Structural and Functional Characterisation of Human Lysosomal Glycoside Hydrolases: Insights into Lysosomal Storage Disorders

Rhianna Jayne Rowland

PhD

University of York

Chemistry

September 2021

Abstract

Carbohydrates are one of the most abundant and diverse class of biomolecules on earth, and their importance in life is reflected by the range of carbohydrate active enzymes (CAZymes) which have evolved to process glycans in nature. Unsurprisingly, defects in these CAZymes have been linked to a range of human diseases. Of note, lysosomal storage disorders (LSDs) are a group of inherited metabolic disorders which result from deficiencies in particular lysosomal glycoside hydrolases. Understanding the enzymes which underpin these disorders is crucial to the development of effective diagnostic and therapeutic strategies. In this regard, the work described here sought to aid in the development of novel activity-based probes (ABPs), inhibitors and molecular chaperones for GBA and α-GAL, the enzymes associated with the LSDs Gaucher Disease (GD) and Fabry Disease (FD) respectively, to support the study of these enzymes in disease pathogenesis, diagnosis and treatment.

In **Chapter 2**, the first ever co-crystal structures of human α-GAL in complex with several *galacto*-configured cyclophellitol-based inhibitors and ABPs for FD are described, providing key mechanistic and structural insight into their binding mode, reactivity and conformation. This work also expands to the characterisation of a non-covalent *galacto*-cyclophellitol cyclosulfamidate inhibitor, which exhibits chaperoning behaviour towards α-GAL as a potential pharmacological chaperone for the treatment of FD. In a parallel vein, the work presented in **Chapter 3** describes a series of crystal structures in which human GBA is complexed with a variety of cyclophellitol ABPs and inhibitors for GD. Indeed, these co-crystal complexes enabled a structure-guided approach to the development of more selective GBA inhibitors, which proved suitable for the generation of neuropathic GD animal models through chemical knockdown. To support such work, **Chapter 4** outlines the development of an insect-baculovirus expression vector system for the production of non-clinical human GBA, a long-standing academic target which has previously only been reliably produced in industrial labs by unknown means. Lastly, using this recombinant GBA formulation, **Chapter 5** discusses the structural analysis of a novel class of allylic carbasugar inhibitors of GBA, revealing multiple covalent and non-covalent mechanisms which may provide structural bases for their stabilising behaviour as potential chaperones for GD. It is ultimately hoped that this work will support future developments in the field of LSD research, particularly with regard to the diagnosis and treatment of such disorders.

*Part of the work presented in thesis has been published in the following papers: (**1**) R. Rowland et al., *Chem. Eur. J.,* **2021,** 27, 16377-16388 (**2**) R. Rowland et al., *Acta Cryst D*, 2020, **76**, 565-580, (**3**) M. Artola et al., *Chem Sci*, 2019, **10**, 9233-9243, (**4**) M. Artola et al., *J. Am. Chem. Soc*., 2019, **141**, 4214-4218 (**5**) S. P. Schröder et al., *ACS Cent. Sci*., 2019, **5**, 1067-1078.

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to give a heartfelt thank you to my supervisor Professor Gideon Davies for providing me with this opportunity and creating such a positive and enjoyable research environment. Aside from being an outstanding Professor with boundless knowledge of the field, he is also a warm-hearted and humble person whose support, encouragement and positivity have been second to none. I most certainly would not be here today, having achieved all that I have, without his unconditional willingness to help, unlimited patience and invariable belief in my abilities. Even as I come to the end of this journey, I feel wholly undeserving of such an incredible opportunity, and I will be forever grateful for your supervision and expertise.

Secondly, I wish to thank to Dr Liang Wu, who had the most arduous task of teaching me absolutely everything lab related! Having joined YSBL with no biochemical or structural experience, Liang taught me everything I needed to know with unbelievable patience and kindness. I cannot thank him enough for his invaluable expertise, experimental guidance and willingness to answer every question no matter how small or stupid. Indeed, most of the work presented here would not have been possible without his support.

I would also like to thank Jane and Kate who formed the incredible YSBL admin team for many years. I fondly remember Kate bringing me and all the new first year students into her office for a cup of tea and a chat during our first week in YSBL making us all feel at home. They were kind, cheerful and caring people who brightened the YSBL group and were always prepared to help with any problem.

Regarding scientific support, some specific thanks go to Dr Zach Armstrong, Dr Richard Meek and Dr Nicholas McGregor for their experimental guidance and willingness to help with any problem. A big thank you also goes to Sam Hart and Dr Johan Turkenburg for their countless hours spent collecting all the crystallography data presented here. Additionally, I would like to thank Chris Taylor from the Technology Facility for his expertise in mass spectrometry and Andrew Leech for his endless knowledge on variety of biochemical characterisation techniques. I must also acknowledge our amazing technicians, Louise, Simon and Jules, who work tirelessly behind the scenes to support our research and ensure labs run as efficiently as possible. I would also like to thank Dr Alison Parkin for her positivity and words of encouragement over the course of my PhD. Lastly, I would like to thank Dr Martin Fascione for his excellent scientific input during my TAP meetings.

On a more personal note, there are many other people I would like to acknowledge for their friendship and support. Of course, I want to thank all members of the GJD group, both past and present, who made the group such a wonderful research group to be a part of. I also want to acknowledge my amazing DnD team; Ben, Wendy, Alex, Zoe, Malcom and Tim K, who have provided many nights of true entertainment and escapism through questionable roleplay. I feel it is also necessary to give an extra special thank you to Dorothy, Alexandra and Saioa, who have become true and hopefully lifelong friends. They have on many occasions gone above and beyond to help me and I cannot thank them enough for their unconditional support and kindness.

Lastly, a huge thank you goes to my parents, Wayne and Amanda, and my sister Megan, who have always supported and encouraged my academic endeavours. Whilst I know you probably won't understand much in this thesis, you will still take the time to flick through it and be proud of what I have achieved all the same. I also want to acknowledge my grandmas; Grandma Daphne for her love, constant words of encouragement and her unwavering belief in me, and also my Grandma Cheryl for our many late-night telephone conversations in which we spent hours putting the world to right! Last but by no means least, my partner Zak; even though we have spent most of the time apart, your encouragement, love and companionship have been a constant and I am truly grateful to have shared this time with you.

Author's Declaration

All the work presented herein was performed by the author with the exception of collaborative work which is clearly stated in the text and outlined below. I declare that this thesis is a presentation of original work written solely by the author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Work conducted by others:

- All X-ray data were collected remotely at the Diamond Light Source Facility as organised by Sam Hart and Dr Johan Turkenburg.
- Synthesis of activity-based probes and inhibitors discussed in **Chapters 2** and **3** was conducted by researchers under the supervision of Prof. Hermen Overkleeft (Leiden University).
- Prof. Marta Artola (Leiden University) provided the *in-situ* cell data described in Figure 2.29 of **Chapter 2.**
- The co-complex crystal structure described in Figure 3.14 of **Chapter 3** was obtained by Dr Imogen Breen prior to the commencement of this work.
- SEC-MALLS analysis of recombinant human GBA discussed in **Chapter 4** was performed under the supervision of Andrew Leech, York Technology Facility
- Tryptic Digest MALDI-MS analysis discussed in **Chapter 4** was performed by Chris Taylor, York Technology Facility.
- The 2,4-DNP-2-Deoxy-2-fluoro-beta-D-glucopyranoside inhibitor discussed in **Chapter 4** was synthesised by Feng Lui under the supervision of Prof. Stephen Withers.
- Synthesis of the allylic carbasugars discussed in **Chapter 5** was performed by researchers under the supervision of Prof. Andrew Bennet and Prof. David Vocadlo (Simon Fraser University).
- Prof. David Vocadlo (Simons Fraser University) provided Figure 5.9 in **Chapter 5.**
- Dr Olga Moroz and Dr Lena Blagova performed the crystallisation of the GH3 β-xylosidase and GH10 β-xylanase described in Figure 6.8. Enzymes supplied by Novozymes.

Chapter 1: Introduction to Carbohydrate Active Enzymes and Lysosomal Storage Disorders

1.1 Abstract

Carbohydrates are one of the most abundant class of biomolecules on earth and are ubiquitous across all forms of life. The complexity and diversity of carbohydrates combined with their ability to form glycoconjugates by linkage to other biomolecules allows carbohydrates to fulfil a vast array of biological functions. The importance of carbohydrates is reflected in the range of carbohydrate active enzymes (CAZymes) which have evolved to synthesise, process and degrade glycans in nature. Unsurprisingly, defects in these CAZymes and carbohydrate processing pathways have been linked to a range of human diseases. Of note, lysosomal storage disorders (LSDs) are a group of inherited metabolic disorders which result from deficiencies in certain lysosomal glycoside hydrolases which are required for glycolipid catabolism. Subsequently, these diseases are primarily characterised by the cellular accumulation of glycolipids throughout the body, leading to multisystemic clinical symptoms which are not only severely debilitating but fatal if not treated effectively. Therefore, understanding the enzymes which underpin these diseases is crucial to the development of effective diagnostic and therapeutic strategies.

Firstly, this chapter will give an overview of carbohydrates and glycoconjugates, including the selective synthesis of glycoproteins and glycolipids, their degradation in the lysosomal system and their roles in human health and disease. Thereafter, this chapter will focus on the function of the lysosome and the roles of lysosomal glycoside hydrolases in the pathophysiology of LSDs. The most common LSDs: Gaucher disease and Fabry disease, will be discussed at length, focussing on the genetic and biochemical defects which underpin these diseases, as well as current diagnostic and therapeutic strategies.

1.2 Carbohydrates and Glycoconjugates

1.2.1 Overview

Broadly speaking, there are four major classes of biomolecules; nucleotides, amino acids, carbohydrates and lipids[1]. In contrast to nucleic acids and proteins, the "glyco-code" for the biosynthesis of carbohydrates lacks template-encoding and is disconnected from direct gene sequence control[2]. Instead, carbohydrate biosynthesis is achieved from monosaccharide building blocks, Figure 1.1, by the concerted action of a range of carbohydrate active enzymes (CAZymes); specifically glycosyltransferases (GTs) and glycoside hydrolases (GHs)[3]. In fact, a substantial part of the human genome (~3-4%) is dedicated to genes which encode for proteins involved in carbohydrate synthesis, modification, assembly, transport and degradation[4,5].

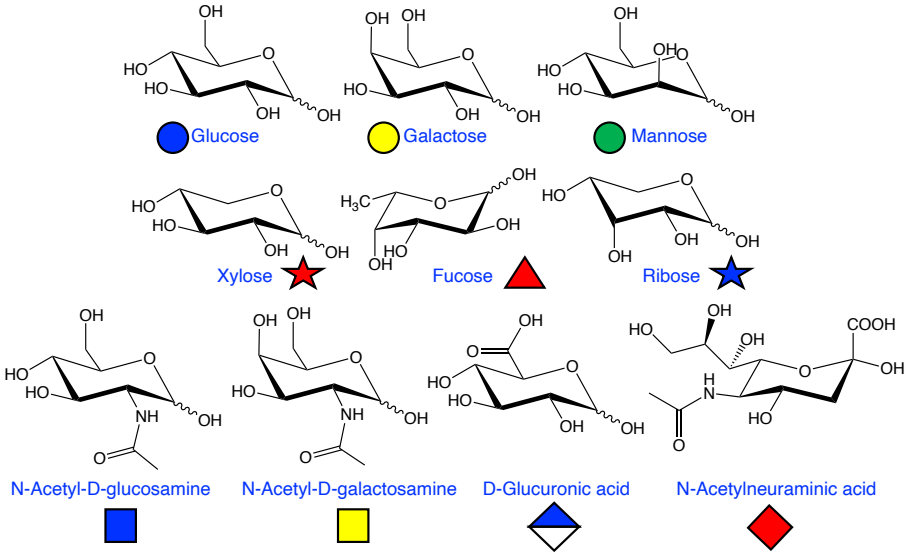


Figure 1.1: Conformational structures of the 10 common monosaccharides used to build the human glycome, where the C1 hydroxyl can adopt α or β configurations. The associated symbols are stylised using the Symbol Nomenclature for Glycans (SNFG)[6].

Owing to the presence of chemically equivalent hydroxyl groups, multiple chiral centres and α/β stereochemistry; monosaccharides can be connected together through various linkages to produce many isomers[3,4]. These isomers can be further diversified through branching and the introduction of substituents by hydroxyl group functionalisation[3]. In fact, it has been calculated that a hexasaccharide can theoretically form more than 1×10^{12} linear and branched isomers[7]; although, the diversity of glycan structures produced in reality is regulated by the enzymes and substrates available for glycan synthesis[8]. Conversely, nucleic acids and proteins are almost exclusively linear with a single type of linkage[1]. Therefore, the monosaccharide units required for carbohydrate biosynthesis surpass both nucleic acids

and amino acids in coding capacity and complexity[9]. This diversity, in combination with the ability to form glycoconjugates by linkage to other biomolecules, such as proteins (glycoproteins[10,11]) and lipids (glycolipids[12]), allows carbohydrates to fulfil many biological functions[10]. Consequently, the field of glycobiology focusses on elucidating the structure, chemistry, biosynthesis and biological functions of glycans and their glycoconjugates[13].

1.2.2 Importance of Carbohydrates and Glycoconjugates

Carbohydrates are ubiquitous across all forms of life, playing critical roles in a vast array of biological functions, such as cellular recognition[14], cell signalling[15,16], molecular trafficking[17], energy storage[18] and regulation of protein conformation, stability and function[1,19,20]. On the cell surface, glycans and glycoconjugates constitute the glycocalyx, a protective carbohydrate barrier which shields the plasma membrane and mediates inter- and intra-cellular interactions[19,21]. Additionally, these cell surface carbohydrates often act as ligands in biorecognition processes in host-pathogen interactions[22,23], inflammation[24] and metastasis[25,26]. However, one of the earliest clinical implications of cell surface glycans was the discovery that the A, B and H(O) blood group antigens are terminal neutral glycans, Figure 1.2, which control the ABO polymorphisms of the human blood groups[27].

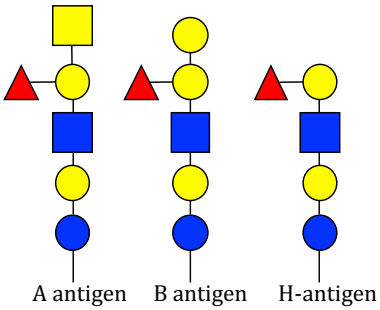


Figure 1.2: Structure of blood cell surface glycan antigens A, B and H.

Another important function of carbohydrates is to provide structural integrity through the formation of extensive carbohydrate polymers. For example, the polysaccharide cellulose, which consist of β(1→4) linked D-glucose units, underpins the structural integrity of plant cell walls[28], whilst chitin, a long-chain polysaccharide comprising of β(1→4) linked *N*-acetylglucosamine residues, provides structural integrity to the exoskeleton of arthropods[29]. On the cellular level, oligosaccharides also provide structural integrity to the cellular matrix and are required for the organisation of the cell membrane[19]. For instance, heparan sulfate, a linear sulphated polysaccharide composed of repeating glucuronic acid and D-glucosamine disaccharide units[30], provides structural organisation to the basement membrane and extracellular matrix of mammalian cells[31–33].

1.2.3 Protein Glycosylation

The glycosylation of proteins has the power to greatly amplify the capacity of the proteome by generating diverse glycoforms with different properties and functions[13,34]. Consequently, glycoproteins are present throughout nature, occurring in animals, plants, unicellular organisms and even viruses[11]. Indeed, the majority of eukaryotic cell surface and secreted proteins are co- or post-translationally glycosylated through the concerted action of substrate specific GTs and GHs[18]. In eukaryotic systems, protein glycosylation is achieved by the covalent attachment of carbohydrates through four mechanisms[11]; linkage to a nitrogen atom on an asparagine residue (*N*-glycosylation)[35], to an oxygen atom on a serine or threonine residue (*O*-glycosylation)[36], through a carbon atom on a tryptophan residue (*C*-mannosylation)[37] or through the formation of a covalent glycosylphosphatidylinositol (GPI) anchor[38]. In fact, the human glycome, which comprises the entire collection of sugars within the human body, is generated through 16 glycosylation pathways; 14 of which are dedicated to protein glycosylation[5]. These glycosylation pathways are dynamic, enzymatic processes which are broadly defined by the protein-sugar linkage (*O-, N-* or *C*-linked), the monosaccharide covalently linked to the protein and the initiating enzyme[39].

1.2.3.1 Protein-Glycan Synthesis

Approximately 173 unique GTs have been identified in the enzymatic synthesis of glycans across the various human protein glycosylation pathways[5]. This group of CAZymes are highly efficient glycosidic bond forming enzymes, which catalyse the transfer of monosaccharides from activated sugar donors to acceptor molecules such as proteins, lipids or other sugars[40]. Donor substrates are typically activated nucleotide sugars, including nucleotide uridine diphosphate (UDP) sugars, guanosine diphosphate (GDP) sugars and cytidine monophosphate (CMP) sugars, however, activated dolichol phosphate (Dol) sugars may also be used[41], Figure 1.3. Importantly, GTs show strong specificity for the sugar donor, the acceptor residue to which the sugar is to be transferred and the resulting glycosidic linkage, permitting selective glycosylation of proteins and lipids[40].

To date, 47 GTs have been implicated in the initiation steps of protein glycosylation in which a monosaccharide is transferred to a Asn/Ser/Thr or Trp residue of an accepting protein[5]. These initiation steps commonly occur in the lumen of the endoplasmic reticulum (ER); however, specific forms of glycosylation start in the early Golgi and cytoplasm, namely *O*-GalNAc and *O*-Xyl types of *O*-glycosylation[5]. In the case of *N*-glycosylated proteins, the initial

steps invariably occur in the ER, with the addition of a -$GlcNAc_2Man_9Glc_3$ glycan tree which is sequentially modified by various GHs and acts as an error-checking marker for the calnexin-calreticulin chaperoning pathway[42,43]. This pathway can temporarily retain glycoproteins and give them time to re-fold, therefore, the role of *N*-glycans in mediating entry to this pathway makes them vital in ensuring correct protein folding and stability[39]. Following initial glycosylation, a number of extension steps occur in the Golgi and *trans*-Golgi network (TGN) to produce core glycan structures which can be further expanded by elongation, branching and capping[5], Figure 1.4. The core extension steps tend to be pathway specific and unique to different types of protein glycosylation[13,39]. In contrast, the elongation, branching and capping steps may be non-specific and shared among different types of glycosylation[5,39]. Once the glycans have been constructed, side chain modifications may be made through sulfonation, phosphorylation and acetylation, which further control the behaviour, trafficking and function of the appended protein[5].

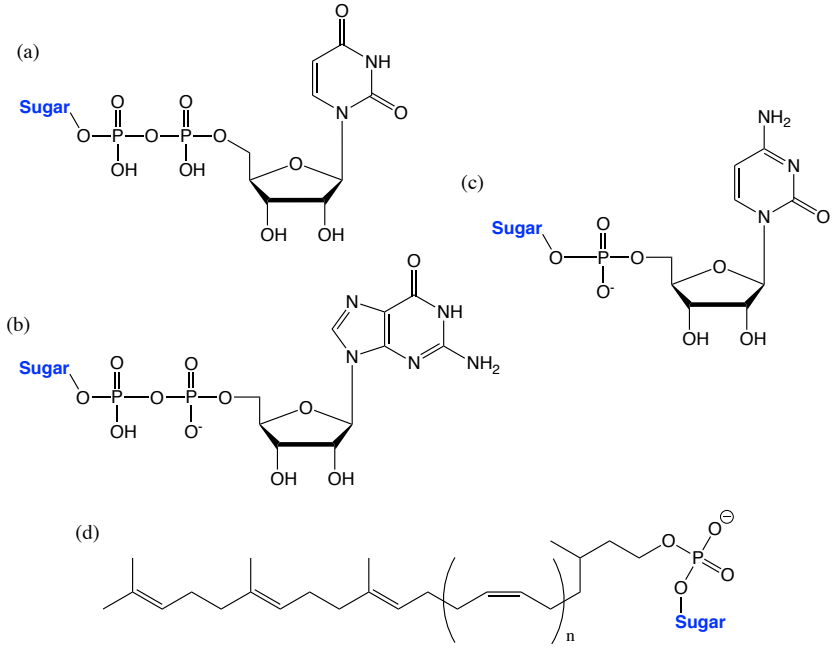


Figure 1.3: Chemical structure of activated sugar donors (a) uridine diphosphate (b) guanosine diphosphate (c) cytidine monophosphate and (d) dolichol phosphate sugar donor used by GTs for the synthesis of protein glycosylation.

The specificity of GTs and GHs required for glycan biosynthesis is defined by the glycan structures rather than the biomolecule to which they are attached[41]. Therefore, a single GT can glycosylate a range of seemingly unrelated proteins[17]. The great variability observed in mammalian glycan structures is primarily achieved by tissue-specific regulation of genes encoding enzymes involved in the glycosylation processes. However, glycosylation is also controlled by the availability of sugar donors, competition between GTs and the microenvironment of the growing glycan structure[5,41]. Moreover, changes in the expression

of molecular chaperones and altered trafficking of GTs between the ER and Golgi can result in drastic changes in the glycome of an organism[44,45].

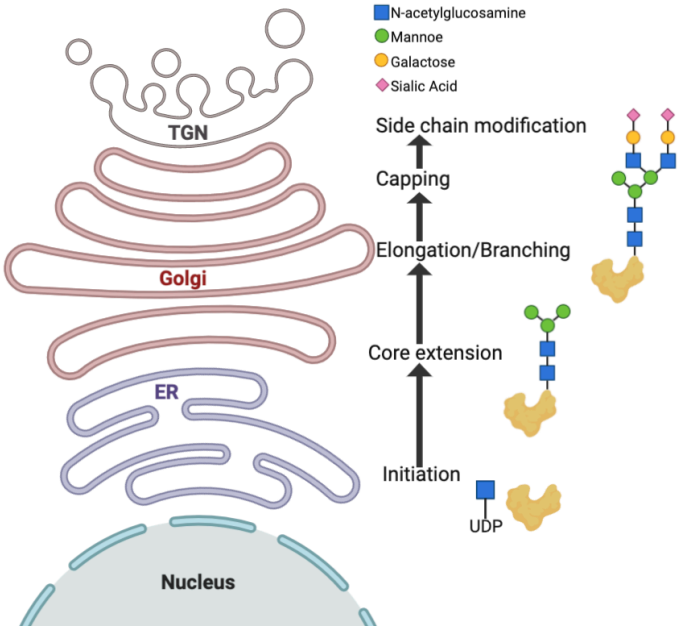


Figure 1.4: Subcellular localisation of protein glycosylation; the initiation steps typically occur in the endoplasmic reticulum (ER), followed by core extension, elongation, branching and capping in the Golgi and *trans*-Golgi network (TGN). Figure created in BioRender (www.app.biorender.com).

1.2.3.2 Importance of Protein Glycosylation

It is estimated that glycoproteins may comprise anywhere from 1-80% of carbohydrates by weight[11]. Given the large size and complexity of oligosaccharides, protein glycosylation is arguably the most complicated post-translational modification and the most significant factor in altering the properties of the proteins to which they are attached[10,11]. However, the exact impact of glycosylation on the biophysical properties of proteins depends on the location of the glycosylation sites and the surrounding environment[1]. Furthermore, intramolecular interactions between the peptide and glycan can affect the conformation of both the glycan and peptide, which in turn may control the activity of the protein[46]. Such interactions exemplify the complex structure-function relationships that exist between protein glycosylation and protein behaviour. Nevertheless, the importance of protein glycosylation is demonstrated by the fact many eukaryotic proteins require defined glycosylation patterns in order to achieve correct protein folding and activity[46]. In fact, 'incorrectly' glycosylated proteins are often identified by molecular chaperones and directed to the endoplasmic reticulum-associated degradation (ERAD) pathway to be removed by the proteasome[47]. Protein glycosylation has also been shown to protect proteins against proteolytic degradation and denaturation[11].

1.2.3.3 Protein Glycosylation and Disease

Many studies have linked defective protein glycosylation to a number of human diseases[17]. Of note, changes in cell glycan profiles have been identified in various forms of cancer[25,26], with aberrant glycosylation affecting tumour metastasis and invasive potential[48–50]. Specifically, it has been reported that the metastatic potential of many malignant tumours is directly related to the expression of the Sialyl Lewis X antigen (sLeX), a tetrasaccharide composed of sialic acid, fucose and *N*-acetyllactosamine residues[51]. Additionally, elevated levels of truncated *O*-glycans and high complex-type *N*-glycans have been associated with shorter survival rates of patients with some forms of cancer[48].

A range of metabolic diseases resulting from erroneous protein glycosylation have also been reported. For example, a large group of over 130 metabolic disorders, termed congenital disorders of glycosylation (CDG), result from inherited defects and deficiencies in proteins required for glycosylation[52–54]. In fact, CDGs resulting from defects in all four categories of protein glycosylation have been identified (*N-*, *O-*, *C-*, GPI anchor)[55]. These disorders are multisystemic in nature, with disease phenotypes ranging from mild to severely debilitating to lethal[52,55]. Symptoms most commonly include developmental delay, decreased muscle mass, neurological abnormalities and visceral disease affecting the liver, heart and skin[55].

In the early 1970s, the identification of inclusion-cell (I-cell) disease, another rare human genetic disorder, further demonstrated the importance of glycans in ensuring correct protein trafficking and activity[56]. I-cell disease is a slowly progressive metabolic disorder which develops in childhood and results in severe growth retardation, skeletal disease, and premature death in the first decade of life[56,57]. Initial investigations into I-cell disease led to the hypothesis that lysosomal degradative enzymes share a common recognition marker that facilitates their trafficking and uptake into the lysosome[58]. These enzymes were later found to possess a phosphomannosyl marker appended to their glycans, which mediates intra- and inter-cellular trafficking to the lysosome[59]. Following extensive research into the structures of the glycans involved[60], it was concluded that the biochemical defect in I-cell disease is faulty lysosomal trafficking due to inappropriate phosphorylation of the protein glycans[61]. This was arguably the first demonstration of a specific yet critical biological role of protein glycosylation in ensuring appropriate protein trafficking. The discovery of such degradative enzymes and their defective lysosomal trafficking subsequently led to the understanding of another group of rare inherited metabolic disorders called lysosomal storage disorders, which result from abnormal lysosomal function[62].

1.2.4 Glycosphingolipids

In addition to protein glycosylation, it is also important to consider the impact of glycans attached to other biomolecules, namely lipids. Of note, glycosphingolipids (GSLs), which are the most common subclass of glycolipids found in vertebrates[63], play important roles in almost every stage of the cellular life cycle from growth to differentiation to apoptosis[16,64–66]. Structurally, GSLs are amphipathic molecules consisting of a ceramide lipid chain with an *N*-acetylated sphingosine group attached to a monosaccharide core through a β-glycosidic linkage[67–69]. In humans, the monosaccharide core comprises a glucose or galactose moiety, which may be modified by the addition of other sugars to yield a diverse range of gangliosides, globosides and other series of GSLs[67,70].

Since their initial discovery in 1884, hundreds of GSLs have been identified, with huge heterogeneity observed in both the ceramide backbone and sugar head group[67,71]. Approximately 12 sugars are known to constitute the sugar component of GSLs and whilst any two residues can theoretically be linked, only certain sugar combinations are observed in vertebrates[63]. For example, in disaccharide GSLs, glucose is always linked galactose and fucose is invariably a terminal sugar[68]. The fatty acid chain of the ceramide component can also vary significantly, exhibiting diversity in both carbon chain length and saturation[72]. These variations in the ceramide moiety are thought to modulate the membrane associated functions of GSLs[71]. Owing to the incredible complexity and diversity of GSLs, the LIPID MAPS Structure Database (LMSD, https://www.lipidmaps.org/data/structure/) was developed to collate and record the structures of biologically relevant lipids and glycolipids in a consistent and accessible format[70].

1.2.4.1 Glycosphingolipid Synthesis

Similar to protein glycosylation, the assembly of glycans into GSLs is not template-based and is driven by the concerted action of GTs and GHs[12,63]. In fact, GSL synthesis is a stepwise process which begins in the ER with the condensation of a sphingoid base with an acyl-CoA to yield a ceramide product[72,73], Figure 1.5. This reaction is catalysed by a group of ceramide synthases which exhibit different specificities for the acyl-CoA substrate depending on the length of the acyl chain[74]. The resulting ceramide product may be galactosylated by galactosyltransferase in the lumen of ER to produce galactosylceramide (GalCer)[75] or it may be transferred to the *cis*-Golgi where it is glucosylated by glucosylceramide synthase (GCS) to form glucosylceramide (GlcCer)[76,77], Figure 1.6 (a,b). Alternatively, ceramide may be

transported to the *trans*-Golgi network (TGN) by the ceramide-transfer protein (CERT) for the synthesis of sphingomyelin[78,79], Figure 1.6 (c), which plays a structural role in the cell membrane of most eukaryotic cells and the myelin sheath that surrounds nerve axons[80].

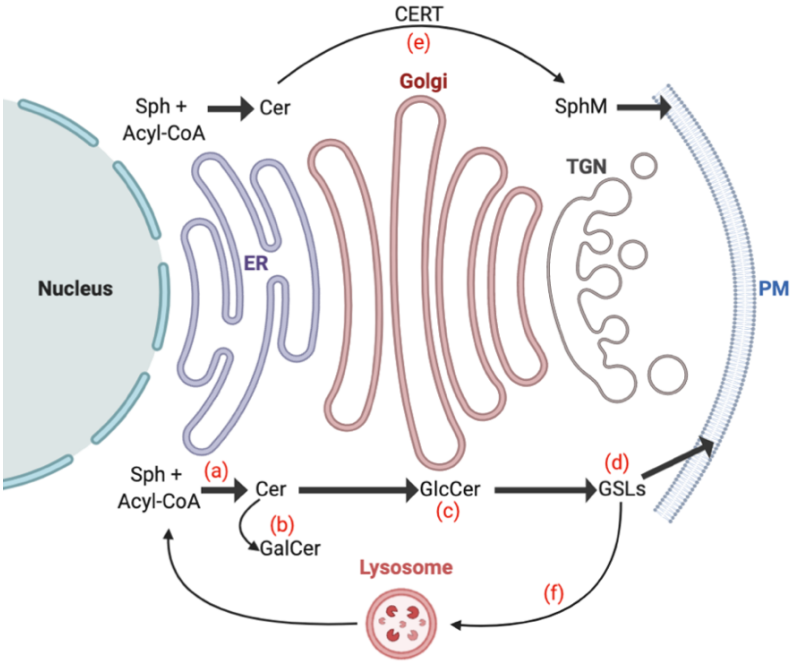


Figure 1.5: Process of glycosphingolipid (GSL) synthesis: (a) Sphingoid base (Sph) and acetyl-CoA react to form ceramide (Cer) which may be (b) galactosylated in the ER to form galactosylceramide (GalCer) or (c) glucosylated in the Golgi to produce glucosylceramide (GlcCer). (d) GlcCer can be glycosylated to form more complex GSLs which are transported to the plasma membrane (PM). (e) Cer may be transported to the *trans*-Golgi network (TGN) by the ceramide-transfer protein (CERT) to form sphingomyelin (SphM) which is exported to the PM. (f) GSLs are transported to the lysosome for degradation. Figure created in BioRender (www.app.biorender.com).

GlcCer and GalCer are the simplest of GSLs which serve as common precursors to all other GSLs that are synthesised in the Golgi complex[12], Figure 1.6 (a,b). In fact, GlcCer is the precursor to ~90% of mammalian GSLs, with GalCer accounting for the remaining 10%[81]. Once synthesised, GlcCer may be trafficked from the *cis*-Golgi to the luminal membrane leaflet of the Golgi and be further glycosylated by the transfer of monosaccharide units from activated sugar donors by the action of multiple GTs[63,82]. Specifically, GlcCer may be galactosylated to produce lactosylceramide (LacCer), which serves as the precursor to the GSL series found in vertebrates[81,83]. Additionally, LacCer can be modified by the addition of *N*-acetylneuraminic acid to yield monosialodihexosylganglioside (GM3), the simplest of gangliosides[63]. From the Golgi, GSLs are transported by membrane-bound transporters to the plasma membrane, where they may undergo further modification by GHs and GTs[84].

Alternatively, GSLs can be transported to the lysosome where they are degraded by various lysosomal GHs in a stepwise fashion to yield free monosaccharides and ceramide[69,85], Figure 1.5. The free ceramide may then serve in the resynthesis of GSLs in the ER or be converted to sphingosine by the enzyme acid ceramidase and further metabolised by sphingosine kinases to yield sphingosine-1-phosphate[69,86].

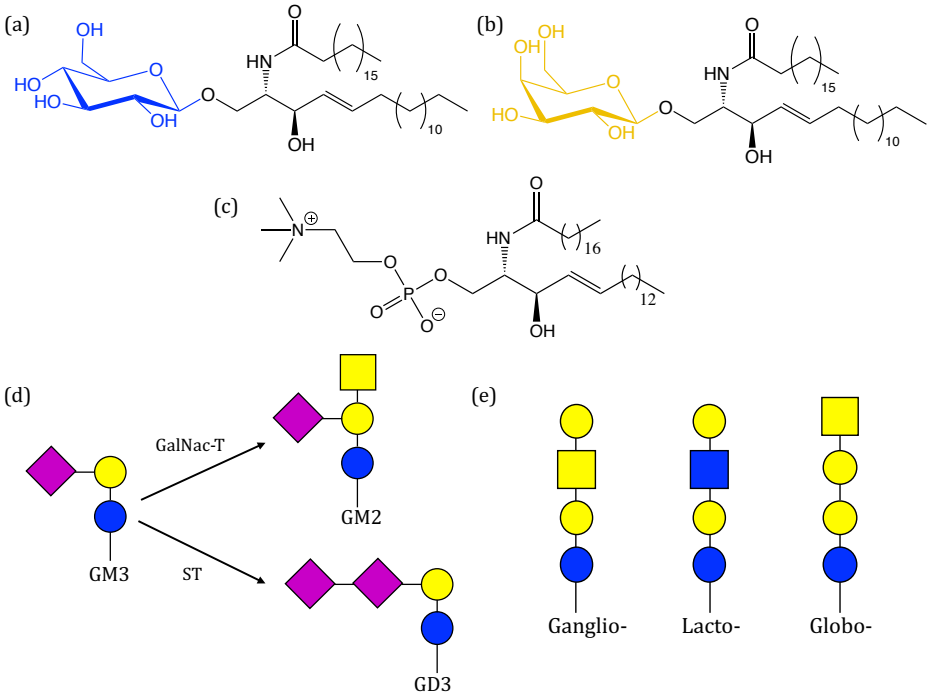


Figure 1.6: Chemical structure of (a) glucosylceramide (b) galactosylceramide and (c) sphingomyelin. (d) Competition between *N*-acetylgalactosaminyl-transferase (GalNac-T) and sialyltransferase (ST) for the synthesis of GM2 and GD3 gangliosides from GM3. (e) Sugar cores of the ganglio-, lacto- and globo-series of GSLs using SNFG[6].

The diversity and specificity achieved during GSL synthesis is primarily controlled by the expression of GTs and GHs but is also influenced by their subcellular localisation[82,83]. In fact, individual GTs are commonly localised to specific Golgi sub-compartments where they interact with one another to develop metabolic production lines for the synthesis of complex GSLs[83,87]. Of note lactosylceramide synthase interacts with both GM3 synthase and globotriaosylceramide (Gb3) synthase to transport LacCer to various Golgi sub-compartments for further glycosylation[63]. Competition between GTs is also important in controlling GSL synthesis. For example, the ganglioside GM3 may serve as a substrate for *N*-acetylgalactosaminyltransferase to yield GM2 ganglioside or as a substrate for sialyltransferase to form the ganglioside GD3[68], Figure 1.6 (d).

In vertebrates, there are three major series of GSLs; the ganglio-, globo- and neolacto-series, members of which share a common sugar core[68], Figure 1.6 (e). Moreover, these series of

GSLs are often expressed in tissue-specific patterns. Of note, the ganglio-series of GSLs are predominant in the brain, whereas the lacto-series are prominent in the secretory organs (glands)[68].

1.2.4.2 Glycosphingolipid Degradation

GSLs are transported to the lysosome for degradation through four routes: endocytosis, phagocytosis, autophagy or direct transport[84]. Once in the lysosome, GSLs are catabolised in a stepwise process catalysed by GHs which cleave sugar residues sequentially from the non-reducing end of the sugar core[85,88]. This stepwise cleavage ultimately yields free ceramide, sphingosine, fatty acids and monosaccharides which leave the lysosome by diffusion or with the aid of specialised transport systems[84]. These degradation products may be utilised in the resynthesis of biomolecules or further catabolised for energy[84,89].

A key difference between GSL metabolism and catabolism is that during GSL synthesis, the required substrates and enzymes are typically membrane bound, whereas during GSL degradation, the necessary lysosomal GHs are water soluble[83]. Therefore, the process of GSL degradation is typically facilitated by the action of 'activator' proteins called saposins[90,91]. For example, GlcCer is degraded into ceramide and glucose by β-glucocerebrosidase (GBA), which requires saposin C for effective *in vivo* function[92,93], Figure 1.7.

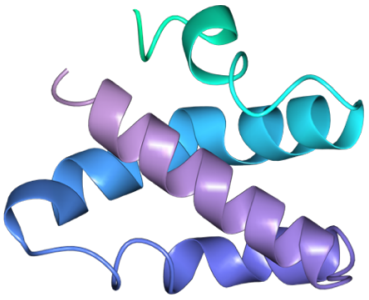


Figure 1.7: Crystal structure of human saposin C which is required for efficient *in vivo* function of human β-glucocerebrosidase. Structure consists of four α-helices folded into an ellipsoid. Figure generated in CCP4mg[94] using deposited PDB 2GTG coordinates[95].

The saposin proteins are commonly synthesised as larger precursors which upon transport to the lysosome are cleaved into four small (~80 amino acids) saposin proteins; A, B, C and D[88,91]. To date, there are two models proposed for the action of saposins; the 'solubiliser' model which suggests that saposins work by extracting the lipid portion of GSLs from the membrane, and the 'liftase' model which proposes that saposins bind directly to the membrane and improve lipid accessibility by distorting and remodelling the local environment of the cell membrane[96,97]. Whilst the exact interactions between lysosomal

GHs, GSLs and saposins remains unclear, it is understood that each saposin fulfils a specific function that cannot be compensated by another[98]. The importance of saposin proteins in facilitating efficient GSL catabolism is demonstrated by the fact that deficiencies in certain saposin proteins result in human disease. For example, a deficiency in saposin B leads to an atypical form of leukodystrophy[99], whilst a deficiency in saposin C results in an atypical form of Gaucher disease, the most common lysosomal storage disorder[92,100].

1.2.4.3 Importance of Glycosphingolipids

GSLs are primarily located in the outer leaflet of the plasma membrane of cells, forming dynamic lipid raft assemblies in which the lipid chains embed into the cell membrane whilst the polar sugar moieties protrude into the extracellular space[101,102]. The self-association of hundreds of lipids into these lipid rafts, which are typically 10-50 nm in diameter[68], is driven by their amphipathic properties provided by their long lipid chains[84,103]. In addition to providing structural organisation to cell membranes, these GSLs also constitute part of the glycocalyx which coats and protects the cell membrane[81]. Functionally, these cell surface GSLs act receptors and co-receptors for a variety of ligands including cytokines, hormones, growth factors, toxins and viruses[14,19,104]. Of note, certain GSLs have been shown to regulate the activity of a number of signalling receptors, such as the epidermal growth factor receptor (EGFR)[105], the insulin receptor[106] and the T-cell receptor[107]. Additionally, GSLs actively regulate many cellular processes, including cell proliferation, apoptosis, endocytosis, intracellular transport and inflammation[12,63], all of which can contribute to tumour formation and cancer progression. Indeed, abnormal production of gangliosides GD2 and GD3 has been shown to promote tumour growth and enhance the invasive potential of both small cell lung cancer and breast cancer[108,109], whilst aberrant Gb3 production has been linked to the metastatic potential of colon cancer[110].

1.2.4.4 Glycosphingolipids and Disease

Our understanding of the roles of GSLs in mammalian disease has largely derived from mouse genetics and the study of mutant mouse models[111–113]. For example, studies have shown that knockdown of GlcCer synthase in mice results in embryonic lethality due to widespread apoptosis[113], whilst ablation of GM2 synthase leads to male infertility, motor defect and Parkinsonism[114]. In fact, the rapid expansion of our knowledge of GSLs over the past 20-30 years has in part been driven by their pathogenic roles in diseases such as cancer[65,115], diabetes[116,117] and Alzheimer's[118,119] and their involvement in regulating the immune system for infection and inflammation[120,121]. Of note, numerous pathogens

including viruses, bacteria and toxins, directly interact with the carbohydrate moieties of GSLs to gain access to cells for infection[122]. A well-studied example is the human immunodeficiency virus (HIV) which binds to receptors on the host's cells via the viral envelope protein gp120[123]. Studies suggest that certain cell surface GSLs interact with HIV gp120 and facilitate infection by acting as a cell entry points for HIV[107].

In addition to their negative implications in facilitating pathogen cell entry, GSLs are also essential in regulating the innate and adaptive immune responses to infection[120,121]. Specifically, GSLs can transduce signals by binding to ligands, such as bacterial toxins and antibodies[124], to initiate a cascade of signalling events that regulate T-cell production[107,125] and activate natural killer T-cells[126]. This regulatory role in T-cell biology is important in controlling downstream processes implicated in allergic responses, autoimmune diseases and cancer[126]. However, the involvement of GSLs in lysosomal storage disorders is arguably their most distinguished pathological role in human health and disease[127,128].

1.3 Lysosomal Storage Disorders

1.3.1 Lysosome: Structure and Function

The lysosome, originally discovered in 1955 by Dr Christian de Duve[129], is a membrane bound degradative organelle found in all eukaryotic cells[130]. During investigations into the action of insulin, de Duve reported a biochemical cell fraction enriched in hydrolytic activities towards proteins and lipids[131]. Following biochemical, cytological and structural analyses, a membrane-bound compartment, now called the lysosome, was identified and found to be responsible for the degradation and recycling of various macromolecules and cellular components[89,129].

1.3.1.1 Lysosome Formation and Structure

Lysosomes are formed by the merging of post-Golgi traffic vesicles, containing the required hydrolytic enzymes, with endocytic vesicles generated at the plasma membrane, containing materials to be degraded[130,132], Figure 1.8. These cargo-filled vesicles merge to form early endosomes which undergo maturation to form late endosomes with associated changes in the protein-lipid coating[133]. Throughout this process, additional hydrolytic enzymes are delivered in vesicles to the maturing endosomes to eventually form lysosomes[89]. The maturation of endosomes to lysosomes is accompanied by acidification of the lumen, which is required to permit the uptake of soluble hydrolases from the M6P pathway[134,135].

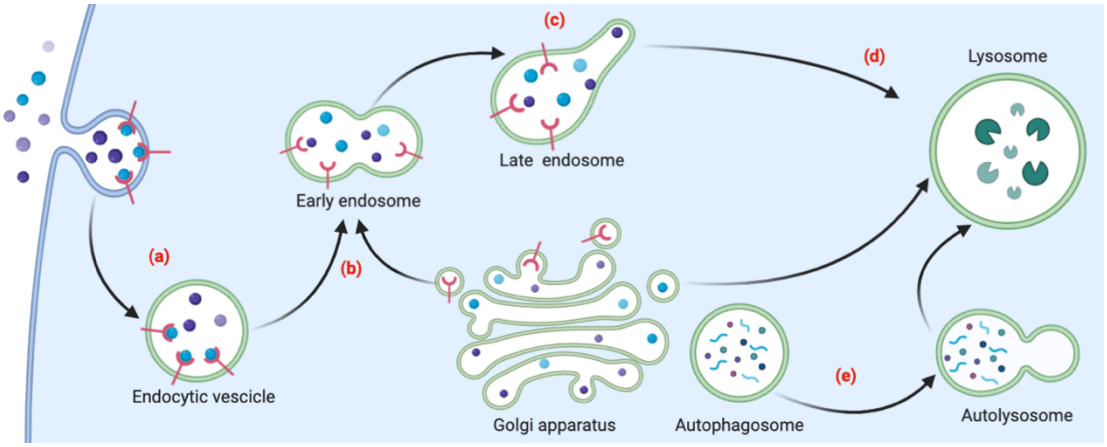


Figure 1.8: Simplified flow diagram of the endolysosomal system. (a) Extracellular material is endocytosed and trafficked to (b) early endosomes by fusion with post-Golgi traffic vesicles. (c) The material is retained for degradation and the early endosomes gradually mature to late endosomes which progress to (d) lysosomes with concomitant lumen acidification. (e) Intracellular materials may be directed for degradation by autophagy in which autophagosomes fuse with the lysosome. Figure created in BioRender (www.app.biorender.com).

The outer membrane of the lysosome is a single phospholipid bilayer which is decorated with a number of transmembrane proteins[130,134], the most notable being the lysosome-associated membrane protein 1 (LAMP-1) and 2 (LAMP-2) and the lysosomal integral membrane protein 2 (LIMP2)[89,132]. These transmembrane proteins are heavily glycosylated on the luminal side of the membrane, forming the glycocalyx which protects the membrane against digestion by the resident hydrolytic enzymes[134]. These enzymes require an acidic pH (pH 4.5-5.0)[135] to function, which is provided by an ATP driven proton pump called vacuolar H^+-ATPase[136].

1.3.1.2 Function of the Lysosome

Once formed, lysosomes exist in a dynamic state, fusing with the plasma membrane, endosomes and phagosomes to play key roles in cellular trafficking, phagocytosis, macroautophagy and cell-extracellular matrix communication[89,134]. Additionally, lysosomes can 'tether' to other organelles, such as mitochondria and the ER, using specialised tethering proteins to form microdomain structures which allow the transfer of lipids, metabolites and ions between organelles[134,136]. Consequently, the lysosome is considered the metabolic hub of the cell, influencing processes such as nutrient sensing, amino acid homeostasis, ion signalling and energy metabolism[89,134,136]. The primary and arguably most important function of the lysosome is the degradation of intra- and extra-cellular materials[129,136].

The lysosome is known to contain approximately 60 water soluble proteases, lipases and nucleases, which degrade biomolecules such as carbohydrates, proteins, nucleic acids and lipids[85,89,137]. The trafficking of degradative enzymes to the lysosome is typically mediated by mannose-6-phosphate (M6P) residues attached to their glycans, which allow the enzymes to bind to the M6P receptor for transport to the lysosome via the M6P pathway[138,139]. This involves packaging enzymes into vesicles with the aid of clathrin adaptor proteins (adaptins) to facilitate their transport to the lysosome through the endosomal system[139]. However, certain lysosomal enzymes, namely β-glucocerebrosidase (GBA), are targeted to the lysosome via M6P independent pathways, such as the LIMP2 pathway[138,140]. In contrast, materials to be degraded are transported to the lysosome via endocytosis and autophagy in membrane bound vesicles[89,134]. Following substrate degradation, lysosomes secrete their catabolised contents through lysosomal exocytosis, which is a Ca^{2+}-mediated process in which lysosomes fuse with the plasma membrane to release their contents into the extracellular matrix[141,142].

Over 1000 genes have been identified in regulating lysosomal function, with disease causing mutations linked to lysosomal GHs, lysosomal integral membrane proteins, lysosome-related organelles and lipofuscin production[62,143] (a lipid pigment which accumulates with age in the lysosome[144]). Changes in endocytic trafficking and lysosomal function as a result of such mutations have been implicated in a number of human diseases[145]. For example, some endolysosomal gene mutations are known risk factors for late-onset neurodegenerative disorders[146], including Alzheimer's, Parkinson's and Huntington's disease[118,119,145], whilst mutations in lysosomal GHs underpin lysosomal storage disorders.

1.3.2 Overview of Lysosomal Storage Disorders

Lysosomal storage disorders (LSDs) are a group of ~70 inherited metabolic disorders which result from defective lysosomal function[62,128]. Collectively, these diseases have an estimated frequency of 1:5,000-5,500 in the general population, however, individual LSDs are much more rare, with incidences ranging from 1:50,000 to 1:250,000 live births[62,147]. Unfortunately, these rates are considered underestimates because they assume all cases are accounted for and that accurate carrier frequencies are known[62,145]. In reality, many patients go undiagnosed and carriers often go undetected due to testing limitations and unreliable diagnostic methods[148,149]. Furthermore, specific populations are thought to be at much higher risk for certain LSDs. For example, compared to the general population, the Ashkenazi Jewish population has a much higher incidence of Gaucher disease (1:855).

Broadly speaking, LSDs are sphingolipidoses which are caused by recessively inherited mutations in genes encoding for hydrolytic lysosomal enzymes[98]. These mutations result in deficiencies in specific glycosidase activities, which leads to the toxic accumulation of GSL substrates within cells[127,150]. Of note, Gaucher disease (GD) and Fabry disease (FD) are caused by deficiencies in the lysosomal β-glucocerebrosidase (GBA) and α-galactosidase A (α-GAL) respectively[151–153]. In some LSDs, the primary defect lies in the function of other proteins, such as lysosomal membrane proteins, transporters or enzyme activators[62]. In these cases, disease typically results from the storage of GSLs due to defects in GSL trafficking rather than defective GSL catabolism[62]. For example, Niemann-Pick type C disease results from aberrant cholesterol transport from the lysosome[154,155].

1.3.2.1 Brief History of Lysosomal Storage Disorders

Although Gaucher disease (GD) was the first LSD to be described[156], Pompe disease was the first to be classified as a LSD in 1963[127,157], when De Duve and Hers determined that a deficiency in the enzyme acid α-glucosidase results in the toxic storage of glycogen in the lysosome[155]. This work was swiftly followed by electron microscopy studies which revealed the presence of "extremely overcrowded lysosomes" in the endothelial, smooth muscle and perivascular cells of patients with the LSD Fabry disease (FD)[158]. When placed in context with the work by De Duve and Hers, it was concluded that these overcrowded lysosomes were the result of substrate accumulation in the lysosomal system due disturbances in specific lysosomal enzymes[127]. In the same year, Dempsey and co-workers established that a sex-linked genetic mutation was responsible for these changes in lysosomal function in FD patients[159]. Therefore, it was determined that LSDs are hereditary metabolic disorders primarily caused by recessively inherited mutations in genes encoding for lysosomal enzymes, which leads to abnormal lysosomal function and storage of substrates within the lysosome.

1.3.2.2 Classification and Presentation of Lysosomal Storage Disorders

LSDs are primarily classified according to the GSL that accumulates, such as GM2 ganglioside in Tay-Sachs disease, GalCer in Krabbe disease, Gb3 in FD and GlcCer in GD[62,98], Table 1.1, however, demonstration of a specific enzyme deficiency is required for diagnosis.

Table 1.1: List of lysosomal storage disorders, the defective lysosomal enzymes and stored substrates which underpin their pathology[62,145,160]

Disease	Defective lysosomal Protein	Stored Substrate
Gaucher Disease	β-Glucocerebrosidase (GBA)	Glucosylceramide
Fabry Disease	α-Galactosidase A (α-GAL)	Globotriaosylceramide
Krabbe Disease	Galactocerebrosidase (GALC)	Galactosylceramide
Pompe Disease	α-Glucosidase (GAA)	Glycogen
Sandhoff Disease	β-Hexosaminidase A (HEXA) β-Hexosaminidase B (HEXB)	GM2 ganglioside
Tay-Sachs Disease (B-variant)	β-Hexosaminidase A (HEXA) β-Hexosaminidase S (HEXS)	GM2 ganglioside
Niemann Pick (Types A/B)	Acid sphingomyelinase (ASM)	Sphingomyelin
Niemann-Pick (Type C2)	NPC2 soluble cholesterol binding protein	Cholesterol
Niemann-Pick (Type C1)	NPC1 membrane protein	Cholesterol
Farber Disease	Lysosomal acid ceramidase (AC)	Sphingosine
Metachromatic Leukodystrophy	Arylsulphatase A (ASA)	Sulfogalactosylceramide

Although LSDs are genetically and clinically distinct disorders, they typically share a multisystemic nature, presenting with visceral, skeletal and neurological disease[62,145]. However, the age of onset for most LSDs varies considerably, from early infancy to adolescence to late adulthood. Moreover, a continuum of disease phenotypes are observed within each LSD, ranging from asymptomatic to slowly progressive disease over decades to death in utero[161–163]. Whilst loose associations can be made between the specific genetic mutation, the defective enzyme, the amount of residual enzymatic activity and the disease phenotype, providing an early diagnosis and reliable disease prognosis remains incredibly difficult[164,165]. As the most common LSDs, GD and FD have received the most attention.

1.4 Gaucher Disease

1.4.1 Overview and History

Gaucher disease (GD, OMIM (Online Mendelian Inheritance in Man®) 230800) is an autosomal recessively inherited LSD which was first described in the medical thesis of Philippe Gaucher in 1882[156]. His work described a young female who presented with hepatosplenomegaly and cachexia (weakness and wasting of the body). Initially, Philippe proposed a rare epithelioma of the spleen due to the presence of unusual swollen cells. However, a number of patients with similar pathological findings were later discovered, leading to the recognition of the systemic and familial nature of the disease by Nathan Brill

in the early 1900s[166]. Brill termed the disease 'Gaucher disease' and the unusual cells became known as 'Gaucher cells'[167]. GD was subsequently acknowledged as a multi-systemic chronic disease affecting the liver, spleen, bone marrow and lymph nodes[168]. In 1927, a similar multi-systemic disease with rapidly progressive neurodegeneration was reported[169], swiftly followed by another case with much slower neurodegeneration in Northern Sweden[170]. However, the biochemical nature of GD was not recognised until 1934, when Aghion discovered the accumulation of GlcCer within the Gaucher cells[171]. These cells were observed in bone marrow aspirate and other tissues with a distinct 'wrinkled tissue paper' appearance resulting from the accumulation of GlcCer in the lysosomal system[167,168]. However, the reason for GlcCer accumulation remained unknown.

In 1965, Brady and co-workers began studying the metabolism of GlcCer using ^{14}C radiolabelled precursors[172]. They published a series of papers in which the catabolism of GlcCer was outlined[172–174]. They first demonstrated that radiolabelled GlcCer was metabolised by an enzyme in the human spleen[172]. Secondly, this enzyme cleaved GlcCer to generate radiolabelled glucose and *N*-stearoylsphingosine (C18 ceramide)[172]. In further studies, the level of the GlcCer-cleaving enzyme was measured in human spleen tissue, revealing diminished activity in GD patients[173,174]. Subsequently, the biochemical basis of GD was established as an inherited deficiency in the GlcCer cleaving enzyme (β-glucocerebrosidase, GBA) which results in the accumulation of GlcCer within the cell[174]. This enzyme was later identified as a β-glucosidase which localises to ultracentrifuge fractions associated with the lysosome[175,176]. Consequently, GBA was classified as a lysosomal β-glucosidase and GD was acknowledged as a LSD.

1.4.1.1 Epidemiology

GD is the most common LSD, with an estimated incidence of 1:40,000-60,000 in the general population[177,178], however, this varies considerably across different populations. For example a much higher incidence of 1:855[179] is observed in the Ashkenazi Jewish population, with a carrier frequency of 1:16[180]. A higher frequency has also been observed in new-born screening studies, suggesting the frequency rate in the general population is a considerable underestimate[181]. However, it is difficult to precisely estimate the frequency of GD due to the presence of asymptomatic and late onset patients[182]. The only comprehensive population analysis was performed in Australia, yielding a frequency of 1:40,000 amongst a predominantly white population of European origin[147]. The frequency of GD in large populations of China, India, Indonesia and Africa is relatively unknown[178].

1.4.2 Clinical Manifestations

As a result of Brady's work in establishing the biochemical basis of GD[172–174], it was concluded that defects and deficiencies in GBA lead to reduced enzymatic activity and subsequent accumulation of GlcCer within the lysosome. This intracellular accumulation of GlcCer is believed to be responsible for the clinical manifestations of GD through a complex cascade of biochemical and cellular pathways[183,184]. Traditionally, GD is classified into three clinical phenotypes based on the presence of neurological manifestations and the rate of neuronopathic disease progression[185,186], Table 1.2.

1.4.2.1 Gaucher Disease Type 1

GD type 1 (GD1) is the most common form of GD, accounting for over 90% of GD cases[179]. GD1 is usually distinguished by an absence of neurological symptoms, however, the visceral manifestation of this phenotype vary considerably from essentially asymptomatic to severe visceral disease which is fatal in the first two decades of life[185,187]. Hepatosplenomegaly, liver failure, skeletal disease, anaemia, thrombocytopenia (low platelet levels) and malignancies are all associated with GD1[152,188]. However, skeletal symptoms such as bone pain, bone marrow failure, fractures and infections are often the most debilitating symptoms[183,189]. Furthermore, skeletal complications are often irreversible leading to long-term disabilities[168]. Consequently, skeletal disease has become the hallmark of GD1. Less common manifestations include pulmonary hypertension, interstitial lung disease and cardiac complications[190,191]. GD1 significantly impacts quality of life, however, it is rarely life threatening now that treatments such as enzyme replacement therapy are available.

1.4.2.2 Gaucher Disease Types 2 and 3

GD types 2 and 3 are the less common neuronopathic forms of GD, with an estimated incidence of <1 in 100,000[192]. Types 2 and 3 characteristically involve the central nervous system (CNS) , with a continuum of phenotypes ranging from death in utero to rapidly progressive neuronopathic disease to more slowly progressive CNS deterioration over a couple of decades[185].

Clinically, type 2 (GD2, OMIM 230900) is referred to as the 'acute-neuronopathic' form of GD which presents within the first few months of life[193]. GD2 typically accounts for < 5% of GD cases and is characterised by early neurological impairment with additional visceral manifestations[187]. Initially, hepatosplenomegaly is observed followed by delayed growth

and development[194]. Neurological symptoms such as oculomotor abnormalities follow swiftly, with premature death occurring in the first year or so of life due to neurological deterioration[194,195]. Consequently, GD2 is considered the most severe form of GD. However, an additional phenotype, termed perinatal lethal GD2 (OMIM 6080130), may also be distinguished as a distinct form of GD2. This is the rarest form of GD (<1%) but the most lethal, typically resulting in death in utero or shortly after birth[161]. The first case of perinatal lethal GD2 was believed to be reported by Drukker et al. (1970)[196], who described a Sephardic-Jewish infant who died 48 hrs after birth following an intracranial haemorrhage. This was followed by a number of cases across the world, reporting symptoms such as hepatosplenomegaly, thrombocytopenia, collodion skin (parchment-like membrane covering the skin[197,198]), pulmonary hypoplasia (incomplete development of the lungs), muscular atrophy, facial dysmorphism and neurological symptoms typical of GD2[199–202].

Type 3 (GD3, OMIM 231000), is the 'subacute-neuronopathic' form of GD[193] which typically accounts for ~5% of GD patients[203]. GD3 is prevalent in Egypt, East Asia and Europe[204] with a much higher incidence being reported in Norrbottnian patients in Sweden[205]. GD3 is characterised by similar visceral manifestations to GD1 but with additional neurological symptoms which present at a later age than in GD2[183]. Typically, GD3 manifests before the age of two, with slow progression and milder neurological involvement. Visceral and bone marrow manifestation usually develop first, with highly variable phenotypes[203]. Neurological symptoms tend to occur several years after the visceral manifestations[187], however, neurological symptoms can manifest before the age of 2[204]. Common symptoms include oculomotor abnormalities, seizures, ataxia (lack of voluntary coordination), progressive myoclonus epilepsy and dementia[193,204]. GD3 has been further divided into three sub-classifications; type 3a is a mild visceral disease with rapidly progressive neurological degeneration, type 3b involves severe visceral disease with mild neurological involvement and type 3c is characterised by mild visceral and neurological disease but with associated cardia complications[206,207]. In the case of GD3 individuals, premature death commonly occurs in the third to fourth decade of life[193].

Despite the fact GD is classified into three distinct clinical phenotypes, the boundaries between these phenotypes are often unclear, particularly in the early stages of GD3 when neurological symptoms are less obvious. This is further complicated by the fact that individuals with saposin C deficiencies also exhibit neurological symptoms comparable to GD3 patients[92,100,208].

Table 1.2: Summary of GD phenotypes. *Frequencies reported with regard to worldwide population

Clinical Features	GD1	GD2	GD3
Age of onset	Childhood-adulthood	Infancy	Childhood-adulthood
Life span	6-80 years	< 2 years	2-60 years
Frequency*	1:400-60:000	<1:100,000	<1:50,000
Hepatomegaly	Yes	Yes	Yes
Splenomegaly	Yes	Yes	Yes
Skeletal disease	Yes	No	Yes
Neurological disease	No	Yes (rapid progression)	Yes (slow progression)

1.4.3 Gaucher Disease and *GBA1*

In light of the biochemical basis of GD, the *GBA1* gene encoding human GBA was isolated and characterised from a variety of sources[209–214]. After some dispute, the *GBA1* gene was mapped to 1q21 on the long arm of chromosome 1, Figure 1.9. A glucocerebrosidase pseudogene (GBAP) was also identified 16 kb downstream of the functioning *GBA1* gene[215,216]. This pseudogene shares 96% sequence identity with the active gene[215,217], however, it cannot be translated to yield active GBA due to small deletions in several introns and the loss of some exons during mRNA processing[216]. Although inactive, understanding the pseudogene structure has proved important to the analysis of the functional *GBA1* gene.

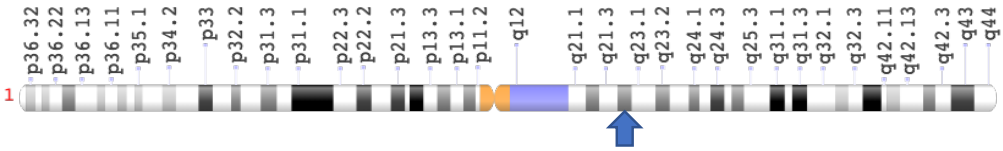


Figure 1.9: Location of *GBA1* gene at Xq21 on chromosome 1 indicated by the blue arrow. Centromere highlighted in orange and variable region highlighted in purple. Figure prepared using NCBI genome decoration tool with GRCh38.p12 (Genome Reference Consortium Human Build) representation (www.ncbi.nlm.nih.gov/genome/tools/gdp).

GD is caused by a range of mutations in the *GBA1* gene, and according to The Human Gene Mutation Database[218] (www.hgmd.org, Institute of Medical Genetics in Cardiff) over 500 genetic mutations at the *GBA1* locus have been identified. Missense mutations are the most prevalent, accounting for 75% of GD causing mutations. However, frame-shift, splicing, insertion and deletion mutations have also been described[219,220,194]. Additionally, mutations resulting from gene conversion between the functional gene and the pseudogene have been identified[221]. The close proximity of the pseudogene to the functional gene and the high level

of sequence homology is believed to result in portions of the pseudogene being non-reciprocally integrated into the *GBA1* gene sequence[217]. This is supported by the identification of GD alleles containing mutations identical to those found in the pseudogene[185]. In fact, one of the most common mutations, the L444P mutation, arises from apparent gene conversion with the pseudogene[185].

Mutations L444P, N370S, D409H, 84GG and IVS2+1 are the most prevalent mutations[222], with N370S and 84CG mutations accounting for 80-89% of disease producing alleles in the Ashkenazi Jewish population[223]. The N370S mutation is also the most common mutation amongst non-Jewish patients, accounting for 30% of disease causing alleles[224]. The N370S mutation is a missense mutation which results in the substitution of an asparagine for serine[225]. Although stable GBA can be expressed with this mutation, its catalytic activity is drastically reduced[226]. The L444P mutation is also a missense mutation that results in the substitution of a leucine for a proline which generates a new restriction site for the NciI restriction enzyme[227]. In contrast, the 84GG mutant is an insertion mutation where an additional glycine residue is added[228,229]. The frameshift which results from this nucleotide insertion in the cDNA causes a shift in protein translation, resulting in a stop codon 18 amino acid residues downstream of the insertion site[229]. This in turn results in premature termination of the protein[222,229]. Consequently, the 84GG mutations is known as a 'null' mutation which does not produce any mature GBA. The IVS2+1 mutant is a splice mutation in which the 5' donor splice site is destroyed. In simplified terms, this results in the synthesis of abnormal mRNAs which prevent the production of active GBA[230]. Lastly, the D409H mutation is a missense mutation resulting in the substitution of an aspartic acid for a histidine[224]. This mutation exists in the pseudogene and usually arises from gene conversion or unequal homologous recombination with the functional gene[222]. Of note, this mutation is associated with the Norrbottnian type GD reported in Sweden[205].

1.4.3.1 Genotype-Phenotype Relationships

Despite the characterisation of the *GBA1* gene and identification of over 500 genetic mutations, no definitive genotype-phenotype relationships have been established. The extreme variations in disease manifestations, even amongst siblings with the same genetic mutations, has complicated the analysis of genotype-phenotype relationships[231,232]. Additionally, the rarity of many GD mutations has limited investigations to the major mutant alleles, N370S, L444P and D409H[185]. Following extensive research, some degree of genotype-phenotype relationship, albeit vague, has been established for these mutants[232].

The presence of the N370S mutant, in the homoallelic or heteroallelic state, appears to preclude the development of neurological disease and commonly correlates with non-neuronopathic GD1[225,222]. In fact, homozygotes for the N370S mutation are often asymptomatic and may go undiagnosed for most of their life[221,233]. This general association has been attributed to the fact that this mutation yields some active GBA (~10-20% of the normal range[226]) which is sufficient to prevent CNS involvement[152]. In contrast, GD patients with the N370S mutation in a heteroallelic state (with another mutant allele), such as N370S/L444P, experience early onset of more severe GD manifestations[231]. Individuals carrying homozygous L444P or D409H mutations appear to suffer from neuronopathic GD[232]. Specifically, the homozygous L444P mutation is commonly associated with GD3, leading to severe neurological and visceral manifestations[227]. The L444P mutation has also been observed in GD1 and GD2 but in the heteroallelic state[234]. In addition to neurological manifestations, patients homozygous for the D409H mutation also exhibit cardiovascular complications but with milder visceral disease[206]. Lastly, the null 84GG and IVS2+1 mutations are associated with severe phenotypes, even in heteroallelic states[222], with most homozygotes for these rare mutations dying in the perinatal period[235]. These severe phenotypes are thought to be a consequence of such null mutations completely abrogating GBA activity[229].

Although our understanding of GD genotypes and phenotypes has grown, there is still considerable phenotypic variability amongst the same genotypes[149,232]. Whilst there are some appropriate generalisations between the two, there are many exceptions which have hindered further characterisation of genotype-phenotype relationships. Therefore, it has become apparent that other non-genetic factors play a role in determining disease phenotype and severity. It is likely that the genotype determines the residual activity of endogenous GBA which sets the boundaries for potential phenotypic variation in response to other influences[185]. Consequently, the use of genotypes to diagnose patients and establish a prognosis is difficult and unreliable.

1.4.4 Gaucher Disease and Parkinsonism

Parkinson's Disease (PD) is the second most common neurodegenerative disease and is pathologically characterised by an accumulation of α-synuclein (a presynaptic neuronal protein) in the CNS[236,237]. This α-synuclein accumulation leads to abnormal motor symptoms such as tremors and rigidity, as well as non-motor symptoms including cognitive decline, sleep disorders and depression[236]. In 2009, analysis of 5691 PD patients confirmed

an association between *GBA1* mutations and PD[238]. In fact, GD patients are reported to have a 26-fold higher life-time risk of developing PD compared to the general population[239,240]. Moreover, mutations in the *GBA1* gene occur in a considerable number of all patients with sporadic PD, varying between 8-12% across the world's population[241].

Several publications have hypothesised a link between reduced GBA activity and α-synuclein accumulation following the discovery that lysosomal dysfunction and depleted GBA activity promotes propagation of α-synuclein aggregates[242–245]. Specifically, in the loss-of-function hypothesis, the loss of GBA activity is thought to compromise lysosomal α-synuclein degradation, resulting in an accumulation of lipid substrates which perturbs the clearance of α-synuclein aggregates[237,246]. Additionally, in the gain-of-function hypothesis, accumulated GlcCer and misfolded GBA are thought to directly interact with α-synuclein and promote amyloid formation by stabilising soluble α-synuclein oligomers which aggregate and form Lewy bodies in the nerve cells[237,244]. Lastly, in the third bidirectional loop hypothesis, α-synuclein aggregates are proposed to have an inhibitory effect on GBA[237,247] which feeds into a viscous cycle of further α-synuclein aggregation, Figure 1.10.

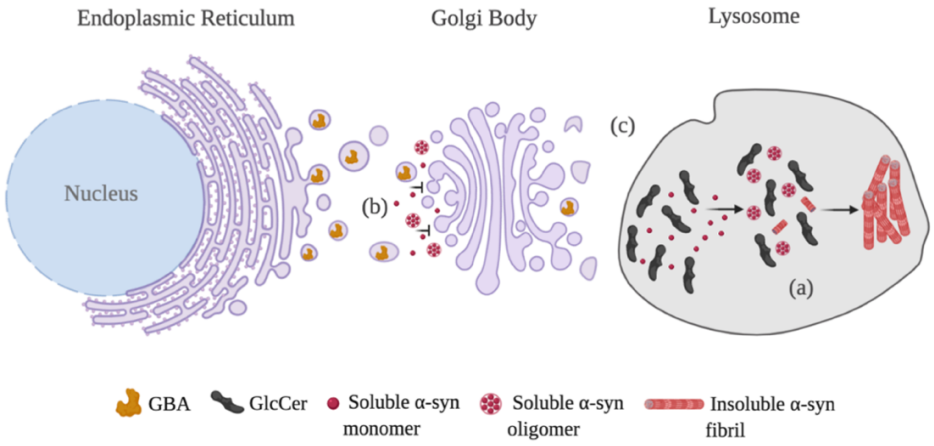


Figure 1.10: Potential pathogenic mechanism between GBA and α-synuclein accumulation in PD (a) Lysosomal accumulation of GlcCer stabilizes soluble α-synuclein oligomers which are processed to amyloid fibrils. (b) Accumulation of soluble α-synuclein monomers and oligomers blocks the ER-Golgi trafficking of GBA. (c) Reduced transport of GBA to lysosome amplifies GlcCer accumulation and the stabilisation of soluble α-synuclein oligomers, resulting in stronger inhibition of GBA trafficking with each pathogenic cycle. Figure created in BioRender (www.app.biorender.com).

Despite the unequivocal correlation between *GBA1* mutations and the risk of developing PD, many GD patients do not suffer from PD, therefore, a reduction in GBA activity alone is not sufficient to account for the association[241]. Furthermore, the concept that accumulating GlcCer substrate may play a role in PD was brought into question in 2015, when Gegg et

al.[248] reported no evidence of substrate accumulation in the brain of PD patients with *GBA1* mutations. Following years of research, the exact link between *GBA1* mutations and PD remains unclear, but mutations in the *GBA1* gene are considered the most common genetic risk factors for PD[245], specifically N370S, L44P, 84GG, IVS2+1 and V394L mutations[249,250].

1.4.5 Gaucher Disease and Cancer

An increased risk of developing various forms of cancer, specifically melanoma, non-Hodgkin lymphoma, pancreatic and hepatocellular cancer, has been associated with GD[251,252]. However, the pathophysiology of cancer development in GD is poorly understood. The most common hypothesis is that the accumulation of GlcCer alters cellular function, resulting in abnormal cytokine and chemokine production and the formation of tumours[253,254]. The second hypothesis proposes that tumour formation is facilitated by perturbed sphingolipid metabolism in pre-existing cancerous cells, resulting unfavourable changes in the proliferative and anti-proliferative balance[255]. Regardless of the pathogenic mechanism, the strongly increased risk of developing forms of cancer further exemplifies the need for early detection and treatment of GD.

1.4.6 Gaucher Disease Pathogenesis

Initially, there were two main hypotheses regarding the pathogenesis of GD. The first hypothesis, termed the "constipated lysosome hypothesis", presumed a passive role for the accumulation of GSLs in the lysosome and considered Gaucher cells as inert lesions[185]. The second hypothesis, called the "transduction box hypothesis", proposed that a deficiency in an essential signal, which would normally leave the lysosome following GlcCer hydrolysis, effects cellular functions[185]. This hypothesis assumes a normal active role for GlcCer hydrolysis products rather than a toxic effect of accumulated substrate[185]. Both hypotheses lacked sufficient *in vivo* evidence, likely resulting from the shortage of viable animal models. However, over the past few decades, advances in clinical and animal model studies have unambiguously demonstrated a role for immune activation in GD pathophysiology[253,256].

1.4.6.1 Involvement of the Immune System

In 2002, Zhao and Grabowski put forward their own hypothesis for the pathophysiology of GD following the observation that defective GBA activity alters the release of cytokines, proteases and antigens from activated macrophages[185,253]. Zhao and Grabowski reasoned that normal activation of the macrophage system, by some form of lipid component, leads

to a cascade of transcriptome and proteome effects which are disrupted in GD due to the lack of GlcCer hydrolysis[185]. Overall their hypothesis includes the presentation of GlcCer to the lysosome, the accumulation of GlcCer due to deficient GBA activity and subsequent lack of lipid effector egress from the lysosome leading to a series of biochemical events involving altered expression of macrophage activation markers[185]. Specifically, the lack of ceramide release from the lysosome is believed to increase the production of cytokines through a loss of feedback signal mechanism[185, 257]. Therefore, the accumulation of GlcCer may be viewed as a secondary event; the primary cause could be the lack of lipid effector egress which induces swelling of the macrophages and altered expression of activation markers[185]. In fact, elevated cytokines are observed in Gaucher cells and the plasma of GD patients, with a trend between increasing plasma cytokine levels and disease severity being identified[185]. Consequently, Gaucher cells are not inert storage containers as originally proposed in the constipated lysosome hypothesis. Instead Gaucher cells are metabolically active cells which are surrounded by pro-inflammatory macrophages[258,259].

More recently, Pandey et al. (2017)[260] identified a role for the complement system, specifically the activation of complement C5a, in controlling the accumulation of GlcCer and the subsequent inflammation in GD. The complement system is a component of the immune system which is vital to the detection and destruction of invading pathogens[261]. Complement component C5 is the fifth component of the complement which plays an important role in inflammation and apoptosis. Specifically, C5a is a small activator peptide of the anaphylatoxin family, which is cleaved from C5 by cell derived proteases[261,262]. Once released, C5a exerts a proinflammatory response by binding to its two receptors, C5aR1 and C5aR2, located on immune cells which upregulates the expression of co-stimulatory molecules[262,263]. Pandey et al. observed local and systemic C5a complement activation in GBA-deficient mice and mice treated with GBA inhibitors[260]. This was associated with GlcCer storage, tissue inflammation and proinflammatory cytokine production. In fact, a GBA deficiency in both mice and humans was linked to the production of GlcCer-specific IgG autoantibodies which induce complement activation and the generation of C5a[260]. Not only does this result in the production of proinflammatory cytokines and chemokines, it's also reported that binding of C5a to C5aR1 controls UDP-glucose ceramide glucosyltransferase production, disrupting the balance between GlcCer formation and degradation[260]. This feeds into a cycle of cellular GlcCer accumulation, complement activation, C5a production and innate and adaptive immune cell response. This work exemplifies the complex involvement of the immune system in GD pathogenesis.

1.4.6.2 *Importance of Glucosylsphingosine*

In addition to GlcCer, another glycosphingolipid called glucosylsphingosine (GlcSph) accumulates in GD patients[264]. In the absence of GBA activity, GlcCer can enter an alternative pathway in which a ceramidase converts it to GlcSph by de-acteylation[265], Figure 1.11. Analysis of tissues from healthy individuals and GD patients has demonstrated elevated splenic and hepatic levels of GlcSph in patients with all types of GD[266]. Interestingly, elevated GlcSph levels were also observed in the brain of patients suffering from neuronopathic GD[267,266], suggesting that GlcSph may contribute to the neurological manifestations associated with GD2 and GD3[268,265]. Moreover, GlcSph may be further metabolised in the cytoplasm by non-lysosomal GBA (GBA2) to generate sphingosine and sphingosine-1-phosphate[269], Figure 1.11. Sphingosine is particularly toxic to the bone and has been identified as a molecular mediator of GD1 pathophysiology[269]. In fact, deletion of the GBA2 gene prevents formation of sphingosine and appears to improve the clinical phenotype of GD1 in knockout mice[269]. Specifically, improvements in visceral, haematological and skeletal symptoms have been reported, demonstrating a pathological role for sphingosine[269]. A potential pathogenic role for S1P in LSDs has also recently surfaced, with elevated S1P levels being linked to the cardiac pathology of FD[270] and the neuronal pathology of Sandhoff disease[271]. However, further research is required to fully understand the roles of the multiple GlcCer metabolic pathways in GD pathophysiology.

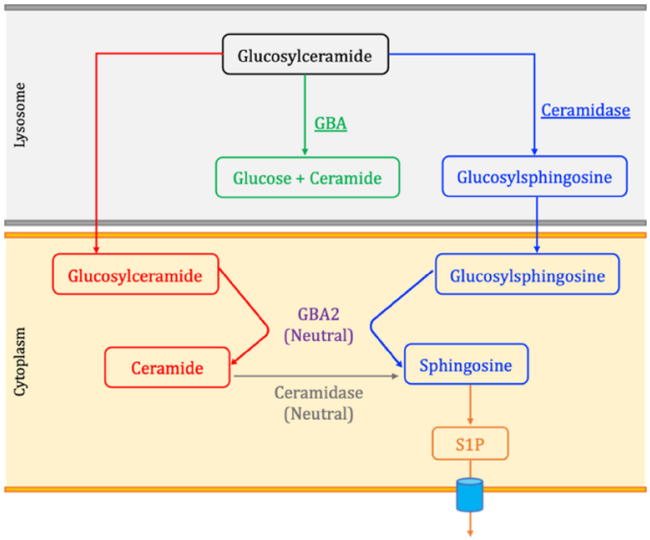


Figure 1.11: Metabolic pathways of glucosylceramide (GlcCer). In the presence of GBA activity, GlcCer is hydrolysed to glucose and ceramide (green pathway). In the absence of GBA activity, a ceramidase converts GlcCer to glucosylsphingosine (GlcSph) which may pass to the cytoplasm and be converted to sphingosine and sphingosine-1-phosphate (S1P) by cytoplasmic GBA2 (non-lysosomal GBA) (blue pathway). GlcCer may also be degraded to ceramide by GBA2 (red pathway) which feeds into the production of S1P.

1.4.7 Diagnosis of Gaucher Disease

Unfortunately, diagnosis of GD typically occurs several years after the onset of symptoms due to the non-specific nature of clinical manifestations and the rarity of the disease.

1.4.7.1 Histological Diagnosis

Traditionally, GD was diagnosed by identification of lipid engorged Gaucher cells in bone marrow aspirate or other tissue biopsies[223]. However, Gaucher-like cells have been described in other disorders, such as multiple myeloma, Hodgkin disease and lymphoma[191,258]. Therefore, a diagnosis based on the presence of Gaucher cells alone was not very reliable and great caution in the interpretation of lipid-laden macrophages was required. This approach is also invasive and is no longer used for the diagnosis of GD[203].

1.4.7.2 Enzymatic Diagnosis

The identification of deficient GBA activity in GD patients permitted the development of more reliable diagnostic techniques. In particular, the Brady group developed an enzymatic method in which the activity of GBA in blood leukocytes is quantified through a fluorometric assay using the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside[272]. Demonstrating reduced GBA activity through this approach is considered diagnostic, with the majority of GD patients exhibiting GBA activity that is 0-15% of the mean normal activity[223]. Such activity assays can be performed on a variety of samples, namely blood plasma, skin fibroblasts, amniotic fluid cells and chorionicvilli[191,273], and they are less invasive than traditional histological methods. Indeed, this approach has become the gold standard for GD diagnosis.

Despite its success, there are a number of limitations to the enzymatic assay which must be considered; firstly, there is no correlation between the level of GBA activity and disease severity, therefore, little prognostic information can be deduced[274]. Secondly, it is estimated that most GD carriers and approximately 10-33% of GD heterozygotes exhibit GBA activity levels which overlap with the normal range[168,275]. Consequently, this approach is not wholly reliable for the diagnosis of heterozygotes and cannot be used for the detection of GD carriers. Thirdly, the activity of GBA is labile, so blood samples need to be tested within 24hrs of collection for optimal results[223]. Another factor which typically complicates *in vitro* assays for GBA, is the presence of functionally related β-glucosidases, namely GBA2 (GH116) and GBA3 (GH1), which also cleave the fluorogenic substrate. Consequently,

subtractive assays using inhibitors to inactivate related glucosidases are often required[276–278]. This involves additional sample handling and manipulation which may compromise the accuracy of the measurements. Such subtractive assays are now commonplace in GBA research and exemplify the need for higher specificity GBA substrates.

1.4.7.3 Genetic Diagnosis

Advances in DNA analysis and improvements in our understanding of the *GBA1* gene have permitted the development of genetic diagnosis. A number of polymerase chain reaction (PCR) based techniques have been developed to facilitate the detection of known *GBA1* mutations[223]. In such techniques, the gene fragment of interest is targeted with primers to amplify the structural gene; mutations can then be identified by restriction digestion[227], an amplification refractory mutation system[279,280], allele-specific oligonucleotide hybridization[281] or by mismatched PCR[282].

DNA analysis holds a number of advantages over enzyme activity assays; firstly, the samples required for DNA analysis are highly stable and can be transported at ambient temperature without time constraints[223]. Secondly, it permits the detection of GD carriers and certain GD heterozygotes who cannot be diagnosed through enzyme activity assays[168]. In fact, prenatal diagnoses can also be performed as early as 10-12 weeks by genetic analysis of amniotic fluid cells[283]. Additionally, mutation analysis has some ability to predict disease prognosis based on known genotype-phenotype relationships[223]. However, due to the extreme phenotypic variability observed within a single genotype, caution should be taken when predicting disease prognosis. Whilst DNA analysis is highly reliable for the diagnosis of Jewish patients, due to the fact that four major mutations (N370S, L444P, 84GG and IVS2+1) account for over 90% of the mutant alleles in the Jewish population[168], it is considered much less conclusive for non-Jewish patients in which a wide range of mutant alleles are observed[168]. Perhaps the biggest drawback of DNA analysis is that its success is limited by our knowledge of *GBA1* mutations. Although many GD alleles have been identified[284], it is likely that many unknown mutants exist, therefore, a negative result in DNA analysis does not guarantee the absence of GD alleles.

1.4.7.4 Use of Biomarkers

A range of macrophage specific markers are known to be elevated in the blood plasma of GD patients and these markers may be used to support the diagnosis of GD. Of note, chitotriosidase (CHIT1) has been used as a biomarker for GD since 1994[285]. CHIT1 is a

chitinase which is secreted by GlcCer-laden macrophages and it's activity is significantly elevated in GD individuals, reportedly up to 1000-fold[285,286]. Subsequently, CHIT1 is considered one of the most specific and sensitive bio-markers of GD[285]. Although CHIT1 levels in blood plasma does not necessarily correlate with specific clinical manifestations, it does reflect the total body burden of Gaucher cells and generally correlates well with the severity of hepatomegaly[287]. Subsequently, CHIT1 activity measurements provide a sensitive and reliable method for evaluating disease progression and treatment efficacy, with a reduction in CHIT1 activity indicating a positive therapeutic effect[288]. However, raised CHIT1 activity alone is not diagnostic of GD because CHIT1 may be elevated in other pathological conditions including Niemann-Pick disease, multiple sclerosis and Alzheimer's disease[187,288]. Additionally, CHIT1 levels vary significantly among patients, with a third of individuals exhibiting low levels which are difficult to interpret[288]. In fact, it is estimated that 6% of individuals have no detectable CHIT1 activity due to a null mutation in the cognate gene[289]. Such limitations have hindered its use in inter-patient comparisons.

In cases where GD is suspected and CHIT1 levels are normal or undetectable, other markers such as the angiotensin converting enzyme (ACE) or tartrate resistant isoenzyme (TRAP) may be used[203]. More recently, chemokine ligand 18 (CCL18), a small cytokine belonging to the CC cytokine family, was found to be a novel biomarker of GD[290]. CCL18 is secreted by various cell types, such as macrophages and dendritic cells, and promotes the recruitment of T lymphocytes[291]. Gaucher cells also produce CCL18, with a 20-50-fold increase in CCL18 levels being observed in GD patients[290,292]. Although the elevation of plasma CCL18 is typically less than that of CHIT1, its levels are more consistent across GD patients, allowing for inter-patient comparisons to be made[292]. However, elevated CCL18 plasma levels have been observed in other chronic inflammatory diseases[293], therefore, CCL18 levels are commonly analysed alongside CHIT1 to support of GD diagnosis.

Following the discovery of elevated GlcSph levels in the spleen, liver and brain of GD patients[267], GlcSph has also been proposed as a biomarker for GD. A number of studies have indicated that a moderate relationship between GlcSph levels and clinical symptoms exists[294]. Additionally, correlations between GlcSph levels, CHIT1 activity and CCL18 levels have been identified[266,267,295]. Indeed, GlcSph levels have already been used to differentiate GD patients from healthy individuals and GD carriers[296]. Therefore, GlcSph presents as a sensitive and reliable biomarker for GD which may also serve as an indicator of disease burden and response to treatment

1.4.8 Therapeutic Strategies for Gaucher Disease

1.4.8.1 Enzyme Replacement Therapy

Up until the early 1990s, treatment of GD was purely palliative, focussing on relieving symptoms associated with certain organs[223]. However, after establishing the biochemical basis of GD as a deficiency in GBA activity, Brady and co-workers proposed that GD may be amenable to enzyme replacement therapy (ERT), in which patients are administered with active exogenous enzyme to compensate for their inherent deficiency[297,298].

To support research of ERT it became apparent that considerable quantities of GBA would be required, therefore, Brady's group began exploring sources of GBA, with human placenta being a rich source. Subsequently, GBA was purified from human placental tissue and administered to two GD patients in an attempt to correct their disease phenotypes[299–301]. A 26% reduction in hepatic GlcCer levels was demonstrated within 24 hours of injection, with normal GlcCer blood levels being achieved within 72 hours[300]. Gradual re-accumulation of GlcCer was observed over several weeks but this result was highly encouraging. Consequently, another GD1 patient, with 10-20-fold higher hepatic GlcCer levels than previously studied patients, was injected with placental GBA. Unfortunately, only an 8% reduction in hepatic GlcCer was observed[301]. It was deduced that insufficient enzyme was administered to induce a therapeutic response and that GD individuals may require different doses. However, the GBA purification procedure could not be scaled-up to meet demand and research was halted by a shortage of pure enzyme. Eventually, a larger scale placental purification procedure was developed and the new enzyme preparation was administered to seven GD patients[302,303]; disappointingly, only three of the patients exhibited beneficial effects[303]. On further investigation it was determined that the glycosylation profile of this new formulation promoted uptake of the enzyme by liver hepatocytes rather than macrophages[302,304,305]. Therefore, the undesirable biodistribution was thought to account for the poor therapeutic response.

Around this time, it was discovered that macrophages display lectins on their surface which exhibit high affinity for mannose-terminated glyconconjugates[304]. Consequently, the GBA formulation was treated with a series of *exo*-glycosidases to yield mannose-terminated glycans and encourage macrophage uptake[306]. Incredibly, this change in glycosylation resulted in a 50-fold increase in GBA uptake by liver macrophages[307]. In subsequent dose-dependent clinical trials, consistent improvements in disease symptoms were observed at

a dose of 60 U/kg (units of enzyme per kg bodyweight)[301,308]. Therefore, weekly administration of 60 U/kg became the recommended dose when the US Food and Drug administration approved this mannose-terminated GBA formulation for the treatment of GD in 1991 under the trade name Ceredase® (Alglucerase), Table 1.3. Clinically, Ceredase® was well tolerated, yielding substantial therapeutic results with few adverse side effects[309]. Initially, there were concerns over a number of patients who had developed IgG antibodies to Ceredase®, however, these antibodies appeared to have little effect on the clinical efficacy[310]. Another concern was contamination by mannosylated proteins, namely human gonadotrophin (hCG) which diminishes the biological effect of the enzyme administered[311]. However, the biggest issue was the thousands of metric tonnes of human placenta required to support Ceredase® production[309]. Consequently, there was a strong incentive to develop recombinant strategies for GBA production.

A few years later, the Genzyme Corporation (Sanofi Genzyme, Cambridge, MA, USA) produced GBA using recombinant technology in a Chinese hamster ovary cell line (CHO)[312]. To ensure macrophage uptake, the resulting recombinant enzyme was treated with a series of *exo*-glycosidases to yield terminal mannose residues. In contrast to Ceredase® which contained fucose and *N*-acetylglucosamine residues, this recombinant enzyme retains high-mannose glycans at all *N*-glycosylation sites[309] and exhibits similar tissue pharmacokinetics to the placenta-derived formulation[313,314]. This formulation was approved for the treatment of GD in 1994 under the trade name Cerezyme® (Imiglucerase), which remains one of the most prominent ERTs to date[312], Table 1.3. Importantly, the use of Cerezyme® in ERT has proved safe and effective in improving and even reversing GD manifestations with fewer side effects than Ceredase®[313]. Specifically, significant improvements in hepatosplenomegaly and haematological parameters are typically observed within a year of commencing treatment[315,316]. Skeletal manifestations are usually slower to respond[317], however, bone pain and bone marrow parameters may begin to improve after 6 months of therapy[318]. The slow skeletal response has been attributed to the fact that skeletal manifestations often result from end-organ damage and require additional therapies[318,319].

In 2009, a vesivirus infection at Genzyme's production facility halted the production of Cerezyme® leading to a worldwide shortage. This made way for a number of alternative ERT drugs and novel therapeutics, Table 1.3. In 2010, another recombinant product called Velaglucerase alfa (Vpriv®, Shire HGT Inc, Lexington, MA, USA) was licensed for use in the USA and Europe. This form of GBA is produced by specific gene activation in a HT-1080 cell line derived from a human fibrosarcoma[309,320]. Importantly, this cell line is cultured in the

presence of kifunensine, a mannosidase I inhibitor, which results in the synthesis of recombinant GBA with immature high mannose-type glycans[320]. Consequently, no further glycosylation modifications are required for macrophage uptake, providing a significant advantage over Cerezyme®. Additionally, Velaglucerase alfa is reportedly internalised by macrophages 2.5-fold faster than Cerezyme®, likely resulting from its glycosylation pattern[320]. Owing to a shortage of Cerezyme®, many patients were switched to Velaglucerase alfa with considerable therapeutic benefit and no notable adverse changes in clinical parameters. In fact, Velaglucerase alfa has proven as therapeutically effective as Cerezyme® for the treatment of GD[321,322].

Just a couple of years later, a novel plant-derived variant of human GBA, called Taliglucerase alfa (Elelyso®, Pfizer, New York, USA), was developed by Pfizer and Protalix and approved for use in the US, Table 1.3. This product is expressed in carrot-root cells using plant specific *C*- and *N*-terminal sorting signals which facilitate targeting of the nascent protein to storage vacuoles[323]. To permit secretion from the vacuoles, the enzyme is equipped with glycosylation sequences containing core β-1,2-xylose and α-1,3-fucose residues, typical of plant systems[324]. These sugars are not usually observed in mammalian systems, so there were concerns regarding efficacy and immunogenicity, however, no evidence of increased immunogenicity or reduced macrophage uptake have been found[323,325]. This may be attributed to that fact that high mannose-terminated glycans account for over 90% of its *N*-glycans[323], which is likely due to the action of various vacuole-based glycosidases[326]. Therefore, this formulation does not require additional glycosylation modifications to ensure macrophage uptake. Clinical trials with Taliglucerase alfa revealed significant reductions in liver and spleen volumes, improved haemoglobin and platelet levels and clearance of Gaucher cells from the bone marrow[325,327–329]. Subsequently, Taliglucerase alfa was approved for clinical use in the US, Canada, Israel and Brazil but has not been approved in Europe[328].

There are currently no criteria for the preferential use of one ERT formulation over another to treat GD1 patients, however, guidelines for individualisation of dosing and maintenance have been published[330]. In contrast, Cerezyme® is the only ERT with authorisation for the treatment of GD3 individuals, whilst none of the ERTs are indicated for GD2 because ERT therapy has no impact on the rapid progression of neurological symptoms.

Table 1.3: Summary of commercially available GBA formulations for ERT

	Alglucerase	Imiglucerase	Velaglucerase Alfa	Taliglucerase Alfa
Manufacturer	Genzyme Corp	Genzyme Corp	Pfizer and Protalix	Shire HGT inc
Source	Human placenta	CHO cell line	HT-1080 cell line	Carrot root cells
Trade name	Ceredase®	Cerezyme®	Vpriv®	Elelyso®

Despite the positive clinical effects of ERT, there are still a number of limitations to this approach. Firstly, ERT is extremely costly with an estimated annual cost of $200,000 per patient[331,332]. Whilst many GD patients benefit considerably from ERT, a wide variation in therapeutic response is observed. In fact, no correlation between genotype, phenotype and response to ERT has been identified[333], but a number of factors have been associated with poor ERT response, namely haematological malignancies, liver cirrhosis and pulmonary disease[223]. Patients with such complications may respond poorly to ERT, requiring higher doses of enzyme with limited benefit[316]. Another drawback of ERT is the commitment to lifelong bi-weekly injections, which are painful and inconvenient. Furthermore, administration of exogenous enzyme has no effect on the neurological manifestations of GD because the enzyme is unable to cross the blood-brain barrier[334,335]. Therefore, ERT is unsuitable for the treatment of patients suffering from neuronopathic GD. This has encouraged the development of alternative therapies with the ability to alleviate neurological symptoms

1.4.8.2 Substrate Reduction Therapy

In 1980, Vunnman and Radin first proposed the use of substrate reduction therapy (SRT) as a therapeutic strategy for LSDs[336]. Akin to ERT, the aim of SRT is to reduce the accumulation of glycolipid substrates within the macrophage system. However, the concept of SRT involves inhibiting the initial synthesis of the substrate rather than facilitating substrate hydrolysis[337]. In patients who retain some residual activity, inhibition of the enzyme responsible for substrate synthesis should prevent further accumulation of substrate and permit the slow hydrolysis of accumulated substrate by endogenous enzymatic activity[338]. Over time, the balance between substrate synthesis and degradation should be restored, however, this is a slow process meaning clinical benefits present much later than ERT[339].

In the case of GD, SRT involves inhibition of glucosylceramide synthase (EC 2.4.1.80, GCS) which catalyses the synthesis of GlcCer by the transfer of glucose to ceramide[340]. It should be noted that GlcCer synthesis is the rate-determining step in the biosynthesis of most

GSLs[160], therefore inhibition of GCS may hold promise for treating other glycosphingolipid diseases[341]. Currently, two commercial SRTs are available for GD: Miglustat (N-butyl-1-deoxynojirimycin, Zavesca®, Actelion Pharmaceuticals US Inc, USA) and Eliglustat (Cerdelga®, Sanofi Genzyme)[338], Figure 1.12.

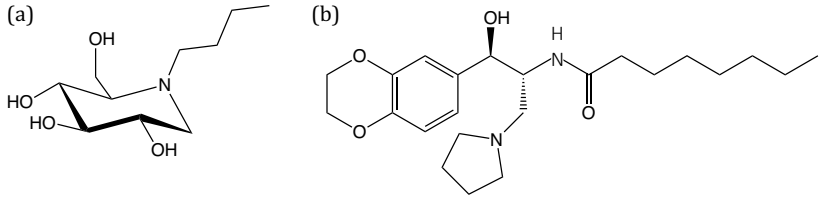


Figure 1.12: Structure of (a) Miglustat (N-butyl-1-deoxynojirimycin) which mimics the glucose moiety of the natural GlcCer substrate and (b) Eliglustat which mimics the ceramide portion of the natural substrate.

Miglustat was the first SRT to be approved for the treatment of GD1 patients for whom ERT is unsuitable[342]. Miglustat is a weak, non-specific inhibitor of GCS which mimics the glucose moiety of GlcCer[343], Figure 1.12 (a). In initial animal studies, Miglustat was shown to reduce the accumulation of GlcCer and delay the onset of GD symptoms[344]. In a 12-month clinical trial, oral administration of Miglustat resulted in improvements in hepatosplenomegaly, thrombocytopenia, haemoglobin levels and platelet count[345]. In a subsequent 36-month extension study, the long-term efficacy and safety of Miglustat was demonstrated, with further therapeutic benefits and reduction in CHIT1 activity, indicating a relief in GlcCer-laden macrophages[339,346]. Additionally, Miglustat was shown to cross the blood-brain barrier, achieving significant tissue distribution within the brain. However, in a phase II clinical trial with GD3 patients, Miglustat proved ineffective for the treatment of neurological manifestations[347]. Furthermore, a number of adverse side effects were reported resulting from its broad specificity and off-target inhibition of other glycosidases[348]. In particular, Miglustat is known to inhibit intestinal glycosidases leading to gastrointestinal complaints which caused many patients to withdraw from treatment in clinical trials[342,349,350]. Fortunately, these side effects were overcome by Eliglustat, the second SRT to be approved for the treatment of GD.

Eliglustat was approved in the US in 2014 and later by the EU in 2015 as a first-line SRT therapy for GD1 patients with compatible CYP2D6 metaboliser phenotypes[337,351]. Eliglustat is a more potent and specific inhibitor of GCS than Miglustat and functions by mimicking the ceramide moiety of GlcCer[351,352], Figure 1.12 (b). It is preferentially metabolised by the CYP2D6 enzyme of the cytochrome P450 pathway[352], therefore, suitability for Eliglustat treatment is currently based on the CYP2D6 metaboliser status of the patient[337].

Additionally, Eliglustat is contraindicated in patients with certain cardiac conditions[351], nevertheless, Eliglustat is suitable for the majority of GD1 patients. The clinical development programme for Eliglustat is the largest ever clinical study conducted for GD, involving 393 patients from 12 countries[351]. In a number of phase II and phase III clinical trials, Eliglustat was orally administered to treatment-naive and ERT patients to establish its safety, tolerability and pharmacokinetic profiles[351,352]. Significant reductions in liver and spleen volumes were observed as well as improvements in haematological and skeletal parameters[351,353]. Furthermore, the rate of adverse side effects was drastically reduced compared to Miglustat, with only 3% of patients withdrawing due to side effects[351]. This improved safety profile is attributed to its specificity for GCS and minimal off-target inhibition[348]. Importantly, Eliglustat also displays good oral bioavailability and broad tissue distribution[352,354]. However, the ceramide-like nature of Eliglustat hinders its ability to cross the blood brain barrier, which means very poor distribution is observed in the brain and CNS[352]. This is further exacerbated by the multidrug transporter Pgp-1 (p-glycoprotein 1), which immediately transports Eliglustat out of the CNS[337]. Therefore, Eliglustat is ineffective against the neurological manifestations of GD2 and GD3 patients[355].

Recent research regarding the development of CNS-accessible inhibitors has produced some promising results[356], and investigations into the combined use of ERT and SRT in mouse models have indicated that combination therapy may offer greater therapeutic benefits[357]. Despite the success of ERT and SRT, both therapeutic approaches are limited by their inability to alleviate neurological manifestations. Consequently, current research regarding the treatment of GD is heavily focussed on the development of novel therapeutic approaches with the ability to treat neuronopathic forms of GD.

1.4.8.3 Pharmacological Chaperone Therapy

It is generally accepted that the accumulation of substrate in most LSDs occurs when the residual enzymatic activity falls below a certain threshold[358]. A threshold activity of ~ 10% the normal level is thought to be sufficient to prevent the accumulation of substrate in most LSDs[359]. Therefore, depending on their genotype, some patients may only require a small increase in enzymatic activity to achieve therapeutic benefit. One approach to increasing residual enzymatic activity is pharmacological chaperone therapy (PCT).

In some cases of GD, mutations in the *GBA1* gene result in the synthesis of misfolded GBA[360] These mis-folded mutants may still exhibit partial catalytic activity but are degraded by the quality control system of the ER before they can be trafficked to the lysosome[360]. Therefore,

the aim of PCT is to stabilise partially active, misfolded GBA so it can pass through the ER and be transported to the lysosome to degrade accumulating substrates[361]. This approach involves the use of pharmacological chaperones (PCs), which are reversible non-covalent inhibitors that when used at sub-inhibitory concentrations bind to the misfolded enzyme of interest and stabilise its conformation[360]. This prevents premature degradation in the ER and permits further maturation of the mutant enzyme, allowing it to be trafficked to the lysosome[355]. This results in an increase in residual enzymatic activity that facilitates the hydrolysis of accumulated substrate[360,362].

Initially, a variety of alkylated derivatives of the iminosugar 1-deoxynojirimycin (DNJ), Figure 1.13(a), were highlighted as potential PCs for GD[363]. DNJs are known to inhibit numerous enzymes including the ER oligosaccharide-processing enzymes α-glucosidase I and II, glucosylceramide synthase, non-lysosomal β-glucosidase (GBA2) and lysosomal GBA[363]. However, alkylation of DNJ was found to drastically improve selective inhibition of GBA over related enzymes[363]. Subsequently, a number of alkylated DNJ derivatives have been tested as putative chaperones for GBA. In a crucial study, a GBA N370S mutant fibroblast was incubated with sub-inhibitory concentrations of N-nonyl-1-deoxy-nojirimycin (NN-DNJ), Figure 1.13 (b), resulting in a 1.65-fold increase in GBA activity which persisted for 6-days following removal of NN-DNJ[363]. It became evident that the length of the alkyl chain significantly impacts the ability of DNJ to bind and stabilise GBA[363]. Specifically, it was proposed that the hydrophobic alkyl chain mimics the ceramide moiety of the natural substrate and improves the affinity for the enzyme by exploiting hydrophobic recognition and binding elements of the active site[363].

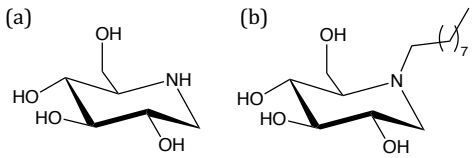


Figure 1.13: Chemical structure of (a) 1-deoxynojirimycin (DNJ) (b) alkylated iminosugar *N*-nonyl-1-deoxynojirimycin (NN-DNJ).

Alternative non-alkylated iminosugar based inhibitors have also been investigated as PCs for GBA. Of note, Isofagomine (IFG Afegostat, Amicus Therapeutics and Shire plc), a known active site directed GBA inhibitor, remains one of the most studied iminosugar inhibitors for PCT[364–367], Figure 1.14 (a). Structural studies with GBA suggest that IFG induces changes to the surface topology of GBA which results in the formation of a new hydrogen bonding network between core residues that in turn stabilises a distinct GBA conformation[368]. This is thought to stabilise GBA by locking the enzyme into a substrate-bound conformation[368].

During primary *in vitro* investigations, N370S fibroblasts incubated with IFG exhibited a 2-fold increase in GBA activity[365]. This activity was localised to the lysosome using the lysosomal marker LAMP1, suggesting that IFG increases the activity of the N370S mutant and enhances its transport to the lysosome[365]. In subsequent studies, increased GBA activity was also observed in IFG treated mice with V394L, D409H, D409V and L444P mutations[365–367]. Promisingly, IFG was able to access the CNS with a broad tissue distribution. Based on a number of pre-clinical studies, IFG advanced to phase II clinical trials but was later dropped when it failed to meet clinical end-points[369] (partially attributed to its broad specificity[370]).

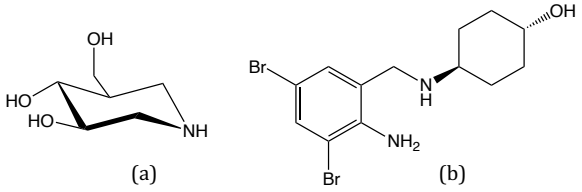


Figure 1.14: Chemical structure of (a) Isofagomine (IFG) and (b) Ambroxol.

In a more recent effort to identify novel PCs for GD, a high through-put screening assay of the FDA approved drug library identified Ambroxol, a widely used expectorant, as a pH-dependent inhibitor of GBA[371], Figure 1.14 (b). Subsequent *in vitro* testing with N370S and F213I mutant fibroblasts, demonstrated that Ambroxol enhances GBA activity in the lysosome[371,372]. Increased levels of the lysosomal membrane protein LIMP-2 have also been observed[372,373]. Although the mechanism by which Ambroxol improves LIMP-2 activity is yet to be elucidated, this phenomenon may account for its ability to improve the lysosomal trafficking and activity of GBA. In clinical studies, Ambroxol proved to be safe at high doses with good tolerability and few side effects[374]. Specifically, therapeutic benefits for visceral and haematological manifestations have been reported[375], as well as considerable improvements in gross motor function for patients suffering with GD3[374]. This suggests Ambroxol is able to cross the blood-brain barrier and access the CNS, as confirmed by a reduction in GlcSph levels in cerebrospinal fluid[374]. Therefore, Ambroxol presents as the most promising PC for GD to date, with considerable potential for the treatment of neuronopathic GD3[374,376]. Additionally, Ambroxol has recently been identified as a potential neuroprotective compound for the treatment of PD[377]. Consequently, a great deal of research into its development has been published[371–375].

Other putative GBA chaperones include bicyclic L-idonojirimycin[378], which has been proposed for patients homozygous for the L444P mutation, N-octyl-β-valienamine, which has proved a modest chaperone for the F213I mutant, and 2-alkyl trihydroxypiperidines which have chaperoning activity towards heterozygous N370S and homozygous L444P[379].

The main advantages of PCs is their ability to access the CNS and provide therapeutic benefits for neurological manifestations[358]. Additionally, most PCs are administered orally[338], eliminating the need for bi-weekly injections. Furthermore, PCs can be administered in combination with ERT to improve the stability, activity and trafficking of the recombinant enzyme[298]. For example, a recent study regarding the combined used of Ambroxol and ERT for neuronopathic GD has demonstrated that high doses of Ambroxol can be safely administered to enhance the activity of recombinant GBA, resulting in considerable improvements in neurological manifestations[374]. PCs may also permit the use of lower enzyme doses during ERT, helping to lower the cost of treatment. However, not all GD patients will benefit from PCs because PCT is only suitable for 'responsive' GBA mutants which yield misfolded GBA with partial catalytic activity.

1.4.8.4 Strategies for the Future

In light of the various limitations of ERT, SRT and PCT, there is a strong incentive to develop "one-time", permanent therapies for GD which alleviate the requirement for costly, regularly administered treatments. In this regard, allogeneic hematopoietic stem-cell transplantation (HSCT) has been investigated as a curative approach for GD1[380]. Indeed, hematopoietic stem cells (stem cells that can develop into all types of blood cells) are considered ideal targets for gene transfer due to their longevity and the capacity for self-renovation[381]. Typically, hematopoietic stem cells are harvested from the patient, gene edited (through a viral vector) and then infused back into the patient for engraftment380. HSCT for GD was initially tested in the early 1980s using bone marrow transplantation in a GD3 child[382]. Despite promising post-transplant results and enhanced plasma GBA activity, no long-term significant changes in Gaucher cell infiltration or clinical status were observed. Nevertheless, numerous studies on bone marrow HSCT were performed in the 1980s-1990s with varying degrees of success[382–386]. Some patients showed rapid improvements in GD manifestations and continued to lead active lives for many years following transplantation, however, others exhibited far less success[380]. Owing to the significant transplant-related morbidity and mortality, HSCT was quickly surpassed by ERT and has only been reserved for individuals with severe neurological disease that are unresponsive to ERT[387]. However, given the limitations of ERT, namely the financial burden, need for bi-weekly injections and ineffectiveness against neurological disease, people are seeking to establish if HSCT could be a safe, one-time curative approach for GD[380].

A similar but potentially safer approach to HSCT is engineering the expression of GBA in the patient's own hematopoietic stem and progenitor cells (HSPCs) which eliminates graft-versus-host disease (a condition which arises when the recipients body attacks the donated bone marrow)[388]. Non-targeted gene addition into HSPCs using retro- and lenti-viral vectors has been explored for GD with some promising results in murine models[389,390]. However, concerns regarding potential mutagenesis and malignant transformation during viral gene transfer is driving the need for targeted gene addition strategies. For example, a recent study reported a CRISPR/Cas9-based approach to successfully gene edit HSPCs to produce GBA-expressing macrophages and monocytes[391]. It was reasoned that the primary manifestations of GD arise from pathologies in monocytes and macrophages, therefore, restoring GBA activity in such cells should be sufficient to provide some phenotypic correction[391]. Whilst these studies remain in their infancy, this approach appears suitable for gene editing of HSPCs to induce specific protein expression which may show flexibility for the treatment of other LSDs.

Clearly gene therapy represents a future therapeutic option for GD and will likely gain considerable academic and pharmaceutical attention as a curative approach. Nevertheless, the associated risks and ethical issues will complicate progress in gene therapy[381].

1.5 Fabry Disease

1.5.1 Overview and History

Fabry's disease (FD, Anderson-Fabry disease, OMIM 301500) is an X-linked inherited LSD first reported in 1898 by dermatologists Johannes Fabry and William Anderson who independently described patients with red-purple lesions on the skin (angiokeratoma corporis diffusum)[392]. Initially, FD was recognised as a systemic vascular disease due to the presence of abnormal vacuoles found in cells of affected individuals[393]. It was later identified as a lipid storage disorder when the cellular accumulation of globotriaosylceramide (Gb3) was discovered[393,394]. In 1965, electron microscopy studies revealed the presence of unusual bodies in the endothelial smooth muscle and perivascular cells of FD individuals[158]. These bodies were described as 'extremely overcrowded lysosomes' which result from defective lysosomal function and subsequent lipid accumulation[158]. In the same year, Dempsey and co-workers reported that a sex-linked deficient gene, with constant penetrance in homozygous males and occasional penetrance in heterozygous females, was the underlying genetic abnormality[159]. Soon after, the biochemical basis of FD was

established by Brady and co-workers, who discovered that deficiencies in human α-galactosidase A (α-GAL) lead to the accumulation of Gb3 throughout the body[395].

1.5.1.1 Epidemiology

Following GD, FD is the second most common lysosomal storage disorder with an estimated incidence of 1:40,000-117,000 worldwide[147,396]. However, new-born genetic screening studies have reported incidences up to 1:1,250[397], suggesting a drastic underestimation in the number of cases due to undiagnosed and late-onset patients[148,398] Additionally, the incidence rate appears to vary with race and geographical location, with Italy and Taiwan reporting surprisingly high frequencies of 1:3,200[398] and 1:1,500[399] respectively. Despite being an X-linked inherited disorder, FD effects both men and women, however, it is estimated that only 70% of females with α-GAL defects display symptoms, whilst almost 100% of males exhibit clinical manifestations[162,400].

1.5.2 Clinical Manifestations

In FD patients, the loss of functional α-GAL leads to a progressive build-up of Gb3 within the lysosome of cells throughout the body and this is thought to be responsible for the clinical symptoms[401]. Gb3 primarily accumulates in cells of the heart, kidney, skin, brain and eyes, initially resulting in asymptomatic cellular damage and dysfunction[401,402]. This cellular dysfunction is thought to trigger a cascade of events including cellular death, oxidative stress[403], compromised energy metabolism[404], blood vessel injury and tissue ischemia leading to irreversible cardiac damage and renal fibrosis[405,406]. As FD progresses, the incessant damage to vital organs results in organ failure with end-stage renal disease and cardiovascular/cerebrovascular complications[407]. Although this general disease progression has been established, the clinical manifestations of FD can vary in frequency and severity, Table 1.4. Currently, FD is categorised into "classical" and "non-classical" phenotypes depending on the amount of residual α-GAL activity present[401,408,409].

1.5.2.1 Fabry Disease Type 1

Type 1 "classical" phenotype is associated with patients who have little or no detectable α-GAL activity, typically < 1% the normal activity. These individuals are most severely affected, experiencing early symptoms in childhood or adolescence due to the accumulation of Gb3 in capillaries and small blood vessels[401,410]. Early symptoms in children include severe burning pain in the hands and feet (acroparesthesias), gastrointestinal problems,

diminished ability to sweat (hypohydrosis), chronic fatigue and poor growth[409,411]. However, the most visible early clinical feature is small dark spots on the skin (angiokeratoma) resulting from benign lesions of the capillaries[411,412]. With increasing age, the continuous deposition of Gb3 leads to vascular disease of the heart, kidneys and brain[411,413,414]. Major manifestations of classic FD include progressive kidney dysfunction leading to renal failure, strokes and cardiac disease, which eventually result in premature death in the fourth or fifth decade of life[415,416]

1.5.2.2 Fabry Disease Type 2

In contrast to the classic phenotype, patients with type 2 "non-classic" FD have some residual α-GAL activity, varying between 2-20% the normal range[409,411,414]. These patients do not exhibit the early manifestations of FD, presumably due to the presence of residual α-GAL activity which is sufficient to prevent the initial accumulation of Gb3. Non-classic FD patients experience milder, late-onset symptoms which present in the fourth to sixth decade of life and are typically confined to one organ[417]. Symptoms commonly include cardiac or renal disease, however, some individuals experience no clinical disease at all[411,418,419,] As result of the late onset of mild symptoms, many non-classic FD patients go undiagnosed, in fact, most type 2 individuals are only identified through screening in cardiac, haemodialysis, renal transplant and stroke clinics[420]. Consequently, the prevalence of non-classic FD has been drastically underestimated. The true prevalence is still uncertain, however, based on new born screening studies, it is estimated that type 2 FD is 5-10 time more frequent than classic FD in males from the same region, ethnic group or race[398,399,421].

1.5.2.3 Fabry Disease in Females

In view of the fact that FD is a hereditary X-linked lysosomal storage disorder, heterozygous females were long thought to be carriers of the disease and generally resistant to developing clinical symptoms[422]. However, as our knowledge of the disease advanced, it became apparent that FD exhibits a wide spectrum of heterogeneous, progressive clinical phenotypes in females. The clinical manifestations for females range from asymptomatic to symptoms as severe as those observed in males[423]. In fact, heterozygous females can display severe manifestations affecting the kidneys[424], heart[425], brain[426], intestines[427] and respiratory system[428]. The exact mechanism by which female heterozygotes develop symptoms is yet to be elucidated, however, the variation in phenotypes is thought to result from random X-chromosome inactivation[429,430].

Table 1.4: Overview of typical FD symptoms by age of onset[431]

Age of Onset	Symptoms
Childhood/Adolescence (<16 years)	Angiokeratomas
	Gastrointestinal disturbances
	Ophthalmological abnormalities
	Hearing impairment
	Hypohidrosis
	Lethargy
	Neuropathic Pain
	Onset of renal/cardiac abnormalities
Early adulthood (17-30 years)	Worsening of the above
	Progressive renal failure
	Cardiomyopathy
	Strokes/ischemic attacks
	Nervous system problems
Late adulthood (>30 years)	Worsening of the above

If left untreated, FD reduces male life expectancy by about 20 years, with a sharp decline in survival after the age of 35[414]. The effect on female life expectancy is less well-established but it is thought to be reduced by approximately 10 years[414–416].

1.5.3 Fabry Disease and α-GAL

In the late 1980s, the entire genomic sequence and full length complimentary DNA (cDNA) encoding α-GAL were isolated and characterised[432,433]. The full length 12 kb *GLA* gene was sequenced and mapped to chromosome Xq22.1 located on the long arm of the X-chromosome[433], Figure 1.15. Numerous other genetic disorders have been localised to the Xq22 region, namely Alport syndrome[434,435], Burton's agammaglobulinemia[436] and Megalocornea syndrome[437]. The full-length cDNA encodes the 429 amino acid polypeptide α-GAL which contains a 31-residue signalling sequence[432,438].

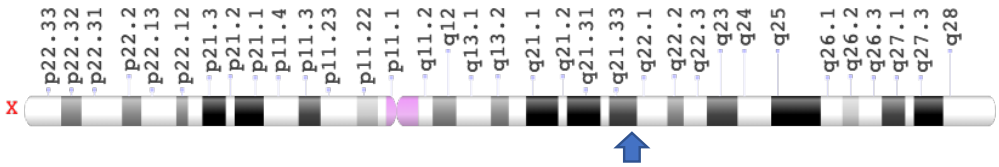


Figure 1.15: Location of *GLA* gene at Xq21 on the X-chromosome indicated by the blue arrow. Centromere highlighted in pink. Figure prepared using NCBI genome decoration tool (www.ncbi.nlm.nih.gov/genome/tools/gdp) with GRCh38.p12 representation.

It is understood that mutations in the *GLA* gene result in defects in α-GAL, which lead to complete or partial loss of enzymatic activity[397,439]. Subsequently, Gb3 and related GSLs accumulate in the lysosome, leading to the clinical manifestations of FD. Disease causing

mutations have been identified throughout the enzyme structure and can be categorised into three classes; (i) mutations near the active site which directly interfere with substrate binding[440] (ii) mutations in burial regions which disrupt the hydrophobic core of the enzyme and induce misfolding[410,440] and (iii) mutations which have other effects on the folded state of the enzyme, such as those which interfere with dimerisation[441].

A variety of techniques have been employed to identify the nature and frequency of FD causing mutations[442]. Initial studies assessed the frequency of disease causing rearrangements in 165 unrelated FD patients using PCR amplification of the entire *GLA* coding region[153,443]. One partial gene duplication and five partial gene deletions were highlighted at a frequency of 5%, typical of X-linked disorders[443]. In addition, small deletion and insertion mutations were identified which resulted in premature peptide chain termination. Single base substitutions, splice and deletion mutations affecting RNA processing of the *GLA* transcript were also detected[442]. According to The Human Gene Mutation Database[218] (www.hgmd.org, Institute of Medical Genetics in Cardiff) over 1,000 *GLA* gene mutations have been identified to date, including missense/nonsense mutations, splicing mutations, deletions and insertions as well as complex rearrangements.

1.5.3.1 Genotype-Phenotype Relationships

Attempts to correlate genetic mutations with clinical phenotypes have been relatively unsuccessful due to the variability of disease onset, symptoms and progression[411]. Genotype-phenotype studies have also been hindered by the rarity of the disease; however, some general associations have been observed. For example, N215S, Q279E, M296V and R301Q mutations are atypical mild mutations associated with individuals who are asymptomatic or experience a mild form of the disease[444]. In contrast, R227Q, R227X and S297F mutations are associated with the severe classical phenotype[444]. It is generally thought that mutations leading to little or no α-GAL activity result in the classic phenotype, whilst mutations which permit residual α-GAL activity lead to the non-classic phenotype[442].

Despite our extensive knowledge of *GLA* mutations, novel mutants continue to be identified. For example, in a recent sequence analysis study, Onay et al (2020) reported two novel *GLA* mutations in the Turkish population[445]. In fact, this study provides the largest reported *GLA* mutation spectrum in the Turkish population to date, with the aim of providing insight into genotype-phenotype correlations. Although it is not yet possible to predict a phenotype from the location and type of mutation, there are databases (www.dbFGP.org and Fabry-Database.org[439]) which provide phenotype assignments for reported *GLA* mutations. These

databases were curated to provide health-care workers, patients and researchers with easily accessible information about currently known genotype-phenotype correlations[439].

1.5.4 Fabry Disease Pathogenesis

FD is typically characterised by the accumulation of Gb3 within the lysosome of cells throughout the body, however, other glycolipids, such as globotriaosylsphingosine (lyso-Gb3), also accumulate as a result of α-GAL deficiencies[446]. The accumulation of these glycolipids alone is not sufficient to account for the full pathophysiology of the disease and the full pathogenic mechanisms behind FD have not been fully elucidated[447].

Historically, the 'cytotoxicity hypothesis' was adopted to rationalise the pathophysiology of most LSDs[448]. This hypothesis presumes that lipid-laden lysosomes cause the clinical symptoms by somehow inducing abnormal cell function or cell death[448]. In the case of FD, elevated Gb3 levels have been observed in the lysosomes of endothelial, epithelial, smooth muscle, corneal, myocardial and renal cells as well as in the meninges, cerebral blood vessels, connective tissue and ganglia of the central nervous system[402,412]. This widespread deposition of Gb3 is therefore believed to be responsible for the multisystemic clinical manifestations. Specifically, accumulation of Gb3 within the vascular system is thought to cause narrowing of the blood vessels, leading to ischemia and infarction (tissue death due to inadequate blood supply) in affected organs[447]. However, this hypothesis does not fully explain the pathogenic role of accumulated glycolipids and fails to explore the subsequent molecular cascades that lead to the widespread cellular dysfunction and devastating clinical symptoms. Crucially, this hypothesis does not consider the influence of the lysosome on other cellular processes, such as immune system regulation and homeostatic control[447], which may also contribute to clinical manifestations if perturbed. Consequently, the 'cytotoxicity hypothesis' is now considered a drastic oversimplification of FD pathogenesis.

In order to understand the mechanisms by which α-GAL deficiencies cause FD, interactions of glycolipids and the lysosome with other cellular systems and processes must considered. This became of added importance following the finding that Gb3 also accumulates in other cellular structures, including the ER, cell membrane and the cell nucleus[402]. Additionally, defective lysosomal function is thought to impact the regulation of the immune system by altering multiple stages of the immune response, including antigen presentation[449], phagocytosis and release of pro-inflammatory mediators[447]. Therefore, it is important to consider the involvement of the immune system in FD pathogenesis.

1.5.4.1 Involvement of the Immune System

Studies regarding immune function in LSDs have shown that accumulated glycolipids stimulate a variety of pathogenic pathways and inflammatory responses by triggering the production of proinflammatory cytokines and inflammation markers[256,450,451]. In the case of FD, increased tissue infiltration of lymphocytes and macrophage related markers has been identified in endomyocardial biopsies, indicating increased inflammation in the cardiovascular system[452]. In addition, the widespread accumulation of Gb3 has been linked to the production of reactive oxygen species (ROS) which cause oxidative stress and damage to vascular endothelial cells[403]. These ROS can also induce protein damage[453]. It is postulated that these damaged proteins may act as neo-antigens, evoking an autoimmune response which further exacerbates inflammation and contributes to the autoimmune disorders associated with FD[453,454].

It is evident that disruption of lysosomal function by abnormal glycolipid storage has the potential to compromise many normal cell operations, resulting in a range of downstream consequences for the immune system, CNS and vascular system[450]. By promoting chronic inflammation, cell damage can occur through multiple mechanisms leading to the development of clinical manifestations which are otherwise unrelated to the initial pathology[447]. Additionally, the accumulation of Gb3 is known to contribute to endothelial dysfunction, impaired blood vessel structure, cardiac thromboembolic events and abnormal blood components, all of which can lead to life threatening strokes and ischemic attacks[447]. Consequently, the pathogenesis of FD is far more complex than initially believed.

1.5.5 Diagnosis of Fabry Disease

Early diagnosis of FD is vital in ensuring effective treatment, however, diagnosis is often hindered by the multisystemic and non-specific nature of the symptoms[411,455]. In fact, the diagnostic delay for most FD patients is ~ 15 years[456].

1.5.5.1 Enzymatic Diagnosis

Before the biochemical basis of FD was established, a diagnosis was made by the presence of angiokeratomas with a positive family history of FD or fatal kidney/heart disease[412]. Diagnoses were typically supported by the presence of other cutaneous markers as well as cardiovascular, gastrointestinal, renal or respiratory manifestations[412]. Nowadays, a biochemical diagnosis can be made by screening for reduced or absent α-GAL activity in

blood plasma or cultured skin fibroblasts using a fluorogenic substrate activity assay[457]. Generally, activity less than 20% of the normal is considered diagnostic, however, levels below 35% can be indicative[401,458]. Even prenatal diagnoses can be made, as early as the 17th week of pregnancy, by demonstration of reduced α-GAL in cultured fetal cells[459]. Although this enzymatic approach is generally reliable for male patients, diagnosis by α-GAL activity alone is not always sufficient for female patients because they may exhibit deficient to normal α-GAL activity due to random X-chromosome inactivation in the sample[165,430]. This is further complicated by the fact that female patients can experience mild symptoms[423], therefore, alternative diagnostics are required.

1.5.5.2 Genetic Diagnosis

Advances in DNA analysis and characterisation of the *GLA* gene have permitted the development of genetic diagnosis[153,433]. Given the extreme variation in the type and number of *GLA* mutations, no single protocol is able to detect all mutations, consequently, a number of polymerase chain reaction (PCR) based techniques have been employed to identify *GLA* mutations[410,460–462]. The Multiplex Ligation-dependent Probe Amplification (MLPA) method, developed in 2002, can readily detect rearrangements and is ideal for the detection of deletion and duplication mutations[463,464]. In MLPA, multiple pairs of oligonucleotide probes containing universal primer sequences are hybridised to their perspective targets on genomic target sequences. Once hybridised, adjacent probes are ligated with a thermostable ligase and the fragments are amplified by PCR[463]. In this way, only ligated oligonucleotides are amplified, ensuring the amount of PCR product is proportional to the amount of DNA target. Importantly, each probe has a unique length and a fluorescent label which allows the resulting amplicons to be separated and quantified on an automatic sequencer[463]. This overcomes the resolution limitation of multiplex PCR whilst allowing multiple different sequences to be targeted in a single PCR-based experiments. The MLPA approach has been successful in identifying new FD deletion mutations which were undetectable through more traditional sequencing analysis[461]. However, the success of DNA diagnoses is limited by our knowledge of current *GLA* mutations[439]. As such, it is common practice to use both enzymatic and genetic tests in tandem to make a diagnosis, Figure 1.16.

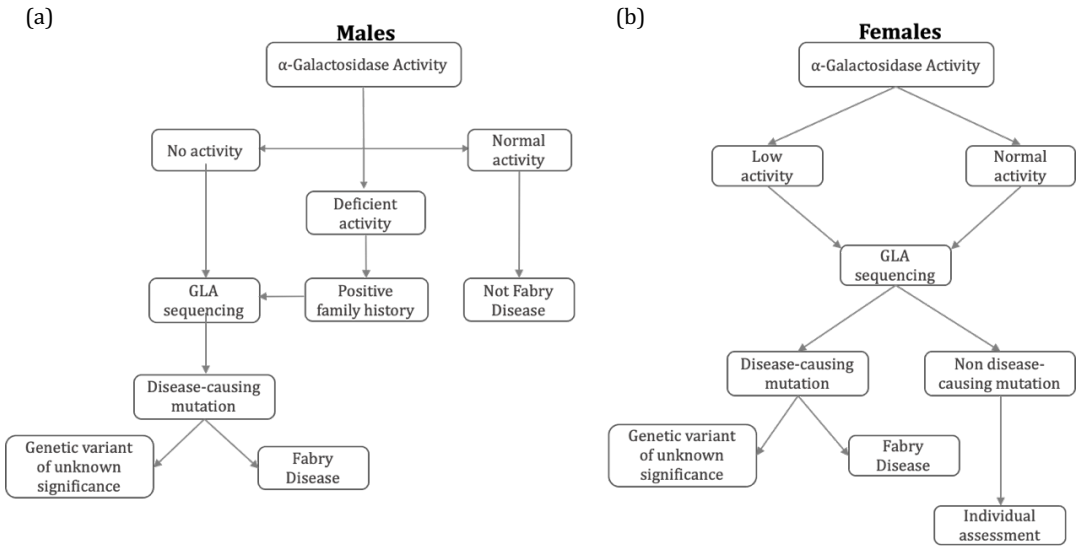


Figure 1.16: General pathway to the diagnosis of (a) male and (b) female individuals suspected of having Fabry disease through enzymatic and genetic diagnoses.

1.5.5.3 Use of Biomarkers

More recently, biomarkers of FD have been considered useful in supporting a diagnosis[287]. As outlined previously, Gb3 is known to accumulate in various organs of FD patients, therefore, identification of elevated Gb3 levels in plasma or urine can support a diagnosis[465]. However, Gb3 is not considered a particularly useful biomarker of FD because its levels do not always correlate with disease severity[411,414]. In fact, plasma Gb3 levels in female patients are usually within the normal range[422]. Urinary Gb3 levels are generally more reliable[466,467], however, some patients with late onset variants exhibit normal urinary Gb3 levels[150]. Furthermore, significant Gb3 accumulation is observed before birth, long before clinical symptoms develop[468]. The discrepancy between Gb3 storage and clinical symptoms suggests that Gb3 is not a reliable biomarker. Consequently, there is an urgent need for more reliable and validated FD biomarkers.

Globotriaosylsphingosine (lyso-Gb3), the deacetylated version of Gb3, has become a more reliable biomarker for FD following the observation that FD patients exhibit increased lyso-Gb3 blood plasma levels[446]. As a result of reduced α-GAL activity, accumulating Gb3 can be converted to lyso-Gb3, Figure 1.17, which is a potent inhibitor of α-GAL[446]. Consequently, lyso-Gb3 can contribute to the pathophysiology of FD by promoting cellular storage of Gb3 through an aggressive cycle[446,469]. A pathological role for lyso-Gb3 in FD is supported by the observation that increased exposure to lyso-Gb3 leads to more severe disease in male and female patients[470,471]. The apparent correlation between lyso-Gb3 and disease severity, renders lyso-Gb3 a more prudent marker of FD[470,472], particularly in women. In fact, lyso-

Gb3 is currently used to evaluate disease progression and monitor the efficacy of therapeutic treatment[473–475].

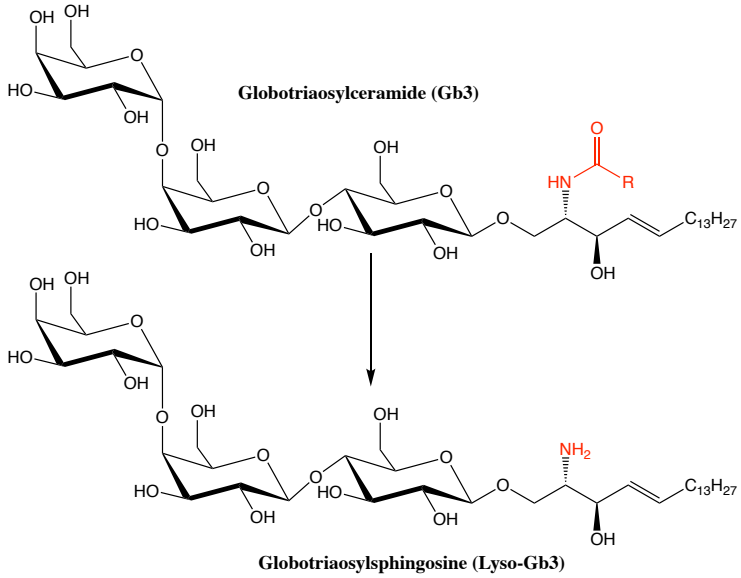


Figure 1.17: Structure of globotriaosylceramide (Gb3) and the deacetylated derivative globotriaosylsphingosine (lyso-Gb3).

A potential alternative biomarker for FD is sphingosine-1-phosphate (S1P)[476]. S1P is a membrane derived signalling lipid which interacts with a group of G-protein receptors to aid in the regulation of cell proliferation and immune cell trafficking[476–478]. However, a potential pathogenic role for S1P in LSDs has recently surfaced. Normal lysosomal GSL catabolism leads to the production of sphingosine, which can be converted to S1P by sphingosine kinases[479], Figure 1.18. It has been suggested that the abnormal GLS catabolism associated with LSDs may perturb S1P levels which in turn contributes to the pathophysiology[480]. In fact, elevated S1P levels have been detected in the plasma of FD patients, with a strong correlation between plasma S1P levels and ventricular hypertrophy (enlargement/thickening of the left ventricle)[270]. Furthermore, development of cardiovascular abnormalities, similar to those of FD patients, has been observed in mice treated with S1P[270]. Therefore, S1P may present as a useful biomarker of FD. However, some studies have demonstrated highly variable S1P levels in plasma samples from FD patients[480], which obscures the pathogenic role of S1P. Additionally, current literature on S1P abnormalities in LSDs is complicated. For example, whilst elevated S1P levels have been implicated in the cardiac pathology of FD[270] and the neuronal pathology of Sandhoff disease[271], deficient S1P levels have been linked to the pathogenesis of Niemann-Pick type C disease[131]. Therefore, further analysis of S1P in relation to FD pathogenesis is required.

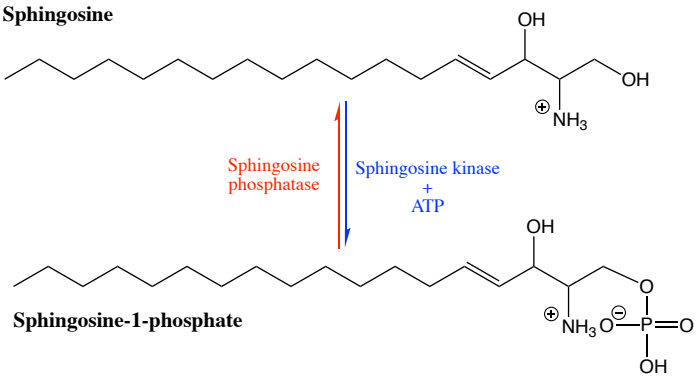


Figure 1.18: Structure and biosynthesis of sphingosine-1-phopshate (S1P) from sphingosine by sphingosine kinase.

1.5.6 Therapeutic Strategies for Fabry Disease

1.5.6.1 Enzyme Replacement Therapy

Prior to 2001, the treatment of FD was purely palliative; fortunately, ERT was approved for FD in Europe in 2001[411,483]. The aim of ERT with respect to the treatment of FD is to prevent disease progression by compensating for the underlying α-GAL deficiency and reducing the accumulation of Gb3[484,485]. This involves intravenous administration of recombinant human α-GAL (r-αGAL) of which there are two available forms; Replagal® (agalsidase alfa), produced by Shire Pharmaceuticals in a genetically engineered human cell line[484], and Fabrazyme® (agalsidase β), manufactured by Genzyme using recombinant DNA technology in a Chinese hamster ovary (CHO) cell line[485]. These products have identical amino acid sequences to the native enzyme and only differ from one another by their glycosylation patterns resulting from the use of different production cell lines[486].

A number of studies have shown that biweekly administration of r-αGAL can lead to clinical benefits for FD patients[487–492]. Once intravenously injected, r-αGAL is internalised by cells through the M6P pathway and is trafficked to the lysosome to degrade accumulated Gb3. Importantly, injection of r-αGAL reduces the level of Gb3 in plasma, urine, skin and the endothelial cells of the liver, kidneys and heart in a dose-dependent manner[137,493]. Improvements in neuropathic pain, kidney function, cardiac function and gastrointestinal complaints have been reported[137], with the risk of major clinical events such as renal disease, cardiac events or death being significantly reduced[484,485,494]. Importantly, ERT provides long term stabilisation of renal complications with a favourable safety profile[495,496]. Indeed, intravenous administration of r-αGAL is generally well-tolerated, with mild infusion reactions being the most common side-effects[137]. It is generally accepted that the earlier

treatment is begun, the greater the therapeutic benefits will be for the patient because ERT has the potential to prevent adverse changes in the heart and kidneys if started early enough. However, ERT is not a cure for FD and is not necessarily beneficial for all patients because irreversible organ damage often occurs before symptoms and treatment start[409,411].

One major drawback of ERT is the short circulating half-life and variable uptake of r-αGAL by disease affected tissues[497]. Notably, both forms of recombinant enzyme are unable to cross the blood-brain barrier and have little effect on neurological symptoms[497]. Furthermore, intravenous administration is often painful and inconvenient for patients, requiring weekly hospital trips[411]. ERT is also very expensive, with an estimated annual cost of €210,000 per patient, imposing a significant cost burden on health care systems[487]. Another major concern regarding ERT is the development of antibodies towards the recombinant enzyme. The production of immunoglobulin G (IgG) antibodies to r-αGAL is common amongst patients undergoing ERT, in fact as many as 50% of all male patients with classic FD develop anti-drug antibodies to the recombinant enzyme[498]. Not only can these antibodies cause hypersensitivity reactions which require longer infusion times and additional medications[499,500], but they also have the potential to reduce treatment efficacy by binding to the administered enzyme and modifying its tissue distribution, metabolic clearance, subcellular trafficking and activity[499]. In particular, the cross-reactivity of antibodies to r-αGAL has been shown to affect the activity of the recombinant enzyme *in vitro*[500]. As a result, IgG antibodies may reduce the exposure of target organs to the administered enzyme and reduce treatment efficacy[500]. Whilst recent pre-treatment prediction models for anti-body development have been reported[498], the lack of appropriate biomarkers for FD complicates the evaluation of the effects of antibodies on treatment efficacy. Consequently, the clinical relevance of antibody production remains poorly understood[500].

1.5.6.2 Pharmacological Chaperone Therapy

Another therapeutic approach is pharmacological chaperone therapy (PCT). In some cases of FD, specific gene mutations result in misfolded α-GAL mutants. *In vitro* studies have demonstrated that these mutants still exhibit partial catalytic activity, however, they are recognised by the quality control system of the ER and degraded before they are trafficked to the lysosome[501]. PCT has the potential to stabilise these mutants and ensure correct trafficking to the lysosome for substrate degradation[502].

The use of PCs to treat FD patients was first demonstrated in 2001, when a patient with Fabry's cardiomyopathy was administered with infusions of galactose[411,503], Figure 1.19 (a). These infusions were shown to increase α-GAL activity and significantly improve cardiac manifestations[503]. Furthermore, these infusions were well tolerated and clinical improvements were sustained for two years post-treatment[503]. The mechanisms by which galactose and other reversible inhibitors enhance the stability and activity of α-GAL has been studied extensively[504,505]. Indeed galactose is thought to behave as a chemical chaperone by binding to the active site of mutant α-GAL and promoting proper folding, dimerization and processing of the enzyme[503]. This reduces the formation of misfolded mutant enzyme and facilitates the transport of stabilised enzyme to the lysosome[411,503]. Once in the lysosome, α-GAL is further stabilised by the acidic environment and the PC (galactose) is displaced from the active site, allowing α-GAL to hydrolyse its glycolipid substrates and reduce the storage burden on the lysosome.

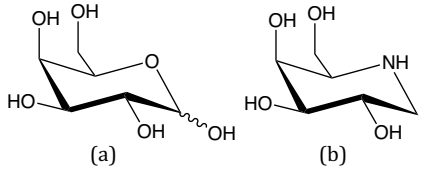


Figure 1.19: Structures of active-site-specific chaperones (a) galactose and (b) 1-deoxygalactonojirimycin DGJ (Migalastat). At lysosomal pH the endocyclic nitrogen of Migalastat is protonated, allowing it to act as a transition state mimic

Further research into molecular chaperones for α-GAL led to the development of the active-site-specific chaperone 1-deoxygalactonojirimycin (DGJ)[505,506], Figure 1.19 (b). DGJ was first isolated in 1988 as a fermentation product of the soil bacterium Streptomyces lydicus[507] and was later identified as a small molecular chaperone for α-GAL. Subsequently, DGJ was further developed by Amicus Therapeutics under the tradename Migalastat[508,509]. Migalastat is a product/transition-state analogue (the latter by virtue of its positive charge at lysosomal pH) which mimics the terminal galactose moiety of the natural Gb3 substrate. When used at suitable concentrations, Migalastat binds reversibly to the active site of α-GAL and stabilises certain mutant forms of the enzyme[506,510]. Following transport to the lysosome, the change in pH induces dissociation of Migalastat, yielding free enzyme which is able to hydrolyse its natural substrates[504,510]. In initial pre-clinical studies, Migalastat was shown to increase α-GAL activity in cultured fibroblasts derived from FD patients[504,509] and in transgenic FD mice expressing mutant α-GAL[511–513]. In a consequent phase 1 trail with healthy individuals, treatment with Migalastat resulted in a 2-fold increase in α-GAL activity in white blood cells with no serious adverse side effects[514]. In phase 2 trials with FD patients,

oral administration of Migalastat also lead to a 2-fold increase in α-GAL activity in the blood, skin and kidneys, with a concomitant reduction in Gb3 and lyso-Gb3 levels[515–518]. Consequently, Migalastat was approved by the EU in 2006 for the long-term treatment of FD patients with amenable mutations[517,519]. However, only certain α-GAL mutations are amenable to PCT, specifically, mutants which are prone to misfolding are likely to respond to PCT and are termed "responsive". In contrast, mutations which impair the full synthesis of α-GAL or directly alter substrate binding are unlikely to respond to PCT and are termed ''non-responsive"[520]. Fortunately, it is estimated that 65% of mutants associated with classic FD and 90% of mutants associated with non-classic FD are responsive[509,521], therefore, PCT has considerable therapeutic potential for FD.

One of the main benefits of Migalastat is that it is orally available and more easily managed in comparison to ERT[517]. Migalastat also exhibits a broader tissue distribution than the recombinant enzymes. Of note, Migalastat is able to cross the blood-brain barrier and access the CNS[522], making it more appropriate for patients suffering with neurological disease. However, structural analysis of α-GAL mutations has revealed that the most disruptive mutants, associated with severe forms of FD, are on average the least responsive to treatment with Migalastat[521]. The requirement for "responsive" mutants presents one limitation to the use of PCs for the treatment of FD.

1.5.6.3 Combination Therapy

More recently, adjunctive therapies have been explored in an attempt to improve the efficacy of FD treatment. Specifically, a combination of ERT and PCT has been investigated[520]. The use of PCs to stabilise the recombinant enzyme administered in ERT was first investigated for Pompe disease, which results from a deficiency in α-glucosidase (α-GAA)[523]. In initial studies it was found that administering the molecular chaperone N-butyl-deoxynojirimycin (NB-DNJ) with recombinant α-GAA (lysosomal α-glucosidase) results in improved lysosomal trafficking and cellular activity of r-αGAA compared to ERT alone[523]. Subsequently, studies on the combined use of DGJ (Migalastat) and r-αGAL demonstrated a considerable increase in r-αGAL activity in the plasma and skin of FD patients treated with ERT and PCT compared to ERT alone[510,522,524]. In fact, Migalastat was shown to stabilise administered r-αGAL against proteolysis and against pH and temperature dependent denaturation. This improved the circulating half-life of the recombinant enzyme, leading to enhanced uptake by disease affected tissues and a reduction in the EC_{50} value[522]. This suggests that DGJ may permit the use of smaller recombinant enzyme doses and/or

less frequent administration in ERT whilst maintaining the same therapeutic efficacy. Given the positive clinical effects of combination therapy and the high cost of ERT, the use of PCs to supplement r-αGAL has become a recent focus point in the pharmaceutical industry.

1.5.6.4 Substrate Reduction Therapy

As highlighted previously, glucosylceramide synthase (GCS) is a key glycosyltransferase in the biosynthesis of GSLs, including Gb3[77]. Therefore, GCS is a potential therapeutic target for substrate reduction therapy (SRT) for the treatment of FD and other sphingolipidoses[525]. Specifically, inhibition of GCS should block its glycosyltransferase activity and prevent further synthesis of Gb3. In turn, this should allow any endogenous α-GAL to slowly degrade accumulated Gb3 and reduce the storage burden[526]. N-butyldeoxynojirimycin (NB-DNJ, Miglustat, Figure 1.12 (a)) was one of the first compounds identified as a GCS inhibitor[527]. When fed to knockout mice exhibiting Tay-Sachs[527] and Sandhoff disease[163], NB-DNJ partially corrected the lipid accumulation and disease phenotypes, providing proof of concept to targeting GCS in SRT. However, the low inhibitory potency of this NB-DNJ limited its use.

A well-studied and more potent GCS inhibitor is D-threo-1-phenyl-2-decanoylamino-3-morpholino-1-propanol (PDMP)[528–530]. PDMP possess three functional groups, the phenyl group, the fatty acid acylamide group and cyclic amine, Figure 1.20 (a), all of which can be modified to improve specificity towards GCS[531]. For example, substitution of the fatty acid acylamide group for a palmitoyl group and substitution of the morpholino group for a pyrrolidino function yields a significantly more potent inhibitor (D-t-P4)[532], Figure 1.20 (b). Additional substitution at the phenyl group with hydroxy (D-t-pOH-P4) or ethylenedioxy groups (D-t-EtDO-P4), Figure 1.20 (c,d), further improves the inhibitory potency, generating inhibitors that are ~2000 times more potent than PDMP[533]. Whilst the exact mode of action of such PDMP inhibitors is poorly understood, mutagenesis studies on murine GCS have suggested that PDMP may bind to the UDP-glucose binding region of GCS[534]. Regardless of the inhibition mechanism, treatment of FD lymphocytes with such PDMP inhibitors is reported to reduce the levels of GlcCer and Gb3 by up to 70-80%[531]. In fact, FD knockout mice treated with D-t-EtDO-P4 exhibited a concentration-dependent depletion of GlcCer in the liver, spleen and kidneys[535] and a reduction in renal and hepatic Gb3 levels[535]. Unfortunately, D-t-EtDO-P4 had little effect on the level of GlcCer in the brain due to its inability to cross the blood-brain barrier[535]. Nevertheless, these studies demonstrated that FD should be amenable to SRT.

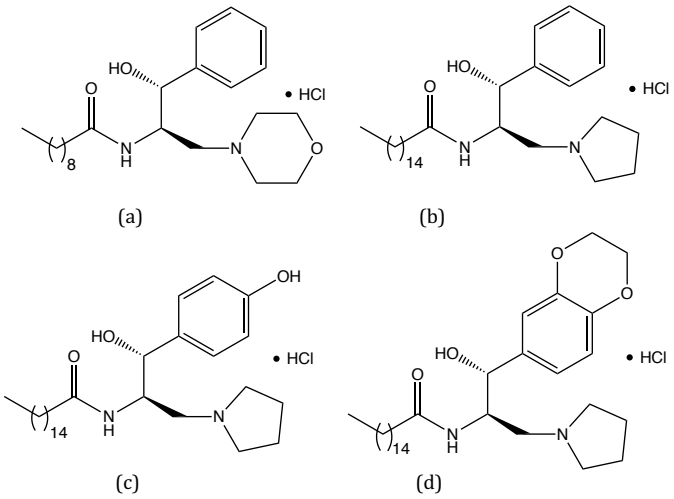


Figure 1.20: Chemical structure of glucosylceramide synthase inhibitors (a) PDMP (b) D-t-P4 (c) D-t-pOH-P4 and (d) D-t-EtDO-P4.

More recently, Ibiglustat (Genz-682452, Venglustat) has proved a potentially useful GCS inhibitor for the treatment of FD[536], Figure 1.21 (b). Indeed, Ibiglustat has been shown to prevent the accumulation of Gb3 in cardiomyocytes of FD patients and significantly reduce tissue levels of Gb3 and lyso-Gb3 in Fabry mice[536,537]. Furthermore, a greater reduction in glycolipid accumulation has been observed in mice treated with a combination of Ibiglustat and r-αGAL compared to r-αGAL alone[536], suggesting Ibiglustat may be suitable for combination therapy. This can be rationalised by the different biodistribution of oral Ibiglustat in comparison to intravenous r-αGAL. Notably, Ibiglustat has a more profound effect in lowering renal and intestinal Gb3 levels and is able to cross the blood-brain barrier[536]. Therefore, Ibiglustat may prove effective in combination with ERT to alleviate neurological symptoms that cannot be treated by ERT alone.

Lucerastat (N-butyldeoxygalactonojirimycin) is a soluble, low molecular weight iminosugar inhibitor of GCS[538,539], which is currently the most promising candidate for the treatment of FD by SRT, Figure 1.21. Lucerstat has proved a selective inhibitor of GCS at physiologically relevant pH values, showing no off-target inhibition with other related enzymes, such as lysosomal glucocerebrosidase (GBA)[539,540]. Importantly, Lucerastat is able to cross the blood-brain barrier and has been shown to reduce the accumulation of GM2 gangliosides in the brain of Sandhoff mouse models[541] and deplete the storage of Gb3 in the ganglia of Fabry mice[538]. Lucerastat was first tested in healthy individuals in a phase 1 clinical trial, which demonstrated good safety, tolerability and pharmacokinetic properties[538]. It has also been tested in a phase 1 open-label exploratory clinical trial with FD patients who were being treated by ERT[541]. In this trial, oral administration of Lucerastat in combination with ERT resulted in a significant reduction in plasma Gb3 and lyso-Gb3 levels compared to ERT

alone. Importantly Lucerastat was also well-tolerated by the FD patients[541]. A phase 3 double-blind clinical trial is currently underway (NCT03425539) to assess the safety and efficacy of Lucerastat monotherapy in FD individuals. This study is expected to end in 2022 and holds considerable promise for the development of a reliable SRT for FD.

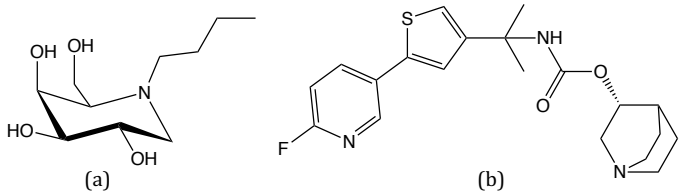


Figure 1.21: Chemical structure of glucosylcermide synthase inhibitors (a) Lucerastat (N-butyldeoxygalactonojirimycin) (b) Ibiglustat (Genz-682452).

1.5.6.5 Therapies for the Future

As highlighted in *section 1.4.8.4,* there is considerable interest in developing gene therapies for LSDs. Perhaps the most appealing aspect of gene therapy is that, in theory, it is a one-time, curative approach which circumvents the need for lifelong treatment[381]. Indeed, current therapeutic strategies for FD are incredibly costly, therefore a single gene therapy treatment may also alleviate the financial burden currently associated with FD therapy.

In a recent 2020 study, 5 male FD patients underwent a stem cell transplant with hematopoietic cells that were transduced with a lentiviral vector encoding the human α-GAL enzyme[542]. The primary aim of this study was to evaluate the safety of stem cell transplantation and determine whether α-GAL activity could be enhanced for phenotypic benefit. Promisingly, α-GAL activity was detected 6 days post-transplantation and whilst the plasma activity decreased over time, it remained above the levels typically associated with FD over the course of the study (from Jan 2017 to Feb 2020)[542]. Additionally, all participants were eligible to discontinue their ERT owing to sufficient α-GAL activity. Unfortunately, changes in Gb3 and lyso-Gb3 levels were less convincing, nevertheless, these levels remained generally stable and no significant safety or toxicity events were identified[542]. This ground-breaking study highlights the considerable promise lentivirus-mediated gene therapy may hold for the future of FD therapy. In fact, small scale success such as these solidify the viability of gene therapy treatments for LSDs and other congenital disorders.

The future of gene therapy for LSDs is uncertain but it is clear that new experimental vectors with enhanced efficiency and specificity will be required[381]. Additionally, a deeper understanding of downstream effects on the immune system will be necessary to determine the safety of such an approach to balance this with the potential benefits[381]

1.6 Activity-Based Protein Profiling

In order to improve both diagnostic and therapeutic technologies for LSDs, a detailed understanding of how the activity of certain enzymes influence the disease phenotype is required. Whilst significant advances have been made in proteomic and genomic approaches[543], the complexity of the human proteome limits the amount of information that can be obtained from these techniques[544]. Specifically, genomic and proteomic approaches attempt to determine the importance of a protein to a given biological process through changes in protein expression and abundance[543]. Given protein activities tend to be regulated by co- and post-translational modifications, protein levels often do not directly correlate with protein activity[543,545], therefore, these methods provide limited information on the impact of protein activity in disease states. Fortunately, the technology of activity-based protein profiling (ABPP) has expanded rapidly over the past few decades and has proved indispensable in improving our understanding of the physiological roles of proteins in health and disease[544,546]. Here, an overview of ABPP will be given; details regarding the development of ABPs for lysosomal glycosidases and their application in the study of LSDs will be covered in **Chapters 2** and **3.**

1.6.1 Overview of Activity-Based Protein Profiling

ABPP offers an attractive approach for the simplification of complex proteomes by allowing the activity of specific enzymes to be profiled within native cellular environments[545,547]. Specifically, this approach makes use of small chemical probes which are designed to react with an active site residue of a target enzyme in a covalent manner to permit labelling of only catalytically active enzyme[545,548,549]. These molecular probes, termed activity-based probes (ABPs), comprise of three fundamental components; a reactive "warhead" group which reacts with a catalytic residue of the target enzyme, a recognition element or linker region which serves as a spacer between the warhead and the tag (but may also improve the selectivity of the probe) and lastly the reporter group such as a fluorescent tag, affinity tag, radioisotope or ligation handle[545,547,550,551], Figure 1.22.

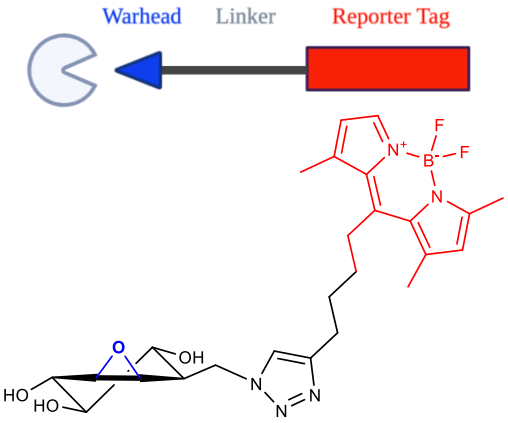


Figure 1.22: (a) Generic activity-based probe structure: reactive warhead (blue), linker (grey) and reporter tag (red). (b) Example of cyclophellitol-based ABP where the epoxide constitutes the warhead, the triazole-amide forms the linker and Cy5 tag is the reporter.

Broadly speaking, any enzyme which forms a covalent enzyme-substrate complex during its catalytic mechanism may be targeted by ABPs[547,552]. Specifically, this covalent intermediate can be exploited for the design of ABPs which mimic the substrate and react in a similar manner with a catalytic residue to form a covalent and irreversible enzyme-ABP complex[547]. However, one of the biggest challenges is selecting an appropriate scaffold onto which the warhead and reporter group is to be installed. Typically, natural substrates or inhibitors are used as the initial scaffold[546,547]; the most notable examples being conduritol B epoxide (CBE)[553] and cyclophellitol (CP)[554], Figure 1.23 (a,b). These compounds are mechanism-based β-glucosidase inhibitors which employ an electrophilic epoxide trap to covalently modify the enzymatic catalytic nucleophile[547], Figure 1.23 (c). These inhibitors have been used as scaffolds for the development of numerous glycosidase ABPs by alterations to the hydroxyl stereochemistry and functionalisation with suitable reporter tags[277,555–557].

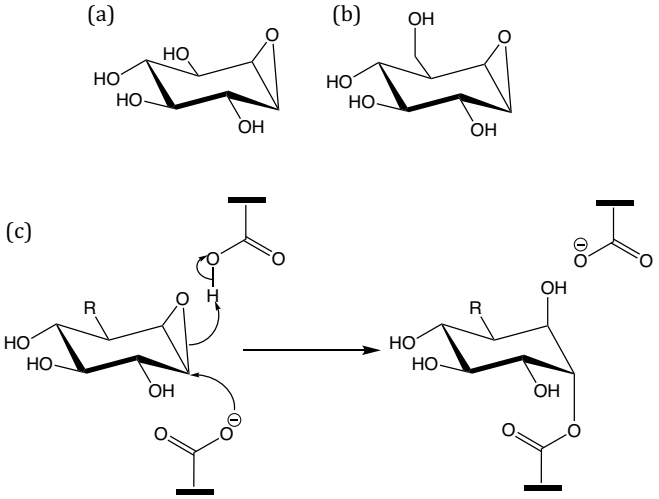


Figure 1.23: Chemical structure of (a) conduritol B epoxide (CBE) and (b) cyclophellitol (c) Generic inhibition of a mechanism of β-glycosidase by CBE or cyclophellitol.

Depending on the purpose of the experiment, a target enzyme may be isolated and purified by equipping the ABP with an affinity tag or the enzyme may be directly visualised and its activity profiled using a fluorescent tag[549,558,559]. Alternatively, structural analysis of the ABP in complex with the enzyme of interest may be performed to guide the development of more potent and selective probes[560–562], Figure 1.24. By selectively labelling only catalytically active enzyme, ABPP holds advantage over proteomic techniques by permitting the differentiation between active and non-active enzyme and allowing associations between enzyme activity and disease phenotypes to be deduced[555,558,563].

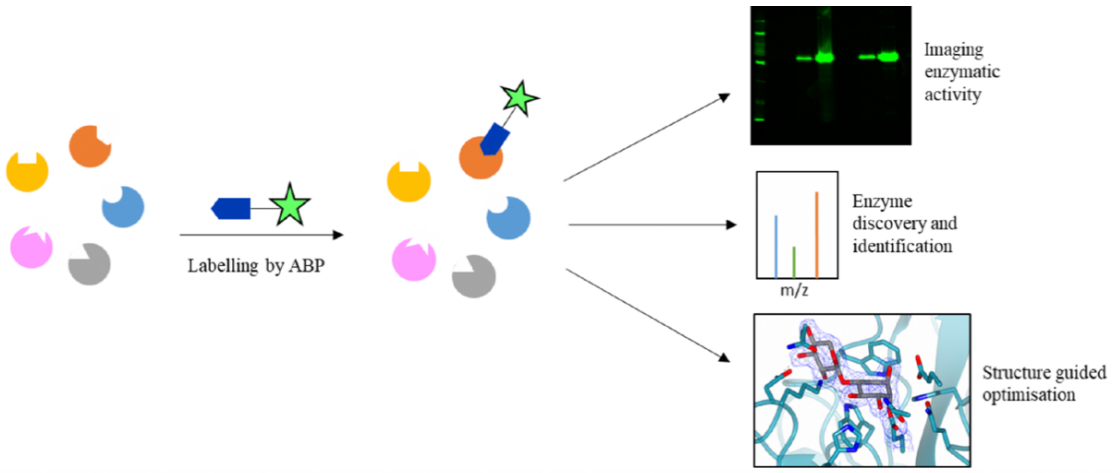


Figure 1.24: General workflow of activity-based protein profiling. Use of activity-based probes to selectively label active enzymes in complex mixtures.

1.6.2 Applications of Activity-Based Protein Profiling

Initially, ABPP focused on serine and cysteine proteases and hydrolases[564,565], however, advances in the development of highly class-specific probes has led to their use in a wider range of enzyme classes. In fact, ABPs have been developed for glycoside hydrolases[546,566], kinases[567], phosphatases[568], methyltransferases[569,570] and ubiquitin ligases[571]. The expansion of ABPs to the study of GHs is arguably the most important advance in the field of ABPP and has permitted the study of many glycosidases, ranging from human lysosomal enzymes[277,555,558] to plant biomass degrading enzymes[572,573]. The ability to generate both broad-spectrum and enzyme specific ABPs for GHs has resulted in ABPP finding applications in biomedical diagnostics and imaging, inhibitor and target discovery and biotechnology[547]. For example, β-glucuronidase ABPs have been used to detect and monitor the overexpression of heparanase in human tissue[574], which has the potential to be used as a diagnostic tool to track abnormal heparanase regulation in cancer development and inflammation. Additionally, fluorescent glycosidase ABPs have been utilised for high-throughput screening of compound libraries to identify novel inhibitors and therapeutic

compounds[575]. Specifically, this approach makes use of fluorescence polarisation ABPP assays (FluoPol-ABPP) in which fluorescent ABPs are excited with plane polarised light and the fluorescence emission is monitored[575]. The fluorescence polarisation signal depends on whether the ABPs are bound to the enzyme, resulting in high signal, or if they are free in solution, which results in a low signal. Consequently, when an enzyme is incubated with a mixture of ABPs and potential inhibitors, any compounds which effectively inhibit the enzyme will reduce ABP binding, leading to a loss in fluorescence polarisation signal and hit identification[575]. This approach has already been employed to identify potent inhibitors of non-lysosomal glucocerebrosidase (GBA2)[575]. In addition to biomedical applications, ABPP is becoming increasingly popular in biotechnology, with fruitful applications in the study of bacterial and fungal secretomes for the identification of biomass degrading enzymes[572,576].

1.7 This Work

In summary, lysosomal storage disorders are rare but debilitating metabolic disorders which can be fatal if not detected and treated effectively. Despite advances in the field, diagnostics and therapeutics for these disorders remain sub-optimal, with many patients receiving a late diagnosis and treatments which are not best suited to their disease genotype and phenotype. Improving both diagnostic and therapeutic strategies for these disorders requires extensive knowledge of the underlying genomic and enzymatic defects and the interplay of these in controlling disease phenotype. As discussed, ABPP is a rapidly evolving technology which offers the ability to detect and profile enzymatic activity *in vitro, in situ* and *in vivo*, therefore it is ideally suited to such a problem.

1.7.1 Aims of this Work

In collaboration with Prof. Hermen Overkleeft at Leiden University, the overarching aim of this work was to aid in the development of activity-based probes (ABPs) designed for human β-glucocerebrosidase (GBA) and α-galactosidase A (α-GAL), the lysosomal enzymes which underpin Gaucher disease (GD) and Fabry disease (FD) respectively. Specifically, this work aimed to use protein x-ray crystallography to analyse these ABPs at the 3D level and provide structural data to inform the development of more potent and selective compounds.

Through separate collaborations with the Overkleeft Lab (Leiden University) and the Vocadlo Lab (Simon Fraser University), this work further aimed to expand our structural approach to the study of novel inhibitors and molecular chaperones for these enzymes.

Specifically, it was hoped that structure-based optimisation supported by biochemical methods could be used to develop more potent and selective GBA and α-GAL compounds for use in applications such as chaperone therapy and in the generation of animal models. To support such work, a source of active GBA enzyme would be required. Unfortunately, our supply of commercial GBA (Imiglucerase, Cerezyme®, Sanofi Genzyme) was severely limited and prior to this work the reliable over production of GBA had only been achieved in industrial labs by unknown and irreproducible means. Consequently, this work also aimed to establish an insect-baculoviral expression system for the production of non-clinical GBA.

The aims of this work are summarised as follows:

- Structurally and biochemically analyse a range of *galacto*-configured cyclophellitol inhibitors developed for human α-GAL and investigate any pharmacological chaperone potential for FD– Collaboration with Overkleeft Lab (Leiden University) – **Chapter 2.**

- Guide the structure-based development of more potent activity-based probes and inhibitors for GBA - Collaboration with Overkleeft Lab (Leiden University) – **Chapter 3.**

- Establish a reliable in-house expression system for the production of human GBA to support continued efforts in developing novel GBA active compounds– **Chapter 4.**

- Structurally investigate a novel class of cyclic allylic carbasugar inhibitors which show pharmacological chaperone potential towards GBA– Collaboration with Vocadlo Lab (SFU) – **Chapter 5**

To date, five publications and one patent have arisen from this work:

1. R. J. Rowland, Y. Chen, I. Breen, L. Wu, W. Offen, T. J. Beenakker, Q. Su, A. M. C. H. can den Nieuwendijk, M. Artola, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, *Design, Synthesis and Structural Analysis of Glucocerebrosidase Imaging Agents,* Chem. Eur. J., **2021**, 27, 16377-16388.

2. R. J. Rowland, L. Wu, F. Liu, G. J. Davies, *A baculoviral system for the production of human β-glucocerebrosidase enables atomic resolution analysis*, Acta Cryst. D, **2020**, 76, 565-580.

3. M. Artola, C. Hedberg, R. J. Rowland, L. Raich, K. Kytidou, L. Wu, A. Schaaf, M. J. Ferraz, G. A. van der Marel, J. D. C. Codée, C. Rovira, J. M. F. G. Aerts, G. J. Davies and H. S. Overkleeft, *α-D-Gal-cyclophellitol cyclosulfamidate is a Michaelis complex analog that stabilizes therapeutic lysosomal α-galactosidase A in Fabry disease,* Chem. Sci., **2019**, 10, 9233-9243.

4. M. Artola, C-L. Kuo, L. T. Lelieveld, R. J. Rowland, G. A. van der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *Functionalized Cyclophellitols Are Selective Glucocerebrosidase Inhibitors and Induce a Bona Fide Neuropathic Gaucher Model in Zebrafish*, J. Am. Chem. Soc., **2019**, 141, 4214-4218.

5. S. P. Schröder, C. de Boer, N. G. S. McGregor, R. J. Rowland, O. Moroz, E. Blagova, J. Reijngoud, M. Arentshorst, D. Osborn, M. D. Morant, E. Abbate, M. A. Stringer, K. B. R. M. Krogh, L. Raich, C. Rovira, J-G. Berrin, G. P. van Wezel, A. F. J. Ram, B. I. Florea, G. A. van der Marel, J. D. C. Codée, K. S. Wilson, L. Wu, G. J. Davies and H.S. Overkleeft, *Dynamic and Functional Profiling of Xylanase-Degrading Enzymes in Aspergillus Secretomes Using Activity-Based Probes,* ACS Cent. Sci., **2019**, 5, 1067-1078.

6. H. S. Overkleeft, G. J. Davies, J. M. F. G. Aerts, M. Artola, R. J. Rowland, L. Wu; *Pharmacological Chaperones for Glycosidase Treatment Therapy*; WO/2020/046132A1; 2020. Also published as NL2021840.

Chapter 2: Inhibitors for α-Galactosidase A: Pharmacological Chaperones for Fabry Disease

2.1 Abstract

Fabry disease (FD) is a lysosomal storage disorder caused by inherited deficiencies in lysosomal α-galactosidase A (α-GAL). Current treatments involve enzyme replacement therapy and pharmacological chaperone therapy; however, both require an early diagnosis and a means of monitoring treatment efficacy. In this regard, the Overkleeft lab (Leiden University) synthesised a range of α-*galacto* configured cyclophellitol epoxide and aziridine activity-based probes (ABPs) to profile α-GAL activity in native cellular environments. More recently, a new class of *galacto*-configured α-1,6-cyclophellitol cyclosulfate inhibitors were also developed. Here, the first ever structural analyses of these α-*galacto*-cyclophellitol inactivators on the structure of recombinant human α-GAL is reported.

The α-*galacto*-configured cyclophellitol epoxide and aziridine compounds were shown to inhibit r-αGAL in a mechanism-based manner, covalently modifying the catalytic nucleophile through reaction with epoxide or aziridine warhead. Interestingly, the *N*-acyl functionalised aziridine inhibited r-αGAL in the same covalent manner; whilst the *N*-alkyl analogue bound non-covalently in the active site. This may provide some structural rationale for the reduced labelling efficiency reported for *N*-alkyl aziridine ABPs. Importantly, the co-crystal structure obtained with the cyclosulfate inhibitor, revealed covalent inhibition of r-αGAL through the anticipated mechanism in which the cyclosulfate moiety covalently modifies the catalytic nucleophile; demonstrating the potential to expand the suite of cyclophellitol-based inhibitors by use of alternative electrophilic warheads. In contrast, the cyclosulfamidate analogue behaves as a Michaelis Complex mimic, binding non-covalently in the active site of r-αGAL. In preliminary *in vitro* thermal stability studies, this cyclosulfamidate inhibitor was found to stabilise r-αGAL against thermal denaturation, and our collaborators in Leiden later demonstrated a chaperoning effect *in situ* in various FD fibroblasts. Therefore, this *galacto*-configured cyclophellitol cyclosulfamidate inhibitor may show pharmacological chaperone potential towards α-GAL for the treatment of FD.

*Some of this work is published in (**1**) M. Artola, C. Hedberg, R. J. Rowland, L. Raich, K. Kytidou, L. Wu, A. Schaaf, M. J. Ferraz, G. A. van der Marel, J. D. C. Codee, C. Rovira, J. M. F. G. Aerts, G. J. Davies, H. S. Overkleeft, *Chem. Sci.,* 2019, **10**, 9233-9243 and (**2**) *NL Pat.*, WO20046132A1, 2020.

2.2 Introduction

2.2.1 α-Galactosidase A

The human enzyme α-galactosidase A (α-GAL, E.C. 3.2.1.22) is a lysosomal glycoside hydrolase belonging to the GH27 family of retaining α-galactosidases. α-GAL plays a key role in glycosphingolipid metabolism by catalysing the removal of terminal α-linked galactosyl moieties from glycoconjugates[577,578]. Specifically, α-GAL is responsible for hydrolysing globotriaosylceramide (Gb3), the primary glycosphingolipid which underpins the pathogenesis of Fabry disease, Figure 2.1.

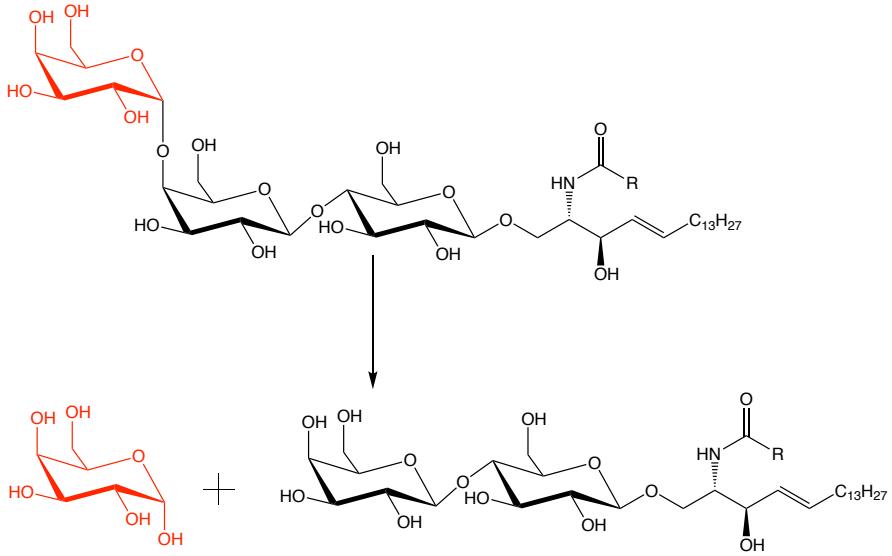


Figure 2.1: Hydrolysis of globotriaosylceramide (Gb3) by α-GAL to yield lactosylceramide (black) and galactose (red) with retention of α-anomeric stereochemistry.

α-GAL is encoded by the *GLA* gene located on the long arm of the X-chromosome (Xq22.1)[579]. It is primarily synthesised in the endoplasmic reticulum as a 55 kDa precursor with a 31 amino acid signal sequence that is cleaved during transit through the ER[432,580]. The precursor is transferred to the Golgi apparatus where the oligosaccharides undergo modification by the addition of mannose-6-phosphate (M6P) residues[581,433]. Consequently the enzyme binds to the M6P receptor and is trafficked to the lysosome via the M6P pathway[138,440,502]. Once in the pre-lysosomal compartment, the acidic pH of the lysosome induces dissociation of the enzyme from the M6P receptor to form a functional dimer[580].

2.2.1.1 CAZy Glycoside Hydrolase 27 (GH27) Family

The GH27 family belongs to the GH-D CAZy clan along with 2 other GH families (GH31 and GH36) which share the characteristic $(\beta/\alpha)_8$ fold. Currently, the GH27 family contains

enzymes from archaea, bacteria, eukaryote and viruses with known α-galactosidase, α-*N*-acetylgalactosaminidase, isomalto-dextranase, β-L-arabinopyranosidase and galactan:galactan galactosyltransferase activity. Perhaps the most notable enzymes of this family are human α-GAL and α-*N*-acetylgalactosaminidase, which are associated with Fabry disease and Schindler disease respectively[441]. Members of the GH27 family are *exo*-acting enzymes, which cleave terminal α-galactose moieties from the non-reducing end of their substrates[582]. These enzymes are also anomeric-configuration retaining enzymes which operate through a double-displacement catalytic mechanism[577,583].

2.2.1.2 Catalytic Mechanism

Retention of α-anomeric stereochemistry by GH27 members was first demonstrated in 1999 through proton NMR studies of the hydrolysis of *p*-nitrophenyl α-galactopyranoside by an α-galactosidase isolated from the white-rot fungus *Phanerochaete chrysosporium*[583]. The catalytic nucleophile of this enzyme was later identified by enzymatic trapping with fluorinated glycosides followed by pepsin digestion and mass spectrometry analysis[584,585]. In the case of human α-GAL, a covalently trapped 2-deoxy-2,2-difluorogalactosyl-enzyme intermediate revealed the catalytic nucleophile to be Asp170 and the general acid/base residue to be Asp231. Consequently, α-GAL employs the Koshland double displacement mechanism, using catalytic residues Asp170 and Asp231, to hydrolyse its substrates with retention of α-anomeric stereochemistry, Figure 2.2.

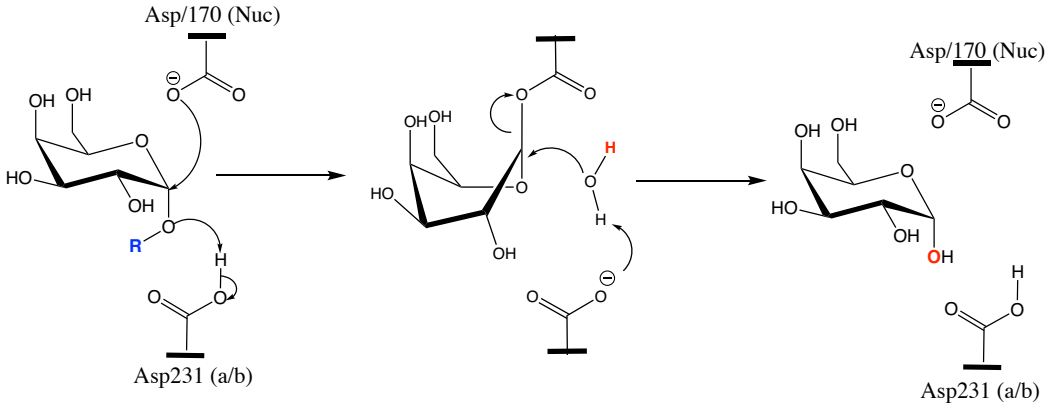


Figure 2.2: Koshland double-displacement hydrolysis mechanism of retaining α-GAL which proceeds via a covalent glycosyl-enzyme intermediate. Asp170 is the active site nucleophile (Nuc) and Asp231 behaves as the general acid/base residue (a/b).

Throughout the double displacement mechanism the sugar substrate undergoes substantial changes in conformation to satisfy the stereoelectronic and orbital overlap requirements of glycoside hydrolysis[370]. In 2010, a set of α-GAL structures were published including a native substrate-bound Michaelis complex, a trapped covalent 2-deoxy-2,2-difluorogalactosyl-

enzyme intermediate and product-bound structure[577]. The conformation adopted by the sugar substrate in each of these structures provided the first complete analysis of substrate distortion throughout the enzymatic reaction. According to a Michaelis complex→[Transition state]‡→Covalent intermediate reaction coordinate, GH27 enzymes employ a $^4C_1$→$[^4H_3]$‡→$^1S_3$ conformational itinerary[586], Figure 2.3.

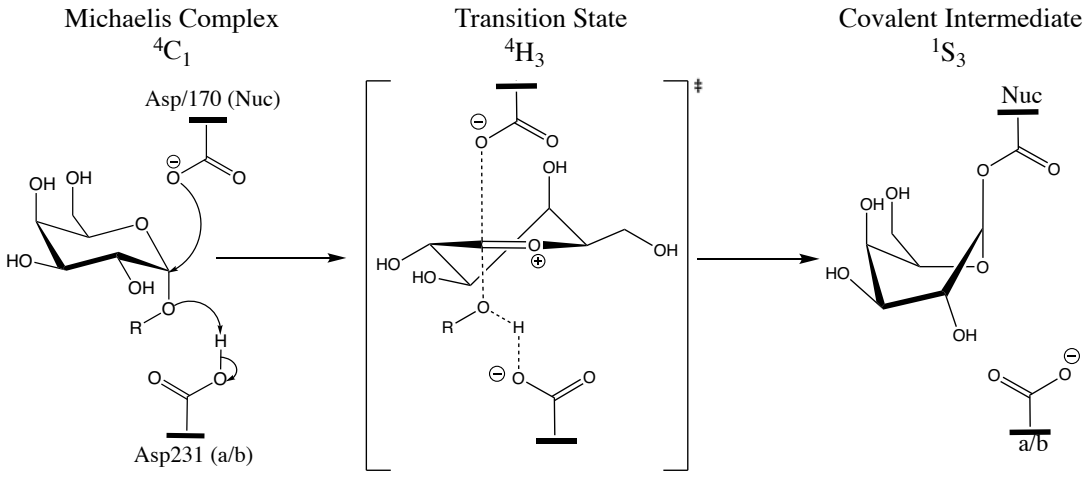


Figure 2.3: Conformational reaction itinerary of retaining α-galactosidases. Upon binding in the Michaelis complex the substrate adopts a $^4C_1$ conformation to ensure the aglycon leaving group is axially positioned for in-line nucleophilic attack. On nucleophilic attack, a high energy transition state is formed in $^4H_3$ conformation allowing for partial oxocarbenium double bond formation between O5 and C1. Subsequently, a covalent glycosyl-enzyme intermediate in $^1S_3$ conformation is formed.

2.2.1.3 Structure of Human α-Galactosidase A

X-ray crystallography analysis of GH27 members was initially hindered by the high levels of protein glycosylation. However, in 2002, the 3D structure of chicken *N*-acetylgalactosaminidase (α-NAGAL) was solved, providing the first ever structure of a GH27 enzyme[587]. Subsequently, structures of the rice α-galactosidase[588], human α-galactosidase A[440], and fungal Hypocrea jecorina α-galactosidase[589] were solved.

The first x-ray structure of human α-GAL was solved in 2004 (PDB 1R46)[440] using a commercially available recombinant formulation, Replagal ™ (Agalisdase Alfa, Shire Plc). The structure was solved to 3.25 Å by molecular replacement using a homology model from the crystal structure of chicken *N*-acetylgalactosaminidase[587], which shares 51% sequence identity[587]. The refined structure revealed human α-GAL to be a homodimeric glycoprotein, comprising of two 398-residue monomers[440], Figure 2.4 (a). The two monomers packed together in the asymmetric unit to form a broad 75 Å^2 dimer interface formed by 30-

residues from each monomer, involving loops β_1-α_1, β_6-α_6, β_7-α_7, β_8-α_8, β_{11}-β_{12} and β_{15}-β_{16}[440], Figure 2.4. Each monomer consisted of two domains; a characteristic $(\beta/\alpha)_8$ TIM domain and a C-terminal domain comprised of eight antiparallel β-strands arranged in a β-sandwich[440], Figure 2.4. (b). The active site was located at the centre of the TIM barrel domain which is characteristic of GH-D clan members. Specifically, the active site was formed by the side chains of 15 residues from loops β_1-α_1, β_2-α_2, β_3-α_3, β_4-α_4, β_5-α_5, β_6-α_6 and β_7-α_7. The distance between the two catalytic carboxylate residues (calculated as the average distance between the four oxygen atoms) was reported to be 7.3 Å, which is considerably longer than the average distance (~4.8 Å) observed for retaining α-glycosidases[590]. However, members of the GH27 family exhibit some of the largest distances seen for retaining glycosidases, suggesting the distance between catalytic carboxylate residues could vary more than originally proposed.

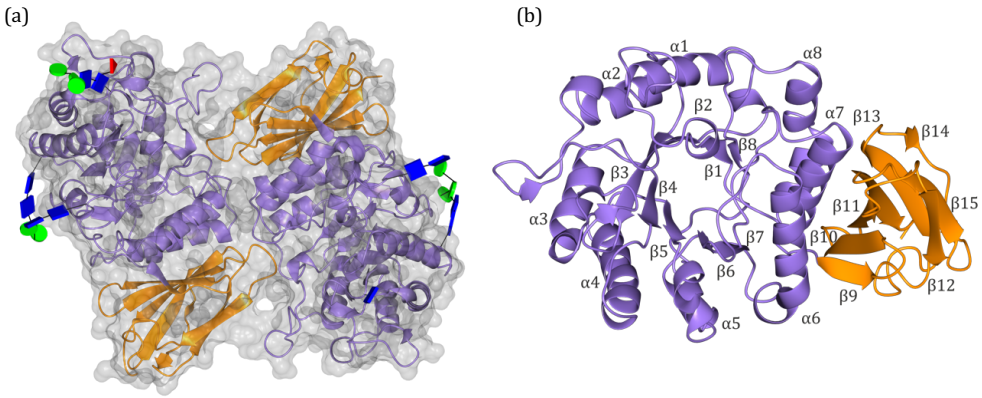


Figure 2.4: First structure of human α-GAL solved by Garboczi *et al.* (2004)[440]. (a) Surface and ribbon diagram of α-GAL dimer. *N*-linked glycans illustrated using Glycoblocks[591]. (b) Ribbon diagram of α-GAL monomer with TIM barrel domain I (residues 32-330) depicted in purple and domain II (residues 331-429) consisting of an eight stranded β-sandwich in orange. Figure created in CCP4mg[94] using deposited PDB 1R46 coordinates.

To provide insight into substrate specificity, a crystal structure of α-GAL in complex with α-galactose was also obtained[440]. This structure revealed that α-GAL makes hydrogen bonding interactions to each functional group of the α-galactose ligand, demonstrating high specificity. However, little specificity was observed for the aglycon portion of the substrate where the active site cleft extends to a broad opening at the dimer interface.

Overall, the known structures of GH27 members are highly conserved, sharing a common tertiary structure. Indeed, the active sites are extremely well conserved, exhibiting a pocket shape that is consistent with the *exo*-mode of action of the GH27 family[440,587–589]. The only dissimilar residues, which tend to reside in the β_5-α_5 loop, are thought to alter substrate specificity by defining a binding site for a hydroxyl group or *N*-acetyl substituent at the C2

position of the galactose ring. To date, 21 crystal structures of human α-GAL have been deposited in the PDB, Table 2.1.

2.2.1.4 Glycosylation of Human α-Galactosidase A

In 1992, Desnick and co-workers achieved overexpression of human α-GAL in Chinese Hamster Ovary (CHO) cells[592], permitting biochemical analysis of α-GAL glycosylation. The purified enzyme was treated with a number of glycosidases, including *N*-glycanase, *O*-glycanase and endoglycosidase-F, to reveal that α-GAL is purely *N*-glycosylated with four putative *N*-glycosylation sites (Asn139, Asn192, Asn215 and Asn408). Subsequent studies involving exoglycosidase digestion and metabolic [^{3}H] mannose radiolabelling, indicated that only the first three glycosylation sites are occupied[593]. The fourth putative (Asn408) site was shown to be unoccupied, presumably due to the fact that this glycosylation site contains the amino acid sequence Asn-Pro-Thr, which is not usually recognised by the carbohydrate attachment machinery. This glycosylation profile was later observed in the first x-ray structure of human α-GAL, in which each monomer exhibited three occupied N-linked glycosylation sites at Asn139, Asn192, and Asn215[440], Figure 2.5. Further studies into the glycosylation of human α-GAL revealed that occupancy of the Asn215 site is vital to enzymatic activity and efficient lysosomal trafficking[593]. Specifically, site directed mutagenesis and immunofluorescence studies demonstrated a complete loss of enzymatic activity on elimination of the Asn215 site in combination with either site 1 (Asn139) or site 2 (Asn192). Furthermore, mutation of the Asn215 site resulted in the synthesis of defective polypeptides which were localized to the ER and presumably degraded[593]. In combination with amino acid sequence analysis, it was concluded that occupation of the Asn215 glycosylation site is required to prevent aggregation of a nearby hydrophobic region, thereby enhancing enzyme stability, solubility and subsequent lysosomal trafficking[593].

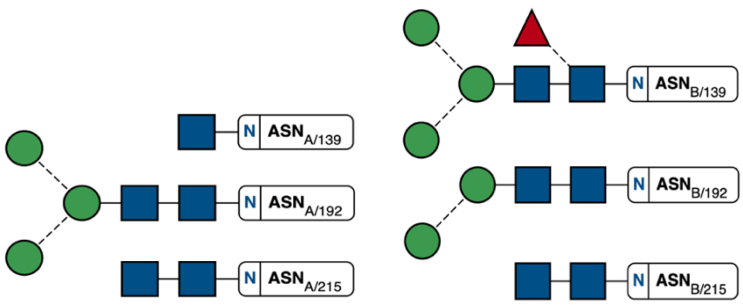


Figure 2.5: Glycosylation profile of first human α-GAL crystal structure (PDB 1R46)[440] with occupation of Asn139, Asn192 and Asn215 *N*-glycosylation sites. Figure generated in Privateer[594] using the Glycoblock format[591].

Table 2.1: Information for all human α-GAL structures deposited in the PDB to date. CHO = Chinese Hamster Ovary

PDB	Source	Drug Name	Ligand	Res (Å)	Reference
1R46	Human	Replagal	None	3.25	Garman et al (2004)[440]
1R47	Human	Replagal	α-Galactose	3.45	Garman et al (2004)[440]
3GXN	CHO	Fabrazyme®	None	3.02	Lieberman et al (2009)[595]
3GXP	CHO	Fabrazyme®	α-Galactose	2.20	Lieberman et al (2009)[595]
3GXT	CHO	Fabrazyme®	1-deoxygalactonijirimycin	2.70	Lieberman et al (2009)[595]
3HG2	Insect	-	None	2.30	Guce et al (2010)[577]
3HG3	Insect	-	Melibiose	1.90	Guce et al (2010)[577]
3HG4	Insect	-	2',4',6'-trinitrophenyl-2-deoxy-2,2-difluoro-α-D-galactopyranoside	2.30	Guce et al (2010)[577]
3HG5	Insect	-	α-Galactose	2.30	Guce et al (2004)[577]
3LX9	Insect	-	*N*-Acetylgalactosamine (GALNAc)	2.04	Tomasic et al (2004)[596]
3LXA	Insect	-	α-Galactose	3.04	Tomasic et al (2004)[596]
3LXB	Insect	-	Glycerol	2.85	Tomasic et al (2004)[596]
3LXC	Insect	-	Glycerol	2.35	Tomasic et al (2004)[596]
3S5Y	CHO	Fabrazyme®	1-deoxygalactonijirimycin	2.40	Guce et al (2011)[597]
3S5Z	CHO	Fabrazyme®	α-Galactose	2.00	Guce et al (2011)[597]
3TV8	CHO	Fabrazyme®	1-deoxygalactonijirimycin	2.64	Guce et al (2011)[597]
4NXS	CHO	-	(2R,3S,4R,5S)-*N*-(4-fluorophenyl)-3,4,5-trihydroxy-2-(hydroxymethyl)piperidine-1-carbothioamide	2.55	Yu et al (2014)[598]
6IBK	CHO	Fabrazyme®	α-galactose configured cyclosulfamidate	1.99	Artola et al (2019)[561]**(This Thesis)**
6IBM	CHO	Fabrazyme®	α-galactose configured cyclosulfate	2.07	Artola et al (2019)[561] **(This Thesis)**
6IBR	CHO	Fabrazyme®	α-galactose configured cyclophellitol epoxide	2.02	Artola et al (2019)[561] **(This Thesis)**
6IBT	CHO	Fabrazyme®	α-galactose configured aziridine	2.04	Artola et al (2019)[561] **(This Thesis)**

*Table updated 26/08/21

2.2.2 Activity-Based Probes Retaining α-Galactosidases

Activity based protein profiling has found application in the study of lysosomal storage disorders as a unique research tool for investigating the activity of lysosomal enzymes. Initial efforts focused on the development of inhibitors and activity based probes (ABPs) for retaining β-glucosidases, such as β-glucocerebrosidase which is implicated in Gaucher disease[599–601]. The success of the tagged-cyclophellitol β-glucosidase probes pioneered by the Overkleeft lab[556,599,600], led to the development of ABPs for other glycosidases, namely for retaining α-galactosidases such as lysosomal α-GAL[602].

2.2.2.1 Galacto-Configured Cyclophellitol Activity-Based Probes

To initiate the development of cyclophellitol-based ABPs for retaining α-galactosidases, Overkleeft and co-workers first synthesised cyclophellitol epoxide and cyclophellitol aziridine isomers with α-galactopyranose configuration[602]. These α-*galacto*-configured cyclophellitol isomers covalently inhibit retaining α-galactosidases through the mechanism-based mode of action depicted in Figure 2.6 (a), in which the enzymatic nucleophile attacks the epoxide or aziridine warhead through trans-diaxial ring opening to form a covalent enzyme-inhibitor complex. These inhibitors are also thought to benefit from mimicking the half-chair ($^4H_3$) conformation of the transition state of the retaining α-galactosidase reaction itinerary. Specifically, this $^4H_3$ conformation places the electrophilic warhead in line for nucleophilic attack.

The *galacto*-configured cyclophellitol epoxide was found to be a micromolar inhibitor of α-GAL (apparent IC_{50} = 13 μM), whilst the aziridine analogue proved to be a more potent nanomolar inhibitor (apparent IC_{50} = 40 nM). Subsequently, these *galacto*-configured cyclophellitol isomers were used for ABP development by the attachment of suitable reporter groups at the C6 position of α-*galacto*-cyclophellitol epoxides or at the aziridine nitrogen of α-*galacto*-cyclophellitol aziridines[556,558,603], Figure 2.6. Of note, *N*-tagged α-*galacto*-cyclophellitol aziridine probes have proved selective, nanomolar inhibitors of α-GAL, with a 10^3-10^4-fold higher potency than the epoxide derivatives (apparent IC_{50} = 2-3 nM for *N*-tagged aziridines vs IC_{50} = 13 μM for epoxide)[556,558]. Specifically, fluorescently tagged *N*-acyl aziridine ABPs have been used to label endogenous human α-GAL and *N*-acetylgalactosaminidase (α-GAL B) activity in wild-type fibroblasts with good specificity and no cross-reactivity with retaining β-glucosidases[556,558]. However, the *N*-acyl moiety used to introduce functionality is prone to hydrolysis, putting strain on the synthesis,

purification and handling of these ABPs. Consequently, *N*-alkyl *galacto*-cyclophellitols were synthesised, which have proved more stable in mildly acidic and basic conditions, making them much easier to synthesise and handle[556]. Such *N*-alkyl probes are nanomolar inhibitors of α-GAL (apparent IC_{50} = 3 nM[556]) and have been used to directly visualise and profile α-GAL activity both *in vitro* and *in situ* in wild-type fibroblasts[556] . Although *N*-acyl *galacto*-configured cyclophellitol aziridine ABPs exhibit a higher labelling efficiency, the improved stability of *N*-alkyl aziridines makes them viable ABPs for retaining α-galactosidases[556].

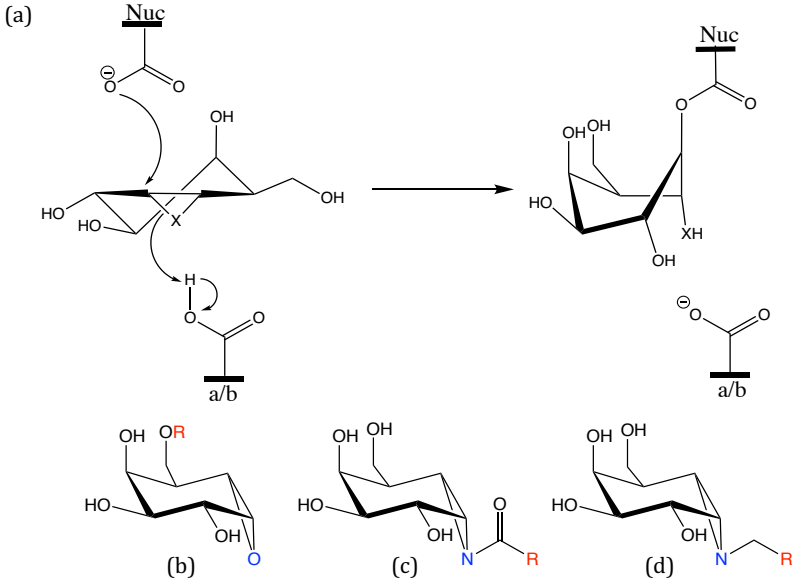


Figure 2.6: (a) General mode of action of α-*galacto*-configured cyclophellitol epoxide (X = O) and aziridine probes (X = NH or NR) for retaining α-galactosidases. (b) α-*galacto*-cyclophellitol epoxide functionalised at C6 position (c) *N*-acyl functionalised and (d) *N*-alkyl functionalised α-*galacto*-cyclophellitol aziridine where R = reporter group.

These tagged *galacto*-configured cyclophellitol epoxide and aziridine probes offer a powerful activity-based protein profiling approach to the study of α-GAL in FD because they provide a route to analysing α-GAL activity within different phenotypic variants of the disease. These probes also show considerable therapeutic potential, as they may be used to monitor and evaluate the efficacy of different therapeutic approaches or aid in the screening of pharmacological chaperones for α-GAL. Consequently, ABPs which target lysosomal enzymes such as α-GAL are of considerable interest.

2.2.3 Inhibitors for Retaining α-Galactosidases

Specific inhibitors of glycosidases are of interest for a number of different applications, namely mechanistic studies on the enzymes themselves, control of enzymatic activity, study of glycoprotein processing and potential therapeutic behaviour. Several active site-directed

inactivators of glycosidases have been described, which generally consist of a glycoside containing reactive functionalities such as isothiocyanates, epoxides and α-halocarbonyls[604–606]. However, one of the most influential classes of retaining glycosidase inhibitors are the activated fluoro-glycosides which were pioneered by the Withers Lab.

2.2.3.1 Activated-Fluoroglycosides

2-deoxy-2-fluoro-D-glycosyl fluorides are mechanism-based inhibitors which function by trapping the enzymatic nucleophile of retaining glycosidases in a covalent fluoroglycosyl intermediate[607]. The mode of action of these inhibitors is based on the general catalytic mechanism of retaining glycosidases, which hydrolyse their substrates with overall retention of anomeric stereochemistry[608]. This is achieved through a 2-step double displacement mechanism in which the glycosidic bond of the substrate is cleaved to form a covalent glycosyl-enzyme intermediate (glycosylation) which is then hydrolysed in a second step to complete the enzymatic reaction and free the enzymatic nucleophile (deglycosylation)[609], Figure 2.2. Both the glycosylation and deglycosylation steps of this mechanism proceed via transition states with substantial oxocarbenium ion character[608]. Therefore, Withers and co-workers suggested that substitution of the C2-hydroxyl (adjacent to the reaction centre) with an electronegative fluorine atom should destabilize these transition states and decrease the rates of both glycosylation and deglycosylation[607]. Thus, 2-deoxy2-fluoroglycosides are very slow substrates, but with the inclusion of a relatively reactive aglycone leaving group, it is possible to increase the glycosylation rate to permit accumulation of the 2-deoxy-2-fluoroglycosyl enzyme intermediate. Therefore, if the deglycosylation rate is sufficiently slow then this glycosyl enzyme intermediate will be "trapped" and inhibit the enzyme[607].

Over the years, the Withers' lab has demonstrated the success of 2-deoxy-2-fluoro-D-glycosyl fluorides through the use of differently configured glycosides to inhibit a range of retaining glucosidases, galactosidases and mannosidases[607,610,611]. Indeed, galactosyl fluorides have been used to identify the catalytic nucleophile of numerous GH27 family members by the generation of trapped fluoroglycoside intermediates[607,610,612]. In this regard, activated 2,2-difluorogalactosides were employed to identify the active site nucleophile of human α-GAL[577] and α-galactosidase from the white rot fungi *Phanerochaete chrysosporium*[584], whilst 5-fluorogalactosides were used to determine the catalytic nucleophile of the green coffee bean α-galactosidase[585], Figure 2.7. In the latter case, the 5-fluoro-α-D-galactopyranosyl inhibitor was found to inhibit the green coffee beans α-

galactosidase in a competitive and reversible manner, Figure 2.7 (b). Nevertheless, the intermediate was sufficiently long lived to allow for proteolysis, isolation and identification of the labelled peptide through LC-MS[585]. Although such fluoroglycosides have proved incredibly useful in elucidating the mechanistic chemistry of various glycosidases, the hydrolytic susceptibility of fluoroglycosyl-enzyme intermediates renders these inhibitors unsuitable for certain applications.

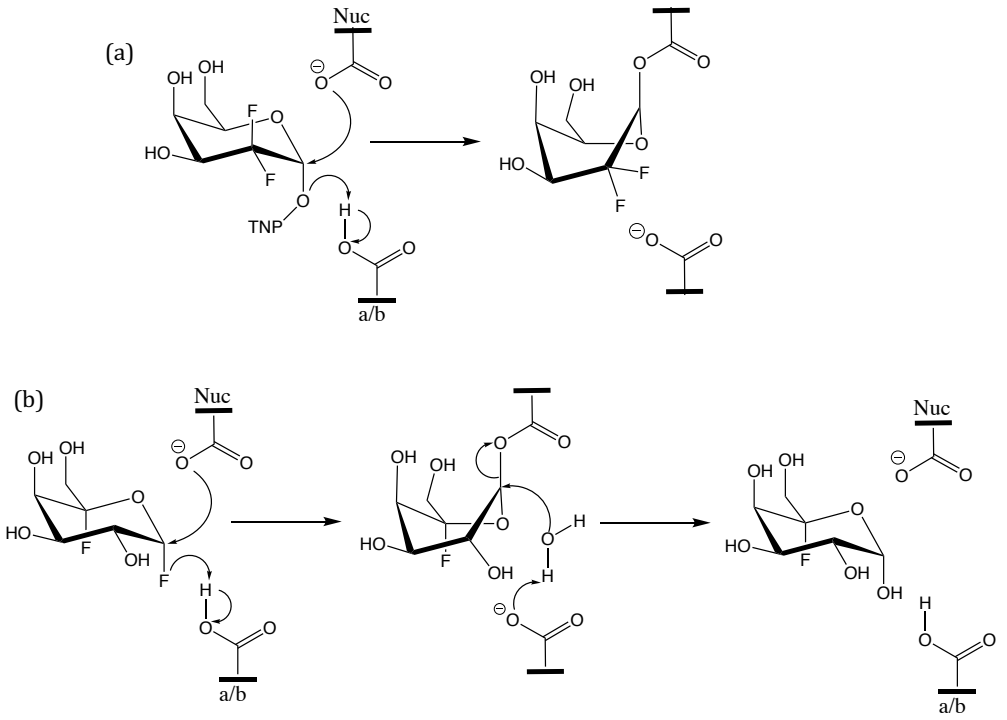


Figure 2.7: Reaction mechanism of a retaining α-galactosidase with (a) 2‘,4‘,6‘-Trinitrophenyl-2-Deoxy-2,2-difluoro-α-galactoside to form a trapped fluoroglycosyl-enzyme intermediate and (b) 5-fluoro-α-D-galactopyranosyl fluoride to form a covalent intermediate which is turned over by hydrolysis.

2.2.3.2 Cyclophellitol-Based Inhibitors

Another important class of covalent α-galactosidase inhibitors is the cyclophellitol derived compounds. As part of their venture to design and synthesise a diverse suit of ABPs for retaining α-galactosidases, the Overkleeft lab has produced an impressive range of *galacto*-configured cyclophellitol based inhibitors[556,558,560]. Their approach is somewhat derived from early work by Tong and Ganem who reported a potent *galacto*-configured cyclic aziridine-based inactivator which interferes with α-galactosyl hydrolysis by covalent modification of the catalytic nucleophille[613]. However, the key nitrogen atom of this inactivator is part of both the piperidine and the aziridine, rendering it unavailable for modification, Figure 2.8 (a). In contrast, the *galacto*-configured cyclophellitol aziridine inhibitors developed by the Overkleeft group, which also work by covalent modification of

the catalytic nucleophile, allow the aziridine nitrogen to be functionalised[558], Figure 2.8 (b). This has permitted the development of a range of *N*-functionalised α-*galacto*-cyclophellitol aziridine inhibitors for retaining α-galactosidases, with the potential to fine tune the inhibitory potency and selectivity of such inhibitors by altering the *N*-functionality[556]. Indeed, these *N*-functionalised inhibitors provided the basis for the *galacto*-configured cyclophellitol aziridine ABPs outlined previously in *section 2.2.2.*

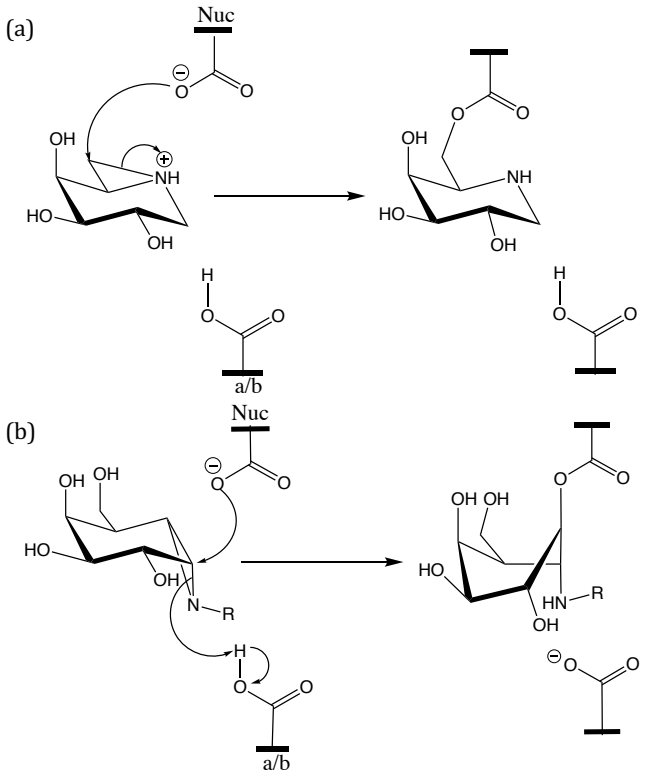


Figure 2.8: Chemical structure and mechanistic action of (a) the cyclic aziridine based α-galactosidase inactivator reported by Tong et al.[613] and (b) general *galacto*-configured cyclophellitol aziridine inhibitor developed by the Overkleeft lab (R = *N*-acyl/*N*-alkyl).

Having established a number of α-*galacto* configured cyclophellitol epoxide and aziridine inhibitors, the Overkleeft group expanded the range of α-galactosidase inhibitors by synthesising an α-*galacto* configured cyclophellitol cyclosulfate[561]. This was achieved by substitution of the α-configured epoxide with an electrophilic cyclosulfate equivalent, Figure 2.9 (a). Contrary to the transition state mimicking epoxide and aziridine inhibitors, this cyclosulfate analogue adopts the $^4C_1$ conformation, behaving as a Michaelis complex mimic[560,561]. Such cyclosulfate inhibitor is believed to covalently modify the catalytic nucleophile of retaining α-galactosidases according to the mechanism depicted in Figure 2.9 (b), in which the cyclic sulfate warhead is ring opened to form a covalent enzyme-inhibitor complex[561]. However, prior to this work, no co-crystal structures of this inhibitor were available to support this mechanism. Nevertheless, initial inhibition studies performed by

the Overkleeft lab revealed this α-*galacto*-cyclophellitol cyclosulfate to be a selective and irreversible inhibitor of human α-GAL, with affinity on par with the analogous cyclophellitol epoxide (apparent IC_{50} = 25 μM compared to 13 μM for equivalent epoxide)[561]. The success of this cyclosulfate inhibitor demonstrates how the panel of cyclophellitol-based α-GAL inhibitors may be readily expanded by using alternative electrophilic warheads.

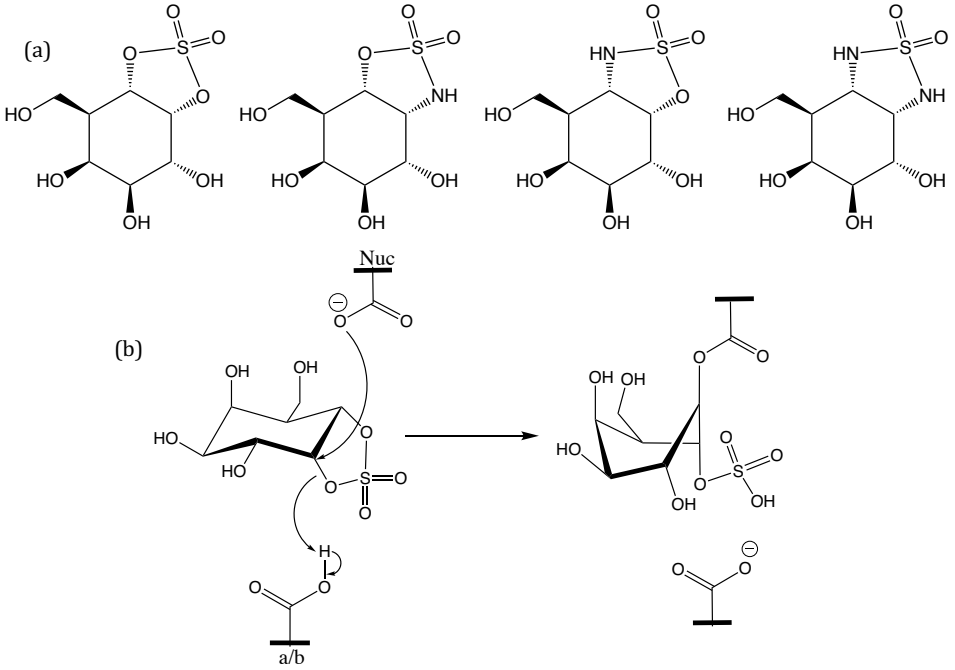


Figure 2.9: (a) Chemical structure of α-galactose configured cyclosulfate, cyclosulfamidate and cyclosulfamide compounds. (b) Predicted reaction mechanism of α-*galacto*-configured cyclophellitol cyclosulfate inhibitor with a retaining α-galactosidase.

2.2.3.3 Inhibitor for Pharmacological Chaperone Therapy

In light of the considerable therapeutic potential of PCT, it is evident that potent yet reversible inhibitors of α-GAL are required to expand the repertoire of molecular chaperones for FD. In this regard, Overkleeft and co-workers hypothesised that substitution of one or both of the cyclosulfate ring oxygens of the α-*galacto*-cyclosulfate inhibitor with a nitrogen atom, should reduce the leaving group capacity and generate reversible α-GAL inhibitors with potential chaperoning behaviour. Consequently, α-*galacto* configured cyclosulfamidate and cyclosulfamide compounds were synthesised[561], Figure 2.9 (a). Unfortunately, the cyclosulfamidate compound, in which the nitrogen atom occupied the position of the anomeric oxygen in the natural substrate, proved inactive against α-GAL and the cyclosulfamide was found to be a very weak, reversible inhibitor of α-GAL (apparent IC_{50} = 423 μM). However, the alternative cyclosulfamidate compound, in which the nitrogen occupies the C1 position, proved to be a rather good reversible α-GAL inhibitor (IC_{50} = 67 μM, *K*i = 110 μM)[561]. This suite of α-*galacto* configured cyclosulfate and cyclosulfamidate

compounds constitute a new generation of α-galactosidase inhibitors, which hold potential for the development of molecular chaperones for human α-GAL. However, until the work described in this chapter, there was no structural information on the conformation, reactivity and binding mode of these compounds.

2.2.4 Research Aims

The aim of this work was to address the lack of structural data on the *galacto*-configured cyclophellitol inhibitors designed by the Overkleeft lab. Specifically, this work aimed to obtain 3D crystal structures of human α-GAL in complex with the classic α-*galacto*-configured cyclophellitol epoxide, aziridine and *N*-functionalised aziridine inhibitors, as well as the novel *α-galacto*-configured cyclosulfate and cyclosulfamidate inhibitors, Figure 2.10. It was hoped this work would provide structural insight into the binding and conformation of these inhibitors to further explore the reported differences in reactivity and potency. This work also aimed functionally characterise the cyclosulfate and cyclosulfamidate inhibitors and investigate for potential chaperoning behaviour.

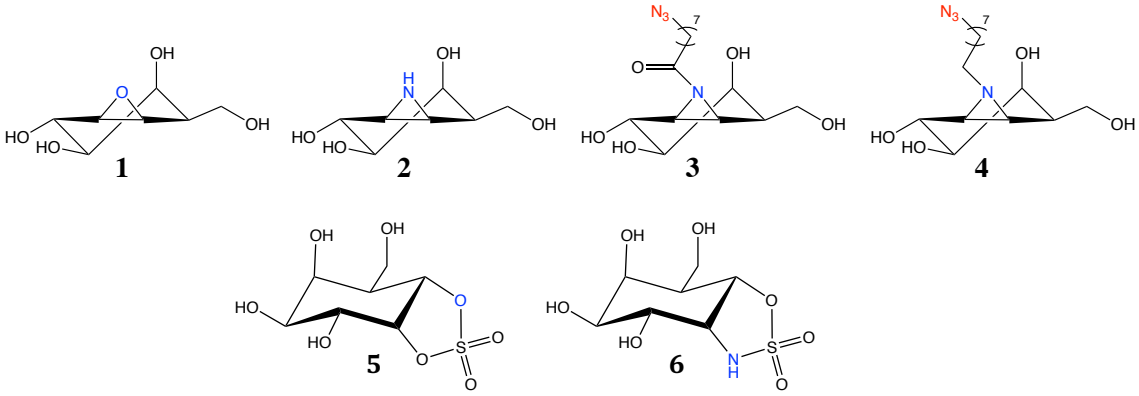


Figure 2.10: Series of α-*galacto*-configured cyclophellitol inhibitors investigated for human α-GAL in this work.

2.3 Materials and Methods

2.3.1 Crystallisation of Fabrazyme®

Initial crystallisation screening of Fabrazyme® (generously provided by Professor Johannes Aerts, Leiden University) was carried out using sitting-drop vapour-diffusion, set up using a *Mosquito Crystal* liquid handling robot in 96 well plates. Optimisation was performed from conditions taken from Guce et al. (2009)[577] and JCSG D4 with microseeding. A seed stock of crushed Fabrazyme® crystals was prepared in 0.2 M lithium sulfate, 0.1 M sodium acetate (pH 4.6) and 30% PEG 8K, according to previously published protocols[614]. Working seed stocks were prepared at 1, 1:10, 1:100, 1:1000, 1:10000 and 1:100000 dilutions. The well solution comprised 25% PEG 4K (50%), 0.1 M sodium acetate (pH 4.6) and 0.2 M lithium sulfate. Crystallization drops consisted of 500 nL Fabrazyme® (20 mg mL^{-1}), 100 nL microseeding solution and 400 nL well solution. Microseeding was optimised by varying the seeding concentration, yielding suitable crystals under the conditions: 25% PEG 4K (50%), 0.1 M sodium acetate (pH 4.6) and 0.2 M lithium sulfate and 1:1000 microseeding.

2.3.1.1 Unliganded Crystal Structure

Unliganded crystals were fished from the crystallisation drops and briefly transferred to an ethylene glycol (EG) cryoprotectant, containing 25% EG (100 %), 50 % PEG 4K (50%), 0.1 M sodium acetate (pH 4.6) and 0.2 M lithium sulfate, before flash freezing in liquid nitrogen. Diffraction quality was tested in-house using a Rigaku micromax-007HF x-ray generator in conjunction with an Actor robotic sample changer. Suitable crystals were sent to the Diamond Light Source facility for full data collection.

2.3.1.2 Inhibitor Complex Structures

The lyophilized inhibitors, synthesised by members of the Overkleeft lab, were resuspended at 20 mM in HEPES buffer (20 mM, pH 7). Mother liquor solutions were prepared (25% PEG 4K (50%), 0.1 M sodium acetate (pH 4.6) and 0.2 M lithium sulfate) and supplemented with each inhibitor to a final concentration of 4 mM. Unliganded crystals were soaked in the inhibitor spiked mother liquor solutions for 4 hours before briefly transferring the crystals to an EG cryoprotectant, containing 25% EG (100 %), 50 % PEG 4K (50%), 0.1 M sodium acetate (pH 4.6) and 0.2 M lithium sulfate. The cryoprotected crystals were flash frozen in liquid nitrogen for full data collection.

2.3.1.3 Data Collection and Processing

Data for all crystals were collected at the at either the i04 or i03 beamline of the Diamond Light Source facility. The data were processed using XIA2[615,616]and AIMLESS[617,618] data reduction pipelines through the CCP4i2 software[619]. All structures were re-indexed to the appropriate space group (P 3_2 2 1) and solved by molecular replacement using MOLREP[620] with PDB 1R46[440] as the search model. Refinement was performed using REFMAC[621], followed by several rounds of manual model building with COOT[622]. Idealised coordinate sets and refinement dictionaries for the inhibitors were generated using JLIGAND[623] and sugar conformations were validated using Privateer[594]. Data collection and refinement statistics for all the structures discussed in this work are provided in Table 2.2. All crystal structure figures were generated in CCP4mg[94].

2.3.2 Thermal Shift Assays

2.3.2.1 Effect of Inhibitors on Thermal Stability

Reactions containing Fabrazyme® (8.43 μM) and inhibitor (0 μM-1000 μM) were prepared in PCR tubes in triplicate. Buffer (20mM HEPES, 100 mM NaCl, pH 7.4) was added to give a reaction volume of 24 μL and a final enzyme concentration of 1 μM. The enzyme and inhibitor were incubated for 1 hour at room temperature, after which SYPRO orange dye (1 μL, ×125) was added. The PCR tubes were immediately placed in a Stratagene Mx3005P qPCR instrument and the SYPRO orange dye was excited at λ_{ex} 517 nm. The resulting fluorescence signal was monitored at λ_{em} 585 nm as the temperature was ramped from 25-95 °C at a rate of 2 °C min^{-1}. The fluorescence signal of each sample was measured in triplicate at each temperature increment.

Initial data analysis was performed using the JTSA software (Bond PS. JTSA. (2017). http://paulsbond.co.uk/jtsa)[624]. The averaged fluorescence signal was plotted against temperature for each inhibitor concentration and the data fitted to a sigmoid-5 function to yield fluorescence thermal stability curves. The protein melting temperature (Tm) was determined at each inhibitor concentration from the mid-point of the stability curves. The change in melting temperature (ΔTm) was calculated relative to a control with no inhibitor. Further analysis was performed in Origin graphing software. A plot of change in protein melting temperature vs inhibitor concentration was constructed and fitted to a 4-parameter logistic function, allowing the theoretical maximum change in Tm (ΔTm_{max}) and the ligand concentration required to achieve half maximal stabilisation (x_0) to be determined.

2.3.2.2 *Effect of pH on Thermal Stability*

The thermal shift "Thermofluor" assay procedure outlined in *section 2.3.2.1* was repeated using McIlvaine buffer (20 mM Na_2HPO_4 and 10 mM Citric acid) at pH 4.5, pH 5.5 and pH 7.4 in place of the HEPES/NaCl buffer to determine the effect of pH on thermal stability.

2.3.3 Enzyme Kinetics

4-Methylumbelliferyl-α-D-galactopyranoside (4-MU-αGal) was prepared at 2 mM in McIlvaine pH 4.5 buffer and diluted two-fold to 15.6 µM. Aliquots (25 µL) of each substrate solution were added to a black 96-microwell polystyrene plate in triplicate. Fabrazyme® was prepared at 40 nM in McIlvaine pH 4.5 buffer and 25 µL was added to each well of substrate solution, giving a final enzyme concentration of 20 nM. Activity against 4-MU-αGal was monitored continuously over 5 minutes at room temperature by measuring the fluorescence of liberated 4-MU (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) using a CLARIOstar monochromator microplate reader (BMG LabTech). A linear calibration was obtained by measuring the fluorescence of the 4-MU product (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) prepared at serial dilutions in McIlvaine pH 4.5 buffer.

All data were processed in Origin graphing software. A calibration curve was constructed by plotting measured 4-MU fluorescence signal against [4-MU] concentration. The rate of substrate hydrolysis (V) was determined at each substrate concentration and fitted by non-linear regression to the Michaelis Menten equation ($\upsilon = V_{max}[S]/(K_M + [S])$), from which the maximum rate of reaction (V_{max}) and Michaelis constant (K_M) were determined.

2.3.4 Thermal Stability Assay

Fabrazyme® was prepared at 400 nM in kinetics buffer (McIlvaine pH 4.5, supplemented with 1% BSA) and added to wells of a black 96-microwell polystyrene plate. Inhibitor was prepared at 200 µM in kinetics buffer and added to each well containing Fabrazyme® (400 nM) in a 1:1 ratio. The enzyme and inhibitor were mixed and 5 µL was immediately transferred to 45 µL substrate prepared at 1.11 mM in kinetics buffer (time point 0 min). The reaction was quenched with 50 µL Na_2CO_3 (1M, pH 10.3) after 2 minutes and the fluorescence signal of liberated 4-MU was measured using a CLARIOstar monochromator microplate reader (λ_{ex} 360 nm, λ_{em} 450 nm). The residual activity was determined relative to a control using the 4-MU calibration prepared in *section 2.3.3*. The above procedure was repeated following incubation with the inhibitor for 120-minutes.

Table 2.2: Collection and refinement statistics for all r-αGAL crystal structures obtained in this chapter.

	Apo	**1**	**2**	**3**	**4**	**5**	**6**	**Migalastat**
Data collection								
Space group	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1	*P* 3_2 2 1
Cell dimensions								
a, *b*, *c* (Å)	90.9, 90.9, 216.7	90.4, 90.4, 216.3	90.4, 90.4, 216.6	90.3, 90.3, 216.0	90.4, 90.4, 215.8	90.3, 90.3, 216.0	90.6, 90.6, 216.5	90.6, 90.6, 216.4
α, β, γ, (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	78.71-1.75 (1.78-1.75)*	63.48-2.02 (2.07-2.02)	63.52-2.04 (2.09-2.04)	78.19-1.94 (1.97-1.94)	107.89-1.73 (1.76-1.73)	78.36-2.07 (2.12-2.07)	63.59-1.99 (2.03-1.99)	45.12-2.30 (2.38-2.30)
R_{merge}	0.11 (0.94)	0.12 (2.52)	0.13 (2.70)	0.14 (0.98)	0.11 (0.93)	0.12 (2.17)	0.11 (2.18)	0.20 (3.60)
R_{pim}	-	0.034 (0.778)	0.042 (0.844)	-	-	0.027 (0.669)	0.025 (0.684)	-
$CC_{1/2}$	0.999(0.591)	0.999 (0.822)	0.986 (0.810)	0.997 (0.544)	0.997 (0.664)	0.998 (0.745)	0.999 (0.732)	0.997 (0.616)
I / σI	10.8 (1.6)	11.7 (1.1)	5.6 (0.9)	8.1 (1.7)	11.0 (1.8)	9.8 (1.1)	11.5 (1.2)	8.4 (0.7)
Completeness (%)	99.9 (97.0)	100.0 (100.0)	100 (99.9)	100.0 (97.8)	100.0 (99.9)	100.0 (100.0)	100.0 (99.8)	100.0 (100.0)
Redundancy	11.7 (10.6)	12.1 (12.2)	12.3 (11.9)	11.9 (12.3)	11.9 (12.2)	12.2 (12.4)	12.2 (11.9)	11.9 (12.1)
Refinement								
Resolution (Å)	78.71-1.75	63.48-2.02	63.52-2.04	78.19-1.94	107.89-1.73	78.36-2.07	63.59-1.99	45.12-2.30
No. reflections	104048	67994	66147	75849	107192	63129	71432	46243
R_{work} / R_{free}	0.21/0.25	0.19/0.26	0.20/0.26	0.23/0.29	0.20/0.24	0.18/0.24	0.18/0.24	0.23/0.29
No. atoms								
Protein	6279	6229	6242	6279	6241	6242	6245	6240
Ligand/ion	274	345	317	293	253	409	402	169
Water	357	233	240	237	345	272	274	51
B-factors (Å^2)								
Protein	36	55	58	46	37	58	60	70
Ligand/ion	65	82	88	74	56	92	91	96
Water	41	57	58	45	42	59	63	53
R.m.s. deviations								
Bond lengths (Å)	0.012	0.008	0.007	0.010	0.012	0.008	0.008	0.007
Bond angles (°)	1.55	1.54	1.53	1.53	1.63	1.57	1.58	1.59
Ramachandran								
Outliers (%)	0.9	1.2	0.8	1.1	0.9	0.9	0.9	2.0
PDB	**-**	6IBR	6IBT	-	-	6IBM	6IBK	-

2.4 Results and Discussion

2.4.1 Crystallisation and Structure of Fabrazyme®

Firstly, the crystallisation of r-αGAL (Fabrazyme® , a kind gift from Professor Johannes Aerts, Leiden University) was optimised from a combination of JCSG D4 screening conditions outlined by Guce et al. (2009)[577] using sitting drop vapour diffusion with microseeding. The final optimised crystallisation conditions were 0.2 M lithium sulfate, 0.1 M sodium acetate pH 4.6 and 25% PEG 4000 with 1:1000 microseeding, Figure 2.11 (a).

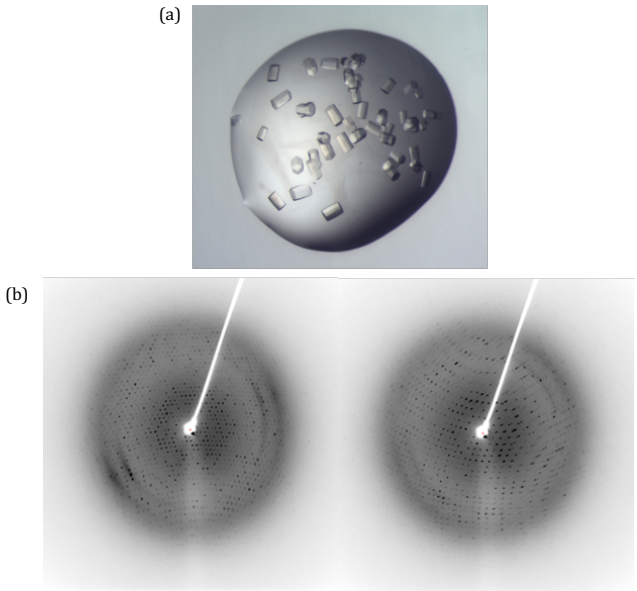


Figure 2.11: (a) Crystals of r-αGAL generated with 1:1000 microseeding (b) Diffraction images of r-αGAL crystals taken in house at 0° and 90° respectively.

An unliganded structure of r-αGAL was obtained at 1.75 Å resolution to reveal a structure that is essentially identical to that of the previously reported structure (PDB 1R46)[440] but at a much higher resolution, Figure 2.12 (a). Specifically, r-αGAL crystallised as a homodimer in which each monomer consists of two domains: a $(\beta/\alpha)_8$ domain spanning residues 32-330 and a C-terminal domain spanning residues 332-422, Figure 2.12 (b). The second domain contains eight antiparallel β strands arranged in a β sandwich that packs against the α_6, α_7 and α_8 helices of the first domain. The active site is located in the first domain at the centre of the β-barrel, Figure 2.12 (b), as expected for a GH-D CAZy clan member. Additionally, both α-GAL monomers display three *N*-linked glycosylation sites at residues Asn139, Asn192 and Asn215. Of note, a high mannose *N*-glycan ($Man_2GlcNAc_2$-Asn) could be modelled at Asn192 of molecule B, consistent with the importance of this *N*-glycosylation site in ensuring lysosomal trafficking of r-αGAL via the M6P pathway.

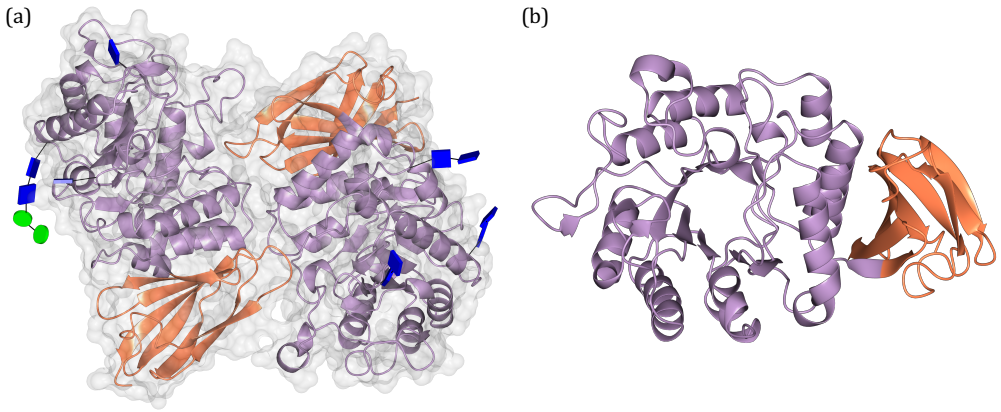


Figure 2.12: (a) Ribbon and surface diagram of r-αGAL (Fabrazyme®) dimer. *N*-linked glycans are illustrated in Glycoblock format[591]. (b) Each monomer comprises of two domains: a $(\beta/\alpha)_8$ domain (lilac) containing the active site and a C-terminal domain (orange) comprising of eight antiparallel β-strands

2.4.2 3D Complexes with *Galacto*-configured Cyclophellitol Epoxide and Cyclophellitol Aziridine Inhibitors

Once suitable unliganded crystals of r-αGAL were obtained, ligand-binding studies were performed with α-*galacto* configured cyclophellitol epoxide (**1**) and aziridine (**2**).

2.4.2.1 Complex with Galacto-configured Cyclophellitol Epoxide Inhibitor

Data for the cyclophellitol epoxide (**1**) complex were collected at the i04 beamline to a resolution of 2.02 Å and solved to reveal a single molecule of **1** covalently bound to the catalytic nucleophile (Asp170) of both r-αGAL molecules in the asymmetric unit. The observed electron density unambiguously shows that the inhibitor has reacted with Asp170 through trans-diaxial ring opening of the epoxide trap to form a covalent enzyme-inhibitor complex, Figure 2.13 (a). The reacted *galacto*-configured cyclophellitol ring adopts the $^1S_3$ skew-boat conformation in the covalent complex, which is consistent with the conformation of the covalent intermediate in the α-galactosidase reaction itinerary, Figure 2.3. The reacted cyclitol also forms an extensive hydrogen bonding network in the active site, making hydrogen bonds to Asp92, Asp93, Lys168, Glu203 and the general acid-base residue Asp231, Figure 2.13 (b). This structure clearly demonstrates that cyclophellitol epoxide (**1**) inhibits r-αGAL according to the expected mechanism-based mode of action and with a ring conformation that is consistent with the known reaction coordinate for α-galactosidases[586].

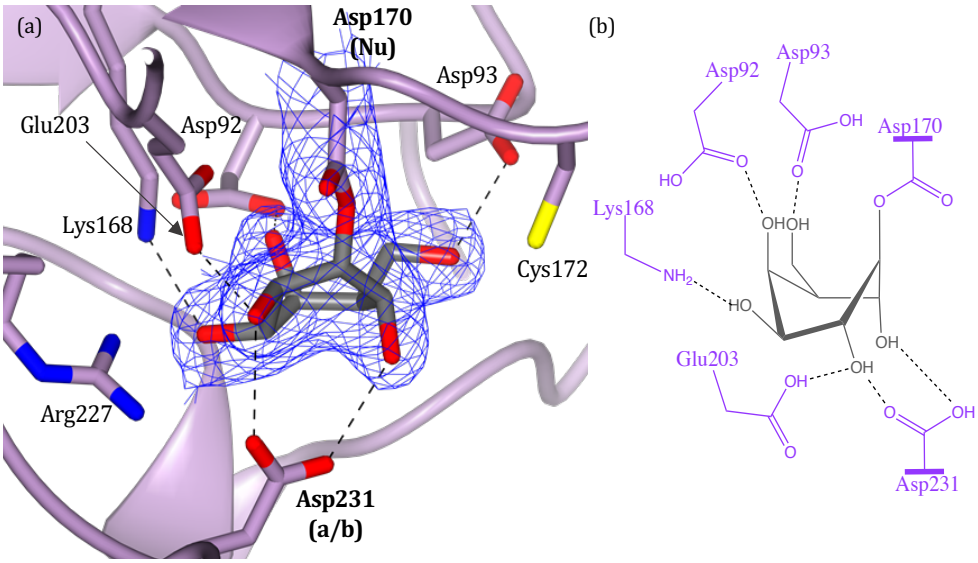


Figure 2.13: (a) 3D active site structure of cyclophellitol epoxide (**1**) covalently bound to the catalytic nucleophile (Asp170) of r-αGAL by trans diaxial ring opening of the epoxide to generate a covalent enzyme-inhibitor complex in a skew-boat $^{1}S_{3}$ conformation. 3. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Asp170 contoured to 1.2σ (0.22 electrons/Å^3). (b) Schematic representation of the hydrogen bonding network of **1** bound covalently to Asp170.

2.4.2.2 Complex with Galacto-configured Cyclophellitol Aziridine Inhibitor

Configurationally isomeric cyclophellitol epoxides and aziridines are often considered interchangeable, but not equally potent, as inhibitors of retaining glycosidases[93,599,600]. To investigate whether this holds true for human αGAL, a co-crystal structure in complex with cyclophellitol aziridine (**2**) was obtained for comparison with epoxide (**1**). The co-crystal structure was solved to 2.04 Å resolution to reveal a single molecule of **2** bound covalently to the catalytic nucleophile of r-αGAL (Asp170) through the same mechanism observed for the epoxide (**1**). Specifically, aziridine warhead covalently modifies the catalytic nucleophile through trans-diaxial ring opening to form a covalent enzyme-inhibitor complex in the skew-boat ($^{1}S_{3}$) conformation, Figure 2.14 (a). The cyclophellitol ring also forms an identical hydrogen bonding network to that of the epoxide equivalent, Figure 2.14 (b). Therefore, the identical mode of binding observed for the epoxide and aziridine inhibitors suggests α-*galacto* configured cyclophellitol epoxides and aziridines may be used somewhat interchangeably as α-galactosidase inhibitors.

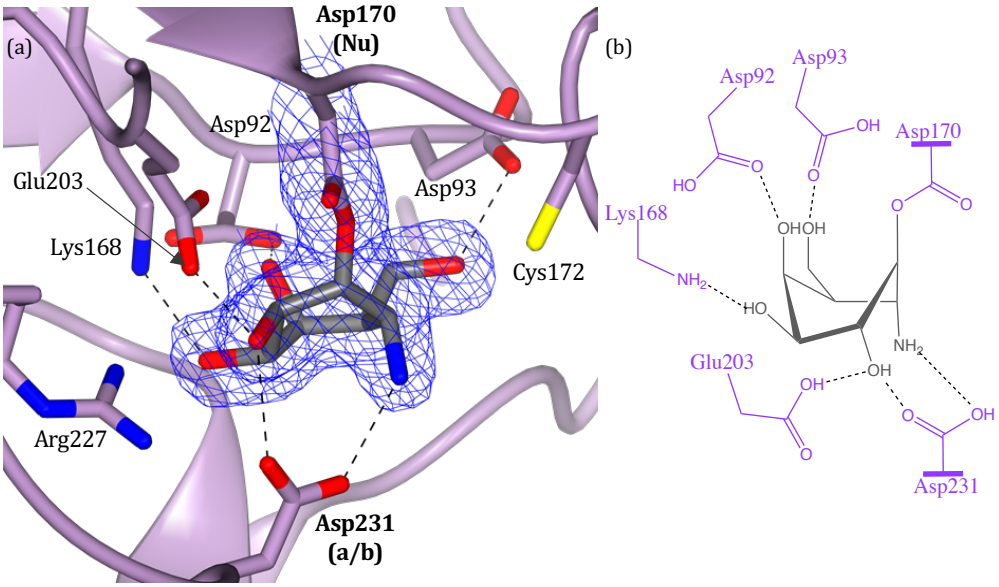


Figure 2.14: (a) 3D active site structure of cyclophellitol aziridine (**2**) covalently bound to the catalytic nucleophile (Asp170) of r-αGAL by trans diaxial ring opening of the aziridine warhead to generate a covalent enzyme-inhibitor complex in a skew-boat $^{1}S_{3}$ conformation. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Asp170 contoured to 1.2σ (0.22 electrons/Å^3). (b) Schematic representation of the hydrogen bonding network of **2** bound covalently to Asp170.

2.4.3 3D Complexes with *N*-acyl and *N*-alkyl Functionalised *Galacto*-configured Cyclophellitol Aziridine Inhibitors

Prior to this work, *N*-acyl and *N*-alkyl functionalised cyclophellitol aziridine compounds had already surfaced as more potent inhibitors of human GBA than the unsubstituted aziridine equivalent (eg. IC_{50} = 17 nM for *N*-alkyl cyclophellitol aziridine vs IC_{50} = 0.5 μM for unsubstituted aziridine)[599]. Consequently, the Overkleeft Lab synthesised a number of *N*-functionalised *galacto*-configured cyclophellitol aziridine inhibitors for α-GAL by making appropriate changes to the stereochemistry of the cyclophellitol moiety[556]. Time-course labelling assays subsequently revealed that *N*-acyl *galacto*-cyclophellitol ABPs label α-GAL more efficiently than the equivalent *N*-alkyl probes[556]. This difference in labelling efficiency was not observed for the equivalent *gluco*-configured cyclophellitol aziridine ABPs developed for GBA[556]. Therefore, structural analysis of the *N*-acyl and *N*-alkyl *galacto*-configured cyclophellitol aziridine inhibitors on the 3D structure of r-αGAL was performed in this work in hope of providing an explanation for the reduced labelling efficiency of *N*-alkyl probes.

2.1.1.1 *Complex with N-acyl Galacto-configured Cyclophellitol Aziridine Inhibitor*

Data for the complex with *N*-acyl *galacto*-configured cyclophellitol aziridine (**3**) were collected at the i04 beamline to 1.94 Å. The resulting electron density unambiguously shows binding of the inhibitor to the catalytic nucleophile of r-αGAL (Asp170) by trans-diaxial ring opening of the *N*-acyl aziridine moiety to form a covalent enzyme-inhibitor complex in the expected skew-boat ($^1S_3$) conformation, Figure 2.15. Unfortunately, the electron density for the *N*-acyl aziridine was limited to the carbonyl group and subsequent 3 carbon atoms, Figure 2.15 (a). No electron density for the remaining alkyl chain and azide tag was observed, likely owing to the high flexibility and subsequent disorder of the alkyl chain. Nevertheless, the ring opened *N*-acyl aziridine was modelled extending out towards the dimer interface. It would be prudent to note, that in contrast to GBA which is reported to accommodate aziridine *N*-functionalisation in a distinct active site cleft, αGAL shows minimal interaction with the carbon chain of the N-acyl group of **3.** In fact, the carbon chain is relatively surface exposed, protruding out from the protein. The only notable interaction is the formation of a hydrogen bond between the carbonyl oxygen of the *N*-acyl group and Cys172. Overall, this structure demonstrates that inhibition of r-αGAL by *N*-acyl *galacto*-configured cyclophellitol aziridine **3** occurs according to the same mechanism as non-functionalised aziridine **2** to form an almost identical covalent enzyme-inhibitor complex.

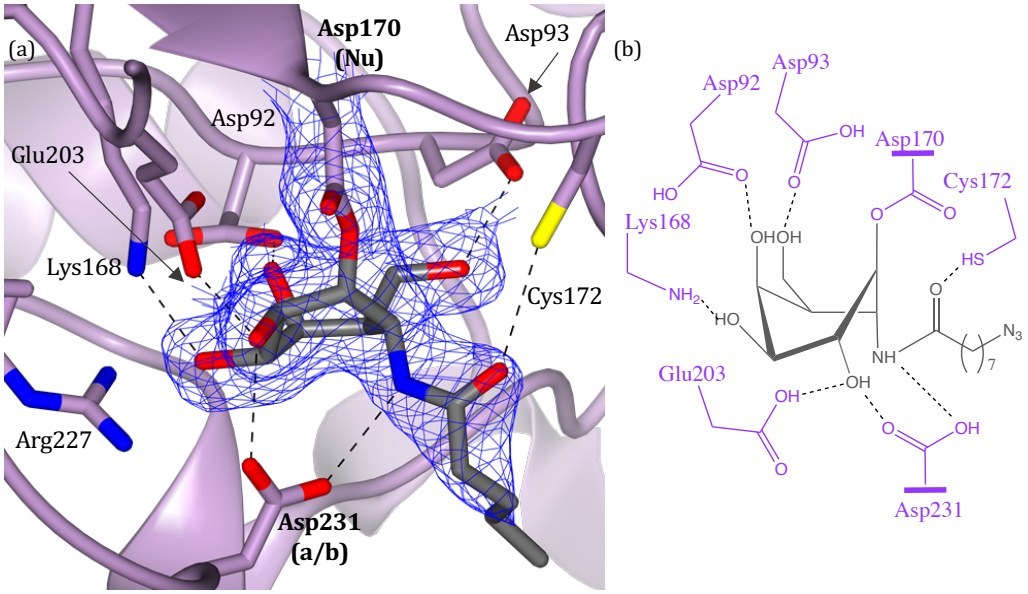


Figure 2.15: (a) 3D active site structure of *N*-acyl cyclophellitol aziridine (**3**) covalently bound to Asp170 of r-αGAL by trans diaxial ring opening of the aziridine warhead to generate a covalent enzyme-inhibitor complex in a skew-boat $^1S_3$ conformation. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Asp170 contoured to 1.2σ (0.26 electrons/Å^3). (b) Schematic of hydrogen bonding network of **3** bound to Asp170 in a skew-boat conformation.

2.4.3.1 *Complex with N-alkyl Galacto-configured Cyclophellitol Aziridine Inhibitor*

The co-crystal structure obtained with *N*-alkyl aziridine inhibitor (**4**) perhaps provides more insight into the reported difference in potency of *N*-alkyl and *N*-acyl ABPs. Indeed, the *N*-alkyl inhibitor (**4**) complex was solved to 1.73 Å resolution to reveal a single molecule of inhibitor **4** bound non-covalently in the active site of both molecules in the asymmetric unit, Figure 2.16 (a). In contrast to *N*-acyl analogue (**3**), this *N*-alkyl inhibitor did not covalently modify the catalytic nucleophile of r-αGAL. This was an unanticipated finding because ABP iterations of **4** have been shown to covalently label r-αGAL, albeit with reduced efficiency compared to the *N*-acyl equivalents[556]. The absence of covalent binding is further mystified by the fact the *galacto*-configured cyclophellitol ring adopts the expected transition state mimicking half-chair ($^4H_3$) conformation and forms an almost identical hydrogen bonding network to the non-functionalised aziridine inhibitor **2**, Figure 2.16 (b). Furthermore, on inspection of the distance between the aziridine moiety and the two catalytic residues (Asp170 and Asp231), the aziridine warhead appears readily poised for in-line nucleophilic attack, providing no insight as to why inhibitor **4** did not covalently inhibit r-αGAL.

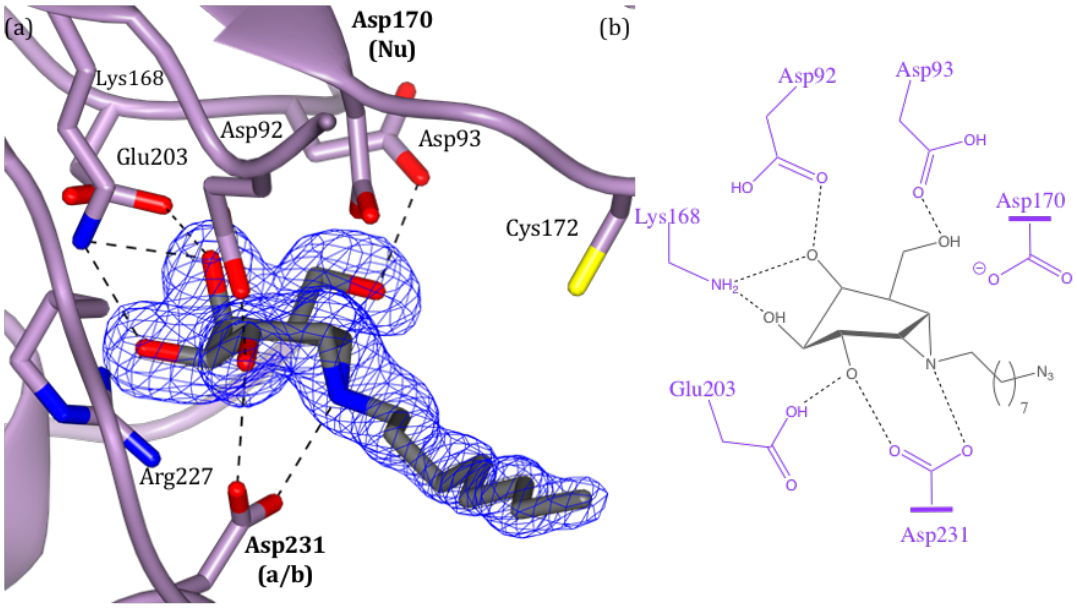


Figure 2.16: (a) 3D active site structure of *N*-alkyl cyclophellitol aziridine (**4**) bound non-covalently in r-αGAL active site adopting the half-chair $^4H_3$ conformation. Insufficient electron density was observed to model the azide tag. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand contoured to 1.2σ (0.30 electrons/Å^3). (b) Schematic representation of the hydrogen bonding network of **4**.

The fact **4** was observed to bind non-covalently in the active site of r-αGAL may provide some reasoning for the reduced labelling efficiency reported for *N*-alkyl probes in comparison to the *N*-acyl equivalents[556]. Additionally, this crystal structure also provides a

rare Michaelis Complex structure of such *galacto*-configured cyclophellitol inactivators, showing how they bind in the active site before they react with the catalytic nucleophile.

2.4.4 3D Complexes with *Galacto*-configured Cyclophellitol Cyclosulfate and Cyclosulfamidate Inhibitors

Following the success of cyclophellitol epoxide and aziridine glycosidase inhibitors[93,556,600,602], the concept was expanded to other electrophilic warheads such as cyclic sulfates. Previous work on an α-configured 1,6-cyclophellitol cyclosulfate inhibitor for retaining α-glucosidases, Figure 2.17 (a), demonstrated that the insertion of an α-configured 1,2-cis-cyclic sulfate, in place of the epoxide, yields a compound favouring the chair $^4C_1$ conformation[560]. This conformation mimics that of the Michaelis complex in the reaction itinerary of α-glucosidases, readily positioning the cyclic sulfate for nucleophilic attack, rendering this compound a potent α-glucosidase inhibitor[560]. Subsequent to the α-glucosidase cyclosulfate inhibitor, a new class of α-galactosidase cyclosulfate and cyclosulfamidate inhibitors were synthesised by the Overkleeft lab by making appropriate changes to the stereochemistry of the cyclophellitol moiety, Figure 2.17 (b). The work discussed in this section aimed to structurally investigate the binding mechanism of these new inhibitors.

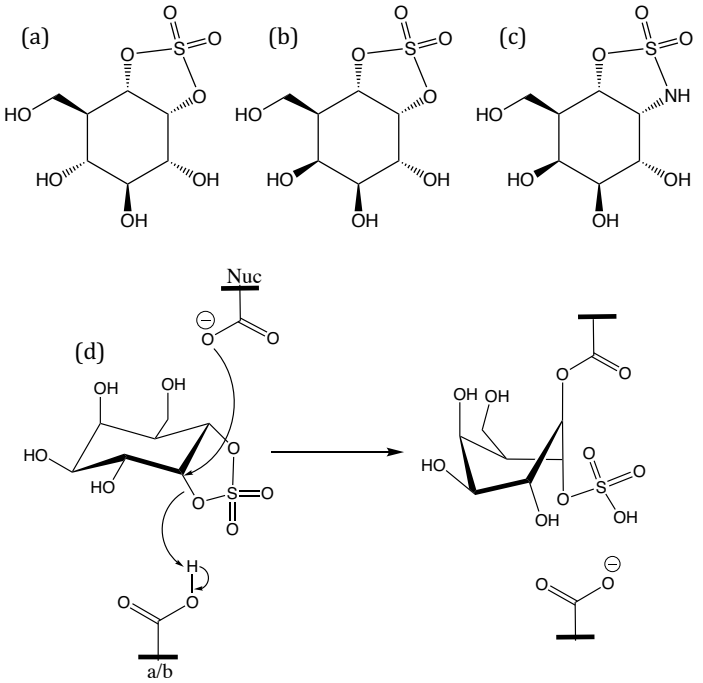


Figure 2.17: Chemical structure of (a) *Gluco*-configured α-1,6-cyclophellitol cyclosulfate (b) *Galacto*-configured α-1,6-cyclophellitol cyclosulfate inhibitor (**5**) (c) *Galacto*-configured α-1,6-cyclophellitol cyclosulfamidate inhibitor (**6**). (d) Predicted inhibition mechanism of the *galacto*-cyclosulfate inhibitor (**5**).

In initial *in vitro* inhibition assays performed by the Overkleeft group, α-galactosidase cyclosulfate **5** proved to be a selective, irreversible inhibitor of human α-GAL (IC_{50} = 25 μM), with affinity on par with the analogous epoxide inhibitor[561]. Consequently, this cyclosulfate inhibitor was expected to inhibit r-αGAL through a similar mechanism observed for the cyclophellitol epoxide, in which the cyclic sulfate warhead is ring opened to form a covalent enzyme-inhibitor complex, Figure 2.17 (d).

2.4.4.1 Complex with α-1,6-Cyclophellitol Cyclosulfate Inhibitor

In order to elucidate the inhibition mechanism, a co-crystal structure of r-αGAL in complex with inhibitor **5** was obtained. The structure was solved to 2.07 Å to reveal a single molecule of **5** covalently bound to the catalytic nucleophile. As anticipated, the inhibitor had reacted with Asp170 by ring opening of the α-1,6-cis-cyclosulfate warhead to form a covalent enzyme-inhibitor complex in the skew-boat ($^1S_3$) conformation, Figure 2.18 (a). This is consistent with the conformation of the covalent intermediate of the α-galactosidase reaction itinerary and with the previous co-crystal structures of the epoxide and aziridine inhibitors. Additionally, the reacted *galacto*-configured cyclophellitol ring forms hydrogen bonds with active site residues Asp92, Asp93, Lys168, Cys172, Glu203, Arg227 and the general acid-base residue Asp231, Figure 2.18 (b), which is almost identical to that observed for the epoxide analogue (**1**) but with additional interactions with Cys172 and Arg227.

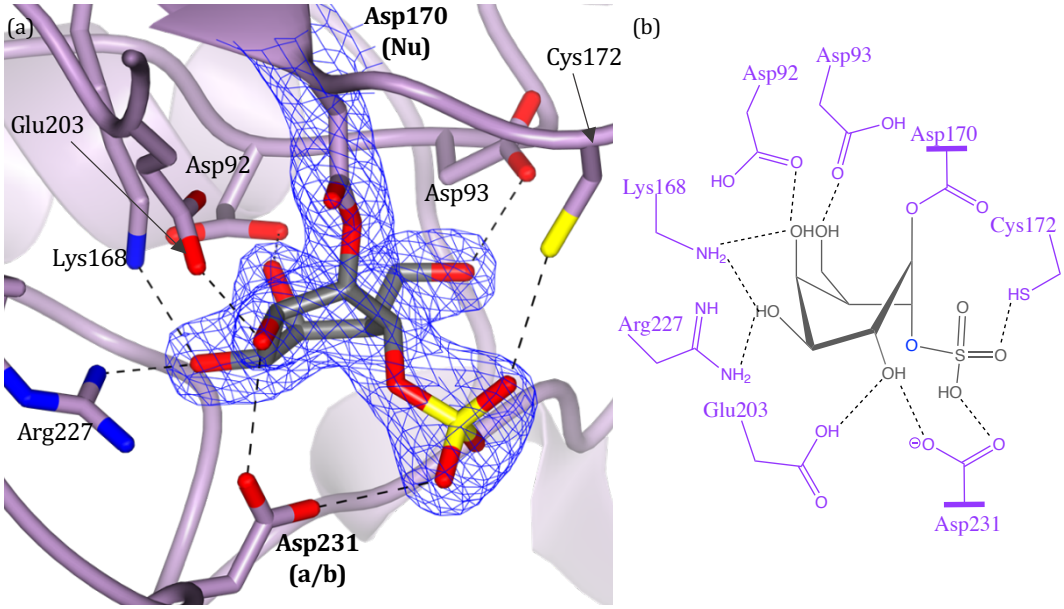


Figure 2.18: (a) 3D active site structure with observed electron density for α-cyclosulfate inhibitor (**5**) bound to Asp170 of r-αGAL to yield an enzyme-inhibitor complex adopting a skew-boat $^1S_3$ conformation. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Asp170 contoured to 1.2σ (0.21 electrons/Å^3). (b) Schematic representation of the hydrogen bonding network of **5**.

This structure clearly shows that cyclosulfate **5** reacts according to the same mechanism-based mode of action observed for the epoxide and aziridine analogues, providing validity to the expansion of the Overkleeft probes by the use of alternative electrophilic warheads. The successful development of this irreversible cyclosulfate inhibitor, sparked interest in the synthesis of a non-covalent, reversible α-galactosidase inhibitor. Specifically, it was hoped that non-covalent inhibitors with chaperone potential could be developed. Consequently, the Overkleeft lab synthesised a cyclosulfamidate analogue (**6**), Figure 2.17 (c), by substituting the cyclosulfate oxygen at the C1 position of the cyclophellitol ring with a nitrogen atom. It was proposed that this substitution would reduce the leaving group capacity and prevent nucleophilic attack at the anomeric position. Preliminary inhibition assays performed by the Overkleeft lab revealed **6** to be a rather good reversible inhibitor (IC_{50} = 67 μM, Ki = 110 μM)[561]. Therefore, a crystal structure of r-αGAL in complex with this cyclosulfamidate inhibitor (**6**) was obtained.

2.4.4.2 Complex with α-1,6-Cyclophellitol Cyclosulfamidate Inhibitor

A co-crystal structure in complex with cyclosulfamidate (**6**) was solved to 1.99 Å resolution to reveal unambiguous electron density for a single molecule of **6** bound non-covalently in the active site of r-αGAL, Figure 2.19 (a). It is evident that replacing the C1 cyclosulfate oxygen with a nitrogen does prevent nucleophilic attack on the *galacto*-configured cyclophellitol warhead. As a result, the cyclosulfamidate moiety remains intact and the inhibitor binds non-covalently in the active site, Figure 2.19.

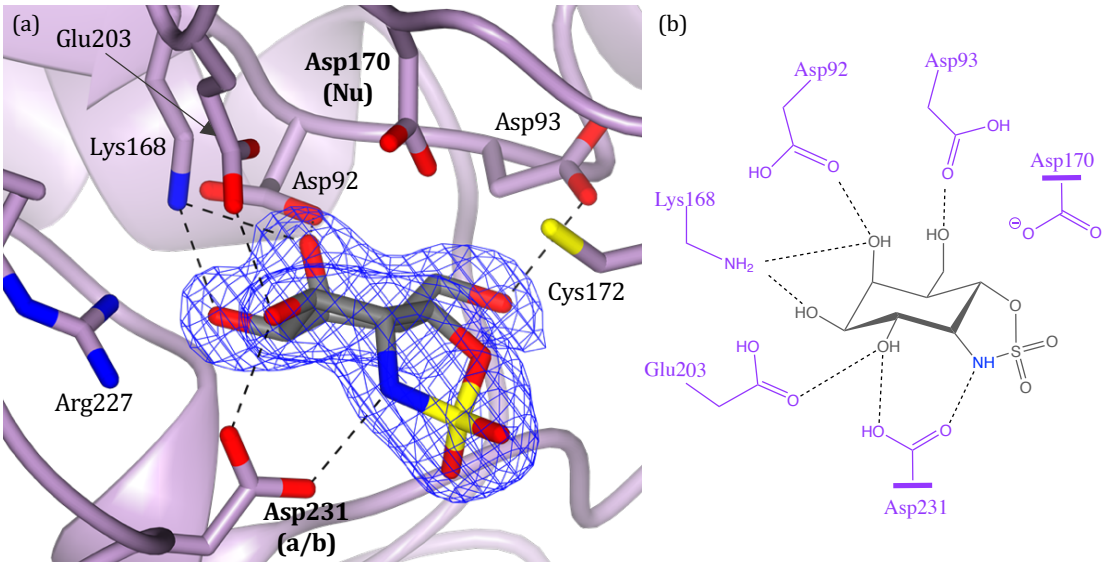


Figure 2.19: (a) Observed electron density for cyclosulfamidate (**6**) bound non-covalently in the active of r-αGAL, adopting a chair $^4C_1$ conformation. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand contoured to 1.2σ (0.21 electrons/Å^3). (b) Schematic representation of hydrogen bonding network of **6.**

On further inspection, it was found that the *galacto*-configured cyclophellitol ring of inhibitor **6** forms hydrogen bonds with active site residues Asp92, Asp93, Lys168, Glu203 and the catalytic acid-base residue Asp231, Figure 2.19 (b). Additionally, in contrast to the skew-boat ($^{1}S_{3}$) conformation adopted by cyclosulfate **5**, this cyclosulfamidate inhibitor adopts a chair ($^{4}C_{1}$) conformation, which is the expected conformation for a Michaelis complex mimic. Indeed, the observed change in ring conformation from $^{4}C_{1}$ to $^{1}S_{3}$ on covalent modification of the catalytic nucleophile, Figure 2.20 (c), demonstrates how these inhibitors take advantage of the $^{4}C_{1}\rightarrow{}^{4}H_{3}\rightarrow{}^{1}S_{3}$ reaction itinerary of α-galactosidases. Furthermore, this structural analysis suggests that configurational and conformational mimicry of the Michaelis complex may provide a powerful strategy for the development of potent, non-covalent inhibitors.

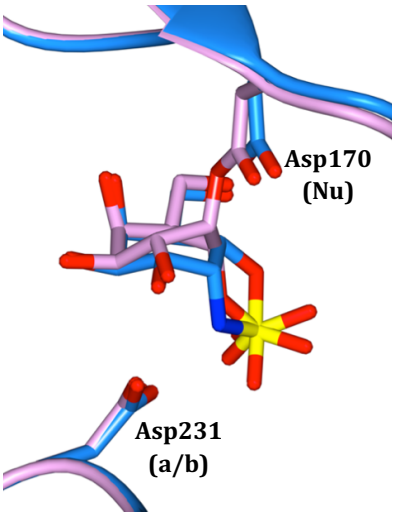


Figure 2.20: Overlay of inhibitor **5** (pink) covalently bound to Asp170 in $^{1}S_{3}$ conformation and inhibitor **6** (blue) bound non-covalently in $^{4}C_{1}$ conformation. Change in ring conformation from $^{4}C_{1}$ to $^{1}S_{3}$ on covalent modification of the catalytic nucleophile.

2.4.5 Investigating the Effect of α-1,6-Cyclophellitol Cyclosulfate on r-αGAL Stability

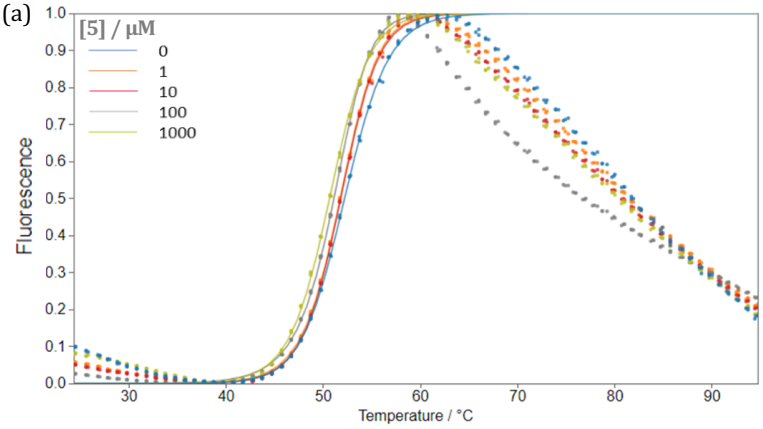
Recently, a number of the Overkleeft β-glucosidase cyclophellitol ABPs were shown to stabilise GBA, the lysosomal enzyme associated with Gaucher's disease, both *in vitro* and *in situ*[625]. Binding of these covalent ABPs to the enzyme active site resulted in improved GBA thermal stability and greater resistance to tryptic digestion. This stabilising effect was also noted in cultured cells, resulting in a marked increase in lysosomal GBA activity in wild type and mutant cell lines[108]. These cyclophellitol ABPs were proposed to improve the stability of GBA by stabilising the "correct" enzyme folding and conformation[625]. Following this work, it was postulated that the covalent cyclosulfate inhibitor **5** may exert a stabilising effect on

r-αGAL. Therefore, the effect of **5** on the stability of r-αGAL was investigated in this work using a "ThermoFluor" assay, which is a quick, temperature-based assay used to determine the effect of varying conditions on protein stability by monitoring changes in protein melting temperature (Tm)[626]. This approach makes use of fluorescent dyes which bind non-specifically to the hydrophobic surfaces of proteins. When proteins denature and unfold, the dye binds to exposed hydrophobic surfaces resulting in the expulsion of water and an increase in fluorescence. Recording this increase in fluorescence as a function of temperature yields a fluorescence thermal stability curve, the mid-point of which corresponds to the Tm of the protein.

2.4.5.1 Thermal Stability Analysis

The effect of the cyclosulfate inhibitor (**5**) on the thermal stability of r-αGAL was investigated by incubating Fabrazyme® with varying concentrations of **5** for 1 hour before adding Sypro Orange dye and monitoring the fluorescence emission as the protein was denatured by heating. Unfortunately, the resulting thermal shift curves indicate that inhibitor **5** does not stabilise r-αGAL against thermal denaturation because no shift in the fluorescence curves or change in protein melting temperature were observed in the presence of the inhibitor, Figure 2.21 (b).

(a)



(b)

[**5**] / µM	T_m / °C
0	52.2
1	51.7
10	51.8
100	51.1
1000	50.6

Figure 2.21: (a) Thermal stability curves recorded for Fabrazyme® pre-incubated with various concentrations of cyclosulfate inhibitor (**5**) for 1-hour at pH 7.4 (b) Melting temperatures (Tm) of Fabrazyme® in the presence of various concentrations of inhibitor **5.** Tm determined from mid-point of the fluorescence stability curve.

To ensure this result was genuine and the inhibitor was still functioning, the inhibitory activity of cyclosulfate **5** against r-αGAL was investigated. Fabrazyme® (200 nM) was pre-incubated with 100 µM of inhibitor **5** for at room temperature for 2 hours. The residual enzymatic activity was subsequently determined through a fluorogenic substrate assay using 4-methylumbelliferyl-α-D-galactopyranoside (4-MU-α-GAL). A 45% loss in α-GAL

activity was observed in the presence of inhibitor **5** relative to a control, Figure 2.22, suggesting that the cyclosulfate compound does inhibit r-αGAL. However, as only 45% inhibition was observed after 2 hours, the Thermofluor assay was repeated with a 4-hour pre-incubation period to ensure maximal inhibition. Unfortunately, no thermal stabilisation of r-αGAL was observed after this extended pre-incubation, confirming that cyclosulfate **5** binds but does not stabilise r-αGAL against thermal denaturation. This was a disappointing finding, however, there is far more interest in reversible, non-covalent inhibitors with stabilising behaviour, hence our attention turned to the cyclosulfamidate inhibitor **6**.

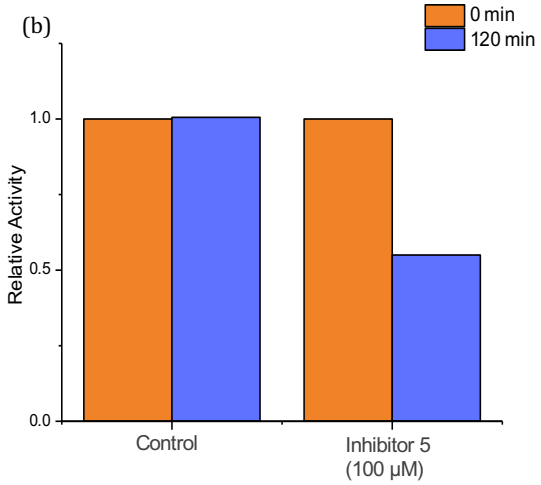


Figure 2.22: Inhibition of r-αGAL by inhibitor **5**. Residual enzymatic activity of r-αGAL after 2hr pre-incubation with 100 μM inhibitor **5** showing a 45% drop-in residual activity.

2.4.6 Chaperoning Behaviour of α-1,6-Cyclophellitol Cyclosulfamidate

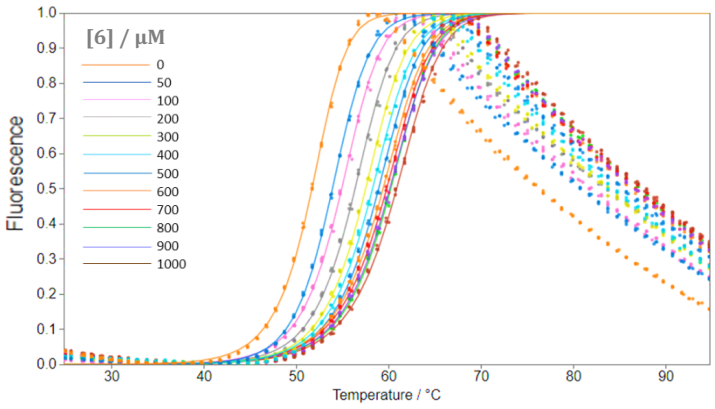
Currently, the iminosugar 1-deoxygalactonijirimycin (DGJ, Migalastat) is used as a pharmacological chaperone for the treatment of FD[519]. Migalastat is known to bind non-covalently and reversibly to the active site of α-GAL and when co-administered with ERT it stabilises the enzyme and enhances its intracellular activity[522]. Following the successful development of our non-covalent cyclosulfamidate α-galactosidase inhibitor (**6**), we pondered whether this reversible inhibitor may also show pharmacological chaperone potential towards α-GAL. Consequently, the chaperoning behaviour of compound **6** was investigated through a ThermoFluor assay, as described for inhibitor **5**.

2.4.6.1 Thermal Stability Analysis

Fabrazyme® was pre-incubated with various concentrations of inhibitor **6** at pH 7.4 for 1 hour before determining the Fabrazyme® melting temperature through a ThermoFluor assay (protocol outlined previously). Encouragingly, the resulting thermal stability curves exhibited a shift to higher temperatures with increasing inhibitor concentration,

demonstrating an increase in Tm, Figure 2.23. This indicates that the cyclosulfamidate inhibitor exerts a stabilising effect, protecting Fabrazyme® against thermal denaturation. In the absence of the inhibitor, the melting temperature of Fabrazyme® was calculated to be 51.7 °C, however, pre-incubation with 100 µM inhibitor **6** resulted in a 3.2 °C increase in Tm. Further increasing the inhibitor concentration resulted in a concomitant increase in Tm, with a 9.1 °C increase observed at 1000 µM inhibitor, Figure 2.23 (b).

(a)



(b)

[6] / µM	Tm/ °C	ΔTm/ °C
0	51.7	-
50	53.9	2.2
100	54.9	3.2
200	56.3	4.6
300	57.5	5.8
400	58.2	6.5
500	58.9	7.2
600	59.4	7.7
700	60.2	8.5
800	59.8	8.1
900	60.1	8.4
1000	60.8	9.1

Figure 2.23: Effect of cyclosulfamidate inhibitor (**6**) on the thermal stability of r-αGAL (a) Thermal stability curves recorded for Fabrazyme® pre-incubated with various concentrations of **6** for 1 hour at pH 7.4. The stability curves shift to higher temperature with increasing inhibitor concentration. (b) Table of Fabrazyme® melting temperatures (Tm) and the change in Tm (ΔTm) calculated at each inhibitor concentration.

A plot of the change in protein melting temperature (ΔTm) vs inhibitor concentration was constructed and fitted to a 4-parameter logistic function, Figure 2.24. This function is commonly used in pharmacology to calculate the IC_{50} and EC_{50} values of drugs[627]. By fitting the thermal shift data to a 4-parameter logistic function the theoretical maximum change in protein melting temperature (ΔTm_{max}) and the ligand concentration required to achieve half maximal stabilisation (x_0) could be estimated, Figure 2.24. Consequently, ΔTm_{max} and x_0 were calculated to be 12.7 $\pm$ 1.1 °C and 372 $\pm$ 6 µM respectively. For a small molecular inhibitor, this theoretical maximum increase in Tm is quite significant, suggesting that inhibitor **6** stabilises r-αGAL against thermal denaturation. Therefore, unlike the cyclosulfate analogue (**5**), the cyclosulfamidate inhibitor may show potential as a novel pharmacological chaperone for Fabry's disease.

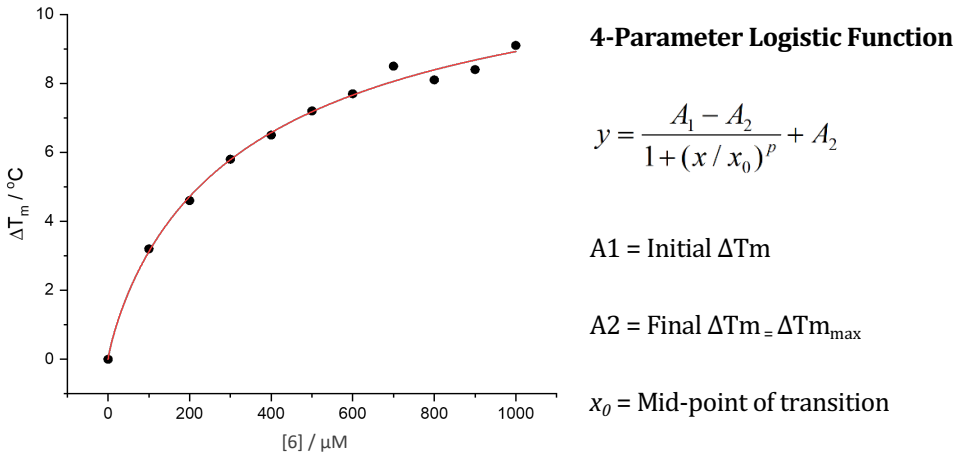


Figure 2.24: Plot of ΔTm versus cyclosulfamidate inhibitor (**6**) concentration fitted to a 4-parameter logistic function. ΔTm_{max} = 12.7 ± 1.1 °C, x_0 =372 ± 6 μM. Errors are fitting errors.

2.4.6.2 Structural Comparison of α-Cyclosulfate and α-Cyclosulfamidate Inhibitors

In an attempt to rationalise the thermal stabilisation of r-αGAL by cyclosulfamidate **6** but not cyclosulfate **5**, the two crystal structures were overlaid to investigate any structural differences. This structural comparison revealed no significant differences in tertiary structure or active site, Figure 2.25 (a), suggesting that the stabilisation of WT r-αGAL by the cyclosulfamidate does not result from the enforcement of a specific protein conformation.

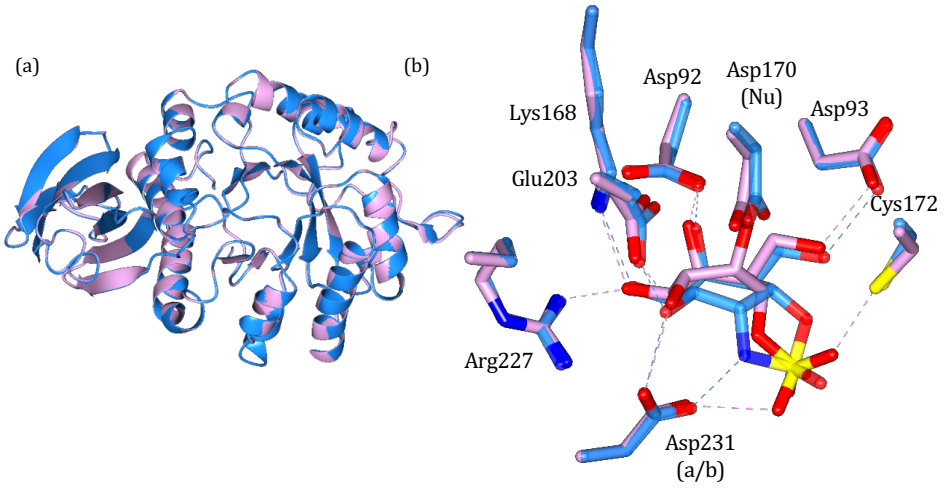


Figure 2.25: Overlay of the cyclosulfate inhibitor (**5**) complex crystal structure (pink) with the cyclosulfamidate inhibitor (**6**) complex crystal structure (blue). No significant difference in (a) tertiary structure or (b) active site structure is observed.

Pharmacological chaperones (PCs) are thought to enhance protein function in a number of ways, such as inducing thermodynamic stabilisation, enforcing correct protein folding through template-based induction or provoking changes in the folding-unfolding kinetics[628]. Therefore, enforcing a specific protein conformation is not necessarily a

feature of all PCs, as reported for the current PC Migalastat[629]. Whilst no considerable changes in protein conformation were observed in complex with the cyclosulfamidate inhibitor, it should be noted that this co-crystal structure was obtained with wild-type r-αGAL, which likely already exists in a stable conformation. Therefore, its possible that the reported increase in protein melting temperature results from the ability of the cyclosulfamidate inhibitor to lock-in and stabilise the existing protein fold. Perhaps this inhibitor would induce changes in protein conformation/folding when complexed to mutant r-αGAL.

2.4.6.3 Structural Comparison of α-Cyclosulfamidate and Migalastat

To compare the binding of cyclosulfamidate **6** to the binding of Migalastat, a crystal structure of r-αGAL in complex with Migalastat was obtained at 2.3 Å resolution. Following refinement, sufficient electron density was observed to model a single molecule of Migalastat bound non-covalently in the active site, Figure 2.26 (a). Migalastat was modelled in the expected chair $^4C_1$ conformation, however, the lower resolution of this structure renders the ring conformation a little ambiguous. Nevertheless, Migalastat forms an extensive hydrogen bonding network within the active site; of note is the interaction between the endocyclic nitrogen atom and the catalytic nucleophile Asp170, Figure 2.26 (b). Importantly, this structure is consistent with the previously deposited structures of r-αGAL in complex with Migalastat (PDB 3GXT[595], 3S5Y[597]) but at improved resolution.

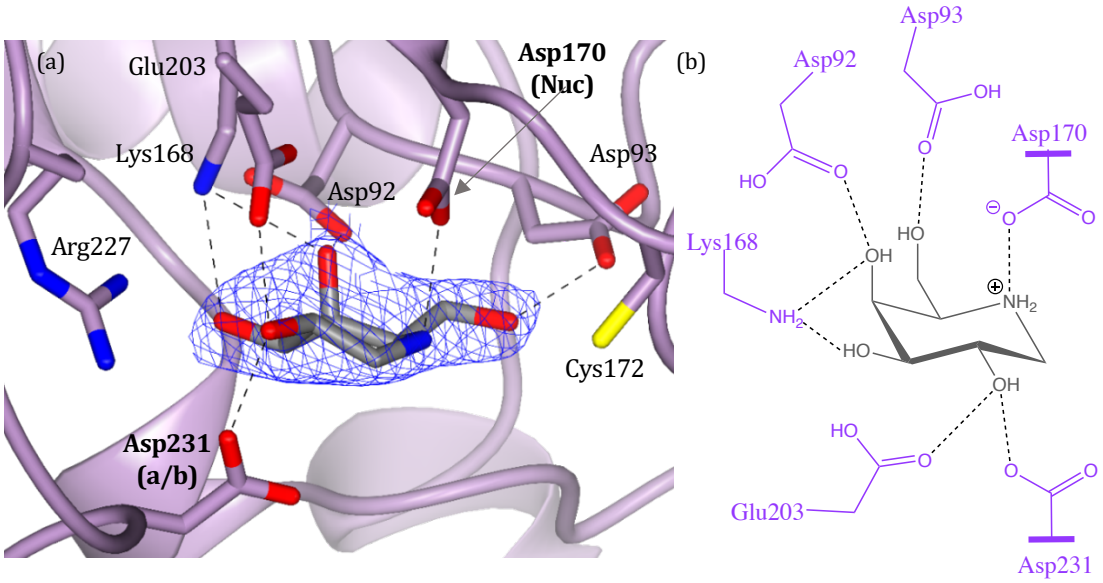


Figure 2.26: (a) 3D active site structure showing the non-covalent binding of Migalastat in the active site of r-αGAL. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand contoured to 1.3 σ (0.22 electrons/Å^3). (b) Schematic representation of hydrogen bonding network of Migalastat. The endo-cyclic nitrogen of Migalastat is likely protonated, due to the acidic pH of the crystallisation conditions.

For structural comparison, the Migalastat and cyclosulfamidate co-crystal structures were overlaid, revealing no substantial differences in tertiary or active site structure of r-αGAL, Figure 2.27. In fact, all active site residues adopt almost identical conformations, with both inhibitors binding in a similar non-covalent manner. The only notable difference lies in the hydrogen bonding network of the two inhibitors, Figure 2.27 (b). Whilst both inhibitors form hydrogen bonds with Asp92, Asp93, Lys168, Glu203 and Asp231, the cyclosulfamidate forms a hydrogen bonding interaction with Asp231 through its cyclosulfamidate nitrogen whereas Migalastat forms a hydrogen bond with Asp170 through the endocyclic nitrogen.

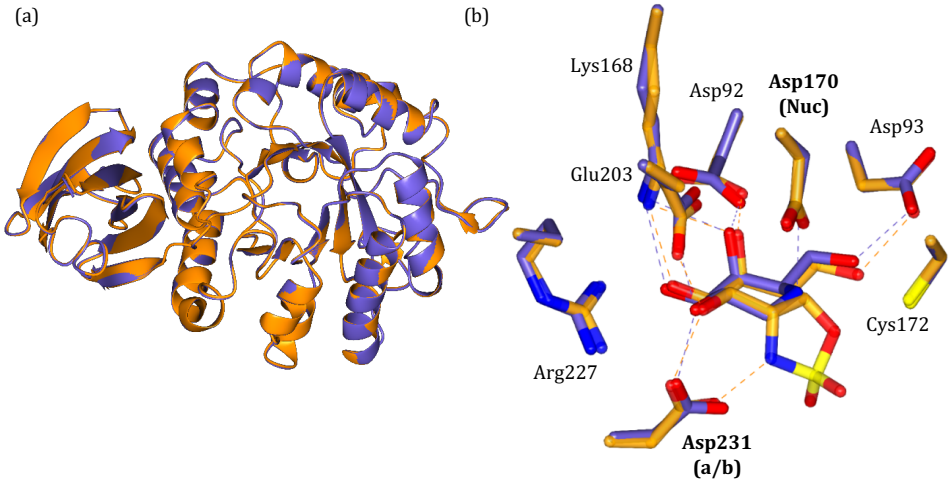
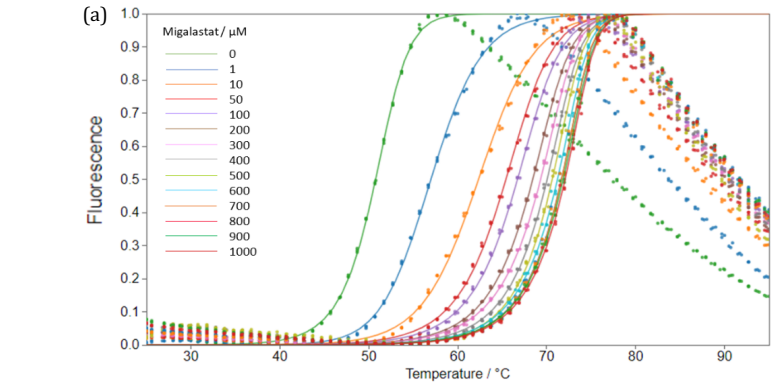


Figure 2.27: Overlay of Migalastat (purple) and cyclosulfamidate (orange) co-crystal structures showing similar (a) tertiary structures and (b) active site structures of r-αGAL.

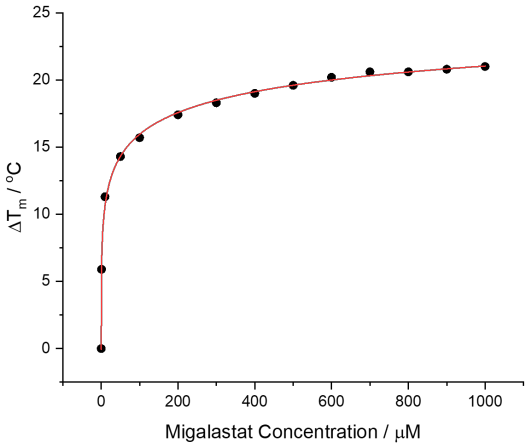
The hydrogen bonding interaction of Migalastat with Asp170 has proved vital not only to its binding but also to its controlled release when in the lysosome[597]. In a neutral environment, such as the ER, the endocyclic nitrogen is likely protonated whilst the carboxyl sidechain of Asp170 is likely deprotonated. This creates a strong ionic interaction which is thought to underpin the potent binding of Migalastat. On the contrary, under acidic conditions such as the lysosome, Asp170 is more likely to be protonated, losing the ionic interaction which weakens the binding and perhaps permits the displacement of Migalastat from the active site. Therefore, it has been proposed that the chaperoning behaviour of Migalastat is dependent upon the protonation state of its endocyclic nitrogen and the catalytic nucleophile. In light of this information, the hydrogen bonding interaction between the nitrogen of the cyclosulfamidate inhibitor and catalytic acid-base residue may also be important to the binding and chaperoning behaviour of this cyclosulfamidate inactivator. Overall, the lack of structural differences between the two structures suggests these inhibitors may bind and stabilise r-αGAL in a similar manner which does not involve enforcing a specific protein conformation.

2.4.6.4 *Comparison of the Chaperoning Behaviour of α-Cyclosulfamidate and Migalastat*

Following structural analysis, the chaperoning behaviour of Migalastat was evaluated through the same ThermoFluor analysis for comparison with cyclosulfamidate **6**. The resulting thermal shift curves showed a dramatic shift to higher temperatures with increasing Migalastat concentration, Figure 2.28 (a), demonstrating a considerable increase in Tm. For example, in the presence of 200 μM Migalastat, a 17.4 °C increase in Tm was observed compared with the 4.6 °C increase for 200 μM **6**. Subsequently, a plot of change in protein melting temperature (ΔTm) vs Migalastat concentration was constructed, Figure 2.28 (b), to reveal ΔTm_{max} and x_0 values of 30 ± 2.2°C and 66 ± 33 μM respectively. Comparing these values with those reported for cyclosulfamidate **6** (12.7 ± 1.1 °C and 372 ± 6 μM), suggests that Migalastat is better at binding and stabilising Fabrazyme® against thermal denaturation. However, the promising *in vitro* stabilisation of the cyclosulfamidate inhibitor (**6**) warranted further investigations into its stabilising effect *in situ*.



(b)



(c)

[Migalastat]/ μM	Tm/ °C	ΔTm/ °C
0	51.1	-
1	57.0	5.9
10	62.3	11.3
50	65.4	14.3
100	66.7	15.7
200	68.4	17.4
300	69.3	18.3
400	70.1	19.0
500	70.6	19.6
600	71.2	20.2
700	71.7	20.6
800	71.7	20.6
900	71.9	20.8
1000	72.1	21.0

Figure 2.28: Effect of Migalastat on the thermal stability of r-αGAL (a) Thermal stability curves recorded for Fabrazyme® pre-incubated with varying concentrations of Migalastat for 1 hour. (b) Plot of ΔTm versus Migalastat concentration fitted to a 4-parameter logistic curve. ΔTm_{max} = 30.0 ± 2.2 °C and x_0 = 66 ± 33 μM. Errors given are fitting errors. (c) Table of Tm and ΔTm observed at each Migalastat concentration.

2.4.7 Treatment of FD Fibroblasts with α-Cyclosulfamidate and Migalastat

All *in situ* work discussed in this section was performed by researchers at the Overkleeft Lab (Leiden University) but has been included to provide greater context for the structural and biochemical work described so far.

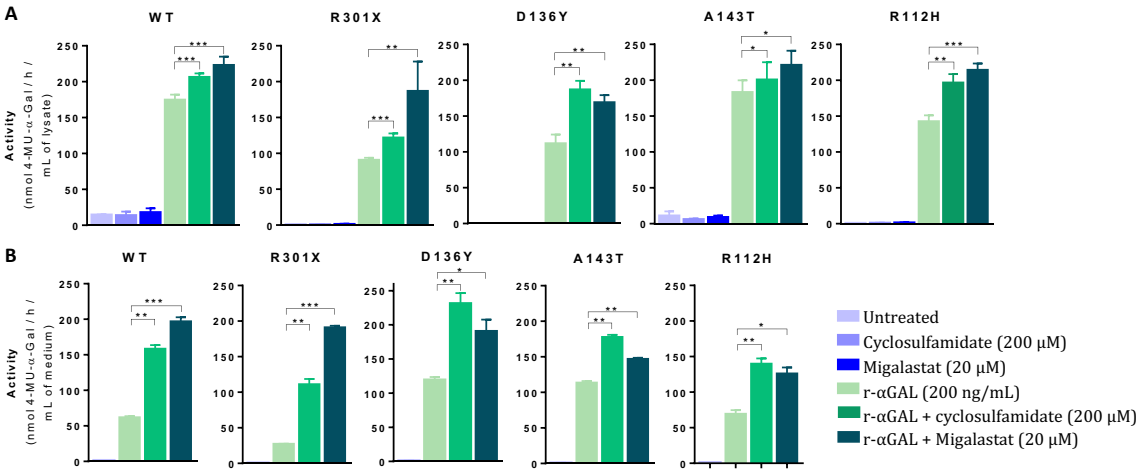


Figure 2.29: *In situ* stabilising effect of cyclosulfamidate and Migalastat (a) Fibroblasts of WT, classic Fabry (R301X and D136Y) and variant Fabry (A143T and R112H) were untreated or incubated with cyclosulfamidate, Migalastat, Fabrazyme®, or a combination of enzyme and inhibitor for 24 h. The medium was collected and α-GAL activity was measured in the cell homogenates by a 4-MU-α-gal assay. In all cell lines co-administration of cyclosulfamidate or Migalastat with Fabrazyme® increased intracellular α-GAL activity compared Fabrazyme® alone. (b) α-Gal A activity is at least two times higher in the cell culture medium of lines treated with cyclosulfamidate or Migalastat**.** Reported activities are mean ± stdev of two biological replicates, each with two technical replicates. Data and figure supplied by M. Artola (Leiden University).

Fibroblasts from 5 different Fabry cell lines (wild-type, classic Fabry R301X and D136Y, and variant Fabry A143T and 112H) were treated *in situ* with inhibitor **6** or Migalastat alone, Fabrazyme® alone and a combination of Fabrazyme® and inhibitors (**6** at 200 μM or Migalastat at 20 μM), Figure 2.29. The activity of α-GAL in cell lysates and cell culture medium was then quantified using a fluorogenic substrate assay for each cell line. Treatment with inhibitor **6** or Migalastat alone for 24 hours had no effect on intra-cellular α-GAL activity. However, compared to cells treated with Fabrazyme® alone, a notable increase in intracellular activity was observed in cells treated with a combination of Fabrazyme® and either inhibitor, Figure 2.29 (a). This increase in intracellular activity was also accompanied by a 2-fold increase in activity in the cell culture medium, Figure 2.29 (b).

This suggests that inhibitor **6** binds and stabilises Fabrazyme® in the cell culture medium, leading to an increase in the amount of active enzyme available for uptake. Subsequently, this stabilised enzyme is internalised by the cells where the reversible inhibitor dissociates to enhance the intracellular activity. Interestingly, the stabilising effect of inhibitor **6** was found to be time dependent, being more pronounced with longer incubation times. In fact, α-GAL activity was found to be 1.5-2 times higher in fibroblasts treated with a combination of Fabrazyme® and inhibitor **6** for 4 days than those treated for 24hrs. Importantly, when co-administered with inhibitor **6** (or Migalastat), similar α-GAL activity and subsequent correction of toxic metabolites (Gb3 and lyso-Gb3) was achieved using half the concentration of Fabrazyme® . This demonstrates the synergy between Fabrazyme® and Migalastat and supports the use of other pharmacological chaperones, such as the cyclosulfamidate inhibitor **6,** in combination therapy with ERT to reduce the amount of recombinant enzyme required. These cell culture experiments are concordant with the *in vitro* ThermoFluor assay results and suggest that cyclosulfamidate inhibitor (**6**) may behave as a molecular chaperone for r-αGAL.

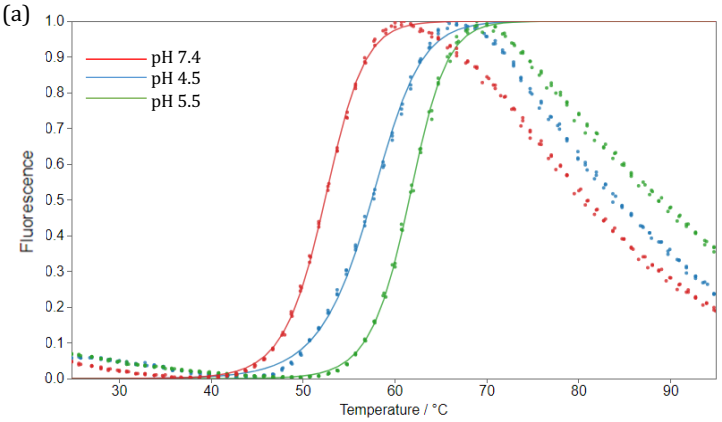
2.4.8 Effect of pH on the Chaperoning Behaviour of α-Cyclosulfamidate and Migalastat

The improved efficacy of combination therapy was recently demonstrated through the co-administration of Migalastat and Fabrazyme®[510]. An important advantage of co-administering ERT with a stabilising pharmacological chaperone (PC) is its potential to be effective for all Fabry patients independent of their α-GAL mutation. Furthermore, the use of PCs may allow the recombinant enzyme dose to be reduced (as demonstrated in the cell culture experiments by our collaborators in Leiden, Figure 2.29) or permit extended intervals between injections. This has the potential to improve therapeutic efficacy, reduce side effects and decrease treatment costs. However, the main drawback of Migalastat as a PC is that it is reported to stabilise r-αGAL at both neutral and acidic pH[522,626]. This suggests that Migalastat does not completely dissociate from the enzyme active site when it reaches the acidic environment of the lysosome, which is detrimental to its chaperoning effect. Ideally, a PC for r-αGAL should stabilise the enzyme at neutral pH during transit through the ER but not at acidic pH when it reaches the lysosome. Consequently, the stabilising effect of the cyclosulfamidate inhibitor was investigated at neutral and acidic pH to determine whether its chaperoning behaviour is pH dependent. If significantly reduced stabilisation of r-αGAL is observed at acidic pH, then this compound may be a superior chaperone.

2.4.8.1 *pH Dependent Stability of r-αGAL*

Firstly, the thermal stability of Fabrazyme® was characterised at pH 4.5 and pH 5.5, to mimic lysosomal pH ranges, and also at pH 7.4 using a McIlvaine buffer system. McIlvaine is a universal buffer composed of citric acid and disodium hydrogen phosphate which can operate over a pH range of 2-8. Therefore, all ThermoFluor assays could be performed under the same buffer system. The thermal shift curves obtained for the thermal denaturation of Fabrazyme® at pH 4.5, 5.5 and 7.4 demonstrate an inherent difference in thermal stability with varying pH, Figure 2.30 (a). Fabrazyme® appears most stable at pH 5.5, with an average Tm of 62.4 ± 1.4 °C, and least stable at pH 7.4 with an average Tm of 52.6 ± 0.5 °C, Figure 2.30 (b). This pH dependent thermal stability is consistent with α-GAL being a lysosomal enzyme which operates in the acidic environment of the lysosome. Furthermore, these melting temperatures are concordant with those reported in literature[522,595,626].

(a)



(b)

pH	Tm / °C
4.5	57.8 ± 0.4
5.5	62.4 ± 1.4
7.4	52.6 ± 0.5

Figure 2.30: pH dependent thermal stability profile of r-αGAL (a) Thermal stability curves recorded for Fabrazyme® at pH 4.5, 5.5 and 7.4, demonstrating an inherent difference in thermal stability at each pH. (b) Calculated Tm of Fabrazyme® at pH 4.5, pH 5.5 and pH 7.4. Values reported as average ± the stdev of 4 replicates.

2.4.8.2 *Investigating the pH Dependence of r-αGAL Stabilisation by α-Cyclosulfamidate*

Once the pH dependent thermal stability profile of Fabrazyme® had been established, the effect of pH on the stabilisation of Fabrazyme® by the cyclosulfamidate inhibitor (**6**) was investigated. Fabrazyme® was incubated with varying concentrations of inhibitor **6** at pH 4.5, 5.5 and 7.4 for 1 hour before repeating the ThermoFluor assay outlined previously. The resulting thermal shift curves were used to calculate the Tm of Fabrazyme® at each inhibitor concentration at each pH, Figure 2.31. The change in Fabrazyme® Tm (ΔTm) was then determined relative to a control with no inhibitor.

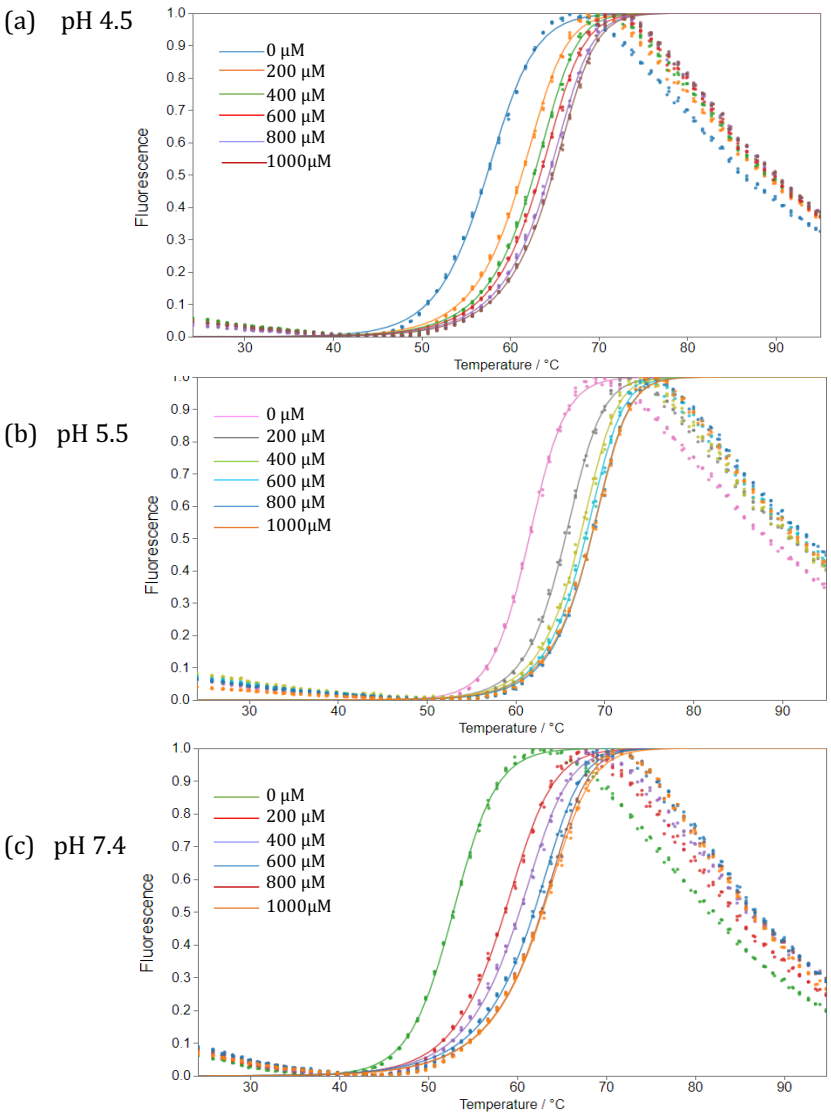
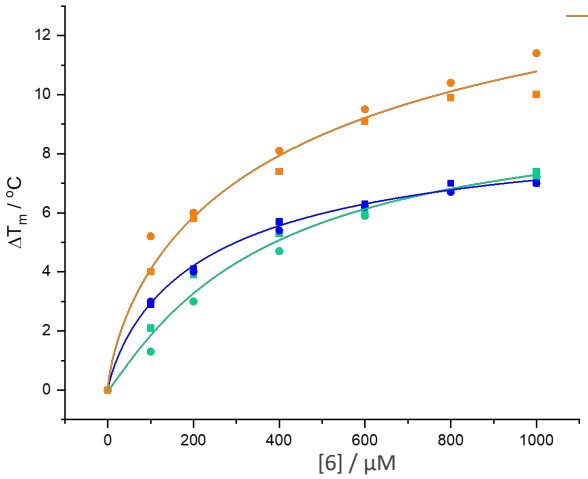


Figure 2.31: Thermal shift curves obtained for Fabrazyme® pre-incubated with varying concentrations of the cyclosulfamidate inhibitor for 1 hour at pH (a) 4.5 (b) 5.5 (c) 7.4. Larger stabilising effect against thermal denaturation evident at pH 7.4.

For a given inhibitor concentration, a comparison of the average ΔTm at each pH indicates that the stabilising effect of inhibitor **6** is greater at neutral pH (pH 7.4) than acidic pH (pH 4.5 and pH 5.5), Figure 2.32 (b). Little difference in the thermal stabilisation of Fabrazyme® was observed between pH 4.5 and pH 5.5, however, a pronounced increase in stabilisation was observed at pH 7.4. For example, in the presence of 1000 µM inhibitor **6**, the Tm of Fabrazyme® increased by 7.3 °C and 7.0 °C at pH 4.5 and pH 5.5 respectively, however, a 10.7 °C increase was observed at pH 7.4. Moreover, this greater increase in Tm at neutral pH was observed consistently across all the inhibitor concentrations, Figure 2.32 (b). Consequently, a plot of ΔTm versus cyclosulfamidate concentration was constructed at each pH and the average fitted to a 4-parameter logistic function, Figure 2.32 (a). From this plot, ΔTm_{max} was calculated to be 9.7 °C ± 1.2 °C, 9.3 °C ± 1.0 °C and 17.4 °C ± 4.1 °C at pH 4.5,

5.5 and 7.4 respectively. These values further demonstrate the enhanced stabilising effect of inhibitor **6** at neutral pH compared to acidic pH, which is desirable for an r-αGAL chaperone. However, the fact some stabilisation was observed at acidic pH suggests that **6** may not fully dissociate from the enzyme when it reaches the lysosome, as has been reported for Migalastat[522,626].

(a)



(b)

	ΔTm / °C		
[Inhibitor] / μM	pH 4.5	pH 5.5	pH 7.4
100	1.7 ± 0.3	3.0 ± 0.1	4.6 ± 0.5
200	3.5 ± 0.4	4.1 ± 0.1	5.9 ± 0.1
400	5.0 ± 0.2	5.6 ± 0.1	7.8 ± 0.3
600	6.0 ± 0.1	6.3 ± 0.1	9.3 ± 0.2
800	6.9 ± 0.1	6.9 ± 0.1	10.2 ± 0.2
1000	7.3 ± 0.1	7.0 ± 0.0	10.7 ± 0.6

Figure 2.32: (a) Plot of ΔTm versus cyclosulfamidate inhibitor (**6**) concentration at pH 4.5, 5.5 and pH 7.4. Average of duplicate experiments fitted to a 4-parameter logistic curve; pH 4.5 ΔTm_{max} = 9.7 °C ± 1.2 °C, pH 5.5 ΔTm_{max} = 9.3 °C ± 1.0 °C, and pH 7.4 ΔTm_{max} = 17.4 °C ± 4.1 °C. Errors given are fitting errors (b) Average ΔTm values calculated for each concentration of **6** at each pH. Values reported as average ± stdev of duplicate experiments.

2.4.8.3 Investigating the pH Dependence of r-αGAL Stabilisation by Migalastat

For comparison, the effect of pH on the stabilisation of Fabrazyme® by Migalastat was also investigated, Figure 2.33. Contrary to the literature, Migalastat was shown to stabilise Fabrazyme® better at neutral pH than acidic pH. For example, in the presence of 1000 μM Migalastat, the Tm of Fabrazyme® increased by 16.9 °C at pH 4.5, whist a 23.2 °C increase in Tm was observed at pH 7.4, Figure 2.34. This greater increase in Tm at neutral pH was observed consistently across the range of Migalastat concentrations tested. Consequently, a plot of ΔTm versus Migalastat concentration was constructed at each pH and fitted to a 4-parameter logistic function, Figure 2.34 (a). At pH 4.5, 5.5 and 7.4 the ΔTm_{max} was calculated to be 21.1 °C ± 4.0 °C, 22.3 °C ± 1.4 °C and 34.5 °C ± 9.0 °C respectively. Therefore, in contrast to the literature which reports no pH dependent stabilisation by Migalastat[522,626], these thermal stability results demonstrate that under these experimental conditions Migalastat stabilises Fabrazyme® better at neutral pH. This discrepancy may arise from the use of

different buffering systems. During optimisation of the ThermoFluor assay in this work, it was found that changing the buffer system could affect the extent of stabilisation by both Migalastat and inhibitor **6**. In the literature, a number of different buffers were used, including HEPES[626], sodium acetate[626] and sodium phosphate[522] buffers, whilst in this work the McIlvaine buffer system was used. Furthermore, the thermal shift assays reported in literature were performed using different buffers at different pH values, suggesting the thermal stability results may have been subject to interference from the buffer components[626].

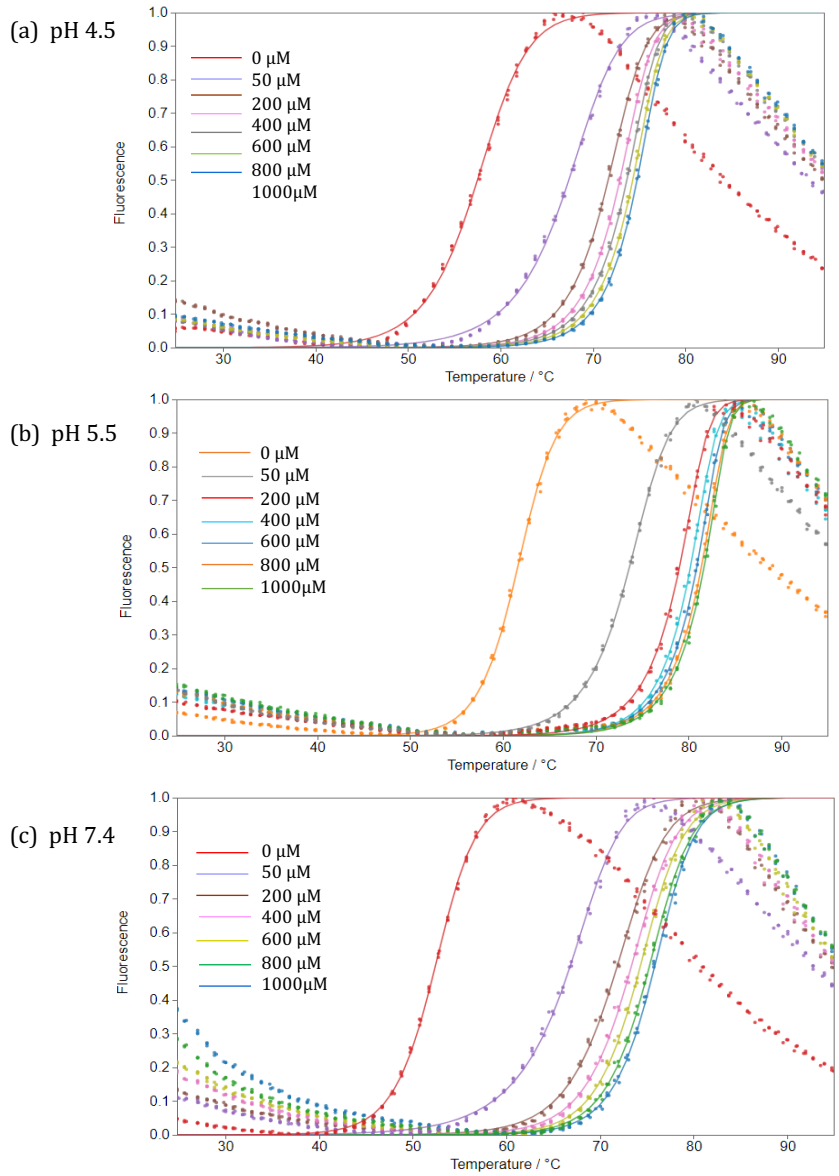
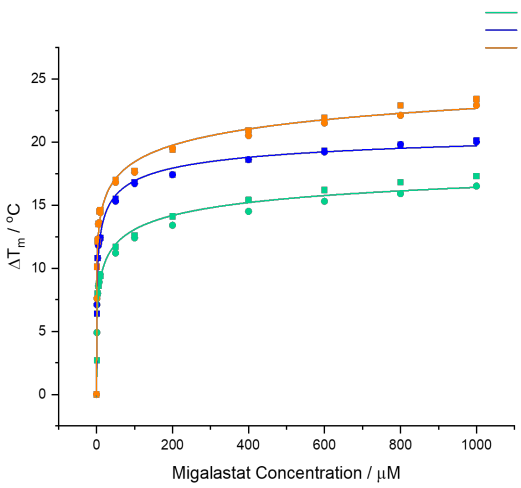


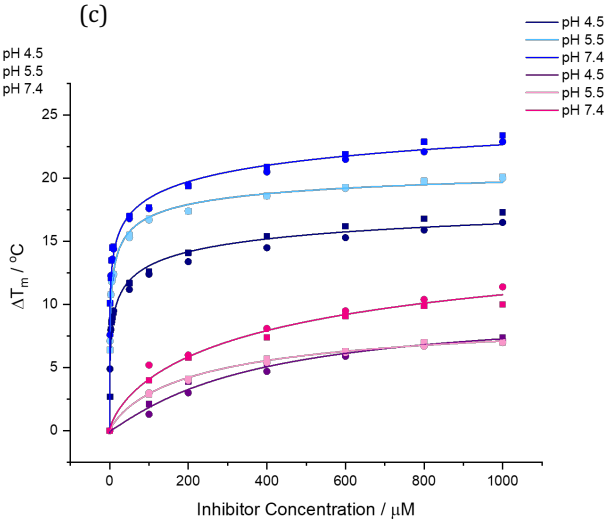
Figure 2.33: Thermal shift curves obtained for Fabrazyme® pre-incubated with varying concentrations of Migalastat for 1 hour at (a) pH 4.5 (b) pH 5.5 (c) pH 7.4. Larger stabilising effect against thermal denaturation evident at pH 7.4.

A comparison of the ΔTm_{max} values and pH dependent chaperoning profiles of cyclosulfamidate **6** and Migalastat, Figure 2.34 (c), suggest both compounds stabilise r-αGAL against thermal denaturation at neutral and acidic pH; but both are more effective stabilisers at neutral pH. Although the stabilising effect of Migalastat appears more prominent, Figure 2.34 (c), the cyclosulfamidate inhibitor shows considerable potential as a PC for FD because it significantly stabilises r-αGAL against thermal denaturation and exhibits a similar pH dependent profile to that of Migalastat both *in vitro* and *in-situ*.

(a)



(c)



(b)

	ΔTm / °C		
[Migalastat]/μM	pH 4.5	pH 5.5	pH 7.4
1	3.8 ± 0.9	6.7 ± 0.3	8.8 ± 1.0
3	8.0 ± 0.0	10.8 ± 0.0	12.2 ± 0.1
5	8.6 ± 0.0	11.8 ± 0.1	13.5 ± 0.0
7	9.0 ± 0.1	12.1 ± 0.1	14.0 ± 0.4
10	9.5 ± 0.1	12.4 ± 0.0	14.5 ± 0.1
50	11.5 ± 0.2	15.5 ± 0.1	16.9 ± 0.1
100	12.5 ± 0.1	16.7 ± 0.1	17.6 ± 0.1
200	13.8 ± 0.3	17.4 ± 0.0	19.5 ± 0.1
400	15.0 ± 0.4	18.6 ± 0.0	20.7 ± 0.2
600	15.8 ± 0.4	19.2 ± 0.1	21.7 ± 0.2
800	16.4 ± 0.4	19.8 ± 0.0	22.5 ± 0.3
1000	16.9 ± 0.3	20.0 ± 0.1	23.2 ± 0.2

Figure 2.34: (a) Plot of ΔTm versus Migalastat concentration at pH 4.5, 5.5 and 7.4. Average of duplicate experiments fitted to a 4-parameter logistic function; pH 4.5 ΔTm_{max} = 21.1 °C ± 4.0 °C, pH 5.5 ΔTm_{max} = 22.3 °C ± 1.4 °C, and pH 7.4 ΔTm_{max} = 34.5 °C ± 9.0 °C (b) Average ΔTm values calculated for each Migalastat concentration at each pH (± stdev of duplicate experiments). (c) Comparative plot of change in protein melting temperature (ΔTm) versus inhibitor concentration at pH 4.5, 5.5 and 7.4 for Migalastat (blue) and cyclosulfamidate inhibitor **6** (pink).

2.5 Summary

In this work, recombinant human α-galactosidase A (Fabrazyme®) was crystallised and the binding of a number of α-*galacto* configured cyclophellitol based inhibitors (developed by the Overkleeft lab, Leiden University) was investigated through crystallographic studies to address the lack of structural data on these compounds.

The classic α-*galacto* configured cyclophellitol epoxide and aziridine compounds were shown to inhibit r-αGAL by covalent modification of the catalytic nucleophile to form an enzyme-inhibitor complex in the expected skew-boat ($^{1}S_{3}$) conformation. The *N*-acyl functionalised aziridine inhibitor was also shown to covalently inhibit r-αGAL by the same mechanism-based mode of action. Surprisingly, the analogous *N*-alkyl functionalised aziridine inhibitor was found to bind non-covalently in the active site. Although there is no obvious reason as to why this *N*-alkyl functionalised aziridine inhibitor did not covalently react, this observation may account for the reduced labelling efficiency reported for *N*-alkyl *galacto*-configured cyclophellitol aziridine ABPs in comparison to the *N*-acyl equivalents. Additionally, the co-crystal structure obtained with the new *galacto*-configured cyclosulfate inhibitor, revealed covalent inhibition of r-αGAL through the anticipated mechanism in which the cyclosulfate moiety is ring opened to covalently modify the catalytic nucleophile. This demonstrates the potential to expand the suite of *galacto*-configured cyclophellitol-based inhibitors for αGAL by the use of alternative electrophilic warheads.

In contrast to the covalent inhibitors, the novel cyclosulfamidate analogue was found to bind non-covalently in the active site of r-αGAL adopting the chair ($^{4}C_{1}$) conformation. In preliminary thermal stability studies, this inhibitor was shown to stabilise r-αGAL against thermal denaturation, with a greater stabilising effect at neutral pH. Consequently, our collaborators in Leiden demonstrated that this inhibitor is capable of increasing r-αGAL activity *in-situ* in Fabry Disease (FD) fibroblasts and allow for partial correction of accumulated toxic metabolites. These results indicate that *galacto*-configured α-1,6-cyclophellitol cyclosulfamidate exhibits chaperoning behaviour towards r-αGAL and may prove valuable in the development of a new class of pharmacological chaperones for the treatment of FD.

Chapter 3: Activity-Based Probes and Inhibitors for β-Glucocerebrosidase: Generation of Gaucher Animal Models

3.1 Abstract

Gaucher disease (GD) is the most common lysosomal storage disorder caused by inherited deficiencies in lysosomal β-glucocerebrosidase (GBA). The development of novel therapeutic strategies for this disorder has been somewhat limited by the lack of reliable Gaucher animal models. Such models have previously been generated through chemical knockdown of GBA activity using the β-glucosidase inhibitors conduritol B epoxide (CBE) and cyclophellitol, however, the broad specificity of these inhibitors has hindered their use. In this work, a range of C6-functionalised cyclophellitol-based inhibitors and activity-based probes (ABPs) are structurally investigated as more selective GBA inhibitors for use in chemical knockdown studies. Structural analysis of a bi-functional cyclophellitol ABP is also described, highlighting the potential for future ABP and inhibitor development.

Initial crystallographic studies of GBA in complex with C6 fluorescently tagged cyclophellitol ABPs revealed a hydrophobic pocket at the dimer interface of GBA which is capable of binding bulky hydrophobic moieties. This hydrophobic cavity appears unique to GBA, providing a structural basis for the improved selectivity of C6-functionalised inhibitors. Consequently, new C6-functionalised cyclophellitol inhibitors were synthesised, which proved to be more potent and selective for GBA than CBE and cyclophellitol. Structural evaluation of a C6 adamantane-substituted cyclophellitol inhibitor revealed binding of the adamantyl group to the unique hydrophobic pocket, supporting the concept that C6-functionalisation of cyclophellitol inhibitors improves the potency and selectivity for GBA by taking advantage of its unique hydrophobic pocket. Consequently, our collaborators in Leiden were able to use this inhibitor in chemical knockdown studies to create *Dario rerio* zebrafish GD models with the potential to explore neuropathic disease.

*Some of the work discussed in this chapter is published in (**1**) M. Artola, C-L. Kuo, L. T. Lelieveld, R. J. Rowland, G. A. van der Marel, J. D. C. Codee, R. G. Boot, G. J. Davies, J. M. F. G. Aerts, H. S. Overkleeft, *J. Am. Chem. Soc.* 2019, **141**, 4214-4218 and (**2**) R. J. Rowland, Y. Chen, I. Breen, L. Wu, W. Offen, T. J. Beenakker, A. M. C. H. can den Nieuwendijk, M. Artola, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, *Chem. Eur. J.,* 2021, **27**, 16377-16388.

3.2 Introduction

3.2.1 β-Glucocerebrosidase (GBA)

β-Glucocerebrosidase (GBA, EC 3.2.1.45) is a 497 amino acid membrane-associated lysosomal enzyme belonging to the GH30 family of retaining β-glucosidases[630,631]. GBA is primarily responsible catalysing the degradation of glucosylceramide (GlcCer) and glucosylsphingosine (GlcSph) by hydrolytic cleavage of the terminal β-linked glucose unit from the sphingolipid aglycone[172,632], Figure 3.1. Deficiencies in GBA activity, and subsequent glycolipid accumulation, underpin the lysosomal storage disorder Gaucher Disease (GD)[172,173,174]. Aside from fulfilling a role in lysosomal glycolipid catabolism, GBA also supports optimal skin function by generating free ceramide which is a key component of the outer skin[633,634].

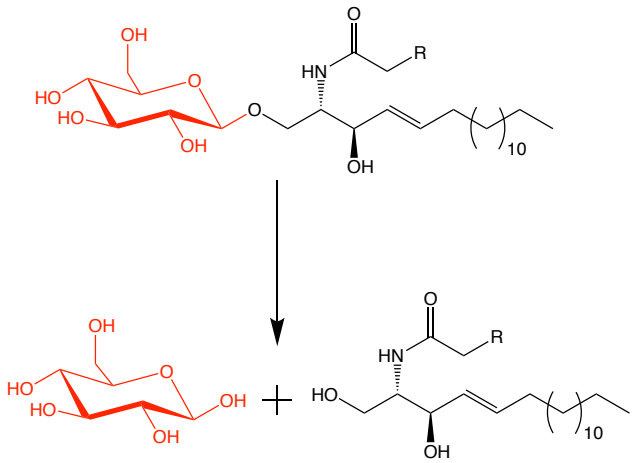


Figure 3.1: Hydrolysis of glucosylceramide into ceramide and glucose by GBA. Terminal β-linked glucose moiety is cleaved with retention of β-anomeric stereochemistry.

GBA is encoded by the *GBA1* gene, a 7.5 kb gene located on chromosome 1q21[185]. Initially, GBA is synthesised in the endoplasmic reticulum (ER) as a 58 kDa polypeptide with a 2 kDa secretory signal sequence[635]. This signalling peptide is cleaved during transit through the ER membrane where the protein is co-translationally glycosylated to produce a 63 kDa high-mannose precursor[185,635,234]. This precursor is subsequently transferred to the Golgi body apparatus where the oligosaccharides undergo further modification to yield mature GBA as a 56 kDa complex-type glycoprotein[185,635,234]. Unlike many other lysosomal enzymes, GBA is not phosphorylated and does not contain mannose-6-phosphate residues for trafficking to the lysosome via the M6P pathway[140]. Instead, transport of nascent GBA to the lysosome is mediated by the lysosomal membrane protein LIMP-2[287,140,309]. To achieve maximal activity, the 80 amino acid activator protein, Saposin C, is also required[92].

3.2.1.1 CAZy Glycoside Hydrolase 30 (GH30) Family

The GH30 family belongs to the GH-A CAZy clan along with 24 other GH families sharing the characteristic $(\beta/\alpha)_8$ TIM barrel fold and C-terminal active site[636]. Following work by St John et al. (2010)[637] on the phylogenetic analysis of GH5 family members, a number of GH5 subfamilies, including *endo*-β-galactanase, xylanase and glucuronoxylan xylan hydrolase enzymes, were reassigned to the GH30 family based on sequence and structural similarities. Currently, the GH30 family comprises of 9 subfamilies containing enzymes from bacteria and eukaryotes with known *endo*-β-1,4-xylanase, β-glucosidase, β-glucuronidase, β-xylosidase, β-fucosidase, glucosylceramidase, β-1,6-glucanase, glucuronoarabinoxylan *endo*-β-1,4-xylanase and *endo*-β-1,6-galactanase activity. Of all the GH30 enzymes, human GBA is the most studied, likely resulting from its pathogenic involvement in GD. Members of the GH30 family are anomeric configuration-retaining enzymes which operate through a double-displacement catalytic mechanism, Figure 3.2.

3.2.1.2 Catalytic Mechanism

Although retention of β-stereochemistry by GH30 family members has not been formally demonstrated through NMR analysis, covalent trapping of GBA with a 2-deoxy-2-fluoro-β-D-glucopyranosyl fluoride inhibitor unambiguously demonstrated the retaining nature of this family of enzymes[611]. Initially, the catalytic nucleophile of GBA was incorrectly identified as Asp443 through inhibition studies with ^{3}H-labelled bromo-conduritol B epoxide (Br-CBE)[638]. However, in a later study, trapping of the enzyme with a 2,4-dinitrophenyl-β-D-2-deoxy-2-fluoroglucopyranoside (2,4,DNP-2F-Glc) inhibitor revealed Glu340 to be the true catalytic nucleophile[611]. The catalytic acid-base residue was later identified as Glu235 through site directed mutagenesis[639].

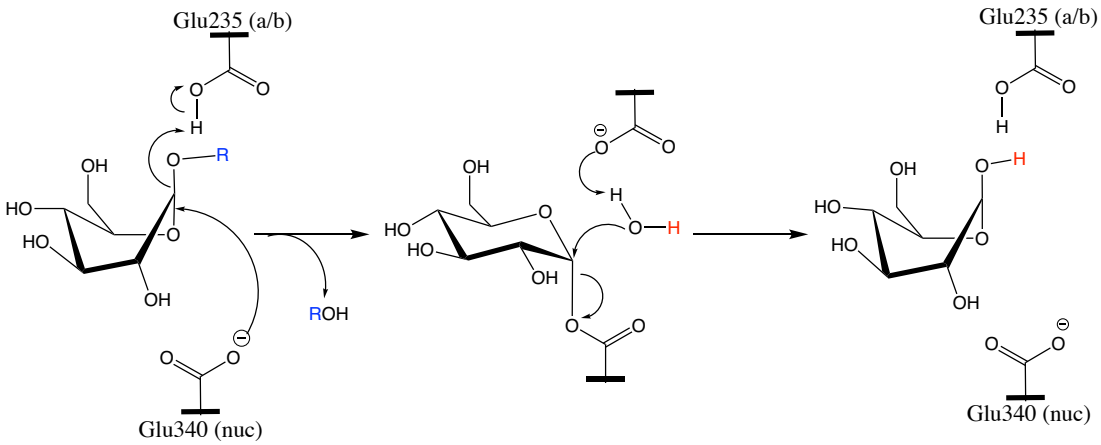


Figure 3.2: Koshland double-displacement mechanism of GBA using Glu340 as the catalytic nucleophile (Nuc) and Glu235 as the general acid/base residue (a/b).

Throughout the double displacement mechanism, the carbohydrate substrate undergoes substantial changes in conformation to satisfy orbital overlap and stereochemical requirements. According to a Michaelis complex → [Transition State]‡ → Covalent intermediate reaction coordinate, GH30 enzymes employ a $^{1}S_{3}$ → $^{4}H_{3}$‡ →$^{4}C_{1}$ conformational itinerary[560], Figure 3.3. Upon binding to the enzyme, the substrate adopts a $^{1}S_{3}$ conformation to ensure the aglycon leaving group is axially positioned for in-line nucleophilic attack[560]. On nucleophilic attack, a high energy transition state is formed in a $^{4}H_{3}$ conformation which is stabilised by lone-pair delocalisation over the partial O5-C1 double bond[560,640]. Subsequently, a covalent glycosyl-enzyme intermediate is formed in a $^{4}C_{1}$ conformation[560].

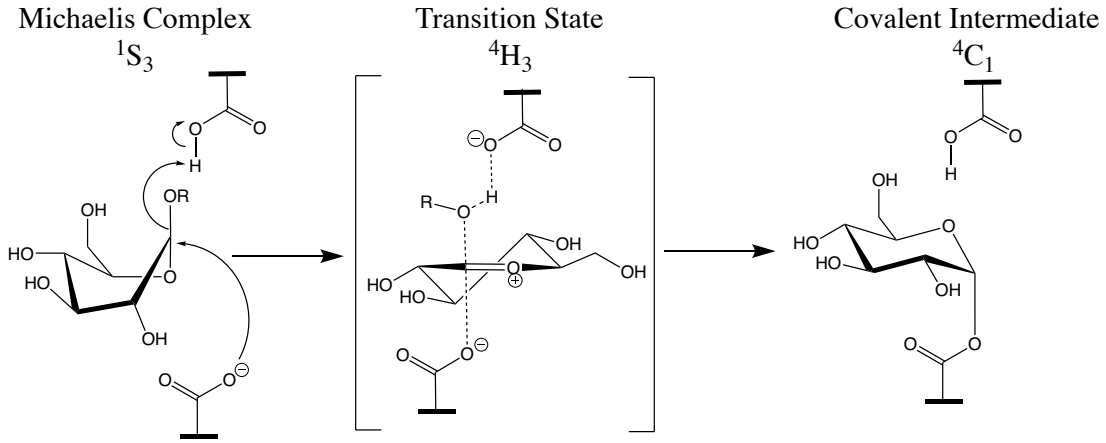


Figure 3.3: The Michaelis complex → [Transition State]‡ → Covalent intermediate itinerary of GH30 β-glucosidases following $^{1}S_{3}$ → $[^{4}H_{3}]^{‡}$ → $^{4}C_{1}$ conformations.

3.2.1.3 Structure of Human GBA

In 2003, the Sussman group solved the first ever structure of human GBA to 2.0 Å (PDB 1OGS)[641] using a mercury heavy atom derivative of the recombinant enzyme Cerezyme® (Sanofi Genzyme, USA). To aid crystallisation, Cerezyme® was treated with *N*-glycosidase F to remove the oligosaccharides before crystallising using hanging-drop vapour diffusion[641].

The refined x-ray structure revealed human GBA to be a homodimeric glycoprotein which, following removal of the 40-amino acid signalling sequence, consists of two 497-residue monomers[641], Figure 3.4 (a). Each monomer comprised of three non-contiguous domains; a three-stranded anti-parallel β-sheet domain flanked by a perpendicular amino-terminal strand and loop, an immunoglobulin-like (Ig) domain consisting of two closely associated β-sheets, and a characteristic $(\beta/\alpha)_{8}$ TIM barrel domain, Figure 3.4 (b). The active site was located at the centre of the β-barrel, demonstrating that Asp443, which was originally proposed as the catalytic nucleophile, cannot be directly involved in catalysis because it is located within the immunoglobulin like domain far away from the active site. The average

distance between the two catalytic carboxylate residues of the catalytic dyad was estimated to be ~5.2 Å, which is consistent with the average distance (~5.3 Å) observed for retaining β- glucosidases[582,590].

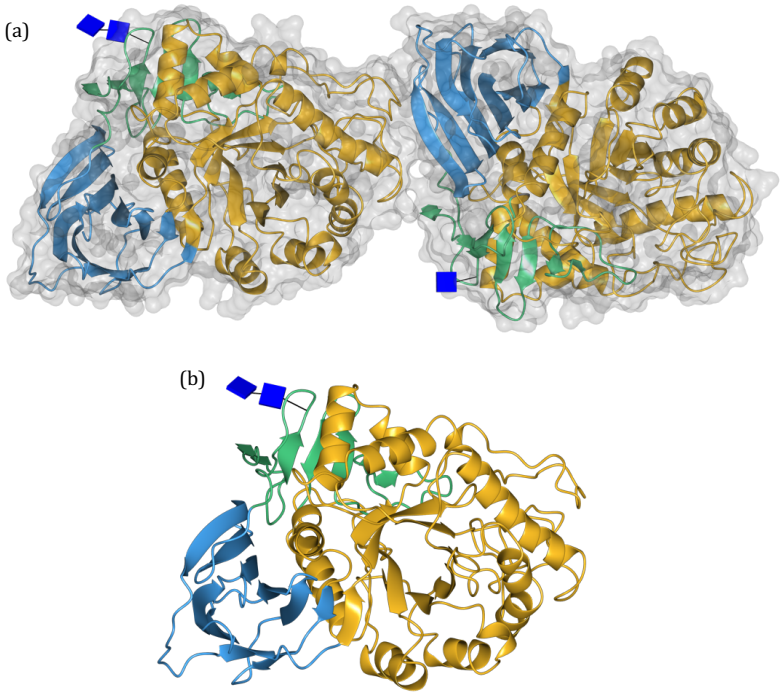


Figure 3.4: First structure of GBA solved by the Sussman group[641]. (a) Surface and ribbon diagram of GBA dimer with a single *N*-glycosylation site at Asn19. (b) Ribbon diagram of GBA monomer with a three-stranded anti-parallel β-sheet domain (green), an immunoglobulin like domain (blue) and a $(\beta/\alpha)_8$ TIM barrel domain shown (yellow). *N*-linked glycans are illustrated using Glycoblocks[591]. Figure created in CCP4mg[94] using deposited 1OGS PDB coordinates.

Several polar residues have been identified in and around the active site which stabilise the active site and hold the substrate in place through an extensive hydrogen bonding network[368]. Indeed, most active site residues are reportedly sensitive to ligand binding and are capable of adopting multiple conformations depending on the substrate bound[642]. Additionally, numerous aromatic residues (Phe128, Trp179, Tyr244, Phe246, Tyr313, Trp381 and Phe397) were found to line one side of the active site pocket, potentially controlling substrate recognition and specificity. Further to this, substrate docking studies with the natural GlcCer substrate, revealed that only the glucose moiety can be accommodated in the active-site pocket[641]. The binding of the fatty acid portion remains poorly understood, however, the fatty acid chains are thought to project out from GBA and embed into the lipid bi-layer or interact with the activator protein, saposin C (SapC)[641]. This is supported by the identification of a ring of hydrophobic residues around the active-site entrance which may facilitate interactions of GBA with the lysosomal membrane or SapC[641].

This first structure of GBA was partially deglycosylated, exhibiting a single *N*-linked glycosylation site at Asn19[641]. However, a crystal structure of fully glycosylated Cerezyme® (PDB 2J25) was later obtained and the protein structure was found to be virtually identical, alleviating concerns that partial deglycosylation might alter the 3D structure of GBA[643]. Interestingly, structural alignment of this Cerezyme® crystal structure with that of xylanase (PDB 1NOF)[644], a bacterial GH reassigned to the GH30 family, shows a high level of similarity between the active-site structures, Figure 3.5. Specifically, this alignment yields nine identical residues and two similar residues within 5 Å of the catalytic dyads; Glu340 and Glu235 for Cerezyme® and Glu253 and Glu165 for xylanase. This structural comparison not only supports the reassignment of xylanase from the GH5 family to the GH30 family but also demonstrates the structural similarity observed across GH30 family members.

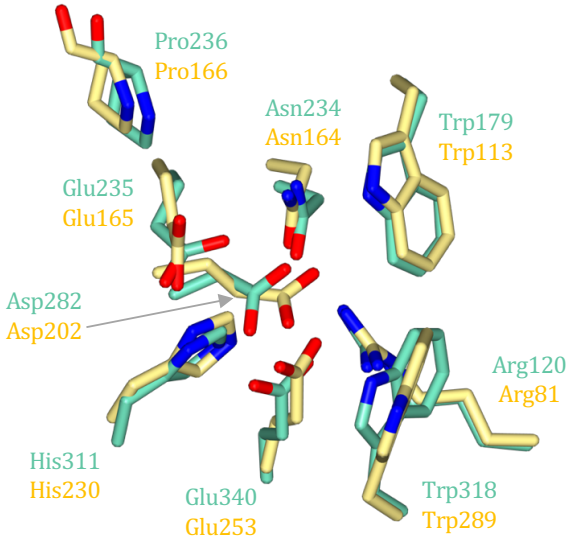


Figure 3.5: Structural alignment of the active sites of Cerezyme® (PDB 2J25)[643] in green and bacterial xylanase (PDB 1NOF)[644] in yellow. The nine identical active site residues which occupy almost identical conformations are displayed, showing high correspondence between the two active sites. Figure created in CCP4mg[94].

Since the first structure of GBA obtained in 2003, numerous structures of GBA in complex with a variety of covalent and non-covalent ligands have been reported, with 41 crystal structures of human GBA deposited in the PDB to date, Table 3.1.

3.2.1.4 N-Glycosylation of GBA

In an initial composition study, Takasaki et al. (1984)[305] characterised the oligosaccharide composition of human placental GBA, revealing *N*-linked high mannose-type oligosaccharides and a series of complex-type triantennary and biantennary oligosaccharides. In contrast to most other lysosomal enzymes, GBA was shown to exhibit incomplete outer chains, with no phosphorylation or mannose-6-phosphate residues[305].

This observation led to the hypothesis that GBA is transported to the lysosome via a M6P independent pathway, which was later identified to be mediated by LIMP-2[287,140,309].

Five potential *N*-glycosylation sites have been identified for GBA (Asn19, Asn59, Asn146, Asn270 and Asn462) all of which are conserved between the human and murine enzyme[645]. However, only the first four *N*-glycosylation sites are thought to be occupied[646]. Although post-synthesis modification of the *N*-glycan composition has little effect on the stability or activity of GBA[647], inhibiting initial glycan synthesis by expressing human cDNA in bacterial systems (or using tunicamycin-treated insect cells), results in the generation of catalytically inactive enzyme forms[647]. This demonstrates that the glycosylation of GBA is essential for the synthesis and maintenance of a catalytically active conformer. In particular, occupancy of the Asn19 glycosylation site has proved vital for the production of catalytically-active GBA by facilitating co-translational folding of the enzyme during synthesis[648,219]. Additionally, molecular dynamic simulations have shown that occupancy of the Asn19 glycosylation site lowers the root-mean-squared deviation (RMSD) of the catalytic residues Glu340 and Glu235, bringing them closer together and potentially stabilising the catalytic dyad[649]. In contrast, the other four glycosylation sites are thought to aid proteolytic stability but are not essential for catalytic activity[648].

In light of these glycosylation requirements, it became apparent that GBA cannot be produced in prokaryotic systems and more complex eukaryotic expression systems which retain all the necessary post-translational modifications are required[647,650]. For example, Cerezyme®®, one commercially available form of recombinant GBA (rGBA), is expressed in a Chinese hamster ovary cell line (CHO)[313]. Whilst the resultant rGBA is catalytically active, further glycosylation modifications are required to ensure the enzyme is internalised by macrophages when administered to GD patients in enzyme replacement therapy (ERT). Specifically, Cerezyme® is sequentially deglycosylated with a range of enzymes to yield high mannose-terminated *N*-glycans[320,651,652]. Although Cerezyme® exhibits five potential *N*-glycosylation sites[219], a crystal structure of fully glycosylated Cerezyme® (PDB 2J25) revealed that only three of the *N*-glycosylation sites, Asn19, Asn59 and Asn146, were occupied[643]. This structure also showed that all the glycosylation sites are adjacent to empty cavities in the crystal, thus allowing placement of the glycans in these spaces without generating steric clashes that would hinder crystallisation[643].

Table 3.1: Information for all the human GBA structures deposited in the PDB to date. CHO = Chinese hamster ovary, Baculo = Baculovirus expression

PDB	Source	Drug Name	Ligand	Res (Å)	Reference
1OGS	CHO	Cerezyme®	None	2.00	Dvir et al. (2003)[641]
1Y7V	CHO	Cerezyme®	Conduritol-B-epoxide	2.40	Premkumar et al. (2005)[653]
2F61	CHO	Cerezyme®	None	2.50	Liou et al. (2005)[654]
2J25	CHO	Cerezyme®	None	2.90	Brumshtein et al. (2006)[643]
2NSX	CHO	Cerezyme®	Isofagomine	2.11	Lieberman et al. (2007)[368]
2NT0	CHO	Cerezyme®	Glycerol	1.79	Lieberman et al. (2007)[368]
2NT1	CHO	Cerezyme®	None	2.30	Lieberman et al. (2007)[368]
2V3D	Plant	Taliglucerase	*N*-butyl-deoxynojirimycin	1.96	Brumshtein et al. (2007)[655]
2V3E	Plant	Taliglucerase	*N*-nonyl-deoxynojirimycin	2.00	Brumshtein et al. (2007)[655]
2V3F	Plant	Taliglucerase	None	1.95	Shaaltiel et al. (2007)[323]
2VT0	Plant	Taliglucerase	Conduritol-B-epoxide	2.15	Kacher et al. (2008)[656]
2WCG	Plant	Taliglucerase	*N*-octyl(cyclic guanidine)-nojirimycin	2.30	Brumshtein et al. (2009)[657]
2WKL	Human	Velaglucerase	None	2.70	Brumshtein et al. (2010)[320]
2XWD	CHO	Cerezyme®	5-*N*,6-*O*-[*N*'-(n-octyl)iminomethylidene]nojirimycin	2.66	Brumshtein et al. (2011)[658]
2XWE	CHO	Cerezyme®	5-*N*,6-*S*-[N'-(n-octyl)iminomethylidene]-6-thionojirimycin	2.31	Brumshtein et al. (2011)[658]
3GXD	CHO	Cerezyme®	None	2.50	Lieberman et al. (2009)[595]
3GXF	CHO	Cerezyme®	Isofagomine	2.40	Lieberman et al. (2009)[595]
3GXI	CHO	Cerezyme®	None	1.84	Lieberman et al. (2009)[595]
3GXM	CHO	Cerezyme®	None	2.20	Lieberman et al. (2009)[595]
3KE0	Baculo	-	None	2.70	Wei et al. (2011)[659]
3KEH	Baculo	-	None	2.80	Wei et al. (2011)[659]
3RIL	CHO	Cerezyme®	(3S,4R,5R,6S)-azepane-3,4,5,6-tetrol	2.40	Orwig et al. (2011)[660]
3RIK	CHO	Cerezyme®	(3S,4R,5R,6S)-1-(2-hydroxyethyl)azepane-3,4,5,6-tetrol	2.48	Orwig et al. (2011)[660]
5LVX	Human	-	Quinazoline modulator	2.20	Zheng et al. (2018)[661]
6MOZ	CHO	Cerezyme®	Pyrrolidine triazole iminosugar	2.10	M-Bailen et al. (2019)[662]
6Q1P	CHO		Pharmacological chaperone norIMX8	1.80	Santana et al (to be published)
6Q1N	CHO		Pharmacological chaperone IMX8	2.53	Santana et al (to be published)
6Q6K	CHO	Cerezyme®	Cy5 tagged cyclophellitol activity-based probe	1.92	Artola et al. (2019)[562] **(This Thesis)**
6Q6L	CHO	Cerezyme®	Adamantyl cyclophellitol inhibitor	1.81	Artola et al. (2019)[562] **(This Thesis)**
6Q6N	CHO	Cerezyme®	Biphenyl cyclophellitol inhibitor	1.63	Artola et al. (2019)[562] **(This Thesis)**
6TJK	CHO	Cerezyme®	None	1.56	Rowland et al. (2020)[663] **(This Thesis)**

6TJJ	Baculo	-	Bis-tris-propane	1.59	Rowland et al. (2020)[663] **(This Thesis)**
6TJQ	Baculo	-	2-deoxy-2-fluoro-β-D-glucopyranosyl fluoride	1.41	Rowland et al. (2020)[663] **(This Thesis)**
6TN1	Baculo	-	None	0.98	Rowland et al. (2020)[663] **(This Thesis)**
6YTP	CHO	Cerezyme®	Azide tagged cyclophellitol epoxide	1.70	Rowland et al. (2021)[664]
6YTR	CHO	Cerezyme®	Cyclophellitol Aziridine	1.70	Rowland et al. (2021)[664] **(This Thesis)**
6YV3	CHO	Cerezyme®	Galacto-configured cyclophellitol aziridine	1.80	Rowland et al. (2021)[664]
6YUT	CHO	Cerezyme®	*N*-acyl functionalised cyclophellitol aziridine	1.76	Rowland et al. (2021)[664]
6Z39	CHO	Cerezyme®	BODIPY-tagged cyclophellitol and *N*-acyl cyclophellitol aziridine	1.70	Rowland et al. (2021)[664] **(This Thesis)**
6Z3I	Baculo	-	Bi-functional cyclophellitol aziridine activity-based probe	1.80	Rowland et al. (2021)[664] **(This Thesis)**
7NWV	Baculo	-	BODIPY-tagged cyclophellitol activity-based probe	1.86	To be published **(This Thesis)**

*Table updated 26/08/21

3.2.2 Activity-Based Probes and Inhibitors for β-Glucosidases

3.2.2.1 Quinone Methide Probes

In early work, activity-based probes (ABPs) making use of quinone methide chemistry were developed to trap and label a variety of β-glucosidases[552]. Such probes typically comprised a glucose recognition head, a hydroxy-benzylic fluoride moiety, a linker and a biotin reporter group[552], Figure 3.6. These quinone methide probes function through a rather convoluted mechanism in which the glycosidic bond between the glucose and hydroxy-benzylic fluoride moiety is cleaved by the target enzyme, leading to 1,6-elimination of the fluoride group and the generation of a reactive quinone methide intermediate[552]. This intermediate then traps a nearby nucleophile of the target enzyme, resulting in labelling of the enzyme with a biotin group which can be used to isolate, detect and enrich the protein of interest[552]. These probes appeared to be effective in labelling purified glycosidases but cross-reactivity with other glycosidases was observed in complex mixtures[552]. This was attributed to the indiscriminate and diffuse nature of the quinone methide group which is released upon probe hydrolysis. Therefore, ABPs with greater selectivity and specificity were desired.

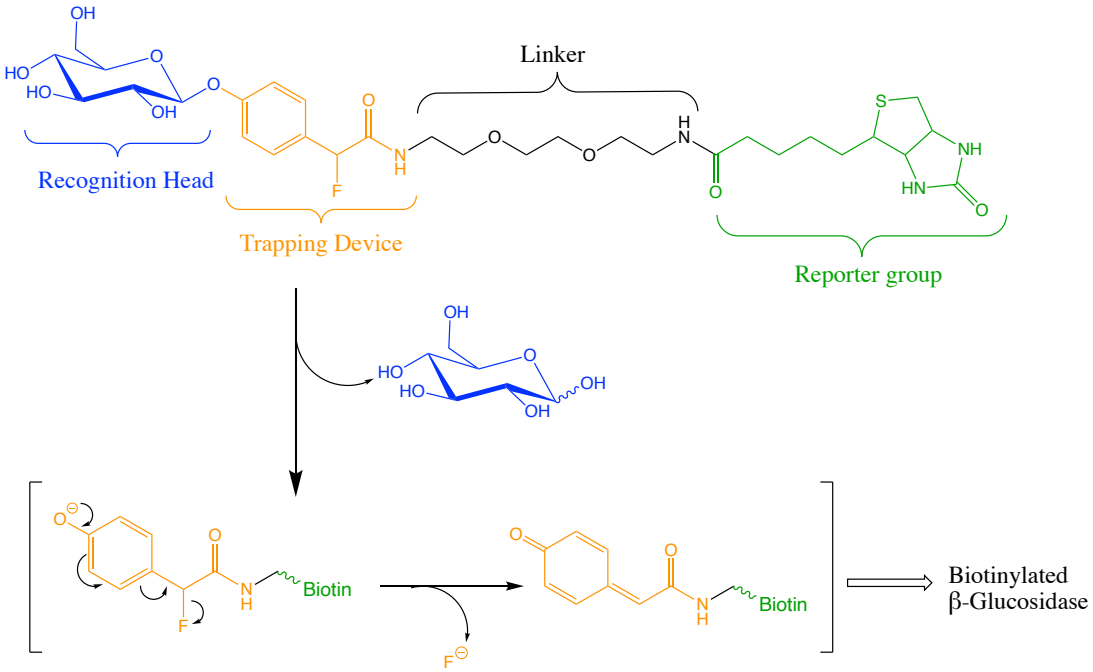


Figure 3.6: General reaction mechanism of quinone methide ABP with retaining a β-glycosidase. Glucose moiety binds to the enzyme and is hydrolysed to release fluoride and reactive quinone methide intermediate which reacts with a nearby enzymatic nucleophile to label the target enzyme.

3.2.2.2 Fluoroglycoside Inhibitors and ABPs

In 1988, Withers and co-workers reported a novel class of mechanism-based inhibitors for retaining glycosidases. These inhibitors were based on the 2-deoxy-2-fluoroglycoside scaffold, which reacts with the catalytic nucleophile of retaining glycosidases to form a long-lived covalent intermediate that accumulates due to the destabilising effect of the fluorine atom on deglycosylation[607,665]. Further development resulted in the generation of probes with radioactive ^{18}F reporter tags or a C6 azide tags for two step labelling with phosphine reporter groups, Figure 3.7. These probes permitted the visualisation of retaining β-glucosidases in complex mixtures, with two-step ABPP being successfully utilised for the detection of β-glucosidases in cell lysates[665].

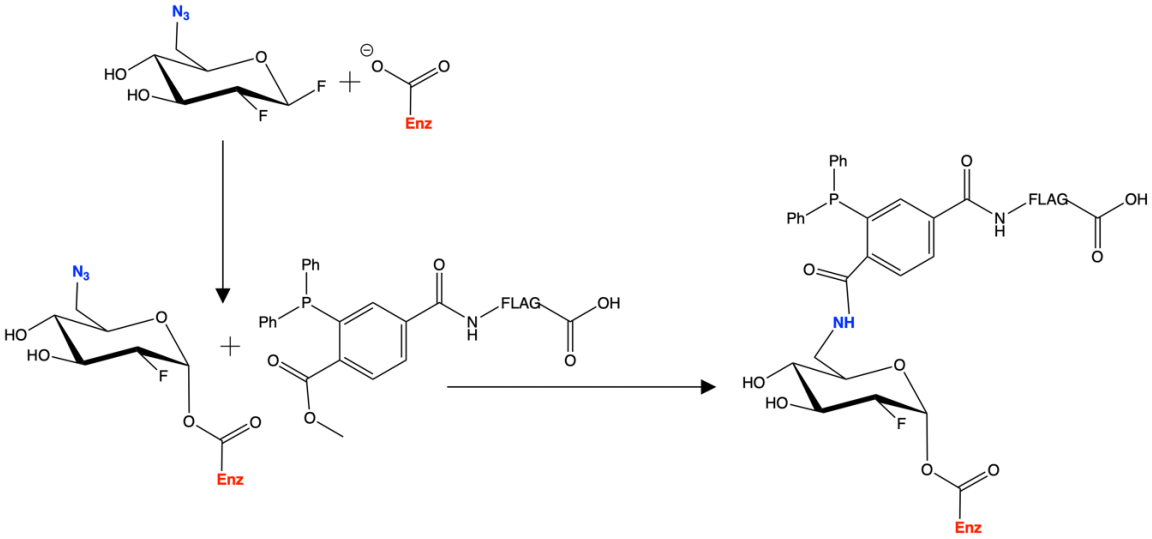


Figure 3.7: Two-step activity-based labelling of a generic β-glucosidase with azide equipped fluoroglycoside inhibitor and phosphine reporter group through the bioorthogonal Staudinger-Bertozzi ligation reaction.

More recently, the Withers lab reported the development of a new class of fluorosugar glycosidase inactivators, which bear tuneable phosphorous-based leaving groups[666], Figure 3.8. These inactivators react with GBA over 4000-time faster than the parent 2-deoxy-2-fluoro-β-D-glucosyl fluoride and have the potential to incorporate lipid-like substituents at the phosphate head to mimic the ceramide portion of the natural substrate[666]. In a similar fashion to the Withers' fluoroglycosides, Hekmat and colleagues also developed a fluorosugar-based probe equipped with a biotin tag via a cleavable disulfide linker[667]. This probe was used to isolate and profile a number of retaining β-glycosidases, leading to the discovery of a new β-1,4-glycanase from the soil bacterium *Cellulomonas fimi*[667]. However, a potential disadvantage

of 2-deoxy-2-fluoroglycoside ABPs is that the enzyme-inhibitor complex is slowly hydrolysed, with reported lifetimes ranging from seconds to months[585,668]. Nevertheless, 2-deoxy-2-fluoroglucosidase probes have been successfully converted into probes for other glycosidases, including galactosidases[665], hexosaminidases[668], xylanases and cellulases[576].

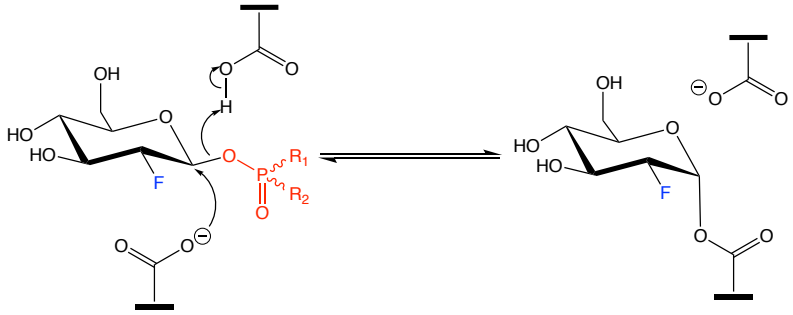


Figure 3.8: Inactivation of GBA by phosphorous bearing 2-deoxy-2-fluoro-β-D-glucoside where R_1/R_2 = O-alkyl or O-benzyl substituents.

3.2.2.3 Conduritol B Epoxide and Cyclophellitol-Based Inhibitors and ABPs

As discussed in **Chapter 1** (*section 1.6.1*), conduritol B epoxide (CBE) and cyclophellitol are irreversible mechanism-based β-glucosidases inhibitors which covalently react with the catalytic nucleophile of a target enzyme through trans-diaxial ring opening of the electrophilic epoxide trap[669]. This results in the formation of a covalent enzyme-inhibitor complex which, unlike the fluoroglycoside-enzyme intermediate, cannot be hydrolysed. By equipping CBE with various reporter groups, this inhibitor has been used to study the activity of wild-type and mutant GBA[670,653]. However, the poor selectivity of CBE results in off-target inhibition of other glycosidases, specifically the related class of α-glucosidases. In contrast, cyclophellitol is a more potent and selective mechanism-based inhibitor of retaining β-glucosidases owing to its differing C5 substituent[669]. CBE features a hydroxy group at the C5 position, whilst cyclophellitol possesses a hydroxymethylene moiety. This change in C5 substituent disrupts the symmetry of the molecule and appears to considerably enhance the potency and specificity of cyclophellitol towards retaining β-glucosidases[559]. In light of this, Overkleeft and co-workers developed a number of cyclophellitol-based ABPs by incorporation of fluorescent reporter groups and ligation handles at the C6 position of the cyclophellitol moiety[600], Figure 3.9 (a). For example, substitution of the C6 hydroxyl with an azide ligation handle results in a potent inhibitor which was used to profile GBA activity through two step labelling using copper(I)-catalysed azide-alkyne cycloaddition[600,601]. Interestingly, direct instalment of a BODIPY fluorophore at the C6-position of the cyclophellitol moiety resulted in a one-step ABP which

labelled human GBA with a much higher potency than the two-step azide ABP[93]. These "Overkleeft" probes have proved superior to CBE probes and have since been used to directly visualise GBA activity *in vitro* in cultured cells and *in vivo* in mice[600,555]. Applications of these cyclophellitol ABPs in monitoring and evaluating treatment efficacy for GD have also been demonstrated[600].

Following the success of the cyclophellitol based ABPs, Overkleeft and co-workers expanded their panel of β-glucosidase ABPs by substituting the epoxide warhead with an aziridine functionality[559,599], Figure 3.9 (b). These aziridine ABPs have proved more potent β-glucosidase inhibitors than the equivalent epoxides and have been used to selectively label multiple β-glucosidases, including GBA and GBA2[599]. A considerable advantage of the cyclophellitol aziridine ABPs is they permit additional functionalisation of the probe by substitution at the aziridine nitrogen[549], allowing the potency and specificity of the probes to be fine-tuned for different glucosidases. Subsequently, a number of *N*-acyl and *N*-alkyl functionalised cyclophellitol aziridine ABPs were synthesised[599,556], Figure 3.9 (c, d), which proved to be nanomolar inhibitors of GBA *in vitro* and *in situ*[556]. However, the *N*-acyl aziridine ABPs were prone to hydrolysis and more difficult to synthesise and handle than the *N*-alkyl analogues[599].

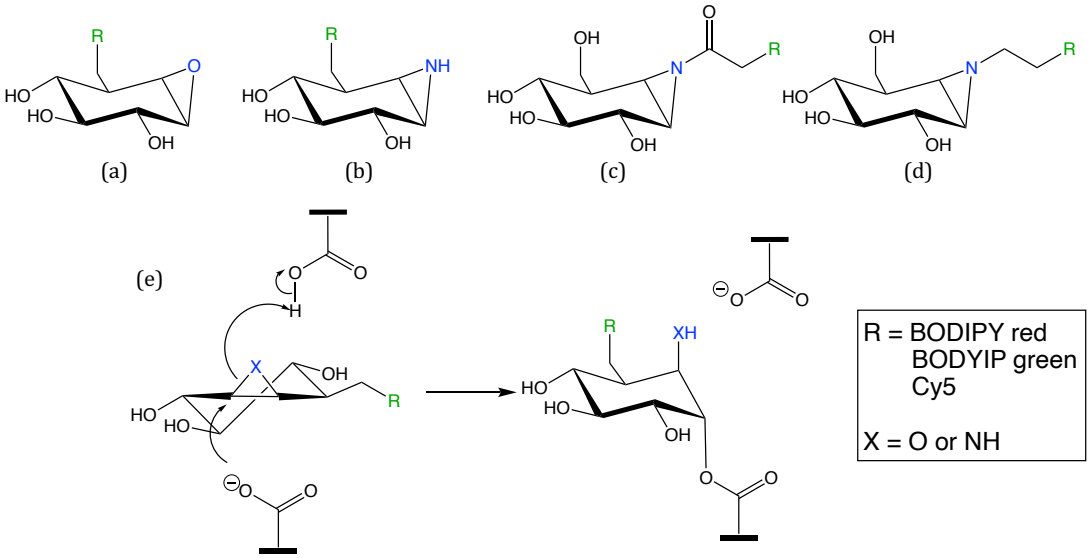


Figure 3.9: Chemical structure of (a) C6 functionalised cyclophellitol epoxide ABP (b) C6 functionalised cyclophellitol aziridine ABP (c) *N*-acyl aziridine ABP (d) *N*-alkyl aziridine ABP. (e) General reaction mechanism of cyclophellitol based ABP with a retaining β-glucosidase.

The discussed cyclophellitol epoxide and aziridine ABPs function by taking advantage of the β-glucosidase conformational reaction itinerary[560]. Specifically, the cyclophellitol moiety of these ABPs adopts the transition state mimicking half-chair ($^4H_3$) conformation, which ensures that

the epoxide or aziridine warhead is readily positioned for nucleophilic attack when ABP binds to the enzyme active site[560,556], Figure 3.9 (e). Following covalent modification of the catalytic nucleophile, the cyclophellitol ring adopts the chair $^4C_1$ conformation, mimicking the conformation of the covalent intermediate in the β-glucosidase reaction itinerary[671]. Consequently, these ABPs behave as covalent mechanism-based inactivators.

3.2.3 Animal Models for Gaucher Disease

It has long been appreciated that animal models are extremely important tools in studying disease pathogenesis, diagnosis and treatment. In the case of GD, mice are commonly used to generate Gaucher models because the nucleotide and corresponding amino acid sequence of murine GBA is > 80 % identical to the human protein[645]. However, the generation of viable Gaucher animal models has proved challenging, Table 3.2.

The first non-genetic mouse model for GD was generated in 1975 through a chemical-knockdown approach using CBE[672]. In this seminal work, Kanfer et al. treated mice with CBE through daily intravenous injection[672]. Following 3-weeks of treatment, the tissue and plasma activity of GBA was reduced by 93% with a concomitant accumulation of GlcCer in the spleen, liver and brain[672]. This chemically-induced model was further enhanced by injecting mice with a combination of CBE and GlcCer, which resulted in increased glycolipid burden[673]. This optimised model was subsequently used to demonstrate the success of gene therapy in restoring GlcCer levels in liver macrophages[674]. Interestingly, neurological manifestations could also be induced by increasing the CBE dosage and frequency of administration[675]. However, these neurological symptoms were not reversed upon removal of CBE and the mice died within weeks of the last injection[675]. The irreversibility of these neurological manifestations suggested that neurological disease may not be reversible in humans, further emphasising the importance of early diagnosis and therapeutic intervention. Unfortunately, the broad specificity of CBE has since limited its use in the generation of GD animal models through chemical knockdown.

3.2.3.1 Genetic Models

The first genetic mouse model of GD was generated in 1992 using a genetic knockdown strategy with a null allele[676]. This model mimicked the GD2 phenotype with significantly reduced GBA activity and accumulation of GlcCer in the liver, spleen, brain and lungs. However,

these mice died during the neonatal period due to skin permeability issues[676,677], severely restricting their use. Nevertheless, these models provided some insight into the skin phenotypes associated with GD2 and indicated that defects in epidermal biochemistry could be used to discriminated between GD2 and GD3 patients[678].

An alternative approach utilising single insertion mutagenesis has also been used to generate Gaucher mouse models[678]. Specifically, mice with the RecNcil (L444P/A456P) mutation and mice homozygous for the L444P mutation have been produced through this mutagenesis approach[679]. Both models exhibited reduced GBA activity in the liver, brain and skin but died soon after birth due to compromised skin function[679]. Although the RecNcil model was sufficiently long lived to investigation the role of GBA in skin function[680], the lethality of these models limited their applications. However, in 2002, the Proia group generated longer lived L444P homozygous mice by crossing L444P heterozygous mice with heterozygous genetic knock-out mice lacking the gene for glucosylceramide synthase (GCS)[681]. These models exhibited some GD manifestations and survived beyond the neonatal period[681]. However, no accumulation of GlcCer was detected and a number of GD features were not observed[681]. Nevertheless, this model did provide some insight into the involvement of the immune system in GD pathophysiology and was also used in initial PCT studies[366].

Mice with other missense mutations, namely V294L, D409H and D409V, have also been generated[678]. These models had considerably improved lifetimes with residual GBA activity 4-10% of the normal range. However, no considerable phenotypic abnormalities were observed[678,682]. Surprisingly, generating a model homozygous for the N370S mutation proved incredibly challenging. Whilst this is a relatively mild mutant in humans, mouse models with this mutation typically died within 24 hours of birth due to skin defects[682]. Interestingly, no relationship between skin phenotype and GBA activity could be deduced, suggesting that mice and humans degrade GlcCer in the skin differently. In fact, a number of mouse strains have resulted in phenotypes which are not analogous to those seen in GD patients with the same mutations[678]. This highlights the importance of environmental factors and other modifiers in determining phenotype. Nevertheless, these mutant mice have been used to test various experimental treatments, such as L-type calcium channel blockers to enhance GBA folding[683].

In an alternative strategy, mice carrying *GBA1* point mutations were crossed with mice carrying null alleles to generate mice with significantly reduced GBA activity. These mice

experienced significant GlcCer accumulation and extreme phenotypic responses[682]. Consequently, these models proved incredibly useful in studying therapeutic approaches, such as adeno-associated viral (AAV) mediated gene therapy[684] and SRT[348,357].

In 2006, the first conditional GD mouse model was generated through genetic knock-out of the *GBA1* gene using *Cre-LoxP* recombination[390]. *Cre-LoxP* combination is a site-specific recombinase technology which makes use of Cre recombinase to perform deletions, insertions translocations or inversions at specific DNA sites dictated by the *loxP* site[685]. This approach has generated viable and long-lived GD mouse models[390] which have been used to assess the therapeutic benefits of bone marrow transplantation and gene therapying in treating the skeletal symptoms associated with GD[678]. A similar *Cre-LoxP* system has also been used to generate mice exhibiting more severe visceral phenotypes[264] and neuronal mouse models exhibiting rapidly-progressive neurological disease[686]. In fact, this was the first mouse model to recapitulate the neurological aspects of GD, providing a basis for investigations into the underlying pathological mechanism and experimental treatments[678]. A number of other conditional neuronal mouse models have since been established to study the progression of neurological manifestations and altered macrophage marker expression[687,688].

More recently, Jackson et al. (2019)[689] developed a novel neuronopathic GD model from an existing GD model. The original D409V homozygous model yields mice with mild neurological symptoms which better represent GD1. In an effort to increase the severity of the neurological symptoms, GBA activity in the CNS of D409V homozygous mice was reduced by administering adeno-associated virus encoding a microRNA against *GBA1* (AAV- miR-GBA)[689]. This microRNA is a small non-coding RNA molecule which prevents translation of GBA by binding to the messenger RNA[690]. The administration of AAV-miR-GBA was shown to cause progressive neurological manifestations, which typically occurred at ~ 10 weeks of age. These mice were viable for up to 36 weeks which, in combination with the progressive nature of neurological symptoms, allowed for pre-symptomatic, early stage and late stage administration of therapeutics[689]. Typical neurological symptoms included neurological impairment, motor dysfunction, inflammation, hyperactivity and ataxia. Furthermore, these symptoms could be prevented in the pre-symptomatic stage by administering miRNA resistant GBA, demonstrating that the neuropathological effects are due to miRNA induced GBA reduction[689]. More importantly, these neurological manifestations are consistent with those observed in

neuronopathic GD3 pateints[689]. It is hoped the slow progression of neurological disease and increased lifespan of these mice will render them suitable for therapeutic testing.

3.2.3.2 Non-Murine Models

Whilst mouse models remain the most common animal model, two non-murine GD models have been described. The first was an Australian Sydney Silky dog which exhibited reduced GBA activity, increased GlcCer levels and neurological manifestations[691,692]. The second non-mouse model was a naturally occurring neuronal GD sheep[693]. Mutation analysis revealed the presence of two homozygous missense mutations, making this sheep the first neuronal GD model to result from spontaneous mutation[693].

Although significant progress has been made in the development of GD animal models, considerable work is still required to ensure such models can faithfully recapitulate the human disease. Ideally, animal models should have long enough life spans to demonstrate differing disease states and severity, whilst permitting the evaluation of various therapeutic strategies.

Table 3.2: Timeline and summary of Gaucher disease murine models discussed in this work.

Year	Generation Method	Genotype	Application
1975	Chemical knockdown with CBE[672]	-	Efficacy of gene therapy
1992	Genetic knockout with null allele[676]	GBA^{+}	Investigate skin pathology
1998	Point mutation[679]	RecNcil,	Investigate skin pathology
2002	Point mutation cross with null allele[681]	L444P/L444P	Therapeutic evaluation
2003	Point mutation[682]	V394L	Testing therapeutic options
		D409H	Testing therapeutic options
		D409V	Testing therapeutic options
2003	Point Mutation[682]	N370S	-
2006	Conditional knockout *Cre-LoxP*[390]	-	Evaluate gene therapy
2007	Conditional knockout[686]	-	Investigate neuropathology
2008	Conditional knockout[675]	-	-
2010	Conditional knockout Mx1-*Cre-LoxP*[264]	-	Testing therapeutic options
2019	AAV-mediated miRNA interference[689]	D409V/D409V	Investigate neuropathology

3.2.3.3 Potential for PD Models

It should be noted that certain mutant *GBA1* models are also of interest as models for PD. Specifically, the D409V mutant has been identified as a particularly useful mutant for PD because it exhibits drastically reduced GBA activity which is thought to contribute to the accumulation and aggregation of α-synuclein[694,695]. However, most of the D409V models

described previously are not widely accessible to the wider research community. Consequently, the Michael J. Fox Foundation for Parkinson's Disease (MJFF), founded in 2000, has sponsored the development and distribution of a wide range of PD tools (https://www.michaeljfox.org/research-tools-catalog)[696], including the generation of a new *GBA1* D409V knockout model which is widely available through The Jackson Laboratory repository (www.jax.org/strain/019106).

3.2.4 Inhibitors for Chemical Knockdown Animal Models

There is considerable demand for more potent and selective GBA inhibitors for GD research. In particular, highly specific GBA inhibitors are required for the generation of GD animal models through chemical knockdown. As highlighted previously, chemical knockdown approaches using CBE and cyclophellitol have yielded unreliable models with poor viability. Such problems are thought to be a consequence of the low specificity of these inhibitors which leads to off-target inhibition of other glycosidases. Specifically, cyclophellitol inhibits GBA and non-lysosomal glucosylceramidase (GBA2) with approximately equal efficiency[277], whilst CBE exhibits off target inhibition with lysosomal α-glucosidase (GAA)[697], GBA2[698] and lysosomal β-glucuronidase (GUSB)[699].

Although the Overkleeft cyclophellitol-based ABPs can induce changes in the system at hand when covalently bound, it has been proposed that these ABPs and their equivalent inhibitors may be suitable for the generation of GD animal models through chemical knockdown. This strategy has traditionally employed CBE to inhibit endogenous GBA and induce GD manifestations. However, given that cyclophellitol is a more potent and selective β-glucosidase inhibitor, it is hoped that cyclophellitol-based inhibitors will be more reliable for the generation of GD animal models. Indeed, following primary studies within the Davies group, which indicated that the presence of bulky hydrophobic tags at the C6 position of cyclophellitol ABPs drastically improves their specificity for GBA[599], we questioned whether cyclophellitols bearing hydrophobic moieties at the C6-position would be suitable inhibitors for generating chemical knockdown Gaucher animal models. This hypothesis was further incentivised by reports from the Vocadlo lab (Simon Fraser University) on a range of fluorogenic substrates, featuring fluorophores at C6 of β-glucoside, which proved to be very selective GBA substrates *in situ*[700]. Therefore, the Overkleeft lab synthesised a number of C6-substituted cyclophellitols as potentially more selective GBA inhibitors for use in chemical knockdown studies.

3.2.5 Research Aims

This work aimed to structurally analyse a number of C6-functionalised cyclophellitol-based compounds, Figure 3.10, which were synthesised by the Overkleeft lab as more potent GBA inhibitors for use in chemical knockdown studies. Specifically, this work sought to establish how GBA is able to structurally accommodate bulky C6-substituents and how this may be exploited for the development of more selective inhibitors. In addition, this work also aimed to structurally evaluate a bi-functionalised aziridine ABP, which was synthesised by the Overkleeft lab following the hypothesis that bi-functionalised ABPs may exhibit enhanced potency for GBA in comparison to the C6-functionalised cyclophellitol analogues.

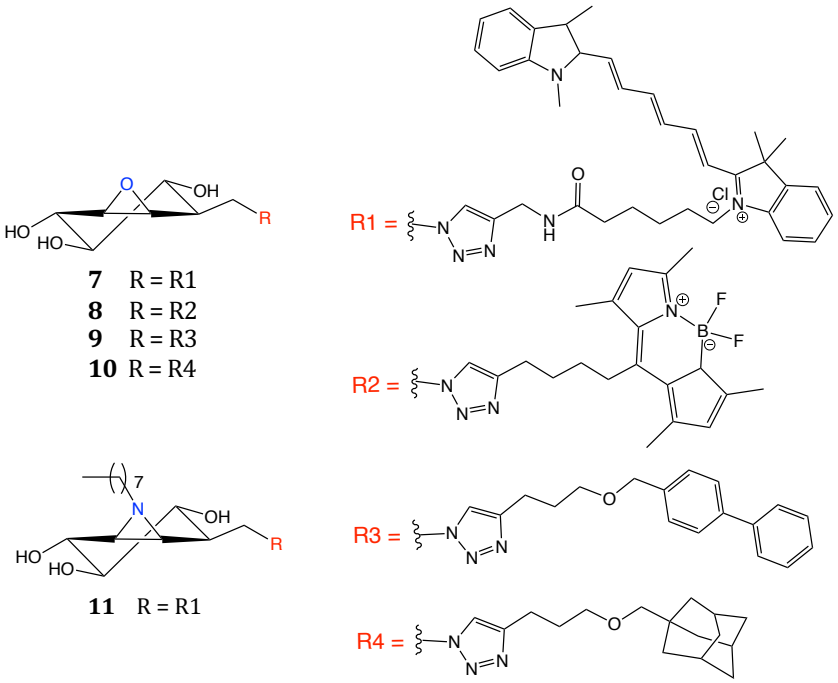


Figure 3.10: Chemical structure of cyclophellitol based ABPs (**7**, **8** and **11**) and inhibitors (**9** and **10**) investigated in this work.

3.3 Materials and Methods

3.3.1 Crystallisation of Cerezyme®

Prior to crystallisation, Cerezyme® (a kind gift from Professor Johannes Aerts, Leiden University) was deglycosylated with PNGase F (New England Biolabs (NEB)) for 5 days at room temperature (performed by Liang Wu (University of York)). Crystallisation screening was performed using hanging drop vapour diffusion based on conditions outlined by Dvir et al. (2003)[641]. Crystals of rGBA were obtained using a 1:1 (v/v) ratio of Cerezyme® (1µL, 9.1 mg mL^{-1}) and well solution comprised of 1.1 M ammonium sulfate, 0.1 M sodium acetate, 0.04 M potassium chloride and 0.19 M guanidine hydrochloride.

3.3.1.1 Unliganded Crystal Structure

Unliganded crystals were fished from the crystallisation drops and briefly transferred to 2M lithium sulfate cryoprotectant containing 0.1 M sodium acetate buffer (pH 4.6), 0.04 M potassium chloride and 0.17 M guanidine hydrochloride. Crystals were flash frozen in liquid nitrogen prior to in-house diffraction testing using a Rigaku micromax-007HF X-ray generator in conjunction with an Actor robotic sample changer. Suitable crystals were sent to the Diamond Light Source facility for full data collection.

*3.3.1.2 Complexes with C6-functionalised ABPs (**7**) and Inhibitors (**9, 10**) and*

Each inhibitor was prepared at 20 mM in HEPES buffer (20 mM, pH 7.4) and diluted to 2mM (in mother liquor comprising 1.5 M ammonium sulfate, 0.125 M sodium acetate (pH 4.6), 0.05 M potassium chloride and 0.19 M guanidine hydrochloride. Unliganded crystals were soaked in the inhibitor spiked mother liquor solutions for 4 hours before transferring to a lithium sulfate cryo-protectant and freezing in liquid nitrogen for data collection.

*3.3.1.3 Complex with BODIPY-ABP (**8**) and Bi-functionalised ABP (**11**)*

Each ABP was prepared at 20mM in 100 % DMSO before diluting to 2 mM in mother liquor containing 0.2 M sodium sulfate, 0.25 M HEPES pH 7.0, 14% (v/v) PEG 3350 and a final DMSO concentration of 10%. Unliganded recombinant GBA crystals (prepared in **Chapter 4**) were briefly soaked in the ABP spiked mother liquor solution and fished from the drop as soon as

crystal damage from the high DMSO concentration became apparent. The crystals were then transferred to an ethylene glycol cryoprotectant (25%) before flash freezing in liquid nitrogen

3.3.1.4 Data Processing and Structure Determination

Data for the unliganded crystal were collected at the i03 beamline whilst data for the co-crystal complexes were collected at the i04_1 (**7**), i04 (**8, 11**) and i24 (**9, 10**) beamlines. For all structures, data were processed using the XIA2[615,616] and AIMLESS[617,618] data reduction pipelines through the CCP4i2 software[619]. The unliganded and C6-functionalised inhibitor complex structures were solved by molecular replacement using MOLREP[620] with the previous GBA PDB (2NT0) as the homologous search model. The bi-functional ABP complex structure was solved by molecular replacement using PDB 6TJK. For all structures, refinement was performed using REFMAC[621] followed by several rounds of manual model building with COOT[622] Idealised coordinate sets and refinement dictionaries for the ligands were generated using JLIGAND[623] and sugar conformations were validated using Privateer[594]. Data collection and refinement statistics for all structures discussed in this chapter can be found in Table 3.3.

3.3.2 In-Solution Fluorescence Labelling

3.3.2.1 Time dependent labelling

Cerezyme® was prepared at 200 nM in 150 mM McIlvaine buffer pH 5.2 (with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) and incubated at 37 °C with 200 nM **7**. A 5 µL sample was taken at 10, 20, 30 and 60 min and denatured with Laemmli dye (3x) by heating at 95 °C for 5 minutes.

3.3.2.2 Concentration dependent labelling

GBA produced in an insect-baculovirus expression system[663] (see **chapter 4**) was diluted to 200 nM in 150 mM McIlvaine buffer pH 5.2 (with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) and ABP **7** was added to 150, 100, 50, 10, 1, 0.1, 0.01 or 0.001 nM in a reaction volume of 10 µL. The reactions were incubated at 37 °C for 30 mins and then denatured with Laemmli (x3) sample buffer at 95 °C for 5 minutes.

All ABPP samples were resolved by electrophoresis on a 10% SDS-PAGE gel and scanned on fluorescence using an Amersham Typhoon 5 Imager (GE Healthcare) with λ_{EX} 635 ; λ_{Em} 665 nm.

Table 3.3: Data collection and refinement statistics for all rGBA crystal structures discussed in this chapter. BEVS = insect-baculovirus expression

	Apo	**7**	**8**	**9**	**10**	**11**
Protein Source	**Cerezyme®**	**Cerezyme®**	**BEVS**	**Cerezyme®**	**Cerezyme®**	**BEVS**
Data collection			***Chapter 4**			***Chapter 4**
Space group	*C* 2 2 2_1	*C* 2 2 2_1	*P* 1 2_1 1	*C* 2 2 2_1	*C* 2 2 2_1	*P* 1 2_1 1
Cell dimensions						
a, *b*, *c* (Å)	110.1, 285.8, 91.9	110.6, 285.9, 92.3	52.9, 158.4, 68.2	110.4, 285.2, 91.9	110.2, 285.1, 92.0	53.1, 76.7, 68.0
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 102, 90	90, 90, 90	90, 90, 90	90, 102, 90
Resolution (Å)	77.43-1.71 (1.74-1.71)*	77.52-1.92 (1.95-1.92)	158.42-1.86 (1.89-1.86)	72.13-1.63 (1.66-1.63)	71.98-1.81 (1.84-1.81)	76.67-1.80 (1.80-1.84)
R_{merge}	0.115 (2.226)	0.161 (1.623)	0.119 (0.843)	0.101 (1.609)	0.132 (1.661)	0.207 (2.351)
R_{pim}	0.046 (0.873)	0.064 (0.647)	0.049 (0.345)	0.047 (0.735)	0.061 (0.757)	0.084 (0.971)
I / σI	9.34 (0.98)	6.5 (1.1)	9.1 (2.0)	8.8 (1.1)	7.7 (1.0)	6.7 (0.8)
$CC_{1/2}$	0.998 (0.693)	0.990 (0.707)	0.996 (0.859)	0.989 (0.476)	0.989 (0.529)	0.995 (0.393)
Completeness (%)	91.9 (98.0)	99.9 (99.8)	100 (100)	99.9 (98.8)	100.0 (98.7)	100 (100)
Redundancy	7.9 (8.1)	8.3 (8.1)	6.9 (7.0)	6.49 (6.50)	6.53 (6.58)	7.0 (6.8)
Refinement						
Resolution (Å)	77.43-1.71	77.52-1.92	158.42-1.86	72.13-1.63	71.98-1.81	76.67-1.80
No. reflections	156023	925691	627714	1167262	859243	348887
R_{work} / R_{free}	0.20/0.23	0.18/0.22	0.18/0.23	0.18/0.21	0.18/0.21	0.18/0.19
No. atoms						
Protein	7911	7939	7920	7870	7960	3972
Ligand/ion	234	354	384	273	334	193
Water	655	699	744	966	936	340
B-factors (Å^2)						
Protein	34	39	25	26	28	26
Ligand/ion	69	65	46	52	55	48
Water	42	47	35	39	39	37
R.m.s. deviations						
Bond lengths (Å)	0.010	0.009	0.009	0.012	0.010	0.005
Bond angles (°)	1.59	1.60	1.56	1.67	1.63	1.39
Ramachandran Plot						
Favourable regions (%)	94.8	94.5	95.1	94.7	94.6	96.4
Allowed regions (%)	4.3	4.3	4.0	4.2	4.3	3.3
PDB code	6TJJ	6Q6K	7NWV	6Q6N	6Q6L	6Z31

3.4 Results and Discussion

3.4.1 Crystallisation and Structure of Cerezyme®

Crystallisation of rGBA (Cerezyme®, a kind gift from Professor Johannes Aerts, Leiden University) was optimised from conditions outlined by Dvir et al. (2003)[641] using hanging drop vapour diffusion, Figure 3.11.

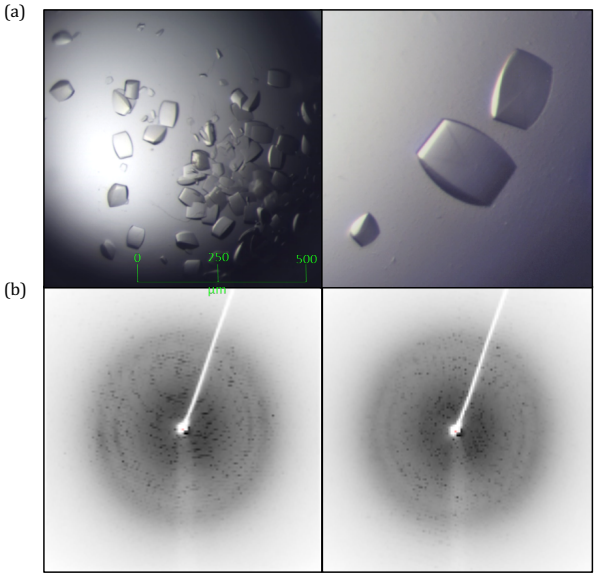


Figure 3.11: (a) Crystals of rGBA grown under optimised conditions. (b) Diffraction images of rGBA crystals taken in house at 0° and 90° respectively.

Following crystallisation, an unliganded rGBA structure was obtained at 1.71 Å and the structure was solved to reveal a homo-dimeric glycoprotein comprising of two 497-residue monomers, Figure 3.12 (a). Each GBA monomer comprised of the expected three non-contiguous domains, Figure 3.12 (b). Domain I consists of a three-stranded anti-parallel β-sheets which tightly interact with a number of α-helices from domain III. Domain II is formed of two β-sheets associated in an immunoglobulin-like fold and domain III comprises a $(\beta/\alpha)_8$ TIM barrel, with the active site located at the centre of the β-barrel.

As previously highlighted, the glycosylation of rGBA is vital to its activity and trafficking[313,335], however, rGBA was deglycosylated with EndoH in this work to aid crystallisation. Despite this, two *N*-linked glycosylation sites were observed at Asn19 and Asn146 in both molecules of the crystallographic dimer. A single *N*-acetylglucosamine (NAG) residue could be modelled at Asn146, whilst two NAG residues were modelled at Asn19. Of note, occupation of the Asn19 *N*-

glycosylation site has proved crucial to GBA activity. Overall, this unliganded structure of rGBA is consistent with the initial Cerezyme® structure (PDB 1OGS) deposited by Dvir et al. (2003)[641] but at a higher resolution and with an additional occupied *N*-glycosylation site at Asn146.

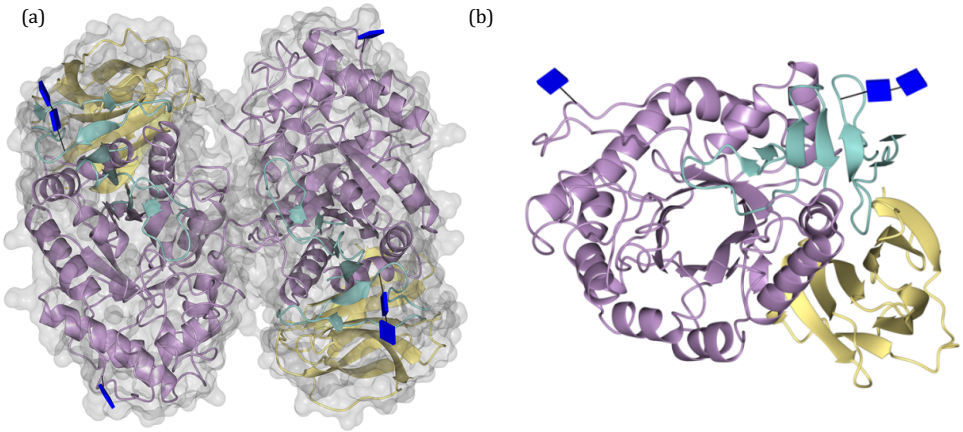


Figure 3.12: Crystal structure of recombinant GBA. (a) Surface and ribbon diagram of rGBA dimer (b) Ribbon diagram of rGBA monomer comprising of three domains; Domain I consists of an anti-parallel β-sheet (teal), Domain II consists of two β-sheets forming an immunoglobulin-like fold (yellow) and Domain III comprises a $(\beta/\alpha)_8$ TIM barrel containing the catalytic active site (lilac). *N*-linked glycans illustrated using Glycoblock format[591].

3.4.2 3D Complexes with Fluorescently Tagged Cyclophellitol Activity-Based Probes

Cyclophellitol epoxides and aziridines offer a powerful activity-based probe approach to the study of β-glucosidase activities. In the case of GBA, attachment of bulky hydrophobic substituents at the C6 position of the cyclophellitol unit has proved beneficial for GBA inhibition, yielding inhibitors and ABPs with nanomolar affinity compared to the high nanomolar-micromolar affinity observed for unsubstituted equivalents[599]. Moreover, the presence of bulky C6-groups appears detrimental to the inhibition of other retaining β-glucosidases. Consequently, C6-substituted cyclophellitols have proved very potent GBA inhibitors with considerably enhanced selectivity for GBA *in situ, in vitro* and *in vivo*[555,600].

Previous structural studies performed within the Davies group on a C6-functionalised cyclophellitol ABP (**8**), indicated that GBA may be able to accommodate hydrophobic substituents in a large cavity at the dimer interface[562]. This may provide some structural rationale for the improved potency and selectivity of C6-functionalised cyclophellitols. Therefore, to investigate this further, structural analysis of rGBA in complex with a C6 Cy5-

functionalised cyclophellitol-based ABP (**7**) was performed in this work. Data for the Cy5-functionalised ABP (**7**) complex were solved 1.92 Å resolution to reveal a single molecule of **7** bound covalently in the active site of both rGBA molecules in the crystallographic dimer, Figure 3.13. Specifically, clear electron density was observed for reacted ABP **7** bound to the catalytic nucleophile (Glu340) through trans-diaxial ring opening of the epoxide warhead to form a covalent enzyme complex in the $^4C_1$ chair conformation, Figure 3.13 (a). This conformation is consistent with the covalent intermediate in the retaining β-glucosidase conformational reaction itinerary, Figure 3.3. Additionally, the reacted cyclophellitol moiety of ABP **7** forms an extensive hydrogen bonding network in the active site, making hydrogen bonds with Asp127, Trp179, Asn234 and Trp381. In contrast, the C6-triazole linker and subsequent amide group bind through predominantly hydrophobic interactions in a broad hydrophobic active site cleft formed by Tyr244, Phe246 and Tyr313. These residues were highlighted in the first crystal structure of Cerezyme® as being important to the formation of the active site pocket[641].

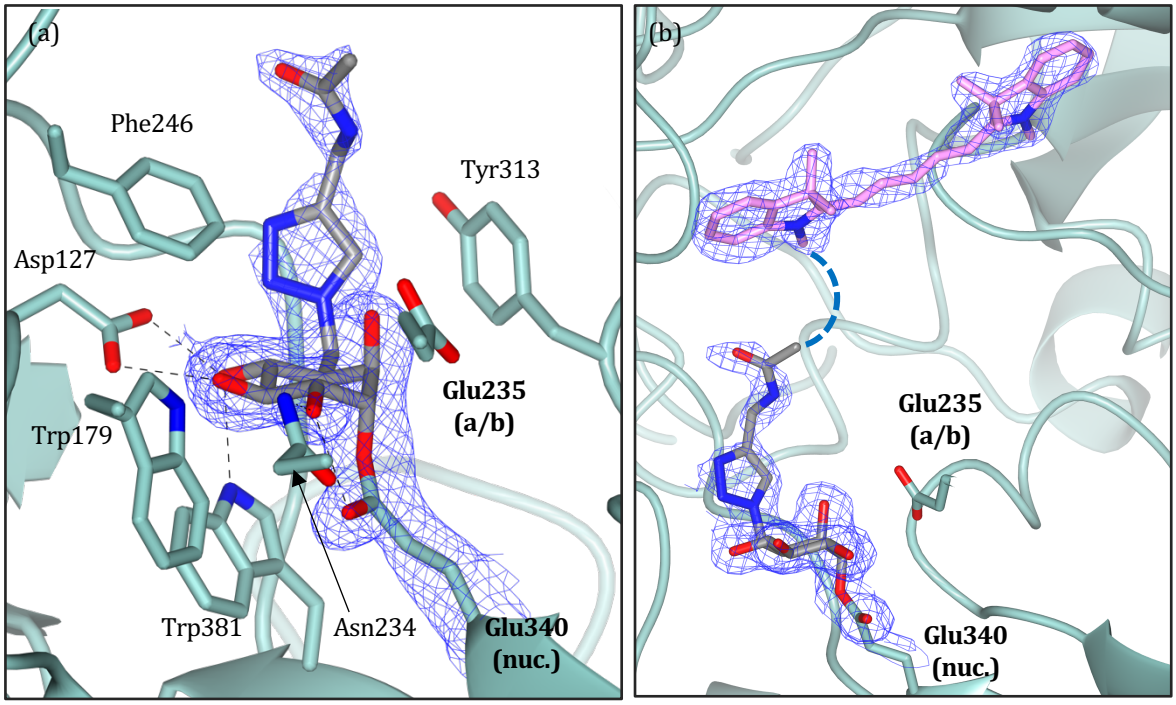


Figure 3.13 (a) 3D active site structure of ABP **7** covalently bound to the catalytic nucleophile (Glu340) of rGBA by trans-diaxial ring opening of the epoxide trap to form a covalent complex in $^4C_1$ conformation. (b) Electron density for Cy5 fluorophore (pink) of ABP **7** bound in a hydrophobic pocket at the dimer interface. Fluorophore lies close to the molecule of **7** bound in chain A. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Glu340 contoured to 1 σ (0.29 electrons/Å^3).

Unfortunately, no electron density was observed for the butyl chain that links the triazole group to the Cy5 fluorescent tag, Figure 3.13 (b), likely due to disorder of the alkyl chain or probe decomposition. However, unambiguous electron density for a complete Cy5 tag was observed close to the ABP molecule bound in the active site of chain A, Figure 3.13 (b). Specifically, the Cy5 fluorophore lies at the dimer interface in a hydrophobic pocket formed by residues Leu241, Tyr244, Pro245, Phe246 and Tyr313. It is important to note that due to the lack of electron density between the triazole linker and Cy5 tag, it is not possible to determine if the fluorophore observed at the dimer interface belongs to the ABP molecule bound in the active site of chain A, chain B or if it is from a separate decomposed ABP molecule. However, on inspecting the distance between the extended amide group of the probe bound in chain A and the fluorescent tag, it is conceivable that these two components are linked by a butyl chain. Nevertheless, there is no electron density to support this, and it is possible that the fluorophore comes from a third, decomposed ABP molecule.

Binding of the Cy5 tag of ABP **7** at the dimer interface of rGBA is consistent with a co-crystal structure of a BODIPY-tagged ABP (**8**) which had been reported previously in Davies group. In this previous study, crystals of rGBA were accidentally exposed to a BODIPY functionalised cyclophellitol probe following pre-incubation with a cyclophellitol aziridine inhibitor. In the resulting co-complex, the *N*-acyl aziridine inhibitor bound covalently to the catalytic nucleophile, whilst the BODIPY ABP was observed in the hydrophobic pocket at the dimer interface, Figure 3.14. Whilst the binding of this BODIPY ABP could be considered artefactual, it is notable that very simply torsional rotation of the linker allows the linker and the cyclophellitol to be placed in the active centre channel, putting the epoxide warhead in perfect super-position with the trapped aziridine without any movement of the BODIPY group itself. In combination with the Cy5-functionalised ABP **7** co-crystal complex reported in this work, these structures suggest that GBA can accommodate bulky, hydrophobic substituents in a hydrophobic pocket at the dimer interface, which presents as a genuine allosteric binding site, Figure 3.14 (b).

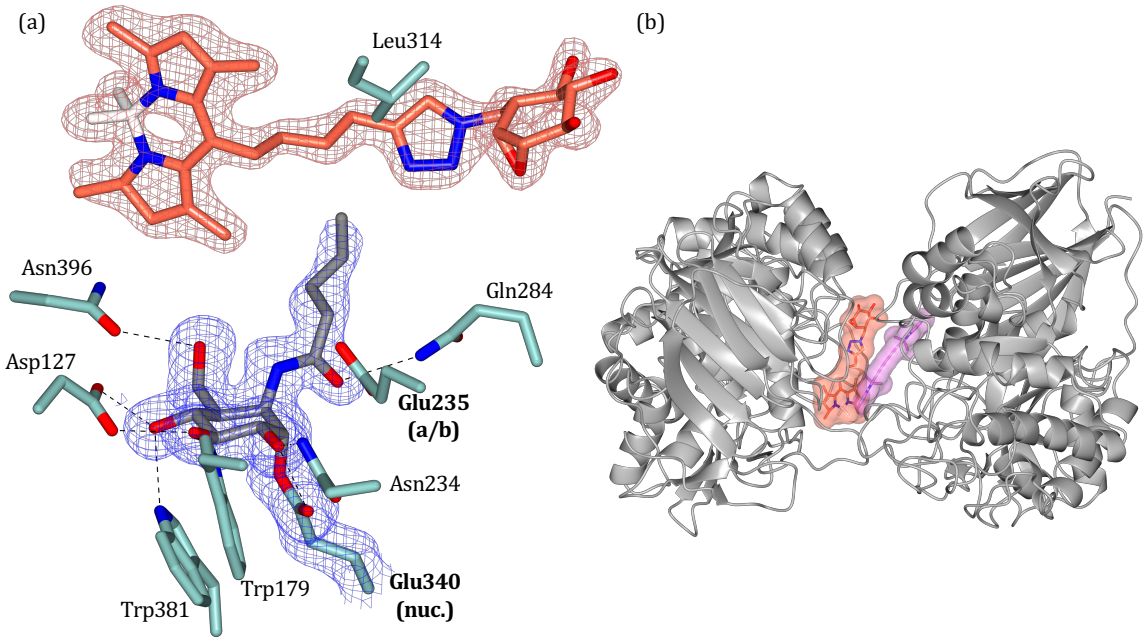


Figure 3.14: Co-complex of rGBA with an *N*-acyl cyclophellitol aziridine and a BODIPY-tagged cyclophellitol ABP (**8**). Data obtained by Dr Imogen Breen prior to this PhD. (a) Observed electron density for *N*-acyl cyclophellitol aziridine (grey) bound to the catalytic nucleophile (Glu340) of rGBA and for the intact BODIPY-tagged ABP **8** (orange) bound at the dimer interface. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand and Glu340 contoured to 1 σ (0.29 electrons/Å^3). (b) Ribbon and surface diagram demonstrating the binding of fluorescently tagged ABPs at the dimer interface: Cy5 fluorophore of ABP **7** in pink and BODIPY fluorophore of ABP **8** in orange.

*3.4.2.1 BODIPY-functionalised cyclophellitol ABP (**8**)*

To verify that binding of the BODIPY tag was not artefactual, a co-crystal structure of ABP **8** alone was obtained 1.86 Å resolution. The resulting co-crystal complex revealed clear electron density for the reacted cyclophellitol moiety of ABP **8** bound covalently to the catalytic nucleophile of GBA through classical trans-diaxial ring opening of the epoxide trap. Importantly, there was sufficient electron density to model the full, intact molecule of ABP **8** bound in molecule A of the rGBA dimer, Figure 3.15 (a). Specifically, the C6-triazole linker and subsequent 4-carbon alkyl chain was modelled through a broad active site cleft, formed by Tyr244, Pro245, Phe246, Tyr313 and Asn396, which extends towards the dimer interface, Figure 3.15 (b). Indeed, the BODIPY tag of ABP **8** is 'sandwiched' at the dimer interface between residues Leu241, Leu314, Phe316, Phe347 and Trp348 of chain A and residues Leu241, Leu314, Phe316 and Leu317 of chain B, which is consistent with the binding of the C6-substituent of **7**.

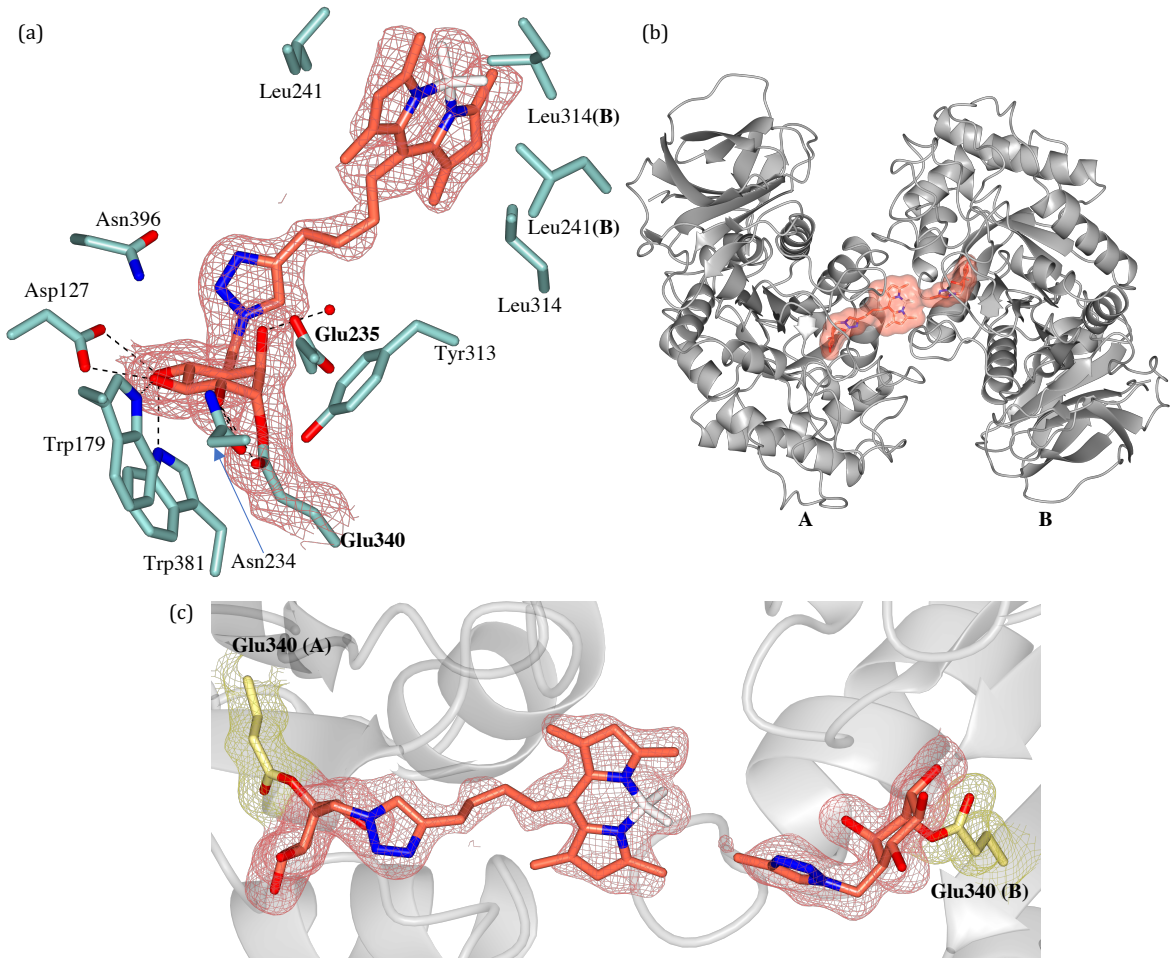


Figure 3.15: (a) Active site structure of ABP **8** bound to the catalytic nucleophile (Glu340) of rGBA. The C6-triazole linker and alkyl chain extend through an active site cleft towards the dimer interface where the BODIPY tag binds (b) Ribbon diagram of rGBA dimer highlighting the binding of the BODIPY-tag of ABP **8** (orange) in a hydrophobic cavity at the dimer interface. (c) Electron density for ABP **8** bound to Glu340 (yellow) in both chains of the GBA dimer, showing the C6-substituent of both ABP molecules extending towards the dimer interface. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand and Glu340 contoured to 0.8 σ.

It should also be noted that the triazole linker of the molecule of **8** bound in the active site of chain B also extends towards the dimer interface; however, the density for the alkyl chain is less clear and only one BODIPY tagged could be modelled at the dimer interface, Figure 3.15 (c). This poor electron density may result from disorder of the alkyl chain or probe decomposition. Nevertheless, in combination with the co-crystal structure of Cy5 tagged ABP **7**, this complex provides further evidence for a unique binding mode of this hydrophobic binding cavity. More recently, this allosteric site has been exploited for the binding of a novel class of pyrrozo-

pyrazine activators with chaperoning behaviour[276], indicating this binding site may hold promise for the development of GBA active compounds with therapeutic potential.

To further investigate the importance of this hydrophobic binding site at the dimer interface, a quick comparative structural study against related β-glucosidases was performed to reveal that this hydrophobic pocket is unique to GBA. It appears other β-glucosidases do not exhibit this hydrophobic cavity, so are less able to accommodate large C6-substituents..

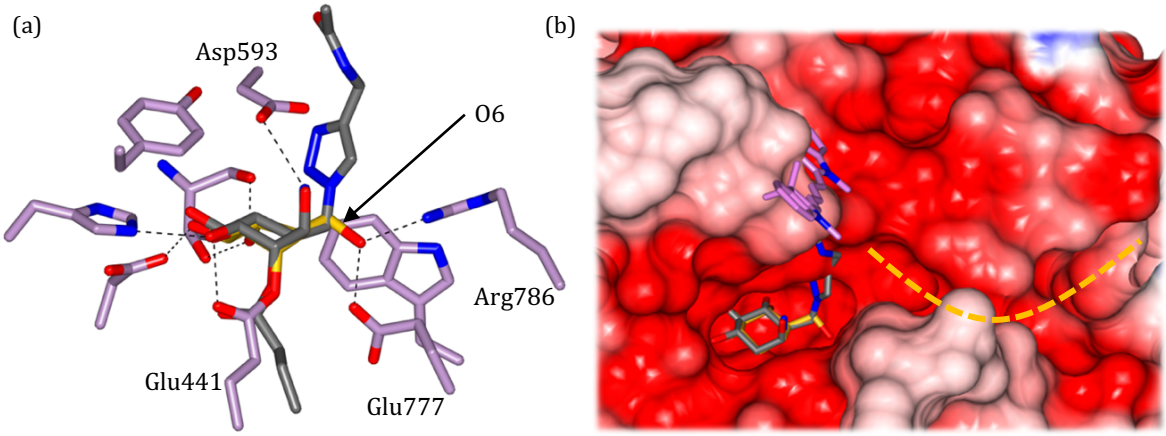


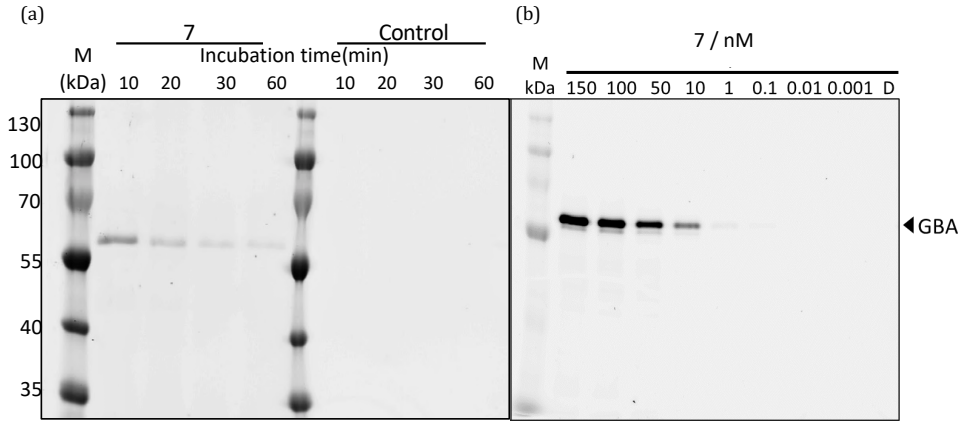
Figure 3.16: Accommodation of C6-derivatised ligands in GBA and TxGH116 (a) Overlay of bacterial TxGH116 structure (PDB 5NCX[575], active site residues in lavender, aziridine inhibitor in yellow) with structure of GBA in complex with **7** (ABP in grey). Clear "inwards" orientation of the C6 hydroxyl of the TXGH116 aziridine ligand, resulting from interactions with Arg786 and Glu777. (b) Partial electrostatic surface of TxGH116 overlaid with the ligand coordinates for ABP **7** (grey) bound to GBA, indicating potential steric clashes with the C6 substituent of **7** pink). Narrow hydrophobic pocket of TxGH116 highlighted by dashed yellow line.

In fact, superposition of the co-crystal structure of rGBA in complex ABP **7** with a structure of bacterial β-glucosidase TxGH116 (*Thermoanaerobacterium xylanolyticum*, PDB 5NCX[575]), which is a close homologue of human GBA2[701], demonstrates that the C6-functionalised ABP would not be readily accommodated, Figure 3.16. When **7** is superimposed into the active site of TxGH116, the C6 position points upwards and away from the hydrophobic pocket of the enzyme, which is considerably more narrow and less hydrophobic than that observed in GBA, Figure 3.16 (b). In contrast, the C6 position of the aziridine ligand already bound in the active site of TxGH116 points inwards and towards the narrow hydrophobic pocket, presumably as a result of interactions with Arg786 and Glu777, Figure 3.16 (a). Additionally, Asp593 considerably restricts the space available for the C6-substituent. A combination of all these

factors likely hinders the binding of C6-functionalised cyclophellitols to the human GBA2 homologue TxGH116, thus allowing sufficient discrimination between GBA and GBA2 to be achieved[701]. Therefore, the attachment of bulky hydrophobic moieties at the C6-position of cyclophellitol appears to impart improved selectivity for GBA by taking advantage of this hydrophobic pocket.

3.4.2.2 In-Solution Fluorescence Labelling

Subsequent to structural analysis, activity-based labelling of rGBA by ABP **7** was investigated through a time-course labelling experiment and through a ABP titration experiment to estimate the in-gel detection limit of this probe. The resulting samples were separated by SDS-PAGE and the fluorescently labelled protein was identified by fluorescent readout of the slab gel.



(c)

	IC50 / μM		
In vitro	CBE	CP	7
r-GBA	4.28 ± 0.5	0.030 ± 0.002	0.003 ± 0.000
rGBA2	101 ± 20	0.030 ± 0.003	41.2 ± 10.1
r-GAA	1900 ± 192	>100	>100

Figure 3.17: Time course labelling of rGBA (200 nM) by ABP **7** (200 nM) demonstrating a drop in labelling over time. (b) Concentration dependent labelling of rGBA by ABP **7** showing gel detection limit of 0.1 nM ABP **7**. D = Denatured sample (c) Apparent *in vitro* IC_{50} values against GBA, GBA2 and GAA (IC_{50} data supplied by M. Artola - University of Leiden)[562].

Interestingly, labelling of rGBA reached saturation within 10 minutes, with a drop in labelling observed thereafter, Figure 3.17 (a). This was a surprising observation, as labelling was expected to be irreversible and accumulative over time. However, these results suggests that this ABP reversibly inhibits GBA and is slowly hydrolysed by the catalytic machinery of the enzyme. This behaviour should be considered when performing future experiments with this probe. Nevertheless, this ABP has proved incredibly potent for GBA *in vitro* (IC_{50} ~3 nM),

showing concentration dependent labelling with a detection limit of 0.1 nM ABP **7** by SDS-PAGE, Figure 3.17 (b). Additionally, this probe has proved more potent than both CBE (IC_{50} ~5 μM) and cyclophellitol (IC_{50} ~30 nM), with considerably improved selectivity for GBA over other glucosidases, including GBA2 (non-lysosomal glucosylceramide) and GAA (lysosomal α-glucosidase), Figure 3.17 (c). These apparent IC_{50} values demonstrate the increased potency and selectivity of C6-functionalised cyclophellitols for GBA, which likely results from their ability to bind to the hydrophobic allosteric binding site of at the dimer interface of GBA.

3.4.3 Structure-Based Development of C6-Substituted Cyclophellitols

The potential to generate more selective cyclophellitol-based GBA inactivators by functionalisation of the C6 hydroxyl holds considerable promise for improving the application of such compounds in the study of GD. Notably, access to more selective ABPs will permit improved labelling and monitoring of active GBA, which may benefit applications in diagnostic and therapeutic contexts. In regard to the study of GD pathogenesis and treatment, there is an urgent demand for appropriate GD animal models which link impaired GBA function to GD. Numerous murine GD models have been generated through a chemical knockdown strategy in which CBE and cyclophellitol were used to irreversibly inhibit endogenous GBA and induce GD phenotypes[702,678]. However, the poor selectivity of these inhibitors has hindered their use in the generation of reliable Gaucher animal models. In light of the improved potency and selectivity of C6-functionalised cyclophellitol ABPs, we postulated that C6-functionalised cyclophellitol-based inhibitors may be more suitable inhibitors for the generation of viable GD animal models through a chemical knockdown strategy. Subsequently, the Overkleeft lab synthesised a range of C6-functionalised cyclophellitols, including a C6-biphenyl substituted cyclophellitol (**9**) and a C6-adamantly substituted cyclophellitol (**10**), Figure 3.10, as potentially more selective GBA inhibitors[562]. Herein, the conformation and reactivity of these inhibitors was structurally evaluated on the 3D structure of GBA to further investigate the binding of the C6 substituents.

3.4.3.1 3D Complex Structure with C6 Biphenyl-Substituted Inhibitor

Initially, a co-crystal structure of rGBA in complex with the biphenyl-cyclophellitol inhibitor (**9**) was obtained in hopes of analysing the binding of the hydrophobic biphenyl group. Data for this complex were refined to 1.63 Å resolution to reveal a single molecule of **9** bound covalently to the catalytic nucleophile of both GBA molecules in the asymmetric unit, Figure 3.18. The reacted

cyclophellitol moiety was found to adopt the $^4C_1$ chair conformation, forming hydrogen bonds with active site residues Asp127, Trp179, Asn234, Trp381 and Asn396. Whilst the C6-triazole linker could be modelled, there was insufficient electron density to model the biphenyl-substituent. Given the rigid and planar nature of the biphenyl substituent, the absence of electron density most likely reflects decomposition of the inhibitor rather than significant disorder, however, the reason for decomposition is unknown. Despite demonstrating the expected mechanism-based mode of inhibition, this structure provides no information on the binding of the C6-biphenyl substituent. Therefore, structural analysis of rGBA in complex with the adamantly-cyclophellitol inhibitor was performed in an effort to elucidate the binding mode of the C6-adamantyl substituent.

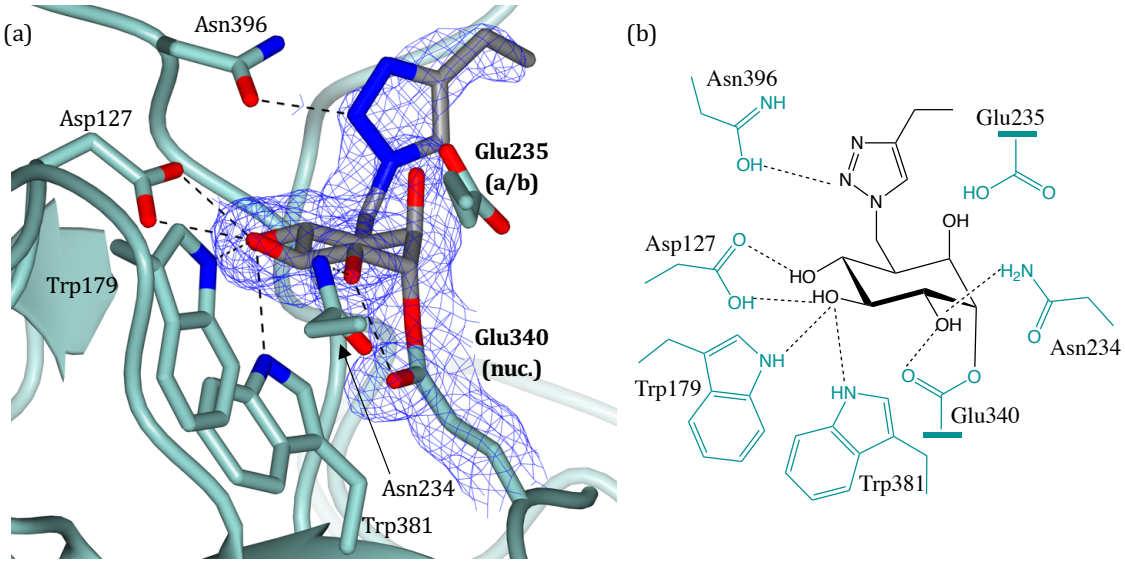


Figure 3.18: (a) 3D active site structure of inhibitor **9** covalently bound to the catalytic nucleophile (Glu340) of rGBA in a $^4C_1$ chair conformation. Only the cyclophellitol moiety and triazole linker could be modelled. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand and Glu340 contoured to 1σ (0.22 electrons/Å^3). (b) Schematic of the hydrogen bond network of **9** in the active site of GBA.

3.4.3.2 3D Complex Structure with C6 Adamantyl-Substituted Inhibitor

Data for the C6-adamantyl inhibitor (**10**) co-crystal complex were collected to 1.81 Å resolution and solved by molecular replacement to reveal a single molecule of **10** bound covalently in the active site of both rGBA molecules in the asymmetric unit. Specifically, unambiguous electron density was observed for the ring-opened cyclophellitol species bound covalently to the catalytic nucleophile, Figure 3.19. Importantly, sufficient electron density for

the triazole moiety, ether linker and adamantane substituent was observed, allowing the full inhibitor to be modelled in both molecules, Figure 3.19. The triazole linker binds to the broad active site cleft formed by Tyr313, Phe246 and Trp348, which was reported to accommodate the triazole linker of ABP **7**. Additionally, the adamantane group was modelled to bind in the hydrophobic cavity at the dimer interface of rGBA, which is consistent with the binding of the Cy5-tag of ABP **7**, Figure 3.20. Indeed, the triazole linker of both ABP **7** and inhibitor **10** were modelled in almost identical conformations, with the Cy5 tag and adamantane groups binding in a similar region of the hydrophobic pocket at the dimer interface, Figure 3.20.

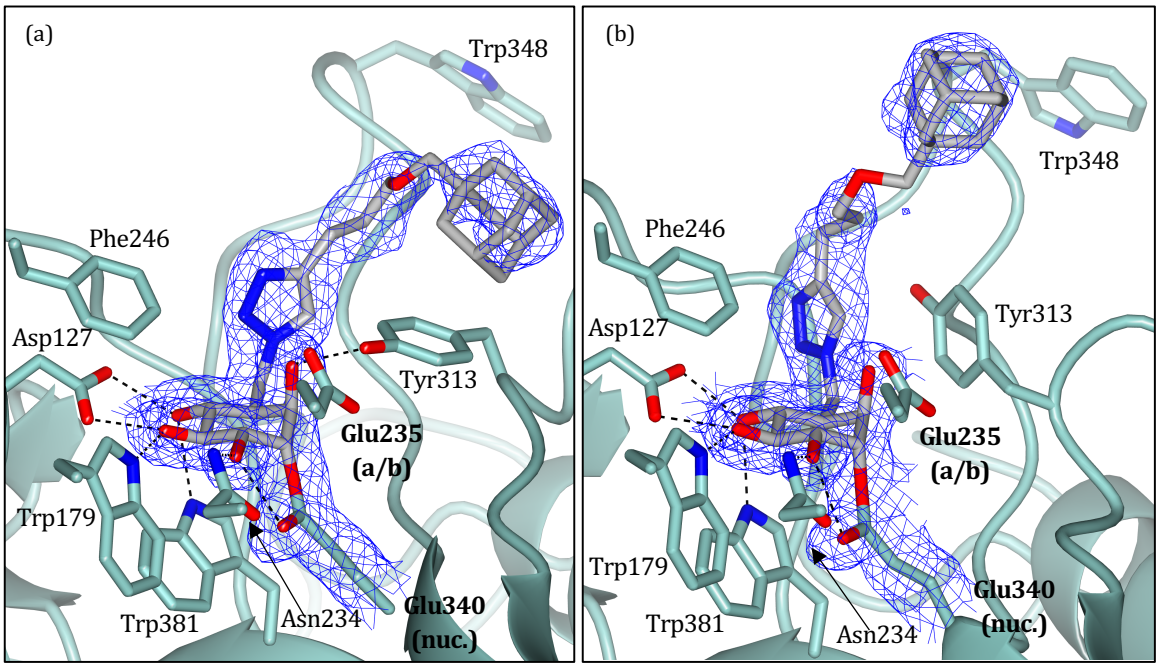


Figure 3.19: 3D active site structure showing covalent binding of inhibitor **10** in the active site of (a) chain B and (b) chain A, demonstrating a change in orientation of the adamantly substituent (likely owing to binding through predominantly hydrophobic interactions). Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand and Glu340 contoured to 1 σ (0.21 electrons/Å^3).

Interestingly, the linker and adamantane moiety of inhibitor **10** were observed to bind in slightly different orientations in each molecule of GBA in the dimer, likely resulting from binding via predominantly hydrophobic interactions. Of note, is the "upward" extension of the adamantyl substituent in chain A, Figure 3.19 (b). Additionally, a change in the orientation of Tyr313 was observed, with Tyr313 adopting a "downward" conformation in chain B, presumably as a result of the alternative conformation of the C6-adamantane group. Moreover, this change in Tyr313 conformation results in the formation of an additional hydrogen bond

with the ring opened epoxide, Figure 3.19 (a). It is clear from this crystal structure that the previously identified hydrophobic cavity at the dimer interface, which is absent in related β-glucosidases, is capable of accommodating a variety of hydrophobic substituents. This provides a structural basis for the inhibitory preference of GBA and the improved potency of C6-functionalised inhibitors.

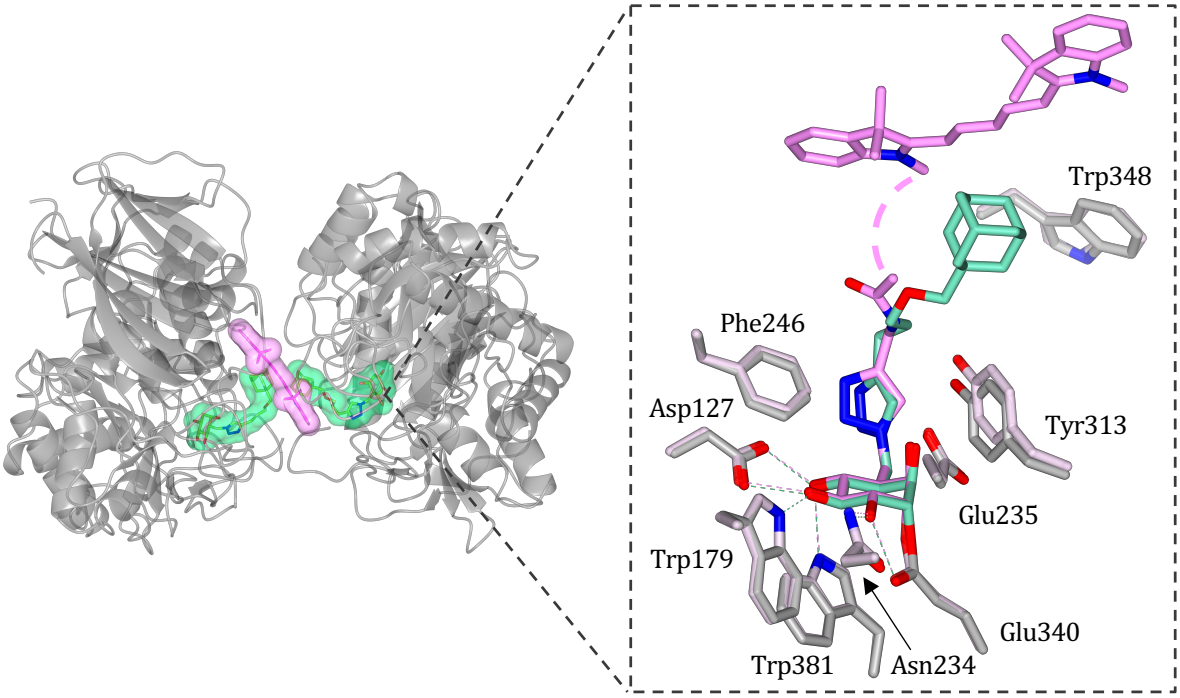


Figure 3.20: Ribbon and surface diagram of rGBA in complex with **10** showing the binding of the adamantyl group (green surface) to a similar region of the hydrophobic pocket occupied by the Cy5 tag (pink surface) of ABP **7**. Overlay shows the active site residues of the biphenyl complex (residues in grey, inhibitor in green) superimpose with the active site residues of the ABP **7** complex (residues in pale pink, ABP in pink), including residues Tyr313, Phe246 and Trp348 which form the hydrophobic cavity that accommodates the C6-triazole linker.

3.4.3.3 Inhibitory Potency and Selectivity of C6-functionalised Cyclophellitol Inhibitors

Following structural investigations, the *in vitro* inhibitory activity of the C6-functionalised cyclophellitol inhibitors was evaluated by our collaborators at Leiden University[562]. The inhibitors were incubated with rGBA and two major off-target glycosidases; human GBA2 and human GAA. Following incubation, the residual enzymatic activity was determined through fluorogenic substrate assays. Both **9** and **10** were shown to be nanomolar inhibitors of rGBA, being 4000-times and 200-times more selective than CBE respectively. More importantly, both inhibitors were inactive towards GBA2 and rGAA (IC_{50} > 100 μM)[562], Table 3.4.It is evident from

the IC_{50} values that both C6-functionalised cyclophellitol inhibitors **9** and **10** exhibit improved potency and selectivity for GBA than CBE and cyclophellitol. In combination with the structural information uncovered in this work, these values further support the concept that C6-functionalised cyclophellitols benefit from enhanced GBA potency and selectivity as a result of their ability to bind to the unique hydrophobic binding cavity of GBA.

Table 3.4: Apparent IC_{50} values for *in vitro* inhibition of recombinant GBA, GBA2 and GAA by CBE, cyclophellitol (CP), **8** and **9**. Error ranges depict stdev from biological triplicates. Data supplied by M. Artola (University of Leiden)[562].

	IC_{50} / μM			
In vitro	**CBE**	**CP**	**9**	**10**
rGBA	4.28 ± 0.5	0.030 ± 0.002	0.001 ± 0.000	0.001 ± 0.000
rGBA2	101 ± 20	0.030 ± 0.003	>100	>100
rGAA	1900 ± 192	>100	>100	>100

3.4.4 3D Complex with Bi-functional Cyclophellitol Aziridine Activity-Based Probe

In previous work by the Overkleeft lab, *N*-functionalised cyclophellitol aziridines proved to be more potent GBA inhibitors than the analogous un-functionalised aziridines[556]. In combination with the enhanced potency and selectivity of the C6-functionalised inhibitors analysed in this work, it was hypothesised that bi-functional cyclophellitol aziridine probes, which are functionalised at the C6-position and the aziridine nitrogen, may exhibit further improvements in potency and selectivity. Consequently, the Overkleeft lab synthesised a bi-functional cyclophellitol aziridine ABP (**11**), which is substituted at the C6-position with a Cy5 tag and substituted at the aziridine nitrogen with an octyl chain, Figure 3.10. It was hoped this ABP would be the most potent cyclophellitol-based GBA inhibitor to date. Unfortunately, in primary *in vitro* studies, this bi-functional ABP was found to label GBA with efficiency on par with uni-functionalised APB **7** (apparent IC_{50} of 3.2 nM for ABP **7** vs 53 nM for ABP **11**), suggesting that bi-functional probes do not benefit from improved potency. Although this was a disappointing finding, structural analysis of ABP **11** in complex with rGBA (produced in the insect-baculovirus expression vector system described in **Chapter 4**) was performed to elucidate its binding mode.

Initially, structural analysis of the bi-functional probe was hindered by its low solubility, requiring a minimum of 10% DMSO in the crystal drop to ensure the probe was sufficiently

dissolved. This put the crystals under significant strain, nevertheless, a suitable co-crystal structure was eventually obtained. It would be prudent to note that this poor solubility may also be a limiting factor in the labelling efficiency of ABP **11**.

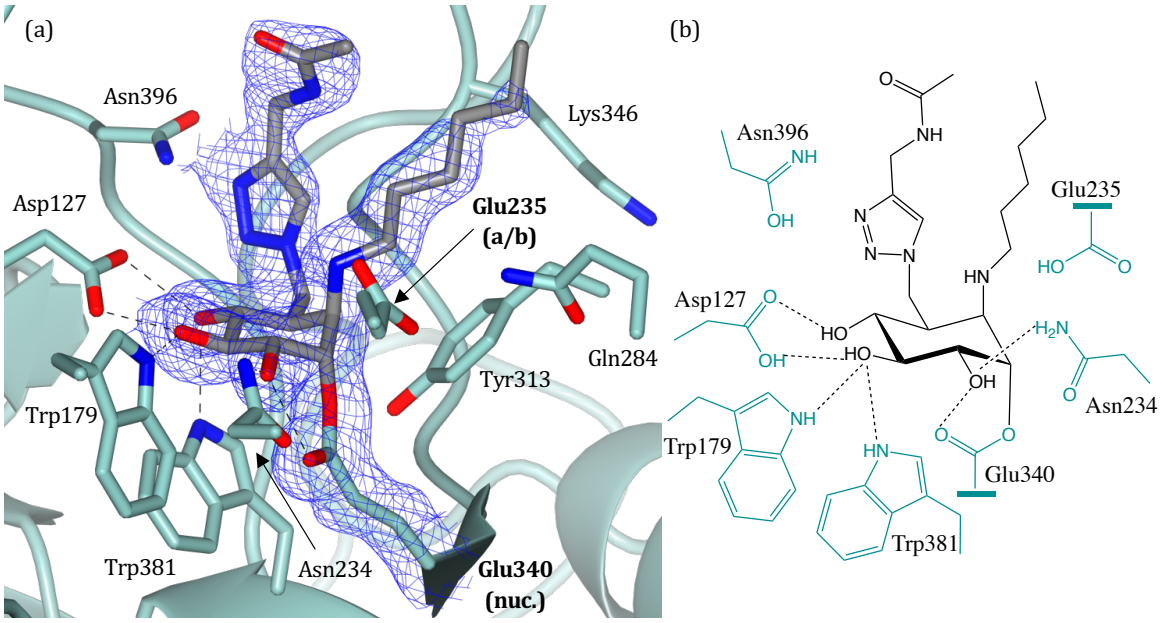


Figure 3.21: (a) 3D active site structure of ABP **11** covalently bound to Glu340 of rGBA forming a covalent complex in $^4C_1$ conformation. The ring opened *N*-alkyl aziridine extends through a narrow active site pocket formed by residues Tyr313, Gln384 and Lys346. Insufficient density was observed to model the C6-alkyl linker and Cy5 tag. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand and Glu340 contoured to 1 σ (0.22 electrons/Å^3). (b) Schematic of the hydrogen bonding network of ABP **11**.

Data for the bi-functional ABP (**11**) complex were solved to 1.80 Å resolution and to reveal a single molecule of **11** bound covalently to the catalytic nucleophile of rGBA, Figure 3.21. The reacted cyclophellitol adopts the expected $^4C_1$ chair conformation and forms hydrogen bonds with a number of active site residues including Asp127, Trp179, Asn234 and Trp381. Importantly, sufficient electron density for the ring opened *N*-alkyl aziridine warhead was observed, allowing the first 6 carbons of the octyl chain to be modelled. This was sufficient to establish binding of the alkyl chain to a narrow active site pocket formed by residues Tyr313, Lys346 and Gln284. Specifically, the alkyl chain extends through this cleft towards the surface of the protein, which may reflect the binding of the fatty acid portion of the natural GlcCer substrate which is thought to project out from the protein and interact with the lipid bilayer. Whilst clear electron density was observed for the C6-triazole linker and subsequent amide group, the Cy5 tag of ABP **11** could not be modelled. This likely results from decomposition of

the probe. Nevertheless, when overlayed with the Cy5-functionalised ABP (**7**) co-crystal structure, it is apparent that the triazole linker of ABP **11** binds in a similar conformation in the same active site cleft formed by Trp348, Phe246 and Tyr313, Figure 3.22 (a). This cleft extends towards the broader hydrophobic allosteric site at the dimer interface where the Cy5 tag of ABP 7 was observed to bind, Figure 3.22 (b).

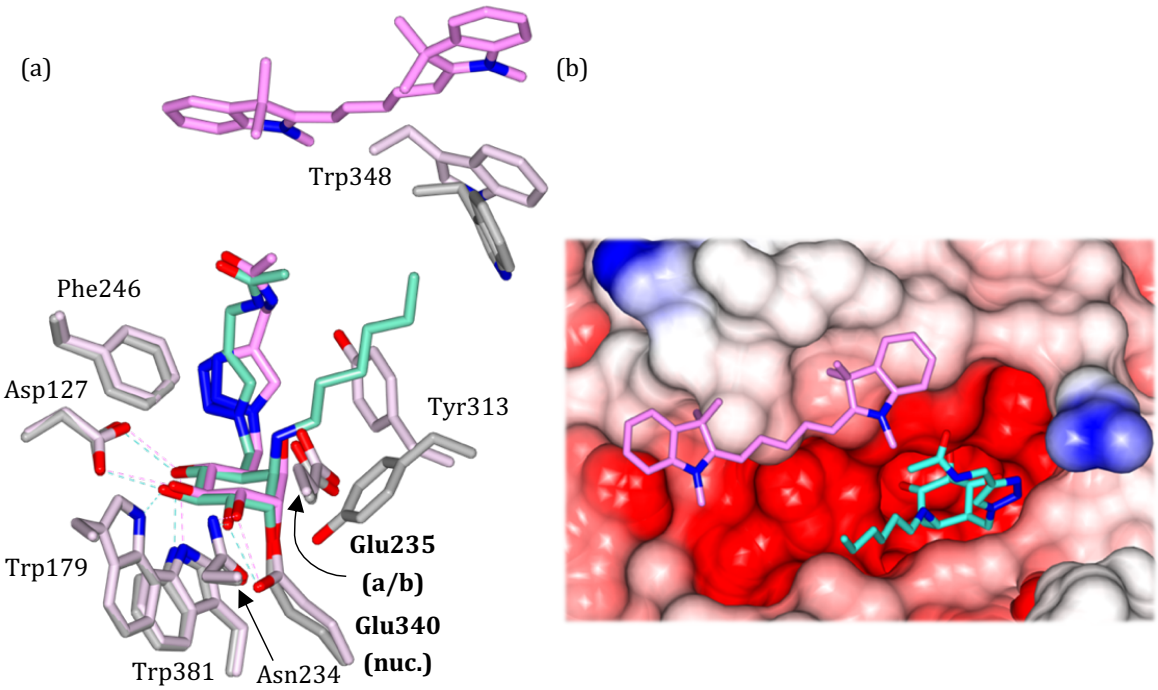


Figure 3.22: (a) Overlay of **11** co-crystal structure (ABP in teal, active site residues in grey) and **7** co-crystal structure (ABP in pink, active site residues in light pink). Triazole linkers of both ABPs adopt similar conformations and occupy the same hydrophobic pocket formed by Trp348, Phe246 and Tyr313. (b) Surface diagram showing superposition of **11** into active site of ABP **7** structure, exposing how the *N*-alkyl aziridine (teal) is accommodated in a separate, narrow and less hydrophobic pocket which extends towards the protein surface.

Despite demonstrating that the O6- and aziridine nitrogen substituents of ABP **11** are structurally exclusive and are accommodated in two unique active site clefts of GBA, ABP **11** exhibits no further improvements in potency or selectivity compared to the uni-functionalised equivalents. Nevertheless, this bi-functional ABP is a nanomolar inhibitor of GBA, which provides future opportunities for ABP development through both the C6- and aziridine nitrogen

Given no improvements in selectivity or potency were observed for the bi-functional ABP, our focussed turned back to the C6 uni-functionalized cyclophellitols. Specifically, our collaborators at Leiden University investigated the C6-functionalised cyclophellitol inhibitors **9** and **10** for

use in chemical knockdown studies to generate Gaucher models in zebrafish. All animal model experiments discussed in the following section were performed by researchers at Leiden University; however, they have been included for completeness.

3.4.5 Chemical Knockdown of GBA Activity in Zebrafish

Firstly, *in vivo* GBA inhibition was investigated in *Dario rerio* zebrafish embryos by incubation with inhibitors **9** or **10** for 5 days, after which enzyme selectivity analysis was performed using appropriate ABPs for GBA, GBA2 and GAA[562]. Quantification of the ABP labelled bands revealed both inhibitors to be 4,000-fold more potent than CBE, with *in vivo* apparent IC_{50} values of 4-10 nM, Table 3.5. Furthermore, improved selectivity for GBA was observed; at 0.1-10 μM of **9** or **10**, labelling of rGBA with broad-spectrum retaining β-glucosidase ABPs was abolished, whilst labelling of GBA2, GAA, ER α-glucosidase (GANAB) and lysosomal β-glucuronidase (GUSB) was unaffected, Figure 3.23. Interestingly, a 10-30-fold increase in glucosylsphingosine (GlcSph) levels was also observed, strongly indicating efficient *in vivo* inhibition of GBA was achieved with subsequent accumulation of GlcCer and GlcSph. Comparatively, CBE was required at 1,000-10,000-fold higher concentrations to achieve similar a similar elevation in GlcSph levels.

Table 3.5: Apparent IC_{50} values for *in vivo* GBA inhibition in 5-day treated zebrafish embryos with CBE, cyclophellitol, **8** and **9.** Data supplied by M. Artola (University of Leiden)[562]

	IC_{50} / nm			
In vivo	**CBE**	**CP**	**9**	**10**
rGBA	44.1×10^4	83	5.85 ± 2.44	3.94 ± 1.2
rGBA2	8.9×10^4	59	$>10^4$	$>10^4$
rGAA	9.55×10^4	$> 10^5$	$>10^4$	$>10^4$

Lastly, the brain permeability of inhibitor **10** was investigated as its ability to cross the blood brain barrier is vital for the generation of neuropathic GD models. Consequently, adult zebrafish were administered with DMSO as controls, BODIPY-functionalised ABP **12** or inhibitor **10** via food intake, Figure 3.23. After 16 hours, brains and other organs were isolated, homogenised and analysed using a range of β-glucosidase ABPs to visualise residual GBA activity. ABPP of the brain homogenates revealed considerable GBA activity in the control and ABP **12** treated fish, but no labelling of GBA was observed in fish treated with inhibitor **10**[562], Figure 3.23. This indicates that inhibitor **10** is able to cross the blood brain barrier and abrogate GBA activity in the brain. Additionally, labelling by a broad-spectrum ABP revealed that GBA2 is not a target of **10**, Figure 3.23, demonstrating selective inhibition of GBA

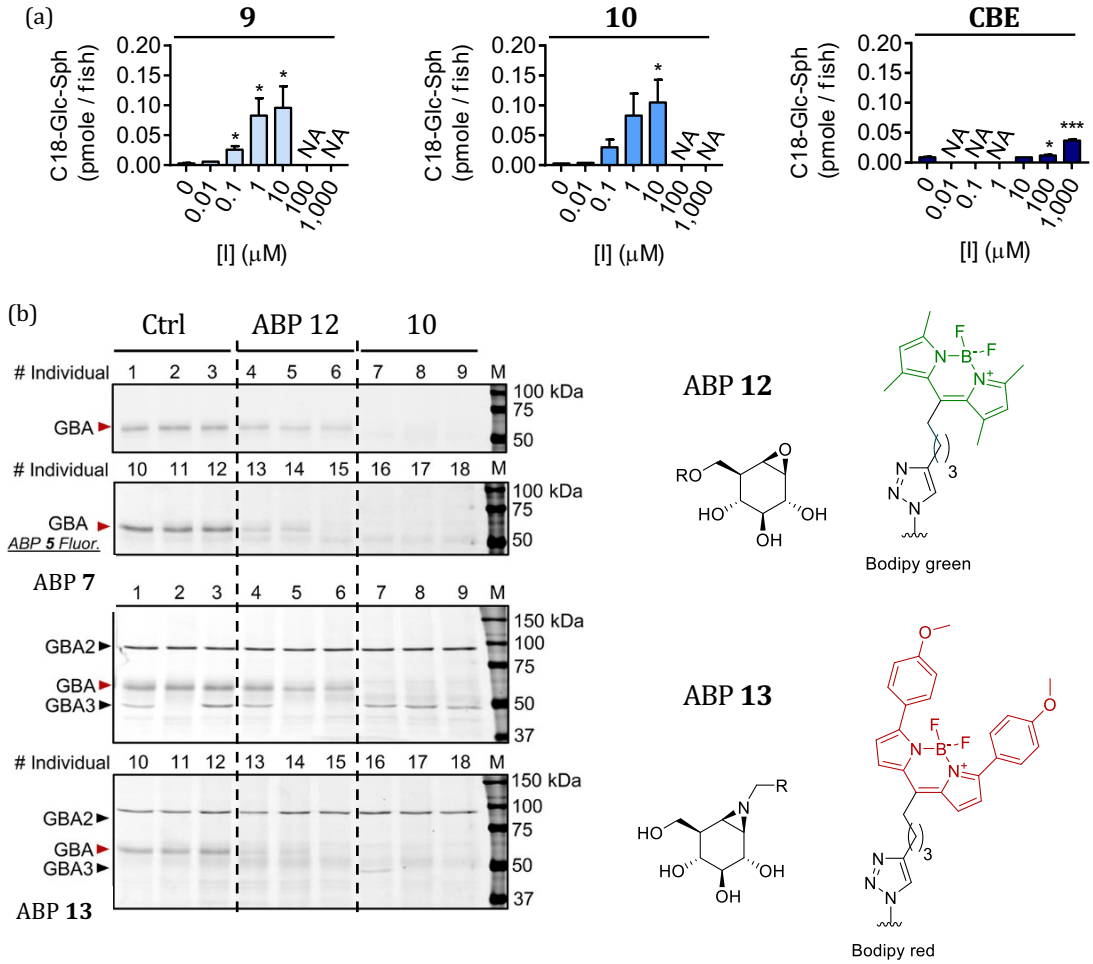


Figure 3.23: *In vivo* action of inhibitors **9** and **10** (a) Increase in GlcSph levels in zebrafish embryos treated for 5 days with inhibitors **9, 10** or **CBE**. (b) *In vivo* targets of inhibitor **10** in brains of adult zebrafish as visualised by competitive ABPP using GBA selective ABP **7** or broad-spectrum ABP **13**. I*n vivo* inhibition profile of ABP **12** used for comparison. Visualisation of labelled GBA by fluorescent readout, showing selective inhibition of GBA by **10**. Data supplied by L. Lelieveld (Leiden University).

These results demonstrate that C6-functionalisation of cyclophellitol-based inactivators yields potent and selective GBA inhibitors. In particular, the adamantane- and biphenyl-functionalised cyclophellitol inactivators are selective, nanomolar inhibitors of GBA both *in vitro* and *in vivo.* Given the growing demand for viable GD animal models to aid in the study of GD pathophysiology, these C6-functionalised inhibitors are superior to both CBE and cyclophellitol for generating GBA deficient zebrafish models through chemical knockdown. Moreover, owing to their improved selectivity, potency and ability to cross the blood-brain barrier, these inhibitors also offer a route to generating neuropathic GD models which have previously been very difficult to obtain without compromising animal model viability.

3.5 Summary

Tagged cyclophellitols offer a powerful activity-based protein profiling approach to the visualisation and quantification of specific enzymatic activities. Here, the design and structural analysis of range of cyclophellitol epoxide and aziridine activity-based probes (ABPs) and inhibitors for human β-glucocerebrosidase (GBA) is described.

Initial crystallographic studies of GBA in complex with a Cy5-functionalised cyclophellitol ABP, exposed a hydrophobic pocket at the dimer interface of GBA which can accommodate large hydrophobic moieties. This hydrophobic cavity was found to be unique to GBA, providing a structural basis for the improved selectivity of C6-functionalised inhibitors and ABPs. This structural information subsequently informed the design of new C6-functionalised cyclophellitol inhibitors, including a C6 biphenyl- and C6 adamantly-substituted cyclophellitol, which we hoped would offer more potent and selective inhibition of GBA. Structural analysis of the adamantane inhibitor revealed binding of the adamantyl group to the same hydrophobic pocket occupied by the Cy5 tag of the cyclophellitol ABP, providing further evidence for a recognition or binding role of this hydrophobic pocket. Importantly, both C6-functionalised cyclophellitols were found to be potent, nanomolar inhibitors of GBA, with considerably improved selectivity compared to CBE and cyclophellitol. Therefore, this work suggests that attachment of bulky hydrophobic moieties at the C6-position of cyclophellitol imparts improved selectivity for GBA by taking advantage of its unique hydrophobic cavity.

To investigate the *in vivo* inhibitory activity of the C6-functionalised inhibitors, our collaborators at Leiden University performed chemical knockdown studies in *Dario rerio* zebrafish embryos. Of note, the adamantane-substituted inhibitor was shown to be 4000-fold more potent that CBE, with considerably improved selectivity for GBA. Additionally, treatment with this inhibitor results in a 10-30-fold increase in glucosylsphingosine levels *in vivo,* demonstrating effective GBA inhibition. Furthermore, this inhibitor selectively blocked GBA activity in the brain of adult zebra fish, demonstrating the ability to cross the blood-brain barrier. Therefore, this C6-adamantyl substituted cyclophellitol presents as a superior inhibitor to both CBE and cyclophellitol for generating GBA deficient zebrafish models with the potential to investigate neuropathic Gaucher disease.

Chapter 4: Insect-Baculoviral Expression of Recombinant Human β-Glucocerebrosidase

4.1 Abstract

The lysosomal glycoside hydrolase, β-glucocerebrosidase (GBA) catalyses the hydrolysis of glycosphingolipids. Inherited mutations in the *GBA1* gene cause defects and deficiencies in GBA activity, which subsequently lead to the lysosomal storage disorder Gaucher disease (GD). More recently, associations between *GBA1* mutations and Parkinson disease have also been identified. Consequently, GBA is of considerable clinical and academic interest, with continuous advances in the development of inhibitors, chaperones and activity-based probes. Development of novel GBA active compounds requires a source of functional protein, however, due to the lack of reliable expression systems for GBA, the majority of non-clinical structural and mechanistic studies on this enzyme today rely on expired enzyme replacement therapy (ERT) formulations. Such formulations are incredibly costly and often difficult to obtain in adequate supply, typically requiring material transfer agreements.

Herein, the production of active non-clinical GBA in insect cells using the baculovirus expression vector system is described. This formulation exhibits comparable activity and biophysical properties to ERT preparations and is readily crystallised for x-ray crystallography studies. Indeed, 3D crystal structures of this formulation in complex with a number of GBA active compounds were obtained, including a co-crystal complex with the glucoside inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside, demonstrating the utility of this GBA formulation for ligand-binding studies. Furthermore, a previously unseen crystal form of GBA was obtained which diffracted to give a 0.98 Å unliganded structure, the highest resolution structure of this protein deposited to date.

In light of the purity, stability and biochemical properties of this GBA formulation, the production protocol described here should circumvent the need for ERT formulations and provide a source of non-clinical GBA for biochemical and structural studies within GD research.

*Some of the work discussed in this chapter is published in R. J. Rowland, L. Wu, F. Liu, G. J. Davies, *Acta Cryst*, 2020, **D76**, 565-580.

4.2 Introduction

4.2.1 Importance of β-Glucocerebrosidase (GBA)

As discussed in **Chapters 1** and **3**, β-Glucocerebrosidase (GBA, EC 3.2.1.45) is a membrane-associated lysosomal glycoside hydrolase which is primarily responsible for catalysing the degradation of glucosylceramide (GlcCer)[172]. Inherited deficiencies in GBA activity cause an accumulation of GlcCer within cells throughout the body, subsequently leading to the most common lysosomal storage disorder Gaucher Disease (GD)[174,219]. GD can severely impact patients' lives with multi-organ disease manifestations and premature death[189]. The considerable variation in disease phenotypes and lack of reliable genotype-phenotype relationships often hinder the diagnosis and treatment of GD. This is further complicated by the large number and rarity of GD causing *GBA1* mutations[149,232,274]. More recently, mutations in the *GBA1* gene have also been identified as the highest known genetic risk factor for Parkinson disease[245,355]. Consequently, GBA is of considerable clinical and academic interest, with continuous advances in the development of inhibitors[277,562,703,704], chaperones[705–708] and activity-based probes[557,560,562,704] to study this enzyme in disease pathogenesis. However, the development of new GD diagnostics, therapeutics and GBA active compounds requires a source of functional protein for biochemical and structural analyses.

4.2.2 Natural Sources of GBA

Identification of defective GBA function as the biochemical basis of GD initiated studies into the purification and characterisation of human GBA[174]. This proved challenging due to the relatively low abundance of the protein in natural sources, its tendency to adhere to membranes and its intrinsic glycan heterogeneity[299]. However, in seminal work by Pentchev et al. (1973)[299], GBA was purified from placental tissue by ammonium sulfate fractionation, gel filtration and weak cation exchange chromatography. The poor stability of GBA proved to be the biggest challenge, requiring the presence of detergents to keep the protein stable during purification. Nevertheless, GBA was purified to homogeneity with a typical yield of 330 μg from 1.7 kg of fresh placental tissue[299]. Whilst this yield was low and the ability to scale up was limited, there was sufficient protein for characterisation of enzymatic activity, protein stability and glycosylation. Furthermore, this work provided a strong basis for the development of enzyme

replacement therapy (ERT). Specifically, the ERT formulation Ceredase® was developed from this work following improvements to the purification procedure and modification of protein glycosylation[709]. However, the thousands of metric tonnes of human placenta required to support Ceredase® production imposed significant practicality and cost issues[309,709]. Consequently, there was a strong incentive to develop recombinant strategies for GBA production.

4.2.3 Recombinant Sources of GBA

Given the post-translational glycosylation required to ensure proper folding, GBA cannot be produced in prokaryotic systems[647]. Consequently, eukaryotic systems with the necessary post-translational modification capabilities must be employed. This is exemplified by existing ERT expression systems for clinical GBA (see **Chapter 1** *section 1.4.8.1*).

4.2.3.1 Clinical Expression Systems

In light of the therapeutic potential of ERT, the Genzyme Corporation (Sanofi Genzyme, USA) developed Cerezyme®, a mannose-terminated recombinant GBA formulation produced in a Chinese hamster ovary cell line (CHO)[309,312]. Cerezyme® has been used as an ERT formulation for the treatment of GD since 1994 and remains the most prominent ERTs to date[312]. Alternative recombinant formulations include Velaglucerase alfa (Vpriv®, Shire HGT Inc, USA) produced by specific gene activation in a HT-1080 cell line[320,322] and Taliglucerase alfa (Elelyso®, Pfizer, USA) expressed in carrot-root cells.[323,325,328]. As discussed in **Chapter 1**, these formulations have proved safe and effective for the treatment of GD, demonstrating the ability to improve and even reverse GD manifestations[313]. However, these formulations are incredibly costly[331].

4.2.3.2 Non-Clinical Expression Systems

In addition to the ERT expression platforms, production of non-clinical GBA has been attempted in mammalian systems such murine cells[710] and COS-1 cells[711], plant systems including the *Arabidopsis thaliana* plant[326] and glycoengineered *Nicotiana benthamiana* plants[712], as well as *Pichia pastoris*[713] and insect-baculoviral expression vector systems[650,714,715]. In regard to the production of non-clinical human GBA in mammalian systems, early work by Grabowski et al. (1989)[711] made a considerable breakthrough by successfully producing GBA in COS-1 cells, derived from monkey kidney tissue[711]. Whilst the GBA produced was active against the

artificial substrate 4-MU-glucopyranoside, analysis of the oligosaccharide composition revealed that the post-translational glycosylation profile of GBA in COS-1 cells differs significantly from that found in human fibroblasts[711]. This work highlighted the importance of protein glycosylation and the impact of altered post-translational oligosaccharide processing in different expression hosts. Later, Febrega et al. (2002)[710] employed recombinant retroviruses encoding wild-type GBA and E235A and E340A mutant proteins to produce active and inactive GBA in murine cells[710]. Whilst no protein purification was performed, this work revealed that both mutants are catalytically inactive despite being correctly processed and sorted to the lysosome[710]. This suggested an important role for Glu235 in the catalytic mechanism, but not in the protein folding and processing, which supported previous computer-based predictions concerning the role of Glu235 as the putative general acid-base residue.

In light of the success of Velaglucerase alfa, a number of studies have investigated various plant systems for recombinant GBA production, including *Arabidopsis thaliana*[326], rice suspension culture[716] and *Nicotiana benthamiana*[712,717]. Indeed, plant-based expression platforms have many advantages for the production of pharmaceutical recombinant proteins, including low cultivation costs, scale-up potential, ability for complex protein production with post-translational modifications, and a low risk of contamination by human pathogens[718]. However, one of the major challenges of using plants for the production of human glycoproteins is preventing the biosynthesis of plant *N*-glycans, such as β-1,2-xylose and α-1,3-fucose, which can induce an immunogenic response in humans[719].

A number of strategies have been developed to reduce plant-specific *N*-glycan maturation. Of note, He et al. (2012)[326] reported a unique system for the production of active human GBA in seeds of the *Arabidopsis thaliana complex-glycan-deficient* (*cgl*) mutant[326]. This mutant is deficient in *N*-acetylglucosaminyl transferase I, which is responsible for adding *N*-acetylglucosamine (GlcNAc) to trimmed *N*-glycans[720]. Therefore, this mutant is unable to transfer GlcNAc to *N*-glycans which in turn prevents the addition of immunogenic β-1,2-xylose and α-1,3-fucose units. Specifically, this approach made use of *Agrobacterium*-mediated gene transfer to transform the *cgl* mutant seeds and permit expression of the human *GBA1* gene[326]. The resulting recombinant protein was found to accumulate in the apoplast and was subsequently purified in a 3-step procedure to reveal a variety of glycoforms. Importantly, this recombinant formulation exhibited similar kinetic parameters and thermal stability to the

commercial product Cerezyme®. However, some xylose and fucose containing glycans were detected which required further purification, and no information on post-purification yield was provided[326].

Soon after this work, an alternative approach to reduce the presence of plant specific *N*-glycans was reported by Limkul et al. (2016)[712]. In this work, RNA interference (RNAi) was used to down-regulate endogenous *N*-acetylglucosaminyl transferase I activity in *Nicotiana Benthamian* to generate a glycoengineered mutant. This glycoengineered *N. benthamiana* plant was then cross-pollinated with a GBA-expressing plant for the production of human GBA[712]. Using this approach, a considerable reduction in plant specific *N*-glycans was achieved, with the result that > 90% of the total *N*-glycans were of high mannose type. Although the resulting GBA formulation had a lower cellular activity compared to Cerezyme®, this plant derived formulation was taken up by macrophages during *in vivo* studies[712]. Most recently, Uthailik et al. (2021)[721] employed this glycoengineered N. *benthamiana* mutant to produce recombinant GBA with mannosidic-type *N*-glycans using *Agrobacterium*-mediated transient expression[721]. The resulting recombinant GBA was found to have comparable activity to mammalian derived GBA and was shown to exhibit mannosidic-type *N*-glycan structures without plant-specific *N*-glycans[721]. Moreover, compared to the stable expression reported by Limkul et al. (2016)[712], this transient expression system was reported to provide ~2.3-fold higher crude GBA activity, which was attributed to the fact that *Agrobacterium*-mediated transient expression is often more efficient than that of stable expression by gene integration[721]. Nevertheless, a purified GBA yield was not provided, and such a system remains poorly accessible to the wider research community. Consequently, a robust plant expressions system for GBA is yet to be established.

Given the involvement of GBA in GD and other pathologies, there is a pressing need for a reliable and affordable source of recombinant GBA to meet research demands. However, the diversity in non-clinical expression systems and the lack of information on protein purity and yield, demonstrates the current lack of consensus on a robust and economical platform for non-clinical GBA production. One platform with considerable potential is the insect-baculovirus expression system.

4.2.4 Baculoviral Expression Vector Systems (BEVS)

4.2.4.1 Baculoviruses

Baculoviruses, of the *Baculoviridae* family, are the most prominent viruses known to affect the insect population[722]. Specifically, baculoviruses are a diverse group of double-stranded, circular, super-coiled DNA viruses, typically 80-180 kb in size[723], encased in a rod-shaped capsid[724–726]. These are arthropod specific viruses, with no known homologues in any other organism including plants, fungi and bacteria[727]. Nevertheless, a high degree of diversity in the genome size, organisation and content has been observed[728–730].

Initial interest in baculoviruses was driven by the threat posed to the silk industry through inflicting various diseases in silkworms (*Bombyx mori*)[731,732]. In the mid 1800s, the presence of highly refractive, polyhedron shaped occlusion bodies were identified in disease affected insects using light microscopy[726]. However, it wasn't until the late 1940s that rod-shaped virus particles, called virions, were unambiguously demonstrated using electron microscopy[733]. Subsequently, two types of baculoviruses were identified: nucleopolyhedroviruses (NPVs) which produce polyhedral occlusion bodies in the nucleus, and granuloviruses (GVs) which form ovicylindrical (granular) occlusion bodies found in infected cells following rupture of the nuclear membrane[734,735]. In the case of NPVs, multiple virions are embedded in large occlusion bodies, whereas GVs produce smaller occlusion bodies typically containing a single virion[736,737]. Of note, NPVs are the causative agents of nuclear polyhedrosis disease in the silkworm, which remains incredibly detrimental to the silk production industry today[731,738].

4.2.4.2 Baculovirus Infection and Replication

An unusual behaviour of baculoviruses is their ability to produce two distinct types of enveloped virus during their life cycle[723,727]; occlusion-derived viruses (ODV) which are encapsulated in a crystalline protein matrix and are responsible for host-to-host transmission, and budded viruses (BV) which are single nucleocapsids released from infected cells and are responsible for systemic cell-to-cell transmission[723], Figure 4.1.

Typically, naturally occurring *in vivo* infection occurs when an insect feeds on a plant contaminated with occluded virus. The ingested virus is then transported to the mid-gut of the host where the polyhedrin protein matrix dissolves to release ODV[723,739]. In the early phase of

the baculoviral life cycle, ODV fuse to the epithelial cell membrane of the host intestine, penetrate the host cells through the endosomal system and are transported to the nucleus where the expression of early viral genes is initiated and host gene expression is terminated[735,740,741]. In the late phase, genes which code for viral DNA replication and virus assembly are expressed, resulting in the production of BV which egress from the nucleus to spread the infection systematically from cell-to-cell[740,741,721]. During budding from the cell, BV acquire an envelope which is decorated at one end with the glycoprotein gp64[742]. This glycoprotein is a homotrimeric protein that mediates fusion of the virus with the host endosomal membrane to permit cell entry[739,742,743]. In the very late phase of viral replication, the polyhedrin[744] and p10[745] genes, which control the expression of viral proteins, are expressed[723]. Lastly ODV are produced in the nucleus of infected cells by encapsulating multiple viral nucleocapsids in a large, polyhedrin matrix to form occlusion bodies. During cell lysis, these occlusion bodies are released into the environment to initiate a new infection cycle in a new host[723], Figure 4.1.

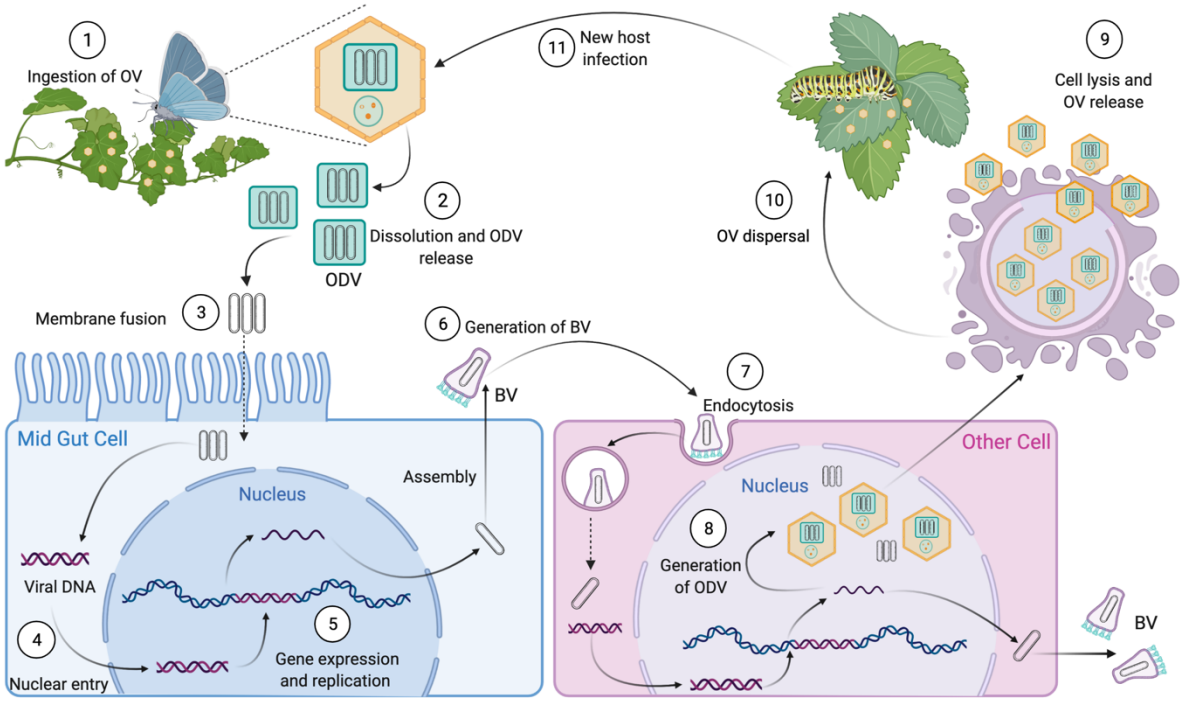


Figure 4.1: Infection and replication cycle of a generic NPV in Lepidoptera hosts. The primary infection is described by events 1-6 and the secondary infection is described by stages 7-11. OV = occluded virus, ODV = occlusion derived virus, BV = budded virus. Figure created in BioRender (www.app.biorender.com).

4.2.4.3 Applications of Baculoviruses

Baculoviruses exhibit a very narrow host range, typically limited to a few related insect species[730]. Owing to this high degree of host specificity, baculoviruses have proved beneficial in a number of applications. For example, baculoviruses have been employed as insecticides for controlling pests in agriculture and forestry[556]. In a similar vein, baculovirus genes that encode for enzymes which target the intestinal membrane of insects have been used to generate transgenic, insect-resistant plants[748]. Baculoviruses have also proved extremely useful lab tools, with several commercially available baculoviral vectors being employed in antiviral and cancer gene therapies[749], drug screening[750] and vaccine production[751,752]. Additionally, baculoviruses have played a pivotal role in the overexpression of eukaryotic proteins[753–756]. Indeed, the insect-baculovirus expression system was recently used to successfully produce four version of the SARS-CoV-2 (COVID-19) spike protein as potential vaccination targets[757], demonstrating the adaptability of BEVS to novel and highly desirable therapeutic targets.

4.2.4.4 Recombinant Baculoviruses for Protein Expression

Historically, eukaryotic proteins have proved incredibly difficult to recombinantly produce due to their large size and requirements for post-translational processing[755]. Of note, glycosylation and phosphorylation are vital to the stability and biological activity of most eukaryotic proteins, however, these processes are not typically supported by a prokaryotic host[755]. Considerable effort has been devoted to improving bacterial systems for eukaryotic expression through modification of *E.coli* host strains[758], however, the production of most eukaryotic proteins still requires a eukaryotic host; this is especially true for pharmaceutical proteins.

The use of baculovirus expression vectors was first described in the early 1980s and has since proved useful for the production of many recombinant eukaryotic proteins in insect cells[759–762]. The major difference between the naturally occurring *in vivo* infection and the *in vitro* infection for recombinant protein production, is that the naturally occurring polyhedrin gene within the baculovirus genome is replaced with a foreign gene of interest[763]. In the very late phase of *in vivo* viral replication, the polyhedrin gene is expressed to produce large amounts of polyhedrin protein which is required to package virions into occlusion bodies[726]. However, the polyhedrin protein is not required for recombinant *in vitro* infection because viral infection is propagated in cell culture by budded virions[764]. Therefore, the polyhedrin gene can be replaced

with a foreign gene of interest which is subsequently put under control of the strong polyhedrin promoter[755,762]. This promoter drives efficient, high-level expression of the recombinant protein which is then processed, modified and targeted to the appropriate cellular location[762]. Additionally, baculoviruses arrest most host gene transcription, thereby prioritizing viral gene expression[765].

The *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV), originally isolated from the alfafa looper (*A. californica*), was the first baculovirus to be fully sequenced[766]. In contrast to most baculoviruses, AcMNPV is able to infect around 30 species from the Lepidoptera order and has become the most studied baculovirus for recombinant production[763,767–769]. AcMNPV is a particularly advantageous baculoviral vector because it's genome contains multiple non-essential regions into which foreign DNA can be effectively inserted[766]. Additionally, its polyhedrin/p10 genes provide very strong promoters which direct transcription of recombinant genes[744,745,762]. In fact, the first reports of baculoviral systems for the production of recombinant proteins used AcMNPV derived vectors to produce human β-interferon[770] and *E.coli* β-galactosidase[771] in *Spodoptera frugiperda* (Sf9) cells. These studies marked the emergence of the baculovirus expression vector system (BEVS), which has advanced considerably in the last 30-40 years, as reviewed extensively by van Oers et al. (2015)[772] and more recently by Gorda et al (2021)[765].

4.2.4.5 Production of Recombinant Baculoviruses

Traditionally, recombinant baculovirus was generated in a two-step procedure, involving cloning the foreign gene into a transfer vector and then inserting this into the baculovirus genome by homologous recombination in insect cells[773]. This approach was limited by the low yield of recombinant virus and the requirement for multiple rounds of plaque purification[770]. However, in 1993, the generation of recombinant baculovirus was drastically simplified by the engineering of the baculoviral genome into an artificial bacterial chromosome known as a bacmid[765,774]. The bacmid can be maintained and replicated in bacterial cells (e.g. DH10β), allowing for the production and amplification of recombinant bacmids in *E. coli*[765]. In this approach, the gene of interest is cloned into a transfer vector which is then transformed into a DH10β *E.coli* strain containing the baculovirus shuttle vector to generate the recombinant bacmid[774]. Nowadays, there are numerous commercial kits available to facilitate recombinant bacmid production in bacterial cells, including the Bac-to- Bac® Baculovirus Expression System

(Thermo Fisher Scientific), flashBAC (Oxford Expression Technologies), the BaculoDirect™ Baculovirus Expression System (Thermo Fisher Scientific) and BaculoGold™ (BD Biosciences)[765,775]. Of note, the Bac-to-Bac® system uses the Tn7 transposition method which involves transferring the foreign gene of interest from a mini-Tn7 transposase site in the transfer plasmid to the Tn7 transposase site in the baculovirus shuttle vector using a transposase enzyme[765]. More recently, BEVS has seen the introduction of the MultiBac system[755], which allows for the synthesis of multi-subunit protein complexes using a single baculoviral vector[765].

4.2.4.6 Insect-Baculoviral Expression Systems

Since the isolation of the first insect cell line in the early 1960s[776], over 500 insect cell lines have become available[775,777]. Nowadays, the Lepidopteran cell lines derived from *Spodoptera frugiperda* (Sf21 and Sf9) and *Trichoplusia ni* (High Five, BTI-Tn-5BI-4) are commonplace in BEVS[754,756,778,779], however, Diptera cell lines from *Drosophila melanogaster* (S2) are becoming increasingly popular[775]. Other less common host cell lines include Bm5 derived from *Bombyx mori*, Tn368 obtained from *Trichoplusia ni* and Ea88 isolated from *Estigmene acrea*; however, the utility of these lines has been limited by their poor growth in suspension culture[780].

One advantage of insect cells is their ability perform post-translational modifications[781], namely glycosylation and phosphorylation, which is vital to the stability and activity of many human glycoproteins[782]. However, it should be noted that insect cells generally exhibit less complex *N*-glycosylation profiles than mammals[775,783], Figure 4.2. Insect *N*-glycans are typically mannose terminated, whilst mammalian *N*-glycans exhibit terminal sialic acid residues and show more antennal diversity[775,784]. Additionally, insect *N*-glycans are decorated with core α-1,3-fucose units which can be immunogenic to humans[784,785]. Several strategies have been developed to address the issues of incompatible *N*-glycosylation in insect cell expression systems[775], and such approaches have led to the creation of insect cell lines that express genes for the enzymes required to produce mammalian glycosylation patterns[786,787]. Nevertheless, these systems remain in their infancy and most proteins produced in BEVS are limited to preclinical research[783]. However, the high levels of recombinant gene expression, ease of maintenance and scale up[759,788], ability to perform post-translational modifications and inherent safety[755], means the insect-baculovirus expression system has become one of the most versatile systems for recombinant protein production[765,780].

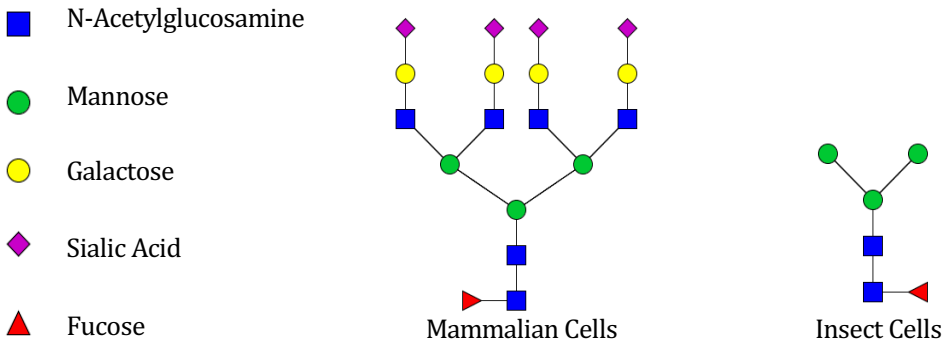


Figure 4.2: Mammalian and insect cell *N*-glycan profiles demonstrating the truncated nature of insect cell glycosylation with core α-1,3-fucosylation. *N*-glycans constructed in GlycanBuilder[789] (https://sugarbind.expasy.org/builder) using Glycoblock format[591].

4.2.5 Baculoviral Expression Systems for GBA

4.2.5.1 Insect-Baculoviral Expression Systems for GBA

Previous studies on the use of BEVS platforms for GBA production have shown some success, albeit with inconsistent results regarding protein quantity and quality. Early work by Martin et al. (1988)[714] using the AcMNPV vector demonstrated successful GBA expression in Sf9 cells, with 40% of the recombinant product being secreted into the medium. The protein was only partially purified and just 15-40% of the original protein was recovered. Nevertheless, this study demonstrated that human GBA can be produced in BEVS, and this was later supported by another investigation in which secretion of active GBA from Sf9 cells was achieved[790]. Conversely, other studies reported minimal secretion of GBA from S*f*9 cells using the same A*c*MNPV vector and promoter[648,711] and subsequent investigations demonstrated that recombinantly produced GBA can be stored intracellularly rather than being secreted[791]. However, no purification or biochemical analysis of the recombinant protein from cell homogenates was performed.

More recently, Sinclair et al. (2006)[650] employed the *Orgyia pseudotsugata* multi-capsid nucleopolyhedrovirus (OpNPV) for GBA production in Sf9 cells to investigate the effect of the full-length and shortened native signal sequences on GBA secretion. The full length signalling construct was reported to produce 30% more enzymatic activity than the shortened construct but both resulted in secretion of GBA into the media[650]. In the same year, wild-type GBA and the N370S GD causing mutant were expressed in High Five (Hi5) cells by Sawkar et al. (2006)[715], allowing the stability, pH sensitivity, cellular localisation and chaperone mediated stabilisation

of the mutant to be studied[659]. Despite this success, little information on the expression system was provided (likely owing to the fact this work was done in collaboration with Sanofi Genzyme) and no information on protein yield were reported. Therefore, in the absence of a reliable expression platform for non-clinical GBA, there is considerable ongoing reliance on expired ERT formulations for biochemical, mechanistic and structural studies. Indeed, prior to this work, the Davies group was dependent on expired Cerezyme® samples for structural studies. Unfortunately, these ERT formulations are incredibly costly and often only obtainable in limited supply under a Material Transfer Agreement (MTA).

4.2.6 Research Aims

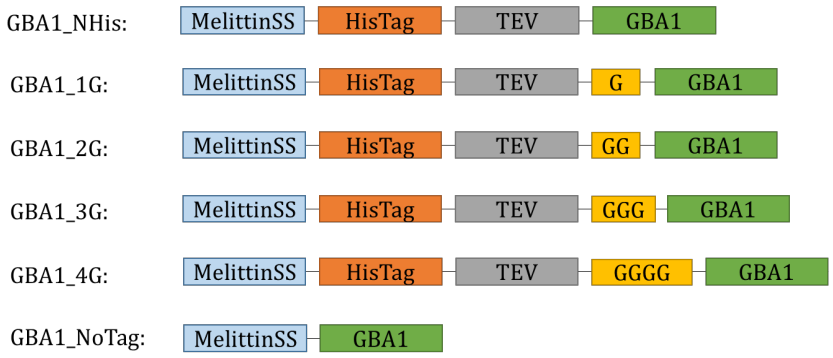
To support our long-standing interest in the development of inhibitors and activity-based probes for GBA, this work aimed to establish a reliable insect-baculoviral expression vector system (BEVS) for the production of recombinant human GBA. Specifically, this work sought to utilise a Bac-to-Bac® approach to develop an AcMNPV-derived baculoviral expression system using *Trichoplusia ni* (High Five, Hi5) cells as the expression host. This work further aimed to generate recombinant GBA suitable for crystallography studies to permit structural analyses with novel GBA active compounds.

4.3 Materials and Methods

4.3.1 *GBA1* Constructs

Several *GBA1* constructs were produced to find a construct suitable for effective GBA production, purification and crystallisation. Initially, an *N*-terminally His_6-tagged construct was generated, followed by several constructs containing successively long glycine linkers to make the TEV cleavage site more accessible. A non-tagged construct was also generated, Figure 4.3. Details of the required primers and backbones can be found in Appendix 1.

(a)



(b)

Construct	GBA1 Source Plasmid	Forward Primer	Reverse Primer	Backbone
GBA1_NHis	pGEn1-GBA	P262	P387	B24
GBA1_1G	pOMNI-GBA1_NHis	F1G	P387	B24
GBA1_2G	pOMNI-GBA1_NHis	F2G	P387	B24
GBA1_3G	pOMNI-GBA1_NHis	F3G	P387	B24
GBA1_4G	pOMNI-GBA1_NHis	F4G	P387	B24
GBA1_NoTag	pOMNI-GBA1_NHis	FCHis	P387	B41

Figure 4.3: (a) *GBA1* constructs produced in this study, where MelittinSS represents the honeybee melittin signal sequence, HisTag depicts a hex-histidine affinity tag, TEV indicates a Tobacco Etch Virus cleavage site, G*BA1* represents the *N*-terminally truncated *GBA1* gene and G indicates a glycine linker. (b) Table of backbones, primers and source plasmids required for each construct. Details of these can be found in Appendix 1.

4.3.2 Generation of Recombinant Transfer Plasmids Encoding *GBA1*

4.3.2.1 Obtaining the GBA1 Gene

The *N*-terminally truncated *GBA1* gene was initially subcloned from the pGEn1-GBA plasmid (DNASU Clone ID: HsCD00413213[792]), Figure 4.4, obtained from the Glycoenzyme repository

(http://glycoenzymes.ccrc.uga.edu/) using Phusion® (New England Biolabs) polymerase chain reaction (PCR) with forward primer P262 and reverse primer P387, Figure 4.3 (b), to introduce a His_6-tag at the *N*-terminus. The pGEn1-GBA template was mixed with Phusion HF buffer (1x), d-NTPs (0.2 mM), P262 (0.5 µM), P387 (0.5 µM) and Phusion DNA polymerase enzyme (2U) in a 100 µL reaction. The thermocycling conditions used for Phusion PCR are summarised in Table 4.1. Successful amplification was confirmed by running the PCR product on an agarose gel (1%) with SYBR Safe DNA gel stain (0.1x). The insert was gel extracted using the QIAquick Gel Extraction Kit (50) according to the manufacturers protocol and analysed by sanger sequencing to confirm the presence of the human *GBA1* gene, HisTag and TEV cleavage site. This insert was used to generate the GBA1_NHis construct only.

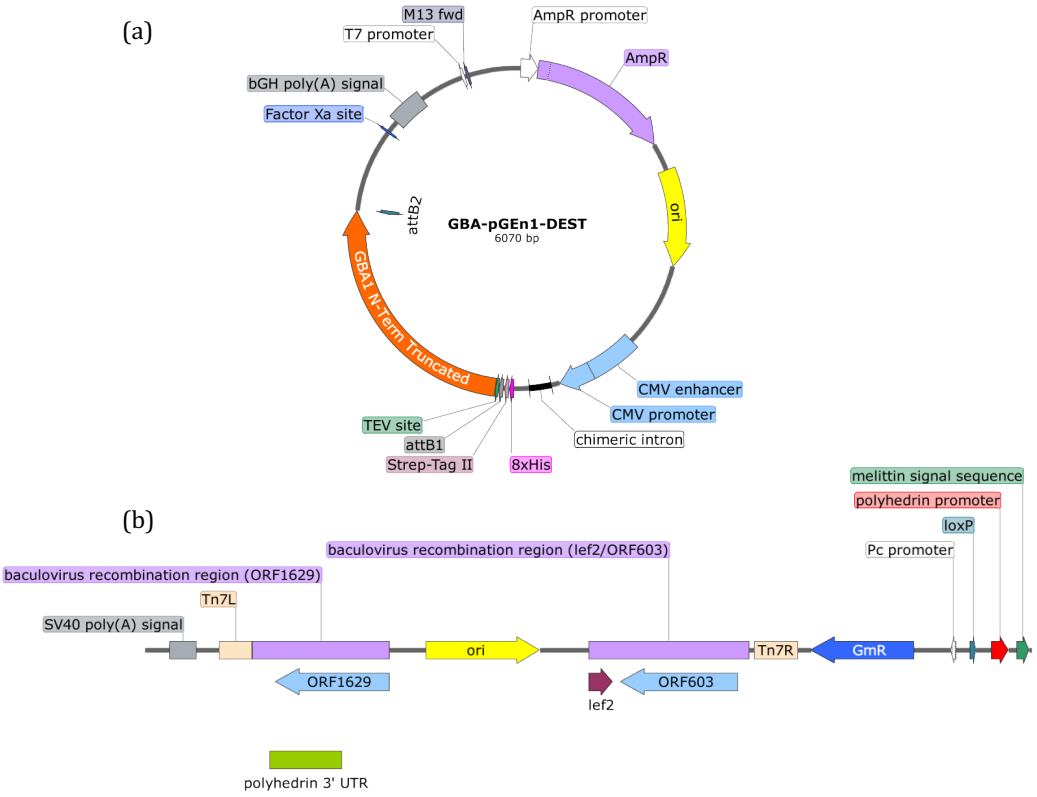


Figure 4.4: (a) GBA-pGEn1-DEST plasmid (DNASU Clone ID: HsCD00413213[792]), containing the N-term truncated *GBA1* gene. (b) p-OMNI B24 backbone required for SLIC Maps generated in SnapGene Viewer.

Table 4.1: Thermocycling conditions employed for Phusion PCR

Stage	Temperature / °C	Duration / s	Number of cycles
Initial Denaturation	98	30	1
Denaturation	98	10	32
Annealing	68.7	45	32
Extension	72	420	32
Final Extension	72	600	1

4.3.2.2 *Generating GBA1_NHis Transfer Plasmid by Sequence and Ligation Independent Cloning (SLIC)*

The *GBA1* insert obtained from the pGEN1-GBA plasmid (1 ng μL^{-1}) was treated with T4 DNA polymerase (0.5 U) in the presence of BSA (0.1 mg mL^{-1}) for 30 minutes at room temperature. The reaction was halted by adding d-cNTP (1 mM). The linearised B24 pOMNI backbone containing the honeybee melittin signal sequence, Figure 4.4 (b), was obtained by restriction digest of an existing pOMNI plasmid using HindIII-HF and XmaI (New England Biolabs). The original pOMNI vector, containing the Tn7 transposon sequences (Tn7L and Tn7R), was kindly provided to the York Structural Biology Laboratory by the Berger Lab (University of Bristol). The linearised pOMNI backbone was subsequently treated with T4 DNA polymerase as described above. All DNA fragments were analysed on an agarose gel (1%) and purified by gel extraction using the QIAquick Gel Extraction Kit (Qiagen).

The insert (1.0 ng μL^{-1}) and backbone (1.5 ng μL^{-1}) were subsequently treated together with RecA (2 ng μL^{-1}), in the presence of RecA buffer (1x) and ATP (1 mM), for 1 hour at 37 °C. The RecA treated *GBA1* insert and B24 backbone were incubated with DH5α *E. coli* cells on ice for 30 minutes before transformation by heat shock at 42 °C for 45 seconds. The cells were incubated at 37 °C for 1-hour, Super Optimal Broth medium supplemented with 20 mM glucose (SOC medium) (300 µL) was added and the cells were plated on Luria–Bertani broth (LB) agar plates containing gentamicin (15 µg ml^{-1}). The plates were incubated overnight at 37 °C. Individual colonies were re-streaked and incubated overnight at 37 °C. An overnight 5 mL LB culture containing gentamicin (15 µg ml^{-1}) was grown and the plasmid DNA was extracted and purified using a QIAprep Spin Miniprep Kit (Qiagen, 250) according to the manufacturers protocol.

4.3.2.3 *Verification GBA1_NHis Transfer plasmid*

To confirm the desired transfer plasmid had been formed, colony PCR was performed on each re-streaked colony using forward primer 5'-CAGCAGCGAAGTCGCCATAAC-3' (P51) and reverse primer 5'-CAGCCGGATCTTCTAGGCTC-3' (P52) to amplify the *GBA1* gene. The re-streaked colonies were transferred to 100 µL of ultra-pure water and boiled at 95 °C for 5 minutes. A 20 µL PCR reaction was prepared for each colony containing P51 (0.5 µM), P52 (0.5 µM), d-NTPs (0.2 mM), Phusion HF buffer (1x), Phusion enzyme (2 U) and colony DNA (1 µL). The reactions

were subjected to the thermocycling conditions outlined in Table 4.1. The PCR products were analysed on an agarose gel (1%) with SYBR Safe DNA gel stain (0.1x).

To further verify the transfer plasmid, HindIII restriction digest was performed. According to computer simulated digest using ApE plasmid editor, treatment with HindIII should yield two fragments; one of ~1.6 kb containing the full *GBA1* gene and a second fragment of ~6.1 kb containing the rest of the plasmid. The purified GBA1_NHis plasmid DNA (~1 µg) was treated with HindIII-HF (10 U) and CutSmart buffer (1x) for 2 hours at 37 °C. The digested material was analysed on an agarose gel (1%) with SYBR Safe DNA gel stain (0.1x) to confirm the correct construct was formed. Additionally, the purified GBA1_NHis plasmid was diluted to 72 ng μL^{-1} in ultrapure water (30 µL) and confirmed by Sanger sequencing using P51 and P52.

4.3.2.4 Generating GBA1_1G, GBA1_2G, GBA1_3G, GBA1_4G and GBA1_NoTag Transfer Plasmid by SLIC

To obtain the desired *GBA1* insert for the remaining constructs, the *GBA1* gene was copied and amplified from the pOMNI-GBA1-NHis transfer plasmid by Phusion® PCR using the primers listed in Figure 4.3 (b) to introduce the desired modification. The pOMNI-GBA1_NHis template was mixed with Phusion HF buffer (1x), d-NTPs (0.2 mM), Phusion DNA polymerase enzyme (2U) and the appropriate forward and reverse primers (0.5 µM) in a 100 µL reaction. The thermocycling conditions used for Phusion PCR are summarised in Table 4.1. Successful amplification was confirmed by analysing the PCR products on an agarose gel (1%) with SYBRSafe DNA gel stain (0.1x). The *GBA1* inserts were gel extracted using the QIAquick Gel Extraction Kit (50) according to the manufacturers protocol.

The *GBA1* inserts (1 ng μL^{-1}) were treated with T4 DNA polymerase (0.5 U) as described previously and diluted with water to 5 ng μL^{-1}. The relevant backbones (5 ng μL^{-1}), Figure 4.3 (b), were supplied by Liang Wu (University of York) pre-treated with T4 DNA polymerase. The *GBA1* inserts (1.0 ng μL^{-1}) and relevant backbones (1.5 ng μL^{-1}) were subsequently treated together with RecA (2 ng μL^{-1}) as described previously. Following RecA treatment, the backbones and inserts were incubated with ONESHOT Top10 *E. coli* cells on ice for 20 minutes before transformation by heat shock at 42 °C for 45 seconds. The cells were incubated at 37 °C for 1 hour, SOC medium (300 µL) was added and the cells were plated on LB Agar plates with gentamicin (15 µL mL^{-1}). The plates were incubated overnight at 37 °C and individual colonies

were re-streaked before preparing overnight 5 mL LB cultures supplemented with gentamicin (15 μg mL^{-1}). The resulting transfer plasmids were extracted and purified from the overnight cultures using a QIAprep Spin Miniprep Kit (Qiagen, 250) according to the manufacturers protocol. These transfer plasmids were verified by colony PCR, HindIII digest and Sanger sequence as outlined for the GBA1_NHis construct.

4.3.3 Generation of Recombinant DH10EMBacY Bacmids Encoding *GBA1*

The DH10EMBacY *E. coli* strain was generously provided by the Berger Lab (University of Bristol). The DH10EMBacY contains the EMBacY baculovirus shuttle vector (bacmid bMON14272) with a mini-attTn7 target site, a tetracycline resistant helper plasmid (pMON7124) encoding the transposase enzyme, a yellow fluorescent protein (YFP) reporter gene, the LacZα gene and a kanamycin resistance selection marker.

Recombinant bacmid for each construct was produced using the Tn7 transposition method in DH10EMBacY[754,756] (Geneva Biotech). Briefly, purified transfer plasmid DNA was transformed into DH10EMBacY cells by electroporation at 1.8 kV. SOC medium was added immediately and the cells were incubated for 4 hours at 37 °C before blue/white screening on LB agar plates containing kanamycin (50 μg mL^{-1}), gentamicin (15 μg mL^{-1}), tetracycline (15 μg mL^{-1}), IPTG (1 mM) and x-Gal (1x). The plates were incubated for 2 days at 37 °C until blue and white colonies could be identified. Successful white colonies were re-streaked and confirmed by Phusion® colony PCR. Colony PCR reactions were prepared containing forward primer 5'-CCCAGTCACGACGTTGTAAAACG-3' (P55) (0.5 μM), reverse primer 5'-AGCGGATAACAATTTCACACAGG-3' (P56) (0.5 μM), d-NTPs (0.2 mM), Phusion HF buffer (1x), DMSO (3 %), Phusion enzyme (2 U) and boiled colony DNA (1 μL). The reactions were subjected to the same thermocycling conditions outlined in Table 4.1. The PCR products were analysed on an agarose gel (1%) with SYBR Safe DNA gel stain (0.1x). Successful colonies were grown in 10 mL LB overnight cultures containing kanamycin (50 μg mL^{-1}), gentamicin (15 μg mL^{-1}) and tetracycline (15 μg mL^{-1}). The recombinant bacmids were isolated and purified from the LB cultures using the Pure Link™ HiPure Plasmid DNA Purification Kit (Invitrogen) and verified by Phusion® PCR using forward primer P51 and P52 to amplify the GBA gene, and P55 and P56 to amplify across the Tn7 insertion site.

4.3.4 Generation of Recombinant Baculovirus Encoding *GBA1*

4.3.4.1 General Notes

All tissues culture experiments were conducted in a laminar flow hood with an air flow of ~ 0.75 m s^{-1}. The flow hood was irradiated with UV light for 10 minutes before and after use. The hood was wiped with 70% ethanol and ChemGene™ disinfectant before and after use. Incubation of cell cultures was performed at 28 °C and 87 rpm.

4.3.4.2 Transfection of Sf9 cells and Production of Recombinant Baculovirus

Recombinant baculovirus was generated and amplified in S*f*9 cells (clonal isolate of *Spodoptera frugiperda* S*f*21 cells (IPLB-Sf21-AE)) purchased from Invitrogen. Adherent S*f*9 cells were grown at 28 °C for 2 days in 60 mL Insect-XPRESS™ protein free media (Lonza Bioscience) supplemented with 2% fetal bovine serum (FBS). At log phase growth, 2 mL of suspension S*f*9 cells were seeded into each well of a 6-well tissue culture plate at a density of 0.45 × 10^6 cells mL^{-1} and allowed to settle for 10 minutes in a humidified incubator at 28 °C. 180 µL of a transfection mix containing Insect-XPRESS media (1.05 mL), recombinant bacmid DNA (~ 100 µg) and FuGENE HD (Promega)[793] transfection agent (31.5 µL) was added dropwise to each well and the cells were incubated in a static humidified incubator at 28 °C until ~ 95 % baculoviral transduction was achieved, as indicated by expression of the EMBacY YFP marker gene. The supernatant was collected by centrifugation at 200 g for 5 minutes and FBS (0.2 mL) was added to yield the viral P1 stock. A 50 mL culture of S*f*9 cells was prepared at 1 × 10^6 cells mL^{-1} in Insect-XPRESS media and infected with 1 mL of viral P1 stock. The culture was incubated at 28 °C with shaking at 87 rpm until 95 % transfection was achieved. The supernatant was collected by centrifugation at 200 g for 5 minutes and FBS (1 mL) was added to yield the viral P2 stock.

4.3.5 Expression of Recombinant Human GBA in High Five Cells

4.3.5.1 Test Expression

An adherent High Five™ cell line (Hi5, BTI-Tn-5B1-4, Invitrogen) was prepared in a 60 mL culture in Express Five™ Serum Free Media supplemented with 20 mM L-Glutamine (Thermo Fischer Life Technologies Ltd). The culture was incubated at 28 °C and 87 rpm for ~24 hours. Once critical cell density (> 2 × 10^6 cells mL^{-1}) was reached, the culture was split back and

prepared at ~1-2 × 10^6 cells mL^{-1} in 50 mL. Baculovirus P2 stock (500 µL, prepared in *section 4.3.4)* encoding human *GBA1* was added and the cultures were incubated at 28 °C and 87 rpm until YFP fluorescence was observed in 95% of the cells (~2-3 days). The supernatant was harvested by centrifugation at 400 g for 15 min at 4°C. DTT and PMSF were added to achieve a final concentration of 1 mM and 0.1 mM respectively.

4.3.5.2 Full Scale Expression

An adherent High Five™ cell line (BTI-Tn-5B1-4, Invitrogen) was prepared in a 60 mL culture in Express Five™ Serum Free Media supplemented with 20 mM L-Glutamine (Thermo Fischer Life Technologies Ltd). The culture was incubated at 28 °C and 87 rpm for ~24 hours. Once critical cell density (> 2 × 10^6 cells mL^{-1}) was reached, the culture was successively passaged to 100 mL, 600 mL, 1.8 L and 3.6 L in Express Five™ Serum Free Media. The 3.6 L culture (~1-2 × 10^6 cells mL^{-1}) was prepared in 6 × 600 mL cultures and infected with 750 µL of baculovirus P2 stock (prepared in *section 4.3.4*). The cultures were incubated at 28 °C and 87 rpm until 95% transfection was achieved, as indicated by YFP fluorescence. The supernatant was harvested by centrifugation at 200 g for 15 min at 4°C, followed by further clearing of debris by centrifugation at 4000 g for 60 min at 4°C. DTT and PMSF were added to a final concentration of 1 mM and 0.1 mM respectively.

4.3.6 Purification of Recombinant Human GBA

4.3.6.1 Purification of Tagged Constructs

Immobilised Metal Affinity Chromatography (HisTrap Purification): For all his_6-tagged constructs, initial purification was performed using a 5 mL HisTrap Excel HP column (GE Healthcare). The HisTrap column was equilibrated with high imidazole buffer B (20 mM Tris, 500 mM NaCl 1 M imidazole pH 8.0, 0.05% N-dodecyl-β-D-maltoside (DDM)) followed by equilibration with low imidazole buffer A (20 mM Tris, 500 mM NaCl, 20 mM imidazole pH 8.0, 0.05% DDM) before the cell culture medium was loaded. The column was eluted in a buffer gradient from 0% to 100% buffer B over 20 CVs. The column was then run in 100% buffer B for a further 5 CVs. The eluate was collected as 1.6 mL fractions and all fractions with an A_{280} response were analysed by SDS-PAGE. Fractions which appeared to contain the GBA protein (~60 kDa) were pooled and subjected to further purification.

Cation Exchange Chromatography (CatIEX): Fractions pooled from HisTrap purification were diluted 15-fold in buffer A (20 mM MES, pH 6.0) to lower the salt concentration and adjust the pH to blow the estimated PI of the protein (PI = 6.9). The diluted fractions were purified through cation exchange chromatography using a 1 mL HiTrap SP (sulfopropyl) HP cation exchange column (GE Healthcare) or with a higher resolution 1 mL RESOURCE S cation exchange column. The column was washed with high salt buffer B (20 mM MES, 1.5 M NaCl, 0.05% DDM, pH 6.0) and equilibrated in low salt buffer A (20 mM MES, 0.05% DDM, pH 6.0). The sample was loaded onto the column and eluted over a linear gradient from 0% to 100% buffer B over 20-40 CVs. The column was then run in 100% buffer B for a further 5 CVs. The eluate was collected in 1.6 mL and all fractions with an A_{280} response were analysed by SDS-PAGE. Fractions containing GBA protein were pooled and subjected to further purification.

Size exclusion Chromatography (SEC): Fractions containing GBA were pooled and concentrated to 2 mL using a 30 kDa molecular weight cut-off centrifugal filter (VivaSpin™ 20). A S75 or S200 Superdex size exclusion column (GE Healthcare) was equilibrated in SEC buffer 20 mM MES, 200 mM NaCl, 0.05% DDM pH 6.5) and the sample was manually injected. The sample was eluted in SEC buffer for 1.5 CVs and collected in 1.6 mL fractions in a 96 well plate. All fraction with an A_{280} response were analysed by SDS-PAGE. Fractions containing GBA were pooled and buffer exchanged by a second round of SEC using a buffer without DDM detergent (20 mM MES, 200 mM NaCl, pH 6.5). Following buffer exchange, the pooled fraction containing GBA were concentrated to ~10 mg mL^{-1} using a 30 kDa Vivaspin concentrator. Typical yields of 3-4 mg (0.83-1.1 mg L^{-1} of culture media).

TEV Protease Treatment: For the His-tagged proteins, the protein fractions pooled from the first round of HisTrap purification were treated with AcTEV protease (1000 U μL^{-1}) in the presence of DTT (1 mM). A negative control without AcTEV and a positive control containing 10x AcTEV were prepared. The reaction and controls were incubated over night at room temperature and were analysed by SDS-PAGE

4.3.6.2 Purification of Non-tagged Construct

In the case of the non-tagged protein, the conditioned supernatant (~ 3.6 L) was concentrated to ~1 L using a KrosFlo® Research II*i* Tangential Flow Filtration (TFF) System with a 30 kDa mPES hollow fibre filter module prior to purification. The tagless protein was purified using a

procedure previously described by Sawkar et al. (2006)[715], with the addition of a size-exclusion step to yield pure protein suitable for crystallisation.

Hydrophobic Interaction Chromatography: rGBA was extracted from the concentrated media by hydrophobic interaction chromatography using a TOYOPEARL Butyl-650C column (Tosoh Bioscience). The column was pre-equilibrated with 1.5 CV Buffer A (20 mM NaOAc, 150 mM NaCl, pH 5.0) before loading of the cell culture media. The protein was isocratically eluted into 100% Buffer B (20 mM NaOAc, 150 mM NaCl, 50% (v/v) ethylene glycol, pH 5.0) over 5 CVs and collected as 1.6 mL fractions. All fractions with an A_{280} response were analysed by SDS-PAGE. Fractions containing GBA protein were pooled.

Cation Exchange Chromatography (CatIEX): The fractions pooled from HIC were diluted 3-fold in deionised water and purified by cation-exchange chromatography using a HiTrap Heparin Sepharose FF column (GE Healthcare). The column was pre-equilibrated in Buffer A (20 mM NaOAc, 50 mM NaCl, 20% (v/v) ethylene glycol, pH 5.0) before the sample was loaded. The protein was eluted with a linear gradient from 0-100% Buffer B (20 mM NaOAc, 1 M NaCl, 20% (v/v) ethylene glycol, pH 5.0) over 20 CV. All fractions with an A_{280} response were analysed by SDS-PAGE. Fractions containing GBA protein were pooled and diluted 15-fold in 20% ethylene glycol and further purified by weak cation-exchange with a HiTrap CM Sepharose FF column (GE Healthcare). The column was pre-equilibrated with Buffer A (30 mM Na Citrate, 0.01% Tween-80, pH 5.7) before the sample was loaded. The protein was eluted in a linear gradient over 20 CV into 100% Buffer B (55 mM Na Citrate, 0.01% Tween-80, pH 6.3) and collected in 1.6 mL fractions. All fractions with an A_{280} response were analysed by SDS-PAGE.

Size exclusion Chromatography (SEC): GBA containing fractions were pooled, concentrated to ~1.5 mL using a 30 kDa Vivaspin concentrator (GE Healthcare) and purified using a Superdex S200 16/600 column (GE Healthcare) in SEC buffer (10 mM MES, 100 mM NaCl, 1 mM TCEP, pH 6.5). The protein was eluted over 1.5 CVs and collected as 1.6 mL fractions. All fractions with an A_{280} response were analysed by SDS-PAGE and fractions containing GBA protein were pooled and concentrated to ~10 mg L^{-1} using a 30 kDa Vivaspin concentrator. Typical yields were 13-16.7 mg per preparation (3.6-4.6 mg L^{-1} of culture media).

4.3.7 Expression and Purification of Recombinant Human Saposin C

The gene encoding human Saposin C (SapC), with an additional methionine and aspartate at the N-terminal, was codon optimised for *E. coli* and purchased from Genscript subcloned into the NcoI and BamHI sites of the pET-16b vector. The resulting vector was transformed into *E. coli* Origami B (DE3) cells (Novagen) by heatshock. Cultures were grown at 37 °C in Luria-Bertani (LD) media supplemented with Ampicillin (100 μg mL^{-1}) to an OD_{600} of 0.8-1.0 before induction with 0.8 mM IPTG. The cultures were grown for a further 4 hours at 37 °C after which the cells were harvested by centrifugation (4000 g for 20mins).

The cell pellet was resuspended in anion-exchange buffer (25 mM NaCl, 25 mM Tris-HCl, pH 7.5) and lysed by sonication. The lysate was clarified by centrifugation at 16,000g for 10 min and the supernatant was heated at 85°C for ~30 min. Precipitated proteins were removed by centrifuged at 16,000g for 10 min. The resulting supernatant was applied directly to a Q-Sepharose column (GE Healthcare) that had been pre-equilibrated in binding buffer (25 mM NaCl, 25 mM Tris-HCl at pH 7.5). The protein was eluted with a linear gradient of 0-100% elution buffer (1 M NaCl, 25 mM Tris-HCl at pH 7.5) over 20 CVs followed by a further 5 CVs at 100% elution buffer. The peak fractions containing SapC were pooled, concentrated, and applied to a Superdex S75 16/600 (column GE Healthcare) in 50 mM Tris, 150 mM NaCl, pH 7.4 buffer. Fractions containing SapC were pooled and concentrated to ~9 mg mL^{-1} (protein concentration determined by Bradford assay). Protein identity was confirmed by LC-MS/MS. Summary of protein purification and identification can be found in Appendix 3.

4.3.8 Biochemical Characterisation of Recombinant GBA

4.3.8.1 Activity-Based Labelling

To confirm the expression of GBA and production of active recombinant protein, labelling assays were performed with a Cy5 labelled cyclophellitol-aziridine activity-based probe (JJB376, supplied by the Overkeeft lab, Leiden University). The supernatant (5 μL) and re-suspended cells (5 μL) were treated with JJB367 (1 μL, 10 μM) at pH 4.5 and pH 7.4 using appropriately adjusted McIlvaine buffers (150 mM disodium hydrogen phosphate, 50 mM citric acid) in a 10 μL reaction. The reactions were incubated at 37 °C for 30 minutes with shaking. The labelling reaction was halted by boiling at 98 °C for 5 minutes with loading dye (60 mM

Tris HCl pH 6.8, 10 % (v/v) glycerol, 2% (v/v) SDS, 0.05 % bromophenol blue and 5% (v/v) β-mercaptoethanol). The denatured samples were separated by electrophoresis on a 10% SDS-PAGE gel by running continuously at 200 V for 50 minutes. The wet slab gel was scanned on fluorescence using an Amersham Typhoon 5 at λ_{ex} 635 nm and λ_{em} > 665 nm to visualise fluorescently labelled GBA. PageRuler Plus prestained protein ladder was used as a marker. Full quantification of protein labelling was not performed.

4.3.8.2 Enzyme Kinetics

Michaelis-Menten kinetics were assayed with the fluorogenic substrate 4-methylumbelliferyl β-D-glucopyranoside (4-MU-Glc). GBA was prepared at 20 nM in kinetics buffer (McIlvaine buffer: 150 mM disodium hydrogen phosphate, citric acid [pH 5.2], supplemented with 0.2% (v/v) Taurocholate, 0.1% (v/v) Triton X-100 and 0.1% (v/v) bovine serum albumin (BSA)). 4-MU-Glc was prepared at 5mM in kinetics buffer and diluted 2-fold to yield solutions at 2.5, 1.25, 0.625, 0.313, 0.156, 0.078 and 0.039 mM. Each substrate solution (25 µL) was added to the wells of a black 384 well polystyrene plate in quadruplicate. GBA (25 µL, 20 nM) was added to each well to give a final enzyme concentration of 10 nM. Activity against 4-MU-Glc was monitored continuously over 5-minutes at 37 °C or room temperature by measuring the fluorescence of liberated 4-MU (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) using a CLARIOstar® *Plus* microplate reader (BMG LabTech). A linear calibration was generated by measuring the fluorescence of the 4-MU product (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) prepared at serial dilutions in kinetics buffer. Each 4-MU concentration was measured in quadruplicate. Using the 4-MU calibration, the average initial rate of substrate hydrolysis (V) was determined at each substrate concentration. The initial rates (V) were plotted against substrate concentration [S] and fitted by nonlinear regression to the Michaelis-Menten equation ($\upsilon = V_{max}[S] / (K_M + [S])$) to generate values of K_M, V_{max} and k_{cat} using the relationship $k_{cat} = V_{max}/[Enz]$. All data were processed in Origin graphing software.

4.3.8.3 GBA Activation by Saposin C

GBA was prepared at 30 nM in 100 mM NaOAc (pH 5.2) buffer supplemented with 0.1% (v/v) Triton X-100 and 0.1% (w/v) BSA. Additionally, SapC was prepared in serial dilutions from 132 µM – 64 nM in 100 mM NaOAc (pH 5.2) buffer supplemented with 0.1% (v/v) Triton X-100 and 0.1% (w/v) BSA. In a black 384 well plate, GBA (30 nM) and SapC (132 µM–64 nM) were mixed

in a 40 µL reaction volume in a ratio of 1:1 and incubated for 15 minutes at room temperature. To initiate the reaction, 4-methylumbelliferyl β-D-glucopyranoside (prepared at 15 mM) was added to a final concentration of 5 mM. The reactions were incubated at 37 °C for 15 minutes and stopped by the addition of 10 µL Na_2CO_3. The extent of substrate hydrolysis was determined by measuring the fluorescence of liberated 4-MU (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) using a CLARIOstar® *Plus* microplate reader (BMG LabTech). A linear calibration was generated by measuring the fluorescence of the 4-MU product (λ_{ex} 360/20 nm, λ_{em} 450/30 nm) prepared at serial dilutions in 100 mM NaOAc (pH 5.2) (buffer supplemented with 0.1% (v/v) Triton X-100 and 0.1% (w/v) BSA) + 10 µL Na_2CO_3. Using the 4-MU calibration, the amount of liberated 4MU at each SapC concentration was determined and fitted to a 4-parameter logistic function.

4.3.8.4 IC_{50} of Bis-Tris-Propane

The IC_{50} of bis-Tris propane (BTP) was determined using 4-MU-Glc. GBA was prepared at 58 nM in kinetics buffer [McIlvaine buffer; 150 mM disodium hydrogen phosphate, citric acid pH 5.2 supplemented with 0.2% (v/v) taurocholate, 0.1% (v/v) Triton X-100 and 0.1% (v/v) bovine serum albumin (BSA)]. BTP was prepared at 500 mM in kinetics buffer (pH 7.0) and diluted two-fold in serial dilutions to 488 nM. The BTP solutions (20 µl) were added to the wells of a black 384-well polystyrene plate followed by the addition of 4-MU-Glc (20 µl) prepared at 4 mM in kinetics buffer. GBA (5 µl) was added to each well to yield a final enzyme concentration of 6.5 nM. The activity against 4-MU-Glc was monitored continuously over 5 min at room temperature by measuring the fluorescence of liberated 4-MU (λ_{ex} = 360/20 nm, λ_{em} = 450/30nm) using a CLARIOstar Plus microplate reader (BMG Labtech). The assays were performed in quadruplicate for each BTP concentration. A linear calibration was generated by measuring the fluorescence of the 4-MU product (λ_{ex} = 360/20 nm, λ_{em} = 450/30 nm) prepared at serial dilutions in kinetics buffer. Using the 4-MU calibration, the initial rate of substrate hydrolysis (*V*) was determined at each concentration of BTP. The rates (*V*) were plotted against the log of the BTP concentration and were fitted by nonlinear regression to the four-parameter logistic function to determine the IC_{50}.

4.3.8.5 Thermofluor Analysis

Triplicate 25 µL reactions containing 2 µM recombinant GBA and 5x SYPROrange dye were prepared in McIlvaine buffer at pH 4.5, 5.2 and 7.0. The thermofluor assay was performed in a Stratagene Mx3005P qPCR instrument. The SYPROrange dye was excited at λ_{ex} 517 nm and the resulting fluorescence was monitored at λ_{em} 585 nm as the temperature was ramped from 25-95 °C at a rate of 2 °C min^{-1}. Data analysis was performed using the JTSA software (Bond, PS. JTSA. (2017). at http://paulsbond.co.uk/jtsa)[624]. The average fluorescence was plotted against temperature and fitted to a sigmoid-5-function at each pH.

4.3.8.6 Nano Differential Scanning Fluorimetry (NanoDSF) Analysis

Recombinant GBA was prepared at 1 mg mL^{-1} in McIlvaine pH 5.2 buffer in triplicate 20 µL reactions. The samples were loaded into Prometheus™ high sensitivity capillaries and analysed in a Prometheus NT.48 instrument (Nano Temper Technologies). The thermal unfolding was monitored by tryptophan fluorescence (λ_{ex} 260 nm, λ_{em} 330 and 350 nm, 35 % gain) as the temperature was ramped from 25-90 °C at a rate of 1 °C min^{-1}. The light scattering was also monitored. Protein denaturation curves were subsequently generated by plotting the ratio of λ_{em} 330/50 nm against temperature, allowing the protein melting temperature to be calculated from the mid-point of the transition.

4.3.8.7 Multiple Angle Laser Light Scattering (SEC-MALLS)

SEC-MALLS was performed by Andrew Leech (York Technology Facility). GBA was prepared at 1-3 mg mL^{-1} in water and loaded on a Superdex 20 HR 10/300 column connected to a HPLC system and run in SEC buffer (10 mM MES, 100 mM NaCl, 1 mM TCEP, pH 6.5). The HPLC system was connected to a Heleos lighter scattering instrument with a laser wavelength of 658 nm (Wyatt Technology) and an Optilab dRI detector to perform the multiple angle laser light scattering. Data analysis was performed using the ASTRA software interface from Wyatt technology with a BSA sample as a standard.

4.3.8.8 Mass Spectrometry

The contaminant band observed during protein purification at ~90 kDa was isolated by SDS-PAGE and sent for mass spectrometry analysis which was performed by Chris Taylor (York Technology Facility). The contaminant band was digested with trypsin and data were acquired

on an Orbitrap Fusion LC-MS system. In order to reduce contamination, two blanks were run prior to the sample and any matches persisting into the gel band run were subtracted from the final result. The spectra were searched against a combination of the *Trichoplusia sp* and Uniprot databases using PEAKS Studio X+ for unique peptides and subsequent matches were adjusted to a target false discovery rate of 1%.

Intact Solution Mass Spectrometry: Recombinant GBA was diluted to 0.7 mg mL^{-1} in 1% formic acid and 10% acetonitrile. The acidified sample (5 µL) was injected over an MSPac DS-10 Desalting Cartridge flowing at 30 µL min^{-1} using a NanoAcquity HPLC (Waters). Following a 5-minute wash with 20% acetonitrile and 0.1% formic acid in water, protein was eluted into a maXis UHR-ToF (Bruker) with a 10-minute gradient from 20-55% acetonitrile. The column was washed for 2 minutes with 80% acetonitrile and equilibrated for 3 minutes with 20% acetonitrile between runs. Following protein signal integration and baseline subtraction, spectra were deconvoluted using the maximum entropy algorithm within COMPASS to calculate protein mass.

Time Dependent Mass Spectrometry Analysis: A 50 µL reaction containing recombinant GBA (1.4 mg mL^{-1}), ABP (160 µM) and McIlvaine pH 5.2 buffer was prepared and incubated at room temperature. A 5 µL sample was taken at various time points and the labelling reaction was halted by acidification with 20 µL 1% formic acid and 10% acetonitrile. The samples were analysed by intact mass spectrometry by injection over an MSPac DS-10 Desalting Cartridge flowing at 30 µL min^{-1} using a NanoAcquity HPLC (Waters) and analysis on a maXis UHR-ToF (Bruker) as described previously. Following protein signal integration and baseline subtraction, spectra were deconvoluted using the maximum entropy algorithm within COMPASS to calculate protein mass.

4.3.9 Crystallisation of Recombinant GBA

4.3.9.1 Crystallisation of Tagged GBA

Purified GBA1_NHis (10 mg mL^{-1}) was tested against a range of commercial crystallisation screens including Index, PACT premier HT-96 and JCSG+ screen from Molecular Dimensions and against an ammonium sulfate screen previously used to crystallise Cerezyme®, See **Chapter 3** (*section 3.2.1*). A hit was identified in well B1 of the PACT premier screen with 0.1 M

MIB (sodium malonate dibasic, imidazole and boric acid) and 25 % PEG 1.5 K. Optimisation of these conditions failed to produce suitable crystals. The protein was treated with EndoH (50 units) and PNGaseF (50 units) at 37 °C for 4 hours in an attempt to deglycosylate the protein for crystallisation. However, neither enzymes were able to deglycosylate the recombinant protein. The 3-glycine linker construct was screened against Index, PACT premier HT-96, JCSG+, Xtal Screen and Salt RX screens. Unfortunately, no suitable crystals were obtained.

4.3.9.2 Crystallisation of Non-Tagged GBA

Purified tagless GBA (10 mg mL^{-1}) was tested against a range of commercial crystallisation screens. An initial hit was found in well H8 of the PACT premier HT-96 screen from Molecular Dimensions[794] with conditions 0.2 M Na_2SO_4, 20% (w/v) PEG3350, 0.1 M bis-Tris-propane pH 8.5. Optimisation of PEG3350 concentration and buffer pH was performed in a 48-well MRC sitting-drop vapour-diffusion format to yield thin, rod-like crystals at pH 7 and 7.5. Further optimisation of PEG3350 and protein concentration resulted in larger crystals of the same morphology. Final optimised conditions were 0.3 µL GBA (10 mg mL^{-1}) + 0.5 µL well solution (0.2 M Na_2SO_4, 14% (v/v) PEG3350, 0.1 M bis-Tris-propane pH 7.0).

Sequential Seeding to Avoid Bis-Tris-Propane: As bis-Tris-propane and related compounds (see Roberts & Davies, 2012[795] for review) are glycosidase inhibitors that would interfere with soaking experiments, sequential seeding was used to obtain crystals in non bis-Tris-propane conditions. Crystals obtained in the presence of bis-Tris-propane were used to generate a concentrated seed stock, according to previously published protocols[614]. In a 48-well MRC sitting-drop vapour-diffusion format, dilutions of the concentrated seed stock (1:100 and 1:1000) were used to screen into the PACT H8 well conditions in which the bis-Tris-propane was substituted with HEPES buffer (pH 7 and 7.5). Through optimisation of PEG3350 concentration, HEPES concentration, protein volume and seeding ratios, crystals suitable for generating new seed stocks were obtained using 0.2 µL GBA (10 mg mL^{-1}) + 0.2 µL well solution (0.2 M Na_2SO_4, 14 % (v/v) PEG3350, 0.25 M HEPES (pH 7.0)) + 50 nL seed solution (1:1000 dilution). These seed stocks were used to re-screen into previous HEPES conditions in a 48-well MRC sitting-drop format, resulting in crystals suitable for analysis under the conditions 0.2 µL GBA (10 mg mL^{-1}) + 0.4 µL well solution (0.2 M Na_2SO_4, 14% (v/v) PEG3350, 0.25 M HEPES (pH7)) + 0.1 µL seed (1:1000).

Crystallisation of Unliganded GBA : An initial hit was also identified in well A5 of the JCSG+ screen from Molecular Dimensions [794] with conditions 0.2 M Mg formate, 20% (w/v) PEG3350. Optimisation of Mg formate, PEG3350 and protein concentration resulted in larger crystals of the same morphology. Final optimised conditions were 0.6 µL GBA (10 mg mL^{-1}) + 0.5 µL well solution (0.2 M Mg formate, 19% (v/v) PEG3350).

*4.3.9.3 Complex with 2-deoxy-2-fluoro-β-D-glucopyranoside (**14**)*

The 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside (2F-DNPGlc, **14**) was prepared using well-established literature protocols by Fen Lui (University of British Colombia). To obtain a co-crystal complex, solid 2F-DNPGlc was added directly to a crystal drop containing crystals generated in HEPES conditions (*section 4.3.8.2*) and soaking overnight.

4.3.9.4 Cryoprotection

All crystals were cryoprotected with well solution supplemented with 25% (v/v) ethylene glycol prior to flash freezing in liquid N_2 for data collection.

4.3.9.5 Data Collection, Structure Solution and Refinement

Data for the bis-Tris-propane (BTP) and 2F-Glc complexes were collected at the i04 beamline of the Diamond Light Source to of 1.56 Å and 1.41 Å resolution respectively. The data were integrated using the –dials pipeline of XIA2[615,616] and data reduction was performed using AIMLESS[617,618] in the CCP4 software suite[796]. The BTP complexed structure was solved by molecular replacement using a previously obtained structure of Cerezyme®[562] as the search model using PHASER[797]. The 2F-Glc co-crystal structure was solved by refining against the bis-Tris-propane complex structure using REFMAC[621].

Data for the unliganded crystal were collected at the i04-1beamline of the Diamond Light Source and integrated using the *autoPROC* pipeline[798]. Data reduction was performed in AIMLESS[617,618] and the data were processed to a resolution of 0.98 Å. The structure was solved by molecular replacement using MOLREP[620] with PDB 2NT1[368] as the search model. All structures were refined using REFMAC[621] followed by multiple rounds of manual model building with COOT[622]. The 0.98 Å unliganded structure was anisotropically refined with multiple TLS refinement cycles using the automatic REFMAC option. Idealised coordinate sets and refinement dictionaries for ligands were generated using AceDRG[799,800] or JLIGAND[623].

Sugar conformations were validated using Privateer[594] and MolProbity[801] was used to assess model validity before deposition to the PDB. Data collection and refinement statistics for all rGBA crystal structures discussed in this work are given in Table 4.2.

Table 4.2: Data collection and refinement statistics for all rGBA crystal structures discussed in this chapter

	BTP Complex	**2F-DNP-Glc (14)**	**Unliganded**
Data collection			
Space group	$P2_1$	$P2_1$	*P*1
Cell dimensions			
a, *b*, *c* (Å)	52.7, 156.2, 68.3	53.1, 76.4, 68.2	44.5, 46.2, 64.2
α, β, γ (°)	90, 102, 90	90, 102, 90	86, 75, 83
Resolution (Å)	66.75-1.56 (1.59-1.56)	52.01-1.41 (1.43-1.41)	31.70-0.98 (1.00-0.98)
R_{merge}	0.105 (1.670)	0.135 (4.033)	0.043 (0.486)
R_{pim}	0.057 (0.90)	0.056 (1.66)	0.037 (0.49)
$I / \sigma I$	7.3 (0.7)	7.2 (0.4)	8.7 (1.0)
$CC_{1/2}$	0.987 (0.336)	0.999 (0.357)	0.994 (0.685)
Completeness (%)	96.8 (94.7)	100 (99.7)	74.2 (10.9)
Redundancy	4.3 (4.4)	6.7 (6.7)	1.9 (1.2)
Refinement			
Resolution (Å)	55.75-1.56	52.01-1.41	31.70-0.98
No. reflections	637325 (31169)	689440 (33887)	394040 (1815)
R_{work} / R_{free}	0.17/0.2	0.18/0.21	0.11/0.13
No. atoms			
Protein	7987	4065	4486
Ligand/ion	418	173	175
Water	907	484	709
B-factors (Å^2)			
Protein	23	21	8
Ligand/ion	39	42	16
Water	34	35	29
R.m.s. deviations			
Bond lengths (Å)	0.014	0.013	0.009
Bond angles (°)	1.80	1.71	1.61
Ramachandran Plot Residues			
In most favourable regions (%)	95	95	95
In allowed regions (%)	4	4	4
PDB code	6TJK	6TJQ	6TN1

4.4 Results and Discussion

In light of the extensive post-translational modifications required to produce active GBA protein, eukaryotic systems which retain all the necessary post-translational modification requirements must be employed for the production of human GBA. Unfortunately, the quantity and quality of GBA produced by a range of eukaryotic systems in non-clinical contexts has been inconsistent[648,711,714,790,791], demonstrating a lack of consensus on a reliable and robust platform for GBA production. Consequently, there is an ongoing reliance on incredibly costly ERT formulations for biochemical, mechanistic and structural studies of GBA. To support our long-standing interest in developing novel ABPs and inhibitors for GBA, a baculoviral expression vector system (BEVS) was developed in this work to circumvent the need for ERT formulations. A number of GBA constructs had to be investigated and optimised to achieve this:

- N-terminally His_6-tagged construct (*section 4.4.1*, pages 195-202) – unsuccessful
- 3-glycine linker His_6-tagged construct (*section 4.4.3*, pages 203-205) – unsuccessful
- Non-tagged construct (*section 4.4.4*, pages 205-224) – successful

However, all constructs employed the honeybee melittin signal sequence, one of the strongest signalling sequences known for insect cells[802], to direct the recombinant protein to the insect cell secretory pathway and ensure the necessary post-translation glycosylation is performed to produce active GBA. A simplified workflow of the insect-baculovirus system employed in this work is summarised in Figure 4.5.

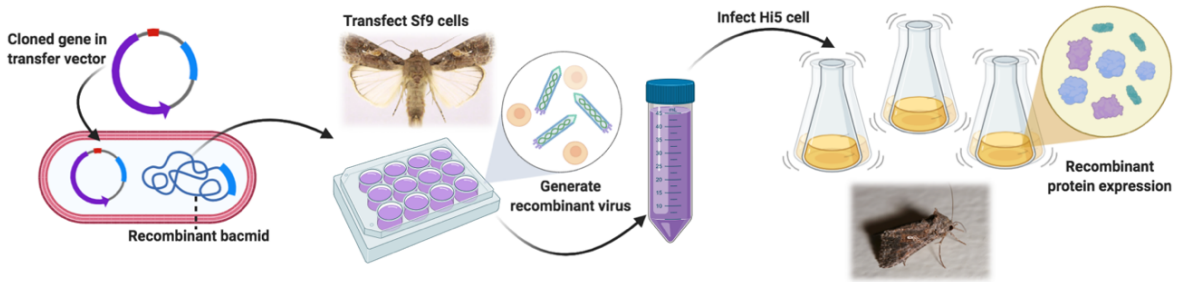


Figure 4.5: Simplified workflow of the insect-baculovirus system employed in this work for the production of recombinant GBA. Figure created in BioRender (www.app.biorender.com).

4.4.1 *N*-Terminally His_6-Tagged GBA Construct

4.4.1.1 Generation of Recombinant Bacmid

Initially, a *N*-terminally His_6-tagged GBA construct (GBA1_NHis) was investigated in hopes that the His_6-tag would allow for facile purification of the resulting protein. The *N*-term truncated *GBA1* gene, lacking its 40 amino acid native signalling sequence, was originally copied from the pGEn1-GBA plasmid (DNASU Clone ID: HsCD00413213[792]) and a tailored *GBA1* insert was generated in which a His_6 tag and a TEV cleavage site were introduced directly downstream of the *GBA1* coding sequence. A transfer plasmid encoding the *GBA1* gene was formed by sequence and ligation independent cloning (SLIC) of the *GBA1* insert with a linearised pOMNI backbone (B24), Figure 4.6. Of note, the backbone contained the honeybee melittin signal sequence, to direct the nascent GBA to the insect cell secretary pathway for secretion, and the two Tn7 transposon sequences (Tn7L and Tn7R) which are required to generate the recombinant bacmid through Tn7 transposition, Figure 4.4 (b). The resulting transfer plasmid was confirmed by HindIII restriction digest and Sanger sequencing before further use.

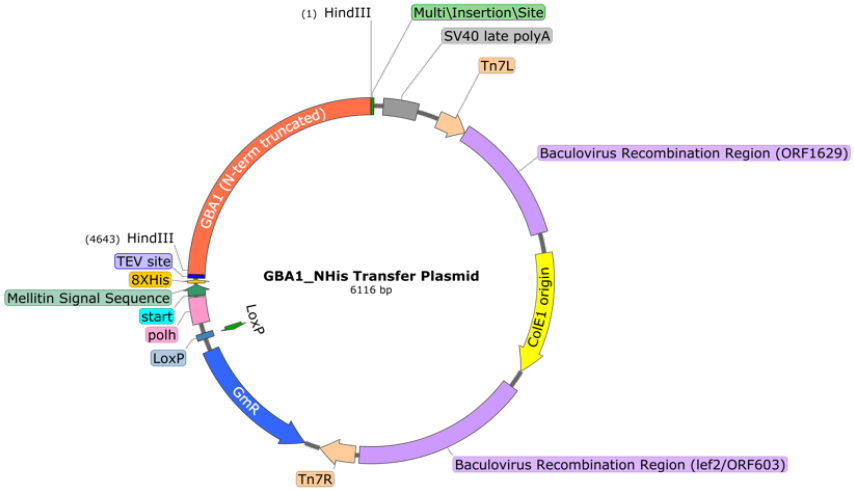


Figure 4.6: Sequenced GBA transfer plasmid generated by SLIC of His_6-tagged *GBA1* insert and linearized B24 plasmid backbone. The transfer plasmid contained the N-term truncated *GBA1* gene, His_6-tag, TEV cleavage site and melittin signalling sequence between the Tn7L and Tn7R transposon sequences. Plasmid map generated in SnapGene Viewer.

The Tn7 transposition method developed by researchers at Geneva Biotech[754,756] was employed in this work to generate a baculoviral vector encoding the human *GBA1* gene. Specifically, the DH10EMBacY *E.coli* strain was used, which contains the EMBacY baculovirus shuttle vector with a mini-attTn7 target site, a tetracycline resistant helper plasmid encoding the transposase

enzyme, a yellow fluorescent protein (YFP) reporter gene and LacZα gene. The GBA_NHis transfer plasmid was transformed into DH10EMBacY cells by electroporation, resulting in transposition of the full *GBA1* cassette from the transfer plasmid into the bacmid by action of the Tn7 transposase enzyme. Upon productive integration of the *GBA1* cassette into the bacmid, the LacZα gene is disrupted which permits selection of successfully transformed colonies via blue-white screening. White colonies were re-streaked and further confirmed by colony PCR using primers (P55 and P56) which amplified across the bacmid Tn7 insertion site. The recombinant bacmid was subsequently purified from overnight cultures of successful colonies and verified by PCR using two sets of primers; one set of primers to amplify the *GBA1* gene (P51 and P52) and another set of primers (P55 and P56) to amplify across the transposition site.

4.4.1.2 Generation of Recombinant Bacmid

S*f*9 and S*f*21 cell lines (*Spodoptera frugiperda)* are often recommended for initial transfection and amplification of recombinant baculovirus because they are reported to be more efficient in generating infectious viral particles compared to other cell lines, including Hi5[778]. Consequently, Sf9 cells were employed in this work to generate recombinant baculovirus encoding His_6-tagged human GBA. Sf9 cells were cultured to log phase growth and transfected with purified recombinant bacmid using a non-liposomal transfection agent[793]. Once > 95% transfection was achieved, as indicated by expression of the EMBacY YFP marker gene, the supernatant containing the recombinant baculovirus was harvested. This recombinant virus was further amplified in Sf9 cells to produce a second-generation viral stock.

4.4.1.3 Protein Production and Purification

Hi5 (*Trichoplusia ni*) cells are commonly employed for the expression of recombinant proteins because they are reported to exhibit 5-10-fold higher expression levels and subsequently greater protein yields in comparison to Sf9 cells[778]. Therefore, a Hi5 cell line was employed for the expression of recombinant GBA in this work. Hi5 cells were cultured to 3.6 L and infected with the second-generation recombinant viral stock. The cultures were allowed to express until YFP fluorescence was observed in 95% of the cells, after which the supernatant was harvested.

Following expression, the *N*-terminally His_6-tagged protein was purified from the cell culture media according to the procedure outlined in Figure 4.7. Firstly, the protein was extracted from the media by immobilised metal affinity chromatography (Ni^{2+} affinity chromatography) which

resulted in a protein band at ~60 kDa corresponding to GBA. A coeluting contaminant band was also observed at ~90 kDa, which proved difficult to remove during purification. Moreover, when this contaminant was successfully removed, the GBA protein precipitated out of solution. Consequently, various detergents including cholic acid, Triton X-100 and N-dodecyl-β-D-maltoside (DDM) were used to keep the protein stable during purification. Of all the detergents tested, DDM proved to be the most effective and was used in all purification buffers at a concentration below the critical micelle concentration (CMC).

Table 4.3: Matches identified for the major contaminant according to unique peptide matches identified by Orbitrap Fusion LC-MS analysis

Accession	Mass / Da	No. matches	Description	Origin
Q9BLC5	82713	13	Heat shock protein 83 (Hsp 83)	*Bombyx Mori*
Q1HPK6	95750	7	Translation elongation factor 2 (tef2)	*Bombyx Mori*
P08161	95279	9	Early 94 kDa protein	*AcMNPV*

Given the persistence of the contaminant and the undesirable protein behaviour induced on its removal, the contamination band was analysed by tryptic digest mass spectrometry. The resulting spectra were searched against *Trichoplusia sp* and Uniprot databases for unique peptide matches. Three matches were identified, including the heat shock protein 83 (Hsp83), the translation elongation factor 2 (tef2) and the early 94 kDa protein, Table 4.3. Based on the high number of peptide matches, Hsp83 is the most likely protein to account for the contaminant band. This is further supported by the stabilising function of this heatshock protein and the drop in GBA stability observed on removing the contaminant.

Following initial purification by Ni^{2+} affinity chromatography, GBA was further purified through cation exchange, size exclusion chromatography (SEC) and a final buffer exchange step to remove the DDM detergent and generate protein suitable for crystallisation, Figure 4.7. A typical yield of 3 mg per 3.6 L (0.83 mg L^{-1}) expression media was achieved. A previous study in which GBA was produced in insect cells failed to report a yield for comparison[715] and only estimated yields have been provided in the very few studies in which GBA has been purified[650]. Therefore, it is not possible to comment on the expression yield of this system relative to previous studies. Nevertheless, for a complex, human glycosylated protein, this yield was considered reasonable and there was there was sufficient pure protein to perform some biochemical characterisation.

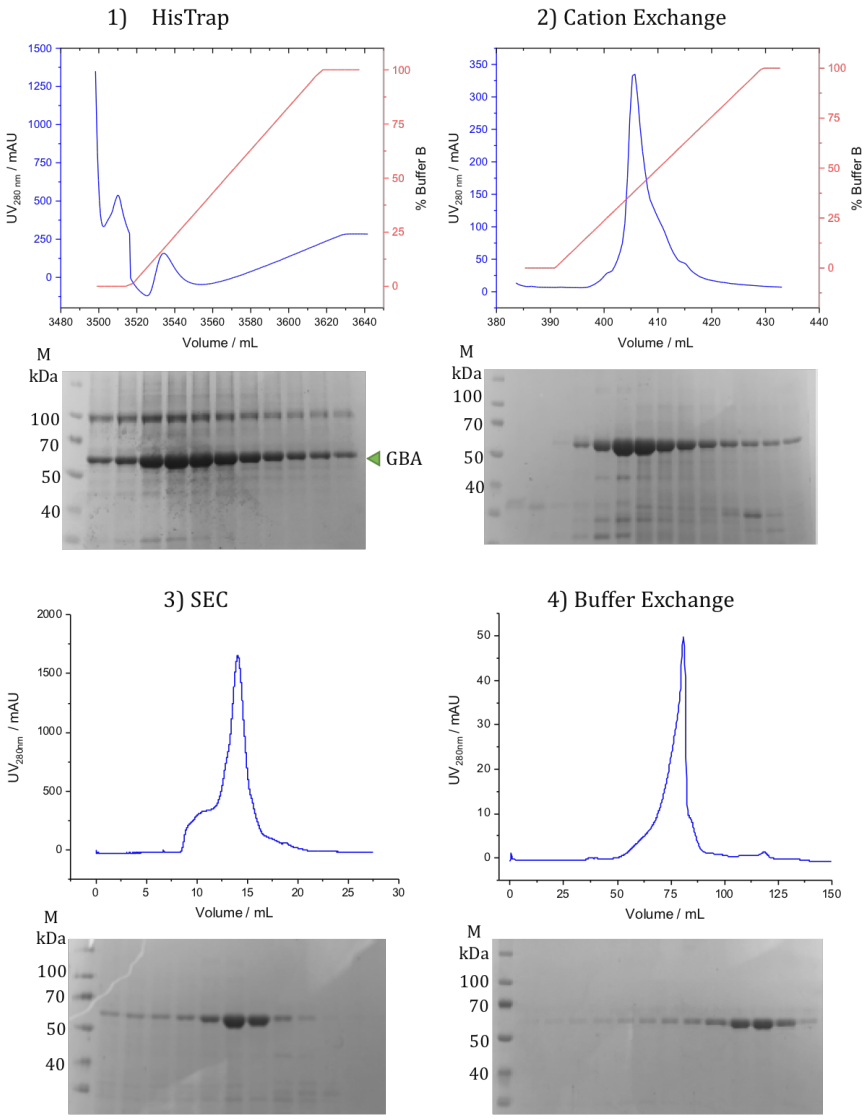


Figure 4.7: Purification of recombinant His_6-tagged GBA (1) GBA (~60 kDa) was extracted from the cell culture media by IMAC with a His-trap column, followed by (2) cation exchange chromatography with a sulfopropyl column and (3) a size-exclusion chromatography in the presence of 0.05% DDM. (4) A final buffer exchange step was performed on an S200 Superdex column to remove the DDM detergent. Typical yield of 3 mg per expression (0.83 mg L^{-1}).

4.4.1.4 Biochemical Characterisation

Activity-Based Labelling

Firstly, in-solution labelling assays of the recombinant protein were performed with a fluorescently tagged cyclophellitol epoxide ABP (ME569) and cyclophellitol aziridine ABP (JJB367) to rapidly determine if the protein was active. The labelling reactions were performed at 37 °C for 1 hour and analysed by SDS-PAGE followed by fluorescent readout. Fluorescently

labelled GBA was observed ~60 kDa, Figure 4.8, demonstrating that the recombinantly produced protein is catalytically active and is labelled by the Overkleeft ABPs. Furthermore, greater labelling was observed with the aziridine ABP, which is concordant with the improved labelling efficiency known for cyclophellitol aziridine probes.

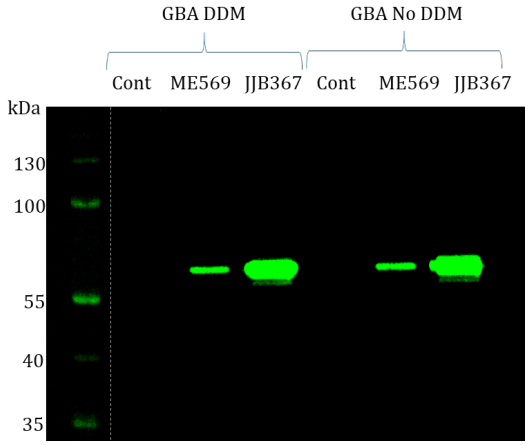


Figure 4.8: In-solution labelling assay of rGBA (200 nM) in the presence and absence of DDM detergent using fluorescently labelled cyclophellitol epoxide ABP (ME569) and cyclophellitol aziridine ABP (JJB367) at 10 μM. Fluorescently labelled GBA was observed ~60 kDa indicating the recombinant protein is catalytically active.

Kinetics

Following activity-based labelling, the enzyme kinetics were assayed in the absence and presence of DDM detergent using the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside (4-MU-Glc). The initial reaction rates were fitted by non-linear regression to the Michaelis-Menten equation allowing K_M and k_{cat} values to be determined, Figure 4.9 (a). Encouragingly, the catalytic activity of GBA in the absence of DDM detergent (k_{cat} = 126 $\pm$ 3 min^{-1}) was found to be comparable to GBA in the presence of DDM (k_{cat} = 160 $\pm$ 13 min^{-1}). Furthermore, the enzymatic activity of both protein samples remained stable for ~ 5-weeks when stored in the fridge at 4 °C, demonstrating reasonable long-term stability, Figure 4.9 (b). The comparable activity and long-term stability of GBA in the presence and absence of DDM suggests the detergent may not be required to keep the protein stable once purified. This holds a particular advantage for crystallography studies, as the presence of detergents typically hinders crystallisation.

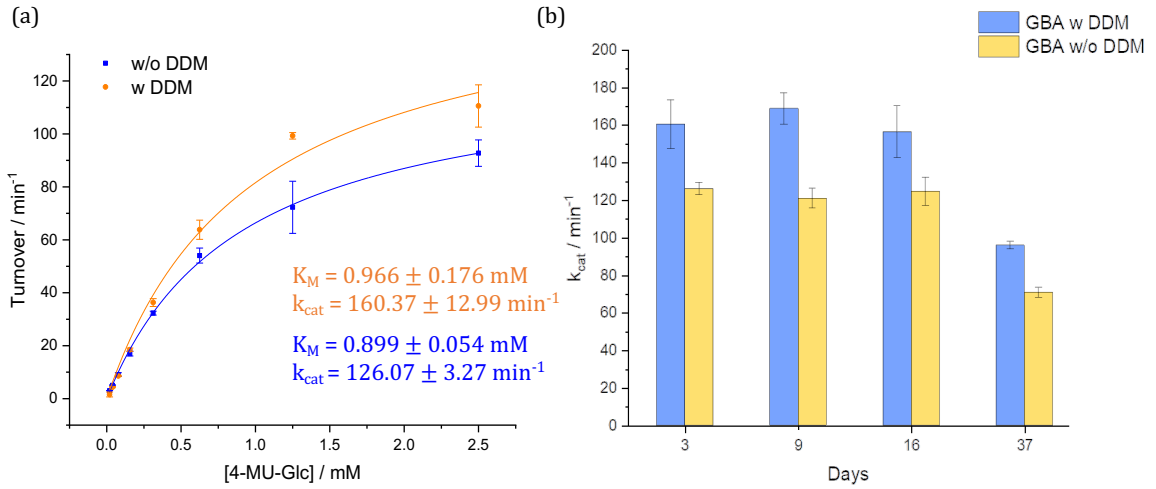


Figure 4.9: (a) Michaelis-Menten kinetics of GBA with and without DDM using 4-MU-Glc at room temperature. (b) Long term catalytic activity of GBA when stored in the fridge at 4 °C. A 40% and 44% decrease in activity was observed in the presence and absence of DDM respectively at day 37. All data plotted as the average ± stdev of 4 replicates.

Thermal Stability

The thermal stability of the recombinant protein was initially investigated through a Thermofluor assay using Sypro Orange dye. However, considerably high background fluorescence was observed which prevented the generation of suitable thermal shift curves. This high background fluorescence could be attributed to residual DDM in the samples. Consequently, Nano Differential Scanning Fluorimetry (NanoDSF), which relies on intrinsic protein fluorescence from tryptophan and tyrosine residues, was employed to evaluate the thermal stability of the recombinant GBA formulation.

The protein fluorescence was monitored at λ_{em} 330 nm and 350 nm as the protein was denatured by heating from 20-95 °C. The ratio of 350/330 nm fluorescence was plotted against temperature to yield a fluorescence thermal stability curve from which the protein melting temperature (Tm) could be determined. At pH 7.0, the Tm of GBA was calculated to be 51°C, Figure 4.10 (a). Unfortunately, due to the limited supply of Cerezyme®, NanoDSF analysis could not be performed for direct comparison, however, a previous ThermoFluor assay revealed the Tm of Cerezyme® to be 50 °C at pH 7.0, Figure 4.10 (b). These preliminary results indicate that the recombinant GBA produced in insect cells in this work exhibits comparable thermal stability to that of Cerezyme® produced in CHO cells.

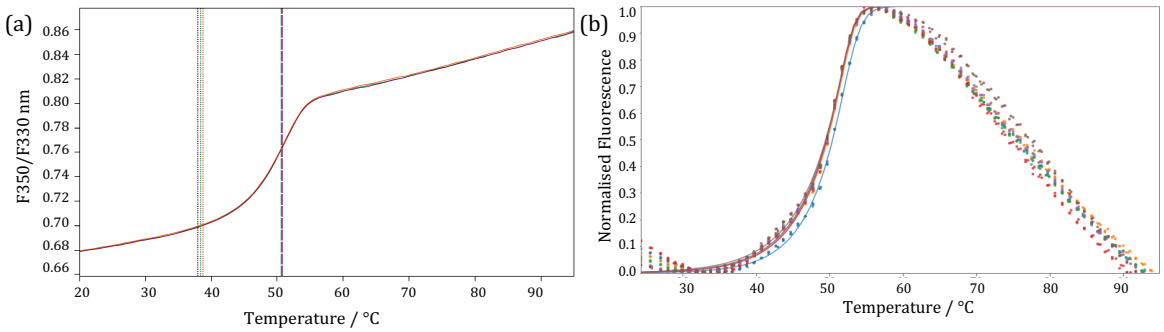


Figure 4.10: (a) NanoDSF analysis of recombinant His_6-tagged GBA at pH 7.4. Protein melting temperature determined from the midpoint of fluorescence transition, Tm = 50 °C (average of triplicate). (b) Thermofluor analysis of Cerezyme® at pH 7.4. Protein melting temperature determined from midpoint of the fluorescence transition, Tm = 51 °C (average of 5 replicates)

4.4.1.5 Crystallisation

Subsequent to brief biochemical characterisation, crystallisation of the recombinant His_6-tagged GBA was attempted. Unfortunately, despite extensive screening and optimisation, crystals suitable for crystallography studies could not be obtained. Several factors could have hindered crystallisation, including residual DDM detergent, protein glycosylation and the presence of the His_6-tag. Removal of residual DDM detergent is incredibly challenging because DDM is not dialysable. In regard to protein glycosylation, attempts were made to cleave the *N*-glycans using deglycosylating enzymes such as Endo H (cleaves *N*-linked glycans between the two GlcNAc units of the chitobiose core) and PNGase F (cleaves between the innermost GlcNAc and asparagine residues). However, both enzymes proved ineffective under non-denaturing conditions and the *N*-glycosylation could not be altered, Figure 4.11. Owing to expression in insect cells, it was postulated that the *N*-glycosylation profile of this GBA formulation may contain core α-1,3 or α-1,6-fucose residues which prevent the cleavage of *N*-glycans by PNGase F. This is further compounded by the fact that Endo H is unable to cleave complex glycans. Consequently, removal of the *N*-glycans also proved difficult. Lastly, the protein was treated with TEV protease in an effort to cleave the *N*-terminal His_6-tag. Unfortunately, the tag could not be cleaved, even in the presence of excess TEV protease and prolonged incubation periods, Figure 4.11. The TEV site (ENLYFQG) had been confirmed by sequencing of the original transfer plasmid, therefore, it was suggested that the cleavage site was inaccessible, potentially due to its close proximity to the protein.

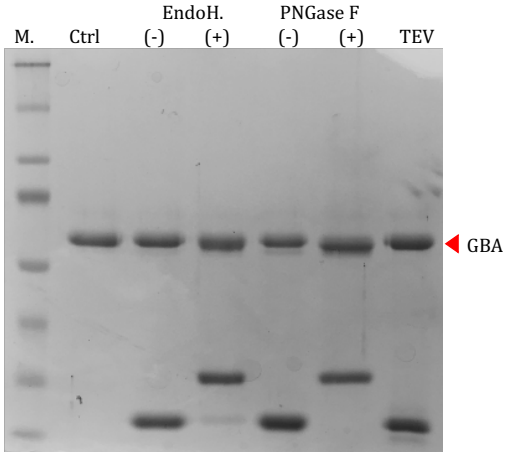


Figure 4.11: Treatment of His_6-tagged GBA with TEV protease, Endo H and PNGase F under denaturing (+) and non-denaturing (-) conditions. Some change in protein glycosylation observed under denaturing conditions with Endo H and PNGase F. No change under non-denaturing conditions. No change observed when treated with TEV protease.

4.4.2 Alternative GBA Constructs

Considering all the problems highlighted with the His_6-tagged construct, several new constructs were designed in an attempt to overcome the tag cleavage problem. Four constructs with progressively longer glycine linkers between the GBA coding sequence and TEV cleave site (1G, 2G, 3G and 4G) were synthesised to improve the accessibility of TEV site. A tagless construct was also generated; however, purification of untagged proteins is inherently more complex and labour intensive; therefore, it was hoped that one of the glycine-linked His_6-tagged constructs would be suitable. To simplify the workflow, the recombinant transfer plasmids, bacmids and baculoviral stocks for all constructs were generated in tandem using the procedures described previously for the N-terminally His_6-tagged construct and with the primers and backbones outlined in Figure 4.3 (b).

4.4.2.1 Test Expressions

To ensure each construct could produce active GBA, small-scale test expressions were performed in Hi5 cells. Following expression, the supernatant was harvested and the cells were resuspended in water. Activity-based labelling with ABP JJB367 was performed on the supernatant and resuspended cells to rapidly detect GBA activity. The reactions were separated by SDS- PAGE and analysed by fluorescent readout to reveal active GBA at ~60 kDa in both the supernatant and resuspended cells of all constructs, Figure 4.12. Although greater fluorescence

signal was observed in the resuspended cells, it is important to note that this results from resuspension of the cells in ~5 mL volume which is considerably smaller than the 50 mL of media from which the supernatant labelling reaction sample was taken. Therefore, any GBA in the cells would be at a substantially higher concentration. In light of this, it is clear that a significant portion of recombinantly produced GBA is secreted into the media by the action of the melittin signal sequence, but some GBA is stored intracellularly. Further quantification of GBA secretion was not performed in this work. Most importantly, these labelling reactions demonstrate that all constructs produce active GBA, therefore all constructs should be suitable for further full-scale expression studies.

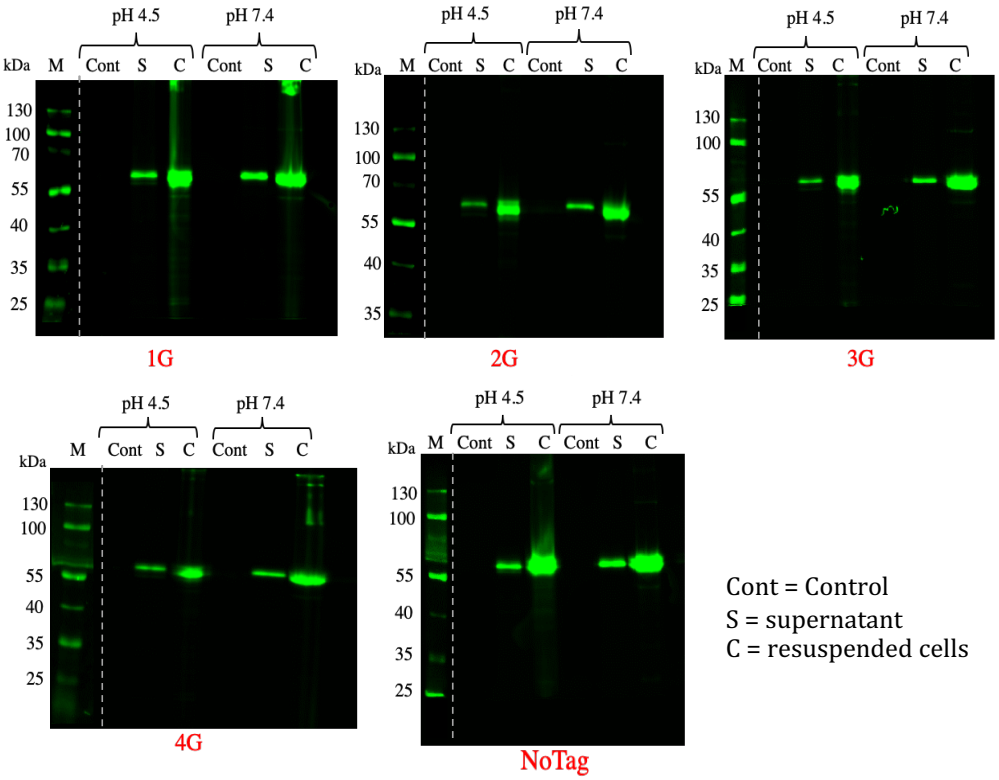


Figure 4.12: In-solution labelling reactions of the supernatant (S) and resuspended cells (C) of each test expression with ABP JJB367. Fluorescent readout of SDS-PAGE analysis shows labelled GBA at ~60 kDa, indicating active GBA is produced by all constructs. 1G-4G = glycine linker constructs, NoTag = tagless construct.

4.4.3 3G Glycine Linker His_6-Tagged Construct

The three-glycine linker construct (GBA_3G) was chosen as the first alternative construct to be investigated, as it was hoped the glycine linker would be sufficiently long to render the TEV cleavage site accessible and overcome the His_6-tag problem. It was also hoped that if this

construct was successful, the 1-glycine and 2-glycine linker constructs could be investigated to reduce the length of the glycine chain that would remain appended to the *N*-terminus of the protein. All the necessary cloning required to produce the GBA_3G recombinant transfer plasmid and bacmid was performed as described for the N-terminally His_6-tagged construct using the relevant primers and backbone, Figure 4.3 (b).

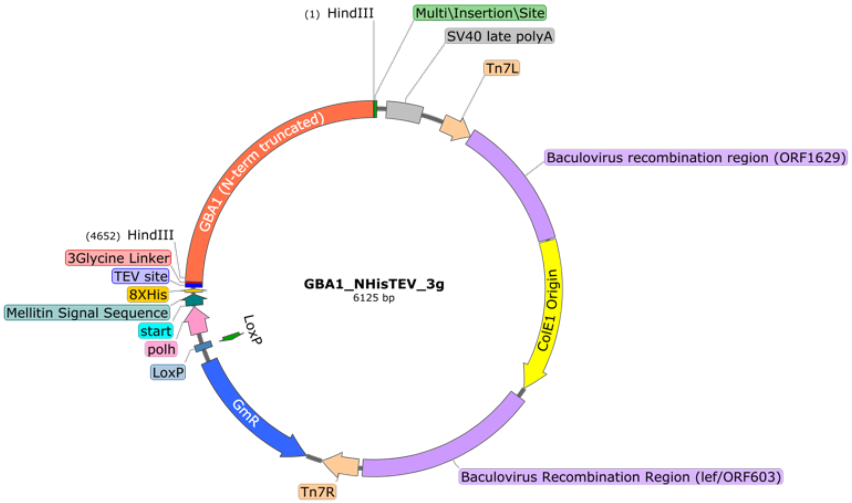


Figure 4.13: Sequenced transfer plasmid containing the N-term truncated *GBA1* gene, a 3-glycine linker, TEV cleavage site and His_6-tag. Plasmid map generated in SnapGene Viewer.

4.4.3.1 Protein Production and Purification

Following a full-scale expression in Hi5 cells, recombinant GBA_3G was purified from the supernatant according to the same procedure outlined for the *N*-terminally His_6-tagged construct, Figure 4.7. However, the final buffer exchange step to remove the DDM detergent was not performed as this was not considered vital to determining the effect of the linker. Following HisTrap, cation exchange and size exclusion chromatography, sufficiently pure protein was obtained at a yield of 3.5 mg (0.97 mg mL^{-1}), Figure 4.14. Importantly, there was adequate protein to perform several optimisation assays with TEV protease to assess whether the His_6-tag could be cleaved. Unfortunately, despite introducing the 3-glycine linker, the His_6-tag could not be cleaved by treatment with TEV protease. The reason for this remains unclear, however, given the hydrophobicity of GBA, it's possible that the tag folds into the protein and binds through a hydrophobic surface or pocket rendering it inaccessible to TEV protease. Consequently, it seemed unlikely that the remaining glycine linker constructs would be suitable and efforts turned to expression and purification of the tagless construct.

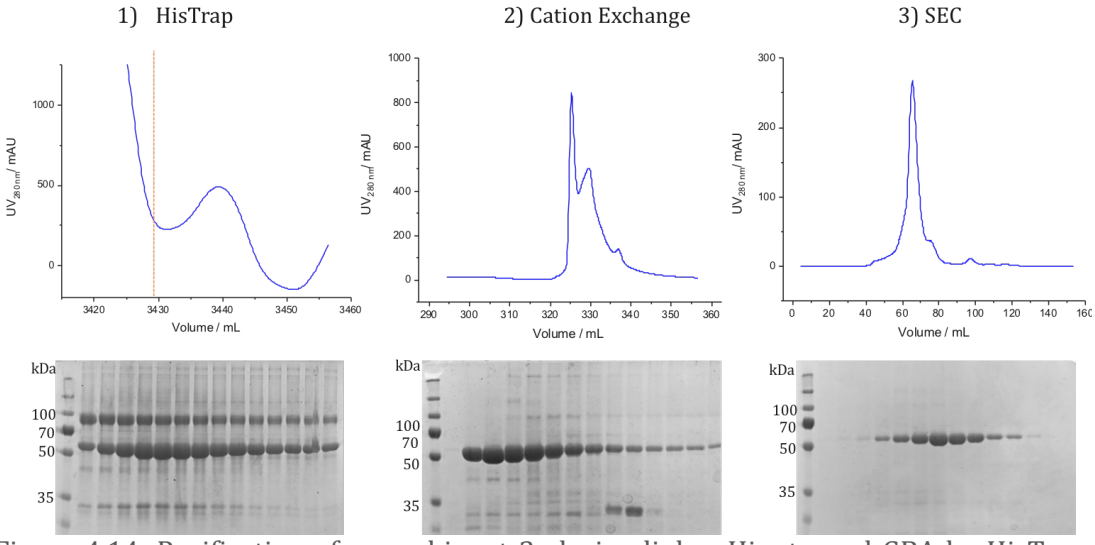


Figure 4.14: Purification of recombinant 3-glycine linker His_6-tagged GBA by HisTrap (1), cation exchange chromatography (2) and size-exclusion chromatography (3) in the presence of 0.05% DDM. Typical yield of 3.5 mg from 3.6 L expression volume (0.97 mg L^{-1}).

4.4.4 Non-Tagged Construct

4.4.4.1 Cloning

All the necessary cloning required to produce the transfer plasmid and bacmid for the tagless GBA construct was performed according to the same procedure employed for the His_6-tagged constructs using the primers and backbone outlined in Figure 4.3. The recombinant transfer plasmid was verified by HindIII digest and sanger sequencing, whilst the recombinant bacmid was verified by PCR, Figure 4.15. The recombinant baculovirus was generated in Sf9 cells.

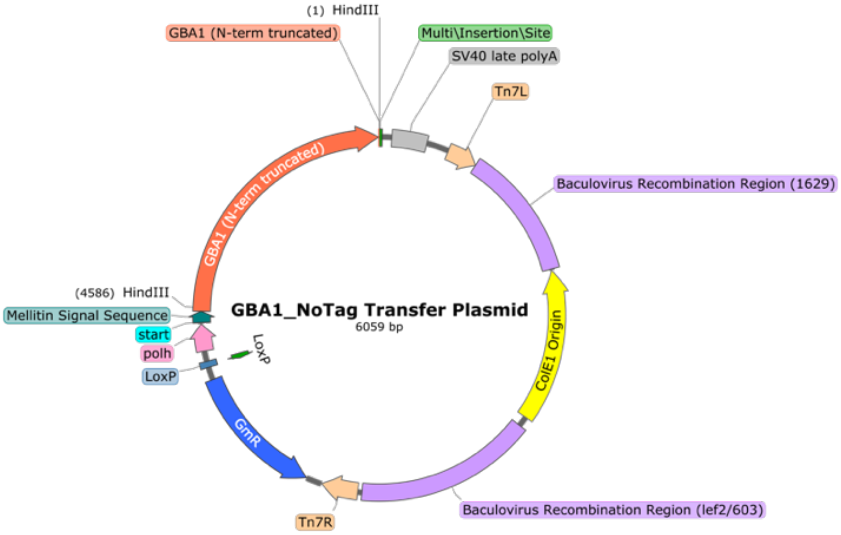


Figure 4.15: Sequenced transfer plasmid encoding the non-tagged N-terminally truncated *GBA1* gene. Plasmid map generated in SnapGene Viewer

4.4.4.2 Protein Production and Purification

Tagless rGBA was expressed in Hi5 cells and purified from the cell culture media using a modified procedure of the protocol described by Sawkar et al. (2006)[715]. Specifically, rGBA was extracted from the cell culture media using hydrophobic interaction chromatography and then purified on a heparin affinity column using its cation exchange properties. A second round of weak cation exchange was performed using a carboxymethyl sepharose column in the presence of 0.01 % TWEEN® 80 (polysorbate 80), which is a commonly used detergent for the extraction of membrane proteins. A final SEC step was performed to remove the detergent and generate protein suitable for x-ray crystallography, Figure 4.16.

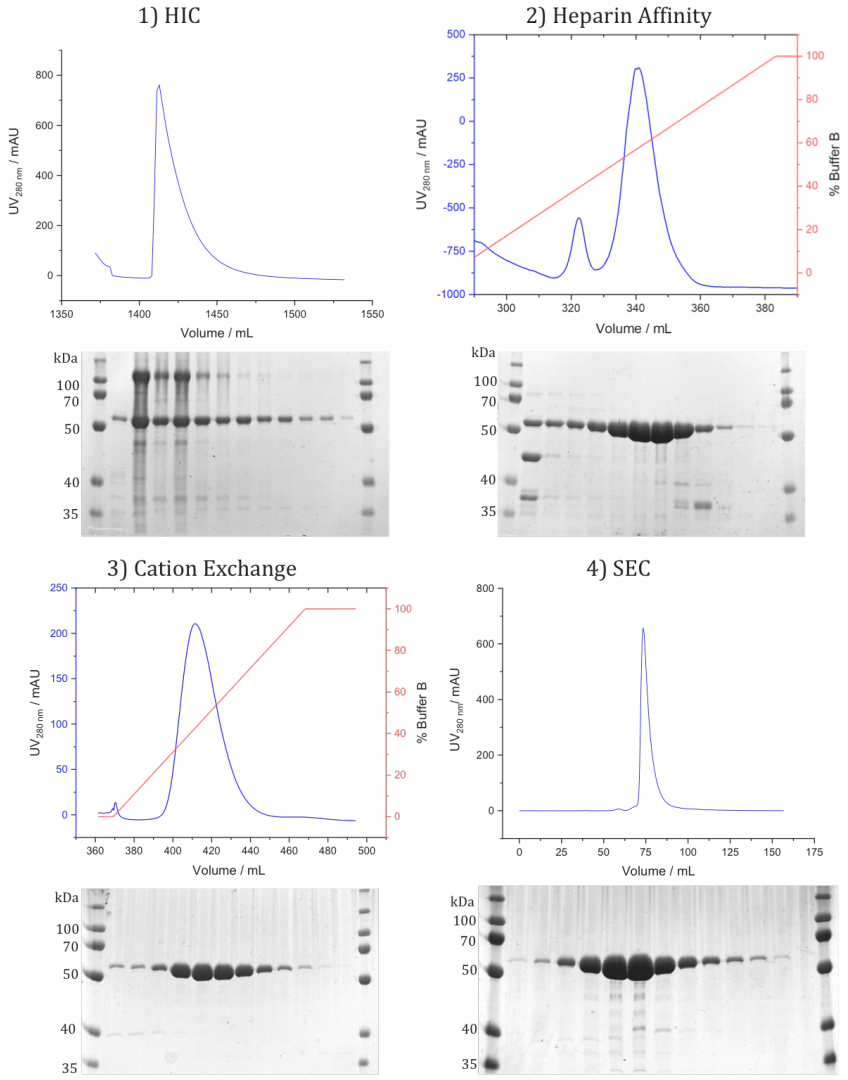


Figure 4.16: Purification of recombinant tagless GBA by (1) hydrophobic interaction chromatography, (2) cation exchange chromatography with a heparin affinity column (3) cation exchange with a carboxymethyl sepharose column in presence of 0.01 % TWEEN 80 and (4) size exclusion chromatography. Typical yield of 13-27.0 mg (3.6-7.5 mg L^{-1}).

Following purification, a typical yield of 13.0-27.0 mg (3.6-7.5 mg mL^{-1}) was achieved, which is a considerable improvement on the 3.0-3.5 mg obtained for the His_6-tagged constructs. Unfortunately, the study in which the original purification was described failed to report a yield for comparison[715] and only estimated yields have been provided in the very few studies in which GBA has been purified[650]. Thus, is it not possible to comment on the expression yield of this system relative to previous studies. Nevertheless, this expression system generates sufficient protein for both biochemical and structural studies.

4.4.4.3 Biochemical Characterisation

The biophysical and biochemical properties of the tagless GBA formulation were investigated to assess whether this recombinant product could be a viable alternative to ERT formulations for non-clinical use.

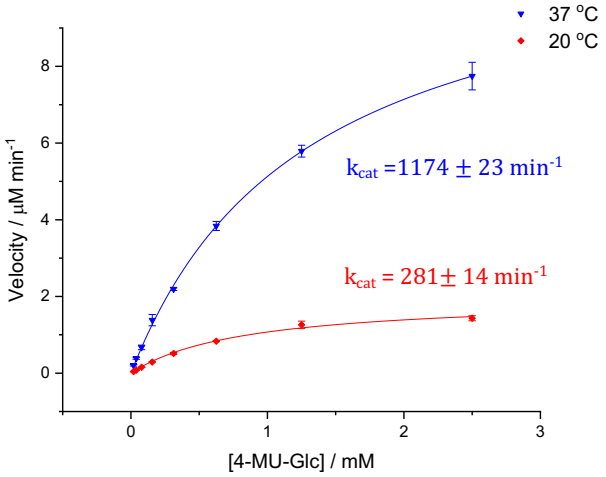
Enzyme kinetics

Firstly, the Michaelis Menten kinetics were assayed at room temperature and 37 °C using the fluorogenic substrate 4-MU-Glc. Initial reaction rates were fitted to the Michaelis-Menten equation by non-linear regression and K_M, V_{max} and k_{cat} were calculated, Figure 4.17. A 4-fold increase in k_{cat} was observed on increasing the temperature from 20 °C to 37 °C, with all kinetic parameters comparing favourably to those reported for Cerezyme®[651], Figure 4.17. Indeed, these K_M, V_{max} and values suggest that the GBA formulation produced in insect cell in this work exhibits similar kinetic properties to the commercial product Cerezyme® which is produced in CHO cells. Additionally, the k_{cat} of this recombinant enzyme (k_{cat} = 1174 $\pm$ 23 min^{-1}) is comparable to that of the GBA formulation produced in insect-cells by Sawkar et al. (2006)[715] (k_{cat} = 868 $\pm$ 28 min^{-1}); however, no K_M or V_{max} values were reported in that study.

It would be prudent to note that there have been concerns in the literature regarding the impact of the inner filter effect (IFE) on the kinetic parameters determined using 4-MU substrates. It is known that the inner filter effect reduces the detected fluorescence at high 4-MU concentrations due to an increase in optical density and excess fluorescence absorbance[803,804]. Consequently, the IFE was briefly investigated in this work to determine its impact on the reported kinetic parameters, Appendix 2. The kinetic assay was repeated at a higher excitation wavelength (390 nm) to minimise the excess absorbance of excitation light by the fluorophore and eliminate primary IFEs. Importantly, the resulting kinetic parameters (K_M = 1.371 $\pm$ 0.244

mM, V_{max} = 21.26 ± 1.88 µM min^{-1}, k_{cat} = 2126 min^{-1}) were consistent with those obtained in the original assay, indicating that the reported kinetic parameters are minimally affected by the IFE. Furthermore, the majority of kinetic studies on this enzyme, including those cited in this work for comparison, use 4-MU-Glc as the fluorogenic substrate.

(a)



(b)

At 37 °C	Recombinant GBA	Cerezyme	rhWT-GBA
K_M (mM)	1.288 ± 0.051	1.127 ± 0.052	-
V_{max} (µM min^{-1})	11.74 ± 0.23	2.21 ± 0.03	-
k_{cat} (min^{-1})	1174 ± 23	1325 ± 2	868 ± 28

Figure 4.17: (a) Michaelis-Menten kinetics of tagless GBA using 4-MU-Glc at 20 °C and 37 °C. Data plotted as the average ± stdev of 4 replicates. (b) Comparison of kinetic parameters with those reported for Cerezyme®[651] and GBA produced in insect cells by Sawkar et al. (2006)[715] (rhWT-GBA) also determined using 4-MU-Glc at 37 °C.

Activity Enhancement with Saposin C

Perhaps more interestingly, the activity of this GBA formulation against the artificial substrate 4-MU-Glc was enhanced 6.4-fold when in the presence of the activator protein Saposin C (SapC). Saposins are a family of small, heat stable, sphingolipid activator proteins which are required to supplement the activity of certain lysosomal glycoside hydrolases[805,806]. Specifically, SapC has been shown to activate human GBA for efficient glucosylceramide (GlcCer) hydrolysis *in vivo*[95,807,808]. In fact, patients with normal GBA activity but deficient SapC levels have been reported to exhibit Gaucher-like symptoms, demonstrating the *in vivo* significance of SapC in GlcCer metabolism[92,100]. Although the mechanism of activation of SapC towards GBA is still unknown, a model has been proposed in which SapC and GBA bind each other in the lysosomal membrane where the sphingolipid is located, thereby enhancing the accessibility of GBA to the sphingolipid[809]. However, SapC has also been shown to enhance the activity of GBA *in vitro*

in the absence of a (lysosomal) membrane but in the presence of detergents[807]. Therefore, the effect of SapC on this GBA formulation was crudely investigated through an activity assay.

Recombinant human SapC was expressed in Origami (DE3) cells, which contain mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes to enhance the formation of disulfide bonds that are required for correct folding of SapC. Following heat treatment to precipitate out larger cytosolic proteins, SapC was purified through anion-exchange chromatography and gel filtration, see Appendix 3 for full details. Subsequently, GBA (10 nM) was incubated with increasing concentrations of SapC (20 nM – 44 μM) in the presence of 0.1% Triton X-100 detergent for 15 mins at 37°C. Following incubation, 4-MU-Glc substrate was added and the reactions were stopped after 15 minutes by the addition of Na_2CO_3 (pH 11). At each SapC concentration, the extent of substrate hydrolysis was determined by measuring the fluorescence of liberated 4-MU product, Figure 4.18. Subsequently, SapC was found to enhance the activity of this GBA formulation 6.4-fold, which consistent with a previous study in which native SapC was shown to stimulate the activity of Ceredase® (a therapeutic GBA formulation) by 7-fold[807]. Therefore, it is encouraging to find that the recombinant GBA produced in insect cells in this work is enhanced by its activator protein SapC and that this increase in GBA activity is comparable to that achieved with a therapeutic formulation. This unearths future opportunities to further probe the interactions of GBA and SapC, which is of considerable interest in Gaucher disease research.

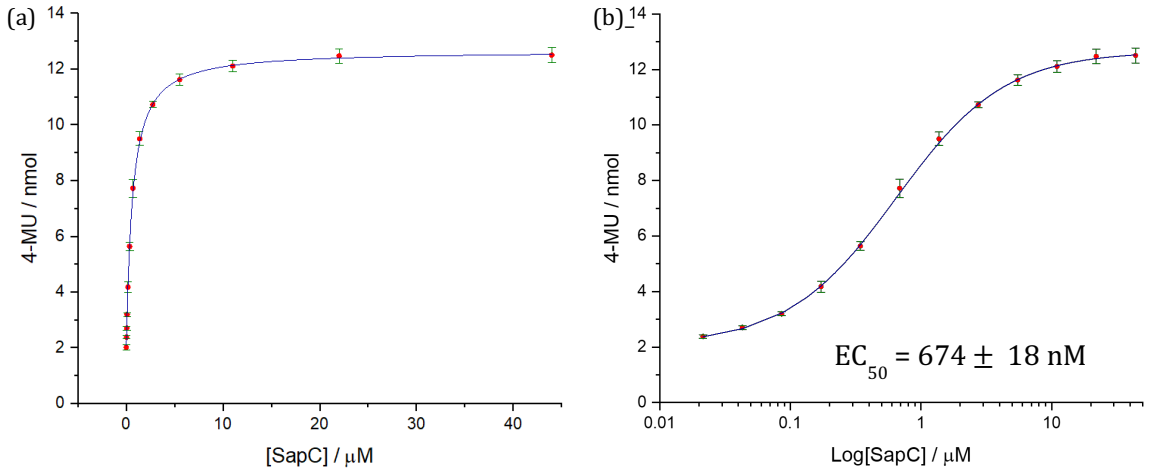


Figure 4.18: Effect of increasing SapC concentrations (20 nM-44 μM) on the activity of rGBA (10 nM) against 4-MU-Glc. Liberated 4-MU product (nmol) plotted against (a) SapC concentration and (b) Log of the SapC concentration and fitted to 4-parameter logistic function. A 6.4-fold increased in GBA activity was observed with an EC_{50} = 674 ± 18 nM. Data plotted as average ± stdev of 4 replicates.

Long Term Stability

The long-term stability of the recombinant GBA formulation was also investigated by monitoring the catalytic activity over time. When stored at 4 °C, the enzymatic activity remained stable for ~5-6 weeks, after which the activity gradually dropped to 50% by day 82, Figure 4.19. This suggests the recombinant enzyme exhibits good long-term stability once purified. Additionally, no considerable change in k_{cat} was observed upon a freeze-thaw cycle, indicating this protein is suitable for freezing for long-term storage, Figure 4.19.

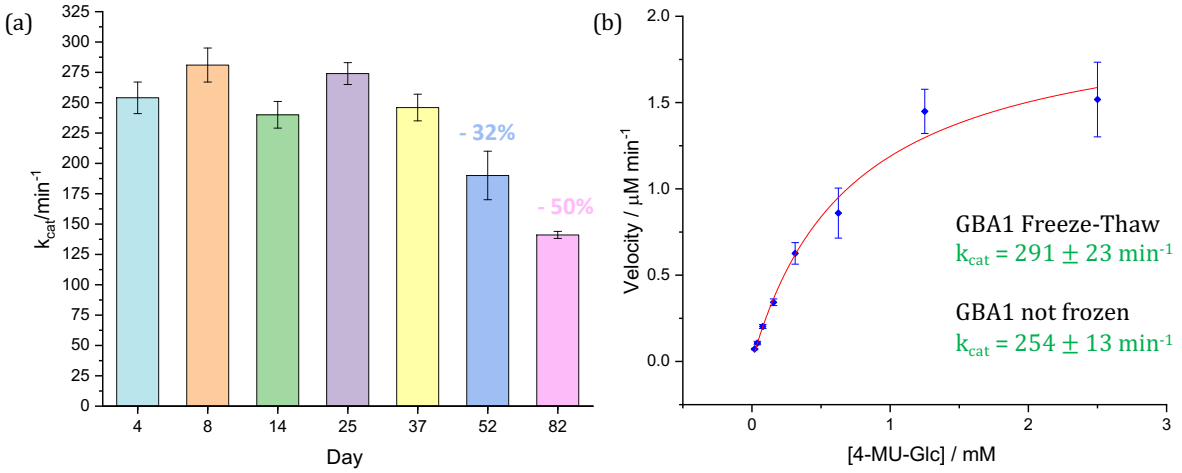
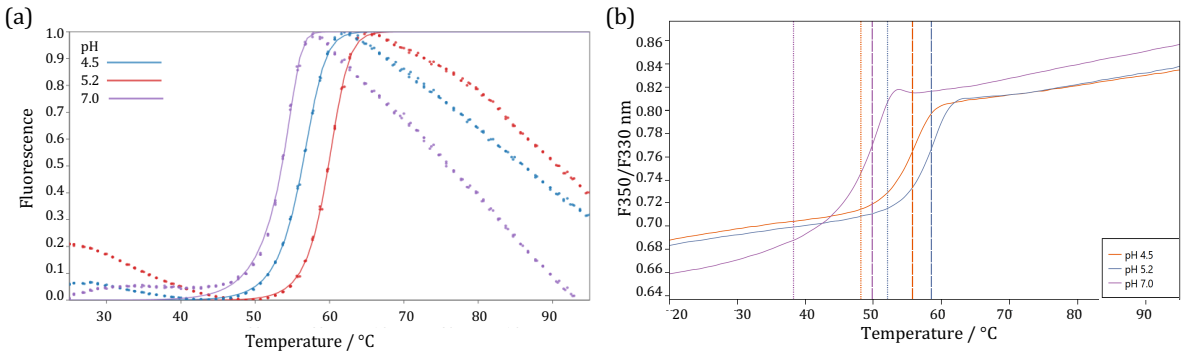


Figure 4.19: (a) Long term activity of GBA stored at 4 °C. Data plotted as average ± stdev of 4 replicates. Assays performed at 20 °C (b) Michaelis-Menten kinetic assay of GBA performed at 20 °C following a freeze-thaw cycle. No significant change in k_{cat} (291 ± 23 min^{-1}) compared to GBA which has not been frozen (254 ± 13 min^{-1}).

Thermal Stability

Following kinetic analysis, the thermal stability of the recombinant GBA formulation was evaluated. In contrast to the *N*-terminally His_6-tagged formulation, ThermoFluor analysis could be performed on the tagless GBA protein as the background fluorescence was sufficiently low to generate suitable thermal shift curves. This indicates that the TWEEN-80 detergent was successfully removed from the protein during the purification procedure. Consequently, the thermal stability of the tagless recombinant protein was evaluated using both ThermoFluor and NanoDSF analysis at pH 4.5 and 5.2, to mimic the lysosomal pH range, and at pH 7.0. The melting temperatures determined by each technique were concordant, demonstrating an inherent difference in thermal stability with varying pH, Figure 4.20. Specifically, GBA proved more stable at acidic pH than neutral pH, with optimum stability at pH 5.2 (NanoDSF Tm = 58.4 ± 0.0 °C). This pH dependent thermal stability profile is expected for a lysosomal enzyme.



	Tm (°C)	
pH	Thermofluor	NanoDSF
4.5	56.3 ± 0.0	55.7 ± 0.0
5.2	60.0 ± 0.2	58.4 ± 0.0
7.0	53.8 ± 0.1	49.9 ± 0.1

Figure 4.20: Thermal stability analysis of tagless GBA by (a) ThermoFluor and (b) NanoDSF at pH 4.5, 5.2 and 7.0. (c) Melting temperature of GBA at each pH determine by ThermoFluor and NanoDSF. Data are reported as the average ± stdev of 3 replicates.

The lower melting point of GBA at neutral pH has been noted previously and is supported by proteolysis studies in which GBA was found to be resistant to tryptic digestion at pH 5.2 but not at pH 7.4[625]. This behaviour is thought to arise from destabilising changes in the enzymes native fold at neutral pH. Therefore, this recombinant GBA formulation exhibits the expected pH dependent thermal stability profile. Moreover, the melting temperatures reported for our GBA formulation are consistent with those reported for Cerezyme®[625], which also exhibits a pH dependent thermal stability profile, Table 4.4. In contrast, no pH dependent thermal stability was reported for the GBA formulation produced in insect cells by Sawkar et al. (2006)[715]

Table 4.4: Tm values for (a) recombinant GBA produced in this work as determined by ThermoFluor analysis. Data reported as an average ± stdev of 3 replicates (b) Cerezyme® determined by circular dichroism by Bdira et al. (2017)[625] and (c) GBA produced in insect cells by Sawkar et al. (2006)[715] determined by circular dichroism.

Protein	Tm (°C)		
Recombinant GBA[a]	56.3 ± 0.0 (pH 4.5)	60.0 ± 0.2 (pH 5.2)	53.8 ± 0.1 (pH 7.0)
Cerezyme®[b]	-	61 (pH 5.2)	57 (pH 7.4)
rhWT-GBA[c]	-	49.3 (pH 5.3)	49.2 (pH 7.0)

Size Exclusion Chromatography-Multiple Angle Light Scattering (SEC-MALLS)

SEC-MALLS was performed to determine the molecular mass of the recombinant protein and evaluate its behaviour in solution. Surprisingly, SEC-MALLS analysis revealed three peaks at 120 kDa, 59 kDa and 62 kDa, Figure 4.21 (a). The 120 kDa peak likely corresponds to GBA dimer, whilst the 59 kDa peak represents monomeric GBA. However, the large peak at 62 kDa could not be accounted for.

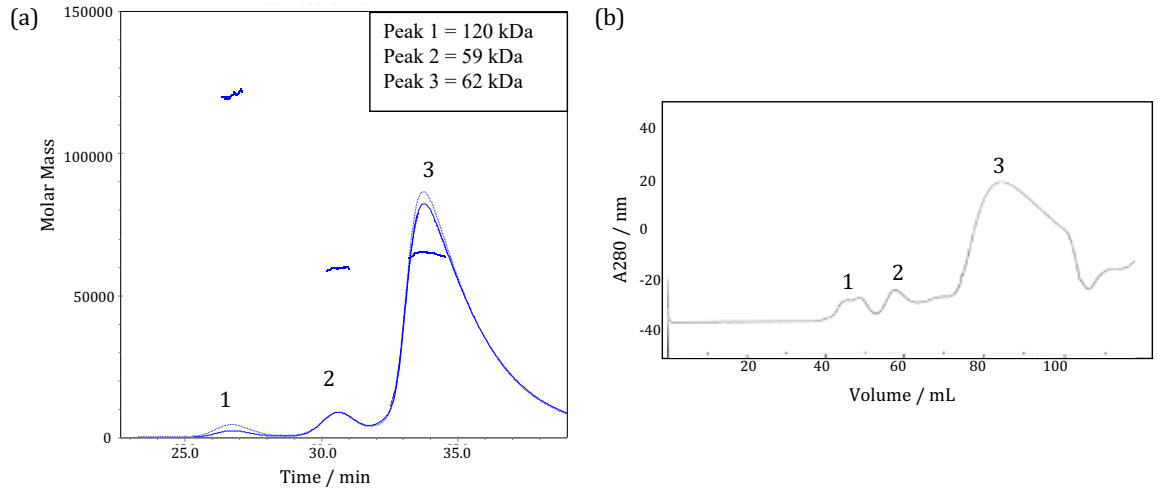


Figure 4.21: (a) SEC-MALLS analysis of tagless GBA. Peak 1 at 120 kDa corresponds to GBA dimer, peak 2 at 59 kDa corresponds to GBA monomer, peak 3 at 62 kDa is unknown. (b) Analysis of Cerezyme® on S200 Superdex column demonstrating same 3-peak elution profile.

It was originally proposed that the unknown peak at 62 kDa may represent a specific, heavily glycosylated form of GBA, however, analysis of the commercial product Cerezyme® on a Superdex column revealed the same 3-peak behaviour, Figure 4.21 (b). The exact cause of this elution behaviour remains unclear; however, it is possible that the elution is altered by interactions between glycosylated GBA and the dextran covalently bound to the agarose particles of the Superdex matrix. This is strongly supported by the fact that this unusual behaviour could be controlled by altering the salt concentration of the SEC buffer. Increasing the salt concentration from 100 mM to 300 mM led to a reduction in the unknown peak at 62 kDa and a concomitant increase in the monomer peak at 59 kDa, Figure 4.22. It is also important to note that relatively minor changes in protein structure, such as glycosylation, may affect protein solubility and encourage secondary interactions with the resin. This in turn may cause similarly sized protein molecules to elute at different times, resulting in broadening of the peaks and poor definition between monomers and oligomers. Therefore, the unusual elution profile may be further complicated by the presence of multiple glycoforms within the protein sample.

Importantly, all 3 SEC-MALLS peaks contained active GBA protein, therefore, combined with the change in elution profile observed on increasing salt concentration, it was concluded that the unusual elution behaviour results from unfavourable interactions of the protein with the Superdex resin rather than the presence of multiple protein oligomers in the sample. It is likely that the SEC step in the initial purification procedure could be performed due to the presence of TWEEN-80 detergent in the sample from the previous cation-exchange step.

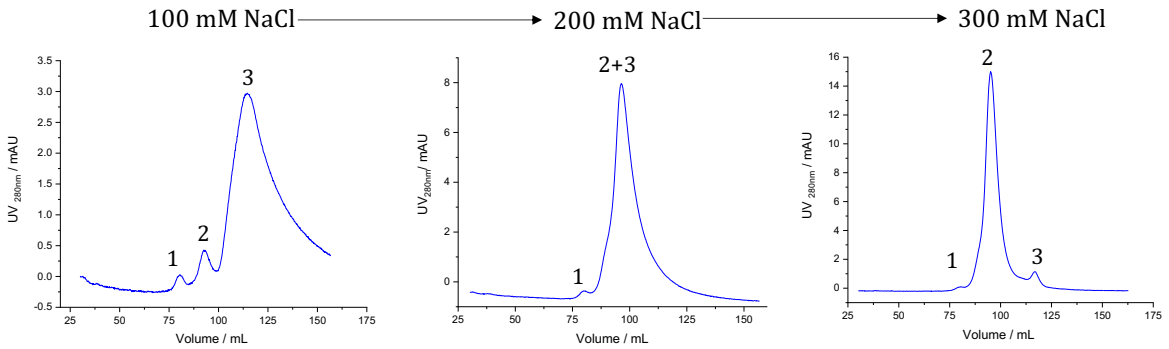


Figure 4.22: (a) Effect of increasing salt concentration on the elution profile of recombinant GBA from S200 Superdex column.

Intact Mass Spectrometry

Following SEC-MALLS analysis, intact mass spectrometry was performed to provide some insight into the glycosylation profile of the protein. The resulting mass spectrum revealed a predominant glycoform at 59392.83 Da and a series of glycoforms varying by mannose and fucose units, Figure 4.23. Given this protein was expressed in insect cells, the presence of core fucosylation was expected and evidence of this in the mass spectrum further supports the previous observation that PNGase F is unable to remove the *N*-glycans from this formulation. Additionally, the large number of glycoforms was unsurprising given that the protein was expressed in insect cells with no attempts to control post-translational glycosylation.

For therapeutic glycoproteins, glycosylation is vital to ensuring optimum biological activity, therapeutic efficacy and tolerability. Therefore, selection of expression host, glycoengineering and upstream processes are extensively optimised to control the glycosylation of therapeutic proteins[719]. However, to the authors knowledge, no insect cell made GBA has been approved for therapeutic use to date because glycoforms produced by insect cells can be immunogenic[810]. Of note, Hi5 cells modify glycoproteins with core fucosylation, as observed in the intact mass spectrum of the protein produced in this work, which can induce immunogenic and allergic

responses in humans[784]. Several strategies have been developed to address the issue of incompatible *N*-glycosylation in insect cell expression systems, however, these remain in their infancy[775]. For the purposes of this work, the glycosylation of GBA need not be homogeneous, however, there were concerns that the presence of multiple glycoforms would hinder crystallisation and subsequent structural studies. It should also be noted that this GBA formulation is not suitable for human administration due to the variable glycosylation profile and insect-cell like glycosylation.

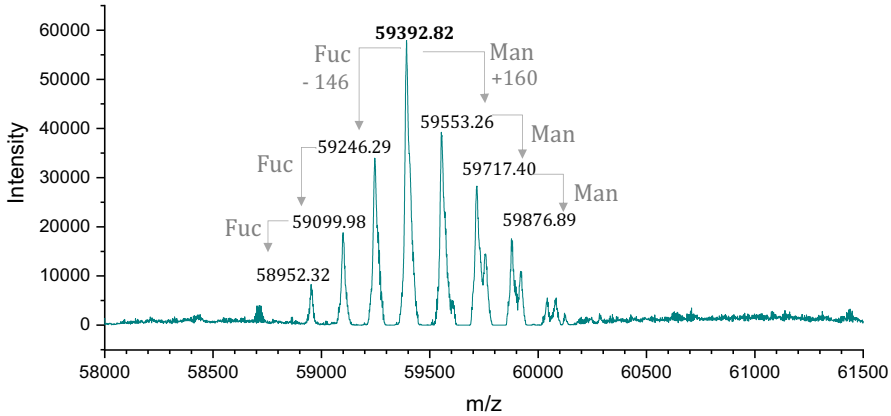


Figure 4.23: Intact mass spectrum of non-tagged GBA protein, revealing the presence of multiple GBA glycoforms varying by fucose and mannose units.

Activity-Based Labelling

Time dependent intact MS analysis was also used to further investigate the labelling of recombinant GBA by a Cy5-tagged cyclophellitol epoxide ABP (ME569) and a Cy5-tagged cyclophellitol aziridine ABP (JJB367). GBA was incubated at optimum pH (pH 5.2) with each ABP and aliquots were taken at 1-min and 5-min for analysis by intact MS. The resulting spectra demonstrated a shift in the MS profile of GBA by 848 Da or 721 Da, corresponding to covalent labelling by JJB367 and ME569 respectively, Figure 4.24. Within 1 minute, approximately 80% of GBA was labelled by the aziridine probe JJB367, compared to 67% labelling by the epoxide probe ME569. This exemplifies the enhanced labelling efficiency of cyclophellitol aziridines compared to the epoxide equivalents. Nevertheless, labelling by both probes reached completion within 5 minutes. These labelling assays further confirm the catalytic activity of this GBA formulation and demonstrate how such mass spectrometry studies may be used to probe inactivator binding.

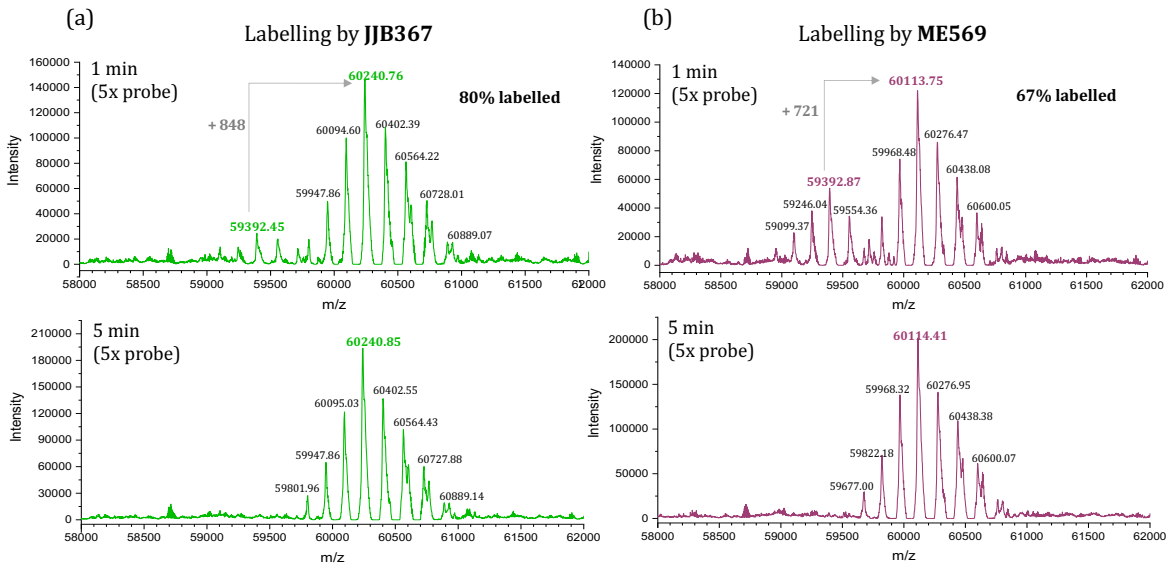


Figure 4.24: Intact mass spectra of GBA labelling with ABP JJB367 and ME569. A shift in the GBA profile of 848 Da and 721 Da is observed, corresponding to labelling by JJB367 and ME569 respectively. Both labelling reactions reached completion within 5 minutes.

4.4.4.4 Crystallisation

Following biochemical characterisation, the tagless GBA formulation was tested against a range of commercial crystallisation screens, with an initial hit identified in well H8 of the PACT premier HT-96 from Molecular Dimensions[794] containing 0.2 M sodium sulfate, 20% (w/v) PEG 3350, 0.1 M bis-Tris propane (BTP) pH 8.5. However, this pH was considered too high for GBA, which is most stable at acidic pH (< 6)[93]. Additionally, this pH is at the upper limit of which the Overkleeft cyclophellitol probes function[93]. Therefore, optimisation of crystallisation components was performed to generate crystals at a lower pH (pH 7.0), Figure 4.25.

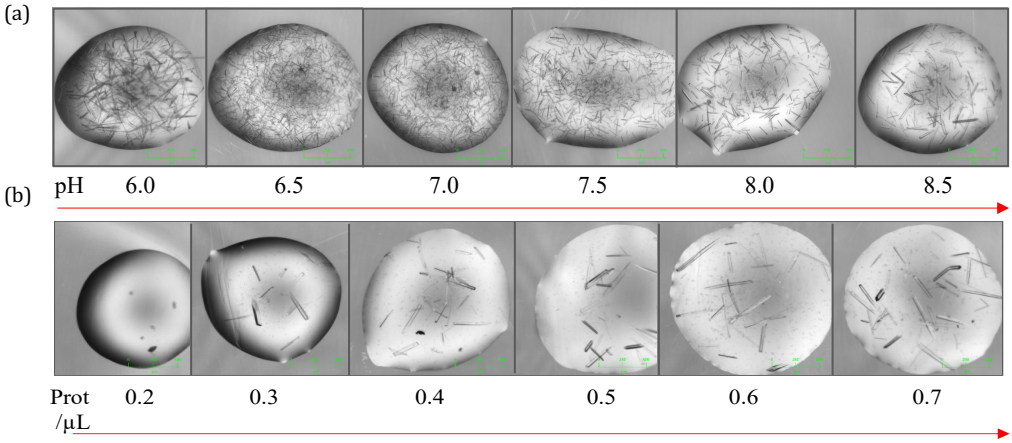


Figure 4.25: Optimisation of GBA crystallisation conditions.(a) altering pH of bis-Tris-propane buffer and (b) varying protein concentration to generate suitable crystals at pH 7.0.

rGBA in Complex with Bis-Tris-Propane

Through alteration of buffer pH, crystals of rGBA were obtained at pH 7.0 and tested at the Diamond Light Source. The resulting data were processed to 1.56 Å resolution to reveal that GBA had crystallised in spacegroup $P2_1$ with two molecules in the asymmetric unit, Figure 4.26 (a). This homodimeric crystal form has been reported previously for GBA[595,657,660], however, in contrast to earlier studies[595,641,654], this protein formulation was not deglycosylated prior to crystal screening. Consequently, this structure exhibited visible *N*-glycosylation at Asn19, Asn59 and Asn146 in chain A and Asn19 and Asn146 in chain B, Figure 4.26 (a). It is understood that occupation of the Asn19 *N*-glycosylation site is vital for GBA activity[647,648] and in this structure, Asn19 is occupied by a chitobiose core with a β-1,4-mannose unit in chain B and an additional α-1,3-mannose unit in chain A, Figure 4.26 (a). The absence of core fucosylation in the glycan profile was surprising, especially as it was observed so prominently in the intact MS spectrum. However, it is possible that core fucosylation is present but is not observable in the electron density or that this specific non-fucosylated glycoform crystallises more readily.

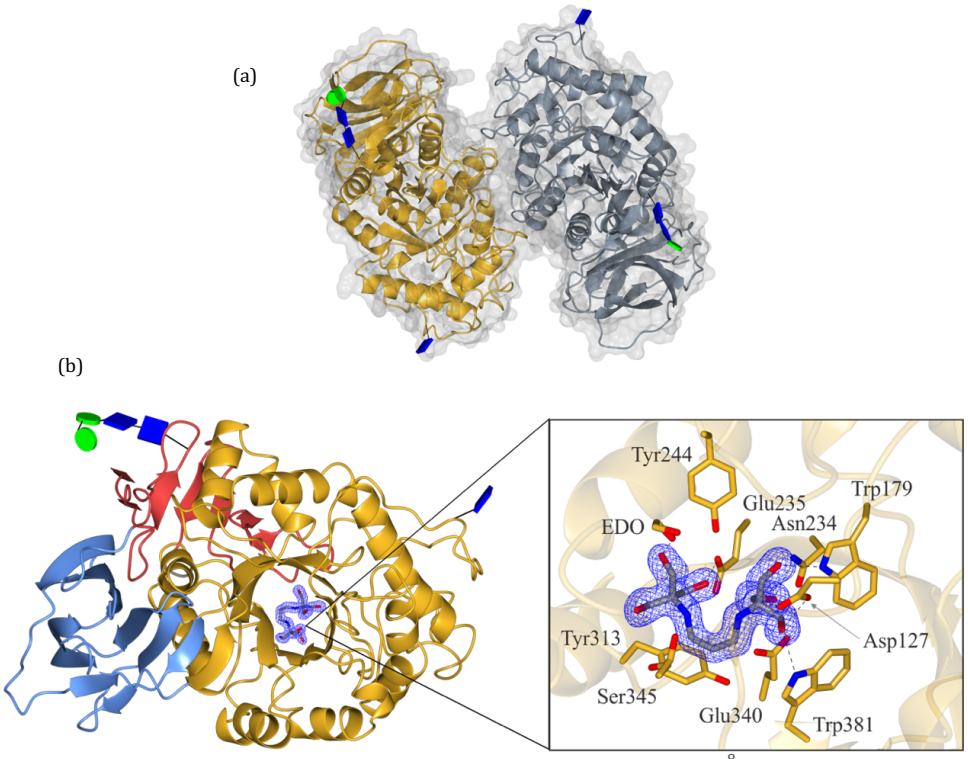


Figure 4.26: (a) Crystal structure of rGBA dimer obtained at 1.56 Å resolution (PDB 6TJK). *N*-glycans depicted in glycoblock format[591]. (b) GBA monomer comprises of 3 domains; domain I (residues 1-27 and 383-414) in red, domain II (residues 30-75 and 431-497) in blue and domain III (residues 76–381 and 416-430) in gold. Zoom panel shows active site structure with BTP bound non covalently. Electron density ($2F_o$-F_c) of BTP contoured to 1σ (0.34 e/Å^3).

Each GBA monomer in the asymmetric unit comprised of the expected three non-contiguous domains: an anti-parallel β-sheet domain (domain I), an immunoglobulin like domain (domain II) and a TIM barrel domain containing the active site (domain III), Figure 4.26 (b). Importantly, the tertiary structure of this recombinant protein compares favourably with an unliganded structure of Cerezyme® obtained for comparison, Figure 4.27. An overlay of the two structures using the online PDBeFold tool revealed a Cα root mean square deviation (RMSD) of 0.57 Å (Q-score 0.95) and 0.50 Å (Q-score 0.96) for overlay of the A chains and B chains respectively, indicating good structural similarity of the protein back bone. Some deviations were observed in the flexible loop regions of residues 27-30, 60-64, 313-319 and 395-398. However, these loops have been observed in multiple conformations in previously published crystal structures of GBA, suggesting these regions exhibit considerable dynamic flexibility. Of note, loop 1 containing residues 313-319 has been observed in an extended (open) and helical (closed) conformation[811]. Specifically, the helical (closed) conformation is commonly seen when a substrate is bound in the active site, and this change in loop conformation from extended to helical is believed to play a role in controlling active site access by inducing changes in nearby hydrophobic regions of the protein[812]. It is possible that the closed (helical) conformation of loop 1 observed in this rGBA structure may result from non-covalent binding of a molecule of bis-Tris-propane BTP, from the crystallisation conditions, in the active site.

Unfortunately, a true ligand-free structure was not obtained due to the presence of BTP in the active site, Figure 4.26 (b). Binding of BTP to glycosidases has been observed previously and results from a superficial similarity between the hydroxylated and positively charged BTP molecule and the oxocarbenium ion transition state of glycoside hydrolysis, which is strongly stabilized by glycosidase enzymes[795,813,814]. Nevertheless, when aligned with Cerezyme®, it is clear that the active site of rGBA is extremely well conserved, with most active site residues, including the catalytic acid-base residue (Glu235) and catalytic nucleophile (Glu340), adopting almost identical conformations, Figure 4.27 (b). The only notable difference is the 'downwards' displacement of Tyr313 in the BTP complex, presumably to avoid clashing with the hydroxyl groups of the BTP molecule.

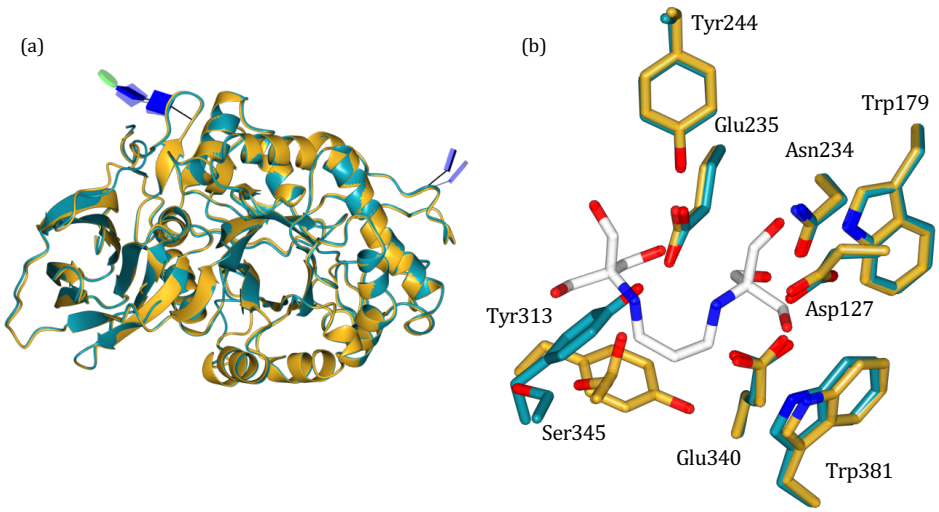


Figure 4.27: (a) Tertiary structure and (b) active site overlay of recombinant GBA (PDB 6TJK) obtained at pH 7.0 (gold) and Cerezyme® (PDB 6TJJ) obtained at pH 4.6 (teal). A molecule of BTP (white) occupies the active site of PDB 6TJK.

Although good tertiary and active site structural similarity was observed with Cerezyme®, there were concerns that binding of BTP to the active site would interfere with ligand soaking experiments. Unfortunately, this was confirmed to be the case when attempts to use these crystals for ligand binding studies, by soaking with other ligands to displace BTP, proved unsuccessful. The inability to displace BTP from the active site can be rationalised by the high concentration of BTP used in the crystallisation conditions (100 mM) and its comparatively potent IC_{50} (4.31 $\pm$ 0.42 mM against 4-MU-Glc), Figure 4.28. Consequently, crystallisation conditions without BTP were required to generate crystals suitable for ligand binding studies.

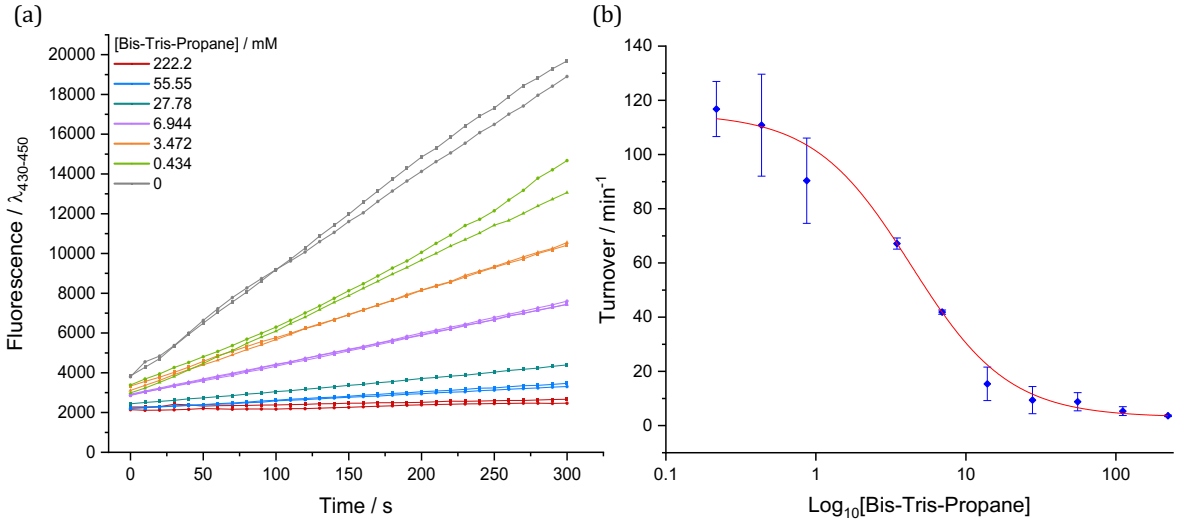


Figure 4.28: (a) Raw fluorescence data for the hydrolysis of 4-MU-Glc by GBA in the presence of varying concentration of BTP. (b) A plot of the rate of 4-MU-Glc hydrolysis vs BTP concentration fitted to a 4-parameter logistic function to estimate IC_{50} of bis-tris propane.

Removal of Bis-Tris-Propane

In an effort to remove BTP from the crystallisation conditions, screening was performed in which BTP was substituted with HEPES (pH 7.0 and 7.5) or PIPES (pH 6.5 and 7.0) buffers. Unfortunately, these primary screens failed to produce any crystals, Figure 4.29. Therefore, a seed stock was generated from crystals obtained under BTP conditions and used for microseeding into HEPES containing conditions. Following optimisation, microcrystals were formed and used to prepare a second seed stock without BTP. Extensive optimisation was performed to generate crystals suitable for structural studies using 0.2 M Na_2SO_4, 14% (v/v) PEG3350, 0.25 M HEPES (pH 7.0) and 0.1 µL seed solution (1:1000), Figure 4.29. Given these crystals were obtained in the absence of BTP they were assumed to be unliganded and therefore suitable for ligand binding studies.

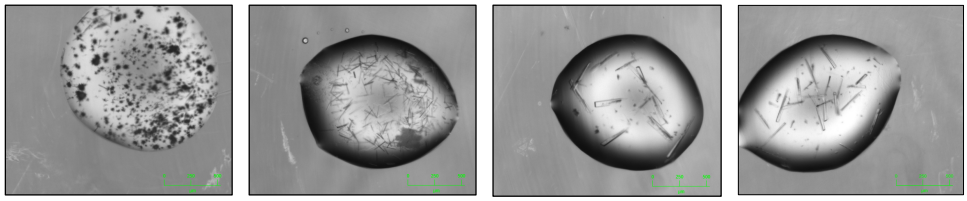


Figure 4.29: Removal of BTP from crystallisation conditions by (a) substitution of BTP with HEPES buffer (b) use of BTP crystals for microseeding (c-d) use of HEPES microcrystals for further optimisation of HEPES conditions.

Covalent Complex with 2F-DNPGlc (14)

2,4-Dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside (2F-DNPGlc, **14**) is a well characterised mechanism-based inactivator which functions by trapping retaining glucosidases in a covalent fluoroglycosyl-enzyme complex. Specifically, substitution of the C2 hydroxyl group with an electronegative fluorine atom destabilises both transition states to enzyme active site glycosylation and deglycosylation[607,610]. However, the addition of a reactive DNP leaving group increases the rate of glycosylation, allowing a trapped enzyme-inhibitor complex to accumulate following reaction with the enzyme[607], Figure 4.30. Despite this being a mechanistically important inhibitor, there have been no reported crystal structures of GBA in complex with this inhibitor. Consequently, this work aimed to generate a co-crystal structure of GBA in complex with 2F-DNPGlc to demonstrate the suitability of these crystals for ligand binding studies.

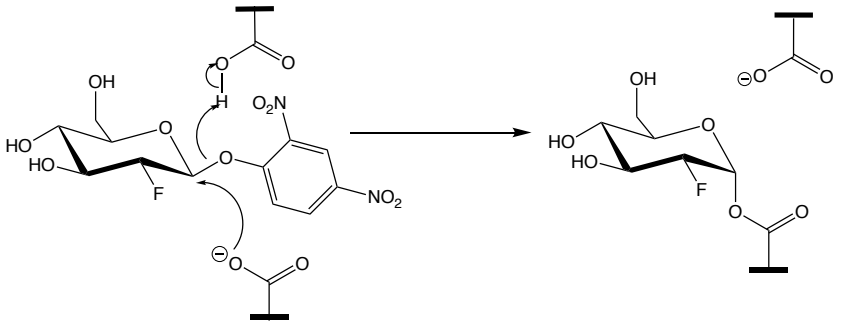


Figure 4.30: Mechanism of 2F-DNPGlc hydrolysis by retaining β-glucosidase to generate a trapped covalent glycosyl-enzyme intermediate.

A structure of the 2-deoxy-2-fluoroglucopyranosyl-GBA intermediate was obtained by soaking unliganded GBA crystals overnight in 2F-DNPGlc. The data were solved to 1.41 Å resolution to reveal covalent binding of the 2-deoxy-2-fluoro glucose moiety to the catalytic nucleophile by cleavage of the 2,4-DNP leaving group, Figure 4.31. The covalent bond length was measured to be 1.42 Å, with the gluco-configured ring adopting the $^4C_1$ chair conformation. This is consistent with the covalent intermediate of the conformational reaction itinerary of GBA. Additionally, the glycosyl moiety forms hydrogen bonds with a number of active site residues, including Trp179, Asp127, Trp179, Asn234, Glu340, Trp381 and Asn396. Interestingly, two conformations of the catalytic nucleophile could be modelled, Figure 4.31. It is proposed that electrostatic repulsion between the carboxylate of the catalytic nucleophile and the C2 fluorine atom of the 2F-Glc inactivator, enforces a 28° rotation about Cγ of the nucleophilic residue, resulting in movement of the O1 atom of the carboxylate residue away from the C2 linked fluorine atom.

Aside from providing a novel structure in complex with a mechanistically relevant glucosidase inhibitor, this complex demonstrates the potential of our GBA formulation to be used as an alternative to ERT preparations for ligand binding studies. Importantly, these crystals have permitted continued collaboration with the Overkleeft lab in regard to the development of ABPs GBA. For example, these GBA crystals were used to structurally evaluate the bifunctional Cy5-functionalised cyclophellitol aziridine ABP discussed in **Chapter 3**, further highlighting the potential of this GBA formulation to support the development of novel GBA active compounds.

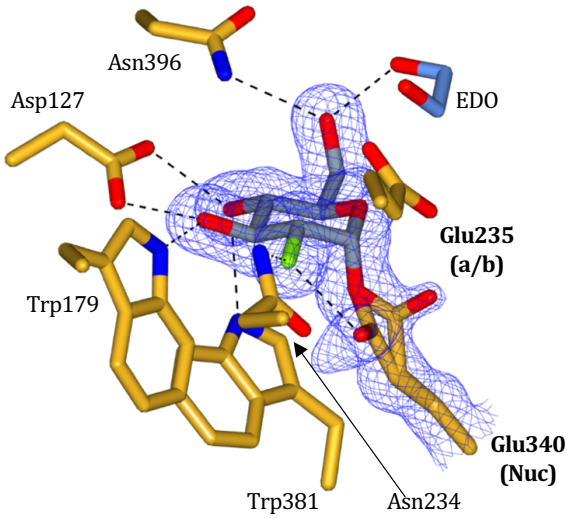


Figure 4.31: Active site structure of the 2-deoxy-2-fluoro-β-D-glucopyranoside-GBA covalent intermediate (PDB 6TJQ). The 2F-Glc moiety is covalently bound to Glu340 in the $^4C_1$ chair conformation. The catalytic nucleophile adopts two conformations. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Glu340 contoured to 1.1 σ (0.40 e/Å^3).

Atomic Resolution Ligand Free Structure

Given the tight binding of BTP in the originally identified GBA crystal form, optimisation of alternative, non-BTP containing crystallisation conditions was performed in parallel with the efforts to remove BTP by microseeding. During initial screens, a hit was identified in the JCSG$^+$ screen[794] and following optimisation, crystals suitable for structural analysis were generated using 0.2 M magnesium formate and 19% (v/v) PEG3350. These crystals diffracted well at the Diamond Light Source facility, yielding a 0.98 Å unliganded structure of GBA (PDB 6TN1), Figure 4.32 (a). Not only is this the highest resolution structure of GBA deposited to date, it also exists in a previously unreported crystal form. GBA has been crystallised in C 222_1 and P 2_1 space groups, however this unliganded structure crystallises in spacegroup P1. This new crystal form contains one GBA molecule in the asymmetric unit which comprises of the expected three non-contiguous domains, with N-glycosylation at Asn19 and Asn146, Figure 4.32 (a).

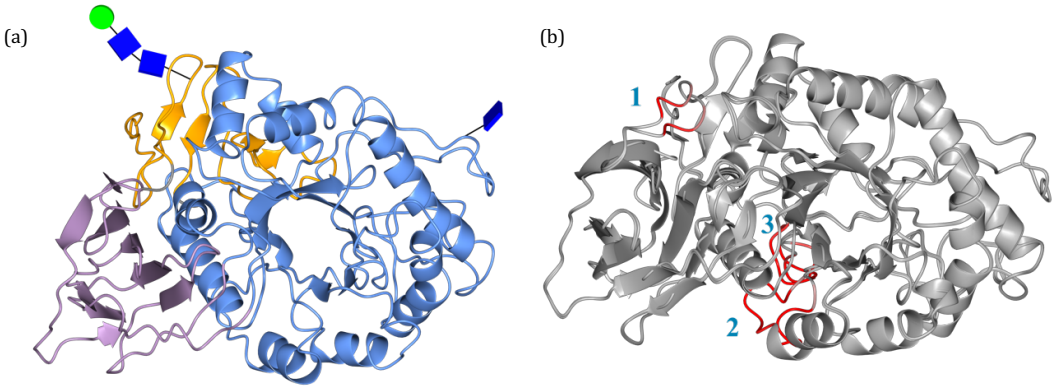


Figure 4.32: (a) Crystal structure of GBA obtained at 0.98 Å resolution. Domain I (residues 1–27 and 383–414) = immunoglobulin-like domain in lilac, domain II (residues 30–75 and 431–497) = anti-parallel β-sheet domain in orange and domain III (residues 76–381 and 416–430) = $(\beta/\alpha)_8$ TIM barrel in blue. (b) Overlay of the unliganded GBA structure with the BTP-complexed structure. Red indicates areas of high r.m.s.d. between the protein backbone: 1 = residues 27–31, 2 = 314–319 and 3 = 344–350.

Overall, the three-domain tertiary structure is highly similar to that of the BTP co-crystal and Cerezyme® structures, with a Cα RMSD of 0.49 Å (Q-score 0.94) and 0.60 Å (Q-score 0.94) respectively. However, some deviations in the protein backbone were observed in the three flexible loop regions of residues 26-31, 314-319 and 344-350, Figure 4.32 (b). Despite the sub-Ångström resolution, residues 26-31 and 314-319 were challenging to model, further reflecting the flexibility and disorder of these loops. Specifically, in contrast to the BTP complexed structure, loop 1 containing residues 311-319 is observed in the extended (open) conformation in this unliganded structure, Figure 4.33, which is consistent with reports in literature that loop 1 changes conformation from extended (open) to helical (closed) on ligand binding[812]. This also suggests these crystals will be suitable for ligand binding studies as the extended loop conformation ensures the active site remains accessible. Importantly, the active site of this unliganded structure compares well with the active site of Cerezyme® and the BTP complexed GBA structures, Figure 4.34. In fact, most active site residues occupy identical conformations, with the exception of Tyr313 which restores its "upwards" conformation in the absence of BTP. This Tyr313 residue appears particularly mobile, occupying a different conformation in each GBA structure, Figure 4.34.

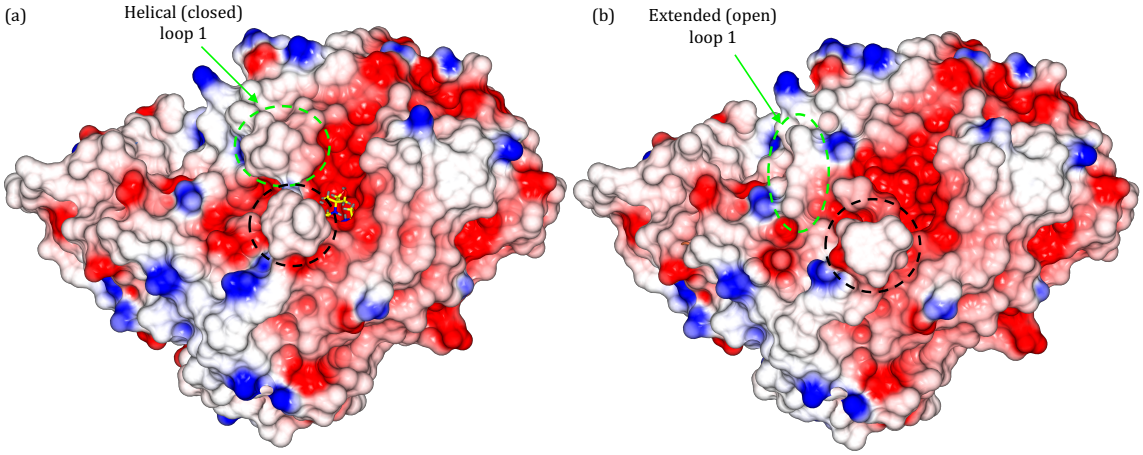


Figure 4.33: Electrostatic potential surface of (a) BTP complexed GBA structure in which loop 1 (green) is observed in the helical (closed) conformation and (b) GBA apo structure in which loop 1 is observed in the extended (open) conformation. Changes in loop 1 also cause shifts in the conformation of loop 2 (black).

Although numerous unliganded GBA structures have been solved previously, this sub-Ångström resolution structure permits the first ever atomic resolution analysis of GBA, revealing finer details in its structure. For example, two conformations of the catalytic acid-base residue could be modelled, Figure 4.34 (b). In fact, many alternative side chain conformations could be observed throughout the structure, providing more detail on side chain mobility and interactions. Furthermore, electron delocalisation over carbonyl groups and double bonds could be readily observed, as well as proton positions in the difference electron density map, Figure 4.34 (c). It is anticipated that this new crystal form will be utilised in structural studies to provide atomic resolution analysis of ligand-binding interactions with GBA.

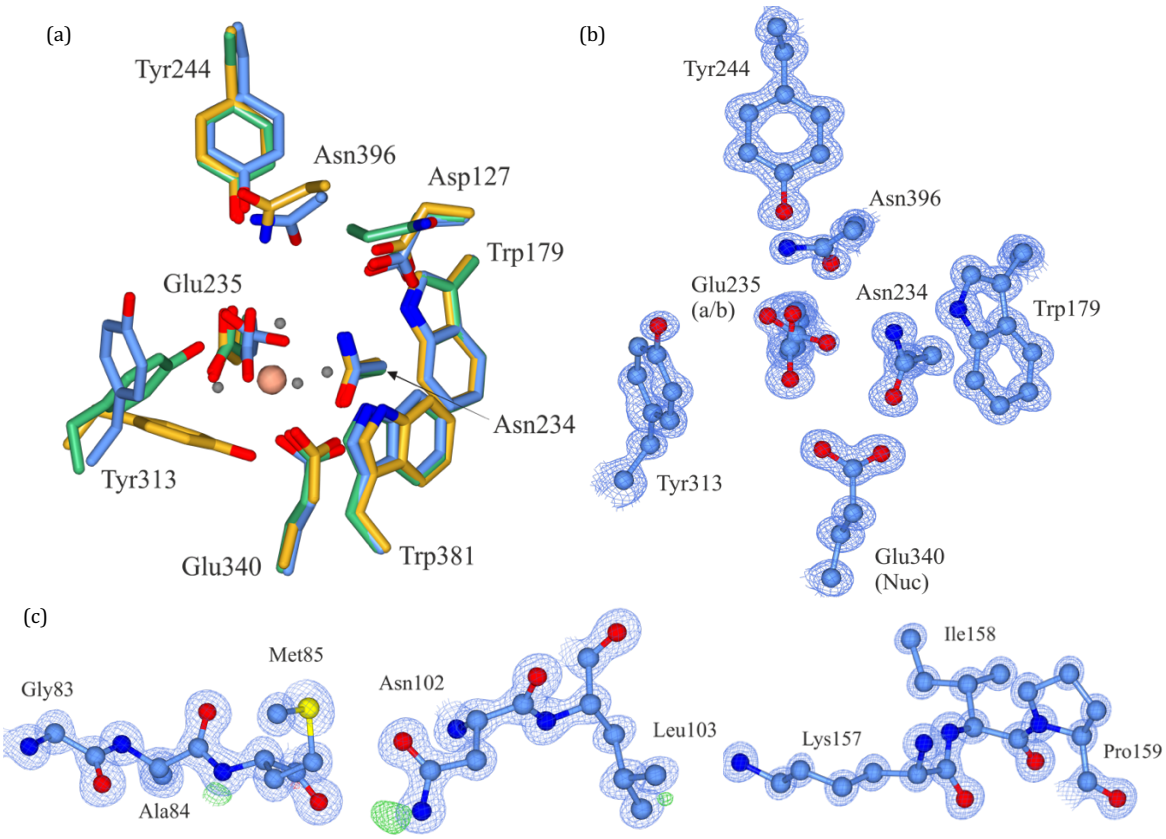


Figure 4.34: (a) Active site structure of unliganded GBA (blue) overlaid with active-site structure of the BTP-complex structure (gold) and Cerezyme® (green). Magnesium ion (peach) coordinated by 4 waters occupies the active site. (b) Active site structure of unliganded GBA with observed electron density for each active site residue. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for selected residues contoured to 2 σ (1.0 e/Å^3). (c) Selection of modelled residues with difference electron density in green (contoured to 3 σ (0.37 e/Å^3) highlighting proton positions.

4.5 Summary

In this work, the development of an insect-baculovirus expression system for the production of non-clinical, human GBA is described. Following investigations into numerous unsuccessful His_6-tagged constructs, a tagless formulation was found to be suitable for expression, purification and crystallisation. Importantly, this GBA formulation exhibits comparable biochemical and biophysical properties to commercial therapeutic products (Cerezyme®), demonstrating good activity against the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside and reasonable long-term stability. This GBA formulation also crystallises in multiple crystal forms with high resolution. Of note, these crystals have proved suitable for ligand-binding studies and were used to investigate several GBA active compounds, including the glucosidase inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside and a bi-functional fluorescently tagged activity-based probe (Overkleeft Lab, Leiden University). These structures demonstrate the utility of this GBA formulation in ligand binding studies for the analysis and development of novel GBA active compounds. Furthermore, a novel crystal form of GBA was obtained in this work, which diffracted to give a 0.98 Å resolution unliganded structure. Not only is this the highest resolution GBA structure deposited to date, it also permitted the first ever atomic resolution analysis of GBA, revealing finer details in its structure.

In light of its purity, stability and activity, the GBA production protocol described herein provides an alternative source of active, non-clinical GBA. Consequently, this formulation should alleviate reliance on ERT preparations for biochemical and structural studies and support the development of novel GBA active compounds for Gaucher Disease research.

Chapter 5: Structural Analysis of Novel β-Glucocerebrosidase Pharmacological Chaperones

5.1 Abstract

The role of defective β-glucocerebrosidase (GBA) activity in the pathology Gaucher Disease (GD), and the more recent association with Parkinson's disease, has resulted in considerable interest in the development of molecular chaperones for GBA. Unfortunately, traditional, competitive chaperones have struggled to progress from positive pre-clinical results to real-life patient benefit. Consequently, the Vocadlo and Bennet laboratories (Simon Fraser University, SFU) have developed a powerful new approach to the design of single turn-over, allylic carbasugar inhibitors with pharmacological chaperone potential. Herein, the structural analysis of these carbasugar inhibitors on the 3D structure of human GBA is described.

Using the recombinant GBA formulation produced in **Chapter 4**, co-crystal structures of rGBA in complex with a number of chloro- (SRK049, SRK121) and fluoro-substituted (SBB006) allylic carbasugar inhibitors were obtained at 1.49-1.59 Å resolution. These co-crystal structures revealed two distinct inhibition mechanisms in which the inhibitors covalently modify the catalytic nucleophile of GBA to yield enzyme-inhibitor complexes in the envelope conformation. Specifically, SRK049 was shown to react by nucleophilic attack at C1 (the carbon centre at which the leaving group is attached) and direct cleavage of the C—Cl bond, whilst its enantiomer SRK121 reacted through the endocyclic double bond, resulting in an allylic rearrangement and subsequent release of Cl^-. Whilst a covalent complex of the fluorine derivative SBB006 could not be obtained, the co-crystal complex reported here provides insight into the non-covalent binding of these inhibitors in the active site, potentially describing the Michaels complex. Additionally, this work also uncovered a seemingly enantiospecific allosteric binding site in the immunoglobulin-like domain of GBA which binds SRK121 and its fluoride equivalent SBB006 through hydrogen bonding interactions. Indeed, binding to this site may provide a structural rationale for the improved ability of SRK121 to chaperone mutant L44P/P415R GBA over its enantiomer SRK049. It is hoped that the structural and mechanistic information generated in this work will enable a structure-guided approach to improving the pharmacological chaperone properties of such inhibitors.

5.2 Introduction

5.2.1 Protein Misfolding: Implications for Lysosomal Storage Disorders

5.2.1.1 Endoplasmic Reticulum-Associated Protein Degradation Pathway

Protein folding is an essential, complex process which in the case of secretory and membrane proteins occurs in the endoplasmic reticulum (ER)[815]. Unfortunately, folding of proteins is commonly error-prone, with factors such as genetic mutations, heat shock and oxidative stress leading to improper folding[815]. Moreover, accumulation of unfolded or misfolded proteins is often cytotoxic and can lead to a range of protein misfolding diseases, namely Parkinson's Disease (PD) and Alzheimer's[816,817]. Fortunately, the ER is equipped with a protein quality control system which has evolved to manage protein misfolding events and enhance the fraction of correctly folded proteins[501,815]. Specifically, this quality control system is able to distinguish between correctly folded proteins, which are subsequently transported to their required destination, and misfolded proteins which are flagged by the quality control system and retained within the ER for refolding or degradation[501]. This system is underpinned by the ER-associated protein degradation (ERAD) pathway, which selectively removes improperly folded proteins[47,815]. In this pathway, proteins to be degraded are transported to the ER cytosol where they are ubiquitinated and degraded by the proteasome[47]. In addition, the accumulation of misfolded proteins in the ER activates the unfolded protein response, which induces expression of molecular chaperones and other ERAD components to enhance the folding capacity of the ER[818,819].

5.2.1.2 Implications for Gaucher Disease

As highlighted previously in **Chapter 1**, lysosomal glycoside hydrolases (GHs) are biosynthesised in the ER and following correct folding are transported to the lysosome where they catalyse the hydrolysis of various glycolipid substrates. However, recessively inherited mutations in the genes encoding for these lysosomal enzymes can cause misfolding, which ultimately results in reduced enzymatic activity in the lysosome, substrate accumulation and a lysosomal storage disorder (LSD)[151,160,820]. Depending on the mutation, some patients may exhibit sufficient residual enzymatic activity to negate severe disease, but most patients have inadequate activity to process the substrate which accumulates with negative implications for

a range of downstream cellular processes. For example, patients with homozygous loss-of-function mutations in the *GBA1* gene exhibit significantly reduced β-glucocerebrosidase (GBA) activity, which is required for the degradation of glucosylceramide (GlcCer) and glucosylsphingosine (GlcSph)[172,173]. Therefore, in the absence of adequate GBA activity, GlcCer and GlcSph accumulate in cells and organs throughout the body, leading to Gaucher disease (GD)[168]. Furthermore, heterozygotes carrying just one mutant *GBA1* allele have been found to have a significantly increased life-time risk of developing PD with potentially more severe symptoms[250,821,822]. Whilst the pathology of PD is primarily characterized by abnormal accumulation of α-synuclein protein in the brain and consequent loss of dopaminergic neurons, studies have shown a correlation between reduced GBA activity and increased α-synuclein aggregates (as discussed in detail in **Chapter 1** (*section 1.4.4*))[243,246,823]. Consequently, therapeutic approaches to enhance GBA activity are of considerable pharmaceutical interest for both GD and PD. Indeed, it has been shown that wild-type GBA can be chaperoned in the brain and central nervous system to increase lysosomal GBA activity[824]. Moreover, overexpression of GBA in the brain has been shown to hinder α-synuclein aggregation, which may hold promise for PD therapy[825,826].

5.2.2 Therapeutic Approaches

Over the past 3 decades, therapeutic strategies for GD and other LSDs have advanced considerably, with the development of enzyme replacement therapy (ERT)[652], substrate reduction therapy (SRT)[356] and pharmacological chaperone therapy (PCT)[827] (see **Chapter 1** *section 1.4.8* for more details). Whilst ERT is the most common therapeutic approach, each of these strategies can be preferentially used depending on the genotype and phenotype of the patient[338]. For example, ERT is the most widely used therapy which is very effective at treating visceral and skeletal manifestations of GD type 1 by enhancing the amount of active lysosomal GBA[312,331,828]. However, the recombinant enzyme administered during treatment is unable to cross the blood brain barrier, therefore, ERT is ineffective at treating neurological disease associated with type 2 and 3 GD[334,335]. Some patients also develop anti-protein antibodies to the recombinant enzyme, which can drastically reduce the efficacy of ERT[310,820]. SRT takes an entirely different therapeutic approach in which small molecular inhibitors are administered to prevent the initial biosynthesis of glycolipid substrates and stop their accumulation[336]. SRT holds a major advantage over ERT in that it can be administered orally without concerns

regarding anti-protein antibodies[354]. However, glycolipids are critical components of the cell membrane and are required for normal brain function[829]. Therefore, complete abrogation of glycolipid synthesis may cause side effects in itself. Additionally, many of the inhibitors used for SRT, for example Eliglustat, are unable to cross the blood brain barrier and are therefore unsuitable for the treatment of neurological symptoms[337,352]. In contrast, small pharmacological chaperones used in PCT are typically more amenable to neurological disease as they are able to cross the blood brain barrier and augment GBA activity in the brain[827]. Indeed, the ability of pharmacological chaperones to treat neurological manifestations is a major advantage of PCT.

All therapeutic approaches for GD are incredibly costly, for example, the therapeutic ERT formulation Cerezyme® costs approximately ~$200,000 per patient per year[332], whilst the substrate reduction inhibitor Eliglustat costs between $250,000–500,000 per year[830]. Given the potential therapeutic benefits, especially for PD for which no disease modifying therapy is known, there is a compelling need for novel, efficacious treatments for addressing defective GBA activity. Moreover, therapies targeting GBA may also open the door to treating other LSDs through a similar approach.

5.2.2.1 Pharmacological Chaperone Therapy

The academic and pharmaceutical interest in PCT for GD has grown considerably over the last 2 decades, especially as the pathological consequences of defective GBA activity have been linked to other disorder such as Parkinson's. Typically, PCT involves administering a small molecule (pharmacological chaperone, PC) that binds to the enzyme of interest as it is being synthesized in the ER to lower the free energy of folding, encourage the formation of correctly folded protein and enhance the amount of active enzyme transported from the ER to the desired destination[827]. In this regard, several small molecule PCs, such as Isofagomine, have been developed and tested for GBA in various cell models with some promising results[365,367,376,595]. However, translation of these compounds into clinical therapy has been incredibly problematic. For example, whilst Isofagomine was found to have broad tissue distribution with access to the central nervous system, it was eventually dropped in phase II clinical trials when it failed to meet clinical requirements[369]. This was partially attributed to its broad specificity, which resulted in off target inhibition of other GHs[370]. Another common problem is that PCs are typically weakly basic, polar molecules which tend to accumulate in the lysosome[831]. Not only can this result in low clearance from the lysosome but also prolonged target engagement,

meaning these molecules remain associated with the target enzyme and inhibits its activity even once the enzyme has reached the lysosome[831], Figure 5.1 (a). This behaviour is a major limiting factor for PCT and is arguably the reason that this approach commonly fails to translate into real clinical benefit. Therefore, there remains considerable demand for PCs which efficiently clear the lysosome and whose mode of action is not limited to cell and animal models.

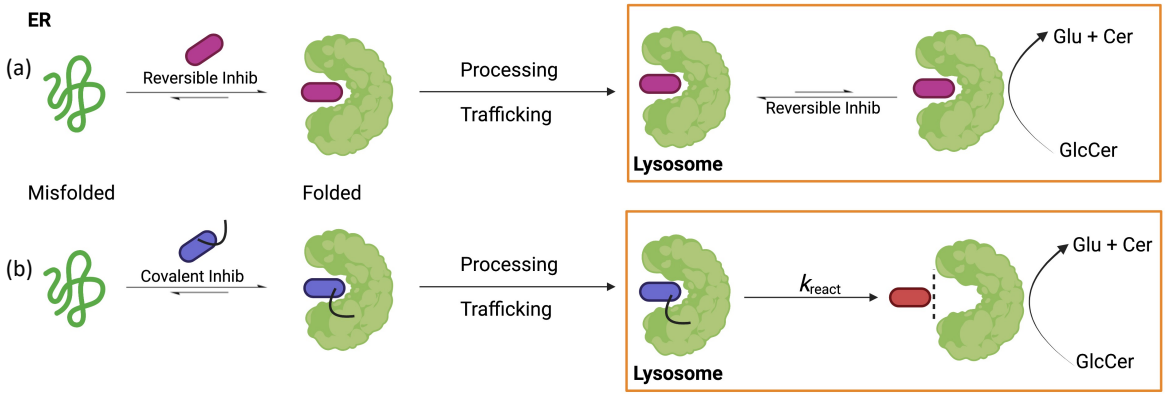


Figure 5.1: Chaperone assisted folding of GBA. (a) Reversible, non-covalent chaperone assists folding of GBA in endoplasmic reticulum (ER) to enhance proportion of correctly folded GBA in lysosome. However, the PC continues to bind GBA in the lysosome, reducing the attainable activity. (b) Reversible, covalent inhibitor binds to GBA and enhances trafficking to the lysosome. The inhibitor is then hydrolysed to release free enzyme and an inhibitor product which can no longer bind GBA, resulting in increased activity in the lysosome.

5.2.2.2 Addressing the Problems of Pharmacological Chaperones

The Vocadlo and Bennet laboratories (Simon Fraser University, SFU) have suggested that the failure to progress from positive results in cell and animal models to real clinical benefit may result from limitations early in the experimental protocols of PC development. Specifically, when PCs are under investigation, the cells and tissues are typically lysed before the chaperoned activity is measured[832,833]. During lysis, the cellular contents, including the inhibitor and enzyme, are diluted which means a greater fraction of the enzyme is unbound relative to the amount in the lysosome. Additionally, weakly basic PC inhibitors are retained within the lysosome where they continue to inhibit the enzyme of interest, Figure 5.1 (a); however, this effect is generally not observed due to the dilution effects. Therefore, it is often challenging to determine the extent to which the PCs are retained within the lysosome and their prolonged target engagement. This is particularly true for traditional tight-binding, competitive inhibitors which target GHs throughout the complete cycle of folding, processing, trafficking and

lysosomal residency. In this regard, it has been shown that GBA takes ~8 hours to be trafficked to lysosomes where it has a $t_{1/2}$ of between 24 and 48 hours[834]. Thus, an increase in the amount of correctly folded protein in lysosomes does not necessarily result in an increase in lysosomal activity if the PC does not dissociate efficiently, Figure 5.1. Consequently, researchers at the Vocadlo and Bennet labs have developed a powerful new approach to the design and evaluation of selective, single turn over PCs which capitalize on enzyme transition state mimicry.

Using various generation of their innovative fluorescence-quenched substrates[700,835], the Vocadlo group have pioneered a strategy to measure lysosomal GBA activity directly within live cells, circumventing the dilution effects associated with cell lysis measurements. Indeed, a glucoside equipped with a suitable fluorophore and quencher pair permits the generation of substrates that allow convenient time-dependent monitoring of endogenous GBA activity within cells[700,835], Figure 5.2. Moreover, these quenched fluorescent substrates facilitate the measurement of GBA inhibition and chaperoned lysosomal activity in live cells through confocal microscopy or high content imaging, allowing the lysosomal residency of PCs to be more reliable evaluated.

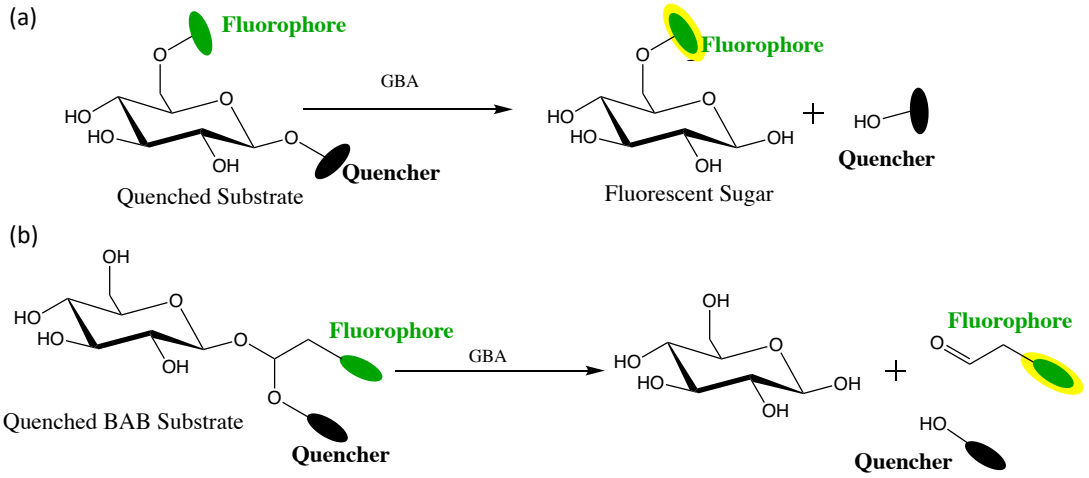


Figure 5.2: General reaction scheme showing the design of the Vocadlo fluorescence-quenched substrates (a) Fluorescence-quenched glucosidase substrate where a fluorescent signal is only produced upon substrate hydrolysis to yield the fluorescent glucoside and (b) 'bis-acetal based' (BAB) fluorescence-quenched substrates which are hydrolysed to yield a hemiacetal that rapidly gives rise to an aldehyde that diffuses from the quencher to give a fluorescence signal.

5.2.3 Glycoside Hydrolase Inhibitors

2.1.1.1 Novel Allylic Carbasugar Chaperones

Using their innovative fluorescence-quenched substrates[700], the Vocadlo and Bennet labs have developed a novel class of carbasugar PCs, Figure 5.3, which in contrast to traditional, tight-binding competitive inhibitors, covalently modify GBA in a transient manner. Specifically, these single turn-over allylic carbasugars covalently modify GBA and are turned over to release a product which is no longer capable of binding to the enzyme[836], Figure 5.3. As a result, these PCs are unlikely to exhibit prolonged engagement with GBA in the lysosome, Figure 5.1 (b).

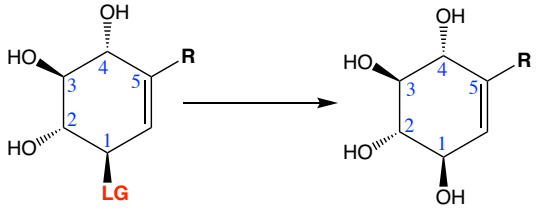
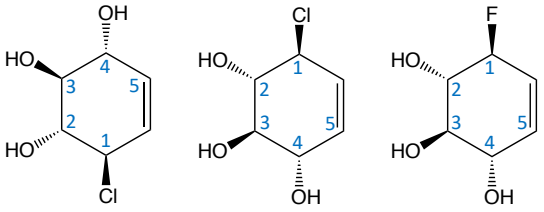


Figure 5.3: General chemical structure of GBA allylic carbasugar inhibitors in which the endocyclic oxygen is replaced with a carbon and a double bond is introduced. Typical leaving groups (LG) include F, Cl and substituted aromatics. Once the inhibitor has been turned over by GBA, the resulting hydrolysed product is no longer capable of inhibiting the enzyme.

In contrast to the 2-deoxy-2-fluoroglucosides, which can typically only be modified through the 2-fluoro leaving group[666], the allylic carbasugar analogues have the potential to be tailored to tune the rate constant for covalent binding (k_{inact}) and intermediate hydrolysis to yield free enzyme (k_{react})[836]. Indeed, the Vocadlo lab has already shown that by substituting the endocyclic oxygen with a carbon and introducing a double bond, they can generate PCs which react via positively charged transition states that are stabilized by the endocyclic C=C bond rather then the endocyclic oxygen[837,838]. Additionally, by changing the group attached to C5 of the carbocycle, both k_{inact} and k_{react} can be altered, whilst changing the leaving group at C1 (ie the anomeric carbon equivalent at which the group to be cleaved is attached) alters k_{inact} only, Figure 5.3. Therefore, it is possible to control the half-life of these carbasugar PCs, which is of benefit to maximizing their chaperoning behaviour. However, perhaps the most significant advantage of such inhibitors is that after hydrolysis of the covalent enzyme-inhibitor complex, the resulting product is inactive and can no longer inhibit the enzyme. Therefore, it is hoped these inhibitors will efficiently clear the enzyme and will not experience prolonged target engagement in the lysosome.

5.2.4 Research Aims

The work discussed in this chapter aimed to structurally characterise a number of halide carbasugar inhibitors recently developed by the Vocadlo and Bennet groups as potential pharmacological chaperones for GBA, Figure 5.4. Specifically, it was hoped that solving the crystal structures of these inhibitors in complex with human GBA would reveal key structural and mechanistic information to improve our understanding of their mode of action. It is also envisaged that such structural information may enable a structure-guided approach to improving the pharmacological chaperone properties of these inhibitors by highlighting important molecular interactions with GBA.



15 **16** **17**

Figure 5.4: Chemical structure of halide carbasugar inhibitors (**15, 16, 17**) investigated in this work. Carbon centres numbered according to IUPAC priorities.

5.3 Materials and Methods

5.3.1 Protein Production and Crystallisation

Human β-glucocerebrosidase (GBA) was produced in an insect-baculovirus expression vector system and purified according to procedures outlined in **Chapter 4**. GBA was crystallised under the HEPES conditions described in **Chapter 4** (*section 4.3.8.2*), with 0.2 µl GBA (10 mg ml^{-1}) + 0.4 µl well solution [0.2 M sodium sulfate, 14% (v/v) PEG 3350, 0.25 M HEPES pH 7] + 0.1 µl seed solution (1:1000 dilution).

5.3.2 Inhibitor Complex Structures

GBA co-crystal complexes were generated by soaking unliganded GBA crystals in SRK049, SRK121 and SBB006. In each case, a small amount of solid compound was dissolved in mother liquor drops containing the unliganded GBA crystals. The crystals were soaked for 10, 30 and 60 minutes before briefly transferring to a cryoprotectant solution comprised of mother liquor supplemented with 20% (v/v) glycerol. The crystals were flash frozen in liquid nitrogen for data collection.

5.3.2.1 Data Collection and processing

Data were collected at the i03 (SRK121 and SBB006) and i04 (SRK049) beamlines of the Diamond Light Source facility (DLS) and integrated using the -dials (SRK049 and SRK121) or -3dii (SBB006) pipelines in XIA2[615,616]. Data reduction was performed using AIMLESS[617,618] through the CCP4i2 software[619]. The SRK049 complex was solved by molecular replacement using MOLREP[620] with a previously obtained unliganded GBA structure (PDB 6TJK)[663] as the homologous search model. Refinement was performed using REFMAC[621] followed by several rounds of manual model building with COOT[622]. SRK121 and SBB006 complexes were solved against the refined SRK049 co-crystal structure. Idealized coordinate sets and refinement dictionaries for the ligands were generated in AceDRG[799,800]. Sugar conformations were validated using Privateer[594]. Crystal structure figures were generated using CCP4mg[94]. Data collection and refinement statistics are summarised in Table 5.1.

Table 5.1: Collection and refinement statistics for all structures discussed in this chapter

	SRK049 (15)	**SRK121 (16)**	**SBB006 (17)**
Data collection			
Beamline	Diamond i04	Diamond i03	Diamond i03
Space group	*P* 2_1	*P* 2_1	*P* 2_1
Cell dimensions			
a, *b*, *c* (Å)	53.2, 76.5, 68.5	53.1, 76.5, 68.1	52.9, 156.0, 68.0
α, β, ɣ (°)	90, 102, 90	90, 102, 90	90, 102, 90
Resolution (Å)	52.10-1.58 (1.61-1.58)*	51.96-1.49 (1.52-1.49)	51.99-1.59 (1.62-1.59)
R_{merge}	0.131 (1.569)	0.112 (1.795)	0.150 (1.998)
R_{pim}	0.055 (0.679)	0.046 (0.728)	0.062 (0.801)
I / σI	8.8 (1.1)	8.6 (0.9)	8.6 (1.1)
CC1/2	0.998 (0.556)	0.997 (0.608)	0.997 (0.544)
Completeness (%)	100 (99.9)	98.9 (97.0)	99.6 (99.1)
Redundancy	6.7 (6.3)	6.7 (6.9)	6.9 (7.1)
Refinement			
Resolution (Å)	52.10-1.58	51.96-1.49	51.99-1.59
No. reflections	73507	85814	143489
R_{work} / R_{free}	0.15/0.18	0.17/0.21	0.18/0.22
No. atoms			
Protein	4096	3985	7952
Ligand/Ion	204	217	412
Water	543	476	874
B-factors (Å^2)			
Protein	19	21	20
Ligand	47	42	38
Water	35	35	32
R.m.s. deviations			
Bond lengths (Å)	0.019	0.007	0.007
Bond angles (°)	1.45	1.46	1.39
Ramachandran Plot Residues			
Most favourable regions (%)	96.0	95.2	95.1
Allowed regions (%)	3.0	3.8	3.9
PDB code	-	-	-

*Values in parentheses are for the outer shell

5.4 Results and Discussion

All crystal structures described in this chapter were obtained using the recombinant GBA (rGBA) formulation produced in the insect-baculovirus expression system (BEVS) described in **Chapter 4**. In order to obtain co-crystal structures with the reversible carbasugars, a number of time dependent ligand soaks were performed to capture the transient covalent species.

5.4.1 SRK2049 Co-Crystal Structure

Following a 30-minute soak, a co-crystal structure of rGBA in complex with SRK049 (**15**) was obtained at 1.58 Å resolution, revealing unambiguous electron density for SRK049 bound covalently to the catalytic nucleophile (Glu340) of GBA, Figure 5.5 (a). Specifically, SRK049 had reacted with Glu340 via nucleophilic attack at the anomeric carbon and subsequent cleavage of the C—Cl bond, Figure 5.5 (b). The resulting enzyme-inhibitor complex was modelled at 75% occupancy, with a covalent bond length of 1.38 Å between the anomeric carbon and the Glu340 carboxylate. The lower occupancy likely reflects reversible inhibition by SRK049 and partial hydrolytic turnover of the covalent complex, nevertheless, a covalent complex was captured. Importantly, the electron density was sufficiently clear to model the reacted carbasugar ring in the envelope ^{2}E conformation (where C1 is attached to the chlorine leaving group), Figure 5.5 (a). This conformation is not a member of the traditional β-glucosidase conformational reaction itinerary but is likely enforced by the planar conformational requirement of the endocyclic double bond. The carbasugar also forms an extensive hydrogen bonding network, making hydrogen bonding interactions with active site residues Asp127, Trp179, Asn234, Glu340 and Trp381 through its hydroxyl substituents. This highlights the specificity with which the carbasugar binds in the active site.

Prior to this work, the exact nucleophilic residue with which SRK049 reacts was uncertain, as was the nature of the covalent complex. Therefore, this co-crystal structure firstly demonstrates that SRK049 covalently modifies the catalytic nucleophile of GBA and secondly describes the transient enzyme-inhibitor covalent complex. It is evident that this inhibitor binds in the active site with good specificity, as indicated by its hydrogen bonding network; however, it holds the significant advantage that once it has been turned over by the enzyme it releases a product that is no longer capable of binding to the enzyme. Therefore, in contrast to most other PCs, SRK049 should not exhibit prolonged GBA engagement or inhibition.

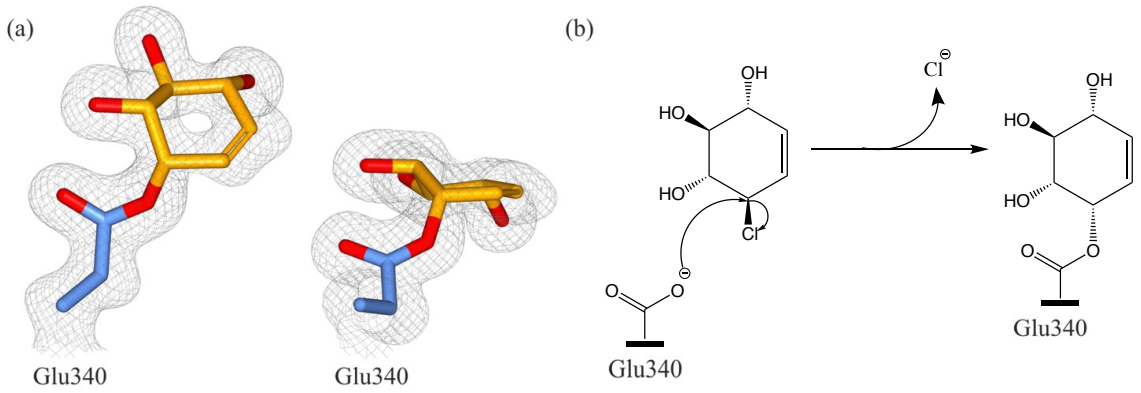


Figure 5.5: (a) Co-crystal structure of rGBA in complex with SRK049 (**15**), showing two orientations of the covalent complex in which SRK049 adopts an ^{2}E envelope conformation. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Glu340 contoured to 1 σ (0.37 e/Å^3). (b) Proposed reaction mechanism for the inhibition of rGBA by SRK049.

5.4.2 SRK121 Co-Crystal Structure

Following structural analysis of SRK049, we next sought to investigate its enantiomer SRK121 (**16**). On initial inspection, the obtained co-crystal structure revealed a covalent enzyme-inhibitor complex which is seemingly identical to that of SRK049 with regard to hydroxyl stereochemistry and hydrogen bonding network, Figure 5.6 (a). However, being the enantiomer of SRK049, this inhibitor must have reacted through an alternative mechanism to generate this covalent complex with the observed hydroxyl stereochemistry. Therefore, in contrast to direct nucleophilic attack at C1 to break the C—Cl bond, it appears SRK121 has reacted at the endocyclic double bond via C5, resulting in allylic rearrangement and subsequent elimination of Cl^-, Figure 5.6 (b). This mechanism yields the observed enzyme-inhibitor complex in a ^{4}E envelope conformation (where C1 is attached to the chlorine leaving group, or equivalent to ^{2}E conformation if given glucose numbering) with a covalent bond length of 1.40 Å to Glu340. Similarly, to SRK049, this complex was the modelled at 70% occupancy, suggesting some non-covalent binding was present. Nevertheless, this covalent complex provides evidence for a novel allylic rearrangement mechanism which has not previously been reported for GBA.

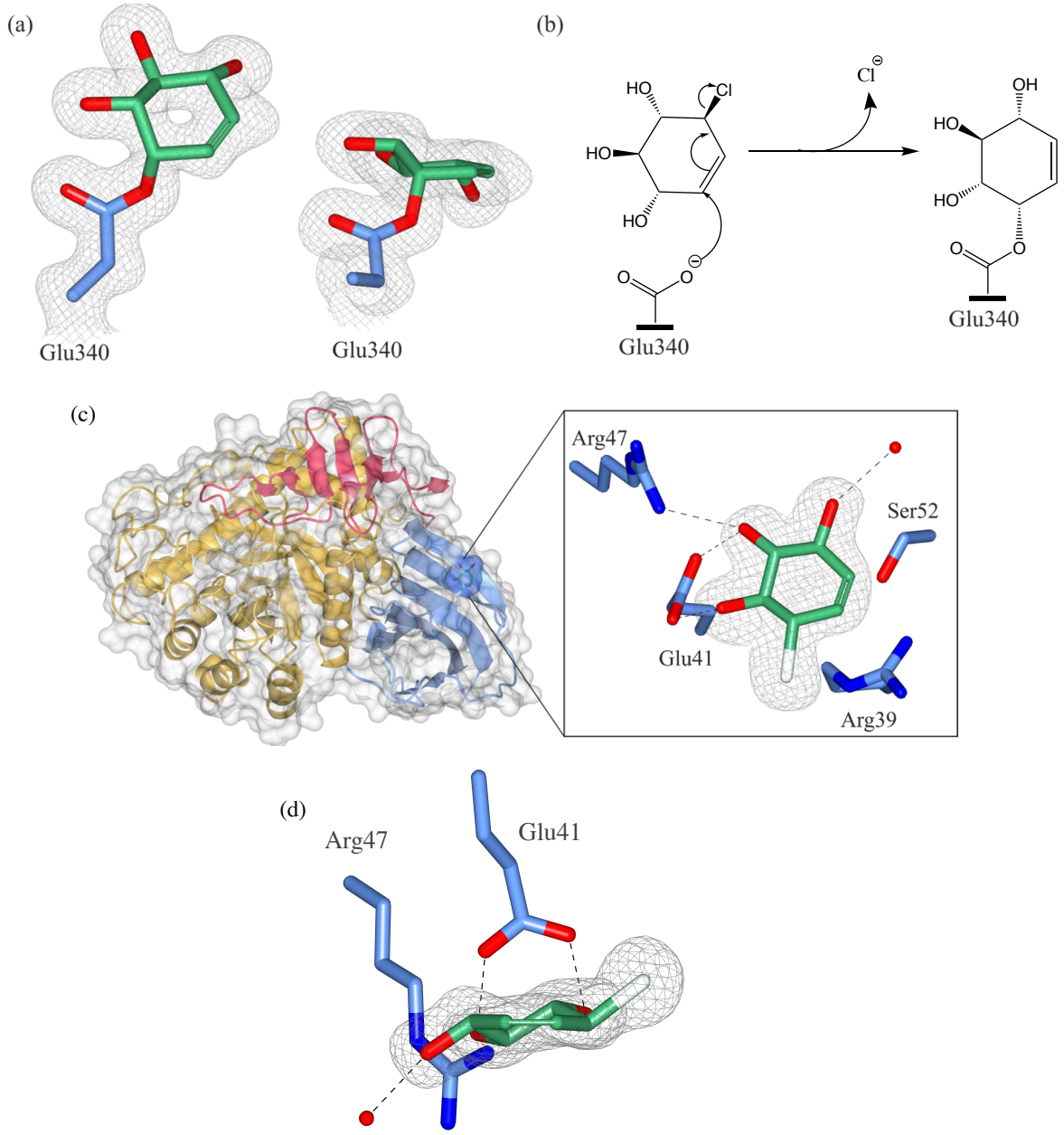


Figure 5.6: (a) Co-crystal structure of rGBA in complex with SRK121 (**16**), showing two orientations of the covalent complex in which SRK121 adopts an ^{4}E envelope conformation (equivalent to the ^{2}E conformation if numbering is changed to glucose such that the reactive centre is numbered C1). (b) Proposed reaction mechanism for the inhibition of GBA by SRK121 (c) Surface and ribbon diagram of GBA highlighting the allosteric binding site at the surface of immunoglobulin-like domain. (d) Electron density for the non-covalent binding of SRK121 in the allosteric binding site. SRK121 adopts the $^{2}H_{3}$ half-chair conformation (equivalent to $^{4}H_{3}$ if applying glucose numbering). White atom = chlorine. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) contoured to 1 σ (0.36 e/Å^3).

Intriguingly, a single molecule of SRK121 was also observed to bind non-covalently in a distant, allosteric site at the surface of the immunoglobulin-like domain of GBA, Figure 5.6 (c). In contrast to the envelope conformation adopted in the active site, the carbasugar ring of SRK121 adopts the half-chair $^{2}H_{3}$ conformation in this allosteric site, forming hydrogen bonding interactions with Arg47, Glu41 and a nearby water molecule, Figure 5.6 (c, d). A potential π- π stacking interaction between the endocyclic double bond of SRK121 and the sidechain of Arg39 was also highlighted, with a separation distance of 3.3 Å between the π- π planes. Interestingly, the enantiomer SRK049 did not bind to this site, indicating that the hydroxyl group stereochemistry and resulting hydrogen bonding network strongly influence non-covalent binding in this site.

5.4.3 SBB006 Co-Crystal Structure

Following analysis of the chlorine substituted allylic carbasugars, focus turned to the fluorine analogue SBB006. Unfortunately, the resulting co-crystal structure proved harder to analyse, yielding a more ambiguous result. A co-crystal structure of rGBA in complex with SBB006 (**17**) was obtained at 1.59 Å, revealing a single molecule of SBB006 bound non-covalently in the enzyme active site of GBA, Figure 5.7. Due to our inability to distinguish between F and OH in the electron density map, it is unclear from whether the molecule of SBB006 is unreacted, and therefore bound in the "Michaelis complex", or if the inhibitor has been turned over and the hydrolysed product is observed in the active site; the former is most probable given that once these inhibitors are turned over by GBA they are no longer capable of inhibiting the enzyme.

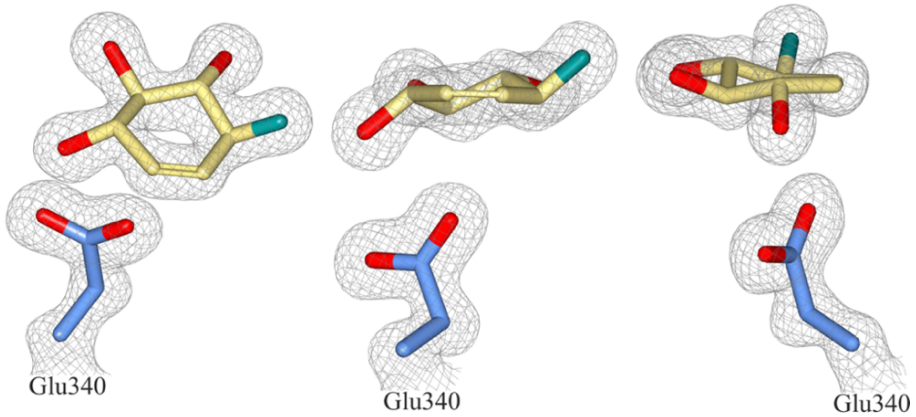


Figure 5.7: Crystal structure of rGBA in complex with SBB006 (**17**), showing three orientations of SBB006 bound non-covalently in the active site in a half-chair conformation. Green atom = F or OH. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) selected for the ligand and Glu340 contoured to 1 σ (0.37 e/Å^3).

Despite trying various ligand soaking times, a covalent complex with SBB006 could not be obtained. Nevertheless, this structure provides insight into the non-covalent binding and conformation of SBB006 in the active site, potentially describing the Michaels complex of such compounds. Indeed, the electron density unambiguously shows that SBB006 has bound in the half-chair conformation with the endocyclic double bond (that somewhat "mimics" the partial double bond of the oxocarbenium transition state) positioned over the catalytic nucleophile Glu340, Figure 5.7. The carbasugar ring also forms hydrogen bonding interactions with active site residues Asp127, Trp179, Asn234, Glu340 and Trp396, resulting in a similar hydrogen bonding network to that of SRK121 in its covalent complex.

Similarly, to SRK121, SBB006 was also observed to bind non-covalently in the allosteric site at the surface of the immunoglobulin-like domain of GBA, Figure 5.8. In this instance, the inhibitor was assumed to be unreacted and was modelled in the same orientation as observed for SRK121. Under this assumption, the electron density shows that SBB006 adopts the $^{2}H_{3}$ half-chair conformation, forming the same hydrogen bonding network as SRK121. Specifically, the carbasugar ring makes hydrogen bonds with Glu41, Arg47 and a nearby water molecule, and forms a potential π- π stacking interaction with Arg39 through the endocyclic double bond, Figure 5.8. This binding mode is identical to that observed for SRK121 which in combination with the lack of binding of SRK049, suggests this binding site is enantiospecific and is underpinned by the hydroxyl group stereochemistry and resulting hydrogen bonding network.

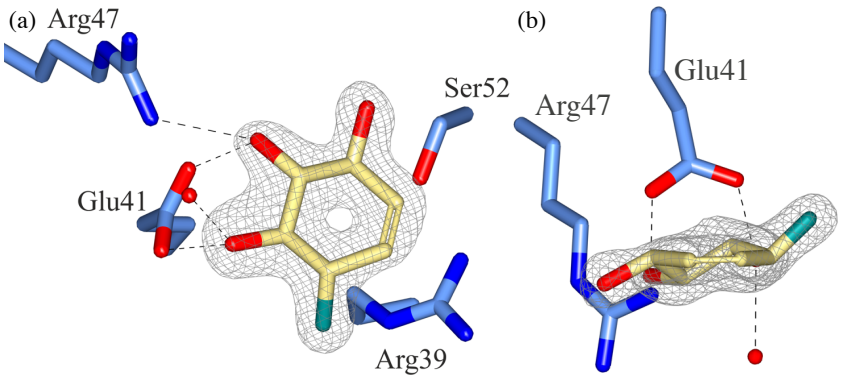


Figure 5.8: Electron density showing non-covalent binding of SBB006 (**17**) in the allosteric binding site in the immunoglobulin-like domain of GBA. (a) SBB006 forms hydrogen bonds with Glu41, Arg47 and a nearby water molecule, as well as π- π stacking interactions with Arg39. (b) The carbasugar ring adopts the $^{2}H_{3}$ half-chair conformation (equivalent to $^{4}H_{3}$ if applying glucose numbering). Green atom = fluorine atom. Maximum-likelihood/σA weighted electron density map ($2F_o$-F_c) for the ligand contoured to 1 σ (0.37 e/Å^3).

5.4.4 Chaperoning Behaviour

The live cell chaperone assays described here were performed by researchers in the Vocadlo lab but have been included for completeness and to link with the structural and mechanistic information generated in this work.

5.4.4.1 Live Cell Chaperone Assays

Using their innovative GBA fluorescence-quenched substrates[700], the Bennet group has already demonstrated the chaperoning behavior of these carbasugars in live cells. Following a 48-hour treatment, SRK121 and SRK049 were shown to enhance the lysosomal GBA activity in wild type and L444P/P415R mutant fibroblasts, Figure 5.9. Moreover, these PCs appear to chaperone GBA differentially, with SRK049 being more effective in chaperoning WT GBA, whilst SRK121 provides a greater chaperoning effect for mutant L444P/P415R. This is of particular interest because the L444P/P415R GBA mutation has been linked to α-synuclein accumulation which contributes to PD[839,840]. Therefore, this chaperone may also show therapeutic potential for PD.

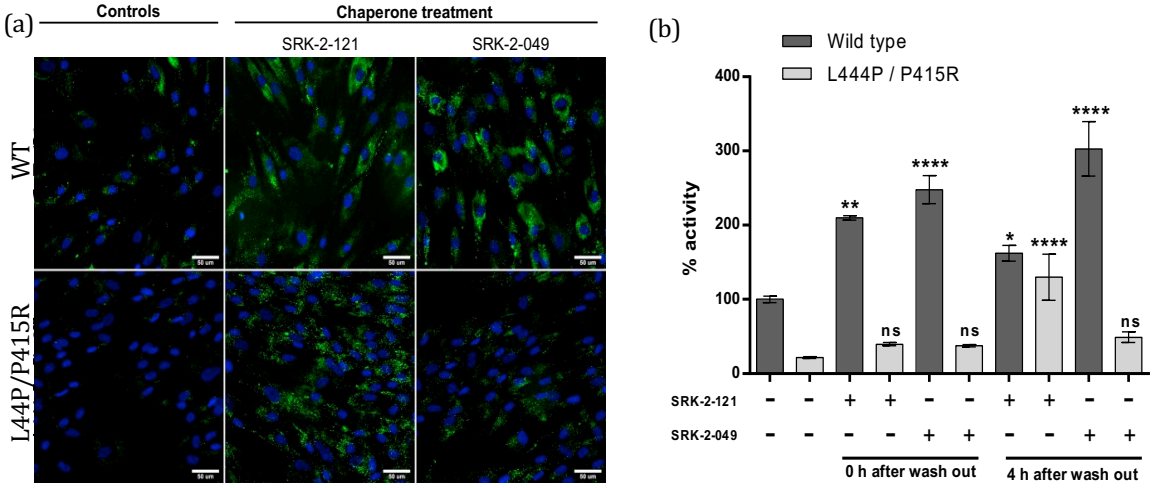


Figure 5.9: Wild type (WT) and L444P/P415R GBA fibroblasts cultured for 48 h with chaperone SRK121 or SRK049 (30 μM). 0 or 4 h after washing with PBS, cells were treated with 5 μM of fluorescent GBA substrate for 1 h and imaged along with Hoechst DNA stain. (A) Images of WT (top) and L444P/P415R (bottom) fibroblasts showing considerable increase in GBA activity (green fluorescence) compared to control (B) GBA activity in lysosomes of WT (dark grey) or L444P/P415R (light grey) fibroblasts treated with SRK121 or SRK049 measured 0 or 4 h after washing out the PC (normalized to untreated cells). Error bars represent standard error of the mean. *Data and figure supplied by Prof. David Vocadlo.

5.4.4.2 Links to Structural Data

Although these PCs require further analysis, the preliminary chaperone studies are very promising and combined with the structural information generated in this work, the different mode of actions of these PCs are now better understood. In addition to revealing two distinct covalent inhibition mechanisms, this work also identified a novel enantiospecific allosteric binding site in the immunoglobulin-like domain of GBA, Figure 5.10. Indeed, it is possible that the enhanced chaperoning effect of SRK121 on L444P/P415R mutant GBA may be structurally rationalized by its ability to bind to this allosteric site. Specifically, the L444P and P415R mutations reside near this allosteric binding site, with L444P located on one side of the site on a β-sheet of the immunoglobulin domain, and the P415R mutation sitting on the other side of the site on an α-helix of the TIM barrel domain, Figure 5.10. The ability of SRK121, but not SRK049, to bind to this allosteric site may provide some stabilization to this region and encourage correct protein folding, allowing SRK121 to chaperone this specific mutant more effectively than SRK049. Consequently, these co-crystal structures may provide some insight into the differential chaperoning behaviour of these PCs. Furthermore, considering the L444P/P415R GBA mutant has been linked to increased α-synuclein accumulation and parkinsonism, the structural information presented here could be of considerable value for the development of GBA chaperones with potential therapeutic benefits for PD.

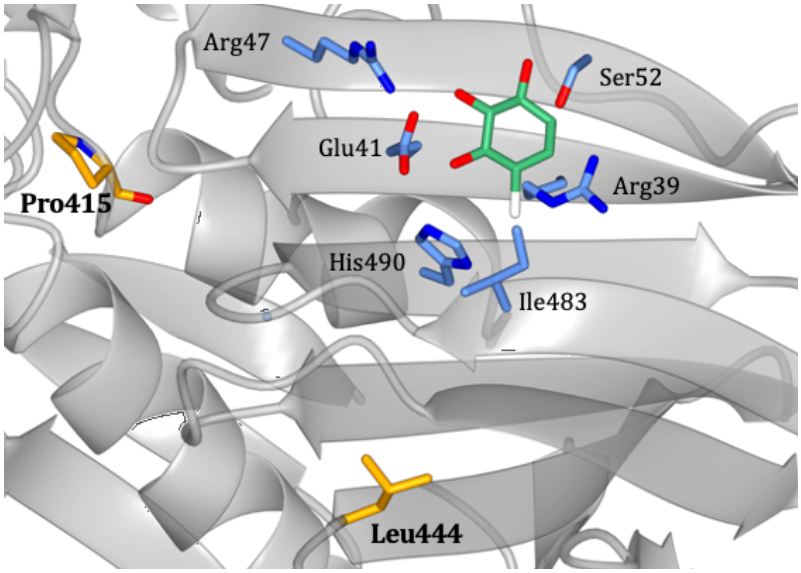


Figure 5.10: Location of Pro415 and Leu444 residues (which are mutated in the L444P/P415R GBA mutant) shown in orange relative to allosteric binding site in which SRK121 binds (ligand in green, binding site residues in blue). Leu444 resides on a β-strand of the immunoglobulin domain, and the P415R mutation sits on an α-helix of the TIM domain.

5.5 Summary

To address the growing demand for molecular chaperones of GBA, the Vocadlo and Bennet laboratories (SF University) have developed a novel class of single turn-over allylic carbasugar inhibitors which chaperone GBA. In order to improve our understanding of their mode of action, structural analysis of these chaperoning inhibitors in complex with human GBA (rGBA) was performed in this work.

The co-crystal structures obtained here reveal that these carbasugar inhibitors covalently modify the catalytic nucleophile (Glu340) of GBA through two distinct inhibition mechanisms. Specifically, SRK049 was shown to react with the enzymatic nucleophile by direct attach at C1 and cleavage of the C—Cl bond, whilst its enantiomer SRK121, was found to react through the endocyclic double bond via C5, resulting in an allylic rearrangement and release of Cl^-. Regardless of the reaction mechanism, both SRK049 and SRK121 form seemingly identical covalent complexes, in which the carbasugar adopts an envelope conformation. Whilst a covalent complex of the fluorine derivative SBB006 could not be obtained, the resulting co-crystal complex provides insight into the non-covalent binding of these inhibitors in the active site. Indeed, this complex may describe the Michaels complex, in which the carbasugar ring adopts the half-chair conformation with the endocyclic double bond positioned over Glu340 for nucleophilic attack. Additionally, this work also uncovered an allosteric site at the surface of the immunoglobulin-like domain of GBA which binds SRK121 and its fluoride equivalent SBB006. In contrast to the envelope conformation observed in the active site, the unreacted carbasugars adopt the $^2H_3$ half-chair conformation (equivalent to $^4H_3$ conformation with glucose numbering) in this alternative site, and bind through predominantly hydrogen bonding interaction. In combination with the lack of binding of the enantiomer SRK049, it appears this allosteric site is enantiospecific. Moreover, binding to this site may provide a structural rationale for the improved ability of SRK121 to chaperone mutant L44P/P415R GBA owing to the fact these mutations lie close to this allosteric site.

By uncovering two distinct covalent inhibitions mechanisms and exposing a distant allosteric binding site, this work provides key structural and mechanistic information which may enable a structure-guided approach to improving the pharmacological chaperone properties of these allylic carbasugar inhibitors.

Chapter 6: Conclusions and Future Perspectives

6.1 Summary

Carbohydrates are one of the most diverse class of biomolecules on earth, and their wide-spread biological importance is reflected by the range of carbohydrate active enzymes (CAZymes) which have evolved to process carbohydrates in nature[4,18,631]. It is therefore no surprise that defects in these CAZymes have been linked to a range of human diseases[17,53]; of note, is the group of metabolic disorders called lysosomal storage disorders (LSDs)[127]. As discussed in **Chapter 1**, these disorders result from inherited deficiencies in certain lysosomal glycoside hydrolases (GHs) which are required for efficient glycolipid catabolism *in vivo.* Subsequently, these diseases are primarily characterised by the cellular accumulation of glycolipids throughout the body which leads to multisystemic clinical symptoms that can be severely debilitating and fatal if not treated effectively[145,167]. Therefore, understanding the enzymes which underpin these diseases is crucial to the development of diagnostic and therapeutic strategies.

The work in this thesis has focussed on the lysosomal GHs α-galactosidase A (α-GAL) and β-glucocerebrosidase (GBA), which underpin the lysosomal storage disorders Fabry disease (FD) and Gaucher disease (GD) respectively. As the most common LSDs, the biochemical and genetic bases of these disorders have been thoroughly investigated, however, there is still considerable work to be done on improving our understanding of the structure and function of these enzymes in disease pathogenesis and phenotype. In fact, relationships between genetic mutations, defective enzymatic activity, extent of glycolipid storage and severity of clinical manifestations remain unclear[149,231,232]. Consequently, there has been persistent academic and pharmaceutical interest in developing inhibitors, activity-based probes (ABPs) and molecular chaperones to study these enzymes in disease pathogenesis, diagnosis and treatment.

Prior to the onset of this work, activity-based protein profiling (ABPP) had emerged as a powerful approach for the simplification of complex proteomes, allowing the activity of specific enzymes to be profiled within native cellular environments[29,531]. Initially, ABPP focused on serine and cysteine proteases and hydrolases[564,565], however, the development of ABPs for GHs was arguably the most important advance in the field of ABPP and has permitted the study of

many glycosidases, including lysosomal α-GAL and GBA[546,556,599]. Building on the Withers' glycosides[607], the Overkleeft lab took advantage of the selectivity of cyclophellitol towards β-glucosidases[669] to develop a range of inhibitors and ABPs, which have proved effective for the study of GBA both *in situ* and *in vitro*[555]. Moreover, the adaptation of such cyclophellitol-based inactivators to other retaining glycosidases, namely α-GAL, was achieved by synthesising cyclophellitol isomers with alternative sugar conformations. Indeed, a range of g*alacto*-configured cyclophellitol epoxide and *N*-functionalised aziridine ABPs have been used to detect and quantify α-GAL activity *in situ*[558]. Coincident with commencement of the work discussed in this thesis, the Overkleeft lab had also expanded their suite of cyclophellitol inhibitors by the use of alternative electrophilic warheads, specifically cyclosulfate and cyclosulfamidate moieties[560,561]. However, most of these cyclophellitol inactivators and their ABP iterations had not been observed on the 3D structure of human GBA or α-GAL, thus hindering the fundamental understanding of ABP reactivity, specificity and conformation.

6.2 This Work and Future Avenues

In collaboration with Prof. Hermen Overkleeft and his lab at Leiden University, one aim of this work was to aid in the development of cyclophellitol inactivators and ABPs for both GBA and α-GAL. Specifically, this work sought to use protein x-ray crystallography to analyse these compounds at the 3D level and provide key structural data to inform the design and development of more potent and selective inhibitors and ABPs.

6.2.1 Structural Analysis of Cyclophellitol Inhibitors for α-GAL

In **Chapter 2**, the first co-crystal structures of recombinant human α-GAL (Fabrazyme®) in complex with a range of α-*galacto* configured cyclophellitol epoxide, aziridine, cyclosulfate and cyclosulfamidate inhibitors were reported. These crystal structures provided the first analyses of such α-*galacto* configured cyclophellitols on the 3D structure of human α-GAL, revealing the covalent mechanism-based mode of action of the unsubstituted epoxide, aziridine and cyclosulfate inhibitors. These studies also provided a structural rationale for the reduced potency reported for *N*-alkyl aziridines, which were found to bind non-covalently, compared to their *N*-acyl analogues. Perhaps the most interesting findings came from analyses of the cyclosulfamidate inhibitor which, in contrast to the cyclosulfate, was found to bind non-

covalently in the active site of α-GAL as a conformational "Michaelis Complex isostere". In preliminary thermal stability studies, this inhibitor was shown to stabilise r-αGAL against thermal denaturation and in subsequent cell-based assays (performed by the Overkleeft Lab) was shown to increase r-αGAL activity in FD fibroblasts and partially correct glycolipid accumulation. These results indicate that *galacto*-α-1,6-cyclophellitol cyclosulfamidate exhibits chaperoning behaviour towards α-GAL, and may be used in combination with ERT to enhance the activity of the recombinant enzyme. Subsequently, we patented this class of compounds as potential pharmacological chaperones for LSDs.

It is hoped the structural information generated in this work will inform the design of more potent ABPs and inhibitors for α-GAL in the future. For example, whilst *N*-alkyl *galacto*-cyclophellitol aziridine ABPs are easier to synthesise and handle over their *N*-acyl counterparts, these structural studies suggest *N*-alkyl ABPs may be less potent due to partial non-covalent binding. These studies also demonstrate the ability to expand the panel of α-GAL inhibitors by the use of alternative electrophilic warheads, specifically cyclosulfates. Additionally, the reported structural analysis of the *galacto*-configured cyclosulfamidate inhibitor suggests that configurational and conformational mimicry of the Michaelis complex is a powerful strategy for the development of competitive non-covalent inhibitors. We believe that transferring the structural characteristics of this *galacto*-cyclosulfamidate to differently configured analogues may yield potent, competitive glycosidase inhibitors that may have biological or biomedical value in their own right; be it as stabilising agents or enzyme inhibitors. Indeed, our work on α-GAL for Fabry disease has been applied to *gluco*-configured cyclosulfates, which appear to be potent α-glucosidase inhibitors[560,841]. Such compounds are now under assessment for human glucosidase inhibition in the context of anti-COVID virus strategies through disruption of viral glycan maturation. Specifically, α-glucosidases I and II are essential in trimming the *N*-linked glycans of the SARS-CoV2 spike protein to form truncated glycans that play critical roles in correct protein folding, quality control and maturation[842]. Furthermore, the truncated glycans produced by the action of α-glucosidases I and II are required for recognition by chaperones calnexin (CNX) and calreticulin (CRT), which feed into a rescue cycle to salvage incorrectly folded proteins[843]. Therefore, the ability of α-glucosidases I and II to regulate glycoprotein entry into the CNX/CRT folding-cycle makes them a potential target for disrupting the production of essential viral glycoproteins of the SARS-CoV2 spike protein[844,845]

6.2.2 Structure-Guided Development of Potent GBA Inhibitors

The "Overkleeft" cyclophellitol-based ABPs have proved powerful tools for the visualisation and profiling of GBA activity. Indeed, these ABPs have potential applications in disease diagnostics, monitoring of disease progression and evaluation of therapeutic intervention. However, applications of the equivalent untagged-inactivators have been more restricted.

Animal models linking impaired GBA function to GD have long been sought after, however, the generation of such models has proved challenging. Specifically, chemical knockdown strategies using the mechanistic inhibitors conduritol-B-epoxide (CBE) and cyclophellitol have been limited by the viability of the resulting animal models due to off-target inhibition of other β-glucosidases. Evidently, more potent and selective GBA inhibitors are required. Therefore, we questioned whether C6-substituted cyclophellitols, bearing bulky hydrophobic moieties at the C6-position, would be more selective GBA inhibitors for the generation of Gaucher animal models through chemical knockdown.

In **Chapter 3**, crystal structures of GBA in complex with a number of C6-substituted cyclophellitol ABPs and inactivators were described. These ABP complexes exposed a hydrophobic pocket at the dimer interface of GBA which is capable of accommodating large hydrophobic moieties. Furthermore, this hydrophobic cavity was shown to be unique to GBA, providing a structural basis for the improved GBA selectivity of C6-functionalised inactivators. This information subsequently guided the design of new cyclophellitol inhibitors; of note, a C6 adamantane-substituted cyclophellitol inactivator was found to be a nanomolar inhibitor of GBA, with considerably improved selectivity over CBE and cyclophellitol. Importantly, our collaborators in Leiden successfully employed this inhibitor in chemical knockdown studies to selectively block endogenous GBA activity in *Dario rerio* zebrafish. We believe the improved selectivity of this C6-adamantane substituted inhibitor, in combination with its ability to cross the blood-brain barrier, make it a promising candidate for use in chemical knockdown studies. Its ability to abrogate GBA activity in the brain also provides a route to the development of neuropathic GD animal models, which may aid in the study and treatment of neurological abnormalities in GD.

6.2.3 Over Expression of Non-Clinical Human GBA

Prior to the development of the insect-baculovirus GBA expression system discussed in **Chapter 4**, mechanistic and structural studies of human GBA, particularly in academic settings, were limited by the supply of active GBA enzyme. In fact, reliable over-production of GBA had only been achieved in pharmaceutical labs through unknown and hence irreproducible means. Consequently, there was a considerable reliance on costly therapeutic formulations for GBA studies. Indeed, at the start of my PhD, the Davies group was reliant on expired Cerezyme® (Sanofi Genzyme) samples which were difficult to obtain in adequate supply. It became apparent that a more reliable source of GBA would be required to support our long-standing interest in developing novel inhibitors, probes and chaperones for this enzyme. Therefore, a major goal of this work was to establish an in-house expression system for GBA. Owing to the complexity of GBA as a human membrane-associated glycoprotein, GBA cannot be produced in prokaryotic systems. This immediately ruled out simple bacterial expression. Consequently, an insect-baculovirus expression vector system (BEVS) was established for the production of human GBA, as described in **Chapter 4.**

Following investigations into numerous unsuccessful tagged constructs, a tagless formulation was found to be suitable for expression, purification and crystallisation. Importantly, this formulation exhibited comparable activity and biophysical properties to commercial products with reasonable long-term stability. This GBA formulation also crystallised in multiple forms with high resolution, providing the first ever sub-angstrom resolution structure of this protein. More importantly, the utility of these crystals for ligand binding studies was demonstrated through co-crystal structures with various inhibitors, including 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside, the Overkleeft bi-functional ABP (discussed in **Chapter 3**) and a novel class of allylic carbasugars developed by the Vocadlo lab (discussed in **Chapter 5**). These structures not only demonstrate successful application of this GBA formulation in structural studies but also exemplify the demand for active, crystallisable GBA in GD research. In light of its purity, stability and activity, this GBA BEVS platform provides an alternative source of non-clinical GBA which we hope will relieve reliance on therapeutic formulations and support the development of novel GBA active compounds for Gaucher research.

6.2.4 Structural Analysis of Novel GBA Chaperones

The role of defective GBA activity in the pathology of GD, and recent association with Parkinson's disease (PD), has resulted in considerable interest in the development of molecular chaperones for GBA. Unfortunately, traditional, competitive, tight-binding chaperones have struggled to progress from positive pre-clinical results to real-life patient benefit, likely owing to lysosomal retention and prolonged target engagement[831]. Therefore, there remains a compelling need for selective GBA inhibitors which chaperone the enzyme during its transport to the lysosome, but which efficiently dissociate from the enzyme once in the lysosome. In this regard, the Vocadlo and Bennet laboratories developed a novel class of single turn-over, allylic halide carbasugar inhibitors which show pharmacological potential towards human GBA.

In **Chapter 5**, analysis of these carbasugar inhibitors on the 3D structure of recombinant GBA (produced in the BEVS described in **Chapter 4**) was performed to provide mechanistic and structural insight into their mode of action. The obtained co-crystal structures revealed two distinct inhibition mechanisms in which the inhibitors covalently modify the catalytic nucleophile of GBA. Specifically, SRK049 was shown to react by direct cleavage of the carbon-chlorine bond, whilst the enantiomer SRK121 reacted through the endocyclic double bond, resulting in an allylic rearrangement and subsequent release of Cl^-. Not only do these complexes describe the nature and conformation of the transient covalent species, but they also revealed a novel allylic rearrangement mechanism which has not previously been reported for GBA. Additionally, this work also uncovered a seemingly enantiospecific allosteric site in the immunoglobulin-like domain of GBA, which binds SRK121 and SBB006 through predominantly hydrogen bonding interactions. Importantly, the ability to bind to this site may provide a structural rationale for the chaperoning behaviour towards L44P/P415R GBA. Perhaps more promisingly, the L444P/P415R GBA mutant has been linked to increased α-synuclein accumulation and parkinsonism, therefore the structural information presented here could be of considerable value for the development of GBA chaperones with potential therapeutic benefits for PD for which no disease modifying therapy is known.

It is ultimately hoped that the structural and mechanistic information generated in this work may enable a structure-guided approach to improving the pharmacological chaperone properties of these inhibitors. For example, by understanding the different covalent inhibition mechanisms and the nature of the covalent complex, it may be possible to further tune the rate

constant for covalent binding (k_{inact}) and for intermediate hydrolysis (k_{react}) to alter the half-life. Indeed, access to inhibitors with controllable half-lives are of considerable interest for the development of pharmacological chaperones because one of the key limiting factors of traditionally tested chaperones is their retention within the lysosome and prolonged target engagement[831]. Therefore, these allylic carbasugars hold a number of advantages over traditional chaperones; firstly, once the carbasugar reacts with GBA, it releases a product which is no longer able to bind to the enzyme, secondly, by altering the leaving group and carbasugar substituents, there is potential to control the reactivity and half-life of these inhibitors. Therefore, these carbasugars hold considerable promise for further development to yield chaperones which may translate into real clinical benefit for GD and PD.

6.2.5 Inactive GBA Mutants: Potential for Glycosynthase Activity

Towards the end of this PhD, I also expanded the GBA BEVS platform to the production of inactive GBA mutants. Whilst these mutants have not been reported in this thesis, I was able to successfully produce, purify and crystallise a number of GBA inactive mutants by mutating either the catalytic nucleophile or catalytic acid-base residue. Our main goal was to investigate these mutants for glycosynthase behaviour as a possible enzymatic route to glycolipid synthesis. Recently, glycosphingolipids (GSLs) have emerged as potential therapeutic compounds for alleviating symptoms of cancer[846,847], diabetes[16,116] and Alzheimer's[848–850]. Consequently, access to sufficient quantities of GSLs is essential in supporting the development of therapeutically valuable products. In nature, glycoconjugates are synthesised by glycosyltransferases (GTs), which are efficient glycosidic bond forming enzymes *in vivo;* however, their applications for *in vitro* synthesis have been limited by difficulties associated with their expression and purification. Alternatively, retaining GHs have been explored for glycosidic bond formation forcing GHs to operate in a transglycosylation mode[851,852], Figure 6.1. A common approach involves putting the GH under kinetic control by using an activated glycosyl donor that reacts with the catalytic nucleophile to form a covalent complex which may be intercepted by a suitable aglycone acceptor. However, yields of such transglycosylation reactions tend to be poor due to competing hydrolytic activity[853], Figure 6.1.

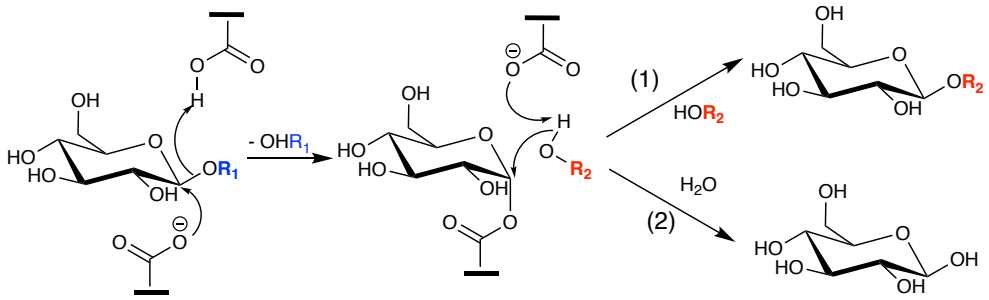


Figure 6.1: Mechanism of (1) transglucosylation (with HOR acceptor) or (2) hydrolysis (H_2O) of a generic retaining β-glucosidase.

An alternative approach is to generate GH mutants which are catalytically incompetent[854–856]. The first class of GH mutants to be engineered for this purpose, termed glycosynthases, were prepared by the Withers lab in 1998[854]. In this seminal work, inactive mutants of the GH1 retaining β-glucosidase from *Agrobacterium sp.* (Abg) were generated by mutating the catalytic nucleophile to amino acids such as alanine, serine or glycine, which are unable to partake in the hydrolysis mechanism, Figure 6.2. Used in conjunction with a glycosyl fluoride donor of opposite anomeric configuration to the natural substrate, these glycosynthase mutants were capable of synthesising complex saccharides with yields on the gram-scale[854]. A number of glycosynthases have since been generated from various GHs, including mutants of endoglycoceramidase II (EGC II) which have proved effective at catalysing the transfer of glycosyl-fluorides to various sphingosine acceptors, resulting in the synthesis of GM gangliosides[857,858]. Given the therapeutic potential of these gangliosides, such enzymatic synthesis is an attractive approach to the synthesis of biomedically valuable glycolipids.

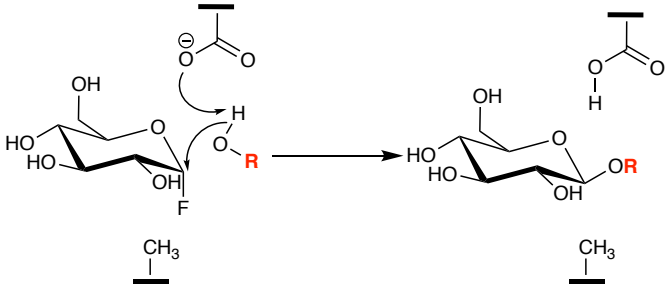


Figure 6.2: Mechanism of glycosynthase derived from a retaining β-glucosidase by mutation of the catalytic nucleophile to a residue that is unable to partake in the catalytic hydrolysis mechanism. A α-fluoro-glycosyl donor is transferred to a suitable acceptor (HOR) with inversion of anomeric stereochemistry.

Several retaining β-glycosidases are reported to perform transglycosylation reactions when provided with a suitable acceptor, and such activity has been reported for GBA[851,859,860].

Therefore, we sought to test if glycosynthase mutants of GBA could be generated. Consequently, I expressed, purified and crystallised GBA mutants in which the catalytic nucleophile (Glu340) was mutated to an alanine (E340A), serine (E340S), glycine (E340G) or glutamine (E340Q) through a QuikChange site directed mutagenesis approach. A mutant in which the catalytic acid-base (Glu235) was mutated to an alanine (E235A) was also prepared. Successful mutation was initially confirmed by intact mass spectrometry analysis, Figure 6.4, followed by structure determination through protein x-ray crystallography to model the mutated residue, Figure 6.3. To the authors knowledge no such mutants of GBA have been produced in a BEVS system before, demonstrating the expansion of this BEV system to the production of GBA mutants.

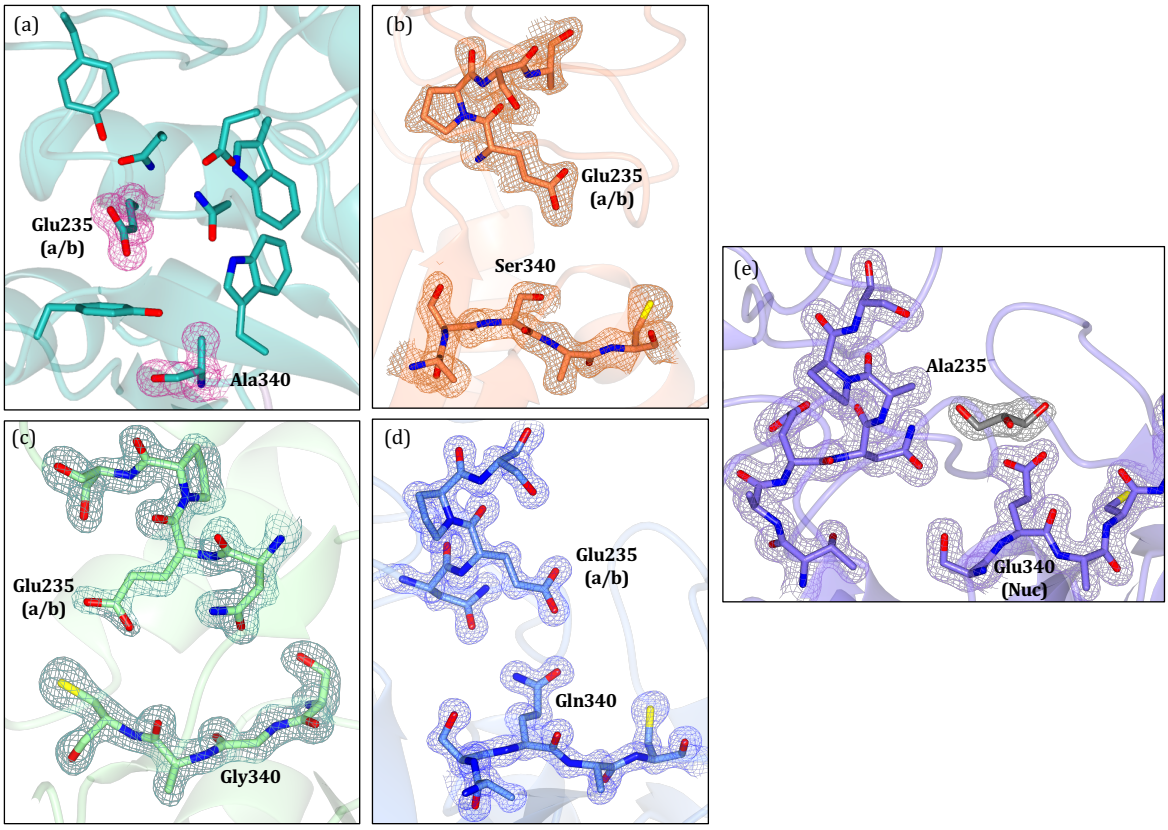


Figure 6.3: Crystal structures of inactive GBA mutants. Electron density for each mutant showing successful mutation of the catalytic nucleophile (Glu340) to (a) Ala340, (b) Ser340 (c) Gly340 (d) Gln340. An acid-base mutant was also produced and crystallised in which Glu235 was mutated to (e) Ala235. Maximum-likelihood/σA weighted electron density maps contoured to 1σ.

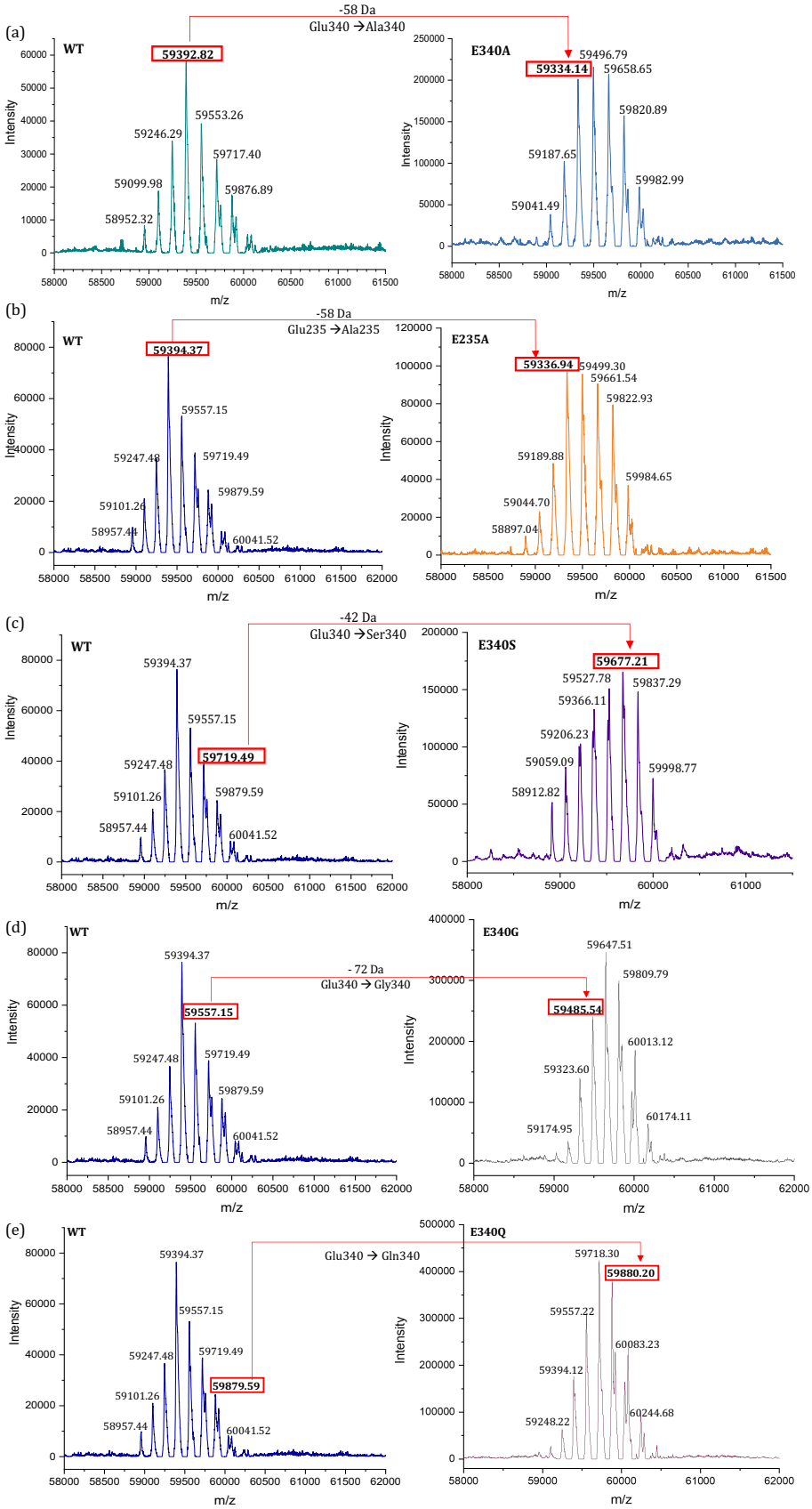


Figure 6.4: Comparison of intact mass spectrum profile of WT GBA with each GBA mutant with showing associated mass shift for (a) E340A (b) E235A (c) E340S (d) E340G (e) E340Q.

Unfortunately, none of these inactive mutants showed transglucosylation behaviour in preliminary TLC assays using various ceramide acceptors and a 1-deoxy-1-fluoro-α-D-glucose donor (synthesis in Appendix 4), Figure 6.5. Whilst some optimisation to encourage glycosynthase activity was performed in regard to donor/acceptor ratios, additives and temperature, no notable transglucosylation was observed and these studies were dropped in light of COVID-19 restrictions. However, it's possible with more testing that these mutants may be capable of performing transglucosylation reactions. Indeed, in 2016, Aerts and co-workers showed that wild-type GBA is capable of glucosylating cholesterol; although the level of transglucosylation was low and fluorescent acceptors were required to detect transglucosylated products[859]. Nevertheless, Aerts and colleagues showed that GBA also mediates the formation of xylosyl-cholesterol by trans-xylosidase reactions[861]. Furthermore, the Overkleeft lab recently reported xylose-configured cyclophellitol and cyclophellitol aziridines that selectively inhibit GBA over GBA2 and GBA3 *in vitro* and *in vivo*[862]*.* Indeed, the xylose configured cyclophellitol appears more potent and more selective for GBA than the classical GBA inhibitor, conduritol B-epoxide (CBE)[862]. Therefore, there is a strong incentive to further investigate these mutants for glycosynthase activity, particularly with xylose configured donors.

Aside from their glycosynthase potential, we envisage these mutants will facilitate the structural analysis of the Michaels complexes of various GBA inhibitors and ABPs. Such Michaelis complexes are difficult to obtain with wild-type enzyme, therefore, crystals of these mutants, Figure 6.3, may provide a route to analysing the non-covalent binding of GBA active compounds before they covalently react. Lastly, it may also be possible to apply this BEVS platform and mutagenesis strategy to produce GD causing GBA mutants. Whilst the GD N370S mutant has been expressed previously in insect cells by Sawkar et al. (2006)[715], the vast majority of GD mutants have not been over-expressed. Establishing a BEVS system for the production of such mutants would provide a route their structural and functional characterisation, which in turn may provide insight into the complex structure-function relationships that exist between GD phenotypes and GBA mutations. Access to GBA mutants would also offer the opportunity to test certain therapeutic strategies, such as chaperone mediated therapies, on specific GBA mutants of clinical interest.

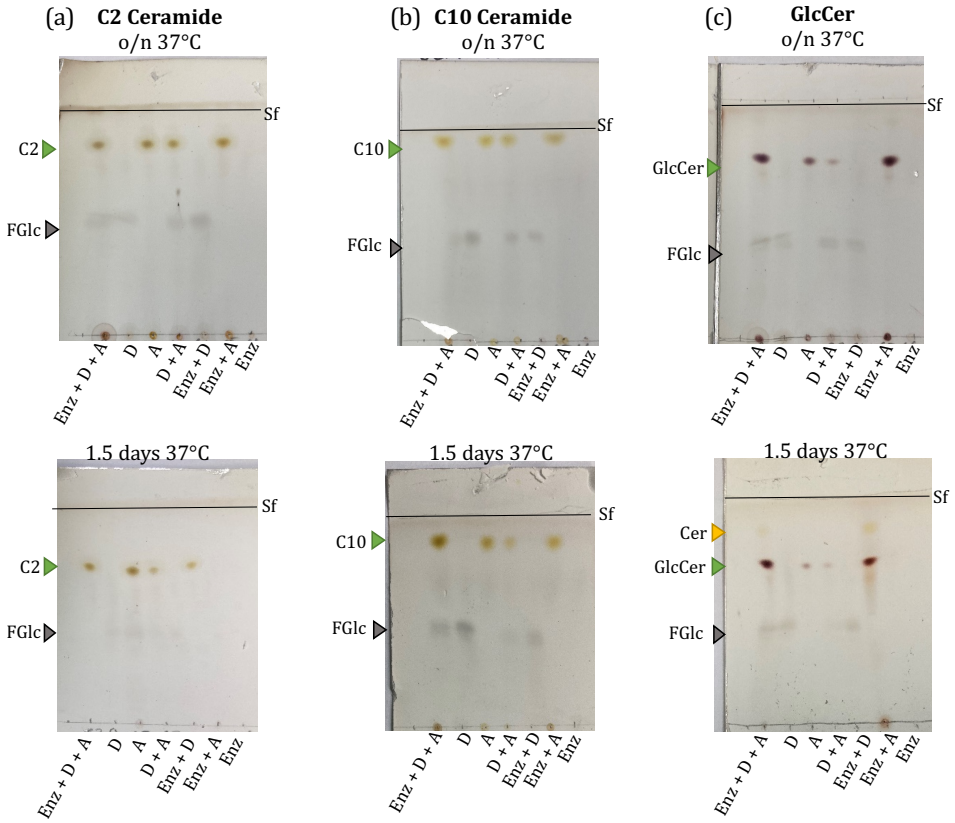


Figure 6.5: TLC analysis of transglucosylation reactions performed with E340S GBA mutant using a mobile phase of ethylacetate:methanol:water (7:2:1). 1-deoxy-1-fluoro-α-D-glucose (FGlc) used as the glycoside donor with (a) C2-Ceramide (b) C10-ceramide and (c) Glucosylceramide GlcCer acceptors. No transglucosylation observed after 1.5 days (would anticipate a new band between donor and acceptor if transglucosylation had occurred) . Some residual hydrolytic activity resulted in the formation of ceramide (Cer) from GlcCer.

6.2.6 Future Perspectives for Cryo-EM

Evidently the GBA BEVS platform established in this work provides many avenues for future exploration. One potential investigative route is to elucidate the binding of GBA with its activator protein Saposin C (SapC). This is of considerable interest in GD research because SapC is required for efficient glucosylceramide metabolism *in vivo*. Additionally, deficiencies in SapC can lead to an atypical form of GD[92,100]. In light of the successful production and purification of both recombinant GBA and SapC reported in **Chapter 4**, it may be possible to obtain a structure of GBA in complex with SapC for structural and functional characterisation. Indeed, cryo-EM studies to obtain such a complex are currently underway, but the small size of both GBA (~55 kDa) and SapC (~9 kDa) has been a limiting factor in cryo-EM data processing.

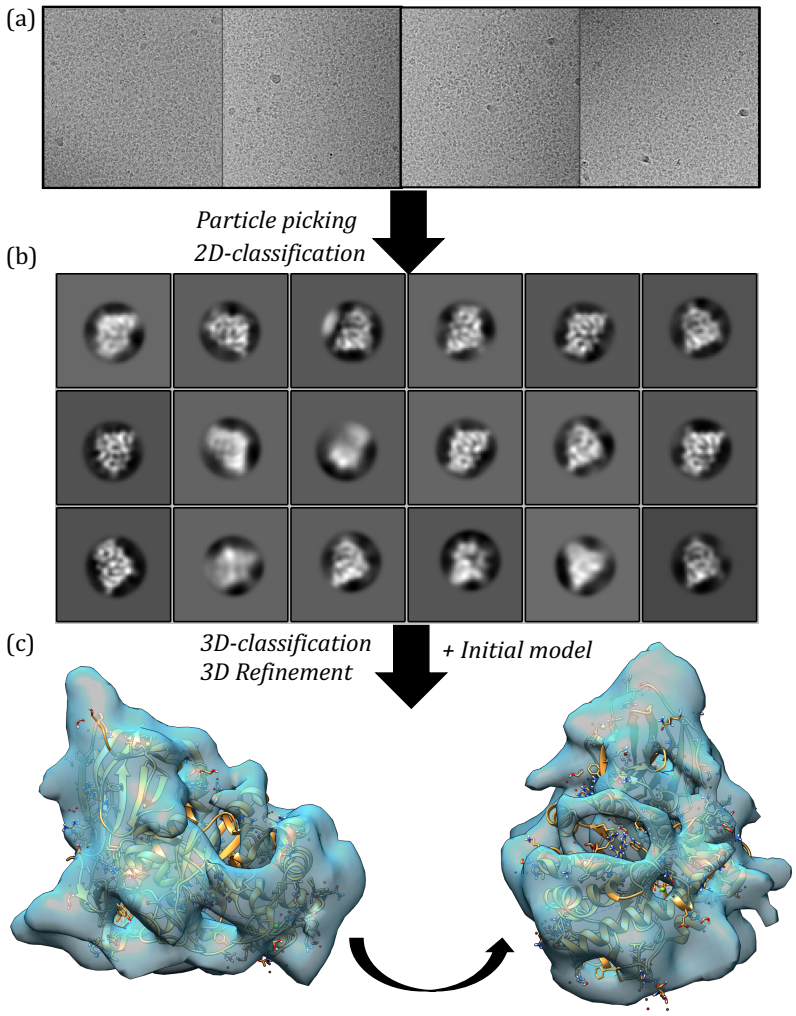


Figure 6.6: Progress on structure determination of GBA by cryo-EM. (a) Selection of micrographs (b) Representative 2D class averages of GBA generated from auto-picked particles. Preferred orientation evident (c) Preliminary 3D reconstructed map of GBA resulting from 3D-classification and 3D-refinement of auto-picked particles. Resulting map comprises of 2 inseparable orientations of GBA. PDB model 6TN1 docked in map.

In preliminary cryo-EM studies, GBA was found to exhibit considerable preferential orientation on the cryo-EM grids. This was ultimately detrimental to its structure solution because an insufficient number of different 2D particle views were obtained for reliable 3D reconstruction, Figure 6.6. Consequently, the preliminary 3D model generated thus far comprises of two different orientations of GBA which we have been unable to separate in data processing, likely owing to its small size and preferential orientation. Figure 6.6. In an effort to overcome such issues, we recently collected a new data set for rGBA using a Volta phase plate, which permits in-focus phase contrast and boosts lower frequency information that is considered vital for small targets[863,864]. Whilst this considerably enhanced the image contrast, Figure 6.7, no

improvement in the reconstructed 3D model could be obtained, indicating that preferential orientation of GBA is the real limiting factor.

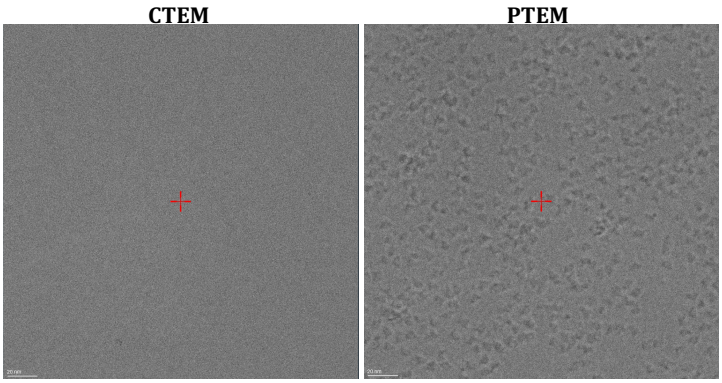


Figure 6.7: Effect of the Volta phase plate; representative in-focus micrograph of rGBA under conventional transmission electron microscopy (CTEM) compared to in-focus image under phase plate transmission electron microscopy (PTEM). Considerable enhancement in image contrast is obtained using PTEM allowing GBA particles to be visualised whilst in focus.

It is hoped that a GBA-SapC complex may be obtained in the future to alleviate both the preferential orientation and size issues limiting our cryo-EM data processing efforts. Indeed, we have already shown that a 6.4-fold enhancement in GBA activity can be achieved in the presence of recombinant SapC (see **Chapter 4** *section 4.4.3*), suggesting some interaction between the two exists in solution. Nevertheless, a complex of this size (~65 kDa) still sits at the cutting-edge limit of cryo-EM, therefore, attempts to generate a complex through x-ray crystallography are also ongoing. It is hoped that this preliminary work will lay a good foundation for future advances in understanding the interactions between GBA and SapC.

6.3 Current and Future Developments in Activity-Based Protein Profiling (ABPP)

The field of ABPP has advanced rapidly over the past couple of decades, resulting in the development of ABPs for GHs[546,566], kinases[567], phosphatases[568], methyltransferases[569,570] and ubiquitin ligases[571]. Aside from biomedical applications, ABPP has found utility in other areas, namely industrial biotechnology. The need for cost-effective biomass processing strategies is of utmost importance as the demand for renewable energy and sustainable products grows. However, the recalcitrance of biomass materials necessitates the search for enzymes that degrade these materials efficiently under industrially relevant conditions. Fortunately, ABPs

provide a route to high-throughput screening, identification and characterisation of enzymes with desirable biomass degrading activities in complex mixtures. In 2019, we showed the detection and identification of industrially useful β-xylosidases and endo-β-1,4-xylanases in the secretomes of *Aspergillus niger* (a model fungal saprophyte) through the use of cyclophellitol ABPs[865]; work for which I contributed structure solution and analysis, Figure 6.8. We demonstrated that simple elongation of monomeric xylose ABPs to mimic a xylobiose moiety is a viable strategy for expanding the scope of cyclophellitol ABPs from β-xylosidases to β-xylanases. Moreover, we demonstrated the use of such ABPs to assess enzyme–substrate specificities, thermal stabilities and other biotechnologically relevant parameters[865], Appendix 5. This study not only highlights the utility of cyclophellitol ABPs as tools for the discovery of biomass degrading enzymes but also provides a basis for further ABP elaboration as a route to interrogating more glycosidases with distinct substrate specificities.

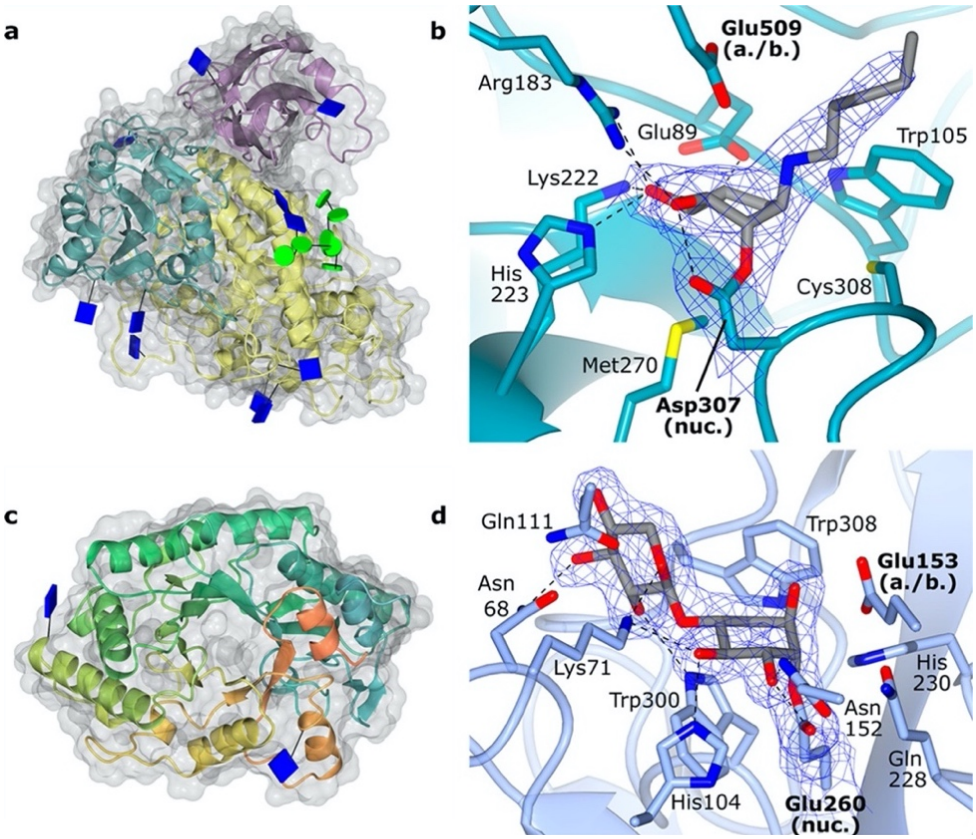


Figure 6.8: Crystal structures of GH3 β-xylosidase and GH10 β-xylanase identified by ABPP (a) Ribbon representation of GH3 β-xylosidase *Anid*XlnD from *A nidulans*. (b) Active site of *Anid*XlnD bound to xylobiose aziridine inhibitor. (c) Ribbon representation of GH10 β-xylanase ASPACDRAFT_127619 catalytic domain. (d) Active site of ASPACDRAFT_127619 bound to xylobiose-epoxide inhibitor. Maximum-likelihood/σA weighted electron density maps contoured to 1.1σ (0.32 e/Å^3).

Following this work, the Overkleeft lab has continued to develop ABPs for the characterisation of enzymes involved in biomass degradation. Specifically, a series of glycosylated cyclophellitol ABPs mimicking β-1,4-glucan oligosaccharides have been synthesised and applied for the detection of cellulases and β-1,4-glucanases required for lignocellulose degradation[866]. Additionally, five-membered α-L-arabinofuranosidase inhibitors have been synthesised following a route inspired by the six-membered cyclophellitol derivatives[867]. α-l-Arabinofuranoside units are commonly found on hemicellulosic and pectinaceous plant polysaccharides, and their efficient removal is required for the breakdown of xylan-rich biomass. Therefore, ABPs were developed from the 5-membered α-L-arabinofuranoside inhibitors, using the aziridine warhead to introduce a reporter group, and applied for the detection of α-L-arabinofuranosidases within *A. niger* and basidiomycetee secretomes[867]. The broad applicability of these inhibitors and ABPs makes them valuable tools for the high throughput characterisation of carbohydrate degrading enzymes in complex systems.

In addition to biomass degrading enzymes, there are a number of other GHs of interest in industrial biotechnology, of note are amylases. Amylases of the GH13 family[631] are hydrolytic enzymes often applied as industrial catalysts in food processing and detergent production[868,869]. However, first and foremost, amylases are required for the processing of starch in many kingdoms of life[870]. In humans, salivary and pancreatic amylases aid in the digestion of starch to glucose, and are commonly targeted in therapeutic approaches for type 2-diabetes. For example, anti-diabetic drugs currently in clinical use include the α-glucosidase inhibitors miglitol and acarbose, which inhibit human pancreatic amylases and control postprandial glucose levels[871–873]. However, the relatively poor selectivity of these inhibitors leads to off target inhibition of intestinal amylases, resulting in undesirable side effects. Therefore, more specific inhibitors of pancreatic α-amylases are of considerable interest[874], in turn requiring efficient and sensitive assays that report specifically on amylase activities in complex biological materials. Fortunately, ABPP are inherently suited for such applications. Consequently the Overkleeft lab synthesised a panel of tagged maltobiose-configured 1,6-epi-cyclophellitols as ABPs for retaining amylases[841], Figure 6.9. Such ABPs have proved effective against Taka-amylase (*Aspergillus oryzae* α-amylase), with cyclosulfate ABPs being more potent than the epoxide and aziridine counterparts. Interestingly, whilst the cyclosulfate and epoxide inhibitors were found to bind in the expected $^4C_1$ chair conformation, the aziridine derivatives bound in an unprecedented E_3 conformation[841], which may account for the inability of cyclophellitol

aziridines to label human saliva amylase. In contrast, the epoxide equivalents label human saliva amylase in a concentration and time dependent manner and are capable of effectively detecting α-amylases in fungal secretomes[841]. Therefore, it is hoped that these labelled maltobiose epi-cyclophellitols will find use in the discovery of new amylase inhibitors with therapeutic potential for type 2 diabetes. Additionally, these ABPs may also aid in the discovery of new microbial amylases with advantageous properties for biotechnological application.

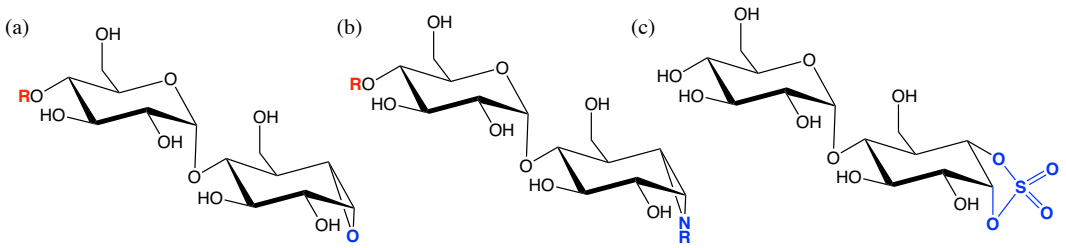


Figure 6.9: Design of mechanism-based retaining cyclophellitol (a) epoxide (b) aziridine and (c) cyclosulfate α-endoglucosidase inhibitors and/or ABPs when R = reporter group.

Whilst it appears that ABPP has reached maturity with regard to retaining glycosidases, there remain many challenges to address in the field of ABP. For example, whilst monosaccharide ABPs targeting retaining *exo*-glycosidases are relatively well established, profiling of retaining *endo*-glycosidases requires more complex ABPs which mimic the length of the natural substrates[875,876]. Another challenge lies in the synthesis of broad-spectrum glycosidase ABPs. In many cases enzyme specific ABPs are desired, however, broad-spectrum ABPs are particularly useful for high-throughput screening of multiple enzyme classes in a relatively simple experimental setup. Unfortunately, one approach to generation of more broad-spectrum probes by removal of cyclophellitol-hydroxyl groups from the ABP scaffold, was found to drastically reduce the potency of the ABPs[549]. Such problem may be overcome by the use of multiple class-specific ABPs in a single experiment, but further work on the design and synthesis of potent, broad-specificity ABPs is required. Lastly, and perhaps most importantly, the biggest weakness of most glycosidase ABPs is the requirement for the target enzyme to form a covalent enzyme-substrate intermediate. Consequently, ABPs for inverting enzymes have been much more difficult to design. Some progress has been made with photo-affinity and quinone methide-based probes[877], however, these probes commonly suffer from poor efficacy and selectivity. In the future, development of ABPs for inverting glycosidases will likely require a combination of high-level synthetic chemistry, predictive computational methods and structural analysis of the proteins of interest.

Appendix 1: Generation of *GBA1* Constructs

A1.1: Oligonucleotide primers required for generation of GBA1 constructs. Oligonucleotide sequence given 5' → 3'

P51	CAGCAGCGAAGTCGCCATAAC
P52	CAGCCGGATCTTCTAGGCTC
P55	CCCAGTCACGACGTTGTAAAACG
P56	AGCGGATAACAATTTCACACAGG
P262	CATCACCACCATCATGGTACCGCAGAAAACTTGTACTTTCAAGGC
P387	CTAGTACTTCTCGACAAGCTTCTACTGGCGACGCCACAGGTAG
F1G	CATCATCACCACCATCATGGTACCGCAGAAAACTTGTACTTTCAAGGCGGAGCCCG CCCCTGCATCCCTAAAAGC
F2G	CATCATCACCACCATCATGGTACCGCAGAAAACTTGTACTTTCAAGGCGGAGGCGC CCGCCCCTGCATCCCTAAAAGC
F3G	CATCATCACCACCATCATGGTACCGCAGAAAACTTGTACTTTCAAGGCGGAGGCGG TGCCCGCCCCTGCATCCCTAAAAGC
F4G	CATCATCACCACCATCATGGTACCGCAGAAAACTTGTACTTTCAAGGCGGAGGCGG TGGAGCCCGCCCCTGCATCCCTAAAAGC
FCHis	TACATTAGCTACATTTATGCGGCCCGCCCCTGCATCCCTAAAAGC
RCHis	CCTCTAGTACTTCTCGACAAGCTTCTAATGATGGTGGTGATGATGCT GGCGACGCCACAG

F = forward primer, R = reverse primer, NHis = N-terminal His-tag, NHis_1G = 1-glycine linker, NHis_2G = 2-glycine linker, NHis_3G = 3-glycine linker, NHis_4G = 4-glycine linker, CHis = C-terminal His-tag, Nuc = nucleophile mutant, ab = acid-base mutant

A1.2: Gene sequence for pGEn1-GBA plasmid (DNASU Clone ID: HsCD00413213[792])

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACA
GGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATACGCGTACCGCTAGCCAGGAAGAGTTTGTAGAAACGCAAAAAGGCCA
TCCGTCAGGATGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTTATGGCGGGCGTCCTGCCCGCCACCCTCCGGGCCGTTGCTTCACAACGTTC
AAATCCGCTCCCGGCGGATTTGTCCTACTCAGGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTCCGACTGAGCCTT
TCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCGTTAACGCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGT
CTTAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACACATTGATGAGCAATGCTTTTTTATAATGCC
AACTTTGTACAAAAAAGCAGGCTCTGAAAACTTGTACTTTCAAGGC**GCCCGCCCCTGCATCCCTAAAAGCTTCGGCTACAGCTCGGTGGT**
GTGTGTCTGCAATGCCACATACTGTGACTCCTTTGACCCCCCGACCTTTCCTGCCCTTGGTACCTTCAGCCGCTATGAGAGTACACGCA

GTGGGCGACGGATGGAGCTGAGTATGGGGCCCATCCAGGCTAATCACACGGGCACAGGCCTGCTACTGACCCTGCAGCCAGAACAGAA
GTTCCAGAAAGTGAAGGGATTTGGAGGGGCCATGACAGATGCTGCTGCTCTCAACATCCTTGCCCTGTCACCCCCTGCCCAAAATTTG
CTACTTAAATCGTACTTCTCTGAAGAAGGAATCGGATATAACATCATCCGGGTACCCATGGCCAGCTGTGACTTCTCCATCCGCACCT
ACACCTATGCAGACACCCCTGATGATTTCCAGTTGCACAACTTCAGCCTCCCAGAGGAAGATACCAAGCTCAAGATACCCCTGATTCA
CCGAGCCCTGCAGTTGGCCCAGCGTCCCGTTTCACTCCTTGCCAGCCCCTGGACATCACCCACTTGGCTCAAGACCAATGGAGCGGTGA
ATGGGAAGGGGTCACTCAAGGGACAGCCCGGAGACATCTACCACCAGACCTGGGCCAGATACTTTGTGAAGTTCCTGGATGCCTATGC
TGAGCACAAGTTACAGTTCTGGGCAGTGACAGCTGAAAATGAGCCTTCTGCTGGGCTGTTGAGTGGATACCCCTTCCAGTGCCTGGG
CTTCACCCCTGAACATCAGCGAGACTTCATTGCCCGTGACCTAGGTCCTACCCTCGCCAACAGTACTCACCACAATGTCCGCCTACTCA
TGCTGGATGACCAACGCTTGCTGCTGCCCCACTGGGCAAAGGTGGTACTGACAGACCCAGAAGCAGCTAAATATGTTCATGGCATTGC
TGTACATTGGTACCTGGACTTTCTGGCTCCAGCCAAAGCCACCCTAGGGGAGACACACCGCCTGTTCCCCAACACCATGCTCTTTGCCT
CAGAGGCCTGTGTGGGCTCCAAGTTCTGGGAGCAGAGTGTGCGGCTAGGCTCCTGGGATCGAGGGATGCAGTACAGCCACAGCATCA
TCACGAACCTCCTGTACCATGTGGTCGGCTGGACCGACTGGAACCTTGCCCTGAACCCCGAAGGAGGACCCAATTGGGTGCGTAACTT
TGTCGACAGTCCCATCATTGTAGACATCACCAAGGACACGTTTTACAAACAGCCCATGTTCTACCACCTTGGCCACTTCAGCAAGTTC
ATTCCTGAGGGCTCCCAGAGAGTGGGGCTGGTTGCCAGTCAGAAGAACGACCTGGACGCAGTGGCACTGATGCATCCCGATGGCTCTG
CTGTTGTGGTCGTGCTAAACCGCTCCTCTAAGGATGTGCCTCTTACCATCAAGGATCCTGCTGTGGGCTTCCTGGAGACAATCTCACC
TGGCTACTCCATTCACACCTACCTGTGGCGTCGCCAGTGACCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATT
TGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGCCATCCAGCTGATATCCCCTATAGTGAGTCGTATTACATGGTCAT
AGCTGTTTCCTGGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAATAATATCATCATGAACAATAAAACTGT
CTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCTGA
TTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGCCAGAGTTGT
TTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACC
ATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGAAAAACAGCATTCCAGGTATTAGAAGAATATCC
TGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATC
GCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTT
GAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTT
TGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCG
GTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTC
GATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGCAAGCTCATGACCAA
AATCCCTTAACGTGAGTTACGCGTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAC
TGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT
ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGG
GCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT

Orange = *GBA1* gene

A1.3: Gene sequence for p-OMNI plasmid

AAGCTTGTCGAGAAGTACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCT
GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAAT
TTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGATCACTGCT
TGAGCCTAGAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTATCATAACCCCTAGGG
TATACCCATCTAATTG**GAACCAGATAAGTGAAATCTAGTTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTAC**
ACCCAGTTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCAACTATTTTGTCCG
CCCACAACCGGTGGAGGAAATTCTCCTTGAAGTTTCCCTGGTGTTCAAAGTAAAGGAGTTTGCACCAGACGCACCTCTGTTCACTGGTCCGG
CGTATTAAAACACGATACATTGTTATTAGTACATTTATTAAGCGCTAGATTCTGTGCGTTGTTGATTTACAGACAATTGTTGTACGTATTT
TAATAATTCATTAAATTTATAATCTTTAGGGTGGTATGTTAGAGCGAAAATCAAATGATTTTCAGCGTCTTTATATCTGAATTTAAATATT
AAATCCTCAATAGATTTGTAAAATAGGTTTCGATTAGTTTCAAACAAGGGTTGTTTTTCCGAACCGATGGCTGGACTATCTAATGGATTTT
CGCTCAACGCCACAAAACTTGCCAAATCTTGTAGCAGCAATCTAGCTTTGTCGATATTCGTTTGTGTTTTGTTTTGTAATAAAGGTTCGACG
TCGTTCAAAATATTATGCGCTTTTGTATTTCTTTCATCACTGTCGTTAGTGTACAATTGACTCGACGTAAACACGTTAAATAGAGCTTGGAC
ATATTTAACATCGGGCGTGTTAGCTTTATTAGGCCGATTATCGTCGTCGTCCCAACCCTCGTCGTTAGAAGTTGCTTCCGAAGACGATTTTG
CCATAGCCACACGACGCCTATTAATTGTGTCGGCTAACACGTCCGCGATCAAATTTGTAGTTGAGCTTTTTGGAATTACCGGTTGACTTGGG
TCAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC

TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACG
ATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGAT
ACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTC
TTTCCTGCGTTATCCCCTGATTGACTTGGGTCGCTCTTCCTGTGGATGCGCAGGTATGTACAGGAAGAGGTTTATACTAAACTGTTACATTG
CAAACGTGGTTTCGTGTGCCAAGTGTGAAAACCGATGTTTAATCAAGGCTCTGACGCATTTCTACAACCACGACTCTAAGTGTGTGGGTGAA
GTCATGCATCTTTTAATCAAATCCCAAGATGTGTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTCTGTCCGTTTGCTGG
CAACTGCAAGGGTCTCAATCCTATTTGTAATTATTGAATAATAAAACAATTATAAATGTCAAATTTGTTTTTTATTAACGATACAAACCAA
ACGCAACAAGAACATTTGTAGTATTATCTATAATTGAAAACGCGTAGTTATAATCGCTGAGGTAATATTTAAAATCATTTTCAAATGATTC
ACAGTTAATTTGCGACAATATAATTTTATTTTCACATAAACTAGACGCCTTGTCGTCTTCTTCTTCGTATTCCTTCTCTTTTTCATTTTTCT
CTTCATAAAAATTAACATAGTTATTATCGTATCCATATATGTATCTATCGTATAGAGTAAATTTTTTGTTGTCATAAATATATATGTCTTT
TTTAATGGGGTGTATAGTACCGCTGCGCATAGTTTTTCTGTAATTTACAACAGTGCTATTTTCTGGTAGTTCTTCGGAGTGTGTTGCTTTAA
TTATTAAATTTATATAATCAATGAATTTGGGATCGTCGGTTTTGTACAATATGTTGCCGGCATAGTACGCAGCTTCTTCTAGTTCAATTACA
CCATTTTTTAGCAGCACCGGATTAACATAACTTTCCAAAATGTTGTACGAACCGTTAAACAAAAACAGTTCACCTCCCTTTTCTATACTATT
GTCTGCGAGCAGTTGTTTGTTGTTAAAAATAACAGCCATTGTAATGAGACGCACAAACTAATATCACAAACTGGAAATGTCTATCAATATA
TAGTTGCTGATTGCGCAGATGCCCTGCGTAAGCGGG**TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACA**
ATAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGC
TAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATGTAAAGACTATATTCGCG
GCGTTGTGACAATTTACCGAACAACTCCGCGGCCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGGTCGATATCAA
AGTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGAGCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGCA
CCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAATGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGAT
ATAGATCTCACTACGCGGCTGCTCAAACTTGGGCAGAACGTAAGCCGCGAGAGCGCCAACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGA
TGAATGTCTTACTACGGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTGGCTACGTCTCCGAACTCACGACCGAAA
AGATCAAGAGCAGCCCGCATGGATTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTGAGCCACCTAACTTTGTTTT
AGGGCGACTGCCCTGCTGCGTAACATCGTTGCTGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGCGTAACGCGCTT
GCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAACAGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGTTCG
GTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTT
CATCCGTTTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGCCATAACTTCGTATAGCATACATTATACGAAGTTATCTG
TAACTATAACGGTCCTAAGGTAGCGAGTTTAAACACTAGTATCGATTCGCGACCTACTCCGGAATATTAATAGATCATGGAGATAATTAAA
ATGATAACCATCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTATTCATA
CCGTCCCACCATCGGGCGCGGATCCATG**AAATTTTTGGTGAACGTGGCCTTGGTGTTTATGGTGGTTTACATTAGCTACATTTATGCG**
GACCCGGGCCATCATCACCACCATCATGGTACC

Red = Tn7 transposition sites , Yellow = Melittin Signal Sequence

A1.4: Gene sequence for B24-backbone

AGCTTGTCGAGAAGTACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTG
AACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT
TCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGATCACTGCTT
GAGCCTAGAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTATCATAACCCCTAGGGT
ATACCCATCTAATTG**GAACCAGATAAGTGAAATCTAGTTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACA**
CCCAGTTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCAACTATTTTGTCCGC
CCACAACCGGTGGAGGAAATTCTCCTTGAAGTTTCCCTGGTGTTCAAAGTAAAGGAGTTTGCACCAGACGCACCTCTGTTCACTGGTCCGGC
GTATTAAAACACGATACATTGTTATTAGTACATTTATTAAGCGCTAGATTCTGTGCGTTGTTGATTTACAGACAATTGTTGTACGTATTTT
AATAATTCATTAAATTTATAATCTTTAGGGTGGTATGTTAGAGCGAAAATCAAATGATTTTCAGCGTCTTTATATCTGAATTTAAATATTA
AATCCTCAATAGATTTGTAAAATAGGTTTCGATTAGTTTCAAACAAGGGTTGTTTTTCCGAACCGATGGCTGGACTATCTAATGGATTTTC
GCTCAACGCCACAAAACTTGCCAAATCTTGTAGCAGCAATCTAGCTTTGTCGATATTCGTTTGTGTTTTGTTTTGTAATAAAGGTTCGACGT
CGTTCAAAATATTATGCGCTTTTGTATTTCTTTCATCACTGTCGTTAGTGTACAATTGACTCGACGTAAACACGTTAAATAGAGCTTGGACA
TATTTAACATCGGGCGTGTTAGCTTTATTAGGCCGATTATCGTCGTCGTCCCAACCCTCGTCGTTAGAAGTTGCTTCCGAAGACGATTTTGC
CATAGCCACACGACGCCTATTAATTGTGTCGGCTAACACGTCCGCGATCAAATTTGTAGTTGAGCTTTTTGGAATTACCGGTTGACTTGGGT
CAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTG
ATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT
TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTT

TCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC
ACCGCCTACATACCTCGCTCTGCTTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAG
CGCACGAGGGAGCTTCCAGGGGGAAACGCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGC
TCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT
TCCTGCGTTATCCCCTGATTGACTTGGGTCGCTCTTCCTGTGGATGCGCAGGTATGTACAGGAAGAGGTTTATACTAAACTGTTACATTGCA
AACGTGGTTTCGTGTGCCAAGTGTGAAAACCGATGTTTAATCAAGGCTCTGACGCATTTCTACAACCACGACTCTAAGTGTGTGGGTGAAGT
CATGCATCTTTTAATCAAATCCCAAGATGTGTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTCTGTCCGTTTGCTGGCA
ACTGCAAGGGTCTCAATCCTATTTGTAATTATTGAATAATAAAACAATTATAAATGTCAAATTTGTTTTTTATTAACGATACAAACCAAAC
GCAACAAGAACATTTGTAGTATTATCTATAATTGAAAACGCGTAGTTATAATCGCTGAGGTAATATTTAAAATCATTTTCAAATGATTCAC
AGTTAATTTGCGACAATATAATTTTATTTTCACATAAACTAGACGCCTTGTCGTCTTCTTCTTCGTATTCCTTCTCTTTTTCATTTTTCTCT
TCATAAAAATTAACATAGTTATTATCGTATCCATATATGTATCTATCGTATAGAGTAAATTTTTTGTTGTCATAAATATATATGTCTTTTT
TAATGGGGTGTATAGTACCGCTGCGCATAGTTTTTCTGTAATTTACAACAGTGCTATTTTCTGGTAGTTCTTCGGAGTGTGTTGCTTTAATT
ATTAAATTTATATAATCAATGAATTTGGGATCGTCGGTTTTGTACAATATGTTGCCGGCATAGTACGCAGCTTCTTCTAGTTCAATTACACC
ATTTTTTAGCAGCACCGGATTAACATAACTTTCCAAAATGTTGTACGAACCGTTAAACAAAAACAGTTCACCTCCCTTTTCTATACTATTGT
CTGCGAGCAGTTGTTTGTTGTTAAAAATAACAGCCATTGTAATGAGACGCACAAACTAATATCACAAACTGGAAATGTCTATCAATATATA
GTTGCTGATTGCGCAGATGCCCTGCGTAAGCGGG**TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA**
TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGCT
AAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATGTAAAGACTATATTCGCGG
CGTTGTGACAATTTACCGAACAACTCCGCGGCCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGGTCGATATCAAA
GTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGAGCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGCAC
CAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAATGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATA
TAGATCTCACTACGCGGCTGCTCAAACTTGGGCAGAACGTAAGCCGCGAGAGCGCCAACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGAT
GAATGTCTTACTACGGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTGGCTACGTCTCCGAACTCACGACCGAAAA
GATCAAGAGCAGCCCGCATGGATTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTGAGCCACCTAACTTTGTTTTA
GGGCGACTGCCCTGCTGCGTAACATCGTTGCTGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGCGTAACGCGCTTG
CTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAACAGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGTTCGG
TCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTC
ATCCGTTTCCACGGTGTGCGTCACCCGGCAACCTTGGG**CAG**CAGCGAAGTCGCCATAACTTCGTATAGCATACATTATACGAAGTTATCTGT
AACTATAACGGTCCTAAGGTAGCGAGTTTAAACACTAGTATCGATTCGCGACCTACTCCGGAATATTAATAGATCATGGAGATAATTAAAA
TGATAACCATCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTATTCATAC
CGTCCCACCATCGGGCGCGGATCCATG**AAATTTTTGGTGAACGTGGCCTTGGTGTTTATGGTGGTTTACATTAGCTACATTTATGCG**G
ACCCGGGC**CATCATCACCACCATCAT**GGTACC

Red = Tn7 transposition sites, Yellow = Melittin Signal Sequence, Blue = His_6-tag

A1.5: Gene sequence for B41-backbone

AACACGATACATTGTTATTAGTACATTTATTAAGCGCTAGATTCTGTGCGTTGTTGATTTACAGACAATTGTTGTACGTATTTTAATAATT
CATTAAATTTATAATCTTTAGGGTGGTATGTTAGAGCGAAAATCAAATGATTTTCAGCGTCTTTATATCTGAATTTAAATATTAAATCCTC
AATAGATTTGTAAAATAGGTTTCGATTAGTTTCAAACAAGGGTTGTTTTTCCGAACCGATGGCTGGACTATCTAATGGATTTTCGCTCAAC
GCCACAAAACTTGCCAAATCTTGTAGCAGCAATCTAGCTTTGTCGATATTCGTTTGTGTTTTGTTTTGTAATAAAGGTTCGACGTCGTTCAA
AATATTATGCGCTTTTGTATTTCTTTCATCACTGTCGTTAGTGTACAATTGACTCGACGTAAACACGTTAAATAGAGCTTGGACATATTTAA
CATCGGGCGTGTTAGCTTTATTAGGCCGATTATCGTCGTCGTCCCAACCCTCGTCGTTAGAAGTTGCTTCCGAAGACGATTTTGCCATAGCC
ACACGACGCCTATTAATTGTGTCGGCTAACACGTCCGCGATCAAATTTGTAGTTGAGCTTTTTGGAATTACCGGTTGACTTGGGTCAACTGT
CAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT
CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC
TGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAG
GTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC
TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTAC
CGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAG
CGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCA
GGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCG
TTATCCCCTGATTGACTTGGGTCGCTCTTCCTGTGGATGCGCACCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTGACTTGGGTC
GCTCTTCCTGTGGATGCGCAGGTATGTACAGGAAGAGGTTTATACTAAACTGTTACATTGCAAACGTGGTTTCGTGTGCCAAGTGTGAAAAC
CGATGTTTAATCAAGGCTCTGACGCATTTCTACAACCACGACTCTAAGTGTGTGGGTGAAGTCATGCATCTTTTAATCAAATCCCAAGATGT

GTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTCTGTCCGTTTGCTGGCAACTGCAAGGGTCTCAATCCTATTTGTAATT
ATTGAATAATAAAACAATTATAAATGTCAAATTTGTTTTTTATTAACGATACAAACCAAACGCAACAAGAACATTTGTAGTATTATCTATA
ATTGAAAACGCGTAGTTATAATCGCTGAGGTAATATTTAAAATCATTTTCAAATGATTCACAGTTAATTTGCGACAATATAATTTTATTTT
CACATAAACTAGACGCCTTGTCGTCTTCTTCTTCGTATTCCTTCTCTTTTTCATTTTTCTCTTCATAAAAATTAACATAGTTATTATCGTAT
CCATATATGTATCTATCGTATAGAGTAAATTTTTTGTTGTCATAAATATATATGTCTTTTTTAATGGGGTGTATAGTACCGCTGCGCATAG
TTTTTCTGTAATTTACAACAGTGCTATTTTCTGGTAGTTCTTCGGAGTGTGTTGCTTTAATTATTAAATTTATATAATCAATGAATTTGGG
ATCGTCGGTTTTGTACAATATGTTGCCGGCATAGTACGCAGCTTCTTCTAGTTCAATTACACCATTTTTTAGCAGCACCGGATTAACATAAC
TTTCCAAAATGTTGTACGAACCGTTAAACAAAAACAGTTCACCTCCCTTTTCTATACTATTGTCTGCGAGCAGTTGTTTGTTGTTAAAAATA
ACAGCCATTGTAATGAGACGCACAAACTAATATCACAAACTGGAAATGTCTTCAATATATAGTTGCTGATTGCGCAGATGCCCTGCGTAAGC
GGG**TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGTCTTAAACTAGACAGAATAGTTGT**
AAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAA
ATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATGTAAAGACTATATTCGCGGCGTTGTGACAATTTACCGAACAACTCCGCGGC
CGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGGTCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAACTTTG
TATAGAGAGCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGCACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTG
ATGAGCGCGGTGGCAATGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATCTCACTACGCGGCTGCTCAAACTTGGG
CAGAACGTAAGCCGCGAGAGCGCCAACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTACGGAGCAAGTTCCCGAGGT
AATCGGAGTCCGGCTGATGTTGGGAGTAGGTGGCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGATTTGACTTGG
TCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTGAGCCACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGCT
GCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGCGTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACA
AAAAAACAGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCAT
ACGCTACTTGCATTACAGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTCATCCGTTTCCACGGTGTGCGTCACCCGGCAAC
CTTGGGCAGCAGCGAAGTCGCCATAACTTCGTATAGCATACATTATACGAAGTTATCTGTAACTATAACGGTCCTAAGGTAGCGAGTTTAAA
CACTAGTATCGATTCGCGACCTACTCCGGAATATTAATAGATCATGGAGATAATTAAAATGATAACCATCTCGCAAATAAATAAGTATTTT
ACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTATTCATACCGTCCCACCATCGGGCGCGGATCCATG**AAATT**
TTTGGTGAACGTGGCCTTGGTGTTTATGGTGGTTTACATTAGCTACATTTATGCGGACCCGGG

Red = Tn7 transposition sites , Yellow = Melittin Signal Sequence

Appendix 2: Investigation of Inner Filter Effect

A2.1: Investigating primary inner filter effect

4-MU was prepared at concentrations ranging 2.5 mm to 2.4 µM in pH 5.2 kinetics buffer (McIlvaine buffer: 150 mM disodium hydrogen phosphate, citric acid [pH 5.2], 0.2% (v/v) Taurocholate, 0.1% (v/v) Triton X-100 and 0.1% (v/v) bovine serum albumin) to cover the concentration range of 4-methylumbelliferyl β-D-glucopyranoside (4-MU-Glc) used in the Michaelis Menten assays. The epifluorescence of 4-MU was measured (λ_{em} 450-430 nm) in 50 uL volumes in quadruplicate in a black 384-well plate using different excitation wavelengths (λ_{ex} 320, 340, 360, 380, 390, 400 nm). The average 4-MU fluorescence was plotted against [4-MU] concentration. The 4-MU fluorescence proved non-linear over the tested concentrations using 360 nm excitation wavelength, indicating the inner filter effect impacts the kinetic assays performed at this wavelength. However, the 4-MU fluorescence appeared linear using 380, 390 and 400 nm excitation wavelengths, indicating kinetic assays performed under these conditions should not be impacted by inner filter effects, Figure A.1

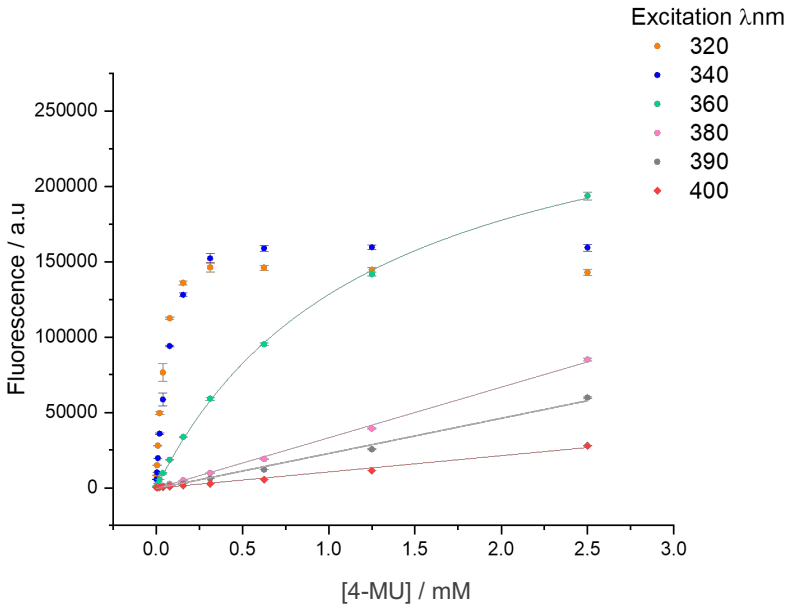


Figure A2.1: Plot of observed epifluorescence (λ_{em} 450/30 nm) vs 4-MU concentration using different excitations wavelengths. Data plotted as average ± standard deviation of 4 replicates. Linear best fit constructed for data collected using 380 nm, 390 nm and 400 nm excitation wavelengths. Fluorescence data at 360 nm fitted to Michaelis Menten equation.

A2.2: Michaelis Menten Assay at λ_{ex} 390 nm

The fluorogenic substrate 4-methylumbelliferyl β-D-glucopyranoside (4-MU-Glc) was prepared at 10 mM in kinetics buffer and diluted 2-fold to 0.019 mM. Each substrate solution (25 µL) was added to the wells of a black 384 well polystyrene plate in quadruplicate. GBA (25 µL, 20 nM) was added to each well to give a final enzyme concentration of 10 nM. Activity against 4-MU-Glc was monitored continuously over 5-minutes at 37 °C by measuring the fluorescence of liberated 4-MU (λ_{ex} 390/15 nm, λ_{em} 450/30 nm) using a CLARIOstar® *Plus* microplate reader (BMG LabTech). A linear calibration was generated by measuring the fluorescence of 4-MU (λ_{ex} 390, λ_{em} 450/30 nm) prepared at serial dilutions from 2.5 mm to 2.4 µM in kinetics buffer. Each 4-MU concentration was measured in quadruplicate. All data were processed in Origin graphing software. Using the 4-MU calibration, the rate of substrate hydrolysis (V) was determined at each substrate concentration. The rates (V) were plotted against substrate concentration [S] and fitted by nonlinear regression to the Michaelis-Menten equation (rate = V_{max} [S] / (K_M + [S])) to generate values of K_M, V_{max} and k_{cat} using the relationship k_{cat} = V_{max}/[Enz]. The determined kinetic parameters proved comparable to those obtained in the original assay, indicating the reported K_M value is correct.

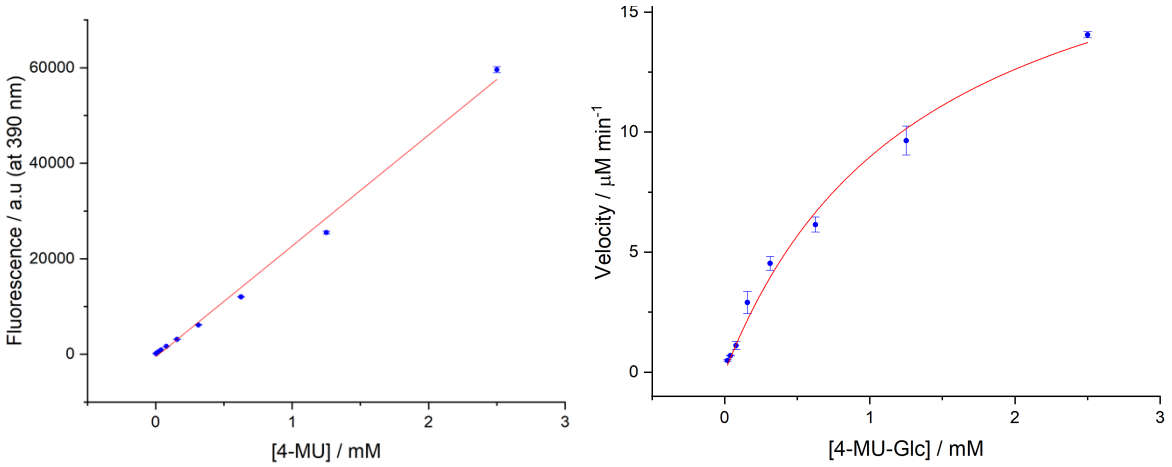


Figure A2.2: (a) Calibration curve of measured 4M fluorescence vs [4-MU] using λ_{ex} 390 nm. (b) Michaelis-Menten kinetic assay of non-tagged GBA at 37 °C using λ_{ex} 390 nm. Data plotted as the average ± standard deviation of 4 replicates. K_M = 1.371 ± 0.244 mM, V_{max} = 21.26 ± 1.88 µM min^{-1}, k_{cat} = 2126 min^{-1}.

Appendix 3: Production of Recombinant Human Saposin C

A3.1: Test expression in various E. coli strains

The gene encoding human Saposin C (SapC), with an additional methionine and aspartate at the N-terminal, was codon optimised for *E. coli* and purchased from Genscript subcloned into the NcoI and BamHI sites of the pET-16b vector. To investigate the best cell line for recombinant Saposin C production, the resulting vector was transformed into *E. coli* AD494 (DE3), Origami B (DE3) cells, BL21trxB (DE3) and Shuffle (DE3) cells by heatshock. Cultures (500 mL) were grown at 37 °C (AD494, BL21trxB, Origami cells) or 30 °C (Shuffle cells) in Luria-Bertani (LD) media supplemented with Ampicillin (100 μg mL^{-1}) to an OD_{600} of 0.8-1.0 before induction with 0.8 mM IPTG. The cultures were grown for a further 4 hours after which the cells were harvested by centrifugation (4000g for 20mins). The cell pellet was resuspended in anion-exchange buffer (25 mM NaCl, 25 mM Tris-HCl, pH 7.5) and lysed by sonication. The lysate was clarified by centrifugation at 16,000g for 10 min and the supernatant was heated at 85°C for ~30 min. Pre- and post-heat-treatment samples were analysed by SDS-PAGE (20%) to reveal greater SapC production in Origami cells.

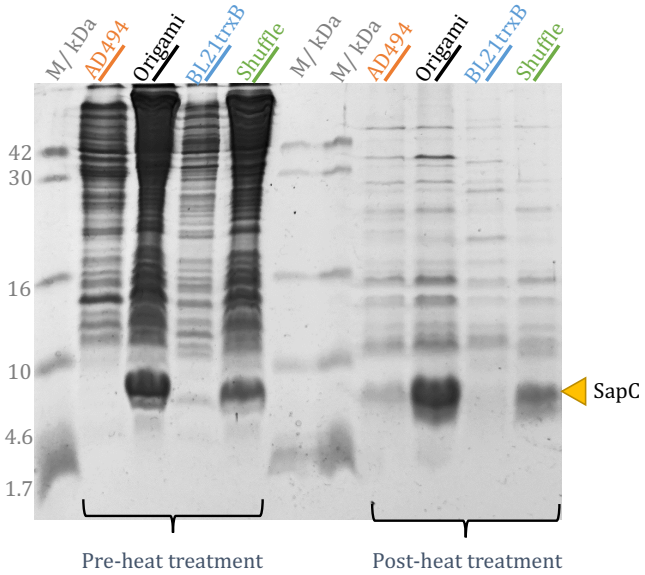


Figure A3.1: SDS-PAGE analysis of pre- and post-heat treatment samples from test expressions in *E. coli* AD494, Origami, BL21trxB and Shuffle cells. Greatest expression of recombinant SapC achieved by Origami cells.

A3.1: Full scale expression and purification

Origami(DE3) cells transformed with the SapC encoding pET-16b vector were grown at 37 °C in Luria-Bertani (LD) media supplemented with Ampicillin (100 μg mL^{-1}) to an OD_{600} of 0.8-1.0 before induction with 0.8 mM IPTG. The cultures were grown for a further 4 hours at 37 °C after which the cells were harvested by centrifugation. The cell pellet was resuspended in anion-exchange buffer (25 mM NaCl, 25 mM Tris-HCl, pH 7.5) and lysed by sonication. The lysate was clarified by centrifugation and the supernatant was heated at 85°C for ~30 min. Precipitated proteins were removed, and the resulting supernatant was applied directly to a Q-Sepharose column (GE Healthcare) that had been pre-equilibrated in binding buffer (25 mM NaCl, 25 mM Tris-HCl at pH 7.5). The protein was eluted with a linear gradient of 0-100% elution buffer (1 M NaCl, 25 mM Tris-HCl at pH 7.5) over 20 CVs followed by a further 5 CVs at 100% elution buffer. The peak fractions containing SapC were pooled, concentrated, and applied to a Superdex S75 16/600 (column GE Healthcare) in 50 mM Tris, 150 mM NaCl, pH 7.4 buffer.

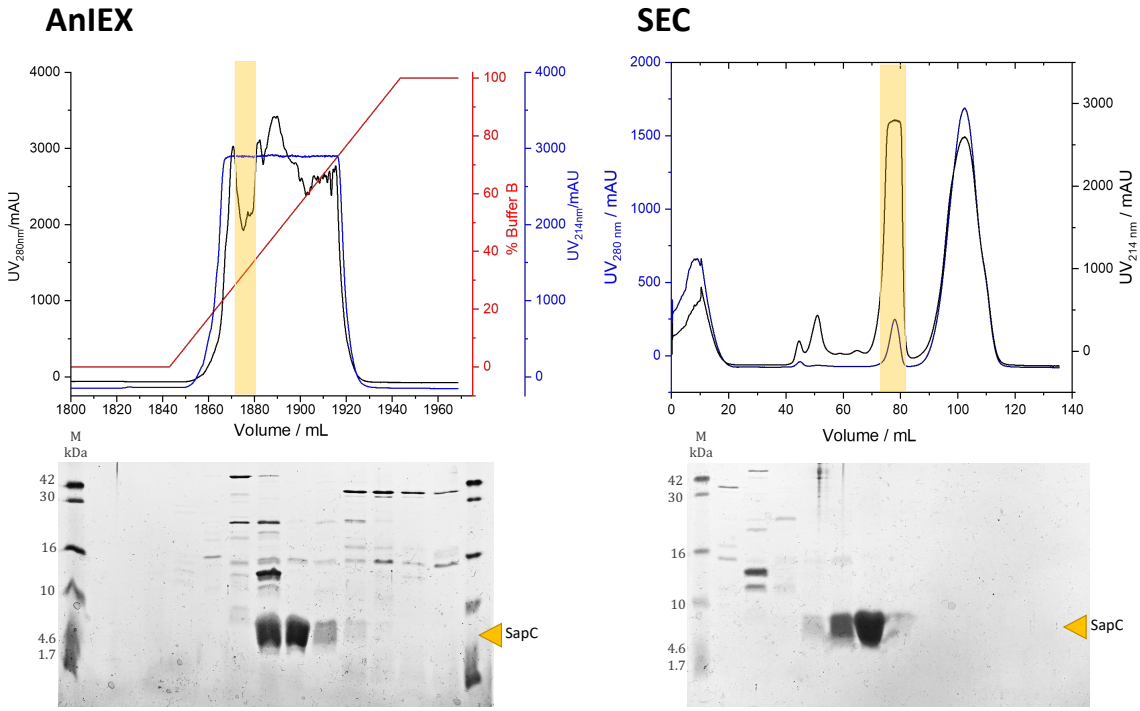


Figure A3.2: Purification of recombinant SapC expressed in Origami(DE3) cells. Following heat treatment, SapC was purified from the cell lysate by anion exchange and size exclusion chromatography. SapC has no tryptophan residues so does not give an A280 nm response and was subsequently monitored at A214 nm during purification. Chromatogram peaks containing recombinant SapC (as determined by SDS-PAGE) are highlighted in yellow.

Appendix 4: Synthesis of 1-deoxy-1-fluoro-α-D-glucose

A4.1: Synthesis

1-deoxy-1-fluoro-α-D-glucose was synthesised according to previously published procedures[878]. Briefly, 2,3,4,6-tetra-O-acetyl-α-D-glucopyranosyl fluoride was dissolved in methanol and sodium methoxide in methanol was added dropwise. The reaction was stored overnight at 4 ^{o}C. The reaction was quenched by the addition of silica gel and evaporated to dryness. The resulting slurry was purified by flash chromatography using ethyl acetate:ethanol 5:2. The resulting compound was analyse by NMR and MS to confirm its identity.

A4.2: Characterisation

NMR: ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.54 (dd, $J_{H1\text{-}F}$ = 53.9 Hz, $J_{H1\text{-}H2}$ = 2.8 Hz, 1H, H-1), 3.81 (m, 1H, H-6a), 3.70 (m, 2H, H-6b, H-5), 3.64 (dd, $J_{H2,H3}$ = 9.4 Hz, $J_{H3,H4}$ = 9.4 Hz, H-3), 3.42 (m, 2H, H-2, H-4). ^{13}C NMR (126 MHz, $CDCl_3$) δ 107.6 (d, J = 224.5 Hz, C-1), 74.7 (d, J = 3.3 Hz, C-5), 73.0 (s, C-4), 71.8 (d, J = 25.1 Hz, C-2), 69.2 (s, C-3), 60.7 (s, C-6).

Mass spectroscopy: (ESI^{+})- Calculated $C_6H_{11}O_5F$: 182.06. Found $[M+Na]^{+}$ 205.0481

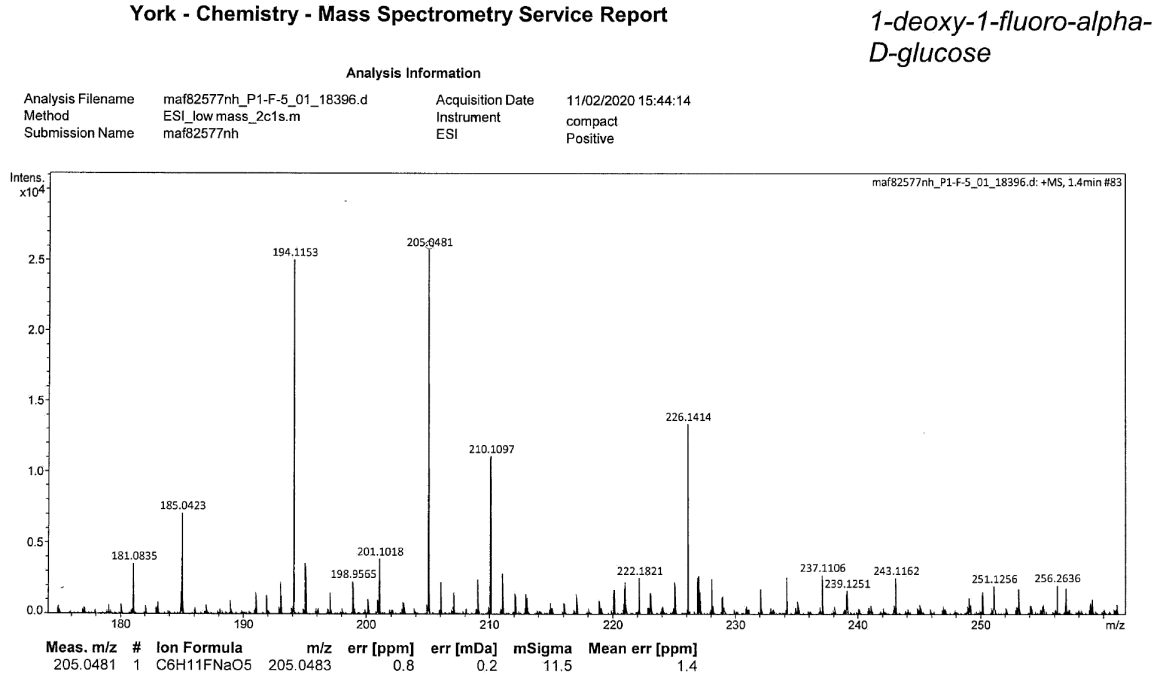


Figure A4.1: Mass spectrum of 1-deoxy-1-fluoror-alpha-D-glucose product.

Appendix 5: Publications

To date, five publications and one patent have arisen from this work:

1. R. J. Rowland, Y. Chen, I. Breen, L. Wu, W. Offen, T. J. Beenakker, Q. Su A. M. C. H. can den Nieuwendijk, M. Artola, J. M. F. G. Aerts, H. S. Overkleeft, G. J. Davies, *Design, Synthesis and Structural Analysis of Glucocerebrosidase Imaging Agents,* Chem. Eur. J., **2021,** 27, 16377-16388.

2. R. J. Rowland, L. Wu, F. Liu, G. J. Davies, *A baculoviral system for the production of human β-glucocerebrosidase enables atomic resolution analysis*, Acta Cryst. D, **2020**, 76, 565-580.

3. M. Artola, C. Hedberg, R. J. Rowland, L. Raich, K. Kytidou, L. Wu, A. Schaaf, M. J. Ferraz, G. A. van der Marel, J. D. C. Codée, C. Rovira, J. M. F. G. Aerts, G. J. Davies and H. S. Overkleeft, *α-D-Gal-cyclophellitol cyclosulfamidate is a Michaelis complex analog that stabilizes therapeutic lysosomal α-galactosidase A in Fabry disease,* Chem. Sci., **2019**, 10, 9233-9243.

4. M. Artola, C-L. Kuo, L. T. Lelieveld, R. J. Rowland, G. A. van der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *Functionalized Cyclophellitols Are Selective Glucocerebrosidase Inhibitors and Induce a Bona Fide Neuropathic Gaucher Model in Zebrafish*, J. Am. Chem. Soc., **2019**, 141, 4214-4218.

5. S. P. Schröder, C. de Boer, N. G. S. McGregor, R. J. Rowland, O. Moroz, E. Blagova, J. Reijngoud, M. Arentshorst, D. Osborn, M. D. Morant, E. Abbate, M. A. Stringer, K. B. R. M. Krogh, L. Raich, C. Rovira, J-G. Berrin, G. P. van Wezel, A. F. J. Ram, B. I. Florea, G. A. van der Marel, J. D. C. Codée, K. S. Wilson, L. Wu, G. J. Davies and H.S. Overkleeft, *Dynamic and Functional Profiling of Xylanase-Degrading Enzymes in Aspergillus Secretomes Using Activity-Based Probes,* ACS Cent. Sci., **2019**, 5, 1067-1078.

6. H. S. Overkleeft, G. J. Davies, J. M. F. G. Aerts, M. Artola, R. J. Rowland, L. Wu; *Pharmacological Chaperones for Glycosidase Treatment Therapy*; WO/2020/046132A1; **2020**. Also published as NL2021840.

ACS central science

Research Article

Cite This: *ACS Cent. Sci.* 2019, 5, 1067−1078

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

Dynamic and Functional Profiling of Xylan-Degrading Enzymes in *Aspergillus* Secretomes Using Activity-Based Probes

Sybrin P. Schröder,[†,○] Casper de Boer,[†,○] Nicholas G. S. McGregor,[‡] Rhianna J. Rowland,[‡] Olga Moroz,[‡] Elena Blagova,[‡] Jos Reijngoud,[§] Mark Arentshorst,[§] David Osborn,[∥] Marc D. Morant,[⊥] Eric Abbate,[∥] Mary A. Stringer,[⊥] Kristian B. R. M. Krogh,[⊥] Lluís Raich,[#] Carme Rovira,[#,▽] Jean-Guy Berrin,[◆] Gilles P. van Wezel,[§] Arthur F. J. Ram,[§] Bogdan I. Florea,[†] Gijsbert A. van der Marel,[†] Jeroen D. C. Codée,[†] Keith S. Wilson,[‡] Liang Wu,[*,‡] Gideon J. Davies,[*,‡] and Herman S. Overkleeft[*,†]

[†]Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2300 RA Leiden, The Netherlands

[‡]York Structural Biology Laboratory, Department of Chemistry, The University of York, York YO10 5DD, U.K.

[§]Molecular Microbiology and Biotechnology, Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands

[∥]Novozymes Inc., 1445 Drew Avenue, Davis, California 95618, United States

[⊥]Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark

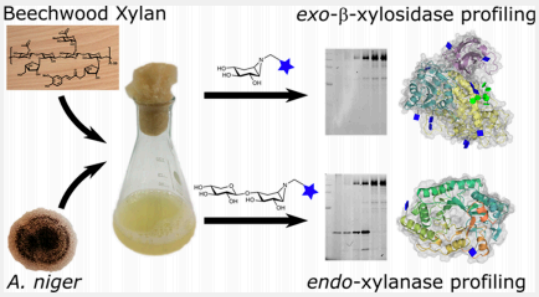
[#]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

[▽]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08020 Barcelona, Spain

[◆]Biodiversité et Biotechnologie Fongiques (BBF), UMR1163, INRA, Aix Marseille University, F-13009 Marseille, France

Supporting Information

ABSTRACT: Plant polysaccharides represent a virtually unlimited feedstock for the generation of biofuels and other commodities. However, the extraordinary recalcitrance of plant polysaccharides toward breakdown necessitates a continued search for enzymes that degrade these materials efficiently under defined conditions. Activity-based protein profiling provides a route for the functional discovery of such enzymes in complex mixtures and under industrially relevant conditions. Here, we show the detection and identification of β-xylosidases and *endo*-β-1,4-xylanases in the secretomes of *Aspergillus niger*, by the use of chemical probes inspired by the β-glucosidase inhibitor cyclophellitol. Furthermore, we demonstrate the use of these activity-based probes (ABPs) to assess enzyme−substrate specificities, thermal stabilities, and other biotechnologically relevant parameters. Our experiments highlight the utility of ABPs as promising tools for the discovery of relevant enzymes useful for biomass breakdown.



INTRODUCTION

Reflecting their biological roles as structural molecules, plant polysaccharides often show extraordinary recalcitrance to chemical and enzymatic degradation. Because of the considerable potential of plant biomass as a renewable feedstock, there is a continual hunt for new biomass-degrading enzymes that work well on complex substrates and under industrial process conditions. Fungal saprophytes have evolved to utilize plant biomass as a source of nutrients and thus produce many enzymes suitable for industrial biomass degradation.[1] While advances in DNA sequencing technologies have provided a wealth of information on the genomes of fungal saprophytes, this abundance of genomic information has not been matched by a commensurate increase in our ability to determine the functions of their encoded gene products.

Glycoside hydrolases (or glycosidases) are ubiquitous enzymes responsible for the hydrolytic breakdown of polysaccharides and glycoconjugates.[2] Along with lytic polysaccharide monooxygenases[3] and polysaccharide lyases,[4] glycosidases comprise one of the enzymatic cornerstones of biomass catabolism and are highly sought after as tools for industrial biocatalytic process development. The Carbohydrate Active EnZymes (CAZy) database (www.cazy.org) lists

Received: March 5, 2019
Published: May 24, 2019

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

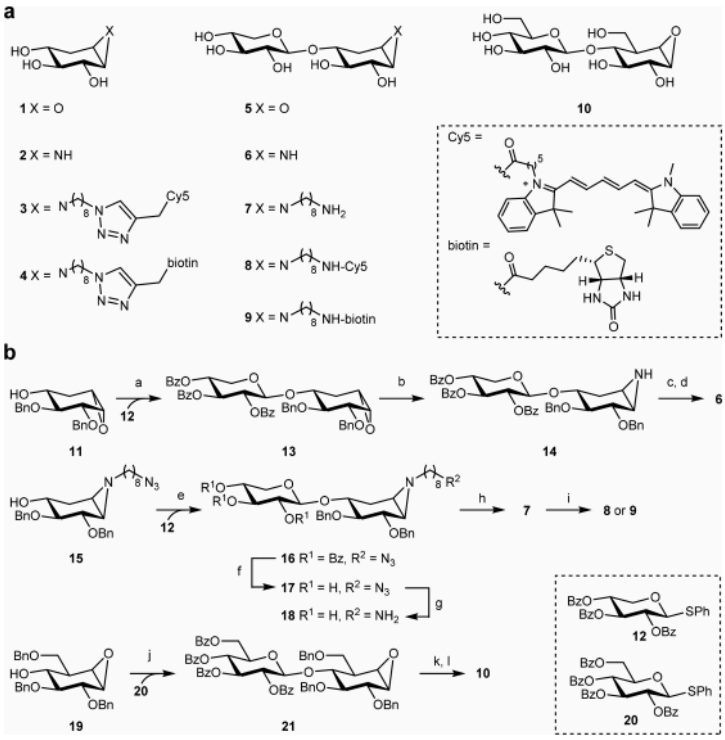


Figure 1. Design of mono-xylo and xylobiose mechanism-based inhibitors and ABPs. (a) Structures of mono-xylo and xylobiose inhibitors and ABPs used in this work. Structures of additional molecules used are shown in Supplemental Figure 3. (b) Synthetic strategy for chemical glycosylation of xylo-configured cyclophellitol derivatives. Reagents and conditions: (a) **12**, NIS, TMSOTf (cat.), DCM, 4 Å MS, −40 °C, 4 h, 17%; (b) 1. NaN_3, Et_3N·HCl, DMF, 100 °C, 16 h; 2. Polymer-bound PPh_3, MeCN, 70 °C, 16 h, 69%; (c) NaOMe, DCM, MeOH, 84%; (d) Na, ^{t}BuOH, THF, NH_3, −60 °C, 1 h, quant.; (e) **12**, NIS, TMSOTf (1.4 equiv), DCM, 4 Å MS, −40 °C, 4 h, 77%; (f) NaOMe, DCM, MeOH, rt, 16 h, 87%; (g) polymer-bound PPh_3, H_2O, MeCN, 70 °C, 20 h, 93%; (h) Li, NH_3, THF, −60 °C, 1 h, 85%; (i) Cy5-OSu or biotin-OSu, DIPEA, DMF, rt, 16 h, yield **8**: 22%, yield **9**: 25%; (j) **20**, NIS, TMSOTf (cat.), DCM, 4 Å MS, −30 to −10 °C, 2 h, 53%; (k) NaOMe, MeOH, DCM, rt, 16 h, quant.; (l) $Pd(OH)_2/C$, H_2, H_2O, MeOH, dioxane, rt, 2.5 h, quant. Detailed synthetic procedures can be found in the Supporting Information.

hundreds of thousands of putative glycosidase open reading frames (ORFs) identified by homology to known glycosidases.[5] Because of the slow pace of enzyme characterization, the vast majority of these sequences remain uncharacterized, making the identification of the right combination of enzymes for new biomass-degrading processes challenging.

Holding promise to accelerate enzyme discovery, activity-based protein profiling (ABPP) is a chemical proteomic strategy that enables the identification and quantification of specific enzymes of interest in complex mixtures.[6,7] ABPP relies on the availability of suitable activity-based probes (ABPs)—selective, covalent, and irreversible enzyme inhibitors endowed with a reporter entity. This reporter may be a fluorophore, affinity tag, or another bioorthogonal group, enabling a variety of downstream detection methods. When suitable ABPs are available, ABPP can provide high-throughput proteomic data complementary to genomic approaches, enabling the rapid identification of enzymes of interest.

Considerable work undertaken to understand the mechanisms of glycosidase activity has facilitated the development of many classes of glycosidase inhibitors suitable for use as ABPs. It is now widely appreciated that glycosidases carry out bond cleavage primarily through acid/base assisted reaction mechanisms leading to either net inversion or net retention of anomeric configuration, the latter occurring via the formation of a key covalent glycosyl-enzyme intermediate (Supplemental Figure 1a).[2] The groups of Withers, Vocadlo, and Bertozzi have pioneered tagged fluoroglycosides, which form trapped glycosyl-enzyme intermediates, as ABPs for retaining exo- and endo-glycosidases.[8−11] In a similar vein, Lo and colleagues have reported the design of ABPs that act in situ within a glycosidase active site to generate reactive electrophiles that label nearby nucleophiles.[12] Wright and co-workers have recently reported the use of multiple ABP chemistries to study lignocellulose degradation by *Trichoderma reesei* and *Clostridium thermocellum*.[13−16]

We have previously reported a suite of ABPs inspired by cyclophellitol, a mechanism-based β-glucosidase inhibitor isolated from *Phellinus sp.*[17] Via its reactive epoxide ring, cyclophellitol can react with β-glucosidases to produce a stable enzyme-adduct (Supplemental Figure 1b).[18] This generic inhibition modality has subsequently been expanded to produce a panel of epoxide and aziridine tools for the study of a number of retaining glycosidases, primarily in biomedical contexts.[19] Cyclophellitol-derived ABPs have been used to probe β- and α-glucosidases in tissue samples from Gaucher[20] and Pompe[21] disease patients, respectively. We have also reported the use of cyclophellitol-derived ABPs in the

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

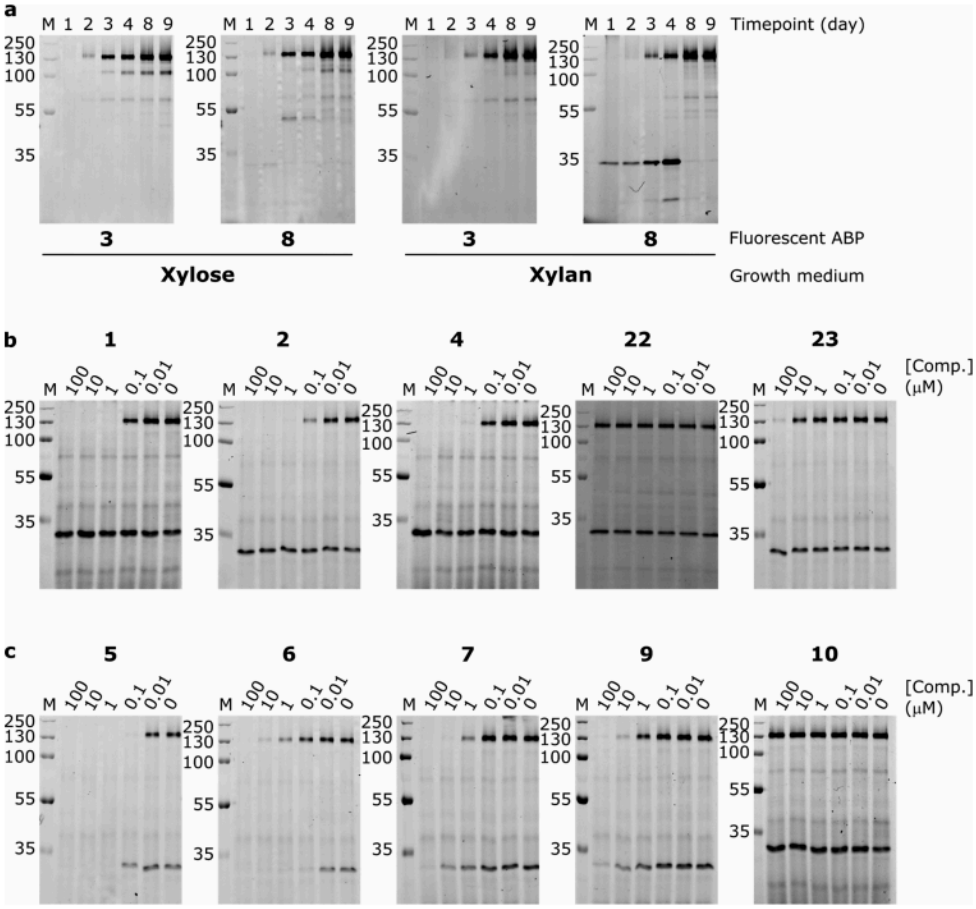


Figure 2. Fluorescent scans of Cy5 ABP labeled *A. niger* secretomes induced by xylose or BX. (a) Secretome samples were collected at indicated time points and labeled with the indicated probes. A ~30 kDa band labeled by **8** was only present in xylan-induced secretomes. (b) Day 6 BX-induced secretomes preincubated with "monosaccharide" competitors, before labeling with **8**. Mono-xylo competitors **1**, **2**, and **4** inhibit labeling of the ~130 kDa band, suggesting this band corresponds to a β-xylosidase. Glucose configured competitors **22** and **23** have little effect on labeling. (c) BX-induced secretomes preincubated with "disaccharide" competitors, before labeling with **8**. Xylobiose competitors **5**, **6**, **7**, and **9** inhibit labeling of both ~130 kDa and ~30 kDa bands, while cellobiose configured **10** shows no effect on labeling. Gel molecular weight markers are given in kilodaltons. Comp. − competitor.

discovery and quantification of α-L-fucosidases,[22] 6-phospho-β-D-glucosidases,[23] and β-D-glucuronidases in the context of heparan degradation.[24]

Here we present cyclophellitol-derived ABPs configured for detection of endo-β-1,4-xylanases (hereafter β-xylanases) and β-xylosidases involved in biomass breakdown. Xylan is a β-1,4-linked polymer of D-xylose decorated with α-D-glucuronopyranose, 4-*O*-methyl-α-D-glucuronopyranose, and α-L-arabinofuranose residues, which may be further modified with acetate, ferulate, and coumarate esters (Supplemental Figure 2). β-Xylosidases and β-xylanases are considered to be the principal enzymes responsible for xylan degradation, as these enzymes directly work to hydrolyze the xylan polysaccharide backbone. Initial xylan degradation by β-xylanases releases short xylose oligosaccharides from polymeric xylan, which are then further hydrolyzed to monomeric xylose by β-xylosidases.[25]

Xylan is the main hemicellulosic component of plant biomass,[26] hence its breakdown is of major importance to the biofuel, animal feed, and pulp and paper industries, and there is a continual search for novel β-xylanases and β-xylosidases with industrially useful properties.[27] Using the model fungal saprophyte *Aspergillus niger*,[28,29] we demonstrate here the application of xylose-configured ABPs for the rapid detection and identification of β-xylosidase and β-xylanases in complex fungal secretome samples. Our results illustrate the applicability of cyclophellitol-derived ABPs beyond the biomedical arena, into the field of biotechnologically relevant enzyme discovery and functional dissection.

RESULTS

Design and Synthesis of Xylosidase and Xylanase Probes. We have previously described the synthesis of *xylo*-cyclophellitol epoxide **1** and aziridine **2**, as mechanism-based β-xylosidase inhibitors.[30] Following developed strategies, compound **2** was equipped with a Cy5 fluorophore or biotin in order to produce β-xylosidase ABPs **3** and **4** respectively

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

(Figure 1a; see Supporting Information for detailed synthetic procedures).

Our previously reported cyclophellitol-derived ABPs have all consisted of monomeric cyclitols that mimic monosaccharide substrates. While these "monosaccharide" ABPs react well with exo-glycosidases, endo-glycosidases typically remain inert toward these molecules, due to a lack of sufficient interactions within the larger endo-glycosidase active sites. We hypothesized that nonreducing end extension of *xylo*-cyclophellitol ABPs with a β-1,4-linked xylose would enable profiling of endo-xylanase activity. Thus, *epi-xylo*-cyclophellitol **11** was glycosylated with thiophenyl donor **12** using *N*-iodosuccinimide (NIS) and catalytic trimethylsilyl trifluoromethanesulfonate (TMSOTf) to afford pseudodisaccharide epoxide **13** (Figure 1b). This epoxide was converted to aziridine **14** by treatment with azide followed by a Staudinger-type ring closure[31] and deprotected to give untagged "disaccharide" aziridine **6**.

We next constructed alkylated xylobiose cyclophellitol aziridines via chemical glycosylation of a protected mono-xylo cyclophellitol aziridine derivative.[30] Alkyl-aziridine **15** was readily amenable to chemical glycosylation with donor **12** and NIS/TMSOTf, giving pseudodisaccharide **16** in good yield. Debenzoylation, azide reduction, and debenzylation of **16** under Birch conditions afforded alkyl-aziridine inhibitor **7**, which we conjugated with a Cy5 fluorophore or biotin moiety to produce ABPs **8** and **9**, respectively. The same methodology was applied to construct *cellobiose*-cyclophellitol **10** via glycosylation of acceptor **19** with donor **20**, demonstrating the versatility of chemical glycosylation for the synthesis of cyclophellitol-derived "disaccharides".

ABP Labeling of *Aspergillus niger* Secretomes. *A. niger* strain N402 (a derivative of NRRL3/ATCC 9029/CBS 120.49)[32] was grown in minimal medium containing either 50 mM (0.75% w/v) xylose or 1% w/v beechwood xylan (BX) as the sole carbon source. Secretome samples from each culture were taken at 1, 2, 3, 4, 8, and 9 day time points, adjusted to pH 4.5 using phosphate/citrate buffer, and screened using Cy5 fluorescent ABPs **3** or **8** without further sample concentration.

ABP labeling experiments revealed a dynamic secretome, with labeled bands generally increasing in intensity over the course of the experiment (Figure 2a). In contrast to the ready visualization of bands following ABP labeling by **8**, Coomassie staining of secretome gels was not sensitive enough to detect any bands, even at day 9 where ABP labeling was most intense. Silver staining resolved some faint bands from day 8, showing accumulation of protein within the secretome over time. However, these bands were still well below the intensities observed for ABP labeling (Supplemental Figure 4).

The patterns of ABP labeled enzymes were specific to the carbon source present in the culture medium. Secretomes from cultures induced by xylose and BX both presented a strong band at ~130 kDa when probed by either mono-xylo ABP **3** or xylobiose ABP **8**, whereas a ~30 kDa band was observed only in BX-induced secretomes labeled by **8**, indicating a specific enzyme induced by growth on BX. The ~30 kDa band disappeared from BX-induced secretomes by day 8, whereas the ~130 kDa band continued to grow in intensity up to the end of the time course. This band pattern is consistent with a model of initial xylan degradation by a β-xylanase (~30 kDa band), followed by subsequent breakdown of the released xylobiose by an exo-acting β-xylosidase (~130 kDa band). Day 4 BX-induced secretome (which contains both ~30 kDa and ~130 kDa bands) was able to efficiently hydrolyze the artificial fluorogenic substrates 4-methylumbelliferyl (4MU)-β-D-xyloside and 4MU-β-D-xylobioside, confirming the ability of this mixture to degrade xylosidic substrates (Supplemental Figure 5). Labeling of a β-xylosidase by **8** can be readily explained by initial enzymatic cleavage of the terminal xylose of **8**, followed by reaction with the subsequently liberated mono-xylo ABP **8a** (Supplemental Figure 3b). A broad pH range (pH 4–7) was observed for labeling of the ~30 kDa band by **8**, while the ~130 kDa band showed a narrower labeling pH range, peaking around pH 4.5 (Supplemental Figure 6a).

Preincubation of day 6 xylan-induced secretomes with mono-xylo-epoxide **1** prior to treatment with ABP **8** abolished ~130 kDa labeling in a concentration-dependent fashion, but did not affect labeling of the ~30 kDa band. Furthermore, these secretomes lost the ability to hydrolyze 4MU-β-D-xyloside, but not 4MU-β-D-xylobioside, strongly suggesting the presence of an exo-acting β-xylosidase at ~130 kDa (Supplemental Figure 5). We also observed selective competition of the ~130 kDa band by mono-xylo aziridines **2** and **4**, reflecting their similar selectivity to **1**. No labeling inhibition was seen after preincubation with cyclophellitol **22** (Supplemental Figure 3a), although some loss of intensity for the ~130 kDa band was seen after preincubation with 100 μM cyclophellitol aziridine **23**, presumably due to off-target reactivity of **23** at higher concentrations (Figure 2b).

Both the ~30 kDa and ~130 kDa bands were inhibited by preincubation with xylobiose epoxide **5** or nonfluorescent xylobiose aziridines **6**, **7**, or **9**, indicating that the ~30 kDa band likely corresponded to a β-xylanase. Preincubation with **5** also abrogated the ability of the secretome to hydrolyze both 4MU-β-D-xyloside and 4MU-β-D-xylobioside (Supplemental Figure 5). No labeling inhibition for any band was observed following preincubation with cellobiose epoxide **10** (Figure 2c). Interestingly, while untagged xylobiose aziridine **6** competed with labeling of the ~30 kDa band with similar potency to xylobiose epoxide **5**, functionalization of the aziridine with an alkyl tail (e.g., **7** and **9**) caused a marked loss of potency for inhibition of ~30 kDa band labeling, indicating that alkylation of the aziridine nitrogen was unfavorable for reactivity with this enzyme.

Identification of ABP Labeled Bands. On the basis of their observed molecular masses and competition profiles, we postulated that the ~130 kDa band most likely corresponded to the *A. niger* GH3 exo-β-xylosidase XlnD (peptide molecular weight 85 kDa, but with 15 N-glycosylation sites),[33] and the ~30 kDa band most likely corresponded to the *A. niger* GH10 β-xylanase XlnC. Treatment of day 6 xylan-induced secretome with peptide N-glycosidase F (PNGase F) after labeling by **8** caused electrophoretic migration of the ~130 kDa band to shift below 100 kDa (Supplemental Figure 6b), consistent with removal of multiple N-glycans from XlnD.

The identity of the labeled enzymes was investigated by an activity-based protein pulldown. Treatment of day 4 xylan-induced secretome with biotinylated ABPs **4** or **9** resulted in the labeling of both ~130 kDa and ~30 kDa bands with biotin, as evidenced by western blotting using HRP-conjugated streptavidin. The patterns of labeling produced by **4** and **9** were identical to those produced by equivalent Cy5-tagged ABPs **3** and **8** (Supplemental Figure 7).

Pulldown of labeled enzymes onto streptavidin beads followed by on-bead tryptic digestion and LC-MS/MS analysis

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

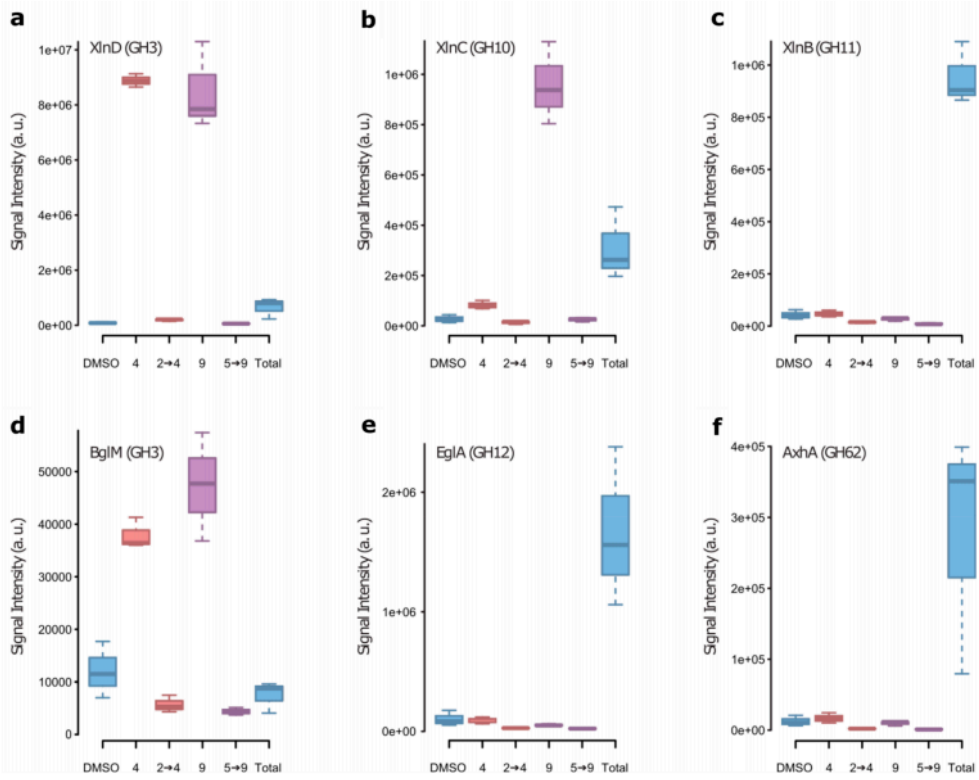


Figure 3. Peptide signal intensities observed following activity-based protein pulldown from a xylan-induced *A. niger* secretome. Total MS signal intensity from nonconflicting peptides is shown for (a) XlnD (*β*-xylosidase), (b) XlnC (GH10 *β*-xylanase), (c) XlnB (GH11 *β*-xylanase), (d) BglM (glucosidase), (e) EglA (glucanase), and (f) AxhA (arabinofuranosidase). CAZy GH family names are given next to each enzyme. Box plots show the range of peptide intensities measured in three replicates each of a negative control pulldown with no ABP (DMSO), a *β*-xylose-configured probe pulldown with (**2** → **4**) and without (**4**) competitor pretreatment, a *β*-xylobiose-configured probe pulldown with (**5** → **9**) and without (**9**) competitor pretreatment, and the total secretome (Total). Full proteomics results can be found in Supplemental File 1. a.u. − arbitrary units.

of the resulting peptides supported our hypothesized identification of the ∼130 kDa and ∼30 kDa bands. Each protein identified in the total secretome was quantified using spectral intensity-based relative quantification using nonconflicting peptides to assess its abundance in the pulldown, compared to its abundance in either the total secretome, a DMSO only no-probe pulldown control, or a competitor-inhibited pulldown control following treatment by **2** or **5** (Figure 3, Supplemental File 1). As expected, XlnC (NRRL3_08708; Genbank: AIC36735.1) and XlnD (NRRL3_02451; Genbank: AIC36733.1) were clearly abundant in the pulldown following treatment with **9**, whereas XlnD, but not XlnC, was highly enriched following treatment with **4**. Conversely, EglA (NRRL3_00819; Genbank: CAK46524.1), a retaining endo-glucanase, and AxhA (NRRL3_08707; Genbank: AIC36734.1), an inverting arabinofuranosidase, were both abundant in the secretome, but not pulled down by either **4** or **9**. Although XlnD was the only *β*-xylosidase found in the secretome, BglM (NRRL3_10133; Genbank: GAQ41973.1), a low abundance GH3 *β*-glucosidase in total secretome, was also enriched by **4** and **9**, suggesting some off-target reactivity between GH3 enzymes. Surprisingly, XlnB (NRRL3_01648; Genbank: CAK43456.1), a GH11 *β*-xylanase abundant in the total secretome digest, was not detected in the pulldown (discussed further below). Pretreatment of the secretome with either **2** or **5** also caused a significant loss of peptide abundance for XlnD, XlnC, and BglM, demonstrating that these enzymes were labeled selectively. Comparing these hits to the bands observed in Supplemental Figure 7, we can account for the major band observed at ∼130 kDa as XlnD and the major band at ∼30 kDa as XlnC. BglM (∼80 kDa plus N-glycosylation), present at a much lower level, would likely migrate similarly to XlnD.

Kinetics of GH3 and GH10 ABP Labeling. To better understand the structure−activity relationships governing reactivity of our ABPs, we tested **1**−**6**, **8**, and **9** (encompassing mono-xylo and xylobiose epoxides, untagged aziridines, and biotin or Cy5 functionalized aziridines) for their ability to react with recombinant GH3 and GH10 enzymes. Kinetics of probe/inhibitor reactivity were determined by measuring the inhibition of enzymatic activity over time, since covalent labeling of a glycosidase nucleophile necessarily leads to enzyme inhibition.

In the absence of recombinant *A. niger* xylosidase/xylanase XlnD and XlnC, we utilized two enzymes from closely related *Aspergillus* species: XlnD from *A. nidulans* (GH3; Genbank: CAA73902.1; hereafter referred to as *Anid*XlnD), which shares 66% sequence identity with *A. niger* XlnD, and ASPACDRAFT_127619 from *A. aculeatus* (GH10; Genbank: XP_020051463.1; catalytic domain only), which shares 89%

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

identity with *A. niger* XlnC. *Anid*XlnD and ASPACDRAFT_127619 efficiently hydrolyzed 4MU-β-D-xyloside and 4MU-β-D-xylobioside (Supplemental Figure 8a), verifying their activities as β-xylosidase and β-xylanase, respectively.

We observed substantial variability in the rates of reactivity across all enzyme/inhibitor combinations tested (Table 1, Supplemental Figure 8b). β-Xylosidase *Anid*XlnD was inhibited by mono-xylo probes/inhibitors **1**, **2**, **3**, and **4** with k_{inact}/K_I values of 0.01, 0.14, 0.007, and 0.007 min^{-1} μM^{-1} respectively, whereas these molecules were totally ineffective against ASPACDRAFT_127619. In contrast, xylobiose probes/inhibitors **5**, **6**, **8**, and **9** were effective inhibitors of ASPACDRAFT_127619, with k_{inact}/K_I values of 0.26, 0.45, 0.02, and 0.27 min^{-1} μM^{-1} respectively, and also inhibited *Anid*XlnD, with apparent k_{inact}/K_I values of 0.042, 0.004, 0.002, and 0.0002 min^{-1} μM^{-1} respectively. In line with gel labeling experiments (Figure 2b,c), the values from inhibition kinetics showed aziridine alkylation was typically detrimental for ABP/inhibitor reactivity. Approximately 20 times lower k_{inact}/K_I values were measured for the Cy5-tagged aziridines **3** and **8** (against *Anid*XlnD and ASPACDRAFT_127619 respectively), compared to their cognate untagged aziridines **2** and **6**. In contrast, biotin-tagged xylobiose aziridine **9** showed only a 2-fold lower k_{inact}/K_I compared to **6**, with poor k_{inact} for **9** largely offset by a substantial decrease in K_I. Taken together, our results show that the identity of the linker moiety can markedly affect cyclophellitol-derived ABP/inhibitor reactivity, at least in the case of retaining β-xylosidases and β-xylanases.

Table 1. Kinetic Parameters for Covalent Inhibition of *Anid*XlnD and ASPACDRAFT_127619 by Mono-xylo and Xylobiose ABPs/Inhibitors[a]

	*Anid*XlnD (GH3)			ASPACDRAFT_127619 (GH10)		
compound	K_I (μM)	k_{inact} (min^{-1})	k_{inact}/K_I (min^{-1} μM^{-1})	K_I (μM)	k_{inact} (min^{-1})	k_{inact}/K_I (min^{-1} μM^{-1})
1	n.d.	n.d.	0.010	n.d.	n.d.	n.d.
2	36.7	5.02	0.14	n.d.	n.d.	n.d.
3	n.d.	n.d.	0.007	n.d.	n.d.	n.d.
4	64.7	0.46	0.007	n.d.	n.d.	n.d.
5	62.1	2.61	0.042	44.7	11.82	0.26
6	71.7	0.26	0.004	1.0	0.46	0.45
8	167.6	0.35	0.002	0.7	0.013	0.02
9	53.3	0.012	0.0002	0.04	0.011	0.28

[a]For *Anid*XlnD with **1** and **3**, it was not possible to obtain separate k_{inact} and K_I parameters. Only the combined k_{inact}/K_I parameter is shown for these cases (see methods). n.d.: not determinable.

Structural Basis of GH3 and GH10 ABP Labeling. To complement biochemical characterization of ABP reactivity from inhibition kinetics, we sought to obtain structural insights into ABP interactions with recombinant GH3 and GH10 enzymes via X-ray crystallography of reacted enzyme−ABP complexes.

The three-dimensional structure of apo *Anid*XlnD (PDB accession code 6Q7I) was solved at 1.50 Å resolution by molecular replacement using a GH3 from *Hypocrea jecorina* (*Trichoderma reesei*) as a search model (PDB: 5A7M). *Anid*XlnD crystallized with two molecules in the asymmetric unit (ASU). Each protein monomer showed a typical 3 domain GH3 structure, with a predominantly α-helical *N*-terminal domain containing the enzyme active site (residues 19−410), a 3-layer $\alpha/\beta/\alpha$ sandwich domain (411−638), and a C-terminal fibronectin type III (FnIII) like domain (639−788). As with *A. niger* XlnD, *Anid*XlnD contains a large number of N-glycosylation sites. While endo-glycosidase H (EndoH) treatment prior to crystallization resulted in most of the observed *Anid*XlnD N-glycans being truncated to a single GlcNAc, one site (Asn140) appeared in crystallo as an extensive high-mannose N-glycan, with seven or eight mannose sugars modeled in the two independent monomers of the ASU (Figure 4a; Supplemental Table 1).

To generate an ABP complex, *Anid*XlnD was cocrystallized with xylobiose aziridine **8**, and the resulting structure was solved at 2.14 Å resolution (PDB accession code 6Q7J). As expected, the ABP ligand observed within the *Anid*XlnD active site was not **8** itself, but mono-xylo aziridine **8a** released by cleavage of the terminal xylose of **8**. Reacted **8a** was bound to the enzyme nucleophile Asp307 in a $^4C_1$ chair conformation, with direct H-bonding interactions to Arg183, Lys222, His223, and Glu107. The C5 methylene of the xylose probe resided ~4.5 Å from Trp105, consistent with a C−H-π stacking interaction at this position (Figure 4b). Although the *N*-linked fluorophore tail of **8a** was observed to project out of the enzyme active site, only the first five carbon atoms of the linker were modeled due to increasing disorder along the alkyl chain.

The crystal structure of unliganded GH10 ASPACDRAFT_127619 catalytic domain (PDB accession code 6Q8M) was solved at 1.42 Å resolution using a *Fusarium oxysporum* GH10 (PDB: 3U7B) as a molecular replacement search model, with which ASPACDRAFT_127619 shares 51% sequence identity. ASPACDRAFT_127619 crystallized with two molecules in the ASU, each comprising a single $(\beta/\alpha)_8$ barrel containing an extended central cleft capable of binding xylan oligosaccharide chains. Two Asn-linked GlcNAcs resulting from EndoH digest of N-glycans were observed on each protein monomer (Figure 4c).

We were unable to obtain a complex of ASPACDRAFT_127619 with **8**, possibly due to the slow reactivity of **8** with ASPACDRAFT_127619 (as measured by inactivation kinetics; Table 1). However, cocrystallization of ASPACDRAFT_127619 with the substantially more reactive xylobiose epoxide **5** furnished a 1.76 Å complex with a single molecule of **5** in each enzyme active site (PDB accession code 6Q8M). The reacted epoxide warhead bound to the ASPACDRAFT_127619 nucleophile Glu260 in a $^4C_1$ chair conformation, with the terminal xylose also adopting a $^4C_1$ chair in the −2 subsite of the binding cleft.[34] Reacted **5** made direct H-bonds to the side chains of Asn152, His104, Lys71, Asn68, and Trp300 of ASPACDRAFT_127619, as well as a C−H-π stacking interaction between the −1 subsite C5 methylene and Trp308 (Figure 4d).

We also calculated homology models for *A. niger* XlnD and XlnC using the Phyre2 server[35] and the apo *Anid*XlnD and ASPACDRAFT_127619 crystal structures as templates. In line with the close sequence identity between the *A. niger* enzymes and their homologues, we obtained 100% confidence models for both XlnD and XlnC, which showed high structural similarity to *Anid*XlnD and ASPACDRAFT_127619 respectively. The *A. niger* XlnD model displayed full conservation of active site residues compared to *Anid* XlnD, while the XlnC model showed near complete conservation of active site residues compared to ASPACDRAFT_127619 (Trp308 of ASPACDRAFT_127619 substituted by an Arg(302) in the XlnC model; Supplementary Figure 9a,b). The close similarity

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

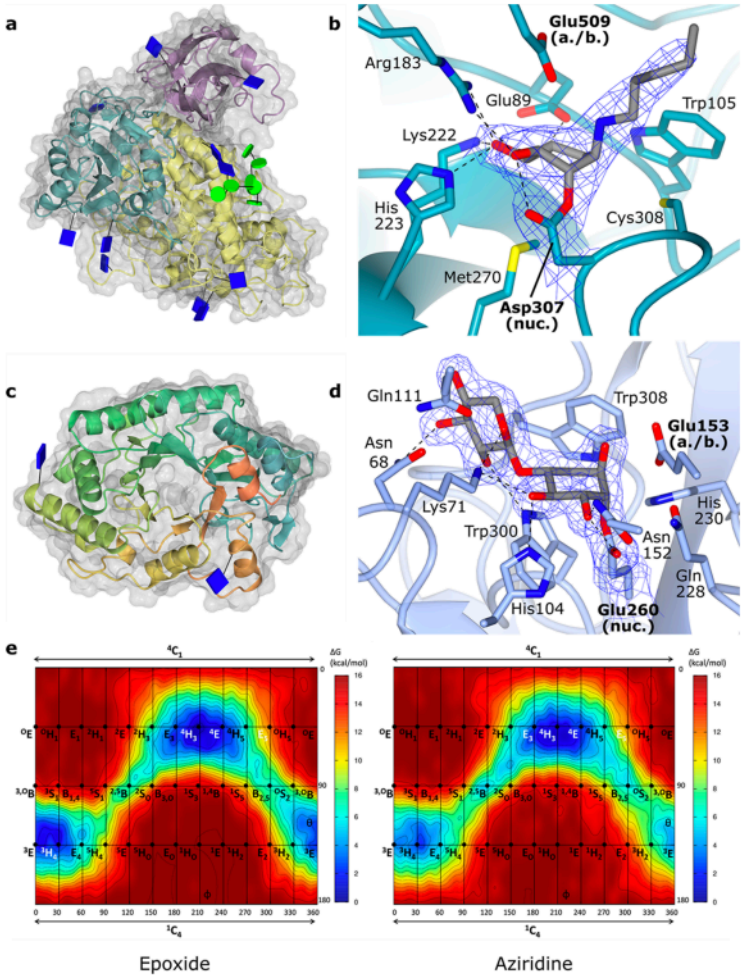


Figure 4. Crystal structures of representative GH3 β-xylosidase and GH10 β-xylanase enzymes. (a) Ribbon and surface representation of the structure of GH3 β-xylosidase *Anid*XlnD from *A nidulans*. N-glycans are shown using the Glycoblocks[42] representation (blue squares and green circles). (b) Active site of *Anid*XlnD bound to **8a** after hydrolysis of the terminal xylan of **8**. Electron density is REFMAC5 maximum-likelihood/σ_A weighted $2F_o-F_c$ contoured to 1.1 σ (0.24 e$^-$/Å^3). (c) Ribbon and surface representation of the GH10 β-xylanase ASPACDRAFT_127619 core catalytic domain. (d) Active site of ASPACDRAFT_127619 bound to reacted **5**. Electron density shown is REFMAC5 maximum-likelihood/σ_A weighted $2F_o-F_c$ contoured to 1.1σ (0.32 e$^-$/Å^3). (e) Conformational FELs for mono-xylo-epoxide (**1**) and aziridine (**2**) ABPs indicate these molecules have a ground state conformation centered around $^4H_3$.

between the *A. niger* enzyme models and the *A. nidulans*/*aculeatus* crystal structures suggests that structural and kinetics data obtained using the *Aspergillus* homologues are likely to reflect activities of the native *A. niger* enzymes.

One unexpected result from our proteomics experiments was the lack of ABP reactivity displayed by the *A. niger* GH11 β-xylanase XlnB, which was present in total xylan-induced secretomes but not pulled down by biotin ABP **9** (Figure 3c). This behavior contrasted sharply with the strong ABP reactivity displayed by the GH10 β-xylanase XlnD. Although GH10 and GH11 β-xylanases catalyze similar reactions, these two CAZy families utilize distinct conformational itineraries to process their substrates during hydrolysis.[36] GH10 β-xylanases are proposed to utilize a $^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$ conformational itinerary proceeding from the Michaelis complex to formation of the covalent intermediate[37,38] (a $^4C_1$ covalent intermediate conformation is reflected by our structures of ASPACDRAFT_127619 with **5**). In contrast, GH11 β-xylanases are proposed to operate via a $[^{2,5}B]^{\ddagger}$ transition state conformation, likely accessed from a ^2S_O skew boat conformation in the Michaelis complex.[39]

To rationalize the basis of ABP reactivity differences between GH10 and GH11 enzymes, we calculated conformational free energy landscapes (FELs) for both xylo-epoxide (**1**) and aziridine (**2**) ABP warheads, using ab initio metadynamics with Cremer–Pople puckering coordinates θ and φ as collective variables (Figure 4e).[39,40] These FELs showed both xylo-epoxide and xylo-aziridine ABPs to favor a $^4H_3$ ground state conformation, matching the proposed $[^4H_3]^{\ddagger}$ transition state conformation utilized by GH10 enzymes and rationalizing the strong reactivity of our ABPs toward this family. In contrast, while $^{2,5}B$ (the transition state conformation

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

utilized by GH11 enzymes) was still a relatively low energy ABP conformer, it was calculated to be approximately 6 kcal·mol^{-1} higher in energy than $^{4}H_{3}$, which may, at least in part, contribute to the poor reactivity of xylobiose ABPs with GH11 enzymes. Longer xylan mimicking ABPs may also be required to effectively label GH11 enzymes. This is consistent with the surprising observation that secretomes pretreated with **5** do not hydrolyze 4MU-β-D-xylobioside (Supplemental Figure 5), suggesting XlnB is not active on this substrate.

Supporting the notion that the $^{2,5}B$ conformation is unfavored but not necessarily inaccessible for xylo-configured ABPs, we were able to obtain an in crystallo complex of **5** reacted with the *A. niger* GH11 enzyme XynA[41] (41% sequence identity with XlnB; PDB accession code: 6QE8), although extended ligand soaking times (>24 h) were required to achieve acceptable active site occupancy. The reacted warhead of **5** adopted a distorted $^{5}S_{1}$ skew boat bound to the XynA catalytic nucleophile Glu106, supporting computational predictions proposing $^{5}S_{1}$ as the covalent intermediate conformation utilized by GH11 enzymes[39] (Supplemental Figure 9c,d; Note, herein we have numbered XynA residues according to the full protein sequence, including the cleaved 27 amino acid signal peptide. Glu106 in our structure corresponds to Glu79 in Vandermarliere et al.[41]).

Substrate Competition of ABP Labeling. Competitive ABP labeling is a well-established technique for the identification of enzyme ligands or substrates.[6] In this experiment format, a molecule that can interact with the enzyme is allowed to compete with the ABP for active site binding, leading to concentration-dependent loss of ABP labeling intensity. To establish a proof of principle for ABP-mediated discovery of substrates and inhibitors for biomass-degrading enzymes, we investigated whether xylo-ABP labeling could be competitively inhibited by a set of known xylanase ligands.

Initial competitive ABP experiments were carried out using the recombinant GH10 β-xylanase ASPACDRAFT_127619. Labeling reactions were carried out for 10 min at room temperature, using a low amount (100 nM) of ABP **8** to maximize sensitivity of the ABP-enzyme interaction to competition. Under these conditions, ASPACDRAFT_127619 labeling by **8** was effectively inhibited by 4MU-β-D-xylobioside with an IC_{50} of 204 μM. This IC_{50} value compared well with the K_M for hydrolysis of 4MU-β-D-xylobioside by ASPACDRAFT_127619 (279 μM), suggesting labeling inhibition reflected genuine active site competition of **8** by 4MU-β-D-xylobioside (Figure 5a; Supplemental Figure 8a).

We next explored competing ASPACDRAFT_127619 labeling using polysaccharide substrates. Labeling by **8** was effectively inhibited by solubilized BX (the substrate with which the *A. niger* secretome had been induced; Figure 5b), as well as by solubilized wheat arabinoxylan (WAX; Figure 5c), with IC_{50}'s of ~57 μg/mL and ~67 μg/mL respectively. No effect on labeling was observed after competition with either mixed linkage barley β-glucan or konjac glucomannan, which are not β-xylanase substrates (Supplemental Figure 10a,b). Analysis of BX and WAX hydrolysis kinetics by ASPACDRAFT_127619 using a bicinchoninic acid (BCA) reducing end assay gave apparent K_M values of 196 and 357 μg/mL respectively (Supplemental Figure 10c). The more potent IC_{50} values measured for BX and WAX ABP labeling inhibition, compared to the K_M values for their hydrolysis, may reflect ASPACDRAFT_127619's interaction with "nonconsensus" xylan sites, which are not turned over enzymatically but can still compete with **8** for active site occupancy.

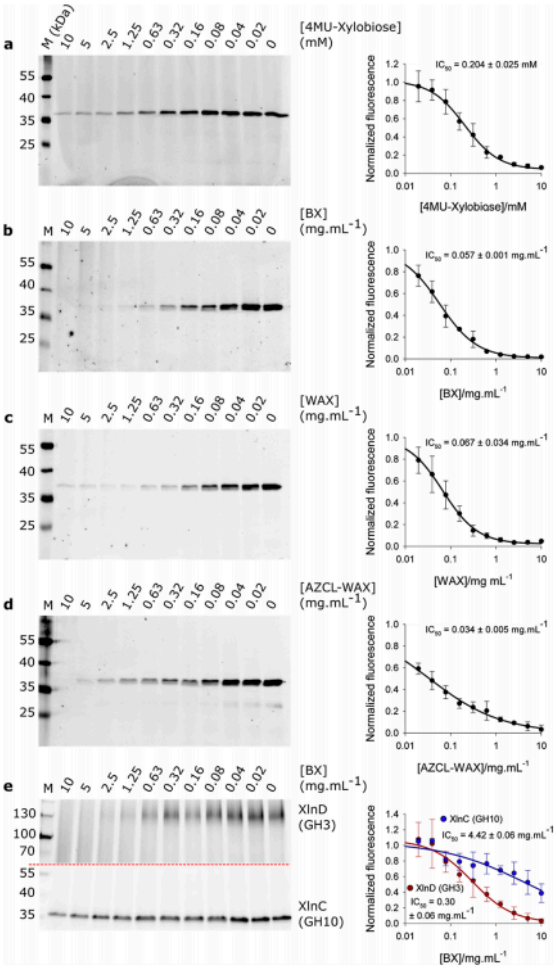


Figure 5. ABP labeling of β-xylosidases and β-xylanases is inhibited by competition with xylanase substrates. ASPACDRAFT_127619 labeling by **8** is inhibited by competition with (a) 4MU-β-D-xylobioside, (b) BX, (c) WAX, and (d) insoluble AZCL-linked WAX. (e) ABP Labeling of XlnC and XlnD in *A. niger* secretomes is competed by BX. This gel has been contrast adjusted at the dotted line to best show change in labeling intensity for each band. Data points are mean ± standard deviation from three (e) or four (a–d) technical replicates.

Since xylans are typically found in insoluble form, we also investigated whether labeling by **8** could be inhibited by insoluble xylan substrates. Using AZCL dye conjugated insoluble WAX as a competitor, clear inhibition of ABP labeling was observed with an IC_{50} of 34 μg/mL (Figure 5d). Interaction of AZCL-WAX with ASPACDRAFT_127619 was confirmed by the colorimetric release of solubilized AZCL upon enzymatic degradation of the insoluble xylan, although the apparent K_M for this reaction was substantially higher (11.44 mg/mL) than the IC_{50} from competitive ABP labeling (Supplemental Figure 10d). As with hydrolysis of soluble BX and WAX, this discrepancy between K_M and IC_{50} may reflect

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067–1078

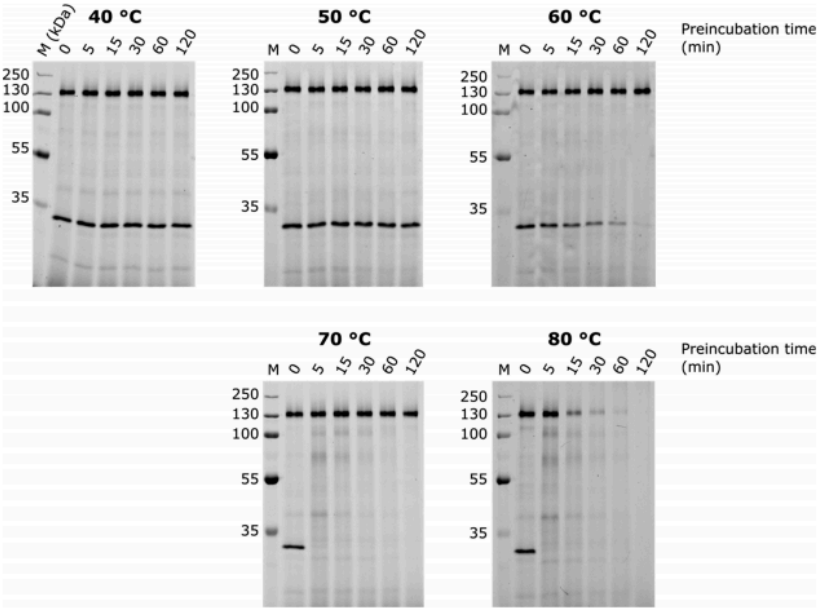


Figure 6. ABP analysis of enzyme thermal denaturation in *A. niger* secretomes. Both XlnC and XlnD bands are stable up to 60 °C, whereupon XlnC slowly loses activity with increased heating times. XlnC activity was completely abolished within 5 min at 70 °C, whereas extended heating at 80 °C was required to abolish XlnD activity.

unproductive enzyme−substrate binding events that do not result in turnover.

Lastly, we investigated whether we could compete ABP labeling of enzymes within *A. niger* secretomes. To achieve optimum labeling of secretome enzymes, the competitive labeling reaction was adjusted to 20 min incubation at 37 °C with 1 μM ABP **8** (Supplemental Figure 10e). We controlled for changes to IC_{50} values arising from these more stringent labeling conditions by reassessing competition of ASPACDRAFT_127619 labeling by BX. Under the more stringent labeling conditions, IC_{50} for competition of **8** by BX was ~10× less potent than previously calculated (596 μg/mL vs. 57 μg/mL), consistent with a 10× increase in the concentration of **8** (Figure 5b vs. Supplemental Figure 10f).

BX inhibited labeling of both XlnD and XlnC in BX-induced secretomes, with IC_{50} values of ~300 μg/mL and ~4.42 mg/mL respectively (Figure 5 e). It is likely that some of the observed XlnD labeling inhibition was effected by short xylo-oligosaccharides released from the XlnC reaction, rather than a direct interaction with BX. Interestingly, IC_{50} for BX competition of XlnC labeling in secretomes was approximately an order of magnitude less potent than that observed for ASPACDRAFT_127619 under the same conditions (4.42 mg/mL vs 596 μg/mL; Figure 5e versus Supplemental Figure 10f), indicating poorer xylan affinity for XlnC compared to ASPACDRAFT_127619. This was confirmed by competitive ABP labeling of *A. niger* secretomes spiked with ASPACDRAFT_127619—at high BX concentrations, labeling of ASPACDRAFT_127619 in spiked secretomes was substantially more inhibited than native XlnC (Supplemental Figure 10g), consistent with stronger affinity between ASPACDRAFT_127619 and BX, compared to *A. niger* XlnC and BX.

Evaluation of Enzyme Stability. Industrial breakdown of plant biomass often necessitates the use of harsh chemical conditions and/or high temperatures[43] which can denature many otherwise useful enzymes. ABP profiling is well placed to assess the stabilities of multiple enzymes within their native context and thus provides a useful tool for the discovery of enzymes that are stable under industrially relevant conditions.

We assessed *A. niger* secretome enzyme stability by preincubating day 6 xylan-induced secretomes at defined temperatures, before cooling and labeling with **8** (Figure 6). Both XlnD and XlnC bands remained stable after 60 min preincubations at 40 and 50 °C. At 60 °C, the lower XlnC band gradually lost labeling intensity with increasing preincubation times, indicating gradual thermal denaturation of this enzyme. By 70 °C, all XlnC activity was abolished within 5 min. Consistent with previous studies,[44] XlnD was found to be to be highly thermostable, retaining full activity-based labeling even after 2 h preincubation at 70 °C. As expected from ABP labeling, secretomes preincubated 70 °C for 30 min retained the ability to hydrolyze 4MU-β-D-xyloside but not 4MU-β-D-xylobioside (Supplemental Figure 5). Heating to 80 °C was required to abolish XlnD labeling, although a small amount of labeling was still observed even after 30 min at this temperature. Similar analyses of *Anid*XlnD and ASPACDRAFT_127619 showed the recombinant enzymes to be substantially less thermally resistant than their *A. niger* counterparts, with loss of labeling intensity observed for both enzymes even after 40 °C preincubations (Supplemental Figure 11). Taken together, these experiments demonstrate the suitability of ABPs for assessing the stability of enzymes in isolation as well as within complex secretome mixtures.

■ DISCUSSION

Plant polysaccharides are the most abundant biopolymers on earth. Recent analyses indicate an annual production of plant biomass on the >10 gigatonne scale suggesting a global

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

terrestrial pool of greater than 2000 gigatonnes.[45,46] Plant biomass therefore has considerable potential for exploitation as a renewable energy source.[47−50] Effecting the breakdown of plant biomass into useful small molecules will form an important part of the sustainable biofuel and chemical industries of the future. Research toward this goal necessitates the discovery of enzymes that can catabolize cellulose and hemicelluloses in a controlled and well-defined manner.

Traditional approaches toward enzyme detection, such as the use of fluorogenic or colorimetric activity assays to detect glycosidases, are limited in sensitivity and lack the ability to resolve multiple overlapping enzyme activities in complex samples. Additionally, there is limited scope for leveraging information from activity assays toward direct identification of the enzymes responsible for such activities. Conversely, omics-based approaches, which aim to detect changes in gene expression patterns in response to changes in the external environment, offer limited insight into the functions of numerous identified up- or down-regulated genes and secreted proteins. Recent transcriptomic and proteomic analysis of the response of *Pycnoporus coccineus* to different biomass substrates revealed four gene clusters which were upregulated in response to lignocellulosic biomass.[51] While many of these genes and secreted proteins could be confidently annotated as carbohydrate-active enzymes, their potential utility, as dictated by pH−activity profile, thermal stability, and substrate specificity, remains unknown.

The use of ABPs for the simultaneous characterization of multiple enzyme activities fills this gap. Here, we have presented a suite of cyclophellitol-derived ABPs, which provide a rapid and convenient method for screening complex samples for xylan-specific enzymes in their native contexts. Using biotin-tagged ABPs, we were able to identify a specific subset of three enzymes with known β-xylanase, β-xylosidase, or β-glucosidase activities out of all the proteins present in a BX-induced *A. niger* secretome. Using a single fluorescent ABP, we were able to track changes in the enzymatic composition of BX-induced *A. niger* secretomes over several days, visualizing the dynamic accumulation and disappearance of multiple enzyme activities over time. Such information on the response of saprophytes grown under different conditions provides valuable information on the strategies employed by these organisms to catabolize complex substrates.

One of our key motivations for the development of β-xylanase and β-xylosidase ABPs was the desire to create tools for identification of enzymes that may be useful in industrial biomass breakdown. Industrial processes typically require enzymes with well-defined activities and high stability, properties which are well suited for investigation using ABP-based approaches. Enzyme specificity can be readily characterized using competitive ABP labeling strategies, which measure the susceptibility of ABP labeled enzymes to inhibition by substrates or ligands of interest. Enzyme stability can be assessed by preincubating secretomes under the (potentially) denaturing conditions of interest and observing which enzymes remain responsive to subsequent ABP labeling. We have shown here that XlnC and XlnD in *A. niger* secretomes are substantially more resistant to thermal denaturation than their recombinant homologues ASPAC-DRAFT_127619 and *Anid*XlnD, which may reflect both intrinsic differences in enzyme stabilities and/or the presence of stabilizing factors within the *A. niger* secretome, which protect XlnC and XlnD from the effects of higher temperatures.

Although *A. niger* itself is a well-characterized model organism, the ABP workflow is generally applicable and adaptable to the discovery of unknown enzymes from novel organisms. Work to discover such enzymes will also necessitate the development of further ABPs with distinct specificities, which can target the wide range of glycosidase activities involved in breakdown of plant polysaccharides. We have demonstrated here that simple elongation of monomeric xylose ABPs to mimic a xylobiose moiety is a viable strategy for expanding the scope of cyclophellitol-derived ABP reactivity from β-xylosidases to β-xylanases. Further ABP elaboration, such as the development of structures that mimic branched oligosaccharide substrates, may allow for the precise interrogation of yet more glycosidases with distinct substrate specificities.

One limitation of our current generation of ABPs is their lack of reactivity with retaining glycosidases that utilize conformational itineraries not involving the $^4H_3$ conformation, the ground state conformation typically favored by cyclophellitol-derived ABPs.[52] Inverting glycosidases, which do not utilize a covalent enzyme−substrate intermediate during their catalytic cycles, are also invisible to cyclophellitol and its derivatives. Development of novel warhead chemistries that effectively label the active site of these diverse glycosidases is an obvious area where future ABP design efforts should be focused. Ultimately, we envision that the creation of a comprehensive library of suitable ABPs will allow for dynamic analysis of all glycosidases produced by a saprophyte in response to growth on defined substrates. Such proteomic level information will complement the increasing plethora of genomic level information that has become available since the advent of high-throughput DNA sequencing technologies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00221.

Supplemental Figures 1−11, Supplemental Table 1, synthetic protocols (PDF)

Proteomics data sets (XLSX)

Movies of GH3, GH10 and GH11 enzyme−ABP complexes (MPG1, MPG2, and MPG3)

■ AUTHOR INFORMATION

Corresponding Authors

*(L.W.) E-mail: liang.wu@york.ac.uk.
*(G.J.D.) E-mail: gideon.davies@york.ac.uk.
*(H.S.O.) E-mail: h.s.overkleeft@chem.leidenuniv.nl.

ORCID

Sybrin P. Schröder: 0000-0002-7042-0911
Carme Rovira: 0000-0003-1477-5010
Gilles P. van Wezel: 0000-0003-0341-1561
Jeroen D. C. Codée: 0000-0003-3531-2138
Keith S. Wilson: 0000-0002-3581-2194
Gideon J. Davies: 0000-0002-7343-776X

Author Contributions

○S.P.S. and C.d.B. contributed equally to this work.

DOI: 10.1021/acscentsci.9b00221
ACS Cent. Sci. 2019, 5, 1067−1078

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Cite This: J. Am. Chem. Soc. 2019, 141, 4214−4218
Communication
pubs.acs.org/JACS

Functionalized Cyclophellitols Are Selective Glucocerebrosidase Inhibitors and Induce a Bona Fide Neuropathic Gaucher Model in Zebrafish

Marta Artola,[†,∥] Chi-Lin Kuo,[‡,∥] Lindsey T. Lelieveld,[‡] Rhianna J. Rowland,[§] Gijsbert A. van der Marel,[†] Jeroen D. C. Codée,[†] Rolf G. Boot,[‡] Gideon J. Davies,[§] Johannes M. F. G. Aerts,*[,‡] and Herman S. Overkleeft*[,†]

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

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

Supporting Information

ABSTRACT: Gaucher disease is caused by inherited deficiency in glucocerebrosidase (GBA, a retaining β-glucosidase), and deficiency in GBA constitutes the largest known genetic risk factor for Parkinson's disease. In the past, animal models of Gaucher disease have been generated by treatment with the mechanism-based GBA inhibitors, conduritol B epoxide (CBE), and cyclophellitol. Both compounds, however, also target other retaining glycosidases, rendering generation and interpretation of such chemical knockout models complicated. Here we demonstrate that cyclophellitol derivatives carrying a bulky hydrophobic substituent at C8 are potent and selective GBA inhibitors and that an unambiguous Gaucher animal model can be readily generated by treatment of zebrafish with these.

Glucocerebrosidase (acid glucosylceramidase, GBA, EC 3.2.1.45) is a lysosomal retaining β-glucosidase that belongs to the glycoside hydrolase (GH) 30 family (www.cazy.org)[1] and degrades the glycosphingolipid, glucosylceramide, through a two-step Koshland double displacement mechanism (Figure 1a). Inherited deficiency in GBA causes the most common autosomal recessive lysosomal storage disorder, Gaucher disease.[2] Individuals carrying heterozygous mutations in the gene coding for GBA do not develop Gaucher disease but have a remarkable increased risk for developing Parkinson's disease (PD) and Lewy-body dementia.[3−5] Appropriate animal models linking impaired GBA functioning to Gaucher disease and Parkinson's disease are imperative both for understanding the pathophysiology of these diseases and for the development of effective treatments for these. Because complete genetic abrogation of GBA hampers animal viability due to skin permeability problems,[6] research models have been generated in the past in a chemical knockdown strategy by making use of the mechanism-based, covalent, and irreversible retaining β-glucosidase inhibitor, conduritol B epoxide (CBE, **1**, Figure 1b), or its close structural analogueue, cyclophellitol (**2**, Figure 1b).[7,8] One complication in the use of these compounds is their relative lack of selectivity.[9] We found that cyclophellitol **2** is unsuited for creating a reliable Gaucher animal model because it targets GBA and GBA2 with about equal efficiency.[9] On the other hand, CBE **1** exhibits some GBA selectivity but it also inhibits lysosomal α-glucosidase (GAA),[10−13] non-lysosomal glucosylceramidase (GBA2),[14,15] and lysosomal β-glucuronidase (GUSB).[16] Effective mouse models can be generated with CBE **1**, but the therapeutic window is rather narrow and varies in cellular and animal models.

Recent research from our group has revealed that functionalized cyclophellitol derivatives carrying a BODIPY substituent at C8 (cyclophellitol numbering, the primary carbon corresponding to C6 in glucose) are very potent and very selective activity-based probes (ABPs) for monitoring GBA activity in vitro, in situ, and in vivo.[17,18] The presence of a bulky and hydrophobic substituent at this position at once proved beneficial for GBA inactivation (ABPs **3** and **4**, Figure 1c,d) proved to inhibit GBA in the nanomolar range, whereas cyclophellitol **2** is a high nanomolar to micromolar GBA inactivator) and detrimental to inhibition of other retaining β-glucosidases. Following these studies, Vocadlo and co-workers designed a set of fluorogenic substrates featuring a fluorophore at C6 of a β-glucoside, the aglycon of which carried a fluorescence quencher, compounds that proved to be very selective GBA substrates in situ.[19] These results altogether evoked the question whether cyclophellitols bearing a simple, hydrophobic moiety at C8, such as compounds **6** and **7** (Figure 1d), would be suitable compounds for generating chemical knockdown Gaucher animal models. We show here the validity of this reasoning in the generation of a GBA-deficient *Dario rerio* zebrafish model, as revealed by the accumulation of elevated levels of the Gaucher harbinger lysolipid, glucosylsphingosine, using cyclophellitol derivatives **6** and **7**.

At the onset of our studies, we sought for structural support for the design of compounds **6** and **7**. We have in the recent past synthesized Cy5-functionalized cyclophellitol **5** (unpublished) and obtained a crystal structure of human recombinant GBA soaked with this ABP (reported here). As expected (Figure 2a), the active site nucleophile (in both molecules of

Received: January 3, 2019
Published: February 27, 2019

Published: February 27, 2019

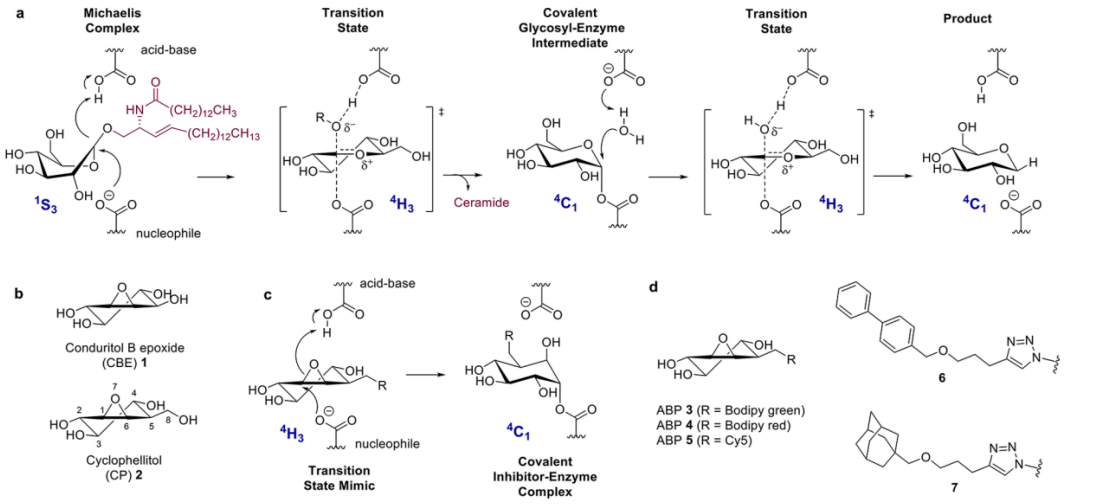


Figure 1. (a) Glucocerebrosidase (GBA) hydrolyses glucosylceramide in a two-step double displacement mechanism to yield glucose and ceramide. (b) Chemical structure of CBE **1** and cyclophellitol **2**. (c) Mechanism-based inactivation of GBA by glucopyranoside-configured cyclitol epoxides (shown for cyclophellitol). (d) Structures of C8-extended cyclophellitol derivatives used in the here-presented studies: GBA activity-based probes ABPs **3**–**5** and selective inhibitors **6** and **7** (see the full chemical structures of ABPs **3**–**5** and **8**–**14** in the Supporting Information (SI)).

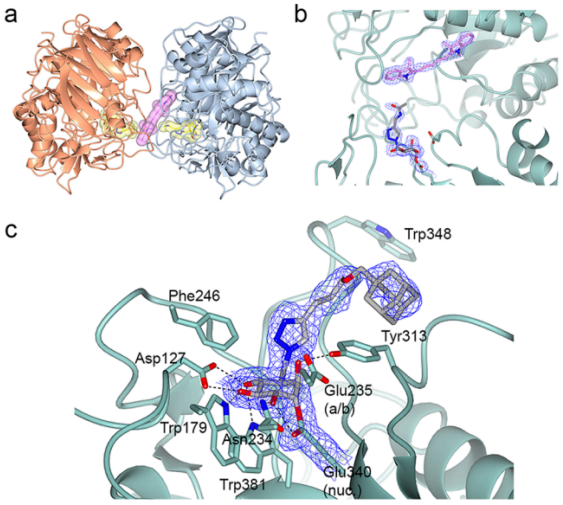


Figure 2. Structure of GBA reacted with ABP **5** and adamantyl-cyclophellitol **7**. (a) GBA dimer, with the cyclophellitol and linker moiety of **5** shaded in yellow and a single observed Cy5 in pink. (b) Zoomed view of a GBA monomer reacted with ABP **5**. (c) Structure of GBA with adamantyl-cyclophellitol **7**. The linker-adamantyl moiety of **7** is observed in slightly different positions in the two molecules of the asymmetric unit (PDB 6Q6L, SI, Figure S2), reflecting its binding through predominantly hydrophobic interactions.

the asymmetric unit) had reacted with the epoxide to yield the covalently bound cyclitol in $^4C_1$ conformation, with the Cy5 moiety, via its flexible linker, clearly bound in one molecule of the asymmetric unit (the differences may reflect crystal packing constraints in a soaking experiment) accommodated by a hydrophobic pocket in GBA. Previous studies by us on the bacterial glycoside hydrolase, *Thermoanaerobacterium xylanolyticum* *Tx*GH116 β-glucosidase, a close homologue of human GBA2 with a conserved active site, instead showed an "inwards" position of O6 (SI, Figure 2a) and a narrower and less hydrophobic pocket (SI, Figure 2b), which may partially mitigate against the binding of O6-functionaised probes, thus allowing sufficient discrimination for GBA over GBA2.[20,21]

Biphenyl-cyclophellitol **6** and adamantyl-cyclophellitol **7** were synthesized following adaptations of literature cyclophellitol syntheses (see SI for synthesis details of Cy5-cyclophellitol ABP **5** and compounds **6** and **7**) to generate superior selective GBA inhibitors for the generation of a Gaucher model zebrafish.[22,23]

Although soaking of GBA crystals with **6** did not yield suitable structures for structural analysis (SI, Figure S2), soaking with **7** did (Figure 2c) and again revealed binding of the hydrophobic moiety (here, the adamantane) to the same hydrophobic cavity and pocket occupied by the O6 linker on Cy5 ABP **5**. Several hydrophobic residues, including Tyr313, Phe246, and Trp348 provide the wide cavity that is able to accommodate different hydrophobic O6 substituents which is absent in other human β-glucosidases and which provides the structural basis for the inhibitory (and substrate) preferences of GBA.

We evaluated the in vitro activity and selectivity of compounds **6** and **7** toward GBA and the two major off-target glycosidases of CBE **1**, GBA2 and GAA, by preincubating the inhibitors with recombinant human GBA (rGBA, Cerezyme), human GBA2 (from lysates of GBA2 overexpressed cells), and recombinant human GAA (rGAA, Myozyme) for 3 h, followed by enzymatic activity measurement. We showed that both compound **6** and **7** were nanomolar inhibitors of rGBA (apparent IC_{50} values = 1.0 nM), which were 4000-fold more potent than CBE **1** (apparent IC_{50} values = 4.28 μM) (Figure 3a, SI, Figure S3) with improved lipophilic ligand efficiencies (LipE) (SI, Table S2). Both compounds **6** and **7** were rather inactive toward GBA2 and GAA (apparent IC_{50} values >100 μM), similar to ABP **3** and **5** (Figure 3a, SI, Figure S4). When comparing their selectivity toward GBA, both compounds **6** and **7** exhibited IC_{50} ratio (GBA2/GBA and GAA/GBA) of >100000, thus making them 4000 times and 200 times more

a	CBE **1** (nM)	CP **2** (nM)	ABP **3** (nM)	ABP **5** (nM)	**6** (nM)	**7** (nM)
In vitro						
rGBA	4,280 ± 500 [a]	29.6 ± 2.40 [a]	1.20 ± 0.30	3.20 ± 0.17	1.06 ± 0.19	0.96 ± 0.17
GBA2 (HEK293T lysates)	101,000 ± 20,100	29.7 ± 3.13	> 10^5	412,000 ± 10,100	> 10^5	> 10^5
rGAA	1,900,000 ± 192,000	> 10^5	> 10^5	> 10^5	> 10^5	> 10^5
Ratio *(in vitro)*						
GBA2/ GBA	24	1	> 10^5	> 10^4	> 10^5	> 10^5
GAA/ GBA	444	> 10^3	> 10^6	> 10^6	> 10^6	> 10^6

b	CBE **1** (nM)	CP **2** (nM)	ABP **3** (nM)	ABP **5** (nM)	**6** (nM)	**7** (nM)
In vivo (Danio rerio larvae)						
GBA	4.41 x10^4 [a]	83 [a]	31.6 ± 8.88	284 ± 31.5	5.85 ± 2.44	3.94 ± 1.21
GBA2	8.90 x10^5 [a]	59 [a]	> 10^4	> 10^4	> 10^4	> 10^4
GAA	9.55 x 10^6 [a]	> 10^5 [a]	> 10^4	> 10^4	> 10^4	> 10^4
Ratio *(in vivo)*						
GBA2/ GBA	22	0.714	> 316	> 35	> 1710	> 2540
GAA/ GBA	233	> 120	> 316	> 35	> 1710	> 2540

[a] Values from ref. 9

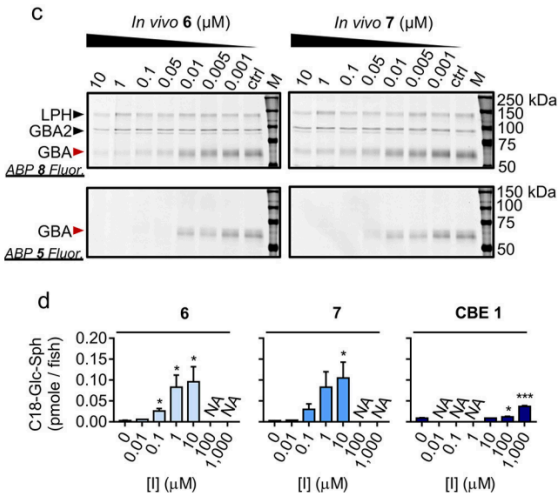


Figure 3. (a) Apparent IC_{50} values for in vitro inhibition of GBA, GBA2, and GAA in recombinant enzymes (rGBA and rGAA) or overexpressed cell lysates (GBA2) by compounds **1**, **2**, **3**, **5**, **6**, and **7**. Error ranges depict standard deviations from biological triplicates. (b) Apparent IC_{50} values for in vivo inhibition in 5-day treated zebrafish embryo with compounds **1**, **2**, **3**, **5**, **6**, and **7**. Error ranges depict standard deviations from n = 12–24 individuals. (c) Competitive ABPP in lysates of zebrafish treated in vivo with compounds **6** and **7** using broad-spectrum retaining β-glucosidase ABP **8** and selective GBA ABP **5** as readout. (d) Glucosylsphingosine levels produced in zebrafish embryos treated for 5 days with inhibitors **6**, **7** or CBE **1**.[9] Error ranges depict standard deviations from n = 3 individuals. N/A, not analyzed; *, $p < 0.5$; ***, $p < 0.001$.

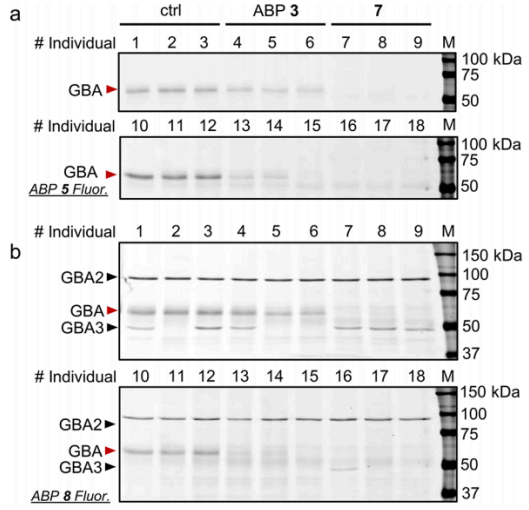


Figure 4. In vivo targets of ABP **3** and **7** in brains of adult zebrafish. Competitive ABPP in adult zebrafish homogenates with selective GBA ABP **5** (a) or broad-spectrum retaining β-glucosidase ABP **8** (b) as read-out.

selective than CBE **1** (IC_{50} ratio = 23.6 for GBA2/GBA and 444 for GAA/GBA) (Figure 3a). To evaluate the in vivo activity of compound **6** and **7**, compounds were added to the egg–water containing zebrafish (*Danio rerio*) embryos and incubated for 5 days at 28 °C before subsequent homogenization and enzyme selectivity analysis by appropriate ABP labeling.[9,24] Quantification of ABP-labeled bands revealed that compounds **6** and **7** had in vivo apparent IC_{50} values toward GBA of 4–6 nM, and that they were 5–70-fold more potent than ABP **3** or **5** and 7500-fold more potent than CBE **1** (Figure 3b, SI, Figure S5) in the zebrafish larvae. More importantly, an improved selective inactivation of GBA was achieved with both compounds **6** and **7**. At a concentration of 0.1–10 μM of compound **6** or **7**, ABP labeling of GBA with broad-spectrum retaining β-glucosidase ABP **8** (SI, Figure S1) and GBA specific ABP **5** was abrogated (Figure 3c), while other enzymes such as GBA2, LPH (Figure 3c), or GAA, ER α-glucosidase GANAB, and lysosomal β-glucuronidase GUSB (SI, Figure S6a, S6b) were not affected.

At 0.1–10 μM of inhibitor **6** or **7**, we also observed 10–30-fold elevation in the level of glucosylsphingosine (GlcSph), which is known to be formed by acid ceramidase-mediated conversion of accumulating GlcCer in lysosomes.[25,26] Therefore, this observation also strongly points to in vivo inactivation of lysosomal GBA. For comparison, reaching similar GlcSph levels in the zebrafish with CBE required 1000–10000-fold higher concentration in contrast with compounds **6** or **7** (Figure 3d), concentrations at which GBA2 and GAA may also be targeted (Figure 3b). Finally, we investigated the brain permeability of these new inhibitors, a crucial feature for their future application in the study of neuropathic Gaucher disease and Parkinson's disease. Adult zebrafish of 3 months' age were treated with DMSO, ABP **3**, or compound **7** (1.6 nmol/fish, approximately 4 μmol/kg) administered via food intake, and after 16 h brains and other organs were isolated, homogenized, and analyzed by ABP labeling using ABP **5** (GBA), ABP **8** (GBA2 + GBA), ABP **11** (GAA at pH 4.0 and ER α-glucosidase GANAB at pH 7.0),

and ABP **13** (lysosomal β-glucuronidase GUSB) (SI, Figure S1). Labeling of brain homogenate of adult zebrafish with ABP **5** resulted in considerable GBA labeling in control and ABP **3**-treated fish, but no labeling in brain homogenates from fish treated with compound **7** (Figure 4). Labeling by the broad-spectrum β-glucosidase ABP **8** showed that GBA2 was not a target of compound **7**, nor was the lower running band (48 kDa), which we hypothesize to be the cytosolic β-glucosidase, GBA3. We noted that the expression level of this protein is likely variable among individual fish, as 4 out of 6 fishes in the control group lacked this band. In the visceral organs (both liver and spleen), both ABP **3** and compound **7** selectively abrogated GBA while not affecting the labeling on other tested glycosidases (SI, Figure S7).

To summarize, crystallographic studies aided the rational design of novel cyclophellitol analogues **6** and **7**, which turned out to be very potent and selective GBA inhibitors, also in zebrafish embryos and adult zebrafish (GBA2/GBA inhibition ratio >1000). Compound **7**, which also completely block GBA activity in the brain, is in our opinion superior to CBE **1** and CP **2** for generating GBA deficiency on demand in zebrafish, thus to create zebrafish models for neuropathic Gaucher, to assist research in the context of neuropathic GD and PD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b00056.

Experimental data and procedures, crystallographic data, and synthesis (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*h.s.overkleeft@lic.leidenuniv.nl
*j.m.f.g.aerts@lic.leidenuniv.nl

ORCID

Marta Artola: 0000-0002-3051-3902
Chi-Lin Kuo: 0000-0003-3748-5008
Jeroen D. C. Codée: 0000-0003-3531-2138
Gideon J. Davies: 0000-0002-7343-776X
Herman S. Overkleeft: 0000-0001-6976-7005

Author Contributions

‖M.A. and C.-L.K. contributed equally to this paper.

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. for financial support and postdoctoral contract to M.A.). R.J.R. is supported by the BBSRC (BB/M011151/1). We thank the Diamond Light Source for access to beamline i02 and i04 (proposal no. mx-13587). G.J.D. is supported by the Royal Society though the Ken Murray Research Professorship.

■ REFERENCES

(1) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42* (D1), D490−D495.

(2) Brady, R. O.; Kanfer, J. N.; Bradley, R. M.; Shapiro, D. Demonstration of a Deficiency of Glucocerebroside-Cleaving Enzyme in Gaucher's Disease. *J. Clin. Invest.* **1966**, *45*, 1112−1115.

(3) Schapira, A. H. V. Glucocerebrosidase and Parkinson Disease: Recent Advances. *Mol. Cell. Neurosci.* **2015**, *66*, 37−42.

(4) Tsuang, D.; Leverenz, J. B.; Lopez, O. L.; Hamilton, R. L.; Bennett, D. A.; Schneider, J. A.; Buchman, A. S.; Larson, E. B.; Crane, P. K.; Kaye, J. A.; Kramer, P.; Woltjer, R.; Kukull, W.; Nelson, P. T.; Jicha, G. A.; Neltner, J. H.; Galasko, D.; Masliah, E.; Trojanowski, J. Q.; Schellenberg, G. D.; Yearout, D.; Huston, H.; Fritts-Penniman, A.; Mata, I. F.; Wan, J. Y.; Edwards, K. L.; Montine, T. J.; Zabetian, C. P. GBA Mutations Increase Risk for Lewy Body Disease with and without Alzheimer Disease Pathology. *Neurology* **2012**, *79* (19), 1944−1950.

(5) 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.; Dürr, 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.; Ziegler, S. G. Multicenter Analysis of Glucocerebrosidase Mutations in Parkinson's Disease. *N. Engl. J. Med.* **2009**, *361* (17), 1651−1661.

(6) Holleran, W. M.; Ginns, E. I.; Menon, G. K.; Grundmann, J. U.; Fartasch, M.; McKinney, C. E.; Elias, P. M.; Sidransky, E. Consequences of β-Glucocerebrosidase Deficiency in Epidermis. Ultrastructure and Permeability Barrier Alterations in Gaucher Disease. *J. Clin. Invest.* **1994**, *93* (4), 1756−1764.

(7) Farfel-Becker, T.; Vitner, E. B.; Futerman, A. H. Animal Models for Gaucher Disease Research. *Dis. Models & Mech.* **2011**, *4*, 746−752.

(8) Vardi, A.; Zigdon, H.; Meshcheriakova, A.; Klein, A. D.; Yaacobi, C.; Eilam, R.; Kenwood, B. M.; Rahim, A. A.; Massaro, G.; Merrill, A. H.; Vitner, E. B.; Futerman, A. H. Delineating Pathological Pathways in a Chemically Induced Mouse Model of Gaucher Disease. *J. Pathol.* **2016**, *239*, 496−509.

(9) Kuo, C.-L.; Kallemeijn, W. W.; Lelieveld, L.; Mirzaian, T. M.; Zoutendijk, I.; Vardi, A.; Futerman, A. H.; Meijer, A. H.; Spaink, H. P.; Overkleeft, H. S.; Aerts, J. M. F. G.; Artola, M. *In Vivo* Inactivation of Glycosidases by Conduritol B Epoxide and Cyclophellitol as Revealed by Activity-Based Protein Profiling. *FEBS J.* **2019**, *286* (3), 584−600.

(10) Quaroni, A.; Gershon, E.; Semenza, G. Affinity Labeling of the Active Sites in the Sucrase-Isomaltase Complex from Small Intestine. *J. Biol. Chem.* **1974**, *249*, 6424−6433.

(11) Shou-jun, Y.; Su-guo, G.; Yu-cheng, Z.; Shu-zheng, Z. Inactivation of α-Glucosidase by the Active-Site-Directed Inhibitor, Conduritol B Epoxide. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1985**, *828* (3), 236−240.

(12) Hermans, M. M. P.; Krooss, M. A.; van Beeurnens, J.; Oostras, B. A.; Reusersll, A. J. J. Human Lysosomal a-Glucosidase Characterization of the Catalytic Site. *J. Biol. Chem.* **1991**, *266* (21), 13507−13512.

(13) Braun, H.; Legler, G.; Deshusses, J.; Semenza, G. Stereospecific Ring Opening of Conduritol-B-Epoxide by an Active Site Aspartate Residue of Sucrase-Isomaltase. *BBA - Enzymol.* **1977**, *483* (1), 135−140.

(14) van Weely, S.; Brandsma, M.; Strijland, A.; Tager, J. M.; Aerts, J. M. F. G. Demonstration of the Existence of a Second, Non-Lysosomal Glucocerebrosidase That Is Not Deficient in Gaucher

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2019, **10**, 9233

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 5th July 2019
Accepted 19th August 2019

DOI: 10.1039/c9sc03342d

rsc.li/chemical-science

α-D-Gal-cyclophellitol cyclosulfamidate is a Michaelis complex analog that stabilizes therapeutic lysosomal α-galactosidase A in Fabry disease†

Marta Artola,[ab] Christinne Hedberg,[a] Rhianna J. Rowland,[c] Lluís Raich,[d] Kassiani Kytidou,[b] Liang Wu,[c] Amanda Schaaf,[a] Maria Joao Ferraz,[b] Gijsbert A. van der Marel,[a] Jeroen D. C. Codée,[a] Carme Rovira,[de] Johannes M. F. G. Aerts,[b] Gideon J. Davies *[c] and Herman S. Overkleeft *[a]

Fabry disease is an inherited lysosomal storage disorder that is characterized by a deficiency in lysosomal α-D-galactosidase activity. One current therapeutic strategy involves enzyme replacement therapy, in which patients are treated with a recombinant enzyme. Co-treatment with enzyme active-site stabilizers is advocated to increase treatment efficacy, a strategy that requires effective and selective enzyme stabilizers. Here, we describe the design and development of an α-D-gal-cyclophellitol cyclosulfamidate as a new class of neutral, conformationally constrained competitive glycosidase inhibitors that act by mimicry of the Michaelis complex conformation. We found that D-galactose-configured α-cyclosulfamidate **4** effectively stabilizes recombinant human α-D-galactosidase (agalsidase beta, Fabrazyme®) both *in vitro* and *in cellulo*.

Introduction

Deficiency of α-galactosidase A (α-gal A, EC 3.2.1.22, a retaining glycosidase of the GH27 glycoside hydrolase family (http://www.cazy.org)[1]) underlies Fabry disease. This inherited lysosomal storage disorder is characterized by toxic accumulation of glycosphingolipid globotriaosylceramide (Gb3) in lysosomes and its sphingoid base, globotriaosylsphingosine (lyso-Gb3) in plasma and tissues.[2,3] Several mutations in the GLA gene encoding α-gal A can result in diminished protein levels and/or enzyme activity, leading to altered metabolite levels and a range of Fabry disease phenotypes. The accumulation of glycosphingolipid metabolites is thought to cause progressive renal and cardiac insufficiency and CNS pathology in Fabry patients.[4] Enzyme replacement therapy (ERT) for Fabry disease involves intravenous treatment with recombinant human α-gal A (agalsidase beta, Fabrazyme® or agalsidase alpha, Replagal®), but the clinical efficacy of this therapy is limited.[5–7] 1-Deoxygalactonojirimycin (Gal-DNJ **8**, Migalastat®, Fig. 1B) has recently been approved as a pharmacological chaperone (PC) for the treatment of Fabry disease in patients with amenable mutations.[8] Gal-DNJ **8** binds mutant forms of α-gal A, which are catalytically competent but otherwise targeted for degradation due to misfolding. Gal-DNJ **8** stabilizes the protein fold, allowing the mutant α-gal A to be trafficked to lysosomes. However, PC therapy for Fabry disease is limited to specific mutations and its efficacy is hotly debated.[9–13] For this reason, an attractive alternative therapeutic intervention strategy, proposed recently, comprises jointly administering a recombinant enzyme and a pharmacological chaperone.[14–16] This strategy aims to stabilize the recombinant enzyme in circulation such that larger proportions may reach disease affected tissues, permitting the use of extended intervals between injections and lower enzyme dosages, which should diminish side effects, improve the patient's lifestyle and reduce treatment costs.[17,18] For this strategy to become clinical practice, allosteric enzyme stabilizers or orthosteric competitive α-gal A inhibitors that prevent enzyme unfolding and are displaced by the accumulated metabolites in the lysosome recovering the glycosidase activity, with good selectivity and pharmacokinetic/pharmacodynamic properties, are required.[16,18,19] We argue that the discovery of

[a]*Department of Bio-organic Synthesis, Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands. E-mail: h.s.overkleeft@lic.leidenuniv.nl*

[b]*Department of Medical Biochemistry, Leiden Institute of Chemistry, Einsteinweg 55, 2333 CC Leiden, The Netherlands*

[c]*Department of Chemistry, York Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, UK. E-mail: gideon.davies@york.ac.uk*

[d]*Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica), Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franques 1, 08028 Barcelona, Spain*

[e]*Fundació Catalana de Recerca i Estudis Avancats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain*

† Electronic supplementary information (ESI) available: Supplementary Fig. S1 to S7, Tables S1 and S2, materials and methods (biological and biochemical methods, DFT calculations, crystallography and chemical synthesis), and NMR spectra. See DOI: 10.1039/c9sc03342d

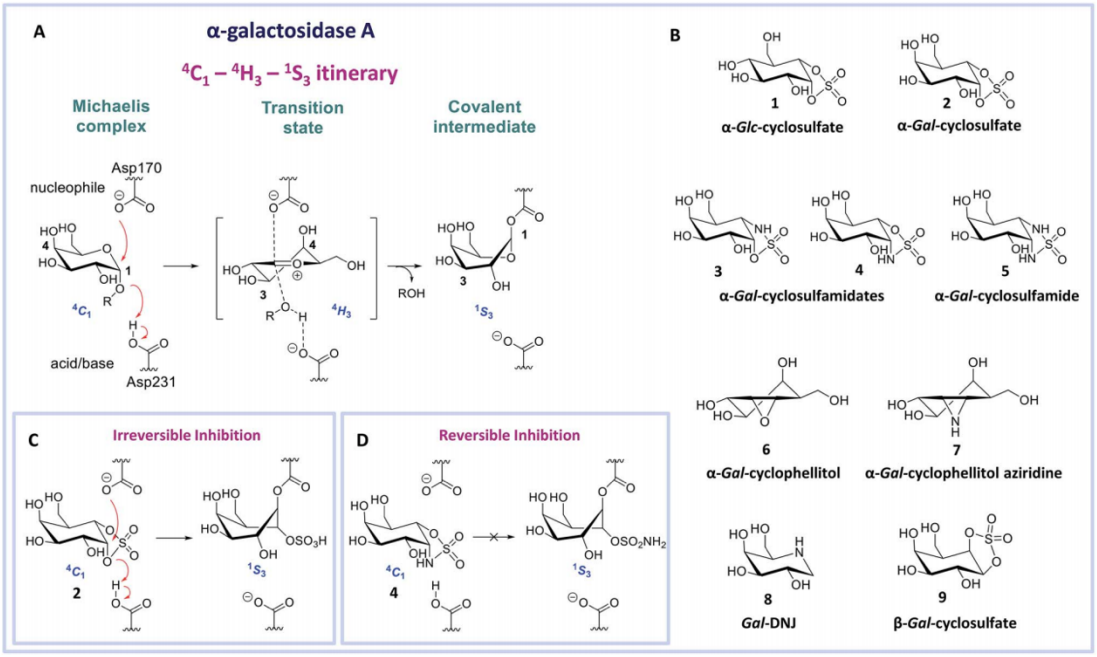


Fig. 1 Reaction coordinates of α-galactosidases and inhibitors. (A) Reaction itinerary of retaining α-galactosidase, showing conformations of the Michaelis complex, transition state, and covalent intermediate. (B) Chemical structures of α-glucose configured cyclosulfate **1**, α-galactose configured cyclosulfate **2**, cyclosulfamidates **3** and **4**, cyclosulfamide **5**, cyclophellitol **6**, cyclophellitol aziridine **7**, 1-deoxygalactonojirimycin **8** and β-galactose configured cyclosulfate **9**. Galactose configured cyclosulfate **2** and cyclosulfamidate **4** inhibit α-galactosidases irreversibly (C) or reversibly (D) by mimicking the $^4C_1$ "Michaelis-like" conformation.

such commodities would be greatly facilitated by the design of new inhibitor templates.

Human α-gal A hydrolyzes its substrate following a Koshland double displacement mechanism, resulting in net retention of stereochemistry at the anomeric center of the produced galactopyranose.[20,21] The reaction coordinates by which α-gal A processes its substrate to form the intermediate covalent adduct are $^4C_1 \rightarrow {}^4H_3^{\ddagger} \rightarrow {}^1S_3$ with respect to the conformation of the galactopyranose moiety in the Michaelis complex → TS → covalent intermediate complex (Fig. 1A).[22–24] This same reaction itinerary is also employed by GH31 retaining α-glucosidases, with the difference that a glucopyranose, rather than a galactopyranose, is captured in the enzyme active site.[25] We have recently shown that α-glc-cyclosulfate **1** (Fig. 1B) potently, selectively and irreversibly inhibits retaining α-glucosidases. Compound **1** in free solution predominantly resides in a $^4C_1$ chair conformation, thus mimicking the initial Michaelis complex conformation utilized by α-glucosidases[25] allowing facile interception by the catalytic nucleophile.

We reasoned that α-gal-cyclosulfate **2** would covalently and irreversibly inhibit α-gal A with equal efficiency and selectivity following the same mode of action (Fig. 1B and C). Building on this concept, we further hypothesized that substitution of one or both of the cyclosulfate ring oxygens for nitrogen, as in compounds **3–5**, would lead to competitive α-gal A inhibitors because of the decreased leaving group capacity of cyclosulfamidates/cyclosulfamides, when compared to cyclosulfates (Fig. 1D). Such compounds would then offer competitive enzyme inhibitors to be tested as stabilizers of recombinant α-galactosidase A for Fabry treatment. Here, we show the validity of this reasoning by revealing α-gal-cyclosulfamidate **4** as a first-in-class, effective and selective, competitive α-gal A inhibitor. Structural and computational analysis of the conformational behavior of compound **4** in solution and in the active site of human α-gal A supports our design and provides a molecular rationale why compound **4** is an effective α-gal A inhibitor. We also show compound **4** to be effective in stabilizing recombinant α-gal A *in vitro* and *in cellulo* and that sphingolipid levels in Fabry fibroblasts are effectively corrected by co-treatment with α-gal A and **4**.

Results

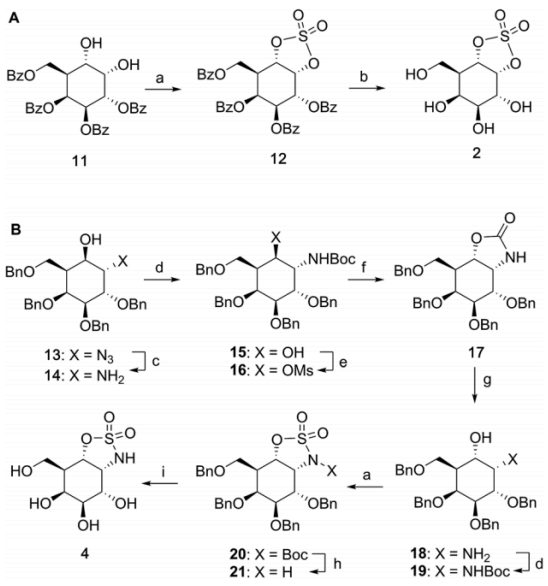
Synthesis of α-D-galactose configured cyclosulfate 2 and cyclosulfamidate 4

Benzoylated diol **11** (see the ESI and Scheme S1† for its synthesis) was treated with thionyl chloride and subsequently oxidized with ruthenium trichloride and sodium periodate to afford protected cyclosulfate **12**. α-Gal-cyclosulfate **2** was obtained after benzoyl removal using methanolic ammonia

(Scheme 1A). The synthesis of *cis*-1-amino-6-hydroxy cyclohexane **18**, a key intermediate in the synthesis of α-gal-cyclosulfamidate **4** proceeded through oxazolidinone **17**, which was obtained from *trans*-azido alcohol **13** (itself obtained from perbenzylated galacto-cyclophellitol, see the ESI†) as depicted in Scheme 1B. Hydrolysis of the carbamate in **17** and *N*-bocylation provided **19**, which was transformed into fully protected cyclosulfamidate **20**. Global deprotection finally yielded the target compound **4**. The synthesis of compounds **3**, **5** and **9** (Fig. 1B) and intermediates follows related strategies, as is described in the ESI.†

α-D-Galactose configured cyclosulfate 2, cyclosulfamidates 3 and 4, and cyclosulfamide 5 are predominantly in the $^4C_1$ conformation

Free energy landscapes (FELs) of inhibitors report the conformational behavior in solution well, and can therefore be used to predict the selectivity for GH active sites.[26] We calculated the conformational FELs of compounds **2–5** using *ab initio* metadynamics (Fig. 2A, S1 and S2†). The FEL of α-gal-cyclophellitol cyclosulfate **2** is strongly biased towards $^4C_1$, with a secondary minimum around $B_{2,5}$. This $B_{2,5}$ conformer is unlikely to be enzyme active-site-reactive as it exhibits an equatorial C1–O bond (Fig. S2†). The $^4C_1$ minimum of the substrate extends towards the TS-like $^4H_3$ conformation, indicating that cyclosulfate **2** in a $^4H_3$ conformation could be transiently populated on-enzyme, favoring the nucleophilic attack and formation of a glycosyl-enzyme adduct. The FELs of **3–5** show that substitution of the cyclic sulfate trap by cyclic sulfamidates (**3** and **4**) or sulfamide (**5**) does not significantly affect the conformational preferences. The local $B_{2,5}$ minimum in **4** is more pronounced, probably due to a hydrogen bond between the 2-OH and one cyclosulfamidate oxygen (Fig. S2†).



Scheme 1 Synthesis of α-D-galactose configured cyclosulfate **2** (A) and cyclosulfamidate **4** (B). Reagents and conditions: (a) (i) $SOCl_2$, Et_3N, imidazole, DCM, 0 °C; (ii) $RuCl_3$, $NaIO_4$, CCl_4, MeCN, 0 °C, 3 h, **12**: 56% and **20**: 59%; (b) NH_3, MeOH, rt, 3 h, 34%; (c) PtO_2, H_2, THF, rt, 4 h, 99%; (d) Boc_2O, Et_3N, DCM, rt, 16 h, **15**: 93% and **19**: 99%; (e) imidazole, MsCl, Et_3N, $CHCl_3$, rt, 16 h; (f) DMF, 120 °C, 2 days, 47% over 2 steps; (g) 1 M NaOH, EtOH, 70 °C, 3 h, to rt, 16 h, 86%; (h) TFA, DCM, rt, 16 h, 71%; (i) $Pd(OH)_2$, H_2, MeOH, rt, 18 h, 57%.

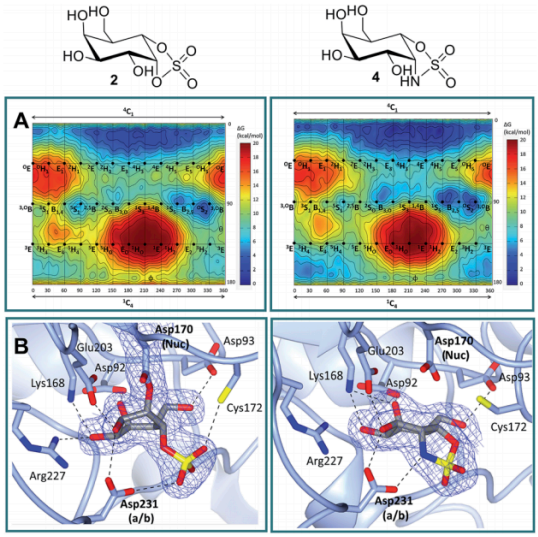


Fig. 2 Conformational free energy landscapes and crystal structures of α-gal-cyclosulfate **2** and α-gal-cyclosulfamidate **4** in α-gal A (agalsidase beta). (A) α-Gal-cyclosulfate **2** and α-gal-cyclosulfamidate **4** adopt $^4C_1$ ground state conformations. The *x* and *y* axes of each graph correspond to the φ and θ Cremer–Pople puckering coordinates (in degrees), respectively. Isolines are 1 kcal mol^{-1}. (B) α-Gal-cyclosulfate **2** (left) reacts with the Asp170 nucleophile and adopts a $^1S_3$ conformation covalent adduct (*i.e.*, "intermediate-like") in complex with agalsidase beta. Unreacted **4** (right) in complex with agalsidase beta adopts a $^4C_1$ "Michaelis-like" conformation in the active site. Electron density for protein side chains and ligands is REFMAC maximum-likelihood/σA-weighted $2F_o - F_c$ contoured to 0.21 electrons per Å^3 for **2** and **4**. Nuc = nucleophile; a/b = acid/base.

α-D-Galactose configured cyclosulfate 2 and isosteres 4 and 5 inhibit α-gal A *in vitro*

α-Gal-cyclosulfate **2**, α-gal-cyclosulfamidates **3** and **4**, and α-gal-cyclosulfamide **5**, as well as the known α-galactosidase inhibitors α-gal-cyclophellitol **6**,[27] α-gal-cyclophellitol aziridine **7**,[27] Gal-DNJ **8**[28] and β-gal-cyclosulfate **9** were evaluated on their inhibitory potential against recombinant human GH27 α-galactosidase A (α-gal A, Fabrazyme®, agalsidase beta) and their selectivity over human β-galactosidases, galactosidase beta 1 (GLB1, GH35) and galactosylceramidase (GALC, GH59). We first determined the apparent IC_{50} values by using fluorogenic 4-methylumbelliferyl (4MU)-α- or -β-D-galactopyranose substrates (Table 1). α-Gal-cyclosulfate **2** effectively inhibits α-gal A on a par with α-gal-cyclophellitol **6** (apparent IC_{50} = 25 *vs.* 13 μM, respectively), although less potently than α-gal-cyclophellitol

Table 1 Apparent IC_{50} values for *in vitro* inhibition of human recombinant α-galactosidase A (agalsidase beta), β-galactosidase GLB1 in human fibroblast lysates and GALC overexpressed in HEK293 cells. Inactivation rates and inhibition constants (k_{inact} and K_I) in human recombinant α-galactosidase A (agalsidase beta); N.D., not determined

Compd.	*In vitro* α-gal A (agalsidase beta) apparent IC_{50} (μM)	*In vitro* β-gal (GLB1) apparent IC_{50} (μM)	*In vitro* β-gal (GALC) apparent IC_{50} (μM)	Kinetic parameters α-gal A (agalsidase beta) k_{inact} (min^{-1}) and K_I (μM) or K_i (μM)	Kinetic parameters α-gal A (agalsidase beta) k_{inact}/K_I (min^{-1} μM^{-1})
α-Gal-cyclosulfate **2**	25 ± 2.5	>1000	>1000	Irreversible $K_I = 237$ $k_{inact} = 0.06$	0.25
α-Gal-cyclosulfamidate **3**	>1000	39 ± 4.6	95 ± 14	N.A.	N.A.
α-Gal-cyclosulfamidate **4**	67 ± 4.7	>1000	>1000	Competitive $K_i = 110$	—
α-Gal-cyclosulfamide **5**	423 ± 58	38 ± 1.7	191 ± 5.5	N.D.[a]	N.D.[a]
α-Gal-cyclophellitol **6**	13 ± 0.95	0.84 ± 0.13	4.2 ± 0.51	Irreversible $K_I = 430$ $k_{inact} = 0.24$	0.55
α-Gal-cyclophellitol aziridine **7**	0.040 ± 0.005	0.93 ± 0.06	1.1 ± 0.30	Irreversible N.D.[b]	16.4
Gal-DNJ **8**	0.079 ± 0.004	42 ± 0.72	433 ± 39	Competitive $K_i = 0.23$	—
β-Gal-cyclosulfate **9**	>1000	>1000	>1000	N.A.	N.A.

[a] Due to weak inhibition. [b] Due to fast inhibition; N.A., not applicable. Reported values are mean ± standard deviation from 3 technical replicates.

aziridine **7** (apparent $IC_{50} = 40$ nM). Cyclosulfamidate **4**, with the sulfamidate nitrogen taking up the position occupied by the anomeric oxygen in the natural substrate, proved to be a rather good inhibitor ($IC_{50} = 67$ μM), whereas isomer **3** is inactive and sulfamide **5** considerably less potent ($IC_{50} = 423$ μM).

We also measured the apparent IC_{50} values for inhibition against two human β-galactosidases: GLB1 (measured in human fibroblast lysates) and GALC (measured in over-expressing HEK293 cell medium). Cyclosulfate **2** and cyclosulfamidate **4** appear to be more selective than cyclophellitol epoxide **6** and aziridine **7**, and we reason that this is due to the $^4C_1$ conformation adopted by **2** and **4**, which corresponds to the Michaelis complex conformation in α-galactosidases, but not in β-galactosidases (compound **2** is inactive up to 1 mM whereas **6** and **7** display low micromolar activity towards GLB1 and GALC). β-Gal-cyclosulfate **9**, which in principle neither mimics the Michaelis complex nor the transition state conformation of β-galactosidases,[29,30] is inactive against β- and α-galactosidases up to 1 mM. Cyclosulfamidate **4** and Gal-DNJ **8** show selectivity over α-glucosidase GAA, but both inhibit human recombinant β-glucosidase (GBA) (Table S1†).

We next explored the reversibility of inhibition by our new cyclic sulfate analogues. Enzymes were pre-incubated for different time periods (30, 60, 120, and 240 min) with inhibitors at concentrations of their corresponding apparent IC_{50} values, after which residual α-gal A activity was determined (Fig. S3†). Whilst cyclosulfate **2** is an irreversible inhibitor (showing a decrease in α-galactosidase activity with longer incubation time), cyclosulfamidate **4** and cyclosulfamide **5** are reversible inhibitors as revealed by a constant residual activity with extended incubation times. This was confirmed by kinetic studies monitoring the absorbance generated by the hydrolysis of the 2,4-dinitrophenyl-α-D-galactopyranoside substrate (2,4-DNP-α-gal) (Table 1). The irreversible inhibitors **2**, **6** and **7** follow pseudo-first order kinetics. Although α-cyclosulfate **2** has a similar k_{inact}/K_I ratio to α-cyclophellitol **6** ($k_{inact}/K_I = 0.25$ *vs.* 0.55 min^{-1} mM^{-1}, respectively), it has a stronger initial binding constant (K_I) and a slower inactivation rate constant (k_{inact}) (**2**: $K_I = 237$ μM and $k_{inact} = 0.06$ min^{-1}; **6**: $K_I = 430$ μM and $k_{inact} = 0.24$ min^{-1}). Only a k_{inact}/K_I ratio could be measured for α-aziridine **7** due to fast inhibition ($k_{inact}/K_I = 16.4$ min^{-1} mM^{-1}). Kinetics with increasing 2,4-DNP-α-gal concentrations showed that cyclosulfamidate **4** reversibly inhibits α-galactosidase with a K_i of 110 μM.

Structural analysis of α-gal-cyclosulfate 2 and α-gal-cyclosulfamidate 4 in complex with agalsidase beta

Firstly, in order to confirm the covalent inhibition by the cyclic sulfate, the X-ray structure of agalsidase beta in complex with **2** (PDB:6IBM) was determined to 1.99 Å, revealing a single

molecule of **2** covalently bound to the enzyme active site (Fig. 2B). The observed electron density unambiguously shows that α-cyclosulfate **2** has reacted by attack of the catalytic nucleophile Asp170 to form a covalent enzyme–inhibitor complex. This covalent complex adopts a $^{1}S_{3}$ conformation, consistent with the conformation of the covalent intermediate in the α-galactosidase reaction itinerary (Fig. 1A). Upon nucleophilic attack to the cyclic sulfate, the sulfate forms hydrogen bonds with Asp231 and Cys172, the latter suffering a shift in position.

Armed with the knowledge that **2** indeed forms a covalent adduct to agalsidase beta, we moved on to ascertain if the cyclosulfamidate **4** would, as envisaged, function as a non-covalent, active-centre-directed, inhibitor of the enzyme-replacement enzyme. In contrast to cyclosulfate **2**, cyclosulfamidate **4** (PDB:6IBK determined to 2.07 Å) was indeed shown to reversibly bind the catalytic site (Fig. 2B). As expected, the α-galactose configured cyclosulfamidate **4** adopts a $^{4}C_{1}$ "Michaelis-like" complex conformation in the active site. Interestingly, the NH from the cyclosulfamidate moiety forms a hydrogen bond with the acid/base Asp231.

Thermostability of agalsidase beta in the presence of α-cyclosulfamidate 4 and Gal-DNJ 8

Competitive α-galactosidase inhibitors, including Gal-DNJ **8**, are currently investigated in clinical studies as stabilizers of the recombinant enzyme, where they are employed to enhance enzyme replacement efficacy. In such a treatment regime, the enzyme and active site inhibitor are administered jointly.[11,14,16] The stability of a recombinant enzyme relative to the temperature is considered to reflect well its stability in body circulation,[31] and can be measured with ease, also in the presence of competitive inhibitors designed to stabilize the protein fold.[31,32] Accordingly, we performed thermal stability assays (TSAs) on agalsidase beta alone and in the presence of increasing concentrations of **2**, **4** or **8**.

Thermal melting profiles of lysosomal α-gal A revealed that α-gal A is most thermostable at pH 5.5 (Fig. 3A and S4†), which is consistent with α-gal A being a lysosomal enzyme. α-Gal-cyclosulfamidate **4** stabilizes α-gal A at pH 7.4 with a ΔTm_{max} of 17.4 °C, compared to a ΔTm_{max} of 34.3 °C produced by Gal-DNJ **8** (Fig. 3A and S4†). TSA effects at acidic pHs were lower for both **4** and Gal-DNJ **8**, with recorded ΔTm_{max} values of 9.3 °C and 22.3 °C for **4** and Gal-DNJ **8**, respectively at pH 5.5, and ΔTm_{max} values of 9.7 °C and 21.2 °C for the same compounds at pH 4.5. Surprisingly, we observed no thermostabilization effect on α-gal A in the presence of α-gal-cyclosulfate **2**, despite this compound being an irreversible α-galactosidase inhibitor. Possibly, the sulfate group does not provide the optimal enzyme–ligand interactions to induce stabilization of α-galactosidase when the ring is in the $^{1}S_{3}$ conformation adopted by covalently bound **2**, compared to the $^{4}C_{1}$ conformations adopted by both **4** and Gal-DNJ **8**.

Stabilization of agalsidase beta by α-cyclosulfamidate 4 in cell culture medium

Agalsidase beta shows poor stability in plasma at a pH of 7.3–7.4, with only ∼25% of the hydrolytic activity being retained after incubation of the enzyme at 1 μg mL^{-1} in human plasma for 30 minutes.[33] Given the stabilizing effect observed for **4** in the above-described TSAs, we investigated the ability of this compound to stabilize agalsidase beta in culture medium at physiological pH compared to Gal-DNJ **8**.[34] We first investigated the stabilization effect of the inhibitors in culture medium alone, as a surrogate measure for plasma stability (Fig. 3B and C). Incubation of agalsidase beta (25 μL of 2.5 μg μL^{-1}) in cell culture medium (Dulbecco's Modified Eagles Medium/Nutrient Mixture F-12 (DMEM/F12), supplemented with 10% fetal calf serum and 1% penicillin/streptomycin) led to 80% loss of activity within 15 min, in line with the poor stability of this enzyme in blood plasma. To assess the stabilizing effects of **2**, **4** and Gal-DNJ **8** in cell culture media, agalsidase beta was incubated with increasing concentrations of these compounds, followed by capture of the enzyme on concanavalin A (ConA) sepharose beads, washing to remove the bound inhibitor, and quantification of residual α-galactosidase activity with the 4MU-α-gal substrate. Media stabilization of agalsidase beta followed the same trend as observed in TSAs, with **2** failing to stabilize the enzyme and instead irreversibly inhibiting agalsidase beta. In contrast, α-gal-cyclosulfamidate **4** and Gal-DNJ **8** both prevented inactivation of agalsidase beta in cell culture media (pH 7.2) and ∼75% residual α-gal-A activity was retained after incubation with **4** (at 500 μM) or Gal-DNJ **8** (at 50 μM) (Fig. 3C).

Competitive activity-based protein profile (ABPP) on recombinant α-galactosidases

We studied the binding of α-cyclosulfamidate **4** and Gal-DNJ **8** to the commercial α-galactosidases: agalsidase beta (Fabrazyme®), agalsidase alpha (Replagal®) and α-galactosidase B (*N*-acetylgalactosaminidase, NAGA) by competitive activity-based protein profiling (ABPP, Fig. 4). Enzymes were incubated with increasing concentrations (ranging from 0 to 1000 μM) of both α-cyclosulfamidate **4** and Gal-DNJ **8** for 30 min at 37 °C, followed by incubation with 0.2 μM of an α-galactosidase Cy5 activity-based probe (ABP **10**, Fig. S5†) for 30 min at 37 °C. After incubation, the samples were analyzed using sodium dodecyl sulfate–polyacrylamide gels (SDS–PAGE), followed by a fluorescent scan of the gels as previously described.[27,33] Competitive ABPP revealed that α-cyclosulfamidate **4** (100–500 μM) and Gal-DNJ **8** (1–10 μM) inhibit both recombinant human α-galactosidases and *N*-acetylgalactosaminidase.

In situ treatment of cultured fibroblasts from patients with Fabry disease

We next investigated whether the stabilizing effect of α-cyclosulfamidate **4** towards agalsidase beta produced an improvement in the cellular uptake of the enzyme by fibroblasts. We performed *in situ* studies in 5 different primary cell lines obtained from adult male volunteers: wild-type (WT, control) representing normal α-gal A activity, 2 classic Fabry mutant fibroblasts (R301X and D136Y) with no α-gal A activity and 2 atypical variant Fabry mutants (A143T and R112H) with substantially lowered residual α-gal A activity. Fibroblasts were

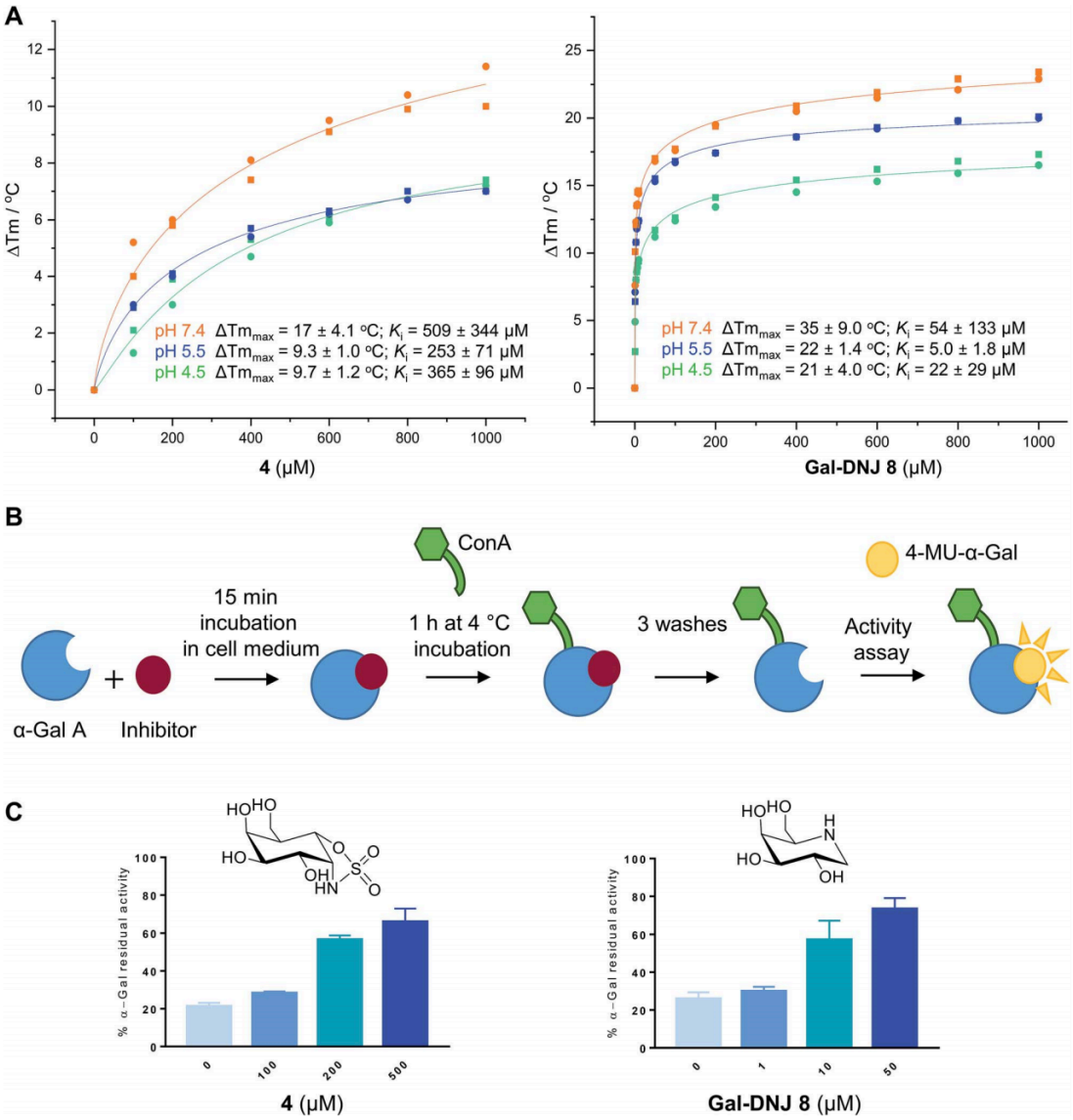


Fig. 3 Effect of α-cyclosulfamidate **4** and Gal-DNJ **8** on the thermal stability and cell culture medium stability of agalsidase. (A) Heat-induced melting profiles of lysosomal α-gal A recorded by thermal shift assay, measured at pH 4.5, 5.5 and 7.4 in the presence of α-cyclosulfamidate **4** and Gal-DNJ **8**. Melting points (Tm) were determined through thermal shift analyses by monitoring the fluorescence of Sypro Orange dye (λ_{em} 585 nm) as a function of temperature (see the ESI†). (B) Schematic representation of stabilization effect assay. Agalsidase beta was incubated with an inhibitor for 15 min in DMEM/F-12 medium and subsequently incubated with ConA sepharose beads for 1 h at 4 °C and washed to remove the inhibitor. Residual α-gal activity was quantified with the 4-MU-α-gal substrate. (C) Percentage of α-gal A residual activity after 15 min of incubation in DMEM/F-12 medium in the presence of inhibitors α-gal-cyclosulfamidate **4** (at 0, 100, 200, and 500 μM) and Gal-DNJ **8** (0, 1, 10 and 50 μM), followed by post final ConA purification. Percentages are calculated considering the 100% activity of α-gal A obtained at 0 min incubation time ($n = 2$, error bars indicate mean ± standard deviation).

incubated with 0.5% DMSO (untreated) or either **4** (200 μM) or Gal-DNJ **8** (20 μM) (in blue), agalsidase beta (100 ng) or with a combination of both enzyme and stabilizing agent (in green) (Fig. 5A). After 24 h treatment, the cells were harvested and homogenized, and the intracellular α-gal A activity of the corresponding cell lysates was measured. The WT cell line presented normal α-gal A activity while untreated classic Fabry patients (R301X and D136Y) and variant mutation samples

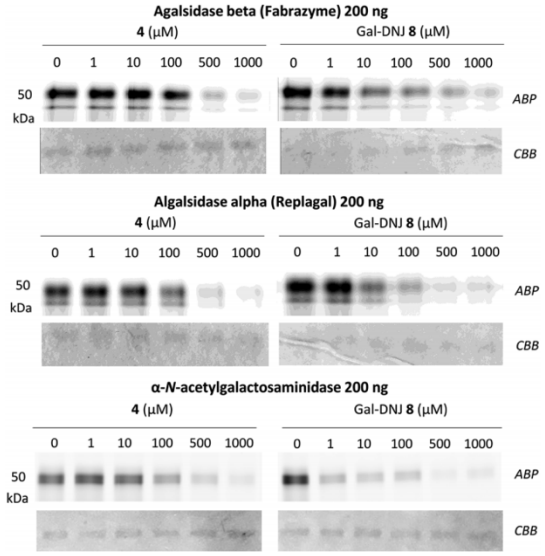


Fig. 4 Competitive ABPP in α-galactosidases. α-Galactosidases (agalsidase beta 200 ng and agalsidase alpha 200 ng) and α-*N*-acetylgalactosaminidase (NAGA, 200 ng) were pre-incubated with α-cyclosulfamidate **4** (0–1000 μM) or Gal-DNJ **8** (0–1000 μM) for 30 min followed by fluorescent labelling with Cy5 α-galactosidase ABP **10**. ABP: activity-based probe, CBB: Coomassie brilliant blue staining.

(A143T and R112H) showed reduced enzymatic activity. None of the cell lines, not even classical Fabry fibroblasts R301X and D136Y, showed a significant increase in α-gal A activity when incubated with **4** (200 μM) or **8** (20 μM) alone for 24 h. Of note, Gal-DNJ **8** is known to enhance α-gal A activity in R301Q lymphoblasts after *in situ* 4 day treatment of a 100 μM daily dose.[34] Treatment with agalsidase beta showed a considerable increase in α-gal A activity in all the studied cell lines. This effect was improved in most cases with the combinatorial treatment of agalsidase beta and **4** or **8** after 24 h of incubation. We also measured α-gal A activity in media in order to confirm that the increase in α-gal A activity in cell lysates is due to stabilization of the enzyme (Fig. 5B). Thus, culture media were collected before harvesting the cells and α-gal A activity was measured after ConA purification. α-Gal A activity in media was at least double in all the cell lines treated with α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM), consistent with these compounds preventing α-gal A degradation during cell culture.

Gb3 and lysoGb3 levels are corrected by α-cyclosulfamidate 4

Generally, Fabry patients present elevated Gb3 which is further metabolized by acid ceramidase into lysoGb3 in lysosomes.[35] LysoGb3 constitutes a signature of Fabry disease and allows diagnostic monitoring of disease progression,[2,3,36–38] and has been linked to neuronopathic pain and renal failure through its effect on nociceptive neurons and podocytes.[39–42] We investigated whether co-administration of α-cyclosulfamidate **4** and

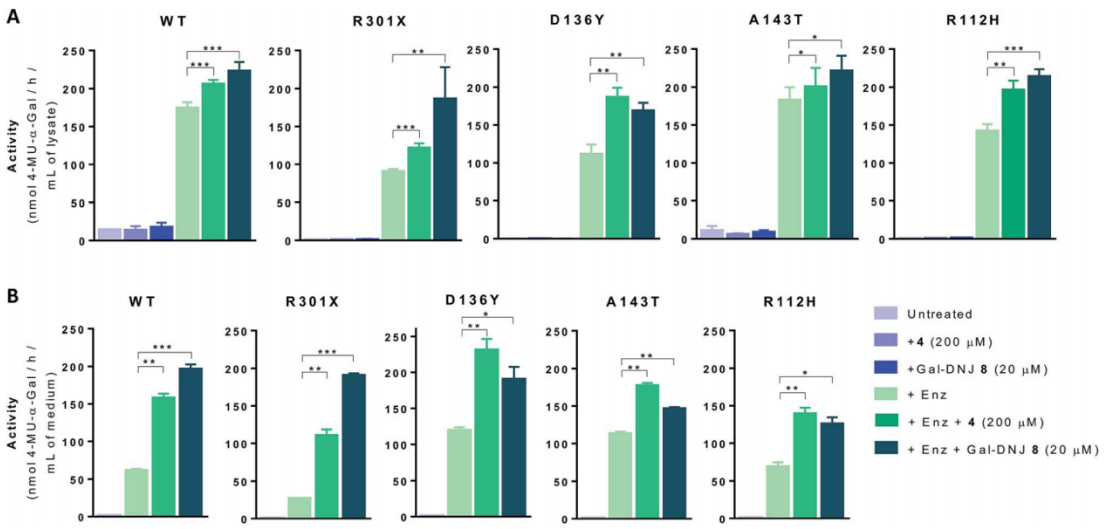


Fig. 5 Effect on α-gal A activity in fibroblast culture and medium following treatment with α-cyclosulfamidate **4** and Gal-DNJ **8**. (A) Fibroblasts of WT, classic Fabry (R301X and D136Y) and variant Fabry (A143T and R112H) were untreated or incubated with α-cyclosulfamidate **4** (200 μM), Gal-DNJ **8** (20 μM), agalsidase beta (200 ng mL^{-1}) or a combination of enzyme and stabilizing agent for 24 h. Then, the medium was collected, cells were harvested, and α-gal A activity was measured in the cell homogenates by 4-MU-α-gal assay. In all cell lines co-administration of α-cyclosulfamidate **4** or Gal-DNJ **8** with agalsidase beta increased intracellular α-gal A activity when compared to cells treated with only agalsidase beta. (B) α-Gal A activity in cell culture medium samples was measured after ConA purification. α-Gal A activity is at least two times higher in all the cell lines treated with α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM). Reported activities are mean ± standard deviation from two biological replicates, each with two technical replicates, *$p < 0.5$; **$p < 0.01$; ***$p < 0.001$.

Gal-DNJ **8** with agalsidase beta would also have a positive effect in correcting these toxic metabolite levels. Gb3 and lysoGb3 levels from *in situ* treated cells were measured by LC-MS/MS (Fig. 6). Normal Gb3 and lysoGb3 levels observed in wild-type cells are in the range of around 2000 pmol mL^{-1} and 2 pmol mL^{-1} of Gb3 and lysoGb3, respectively. Cultured fibroblasts from classic Fabry patients (R301X and D136Y) treated with agalsidase beta resulted in a reduction of Gb3 and lysoGb3. This reduction was similar when fibroblasts were treated with the combination of α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) and agalsidase beta. A variant Fabry A143T cell line presents normal Gb3 and lysoGb3 basal levels, whereas in R112H fibroblasts, these metabolites are increased and not corrected by agalsidase beta itself or inhibitor–agalsidase beta combination treatment (Fig. S6†).

In situ 4 day treatment of cultured fibroblasts: increased α-gal A activity and Gb3 metabolite correction by α-cyclosulfamidate 4

We next investigated whether the beneficial effect could be potentiated by extended incubation treatments. Thus, WT and classic Fabry (R301X) fibroblasts were treated with agalsidase beta (200 ng mL^{-1}) or with a combination of enzyme (200 ng mL^{-1}) and α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) for 4 days. Fibroblasts were treated every 24 h with new medium and/or enzyme with or without inhibitor, and medium samples were collected for α-gal A activity quantification (see ESI Fig. S7†). α-Gal A activity was 3–4 times higher in fibroblasts treated with the combination of recombinant α-gal A and α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) than those treated with agalsidase beta alone (Fig. 7A). This increase in α-gal A activity correlates with the reduction of lyso-Gb3 from ∼14 pmol mL^{-1} to ∼4 pmol mL^{-1} in the cell lysates of Fabry (classic R301X) fibroblasts (Fig. 7B). We finally studied whether the amount of required ERT could be decreased when stabilized with **4** or Gal-DNJ **8** and still produce a similar effect. WT and Fabry (classic R301X) fibroblasts were treated with agalsidase beta at 200 ng mL^{-1} and 100 ng mL^{-1}. A reduction in toxic metabolites can be achieved in 4 days with half the concentration of enzyme (100 ng mL^{-1}) when either α-cyclosulfamidate **4** (200 μM) or Gal-DNJ (20 μM) is added (Fig. 7C and D), with a similar reduction of toxic lyso-Gb3 from ∼10 pmol mL^{-1} to ∼5–6 pmol mL^{-1} in the cell lysates of Fabry (classic R301X) fibroblasts (Fig. 7E).

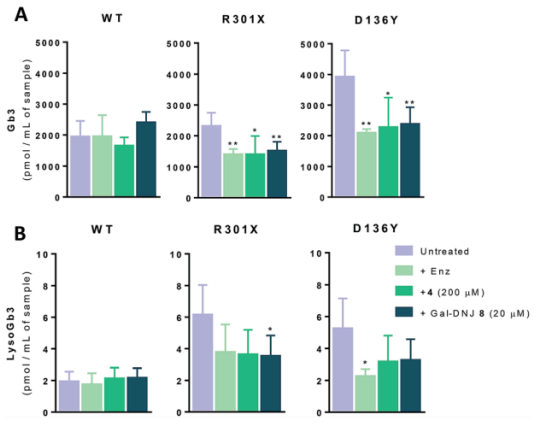


Fig. 6 Gb3 and lyso-Gb3 quantification in cultured fibroblasts treated with agalsidase beta co-administrated with α-cyclosulfamidate **4** and Gal-DNJ **8**. Gb3 (A) and lysoGb3 (B) levels (pmol mL^{-1} of sample) measured by LC-MS/MS in Fabry fibroblasts from WT and classic Fabry patients (R301X and D136Y) treated with agalsidase beta (200 ng mL^{-1}) with or without α-cyclosulfamidate **4** (200 μM) and Gal-DNJ **8** (20 μM) for 24 h. Reported activities are mean ± standard deviation from two biological replicates, each with two technical replicates, $*p < 0.5$; $**p < 0.01$.

Discussion

ERT with intravenous administration of recombinant human α-D-galactosidase (agalsidase beta, Fabrazyme® or agalsidase alpha, Replagal®) reduces the levels of Gb3 and lyso-Gb3 in some tissues of Fabry patients, but its clinical efficacy is still limited.[5–7,43] The limited enzyme stability in plasma is a major drawback, and it is for this reason that enzyme active site binders that stabilize recombinant enzyme in circulation are pursued – with Gal-DNJ **8** (Migalastat®) currently in use in the clinic as the benchmark. Here we report the design and synthesis of the first-in-class conformational glycosidase inhibitor and α-gal A stabilizing agent, α-cyclosulfamidate **4**. We show that this compound reversibly and selectively inhibits agalsidase beta with an IC_{50} value of 67 μM and a K_i of 110 μM. *Ab initio* metadynamics calculations and structural analysis of α-cyclosulfamidate **4** in complex with agalsidase beta show that this inhibitor binds in a $^4C_1$ conformation mimicking the Michaelis complex conformation. We demonstrate that α-cyclosulfamidate **4** stabilizes recombinant human α-D-galactosidase (agalsidase beta, Fabrazyme®) in thermal stabilization assays and show that this prevents its degradation in cell culture medium. We further show that both α-gal-cyclosulfamidate **4** and Gal-DNJ **8** stabilize the enzyme more significantly at neutral pH than under acidic conditions (ΔTm_{max} difference of 8.1 °C for **4** and ΔTm_{max} difference of 12.2 °C for Gal-DNJ **8**).

To further study the stabilizing effect, we investigated whether α-gal-cyclosulfamidate **4** and Gal-DNJ **8** would stabilize α-gal A activity under *in situ* cell conditions. Treatment of fibroblasts (WT, classic and variant Fabry cell lines) with only α-gal-cyclosulfamidate **4** (at 200 μM) and Gal-DNJ **8** (at 20 μM) for 24 h shows no effect on α-D-galactosidase activity. However, we observe an increased α-D-galactosidase activity in all cells treated with the combination of agalsidase beta and stabilizing agents (**4** at 200 μM and **8** at 20 μM) when compared to the cells treated only with agalsidase beta. This result also correlates with an increased α-D-galactosidase activity in the cell medium of the cells treated with enzyme and **4** or **8**. The stabilizing effect is more pronounced when cells are treated for longer time (4 days), suggesting that the agalsidase beta complexed with α-gal-cyclosulfamidate **4** or Gal-DNJ **8** is stabilized in the cell medium, internalized and dissociated from the reversible inhibitor in the lysosomes. Finally, co-administration of α-cyclosulfamidate **4** or

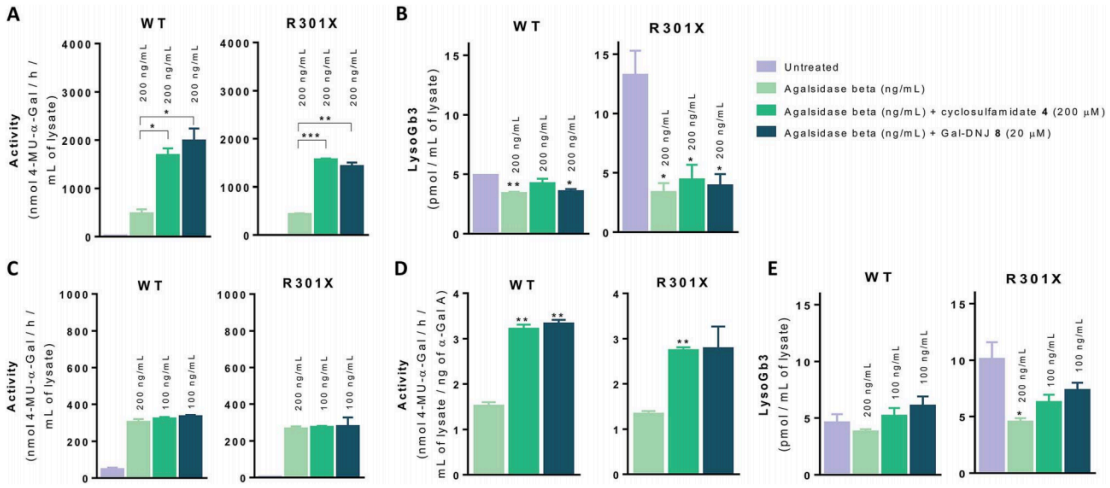


Fig. 7 Effect on α-gal A activity and lyso-Gb3 correction in cultured fibroblasts treated with α-cyclosulfamidate **4** and Gal-DNJ **8**. Fibroblasts of WT and classic Fabry (R301X) were incubated with agalsidase beta (200 ng mL^{-1}) or the combination of enzyme (200 ng mL^{-1} or 100 ng mL^{-1}) and stabilizing agent (**4** 200 μM or **8** 20 μM) for 4 days. Then, the medium was collected, cells were harvested, and α-gal A activity was measured in the cell homogenates by 4-MU-α-gal assay. (A) Intracellular α-gal A activity in fibroblasts treated with agalsidase beta (200 ng mL^{-1}) or the combination of α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) with agalsidase beta (200 ng mL^{-1}) for 4 days. (B) LysoGb3 levels measured by LC-MS/MS in Fabry fibroblasts from panel A. (C) Intracellular α-gal A activity is comparable in cell lines treated with the combination of α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) but this requires only half the concentration of agalsidase beta (100 ng mL^{-1}). (D) Intracellular α-gal A activity per ng of agalsidase beta in fibroblasts treated with agalsidase beta (100 ng mL^{-1}) or the combination of α-cyclosulfamidate **4** (200 μM) or Gal-DNJ **8** (20 μM) with half the concentration of agalsidase beta (100 ng mL^{-1}). (E) LysoGb3 levels measured by LC-MS/MS in Fabry fibroblasts from panel C or D. Reported lipid levels are mean ± standard deviation from two biological replicates, each with two technical replicates, *$p < 0.5$; **$p < 0.01$; ***$p < 0.001$.

Gal-DNJ **8** with agalsidase beta highlights a similar correction of toxic Gb3 and lyso-Gb3 metabolite levels as with the ERT alone. Importantly, similar α-gal A activity and correction of toxic metabolites is achieved with half the concentration of agalsidase beta when this is stabilized by α-cyclosulfamidate **4** or Gal-DNJ **8**. The synergy between Gal-DNJ **8** and the human recombinant α-gal A in cultured fibroblasts from Fabry patients has recently been demonstrated both in agalsidase alpha and beta.[16,19] This synergism, together with our agalsidase beta stabilization results, supports the idea that the efficacy of a combination treatment may be superior to ERT or PC alone for several reasons. Co-administration of ERT and the active site inhibitor may be effective in all Fabry patients, independent of mutations in their endogenous α-gal A. Furthermore, stabilization of the recombinant human α-gal A by a stabilizing agent may reduce the required enzyme dosages or extend IV injection intervals, and therefore improve the patient's lifestyle and reduce side effects and treatment costs.

Conclusions

In conclusion, we have developed a new class of α-D-galactosidase inhibitors based on cyclophellitol cyclosulfamidate as a conformational Michaelis complex isostere. Although cyclosulfamidate **4** is a 1000-fold weaker inhibitor of recombinant α-gal A compared to Gal-DNJ **8** *in vitro*, it stabilizes α-gal A *in cellulo* at only 10 fold higher concentration, and we argue that non-basic, competitive glycosidase inhibitors are attractive starting points for clinical development as stabilizers of (recombinant) glycosidases in the context of lysosomal storage disorders. Also, compound **4** together with its structural isosteres (**3** and **5**) comprise a new class of competitive glycosidase inhibitors, and stabilizes agalsidase beta for the first time, not by the glycoside configurational mimicry and basic nature that is the hallmark of iminosugars (including Migalastat®), but by configurational and conformational mimicry of the Michaelis complex. Michaelis complex or product-like conformational competitive inhibitors have been reported to act on other glycosidases, for instance, thio-oligosaccharides[44–46] and kifunensine.[47,48] We believe that transferring the structural characteristics of our cyclosulfamidates to differently configured structural analogues may yield potent and selective competitive inhibitors targeting glycosidases and that these may have biological or biomedical value in their own right, be it as stabilizing agents or as classical enzyme inhibitors.

Conflicts of interest

There are no conflicts to declare.

Author contributions

M. A., J. M. F. G. A., H. S. O. and G. J. D. conceived and designed the experiments. M. A., C. H. and A. S. carried out synthesis of

inhibitors under supervision of G. A. v. d. M and J. D. C. C. L. R. performed *ab initio* metadynamics calculations under supervision of C. R. R. R. and L. W. carried out structural studies on enzyme–inhibitor complexes and thermostability assays under supervision of G. J. D. M. A., M. J. F and K. K. determined the IC_{50} values and kinetic parameters, and performed agalsidase beta stabilization studies *in vitro* and *in cellulo*, and lipid metabolite quantification. M. A., J. M. F. G. A., G. J. D and H. S. O. wrote the manuscript with input from all authors.

Acknowledgements

We thank The Netherlands Organization for Scientific Research (NWO-CW, ChemThem grant to J. M. F. G. A. and H. S. O., and NWO TOP grant 2018-714.018.002 to 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. for financial support and postdoctoral contract to M. A.), the Spanish Ministry of Science, Innovation and Universities (MICINN) and the European Regional Development Fund (Fondo Europeo de Desarrollo Regional, FEDER) (CTQ2017-85496-P to C. R.), the Agency for Management of University and Research Grants of Generalitat de Catalunya (AGAUR) (2017SGR-1189 to C. R.) and the Spanish Structures of Excellence María de Maeztu (MDM-2017-0767 to C. R.). C. H. is supported by Villum Foundation (VKR023110). R. J. R. is supported by the BBSRC (BB/M011151/1). We thank the Diamond Light Source for access to beamline i02 and i04 (proposal number mx-13587) and the Barcelona Supercomputing Center for providing the computer resources at MareNostrum and the technical support (proposal RES-QCM-2017-2-0011). G. J. D. is supported by the Royal Society though the Ken Murray Research Professorship.

References

1 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, 490–495.
2 J. M. Aerts, J. E. Groener, S. Kuiper, W. E. Donker-Koopman, A. Strijland, R. Ottenhoff, C. van Roomen, M. Mirzaian, F. A. Wijburg, G. E. Linthorst, A. C. Vedder, S. M. Rombach, J. Cox-Brinkman, P. Somerharju, R. G. Boot, C. E. Hollak, R. O. Brady and B. J. Poorthuis, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2812–2817.
3 H. Gold, M. Mirzaian, N. Dekker, M. J. Ferraz, J. Lugtenburg, J. D. C. Codée, G. A. Van Der Marel, H. S. Overkleeft, G. E. Linthorst, J. E. M. Groener, J. M. Aerts and B. J. H. M. Poorthuis, *Clin. Chem.*, 2013, **59**, 547–556.
4 R. J. Desnick, Y. A. Ioannou, C. M. Eng, α-Galactosidase A Deficiency: Fabry Disease, in *The Online Metabolic and Molecular Bases of Inherited Diseases*, ed. David Valle, *et al.*, McGraw-Hill, New York, NY, 2014, DOI: 10.1036/ommbid.181.
5 M. Arends, M. Biegstraaten, D. A. Hughes, A. Mehta, P. M. Elliott, D. Oder, O. T. Watkinson, F. M. Vaz, A. B. P. Van Kuilenburg, C. Wanner and C. E. M. Hollak, *PLoS One*, 2017, **12**, 1–17.
6 M. Arends, F. A. Wijburg, C. Wanner, F. M. Vaz, A. B. P. van Kuilenburg, D. A. Hughes, M. Biegstraaten, A. Mehta, C. E. M. Hollak and M. Langeveld, *Mol. Genet. Metab.*, 2017, **121**, 157–161.
7 T. Alegra, F. Vairo, M. V de Souza, B. C. Krug and I. V. D. Schwartz, *Genet. Mol. Biol.*, 2012, **35**, 947–954.
8 A. Markham, *Drugs*, 2016, **76**, 1147–1152.
9 F. E. Mohamed, L. Al-Gazali, F. Al-Jasmi and B. R. Ali, *Front. Pharmacol.*, 2017, **8**, 448.
10 S. Ishii, H.-H. Chang, K. Kawasaki, K. Yasuda, H.-L. Wu, S. C. Garman and J.-Q. Fan, *Biochem. J.*, 2007, **406**, 285–295.
11 E. R. Benjamin, J. J. Flanagan, A. Schilling, H. H. Chang, L. Agarwal, E. Katz, X. Wu, C. Pine, B. Wustman, R. J. Desnick, D. J. Lockhart and K. J. Valenzano, *J. Inherited Metab. Dis.*, 2009, **32**, 424–440.
12 D. A. Hughes, K. Nicholls, S. P. Shankar, G. Sunder-Plassmann, D. Koeller, K. Nedd, G. Vockley, T. Hamazaki, R. Lachmann, T. Ohashi, I. Olivotto, N. Sakai, P. Deegan, D. Dimmock, F. Eyskens, D. P. Germain, O. Goker-Alpan, E. Hachulla, A. Jovanovic, C. M. Lourenco, I. Narita, M. Thomas, W. R. Wilcox, D. G. Bichet, R. Schiffmann, E. Ludington, C. Viereck, J. Kirk, J. Yu, F. Johnson, P. Boudes, E. R. Benjamin, D. J. Lockhart, C. Barlow, N. Skuban, J. P. Castelli, J. Barth and U. Feldt-Rasmussen, *J. Med. Genet.*, 2017, **54**, 288–296.
13 G. Sunder-Plassmann, R. Schiffmann and K. Nicholls, *Expert Opin. Orphan Drugs*, 2018, **6**, 301–309.
14 C. Porto, M. Cardone, F. Fontana, B. Rossi, M. R. Tuzzi, A. Tarallo, M. V. Barone, G. Andria and G. Parenti, *Mol. Ther.*, 2009, **17**, 964–971.
15 E. R. Benjamin, R. Khanna, A. Schilling, J. J. Flanagan, L. J. Pellegrino, N. Brignol, Y. Lun, D. Guillen, B. E. Ranes, M. Frascella, R. Soska, J. Feng, L. Dungan, B. Young, D. J. Lockhart and K. J. Valenzano, *Mol. Ther.*, 2012, **20**, 717–726.
16 C. Porto, A. Pisani, M. Rosa, E. Acampora, V. Avolio, M. R. Tuzzi, B. Visciano, C. Gagliardo, S. Materazzi, G. La Marca, G. Andria and G. Parenti, *J. Inherited Metab. Dis.*, 2012, **35**, 513–520.
17 D. G. Warnock, D. G. Bichet, M. Holida, O. Goker-Alpan, K. Nicholls, M. Thomas, F. Eyskens, S. Shankar, M. Adera, S. Sitaraman, R. Khanna, J. J. Flanagan, B. A. Wustman, J. Barth, C. Barlow, K. J. Valenzano, D. J. Lockhart, P. Boudes and F. K. Johnson, *PLoS One*, 2015, **10**, 1–17.
18 J. Castelli, S. Sitaraman, D. J. Lockhart, K. Nicholls, P. F. Boudes, R. Giugliani, D. P. Germain, D. A. Hughes, C. J. Jennette, A. Mehta, A. Bragat and L. Barisoni, *Orphanet J. Rare Dis.*, 2012, **7**, 91.
19 A. Pisani, C. Porto, G. Andria and G. Parenti, *J. Inherited Metab. Dis.*, 2014, **37**, 145–146.
20 H. Brumer, P. F. G. Sims and M. L. Sinnott, *Biochem. J.*, 1999, **339**, 43–53.
21 A. I. Guce, N. E. Clark, E. N. Salgado, D. R. Ivanen, A. A. Kulminskaya, H. Brumer and S. C. Garman, *J. Biol. Chem.*, 2010, **285**, 3625–3632.

research papers

STRUCTURAL BIOLOGY

ISSN 2059-7983

A baculoviral system for the production of human β-glucocerebrosidase enables atomic resolution analysis

Rhianna J. Rowland,[a] Liang Wu,[a] Feng Liu[b] and Gideon J. Davies[a]*

[a]Department of Chemistry, York Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, United Kingdom, and [b]Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. *Correspondence e-mail: gideon.davies@york.ac.uk

Received 17 December 2019
Accepted 9 April 2020

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: β-glucocerebrosidase; Gaucher disease; insect–baculovirus expression; crystallization.

PDB references: β-glucocerebrosidase, 6tn1; complex with bis-Tris propane, 6tjk; complex with 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside, 6tjq; Cerezyme, 6tjj

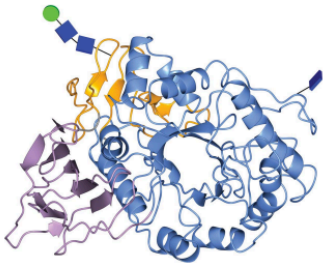
Supporting information: this article has supporting information at journals.iucr.org/d

The lysosomal glycoside hydrolase β-glucocerebrosidase (GBA; sometimes called GBA1 or GC_{ase}) catalyses the hydrolysis of glycosphingolipids. Inherited deficiencies in GBA cause the lysosomal storage disorder Gaucher disease (GD). Consequently, GBA is of considerable medical interest, with continuous advances in the development of inhibitors, chaperones and activity-based probes. The development of new GBA inhibitors requires a source of active protein; however, the majority of structural and mechanistic studies of GBA today rely on clinical enzyme-replacement therapy (ERT) formulations, which are incredibly costly and are often difficult to obtain in adequate supply. Here, the production of active crystallizable GBA in insect cells using a baculovirus expression system is reported, providing a nonclinical source of recombinant GBA with comparable activity and biophysical properties to ERT preparations. Furthermore, a novel crystal form of GBA is described which diffracts to give a 0.98 Å resolution unliganded structure. A structure in complex with the inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside was also obtained, demonstrating the ability of this GBA formulation to be used in ligand-binding studies. In light of its purity, stability and activity, the GBA production protocol described here should circumvent the need for ERT formulations for structural and biochemical studies and serve to support GD research.

1. Introduction

β-Glucocerebrosidase (glucosylceramidase; GBA; EC 3.2.1.45; often termed GBA1 or GC_{ase}) is a membrane-associated lysosomal enzyme belonging to the GH30 family of glycoside hydrolases (http://www.cazy.org; Lombard *et al.*, 2014). GBA is responsible for catalysing the hydrolysis of glycosphingolipids, specifically glucosylceramide (GlcCer), by the hydrolytic cleavage of β-linked glucose moieties from sphinogolipid aglycones (Brady *et al.*, 1965). Mechanistically, GBA is a retaining β-glucosidase which hydrolyses its substrates with net retention of β-anomeric stereochemistry (mechanisms are reviewed in Gloster & Davies, 2010; Zechel & Withers, 2000). Retention of anomeric configuration is achieved through the Koshland double-displacement mechanism (Koshland, 1953) using two key carboxylate-containing residues; in the case of human GBA, Glu340 serves as the catalytic nucleophile and Glu235 acts as the general acid/base (Fig. 1).

Inherited deficiencies in GBA activity cause an accumulation of GlcCer within lysosomes, subsequently leading to the most common lysosomal storage disorder, Gaucher disease (GD; Grabowski & Horowitz, 1997; Brady *et al.*, 1966). Traditionally, GD is classified into three clinical phenotypes



based on the presence of neurological manifestations and the rate of neuronopathic disease progression (Zhao & Grabowski, 2002). The non-neuropathic form, GD type 1, is the most common clinical type, accounting for over 90% of GD cases (Mistry *et al.*, 2011; Mehta, 2006). Types 2 and 3 are less common neuronopathic forms of GD involving the central nervous system (CNS), with a continuum of phenotypes ranging from death *in utero* to slowly progressive CNS deterioration over decades (Davies *et al.*, 2007; Erikson *et al.*, 1997). Consequently, GBA is of significant clinical importance.

Enzyme-replacement therapy (ERT) was introduced in the early 1990s as a first-line treatment for GD. The aim of ERT is to correct for the underlying GBA deficiency and alleviate clinical symptoms by supplying patients with active exogenous enzyme (Beck, 2018; Mistry & Abrahamov, 1997). In 1991, Ceredase (alglucerase), a glyco-modified GBA formulation purified from human placenta (Barton *et al.*, 1990, 1991; Furbish *et al.*, 1977), was approved for ERT. Subsequently, the Genzyme Corporation (Sanofi Genzyme, Cambridge, Massachusetts, USA) developed Cerezyme (imiglucerase), a recombinant formulation with mannose-terminated N-glycans produced in a Chinese hamster ovary (CHO) cell line (Serratrice *et al.*, 2016; Deegan & Cox, 2012). However, Cerezyme production was halted in 2009 by a vesivirus infection at Genzyme's production facility (Qiu *et al.*, 2013), promoting the development of alternative ERTs. In 2010, velaglucerase alfa (VPRIV; Shire HGT, Lexington, Massachusetts, USA), a recombinant product produced by specific gene activation in an HT-1080 cell line, was licenced for use (Ben Turkia *et al.*, 2013; Brumshtein *et al.*, 2010). Later, taliglucerase alfa (Elelyso; Pfizer, New York, USA), a plant-derived variant expressed in carrot-root cells, was approved in the US (Zimran *et al.*, 2016; Aviezer *et al.*, 2009; Shaaltiel *et al.*, 2007).

Given the hydrophobic nature of GBA and the post-translational glycosylation that is required to ensure proper folding, GBA cannot be produced in prokaryotic systems (Grace & Grabowski, 1990). Consequently, eukaryotic systems with the necessary post-translational modification capabilities must be employed, as exemplified by the existing ERT expression systems. In addition to these expression platforms, GBA production has been attempted in murine cells (Fabrega *et al.*, 2000), COS-1 cells (Grabowski *et al.*, 1989), seeds of the *Arabidopsis thaliana* plant (He *et al.*, 2012), glycoengineered *Nicotiana benthamiana* plants (Limkul *et al.*, 2016), *Pichia pastoris* (Sinclair & Choy, 2002) and baculoviral expression vector systems (BEVS; Martin *et al.*, 1988; Sinclair *et al.*, 2006; Sawkar *et al.*, 2006). Such diversity demonstrates the current lack of consensus on a robust and economical platform for GBA production for nonclinical use. Accordingly, there is a considerable ongoing reliance on expired ERT formulations for biochemical, mechanistic and structural studies. These ERT preparations can be incredibly costly, and are often only obtainable under a Material Transfer Agreement (MTA) and in limited supply. Given the clinical importance of GBA and the continuing development of novel GBA chaperones (Goddard-Borger *et al.*, 2012; Diot *et al.*, 2011; Hill *et al.*, 2011; Marugan *et al.*, 2011), inhibitors (Artola *et al.*, 2019; Kuo *et al.*, 2019; Zoidl *et al.*, 2019; Schröder *et al.*, 2018) and activity-based probes (ABPs; Artola *et al.*, 2017, 2019; Schröder *et al.*, 2018; Beenakker *et al.*, 2017), there is a pressing need for reliable sources of recombinant GBA to meet research demands and reduce the reliance upon ERT formulations.

The use of baculovirus expression vectors was first described in the 1980s and has since proved to be useful for the production of many recombinant proteins (Chambers *et al.*, 2018; Bonning & Hammock, 1996; Cameron *et al.*, 1989; Luckow & Summers, 1988). The *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) of the *Baculoviridae* family is the most applied baculoviral vector for recombinant protein production (Jarvis, 2003; Possee, 1997; Jarvis *et al.*, 1996; Blissard & Rohrmann, 1990). Lepidopteran insect-cell lines, such as *Spodoptera fruigiperda* (Sf9 cells) and *Trichoplusia ni* (BTI-Tn-5B1-4, Tn5, High Five cells), have been studied extensively as hosts for this viral vector (Bieniossek *et al.*, 2012; Fitzgerald *et al.*, 2006; Wilde *et al.*, 2014; Jarvis *et al.*, 1990). However, to our knowledge, no

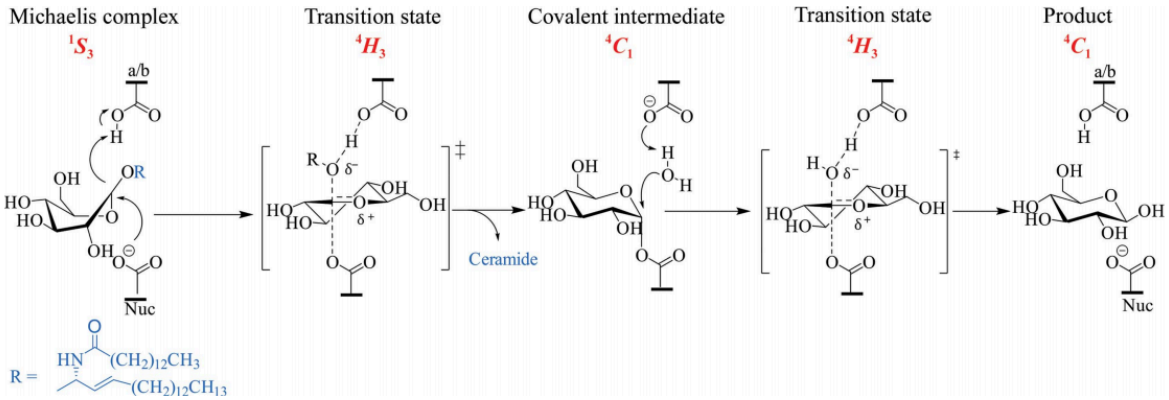


Figure 1
GBA hydrolyses glucosylceramide in a two-step double-displacement mechanism to yield ceramide and glucose with retention of β-stereochemistry.

insect-cell-produced GBA has been approved for therapeutic use because the glycoforms produced by insect cells can be immunogenic (Hancock *et al.*, 2008). Of note, High Five cells modify glycoproteins with a core α-1,3-fucose, which can induce immunogenic and allergic responses in humans. Several strategies have been developed to address the issue of incompatible N-glycosylation in insect-cell expression systems; however, these remain in their infancy (Zitzmann *et al.*, 2017).

Previous studies on the use of BEVS platforms for GBA production have shown some success, albeit with inconsistent results regarding protein quantity and quality. Early work by Martin *et al.* (1988) using the AcMNPV vector and polyhedrin promoter in Sf9 cells demonstrated GBA expression, with enzymatic activity detected in both the cell extract and culture medium. However, the efficiency of protein secretion was called into question in subsequent studies (Berg-Fussman *et al.*, 1993; Grabowski *et al.*, 1989). Additionally, Choy *et al.* (1996) demonstrated that GBA produced in AcMNPV-transfected Sf9 cells can be stored intracellularly. More recently, Sinclair *et al.* (2006) employed the *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus (OpNPV) for GBA production in Sf9 cells to investigate the effect of the full-length and shortened native signal sequences on GBA secretion. The full-length signalling construct was reported to produce 30% more enzymatic activity than the shortened construct but, in contrast to previous studies (Xu & Grabowski, 1998; Choy *et al.*, 1996), both constructs resulted in the secretion of the majority of the GBA into the medium (Sinclair *et al.*, 2006). In addition, wild-type GBA and an N370S GD-associated mutant have been expressed in High Five cells (Sawkar *et al.*, 2006; Wei *et al.*, 2011) to evaluate the stability of the protein and its susceptibility to chaperone-mediated stabilization; however, no information on protein yield or secretion efficiency have been reported for this expression system. Despite this success, little has since been published with regard to the production of GBA using BEVS, and there remains a lack of literature describing the establishment of a reliable baculoviral GBA expression platform for academic research purposes.

As part of our long-standing interest in the development of ABPs to study GBA and other glycosidases, we established a BEVS platform for the production of recombinant GBA for biochemical and structural studies. Here, we describe the production of active human GBA in insect cells using a MultiBac AcNPV-derived expression system (Bieniossek *et al.*, 2012), which circumvents the need for commercial sources of GBA. In this approach, Sf9 cells were used to generate the recombinant baculovirus, whilst High Five cells were employed for the production of recombinant GBA. An N-terminally truncated GBA gene, lacking its native signalling sequence, was used in conjunction with the honeybee melittin signal sequence (Tessier *et al.*, 1991), a widely used secretion signal in insect-cell expression, to drive the secretion of recombinant GBA into the cell medium. This recombinant GBA was active against the artificial substrate 4-methylumbelliferyl-β-D-glucopyranoside and exhibited an optimum thermal stability at pH 5.2, as expected for a lysosomal enzyme. Furthermore, our recombinant GBA was purified with a typical yield of 3.6–4.6 mg per litre of cell medium, providing sufficient protein for biochemical and structural analysis. Consequently, our GBA was studied by X-ray crystallography to generate a 0.98 Å resolution unliganded structure in a novel crystal form. This is the highest resolution structure of GBA deposited to date, allowing exquisite atomic resolution analysis of GBA that reveals two conformations of the catalytic acid/base residue. A structure in complex with the β-glucosidase inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside was also obtained, demonstrating the utility of this recombinant GBA for ligand-binding studies. Taken together, our results provide a standard method for nonclinical GBA production, which should help to reduce the reliance on ERT preparations in future *in vitro* studies.

2. Materials and methods

2.1. Generation of the recombinant transfer plasmid

The N-terminally truncated GBA gene was subcloned from the pGEn1-GBA plasmid (DNASU Clone ID HsCD00413213; Fig. 2*a*; gene obtained from the Glycoenzyme repository; http://glycoenzymes.ccrc.uga.edu/; Moremen *et al.*, 2018) using a Phusion (New England Biolabs) polymerase chain reaction (PCR) with the forward primer 5′-TACATTAGCTACATTT ATGCGGCCCGCCCCTGCATCCCTAAAAGC-3′ and the reverse primer 5′-CTAGTACTTCTCGACAAGCTTCTACT GGCGACGCCACAGGTAG-3′.

The linearized pOMNI plasmid backbone, containing the honeybee melittin signal sequence immediately following the translation start codon (Fig. 2*b*), was obtained by restriction digestion of an existing pOMNI plasmid using HindIII-HF and XmaI (New England Biolabs). The original pOMNI vector, containing the Tn7 transposon sequences (Tn7L and Tn7R) for Tn7 transposition, was kindly provided to the York Structural Biology Laboratory by the Berger Laboratory, University of Bristol. All DNA fragments were analysed on an agarose gel (1%) and were purified by gel extraction using a QIAquick Gel Extraction Kit (Qiagen).

The recombinant transfer plasmid was generated by sequence- and ligation-independent cloning (SLIC) of the GBA insert and linearized pOMNI backbone in One Shot TOP10 *Escherichia coli* cells (Invitrogen) using standard protocols (Li & Elledge, 2007, 2012). Briefly, the GBA insert and linearized backbone were independently treated with T4 DNA polymerase (New England Biolabs) for 30 min, followed by the addition of d-CTP. The GBA insert (1.0 ng μl^{-1}) and backbone (1.5 ng μl^{-1}) were treated together with RecA (New England Biolabs) in the presence of RecA buffer and ATP for 1 h at 37°C. The GBA insert and backbone were transformed into One Shot TOP10 *E. coli* cells by heat shock. The transfer-plasmid DNA was extracted and purified from overnight cultures of successful colonies in Luria–Bertani broth (LB) using a QIAprep Spin Miniprep Kit (Qiagen). The transfer plasmid was verified by restriction-digest analysis with HindIII and Sanger sequencing (Fig. 1*c*) using the forward

primer 5′-CAGCAGCGAAGTCGCCATAAC-3′ and the reverse primer 5′-CAGCCGGATCTTCTAGGCTC-3′.

2.2. Generation of the recombinant bacmid

The DH10EMBacY *E. coli* strain was generously provided by the Berger Laboratory, University of Bristol. The DH10EMBacY strain contains the EMBacY baculovirus shuttle vector (bacmid bMON14272) with a mini-attTn7 target site, a tetracycline-resistant helper plasmid (pMON7124) encoding the transposase enzyme, a yellow fluorescent protein (YFP) reporter gene, the LacZα gene and a kanamycin-resistance selection marker.

The recombinant bacmid was produced using the Tn7 transposition method in DH10EMBacY cells (Bieniossek *et al.*, 2012; Fitzgerald *et al.*, 2006; Geneva Biotech). Briefly, purified transfer-plasmid DNA was transformed into DH10EMBacY cells by electroporation. Super Optimal Broth medium supplemented with 20 m*M* glucose (SOC medium) was added and the cells were incubated for 4 h at 37°C before blue/white screening on LB agar plates containing kanamycin (50 µg ml^{-1}), gentamicin (15 µg ml^{-1}), tetracycline (15 µg ml^{-1}), IPTG (1 m*M*) and x-Gal (1×). White colonies were re-streaked and confirmed by Phusion colony PCR using the forward primer 5′-CCCAGTCACGACGTTGTAAAACG-3′ and the reverse primer 5′-AGCGGATAACAATTTCACAC AGG-3′. The recombinant bacmid was purified from LB cultures of successful colonies using a PureLink HiPure Plasmid DNA Purification Kit (Invitrogen) and was verified by Phusion PCR using the forward primer 5′-CAGCAGCG AAGTCGCCATAAC-3′ and the reverse primer 5′-CAGCC GGATCTTCTAGGCTC-3′ to amplify the GBA gene and the forward primer 5′-CCCAGTCACGACGTTGTAAAACG-3′ and the reverse primer 5′-AGCGGATAACAATTTCACAC AGG-3′ to amplify across the bacmid Tn7 insertion site.

2.3. Production of the recombinant baculovirus

The recombinant baculovirus was generated and amplified in Sf9 cells (clonal isolate of *S. frugiperda* Sf21 cells; IPLB-Sf21-AE) purchased from Invitrogen. Adherent Sf9 cells were grown at 28°C for two days in 60 ml Insect-XPRESS protein-free medium (Lonza Bioscience) supplemented with 2% fetal bovine serum (FBS). At log-phase growth, 2 ml of suspended Sf9 cells was seeded into each well of a six-well tissue-culture plate at a density of 0.45 × 10^6 cells ml^{-1} and allowed to settle for 10 min in a humidified incubator at 28°C to establish an adherent culture. 180 µl of a transfection mixture containing Insect-XPRESS medium (1.05 ml), recombinant bacmid DNA (~100 µg) and FuGENE HD (Promega) transfection agent

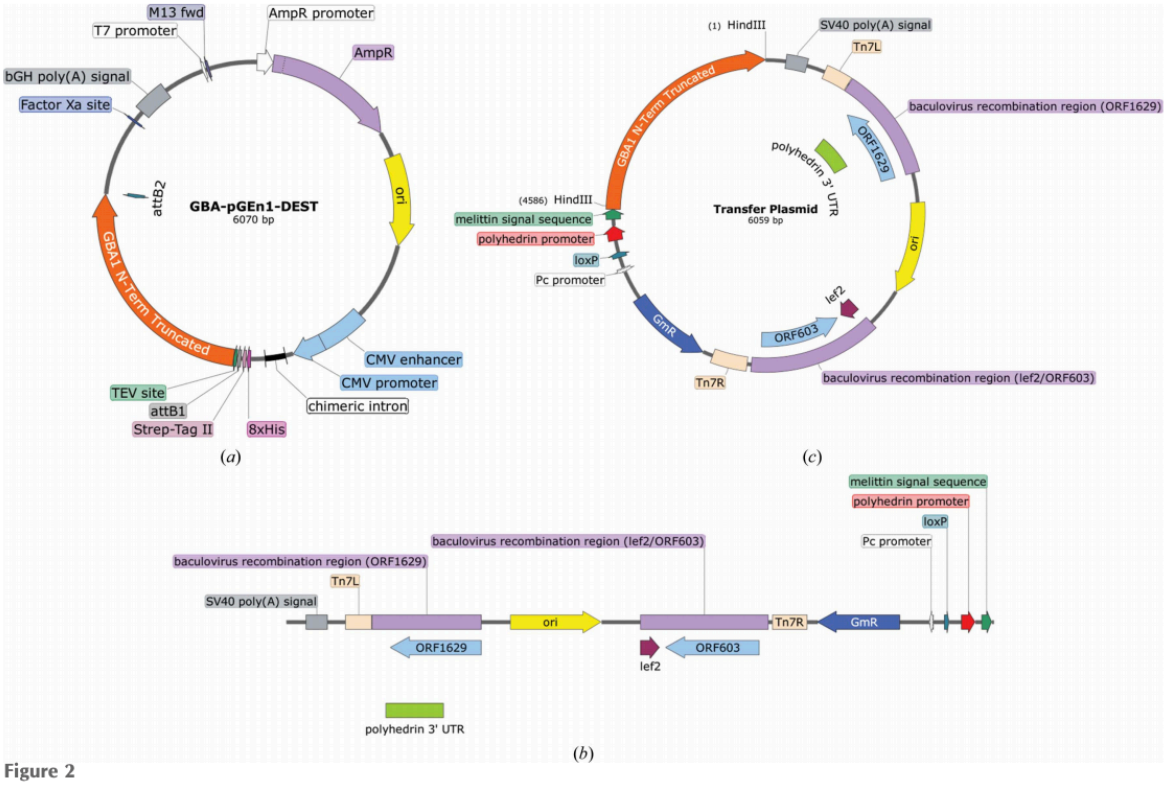


Figure 2
(*a*) The GBA-pGEn1-DEST plasmid containing the N-terminally truncated GBA gene. (*b*) The linearized plasmid backbone used for SLIC. (*c*) Sequenced GBA transfer plasmid generated by SLIC of the GBA (N-terminally truncated) insert and linearized plasmid backbone.

(31.5 µl) was added dropwise to each well of the six-well tissue-culture plate. The cells were incubated in a static humidified incubator at 28°C until ∼95% baculoviral transduction was achieved (∼2–3 days), as indicated by expression of the EMBacY YFP marker gene. The supernatant was collected by centrifugation at 200*g* for 5 min and FBS (0.2 ml) was added to yield the viral P1 stock. A 50 ml culture of Sf9 cells was prepared at 1×10^6 cells ml^{-1} in Insect-XPRESS medium and infected with 1 ml of viral P1 stock. The culture was incubated in a shaker incubator at 28°C and 87 rev min^{-1} until 95% transfection was achieved (∼2–3 days). The supernatant was collected by centrifugation at 200*g* for 5 min and FBS (1 ml) was added to yield the viral P2 stock.

2.4. Expression of GBA in High Five cells

A suspension adapted High Five cell line (BTI-Tn-5B1-4, Invitrogen) was prepared in a 60 ml culture in Express Five Serum Free Medium supplemented with 20 m*M* L-glutamine (Thermo Fisher Life Technologies). The culture was incubated at 28°C and 87 rev min^{-1} for ∼24 h. When a critical cell density (>2×10^6 cells ml^{-1}) had been reached, the culture was successively passaged into 100 ml, 600 ml, 1.8 l and 3.6 l Express Five Serum Free Medium. The 3.6 l culture (∼$1–2 \times 10^6$ cells ml^{-1}) was split into 6 × 600 ml cultures and infected with 750 µl of baculovirus P2 stock. The cultures were incubated at 28°C and 87 rev min^{-1} until YFP fluorescence was observed in 95% of the cells (∼2–3 days). The supernatant was harvested by centrifugation at 400*g* for 15 min at 4°C, followed by further clearing of debris by centrifugation at 4000*g* for 60 min at 4°C. DTT and PMSF were added to achieve final concentrations of 1 and 0.1 m*M*, respectively.

2.5. Protein purification

The conditioned supernatant was concentrated using a KrosFlo Research IIi Tangential Flow Filtration System with a 30 kDa mPES hollow-fibre filter module. GBA was purified using a previously outlined procedure (Sawkar *et al.*, 2006), with the addition of a size-exclusion step. Recombinant GBA was extracted from the medium by hydrophobic interaction chromatography using a TOYOPEARL Butyl-650C column (Tosoh Bioscience). The column was pre-equilibrated with 1.5 column volumes (CV) of buffer *A* (20 m*M* sodium acetate, 150 m*M* NaCl pH 5.0) and the protein was isocratically eluted into buffer *B* [20 m*M* sodium acetate, 150 m*M* NaCl, 50%(*v*/*v*) ethylene glycol pH 5.0] over 5 CV. GBA-containing fractions were pooled, diluted threefold in deionized water and purified by cation-exchange chromatography using a HiTrap Heparin Sepharose FF column (GE Healthcare) pre-equilibrated in buffer *A* [20 m*M* sodium acetate, 50 m*M* NaCl, 20%(*v*/*v*) ethylene glycol pH 5.0]. The protein was eluted with a linear gradient over 20 CV into buffer *B* [20 m*M* sodium acetate, 1 *M* NaCl, 20%(*v*/*v*) ethylene glycol pH 5.0]. Fractions containing GBA were pooled, diluted 15-fold in 20% ethylene glycol and purified by weak cation exchange on a HiTrap CM Sepharose FF column (GE Healthcare) pre-equilibrated with buffer *A* (30 m*M* sodium citrate, 0.01% Tween 80 pH 5.7). The protein was eluted in a linear gradient over 20 CV into buffer *B* (55 m*M* sodium citrate, 0.01% Tween 80 pH 6.3). GBA-containing fractions were pooled, concentrated to ∼1.5 ml using a 30 kDa Vivaspin concentrator (GE Healthcare) and purified using a Superdex S200 16/600 column (GE Healthcare) in SEC buffer (10 m*M* MES, 100 m*M* NaCl, 1 m*M* TCEP pH 6.5). GBA-containing fractions were concentrated to ∼10 mg ml^{-1} using a 30 kDa Vivaspin concentrator. Typical yields were 13–16.7 mg per preparation (3.6–4.6 mg per litre of culture medium). Macromolecule-production information is summarized in Table 1.

Table 1
Macromolecule-production information.

The 5′ end of the forward primer was designed to be complementary to the melittin signal sequence of the linearized p-OMNI backbone and the 5′ end of the reverse primer was designed to be complementary to the multi-insertion site of the linearized p-OMNI backbone (denoted in bold).

Source organism	*Homo sapiens*
DNA source	GBA-pGEn (DNASU: HsCD00413213; Moremen *et al.*, 2018)
Forward primer	**TACATTAGCTACATTTATGCG**GCCCGCCC CTGCATCCCTAAAAGC
Reverse primer	**CTAGTACTTCTCGACAAGCTT**CTACTGGCG ACGCCACAGGTAG
Cloning vector	pOMNI-derived vector (Sari *et al.*, 2016)
Expression vector	DH10EMBacY AcNPV-derived vector (Fitzgerald *et al.*, 2006)
Expression host	*Trichoplusia ni* (BTI-Tn-5B1-4, High Five cells)
Complete amino-acid sequence of the construct produced	ARPCIPKSFGYSSVVCVCNATYCDSFDPPT FPALGTFSRYESTRSGRRMELSMGPIQA NHTGTGLLLTLQPEQKFQKVKGFGGAMT DAAALNILALSPPAQNLLLKSYFSEEGI GYNIIRVPMASCDFSIRTYTYADTPDFQ LHNFSLPEEDTKLKIPLIHRALQLAQRP VSLLASPWTSPTWLKTNGAVNGKGSLKG QPGDIYHQTWARYFVKFLDAYAEHKLQF WAVTAENEPSAGLLSGYPFQCLGFTPEH QRDFIARDLGPTLANSTHHNVRLLMLDD QRLLLPHWAKVVLTDPEAAKYVHGIAVH WYLDFLAPAKATLGETHRLFPNTMLFAS EACGSKFWEQSVRLGSWDRGMQYSHSII TNLLYHVVGWTDWNLALNPEGGPNWVRN FVDSPIIVDITKDTFYKQPMFYHLGHFS KFIPEGSQRVGLVASQKNDLDAVALMHP DGSAVVVVLNRSSKDVPLTIKDPAVGFL ETISPGYSIHTYLWRRQ

2.6. Michaelis–Menten kinetics

Michaelis–Menten kinetics were assayed using the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside (4-MU-Glc). GBA was prepared at 20 n*M* in kinetics buffer [McIlvaine buffer; 150 m*M* disodium hydrogen phosphate, citric acid pH 5.2 supplemented with 0.2%(*v*/*v*) taurocholate, 0.1%(*v*/*v*) Triton X-100 and 0.1%(*v*/*v*) bovine serum albumin (BSA)]. 4-MU-Glc was prepared at 5 m*M* in kinetics buffer and diluted twofold to yield solutions at 2.5, 1.25, 0.625, 0.313, 0.156, 0.078 and 0.039 m*M*. Each substrate solution (25 µl) was added to the wells of a black 384-well polystyrene plate. GBA (25 µl, 20 n*M*) was added to each well to give a final enzyme concentration of 10 n*M* and final substrate concentrations of 2.5, 1.25, 0.625, 0.313, 0.156, 0.078, 0.039 and 0.0195 m*M*. Activity against 4-MU-Glc was monitored continuously over

5 min at 37°C by measuring the fluorescence of liberated 4-MU (λ_{ex} = 360–320 nm, λ_{em} = 450–430 nm) using a CLARIOstar Plus microplate reader (BMG Labtech). Assays were performed in quadruplicate for each substrate concentration. A linear calibration was generated by measuring the fluorescence of the 4-MU product (λ_{ex} = 360–320 nm, λ_{em} = 450–430 nm) prepared at serial dilutions of 125, 62.5, 31.25, 15.63, 7.81, 3.91, 1.95 and 0.98 µ*M* in kinetics buffer. Each 4-MU concentration was measured in quadruplicate.

All data were processed using the *Origin* graphing software. Using the 4-MU calibration, the rate of substrate hydrolysis (V) was determined at each substrate concentration. The rates (V) were plotted against substrate concentration [S] and fitted by nonlinear regression to the Michaelis–Menten equation {rate = V_{max}[S]/(K_m + [S])} to generate values of K_m, V_{max} and k_{cat} using the relationship k_{cat} = V_{max}/[Enz].

2.7. Thermal stability

Triplicate 25 µl reactions of 2 µ*M* GBA and 5× SYPRO Orange dye (Fisher Scientific) were prepared in McIlvaine buffer at pH 5.2 and pH 7.0. The Thermofluor assay was performed using a Stratagene Mx3005P qPCR instrument. The SYPRO Orange dye was excited at λ_{ex} = 517 nm and the resulting fluorescence was monitored at λ_{em} = 585 nm with a temperature ramp from 25 to 95°C at a rate of 2°C min^{-1}. Data analysis was performed using *JTSA* (http://paulsbond.co.uk/jtsa; Schulz *et al.*, 2013). The average fluorescence was plotted against temperature and fitted to a sigmoid 5 function at each pH value. The melting temperature was estimated from the midpoint of the transition.

2.8. Crystallization

2.8.1. Crystallization of recombinant GBA. Purified GBA (10 mg ml^{-1}) was tested against a range of commercial crystallization screens. An initial hit was found in well H8 of the PACT *premier* HT-96 screen from Molecular Dimensions (Newman *et al.*, 2005) with conditions consisting of 0.2 *M* sodium sulfate, 20%(*w*/*v*) PEG 3350, 0.1 *M* bis-Tris propane pH 8.5. Optimization of the PEG 3350 concentration and the buffer pH was performed in a 48-well MRC sitting-drop vapour-diffusion format to yield thin, rod-like crystals at pH 7 and 7.5. Further optimization of the PEG 3350 and protein concentrations resulted in larger crystals with the same morphology. The final optimized conditions were 0.3 µl GBA (10 mg ml^{-1}) plus 0.5 µl well solution [0.2 *M* sodium sulfate, 14%(*v*/*v*) PEG 3350, 0.1 *M* bis-Tris propane pH 7.0].

2.8.2. Sequential seeding to avoid the presence of bis-Tris propane in the active site. As bis-Tris propane and related compounds are glycosidase inhibitors (for a review, see Roberts & Davies, 2012) that would interfere with soaking experiments, seeding was used to obtain crystals in non-bis-Tris propane conditions. Crystals obtained in the presence of bis-Tris propane were used to generate a concentrated seed stock according to previously published protocols (Shaw Stewart *et al.*, 2011). In a 48-well MRC sitting-drop vapour-diffusion format, dilutions of the concentrated seed stock (1:100 and 1:1000) were used to screen into PACT *premier* HT-96 well H8 conditions in which the bis-Tris propane had been substituted with HEPES buffer pH 7 and 7.5. Through optimization of the PEG 3350 concentration, HEPES concentration, protein volume and seeding ratios, crystals suitable for generating new seed stocks were obtained using 0.2 µl GBA (10 mg ml^{-1}) plus 0.2 µl well solution [0.2 *M* sodium sulfate, 14%(*v*/*v*) PEG 3350, 0.25 *M* HEPES pH 7.0] plus 50 nl seed solution (1:1000 dilution). These seed stocks were used to rescreen previous HEPES-containing conditions in a 48-well MRC sitting-drop format, resulting in crystals that were suitable for analysis under the conditions 0.2 µl GBA (10 mg ml^{-1}) plus 0.4 µl well solution [0.2 *M* sodium sulfate, 14%(*v*/*v*) PEG 3350, 0.25 *M* HEPES pH 7] plus 0.1 µl seed solution (1:1000 dilution).

2.8.3. Complex with 2-deoxy-2-fluoro-β-D-glucopyranoside. 2,4-Dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside (2F-DNPGlc) was prepared using well established literature protocols. Briefly, 3,4,6-tri-*O*-acetyl glucal was fluorinated using Selectfluor in acetonitrile/H_2O (Burkart *et al.*, 1997) and was then coupled with 1-fluoro-2,4-dinitrophenylbenzene, separated and deprotected according to published procedures (Namchuk *et al.*, 2000). Instead of peracetylation, column purification, crystallization and anomeric deacetylation to give 3,4,6-Ac-2F-Glc, DNP coupling on the Selectfluor reaction mixture was performed to yield a mixture of 3,4,6-Ac-DNP-α/β-2F-glucoside/mannoside, which were separated using 2:9:9 ethyl acetate:dichloromethane:hexanes as the mobile phase and recrystallization.

Crystals generated in HEPES conditions following multiple rounds of seeding were soaked with 2F-DNPGlc (synthesized as above) overnight. The crystals were cryoprotected by soaking in well solution supplemented with 25% ethylene glycol prior to flash-cooling in liquid nitrogen for data collection.

2.8.4. Crystallization of unliganded GBA. An initial hit was also identified in well A5 of the JCSG-*plus* screen from Molecular Dimensions (Newman *et al.*, 2005) with conditions consisting of 0.2 *M* magnesium formate, 20%(*w*/*v*) PEG 3350. Optimization of the magnesium formate, PEG 3350 and protein concentrations resulted in larger crystals with the same morphology. The final optimized conditions were 0.6 µl GBA (10 mg ml^{-1}) plus 0.5 µl well solution [0.2 *M* magnesium formate, 19%(*v*/*v*) PEG 3350].

2.8.5. Cryoprotection. All crystals were cryoprotected with well solution supplemented with 25%(*v*/*v*) ethylene glycol prior to flash-cooling in liquid nitrogen for data collection.

2.8.6. Crystallization of Cerezyme. Prior to crystallization, Cerezyme (a generous gift from Professor Hans Aerts, Leiden) was deglycosylated with PNGase F (20 µl; New England Biolabs) for five days at room temperature. The digested material was purified by size-exclusion chromatography on a Superdex 75 16/600 column. Crystals were obtained using hanging-drop vapour diffusion, based on conditions outlined by Dvir *et al.* (2003). The drops consisted of 1 µl Cerezyme (9.1 mg ml^{-1}) and 1 µl well solution consisting of 1.1 *M* ammonium sulfate, 0.19 *M* guanidine–HCl,

Table 2
Crystallization.

	PDB entry 6tjk	PDB entry 6tjj	PDB entry 6tjq	PDB entry 6tn1
Method	Sitting-drop vapour diffusion	Hanging-drop vapour diffusion	Sitting-drop vapour diffusion	Sitting-drop vapour diffusion
Plate type	MRC Maxi 48-well	24-well XRL	MRC Maxi 48-well	MRC Maxi 48-well
Temperature (K)	293	293	293	293
Protein concentration (mg ml^{-1})	10	10	10	10
Buffer composition of protein solution	10 m*M* MES, 100 m*M* NaCl, 1 m*M* TCEP pH 6.5	20 m*M* MES, 100 m*M* NaCl pH 6.5	10 m*M* MES, 150 m*M* NaCl, 1 m*M* TCEP pH 6.5	10 m*M* MES, 100 m*M* NaCl, 1 m*M* TCEP pH 6.5
Composition of reservoir solution	0.2 *M* Na_2SO_4, 14%(*v*/*v*) PEG 3350, 0.1 *M* bis-Tris propane pH 7.0	1.1 *M* $(NH_4)_2SO_4$, 0.19 *M* guanidine–HCl, 0.04 *M* KCl, 0.1 *M* sodium acetate pH 4.6	0.2 *M* Na_2SO_4, 14%(*v*/*v*) PEG 3350, 0.25 *M* HEPES pH 7.0	0.2 *M* magnesium formate, 19%(*v*/*v*) PEG 3350
Volume and ratio of drop (protein:well) (µl)	0.8 (0.3:0.5)	2 (1:1)	0.7 (0.2:0.4, + 0.1 seed)	1.1 (0.6:0.5)
Volume of reservoir (µl)	100	500	100	100

Table 3
Data collection and processing.

Values in parentheses are for the outer resolution shell.

	PDB entry 6tjk	PDB entry 6tjj	PDB entry 6tjq	PDB entry 6tn1
Diffraction source	Beamline I04, DLS	Beamline I03, DLS	Beamline I04, DLS	Beamline I04-1, DLS
Wavelength (Å)	0.9795	0.9763	0.9795	0.9159
Temperature (K)	100	100	100	100
Detector	EIGER2 XE 16M	EIGER2 XE 16M	EIGER2 XE 16M	PILATUS 6M-F
Rotation range per image (°)	0.1	0.1	0.1	0.1
Total rotation range (°)	360	360	360	360
Exposure time per image (s)	0.020	0.020	0.010	0.040
Space group	$P2_1$	$C222_1$	$P2_1$	$P1$
a, *b*, *c* (Å)	52.7, 156.2, 68.3	110.1, 285.9, 91.9	53.1, 76.4, 68.2	44.5, 46.2, 64.2
α, β, γ (°)	90, 102, 90	90, 90, 90	90, 102, 90	86, 75, 83
Resolution range (Å)	66.75–1.56 (1.59–1.56)	77.44–1.59 (1.62–1.59)	52.01–1.41 (1.43–1.41)	31.70–0.98 (1.00–0.98)
Total No. of reflections	637325 (31169)	1531796 (74996)	689440 (33887)	394040 (1815)
No. of unique reflections	147746 (7162)	193710 (9517)	102363 (5031)	208682 (1518)
Completeness (%)	96.8 (94.7)	100 (100)	100 (99.7)	74.2 (10.9)†
Multiplicity	4.3 (4.4)	7.9 (7.9)	6.7 (6.7)	1.9 (1.2)
$\langle I/\sigma(I)\rangle$	7.3 (0.7)	7.2 (0.5)	7.2 (0.4)	8.7 (1.0)
$R_{p.i.m.}$	0.057 (0.90)	0.055 (2.34)	0.056 (1.66)	0.037 (0.49)
$CC_{1/2}$	0.99 (0.34)‡	0.99 (0.68)	0.99 (0.36)‡	0.99 (0.69)
Overall *B* factor from Wilson plot (Å^2)	18	24	18	7

† The low completeness for the outer resolution shell (also reflected in the overall statistics) for PDB entry 6tn1 reflected integration into the corners of a square detector. ‡ Data with a low outer bin $I/\sigma(I)$ (with a $CC_{1/2}$ of approximately 0.35) were used as they were reflected by improved maps with appropriate model-refinement statistics at these resolutions.

0.04 *M* KCl, 0.1 *M* sodium acetate pH 4.6. Crystals were transferred to a lithium sulfate cryoprotectant (0.2 *M* lithium sulfate, 0.17 *M* guanidine–HCl, 0.04 *M* KCl, 0.1 *M* sodium acetate pH 4.6) before flash-cooling in liquid nitrogen.

Crystallization information is summarized in Table 2.

2.9. Data collection, structure solution and refinement

Data for the bis-Tris propane complex were collected on the I04 beamline at the Diamond Light Source (DLS) and were integrated using the *DIALS* pipeline in *xia*2 (Winter, 2010; Winter *et al.*, 2018). Data reduction was performed using *AIMLESS* (Evans, 2006; Evans & Murshudov, 2013) from the *CCP*4 suite (Winn *et al.*, 2011) and the data were processed to a resolution of 1.56 Å. Molecular replacement using a previously obtained structure of Cerezyme (Artola *et al.*, 2019) as the search model was conducted with *Phaser* (McCoy *et al.*, 2007).

Data for the 2F-Glc complex were collected on the I04 beamline at DLS and were integrated using the *DIALS* pipeline in *xia*2. Data reduction was performed in *AIMLESS* and the data were processed to a resolution of 1.41 Å. The structure was solved by molecular replacement with *MOLREP* (Vagin & Teplyakov, 2010) using the bis-Tris propane complex structure obtained in this work as the search model.

Data for the unliganded crystal were collected on the I04-1 beamline at DLS and were integrated using the *autoPROC* pipeline (Vonrhein *et al.*, 2011). Data reduction was performed in *AIMLESS* and the data were processed to a resolution of 0.98 Å. Molecular replacement using PDB entry 2nt1 (Lieberman *et al.*, 2007) as the search model was conducted using *MOLREP*.

Data for the Cerezyme crystal were collected on the I03 beamline at DLS to 1.71 Å resolution and were integrated using the *DIALS* pipeline in *xia*2. The structure was solved by molecular replacement with *MOLREP* using PDB entry 2nt0 (Lieberman *et al.*, 2007) as the search model.

Data-collection and processing statistics are summarized in Table 3.

All structures were refined using *REFMAC* (Murshudov *et al.*, 2011) followed by multiple rounds of manual model

Table 4
Structure solution and refinement.

Values in parentheses are for the outer resolution shell.

	PDB entry 6tjk	PDB entry 6tjj	PDB entry 6tjq	PDB entry 6tn1
Resolution range (Å)	66.75–1.56 (1.59–1.56)	77.44–1.59 (1.62–1.59)	52.01–1.41 (1.43–1.41)	31.70–0.98 (1.00–0.98)
Completeness (%)	96.8 (94.7)	100 (100)	100 (99.7)	74.2 (10.9)
Final R_{cryst}	0.17	0.22	0.18	0.11
Final R_{free}	0.20	0.25	0.21	0.13
Cruickshank DPI (Å)	0.08	0.09	0.06	0.02
No. of non-H atoms				
Protein	7987	8008	4065	4486
Ion	—	2	—	8
Ligand	418	242	173	167
Water	907	721	484	709
Total	9312	8973	4722	5370
R.m.s. deviations				
Bonds (Å)	0.014	0.014	0.013	0.009
Angles (°)	1.80	1.84	1.71	1.61
Average B factors (Å^2)				
Protein	23	34	21	8
Ion	—	42	—	16
Ligand	39	67	42	15
Water	35	42	35	29
Ramachandran plot				
Most favoured (%)	95	95	95	95
Allowed (%)	4	4	4	4

building with *Coot* (Emsley *et al.*, 2010). The 0.98 Å resolution unliganded structure was anisotropically refined with multiple TLS refinement cycles using the automatic *REFMAC* option. Idealized coordinate sets and refinement dictionaries for ligands were generated using *AceDRG* (Long *et al.*, 2017*a*,*b*) or *JLigand* (Lebedev *et al.*, 2012). Sugar conformations were validated using *Privateer* (Agirre *et al.*, 2015), and *MolProbity* (Chen *et al.*, 2010) was used to assess model validity before deposition in the PDB. Refinement statistics are summarized in Table 4.

3. Results and discussion

3.1. Recombinant protein production and purification

In an attempt to circumvent the need for ERT sources of GBA for structural and biochemical studies, we sought to establish an in-house BEVS expression system for GBA. We generated a construct in which an N-terminally truncated GBA gene, lacking its native signal sequence, was used in conjunction with a honeybee melittin secretion signal (Soejima *et al.*, 2013; Tessier *et al.*, 1991) to promote the secretion of recombinant GBA into the medium. High Five cells were used for the production of recombinant GBA because they have been reported to be more efficient than Sf9 cells in secreting certain proteins (however, this effect will be protein-dependent; Wilde *et al.*, 2014).

A recombinant bacmid encoding the N-terminally truncated GBA gene was produced using the established Tn7 transposition method in DH10EMBacY cells (Bieniossek *et al.*, 2012; Fitzgerald *et al.*, 2006; Geneva Biotech; Table 1). Sf9 cells were transfected with the recombinant bacmid to generate recombinant baculovirus encoding the human GBA gene.

Table 5
Kinetic analysis of recombinant GBA and comparison with Cerezyme and with GBA produced in insect cells by Sawkar *et al.* (2006).

	Recombinant GBA†	Cerezyme‡	rhWT-GBA§
K_m (mM)	1.288 ± 0.051	1.127 ± 0.052	—
V_{max} (μM min^{-1})	11.74 ± 0.23	2.21 ± 0.03	—
k_{cat} (min^{-1})	1174 ± 23	1325 ± 2	868 ± 28

† Values for recombinant GBA produced in this study. The data shown are the average ± standard deviation of four replicates. ‡ Values reported for Cerezyme (Tekoah *et al.*, 2013). § k_{cat} value for GBA produced in insect cells by Sawkar *et al.* (2006) determined by conversion of the reported specific activity.

Functional GBA was subsequently produced in High Five cells by infection with the recombinant baculovirus, resulting in secretion of the recombinant product into the cell medium. GBA was purified from the cell medium in the presence of Tween 80 detergent according to a previously outlined procedure (Sawkar *et al.*, 2006), followed by a size-exclusion step to remove the detergent and yield pure protein suitable for X-ray crystallography (Fig. 3).

Following purification, a typical yield of 3.6–4.6 mg l^{-1} was achieved, generating 13.0–16.7 mg of protein per expression. This production protocol generates sufficient purified protein for both biochemical and structural studies. Unfortunately, the study from which the purification procedure was taken failed to report a yield for comparison (Sawkar *et al.*, 2006), and only estimated yields have been provided in the very few studies in which GBA has been purified (Sinclair *et al.*, 2006). Thus, we are unable to comment on the expression yield of our GBA expression system relative to previous studies.

3.2. Biochemical characterization

The biophysical properties of our GBA were investigated to evaluate whether this recombinant product could be a viable alternative to ERT formulations for nonclinical academic use. The kinetics of our GBA were assayed using the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside (4-MU-Glc) and the initial reaction rates were fitted to the Michaelis–Menten equation (Fig. 4*a*). Our recombinant enzyme exhibited comparable K_m, V_{max} and k_{cat} values to those reported for Cerezyme (Tekoah *et al.*, 2013; Table 5), suggesting that our GBA produced in insect cells exhibits similar kinetic properties to Cerezyme produced in CHO cells. Furthermore, the k_{cat} of this recombinant enzyme compares favourably with that of GBA produced in insect cells by Sawkar *et al.* (2006) (Table 5), although no K_m or V_{max} values were reported in that study.

The pH-dependent stability of the recombinant protein was evaluated through a thermal shift assay at pH 5.2 and pH 7.0 (Fig. 4*b*). Our GBA exhibited optimum stability at pH 5.2 (T_m = 60°C), as expected for an enzyme which operates in the acidic environment of the lysosome. This T_m compares favourably with that of Cerezyme (Ben Bdira *et al.*, 2017) and is 10°C higher than that of GBA produced in insect cells by Sawkar *et al.* (2006) (Table 6). The recombinant GBA demonstrated pH-dependent thermal stability, exhibiting a 6.2°C decrease in T_m at pH 7.0 compared with pH 5.2. This is

consistent with the 4°C decrease that has been reported for Cerezyme (Ben Bdira *et al.*, 2017; Table 6). A decrease in thermal stability at neutral pH has been supported by proteolysis studies, in which GBA was found to be resistant to tryptic digestion at pH 5.2 but not at pH 7.4 (Ben Bdira *et al.*, 2017). This behaviour is thought to arise from changes in the native fold of the enzyme at neutral pH. Therefore, our recombinant protein exhibits the pH-dependent thermal stability profile expected for GBA. In contrast, pH-dependent thermal stability was not reported for the GBA formulation produced by Sawkar *et al.* (2006).

3.3. Crystallization and structure solution

3.3.1. Structure of GBA in complex with bis-Tris propane. Crystals of recombinant GBA were initially obtained in well H8 [0.2 *M* sodium sulfate, 20%(*w*/*v*) PEG 3350, 0.1 *M* bis-Tris propane] of the PACT screen (Newman *et al.*, 2005) at pH 8.5. Further optimization of the buffer pH, the precipitant concentration and the protein concentration generated crystals at pH 7.0 (Figs. 5*a* and 5*b*), at which GBA is more active. Using 0.2 *M* sodium sulfate, 14%(*v*/*v*) PEG 3500 and 0.1 *M* bis-Tris propane (BTP) pH 7.0, GBA crystallized in space group $P2_1$, with two molecules in the asymmetric unit, and the crystals diffracted to give a 1.56 Å resolution data set (PDB entry 6tjk; Fig. 5*c*). Unlike in some earlier studies on GBA, we did not deglycosylate the GBA prior to crystal screening; consequently, the resulting structure exhibited visible N-glycosylation at Asn19, Asn59 and Asn146 in chain *A* ,and at Asn19 and Asn146 in chain *B*. In contrast, only glycosylation at the Asn19 site had been modelled in a number of previous studies (Dvir *et al.*, 2003; Liou *et al.*, 2006; Lieberman *et al.*, 2009). Occupancy of the Asn19 N-glycosylation site is known to be vital for GBA activity (Berg-Fussman *et al.*, 1993;

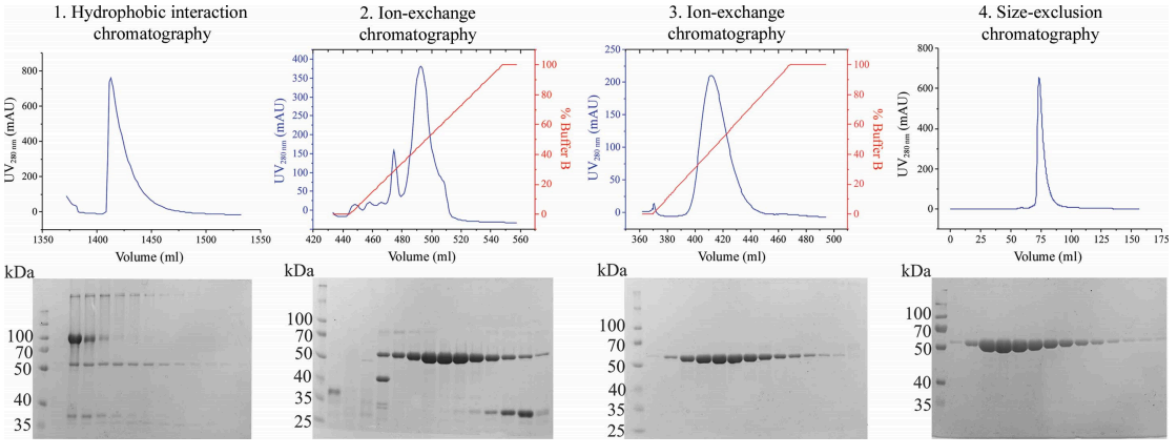


Figure 3
Purification of recombinant GBA from cell-culture medium with purification chromatograms and SDS–PAGE analyses for each purification step. GBA (~55 kDa) was extracted from the medium by hydrophobic interaction chromatography followed by two rounds of cation-exchange chromatography with the addition of Tween 80 detergent. Purification was completed by a size-exclusion step to remove Tween 80.

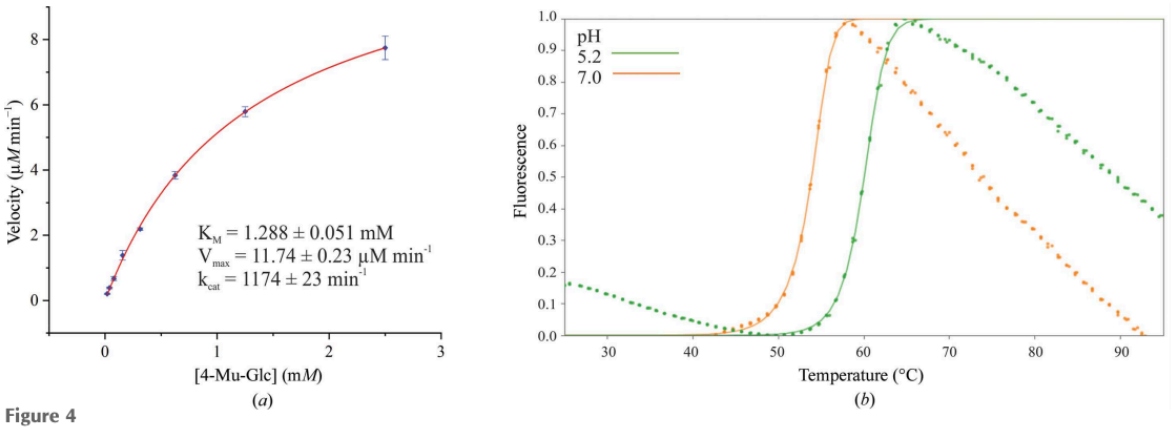


Figure 4
(*a*) Michaelis–Menten kinetic assay of GBA using the fluorogenic substrate 4-MU-Glc. Data are plotted as the average ± standard deviation of four replicates. (*b*) Heat-induced melting profile of GBA at pH 5.2 and pH 7.0 recorded by thermal shift assay using SYPRO Orange dye.

Table 6
T_m values for the thermal denaturation of recombinant GBA, Cerezyme (Lieberman *et al.*, 2009) and rhWT-GBA produced in insect cells by Sawkar *et al.* (2006).

Protein	T_m (°C)	
Recombinant GBA†	60.0 ± 0.2 (pH 5.2)	53.8 ± 0.1 (pH 7.0)
rhWT-GBA‡	49.3 (pH 5.3)	49.2 (pH 7.0)
Cerezyme§	61 (pH 5.2)	57 (pH 7.4)

† T_m values of recombinant GBA produced in this work as determined by a Thermofluor assay. Data are reported as the average ± standard deviation of three replicates. ‡ T_m values of GBA produced in insect cells by Sawkar *et al.* (2006) as determined by circular dichroism (Sawkar *et al.*, 2006). § T_m values reported for Cerezyme as determined by circular dichroism (Ben Bdira *et al.*, 2017).

Grace & Grabowski, 1990), and in this structure the Asn19 site is occupied by a chitobiose core with a β-1,4-mannose unit in chain *B* and an additional α-1,3-mannose unit in chain *A*. Only single N-linked GlcNAc residues could be modelled at the other N-glycosylation sites.

A crystal structure of Cerezyme in space group $C222_1$ was obtained at 1.59 Å resolution (PDB entry 6tjj) for comparison with the structure of our GBA in complex with BTP. The tertiary structure of our recombinant enzyme is similar to that of Cerezyme, exhibiting the same three domains as observed in previous studies. Domain I spans residues 1–27 and 383–414, forming an antiparallel β-sheet, domain II consists of residues 30–75 and 431–497, which form an immunoglobulin-like fold, and domain III comprises a $(\beta/\alpha)_8$ TIM barrel formed by residues 76–381 and 416–430 (Fig. 5*d*). There is one amino-acid change at residue 495, which is an arginine in our recombinant GBA sequence but is a histidine in Cerezyme. This is a known mutation in Cerezyme, which deviates from human placental GBA (Wei *et al.*, 2011). Overall, the tertiary structure of the recombinant GBA produced in this work compares well with the unliganded structure of Cerezyme (Fig. 5*e*), with C^α root-mean-square deviations (r.m.s.d.s) of 0.57 Å (*Q*-score of 0.95) and 0.50 Å (*Q*-score of 0.96) for overlay of the *A* chains and *B* chains, respectively. However, some deviations in the protein backbone were observed in the flexible loop regions of residues 27–30, 60–64, 319–313 and 395–398. In previously published crystal structures of GBA, the loops formed by residues 311–319 and 394–399 are present in multiple conformations, suggesting dynamic flexibility of these regions. Nevertheless, this structure is also comparable with the deposited structure of Cerezyme obtained at pH 7.5 (PDB entry 2nt1; Lieberman *et al.*, 2007), with a C^α r.m.s.d. of $\leq$0.61 Å (*Q*-score of $\geq$0.94) for all chains.

Unfortunately, a true ligand-free structure was not obtained owing to the binding of BTP from the crystallization conditions in the active site (Fig. 5*d*). The binding of BTP to glycosidases has been observed previously (Thompson *et al.*, 2012; Roberts & Davies, 2012; Brunzelle *et al.*, 2008) and results from a superficial similarity between the hydroxylated

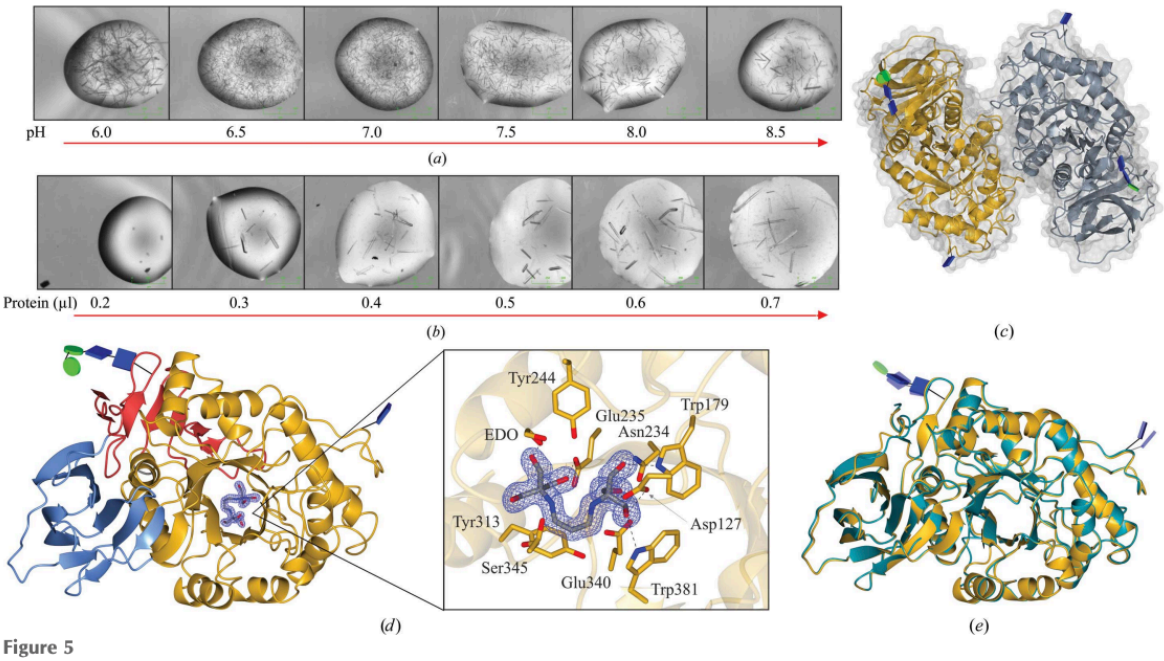


Figure 5
(*a*) Optimization of the crystallization pH using bis-Tris propane buffer. (*b*) Optimization of the protein concentration. (*c*) Crystal structure of the GBA dimer obtained at 1.56 Å resolution (PDB entry 6tjk). N-Glycans are depicted in glycoblock format (McNicholas & Agirre, 2017). (*d*) GBA monomer comprising of three domains: domain I (residues 1–27 and 383–414) is shown in blue, domain II (residues 30–75 and 431–497) in red and domain III (residues 76–381 and 416–430) in gold. The active site contains bound bis-Tris propane, which forms hydrogen bonds to Trp179, Asn234, Glu235, Glu340, Trp381 and an ethylene glycol (EDO) cryoprotectant molecule. Electron density is contoured to 1σ (0.34 e Å^{-3}). (*e*) Overlay of recombinant GBA (gold) obtained at pH 7.0 and Cerezyme (teal) obtained at pH 4.6 (PDB entry 6tjj).

and positively charged BTP molecule and the oxocarbenium-ion transition state of glycoside hydrolysis, which is strongly stabilized by glycosidase enzymes. Although the active site is occupied by BTP, when aligned with the active site of Cerezyme (Fig. 6) it is clear that most active-site residues, including the catalytic Glu235 (acid/base) and Glu340 (nucleophile), adopt almost identical conformations. However, Tyr313 is displaced downwards in the BTP complex, presumably to avoid clashing with the hydroxyl groups of the BTP molecule.

Attempts to use these crystals for ligand-binding studies by soaking with other ligands to displace BTP were unsuccessful. The inability to displace BTP from the active site can be rationalized by the high concentration of BTP used in the crystallization conditions (100 m*M*) and its comparatively potent IC_{50} (IC_{50} = 4.31 ± 0.42 m*M*) against 4-MU-Glc (Appendix *A*). Consequently, crystals obtained under BTP-containing conditions were used for microseeding into conditions in which BTP was substituted with HEPES buffer pH 7.0. Following multiple rounds of seeding and optimization, crystals suitable for structural studies were generated using 0.2 *M* sodium sulfate, 14%(*v*/*v*) PEG 3350, 0.25 *M* HEPES pH 7.0 and 0.1 µl seed solution (1:1000 dilution).

3.3.2. Trapped covalent intermediate structure. To demonstrate the potential of these GBA crystals for use in structural studies, optimized BTP-free crystals were soaked with 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside (2F-DNPGlc) to generate a novel GBA complex structure. 2F-DNPGlc is a well characterized β-glucosidase inhibitor in which substitution of the C2 hydroxyl group with an electronegative F atom destabilizes the oxocarbenium-ion transition states for both enzyme active-site glycosylation and deglycosylation (Street *et al.*, 1992; Withers *et al.*, 1988; Fig. 1). However, the addition of a reactive DNP leaving group to the aglycone increases the rate of glycosylation, allowing a trapped enzyme–inhibitor complex to accumulate after reaction with the enzyme (Withers *et al.*, 1988). Activated 2-deoxy-2-fluoroglycosides have been used in combination with X-ray crystallography to gain mechanistic insights into retaining glycosidases, with 2-deoxy-2-fluoro-β-D-glucopyranosyl fluoride notably having been used to correct the identity of the catalytic nucleophile of GBA (Miao *et al.*, 1994), but no co-crystal complex of GBA with this inhibitor has been previously reported.

A structure of the 2-deoxy-2-fluoroglucopyranosyl-GBA intermediate was obtained at 1.41 Å resolution (PDB entry 6tjq), showing unambiguous electron density for covalent binding of the 2-deoxy-2-fluoroglucose moiety to the catalytic nucleophile, with a covalent bond length of 1.42 Å (Fig. 7*a*). The glucose-configured ring adopts a $^4C_1$ chair conformation, consistent with the conformation of the covalent glycosyl-enzyme intermediate in the GBA conformational itinerary (Fig. 1). The bound 2F-glycone moiety also forms hydrogen bonds to Trp179, Asp127, Trp179, Asn234, Glu340, Trp381, Asn396 and an ethylene glycol cryoprotectant molecule. Interestingly, two conformations of the catalytic nucleophile can be observed (Fig. 7*a*). We postulate that electrostatic repulsion between the carboxylate of the catalytic nucleophile and the C2-linked F atom of the 2F-Glc inactivator enforces a 28° rotation about C^{γ} of the nucleophilic residue, resulting in movement of the O1 atom of the carboxylate residue away from the C2-linked F atom. Aside from providing a novel

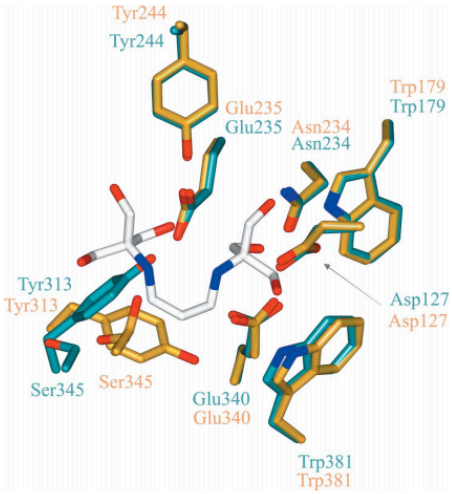


Figure 6
(*a*) Active-site overlay of Cerezyme (teal) and recombinant GBA (gold) with bis-Tris propane occupying the active site (white).

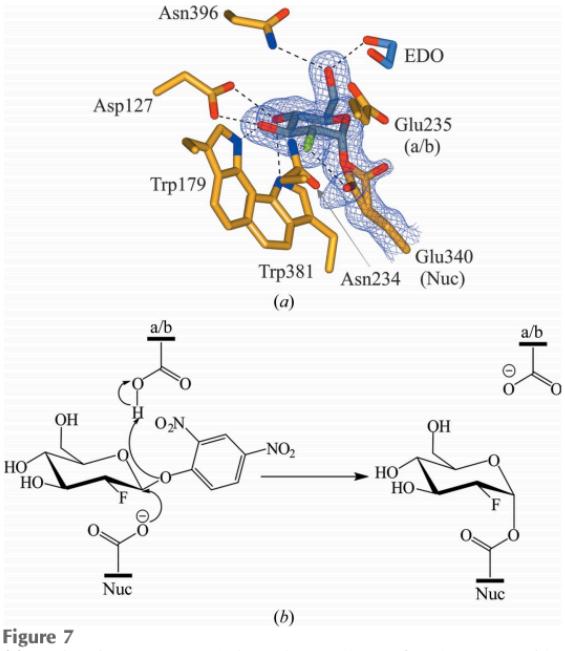


Figure 7
(*a*) Active-site structure of the 2-deoxy-2-fluoro-β-D-glucopyranoside-GBA covalent intermediate (PDB entry 6tjq). The 2F-Glc moiety is covalently bound to the catalytic nucleophile (Glu340), which occupies two conformations. a/b, catalytic acid/base; Nuc, catalytic nucleophile; EDO, ethylene glycol. Electron density is contoured to 1.1σ (0.40 e Å^{-3}). (*c*) Mechanism of the hydrolysis of 2F-DNPGlc by GBA to generate the covalent glycosyl-enzyme intermediate

structure in complex with a mechanistically relevant glucosidase inhibitor, this complex demonstrates the ability of our GBA to be used as an alternative to ERT preparations in the structure-based development of new inhibitory compounds for GBA.

3.3.3. Atomic resolution ligand-free structure. Given the tight binding of BTP in our originally identified GBA crystal form, we also sought to identify non-BTP-containing crystallization conditions in parallel with our efforts to remove BTP by microseeding. During initial screening, crystals were also found under condition A5 [0.2 *M* magnesium formate, 20%(*w*/*v*) PEG 3350] of the JCSG-*plus* screen (Newman *et al.*, 2005). Optimized crystals suitable for structural analysis were obtained using 0.2 *M* magnesium formate, 19%(*v*/*v*) PEG 3350. Subsequently, a 0.98 Å resolution unliganded structure of GBA was obtained (PDB entry 6tn1). Not only is this the highest resolution structure of GBA deposited to date, it also exists in a previously unreported crystal form. Previously, GBA has been crystallized in space groups $C222_1$ and $P2_1$; however, this unliganded structure crystallized in space group $P1$. The new structure contains one molecule in the asymmetric unit, which comprises three noncontiguous domains, with N-glycosylation at Asn19 and Asn146 (Fig. 8*a*). Overall, the three-domain tertiary structure is highly similar to that of the BTP complex and Cerezyme, with C^{α} r.m.s.d.s of 0.49 Å (*Q* score of 0.94) and 0.60 Å (*Q* score of 0.94), respectively. However, some deviations in the protein backbone were observed in the flexible loop regions consisting of residues 26–31, 314–319 and 344–350 (Fig. 8*b*). Despite the sub-Ångström resolution, residues 26–31 and 314–319 were challenging to model, reflecting the flexibility and disorder of these loops, which has also been observed in previous GBA structures. Importantly, the active site of this unliganded structure compares well with the active sites of Cerezyme and the BTP complex. The majority of active-site residues occupy essentially identical conformations, with the exception of Tyr313, which restores its 'upwards' conformation in the absence of BTP (Fig. 8*c*). In fact, Tyr313 appears to be particularly mobile, occupying a different conformation in each GBA structure.

The sub-Ångström resolution of this unliganded structure permits the first ever atomic resolution analysis of GBA, uncovering finer details in its structure. For example, two conformations of the catalytic acid/base residue (Glu235) can be observed (Fig. 9*a*). In fact, many alternative side-chain conformations could be modelled throughout the structure, providing more detail on side-chain mobility and interactions. We also observed proton positions for some residues in the difference electron density, as well as electron delocalization over carbonyls and double bonds (Figs. 9*b*–9*f*). We anticipate

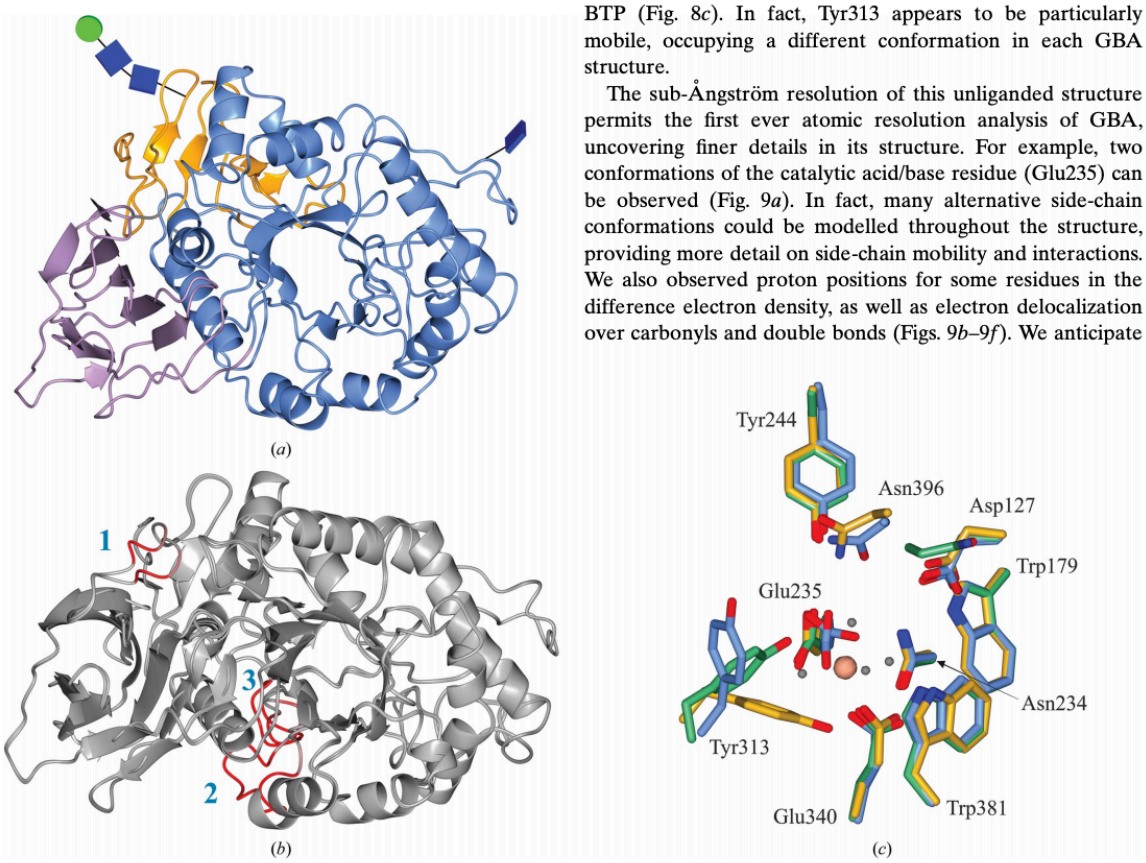


Figure 8
(*a*) Crystal structure of the GBA monomer obtained at 0.98 Å resolution (PDB entry 6tn1). Domain I (residues 1–27 and 383–414) is shown in lilac, domain II (residues 30–75 and 431–497) in orange and domain III (residues 76–381 and 416–430) in blue. N-Glycans are depicted in glycoblock format (McNicholas & Agirre, 2017). (*b*) Overlay of the unliganded GBA structure with the BTP-complexed structure (PDB entry 6tjk). Red indicates areas of high r.m.s.d. between the protein backbones. Loop 1 contains residues 27–31, loop 2 comprises residues 314–319 and loop 3 contains residues 344–350. (*c*) Active site of the unliganded GBA crystal structure (blue) overlaid with active-site residues of the BTP-complex structure (gold; PDB entry 6tjk) and Cerezyme (green; PDB entry 6tjj). A magnesium ion (peach) coordinated by four waters (grey), Glu340 (Nuc) and Glu235 (a/b) occupies the active site.

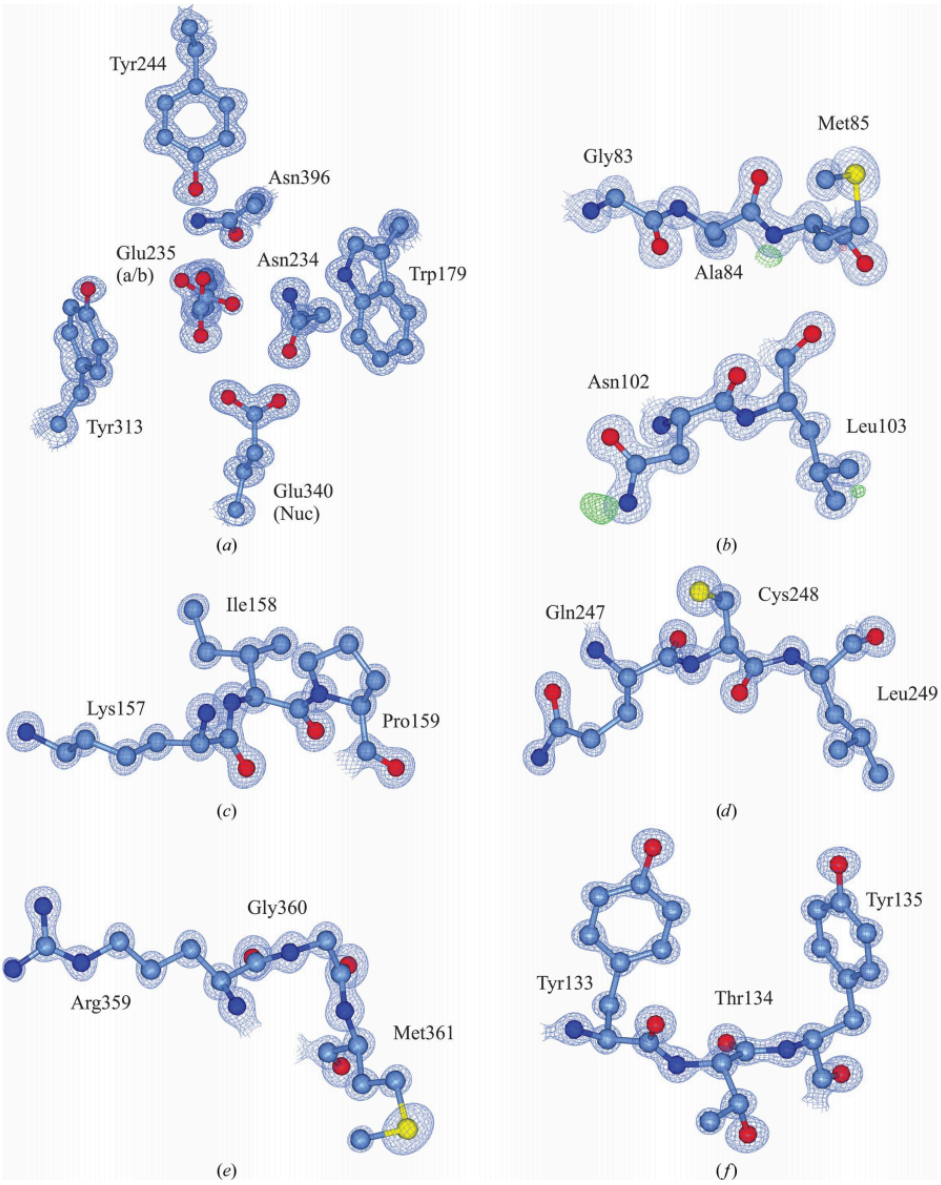


Figure 9
(*a*) Electron density for active-site residues, including the catalytic nucleophile (Nuc) Glu340 and catalytic acid/base (a/b) Glu235, contoured to 2σ (1.0 e Å^{-3}). (*b*) Selection of modelled residues with difference electron density [green; contoured to 3σ (0.37 e Å^{-3})] highlighting proton positions. (*c*)–(*f*) Modelled residues from domain I of GBA (PDB entry 6tn1) demonstrating atomic resolution (electron density contoured to 3.5σ (1.75 e Å^{-3}).

that this new crystal form will be utilized in structural studies to provide atomic resolution analysis of ligand binding and interactions with GBA.

4. Conclusions

This work describes a detailed approach to the production of active human GBA in an AcNPV-derived baculoviral expression system, providing an alternative source to ERT formulations. Recombinant GBA was produced and secreted from baculovirus-transduced insect cells by the action of the honeybee melittin signal sequence, and was purified with a typical yield of 3.6–4.6 mg l^{-1}. The recombinant protein was shown to be active against the artificial substrate 4-methylumbelliferyl-β-D-glucopyranoside (K_m = 1.3 mM, k_{cat} = 1174 min^{-1}) and exhibited optimum thermal stability at pH 5.2 (T_m = 60°C), consistent with the biophysical properties of the commercial ERT Cerezyme. Moreover, our recombinant

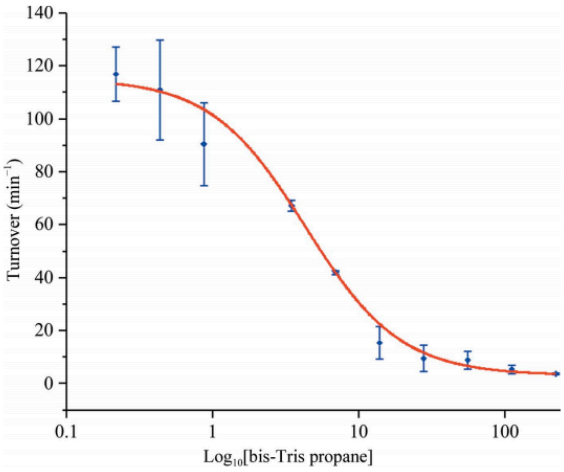


Figure 10
Rate of hydrolysis of 4-MU-Glc at pH 7.0 versus bis-Tris propane concentration fitted to a four-parameter logistic function to estimate the IC_{50} of bis-Tris propane.

GBA crystallizes readily and is amenable to structural ligand binding, as demonstrated by a novel structure in complex with the glucosidase inhibitor 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-D-glucopyranoside. We also identified a novel crystal form of GBA which diffracts to 0.98 Å resolution. This is the highest resolution structure of GBA deposited to date, permitting exquisite atomic resolution analysis. Multiple alternative residue conformations were observed throughout the structure, including two conformations of the catalytic acid/base residue. We envision that the BEVS GBA production system described in this work will alleviate the over-reliance on ERT formulations, aid in future biochemical studies of GBA and support structure-guided development of novel GBA ligands.

APPENDIX *A*
The IC_{50} of bis-Tris propane

The IC_{50} of bis-Tris propane was determined using the fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside (4-MU-Glc). GBA was prepared at 58 n*M* in kinetics buffer [McIlvaine buffer; 150 m*M* disodium hydrogen phosphate, citric acid pH 7.0 supplemented with 0.2%(*v*/*v*) taurocholate, 0.1%(*v*/*v*) Triton X-100 and 0.1%(*v*/*v*) bovine serum albumin (BSA)]. Bis-Tris propane was prepared at 500 m*M* in kinetics buffer pH 7.0 and diluted twofold in serial dilutions to 0.488 µ*M*. The solutions (20 µl) were added to the wells of a black 384-well polystyrene plate followed by the addition of 4-MU-Glc (20 µl) prepared at 4 m*M* in kinetics buffer. GBA (5 µl) was added to each well to yield a final enzyme concentration of 6.5 n*M*. The activity against 4-MU-Glc was monitored continuously over 5 min at room temperature by measuring the fluorescence of liberated 4-MU (λ_{ex} = 360–320 nm, λ_{em} = 450–430 nm) using a CLARIOstar Plus microplate reader (BMG Labtech). The assays were performed in quadruplicate for each bis-Tris propane concentration. A linear calibration was generated by measuring the fluorescence of the 4-MU product (λ_{ex} = 360–320 nm, λ_{em} = 450–430 nm) prepared at serial dilutions of 62.5, 31.25, 15.63, 7.81, 3.91 and 1.95 µ*M* in kinetics buffer. The data were processed using the *Origin* graphing software. Using the 4-MU calibration, the rate of substrate hydrolysis (*V*) was determined at each concentration of bis-Tris propane. The rates (*V*) were plotted against the log of the bis-Tris propane concentration and were fitted by nonlinear regression to the four-parameter logistic function to determine the IC_{50} (IC_{50} = 4.31 ± 0.42 m*M*; Fig. 10).

Acknowledgements

We thank the Berger Laboratory, University of Bristol for generously providing the p-OMNI and DH10EMBacY vectors for the production of recombinant baculovirus. We thank Diamond Light Source for access to beamlines I03, I04 and I04-1, which contributed to the results presented here (proposal No. mx-18598), and Johan Turkenburg and Sam Hart for coordinating data collection.

Funding information

GJD is supported by the Royal Society through provision of the Ken Murray research professorship, LW was funded by the ERC (Glycopoise) and by the Royal Society PDRA support provided through the Ken Murray Research Professorship, and RJR was funded by the BBSRC through the White Rose Doctoral Training program.

References

Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G. J., Wilson, K. S. & Cowtan, K. D. (2015). *Nat. Struct. Mol. Biol.* **22**, 833–834.
Artola, M., Kuo, C.-L., Lelieveld, L. T., Rowland, R. J., van der Marel, G. A., Codée, J. D. C., Boot, R. G., Davies, G. J., Aerts, J. M. F. G. & Overkleeft, H. S. (2019). *J. Am. Chem. Soc.* **141**, 4214–4218.
Artola, M., Wu, L., Ferraz, M. J., Kuo, C.-L., Raich, L., Breen, I. Z., Offen, W. A., Codée, J. D. C., van der Marel, G. A., Rovira, C., Aerts, J. M. F. G., Davies, G. J. & Overkleeft, H. S. (2017). *ACS Cent. Sci.* **3**, 784–793.
Aviezer, D., Brill-Almon, E., Shaaltiel, Y., Hashmueli, S., Bartfeld, D., Mizrachi, S., Liberman, Y., Freeman, A., Zimran, A. & Galun, E. (2009). *PLoS One*, **4**, e4792.
Barton, N. W., Brady, R. O., Dambrosia, J. M., Di Bisceglie, A. M., Doppelt, S. H., Hill, S. C., Mankin, H. J., Murray, G. J., Parker, R. I., Argoff, C. E., Grewal, R. P. & Yu, K. (1991). *N. Engl. J. Med.* **324**, 1464–1470.
Barton, N. W., Furbish, F. S., Murray, G. J., Garfield, M. & Brady, R. O. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 1913–1916.
Beck, M. (2018). *Dev. Med. Child Neurol.* **60**, 13–18.
Beenakker, T. J. M., Wander, D. P. A., Offen, W. A., Artola, M., Raich, L., Ferraz, M. J., Li, K.-Y., Houben, J. H. P. M., van Rijssel, E. R., Hansen, T., van der Marel, G. A., Codée, J. D. C., Aerts, J. M. F. G., Rovira, C., Davies, G. J. & Overkleeft, H. S. (2017). *J. Am. Chem. Soc.* **139**, 6534–6537.
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. & Aerts, J. M. F. G. (2017). *ACS Chem. Biol.* **12**, 1830–1841.
Ben Turkia, H., Gonzalez, D. E., Barton, N. W., Zimran, A., Kabra, M., Lukina, E. A., Giraldo, P., Kisinovsky, I., Bavdekar, A., Ben Dridi, M.-F., Gupta, N., Kishnani, P. S., Sureshkumar, E. K., Wang,

Chemistry–A European Journal | Full Paper
doi.org/10.1002/chem.202102359
Chemistry Europe
European Chemical Societies Publishing

Design, Synthesis and Structural Analysis of Glucocerebrosidase Imaging Agents

Rhianna J. Rowland[+],[a] Yurong Chen[+],[b] Imogen Breen,[a] Liang Wu,[a] Wendy A. Offen,[a] Thomas J. Beenakker,[b] Qin Su,[c] Adrianus M. C. H. van den Nieuwendijk,[b] Johannes M. F. G. Aerts,[c] Marta Artola,[c] Herman S. Overkleeft,*[b] and Gideon J. Davies*[a]

Abstract: Gaucher disease (GD) is a lysosomal storage disorder caused by inherited deficiencies in β-glucocerebrosidase (GBA). Current treatments require rapid disease diagnosis and a means of monitoring therapeutic efficacy, both of which may be supported by the use of GBA-targeting activity-based probes (ABPs). Here, we report the synthesis and structural analysis of a range of cyclophellitol epoxide and aziridine inhibitors and ABPs for GBA. We demonstrate their covalent mechanism-based mode of action and uncover binding of the new *N*-functionalised aziridines to the ligand binding cleft. These inhibitors became scaffolds for the development of ABPs; the O6-fluorescent tags of which bind in an allosteric site at the dimer interface. Considering GBA's preference for O6- and *N*-functionalised reagents, a bi-functional aziridine ABP was synthesized as a potentially more powerful imaging agent. Whilst this ABP binds to two unique active site clefts of GBA, no further benefit in potency was achieved over our first generation ABPs. Nevertheless, such ABPs should serve useful in the study of GBA in relation to GD and inform the design of future probes.

Introduction

Gaucher disease (GD) is the most common lysosomal storage disorder which is caused by inherited deficiencies in β-glucocerebrosidase (glucosylceramidase, GCase, GBA, EC 3.2.1.45). This lysosomal glycoside hydrolase is encoded by the *GBA1* gene[1] and according to The Human Gene Mutation Database (www.hgmd.org, Institute of Medical Genetics in Cardiff[2]) over 500 genetic mutations at the *GBA1* locus are known. Moreover, mutations in the *GBA1* gene have recently been identified as the highest known genetic risk factor for Parkinson's disease (PD),[3–6] further intensifying therapeutic interest in the *GBA1* gene and enzyme.

[a] *R. J. Rowland,[+] Dr. I. Breen, Dr. L. Wu, W. A. Offen, Prof. G. J. Davies*
Department of Chemistry, York Structural Biology Laboratory (YSBL)
University of York
Heslington, York, YO10 5DD (UK)
E-mail: gideon.davies@york.ac.uk

[b] *Y. Chen,[+] Dr. T. J. Beenakker, A. M. C. H. van den Nieuwendijk, Prof. H. S. Overkleeft*
Department of Bio-organic Synthesis, Leiden Institute of Chemistry
Leiden University
Einsteinwegg 55, 2300 RA Leiden (Netherlands)
E-mail: h.s.overkleeft@chem.leidenuniv.nl

[c] *Q. Su, Prof. J. M. F. G. Aerts, Dr. M. Artola*
Department of Medicinal Biochemistry, Leiden Institute of Chemistry
Leiden University
Einsteinwegg 55, 2300 RA Leiden (Netherlands)

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202102359

GD is primarily characterized by the cellular accumulation of glucosylceramide (GlcCer), and its deacylated derivative glucosylsphingosine (GlcSph), as a result of deficient GBA activity.[7–9] The multisystemic storage of these glycolipids leads to the clinical symptoms of GD, which can vary considerably in frequency and severity. Clinical manifestation of GD type 1 and GD type 2 commonly include skeletal disease and visceral disease affecting the spleen, kidneys, liver and heart.[10–13] In more severe cases (GD type 3), neurological disorders also arise due to GlcCer deposition in the brain.[14,15]

GBA is a 497 amino-acid membrane-associated glycoprotein belonging to GH30 CAZy family (www.cazy.org) of retaining β-glucosidases.[16] GBA is primarily responsible for catalyzing the degradation of GlcCer by hydrolytic cleavage of the β-glucose moiety from the aglycone to yield free ceramide and glucose.[7,17,18] This is achieved with retention of β-anomeric stereochemistry of the released glucose unit through a Koshland double displacement mechanism using Glu340 as the catalytic nucleophile and Glu235 as the catalytic acid-base. The enzymatic nucleophile of GBA was identified (and corrected from Asp443[19]) by the Wither's lab through covalent-trapping of the enzyme with a 2-fluoro-2-deoxy glucoside inactivator.[20] Given the clinical importance of GBA in both GD and PD, it is arguably the most widely studied human glucosidase, with relentless interest in developing novel chaperones,[21–24] inhibitors[25–27] and activity-based probes (ABPs)[28,29] to study this enzyme in disease pathogenesis, diagnosis and treatment.

The seminal 3D structure of GBA was reported in 2003,[30] followed by a number of co-crystal complexes with imino-sugar inhibitors *N*-butyldeoxynojirimycin and *N*-nonyldeoxynojirimycin.[31] Later studies have investigated GBA at the 3D level to obtain insight into GD mutations and potential molecular chaperone binding motifs.[32–35] For example, active

site directed quinazoline modulators,[21,36] competitive 3,4,5,6-tetra-hydroxyazepane inhibitors[37] and uncompetitive pyrrolo [2,3-b]pyrazines inhibitors[38] have recently been reported.

Cyclophellitol, a natural product originally isolated from the mushroom *Phellinus sp.*,[39] is a potent and irreversible β-glucosidase inhibitor which exhibits considerably improved selectivity over its close structural homologue conduritol-B-epoxide (CBE[40,41]).[42] Building on this enhanced selectivity, and with inspiration from the Withers' fluoro-glycosides,[43–45] we have developed a range of cyclophellitol-based inhibitors and ABPs for GBA (Figure 1a,b). Indeed, tagged-cyclophellitol epoxides and aziridines provide a powerful activity-based protein profiling (ABPP) approach to the study of GBA both *in situ* and *in vitro*,[28,46] with potential applications in diagnostics and therapeutic evaluation. In a previous study, we touched upon the binding of a O6-adamantane substituted cyclophellitol inhibitor and a Cy5-tagged cyclophellitol ABP for GBA,[27] however, these structures presented just a fraction of our structural work on the cyclophellitol-based inactivators. In fact, most of these inhibitors and their recent ABP iterations have not been observed on the 3D structure of GBA, hindering a fundamental understanding of ABP reactivity, specificity and conformation.

Here, we sought to structurally dissect the binding of various cyclophellitol-based inhibitors and ABPs with human GBA, comparing epoxides vs. aziridines (**1**, **5** vs. **2**, **3**, **4**, **6**), *gluco*- vs. *galacto*-configuration (**1**, **2**, **3**, **5**, **6** vs. **4**, **7**), aziridine nitrogen functionalization (**3**, **6**), O6-functionalisation with fluorescent reporter groups (BODIPY **5** and Cy5 **6**) and bi-functionalization at both the O6-position and aziridine nitrogen (**6**), (Figure 1b,c). In doing so, we hoped to establish the binding mode of these inactivators and uncover key mechanistic and structural information to inform the design of more efficacious probes for studying human GBA with regards to GD.

Results and Discussion

Tagged cyclophellitol epoxides and aziridines are valuable tools for studying the activity of retaining glycosidases in a wide range of applications, from biomedical purposes in human health and disease[27,47,48] to biotechnology for the study of

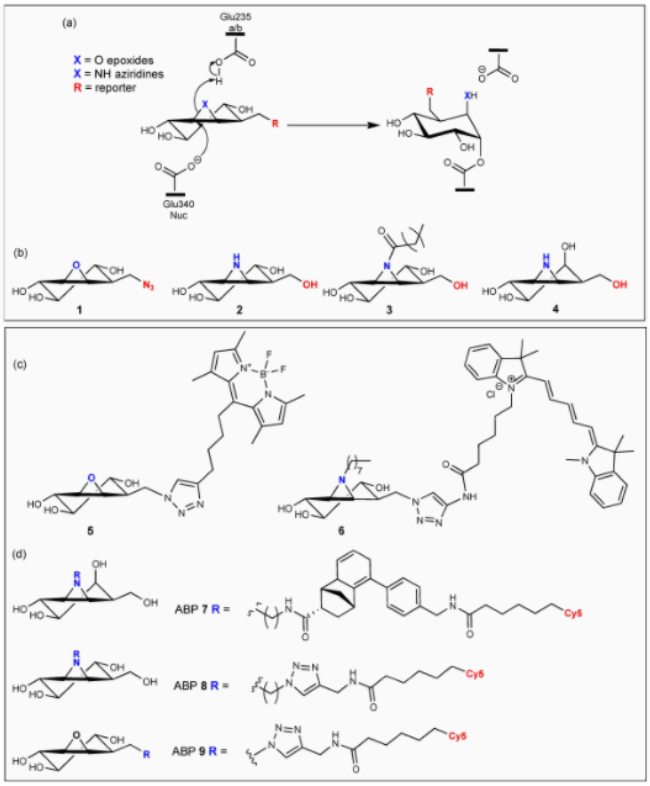


Figure 1. (a) Reaction mechanism of GBA inactivation by irreversible cyclophellitol epoxide and aziridine inhibitors. Panel of (b) cyclophellitol epoxide and aziridine inhibitors (**1**–**4**) and (c) activity-based probes (ABPs) (**5**–**6**) structurally investigated in this work. (d) Range of *galacto*- and *gluco*-configured cyclophellitol ABPs (**7**–**9**) employed for comparative ABPP studies in this work.

biomass degrading enzymes.[49–51] Here, we report the design, synthesis and structural analysis of a panel of β-glucose configured cyclophellitol-based inhibitors and ABPs on the 3D structure of GBA with supportive ABPP studies (Figure 1).

Unsubstituted cyclophellitol epoxide (1)

A co-crystal structure of recombinant human GBA (rhGBA) in complex with azide-tagged cyclophellitol **1** was obtained at 1.70 Å resolution, revealing classical trans-diaxial ring-opening of the epoxide warhead to form a covalent enzyme-inhibitor complex with the enzymatic nucleophile (Glu340), (Figure 2a). The covalent bond length was measured to be 1.42 Å, with the reacted cyclitol adopting the $^4C_1$ chair conformation. This enzyme-inhibitor complex is consistent with the β-glucoside conformational reaction itinerary (which follows a Michaelis Complex→Transition state‡→Covalent Intermediate itinerary of $^1S_3$→$^4H_3$→$^4C_1$) and is consistent with the revised conduritol-β-epoxide (CBE) complex,[52] which was corrected from the originally reported boat conformation.[53] Additionally, the cyclophellitol moiety forms an extensive hydrogen bonding network within the active site, making hydrogen bonds with Asp127, Trp179, Asn234, Tyr313 and Trp381. Whilst electron density for the C6 azido-tag was weak, likely due to flexibility and disorder, there was sufficient density to model the azide substituent extending 'upwards' into a relatively spacious cavity at the back of the active site. An absence of azide electron density for inhibitor **1** has been reported previously when in complex with an unrelated bacterial β-glucoside (*Thermotoga maritima*, T*m*GH1).[54] In contrast to this bacterial co-complex, the improved azido-electron density of **1** in complex with rhGBA reported here provides some insight into the conformation of the azide-tag, which may serve as a ligation handle for two step activity-based protein profiling (ABPP).[55,56] Indeed, further functionalization of this azide-tag for two-step labelling is structurally supported by the relatively large open cavity in which the azide-substituent binds, which would likely accommodate larger reporter groups. However, in the case of GBA, direct one-step ABPs have proved more effective than two-step strategies for activity-based profiling.[56]

Unsubstituted cyclophellitol aziridine (2)

Configurationally isomeric cyclophellitol epoxides and aziridines are often considered interchangeable, although not always equally potent,[29,57] as inhibitors of retaining glycosidases. Therefore, to probe whether this holds true from a mechanistic inhibition point of view, a crystal structure rhGBA in complex with cyclophellitol aziridine **2** was obtained at 1.70 Å resolution to enable a detailed comparison with cyclophellitol epoxide **1**.

The electron density of the resulting co-crystal structure unambiguously shows that **2** reacts with the enzymatic nucleophile of GBA (Glu340) through the aziridine warhead to form a covalent trans-diaxial ring opened complex which is essentially identical to reacted cyclophellitol epoxide **1**, (Figure 2b). Indeed, **2** forms a covalent complex in the $^4C_1$ chair conformation with a covalent bond length of 1.46 Å to Glu340. The only notable difference between the complex of **2** and **1** lies in their hydrogen bonding networks, where a hydrogen bond is formed between the O6-hydroxyl of **2** and Asn396, and the hydrogen bond to Tyr313 is lost, likely due to the "downwards" displacement of Tyr313 in complex with **2**. The flexibility of Tyr313 has been reported previously[58] and the "downwards" conformation observed in complex with **2** appears to preclude hydrogen bonding with the ring-opened aziridine. Instead, the aziridine warhead forms a hydrogen bond to a nearby water molecule in the active site. Overall, this co-crystal structure demonstrates the almost identical binding mode of cyclophellitol epoxides and aziridines, which supports their interchangeable use in the development of ABPs for GBA. Indeed, both fluorescently-tagged cyclophellitol epoxides and aziridines have proved effective, nanomolar probes for GBA (apparent IC_{50} of 1.24 nM for BODIPY-functionalized epoxide compared to IC_{50} of 1.15 nM for BODIPY-functionalized aziridine[29]), with great potential for the study of GBA in regards to GD.

N-Acyl cyclophellitol aziridine (3)

One route to modifying cyclophellitol aziridines and facilitating their conversion to ABPs is functionalization of the aziridine

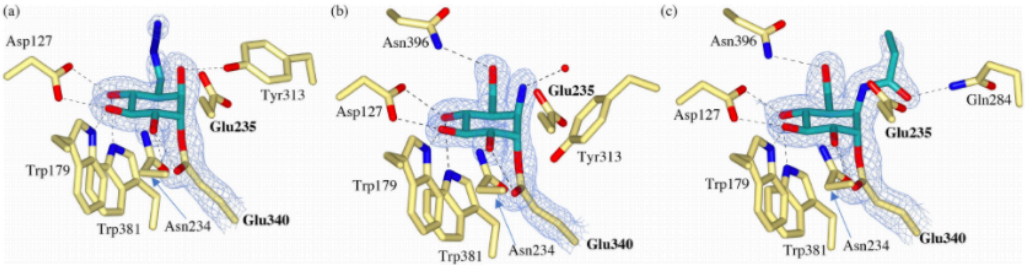


Figure 2. Observed electron density for **1**, **2** and **3** bound covalently to the catalytic nucleophile (Glu340) of rhGBA through trans-diaxial ring opening of the epoxide or aziridine warhead. Weak electron density observed for the C6-azide tag of **1** but clear density for ring-opened *N*-acyl aziridine of **3**. Maximum-likelihood/σA weighted ($2F_o$–F_c) electron density maps contoured to 1.2 σ (a = 1.30 e /Å^3, b = 1.38 e /Å^3, c = 1.30 e /Å^3).

nitrogen.[59] Consequently, *N*-acyl functionalized aziridine **3** was synthesized and found to be a more potent inhibitor than the unsubstituted aziridine analogue **2** (apparent IC_{50} of 0.07 μM for *N*-acyl **3** compared to the 0.5 μM for unsubstituted **2**[57]). In light of this improved potency, we sought to investigate the structural accommodation of the *N*-acyl functionalization by GBA.

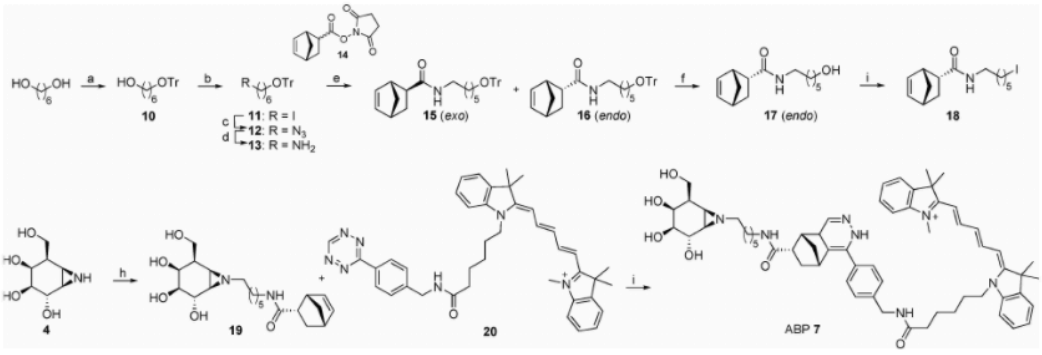
A co-crystal structure of rhGBA in complex with **3** was obtained at 1.76 Å resolution to reveal unambiguous electron density for the reacted cyclitol bound covalently to the enzymatic nucleophile via trans-diaxial ring opening of the *N*-acyl aziridine warhead, (Figure 2c). The cyclophellitol ring adopts the expected $^4C_1$ chair conformation in the resulting complex, with a covalent bond length of 1.42 Å to Glu340. This complex is superimposable with that formed by the unsubstituted aziridine analogue **2**. Importantly, the electron density for the *N*-acyl group and subsequent 2–3 carbon tail was sufficient to model binding of the ring opened *N*-acyl chain to a narrow active site cleft flanked by Tyr313 and Glu284. The carbonyl oxygen of the *N*-acyl aziridine also forms a hydrogen bond with Gln284, introducing an additional hydrogen bond to the cyclophellitol aziridine network. It's possible that binding of this *N*-acyl moiety somewhat mimics the binding of one of the two acyl chains of the natural GlcCer substrate, which may provide a structural basis for the improved potency reported for *N*-acyl aziridines over the unsubstituted aziridine analogue.[57,59] Furthermore, *N*-alkyl functionalized aziridines have also proved to be potent and selective inhibitors of GBA (apparent IC_{50} of 0.017 μM for *N*-alkyl compared to IC_{50} of 0.07 μM for *N*-acyl[57]), with improved chemical stability over their *N*-acyl equivalents.[59] This highlights the potential for future inhibitor and ABP development through *N*-acylation/alkylation of cyclophellitol aziridines.

Cross-reactivity with *galacto*-configured aziridines

One issue surrounding ABPs is their occasional cross-reactivity with related glycosidases. Of note, GBA can be by inhibited both *gluco*- and *xylo*-configured substrates, with demonstrated activity against 4-methylumbelliferyl-β-xyloside;[60] conversely, *galacto*-configured cyclophellitol epoxides have proved inactive against GBA.[61] Nevertheless, studies on other glycosidases have shown that *galacto*-configured inactivators occasionally bind to glucosidases. For example, Gloster et al., 2007 reported binding of T*m*GH1 β-glucoside to *galacto*-hydroximolactam.[62] Therefore, whilst we know *galacto*-configured cyclophellitol epoxides do not bind GBA, we sought to assess if there is potential for cross-reactivity with *galacto*-configured aziridine inhibitor **4** and ABP **7**.

Synthesis of Cy5 Tagged *Galacto*-Configured Cyclophellitol Aziridine (7)

β-Galactose-configured cyclophellitol aziridine **4**, prepared as described previously,[63] was alkylated with 8-azido-1-iodooctane under basic conditions (Scheme 1). Unfortunately, Cu(I)-catalyzed azide/alkyne cycloaddition (CuAAC) with Cy5-alkyne proved abortive and a complex mixture was obtained. In an attempt to obtain a β-galactosidase ABP, the inverse-electron-demand Diels-Alder (IEDDA) ligation strategy was investigated to introduce the fluorophore at the final synthetic step. For this purpose, norbornene-modified cyclophellitol aziridine **19** was synthesized and reacted with tetrazine-Cy5 **20** to obtain ABP **7**. The synthesis of norbornene-functionalized aziridine started with monotritylation of 1,6-hexanediol to give **10** followed by iodination of the primary alcohol with iodine and substitution with sodium azide to afford **12**. Reduction of azide **12** using triphenylphosphine on beads gave amine **13**. Norbornene-OSu ester **14** was obtained according to the literature procedure[64] as a mixture of *endo*- and *exo*-isomers. This mixture was



Scheme 1. Synthesis of ABP **7**. Reagents and conditions: a) TrCl, pyridine, CH_2Cl_2, rt, 90 min, 95%; b) imidazole, PPh_3, I_2, Et_2O, CH_3CN, rt, overnight, 80%; c) NaN_3, DMF, 80 °C, overnight, quant; d) polymer-bound PPh_3, H_2O, THF, 48 h, quant; e) norbornene-OSu, DIPEA, DCE, rt, overnight, **15**: 28%, **16**: 68%; f) *p*-toluenesulfonic acid, CH_2Cl_2, MeOH, rt, overnight, 86%; g) PPh_3, I_2, imidazole, THF, reflux, 1.5 h, 73%; h) **18**, K_2CO_3, DMF, 75 °C, overnight, 12%; i) Cy5 tetrazine **20**, MeOH, overnight, 87%.

condensed with amine **13** resulting in a mixture of norbornene-trityl products (*exo/endo*, 1 : 2.3), which were easily separated by silica gel column chromatography. Deprotection of *endo*-product **16** under acidic conditions gave alcohol **17** in 86% yield. Treatment of **17** with triphenylphosphine, iodine and imidazole at elevated temperatures in tetrahydrofuran. afforded iodonorbornene **18**. Cyclophellitol aziridine **7** was then *N*-alkylated with iodonorbornene linker **18** using potassium carbonate as base in DMF. Subsequent ligation[65] with Cy5 tetrazine **19** gave ABP **7** as mixture of isomers.

In-solution labelling of rhGBA by *galacto*-configured cyclophellitol aziridine ABP 7

To rapidly investigate for cross-reactivity with *galacto*-configured cyclophellitol aziridines, time-dependent labelling assays of rhGBA with *galacto*-configured Cy5-tagged aziridine ABP **7** were performed (Figure 3).

To our surprise, ABP **7** (150 nM) rapidly labelled rhGBA within 2 minutes and reached saturation within 30-minutes (Figure 3). In comparison, labelling by *gluco*-configured ABP **8** (150 nM) achieved saturation within 2 minutes (Figure 3). Whilst *galacto*-configured ABP **7** is slower to label rhGBA compared to *gluco*-configured ABP **8**, this simple labelling assay demonstrates the cross-reactivity potential of β-*galacto*-configured aziridine ABPs with GBA and suggests that such ABPs should be carefully analysed for cross reactivity before interpretation *in vivo*. However, it should also be noted that broad spectrum ABPs have proved useful in enzyme and inhibitor discovery, illustrating that not all ABPs need to be highly specific.[66] Nevertheless, it is important to establish the reactivity and binding of such ABPs to understand their limitations and identify areas for future improvements. Therefore, cross-reactivity data such as this are important in guiding the selection of ABPs for a desired application.

Galacto-configured cyclophellitol aziridine (4)

To further understand the binding of *galacto*-configured cyclophellitol aziridines with rhGBA, a co-crystal structure with *galacto*-configured **4** was obtained at 1.80 Å. The resulting complex revealed that **4** covalently modifies the catalytic nucleophile of GBA in an almost identical fashion to *gluco*-configured reagents **1** and **2**. Specifically, **4** reacts with Glu340 via trans-diaxial ring opening of the aziridine trap to form a covalent complex in the $^4C_1$ conformation with a bond length of 1.45 Å (Figure 4a). The key difference between this *galacto*-configured complex and the *gluco*-configured analogues is the axial conformation of the O4 hydroxyl, which is still able to form a hydrogen bond with Asp127. Indeed, the hydrogen bonding network of **4** is very similar to that of *gluco*-configured **1** and **2**, with both O3 and O4 able to form hydrogen bonds to Asp127 regardless of whether the sugar is *gluco*- or *galacto*-configured. The only notable difference is the absence of a hydrogen bond between the O4 hydroxyl and Trp381, presumably owing to its axial conformation.

This led us to ponder why *galacto*-aziridines appear to be reasonably potent inhibitors of GBA whereas the equivalent O6-modified *galacto*-epoxides are essentially unreactive (as reported by Marques et al., 2017). Overlay of *galacto*-aziridine **4** and *gluco*-epoxide **1** provides an immediate clue: the O6 hydroxyl of *galacto*-aziridine **4** has rotated such that it points 'downwards', perhaps to avoid steric clash with the axial *galacto*-O4, (Figure 4b). Thus, one possible explanation for the cross-reactivity of *galacto*-aziridines but not O6-substituted *galacto*-epoxides, is the requirement to displace the O6-hydroxyl for covalent binding which would not be possible were the O6-substituted. Therefore, it appears that the placement of the reporter group is important in controlling the selectivity and cross-reactivity of cyclophellitol-based ABPs. In combination with the ABP labelling assays of rhGBA, this co-crystal structure serves as a caution when assuming similar configurational specificity of cyclophellitol epoxides and aziridines, which evidently does not always translate to similar glycosidase specificity.

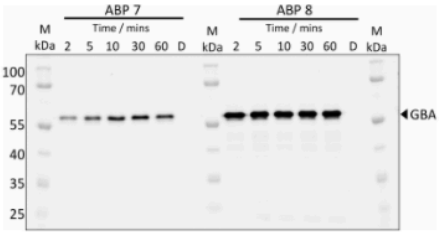


Figure 3. Time dependent labelling of rhGBA (700 nM) by Cy5-tagged *galacto*-ABP **7** (150 nM) after 2, 5, 10, 30 and 60 mins showing that labelling reaches saturation at 30 mins. In comparison, labelling by Cy5-tagged *gluco*-ABP **8** (150 nM) reaches saturation in 2mins. D = denatured protein sample demonstrating there is no unspecific labelling with inactive rhGBA. ABP labelled rhGBA (60 kDa) visualized by Cy5 fluorescent readout.

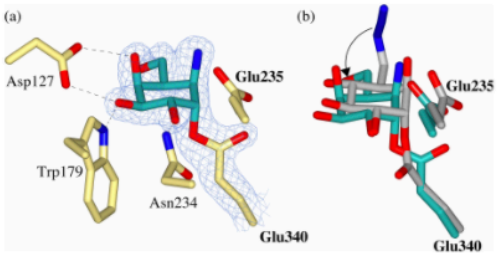


Figure 4. (a) Observed electron density for *galacto*-configured aziridine **4** bound covalently to the catalytic nucleophile (Glu340) of rhGBA through trans-diaxial ring opening of the aziridine warhead. Maximum-likelihood/σA weighted ($2F_o$–F_c) electron density map contoured to 1.2 σ (1.31 e /Å^3) (b) Overlay of *galacto*- configured aziridine **4** (teal) and *gluco*-configured azide-tagged epoxide **1** (grey) demonstrating rotation of the O6 hydroxyl of **4**.

BODIPY-tagged cyclophellitol epoxide (5) in complex with GBA inactivated with *N*-acyl aziridine (3)

In addition to a previously reported structure of rhGBA in complex with the O6- Cy5 tagged cyclophellitol epoxide ABP **9**,[27] a serendipitous co-complex of O6- BODIPY tagged ABP **5** and *N*-acyl aziridine **3** was obtained in this study by accidental soaking of rhGBA crystals in complex *N*-acyl aziridine **3** with additional ABP **5**. In the resulting co-complex, clear electron density was observed for **3** bound covalently to Glu340, with its ring-opened *N*-acyl chain bound in the narrow active site cavity flanked by Gln283 and Tyr313 (Figure 5a). Additionally, unambiguous electron density for an intact molecule of ABP **5** was observed, allowing the full ABP to be modelled at the surface of molecule A of the crystallographic dimer, (Figure 5b). Specifically, the BODIPY tag binds at the dimer interface in a hydrophobic cavity formed by Leu241, Tyr244, Pro245, Phe246 and Tyr313, which we reported previously to accommodate the Cy5 tag of ABP **9**.[27] The unreacted cyclophellitol of **5** sits exposed at the surface of the protein, making no hydrogen bonding interactions with the enzyme; however, the hydrophobic face of the cyclophellitol unit lies above the exposed side chain of Leu241 (Figure 5a). Whilst the binding of ABP **5** could be considered artefactual, it is notable that very simple torsional rotation of the linker allows the linker and cyclophellitol unit to be placed in the active site, putting the epoxide warhead in perfect super-position with the trapped aziridine without any movement of the BODIPY group itself. In combination with the previously reported structure of rhGBA in complex with a Cy5 tagged ABP **9**,[27] this ABP **5** co-complex provides further evidence for a unique binding mode of this hydrophobic cavity, which has recently been exploited for the binding of a novel class of pyrrozopyrazine activators with chaperoning potential.[38]

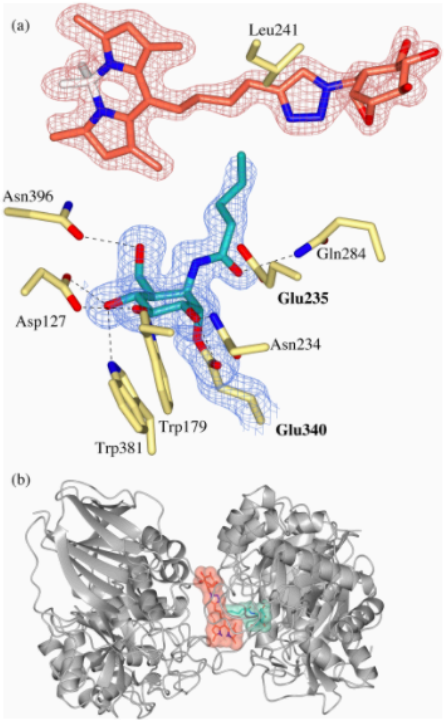


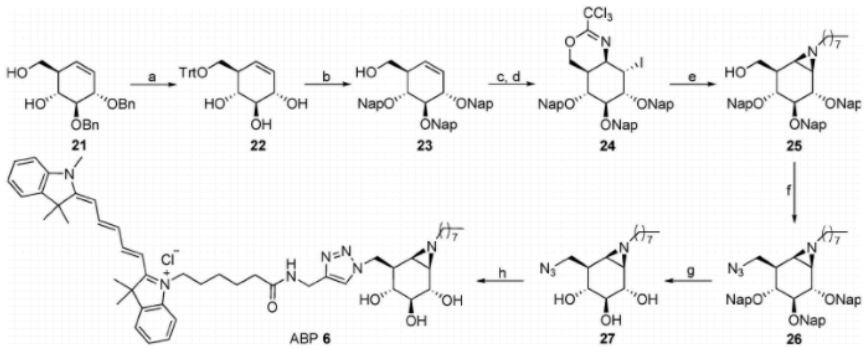
Figure 5. (a) Observed electron density for binding of *N*-acyl aziridine **3** in the active site of rhGBA and intact BODIPY-tagged ABP **5** at dimer interface. Maximum-likelihood/σA weighted ($2F_o–F_c$) electron density map contoured to 1.0 σ (a = 1.29 e /Å^3) (b) Ribbon diagram of rhGBA dimer shows binding of ABP **5** in a hydrophobic cavity at the dimer interface (red surface), with *N*-acyl **3** bound in the active site (blue surface).

GBA is notable for its tolerance, indeed preference, for O6-substituted reagents which exhibit increased specificity for GBA over other β-glucosidases, including non-lysosomal GBA (GBA2) and generic GH1 β-glucosidases.[27] GBA also favours iminosugar inhibitors and cyclophellitol aziridines extended at the aziridine nitrogen position. Considering these preferences and the structural information provided by the co-complex of **3** and **5**, we suggested that binding of the O6-substituent and aziridine *N*-functionalization are structurally exclusive and may reflect the enzymes specificity for a lipid substrate with two acyl tails. This led us to ponder whether a new generation of bifunctional aziridine ABPs, which are functionalised at the O6-position and aziridine nitrogen, may exhibit further improvements in potency and selectivity for GBA. Consequently, the O6- Cy5 tagged *N*-octyl bifunctional aziridine ABP **6** was synthesised and its GBA activity and selectivity were compared to the parent Cy5 tagged cyclophellitol aziridine ABP **8** and cyclophellitol epoxide ABP **9**.

Cy5 Tagged Bi-Functional Cyclophellitol Aziridine (6)

Synthesis of Cy5 and N*-octyl bifunctional cyclophellitol aziridine (6)*

In this work, we developed a new synthetic route towards 6-azido octyl aziridine **27** following a key 2-napthylmethyl ether (Nap) protecting group strategy (Scheme 2). Briefly, starting intermediate **21** was synthesized in nine steps from D-xylose based on chemistry developed by Madsen and co-workers.[67,68] Benzyl deprotection was not compatible with the required azide functionality and electrophilic aziridine, therefore, debenzylation of **21** with boron trichloride (BCl_3) and selective tritylation of the primary alcohol afforded intermediate **22**. The secondary alcohols were protected as Nap ethers followed by detritylation of the primary alcohol to afford **23**. Treatment of **23** with trichloroacetonitrile yielded a primary imidate intermediate, which was treated with *N*-iodosuccinimide (NIS) to stereospecifically afford the cyclic imidate **24**. Acidolysis and base treatment of **24** resulted in the formation of a free aziridine, which was then alkylated with an octyl linker to form compound **25**. Tosylation of the primary alcohol of **25** followed by azide substitution resulted in **26**. Napthylmethyl ethers were then removed by DDQ to afford compound **27**, which was



Scheme 2. Synthesis of ABP **6**. Reagents and conditions: a) *i*) BCl_3, DCM, - 78°C, 2 h; *ii*) TrCl, Et_3N, DMAP, DMF, rt, 19 h, 30% over two steps; b) *i*) NapBr, NaH, TBAI, DMF, 0°C-rt, 5 h; *ii*) TsOH, DCM/MeOH (1/1), rt, overnight, 73% over two steps; c) Cl_3CCN, DBU, DCM, rt, overnight; d) NIS, $CHCl_3$, rt, 17 h, 40% over two steps; e) *i*) HCl, DCM/MeOH (1/1), rt, overnight, then Amberlite IRA-67, 20 h; *ii*) 1-iodooctane, K_2CO_3, DMF, 80°C, 16 h, 35% over two steps; f) *i*) TsCl, Et_3N, 1-methyl imidazole, DCM, rt, 28 h; *ii*) NaN_3, DMF, 50 °C, 40 h, 68% over two steps; g) DDQ, DCM/H_2O (10/1), rt, 24 h, 66%; h) Cy5-alkyne, $CuSO_4$, NaAsc, DMF, rt, overnight, 30%.

finally coupled with Cy5 using click chemistry to yield ABP **6** (Scheme 2).

Cy5-tagged bifunctional cyclophellitol aziridine (6)

To investigate the accommodation of the two functionalities of ABP **6** by GBA, a co-crystal structure in complex with bifunctional ABP **6** was obtained at 1.80 Å resolution, demonstrating covalent binding of the cyclophellitol aziridine to the catalytic nucleophile of GBA (Figure 6a). Specifically, the reacted cyclophellitol adopts the expected $^4C_1$ chair conformation, with a covalent bond length of 1.47 Å to Glu340. Furthermore, unambiguous electron density for the ring opened *N*-alkyl aziridine warhead was observed, allowing the first 5 carbons of the *N*-octyl chain to be modelled. This was sufficient to establish binding of the *N*-alkyl chain to the narrow active site channel formed by Gln284, Tyr313, Lys346 and Trp348, which is consistent with the complex of *N*-acyl aziridine **3**. In fact, the *N*-alkyl chain of ABP **6** extends through this pocket towards the surface of the protein, which may provide some indication into the binding of the fatty acid portion of the natural GlcCer substrate which is thought to project out from the protein and interact with the lipid bilayer.[30] Unfortunately, whilst sufficient electron density for the O6-triazole linker and subsequent amide group was observed, the Cy5 tag could not be modelled. Nevertheless, the O6-triazole linker was found to bind in the hydrophobic cavity formed by Trp348, Phe246 and Tyr313, which we reported previously to accommodate the triazole linker of ABP **9**.[27] Additionally, this binding cavity extends towards the broader hydrophobic allosteric site at the dimer interface where we reported the BODIPY tag of ABP **5** to bind.

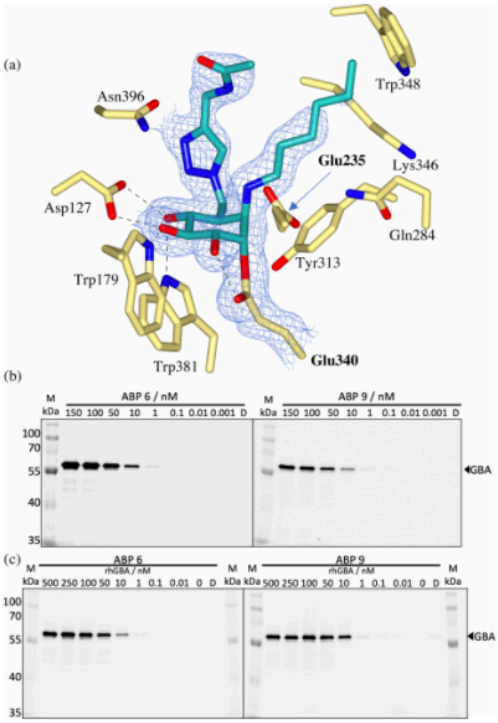


Figure 6. (a) Observed electron density for ABP **6** bound covalently to the catalytic nucleophile (Glu340) of rhGBA by trans-diaxial ring opening of the *N*-alkyl aziridine warhead. Maximum-likelihood/σA weighted ($2F_o-F_c$) electron density map contoured to 1.0 σ (a = 1.31 e⁻/Å³). (b) Labeling of rhGBA (200 nM) with decreasing concentrations of ABP **6** or ABP **9** (150–0.001 nM) at 37 °C for 30 mins followed SDS-PAGE separation. (c) Incubation of ABP **6** or ABP **9** (150 nM) with decreasing concentrations of rhGBA (500–0.01 nM) followed by SDS-PAGE analysis. Fluorescently labelled rhGBA visualized by Cy5 fluorescent readout. D = denatured protein sample.

Activity-based labelling of rhGBA with ABP 6 and ABP 9

Activity-based labelling of rhGBA (produced in an insect-baculovirus expression vector system[58]) with bi-functional ABP **6** was performed and compared to labelling by its mono-functionalized epoxide derivative ABP **9**. Firstly, in solution labelling of excess but constant rhGBA (200 nM) was performed in the presence of decreasing ABP concentration (150–0.001 nM), demonstrating clear concentration dependent labelling with a gel-detection limit of 1 nM for ABP **6** and 0.1 nM for ABP **9** (Figure 6b). Secondly, labelling assays in which ABP **6** and ABP **9** (150 nM) were incubated with decreasing rhGBA concentrations (500–0.01 nM) were performed to further demonstrate the concentration dependent labelling down to 1 nM rhGBA with ABP **6** and 0.1 nM rhGBA with ABP **9**, (Figure 6c). These in gel detection limits are concordant across both assays and demonstrate that ABP **9** exhibits ca. 10-fold increase in potency.

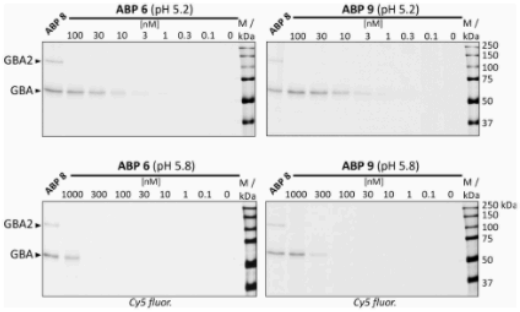


Figure 7. Fluorescent labeling of mouse brain lysate (25 μg total protein) with different concentrations of bi-functional ABP **6** and mono-functional ABP **9** at pH 5.2 (upper panels) or pH 5.8 (lower panels) after incubation for 30 min at 37 °C. Labeling by broad-spectrum β-glucosidase ABP **8** is shown for comparison (100 nM, pH 5.2 or pH 5.8, 30 min, 37 °C).

In vitro *activity and selectivity of Cy5 biofunctionalized cyclophellitol aziridine ABP 6*

To further investigate the potency and selectivity of ABP **6**, *in vitro* activity assays against GBA and two other related β- and α-glucosidases (GBA2 and GAA) were performed and compared to ABP **9**.

ABP **6** and ABP **9** were pre-incubated with recombinant human GBA (rhGBA, Imiglucerase), human GBA2 (from lysates of GBA2 overexpressed cells) and recombinant human GAA (rhGAA, Myozyme) for 3 hr followed by enzymatic activity measurement using 4-methylumbelliferyl-β- and α-glucosides as fluorogenic substrates. As shown in Table 1, ABP **6** proved to be a nanomolar inhibitor of GBA (with an apparent IC_{50} value of 53 nM) and is inactive toward GBA2 and GAA (apparent IC_{50} values >100 μM), thus exhibiting comparable selective inhibition of GBA (IC_{50} ratio $>10^3$ for both GBA2/GBA and GAA/GBA) as we reported previously for ABP **9**.[27]

We next investigated their labeling efficiency and selectivity toward GBA in mouse brain lysate at pH 5.2 (containing 0.2% taurocholate and 0.1% Triton-100) and pH 5.8 as respective optimal conditions for GBA and GBA2 activities (Figure 7). As expected, ABP **6** and **9** selectively labeled GBA in a concentration-dependent manner under pH 5.2 (upper panels), with significant labeling observed at 10 nM ABP. Under pH 5.8 (lower panels), the labeling efficiency of both ABPs towards GBA decreased and significant labeling can only be observed at 1000 nM (ABP **6**) or 300 nM (ABP **9**). Importantly, no labelling of GBA2 was observed up to 1000 nM, showing good GBA selectivity. For comparison, broad-spectrum β-glucosidase ABP **8** efficiently labelled both GBA and GBA2 at 100 nM under both pHs.

Table 1. Apparent IC_{50} values for *in vitro* inhibition of rhGBA, rhGAA and GBA2 from overexpressed cell lysates by ABP **6** and **9**. Error ranges depict standard deviations from technical duplicates.

Compound	*In vitro* rhGBA IC_{50} (nM)	*In vitro* GBA2 (HEK293T lysate) IC_{50} (nM)	*In vitro* rhGAA IC_{50} (nM)
ABP 6	53.1 ± 2.65	$>10^5$	$>10^5$
ABP 9[27]	3.20 ± 0.17	$412\times10^3\pm 10.1\times10^3$	$>10^5$

In situ *labeling of GBA and GBA2 in living cells by ABP 6, 8 and 9*

To further evaluate the selectivity of ABP **6**, **8** and **9**, *in situ* labeling of GBA and GBA2 in living cells were investigated. HEK293T cells containing endogenous GBA and overexpressed GBA2 were treated with the ABPs at different concentrations (1–1000 nM) for 24 hr. Cells were then washed, lysed and the fluorescence was visualized by gel-based ABPP (Figure 8). Treatment with broad spectrum ABP **8** resulted in unbiased labeling of GBA and GBA2 at 10 nM, with labeling of both enzymes reaching saturation at 100 nM after 24 h incubation. In comparison, selective labeling of GBA in ABP **6** treated cells was observed at 10 nM, with some GBA2 labelling observed at higher probe concentrations (100 nM). More selective labeling

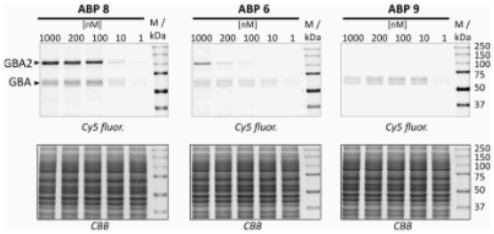


Figure 8. *In situ* labeling of GBA and GBA2 in HEK293T cells with ABP **6**, ABP **8** and ABP **9** at varying concentrations at 37 °C for 24 h, followed by SDS-PAGE separation and fluorescent readout (top panel). *CBB*, Coomassie Brilliant Blue staining (bottom panel).

of GBA was achieved with ABP **9**, which did not label GBA2 even at the highest concentration of probe applied (1000 nM).

Therefore, despite demonstrating that the O6- and aziridine nitrogen substituents of ABP **6** are structurally exclusive and are accommodated by GBA in two unique active site clefts, ABP **6** exhibits no further improvements in potency or selectivity for GBA over the O6 mono-functionalized ABPs. Nevertheless, this bi-functional ABP remains a nanomolar inhibitor of GBA which provides future opportunities for ABP development through modification of both the O6- and aziridine nitrogen substituents. This work also exemplifies the complexity of ABP development.

Conclusion

Tagged cyclophellitols offer a powerful activity-based protein profiling approach for the visualization and quantification of specific enzymatic activities. Here, we report the design, synthesis, and structural analysis of a range of cyclophellitol epoxide and aziridine inactivators and activity-based probes (ABPs) for human β-glucocerebrosidase (GBA). These studies not only demonstrate the mechanism-based mode of action of these compounds as covalent inactivators, but also highlight binding of *N*-functionalized aziridines to the active site cleft. The cyclophellitol-based inhibitors subsequently served as scaffolds for the development of ABPs; the O6-fluorescent tags of which bind to an allosteric site at the dimer interface. In light of the accommodation of *N*-functionalized aziridines and O6-substituents by GBA, we synthesized a bi-functional O6- Cy5 *N*-octyl aziridine ABP which we hoped would offer a more powerful imaging agent. Whilst we structurally validated that the O6- and aziridine functionalities are structurally exclusive and bind in two distinct active site clefts, this bi-functional ABP showed no benefit in potency or selectivity over the mono-functionalized ABPs. Nevertheless, this study provides fundamental insight into ABP reactivity, specificity, and conformation with a tale of caution on ABP cross-reactivity when assuming similar glycosidase specificity of configurationally isomeric cyclophellitol epoxides and aziridines. We envisage these inhibitors and ABPs will serve useful in the study of GBA in relation to Gaucher Disease and inform the design of next-generation inhibitors and probes.

Experimental Section

Synthesis: General experimental details

General synthetic details can be found in the Supporting Information (page 3).

Experimental procedures and characterization data of products

Inhibitors **1**,[28] **2**,[68] **3**,[57] **4**,[63] and ABPs **5**[28] and **8**[66] were synthesized according to previously published procedures and their spectroscopic data are in agreement with those reported previously. ABPs **6** and **7** were synthesised and characterised according to the procedures outlined in the Supporting Information (pages 3–7).

Biochemical methods: General experimental details

Recombinant human GBA (rhGBA, imiglucerase, Cerezyme®) and GAA (rhGAA, alglucosidase alfa, Myozyme) were obtained from Sanofi Genzyme (Cambridge, MA, USA). rhGBA was also produced in an insect-baculovirus expression vector system (BEVS) and purified according to previously published procedures.[58] Other general biochemical details can be found in the Supporting Information (page 7).

In vitro activity of ABP 6 on rhGBA, GBA2 and rhGAA

In vitro apparent IC_{50} measurements of ABP **6** with rhGBA and rhGAA were determined using the fluorogenic substrate methods described previously.[27] For *in vitro* apparent IC_{50} measurements of GBA2, 8 volumes of cell lysates (4 μg total protein/μL) containing overexpressed human GBA2 were firstly pre-incubated with 1 volume of MDW941 (100 nM final concentration, 0.5% (v/v DMSO)) for 30 min at 37 °C to selectively inhibit GBA activity. Lysates were then incubated with 1 volume of ABP **6** at various concentrations for 3 h at 37 °C, before subsequent enzymatic assay for GBA2 activity as described earlier.[27,69] All assays were performed in duplicate sets, each with 3 technical replicates at each inhibitor concentration. DMSO concentration was kept at 0.5%–1% (v/v) in all assays during incubation with compounds. *In vitro* apparent IC_{50} values were calculated by fitting data with [inhibitor] vs. response-various slope (four parameters) function using Graphpad Prism 7.0 software. Average values and standard deviations were calculated from the two sets.

Time-dependent labelling assays of rhGBA with *galacto*-ABP 7 and *gluco*-ABP 8

rhGBA produced in BEVS[58] was prepared at 700 nM in 150 mM McIlvaine buffer pH 5.2 (containing 0.1% (v/v) Triton X-100 and 0.2% (w/v) sodium taurocholate) and ABP **7** or ABP **8** were added to 150 nM. The reactions were incubated at 37 °C and aliquots were taken at 2, 5, 10, 30 and 60 mins. The aliquots were immediately denatured with Laemmli (x3) sample buffer by heating at 95 °C for 5 minutes. The samples were resolved by electrophoresis in 10% SDS-PAGE gels, running at 200 V for ~50 minutes. Wet slab gels were scanned on fluorescence using an AmershamTyphoon 5 Imager (GE Healthcare) with λ_{EX} 635 nm; λ_{EM} >665 nm.

Titration of ABP 6 and ABP 9 with rhGBA

rhGBA produced in BEVS[58] was diluted to 200 nM in 150 mM McIlvaine buffer pH 5.2 (containing 0.1% (v/v) Triton X-100 and 0.2% (w/v) sodium taurocholate) and ABP **6** or ABP **9** were added to 150, 100, 50, 10, 1, 0.1, 0.01 or 0.001 nM final concentration. The reactions were incubated at 37 °C for 30 mins and denatured with Laemmli (x3) sample buffer at 95 °C for 5 minutes. The samples were resolved by electrophoresis in 10% SDS-PAGE gels, running at 200 V for ~50 minutes. Wet slab gels were scanned on fluorescence using an Amersham Typhoon 5 Imager (GE Healthcare) with λ_{EX} 635 nm; λ_{EM} >665 nm.

List of Abbreviations

AAV	Adeno-associated virus
AcMNPV	Autographa californica multiple nucleopolyhedrovirus
ABP	Activity based probe
ABPP	Activity based protein profiling
Ala	Alanine
AnIEX	Anion exchange
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine triphosphate
BEVS	Baculovirus expression vector system
BODIPY	Boron-dipyrromethene
BSA	Bovine serum albumin
BTP	Bis-Tris propane
BV	Budded virus
C	Celsius
CatIEX	Cation exchange
CAZy	Carbohydrate active enzymes database
CBE	Conduritol B epoxide
CCL18	Chemokine ligand 18
CCP4i2	Collaborative computational project No. 4 (interface 2)
CCP4mg	Collaborative computational project molecular graphics
CDG	Congenital disorder of glycosylation
cDNA	Chromosomal deoxyribonucleic acid
CERT	Ceramide transfer protein
CHIT1	Chitotriosidase
CHO	Chinese Hamster Ovary
CMC	Critical micelle concentration
CMP	Cytidine monophosphate
CNS	Central nervous system
COOT	Crystallographic Object-Oriented Toolkit
CP	Cyclophellitol
CRIPSR	Clustered regularly interspaced short palindromic repeats
CV	Column volume
Cys	Cysteine
Cy5	Cyanine dye 5
Da	Dalton
DDM	N-dodecyl-B-D-Maltoside
DGJ	1-deoxygalactonojirimycin
DLS	Diamond Light Source
DMSO	Dimethyl sulfoxide
DNJ	Deoxynojirimycin
DNA	Deoxyribonucleic acid
DNP	2,4-Dinotrophenol
DSF	Differential scanning fluorimetry

E.coli	Escherichia coli
EC	Enzyme commission (number)
EC_{50}	Half maximal effective concentration
EG	Ethylene glycol
EGC	Endoglycoceramidase
EGFR	Epidermal growth factor
EM	Electron microscopy
EndoH	Endoglycosidase H
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated degradation
ERT	Enzyme replacement therapy
ESI	Electrospray ionisation
FBS	Fetal Bovine Serum
FD	Fabry Disease
Fuc	Fucose
GAA	Lysosomal alpha-glucosidase
Gal	Galactose
GBAP	Glucocerebrosidase pseudo gene
GBA1	β-Glucocerebrosidase 1 (lysosomal) (Homo sapiens)
GBA2	β-Glucocerebrosidase 2 (non-lysosomal) (Homo sapiens)
GCS	Glucosylceramide synthase
Gb3	Globotriaosylceramide
GD	Gaucher Disease
GDP	Guanosine diphosphate
GlcCer	Glucosylceramide
GlcSph	Glucosylsphingosine
GH	Glycoside hydrolase
GLA	Galactosidase alpha
Glc	Glucose
GlcNAC	N-acetyl-glucosamine
GM3	Monosialodihexosylganglioside
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
Gor	Glutathione reductase
GPI	Glycosylphosphatidylinositol
GSL	Glycosphingolipid
GT	Glycosyltransferase
GUSB	Lysosomal beta-glucuronidase
GV	Granulovirus
h	Hours
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
His_6-tag	Hexa histidine tag
HIV	Human immunodeficiency virus
Hi5	High Five Cells
HPLC	High performance liquid chromatography

HSCT	Hematopoietic stem cell transplantation
HSPC	Hematopoietic stem and progenitor stem cell
IC_{50}	Half maximal inhibitory concentration
IEX	Ion exchange
IFE	Inner filter effect
IFG	Isofagomine
IgG	Immunoglobulin G antibody
Ile	Isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
*k*cat	Catalytic efficiency
kDa	Kilodalton
*K*i	Inhibition constant
K_M	Michaelis constant
L	Litre
LacCer	Lactosylceramide
LAMP	Lysosome associated membrane protein
LB	Lysogeny Broth
LC-MS	Liquid chromatography – mass spectrometry
Leu	Leucine
LIMP	Lysosomal integral membrane protein
LSD	Lysosomal storage disorder
Lyso-Gb3	Globotriaosylceramide
m	Minutes
M	Molar
MALLS	Multi-Angle Laser Light Scattering
Man	Mannose
MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligram
MLPA	Multiplex ligation dependent probe amplification
mL	Millilitre
nL	Nanolitre
mM	Millimolar
Mol	Moles
MS	Mass spectrometry
M6P	Mannose-6-phosphate
NAGAL	N-acetylgalactosaminidase
NB-DNJ	N-butyldeoxynojirimycin
ng	Nanogram
nm	Nanometre
nM	Nanomolar
NN-DNJ	N-nonyldeoxynojirimycin
NPV	Nucleopolyhedrovirus
ODV	Occlusion derived virus
OMIM	Online Mendelian Inheritance in Man
OpNPV	*Orgyia pseudotsugata* multi-capsid nucleopolyhedrovirus (OpNPV)
PAGE	Polyacrylamide gel electrophoresis
PC	Pharmacological chaperone

PCR	Polymerase chain reaction
PCT	Pharmacological chaperone therapy
PD	Parkinson disease
PDB	Protein databank
PDMP	D-threo-1-phenyl-2-decanoylamino3-morpholino-1-propanol
PEG	Polyethylene glycol
Phe	Phenylalanine
PNGase	Peptide-N-glycosidase
Pro	Proline
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPM	Revolutions per minute
s	Seconds
SapC	Saposin C (Homo Sapien)
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Ser	Serine
Sf9	Spodoptera frugiperda
sLeX	Sialyl lewis X
SLIC	Sequence and ligation independent cloning
SNFG	Signature nomenclature for glycans
SRT	Substrate reduction therapy
Stdev	Standard deviation
S1P	Sphingosine-1-phosphate
TCEP	Tris(2-carboxyethyl)phosphine
TGN	Trans-Golgi Network
Thr	Threonine
TLC	Thin layer chromatography
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
TrxB	Thioredoxin reductase
Trp	Tryptophan
TSA	Thermal Shift Assay
Tyr	Tyrosine
Tx	*Thermoanaerobacterium xylanolyticum*
UDP	Uridine diphosphate
Vmax	Maximal enzyme velocity
v/v	volume/volume
YFP	Yellow fluorescent protein
WT	Wild type
w/v	weight/volume
4MU	4-methylumbelliferone
µg	Microgram
µL	Microlitre
µM	Micromolar
µmol	Micromole

References

1 R. A. Dwek, *Chem. Rev.* , 1996, **96**, 683–720.
2 S. Dedola, M. D. Rugen, R. J. Young and R. A. Field, *ChemBioChem*, 2020, **21**, 423–427.
3 H. J. Gabius, H. C. Siebert, S. André, J. Jiménez-Barbero and H. Rüdiger, *ChemBioChem*, 2004, 5, 740–764.
4 H. J. Gabius and J. Roth, *Histochem. Cell Biol.*, 2017, **147**, 111–117.
5 K. T. Schjoldager, Y. Narimatsu, H. J. Joshi and H. Clausen, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 729–749.
6 A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. Vliegenthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi and S. Kornfeld, *Glycobiology*, 2015, **25**, 1323–1324.
7 R. A. Laine, *Glycobiology*, 1994, **4**, 759–767.
8 A. V. Nairn, W. S. York, K. Harris, E. M. Hall, J. M. Pierce and K. W. Moremen, *J. Biol. Chem.*, 2008, **283**, 17298–17313.
9 R. A. Laine, in *Glycosciences*, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2008, pp. 1–14.
10 G. W. Hart, *Curr. Opin. Cell Biol.*, 1992, **4**, 1017–1023.
11 R. G. Spiro, *Adv. Protein Chem.*, 1973, **27**, 349–467.
12 J. Kopitz, *Histochem. Cell Biol.*, 2017, **147**, 175–198.
13 R. G. Spiro, *Glycobiology*, 2002, **12**, 43–56.
14 P. R. Crocker and T. Feizit, *Curr. Opin. Struct. Biol.*, 1996, **6**, 679–691.
15 K. Furukawa, Y. Ohkawa, Y. Yamauchi, K. Hamamura, Y. Ohmi and K. Furukawa, *J. Biochem.*, 2012, **151**, 573–578.
16 M. L. Allende and R. L. Proia, *Curr. Opin. Struct. Biol.*, 2002, **12**, 587–592.
17 K. Ohtsubo and J. D. Marth, *Cell*, 2006, **126**, 855–867.
18 A. Varki, *Glycobiology*, 2017, **27**, 3–49.
19 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
20 T. W. Rademacher, R. B. Parekh and R. A. Dwek, *Annu. Rev. Biochem.*, 1988, **57**, 785–838.
21 H. H. Freeze, G. W. Hart and R. L. Schnaar, *Essentials of Glycobiology: Chapter 5 Glycosylation Precursors*, Cold Spring Harbor Laboratory Press, New York, 3rd edn., 2017.
22 W. Van Breedam, S. Pöhlmann, H. W. Favoreel, R. J. de Groot and H. J. Nauwynck, *FEMS Microbiol. Rev.*, 2014, **38**, 598–632.
23 F. Y. Y. Tan, C. M. Tang and R. M. Exley, *Trends Biochem. Sci.*, 2015, **40**, 342–350.
24 R. L. Schnaar, *J. Leukoc. Biol.*, 2016, **99**, 825–838.
25 S. Chen and M. Fukuda, *Methods Enzymol.*, 2006, **416**, 371–380.
26 Y. Gu, W. Mi, Y. Ge, H. Liu, Q. Fan, C. Han, J. Yang, F. Han, X. Lu and W. Yu, *Cancer Res.*, 2010, **70**, 6344–6351.
27 K. Landsteiner and P. Levine, *J. Exp. Med.*, 1928, **47**, 757–75.
28 A. C. O'Sullivan, *Cellulose*, 1997, **4**, 173–207.
29 R. Minke and J. Blackwell, *J. Mol. Biol.*, 1978, **120**, 167–181.
30 Z. Shriver, I. Capila, G. Venkataraman and R. Sasisekharan, in *Handb Exp Pharmacol*, 2012, pp. 159–176.
31 P. D. Yurchenco, Y.-S. Cheng and J. C. Schittny, *J. Biol. Chem.*, 1990, **265**, 3981–3991.
32 D. J. Carey, D. M. Crumbling, R. C. Stahl and D. M. Evans, *J. Biol. Chem.*, 1990, **265**, 20627–20633.
33 R. G. LeBaron, A. Hook, J. D. Esko, S. Gay and M. Hook, *J. Biol. Chem.*, 1989, **264**, 7950–7956.
34 R. Aebersold, J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlén, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young and B. Zhang, *Nat. Chem. Biol.*, 2018, **14**, 206–214.
35 R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 1985, **54**, 631–664.

36 P. Van Den Steen, P. M. Rudd, R. A. Dwek and G. Opdenakker, *Crit. Rev. Biochem. Mol. Biol.*, 1998, **33**, 151–208.
37 Y. Ihara, Y. Inai, M. Ikezaki, I.-S. L. Matsui, S. Manabe and Y. Ito, in *Glycoscience: Biology and Medicine*, Springer Japan, 2014, pp. 1–8.
38 T. Kinoshita and M. Fujita, *J. Lipid Res.*, 2016, **57**, 6–24.
39 E. Lisowska and E. Jaskiewicz, Protein Glycosylation: An Overview, *LS*, John Wiley & Sons, Ltd, Chichester, UK, 2012.
40 C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča and A. Imberty, *Glycobiology*, 2006, **16**, 29–37.
41 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, **77**, 521–555.
42 G. S. Gupta, in *Animal Lectins: Form, Function and Clinical Applications*, Springer Vienna, Vienna, 2012, 29–56.
43 J.-X. Zhang, I. Braakman, K. E. S. Matlack and A. Helenius, *Mol. Biol. Cell*, 1997, **8**, 1943–1954.
44 X. Wu, R. A. Steet, O. Bohorov, J. Bakker, J. Newell, M. Krieger, L. Spaapen, S. Kornfeld and H. H. Freeze, *Nat. Med.*, 2004, **10**, 518–523.
45 T. Ju and R. D. Cummings, *Nature*, 2005, **437**, 1252.
46 J. Roth, C. Zuber, S. Park, I. Jang, Y. Lee, K. G. Kysela, V. Le Fourn, R. Santimaria, B. Guhl and J. W. Cho, *Mol. Cells*, 2010, **30**, 497–506.
47 S. S. Vembar and J. L. Brodsky, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 944–957.
48 S. Hakomori, *PNAS*, 2002, **99**, 10231–10233.
49 S. Hakomori, *Cancer Res.*, 1996, **56**, 5309–5318.
50 A. Kobata and J. Amano, *Immunol. Cell Biol.*, 2005, **83**, 429–439.
51 R. Kannagi, *Glycoconj. J.*, 2003, **20**, 353–364.
52 B. G. Ng and H. H. Freeze, *Trends Genet.*, 2018, **34**, 466–476.
53 H. J. Joshi, L. Hansen, Y. Narimatsu, H. H. Freeze, B. Henrissat, E. Bennett, H. H. Wandall, H. Clausen and K. T. Schjoldager, *Glycobiology*, 2018, **28**, 284–294.
54 J. Jaeken and R. Péanne, *J. Inherit. Metab. Dis.*, 2017, **40**, 569–586.
55 I. J. Chang, M. He and C. T. Lam, *Ann. Transl. Med.*, 2018, **6**, 477–477.
56 J. G. Leroy, M. W. Ho, M. C. Macbrinn, K. Zielke, J. Jacob and J. S. O'Brien, *Pediatr. Res.*, 1972, **6**, 752–757.
57 J. G. Leroy, J. W. Spranger, M. Feingold, J. M. Opitz and A. C. Crocker, *J. Pediatr.*, 1971, **79**, 360–365.
58 S. Hickman and E. F. Neufeld, *Biochem. Biophys. Res. Commun.*, 1972, **49**, 992–999.
59 M. R. Natowicz, M. M. Y. Chi, O. H. Lowry and W. S. Sly, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 4322–4326.
60 A. Varki and S. Kornfeld, *J. Biol. Chem.*, 1980, **255**, 10847–10858.
61 A. Hasilik, A. Waheed and K. von Figura, *Biochem. Biophys. Res. Commun.*, 1981, **98**, 761–767.
62 F. M. Platt, A. D'Azzo, B. L. Davidson, E. F. Neufeld and C. J. Tifft, *Nat. Rev. Dis. Prim.*, 2018, **4**, 27.
63 G. D'Angelo, S. Capasso, L. Sticco and D. Russo, *FEBS J.*, 2013, **280**, 6338–6353.
64 E. A. Miljan and E. G. Bremer, *Sci. Signal.*, 2002, **2002**, 15.
65 S. Hakomori, *Annu. Rev. Biochem.*, 1981, **50**, 733–764.
66 F. Malisan and R. Testi, *Exp. Gerontol.*, 2002, **37**, 1273–1282.
67 A. H. Merrill, *Chem. Rev.*, 2011, **111**, 6387–6422.
68 R. L. Schnaar and K. Kinoshita, in *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, 3rd edn., 2017, vol. 10.
69 T. Wennekes, R. J. B. H. N. Van Den Berg, R. G. Boot, G. A. Van Der Marel, H. S. Overkleeft and J. M. F. G. Aerts, *Angew. Chemie - Int. Ed.*, 2009, **48**, 8848–8869.
70 M. Sud, E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. H. Raetz, D. W. Russell and S. Subramaniam, *Nucleic Acids Res.*, 2007, **35**, D527–D532.
71 S. Degroote, J. Wolthoorn and G. Van Meer, *Semin. Cell Dev. Biol.*, 2004, **15**, 375–387.
72 W. Zheng, J. Kollmeyer, H. Symolon, A. Momin, E. Munter, E. Wang, S. Kelly, J. C. Allegood, Y. Liu, Q. Peng, H. Ramaraju, M. C. Sullards, M. Cabot and A. H. Merrill, *Biochim. Biophys. Acta - Biomembr.*, 2006, **1758**, 1864–1884.
73 S. T. Pruett, A. Bushnev, K. Hagedorn, M. Adiga, C. A. Haynes, M. C. Sullards, D. C. Liotta and A. H. Merrill, *J. Lipid Res.*, 2008, **49**, 1621–1639.
74 Y. Pewzner-Jung, S. Ben-Dor and A. H. Futerman, *J. Biol. Chem.*, 2006, **281**, 25001–25005.
75 A. Bosio, E. Binczek, M. M. Le Beau, A. A. Fernald and W. Stoffel, *Genomics*, 1996, **34**, 69–75.
76 S. Ichikawa, H. Sakiyama, G. Suzuki, K. I. Hidari and Y. Hirabayashi, *Proc. Natl. Acad. Sci.*, 1996,

93, 4638–4643.
77 J. A. Shayman and A. Abe, *Methods Enzymol.*, 2000, **311**, 42–49.
78 K. Hanada, K. Kumagai, S. Yasuda, Y. Miura, M. Kawano, M. Fukasawa and M. Nishijima, *Nature*, 2003, **426**, 803–809.
79 K. Hanada, K. Kumagai, N. Tomishige and M. Kawano, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, 2007, **1771**, 644–653.
80 J. P. Slotte and B. Ramstedt, *Eur. J. Lipid Sci. Technol.*, 2007, **109**, 977–981.
81 C. A. Lingwood, *Cold Spring Harb. Perspect. Biol.*, 2011, **3**, 4788
82 H. J. F. Maccioni, R. Quiroga and M. L. Ferrari, *J. Neurochem.*, 2011, **117**, 589-602
83 T. Kolter, R. L. Proia and K. Sandhoff, *J. Biol. Chem.*, 2002, **277**, 25859–25862.
84 M. Aureli, N. Loberto, V. Chigorno, A. Prinetti and S. Sonnino, *Neurochem. Res.*, 2011, **36**, 1636–1644.
85 T. Kolter and K. Sandhoff, *FEBS Lett.*, 2010, **584**, 1700–1712.
86 K. Kitatani, J. Idkowiak-Baldys and Y. A. Hannun, *Cell. Signal.*, 2008, **20**, 1010–1018.
87 H. J. F. Maccioni, R. Quiroga and W. Spessott, *FEBS Lett.*, 2011, **585**, 1691–1698.
88 T. Kolter and K. Sandhoff, *Annu. Rev. Cell Dev. Biol.*, 2005, **21**, 81–103.
89 R. M. Perera and R. Zoncu, *Annu. Rev. Cell Dev. Biol.*, 2016, **32**, 223–253.
90 K. Sandhoff and T. Kolter, *Clin. Chim. Acta*, 1997, **266**, 51–61.
91 C. G. Schuette, B. Pierstorff, S. Huettler and K. Sandhoff, *Glycobiology*, 2001, **11**, 81-90.
92 R. J. Tamargo, A. Velayati, E. Goldin and E. Sidransky, *Mol. Genet. Metab.*, 2012, **106**, 257–263.
93 W. W. Kallemeijn, K.-Y. Li, M. D. Witte, A. R. A. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. A. A. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. C. Walvoort, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts and H. S. Overkleeft, *Angew. Chemie Int. Ed.*, 2012, **51**, 12529–12533.
94 S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 386–394.
95 V. E. Ahn, P. Leyko, J.-R. Alattia, L. Chen and G. G. Privé, *Protein Sci.*, 2006, **15**, 1849–1857.
96 C. H. Hill, G. M. Cook, S. J. Spratley, S. Fawke, S. C. Graham and J. E. Deane, *Nat. Commun.*, 2018, **9**, 151.
97 A. Darmoise, P. Maschmeyer and F. Winau, *Advcances Immunol.*, 2010, **105**, 25–62.
98 H. Schulze and K. Sandhoff, *Cold Spring Harb. Perspect. Biol.*, 2011, **3**, 4804
99 K. A. Kretz, G. S. Carson, S. Morimoto, Y. Kishimoto, A. L. Fluharty and J. S. O'Brien, *Proc Natl Acad Sci USA*, 1990, **87**, 2541–2544.
100 A. Tylki-Szymańska, B. Czartoryska, M. T. Vanier, B. J. M. H. Poorthuis, J. A. E. Groener, A. Ługowska, G. Millat, A. M. Vaccaro and E. Jurkiewicz, *Clin. Genet.*, 2007, **72**, 538–542.
101 D. A. Brown and E. London, *Annu. Rev. Cell Dev. Biol.*, 1998, **14**, 111–136.
102 K. Simons and D. Toomre, *Nat. Rev. Mol. Cell Biol.*, 2000, **1**, 31–39.
103 G. van Meer and Q. Lisman, *J. Biol. Chem.*, 2002, **277**, 25855–25858.
104 J. M. F. G. Aerts, M. Artola, M. van Eijk, M. J. Ferraz and R. G. Boot, *Front. Cell Dev. Biol.*, 2019, **7**, 324.
105 Ü. Coskun, M. Grzybek, D. Drechsel and K. Simons, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 9044–9048.
106 K. Kabayama, T. Sato, K. Saito, N. Loberto, A. Prinetti, S. Sonnino, M. Kinjo, Y. Igarashi and J. I. Inokuchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13678–13683.
107 J. I. Inokuchi, M. Nagafuku, I. Ohno and A. Suzuki, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, 2015, **1851**, 98–106.
108 K. Hamamura, K. Furukawa, T. Hayashi, T. Hattori, J. Nakano, H. Nakashima, T. Okuda, H. Mizutani, H. Hattori, M. Ueda, T. Urano, K. O. Lloyd and K. Furukawa, *Proc. Natl. Acad. Sci.*, 2005, **102**, 11041–11046.
109 S. Yoshida, S. Fukumoto, H. Kawaguchi, S. Sato, R. Ueda and K. Furukawa, *Cancer Res.*, 2001, **61**, 4244–52.
110 O. Kovbasnjuk, R. Mourtazina, B. Baibakov, T. Wang, C. Elowsky, M. A. Choti, A. Kane and M. Donowitz, *Proc. Natl. Acad. Sci.*, 2005, **102**, 19087–19092.
111 S. Ngamukote, M. Yanagisawa, T. Ariga, S. Ando and R. K. Yu, *J. Neurochem.*, 2007, **103**, 2327–2341.
112 S. Ichikawa, N. Nakajo, H. Sakiyama and Y. Hirabayashi, *Proc. Natl. Acad. Sci.*, 1994, **91**, 2703–2707.

113 T. Yamashita, R. Wada, T. Sasaki, C. Deng, U. Bierfreund, K. Sandhoff and R. L. Proia, *Proc. Natl. Acad. Sci.*, 1999, **96**, 9142–9147.
114 G. Wu, Z.-H. Lu, N. Kulkarni, R. Amin and R. W. Ledeen, *Neurochem. Res.*, 2011, **36**, 1706–1714.
115 C. Tringali, I. Silvestri, F. Testa, P. Baldassari, L. Anastasia, R. Mortarini, A. Anichini, A. López-Requena, G. Tettamanti and B. Venerando, *BMC Cancer*, 2014, **14**, 560.
116 S. Tagami, J. I. Inokuchi, K. Kabayama, H. Yoshimura, F. Kitamura, S. Uemura, C. Ogawa, A. Ishii, M. Saito, Y. Ohtsuka, S. Sakaue and Y. Igarashi, *J. Biol. Chem.*, 2002, **277**, 3085–3092.
117 J. M. Aerts, R. Ottenhoff, A. S. Powlson, A. Grefhorst, M. Van Eijk, P. F. Dubbelhuis, J. Aten, F. Kuipers, M. J. Serlie, T. Wennekes, J. K. Sethi, S. O'Rahilly and H. S. Overkleeft, *Diabetes*, 2007, **56**, 1341–1349.
118 T. Ariga, M. P. McDonald and R. K. Yu, *J. Lipid Res.*, 2008, **49**, 1157–1175.
119 R. K. Yu, T. Ariga and C. Wakade, *Int. J. Alzheimers. Dis.*, 2011, **2011**, 1-14
120 H. Nakayama, M. Nagafuku, A. Suzuki, K. Iwabuchi and J. Inokuchi, *FEBS Lett.*, 2018, **592**, 3921–3942.
121 T. Zhang, A. A. De Waard, M. Wuhrer and R. M. Spaapen, *Front. Immunol.*, 2019, **10,** 90.
122 K. Hanada, *Jpn. J. Infect. Dis.*, 2005, **58**, 131–148.
123 E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe and P. H. Seeberger, *Chem. Biol.*, 2004, **11**, 875–881.
124 T. I. Klokk, S. Kavaliauskiene and K. Sandvig, *Cell. Mol. Life Sci.*, 2016, **73**, 1301–1316.
125 M. Nagafuku, K. Okuyama, Y. Onimaru, A. Suzuki, Y. Odagiri, T. Yamashita, K. Iwasaki, M. Fujiwara, M. Takayanagi, I. Ohno and J. I. Inokuchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E336–E342.
126 A. Kumar, N. Suryadevara, T. M. Hill, J. S. Bezbradica, L. V. Van Kaer and S. Joyce, *Front. Immunol.*, 2017, **8**, 1.
127 A. Mehta, M. Beck, A. Linhart, G. Sunder-Plassmann and U. Widmer, *History of lysosomal storage diseases: an overview*, Oxford PharmaGenesis, 2006.
128 E. Beutler, *Mol. Genet. Metab.*, 2006, **88**, 208–215.
129 C. de Duve, *Nat. Cell Biol.*, 2005, **7**, 847–849.
130 S. Kornfeld and I. Mellman, *Annu. Rev. Cell Biol.*, 1989, **5**, 483–525.
131 C. Duve, *Science (80-.).*, 1975, **189**, 186–194.
132 P. Saftig and J. Klumperman, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 623–635.
133 J. Rink, E. Ghigo, Y. Kalaidzidis and M. Zerial, *Cell*, 2005, **122**, 735–749.
134 C. Settembre, A. Fraldi, D. L. Medina and A. Ballabio, *Nat. Rev. Mol. Cell Biol.*, 2013, **14**, 283–296.
135 S. Ohkuma and B. Poole, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 3327–3331.
136 R. E. Lawrence and R. Zoncu, *Nat. Cell Biol.*, 2019, **21**, 133–142.
137 R. Schiffmann, J. B. Kopp, H. A. Austin, S. Sabnis, D. F. Moore, T. Weibel, J. E. Balow and R. O. Brady, *J. Am. Med. Assoc.*, 2001, **285**, 2743–2749.
138 T. Braulke and J. S. Bonifacino, *Biochim. Biophys. Acta - Mol. Cell Res.*, 2009, **1793**, 605–614.
139 P. Ghosh, N. M. Dahms and S. Kornfeld, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 202–212.
140 D. Reczek, M. Schwake, J. Schröder, H. Hughes, J. Blanz, X. Jin, W. Brondyk, S. Van Patten, T. Edmunds and P. Saftig, *Cell*, 2007, **131**, 770–783.
141 N. W. Andrews, *Trends Cell Biol.*, 2000, **10**, 316–321.
142 A. Rodríguez, P. Webster, J. Ortego and N. W. Andrews, *J. Cell Biol.*, 1997, **137**, 93–104.
143 G. Di Fruscio, A. Schulz, R. De Cegli, M. Savarese, M. Mutarelli, G. Parenti, S. Banfi, T. Braulke, V. Nigro and A. Ballabio, *Autophagy*, 2015, **11**, 928–938.
144 M. L. Katz and W. G. Robison, *Arch. Gerontol. Geriatr.*, 2002, **34**, 169–184.
145 F. M. Platt, B. Boland and A. C. van der Spoel, *J. Cell Biol.*, 2012, **199**, 723–734.
146 D. J. Metcalf, M. García-Arencibia, W. E. Hochfeld and D. C. Rubinsztein, *Exp. Neurol.*, 2012, **238**, 22–28.
147 P. J. Meikle, *JAMA*, 1999, **281**, 249.
148 A. Nowak and R. J. Desnick, *Mol. Genet. Metab.*, 2019, **126**, S110.
149 G. A. Grabowski, *Genet. Test.*, 1997, **1**, 5–12.
150 E. Young, K. Mills, P. Morris, A. Vellodi, P. Lee, S. Waldek and B. Winchester, *Acta Paediatr. Suppl*, 2005, **94**, 51–54.
151 A. Vellodi, *Br. J. Haematol.*, 2005, **128**, 413–431.
152 A. Mehta, *Eur. J. Intern. Med.*, 2006, **17**, S2–S5.
153 R. Kornreich and R. J. Desnick, *Hum. Mutat.*, 1993, **2**, 108–111.

154 V. Puri, R. Watanabe, M. Dominguez, X. Sun, C. L. Wheatley, D. L. Marks and R. E. Pagano, *Nat. Cell Biol.*, 1999, **1**, 386–388.
155 H. Hers, *Biochem. J.*, 1963, **86**, 11–16.
156 P. C. . Gaucher, Paris, 1882.
157 E. J. Parkinson-Lawrence, T. Shandala, M. Prodoehl, R. Plew, G. N. Borlace and D. A. Brooks, *Physiology*, 2010, **25**, 102–115.
158 K. Hashimoto, B. G. Gross and W. F. Lever, *J. Invest. Dermatol.*, 1965, **44**, 119–128.
159 H. Dempsey, M. W. Hartley, J. Carroll, J. Balint, R. E. Miller and W. B. Frommeyer, *Ann. Intern. Med.*, 1965, **63**, 1059–1068.
160 F. M. Platt, *Nature*, 2014, **510**, 68–75.
161 C. Mignot, A. Gelot, B. Bessières, F. Daffos, M. Voyer, F. Menez, C. Fallet Bianco, S. Odent, D. Le Duff, P. Loget, P. Fargier, J. Costil, P. Josset, J. Roume, M. Vanier, I. Maire and T. Billette de Villemeur, *Am. J. Med. Genet. Part A*, 2003, **120A**, 338–344.
162 C. M. Eng, J. Fletcher, W. R. Wilcox, S. Waldek, C. R. Scott, D. O. Sillence, F. Breunig, J. Charrow, D. P. Germain, K. Nicholls and M. Banikazemi, *J. Inherit. Metab. Dis.*, 2007, **30**, 184–192.
163 M. Jeyakumar, T. D. Butters, M. Cortina-Borja, V. Hunnam, R. L. Proia, V. H. Perry, R. A. Dwek and F. M. Platt, *Proc. Natl. Acad. Sci.*, 1999, **96**, 6388–6393.
164 M. Fuller, P. J. Meikle and J. J. Hopwood, *Epidemiology of lysosomal storage diseases: an overview*, Oxford PharmaGenesis, 2006.
165 G. E. Linthorst, A. C. Vedder, J. M. F. G. Aerts and C. E. M. Hollak, *Clin. Chim. Acta*, 2005, **353**, 201–203.
166 N. E. Brill, *Am J Med Sci*, 1901, **121**, 377–392.
167 S. U. Walkley, *Brain Pathol.*, 1998, **8**, 175–93.
168 H. N. Baris, I. J. Cohen and P. K. Mistry, *Pediatr Endocrinol Rev.*, 2014, **12**, 72–81.
169 C. Oberling and P. Woringer, *Rev Fr Pédiatr.*, 1927, **3**, 457–532.
170 L. Svennerholm, S. Dreborg, A. Erikson, C. G. Groth, P. O. Hillborg, G. Håkansson, O. Nilsson and E. Tibblin, *Prog. Clin. Biol. Res.*, 1982, **95**, 67–94.
171 H. Aghion, Paris University, 1934.
172 R. O. Brady, J. N. Kanfer and D. Shapiro, *J. Biol. Chem.*, 1965, **240**, 39–43.
173 R. O. Brady, J. N. Kanfer and D. Shapiro, *Biochem. Biophys. Res. Commun.*, 1965, **18**, 221–225.
174 R. O. Brady, J. N. Kanfer, R. M. Bradley and D. Shapiro, *J. Clin. Invest.*, 1966, **45**, 1112–1115.
175 P. A. Ockerman, *Biochim. Biophys. Acta*, 1968, **165**, 59–62.
176 N. J. Weinreb, R. O. Brady and A. L. Tappel, *Biochim. Biophys. Acta - Enzymol.*, 1968, **159**, 141–146.
177 P. K. Mistry, J. A. Weinthall and N. J. Weinreb, *Cinical Adv. Hematol. Oncol. H&O*, 2012, **10**, 1–16.
178 G. A. Grabowski, *Lancet*, 2008, **372**, 1263–1271.
179 P. K. Mistry, M. D. Cappellini, E. Lukina, H. Özsan, S. Mach Pascual, H. Rosenbaum, M. Helena Solano, Z. Spigelman, J. Villarrubia, N. P. Watman and G. Massenkeil, *Am. J. Hematol.*, 2011, **86**, 110–115.
180 L. Peleg, A. Frisch, B. Goldman, M. Karpaty, R. Narinsky, S. Bronstein and M. Frydman, *Eur. J. Hum. Genet.*, 1998, **6**, 185–186.
181 B. J. H. M. Poorthuis, R. A. Wevers, W. J. Kleijer, J. E. M. Groener, J. G. N. de Jong, S. van Weely, K. E. Niezen-Koning and O. P. van Diggelen, *Hum. Genet.*, 1999, **105**, 151–156.
182 T. M. Cox and J. P. Schofiled, *Baillieres. Clin. Haematol.*, 1997, **10**, 657–689.
183 M. Jmoudiak and A. H. Futerman, *Br. J. Haematol.*, 2005, **129**, 178–188.
184 G. M. Pastores, *Baillieres. Clin. Haematol.*, 1997, **10**, 739–49.
185 G. A. Grabowski and H. Zhao, *C. Cell. Mol. Life Sci.*, 2002, **59**, 694–707.
186 E. H. Kolodny, M. D. Ullman, H. J. Mankin, S. S. Raghavan, J. Topol and J. L. Sullivan, *Prog. Clin. Biol. Res.*, 1982, **95**, 33–65.
187 J. Stirnemann, N. Belmatoug, F. Camou, C. Serratrice, R. Froissart, C. Caillaud, T. Levade, L. Astudillo, J. Serratrice, A. Brassier, C. Rose, T. Billette de Villemeur and M. Berger, *Int. J. Mol. Sci.*, 2017, **18**, 441.
188 J. Goldblatt, *Syndr. Mon. J. Med. Genet.*, 1988, **25**, 415–418.
189 A. H. Futerman and A. Zimran, *Gaucher Disease*, CRC Press, Boca Raton, 1st edn., 2006.
190 G. A. Grabowski, N. Leslie and R. Wenstrup, *Blood Rev.*, 1998, **2**, 115–133.
191 E. Beutler and G. A. Grabowski, *No Title*, McGraw-Hill Medical, New York, 8th edn., 2001.
192 E. H. Davies, R. Surtees, C. DeVile, I. Schoon and A. Vellodi, *J. Inherit. Metab. Dis.*, 2007, **30**, 768–

782.
193 A. Erikson, B. Bembi and R. Schiffmann, *Baillieres. Clin. Haematol.*, 1997, **10**, 711–23.
194 D. L. Stone, N. Tayebi, E. Orvisky, B. K. Stubblefield, V. Madike and E. Sidransky, *Hum. Mutat.*, 2000, **15**, 181–188.
195 E. Sidransky, *Mol. Genet. Metab.*, 2004, **83**, 6–15.
196 A. Drukker, M. I. Sacks and S. Gatt, *Pediatrics*.
197 A. K. Simalti and H. Sethi, *Med. J. Armed Forces India*, 2017, **73**, 197–199.
198 K. Lui, C. Commens, R. Choong and R. Jaworski, *Arch. Dis. Child.*, 1988, **63**, 854–856.
199 M. Owada, T. Sakiyama and T. Kitagawa, *Pediatr. Res.*, 1977, **11**, 641–646.
200 E. Sidransky, D. M. Sherer and E. I. Ginns, *Pediatr. Res.*, 1992, **32**, 494–498.
201 A. Fujimoto, N. Tayebi and E. Sidransky, *Am. J. Med. Genet.*, 1995, **59**, 356–358.
202 F. S. Finn, M. Zhang, S.-H. Chen and R. C. Scott, *Am. J. Med. Genet. Part A*, 2000, **91**, 222–226.
203 A. Nagral, *J. Clin. Exp. Hepatol.*, 2014, **4**, 37–50.
204 A. Tylki-Szymańska, A. Vellodi, A. El-Beshlawy, J. A. Cole and E. Kolodny, *J. Inherit. Metab. Dis.*, 2010, **33**, 339–346.
205 N. Dahl, M. Lagerstrom, A. Erikson, U. Petterson, M. Lagerström, A. Erikson and U. Pettersson, *Am. J. Hum. Genet.*, 1990, **47**, 275–278.
206 A. Abrahamov, D. Elstein, A. Zimran, B. Farber, Y. Glaser, V. Gross-Tsur, I. Hadas-Halpern, S. Ronen, M. Horowitz and M. Tafakjdi, *Lancet*, 1995, **346**, 1000–1003.
207 R. O. Brady, N. W. Barton and G. A. Grabowski, *Arch. Neurol.*, 1993, **50**, 1212–1224.
208 H. Christomanou, A. Aignesberger and R. P. Linke, *Biol. Chem. Hoppe. Seyler.*, 1986, **367**, 879–890.
209 R. A. Barneveld, W. Keijzer, F. P. W. Tegelaers, E. I. Ginns, A. Geurts Van Kessel, R. O. Brady, J. A. Barranger, J. M. Tager, H. Galjaard, A. Westerveld and A. J. J. Reuser, *Hum Genet*, 1983, **64**, 227–231.
210 E. I. Ginns, P. V. Choudary, B. M. Martin, S. Winfield, B. Stubblefield, J. Mayor, D. Merkle-Lehman, G. J. Murray, L. A. Bowers and J. A. Barranger, *Biochem. Biophys. Res. Commun.*, 1984, **123**, 574–580.
211 E. I. Ginns, P. V. Choudary, S. Tsuji, B. Martin, B. Stubblefield, J. Sawyer, J. Hozier and J. A. Barranger, *Proc. Natl. Acad. Sci.*, 1985, **82**, 7101–7105.
212 S. Tsuji, P. V Choudary, B. M. Martin, S. Winfield, J. A. Barranger and E. I. Ginns, *J. Biol. Chem.*, 1986, **261**, 50–53.
213 O. Reiner, M. Wigderson and M. Horowitz, *DNA*, 1988, **7**, 107–116.
214 J. Sorge, C. West, B. Westwood and E. Beutler, *Proc. Natl. Acad. Sci.*, 1985, **82**, 7289–7293.
215 M. Horowitz, S. Wilder, Z. Horowitz, O. Reiner, T. Gelbart and E. Beutler, *Genomics*, 1989, **4**, 87–96.
216 J. Sorge, E. Gross, C. West and E. Beutler, *J. Clin. Invest.*, 1990, **86**, 1137–1141.
217 A. Zimran, J. Sorge, E. Gross, M. Kubitz, C. West and E. Beutler, *J. Clin. Invest.*, 1990, **85**, 219–222.
218 P. D. Stenson, M. Mort, E. V. Ball, K. Evans, M. Hayden, S. Heywood, M. Hussain, A. D. Phillips and D. N. Cooper, *Hum. Genet.*, 2017, 136, 665–677.
219 G. A. Grabowski and M. Horowitz, *Baillieres. Clin. Haematol.*, 1997, **10**, 635–656.
220 M. E. Grace, P. Ashton-Prolla, G. M. Pastores, A. Soni and R. J. Desnick, *J. Clin. Invest.*, 1999, **103**, 817–823.
221 A. Zimran, E. Gross, C. West, J. Sorge, M. Kubitz and E. Beutler, *Lancet*, 1989, **334**, 349–352.
222 M. Horowitz and A. Zimran, *Hum. Mutat.*, 1994, **3**, 1–11.
223 P. K. Mistry and A. Abrahamov, *Baillieres. Clin. Haematol.*, 1997, **10**, 817–838.
224 N. Eyal, S. Wilder and M. Horowitz, *Gene*, 1990, **96**, 277–283.
225 S. Tsuji, B. M. Martin, J. A. Barranger, B. K. Stubblefield, M. E. LaMarca and E. I. Ginns, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 2349–52.
226 M. E. Grace, P. N. Gravesll, F. I. Smithll and G. A. Grabowskis, *J. Biol. Chem.*, 1990, **265**, 6827–6835.
227 S. Tsuji, P. V Choudary, B. M. Martin, B. K. Stubblefield, J. A. Mayor, J. A. Barranger and E. I. Ginns, *N. Engl. J. Med.*, 1987, **316**, 570–5.
228 J. F. Desforges and E. Beutler, *N. Engl. J. Med.*, 1991, **325**, 1354–1360.
229 E. Beutler, T. Gelbart, W. Kuhl, J. Sorge and C. West, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 10544–7.
230 G. S. He and G. A. Grabowski, *Am. J. Hum. Genet.*, 1992, **51**, 810–20.

231 E. Beutler and T. Gelbart, *Br. J. Haematol.*, 1993, **85**, 401–5.
232 P. Mistry, *Lancet*, 1995, **346**, 982–983.
233 M. Balwani, L. Fuerstman, R. Kornreich, L. Edelmann and R. J. Desnick, *Arch. Intern. Med.*, 2010, **170**, 1463–9.
234 I. Ron and M. Horowitz, *Hum. Mol. Genet.*, 2005, **14**, 2387–2398.
235 G. A. Grabowski, A. Zimran and H. Ida, *Am. J. Hematol.*, 2015, **90**, S12–S18.
236 W. Dauer and S. Przedborski, *Neuron*, 2003, **39**, 889–909.
237 A. Lee and R. M. Gilbert, *Neurol. Clin.*, 2016, **34**, 955–965.
238 E. Sidransky, M. A. Nalls, J. O. Aasly, J. Aharon-Peretz, G. Annesi, E. R. Barbosa, A. Bar-Shira, D. Berg, J. Bras, A. Brice, C.-M. Chen, L. N. Clark, C. Condroyer, E. V. De Marco, A. Dürr, M. J. Eblan, S. Fahn, M. J. Farrer, H.-C. Fung, Z. Gan-Or, T. Gasser, R. Gershoni-Baruch, N. Giladi, A. Griffith, T. Gurevich, C. Januario, P. Kropp, A. E. Lang, G.-J. Lee-Chen, S. Lesage, K. Marder, I. F. Mata, A. Mirelman, J. Mitsui, I. Mizuta, G. Nicoletti, C. Oliveira, R. Ottman, A. Orr-Urtreger, L. V. Pereira, A. Quattrone, E. Rogaeva, A. Rolfs, H. Rosenbaum, R. Rozenberg, A. Samii, T. Samaddar, C. Schulte, M. Sharma, A. Singleton, M. Spitz, E.-K. Tan, N. Tayebi, T. Toda, A. R. Troiano, S. Tsuji, M. Wittstock, T. G. Wolfsberg, Y.-R. Wu, C. P. Zabetian, Y. Zhao and S. G. Ziegler, *N. Engl. J. Med.*, 2009, **361**, 1651–1661.
239 N. Tayebi, M. Callahan, V. Madike, B. K. Stubblefield, E. Orvisky, D. . Krasenwich, J. J. Fillano and E. Sidransky, *Mol. Genet. Metab.*, 2001, **73**, 313–321.
240 B. Rosenbloom, M. Balwani, J. M. Bronstein, E. Kolodny, S. Sathe, A. R. Gwosdow, J. S. Taylor, J. A. Cole, A. Zimran and N. J. Weinreb, *Blood Cells, Mol. Dis.*, 2011, **46**, 95–102.
241 M. Avenali, F. Blandini and S. Cerri, *Front. Aging Neurosci.*, 2020, **0**, 97.
242 L. Alvarez-Erviti, Y. Seow, A. H. Schapira, C. Gardiner, I. L. Sargent, M. J. A. Wood and J. M. Cooper, *Neurobiol. Dis.*, 2011, **42**, 360–367.
243 E.-J. Bae, N. Y. Yang, C. Lee, H.-J. Lee, S. Kim, S. P. Sardi and S.-J. Lee, *Exp. Mol. Med.*, 2015, **47**, e153–e153.
244 J. R. Mazzulli, Y. H. Xu, Y. Sun, A. L. Knight, P. J. McLean, G. A. Caldwell, E. Sidransky, G. A. Grabowski and D. Krainc, *Cell*, 2011, **146**, 37–52.
245 M. Toffoli, L. Smith and A. H. V. Schapira, *J. Neurochem.*, 2020, **154**, 11–24.
246 K. E. Murphy, A. M. Gysbers, S. K. Abbott, N. Tayebi, W. S. Kim, E. Sidransky, A. Cooper, B. Garner and G. M. Halliday, *Brain*, 2014, **137**, 834–848.
247 T. L. Yap, A. Velayati, E. Sidransky and J. C. Lee, *Mol. Genet. Metab.*, 2013, **108**, 56–64.
248 M. E. Gegg, L. Sweet, B. H. Wang, L. S. Shihabuddin, S. P. Sardi and A. H. V. Schapira, *Mov. Disord.*, 2015, **30**, 1085–1089.
249 J. Aharon-Peretz, S. Badarny, H. Rosenbaum and R. Gershoni-Baruch, *Neurology*, 2005, **65**, 1460–1461.
250 C. Sato, A. Morgan, A. E. Lang, S. Salehi-Rad, T. Kawarai, Y. Meng, P. N. Ray, L. A. Farrer, P. St George-Hyslop and E. Rogaeva, *Mov. Disord.*, 2005, **20**, 367–370.
251 O. Landgren, I. Turesson, G. Gridley and N. E. Caporaso, *Arch. Intern. Med.*, 2007, **167**, 1189–1194.
252 M. de Fost, S. Vom Dahl, G. J. Weverling, N. Brill, S. Brett, D. Haussinger and C. E. M. Hollak, *Blood Cells, Mol. Dis.*, 2006, **36**, 53–58.
253 M. J. Allen, B. J. Myer, A. M. Khokher, N. Rushton and T. M. Cox, *QJM*, 1997, **90**, 19–25.
254 V. Barak, M. Acker, B. Nisman, I. Kalickman, A. Abrahamov, A. Zimran and S. Yatziv, *Eur. Cytokine Netw.*, 1999, **10**, 205–10.
255 L. Astudillo, N. Therville, C. Colacios, B. Ségui, N. Andrieu-Abadie and T. Levade, *Biochimie*, 2016, **125**, 267–280.
256 J. A. Castaneda, M. J. Lim, J. D. Cooper and D. A. Pearce, *Acta Neuropathol.*, 2008, **115**, 159–74.
257 H. Du, M. Heur, M. Duanmu, G. A. Grabowski, D. Y. Hui, D. P. Witte, J. Mishra, M. Heur, M. Duanmu, G. A. Grabowski, D. Y. Hui, D. P. Witte and J. Mishra Lysosomal, *J. Lipid Res.*, 2001, **42**, 489–500.
258 A. P. Bussink, M. van Eijk, G. H. Renkema, J. M. Aerts and R. G. Boot, *Int. Rev. Cytol.*, 2006, **252**, 71–128.
259 L. A. Boven, M. van Meurs, R. G. Boot, A. Mehta, L. Boon, J. M. Aerts and J. D. Laman, *Am. J. Clin. Pathol.*, 2004, **122**, 359–69.
260 manoj K. Pandey, T. A. Burrow, R. Rani, L. J. Martin, D. Witte, K. D. Setchell, mary A. Mckay, A. F. Magnusen, W. Zhang, B. Liou, J. Köhl and G. A. Grabowski, *Nature*, 2017, **543**, 108–112.
261 M. Kolev, G. Le Friec and C. Kemper, *Nat. Rev. Immunol.*, 2014, **14**, 811–820.

262 H. D. Manthey, T. M. Woodruff, S. M. Taylor and P. N. Monk, *Int. J. Biochem. Cell Biol.*, 2009, **41**, 2114–2117.
263 A. Klos, E. Wende, K. J. Wareham and P. N. Monk, *Pharmacol. Rev.*, 2013, 65, 500–543.
264 P. K. Mistry, J. Liu, M. Yang, T. Nottoli, J. McGrath, D. Jain, K. Zhang, J. Keutzer, W.-L. Chuang, W.-L. Chuang, W. Z. Mehal, H. Zhao, A. Lin, S. Mane, X. Liu, Y. Z. Peng, J. H. Li, M. Agrawal, L.-L. Zhu, H. C. Blair, L. J. Robinson, J. Iqbal, L. Sun and M. Zaidi, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19473–8.
265 K. Suzuki, *Neurochem Res*, 1998, **23**, 251–259.
266 E. Orvisky, K. Park, J, M. E. LaMarca, E. I. Ginns, B. M. Martin, N. Tayebi and E. Sidransky, *Mol. Genet. Metab.*, 2002, **76**, 262–270.
267 O. Nilsson and L. Svennerholm, *J. Neurochem.*, 1982, **39**, 709–718.
268 E. Orvisky, E. Sidransky, C. E. McKinney, M. E. Lamarca, R. Samimi, D. Krasnewich, B. M. Martin and E. I. Ginns, *Pediatr. Res.*, 2000, **48**, 233–7.
269 P. K. Mistry, J. Liu, L. Sun, W.-L. Chuang, T. Yuen, R. Yang, P. Lu, K. Zhang, J. Li, J. Keutzer, A. Stachnik, A. Mennone, J. L. Boyer, D. Jain, R. O. Brady, M. I. New and M. Zaidi, *PNAS*, 2014, **111**, 4934–4939.
270 N. Brakch, O. Dormond, S. Bekri, D. Golshayan, M. Correvon, L. Mazzolai, B. Steinmann and F. Barbey, *Eur. Heart J.*, 2010, **31**, 67–76.
271 Y.-P. Wu, K. Mizugishi, M. Bektas, R. Sandhoff and R. L. Proia, *Hum. Mol. Genet.*, 2008, **17**, 2257–2264.
272 J. P. Kampine, R. O. Bardy, J. N. Kanfer, M. Feld and D. Shapiro, *Science (80-.).*, 1967, **155**, 86–88.
273 B. M. Turner and K. Hirschhorn, *Am. J. Hum. Genet.*, 1978, **30**, 346–58.
274 G. A. Grabowski, J. Goldblatt, T. Dinur, J. Kruse, L. Svennerholm, S. Gatt, R. J. Desnick, J. M. Opitz and J. F. Reynolds, *Am. J. Med. Genet.*, 1985, **21**, 529–549.
275 S. S. Raghavan, J. Topol and E. H. Kolodny, *Am. J. Hum. Genet.*, 1980, **32**, 158–73.
276 J. R. Mazzulli, F. Zunke, T. Tsunemi, N. J. Toker, S. Jeon, L. F. Burbulla, S. Patnaik, E. Sidransky, J. J. Marugan, C. M. Sue and D. Krainc, *J. Neurosci.*, 2016, **36**, 7693–7706.
277 C.-L. L. Kuo, W. W. Kallemeijn, L. T. Lelieveld, M. Mirzaian, I. Zoutendijk, A. Vardi, A. H. Futerman, A. H. Meijer, H. P. Spaink, H. S. Overkleeft, J. M. F. G. Aerts and M. Artola, *FEBS J.*, 2019, **286**, 584–600.
278 J. Benz, A. C. Rufer, S. Huber, A. Ehler, M. Hug, A. Topp, W. Guba, E. C. Hofmann, R. Jagasia, R. M. R. Sarmiento and R. M. Rodríguez Sarmiento, *Angew. Chemie Int. Ed.*, 2021, **60**, 5436–5442.
279 S. Little, in *Current Protocols in Human Genetics*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2001.
280 P. K. Mistry, S. J. Smith, M. Ali, T. M. Cox, C. S. R. Hatton and N. McIntyre, *Lancet*, 1992, **339**, 889–892.
281 M. Horowitz, G. Tzuri, N. Eyal, A. Berebi, E. H. Kolodny, R. O. Brady, N. W. Barton, A. Abrahamov and A. Zimran, *Am. J. Hum. Genet.*, 1993, **53**, 921–30.
282 E. Beutler, T. Gelbart and C. West, *Clin. Chim. Acta*, 1990, **194**, 161–166.
283 S. Yoshida, J. Kido, S. Matsumoto, K. Momosaki, H. Mitsubuchi, T. Shimazu, K. Sugawara, F. Endo and K. Nakamura, *Pediatr. Int.*, 2016, **58**, 946–949.
284 K. S. Hruska, M. E. LaMarca, C. R. Scott and E. Sidransky, *Hum. Mutat.*, 2008, **29**, 567–583.
285 C. E. Hollak, S. van Weely, M. H. van Oers and J. M. Aerts, *J. Clin. Invest.*, 1994, **93**, 1288–92.
286 G. H. Renkema, R. G. Boot, A. Strijland, W. E. Donker-Koopman, M. van den Berg, A. O. Muijsers and J. M. Aerts, *Eur. J. Biochem.*, 1997, **244**, 279–85.
287 M. J. Ferraz, W. W. Kallemeijn, M. Mirzaian, D. H. Moro, A. Marques, P. Wisse, R. G. Boot, L. I. Willems, H. S. Overkleeft and J. M. F. G. Aerts, *Biochim. Biophys. Acta*, 2014, **1841**, 811–825.
288 Y. Guo, W. He, A. M. Boer, R. A. Wevers, A. M. de Bruijn, J. E. M. Groener, C. E. M. Hollak, J. M. F. G. Aerts, H. Galjaard and O. P. van Diggelen, *J. Inherit. Metab. Dis.*, 1995, **18**, 717–722.
289 A. P. Bussink, M. Verhoek, J. Vreede, K. Ghauharali-Van Der Vlugt, W. E. Donker-Koopman, R. R. Sprenger, C. E. Hollak, J. M. F. G. Aerts and R. G. Boot, *FEBS J.*, 2009, **276**, 5678–5688.
290 K.-L. Chang, W.-L. Hwu, H.-Y. Yeh, N.-C. Lee and Y.-H. Chien, *Blood Cells, Mol. Dis.*, 2010, **44**, 38–40.
291 S. A. Islam, M. F. Ling, J. Leung, W. G. Shreffler and A. D. Luster, *J. Exp. Med.*, 2013, **210**, 1889–1898.
292 R. G. Boot, *Blood*, 2004, **103**, 33–39.
293 F. Bonella and U. Costabel, *Semin. Respir. Crit. Care Med.*, 2014, **35**, 181–200.
294 D. Elstein, B. Mellgard, Q. Dinh, L. Lan, Y. Qiu, C. Cozma, S. Eichler, T. Böttcher and A. Zimran, *Mol.*

Genet. Metab., 2017, **122**, 113–120.
295 N. Dekker, L. van Dussen, C. E. M. Hollak, H. Overkleeft, S. Scheij, K. Ghauharali, M. J. van Breemen, M. J. Ferraz, J. E. M. Groener, M. Maas, F. A. Wijburg, D. Speijer, A. Tylki-Szymanska, P. K. Mistry, R. G. Boot and J. M. Aerts, *Blood*, 2011, **118**, 118–127.
296 A. Rolfs, A.-K. Giese, U. Grittner, D. Mascher, D. Elstein, A. Zimran, T. Böttcher, J. Lukas, R. Hübner, U. Gölnitz, A. Röhle, A. Dudesek, W. Meyer, M. Wittstock and H. Mascher, *PLoS One*, 2013, **8**, 79732.
297 R. O. Brady, *N. Engl. J. Med.*, 1966, **275**, 312–318.
298 M. Beck, *Dev. Med. Child Neurol.*, 2018, **60**, 13–18.
299 P. G. Pentchev, R. O. Brady, S. R. Hibbert, A. E. Gal and D. Shapiro, *J. Biol. Chem.*, 1973, **248**, 5256–5261.
300 R. O. Brady, P. G. Pentchev, A. E. Gal, S. R. Hibbert and A. S. Dekaban, *N. Engl. J. Med.*, 1974, **291**, 989–993.
301 R. O. Brady, *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, 2003, **358**, 915–919.
302 F. S. Furbish, H. E. Blairt, J. Shiloacht, P. G. Pentchev and R. 0 Brady, *Med. Sci.*, 1977, **74**, 3560–3563.
303 R. O. Brady, *Baillieres. Clin. Haematol.*, 1997, **10**, 621–634.
304 P. D. Stahl, J. S. Rodman, M. J. Miller and P. H. Schlesinger, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 1399–403.
305 S. Takasaki, G. J. Murray, F. S. Furbish, R. O. Brady, J. A. Barranger and A. Kobata, *J. Biol. Chem.*, 1984, **259**, 10112–10117.
306 G. J. Murray, *Methods Enzymol.*, 1987, **149**, 25–42.
307 R. O. Brady and F. S. Furbish, in *Membranes and transport*, Plenum Press, New York, Ed A. N. M., 1982, pp. 587-592.
308 N. W. Barton, R. O. Brady, J. M. Dambrosia, A. M. Di Bisceglie, S. H. Doppelt, S. C. Hill, H. J. Mankin, G. J. Murray, R. I. Parker and C. E. Argoff, *N. Engl. J. Med.*, 1991, **324**, 1464–70.
309 P. B. Deegan and T. M. Cox, *Drug Des. Devel. Ther.*, 2012, **6**, 81–106.
310 S. M. Richards, T. A. Olsen and J. M. McPherson, *Blood*, 1993, **82**, 1402–1409.
311 Y. Cohen, D. Elstein, A. Abrahamov, H. Hirsch and A. Zimran, *Am. J. Hematol.*, 1994, **47**, 235–236.
312 C. Serratrice, S. Carballo, J. Serratrice and J. Stirnemann, *Core Evid.*, 2016, **11**, 37–47.
313 G. A. Grabowski, N. W. Barton, G. . Pastores, J. M. Dambrosia, T. K. Benerjee, M. A. McKee, C. Parker, R. Schiffmann, S. C. Hill and R. O. Brady, *Ann Intern Med.*, 1995, **122**, 33–39.
314 A. Zimran, D. Elstein, E. Levy-Lahad, S. Zevin, I. Hadas-Halpern, A. Abrahamov, Y. Bar-Ziv, A. Schwartz and J. Foldes, *Lancet*, 1995, **345**, 1479–1480.
315 G. A. Grabowski, *Drugs*, 1996, **52**, 159–167.
316 G. . Pastores, A. R. Sibille and G. A. Grabowski, *Blood*, 1993, **82**, 408–416.
317 D. I. Rosenthal, S. H. Doppelt, H. J. Mankin, J. M. Dambrosia, R. J. Xavier, K. A. McKusick, B. R. Rosen, J. Baker, L. T. Niklason and S. C. Hill, *Pediatrics*, 1995, **96**, 629–37.
318 P. K. Mistry, S. Davies, A. Corfield, A. K. Dixon and T. M. Cox, *Q. J. Med.*, 1992, **83**, 541–6.
319 P. K. Mistry, E. P. Wraight and T. M. Cox, *Lancet*, 1996, **348**, 1555–1559.
320 B. Brumshtein, P. Salinas, B. Peterson, V. Chan, I. Silman, J. L. Sussman, P. J. Savickas, G. S. Robinson and A. H. Futerman, *Glycobiology*, 2010, **20**, 24–32.
321 A. Zimran, G. M. Pastores, A. Tylki-Szymanska, D. A. Hughes, D. Elstein, R. Mardach, C. Eng, L. Smith, M. Heisel-Kurth, J. Charrow, P. Harmatz, P. Fernhoff, W. Rhead, N. Longo, P. Giraldo, J. A. Ruiz, D. Zahrieh, E. Crombez and G. A. Grabowski, *Am. J. Hematol.*, 2013, **88**, 172–178.
322 H. Ben Turkia, D. E. Gonzalez, N. W. Barton, A. Zimran, M. Kabra, E. A. Lukina, P. Giraldo, I. Kisinovsky, A. Bavdekar, M.-F. Ben Dridi, N. Gupta, P. S. Kishnani, E. K. Sureshkumar, N. Wang, E. Crombez, K. Bhirangi and A. Mehta, *Am. J. Hematol.*, 2013, **88**, 179–184.
323 Y. Shaaltiel, D. Bartfeld, S. Hashmueli, G. Baum, E. Brill-Almon, G. Galili, O. Dym, S. A. Boldin-Adamsky, I. Silman, J. L. Sussman, A. H. Futerman and D. Aviezer, *Plant Biotechnol. J.*, 2007, **5**, 579–590.
324 V. Gomord, A.-C. Fitchette, L. Menu-Bouaouiche, C. Saint-Jore-Dupas, C. Plasson, D. Michaud and L. Faye, *Plant Biotechnol. J.*, 2010, **8**, 564–587.
325 D. Aviezer, E. Brill-Almon, Y. Shaaltiel, S. Hashmueli, D. Bartfeld, S. Mizrachi, Y. Liberman, A. Freeman, A. Zimran and E. Galun, *PLoS One*, 2009, **4**, 4792.
326 X. He, J. D. Galpin, M. B. Tropak, D. Mahuran, T. Haselhorst, M. von Itzstein, D. Kolarich, N. H. Packer, Y. Miao, L. Jiang, G. A. Grabowski, L. A. Clark and A. R. Kermode, *Glycobiology*, 2012, **22**,

492–503.
327 A. Zimran, E. Brill-Almon, R. Chertkoff, M. Petakov, F. Blanco-Favela, E. T. Munoz, S. E. Solorio-Meza, D. Amato, G. Duran, F. Giona, R. Heitner, H. Rosenbaum, P. Giraldo, A. Mehta, G. Park, M. Phillips, D. Elstein, G. Altarescu, M. Szleifer, S. Hashmueli and D. Aviezer, *Blood*, 2012, **119**, 4577–4577.
328 A. Zimran, G. Durán, A. Mehta, P. Giraldo, H. Rosenbaum, F. Giona, D. J. Amato, M. Petakov, E. T. Muñoz, S. E. Solorio-Meza, P. A. Cooper, S. Varughese, R. Chertkoff and E. Brill-Almon, *Am. J. Hematol.*, 2016, **91**, 656–660.
329 L. van Dussen, A. Zimran, E. M. Akkerman, J. M. F. G. Aerts, M. Petakov, D. Elstein, H. Rosenbaum, D. Aviezer, E. Brill-Almon, R. Chertkoff, M. Maas and C. E. M. Hollak, *Blood Cells, Mol. Dis.*, 2013, **50**, 206–211.
330 H. C. Andersson, J. Charrow, P. Kaplan, P. Mistry, G. M. Pastores, A. Prakesh-Cheng, B. E. Rosenbloom, C. R. Scott, R. S. Wappner and N. J. Weinreb, *Genet. Med.*, 2005, **7**, 105–110.
331 E. Beutler, A. Kay, A. Saven, P. Garver, D. Thurston, A. Dawson and B. Rosenbloom, *Am. Soc. Hematol.*, 1991, **78**, 1183–1189.
332 A. B. Engelberg, A. S. Kesselheim and J. Avorn, *N. Engl. J. Med.*, 2009, **361**, 1917–1919.
333 A. Zimran, D. Elstein, R. Kannai, S. Zevin, I. Hadas-Halpern, E. Levy-Lahad, Y. Cohen, M. Horowitz and A. Abrahamov, *Am. J. Med.*, 1994, **97**, 3–13.
334 R. Schiffmann, M. P. Heyes, J. M. Aerts, J. M. Dambrosia, M. C. Patterson, T. DeGraba, C. C. Parker, G. C. Zirzow, K. Oliver, G. Tedeschi, R. O. Brady and N. W. Barton, *Ann. Neurol.*, 1997, **42**, 613–621.
335 G. C. Zirzow, O. A. Sanchez, G. J. Murray, R. O. Brady and E. H. Oldfield, *Neurochem. Res.*, 1999, **24**, 301–305.
336 R. R. Vunnam and N. S. Radin, *Chem. Phys. Lipids*, 1980, **26**, 265–78.
337 N. Belmatoug, M. Di Rocco, C. Fraga, P. Giraldo, D. Hughes, E. Lukina, P. Maison-Blanche, M. Merkel, C. Niederau, U. Plöckinger, J. Richter, T. M. Stulnig, S. vom Dahl and T. M. Cox, *Eur. J. Intern. Med.*, 2017, **37**, 25–32.
338 R. M. Shawky and S. M. Elsayed, *Egypt. J. Med. Hum. Genet.*, 2016, **17**, 281–285.
339 D. Elstein, C. E. M. Hollak, J. M. F. G. Aerts, S. van Weely, M. Maas, T. M. Cox, R. H. Lachman, M. Hrebicek, F. M. Platt, T. D. Butters, R. A. Dwek and A. Zimran, *J. Inherit.Metab. Dis.*, 2004, **27**, 757–766.
340 S. Basu, B. Kaufman and S. Roseman, *J. Biol. Chem.*, 1968, **243**, 5802–5804.
341 H. Zhao, M. Przybylska, I.-H. Wu, J. Zhang, C. Siegel, S. Komarnitsky, N. S. Yew and S. H. Cheng, *Diabetes*, 2007, **56**, 1210–8.
342 G. . Pastores, N. L. Barnett and E. H. Kolodny, *Clin. Ther.*, 2005, **27**, 1215–1227.
343 T. Wennekes, A. J. Meijer, A. K. Groen, R. G. Boot, J. E. Groener, M. van Eijk, R. Ottenhoff, N. Bijl, K. Ghauharali, H. Song, T. J. O'Shea, H. Liu, N. Yew, D. Copeland, R. J. van den Berg, G. A. van der Marel, H. S. Overkleeft and J. M. Aerts, *J. Med. Chem.*, 2010, **53**, 689–698.
344 F. M. Platt, M. Jeyakumar, U. Andersson, T. Heare, R. A. Dwek and T. D. Butters, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 2003, **358**, 947–54.
345 T. Cox, R. Lachmann, C. Hollak, J. Aerts, S. van Weely, M. Hrebícek, F. Platt, T. Butters, R. Dwek, C. Moyses, I. Gow, D. Elstein and A. Zimran, *Lancet*, 2000, **355**, 1481–5.
346 R. Heitner, D. Elstein, J. Aerts, S. van Weely and A. Zimran, *Blood Cells. Mol. Dis.*, 1991, **28**, 127–33.
347 R. Schiffmann, E. J. FitzGibbon, C. Harris, C. DeVile, E. H. Davies, L. Abel, I. N. Van Schaik, W. S. Benko, M. Timmons, M. Ries and A. Vellodi, *Ann. Neurol.*, 2008, **64**, 514–522.
348 K. A. McEachern, J. Fung, S. Komarnitsky, C. S. Siegel, W. Chaung, E. Hutto, J. A. Shayman, G. A. Grabowski, J. M. F. G. Aerts, S. H. Cheng, D. P. Copeland and J. Marshall, *Mol. Genet. Metab.*, 2007, **91**, 259–267.
349 D. J. Kuter, A. Mehta, C. E. M. Hollak, P. Giraldo, D. Hughes, N. Belmatoug, M. Brand, A. Muller, B. Schaaf, G. R. and A. Zimran, *Blood Cells, Mol. Dis.*, 2013, **51**, 116–124.
350 M. Machaczka, R. Hast, I. Dahlman, R. Lerner, M. Klimkowska, M. Engvall and H. Hägglund, *Ups. J. Med. Sci.*, 2012, **117**, 28–34.
351 M. J. Peterschmitt, G. F. Cox, J. Ibrahim, J. MacDougall, L. H. Underhill, P. Patel and S. J. M. Gaemers, *Blood Cells, Mol. Dis.*, 2018, **68**, 185–191.
352 J. A. Shayman, *Drugs Future*, 2010, **35**, 613.
353 P. K. Mistry, E. Lukina, H. Ben Turkia, D. Amato, H. Baris, M. Dasouki, M. Ghosn, A. Mehta, S.

Packman, G. Pastores, M. Petakov, S. Assouline, M. Balwani, S. Danda, E. Hadjiev, A. Ortega, S. Shankar, M. H. Solano, L. Ross, J. Angell and M. J. Peterschmitt, *JAMA*, 2015, **313**, 695.
354 T. M. Cox, *Curr. Opin. Investig. Drugs*, 2010, **11**, 1169–81.
355 M. Barkhuizen, D. G. Anderson and A. F. Grobler, *Neurochem. Int.*, 2016, **93**, 6–25.
356 J. Marshall, Y. Sun, D. S. Bangari, E. Budman, H. Park, J. B. Nietupski, A. Allaire, M. A. Cromwell, B. Wang, G. A. Grabowski, J. P. Leonard and S. H. Cheng, *Mol. Ther.*, 2016, **24**, 1019–1029.
357 J. Marshall, K. A. McEachern, W.-L. Chuang, E. Hutto, C. S. Siegel, J. A. Shayman, G. A. Grabowski, R. K. Scheule, D. P. Copeland and S. H. Cheng, *J. Inherit. Metab. Dis.*, 2010, **33**, 281–9.
358 A. R. Sawkar, S. L. Adamski-Werner, W. Chen, C. Wong, E. Beutler, K. Zimmer and J. W. Kelly, *Chem. Biol.*, 2005, **12**, 1235–1244.
359 U. H. Schueler, T. Kolter, C. R. Kaneski, G. C. Zirzow, K. Sandhoff and R. O. Brady, *J. Inherit.Metab. Dis. 27*, 2004, **27**, 649–58.
360 J.-Q. Q. Fan, *Trends Pharmacol. Sci.*, 2003, **24**, 355–360.
361 A. Ulloa-Aguirre, J. A. Janovick, S. P. Brothers and P. M. Conn, *Traffic*, 2004, **5**, 821–837.
362 J. P. Morello, U. E. Petäjä-Repo, D. G. Bichet and M. Bouvier, *Trends Pharmacol. Sci.*, 2000, **21**, 466–9.
363 A. R. Sawkar, W. Chen, E. Beutler, C. Wong, W. E. Balch and F. W. Kelly, *PNAS*, 2002, **99**, 15428–15433.
364 J. A. Shayman and S. D. Larsen, *J. Lipid Res.*, 2014, **55**, 1215–25.
365 H.-H. Chang, N. Asano, S. Ishii, Y. Ichikawa and J.-Q. Fan, *FEBS J.*, 2006, **273**, 4082–92.
366 R. Khanna, E. R. Benjamin, L. Pellegrino, A. Schilling, B. A. Rigat, R. Soska, H. Nafar, B. E. Ranes, J. Feng, Y. Lun, A. C. Powe, D. J. Palling, B. A. Wustman, R. Schiffmann, D. J. Mahuran, D. J. Lockhart and K. J. Valenzano, *FEBS J.*, 2010, **277**, 1618–38.
367 R. Steet, S. Chung, W.-S. Lee, C. W. Pine, H. Do and S. Kornfeld, *Biochem. Pharmacol.*, 2007, **73**, 1376–1383.
368 R. L. Lieberman, B. A. Wustman, P. Huertas, A. C. Powe, C. W. Pine, R. Khanna, M. G. Schlossmacher, D. Ringe and G. A. Petsko, *Nat. Chem. Biol.*, 2007, **3**, 101–107.
369 P. K. Mistry, G. Lopez, R. Schiffmann, N. W. Barton, N. J. Weinreb and E. Sidransky, *Mol. Genet. Metab.*, 2017, **120**, 8–21.
370 T. M. Gloster and G. J. Davies, *Org. Biomol. Chem.*, 2010, **8**, 305–320.
371 G. H. B. Maegawa, M. B. Tropak, J. D. Buttner, B. A. Rigat, M. Fuller, D. Pandit, L. Tang, G. J. Kornhaber, Y. Hamuro, J. T. R. Clarke and D. J. Mahuran, *J. Biol. Chem.*, 2009, **284**, 23502–23516.
372 A. McNeill, J. Magalhaes, C. Shen, K.-Y. Chau, D. Hughes, A. Mehta, T. Foltynie, J. M. Cooper, A. Y. Abramov, M. Gegg and A. H. V. Schapira, *Brain*, 2014, **137**, 1481–1495.
373 G. Ambrosi, C. Ghezzi, R. Zangaglia, G. Levandis, C. Pacchetti and F. Blandini, *Neurobiol. Dis.*, 2015, **82**, 235–242.
374 A. Narita, K. Shirai, S. Itamura, A. Matsuda, A. Ishihara, K. Matsushita, C. Fukuda, N. Kubota, R. Takayama, H. Shigematsu, A. Hayashi, T. Kumada, K. Yuge, Y. Watanabe, S. Kosugi, H. Nishida, Y. Kimura, Y. Endo, K. Higaki, E. Nanba, Y. Nishimura, A. Tamasaki, M. Togawa, Y. Saito, Y. Maegaki, K. Ohno and Y. Suzuki, *Ann. Clin. Transl. Neurol.*, 2016, **3**, 200–215.
375 A. Zimran, G. Altarescu and D. Elstein, *Blood Cells, Mol. Dis.*, 2013, **50**, 134–137.
376 B. Charkhand, M. H. Scantlebury, A. Narita, A. Zimran and W. Al-Hertani, *Mol. Genet. Metab. Reports*, 2019, **20**, 100476.
377 S. Mullin, L. Smith, K. Lee, G. D'Souza, P. Woodgate, J. Elflein, J. Hällqvist, M. Toffoli, A. Streeter, J. Hosking, W. E. Heywood, R. Khengar, P. Campbell, J. Hehir, S. Cable, K. Mills, H. Zetterberg, P. Limousin, V. Libri, T. Foltynie and A. H. V. Schapira, *JAMA Neurol.*, 2020, **77**, 427–434.
378 P. Alfonso, V. Andreu, A. Pino-Angeles, A. A. Moya-García, M. I. García-Moreno, J. C. Rodríguez-Rey, F. Sánchez-Jiménez, M. Pocoví, C. Ortiz Mellet, J. M. García Fernández and P. Giraldo, *ChemBioChem*, 2013, **14**, 943–949.
379 F. Clemente, C. Matassini, C. Faggi, S. Giachetti, C. Cresti, A. Morrone, P. Paoli, A. Goti, M. Martínez-Bailén and F. Cardona, *Bioorg. Chem.*, 2020, **98**, 103740.
380 S. Ito and A. J. Barrett, *Pediatr. Hematol. Oncol.*, 2013, **30**, 61–70.
381 G. A. R. Gonçalves and R. de M. A. Paiva, *Einstein*, 2017, **15**, 369.
382 J. M. Rappeport and E. I. Ginns, *N. Engl. J. Med.*, 1984, **311**, 84–88.
383 C. S. August, M. Palmieri, P. Nowell, W. L. Elkins, G. D'Angio, R. H. Glew and L. Daniels, *Pediatr. Res. 1984 184*, 1984, **18**, 236–236.
384 J. R. Hobbs, P. J. Shaw, K. H. Jones, I. Lindsay and M. Hancock, *Lancet*, 1987, **329**, 1111–1115.

385 R. O, G. CG, E. A, B. L, G. S, M. JE and S. L, *Transplantation*, 1988, **46**, 66–70.
386 K. W. Chan, L. T. Wong, D. Applegarth and A. G. Davidson, *Bone Marrow Transplant.*, 1994, **14**, 327–30.
387 P. M. Hoogerbrugge, O. F. Brouwer, P. Bordigoni, G. Cornu, P. Kapaun, J. J. Ortega, A. O'Meara, G. Souillet, D. Frappaz, S. Blanche, A. Fischer and O. Ringden, *Lancet*, 1995, **345**, 1398–1402.
388 J. L. Ferrara, J. E. Levine, P. Reddy and E. Holler, *Lancet*, 2009, **373**, 1550–1561.
389 M. Dahl, A. Doyle, K. Olsson, J.-E. Månsson, A. R. A. Marques, M. Mirzaian, J. M. Aerts, M. Ehinger, M. Rothe, U. Modlich, A. Schambach and S. Karlsson, *Mol. Ther.*, 2015, **23**, 835–844.
390 I. B. Enquist, E. Nilsson, A. Ooka, J.-E. Månsson, K. Olsson, M. Ehinger, R. O. Brady, J. Richter and S. Karlsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 13819–24.
391 S. G. Scharenberg, E. Poletto, K. L. Lucot, P. Colella, A. Sheikali, T. J. Montine, M. H. Porteus and N. Gomez-Ospina, *Nat. Commun.*, 2020, **11**, 3327.
392 W. ANDERSON, *Br. J. Dermatol.*, 1898, **10**, 113–117.
393 A. W. Pompen, M. Ruiter and H. J. Wyers, *Acta Med Scand*, 1947, **128**, 234–255.
394 C. C. Sweeley and B. Klionsky, *J. Biol. Chem.*, 1963, **238**, 3148–50.
395 R. O. Brady, A. E. Gal, R. M. Bradley, E. Martensson, A. L. Warshaw and L. Laster, *N. Engl. J. Med.*, 1967, **276**, 1163–1167.
396 R. J. Desnick, Y. A. Ioannou and C. M. Eng, *McGraw-Hill Med.*, 2014, 1–62.
397 L. Van Der Tol, B. E. Smid, B. J. H. M. Poorthuis, M. Biegstraaten, R. H. L. Deprez, G. E. Linthorst and C. E. M. Hollak, *J. Med. Genet.*, 2014, **51**, 1–9.
398 M. Spada, S. Pagliardini, M. Yasuda, T. Tukel, G. Thiagarajan, H. Sakuraba, A. Ponzone and R. J. Desnick, *Am. J. Hum. Genet.*, 2006, **79**, 31–40.
399 W.-L. Hwu, Y.-H. Chien, N.-C. Lee, S.-C. Chiang, R. Dobrovolny, A.-C. Huang, H.-Y. Yeh, M.-C. Chao, S.-J. Lin, T. Kitagawa, R. J. Desnick and L.-W. Hsu, *Hum. Mutat.*, 2009, **30**, 1397–1405.
400 A. Mehta, R. Ricci, U. Widmer, F. Dehout, A. Garcia de Lorenzo, C. Kampmann, A. Linhart, G. Sunder-Plassmann, M. Ries and M. Beck, *Eur. J. Clin. Invest.*, 2004, **34**, 236–242.
401 R. J. Desnick, Y. A. Ioannou and C. M. Eng, *The metabolic and molecular bases of inherited disease*, McGraw-Hill, New York, 2001.
402 H. Askari, C. R. Kaneski, C. Semino-Mora, P. Desai, A. Ang, D. E. Kleiner, L. T. Perlee, M. Quezado, L. E. Spollen, B. A. Wustman and R. Schiffmann, *Virchows Arch.*, 2007, **451**, 823–834.
403 J.-S. S. Shen, X.-L. L. Meng, D. F. Moore, J. M. Quirk, J. A. Shayman, R. Schiffmann and C. R. Kaneski, *Mol. Genet. Metab.*, 2008, **95**, 163–168.
404 A. M. Das and H. Y. Naim, *Adv. Clin. Chem.*, 2009, **49**, 57–71.
405 F. Weidemann, F. Breunig, M. Beer, J. Sandstede, S. Störk, W. Voelker, G. Ertl, A. Knoll, C. Wanner and J. M. Strotmann, *Eur. Heart J.*, 2005, **26**, 1221–7.
406 R. Torra, in *Kidney International*, Nature Publishing Group, 2008, vol. 74, pp. S29-32.
407 R. Schiffmann, D. G. Warnock, M. Banikazemi, J. Bultas, G. E. Linthorst, S. Packman, S. A. Sorensen, W. R. Wilcox and R. J. Desnick, *Nephrol. Dial. Transplant.*, 2009, **24**, 2102–2111.
408 M. H. Branton, R. Schiffman, S. G. Sabnis, G. J. Murray, J. M. Quirk G. Altarescu, L. Ggoldfarb R. O. Brady, J. E. Balow, H. A. Austin and J. B. Kopp, *Medicine (Baltimore).*, 2002, **81**, 122–138.
409 R. El-Abassi, D. Singhal and J. D. England, *J. Neurol. Sci.*, 2014, **344**, 5–19.
410 J. Shabbeer, M. Yasuda, S. D. Benson and R. J. Desnick, *Hum. Genomics*, 2006, **2**, 297–309.
411 Y. A. Zarate and R. J. Hopkin, *Lancet*, 2008, **372**, 1427–1435.
412 M. Möhrenschlager, M. Braun-Falco, J. Ring and D. Abeck, *Am. J. Clin. Dermatol.*, 2003, **4**, 189–196.
413 C. Kampmann, C. M. Wiethoff, C. Whybra, F. A. Baehner, E. Mengel and M. Beck, *Acta Paediatr. Int. J. Paediatr.*, 2008, **97**, 463–469.
414 R. Schiffmann, *Pharmacol. Ther.*, 2009, **122**, 65–77.
415 K. D. MacDermot, *J. Med. Genet.*, 2001, **38**, 750–760.
416 K. D. MacDermot, *J. Med. Genet.*, 2001, **38**, 769–775.
417 M. Arends, C. Wanner, D. Hughes, A. Mehta, D. Oder, O. T. Watkinson, P. M. Elliott, G. E. Linthorst, F. A. Wijburg, M. Biegstraaten and C. E. Hollak, *J. Am. Soc. Nephrol.*, 2017, **28**, 1631–1641.
418 W. von Scheidt, C. M. Eng, T. F. Fitzmaurice, E. Erdmann, G. Hübner, E. G. J. Olsen, H. Christomanou, R. Kandolf, D. F. Bishop and R. J. Desnick, *N. Engl. J. Med.*, 1991, **324**, 395–399.
419 G. Houge, C. Tøndel, Ø. Kaarbøe, A. Hirth, L. Bostad and E. Svarstad, *Eur. J. Hum. Genet.*, 2011, **19**, 1111–1111.
420 G. E. Linthorst, M. G. Bouwman, F. A. Wijburg, J. M. F. G. Aerts, B. J. H. M. Poorthuis and C. E. M.

Hollak, *J. Med. Genet.*, 2010, **47**, 217–222.
421 M. P. Wasserstein, M. Caggana, S. M. Bailey, R. J. Desnick, L. Edelmann, L. Estrella, I. Holzman, N. R. Kelly, R. Kornreich, S. G. Kupchik, M. Martin, S. M. Nafday, R. Wasserman, A. Yang, C. Yu and J. J. Orsini, *Genet. Med.*, 2019, **21**, 631–640.
422 D. P. Germain, *Orphanet J. Rare Dis.*, 2010, **5**, 30.
423 C. Whybra, C. Kampmann, I. Willers, J. Davies, B. Winchester, J. Kriegsmann, K. Brühl, A. Gal, S. Bunge and M. Beck, *J. Inherit. Metab. Dis.*, 2001, **24**, 715–24.
424 B. R. Migeon, *J. Am. Soc. Nephrol.*, 2008, **19**, 2052–2059.
425 C. Kampmann, F. Baehner, C. Whybra, C. Martin, C. M. Wiethoff, M. Ries, A. Gal and M. Beck, *J. Am. Coll. Cardiol.*, 2002, **40**, 1668–74.
426 K. Wendrich, C. Whybra, M. Ries, A. Gal and M. Beck, in *Rare Kidney Diseases*, KARGER, Basel, 2001, **136**, 241–244.
427 B. Hoffmann, M. Schwarz, A. Mehta and S. Keshav, *Clin. Gastroenterol. Hepatol.*, 2007, **5**, 1447–1453.
428 S. Magage, J. C. Lubanda, Z. Susa, J. Bultas, D. Karetová, R. Dobrovolný, M. Hřebíček, D. P. Germain and A. Linhart, *J. Inherit. Metab. Dis.*, 2007, **30**, 790–799.
429 P. Juchniewicz, A. Kloska, A. Tylki-Szymańska, J. Jakóbkiewicz-Banecka, G. Węgrzyn, M. Moskot, M. Gabig-Cimińska and E. Piotrowska, *Gene*, 2018, **641**, 259–264.
430 E. Maier, S. Osterrieder, C. Whybra, M. Ries, A. Gal, M. Beck, A. Roscher and A. Muntau, *Acta Paediatr.*, 2006, **95**, 30–38.
431 A. Mehta, M. Beck, F. Eyskens, C. Feliciani, I. Kantola, U. Ramaswami, A. Rolfs, A. Rivera, S. Waldek and D. P. Germain, *QJM*, 2010, **103**, 641–659.
432 D. F. Bishop, D. H. Calhount, H. S. Bernstein, P. Hantzopoulost, M. Quinnt and R. J. Desnick, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 4859–4863.
433 R. Kornreich, R. J. Desnick and D. F. Bishop, *Nucleic Acids Res.*, 1989, **17**, 3301–3302.
434 S. L. Hostikka, R. L. Eddy, M. G. Byers, M. Hoyhtya, T. B. Shows and K. Tryggvason, *Proc. Natl. Acad. Sci.*, 1990, **87**, 1606–1610.
435 D. F. Barker, S. L. Hostikka, J. Zhou, L. T. Chow, A. R. Oliphant, S. C. Gerken, M. C. Gregory, M. H. Skolnick, C. L. Atkin and K. Tryggvason, *Science (80-.).*, 1990, **248**, 1224–1227.
436 D. Vetrie, I. Vořechovský, P. Sideras, J. Holland, A. Davies, F. Flinter, L. Hammarström, C. Kinnon, R. Levinsky, M. Bobrow, C. I. E. Smith and D. R. Bentley, *Nature*, 1993, **361**, 226–233.
437 J. D. Chen, D. Mackey, H. Fuller, S. Serravalle, J. Olsson and M. J. Denton, *Hum. Genet.*, 1989, **83**, 292–294.
438 D. F. Bishop, R. Kornreich and R. J. Desnick, *Genetics*, 1988, **85**, 3903–3907.
439 S. Saito, K. Ohno and H. Sakuraba, *J. Hum. Genet.*, 2011, **56**, 467–468.
440 S. C. Garman and D. N. Garboczi, *J. Mol. Biol.*, 2004, **337**, 319–335.
441 S. C. Garman, *Biocatal. Biotransformation*, 2006, **24**, 129–136.
442 C. M. Eng and R. J. Desnick, *Hum. Mutat.*, 1994, **3**, 103–111.
443 H. S. Bernstein, D. F. Bishop, K. H. Astrin, R. Kornreich, C. M. Eng, H. Sakuraba and R. J. Desnick, *J. Clin. Invest.*, 1989, **83**, 1390–1399.
444 C. M. Eng, L. A. Resnick-Silverman, D. J. Niehaus, K. H. Astrin and R. J. Desnick, *Am. J. Hum. Genet.*, 1993, **53**, 1186–97.
445 H. Onay, H. Bolat, G. K. Yıldırım, E. Kose, S. K. Uçar, S. Aşıkovalı, F. Özkınay and M. Çoker, *J. Pediatr. Endocrinol. Metab.*, 2020, **33**, 1245–1250.
446 J. M. Aerts, J. E. Groener, S. Kuiper, W. E. Donker-Koopman, A. Strijland, R. Ottenhoff, C. van Roomen, M. Mirzaian, F. A. Wijburg, G. E. Linthorst, A. C. Vedder, S. M. Rombach, J. Cox-Brinkman, P. Somerharju, R. G. Boot, C. E. Hollak, R. O. Brady and B. J. Poorthuis, *Proc. Natl. Acad. Sci.*, 2008, **105**, 2812–2817.
447 P. Rozenfeld and S. Feriozzi, *Molecular Genetics and Metabolism,* 2017, **122**, 19-27
448 R. J. Desnick, S. R. Thorpe and M. B. Fiddler, *Physiol. Rev.*, 1976, **56**, 57–99.
449 L. C. Hsing and A. Y. Rudensky, *Immunol. Rev.*, 2005, **207**, 229–241.
450 S. U. Walkley, in *Journal of Inherited Metabolic Disease*, John Wiley & Sons, Ltd, 2009, **32,** pp. 181–189.
451 P. N. De Francesco, J. M. Mucci, R. Ceci, C. A. Fossati and P. A. Rozenfeld, *Mol. Genet. Metab.*, 2013, **109**, 93–99.
452 Y. Hayashi, H. Hanawa, S. Jiao, G. Hasegawa, Y. Ohno, K. Yoshida, T. Suzuki, T. Kashimura, H. Obata, K. Tanaka, T. Watanabe and T. Minamino, *Inflammation*, 2015, **38**, 2288–2299.

453 G. B. Biancini, C. S. Vanzin, D. B. Rodrigues, M. Deon, G. S. Ribas, A. G. Barschak, V. Manfredini, C. B. O. Netto, L. B. Jardim, R. Giugliani and C. R. Vargas, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2012, **1822**, 226–232.
454 P. Martinez, M. Aggio and P. Rozenfeld, *J. Inherit. Metab. Dis.*, 2007, **30**, 365–369.
455 C. Filoni, A. Caciotti, L. Carraresi, C. Cavicchi, R. Parini, D. Antuzzi, A. Zampetti, S. Feriozzi, P. Poisetti, S. C. Garman, R. Guerrini, E. Zammarchi, M. A. Donati and A. Morrone, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2010, **1802**, 247–252.
456 W. R. Wilcox, J. P. Oliveira, R. J. Hopkin, A. Ortiz, M. Banikazemi, U. Feldt-Rasmussen, K. Sims, S. Waldek, G. M. Pastores, P. Lee, C. M. Eng, L. Marodi, K. E. Stanford, F. Breunig, C. Wanner, D. G. Warnock, R. M. Lemay and D. P. Germain, *Mol. Genet. Metab.*, 2008, **93**, 112–128.
457 J. S. Mayes, J. B. Scheerer, R. N. Sifers and M. L. Donaldson, *Clin. Chim. Acta*, 1981, **112**, 247–251.
458 T. Kitagawa, K. Suzuki, N. Ishige, T. Ohashi, M. Kobayashi, Y. Eto, A. Tanaka, H. Odaka and M. Owada, *Pediatr. Nephrol.*, 2008, **23**, 1461–1471.
459 R. O. Brady, B. W. Uhlendorf and C. B. Jacobson, *Science (80-.).*, 1971, **172**, 174–175.
460 E. Schäfer, K. Baron, U. Widmer, P. Deegan, H. P.H. Neumann, G. Sunder-Plassmann, J.-O. Johansson, C. Whybra, M. Ries, G. M. Pastores, A. Mehta, M. Beck and A. Gal, *Hum. Mutat.*, 2005, **25**, 412–412.
461 A. Schirinzi, M. Centra, C. Prattichizzo, M. Gigante, M. De Fabritiis, V. Giancaspro, F. Petrarulo, R. Santacroce, M. Margaglione, L. Gesualdo and E. Ranieri, *Mol. Genet. Metab.*, 2008, **94**, 382–385.
462 A. A. Hagège, E. Caudron, T. Damy, R. Roudaut, A. Millaire, C. Etchecopar-Chevreuil, T. C. Tran, F. Jabbour, C. Boucly, P. Prognon, P. Charron and D. P. Germain, *Heart*, 2011, **97**, 131–136.
463 J. P. Schouten, C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens and G. Pals, *Nucleic Acids Res.*, 2002, **30**, e57.
464 A. De Luca, I. Bottillo, M. C. Dasdia, A. Morella, V. Lanari, L. Bernardini, L. Divona, S. Giustini, L. Sinibaldi, A. Novelli, I. Torrente, A. Schirinzi and B. Dallapiccola, *J. Med. Genet.*, 2007, **44**, 800–808.
465 W. J. L. Cable, R. H. McCluer, E. H. Kolodny and M. D. Ullman, *Neurology*, 1982, **32**, 1139–1145.
466 S. Roy, K. Gaudin, D. P. Germain, A. Baillet, P. Prognon and P. Chaminade, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2004, **805**, 331–337.
467 C. Auray-Blais, D. Cyr, K. Mills, R. Giguère and R. Drouin, *J. Inherit. Metab. Dis.*, 2007, **30**, 106.
468 A. C. Vedder, A. Strijland, M. A. vd B. Weerman, S. Florquin, J. M. F. G. Aerts and C. E. M. Hollak, *J. Inherit. Metab. Dis.*, 2006, **29**, 106–111.
469 M. D. Sanchez-Nino, A. B. Sanz, S. Carrasco, M. A. Saleem, P. W. Mathieson, J. M. Valdivielso, M. Ruiz-Ortega, J. Egido and A. Ortiz, *Nephrol. Dial. Transplant.*, 2011, **26**, 1797–1802.
470 S. M. Rombach, N. Dekker, M. G. Bouwman, G. E. Linthorst, A. H. Zwinderman, F. A. Wijburg, S. Kuiper, M. A. vd Bergh Weerman, J. E. M. Groener, B. J. Poorthuis, C. E. M. Hollak and J. M. F. G. Aerts, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2010, **1802**, 741–748.
471 A. Nowak, T. P. Mechtler, T. Hornemann, J. Gawinecka, E. Theswet, M. J. Hilz and D. C. Kasper, *Mol. Genet. Metab.*, 2018, **123**, 148–153.
472 T. Togawa, T. Kodama, T. Suzuki, K. Sugawara, T. Tsukimura, T. Ohashi, N. Ishige, K. Suzuki, T. Kitagawa and H. Sakuraba, *Mol. Genet. Metab.*, 2010, **100**, 257–261.
473 A. Nowak, T. P. Mechtler, R. J. Desnick and D. C. Kasper, *Mol. Genet. Metab.*, 2017, **120**, 57–61.
474 M. J. van Breemen, S. M. Rombach, N. Dekker, B. J. Poorthuis, G. E. Linthorst, A. H. Zwinderman, F. Breunig, C. Wanner, J. M. Aerts and C. E. Hollak, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2011, **1812**, 70–76.
475 G. Effraimidis, U. Feldt-Rasmussen, Å. K. Rasmussen, P. Lavoie, M. Abaoui, M. Boutin and C. Auray-Blais, *J. Med. Genet.*, 2021, **58**, 692-700.
476 V. A. Blaho and T. Hla, *J. Lipid Res.*, 2014, **55,** 1596–1608.
477 S. Spiegel and S. Milstien, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 397–407.
478 M. L. Allende and R. L. Proia, in *Sphingolipid Biology*, Springer Japan, 2006, 385–402.
479 A. J. Snider, K. Alexa Orr Gandy and L. M. Obeid, *Biochimie*, 2010, **92**, 707–715.
480 M. Mirzaian, P. Wisse, M. J. Ferraz, A. R. A. Marques, T. L. Gabriel, C. P. A. A. van Roomen, R. Ottenhoff, M. van Eijk, J. D. C. Codée, G. A. van der Marel, H. S. Overkleeft and J. M. Aerts, *Clin. Chim. Acta*, 2016, **459**, 36–44.
481 E. Lloyd-Evans and F. M. Platt, *Traffic*, 2010, **11**, 419–428.
482 E. Lloyd-Evans, A. J. Morgan, X. He, D. A. Smith, E. Elliot-Smith, D. J. Sillence, G. C. Churchill, E. H. Schuchman, A. Galione and F. M. Platt, *Nat. Med.*, 2008, **14**, 1247–1255.

483 O. Lidove, D. JolY, F. Barbey, S. Bekri, J. F. Alexandra and V. Peigne, *Int J Clin Pr.*, 2007, **61**, 293–302.
484 G. M. Keating, *BioDrugs*, 2012, **26**, 335–354.
485 G. M. Keating and D. Simpson, *Drugs*, 2007, **67**, 435–455.
486 K. Lee, X. Jin, K. Zhang, L. Copertino, L. Andrews and J. Baker-Malcolm, *Glycobiology*, 2003, **13**, 305–313.
487 C. O. Kim, E. A. Oh and M. S. Park, *J. Clin. Pharmacol.*, 2014, **54**, 675–681.
488 D. A. Hughes, P. M. Elliott, J. Shah, J. Zuckerman, G. Coghlan, J. Brookes and A. B. Mehta, *Heart*, 2008, **94**, 153–158.
489 R. Schiffmann, P. Hauer and B. Freeman, *Muscle Nerve*, 2006, **34**, 53–56.
490 R. Schiffmann, M. Ries, M. Timmons, J. T. Flaherty and R. O. Brady, *Nephrol. Dial. Transplant.*, 2006, **21**, 345–354.
491 A. C. Vedder, G. E. Linthorst, G. Houge, J. E. M. Groener, E. E. Ormel, B. J. Bouma, J. M. F. G. Aerts, A. Hirth and C. E. M. Hollak, *PLoS One*, 2007, **2**, e598.
492 R. Schiffmann, H. Askari and M. Timmons, *J Am Soc Nephrol*, 2007, **18**, 1576–1583.
493 C. M. Eng, N. Guffon, W. R. Wilcox, D. P. Germain, P. Lee, S. Waldek, L. Caplan, G. E. Linthorst and R. J. Desnick, *N. Engl. J. Med.*, 2001, **345**, 9–16.
494 C. Kampmann, A. Linhart, R. B. Devereux and R. Schiffmann, *Clin. Ther.*, 2009, **31**, 1966–1976.
495 W. R. Wilcox, M. Banikazemi, N. Guffon, S. Waldek, P. Lee, G. E. Linthorst, R. J. Desnick and D. P. Germain, *Am. J. Hum. Genet.*, 2004, **75**, 65–74.
496 D. P. Germain, S. Waldek, M. Banikazemi, D. A. Bushinsky, J. Charrow, R. J. Desnick, P. Lee, T. Loew, A. C. Vedder, R. Abichandani, W. R. Wilcox and N. Guffon, *J. Am. Soc. Nephrol.*, 2007, **18**, 1547–1557.
497 R. J. Desnick and E. H. Schuchman, *Nat. Rev. Genet*, 2002, **3**, 954–966.
498 S. J. van der Veen, W. J. Vlietstra, L. van Dussen, A. B. P. van Kuilenburg, M. G. W. Dijkgraaf, M. Lenders, E. Brand, C. Wanner, D. Hughes, P. M. Elliott, C. E. M. Hollak and M. Langeveld, *Int. J. Mol. Sci. 2020, Vol. 21, Page 5784*, 2020, **21**, 5784.
499 B. Bénichou, S. Goyal, C. Sung, A. M. Norfleet and F. O'Brien, *Mol. Genet. Metab.*, 2009, **96**, 4–12.
500 G. E. Linthorst, C. E. M. Hollak, W. E. Donker-Koopman, A. Strijland and J. M. F. G. Aerts, *Kidney Int.*, 2004, **66**, 1589–1595.
501 L. Ellgaard and A. Helenius, *Nat. Rev. Mol. Cell Biol.*, 2003, **4**, 181–191.
502 K. Sugawara, Y. Tajima, I. Kawashim, T. Tsukimura, S. Saito, K. Ohno, K. Iwamoto, T. Kobayashi, K. Itoh and H. Sakuraba, *Mol. Genet. Metab.*, 2009, **96**, 233–238.
503 A. Frustaci, C. Chimenti, R. Ricci, L. Natale, M. A. Russo, M. Pieroni, C. M. Eng and R. J. Desnick, *N. Engl. J. Med.*, 2001, **345**, 25–32.
504 J. Q. Fan, S. Ishii, N. Asano and Y. Suzuki, *Nat. Med.*, 1999, **5**, 112–115.
505 N. Asano, S. Ishii, H. Kizu, K. Ikeda, K. Yasuda, A. Kato, O. R. Martin and J.-Q. Fan, *Eur. J. Biochem.*, 2000, **267**, 4179–4186.
506 J. Q. Fan, S. Ishii, N. Asano and Y. Suzuki, *Nat. Med*, 1999, **5**, 112–115.
507 Y. Miyake and M. Ebata, *Agric. Biol. Chem.*, 1988, **52**, 661–666.
508 G. H.-F. Yam, C. Zuber and J. Roth, *FASEB J.*, 2005, **19**, 12–18.
509 E. R. Benjamin, J. J. Flanagan, A. Schilling, H. H. Chang, L. Agarwal, E. Katz, X. Wu, C. Pine, B. Wustman, R. J. Desnick, D. J. Lockhart and K. J. Valenzano, *Journal of Inherited Metabolic Disease*, 2009, **32**,. 424–440.
510 D. G. Warnock, D. G. Bichet, M. Holida, O. Goker-Alpan, K. Nicholls, M. Thomas, F. Eyskens, S. Shankar, M. Adera, S. Sitaraman, R. Khanna, J. J. Flanagan, B. A. Wustman, J. Barth, C. Barlow, K. J. Valenzano, D. J. Lockhart, P. Boudes and F. K. Johnson, *PLoS One*, 2015, **10**, e0134341.
511 B. Young-Gqamana, N. Brignol, H.-H. Chang, R. Khanna, R. Soska, M. Fuller, S. A. Sitaraman, D. P. Germain, R. Giugliani, D. A. Hughes, A. Mehta, K. Nicholls, P. Boudes, D. J. Lockhart, K. J. Valenzano and E. R. Benjamin, *PLoS One*, 2013, **8**, e57631.
512 R. Khanna, R. Soska, Y. Lun, J. Feng, M. Frascella, B. Young, N. Brignol, L. Pellegrino, S. A. Sitaraman, R. J. Desnick, E. R. Benjamin, D. J. Lockhart and K. J. Valenzano, *Mol. Ther.*, 2010, **18**, 23–33.
513 S. Ishii, H. H. Chang, H. Yoshioka, T. Shimada, K. Mannen, Y. Higuchi, A. Taguchi and J. Q. Fan, *J. Pharmacol. Exp. Ther.*, 2009, **328**, 723–731.
514 F. K. Johnson, P. N. Mudd, A. Bragat, M. Adera and P. Boudes, *Clin. Pharmacol. Drug Dev.*, 2013, **2**, 120–132.

515 D. P. Germain, R. Giugliani, D. A. Hughes, A. Mehta, K. Nicholls, L. Barisoni, C. J. Jennette, A. Bragat, J. Castelli, S. Sitaraman, D. J. Lockhart and P. F. Boudes, *Orphanet J. Rare Dis.*, 2012, **7**, 91.
516 R. Giugliani, S. Waldek, D. P. Germain, K. Nicholls, D. G. Bichet, J. K. Simosky, A. C. Bragat, J. P. Castelli, E. R. Benjamin and P. F. Boudes, *Mol. Genet. Metab.*, 2013, **109**, 86–92.
517 A. Markham, *Drugs*, 2016, **76**, 1147–1152.
518 D. A. Hughes, K. Nicholls, S. P. Shankar, G. Sunder-Plassmann, D. Koeller, K. Nedd, G. Vockley, T. Hamazaki, R. Lachmann, T. Ohashi, I. Olivotto, N. Sakai, P. Deegan, D. Dimmock, F. Eyskens, D. P. Germain, O. Goker-Alpan, E. Hachulla, A. Jovanovic, C. M. Lourenco, I. Narita, M. Thomas, W. R. Wilcox, D. G. Bichet, R. Schiffmann, E. Ludington, C. Viereck, J. Kirk, J. Yu, F. Johnson, P. Boudes, E. R. Benjamin, D. J. Lockhart, C. Barlow, N. Skuban, J. P. Castelli, J. Barth and U. Feldt-Rasmussen, *J. Med. Genet.*, 2017, **54**, 288–296.
519 D. P. Germain, D. A. Hughes, K. Nicholls, D. G. Bichet, R. Giugliani, W. R. Wilcox, C. Feliciani, S. P. Shankar, F. Ezgu, H. Amartino, D. Bratkovic, U. Feldt-Rasmussen, K. Nedd, U. Sharaf El Din, C. M. Lourenco, M. Banikazemi, J. Charrow, M. Dasouki, D. Finegold, P. Giraldo, O. Goker-Alpan, N. Longo, C. R. Scott, R. Torra, A. Tuffaha, A. Jovanovic, S. Waldek, S. Packman, E. Ludington, C. Viereck, J. Kirk, J. Yu, E. R. Benjamin, F. Johnson, D. J. Lockhart, N. Skuban, J. Castelli, J. Barth, C. Barlow and R. Schiffmann, *N. Engl. J. Med.*, 2016, **375**, 545–555.
520 X. Wu, E. Katz, M. C. Della Valle, K. Mascioli, J. J. Flanagan, J. P. Castelli, R. Schiffmann, P. Boudes, D. J. Loackhart, K. J. Valenzano and E. R. Benjamin, *Hum. Mutat.*, 2011, **32**, 965–977.
521 A. Siekierska, G. De Baets, J. Reumers, R. Gallardo, S. Rudyak, K. Broersen, J. Couceiro, J. Van Durme, J. Schymkowitz and F. Rousseau, *J. Biol. Chem.*, 2012, **287**, 28386–28397.
522 E. R. Benjamin, R. Khanna, A. Schilling, J. J. Flanagan, L. J. Pellegrino, N. Brignol, Y. Lun, D. Guillen, B. E. Ranes, M. Frascella, R. Soska, J. Feng, L. Dungan, B. Young, D. J. Lockhart and K. J. Valenzano, *Mol. Ther.*, 2012, **20**, 717–726.
523 C. Porto, M. Cardone, F. Fontana, B. Rossi, M. R. Tuzzi, A. Tarallo, M. V. Barone, G. Andria and G. Parenti, *Mol. Ther.*, 2009, **17**, 964–971.
524 A. Pisani, C. Porto, G. Andria and G. Parenti, *J. Inherit. Metab. Dis.*, 2014, **37**, 145–146.
525 N. S. Radin, *Glycoconj. J.*, 1996, **13**, 153–157.
526 N. S. Radin, J. A. Shayman and J. Inokuchi, *Adv. Lipid Res.*, 1993, **26**, 183–213.
527 F. M. Platt, G. R. Neises, R. A. Dwek and T. D. Butters, *J. Biol. Chem.*, 1994, **269**, 8362–8365.
528 J. A. Shayman, L. Lee, A. Abe and L. Shu, *Methods Enzymol.*, 2000, **311**, 373–387.
529 J.-I. Inokuchi and N. S. Radin, *J. Lipid Res.*, 1987, **28**, 565–571.
530 A. Abe, N. S. Radin and J. A. Shayman, *Biochim. Biophys. Acta - Lipids Lipid Metab.*, 1996, **1299**, 333–341.
531 A. Abe, L. J. Arend, L. Lee, C. Lingwood, R. O. Brady and J. A. Shayman, *Kidney Int.*, 2000, **57**, 446–454.
532 A. Abe, N. S. Radin, J. A. Shayman, L. L. Wotring, R. E. Zipkin, R. Sivakumar, J. M. Ruggieri, K. G. Carson and B. Ganem, *J. Lipid Res.*, 1995, **36**, 611–621.
533 L. Lee, A. Abe and J. A. Shayman, *J. Biol. Chem.*, 1999, **274**, 14662–14669.
534 D. L. Marks, M. Dominguez, K. Wu and R. E. Pagano, *J. Biol. Chem.*, 2001, **276**, 26492–26498.
535 A. Abe, S. Gregory, L. Lee, P. D. Killen, R. O. Brady, A. Kulkarni and J. A. Shayman, *J. Clin. Invest.*, 2000, **105**, 1563–1571.
536 K. M. Ashe, E. Budman, D. S. Bangari, C. S. Siegel, J. B. Nietupski, B. Wang, R. J. Desnick, R. K. Scheule, J. P. Leonard, S. H. Cheng and J. Marshall, *Mol. Med.*, 2015, **21**, 389–399.
537 J. M. Itier, G. Ret, S. Viale, L. Sweet, D. Bangari, A. Caron, F. Le-Gall, B. Bénichou, J. Leonard, J. F. Deleuze and C. Orsini, *J. Inherit. Metab. Dis.*, 2014, **37**, 1013–1022.
538 N. Guérard, O. Morand and J. Dingemanse, *Orphanet J. Rare Dis.*, 2017, **12**, 9.
539 T. D. Butters, L. A. G. M. Van Den Broek, G. W. J. Fleet, T. M. Krulle, M. R. Wormald, R. A. Dwek and F. M. Platt, *Tetrahedron Asymmetry*, 2000, **11**, 113–124.
540 U. Andersson, T. D. Butters, R. A. Dwek and F. M. Platt, *Biochem. Pharmacol.*, 2000, **59**, 821–829.
541 R. C. Baek, J. L. Kasperzyk, F. M. Platt and T. N. Seyfried, *Neurochem. Int.*, 2008, **52**, 1125–1133.
542 A. Khan, D. L. Barber, J. Huang, C. A. Rupar, J. W. Rip, C. Auray-Blais, M. Boutin, P. O'Hoski, K. Gargulak, W. M. McKillop, G. Fraser, S. Wasim, K. LeMoine, S. Jelinski, A. Chaudhry, N. Prokopishyn, C. F. Morel, S. Couban, P. R. Duggan, D. H. Fowler, A. Keating, M. L. West, R. Foley and J. A. Medin, *Nat. Commun. 2021 121*, 2021, **12**, 1–9.
543 A. E. Speers and B. F. Cravatt, *ChemBioChem*, 2004, **5**, 41–47.
544 A. B. Berger, P. M. Vitorino and M. Bogyo, *Am. J. Pharmacogenomics*, 2004, **4**, 371–81.

545 L. I. Willems, H. S. Overkleeft and S. I. van Kasteren, *Bioconjugate Chem*, 2014, **25**, 1181–1191.
546 L. I. Willems, J. Jiang, K.-Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. N. Beenakker, S. P. Schröder, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée and H. S. Overkleeft, *Chem. A Eur. J.*, 2014, **20**, 10864–10872.
547 L. Wu, Z. Armstrong, S. P. Schröder, C. de Boer, M. Artola, J. M. Aerts, H. S. Overkleeft and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2019, **53**, 25–36.
548 B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.*, 2008, **77**, 383–414.
549 S. P. Schroder, J. W. Van De Sande, W. W. Kallemeijn, C. L. Kuo, M. Artola, E. J. van Rooden, J. Jiang, T. J. N. Beenakker, B. I. Florea, W. A. Offen, G. J. Davies, A. J. Minaars, J. M. F. G. Aerts, J. D. C. Codee, G. A. Van Der Marel and H. H. S. Overkleeft, *R. Soc. Chem.*, 2017, **53**, 12528–12531.
550 L. I. Willems, W. A. van der Linden, N. Li, K. Y. Li, S. Hoogendoorn, G. A. van der Marel, B. I. F. A and H. S. Overkleeft, *Acc. Chem. Res.*, 2011, **44**, 718–729.
551 J. Jiang, C.-L. Kuo, L. Wu, C. Franke, W. W. Kallemeijn, B. I. Florea, E. van Meel, G. A. van der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, H. S. Overkleeft and J. M. F. G. Aerts, *ACS Cent. Sci.*, 2016, **2**, 351–358.
552 C.-S. Tsai, Y.-K. Li and L.-C. Lo, *Org. Lett.*, 2002, **4**, 3607–3610.
553 G. Legler, *Biochim. Biophys. Acta - Enzymol.*, 1968, **151**, 728–729.
554 S. Atsumi, K. Umezawa, H. Linuma, H. Naganawa, H. Nakamura, Y. Iitaka and T. J. Takeuchi, *Antibiot.*, 1990, **43**, 49–53.
555 D. Herrera Moro Chao, W. W. Kallemeijn, A. R. A. Marques, M. Orre, R. Ottenhoff, C. van Roomen, E. Foppen, M. C. Renner, M. Moeton, M. van Eijk, R. G. Boot, W. Kamphuis, E. M. Hol, J. Aten, H. S. Overkleeft, A. Kalsbeek and J. M. F. G. Aerts, *PLoS One*, 2015, **10**, e0138107.
556 J. Jiang, T. J. M. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codée, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. - A Eur. J.*, 2015, **21**, 10861–10869.
557 T. J. M. Beenakker, D. P. A. Wander, W. A. Offen, M. Artola, L. Raich, M. J. Ferraz, K. Y. Li, J. H. P. M. Houben, E. R. Van Rijssel, T. Hansen, G. A. Van Der Marel, J. D. C. Codée, J. M. F. G. Aerts, C. Rovira, G. J. Davies and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2017, **139**, 6534–6537.
558 L. I. Willems, T. J. M. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. Van Rijssel, B. I. Florea, J. D. C. C. Codée, G. A. Van Der Marel, J. M. F. G. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2014, **136**, 11622–11625.
559 K. Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. Van Den Elst, C. S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. Van Der Marel and H. S. Overkleeft, *European J. Org. Chem.*, 2014, **2014**, 6030–6043.
560 M. Artola, L. Wu, M. J. Ferraz, C.-L. L. Kuo, L. Raich, I. Z. Breen, W. A. Offen, J. D. C. C. Codée, G. A. Van Der Marel, C. Rovira, J. M. F. G. F. G. Aerts, G. J. Davies and H. S. Overkleeft, *ACS Cent. Sci.*, 2017, **3**, 784–793.
561 M. Artola, C. Hedberg, R. J. Rowland, L. Raich, K. Kytidou, L. Wu, A. Schaaf, M. J. Ferraz, G. A. Van Der Marel, J. D. C. Codée, C. Rovira, J. M. F. G. Aerts, G. J. Davies and H. S. Overkleeft, *Chem. Sci.*, 2019, **10**, 9233–9243.
562 M. Artola, C.-L. Kuo, L. T. Lelieveld, R. J. Rowland, G. A. van der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2019, **141**, 4214–4218.
563 W. W. Kallemeijn, S. Scheij, S. Hoogendoorn, M. D. Witte, D. Herrera Moro Chao, C. P. A. A. van Roomen, R. Ottenhoff, H. S. Overkleeft, R. G. Boot and J. M. F. G. Aerts, *PLoS One*, 2017, **12**, 170268.
564 Y. Liu, M. P. Patricelli and B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14694–14699.
565 D. Greenbaum, K. F. Medzihradszky, A. Burlingame and M. Bogyo, *Chem. Biol.*, 2000, **7**, 569–581.
566 B. P. Rempel and S. G. Withers, *Glycobiology*, 2008, **18**, 570–586.
567 M. P. Patricelli, A. K. Szardenings, M. Liyanage, T. K. Nomanbhoy, M. Wu, H. Weissig, A. Aban, D. Chun, S. Tanner and J. W. Kozarich, *Biochemistry*, 2007, **46**, 350–358.
568 L.-C. Lo, T.-L. Pang, C.-H. Kuo, Y.-L. Chiang, H.-Y. Wang and J.-J. Lin, *J. Proteome Res.*, 2002, **1**, 35–40.
569 O. Obianyo, C. P. Causey, J. E. Jones and P. R. Thompson, *ACS Chem. Biol.*, 2011, **6**, 1127–1135.
570 H. Ü. Kaniskan and J. Jin, *ACS Chem. Biol.*, 2015, **10**, 40–50.
571 M. P. C. Mulder, K. Witting, I. Berlin, J. N. Pruneda, K.-P. Wu, J.-G. Chang, R. Merkx, J. Bialas, M. Groettrup, A. C. O. Vertegaal, B. A. Schulman, D. Komander, J. Neefjes, F. El Oualid and H. Ovaa, *Nat. Chem. Biol.*, 2016, **12**, 523–530.
572 L. M. Chauvigné-Hines, L. N. Anderson, H. M. Weaver, J. N. Brown, P. K. Koech, C. D. Nicora, B. A.

Hofstad, R. D. Smith, M. J. Wilkins, S. J. Callister and A. T. Wright, *J. Am. Chem. Soc.*, 2012, **134**, 20521–20532.
573 O. Hekmat, Y.-W. Kim, S. J. Williams, S. He and S. G. Withers, *J. Biol. Chem.*, 2005, **280**, 35126–35135.
574 L. Wu, J. Jiang, Y. Jin, W. W. Kallemeijn, C.-L. Kuo, M. Artola, W. Dai, C. van Elk, M. van Eijk, G. A. van der Marel, J. D. C. Codée, B. I. Florea, J. M. F. G. Aerts, H. S. Overkleeft and G. J. Davies, *Nat. Chem. Biol.*, 2017, **13**, 867–873.
575 D. Lahav, B. Liu, R. J. B. H. N. Van Den Berg, A. M. C. H. Van Den Nieuwendijk, T. Wennekes, A. T. Ghisaidoobe, I. Breen, M. J. Ferraz, C. L. Kuo, L. Wu, P. P. Geurink, H. Ovaa, G. A. Van Der Marel, M. Van Der Stelt, R. G. Boot, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2017, **139**, 14192–14197.
576 O. Hekmat, C. Florizone, Y.-W. Kim, L. D. Eltis, R. A. J. Warren and S. G. Withers, *ChemBioChem*, 2007, **8**, 2125–2132.
577 A. I. Guce, N. E. Clark, E. N. Salgado, D. R. Ivanen, A. A. Kulminskaya, H. Brumer and S. C. Garman, *J. Biol. Chem.*, 2010, **285**, 3625–3632.
578 A. E. Gal and R. O. Y. M. Bradley, *J. Biol. Chem.,* 1967, **242**, 1021-1026
579 M. F. Fox, D. L. DuToit, L. Warnich and A. E. Retief, *Cytogenet. Genome Res.*, 1984, **38**, 45–49.
580 S. Bekri, *Importance of glycosylation in enzyme replacement therapy*, Oxford PharmaGenesis, 2006., Chapter 5.
581 N. C. LeDonne, J. L. Fairley and C. C. Sweeley, *Arch. Biochem. Biophys.*, 1983, **224**, 186–195.
582 G. Davies and B. Henrissat, *Structure*, 1995, **3**, 853–859.
583 H. Brumer, P. F. Sims and M. L. Sinnott, *Biochem. J.*, 1999, **339**, 43–53.
584 D. O. Hart, S. He, C. J. Chany, S. G. Withers, P. F. G. G. Sims, M. L. Sinnott and H. Brumer, *Biochemistry*, 2000, **39**, 9826–9836.
585 H. D. Ly, S. Howard, K. Shum, S. He, A. Zhu and S. G. Withers, *Carbohydr. Res.*, 2000, **329**, 539–547.
586 G. J. Davies and S. J. Williams, *Biochem. Soc. Trans.*, 2016, **44**, 79–87.
587 S. C. Garman, L. Hannick, A. Zhu and D. N. Garboczi, *Structure*, 2002, **10**, 425–434.
588 Z. Fujimoto, S. Kaneko, M. Momma, H. Kobayashi and H. Mizuno, *J. Biol. Chem.*, 2003, **278**, 20313–20318.
589 A. M. Golubev, R. A. Nagem, J. R. B. Neto, K. N. Neustroev, E. V. Eneyskaya, A. A. Kulminskaya, K. A. Shabalin, A. N. Savel'ev and I. Polikarpov, *J. Mol. Biol.*, 2004, **39**, 413–422.
590 J. D. Mccarter and S. G. Withers, *Current Opinions in Structural Biology,* 1994, **4**, 885-895
591 J. Agirre and S. McNicholas, *Acta Crystallogr D Struct Biol*, 2017, **73**, 187–194.
592 Y. A. Ioannou, D. F. Bishop and R. J. Desnick, *J. Cell Biol.*, 1992, **119**, 1137–1150.
593 Y. A. Ioannou, K. M. Zeidner, M. E. Grace and R. J. Desnick, *Biochem. J.*, 1998, **332**, 789–797.
594 J. Agirre, J. Iglesias-Fernández, C. Rovira, G. J. Davies, K. S. Wilson and K. D. Cowtan, *Nat. Struct. Mol. Biol.*, 2015, **22**, 833–834.
595 R. L. Lieberman, J. A. D'aquino, D. Ringe, G. A. Petsko, J. Alejandro D'aquino, D. Ringe, G. A. Petsko, J. A. D'Aquino, D. Ringe and G. A. Petsko, *Biochemistry*, 2009, **48**, 4816–4827.
596 I. B. Tomasic, M. C. Metcalf, A. I. Guce, N. E. Clark and S. C. Garman, *J. Biol. Chem.*, 2010, **285**, 21560–21566.
597 A. I. Guce, N. E. Clark, J. J. Rogich and S. C. Garman, *Chem. Biol.*, 2011, **18**, 1521–1526.
598 Y. Yu, T. Mena-Barragán, K. Higaki, J. L. Johnson, J. E. Drury, R. L. Lieberman, N. Nakasone, H. Ninomiya, T. Tsukimura, H. Sakuraba, Y. Suzuki, E. Nanba, C. O. Mellet, J. M. García Fernández and K. Ohno, *ACS Chem. Biol.*, 2014, **9**, 1460–1469.
599 K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, W. E. Donker-Koopman, R. G. Boot, J. M. F. G. Aerts, J. D. C. Codée, G. A. van der Marel and H. S. Overkleeft, *Org. Biomol. Chem.*, 2014, **12**, 7786–91.
600 M. D. Witte, W. W. Kallemeijn, J. Aten, K. Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. Hollak, R. Ottenhoff, R. G. Boot, G. A. va. der Marel, H. S. Overkleeft and J. M. Aert, *Nat Chem Biol*, 2010, **6**, 907–913.
601 M. D. Witte, M. T. C. Walvoort, K.-Y. Li, W. W. Kallemeijn, W. E. Donker-Koopman, R. G. Boot, J. M. F. G. Aerts, J. D. C. Codée, G. A. van der Marel and H. S. Overkleeft, *ChemBioChem*, 2011, **12**, 1263–1269.
602 L. I. Willems, T. J. M. Beenakker, B. Murray, B. Gagestein, H. van den Elst, E. R. van Rijssel, J. D. C. Codée, W. W. Kallemeijn, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *European J. Org. Chem.*, 2014, **2014**, 6044–6056.

603 C.-L. Kuo, E. van Meel, K. Kytidou, W. W. Kallemeijn, M. Witte, H. S. Overkleeft, M. E. Artola and J. M. Aerts, *Methods Enzymol.*, 2018, **598**, 217–235.
604 M. L. Shulman, S. D. Shiyan and A. Y. Khorlin, *BBA - Enzymol.*, 1976, **445**, 169–181.
605 P. J. Marshall, M. L. Sinnott, P. J. Smith and D. Widdows, *J. Chem. Soc. Perkin Trans. 1*, 1981, 366–376.
606 G. Legler and E. Bieberich, *Arch. Biochem. Biophys.*, 1988, **260**, 437–442.
607 S. G. Withers, K. Rupitz and I. P. Street, *J. Biol. Chem.*, 1988, **263**, 7929–32.
608 G. J. Davies, A. Planas and C. Rovira, *Acc. Chem. Res.*, 2012, **45**, 308–316.
609 D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
610 I. P. Street, J. B. Kempton and S. G. Withers, *Biochemistry*, 1992, **31**, 9970–9978.
611 S. Miao, J. D. McCarter, M. . Grace, G. A. Grabowski, R. Aebersold and S. G. Withers, *J. Biol. Chem.*, 1994, **269**, 10975–10978.
612 M. D. Burkart, Z. Zhang, S.-C. Hung and C.-H. Wong, *J. Am. Chem. Soc.*, 1997, **119**, 11743–11746.
613 M. K. Tong and B. Ganem, *J. Am. Chem. Soc.*, 1988, **110**, 312–313.
614 P. D. Shaw Stewart, S. A. Kolek, R. A. Briggs, N. E. Chayen and P. F. M. Baldock, *Cryst. Growth Des.*, 2011, **11**, 3432–3441.
615 W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 125–132.
616 G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186–190.
617 P. Evans, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2006, **62**, 72–82.
618 P. R. Evans and G. N. Murshudov, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2013, **69**, 1204–1214.
619 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 235–242.
620 A. Vagin and A. Teplyakov, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 22–25.
621 G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 1997, **53**, 240–255.
622 P. Emsley and K. Cowtan, *Acta Crystallogr D*, 2004, **60**, 2126–2132.
623 A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2012, **68**, 431–440.
624 M. N. Schulz, J. Landström and R. E. Hubbard, *Anal. Biochem.*, 2013, **433**, 43–47.
625 F. B. Bdira, W. W. Kallemeijn, S. V. Oussoren, S. Scheij, B. Bleijlevens, B. I. Florea, C. P. A. A. van Roomen, R. Ottenhoff, M. J. F. M. van Kooten, M. T. C. Walvoort, M. D. Witte, R. G. Boot, M. Ubbink, H. S. Overkleeft, J. M. F. G. Aerts, F. Ben Bdira, W. W. Kallemeijn, S. V. Oussoren, S. Scheij, B. Bleijlevens, B. I. Florea, C. P. A. A. van Roomen, R. Ottenhoff, M. J. F. M. van Kooten, M. T. C. Walvoort, M. D. Witte, R. G. Boot, M. Ubbink, H. S. Overkleeft and J. M. F. G. Aerts, *ACS Chem. Biol.*, 2017, **12**, 1830–1841.
626 G. Andreaotti, M. Monticelli and M. V. Cubellis, *Drug. Test. Anal.*, 2017, **7**, 831–834.
627 J. L. Sebaugh, *Pharm. Stat.*, 2011, **10**, 128–134.
628 D. M. Pereira, P. Valentão and P. B. Andrade, *Chem. Sci.*, 2018, **9**, 1740–1752.
629 G. H.-F. Yam, N. Bosshard, C. Zuber, B. Steinmann and J. Roth, *Am. J. Physiol. Physiol.*, 2006, **290**, 1076–1082.
630 A. H. Futerman and F. M. Platt, *Mol. Genet. Metab.*, 2017, **120**, 22–26.
631 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490–D495.
632 E. Beutler, *Science (80-.).*, 1992, **256**, 794–799.
633 W. M. Holleran, Y. Takagi, G. K. Menon, G. Legler, K. R. Feingold and P. M. Elias, *J. Clin. Invest.*, 1993, **91**, 1656–1664.
634 W. M. Holleran, Y. Takagi, G. K. Menon, S. M. Jackson, J. M. Lee, K. R. Feingold and P. M. Elias, *J. Lipid Res.*, 1994, **35**, 905–912.
635 C. Liu, A. B. Bahnsons, J. T. Dunigan, C. Walkins, S and J. A. Barranger, *J Mol Med*, 1998, **76**, 773–781.
636 B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon and G. Davies, *Proc. Natl. Acad. Sci.*, 1995, **92**, 7090–7094.
637 F. J. St John, J. M. González and E. Pozharski, *FEBS Lett.*, 2010, **584**, 4435–4441.
638 T. Dinur, K. M. Osiecki, G. Legler, S. Gatt, R. J. Desnick and G. A. Grabowski, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 1660–4.

639 S. Fabrega, P. Durand, J.-P. Mornon and P. Lehn, *J. Soc. Biol.*, 2002, **196**, 151–160.
640 G. J. Davies, V. M.-A. Ducros, A. Varrot and D. L. Zechel, *Biochem. Soc. Trans.*, 2003, **31**, 523–527.
641 H. Dvir, M. Harel, A. A. McCarthy, L. Toker, I. Silman, A. H. Futerman and J. L. Sussman, *EMBO Rep.*, 2003, **4**, 704–709.
642 R. L. Lieberman, *Enzyme Res.*, 2011, **2011**, 1–15.
643 B. Brumshtein, M. R. Wormald, I. Silman, A. H. Futerman and J. L. Sussman, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2006, **62**, 1458–1465.
644 S. B. Larson, J. Day, A. P. B. De la Rosa, N. T. Keen and A. McPherson, *Biochemistry*, 2003, **42**, 8411–8422.
645 R. R. O'Neill, T. Tokoro, C. A. Kozak and R. O. Brady, *Proc. Natl. Acad. Sci.*, 1989, **86**, 5049–5053.
646 M. . Grace, K. M. Newman, V. Scheinker, A. Berg-Fussman and G. A. Grabowski, *J. Biol. Chem.*, 1994, **269**, 2283–2291.
647 M. E. Grace and G. A. Grabowski, *Biochem. Biophys. Res. Commun.*, 1990, **168**, 771–7.
648 A. Berg-Fussman, M. . E. Grace, Y. Ioannou and G. A. A. Grabowski, *J. Biol. Chem.*, 1993, **268**, 14861–6.
649 L. Pol-Fachin, M. Siebert, H. Verli and M. L. Saraiva-Pereira, *Glycoconj. J.*, 2016, **33**, 237–244.
650 G. Sinclair, T. A. Pfeifer, T. A. Girgliatti and F. Y. M. Choy, *Biochem. Cell Biol.*, 2006, **84**, 148–156.
651 Y. Tekoah, S. Tzaban, T. Kizhner, M. Hainrichson, A. Gantman, M. Golembo, D. Aviezer and Y. Shaaltiel, *Biosci. Rep.*
652 R. O. Brady, *Annu. Rev. Med.*, 2006, **57**, 283–296.
653 L. Premkumar, A. R. Sawkar, S. Boldin-Adamsky, L. Toker, I. Silman, J. W. Kelly, A. H. Futerman and J. L. Sussman, *J. Biol. Chem.*, 2005, **280**, 23815–23819.
654 B. Liou, A. Kazimierczuk, M. Zhang, C. R. Scott, R. S. Hedge and G. A. Grabowski, *J. Biol. Chem.*, 2006, **281**, 4242–4253.
655 B. Brumshtein, H. M. Greenblatt, T. D. Butters, Y. Shaaltiel, D. Aviezer, I. Silman, A. H. Futerman and J. L. Sussman, *J. Biol. Chem.*, 2007, **282**, 29052–29058.
656 Y. Kacher, B. Brumshtein, S. Boldin-Adamsky, L. Toker, A. Shainskaya, I. Silman, J. L. Sussman and A. H. Futerman, *Biol. Chem.*, 2008, **389**, 1361–1369.
657 B. Brumshtein, M. Aguilar-Moncayo, M. I. García-Moreno, C. Ortiz Mellet, J. M. García Fernández, I. Silman, Y. Shaaltiel, D. Aviezer, J. L. Sussman and A. H. Futerman, *ChemBioChem*, 2009, **10**, 1480–1485.
658 B. Brumshtein, M. Aguilar-Moncayo, J. M. Benito, J. M. García Fernandez, I. Silman, Y. Shaaltiel, D. Aviezer, J. L. Sussman, A. H. Futerman and C. Ortiz Mellet, *Org. Biomol. Chem.*, 2011, **9**, 4160.
659 R. R. Wei, H. Hughes, S. Boucher, J. J. Bird, N. Guziewicz, S. M. Van Patten, H. Qiu, C. Q. Pan and T. Edmunds, *J. Biol. Chem.*, 2011, **286**, 299–308.
660 S. D. Orwig, Y. L. Tan, N. P. Grimster, Z. Yu, E. T. Powers, J. W. Kelly and R. L. Lieberman, *Biochemistry*, 2011, **50**, 10647–10657.
661 J. Zheng, L. Chen, O. S. Skinner, D. Ysselstein, J. Remis, P. Lansbury, R. Skerlj, M. Mrosek, U. Heunisch, S. Krapp, J. Charrow, M. Schwake, N. L. Kelleher, R. B. Silverman and D. Krainc, *J. Am. Chem. Soc.*, 2018, **140**, 5914–5924.
662 M. Martínez-Bailén, A. T. Carmona, A. C. Patterson-Orazem, R. L. Lieberman, D. Ide, M. Kubo, A. Kato, I. Robina and A. J. Moreno-Vargas, *Bioorg. Chem.*, 2019, **86**, 652–664.
663 R. J. Rowland, L. Wu, F. Liu and G. J. Davies, *Acta Crystallogr. Sect. D Struct. Biol.*, 2020, **76**, 565–580.
664 G. J. Davies, R. J. Rowland, Y. Chen, I. Breen, L. Wu, W. A. Offen, T. Beenakker, Q. Su, A. M. C. H. van den Nieuwendijk, J. M. F. G. Aerts, M. Artola and H. S. Overkleeft, *Chem. – A Eur. J.*, , DOI:10.1002/chem.202102359.
665 D. J. Vocadlo and C. R. Bertozzi, *Angew. Chemie Int. Ed.*, 2004, **43**, 5338–5342.
666 B. P. Rempel, M. B. Tropak, D. J. Mahuran and S. G. Withers, *Angew. Chemie Int. Ed.*, 2011, **50**, 10381–10383.
667 O. Hekmat, Y.-W. Kim, S. J. Williams, S. He and S. G. Withers, *J. Biol. Chem.*, 2005, **280**, 35126–35.
668 K. A. Stubbs, A. Scaffidi, A. W. Debowski, B. L. Mark, R. V Stick and D. J. Vocadlo, *J. Am. Chem. Soc.*, 2008, **130**, 327–335.
669 T. M. Gloster, R. Madsen and G. J. Davies, *Org. Biomol. Chem.*, 2007, **5**, 444–446.
670 G. A. Grabowski, K. Osiecki-Newman, T. Dinur, D. Fabbro, G. Legler, S. Gatt and R. J. Desnick, *J. Biol. Chem.*, 1986, **261**, 8263–8269.
671 G. Speciale, A. J. Thompson, G. J. Davies and S. J. Williams, *Curr. Opin. Struct. Biol.*, 2014, **28,** 1–13.

672 J. N. Kanfer, G. Legler, J. Sullivan, S. S. Raghavan and R. A. Mumford, *Biochem. Biophys. Res. Commun.*, 1975, **67**, 85–90.
673 S. C. Datta and N. S. Radin, *Lipids*, 1986, **21**, 702–9.
674 J. Marshall, K. A. McEachern, J. A. C. Kyros, J. B. Nietupski, T. Budzinski, R. J. Ziegler, N. S. Yew, J. Sullivan, A. Scaria, N. van Rooijen, J. A. Barranger and S. H. Cheng, *Mol. Ther.*, 2002, **6**, 179–89.
675 Y.-H. Xu, R. Reboulet, B. Quinn, J. Huelsken, D. Witte and G. A. Grabowski, *Mol. Genet. Metab.*, 2008, **94**, 190–203.
676 V. L. Tybulewicz, M. L. Tremblay, M. E. LaMarca, R. Willemsen, B. K. Stubblefield, S. Winfield, B. Zablocka, E. Sidransky, B. M. Martin and S. P. Huang, *Nature*, 1992, **357**, 407–10.
677 W. M. Holleran, E. I. Ginns, G. K. Menon, J. U. Grundmann, M. Fartasch, C. E. McKinney, P. M. Elias and E. Sidransky, *J. Clin. Invest.*, 1994, **93**, 1756–64.
678 T. Farfel-Becker, E. B. Vitner and A. H. Futerman, *Dis. Model. Mech.*, 2011, **4**, 746–52.
679 Y. Liu, K. Suzuki, J. D. Reed, A. Grinberg, H. Westphal, A. Hoffmann, T. Döring, K. Sandhoff and R. L. Proia, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 2503–8.
680 T. Doering, R. L. Proia and K. Sandhoff, *FEBS Lett.*, 1999, **447**, 167–170.
681 H. Mizukami, Y. Mi, R. Wada, M. Kono, T. Yamashita, Y. Liu, N. Werth, R. Sandhoff, K. Sandhoff and R. L. Proia, *J. Clin. Invest.*, 2002, **109**, 1215–21.
682 Y.-H. Xu, B. Quinn, D. Witte and G. A. Grabowski, *Am. J. Pathol.*, 2003, **163**, 2093–2101.
683 B. Rigat and D. Mahuran, *Mol. Genet. Metab.*, 2009, **96**, 225–32.
684 K. A. McEachern, J. B. Nietupski, W.-L. Chuang, D. Armentano, J. Johnson, E. Hutto, G. A. Grabowski, S. H. Cheng and J. Marshall, *J. Gene Med.*, 2006, **8**, 719–729.
685 R. Kühn and R. M. Torres, in *Transgenesis Techniques*, Humana Press, New Jersey, pp. 175–204.
686 I. B. Enquist, C. Lo Bianco, A. Ooka, E. Nilsson, J.-E. Månsson, M. Ehinger, J. Richter, R. O. Brady, D. Kirik and S. Karlsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 17483–8.
687 T. Farfel-Becker, E. B. Vitner, S. N. R. Pressey, R. Eilam, J. D. Cooper and A. H. Futerman, *Hum. Mol. Genet.*, 2011, **20**, 1375–86.
688 E. B. Vitner, H. Dekel, H. Zigdon, T. Shachar, T. Farfel-Becker, R. Eilam, S. Karlsson and A. H. Futerman, *Hum. Mol. Genet.*, 2010, **19**, 3583–90.
689 K. L. Jackson, C. Viel, J. Clarke, J. Bu, M. Chan, B. Wang, L. S. Shihabuddin and S. P. Sardi, *Neurobiol. Dis.*, 2019, **130**, 104513.
690 F. Wahid, A. Shehzad, T. Khan and Y. Y. Kim, *Biochim. Biophys. Acta - Mol. Cell Res.*, 2010, 1803, 1231–1243.
691 W. J. Hartley and W. F. Blakemore, *Vet. Pathol.*, 1973, **10**, 191–201.
692 N. S. Van De Water, R. D. Jolly and B. R. Farrow, *Aust. J. Exp. Biol. Med. Sci.*, 1979, **57**, 551–4.
693 L. Karageorgos, M. J. Lancaster, J. S. Nimmo and J. J. Hopwood, *J. Inherit. Metab. Dis.*, 2011, **34**, 209–15.
694 N. K. Polinski, T. N. Martinez, A. Gorodinsky, R. Gareus, M. Sasner, M. Herberth, R. Switzer, S. O. Ahmad, M. Cosden, M. Kandebo, R. E. Drolet, P. D. Buckett, W. Shan, Y. Chen, L. J. Pellegrino, G. D. Ellsworth, L. B. Dungan, W. D. Hirst, S. W. Clark and K. D. Dave, *PLoS One*, 2021, **16**, e0252325.
695 E. Clarke, C. Jantrachotechatchawan, Y. Buhidma, M. Broadstock, L. Yu, D. Howlett, D. Aarsland, C. Ballard and P. T. Francis, *Neurochem. Int.*, 2019, **129**, 104502.
696 M. A. S. Baptista, K. D. Dave, N. P. Sheth, S. N. De Silva, K. M. Carlson, Y. N. Aziz, B. K. Fiske, T. B. Sherer and M. A. Frasier, *Dis. Model. Mech.*, 2013, **6**, 1316–1324.
697 Y. Shou-jun, G. Su-guo, Z. Yu-cheng and Z. Shu-zheng, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, 1985, **828**, 236–240.
698 C. M. Ridley, K. E. Thur, J. Shanahan, N. B. Thillaiappan, A. Shen, K. Uhl, C. M. Walden, A. A. Rahim, S. N. Waddington, F. M. Platt and A. C. Van Der Spoel, *J. Biol. Chem.*, 2013, **288**, 26052–66.
699 A. Hara, *Biochim. Biophys. Acta - Gen. Subj.*, 1979, **582**, 423–433.
700 A. K. Yadav, D. L. Shen, X. Shan, X. He, A. R. Kermode and D. J. Vocadlo, *J. Am. Chem. Soc.*, 2015, **137**, 1181–1189.
701 R. Charoenwattanasatien, S. Pengthaisong, I. Breen, R. Mutoh, S. Sansenya, Y. Hua, A. Tankrathok, L. Wu, C. Songsiriritthigul, H. Tanaka, S. J. Williams, G. J. Davies, G. Kurisu and J. R. K. Cairns, *ACS Chem. Biol.*, 2016, **11**, 1891–900.
702 M. C. Stephens, A. Bernatsky, V. Burachinsky, G. Legler and J. N. Kanfer, *J. Neurochem.*, 1978, **30**, 1023–7.
703 M. Zoidl, A. Wolfsgruber, M. Schalli, S. A. Nasseri, P. Weber, A. E. Stütz, S. G. Withers and T. M. Wrodnigg, *Monatshefte für Chemie - Chem. Mon.*, 2019, **150**, 831–842.

704 S. P. Schröder, L. Wu, M. Artola, T. Hansen, W. A. Offen, M. J. Ferraz, K. Y. Li, J. M. F. G. Aerts, G. A. Van Der Marel, J. D. C. Codée, G. J. Davies and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2018, **140**, 5045–5048.
705 E. D. Goddard-Borger, M. B. Tropak, S. Yonekawa, C. Tysoe, D. J. Mahuran and S. G. Withers, *J. Med. Chem.*, 2012, **55**, 2737–2745.
706 J. D. Diot, I. G. Moreno, G. Twigg, C. O. Mellet, K. Haupt, T. D. Butters, J. Kovensky and S. G. Gouin, *J. Org. Chem.*, 2011, **76**, 7757–7768.
707 T. Hill, M. B. Tropak, D. Mahuran and S. G. Withers, *ChemBioChem*, 2011, **12**, 2151–2154.
708 J. J. Marugan, W. Zheng, O. Motabar, N. Southall, E. Goldin, W. Westbroek, B. K. Stubblefield, E. Sidransky, R. A. Aungst, W. A. Lea, A. Simeonov, W. Leister and C. P. Austin, *J. Med. Chem.*, 2011, **54**, 1033–1058.
709 D. P. Goldman, A. E. Clarke and A. M. Garber, *Int. J. Technol. Assess. Health Care*, 1992, **8**, 583–597.
710 S. Fabrega, P. Durand, P. Codogno, C. Bauvy, A. Delomenie, B. Henrissat, B. M. Martin, C. McKinney, E. I. Ginns, J. Mornon and P. Lehn, *Glycobiology*, 2000, **10**, 1217–1224.
711 G. A. Grabowski, W. R. White and M. E. Grace, *Enzyme*, 1989, **41**, 131–42.
712 J. Limkul, S. Lizuka, Y. Sato, R. Misaki, T. Ohashi, T. Ohashi and K. Fujiyama, *Plant Biotechnol. J.*, 2016, **14**, 1682–1694.
713 G. Sinclair and F. Y. M. Choy, *Protein Expr. Purif.*, 2002, **26**, 96–105.
714 B. M. Martin, S. Tsuji, M. E. LaMarca, K. Maysak, W. Eliason and E. I. Ginns, *DNA*, 1988, **7**, 99–106.
715 A. R. Sawkar, M. Schmitz, K. P. Zimmer, D. Reczek, T. Edmunds, W. E. Balch and J. W. Kelly, *ACS Chem. Biol.*, 2006, **1**, 235–251.
716 J. W. Jung, H. Y. Choi, N. X. Huy, H. Park, H. H. Kim, M. S. Yang, S. H. Kang, D. Il Kim and N. S. Kim, *Protein Expr. Purif.*, 2019, **158**, 81–88.
717 U. Naphatsamon, T. Ohashi, R. Misaki and K. Fujiyama, *Int. J. Mol. Sci.*, 2018, **19**, 1972.
718 B. Shanmugaraj, C. J. I. Bulaon and W. Phoolcharoen, *Plants,* 2020, **9**, 842.
719 S. K. Gupta and P. Shukla, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 10457–10468.
720 R. Strasser, J. Stadlmann, B. Svoboda, F. Altmann, J. Glössl and L. Mach, *Biochem. J.*, 2005, **387**, 385–391.
721 N. Uthailak, H. Kajiura, R. Misaki and K. Fujiyama, *Front. Plant Sci.*, 2021, **0**, 1039.
722 R. J. Clem and A. L. Passarelli, *PLoS Pathog.*, 2013, **9**, 1003729.
723 G. W. Blissard, *Cytotechnology*, 1996, **20**, 73-93
724 M. Jurkovíčová, J. H. van Touw, J. S. Sussenbach and J. Ter Schegget, *Virology*, 1979, **93**, 8–19.
725 M. D. Summers, D. L. Anderson, I. P. Kok, A. V Chistyakova, A. P. Gudz-Gorban, A. P. Solomko and A. Int, *Journal of Virology* , 1973, **12**, 127
726 J. Slack and B. M. Arif, *Advances in Virus Research*, Elsevier Inc., Ontario, 2006, pp. 99–165.
727 L. E. Volkman and B. A. Keddie, *Semin. Virol.*, 1990, **1**, 249–256.
728 S. A. B. Miele, M. J. Garavaglia, M. N. Belaich and P. D. Ghiringhelli, *Int. J. Evol. Biol.*, 2011, **2011**, 1–15.
729 Z. Nv Yang, *J. Biotechnol. Biomater.*, 2012, **2**, 1–7.
730 M. Ikeda, R. Hamajima and M. Kobayashi, *Entomol. Sci.*, 2015, **18**, 1–20.
731 L. Jiang and Q. Xia, *Insect Biochem. Mol. Biol.*, 2014, **48**, 1–7.
732 E. A. Herniou, J. A. Olszewski, J. S. Cory and D. R. O'Reilly, *Annu. Rev. Entomol.*, 2003, 48, 211–234.
733 G. H. Bergold, *Can. J. Res.*, 1950, **28e**, 5–11.
734 F. T. Bird and M. M. Whalen, *Can. J. Microbiol.*, 1954, **1**, 170–174.
735 F. T. Bird, *Biochim. Biophys. Acta*, 1952, **8**, 360–368.
736 H. W. Ackermann and W. A. Smirnoff, *J. Invertebr. Pathol.*, 1983, **41**, 269–280.
737 F. T. Bird, *Virology*, 1957, **3**, 237–242.
738 E. V. Subbaiah, C. Royer, S. Kanginakudru, V. V. Satyavathi, A. S. Babu, V. Sivaprasad, G. Chavancy, M. Darocha, A. Jalabert, B. Mauchamp, I. Basha, P. Couble and J. Nagaraju, *Genetics*, 2013, **193**, 63–75.
739 S. A. Monsma, A. G. P. Oomens and G. W. Blissard, *Journal of Virology*, 1996, **70**, 4607-4616.
740 B. A. Keddie, G. W. Aponte and L. E. Volkman, *Science (80-.).*, 1989, **243**, 1728–1730.
741 R. R. Granados and K. A. Lawler, *Virology*, 1981, **108**, 297–308.
742 A. G. P. Oomens, S. A. Monsma and G. W. Blissard, *Virology*, 1995, **209**, 592–603.
743 Y. Katou, M. Ikeda and M. Kobayashi, *Virology*, 2006, **347**, 455–465.
744 B. J. L. Hooft van Iddekinge, G. E. Smith and M. D. Summers, *Virology*, 1983, **131**, 561–565.
745 M. M. van Oers, J. T. M. Flipsen, C. B. E. M. Reusken and J. M. Vlak, *Virology*, 1994, **200**, 513–523.

746 F. Moscardi, *Annu. Rev. Entomol.*, 1999, **44**, 257–289.
747 B. Szewczyk, L. Rabalski, E. Krol, W. Sihler and M. L. De Souza, *J. Biopestic.* 2009, **2**, 209-216
748 S. Liu, H. Li, S. Sivakumar and B. C. Bonning, in *Advances in Virus Research*, 2006, **68**, 427–457.
749 C.-Y. Chen, C.-Y. Lin, G.-Y. Chen and Y.-C. Hu, *Biotechnol. Adv.*, 2011, **29**, 618–631.
750 J. P. Condreay, R. S. Ames, N. J. Hassan, T. A. Kost, R. V. Merrihew, D. E. Mossakowska, D. J. Pountney and M. A. Romanos, in *Advances in Virus Research*, 2006, **68**, 255–286.
751 M. M. van Oers, *Adv. Virus Res.*, 2006, **68**, 193–253.
752 C. C. Lai, Y. C. Cheng, P. W. Chen, T. H. Lin, T. T. Tzeng, C. C. Lu, M. S. Lee and A. Y. C. Hu, *J. Biol. Eng.*, 2019, **13**, 78.
753 I. Berger, D. J. Fitzgerald and T. J. Richmond, *Nat. Biotechnol.*, 2004, **22**, 1583–1587.
754 C. Bieniossek, T. Imasaki, Y. Takagi and I. Berger, *Trends Biochem. Sci.*, 2012, 37, 49–57.
755 D. Sari, K. Gupta, D. B. T. G. Raj, A. Aubert, P. Drncová, F. Garzoni, D. Fitzgerald and I. Berger, in *Advances in Experimental Medicine and Biology*, Springer New York LLC, 2016, vol. 896, pp. 199–215.
756 D. J. Fitzgerald, P. Berger, C. Schaffitzel, K. Yamada, T. J. Richmond and I. Berger, *Nat. Methods*, 2006, **3**, 1021–1032.
757 T. Li, Q. Zheng, H. Yu, D. Wu, W. Xue, H. Xiong, X. Huang, M. Nie, M. Yue, R. Rong, S. Zhang, Y. Zhang, Y. Wu, S. Wang, Z. Zha, T. Chen, T. Deng, Y. Wang, T. Zhang, Y. Chen, Q. Yuan, Q. Zhao, J. Zhang, Y. Gu, S. Li and N. Xia, *Emerg. Microbes Infect.*, 2020, **9**, 2076–2090.
758 O. Khow and S. Suntrarachun, *Asian Pac. J. Trop. Biomed.*, 2012, **2**, 159–162.
759 A. C. Chambers, M. Aksular, L. P. Graves, S. L. Irons, R. D. Possee and L. A. King, *Curr. Protoc. Protein Sci.*, 2018, **91**, 541-546.
760 B. C. Bonning and B. D. Hammock, *Annu. Rev. Entomol.*, 1996, **41**, 191–210.
761 I. R. Cameron, R. D. Possee and D. H. L. Bishop, *Trends Biotechnol.*, 1989, 7, 66–70.
762 V. A. Luckow and M. D. Summers, *Nat. Biotechnol.*, 1988, **6**, 47–55.
763 R. D. Possee, *Curr. Opin. Biotechnol.*, 1997, **8**, 569–572.
764 G. E. Smith, M. J. Fraser and M. D. Summers, *J. Virol.*, 1983, **46**, 584–593.
765 B. Gorda, C. Toelzer, F. Aucilino and I. Berger, in *Methods in Enzymology*, 2021, pp. 2–23.
766 M. D. Ayres, S. C. Howard, M. Lopez-Ferber, M. Lopez-Ferber and R. D. Possee, *Virology*, 1994, **202**, 586–605.
767 D. L. Jarvis, *Virology*, 2003, 310, 1–7.
768 D. L. Jarvis, C. Weinkauf and L. A. Guarino, *Protein Expr. Purif.*, 1996, **8**, 191–203.
769 G. W. Blissard and G. F. Rohrmann, *Annu. Rev. Entomol.*, 1990, **35**, 127–155.
770 G. E. Smith, M. D. Summers and M. J. Fraser, *Mol. Cell. Biol.*, 1983, **3**, 2156–2165.
771 G. D. Pennock, C. Shoemaker and L. K. Miller, *Mol. Cell. Biol.*, 1984, **4**, 399–406.
772 M. M. van Oers, G. P. Pijlman and J. M. Vlak, *J. Gen. Virol.*, 2015, **96**, 6–23.
773 D. R. O'Reilly, L. K. Miller and V. A. Luckow, *Baculovirus Expression Vectors: A Laboratory Manual*, Oxford University Press, New York, Oxford, 1994.
774 V. A. Luckow, S. C. Lee, G. F. Barry and P. O. Olins, *J. Virol.*, 1993, **67**, 4566–4579.
775 J. Zitzmann, G. Sprick, T. Weidner, C. Schreiber and P. Czermak, in *New Insights into Cell Culture Technology*, InTech, 2017.
776 T. D. C. Grace, *Nature*, 1962, **195**, 788–789.
777 D. E. Lynn and R. L. Harrison, in *Methods in Molecular Biology*, Humana Press Inc., 2016, **1350**, pp. 119–142.
778 M. Wilde, M. Klausberger, D. Palmberger, W. Ernst and R. Grabherr, *Biotechnol. Lett.*, 2014, **36**, 743–749.
779 D. L. Jarvis, J. G. W. Fleming, G. R. Kovacs, M. D. Summers and L. A. Guarino, *Bio/Technology*, 1990, **8**, 950–955.
780 A. Contreras-Gómez, A. Sánchez-Mirón, F. García-Camacho, E. Molina-Grima and Y. Chisti, *Biotechnol. Prog.*, 2014, **30**, 1–18.
781 N. Chen, X. Kong, S. Zhao and W. Xiaofeng, *Virus Res.*, 2020, **279**, 197865.
782 J. Yin, G. Li, X. Ren and G. Herrler, *J. Biotechnol.*, 2007, 127, 335–347.
783 T. A. Kost, J. P. Condreay and D. L. Jarvis, *Nat. Biotechnol.*, 2005, **23**, 567–575.
784 R. L. Harrison and D. L. Jarvis, *Adv. Virus Res.*, 2006, **68**, 159–191.
785 F. Altmann, *Int. Arch. Allergy Immunol.*, 2007, **142**, 99–115.
786 J. R. Hollister and D. L. Jarvis, *Glycobiology*, 2001, **11**, 1–9.
787 J. J. Aumiller, H. Mabashi-Asazuma, A. Hillar, X. Shi and D. L. Jarvis, *Glycobiology*, 2012, **22**, 417–

428.
788 C. L. Merrington, M. J. Bailey and R. D. Possee, *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.*, 1997, **8**, 283–297.
789 A. Ceroni, A. Dell and S. M. Haslam, *Source Code Biol. Med.*, 2007, 2, 3.
790 Y. H. Xu and G. A. Grabowski, *Mol. Genet. Metab.*, 1998, **64**, 87–98.
791 F. Y. Choy, C. Wei and D. Levin, *Am. J. Med. Genet.*, 1996, **65**, 184–9.
792 K. W. Moremen, A. Ramiah, M. Stuart, J. Steel, L. Meng, F. Forouhar, H. A. Moniz, G. Gahlay, Z. Gao, D. Chapla, S. Wang, J. Y. Yang, P. K. Prabhakar, R. Johnson, M. Dela Rosa, C. Geisler, A. V. Nairn, J. Seetharaman, S. C. Wu, L. Tong, H. J. Gilbert, J. Labaer and D. L. Jarvis, *Nat. Chem. Biol.*, 2018, **14**, 156–162.
793 L. B. Jacobsen, S. A. Calvin and E. K. Lobenhofer, *Biotechniques*, 2009, **47**, 617–624.
794 J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart and A. Perrakis, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2005, **61**, 1426–1431.
795 S. M. Roberts and G. J. Davies, in *Methods in Enzymology*, Academic Press Inc., 2012, vol. 510, pp. 141–168.
796 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 235–242.
797 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
798 C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack and G. Bricogne, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 293–302.
799 F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus and G. N. Murshudov, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 103–111.
800 F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus and G. N. Murshudov, *Acta Crystallogr. Sect. D Struct. Biol.*, 2017, **73**, 112–122.
801 V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 12–21.
802 D. C. Tessier, D. Y. Thomas, H. E. Khouri, F. Laliberté and T. Vernet, *Gene*, 1991, **98**, 177–83.
803 A. V. Fonin, A. I. Sulatskaya, I. M. Kuznetsova and K. K. Turoverov, *PLoS One*, 2014, **9**, e103878.
804 S. Kumar Panigrahi and A. Kumar Mishra, *J. Photochem. Photobiol. C Photochem. Rev.*, 2019, 41, 100318.
805 Y. Kishimoto, M. Hiraiwa and J. S. O'Brien, *J. Lipid Res.*, 1992, **33**, 1255–1267.
806 A. M. Vaccaro, R. Salvioli, M. Tatti and F. Ciaffoni, *Neurochem. Res. 1999 242*, 1999, **24**, 307–314.
807 S. Weller, W. Carson, Y. Lee, D. B. Teplow, Y. Kishimoto, J. S O'Brien J. A. Barranger, J. M. Tomich, *J. Mol. Neurosci.*, 1993, **4**, 161–172.
808 E. de Alba, S. Weiler and N. Tjandra, *Biochemistry*, 2003, **42**, 14729–14740.
809 J.-R. Alattia, J. E. Shaw, C. M. Yip and G. G. Privé, *Proc. Natl. Acad. Sci.*, 2007, **104**, 17394–17399.
810 K. Hancock, S. Narang, S. Pattabhi, M. L. Yushak, A. Khan, S. ching Lin, R. Plemons, M. J. Betenbaugh and V. C. W. Tsang, *J. Immunol. Methods*, 2008, **330**, 130–136.
811 R. Romero, A. Ramanathan, T. Yuen, D. Bhowmik, M. Mathew, L. B. Munshi, S. Javaid, M. Bloch, D. Lizneva, A. Rahimova, A. Khan, C. Taneja, S.-M. Kim, L. Sun, M. I. New, S. Haider and M. Zaidi, *Proc. Natl. Acad. Sci.*, 2019, **116**, 5086–5095.
812 R. Romero, A. Ramanathan, T. Yuen, D. Bhowmik, M. Mathew, L. B. Munshi, S. Javaid, M. Bloch, D. Lizneva, A. Rahimova, A. Khan, C. Taneja, S.-M. Kim, L. Sun, M. I. New, S. Haider and M. Zaidi, *Proc. Natl. Acad. Sci.*, 2019, **116**, 5086–5095.
813 A. J. Thompson, R. J. Williams, Z. Hakki, D. S. Alonzi, T. Wennekes, T. M. Gloster, K. Songsrirote, J. E. Thomas-Oates, T. M. Wrodnigg, J. Spreitz, A. E. Stutz, T. D. Butters, S. J. Williams and G. J. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 781–786.
814 J. S. Brunzelle, D. B. Jordan, D. R. McCaslin, A. Olczak and Z. Wawrzak, *Arch. Biochem. Biophys.*, 2008, **474**, 157–166.
815 J. Hoseki, R. Ushioda and K. Nagata, *J. Biochem.*, 2010, **147**, 19–25.
816 P. J. Thomas, B. H. Qu and P. L. Pedersen, *Trends Biochem. Sci.*, 1995, **20**, 456–459.
817 N. Gregersen, P. Bross, M. M. Jørgensen, T. J. Corydon and B. S. Andresen, *J. Inherit. Metab. Dis.*, 2000, **23**, 441–447.
818 J. D. Malhotra and R. J. Kaufman, *Semin. Cell Dev. Biol.*, 2007, **18**, 716–731.

819 C. Hetz, K. Zhang and R. J. Kaufman, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 421–438.
820 G. Parenti, G. Andria and A. Ballabio, *Annu. Rev. Med.*, 2015, **66**, 471–486.
821 J. Mitsui, I. Mizuta, A. Toyoda, R. Ashida, Y. Takahashi, J. Goto, Y. Fukuda, H. Date, A. Iwata, Y. M., N. Hattori, M. Murata and T. Toda, *Arch Neurol.*, 2009, **66**, 571–576.
822 E. Sidransky and G. Lopez, *Lancet Neurol.*, 2012, **11**, 986–998.
823 I. Stojkovska, D. Krainc and J. R. Mazzulli, *Cell Tissue Res.*, 2018, **373**, 51–60.
824 F. Richter, S. M. Fleming, M. Watson, V. Lemesre, L. Pellegrino, B. Ranes, C. Zhu, F. Mortazavi, C. K. Mulligan, P. C. Sioshansi, S. Hean, K. De La Rosa, R. Khanna, J. Flanagan, D. J. Lockhart, B. A. Wustman, S. W. Clark and M. F. Chesselet, *Neurotherapeutics*, 2014, **11**, 840–856.
825 S. P. Sardi, J. Clarke, C. Kinnecom, T. J. Tamsett, L. Li, L. M. Stanek, M. A. Passini, G. A. Grabowski, M. G. Schlossmacher, R. L. Sidman, S. H. Cheng and L. S. Shihabuddin, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 12101–12106.
826 S. P. Sardi, J. Clarke, C. Viel, M. Chan, T. J. Tamsett, C. M. Treleaven, J. Bu, L. Sweet, M. A. Passini, J. C. Dodge, W. H. Yu, R. L. Sidman, S. H. Cheng and L. S. Shihabuddin, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 3537–3542.
827 G. Parenti, *EMBO Mol. Med.*, 2009, 1, 268–279.
828 E. Beutler, *Bailliere's Clin. Haematol. Clin. Haematol.*, 1997, **10**, 751–763.
829 W. Stoffel and A. Bosio, *Curr. Opin. Neurobiol.*, 1997, **7**, 654–661.
830 *Pharmacoeconomic Review report: Eliglustat (Cerdelga) Sanofi Genzyme* , Canadian Agency for Drugs and Technologies in Health, Ottawa, 2017.
831 M. Stark, T. F. D. Silva, G. Levin, M. Machuqueiro and Y. G. Assaraf, *Cells*, 2020, **9**, 1082.
832 Y. Sun, B. Liou, Y. H. Xu, B. Quinn, W. Zhang, R. Hamler, K. D. R. Setchell and G. A. Grabowski, *J. Biol. Chem.*, 2012, **287**, 4275–4287.
833 S. Front, A. Biela-Banaś, P. Burda, D. Ballhausen, K. Higaki, A. Caciotti, A. Morrone, J. Charollais-Thoenig, E. Gallienne, S. Demotz and O. R. Martin, *Eur. J. Med. Chem.*, 2017, **126**, 160–170.
834 L. M. Jonsson, G. J. Murray, S. H. Sorrell, A. Strijland, J. F. Aerts, E. I. Ginns, J. A. Barranger, J. M. Tager and A. W. Schram, *Eur. J. Biochem.*, 1987, **164**, 171–179.
835 S. Cecioni and D. J. Vocadlo, *J. Am. Chem. Soc.*, 2017, **139**, 8392–8395.
836 C. Adamson, R. J. Pengelly, S. Shamsi Kazem Abadi, S. Chakladar, J. Draper, R. Britton, T. M. Gloster and A. J. Bennet, *Angew. Chemie Int. Ed.*, 2016, **55**, 14978–14982.
837 W. Ren, R. Pengelly, M. Farren-Dai, S. S. K. Abadi, V. Oehler, O. Akintola, J. Draper, M. Meanwell, S. Chakladar, K. Świderek, V. Moliner, R. Britton, T. M. Gloster and A. J. Bennet, *Nat. Commun. 2018 91*, 2018, **9**, 1–12.
838 S. Shamsi Kazem Abadi, M. Tran, A. K. Yadav, P. J. P. Adabala, S. Chakladar and A. J. Bennet, *J. Am. Chem. Soc.*, 2017, **139**, 10625–10628.
839 S. P. Yun, D. Kim, S. Kim, S. Kim, S. S. Karuppagounder, S.-H. Kwon, S. Lee, T.-I. Kam, S. Lee, S. Ham, J. H. Park, V. L. Dawson, T. M. Dawson, Y. Lee and H. S. Ko, *Mol. Neurodegener.*, 2018, **13**, 1.
840 A. Migdalska-Richards, M. Wegrzynowicz, R. Rusconi, G. Deangeli, D. A. Di Monte, M. G. Spillantini and A. H. V Schapira, *Brain*, 2017, **140**, 2706.
841 Y. Chen, Z. Armstrong, M. Artola, B. I. Florea, C.-L. Kuo, C. de Boer, M. S. Rasmussen, M. A. Hachem, G. A. van der Marel, J. D. C. Codée, J. M. F. G. Aerts, G. J. Davies and H. S. Overkleeft, *J. Am. Chem. Soc.*, 2021, **143**, 2423–2432.
842 W. SJ and G.-B. ED, *Biochem. Soc. Trans.*, 2020, **48**, 1287–1295.
843 W. SJ and G.-B. ED, *Biochem. Soc. Trans.*, 2020, **48**, 1287–1295.
844 H. Reyes, Y. Du, T. Zhou, X. Xie, P. Y. Shi, S. Weiss and T. M. Block, *bioRxiv*, 2021, 2021.05.14.444190.
845 S. Rajasekharan, R. M. Bonotto, L. N. Alves, Y. Kazungu, M. Poggianella, P. Martinez-Orellana, N. Skoko, S. Polez and A. Marcello, *Viruses*, 2021, **13**, 808.
846 H. Qu, J. M. Liu, J. Wdzieczak-Bakala, D. Lu, X. He, W. Sun, M. Sollogoub and Y. Zhang, *Eur. J. Med. Chem.*, 2014, **75**, 247–257.
847 T. W. Chung, S. J. Kim, H. J. Choi, K. J. Kim, M. J. Kim, S. H. Kim, H. J. Lee, J. H. Ko, Y. C. Lee, A. Suzuki and C. H. Kim, *Glycobiology*, 2009, **19**, 229–239.
848 K. Yuyama, H. Sun, S. Sakai, S. Mitsutake, M. Okada, H. Tahara, J. I. Furukawa, N. Fujitani, Y. Shinohara and Y. Igarashi, *J. Biol. Chem.*, 2014, **289**, 24488–24498.
849 T. Mutoh, Y. Hirabayashi, T. Mihara, M. Ueda, H. Koga, A. Ueda, T. Kokura and H. Yamamoto, *CNS Neurol. Disord. - Drug Targets*, 2008, **5**, 375–380.
850 M. Amaro, R. Šachl, G. Aydogan, I. I. Mikhalyov, R. Vácha and M. Hof, *Angew. Chemie Int. Ed.*,

2016, **55**, 9411–9415.
851 P. M. Danby and S. G. Withers, *ACS Chem. Biol.*, 2016, **11**, 1784–1794.
852 F. A. Shaikh and S. G. Withers, *Biochem. Cell Biol.*, 2008, **86**, 169–177.
853 L.-X. Wang and W. Huang, *Curr. Opin. Chem. Biol.*, 2009, **13**, 592–600.
854 L. F. Mackenzie, Q. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
855 S. J. Williams and S. G. Withers, *Aust. J. Chem.*, 2002, **55**, 3.
856 B. Rakić and S. G. Withers, *Aust. J. Chem.*, 2009, **62**, 510.
857 M. D. Vaughan, K. Johnson, S. DeFrees, X. Tang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 2006, **128**, 6300–6301.
858 J. R. Rich, A.-M. Cunningham, M. Gilbert and S. G. Withers, *Chem. Commun.*, 2011, **47**, 10806.
859 A. A. Marques, M. Mirzaian, H. Akiyama, P. Wisse, M. J. Ferraz, P. Gaspar, K. Ghauharali-van der Vlugt, R. Meijer, P. Giraldo, P. Alfonso, P. Irún, M. Dahl, S. Karlsson, E. V. Pavlova, T. M. Cox, S. Scheij, M. Verhoek, R. Ottenhoff, C. A. A. van Roomen, N. S. Pannu, M. van Eijk, N. Dekker, R. G. Boot, H. S. Overkleeft, E. Blommaart, Y. Hirabayashi and J. M. Aerts, *J. Lipid Res.*, 2016, **57**, 451–463.
860 H. Akiyama, S. Kobayashi, Y. Hirabayashi and K. Murakami-Murofushi, *Biochem. Biophys. Res. Commun.*, 2013, **441**, 838–843.
861 D. E. Boer, M. Mirzaian, M. J. Ferraz, K. C. Zwiers, M. V. Baks, M. D. Hazeu, R. Ottenhoff, A. R. A. Marques, R. Meijer, J. C. P. Roos, T. M. Cox, R. G. Boot, N. Pannu, H. S. Overkleeft, M. Artola and J. M. Aerts, *J. Lipid Res.*, 2021, **62**, 100018.
862 Q. Su, S. P. Schröder, L. T. Lelieveld, M. J. Ferraz, M. Verhoek, R. G. Boot, H. S. Overkleeft, J. M. F. G. Aerts, M. Artola and C.-L. Kuo, *ChemBioChem*, 2021, **22**, 1–10.
863 R. Danev and W. Baumeister, *Elife*.
864 H. W. Wang and X. Fan, *Curr. Opin. Struct. Biol.*, 2019, **58**, 175–182.
865 S. P. Schröder, C. de Boer, N. G. S. McGregor, R. J. Rowland, O. Moroz, E. Blagova, J. Reijngoud, M. Arentshorst, D. Osborn, M. D. Morant, E. Abbate, M. A. Stringer, K. B. R. M. Krogh, L. Raich, C. Rovira, J.-G. Berrin, G. P. van Wezel, A. F. J. Ram, B. I. Florea, G. A. van der Marel, J. D. C. Codée, K. S. Wilson, L. Wu, G. J. Davies and H. S. Overkleeft, *ACS Cent. Sci.*, 2019, **5**, 1067–1078.
866 C. de Boer, N. G. S. McGregor, E. Peterse, S. P. Schröder, B. I. Florea, J. Jiang, J. Reijngoud, A. F. J. Ram, G. P. van Wezel, G. A. van der Marel, J. D. C. Codée, H. S. Overkleeft and G. J. Davies, *RSC Chem. Biol.*, 2020, **1**, 148–155.
867 N. G. S. McGregor, M. Artola, A. Nin-Hill, D. Linzel, M. Haon, J. Reijngoud, A. Ram, M.-N. Rosso, G. A. van der Marel, J. D. C. Codée, G. P. van Wezel, J.-G. Berrin, C. Rovira, H. S. Overkleeft and G. J. Davies, *J. Am. Chem. Soc.*, 2020, **142**, 4648–4662.
868 O. Kirk, T. V. Borchert and C. C. Fuglsang, *Curr. Opin. Biotechnol.*, 2002, **13**, 345–351.
869 A. Sundarram and T. P. K. Murthy, *J. Appl. Environ. Microbiol.*, 2014, **2**, 166–175.
870 M. J. E. C. Van Der Maarel, B. Van Der Veen, J. C. M. Uitdehaag, H. Leemhuis and L. Dijkhuizen, *J. Biotechnol.*, 2002, **94**, 137–155.
871 A. D. Mooradian and J. E. Thurman, *Drugs 1999 571*, 2012, **57**, 19–29.
872 J.-L. Chiasson, R. G. Josse, R. Gomis, M. Hanefeld, A. Karasik and M. Laakso, *Lancet*, 2002, **359**, 2072–2077.
873 L. J. Scott and C. M. Spencer, *Drugs*, 2000, **59**, 521–549.
874 C. A. Tarling, K. Woods, R. Zhang, H. C. Brastianos, G. D. Brayer, R. J. Andersen and S. G. Withers, *ChemBioChem*, 2008, **9**, 433–438.
875 N. Jain, M. A. Attia, W. A. Offen, G. J. Davies and H. Brumer, *Org. Biomol. Chem.*, 2019, **17**, 398–398.
876 S. J. Williams, O. Hekmat and S. G. Withers, *ChemBioChem*, 2006, **7**, 116–124.
877 M. N. Gandy, A. W. Debowski and K. A. Stubbs, *Chem. Commun.*, 2011, **47**, 5037.
878 S. M. Wilkinson, C. W. Liew, J. P. Mackay, H. M. Salleh, S. G. Withers and M. D. McLeod, *Org. Lett.*, 2008, **10**, 1585–1588.