibitors.[5,6] *endo*-α-1,2-Mannanases are produced by *Bacteroides* spp., bacterial residents of the gut microbiota.[7] They facilitate the degradation of dietary yeast mannan consumed in bread and fermented foods, facilitating the breakdown of these complex carbohydrates, with beneficial effects on the gastrointestinal tract and, possibly, mitigating the symptoms of Crohn's disease.[8] Given the importance of family GH99 enzymes in N-linked glycan maturation and carbohydrate processing by the microbiota, the development of inhibitors has been of particular importance to allow assessment and manipulation of their roles in these complex processes. In this work, we investigate several mechanism-inspired inhibitor design concepts for family GH99 *endo*-α-1,2-mannanases from the gut microbiota constituents *Bacteroides thetaiotaomicron* and *Bacteroides xylanisolvens*; *BtGH99* and *BxGH99*, respectively. Our results cast light on the importance of structural mimicry of shape

Received: September 25, 2016
Published: December 19, 2016

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

and charge of species along the reaction coordinate for achieving potent inhibition of this enzyme family.

Glycosidases that cleave their substrates with retention of anomeric configuration typically operate through a two-step mechanism that proceeds via a covalent glycosyl-enzyme intermediate. Such enzymes utilize enzymatic amino acid residues that in the first step act as general acid and nucleophile to assist in departure of the anomeric substituent while simultaneously substituting the anomeric group; in the second step the first carboxylate acts as a general base to deprotonate a nucleophilic water molecule that hydrolyzes the covalent glycosyl enzyme intermediate (Figure 1A).[9,10] Important

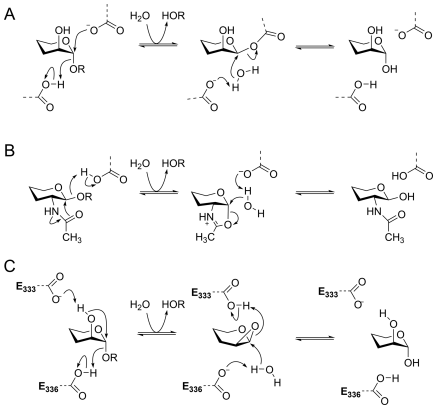


Figure 1. (A) Mechanism for a canonical retaining α-glycosidase that proceeds through a covalent glycosyl-enzyme intermediate. (B) Mechanism for a retaining β-hexosaminidase involving neighboring group participation by the 2-acetamido group, via an oxazolinium ion intermediate. (C) Proposed mechanism for family GH99 α-mannosidases involving neighboring group participation by the 2-OH group, via a 1,2-anhydro sugar (epoxide). Numbering is for *Bx*GH99.

exceptions include a range of β-hexosaminidases that perform catalysis through mechanisms involving neighboring group participation by the 2-acetamido group of the substrate (Figure 1B).[11] These enzymes also operate through a two-step mechanism: in the first step an amino acid residue acts as a general acid to assist in departure of the leaving group while the 2-acetamido group performs a nucleophilic attack on the anomeric center, forming a bicyclic oxazoline/oxazolinium ion intermediate. In the second step the same amino acid residue acts as a general base, assisting nucleophilic attack by a water molecule that opens the oxazolinium ion ring, reforming the 2-acetamido group and completing the hydrolysis reaction.

Family GH99 enzymes cleave glycosides with an overall retention of anomeric configuration.[12] X-ray structures are available for GH99 enzymes in complex with a variety of ligands based on sugar-shaped heterocycles.[7,8,12] However, in X-ray structures of *B. xylanisolvens Bx*GH99 with various substrate-mimicking ligands, it was not possible to identify an appropriately positioned enzymatic nucleophile within the typical <3 Å distance from the reactive anomeric center, leading to the proposal of a nonclassical mechanism.[12] In particular, in structures of GlcIFG (**2**) and ManIFG (**3**) with *Bx*GH99, there were no close contacts with a likely candidate enzymatic nucleophile,[7,12] at odds with the usual observation of a carboxylate situated typically 2.6−2.7 Å away in classical retaining glycosidases. Moreover, in complexes of GlcDMJ (**1**) with *Bx*GH99 a conserved carboxyl residue (E333; numbering refers to *Bx*GH99) was located 2.7 Å from the 2-OH group;[12] similar observations extend to the binding of a substrate (α-Man-1,3-α-ManMU) to the carboxamide mutant *Bx*GH99 E333Q.[7] Collectively these data supported the proposal of a two-step reaction involving in the first step the formation of a bicyclic 1,2-anhydro sugar intermediate, through E333 acting as a general base residue to deprotonate the 2-OH and facilitating a nucleophilic substitution at C1 coincident with departure of the leaving group, assisted by E336 acting as general acid (Figure 1C).[12] In the second step of this proposed mechanism, E333 acts as a general acid, assisting ring opening of the epoxide, while E336 acts as a general base, promoting nucleophilic attack by a water molecule. While such a mechanism lacks precedent in enzymes, there is strong evidence that the base-catalyzed solvolysis of 4-nitrophenyl α-D-mannoside and α-mannosyl fluoride proceed through similar neighboring group participation mechanisms.[13−16] The most stable conformation of 1,2-anhydro-β-D-mannose is a $^4H_5$ half-chair; applying the principle of least nuclear motion, a $^4C_1 \rightarrow {}^4E^{\ddagger} \rightarrow {}^4H_5$ conformational itinerary has been proposed for the first step of the family GH99 reaction coordinate.[17]

Intensive efforts have been invested in the rational development of glycosidase inhibitors, and many fundamental principles have been articulated inspired by our deep mechanistic understanding of this class of enzyme. Based on pioneering insights from Pauling[18] and Wolfenden,[19] it is recognized that a common principle underpinning catalysis is the selective affinity of an enzyme for a reaction transition state, relative to the ground state. Accordingly, inhibitor design by transition state mimicry, which can take advantage of the high transition state affinity of a glycosidase, has proven a useful guiding strategy.[20] While it is widely appreciated that perfect transition state mimics are chemically unstable and thus unattainable, a general design principle is to develop analogues incorporating features that mimic the shape and charge of the transition state.[20] Three features have been highlighted for consideration in the development of effective glycosidase inhibitors: configuration, conformation and charge.[21] Configuration is the simplest to address and not surprisingly it is usually found that glycosidases are normally best targeted by inhibitors with stereochemistry matching that of the substrate. In the case of *Bx*GH99, an enzyme that has the ability to cleave both α-Glc-1,3-α-Man-OR and α-Man-1,3-α-Man-OR configured substrates (with a preference for the latter),[7] optimal inhibition is achieved by inhibitors matching the preferred substrate configuration. Glycosidases typically operate through transition states with substantial oxocarbenium ion character, and partial double bond development between O5 and C1, leading to a flattened conformation at the transition state. Consequently, mimicry of the flattened conformation expected at the transition state has proven a second effective strategy, with inhibitors bearing sp^2-hybridized atoms at the anomeric or endocyclic oxygen positions, such as glyconolactones and -lactams, identified as fairly broad spectrum glycosidase inhibitors. Finally, partial charge development at C1 and the endocyclic oxygen at the transition state can be mimicked by the protonated forms of nitrogen-containing heterocycles, exemplified by deoxymannojirimycin (DMJ) with

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

a nitrogen in place of the endocyclic oxygen, and isofagomine with a nitrogen in place of C1.

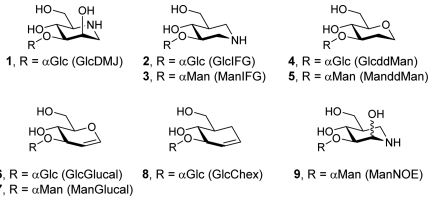
Motivated by the unusual mechanism proposed for family GH99 enzymes and the complexity of their biochemical roles, our understanding of which should benefit from the development of potent inhibitors, we undertook a study of several inhibitor designs inspired by the concepts of charge and shape mimicry. GlcDMJ (**1**), an inhibitor reported by Spiro and co-workers,[22] was the first effective inhibitor of mammalian *endo*-α-1,2-mannosidase,[6] and was subsequently shown to bind to and inhibit *Bx*GH99 and *Bt*GH99 (Table 1).[12] More potent

Table 1. Dissociation Constants for GH99 *endo*-Mannanase Inhibitors

	K_D values (μM)		
compd	*Bt*GH99	*Bx*GH99	method
1 (GlcDMJ)	24	ND[a]	ITC[12]
2 (GlcIFG)	0.63	ND[a]	ITC[12]
3 (ManIFG)	0.14	0.27	ITC[6]
5 (ManddMan)	53 ± 5	221 ± 11	NMR
7 (ManGlucal)	15 ± 1.9	111 ± 11	NMR
8 (GlcChex)	no binding	no binding	NMR/ITC
9 (ManNOE)	0.03 ± 0.01	0.013 ± 0.002	ITC

[a]ND, not determined.

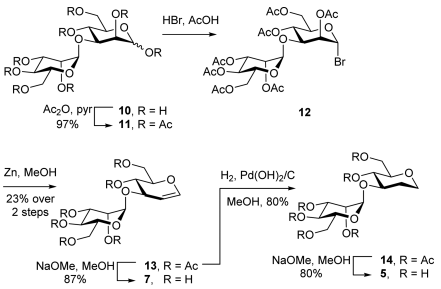
inhibition was achieved by GlcIFG (**2**),[12] which also proved to be a more effective inhibitor than GlcDMJ of mammalian *endo*-α-1,2-mannosidase in cell-based studies,[23] demonstrating that varying the position of charge can provide improvements in potency. Further, upon identification of the *Bacteroides* spp. enzymes as preferential *endo*-α-1,2-mannanases,[7,8] we were able to configurationally match the substrate and develop the inhibitor ManIFG (**3**), the most potent inhibitor yet reported for any GH99 enzyme.[7] However, ManIFG and GlcIFG lack the 2-OH group of the substrate and thus cannot benefit from specific interactions with the putative acid/base E333. Separately, Spiro and co-workers reported that two other neutral compounds were almost as effective as GlcDMJ in the inhibition of mammalian *endo*-α-1,2-mannosidase, namely, GlcddMan (**4**) and GlcGlucal (**6**).[6,24] We were intrigued by these observations and sought to investigate whether the equivalent configurationally matched species, ManddMan (**5**) and ManGlucal (**7**), and the related cyclohexene derivative (**8**) were inhibitors of bacterial GH99 enzymes and to understand how they bind to the enzyme.



RESULTS AND DISCUSSION

Synthesis of ManGlucal and ManddMan. ManddMan (**5**) and ManGlucal (**7**) were prepared from α-1,3-mannobiose (**10**) (Scheme 1). Acetylation, followed by bromination afforded mannobiosyl bromide (**12**), which was converted to the

Scheme 1. Synthesis of ManddMan (5) and ManGlucal (7)



protected glucal (**13**) using Zn in MeOH. Zemplén transesterification afforded ManGlucal (**7**). Alternatively, reduction of **13** using H_2/Pd−C, followed by Zemplén transesterification afforded ManddMan (**5**). The preparation of GlcChex (**8**) will be described elsewhere.

Binding and 3-D Structural Analyses of Putative "Shape Mimics". Dissociation constants for the binding of compounds **5** and **7** to *Bt*GH99 and *Bx*GH99 were determined by NMR spectroscopy (Figure S1). The 2D NMR SOFAST-HMQC spectra of ^{15}N-labeled enzymes determined in the presence or absence of a saturating amount of the ligands revealed several H−N peaks that displayed significant chemical shift perturbations. For instance, new signals for an arginine residue (assigned as R295 in *Bx*GH99 and R291 in *Bt*GH99 on the basis of analysis of inter-residue nOes from the 3D-HSQC-NOESY spectra; see annotation to Figure 4A) appeared during the titration experiments, which were in slow exchange with the initial ones in the chemical shift time scale (Figure 2). Therefore, since the relative intensities of these signals are proportional to the populations of the bound and unbound forms (see Experimental Section and Supporting Information (SI)), the dissociation constants (K_D) were readily calculated. The binding constants are shown in Table 1. No evidence for binding of GlcChex could be obtained by either NMR or ITC.

In order to analyze the mode of binding of the conformationally restricted compounds, 3-D structures of complexes of *Bx*GH99 with **7** and **8** were determined by X-ray crystallography at near atomic (approximately 1.0 Å) resolutions (Table 2, Figure 3A,B). ManGlucal **7** (K_D 111 μM) binds to *Bx*GH99 in the −2 and −1 subsites, with the −1 glucal ring intact in a $^4H_5$ conformation. For reasons most likely related to its poor affinity for the enzyme, we were unable to obtain a complex of GlcChex with wildtype enzyme, but were serendipitously successful in obtaining a complex with the catalytically inactive *Bx*GH99 E333Q mutant. In this complex GlcChex **8** also bound in the −2/−1 subsite with the cyclohexene ring in a $^4H_5$ conformation (Figure 3B). Relative to ManGlucal, GlcChex suffers both by replacement of the endocyclic oxygen with methylene and by the presence of a nonreducing-end glucosyl moiety, the latter of which is known to reduce binding to the *Bacteroides* spp. enzymes by 4−10-fold.[7] Owing to the unmeasurable binding of GlcChex, the synthesis of the mannose analogue was not pursued. The lack of oxygen atoms within the GlcChex ring means it cannot form hydrogen bonds with active site residues Y252, E333 or E336. MS experiments indicated that the compound is not affected by

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

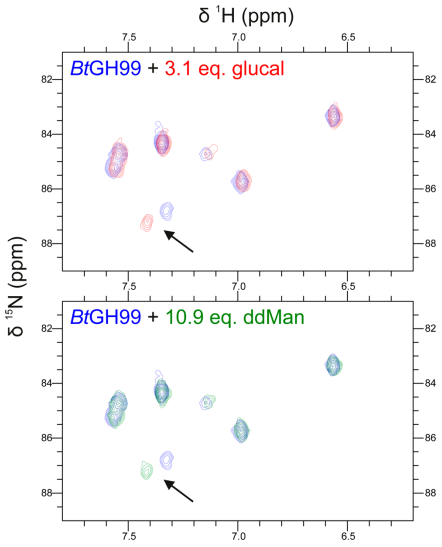


Figure 2. Excerpt of the side chain HN ε-arginine region in the 2D SOFAST-HMQC NMR spectra of ^{15}N-labeled *Bt*GH99 in the absence (blue) or presence of an excess (top) of ManGlucal (**7**; red) or (bottom) ManddMan (**5**; green). The arrow highlights the chemical shift perturbation observed for the HN-ε signal corresponding to R295.

the enzyme or its variants so the reason for binding to only the inactive variant is unclear.

Glycals are often effective inhibitors of classical retaining glycosidases. Inhibition is typically found to be time-dependent, owing to a chemical reaction in which the conjugate acid of the nucleophile protonates the enol ether of the glycal, and forms a 2-deoxy-glycosyl enzyme.[25,26] This mode of reactivity has been exploited to allow the identification of the catalytic nucleophile by peptide sequencing,[25] and the 2-deoxyglycosyl-enzymes are sufficiently stable to be studied by X-ray crystallography.[27] To date, only two classes of retaining glycosidases have been identified upon which glycals bind as competitive inhibitors without exhibiting this mode of reactivity. These are *N*-acetylhexosaminidases that use neighboring group participation and retaining sialidases, and in both cases these enzymes lack a typical carboxylate nucleophile. Retaining *N*-acetylhexosaminidases use a 2-acetamido group, which is able to catalyze the slow hydration of the enol ether.[30] Retaining sialidases lack a carboxylate nucleophile capable of protonating the glycal, and instead use a less acidic tyrosine residue as the catalytic nucleophile.[31] In the case of sialidases, the corresponding glycal 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid has been elaborated into the extraordinarily potent inhibitor zanamivir,[32] a clinically used antiinfluenza drug. Despite its potency, quantitative examination of transition state mimicry by zanamivir reveals it to be a poor transition state mimic.[33] The observation that ManGlucal binds to *Bx*GH99 without a chemical reaction despite the retaining mechanism of the enzyme provides further evidence in favor of the unique neighboring group participation mechanism proposed for this family.

Binding and 3-D Structural Analyses of ManddMan: Probing the Role of O2 Interactions. In order to harness the tighter binding of the bacterial enzymes with a mannoside in the −2 subsite[7] ManddMan (**5**) was synthesized (Scheme 1). NMR titration revealed ManddMan to bind to *Bx*GH99 and *Bt*GH99 with K_D values of 221 and 53 μM, respectively (Table 1). By comparison, GlcddMan (**4**) is an inhibitor for rat endomannosidase[6] with an IC_{50} value of 3.8 μM for inhibition of cleavage of ^{14}C-labeled $Glc_1Man_9GlcNAc$, only slightly worse than that of GlcDMJ (**1**) (IC_{50} = 1.7 μM). The structure of *Bx*GH99 in complex with ManddMan was determined at 1.03 Å resolution. Compound **5** binds in the −2/−1 subsites of the enzyme in an undistorted $^4C_1$ conformation. This conformation matches that of the ground state of the substrate, and while this is consistent with the modest dissociation constant, it is noteworthy that the tight-binding inhibitor ManIFG **3** also binds in a $^4C_1$ chair.

Design, Synthesis, and Characterization of ManNOE as a GH99 Inhibitor. As ManIFG (**3**), ManddMan (**5**), and ManGlucal (**7**) all lack an O2 group, in considering the contribution of shape and charge to inhibition, it would appear that the cationic nature of ManIFG contributes most significantly to inhibition, in spite of its conformational resemblance to the ground state. We therefore decided that it would be appropriate to investigate a charged inhibitor based on ManIFG that was able to make the correct O2 interactions, in particular with E333. Inspired by the work of Bols and co-workers on the development of noeuromycin, a 2-hydroxy analogue of isofagomine that binds 2−4000 times more tightly than isofagomine to various glycosidases,[34] we therefore synthesized the noeuromycin derivative, ManNOE. This inhibitor was synthesized by the regioselective mannosylation of the nitrile diol **16**[35] by trichloroacetimidate **15** Scheme 2.[36] The sole acetate group of the glycoside **17** was cleaved by treatment with HCl/MeOH to afford alcohol **18**, and the nitrile group was reduced using BH_3·Me_2S, followed by protection as the Boc derivative **19**. Hydrogenolysis of the benzyl ethers of **19** using H_2/Pd−C, and then cleavage of the Boc group with HCl, afforded ManNOE.HCl (**9**) [as a mixture of α-hydroxypiperidine and pyranose isomers (not drawn); see the SI].

Table 2. X-ray Data and Structure Summary

	5 (ManddMan)	**7** (ManGlucal)	**8** (GlcChex)	**10** (ManNOE)	**10** + M_2 (ManNOE + 1,2-α-mannobiose)
resolution (outer shell) (Å)	76.77−1.03 (1.05−1.03)	76.85−1.07 (1.09−1.07)	39.54−1.2 (1.22−1.2)	76.73−1.14 (1.16−1.14)	57.21−1.05 (1.07−1.05)
R_{merge} (outer)	0.052 (0.989)	0.052 (1.748)	0.059 (0.955)	0.051 (1.158)	0.054 (1.314)
R_{cryst}/R_{free}	0.117/0.130	0.124/0.141	0.119/0.137	0.124/0.143	0.115/0.133
rmsd bonds (Å)	0.010	0.011	0.012	0.011	0.013
rmsd angles (deg)	1.53	1.54	1.59	1.50	1.67
PDB code	5M17	5M5D	5MEL	5LYR	5M03

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

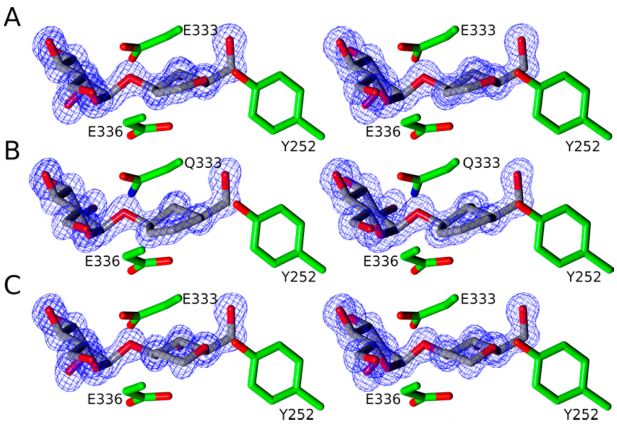
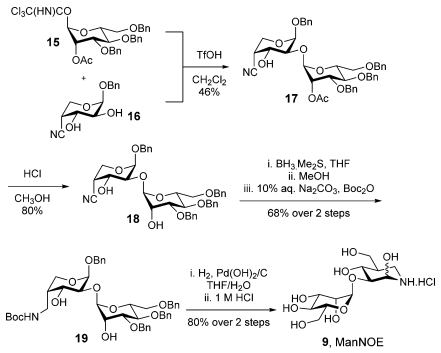


Figure 3. Stereoview of active site of *B. xylanisolvens* family GH99 enzyme complexes. (A) *Bx*GH99 with ManGlucal (**7**), (B) *Bx*GH99-E333Q with GlcChex (**8**), and (C) *Bx*GH99 with ManddMan (**5**). Depicted electron density maps are REFMAC[28] maximum-likelihood/σA weighted $2F_o - F_c$ syntheses contoured at 1.5σ (0.57, 0.59, and 0.62 eÅ^{-3}, respectively). All panels were assembled using CCP4mg.[29]

Scheme 2. Synthesis of ManNOE (9)



Isothermal titration calorimetry revealed ManNOE (**9**) to bind to *Bx*GH99 with K_D = 13 nM, and to *Bt*GH99 with a K_D = 30 nM, 17- and 5-fold more tightly than ManIFG (**3**) to the respective enzymes, commensurate with improvements seen for binding of IFG versus NOE for other enzymes,[34,37] and demonstrating that better matching of the substrate by reinstatement of the 2-OH group absent in the latter compound provides more effective inhibition (Table 1, Figure S4). 3-D structures were solved of a binary complex of *Bx*GH99−ManNOE, and a ternary complex of *Bx*GH99−ManNOE−α-1,2-mannobiose at resolutions of 1.14 and 1.05 Å, respectively (Figure 4). The poses of ManNOE in both complexes were essentially identical and the more informative ternary complex, with ManNOE in the −2/−1 subsites, and α-1,2-mannobiose in the +1/+2 subsites will therefore be discussed. The NOE heterocycle binds in a $^4C_1$ conformation, similar to that seen for ManIFG with the same enzyme. A close contact with E333 Oδ··· O2 of 2.58 Å is evident, similar to that seen in the complex of *Bx*GH99 with GlcDMJ (2.54 Å, PDB 4AD3).

Conformational Analyses of Glucal, Chex, ddMan, and NOE. In order to understand the intrinsic conformational preferences of the D-glucal, 1,2-dideoxymannose (ddMan), Chex, and noeuromycin (NOE) inhibitor warheads, so as to help

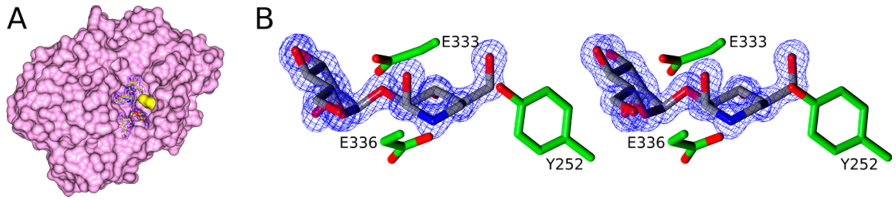


Figure 4. (A) X-ray structure of ternary complex of *Bx*GH99 with ManNOE (above) and α-1,2-mannobiose (below). The residue used for NMR titrations, R295, is shown in yellow. $2F_o - F_c$ map contoured at 1.0 σ (0.42 eÅ^{-3}). The +1 subsite mannose residue electron density is best modeled by two mannose conformations with 0.6/0.4 occupancy, rotated by about 30° with respect to +2 mannose. (B) Stereoview of ManNOE in the active site. $2F_o - F_c$ map contoured at 1.5 σ (0.58 eÅ^{-3}). Assembled using CCP4mg.[29]

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

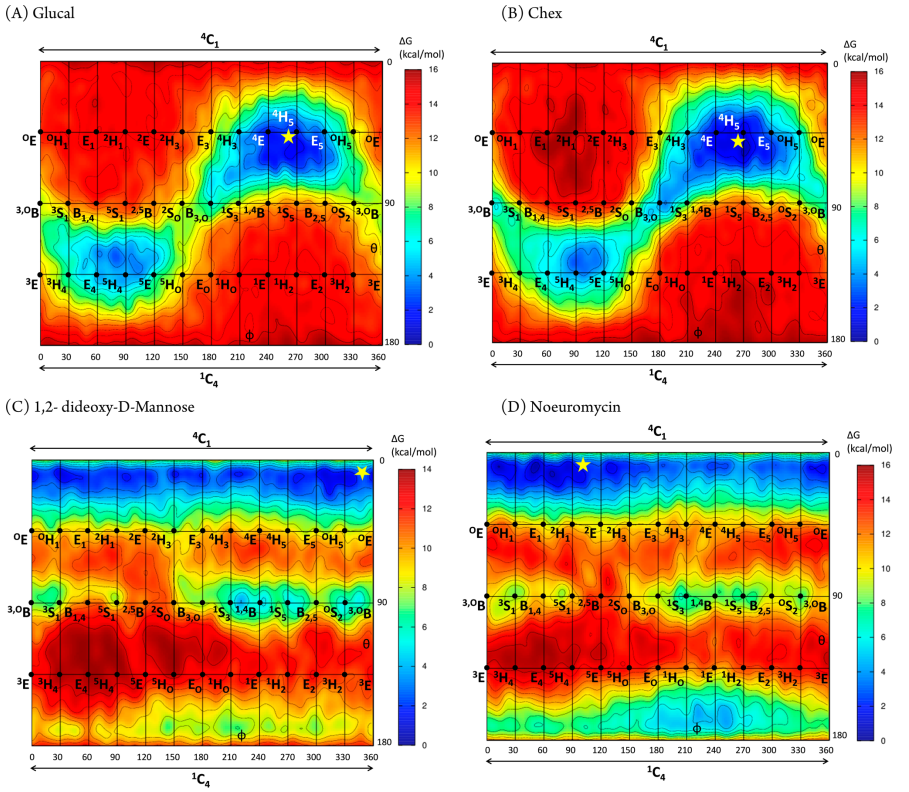


Figure 5. Conformational free-energy landscapes (FELs, Mercator projection) of isolated D-glucal (A), Chex (B), ddMan (C), and NOE (D), contoured at 1 kcal mol^{-1}. FELs have been annotated (yellow star) with the inhibitor conformations of ManGlucal **7** (for A), GlcChex **8** (for B), ManddMan **5** (for C), and ManNOE **9** (for D) that have been observed on-enzyme in this work.

rationalize the conformations observed on-enzyme, we calculated their conformational free energy landscapes (FELs). FELs were computed by ab initio metadynamics (see Experimental Section), and the Cremer−Pople puckering coordinates θ and ϕ were used as collective variables, yielding a Mercator representation[38] for each inhibitor FEL (Figure 5). The same procedure has been previously used to analyze the conformational preferences of related GH inhibitors (mannoimidazole, glucoimidazole, and IFG).[39,40]

The FELs of both D-glucal and Chex (Figure 5A,B) exhibit a main energy minimum centered at $^4H_5$. The free energy well is quite large and extends through ^4E−$^4H_5$−E_5, with an "arm" toward $B_{3,O}$, indicating substantial conformational freedom around $^4H_5$. There is also a local minimum in the southern hemisphere, centered at $^5H_4$. However, this is 3 kcal mol^{-1} higher in energy and thus less populated at room temperature. Interconversion between the two minima is hindered by an 8 kcal mol^{-1} energy barrier. Therefore, both D-glucal and Chex display only one main accessible conformation at room temperature, $^4H_5$, but can readily adopt the nearby ^4E conformation predicted for the transition states leading to and from the proposed 1,2-anhydrosugar intermediate. These data suggest that D-glucal and Chex provide good shape mimicry of the transition state or intermediate for the GH99 catalyzed reaction.

That both D-glucal and Chex adopt a $^4H_5$ conformation when complexed to *Bx*GH99 (Table 2 and yellow star in Figure 5A,B) indicates that the conformational preference of the isolated molecules are not significantly perturbed on-enzyme. This is also the case for ddMan, for which the FEL (Figure 5C) is strongly biased toward the $^4C_1$ chair (the local minima on the equator are ≈8 kcal/mol higher in energy), as observed in the X-ray structure on-enzyme. However, the FEL of ddMan does not exhibit any stable minimum around $^4H_5$, thus it cannot be considered a GH99 transition state shape mimic, and should instead be considered a mimic of the substrate conformation in the Michaelis complex.

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

NOE differs from the other inhibitors considered here as it is a basic molecule and was therefore considered as both the neutral (Figure 5D) and protonated species (Figure S5). While the topographies of the FELs for the two protonation states are broadly similar, they differ greatly in relative energies, most importantly for the global and local minima, such that the most stable state in NOE, a $^4C_1$ chair, becomes the "ring-flipped" $^1C_4$ conformer in protonated NOE. In that case, the two most stable species, the $^1C_4$ and $^1S_5$ states, are characterized by the presence of transannular hydrogen bonds between NH_2^+ and O6 or O3, respectively. What then is the most appropriate FEL to consider in relation to the enzyme-bound state, which is expected to be protonated, but which in a $^4C_1$ conformation lacks a transannular hydrogen bond? It is known that in the absence of solvation (in the gas phase) flexible molecules in low charge states tend to compensate charge effects by forming stabilizing intramolecular interactions that do not take place in other environments (e.g., in solution). For example, low-charge state proteins in the gas phase tend to adopt more compact structures owing to increased intramolecular interactions.[41] On balance, we consider the FEL of neutral NOE to be a more relevant representation of the enzyme-bound conformations, because interactions of the inhibitor with active site residues prevent the formation of intramolecular hydrogen bonds that dominate the conformations of charged species. Accordingly, the FEL displays a wide main minimum situated close to $^4C_1$ (Figure 5D), and is thus similar to that of ddMan. The NOE FEL is also reminiscent of the one previously computed for the closely related neutral IFG inhibitor[39] (which differs from NOE by the absence of the 2-hydroxyl group). Interestingly, the transition state region between the north pole and the equator in NOE ($^4H_3/^4E$) is shifted by 60° in ϕ in ddMan. This is likely due to a vicinal intramolecular hydrogen bond formed between the 2-OH and the 3-OH, which stabilizes the $^4H_3/^4E$ conformations in NOE; this interaction is not present in ddMan or IFG as they both lack a 2-OH. Overall, NOE most closely resembles the conformation of the substrate in the Michaelis complex and in its protonated state on-enzyme provides mimicry of an oxocarbenium-ion-like transition state; it is therefore best considered a "charge"-mimicking inhibitor with no significant shape mimicry of the transition state.

CONCLUSIONS

The proposed GH family 99 neighboring group participation mechanism, via a 1,2-anhydro sugar, allows prediction of a $^4C_1 \rightarrow {}^4E^{\ddagger} \rightarrow {}^4H_5$ conformational itinerary for the first step of the reaction coordinate. The FELs for Glucal and Chex suggest that when these inhibitor warheads are extended to ManGlucal and GlcChex, their flattened conformations provide mimicry of the ^4E transition state and $^4H_5$ intermediate conformations. Their nonbasic nature provides shape but not charge mimicry of the transition state. X-ray structures of these compounds in complex with *Bx*GH99 revealed them to bind in a $^4H_5$ conformation, most closely matching the proposed intermediate conformation. The modest dissociation constant of ManGlucal, and lack of detectable binding for GlcChex, suggests that shape mimicry of the intermediate or transition state provides only weak affinity for the enzyme. On the other hand the FEL for the neutral sugar ddMan reveals a preference for a $^4C_1$ conformation, and the corresponding complex of ManddMan with *Bx*GH99 revealed this compound to bind in the same conformation, albeit with an affinity relative to ManGlucal 8- or 18-fold worse for *Bx* or *Bt*GH99 enzymes, respectively. As ManGlucal, GlcChex and ManddMan all lack a 2-OH group, these ratios provide an estimate for the contribution of shape-mimicry of the TS or intermediate to enzyme binding.

Based on the above analysis, and combined with the previous discovery that the best inhibitor for GH99 *endo*-α-mannanases is ManIFG,[7] which provides charge mimicry, but poor transition state shape mimicry, we were inspired to reinstate the 2-hydroxyl group missing in this compound. ManNOE was synthesized and shown to bind to *Bx* and *Bt*GH99 with K_D values of 30 and 13 nM, the most tightly binding ligand for these enzymes yet reported, and a 20-fold enhancement of affinity relative to ManIFG. The X-ray structure of ManNOE in complex with BxGH99 reveals a $^4C_1$ ground-state conformation mimicking the Michaelis complex, and we conclude that this inhibitor acts primarily to mimic the charge of the transition state.

EXPERIMENTAL SECTION

X-ray Crystallography. Data reduction was performed using xia2[42] or XDS.[43] Data set HKL index was matched to a previous solution using Aimless[44,45] software and FreeR set was generated from *Bx*GH99−ManddMan data, and then used for every other data set. Initial polypeptide chain model was obtained from the same previous solution, refined against *Bx*GH99−ManddMan data and used as starting model for other structure solutions. Refmac5[28] with ProSmart was used for restrained refinement and Coot[46] for real-space refinement. During model rebuilding the $F_{obs} - F_{calc}$ difference map was examined at 3σ. Validation was performed using Coot and edstats.[47] Sugar and pseudopyranose conformations and density correlation were validated by Privateer.[48,49]

Isothermal Titration Calorimetry. ManNOE binding to *Bt* and *Bx*GH99 was measured using a MicroCal AutoITC200 instrument (Malvern Instruments, formerly GE Healthcare) at 25 °C in 25 mM HEPES pH 7.0, 100 mM NaCl buffer. Protein concentration was kept at 5 μM and ligand concentration at 50 μM. As ManNOE exists as a 5:2 ratio of D-gluco and D-manno NOE isomers; the effective ManNOE concentration was adjusted accordingly.

2D NMR Titrations. Two-dimensional $^1H-^{15}N$ SOFAST-HMQC[50] spectra were recorded at 298 K for 1 h using ^{15}N-labeled *Bt*GH99 and *Bx*GH99 on a Bruker AVANCE III 800 MHz spectrometer with cryoprobe. Upon binding of ManGlucal or ManddMan, chemical shift perturbations were observed in slow exchange regime. A signal corresponding to Nε-Hε of R295 (numbering based on *Bx*GH99), and which is close to the enzyme active site, was chosen as binding reporter. In the case of *Bx*GH99, this signal shifts from δ 7.33, 87.5 ppm (^1H,^{15}N) in the free state to δ 7.41, 87.2 ppm (^1H,^{15}N) in the bound state. Bound and free protein populations at different protein/ligand ratios were calculated from peak intensities. NMR measurements were made in 50 mM potassium phosphate pH 7.0, 50 mM KCl with 5% D_2O added. Protein concentration was determined by measuring 280 nm UV absorbance after denaturing the solution in 6 M guanidinium chloride (average protein concentration: 58 μM). Ligand concentrations were cross verified by integrating the ^1H peaks against the internal standard TSP-d_4 (Sigma-Aldrich). The NMR data were processed and integrated using NMRPipe.[51] The dissociation constants (K_D values) were estimated using in-house Matlab 2015b scripts, using the following equation:

$$\frac{[PL]}{P_T} = \frac{[L]}{[L] + K_D}$$

where L is the free ligand concentration and $([PL]/P_T)$ is the ligand-bound protein fraction. Duplicate experimental points were used for error-bar estimation. Errors were propagated using a Monte Carlo algorithm to estimate the uncertainty in the K_D values. A distribution of K_D values (n = 10 000) were obtained from data sets randomly varying within the error bars, and the standard deviation of was used for the K_D error estimation.

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

Quantum Chemical Calculations. To obtain the conformational free energy landscapes of ddMan, Glucal, Chex, and NOE, quantum mechanical calculations were performed using density functional theory based molecular dynamics (MD), according to the Car−Parrinello (CP) method.[52] Each molecule was enclosed in an isolated cubic box of 12.0 Å × 12.0 Å × 12.0 Å. A fictitious electron mass of 700 au was used for the CP Lagrangian and a time step of 0.12 fs was used in all CPMD simulations. The Kohn−Sham orbitals were expanded in a plane wave (PW) basis set with a kinetic energy cutoff of 70 Ry. Ab initio pseudopotentials, generated within the Troullier−Martins scheme, were employed.[53] The Perdew, Burke, and Ernzerhoff generalized gradient-corrected approximation (PBE)[54] was selected in view of its good performance in previous work on isolated sugars,[55] glycosidases,[56] and glycosyltransferases.[57]

The metadynamics algorithm,[58] provided by the Plumed 2 plugin,[59] was used to explore the conformational free energy landscape of the systems, taking as collective variables θ and φ of the puckering coordinates of Cremer and Pople,[60] in the spirit of the pioneering work by Dowd, French, and Reilly.[61] Initially, the height of these Gaussian terms was set at 0.6 kcal mol^{-1} and a new Gaussian-like potential was added every 250 MD steps. Once the whole free energy space was 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 width of the Gaussian terms was set to 0.10 Å. The simulations were stopped when energy differences among wells remain constant, which was further confirmed by a time-independent free energy estimator.[62] For all molecules, the phase space was fully explored in less than 60 ps and the simulations were further extended up to 140 ps for Chex and Glucal, 160 ps for ddMan, and 240 ps for NOE. The errors in the principal minima, taken as a standard deviation (SD) from the last 60 ps, are below 0.6 kcal mol^{-1} (Figure S6).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b10075.

Experimental procedures, analytical data (^{1}H and ^{13}C NMR, MS) for all new compounds; detailed crystallographic experimental and data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*c.rovira@ub.edu
*gideon.davies@york.ac.uk
*sjwill@unimelb.edu.au

ORCID

Andrew J. Thompson: 0000-0001-7865-1856
Ganeko Bernardo-Seisdedos: 0000-0002-1372-3844
Carme Rovira: 0000-0003-1477-5010
Spencer J. Williams: 0000-0001-6341-4364

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Australian Research Council (FT130100103; DP120101396), the BBSRC (BB/G016127/1) and the ERC (ERC-2012-AdG-322942), the Spanish Ministry of Economy and Competitiveness (CTQ2014-55174, CTQ2015-64597-C2-1P, CTQ2015-68756-R) and the Generalitat de Catalunya (2014SGR-987). We thank Diamond Light Source for access to beamlines i02, i04, i04-1, and i24 (mx-9948) that contributed to the results presented here. We are grateful to Sivanandam Veeramuthu for his help with NMR. In-house crystal screening and testing was performed on X-ray equipment provided, in part, by the Wellcome Trust. The authors gratefully acknowledge the computer resources at *MareNostrum* and the technical support provided by BSC-CNS (RES-QCM-2016-3-00017). L.R. thanks the University of Barcelona for an APIF predoctoral fellowship.

■ REFERENCES

(1) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490−5.
(2) Lubas, W. A.; Spiro, R. G. *J. Biol. Chem.* **1987**, *262*, 3775−81.
(3) Lubas, W. A.; Spiro, R. G. *J. Biol. Chem.* **1988**, *263*, 3990−8.
(4) Moore, S. E.; Spiro, R. G. *J. Biol. Chem.* **1990**, *265*, 13104−12.
(5) Moore, S. E.; Spiro, R. G. *J. Biol. Chem.* **1992**, *267*, 8443−51.
(6) Hiraizumi, S.; Spohr, U.; Spiro, R. G. *J. Biol. Chem.* **1993**, *268*, 9927−35.
(7) Hakki, Z.; Thompson, A. J.; Bellmaine, S.; Speciale, G.; Davies, G. J.; Williams, S. J. *Chem. - Eur. J.* **2015**, *21*, 1966−77.
(8) Cuskin, F.; Lowe, E. C.; Temple, M. J.; Zhu, Y.; Cameron, E. A.; Pudlo, N. A.; Porter, N. T.; Urs, K.; Thompson, A. J.; Cartmell, A.; Rogowski, A.; Hamilton, B. S.; Chen, R.; Tolbert, T. J.; Piens, K.; Bracke, D.; Vervecken, W.; Hakki, Z.; Speciale, G.; Munoz-Munoz, J. L.; Day, A.; Pena, M. J.; McLean, R.; Suits, M. D.; Boraston, A. B.; Atherly, T.; Ziemer, C. J.; Williams, S. J.; Davies, G. J.; Abbott, D. W.; Martens, E. C.; Gilbert, H. J. *Nature* **2015**, *517*, 165−9.
(9) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171−202.
(10) Zechel, D. L.; Withers, S. G. *Acc. Chem. Res.* **2000**, *33*, 11−8.
(11) Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539−55.
(12) Thompson, A. J.; Williams, R. J.; Hakki, Z.; Alonzi, D. S.; Wennekes, T.; Gloster, T. M.; Songsrirote, K.; Thomas-Oates, J. E.; Wrodnigg, T. M.; Spreitz, J.; Stutz, A. E.; Butters, T. D.; Williams, S. J.; Davies, G. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 781−6.
(13) Gasman, R. C.; Johnson, D. C. *J. Org. Chem.* **1966**, *31*, 1830−8.
(14) Micheel, F.; Borrmann, D. *Chem. Ber.* **1960**, *93*, 1143−7.
(15) Kyosaka, S.; Murata, S.; Tanaka, M. *Chem. Pharm. Bull.* **1983**, *31*, 3902−5.
(16) Speciale, G.; Farren-Dai, M.; Shidmoossavee, F. S.; Williams, S. J.; Bennet, A. J. *J. Am. Chem. Soc.* **2016**, *138*, 14012−9.
(17) Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. *Curr. Opin. Struct. Biol.* **2014**, *28*, 1−13.
(18) Pauling, L. *Am. Sci.* **1948**, *36*, 51−8.
(19) Wolfenden, R. *Acc. Chem. Res.* **1972**, *5*, 10−8.
(20) Heightman, T. D.; Vasella, A. T. *Angew. Chem., Int. Ed.* **1999**, *38*, 750−70.
(21) Gloster, T. M.; Vocadlo, D. J. *Nat. Chem. Biol.* **2012**, *8*, 683−94.
(22) Spohr, U.; Bach, M.; Spiro, R. G. *Can. J. Chem.* **1993**, *71*, 1928−42.
(23) Alonzi, D. S.; Kukushkin, N. V.; Allman, S. A.; Hakki, Z.; Williams, S. J.; Pierce, L.; Dwek, R. A.; Butters, T. D. *Cell. Mol. Life Sci.* **2013**, *70*, 2799−814.
(24) Spohr, U.; Bach, M.; Spiro, R. G. *Can. J. Chem.* **1993**, *71*, 1919−27.
(25) Legler, G.; Roeser, K. R.; Illig, H. K. *Eur. J. Biochem.* **1979**, *101*, 85−92.
(26) Arribas, J. C.; Herrero, A. G.; Martin-Lomas, M.; Canada, F. J.; He, S.; Withers, S. G. *Eur. J. Biochem.* **2000**, *267*, 6996−7005.
(27) Hill, C. H.; Graham, S. C.; Read, R. J.; Deane, J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20479−84.
(28) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−67.
(29) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 386−94.
(30) Santana, A. G.; Vadlamani, G.; Mark, B. L.; Withers, S. G. *Chem. Commun.* **2016**, *52*, 7943−6.
(31) Watts, A. G.; Damager, I.; Amaya, M. L.; Buschiazzo, A.; Alzari, P.; Frasch, A. C.; Withers, S. G. *J. Am. Chem. Soc.* **2003**, *125*, 7532−3.

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

(32) von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. *Nature* **1993**, *363*, 418−23.
(33) Shidmoossavee, F. S.; Watson, J. N.; Bennet, A. J. *J. Am. Chem. Soc.* **2013**, *135*, 13254−7.
(34) Liu, H.; Liang, X.; Søhoel, H.; Bülow, A.; Bols, M. *J. Am. Chem. Soc.* **2001**, *123*, 5116−7.
(35) Goddard-Borger, E. D.; Stick, R. V. *Aust. J. Chem.* **2007**, *60*, 211−3.
(36) Mayer, T. G.; Schmidt, R. R. *Eur. J. Org. Chem.* **1999**, *1999*, 1153−65.
(37) Meloncelli, P. J.; Gloster, T. M.; Money, V. A.; Tarling, C. A.; Davies, G. J.; Withers, S. G.; Stick, R. V. *Aust. J. Chem.* **2007**, *60*, 549−65.
(38) Davies, G. J.; Planas, A.; Rovira, C. *Acc. Chem. Res.* **2012**, *45*, 308−16.
(39) Williams, R. J.; Iglesias-Fernandez, J.; Stepper, J.; Jackson, A.; Thompson, A. J.; Lowe, E. C.; White, J. M.; Gilbert, H. J.; Rovira, C.; Davies, G. J.; Williams, S. J. *Angew. Chem., Int. Ed.* **2014**, *53*, 1087−91.
(40) Tankrathok, A.; Iglesias-Fernández, J.; Williams, R. J.; Pengthaisong, S.; Baiya, S.; Hakki, Z.; Robinson, R. C.; Hrmova, M.; Rovira, C.; Williams, S. J.; Ketudat Cairns, J. R. *ACS Catal.* **2015**, *5*, 6041−51.
(41) Warnke, S.; von Helden, G.; Pagel, K. *J. Am. Chem. Soc.* **2013**, *135*, 1177−80.
(42) Winter, G. *J. Appl. Crystallogr.* **2010**, *43*, 186−90.
(43) Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−32.
(44) Evans, P. R. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 282−92.
(45) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204−14.
(46) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(47) Tickle, I. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 454−67.
(48) Agirre, J.; Davies, G.; Wilson, K.; Cowtan, K. *Nat. Chem. Biol.* **2015**, *11*, 303.
(49) Agirre, J.; Iglesias-Fernandez, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D. *Nat. Struct. Mol. Biol.* **2015**, *22*, 833−4.
(50) Schanda, P.; Kupče, Ē.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199−211.
(51) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−93.
(52) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−4.
(53) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993−2006.
(54) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−8.
(55) Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129*, 10686−93.
(56) Raich, L.; Borodkin, V.; Fang, W.; Castro-Lopez, J.; van Aalten, D. M.; Hurtado-Guerrero, R.; Rovira, C. *J. Am. Chem. Soc.* **2016**, *138*, 3325−32.
(57) Ardèvol, A.; Rovira, C. *J. Am. Chem. Soc.* **2015**, *137*, 7528−47.
(58) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−6.
(59) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604−13.
(60) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354−8.
(61) Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1994**, *264*, 1−19.
(62) Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2015**, *119*, 736−42.

DOI: 10.1021/jacs.6b10075
J. Am. Chem. Soc. 2017, 139, 1089−1097

OPEN BIOLOGY

rsob.royalsocietypublishing.org

Research

Cite this article: Hemsworth GR *et al*. 2016 Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut. *Open Biol.* **6**: 160142.
http://dx.doi.org/10.1098/rsob.160142

Received: 10 May 2016
Accepted: 1 July 2016

Subject Area:
biochemistry/structural biology

Keywords:
xyloglucan, polysaccharide utilization loci, glycoside hydrolases

Authors for correspondence:
Harry Brumer
e-mail: brumer@msl.ubc.ca
Gideon J. Davies
e-mail: gideon.davies@york.ac.uk

†These authors contributed equally to this study.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsob.160142.

THE ROYAL SOCIETY PUBLISHING

Structural dissection of a complex *Bacteroides ovatus* gene locus conferring xyloglucan metabolism in the human gut

Glyn R. Hemsworth[1,†], Andrew J. Thompson[1,†], Judith Stepper[1], Łukasz F. Sobala[1], Travis Coyle[2], Johan Larsbrink[3,4], Oliver Spadiut[3,5], Ethan D. Goddard-Borger[6], Keith A. Stubbs[2], Harry Brumer[3,4] and Gideon J. Davies[1]

[1]Department of Chemistry, York Structural Biology Laboratory, The University of York, Heslington, York YO10 5DD, UK
[2]School of Chemistry and Biochemistry, The University of Western Australia, Crawley, Western Australia 6009, Australia
[3]Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden
[4]Michael Smith Laboratories and Department of Chemistry, University of British Columbia, 2185 East Mall, Vancouver, British Columbia, Canada V6T 1Z4
[5]Wallenberg Wood Science Center, Royal Institute of Technology (KTH), Teknikringen 56–58, 100 44 Stockholm, Sweden
[6]The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville Victoria 3052, Australia

GRH, 0000-0002-8226-1380

The human gastrointestinal tract harbours myriad bacterial species, collectively termed the microbiota, that strongly influence human health. Symbiotic members of our microbiota play a pivotal role in the digestion of complex carbohydrates that are otherwise recalcitrant to assimilation. Indeed, the intrinsic human polysaccharide-degrading enzyme repertoire is limited to various starch-based substrates; more complex polysaccharides demand microbial degradation. Select Bacteroidetes are responsible for the degradation of the ubiquitous vegetable xyloglucans (XyGs), through the concerted action of cohorts of enzymes and glycan-binding proteins encoded by specific xyloglucan utilization loci (XyGULs). Extending recent (meta) genomic, transcriptomic and biochemical analyses, significant questions remain regarding the structural biology of the molecular machinery required for XyG saccharification. Here, we reveal the three-dimensional structures of an α-xylosidase, a β-glucosidase, and two α-L-arabinofuranosidases from the *Bacteroides ovatus* XyGUL. Aided by bespoke ligand synthesis, our analyses highlight key adaptations in these enzymes that confer individual specificity for xyloglucan side chains and dictate concerted, stepwise disassembly of xyloglucan oligosaccharides. In harness with our recent structural characterization of the vanguard endo-xyloglucanse and cell-surface glycan-binding proteins, the present analysis provides a near-complete structural view of xyloglucan recognition and catalysis by XyGUL proteins.

1. Background

The metabolism of complex carbohydrates in the distal gastrointestinal (GI) tract is central to human nutrition and health [1,2]. It is widely understood that a well-balanced human diet consists of a significant proportion of fruits and vegetables, the cell walls of which are primarily (approx. 90% of the dry weight) comprised of a structurally diverse array of intrinsically non-digestible polysaccharides popularly referred to as 'dietary fibre' [1–5]. The human genome is, however, remarkably bereft of genes encoding the enzymes

necessary to digest the manifold plant polysaccharides we ingest, with the exception of the α-glucans, amylose and amylopectin, that constitute starch [6]. Even in this case, certain structurally compact, recalcitrant forms, known as 'resistant starches' (RS), may reach the colon intact [3]. Both RS and the diverse non-starch polysaccharides (NSP) of plant cell walls are instead metabolized, to various extents, by our symbiotic gut microbiota. Microbial fermentation of monosaccharides in the gut produces short chain fatty acids (SCFAs), which provide a notable proportion (up to 10%) of our daily caloric intake. In addition, localized butyrate production is particularly required to maintain a healthy colonic epithelium [7–9]. There is, therefore, intense current research focus on (and considerable popular interest in) potential causal links between imbalance of the microbiota (dysbiosis) and a wide array of human diseases, including irritable bowel diseases, persistent *Clostridium difficile* infection, metabolic syndrome, diabetes, atopy and neurological disorders [10–14].

Thus, human health is crucially dependent on the population dynamics of the gut ecosystem, which is, in turn, rooted in the capacity of the microbiota to utilize the complex carbohydrates that we are otherwise incapable of accessing [15,16]. Strikingly, many *individual* microbiotal species, especially from the phylum Bacteroidetes, possess the genetic capacity to produce *hundreds* of predicted carbohydrate-active enzymes (CAZymes) [6,17]. This tremendous diversity is directly reflective of the natural structural complexity of plant, fungal and animal oligosaccharides and polysaccharides in the human diet [5,16]. Numerous (meta)genomic, transcriptomic and proteomic studies are continuing to provide a wealth of information on the genetic potential and dynamic response of the human gut microbiome with regard to complex carbohydrate catabolism [9,17–22]. However, our functional understanding of the molecular mechanisms fuelling this ecosystem is currently only in its infancy, due to a comparative paucity of enzymology and structural biology [23,24]. Indeed, among glycoside hydrolases (GH) from all organisms, biochemically and structurally characterized examples total only approximately 5% and 0.5%, respectively, of known open-reading frames (ORFs) [25]; these values are much lower for gut bacterial species.

The two dominant phyla in the colon of healthy adult humans are the Gram-positive Firmicutes and the Gram-negative Bacteroidetes [26], individual species of which have been implicated as key contributors to the breakdown of NSP in the diet [17,19,27,28]. Bacteroidetes are particularly notable for organizing cohorts of CAZymes and binding, transport and sensor/regulator proteins into contiguous polysaccharide utilization loci (PULs) [23,29,30]. Bacteroidetes PUL complexity generally scales with the monosaccharide and linkage complexity of the cognate substrate, especially with regard to the number of GHs and polysaccharide lyases (PLs) [17,19,23]. As such, PULs often encode complete molecular systems for the specific utilization of individual polysaccharides. Likewise, intimate coordination of substrate adherence and initial backbone cleavage at the cell surface, followed by complete oligosaccharide hydrolysis in the confines of the periplasmic space, represents a particularly elegant evolutionary strategy to limit loss of monosaccharides to the competitive gut environment [31] (figure 1).

Transcending 'omics' surveys of the gut microbiota, an emerging methodology for the in-depth functional characterization of PULs combines bacterial genetics, biochemistry and enzymology, and structural biology. A growing number of such system-based approaches have been used to elucidate the complex molecular details of fructan [36], seaweed porphyran [37], yeast mannan [38] and cereal xylan [39] utilization by symbiotic human gut *Bacteroides* species. In this context, we recently reported the characterization of a novel xyloglucan utilization locus (XyGUL) that confers *Bacteroides ovatus*, and species harbouring syntenic XyGULs, with the ability to utilize this abundant vegetable polysaccharide across sampled human populations [32]. In this work, the complete biochemical and crystallographic characterization of the vanguard *endo*-xyloglucanase responsible for initiating substrate backbone cleavage was presented, in addition to biochemical data revealing the substrate specificity of the six downstream *exo*-glycosidases. Together, these data allowed us to outline a general pathway for dietary xyloglucan saccharification to monosaccharides for primary metabolism. Until now, however, molecular-level insight into xyloglucan oligosaccharide (XyGO) recognition and hydrolysis by these key downstream enzymes has been lacking. Here, we present the three-dimensional structures of *Bo*GH31, *Bo*GH43A, *Bo*GH43B and *Bo*GH3B, expanding our knowledge of the structural determinants required for xyloglucan degradation (figure 1). Our analyses highlight key adaptations in these enzymes that confer their specificity for xyloglucan oligosaccharides, while also providing a rationale for the maintenance of two divergent genes coding for GH3 enzymes, and similarly two divergent genes for GH43 family members, within the same PUL.

2. Material and methods

2.1. Cloning, over-expression and structure determination of *Bo*GH31

For structural characterization, the gene encoding *Bo*GH31 was recloned from pET21a(GH31) [32] into a modified pET28a vector (pET-YSBL3C) containing an N-terminal His_6-tag for immobilized metal affinity purification (IMAC) and 3C-cleavage site to allow subsequent removal of the tag [40]. The GH31 ORF was amplified from the pET21a(GH31) template and cloned into linearized pET-YSBL3C using the InFusion-HD cloning kit (ClonTech), according to the manufacturer's instructions, to give pET-YSBL3C(GH31).

Chemically competent *Escherichia coli* TUNER(DE3) cells were transformed with the pET-YSBL3C(GH31) vector and grown in LB medium containing 50 μg ml^{-1} kanamycin at 37°C. Once the cells reached an OD_{600nm} of 0.8–1.0, the temperature was lowered to 16°C and expression was induced by the addition of isopropyl β-D-galactopyranoside (IPTG) to a final concentration of 200 μM and the expression was allowed to proceed overnight. Cells were harvested by centrifugation at 10 800*g* for 20 min at 4°C. Spent medium was discarded and the cells were resuspended in 5× volumes of Buffer A (50 mM HEPES pH 7, 0.3 M NaCl, 10 mM imidazole). Cells were lysed with four 20 s pulses of sonication at maximum amplitude in an MSE Soniprep 150 sonicator on ice. Cell debris was removed by centrifugation at 3900*g* in a cooled bench top centrifuge and the cleared lysate was applied directly to a 5 ml HisTrap FF Crude column (GE Healthcare). After washing with 5–6 volumes of Buffer A, protein was eluted with a linear gradient from 0 to 100% Buffer B (50 mM HEPES pH 7, 0.3 M NaCl, 500 mM imidazole) over

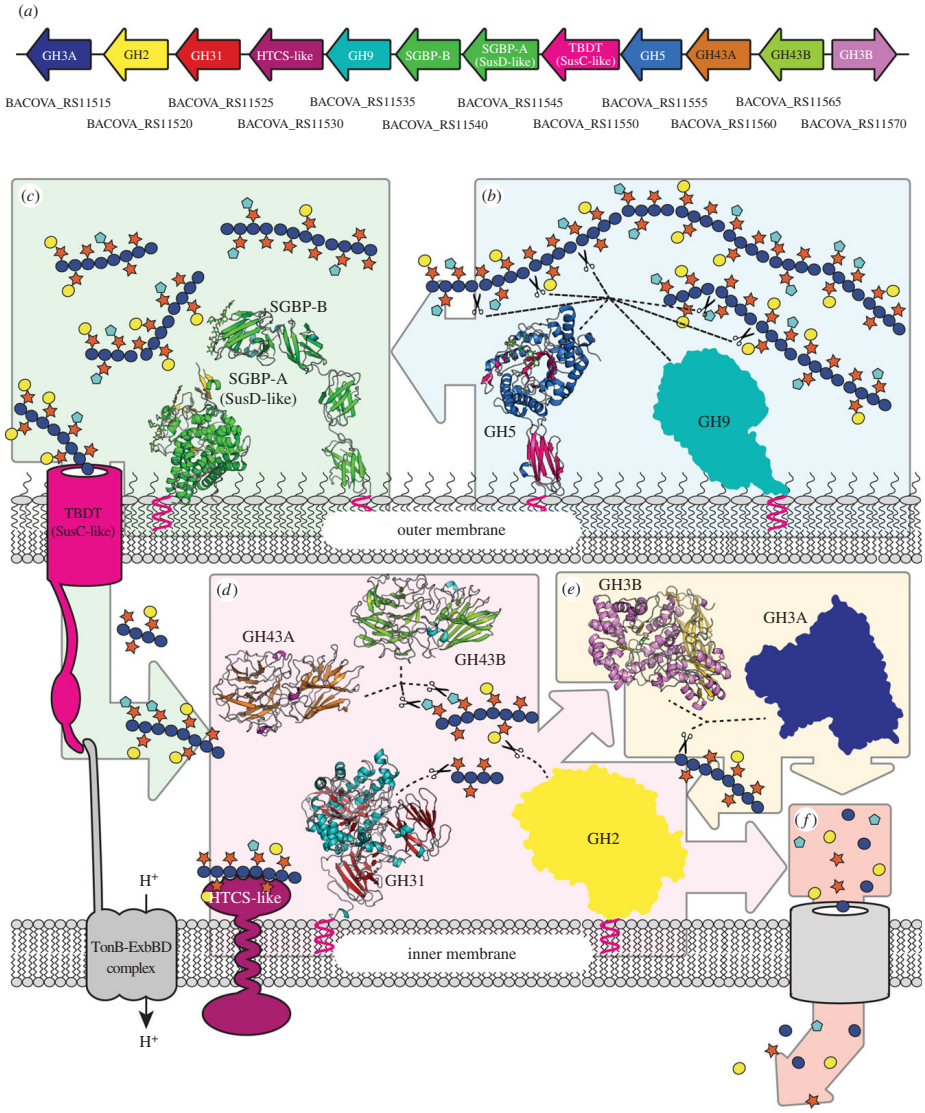


Figure 1. Summary of the xyloglucan saccharification system encoded by the *Bo*XyGUL. (*a*) Gene organization of the *Bo*XyGUL, coloured with reference to the proteins shown in subsequent panels. GenBank locus tag numbers are indicated below each gene. (*b*) *endo*-Xyloglucanases GH5 (structure from [32]) and GH9 localized to the surface of the cell cleave long xyloglucan polysaccharides into smaller fragments, with glycan capture facilitated by cell-surface glycan-binding proteins SGBP-A and SGBP-B (structures from [33]). (*c*) The resulting shorter oligosaccharides are imported into the periplasm via the TonB-dependent transporter (TBDT) for further processing. (*d*) *exo*-Glycosidases GH31, GH43A, GH43B and GH2 remove pendant xylosyl, arabinofuranosyl and galactosyl side chains. (*e*) GH3A and GH3B β-glucosidases act from the non-reducing end liberating individual glucose residues. The oligosaccharides can be further hydrolysed, by these individual enzymes, into monosaccharides. (*f*) The liberated sugars are imported into the cell and metabolized. See [34] for a detailed summary of XyG side-chain structures; monosaccharides are represented using standard Consortium of Functional Glycomics symbols [35].

20 column volumes, collecting 6 ml fractions. Peak fractions containing *Bo*GH31 were combined and concentrated to less than 2 ml using a 50 kDa cut-off Sartorius concentrator before being applied to a HiTrap 16/60 superdex 200 column (GE Healthcare), which had been equilibrated with 25 mM HEPES pH 7, 100 mM NaCl and 1 mM DTT. After a void volume of 40 ml, 2 ml fractions were collected and those containing *Bo*GH31 were combined and concentrated using a

50 kDa cut-off Sartorius concentrator. Protein concentration was determined to be 35 mg ml^{-1} as judged by A_{280nm} using an extinction coefficient of 238735 M^{-1} cm^{-1} and a molecular weight of 109 815.6 Da.

Crystals of *Bo*GH31 were obtained by hanging drop vapour diffusion (19°C) using 0.2 M potassium thiocyanate, 14–20% (w/v) PEG-3350 as mother liquor and were used for subsequent structure determination. Crystals were cryo-cooled for data collection at 100 K by plunging in liquid nitrogen after a 30 s soak in mother liquor supplemented with 20% ethylene glycol. Crystals of *Bo*GH31 in complex with a covalent inhibitor, 5FIdoF [41,42], were obtained by soaking native crystals in 10 mM (final) 5FIdoF supplemented with mother liquor for 30 s, immediately prior to cryocooling. Diffraction data for native *Bo*GH31 were collected at Diamond Light Source, beamline i04-1 at a wavelength of 0.920 Å, while data for the covalent 5FIdoF complex were collected at beamline i04 (also Diamond Light Source, $\lambda = 0.9795$ Å). All data were indexed and integrated using XDS [43] with all subsequent processing steps performed using the CCP4 software suite [44]. The structure was solved by molecular replacement in MOLREP [44] using the protein chain in PDB entry 2xvg as the search model. An initial model was generated using ARP-WARP [45] before subsequent model building and refinement were performed in COOT [46] and REFMAC [47], respectively.

2.2. Cloning, over-expression and structure determination of *Bo*GH43A

For structural characterization, the gene encoding *Bo*GH43A was recloned from pET21a [32] into pET28a containing an N-terminal His_6-tag for IMAC. The *Bo*GH43A ORF was amplified from the pET21a(*Bo*GH43A) template and cloned into linearized pET28a using the InFusion-HD cloning kit (ClonTech) according to the manufacturer's instructions. Protein expression and purification were performed exactly as described above for *Bo*GH31. The final *Bo*GH43A sample was concentrated on a 30 kDa cut-off Sartorius concentrator to 103 mg ml^{-1} as judged by A_{280nm} using an extinction coefficient of 105 450 M^{-1} cm^{-1} and a molecular weight of 57 965.1 Da.

Crystals of *Bo*GH43A were obtained by hanging drop vapour diffusion (19°C) using 0.1 M Tris pH 7.2–7.8, 0.18 M magnesium chloride and 12% (w/v) PEG-6000 as mother liquor and were used for subsequent structure determination. Crystals were cryo-cooled for data collection at 100 K by plunging in liquid nitrogen after a 30 s soak in mother liquor supplemented with 20% ethylene glycol. Crystals of *Bo*GH43A in complex with AraDNJ and AraLOG were obtained by soaking native crystals in 10 mM (final) solutions of respective compounds supplemented with mother liquor for 60 min, prior to cryocooling. Diffraction data for native *Bo*GH43A were collected at Diamond Light Source, beamline i04-1 at a wavelength of 0.920 Å, while datasets for AraDNJ and AraLOG complexes were both collected at beamline i03 ($\lambda = 0.9795$ Å). All data were indexed and integrated using XDS [43] with all subsequent processing steps performed using the CCP4 software suite [44]. The structure was solved by molecular replacement in PHASER [48] using the protein chain from previously solved *Bo*GH43B as the search model. An initial model was generated using BUCCANEER [49,50] before subsequent model building and refinement were performed in COOT [46] and REFMAC [47], respectively.

2.3. Over-expression and structure determination of *Bo*GH43B

Chemically competent *E. coli* BL21 (DE3) cells were transformed with pET21a(*Bo*GH43B) [32] and grown in LB medium containing 100 μg ml^{-1} ampicillin at 37°C. Once the cells reached an OD_{600} of 0.4–0.6, the temperature was lowered to 16°C and expression was induced by the addition of IPTG to a final concentration of 100 μM and the expression was allowed to proceed overnight. Cells were harvested by centrifugation at 10 800*g* for 20 min at 4°C. Spent medium was discarded and the cells were resuspended in 5× volumes of Buffer A (50 mM HEPES pH 7, 0.5 M NaCl, 30 mM imidazole). Cells were lysed with four 20 s pulses of sonication at maximum amplitude in an MSE Soniprep 150 sonicator on ice. Cell debris was removed by centrifugation at 39 000*g* and the supernatant was applied directly to a 5 ml HisTrap FF Nickel NTA column (GE HEalthcare). After washing with five volumes of Buffer A, protein was eluted with a linear gradient from 0 to 100% Buffer B (50 mM HEPES pH 7, 0.5 M NaCl, 300 mM imidazole) over 20 column volumes, collecting 1.6 ml fractions. Peak fractions containing *Bo*GH43B were combined and concentrated to less than 1 ml using a 30 kDa cut-off Sartorius concentrator before being applied to a HiTrap 16/60 superdex 200 column (GE Healthcare) which had been equilibrated with 10 mM HEPES pH 7, 250 mM NaCl. After a void volume of 40 ml, 1.6 ml fractions were collected and those containing *Bo*GH43B were combined, concentrated and buffer exchanged with 10 mM HEPES pH 7 on a 30 kDa cut-off Sartorius concentrator. Protein concentration was determined to be 10 mg ml^{-1} as judged by A_{280nm} using an extinction coefficient of 102 790 M^{-1} cm^{-1} and a molecular weight of 57 243.3 Da.

Crystals of *Bo*GH43B were obtained by hanging drop vapour diffusion using 0.2 M sodium acetate pH 5, 20–30% PEG-3350 as mother liquor and they were used for subsequent structure determination. Crystals were cryo-cooled for data collection at 100 K by plunging in liquid nitrogen after a 30 s soak in mother liquor supplemented with 20% ethylene glycol. Diffraction data were collected at Diamond Light Source, beamline i02 at a wavelength of 0.980 Å. The data were indexed and integrated using XDS [43] with all subsequent processing steps performed using the CCP4 software suite [44]. The structure was solved by molecular replacement in PHASER [48] using the protein chain in PDB entry 1yrz as the search model. The initial phases were improved using PARROT [51] and an initial model generated using BUCCANEER [49,50] before subsequent model building and refinement were performed in COOT [46] and REFMAC [47], respectively.

2.4. Over-expression and structure determination of GH3B

GH3B expression and purification from the pET21a(GH3B) construct created by Larsbrink *et al*. [32] was performed as described above for *Bo*GH43B. The final sample was prepared at 10 mg ml^{-1} as judged by the A_{280nm} using an extinction coefficient of 142 670 M^{-1} cm^{-1} and a molecular weight of 86 512.6 Da.

Crystals were obtained by hanging drop vapour diffusion using 0.2 M sodium acetate and 15–25% PEG-3350 as the

mother liquor. Crystals were cryo-cooled by plunging in liquid nitrogen using mother liquor supplemented with 20% ethylene glycol as the cryo-protectant prior to data collection at Diamond Light Source, beamline i04-1 at a wavelength of 0.920 Å. Indexing and integration of diffraction data was performed with XDS [43] with all subsequent data processing performed using the CCP4 software suite [44]. Data were phased by molecular replacement in PHASER [48] using the barley β-glucosidase structure 1ex1 [52] as the search model. Phase improvement was performed using PARROT [51] before generation of an initial model using BUCCANEER [49,50]. Subsequent model building and refinement were performed in COOT [46] and REFMAC [47], respectively. TLS refinement using two TLS groups per protein chain was invoked towards the end of structure refinement.

2.5. Synthesis of arabinofuranosidase inhibitors

2.5.1. General

^{1}H and ^{13}C nuclear magnetic resonance spectra were obtained on Bruker ARX500 (500 MHz for ^{1}H and 125 MHz for ^{13}C) or Bruker AV600 (600 MHz for ^{1}H and 150 MHz for ^{13}C) spectrometers (see the electronic supplementary material). Mass spectra were recorded with a Waters GCT Premier spectrometer using electrospray ionization (ES).

2.5.2. (*E*) and (*Z*)-2,3,5-Tri-*O*-acetyl-L-arabinofuranose oxime (**2**)

Hydroxylamine hydrochloride (240 mg, 3.45 mmol) was added to a solution of the hemiacetal **1** [53] (610 mg, 2.21 mmol) and pyridine (0.45 ml, 5.5 mmol) in MeOH (20 ml) and the mixture was stirred at reflux (2 h). Concentration of the solution by co-evaporation with toluene (3 × 15 ml) followed by flash chromatography of the residue (6 : 4 EtOAc/hexanes) produced the presumed oxime **2** as a white solid (575 mg, 94%). R_f 0.40 (7 : 3 EtOAc/hexanes). This solid was used without further purification.

2.5.3. (*Z*)-2,3,5-Tri-*O*-acetyl-L-arabinonhydroximo-1,4-lactone (**3**)

1,8-Diazabicyclo[5.4.0]undec-7-ene (0.35 ml, 2.3 mmol) was added to a solution of the oxime **2** (575 mg, 2.08 mmol) and NCS (305 mg, 2.28 mmol) in CH_2Cl_2 (21 ml) at −40°C, in such a way that the temperature did not rise above −35°C, and the resulting mixture was stirred at −40°C for 1 h before being allowed to warm to room temperature over 2 h. The resulting solution was quenched with water and diluted with CH_2Cl_2 (20 ml). The organic layer was separated and washed with water (3 × 15 ml), brine, dried ($MgSO_4$), filtered and concentrated. Flash chromatography of the residue (3 : 2 EtOAc/hexanes) yielded the triacetate **3** as a colourless oil (410 mg, 71%). R_f 0.38 (7 : 3 EtOAc/hexanes). ^{1}H NMR (500 MHz, $CDCl_3$): δ 6.96 (*br* s, 1H), 5.74 (d, 1H, J = 2.8 Hz), 5.22–5.20 (m, 1H), 4.68–4.63 (m, 1H), 4.42 (dd, 1H, J = 4.5, 12.0 Hz), 4.31 (dd, 1H, J = 6.0, 12.0 Hz), 2.15 (s, 3H), 2.13–2.11 (m, 6H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 170.66, 169.89, 169.28, 154.26, 83.37, 74.90, 72.46, 62.39, 20.69, 20.65. HRMS (ES): m/z = 312.0683; $[M + Na]^+$ requires 312.0695.

2.5.4. (*Z*)-L-Arabinonhydroximo-1,4-lactone (AraLOG)

Saturated ammonia in MeOH (5 ml) was added to a solution of the triacetate **3** (100 mg, 0.346 mmol) in MeOH (5 ml) at 0°C and the solution was allowed to stand (0°C, 2 h). Concentration of the solution followed by flash chromatography of the residue (3 : 7 MeOH/EtOAc) yielded the title compound (39 mg, 68%). R_f 0.37 (3 : 7 MeOH/EtOAc). ^{1}H NMR (600 MHz, D_2O): δ 4.70 (d, 1H, J = 7.2 Hz), 4.39–4.33 (m, 1H), 4.20 (dd, 1H, J = 7.2 Hz), 4.00–3.95 (m, 1H), 3.81 (dd, 1H, J = 4.8, 13.2 Hz); ^{13}C NMR (150 MHz, D_2O): δ 159.00, 84.79, 73.95, 73.71, 59.99. HRMS (ES): m/z = 164.0551; $[M + H]^+$ requires 164.0559.

2.5.5. (*Z*)-*O*-(2,3,5-Tri-*O*-acetyl-L-arabinosylidene)amino *N*-phenylcarbamate (**4**)

Phenyl isocyanate (50 µl, 0.46 mmol) was added to a solution of the triacetate **3** (105 mg, 0.363 mmol) and Et_3N (0.16 ml, 1.2 mmol) in THF (5 ml) at 0°C and the solution was stirred (0°C, 2 h). Concentration followed by flash chromatography of the residue (1 : 1 EtOAc/hexanes) produced the carbamate **4** as a colourless foam (90 mg, 57%). R_f 0.31 (1 : 1 EtOAc/hexanes). ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.76 (*br* s, 1H), 7.49–7.44 (m, 2H), 7.36–7.30 (m, 2H), 7.14–7.09 (m, 1H), 5.86 (d, 1H, J = 3.0), 5.24 (dd, 1H, J = 2.5, 3.0 Hz), 4.77–4.74 (m, 1H), 4.46 (dd, 1H, J = 4.5, 12.5 Hz), 4.34 (dd, 1H, J = 6.0, 12.5 Hz), 2.20 (s, 3H), 2.15 (s, 3H), 2.14 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 170.38, 169.70, 168.88, 157.69, 151.21, 137.00, 129.07, 124.14, 119.36, 85.25, 77.16, 74.70, 72.85, 62.02, 20.60, 20.52. HRMS (ES): m/z = 409.1248; $[M + H]^+$ requires 409.1247.

2.5.6. (*Z*)-*O*-(L-Arabinosylidene)amino *N*-phenylcarbamate (AraPUG)

Saturated ammonia in MeOH (5 ml) was added to a solution of the carbamate **4** (80 mg, 0.20 mmol) in MeOH (5 ml) at 0°C and the solution was allowed to stand (0°C, 2 h). The resulting solution was concentrated to yield a white solid. Trituration of the solid (1 : 4 : 95 H_2O/MeOH/EtOAc) yielded the title compound as a white powder (43 mg, 78%). R_f 0.26 (1 : 9 MeOH/EtOAc). ^{1}H NMR (600 MHz, $(CD_3)_2SO$): δ 9.78 (*br* s, 1H), 7.52–7.47 (m, 2H), 7.32–7.26 (m, 2H), 7.03–6.99 (m, 1H), 6.21 (*br* s, 1H), 5.85 (*br* s, 1H), 5.14 (*br* s, 1H), 4.46 (d, 1H, J = 4.8 Hz), 4.26–4.22 (m, 1H), 4.01 (m, 1H), 3.71 (m, 1H), 3.58 (m, 1H); ^{13}C NMR (150 MHz, $(CD_3)_2SO$): δ 163.17, 151.81, 138.71, 128.75, 122.71, 118.58, 88.38, 74.45, 73.77, 59.91. HRMS (ES): m/z = 283.0928; $[M + H]^+$ requires 283.0930.

2.6. Binding constant determination for AraF inhibitors

Binding of two arabinofuranosidase inhibitors, AraDNJ and AraLOG, to *Bo*GH43A and *Bo*GH43B was investigated by isothermal titration calorimetry (ITC) in a MicroCal Auto-ITC200 system (GE Healthcare/Malvern Instruments). *Bo*GH43A titrations were performed in 25 mM HEPES pH 7.0, 100 mM NaCl and 1 mM DTT, while *Bo*GH43B titrations used 25 mM HEPES pH 7.0, 100 mM NaCl. Ligands were prepared by dilution in the identical buffer as used for protein sample preparation. AraLOG binding could not be detected to either *Bo*GH43A or B with titrations performed in triplicate at 25°C, with 1 mM AraLOG titrated into 100 µM pure protein. An interaction between AraDNJ and both proteins, however, could be detected but appeared to be weak and so low c-value ITCs were performed to obtain binding data [54]. Assays were conducted in triplicate at

25°C, with 2 mM AraDNJ titrated into approximately 100 µM protein (more precise protein concentrations were measured for each sample immediately before performing the titrations and these values were used for data fitting in ORIGIN). To obtain saturation, titrations were split into two runs, the first consisting of a single 1 µl injection at the start of the run (discarded during the analysis) followed by 19× 2 µl injections of ligand. At the end of this run 39 µl was removed from the cell, the syringe was refilled with ligand and the titration was continued with 20 additional 2 µl injections. CONCAT32 (MicroCal) was then used to concatenate the data together into a single titration. To account for heats of dilution, an additional titration was performed in exactly the same way, titrating ligand into buffer. These reference data were then subtracted from all experimental data which were subsequently used to calculate dissociation constants (K_d) using the ORIGIN 7 software package by fixing the *N*-value at 1.0 during the fitting (MicroCal, see figure 3*d*).

3. Results and discussion

3.1. Structure of the α-xylosidase *Bo*GH31

As with many of the glycoside hydrolase families represented within the *Bo* xyloglucan PUL (XyGUL), GH31 forms a large (currently over 3800 sequences) and functionally diverse collection of enzymes, with many α-glucosidases, α-xylosidases and α-galactosidases featuring prominently [25]. Within XyGULs, GH31 α-xylosidases play an essential role removing xylose from the non-reducing end of processed xyloglucan oligosaccharides (illustrated in figure 1*d*). Such activity permits enzymatic access to the β-1,4-linked glucose moieties of the XyGO backbone. Indeed, deletion of the gene encoding GH31 from the XyGUL completely abrogates the ability of *B. ovatus* to grow on XyG and XyGOs [32]. Consistent with this role, the GH31 α-xylosidase present within the *Bo* XyGUL (*Bo*GH31) has been shown to be highly active against native XyGO substrates (XXXG and XLLG, nomenclature according to [34]), rather than disaccharide-configured activity probes, such as Xyl-α-PNP [32], despite the presence of optimized chemical leaving groups requiring little protonic assistance from the enzyme. These observations suggest substrate binding by XyGO-active GH31 enzymes to be a both complex and highly specific process, requiring recognition and occupancy of multiple sub-sites distal to the catalytic centre.

The crystal structure of *Bo*GH31 was determined to a resolution of 1.5 Å by molecular replacement using the coordinates of *Cj*Xyl31A, a functional homologue present in *Cellvibrio japonicus* (PDB ID: 2xvg, see [55]), as the search model (for X-ray data collection and refinement statistics, see the electronic supplementary material, table S1). A structural comparison of the refined *Bo*GH31 atomic model using PDBEFOLD [56] revealed close similarity to several other GH31 enzymes, including YicI from *E, coli* (currently the only other structurally characterized α-xylosidase [57]). However, by far the closest structural match to *Bo*GH31 was *Cj*Xyl31A (*Z* score = 33.1, with RMSD = 1.15 Å across 888 matched Cα positions). As observed for *Cj*Xyl31A, *Bo*GH31 presents with an extensive, modular structure featuring several accessory domains appended to a well-conserved TIM barrel-like structure (figure 2*a*) (for a full description of terms and domain nomenclature see [55]). The catalytic core of *Bo*GH31 is composed of residues 384 to 758, which form the central $(\beta/\alpha)_8$ (TIM) barrel fold and harbour the active site (discussed below). The domains decorating the central catalytic unit include an N-terminal β-sandwich domain formed by residues 16 to 213 with additional strands contributed by residues 363 to 383 when the peptide chain returns from a PA14 domain (residues 214 to 362). The presence of PA14 has been observed previously for GH31 in *Cj*Xyl31A and is believed to contribute to the recognition and binding of extended XyGO substrates, as was indicated by NMR spectroscopy and molecular docking studies [55,58]. C-terminal to the central catalytic unit, are two additional domains—the C-terminal proximal (residues 759–839) and distal (residues 840–954) β-sandwiches. While these accessory regions can be thought of as distinct subdomains, extensive interactions and packing of secondary structure elements against the central $(\beta/\alpha)_8$ barrel are strongly suggestive of a low-flexibility, monolithic structure.

The location of the *Bo*GH31 active site and identity of the catalytic amino acids were confirmed through analysis of a covalent enzyme-glycoside intermediate formed between crystals of native *Bo*GH31 and a nucleophile-trapping glycosyl fluoride, 5-fluoro-β-L-idosyl fluoride (5FIdoF) (figure 2*a–c*). Within the complex structure, 5FIdoF forms an α-glycosidic linkage to the side-chain carboxylate of Asp553 at the centre of the $(\beta/\alpha)_8$ barrel. 5FIdoF makes H-bonding interactions to Asp553, Arg613, Asp630 (O2 of the sugar ring), His709 and a highly coordinated water molecule positioned between Asp630 and Asp659 (O3) and Asp441 (O4 and the axially positioned F5 atom). Interestingly, the enzyme-bound 5FIdoF shows significant distortion away from the $^1C_4$ ground state expected for L-sugars, appearing in a $^1S_3$ conformation. Such a conformation is also reflected in various other covalent intermediates with GH31 enzymes, including *Cj*Xyl31A in complex with 5-fluoro-α-D-xylosyl fluoride (5FXylF; also $^1S_3$, see 2xvk [55]) and *Cj*Agd31B, a GH31 α-1,4-transglucosylase, in complex with 5-fluoro-α-D-glucosyl fluoride (5FGlcF; ligand appears midway between $^4C_1$ and $^1S_3$, see 4ba0 [59]).

The *Bo*GH31 covalent glycosyl-enzyme intermediate structure lends further support to the role of the PA14 domain in ligand binding [55]. This domain is in close proximity to the enzyme-bound 5FIdoF, with the side chain of Trp316 approximately 6.5 Å from the ligand (figure 2*d*). Furthermore, a fortuitously bound HEPES molecule, present in the protein buffer, can also be observed in the active site pocket below the plane of the 5FIdoF sugar ring and bridging the gap between ligand and PA14 (figure 2*c*). Within xyloglucan from both dicot and solanaceous species, side-chain xylose moieties are linked α-1,6 to the glucan backbone. Thus backbone sugars occupying the +1, and other potential positive sub-sites, would also highly likely be coordinated below the plane of a −1 xyloside, extending across and out of the catalytic $(\beta/\alpha)_8$ barrel. The positioning of HEPES therefore appears prescient, with the piperazine ring of the ligand engaged in a van der Waals' stacking interaction with Trp513 (catalytic domain) from above, and Trp316 of PA14 from below. The positioning of these aromatic side chains, in addition to numerous other amino acids capable of forming hydrogen bonds, is highly suggestive of a carbohydrate-binding motif, and therefore a direct role for PA14 in the coordination of extended XyGO substrates. A homologous role was proposed for the

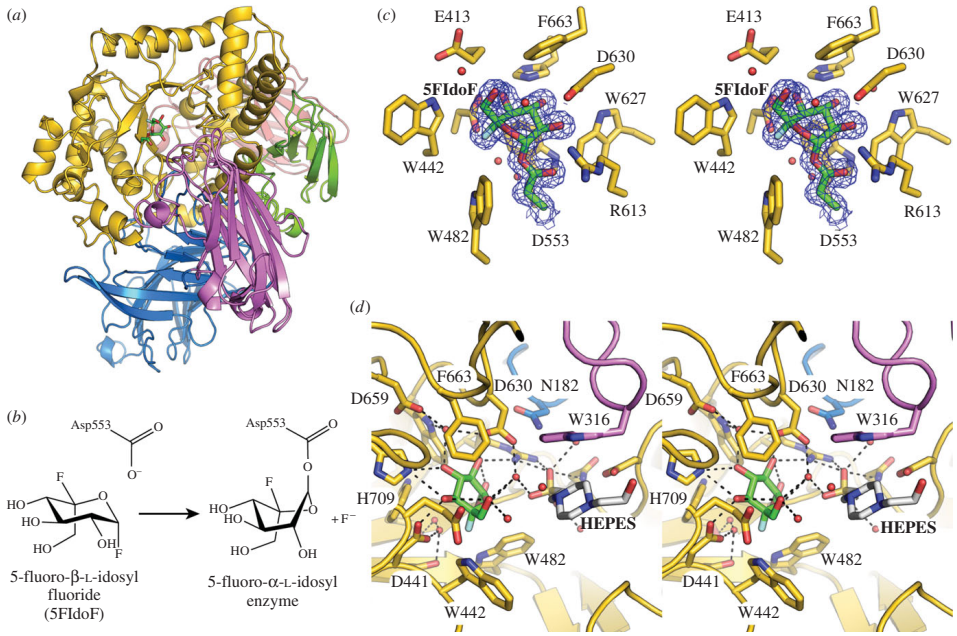


Figure 2. *Bo*GH31 α-xylosidase overall and active site structure. (*a*) Overall structure of *Bo*GH31 coloured by domain: N-terminal β sandwich in blue, the PA14 domain in purple, TIM barrel domain in gold, C-terminal proximal β-sandwich in green and C-terminal distal β-sandwich in red. The location of the active site revealed by the covalent glycosyl-enzyme intermediate is shown as sticks coloured by atom type with green carbons. (*b*) Mechanism of formation of the glycosyl-enzyme intermediate for the *Bo*GH31–5FIdoF complex. (*c*) Wall-eyed stereo view of the active site pocket. The active site nucleophile (Asp553) and 5FIdoF are coloured with green carbon atoms, with the surrounding active site side chains shown with gold carbon atoms. (*d*) Wall-eyed stereo view of the wider active site with the additional fortuitous HEPES molecule (white carbon atoms) shown revealing the likely role of the PA14 domain (purple) in extending the active site pocket to allow binding of longer xyloglucan oligosaccharides.

PA14 domain in the structurally similar, XyGO-specific *Cj*Xyl31A from the saprophyte *C. japonicus* [55,58].

3.2. Structures of the α-L-arabinofuranosidases *Bo*GH43A and *Bo*GH43B

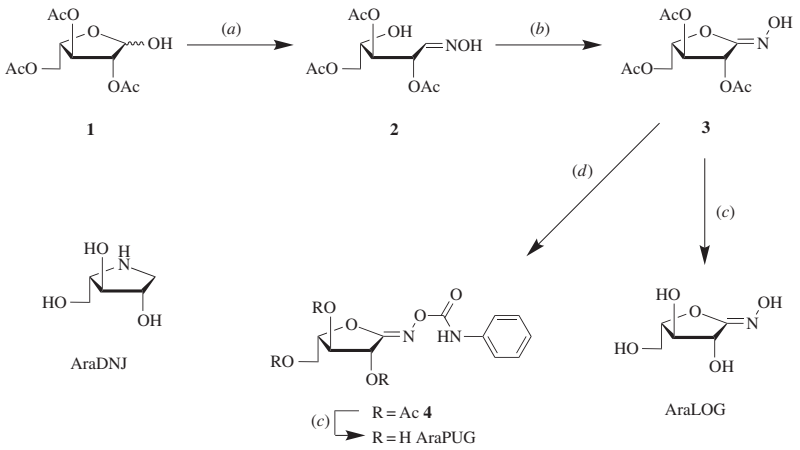
GH43 is a large and diverse family of CAZymes with members identified with β-xylosidase, α-L-arabinofuranosidase, arabinanase, xylanase, galactan 1,3-β-galactosidase, α-1,2-L-arabinofuranosidase, exo-α-1,5-L-arabinofuranosidase, exo-α-1,5-L-arabinanase and β-1,3-xylosidase activities. There are two GH43 family members represented in the *B. ovatus* xyloglucan PUL: *Bo*GH43A and *Bo*GH43B [32]. Both enzymes have demonstrable activity on L-Araf-α-PNP, though *Bo*GH43A was considerably more active, and both are thought to be responsible for the removal of pendant arabinofuranoside side chains from solanaceous xyloglucan substrates, thereby converting S to X for further processing by the α-xylosidase and other members of the PUL [32].

3.2.1. Synthesis of arabinofuranosidase inhibitors

To aid in the structural characterization of the *Bo*GH43A and *Bo*GH43B active sites, two new potential inhibitors for these enzymes were synthesized. The compounds were prepared incorporating an sp^2-hybridized carbon at carbon-1, which is thought to allow the carbohydrate ring to potentially adopt a conformation that is similar to the geometry of the transition state of glycosidase-catalysed reactions [60]. The synthesis of these inhibitors proceeded from the hemiacetal **1** (scheme 1) [53]. Treatment of the hemiacetal with hydroxylamine hydrochloride yielded the presumed mixture of oximes **2**, which were used without purification and converted to the hydroximolactone **3** in good overall yield. The inhibitor AraLOG was then prepared by treating **3** with saturated ammonia in methanol. Taking the hydroximolactone **3** and treating with phenyl isocyanate furnished the phenyl carbamate **4**. Deprotection of the carbamate **4** under similar conditions used to prepare AraLOG gave AraPUG in good yield.

3.2.2. *Bo*GH43A structure

The structure of *Bo*GH43A was determined to be 1.6 Å by molecular replacement using the structure of *Bo*GH43B described below as the search model (for X-ray data collection and refinement statistics, see the electronic supplementary material, table S2). Typical of all GH43s, *Bo*GH43A has a two-domain architecture, consisting of an N-terminal



Scheme 1. Synthesis of the putative α-L-arabinofuranosidase inhibitors AraLOG and AraPUG as well as the structure of the iminosugar AraDNJ. (*a*) $HONH_2$·HCl, C_5H_5N, MeOH; (*b*) NCS, DBU, CH_2Cl_2; (*c*) NH_3/MeOH (*d*) PhNCO, Et_3N, THF.

5-bladed β-propeller domain (residues 21 to 321) harbouring the catalytic active site, and a C-terminal β-sandwich domain (residues 322 to 522) which is frequently observed, though can be replaced by carbohydrate binding modules in some family members (see [61] for example) (figure 3*a*). Structural comparisons using PDBeFold [56] reveal close overall matches to other GH43s including XynB from *Bacillus subtilis* subsp. subtilis strain 168 (*Bs*XynB, 1yif; Z score = 17.8, with RMSD = 1.44 Å across 478 matched C_α positions) and XynB from *Bacillus halodurans* C-125 (*Bh*XynB, 1yrz; Z score = 17.7, RMSD = 1.45 Å across 473 matched C_αs), which all share the same two-domain architecture.

Within the native *Bo*GH43A structure, a TRIS molecule from the crystallization solution was observed bound in a shallow, enclosed pocket proposed to form the *Bo*GH43A −1 sub-site. Soaking of native *Bo*GH43A with two putative inhibitors, AraDNJ [62] and AraLOG, yielded respective enzyme–ligand complexes, confirming this as the active site (figure 3*b*). Disappointingly, no complexes were obtained with AraPUG, despite the use of high concentrations of inhibitor. AraDNJ was able to displace TRIS from the −1 sub-site and appeared bound in a low-energy ^3E conformation typical of iminosugar 'furanose' inhibitors. The side-chain carboxylate of Asp140 (O3 and O4 positions), the backbone amino group of Ala94 (O4) and the OD2 atom of Asp34 all directly coordinated the inhibitor (figure 3*b*). GH43 members typically contain three highly conserved acidic residues in their active sites to impart activity [63]. Together with Asp34 as the general base, which activates water to attack the anomeric carbon, Glu189 is ideally poised as the general acid, while Asp140 completes the triplet of residues and is important for modulating the pK_a and orienting the general acid for catalysis. The positions of these residues are absolutely conserved with other GH43 members.

For the AraLOG complex, repeated soaking at concentrations of up to 25 mM AraLOG for several hours failed to displace TRIS from the −1 sub-site. Rather, AraLOG was instead observed at the +1 site, which would normally be occupied by xylose moieties in the XyGO substrate (figure 3*c*). The AraLOG complex therefore highlights key interactions at this +1 sub-site, with the inhibitor stacking against Tyr187 while also H-bonding directly to the side chains of Glu210 and Glu189. In the light of the inability of AraLOG to displace TRIS from the active site, ITC (in the absence of Tris) was used to probe the affinity of both *Bo*GH43A and *Bo*GH43B (discussed below) for these inhibitors. AraDNJ binds to *Bo*GH43A with $K_d = 35 \pm 4$ μM (figure 3*d*), while AraLOG binding was too weak to be measured using this technique, consistent with its inability to displace TRIS during crystal soaking.

3.2.3. *Bo*GH43B structure

Despite significant functional overlap with *Bo*GH43A, *Bo*GH43B, the second α-L-arabinofuranosidase present in the *Bo*XyGUL, shares just 41% sequence identity with *Bo*GH43A and appears to be significantly less active on the substrates tested [32]. The structure of *Bo*GH43B was determined to 2.3 Å resolution by molecular replacement using a β-1,4-xylosidase from *B. halodurans* (PDB ID 1yrz) as the search model (electronic supplementary material, table S3). Remarkably, given their apparent differences at the amino acid level, the structure of *Bo*GH43B appears extremely similar to that of *Bo*GH43A, which can be superimposed onto *Bo*GH43B, using GESAMT [44], with an RMSD of 1.24 Å over 482 amino acid residues (figure 4*a*). Comparison of tertiary folds reveals few significant differences between the two paralogues, with the most obvious being the presence of a metal binding site, occupied by calcium, towards the C-terminus of *Bo*GH43B. Such an equivalent site appears entirely absent within *Bo*GH43A. In some GH43 members, addition of divalent cations within the catalytic site has led to increased activity and stability for these enzymes [64–66]. However, the Ca^{2+}-binding site in *Bo*GH43B is located in the C-terminal β-sandwich domain, on the opposite side of the molecule from the active site, and similar sites in other family members have not been implicated in catalysis to date [63].

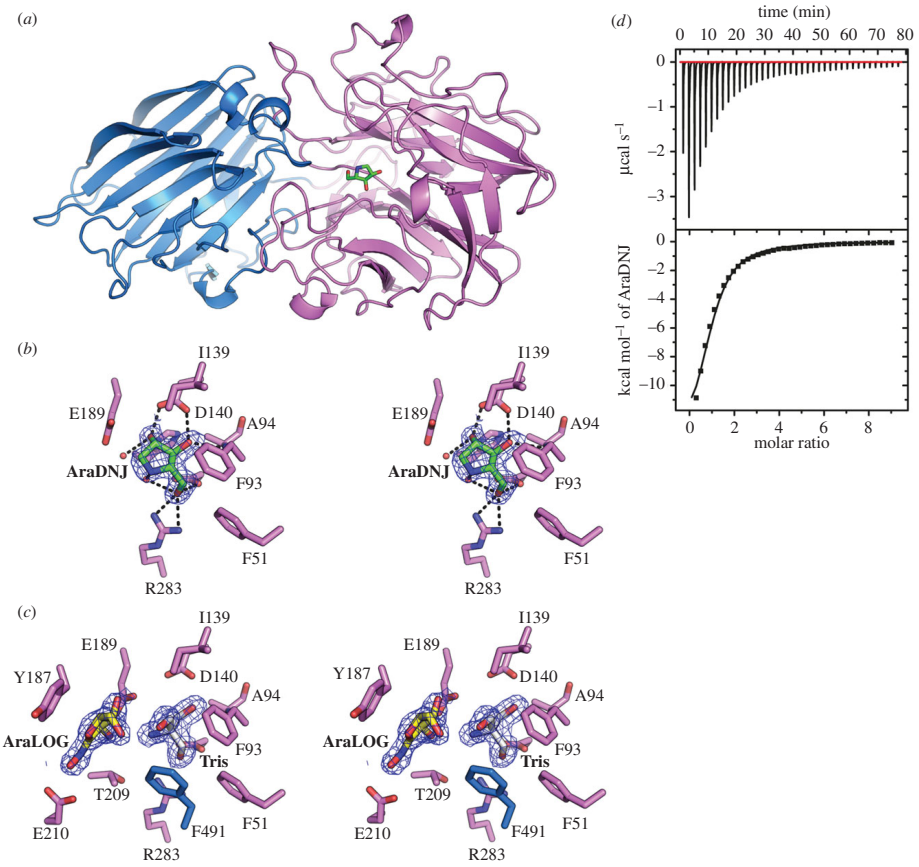


Figure 3. Overall structure and inhibitor binding to *Bo*GH43A α-L-arabinofuranosidase. (*a*) Overall structure of *Bo*GH43A; the N-terminal catalytic domain is coloured purple and C-terminal β-sandwich domain is coloured blue. The location of the active site is indicated by the position of AraDNJ shown in stick representation coloured by atom type with green carbons. (*b*) Wall-eyed stereo view of the active site in the *Bo*GH43A-AraDNJ complex. The final 2Fo-Fc map for the ligand is shown contoured at 1σ in blue. The hydrogen bonding interactions made by the inhibitor are shown as black dashed lines. (*c*) Wall-eyed stereo view of the active site for *Bo*GH43A-AraLOG complex. Binding of the AraLOG inhibitor (yellow carbon atoms) was too weak to displace TRIS (white carbon atoms) from the −1 sub-site and instead occupies +1, revealing key stacking interactions with Tyr187 and other conserved residues. (*d*) ITC thermogram showing the binding of AraDNJ to *Bo*GH43A in solution giving a K_d of 35 ± 4 µM.

Attempts to obtain structures of *Bo*GH43B in complex with the same inhibitors used for *Bo*GH43A were unsuccessful. ITC was used to determine the affinity of *Bo*GH43B for AraDNJ and AraLOG. *Bo*GH43B bound AraDNJ with a K_d of 111 ± 6 µM (figure 4*c*), while the affinity for AraLOG was too weak to be measured, as observed for *Bo*GH43A. This weaker binding affinity for AraDNJ also appears consistent with the lower specific activity of *Bo*GH43B for xyloglucan oligosaccharides when compared to its counterpart [32]. Superposition of apo-*Bo*GH43B with AraDNJ-*Bo*GH43A reveals that the three residues implicated in catalysis (Asp38, Asp148 and Glu198 in *Bo*GH43B) are absolutely conserved. The only difference in the *Bo*GH43B −1 sub-site is the replacement of Phe93 (in *Bo*GH43A) with a tyrosine residue in *Bo*GH43B. The +1 sub-site occupied by AraLOG in *Bo*GH43A, however, is considerably different. AraLOG stacks against Tyr187 in *Bo*GH43A, which is replaced by Ser196 in *Bo*GH43B. This variation means the active site pocket in *Bo*GH43B is considerably more open than in its XyGUL paralogue, possibly resulting in weaker substrate binding affinity and hence lower specific activity against authentic XyGO substrates. The reasoning that *B. ovatus* should harbour two GH43 members in its XyGUL remains unclear, but the differences in the active site architecture away from the −1 sub-site may represent the adaptation of these enzymes to specific substrate sources, possibly with alternate Araf structures on XyG branch termini [34].

3.3. Structure of β-glucosidase *Bo*GH3B

GH3 represents a large family of over 8000 sequences in the CAZy database. Like GH43, there are two GH3 members (*Bo*GH3A and *Bo*GH3B) present in the *Bo* XyGUL, both of

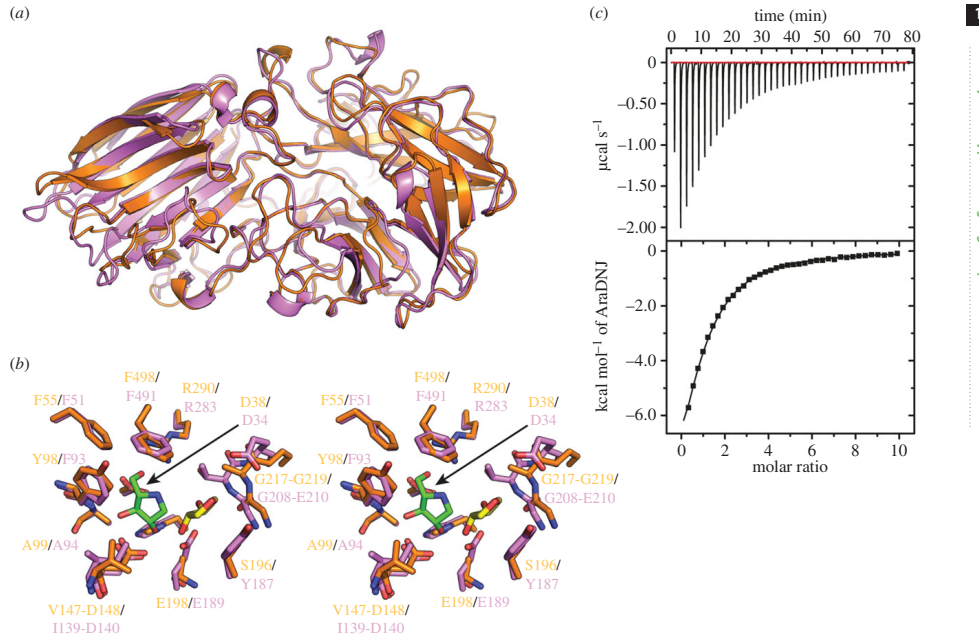


Figure 4. Structural comparison of *Bo*GH43B with *Bo*GH43A. (*a*) Overall superposition of *Bo*GH43B (orange) with *Bo*GH43A (purple) revealing a highly conserved tertiary structure arrangement between the two proteins, even though they share only 41% sequence identity. The largest differences between the two proteins are restricted to altered positioning of loops linking secondary structure elements. (*b*) Comparison of the arrangement of the active site residues between *Bo*GH43B (orange carbons) and *Bo*GH43A (purple carbons). The positions of AraDNJ (green carbons) and AraLOG (yellow carbons) from the *Bo*GH43A structure are shown. The −1 sub-site is mainly conserved between the two proteins, but the +1 sub-site surrounding the AraLOG ligand is significantly different between *Bo*GH43A and B, most notably Tyr187, which stacks against AraLOG in *Bo*GH43A, is replaced by Ser196 in *Bo*GH43B. (*c*) Representative ITC thermogram for the measurement of AraDNJ binding to *Bo*GH43B. Weaker binding was observed to this enzyme compared to *Bo*GH43A giving a K_d of 111 ± 6 µM, providing a rationale for our inability to obtain a *Bo*GH43B-AraDNJ complex structure.

which have been shown to be β-glucosidases with very similar specific activities. Despite apparently duplicated biochemical function, the two enzymes appear to have diverged significantly, sharing only 27% sequence identity at the amino acid level [32]. As for the GH43 enzymes, the functional significance of maintaining two seemingly identical β-glucosidases remains unclear, and so we aimed to structurally characterize both orthologues.

While GH3B proved readily amenably to crystallization, unfortunately, despite intense efforts, a similarly crystallizable form of GH3A could not be produced. The structure of GH3B was determined to 2.3 Å resolution (electronic supplementary material, table S4) by molecular replacement using the coordinates of barley β-glucosidase (PDB ID: 1ex1, see [52]) as the search model. *Bo*GH3B comprises a three-domain architecture, consisting of an N-terminal (TIM) barrel-like domain (residues 26 to 419), a central α/β sandwich domain (residues 420 to 660) and a fibronectin type-III (FN-III)-like domain at the C-terminus (residues 661–782) (figure 5*a*). Structural comparisons using PDBeFold [56] revealed close structural matches to several other GH3 members, the closest match being to a single protomer of a novel homodimeric GH3 identified in a metagenomic analysis of unnamed soil bacteria (PDBs: 3u48 and 3u4a), with RMSDs of 1.22 and 1.21 Å over 742 and 739 residues, respectively. The dimeric organization of this novel enzyme appears potentially important for function, with a large, flexible loop reaching over from one protomer to contact the substrate and fully assemble the active site of the neighbouring molecule. There is no suggestion of such a dimerization occurring for GH3B, which also shows close matches to more typical monomeric family members including the family 3 β-glucosidases from *Thermatoga neapolitana* (PDBs: 2x42 and 2x41 with RMSDs of 1.49 Å and 1.50 Å, respectively, both over 715 residues) [68] and *Hypocrea jecorina*/*Trichoderma reesei* (PDBs: 4i8d and 3zyz with RMSDs of 1.42 and 1.50 Å over 711 and 713 residues, respectively) [69]. All of these structures share the same three-domain architecture as GH3B, though maximum identity is no more than 36% at the primary sequence level.

*Bo*GH3B was found to co-purify with glucose in its active site (figure 5*b*). This could readily be modelled with a $^4C_1$ chair conformation, highlighting the position of the −1 subsite. As is typical for hydrolytic GH3 members, the active site is formed largely by residues from the core TIM barrel, with additional interactions further contributed by loops from the α/β sandwich domain (figure 5*b*). GH3 members are well-known to employ the classical Koshland double-displacement, configuration-retaining mechanism [70]. Within

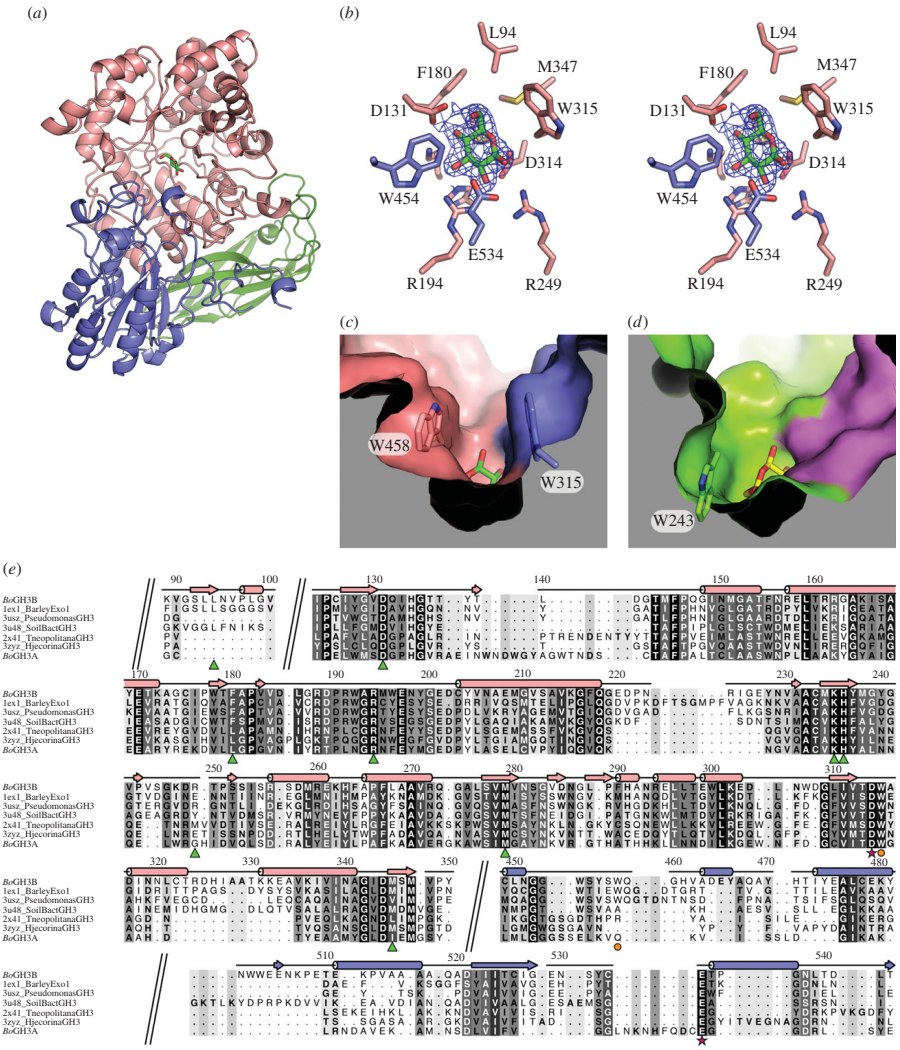


Figure 5. Structural analysis of *Bo*GH3B. (*a*) Overall structure of *Bo*GH3B with the N-terminal TIM-barrel-like domain coloured pink, the α/β sandwich domain coloured purple and the C-terminal FN-III like domain coloured green. The position of the active site is indicated by the presence of glucose, represented as sticks with green carbon atoms. (*b*) Wall-eyed stereo view of the active site in *Bo*GH3B. The final 2Fo-Fc map is shown contoured at 1σ as a blue mesh around the glucose molecule that co-purified with the protein. All side chains within 4 Å of glucose are shown as sticks with the carbons coloured according to the domain from which they are provided. (*c*) Surface representation of *Bo*GH3B around the active site pocket revealing a tight entry to the active site resulting from the presence of Trp315 and Trp458, the positions of which are indicated in stick representation. The surface is coloured by domain as for the previous panels and the surface has been clipped for clarity. See the electronic supplementary material, figure S1*a*, for stereo view. (*d*) Surface representation of the active site pocket from *T. neopolitana* GH3 (PDB 2x41), shown from the same perspective as for *Bo*GH3B in (*c*). The surface is coloured with the TIM-barrel domain shown in green and the α/β domain coloured magenta. This enzyme lacks an equivalent residue to Trp458, and the side chain of Trp243 (shown in stick representation) is positioned differently giving a much more open structure to the active site entrance. See the electronic supplementary material, figure S1*b*, for stereo view. (*e*) Structure-based sequence alignment of *Bo*GH3B with other GH3 family members identified through PDBeFold [56]—sections of the alignment have been removed for brevity and breaks are indicated by diagonal double lines across the alignment (the full sequence alignment can be found in the electronic supplementary material, figure S2). The sequence for *Bo*GH3A was added to the structural alignment using MAFFT [67]. The secondary structure elements (coloured by domain as previously) and residue numbers from *Bo*GH3B are indicated along the top of the alignment, with sequence similarity indicated by the shading behind the individual amino acids. Below the aligned sequences, residues lining the −1 sub-site are indicated with green triangles, the catalytic nucleophile and acid/base are indicated by magenta stars and tryptophan side chains narrowing the active site structure in *Bo*GH3B are shown with orange circles.

the GH3B active site, putative catalytic nucleophile (Asp314) and acid/base (Glu534) residues can be observed in close proximity to the glucose moiety, poised for nucleophilic attack. Together with residues forming the −1 sub-site, these interactions appear well conserved, and are maintained in several other GH3–glucose complexes [52,68,69]. Away from the −1 sub-site, the exterior surface structure of the GH3B active pocket deviates from the most closely related homologues, presenting as a more closed structure (figure 5*c*) similar to that seen in the distantly related barley β-glucosidase [52]. The barley enzyme shows quite narrow specificity for β-1,3- and β-1,4-linked glucans, while closer overall structural matches to *Bo*GH3B, including the *T. neapolitana* and *H. jecorina* enzymes described above, show much broader activities against β-1,2-, β-1,3-, β-1,4- and β-1,6-linked disaccharides [68,69]. Such promiscuous catalytic functionality has been suggested to result from the more open active site architecture maintained by this group, allowing diverse linkages and longer substrates to be accommodated (figure 5*d*) [68]. GH3B has significant activity for glucose-only oligosaccharides but displays far weaker activity on xyloglucan-derived oligos, which retain their xylose side chains [32]. Similar to barley β-glucosidase, such observations might suggest that the narrowing of the active site cleft could be responsible for the high specificity of *Bo*GH3B towards β-1,4-linked glucans.

Analysis of residues forming the GH3B +1 sub-site reveals more discernable differences between the two paralogous GH3 members in the *Bo*XyGUL. Sequence analysis suggests poor conservation of two aromatic residues, Trp315 and Trp458 (BoGH43B numbering), which through π-stacking interactions appear to form the narrow GH3B +1 sub-site. Although the equivalent to Trp315 is maintained in GH3A (Trp274), an equivalent to Trp458 appears absent. We hypothesize therefore that GH3A may present a more open active site architecture, leading to a similar rationale in the presence of two GH3 genes to that described above for the *Bo*XyGUL GH43 paralogues. The closed active site pocket in GH3B appears to result in higher affinity interactions with longer 'cello-oligosaccharides', suggesting that, as for the two *Bo*GH43 members, subtle differences in the active site architecture might confer adaptations to specific substrates. Again, such a proposal would thus provide a reasonable molecular basis for the maintenance of two highly similar genes in the same operon.

4. Conclusion

The absence, within the human genome, of genes encoding enzymes able to metabolize a significant proportion of the complex polysaccharides present in our own diet has thrown into sharp relief the importance of our internal microbial ecosystems [6,71]. The capacity of the gut microbiota to utilize these large, intractable molecules dictates both the composition and correct functioning of this large non-somatic dietary organ, and as such has a direct and crucial impact upon the health of the human host [72]. Recent systems biology approaches have highlighted the many niche roles played by diverse bacteria within the human microbiota [36–39]. While genomics and metagenomics initiatives continue apace, generating increasing amounts of sequence data, further approaches linking sequence data to biological function are essential to understanding the adaptations of individual species that allows them to fulfil their symbiotic role within the human digestive system. Xyloglucan degradation is a niche occupied primarily by the Bacteroidetes, and we have previously highlighted the importance of the specific XyGUL encoded by *B. ovatus* to allow this bacterium to compete for nutrients [32]. Central to this analysis was the tertiary structural characterization of the vanguard endo-xyloglucanase, *Bo*GH5, that catalyses the first backbone hydrolysis step required for xyloglucan polysaccharide metabolism. Recently, we have revealed the key role that two cell-surface glycan-binding proteins (SGBPs) encoded by the XyGUL play in XyG utilization through combined genetic, biophysical and crystallographic analyses [33].

Here, we have significantly extended our knowledge of the structural biology of the XyGUL through crystallography of several exo-glycosidases encoded by the *Bo*XyGUL. This analysis provided insight into the structural features within these enzymes that allow them to interact with and degrade their xyloglucan oligosaccharide substrates. Furthermore, our analysis highlights differences in the structures of two GH43 proteins, which display similar biochemical properties but are maintained within the operon nonetheless. Such observations suggest that these paralogues may play subtly different roles during the degradation of xyloglucans from different sources, or may function most optimally at different stages in the catabolism of XyGOs, for example before or after hydrolysis of certain side-chain moieties. While we were unable to determine a structure for *Bo*GH3A, our structural and sequence analysis of *Bo*GH3B has also allowed us to highlight further potential differences between these two enzymes encoded by the operon. Together with existing biochemical data, our analyses of the three-dimensional structures, and various enzyme-inhibitor complexes, of *Bo*GH31, *Bo*GH43A, *Bo*GH43B and *Bo*GH3B provide molecular-level insight into the stepwise breakdown of xyloglucan by the *Bo*XyGUL. Characterization of key adaptions within these enzymes provides a firm rationale for alternate specificities for XyGOs that may also allow for more efficient degradation of xyloglucan from different sources within the gut.

Data accessibility. All structures and accompanying structure factors have been deposited with the Protein Data Bank (PDB) with accession codes 5JOU, 5JOV, 5JOW, 5JOX, 5JOY, 5JOZ and 5JP0. Individual ITC thermograms and NMR spectra can be found in the electronic supplementary material.

Authors' contributions. G.R.H and A.J.T performed experiments and analysed data. J.S. and Ł.F.S. performed additional cloning and purification and some ITC, respectively. T.C. synthesized arabinofuranosidase inhibitors under the supervision of K.A.S.E.D.G.B. synthesised AraDNJ. J.L. and O.S. performed primary gene cloning and recombinant enzyme production. H.B. and G.J.D. directed the research. The manuscript was written by G.R.H, H.B., A.J.T. and G.J.D. with contributions from all authors.

Competing interests. We declare we have no competing interests.

Funding. Work in the Davies group was supported by the BBSRC (grant no. BB/I014802/1), L.S. is supported by the European Research Council proposal No. 322942—'GlycoPOISE'. Work in the Brumer group during the course of this project was supported by the Mizutani Foundation for Glycoscience, The Swedish Research Council Formas (via CarboMat—the KTH Advanced Carbohydrate Materials Centre), The Swedish Research Council (Vetenskapsrådet), the Knut and Alice Wallenberg Foundation (via the Wallenberg Wood Science Centre), faculty funding from the University of British Columbia, the Natural Sciences and Engineering Research Council of Canada (Discovery Grant), the Canada Foundation for Innovation and the British Columbia Knowledge Development Fund, and the Canadian Institutes for

Health Research (MOP-137134, MOP-142472). Support for this work by the Australian Research Council (K.A.S.), the Australian Government, the University of Western Australia, and the Centre for Microscopy, Characterisation and Analysis at the University of Western Australia (T.C.) is also acknowledged.

Acknowledgements. We thank Diamond Light Source for access to beamlines I02, I03, I04 and I04-1 (proposal nos. mx-7864 and mx-9948) that contributed to the results presented here. We also gratefully acknowledge Johan Turkenburg and Sam Hart for their assistance with synchrotron X-ray data collection.

References

1. Food and Agriculture Organization of the United Nations. 1998 *Carbohydrates in human nutrition*. (FAO Food and Nutrition Paper—66, available at URL http://www.fao.org/docrep/w8079e/w8079e00.htm). Rome, Italy: Food and Agriculture Organization.
2. Mann J *et al.* 2007 FAO/WHO Scientific Update on carbohydrates in human nutrition: conclusions. *Eur. J. Clin. Nutr.* **61**, S132–S137. (doi:10.1038/sj.ejcn.1602943)
3. Cummings JH, Stephen AM. 2007 Carbohydrate terminology and classification. *Eur. J. Clin. Nutr.* **61**, S5–S18. (doi:10.1038/sj.ejcn.1602936)
4. Cummings JH, Mann JI, Nishida C, Vorster HH. 2009 Dietary fibre: an agreed definition. *Lancet* **373**, 365–366. (doi:10.1016/S0140-6736(09)60117-3)
5. Hamaker BR, Tuncil YE. 2014 A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *J. Mol. Biol.* **426**, 3838–3850. (doi:10.1016/j.jmb.2014.07.028)
6. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. 2013 The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol* **11**, 497–504. (doi:10.1038/nrmicro3050)
7. McNeil NI. 1984 The contribution of the large-intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39**, 338–342.
8. Elia M, Cummings JH. 2007 Physiological aspects of energy metabolism and gastrointestinal effects of carbohydrates. *Eur. J. Clin. Nutr.* **61**, S40–S74. (doi:10.1038/sj.ejcn.1602938)
9. Scott KP, Gratz SW, Sheridan PO, Flint HJ, Duncan SH. 2013 The influence of diet on the gut microbiota. *Pharmacol. Res.* **69**, 52–60. (doi:10.1016/j.phrs.2012.10.020)
10. Arrieta MC, Stiemsma LT, Amenyogbe N, Brown EM, Finlay B. 2014 The intestinal microbiome in early life: health and disease. *Front. Immunol.* **5**, 18. (doi:10.3389/fimmu.2014.00427)
11. Biedermann L, Rogler G. 2015 The intestinal microbiota: its role in health and disease. *Eur. J. Clin. Pediatr.* **174**, 151–167. (doi:10.1007/s00431-014-2476-2)
12. Schwabe RF, Jobin C. 2013 The microbiome and cancer. *Nat. Rev. Cancer* **13**, 800–812. (doi:10.1038/nrc3610)
13. Vangay P, Ward T, Gerber JS, Knights D. 2015 Antibiotics, pediatric dysbiosis, and disease. *Cell Host Microbe* **17**, 553–564. (doi:10.1016/j.chom.2015.04.006)
14. Borgia G, Maraolo AE, Foggia M, Buonomo AR, Gentile I. 2015 Fecal microbiota transplantation for *Clostridium difficile* infection: back to the future. *Expert Opin. Biol. Ther.* **15**, 1001–1014. (doi:10.1517/14712598.2015.1045872)
15. David LA *et al.* 2014 Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559. (doi:10.1038/nature12820)
16. Koropatkin NM, Cameron EA, Martens EC. 2012 How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335. (doi:10.1038/nrmicro2746)
17. McNulty NP *et al.* 2013 Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11**, e1001637. (doi:10.1371/journal.pbio.1001637)
18. Walker AW, Duncan SH, Louis P, Flint HJ. 2014 Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol.* **22**, 267–274. (doi:10.1016/j.tim.2014.03.001)
19. Martens EC *et al.* 2011 Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, 16. (doi:10.1371/journal.pbio.1001221)
20. Rogers TE, Pudlo NA, Koropatkin NM, Bell JSK, Balasch MM, Jasker K, Martens EC. 2013 Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures. *Mol. Microbiol.* **88**, 876–890. (doi:10.1111/mmi.12228)
21. Raghavan V, Groisman EA. 2015 Species-specific dynamic responses of gut bacteria to a mammalian glycan. *J. Bacteriol.* **197**, 1538–1548. (doi:10.1128/jb.00010-15)
22. Martens EC, Chiang HC, Gordon JI. 2008 Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457. (doi:10.1016/j.chom.2008.09.007)
23. Martens EC, Kelly AG, Tauzin AS, Brumer H. 2014 The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J. Mol. Biol.* **426**, 3851–3865. (doi:10.1016/j.jmb.2014.06.022)
24. Gordon JI. 2012 Honor thy gut symbionts redux. *Science* **336**, 1251–1253. (doi:10.1126/science.1224686)
25. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014 The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495. (doi:10.1093/nar/gkt1178)
26. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012 Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230. (doi:10.1038/nature11550)
27. Salyers AA, Vercellotti JR, West SEH, Wilkins TD. 1977 Fermentation of mucin and plant polysaccharides by strains of bacteroides from human colon. *Appl. Environ. Microbiol.* **33**, 319–322.
28. Wegmann U, Louis P, Goesmann A, Henrissat B, Duncan SH, Flint HJ. 2013 Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('*Ruminococcus bicirculans*') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ. Microbiol.* **16**, 2879–2890. (doi:10.1111/1462-2920.12217)
29. Hemsworth GR, Dejean G, Davies GJ, Brumer H. 2016 Learning from microbial strategies for polysaccharide degradation. *Biochem. Soc. Trans.* **44**, 94–108. (doi:10.1042/BST20150180)
30. Terrapon N, Lombard V, Gilbert HJ, Henrissat B. 2015 Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647–655. (doi:10.1093/bioinformatics/btu716)
31. Cameron EA, Kwiatkowski KJ, Lee BH, Hamaker BR, Koropatkin NM, Martens EC. 2014 Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. *Mbio* **5**, 12. (doi:10.1128/mBio.01441-14)
32. Larsbrink J *et al.* 2014 A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498–502. (doi:10.1038/nature12907)
33. Tauzin AS, Kwiatkowski KJ, Orlovsky NI, Smith CJ, Creagh AL, Haynes CA, Wawrzak Z, Brumer H, Koropatkin NM. 2016 Molecular dissection of xyloglucan recognition in a prominent human gut symbiont. *MBio* **7**, e02134-15. (doi:10.1128/mBio.02134-15)
34. Tuomivaara ST, Yaoi K, O'Neill MA, York WS. 2015 Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature. *Carbohydr. Res.* **402**, 56–66. (doi:10.1016/j.carres.2014.06.031)
35. Varki A *et al.* 2015 *Essentials of glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
36. Sonnenburg ED, Zheng HJ, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL. 2010 Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, U1241–U1256. (doi:10.1016/j.cell.2010.05.005)

37. Hehemann JH, Kelly AG, Pudlo NA, Martens EC, Boraston AB. 2012 Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19 786–19 791. (doi:10.1073/pnas.1211002109)
38. Cuskin F *et al.* 2015 Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517**, U165–U186. (doi:10.1038/nature13995)
39. Rogowski A *et al.* 2015 Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 15. (doi:10.1038/ncomms8481)
40. Fogg MJ, Wilkinson AJ. 2008 Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.* **36**, 771–775. (doi:10.1042/BST0360771)
41. McCarter JD, Withers SG. 1996 Unequivocal identification of Asp-214 as the catalytic nucleophile of *Saccharomyces cerevisiae* alpha-glucosidase using 5-fluoro glycosyl fluorides. *J. Biol. Chem.* **271**, 6889–6894. (doi:10.1074/jbc.271.12.6889)
42. McCarter JD, Withers SG. 1996 5-Fluoro glycosides: a new class of mechanism-based inhibitors of both α- and β-glucosidases. *J. Am. Chem. Soc.* **118**, 241–242. (doi:10.1021/ja952732a)
43. Kabsch W. 2010 XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132. (doi:10.1107/S0907444909047337)
44. Winn MD *et al.* 2011 Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242. (doi:10.1107/S0907444910045749)
45. Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008 Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179. (doi:10.1038/nprot.2008.91)
46. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010 Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501. (doi:10.1107/S0907444910007493)
47. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011 REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367. (doi:10.1107/S0907444911001314)
48. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007 Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674. (doi:10.1107/S0021889807021206)
49. Cowtan K. 2008 Fitting molecular fragments into electron density. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 83–89. (doi:10.1107/S0907444907033938)
50. Cowtan K. 2006 The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011. (doi:10.1107/S0907444906022116)
51. Cowtan K. 2010 Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478. (doi:10.1107/S0907444909039947X)
52. Varghese JN, Hrmova M, Fincher GB. 1999 Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* **7**, 179–190. (doi:10.1016/S0969-2126(99)80024-0)
53. Zhao M, Wang Y, Huo C, Li C, Zhang X, Peng L, Peng S. 2009 Stereoselective synthesis of novel N-(α-L-arabinofuranos-1-yl)-L-amino acids. *Tetrahedron Asymmetry* **20**, 247–258. (doi:10.1016/j.tetasy.2009.01.014)
54. Turnbull WB, Daranas AH. 2003 On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **125**, 14 859–14 866. (doi:10.1021/ja036166s)
55. Larsbrink J, Izumi A, Ibatullin FM, Nakhai A, Gilbert HJ, Davies GJ, Brumer H. 2011 Structural and enzymatic characterization of a glycoside hydrolase family 31 α-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochem. J.* **436**, 567–580. (doi:10.1042/BJ20110299)
56. Krissinel E, Henrick K. 2004 Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268. (doi:10.1107/S0907444904026460)
57. Lovering AL, Lee SS, Kim Y-W, Withers SG, Strynadka NCJ. 2005 Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J. Biol. Chem.* **280**, 2105–2115. (doi:10.1074/jbc.M410468200)
58. Silipo A, Larsbrink J, Marchetti R, Lanzetta R, Brumer H, Molinaro A. 2012 NMR spectroscopic analysis reveals extensive binding interactions of complex xyloglucan oligosaccharides with the *Cellvibrio japonicus* glycoside hydrolase family 31 alpha-xylosidase. *Chemistry* **18**, 13 395–13 404. (doi:10.1002/chem.201200488)
59. Larsbrink J, Izumi A, Hemsworth GR, Davies GJ, Brumer H. 2012 Structural enzymology of *Cellvibrio japonicus* Agd31B protein reveals α-transglucosylase activity in glycoside hydrolase family 31. *J. Biol. Chem.* **287**, 43 288–43 299. (doi:10.1074/jbc.M112.416511)
60. Gloster TM, Davies GJ. 2010 Glycosidase inhibition: assessing mimicry of the transition state. *Org. Biomol. Chem.* **8**, 305–320. (doi:10.1039/b915870g)
61. Fujimoto Z, Ichinose H, Maehara T, Honda M, Kitaoka M, Kaneko S. 2010 Crystal structure of an exo-1,5-α-L-arabinofuranosidase from *Streptomyces avermitilis* provides insights into the mechanism of substrate discrimination between exo- and endo-type enzymes in glycoside hydrolase family 43. *J. Biol. Chem.* **285**, 34 134–34 143. (doi:10.1074/jbc.M110.164251)
62. Jones DWC, Nash RJ, Bell EA, Williams JM. 1985 Identification of the 2-hydroxymethyl-3,4-dihydroxypyrrolidine (or 1,4-dideoxy-1,4-iminopentitol) from *Angylocalyx boutiqueanus* and from *Arachniodes standishii* as the (2R, 3R, 4S)-isomer by the synthesis of its enantiomer. *Tetrahedron Lett.* **26**, 3125–3126. (doi:10.1016/S0040-4039(00)98635-0)
63. Brüx C, Ben-David A, Shallom-Shezifi D, Leon M, Niefind K, Shoham G, Shoham Y, Schomburg D. 2006 The structure of an inverting GH43 beta-xylosidase from *Geobacillus stearothermophilus* with its substrate reveals the role of the three catalytic residues. *J. Mol. Biol.* **359**, 97–109. (doi:10.1016/j.jmb.2006.03.005)
64. Lee CC, Braker JD, Grigorescu AA, Wagschal K, Jordan DB. 2013 Divalent metal activation of a GH43 β-xylosidase. *Enzyme Microb. Technol.* **52**, 84–90. (doi:10.1016/j.enzmictec.2012.10.010)
65. Jordan DB, Lee CC, Wagschal K, Braker JD. 2013 Activation of a GH43 β-xylosidase by divalent metal cations: slow binding of divalent metal and high substrate specificity. *Arch. Biochem. Biophys.* **533**, 79–87. (doi:10.1016/j.abb.2013.02.020)
66. Hassan N, Kori LD, Gandini R, Patel BKC, Divne C, Tan TC. 2015 High-resolution crystal structure of a polyextreme GH43 glycosidase from *Halothermothrix orenii* with α-L-arabinofuranosidase activity. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **71**, 338–345. (doi:10.1107/S2053230X15003337)
67. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
68. Pozzo T, Pasten JL, Karlsson EN, Logan DT. 2010 Structural and functional analyses of beta-glucosidase 3B from *Thermotoga neapolitana*: a thermostable three-domain representative of glycoside hydrolase 3. *J. Mol. Biol.* **397**, 724–739. (doi:10.1016/j.jmb.2010.01.072)
69. Karkehabadi S *et al.* 2014 Biochemical characterization and crystal structures of a fungal family 3 β-glucosidase, Cel3A from *Hypocrea jecorina*. *J. Biol. Chem.* **289**, 31 624–31 637. (doi:10.1074/jbc.M114.587766)
70. Koshland DE. 1953 Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436. (doi:10.1111/j.1469-185X.1953.tb01386.x)
71. Tasse L *et al.* 2010 Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* **20**, 1605–1612. (doi:10.1101/gr.108332.110)
72. Lattimer JM, Haub MD. 2010 Effects of dietary fiber and its components on metabolic health. *Nutrients* **2**, 1266–1289. (doi:10.3390/nu2121266)

List of Abbreviations

$2mF_o$-DF_c	Maximum likelihood (m, figure of merit)-weighted and σ_A(D)-weighted $2F_o$-F_c electron density map
aa	Amino acid(s)
AA	Auxillary activity
Abs_{280}	Absorbance at 280 nm
Abs_{340}	Absorbance at 340 nm
Ac	Acetate
ALG5	Asparagine-linked glycosylation 5 homolog
Amp^+	With a working concentration of ampicillin added
ASU	Asymmetric unit
BCD	Bromoconduritol
B factor	Debye-Waller factor
BHK	Baby hamster kidney
BiP	Binding immunoglobulin protein
Bt	*Bacteroides thetaiotaomicron*
BVDV	Bovine viral diarrhea virus
Bx	*Bacteroides xylanisolvens*
Cam^+	With a working concentration of chloramphenicol added
CBM	Carbohydrate-binding module
CC	Correlation coefficient
CD4	Cluster of differentiation 4
CDS	Coding DNA sequence
CE	Carbohydrate esterase
CFU	Colony forming unit
Cg	*Cricetulus griseus*
CHO	Chinese hamster ovary
CIP	Cocaine-induced paranoia

Cm^r	Chloramphenicol resistance gene
CNS	Central nervous system
CNX	Calnexin
COPII	Coat protein II
CRT	Calreticulin
CspA	Cold shock protein CspA
CST	Castanospermine
d	Deoxy
Da	Dalton
DAPI	4′,6-Diamidino-2-phenylindole
dd	Dideoxy
DLS	Diamond Light Source
DMJ	1-Deoxymannojirimycin
DNA	2-Deoxyribonucleic acid
DNJ	1-Deoxynojirimycin
dNTP	Deoxiribonucleotide triphosphate
Dol	Dolichol
DOLK	Dolichol kinase
DP	Differential power
DPAGT1	Dolichyl-phosphate *N*-acetylglucosaminephosphotransferase 1
DPM	Dolichol-P-mannose synthase
DPM1	Dolichyl-phosphate mannosyltransferase subunit 1, catalytic
DTT	1,4-Dithiothreitol
EDEM	ER degradation enhancing alpha-mannosidase like protein
EGFP	Enhanced green fluorescent protein
Env	Envelope
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated degradation
ERGIC	ER-Golgi intermediate compartment
ERLEC1	Endoplasmic reticulum lectin 1
ERQC	Endoplasmic reticulum quality control
ESI	Electrospray ionization
FEL	Free energy landscape
FFU	Focus forming units

FOS	Free oligosaccharides
fs	Frameshift
× g	Multiples of gravitational acceleration on the Earth's surface
Gal	D-Galactose
GalNAc	*N*-acetylgalactosamine
GANAB	Glucosidase II alpha subunit
GDP	Guanosine diphosphate
GF	Gel filtration
GH	Glycoside hydrolase
Glc	D-Glucose
GlcNAc	*N*-acetylglucosamine
GTEx	Genotype-Tissue Expression database
GnXMXGn2	$GlcNAc_XMan_XGlcNAc_2$
GXMXGn2	$Glc_XMan_XGlcNAc_2$
GXMX	Glc_XMan_X
G protein	Glycoprotein (viral protein, not related to G protein coupled to transmembrane receptors)
GFP	Green fluorescent protein
GPI	Glycosylphosphatidyinositol
GT	Glycosyltransferase
GWAS	Genome-wide association study
HepC	Hepatitis C
HCV	Hepatitis C virus
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HEPPS	3-[4-(2-Hydroxyethyl)piperazin-1-yl]propane-1-sulfonic acid
HIV	Human immunodeficiency virus
Hs	*Homo sapiens*
HSP70	70-kDa heat shock protein
HSPA5	Heat shock protein family A member 5
IC_{50}	Inhibitor concentration at which an enzyme catalyzes the reaction at 50% of the rate of the uninhibited reaction
IFG	Isofagomine
iGWAS	Integrative genome-wide association study
Im	Imidazole

IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
IUB	International Union of Biochemistry
IUPAC	International Union of Pure and Applied Chemistry
Kan	Kanamycin
Kan^+	With a working concentration of kanamycin added
kd	Knockdown
KIE	Kinetic isotope effects
KIF	Kifunensine
ko	Knockout
LB	Lysogeny broth
LG	Leaving group
LUCA	Last universal common ancestor
M	Molar
MagT1	Magnesium transporter 1
Man	D-Mannose
MAN1A1	mannosidaseGH alpha class 1A member 1
MAN1A2	mannosidase alpha class 1A member 2
MAN1B1	mannosidase alpha class 1B member 1
MAN1C1	mannosidase alpha class 1C member 1
MANEA	Mannosidase endo-alpha
MANEAL	Mannosidase endo-alpha-like
ManIm	Mannoimidazole
MAP1LC3	Microtubule associated protein 1 light chain 3
MCS	Multiple cloning site
MDBK	Madin-Darby bovine kidney
MDCK	Madin-Darby canine kidney
Me	Methyl
MES	2-(*N*-morpholino)ethanesulfonic acid
MGAT1	Mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase
MIB	Malonate/imidazole/boric acid (buffer system)
MOGS	Mannosyl-oligosaccharide glucosidase
MPDU1	Mannose-P-dolichol utilization defect 1

MOPSO	2-Hydroxy-3-morpholinopropanesulfonic acid
MQ	Ultrapure
MS	Mass spectrometry
MUF	4-Methylumbelliferone
Mw	Molecular weight
MX	Man_X
MXGn2	$Man_XGlcNAc_2$
N	Number of sites (in ITC)
NAD	Nicotinamide adenine dinucleotide
*N*B-DNJ	*N*-butyl-deoxynojirimycin
NCS	Non-crystallographic symmetry
ND	Not determined
NGS	Next-Generation Sequencing
NOE	Noeuromycin
NS5A	Non-structural protein 5A
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
OS9	OS9, endoplasmic reticulum lectin
OST	Oligosaccharyltransferase
P	Phosphate
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PL	Polysaccharide lyase
PP	Pyrophosphate
PEG	Polyethylene glycol
PEG-MME	Polyethylene glycol monomethyl ether
PNGase	Peptide:N-glycosidase
ppm	Parts per million
PRKCSH	Protein kinase C substrate 80K-H
PtK1	*Potorous tridactylus* kidney
PTM	Post-translational modification
QCV	Quality control vesicle
R	Radical
Ref.	Reference

RFT1	RFT1 (requiring fifty three 1) homolog
RMS	Root mean square
RMSD	Root mean square difference or root mean square deviation
RPKM	Reads per kilobase per million
RPM	Revolutions per minute
RT	Room temperature
SARS-CoV	Severe acute respiratory syndrome coronavirus
SNP	Single-nucleotide polymorphism
SOC	Super-optimal broth with catabolite repression
STT3	Catalytic subunit of the oligosaccharyltransferase complex
SW	Swainsonine
TEW	Anderson–Evans polyoxotungstate $[TeW_6O_{24}]^{6-}$
TGN	*trans*-Golgi network
TLC	Thin layer chromatography
TPM	Transcripts per million
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
TS	Transition state
TSS	Transformation and storage solution
Tusc3	Tumor suppressor candidate 3
UDP	Uridine diphosphate
UTR	Untranslated region
V3	Variable 3
VSV	Vesicular stomatitis virus
WT	Wild-type
XTP3-B	XTP3-transactivated gene B protein
Xyl	D-Xylose
y	year(s)
ΔH	Change in enthalpy
ΔS	Change in entropy
Δ#	N-terminal truncation before amino acid number #

References

1. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126,** 855–867 (2006).

2. Varki, A. Biological roles of glycans. *Glycobiology* **27,** 3–49 (2017).

3. Hanson, S. R. *et al.* The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proceedings of the National Academy of Sciences* **106,** 3131–3136 (2009).

4. Inatani, M., Irie, F., Plump, A. S., Tessier-Lavigne, M. & Yamaguchi, Y. Mammalian Brain Morphogenesis and Midline Axon Guidance Require Heparan Sulfate. *Science* **302,** 1044–1046 (2003).

5. Wybenga, L. E. *et al.* Glycophorin as a Receptor for Sendai Virus. *Biochemistry* **35,** 9513–9518 (1996).

6. Gagneux, P. & Varki, A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9,** 747–755 (1999).

7. Morrow, A. L. *et al.* Human milk oligosaccharides are associated with protection against diarrhea in breast-fed infants. *The Journal of Pediatrics* **145,** 297–303 (2004).

8. Prigozy, T. I. *et al.* The Mannose Receptor Delivers Lipoglycan Antigens to Endosomes for Presentation to T Cells by CD1b Molecules. *Immunity* **6,** 187–197 (1997).

9. Yu, L.-C., Lee, H.-L., Chan, Y.-S. & Lin, M. The Molecular Basis for the B(A) Allele: An Amino Acid Alteration in the Human Histoblood Group B α-(1,3)-Galactosyltransferase Increases Its Intrinsic α-(1,3)-N-Acetylgalactosaminyltransferase Activity. *Biochemical and Biophysical Research Communications* **262,** 487–493 (1999).

10. Carlin, A. F. *et al.* Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* **113,** 3333–3336 (2009).

11. Olsnes, S. & Pihl, A. Different biological properties of the two constituent peptide chains of ricin a toxic protein inhibiting protein synthesis. *Biochemistry* **12,** 3121–3126 (1973).

12. Stowell, S. R. *et al.* Innate immune lectins kill bacteria expressing blood group antigen. *Nature Medicine* **16,** 295–301 (2010).

13. Spiro, R. G. Protein glycosylation: Nature, distribution, enzymatic formation, and

disease implications of glycopeptide bonds. *Glycobiology* **12,** 43R–56R (2002).

14. Sharon, N. The conquest of the last frontier of molecular and cell biology. *Biochimie* **83,** 555 (2001).

15. Miljković, M. Introduction. in *Carbohydrates - Synthesis, Mechanisms, and Stereoelectronic Effects* 1–25 (Springer, New York, NY, 2010). doi:10.1007/978-0-387-92265-2_1

16. Steinhardt, R. G., Calvin, A. D. & Dodd, E. A. Taste-Structure Correlation with α-D-Mannose and β-D-Mannose. *Science* **135,** 367–368 (1962).

17. Kendrew, J. C. & Moelwyn-Hughes, E. A. The Kinetics of Mutarotation in Solution. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **176,** 352–367 (1940).

18. Miljković, M. Conformational Analysis of Monosaccharides. in *Carbohydrates - Synthesis, Mechanisms, and Stereoelectronic Effects* 27–56 (Springer New York, 2010).

19. Juaristi, E. & Cuevas, G. Recent studies of the anomeric effect. *Tetrahedron* **48,** 5019–5087 (1992).

20. Huang, Y., Zhong, A.-G., Yang, Q. & Liu, S. Origin of anomeric effect: A density functional steric analysis. *The Journal of Chemical Physics* **134,** 084103 (2011).

21. Zhou, X., Yu, D., Rong, C., Lu, T. & Liu, S. Anomeric effect revisited: Perspective from information-theoretic approach in density functional reactivity theory. *Chemical Physics Letters* **684,** 97–102 (2017).

22. Stewart, R. A., Carrico, C. K., Webster, R. L. & Steinhardt, R. G. Physicochemical Stereospecificity in Taste Perception of α-D-Mannose and β-D-Mannose. *Nature* **234,** 220–220 (1971).

23. Hayes, M. R. & Pietruszka, J. Synthesis of Glycosides by Glycosynthases. *Molecules* **22,** 1434 (2017).

24. Schuman, B., Evans, S. V. & Fyles, T. M. Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *PLoS ONE* **8,** (2013).

25. Lairson, L., Henrissat, B., Davies, G. & Withers, S. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry* **77,** 521–555 (2008).

26. Cremer, D. & Pople, J. A. General definition of ring puckering coordinates. *Journal of the American Chemical Society* **97,** 1354–1358 (1975).

27. Davies, G. J., Planas, A. & Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. *Accounts of Chemical Research* **45,** 308–316 (2012).

28. Ardèvol, A., Biarnés, X., Planas, A. & Rovira, C. The Conformational Free-Energy Landscape of β-d-Mannopyranose: Evidence for a 1S5 → B2,5 → OS2 Catalytic Itinerary in β-Mannosidases. *Journal of the American Chemical Society* **132,** 16058–16065 (2010).

29. Males, A., Raich, L., Williams, S. J., Rovira, C. & Davies, G. J. Conformational Analysis of the Mannosidase Inhibitor Kifunensine: A Quantum Mechanical and Struc-

tural Approach. *ChemBioChem* **18,** 1496–1501 (2017).

30. Thompson, A. J. *et al.* The Reaction Coordinate of a Bacterial GH47 α-Mannosidase: A Combined Quantum Mechanical and Structural Approach. *Angewandte Chemie International Edition* **51,** 10997–11001 (2012).

31. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L. & Mornon, J. P. Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81,** 83–95 (1989).

32. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **280,** 309–316 (1991).

33. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Research* **37,** D233–D238 (2009).

34. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42,** D490–D495 (2014).

35. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature and Symbolism for Amino Acids and Peptides. *European Journal of Biochemistry* **138,** 9–37 (1984).

36. Henrissat, B. *et al.* Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 7090–7094 (1995).

37. Beer, T. de, Vliegenthart, J. F. G., Loeffler, A. & Hofsteenge, J. The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase Us Is .alpha.-Mannopyranose. *Biochemistry* **34,** 11785–11789 (1995).

38. Ihara, Y. *et al.* C-Mannosylation: A Modification on Tryptophan in Cellular Proteins. in *Glycoscience: Biology and Medicine* (eds. Endo, T., Seeberger, P. H., Hart, G. W., Wong, C.-H. & Taniguchi, N.) 1–8 (Springer Japan, 2014). doi:10.1007/978-4-431-54836-2_67-1

39. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 8256–8261 (2008).

40. Roth, J. & Zuber, C. Quality control of glycoprotein folding and ERAD: The role of N-glycan handling, EDEM1 and OS-9. *Histochemistry and Cell Biology* **147,** 269–284 (2017).

41. Bause, E. & Legler, G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochemical Journal* **195,** 639–644 (1981).

42. Hubbard, S. C. & Ivatt, R. J. Synthesis and Processing of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry* **50,** 555–583 (1981).

43. Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides.

Annual Review of Biochemistry **54,** 631–664 (1985).

44. Lederkremer, G. Z. Glycoprotein folding, quality control and ER-associated degradation. *Current Opinion in Structural Biology* **19,** 515–523 (2009).

45. Stanley, P. Golgi Glycosylation. *Cold Spring Harbor Perspectives in Biology* **3,** a005199 (2011).

46. Aebi, M. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833,** 2430–2437 (2013).

47. Lannoo, N. & Van Damme, E. J. Review/N-glycans: The making of a varied toolbox. *Plant Science* **239,** 67–83 (2015).

48. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Current Opinion in Cell Biology* **41,** 57–65 (2016).

49. Pavy, F. W. VI. The glucoside constitution of proteid matter. *Proceedings of the Royal Society of London* **54,** 53–57 (1894).

50. Rosevear, J. W. & Smith, E. L. STRUCTURE OF GLYCOPEPTIDES FROM A HUMAN γ-GLOBULIN. *Journal of the American Chemical Society* **80,** 250–251 (1958).

51. Johansen, P. G., Marshall, R. D. & Neuberger, A. Carbohydrates in protein. 3. The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *Biochemical Journal* **78,** 518–527 (1961).

52. Hubbard, S. C. & Robbins, P. W. Synthesis and processing of protein-linked oligosaccharides in vivo. *J. Biol. Chem* **254,** 4568–4576 (1979).

53. Lehle, L., Schulz, I. & Tanner, W. Dolichyl phosphate linked sugars as intermediates in the synthesis of yeast mannoproteins: An in vivo study. *Archives of Microbiology* **127,** 231–237 (1980).

54. Trimble, R. B., Maley, F. & Tarentino, A. L. Characterization of large oligosaccharide-lipids synthesized in vitro by microsomes from Saccharomyces cerevisiae. *Journal of Biological Chemistry* **255,** 10232–10238 (1980).

55. Jelk, J. *et al.* Glycoprotein Biosynthesis in a Eukaryote Lacking the Membrane Protein Rft1. *Journal of Biological Chemistry* **288,** 20616–20623 (2013).

56. Ashida, H., Maeda, Y. & Kinoshita, T. DPM1, the Catalytic Subunit of Dolichol-phosphate Mannose Synthase, Is Tethered to and Stabilized on the Endoplasmic Reticulum Membrane by DPM3. *Journal of Biological Chemistry* **281,** 896–904 (2006).

57. Kranz, C. *et al.* A mutation in the human *MPDU1* gene causes congenital disorder of glycosylation type If (CDG-If). *The Journal of Clinical Investigation* **108,** 1613–1619 (2001).

58. Schenk, B. *et al. MPDU1* mutations underlie a novel human congenital disorder of glycosylation, designated type If. *The Journal of Clinical Investigation* **108,** 1687–1695 (2001).

59. Haeuptle, M. A. & Hennet, T. Congenital disorders of glycosylation: An update

on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Human Mutation* **30,** 1628–1641 (2009).

60. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nature Communications* **4,** 2627 (2013).

61. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).

62. Pfeffer, S. *et al.* Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nature Communications* **5,** 3072 (2014).

63. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nature Structural & Molecular Biology* **24,** 1100–1106 (2017).

64. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science* eaar5140 (2018). doi:10.1126/science.aar5140

65. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica Et Biophysica Acta* **1473,** 4–8 (1999).

66. Zielinska, D. F., Gnad, F., Wiśniewski, J. R. & Mann, M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **141,** 897–907 (2010).

67. Harada, Y., Li, H., Li, H. & Lennarz, W. J. Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 6945–6949 (2009).

68. Kelleher, D. J., Karaoglu, D., Mandon, E. C. & Gilmore, R. Oligosaccharyltransferase Isoforms that Contain Different Catalytic STT3 Subunits Have Distinct Enzymatic Properties. *Molecular Cell* **12,** 101–111 (2003).

69. Shrimal, S., Trueman, S. F. & Gilmore, R. Extreme C-terminal sites are posttranslocationally glycosylated by the STT3B isoform of the OST. *J Cell Biol* **201,** 81–95 (2013).

70. Tai, V. W.-F. & Imperiali, B. Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase. *The Journal of Organic Chemistry* **66,** 6217–6228 (2001).

71. Molinari, M. & Helenius, A. Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **402,** 90–93 (1999).

72. Shrimal, S., Cherepanova, N. A. & Gilmore, R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell & Developmental Biology* **41,** 71–78 (2015).

73. Pless, D. D. & Lennarz, W. J. Enzymatic conversion of proteins to glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **74,** 134–138 (1977).

74. Tomme, P., Kwan, E., Gilkes, N. R., Kilburn, D. G. & Warren, R. A. Characterization of CenC, an enzyme from Cellulomonas fimi with both endo- and exoglucanase

activities. *Journal of Bacteriology* **178,** 4216–4223 (1996).

75. Shailubhai, K., Pukazhenthi, B. S., Saxena, E. S., Varma, G. M. & Vijay, I. K. Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *Journal of Biological Chemistry* **266,** 16587–16593 (1991).

76. Schweden, J., Borgmann, C., Legler, G. & Bause, E. Characterization of calf liver glucosidase I and its inhibition by basic sugar analogs. *Archives of biochemistry and biophysics* **248,** 335–340 (1986).

77. Bause, E., Erkens, R., Schweden, J. & Jaenicke, L. Purification and characterization of trimming glucosidase I from Saccharomyces cerevisiae. *FEBS Letters* **206,** 208–212 (1986).

78. Barker, M. K. & Rose, D. R. Specificity of Processing α-Glucosidase I Is Guided by the Substrate Conformation: Crystallographic and in silico Studies. *Journal of Biological Chemistry* **288,** 13563–13574 (2013).

79. Alarico, S., Empadinhas, N. & Costa, M. S. da. A new bacterial hydrolase specific for the compatible solutes α-D-mannopyranosyl-(1→2)-D-glycerate and α-D-glucopyranosyl-(1→2)-D-glycerate. *Enzyme and Microbial Technology* **52,** 77–83 (2013).

80. Miyazaki, T., Nishikawa, A. & Tonozuka, T. Crystal structure of the enzyme-product complex reveals sugar ring distortion during catalysis by family 63 inverting α-glycosidase. *Journal of Structural Biology* **196,** 479–486 (2016).

81. Deprez, P., Gautschi, M. & Helenius, A. More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle. *Molecular Cell* **19,** 183–195 (2005).

82. Jiang, J. *et al.* Detection of Active Mammalian GH31 α-Glucosidases in Health and Disease Using In-Class, Broad-Spectrum Activity-Based Probes. *ACS Central Science* **2,** 351–358 (2016).

83. Trombetta, E. S., Simons, J. F. & Helenius, A. Endoplasmic Reticulum Glucosidase II Is Composed of a Catalytic Subunit, Conserved from Yeast to Mammals, and a Tightly Bound Noncatalytic HDEL-containing Subunit. *Journal of Biological Chemistry* **271,** 27509–27516 (1996).

84. Koshland, D. E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews* **28,** 416–436 (1953).

85. Frandsen, T. P. & Svensson, B. Plant alpha-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. *Plant Molecular Biology* **37,** 1–13 (1998).

86. Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate Specificity Analysis of Endoplasmic Reticulum Glucosidase II Using Synthetic High Mannose-type Glycans. *Journal of Biological Chemistry* **281,** 31502–31508 (2006).

87. Molinari, M. & Helenius, A. Chaperone Selection During Glycoprotein Translo-

cation into the Endoplasmic Reticulum. *Science* **288,** 331–333 (2000).

88. Zhu, T., Satoh, T. & Kato, K. Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: Crystal structure of third thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase. *Scientific Reports* **4,** 7322 (2014).

89. Roversi, P. *et al.* Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proceedings of the National Academy of Sciences* **114,** 8544–8549 (2017).

90. Blanco-Herrera, F. *et al.* The UDP-glucose: Glycoprotein glucosyltransferase (UGGT), a key enzyme in ER quality control, plays a significant role in plant growth as well as biotic and abiotic stress in Arabidopsis thaliana. *BMC Plant Biology* **15,** 127 (2015).

91. Molinari, M., Galli, C., Vanoni, O., Arnold, S. M. & Kaufman, R. J. Persistent Glycoprotein Misfolding Activates the Glucosidase II/UGT1-Driven Calnexin Cycle to Delay Aggregation and Loss of Folding Competence. *Molecular Cell* **20,** 503–512 (2005).

92. Karaveg, K. *et al.* Mechanism of Class 1 (Glycosylhydrolase Family 47) α-Mannosidases Involved in N-Glycan Processing and Endoplasmic Reticulum Quality Control. *Journal of Biological Chemistry* **280,** 16197–16207 (2005).

93. Ninagawa, S. *et al.* EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step. *J Cell Biol* **206,** 347–356 (2014).

94. Benyair, R. *et al.* Mammalian ER mannosidase I resides in quality control vesicles, where it encounters its glycoprotein substrates. *Molecular Biology of the Cell* **26,** 172–184 (2015).

95. Ogen-Shtern, N., Avezov, E., Shenkman, M., Benyair, R. & Lederkremer, G. Z. Mannosidase IA is in Quality Control Vesicles and Participates in Glycoprotein Targeting to ERAD. *Journal of Molecular Biology* **428,** 3194–3205 (2016).

96. Velasco, A. *et al.* Cell type-dependent variations in the subcellular distribution of alpha-mannosidase I and II. *The Journal of Cell Biology* **122,** 39–51 (1993).

97. Tremblay, L. O. & Herscovics, A. Cloning and expression of a specific human α1, 2-mannosidase that trims Man9GlcNAc2 to Man8GlcNAc2 isomer B during N-glycan biosynthesis. *Glycobiology* **9,** 1073–1078 (1999).

98. Herscovics, A., Romero, P. A. & Tremblay, L. O. The specificity of the yeast and human class I ER alpha 1,2-mannosidases involved in ER quality control is not as strict previously reported. *Glycobiology* **12,** 14G–15G (2002).

99. Lal, A. *et al.* Substrate specificities of recombinant murine Golgi α1,2-mannosidases IA and IB and comparison with endoplasmic reticulum and Golgi processing α1,2-mannosidases. *Glycobiology* **8,** 981–995 (1998).

100. Xiang, Y., Karaveg, K. & Moremen, K. W. Substrate recognition and catalysis by GH47 α-mannosidases involved in Asn-linked glycan maturation in the mammalian

secretory pathway. *Proceedings of the National Academy of Sciences* **113,** E7890–E7899 (2016).

101. Tremblay, L. O. & Herscovics, A. Characterization of a cDNA Encoding a Novel Human Golgi α1,2-Mannosidase (IC) Involved in N-Glycan Biosynthesis. *Journal of Biological Chemistry* **275,** 31655–31660 (2000).

102. Calì, T., Galli, C., Olivari, S. & Molinari, M. Segregation and rapid turnover of EDEM1 by an autophagy-like mechanism modulates standard ERAD and folding activities. *Biochemical and Biophysical Research Communications* **371,** 405–410 (2008).

103. Giménez-Xavier, P., Francisco, R., Platini, F., Pérez, R. & Ambrosio, S. LC3-I conversion to LC3-II does not necessarily result in complete autophagy. *International Journal of Molecular Medicine* **22,** 781–785 (2008).

104. Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *The EMBO Journal* **19,** 5720–5728 (2000).

105. Avezov, E., Frenkel, Z., Ehrlich, M., Herscovics, A. & Lederkremer, G. Z. Endoplasmic Reticulum (ER) Mannosidase I Is Compartmentalized and Required for N-Glycan Trimming to Man5–6GlcNAc2 in Glycoprotein ER-associated Degradation. *Molecular Biology of the Cell* **19,** 216–225 (2008).

106. Aikawa, J.-i., Takeda, Y., Matsuo, I. & Ito, Y. Trimming of glucosylated N-glycans by human ER α1,2-mannosidase I. *The Journal of Biochemistry* **155,** 375–384 (2014).

107. Aikawa, J.-i., Matsuo, I. & Ito, Y. In vitro mannose trimming property of human ER α-1,2 mannosidase I. *Glycoconjugate Journal* **29,** 35–45 (2012).

108. Schallus, T. *et al.* Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Molecular Biology of the Cell* **19,** 3404–3414 (2008).

109. Galli, C., Bernasconi, R., Soldà, T., Calanca, V. & Molinari, M. Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *PLOS ONE* **6,** e16304 (2011).

110. Satoh, T. *et al.* Structural Basis for Oligosaccharide Recognition of Misfolded Glycoproteins by OS-9 in ER-Associated Degradation. *Molecular Cell* **40,** 905–916 (2010).

111. Bernasconi, R., Pertel, T., Luban, J. & Molinari, M. A Dual Task for the Xbp1-responsive OS-9 Variants in the Mammalian Endoplasmic Reticulum INHIBITING SECRETION OF MISFOLDED PROTEIN CONFORMERS AND ENHANCING THEIR DISPOSAL. *Journal of Biological Chemistry* **283,** 16446–16454 (2008).

112. Fujimori, T., Kamiya, Y., Nagata, K., Kato, K. & Hosokawa, N. Endoplasmic reticulum lectin XTP3-B inhibits endoplasmic reticulum-associated degradation of a misfolded α1-antitrypsin variant. *FEBS Journal* **280,** 1563–1575 (2013).

113. Groisman, B., Shenkman, M., Ron, E. & Lederkremer, G. Z. Mannose trimming is required for delivery of a glycoprotein from EDEM1 to XTP3-B and to late ER-associated degradation steps. *Journal of Biological Chemistry* jbc.M110.154849 (2010).

doi:10.1074/jbc.M110.154849

114. Sekiya, M. *et al.* EDEM Function in ERAD Protects against Chronic ER Proteinopathy and Age-Related Physiological Decline in Drosophila. *Developmental Cell* **41,** 652–664.e5 (2017).

115. Molinari, M., Calanca, V., Galli, C., Lucca, P. & Paganetti, P. Role of EDEM in the Release of Misfolded Glycoproteins from the Calnexin Cycle. *Science* **299,** 1397–1400 (2003).

116. Mayer, M. P. & Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* **62,** 670–684 (2005).

117. Blond-Elguindi, S., Fourie, A. M., Sambrook, J. F. & Gething, M. J. Peptide-dependent stimulation of the ATPase activity of the molecular chaperone BiP is the result of conversion of oligomers to active monomers. *Journal of Biological Chemistry* **268,** 12730–12735 (1993).

118. Gething, M. J. Role and regulation of the ER chaperone BiP. *Seminars in Cell & Developmental Biology* **10,** 465–472 (1999).

119. Meusser, B., Hirsch, C., Jarosch, E. & Sommer, T. ERAD: The long road to destruction. *Nature Cell Biology* **7,** 766–772 (2005).

120. Lecker, S. H., Goldberg, A. L. & Mitch, W. E. Protein Degradation by the Ubiquitin–Proteasome Pathway in Normal and Disease States. *Journal of the American Society of Nephrology* **17,** 1807–1819 (2006).

121. Vembar, S. S. & Brodsky, J. L. One step at a time: Endoplasmic reticulum-associated degradation. *Nature Reviews Molecular Cell Biology* **9,** 944–957 (2008).

122. Alonzi, D. S. *et al.* Glycoprotein misfolding in the endoplasmic reticulum: Identification of released oligosaccharides reveals a second ER-associated degradation pathway for Golgi-retrieved proteins. *Cellular and Molecular Life Sciences* **70,** 2799–2814 (2013).

123. Tarentino, A. L., Gomez, C. M. & Plummer, T. H. Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24,** 4665–4671 (1985).

124. Huang, Y. & Orlando, R. Kinetics of N-Glycan Release from Human Immunoglobulin G (IgG) by PNGase F: All Glycans Are Not Created Equal. *Journal of Biomolecular Techniques : JBT* **28,** 150–157 (2017).

125. Schekman, R. & Orci, L. Coat Proteins and Vesicle Budding. *Science* **271,** 1526–1533 (1996).

126. Schweizer, A., Fransen, J. A., Bächi, T., Ginsel, L. & Hauri, H. P. Identification, by a monoclonal antibody, of a 53-kD protein associated with a tubulo-vesicular compartment at the cis-side of the Golgi apparatus. *The Journal of Cell Biology* **107,** 1643–1653 (1988).

127. Pimpaneau, V., Midoux, P., Monsigny, M. & Roche, A. C. Characterization and isolation of an intracellular D-mannose-specific receptor from human promyelocytic

HL60 cells. *Carbohydrate Research* **213,** 95–108 (1991).

128. Appenzeller-Herzog, C. & Hauri, H.-P. The ER-Golgi intermediate compartment (ERGIC): In search of its identity and function. *Journal of Cell Science* **119,** 2173–2183 (2006).

129. Zuber, C., Spiro, M. J., Guhl, B., Spiro, R. G. & Roth, J. Golgi Apparatus Immunolocalization of Endomannosidase Suggests Post-Endoplasmic Reticulum Glucose Trimming: Implications for Quality Control. *Molecular Biology of the Cell* **11,** 4227–4240 (2000).

130. Lubas, W. A. & Spiro, R. G. Golgi endo-alpha-D-mannosidase from rat liver, a novel N-linked carbohydrate unit processing enzyme. *Journal of Biological Chemistry* **262,** 3775–3781 (1987).

131. Tulsiani, D. R., Opheim, D. J. & Touster, O. Purification and characterization of alpha-D-mannosidase from rat liver golgi membranes. *Journal of Biological Chemistry* **252,** 3227–3233 (1977).

132. Tulsiani, D. R., Hubbard, S. C., Robbins, P. W. & Touster, O. Alpha-D-Mannosidases of rat liver Golgi membranes. Mannosidase II is the GlcNAcMAN5-cleaving enzyme in glycoprotein biosynthesis and mannosidases Ia and IB are the enzymes converting Man9 precursors to Man5 intermediates. *Journal of Biological Chemistry* **257,** 3660–3668 (1982).

133. Shah, N., Kuntz, D. A. & Rose, D. R. Golgi α-mannosidase II cleaves two sugars sequentially in the same catalytic site. *Proceedings of the National Academy of Sciences* **105,** 9570–9575 (2008).

134. Ogawa, R. *et al.* Structure and Transcriptional Regulation of Human α-Mannosidase IIx (α-Mannosidase II Isotype) Gene. *European Journal of Biochemistry* **242,** 446–453 (1996).

135. Fukuda, M. N. & Akama, T. O. In vivo role of α-mannosidase IIx: Ineffective spermatogenesis resulting from targeted disruption of the Man2a2 in the mouse. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1573,** 382–387 (2002).

136. Akama, T. O. *et al.* Essential and mutually compensatory roles of α-mannosidase II and α-mannosidase IIx in N-glycan processing in vivo in mice. *Proceedings of the National Academy of Sciences* **103,** 8983–8988 (2006).

137. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1. *Cell* **103,** 467–479 (2000).

138. Stanley, P., Schachter, H. & Taniguchi, N. N-Glycans. in *Essentials of Glycobiology* (eds. Varki, A. et al.) (Cold Spring Harbor Laboratory Press, 2009).

139. Elola, M. T., Blidner, A. G., Ferragut, F., Bracalente, C. & Rabinovich, G. A. Assembly, organization and regulation of cell-surface receptors by lectin–glycan complexes. *Biochemical Journal* **469,** 1–16 (2015).

140. Takahashi, M., Kizuka, Y., Ohtsubo, K., Gu, J. & Taniguchi, N. Disease-associated

glycans on cell surface proteins. *Molecular Aspects of Medicine* **51,** 56–70 (2016).

141. Varki, A. & Kornfeld, S. P-type Lectins. in *Essentials of Glycobiology* (eds. Varki, A. et al.) (Cold Spring Harbor Laboratory Press, 2009).

142. Cooper, G. M. The Golgi Apparatus. (2000).

143. Godelaine, D., Spiro, M. J. & Spiro, R. G. Processing of the carbohydrate units of thyroglobulin. *Journal of Biological Chemistry* **256,** 10161–10168 (1981).

144. Datema, R., Romero, P. A., Legler, G. & Schwarz, R. T. Inhibition of formation of complex oligosaccharides by the glucosidase inhibitor bromoconduritol. *Proceedings of the National Academy of Sciences of the United States of America* **79,** 6787–6791 (1982).

145. Romero, P. A., Saunier, B. & Herscovics, A. Comparison between 1-deoxynojirimycin and N-methyl-1-deoxynojirimycin as inhibitors of oligosaccharide processing in intestinal epithelial cells. *Biochemical Journal* **226,** 733–740 (1985).

146. Lubas, W. A. & Spiro, R. G. Evaluation of the role of rat liver Golgi endo-alpha-D-mannosidase in processing N-linked oligosaccharides. *Journal of Biological Chemistry* **263,** 3990–3998 (1988).

147. Iwamoto, S. *et al.* Measurement of endo-α-mannosidase activity using a fluorescently labeled oligosaccharide derivative. *Bioscience, Biotechnology, and Biochemistry* **78,** 927–936 (2014).

148. Thompson, A. J. *et al.* Structural and mechanistic insight into N-glycan processing by endo-α-mannosidase. *Proceedings of the National Academy of Sciences* **109,** 781–786 (2012).

149. Speciale, G., Farren-Dai, M., Shidmoossavee, F. S., Williams, S. J. & Bennet, A. J. C2-Oxyanion Neighboring Group Participation: Transition State Structure for the Hydroxide-Promoted Hydrolysis of 4-Nitrophenyl α-d-Mannopyranoside. *Journal of the American Chemical Society* **138,** 14012–14019 (2016).

150. Moore, S. E. & Spiro, R. G. Characterization of the endomannosidase pathway for the processing of N-linked oligosaccharides in glucosidase II-deficient and parent mouse lymphoma cells. *Journal of Biological Chemistry* **267,** 8443–8451 (1992).

151. Spiro, M. J. & Spiro, R. G. Use of recombinant endomannosidase for evaluation of the processing of N-linked oligosaccharides of glycoproteins and their oligosaccharide-lipid precursors. *Glycobiology* **10,** 521–529 (2000).

152. Zhao, X. *et al.* Inhibition of Endoplasmic Reticulum-Resident Glucosidases Impairs Severe Acute Respiratory Syndrome Coronavirus and Human Coronavirus NL63 Spike Protein-Mediated Entry by Altering the Glycan Processing of Angiotensin I-Converting Enzyme 2. *Antimicrobial Agents and Chemotherapy* **59,** 206–216 (2015).

153. Hiraizumi, S., Spohr, U. & Spiro, R. G. Characterization of endomannosidase inhibitors and evaluation of their effect on N-linked oligosaccharide processing during glycoprotein biosynthesis. *Journal of Biological Chemistry* **268,** 9927–9935 (1993).

154. Ardron, H. *et al.* Synthesis of 1,5-dideoxy-3-O-(α-D-mannopyranosyl)-1,5-imino-

D-mannitol and 1,5-dideoxy-3-O-(α-D-glucopyranosyl)-1,5-imino-D-mannitol: Powerful inhibitors of endomannosidase. *Tetrahedron: Asymmetry* **4,** 2011–2024 (1993).

155. Burlingham, B. T. & Widlanski, T. S. An Intuitive Look at the Relationship of Ki and IC50: A More General Use for the Dixon Plot. *Journal of Chemical Education* **80,** 214 (2003).

156. Suhara, Y. & Achiwa, K. Disaccharides as Endomannosidase Inhibitors : Syntheses of α-Homomannojirimycin and β-Homomannojirimycin Linked to D-Glucose and D-Mannose. *Chemical and Pharmaceutical Bulletin* **43,** 414–420 (1995).

157. Spreitz, J. & Stütz, A. E. Golgi endomannosidase inhibitor, α-d-glucopyranosyl-(1→3)-1-deoxymannojirimycin: A five-step synthesis from maltulose and examples of N-modified derivatives. *Carbohydrate Research* **339,** 1823–1827 (2004).

158. Quach, T. *et al.* Fleetamine (3-O-α-D-glucopyranosyl-swainsonine): The synthesis of a hypothetical inhibitor of endo-α-mannosidase. *Tetrahedron: Asymmetry* **23,** 992–997 (2012).

159. Hakki, Z. *et al.* Structural and Kinetic Dissection of the endo-α-1,2-Mannanase Activity of Bacterial GH99 Glycoside Hydrolases from Bacteroides spp. *Chemistry – A European Journal* **21,** 1966–1977 (2015).

160. Moore, S. E. & Spiro, R. G. Demonstration that Golgi endo-alpha-D-mannosidase provides a glucosidase-independent pathway for the formation of complex N-linked oligosaccharides of glycoproteins. *Journal of Biological Chemistry* **265,** 13104–13112 (1990).

161. Tulsiani, D. R. P., Coleman, V. D. & Touster, O. Asparagine-linked glycoprotein biosynthesis in rat brain: Identification of glucosidase I, glucosidase II, and an endomannosidase (glucosyl mannosidase). *Archives of Biochemistry and Biophysics* **277,** 114–121 (1990).

162. Rabouille, C. & Spiro, R. G. Nonselective utilization of the endomannosidase pathway for processing glycoproteins by human hepatoma (HepG2) cells. *Journal of Biological Chemistry* **267,** 11573–11578 (1992).

163. Torossi, T., Fan, J.-Y., Sauter-Etter, K., Roth, J. & Ziak, M. Endomannosidase processes oligosaccharides of α1-antitrypsin and its naturally occurring genetic variants in the Golgi apparatus. *Cellular and Molecular Life Sciences CMLS* **63,** 1923–1932 (2006).

164. Dedola, S. *et al.* Direct assay for endo-α-mannosidase substrate preference on correctly folded and misfolded model glycoproteins. *Carbohydrate Research* **434,** 94–98 (2016).

165. Moore, S. E. & Spiro, R. G. Inhibition of glucose trimming by castanospermine results in rapid degradation of unassembled major histocompatibility complex class I molecules. *Journal of Biological Chemistry* **268,** 3809–3812 (1993).

166. Weng, S. & Spiro, R. G. Evaluation of the early processing routes of N-linked oligosaccharides of glycoproteins through the characterization of Man8GlcNAc2 isomers: Evidence that endomannosidase functions in vivo in the absence of a glucosi-

dase blockade. *Glycobiology* **6,** 861–868 (1996).

167. Kim, I. S. *et al.* Mechanism of membrane fusion induced by vesicular stomatitis virus G protein. *Proceedings of the National Academy of Sciences* **114,** E28–E36 (2017).

168. Karaivanova, V. K., Luan, P. & Spiro, R. G. Processing of viral envelope glycoprotein by the endomannosidase pathway: Evaluation of host cell specificity. *Glycobiology* **8,** 725–730 (1998).

169. Kukushkin, N. V., Easthope, I. S., Alonzi, D. S. & Butters, T. D. Restricted processing of glycans by endomannosidase in mammalian cells. *Glycobiology* **22,** 1282–1288 (2012).

170. Negroiu, G., Branza-Nichita, N., Petrescu, A. J., Dwek, R. A. & Petrescu, S. M. Protein specific N-glycosylation of tyrosinase and tyrosinase-related protein-1 in B16 mouse melanoma cells. *Biochemical Journal* **344,** 659–665 (1999).

171. Durrant, C. & Moore, S. E. H. Perturbation of free oligosaccharide trafficking in endoplasmic reticulum glucosidase I-deficient and castanospermine-treated cells. *Biochemical Journal* **365,** 239–247 (2002).

172. Torossi, T., Guhl, B., Roth, J. & Ziak, M. Endomannosidase undergoes phosphorylation in the Golgi apparatus. *Glycobiology* **20,** 55–61 (2010).

173. Kukushkin, N. V., Alonzi, D. S., Dwek, R. A. & Butters, T. D. Demonstration that endoplasmic reticulum-associated degradation of glycoproteins can occur downstream of processing by endomannosidase. *Biochemical Journal* **438,** 133–142 (2011).

174. Moore, S. E., Bauvy, C. & Codogno, P. Endoplasmic reticulum-to-cytosol transport of free polymannose oligosaccharides in permeabilized HepG2 cells. *The EMBO Journal* **14,** 6034–6042 (1995).

175. Katiyar, S., Li, G. & Lennarz, W. J. A complex between peptide:N-glycanase and two proteasome-linked proteins suggests a mechanism for the degradation of misfolded glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 13774–13779 (2004).

176. Spiro, M. J., Bhoyroo, V. D. & Spiro, R. G. Molecular Cloning and Expression of Rat Liver Endo-α-mannosidase, an N-Linked Oligosaccharide Processing Enzyme. *Journal of Biological Chemistry* **272,** 29356–29363 (1997).

177. Dong, Z., Zuber, C., Spiro, M. J., Spiro, R. G. & Roth, J. Immunohistochemical evaluation of endomannosidase distribution in rat tissues: Evidence for cell type-specific expression. *Histochemistry and Cell Biology* **114,** 461–467 (2000).

178. Verde, C. *et al.* Effect of ATP depletion and DTT on the transport of membrane proteins from the endoplasmic reticulum and the intermediate compartment to the Golgi complex. *European Journal of Cell Biology* **67,** 267–274 (1995).

179. Spiro, M. J. & Spiro, R. G. Release of polymannose oligosaccharides from vesicular stomatitis virus G protein during endoplasmic reticulum–associated degradation. *Glycobiology* **11,** 803–811 (2001).

180. Aebi, M. *et al.* Carbohydrate-deficient glycoprotein syndromes become congen-

ital disorders of glycosylation: An updated nomenclature for CDG. *Glycoconjugate Journal* **16,** 669–671 (1999).

181. De Praeter, C. M. *et al.* A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency. *The American Journal of Human Genetics* **66,** 1744–1756 (2000).

182. Völker, C. *et al.* Processing of N-linked carbohydrate chains in a patient with glucosidase I deficiency (CDG type IIb). *Glycobiology* **12,** 473–483 (2002).

183. Sadat, M. A. *et al.* Glycosylation, Hypogammaglobulinemia, and Resistance to Viral Infections. *New England Journal of Medicine* **370,** 1615–1625 (2014).

184. Gañán, S., Cazzulo, J. J. & Parodi, A. J. A major proportion of N-glycoproteins are transiently glucosylated in the endoplasmic reticulum. *Biochemistry* **30,** 3098–3104 (1991).

185. Dairaku, K. & Spiro, R. G. Phylogenetic survey of endomannosidase indicates late evolutionary appearance of this N-linked oligosaccharide processing enzyme. *Glycobiology* **7,** 579–586 (1997).

186. Yu, Y. *et al.* Substance dependence low-density whole genome association study in two distinct American populations. *Human Genetics* **123,** 495 (2008).

187. Farrer, L. A. *et al.* Association of Variants in MANEA With Cocaine-Related Behaviors Mar 01, 2009 JAMA Psychiatry JAMA Network. *Archives of General Psychiatry* **66,** 267–274 (2009).

188. Jensen, K. P. *et al.* The α-endomannosidase gene (MANEA) is associated with panic disorder and social anxiety disorder. *Translational Psychiatry* **4,** e353 (2014).

189. Hu, Y. *et al.* A Pooling Genome-Wide Association Study Combining a Pathway Analysis for Typical Sporadic Parkinson's Disease in the Han Population of Chinese Mainland. *Molecular Neurobiology* **53,** 4302–4318 (2015).

190. Huang, Y.-T., Liang, L., Moffatt, M. F., Cookson, W. O. C. M. & Lin, X. iGWAS: Integrative Genome-Wide Association Studies of Genetic and Genomic Data for Disease Susceptibility Using Mediation Analysis. *Genetic epidemiology* **39,** 347–356 (2015).

191. Romanuik, T. L. *et al.* Identification of novel androgen-responsive genes by sequencing of LongSAGE libraries. *BMC Genomics* **10,** 476 (2009).

192. Hiraizumi, S., Spohr, U. & Spiro, R. G. Ligand affinity chromatographic purification of rat liver Golgi endomannosidase. *Journal of Biological Chemistry* **269,** 4697–4700 (1994).

193. Spiro, R. G., Zhu, Q., Bhoyroo, V. & Söling, H.-D. Definition of the Lectin-like Properties of the Molecular Chaperone, Calreticulin, and Demonstration of Its Copurification with Endomannosidase from Rat Liver Golgi. *Journal of Biological Chemistry* **271,** 11588–11594 (1996).

194. Bause, E. & Burbach, M. Purification and Enzymatic Properties of Endo-α1,2-Mannosidase from Pig Liver Involved in Oligosaccharide Processing. *Biological Chem-*

istry Hoppe-Seyler **377,** 639–646 (1996).

195. Hamilton, S. R. *et al.* Intact α-1,2-endomannosidase is a typical type II membrane protein. *Glycobiology* **15,** 615–624 (2005).

196. Hardt, B. *et al.* Human endo-α1,2-mannosidase is a Golgi-resident type II membrane protein. *Biochimie* **87,** 169–179 (2005).

197. Torossi, T., Roth, J. & Ziak, M. A single tryptophan residue of endomannosidase is crucial for Golgi localization and in vivo activity. *Cellular and Molecular Life Sciences* **64,** 1881–1889 (2007).

198. Brinkrolf, K. *et al.* Chinese hamster genome sequenced from sorted chromosomes. *Nature Biotechnology* **31,** 694 (2013).

199. Stehli, J., Torossi, T. & Ziak, M. Triple arginines in the cytoplasmic tail of endomannosidase are not essential for type II membrane topology and Golgi localization. *Cellular and Molecular Life Sciences* **65,** 1609–1619 (2008).

200. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 10915–10919 (1992).

201. Geneious version 8.1.9 created by Biomatters. Available from https://www.geneious.com.

202. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32,** 1792–1797 (2004).

203. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Research* **43,** D512–D520 (2015).

204. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67,** 386–394 (2011).

205. Matsuda, K. *et al.* Heterologous Expression, Purification, and Characterization of an α-Mannosidase Belonging to Glycoside Hydrolase Family 99 of Shewanella amazonensis. *Bioscience, Biotechnology, and Biochemistry* **75,** 797–799 (2011).

206. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859 (1995).

207. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* **517,** 165–169 (2015).

208. Robb, C. S., Mystkowska, A. A. & Hehemann, J.-H. Crystal structure of a marine glycoside hydrolase family 99-related protein lacking catalytic machinery. *Protein Science* **26,** 2445–2450 (2017).

209. Rathore, U. *et al.* Glycosylation of the core of the HIV-1 envelope subunit protein gp120 is not required for native trimer formation or viral infectivity. *Journal of*

Biological Chemistry jbc.M117.788919 (2017). doi:10.1074/jbc.M117.788919

210. Wang, W. *et al.* A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology* **10,** 14 (2013).

211. Pritchard, L. K. *et al.* Structural Constraints Determine the Glycosylation of HIV-1 Envelope Trimers. *Cell Reports* **11,** 1604–1613 (2015).

212. Behrens, A.-J. *et al.* Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Reports* **14,** 2695–2706 (2016).

213. Coss, K. P. *et al.* HIV-1 Glycan Density Drives the Persistence of the Mannose Patch within an Infected Individual. *Journal of Virology* **90,** 11132–11144 (2016).

214. Yang, Z.-Y. *et al.* pH-Dependent Entry of Severe Acute Respiratory Syndrome Coronavirus Is Mediated by the Spike Glycoprotein and Enhanced by Dendritic Cell Transfer through DC-SIGN. *Journal of Virology* **78,** 5642–5650 (2004).

215. Ritchie, G. *et al.* Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology* **399,** 257–269 (2010).

216. Fischer, P. B. *et al.* The alpha-glucosidase inhibitor N-butyldeoxynojirimycin inhibits human immunodeficiency virus entry at the level of post-CD4 binding. *Journal of Virology* **69,** 5791–5797 (1995).

217. Fischer, P. B., Karlsson, G. B., Dwek, R. A. & Platt, F. M. N-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with impaired gp120 shedding and gp41 exposure. *Journal of Virology* **70,** 7153–7160 (1996).

218. Fischl, M. A. *et al.* The safety and efficacy of combination N-butyldeoxynojirimycin (SC-48334) and zidovudine in patients with HIV-1 infection and 200-500 CD4 cells/mm3. *Journal of Acquired Immune Deficiency Syndromes* **7,** 139–147 (1994).

219. Ficicioglu, C. Review of miglustat for clinical management in Gaucher disease type 1. *Therapeutics and Clinical Risk Management* **4,** 425–431 (2008).

220. Chang, J., Guo, J.-T., Du, Y. & Block, T. Imino sugar glucosidase inhibitors as broadly active anti-filovirus agents. *Emerging Microbes & Infections* **2,** e77 (2013).

221. Beek, W. P. van, Smets, L. A. & Emmelot, P. Increased Sialic Acid Density in Surface Glycoprotein of Transformed and Malignant Cells—a General Phenomenon? *Cancer Research* **33,** 2913–2922 (1973).

222. Yamashita, K., Ohkura, T., Tachibana, Y., Takasaki, S. & Kobata, A. Comparative study of the oligosaccharides released from baby hamster kidney cells and their polyoma transformant by hydrazinolysis. *Journal of Biological Chemistry* **259,** 10834–10840 (1984).

223. Pierce, M. & Arango, J. Rous sarcoma virus-transformed baby hamster kidney cells express higher levels of asparagine-linked tri- and tetraantennary glycopeptides

containing [GlcNAc-beta (1,6)Man-alpha (1,6)Man] and poly-N-acetyllactosamine sequences than baby hamster kidney cells. *Journal of Biological Chemistry* **261,** 10772–10777 (1986).

224. Henderson, N. C. & Sethi, T. The regulation of inflammation by galectin-3. *Immunological Reviews* **230,** 160–171 (2009).

225. Cummings, R. D., Trowbridge, I. S. & Kornfeld, S. A mouse lymphoma cell line resistant to the leukoagglutinating lectin from Phaseolus vulgaris is deficient in UDP-GlcNAc: Alpha-D-mannoside beta 1,6 N-acetylglucosaminyltransferase. *Journal of Biological Chemistry* **257,** 13421–13427 (1982).

226. Yamashita, K., Tachibana, Y., Ohkura, T. & Kobata, A. Enzymatic basis for the structural changes of asparagine-linked sugar chains of membrane glycoproteins of baby hamster kidney cells induced by polyoma transformation. *Journal of Biological Chemistry* **260,** 3963–3969 (1985).

227. Dennis, J. W., Laferte, S., Waghorne, C., Breitman, M. L. & Kerbel, R. S. Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* **236,** 582–585 (1987).

228. Granovsky, M. *et al.* Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nature Medicine* **6,** 306–312 (2000).

229. Demetriou, M., Granovsky, M., Quaggin, S. & Dennis, J. W. Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* **409,** 733–739 (2001).

230. Lagana, A. *et al.* Galectin Binding to Mgat5-Modified N-Glycans Regulates Fibronectin Matrix Remodeling in Tumor Cells. *Molecular and Cellular Biology* **26,** 3181–3193 (2006).

231. Pierce, J. M., Hindsgaul, O. & CA. United States Patent: 5032505 - Inhibitors for glycosaminosyl transferase V. (1991).

232. Tedaldi, L. M., Pierce, M. & Wagner, G. K. Optimised chemical synthesis of 5-substituted UDP-sugars and their evaluation as glycosyltransferase inhibitors. *Carbohydrate Research* **364,** 22–27 (2012).

233. Woods, E. C. *et al.* A bulky glycocalyx fosters metastasis formation by promoting G1 cell cycle progression. *eLife* **6,** e25752 (2017).

234. Kijima-Suda, I., Miyamoto, Y., Toyoshima, S., Itoh, M. & Osawa, T. Inhibition of Experimental Pulmonary Metastasis of Mouse Colon Adenocarcinoma 26 Sublines by a Sialic Acid:Nucleoside Conjugate Having Sialyltransferase Inhibiting Activity. *Cancer Research* **46,** 858–862 (1986).

235. Heifetz, A., Keenan, R. W. & Elbein, A. D. Mechanism of action of tunicamycin on the UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase. *Biochemistry* **18,** 2186–2192 (1979).

236. King, I. A. & Tabiowo, A. Effect of tunicamycin on epidermal glycoprotein and

glycosaminoglycan synthesis in vitro. *Biochemical Journal* **198,** 331–338 (1981).

237. Dennis, J. W. Effects of Swainsonine and Polyinosinic:Polycytidylic Acid on Murine Tumor Cell Growth and Metastasis. *Cancer Research* **46,** 5131–5136 (1986).

238. Sparks, S. E. & Krasnewich, D. M. Congenital Disorders of N-Linked Glycosylation and Multiple Pathway Overview. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 2017).

239. Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. Substrate-Assisted Catalysis Unifies Two Families of Chitinolytic Enzymes. *Journal of the American Chemical Society* **119,** 7954–7959 (1997).

240. Mark, B. L. *et al.* Crystallographic Evidence for Substrate-assisted Catalysis in a Bacterial β-Hexosaminidase. *Journal of Biological Chemistry* **276,** 10330–10337 (2001).

241. Alteen, M. G. *et al.* Mechanism of Human Nucleocytoplasmic Hexosaminidase D. *Biochemistry* **55,** 2735–2747 (2016).

242. Aalten, D. M. F. van *et al.* Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proceedings of the National Academy of Sciences* **98,** 8979–8984 (2001).

243. Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocadlo, D. J. O-GlcNAcase Uses Substrate-assisted Catalysis. Kinetic analysis and development of highly selective mechanism-inspired inhibitors. *Journal of Biological Chemistry* **280,** 25313–25322 (2005).

244. Dennis, R. J. *et al.* Structure and mechanism of a bacterial β-glucosaminidase having *O*-GlcNAcase activity. *Nature Structural & Molecular Biology* **13,** 365 (2006).

245. He, Y., Macauley, M. S., Stubbs, K. A., Vocadlo, D. J. & Davies, G. J. Visualizing the Reaction Coordinate of an O-GlcNAc Hydrolase. *Journal of the American Chemical Society* **132,** 1807–1809 (2010).

246. Cekic, N. *et al.* Analysis of transition state mimicry by tight binding aminothiazoline inhibitors provides insight into catalysis by human O-GlcNAcase. *Chemical Science* **7,** 3742–3750 (2016).

247. Marković-Housley, Z. *et al.* Crystal Structure of Hyaluronidase, a Major Allergen of Bee Venom. *Structure* **8,** 1025–1035 (2000).

248. Abbott, D. W., Macauley, M. S., Vocadlo, D. J. & Boraston, A. B. Streptococcus pneumoniae Endohexosaminidase D, Structural and Mechanistic Insight into Substrate-assisted Catalysis in Family 85 Glycoside Hydrolases. *Journal of Biological Chemistry* **284,** 11676–11689 (2009).

249. Micheel, F. & Borrmann, D. Ein neues Verfahren zur Synthese höherer Saccharide. *Chemische Berichte* **93,** 1143–1147 (1960).

250. Roston, D., Islam, Z. & Kohen, A. Kinetic isotope effects as a probe of hydrogen transfers to and from common enzymatic cofactors. *Archives of Biochemistry and*

Biophysics **544,** 96–104 (2014).

251. Gu, H. & Zhang, S. Advances in Kinetic Isotope Effect Measurement Techniques for Enzyme Mechanism Study. *Molecules* **18,** 9278–9292 (2013).

252. Schramm, V. L. Enzymatic Transition States and Transition State Analog Design. *Annual Review of Biochemistry* **67,** 693–720 (1998).

253. Reeves, R. E. The Shape of Pyranoside Rings. *Journal of the American Chemical Society* **72,** 1499–1506 (1950).

254. Alonso-Gil, S. *et al.* Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α-Mannosidases. *Journal of the American Chemical Society* **139,** 1085–1088 (2017).

255. Zhu, Y. *et al.* Mechanistic insights into a Ca2+-dependent family of α-mannosidases in a human gut symbiont. *Nature Chemical Biology* **6,** 125–132 (2010).

256. Elbein, A. D., Solf, R., Dorling, P. R. & Vosbeck, K. Swainsonine: An inhibitor of glycoprotein processing. *Proceedings of the National Academy of Sciences of the United States of America* **78,** 7393–7397 (1981).

257. Dorling, P. R., Huxtable, C. R. & Colegate, S. M. Inhibition of lysosomal α-mannosidase by swainsonine, an indolizidine alkaloid isolated from Swainsona canescens. *Biochemical Journal* **191,** 649–651 (1980).

258. Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent Escherichia coli: Transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 2172–2175 (1989).

259. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology* **166,** 557–580 (1983).

260. Origin. OriginLab, Northampton, MA.

261. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43,** 186–190 (2010).

262. Waterman, D. G. *et al.* Diffraction-geometry refinement in the DIALS framework. *Acta Crystallographica Section D: Structural Biology* **72,** 558–575 (2016).

263. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography* **69,** 1204–1214 (2013).

264. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography* **67,** 355–367 (2011).

265. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66,** 486–501 (2010).

266. Potterton, L. *et al.* CCP4i2: The new graphical user interface to the CCP4 program

suite. *Acta Crystallographica Section D: Structural Biology* **in press,** (2018).

267. Speciale, G., Thompson, A. J., Davies, G. J. & Williams, S. J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Current Opinion in Structural Biology* **28,** 1–13 (2014).

268. ChemAxon. Marvin.

269. Petricevic, M. *et al.* Contribution of Shape and Charge to the Inhibition of a Family GH99 endo-α-1,2-Mannanase. *Journal of the American Chemical Society* **139,** 1089–1097 (2017).

270. Nerinckx, W., Desmet, T., Piens, K. & Claeyssens, M. An elaboration on the syn–anti proton donor concept of glycoside hydrolases: Electrostatic stabilisation of the transition state as a general strategy. *FEBS Letters* **579,** 302–312 (2005).

271. Serizawa, K., Noguchi, M., Li, G. & Shoda, S.-i. First Detection of Unprotected 1,2-Anhydro Aldopyranoses. *Chemistry Letters* **46,** 1024–1026 (2017).

272. Williams, R. J. *et al.* Combined Inhibitor Free-Energy Landscape and Structural Analysis Reports on the Mannosidase Conformational Coordinate. *Angewandte Chemie International Edition* **53,** 1087–1091 (2014).

273. Tailford, L. E. *et al.* Structural and biochemical evidence for a boat-like transition state in β-mannosidases. *Nature Chemical Biology* **4,** 306–312 (2008).

274. Lebedev, A. A. *et al.* JLigand: A graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68,** 431–440 (2012).

275. Broecker, J., Vargas, C. & Keller, S. Revisiting the optimal c value for isothermal titration calorimetry. *Analytical Biochemistry* **418,** 307–309 (2011).

276. Turnbull, W. B. & Daranas, A. H. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *Journal of the American Chemical Society* **125,** 14859–14866 (2003).

277. Fernandes, P. Z., Petricevic, M., Sobala, L., Davies, G. J. & Williams, S. J. Exploration of Strategies for Mechanism-Based Inhibitor Design for Family GH99 endo-α-1,2-Mannanases. *Chemistry – A European Journal* **24,** 7464–7473 (2018).

278. Mandhapati, A. R. *et al.* Importance of the 6 -Hydroxy Group and Its Configuration for Apramycin Activity. *ChemMedChem* **9,** 2074–2083 (2014).

279. Liu, H., Liang, X., Søhoel, H., Bülow, A. & Bols, M. Noeuromycin, A Glycosyl Cation Mimic that Strongly Inhibits Glycosidases. *Journal of the American Chemical Society* **123,** 5116–5117 (2001).

280. Davies, G. Glycoside Hydrolase Family 99. *CAZYpedia* (2015).

281. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291,** 1304–1351 (2001).

282. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues.

Nature **550,** 204–213 (2017).

283. Shaul, Y. D. *et al.* MERAV: A tool for comparing gene expression across human tissues and cell types. *Nucleic Acids Research* **44,** D560–D566 (2016).

284. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology* **28,** 1248–1250 (2010).

285. The Human Protein Atlas. www.proteinatlas.org.

286. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356,** eaal3321 (2017).

287. Moremen, K., LaBaer, J., Steel, J., Jarvis, D. & Gilbert, H. Glyco-Enzyme Repository.

288. Huang, C.-H. *et al.* UbiSite: Incorporating two-layered machine learning method with substrate motifs to predict ubiquitin-conjugation site on lysines. *BMC Systems Biology* **10,** S6 (2016).

289. Wang, J.-R. *et al.* ESA-UbiSite: Accurate prediction of human ubiquitination sites by identifying a set of effective negatives. *Bioinformatics* **33,** 661–668 (2017).

290. Qing, G. *et al.* Cold-shock induced high-yield protein production in Escherichia coli. *Nature Biotechnology* **22,** 877–882 (2004).

291. Bijelic, A. & Rompel, A. Ten Good Reasons for the Use of the Tellurium-Centered Anderson–Evans Polyoxotungstate in Protein Crystallography. *Accounts of Chemical Research* **50,** 1441–1448 (2017).

292. Bijelic, A. *et al.* Hen Egg-White Lysozyme Crystallisation: Protein Stacking and Structure Stability Enhanced by a Tellurium(VI)-Centred Polyoxotungstate. *Chembiochem* **16,** 233–241 (2015).

293. Molitor, C., Mauracher, S. G. & Rompel, A. Aurone synthase is a catechol oxidase with hydroxylase activity and provides insights into the mechanism of plant polyphenol oxidases. *Proceedings of the National Academy of Sciences* **113,** E1806–E1815 (2016).

294. Pretzler, M., Bijelic, A. & Rompel, A. Heterologous expression and characterization of functional mushroom tyrosinase (AbPPO4). *Scientific Reports* **7,** 1810 (2017).

295. Stein, N. CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **41,** 641–643 (2008).

296. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

297. Tickle, I. J. Statistical quality indicators for electron-density maps. *Acta Crystallographica Section D: Biological Crystallography* **68,** 454–467 (2012).

298. Wiedemann, C., Bellstedt, P. & Görlach, M. CAPITO—a web server-based anal-

ysis and plotting tool for circular dichroism data. *Bioinformatics* **29,** 1750–1757 (2013).

299. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40 (2008).

300. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols* **5,** 725 (2010).

301. Yang, J. *et al.* The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* **12,** 7 (2014).

302. Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Science : A Publication of the Protein Society* **11,** 739–756 (2002).

303. Whitmore, L., Miles, A. J., Mavridis, L., Janes, R. W. & Wallace, B. A. PCDDB: New developments at the Protein Circular Dichroism Data Bank. *Nucleic Acids Research* **45,** D303–D307 (2017).

304. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal* **321,** 557–559 (1997).

305. COSMIC, the Catalogue of Somatic Mutations in Cancer. http://cancer.sanger.ac.uk/.

306. Forbes, S. A. *et al.* COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Research* **45,** D777–D783 (2017).

307. Deng, W. *et al.* GPS-PAIL: Prediction of lysine acetyltransferase-specific modification sites from protein sequences. *Scientific Reports* **6,** 39787 (2016).

308. Shazman, S., Celniker, G., Haber, O., Glaser, F. & Mandel-Gutfreund, Y. Patch Finder Plus (PFplus): A web server for extracting and displaying positive electrostatic patches on protein surfaces. *Nucleic Acids Research* **35,** W526–W530 (2007).

309. Chan, P., Curtis, R. A. & Warwicker, J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific Reports* **3,** (2013).

310. Thompson, A. J. Structural and mechanistic analysis of N-glycan-processing alpha-mannosidases. (University of York, 2012).

311. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* **131,** 209–217 (1984).

312. Belda, O. & Targett-Adams, P. Small molecule inhibitors of the hepatitis C virus-encoded NS5A protein. *Virus Research* **170,** 1–14 (2012).

313. Karlsson, G. B., Butters, T. D., Dwek, R. A. & Platt, F. M. Effects of the imino sugar N-butyldeoxynojirimycin on the N-glycosylation of recombinant gp120. *The Journal of Biological Chemistry* **268,** 570–576 (1993).

314. Jo, J. *et al.* Draft genome of the sea cucumber Apostichopus japonicus and genetic polymorphism among color variants. *GigaScience* **6,** 1–6 (2017).

315. NCBI Resource Coordinators. Database Resources of the National Center for

Biotechnology Information. *Nucleic Acids Research* **45,** D12–D17 (2017).

316. Simion, P. *et al.* A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology* **27,** 958–967 (2017).

317. Laumer, C. E. *et al.* Placozoans are eumetazoans related to Cnidaria. *bioRxiv* 200972 (2017). doi:10.1101/200972

318. OneZoom Tree of Life Explorer, available at http://www.onezoom.org/.

319. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59,** 307–321 (2010).

320. Palaniappan, K. K. & Bertozzi, C. R. Chemical Glycoproteomics. *Chemical Reviews* **116,** 14277–14306 (2016).

321. Aebersold, R. *et al.* How many human proteoforms are there? *Nature Chemical Biology* **14,** 206–214 (2018).

322. Yang, S., Chatterjee, S. & Cipollo, J. The Glycoproteomics-mass Spectrometry for Studying Glycosylation in Cardiac Hypertrophy and Heart Failure. *PROTEOMICS – Clinical Applications* (2018). doi:10.1002/prca.201700075

323. Oliveira, D. *et al.* Use of Mass Spectrometry to Screen Glycan Early Markers in Hepatocellular Carcinoma. *Frontiers in Oncology* **7,** (2018).

324. Bartolazzi, A., Sciacchitano, S. & D'Alessandria, C. Galectin-3: The Impact on the Clinical Management of Patients with Thyroid Nodules and Future Perspectives. *International Journal of Molecular Sciences* **19,** 445 (2018).

325. McNaught, A. D. Nomenclature of carbohydrates. *Carbohydrate Research* **297,** 1–92 (1997).

326. Agirre, J. Strategies for carbohydrate model building, refinement and validation. *Acta Crystallographica Section D: Structural Biology* **73,** 171–186 (2017).