Glucosylceramide metabolism: from 3D structure to the development of selective chemical probes

Imogen Zofia Breen

PhD

University of York

Chemistry

September 2017

Abstract

Gaucher's disease, the most prevalent of the lysosomal storage disorders, is caused by insufficient lysosomal glucocerebrosidase (GBA1) activity. This is the result of point mutations in the encoding gene, GBA1. The consequence of this reduction in activity is an accumulation of GBA1's substrate, glucosylceramide, in the lysosomes, leading to the various pathologies of Gaucher's disease. Treatment approaches for Gaucher's disease range include enzyme replacement therapy, substrate reduction therapy and the use of small molecules to stabilise mutant forms of the enzyme – pharmacological chaperone therapy. Diagnosis and treatment of Gaucher's disease requires regular quantification of the active GBA1 in a patient's tissues, not just the total GBA1 concentration or total β-glucosidase activity. Human cells also contain a secondary, non-lysosomal glucocerebrosidase, GBA2. The activity of GBA2 can affect the pathology of Gaucher's disease, and GBA2 may interact with some chaperones and probes targeted for GBA1. Point mutations occurring in GBA2 are also linked to human diseases including hereditary spastic paraplegia and cerebellar ataxia.

In this thesis, I describe the three-dimensional structures of human GBA1 and a bacterial homologue for GBA2, TxGH116 – both unliganded and bound to a variety of probe and inhibitor compounds. Prior to this work the only structure of GBA1 with a covalently bound inhibitor were complexes with conduritol-β-epoxide. Here I describe GBA1 in complex with both gluco- and galacto-configured aziridines. The reporting of the structure of TxGH116, which was the first GH116 family protein to be structurally characterised, lead to the founding of CAZy structural clan GH-O. The structural knowledge of these two glucocerebrosidase proteins, especially in complex with covalent inhibitors, can be used to aid the design of probes and eventually drugs with improved specificity for GBA1 or GBA2.

Table of Contents

Abstract iii

Table of Figures ix

Table of Tables xi

Acknowledgements xiii

Authors Declaration xiv

Chapter 1. Glucosylceramide metabolism: from 3D structure to the development of selective chemical probes 1

1. Introduction 2

1.1 The Role of Glucosylceramide 2

1.2 Lysosomal Storage Disorders and Gaucher's Disease 6

1.3 Gaucher's disease and GBA1 6

1.4 GBA1 in other disease 13

1.5 History of enzyme replacement therapy 13

1.6 Chaperone therapy 17

1.7 GBA1 inhibitors 20

1.8 Activity Based Probes 23

1.9 Human GBA1 and GBA2 26

1.10 CAZy and the glycosidase hydrolases 27

1.11 Summary 29

Chapter 2. 3D structures of human glucocerebrosidase GBA1 in complex with activity-based probes 31

2.1 Abstract 31

2.2 Introduction 32

2.2.1 The 3D structure of human GBA1 35

2.2.2 Aims of the research in this chapter 41

2.3 Materials and Methods 42

2.3.1 Purification of commercial GBA1 enzyme 42

2.3.2 Crystallisation conditions 42

2.3.3 Collection of x-ray data 43

2.3.4 Data processing and structure determination 43
2.3.4 Labelling with fluorescent probes 43
2.4 Results and Discussion 45
2.4.1 Unliganded GBA1 Structure 47
2.4.2 3D complexes with Cyclophellitol KY170 and cyclophellitol aziridine KY353 47
2.4.3 3D complex of GBA1 with acyl cyclophellitol aziridine KY358 49
2.4.4 3D complex of GBA1 with a *galacto*-configured cyclophellitol, TB562 51
2.4.5 3D complex of GBA1 with KY358 and MDW933 54
2.5 Summary 59
Chapter 3. Purification and crystallisation of a bacterial GH116 enzyme 61
3.1 Abstract 61
3.2 Introduction 62
3.2.1 GBA2 – non-lysosomal glucosylceramidase 62
3.2.2 GBA2 in human disease 64
3.2.3 GH116 enzymes 66
3.2.4 Aims of the research in this chapter 68
3.3 Material and Methods 69
3.3.1 Sequence Alignment 69
3.3.2 Cloning and transformations of bacterial GH116 enzymes 69
3.3.3 Overexpression and purification of TxGH116N 69
3.3.4 SEC-MALLS 70
3.3.5 Activity Based Probes 71
3.3.6 Crystallisation of TxGH116N 71
3.3.7 Data collection 72
3.4 Results 73
3.4.1 Sequence alignment 73
3.4.2 Expression of Bacterial GH116 enzymes 75
3.4.3 Purification of TxGH116N 76
3.4.4 SEC-MALLS 79
3.4.4 Activity based probes 81

3.4.5 Crystallisation and data Collection....82
3.5 Summary....83
Chapter 4. 3D structures of bacterial glucocerebrosidase TxGH116 in complex....85
4.1 Abstract....85
4.2 Introduction....86
4.2.1 Characterisation of GH116 family enzymes....86
4.2.2 Aims of the research in this chapter....89
4.3 Materials and methods....91
4.3.1 Data collection of TxGH116N crystals....91
4.3.2 Data processing and structure determination....91
4.3.3 Overexpression and purification of TxGH116C....91
4.3.4 Crystallisation of TxGH116C and ligand soaking....92
4.3.5 Data collection of TxGH116C crystals....93
4.3.6 Data processing and structure determination....93
4.3.7 Isothermal titration calorimetry (ITC)....94
4.3.8 Homology Modelling....94
4.4 Results and discussion....95
4.4.1 Overexpression and Purification of TxGH116C....95
4.4.2 Unliganded TxGH116 structure....98
4.4.3 Crystallisation of TxGH116C....102
4.4.4 Active Centre of GH116....104
4.4.5 3D complexes of TxGH116C with castanospermine....106
4.4.6 3D complexes of TxGH116C with KY358....108
4.4.7 3D complexes of TxGH116C with KY375....110
4.4.8 TxGH116 and other retaining β-glucosidases....112
4.4.9 Homology Modelling of the Human GBA2 Enzyme....114
4.5 Summary....118
Chapter 5. Conclusions and perspectives....119
Appendix 1. GH116 protein sequence alignment....131
Appendix 2. Bacterial GH116 sequences....139

Appendix 3. TxGH116C protein sequence 145

Appendix 4. Publication: Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). ACS Chem. Biol.2016 147

Appendix 5. Publication: 1, 6-Cyclosulfates: A New Class of Irreversible Glycosidase Inhibitor. ACS Cent. Sci. 2017 156

Appendix 6. Competitive and Covalent Inhibitors of Human Lysosomal Retaining Exoglucosidases. eLS. 2018..... 167

Appendix 7. A Fluorescence Polarization Activity-Based Protein Profiling Assay in the Discovery of Potent, Selective Inhibitors for Human Non-lysosomal Glucosylceramidase. JACS. 2017..... 179

List of Abbreviations..... 182

References:..... 189

Table of Figures

Figure 1.1. The general structures of biological sphingolipids. 4
Figure 1.2. The reaction catabolised by GBA1, GBA2 and GBA3. 5
Figure 1.3 The three-dimensional structure of GBA1 11
Figure 1.4. Substrate reduction therapy drugs 16
Figure 1.5 A general overview of chaperone therapy 18
Figure 1.6. An example of some inhibitors and activity based probes 21
Figure 1.7. The crystal structure of human GBA1 22
Figure 1.8 Schematic diagram for an activity based probe 25
Figure 1.9 Retaining glycosides mechanism 28
Figure 2.1. The reaction catalysed by GBA1. 32
Figure 2.2. Double-displacement reaction mechanism for GBA1 33
Figure 2.3. Development of β-glucosidase-selective ABPs. 35
Figure 2.4. 3D structure of human GBA1 37
Figure 2.5. GBA1 inhibitors observed in published 3D structures. 40
Figure 2.6 Activity-based GBA1 probes studied at a structural level in this thesis 41
Figure 2.7. Florescent probes used in this time course experiment 44
Figure 2.8. GBA1 crystals 45
Figure 2.9. Binding of KY170 and KY353 to GBA1 49
Figure 2.10. Binding of acyl aziridine KY358 to GBA1. 50
Figure 2.11. SDS-PAGE gel showing labelling of GBA1 by the β-galactose configured aziridine 52
Figure 2.12. Binding of galacto-configured TB562 to GBA1 53
Figure 2.13. Binding of TB562 to GBA1 superimposed with the binding of KY170. 54
Figure 2.14. Binding of KY358 to GBA1. 55
Figure 2.15. MDW933 present in the structure of GBA1 56
Figure 2.16 Rotation of the MDW933 ligand 57
Figure 2.17 Surface model GBA1 58
Figure 2.18. Design of putative bi-substituted GBA1 probe based upon 3D structure. 59
Figure 3.1 A reaction schematic for a retaining β-glucosidase. 66
Figure 3.2 The cyclophellitol-aziridine activity based probed KY375 71
Figure 3.3. A section from a multi-alignment of GH116 β-glucosidase protein sequences. 74
Figure 3.4 SDS-PAGE gel showing expression of bacterial GH116 enzymes 75
Figure 3.5 The elution of the TxGH116N-His_6 protein from a HisTrap nickel column. 77
Figure 3.6. The elution of the TxGH116N-His_6 protein from a size exclusion gel filtration column 78

Figure 3.7. SECMALLS of 125 nM TxGH116N and TxGH116C.. 80
Figure 3.8 Tagging of active retaining β-glucosidases in the whole cell lysate 81
Figure 3.9. Crystals of TxGH116N. 82
Figure 4.1. Capturing the nucleophile in the active site of the TxGH116 enzyme. 89
Figure 4.2 Activity-based TxGH116 probes and inhibitors studied at a structural level in this thesis 90
Figure 4.3. Purification of TxGH116C by nickel chromatography. 96
Figure 4.4. Purification of TxGH116C by gel filtration. 97
Figure 4.5. The 3D-fold structure of TxGH116N 101
Figure 4.6. TxGH116C crystals.. 102
Figure 4.7. The active site of TxGH116C in complex. 105
Figure 4.8. Presence of a C6-hydroxymethyl group affects the ability of a ligand of inhibit TxGH116. 106
Figure 4.9. Castanospermine in the active site of TxGH116C. 107
Figure 4.10. Isothermal titration calorimetry performed using TxGH116C and castanospermine 108
Figure 4.11. KY358 bound to the active site of TxGH116C. 109
Figure 4.12. KY375 in the active site of TxGH116C. 110
Figure 4.13. Surface view of TxGH116C bound to KY375 111
Figure 4.14. The structure of TxGH116 compared to other retaining β-glucosidase families 112
Figure 4.15. The 3D-fold structure of TxGH116N in comparison to that of GH52 family proteins.. 113
Figure 4.16. A comparison of the active sites of TxGH116N and GtGH52. 114
Figure 4.17. Human GBA2 model and known pathological mutants 116
Figure 4.18. The new activity based probe, designated "APB 4". 118
Figure. 5.1 Bifunctional GBA1 inhibitors under development in Overkleeft group.. 120
Figure 5.2. An exciting new direction for GBA1/2 inhibitor 123
Figure 5.3. Cyclic sulfates: a new retaining glycosidase inhibitor. 124
Figure 5.4. TSA high throughput screening of TxGH116C. 125
Figure 5.5. Circular dichroism spectroscopy of TxGH116C 127

Table of Tables:

Table 1.1. Pathological outcomes of Gaucher's disease 9
Table 1.2 Certain GBA1 mutations and Gaucher Disease Phenotype 11
Table 2.1: Published 3D structures of human GBA1 38
Table 2.2: Data collection and refinement statistics for human GBA1 structures described in this chapter 46
Table 3.1 GBA enzymes and their roles 63
Table 3.2 Mutations in GBA2 and their associated pathologies 65
Table 3.3 GBA2 sequence identity across species 67
Table 4.1 GH116 Subfamilies 87
Table 4.2 Data collection and refinement statistics for the TxGH116N structure described in this chapter 99
Table 4.3 Data collection and refinement statistics for TxGH116C structures described in this chapter 103
Table 4.4. Pathogenic GBA2 mutations and their effects 115

Acknowledgements

There are so many people that I would like to thank for their assistance over the course of my PhD. First and foremost, my supervisor, Gideon Davies, for providing me with the opportunity to take part in such a fascinating project, not to mention his endless patience, and unlimited support, over the past four years. I would also like to thank Marek Brzozowski for his help and suggestions during my Thesis Advisory Panel meetings.

Huge gratitude also goes out to all the members of the Davies group, both past and present, for their help and friendship. I would especially like to thank Wendy Offen for all of her help with protein crystallisation and soaking experiments, and Andrew Thompson for first showing me around the lab. In addition, special thanks must go to Liang Wu for his help throughout my PhD - I am truly grateful for his time and patience (and willingness to answer even my most minor and ridiculous questions!)

Some specific "scientific" thanks go to Andrew Leech in the York Technology Facility for his help with multiple SEC-MALLS experiments, Simon Grist for all his help in the wet lab, particularly when I first got started, and Louise Haigh for all her help and support in the YSBL. Finally, I would like to thank Tim Kirk for computer assistance (especially the emergency laptop surgery!) and Catherine Jardine for all her advice and administrative care, with special thanks to Johan Tukenburg and Sam Hart for the many hours they spent collecting all the crystal data presented here, both remotely and at the synchrotron. In addition, I am especially grateful to all the researchers supervised by Herman Overkleeft at the University of Leiden, and particularly Martha Artola for all their fantastic work with this project.

On a personal note, I would like to acknowledge a few of the people who have supported me so much during my PhD, especially during times when I doubted myself and my abilities. They made me believe that I could complete this project and their words and encouragement have been most cherished. A huge thank you to Becky, Sherry, Muhi, Claire, Alexandra, Herman, Robin and both Lucys. I would also like to thank my family (my wonderful parents and my siblings - Myszka, Wisia and Edmund) for their unconditional love, support and postcards. *Dziękuję*.

Authors Declaration

All research carried out and presented in this thesis was completed by the author with the exception of collaborative work, which is stated within the main body of the text or is detailed below. I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Work carried out or overseen by others:

- The bacterial GH116-N enzyme constructs in pET28a vectors were designed by Dr Andrew Thompson, University of York, before the commencement of my PhD
- The design and synthesis of the activity based probes discussed in Chapters 2, 3, and 4 was conducted by researchers under the supervision of Prof. Herman Overkleeft, University of Leiden
- Figures **2.4**, **2.5**, and **2.17** were kindly provided by Prof. Gideon Davies, University of York.
- The GBA1 galactose probe labelling information discussed in Chapter 2 and **Figure 2.11** were kindly provided by Dr Liang Wu, University of York
- The TxGH116C construct described in Chapter 4 was kindly provided by Dr. James R. Ketudat Cairns, Suranaree University of Technology, Institute of Science, Protein Structure, Function and Application, Thailand.
- The GBA1 labelling gels and calculated IC_{50} values discussed in Chapter 5 and **Figure 5.1** were kindly provided by Dr Martha Artola, University of Leiden.
- The SEC-MALLS experiments were performed and analysed by Dr Andrew Leech in the York Technology Facility (**Appendix 4**)

Chapter 1. Glucosylceramide metabolism: from 3D structure to the development of selective chemical probes

1. Introduction[1]

1.1 The Role of Glucosylceramide

Amongst the essential structural components of the cell membranes of eukaryotic cells are a host of different sphingolipids and glycosphingolipids. These entities have roles both as essential cellular structure components and as more biochemically active molecules in living organisms.

Glycosphingolipids are found in all kingdoms of life, in the cells membranes of organisms from bacteria to humans. In vertebrate brains, more 80% of glycoconjugates present are in the form of a glycolipid (1).

The first characterisation of a glycolipid, and the identification of sphingosine was performed in 1881 by the German physician Johann L. W. Thudichum (2). Thudichum isolated multiple compounds from brain extracts that he named cerebrosides. One of these compounds, which he named phrenosin and is now identified as galactosylceramide, GalCer, he subjected to acid hydrolysis, resulting in the separation of three distinct components (3).

The galactosylceramide consisted of d-galactose, a fatty acid and a third component of an "alkaloidal nature". Thudichum felt this component presented "many engimas" and he therefore dubbed them "sphingolipids," in a reference to the Sphinx of Greek mythology, famed for their enigmatic nature.

Later, further glycosphingolipids have been identified as they accumulate to a pathological level in various tissues of patients who suffer from lysosomal storage disorders. The sialic acid- containing glycosphingolipid was first isolated from the brain of a sufferer of Tay-Sachs disease by Ernst Klenk in 1942, and called a ganglioside (now identified as GM2-ganglioside) due to its accumulation within the ganglion of the brain (4) . Glucosylceramide was first isolated from the spleen of a Gaucher's disease patient in 1974 (5).

Sphingolipids are made up of a sphingosine backbone with an N-acylated linked sphingosine group (**Figure1.1)**. Glucose or galactose is then linked to the primary hydroxyl group of the sphingosine with a β-glycosidic bond. A glucose linked in this position gives glucosylceramide (GluCer), one of the most common glycosphingolipids (1).

[1] Some of the work in this introduction is described in Breen et al, (2018) eLS. Wiley on-line library.

As a plasma membrane component, the lipid portion is buried in the outer leaflet of the membrane while the sugar moiety can project outwards. In this manner, glycosphingolipids are able to play important roles beyond providing structural integrity in membranes. Ceramide and glycosphingolipids act in many cellular signalling process including cell recognition, development and differentiation (6).

Glycosphingolipid interactions can serve as an initial basis for cell-to-cell recognition, independent of cell surface lectin interactions. Some glycosphingolipids may also provide an interface for viral and bacterial toxins to interact with the cell surface membrane (7,8).

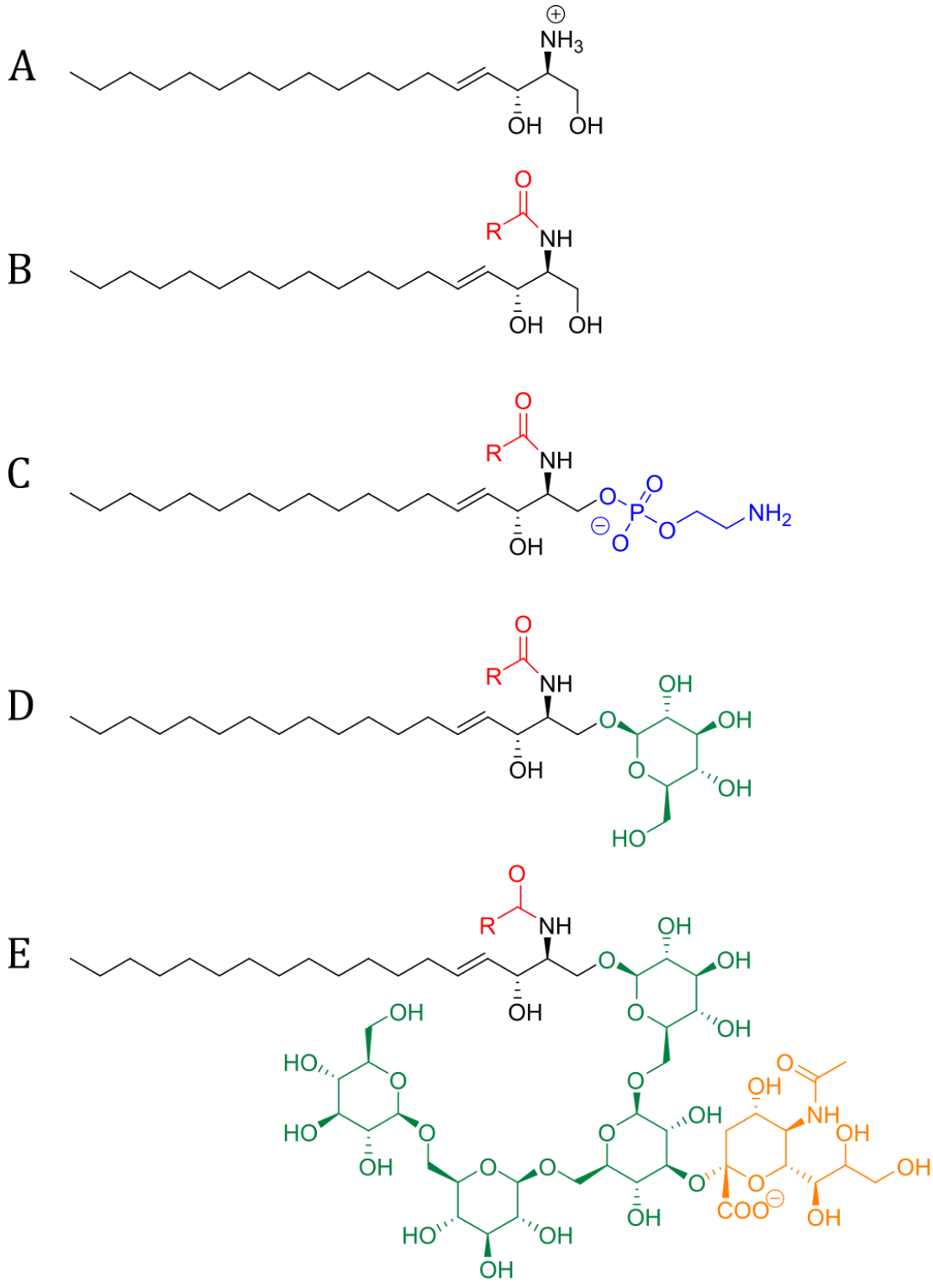


Figure 1.1. The general structures of biological sphingolipids. A) The sphingosine backbone (black). B) A ceramide, the fatty acid residue is shown in red. C) A sphinogomyelin. The phosphoethanolamine group is shown in blue. D) A glucosylceramide with a single sugar residue shown in green. E) An example of a ganglioside - the oligosaccharide is shown in green and sialic acid shown in yellow.

In mammalian cells, GluCer is formed by the glycosyltransferase enzyme glucosylceramide synthase (9) (EC2.4.1.80) by the glycosylation of ceramide using UDP-Glc as the sugar donor. Ceramide is itself an active messenger molecule, playing a role in tumour suppressing and effecting signalling events to drive apoptosis (10). The regulation of internal ceramide concentration therefore must be maintained to prevent adverse effects that may result from elevated ceramide levels. Cells are able to clear excess ceramide rapidly by converting it to GluCer (11) and maintenance of the correct amount of the lipid inside the cells is vital (12). This conversion of ceramide to GluCer by glucosylceramide synthase is also used by cancer cells to neutralize the downstream cell-death signals that are initiated by ceramide (13,14).

GluCer catabolism is primarily through the lysosome (15) where the β-glycosidic bond is hydrolysed to give ceramide and glucose by an acid-β-glucosidase enzyme, GBA1 (**Figure 1.2**). As a consequence of amino acid point mutations in GBA1, GluCer can accumulate in lysosomes, leading to Gaucher's disease.

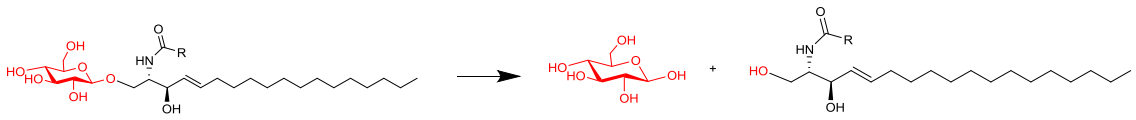


Figure 1.2. The reaction catabolised by GBA1, GBA2 and GBA3. Glycosylceramide is hydrolysed to D-glucose and N-acylsphingosine.

Human cells in fact contain three acid-β-glucosidases capable of degrading GluCer. The lysosomal GBA1 (whose mutations lead to Gaucher's disease), the non-lysosomal cytosolic glucosylceramidase GBA2 and the "Klotho-related protein" GBA3 (16). All three "GBA" enzymes are capable of catabolising glucosylceramide to its constituent parts (**Figure 1.2**).

1.2 Lysosomal Storage Disorders and Gaucher's Disease

Lysosomal storage diseases, as a group of more than 70 disorders, occur in approximately 1 in 7,500 live births (17), however the rates of individual disorders are far fewer. The prevalence of Gaucher's disease, the most common lysosomal storage disorder, is estimated at between 1 per 20,000 and 1 per 60 000 live births in the general population, although this increases in certain high-risk populations. Sialidosis, the deficiency of α-N-acetyl neuraminidase, occurs in approximately 1 per 4.2 million live births worldwide (18).

Although disease phenotypes can differ highly, all lysosomal storage disorders share a common characteristic in that a genetic mutation in a gene encoding for a lysosomal enzyme leads to a loss of function in that enzyme and the subsequent abnormal build-up of substances within the lysosome.

Philippe Charles Ernest Gaucher, a French dermatologist, first described the disease that became known as Gaucher's disease in his MD thesis in 1882 when he encountered a female patient who died of hepatosplenomegaly and cachexia (wasting) (17). By 1912 it was recognised that the disease was multi-systemic, chronic and familial although the link to the lysosome was not made for another 50-odd years. Gaucher's disease is characterised by the build-up of glucosylceramide (GluCer) within the lysosome resulting in enlarged organs, splenomegaly, hepatomegaly and, in some cases, neurological disorders.

The lysosome was discovered in 1955 by Christian de Duve, who, along with others, also described the role it played in intracellular digestion (19). Subsequently, Henri-Géry Hers and co-workers were the first to identify a genetic lysosomal storage disease, called glycogen storage disease type II; now more commonly known as Pompe disease (20).

In 1974, Brady et al identified glucocerebrosidase as the key enzyme deficient in Gaucher's disease patients and successfully intravenously administered unmodified, purified human enzyme to two patients (21). This treatment approach, enzyme replacement therapy, was to become the standard for most lysosomal storage disorders throughout the 20th century (22).

1.3 Gaucher's disease and GBA1

Gaucher's disease is a lysosomal storage disorder linked to mutations in the gene encoding for the lysosomal enzyme acid β-glucocerebrosidase, GBA1. Occurring in 1 in 20,000 live births worldwide, there are huge phenotypic variations within those suffering from

Gaucher's disease. This has led to the disease being classified in to three pathological types – Type I, II and III (28) (**Table 1.1**)

GBA1 (EC 3.2.1.45) is a 497 amino acid membrane-associated glycoprotein belonging CAZy family GH30 (CAzy is described later in this thesis). GBA1 degrades GluCer into glucose and ceramide within the lysosome, which can then be recycled back into the cytoplasm. When mutations in both alleles encoding GBA1 lead to a reduction or loss of function in the enzyme, GluCer can build up in the lysosomes and lead to deleterious effects. All GBA1 mutations and their resulting disease pheonotypes are grouped together as Gaucher's disease.

The gene encoding GBA1 is located on chromosome 1q21 and Gaucher's disease is inherited in an autosomal recessive fashion. It is most common in the Ashkenazi Jewish population where approximately 6.8% of the population is heterozygous for Gaucher's disease and the expected birth frequency is 1 in 1000. (3) The disease also appears at a higher frequency in the Norrbottnian population located in northern Sweden (23).

In certain high risk populations, pre-marriage genetic carrier tests for Gaucher's disease (and another lysosomal storage disorder, Tay-Sach disease) have been offered (24). Commercial genetic analysis companies such as 23andMe, Inc. (https://www.23andme.com) (4) also now offer identification of certain common mutations that an individual can carry - before they may pass it on to their children.

There are three distinct pathological phenotypes of Gaucher's disease(25), characterised by clinical appearance; importantly the disease can manifest with neuropathic or non-neuropathic symptoms (**Table 1.1**). Different mutations in the GBA1 gene may determine the remaining activity level of the β-glucocerebrosidase enzyme, and, to a large extent, the phenotype.

The sub-types of Gaucher's disease are categorised by a patient's symptoms or time of onset rather than the particular amino acid mutation. A host of other factors appear to be able to mitigate the pathological outcome of Gaucher's disease, from the level of GBA2 activity, to onset of other disease in the patient (such as type 1 diabetes mellitus)(26). The effect described as "synergistic heterozygosity", where partial defects or variations occur in a complex metabolic pathway leading to a loss of clear genotype/phenotype correlation has been applied to many metabolic diseases, including Gaucher's (27,28).

Patients with the same mutations, even family members, are seen to have different complications or reactions to treatment. This is due to modifier genes (29), reciprocal recombination (30)and the action of GBA2, the second non-lysosomal acid β-

glucocerebrosidase enzyme. There is at least one recorded case of monozygotic (identical) twins, who exhibit very distinct Gaucher's phenotypes despite carrying the same N188S/N188S genotype (26).

TABLE 1.1. PATHOLOGICAL OUTCOMES OF GAUCHER'S DISEASE

	Type 1: Adult, Non-neuropathic	Type 2: Infantile, Acute Neuropathic	Type 3: Juvenile, Chronic Neuropathic
Age of onset	Young adults to early middle age	Infants	Children and young adults
Occurrence in populations	1 in 100 000 in general population Up to 1 in 450 in Ashkenazi Jewish population	1 in 100 000 live births, No particular ethnicity	1 in 50 000 live births, seen in Norrbottnian population
Symptoms	No CNS involvement, liver, spleen, bone all effected (as in all types)	Early CNS problems and brainstem abnormalities	Later onset of CNS problems including seizures and mental deterioration.
Outcome	Dependent on residual GBA1 activity, may be very mild	Death in infancy (<2 years)	Progresses slower, symptoms become more severe later in childhood.
GBA1 activity	Some, much reduced	Very little	Very little

In Type I Gaucher's disease, there is some residual activity of the enzyme(31). Type 1 makes up over 90% of the cases and is considered non-neuropathic; patients do not exhibit central nervous system involvement and disease is typically diagnosed later in life. Some Type 1 patients may present no symptoms until very late in life, or may even only be diagnosed after a relative is identified as a Gaucher's patient and they too also undergo genetic screening (32).

Patients with Types 2 (acute infantile onset) and 3 (juvenile or early adulthood onset) Gaucher's disease suffer neurological symptoms, as well as the bone lesions, anaemia and other complications seen in Type 1. Only Type 1 and the rare non-neurological Type 3

Gaucher's disease are treatable with enzyme replacement therapy. This is due to the inability of the large protein drug treatment to cross the blood-brain barrier.

Although there are certain mutations generally associated exclusively with "milder" Type 1 Gaucher's disease, mutations such as L444P have been seen in patients with all three Gaucher's types. Patients with heterozygous mutations can manifest unpredictable disease phenotypes, although patients with at least one N370S will not develop primary neurologic disease (33,34).

GBA1 is a 536 amino acid protein and more than 400 different mutations in the enzyme are known to lead to Gaucher's disease. Most of these are missense mutations (35), and novel mutation are continuously being identified in patients (36). Many of these mutations are carried in a single family line (37). The active site amino acids, the nucleophile E340 and the acid/base E235, are located in the catalytic domain, consisting of a $(\beta/\alpha)_8$ TIM barrel, however disease causing mutations have been identified throughout the protein (38).

The predominant point mutation is N370S, which accounts for 75% of the disease cases in Ashkenazi Jewish populations and 30% in non-Jewish population. Genetic analysis suggests that the N370S mutation may have originated with a single progenitor, the frequency in modern populations resulting in a severe population size bottleneck between 1100 CE and 1400 CE (39). Patients with this N370S mutation will not show neuropathic involvement in their disease progression, and are frequently diagnosed later in life (40). *In vitro* studies of 52 recombinantly expressed GBA1 mutants, including the Type 1 N370S mutation show that this mutation gene produces a stable enzyme with some activity against GluCer. (41)

Clinical outcome of Gaucher's disease is dependent on the levels of active GBA1 enzyme present in the patient's lysosome; diagnosis and treatment requiring regular quantification of this. Patients with just 15% of normal, healthy GBA1 activity levels are normally asymptotic, so if the trafficking of mutant GBA1 to the lysosome can be increased and suitably monitored, Gaucher's disease may be ameliorated (42).

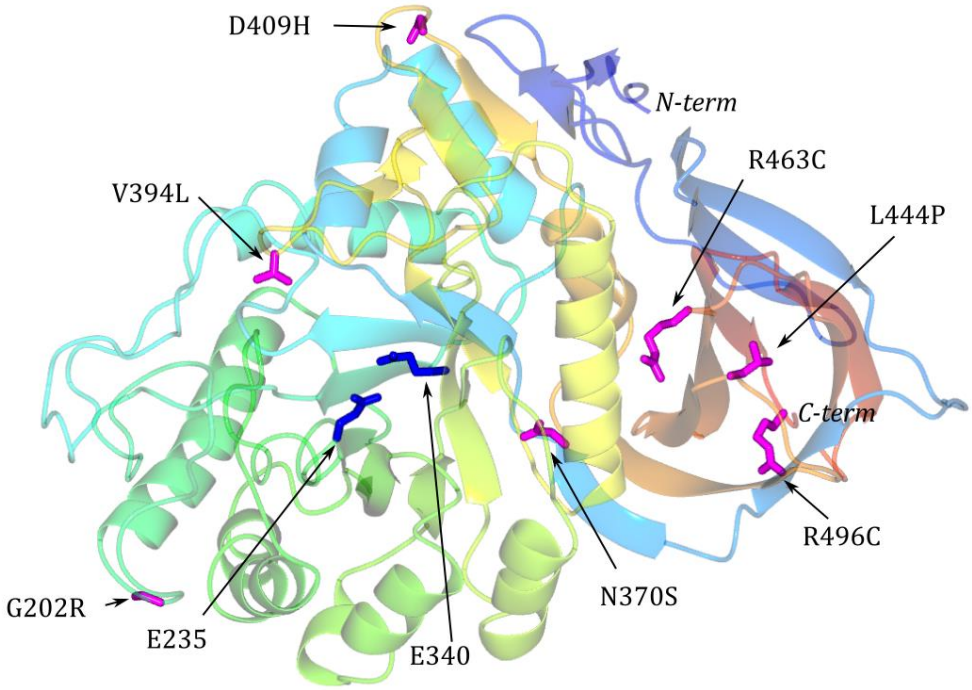


Figure 1.3 The three-dimensional structure of GBA1 (as discussed in Chapter 2) showing the location of the six most common missense mutations (magenta. The active site amino acids (Nucleophile E340, Acid/Base E235) are shown in blue.

TABLE 1.2 CERTAIN GBA1 MUTATIONS AND GAUCHER DISEASE PHENOTYPE

Mutation	Gaucher's disease type
G202R	Type 2 (43)
N370S	Type 1 (44)
V394L	Type 1 or Type 3 if heterozygous with L444P mutation (34,45)
D409H	Type 3 or Type 1 if heterozygous with L444P mutation (33,46)
L444P	Type 1, 2 or 3, dependent on heterozygosity (47,48)
R463C	Type 1 or Type 3 if heterozygous with L444P mutation (33)
R496C	Type 1 (49)

The human cDNA encoding GBA1 was first cloned in 1984 by Ginns et al (50), which eventually lead to the patenting of recombinatly expressed GBA1 from Chinese hamster ovary cells (CHO cells) as Cerezyme® in 1994 by Genzyme(22). Identification of the sequence has allowed gene knock out experiments and in the last three decades, the use of GBA-knockout mouse models (rather than those with chemically induced Gaucher's disease), and even mice with particular point mutations such as N370S have also progressed our understanding of the disease (51,52).

The unliganded three dimensional structure of human GBA1 was initially determined by the Sussmann group in 2003 (38). GBA1 was found to consist of three domains, with the catalytic site located in domain II. The active-site nucleophile had previously been identified as Glu340 by Miao et al in 1994 through the use of a mechanism based inhibitor and electrospray tandem mass spectrometry (53). Modelling with other GH30 family proteins and subsequent site-directed mutagenesis confirmed Glu235 as the catalytic acid/base (54,55). The 3D structure of GBA1 is discussed in further detail in **Chapter 2**.

Gaucher's disease mutations are found across all three domains of the GBA1 enzyme (**Figure 1.3**), not exclusively within the catalytic domain. These mutations can have a range of effects on the enzyme, from the disruption of inter-domain interactions to the prevention of the correct glycosylation of the GBA1 enzyme.

Correct glycosylation of GBA1 is necessary for the cell protein processing systems to accept the enzyme and for the function of the fully active enzyme(56). The development of a catalytically active GBA1 enzyme is dependent on it being correctly glycosylated, particularly the glycosylation of the Asn19 site (57,58). In fact, GBA1 has five putative N-glycosylation sites, four of which undergo posttranslational N-linked oligosaccharide addition (59,60).

N-glycan processing of a nascent polypeptide is carried out in endoplasmic reticulum and the Golgi body. Oligosaccharyltransferase must recognise the consensus sequence (Asn-X-Ser/Thr) within the protein and transfer the precursor glycan to the polypeptide in the endoplasmic reticulum lumen. If the protein is folded correctly, the three glucose residues of the precursor glycan are removed by glucosidase I and II. The removal of the final glucose residue signals that the glycoprotein is ready for transit from the ER to the cis-Golgi.

If the protein is folded incorrectly and either oligosaccharyltransferase fails to recognise a glycosylation site or the final glucose resides are not removed, the GBA1 enzyme will be unable to leave the endoplasmic reticulum. Promotion of this correct folding is the underlying concept of pharmacological chaperone therapy (see section **1.5**) (61).

1.4 GBA1 in other disease

The link between Gaucher's disease and Parkinson's disease first was made from a diagnostic perspective – patients with certain non- neuropathic forms of Gaucher's disease will develop early-onset Parkinsonism. This suggested that perhaps "non-neuropathic" Gaucher's disease still had implications upon the CNS (62).

Parkinson's disease is the second most common neurodegenerative disorder, with a lifetime risk of 3 % in the USA (63), whilst 25 % of patients with Gaucher's disease have a family history of Parkinson's (64). This carries the implication that even heterozygous Gaucher-carrier status may increase the risk of Parkinson's (65).

Between 2.3%–9.4% of patients with Parkinson's disease (increased to 11%–31% in Ashkenazi Jews) carry a GBA1 mutation (66). Two GBA1 mutations, E236K and T369M, which do not appear to cause Gaucher's disease, are more common in patients with Parkinson's (67).

The key component of Parkinson's pathogenesis is α-synuclein, an intrinsically disordered protein. α-synuclein is abundant in brain and able to interact with lipid membranes (68). However it has a susceptibility to aggregation in the form of amyloids in Lewy bodies, one of the pathological features of Parkinson's disease (69).

There appears to be a relationship between the levels of GCase and the behaviour of α-synuclein in cells. When GBA1 is inhibited in human neurons *in vitro,* and GluCer levels subsequently increase, α-synuclein accumulated, resulting in neurotoxicity. In purified form, GluCer directly influenced amyloid formation in α- synuclein. In a reciprocal feedback loop, increased cellular levels of α-synuclein lead to a reduction of lysosomal GBA1 activity due to the disruption of intracellular trafficking (70).

This suggests that treatments that help to promote targeting of GBA1 to lysosomes in Gaucher's patients with presenting with non-neuropathic Type 1 can be expected to diminish the formation of toxic amyloids and potentially prevent Parkinson's onset (64,70,71).

1.5 History of enzyme replacement therapy

Initially, treatment of Gaucher's disease was limited to symptom management; splenectomy was frequently performed for Type 1 patients and patients with Types 2 and 3 had poor prognosis (32). However, by 1965, Dr Roscoe Brady et al had identified that Gaucher's disease was caused by an inherited enzyme deficiency and in 1968 Weinred et al identified the lysosomal nature of the enzyme. This gave clinical researchers a clearer target for disease treatment.

Gaucher's disease the first lysosomal storage disorder to be successfully treated with enzyme replacement therapy (4). Roscoe Brady and colleges were able to successfully treat two patients with purified, unmodified GBA1 in 1974(72). However, although both patients showed a reduction in glucosylceramide accumulation in liver biopsies after treatment, neither patient was able to be stabilized over a long period of time.

Achieving a large enough quantity of purified GBA1 enzyme to viably treat a patient was the main hurdle. This was particularly difficult before the production of recombinant GBA1 proteins and it took almost two decades from the original two person patient trials in 1973 (72) to a commercial GAB1drug being marketed.

The initial method of purification was from human placenta using affinity column chromatography and was developed by Peter Pentchev and colleagues in 1973 (73). As well as the production of protein for therapeutic uses, this increased access to GBA1 allowed researchers to further characterise the enzyme(74).

Post-purification modification of the carbohydrate structure of the glycoprotein was performed, so that it would be better targeted to a patient's macrophages. This allowed physicians to reduce the volume of GBA1 enzyme required per patient. Deglycosylation of the purified GAB1 was performed with the enzymatic removal of N-acetylneuraminic acid, galactose and N-acetylglucosamine to give a mannose-terminal product, which showed an increase liver macrophages in rat models (75).

The modified β-glucocerebrosidase, termed alglucerase, sourced from large volumnes of human placental tissue by the Genzyme Corporation, Cambridge, MA, USA and supplied as a citrate buffered solution for intravenous use was called Ceredase®. Ceredase® was approved for use in Gaucher's disease treatment in 1991 but has since been withdrawn from the market due to the arrival of recombinant expression systems that are capable of producing similar therapeutic proteins without the associated concerns of disease transmission during tissue harvesting and with comparatively lowered costs (22).

By 1992, around 30 percent of all placentas from births in the United States, and tissue from around 70 percent of placentas from Europe, were sent to a Genzyme processing plant in New England to meet demand for GBA1, among other purified enzymes (76). Some patients were forced to travel incredible distances, including internationally, to receive treatment.

Once alglucerase was on the market for Gaucher's treatment, and successfully treating patients with the non-neuropathic Type 1 form of the disease, the effect of enzyme replacement therapy was studied on patients with the chronic neuropathic form (Type 3)

of the disease(77). Although high doses of the enzyme treatments (at first Ceredase® and later Cerezyme®) would reverse the general symptoms of the disease, it became clear that intravenously administered enzyme could not cross the blood-brain barrier and therefore would not alleviate the neurological symptoms.

Imiglucerase is a recombinant Chinese hamster ovary cell produced analogue of human β-glucocerebrosidase, again treated with exoglycosidases to expose terminal mannoses and facilitate macrophage targeting. Imiglucerase also differs from the placental enzyme at amino acid position 495 where arginine has been replaced with a histidine(22).

Imiglucerase was introduced to the market as Cerezyme® (Genzyme Corporation, Cambridge, MA, USA), a freeze-dried medication containing imiglucerase to be given intravenously for long-term enzyme replacement therapy, in 1995. This was the first recombinant protein treatment for Gaucher's disease and remains the most commonly used. The specific activity of highly purified human enzyme is 890,000 units/mg; a typical patient dose being 2.5-5 U/kg biweekly (78).

There are now three human β-glucocerebrosidase analogues which have been approved for use in humans. As well as imiglucerase, there are velaglucerase alfa (VPRIV®, Shire Plc, Dublin) which is purified from a human cell line and taliglucerase alfa (Elelyso®, Protalix Biotherapeuteutics, Karmiel, Israel) which, purified from carrot cells, was the first plant cell-sourced drug to achieve FDA approval. Elelyso® also has the advantage of being a fully kosher medically treament.

Due to the small number of patients and high cost of manufacture, recombinant human β-glucocerebrosidase is one of the most expensive drugs sold, with an average annual cost to U.S. Gaucher's patients of $200,000 (79). For many patients, it is necessary that the financial burden is covered by charities such as the National Gaucher Foundation's CARE program in the USA.

Enzyme replacement therapy (ERT) has some major disadvantages. As well as the incredibly high cost of treatment, it is ineffective for patients suffering from neurological effects. Intravenous injections of the drug treatment must be given once a fortnight, they can be painful and leave patients reliant on frequent hospital assess.

Whilst ERT is generally well tolerated in Type 1 Gaucher's patients, 13 % report adverse reactions to the treatment process itself, especially infections and abscesses at the site of injection. Some patients may also develop IgG antibodies to the recombinant enzyme, and in less than 1% of patients this can lead to anaphylactic reactions (80)

These factors drive a need to develop Gaucher's treatments that can be taken orally, tolerated by those who are unable to take ERT and are able to cross the blood-brain barrier.

The first alternate treatment to enzyme replacement therapy was the use of the iminosugar miglustat (N-butyldeoxynojirimycin, NB-DNJ) which was marketed as Zavasca® (Oxford GlycoSciences/Actelion, Allschwil, Switzerland) and first approved for use as a Gaucher's disease treatment in 2002. Miglustat is an inhibitor of the enzyme glucosyltransferase, and works by the principal of substrate reduction therapy. By reducing the biosynthesis of GluCer, the accumulation in cells is reduced.

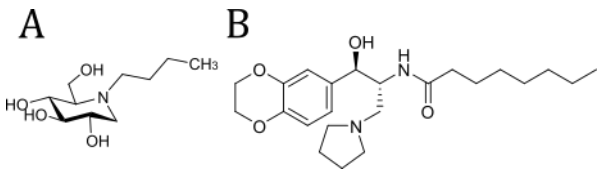


Figure 1.4. Substrate reduction therapy drugs A) Miglustat (Zavasca) B) Eliglustat (Cerdelga)

Although miglustat is licenced for use in Gaucher's patients who are unable to receive enzyme replacement therapy, serious side effects from the treatment are common, including diarrhoea and tremors. Many patients given Zavasca do not tolerate the treatment well and can either require extensive medical support alongside the medication or eventually refuse medication (81,82).

A second substrate reduction therapy treatment was introduced in 2014. Eliglustat (Cerdelga®, Genzyme Corp, Cambridge MA, USA) acts to inhibit glucosylceraminde synthase and thus also reduce GluCer accumulation. Unlike miglustat, eliglustat does not interact with any intestinal enzymes, thereby avoiding some of the main side effects of miglustat that lead to patient non-compliance (83).

Whilst both these small molecule substrate reduction therapy drugs are able to cross the blood-brain barrier, miglustat has some undesirable neuropathic side effects including tremors. Eliglustat is transported out of the central nervous system instantly by the transporter Pgp-1 and no alleviation of Gaucher's disease symptoms is seen (83,84).

As lysosomal storage disorders are monogenic diseases, there has been some consideration of gene therapy as a treatment approach. Sandhoff and Tay-Sachs diseases are both fatal lysosomal storage disorders with no effective treatments, however feline

and mouse models of each disease have been used to model gene therapy as disease treatment (85,86). Lentiviral gene therapy has been shown to cure Type 1 Gauchers disease in mice but it is a long way to human clinical trials (87).

The use of miglustat as a treatment option for other rare lysosomal storage disorders is currently under investigation. The drug has been approved for use in patients with Niemann-Pick disease type C and its use has begun in patients with Sandhoff disease (88) following success in feline and murine models (89-92).

1.6 Chaperone therapy

The third therapeutic pathway for Gaucher's disease is pharmacological chaperone therapy (PCT). Gaucher's disease typically arises when the mutant form of GBA1 is recognised as misfolded in whilst still inside the endoplasmic reticulum. Here it is targeted for degradation by the endoplasmic reticulum- associated degradation (ERAD) system through the proteasome rather than being transported to the lysosome, thus enzyme levels in the lysosome are reduced (93).

In PCT, small molecules are able to stabilise the mutant forms of GBA1 and restore protein trafficking (94) (**Figure 1.5**). As proteins are produced by the ribosomes and secreted in to the endoplasmic reticulum they are in an unfolded state. In the ER a quality control system ensures that only correctly folded and assembled proteins are able to be transported to the Golgi apparatus to be processed further. Improperly processed proteins are held in the ER by chaperone proteins such as *BiP* or *calnexin* and eventually transported back out for degradation by proteasomes (61).

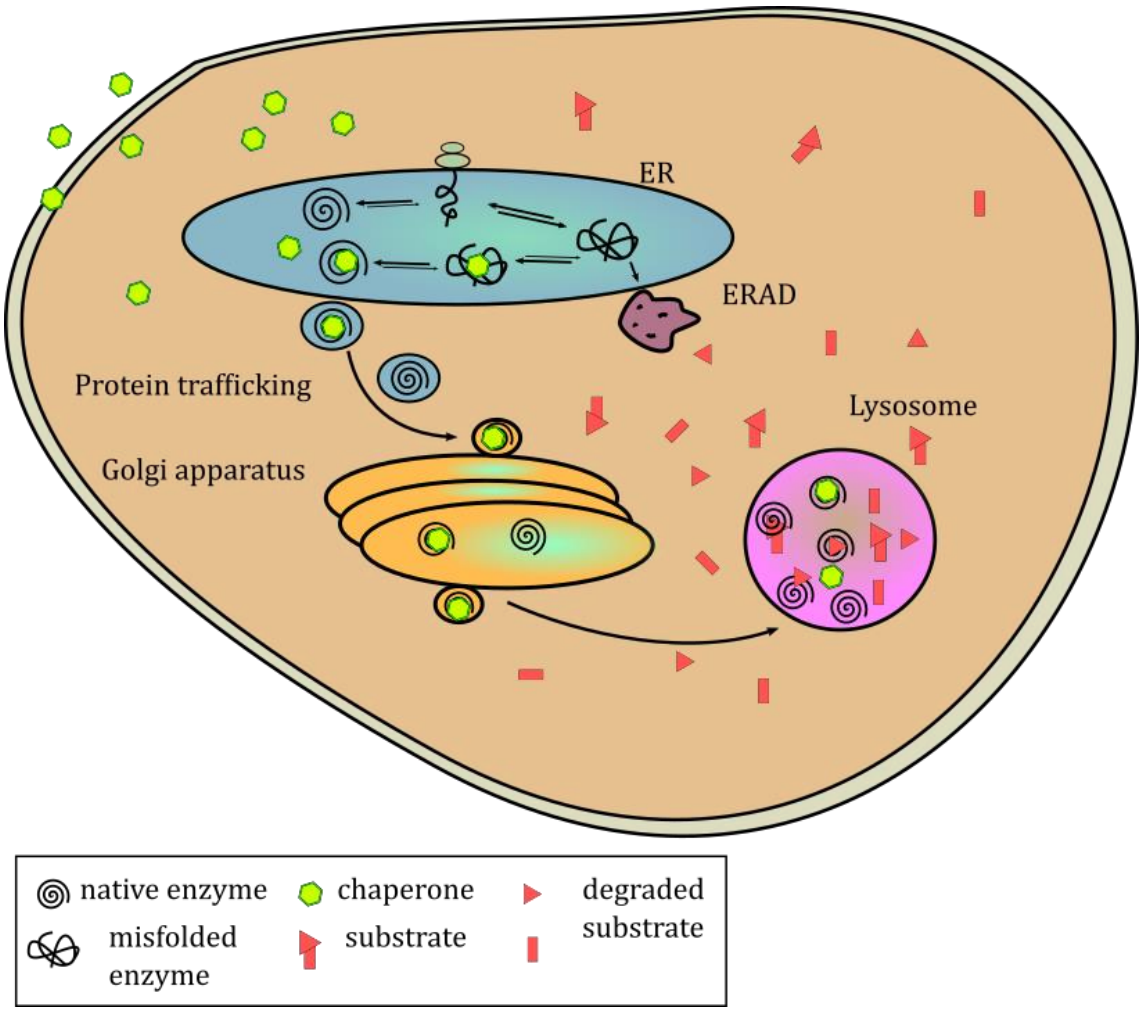


Figure 1.5 A general overview of chaperone therapy. Pharmacological chaperones are able to promote correct folding of the active form of the enzyme, leading to increased trafficking to the lysosome.

As previously mentioned, many mutant GBA1 forms show activity on GluCer. Often these mutants forms exhibit decreased stability at neutral cellular pH, as experienced in the ER, but become more stable at the reduced pH conditions of the lysosome (41,95). These mutant forms will occasionally fold correctly, and therefore be processed normally by the cell, resulting in the low levels of enzyme activity seen in some Gaucher's patients.

Chemical chaperones can act to stabilise potentially active proteins against misfolding, increasing the probability of their achieving the correct fold. This then avoids protein degradation and promotes the correct trafficking through the endoplasmic reticulum to the Golgi, with the enzyme eventually arriving in the lysosome. Here, the enzyme can act with enough activity to achieve a therapeutic effect (42).

Chaperone therapy, sometimes described as enzyme enhancement therapy, makes the use of molecules that are small enough to enter the central nervous system, thus able to treat the neuropathic symptoms that ERT cannot affect. The refinement of Gaucher's disease treatment with small molecule chemical chaperones has a huge impact for patients for whom ERT has proved unsuitable, especially for those with neuropathic symptoms as the small molecules are able to cross the blood-brain barrier and reduce storage cell load in the brain (96).

An active-site specific chaperone, such as a competitive enzyme inhibitor, would bind to the catalytic domain of the enzyme and, once the enzyme has been transported to the lysosome, be replaced by the substrate. In the lysosome, the substrate is at a very high concentration, especially in Gaucher's disease patients where it has been accumulating. However, because of this exchange of inhibitor and substrate, potential chaperone treatments given at too high a concentration could lead to a non-reversed binding and therefore complete inhibition of the enzyme.

PCT is therefore very dose dependent, close monitoring of the levels of active GBA1 in a patient's lysosome is necessary for effective clinical treatment. The use of activity based probes can allow physicians to quantify only active enzyme at any given time and adjust dosage in response to these readings.

In 2005, Alfonso et al (97) showed that the ceramide-specific glucosyltransferase inhibitor miglustat was also able to increase the activity of certain (though not all) GBA1 mutants and wild-type GBA1 in monkey kidney cells. This suggested that the varied success of miglustat treatment in Gaucher's disease patients was not only down to glucosylceramide synthase inhibition but also the stabilization of the mutant GBA1 protein already present in the patients' cells.

More recently, the use of other small molecule chemical chaperones as Gaucher's disease treatment has seen some success. The N370S GBA1 mutant (the most common mutation in patients with Ashkenazi Jewish ancestry (33)) was shown to be stabilized at neutral pHs by N-nonyldeoxynojirimycin (NN-DNJ) leading to a two-fold increase in the enzyme's activity as it was more effectively trafficked to the lysosome and avoided degradation (98). Doses of isofagomine tartrate, an acid β-glucosidase active site inhibitor, were given to mice homozygous for a variety of Gaucher's disease mutations (including certain neuropathic forms) and the mice were shown to respond positively to treatment over the course of several weeks (99).

1.7 GBA1 inhibitors

Glucosidase inhibition can be competitive or covalent – the former being ideal in the design of pharmacological chaperones and the latter for activity based probes. The most effective GBA1 inhibitors are compounds that mimic functional and/or conformational aspects of the parent substrates (i.e. beta-glucosides) (**Figure 1.6**).

Numerous glucosidase inhibitors are found in nature, produced by plants, fungi and bacteria (100). For example, cyclophellitol first isolated from the mushroom strain *Phellinus sp* in 1990 (101). Deoxynojirimycin (DNJ) can be isolated from brewing certain strains of mulberry tea leaves, at concentrations effective to suppress postprandial blood glucose levels, by inhibiting intestinal glucose absorption (102,103).

Many more GBA1-specific inhibitors have since been designed and synthesised in the lab, often taking these naturally sourced inhibitors as inspiration. The most well-known GBA1 inhibitors and potential drug chaperones are imino/aza sugars based around the deoxynojirimycin core (104), and isofagomine (105) **(Figure 1.6.A)**.

As isofagomine is orally available and able to cross the blood/brain barrier, there has been a lot of interest in it as a viable GAB1 chaperone (IC_{50} = ~88 nM for wild type GBA1 (99)). isofagomine has been studied as an effective pharmacological chaperone for both the L444P neuropathic mutant form of GBA1 in fibroblasts (106) and the V394L form in mouse models (96). The structure of GBA1 in complex with isofagomine was published in by Lieberman et al in 2009 (PDB ID: 3GXF) (95).

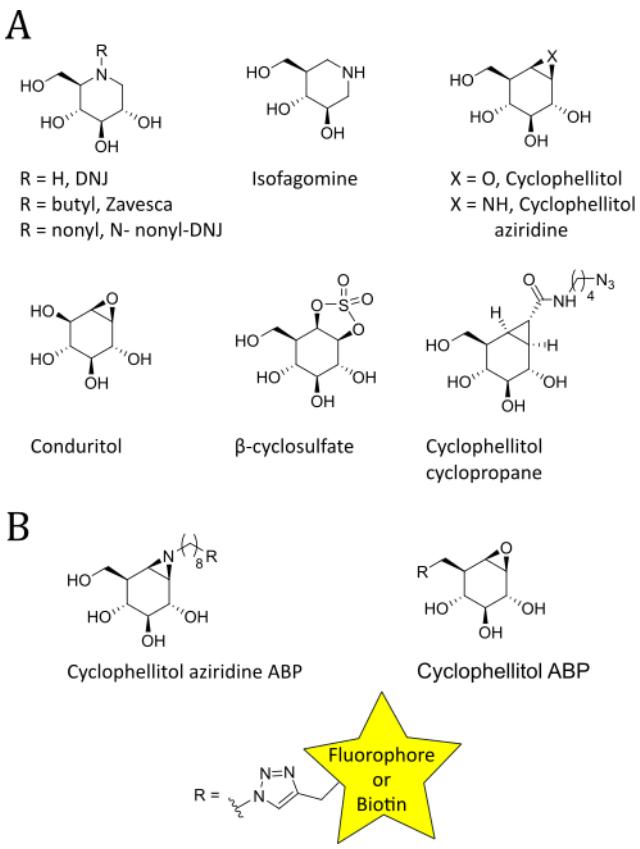


Figure 1.6. An example of some A) inhibitors and B) activity based probes of human lysosomal acid β-glucocerebrosidase, GBA1. Adapted from Breen et al, 2017 (**Appendix 7**)

DNJ is a reasonably potent inhibitor of both retaining and inverting α- and β-glucosidases; this poor selectivity being its main disadvantage. DNJ inhibits both GBA1 and GBA2 (IC_{50} = 250 and 21 μM, respectively) and is also an inhibitor of human acid α-glucosidase (IC_{50} = 1.5 μM) (107). Over recent decades numerous studies have been published, exploring structural and configurational DNJ analogues with an aim to design compounds capable of targeting β-glucosidases alone, and specifically GBA1 (108,109). An important class of structural modifications is the addition of N-alkyl chains.

The SRT drug Zavesca, (*N*-butyl-DNJ, **Figure 1.4, Figure 1.6.A)** is in clinical use for the treatment of Gaucher disease, with GCS inhibition (IC_{50} = 50 μM (107)) as the basis of this therapy. Zavesca is however also a micromolar GBA1 inhibitor (IC_{50} = 400 μM (107)) and the structure of the co-crystal reveals a nice fit of the glucopyranose mimic within the GBA1 enzyme active site. (**Figure 1.7.A,B**) (110).

Along with isofagomine, N-nonyl-DNJ (NN-DNJ) has been studied intensively for application as a pharmacological chaperone. NN-DNJ has been seen to restore protein levels of the N370S mutant GBA1 within fibroblast cells, and binding of the extended N-

alkyl form protected GBA1 against tryptophan digestion, providing evidence of its structural "rigidification" (111).

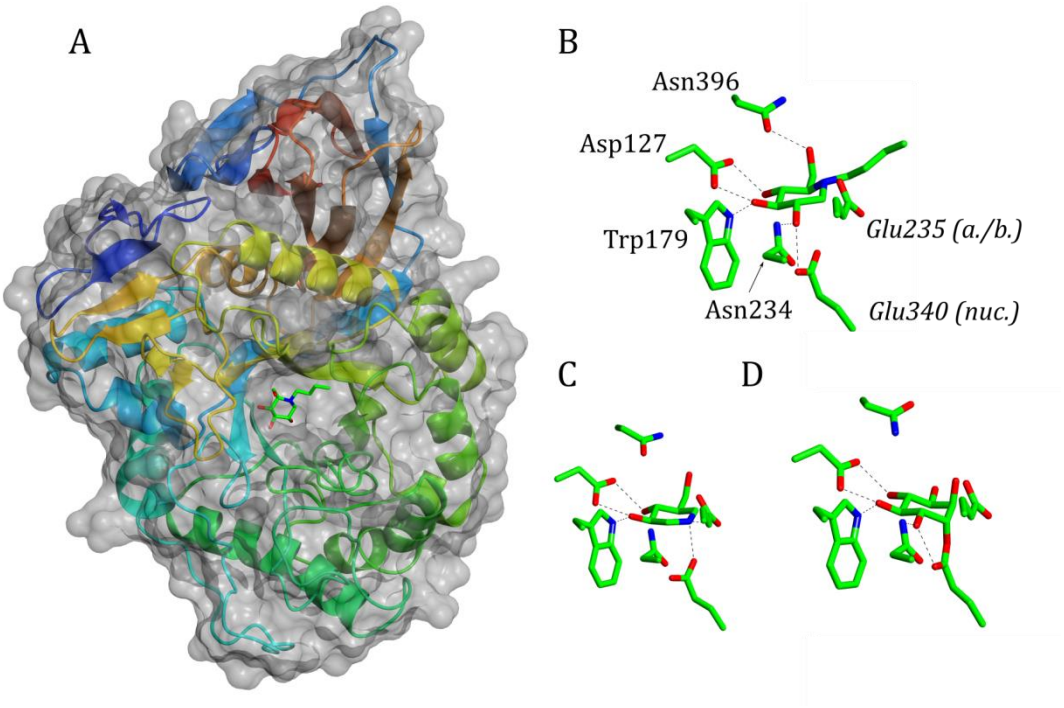


Figure 1.7. The crystal structure of human GBA1, in complex with inhibitors. A) Three dimensional structure of GBA1, showing a molecule of Zavesca in the active site. PDB: 2V3D. B) Active site of GBA1 in complex with Zavesca. Direct interactions between the ligand and protein residues are shown. Residue numbers are annotated (a./b. = acid/base, nuc. = nucleophile. PBD: 2V3D. C) Active site of GBA1 in complex with isofagomine. PDB = 3GXF. D) Active site of GBA1 in complex with conduritol. PDB = 2VT0. Adapted from Breen et al, 2017 (**Appendix 7**)

Cyclophellitol, the natural product isolated from mushrooms, was found to be an effective and irreversible inactivator of β-glucosidases in 1991 by Withers and Umezawa (112). Closely resembling β-glucopyranose in configuration, cyclophellitol is able to neatly position its epoxide moiety within a retaining β-glucosidase active site for nucleophilic attack by the enzyme catalytic nucleophile. The ester adduct formed upon reaction of cyclophellitol with the glycosidase nucleophile is stable, resulting in covalent enzyme inhibition. This makes cyclophellitol a useful "base" for the design of activity based probes for retaining β-glucosidases (**Figure 1.6.B**).

The 3D structures of GBA1 observed with a variety of inhibitors, including potential pharmacological chaperones, have been determined, and a full table of the published

structures is included in **Chapter 2 (Table 2.1).** However prior to this work, the only 3D structures of GBA1 with a covalently-bound inhibitor was the structure with conduritol-β-epoxide (**Figure 1.6.A, Figure 1.7.D**) (PDB ID: 2VT0) (113).

Initial insight into the covalent binding of the β-glucopyranose cyclophellitol was gleened from the structure of the retaining β-glucosidase bacterial homologue TmGH1 from *Thermatoga maritima* (TmGH1) (114) in complex with both cyclophellitol and cyclophellitol cyclopropane (**Figure 1.6.A**).TmGH1 crystals soaked with the non-hydrolysable cyclophellitol analogue show the glucose mimicking portion of the ligand bound within the enzyme active site in a $^{4}H_{3}$ conformation, 'above' the enzyme nucleophile poised for nucleophilic attack (115). When *Tm*GH1 crystals are treated with cyclophellitol itself, a covalent enzyme-inhibitor adduct in $^{4}C_{1}$ conformation can be seen, showing that the epoxide of cyclophellitol has been opened by nucleophilic displacement by the catalytic nucleophile residue (115,116).

Information on reaction itineraries gathered from the 3D crystal structures can be used for the development of conformationally biased competitive inhibitors. Cyclophellitol cyclopropane adopts a $^{4}H_{3}$ conformation, thereby resembling the transition state oxycarbenium ion conformation in retaining exoglycosidases (**Figure 1.9**) and by this virtue has turned out to be a surprisingly potent GBA inhibitor (115).

1.8 Activity Based Probes

One of the main limitations of chaperone therapy is that it is only effective on certain mutations of the GBA1 enzyme. It is also very dosing sensitive, as incorrect dosing of the chaperone can lead of reduction of enzyme function (98). Therefore in clinical trials it is necessary to be able to measure accurately and efficiently the levels of active enzyme in a patient's cells.

In any disease involving a loss of function of an enzyme, methods of quantifying the presence of that enzyme are vital in diagnosis and the monitoring of treatment. Whilst techniques such as proteomics and transcriptomics can give a measure of the level of a given protein being expressed by a cell, interest lies more in the enzyme's activity rather than sheer quantity of the enzyme in the cell. A protein may not have undergone correct the posttranscriptional modifications to function, such as glycosylation or the cleavage of a pro-peptide form. In addition, the enzyme may have failed to locate to the correct target cell organelle.

Activity based probes (ABPs) offer one solution to the activity of active enzyme, and isoforms thereof. They thus allow for the targeting and quantification of only the catalytically active forms for a given enzyme. They can be used in conjunction with measurements such as proteomics to give a percentage of functional enzyme within a cell as the total produced (117). The use of a mechanism based inhibitor means that only active, functioning enzyme is labelled (118) . Alone, the activity based probe will not bind to a misfolded, non-functional enzyme and thus can be used to quantify the levels of active GBA1 in a cell at any time.

APBs are built from three constituent parts – a chemically active warhead component that is designed to target the enzyme of interest specifically and irreversibly, covalently bind. The probe moiety acts as a reporter and may be an affinity tag, for isolation of a specific enzyme from a cell lysate or a fluorescent species to allow qualification and visualisation of active enzyme with a cell or system. Probes can be directly attached to the warhead (often via a linker)– for single step labelling or may incorporate chemical tags for two step reactions (such as azides for click chemistry with alkynes) (**Figure 1.8**) (42).

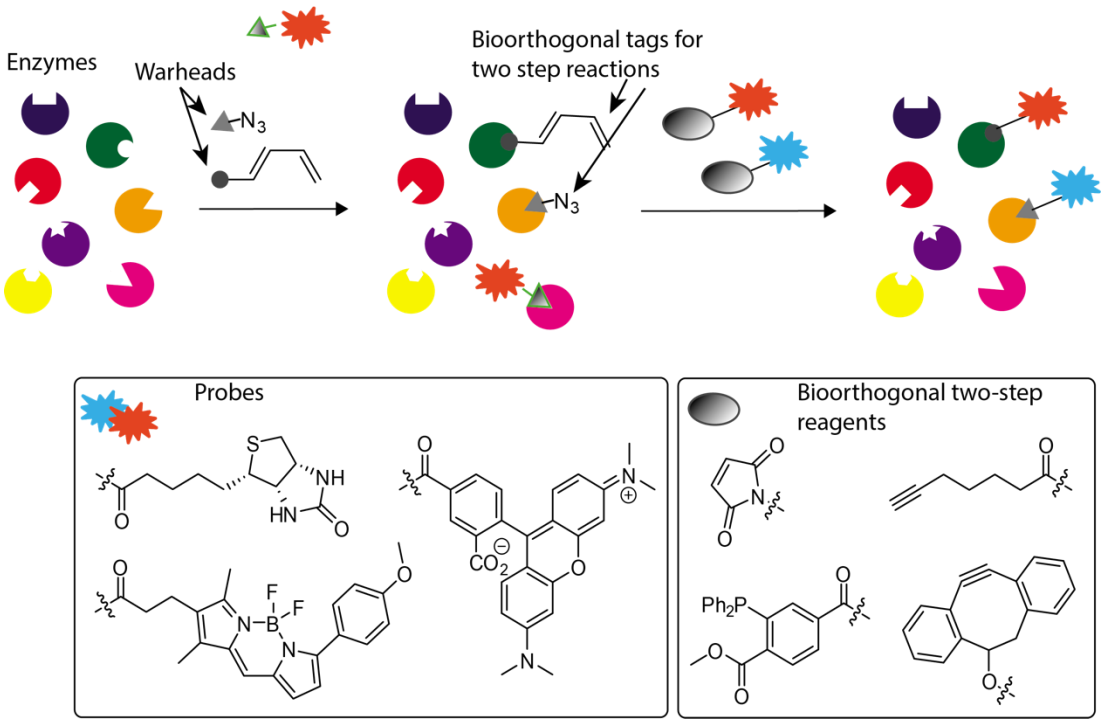


Figure 1.8 Schematic diagram for an activity based probe. A) ABPs are made up of three moieties, a targeting "warhead", a linker region and a probe. B) ABPs can selectively target only the active enzyme in a complex system and can be used in direct one-step labelling or via two step processes with biorthogonal chemical tags. The probe could be a fluorescent group, or a biotin for enzyme capture and enrichment. This thesis focusses on the use of fluorescent probes. Figure adapted from Willems et al (2011) (119)

About two decades ago, the first examples of tagged covalent ABPs for activity-based protein profiling studies on serine hydrolase and cysteine proteases in complex biological samples were published (120,121).

The mechanism-based, covalent and irreversible GBA1 inhibitors described earlier are ideal starting points for the development of ABPs (122,123). Substitution of the epoxide in cyclophellitol for an aziridine yielded cyclophellitol aziridine - a compound at least equally potent in its inhibition of GBA1 (124,125).

In contrast to cyclophellitol, cyclophellitol aziridine can be modified to contain a fluorophore (for gel or cell imaging) or a biotin (for chemical proteomics studies) through aziridine *N*-alkylation or *N*-acylation (**Figure 1.6.B**; (126)). Whilst grafting a reporter moiety onto one of the cyclitol hydroxyls is typically detrimental for reactivity with one notable exception: the substituted cyclophellitol ABP turned out to be a highly specific and very potent GBA1 probe (127).

1.9 Human GBA1 and GBA2

In humans, GBA1 is not the only retaining β-glucosidase. When the β-glucosidase activity in spleen from control subjects and patients with Gaucher's disease was characterized, two β-glucocerebrosidases where able to be separated based on their ability, or inability, to bind anti-(placental β-glucocerebrosidase) antibodies (128).

GBA2 belongs to glycoside hydrolase family 116, a family much less well studied than that of *GBA1* (GH30). Until recently, comparatively little was known about this group of enzymes, initial work on the molecular characterisation of a glycoside hydrolase from the archaeon *Sulfolobus solfataricus* led to the classification of CAZy family GH116 in 2010 (129). The structure of the *GBA2* homologue TxGH116 is described in this thesis.

Whilst GBA1 and GBA2 both act on the same substrate (GluCer), they can be differentiated by their reactions towards particular inhibitor compounds. GBA1 is inhibited by CBE whereas GBA2 is comparatively insensitive to it (130). Conversely, GBA2 is more sensitive to inhibition by deoxynojirimycin analogues compared to its lysosomal counterpart (IC_{50} = 250 and 21 μM, GBA1 and GBA2 respectively) (3).

GBA2 itself is linked to many medical conditions outside of Gaucher's disease. In humans, different mutations in GBA2 can lead to series of hereditary diseases such as autosomal-recessive cerebellar ataxia and hereditary spastic paraplegia (131-133). GBA2 therefore seems to be necessary for healthy neuronal development. Unlike Gaucher's disease, in which a variety of mutations in GBA1 all lead to the same disease outcome, various point mutations in GBA2 are linked to different clinical consequences.

Significantly, Gaucher's disease mouse models show a reduction in symptoms when *GBA2* was knocked out (134). This activity of GBA2 in Gaucher's disease cells is still not fully understood, although it appears it is down-regulated in the absence of GBA1 activity, rather than compensating for the loss of the lysosomal β-glucosidase enzyme (135,136). Deletion of GBA2 in fact partially relives Gaucher's disease pathology in mouse models, suggesting that one factor of Gaucher's pathology is the accumulation of the toxic sphingosine, produced at the cytoplasmic face of the ER and Golgi apparatus by GBA2 activity (134)

The conversion of glucosylceramide to ceramide in excess by overexpressed GBA2 has linked GBA2 to ceramide-driven apoptosis of melanoma cells (137). This suggests that modulation of the behaviour of GBA2 by specific inhibitors could be useful for the treatment of Gaucher's disease whilst the use of activators could possibly be used to combat melanoma, making GBA2 a viable therapeutic target for multiple diseases.

There is a third retaining β-glucosidase within human cells that is capable of the breakdown of GluCer. GBA3 is also known as the "cytosolic β-glucosidase". It is a predominantly liver enzyme and hydrolyses β-D-glucoside and β-D-galactoside and is thought to hydrolyse xenobiotic glycosides. However GBA3 has been shown to have significant neutral glycosylceramidase activity (EC 3.2.1.62); this suggests that the enzyme may be involved in a non-lysosomal catabolic pathway of glucosylceramide (138). So far, no link has been made with GBA3 activity and Gaucher's disease phenotype. Carriers of a common GBA3 mutation show no modification to their Type 1 Gaucher disease manifestation (139).

GBA3 belongs to CAZy family GH1, and the structure of the human protein was published in 2007 (PDB ID: 2JFE) (140). GBA3 has been involved in research into the use of ABPs to identify active site residues in of retaining β-glucosidases and the effect of certain probes (including CBE and conduritol aziridines) on GBA enzymes in general (141), but as protein is not expressed in most tissues and appears to play no pathogenic role in Gaucher's disease there is less focus on it than GBA1 and GBA2 in published research.

1.10 CAZy and the glycosidase hydrolases

All carbohydrate-acting enzymes, including glycoside hydrolases, glycosyltransferases, and carbohydrate binding modules, can be classified into families on the basis of their amino acid sequences(142). These classifications are make-up the online CAZy database. The Carbohydrate-Active Enzymes database (http://www.cazy.org, (143)) was launched in 1999 and currently lists 145 families of glycoside hydrolases alone. CAZymes sharing a similar 3D structure can be further grouped together in clans. CAZy families are founded by an initial enzyme with distinct biochemical characterization, which is joined by those with amino acid sequence similarity. However enzymes in the same family may have different substrate specificities (114).

In glycosidase hydrolase enzymes, hydrolysis of the glycosidic bond may occur with either retention or the inversion of the anomeric configuration – giving the division of retaining glycosidases and inverting glycosidases. These two general mechanisms were described by Daniel Koshland in 1953 (144). GBA1 and GBA2 are both retaining β-glucosidases. Hydrolysis of the glycosidic bond is catalysed by two amino acid residues - a proton donor (acid) and a nucleophile/base.

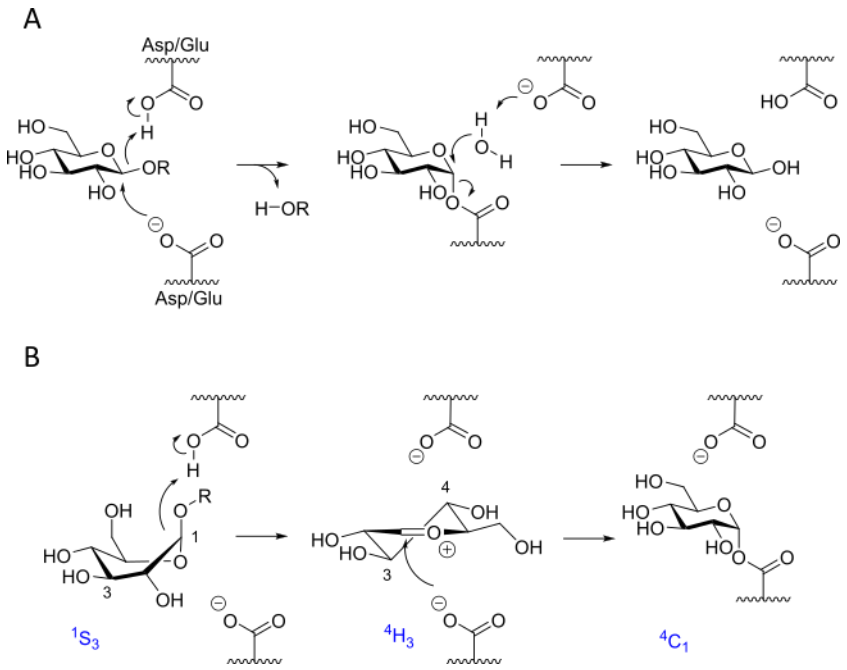


Figure 1.9 Retaining glycosides mechanism. A) Retaining β-glucosidases are able to hydrolyse β-glucopyranose from β-glucosides in a two-step double displacement mechanism, with a net retention of configuration of the anomeric carbon. The acid/base is located at the top, the nucleophile below. B) Conformational itineraries of retaining β-glucosidase. Only the first half of the two-step mechanism is shown. Adapted from Breen et al, 2017 (**Appendix 7**)

Retaining β-glucosidases are found in many CAZy families, e.g. GH1 (including GBA3), GH2, GH3, GH5, GH30 (including GBA1). These GH families can be further sorted into "clans" based upon their three-dimensional structure. For example, GH1, GH2, GH5 and GH30 family enzymes possess $(\beta/\alpha)_8$-barrel catalytic domains and are sorted into clan GH-A, whereas GH3 enzymes are more unusual with an active site between a $(\beta/\alpha)_8$-barrel and a β-sandwich domain.

An enzyme that has been biochemically characterised will also be given at least one Enzyme Commission (EC) number. These numbers are based solely on the reaction that a given enzyme can catalyse and therefore many enzymes across difference GH families, organisms and structural clans may have the same EC number. For example, as GBA1 and GBA2 are both acid β-glucosidases, they are both classified as EC.3.2.1.45.

1.11 Summary

At the start of this thesis, activity-based probes for glycosidases were just emerging as a valuable tool, possibly one with direct clinical application. Major challenges were observing how they bind, how probe and tag groups were accommodated and whether one could engineer better specificity. This was particularly important for distinguishing between GBA1 and GBA2 in living cells. The human GBA1 structure had been reported, with a variety of inhibitors and potential chaperones (Chapter 2) and the Moracci group has performed the first mechanistic and specificity analyses of the GBA2 family GH116. The goals of my PhD were:

- Determine the 3D structure of human GBA1 in complexes with diverse ABPs, in order to analyse specificity and aid the design of better, more specific, ABPs. This work is described in Chapter 2.
- Determine the 3D structure of a bacterial homolog of GBA2 from CAZY family GH116 and, if possible analyse ligand complexes of the enzyme, again to aid probe design and application. This work is described in Chapters 3 and 4.

Possible new directions, new probes and methods, resulting from this work are described in a final "Conclusions and Perspectives" Chapter 5. Four publications from this work are attached as appendices and cited in the relevant chapters.

Chapter 2. 3D structures of human glucocerebrosidase GBA1 in complex with activity-based probes

2.1 Abstract

Gaucher's disease is a lysosomal storage disorder due to mutation in the gene encoding for the lysosomal enzyme acid β-glucocerebrosidase, GBA1. Current treatments involve enzyme replacement, or pharmacological chaperone therapy using small molecule folding aids. Treatment of Gaucher's disease ideally demands early diagnosis and subsequent monitoring of treatment efficacy. Two breakthroughs in these regards were the first 3D structure of GBA1 in 2003 and the development of activity-based probes (ABPs) of enzyme action in 2010. Here, I sought to probe the 3D structure of human GBA1 with diverse covalently-bound ABPs to determine the mode of binding, compare with other known ligands, and ascertain if 3D insight might lead to better, perhaps more specific, probes. Recombinant GBA1 (a gift of Genzyme) was crystallised and the structures solved by X-ray crystallography at resolutions up to 1.49Å resolution. 3D structures of complexes were obtained for reacted *gluco*-epoxide "cyclophellitol-derived" probes and their related aziridines. A complex with a *galacto*-configured aziridine revealed how GBA1 can accommodate galactosyl ligands. A serendipitously-obtained dual complex, with covalent aziridine and non-covalent fluorescent epoxide highlights how the enzyme may accommodate these bulky fluorescent groups. It is clear from inspection of the 3D structures that active centre channels extend out beyond the O6-position of the glucosyl moiety, explaining in part why GBA1 can tolerate substituents at this point in substrate analogues and inhibitors in a way related enzymes cannot. This helps account for the specificity and may lead, ultimately, to the development of next generation inhibitors and activity based probes.

2.2 Introduction

As described more fully in **Chapter 1,** Gaucher's disease is a lysosomal storage disorder caused by diverse mutations in the gene encoding for the lysosomal enzyme acid β-glucocerebrosidase, GBA1(35). Gaucher's disease is characterised by the build-up of glucosylceramide (GluCer) within the lysosome resulting in enlarged organs, splenomegaly, hepatomegaly and, in some cases, neurological disorders (32,39). GluCer is an abundant glucosylceramide, believed to play roles in cell signalling and immune modulation (145). GBA1 (EC 3.2.1.45) is a 497 amino-acid membrane-associated glycoprotein belonging to CAZy (www.cazy.org) family GH30 (114). GBA1 degrades GluCer into glucose and ceramide in the lysosome, which can then be recycled in the cytoplasm. When mutations in both alleles encoding GBA1 lead to a reduction or loss of function in the enzyme, GluCer can build up in the lysosomes and lead to deleterious effects. GBA1 catalyses the hydrolysis of GlcCer with net retention of anomeric configuration,(53) **Figure 2.1**.

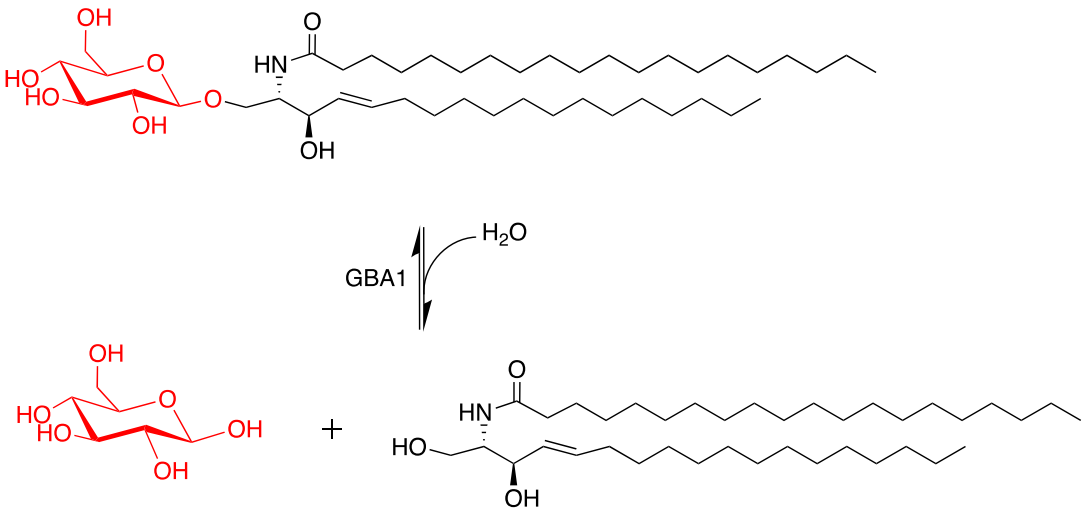


Figure 2.1. The reaction catalysed by GBA1; the hydrolysis of GlcCer with net retention of anomeric configuration.

Prior to structure solution, the reaction mechanism, was first elucidated by the Withers group who demonstrated, in 1994, accumulation of a trapped covalent glycosyl-enzyme intermediate using 2-fluoro-2-deoxy glucosides (53). They were then able to propose a classic Koshland double–displacement mechanism for GBA1 action, with Glu340 (identified by LC MS/MS) acting as the catalytic nucleophile, **Figure 2.2**.

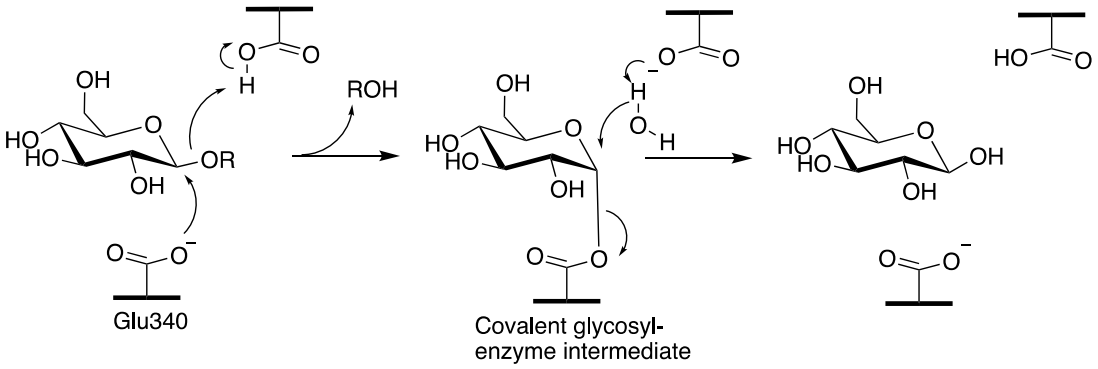


Figure 2.2. Double-displacement reaction mechanism for GBA1 (after (53)). The active centre nucleophile, Glu340, is central to the reactivity with epoxides and aziridines in this chapter.

Building on from this mechanistic breakthrough, in 2010 Herman van Overkleeft and colleagues developed activity-based (protein) probes (ABPs) to study GBA1. Although early covalent inhibitors, such as conduritol epoxides had mislabelled the active centre (requiring the correction by Maio and colleagues described above (53)), Overkleeft took advantage of the better selectivity of cyclophellitol to overcome this problem.

Cyclophellitol is a natural product, isolated in 1990 from the mushrooms of the Phellinus sp. (101,146). It is a is a powerful mechanism-based inhibitor of retaining β-glucosidases and was used in 1992 to induce Gaucher-like symptoms in mice; presumably as a result of GBA1 inhibition (147). In 2007, Gloster, Madsen and Davies revealed how cyclophellitol acted as an irreversible glycosidase inhibitor, **Figure 2.3.A**, (116). Overkleeft subsequently had the foresight to realise that by incorporating additional functional groups (biotin capture probes, azides for click chemistry, fluorescent groups etc) at the 6-position, **Figure 2.3.B**, he could develop β–glucosidase selective ABPs (127). He subsequently extended the cyclophellitol ABP concept to aziridine probes (in which nitrogen sits in place of oxygen in the three-membered ring) allowing substituents to be incorporate on the aziridine nitrogen, **Figure 2.3.C** (for example, (17,124)).

In a landmark 2010 paper, the Overkleeft group demonstrated how fluorescent ABPs could be used to image the activity of GBA1 in cell extracts, including Gaucher disease patients, **Figure 2.3.D**. He further showed that the reagents were not diagnostic tools, but could also measure the efficacy of pharmacological chaperone approaches in cells, **Figure 2.3.E**.

From this point, it became important to study the 3D structure of GBA1 with these ABPs, in order to dissect their mode of action and ultimately – if possible – to aid the design of a

new generation of GBA1 probes. Such work will be described later in this chapter after a brief summary of the 3D structure of human GBA1 and its published complexes.

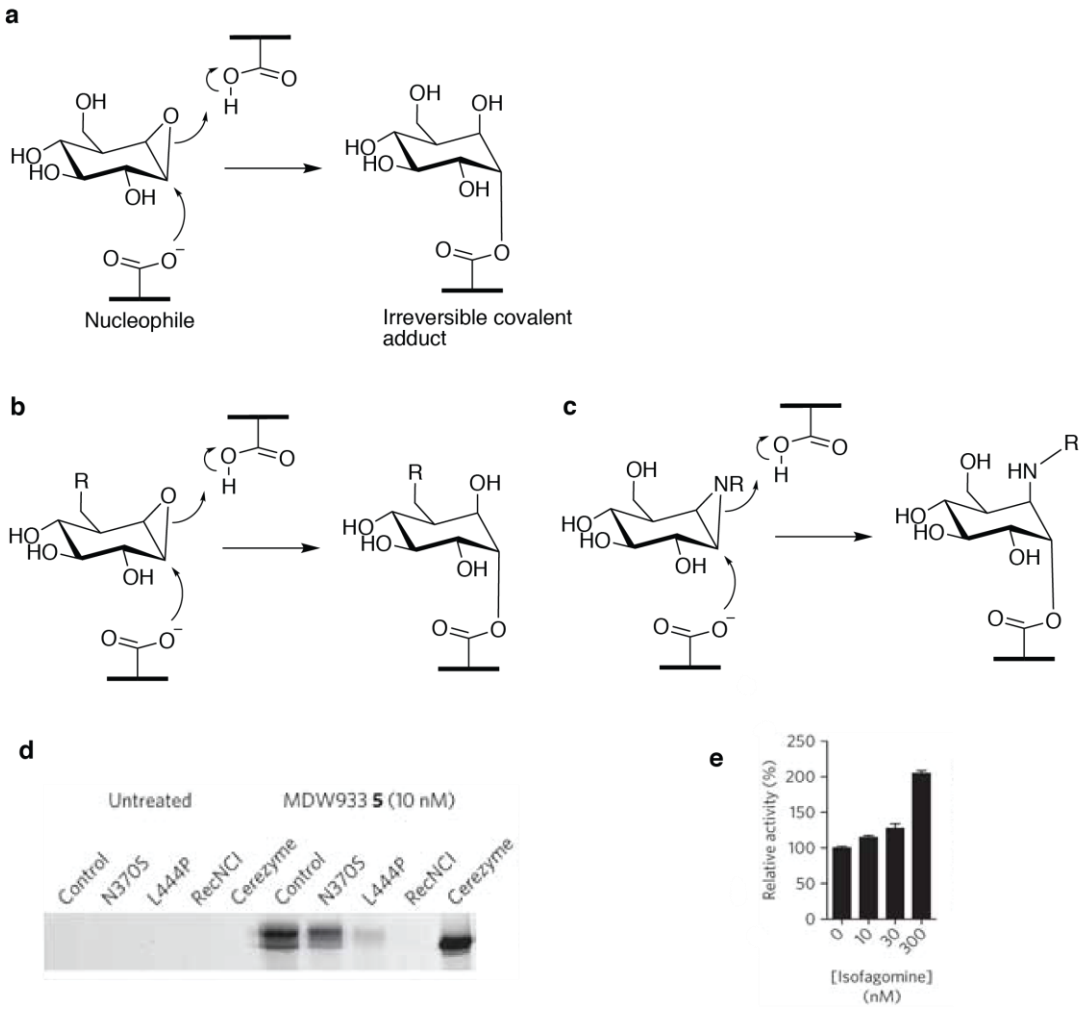


Figure 2.3. Development of β-glucosidase-selective ABPs. A) Gloster and colleagues demonstrate the specific irreversible action of cyclophellitol; B) Overkleeft extends this to make functionalised cyclophellitols and (C) development of aziridine probes in which the functionalization can be incorporated at the ring nitrogen. D) Application of an epoxide ABP, designated MDW933 (described in Figure 2.6), to measure GBA1 activity in human fibroblast extracts: a healthy control group, three Gaucher disease genotypes and a Cerezyme treated patient. E) Experimental demonstration of the molecular chaperone effect of isofagomine, 300nM of which increases GBA1 activity 2-fold in human fibroblasts. (Panels D and E taken from Witte et al., 2010 (127)).

2.2.1 The 3D structure of human GBA1

The unliganded 3D structure of GBA1 was initially determined by the Sussmann group in 2003 using a mercury heavy atom derivative (38). Subsequently the group published key structures with the imino sugars N-butyl and N-nonyl deoxynojirimycin (104). The structure was also re-determined by the Petsko group in 2007, in this case in complex

with the aza sugar isofagomine (105). To date, 23 crystal structures have been deposited on the PDB (see http://www.cazy.org/GH30_structure.html) as listed in Table 2.1.

The 3D structure consists of three non-contiguous domains, with the catalytic site located in domain II (residues 76–381); a $(\beta/\alpha)_8$ "TIM" barrel, **Figure 2.4.A**. GBA1 is activated by lipids, and it is proposed that the "top" surface of the enzyme is the lipid binding face (113). Although the function of the two non-catalytic domains is unknown, mutations that cause Gaucher's disease are found in all three domains (see **Chapter 1**, Figure **1.3**). The $(\beta/\alpha)_8$ barrel is a classic CAZY clan GH-A (148) catalytic centre with the putative acid-base, Glu235, on strand β–4 and the enzymatic nucleophile Glu340 on strand β–7, **Figure 2.4.B**. The active centre is made up with classical hydrogen-binding side chains: Asp127, Trp179, Trp381, and Asn396 whilst Tyr313 forms the base of a hydrophobic platform in which the lipid leaving group is accommodated (illustrated here with the N-butyl DNJ complex; PDB 2V3N).

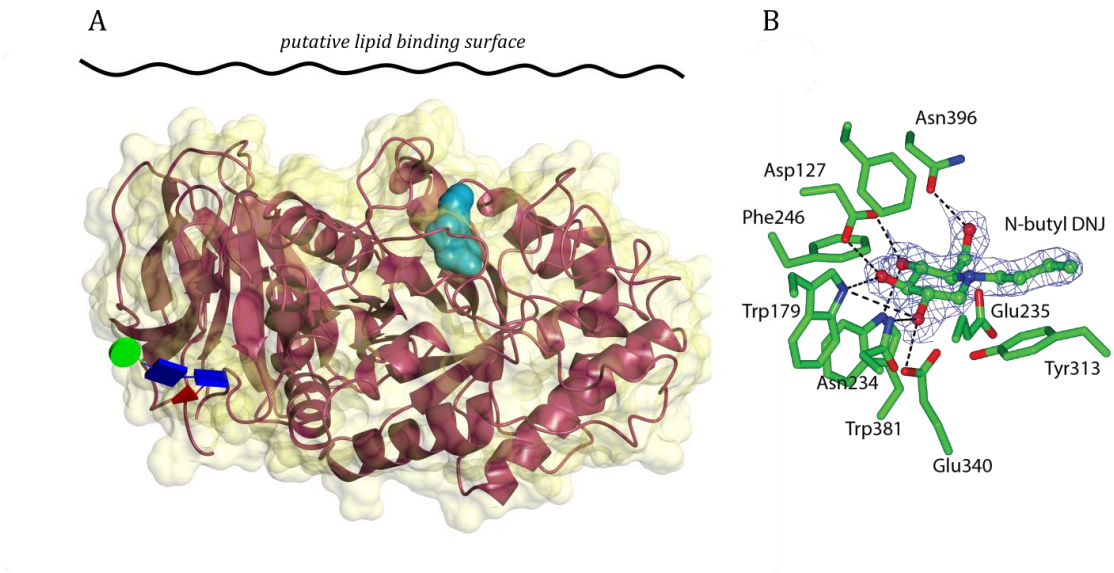


Figure 2.4. 3D structure of human GBA1. A) The 3D structure, show in red, with the active centre indicated as a cyan surface around the N-butyl DNJ ligand, one N-glycan is illustrated using GLYCOBLOCKS (149) and the location of the putative lipid binding face is illustrated. B) The active centre as determined through the complex with N-butyl DNJ ("Zavesca"), PDB ID: 2V3D (104).

GBA1 has been successfully crystallised in two space groups: $C222_1$ (with two copies of GBA1 in the asymmetric unit) and a $P2_1$ form (with four copies). A table of known 3D structures is given overleaf. **Table 2.1** They structures reflect a variety of different expression systems and degrees of glycosylation. The source of the GBA1 enzyme is mainly that used in patient treatment: Cerezyme, Taliglucerase-alfa or Velaglucerase-alfa. The former two have a single point mutation (R495H) that is not present in Velaglucerase-alfa. The first structure of a "Gaucher" mutant form of GBA1, the most common N370S mutation, was solved in 2010 using a protein expressed in baculovirus (150).

TABLE 2.1: PUBLISHED 3D STRUCTURES OF HUMAN GBA1

PDB Code	Source	Commercial drug name	Deglycoslyated?	Ligand?	Resolution	Reference
1OGS	CHO	Cerezyme	Partially	No	2.0 Å	(38)
1Y7V	CHO	Cerezyme	Partially	Conduritrol-B-epoxide	2.4 Å	(151)
2F61	CHO	Cerezyme	Partially	No	2.5 Å	(41)
2J25	CHO	Cerezyme	Partially	No	2.9 Å	(152)
2NSX	CHO	Cerezyme	Partially	Isofagomine	2.11 Å	(105)
2NT0	CHO	Cerezyme	Partially	Glycerol	1.79 Å	(105)
2NT1	CHO	Cerezyme	Partially	No	2.3 Å2	(105)
2V3D	Plant	Taliglucerase-alfa	No	N-Butyl-Deoxynojirimycin	1.96 Å	(104)
2V3E	Plant	Taliglucerase-alfa	No	N-Nonyl-Deoxynojirimycin	2.0 Å	(104)
2V3F	Plant	Taliglucerase-alfa	No	No	1.95 Å	(153)
2VT0	Plant	Taliglucerase-alfa	No	Conduritol-beta-epoxide	2.15 Å	(113)
2WCG	Plant	Taliglucerase-alfa	No	N-Octyl(Cyclic Guanidine)-Nojirimycin	2.3 Å	(154)
2WKL	Human cell line	Velglucerase-alfa	Partially		2.7 Å	(155)
2XWD	CHO	Cerezyme	Partially	5-*N*,6-*O*-[*N*'-(*n*-octyl)iminomethylidene]nojirimycin	2.66 Å	(156)
2XWE	CHO	Cerezyme	Partially	5-*N*,6-*S*-[*N*'-(*n*-octyl)iminomethylidene]-6-thionojirimycin	2.31 Å	(156)
3GXD	CHO	Cerezyme	Partially	No	2.5 Å	(95)

3GXF	CHO	Cerezyme	Partially	Isofagomine	2.4 Å	(95)
3GXI	CHO	Cerezyme	Partially	No	1.84 Å	(95)
3GXM	CHO	Cerezyme	Partially	No	2.2 A	(95)
3KE0	Baculovirus*	No	Partially	No	2.7 Å	(150)
3KEH	Baculovirus*	No	Partially	No	2.8 Å	(150)
3RIL	CHO	Cerezyme	Partially	(3S,4R,5R,6S)-azepane-3,4,5,6-tetrol	2.4 Å	(157)
3RIK	CHO	Cerezyme	Partially	(3S,4R,5R,6S)-1-(2-hydroxyethyl)azepane-3,4,5,6- tetrol	2.48 Å	(157)

Table accurate as of 17th July 2017

*N370S mutant

Many of the 3D structures of GBA1 have been determined in order to understand the binding of inhibitors as potential pharmacological chaperones; key compounds (with their PDB codes) are shown in **Figure 2.5**.

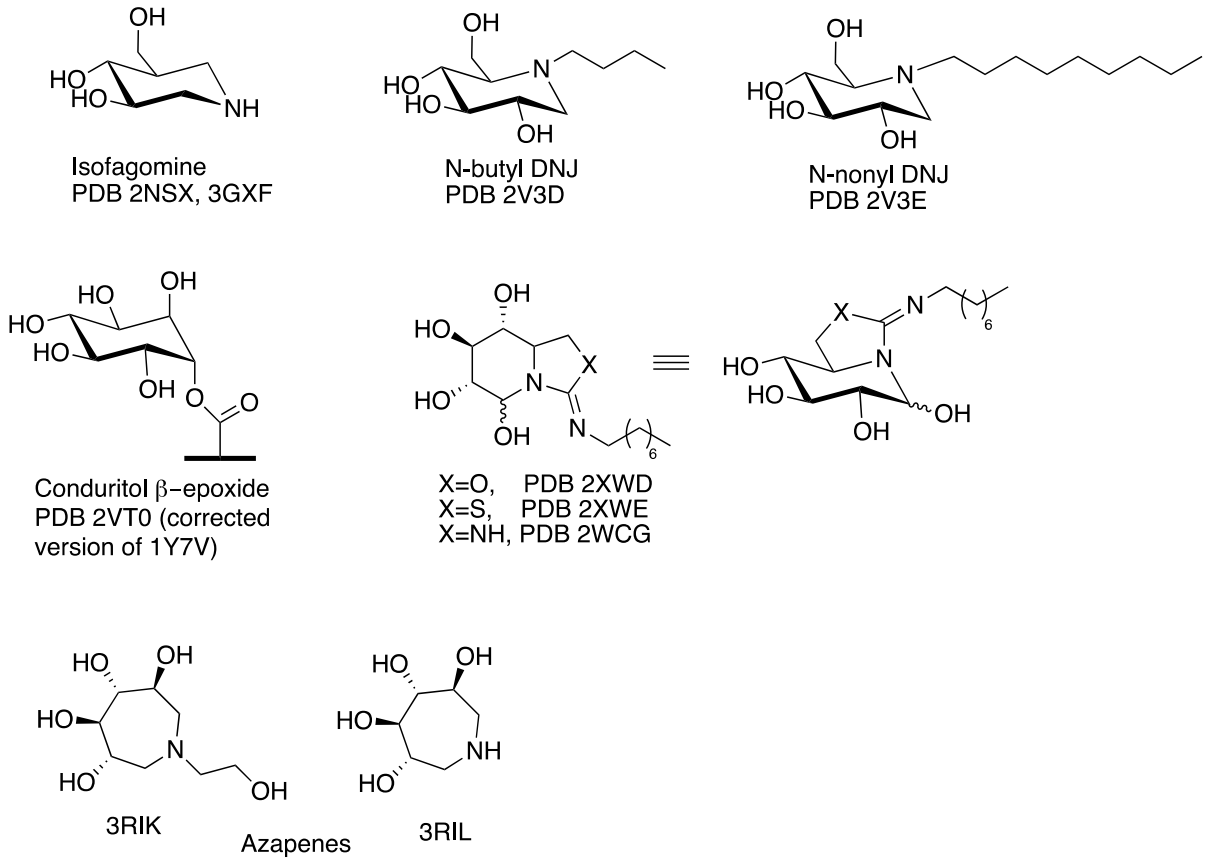


Figure 2.5. GBA1 inhibitors observed in published 3D structures. PDB 2VT0 is the only known covalent adduct complex (which corrected an incorrect interpretation, with the wrong stereochemistry and conformation; published as PDB 1Y7V).

The most well-known GBA1 inhibitors and potential chaperones are the imino/aza sugars based around the deoxynojirimycin core (104), and isofagomine (105). N-butyl DNJ ("Zavesca") finds clinical use for the treatment of Gaucher's disease where it acts through "substrate reduction" via inhibition of the GlcCer synthase. All of these compounds also stabilise GBA1 and can therefore be considered as pharmacological chaperones allowing unstable variants of GBA1 to pass through the secretory pathway. All these compounds, including the bicyclic "cyclic guanidine-nojirimycins" (154,156) and the 7-membered azapenes (157) bind in the active centre "-1" subsite (nomenclature in Ref. (158)) with the pseudo-sugar rings mimicking the binding of glucose.

Prior to this work, the only 3D structures of GBA1 with a covalently-bound inhibitor were the structures with conduritol-β-epoxide. This was incorrectly interpreted in an early

medium resolution paper (151), but subsequently corrected upon collection of higher quality data (113). This paucity of insight into the binding of covalent inhibitors of GBA1, especially give their increasing potential as diagnostics, was one of the motivating factors for the work in this thesis.

2.2.2 Aims of the research in this chapter

The goal of the work described in this chapter, was to solve the 3D structures of human GBA1 (material supplied by Sanofi Genzyme, Cambridge, Ma., USA) in complex with different ABPs selected to reflect different reactive groups, configurations etc. It was hoped that, ultimately, these studies would allow comparison with similar work on GBA2 and allow the development of selective inhibitors for the two enzymes. Five complexes were studied, KY170 and KY353 as representatives of cyclophellitol epoxide and aziridine, KY358 as an example of an acyl-aziridine, TB562 as an example of a galacto-configured ABP that is believed to react with GBA1. Whilst previous findings had shown that galacto-configured epoxides do not interact with GBA1 (159), here it is shown that the galacto-configured aziridine TB562 is able to bind to the enzyme both whilst in solution and in the crystal form. MDW933 is a large fluorescent epoxide (used in Figure 2.3) in an attempt to see how the GBA1 structure might accommodate the bulky fluorescent group, **Figure 2.6**.

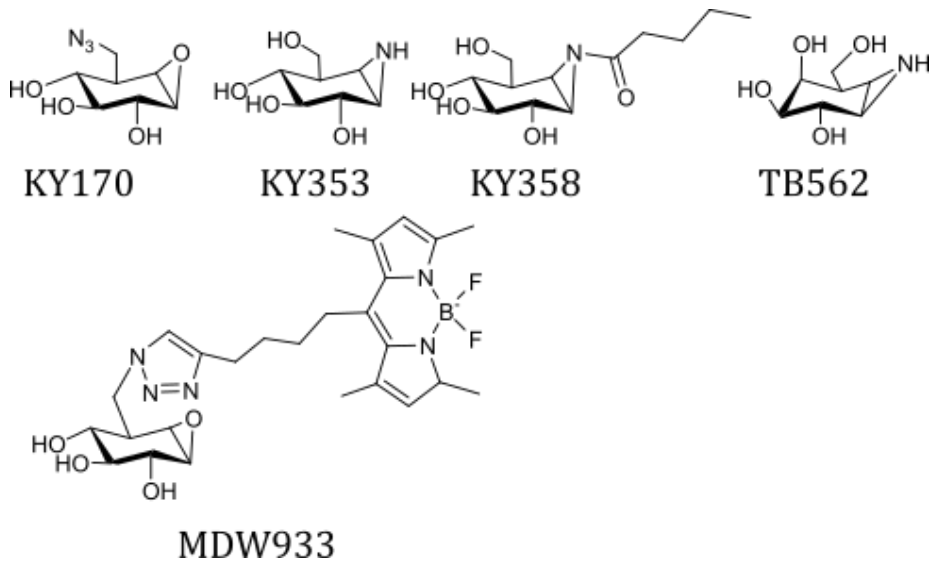


Figure 2.6 Activity-based GBA1 probes studied at a structural level in this thesis

2.3 Materials and Methods

2.3.1 Purification of commercial GBA1 enzyme

Cerezyme, from expired patient samples, was supplied by Sanofi-Genzyme. The supplied lyophilized Cerezyme tablet (containing 5mg active GBA1) was resuspended in phosphate buffered saline (PBS) and dialysed overnight against PBS (pH 7.0) to remove excess salts and preservatives and then buffer exchanged into 50mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer, 100mM NaCl pH 6.6. The contained GBA1 was partially de-gylcosylated using 150 units of N-glycosidase F for 88 h at 25 C following the protocol as defined in Dvir et al (38)) .

Buffer exchanged GBA1 enzyme was concentrated approximately 10 mg/ml and centrifuged for *ca.* 5 min before setting up crystallisation experiments. Crystals of GBA1 were grown by hanging drop vapour diffusion in Greiner Bio One CELLSTAR 24-well plates (set up manually). Crystallisation conditions was based on previously successful conditions (38) varied across four plates. All crystallisation experiments were performed at 18 ◦C.

2.3.2 Crystallisation conditions

Unless otherwise noted, crystallisation drops were set up as 1:1 ratio of protein to mother liquor and GBA1 was added at 9.8 mg/ml.

- *Apo*-GBA1
  Mother liquor: 1 M $(NH_3)_2SO_4$, 0.17 M guanidine HCl, 0.02 M KCl, 0.1 M sodium acetate pH 4.6
- KY170-GBA1
  Mother liquor: 1 M $(NH_3)_2SO_4$, 0.17 M guanidine HCl, 0.02 M KCl, 0.1 M sodium acetate pH 4.6
- KY358-GBA1
  Mother liquor: 1 M $(NH_3)_2SO_4$, 0.17 M guanidine HCl, 0.03 M KCl, 0.1 M sodium acetate pH 4.6
- TB562-GBA1
  Mother liquor: 1 M $(NH_3)_2SO_4$, 0.17 M guanidine HCl, 0.02 M KCl, 0.1 M sodium acetate pH 4.8
- MDW933-GBA1
  Mother liquor: 0.95 M $(NH_3)_2SO_4$, 0.17 M guanidine HCl, 0.02 M KCl, 0.1 M sodium acetate pH 4.6

- KY353-GBA1
  Mother liquor: 1.1 M $(NH_3)_2SO_4$, 0.18 M guanidine HCl, 0.02 M KCl, 0.1 M sodium acetate pH 4.6

2.3.3 Collection of x-ray data

For initial collection of X-ray data, a crystal of GBA1 was transferred to a cryo-protectant solution containing mother liquor supplemented with 20-25% glycerol or ethylene glycol and then flash frozen in liquid N_2. Crystals of GBA1 bound with various ligands were obtained by either adding a solid powder of the compound directly to a drop containing one or more crystals of GBA1 using an acupuncture needle or by adding 1 µl of the compound dissolved in either water or reservoir solution to the drop and then resealing. After incubation from 30 minutes to overnight, the crystals were removed, cryo-protectant containing 15% ethylene glycol and 20-25% glycerol was added and the crystal was cooled using liquid N_2.

All crystals were then sent to the Diamond Light Source in Oxfordshire for data collection

2.3.4 Data processing and structure determination

Data for each crystal set was processed using XIA2(160,161) and the AIMLESS data reduction pipeline through the CCP4i2 software(162).

The previous GBA1 PDB (2NT0) (163) was used to determine the phases and solve the structure of the apo-GBA1 by molecular replacement using MOLREP(164). Coordinates from the apo-GBA1 were used directly for determination of the ligand complex structures as the space groups were isomorphous. R_{free} sets were assigned from the apo-GBA1 structure to maintain the integrity of the set throughout. Idealised coordinate sets and refinement dictionaries for each ligand were generated using ChemDraw3D and JLIGAND(165) or ACEDRG in the CCP4 suite. Idealised coordinate sets and refinement dictionaries for each ligand were generated using ChemDraw3D and JLIGAND(165) in the CCP4 suite.

Refinement of all structures was performed using REFMAC (166) and model building completed using COOT(167), both programs through the CCP4i2 software. Final models for each of the compounds in this chapter were validated using the wwPDB Validation service (validate-rcsb-1.wwpdb.org/) and conformation of sugars were validated using Privateer through CCP4 suite (168).

2.3.4 Labelling with fluorescent probes

A time course experiment was performed with deglycosylated GBA1 in McIlvaine phosphate/citrate buffer pH 6.0, 10% DMSO using the β-galactose aziridine configured

probe TB652 and the β-glucose aziridine JJB367 (**Figure 2.7**). Both probes were used at a final concentration of 100 nM with a GBA1 concentration of 700 nM.

Time points were taken at 30min, 1h, 2h, 4h, 6h, 24h for GBA1-TB652 to confirm the β-galactose binding. A time point for GBA1-JJB367 was taken at 1h to compare known β-glucose binding to the enzyme.

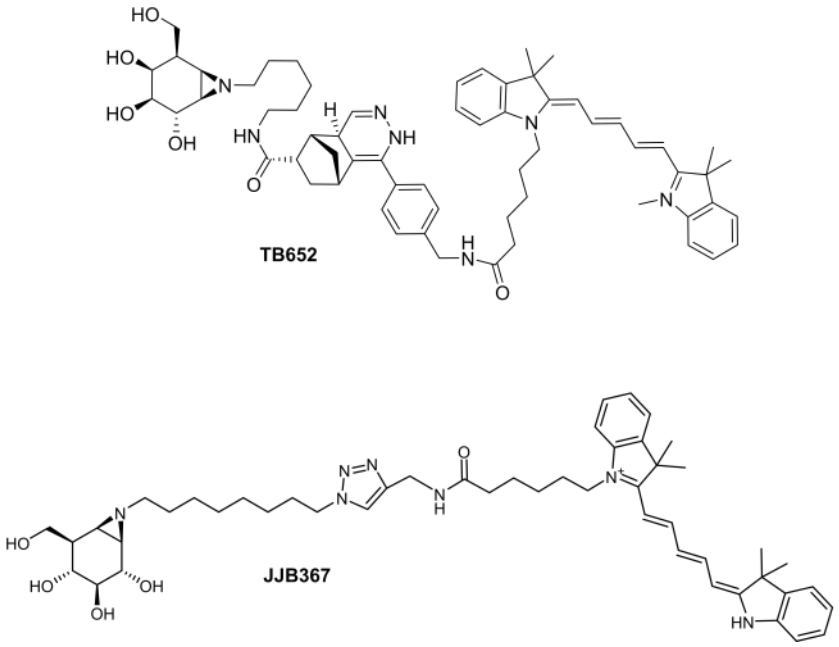


Figure 2.7. Florescent probes used in this time course experiment to probe the anomalous binding of galactosides to GBA1.

2.4 Results and Discussion

GBA1 was successfully crystallised, **Figure 2.8**, following conditions as described by Dvir et al (38). Data reduction and subsequent molecular replacement confirmed the space group of the crystals as $C222_1$ allowing me to refine an unliganded "*apo*" structure of GBA1 at 1.49 Å.

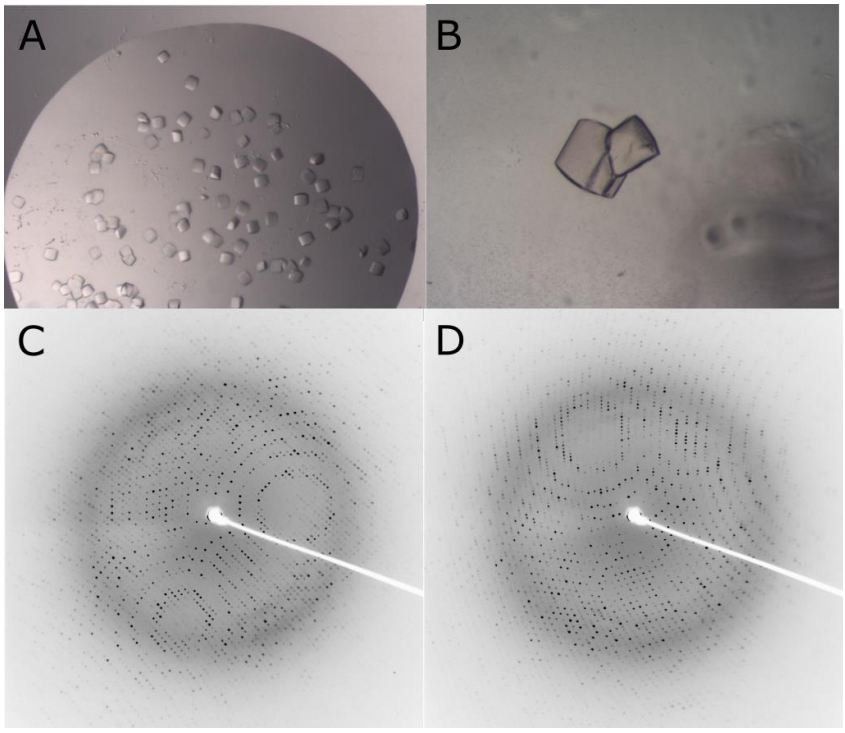


Figure 2.8. GBA1 crystals. A) Crystals of GBA1 B) Crystals of GBA1 after 18 months C and D) Diffraction images taken from unliganded GBA1 crystals in-house. Images were taken at 0-0.5 ° and 90-90.5 ° respectively, with a 5 second exposure.

Data tables for all crystals collected at the Diamond Light Source in Oxfordshire are shown in **Table 2.2** overleaf.

TABLE 2.2: DATA COLLECTION AND REFINEMENT STATISTICS FOR HUMAN GBA1 STRUCTURES DESCRIBED IN THIS CHAPTER

	Cerezyme_apo	**Cerezyme_KY170**	**Cerezyme_KY353**	**Cerezyme_KY358**	**Cerezyme_TB562**	**Cerezyme_ MDW933 / KY358**
Space group	*C* 2 2 2_1	*C* 2 2 2_1	*C* 2 2 2_1	*C* 2 2 2_1	*C* 2 2 2_1	*C* 2 2 2_1
Cell dimensions						
***a*, *b*, *c* (Å)**	109.0, 285.0, 91.0	108.9, 285.8, 91.7	107.4, 286.5, 92.0	110.7, 285.5, 91.9	110.6, 285.3, 91.8	111.2, 285.6, 91.6
α, β, γ(°)	90.00, 90.00, 90.00	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	56.28-1.49(1.52-1.49) *	68.11-1.70(1.73-1.70)	71.62-2.90 (3.06-2.90)	47.59-1.76(1.79-1.76)	72.09-1.80(1.83-1.80)	91.56-1.58(1.62-1.58)
R_{merge}	0.076(2.08)	0.077(1.61)	0.12(0.53)	0.098(0.85)	0.12(1.36)	0.14(2.10)
R_{pim}	0.032(0.88)	0.036(0.74)	0.050(0.22)	0.042(0.37)	0.052(0.62)	0.051(0.79)
CC(1/2)	0.99 (0.53)	0.99(0.52)	0.99(0.96)	0.99(0.84)	0.99(0.59)	0.99(0.50)
***I* /σ (*I*)**	13.4(1.1)	12.1(1.1)	12.2(3.7)	10.0(1.7)	10.9(1.2)	10.7(1.1)
Completeness (%)	100(99.9)	99.1(99.5)	99.9(100)	99.8(99.9)	100(100)	100(100)
Redundancy	6.7(6.5)	5.5(5.7)	6.6(6.9)	6.3(6.2)	5.9(5.8)	8.1(7.9)
Refinement						
Resolution (Å)	56.28-1.49	68.11-1.70	71.62-2.90	47.59-1.76	72.09-1.80	91.56-1.58
No. reflections	229534/11500	156420/7842	31838/1575	143436/7160	134017/6677	198265/9908
R_{work} / R_{free}	0.18/0.20	0.19/0.22	0.20/0.28	0.19/0.21	0.18/0.20	0.17/0.19
***B*-factors (Å²)**						
Protein	26	31	76	30	27	25
Ligand/ion	51	57	104	56	53	44
Water	35	36	N/A	37	37	36
R.m.s. deviations						
Bond lengths (Å)	0.011	0.019	0.010	0.013	0.015	0.014
Bond angles (°)	1.44	1.84	1.44	1.62	1.65	1.59

*Values in parentheses are for highest-resolution shell.

2.4.1 Unliganded GBA1 Structure

X-ray images for the unliganded GBA1 model were collected at the I02 beamline to a resolution of 1.49 Å, whilst data for ligand complexes (described below) were collected at resolutions from 2.4 to 1.58 Å. X-ray data were scaled to resolution limits based on criteria for CC(1/2) of 0.5 in outer resolution shell. The structure can be traced from residues 1-497 in both A and B molecules, although in the latter there are two loop regions with weaker ambiguous density at A 317-318 and at around B344-346. N-glycosylation is observed at Asn19, where two N-acetylglucosamine can be modelled in both molecules . Indeed, there is diffuse density for the β−1,4 linked mannose but this has not been modelled. This glycosylation is consistent with the initial Cerezyme structure by Dvir and colleagues in 2003 (38). At Asn 146 in molecule A, both NAGs of the chitobiosyl core can be modelled; however there is not sufficient density at Asn 146 in molecule B.

The unliganded "apo" structure is essentially identical to past Cerezyme structures, although at a considerably higher resolution than any previously published analysis. Structural similarity searching with PDBeFOLD (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver) shows that, using the A molecule as the representative search, a high level of overlap with past GBA1 structure solutions, with typical main chain r.m.s deviations of just 0.19 Å over all 497 amino-acids (with PDBeFold Q score 1.0 and Z score 29.7). Having secured a good "apo" structure, I then sought to analyse the 3D structures with a variety of covalent activity-based probes.

2.4.2 3D complexes with Cyclophellitol KY170 and cyclophellitol aziridine KY353

The first priority for structure analysis was with the classical cyclophellitol epoxides and aziridines. Data for the 6-azido cyclophellitol KY170 extended to 1.7 Å. A single molecule of covalently reacted KY170 is found on both the A and B molecules with clear electron density, **Figure 2.9A**, for the ring-opened species, covalently attached to the nucleophile Glu340. Although it had been hoped that the 6-azido substituent could be observed, the electron density at this positon is weak, reflecting disorder or decomposition, and the azide could not be modelled with confidence. A similar absence of azide electron density was also recently observed for KY170 bound to an unrelated family GH1 β−glucosidase (see Figure 3B in Ref. (169); again, the reason for the absence of electron density was unknown). The cyclophellitol ring is found in $^4C_1$ chair conformation, and has been opened trans-diaxially as expected, **Figure 2.9B.**

Binding of KY170 to recombinant GBA1 was recorded by Witte et al in 2010, who also first described the synthesis of the ligand (127). KY170 was found to be a far more potent

inhibitor than conduritol B-epoxide (CBE) (**Figure 2.5**), with K_i values of 0.044 ± 0.007 μM for the novel ligand and 53 ± 10.8 for CBE.

The covalent complex with the aziridine KY353 proved more problematic to obtain. For reasons that are unclear, the crystals were often disordered after soaking and credible data were only observed to around 2.9 Å. The electron density shows that the KY353 covalent adduct complex is indeed bound to the nucleophile, **Figure 2.9C**, likely also in $^4C_1$ chair conformation, and has also been opened trans-diaxially, **Figure 2.9D**, but the density is at low resolution and not greatly informative beyond showing covalency.

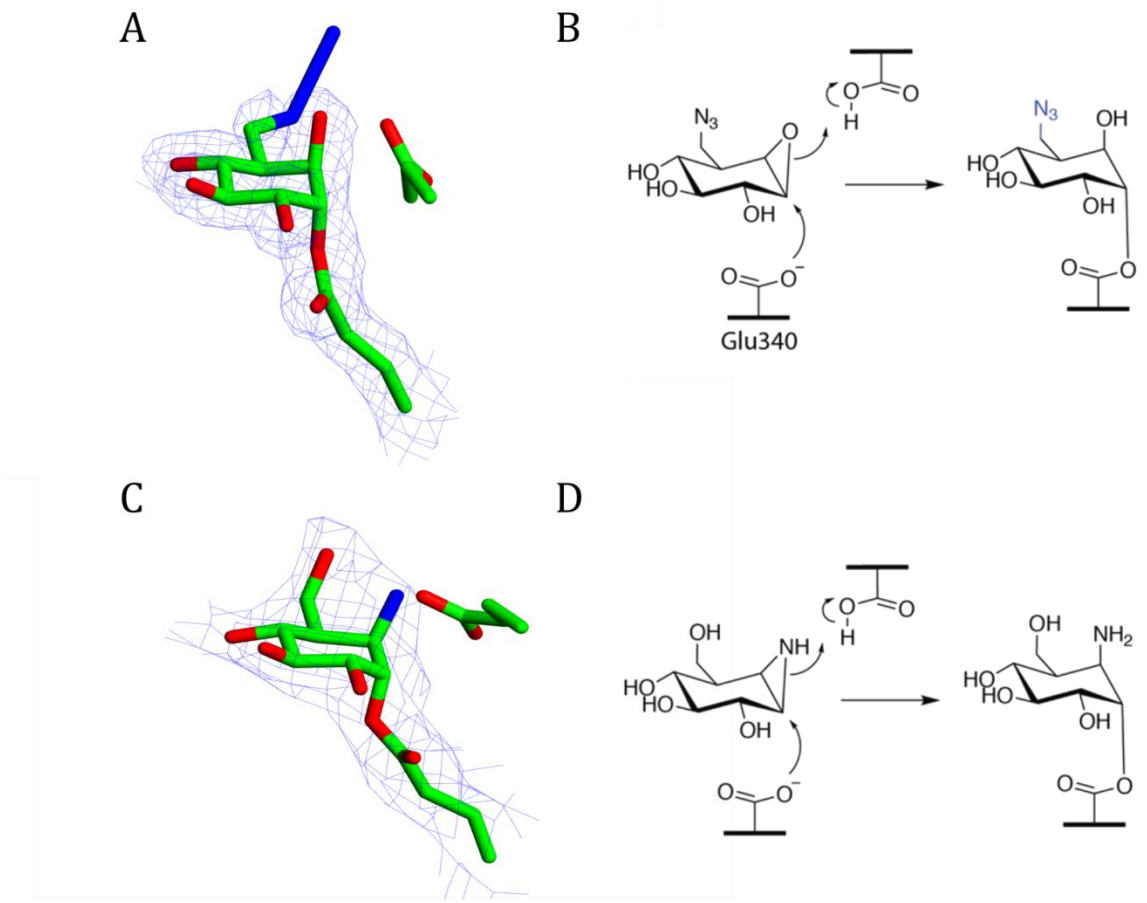


Figure 2.9. Binding of KY170 and KY353 to GBA1. A) Observed electron density for KY170, contoured at 1.2 σ (equivalent to 0.30 electrons / Å^3). B) Trans-diaxial ring opening (the azide coloured blue has scant electron density in the 3D structure). C) Observed electron density for KY353, contoured at 1.2 σ (equivalent to 0.18 electrons / Å^3). D) Trans-diaxial ring opening. Protein structure figures drawn using CCP4MG (170).

The challenges with obtaining a complex structure with the "unmodified" aziridine KY353, coupled to the opportunity to exploit the hydrophobic channel in which the leaving group lipid chain may lie, led us to try and obtain a complex with an acyl aziridine, KY358.

2.4.3 3D complex of GBA1 with acyl cyclophellitol aziridine KY358

Acyl aziridine derivatives, with variety of extensions at the aziridine nitrogen have recently been shown to be potent GBA1 activity-based probes (17). KY358 has previously been shown to be a "better" inhibitor of GBA1 with an apparent IC_{50} of 0.07μM compared to the 0.5 μM for KY353 (124) and acyl aziridine derivatives have indeed found wide use beyond the GBA1 area – such as for profiling of plant glucosidases (171). In order to

obtain a complex with an acyl aziridine, GBA1 crystals were soaked with an "butryamide aziridine", KY358.

KY358 crystals diffracted well and did not appear to suffer the disorder observed with KY353. The structure could be refined at 1.76 Å resolution and revealed clear electron density for the covalently-bound and reacted ligand, **Figure 2.9**. The density for the ring is clear, as is density for the N-acyl group and one, possibly two of the pendant carbons, but beyond that the chain becomes disordered. Subsequently, better insight into the binding of KY358 was obtained in a serendipitous co-complex, so discussion of the interactions of KY358 will be considered later in this chapter in light of these additional data.

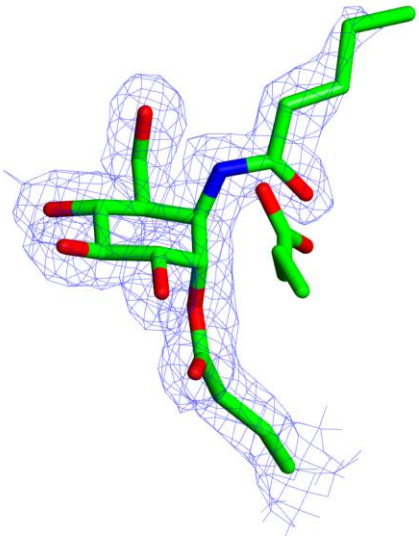


Figure 2.10. Binding of acyl aziridine KY358 to GBA1. Observed electron density for KY170, contoured at 1.2 σ (equivalent to 0.30 electrons / $Å^3$).

2.4.4 3D complex of GBA1 with a *galacto*-configured cyclophellitol, TB562

One of the issues surrounding ABPs is their occasional cross reactivity, indeed promiscuity. GBA1 is believed to be specific for *gluco*- and *xylo*- configured substrates. With TB562 we wanted to assess if there were potential for cross reactivity with a *galacto*-configured covalent inactivator. The synthesis of *galacto*-configured isomers of cyclophellitol was reported in 2014 (172) where it was shown that the *galacto* epoxides did not bind to GBA1. But work on other glycosidases has shown that *galacto*-configured inhibitors can sometime bind to glucosidases; one example (of many) is previous work on the bacterial GBA1 homolog TmGH1 (from *Thermotoga maritima*) by Gloster et al (2007) (173) reported binding of the enzyme to a galacto-configured reagent (for example *galacto*-hydroximolactam (PDB ID: 2J79)).

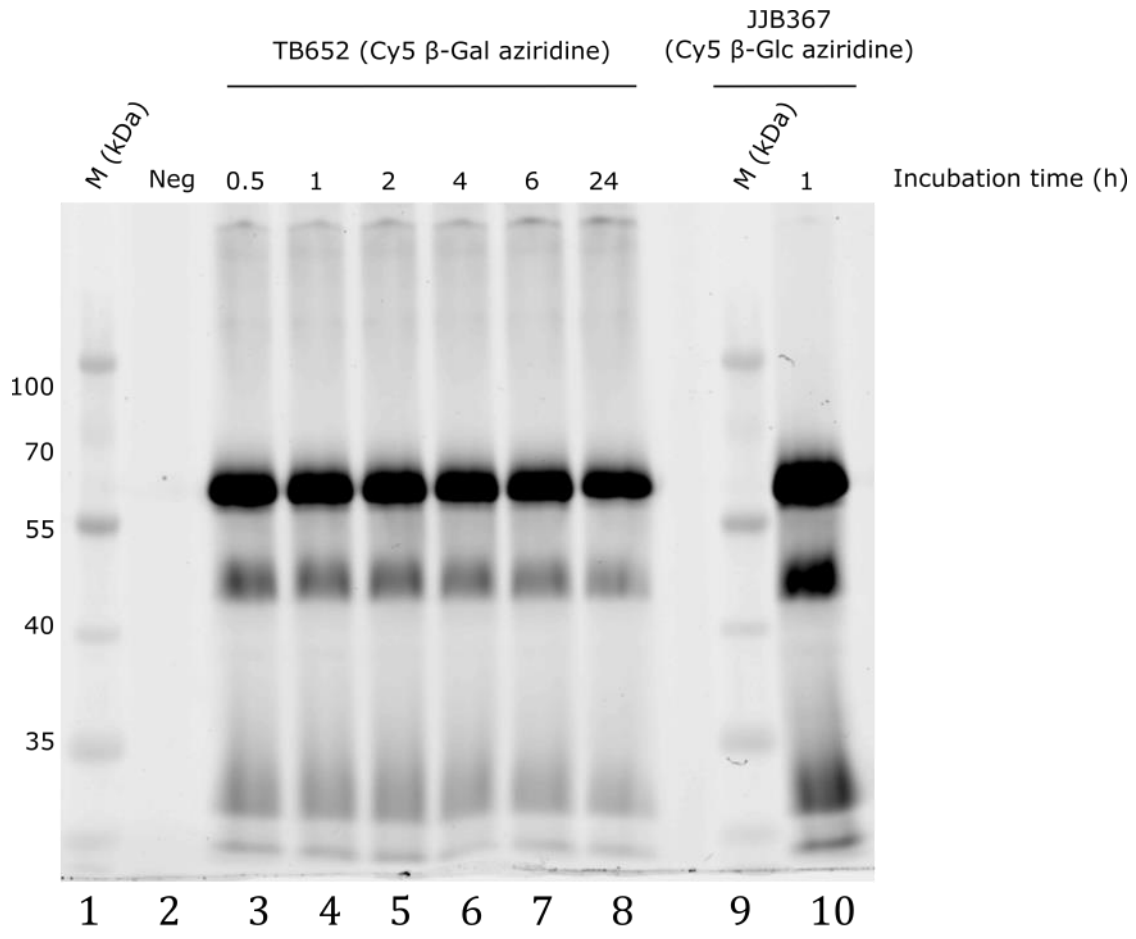


Figure 2.11. SDS-PAGE gel showing the labelling of GBA1 by the β-galactose configured aziridine TB652. Lanes 3-8 are GBA1-TB652, samples taken after 0.5 h, and at subsequent given time points (in hours). Lane 2 is the negative control (GBA1 with no probe). Lane 10 GBA1 labelled with the β-glucose configured aziridine JJB367 after one hour. GBA1 has an expected molecular weight of 56 kDa. Lanes 1 and 9 contain protein weight ladder.

Whilst there was no reported binding of GBA1 to a β-galactose epoxide probe by Willems et al (159), the fluorescent β-galactose aziridine ABP, TB652 was seen to readily bind to GAB1 in solution (**Figure 2.11**) with significant labelling occurring even after the 0.5 hour time point **.** The first observation is that, despite the previous lack of reported activity of GBA1 on galactosides with shorter reaction times, the compound does make a covalent adduct. This immediately highlights the potential cross reactivity of the galactosidase probe, at least the aziridine version which appears more reactive, if it is applied to human cells, especially over longer incubation periods. This strongly suggests that each ABP – whether aziridine or epoxide needs to be seriously analysed for cross reactivity before it is interpreted *in vivo*.

A crystal complex using the probe TB562, a "truncated" form of TB652 without the fluorescent moiety, can give insight in to the enzyme's ability to accommodate this alternative sugar conformation. As previous, a complex with TB562 could be obtained at high resolution (1.8 Å) following a 24 hour soaking experiment.

As seen in **Figure 2.12**, the compound is well ordered and, as far as can be judged by electron density and refinement; it is present at full occupancy. TB562 binds essentially identically to the *gluco*-configured reagents, the only difference being the axial O4 of TB562 which is still able to make a hydrogen bond to Asp127 (distance 2.6 Å), so this residue, as with *gluco* compounds is still able to make H bonds to both O3 and O4 whether the sugar be *gluco* or *galacto* configured.

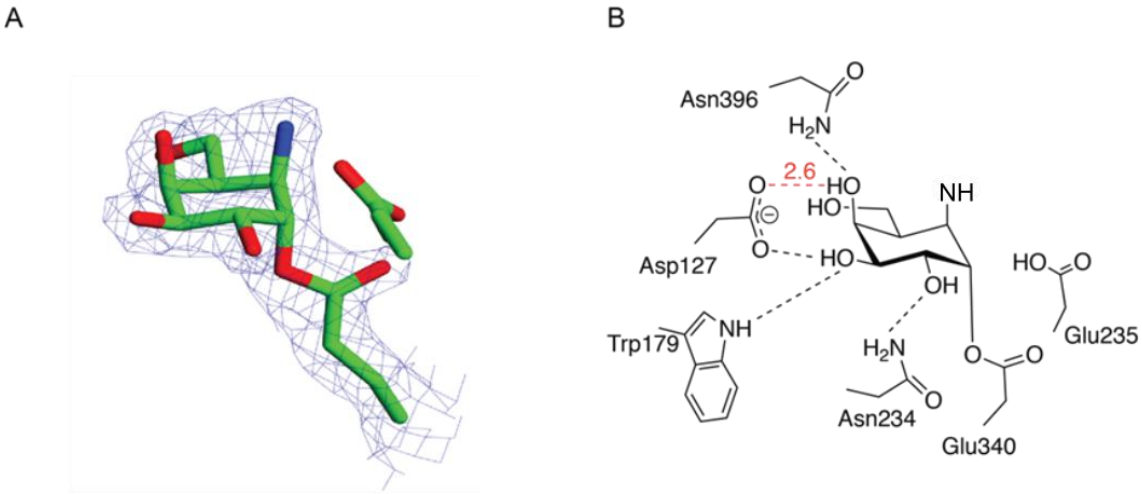


Figure 2.12. Binding of galacto-configured isomer of cyclophellitol, TB562 to GBA1. A) Observed electron density for TB562, contoured at 1.2 σ (equivalent to 0.31 electrons / Å3). B) H-bonding interactions (<3.0A) at the active centre. The only difference with gluco-configured probes is the interaction of an axial O4 with Asp217 (coloured red). Of note is the different position of O6 compared to its position in all gluco-ABP complexes

Of note, however, is that in the *galacto*-aziridine complex O6 has moved, perhaps to minimise steric clash with the axial O4, **Figure 2.12, 2.13**. One possible explanation for the strong cross-reactivity of the *galacto*-aziridine, **Figure 2.11**, but the non-reactivity of the 6-substituted *galacto*-expoxides reported by Willems (159) is this requirement to

displace the O6 hydroxyl for binding. This would not be possible were the O6 substituted, especially with a bulky side chain.

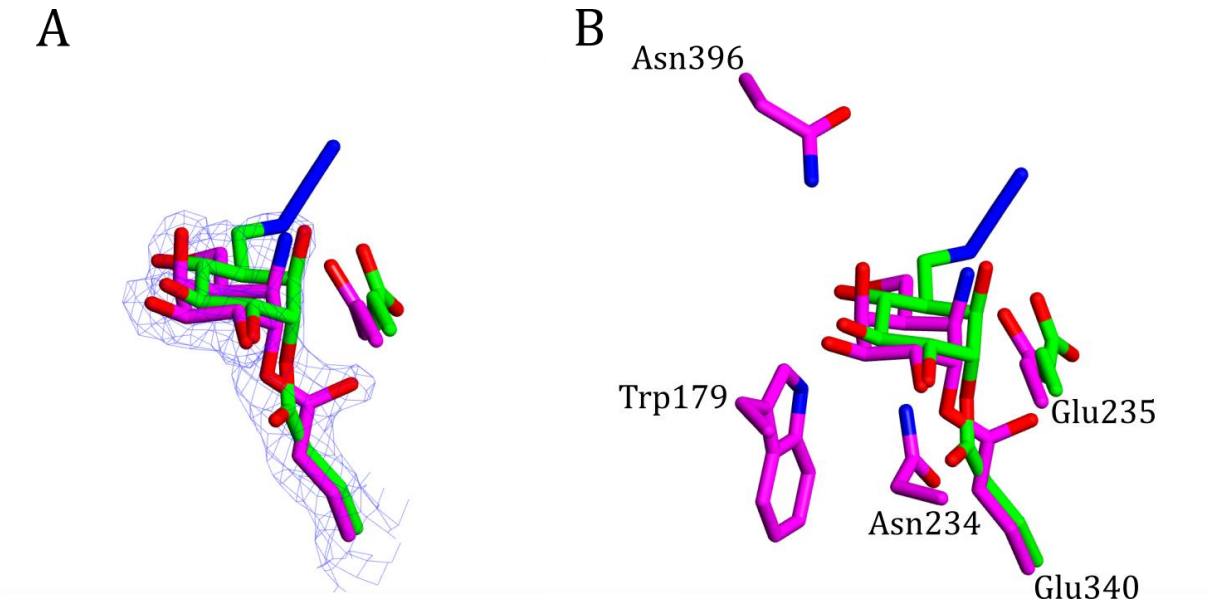


Figure 2.13. Binding of TB562 (in pink) to GBA1 superimposed with the binding of KY170 (shown in green) . A) Observed electron density for TB562, contoured at 1.2 σ (equivalent to 0.31 electrons / Å^3). B) Overlay of TB562 and *gluco*-epoxide KY170 highlighting the similar interactions but different position for O6.

2.4.5 3D complex of GBA1 with KY358 and MDW933

The goal of this chapter was to observe how activity-based probes bound, ideally to visualise how probe 'R-groups' could be accommodated at the 6-position as well as on aziridine nitrogens. Many soaks were performed in addition to those outlined above, often resulting in crystal disintegration. One complex, however, does begin to shed light on the potential interacting surfaces of GBA1. Serendipitously, a crystal of GBA1 was soaked with fluorescent epoxide MDW933 and data collected to 1. 58 Å and upon structure solution it was immediately apparent that the crystal had previously been exposed to / cross-contaminated by acyl aziridine KY358 allowing fortuitous access to a dual ligand structure.

The fluorescent probe MDW933 has been shown to effectively label femtomolar levels of GBA1, for visualisation on SDS-PAGE gels. This is a 10-fold reduction of protein input compared to the levels normally required for detection by Western blot, making the ABP ideal for the accurate detection of the relatively small amounts of recombinant enzyme expressed by transfected mammalian cells (141).

In the KY358/ MDW933 co structure, obtained at 1.58 Å resolution, the acyl aziridine KY358 is bound to the active centre nucleophile, as previously. The electron density is clear and unambiguous, **Figure 2.14.A** and the acyl tail extends through a hydrophobic channel with Tyr244 at its base and is better ordered than when observed previously. Indeed the N-butyl chain aligns well with the chain in the non-covalent inhibitor N-butyl-DNJ *(104),* ***Figure 2.14.B,*** *the amide carbonyl of this acyl aziridine is also able to make a hydrogen bond to the NH_2 of Gln284 presumably also adding to the inhibitory potential compared to KY353.*

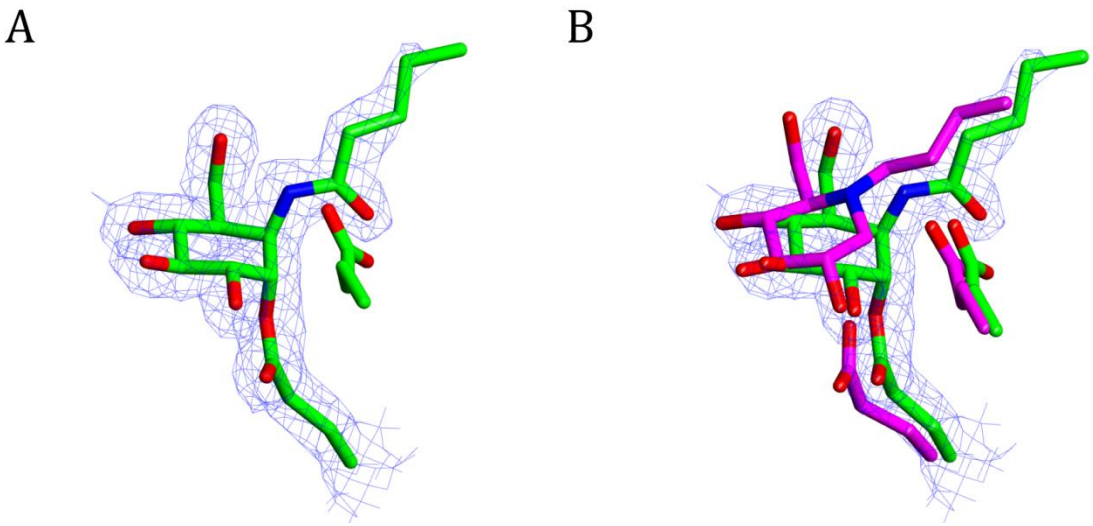


Figure 2.14. Binding of KY358 to GBA1A) Observed electron density for KY358 contoured at 1.2 σ (equivalent to 0.34 electrons / Å3). B) Superposition of the coordinates of 2V3D (104), shown in magenta, and the structure of Cerezyme_KY358, shown in green, atoms are coloured by type. In 2V3D, N-butyl-DNJ is present, unbound, in the active site.

KY358 must have been a cross-contamination prior to addition of MDW933. We were therefore fortunate that MDW933 also bound to the structure, close to the active centre and partially occupying the dimer interface, **Figure 2.15**. The entire length of MDW933 is observed and well ordered; the BODIPY fluorescent group, linker, triazole and cyclophellitol warhead. Whilst the binding may be fortuitous, it may give an indication of how the bulky fluorescent groups are actually accommodated in GBA1 (where indeed O6 substituents add to potency) (174)

Binding of MDW933 to recombinant GBA1 was also recorded by Witte and co-workers in 2010, and it was found to surpass both CBE and KY10 as an irreversible inhibitor of GBA1

with K_i value of 0.007 ± 0.002 (KY170: K_i = 0.044 ± 0.007 μM). A comparison of the relative rate constants demonstrated that the larger MDW933 probe inhibited GBA1 22-fold better than KY170. MDW933 was also used to successfully label GBA1 in mouse tissue lysate; however the probe was also seen to label the lactase-phloridzin hydrolase, LPH (127).

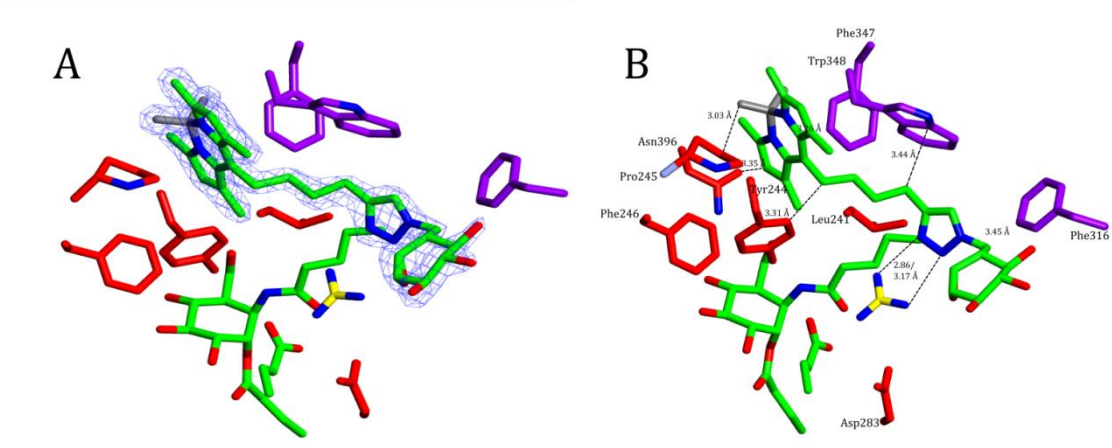


Figure 2.15. MDW933 present in the structure of GBA1, A) Observered electron density for MDW933 at 1.2 σ(equilvalent to 0.34 electons / $Å^3$) B) The surrounding region with side chains from amino acids from molecule A presented in red and those from molecule B presented in purple, atoms are coloured by type. A guanidinium ion from the crystallisation solution is also shown. The catalytic nucleophile and acid/base are presented in green, consistent with previous figures.

The BODIPY group sits at the surface of the A molecule, in a hydrophobic pocket formed by Tyr244 and Pro245 and Phe246, **Figure 2.15**. It is partially "sandwiched" by Trp348 and Phe347 from the adjacent B molecule of the dimer. The alkyl chain and triazole like in this predominantly hydrophobic channel in which Leu241 and Phe316 also contribute. A guanidinium ion (from the crystallisation solution) bridges the triazole to Asp283 of the protein. The epoxide ring sits, surface exposed making no direct H-bonding interactions with the enzyme but the hydrophobic face of the cyclophellitol lies above the exposed side chain of Leu241.

The binding of MDW933, across the enzyme surface may be "artefactual", but lots of evidence suggests that it does provide insight into the binding of the BODIPY group. Modelling, notably very simply torsional rotation of the linker allows the linker and the cyclophellitol to be places through the active centre channel, placing the epoxide in perfect

super-position with the trapped aziridine without *any* movement of the BODIPY group itself, **Figure 2.16**. This may give a good impression of how GBA1 accommodates, indeed prefers, the 6-modified cyclophellitols as activity-based probes.

Unfortunately, soaking of older GBA1 crystals with MDW933 alone resulted in disruption of the crystal lattice, with visible cracking on the crystal surface and no good data could be collected of this complex. This may be due to the large fluorescent moiety disrupting the molecule interfaces within the asymmetric unit or the age of the crystals at the time of testing. It is possible, but untestable, that the presence of the KY358 in the active site of GBA1 may be stabilizing the protein preventing any dramatic movement upon exposure to the larger MDW933 ligand and thus saving the structure of the crystal lattice. MDW933 has previously been confirmed to bind to GBA1 in solution unambiguously by mass spectrometry by Witte et al (127).

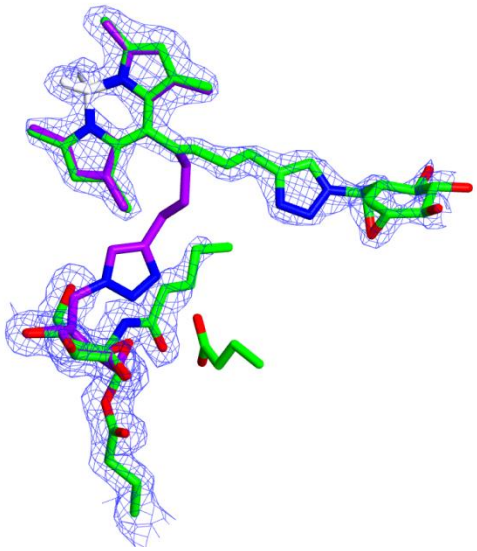


Figure 2.16 Rotation of the MDW933 ligand, leaving the BODIPY group in place, would allow the binding of the epoxide within the active site of the protein. Original placement of MDW933 and KY358 ligands are shown in green, whilst the rotated "alternate" position of MDW933 is shown in blue. Observered electon density for KY358 and MDW933 are shown at 1.2 σ (equilvalent to 0.34 electons / $Å^3$.)

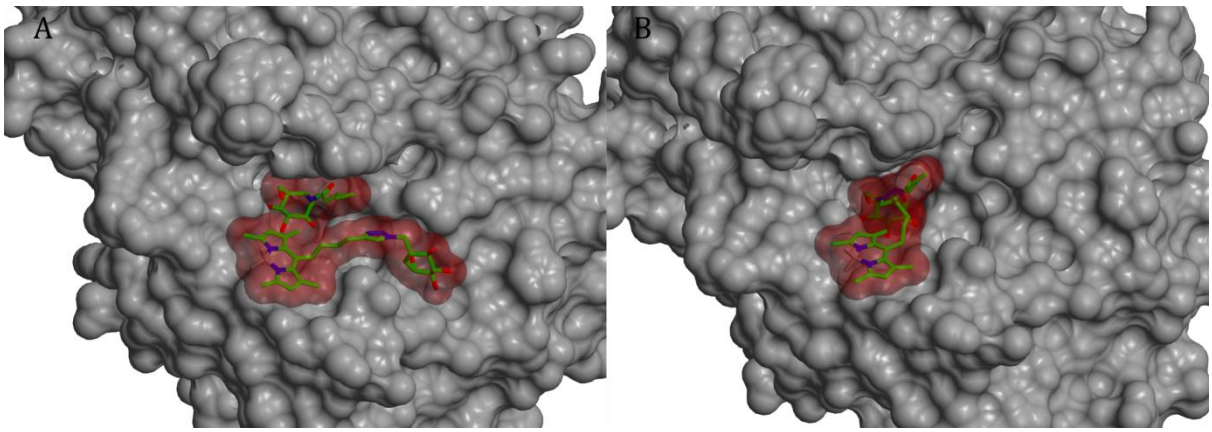


Figure 2.17. A) Surface model GBA1 showing the binding of KY358 in the active site of the protein and MDW933 on the dimer surface. B) MDW933 is also able to rotate and bind into the active site of GBA1, the BODIPY group remaining in place on the surface.

Why is GBA1 able to accommodate substituents at the 6-position (as well as R groups attached to an aziridine nitrogen)? The natural substrate for GBA1 is glucosylceramide, **Figure 2.1**, which has two extended acyl chains that the enzyme needs to accommodate. The aglycon (leaving-group) region of the active site therefore needs to be both hydrophobic and wide enough for two alkyl chains. Inspection of the structure certainly shows that the site can accommodate two chains and, crucially, the O6 of glucose is fully "solvent" exposed and allows free access from the hydroxyl to the hydrophobic surface, **Figure 2.16**, and indeed to the platform occupied by the BODIPY group described above.

2.5 Summary

Crystallographic studies were used to probe the 3D binding of a series of *gluco*-epoxide and *gluco*-aziridine probes and a *galacto*-aziridine with the human GBA1 enzyme, this latter showing the capability of the enzyme to accommodate galactosyl ligands.

A complex captured with both a covalently-bound aziridine in the active site and a non-covalent epoxide with a large florescent group highlight the positions in which the GBA1 enzyme may accommodate other large substrate analogues. This is a key target in the hunt for inhibitor/ABP specificity.

The GBA1 structures reported here immediately suggest a series of "next-generation" inhibitors which could be (a) aziridine-based; allowing hydrophobic chain extension from the nitrogen but also (b) substituted at O6 to allow exploitation of the wider hydrophobic surface and putative dye-binding region (**Figure 2.17**). Such compounds, generic formula below, are currently being synthesised in the Overkleeft group and will be discussed in the final "**Conclusions and Perspectives**" chapter at the end of the thesis.

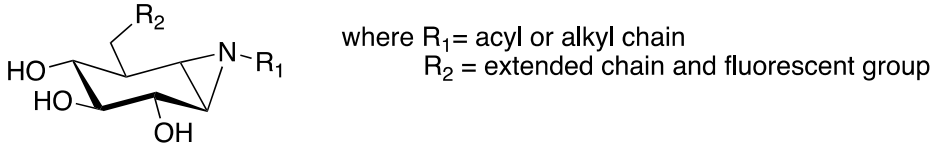


Figure 2.18. Design of putative bi-substituted GBA1 probe based upon 3D structure. Such compounds are currently being synthesised and assessed by the Overkleeft group in *Leiden.*

The ideal goal for monitoring glucosidase activity in human cells would be to obtain specific ABPs for both GBA1 and the cytoplasmic enzyme GBA2; ideally compounds that are specific for each enzyme. The work by the Overkleeft group, illustrated here at the 3D level, has gone a long way to imaging GBA1. At the start of this thesis, no 3D structure for GBA2 or its sequence homologs (CAZY family GH116) were known limiting prospects for rational inhibitor / probe design. Such work is described in the following chapters **3 & 4**.

Chapter 3. Purification and crystallisation of a bacterial GH116 enzyme

3.1 Abstract

The human enzyme GBA2 (CAZy family GH116) is responsible for hydrolysis of glycosphingolipids on the cytoplasmic membrane surface of the Golgi apparatus and the endoplasmic reticulum. In humans, genetic defects in GBA2 may result in hereditary spastic paraplegia and autosomal recessive cerebellar ataxia. Modulation of the activity of GBA2 has been shown to alleviate Gaucher's disease pathology in mice. Therefore GBA2 has become a viable target for treatment pathways. At the start of this thesis, little was known about CAZy family GH116 enzymes, and none had been structurally solved. Bacterial GH116 homologs of the GBA2 enzyme can be used for structural modelling of the human protein and so many were screened for expression. Ultimately two constructs of the GH116 family retaining β-glucosidase from *Thermoanaerobacterium xylanolyticum* were purified to homology for structural characterization.

*Some of the work described in this chapter has been published in Charoenwattanasatien et al., (2016) Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). *ACS Chem Biol* **11**, 1891-1900

3.2 Introduction

3.2.1 GBA2 – non-lysosomal glucosylceramidase

Many important cell surface markers, such as gangliosides which are vital in cell signalling, recognition and play a role in the progression of some infections such as leprosy (8), are constantly degraded, recycled and re-synthesised. One of the key outcomes of this process is the necessarily constant turnover of glucosylceramide with in the cell.

Primarily, the catabolic pathway for glucosylceramide is via the lysosome, where the glycolipid is broken down by GBA1 into glucose and ceramide (see **Chapter 1**). Mutations in GBA1 can lead to a loss of enzyme activity, and thus results in a build-up of glucosylceramide within macrophages, giving them a gritty appearance (described as Gaucher cells). In many other cells types of Gaucher's patients, this build-up is not seen, leading to the conclusion that there is a secondary catabolic pathway for glucosylceramide(175).

Although GBA1 is known as the primary enzyme responsible for glucosylceramide degradation, human cells in fact contain three acid β-glucosidases; the lysosomal GBA1 is the enzyme in which functional mutations can lead to Gaucher's disease. This enzyme is discussed in more detail in **Chapter 2**.

The non-lysosomal acid-β-glucosidase, which had previously been described as a "bile acid β-glucosidase" (176) until it was recognised as being present in all cells(175) is identified GBA2. GBA3 is the neutral, cytosolic beta-glucosidase that is primarily localised to the liver. Although GBA3 does have glucosylceramidase activity, it does not appear to play a significant role in the modulation of Gaucher's disease(139). (**Table 3.1).**

TABLE 3.1 GBA ENZYMES AND THEIR ROLES

	GBA1	**GBA2**	**GBA3**
Alternative names	Lysosomal acid β-glucosidase	Non-lysosomal glucosylceramide, Bile acid β-glucosidase	Cytosolic beta-glucosidase, Klotho-related Protein
EC number	EC 3.2.1.45	EC 3.2.1.45	EC 3.2.1.21, EC 3.2.1.45
CAZy family	GH30	GH116	GH1
Location within the cell	Lysosome	Cytosolic membrane of ER/Golgi	Cytosol, neutral pH
Associated disease	Gaucher's disease(177)	ARCA, HSP(178)	Unknown (138)

GBA2 breaks down glycosylceramide (GluCer) in the cytoplasmic face of the Golgi apparatus and the endoplasmic reticulum. The degradation of GluCer by GBA2 may be responsible for the lower levels of GluCer in some cells of Gaucher's patients and the resulting modulation of overall disease pathology.

Whilst GBA1 and GBA2 both act on the same substrate (GluCer), they differ in their reactions towards particular inhibitor compounds. For example GBA1 is inhibited by conduritol-B-epoxide (CBE) whereas GBA2 is comparatively insensitive to it (130). Conversely, GBA2 is more sensitive to inhibition by deoxynojirimycin analogues compared to its lysosomal counterpart.

GBA2 activity is optimal at cytosolic pH 5.5-6.0, in comparison to GBA1 which is functional at the lower lysosomal pH 3.5-5.5 (179). Activity of GBA2 is affected by whether the enzyme is membrane associated or not. Presence of the membrane can also affect how the enzyme is affected by inhibitors – Ridley et al found that GBA2 was apparently more sensitive to CBE when associated with the cell membranes (179).

The two retaining β-glucosidases have distinct amino acid sequences and 3D structures and belong in separate CAZy glycosidase hydrolase families; GBA1 in family GH30 and GBA2 in family GH116.

Both GBA1 and GBA2 show activity towards a second natural substrate, bile acid-3-*O*-β-glucoside. This was initially thought to be the main substrate of GBA2. However deficiencies in humans of either GBA1 or GBA2 do not lead to the accumulation of bile

acid-3-*O*-β-glucoside within the cell, leading to the hypothesis that one enzyme is able to compensate for the action of the other (180).

Potentially of clinical relevance, the drug miglustat (NB-DNJ, **Chapter 2, Figure 2.5**), used for the treatment of Gaucher's disease, was found to inhibit GBA2, IC_{50} = 6 nM (130,179). The targeted used of miglustat was substrate reduction therapy, with the goal to reduce the synthesis of glucosylceramide (inhibition of the enzyme glucosylceramide synthase).

Miglustat is also a micromolar inhibitor of GBA1 (IC_{50} = 400 μM(107)), and the active site binding of the inhibitor has the secondary effect of stabilising certain mutant forms of GBA1, providing the basis of pharmacological chaperone therapy. However at most therapeutic doses miglustat will be completely inhibiting GBA2 activity (181).

Patients with Niemann-Pick type C1 disease are also prescribed miglustat to reduce glucosylceramide synthase activity. In mouse models, miglustat has been shown to be acting off-target, as an inhibitor of GBA2, leading to increased CNS glucosylceramide levels (182) .

3.2.2 GBA2 in human disease

GBA2 itself is linked to many medical conditions outside of Gaucher's disease. *GBA2*-knockouts in mice show male infertility but no Gaucher's disease phenotype (175). However they do accumulate glucosylceramide in the tissues which most predominantly express GBA2, such as the liver, brain and testes. In comparison, mice with Gaucher's disease, which are GBA1 deficient, present with a variety of issues including splenomegaly and neurological symptoms, a pathology similar to human Gaucher's disease. (183). Zebrafish with the *GBA2* knockout mutation also show some neurological defects. However again it is distinct from Gaucher's disease phenotype. (131)

In humans, different mutations in GBA2 can lead to different hereditary diseases such as autosomal-recessive cerebellar ataxia and hereditary spastic paraplegia (131-133). GBA2 therefore seems to be necessary for healthy neuronal development. Unlike Gaucher's disease, in which a variety of mutations in GBA1 all lead to the same disease outcome, varies point mutations in GBA2 are linked to different clinical consequences.

There are five known missense mutations that may occur in human GBA2 leading to disease (131-133,178,184), shown in **Table 3.2**. Whilst many of these mutations can lead to the similar pathology in patients, often described as spastic paraplegia and cerebellar ataxia, the location of the mutations through the three dimensional protein structure may give further insight into the effect that these changes are having on the enzyme's active site as well as protein stability and potential domain interaction.

Expression of *GBA2* can be upregulated in malignant human melanoma cells *in vitro*, lowering GluCer levels and resulting in higher ceramide levels. This results in decreased cell growth and eventually apoptosis (137). *GBA2* is down-regulated within melanoma cells and could provide a target for cancer treatment.

Significantly, Gaucher's disease mouse models show a reduction in symptoms when *GBA2* was knocked out (134), whilst the overexpression of GBA2 was shown to be toxic to melanoma cells (137). This suggests that modulation of the behaviour of GBA2 by specific inhibitors could be useful for the treatment of Gaucher's diseases whilst the use of activators could possibly be used to combat melanoma.

The activity of GBA2 in Gaucher's disease cells is still not fully understood, although it appears it is down-regulated in the absence of GBA1 activity, rather than compensating for the loss of the lysosomal β-glucosidase enzyme (135,136). Deletion of GBA2 in fact partially relives Gaucher's disease pathology in mouse models, suggesting that one factor of Gaucher's pathology is the accumulation of the toxic sphingosine, produced at the cytoplasmic face of the ER and Golgi apparatus by GBA2. (134)

TABLE 3.2 MUTATIONS IN GBA2 AND THEIR ASSOCIATED PATHOLOGIES

GBA2 mutation	**Associated pathology**
F419V	Hereditary spastic paraplegia/cerebellar ataxia (SPG46)(133)
D594H	Spastic ataxia (184)
R630W	Hereditary spastic paraplegia(131)
G683R	Hereditary spastic paraplegia (178)
R873H	Hereditary spastic paraplegia/cerebellar ataxia (SPG46)(133)

Following research which showed the roles played by GBA2 in spastic paraplegia and cerebellar ataxia (131) as well as the effect that it may have in the modulation of Gaucher's disease (134), understanding and control of the function of GBA2 has become an important step for the treatment of these diseases.

The membrane-associated nature of GBA2 has made it hard to purify the human enzyme. Initial study of GBA2 was limited by low volumes of protein and the necessary use of detergent.

3.2.3 GH116 enzymes

All glycosidases hydrolyse the bond between a carbohydrate and either another carbohydrate or a non-carbohydrate moiety (**Figure 3.1**). Those with a retaining mechanism end their reaction with a net retention of the sugar ring configuration. This is done by a two-step process involving a covalent intermediate (**Chapter 1, Figure 1.9.B**)(118).

Human GBA2 belongs to glycoside hydrolase family 116, a family much less well studied than that of GBA1 (GH30). Until recently, relatively little was known about this group of enzymes. The initial purification of GBA2 from the human liver characterised the enzyme as bile acid β-glucosidase in 1996(176) but it wasn't until 2007 that is was recognised as the non-lysosomal β-glucosidase, GBA2 (175).

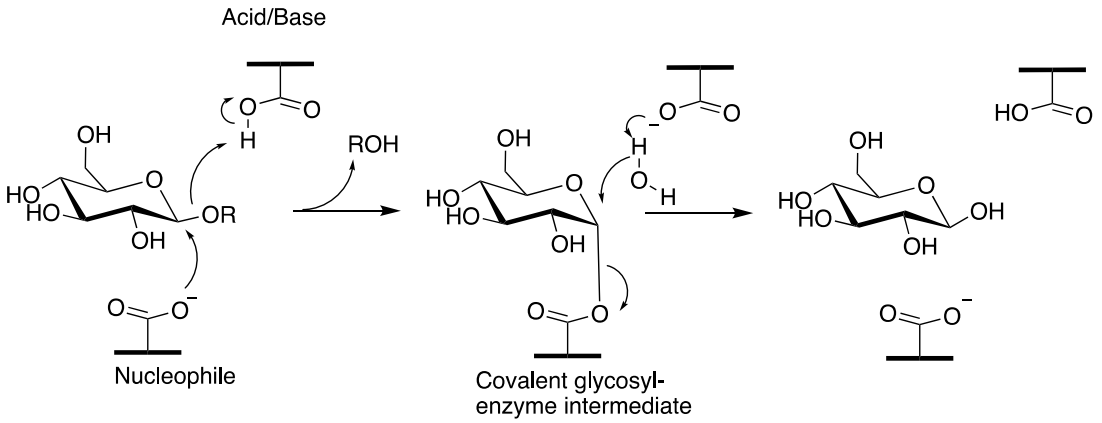


Figure 3.2 A reaction schematic for a retaining β-glucosidase, such as GBA2 or GBA1. The catalytic acid/base is shown at the top whilst the catalytic nucleophile is below.

Work on the molecular characterisation of a glycoside hydrolase from the archaeon *Sulfolobus solfataricus* P2 (SSO1353, by Cobucci-Ponzano et al in 2010), led to the creation and classification of CAZy family GH116 (129). As of August 2017, there are now almost 400 enzymes identified as belonging to the GH116 family, although only five have been functionally characterised.

Whilst all GH116 enzymes are retaining enzymes, with a specificity for β-glucosides, the GH116 family can be further divided in to three "subfamilies" based on other substrate specificities of the enzymes and the sensitivities to the certain competitive inhibitors (185). For example, SSO1353 was seen to be sensitive to conduritol β-epoxide (CBE) inhibition, in contrast to human GBA2 which is insensitive to CBE (129).

GH116 subfamily 1 contains enzymes specific for glucosylceramides, that are sensitive to nM levels of NB-DNJ but are insensitive to CBE; this included the human GBA2 (175). GH116 subfamily 2 was characterised by an enzyme identified as SSO3039, also from the archea *Sulfolobus solfataricus* P2. SSO3039 is a bifunctional β-glucosidase/N-acetyl-β-glucosaminidase, with sensitivity to both NB-DNJ and CBE at μM and mM concentrations respectively (185).

SSO1353, the original GH116 enzyme, belongs to subfamily three and was seen to hydrolyse β-glucosides and β-xylosides bound to hydrophobic groups. Subfamily 3 enzymes have mM sensitivity to both NB-DNJ and CBE. SSO1353 is also capable of the transglycosylation of β-D-xylo-oligosaccharides (129).

Given the absence of a 3D model of any human GH116 family proteins, bacterial homologs can be used to produce a structural model. A Basic Local Alignment Search Tool (BLAST) sequence searches produces two bacterial GH116 enzymes with >37 % sequence identity to the human GBA2, over 830 residues (**Table 3.2**)(186). Over the catalytic domain, sequence identity is higher still, with the GH116 enzyme from *Thermoanaerobacter xylanolyticum* sharing approximately 40% sequence identity with the human counterpart.

TABLE 3.3 GBA2 SEQUENCE IDENTITY ACROSS SPECIES

GH116 Source	**No of amino acids**	**Sequence Identify to Human GBA2 (%)**
Homo sapiens	927	100.00
Pan troglodyte	927	99.46
Mus musculus	918	87.58
Drosophila melanogaster	948	43.91
Arabidopsis thaliana (L.) Heynh.	947	39.15
Thermoanaerobacter xylanolyticum	787	37.98
Caenorhabditis elegans (Bristol N2)	959	37.49
Arthrospira plantensis	799	37.42
Bacteriodes thetaiotamicron	825	24.60
Sulfolobus solfataricus SSO3039	803	23.61
Sulfolobus solfataricus SSO1353	663	22.77

Sequence search performed using BLAST(186).

The full sequence alignment is included in **Appendix 1**.

3.2.4 Aims of the research in this chapter

Here I describe the recombinant expression of three bacterial β-glucosidase GH116 enzymes. Following conformation of the activity of the enzyme in whole cell lysate was performed using activity based probes, a protocol for the production and purification of the β-glucosidase GH116 enzyme from the bacterium *Thermoanaerobacterium xylanolyticum, (Tx*GH116) was designed. Subsequently, TxGH116 was crystalized and the three dimensional structure of the protein solved (described in **Chapter 4**). TxGH116 has a sequence homology of 38% with GBA2, over residues 63 to 893 in the human enzyme (**Table 3.2**). Expression, purification and subsequent study of this enzyme make it an informative homology model for the human protein.

3.3 Material and Methods

3.3.1 Sequence Alignment

A Basic Local Alignment Search Tool (BLAST) search was performed using the human GBA2 protein sequence and five GH116 enzymes (confirmed via the CAZy database) were selected from a range of Kingdoms. Also included were the three bacterial enzymes studied here and the archaeal GH116 from *Sulfolobus solfataricus* SSO1353, which led to the classification of CAZy family GH116 in 2010 (129)

A sequence alignment of ten GH116 family retaining β-glucosidases was performed using the T-Coffee server (tcoffee.crg.cat) and formatted using Espript (espript.ibcp.fr) (187,188).

3.3.2 Cloning and transformations of bacterial GH116 enzymes

In order to optimise chances of successful protein, bacterial GH116 homologs the genes encoding GH116 enzymes from *Thermoanaerobacterium xylanolyticum*, *Arthrospira plantensis* and *Bacteriodes thetaiotamicron* were cloned into pET28a(+) (30 ug/ml kanamycin resistant) between NdeI and XhoI restriction sites (clones synthesised by Genscript, Inc, codon optimised for *E. coli* expression). A His_6-tag was included at the N terminus of the protein sequence, replacing the first 20 amino acid residues, with a thrombin cleavage site between the protein and the tag. These protein constructs were designated TxGH116N, ApGH116N and BtGH116N respectively. Full sequences of the three constructs are in **Appendix 2**.

BL21 (DE3) *E.coli* cells were transformed with the pET-28a(XxGH116) and grown in 5 ml LB media containing 30 mg L^{-1} kanamycin overnight at 37 °C overnight with aeration by shaking at 180 rpm. This culture was then used to inoculated a larger (50 ml) volume of autoclaved LB media containing 30 mg L^{-1} kanamycin and expression of the protein encoding genes were induced by the addition of 1 mg L^{-1} isopropyl-β-D-1-thiogalactopyranoside (IPTG) when the culture reached an OD600 of 0.6. Cells were harvested by centrifugation and subsequently lysed by sonication at 14-18 Hz and expression of the proteins of interest was confirmed by SDS-PAGE gel.

3.3.3 Overexpression and purification of TxGH116N

TxGH116N purification was successfully scaled up. Cells containing pET-28a(TxGH116) were cultured in LB media containing 30 mg L^{-1} kanamycin until an OD600 of 0.6 had been reached. Expression of the TxGH116 gene was induced with 0.5 mg L^{-1} IPTG and the cells

further incubated 16 °C for 18 hr. A reduced concentration of IPTG and slower growth was found to improve protein production.

Cells were subsequently lysed by sonication at 14-18 Hz, four runs of 40-50 second exposure, in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole). Cell debris were removed by centrifugation for 15 minutes in a Sorvall SS-34 rotor at 38 000g.

TxGH116N was removed from the resulting soluble fraction using a two-step purification procedure, consisting of nickel affinity chromatography of the His_6-tagged protein and subsequent size-exclusion chromatographic separation. Total soluble cell lysate was loaded onto a 5 ml GE Healthcare crude HisTrap column, equilibrated with buffer A. Unbound material was removed by washing with three CV of buffer A. Bound protein was eluted with a gradient of buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole) applied over 20 CV.

The fractions containing TxGH116N were pooled, concentrated and buffer exchanged into buffer GF (50 mM HEPES pH 7.5, 500 mM NaCl) to reduce imidazole concentration. The protein was then loaded onto a Superdex S200 gel filtration column equilibrated in buffer GF and the elution of the target protein was performed at 2 ml/min. The fractions containing TxGH116N pooled, concentrated to the required concentration and stored at 5 °C.

3.3.4 SEC-MALLS

Size Exclusion Chromatography - Multi-Angle Laser Light Scattering (SEC-MALLS) was used to determine the molecular mass of both soluble-state TxGH1116N and TxGH116C. A Superdex 20 HR 10/30 column was connected to an HPLC system, which was coupled to a DAWN Heleos lighter scattering instrument with a laser wavelength of 658.00 nm (Wyatt Technology).

TxGH116N was run at 125 nM in buffer containing 50 mM HEPES pH 7.5 150 mM NaCl at 0.5 ml/min. TxGH116C (**Chapter 4**) was run separately at 125 nM in buffer containing 50mM HEPES pH 7.5 150 mM NaCl at 0.5 ml/min. Both samples were centrifuged for *ca.* 5 min before being run.

Data analysis of both experiments was done using ASTRA software interface from Wyatt technology. The absolute masses of the particles in the solution could be obtained in real-time.

3.3.5 Activity Based Probes

Total cell lysate from 5 ml overnight growth cultures of all three bacterial BL21 (DE3)-pET-28a(XxGH116) constructs were incubated with the activity based probe designated KY375 (**Figure 3.2**) at room temperature for 30 mins at a final concentration of 2 mM. The samples were resolved using SDS-PAGE and scanned for fluorescence using BIO-RAD Molecular Imager FX using excitation/emission maxima of 503/512 nm.

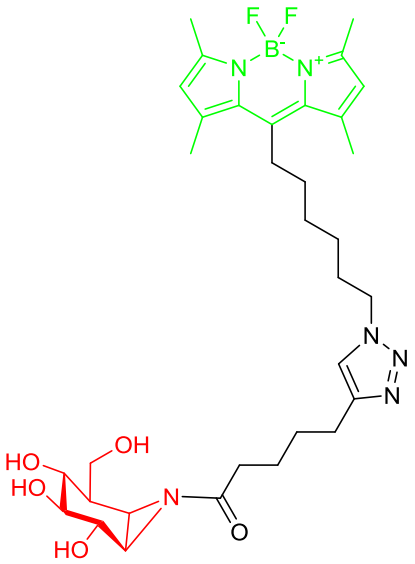


Figure 3.2 The cyclophellitol-aziridine activity based probed KY375 used in this chapter. It is specific towards GH116 enzymes. Red: The "warhead", chemically active targeting moiety. Black: linker. Green: A (boron-dipyrromethene) BODIPY fluorescence group.

3.3.6 Crystallisation of TxGH116N

Purified *Tx*GH116N was screened for crystallisation hits against various commercially available screens using sitting drop vapour diffusion in 96-well plate format. All hits were then fine-screened in 48-well plate format to test and improve crystal size and diffraction quality. Initial crystals of TxGH116N formed after 9-11 weeks and were used as a seed stock to increase the growth speed of *Tx*GH116N crystals.

Best *Tx*GH116N crystals were grown at 19 °C over a period of six weeks with equal volumes of protein (35 mg/ml) and mother liquor solution containing 0.1 M Bistris propane pH8, 0.2 M $(NH_4)_2\ SO_4$ PEG 6K (polyethylene glycol, molecular weight 6 kDa) 15% (w/v).

A selenomethionine variant of *Tx*GH116-N was produced by methionine auxotrophic cells (the same line of *E. coli* BL21 (DE3) cells used) grown in PASM-5052 auto-induction media. The purified SeMetTxGH116 was tested both similar crystallisation conditions as the native protein and also rescreened from initial trials. Once the structure of *Tx*GH116C was published by Sansenya et al (2015)(189) however, further selenomethionine protein growth was deemed unnecessary.

3.3.7 Data collection

Initially crystals of TxGH116N were individually transferred to cryoprotectant solution of the mother liquor combined with 20 % glycerol (v/v) and immediately placed in liquid N_2 at 100K. However, after seeing some issues with crystals cracking on exposure to cryoprotectant, several crystals were frozen without cryoprotectant.

Diffraction quality of all *Tx*GH116N crystals was initially tested in house using a RUH3R rotating anode X-ray generator. Diffraction images were sampled at angles of 0 and 90° using 0.5° oscillations. The best diffracting crystals were then sent to the Diamond Light Source in Oxfordshire for data collection.

3.4 Results

The genes encoding for bacterial β-glucosidases, ApGH116N, BtGH116N and TxGH116N, were successfully expressed in *E.coli.* TxGH116N was successfully purified, and crystallised in a form suitable for further structural study.

3.4.1 Sequence alignment

BLAST (186) sequence search tool with the *Sulfolobus solfataricus* β-glucosidase SSO1353 returns a 23% identity match to the β-glucosidase from *Thermoanaerobacterium xylanolyticum* (TxGH116), with TxGH116 having a 38% similarity across 830 residues when aligned to the human GBA2 enzyme.

The results from the T-Coffee alignment (**Figure 3.3**) show that many potentially key active site residues and as well as the five known GBA2 disease mutations, shown in **Table 4.1**, are conserved between TxGH116 and human GBA2.

The full protein sequence alignment is included in **Appendix 1**.

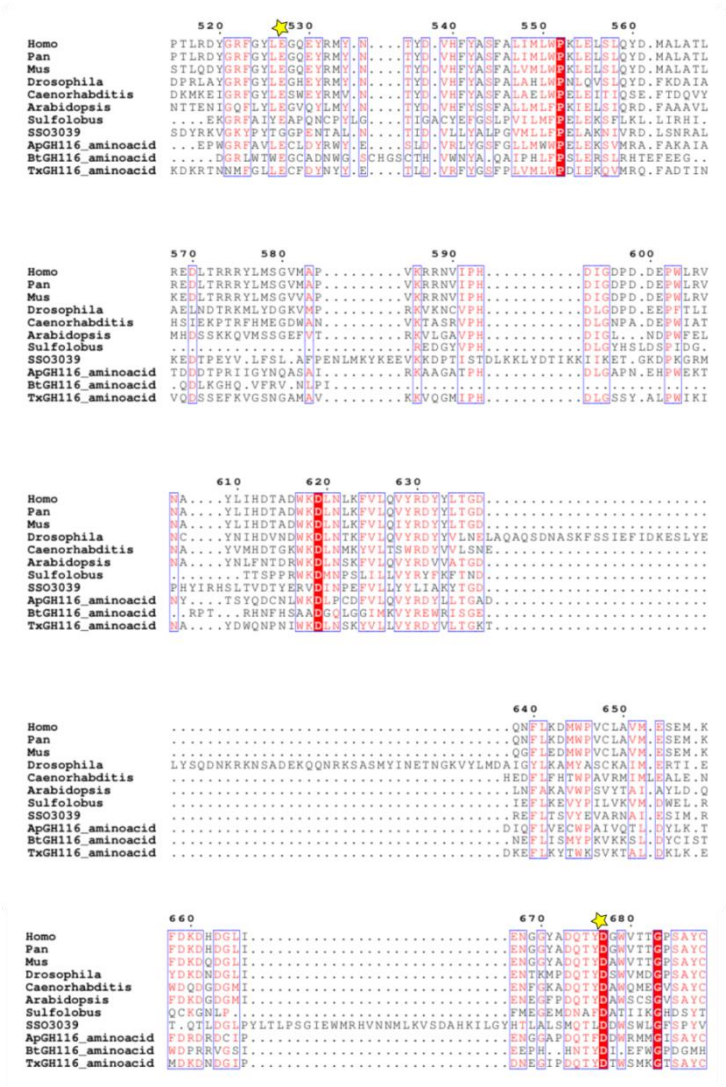


Figure 3.3. A section from a multi-alignment of various GH116 β-glucosidase protein sequences. Invariant residues are highlighted in red. Areas of high conservation are boxed in blue. Identified catalytic residues are marked with a star. Homo is NP_065995.1 from *Homo sapiens.* Pan is XP_001167923.1 from *Pan troglodytes.* Mus is CAM17042.1 from *Mus musculus.* Drosophila is AA041192.2 from *Drosophila melanogaster.* Caenorhabditis is CAD21661.1 from *C. elegans* Bristol N2. Arabidopsis is BAE.99039.1 from *A. thaliana.* Sulfolobus is SSO1353 from *S. solfataricus* P2. SSO3039 is SSO3039 from *S. solfataricus* P2. ApGH116 is BAI91082.1 from *A.platensis* NIES_39. BtGH116 is AA078014.1 from *B. thetaiotaomicron* VP1-5482. TxGH116 is AEF18218.1 from *T. xylanolyticum.* Alignment made using the T-Coffee server (tcoffee.crg.cat) and formatted using Espript (espript.ibcp.fr) (187,188).

3.4.2 Expression of Bacterial GH116 enzymes

Initial expression tests were performed with 50 ml culture volumes of BL21 (DE3) cells containing pET-28a(XxGH116). Growth conditions (temperature post induction, shaking speed), media and IPTG levels were varied to find optimal growth conditions for maximum protein production.

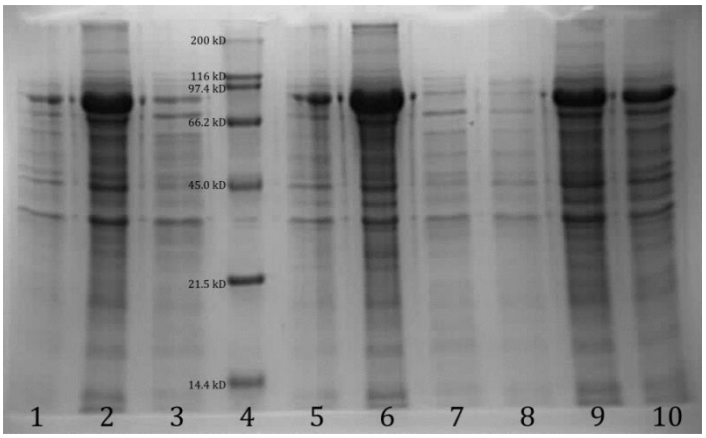


Figure 3.4 SDS-PAGE gel showing expression of bacterial GH116 enzymes by BL21 (DE3) *E. coli* cells. Lanes 1-3) Cell lysate containing pET-28a(ApGH116) pre-induction, 4 hours post-induction with 0.5 mg L^{-1} IPTG, soluble fraction in 50 mM HEPES pH 7.5, 200 mM NaCl only. Lanes 5-7) Cell lysate containing pET-28a(BTGH116) pre-induction, 4 hours post-induction with IPTG, soluble fraction only. Lanes 8-10) Cell lysate containing pET-28a(TxGH116) pre-induction, 4 hours post-induction with IPTG, soluble fraction only. Lane 4 contains Bio-Rad Broad Range Molecular Weight Standards protein ladder.

The SDS-PAGE (**Figure 3.4**) shows that, while all three constructs were expressed, only BL21 (DE3) cells containing pET-28a(TxGH116) produced soluble protein when induced with 0.5-1 mg L^{-1} IPTG. Both ApGH116 and BtGH116, although later proved to be active protein in whole cell lysate, were not readily soluble when the lysed cell pellet was resuspended in a range of buffers and so TxGH116 was selected to scale up protein production.

3.4.3 Purification of TxGH116N

Nickel chromatography was used to obtain the His_6-tagged TxGH116N protein from the cell lysate following sonication (**Figure 3.5**). The protein was then further purified by gel filtration (**Figure 3.6**) to give a final yield of ~ 7 mg /L of cell culture.

Gel filtration of TxGH116N consistently gave two peaks (**Figure 3.6.A)**. However, when samples from each of these peaks were run on an SDS-PAGE gel they showed protein of equal mass, with no obvious contaminents (**Figure 3.6.B**).

A SEC-MALLS run of TXGH116N (**3.4.4**) showed that the protein appeared to be behaving as a dimer in solution when at lower concentrations, and a dimer/monomer mix at high concentrations. The secondary peak resulting from the gel filtration may be the monomer protein eluting later than the larger dimer form, that when denatured and resolved on an SDS-PAGE gel both appear as the same sized molecule.

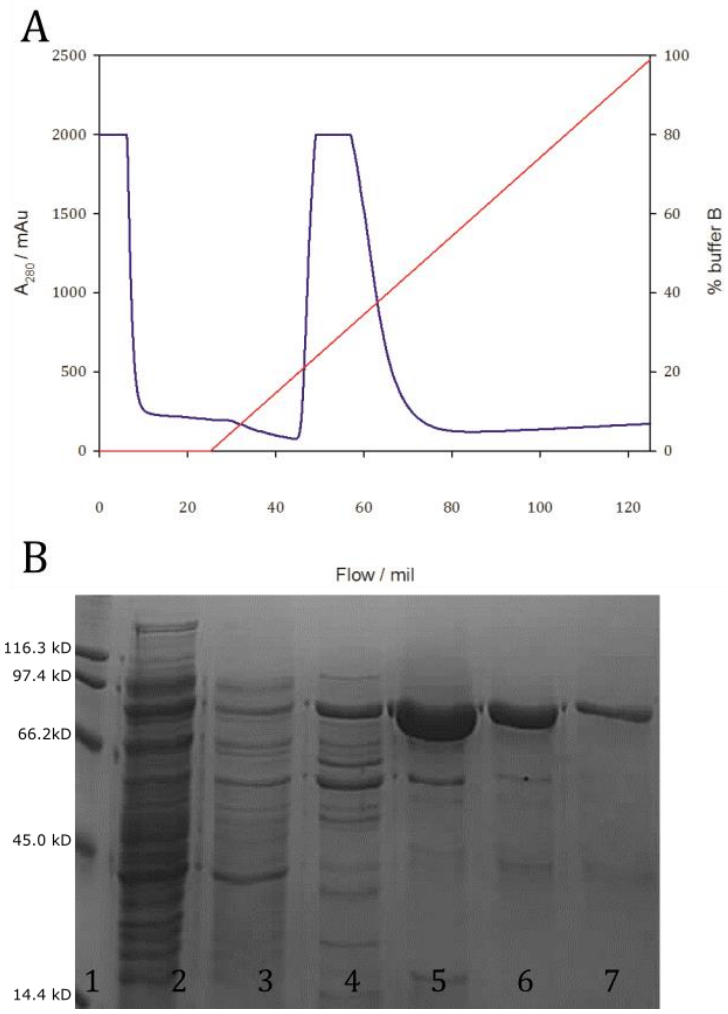


Figure 3.5 A) Chromatogram showing the elution of the TxGH116N-His_6 protein from a HisTrap nickel column with an increasing concentration of buffer containing imidazole. Protein is measured by the UV absorbance at 280 nm (A_{280}). B) An SDS-Page gel of fractions taken from the peak between ~45-70 ml (lanes 5 -7). Lane 1 is a broad range molecular weight ladder (Bio-Rad), lane 2 is a sample of the total cell lysate, lane 3 is a sample of the flow through (1-10 ml) and lane 3 is a sample of the total protein that was loaded on to the column. TxGH116N has an expected MW of 91 kDa.

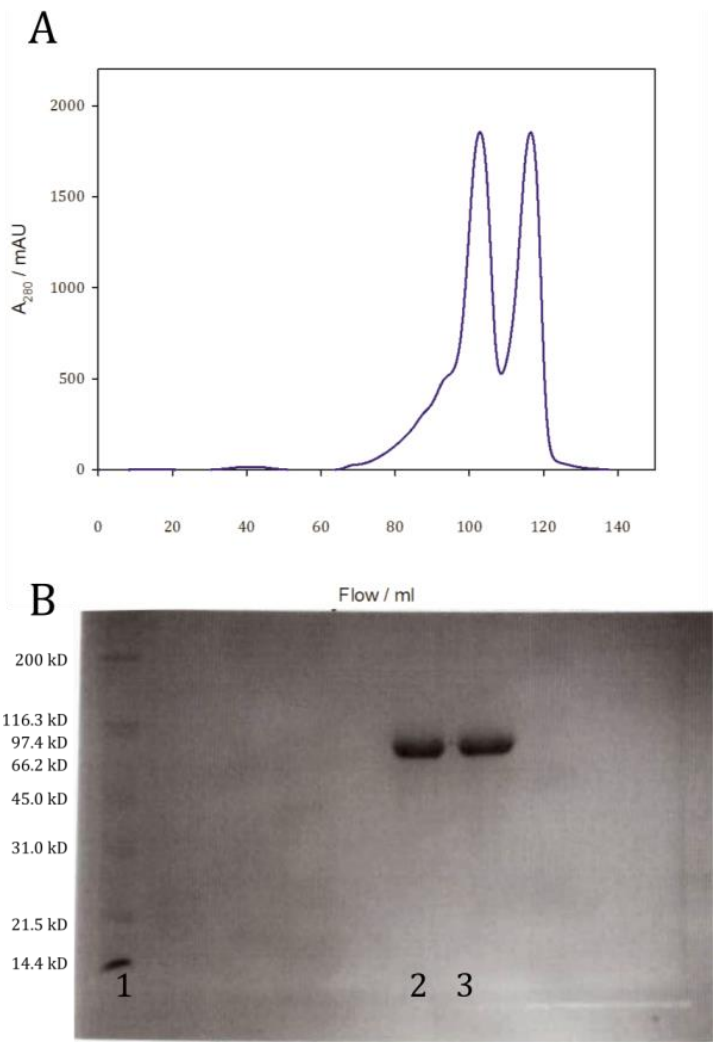


Figure 3.6. A) Chromatogram showing the elution of the TxGH116N-His_6 protein from an S200 size exclusion gel filtration column. Protein is measured by the UV absorbance at 280 nm (A_{280}). B) An SDS-Page gel of fractions taken from the two peaks between ~95-120 ml (lanes 2 -3). Lane 1 is a broad range molecular weight ladder (Bio-Rad). TxGH116N has an expected MW of 91 kDa.

3.4.4 SEC-MALLS

The results of the SEC-MALLS run of TxGH116N showed that the protein was behaving mostly as both a dimer in solution (**Figure 3.7.A**). This is in agreement with the crystal data discussed in **Chapter 4**, where two molecules of TxGH116N were seen in the asymmetric unit of the $P6_1$ crystal form.

However, the $P2_12_12$ crystal form of TxGH116C has only one molecule of the enzyme in each asymmetric unit. SEC-MALLS of the TxGH116C construct of the protein shows the enzyme behaving as a monomer in solution (**Figure 3.7.B**).

The presence of the C-terminal His_6-tag is likely disrupting the dimer interface of the protein. However this has little or no effect on the catalytic activity of the enzyme, as measured by Charoenwattanasatien and co-workers (190).

An earlier SEC-MALLS run of TxGH116N at 250 nM was also run and published in Charoenwattanasatien et al 2015 (190). The secondary run at a lower concentration was performed to determine whether the dimer behaviour was an artefact of a higher protein concentration.

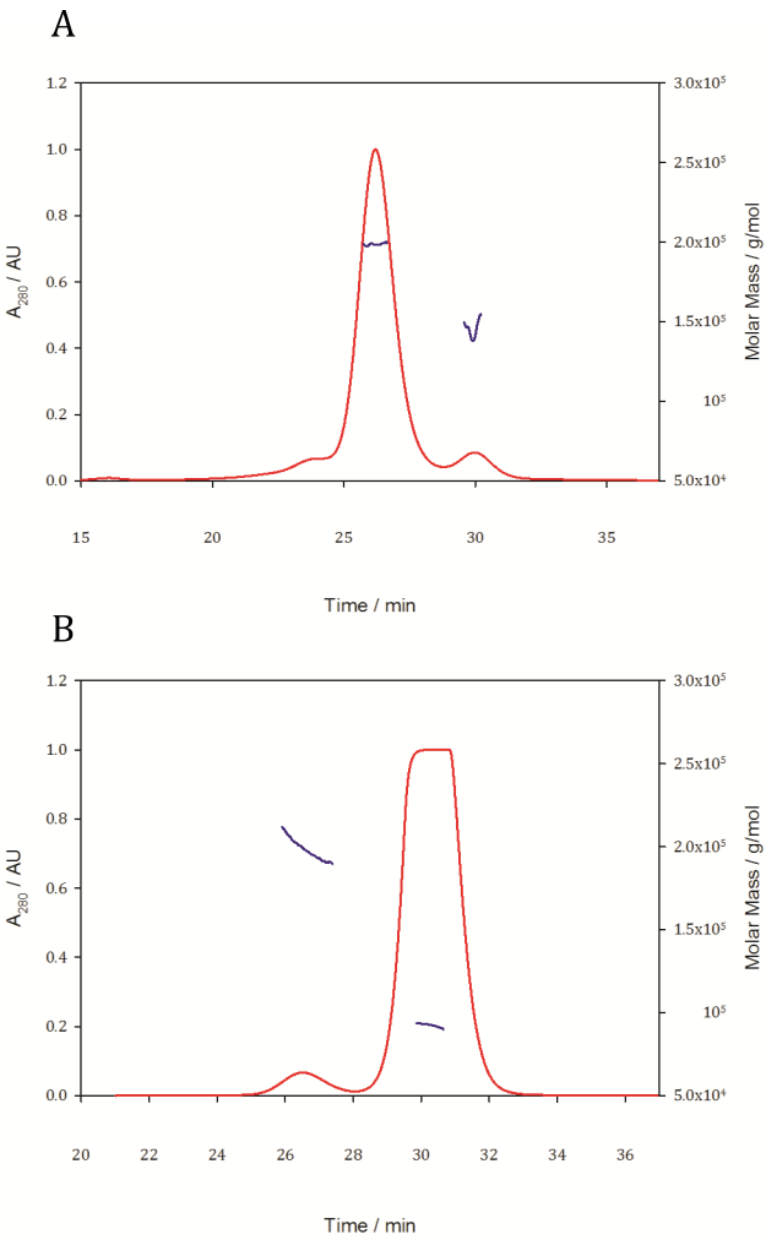


Figure 3.7. Results from SECMALLS of 125 nM TxGH116N and TxGH116C. UV absorbance is shown in red, molar mass is shown in blue. A) TxGH116N – There is a distinct peak at the equivalent of 198.7 kDa, and minor peak at 92.3 kDa. TxGH116N has a predicted mass of 92.7 kDa. B) TxGH116C – there are two distinct peaks, a minor peak at approx. 198 kDa and a major peak at 92.5 kDa.

3.4.4 Activity based probes

The use of activity based probe KY375 (**Figure 3.2)** confirmed the presence of active enzymes in the cell lysate of all three test expression cultures. Successful covalent tagging with KY375 shows that both *AP*GH116 and *Bt*GH116 are expressed in active forms by BL21 (DE3) *E. coli* and that loss protein fold and stability is occurring during purification stages.

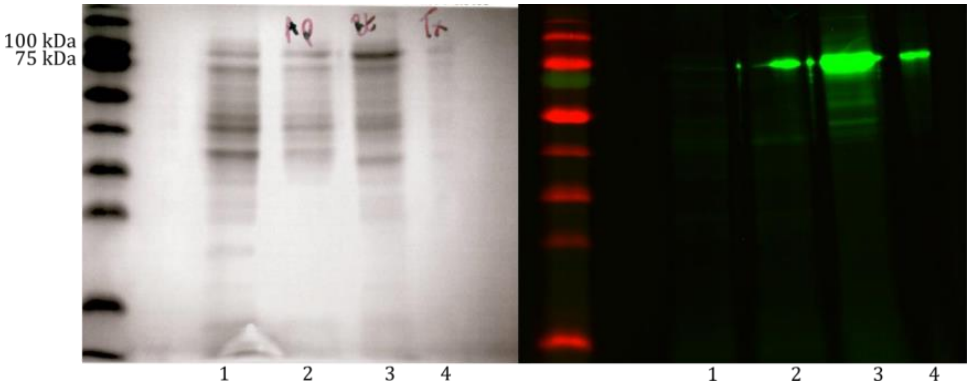


Figure 3.8 Tagging of active retaining β-glucosidases in the whole cell lysate. Only the GH116 enzyme, with the expected molecular weights of 90-91k Da, has been tagged by the fluorescent probe. A) SDS-PAGE gel showing the total cell lysate of *E.*coli transformed with a plasmid containing 1) BtGH116, not exposed to IPTG 2) ApGH116 4 hours post induction with 1 mg L^{-1} IPTG 3) BtGH116 4 hours post induction with 1 mg L^{-1} IPTG 4) TxGH116 4 hours post induction with 1 mg L^{-1} IPTG. B) The same gel, showing the localisation of the BODIPY fluorescent tag to the GH116 enzymes

The KY375 ABP is a large molecule which the enzyme may have struggled to accommodate, but the successful tagging with the KY375 probe (**Figure 3.8**) also severed to confirm TxGH116 as a suitable human protein analogue to pursue further structural studies. Binding of KY375 to the TxGH116 protein is discussed further in **Chapter 4**.

3.4.5 Crystallisation and data Collection

TxGH116N was successfully crystallised in the conditions described above (**Figure 3.9.A,B**). Crystals of TxGH116N were suitable for further analysis, data collection and structure solution (**Figure 3.8.C**). Data reduction and subsequent molecular replacement showed that TxGH116N crystallised in space group $P6_1$. Cell dimensions were measured at a = b = 187.8 Å, c = 99.3 Å; α, β, γ = 90°, 90°, 120°.

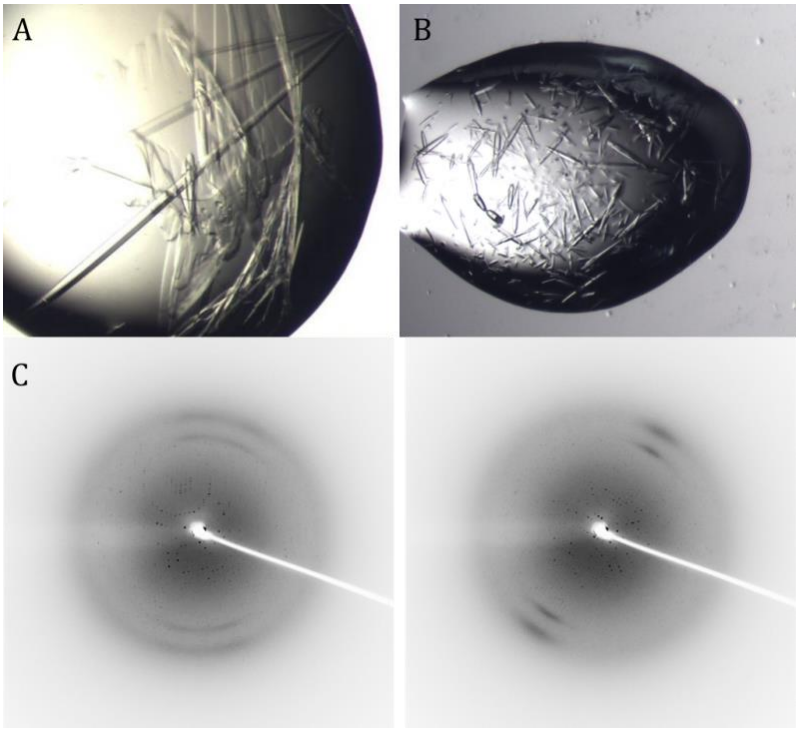


Figure 3.9. Crystals of TxGH116. A) Initial crystals from multi 96 well plate screening. B) TxGH116 crystals following seeding with earlier crystals to speed growth. C) Initial diffraction images, albeit rather weak, taken in-house of the TxGH116N crystal that was used for data collection. Images were taken at 0-0.5 ° and 90-90.5 ° respectively, with a 5 second exposure.

3.5 Summary

The human enzyme, non-lysosomal β-glucosidase GBA2, plays a role in the pathology of human diseases, including hereditary spastic paraplegia, and specific inhibitors of GBA2 may be useful in the treatment of Gaucher's disease. At the start of this thesis, little was known about the structure of proteins in CAZy family GH116, the family of GBA2. In this chapter is described the expression of three bacterial GH116 enzymes, and the purification of TxGH116, a bacterial homolog of the human GBA2 enzyme, from *E.coli* cells.

Use of a cyclophellitol aziridine activity based probe, designated KY375, to label active bacterial GH116 enzymes in solution helps to show that these bacterial enzymes are suitable to replicate the human counterpart. These bacterial enzymes are obviously able to accommodate the large chemical probes that are being studied, like their human equivalent.

Following initial expression tests and screening of purification buffers, scaling up of the initial protein production has provided an efficient protocol for the production of quantities of TxGH116N protein. The protein was purified to a standard suitable for crystallization. Successful crystallisation of TxGH116N lead to the solution of the three-dimensional structure of the protein, as described fully in Chapter 4.

Chapter 4. 3D structures of bacterial glucocerebrosidase TxGH116 in complex

4.1 Abstract

The previous chapter described the over expression of several candidate bacterial GBA2 homologs with the successful crystallization of an enzyme from *Thermoanaerobacterium xylanolyticum,* TxGH116. Here I describe the structure solution of TxGH116. During the course of our work, a Thai group also over-expressed and solved the 3D structure using a different construct, which we term TxGH116C, and we pooled our results for publication*. The TxGH116C construct provided better quality crystals and so was used by me for subsequent complexes – notably with cyclophellitol aziridine derived activity based probes and the glucosidase inhibitor castanospermine at resolutions up to 1.7 Å. The three dimensional structure of the bacterial GH116, TxGH116, gives key insights into the effects that pathogenic human GBA2 mutations may have on the protein. Knowledge of the structure of the bacterial protein, both alone and bound to a range of inhibitors and ligands, can feed into the design of further probes that aim for specificity for GBA2 over GBA1 (or vice versa).

*The work described in this chapter has been published in Charoenwattanasatien et al., (2016) Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). *ACS Chem Biol* **11**, 1891-1900

Aziridine complex: Lahev et al., (2017) A Fluorescence Polarization Activity-Based Protein Profiling Assay in the Discovery of Potent, Selective Inhibitors for Human Non-lysosomal Glucosylceramidase. *J Am Chem Soc*, in press

4.2 Introduction

The lysosomal acid-β-glucosidase that forms the basis of Gaucher's disease, GBA1 of CAZy family GH30, was previous described in **Chapter 2**. The first X-ray crystallographic structure of human *GBA1* was solved in 2003 by Dvir *et al* (38) and since then multiple structures have been published. As described in the introduction and **Chapter 3**, human cells contain a second glucosylcerebrosidase, GBA2 (non-lysosomal glucosylcerebrosidase, E.C. 3.2.1.45)

GBA2 was first identified in humans as a "bile acid β-glucosidase", distinct from earlier identified β-glucosidases, in 1992 by Matern et al (176). A year later, "another" β-glucosidase was recognised as present in many other human tissues as a "conduritol B epoxide (CBE) insensitive β-glucosidase", that was active on glucosylceramide when GBA1 was inhibited by CBE(191). It wasn't until 2007, when Boot et al cloned and expressed the full cDNA of GBA2, that these two enzymes were recognised as the same protein species (175).

Whilst the importance of GBA1 and its role in disease has been widely recognised, the physiological activity of GBA2 has only comparatively recently been documented, notably in male mouse fertility (192), whilst a deficiency of GBA2 has also been linked with the accumulation of glycosylceramide in the spleen and impaired liver regeneration in mice (136,193). In humans, point mutations in GBA2 lead to neurological disorders such as hereditary spastic paraplegia (131).

Due to this role in both Gaucher's disease progression and its involvement in other pathologies, GBA2 has become a viable therapeutic target for disease treatment. To aid the design of new GBA2 probes, three- dimensional structural knowledge of the protein is ideal. GBA2 is a large, membrane-associated protein, with a molecular weight of 104 kDa, unwieldy to work with in a crystallographic context (194) . Bacterial homologs are usually more simply expressed in standard *E. coli* strains, with the advantage of larger protein yields and fewer post-translational modifications such as glycosylation.

4.2.1 Characterisation of GH116 family enzymes

GBA2 is a member of CAZY glycoside hydrolase family GH116; a family first described in the 2009 paper by Cobucci-Ponzano et al (129) with the biochemical analysis of the archaeal β-glucosidase from *Sulfolobus solfataricus* (SSO1353). Since then, CAZy family GH116 has been populated with 377 identified proteins, from all three domains of life. However, as of writing, only 5 have been experimentally analysed and for only one (TxGH116, this chapter) has a 3D structure been published.

Following their work on founding the GH116 family with the functional classification of SSO1353, the Moracci group has further defined three GH116 subfamilies based upon GH116 enzyme substrate specificity and inhibitor sensitivity (**Table 4.1**) (185).

TABLE 4.1 GH116 SUBFAMILIES

GH116 Subfamily	**Subfamily 1 (175)**	**Subfamily 2 (185)**	**Subfamily 3 (129)**
Example	GBA2	SSO3039	SSO1353
Mechanism	Retaining	Retaining	Retaining
Specificity	β-glucosides, glucosylceramides	β-glucosides, N-acetyl-glucosaminides	β-glucosides, β-xylosides
CBE sensitivity?	No	mM	mM
NB-DNJ sensitivity?	nM	μM	mM

In **Chapter 3**, I described the purification of and the crystallisation of a bacterial GBA2 homologue, TxGH116, a β-glucosidase from *Thermoanaerobacterium xylanolyticum*. TxGH116 is both a β-glucosidase and a β-glucosylceramidase, belonging to GH116 subfamily 1 (190); a family which acts with net retention of anomeric configuration (129).

Retaining β-glucosidases, such as GBA2, utilise a catalytic acid/base residue and nucleophile residue to hydrolyse a glycosidic substrate through classic Koshland retaining mechanism. This is a two-step double displacement with the creation of a glycosyl enzyme intermediate (**Figure 3.2**)(144). Firstly the nucleophile residue must attack the anomeric carbon, whilst the acid/base residue is acts as a general acid catalyst, protonating the oxygen of the glycosidic bond to allow leaving group departure and allowing the formation of a covalent glycosyl-enzyme intermediate. In the second step, a water molecule acts as a nucleophile, activated by the acid/base residue which is now acting as a base catalyst, to complete the hydrolysis and deglycoslyate the enzyme (195).

The first archaeal GH116 β-glycosidase, SSO1353, was indeed shown to have a retaining mechanism (129). The identities of the catalytic nucleophile and acid/base of SSO1353 were identified by computational sequence alignment with other GH116 family enzymes and then site directed mutagenesis of the likely target residues.

Asp462 was confirmed as the catalytic acid/base of the enzyme by rescuing of the enzyme activity with sodium azide (which is able to act as an external nucleophile that does not require enzymatic deprotonation for activity). The catalytic nucleophile was identified as through the use of the activity-based inhibitor 2,4-dinithrophenyl-β-D-2-deoxy-2-fluoro-glucopyranoside (2,4,DNp-2F-Glc)(129), which is able to readily undergo the glycosylation step of the enzyme reaction, forming the intermediate, but does not so readily undergo the deglycosylation step, due to the presence of a fluorine substituent at the C2 position. This has the effect of destabilizing the transition states, and so the incorporation of a good leaving group is necessary (in this case 2,4-dinitrophenol) to promote the glycosylation step (141).

As described in Chapters 1 and 2, disruption of this reaction after the first stage can be done by exposing the enzyme to an activity-based probe (ABP) that contains an aziridine analogue of the cyclophellitol "warhead". This process results in an enzyme that is covalently bound to the ligand at the catalytic nucleophile residue (118,141) (**Figure 4.1**). Utilising an ABP with a warhead linked to a probe that is, for example, fluorescent can allow for real time labelling of enzymes in cells with minimal short-term disruption to the function of other biological process, and, importantly, only the labelling of the active, correctly folded enzyme. Probes with a reporting moiety that have fluorescent capacity can be used multiple ways, from *in vivo* protein localisation to fluorescence polarization assays and the analysis of the potency of small molecule inhibitors.

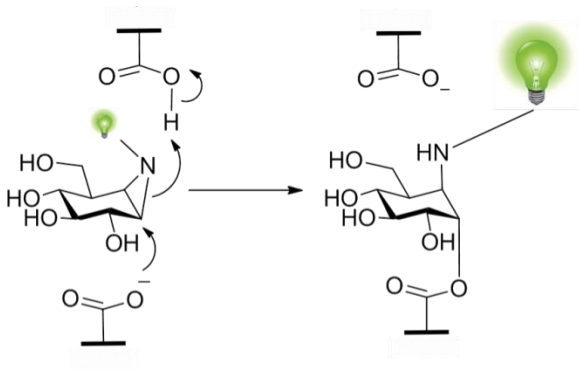


Figure 4.1. Capturing the nucleophile in the active site of the TxGH116 enzyme by disruption of the hydrolysis after the first stage with a cyclophellitol aziridine fluorescent activity based probe (a fluorescent moiety is represented by the green light bulb).

In the human GBA2, the identity of the respective acid/base and nucleophile as Asp 677 and Glu 527 respectively was confirmed by Kallmeijin et al in 2014, using these highly sensitive cyclophellitol-based activity probes, specifically the β-aziridine grafted to a BODIPY (boron-dipyrromethene) fluorescent moiety here described as KY375 (**Figure 4.2**) (141).

4.2.2 Aims of the research in this chapter

The goals of the work described in this chapter were to solve the 3D structure of the bacterial GH116 protein TxGH116 from the bacteria *Thermoanaerobacterium xylanolyticum,* both non-liganded and in complex with different ABPs and known inhibitors. It is hoped that these studies would give further insight into the structure of the human GH116 enzyme GBA2 and allow further development of selective probes more able to differentiate between GBA1 and GBA2.

Five structures were studied, including complexes with the four ligands as shown in **Figure 4.2.** Castanospermine is an indolizidine alkaloid that was first isolated from the seeds of the Moreton Bay Chestnut, *Castanospermum austral,* in 1981 (196) and has inhibitor activity against a variety of glucosidase enzymes including other retaining β-glucosidases such as GH3 enzymes (197) . KY358, previously described in complex with GBA1 in **Chapter 2,** is an acyl-aziridine. KY375, previously described in **Chapter 3**, is a cyclophellitol aziridine functionalised with a boron-dipyrromethene (BODIPY) group.

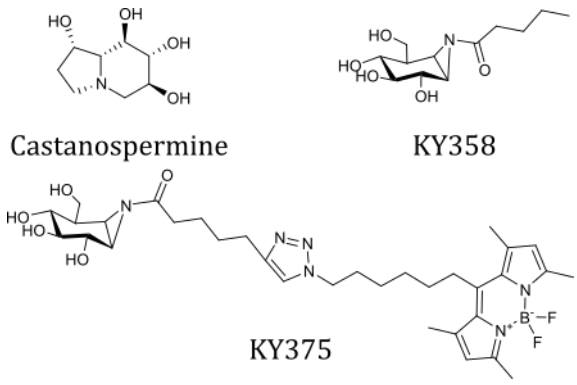


Figure 4.2 Activity-based TxGH116 probes and inhibitors studied at a structural level in this thesis

4.3 Materials and methods

TxGH116N was purified and crystallised as described in Chapter 3.

4.3.1 Data collection of TxGH116N crystals

Initially crystals of TxGH116N were individually transferred to cryoprotectant solution of the mother liquor combined with 20 % glycerol (v/v) and immediately placed in liquid N_2 at 100K. However, after seeing some issues with crystals cracking on exposure to cryoprotectant, several crystals were frozen without cryoprotectant.

Diffraction quality of all *Tx*GH116N crystals was initially tested in house using a RUH3R rotating anode X-ray generator. Diffraction images were sampled at angels of 0 and 90° using 0.5° oscillations. Best diffracting crystals, crystals were then sent to the Diamond Light Source in Oxfordshire for data collection.

4.3.2 Data processing and structure determination

Data for each crystal set was processed using xia2(160,161) and the AIMLESS data reduction pipeline through the CCP4i2 software(162).

After collection of the apo TxGH116N native data, we were contacted by the group of Jim Cairns (Suranaree University of Technology, Thailand) who were working on the same enzyme. Ultimately, they obtained a 3D chain trace, and they kindly provided that to us for structure solution.

The Thai TxGH116C structure (5BVU) (163) was used to solve the structure of the apo-TxGH116N by molecular replacement using MOLREP(164). Refinement of the structure was performed using REFMAC (166) and model building completed using COOT(167), both programs through the CCP4i2 software.

As the "Thai" TxGH116C construct diffracted better, we chose to use that for structures solution of subsequent ligand complexes. Their clone was kindly supplied by James R. Ketudat Cairns, Suranaree University of Technology, Institute of Science, Protein Structure, Function and Application, Thailand. This construct provided a protein that crystallised more readily and at a much lower concentration whilst giving better quality crystal data.

4.3.3 Overexpression and purification of TxGH116C

A pET-30a(+) plasmid containing C-terminal His-tagged TxGH116 gene was supplied by James Cairns. This was a construct consisted of the TxGH116 protein, lacking the first 18 amino acids, with the pET30a(+) N-terminal His_6-tag, S-tag and enterokinase site. At the C-terminal end of the protein was a second His_6-tag from the plasmid.

This construct was designated TxGH116C. A protocol for the purification of TxGH116C was adapted from Charoenwattansatien et al (2016)(190).

Competent *E. coli* BL21 (DE3) cells were transformed with the supplied plasmid and grown in LB media (30 um/ml kanamycin resistance) at 37C until an OD600 of 0.6 had been reached. Expression of the TxGH116C gene induced with 0.5 mM IPTG and the cells incubated at 16 °C for 18 hr. Cells were lysed by sonication at 14-18 Hz, in buffer A (20 mM NaPO4 pH 7.4, 150 mM NaCl). Cell debris was removed by centrifugation.

The resulting supernatant was exposed to 65 ◦C for 20 minute and the resulting precipitated protein removed by centrifugation. The remaining soluble fraction was loaded onto a 5 ml GE sciences crude HisTrap column, equilibrated with buffer A. Unbound material was removed by washing with three CV of buffer A. Bound protein was eluted with a gradient into buffer B (20 mM NaPO4 pH 7.4, 150 mM NaCl, 500 mM imidazole) across 20 CV. The fractions containing TxGH116C were pooled, concentrated and buffer exchanged into Digest buffer (20 mM Tris pH 8, 150 mM NaCl)

The protein was digested with enterokinase (New England Bio Labs) as per the manufacturer's instructions, to remove the N-terminal His_6-tag and S-tag, then buffer exchanged into GF buffer. The protein was then purified over a Superdex S200 gel filtration column, which had been equilibrated with GF buffer (20 mM Tris pH 8, 500 mM NaCl) and the fractions containing TxGH116C pooled, concentrated to 20 mg/ml and stored at 5 °C.

4.3.4 Crystallisation of TxGH116C and ligand soaking

TxGH116C was diluted to ~2 mg/ml and crystallization trials were set up in 48-well plate format sitting drop plates at 18 °C, in conditions adapted from Sansenya et al, 2015 (189) . Before setting up crystallisation experiments TxGH116C centrifuged for *ca.* 5 min.

Unless otherwise noted, crystallisation drops were set up as 1:1 ratio of protein to mother liquor and TxGH116C was added at 2 mg/ml.

- TxGH116C-KY358
  0.2 M ammonium sulfate, 20% (w/v) PEG 3350, 0.1M BisTris pH 6.7
- TxGH116C-ME594
  0.2 M ammonium sulfate, 20% (w/v) PEG 3350, 0.1M BisTris pH 5.5
- TxGH116C-KY375
  0.2 M ammonium sulfate, 15% (w/v) PEG 3350, 0.1M BisTris pH 6.6

- TxGH116C-Castanospermine
  0.2 M ammonium sulfate, 20% (w/v) PEG 3350, 0.1M BisTris pH 6.0
  2:1 protein to mother liquor

Crystals of TxGH116C were soaked in KY358 ligand residue resuspended in the corresponding mother liquor (0.2 M ammonium sulfate, 20% (w/v) PEG 3350, 0.1M BisTris pH 6.7) overnight to allow covalent binding. Crystals of TxGH116C bound with various other ligands were obtained similarly, by adding 1 µl of the compound dissolved in either water or mother liquor to the drop at a concentration of 2 – 5 mM and then resealing. KY353 ligand was added at an unknown concentration due to the nature of the ligand supplied. After incubation the crystals were removed and cooled using liquid N_2.

4.3.5 Data collection of TxGH116C crystals

For initial collection of X-ray data, a crystal of TxGH116C was transferred to a cryo-protectant solution containing mother liquor supplemented with either 20-25% (v/v) glycerol or ethylene glycol and then cooled in liquid N_2. After comparison in-house using a RUH3R rotating anode X-ray generator and selection of the best diffracting crystals were then sent to the Diamond Light Source in Oxfordshire for data collection.

4.3.6 Data processing and structure determination

Data for each crystal set was processed using xia2(160,161) and the AIMLESS data reduction pipeline through the CCP4i2 software(162).

The apoTxGH116C PDB structure (5BVU) from Charoenwattanastien et al (2016) (190) was used to determine the phases and solve the structure. Coordinates from the apo-TxGH116C were used by directly refining against the ligand complex data sets for determination of the ligand complex structures.

Idealised coordinate sets and refinement dictionaries for each ligand were generated using ChemDraw3D and JLIGAND(165) in the CCP4 suite. Refinement of all structures was performed using Refmac (166) and model building completed using Coot(167), both programs through the CCP4i2 software. Final models for each of the compounds in this chapter were validated using the wwPDB Validation service (validate-rcsb-1.wwpdb.org/)

4.3.7 Isothermal titration calorimetry (ITC)

ITC was performed with TxGH116C against castanospermine to get K_D. The protein was used at a cell concentration of100 μM in 50 mM MES buffer (pH 5.5). Titrations were performed with a stock of 2 mM castanospermine over the course of 19 0.5 μL injections.

The experiment was performed using a Micro-Cal Auto-ITC200 machine at 25 ° C and data analysed with the MicroCal PEAQ-ITC analysis software and Excel. Based on crystal structure data, stoichiometry was fixed to 1, giving a revised syringe concentration of 2.76 mM.

4.3.8 Homology Modelling

A three dimensional homology model of the human GBA2 enzyme was produced using the Phyre[2] online server (www.sbg.bio.ic.ac.uk/phyre2)(198) and the coordinates from the known and published TxGH116 structures.

4.4 Results and discussion

The purification and successful crystallisation of TxGH116N was previously described in **Chapter 3**. The structure of TxGH116N was successfully solved using molecular replacement following the provision of the alternative TxGH116C structure (PDB ID: 5BVU)(190).

The second construct, TxGH116C, was also successfully purified, and crystallised in a form suitable for further structural study.

4.4.1 Overexpression and Purification of TxGH116C

As with the TxGH116N construct, nickel chromatography was used to obtain the His_6-tagged protein from the cell lysate following sonication (**Figure 4.3**). The protein was then digested with enterokinase to remove the extensive N-terminal His6-tag and S-tag and then subsequently heat treated at 65 ◦C. After removal of the precipitated, non-thermally stable protein, the remaining protein was further purified by gel filtration (**Figure 4.4**) to give a final yield of ~ 11 mg /L of cell culture.

This is a much higher yield per litre of culture than the TxGH116N construct gave, suggesting that there was some benefits to the extended TxGH116C construct at the protein production stage.

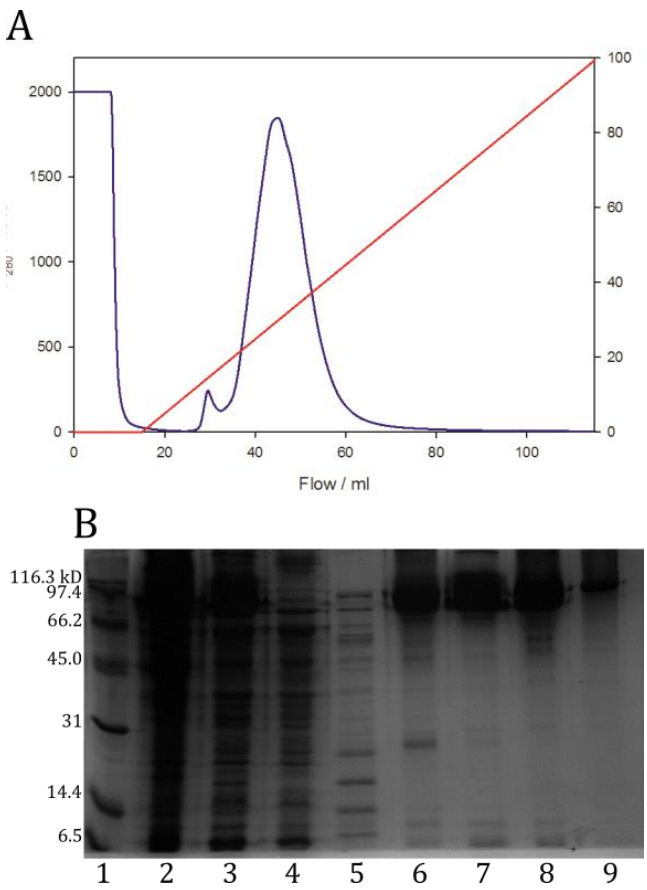


Figure 4.3. Purification of TxGH116C. A) Chromatogram showing the elution of the TxGH116C-His_6 protein from a HisTrap nickel column with an increasing concentration of buffer containing imidazole. Protein is measured by the UV absorbance at 280 nm (A_{280}). B) An SDS-Page gel of fractions taken from the peak between ~30-60 ml (lanes 6 -9). Lane 1 is a broad range molecular weight ladder (Bio-Rad), lane 2 is a sample of the total cell lysate, lane 3 is a sample of the total cell lysate after heat treatment, lane 4 is a sample of the total protein that was loaded on to the column and lane 5 is a sample of the flow through (1-10ml). There is some overloading in the gel.

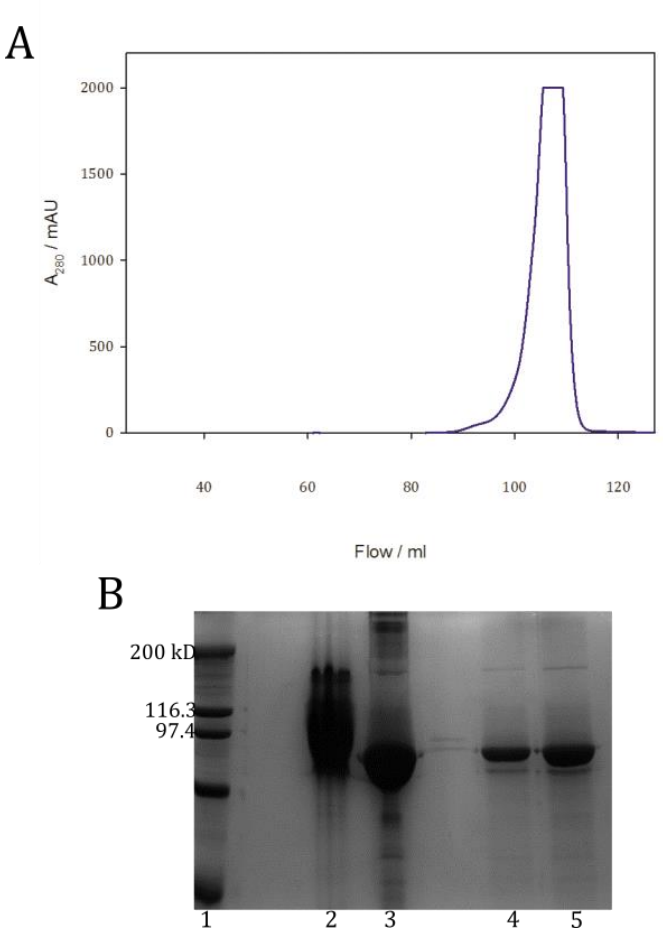


Figure 4.4. Purification of TxGH116C. A) Chromatogram showing the elution of the TxGH116 C protein from an S200 size exclusion gel filtration column. Protein is measured by the UV absorbance at 280 nm (A_{280}). B) An SDS-Page gel of fractions taken from the two peaks between ~95-115 ml (lanes 4-5). Lane 1 is a broad range molecular weight ladder (Bio-Rad). Lane 2 is the total protein from the HisTrap purification and lane 3 is the total protein after digest with enterokinase.

4.4.2 Unliganded TxGH116 structure

Diffraction images of TxGH116N were collected at the I24 beamline at Diamond. Data reduction and subsequent molecular replacement showed that TxGH116N crystallised in space group $P6_1$. The X-ray data was scaled at a resolution of 2.6 Å cut off based on criteria for CC (1/2) of >0.5 in the outer resolution shell. There were two molecules in the asymmetric unit. The full data table for the TxGH116N data set collected at Diamond Light Source in Oxfordshire is shown overleaf (**Table 4.2**).

TABLE 4.2 DATA COLLECTION AND REFINEMENT STATISTICS FOR THE TXGH116N STRUCTURE DESCRIBED IN THIS CHAPTER

	TxGH116N
PDB ID	5FJS
Space group	$P6_1$
Cell dimensions	
a, *b*, *c* (Å)	187.8, 187.8, 99.3
α, β, γ(°)	90,90,120
Resolution (Å)	163-2.60(2.67-2.60)*
R_{merge}	3.7(79.4)
R_{pim}	0.032(0.27)
CC (1/2)	0.997(0.807)
I /σ(I)	51.2(3.0)
Completeness (%)	100(99.9)
Redundancy	9.4(10.5)
Refinement	
Resolution (Å)	162-2.6
No. reflections	61406/3082
R_{work} / R_{free}	0.19/0.25
***B*-factors (Å²)**	
Protein	68
Ligand/ion	68
Water	55
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.7

*Values in parentheses are for highest-resolution shell.

The TxGH116N structure was solved using molecular replacement following the release of the alternate TxGH116C structure (PDB ID: 5BVU) using a SeMet-derivative protein by Charoenwattanasatein et al (190). Our TxGH116N structure can be traced from residues 35 – 803 (of 807) in molecule A in from residues 34 – 803 in molecule B. There is unmodelled density between residues 114-125, 200-206, 234-240 for both molecules A and B, between residues 299-302 in chain A and between residues 132-135, 166-169 in chain B. This is the model of TxGH116N has been with the PDB ID: 5FJS on 2015-10-12.

The structure of TxGH116N is comprised of an N-terminal domain consisting of β-sandwiches and a C-terminal α-helix (solenoid) domain **Figure 4.5.** The catalytic residues are located in the C-terminal domain. The C-terminal domain also contains the binding site for a Ca^{2+} ion. Further description of the active centre similarity, and basis for genetic disease mutations, will be given after description of a different crystal form with higher resolution data.

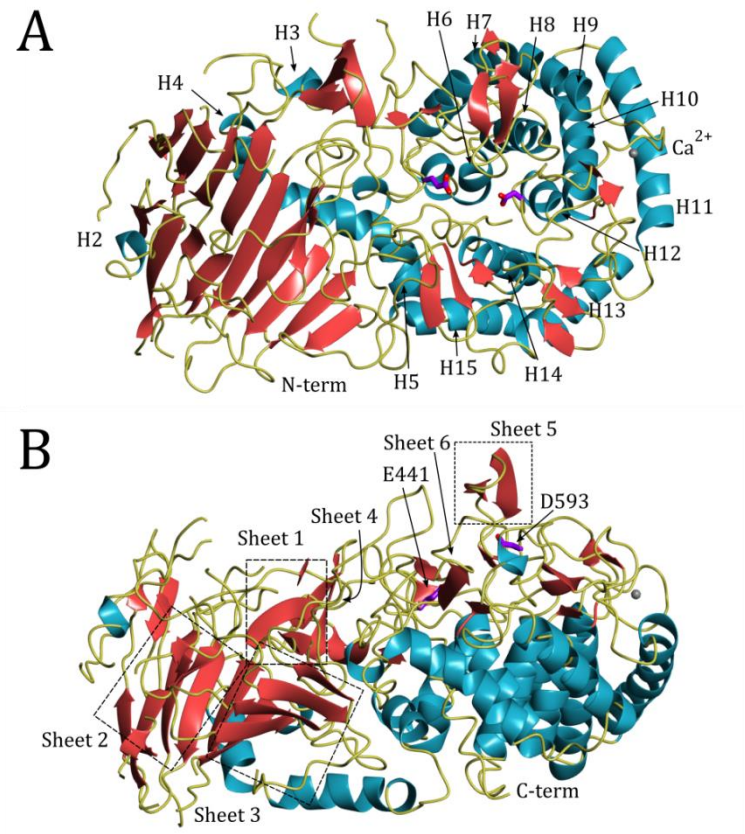


Figure 4.5. The 3D-fold structure of TxGH116N. A) The C-terminal domain, containing the active site residues shown in purple, consists of 11 α-helices labelled H5-H15. These are arranged into six outer helices, surrounding the inner helix pairs, with H5 sitting between the two domains. B). The N-terminal domain is comprised of β-sheets (sheets 1-3) surrounded by α-helices (H1-4)

4.4.3 Crystallisation of TxGH116C

Following successful purification, TxGH116C was crystallised in the conditions as described above, **Figure 4.6.** Data reduction and subsequent refinement showed that TxGH116C crystallised as $P2_1\ 2_1\ 2$ or $P2_12_12_1$, allowing me to refine three liganded TxGH116C structures at 1.70-2.19 Å.

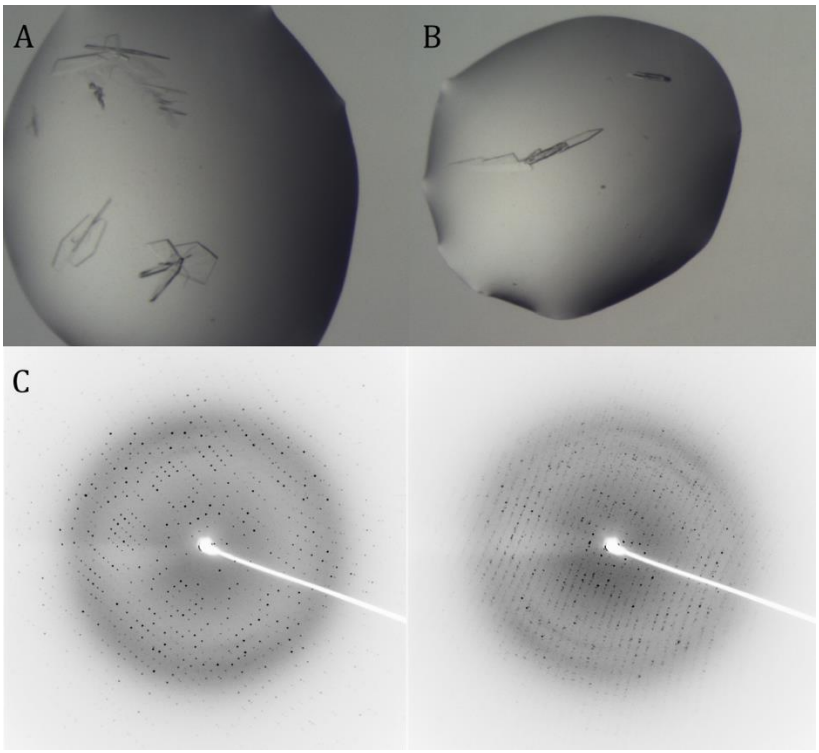


Figure 4.6. TxGH116C crystals. A and B) Crystal forms of TxGH116C. C) Diffraction patterns collected in –house for TxGH116C-KY358 crystal complexes. Images were taken at 0-0.5 ° and 90-90.5 ° respectively, with a 5 second exposure.

Data tables for all the data sets of TxGH116C in this chapter collected at the Diamond Light Source in Oxfordshire are shown on the next page (**Table 4.3**).

TABLE 4.3 DATA COLLECTION AND REFINEMENT STATISTICS FOR TXGH116C STRUCTURES DESCRIBED IN THIS CHAPTER

	TxGH116_ Castanospermine	**TxGH116_KY375**	**TxGH116_KY358**
PDB ID			5NCX
Space group	$P\,2_1\,2_1\,2$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2\,2_1$
Cell dimensions			
a, *b*, *c* (Å)	180.8, 53.9, 84.7	54.0, 164.0, 179.0	53.6, 83.3. 177.5
α, β, γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	90.38-2.19(2.27-2.19)*	49.31-2.16(2.20-2.16)	60.73-1.70(1.73-1.70)
R_{merge}	0.290(3.03)	0.126(1.05)	0.096(1.00)
$\boldsymbol{R_{pim}}$	0.157(1.77)	0.053(0.49)	0.060(0.62)
CC (1/2)	0.968(0.623)	0.997(0.651)	0.998(0.584)
$I/\sigma(I)$	5.1(0.5)	10.4(1.5)	12.4(1.7)
Completeness (%)	96.0(99.7)	99.9(99.8)	99.5(99.2)
Redundancy	5.6(4.9)	6.5(5.6)	5.7(5.6)
Refinement			
Resolution (Å)	90.38-2.19	49.31-2.16	60.73-1.70
No. reflections	47003/2414	86160/4277	87654/4371
R_{work} / R_{free}	0.26/0.30	0.18/0.23	0.15/0.18
***B*-factors (Å²)**			
Protein	28.5	36.1	20.2
Ligand/ion	23.3	132.6	42.6
Water	22.0	29.2	27.7
R.m.s. deviations			
Bond lengths (Å)	0.016	0.015	0.018
Bond angles (°)	2.03	2.21	1.87

*Values in parentheses are for highest-resolution shell.

4.4.4 Active Centre of GH116

The catalytic acid/base and nucleophile residues were previously proposed for the archaeal *Sulfolobus solfataricus SSO1353* β-glucosidase (129) and further insight into the active site and the positioning of the catalytic residues was gathered by collection of data for TxGH116C in complex with trapped inhibitors and probes.

Complexes of TxGH116C with 2-deoxy-2-fluoroglucosyl (DNP2FG), deoxynojirimycin (DNJ), glucoimidazole (GIM) and β-D-glucose were collected by Charoenwattansatien and co-workers and released with the respective PDB IDs: 5BX2-5 (**Figure 4.7.A-D**). The complex with a trapped intermediate DNP2FG, confirmed the catalytic nucleophile as E441, consistent with the equivalent E335 residue previously labelled by DNP2FG in *Sulfolobus solfataricus* β-glucosidase (129)

The catalytic nucleophile, E441, which is equivalent to the human GBA2 E527, and the acid/base D593 (equivalent to D677 in the human GBA2) are located 10.8 Å apart in the TxGH116N $P6_1$ structure, although this is reduced to 8.5 Å in the $P2_12_12$ TxGH116C (PDB ID: 5BVU) structure.

This is a longer distance than can be observed in most other retaining β-glucosidases, where the distance between acid/base and nucleophile is typically 4.5-6 Å. Both subsequent ligand bound structures with TxGH116C and inactive mutations of those residues by Charoenwattanasatien et al (190) confirmed them as the catalytic residues.

The putative acid/base, D593 (equivalent to Asp462 identified by Moracci and colleagues based on azide rescue experiments (129)), is located in an unusual position above the pyranose ring, rather than in in the plane of the sugar ring which is the typical position for configuration-retaining β-glycosidases (199). This location of the acid/base residue prevents efficient protonation of GIM (**Figure 4.7.C**). This explains the lower inhibition of TxGH116C by GIM (K_i = 0.34 ± 0.03 μM) in comparison to DNJ (K_i = 0.13 ± 0.01 μM) (**Figure 4.8.B**) (190).

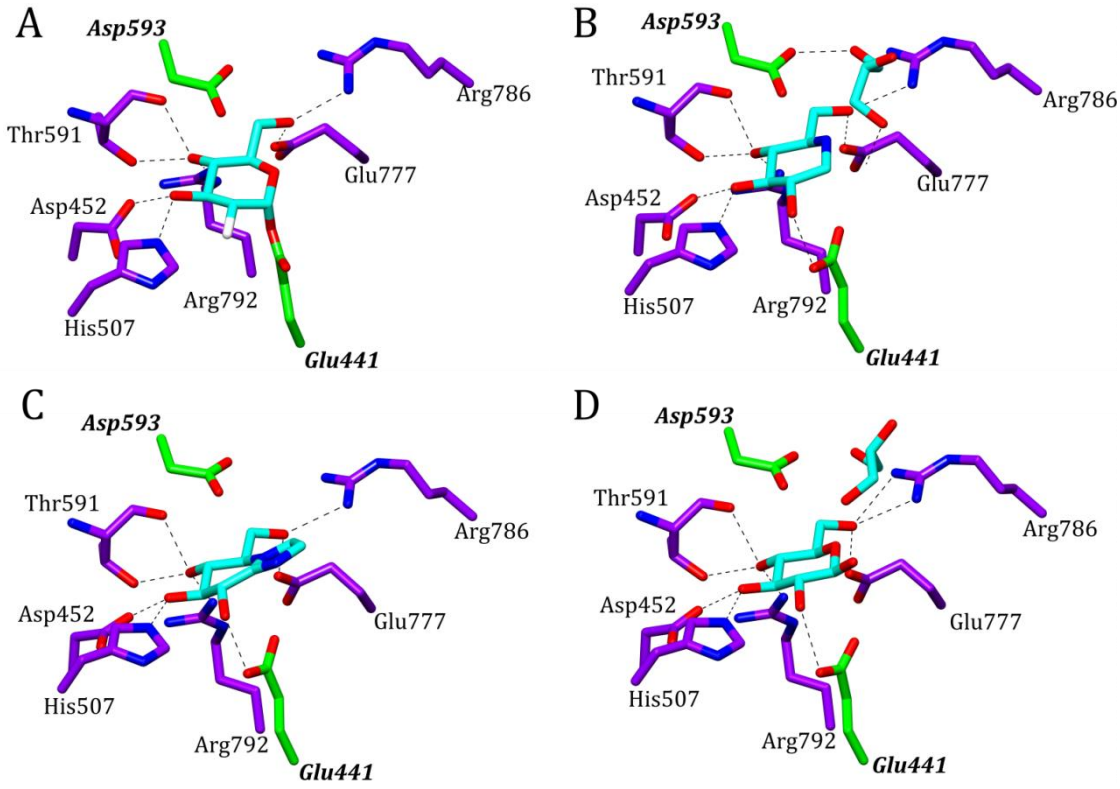


Figure 4.7. The active site of TxGH116C in complex, showing potential hydrogen bonds between ligands and the surrounding residues. Catalytic residues are shown in green and labelled in italics, adjacent residues in purple and the ligand in cyan. A)2-deoxy-2-fluoroglucosyl (PDB ID: 5BX2, data collected to 1.61 Å) B) Deoxynojirimycin (DNJ) (PDB ID: 5BX3, a collected to 1.96 Å) C) Glucoimidazole (PDB ID: 5BX4, a collected to 1.65Å) D)β-D-glucose (PDB ID: 5BX5, a collected to 1.85 Å) (190).

The structure of the enzyme bound to DNP2FG (**Figure 4.7.A**) enables identification of residues interacting with the substrate. The amino acid residues capable of forming hydrogen bonding interactions include H507 and D452 at OH3; D452, T591, and R792 at OH4; and E777 and R786 at the OH6. In the last case, OH6 appears to mediate a salt bridge between E777 and R786.

These active site glucosyl binding residues are conserved between human GBA2 and TxGH116 (**Figure 3.3**), and the potential impact of disease causing mutations in these residues on the human enzyme is discussed later in this chapter.

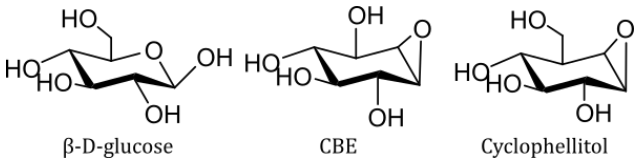


Figure 4.8. Presence of a C6-hydroxymethyl group affects the ability of a ligand of inhibit TxGH116. A) β-D-glucose B) Conduritol B epoxide C) Cyclophellitol

These interactions at OH6 may explain why cyclophellitol compounds which have a hydroxymethyl group at that position are able to inhibit TxGH116, but CBE, which lacks this group, does not (**Figure 4.8**). The binding of two cyclophellitol aziridine based compounds to TxGH116C is described further in later sections of this chapter.

4.4.5 3D complexes of TxGH116C with castanospermine

In parallel with the above work on TxGH116 complexes to map the active site, I also studied complexes with the inhibitor castanospermine. Castanospermine is a wide acting glucosidase inhibitor on a variety of enzymes from GH3 retaining β-glucosidases from A*spergillus aculeatus* (197) to GH31 α-glucosidases such as rat hepatic lysosomal α-glucosidases (200). Castanospermine has also been investigated as a potential inhibitor of breast cancer cell proliferation (201) and as an inhibitor of Dengue virus infection (202). Castanospermine is also an inhibitor of human lysosomal GBA1, (IC_{50} = 19 μM)(203). It was not unknown whether the ligand would bind to TxGH116C crystals.

Crystals that were soaked with castanospermine showed some surface disruption, perhaps due to an overwhelming excess of the ligand added to the drop at a final concentration of 5 mM, and the data collected extended to only 2.10 Å, however the density for the ligand is clear.

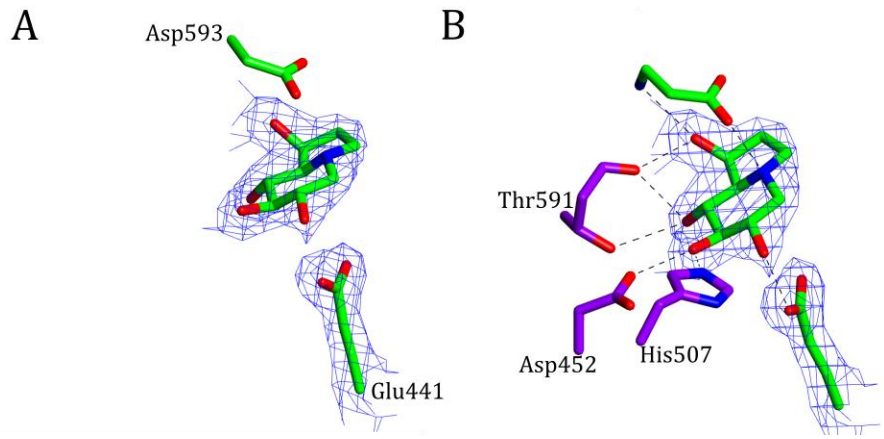


Figure 4.9. Castanospermine in the active site of TxGH116C. A) Observed electron density for castanospermine is contoured at 1.2 σ (equivalent to 0.37 electrons / Å^3). B) Side chains interacting with castanospermine, Active site residues are shown in green; other amino acids interacting with the ligand are shown in purple.

The isothermal titration calorimetry performed with TxGH116C and castanospermine gave an experimentally measured K_D of 103 μM (**Figure 4.10**), consistent with the protein binding at mM concentrations in the crystal soak.

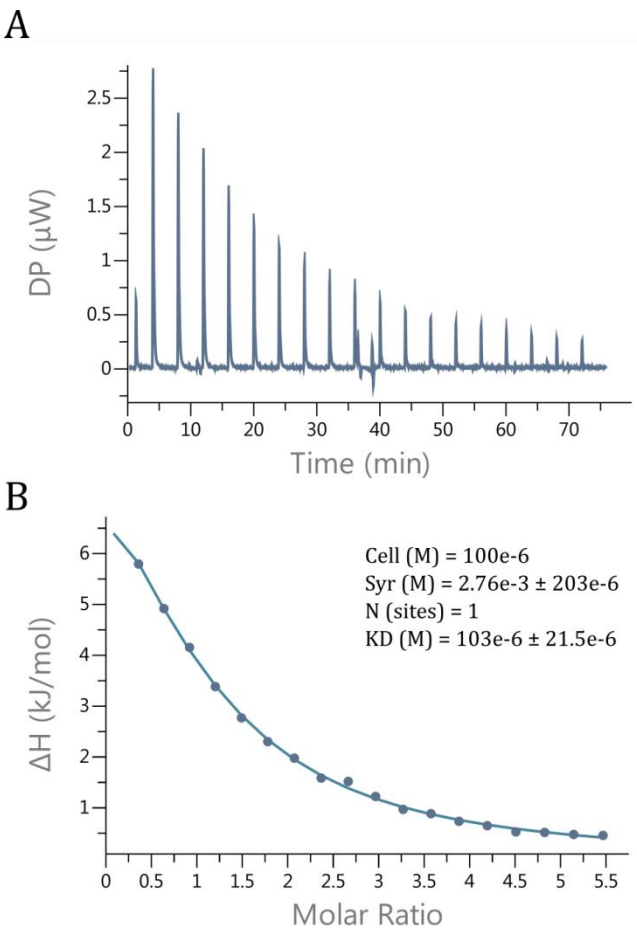


Figure 4.10. Isothermal titration calorimetry (ITC) was performed using TxGH116C and castanospermine. A) The raw titration data (baseline subtracted) B) Integrated data fitted to "One set of sites" model by the MicroCal PEAQ-ITC analysis software

4.4.6 3D complexes of TxGH116C with KY358

In **Chapter 2**, there was observed KY358, an "acyl-aziridine", bound covalently to human GBA1. As acyl aziridines make for potent activity based probes for GBA1 (17), cross-talk between GBA2 and GBA1 can disrupt labelling or even drug treatment.

Crystals of TxGH116C soaked with KY358 diffracted well and did not appear to be disrupted by the ligand. The data for the TxGH116C/KY358 complex could be refined at 1.7 Å, with clear electron density for the covalently bound ring (**Figure 4.11.A**). However, although there is limited density until N-acyl group, there is no density around the extended carbon chain, and it appears disordered. This is similar to the appearance of KY358 as seen bound to GBA1 in **Chapter 2, Figure 2.10**.

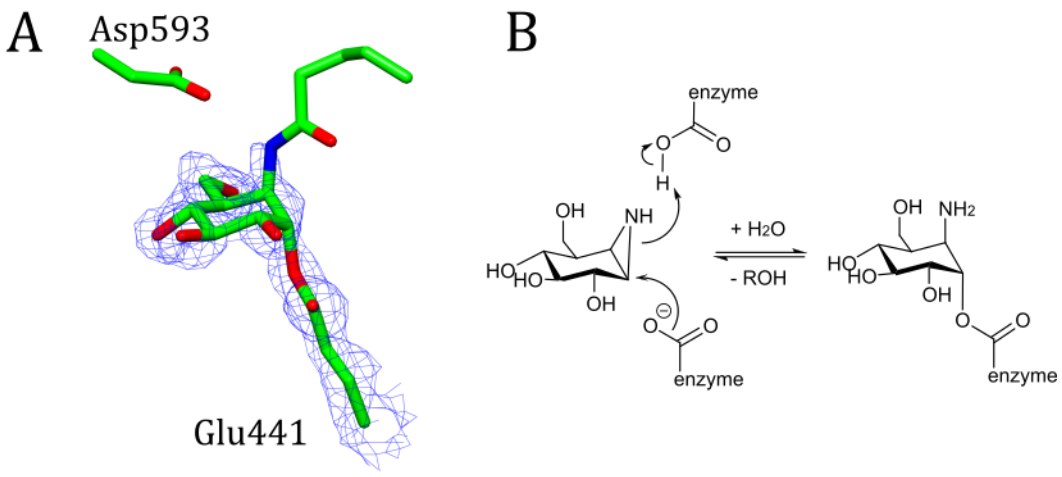


Figure 4.11. KY358 bound to the active site of TxGH116C. A) Observed electron density shown at 1.2 σ (equivalent to 0.46 electrons / Å^3). B) Reaction mechanism for an aziridine probe to covalently bind to a retaining β-glucosidase enzyme.

4.4.7 3D complexes of TxGH116C with KY375

As previously discussed in **Chapter 3,** the fluorescent activity based probe KY375 was confirmed to effectively bind to TxGH116N in solution with a short incubation and thus was an ideal target for crystal soaking. Although the probe contains a large BODIPY moiety, the TxGH116C crystals appeared able to accommodate the ligand upon soaking.

Data for the fluorescent cyclophellitol aziridine complex extended to 2.16 Å. However, in the crystal data the KY375 ligand appears to have hydrolysed at the acyl linker region and there is no density available for any of the ligand beyond the nitrogen atom, (**Figure 4.12).**

Previous SDS-PAGE gel labelling of bacterial GH116 proteins in cell lysates had confirmed that the KY375 probe was stable and accommodated by all three of the tested bacterial enzymes.

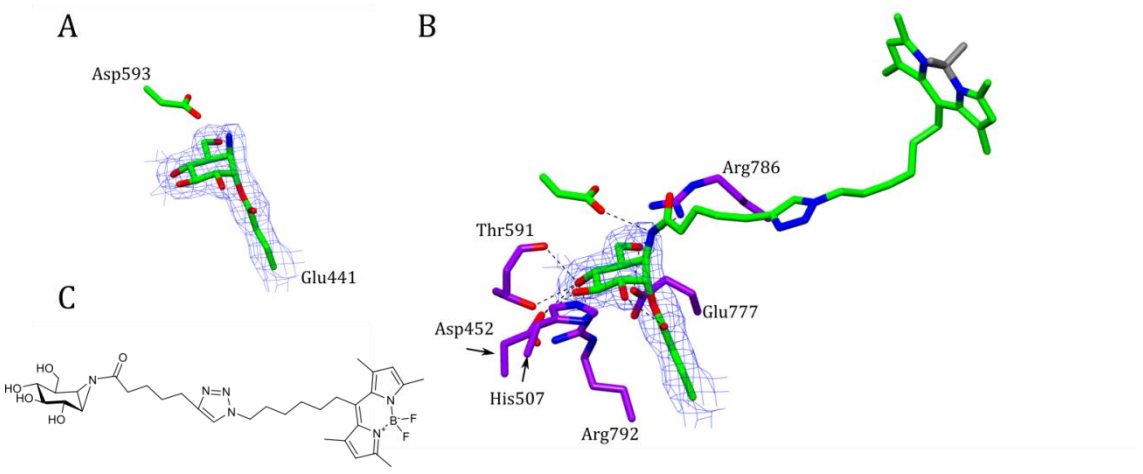


Figure 4.12. KY375 in the active site of TxGH116C. A) Observed electron density is shown at 1.2 σ (equivalent to 0.34 electrons / Å^3) The KY375 probe has been modelled to the density seen. B) The potential location of the full KY375 probe. The interacting side chains are shown in purple, the active site residues in green. C) The structure of KY375.

That TxGH116 (and therefore GBA2) can accommodate such large ligands with these extended alkyl chains and florescent groups can be shown through the effective tagging of the enzymes in solution, and the retention of the tags through the SDS-PAGE gel.

However soaking of the ABP into the crystal lattice may be disrupting the crystal lattice, at the boundary of each molecule, effectively reducing the resolution limits of the data

collection. This may also be reflected in the different space group of this data set ($P2_12_12_1$ rather than $P2_12_12$) due to the shifting of the molecules in the lattice as the enzyme accommodates the ligand.

Nevertheless, the complex of KY375 and TxGH116C does give some insight into how the enzyme may bind large substrate molecules in its substrate binding cleft (**Figure 4.13**). Whilst the covalently bound cyclophellitol aziridine warhead remains in place, the linker and reporter region of the probe may be able to move around in space in the enzyme's substrate binding cleft, allowing for large, bulky reporter groups to be accommodated by the enzyme.

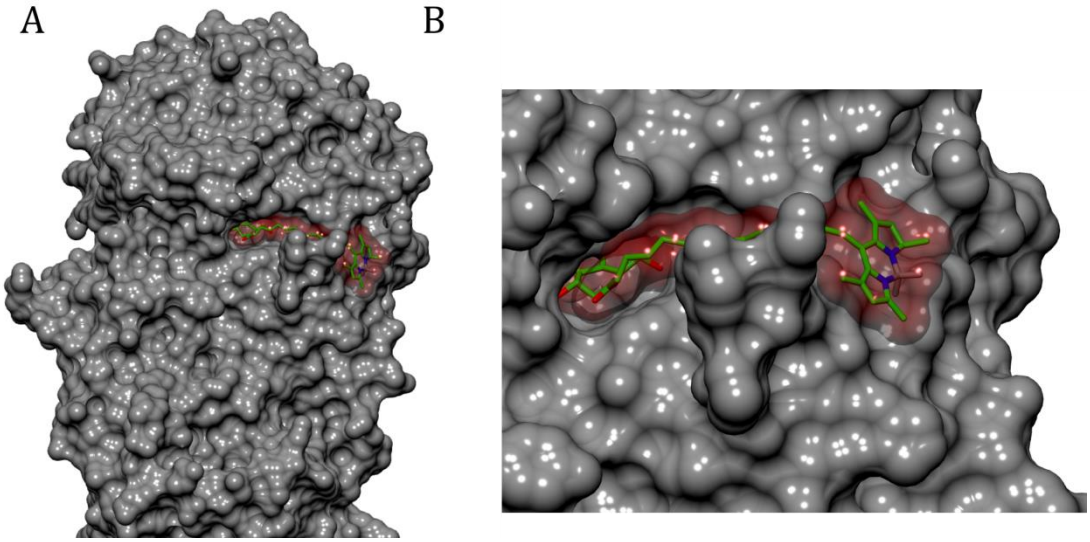


Figure 4.13. Surface view of TxGH116C bound to KY375, showing potential positioning of the ABP in the enzyme. A) There is a cleft on the surface of the protein in which the ABP may lie. B) A closer view of the substrate binding region.

An understanding of how TxGH116 is able to accomdate these larger ABPs can be applied to the human GBA2. KY375 itself has been shown to label GBA2 from whole cell lysates, however it also readily labels GBA1 and GBA3 (17). New probe designs would aim to target GBA2 selectivly.

4.4.8 TxGH116 and other retaining β-glucosidases

Retaining β-glucosidases are found in many CAZy families, e.g. GH1 (including GBA3), GH2, GH3, GH5, GH30 (including GBA1). These GH families can be further sorted into "clans" based upon their three dimensional structure. For example, GH1, GH2, GH5 and GH30 family enzymes possess $(\beta/\alpha)_8$-barrel catalytic domains and are sorted into clan GH-A, whereas GH3 enzymes have an active site between a $(\beta/\alpha)_8$-barrel and a β-sandwich domain (**Figure 4.14**).

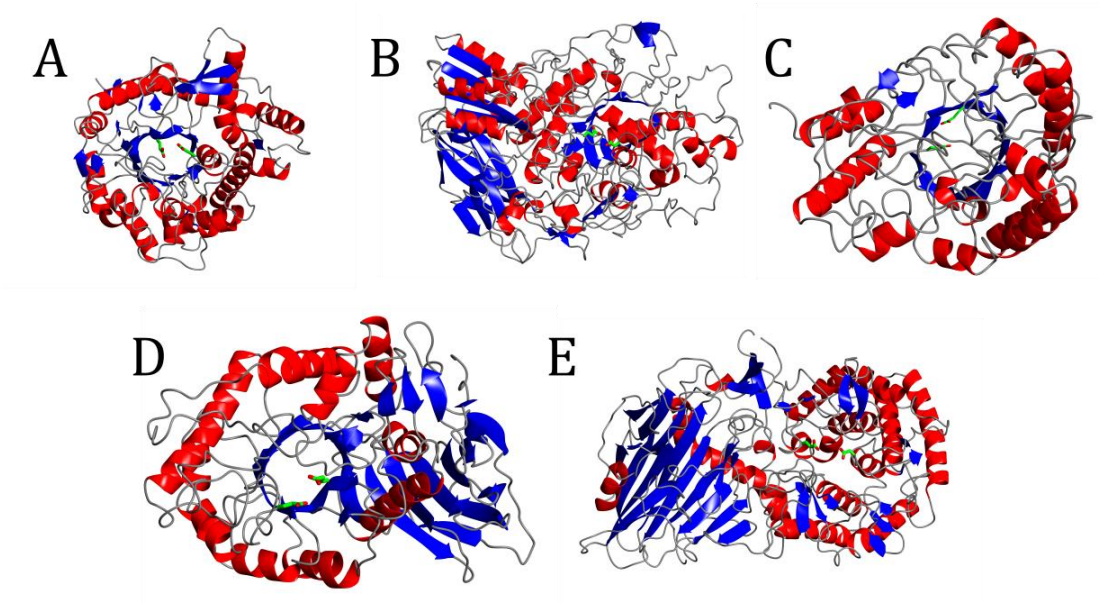


Figure 4.14. The structure of TxGH116 compared to other retaining β-glucosidase families, active site residues are shown in green. A) The structure of human GBA3, a GH1 family enzyme (PDB ID: 2E9L(138)). B) A GH3 family β-glucosidase from *Aspergillus oryzae* (PDB ID: 5FJJ (204) C) A GH5 family β-glucosidase from *Saccharomyces cerevisiae*, (PDB ID: 1H4P(205). D) The structure of human GBA1, as described in **Chapter 2.** E) The structure of TxGH116, GH116.

When TxGH116N is submitted to the PDBe Fold server (206), there is no structural similarity seen to any of the other retaining β-glucosidases, **Figure 4.13.** However, superposition of the GH52 β-xylosidase from *Geobacillus thermoglucosidasius* (PDB ID: 4C1O, GtGH52) (207) gives an RMSD (Root Mean Square Deviation) of 3.1 Å across Cα atoms within molecule A of the structure **Figure 4.15.B.**

Through the PDBeFold server, a direct comparison of chain A from both GtGH52 (PDB: 4C1O) and TxGH116N (PDB: 5FJS) across 556 matched residues, gives Q score of 0.30 and

Z score of 8.1. There is sequence identity between GtGH52 and TxGH116 of 13 %. This has led to the formation of CAZy structural clan GH-O, containing the GH family 52 and 116 enzymes with the common $(\alpha/\alpha)_6$ motif.

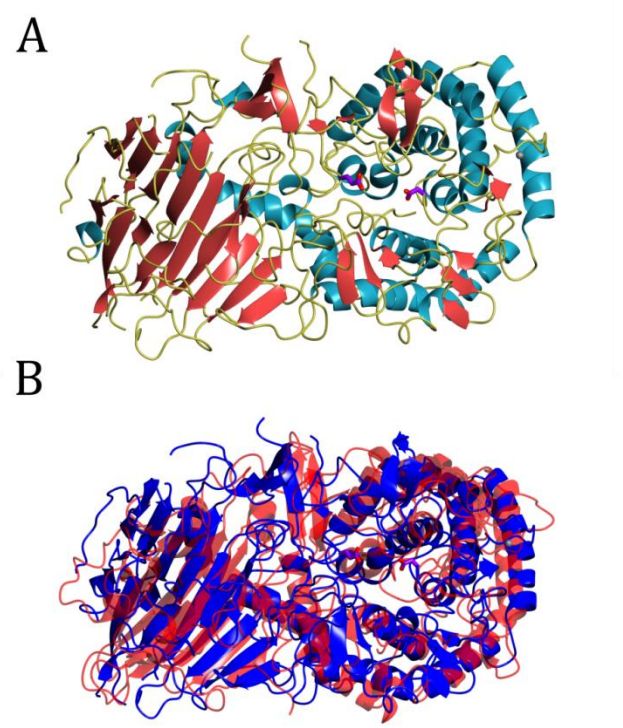


Figure 4.15. The 3D-fold structure of TxGH116N in comparison to that of GH52 family proteins. A) TxGH116N coloured by secondary structure C) An overlay of TxGH116N (blue) with a *Geobacillus thermoglucosidasius* β-xylosidase from family GH52 (PDB ID: 4C1O)(207).

The GH52 enzyme from *Geobacillus thermoglucosidasius* (GtGH52) is a thermophillic β-D-xylopyranosidase that possesses a retaining mechanism to hydrolyse xylo-oligosaccharides to xylose (208). The catalytic nucleophile in GtGH52 has been identified as Glu357, while Asp516 acts as the acid/base through both sequence alignment with other *Geobacillus* family xlyosidases and site-directed mutagenesis (209).

The active residues in GtGH52 are 6.5 Å apart, which is closer to the distance seen in most other retaining β-glucosidases (in comparison to being 10.8 Å apart in the TxGH116N) but still further than seen in most other retaining β-glucosidases (**Figure 4.16**). In the complex of GtGH52 with xylobiose, the substrate can be seen to be making hydrogen bonds with three residues, Asp367, Thr515 and Arg715.

A GESAMT/SSM (210) overlay of the active sites of GtGH52 (residues 340 to 370) and TxGH116N (residues 432 to 462) shows that two of these residues are maintained between the two proteins; Thr515/591 and Asp367/452, GtGH52/TxGH116 respectively (**Figure 4.16.B**). Further, the His418/507 residue is also present; this trio of residues have been earlier described as vital in hydrogen bonding interactions in the TxGH116 active site (**Figure 4.8**)

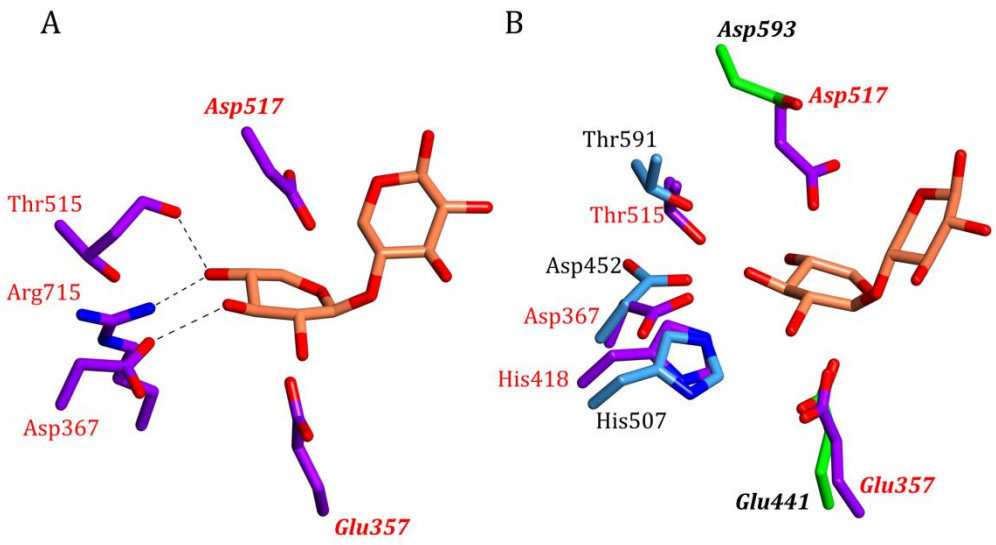


Figure 4.16. A comparison of the active sites of TxGH116N and GtGH52. A) The active site of GtGH52 in complex with xylobiose (coral), PDB ID: 4C1P (207). Active site residues are identified in italics. B) The active site of TxGH116N (green and cyan, labelled in black) superimposed with that of GtGH52 (purple, labelled in red).

4.4.9 Homology Modelling of the Human GBA2 Enzyme

TxGH116 shares 38% sequence identity over residues 63 to 893 with the human GBA2 protein, with notable correspondence in the active site residues with the residues noted as conserved across the two proteins. Therefore, an informative homology model of human GBA2 can generated from the three dimensional structure of TxGH116. The Phyre^2 server (198) was used to create a homology model of GBA2 based upon the 3D structure of TxGH116. Of the human protein, 75% could be modelled at >90% confidence over residues 78-889.

Human GBA2 has two longer loops, one of which lies adjacent to the enzyme binding site and may contribute to the aglycone binding cleft **Figure 4.17.A.** However, these loops cannot be modelled with high confidence, as they are not present in the bacterial protein.

This modelling allows the location and implication of known GBA2 pathogenic mutations to be mapped onto the GBA2 model and their effects on the protein considered, **Figure 4.17.B.** Whilst some GBA2 mutations (D594H, R873H) are located within the protein's active site, and therefore have clear catalytic repercussion for the enzyme, others are considerably distant and their implications are less immediately clear (**Table 4.4**).

Whilst many of these mutations are associated with hereditary spastic paraplegia/cerebellar ataxia (SPG46)(133) in humans, the pathogenic effect of these mutation is discussed in more detail in **Chapter 3** (**Table 3.2**).

TABLE 4.4. PATHOGENIC GBA2 MUTATIONS AND THEIR EFFECTS

GBA2 mutation	**Corresponding TxGH116 residue**	**Effect on the protein**
F419V	F347	Disruption of domain interface
D594H	D508	Disruption of H-bonding with glycosyl 3OH group
R630W	R544	Disruption of salt bridges
G683R	G599	Disruption of cation binding
R873H	R786	Disruption of H-bonding to glucosyl 6OH group

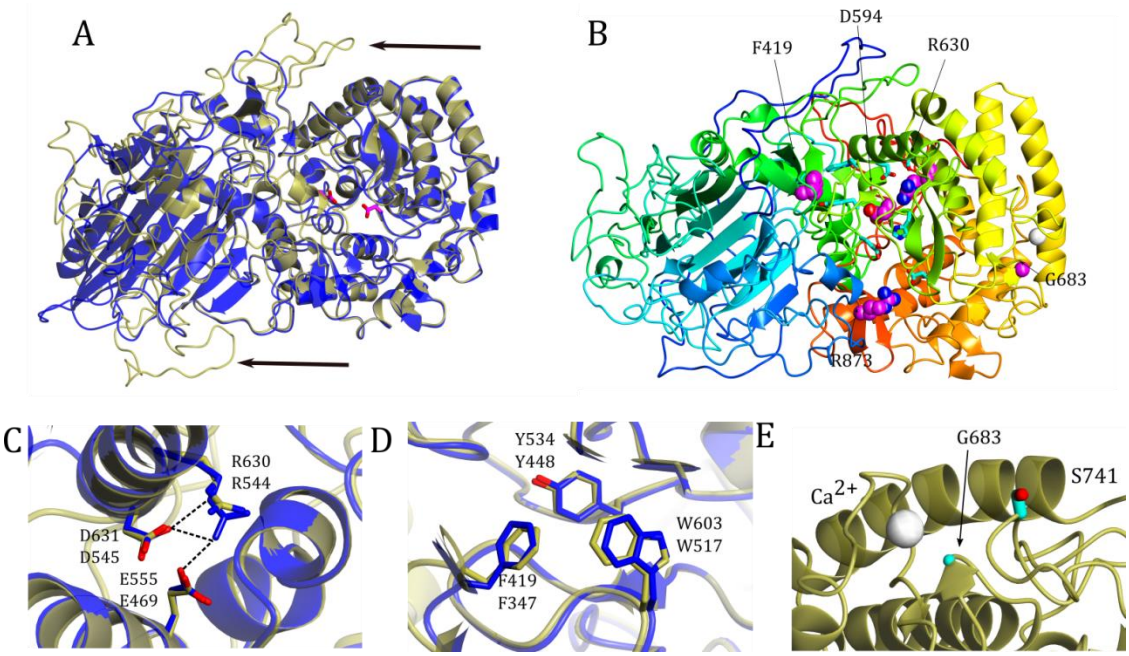


Figure 4.17. Human GBA2 model and known pathological mutants. Potential interactions are shown with dashed lines A) Overall homology model of human GBA2 (gold) superimposed on TxGH116N (blue). There are two distinct loops (marked with arrows) which are notably longer in GBA2. B) The human GBA2 protein homology model with residues involved in disease-causing mutations shown as purple spheres and labelled. C) Human GBA2 (gold) superimposed on TxGH116N (blue) in the area of the R630 residue. Residues are labelled human above/bacterial below, D) Human GBA2 (gold) superimposed on TxGH116N (blue) in the area of the F419 residue. E) Human GBA2 (gold) superimposed on TxGH116N (blue) in the area of the G683 mutation, showing the location of the Ca^{2+} ion, bound in the TxGH116N, respective to the residues. Human model generated on the Phyre2 server(198)

In the active site of TxGH116, R786 is potentially able to make two hydrogen bonds to the glucosyl 6OH group. This can be shown more clearly in the earlier models of TxGH116C bound to ligands (**Figure 4.8**). In human GBA2, this residue is conserved as R873, and the mutation R873H would result in a disruption of this hydrogen bonding.

Likewise, the human D594H mutation disrupts the hydrogen bonding that is shown as TxGH116 from D508 to H507 (conserved as H593 in the human protein), which is then able to make H-bond with OH3 (**Figure 4.8**). These D594H (equivalent to D508) and R873H (equivalent to R786) mutations are seen in patients with autosomal recessive cerebellar ataxia. Whilst these located in close proximity to the active site, other of the more commonly seen human GBA2 mutations are comparatively further away.

The GBA2 R630W mutation, which is seen in some hereditary spastic paraplegia patients, is conserved residue in the TxGH116 enzyme as R544. This is approximately 15 Å from the active site, and is situated between layers of the helices of the $(\alpha/\alpha)_6$- solenoid catalytic domain. Here R544 is able to form salt bridge interactions with D545 and E469, residues that are conserved in the human protein (**Figure 4.17.C**). The mutation of arginine to tryptophan would disrupt this charge balance due to the loss of a charged residue, and the introduction of the larger side chain. This together can result in the destabilising of the protein.

Also recorded in hereditary spastic paraplegia/cerebellar ataxia patients is the F419V mutation, which is located in the N-terminal, β-sandwich domain. It is the only pathogenic mutation located in this domain, although it is located at the interface of the two domains. Here, F419 can be involved in Van der Waals' interactions with the neighbouring aromatic residues, which are also conserved between the human and bacterial proteins (**Figure 4.17.D**). Residue Y534 is located on the long loop between the H5 and H6 helices and W603 is coming from the C-terminal domain, on the loop between the H7 and H8 helices. Disruption of these interactions, by replacing the phenylalanine with a valine, may destabilise the interaction of the two domains.

The G683R mutation, also seen involved in heredity spastic paraplegia, is located 20 Å from the active site, near the surface of the protein, in the outside turn of the loop that includes the catalytic acid/base (D677 in GBA2), **Figure 4.17.E**. The location of G683, very close to the neighbouring loop, is such that if it were to be replaced by an arginine this would result in a clash with S741 on the neighbouring loop. The G683 residue (equivalent to G599) is also neighbouring to the binding site for the Ca^{2+} ion, and replacement of the glycine with a positively charged arginine could disrupt the cation binding.

4.5 Summary

Crystallographic studies were used to discern the three-dimensional structure of TxGH116, a bacterial β-glucosidase from *Thermoanaerobacterium xylanolyticum*, both unliganded and in complex with a series of probes and inhibitors.

At the start of this thesis there were no known structures in the CAZy GH116 family of proteins, and the knowledge of the structure of TxGH116 , a bacterial homolog of the human β-glucosidase GBA2, may aid the design of probes specific to GBA2 over GBA1. This insight into the structure of the human GBA2 protein also allows us to see the potential effects of the known pathogenic mutations, both upon the active site of the enzyme and with effects on protein stability and domain interaction.

The structure of TxGH116 bound to the ABPs KY358 and KY375 supports recent work looking into distinct probes for the three human β-glucosidases. Knowledge of how the TxGH116 enzyme is able accommodate larger probe constructs can be applied to the human GBA2 enzyme models. The cyclophellitol aziridine that makes up for the basis of each of these probes is also the precursor for the "ABP 4" (**Figure 4.18**) a probe that potently and selectively labels GBA2 (Lahav et al, 2017, accepted for publication). Such specific reagents and the promise of GBA1 specific reagents and their applications will be described in the "Conclusions and Perspectives" Chapter 5.

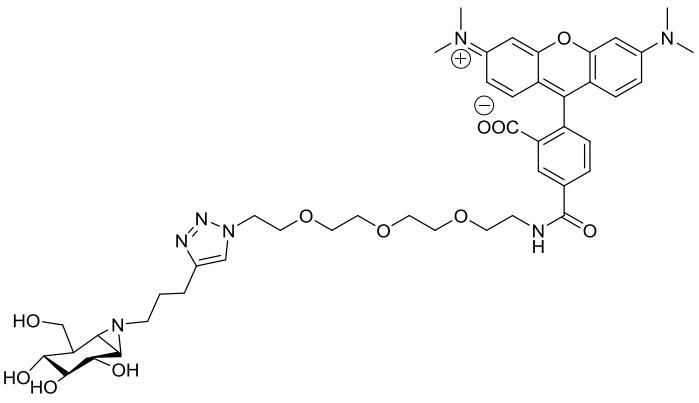


Figure 4.18. The activity based probe, designated "APB 4" has been utilised as an effective reporter in small molecule inhibitor assays for the human GBA2 enzyme.

Chapter 5. Conclusions and perspectives

At the start of this thesis, there were structures for the human glucocerebrosidase GBA1 in complex with diverse inhibitors and putative chaperones. For GBA2 far less was known, the human enzyme had been characterised and Moracci had led the way with characterization of archaeal GH116 enzymes - notably in terms of mechanism, catalytic residues and specificity.

The field of activity-based probes for glucosidases was also just emerging, following pioneering work from Vocadlo and Bertozzi on the human O-GlcNAcase (211)and on diverse glycosidases using 2F glycosides (212). Building on this, Overkleeft introduced the cyclophellitol epoxides and later aziridines, notably to the study of the Gaucher enzyme GBA1 (127).

The goals of my PhD, funded as part of a larger European Research Council grant, was to

- Analyse the 3D structure of human GBA1 in complex with diverse ABPs, in order to understand their binding and provide aid to the design of new inhibitors and probes
- To try and solve a 3D structure for a GH116 family enzyme, again to aid the design of specific inhibitors and probes.

In **Chapter 2**, I described the analysis of GBA1, crystallized using material provided by Sanofi-Genzyme. Despite problems with crystal cracking, I was able to solve the structures in complex with representative probes. The work highlighted how the key specificity feature - bulky substituents - could be accommodated at the O6 position and how hydrophobic substituents at the aziridine nitrogen aid potency. An obvious extension of this work would be to combine the two features; such work has just started in the Overkleeft group, **Figure 5.1**.

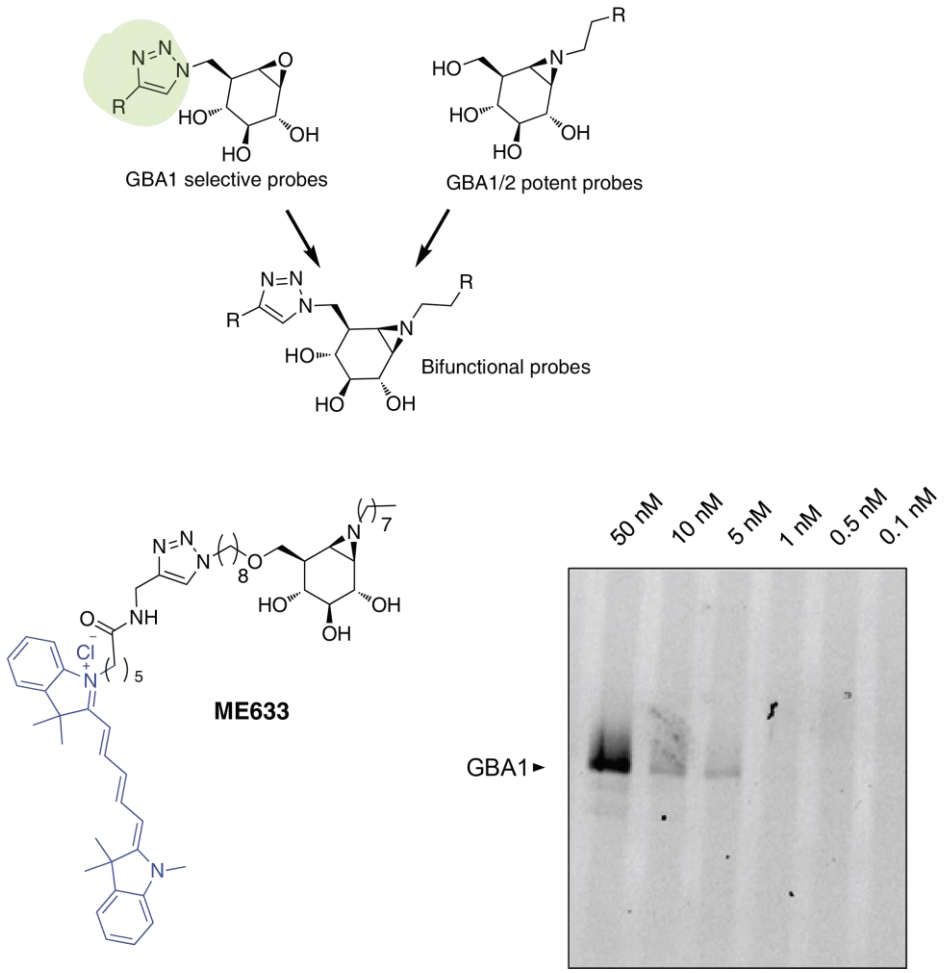


Figure. 5.1 Bifunctional GBA1 inhibitors under development in Overkleeft group. GBA1 structures suggest that bifunctional reagents might benefit from the specificity of the O6 substitutions plus the potency that derives from a functionalized aziridine (see Chapter 2, **Figure 2.10**). The first such species, ME633 has just (Autumn 2017) been synthesized and inhibits GBA1 on gel assays. Gel data were provided by Dr Martha Artola, University of Leiden, unpublished.

The bifunctional probe shown in **Figure 5.1,** ME633, effectively inhibits GBA1 (IC_{50} = 1.32 nM). Importantly, no labelling of GBA2 was reported by the di-alkylated aziridine probe. This does suggest specific, tight binding inhibitors are now a reality; the future will see these extended to cellular studies.

During the course of my work, diverse ABPs have already been synthesized and used to label diverse medically-relevant glycosidases - indeed they are now under investigation for clinical diagnosis. Probes for many classes of enzyme are now available:

- β–D galactosidases, Willems et al, 2014 (213)
- α–L fucosidases, Jiang et al, 2015 (214)
- α–amylases, Caner et al, 2016 (215)
- α–D glucosidases (Pompe disease) , Willems et al, 2014 (213)
- α–galactosidases, Willems et al, 2014 (159)
- β–glucuronidases (including heparanase), Wu et al, 2017 (216)
- β–D-6-phosphoglucosidases, Kwan et al, 2016 (217)

Beyond medicine, one future application of these probes could to use them for biomass degrading enzyme discovery. The CAZY classification (www.cazy.org, (143)) lists hundreds of thousands of ORFs for glycosidases, but industry still struggles to find the correct enzymes for the correct substrates under the correct conditions; the rapid pace of genome sequencing has given us many genes that are predicted to encode proteins with particular activities. The major current barriers are not merely trying to predict activity based on sequence alone, but also to be able to assess, rapidly, whether the encoded proteins have the predicted activity and are robust enough for industrial applications with respect to specific activity, stability, pH optimum, protease sensitivity *etc.*, and ultimately to find required enzymes in complex environmental samples without the need for genomic data. Activity-based probes could provide one solution to these problems for rapid screening of secretomes from industrially relevant organisms.

The main drawback with all current activity-based probes is that they demand an enzymatic nucleophile and are this only suitable for retaining enzymes. Clearly much needs to be done to find probes that will react, specifically, with inverting enzymes.

In **Chapters 3** and **4**, I described the 3D structure and complexes of a bacterial GBA2 homolog, also active on β–glucosides. I am grateful to Professor Jim Cairns for openly sharing and exchanging data to allow a joint publication (190). Ultimately I was able to solve the form of the enzyme expressed in York, and to obtain high resolution complexes on inhibitors and activity-based probes. We were able to define the fold, establishing a "Clan" with the GH52 β–xylosidases and to map the human genetic disease variants. Unfortunately, the TxGh116 isn't active on glucosylceramide and the group, as yet, has been unable to crystallize the human enzyme despite successful over-expression. There is clearly a need for specific GBA2 inhibitors and substrates. The Vocadlo group have recently published on fluorescent-quenched substrates for human GBAases (218). My work on the fluorescent ABPs binding to GBA2 is included in a paper accepted at the

Journal of the American Chemical Society on a fluorescent polarisation assay for GBA1/2 inhibitors, **Figure 5.2**. The principle of this work, well established for drug screening (see for example Lea and Simenov, 2012 (219)) is to use the observation that in fluorescence polarization assays, a fluorophore is excited at a specific wavelength with linear polarized light. The orientation of emitted light and thereby the degree of polarized light depends on the movement the excited fluorophore has undergone, which depends on its size: large molecules (such as GBA2-bound probes) move and rotate slower than small molecules. If binding of probe is competed with small molecule inhibitors, such as from libraries, then one can measure binding constants for compounds through loss of the polarized emission, **Figure 5.2**. Using this technique, we have, through collaboration, started to discover GBA2 specific inhibitors (Lahav et al., *JACS*, in press; see **Appendix 8**)

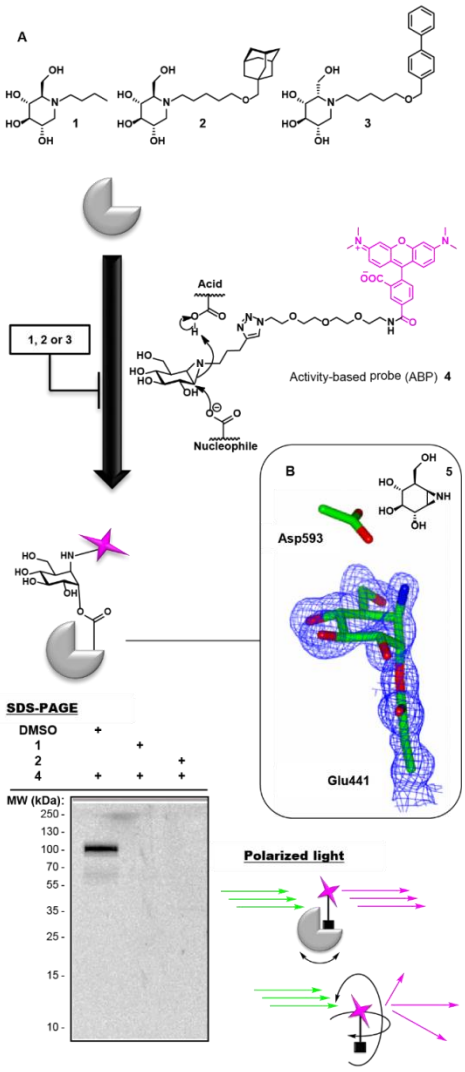


Figure 5.2. An exciting new direction for GBA1/2 inhibitor work is to use the ABPs in fluorescent polarisation assays, illustrated here with human GBA2 (taken from REF Lahav et al., JACS, in press). If the fluorescent group is immobilised, the fluorescent output is polarised, whilst a tumbling fluorophore depolarizes the fluorescent emission. Hence, an immobilised fluorescent ABP can be used to screen for tight(er) binding inhibitors that compete with probe binding. In the example shown, compounds **1** and **2** both compete with the probe (as shown by the gel assay, bottom left). In the fluorescence assay these cause a loss of polarised light. Binding curves give a direct assessment of Kd.

The GBA2 work, has also led to the very recent discovery of new inhibitor classes. Seeking to improve upon aziridines and epoxides, the Davies and Overkleeft group conceived cyclic

sulfates as irreversible glycosidase inhibitors and I used the bacterial GBA2, TxGH116, as one of the testbed enzymes, **Figure 5.3.**

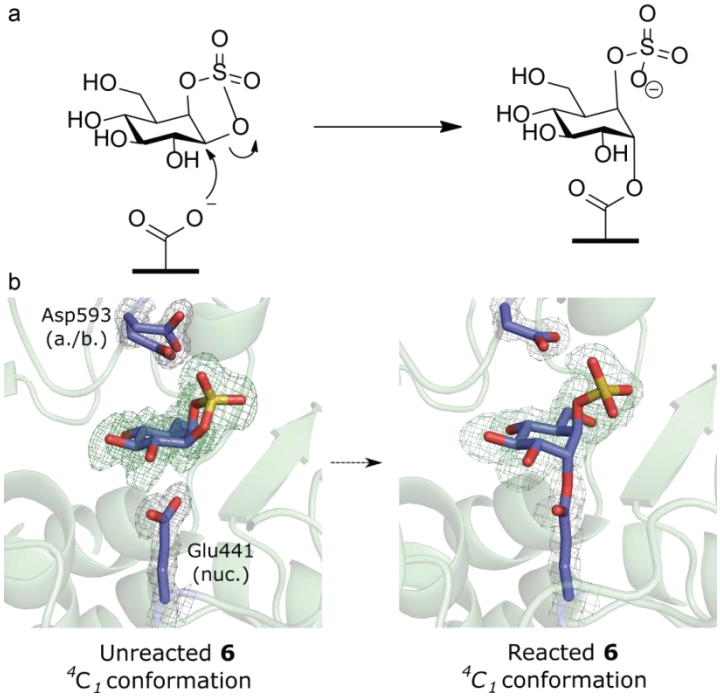


Figure 5.3. Cyclic sulfates (here a cis "beta" glucose" version compound **6**): a new retaining glycosidase inhibitor ((125), also **see Appendix 6**) (a) the reaction of cyclic sulfates (b) 3D structures of unreacted cyclic sulfate on TxGH116C (Liang Wu, short soak, left) and reacted cyclic sulfate on TxGH116C (Breen, long soak, right). Figure adapted from Artola et al, 2017 (125).

Clearly further work will be needed to establish whether substituted cyclic sulfates have application as ABPS, or even as glycosidase inhibitors in a medical context.

To identify further potential ligand targets for TxGH116, a fluorescence-based thermal shift assay (TSA) utilising SYPRO orange dye was used in a high-throughput format (**Figure 5.4**). This allowed the assessment of the effect of ligand binding on protein stability with minimal use of protein recourses. Of 22 ligands assayed, 8 were identified as raising the *Tm* of TxGH116C by > 5°C, These included the known TxGH116 inhibitors, isofagomine (K_i = 2.9 ± 0.1 nM (190)) and DNJ (K_i = 0.13 ± 0.01 μM (190)).

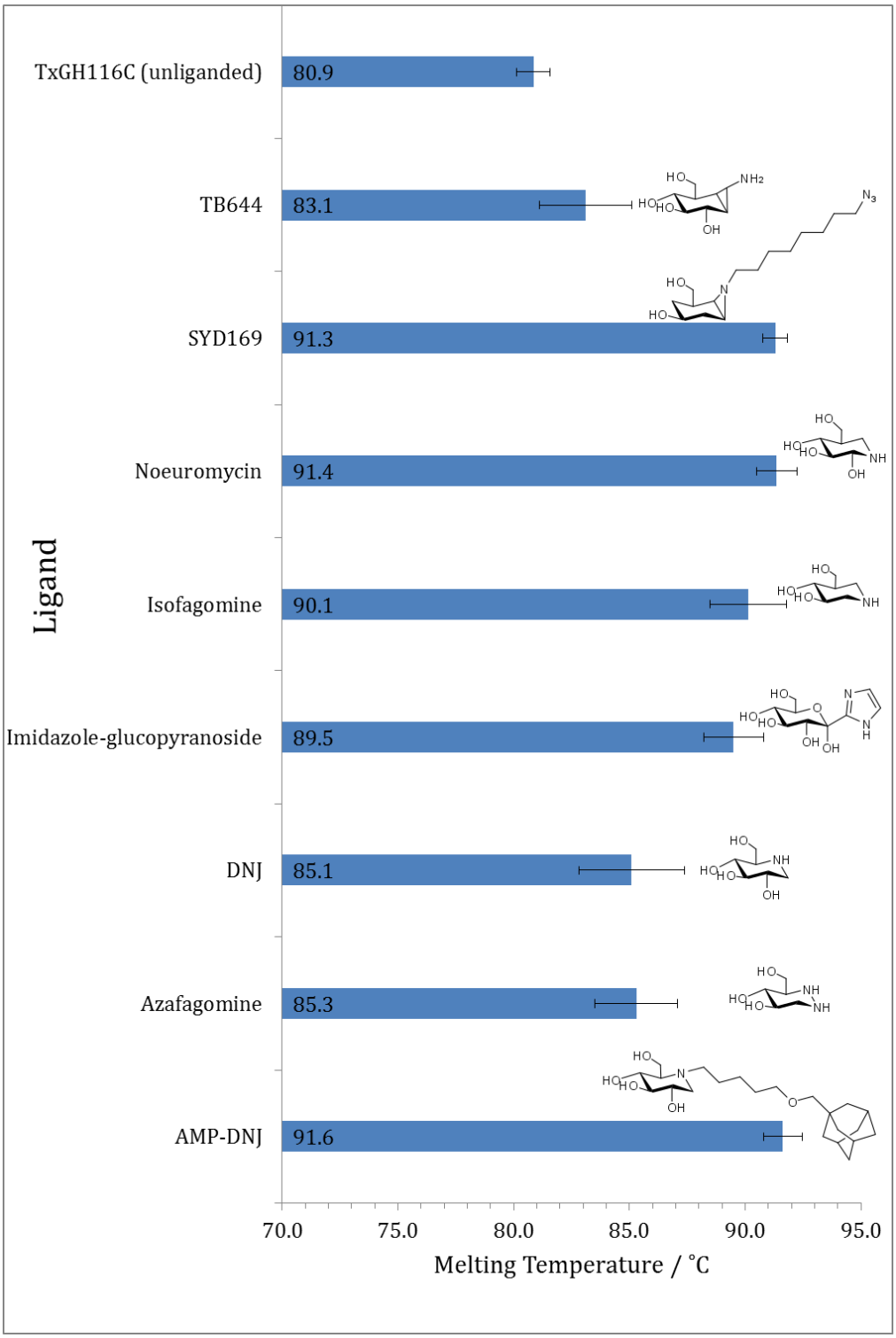


Figure 5.4. TSA high throughput screening of TxGH116C against a selection of potential ligand compounds. Mean average melting temperature is shown in °C, of three repetitions.

Unliganded TxGH116 protein has a melting temperature of 80.9 °C, as the bacteria that TxGH116 is sourced from, *Thermoanaerobacterium xylanolyticum,* is native to the Frying Pan Springs in Yellowstone National Park. The stabilising effect that certain of these ligands (AMP-DNJ, isofagomine, noeuromycin and SYD-169) had on TXGH116 was found to push the limits of the TSA.

To further measure the stabilising effects of the ligands, circular dichroism (CD) was used. The CD was programme to run a full scan from λ= 260 nm to λ = 180 nm at 20 °C and then every 5 °C from 60-95 °C. The machine also took a reading at λ = 222 nm every 5 °C from the start. The machine was able to increase at a maximum of 5 °C per minute, at held each temperature for 2 minutes before taking a reading. (**Figure 5.5**)

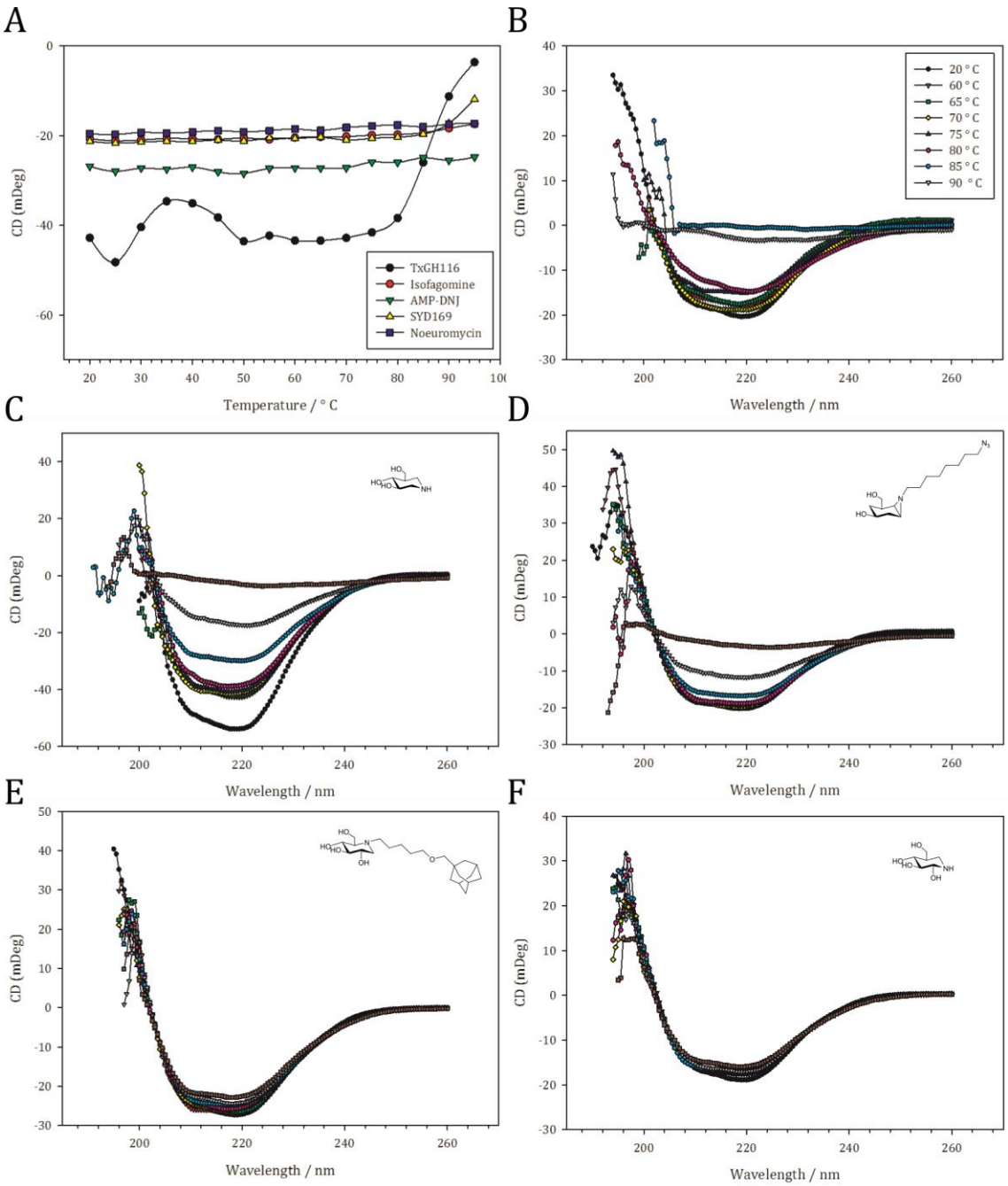


Figure 5.5. Circular dichroism spectroscopy of TxGH116C and the effects of ligand binding on protein stability as CD (mDeg) vs temperature. A) A λ = 222 nm scan of unliganded-TxGH116C and TxGH116C in complex with four ligands as per the key. Readings taken every 5 °C from 20 – 95 °C. B) CD scan of unliganded-TxGH116C, from λ = 180-260 nm. Data has been excluded when HT = >800. Scan temperature is given by the key. C) CD scan of isofagimine-TxGH116C**.** D) CD scan of SYD169-TxGH116C**.** E) CD scan of AMP-DNJ-TxGH116C. F) CD scan of Noeuromycin-TxGH116C **.**

As previously mentioned, isofagomine is a known inhibitor of TxGH116C (K_i = 2.9 ± 0.1 nM (190)), and AMP-DNJ has been shown to inhibit human GBA2 within brain membranes (IC_{50} = 0.8 nM) (179). Whislt isofagomine and SYD169 bound TxGH116C appear to "melt" from 90 °C, both AMP-DNJ- and noeuromycin-TxGH116C has remained in solution, and in the same fold, up to 95 °C.

Soaking experiments of TxGH116C with all four of the ligands has so far resulted in the crystals failing to maintain their order. Binding of these in the molecule during soaking appears to disrupt the crystal lattice. It is possible that co-crystallisation experiments can offer better results.

What effect these ligands are having on overall structure of TxGH116C is unclear, but the disruption seen during soaking suggests that there is some rearrangement within the crystal lattice. Co-crystallisation experiments will likely offer better results.

Treatments for lysosomal storage disorders, such as Gaucher's disease, are still mostly reliant on enzyme replacement therapy, which are expensive, inconvenient and unable to treat neuropathic symptoms. However the use of small molecule drugs in substrate reduction therapies has allowed physicians to treat the neuropathic forms of several lysosomal storage diseases. Niemann-Pick type C disease, for example, in which cholesterol trafficking to lysosomes is disrupted, is treated with the same ceramide glucosyltransferase inhibitor (miglustat) as some forms of Gaucher's disease (220).

The emerging pharmacological chaperone therapy, in which the correct folding of mutant forms of an enzyme is promoted within the ER by small-molecule chaperones, aims to further reduce potential drug side effects that can arise from SRT. PCT can also be used to increase the efficiency of the recombinant enzymes used in ERT; the aim being to reduce the volume of recombinant enzyme required per patient and increase the time between injections and thus reduce the cost (both financial and time spent in hospitals) to the patient.

Both these uses of PCTs are very dose sensitive and accurate methods of measuring any given enzyme levels in a patients cells are necessary, for these therapeutic approaches and all sorts of diagnostic tests (94). The ideal diagnostic tool would measure only the active enzyme within a cell, rather than total, potentially inactive or misfolded, enzyme.

Screening for potential ligand compounds and knowledge about the effects of molecule binding on enzymes can help inspire the design of new ABPs and chaperones for both

GBA1 and GBA2. However to get from an enzyme inhibitor to a potential drug compounds is a long and gruelling process, the hit to lead to final drug development process will often take over a decade and can cost $1 billion (221)

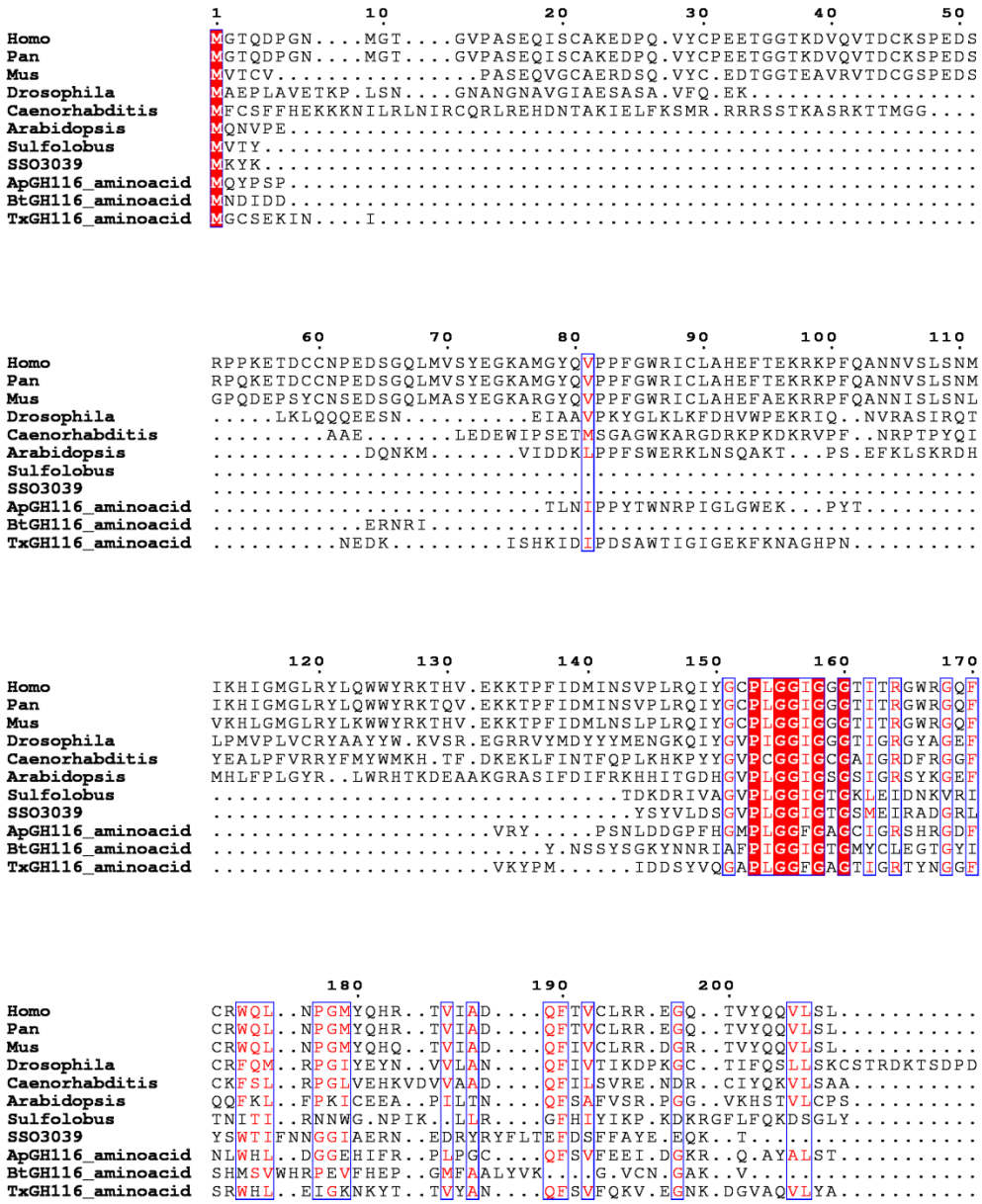
For LSD patients, however, there are some emerging therapies that offer hope of, if not a cure, treatments that could alleviate symptoms entirely. Firstly there is *ex vivo* gene therapy, in which a patient' own cells are harvested, transduced by a viral vector to overexpress the lysosomal enzyme of interest and reintroduced to the patient where they can restore the healthy phenotype. A clinical trial using stem cells from patients with Fabry disease and the retroviral transduction α-galactosidase A cDNA was begun in 2013 but later withdrawn (222).

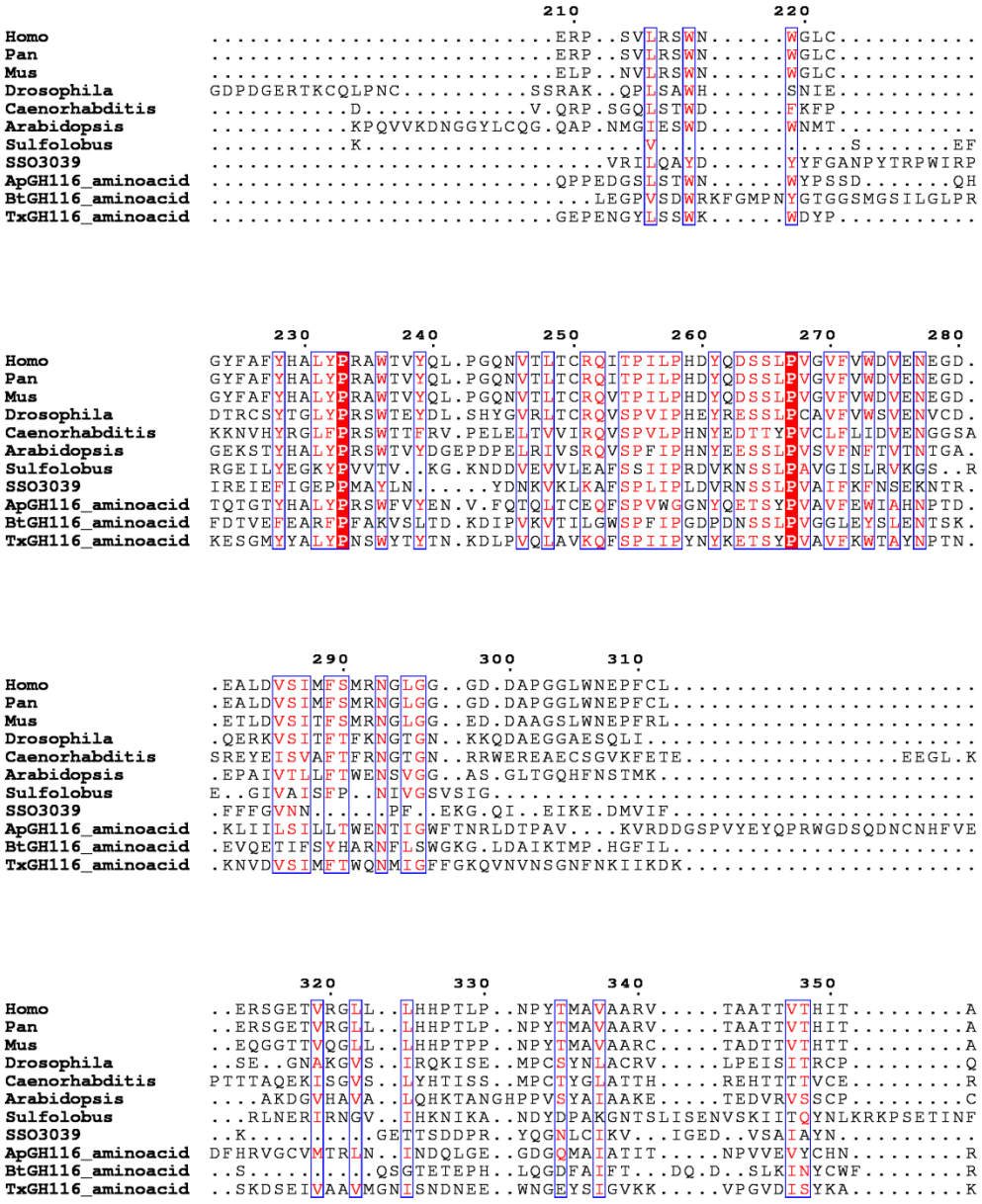
The use of hematopoietic stem cell transplantation for the treatment of Krabbe disease (a deficiency of galactosylceramidase) is now the standard treatment. Healthy donor cells are able to infiltrate organs and cross the blood brain barrier. Once established, the donor cells excrete the required enzyme. Two clinical trials utilising this approach, for Gaucher and Fabry disease have begun (223).

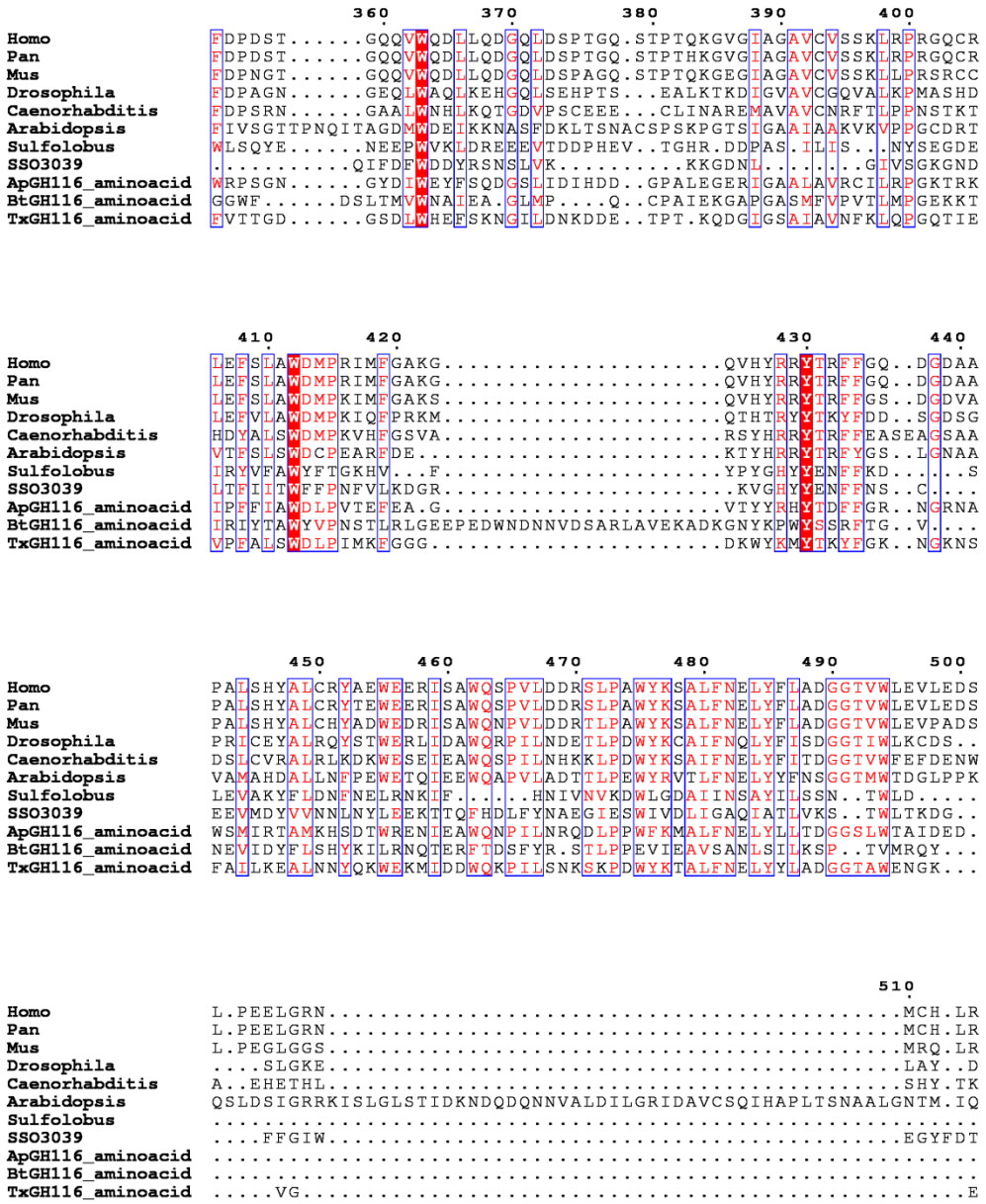
The use of *in vivo* gene therapy in human for lysosomal storage disorders has mostly explored the use of adeno-associated viral vectors and focus has mainly been on those diseases for which no other treatment option is present. Both feline and canine models homologues of lysosomal storage disorders including mucolipidosis II (fatal in humans in infancy) have been involved in trials of *in vivo* gene therapy (224).

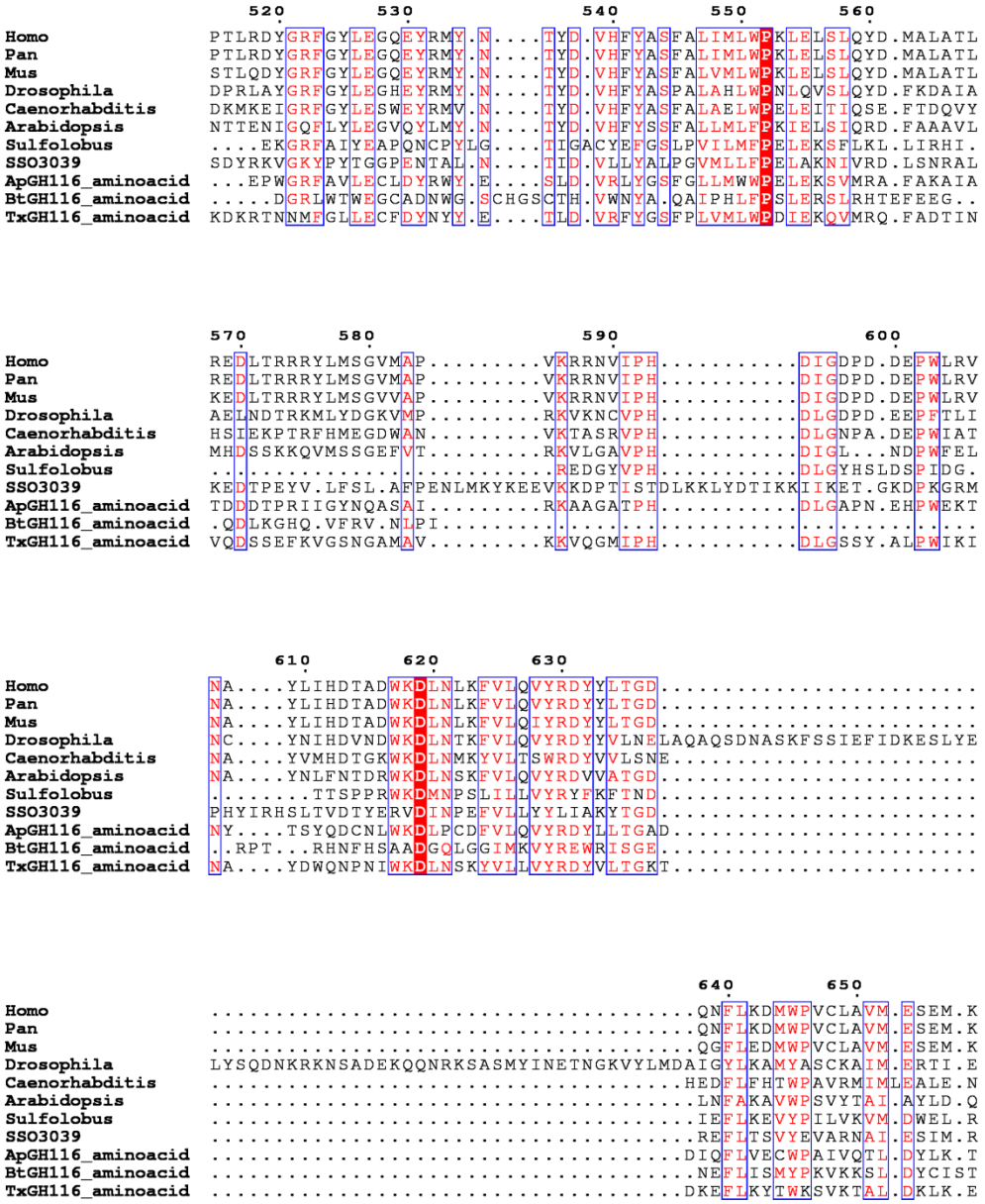
The potential for a cure in gene therapy is still several years in the making for Gaucher's disease. In the meantime, there can be hope that small molecule treatments and the necessary therapeutic tools can be improved to the point that a patient with the illness can live a life free from any symptoms or sickness.

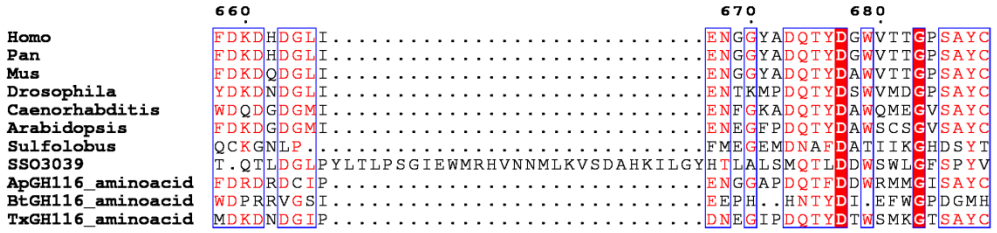
Appendix 1. GH116 protein sequence alignment

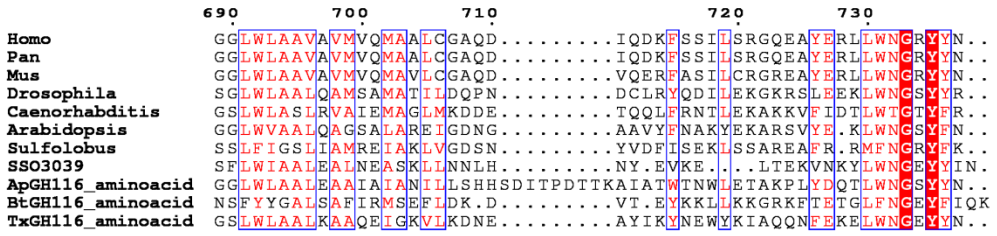


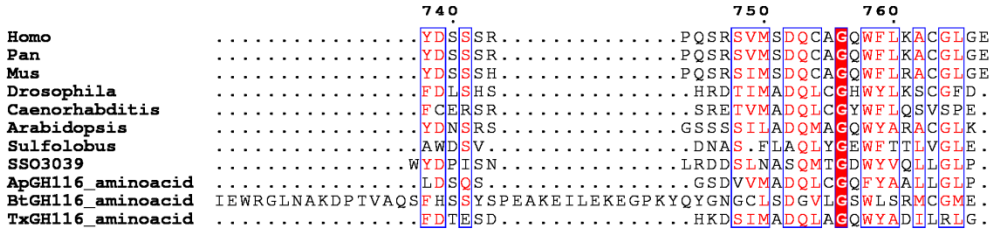

210
220
Homo .....ERP..SVLRSWN.....WGLC.....
Pan .....ERP..SVLRSWN.....WGLC.....
Mus .....ELP..NVLRSWN.....WGLC.....
Drosophila GDPDGERTKCQLPNC.....SSRAK..QPLSAWH.....SNIE.....
Caenorhabditis .....D.....V.QRP.SGQLSTWD.....FKFP.....
Arabidopsis .....KPQVVKDNGGYLCQG.QAP.NMGIESWD.....WNMT.....
Sulfolobus .....K.....V.....S.....EF
SSO3039 .....VRILQAYD.....YYFGANPYTRPWIRP
ApGH116_aminoacid .....QPPEDGSLSTWN.....WYPSSD.....QH
BtGH116_aminoacid .....LEGPVSDWRKFGMPNYGTGGSMGSILGLPR
TxGH116_aminoacid .....GEPENGYLSSWK.....WDYP.....
230
240
250
260
270
280
Homo GYFAFYHALYPRAWTVYQL.PGQNVTLTCRQITPILPHDYQDSSLPVGVFVWDVENEGD.
Pan GYFAFYHALYPRAWTVYQL.PGQNVTLTCRQITPILPHDYQDSSLPVGVFVWDVENEGD.
Mus GYFAFYHALYPRAWTVYQL.PGQNVTLTCRQVTPILPHDYQDSSLPVGVFVWDVENEGD.
Drosophila DTRCSYTGLYPRSWTEYDL.SHYGVRLTCRQVSPVIPHEYRESSLPCAVFVWSVENVCD.
Caenorhabditis KKNVHYRGLFPRSWTTFRV.PELELTVVIRQVSPVLPHNYEDTTYPVCLFLIDVENGGSA
Arabidopsis GEKSTYHALYPRSWTVYDGEPDPELRIVSRQVSPFIPHNYEESSLPVSVFNFTVTNTGA.
Sulfolobus RGEILYEGKYPVVTV..KG.KNDDVEVVLEAFSSIIPRDVKNSSLPAVGISLRVKGS..R
SSO3039 IREIEFIGEPPMAYLN.....YDNKVKLKAFSPLIPLDVRNSSLPVAIFKFNSEKNTR.
ApGH116_aminoacid TQTGTYHALYPRSWFVYEN.V.FQTQLTCEQFSPVWGGNYQETSYPVAVFEWIAHNPTD.
BtGH116_aminoacid FDTVEFEARFPFAKVSLTD.KDIPVKVTILGWSPFIPGDPDNSSLPVGGLEYSLENTSK.
TxGH116_aminoacid KESGMYYALYPNSWYTYTN.KDLPVQLAVKQFSPIIPYNYKETSYPVAVFKWTAYNPTN.
290
300
310
Homo .EALDVSIMFSMRNGLGG..GD.DAPGGLWNEPFCL.....
Pan .EALDVSIMFSMRNGLGG..GD.DAPGGLWNEPFCL.....
Mus .ETLDVSITFSMRNGLGG..ED.DAAGSLWNEPFRL.....
Drosophila .QERKVSITFTFKNGTGN..KKQDAEGGAESQLI.....
Caenorhabditis SREYEISVAFTFRNGTGN..RRWEREAECSGVKFETE.....EEGL.K
Arabidopsis .EPAIVTLLFTWENSVGG..AS.GLTGQHFNSTMK.....
Sulfolobus E..GIVAISFP..NIVGSVSIG.....
SSO3039 .FFFGVNN.....PF..EKG.QI..EIKE.DMVIF.....
ApGH116_aminoacid .KLIILSILLTWENTIGWFTNRLDTPAV....KVRDDGSPVYEYQPRWGDSQDNCNHFVE
BtGH116_aminoacid .EVQETIFSYHARNFLSWGKG.LDAIKTMP.HGFIL.....
TxGH116_aminoacid .KNVDVSIMFTWQNMIGFFGKQVNVNSGNFNKIIKDK.....
320
330
340
350
Homo ..ERSGETVRGLL..LHHPTLP..NPYTMAVAARV.....TAATTVTHIT.....A
Pan ..ERSGETVRGLL..LHHPTLP..NPYTMAVAARV.....TAATTVTHIT.....A
Mus ..EQGGTTVQGLL..LHHPTPP..NPYTMAVAARC.....TADTTVTHTT.....A
Drosophila ..SE..GNAKGVS..IRQKISE..MPCSYNLACRV.....LPEISITRCP.....Q
Caenorhabditis PTTTAQEKISGVS..LYHTISS..MPCTYGLATTH.....REHTTTTVCE.....R
Arabidopsis ....AKDGVHAVA..LQHKTANGHPPVSYAIAAKE.....TEDVRVSSCP.....C
Sulfolobus ...RLNERIRNGV..IHKNIKA..NDYDPAKGNTSLISENVSKIITQYNLKRKPSETINF
SSO3039 ..K.....GETTSDDPR..YQGNLCIKV...IGED..VSAIAYN.....
ApGH116_aminoacid DFHRVGCVMTRLN..INDQLGE..GDGQMAIATIT.....NPVVEVYCHN.....R
BtGH116_aminoacid ..S.....QSGTETEPH..LQGDFAIFT...DQ.D..SLKINYCWF.....R
TxGH116_aminoacid ..SKDSEIVAAVMGNISNDNEE..WNGEYSIGVKK.....VPGVDISYKA.....K

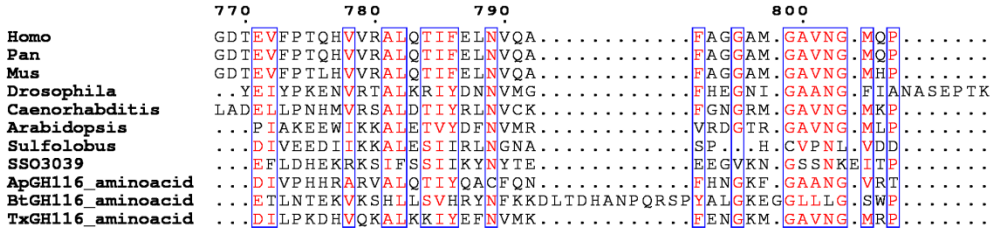

360 370 380 390 400
Homo FDPDST.....GQQVWQDLLQDGQLDSPTGQ.STPTQKGVGIAGAVCVSSKLRPRGQCR
Pan FDPDST.....GQQVWQDLLQDGQLDSPTGQ.STPTHKGVGIAGAVCVSSKLRPRGQCR
Mus FDPNGT.....GQQVWQDLLQDGQLDSPAGQ.STPTQKGEGIAGAVCVSSKLLPRSRCC
Drosophila FDPAGN.....GEQLWAQLKEHGQLSEHPTS...EALKTKDIGVAVCGQVALKPMASHD
Caenorhabditis FDPSRN.....GAALWNHLKQTGDVPSCEEE...CLINAREMAVAVCNRFTLPPNSTKT
Arabidopsis FIVSGTTPNQITAGDMWDEIKKNASFDKLTSNACSPSKPGTSIGAAIAAKVKVPPGCDRT
Sulfolobus WLSQYE.....NEEPWVKLDREEEVTDDPHEV..TGHR.DDPAS.ILIS...NYSEGDE
SSO3039 .........QIFDFWDDYRSNSLVK.....KKGDNL.....GIVSGKGND
ApGH116_aminoacid WRPSGN.....GYDIWEYFSQDGSLIDIHDD..GPALEGERIGAALAVRCILRPGKTRK
BtGH116_aminoacid GGWF.....DSLTMVWNAIEA.GLMP....Q..CPAIEKGAPGASMFVPVTLMPGEKKT
TxGH116_aminoacid FVTTGD.....GSDLWHEFSKNGILDNKDDE..TPT.KQDGIGSAIAVNFKLQPGQTIE
410 420 430 440
Homo LEFSLAWDMPRIMFGAKG.....QVHYRRYTRFFGQ..DGDAA
Pan LEFSLAWDMPRIMFGAKG.....QVHYRRYTRFFGQ..DGDAA
Mus LEFSLAWDMPKIMFGAKS.....QVHYRRYTRFFGS..DGDVA
Drosophila LEFVLAWDMPKIQFPRKM.....QTHTRYYTKYFDD..SGDSG
Caenorhabditis HDYALSWDMPKVHFGSVA.....RSYHRRYTRFFEASEAGSAA
Arabidopsis VTFSLSWDCPEARFDE.....KTYHRRYTRFYGS..LGNAA
Sulfolobus IRYVFAWYFTGKHV...F.....YPYGHYYENFFKD.....S
SSO3039 LTFIITWFFPNFVLKDGR.....KVGHYYENFFNS..C....
ApGH116_aminoacid IPFFIAWDLPVTEFEA.G.....VTYYRHYTDFFGR..NGRNA
BtGH116_aminoacid IRIYTAWYVPNSTLRLGEEPEDWNDNNVDSARLAVEKADKGNYKPWYSSRFTG..V....
TxGH116_aminoacid VPFALSWDLPIMKFGGG.....DKWYKMYTKYFGK..NGKNS
450 460 470 480 490 500
Homo PALSHYALCRYAEWEERISAWQSPVLDDRSLPAWYKSALFNELYFLADGGTVWLEVLEDS
Pan PALSHYALCRYTEWEERISAWQSPVLDDRSLPAWYKSALFNELYFLADGGTVWLEVLEDS
Mus PALSHYALCHYADWEDRISAWQNPVLDDRTLPAWYKSALFNELYFLADGGTVWLEVPADS
Drosophila PRICEYALRQYSTWERLIDAWQRPILNDETLPDWYKCAIFNQLYFISDGGTIWLKCDS..
Caenorhabditis DSLCVRALRLKDKWESEIEAWQSPILNHKKLPDWYKSAIFNELYFITDGGTVWFEFDENW
Arabidopsis VAMAHDALLNFPEWETQIEEWQAPVLADTTLPEWYRVTLFNELYYFNSGGTMWTDGLPPK
Sulfolobus LEVAKYFLDNFNELRNKIF.....HNIVNVKDWLGDAIINSAYILSSN..TWLD.....
SSO3039 EEVMDYVVNNLNYLEEKTTQFHDLFYNAEGIESWIVDLIGAQIATLVKS..TWLTKDG..
ApGH116_aminoacid WSMIRTAMKHSDTWRENIEAWQNPILNRQDLPPWFKMALFNELYLLTDGGSLWTAIDED.
BtGH116_aminoacid NEVIDYFLSHYKILRNQTERFTDSFYR.STLPPEVIEAVSANLSILKSP..TVMRQY...
TxGH116_aminoacid FAILKEALNNYQKWEKMIDDWQKPILSNKSKPDWYKTALFNELYYLADGGTAWENGK...
510
Homo L.PEELGRN.....MCH.LR
Pan L.PEELGRN.....MCH.LR
Mus L.PEGLGGS.....MRQ.LR
Drosophila ....SLGKE.....LAY..D
Caenorhabditis A..EHETHL.....SHY.TK
Arabidopsis QSLDSIGRRKISLGLSTIDKNDQDQNNVALDILGRIDAVCSQIHAPLTSNAALGNTM.IQ
Sulfolobus .....
SSO3039 ....FFGIW.....EGYFDT
ApGH116_aminoacid .....
BtGH116_aminoacid .....
TxGH116_aminoacid .....VG.....E

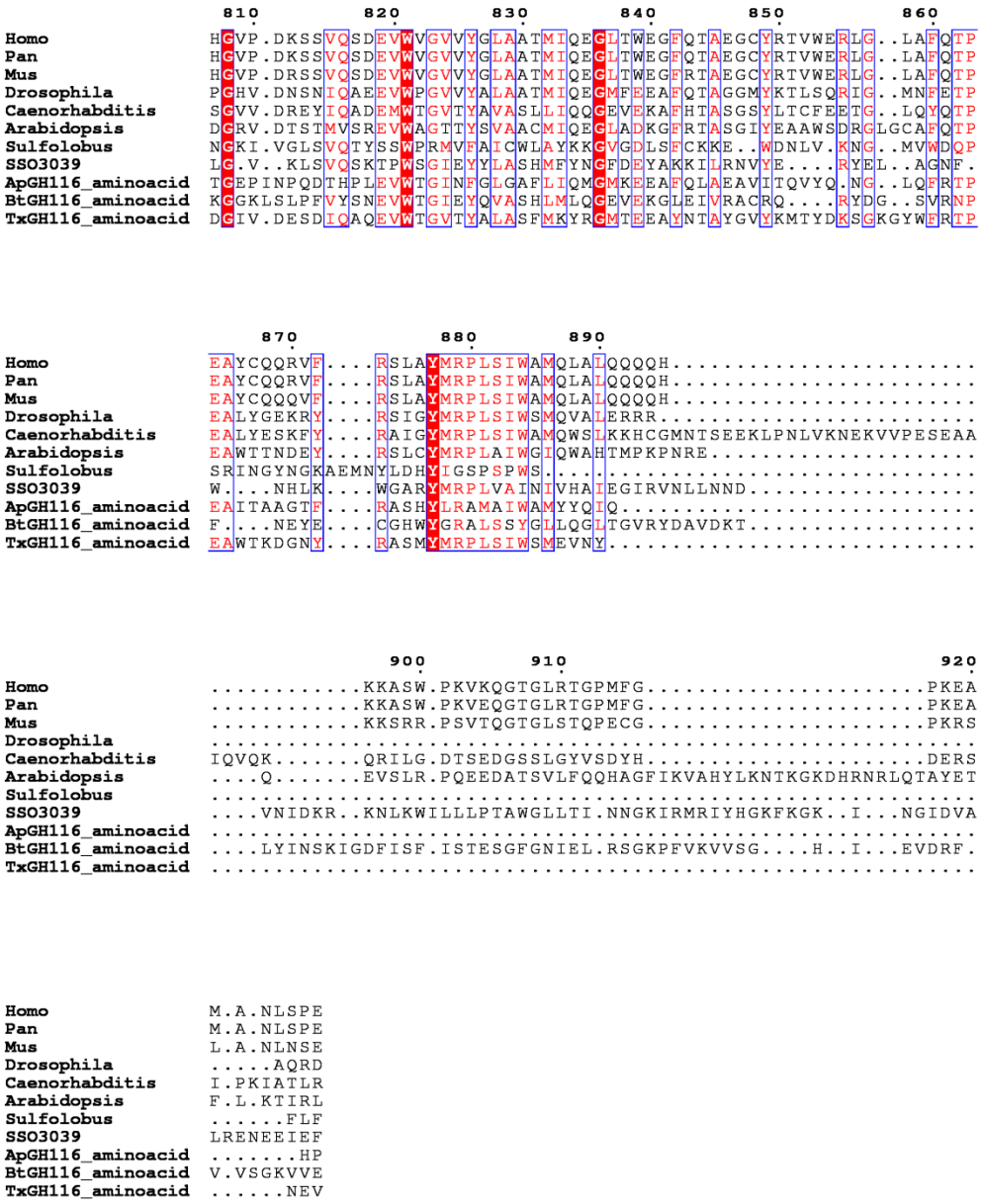

520 530 540 550 560
Homo PTLRDYGRFGYLEGQEYRMY.N....TYD.VHFYASFALIMLWPKLELSLQYD.MALATL
Pan PTLRDYGRFGYLEGQEYRMY.N....TYD.VHFYASFALIMLWPKLELSLQYD.MALATL
Mus STLQDYGRFGYLEGQEYRMY.N....TYD.VHFYASFALVMLWPKLELSLQYD.MALATL
Drosophila DPRLAYGRFGYLEGHEYRMY.N....TYD.VHFYASPALAHLWPNLQVSLQYD.FKDAIA
Caenorhabditis DKMKEIGRFGYLESWEYRMV.N....TYD.VHFYASFALAELWPELEITIQSE.FTDQVY
Arabidopsis NTTENIGQFLYLEGVQYLMY.N....TYD.VHFYSSFALLMLFPKIELSIQRD.FAAAVL
Sulfolobus ....EKGRFAIYEAPQNCPYLG....TIGACYEFGSLPVILMFPELEKSFLKL.LIRHI.
SSO3039 SDYRKVGKYPYTGGPENTAL.N....TID.VLLYALPGVMLLFPELAKNIVRD.LSNRAL
ApGH116_aminoacid ...EPWGRFAVLECLDYRWY.E....SLD.VRLYGSFGLLMWWPELEKSVMRA.FAKAIA
BtGH116_aminoacid .....DGRLWTWEGCADNWG.SCHGSCTH.VWNYA.QAIPHLFPSLERSLRHTEFEEG..
TxGH116_aminoacid KDKRTNNMFGLLECFDYNYY.E....TLD.VRFYGSFPLVMLWPDIEKQVMRQ.FADTIN
570 580 590 600
Homo REDLTRRRYLMSGVMAP.....VKRRNVIPH.....DIGDPD.DEPWLRV
Pan REDLTRRRYLMSGVMAP.....VKRRNVIPH.....DIGDPD.DEPWLRV
Mus KEDLTRRRYLMSGVVAP.....VKRRNVIPH.....DIGDPD.DEPWLRV
Drosophila AELNDTRKMLYDGKVMP.....RKVKNCVPH.....DLGDPD.EEPFTLI
Caenorhabditis HSIEKPTRFHMEGDWAN.....VKTASRVPH.....DLGNPA.DEPWIAT
Arabidopsis MHDSSKKQVMSSGEFVT.....RKVLGAVPH.....DIGL...NDPWFEL
Sulfolobus .....REDGYVPH.....DLGYHSLDSPIDG.
SSO3039 KEDTPEYV.LFSL.AFPENLMKYKEEVKKDPTISTDLKKLYDTIKKIIKET.GKDPKGRM
ApGH116_aminoacid TDDDTPRIIGYNQASAI.....RKAAGATPH.....DLGAPN.EHPWEKT
BtGH116_aminoacid .QDLKGHQ.VFRV.NLPI.....
TxGH116_aminoacid VQDSSEFKVGSNGAMAV.....KKVQGMIPH.....DLGSSY.ALPWIKI
610 620 630
Homo NA....YLIHDTADWKDLNLKFVLQVYRDYYLTGD.....
Pan NA....YLIHDTADWKDLNLKFVLQVYRDYYLTGD.....
Mus NA....YLIHDTADWKDLNLKFVLQIYRDYYLTGD.....
Drosophila NC....YNIHDVNDWKDLNTKFVLQVYRDYYVLNELAQAQSDNASKFSSIEFIDKESLYE
Caenorhabditis NA....YVMHDTGKWKDLNMKYVLTSWRDYVVLSNE.....
Arabidopsis NA....YNLFNTDRWKDLNSKFVLQVYRDVVATGD.....
Sulfolobus .....TTSPPRWKDMNPSLILLVYRYFKFTND.....
SSO3039 PHYIRHSLTVDTYERVDINPEFVLLYYLIAKYTGD.....
ApGH116_aminoacid NY....TSYQDCNLWKDLPCDFVLQVYRDYLLTGAD.....
BtGH116_aminoacid .RPT...RHNFHSAADGQLGGIMKVYREWRISGE.....
TxGH116_aminoacid NA....YDWQNPNIWKDLNSKYVLLVYRDYVLTGKT.....
640 650
Homo .....QNFLKDMWPVCLAVM.ESEM.K
Pan .....QNFLKDMWPVCLAVM.ESEM.K
Mus .....QGFLEDMWPVCLAVM.ESEM.K
Drosophila LYSQDNKRKNSADEKQQNRKSASMYINETNGKVYLMDAIGYLKAMYASCKAIM.ERTI.E
Caenorhabditis .....HEDFLFHTWPAVRMIMLEALE.N
Arabidopsis .....LNFAKAVWPSVYTAI.AYLD.Q
Sulfolobus .....IEFLKEVYPILVKVM.DWEL.R
SSO3039 .....REFLTSVYEVARNAI.ESIM.R
ApGH116_aminoacid .....DIQFLVECWPAIVQTL.DYLK.T
BtGH116_aminoacid .....NEFLISMYPKVKKSL.DYCIST
TxGH116_aminoacid .....DKEFLKYTWKSVKTAL.DKLK.E


660
670
680
Homo FDKDHDGLI...ENGGYADQTYDGWVTTGPSAYC
Pan FDKDHDGLI...ENGGYADQTYDGWVTTGPSAYC
Mus FDKDQDGLI...ENGGYADQTYDAWVTTGPSAYC
Drosophila YDKDNDGLI...ENTKMPDQTYDSWVMDGPSAYC
Caenorhabditis WDQDGDGMI...ENFGKADQTYDAWQMEGVSAYC
Arabidopsis FDKDGDGMI...ENEGFPDQTYDAWSCSGVSAYC
Sulfolobus QCKGNLP...FMEGEMDNAFDATIIKGHDSYT
SSO3039 T.QTLDGLPYLTLPSGIEWMRHVNNMLKVSDAHKILGYHTLALSMQTLDDWSWLGFSPYV
ApGH116_aminoacid FDRDRDCIP...ENGGAPDQTFDDWRMMGISAYC
BtGH116_aminoacid WDPRRVGSI...EEPH..HNTYDI.EFWGPDGMH
TxGH116_aminoacid MDKDNDGIP...DNEGIPDQTYDTWSMKGTSAYC


690
700
710
720
730
Homo GGLWLAAVAVMVQMAALCGAQD.....IQDKFSSILSRGQEAYERLLWNGRYYN..
Pan GGLWLAAVAVMVQMAALCGAQD.....IQDKFSSILSRGQEAYERLLWNGRYYN..
Mus GGLWLAAVAVMVQMAVLCGAQD.....VQERFASILCRGREAYERLLWNGRYYN..
Drosophila SGLWLAALQAMSAMATILDQPN.....DCLRYQDILEKGKRSLEEKLWNGSYYR..
Caenorhabditis GSLWLASLRVAIEMAGLMKDDE.....TQQLFRNTLEKAKKVFIDTLWTGTYFR..
Arabidopsis GGLWVAALQAGSALAREIGDNG.....AAVYFNAKYEKARSVYE.KLWNGSYFN..
Sulfolobus SSLFIGSLIAMREIAKLVGDSN.....YVDFISEKLSSAREAFR.RMFNGRYFK..
SSO3039 SFLWIAALEALNEASKLLNNLH.....NY.EVKE..L..LTEKVNKYLWNGEYYIN.
ApGH116_aminoacid GGLWLAALEAAIAIANILLSHHSDITPDTTKAIATWTNWLETAKPLYDQTLWNGSYYN..
BtGH116_aminoacid NSFYYGALSAFIRMSEFLDK.D.....VT.EYKKLLKKGRKFTETGLFNGEYFIQK
TxGH116_aminoacid GSLWLAALKAAQEIGKVLKDNE.....AYIKYNEWYKIAQQNFEKELWNGEYYN..


740
750
760
Homo .....YDSSSR.....PQSRSVMSDQCACQWFLKACGLGE
Pan .....YDSSSR.....PQSRSVMSDQCACQWFLKACGLGE
Mus .....YDSSSH.....PQSRSIMSDQCACQWFLRACGLGE
Drosophila .....FDLSHS.....HRDTIMADQLCGHWYLKSCGFD.
Caenorhabditis .....FCERSR.....SRETVMADQLCGYWFLQSVSPE.
Arabidopsis .....YDNSRS.....GSSSSILADQMACQWYARACGLK.
Sulfolobus .....AWDSV.....DNAS.FLAQLYGEWFTTLVGLE.
SSO3039 .....WYDPISN.....LRDDSLNASQMTCDWYVQLLGLP.
ApGH116_aminoacid .....LDSQS.....GSDVVMADQLCCQFYAALLGLP.
BtGH116_aminoacid IEWRGLNAKDPTVAQSFHSSSYSPEAKEILEKEGPKYQYGNGCLSDGVLGSWLSRMCGME.
TxGH116_aminoacid .....FDTESD.....HKDSIMADQLACQWYADILRLG.


770
780
790
800
Homo GDTEVFPTQHVVRALQTIFELNVQA.....FAGGAM.GAVNG.MQP.....
Pan GDTEVFPTQHVVRALQTIFELNVQA.....FAGGAM.GAVNG.MQP.....
Mus GDTEVFPTLHVVRALQTIFELNVQA.....FAGGAM.GAVNG.MHP.....
Drosophila ..YEIYPKENVRTALKRIYDNNVMG.....FHEGNI.GAANG.FIANASEPTK
Caenorhabditis LADELLPNHMVRSALDTIYRLNVCK.....FGNGRM.GAVNG.MKP.....
Arabidopsis ...PIAKEEWIKKALETVYDFNVMR.....VRDGTR.GAVNG.MLP.....
Sulfolobus ...DIVEEDIIKKALESIIRLNGNA.....SP..H.CVPNL.VDD.....
SSO3039 ...EFLDHEKRKSIFSSIIKYNYTE.....EEGVKN.GSSNKEITP.....
ApGH116_aminoacid ...DIVPHHRARVALQTIYQACFQN.....FHNGKF.GAANG.VRT.....
BtGH116_aminoacid ...ETLNTEKVKSHLLSVHRYNFKKDLTDHANPQRSPYALGKEGGLLLG.SWP.....
TxGH116_aminoacid ...DILPKDHVQKALKKIYEFNVMK.....FENGKM.GAVNG.MRP.....


810 820 830 840 850 860
Homo HGVP.DKSSVQSDEVWVGVVYGLAATMIQEGLTWEGFQTAEGCYRTVWERLG..LAFQTP
Pan HGVP.DKSSVQSDEVWVGVVYGLAATMIQEGLTWEGFQTAEGCYRTVWERLG..LAFQTP
Mus HGVP.DRSSVQSDEVWVGVVYGLAATMIQEGLTWEGFRTAEGCYRTVWERLG..LAFQTP
Drosophila PGHV.DNSNIQAEEVWPGVVYALAATMIQEGMFEEAFQTAGGMYKTLSQRIG..MNFETP
Caenorhabditis SGVV.DREYIQADEMWTGVTYAVASLLIQQGEVEKAFHTASGSYLTCFEETG..LQYQTP
Arabidopsis DGRV.DTSTMVSREVWAGTTYSVAACMIQEGLADKGFRTASGIYEAAWSDRGLGCAFQTP
Sulfolobus NGKI.VGLSVQTYSSWPRMVFAICWLAYKKGVGDLSFCKKE..WDNLV.KNG..MVWDQP
SSO3039 LG.V..KLSVQSKTPWSGIEYYLASHMFYNGFDEYAKKILRNVYE....RYEL..AGNF.
ApGH116_aminoacid TGEPINPQDTHPLEVWTGINFGLGAFLIQMGMKEEAFQLAEAVITQVYQ.NG..LQFRTP
BtGH116_aminoacid KGGKLSLPFVYSNEVWTGIEYQVASHLMLQGEVEKGLEIVRACRQ....RYDG..SVRNP
TxGH116_aminoacid DGIV.DESDIQAQEVWTGVTYALASFMKYRGMTEEAYNTAYGVYKMTYDKSGKGYWFRTP
870 880 890
Homo EAYCQQRVF....RSLAYMRPLSIWAMQLALQQQQH.....
Pan EAYCQQRVF....RSLAYMRPLSIWAMQLALQQQQH.....
Mus EAYCQQQVF....RSLAYMRPLSIWAMQLALQQQQH.....
Drosophila EALYGEKRY....RSIGYMRPLSIWSMQVALERRR.....
Caenorhabditis EALYESKFY....RAIGYMRPLSIWAMQWSLKKHCGMNTSEEKLPNLVKNEKVVPESEAA
Arabidopsis EAWTTNDEY....RSLCYMRPLAIWGIQWAHTMPKPNRE.....
Sulfolobus SRINGYNGKAEMNYLDHYIGSPSPWS.....
SSO3039 W.....NHLK....WGARYMRPLVAINIVHAIEGIRVNLLNND.....
ApGH116_aminoacid EAITAAGTF....RASHYLRAMAIWAMYYQIQ.....
BtGH116_aminoacid F.....NEYE....CGHWYGRALSSYGLLQGLTGVRYDAVDKT.....
TxGH116_aminoacid EAWTKDGNY....RASMYMRPLSIWSMEVNY.....
900 910 920
Homo .....KKASW.PKVKQGTGLRTGPMFG.....PKEA
Pan .....KKASW.PKVEQGTGLRTGPMFG.....PKEA
Mus .....KKSRR.PSVTQGTGLSTQPECG.....PKRS
Drosophila .....
Caenorhabditis IQVQK.....QRILG.DTSEDGSSLGYVSDYH.....DERS
Arabidopsis ....Q.....EVSLR.PQEEDATSVLFQQHAGFIKVAHYLKNTKGKDHRNRLQTAYET
Sulfolobus .....
SSO3039 ....VNIDKR..KNLKWILLLPTAWGLLTI.NNGKIRMRIYHGKFKGK..I...NGIDVA
ApGH116_aminoacid .....
BtGH116_aminoacid ....LYINSKIGDFISF.ISTESGFGNIEL.RSGKPFVKVVSG....H..I...EVDRF.
TxGH116_aminoacid .....
Homo M.A.NLSPE
Pan M.A.NLSPE
Mus L.A.NLNSE
Drosophila .....AQRD
Caenorhabditis I.PKIATLR
Arabidopsis F.L.KTIRL
Sulfolobus ......FLF
SSO3039 LRENEEIEF
ApGH116_aminoacid .....HP
BtGH116_aminoacid V.VSGKVVE
TxGH116_aminoacid ......NEV

A multi-alignment of various GH116 β-glucosidase protein sequences. Invariant residues are highlighted in red. Areas of high conservation are boxed in blue. Identified catalytic residues are marked with a star. Homo is NP_065995.1 from *Homo sapiens.* Pan is XP_001167923.1 from *Pan troglodytes.* Mus is CAM17042.1 from *Mus musculus.* Drosophila is AA041192.2 from *Drosophila melanogaster.* Caenorhabditis is CAD21661.1 from *C. elegans* Bristol N2. Arabidopsis is BAE.99039.1 from *A. thaliana.* Sulfolobus is SSO1353 from *S. solfataricus* P2. SSO3039 is SSO3039 from *S. solfataricus* P2. ApGH116 is BAI91082.1 from *A.platensis* NIES_39. BtGH116 is AA078014.1 from *B. thetaiotaomicron* VP1-5482. TxGH116 is AEF18218.1 from *T. xylanolyticum.* Alignment made using the T-Coffee server (tcoffee.crg.cat) and formatted using Espript (espript.ibcp.fr) (187,188).

Appendix 2. Bacterial GH116 sequences

These are the full length genes from of each of the bacterial GH116 enzyme that was cloned into pET28a between NdeI and XhoI restriction sites. The full construct includes N-terminal His_6-tag and vector-added residues; they are codon optimised for E coli.

A.1.1 Arthrospira platensis (ApGH116) nucleotide

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCA
GTATCCGTCACCGACCCTGAACATCCCGCCGTACACCTGGAACCGTCCGATTGGCCTGGGCTGGG
AAAAACCGTACACCGTCCGCTATCCGTCCAATCTGGATGACGGTCCGTTTCATGGTATGCCGCTG
GGCGGTTTCGGTGCAGGTTGCATTGGTCGTTCACACCGCGGCGATTTTAACCTGTGGCATCTGG
ACGGCGGTGAACACATCTTCCGTCCGCTGCCGGGTTGTCAGTTTTCTGTGTTCGAAGAAATTGA
TGGCAAACGCCAAGCGTATGCGCTGAGTACGCAGCCGCCGGAAGATGGTAGTCTGTCCACCTGG
AATTGGTACCCGAGCTCTGACCAGCATACCCAAACGGGCACCTATCACGCGCTGTACCCGCGTTC
TTGGTTTGTGTATGAAAACGTTTTCCAGACGCAACTGACCTGCGAACAATTTTCCCCGGTTTGG
GGCGGTAATTATCAGGAAACCTCATACCCGGTTGCAGTCTTCGAATGGATTGCTCATAACCCGA
CCGATAAACTGATTATCCTGAGTATTCTGCTGACGTGGGAAAACACCATCGGTTGGTTTACGAA
TCGTCTGGATACCCCGGCCGTGAAAGTTCGCGATGACGGTTCGCCGGTCTATGAATACCAACCGC
GCTGGGGCGATAGCCAGGACAACTGCAATCATTTTGTTGAAGATTTCCACCGTGTCGGCTGTGT
GATGACCCGCCTGAACATCAATGATCAACTGGGCGAAGGTGACGGCCAGATGGCGATTGCCACG
ATCACCAATCCGGTGGTTGAAGTGTATTGCCATAACCGTTGGCGCCCGTCTGGTAATGGCTATG
ATATTTGGGAATACTTTTCCCAGGATGGTTCACTGATTGACATCCATGATGACGGCCCGGCGCT
GGAAGGTGAACGTATCGGCGCGGCCCTGGCCGTTCGTTGTATTCTGCGCCCGGGTAAAACCCGC
AAAATTCCGTTTTTCATCGCATGGGATCTGCCGGTTACCGAATTTGAAGCTGGCGTCACGTATT
ACCGTCATTATACCGACTTTTTCGGTCGTAACGGCCGCAATGCGTGGTCTATGATCCGTACGGCC
ATGAAACACAGTGATACCTGGCGCGAAAATATTGAAGCGTGGCAAAACCCGATCCTGAATCGTC
AGGACCTGCCGCCGTGGTTTAAAATGGCCCTGTTCAACGAACTGTATCTGCTGACGGATGGCGG
TTCACTGTGGACCGCAATTGACGAAGATGAACCGTGGGGTCGCTTTGCTGTTCTGGAATGCCTG
GATTATCGTTGGTACGAATCGCTGGACGTCCGCCTGTATGGTAGCTTTGGCCTGCTGATGTGGT
GGCCGGAACTGGAAAAATCGGTGATGCGCGCATTCGCTAAAGCGATCGCCACGGATGACGATAC
CCCGCGTATTATCGGCTACAATCAGGCAAGCGCTATTCGCAAAGCAGCTGGTGCAACCCCGCAT
GATCTGGGCGCTCCGAACGAACACCCGTGGGAAAAAACGAATTATACCAGTTACCAGGATTGCA
ACCTGTGGAAAGATCTGCCGTGTGACTTTGTTCTGCAGGTCTATCGTGATTACCTGCTGACCGG
TGCAGACGATATCCAATTCCTGGTGGAATGCTGGCCGGCGATTGTTCAGACGCTGGATTATCTG
AAAACCTTTGACCGTGATCGCGACTGTATCCCGGAAAACGGCGGTGCCCCGGATCAGACCTTCG
ACGATTGGCGCATGATGGGCATTTCCGCATACTGTGGCGGTCTGTGGCTGGCAGCACTGGAAGC
GCTATTGCGATCGCCAATATTCTGCTGTCTCATCACAGTGATATCACCCCGGACACC

ACGAAAGCAATTGCTACGTGGACCAACTGGCTGGAAACGGCAAAACCGCTGTATGATCAGACCC
TGTGGAACGGTAGCTATTACAATCTGGACTCACAGTCGGGCAGCGATGTCGTGATGGCCGACCA
GCTGTGCGGTCAATTTTATGCAGCACTGCTGGGTCTGCCGGATATCGTCCCGCATCACCGTGCGC
GCGTGGCACTGCAGACCATTTACCAAGCGTGTTTTCAGAACTTCCATAATGGTAAATTCGGCGC
AGCTAATGGTGTGCGTACCACGGGCGAACCGATTAACCCGCAGGATACGCACCCGCTGGAAGTT
TGGACCGGTATCAATTTTGGTCTGGGCGCGTTCCTGATTCAAATGGGCATGAAAGAAGAAGCCT
TTCAGCTGGCGGAAGCCGTCATCACGCAGGTGTATCAAAACGGTCTGCAGTTTCGTACCCCGGA
AGCAATTACGGCAGCAGGTACCTTCCGTGCAAGCCATTACCTGCGCGCAATGGCTATTTGGGCG
ATGTATTACCAGATCCAACACCCGTAA

A.1.2 ApGH116 amino acid

MGSSHHHHHHSSGLVPRGSHMQYPSPTLNIPPYTWNRPIGLGWEKPYTVRYPSNLDDGPFHGM
PLGGFGAGCIGRSHRGDFNLWHLDGGEHIFRPLPGCQFSVFEEIDGKRQAYALSTQPPEDGSLST
WNWYPSSDQHTQTGTYHALYPRSWFVYENVFQTQLTCEQFSPVWGGNYQETSYPVAVFEWIA
HNPTDKLIILSILLTWENTIGWFTNRLDTPAVKVRDDGSPVYEYQPRWGDSQDNCNHFVEDFHR
VGCVMTRLNINDQLGEGDGQMAIATITNPVVEVYCHNRWRPSGNGYDIWEYFSQDGSLIDIHDD
GPALEGERIGAALAVRCILRPGKTRKIPFFIAWDLPVTEFEAGVTYYRHYTDFFGRNGRNAWSMI
RTAMKHSDTWRENIEAWQNPILNRQDLPPWFKMALFNELYLLTDGGSLWTAIDEDEPWGRFA
VLECLDYRWYESLDVRLYGSFGLLMWWPELEKSVMRAFAKAIATDDDTPRIIGYNQASAIRKAA
GATPHDLGAPNEHPWEKTNYTSYQDCNLWKDLPCDFVLQVYRDYLLTGADDIQFLVECWPAIV
QTLDYLKTFDRDRDCIPENGGAPDQTFDDWRMMGISAYCGGLWLAALEAAIAIANILLSHHSDIT
PDTTKAIATWTNWLETAKPLYDQTLWNGSYYNLDSQSGSDVVMADQLCGQFYAALLGLPDIVP
HHRARVALQTIYQACFQNFHNGKFGAANGVRTTGEPINPQDTHPLEVWTGINFGLGAFLIQMGM
KEEAFQLAEAVITQVYQNGLQFRTPEAITAAGTFRASHYLRAMAIWAMYYQIQHP

A.1.3 Bacteroides thetaiotaomicron (BtGH116) nucleotide

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGA
ACGACATTGACGACGAACGTAACCGCATCTACAACAGCTCCTACAGTGGTAAATACAACAACCG
ATTGCTTTTCCGATTGGTGGTATTGGCACCGGCATGTATTGCCTGGAAGGCACGGGTTACATCT
CACACATGTCGGTGTGGCACCGCCCGGAAGTTTTTCATGAACCGGGCATGTTCGCGGCCCTGTAT
GTTAAAGGCGTCTGTAACGGTGCAAAAGTGCTGGAAGGTCCGGTTAGCGATTGGCGTAAATTCG
GCATGCCGAATTACGGTACCGGCGGTAGCATGGGCTCTATTCTGGGTCTGCCGCGTTTTGACAC
GGTTGAATTCGAAGCGCGCTTTCCGTTCGCCAAAGTCAGTCTGACCGATAAAGACATTCCGGTG
AAAGTTACGATCCTGGGCTGGTCCCCGTTTATTCCGGGTGATCCGGACAACAGCTCTCTGCCGGT
GGGCGGTCTGGAATATAGCCTGGAAAATACCTCTAAAGAAGTTCAGGAAACGATCTTTTCATAC
CACGCACGCAACTTCCTGTCGTGGGGCAAAGGTCTGGATGCTATTAAAACCATGCCGCATGGCT
TTATCCTGAGTCAATCCGGTACCGAAACGGAACCGCACCTGCAGGGCGATTTTGCCATTTTCACC

ATCAAGACAGCCTGAAAATCAACTATTGCTGGTTTCGTGGCGGTTGGTTCGACTCTCTGACGAT
GGTGTGGAATGCAATTGAAGCTGGTCTGATGCCGCAGTGTCCGGCAATCGAAAAAGGTGCACCG
GGTGCAAGCATGTTTGTCCCGGTGACCCTGATGCCGGGCGAAAAGAAAACCATTCGTATCTATA
CGGCATGGTATGTGCCGAACTCAACGCTGCGCCTGGGTGAAGAACCGGAAGATTGGAATGACAA
CAATGTCGATTCGGCACGTCTGGCTGTGGAAAAAGCGGACAAAGGCAACTATAAACCGTGGTAC
AGTTCCCGCTTTACCGGTGTGAACGAAGTTATCGATTATTTCCTGTCTCATTACAAAATCCTGC
GTAACCAGACCGAACGCTTTACGGACTCTTTCTATCGCAGTACCCTGCCGCCGGAAGTTATTGA
AGCGGTCTCCGCCAATCTGTCAATCCTGAAATCGCCGACCGTGATGCGTCAGTATGATGGCCGCC
TGTGGACGTGGGAAGGTTGCGCAGACAACTGGGGCAGCTGCCATGGTTCTTGTACCCACGTTTG
GAATTACGCGCAAGCCATTCCGCACCTGTTTCCGAGTCTGGAACGTTCCCTGCGCCACACGGAAT
TCGAAGAAGGCCAGGATCTGAAAGGTCACCAAGTCTTTCGTGTGAACCTGCCGATTCGTCCGAC
CCGCCATAATTTCCACAGCGCAGCTGATGGTCAGCTGGGCGGTATCATGAAAGTGTATCGTGAA
TGGCGCATTAGCGGTGAAAACGAATTTCTGATCTCTATGTATCCGAAAGTGAAAAAATCTCTGG
ATTACTGTATCTCCACCTGGGACCCGCGTCGCGTTGGCTCAATTGAAGAACCGCATCACAACACG
TATGATATCGAATTTTGGGGCCCGGACGGTATGCATAATTCCTTCTATTACGGTGCACTGTCAG
CTTTTATCCGTATGTCGGAATTCCTGGATAAAGACGTGACCGAATATAAAAAACTGCTGAAAAA
AGGCCGCAAATTCACCGAAACGGGCCTGTTTAACGGTGAATACTTCATTCAGAAAATCGAATGG
CGTGGTCTGAATGCAAAAGATCCGACCGTGGCTCAAAGCTTTCACTCATCGTATTCTCCGGAAG
CGAAAGAAATTCTGGAAAAAGAAGGCCCGAAATATCAGTACGGCAACGGTTGCCTGAGCGATGG
CGTTCTGGGTAGTTGGCTGTCCCGCATGTGTGGTATGGAAGAAACCCTGAATACGGAAAAAGTT
AAATCACATCTGCTGTCGGTCCACCGTTATAACTTCAAAAAAGATCTGACCGACCATGCGAATC
CGCAGCGCTCACCGTACGCCCTGGGCAAAGAAGGCGGTCTGCTGCTGGGTTCGTGGCCGAAAGG
CGGTAAACTGAGTCTGCCGTTTGTCTATTCCAATGAAGTGTGGACCGGCATTGAATACCAGGTG
GCAAGTCATCTGATGCTGCAAGGCGAAGTGGAAAAAGGTCTGGAAATCGTTCGTGCTTGCCGTC
AGCGCTATGATGGCAGCGTTCGCAACCCGTTTAATGAATATGAATGTGGCCACTGGTACGGTCG
TGCGCTGAGCTCTTACGGTCTGCTGCAAGGCCTGACCGGTGTTCGCTATGATGCCGTCGACAAA
ACGCTGTACATCAACTCTAAAATCGGCGATTTCATCTCATTCATCTCGACCGAAAGCGGCTTTG
GTAATATTGAACTGCGTAGTGGCAAACCGTTCGTCAAAGTGGTTTCCGGTCATATCGAAGTGGA
TCGCTTTGTCGTGAGTGGTAAAGTTGTCGAATAA

A.1.4 BtGH116 amino acid

MGSSHHHHHHSSGLVPRGSHMNDIDDERNRIYNSSYSGKYNNRIAFPIGGIGTGMYCLEGTGYIS
HMSVWHRPEVFHEPGMFAALYVKGVCNGAKVLEGPVSDWRKFGMPNYGTGGSMGSILGLPRF
DTVEFEARFPFAKVSLTDKDIPVKVTILGWSPFIPGDPDNSSLPVGGLEYSLENTSKEVQETIFSYH
ARNFLSWGKGLDAIKTMPHGFILSQSGTETEPHLQGDFAIFTDQDSLKINYCWFRGGWFDSLTM
VWNAIEAGLMPQCPAIEKGAPGASMFVPVTLMPGEKKTIRIYTAWYVPNSTLRLGEEPEDWND
NNVDSARLAVEKADKGNYKPWYSSRFTGVNEVIDYFLSHYKILRNQTERFTDSFYRSTLPPEVIE

AVSANLSILKSPTVMRQYDGRLWTWEGCADNWGSCHGSCTHVWNYAQAIPHLFPSLERSLRHT
FEEGQDLKGHQVFRVNLPIRPTRHNFHSAADGQLGGIMKVYREWRISGENEFLISMYPKVKKSLD
CISTWDPRRVGSIEEPHHNTYDIEFWGPDGMHNSFYYGALSAFIRMSEFLDKDVTEYKKLLKKGR
KFTETGLFNGEYFIQKIEWRGLNAKDPTVAQSFHSSYSPEAKEILEKEGPKYQYGNGCLSDGVLGS
WLSRMCGMEETLNTEKVKSHLLSVHRYNFKKDLTDHANPQRSPYALGKEGGLLLGSWPKGGKL
SLPFVYSNEVWTGIEYQVASHLMLQGEVEKGLEIVRACRQRYDGSVRNPFNEYECGHWYGRALS
SYGLLQGLTGVRYDAVDKTLYINSKIGDFISFISTESGFGNIELRSGKPFVKVVSGHIEVDRFVVSGK
VVE

A.1.5 Thermoanaerobacterium xylanolyticum (TxGH116) nucleotide

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGG
GCTGCTCAGAAAAAATCAACATCAATGAAGACAAAATCTCGCACAAAATCGACATCCCGGACTC
GGCCTGGACCATCGGTATCGGTGAAAAATTTAAAAACGCCGGTCATCCGAATGTTAAATACCCG
ATGATTGATGACAGCTATGTCCAGGGTGCACCGCTGGGCGGTTTTGGTGCAGGCACCATTGGCC
GTACGTATAACGGCGGTTTCTCCCGCTGGCACCTGGAAATCGGCAAAAACAAATACACCACGGT
GTATGCAAATCAGTTTTCAGTCTTCCAAAAAGTGGAGGGTAACAAAGATGGCGTGGCGCAGGTT
CTGTACGCCGGTGAACCGGAAAATGGCTATCTGAGCTCTTGGAAATGGGATTACCCGAAAGAAT
CTGGCATGTATTACGCGCTGTATCCGAACAGTTGGTATACCTACACGAATAAAGATCTGCCGGT
TCAGCTGGCCGTCAAACAATTTTCCCCGATTATCCCGTATAATTACAAAGAAACGTCATACCCG
GTGGCAGTTTTCAAATGGACCGCTTATAACCCGACGAACAAAAATGTCGATGTGTCGATTATGT
TTACCTGGCAGAATATGATCGGCTTTTTCGGCAAACAAGTTAACGTCAATAGCGGTAACTTCAA
CAAAATCATCAAAGACAAAAGCAAAGATTCTGAAATTGTCGCGGCCGTGATGGGCAACATCTCG
AACGATAATGAAGAATGGAATGGTGAATACAGCATTGGCGTTAAAAAAGTCCCGGGTGTGGAT
ATCTCCTATAAAGCGAAATTTGTGACCACGGGTGACGGCTCGGATCTGTGGCATGAATTCAGCA
AAAACGGCATTCTGGACAATAAAGATGACGAAACCCCGACGAAACAGGATGGTATTGGCTCTGC
AATCGCTGTTAACTTCAAACTGCAGCCGGGCCAAACCATTGAAGTGCCGTTTGCCCTGAGTTGG
GACCTGCCGATCATGAAATTCGGCGGTGGCGATAAATGGTACAAAATGTACACCAAATACTTCG
GTAAAAACGGCAAAAACTCTTTCGCAATCCTGAAAGAAGCTCTGAACAACTACCAGAAATGGGA
AAAAATGATCGATGACTGGCAAAAACCGATCCTGAGTAACAAATCCAAACCGGATTGGTATAAA
ACCGCGCTGTTTAATGAACTGTATTACCTGGCAGACGGTGGCACGGCATGGGAAAACGGTAAAG
TTGGCGAAAAAGATAAACGTACCAACAACATGTTCGGTCTGCTGGAATGCTTCGACTACAATTA
TTACGAAACCCTGGATGTGCGTTTTTATGGCAGTTTCCCGCTGGTTATGCTGTGGCCGGATATT
GAAAAACAGGTGATGCGCCAATTTGCCGACACCATCAACGTTCAGGATAGTTCCGAATTCAAAG
TCGGTAGCAATGGCGCGATGGCCGTTAAAAAAGTCCAGGGTATGATTCCGCACGATCTGGGCTC
ATCGTACGCACTGCCGTGGATTAAAATCAACGCTTATGACTGGCAGAACCCGAATATCTGGAAA
GATCTGAATTCGAAATACGTGCTGCTGGTTTACCGCGACTATGTTCTGACCGGTAAAACGGATA
AAGAATTCCTGAAATACACCTGGAAATCTGTGAAAACGGCACTGGACAAACTGAAAGAAATGG

ATAAAGACAACGATGGTATTCCGGATAATGAAGGCATCCCGGACCAGACCTACGATACGTGGAG
CATGAAAGGTACCTCTGCGTATTGTGGCAGTCTGTGGCTGGCAGCTCTGAAAGCGGCCCAGGAA
ATTGGTAAAGTTCTGAAAGATAACGAAGCATACATCAAATACAACGAATGGTACAAAATCGCTC
AGCAAAACTTCGAAAAAGAACTGTGGAACGGCGAATATTACAATTTCGATACCGAATCCGACCA
TAAAGATTCAATTATGGCGGACCAGCTGGCCGGTCAATGGTATGCGGATATTCTGCGTCTGGGC
GACATCCTGCCGAAAGATCACGTTCAGAAAGCCCTGAAGAAAATTTACGAATTCAACGTGATGA
AATTCGAAAACGGTAAAATGGGCGCAGTCAATGGTATGCGCCCGGATGGCATTGTGGACGAATC
TGATATCCAGGCTCAAGAAGTGTGGACCGGTGTTACGTATGCACTGGCTAGTTTCATGAAATAC
CGTGGTATGACCGAAGAAGCATATAACACGGCTTACGGCGTGTATAAAATGACCTACGACAAAT
CCGGTAAAGGCTATTGGTTTCGCACCCCGGAAGCGTGGACGAAAGATGGCAATTACCGTGCCTC
AATGTATATGCGCCCGCTGTCAATTTGGTCGATGGAAGTCAACTATAATGAAGTGTAA

A.1.6TxGH116 amino acid

MGSSHHHHHHSSGLVPRGSHMGCSEKININEDKISHKIDIPDSAWTIGIGEKFKNAGHPNVKYPM
IDDSYVQGAPLGGFGAGTIGRTYNGGFSRWHLEIGKNKYTTVYANQFSVFQKVEGNKDGVAQVL
YAGEPENGYLSSWKWDYPKESGMYYALYPNSWYTYTNKDLPVQLAVKQFSPIIPYNYKETSYPV
AVFKWTAYNPTNKNVDVSIMFTWQNMIGFFGKQVNVNSGNFNKIIKDKSKDSEIVAAVMGNISN
DNEEWNGEYSIGVKKVPGVDISYKAKFVTTGDGSDLWHEFSKNGILDNKDDETPTKQDGIGSAIA
VNFKLQPGQTIEVPFALSWDLPIMKFGGGDKWYKMYTKYFGKNGKNSFAILKEALNNYQKWEK
MIDDWQKPILSNKSKPDWYKTALFNELYYLADGGTAWENGKVGEKDKRTNNMFGLLECFDYN
YYETLDVRFYGSFPLVMLWPDIEKQVMRQFADTINVQDSSEFKVGSNGAMAVKKVQGMIPHDLG
SSYALPWIKINAYDWQNPNIWKDLNSKYVLLVYRDYVLTGKTDKEFLKYTWKSVKTALDKLKE
MDKDNDGIPDNEGIPDQTYDTWSMKGTSAYCGSLWLAALKAAQEIGKVLKDNEAYIKYNEWYK
IAQQNFEKELWNGEYYNFDTESDHKDSIMADQLAGQWYADILRLGDILPKDHVQKALKKIYEFN
VMKFENGKMGAVNGMRPDGIVDESDIQAQEVWTGVTYALASFMKYRGMTEEAYNTAYGVYKM
TYDKSGKGYWFRTPEAWTKDGNYRASMYMRPLSIWSMEVNYNEV

Appendix 3. TxGH116C protein sequence

ThxGH11C-pET30(+) was kindly supplied by James R. Ketudat Cairns, Suranaree University of Technology, Institute of Science, Protein Structure, Function and Application, Thailand. The construct encoded the TxGH116 protein lacking the first 18 amino acids following the pET30(+) N-terminal tag including a His_6 tag, S-tag and enterokinase site. The fusion protein also included the C-terminal His_s-Tag from the plasmid.

TxGH116C amino acid

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK(***)AMALTGCSEKININED
KISHKIDIPDSAWTIGIGEKFKNAGHPNVKYPMIDDSYVQGAPLGGFGAGTIGRTYNGGFSRWHL
EIGKNKYTTVYANQFSVFQKVEGNKDGVAQVLYAGEPENGYLSSWKWDYPKESGMYYALYPNS
WYTYTNKDLPVQLAVKQFSPIIPYNYKETSYPVAVFKWTAYNPTNKNVDVSIMFTWQNMIGFFG
KQVNVNSGNFNKIIKDKSKDSEIVAAVMGNISNDNEEWNGEYSIGVKKVPGVDISYKAKFVTTGD
GSDLWHEFSKNGILDNKDDETPTKQDGIGSAIAVNFKLQPGQTIEVPFALSWDLPIMKFGGGDK
WYKMYTKYFGKNGKNSFAILKEALNNYQKWEKMIDDWQKPILSNKSKPDWYKTALFNELYYL
ADGGTAWENGKVGEKDKRTNNMFGLLECFDYNYYETLDVRFYGSFPLVMLWPDIEKQVMRQF
ADTINVQDSSEFKVGSNGAMAVKKVQGMIPHDLGSSYALPWIKINAYDWQNPNIWKDLNSKYVL
LVYRDYVLTGKTDKEFLKYTWKSVKTALDKLKEMDKDNDGIPDNEGIPDQTYDTWSMKGTSAY
CGSLWLAALKAAQEIGKVLKDNEAYIKYNEWYKIAQQNFEKELWNGEYYNFDTESDHKDSIMA
DQLAGQWYADILRLGDILPKDHVQKALKKIYEFNVMKFENGKMGAVNGMRPDGIVDESDIQAQ
EVWTGVTYALASFMKYRGMTEEAYNTAYGVYKMTYDKSGKGYWFRTPEAWTKDGNYRASMY
MRPLSIWSMEVNYNEVLEHHHHHH

(***) designates the enterokinase cleavage site.

Appendix 4.

ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2)

Ratana Charoenwattanasatien,[†,‡,§,+] Salila Pengthaisong,[‡,§,+] Imogen Breen,[∥] Risa Mutoh,[†] Sompong Sansenya,[⊥] Yanling Hua,[§,#] Anupong Tankrathok,[▽] Liang Wu,[∥] Chomphunuch Songsiriritthigul,[§,○] Hideaki Tanaka,[†] Spencer J. Williams,[◆] Gideon J. Davies,[*,∥] Genji Kurisu,[*,†] and James R. Ketudat Cairns[*,‡,§,¶]

[†]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

[‡]School of Biochemistry, Institute of Science, Suranaree Univerity of Technology, Nakhon Ratchasima 30000, Thailand

[§]Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

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

[⊥]Department of Chemistry, Faculty of Science, Rajamangala University of Technology, Thanyaburi, Pathum Thani 12110, Thailand

[#]Center for Scientific and Technological Equipment, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

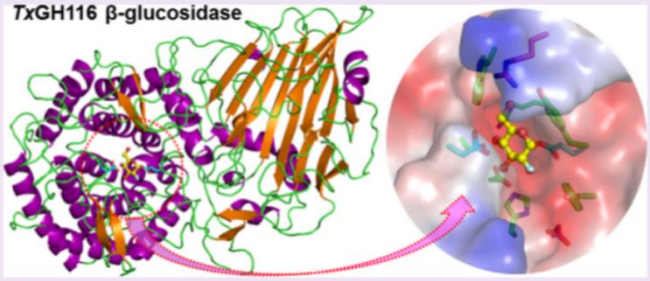
[▽]Department of Biotechnology, Faculty of Agro-Industrial Technology, Kalasin University, Kalasin 46000, Thailand

[○]Synchrotron Light Research Institute, Nakhon Ratchasima 30000, Thailand

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

[¶]Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok 10210, Thailand

Supporting Information



ABSTRACT: Human glucosylcerebrosidase 2 (GBA2) of the CAZy family GH116 is responsible for the breakdown of glycosphingolipids on the cytoplasmic face of the endoplasmic reticulum and Golgi apparatus. Genetic defects in GBA2 result in spastic paraplegia and cerebellar ataxia, while cross-talk between GBA2 and GBA1 glucosylceramidases may affect Gaucher disease. Here, we report the first three-dimensional structure for any GH116 enzyme, *Thermoanaerobacterium xylanolyticum* *Tx*GH116 β-glucosidase, alone and in complex with diverse ligands. These structures allow identification of the glucoside binding and active site residues, which are shown to be conserved with GBA2. Mutagenic analysis of *Tx*GH116 and structural modeling of GBA2 provide a detailed structural and functional rationale for pathogenic missense mutations of GBA2.

Human glucosylcerebrosidase 2 (GBA2, nonlysosomal glucosylcerebrosidase, E.C. 3.2.1.45, bile acid β-glucosidase E.C. 3.2.1.21) is a member of a family of sequence-related glycoside hydrolases (GH) designated GH116 (see www.cazy.org), with members in animals, plants, fungi, archaea, and eubacteria.[1−3] Originally identified as a bile acid β-glucosidase,[4] GBA2 was later found to hydrolyze the glycolipids glucosylceramide (GlcCer) and glucosyllysosphingosine, and its defect was found to disrupt glycolipids in sperm and cause infertility in knockout mice.[5]

GBA2 is a peripheral membrane protein localized to the cytoplasmic face of the endoplasmic reticulum and/or Golgi

Received: February 29, 2016
Accepted: April 26, 2016
Published: April 26, 2016

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

apparatus membrane.[6] GBA2 and the lysosomal family GH30 glucosylcerebrosidase GBA1 (GBA, acid β-glucosidase) are thought to be the primary enzymes responsible for the breakdown of GlcCer.[7] GlcCer is the most abundant glycosylceramide, a common precursor for the majority of more complex glycosphingolipids, and an important messenger in its own right.[8,9] While there is extensive structural and functional knowledge of GBA1, genetic deficiency of which results in the lysosomal storage disease Gaucher disease,[10,11] much less is known of GBA2.

Mutations in the human *GBA2* gene cause hereditary spastic paraplegia (HSP) and autosomal recessive cerebellar ataxia (ARCA).[12−15] Zebrafish with the corresponding gene knockout also show neurological defects.[12] Significantly, knockout of GBA2 in mouse models for Gaucher disease and Niemann-Pick disease, in which there is a secondary deficiency in GBA1, partially alleviated symptoms.[16,17] Furthermore, overexpression of GBA2 was found to be toxic to melanoma cells.[18] Taken together, this suggests that specific inhibitors and activators of GBA2 could be useful for the treatment of Gaucher disease, hereditary ataxias, and possibly melanoma. Although GBA2 inhibitors have been identified,[7,19,20] the lack of any structural insight for GBA2, or indeed any other GH116 protein, has limited both the rational development of more potent inhibitors and the molecular understanding of pathogenic mutations.

Glycoside hydrolase family GH116 includes enzymes with diverse specificities, including a β-xylosidase/β-glucosidase[3] and an *N*-acetylglucosaminidase,[21] the biochemical analysis of which has been useful in defining the basic catalytic machinery, including the general acid/base and catalytic nucleophile. However, these biochemically characterized members are divergent from mammalian GBA2 and lack conservation of residues that have been identified to be mutated in GBA2-related disorders. Here, we describe the three-dimensional structure of a bacterial homologue of the human GBA2, the first 3-D structure for this important family of glycosidases, in native and ligand complexes. Through functional analysis of this highly tractable model, in which all pathogenic GBA2 mutation sites are strictly conserved, we provide detailed insight into the molecular bases of human GBA2 mutations leading to ataxias and paraplegias.

RESULTS AND DISCUSSION

Enzyme Activity and Small Molecule Inhibition. Given the absence of any 3-D model or structure for the human enzyme, we sought to work on a close bacterial homologue of GBA2. *Thermoanaerobacterium xylanolyticum* GH116 (*Tx*GH116) shares approximately 40% sequence identity over the catalytic domain, and we predicted it would share glucose-specificity and identity in the glucose-binding residues in the active site. Recombinant *Tx*GH116 is indeed active on 4-nitrophenyl (4NP) β-D-glucoside (4NPGlc) and other glucosides (Table 1). Michaelis Menten kinetics, determined at 60 °C (the temperature of the Frying Pan Hot Springs, from which *T. xylanolyticum* was isolated[22]), revealed that 4NPGlc is hydrolyzed 17 times as efficiently, in terms of k_{cat}/K_M, than the corresponding β-D-galactoside, 4NPGal. *Tx*GH116 has a pH optimum of 5.5 and shows high activity toward both β-1,3- and β-1,4-linked glucooligosaccharides, with similar k_{cat} and K_M values for both cellobiose and laminaribiose (Table 2). Since *T. xylanolyticum* was isolated based on its growth on plant cell wall polysaccharides,[22] these oligosaccharides are likely natural substrates for the enzyme. Although we were unable to detect hydrolysis of glucosylceramide, *Tx*GH116 does hydrolyze alkyl glycosides including n-heptyl glucoside and n-octyl glucoside. With knowledge that the enzyme was a β-glucosidase, and given its high sequence similarity to GBA2, we were encouraged to determine the three-dimensional structure of *Tx*GH116, as a model for family GH116.

Table 1. Specific Activity of *Tx*GH116 and Catalytic Residue Mutants with Aryl and Alkyl Glycosides and Oligosaccharides

	specific activity (μmol mg^{-1} min^{-1})[a,b]		
substrate	*Tx*GH116 wild type	D593A	E441A
4NP β-D-glucoside	27.3 (100%)	0.0772 (100%)	0.000528
4NP β-D-galactoside	8.90 (33%)	-[c]	-[c]
4NP *N*-acetyl-β-D-glucosaminide	0.123 (0.45%)	-	-
4NP α-D-galactoside	< 0.01 (<0.04%)	-	-
4NP α-L-arabinoside	< 0.01 (<0.04%)	-	-
4NP β-D-fucoside	< 0.01 (<0.04%)	-	-
4NP α-D-glucoside	< 0.01 (<0.04%)	-	-
4NP β-D-mannoside	< 0.01 (<0.04%)	-	-
4NP β-D-xyloside	< 0.01 (<0.04%)	-	-
4NP β-cellobioside	0.302 (1.1%)	-	-
gentiobiose	2.23 (8.2%)	-	-
sophorose	4.10 (15%)	-	-
laminaribiose	11.7 (43%)	0.000245 (0.32%)	-
laminaritriose	16.1 (59%)	0.000478 (0.62%)	-
laminaritetraose	15.0 (55%)	0.000490 (0.62%)	-
laminaripentaose	16.5 (60%)	0.000524 (0.68%)	-
cellobiose	13.4 (49%)	0.000345 (0.45%)	-
cellotriose	15.6 (57%)	0.000538 (0.70%)	-
cellotetraose	14.9 (55%)	0.000498 (0.65%)	-
cellopentaose	13.8 (51%)	0.000448 (0.58%)	-
cellohexaose	13.5 (49%)	0.000422 (0.55%)	-
methyl β-D-glucoside	0.0817 (0.30%)	-	-
n-heptyl-β-D-glucoside	3.74 (14%)	-	-
n-octyl β-D-glucoside	3.58 (13%)	-	-

[a]The assay contained 1 mM substrate in 50 mM MES buffer, pH 5.5, at 60 °C. [b]Percent activity relative to glucose or 4NP released from 4NP β-D-glucopyranoside is given in parentheses. The values for disaccharide hydrolysis were determined by dividing the amount of glucose released by 2, since two glucose molecules are released per glycosidic bond hydrolyzed. For oligosaccharides, which may release more than one glucose molecule per substrate molecule due to sequential cleavage, the values quoted are for total glucose released. [c]The dashes mean that the values were not determined for the mutant enzymes due to low activity.

Three-Dimensional Structures of *Tx*GH116. *Tx*GH116 crystallized in space groups $P2_12_12$ and $P6_1$, and its structure was solved and refined at 1.61 Å (Table 3). The absence of

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

Table 2. Kinetic Parameters of *Tx*GH116 and Its Mutants for Hydrolysis of 4NP-Glycosides and Oligosaccharides

		kinetic parameters (60 °C)			temperature	
protein	substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})	optimum	T_m
WT without N-term tag[a]	4NP β-D-glucoside	0.18 ± 0.008	49.0 ± 0.8	272	75 °C	-
	4NP β-D-galactoside	16.3 ± 0.94	255 ± 5	15.6		
	cellobiose	0.25 ± 0.016	44.4 ± 0.7	178		
	laminaribiose	0.27 ± 0.018	41.4 ± 0.7	153		
WT with N-term tag[b]	4NP β-D-glucoside	0.21 ± 0.012	37.6 ± 0.6	179	75 °C	81.3 °C
D508H	4NP β-D-glucoside	110 ± 3	3.47 ± 0.17	0.031	65 °C	66.5 °C
D508N	4NP β-D-glucoside	41.0 ± 2.5	30.2 ± 0.71	0.74	60 °C	67.2 °C
R544W	4NP β-D-glucoside	0.26 ± 0.011	48.2 ± 0.53	186	60 °C	76.2 °C
R786H	4NP β-D-glucoside	15.7 ± 0.50	193 ± 2.4	12.3	70 °C	73.9 °C

[a]Wild type without the N-terminal tag is used for comparison to R544W and R786H, which were purified in the same way as the crystallized protein. [b]Wild type with the N-terminal tag is used for comparison to D508H and D508N, which could not be digested with enterokinase without internal cleavage of the protein. Temperature optimum curves are shown in SI Figure S8A. Melting temperatures (T_m) for all mutant proteins are those for the fusion protein without removal of the N-terminal tag. All T_m values were determined based on the change in circular dichroism at 220 nm upon heating. The melting curves are provided in SI Figure S8B.

structural homologues necessitated a SeMet-derivatized protein for structure solution by single wavelength anomalous dispersion (see Methods). The structure consists of an N-terminal domain, primarily formed by a two-sheet β-sandwich, and a C-terminal $(\alpha/\alpha)_6$ solenoid domain, Figure 1A. The C-terminal domain contains the residues previously proposed as the catalytic nucleophile and general acid/base in the archaeal β-glycosidase from *Sulfolobus solfataricus* and human GBA2.[3,7]The putative catalytic nucleophile, E441 (equivalent to human GBA2 E527), lies near the end of a long loop between the first (H5) and second (H6) α-helices of the C-terminal domain, while the putative catalytic acid/base, D593 (equivalent to human GBA2 D677), lies in a long loop between the fifth (H9) and sixth (H10) helices of the solenoid (Figure 1A). This loop also contains the binding site for a structural Ca^{2+} ion (Supporting Information, Figure S1A).

GBA2 was previously thought to be an intrinsic membrane protein with residues 689−708 predicted as a transmembrane domain but is easily extracted from cells in buffer without detergents.[2,6] The previously predicted hypothetical transmembrane domain corresponds to H10, an internal helix in the catalytic domain of *Tx*GH116 and the derived GBA2 model, which is incompatible with it being a transmembrane helix. Our model therefore supports the peripheral membrane localization of GBA2, where it may bind to an intrinsic membrane protein or polar lipid head groups.

*Tx*GH116 shows no structural similarity to other retaining β-glucosidases, which fall into families including GH1, GH2, GH3, GH5, and GH30 (SI, Figure S2). Comparison to structures in the PDB database with PDBeFold[2,3] found that the *Tx*GH116 structure is most similar to the family GH52 β-xylosidase from *Geobacillus thermoglucosidasius* (*Gt*GH52, PDB ID: 4C1O, 4C1P),[24] which upon global superposition yields an RMSD of 3.2 Å over 560 matched Cα atoms (PDBeFold Z score 8.3; SI, Figure S3A). However, the sequence similarity of GH52 and GH116 is negligible (GH52 enzymes are not detected using PSI-BLAST with the *Tx*GH116 sequence as a search query), GH52 lacks the N-terminal loop that contributes to the entrance to the active site in GH116, and different numbers of strands are found in the major β-sheets of the N-terminal domain (see SI). Nonetheless, the active sites also share the same geometry for the putative acid/base and nucleophile and a few sugar binding residues (SI Figure S3) allowing us to propose that these two divergent families be grouped as a "clan" (GH-O) of related 3-D structures.[25]

Catalytic Residues and Active Site Structure. GH116 family enzymes act *via* a double-displacement mechanism in which a covalent glycosyl enzyme intermediate is formed and subsequently hydrolyzed *via* oxocarbenium-ion like transition states (SI Figure S4). Such a mechanism demands two key residues: a general acid/base and a nucleophile. The putative acid/base, D593, is located approximately 8 Å (carboxylate to carboxylate) away from the nucleophile (E441; Figure 1B,C) in the $P2_12_12$ structure and 10 Å away in the $P6_1$ structure due to a displacement of the loop containing D593 by 2 Å (not shown). This is a longer distance than is observed in most retaining β-glucosidases, which typically have the acid/base within 4.5−6.5 Å of the nucleophile.[26−29] Given this ambiguity, we verified that the enzyme is retaining by identifying the initial hydrolysis product as β-D-glucose, which retains the anomeric stereochemistry of the substrate, by monitoring the reaction time course by NMR spectroscopy (SI, Figure S5A). Moreover, we demonstrated that E441 is the catalytic nucleophile and D593 the catalytic acid/base by mutation of these residues to alanine (which resulted in large losses of activity, Table 1), followed by chemical rescue[30] using small nucleophiles (see SI, Figures S5B−D).

Several established β-glucosidase inhibitors were examined to identify ligands that could bind in the active site. *Tx*GH116 is inhibited by the covalent inactivator 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucoside (DNP2FG),[31] and by the noncovalent inhibitors isofagomine (IFG), deoxynojirimycin (DNJ), glucoimidazole (GIM), δ-gluconolactone (weak inhibition at 1 mM), and the product D-glucose (Glc). Inhibition by IFG, DNJ, GIM, and Glc is competitive with K_i values of 2.9 ± 0.1 nM, 0.13 ± 0.01 μM, 0.34 ± 0.03 μM, and 4.0 ± 0.3 mM, respectively (SI Figure S6). *Tx*GH116 was not inhibited by conduritol B epoxide (CBE), an inhibitor that has been used to distinguish GBA1 (sensitive) from GBA2 (less sensitive[20]). In contrast, cyclophellitol, a related epoxide that contains a hydroxymethyl group corresponding to the hydroxymethyl group of glucose, rapidly inactivated *Tx*GH116. The similarity of the inhibition profile of human GBA2 and *Tx*GH116 with sugar-shaped inhibitors shows that *Tx*GH116 may be considered an informative functional model for the glycone-binding (−1) subsite of the human enzyme.

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

Table 3. X-ray Data Collection, Phasing, and Refinement Statistics (SAD)

	data set[a] (PDB ID)						
	native-form1 *Tx*GH116 (5FJS)	native-form2 *Tx*GH116 (5BVU)	SeMet *Tx*GH116	*Tx*GH116 G2F (5BX2)	*Tx*GH116 DNJ (5BX3)	*Tx*GH116 GIM (5BX4)	*Tx*GH116 Glc (5BX5)
data collection							
space group	$P6_1$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
cell dimensions							
a, *b*, *c* (Å)	*a* = 187.8, *b* = 187.8, *c* = 99.3	*a* = 177.7, *b* = 54.3, *c* = 83.2	*a* = 177.3, *b* = 54.5, *c* = 83.2	*a* = 177.1, *b* = 54.7, *c* = 83.3	*a* = 177.3, *b* = 54.7, *c* = 83.1	*a* = 177.2, *b* = 54.5, *c* = 83.2	*a* = 177.8, *b* = 54.6, *c* = 83.2
α, β, γ (deg)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
resolution (Å)	163−2.60 (2.67−2.60)	50−1.61 (1.64−1.61)[a]	50−1.90 (1.93−1.90)	50−1.61 (1.67−1.61)[a]	50−1.96 (2.03−1.96)	50−1.65 (1.71−1.65)	50−1.85 (1.88−0.85)
R_{merge} (%)	3.7 (79.4)	5.0 (32.2)	13.7 (68)	6 (39)	7.8 (28.8)	6.8 (50.4)	7.8 (45.3)
$I/\sigma I$	51.2 (3.0)	34.3 (5.9)	34.5 (7.9)	34.5 (6.5)	28.4 (10.2)	27.0 (3.9)	37.7 (8.25)
completeness (%)	100 (99.9)	99.0 (99.2)	99.9 (100)	99.9 (100)	99.3 (99.3)	99.9 (99.9)	95.1 (89.1)
redundancy	9.4 (10.5)	3.5 (3.5)	9.1 (9.5)	7.3 (7.4)	6.7 (7.1)	6.6 (6.1)	6.4 (6.2)
refinement							
resolution (Å)	162−2.6	89−1.61		177−1.61	177−1.96	89−1.65	89−1.85
no. reflections	61406	97906		99862	55692	91591	63333
R_{work}/R_{free} (%)	19/25	15.5/17.5		15.6/17.6	14.8/17.7	15.8/18.1	14.9/17.8
no. atoms							
protein	12083	6283		6292	6273	6295	6268
carbohydrate				11	11	14	12
hetero	2	101		107	95	107	71
water	78	583		628	441	619	474
B-factors (Å^2)							
protein	68	19.4		14.7	17.9	15.4	20.6
carbohydrate		-		9.6	11.8	10.4	17.5
hetero	68	37.5		32.4	39.2	33.6	38.8
water	55	31.5		27.5	29.1	28.4	30.1
RMS deviations							
bond lengths (Å)	0.014	0.008		0.007	0.009	0.007	0.010
bond angles (deg)	1.73	1.29		1.21	1.35	1.26	1.38
Ramachandran plot							
residues in most favorable regions (%)	93.4	88.5		89.3	89.7	88.8	89.3
residues in allowed regions (%)	5.5	11.5		10.7	10.3	11.2	10.7

[a]Each data set comprises data from a single crystal. Values in parentheses are for the highest-resolution shell.

1894
DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

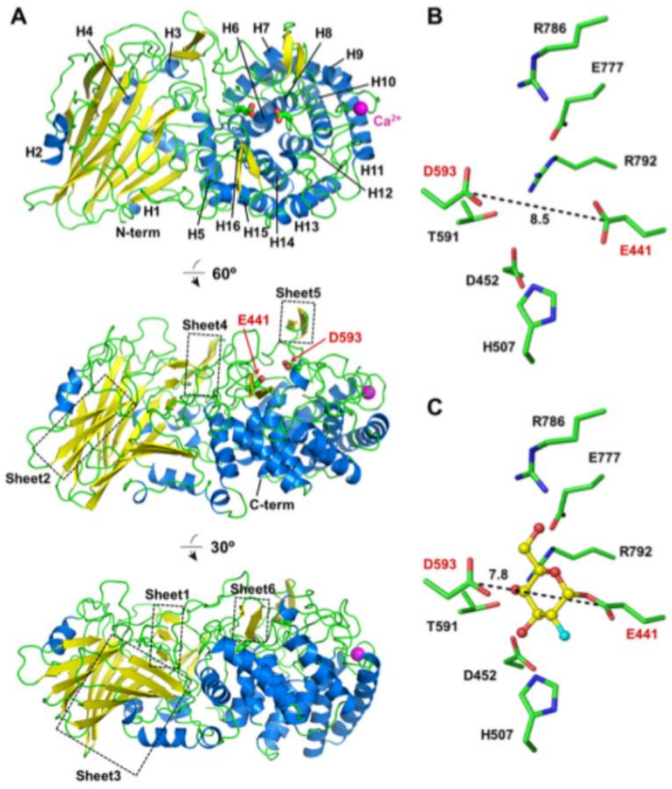


Figure 1. Three-dimensional structure of *Tx*GH116 β-glucosidase. (A) Three views of the three-dimensional structure of *Tx*GH116. The N-terminal domain consists of β-sheets (sheets 1−3) surrounded by α-helices (H1−H4). The C-terminal $(\alpha/\alpha)_6$ solenoid domain contains the active site and consists of 12 α-helices (H5−H16) in six outer helix plus inner helix pairs, with H5 situated between the two domains. A long loop with a two-strand β-sheet (sheet 5) lies between helices H7 and H8, near the active site (E441 and D593). α-Helices are shown in blue and β-strands in yellow. (B) Active site residues of the native structure of *Tx*GH116, showing the distance between the catalytic nucleophile (E441) and acid/base (D593) carbons (red labels). (C) Active site of the *Tx*GH116 covalent intermediate complex with 2-deoxy-2-fluoroglucose, showing the distance between the catalytic residues is decreased in this intermediate. Distances between the geometric means of the carboxyl groups are slightly shorter and are given in the text. Distances are in Å.

Insight into the active site and catalytic residues was obtained by solving the structures of *Tx*GH116 in complexes with a trapped intermediate derived from DNP2FG, the inhibitors DNJ and GIM, and the product (Glc) (Figure 2). The 2-fluoroglucosyl moiety binds to the catalytic nucleophile E441, consistent with the equivalent residue labeled by DNP2FG in studies with *Sulfolobus solfataricus* (E335) and human enzymes (E527),[3,7] and the mutagenesis and chemical rescue studies described in the SI. The putative acid/base, D593, is located in an unorthodox position above the pyranose ring, rather than in a typical position in the plane of the sugar ring which provides for syn- or anti- lateral protonation of the glycosidic oxygen.[32] Its position above the plane of the sugar ring prevents the efficient lateral protonation of GIM that is required for its strong binding, thus explaining the weak inhibition by this molecule.[32]

The structure of the glycosyl enzyme (Figure 2A) enabled identification of residues interacting with the substrate in the −1 subsite. The amino acid residues capable of forming hydrogen bonding interactions include E441 at OH2 (F2 in G2F); H507 and D452 at OH3; D452, T591, and R792 at OH4; and E777 and R786 at the OH6. In the last case, OH6 appears to mediate a salt bridge between E777 and R786. The interactions at OH6 may explain why cyclophellitol, bearing a hydroxymethyl group equivalent to the substrate C6-hydroxymethyl group, inhibits *Tx*GH116, but CBE, which lacks this group, does not. All of the glucosyl binding residues in *Tx*GH116 are conserved with human GBA2, whereas only a few of these could be reliably aligned with a conserved residue in the two biochemically characterized archaeal enzymes, SSO1353 and SSO3039 (SI Figure S7).

The complexes with DNJ and GIM and the product Glc reveal similar interactions to those seen in the G2F intermediate (Figure 2B−D). In the complexes with Glc and DNJ, the pyranose/piperidine rings are in $^4C_1$ chairs, while GIM adopts a $^4H_3$ conformation, corresponding to the preferred low energy conformations of these ligands.[33]

Model of Human GBA2 and Context of Disease-Causing Mutations. Human GBA2 and *Tx*GH116 share 37% sequence identity over the region of GBA2 residues 63 to 893, allowing an informative homology model of GBA2 to be generated from *Tx*GH116. Two longer loops in GBA2, one of which may contribute to the aglycone-binding cleft leading into the active site, as does the corresponding loop in *Tx*GH116

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

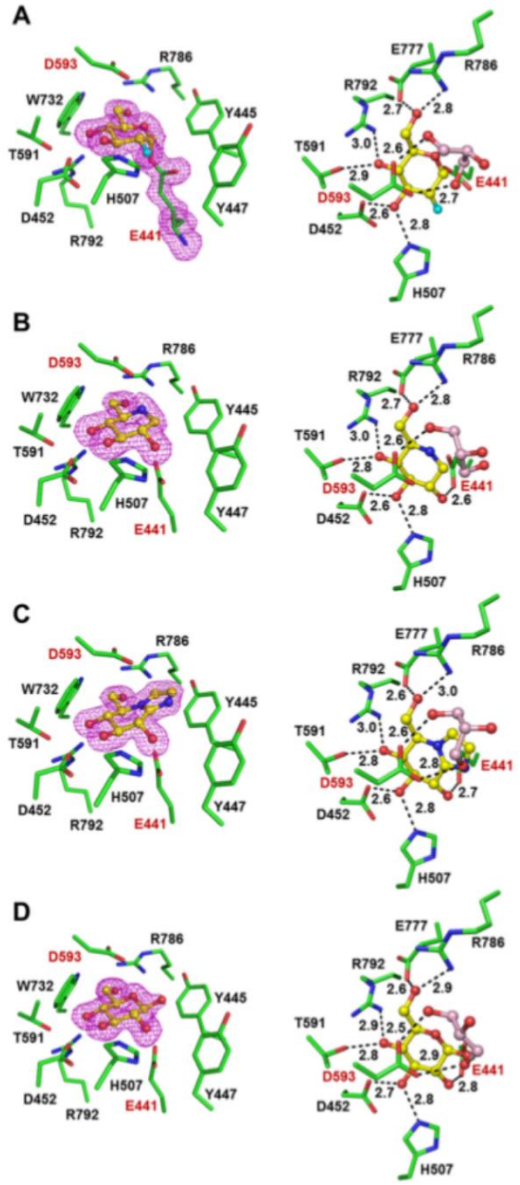


Figure 2. Active site of *Tx*GH116 alone and in complex with inhibitors and glucose. Left panels show the ligand electron density omit maps, contoured at 3σ, in the active sites of the complexes; the right panels show the potential hydrogen bonds between the ligands and surrounding residues, with the distances between the interacting atoms displayed. A glycerol molecule that was omitted from the left panels for clarity is shown in the right panels; catalytic residues are labeled in red. Ligands include (A) 2-deoxy-2-fluoroglucosyl (G2F), (B) deoxynojirimycin (DNJ), (C) glucoimidazole (GIM), and (D) β-D-glucose. Distances are in Å.

(Figure 3A), could not be modeled with high confidence. Notably, there is excellent correspondence between the active site residues, with all the residues noted as binding the sugar in *Tx*GH116 conserved in human GBA2 (Figure 3C).

Known pathogenic human GBA2 mutations include five missense mutations, F419V, D594H, R630W, G683R, and R873H (Figure 3B).[12−15,34] The amino acid residues corresponding to the sites of these mutations are conserved between GBA2 and *Tx*GH116 (F347, D508, R544, G599, and R786, respectively), as well as other GH116 enzymes in subfamily 1, which includes eukaryotic and bacterial enzymes (SI Figure S7).[21] As shown in Figure 3C, the GBA2 D594 and R873 residues (*Tx*GH116: D508 and R786) are located within the active site, providing an immediate catalytic implication for these mutations. As noted earlier, *Tx*GH116 R786 makes two hydrogen bonds to the glucosyl 6OH group, which are predicted to be conserved in GBA2 R873. The mutation R873H will result in a shorter side chain, which will not form an effective hydrogen bond to O6 of Glc. In *Tx*GH116, D508 hydrogen bonds to H507, which in turn makes a hydrogen bond to OH3, and the D594 residue makes the same interaction with the conserved H593 in the GBA2 model. The mutation D594H is expected to disrupt this interaction. The fact that two human pathogenic mutations involve sugar-binding residues emphasizes the importance of this model.

In contrast to the D584H and R873H, the other three missense mutations are located distal to the active site. The R630W mutation is approximately 15 Å away from the active site, between the layers of helices of the catalytic $(\alpha/\alpha)_6$ solenoid domain. The GBA2 model shows that R630 can form salt bridge interactions with two conserved carboxylates, E555 and D631 (Figure 3D). The mutation R630W will result in the loss of a charged residue, disturbing the charge balance, and introduction of a somewhat larger Trp residue, which together may result in destabilization of the protein. The only pathogenic mutation found in the N-terminal β-sandwich domain is F419V. F419 is positioned at the interface of the two domains, and the aromatic ring appears to form dispersive interactions with the conserved neighboring aromatic residues Y534 in the long loop between the H5 and H6 helices and W603 in the loop between the H7 and H8 helices of the C-terminal $(\alpha/\alpha)_6$ solenoid domain (Figure 3B,E). V419 cannot make such interactions, so the F419V mutation may destabilize the interaction between the two domains. The G683R mutation, located approximately 20 Å from the active site, is in an outside turn of the long loop including the catalytic acid/base (D677) and the sugar-binding residue T675 (Figure 3F). G683 is located very close to the neighboring loop, in which S741 is closest to G683, such that a replacement with arginine would result in a clash. It is also next to the binding site for the calcium cation in *Tx*GH116, suggesting that a positively charged Arg could disrupt this binding.

It was recently shown that expression of GBA genes possessing all known pathogenic mutations in COS-7 or HeLa cells resulted in reduced activity in crude cell extracts relative to wild-type.[34] We sought to gain insight into the physical and mechanistic effects of the pathogenic GBA2 D594H, R630W, and R873H mutations by analysis of corresponding mutations in *Tx*GH116. For all three mutations, *Tx*GH116 D508H, R544W, and R786H (corresponding to GBA2 D594H, R630W, and R873H), large decreases in activity were observed at the temperature optimum of the wild type enzyme (Table 2, SI Figure S8). For the *Tx*GH116 R786H mutation (corresponding to GBA2 R873H), a 90-fold increase in the K_M value led to a reduction in k_{cat}/K_M of more than 20-fold (Table 2), which is consistent with the weaker binding of the sugar expected based on this residue's interaction with the

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

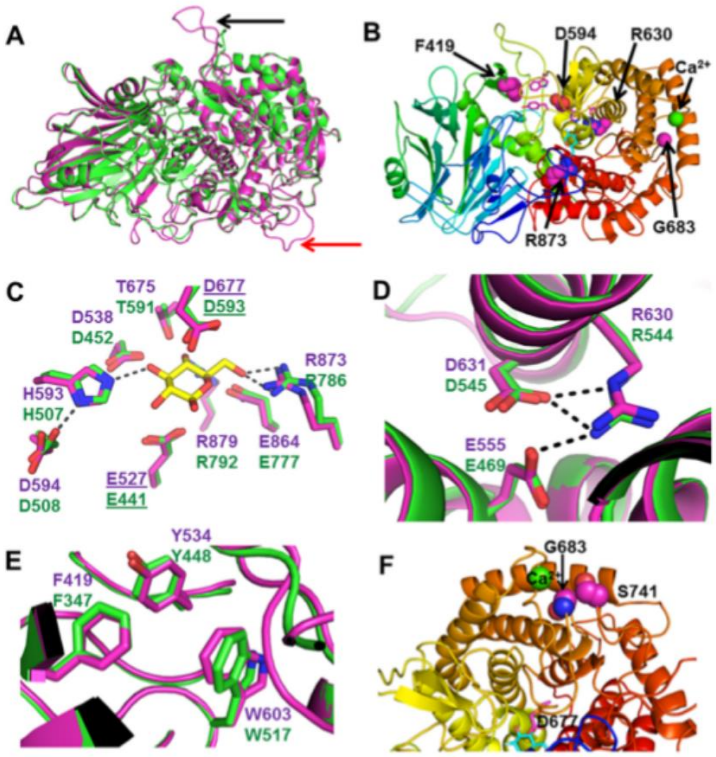


Figure 3. Human GBA2 structural model. (A) Overall structure of the human GBA2 model (purple) superimposed with the *Tx*GH116 structure (green). The loop adjacent to the active site, which is longer in GBA2, is marked with a red arrow, while the longer loop between helices H5 and H6 of the $(\alpha/\alpha)_6$ solenoid domain is marked with a black arrow. (B) Diagram of human GBA2 model showing positions of residues mutated in human disease (space-filling) and interacting residues and ligand (sticks). (C) Superposition of active site residues of human GBA2 and the *Tx*GH116 complex with glucose. All of the sugar binding residues are conserved between *Tx*GH116 and GBA2, as well as the aspartate (D508) that interacts with H507 in *Tx*GH116. The H-bonds between residues involved in human autosomal recessive cerebellar ataxia mutations and the sugar residue are shown as dashed lines. The catalytic nucleophile and acid/base labels are underlined. (D) Superposition of human GBA2 model (purple) on *Tx*GH116 structure (green) in the area of the R630 residue mutated in certain hereditary spastic paraplegia cases. R630 and the nearby carboxylate residues D631 and E555 are conserved with *Tx*GH116 residues. (E) View of GBA2 F419 and the conserved interacting aromatic groups in the overlay of the GBA2 and *Tx*GH116 models. (F) Position of the GBA2 G683R mutation, showing that, although it is near the surface of the protein, G683 is tightly opposed to S741 (both shown in magenta spheres) and the bound calcium cation (green sphere) and is also linked to the catalytic acid/base D677 shown next to the glucosyl residue (cyan carbon sticks).

glucose 6OH. Recapitulation of the GBA2 D594H mutation in *Tx*GH116 D508H resulted in both decreased stability and lowered activity (Table 2). Most notably, the k_{cat}/K_M value decreased 5800-fold, owing mostly to a >500-fold increase in the K_M value, with a small reduction in the k_{cat} value. The large decrease in k_{cat}/K_M and relatively small effect on k_{cat} suggests a large effect on the binding and stabilization of the substrate and transition state of the first committed step of hydrolysis (glycosylation), whereas the rate limiting step is less affected. The wild type k_{cat} values for 4NPGlc (possessing a leaving group with a relatively low pK_a value and thus a relatively low energy barrier for departure) and cellobiose (possessing a poor glucosyl leaving group with a high pK_a value) are similar (Table 2), suggesting that the deglycosylation step (for which the reaction is the same) is rate-limiting. The small change in k_{cat} in the D508H mutant suggests that this is most likely also true for this mutant.[35] To investigate the importance for catalysis of the charge of D508, the isosteric mutation D508N, which should maintain hydrogen bonding to H507, was studied. D508N exhibited a >240-fold reduction in k_{cat}/K_M, and thus, the negative charge on D508 appears to be important for catalysis.

The effect of the GBA2 R630W mutation was studied through the *Tx*GH116 R544W mutant. The R544W mutant exhibited a slightly increased K_M value and little change in the melting temperature but caused a decrease in the temperature of the optimum catalytic activity by 15 °C. These properties suggest the R544W mutation may destabilize the active site, while having little effect on the overall protein stability. This hypothesis is consistent with the observation that a large amount of GBA2 protein accumulated in COS-7 and HeLa cells that expressed the R630W variant, suggesting it is not an unstable protein despite the low level of activity in the standard GBA assay.[34]

■ CONCLUSIONS

GBA2 has risen to prominence due to its involvement in congenital neurological disorders leading to cerebellar ataxia and spastic paraplegia, as well as its interplay with GBA1, defects in which cause Gaucher disease and contribute to Parkinsonism.[36] Modulators of GBA2 activity may therefore provide therapeutic benefits in the treatment of diseases involving defects in GBA1 or GBA2. The *Tx*GH116 structure

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

described here possesses many unique features, particularly in the N-terminal β-sandwich, which is tightly associated with the catalytic domain and contributes to the substrate-binding cleft, and the unusual orientation of the general acid/base residue. The residues binding the glucose in the −1 subsite are highly conserved between *Tx*GH116 and human GBA2, and the structural model provides a useful framework for understanding the molecular basis for binding of ligands to the glycone-binding site. Significantly, key missense mutations causing defects in GBA2 are conserved between the two enzymes, which have allowed identification of likely biochemical and structural reasons underpinning the genetic defects. The present work demonstrates that the missense mutations are located both proximal to and remote from the active site and result in a loss of enzymatic activity by a range of mechanisms including reduction in substrate binding, reduced temperature for optimal catalysis, and reduced enzyme stability.

METHODS

Plasmid Construction and Protein Expression. Two independent constructs presenting *E. coli*-optimized *Tx*GH116 genes were prepared in pET28a(+) and pET30a(+) expression vectors with polyhistidine tags, as described in the Supporting Information (SI). The pET28a(+) construct generated a *Tx*GH116 protein with the first 20 amino acid residues replaced with an N-terminal His_6-tag that was designated *Tx*GH116-N. The pET30a(+) construct encoded the *Tx*GH116 protein lacking the first 18 amino acids following the pET30(+) N-terminal tag including a His_6 tag, S-tag, and enterokinase site. This fusion protein, designated *Tx*GH116Δ1−18, also included the C-terminal His_6 tag from the plasmid. The fusion proteins were produced in *E. coli* strain BL21(DE3) and purified by immobilized metal affinity chromatography, as described in the SI Methods. Production of selenomethionine-labeled *Tx*GH116 (Se*Tx*GH116) was accomplished with the pET30a/*Tx*GH116Δ1−18 plasmid in *E. coli* strain B834(DE3) pLysS cells grown in a SeMet core medium (Nihon Jyunyaku Co.; SI Methods).

Enzymatic Characterization. The *Tx*GH116Δ1−18 protein was used for enzymatic characterization. The optimum pH of *Tx*GH116 for hydrolysis of 4NPGlc was determined to be pH 5.5 by incubating 0.05 μg of enzyme with 1 mM 4NPGlc in 100 mM McIlvaine universal (phosphate-citrate) buffer, pH 2.5−8.0 at 0.5 pH unit intervals, at 60 °C for 30 min. The temperature optimum (75 °C) was determined *via* the hydrolysis of 1 mM 4NPGlc with 0.05 μg of enzyme in 50 mM sodium acetate buffer, pH 5.5, at temperatures of 10−110 °C, for 10 min (SI Figure S8A). For standard activity assays, 1 mM 4NPGlc was incubated with enzyme at 60 °C for 15 min. Chemical rescue activity of E441A and D593A mutants and wild type *Tx*GH116 were assayed under conditions of 50 mM MES, pH 5.5, and 60 °C with 1 mM 4NPGlc, 0.1 μg of wild type for 15 min, 25 μg of E441 for 60 min, and 5 μg for 15 min of D593A with sodium formate and sodium azide concentrations of 0−4 M.

Inhibition constants were determined by assaying triplicate reactions at five 4NPGlc substrate concentrations bracketing the apparent K_M value and five inhibitor concentrations. The inhibitors were preincubated with 0.05 μg of enzyme at 37 °C for 10 min before the addition of 4NPGlc and incubation at 60 °C for 15 min. The K_i values were calculated from the x intercept of the derivative plot of the slopes of Lineweaver−Burk reciprocal plots (1/v vs 1/[4NPGlc]) vs the inhibitor concentration (SI Figure S6). Inhibition was competitive in all cases.

NMR Determination of Retaining Mechanism and Catalytic Rescue Products. The proton and STD NMR experiments were performed on a 500 MHz NMR spectrometer (Bruker AVANCE III HD) with a CPP BBO 500 Cryoprobe and BRUKER TOPSPIN 3.2 software. The ^{1}H NMR spectra were collected at a frequency of 500.366 MHz. The spectra of the 4NPGlc (7.3 mM) hydrolysis reactions were collected before and 10 and 30 min after adding 50 μM *Tx*GH116Δ1−18 enzyme.

Mutagenesis. Site-directed mutagenesis was done by the QuikChange method (Stratagene, Agilent Corp.) to make a catalytic nucleophile (E441A) and acid/base (D593A) mutations and replicate pathogenic human GBA2 mutations (D508H and its variant D508N, R544W, and R786H) in the pET30a/*Tx*GH116Δ1−18 expression vector. Mutagenic primers are listed in SI Table S1. The mutations and entire coding sequence were verified by DNA sequencing (Macrogen).

Protein Crystallization. Crystals were obtained after screening and optimization by hanging drop vapor diffusion at 15 °C in a precipitant of 0.2 M ammonium sulfate, 20% PEG 3000, 0.1 M MES, pH 5.5, for *Tx*GH116Δ1−18, and at 19 °C in 0.2 M ammonium sulfate, 15% w/v PEG 6000, 0.1 M Bis-Tris propane, pH 8, for *Tx*GH116N.

X-ray Diffraction and Structure Solution. Crystals of *Tx*GH116Δ1−18 were soaked briefly in precipitant containing 20% glycerol. For ligand complexes, the soaking solution also contained either 10 mM DNJ, GIM, 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-glucopyranoside (DNPG2F), or 100 mM D-glucose. The crystals were flash vitrified in liquid nitrogen and data collected at 100 K. Diffraction data were collected at the SPring-8 synchrotron beamline BL44XL with 0.9000 Å X-ray radiation for native and 0.97865 Å for SeMet *Tx*GH116Δ1−18 crystals on a MX-300HE detector (Rayonix). The data were processed with the HKL2000 suite.[37] The initial phases were calculated by the single wavelength anomalous dispersion (SAD) method with PHENIX AUTOSOL.[38] The SAD calculation revealed one protein molecule per asymmetric unit in the $P2_12_12$ space group. The Matthews coefficient was 2.20 Å^3/Da, and the solvent content was 44.19%. The SeMet-derivatized protein included 26 SeMet, 25 of which were observed in the anomalous difference Fourier map. X-ray data for *Tx*GH116-N were collected at beamline i24 (wavelength: 0.97 Å) of the Diamond Light Source (Oxfordshire, UK) with a Pilatus3 6 M (100 Hz max. frame rate) detector and processed using MOSFLM and scaled in SCALA[39] to show two protein molecules per asymmetric unit in the $P6_1$ space group. The solvent content was 56%, and the Matthews coefficient was 2.8 Å^3/Da.

The initial phases were transformed to native structure by molecular replacement with the MOLREP program[40] in the CCP4 suite.[39] The structure was built with the COOT graphic program[41] in alternation with refinement with REFMAC5.[42] The complex and *Tx*GH116-N structures were solved by MOLREP[40] and refined in a similar manner. The final models were analyzed with PROCHECK[43] and MolProbity[44] and validated on the PDB Web site. Data and structure refinement statistics are shown in Table 3.

Molecular Modeling. To construct a homology model of the human GBA2 protein, the first 64 and last 34 residues of the human protein, which do not have corresponding sequences in *Tx*GH116, were deleted and the remaining sequence aligned to the *Tx*GH116 sequence and used to build five models within the Modeler program.[45] After the addition of hydrogen and optimization of hydrogen-bonding side chain rotamers with Molprobity,[44] the model with the lowest Molprobity (3.31) and Clash (108.73) scores was used as the human GBA2 model. Comparison of the homology model of human GBA2 with the *Tx*GH116 template structure gave an RMSD of 0.543 Å over 712 Cα atoms that could be aligned.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.6b00192.

Supporting Results and Discussion, Methods, Table S1, Figures S1−S8 (PDF)

AUTHOR INFORMATION

Corresponding Authors

*Tel.: +44 1904 322511. Fax: +44 1904 322516. E-mail: gideon.davies@york.ac.uk.

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

*Tel.: +81 6-6879-8604. Fax: +81-6-6879-8606. E-mail: gkurisu@protein.osaka-u.ac.jp.
*Tel.: +66 44 224304. Fax: +66 44 224185. E-mail: cairns@sut.ac.th.

Author Contributions

[+]Should be considered co-first authors.

Funding

Funding was provided by The Thailand Research Fund (Grant BRG5680012), Suranaree University of Technology, the National Research University project grant from the Commission on Higher Education of Thailand to SUT, the International Collaborative Research Program (ICR-14−03) from the Institute for Protein Research, Osaka University. G.J.D., I.B., and L.W. were supported by the European Research Council through ERC-2012-AdG-32294 "Glycopoise." S.J.W. was supported by the Australian Research Council.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Y. Goto for assistance in circular dichroism and the research staff at Osaka Analytical Center of Rigaku Corp. W. Offen is thanked for assistance and critical reading of the manuscript, and J. Svasti is thanked for advice and support.

■ REFERENCES

(1) Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res. 42*, D490−495.

(2) Matern, H., Boermans, H., Lottspeich, F., and Matern, S. (2001) Molecular cloning and expression of human bile acid β-glucosidase. *J. Biol. Chem. 276*, 37929−37933.

(3) Cobucci-Ponzano, B., Aurilia, V., Riccio, G., Henrissat, B., Coutinho, P. M., Strazzulli, A., Padula, A., Corsaro, M. M., Pieretti, G., Pocsfalvi, G., Fiume, I., Cannio, R., Rossi, M., and Moracci, M. (2010) A new archaeal β-glycosidase from *Sulfolobus solfataricus*: seeding a novel retaining β-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J. Biol. Chem. 285*, 20691−20703.

(4) Matern, H., Heinemann, H., Legler, G. N., and Matern, S. (1997) Purification and characterization of a microsomal bile acid β-glucosidase from human liver. *J. Biol. Chem. 272*, 11261−11267.

(5) Yildiz, Y., Matern, H., Thompson, B., Allegood, J. C., Warren, R. L., Ramirez, D. M., Hammer, R. E., Hamra, F. K., Matern, S., and Russell, D. W. (2006) Mutation of β-glucosidase 2 causes glycolipid storage disease and impaired male fertility. *J. Clin. Invest. 116*, 2985−2994.

(6) Körschen, H. G., Yildiz, Y., Raju, D. N., Schonauer, S., Bonigk, W., Jansen, V., Kremmer, E., Kaupp, U. B., and Wachten, D. (2013) The non-lysosomal β-glucosidase GBA2 is a non-integral membrane-associated protein at the endoplasmic reticulum (ER) and Golgi. *J. Biol. Chem. 288*, 3381−3393.

(7) Kallemeijn, W. W., Witte, M. D., Voorn-Brouwer, T. M., Walvoort, M. T., Li, K. Y., Codee, J. D., van der Marel, G. A., Boot, R. G., Overkleeft, H. S., and Aerts, J. M. (2014) A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining β-glucosidases. *J. Biol. Chem. 289*, 35351−35362.

(8) Lingwood, C. A. (2011) Glycosphingolipid functions. *Cold Spring Harbor Perspect. Biol. 3*, a004788.

(9) Bendelac, A, Savage, P. B., and Teyton, L. (2007) The biology of NKT cells. *Annu. Rev. Immunol. 25*, 297−336.

(10) Jmoudiak, M., and Futerman, A. H. (2005) Gaucher disease: pathological mechanisms and modern management. *Br. J. Haematol. 129*, 178−188.

(11) Butters, T. D. (2007) Gaucher disease. *Curr. Opin. Chem. Biol. 11*, 412−418.

(12) Martin, E., Schule, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J. L., Gonzalez, M. A., Mundwiller, E., Deconinck, T., Wessner, M., Jornea, L., Oteyza, A. C., Durr, A., Martin, J. J., Schols, L., Mhiri, C., Lamari, F., Zuchner, S., De Jonghe, P., Kabashi, E., Brice, A., and Stevanin, G. (2013) Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. *Am. J. Hum. Genet. 92*, 238−244.

(13) Hammer, M. B., Eleuch-Fayache, G., Schottlaender, L. V., Nehdi, H., Gibbs, J. R., Arepalli, S. K., Chong, S. B., Hernandez, D. G., Sailer, A., Liu, G., Mistry, P. K., Cai, H., Shrader, G., Sassi, C., Bouhlal, Y., Houlden, H., Hentati, F., Amouri, R., and Singleton, A. B. (2013) Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am. J. Hum. Genet. 92*, 245−251.

(14) Votsi, C., Zamba-Papanicolaou, E., Middleton, L. T., Pantzaris, M., and Christodoulou, K. (2014) A novel GBA2 gene missense mutation in spastic ataxia. *Ann. Hum. Genet. 78*, 13−22.

(15) Citterio, A., Arnoldi, A., Panzeri, E., D'Angelo, M. G., Filosto, M., Dilena, R., Arrigoni, F., Castelli, M., Maghini, C., Germiniasi, C., Menni, F., Martinuzzi, A., Bresolin, N., and Bassi, M. T. (2014) Mutations in CYP2U1, DDHD2 and GBA2 genes are rare causes of complicated forms of hereditary spastic paraparesis. *J. Neurol. 261*, 373−381.

(16) Mistry, P. K., Liu, J., Sun, L., Chuang, W. L., Yuen, T., Yang, R., Lu, P., Zhang, K., Li, J., Keutzer, J., Stachnik, A., Mennone, A., Boyer, J. L., Jain, D., Brady, R. O., New, M. I., and Zaidi, M. (2014) Glucocerebrosidase 2 gene deletion rescues type 1 Gaucher disease. *Proc. Natl. Acad. Sci. U. S. A. 111*, 4934−4939.

(17) Marques, A. R. A., Aten, J., Ottenhoff, R., van Roomen, C. P. A. A., Herrera Moro, D., Claessen, N., Veloz, M. F. V., Zhou, K., Lin, Z., Mirzaian, M., Boot, R. G., De Zeeuw, C. I., Overkleeft, H. S., Yildiz, Y., and Aerts, J. M. F. G. (2015) Reducing GBA2 activity ameliorates neuropathology in Niemann-Pick Type C mice. *PLoS One 10*, e0135889.

(18) Sorli, S. C., Colie, S., Albinet, V., Dubrac, A., Touriol, C., Guilbaud, N., Bedia, C., Fabrias, G., Casas, J., Segui, B., Levade, T., and Andrieu-Abadie, N. (2013) The nonlysosomal β-glucosidase GBA2 promotes endoplasmic reticulum stress and impairs tumorigenicity of human melanoma cells. *FASEB J. 27*, 489−498.

(19) Overkleeft, H. S., Renkema, G. H., Neele, J., Vianello, P., Hung, I. O., Strijland, A., van der Burg, A. M., Koomen, G. J., Pandit, U. K., and Aerts, J. M. (1998) Generation of specific deoxynojirimycin-type inhibitors of the non-lysosomal glucosylceramidase. *J. Biol. Chem. 273*, 26522−26527.

(20) Ridley, C. M., Thur, K. E., Shanahan, J., Thillaiappan, N. B., Shen, A., Uhl, K., Walden, C. M., Rahim, A. A., Waddington, S. N., Platt, F. M., and van der Spoel, A. C. (2013) β-Glucosidase 2 (GBA2) activity and imino sugar pharmacology. *J. Biol. Chem. 288*, 26052−26066.

(21) Ferrara, M. C., Cobucci-Ponzano, B., Carpentieri, A., Henrissat, B., Rossi, M., Amoresano, A., and Moracci, M. (2014) The identification and molecular characterization of the first archaeal bifunctional exo-beta-glucosidase/N-acetyl-beta-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. *Biochim. Biophys. Acta 1840*, 367−377.

(22) Lee, Y.-E., Jain, M. K., Lee, C., Lowe, S. E., and Zeikus, J. G. (1993) Taxonomic distinction of saccharolytic thermophilic anaerobes: Description of *Thermoanaerobacterium xylanolyticum* gen. nov., sp. nov., and *Thermoanaerobacterium saccharolyticum* gen. nov., sp. nov.; Reclassification of *Thermoanaerobium brockii*, *Clostridium thermosulfurogenes*, and *Clostridium thermohydrosulfiricum* E100−69 as *Thermoanaerobacter brockii* comb. nov., *Thermoanaerobacterium thermosulfurigenes* comb. nov., and *Thermoanaerobacter thermohydrosulfuricus* comb. nov.,respectively; and transfer of *Clostridium*

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

thermohydrosulfuricum 39E to *Thermoanaerobacter ethanolicus*. *Int. J. Syst. Bacteriol.* 43, 41−51.
(23) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2256−2268.
(24) Espina, G., Eley, K., Pompidor, G., Schneider, T. R., Crennell, S. J., and Danson, M. J. (2014) A novel beta-xylosidase structure from *Geobacillus thermoglucosidasius*: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 70, 1366−1374.
(25) Henrissat, B., and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637−644.
(26) Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995) The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3, 951−960.
(27) Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure* 7, 179−190.
(28) Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V., and Schulein, M. (1998) Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. *Biochemistry* 37, 1926−1932.
(29) Dvir, H., Harel, M., McCarthy, A. A., Toker, L., Silman, I., Futerman, A. H., and Sussman, J. L. (2003) X-ray structure of human acid-β-glucosidase, the defective enzyme in Gaucher disease. *EMBO Rep.* 4, 704−709.
(30) Ly, H. D., and Withers, S. G. (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem.* 68, 487−522.
(31) Withers, S. G., Warren, R. A. J., Street, I. P., Rupitz, K., Kempton, J. B., and Aebersold, R. (1990) Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a retaining glycosidase. *J. Am. Chem. Soc.* 112, 5887−5889.
(32) Heightman, T. D., and Vasella, A. (1999) Recent insights into inhibition, structure, and mechanism of configuration-retaining glycosidases. *Angew. Chem., Int. Ed.* 38, 750−770.
(33) Granier, T., Panday, N., and Vasella, A. (1997) Structure-activity relations for imidazo-pyridine-type inhibitors of β-D-glucosidases. *Helv. Chim. Acta* 80, 979−987.
(34) Sultana, S., Reichbauer, J., Schule, R., Mochel, F., Synofzik, M., and van der Spoel, A. C. (2015) Lack of enzyme activity in GBA2 mutants associated with hereditary spastic paraplegia/cerebellar ataxia (SPG46). *Biochem. Biophys. Res. Commun.* 465, 35−40.
(35) MacLeod, A. M., Lindhorst, T., Withers, S. G., and Warren, R. A. J. (1994) The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* 33, 6371−6376.
(36) Bultron, G., Kacena, K., Pearson, D., Boxer, M., Yang, R., Sathe, S., Pastores, G., and Mistry, P. (2010) The risk of Parkinson's disease in type 1 Gaucher disease. *J. Inherited Metab. Dis.* 33, 167−173.
(37) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, In *Methods in Enzymology, Macromolecular Crystallography, part A* (Carter, C. W. J., and Sweet, R. M., Eds.), pp 307−326, Academic Press, New York.
(38) Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L.-W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 65, 582−601.
(39) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67, 235−242.
(40) Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 22−25.
(41) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2126−2132.
(42) Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 67, 355−367.
(43) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283−291.
(44) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 12−21.
(45) Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, In *Current Protocols in Bioinformatics*, p 5.6.1, Chapter 5, John Wiley and Sons, Inc., Hoboken, NJ. DOI: 10.1002/0471250953.bi0506s15.

DOI: 10.1021/acschembio.6b00192
ACS Chem. Biol. 2016, 11, 1891−1900

Appendix 5.

ACS central science

Research Article

http://pubs.acs.org/journal/acscii

1,6-Cyclophellitol Cyclosulfates: A New Class of Irreversible Glycosidase Inhibitor

Marta Artola,[†] Liang Wu,[‡] Maria J. Ferraz,[§] Chi-Lin Kuo,[§] Lluís Raich,[∥] Imogen Z. Breen,[‡] Wendy A. Offen,[‡] Jeroen D. C. Codée,[†] Gijsbert A. van der Marel,[†] Carme Rovira,[∥,⊥] Johannes M. F. G. Aerts,[§] Gideon J. Davies,*[,‡] and Herman S. Overkleeft*[,†]

[†]Department of Bio-organic Synthesis and [§]Department of Medical Biochemistry, Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

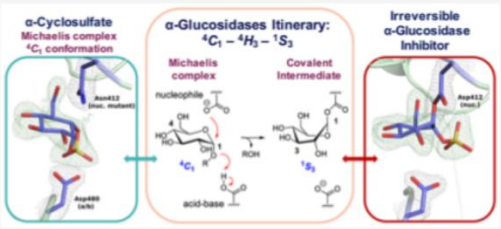
[‡]Department of Chemistry, University of York, Heslington, York, YO10 5DD, U.K.

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

[⊥]Fundació Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

Supporting Information

ABSTRACT: The essential biological roles played by glycosidases, coupled to the diverse therapeutic benefits of pharmacologically targeting these enzymes, provide considerable motivation for the development of new inhibitor classes. Cyclophellitol epoxides and aziridines are recently established covalent glycosidase inactivators. Inspired by the application of cyclic sulfates as electrophilic equivalents of epoxides in organic synthesis, we sought to test whether cyclophellitol cyclosulfates would similarly act as irreversible glycosidase inhibitors. Here we present the synthesis, conformational analysis, and application of novel 1,6-cyclophellitol cyclosulfates. We show that 1,6-*epi*-cyclophellitol cyclosulfate (α-cyclosulfate) is a rapidly reacting α-glucosidase inhibitor whose $^4C_1$ chair conformation matches that adopted by α-glucosidase Michaelis complexes. The 1,6-cyclophellitol cyclosulfate (β-cyclosulfate) reacts more slowly, likely reflecting its conformational restrictions. Selective glycosidase inhibitors are invaluable as mechanistic probes and therapeutic agents, and we propose cyclophellitol cyclosulfates as a valuable new class of carbohydrate mimetics for application in these directions.



INTRODUCTION

The huge diversity of oligosaccharide, polysaccharide, and glycoconjugate structures found in nature reflect their many roles in biological processes. Such diversity is mirrored in the numerous enzymes that have evolved to degrade these biopolymers. Hydrolytic carbohydrate degrading enzymes, glycoside hydrolases or glycosidases, are classified into over 140 distinct sequence (and hence structural) families in the CAZy database.[1] This classification provides a powerful framework upon which aspects of conformational analysis and enzyme inhibition can be constructed.

Most retaining glycosidases employ a conserved mechanism, in which the acetals (occasionally ketals) that make up interglycosidic linkages are hydrolyzed using two key carboxylic acid residues (Asp or Glu), which function as catalytic nucleophile and catalytic acid/base.[2] Central to glycosidase activity are the considerable conformational distortions undergone by the sugar substrate throughout the catalytic cycle, which are necessary to satisfy the stereoelectronic and orbital overlap requirements for glycoside hydrolysis.[3]

Upon binding to the enzyme, a retaining glycosidase substrate in the Michaelis complex initially adopts a conformation in which the leaving group is axially or pseudoaxially positioned, allowing for favorable in-line attack by the catalytic nucleophile. Nucleophilic attack, with concurrent protonation of the glycosidic oxygen by the catalytic acid/base, leads to a high-energy state (the transition state, TS) with considerable oxocarbenium ion character. This transition state must necessarily adopt a conformation in which C5−O5−C1−C2 are coplanar, to allow for partial oxocarbenium double bond development between O5 and C1. Following this transition state, the substrate forms a covalent intermediate with the glycosidase nucleophile, which is intercepted by water (following the reverse conformational itinerary) to liberate free sugar and enzyme. For retaining α- and β-glucosidases, the typical conformational itineraries for the Michaelis complex → TS → intermediate enzymatic half reactions are $^4C_1 \rightarrow {}^4H_3^{\ddagger} \rightarrow {}^1S_3$ and $^1S_3 \rightarrow {}^4H_3^{\ddagger} \rightarrow {}^4C_1$ respectively (Figure 1a), with the

Received: May 19, 2017
Published: July 13, 2017

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

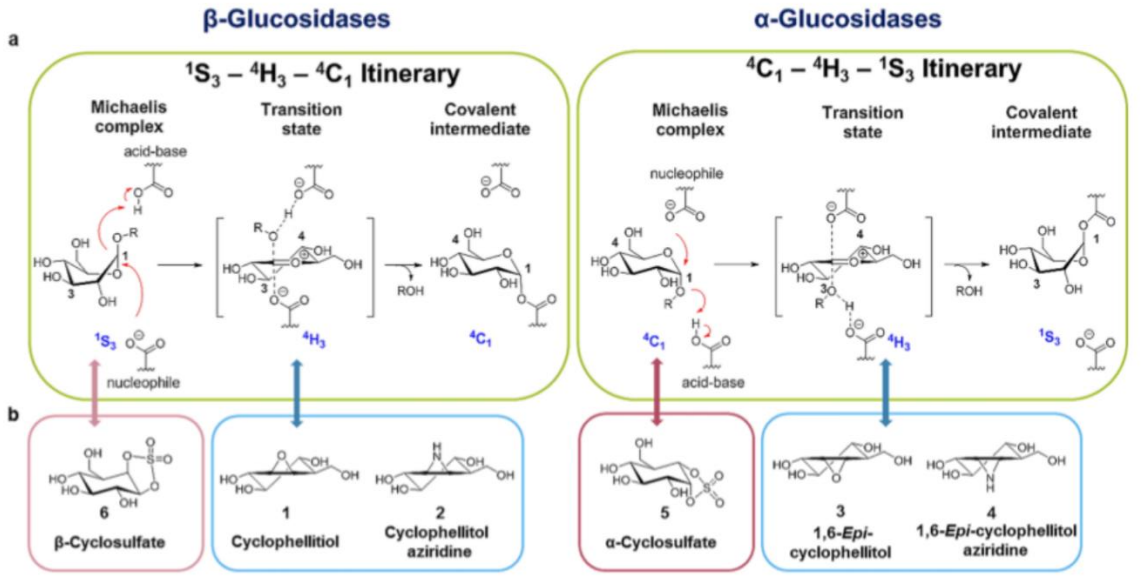


Figure 1. Conformational itinerary of retaining glucosidase reaction pathways and conformation of covalent inhibitors. (a) Reaction itineraries of retaining β-glucosidases and retaining α-glucosidases, showing conformations of the Michaelis complex, transition state, and covalent intermediates. (b) Structures of cyclophellitol (**1**), cyclophellitol aziridine (**2**), 1,6-*epi*-cyclophellitol (**3**), 1,6-*epi*-cyclophellitol aziridine (**4**), α-cyclosulfate (**5**), and β-cyclosulfate (**6**).

$^4H_3$ TS conformation common to both α- and β-glucosidases.[4−7]

Mimicry of transition state features is a powerful strategy for the design of enzyme inhibitors, and thus the majority of competitive glycosidase inhibitors bind by virtue of mimicking either the positive charge or the conformational shape of the transition state.[3,8,9] Conformational mimicry can also aid the reaction of irreversible covalent glycosidase inhibitors,[10] such as cyclitol epoxides and aziridines.[10−12] The natural product, cyclophellitol (**1**), and its synthetic analogue, cyclophellitol aziridine (**2**), are glucose configured inhibitors, which mimic the $^4H_3$ conformation adopted by retaining β-glucosidase substrates in their transition state.[3,13,14] Upon binding to the enzyme active site, nucleophilic attack by the catalytic nucleophile residue opens the epoxide or aziridine ring, irreversibly inhibiting the enzyme via formation of a covalent enzyme−cyclophellitol (aziridine) adduct.

In general, reactive cyclophellitol analogues, locked into TS-like conformations, are rapid irreversible inactivators of retaining glycosidases.[12,15,16] We recently demonstrated that replacement of the cyclophellitol epoxide or aziridine ring by a nonhydrolyzable cyclopropane also yields reversible glycosidase inhibitors with micromolar to nanomolar affinity.[9] Thus, conformation, rather than just electrophilic reactivity, may make a substantial contribution to the inhibitory potency of cyclophellitol analogues.

Given the success of reactive sugar epoxides and aziridines as conformationally enhanced retaining glycosidase inactivators,[10,12,17] we sought to expand the concept to other electrophilic species. Cyclic sulfate electrophiles are often used in synthetic organic chemistry as equivalents of epoxides in nucleophilic additions, sometimes showing superior reactivity and regioselectivity.[18] We therefore sought to test if cyclophellitol configured cyclosulfates could act as a new class of covalent glycosidase inhibitors. Specificity for "α" and "β" enzymes can be conferred simply through choice of the appropriately configured cyclosulfate. We additionally reasoned that substitution of the epoxide in **1** and **3** by a 1,2-*cis*-cyclic sulfate should yield compounds favoring a $^4C_1$ conformation. "α-Configured" 1,6-*epi*-cyclophellitol cyclosulfate **5** (hereafter referred to as α-cyclosulfate) should match, conformationally, the $^4C_1$ Michaelis complex conformation of substrates for α-glucosidases and thus be readily poised for in-line nucleophilic attack by the enzyme, yielding a rapid and potent irreversible inhibitor. In contrast, a "β-configured" 1,6-cyclophellitol cyclosulfate **6** (β-cyclosulfate) locked into a $^4C_1$ conformation would likely disfavor reactivity on retaining β-glucosidases, whose Michaelis complex is $^1S_3$.

Here, we present the synthesis of a panel of 1,6-cyclophellitol cyclosulfates which favor $^4C_1$ chair conformations. α-Cyclosulfate **5** irreversibly inhibits retaining α-glucosidases with affinity on a par with 1,6-*epi*-Cyclophellitol aziridine **4**, whereas β-cyclosulfate **6** is a substantially weaker β-glucosidase inhibitor compared to its cognate aziridine **2**. 3-D structures of covalent adducts from both "α" and "β" cyclosulfates, bound to representative α- and β-glucosidases, combined with inhibitory kinetic studies and competitive activity based protein profiling (ABPP), demonstrate the high specificity and active center−nucleophile reactivity of these molecules.

The development of selective inhibitors of carbohydrate processing enzymes is of major interest both for the understanding of biological processes involved in glycan processing and in the discovery of new therapeutics.[19] Only minor advances have been observed regarding the development of new irreversible glycosidase inhibitors since the discovery of cyclophellitol[20] (aziridines)[21] or 5-fluoro-[22] and 2-deoxy-2-fluoro-[23] glycosides, the latter some 30 years ago.[24] Cyclophellitol cyclosulfates thus provide a conceptually novel and

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

powerful tool for further study of glycosidases in health and disease.

RESULTS AND DISCUSSION

Conformational Free Energy Landscapes of Cyclophellitol Cyclosulfates. Cyclophellitol cyclosulfates **5** and **6** were conceived as putative inhibitors of retaining glucosidases. To determine the likely conformations of these molecules, we first calculated their conformational free energy landscapes (FELs) using *ab initio* metadynamics (Supporting Information).[25] For both **5** and **6**, the calculated ground state conformation was centered at $^4C_1$, with a wide energy minimum expanding toward the $^2H_3$–E_3–$^4H_3$ region (Figure 2a).

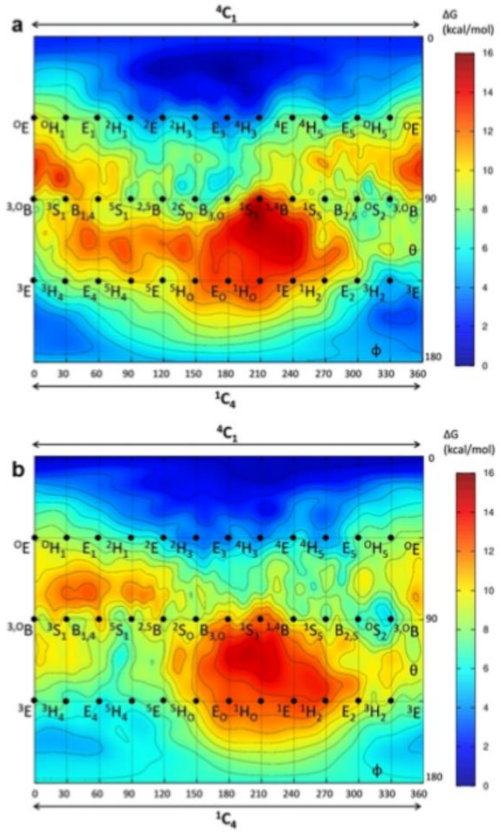


Figure 2. Conformational free energy landscapes of cyclosulfates **5** and **6**. Cyclosulfates **5** (a) and **6** (b) adopt $^4C_1$ ground state conformations, with a broad energy minimum extending toward $^4H_3$. The *x* and *y* axes of each graph correspond to the φ and θ Cremer–Pople puckering coordinates (in degrees), respectively. Isolines are 1 kcal/mol.

We reasoned that the conformational preference of **5** should render it a potent inactivator of α-glucosidases, as its $^4C_1$ conformation matches that of a typical α-glucosidase Michaelis complex, with the $^4H_3$ TS conformation also energetically accessible. In contrast, **6** should not be able to access the $^1S_3$ β-glucosidase Michaelis complex conformation, although the $^4H_3$ TS conformation could still be accessible. Thus, while **6** may still inhibit β-glucosidases, it should do so with less potency compared to **5** vs α-glucosidases.

Synthesis of Cyclic Sulfates 5 and 6. Key intermediate **7** was synthesized from D-xylose in nine steps as optimized in our group based on the total synthesis of cyclophellitol **1** by Madsen and co-workers (Figure 3a).[26,27] Benzylation of the free alcohols in **7** yielded alkene **8**, which was oxidized (ruthenium trichloride/sodium periodate) to afford a mixture of *cis* diols **9** and **10**. Compound **10**, emulating α-glucopyranosides in configuration, was obtained pure after silica gel column chromatography, whereas β-analogue **9** necessitated recrystallization from methanol and diethyl ether in order to obtain homogeneous material. Generation of the cyclic sulfites by treatment of thionyl chloride and subsequent oxidation gave perbenzylated cyclic sulfates **11** and **12**, the benzyl ethers of which were removed by hydrogenolysis using catalytic palladium on carbon to afford compounds **5** and **6**. In line with FEL calculations, analysis of experimental *J*-coupling values for compounds **5** and **6** showed that these analogues indeed adopt a $^4C_1$ conformation in solution (Table S1).

***In Vitro* Inhibition of α- and β-Glucosidases and Kinetic Studies.** To establish the activity of cyclosulfates **5** and **6** in relation to their calculated conformations, we assessed their ability to inhibit representative α- and β-glucosidases compared to cyclophellitol epoxides and aziridines **1**–**4**, utilizing both purified glycosidases and human cell/tissue lysates.

We first tested inhibition against human β-glucosidases GBA1 (recombinant Cerezyme protein from Genzyme, classified into the CAZy glycoside hydrolase family GH30) and GBA2 (GBA2 overexpressing HEK293T lysate, family GH116), and α-glucosidases GAA (recombinant Myozyme from Genzyme, family GH31) and GANAB (Pompe disease fibroblast lysates, family GH31; Table 1). Consistent with our previous findings, cyclophellitol **1** and cyclophellitol aziridine **2** were potent inhibitors of GBA1 and GBA2, and 1,6-*epi*-cyclophellitol **3** and 1,6-*epi*-cyclophellitol aziridine **4** inhibited GAA and GANAB.[10,17,28] In line with our predictions regarding Michaelis complex mimics as strong retaining glycosidase inhibitors, α-cyclosulfate **5** proved to be a potent nanomolar α-glucosidase inhibitor (IC_{50} 82 nM vs GAA; 29 nM vs GANAB) with no reactivity toward β-glucosidases. In contrast, β-cyclosulfate **6** was only a modest inhibitor of β-glucosidases (IC_{50} 119 μM for GBA1 and 58 μM for GBA2). IC_{50}s determined against representative recombinant bacterial glucosidases (β-glucosidases from *Thermotoga maritima*, *Tm*GH1,[3,29] and *Thermoanaerobacterium xylanolyticum*, *Tx*GH116;[30] α-glucosidase from *Cellvibrio japonicus*, *Cj*Agd31B[10,31]) showed the same trend as for the human enzymes: strong inhibition by **5**, and poor inhibition by **6** (Table S2). To confirm that **5** and **6** were stable under the acidic conditions utilized for these enzymatic assays, the solvolytic stability of these compounds in McIlvaine buffer pH 4.0 was assessed by NMR. We observed no spontaneous hydrolysis after 24 h for either **5** or **6** (Figure S2).

We next determined kinetic parameters of inhibition by **1**–**6** against recombinant human GBA1 and GAA, as well as bacterial β-glucosidase *Tm*GH1 (Table 2, Figure S3).[32] Epoxides **1** and **3** irreversibly inhibited GBA and GAA, with initial binding constants (K_I) of 9.2 μM and 1.4 mM, and inactivation rate constants (k_{inact}) of 0.7 and 0.5 min^{-1} respectively. Consistent with their greater reactivity, aziridines **2** and **4** inhibited GBA and GAA considerably faster than epoxides (full inhibition typically within 30 s at higher

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784–793

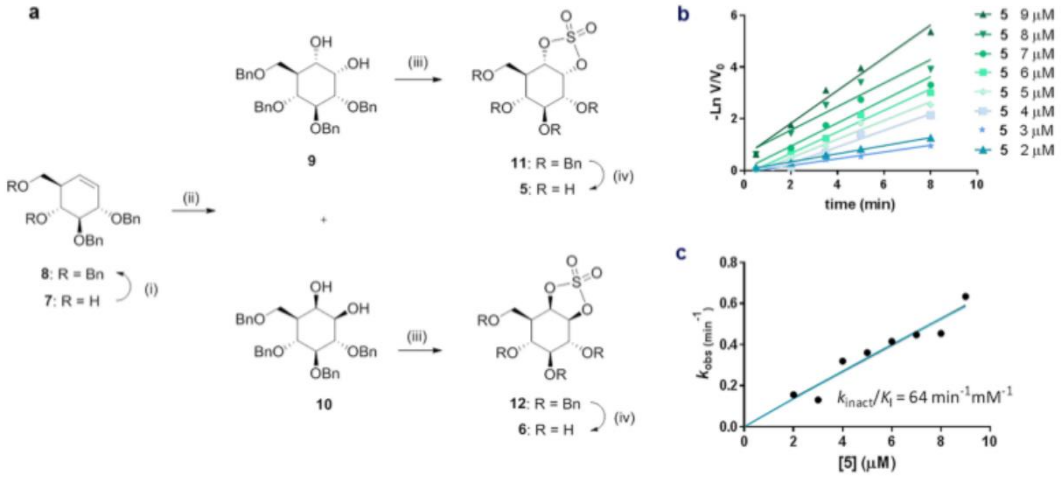


Figure 3. Synthesis of cyclophellitol cyclosulfates **5** and **6**, and inactivation of GAA by compound **5**. (a) Compounds **5** and **6** were prepared from cyclohexene **11**. Reagents and conditions: (i) BnBr, TBAI, NaH, DMF, rt, 18 h, 70%; (ii) $RuCl_3$, $NaIO_4$, EtOAc, ACN, 0 °C, 2 h (**13**, 39%; **14**, 26%); (iii) (a) $SOCl_2$, Et_3N, DCM, 0 °C, (b) $RuCl_3$, $NaIO_4$, CCl_4, ACN, 0 °C, 3 h (**15**, 59%; **16**, 62%); (iv) H_2, Pd/C, MeOH, rt, 18 h (**5**, 71%; **6**, 72%). (b) Semilogarithmic plots of residual activity of GAA versus time at 9, 8, 7, 6, 5, 4, 3, and 2 μM α-cyclosulfate **5**. (c) Plot of pseudo first order rate constants from panel c vs concentration of **5**.

Table 1. Enzyme Inhibition Efficacy of Compounds 1−6[a]

	IC_{50}			
	β-glucosidase		α-glucosidase	
compd	GBA1	GBA2	GAA	GANAB
1	165 ± 4 nM	139 ± 60 nM	>100 μM	>100 μM
2	304 ± 5 nM	14 ± 1 nM	>100 μM	>100 μM
3	>100 μM	>100 μM	15 ± 2 μM	>100 μM
4	21 ± 1 μM	>100 μM	38 ± 3 nM	1.4 ± 0.1 μM
5	>100 μM	>100 μM	82 ± 1 nM	29 ± 2 nM
6	119 ± 9.8 μM	58 ± 4.5 μM	>100 μM	>100 μM

[a]Apparent IC_{50} values for *in vitro* inhibition of GBA1, GBA2, GAA, and GANAB. Reported values are mean ± standard deviation from 3 technical replicates.

concentrations), limiting us to measuring a combined k_{inact}/K_I ratio for these molecules.

α-Cyclosulfate **5** displayed rapid pseudo first order inhibition kinetics against GAA (k_{inact}/K_I = 64 min^{-1} mM^{-1}), comparable to 1,6-*epi*-cyclophellitol aziridine **4** (Figure 3b,c, Table 2), illustrating the potency of α-configured cyclosulfates locked in a Michaelis complex conformation favoring nucleophilic interception. Conversely, irreversible inhibition of GBA1 by **6** was several orders of magnitude slower (K_I = 3.6 mM; k_{inact} = 0.015 min^{-1}; k_{inact}/K_I = 4.33 × 10^{-3} min^{-1} mM^{-1}), suggesting a limited ability of this molecule to adopt the $^1S_3$ conformation required for nucleophilic attack and β-glucosidase reactivity. Comparison of K_I values for **6** and **1** against GBA1 also showed initial binding of **6** to the enzyme to be substantially weaker compared to **1** (3.5 mM and 9 μM respectively). These values are consistent with the tighter binding affinities often displayed by TS shape analogues compared to molecules locked into other conformations.

Interestingly, while irreversible inhibition by aziridines **2** and **4** was largely specific toward α- or β-glucosidases respectively, these molecules were also observed to be less specific *reversible* inhibitors for glucosidases of the opposite anomeric specificity (competitive inhibition of GAA by **2**, K_I = 9.6 μM; competitive inhibition of GBA1 by **4**, K_I = 6.8 μM; Table 2). We reason that the common $^4H_3$ transition state utilized by both α- and β-glucosidases may allow these enzymes to bind aziridines of either "anomeric" configuration. Thus, the conformational preference of cyclophellitol cyclosulfates presents an advantage in specificity when compared to aziridine type inhibitors, by virtue of mimicking a reactive conformation receptive to nucleophilic attack, not shared between α- or β-glucosidases.

Structural Characterization of Enzyme−Cyclosulfate Interactions. Because of the key role of conformation in the contrasting inhibitory potency of cyclophellitol cyclosulfates

Table 2. Kinetic Parameters of Compounds 1−6[a]

	β-glucosidase GBA1			α-glucosidase GAA		
compd	k_{inact} (min^{-1})	K_I	k_{inact}/K_I (min^{-1} mM^{-1})	k_{inact} (min^{-1})	K_I	k_{inact}/K_I (min^{-1} mM^{-1})
1	0.72	9.2 μM	77.7		>500 μM	
2	nd	nd	19.0		9.6 μM*	
3		>500 μM		0.52	1.4 mM	0.37
4		6.8 μM*		nd	nd	58.0
5		>500 μM		nd	nd	64.3
6	0.015	3.6 mM	4.3 × 10^{-3}		>500 μM	

[a]Inactivation rates and inhibition constants (k_{inact} and K_I) for human recombinant β-glucosidase GBA1 and α-glucosidase GAA; nd, not determined due to fast inhibition; *, reversible inhibition observed.

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

against α- vs β-glucosidases, we set out to structurally characterize the conformational itinerary of **5** and **6** when reacted with representative bacterial glucosidases.

Crystal structures of **5** in complex with the GH31 α-glucosidase/tranglycosidase *Cj*Agd31B from *Cellvibrio japonicus*[10,31] were obtained by soaking crystals of wild type *Cj*Agd31B with **5**. We observed density for a covalent adduct bound to the enzyme nucleophile (Asp412) corresponding to ring opened cyclosulfate in a $^{1}S_{3}$ conformation (Figure 4a), matching the conformation previously observed for reacted aziridine **4**,[10] as well as a complex of *Cj*Agd31B reacted with 5-fluoro-α-glucosyl fluoride[31] (Figure 4b; PDB codes 5I24 and 4BA0 respectively). Structures of unreacted **5** were also obtained in complex with an inactive D412N nucleophile mutant of *Cj*Agd31B. Density for unreacted **5** in the *Cj*Agd31B active site matched a $^{4}C_{1}$ conformation as predicted by FEL calculations and NMR analysis (Figure 4a), demonstrating that **5** indeed binds to α-glucosidases as a conformational Michaelis complex analogue, allowing unimpeded access for in-line attack by the enzyme nucleophile. Difference density in the active site of *Cj*Agd31B D412N after soaking with **4** corresponded to an unreacted cyclophellitol aziridine in $^{4}H_{3}$ conformation (Figure 4b), consistent with this molecule acting as a TS shape mimic when unreacted.

We then utilized the β-glucosidase from *Thermoanaerobacterium xylanolyticum* *Tx*GH116 for structural studies with **6**. Due to its slow reactivity as an irreversible β-glucosidase inhibitor, we were able to capture an unreacted complex of **6** with crystals of wild type *Tx*GH116 by utilizing shorter ligand soaking times. After 10−20 min soaks, a single molecule of unreacted **6** was observed within the active site of *Tx*GH116, adopting a $^{4}C_{1}$ conformation in line with FEL calculations and NMR spectra. This $^{4}C_{1}$ conformation renders the equatorial C1−O bond of **6** poorly poised for in-line attack by the enzyme nucleophile Glu441, thus providing a structural rationale for the slow reactivity of this inhibitor against β-glucosidases. Extended ligand soaking times (~24 h) were required to produce a reacted covalent complex of **6** with *Tx*GH116. Reacted **6** was observed bound to the *Tx*GH116 catalytic nucleophile (Glu540) in a $^{4}C_{1}$ covalent intermediate conformation, consistent with the typical retaining β-glucosidase conformational itinerary utilized by this enzyme and matching the $^{4}C_{1}$ conformation previously observed for reacted 2-deoxy-2-fluoroglucoside (PDB code 5BX2).

We also investigated the structural basis for "off-target" inhibition by cyclophellitol aziridines, which we had observed to competitively inhibit glucosidases of "opposing" anomeric specificity. A crystal structure of **2** in complex with *Cj*Agd31B showed a single molecule of inhibitor occupying the enzyme active site, with the aziridine nitrogen H-bonding to the catalytic nucleophile carboxylate (~2.6 Å between aziridine N and Asp O Oδ; Figure S5). The presence of an unreacted aziridine between the nucleophile and the plane of the cyclitol ring induces an E_{3} conformation for the inhibitor, rather than the $^{4}H_{3}$ observed for an unreacted "correct" aziridine (Figure 4b). It is likely that the competitive inhibition of α-glucosidases by **4** is based around a similar interaction between the unreacted aziridine nitrogen and catalytic nucleophile.

Competitive ABPP of α-Glucosidases by Cyclosulfate 5. To assess the utility of cyclosulfates as enzyme inhibitors in complex biological samples, we examined the activity and selectivity of α-cyclosulfate **5** by competitive activity-based protein profiling (ABBP), against **13**[10] (Figure S1), a

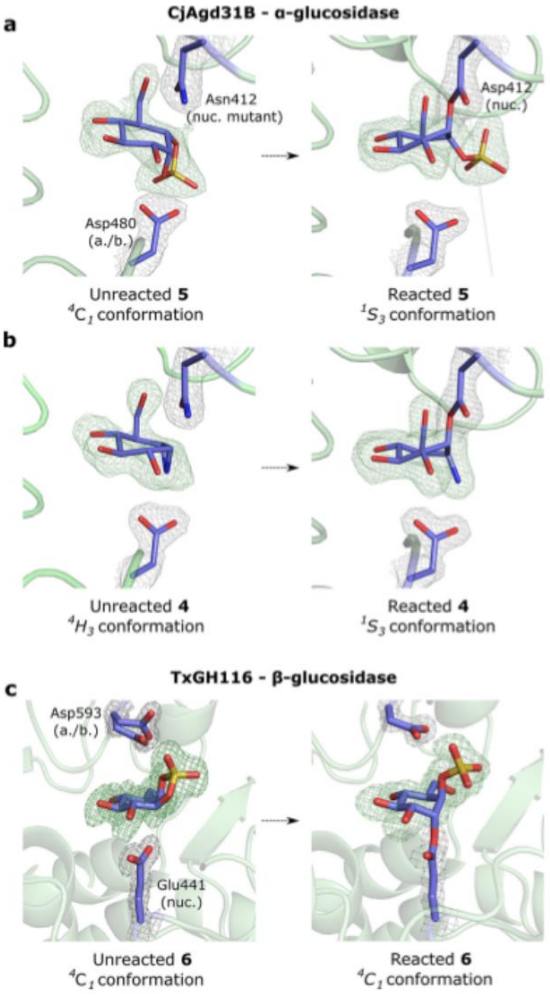


Figure 4. Structures of reacted **4** and **5** bound to wild type *Cj*Agd31B and reacted **6** bound to wild type *Tx*GH116. (a) Unreacted (left) and reacted (right) **5** in complex with *Cj*Agd31B D412N nucleophile mutant and wt *Cj*Agd31B respectively. Unreacted **5** adopts a $^{4}C_{1}$ ring conformation in the active site of *Cj*Agd31B, mimicking the Michaelis complex conformation of GH31 α-glucosidase substrates. Reacted **5** adopts a $^{1}S_{3}$ covalent intermediate conformation. (b) Unreacted (left) and reacted (right) **4** in complex with *Cj*Agd31B D412N nucleophile mutant and wt *Cj*Agd31B respectively. Unreacted **4** adopts a $^{4}H_{3}$ TS conformation. Reacted **4** adopts the same $^{1}S_{3}$ intermediate conformation as observed for **5**. (c) Unreacted (left) and reacted (right) β-cyclosulfate **6** with *Tx*GH116. Unreacted **6** in complex with GH116 β-glucosidase *Tx*GH116 adopts a $^{4}C_{1}$ conformation in the enzyme active site, which is poorly poised for attack by the enzyme nucleophile Glu441, and thus reacts extremely slowly. (Two conformations for the enzyme catalytic acid/base Asp593 can be seen in this complex.) Reacted **6** can be observed after extended soaking, and also adopts a $^{4}C_{1}$ conformation, covalently bound to the enzyme nucleophile. Electron density for protein side chains is *REFMAC* maximum-likelihood/σ_{A}-weighted $2F_{o} - F_{c}$ contoured to 0.44−0.51 and 0.40−0.42 electron/Å^{3} for *Cj*Agd31B and *Tx*GH116 complexes, respectively. Electron density for ligand is $F_{o} - F_{c}$ maps calculated just prior to building in ligand, contoured to 0.17−0.26 and 0.27−0.35 electron/Å^{3} for *Cj*Agd31B and *Tx*GH116 complexes, respectively. nuc. = nucleophile; a./b. = acid/base.

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

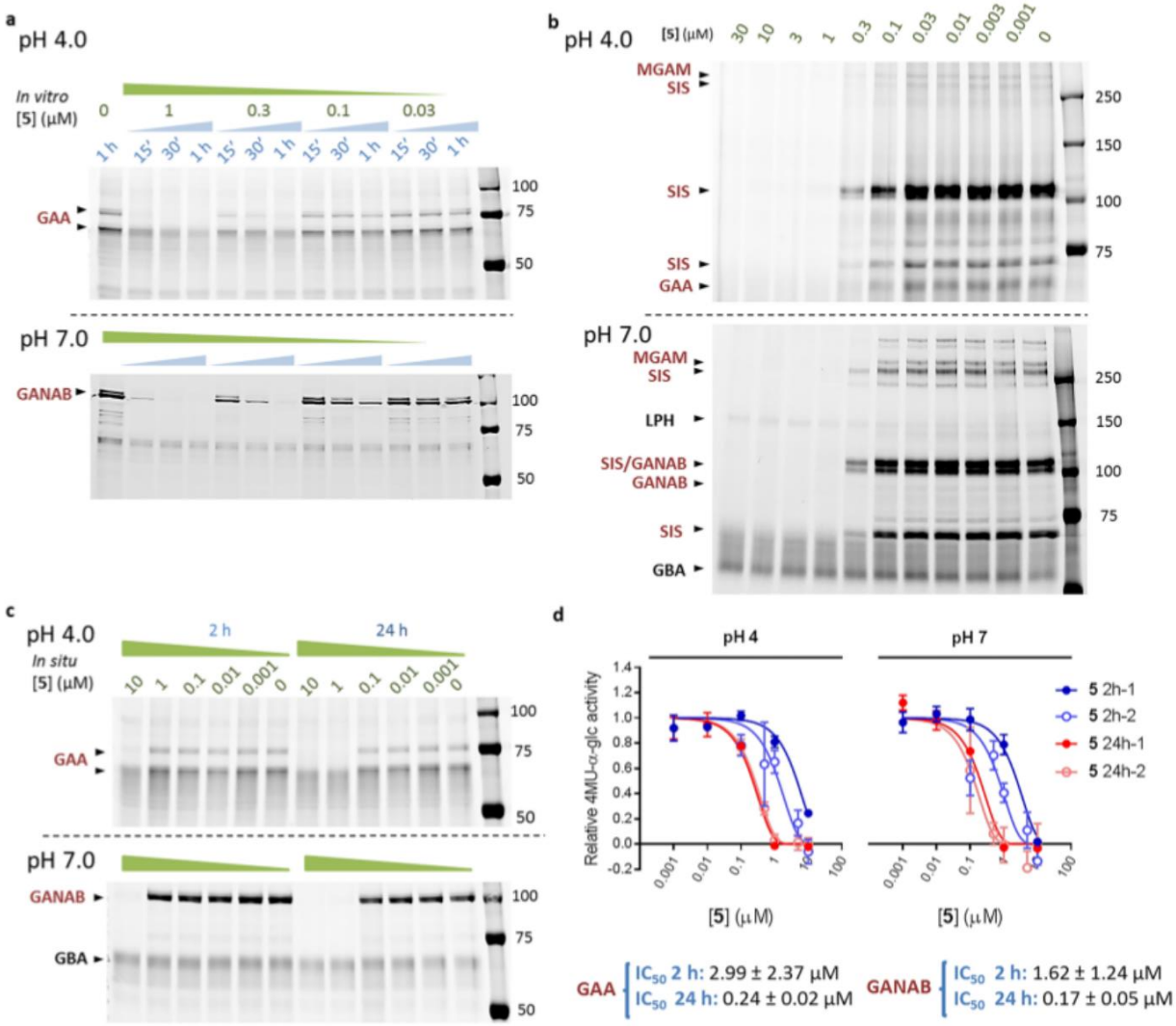


Figure 5. *In vitro* and *in situ* inhibition of GAA and GANAB. (a) **5** inhibits labeling of GAA and GANAB by fluorescent ABP **13** in fibroblast lysates in a concentration and time dependent manner. (b) **5** inhibits labeling of several α-glucosidases in mouse intestine lysate by **13**. Sucrase-isomaltase (Sis), maltase-glucoamylase (MGAM), GAA, and GANAB are labeled by **13**, as well as some off-target β-glucosidase labeling (LPH and GBA). α-Glucosidase labeling can be abrogated by preincubation with **5**, while β-glucosidase labeling persists, demonstrating the superior selectivity of **5** compared to **13**. (c) *In situ* inhibition of GAA and GANAB in fibroblasts at pH 4.0 and 7.0 respectively by **5** at incubation times of 2 and 24 h, followed by labeling of GAA and GANAB by cyclophellitol aziridine Cy5 probe **13**. (d) Apparent IC_{50}s for *in situ* inhibition of GAA and GANAB enzyme activity by **5**. Reported IC_{50}s are mean ± standard deviation from two biological replicates, each with three technical replicates.

fluorescent retaining α-glucosidase aziridine probe which labels GAA at pH 4 (isoforms at 70 and 76 kDa) and both isoforms of GANAB (~100 kDa) at pH 7.

We first preincubated fibroblast lysates with varying concentrations (1, 0.3, 0.1, and 0.03 μM) of **5** for 15, 30, or 60 min at pH 4 or pH 7, and subsequently labeled with **13** (1 μM) for 30 min. SDS−PAGE followed by fluorescence scanning showed time dependent inhibition of both GAA and GANAB by **5** at pH 4 and 7 respectively (Figure 5a). Competitive ABPP in GBA2 and GBA3 overexpressing HEK293T lysates (GBA2+ and GBA3+) against β-glucosidase probe **14** (Figure S1) showed no inhibition of retaining β-glucosidases by **5**, demonstrating the high selectivity of this inhibitor class (Figure S5).

Acarbose, miglitol, and voglibose are α-glucosidase inhibitors (AGIs) widely used in diabetes mellitus type II patients. These inhibitors delay the absorption of carbohydrates, decrease postprandial hyperglycemia and hyperinsulinemia, and thus improve insulin sensitivity and release stress on beta cells.[33−35] Prompted by the potential of AGIs as leads for diabetes mellitus type II drug development, we investigated the inhibition of a group of key metabolic α-glucosidases in murine gastrointestinal tract tissues using competitive ABPP. Our studies showed that, in addition to GAA and GANAB, **5** inhibited two α-glucosidases expressed specifically in intestinal tissue: sucrase-isomaltase (SIS) and maltase-glycoamylase (MGAM) (Figure 5b). Significantly, β-glucosidases lactase-phlorizin hydrolase (LPH) and GBA1, which are labeled in an off-target fashion by

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

13, are not outcompeted by preincubation with **5**, clearly demonstrating superior selectivity of cyclosulfate inhibitors compared to aziridines.

Lastly, we investigated whether α-cyclosulfate **5** is able to cross the cell membrane. Human fibroblasts were exposed to varying concentrations (10, 1, 0.1, 0.01, 0.001, and 0 μM) of **5** for 2 and 24 h, followed by extensive washing and lysate preparation. Lysates were incubated with **13** (0.5 μM) for 30 min and then separated by SDS−PAGE, followed by fluorescence scanning. Time dependent abrogation of ABP **13** fluorescence was observed, illustrating that **5** could cross the cell membrane to act *in situ* as α-glucosidase inhibitor (Figure 5c). Loss of ABP labeling was also accompanied by a corresponding loss of α-glucosidase enzymatic activity in harvested fibroblast lysates. Activity assays utilizing 4-methylumbelliferyl-α-glucose suggested robust *in situ* inhibition of both GAA and GANAB by **5**, with IC_{50}s of 3.0 and 1.6 μM respectively (Figure 5d).

■ CONCLUSION

In this study, we have presented a new class of irreversible glycosidase inhibitors: cyclophellitol cyclosulfates. We have shown that their irreversible action is specific for α- and β-glucosidases in a manner reflecting both the "anomeric" stereochemistry of the cyclophellitol cyclosulfates and the conformational itinerary of the target enzyme.

We have demonstrated, through *ab initio* metadynamics approaches and 3-D structures of enzyme ligand complexes, that unreacted α-cyclosulfate **5** favors a $^4C_1$ "Michaelis complex like" conformation which is perfectly poised for nucleophilic attack by α-glucosidases utilizing a $^4C_1 \rightarrow {}^4H_3 \rightarrow {}^1S_3$ glycosylation itinerary. Thus, **5** is a nanomolar mechanism-based retaining α-glucosidase inactivator, which follows pseudo first order kinetics, and displays superior selectivity compared to existing mechanism-based inhibitors due to its unique conformational behavior. We have also shown that **5** inhibits the intestinal retaining α-glucosidases SIS, MGAM, GAA, and GANAB, without affecting β-glucosidases intestinal LPH or GBA1. Thus, **5**, which is also able to cross the cell membrane, may be a good starting point for the development of agents for the treatment of type II diabetes.

β-Cyclosulfate **6** also adopts a $^4C_1$ conformation in its unreacted state, which does not match the $^1S_3$ conformation of typical β-glucosidase Michaelis complexes. When applied to retaining β-glucosidases following a $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ conformational itinerary, **6** reacts extremely slowly compared to its congener **5**.

FEL calculations show that the ground state conformation of **6** is centered around $^4C_1$, a conformation with an equatorial C1−O bond that precludes in-line attack from the nucleophile. Crystal structures of unreacted **6** in complex with β-glucosidase *Tx*GH116 show that the β-cyclosulfate continues to adopt this $^4C_1$ conformation even within the enzyme active site. However, FEL calculations also suggest that it is energetically possible for **6** to access a TS-like $^4H_3$ conformation. Transient sampling of this TS conformation may allow for occasional reaction of **6** with β-glucosidases, as the cyclosulfate adopts a conformation better suited to nucleophilic attack by the enzyme. However, the slow rate of irreversible inhibition observed for **6** suggests that its predominant conformation within the enzyme active site is one where nucleophilic attack from the enzyme is disfavored.

The last 10−20 years have provided us a deep appreciation of reaction coordinates of diverse glycosidases.[6,7,10,36] This conformational canvas inspires the design and application of chemical species to mimic specific species along the reaction coordinate. Much existing work has focused on the design of noncovalent transition state mimics, or conformationally restricted species that mimic a species along only certain reaction trajectories, such as the specific inhibition of GH47 α-mannosidases by kifunensine.[37,38] Recent years have, however, seen a resurgence in the development of irreversible enzyme inhibitors as clinical probes and diagnostics.[10,11,16,39] With such applications in mind, the development of conformationally restrained irreversible inhibitors primed for nucleophilic attack through mimicry of the Michaelis complex, coupled to more electrophilic reaction centers, inspires creation of conceptually new, potent and selective glycosidase inhibitors. We envision that our design strategy will also be applied for other glycosidases, both those following a similar $^4C_1 \rightarrow {}^4H_3 \rightarrow {}^1S_3$ conformational itinerary and those following other conformational trajectories, guided by the conformational preference of the reaction pathway.

■ MATERIALS AND METHODS

All chemicals were obtained from Sigma-Aldrich, unless otherwise stated. Pompe disease fibroblasts, HEK293T cells (ATCC-CRL-3216), and normal fibroblast cell lines were obtained from the American Type Culture Collection (ATCC) and transfected for GBA2 and GBA3 overexpression (Supporting Information). Cell lines were cultured in DMEM/F-12 (Ham) medium (Invitrogen) supplemented with 10% (v/v) fetal calf serum (FCS; Sigma) and 1% penicillin/streptomycin (Sigma). Mouse tissues were isolated according to guidelines approved by the ethical committee of Leiden University (DEC#13191). Human recombinant enzymes rGBA1 (Cerezyme) and rGAA (Myozyme) were donated by Genzyme, GBA2 was overexpressed in HEK293T lysates, and GANAB was obtained from fibroblasts of Pompe patients diagnosed on the basis of absence of GAA (Supporting Information). Bacterial enzymes *Tm*GH1,[29] *Tx*GH116,[30] and *Cj*Agd31B[31] were expressed as previously described. All cell or tissue lysates were prepared in KPI buffer (25 mM potassium phosphate pH 6.5, supplemented with protease inhibitor 1× cocktail (Roche)) via homogenization on ice with SilentCrusher S equipped with Typ 7 F/S head (30 × 1000 rpm, 3 × 7 s). Lysate protein concentrations were determined with BCA Protein Assay Kit (Pierce). Lysates and proteins were stored in small aliquots at −80 °C until use.

IC_{50} and Inhibition Kinetics Determination. Detailed protocols for IC_{50} and inhibition kinetics measurements can be found in the Supporting Information.

In brief, apparent IC_{50} values were determined by preincubating enzymes with a range of inhibitor concentrations in a 25 μL volume for 30 min at 37 °C (for human enzymes) or 25 °C (for bacterial enzymes). Following preincubation, 25 μL of the enzyme−inhibitor mixture was transferred into 100 μL of the appropriate 4-MU−Glc substrate mixture to determine residual activity. Reactions were quenched with 1 M NaOH−glycine (pH 10.3) upon completion, and 4-MU fluorescence was measured with a LS55 fluorescence spectrophotometer (PerkinElmer) (λ_{EX} 366 nm; λ_{EM} 445 nm). IC_{50} values reported are the mean values from three technical replicates.

In situ IC_{50} values were determined by incubating normal human dermal fibroblasts (grown to confluency) with **5** for 2

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

and 24 h. Cells were washed three times with PBS and harvested by scraping into KPI buffer supplemented with 0.1% (v/v) Triton X-100 and 1× cOmplete protease inhibitor cocktail (Roche). Residual GAA and GANAB activity was measured on the basis of hydrolysis of 4-MU−α-Glc at pH 4 or 7. *In situ* IC_{50} values are mean values from two biological replicates, each with 3 technical repeats (Figure 5d).

For kinetics, enzyme and relevant concentrations of inhibitor were preincubated for 0.5, 2, 3.5, 5, and 8 min for fast inhibitors and 0.5, 20, 60, 120, 180, 240, and 360 min for slow inhibitors. At relevant time points, 5 μL of this enzyme−inhibitor mixture was added to 2,4-DNP−Glc substrate, and release of 2,4-dinitrophenolate was monitored via absorbance at 400 nm to determine the rate of hydrolysis after inhibition (V) compared to a no inhibitor control (V_0).

Pseudo first order rate constants (k_{obs}) were obtained from a linear fit of $-\ln V/V_0$ against time for each value of [I]. A plot of k_{obs} against [I] fitted to the hyperbolic equation $k_{obs} = (k_{inact}[I]/K_I + [I])$ was used to determine k_{inact} and K_I. Where fast inhibition was observed (>50% after 30 s), the k_{inact}/K_I ratio was determined at low values of [I] ($\ll K_I$), using the approximation $k_{obs} \approx k_{inact}[I]/K_I$, where k_{inact}/K_I is the slope of a linear fit of k_{obs} vs [I]. Where reversible inhibition was observed (no variation of $-\ln V/V_0$ with time), K_I was determined by use of Lineweaver−Burk plots at different values of [I].

Time and Concentration Dependent *in Vitro* Inhibition of GAA and GANAB. Homogenates of human normal dermal fibroblast (17.6 μg total protein) were preincubated on ice for 5 min in 150 mM McIlvaine buffer pH 4 or 7 in a total volume of 10 μL. Samples were then incubated for 15, 30, and 60 min at 37 °C with 2.5 μL inhibitor dilutions in McIlvaine buffer pH 4 or 7 to obtain a final concentration of 5, 1.5, 0.5, 0.15, and 0 μM. Afterward, the samples were further incubated with 2.5 μL of ABP **13** (1 μM), diluted in McIlvaine buffer pH 4 or 7 for 30 min at 37 °C. Finally, samples were denatured with 5 μL of 4× Laemmli sample buffer for 5 min at 98 °C, and resolved by 10% (w/v) SDS−PAGE. Wet slab gels were subjected to Cy5 fluorescence scaning (Typhoon FLA9500, GE, $\lambda_{EX} \geq 635$ nm, $\lambda_{EM} \geq 665$ nm).

***In Vitro* Inhibition of α-Glucosidases in Mouse Gastrointestinal Lysate.** Mixtures of mouse duodenum, jejunum, and ileum homogenates containing 40 μg total protein were incubated on ice for 5 min with McIlvaine buffer at pH 4 or 7, in a total volume of 10 μL. Samples were subsequently incubated with 2.5 μL of **5** for 1 h at 37 °C at various inhibitor dilutions in McIlvaine buffer at pH 4.0 or 7.0 to obtain a final concentration of 150, 50, 15, 5, 1.5, 0.5, 0.15, 0.05, 0.015, 0.005, and 0 μM. Then, the samples were further incubated with 2.5 μL of **13** (6 μM, diluted in McIlvaine buffer at pH 4.0 or 7.0) for 30 min at 37 °C. Samples were denatured by boiling with 3.75 μL of 4× Laemmli sample buffer for 5 min, and resolved by 10% (w/v) SDS−PAGE. Wet slab gels were subjected to Cy5 fluorescence scanning (Typhoon FLA9500, GE, $\lambda_{EX} \geq 635$ nm, $\lambda_{EM} \geq 665$ nm).

***In Situ* inhibition of GAA and GANAB.** Human normal dermal fibroblasts were grown to confluence and *in situ* treated with inhibitor **5** at various concentrations (10, 1, 0.1, 0.01, 0.001, and 0 μM) in duplicates for 2 or 24 h. Cells were harvested by first washing three times with PBS and subsequently lysed with KPI buffer (25 mM K_2HPO_4/KH_2PO_4, pH 6.5, 0.1% (v/v) Triton X-100, protease inhibitor cocktail (Roche)) on ice for 30 min. Homogenates were collected by scraping, vortexed, and stored at −80 °C. Lysates containing 4 μg total protein were equilibrated to pH 4 or 7 in 150 mM McIlvaine buffer for 5 min on ice, and incubated with probe **13** (0.5 μM) for 30 min at 37 °C at either pH 4 or 7 in a total volume of 15 μL. Samples were denatured, resolved by SDS−PAGE, and subjected to fluorescent scan (Typhoon FLA9500, GE, $\lambda_{EX} \geq 635$ nm, $\lambda_{EM} \geq 665$ nm)

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.7b00214.

Synthesis and characterization of **5** and **6**, synthesis of all intermediates, molecular modeling, labeling of glucosidases, SDS−PAGE analysis and fluorescence scanning, and crystallographic data collection and refinement statistics (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: gideon.davies@york.ac.uk.
*E-mail: h.s.overkleeft@chem.leidenuniv.nl.

ORCID

Marta Artola: 0000-0002-3051-3902
Jeroen D. C. Codée: 0000-0003-3531-2138
Carme Rovira: 0000-0003-1477-5010
Gideon J. Davies: 0000-0002-7343-776X
Herman S. Overkleeft: 0000-0001-6976-7005

Author Contributions

M.A. synthesized the inhibitors under the guidance of J.D.C.C. and G.A.v.d.M. M.A., M.J.F., and C.-L.K. conducted the enzyme inhibition and biochemistry assays under the guidance of J.M.F.G.A.. L.W., W.A.O., I.Z.B., and G.J.D. designed and performed the X-ray diffraction experiments. L.R. performed metadynamics simulations under guidance of C.R. H.S.O. and G.J.D. conceived the idea, supervised the project, and with M.A. and L.W. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank The Netherlands Organization for Scientific Research (NWO-CW, ChemThem grant to J.M.F.G.A. and H.S.O.), the European Research Council (ERC-2011-AdG-290836 "Chembiosphing", to H.S.O., and ERC-2012-AdG-322942 "Glycopoise", to G.J.D.), Sanofi Genzyme (research grant to J.M.F.G.A. and H.S.O. and postdoctoral contract to M.A.), the Spanish Ministry of Economy and Competitiveness (CTQ2014-55174-P), and the Generalitat de Catalunya (2014SGR-987) for financial support. We thank the Diamond Light Source for access to beamline i02 and i04 (proposal number mx-13587) and the Barcelona Supercomputing Center (BSC-CNS, RES-QCM-2017-1-00014) for technical support and computational resources provided, which contributed to the results presented here. L.R. acknowledges a Ph.D. fellowship from the University of Barcelona (APIF-UB). G.J.D. is supported by the Royal Society though the Ken Murray Research Professorship.

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

REFERENCES

(1) Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490−D495.

(2) Jongkees, S. A. K.; Withers, S. G. Unusual Enzymatic Glycoside Cleavage Mechanisms. *Acc. Chem. Res.* **2014**, *47*, 226−235.

(3) Gloster, T. M.; Davies, G. J. Glycosidase inhibition: assessing mimicry of the transition state. *Org. Biomol. Chem.* **2010**, *8*, 305−20.

(4) Vasella, A.; Davies, G. J.; Bohm, M. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **2002**, *6*, 619−629.

(5) Vocadlo, D. J.; Davies, G. J. Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539−555.

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

(7) Davies, G. J.; Williams, S. J. Carbohydrate-active enzymes: sequences, shapes, contortions and cells. *Biochem. Soc. Trans.* **2016**, *44*, 79−87.

(8) Kallemeijn, W. W.; Witte, M. D.; Wennekes, T.; Aerts, J. M. Mechanism-based inhibitors of glycosidases: design and applications. *Adv. Carbohydr. Chem. Biochem.* **2014**, *71*, 297−338.

(9) Beenakker, T. J. M.; Wander, D. P. A.; Offen, W. A.; Artola, M.; Raich, L.; Ferraz, M. J.; Li, K. Y.; Houben, J.; van Rijssel, E. R.; Hansen, T.; van der Marel, G. A.; Codee, J. D. C.; Aerts, J.; Rovira, C.; Davies, G. J.; Overkleeft, H. S. Carba-cyclophellitols Are Neutral Retaining-Glucosidase Inhibitors. *J. Am. Chem. Soc.* **2017**, *139*, 6534−6537.

(10) Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **2014**, *28*, 1−13.

(11) Witte, M. D.; Kallemeijn, W. W.; Aten, J.; Li, K. Y.; Strijland, A.; Donker-Koopman, W. E.; van den Nieuwendijk, A. M.; Bleijlevens, B.; Kramer, G.; Florea, B. I.; Hooibrink, B.; Hollak, C. E.; Ottenhoff, R.; Boot, R. G.; van der Marel, G. A.; Overkleeft, H. S.; Aerts, J. M. Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat. Chem. Biol.* **2010**, *6*, 907−913.

(12) Jiang, J.; Beenakker, T. J.; Kallemeijn, W. W.; van der Marel, G. A.; van den Elst, H.; Codee, J. D.; Aerts, J. M.; Overkleeft, H. S. Comparing Cyclophellitol *N*-Alkyl and *N*-Acyl Cyclophellitol Aziridines as Activity-Based Glycosidase Probes. *Chem. - Eur. J.* **2015**, *21*, 10861−10869.

(13) Withers, S. G.; Umezawa, K. Cyclophellitol: a naturally occurring mechanism-based inactivator of beta-glucosidases. *Biochem. Biophys. Res. Commun.* **1991**, *177*, 532−537.

(14) Gloster, T. M.; Madsen, R.; Davies, G. J. Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Org. Biomol. Chem.* **2007**, *5*, 444−446.

(15) Willems, L. I.; Beenakker, T. J.; Murray, B.; Scheij, S.; Kallemeijn, W. W.; Boot, R. G.; Verhoek, M.; Donker-Koopman, W. E.; Ferraz, M. J.; van Rijssel, E. R.; Florea, B. I.; Codee, J. D.; van der Marel, G. A.; Aerts, J. M.; Overkleeft, H. S. Potent and selective activity-based probes for GH27 human retaining alpha-galactosidases. *J. Am. Chem. Soc.* **2014**, *136*, 11622−11625.

(16) Jiang, J. B.; Kallemeijn, W. W.; Wright, D. W.; van den Nieuwendijk, A. M. C. H.; Rohde, V. C.; Folch, E. C.; van den Elst, H.; Florea, B. I.; Scheij, S.; Donker-Koopman, W. E.; Verhoek, M.; Li, N.; Schurmann, M.; Mink, D.; Boot, R. G.; Codee, J. D. C.; van der Marel, G. A.; Davies, G. J.; Aerts, J. M. F. G.; Overkleeft, H. S. In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 alpha-L-fucosidases. *Chem. Sci.* **2015**, *6*, 2782−2789.

(17) Li, K. Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; Donker-Koopman, W. E.; Boot, R. G.; Aerts, J. M.; Codee, J. D.; van der Marel, G. A.; Overkleeft, H. S. Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors. *Org. Biomol. Chem.* **2014**, *12*, 7786−7791.

(18) Gao, Y.; Sharpless, K. B. Vicinal Diol Cyclic Sulfates - Like Epoxides Only More Reactive. *J. Am. Chem. Soc.* **1988**, *110*, 7538−7539.

(19) Ernst, B.; Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nat. Rev. Drug Discovery* **2009**, *8*, 661−677.

(20) Atsumi, S.; Umezawa, K.; Iinuma, H.; Naganawa, H.; Nakamura, H.; Iitaka, Y.; Takeuchi, T. Production, isolation and structure determination of a novel beta-glucosidase inhibitor, cyclophellitol, from Phellinus sp. *J. Antibiot.* **1990**, *43*, 49−53.

(21) Tatsuta, K.; Niwata, Y.; Umezawa, K.; Toshima, K.; Nakata, M. Syntheses and enzyme inhibiting activities of cyclophellitol analogs. *J. Antibiot.* **1991**, *44*, 912−914.

(22) McCarter, J. D.; Withers, S. G. 5-fluoro glycosides: A new class of mechanism-based inhibitors of both alpha- and beta-glucosidases. *J. Am. Chem. Soc.* **1996**, *118*, 241−242.

(23) Withers, S. G.; Street, I. P.; Bird, P.; Dolphin, D. H. 2-Deoxy-2-Fluoroglucosides - a Novel Class of Mechanism-Based Glucosidase Inhibitors. *J. Am. Chem. Soc.* **1987**, *109*, 7530−7531.

(24) Rempel, B. P.; Withers, S. G. Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology* **2008**, *18*, 570−586.

(25) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(26) Hansen, F. G.; Bundgaard, E.; Madsen, R. A short synthesis of (+)-cyclophellitol. *J. Org. Chem.* **2005**, *70*, 10139−10142.

(27) Li, K.-Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; van den Elst, H.; Wong, C.-S.; Chander, S. D.; Hoogendoorn, S.; Beenakker, T. J. M.; Codée, J. D. C.; Aerts, J. M. F. G.; van der Marel, G. A.; Overkleeft, H. S. Synthesis of Cyclophellitol, Cyclophellitol Aziridine, and Their Tagged Derivatives. *Eur. J. Org. Chem.* **2014**, *2014*, 6030−6043.

(28) Jiang, J.; Kuo, C. L.; Wu, L.; Franke, C.; Kallemeijn, W. W.; Florea, B. I.; van Meel, E.; van der Marel, G. A.; Codee, J. D.; Boot, R. G.; Davies, G. J.; Overkleeft, H. S.; Aerts, J. M. Correction to "Detection of Active Mammalian GH31 alpha-Glucosidases in Health and Disease Using In-Class, Broad-Spectrum Activity-Based Probes". *ACS Cent. Sci.* **2016**, *2*, 351−358.

(29) Gloster, T. M.; Meloncelli, P.; Stick, R. V.; Zechel, D.; Vasella, A.; Davies, G. J. Glycosidase inhibition: An assessment of the binding of 18 putative transition-state mimics. *J. Am. Chem. Soc.* **2007**, *129*, 2345−2354.

(30) Charoenwattanasatien, R.; Pengthaisong, S.; Breen, I.; Mutoh, R.; Sansenya, S.; Hua, Y.; Tankrathok, A.; Wu, L.; Songsiriritthigul, C.; Tanaka, H.; Williams, S. J.; Davies, G. J.; Kurisu, G.; Cairns, J. R. Bacterial Beta-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). *ACS Chem. Biol.* **2016**, *11*, 1891−1900.

(31) Larsbrink, J.; Izumi, A.; Hemsworth, G. R.; Davies, G. J.; Brumer, H. Structural Enzymology of Cellvibrio japonicus Agd31B Protein Reveals alpha-Transglucosylase Activity in Glycoside Hydrolase Family 31. *J. Biol. Chem.* **2012**, *287*, 43288−43299.

(32) Islam, M. R.; Grubb, J. H.; Sly, W. S. C-terminal processing of human beta-glucuronidase. The propeptide is required for full expression of catalytic activity, intracellular retention, and proper phosphorylation. *J. Biol. Chem.* **1993**, *268*, 22627−22633.

(33) Scheen, A. J. Is there a role for alpha-glucosidase inhibitors in the prevention of type 2 diabetes mellitus? *Drugs* **2003**, *63*, 933−951.

(34) van de Laar, F. A.; Lucassen, P. L.; Akkermans, R. P.; van de Lisdonk, E. H.; Rutten, G. E.; van Weel, C. Alpha-glucosidase inhibitors for patients with type 2 diabetes: results from a Cochrane systematic review and meta-analysis. *Diabetes Care* **2005**, *28*, 154−163.

(35) Wardrop, D. J.; Waidyarachchi, S. L. Synthesis and biological activity of naturally occurring alpha-glucosidase inhibitors. *Nat. Prod. Rep.* **2010**, *27*, 1431−1468.

(36) Davies, G. J.; Planas, A.; Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **2012**, *45*, 308−316.

(37) Vallee, F.; Karaveg, K.; Herscovics, A.; Moremen, K. W.; Howell, P. L. Structural basis for catalysis and inhibition of *N*-glycan processing class I alpha 1,2-mannosidases. *J. Biol. Chem.* **2000**, *275*, 41287−41298.

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

(38) Thompson, A. J.; Dabin, J.; Iglesias-Fernandez, J.; Ardevol, A.; Dinev, Z.; Williams, S. J.; Bande, O.; Siriwardena, A.; Moreland, C.; Hu, T. C.; Smith, D. K.; Gilbert, H. J.; Rovira, C.; Davies, G. J. The reaction coordinate of a bacterial GH47 alpha-mannosidase: a combined quantum mechanical and structural approach. *Angew. Chem., Int. Ed.* **2012**, *51*, 10997−11001.
(39) Schultheis, P. J.; Fleming, S. M.; Clippinger, A. K.; Lewis, J.; Tsunemi, T.; Giasson, B.; Dickson, D. W.; Mazzulli, J. R.; Bardgett, M. E.; Haik, K. L.; Ekhator, O.; Chava, A. K.; Howard, J.; Gannon, M.; Hoffman, E.; Chen, Y.; Prasad, V.; Linn, S. C.; Tamargo, R. J.; Westbroek, W.; Sidransky, E.; Krainc, D.; Shull, G. E. Atp13a2-deficient mice exhibit neuronal ceroid lipofuscinosis, limited alpha-synuclein accumulation and age-dependent sensorimotor deficits. *Hum. Mol. Genet.* **2013**, *22*, 2067−2082.

DOI: 10.1021/acscentsci.7b00214
ACS Cent. Sci. 2017, 3, 784−793

Appendix 6.

Version 1 a0027591

Competitive and Covalent Inhibitors of Human Lysosomal Retaining Exoglucosidases

Imogen Z Breen, *Department of Chemistry, University of York, York, UK*
Marta Artola, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Liang Wu, *Department of Chemistry, University of York, York, UK*
Thomas JM Beenakker, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Wendy A Offen, *Department of Chemistry, University of York, York, UK*
Johannes MFG Aerts, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*
Gideon J Davies, *Department of Chemistry, University of York, York, UK*
Herman S Overkleeft, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Advanced article

Article Contents

- Introduction
- Retaining Glucosidases and Inverting Glucosidases
- Conformational Pathways Followed by Retaining Glucosidases
- Lysosomal Retaining Exoglucosidases and Inherited Disorders
- Competitive and Covalent GBA1 and GAA Inhibitors
- Activity-based Retaining Exoglucosidase Profiling: Diagnostic Tools for Gaucher and Pompe Diseases
- Conclusion and Outlook
- Acknowledgements

Online posting date: 22nd January 2018

Deficiency in human acid glucosylceramidase (GBA1, a retaining β-glucosidase) causes the lysosomal sphingolipid storage disorder Gaucher disease, whereas deficiency in human acid α-glucosidase (GAA, a retaining α-glucosidase) triggers the lysosomal glycogen storage disorder Pompe disease. Both enzymes process their substrate following a two-step double-displacement mechanism involving a covalent enzyme–substrate intermediate. Structural analysis of glycosidases complexed to substrates and inhibitors has provided insight into the reaction coordinates followed by glycosidases during catalytic hydrolysis and has assisted in the design of potent and selective inhibitors. Competitive and covalent inhibitors of both GBA1 and GAA have been developed in the past decades, for fundamental studies, as diagnostics tools, leads for drug development and therapeutic drugs for the clinical treatment of lysosomal storage diseases.

eLS subject area: Biochemistry

How to cite:
Breen, Imogen Z; Artola, Marta; Wu, Liang; Beenakker, Thomas JM; Offen, Wendy A; Aerts, Johannes MFG; Davies, Gideon J and Overkleeft, Herman S (January 2018) Competitive and Covalent Inhibitors of Human Lysosomal Retaining Exoglucosidases. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0027591

Introduction

Carbohydrates and their derivatives, commonly referred to as glycans and encompassing both oligo/polysaccharides and glycoconjugates (e.g. glycolipids and glycoproteins), are widespread in nature and comprise a highly diverse and complex class of biomolecules both in structure and function. This structural and functional diversity is reflected in the wide array of glycoprocessing enzymes expressed by virtually all species spread across the kingdoms of life. Glycoprocessing enzymes are generally classified as glycosyltransferases (GTs; enzymes that create a glycosidic linkage) or glycoside hydrolases (GHs, also known as glycosidases; enzymes that hydrolyse a glycosidic linkage). Both GTs and glycosidases are the subject of intense studies, aided by the classification, based on sequence (and therefore structure), of myriad GT and glycosidase families in the CAZy database (Kötzler *et al.*, 2014). Here, we focus on glycosidases, in particular two retaining exoglycosidases involved in human pathophysiology: lysosomal glucosylceramidase (GBA1 – genetic mutations of which are causative of Gaucher disease) and lysosomal α-glucosidase (GAA – deficiency of which is causative of Pompe disease) (reviewed in Kolter and Sandhoff, 2006; van der Ploeg and Reuser, 2008). The reaction itineraries by which these

enzymes process their substrates will be discussed, also in relation to the mechanism of action of competitive and covalent inhibitors of these enzymes, and the impact of these inhibitors and related activity-based probes (ABPs) on the study of GBA1/GAA activity in human health and disease. **See also**: **Gaucher Disease**; **Glycogen Storage Diseases**

Retaining Glucosidases and Inverting Glucosidases

Glycosidases typically process their substrates following two general mechanisms as described by Koshland in 1953: those that hydrolyse glycosidic linkages with the retention of configuration of the anomeric carbon (retaining glycosidases) and those that do so with the inversion of anomeric configuration (inverting glycosidases) (Koshland, 1953). Thus, nature has evolved two related, though distinct, mechanisms by which a specific substrate glycoside (an acetal or ketal) is hydrolysed to yield a sugar (a hemiacetal or hemiketal) and an alcohol (in the case of *O*-glycosides) (Jongkees and Withers, 2014).

In **Figure 1a**, the hydrolysis of a β-glucopyranosidic linkage by a retaining β-glucosidase (left) and an inverting β-glucosidase (right) is depicted. Both retaining and inverting glycosidase active sites contain a characteristic carboxylic acid/carboxylate pair, which play a key role in enzymatic hydrolysis by functioning as general acid/base or nucleophilic catalysts. Retaining glycosidases act via a two-step double displacement mechanism: the aglycon is first displaced in a formal S_N2 step by the catalytic nucleophile carboxylate residue, assisted by concurrent protonation of the departing glycosidic (exocyclic) oxygen by the catalytic acid/base carboxylic acid, yielding a covalent enzyme–substrate adduct with inverted configuration at the anomeric carbon. Upon leaving the aglycon, water enters the enzyme active site and in a reversal of steps β-glucopyranose is released, again by formal S_N2 displacement with the inversion of anomeric configuration, resulting in the overall retention of anomeric stereochemistry with respect to the initial substrate. By contrast, inverting enzymes produce α-glucopyranose from a β-glucoside directly via a single formal S_N2 displacement. Protonation of the departing aglycon by the catalytic acid residue occurs concurrently alongside nucleophilic attack of the anomeric centre by a water molecule, which is in turn activated by deprotonation by the catalytic base residue, thus yielding a hydrolysed product with the inversion of the anomeric stereocentre with respect to the initial substrate (Davies and Henrissat, 1995).

Therefore, although the products from both enzymatic processes are the same (at least over time when an equilibrium between α-glucopyranose and β-glucopyranose has established), the mechanism of the retaining enzyme is distinct from that of the inverting enzyme: catalysis takes place in two steps and substrate processing by the retaining enzyme proceeds through a covalent intermediate, whereas the inverting enzyme does not. In terms of product formation, retaining glycosidases may, upon expelling the aglycon, allow nucleophiles other than water in their active site. Indeed, retaining glycosidases are often found to be able to catalyse transglycosylations (Hansson *et al.*, 2001). In terms of glycosidase inhibitor design, both competitive and covalent retaining glycosidase inhibitors exist, whereas inverting glycosidase inhibitors are normally of the competitive kind.

Conformational Pathways Followed by Retaining Glucosidases

Figure 1b illustrates the initial enzymatic half-reactions for hydrolysis of a typical glucoside by retaining α- and β-glucosidases, leading to the formation of the covalent enzyme–glycoside intermediate. Closer perusal of the reaction coordinates followed by retaining β-glucosidases and retaining α-glucosidases reveals some telling similarities and differences.

It is thought that both retaining and inverting glycosidases process their substrate via a cyclic oxycarbenium ion-type transition state. The occurrence of such transition states imposes conformational restrictions upon the glycoside substrate, as both anomeric carbon and ring oxygen in an oxycarbenium ion develop considerable sp^2 character, which can only be accommodated by certain conformations of the glycoside ring where the C5–O5–C1–C2 atoms are coplanar (i.e. half-chair, boat or envelope conformations) (Davies and Williams, 2016). Both α-glucosidases and β-glucosidases (**Figure 1b**) proceed via a common $^4H_3$ half-chair oxycarbenium ion transition state, with differences primarily in positioning of the attacking enzyme nucleophile (β-glucosidase: bottom face, α-glucosidase: top face), and acid/base residues within the active site (Vasella *et al.*, 2002; Vocadlo and Davies, 2008).

Differences between α- and β-glucosidases also exist in the conformation of substrates when bound in the Michaelis complex (the initial enzyme–substrate complex, **Figure 1b** – left from the $^4H_3$ transition states) and that of the covalent intermediate (right from the $^4H_3$ transition states). Retaining α-glucosidases process their substrate via a $^4C_1 \rightarrow {}^4H_3 \rightarrow {}^1S_3$ itinerary, in which the substrate binds the enzyme in a $^4C_1$ Michaelis complex conformation, which is the thermodynamic glucose conformation in solution (Davies *et al.*, 2012). Retaining β-glucosidases follow the opposite reaction itinerary: $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$, and here the thermodynamic conformation of the substrate in solution is distorted by the enzyme into a higher energy $^1S_3$ conformation in the Michaelis complex (Davies *et al.*, 1998).

Overall, the oxycarbenium character of the transition state dictates that a conformation compatible with sp^2 hybridised anomeric carbon and ring oxygen atoms must be accommodated during the enzymatic reaction cycle (Thompson *et al.*, 2015), while the substrate configuration influences the specific reaction itinerary followed during hydrolysis. Prediction of reaction itineraries can often be made by calculating the conformational free energy landscape, and such molecular dynamics calculations can assist in the design of covalent and competitive retaining glycosidase inhibitors (Biarnes *et al.*, 2007; Gloster and Davies, 2010).

Version 1 a0027591

Lysosomal Retaining Exoglucosidases and Inherited Disorders

The human lysosomal glucosylceramidase, GBA1, is a retaining β-glucosidase and is responsible for the turnover of glucosylceramide (**Figure 2a**). GBA1 processes its substrate according to the general mechanism for retaining β-glucosidases as depicted in **Figure 1b** and yields ceramide and (initially) β-glucopyranose as products (Beutler, 1992). Genetic mutations in the gene encoding for GBA1 are the primary cause of the human lysosomal storage disorder (LSD) Gaucher disease, characterised by aberrant accumulation of glucosylceramide within various tissues of the body (reviewed in Germain, 2009). The nature of the mutation (at which site and which amino acid is encoded for in place of the wild-type residue) determines the extent to which mutant GBA1 activity is compromised, which in turn impacts the nature and severity of the course of the disease (mild/severe/lethal) (Hruska *et al*., 2008). Depending on the nature of the disease, Gaucher patients may be treated with recombinant GBA1 (enzyme replacement therapy – ERT) and/or by pharmacological inhibition of glucosylceramide synthase (GCS), the enzyme responsible for the biosynthesis of glucosylceramide (substrate reduction therapy – SRT). A third therapeutic strategy currently receiving much attention for Gaucher disease is termed pharmacological chaperone therapy (PCT) (Mistry *et al*., 2017). In PCT, competitive inhibitors stabilise the fold of the mutant enzyme through reversible occupancy of the active site. Such stabilisation may increase the likelihood of mutant enzyme surviving ER quality control and may also play a role in stabilising the mature active enzyme within lysosomes (Bross *et al*., 2001). Several crystal structures of GBA1, in apo form and in complex with various ligands, have now been released. A representative structure is shown in **Figure 3a** (Brumshtein *et al*., 2007). **See also**: **Protein Misfolding and Degradation in Genetic Disease**

These three therapeutic strategies – ERT, SRT and PCT – can also be considered for related LSDs that are characterised by deficiency of other glycosidases, although it should be mentioned that such therapies have to date met with considerable less success for LSDs other than Gaucher (Beck, 2010). One such disease for which no truly effective therapy exists and for which detailed knowledge of enzyme-mediated substrate processing may be key is Pompe disease. In Pompe patients, lysosomal α-glucosidase (GAA) is genetically deficient and as a result the natural GAA substrate, glycogen (**Figure 2a**), accumulates to pathological levels. GAA is a retaining exo-α-glucosidase that processes its substrate as outlined in **Figure 1b** (reviewed in Cox, 2017). A recently released crystal structure of human GAA is shown in **Figure 3b** (Deming and Garman, unpublished).

Competitive and Covalent GBA1 and GAA Inhibitors

Research in the past decades has yielded a large number of structurally diverse glucosidase inhibitors, many of which turn out to be effective inhibitors of the human enzymes, GBA1, GAA or both. Numerous glucosidase inhibitors are found in nature, and many more have also been designed and synthesised in the lab, the latter often taking natural inhibitors as a source of inspiration. In general, the most effective GBA1/GAA inhibitors are compounds mimicking functional and/or conformational aspects of the parent substrates (β-glucosides and α-glucosides, respectively).

Figure 2b depicts a representative set of β- and α-glucosidase inhibitors, and crystal structures of GBA1/GAA or their close homologues in complex with some of these inhibitors are shown in **Figure 3c–h**. 1-Deoxynojirimycin (DNJ, **1**) is the archetypal iminosugar first isolated from plants and microbes, and is a reasonably potent inhibitor of retaining and inverting β- and α-glucosidases. Numerous studies have appeared over the decades, which have explored structural and configurational DNJ analogues (Overkleeft *et al*., 1998; Okumiya *et al*., 2007). One important class of structural modifications pertain to compounds bearing *N*-alkyl chains. *N*-butyl-deoxynojirimycin (Zavesca, **2**) is in clinical use for SRT treatment of Gaucher disease, with GCS inhibition ($IC_{50} = 50\,\mu M$ (Wennekes *et al*., 2010)) as the basis of this therapy. Zavesca is, however, also a micromolar GBA1 inhibitor ($IC_{50} = 400\,\mu M$ (Wennekes *et al*., 2010)) and the structure of the cocrystal reveals a nice fit of **2** within the GBA1 enzyme active site in a $^4C_1$ conformation (Figure **3**c) (Brumshtein *et al*., 2007). Although GBA1 inhibition may initially seem to be counterproductive for a compound aimed at lowering the glucosylceramide levels, active site binding and thereby protein fold stabilisation is the basis of PCT, and thus **2** may actually increase the levels of functional GBA1 in lysosomes (Alfonso *et al*., 2005; Ben Bdira *et al*., 2017). Besides Zavesca, *N*-nonyl-DNJ **3** and isofagomine **4** in particular are also potent GBA1 inhibitors, and have been studied intensively for application in PCT (Ben Bdira *et al*., 2017). One considerable disadvantage of DNJ-type iminosugars is their relatively poor selectivity: retaining and inverting α- and β-glucosidases alike may be targeted by these compounds. Indeed, DNJ inhibits GBA1 and GBA2 ($IC_{50} = 250$ and 21 μM, respectively) and it is a relatively potent GAA inhibitor ($IC_{50} = 1.5\,\mu M$) (Wennekes *et al*., 2010).

The archetypal covalent retaining β-glucosidase inhibitor is represented by the natural product cyclophellitol (**5**) (Atsumi *et al*., 1990). This compound closely resembles β-glucopyranose in configuration and can thus position its epoxide moiety within a retaining β-glucosidase active site for nucleophilic attack by the enzyme catalytic nucleophile. In contrast to the hydrolysable acyl-al linkage that characterises covalent enzyme–substrate adducts in retaining glycosidases, the ester adduct formed upon reaction of cyclophellitol with the glycosidase nucleophile is stable, resulting in covalent enzyme inhibition, rather than substrate processing. Although no crystal structures are available for GBA1 complexed to cyclophellitol, complexes with the retaining β-glucosidase bacterial homologue *Tm*GH1 from *Thermotoga maritima* (which belongs to the same GH-A clan as GBA1 in the CAZy classification, and thus shares mechanistic similarities with it (Lombard *et al*., 2014)) shed some light on the reaction itinerary of this covalent inhibitor with β-glucosidases. *Tm*GH1 crystals soaked with a nonhydrolysable cyclophellitol analogue **10** show the glucose mimicking portion of this ligand bound within the enzyme active site in a $^4H_3$ conformation,

'above' the enzyme nucleophile poised for nucleophilic attack (Beenakker *et al.*, 2017). When *Tm*GH1 crystals are treated with **5** itself, a covalent enzyme–inhibitor adduct in the $^4C_1$ conformation can be seen, showing that the epoxide of **5** has been opened by nucleophilic displacement by the catalytic nucleophile residue (**Figure 4a**). Thus, the latter two conformations of the β-glucosidase $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ catalytic itinerary are reflected by these two enzyme–inhibitor structures (Gloster *et al.*, 2007; Beenakker *et al.*, 2017).

1,6-*Epi*-cyclophellitol **6**, with the epoxide pointing 'down' ('alpha'), is an inhibitor of retaining α-glucosidases, including GAA. As with **5**, a $^4H_3$ transition state conformation is initially adopted by **6**, but upon reaction with α-glucosidases, a $^1S_3$ covalent complex emerges with the reacted aziridine. This $^4H_3 \rightarrow {}^1S_3$ reaction itinerary has been captured in crystal structures of the bacterial GAA homologue *Cj*Agd31B complexed with **6** in both unreacted and reacted forms (**Figure 4b**) (Jiang *et al.*, 2016; Artola *et al.*, 2017). Though cyclophellitol and related cyclitol epoxides are more selective than iminosugars (inverting enzymes are for instance not inhibited), the cross-reactivity between retaining α- and β-glucosidases can sometimes occur, owing to the common $^4H_3$ conformation adopted by both **5** and **6**, which mimics the $^4H_3$ transition state conformation utilised by both α- and β-glucosidases. Exquisite selectivity for retaining α-glucosidases over β-glucosidases can be obtained by substitution of the epoxide in **5** and **6** for a cyclic sulfate electrophile, as in **8** and **9**, respectively. Both cyclosulfates **8** and **9** reside largely in $^4C_1$ conformation, which is the conformation of the Michaelis complex in retaining α-glucosidases, but not retaining β-glucosidases. As a consequence, compound **9** is primed for nucleophilic attack by α-glucosidases and is thus a potent and highly selective GAA inhibitor, whereas **8** is only a weak GBA1 inhibitor owing to its suboptimal conformation for reactivity with β-glucosidases. Crystal structures of representative α- and β-glucosidases in complex with **8** and **9** provide a structural rationale for these observations (**Figure 4c**,**d**) (Artola *et al.*, 2017).

Finally, information on reaction itineraries can also be used for the development of conformationally biased competitive inhibitors. Cyclophellitol cyclopropane **10** adopts a $^4H_3$ conformation, thereby resembling the transition state oxycarbenium ion conformation in retaining exoglycosidases, and by this virtue has turned out to be a surprisingly potent GBA1 inhibitor (Beenakker *et al.*, 2017).

Activity-based Retaining Exoglucosidase Profiling: Diagnostic Tools for Gaucher and Pompe Diseases

Mechanism-based, covalent and irreversible enzyme inhibitors are ideal starting points for the development of activity-based protein profiling (ABPP) methodology (Evans and Cravatt, 2006; Willems *et al.*, 2014a,b). About two decades ago, the first examples of tagged (biotin and fluorophore) covalent inhibitors, termed ABPs, for ABPP studies on serine hydrolase and cysteine protease families in complex biological samples were published (Liu *et al.*, 1999; Greenbaum *et al.*, 2000). Substitution of the epoxide in compounds **5** and **6** for an aziridine yielded cyclophellitol aziridine **11** and 1,6-*epi*-cyclophellitol aziridine **12**, respectively, compounds at least equally potent in inhibition of GBA1 and GAA, respectively (Li *et al.*, 2014; Artola *et al.*, 2017). In contrast to **5**/**6**, compounds **11**/**12** can be modified to contain a fluorophore (for gel/and cell imaging) or a biotin (for chemical proteomics studies) through aziridine *N*-alkylation or *N*-acylation (ABP **13**) (Jiang *et al.*, 2015a,b). Grafting a reporter moiety onto one of the cyclitol hydroxyls is typically detrimental for reactivity with one notable exception: C6-substituted cyclophellitol **14** turned out to be a highly specific and very potent GBA1 probe (Witte *et al.*, 2010). In this way, ABPs selective for retaining β-glucosidases (ABPs **13**/**14**) and retaining α-glucosidases (ABP **15**) were obtained that proved in-class selective, as can be seen from the SDS PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) images reproduced from the original papers (**Figure 5a**) (Witte *et al.*, 2010; Kallemeijn *et al.*, 2012; Jiang *et al.*, 2016). Besides tools for the discovery of enzymatic activities in comparative ABPP experiments, tagged cyclophellitols/cyclophellitol aziridines can be used to diagnose relative enzyme activity content, as shown in **Figure 5b** for monitoring mutant GBA1 activity in relation to various forms of Gaucher disease, and in **Figure 5c** for quantifying the lack of GAA activity in Pompe patient tissue.

Conclusion and Outlook

The design or discovery of glycosidase inhibitors has been actively pursued for more than five decades, yet is still a rewarding field of research. There is an obvious societal need for tailor-made glycosidase inhibitors – both competitive and covalent – given the numerous human diseases that are characterised by malfunctioning of glycosidases and/or aberrations in their substrate/product pools. These diseases include not only inherited lysosomal storage diseases (e.g. Gaucher and Pompe), but also diseases such as type 2 diabetes and Parkinsonism (Asano, 2003). One major issue in the field of glycosidase inhibitors is enzyme selectivity, and close perusal of the reaction itineraries followed by the enzymes may well assist in the future design of potent, selective inhibitors. These may be competitive or covalent, with the latter bringing the additional attractive feature of opening up the field of ABPP to glycobiology. Indeed, besides the retaining α-/β-glucosidase ABPs discussed briefly here, ABPs selective for α-fucosidases (Jiang *et al.*, 2015a,b), α-galactosidases (Willems *et al.*, 2014a,b), β-galactosidases (Marques *et al.*, 2017) and β-glucuronidases (Wu *et al.*, 2017) have seen the light and it is to be expected that many more such probes will emerge in the near future. These probes, combined with libraries of competitive inhibitors, can be applied to human glycosidases involved in health and disease, and may bring solutions for unmet medical needs in the near future, in terms of diagnostics and possibly also leads for drug development.

Acknowledgements

We thank The Netherlands Organisation for Scientific Research (NWO-CW, ChemThem Grant to J.M.F.G.A. and H.S.O), the European Research Council (ERC-2011-AdG-290836 'ChemBioSphing', to H.S.O, end ERC-2012-AdG-322942 'Glycopoise', to G.J.D.) and Sanofi Genzyme (research grant to J.M.F.G.A. and H.S.O and postdoctoral contract to M.A.) for financial support.

Glossary

Activity-based protein profiling It allows the detection and annotation of active enzymes in cells and cell extracts.

Cyclitol epoxides Covalent and mechanism-based glycosidase inactivators.

Enzyme replacement therapy The treatment of patients suffering from a lysosomal enzyme deficiency by administration of recombinant enzyme.

Glycosidases Enzymes that catalyse the hydrolysis of interglycosidic linkages.

Iminosugars Carbohydrate mimetics that competitively inhibit glycosidases.

Pharmacological chaperone therapy The treatment of patients suffering from a lysosomal enzyme deficiency with a chemical able to stabilise the genetically impaired enzyme fold.

Substrate reduction therapy The treatment of patients suffering from a lysosomal storage disorder with a compound able to correct storage metabolite levels by inhibition of its synthesis.

References

Alfonso P, Pampin S, Estrada J, *et al.* (2005) Miglustat (NB-DNJ) works as a chaperone for mutated acid β-glucosidase in cells transfected with several Gaucher disease mutations. *Blood Cells, Molecules & Diseases* **35**: 268–276.

Artola M, Wu L, Ferraz MJ, *et al.* (2017) 1,6-Cyclophellitol cyclosulfates: a new class of irreversible glycosidase inhibitor. *ACS Central Science* **3** (7): 784–793.

Asano N (2003) Glycosidase inhibitors: update and perspectives on practical use. *Glycobiology* **13** (10): 93R–104R.

Atsumi S, Umezawa K, Iinuma H, *et al.* (1990) Production, isolation and structure determination of a novel beta-glucosidase inhibitor, cyclophellitol, from *Phellinus* sp. *Journal of Antibiotics (Tokyo)* **43** (1): 49–53.

Beck M (2010) Therapy for lysosomal storage disorders. *IUBMB Life* **62** (1): 33–40.

Beenakker TJM, Wander DPA, Offen WA, *et al.* (2017) Carba-cyclophellitols are neutral retaining-glucosidase inhibitors. *Journal of the American Chemical Society* **139** (19): 6534–6537.

Ben Bdira F, Kallemeijn WW, Oussoren SV, *et al.* (2017) Stabilization of glucocerebrosidase by active site occupancy. *ACS Chemical Biology* **12** (7): 1830–1841.

Beutler E (1992) Gaucher disease: new molecular approaches to diagnosis and treatment. *Science* **256** (5058): 794–799.

Biarnes X, Ardevol A, Planas A, *et al.* (2007) The conformational free energy landscape of beta-D-glucopyranose. Implications for substrate preactivation in beta-glucoside hydrolases. *Journal of the American Chemical Society* **129** (35): 10686–10693.

Bross P, Andresen BS, Corydon TJ, *et al.* (2001) Protein misfolding and degradation in genetic disease. In: *eLS*. Chichester: John Wiley & Sons, Ltd.

Brumshtein B, Greenblatt HM, Butters TD, *et al.* (2007) Crystal structures of complexes of N-butyl- and N-nonyl-deoxynojirimycin bound to acid beta-glucosidase: insights into the mechanism of chemical chaperone action in Gaucher disease. *Journal of Biological Chemistry* **282** (39): 29052–29058.

Cox TM (2017) Glycogen storage diseases. In: *eLS*. Chichester: John Wiley & Sons, Ltd. http://www.els.net [doi: 10.1002/9780470015902.a0002270.pub2].

Davies G and Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3** (9): 853–859.

Davies GJ, Mackenzie L, Varrot A, *et al.* (1998) Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry* **37** (34): 11707–11713.

Davies GJ, Planas A and Rovira C (2012) Conformational analyses of the reaction coordinate of glycosidases. *Accounts of Chemical Research* **45** (2): 308–316.

Davies GJ and Williams SJ (2016) Carbohydrate-active enzymes: sequences, shapes, contortions and cells. *Biochemical Society Transactions* **44** (1): 79–87.

Evans MJ and Cravatt BF (2006) Mechanism-based profiling of enzyme families. *Chemical Reviews* **106** (8): 3279–3301.

Germain DP (2009) Gaucher disease. In: *eLS*. Chichester: John Wiley & Sons, Ltd. http://www.els.net [doi: 10.1002/9780470015902.a0021432].

Gloster TM and Davies GJ (2010) Glycosidase inhibition: assessing mimicry of the transition state. *Organic & Biomolecular Chemistry* **8** (2): 305–320.

Gloster TM, Madsen R and Davies GJ (2007) Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Organic & Biomolecular Chemistry* **5** (3): 444–446.

Greenbaum D, Medzihradszky KF, Burlingame A, *et al.* (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chemistry & Biology* **7** (8): 569–581.

Hansson T, Andersson M, Wehtje E, *et al.* (2001) Influence of water activity on the competition between β-glycosidase-catalysed transglycosylation and hydrolysis in aqueous hexanol. *Enzyme and Microbial Technology* **29** (8): 527–534.

Hruska KS, LaMarca ME, Scott CR, *et al.* (2008) Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Human Mutation* **29** (5): 567–583.

Jiang JB, Beenakker TJM, Kallemeijn WW, *et al.* (2015a) Comparing cyclophellitol *N*-alkyl and *N*-acyl cyclophellitol aziridines as activity-based glycosidase probes. *Chemistry--A European Journal* **21** (30): 10861–10869.

Jiang JB, Kallemeijn WW, Wright DW, *et al.* (2015b) *In vitro* and *in vivo* comparative and competitive activity-based protein profiling of GH29 alpha-L-fucosidases. *Chemical Science* **6** (5): 2782–2789.

Jiang J, Kuo CL, Wu L, *et al.* (2016) Detection of active mammalian GH31 alpha-glucosidases in health and disease using in-class, broad-spectrum activity-based probes. *ACS Central Science* **2** (5): 351–358.

Jongkees SA and Withers SG (2014) Unusual enzymatic glycoside cleavage mechanisms. *Accounts of Chemical Research* **47** (1): 226–235.

Kallemeijn WW, Li KY, Witte MD, *et al.* (2012) Novel activity-based probes for broad-spectrum profiling of retaining beta-exoglucosidases *in situ* and *in vivo*. *Angewandte Chemie International Edition* **51** (50): 12529–12533.

Kolter T and Sandhoff K (2006) Sphingolipid metabolism diseases. *Biochimica et Biophysica Acta* **1758** (12): 2057–2079.

Koshland DE (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews* **28** (4): 416–436.

Kötzler MP, Hancock SM and Withers SG (2014) Glycosidases: functions, families and folds. In: *eLS*. Chichester: John Wiley & Sons, Ltd. http://www.els.net [doi: 10.1002/9780470015902.a0020548.pub2].

Li KY, Jiang J, Witte MD, *et al.* (2014) Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors. *Organic & Biomolecular Chemistry* **12** (39): 7786–7791.

Liu Y, Patricelli MP and Cravatt BF (1999) Activity-based protein profiling: the serine hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* **96** (26): 14694–14699.

Lombard V, Golaconda Ramulu H, Drula E, *et al.* (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42** (Database issue): D490–D495.

Marques AR, Willems LI, Herrera Moro D, *et al.* (2017) A specific activity-based probe to monitor family GH59 galactosylceramidase, the enzyme deficient in Krabbe disease. *Chembiochem* **18** (4): 402–412.

Mistry PK, Lopez G, Schiffmann R, *et al.* (2017) Gaucher disease: progress and ongoing challenges. *Molecular Genetics and Metabolism* **120** (1–2): 8–21.

Okumiya T, Kroos MA, Vliet LV, *et al.* (2007) Chemical chaperones improve transport and enhance stability of mutant α-glucosidases in glycogen storage disease type II. *Molecular Genetics and Metabolism* **90** (1): 49–57.

Overkleeft HS, Renkema GH, Neele J, *et al.* (1998) Generation of specific deoxynojirimycin-type inhibitors of the non-lysosomal glucosylceramidase. *Journal of Biological Chemistry* **273** (41): 26522–26527.

van der Ploeg AT and Reuser AJ (2008) Pompe's disease. *Lancet* **372** (9646): 1342–1353.

Thompson AJ, Speciale G, Iglesias-Fernandez J, *et al.* (2015) Evidence for a boat conformation at the transition state of GH76 alpha-1,6-mannanases – key enzymes in bacterial and fungal mannoprotein metabolism. *Angewandte Chemie International Edition* **54** (18): 5378–5382.

Vasella A, Davies GJ and Böhm M (2002) Glycosidase mechanisms. *Current Opinion in Chemical Biology* **6** (5): 619–629.

Vocadlo DJ and Davies GJ (2008) Mechanistic insights into glycosidase chemistry. *Current Opinion in Chemical Biology* **12** (5): 539–555.

Wennekes T, Meijer AJ, Groen AK, *et al.* (2010) Dual-action lipophilic iminosugar improves glycemic control in obese rodents by reduction of visceral glycosphingolipids and buffering of carbohydrate assimilation. *Journal of Medicinal Chemistry* **53** (2): 689–698.

Willems LI, Beenakker TJ, Murray B, *et al.* (2014a) Potent and selective activity-based probes for GH27 human retaining alpha-galactosidases. *Journal of the American Chemical Society* **136** (33): 11622–11625.

Willems LI, Jiang J, Li KY, *et al.* (2014b) From covalent glycosidase inhibitors to activity-based glycosidase probes. *Chemistry--A European Journal* **20** (35): 10864–10872.

Witte MD, Kallemeijn WW, Aten J, *et al.* (2010) Ultrasensitive *in situ* visualization of active glucocerebrosidase molecules. *Nature Chemical Biology* **6** (12): 907–913.

Wu L, Jiang J, Jin Y, *et al.* (2017) Activity-based probes for functional interrogation of retaining beta-glucuronidases. *Nature Chemical Biology* **13** (8): 867–873.

Further Reading

Arnold ES (2004) *Iminosugars as Glycosidase Inhibitors: Nojirimycin and Beyond*. Wiley-VCH. DOI: 10.1002/3527601740.

Cravatt BF, Wright AT and Kozarich JW (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annual Review of Biochemistry* **77**: 383–414.

Futerman AH and van Meer G (2004) The cell biology of lysosomal storage disorders. *Nature Reviews Molecular Cell Biology* **5** (7): 554–565.

Jeffery DA and Bogyo M (2003) Chemical proteomics and its application to drug discovery. *Current Opinion in Biotechnology* **14** (1): 87–95.

Wennekes T, van den Berg RJBHN, Boot RG, *et al.* (2009) Glycosphingolipids-nature, function, and pharmacological modulation. *Angewandte Chemie International Edition* **48** (47): 8848–8869.

Willems LI, Overkleeft HS and van Kasteren SI (2014) Current developments in activity-based protein profiling. *Bioconjugate Chemistry* **25** (7): 1181–1191.

Version 1 a0027591

Article Title: Competitive and Covalent Inhibitors of Human Lysosomal Retaining Exoglucosidases

Article ID: a0027591

Article DOI: 10.1002/9780470015902.a0027591

Version: 1

Previous version(s): ****

Article Type: Standard

Readership Level: Advanced article

Top level subject categories: Biochemistry

Keywords: glycosidase # inhibitor # inherited disease # lysosomal storage disorder # enzyme # mechanism # suicide substrate # activity-based protein profiling # glycobiology # chemical biology

Key Concepts

- The structural and functional diversity of glycoconjugates is reflected by the vast number of glycoprocessing enzymes encountered in all domains of life.
- Retaining glycosidases form a covalent intermediate during substrate processing, whereas inverting glycosidases do not.
- Competitive and covalent retaining glycosidase inhibitors exist, whereas inverting glycosidases can normally only be blocked by competitive inhibitors.
- Both covalent and competitive glycosidase inhibitors are found in nature and serve as inspiration for the design of synthetic inhibitors.
- Structural analysis of glycosidases complexed to substrates and inhibitors provides insight into the reaction coordinates followed by glycosidases during the catalytic cycle.
- Conformational analysis of the reaction coordinate of glycosidases is key in the design of potent and selective inhibitors.
- Glycosidase inhibitors may serve as leads for drug development.
- Activity-based glycosidase probes can be applied in the discovery of glycosidase activities.
- Activity-based probes are useful starting points for the development of diagnostics assays in disease areas in which glycosidase activities are involved.
- Genetic deficiency in glycosidases is the basis of numerous inherited disorders including Gaucher disease and Pompe disease.

Author(s) and Affiliation(s):

Imogen Z Breen, *Department of Chemistry, University of York, York, UK*

Marta Artola, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Liang Wu, *Department of Chemistry, University of York, York, UK*

Thomas JM Beenakker, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Wendy A Offen, *Department of Chemistry, University of York, York, UK*

Johannes MFG Aerts, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

Gideon J Davies, *Department of Chemistry, University of York, York, UK*

Herman S Overkleeft, *Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands*

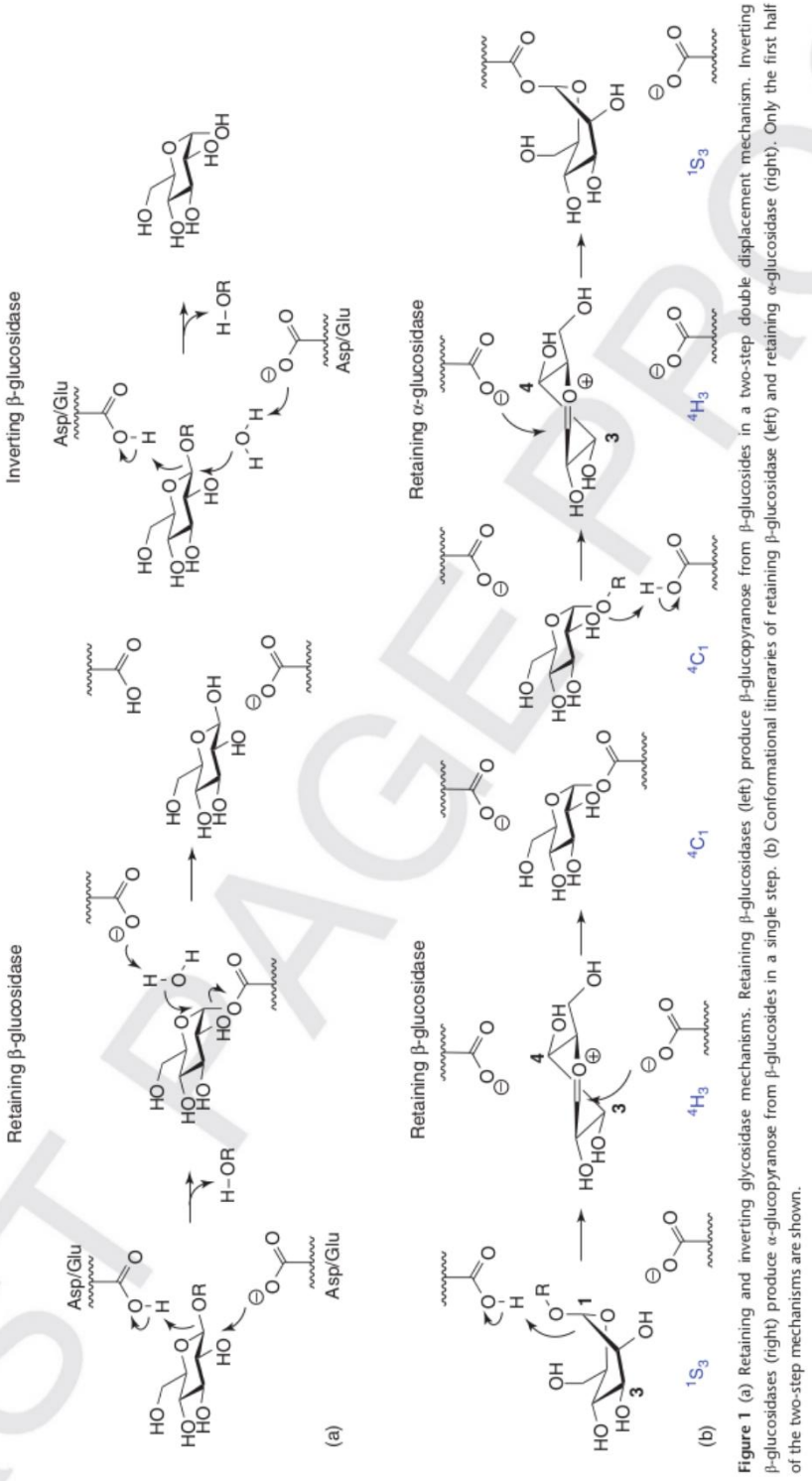


Figure 1 (a) Retaining and inverting glycosidase mechanisms. Retaining β-glucosidases (left) produce β-glucopyranose from β-glucosides in a two-step double displacement mechanism. Inverting β-glucosidases (right) produce α-glucopyranose from β-glucosides in a single step. (b) Conformational itineraries of retaining β-glucosidase (left) and retaining α-glucosidase (right). Only the first half of the two-step mechanisms are shown.

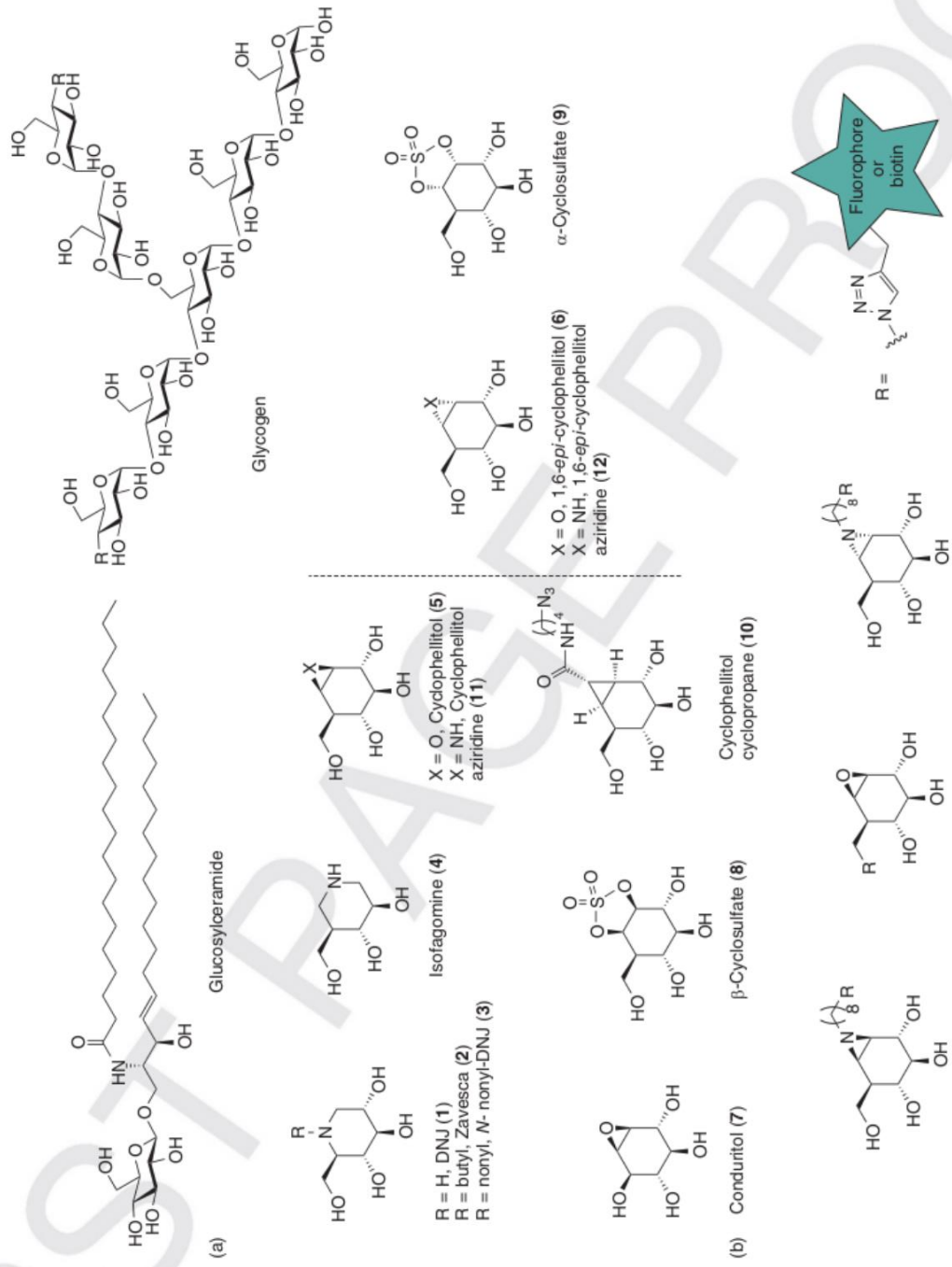


Figure 2 Substrates (a), inhibitors (b) and activity-based probes (c) of human lysosomal glucosylceramidase (GBA1, left) and human lysosomal α-glucosidase (GAA, right).

AU:6

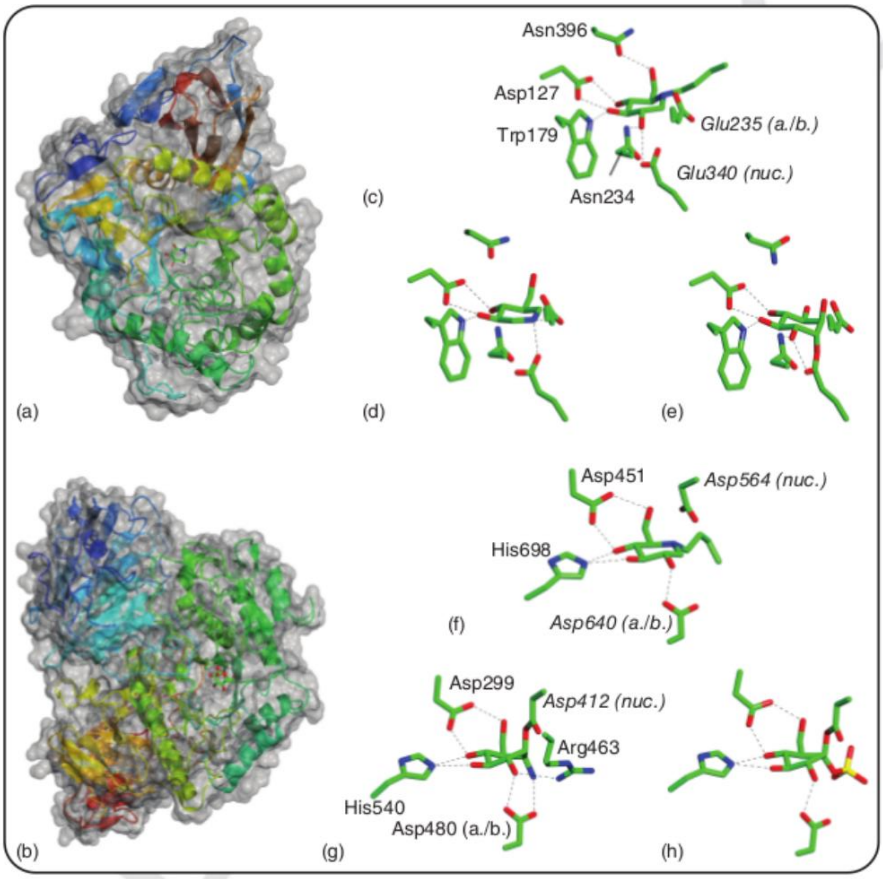


Figure 3 Crystal structures of GBA1, GAA and related enzymes. (a) Three-dimensional structure of human GBA1, showing the overall fold architecture of the enzyme. A molecule of **2** (green) can be observed bound to the enzyme. PDB: 2V3D; (b) three-dimensional structure of human GAA, showing the fold architecture of the enzyme. A molecule of glucose (green) can be observed bound to the enzyme. PDB: 5KZW; (c) active site of GBA1 in complex with **2**. Direct interactions between the ligand and protein residues are shown and sidechain numbers are annotated (nuc., nucleophile; a./b., acid/base). PDB: 2V3D. (d) Active site of GBA1 in complex with **4**. PDB: 3GXF; (e) active site of GBA1 in covalent complex with **7**. PDB: 2VT0; (f) active site of the GAA homologue α-GluI from mouse, in complex with **2**. Sidechain numbers are annotated. PDB: 5IEF; (g) active site of the GAA homologue *Cj*Agd31B from *Cellvibrio japonicus* in covalent complex with **12**. Sidechain numbers are annotated. PDB: 5I24; (h) active site of *Cj*Agd31B in covalent complex with **9**. PDB: 5NPB.

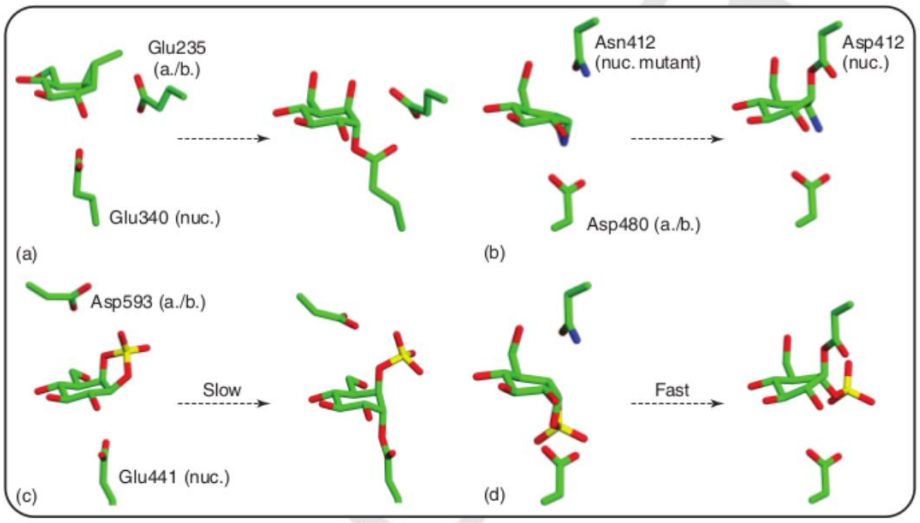


Figure 4 Reaction itineraries of cyclophellitol aziridine and cyclosulfate inhibitors as revealed by X-ray crystallography. For clarity, only nucleophile (nuc.) and acid base (a./b.) residues have been shown in these figures. (a) Reaction itinerary of β-glucosidase configured cyclophellitol probes with the β-glucosidase *Tm*GH1. (left) The sugar analogue of the nonhydrolysable carba-cyclophellitol **10** sits in a $^4H_3$ conformation above the *Tm*GH1 nucleophile, in a position primed for nucleophilic attack. (right) Reaction of *Tm*GH1 with **5** produces a covalent adduct bound to the enzyme nucleophile in a $^4C_1$ conformation. PDB: 5N6S and 2JAL; (b) reaction itinerary of α-glucosidase configured *epi*-cyclophellitol aziridine **12** with the α-glucosidase *Cj*Agd31B. (left) **12** in complex with an inactive *Cj*Agd31B D412N nucleophile mutant sits in a $^4H_3$ conformation within the enzyme active site. (right) Reaction of wild type *Cj*Agd31B with **12** gives a covalent adduct in a $^1S_3$ conformation. PDB: 5NPD and 5I24; (c) reaction itinerary of cyclophellitol cyclosulfate **8** with the β-glucosidase *Tx*GH116 from *Thermoanaerobacterium xylanolyticum*. The $^4C_1$ conformation of this probe does not match the typical β-glucosidase Michaelis complex conformation, hence **8** reacts slowly with β-glucosidases. (left) Unreacted **8** in complex with wild type *Tx*GH116 is observed after short crystal soaking times. (right) Extended soaking times are required to observe *Tx*GH116 reacted with **8**, which forms a covalent adduct in a $^4C_1$ conformation with the enzyme nucleophile. PDB: 5O0S and 5NPF; (d) reaction itinerary of *epi*-cyclophellitol cyclosulfate **9** with the α-glucosidase *Cj*Agd31B. The $^4C_1$ conformation of this probe matches the typical α-glucosidase Michaelis complex conformation, and reacts rapidly. (left) Unreacted **9** in complex with an inactive *Cj*Agd31B D412N nucleophile mutant. (right) Reacted **9** in complex with wild type *Cj*Agd31B forms a covalent adduct on the nucleophile in a $^1S_3$ conformation. PDB: 5NPC and 5NPB.

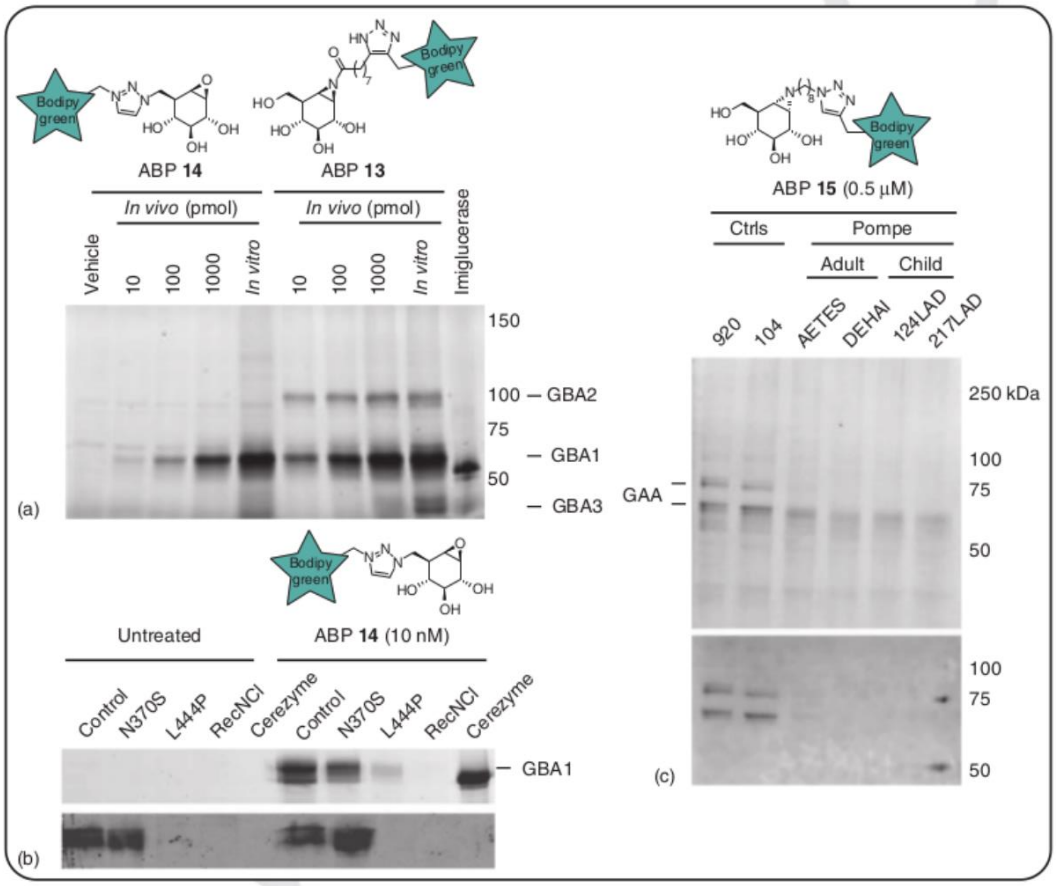


Figure 5 Activity-based protein profiling of β- and α-glucosidases in healthy and diseased (Gaucher and Pompe, respectively) individuals. (a) *In vivo* labelling of endogenous retaining β-exoglucosidases in kidney tissue homogenates of mice using broad spectrum ABP **13** (which labels GBA1, GBA2 and GBA3) and specific GBA1 ABP **14**. (b) Detection of GBA1 in Gaucher fibroblasts by labelling wild-type and homozygous N370S, L444P and RecNCI collodion fibroblast with 10 nM of ABP **14** for 60 min. GBA1 was visualised by in-gel fluorescence scanning (top panel) and by western blotting with 8E4 antibody (bottom panel). (c) *In vitro* GAA labelling at pH 4.0 with ABP **15** visualised by in-gel fluorescence scanning (top panel), followed by Western blot detection (bottom panel) of GAA in various fibroblast lysates, containing wild-type or mutant (Pompe) GAA.

Appendix 7.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

A Fluorescence Polarization Activity-Based Protein Profiling Assay in the Discovery of Potent, Selective Inhibitors for Human Nonlysosomal Glucosylceramidase

Daniël Lahav,[†,#] Bing Liu,[†,#] Richard J. B. H. N. van den Berg,[†] Adrianus M. C. H. van den Nieuwendijk,[†] Tom Wennekes,[†] Amar T. Ghisaidoobe,[†] Imogen Breen,[∥] Maria J. Ferraz,[§] Chi-Lin Kuo,[§] Liang Wu,[∥] Paul P. Geurink,[⊥] Huib Ovaa,[⊥] Gijsbert A. van der Marel,[†] Mario van der Stelt,[‡] Rolf G. Boot,[§] Gideon J. Davies,[∥] Johannes M. F. G. Aerts,*[,§] and Herman S. Overkleeft*[,†]

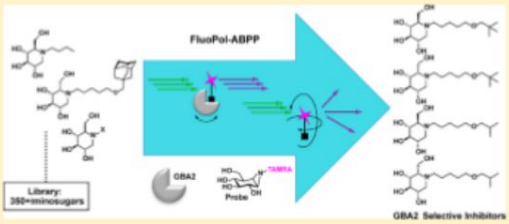
[†]Bioorganic Synthesis, [‡]Molecular Physiology, and [§]Medical Biochemistry, Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

[⊥]Department of Chemical Immunology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands

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

Supporting Information

ABSTRACT: Human nonlysosomal glucosylceramidase (GBA2) is one of several enzymes that controls levels of glycolipids and whose activity is linked to several human disease states. There is a major need to design or discover selective GBA2 inhibitors both as chemical tools and as potential therapeutic agents. Here, we describe the development of a fluorescence polarization activity-based protein profiling (FluoPol-ABPP) assay for the rapid identification, from a 350+ library of iminosugars, of GBA2 inhibitors. A focused library is generated based on leads from the FluoPol-ABPP screen and assessed on GBA2 selectivity offset against the other glucosylceramide metabolizing enzymes, glucosylceramide synthase (GCS), lysosomal glucosylceramidase (GBA), and the cytosolic retaining β-glucosidase, GBA3. Our work, yielding potent and selective GBA2 inhibitors, also provides a roadmap for the development of high-throughput assays for identifying retaining glycosidase inhibitors by FluoPol-ABPP on cell extracts containing recombinant, overexpressed glycosidase as the easily accessible enzyme source.



INTRODUCTION

Human neutral glucosylceramidase (GBA2, nonlysosomal glucocerebrosidase), a member of the glycoside hydrolase (GH) family, GH116 (www.cazy.org), hydrolyzes glucosylceramide in the cytosol to produce glucose and ceramide.[1−3] GBA2 is located at the cytosolic leaflet of the endoplasmic reticulum, Golgi apparatus, and endosomes.[1,4] GBA2 controls levels of glucosylceramide, which is synthesized at the cytosolic leaflet of the Golgi apparatus prior to translocation to the lumen of the organelle for elongation to complex glycosphingolipids. GBA2 activity is complementary to GH30 human acid glucosylceramidase (GBA, glucocerebrosidase), which processes glucosylceramide in lysosomes.[5] Genetic mutations in GBA are at the basis of the lysosomal storage disorder Gaucher disease. Gaucher macrophages are loaded with massive amounts of glucosylceramide.[5] Other tissues in Gaucher, however, have remarkably normal glucosylceramide levels, and it is likely that GBA2 can compensate for reduced GBA activity in these tissues.[5] Glucosylceramide has been linked to inflammation as well,[6] underscoring the biomedical potential of compounds able to modulate glucosylceramide levels.

In addition to its interplay with GBA, GBA2 is itself implicated in many diseases. Mutations in the GBA2 gene cause hereditary spastic paraplegia and autosomal recessive cerebellar ataxia,[7−9] indicating that compounds able to enhance GBA2 activity could have therapeutic potential. Conversely, compounds able to (partially) block GBA2 activity also hold therapeutic value; we recently found that lowering GBA2 activity in Niemann-Pick type C (NPC) mouse models led to elongated life-span and improved motor coordination.[10] This finding we correlated to correction of GBA2 activity levels in Purkinje cells, which are among the most affected neuronal cells in NPC. Slowing-down of neuropathological processes in NPC mice was achieved by genetic ablation of the GBA2 gene and alternatively by pharmacological inhibition of the gene product, the latter with *N*-alkyldeoxynojirimycin derivatives such as *N*-butyldeoxynojirimycin or *N*-(5)-adamantane-1-yl-

Received: July 14, 2017
Published: September 22, 2017

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192−14197

methoxypentyl deoxynojirimycin (respectively Zavesca (**1**) and AMP-DNM (**2**), Figure 1).[10] Indeed, we and others have reported on the development of *N*-alkyldeoxynojirimycin derivatives as potent GBA2 inhibitors. The most potent GBA2 inhibitor at present is AMP-DNM (**2**), with a K_i of about 3 nM; this compound however also inhibits glucosylceramide synthase (GCS) and GBA at high nanomolar concentrations.[11,12]

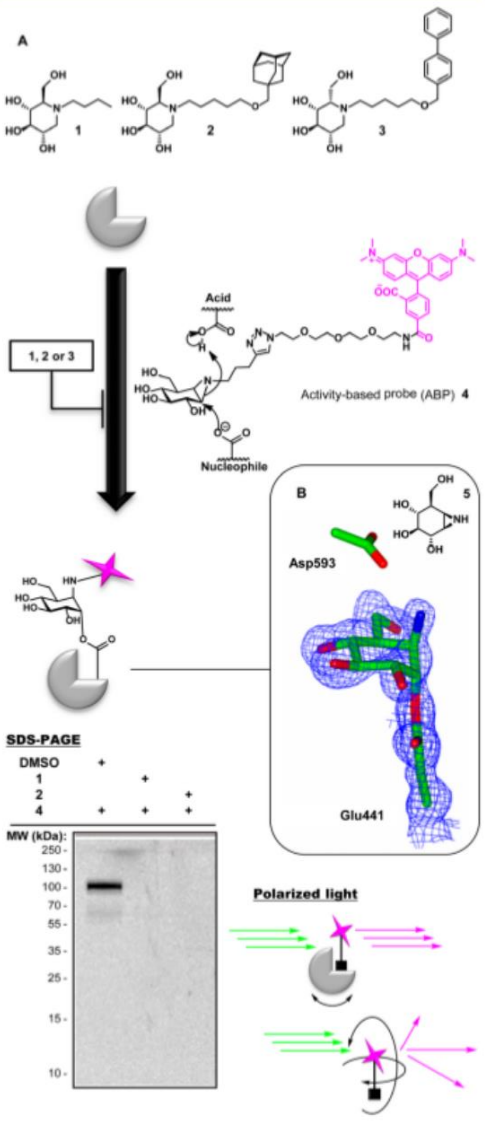


Figure 1. (A) Screening for GBA2 inhibitors (e.g., **1**, **2**, or **3**) using activity-based probe **4** as a read-out in a FluoPol-ABPP assay. (B) 3-D structure (PDB 5NCX) of the *Tx*GH116 β-glucosidase bound to cyclophellitol aziridine **5**; the enzymatic nucleophile Glu441 and putative acid/base Asp593 are shown for reference. The map shown is a maximum-likelihood weighted $2F_{obs} - F_{calc}$ at 0.4 e$^-$/Å^3.

GBA2 inhibitors thus have considerable potential both as cellular probes/tools and as therapeutic agents. Often, though, potent GBA2 inhibitors have significant off-targets; not only promiscuity toward GBA but also the enzyme responsible for the biosynthesis of glucosylceramide: GCS. Selective *in vivo* GBA2 inhibition is therefore complicated. For this reason, and considering the pharmaceutical potential of GBA2 as a drug target, selective GBA2 inhibitors are desirable commodities. Similarly, in order to access specific inhibitors, there is an urgent need for assays that rapidly identify potential GBA2 inhibitors. We here report on the development of a fluorescence polarization activity-based protein-profiling (FluoPol-ABPP) assay for fast identification of GBA2 inhibitors in extracts from cells overexpressing GBA2. Screening of our iminosugar library, which contains over 350 entries, provided lead structures bearing a relatively small (compared to existing nanomolar GBA2 inhibitors, including **2** and **3**) apolar head groups. Building a focused library around this theme and assessment of the inhibitory potential of these against GBA2, GBA, and GCS yielded new leads for, and new insights into, future development of potent, selective GBA2 inhibitors. These leads, combined with our FluoPol-ABPP assay, are a useful addition to existing reagents and tools in unraveling GBA2 functioning in health and disease.

■ RESULTS AND DISCUSSION

Our FluoPol-ABPP assay is based on our activity-based glycosidase profiling methodology that capitalizes on the two-step double displacement mechanism employed by retaining glycosidases. *N*-Alkylcyclophellitol aziridines, with the aziridine nitrogen bearing reporter groups (fluorophores, biotin), are in class, broad-spectrum retaining β-glucosidase activity-based probes (ABPs) reporting on GBA, GBA2, and the cytosolic retaining β-glucosidase GBA3 in human tissue (Figure 1A).[13] In order to support specific reaction and covalency, we solved the structure of the *Thermoanaerobacterium xylanolyticum* *Tx*GH116 β-glucosidase, a close homologue of human GBA2 with a conserved active site, in complex with cyclophellitol aziridine **5** (the precursor of ABP **4**) to ascertain that cyclophellitol aziridines react with the active site nucleophile and thus that competitors of ABP **4** are true active site inhibitors, Figure 1B.[14] The aziridine indeed binds to the invariant catalytic nucleophile Glu441, in an active site pocket on the enzyme surface. The active site pocket is freely accessible to solvent, with little steric hindrance for tagged, functionalized aziridines such as **4** (Supplementary Figure 4a). Binding of tagged ABP **4** to both GBA2 and *Tx*GH116 can also be directly observed by fluorescent scanning of the respective proteins after ABP labeling (Figure 1A and Supplementary Figure 4B).

FluoPol-ABPP studies that have appeared in the literature make use of the tetramethylrhodamine (TMR) dye.[15] In order to stay in tune with these studies, we therefore synthesized TMR-cyclophellitol aziridine **4** (for synthesis details see the Supporting Information, SI Schemes 1 and 2). GBA2 was stably overexpressed in HEK293 cells, constituting the majority of all β-glucosidase activity (the ratio of GBA2 levels over other targets labeled by **4** was established by prolonged imaging of the gel shown in Figure 1A; see the Supporting Information, SI Figure S1). Extracts of HEK293 overexpressing GBA2 were

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192−14197

treated with ABP **4** at 500 nM at pH 7, and the protein content was resolved by SDS PAGE. Fluorescence scanning of the wet gel slab (λ_{ex} 530 nm, λ_{em} 580 nm, Figure 1A) revealed a major band at about 102 kDa, which corresponds to the molecular weight of mature GBA2. Levels of active GBA2 in the overexpressing cells are such that endogenous GBA and GBA3, which are also targets of cyclophellitol aziridine ABPs, are hardly visible, even though labeling of GBA and GBA3 is saturating at the probe concentrations used. GBA2 labeling with **4** can be abolished by preincubation with the known GBA2 inhibitors Zavesca (**1**) (100 μM final concentration) and AMP-DNM (**2**) (10 μM final concentration) in a competitive ABPP experiment (Figure 1A, lanes 2 and 3, respectively). In fluorescence polarization assays, a fluorophore is excited at a specific wavelength with linear polarized light. The orientation of emitted light and thereby the degree of polarized light depend on the movement the excited fluorophore has undergone, which depends on its size: large molecules (such as GBA2-bound **4**) move and rotate slower than small molecules (here: free **4**). Samples are irradiated with linear polarized light (λ_{ex} 530 nm), and the degree of loss of polarized light emitted at λ_{em} 580 nm represents the degree of inhibition of GBA2 labeling with **4**. Initial experiments using fluorescence polarization as a read-out were performed in 96-well plates. The pH optimum of GBA2 is reported[1] to be at pH 5.8, and contemporary assays for measuring GBA2 inhibitor activities are performed at or close to this pH. We (Figure 2A)

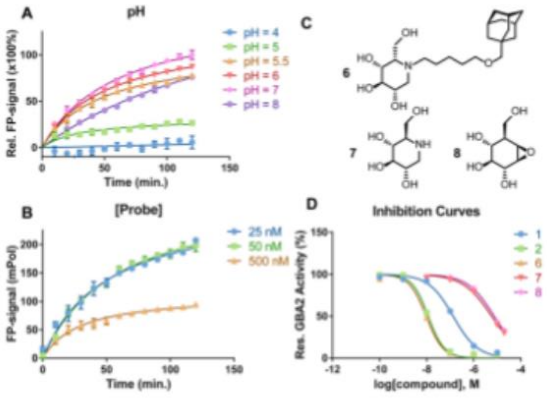


Figure 2. Optimization and validation of FluoPol-ABPP for GBA2. (A) Effect of pH. (B) Effect of probe concentration. (C) Structures of additional established inhibitors. (D) Inhibition by established inhibitors (**1**, **2**, **6**, **7**, and **8**). Error bars represent standard error of the mean (SEM).

established that, in a range from pH 4 to pH 8 (the range at which GBA2 possesses measurable activity), maximal return of polarized light occurs at pH 7. The optimal final probe concentration in the assay conditions used (protein concentration 0.5 mg/mL, V_{final} 75 μL), which intrinsically depends on the amount of active GBA2 present in the extracts (which we keep constant), was found to be 25−50 nM (Figure 2B); in this setting a suitable difference in polarized light emission (Δ about 180 mPol) was achieved between samples derived from GBA2-overexpressing cells and wild-type cells. IC_{50} values of a set of known GBA2 inhibitors, both competitive (Zavesca (**1**), AMP-DNM (**2**), L-*ido*-AMP-DNM (**6**), DNM (**7**)) and mechanism-based (cyclophellitol, **8**), were next established using the FluoPol-ABPP format using 50 nM ABP **4**. Values obtained (Figure 2D) were in line with literature values on the same compound, for both the potent (**2**, **3**, **6**) and less potent (**1**, **7**, **8**), which were all established using fluorogenic substrate assays, thereby validating our FluoPol-ABPP assay.[16]

We next transferred our assay to a 384-well plate format (V_{final} 15 μL) and screened our complete iminosugar library, containing both previously published and until now unpublished compounds (the structures of all 358 compounds and the synthesis of new compounds are in the Supporting Information). All compounds were assessed at 100 nM final concentration, and the percentage of GBA2 inhibition calculated from the relative loss of fluorescence polarization was observed. Besides *N*-alkyldeoxynojirimycin, the corresponding D-*galacto*- and L-*ido*-configured iminosugars are known to be potent GBA2 inhibitors, and library entries adhering to this general description are picked up in our FluoPol-ABPP assay as well (Figure 3).[16] Alternatively

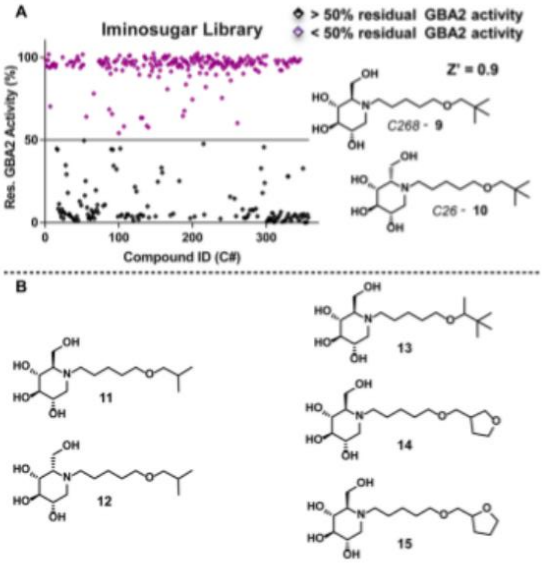


Figure 3. FluoPol-ABPP assay with an iminosugar library (358 entries). (A) Competition at 100 nM. (B) Examples of second-generation analogues of compounds **9** and **10**.

configured piperidines, partially deoxygenated deoxynojirimycin derivatives, and hydroxylated pyrrolidines show up only rarely, if at all, as inhibitors in our assays (for a more detailed grouping of compound classes and GBA2 inhibition see the Supporting Information).

Among the most potent GBA2 inhibitors we identified neopentyloxypentyl-deoxynojirimycin **9** and its L-*ido*-congener **10** (Figure 3). Intriguingly, for these compounds the hydrophobic *N*-alkyl group is relatively small compared to that of **2** and **3**. We previously found that large (as in **2**) and preferably aromatic (as in **3**) alkoxy substituents are preferred by GCS and reasoned that a reduced size of this part of the molecule may be favorable for achieving GBA2 selectivity over the ceramide glucosyltransferase. Accordingly, we synthesized a set of close analogues of compounds **9** and **10** and assessed their inhibitory potency against GBA2, GBA, GBA3, and GCS (examples of second-generation library **11**−**15** in Figure 3; for

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192−14197

Table 1. Inhibition Values and Selectivity Ratios of Deoxynojirimycin Derivatives on GCS, GBA, GBA2, and GBA3

compound	GCS[a]	GBA1[b]	GBA2[b]	GBA3[b]	GCS/GBA2[c]	GBA1/GBA2[c]	GBA3/GBA2[c]
1	50	675	0.326	>1000	153	2071	>3067
2	0.15	0.28	0.0035	27.6	43	80	7886
3	0.008	25.3	0.0027	27.8	3	9370	10296
6	0.15	12.4	0.0083	66.1	18	1494	7964
9	5.05	3.8	0.0051	27.6	990	745	5412
10	2.77	120	0.0077	119	360	15584	15455
11	3.87	10.8	0.0075	128	516	1440	17067
12	0.729	1223	0.0095	177	77	128737	18632
13	6	7.3	0.0059	116	1017	1237	19661
14	53	92	0.0536	180	989	1716	3358
15	>50	122	0.1771	145	>282	689	819

[a]Inhibition value for *in situ* assay is given as IC_{50} (μM). [b]Inhibition value for *in vitro* assay is given as K_i (μM). [c]Relative GBA2 offset (ratio)

synthesis and analytical details see the Supporting Information).

As can be seen in Table 1, compounds **9** and **10** are highly potent and rather selective GBA2 inhibitors. Removal of one methyl group, as in isopropyl derivatives **11** and **12**, led to an increase of inhibitory potency against GBA2 with concomitant (slight) decrease of inhibitory potency against GCS. L-*ido*-configured compounds **10** and **12** are over 100 000-fold selective for GBA2 over GBA1, which is a significant improvement compared to previously reported inhibitors showing a maximal 10 000-fold selectivity (for selectivity ratios of the compounds in Table 1 for GBA2 offset against GBA, GBA3, and GCS, respectively, see the Supporting Information, SI 8). As mentioned earlier, the activity against GCS is much reduced when comparing to our previously reported adamantyl and biphenyl compounds, respectively **2** and **3**. Compounds **9**, **13**, and **14** are about 1000-fold more potent on GBA2 over GCS, while previously reported compounds are at best 150-fold more selective. While substitution of the neopentyl moiety for tetrahydrofuranylmethyl groups, as in **14** and **15**, had a detrimental effect on inhibitory potency toward GBA, GBA2, and GCS, it has no significant effect on the selectivity. Based on both potency and selectivity, neopentyl and isopropyl derivatives **9**–**12** are perhaps the most potent and selective GBA2 inhibitors reported to date.

In a final *in situ* target engagement experiment we performed competitive ABPP on live cells. Cells overexpressing GBA2 or GBA3 and containing endogenous GBA levels were treated with compounds **9**–**12** at various final concentrations, lysed, and then treated with ABP **4**. As can be seen from the images of the SDS PAGE gels (Figure 4), all compounds are cell permeable and selectively block GBA2 over GBA and GBA3 at the concentrations tested.

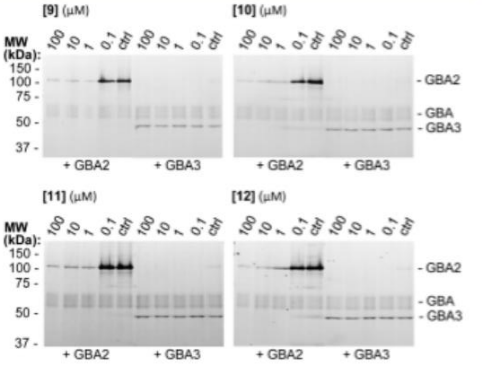


Figure 4. Competitive ABPP experiment of HEK293T cells overexpressing GBA2 or GBA3 (and expressing endogenous GBA) treated with compounds **9**–**12** at various final concentrations prior to cell lysis and ABPP profiling of remaining enzyme activity.

■ CONCLUSIONS

Collectively, this work describes the development of an effective FluoPol-ABPP assay for the rapid screening of GBA2 inhibitors. Fluorescence polarization (FluoPol) is often used for the analysis of the potency of small molecules to inhibit enzymes[17] but has to the best of our knowledge not been used in the context of glycosidase inhibitor assays. This certainly holds true for the combination of FluoPol and ABPP,[18] as we show here. FluoPol-ABPP can be conducted on target enzymes in complex mixtures, as long as the ABP used modifies with considerable selectivity the target enzyme. This can be achieved even with broad-spectrum ABPs such as the one used here, simply by bringing the target enzyme to overexpression. Target selectivity is easily established by running an SDS PAGE on the ABP-labeled sample. One potential caveat of FluoPol-ABPP is the covalent and irreversible nature of the ABP used. Most studies on FluoPol-ABPP inhibitor screening entail the assessment of covalent inhibitors.[18] Here, we show that competitive inhibitors can also be identified using FluoPol-ABPP as an initial screen, after which more in-depth kinetic studies can be executed on selected compounds using conventional assays using fluorogenic substrates. Here, we identified a few compounds from our 358-compound collection that possess considerable GBA2 selectivity, and on the basis of these lead structures we prepared a small focused library to add to our growing library of iminosugar glycosidase inhibitors. In-depth analysis revealed that, indeed, modulating the size of the hydrophobic nitrogen substituent yields inhibitors more potent, and crucially also more selective, for GBA2 compared to existing inhibitors. From a technical point of view, the FluoPol-ABPP assay can be adapted to allow high-throughput screening of large compound selections on GBA2 inhibition. Conceptually, with this study we show that FluoPol-ABPP, as a recently proposed strategy for high-throughput screening of various enzymes, is also eminently feasible for searching for glycosidase inhibitors, especially considering that the list of activity-based glycosidase probes targeting a variety of retaining glycosidases is growing.

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192–14197

EXPERIMENTAL SECTION

Cell Culture. Human embryonic kidney 293T (HEK293T) cells (Sigma) were cultured in DMEM high glucose (Gibco) supplemented with 10% NBS and 100 units/mL penicillin/streptomycin (Gibco) at 37 °C and 5% CO_2. RAW 264.7 cells (American Type Culture Collection) were cultured in RPMI (Gibco) supplemented with 10% fetal calf serum, 1 mM Glutamax, and 100 units/mL penicillin/streptomycin (Gibco) at 37 °C and 5% CO_2.

Transient and Stable Overexpression of Human GBA2 and GBA3. Primers used for transient GBA2-expressing HEK293T were designed based on NCBI reference sequence NM_020944.2. The full-length coding sequences were cloned into pcDNA3.1/Myc-His (Invitrogen, Life Technologies, Carlsbad, CA, USA). Subconfluent HEK293T cells were transfected with empty pcDNA3.1 or GBA2pcDNA3.1 (plasmid:PEI ratio 1:3). Media was refreshed 24 h later, and cells were collected 72 h after transfection in phosphate-buffered saline (PBS) buffer. Cells were centrifuged at 1000 rpm for 10 min, after which the supernatant was removed. Cell pellets were snap-frozen with liquid nitrogen and stored at −80 °C.

Stable GBA2-expressing HEK293T cells were generated as follows. The PCR-amplified human GBA2 (GBA2, acc. no. NM_020944.2) coding sequence (using the following primers: sense 5′-GGGGACAA-GTTTGTACA AAAAAGCAGGCTTAACCACCATGGGGACCC-AGGAT CCAG-3′ and antisense 5′- GGGGACCACTTTGTAC AAGAAAGCTGGGTTTCACTCTGGGCTCAGGTTTG-3′) was cloned into pDNOR-221 and subcloned in pLenti6.3/TO/V5-DEST using the Gateway system (Invitrogen). Correctness of the construct was verified by sequencing. HEK293T cells were transfected with pLenti6.3-GBA2 in combination with the envelope and packaging plasmids pMD2G, pRRE, and pRSV. Subsequently, culture supernatant containing viral particles was collected and used for infection of HEK293T cells. Selection using blasticidin for several weeks rendered cells stably expressing human GBA2 as determined by activity assays. For stable expression of GBA3 in HEK293T cells, the PCR-amplified GBA3 (GBA3, acc. no. NM_020973.4) coding sequence (using the following oligonucleotides: sense 5′-GAATTCGCCGCACCAT-GGCTTTCCCTGCAGGATTTG-3′ and antisense 5′-GCGGCCGC-AGATGTGCTTCAA GGCCATTG-3′) was cloned into pcDNA3.1/Zeo and transfected into HEK293T cells using FuGene 6 Transfection Reagent (Promega Benelux, Leiden, The Netherlands). Selection using Zeocin for several weeks rendered cells stably expressing human GBA3 as determined by activity assays.

Preparation of Lysates. Cell pellets were resuspended in lysis buffer (20 mM hepes, 2 mM DTT, 0.25 M sucrose, 1 mM $MgCl_2$, 2.5 U/mL benzonase, pH 7.0). The homogenate was incubated on ice for 30 min after lysis and homogenization using SilentCrusher (Heidolph). Ultracentrifugation was performed at 32.000 rpm for 30 min at 4 °C. Supernatant fractions were aliquoted in appropriate volumes, after the total protein concentration was determined via a Bradford assay, using BSA (Sigma) for standards and BioRad Quickstart Bradford Reagents. Aliquots were snap-frozen with liquid nitrogen and stored at −80 °C.

Gel-Activity-Based Protein Profiling Experiments. Prepared lysate was diluted in assay buffer (20 mM Hepes, 2 mM DTT at pH 7.0) until appropriate final protein concentration was reached. Samples were incubated with 500 nM probe at 37 °C for 1 h (V_{final} = 20 μL). Protein content was denatured using Laemli buffer (4×) at 100 °C for 3 min. Reactions were resolved by 12.5% SDS-PAGE electrophoresis, and wet slabs were scanned for fluorescence (Molecular Imager Gel Doc XR, Biorad). For the *in situ* competitive ABPP experiments stable GBA2- and GBA3-overexpressing HEK293T cells were incubated for 1 h at 37 °C with various concentrations of the selected inhibitors. Cells were washed two times with PBS and then harvested by scraping in 25 mM potassium phosphate buffer (pH 6.5) containing 0.1% Triton X-100 (v/v). Cells were lysed via snap-freezing in liquid nitrogen. Protein concentration was determined via BCA assay (Thermo Fischer). Labeling of the glucosidases was conducted using 500 nM ABP **4** for 30 min at 37 °C on 20 μg of protein. Protein content was denatured using Laemli buffer (5×) at 100 °C for 5 min. Reactions were resolved by 10% SDS-PAGE electrophoresis, and wet slabs were scanned as described above.

Optimization of the FluoPol-ABPP Assay. Optimal probe concentration on the FluoPol signal was determined by varying probe concentrations from 25 to 500 nM probe at a constant lysate concentration (0.5 mg/mL) and neutral pH. FluoPol-ABPP assays were also performed at different pH values. Hepes was replaced by 20 mM citric acid−citrate buffer for reactions at pH ≤ 6. The pH experiments were performed at optimal probe concentration (50 nM). Competition experiments were conducted by 30 min of preincubation of compounds in the lysates at 37 °C (2.5% DMSO). All reactions (V_{final} 75 μL) containing bovine gamma-globulin (Sigma) and Chaps (Sigma), respectively 0.5 and 1 mg/mL, were carried out in 96-well plates (flat-black bottom, Greiner). FluoPol signals were monitored on an Infinite M1000Pro (Tecan) using λ_{ex} 530 nm and λ_{em} 580 nm. Mock-containing lysates were used as reference samples, samples without probe as blanks to correct for background polarization, and GBA2-containing lysates without inhibitor as controls. All samples were corrected for background polarization, and the residual enzyme activity was calculated based on the polarization signal from the controls. Polarization signals were plotted against time or inhibitor concentration and processed in GraphPad Prism 6.0. IC_{50} values were calculated via nonlinear regression using the mentioned software (N = 2, n = 3).

FluoPol-ABPP Screen of the Iminosugar Library. The screen, using the optimized conditions as described above, was conducted in a 384-well black-bottom plate (Greiner) with reaction volumes set at 15 μL. Concentration of the iminosugars during the screen was 100 nM. The FluoPol signal was measured on a PHERAstar (BMG Labtech). Resulting polarization signals were processed as described above. Residual enzyme activities were plotted against the corresponding iminosugar.

Analysis of Selected Compounds as Inhibitors of Enzymatic Activity of GCS, GBA, GBA2, and GBA3. Enzyme assays of GCS, GBA1, GBA2, and GBA3 were carried out as previously described (refs 16 and 19), and inhibition by selected compounds was assessed. Briefly, for GBA1, pure recombinant human enzyme (Cerezyme from Genzyme) was used. Activity was measured with 3.7 mM 4-methylumbeliferone (4MU)-β-D-glucopyranoside (Sigma) in 150 mM McIlvaine buffer pH 5.2 supplemented with 0.2% sodium taurocholate (w/v), 0.1% Triton X-100 (v/v), and 0.1% bovine serum albumin (BSA) (w/v) for 30 min. The reaction was stopped with excess 1 M NaOH-glycine (pH 10.3), and liberated 4MU fluorescence was measured with an LS55 fluorimeter (PerkinElmer) using λ_{Ex} 366 nm and λ_{Em} 445 nm. For GBA2 measurements, cellular homogenates of a stable HEK293T-overexpressing GBA2 cell line preincubated for 30 min with an inhibitor of GBA1 (1 mM conduritol β epoxide CBE from Sigma) were used. Activity was measured with 3.7 mM 4MU-β-D-glucopyranoside in 150 mM McIlvaine pH 5.8 and 0.1% BSA (w/v) for 1 h. Reactions were stopped as described above for GBA1. For GBA3 measurements, cellular homogenates of a stable HEK293T-overexpressing GBA3 cell line also preincubated for 30 min with an inhibitor of GBA1 (1 mM conduritol β epoxide CBE from Sigma) were used. Reactions were conducted in 100 mM hepes pH 7.0 and 0.1% BSA (w/v) for 1 h. All IC_{50} values were determined in triplicate, and the inhibitors tested were preincubated for 30 min at 37 °C. K_i values were determined in duplicate using a range from 0.05 to 5 mM 4MU-β-D-glucopyranoside in appropriate buffer containing inhibitor. Incubation time and temperature were the same as described above, but without the preincubation step. Observed fluorescence was curve-fitted against inhibitor or substrate concentrations using GraphPad Prism 6.0 in order to obtain IC_{50} or K_i values.

IC_{50} values for GCS were determined *in situ* with 6-[*N*-methyl-*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-ylaminododecanoyl]sphingosine (NBD-ceramide) as substrate. RAW 264.7 cells were grown to confluence in six-well plates and preincubated for 1 h with an inhibitor of GBA1 activity (300 μM CBE), followed by 1 h incubation at 37 °C with 1 μM C6-NBD-ceramide and in the presence of a range of inhibitor concentrations. The cells were washed 3× with PBS and harvested by scraping. After lipid extraction (described by Bligh and

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192−14197

Dyer[20]), the C6-NBD lipids were separated and detected by high-performance liquid chromatography (λ_{Ex} 470 nm and λ_{Em} 530 nm). IC_{50} values were determined in duplicate from the titration curves of observed formed C6-NBD-glucosylceramide, and data were curve-fitted via GraphPad Prism 6.0.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b07352.

Synthetic procedures for ABP **4** and (unpublished) deoxynojirimycin analogues, ^1H and ^{13}C NMR spectra, IR and LC-MS data; figures validating the FluoPol-ABPP method and additional figures for the screen showing a preference for *gluco*- and *ido*-configured iminosugars; compound structures and inhibitory potencies on GCS, GBA1, and GBA2 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*j.m.aerts@chem.leidenuniv.nl

*h.s.overkleeft@chem.leidenuniv.nl

ORCID

Gideon J. Davies: 0000-0002-7343-776X

Herman S. Overkleeft: 0000-0001-6976-7005

Author Contributions

#D. Lahav and B. Liu contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank The Netherlands Organization for Scientific Research (NWO-CW, TOP-PUNT grant to H.S.O.), the China Scholarship Council (CSC, Ph.D. fellowship to B.L.), and the European Research Council (ERC-2011-AdG-290836 "Chembiosphing", to H.S.O., ERC-2012-AdG-32294 "Glycopoise", to G.J.D.; G.J.D. holds the Royal Society Ken Murray Research Professorship) for financial support.

■ REFERENCES

(1) Van Weely, S.; Brandsma, M.; Strijland, A.; Tager, J. M.; Aerts, J. M. F. G. *Biochim. Biophys. Acta, Mol. Basis Dis.* **1993**, *1181*, 55−62.

(2) Yildiz, Y.; Matern, H.; Thompson, B.; Allegood, J. C.; Warren, R. L.; Ramirez, D. M. O.; Hammer, R. E.; Hamra, F. K.; Matern, S.; Russell, D. W. *J. Clin. Invest.* **2006**, *116*, 2985−2994.

(3) Boot, R. G.; Verhoek, M.; Donker-Koopman, W.; Strijland, A.; Van Marle, J.; Overkleeft, H. S.; Wennekes, T.; Aerts, J. M. F. G. *J. Biol. Chem.* **2007**, *282*, 1305−1312.

(4) Körschen, H. G.; Yildiz, Y.; Raju, D. N.; Schonauer, S.; Bönigk, W.; Jansen, V.; Kremmer, E.; Kaupp, U. B.; Wachten, D. *J. Biol. Chem.* **2013**, *288*, 3381−3393.

(5) Ferraz, M. J.; Kallemeijn, W. W.; Mirzaian, M.; Herrera Moro, D.; Marques, A.; Wisse, P.; Boot, R. G.; Willems, L. I.; Overkleeft, H. S.; Aerts, J. M. F. G. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2014**, *1841*, 811−825.

(6) Nagata, M.; Izumi, Y.; Ishikawa, E.; Kiyotake, R.; Doi, R.; Iwai, S.; Zakaria, O.; Yamaji, T.; Miyamoto, T.; Bamba, T.; Yamasaki, S. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3285−3294.

(7) Hammer, M. B.; Eleuch-Fayache, G.; Schottlaender, L. V.; Nehdi, H.; Gibbs, J. R.; Arepalli, S. K.; Chong, S. B.; Hernandez, D. G.; Sailer, A.; Liu, G.; Mistry, P. K.; Cai, H.; Shrader, G.; Sassi, C.; Bouhlal, Y.; Houlden, H.; Hentati, F.; Amouri, R.; Singleton, A. B. *Am. J. Hum. Genet.* **2013**, *92*, 245−251.

(8) Martin, E.; Schüle, R.; Smets, K.; Rastetter, A.; Boukhris, A.; Loureiro, J. L.; Gonzalez, M. A.; Mundwiller, E.; Deconinck, T.; Wessner, M.; Jornea, L.; Oteyza, A. C.; Durr, A.; Martin, J. J.; Schöls, L.; Mhiri, C.; Lamari, F.; Züchner, S.; De Jonghe, P.; Kabashi, E.; Brice, A.; Stevanin, G. *Am. J. Hum. Genet.* **2013**, *92*, 238−244.

(9) Sultana, S.; Truong, N. Y.; Vieira, D. B.; Wigger, J. G. D.; Forrester, A. M.; Veinotte, C. J.; Berman, J. N.; van der Spoel, A. C. *Zebrafish* **2016**, *13*, 177−187.

(10) Marques, A. R. A.; Aten, J.; Ottenhoff, R.; Van Roomen, C. P. A. A.; Moro, D. H.; Claessen, N.; Veloz, M. F. V.; Zhou, K.; Lin, Z.; Mirzaian, M.; Boot, R. G.; De Zeeuw, C. I.; Overkleeft, H. S.; Yildiz, Y.; Aerts, J. M. F. G. *PLoS One* **2015**, *10*, 1−18.

(11) Overkleeft, H. S.; Renkema, G. H.; Neele, J.; Vianello, P.; Hung, I. O.; Strijland, A.; van der Burg, A. M.; Koomen, G.; Pandit, U. K.; Aerts, J. M. F. G. *J. Biol. Chem.* **1998**, *273*, 26522−26527.

(12) Wennekes, T.; van den Berg, R. J. B. H. N.; Boot, R. G.; van der Marel, G. A.; Overkleeft, H. S.; Aerts, J. M. F. G. *Angew. Chem., Int. Ed.* **2009**, *48*, 8848−8869.

(13) Kallemeijn, W. W.; Li, K.-Y.; Witte, M. D.; Marques, A. R. A.; Aten, J.; Scheij, S.; Jiang, J.; Willems, L. I.; Voorn-Brouwer, T. M.; van Roomen, C. P. A. A.; Ottenhoff, R.; Boot, R. G.; van den Elst, H.; Walvoort, M. T. C.; Florea, B. I.; Codée, J. D. C.; van der Marel, G. A.; Aerts, J. M. F. G.; Overkleeft, H. S. *Angew. Chem.* **2012**, *124*, 12697−12701.

(14) Charoenwattanasatien, R.; Pengthaisong, S.; Breen, I.; Mutoh, R.; Sansenya, S.; Hua, Y.; Tankrathok, A.; Wu, L.; Songsiriritthigul, C.; Tanaka, H.; Williams, S. J.; Davies, G. J.; Kurisu, G.; Cairns, J. R. K. *ACS Chem. Biol.* **2016**, *11*, 1891−1900.

(15) Bachovchin, D. A.; Brown, S. J.; Rosen, H.; Cravatt, B. F. *Nat. Biotechnol.* **2009**, *27*, 387−394.

(16) Wennekes, T.; Meijer, A. J.; Groen, A. K.; Boot, R. G.; Groener, J. E.; van Eijk, M.; Ottenhoff, R.; Bijl, N.; Ghauharali, K.; Song, H.; Shea, T. J. O.; Liu, H.; Yew, N.; Copeland, D.; van den Berg, R. J.; van der Marel, G. A.; Overkleeft, H. S.; Aerts, J. M. F. G. *J. Med. Chem.* **2010**, *53*, 689−698.

(17) Lea, W. A.; Simeonov, A. *Expert Opin. Drug Discovery* **2011**, *6*, 17−32.

(18) Niphakis, M. J.; Cravatt, B. F. *Annu. Rev. Biochem.* **2014**, *83*, 341−377.

(19) Dekker, N.; Voorn-Brouwer, T.; Verhoek, M.; Wennekes, T.; Narayan, R. S.; Speijer, D.; Hollak, C. E. M.; Overkleeft, H. S.; Boot, R. G.; Aerts, J. M. F. G. *Blood Cells, Mol., Dis.* **2011**, *46*, 19−26.

(20) Bligh, E. G.; Dyer, W. J. *Can. J. Biochem. Physiol.* **1959**, *37*, 911−917.

DOI: 10.1021/jacs.7b07352
J. Am. Chem. Soc. 2017, 139, 14192−14197

List of Abbreviations

µg	Microgram
µL	Microlitre
µM	Micromolar
µmol	Micromole
ABP	Activity based probe
ABPP	Activity based protein profiling
Ap	*Arthrospira plantensis*
ARCA	Autosomal-recessive cerebellar ataxia
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BLAST	Basic local alignment search tool
BODIPY	Boron-dipyrromethene
Bt	*Bacteriodes thetaiotamicron*
C	Celsius
ca	Circa
CAZy	Carbohydrate active enzymes database
CBE	Conduritol B epoxide
CCP4i2	Collaborative Computational Project No. 4 (interface 2)
cDNA	Chromosomal deoxyribonucleic acid
CHO	Chinese hamster ovary (cell strain)
CNS	Central nervous system
COOT	Crystallographic Object-Oriented Toolkit
CV	Column volumne
Da	Dalton
DNJ	Deoxynojirimycin
DNA	Deoxyribonucleic acid
E.coli	*Escherichia coli*
EC	Enzyme commission (number)
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum- associated degradation
ERT	Enzyme replacement therapy
GBA1	β-Glucocerebrosidase 1 (lysosomal) (*Homo sapiens)*
GBA2	β-Glucocerebrosidase 2 (non-lysosomal) (*Homo sapiens)*
GBA3	β-Glucocerebrosidase 3 (non-lysosomal) (*Homo sapiens)*
GH	Glycoside hydrolase family

GIM	Glucoimidazole
Glc	Glucose
Gln	Glutamine
Glu	Glutamic acid
GluCer	Glucosylceramide
Gt	*Geobacillus thermoglucosidasius*
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HSP	Hereditary spastic paraplegia
Hz	Hertz
IC_{50}	Half maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry (ITC)
kDa	Kilodalton
K_i	Inhibition constant
L	Litre
LB	Lysogeny Broth
LSD	Lysosomal storage disorder
m	Minutes
M	Molar
MES	2-(N-morpholino)ethanesulfonic acid
Mg	Milligram
mL	Millilitre
mM	Millimolar
NB-DNJ	N-butyldeoxynojirimycin, (miglustat, marketed as Zavasca®)
nL	Nanolitre
nm	Nanometre
nM	Nanomolar
NN-DNJ	N-nonyldeoxynojirimycin
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PCT	Pharmacological chaperone therapy
PDB	Protein databank
PEG	Polyethylene glycol

PNGase	Peptide-N-glycosidase
RMSD	Root-mean-square deviation
RPM	Revolutions per minute
s	Seconds
SDS	Sodium dodecyl sulphate
SEC-MALLS	Size Exclusion Chomatography - Multi-Angle Laser Light Scattering
SeMet	SelenoMethioine
Ser	Serine
SPG46	Spastic Paraplegia 46
SRT	Substrate reduction therapy
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
Trp	Tryptophan
TSA	Thermal Shift Assay
Tx	*Thermoanaerobacterium xylanolyticum*
UDP-Glc	Uridine diphosphate glucose

References:

1. Schnaar, R. L., Suzuki, A., and Stanley, P. (2009) *Essentials of Glycobiology*, 2 ed., Cold Spring Harbor Laboratory Press
2. Hannun, Y. A. L. C. R. M. J., A.H.; Bell, R. M.;. (2016) Solving the Riddle of the Role of Sphingolipids in Cell Signaling. in *J Biol Chem*
3. Wennekes, T., van den Berg, R. J., Boot, R. G., van der Marel, G. A., Overkleeft, H. S., and Aerts, J. M. (2009) Glycosphingolipids--nature, function, and pharmacological modulation. *Angewandte Chemie* **48**, 8848-8869
4. Svennerholm, L. (1963) Chromatographic Separation of Human Brain Gangliosides. *J Neurochem* **10**, 613-623
5. Raghavan, S. S., Mumford, R. A., and Kanfer, J. N. (1974) Isolation and characterization of glucosylsphingosine from Gaucher's spleen. *J Lipid Res* **15**, 484-490
6. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130
7. Hakomori, S. (1990) Bifunctional role of glycosphingolipids. Modulators for transmembrane signaling and mediators for cellular interactions. *J Biol Chem* **265**, 18713-18716
8. Ribeiro-Resende, V. T., Ribeiro-Guimarães, M. L., Lemes, R. M., Nascimento, I. C., Alves, L., Mendez-Otero, R., Pessolani, M. C., and Lara, F. A. (2010) Involvement of 9-O-Acetyl GD3 ganglioside in Mycobacterium leprae infection of Schwann cells. *J Biol Chem* **285**, 34086-34096
9. Ichikawa, S., and Hirabayashi, Y. (1998) Glucosylceramide synthase and glycosphingolipid synthesis. *Trends Cell Biol* **8**, 198-202
10. Morad, S. A., and Cabot, M. C. (2013) Ceramide-orchestrated signalling in cancer cells. *Nat Rev Cancer* **13**, 51-65
11. Ito, M., and Komori, H. (1996) Homeostasis of cell-surface glycosphingolipid content in B16 melanoma cells. Evidence revealed by an endoglycoceramidase. *J Biol Chem* **271**, 12655-12660
12. Liu, Y. Y., Hill, R. A., and Li, Y. T. (2013) Ceramide glycosylation catalyzed by glucosylceramide synthase and cancer drug resistance. *Adv Cancer Res* **117**, 59-89
13. Liu, Y. Y., Han, T. Y., Giuliano, A. E., Ichikawa, S., Hirabayashi, Y., and Cabot, M. C. (1999) Glycosylation of ceramide potentiates cellular resistance to tumor necrosis factor-alpha-induced apoptosis. *Exp Cell Res* **252**, 464-470
14. Bleicher, R. J., and Cabot, M. C. (2002) Glucosylceramide synthase and apoptosis. *Biochim Biophys Acta* **1585**, 172-178
15. Platt, F. M. (2014) Sphingolipid lysosomal storage disorders. *Nature* **510**, 68-75
16. Marques, A. R., Mirzaian, M., Akiyama, H., Wisse, P., Ferraz, M. J., Gaspar, P., Ghauharali-van der Vlugt, K., Meijer, R., Giraldo, P., Alfonso, P., Irún, P., Dahl, M., Karlsson, S., Pavlova, E. V., Cox, T. M., Scheij, S., Verhoek, M., Ottenhoff, R., van Roomen, C. P., Pannu, N. S., van Eijk, M., Dekker, N., Boot, R. G., Overkleeft, H. S., Blommaart, E., Hirabayashi, Y., and Aerts, J. M. (2016) Glucosylated cholesterol in mammalian cells and tissues: formation and degradation by multiple cellular β-glucosidases. *J Lipid Res*
17. Kallemeijn, W. W., Li, K. Y., Witte, M. D., Marques, A. R., Aten, J., Scheij, S., Jiang, J., Willems, L. I., Voorn-Brouwer, T. M., van Roomen, C. P., Ottenhoff, R., Boot, R. G., van den Elst, H., Walvoort, M. T., Florea, B. I., Codee, J. D., van der Marel, G. A., Aerts, J. M., and Overkleeft, H. S. (2012) Novel activity-based probes for broad-spectrum profiling of retaining beta-exoglucosidases in situ and in vivo. *Angewandte Chemie* **51**, 12529-12533

18. Meikle, P. J., Hopwood, J. J., Clague, A. E., and Carey, W. F. (1999) Prevalence of lysosomal storage disorders. *JAMA* **281**, 249-254
19. Bainton, D. F. (1981) The discovery of lysosomes. *J Cell Biol* **91**, 66s-76s
20. Hers, H. G. (1963) alpha-Glucosidase deficiency in generalized glycogenstorage disease (Pompe's disease). *Biochem J* **86**, 11-16
21. DEDUVE, C. (1964) FROM CYTASES TO LYSOSOMES. *Fed Proc* **23**, 1045-1049
22. Deegan, P. B., and Cox, T. M. (2012) Imiglucerase in the treatment of Gaucher disease: a history and perspective. *Drug Des Devel Ther* **6**, 81-106
23. Hult, M., Darin, N., von Dobeln, U., and Mansson, J. E. (2014) Epidemiology of lysosomal storage diseases in Sweden. *Acta Paediatr* **103**, 1258-1263
24. Eng, C. M., Schechter, C., Robinowitz, J., and et al. (1997) Prenatal genetic carrier testing using triple disease screening. *JAMA* **278**, 1268-1272
25. Jmoudiak, M., and Futerman, A. H. (2005) Gaucher disease: pathological mechanisms and modern management. *Br J Haematol* **129**, 178-188
26. Biegstraaten, M., van Schaik, I. N., Aerts, J. M., Langeveld, M., Mannens, M. M., Bour, L. J., Sidransky, E., Tayebi, N., Fitzgibbon, E., and Hollak, C. E. (2011) A monozygotic twin pair with highly discordant Gaucher phenotypes. *Blood Cells Mol Dis* **46**, 39-41
27. Vockley, J., Rinaldo, P., Bennett, M. J., Matern, D., and Vladutiu, G. D. (2000) Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. *Mol Genet Metab* **71**, 10-18
28. Sidransky, E. (2004) Gaucher disease: complexity in a "simple" disorder. *Mol Genet Metab* **83**, 6-15
29. Klein, A. D., Ferreira, N. S., Ben-Dor, S., Duan, J., Hardy, J., Cox, T. M., Merrill, A. H., Jr., and Futerman, A. H. (2016) Identification of Modifier Genes in a Mouse Model of Gaucher Disease. *Cell Rep* **16**, 2546-2553
30. Tayebi, N., Stubblefield, B. K., Park, J. K., Orvisky, E., Walker, J. M., LaMarca, M. E., and Sidransky, E. (2003) Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. *Am J Hum Genet* **72**, 519-534
31. Dandana, A., Ben Khelifa, S., Chahed, H., Miled, A., and Ferchichi, S. (2016) Gaucher Disease: Clinical, Biological and Therapeutic Aspects. *Pathobiology* **83**, 13-23
32. Elstein, D., Abrahamov, A., Hadas-Halpern, I., and Zimran, A. (2001) Gaucher's disease. *Lancet* **358**, 324–327
33. Koprivica, V., Stone, D. L., Park, J. K., Callahan, M., Frisch, A., Cohen, I. J., Tayebi, N., and Sidransky, E. (2000) Analysis and classification of 304 mutant alleles in patients with type 1 and type 3 Gaucher disease. *Am J Hum Genet* **66**, 1777-1786
34. Pastores, G. M., and Hughes, D. A. (2000) Gaucher Disease. in *GeneReviews(R)* (Pagon, R. A., Adam, M. P., Ardinger, H. H., Wallace, S. E., Amemiya, A., Bean, L. J. H., Bird, T. D., Ledbetter, N., Mefford, H. C., Smith, R. J. H., and Stephens, K. eds.), Seattle (WA). pp
35. Hruska, K. S., LaMarca, M. E., Scott, C. R., and Sidransky, E. (2008) Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Human mutation* **29**, 567-583
36. Siebert, M., Bock, H., Michelin-Tirelli, K., Coelho, J. C., Giugliani, R., and Saraiva-Pereira, M. L. (2013) Novel mutations in the glucocerebrosidase gene of brazilian patients with Gaucher disease. *JIMD Rep* **9**, 7-16
37. Beutler, E., Gelbart, T., and Scott, C. R. (2005) Hematologically important mutations: Gaucher disease. *Blood Cells Mol Dis* **35**, 355-364
38. Dvir, H., Harel, M., McCarthy, A. A., Toker, L., Silman, I., Futerman, A. H., and Sussman, J. L. (2003) X-ray structure of human acid-beta-glucosidase, the defective enzyme in Gaucher disease. *EMBO reports* **4**, 704-709
39. Chen, M., and Wang, J. (2008) Gaucher disease: review of the literature. *Arch Pathol Lab Med* **132**, 851-853

40. Grabowski, G. A. (2008) Phenotype, diagnosis, and treatment of Gaucher's disease. *Lancet* **372**, 1263-1271
41. Liou, B., Kazimierczuk, A., Zhang, M., Scott, C. R., Hegde, R. S., and Grabowski, G. A. (2006) Analyses of variant acid beta-glucosidases: effects of Gaucher disease mutations. *J Biol Chem* **281**, 4242-4253
42. Goddard-Borger, E. D., Wennekes, T., and Withers, S. G. (2010) Molecular probes: Getting lucky in the lysosome. *Nat Chem Biol* **6**, 881-883
43. Filocamo, M., Mazzotti, R., Stroppiano, M., Seri, M., Giona, F., Parenti, G., Regis, S., Corsolini, F., Zoboli, S., and Gatti, R. (2002) Analysis of the glucocerebrosidase gene and mutation profile in 144 Italian gaucher patients. *Human mutation* **20**, 234-235
44. Tsuji, S., Martin, B. M., Barranger, J. A., Stubblefield, B. K., LaMarca, M. E., and Ginns, E. I. (1988) Genetic heterogeneity in type 1 Gaucher disease: multiple genotypes in Ashkenazic and non-Ashkenazic individuals. *PNAS* **85**, 2349-2352
45. Elstein, D., Scott, C. R., Zeigler, M., Abrahamov, A., and Zimran, A. (2005) Phenotypic heterogeneity in patients with Gaucher disease and the N370S/V394L genotype. *Genet Test* **9**, 26-29
46. Chabas, A., Cormand, B., Grinberg, D., Burguera, J. M., Balcells, S., Merino, J. L., Mate, I., Sobrino, J. A., Gonzalez-Duarte, R., and Vilageliu, L. (1995) Unusual expression of Gaucher's disease: cardiovascular calcifications in three sibs homozygous for the D409H mutation. *J Med Genet* **32**, 740-742
47. Cormand, B., Vilageliu, L., Burguera, J. M., Balcells, S., Gonzalez-Duarte, R., Grinberg, D., and Chabas, A. (1995) Gaucher disease in Spanish patients: analysis of eight mutations. *Human mutation* **5**, 303-309
48. Goker-Alpan, O., Hruska, K. S., Orvisky, E., Kishnani, P. S., Stubblefield, B. K., Schiffmann, R., and Sidransky, E. (2005) Divergent phenotypes in Gaucher disease implicate the role of modifiers. *J Med Genet* **42**, e37
49. Ida, H., Rennert, O. M., Kawame, H., Ito, T., Maekawa, K., and Eto, Y. (1996) Mutation screening of 17 Japanese patients with neuropathic Gaucher disease. *Hum Genet* **98**, 167-171
50. Ginns, E. I., Choudary, P. V., Tsuji, S., Martin, B., Stubblefield, B., Sawyer, J., Hozier, J., and Barranger, J. A. (1985) Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implications for Gaucher disease. *PNAS* **82**, 7101-7105
51. Xu, Y. H., Quinn, B., Witte, D., and Grabowski, G. A. (2003) Viable mouse models of acid beta-glucosidase deficiency: the defect in Gaucher disease. *Am J Pathol* **163**, 2093-2101
52. Tybulewicz, V. L., Tremblay, M. L., LaMarca, M. E., Willemsen, R., Stubblefield, B. K., Winfield, S., Zablocka, B., Sidransky, E., Martin, B. M., Huang, S. P., and et al. (1992) Animal model of Gaucher's disease from targeted disruption of the mouse glucocerebrosidase gene. *Nature* **357**, 407-410
53. Miao, S., McCarter, J. D., Grace, M. E., Grabowski, G. A., Aebersold, R., and Withers, S. G. (1994) Identification of Glu340 as the active-site nucleophile in human glucocerebrosidase by use of electrospray tandem mass spectrometry. *J Biol Chem* **269**, 10975-10978
54. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1996) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *PNAS* **93**, 5674
55. Fabrega, S., Durand, P., Codogno, P., Bauvy, C., Delomenie, C., Henrissat, B., Martin, B. M., McKinney, C., Ginns, E. I., Mornon, J. P., and Lehn, P. (2000) Human glucocerebrosidase: heterologous expression of active site mutants in murine null cells. *Glycobiology* **10**, 1217-1224
56. Pol-Fachin, L., Siebert, M., Verli, H., and Saraiva-Pereira, M. L. (2016) Glycosylation is crucial for a proper catalytic site organization in human glucocerebrosidase. *Glycoconj J* **33**, 237-244

57. Berg-Fussman, A., Grace, M. E., Ioannou, Y., and Grabowski, G. A. (1993) Human acid beta-glucosidase. N-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. *J Biol Chem* **268**, 14861-14866
58. Grace, M. E., and Grabowski, G. A. (1990) Human acid beta-glucosidase: glycosylation is required for catalytic activity. *Biochem Biophys Res Commun* **168**, 771-777
59. Bergmann, J. E., and Grabowski, G. A. (1989) Posttranslational processing of human lysosomal acid beta-glucosidase: a continuum of defects in Gaucher disease type 1 and type 2 fibroblasts. *Am J Hum Genet* **44**, 741-750
60. Erickson, A. H., Ginns, E. I., and Barranger, J. A. (1985) Biosynthesis of the lysosomal enzyme glucocerebrosidase. *J Biol Chem* **260**, 14319-14324
61. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Water, P. (2008) *Molecular Biology of The Cell*, Fifth edition ed., Garland Science
62. Sidransky, E. (2005) Gaucher disease and parkinsonism. *Mol Genet Metab* **84**, 302-304
63. DeMaagd, G., and Philip, A. (2015) Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. *P T* **40**, 504-532
64. Aflaki, E., Westbroek, W., and Sidransky, E. (2017) The Complicated Relationship between Gaucher Disease and Parkinsonism: Insights from a Rare Disease. *Neuron* **93**, 737-746
65. Halperin, A., Elstein, D., and Zimran, A. (2006) Increased incidence of Parkinson disease among relatives of patients with Gaucher disease. *Blood Cells Mol Dis* **36**, 426-428
66. Sidransky, E., Nalls, M. A., Aasly, J. O., Aharon-Peretz, J., Annesi, G., Barbosa, E. R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., Chen, C. M., Clark, L. N., Condroyer, C., De Marco, E. V., Durr, A., Eblan, M. J., Fahn, S., Farrer, M. J., Fung, H. C., Gan-Or, Z., Gasser, T., Gershoni-Baruch, R., Giladi, N., Griffith, A., Gurevich, T., Januario, C., Kropp, P., Lang, A. E., Lee-Chen, G. J., Lesage, S., Marder, K., Mata, I. F., Mirelman, A., Mitsui, J., Mizuta, I., Nicoletti, G., Oliveira, C., Ottman, R., Orr-Urtreger, A., Pereira, L. V., Quattrone, A., Rogaeva, E., Rolfs, A., Rosenbaum, H., Rozenberg, R., Samii, A., Samaddar, T., Schulte, C., Sharma, M., Singleton, A., Spitz, M., Tan, E. K., Tayebi, N., Toda, T., Troiano, A. R., Tsuji, S., Wittstock, M., Wolfsberg, T. G., Wu, Y. R., Zabetian, C. P., Zhao, Y., and Ziegler, S. G. (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-1661
67. Davis, M. Y., Johnson, C. O., Leverenz, J. B., Weintraub, D., Trojanowski, J. Q., Chen-Plotkin, A., Van Deerlin, V. M., Quinn, J. F., Chung, K. A., Peterson-Hiller, A. L., Rosenthal, L. S., Dawson, T. M., Albert, M. S., Goldman, J. G., Stebbins, G. T., Bernard, B., Wszolek, Z. K., Ross, O. A., Dickson, D. W., Eidelberg, D., Mattis, P. J., Niethammer, M., Yearout, D., Hu, S. C., Cholerton, B. A., Smith, M., Mata, I. F., Montine, T. J., Edwards, K. L., and Zabetian, C. P. (2016) Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive Progression in Parkinson Disease. *JAMA Neurol* **73**, 1217-1224
68. Pfefferkorn, C. M., Jiang, Z., and Lee, J. C. (2012) Biophysics of alpha-synuclein membrane interactions. *Biochim Biophys Acta* **1818**, 162-171
69. Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E. S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W. G., Lazzarini, A. M., Duvoisin, R. C., Di Iorio, G., Golbe, L. I., and Nussbaum, R. L. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045-2047
70. Mazzulli, J. R., Xu, Y. H., Sun, Y., Knight, A. L., McLean, P. J., Caldwell, G. A., Sidransky, E., Grabowski, G. A., and Krainc, D. (2011) Gaucher disease glucocerebrosidase and

alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell* **146**, 37-52
71. Jung, O., Patnaik, S., Marugan, J., Sidransky, E., and Westbroek, W. (2016) Progress and potential of non-inhibitory small molecule chaperones for the treatment of Gaucher disease and its implications for Parkinson disease. *Expert Rev Proteomics* **13**, 471-479
72. Brady, R. O., Pentchev, P. G., Gal, A. E., Hibbert, S. R., and Dekaban, A. S. (1974) Replacement therapy for inherited enzyme deficiency. Use of purified glucocerebrosidase in Gaucher's disease. *N Engl J Med* **291**, 989-993
73. Pentchev, P. G., Brady, R. O., Hibbert, S. R., Gal, A. E., and Shapiro, D. (1973) Isolation and characterization of glucocerebrosidase from human placental tissue. *J Biol Chem* **248**, 5256-5261
74. Furbish, F. S., Blair, H. E., Shiloach, J., Pentchev, P. G., and Brady, R. O. (1977) Enzyme replacement therapy in Gaucher's disease: large-scale purification of glucocerebrosidase suitable for human administration. *PNAS* **74**, 3560-3563
75. Mistry, P. K., Lopez, G., Schiffmann, R., Barton, N. W., Weinreb, N. J., and Sidransky, E. (2017) Gaucher disease: Progress and ongoing challenges. *Mol Genet Metab* **120**, 8-21
76. Clarke, T. (2010) Special Report: Inside the battle for Genzyme's future. Reuters, Reuters
77. Altarescu, G., Hill, S., Wiggs, E., Jeffries, N., Kreps, C., Parker, C. C., Brady, R. O., Barton, N. W., and Schiffmann, R. (2001) The efficacy of enzyme replacement therapy in patients with chronic neuronopathic Gaucher's disease. *J Pediatr* **138**, 539-547
78. Weinreb, N. J., Goldblatt, J., Villalobos, J., Charrow, J., Cole, J. A., Kerstenetzky, M., vom Dahl, S., and Hollak, C. (2013) Long-term clinical outcomes in type 1 Gaucher disease following 10 years of imiglucerase treatment. *J Inherit Metab Dis* **36**, 543-553
79. Engelberg, A. B., Kesselheim, A. S., and Avorn, J. (2009) Balancing innovation, access, and profits--market exclusivity for biologics. *N Engl J Med* **361**, 1917-1919
80. Serratrice, C., Carballo, S., Serratrice, J., and Stirnemann, J. (2016) Imiglucerase in the management of Gaucher disease type 1: an evidence-based review of its place in therapy. *Core Evid* **11**, 37-47
81. Cox, T., Lachmann, R., Hollak, C., Aerts, J., van Weely, S., Hrebícek, M., Platt, F., Butters, T., Dwek, R., Moyses, C., Gow, I., Elstein, D., and Zimran, A. (2000) Novel oral treatment of Gaucher's disease with N-butyldeoxynojirimycin (OGT 918) to decrease substrate biosynthesis. *Lancet* **355**, 1481-1485
82. Zimran, A., and Elstein, D. (2003) Gaucher disease and the clinical experience with substrate reduction therapy. *Philos Trans R Soc Lond B Biol Sci* **358**, 961-966
83. Balwani, M., Burrow, T. A., Charrow, J., Goker-Alpan, O., Kaplan, P., Kishnani, P. S., Mistry, P., Ruskin, J., and Weinreb, N. (2016) Recommendations for the use of eliglustat in the treatment of adults with Gaucher disease type 1 in the United States. *Mol Genet Metab* **117**, 95-103
84. Shayman, J. A. (2013) The design and clinical development of inhibitors of glycosphingolipid synthesis: will invention be the mother of necessity? *Trans Am Clin Climatol Assoc* **124**, 46-60
85. Cachón-González, M. B., Wang, S. Z., Lynch, A., Ziegler, R., Cheng, S. H., and Cox, T. M. (2006) Effective gene therapy in an authentic model of Tay-Sachs-related diseases. *PNAS* **103**, 10373-10378
86. Bradbury, A. M., Peterson, T. A., Gross, A. L., Wells, S. Z., McCurdy, V. J., Wolfe, K. G., Dennis, J. C., Brunson, B. L., Gray-Edwards, H., Randle, A. N., Johnson, A. K., Morrison, E. E., Cox, N. R., Baker, H. J., Sena-Esteves, M., and Martin, D. R. (2017) AAV-mediated gene delivery attenuates neuroinflammation in feline Sandhoff disease. *Neuroscience* **340**, 117-125

87. Mistry, P. K., Belmatoug, N., vom Dahl, S., and Giugliani, R. (2015) Understanding the natural history of Gaucher disease. *Am J Hematol* **90 Suppl 1**, S6-11
88. Gray-Edwards, H. L., Brunson, B. L., Holland, M., Hespel, A. M., Bradbury, A. M., McCurdy, V. J., Beadlescomb, P. M., Randle, A. N., Salibi, N., Denney, T. S., Beyers, R. J., Johnson, A. K., Voyles, M. L., Montgomery, R. D., Wilson, D. U., Hudson, J. A., Cox, N. R., Baker, H. J., Sena-Esteves, M., and Martin, D. R. (2015) Mucopolysaccharidosis-like phenotype in feline Sandhoff disease and partial correction after AAV gene therapy. *Mol Genet Metab* **116**, 80-87
89. Lyseng-Williamson, K. A. (2014) Miglustat: a review of its use in Niemann-Pick disease type C. *Drugs* **74**, 61-74
90. Stein, V. M., Crooks, A., Ding, W., Prociuk, M., O'Donnell, P., Bryan, C., Sikora, T., Dingemanse, J., Vanier, M. T., Walkley, S. U., and Vite, C. H. (2012) Miglustat improves purkinje cell survival and alters microglial phenotype in feline Niemann-Pick disease type C. *J Neuropathol Exp Neurol* **71**, 434-448
91. Villamizar-Schiller, I. T., Pabon, L. A., Hufnagel, S. B., Serrano, N. C., Karl, G., Jefferies, J. L., Hopkin, R. J., and Prada, C. E. (2015) Neurological and cardiac responses after treatment with miglustat and a ketogenic diet in a patient with Sandhoff disease. *Eur J Med Genet* **58**, 180-183
92. Ashe, K. M., Bangari, D., Li, L., Cabrera-Salazar, M. A., Bercury, S. D., Nietupski, J. B., Cooper, C. G., Aerts, J. M., Lee, E. R., Copeland, D. P., Cheng, S. H., Scheule, R. K., and Marshall, J. (2011) Iminosugar-based inhibitors of glucosylceramide synthase increase brain glycosphingolipids and survival in a mouse model of Sandhoff disease. *PLoS One* **6**, e21758
93. Maor, G., Rencus-Lazar, S., Filocamo, M., Steller, H., Segal, D., and Horowitz, M. (2013) Unfolded protein response in Gaucher disease: from human to Drosophila. *Orphanet J Rare Dis* **8**, 140
94. Fan, J. Q. (2003) A contradictory treatment for lysosomal storage disorders: inhibitors enhance mutant enzyme activity. *Trends Pharmacol Sci* **24**, 355-360
95. Lieberman, R. L., D'aquino, J. A., Ringe, D., and Petsko, G. A. (2009) Effects of pH and iminosugar pharmacological chaperones on lysosomal glycosidase structure and stability. *Biochemistry* **48**, 4816-4827
96. Sun, Y., Ran, H., Liou, B., Quinn, B., Zamzow, M., Zhang, W., Bielawski, J., Kitatani, K., Setchell, K. D., Hannun, Y. A., and Grabowski, G. A. (2011) Isofagomine in vivo effects in a neuronopathic Gaucher disease mouse. *PLoS One* **6**, e19037
97. Alfonso, P., Pampin, S., Estrada, J., Rodriguez-Rey, J. C., Giraldo, P., Sancho, J., and Pocovi, M. (2005) Miglustat (NB-DNJ) works as a chaperone for mutated acid β-glucosidase in cells transfected with several Gaucher disease mutations. *Blood Cells Mol Dis* **35**, 268-276
98. Sawkar, A. R., Cheng, W.-C., Beutler, E., Wong, C.-H., Balch, W. E., and Kelly, J. W. (2002) Chemical chaperones increase the cellular actuvuty of N370S β-glucosidase:A therapeutic strategy for Gaucher disease. *PNAS* **99**, 15428-15433
99. Sun, Y., Liou, B., Xu, Y. H., Quinn, B., Zhang, W., Hamler, R., Setchell, K. D., and Grabowski, G. A. (2012) Ex vivo and in vivo effects of isofagomine on acid β-glucosidase variants and substrate levels in Gaucher disease. *J Biol Chem* **287**, 4275-4287
100. Pandey, S., Sree, A., Dash, S. S., Sethi, D. P., and Chowdhury, L. (2013) Diversity of marine bacteria producing beta-glucosidase inhibitors. *Microb Cell Fact* **12**, 35
101. Atsumi, S., Umezawa, K., Iinuma, H., Naganawa, H., Nakamura, H., Iitaka, Y., and Takeuchi, T. (1990) Production, isolation and structure determination of a novel beta-glucosidase inhibitor, cyclophellitol, from Phellinus sp. *J Antibiot (Tokyo)* **43**, 49-53
102. Vichasilp, C., Nakagawa, K., Sookwong, P., Higuchi, O., Luemunkong, S., and Miyazawa, T. (2012) Development of high 1-deoxynojirimycin (DNJ) content mulberry tea and use of response surface methodology to optimize tea-making

conditions for highest DNJ extraction. *LWT - Food Science and Technology* **45**, 226-232
103. Li, Y. G., Ji, D. F., Zhong, S., Lin, T. B., Lv, Z. Q., Hu, G. Y., and Wang, X. (2013) 1-deoxynojirimycin inhibits glucose absorption and accelerates glucose metabolism in streptozotocin-induced diabetic mice. *Sci Rep* **3**, 1377
104. Brumshtein, B., Greenblatt, H. M., Butters, T. D., Shaaltiel, Y., Aviezer, D., Silman, I., Futerman, A. H., and Sussman, J. L. (2007) Crystal structures of complexes of N-butyl- and N-nonyl-deoxynojirimycin bound to acid beta-glucosidase: insights into the mechanism of chemical chaperone action in Gaucher disease. *J Biol Chem* **282**, 29052-29058
105. Lieberman, R. L., Wustman, B. A., Huertas, P., Powe, A. C., Pine, C. W., Khanna, R., Schlossmacher, M. G., Ringe, D., and Petsko, G. A. (2007) Structure of acid beta-glucosidase with pharmacological chaperone provides insight into Gaucher disease. *Nat Chem Biol* **3**, 101-107
106. Khanna, R., Benjamin, E. R., Pellegrino, L., Schilling, A., Rigat, B. A., Soska, R., Nafar, H., Ranes, B. E., Feng, J., Lun, Y., Powe, A. C., Palling, D. J., Wustman, B. A., Schiffmann, R., Mahuran, D. J., Lockhart, D. J., and Valenzano, K. J. (2010) The pharmacological chaperone isofagomine increases the activity of the Gaucher disease L444P mutant form of beta-glucosidase. *FEBS J* **277**, 1618-1638
107. Wennekes, T., Meijer, A. J., Groen, A. K., Boot, R. G., Groener, J. E., van Eijk, M., Ottenhoff, R., Bijl, N., Ghauharali, K., Song, H., O'Shea, T. J., Liu, H. L., Yew, N., Copeland, D., van den Berg, R. J., van der Marel, G. A., Overkleeft, H. S., and Aerts, J. M. (2010) Dual-Action Lipophilic Iminosugar Improves Glycemic Control in Obese Rodents by Reduction of Visceral Glycosphingolipids and Buffering of Carbohydrate Assimilation. *J. Med. Chem.* **53**, 689-698
108. Overkleeft, H. S., Renkema, G. H., Neele, J., Vianello, P., Hung, I. O., Strijland, A., van der Burg, A. M., Koomen, G. J., Pandit, U. K., and Aerts, J. M. (1998) Generation of specific deoxynojirimycin-type inhibitors of the non-lysosomal glucosylceramidase. *Journal of Biological Chemistry* **273**, 26522-26527
109. Okumiya, T., Kroos, M. A., Vliet, L. V., Takeuchi, H., Van der Ploeg, A. T., and Reuser, A. J. J. (2007) Chemical chaperones improve transport and enhance stability of mutant α-glucosidases in glycogen storage disease type II. *Mol Genet Metab* **90**, 49-57
110. Brumshtein, B., Greenblatt, H. M., Butters, T. D., Shaaltiel, Y., Aviezer, D., Silman, I., Futerman, A. H., and Sussman, J. L. (2007) Crystal structures of complexes of N-butyl- and N-nonyl-deoxynojirimycin bound to acid beta-glucosidase: insights into the mechanism of chemical chaperone action in Gaucher disease. *Journal of Biological Chemistry* **282**, 29052-29058
111. Ben Bdira, F., Kallemeijn, W. W., Oussoren, S. V., Scheij, S., Bleijlevens, B., Florea, B. I., van Roomen, C. P. A. A., Ottenhoff, R., van Kooten, M. J. F. M., Walvoort, M. T. C., Witte, M. D., Boot, R. G., Ubbink, M., Overkleeft, H. S., and Aerts, J. M. F. G. (2017) Stabilization of Glucocerebrosidase by Active Site Occupancy. *ACS Chemical Biology* **12**, 1830-1841
112. Withers, S. G., and Umezawa, K. (1991) Cyclophellitol: a naturally occurring mechanism-based inactivator of beta-glucosidases. *Biochem Biophys Res Commun* **177**, 532-537
113. Kacher, Y., Brumshtein, B., Boldin-Adamsky, S., Toker, L., Shainskaya, A., Silman, I., Sussman, J. L., and Futerman, A. H. (2008) Acid beta-glucosidase: insights from structural analysis and relevance to Gaucher disease therapy. *Biol Chem* **389**, 1361-1369
114. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495

115. Beenakker, T. J. M., Wander, D. P. A., Offen, W. A., Artola, M., Raich, L., Ferraz, M. J., Li, K. Y., Houben, J., van Rijssel, E. R., Hansen, T., van der Marel, G. A., Codee, J. D. C., Aerts, J., Rovira, C., Davies, G. J., and Overkleeft, H. S. (2017) Carba-cyclophellitols Are Neutral Retaining-Glucosidase Inhibitors. *J. Am. Chem. Soc.* **139**, 6534-6537
116. Gloster, T. M., Madsen, R., and Davies, G. J. (2007) Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Org Biomol Chem* **5**, 444-446
117. Willems, L. I., Overkleeft, H. S., and van Kasteren, S. I. (2014) Current developments in activity-based protein profiling. *Bioconjug Chem* **25**, 1181-1191
118. Withers, S. G., and Aebersold, R. (1995) Approaches to labeling and identification of active site residues in glycosidases. *Protein Sci* **4**, 361-372
119. Willems, L. I., van der Linden, W. A., Li, N., Li, K. Y., Liu, N., Hoogendoorn, S., van der Marel, G. A., Florea, B. I., and Overkleeft, H. S. (2011) Bioorthogonal chemistry: applications in activity-based protein profiling. *Acc Chem Res* **44**, 718-729
120. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-based protein profiling: the serine hydrolases. *Proceedings of the National Academy of Sciences of the U.S.A.* **96**, 14694-14699
121. Greenbaum, D., Medzihradszky, K. F., Burlingame, A., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem Biol* **7**, 569-581
122. Evans, M. J., and Cravatt, B. F. (2006) Mechanism-based profiling of enzyme families. *Chemical Reviews* **106**, 3279-3301
123. Willems, L. I., Jiang, J., Li, K. Y., Witte, M. D., Kallemeijn, W. W., Beenakker, T. J., Schroder, S. P., Aerts, J. M., van der Marel, G. A., Codee, J. D., and Overkleeft, H. S. (2014) From covalent glycosidase inhibitors to activity-based glycosidase probes. *Chemistry-A European Journal* **20**, 10864-10872
124. Li, K. Y., Jiang, J., Witte, M. D., Kallemeijn, W. W., Donker-Koopman, W. E., Boot, R. G., Aerts, J. M., Codee, J. D., van der Marel, G. A., and Overkleeft, H. S. (2014) Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors. *Org Biomol Chem* **12**, 7786-7791
125. Artola, M., Wu, L., Ferraz, M. J., Kuo, C. L., Raich, L., Breen, I. Z., Offen, W. A., Codee, J. D. C., van der Marel, G. A., Rovira, C., Aerts, J., Davies, G. J., and Overkleeft, H. S. (2017) 1,6-Cyclophellitol Cyclosulfates: A New Class of Irreversible Glycosidase Inhibitor. *ACS Central Science* **3**, 784-793
126. Jiang, J. B., Beenakker, T. J. M., Kallemeijn, W. W., van der Marel, G. A., van den Elst, H., Codee, J. D. C., Aerts, J. M. F. G., and Overkleeft, H. S. (2015) Comparing Cyclophellitol *N*-Alkyl and *N*-Acyl Cyclophellitol Aziridines as Activity-Based Glycosidase Probes. *Chemistry-A European Journal* **21**, 10861-10869
127. Witte, M. D., Kallemeijn, W. W., Aten, J., Li, K. Y., Strijland, A., Donker-Koopman, W. E., van den Nieuwendijk, A. M., Bleijlevens, B., Kramer, G., Florea, B. I., Hooibrink, B., Hollak, C. E., Ottenhoff, R., Boot, R. G., van der Marel, G. A., Overkleeft, H. S., and Aerts, J. M. (2010) Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat Chem Biol* **6**, 907-913
128. Aerts, J. M., Donker-Koopman, W. E., van der Vliet, M. K., Jonsson, L. M., Ginns, E. I., Murray, G. J., Barranger, J. A., Tager, J. M., and Schram, A. W. (1985) The occurrence of two immunologically distinguishable beta-glucocerebrosidases in human spleen. *Eur J Biochem* **150**, 565-574
129. Cobucci-Ponzano, B., Aurilia, V., Riccio, G., Henrissat, B., Coutinho, P. M., Strazzulli, A., Padula, A., Corsaro, M. M., Pieretti, G., Pocsfalvi, G., Fiume, I., Cannio, R., Rossi, M., and Moracci, M. (2010) A new archaeal beta-glycosidase from Sulfolobus solfataricus: seeding a novel retaining beta-glycan-specific glycoside hydrolase family along with the human non-lysosomal glucosylceramidase GBA2. *J Biol Chem* **285**, 20691-20703
130. Overkleeft, H. S., Renkema, G. H., Neele, J., Vianello, P., Hung, I. O., Strijland, A., van der Burg, A. M., Koomen, G. J., Pandit, U. K., and Aerts, J. M. (1998) Generation of

specific deoxynojirimycin-type inhibitors of the non-lysosomal glucosylceramidase. *J Biol Chem* **273**, 26522-26527
131. Martin, E., Schüle, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J. L., Gonzalez, M. A., Mundwiller, E., Deconinck, T., Wessner, M., Jornea, L., Oteyza, A. C., Durr, A., Martin, J. J., Schöls, L., Mhiri, C., Lamari, F., Züchner, S., De Jonghe, P., Kabashi, E., Brice, A., and Stevanin, G. (2013) Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. *Am J Hum Genet* **92**, 238-244
132. Hammer, M. B., Eleuch-Fayache, G., Schottlaender, L. V., Nehdi, H., Gibbs, J. R., Arepalli, S. K., Chong, S. B., Hernandez, D. G., Sailer, A., Liu, G., Mistry, P. K., Cai, H., Shrader, G., Sassi, C., Bouhlal, Y., Houlden, H., Hentati, F., Amouri, R., and Singleton, A. B. (2013) Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity. *Am J Hum Genet* **92**, 245-251
133. Sultana, S., Reichbauer, J., Schüle, R., Mochel, F., Synofzik, M., and van der Spoel, A. C. (2015) Lack of enzyme activity in GBA2 mutants associated with hereditary spastic paraplegia/cerebellar ataxia (SPG46). *Biochem Biophys Res Commun* **465**, 35-40
134. Mistry, P. K., Liu, J., Sun, L., Chuang, W. L., Yuen, T., Yang, R., Lu, P., Zhang, K., Li, J., Keutzer, J., Stachnik, A., Mennone, A., Boyer, J. L., Jain, D., Brady, R. O., New, M. I., and Zaidi, M. (2014) Glucocerebrosidase 2 gene deletion rescues type 1 Gaucher disease. *PNAS* **111**, 4934-4939
135. Schonauer, S., Körschen, H. G., Penno, A., Rennhack, A., Breiden, B., Sandhoff, K., Gutbrod, K., Dörmann, P., Raju, D. N., Haberkant, P., Gerl, M. J., Brügger, B., Zigdon, H., Vardi, A., Futerman, A. H., Thiele, C., and Wachten, D. (2017) Identification of a feedback loop involving β-glucosidase 2 and its product sphingosine sheds light on the molecular mechanisms in Gaucher disease. *J Biol Chem* **292**, 6177-6189
136. Yildiz, Y., Hoffmann, P., Vom Dahl, S., Breiden, B., Sandhoff, R., Niederau, C., Horwitz, M., Karlsson, S., Filocamo, M., Elstein, D., Beck, M., Sandhoff, K., Mengel, E., Gonzalez, M. C., Nothen, M. M., Sidransky, E., Zimran, A., and Mattheisen, M. (2013) Functional and genetic characterization of the non-lysosomal glucosylceramidase 2 as a modifier for Gaucher disease. *Orphanet J Rare Dis* **8**, 151
137. Sorli, S. C., Colié, S., Albinet, V., Dubrac, A., Touriol, C., Guilbaud, N., Bedia, C., Fabriàs, G., Casas, J., Ségui, B., Levade, T., and Andrieu-Abadie, N. (2013) The nonlysosomal β-glucosidase GBA2 promotes endoplasmic reticulum stress and impairs tumorigenicity of human melanoma cells. *FASEB J* **27**, 489-498
138. Hayashi, Y., Okino, N., Kakuta, Y., Shikanai, T., Tani, M., Narimatsu, H., and Ito, M. (2007) Klotho-related protein is a novel cytosolic neutral beta-glycosylceramidase. *J Biol Chem* **282**, 30889-30900
139. Dekker, N., Voorn-Brouwer, T., Verhoek, M., Wennekes, T., Narayan, R. S., Speijer, D., Hollak, C. E., Overkleeft, H. S., Boot, R. G., and Aerts, J. M. (2011) The cytosolic β-glucosidase GBA3 does not influence type 1 Gaucher disease manifestation. *Blood Cells Mol Dis* **46**, 19-26
140. Tribolo, S., Berrin, J. G., Kroon, P. A., Czjzek, M., and Juge, N. (2007) The crystal structure of human cytosolic beta-glucosidase unravels the substrate aglycone specificity of a family 1 glycoside hydrolase. *J Mol Biol* **370**, 964-975
141. Kallemeijn, W. W., Witte, M. D., Voorn-Brouwer, T. M., Walvoort, M. T., Li, K. Y., Codée, J. D., van der Marel, G. A., Boot, R. G., Overkleeft, H. S., and Aerts, J. M. (2014) A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining β-glucosidases. *J Biol Chem*
142. Henrissat, B., and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Chem. Biol.* **7**, 637-644
143. www.cazy.org/.

144. Koshland, D. E. (1953) STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews* **28**, 416-436
145. Lingwood, C. A. (2011) Glycosphingolipid functions. *Cold Spring Harb Perspect Biol* **3**
146. Atsumi, S., Iinuma, H., Nosaka, C., and Umezawa, K. (1990) Biological activities of cyclophellitol. *J Antibiot (Tokyo)* **43**, 1579-1585
147. Atsumi, S., Nosaka, C., Iinuma, H., and Umezawa, K. (1992) Inhibition of glucocerebrosidase and induction of neural abnormality by cyclophellitol in mice. *Arch Biochem Biophys* **297**, 362-367
148. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A* **92**, 7090-7094
149. McNicholas, S., and Agirre, J. (2017) Glycoblocks: a schematic three-dimensional representation for glycans and their interactions. *Acta Crystallogr D Struct Biol* **73**, 187-194
150. Wei, R. R., Hughes, H., Boucher, S., Bird, J. J., Guziewicz, N., Van Patten, S. M., Qiu, H., Pan, C. Q., and Edmunds, T. (2011) X-ray and biochemical analysis of N370S mutant human acid β-glucosidase. *J Biol Chem* **286**, 299-308
151. Premkumar, L., Sawkar, A. R., Boldin-Adamsky, S., Toker, L., Silman, I., Kelly, J. W., Futerman, A. H., and Sussman, J. L. (2005) X-ray structure of human acid-beta-glucosidase covalently bound to conduritol-B-epoxide. Implications for Gaucher disease. *J Biol Chem* **280**, 23815-23819
152. Brumshtein, B., Wormald, M. R., Silman, I., Futerman, A. H., and Sussman, J. L. (2006) Structural comparison of differently glycosylated forms of acid-beta-glucosidase, the defective enzyme in Gaucher disease. *Acta Cryst. D* **62**, 1458-1465
153. Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S. A., Silman, I., Sussman, J. L., Futerman, A. H., and Aviezer, D. (2007) Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* **5**, 579-590
154. Brumshtein, B., Aguilar-Moncayo, M., García-Moreno, M. I., Ortiz Mellet, C., García Fernández, J. M., Silman, I., Shaaltiel, Y., Aviezer, D., Sussman, J. L., and Futerman, A. H. (2009) 6-Amino-6-deoxy-5,6-di-N-(N'-octyliminomethylidene)nojirimycin: synthesis, biological evaluation, and crystal structure in complex with acid beta-glucosidase. *Chembiochem* **10**, 1480-1485
155. Brumshtein, B., Salinas, P., Peterson, B., Chan, V., Silman, I., Sussman, J. L., Savickas, P. J., Robinson, G. S., and Futerman, A. H. (2010) Characterization of gene-activated human acid-beta-glucosidase: crystal structure, glycan composition, and internalization into macrophages. *Glycobiology* **20**, 24-32
156. Brumshtein, B., Aguilar-Moncayo, M., Benito, J. M., García Fernandez, J. M., Silman, I., Shaaltiel, Y., Aviezer, D., Sussman, J. L., Futerman, A. H., and Ortiz Mellet, C. (2011) Cyclodextrin-mediated crystallization of acid β-glucosidase in complex with amphiphilic bicyclic nojirimycin analogues. *Org Biomol Chem* **9**, 4160-4167
157. Orwig, S. D., Tan, Y. L., Grimster, N. P., Yu, Z., Powers, E. T., Kelly, J. W., and Lieberman, R. L. (2011) Binding of 3,4,5,6-tetrahydroxyazepanes to the acid-β-glucosidase active site: implications for pharmacological chaperone design for Gaucher disease. *Biochemistry* **50**, 10647-10657
158. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321 (Pt 2)**, 557-559
159. Willems, L. I., Beenakker, T. J., Murray, B., Scheij, S., Kallemeijn, W. W., Boot, R. G., Verhoek, M., Donker-Koopman, W. E., Ferraz, M. J., van Rijssel, E. R., Florea, B. I., Codee, J. D., van der Marel, G. A., Aerts, J. M., and Overkleeft, H. S. (2014) Potent and selective activity-based probes for GH27 human retaining alpha-galactosidases. *JACS* **136**, 11622-11625

160. Kabsch, W. (2010) XDS. *Acta Cryst. D* **66**, 125-132
161. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190
162. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst. D* **67**, 235-242
163. Lieberman, R. L., Wustman, B. A., Huertas, P., Jr, A. C. P., Pine, C. W., Khanna, R., Schlossmacher, M. G., Ringe, D., and Petsko, G. A. (2006) Structure of acid β-glucosidase with pharmacological chaperone provides insight into Gaucger disease. *Nature Chemical Biology* **3**, 101-107
164. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Cryst. D* **66**, 22-25
165. Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A., and Murshudov, G. N. (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Cryst. D* **68**, 431-440
166. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst. D* **53**, 240-255
167. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst. D* **66**, 486-501
168. Agirre, J., Iglesias-Fernandez, J., Rovira, C., Davies, G. J., Wilson, K. S., and Cowtan, K. D. (2015) Privateer: software for the conformational validation of carbohydrate structures. *Nat Struct Mol Biol* **22**, 833-834
169. Beenakker, T. J. M., Wander, D., Offen, W. A., Artola, M., Raich, L. s., Ferraz, M. J., Li, K.-Y., Houben, J. H. P. M., Rijssel, E. R. v., Hansen, T., Marel, G. A. v. d., Codée, J. D. C., Aerts, J. M. F. G., Rovira, C., Davies, G. J., and Overkleeft, H. S. (2017) Carba-Cyclophellitols are Neutral Retaining Glucosidase Inhibitors. *J Am Chem Soc* **accepted, online**
170. McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Cryst. D* **67**, 386-394
171. Chandrasekar, B., Colby, T., Emran Khan Emon, A., Jiang, J., Hong, T. N., Villamor, J. G., Harzen, A., Overkleeft, H. S., and van der Hoorn, R. A. (2014) Broad-range glycosidase activity profiling. *Mol Cell Proteomics* **13**, 2787-2800
172. Willems, L. I., Beenakker, T. J. M., Murray, B., Gagestein, B., van den Elst, H., van Rijssel, E. R., Codée, J. D. C., Kallemeijn, W. W., Aerts, J. M. F. G., van der Marel, G. A., and Overkleeft, H. S. (2014) Synthesis of α- and β-Galactopyranose-Configured Isomers of Cyclophellitol and Cyclophellitol Aziridine. *European Journal of Organic Chemistry* **2014**, 6044-6056
173. Gloster, T. M., Meloncelli, P., Stick, R. V., Zechel, D., Vasella, A., and Davies, G. J. (2007) Glycosidase inhibition: an assessment of the binding of 18 putative transition-state mimics. *J Am Chem Soc* **129**, 2345-2354
174. Ghisaidoobe, A., Bikker, P., de Bruijn, A. C., Godschalk, F. D., Rogaar, E., Guijt, M. C., Hagens, P., Halma, J. M., Van't Hart, S. M., Luitjens, S. B., van Rixel, V. H., Wijzenbroek, M., Zweegers, T., Donker-Koopman, W. E., Strijland, A., Boot, R., van der Marel, G., Overkleeft, H. S., Aerts, J. M., and van den Berg, R. J. (2011) Identification of potent and selective glucosylceramide synthase inhibitors from a library of N-alkylated iminosugars. *ACS Med Chem Lett* **2**, 119-123
175. Boot, R. G., Verhoek, M., Donker-Koopman, W., Strijland, A., van Marle, J., Overkleeft, H. S., Wennekes, T., and Aerts, J. M. (2007) Identification of the non-lysosomal glucosylceramidase as beta-glucosidase 2. *J Biol Chem* **282**, 1305-1312

176. Matern, H., Heinemann, H., Legler, G., and Matern, S. (1997) Purification and characterization of a microsomal bile acid beta-glucosidase from human liver. *J Biol Chem* **272**, 11261-11267
177. Futerman, A. H., and Platt, F. M. (2017) The metabolism of glucocerebrosides - From 1965 to the present. *Mol Genet Metab* **120**, 22-26
178. Citterio, A., Arnoldi, A., Panzeri, E., D'Angelo, M. G., Filosto, M., Dilena, R., Arrigoni, F., Castelli, M., Maghini, C., Germiniasi, C., Menni, F., Martinuzzi, A., Bresolin, N., and Bassi, M. T. (2014) Mutations in CYP2U1, DDHD2 and GBA2 genes are rare causes of complicated forms of hereditary spastic paraparesis. *J Neurol* **261**, 373-381
179. Ridley, C. M., Thur, K. E., Shanahan, J., Thillaiappan, N. B., Shen, A., Uhl, K., Walden, C. M., Rahim, A. A., Waddington, S. N., Platt, F. M., and van der Spoel, A. C. (2013) β-Glucosidase 2 (GBA2) activity and imino sugar pharmacology. *J Biol Chem* **288**, 26052-26066
180. Harzer, K., Blech-Hermoni, Y., Goldin, E., Felderhoff-Mueser, U., Igney, C., Sidransky, E., and Yildiz, Y. (2012) Beta-glucosidase 1 (GBA1) is a second bile acid β-glucosidase in addition to β-glucosidase 2 (GBA2). Study in β-glucosidase deficient mice and humans. *Biochem Biophys Res Commun* **423**, 308-312
181. Platt, F. M., and Jeyakumar, M. (2008) Substrate reduction therapy. *Acta Paediatr* **97**, 88-93
182. Nietupski, J. B., Pacheco, J. J., Chuang, W. L., Maratea, K., Li, L., Foley, J., Ashe, K. M., Cooper, C. G., Aerts, J. M., Copeland, D. P., Scheule, R. K., Cheng, S. H., and Marshall, J. (2012) Iminosugar-based inhibitors of glucosylceramide synthase prolong survival but paradoxically increase brain glucosylceramide levels in Niemann-Pick C mice. *Mol Genet Metab* **105**, 621-628
183. Mistry, P. K., Liu, J., Yang, M., Nottoli, T., McGrath, J., Jain, D., Zhang, K., Keutzer, J., Chuang, W. L., Mehal, W. Z., Zhao, H., Lin, A., Mane, S., Liu, X., Peng, Y. Z., Li, J. H., Agrawal, M., Zhu, L. L., Blair, H. C., Robinson, L. J., Iqbal, J., Sun, L., and Zaidi, M. (2010) Glucocerebrosidase gene-deficient mouse recapitulates Gaucher disease displaying cellular and molecular dysregulation beyond the macrophage. *PNAS* **107**, 19473-19478
184. Votsi, C., Zamba-Papanicolaou, E., Middleton, L. T., Pantzaris, M., and Christodoulou, K. (2014) A novel GBA2 gene missense mutation in spastic ataxia. *Ann Hum Genet* **78**, 13-22
185. Ferrara, M. C., Cobucci-Ponzano, B., Carpentieri, A., Henrissat, B., Rossi, M., Amoresano, A., and Moracci, M. (2014) The identification and molecular characterization of the first archaeal bifunctional exo-β-glucosidase/N-acetyl-β-glucosaminidase demonstrate that family GH116 is made of three functionally distinct subfamilies. *Biochim Biophys Acta* **1840**, 367-377
186. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-410
187. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217
188. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324
189. Sansenya, S., Mutoh, R., Charoenwattanasatien, R., Kurisu, G., and Ketudat Cairns, J. R. (2015) Expression and crystallization of a bacterial glycoside hydrolase family 116 beta-glucosidase from Thermoanaerobacterium xylanolyticum. *Acta crystallographica. Section F, Structural biology communications* **71**, 41-44
190. Charoenwattanasatien, R., Pengthaisong, S., Breen, I., Mutoh, R., Sansenya, S., Hua, Y., Tankrathok, A., Wu, L., Songsiriritthigul, C., Tanaka, H., Williams, S. J., Davies, G. J., Kurisu, G., and Cairns, J. R. (2016) Bacterial β-Glucosidase Reveals the Structural and Functional Basis of Genetic Defects in Human Glucocerebrosidase 2 (GBA2). *ACS Chem Biol* **11**, 1891-1900

191. van Weely, S., Brandsma, M., Strijland, A., Tager, J. M., and Aerts, J. M. (1993) Demonstration of the existence of a second, non-lysosomal glucocerebrosidase that is not deficient in Gaucher disease. *Biochim Biophys Acta* **1181**, 55-62
192. Yildiz, Y., Matern, H., Thompson, B., Allegood, J. C., Warren, R. L., Ramirez, D. M., Hammer, R. E., Hamra, F. K., Matern, S., and Russell, D. W. (2006) Mutation of beta-glucosidase 2 causes glycolipid storage disease and impaired male fertility. *J Clin Invest* **116**, 2985-2994
193. Gonzalez-Carmona, M. A., Sandhoff, R., Tacke, F., Vogt, A., Weber, S., Canbay, A. E., Rogler, G., Sauerbruch, T., Lammert, F., and Yildiz, Y. (2012) Beta-glucosidase 2 knockout mice with increased glucosylceramide show impaired liver regeneration. *Liver Int* **32**, 1354-1362
194. Körschen, H. G., Yildiz, Y., Raju, D. N., Schonauer, S., Bönigk, W., Jansen, V., Kremmer, E., Kaupp, U. B., and Wachten, D. (2013) The non-lysosomal β-glucosidase GBA2 is a non-integral membrane-associated protein at the endoplasmic reticulum (ER) and Golgi. *J Biol Chem* **288**, 3381-3393
195. McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Chem. Biol.* **4**, 885-892
196. Hohenschutz, L. D., Bell, E. A., Jewess, P. J., Leworthy, D. P., Pryce, R. J., Arnold, E., and Clardy, J. (1981) Castanospermine, A 1,6,7,8-tetrahydroxyoctahydroindolizine alkaloid, from seeds of Castanospermum australe. *Phytochemistry* **20**, 811-814
197. Suzuki, K., Sumitani, J., Nam, Y. W., Nishimaki, T., Tani, S., Wakagi, T., Kawaguchi, T., and Fushinobu, S. (2013) Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from Aspergillus aculeatus. *Biochem J* **452**, 211-221
198. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858
199. Heightman, T. D., and Vasella, A. T. (1999) Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases. *Angewandte Chemie International Edition* **38**, 750-770
200. Ellmers, B. R., Rhinehart, B. L., and Robinson, K. M. (1987) Castanospermine: an apparent tight-binding inhibitor of hepatic lysosomal alpha-glucosidase. *Biochemical pharmacology* **36**, 2381-2385
201. Allan, G., Ouadid-Ahidouch, H., Sanchez-Fernandez, E. M., Risquez-Cuadro, R., Fernandez, J. M., Ortiz-Mellet, C., and Ahidouch, A. (2013) New castanospermine glycoside analogues inhibit breast cancer cell proliferation and induce apoptosis without affecting normal cells. *PLoS One* **8**, e76411
202. Whitby, K., Pierson, T. C., Geiss, B., Lane, K., Engle, M., Zhou, Y., Doms, R. W., and Diamond, M. S. (2005) Castanospermine, a potent inhibitor of dengue virus infection in vitro and in vivo. *Journal of virology* **79**, 8698-8706
203. Chang, H.-H., Asano, N., Ishii, S., Ichikawa, Y., and Fan, J.-Q. (2006) Hydrophilic iminosugar active-site-specific chaperones increase residual glucocerebrosidase activity in fibroblasts from Gaucher patients. *FEBS Journal* **273**, 4082-4092
204. Agirre, J., Ariza, A., Offen, W. A., Turkenburg, J. P., Roberts, S. M., McNicholas, S., Harris, P. V., McBrayer, B., Dohnalek, J., Cowtan, K. D., Davies, G. J., and Wilson, K. S. (2016) Three-dimensional structures of two heavily N-glycosylated Aspergillus sp. family GH3 beta-D-glucosidases. *Acta Cryst. D* **72**, 254-265
205. Taylor, S. C., Ferguson, A. D., Bergeron, J. J., and Thomas, D. Y. (2004) The ER protein folding sensor UDP-glucose glycoprotein-glucosyltransferase modifies substrates distant to local changes in glycoprotein conformation. *Nat Struct Mol Biol* **11**, 128-134
206. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst. D* **60**, 2256-2268

207. Espina, G., Eley, K., Pompidor, G., Schneider, T. R., Crennell, S. J., and Danson, M. J. (2014) A novel β-xylosidase structure from Geobacillus thermoglucosidasius: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds. *Acta Cryst. D* **70**, 1366-1374
208. Bravman, T., Zolotnitsky, G., Belakhov, V., Shoham, G., Henrissat, B., Baasov, T., and Shoham, Y. (2003) Detailed kinetic analysis of a family 52 glycoside hydrolase: a beta-xylosidase from Geobacillus stearothermophilus. *Biochemistry* **42**, 10528-10536
209. Bravman, T., Belakhov, V., Solomon, D., Shoham, G., Henrissat, B., Baasov, T., and Shoham, Y. (2003) Identification of the catalytic residues in family 52 glycoside hydrolase, a beta-xylosidase from Geobacillus stearothermophilus T-6. *J Biol Chem* **278**, 26742-26749
210. Krissinel, E. (2012) Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem* **1**, 76-85
211. Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A., and Bertozzi, C. R. (2003) A chemical approach for identifying O-GlcNAc-modified proteins in cells. *PNAS* **100**, 9116-9121
212. Vocadlo, D. J., and Bertozzi, C. R. (2004) A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angewandte Chemie* **43**, 5338-5342
213. Willems, L. I., Beenakker, T. J. M., Murray, B., Gagestein, B., Elst, H. v. d., Rijssel, E. R. v., Codée, J. D. C., Kallemeijn, W. W., Aerts, J. M. F. G., Marel, G. A. v. d., and Overkleeft, H. S. (2014) Synthesis of α- and β-Galactopyranose-Configured Isomers of Cyclophellitol and Cyclophellitol Aziridine. *Eur. J. Org. Chem.*, 6044–6056
214. Jiang, J., Kallemeijn, W. W., Wright, D. W., van den Nieuwendijk, A. M. C. H., Rohde, V. C., Folch, E. C., van den Elst, H., Florea, B. I., Scheij, S., Donker-Koopman, W. E., Verhoek, M., Li, N., Schurmann, M., Mink, D., Boot, R. G., Codee, J. D. C., van der Marel, G. A., Davies, G. J., Aerts, J. M. F. G., and Overkleeft, H. S. (2015) In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 [small alpha]-l-fucosidases. *Chem Sci* **6**, 2782-2789
215. Caner, S., Zhang, X., Jiang, J., Chen, H. M., Nguyen, N. T., Overkleeft, H., Brayer, G. D., and Withers, S. G. (2016) Glucosyl epi-cyclophellitol allows mechanism-based inactivation and structural analysis of human pancreatic alpha-amylase. *FEBS Lett* **590**, 1143-1151
216. Wu, L., Jiang, J., Jin, Y., Kallemeijn, W. W., Kuo, C. L., Artola, M., Dai, W., van Elk, C., van Eijk, M., van der Marel, G. A., Codee, J. D. C., Florea, B. I., Aerts, J., Overkleeft, H. S., and Davies, G. J. (2017) Activity-based probes for functional interrogation of retaining beta-glucuronidases. *Nat. Chem. Biol.*
217. Kwan, D. H., Jin, Y., Jiang, J., Chen, H.-M., Kötzler, M. P., Overkleeft, H. S., Davies, G. J., and Withers, S. G. (2016) Chemoenzymatic synthesis of 6-phospho-cyclophellitol as a novel probe of 6-phospho-β-glucosidases. *FEBS Letters* **590**, 461-468
218. Yadav, A. K., Shen, D. L., Shan, X., He, X., Kermode, A. R., and Vocadlo, D. J. (2015) Fluorescence-quenched substrates for live cell imaging of human glucocerebrosidase activity. *J Am Chem Soc* **137**, 1181-1189
219. Lea, W. A., and Simeonov, A. (2011) Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov* **6**, 17-32
220. Coutinho, M. F., Santos, J. I., and Alves, S. (2016) Less Is More: Substrate Reduction Therapy for Lysosomal Storage Disorders. *Int J Mol Sci* **17**
221. Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011) Principles of early drug discovery. *Br J Pharmacol* **162**, 1239-1249
222. Ruiz de Garibay, A. P., Solinis, M. A., and Rodriguez-Gascon, A. (2013) Gene therapy for fabry disease: a review of the literature. *BioDrugs* **27**, 237-246

223. Kelly, J. M., Bradbury, A., Martin, D. R., and Byrne, M. E. (2017) Emerging therapies for neuropathic lysosomal storage disorders. *Prog. Neurobiol.* **152**, 166-180
224. Haskins, M. (2009) Gene therapy for lysosomal storage diseases (LSDs) in large animal models. *ILAR J* **50**, 112-121