CHAPTER 7

β-Galactose Configured Cyclophellitol Aziridine as Activity-Based Probes for Retaining exo-β-Galactosidases

Based on:

Kuo CL, Beenakker TJM, Profijt R, Marques ARA, Groenewegen N, Offen WA, Sarris AJC, Boot RG, Codée JDC, van der Marel GA, Davies GJ, Overkleeft HS & Aerts JMFG. *To be submitted.*

ABSTRACT

The enzymes β-galactosidase (GLB1) and galactocerebrosidase (GALC) are both retaining exo-glycosidases important in glycoconjugate metabolism, and their hereditary deficiency in man can lead to lysosomal storage disorders with still unmet medical needs. Chemical tools to study these enzymes have been developed in the past, including the cyclophellitol-based activity-based probes (ABPs) for GALC, and broad spectrum (4-deoxy) cyclophellitol aziridine ABPs for both β-glucosidases and β-galactosidases. However, β-galactose-configured N-tagged cyclophellitol aziridines—which might offer specific ABPs for β-galactosidases—have not been generated due to earlier synthetic challenges. Here, ABPs containing such scaffold are evaluated for their reactivities. The analysis shows that the ABPs exhibit expected mechanism-based inhibition and labeling of recombinant bacterial GLB1 homologue, and both GLB1 and GALC in cell lysates, culture medium, and tissue extracts. The Cy5-tagged ABP additionally labels in mouse intestine extracts the dietary enzyme lactase-phlorizin hydrolase (LPH). Pull-down experiments with biotin conjugated ABP and mouse kidney extracts, followed by LC-MS analysis identified GLB1 and GALC as major glycosidase targets, as well as two other putative β-galactosidases GLB1-lile protein 1 and 2 (GLB1L and GLB1L2) that have yet unknown biological functions. The β-galactose-configured N-tagged cyclophellitol aziridine ABPs still label β-glucosidases with considerable affinity, illustrating the limitation of this scaffold for specific β-galactosidase labeling. Nevertheless, by pre-incubating samples with β-glucosidase inhibitors, different β-galactosidases can be simultaneously visualized by gel-based ABP detection. The novel ABPs probes can be employed in future fundamental and clinical research on reactive β-galactosidases

7.1 Introduction

The monosaccharide galactose is incorporated by higher eukaryotes into a variety of glycoconjugates such as glycoproteins, glycolipids, and glycosaminoglycans (GAGs). Among the two possible galactose anomers in glycoconjugates, the predominant one is β-galactose. In humans and other mammals, β-galactose is transferred from UDP-α-galactose to specific acceptors in the Golgi apparatus by the action of dedicated β-1,3- and β-1,4-galactosyltransferases.[1] It is removed from glycoconjugates predominantly by two distinct lysosomal β-galactosidase: acid β-galactosidase (GLB1, E.C. 3.2.1.24) degrading oligosaccharides (glycosaminoglycans and glycoproteins) and gangliosides (such as GM1a and GA1),[2] and galactocerebrosidase (GALC, E.C. 3.2.1.46) degrading mainly galactosylceramide.[3]

GLB1 belongs to Glycoside Hydrolase (GH) family 35.[4] Synthesized as glycosylated 85 kDa precursor, it is targeted to the lysosome through the mannose-6-phosphate (M6P) dependent pathway; proteolytic cleavage at its C terminus results in the mature enzyme containing a large 64 kDa subunit and a small 20 kDa subunit that remain associated with each other without disulfide linkage.[5, 6] The 64 kDa subunit contains a TIM barrel domain containing the catalytic glutamates Glu188 and Glu268, while the 20 kDa subunit is required for the stabilization and functioning of the enzyme.[6, 7, 8] Similar to other lysosomal glycosidases, GLB1's catalysis of glycolipids is assisted by activator proteins, in this case both saposin B and the GM2 activator protein.[9] GLB1 forms in lysosomes a complex together with cathepsin A (PPCA, *CSTA* gene) and neuraminidase (NEU1), resulting in a stable and efficient catalytic machinery for degrading the glycosphingolipid GM1 ganglioside.[10] Alternative splicing of the *GLB1* gene produces a protein named elastin-binding protein (EBP), which does not possess catalytic activity and is instead transported to the extracellular matrix to act as a lectin that modulates elastin fiber formation.[11, 12] Another lysosomal β-galactosidase, GALC, belongs to the GH59 family.[4] It is synthesized as an 80 kDa precursor containing four N-linked glycans, and targeted either directly to the lysosomes via the M6P-mediated pathways or by re-uptake of the secreted proteins through M6P-dependent or -independent pathways.[13, 14] In the lysosome, it is processed to mature enzyme consisting of an N-terminal 50 kDa subunit (containing the TIM barrel domain harboring catalytic residues glu198 and Glu274) and a C-terminal 30 kDa subunit comprising the lectin domain bound with a calcium ion.[13, 15, 16] With the assistance by saposin A,

[17, 18] GALC specifically degrades galactosylceramide that is particularly abundant in myelin, kidney, and epithelial cells of intestine and colon.[19]

The importance of GLB1 and GALC activity is illustrated by diseases caused by their deficiency. Deficiency in GLB1 may result in three distinct lysosomal storage disorders, namely GM1 gangliosidosis, Morquio B syndrome, and galactosialidosis. GM1 gangliosidosis is caused by mutations in the *GLB1* gene and is characterized by an elevated cellular level of GM1 ganglioside, causing lysosomal vacuolization in lymphocytes and neuronal cells, which in turn leads to demyelination and neuronal cell death in the periphery as well as in the central nervous system.[20] Patients are classified into infantile (type I, OMIM # 230500), late infantile/juvenile (type two, OMIM # 230600), and adult onset forms (type III, OMIM # 230650) according to disease feature and progression.[20] Disease severity is inversely related to residual GLB1 activity.[21] Although the molecular mechanisms leading to disease pathogenesis is not completely understood, it has been suggested that abnormal accumulation of GM1 ganglioside in the membranes linking the ER and mitochondria can drive Ca^{2+} efflux from the ER to mitochondria, leading to unfolded protein response (UPR) and mitochondrial stress, followed by apoptosis and ultimately neuronal death.[22-24] Different mutations in *GLB1* gene may lead to another disease entity named Morquio B syndrome (mucopolysaccharidosis type IVB, MPSIVB, OMIM # 253010), where the primary accumulated substrates are oligosaccharides deriving from keratin sulfate and glycoproteins.[2] No central nervous system involvement is observed in these patients, and the major affected tissues is the skeletal system.[2] The third disease related to GLB1 deficiency is galactosialidosis (OMIM # 256540), and is in fact caused by primary deficiency of cathepsin A (PPCA, gene = *CTSA*) that normally forms a protein complex with GLB1 and neuraminidase (NEU1) in the lysosome. Cathepsin A deficiency causes premature degradation of both GLB1 and NEU1, leading to their secondary deficiencies in patients.[10] Deficiency in GALC or its activator protein saposin A, forms the basis of Krabbe disease (globoid cell leukodystrophy, GLD, OMIM # 245200).[3, 25] About 85 % of Krabbe disease patients develop the infantile-onset form that presents with developmental delay and severe neurological damages, with death usually within a few years.[26] Later-onset forms have higher residual GALC activity and milder symptoms, and may have life spans up to the seventh decade of life. The primarily accumulated substrate is galactosylsphingosine (psychosine) in macrophages (globoid cells) and neural cells—particularly oligodendrocytes and Schwann cells, and this cytotoxic compound is believed to cause demyelination and cell death in both the central and peripheral nervous

system.[27, 28] No therapy is yet available for these β-galactose-related lysosomal storage disorders besides bone marrow transplantation (BMT) and hematopoietic stem cell transplantation (HCT) of Krabbe disease, which is only effective before the onset of symptoms (which is usually not the case). Several therapeutic approaches are presently studied including substrate reduction therapies (SRT),[29] enzyme replacement therapies (ERT),[30, 31] pharmacological chaperon therapies (PCT),[32, 33] gene therapies,[34-36] and lysosomal re-acidification therapy[37].

Mechanistically, both GLB1 and GALC are retaining exo-glycosidases using the Koshland double-displacement catalytic mechanism. A covalent glycosidic bond is formed during the reaction itinerary, which makes both enzymes amenable to activity-based glycosidase profiling (general introduction, this thesis). In the past, activity-based probes (ABPs) based on the β-galactose configured cyclophellitol scaffold have been generated (**Fig. 7.1A**, LWA487) and enabled specific labeling of GALC in mouse tissues extracts at high micromolar ABP concentration.[38]

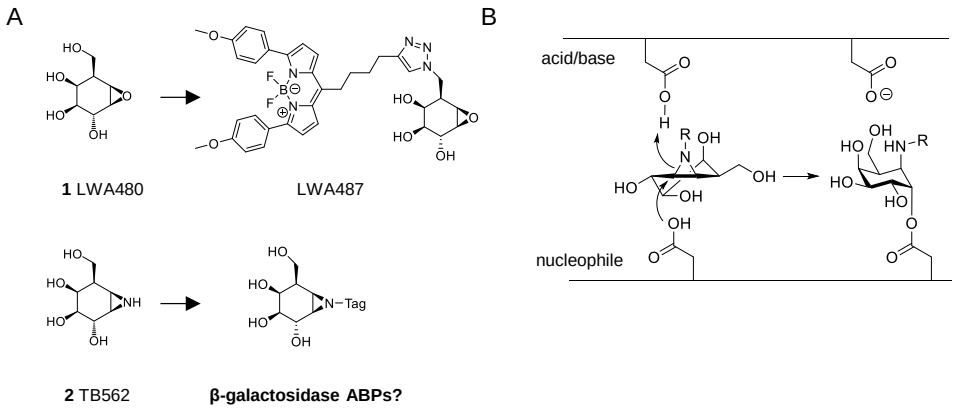


Figure 7.1 Strategies for activity-based labeling towards exo-galactosidases. A) Inhibitor and ABP for GALC (LWA480 and LWA487) and potential inhibitor and ABPs for β-galactosidases B) Proposed reaction mechanism of the β-galactose configured cyclophellitol aziridines towards β-galactosidases.–

On the other hand, ABPs based on the 4-deoxy cyclophellitol aziridine scaffold offers broad spectrum ABPs (**Fig. 7.2**, SYD215) that label both β-glucosidases and β-galactosidases.[39] To develop ABPs specifically towards β-galactosidases, attempts have been made to synthesize the N-alkylated or N-acylated β-galactose configured cyclophellitol aziridines (**Fig. 7.1A**). This was initially hampered by instability of compound intermediates during synthesis.[40] This chapter aims to study the recently available N-alkylated β-galactose configured cyclophellitol

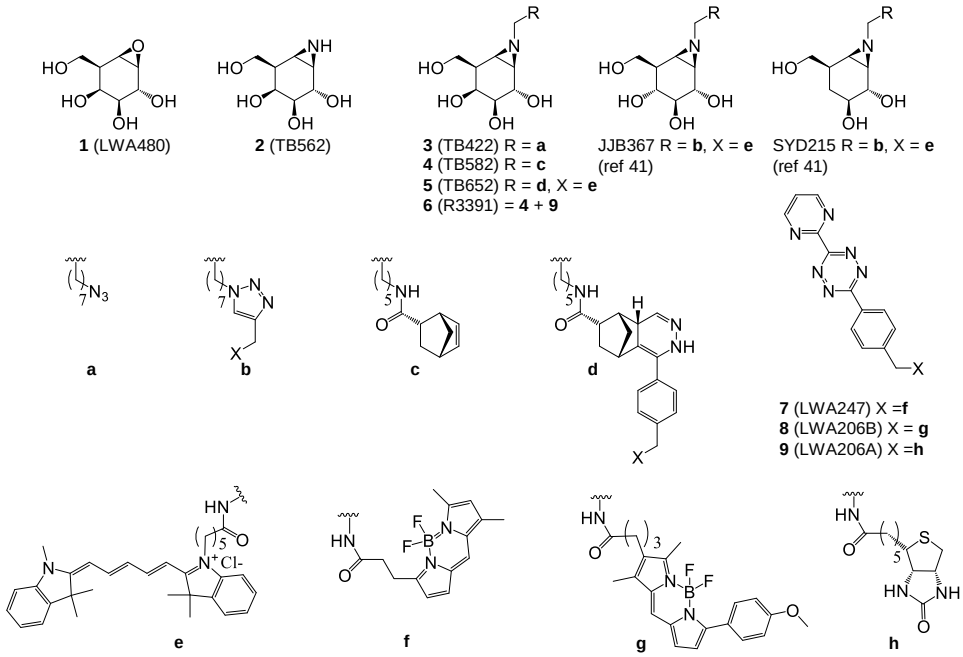


Figure 7.2. Structures of compounds used in this chapter.

aziridine compounds. These ABPs were ultimately generated through an alternative click chemistry involving the reaction between the norbornene-bearing cyclophellitol aziridine and the tetrazine-bearing reporter groups. This chapter reports on a detailed investigation of these compounds regarding inhibitory potency, glycosidase labeling, and target specificity.

7.2 Results

7.2.1 Synthetic strategies for β-galactose configured cyclophellitol aziridines

Compounds were synthesized at the Department of Bio-organic Synthesis, Leiden University. The synthesis of β-galactose configured cyclophellitol **1** and cyclophellitol aziridine **2** has been previously described.[40, 41] Acylation of **2** was unsuccessful due to the rapid decomposition of the products, but alkylation with 8-iodooctylazide was successful, leading to the N-alkylated compound **3**. However, normal Cu(I) catalyzed click reaction with an alkyne-fluorophore did not yield the expected products.[41] Therefore, an alternative strategy was tested, based on the inverse electron demand Diels-Alder (IEEDA) reaction between a norbornene and

a tetrazine.[42] Following this alternative approach, the cyclophellitol aziridine N-alkyl norbornene **4** was firstly synthesized by alkylation of **2** with a norbornene handle, and subsequent click reaction with tetrazine conjugates afforded the desired Cy5 ABP **5**[41] and biotin ABP **6**. In addition, due to the initial unavailability of **5** and **6**, the BODIPY-FL, BODIPY-TMR, and the biotin probe were generated *in situ* by incubation of **4** with compound **7**, **8**, or **9** in DMSO at RT for some period (30 min to overnight), and the mixture was used for several labeling experiments.

7.2.2 *In vitro* inhibition and labeling of compounds **2**-**7** towards β-galactosidases

Compounds were tested for inhibitory potency towards GLB1 (present in human fibroblasts lysates) and mouse GALC (collected culture medium of HEK293T cells overexpressing this protein).[38] The presence of GLB1 and GALC in these materials was first tested with 4-methylumbelliferyl β-D-galactopyranoside (4-MU-β-gal) substrate assay at various pH, and by using 11 μM $AgNO_3$ to selectively inhibit GLB1 over GALC.[43] The β-galactosidase activity in both materials had a pH optimum of 4.5, but at pH 6.0 the culture medium retained 60 % activity, in contrast to the 20 % retained by β-galactosidase in the fibroblast lysates (**Fig. 7.S1A**). In the presence of $AgNO_3$, no activity remained in the fibroblast lysates, while the culture medium retained full β-galactosidase activity (**Fig. 7.S1B**), indicating that the fibroblast lysates contained only GLB1 and that the culture medium contained only GALC. Compounds were next assessed for their inhibitory potency on both enzymes using 4-MU β-gal assay. The GLB1 and GALC enzyme preparations were incubated with substrate and compounds. Compounds **2** and **3** exhibited low nanomolar IC_{50} values towards both GLB1 and GALC; the epoxide **1**[44] and bulkier compound **4**, **5**, and **6** were about ten-fold less potent, but were still nanomolar inhibitors (**Table 7.1, Fig. 7.S2**). For comparison, the 4-deoxy Cy5 ABP SYD215 was a high nanomolar inhibitor for both enzymes, and was eight- to ten-fold less potent than the Cy5 ABP **5**; the β-glucose configured Cy5 ABP JJB367 was a further 50- to 100-fold less reactive, showing apparent IC_{50} values in the micromolar range (**Table 7.1, Fig. 7.S2**).

The compounds also inhibited the bacterial GH35 β-galactosidase from *Cellvibrio japonicus* (CjGH35A, an homologue to human GLB1[45]) with high potency (**Table 7.S1, Fig. 7.S2C**)), and kinetic studies of the compounds towards this enzyme revealed that the N-alkyl azide compound **2** has the highest affinity (lowest K_I) towards the enzyme and the highest inactivation rate constant (k_{inact} / K_I) of 21.71 min^{-1} μM^{-1} (**Table 7.2**, **Fig. 7.S3**, **S4**), with the later value comparable to that of cyclophellitol ABPs towards glucocerebrosidase (GBA).[46] It was also

Table 7.1. Apparent IC_{50} values (nM) of compounds towards retaining exo β-galactosidases. Error range = ± SD, *n* = 2 biological replicates.

	GLB1	GALC
JJB367	7,270 ± 658	53,500 ± 2,770
SYD215	138 ± 24.5	472 ± 37.6
1 (LWA480)	21.7 (ref 44)	39.1 (ref 44)
2 (TB562)	2.55 ± 0.59	5.57 ± 0.36
3 (TB422)	2.55 ± 0.47	12.0 ± 2.27
4 (TB582)	57.8 ± 3.05	98.6 ± 20.8
5 (TB652)	14.6 ± 0.98	61.0 ± 6.89
6 (R3391)	4.79 ± 4.09	9.95 ± 5.73

Table 7.2. Kinetic data for compounds 2-5 towards CjGH35A. Error range = ± SD, *n* = 3 biological replicates.

	K_{inact} (min^{-1})	K_I (μM)	k_{inact}/K_I (min^{-1} μM^{-1})
2 (TB562)	0.99 ± 0.11	1.90 ± 0.39	0.52 ± 0.06
3 (TB422)	0.56 ± 0.11	0.03 ± 0.01	21.71 ± 4.08
4 (TB582)	2.19 ± 0.55	3.37 ± 1.20	0.65 ± 0.16
5 (TB652)	2.03 ± 0.63	0.11 ± 0.04	18.98 ± 5.88

observed that while the maximum potential reaction rate (k_{inact}) of compound **2**-**5** was similar, their affinity (K_I) towards CjGH35A was quite different—with the bare aziridine **2** and the N-alkyl norbornene **4** less favorable than the N-alkyl azide **3** and the Cy5 ABP **5** (**Table 7.2**).

A crystal structure of CjGH35A soaked with compound **2** was resolved and revealed that the compound bound to the enzyme at its catalytic nucleophile and adopted a $^4C_1$ conformation (**Fig. 7.3**).

Next, ABP labeling was attempted in human fibroblast lysates and the culture medium of GALC-overexpressing HEK293T cells. Due to the initial unavailability of the Cy5 ABP **5**, the experiment was performed by firstly pre-incubating the norbornene compound **4** with the tetrazine-BODIPY compounds **8**, and incubating the mixture with the culture medium of

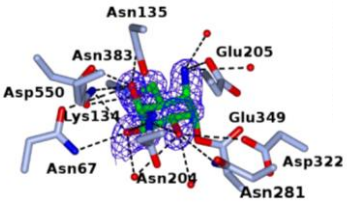


Figure 7.3. Crystal structure of CjGH35A in complex with 2. The map shown is a Fo-Fc map, with phases calculated prior to the inclusion of ligand in the refinement, contoured at 3σ. Carbon atoms are colored green for the ligand and ice blue for the side chains. The interacting residues are annotated, including catalytic residues Glu349 (nucleophile) and Glu205 (acid/base).

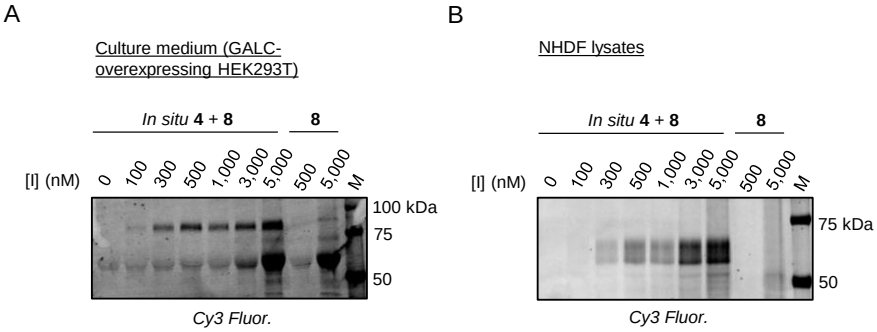


Figure 7.4. *In vitro* ABP labeling using *in situ* generated probe (4 + 8) at different probe concentration. A) Labeling in culture medium of HEK293T cells overexpressing mouse GALC. B) Labeling in human fibroblast (NHDF) lysates.

GALC-overexpressing cells. A band of about 80 kDa in size was observed, (**Fig. 7.4A**), which could correspond to the secreted form of mouse GALC. In human fibroblast lysates, a smear of bands between 50 and 75 kDa was detected (**Fig. 7.4B**), which indicated the labeling of glycoprotein(s).

Because this labeling pattern resembled that of cyclophellitol ABPs towards GBA (Chapter 1 and 2, this thesis), the reactivity of compound **2**-**6** was also examined towards the recombinant human GBA (Imiglucerase) by enzymatic assay. The results showed that all compounds inhibited GBA with nanomolar potency, except for the aziridine compound **2** and the norbornene compound **4**, both being low micromolar inhibitors of GBA **(Table. 7.S1**, **Fig. 7.S5)**. The compounds were also tested for their reactivity towards another retaining β-glucosidase, GBA2. It turned out that they were about two- to ten-fold less potent towards GBA2 compared to GBA **(Table. 7.S1**, **Fig. 7.S5)**. The labeling by the *in situ* generated

BODIPY-TMR probe (**4** + **8**) towards GBA was confirmed in NHDF lysates using a competitive ABPP (cABPP) setup, where its labeling was partially abrogated by pre-incubating the lysates with the GBA-specific ABP ME569 (chapter 3, this thesis) (**Fig. 7.S6**). To further verify if GLB1 and GALC were labeled by the probe, an additional cABPP was performed in mouse tissue extracts (expected to express both GLB1 and GALC) pre-incubated with cyclophellitol (known β-glucosidase inhibitor), compound **3** (inhibitor for both β-glucosidase and β-galactosidase) or **4** (β-galactosidase inhibitor at lower concentrations), subsequently labeled with the GBA-specific ABP ME569, and finally with the *in situ* generated probe. The probe labeled a ~50 kDa band and a ~60 kDa band that overlapped with GBA labeling by ME569 (lane 2 of both gels, **Fig. 7.5A**). Pre-incubation with **3** or **4** (the later at 500 nM), but not cyclophellitol, abolished the labeling by the *in situ* generated BODIPY-FL probe (**Fig. 7.5A**). Contrastingly, cyclophellitol and **3**, but not **4**, abolished the labeling by ME569 (**Fig. 7.5A**). In

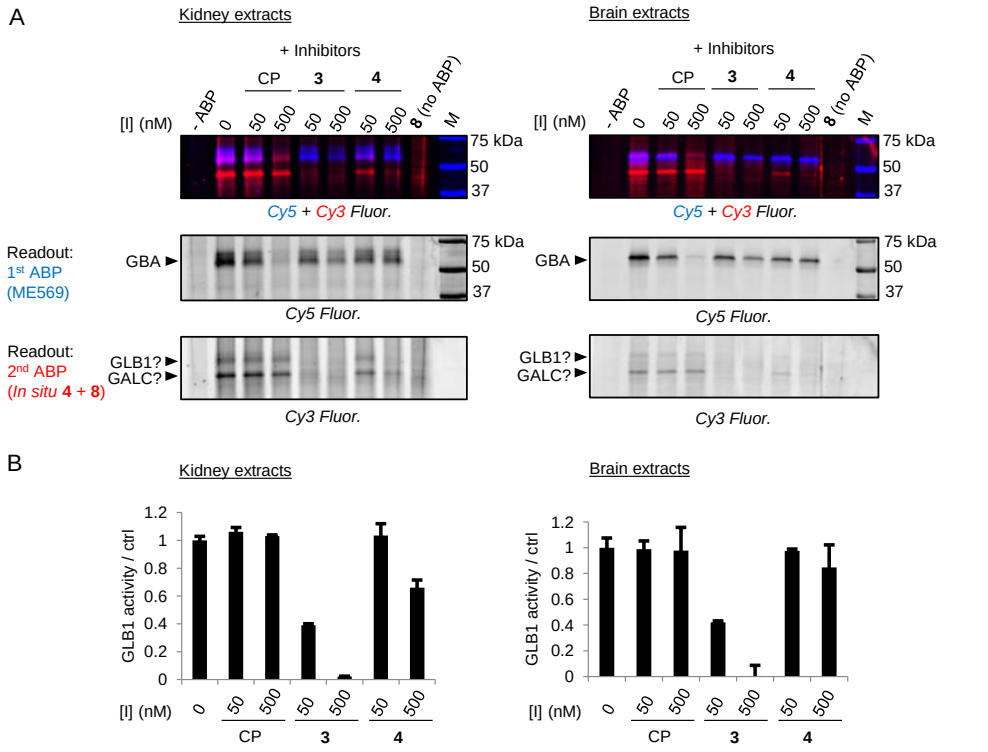


Figure 7.5. Labeling on GLB1 and GBA by the *in situ* generated probe (4 + 8) in extracts of mouse kidney (left) or brain (right). A) Competitive ABPP pre-incubated firstly with cyclophellitol (CP), **3**, or **4**, and secondly with ABP ME569. B) GLB1 activity in the identically treated samples measured by fluorogenic substrate assay. Error range = ± SD, *n* = 3 technical replicates.

parallel, enzymatic assay performed in the identically-treated samples showed that GLB1 activity was abolished by **3** and (at 500 nM) **4** (**Fig. 7.5B**), but not by cyclophellitol. Together, these results strongly suggested that the β-galactose-configured N-tagged cyclophellitol aziridine ABP allows gel-based fluorescence detection for both GLB1 (known size = 64 kDa) and GALC (known size = 50 kDa).

Next, the β-galactose-configured Cy5 ABP **5** was employed for evaluating with mouse kidney extracts (known to contain both GALC and GLB1) its labeling properties. The labeling was performed in samples pre-incubated with the broad-spectrum β-glucosidase BODIPY ABP JJB70, in order to specifically visualize the β-galactosidases (over GBA and GBA2). Labeling occurred in an irreversible and mechanism-based manner: optimal labeling for both GLB1 and GALC occurred at 1 μM ABP 5 (**Fig. 7.6A**), 30 min incubation time (**Fig. 7.6 B**), and at pH from 4.0 to 5.0 (**Fig. 7.6 C**). The labeling was partially abrogated by pre-incubating the samples with 4-MU β-gal, consistent with labeling occurring in the active site of the enzymes (**Fig. 7.6D**).

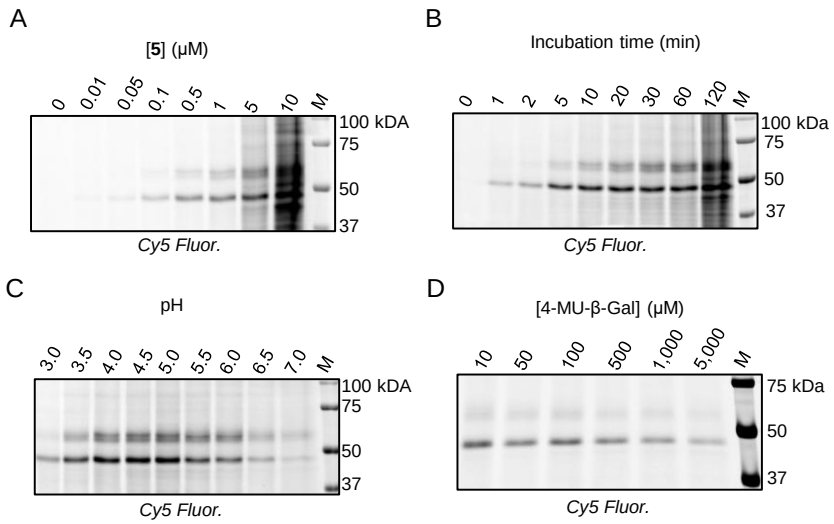


Figure 7.6. Labeling of ABP 5 in mouse kidney extracts. (A) Concentration-dependent labeling. (B) Time-dependent labeling. (C) pH-dependent labeling. (D) Competitive ABPP with 4-MU-β-gal.

7.2.3 Glycosidase target analysis by chemical proteomics and gel-based ABPP in mouse tissue extracts

Glycosidase targets of the biotin ABP **6** in mouse kidney extracts were next examined. For

this purpose, biotin ABP **6** was generated from a mixture of 5 μM **4** and 10 μM **9** reacted O/N. The ABP was with mouse kidney extracts at pH 4.5, and samples were subjected to LC/MS-based protein identification form the streptavidin-enriched and on-bead trypsin-digested peptides. Five glycosidases were identified in samples incubated with the *in situ* generated biotin probe (**4** and **9**), and none in the control (no probe) or the competitive sample (pre-incubated with the Cy5 ABP **5**). The identified glycosidases included the expected targets GALC, GLB1, the β-glucosidase GBA, as well as two other targets: the GLB1-like protein 1 and 2 (GLB1L and GLB1L2, respectively) (**Table 7.3**). The latter two proteins are putative enzymes that are also classified onto GH35 family but with yet unknown biological functions. Both proteins contain catalytic amino acids identical to those of GLB1 (**Fig. 7.S7**). Despite having similar molecular weight to the GLB1 precursor protein (around 70 kDa without N-glycosylation), their post-translational processing, cellular localizations, and tissue-dependent expression are predicted to be different to those of GLB1 (**Table 7.S3)**. To examine whether these features would allow their visualization by the β-galactose configured Cy5 ABP **5**, mouse tissue extracts from brain, epididymis, testis, duodenum, and colon were pre-incubated with the broad-spectrum β-glucosidase ABP JJB70 (to prevent labeling on β-glucosidases by ABP **5**), and next incubated with ABP **5** for gel-based fluorescence detection. The gel showed that each tissue had a distinct labeling profile, and that additional bands were identified at molecular weight or pH range different from those of GLB1 and GALC. In extracts of mouse brain, epididymis,

Table 7.3. List of identified glycosidases in mouse kidney extracts labeled and pulled-down with ABP 6 by LC-MS.

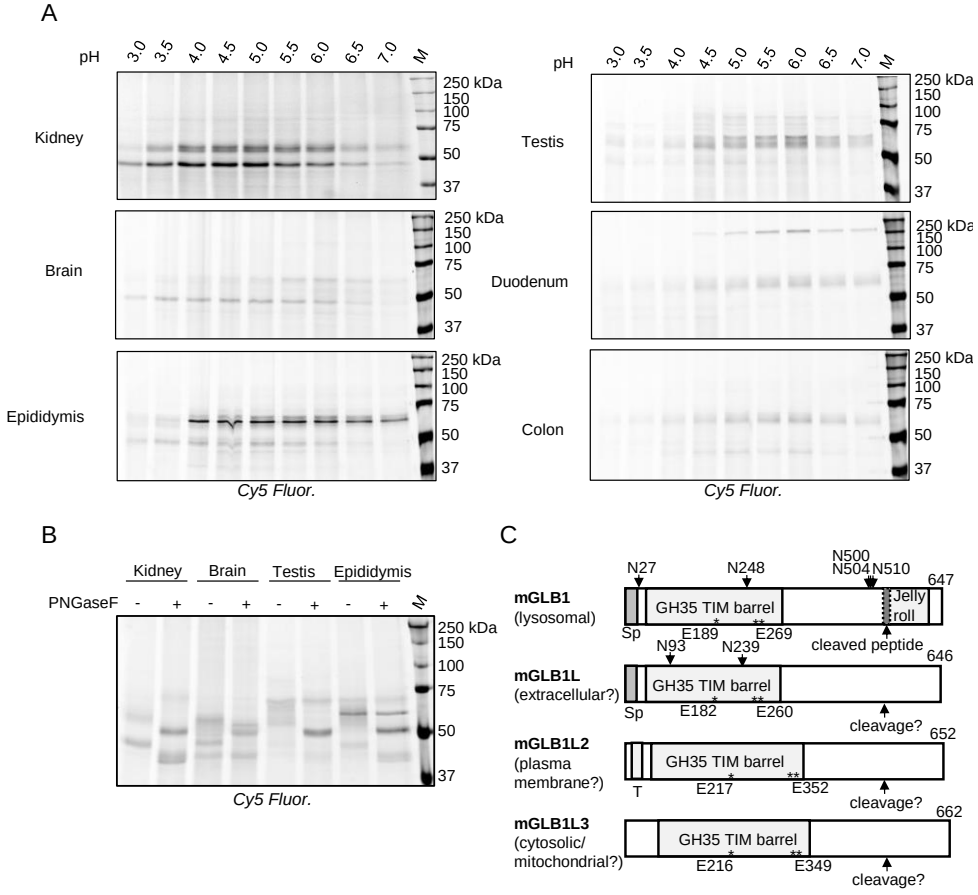
Rank	Accession	Entry	Description	PLGS Score	Coverage (%)
11	P54818	GALC_MOUSE	Galactocerebrosidase OS=Mus musculus GN=GALC PE=1 SV=2	4177	18
13	P23780	BGAL_MOUSE	Beta-galactosidase OS=Mus musculus GN=Glb1 PE=1 SV=1	2851	31
22	Q8VC60	GLB1L_MOUSE	Beta-galactosidase-1-lile protein OS=Mus musculus GN=Glb1l PE =1 SV=1	1783	22
47	P17439	GLCM_MOUSE	Glucosylceramidase OS=Mus musculus GN=Gba PE=1 SV=1	634	27
62	Q3UPY5	GLBL2_MOUSE	Bet-galactosidase-1-lile protein 2 OS=Mus musculus GN=Glb1l2 PE=1 SV=1	367	23

testis, duodenum, and colon, a number of bands at around 60 to 70 kDa displayed pH optimum at 6.0, instead of 4.5 from the one in kidney extracts (**Fig. 7.7A**). In duodenum extracts, an additional band between 150 and 250 kDa was detected optimally at pH 6.0, but it was not presented in extracts of colon. It was reasoned that the later higher band could be the intestinal dietary enzyme lactase phlorizin-hydrolase (LPH), as it contains a lactase pocket that hydrolyzes a β-galactose from the disaccharide lactose.[47] Labeling with ABP **5** in cells expressing either the wild-type human LPH, or LPH mutated at the catalytic nucleophile of phlorizin hydrolase (pocket III) or lactase (IV) revealed that indeed LPH was labeled, and that the labeling largely occurred at the lactase pocket (**Fig. 7.S8**).

The bands detected in mouse kidney, brain, epididymis, and testis were further examined by treating the samples with or without PNGase F, an enzyme that removes the protein N-linked glycans. The treatment in kidney extracts shifted the 64 kDa band (putatively GLB1) to just above 50 kDa, and the 50 kDa band (GALC) to duplet bands at around 40 kDa (**Fig. 7.7B**). In brain extracts, deglycosylation of the 64 kDa band resulted in two bands just above kDa (**Fig. 7.7B**). In testis and epididymis extracts, the ~70 kDa band was not affected by deglycosylation, as well as the prominent ~64 kDa band in the epididymis extracts (**Fig. 7.7B**). Together with the pH experiment, these results suggested that the Cy5 ABP **5** also likely labeled the GLB1-like proteins, as some of the bands between 50 kDa and 75 kDa either has a pH optimum of 6.0 (suggesting non-lysosomal localization), or were not N-glycosylated (GLB1L2 and possibly GLB1L3, also points to non-lysosomal localization). In agreement with these observation, bioinformatics analysis (**Fig. 7.7C and Table 7.S3**) predicts that in contrast to GLB1L (which is likely secreted to the extracellular milieu), the other two GLB1-like proteins are not N-glycosylated. GLB1L2 is predicted to contain no signal peptide, but instead a short transmembrane domain close to its N-terminus. Accordingly, it could be an one-pass plasma membrane protein having its catalytic pocket (C-terminal) facing the extracellular side. GLB1L3 is also predicted to have no signal peptide, and could be a cytosolic/peroxisomal protein; it does have possible N-glycan sites, but if it does not contain an ER-targeting signal, N-glycosylation is not likely to occur. Furthermore, since the GLB1-like proteins are possibly not targeted to the lysosomes, they are unlikely to undergo a similar lysosomal proteolytic cleavage as that of GLB1. This would make their molecular weight larger than the mature, deglycosylated GLB1. The latter corresponds to the band observed just above 50 kDa seen in the PNGase F + lanes in **Fig. 7.7B**. In all the tested tissue homogenates at least one band with higher molecular weight was observed

that was insensitive to PNGase F treatment. These bands are possibly GLB1-like proteins.

Figure 7.7. *In vitro* labeling with ABP 5 in mouse tissue extracts. A) pH-dependent labeling in mouse



tissue extracts. B) Labeling in mouse tissue extracts (pH 6.0) with or without deglycosylation by PNGaseF. C) Comparing domain structure, catalytic sites, and N-glycosylation sites for mouse GLB1 and GLB1-like proteins. *Sp*, signal peptide; *T*, transmembrane domain; *, acid/base; **, nucleophile; downward arrows, N-glycan sites; GH35 TIM barrel, Glycohydrolase family 35 domain (pfam01301); *Jelly roll*, β-galactosidase 4_5 domain (pfam13364).

7.3 Discussion

In this chapter, the newly synthesized β-galactose configured cyclophellitol aziridine ABPs were examined for their mechanism-based labeling of GLB1 and GALC—two retaining exo-β-galactosidases whose lysosomal deficiency underlies the lysosomal storage diseases such as GM1-gangliosidosis, Morquio B syndrome, galactosialidosis, and Krabbe disease.

It is shown that the ABPs successfully label both GLB1 and GALC in mouse kidney extracts in a mechanism-based manner. They exhibit nanomolar inhibitory potency against both enzymes, as well as the bacterial GH35 enzyme CjGH35A. Towards the later enzyme, they also exhibit fast inhibition kinetics, resembling the values of previously reported cyclophellitol-based inhibitors and ABPs towards GBA.[46] When incubated with crystals of CjGH35A, Compound **2** forms a glycosidic bond with the enzyme, as seem in the resolved structure. This further verified its mechanism-based labeling towards β-galactosidases. While all the novel compounds (including the ABPs) also concomitantly inhibited GBA, compound **4** (β-galactose configured cyclophellitol aziridine N-alkyl norbornene) is relatively inactive towards this enzyme (IC_{50} ratio of GBA/GLB1 = 200, highest among the tested compounds). This allows it to be used as a tool to selectively block β-galactosidase activity over β-glucosidases (at a carefully chosen concentration, such as 0.5 μM), a feature exploited in this chapter to verify targets of the newly-generated ABPs. With its concomitant inhibition of β-glucosidases in mind, GLB1 and GALC can be visualized by the Cy5 ABP after pre-incubation of GBA inhibitor, as demonstrate in this chapter. Thus, the Cy5 ABP can be used to rapidly and simultaneously profile GLB1 and GALC activity, as well as their post-translational modification and proteolytic processing. This would make it a very useful tool in the study and diagnosis of lysosomal storage disorders caused by GLB1 or GALC deficiency.

Chemical proteomics with the biotin ABP **6** in mouse kidney extracts revealed that, besides the expected target GLB1, GALC, and GBA, two other putative GH35 β-galactosidases were identified. These are the GLB1-like protein 1 and 2 (GLB1L and GLB1L2), which share homology to GLB1 and contain highly-conserved amino acid sequences near and at the catalytic residues (**Fig. 7.S7**). Gel-based ABPP in a variety of mouse tissue extracts also identifies several proteins labeled by the Cy5 ABP **5** that exhibits properties distinct from the lysosomal protein GLB1 and GALC, such as a more neutral pH optimum (around 6.0), no N-linked glycosylation, and different tissue-dependent expression profile. These observations are corroborated with bioinformatics analysis, which shows that both the mouse and human GLB1L and GLB1L2 are likely not lysosomal proteins, and that GLB1L2 lacks a signal peptide nor possesses any N-glycosylation sites (**Table 7.S3**). It is noteworthy that there exists another GH35 protein, the GLB1-like protein 3 (GLB1L3) in both mouse and human, predicted to be located in the cytosol, mitochondria, or peroxisomes (**Table 7.S3**). This protein was not identified by chemical proteomics in mouse kidney extracts, where it is probably poorly expressed (according to

BioGPS expression database[48], MOE430 dataset[49]). Nevertheless, due to the presence of conserved amino acid sequence at regions homologous to the GLB1's catalytic site (**Table 7.S3**), it is envisioned that future pull-down experiment in tissues highly expressing this protein (such as testis[49]) might also reveal that GLB1L3 also is a *bona fide* β-galactosidase.

Previously, a so-called senescence-associated β-galactosidase (SA-β-gal) is reported in a wide range of senescent cells, from which the β-galactosidase activity can be selectively detected over the non-senescent cells at pH 6.0.[50] The detected pH optimum at 6.0 for GLB1L and GLB1L2 (and possibly also GLB1L3) makes one to speculate the possibility that they might be the SA-β-gal. However, it has been shown with compelling evidence that the identity of SA-β-gal is in fact the lysosomal GLB1, and that upon its upregulation (both mRNA and protein) during senescence its suboptimal activity at pH 6.0 can be readily detected.[51] In fact, such senescence-dependent upregulation is not unique to GLB1, as other lysosomal glycosidases, such as α-L-fucosidase and β-glucuronidase, are also similarly upregulated.[52] Recently, a cell-permeable tellurophene-containing ABP labeling proximal nucleophile has been developed for measuring SA-β-gal activity using an MS-based detection method.[53] It can be envisioned that the here-described cyclophellitol aziridine fluorescent ABPs could also be used to simultaneous profile activity from multiple β-galactosidase across a pH range during senescence, using the simpler gel-based method.

Because the molecular processing, enzymatic activity, and biology of GLB1-like proteins are still largely unknown, the ability of ABP **5** and **6** to profile their activity could offer exciting research directions. These include understanding their cell biology and biochemistry, and potential involvement senescence and in the lysosomal storage diseases caused by GLB1 or GALC deficiency. Additionally, the Cy5 ABP labels the dietary enzyme lactose-phlorizin hydrolase (LPH) in mouse duodenum extracts and in lysates of cells overexpressing this protein. This probe might be also useful to monitor the activity of this enzyme in other research context, such as its age-dependent expression[54] and in diseases such as lactose-intolerance.[55]

To conclude, the novel β-galactose configured cyclophellitol aziridine compounds are true mechanism-based irreversible inhibitors for both GLB1 and GALC. They concomitantly inactivate GBA and GBA2 at higher concentrations, but by pre-incubating the samples with β-glucosidase inhibitors their visualization with the β-galactose ABPs can be prevented. The fluorescent and biotinylated ABPs are useful to investigate activities of GLB1 and GALC across

cell lysates and tissue homogenates. They also labeled the dietary enzyme LPH, and additionally revealed that the GLB1-like proteins are functional β-galactosidases that have different post-translational processing, pH optimum, and tissue-dependent expression to that of GLB1. They should be valuable tools in the future study of β-galactosidases in health and disease.

7.4 Experimental procedures

7.4.1 Chemicals for biological assays

Cyclophellitol (CP) and the β-glucosidase ABPs were synthesized as described earlier.[39, 46] Chemicals were obtained from Sigma-Aldrich, if not otherwise indicated. Protein concentration was measured using Pierce™ BCA assay kit (Thermo Fisher).

7.4.2 Cell culture

HEK293T cells (ATCC, CRL-3216) were cultured in DMEM high glucose (Sigma-Aldrich) supplemented with 10 % FCS, 0.1 (w/v) penicillin/streptomycin, and 1 % (v/v) Glutamax at 37 °C at 7 % CO_2, and sub-cultured at 1:10 ratio twice a week. Normal human dermal fibroblasts (NHDF, Cambrex-Lona, CC-2511) were cultured in HAMF12-DMEM medium (Sigma-Aldrich) supplemented with 10 % FCS and 0.1 (w/v) penicillin/streptomycin at 5 % CO_2, and sub-cultured at 1:3 ratio once per 1-2 weeks. Culture medium was refreshed every 2-3 days for NHDF.

7.4.3 Enzyme activity assays

Recombinant murine GALC was cloned and expressed in HEK293 cells following previously described procedures.[38] The produced protein which was secreted to the culture medium (DMEM high glucose, Gibco) was directly used (5 µL volume) in enzyme activity studies. Human fibroblast lysates were prepared in KPi buffer (25 mM K_2HPO_4/KH_2PO_4, pH 6.5, 0.1 % (v/v) Triton X-100, protease inhibitor cocktail (Roche, version 12, EDTA-free)), and 5.5 µg total protein was used. β-galactosidase from *Cellvibrio japonicus* (CjGH35A)[45] was prepared in KPi buffer, and 8.9 ng protein was used. Enzyme activity assays were performed in 96-well plates (Greiner, black, flat-bottomed, medium-binding). All samples were diluted in McIlvaine buffer (150 mM citrate/Na_2HPO_4, pH 4.5) in a total volume of 25 µL, before incubated with 100 µL substrate mixture (1 mM 4-methylumbelliferyl β-D-galactopyranoside (4-MU-β-gal) in 150 mM McIlvaine buffer (pH 4.5) with 0.2 M NaCl) for 30 min at 37 °C. Reactions were stopped and analyzed as described earlier.[56] To discriminate GLB1 and GALC activity, samples were incubated with McIlvaine buffer with or without 11 µM $AgNO_3$[43], before incubated with substrates. To determine the apparent IC_{50} values, enzymes were equilibrated in 12.5 µL McIlvaine buffer (pH 4.5) and incubated with a range of inhibitor dilutions (12.5 µL) for 30 min at 37 °C, before subsequent incubation with substrates. All assays were performed in triplicate wells, and IC_{50} assays were additionally performed in experimental duplicates per compounds.

IC_{50} values were calculated as described earlier.[56] KPi buffer and culture medium form control HEK293T cells were used as negative control for GLB1 and GALC, respectively.

7.4.4 Inhibition kinetics

An enzyme stock was prepared by diluting CjGH35A to 0.71 ng/μL in McIlvaine buffer (pH 4.5, 0.1 % (w/v) BSA), and pre-warmed at 37℃ for 10 min in a thermoshaker. A series of 2 mL Eppendorf tubes containing compounds **2**-**5** (162.5 μL, various compound concentrations) was added with 4-MU-β-gal (1300 μL, reaction [4-MU-β-gal] = 2345 μM), and identically pre-warmed. The t=0 samples were prepared by taking 112.5 μL from the [compound + 4-MU-β-gal] to the first two columns of a 96-well plate (duplo for each compound concentration), after which stop buffer (200 μL, glycine-NaOH 1M, pH 10.3) and lastly the enzyme (12.5 μL) was added. Then, enzyme (137.5 μL) was added (t=0) to each of the 2 mL Eppendorf tubes (containing mixtures of compounds and substrates). At t=2, 4, 6, 8 and 10 min, an aliquot of 125 μL from each of the reaction mixture was taken in duplo and added to the plate containing stop buffer. After 10 minutes, the plate was measured with a LS-55 Fluorescence Spectrometer (PerkinElmer) using BL Studio with excitation at 366 nm and emission at 445 nm. Kinetic experiments were performed in duplo sets, each set with technical duplicates. Obtained values were plotted against incubation time, and fitted with one phase exponential association function to derive k_{obs} values. The obtained k_{obs} values were plotted against compound concentration in a second graph, and curve-fitted with Michaelis-Menten function to derive K_I' and k_{inact}. K_I was derived from K_I' using the function $K_I (1 + [4\text{-MU-}\beta\text{-gal}]/K_M) = K_I'$, where [4-MU-β-gal] = 2345 μM and K_M = 487.4 μM. The K_M was determined in a separate experiment by incubating the enzyme (same concentration as the kinetic experiments) with a series of substrate concentrations for 10 minutes at 37 °C. The reaction was stopped by adding stop buffer and samples were measured for 4-MU fluorescence. The results were plotted against substrate concentration, and K_M was determined with Michaelis-Menten equation. Results were processed and analyzed using GraphPad Prism 7.0.

7.4.5 Labeling using *in situ* generated probes

The *in situ* generated probe was prepared by incubating 1 mM **4** (in DMSO) with 1 mM **7** or **8** (both in DMSO) for at least 15 min at RT, unless otherwise stated. The produced probe had a concentration of 0.5 mM, assuming the click reaction proceeded to completion. Human

fibroblast lysates (5 μg protein) and culture medium of HEK293T cells (5 μL volume) were diluted in 150 mM McIlvaine buffer (pH 4.5) in a total volume of 10 μL, followed by incubation with 5 μL of the *in situ* generated probe mixture (diluted in McIlvaine buffer) for 30 min at 37 °C, at a final probe concentration of 1 μM. For competitive activity-based protein profiling (cABPP), human fibroblasts lysates were diluted as described above, and pre-incubated with 2.5 μL ME569[57] or the *in situ* generated probe mixture for 30 min at 37 °C at a final probe concentration of 100 nM for ME569 and 1 μM for the *in situ* generated probe. Samples were subsequently incubated with the second probe (ME569 for samples pre-incubated with the in situ generated probe, and vice versa) for another 30 min at 37 °C with identical probe concentration during pre-incubation. Samples were denatured, resolved by SDS-PAGE, and scanned for fluorescence as described previously.[56]

7.4.6 Competitive ABPP in extracts of mouse kidney or brain

25 μg protein from extracts of mouse kidney or brain (prepared as described in chapter 8) was diluted in 10 μL McIlvaine buffer (150 mM, pH 4.5), and pre-incubated with 2 μL cyclophellitol, **3**, or **4** (diluted in DMSO and McIlvaine buffer) for 30 min at 37 °C at 50 or 500 nM concentration for compound during incubation. Samples were subsequently incubated with 2 μL ME569 (100 nM during incubation) for 30 min at 37 °C, then with 2 μL *in situ* generated probe (**4** + **8,** 1 μM of the clicked product during incubation, assuming complete reaction) or 8 (1 μM during incubation, as negative control) for 30 min at 37 °C. Samples were denatured, resolved by SDS-PAGE, and scanned for fluorescence as described previously.[56]

7.4.7 Characterization of labeling by ABP **5** in mouse kidney extracts

25 μg total protein from mouse kidney extracts was diluted in McIlvaine buffer (150 mM, various pH) in a total volume of 10 μL, and incubated with 5 μL ABP **5** prepared at different concentrations and pHs. For labeling at varying pH, kidney extracts were incubated with McIlvaine buffer (pH 3.0 – 7.0) for 15 min at 37 °C, followed by incubating with MDW933[46] (prepared in McIlvaine buffer pH 3.0 - 7.0, 5 μM during incubation) for 30 min at 37 °C to label all GBA, then with 1 μM ABP **5** (prepared in McIlvaine buffer pH 3.0 -7.0, 1 μM during incubation) for 30 min at 37 °C. For labeling at varying ABP concentration, mouse kidney extracts were equilibrated in McIlvaine buffer pH 4.5 for 5 min on ice, incubated with 5 μM MDW933 (end concentration; pH 4.5) for 30 min at 37 °C, then with varying concentration of

ABP **5** (end concentration = 10nM -10μM; pH 4.5) for 30 min at 37 °C. For labeling at varying incubation time, mouse kidney extracts were prepared as above, and incubated with 1 μM ABP **5** (end concentration; pH 4.5) at 37 °C for 1 min to 2 h without MDW933 pre-incubation.

7.4.8 Chemical proteomics

3.5 mg protein from mouse kidney extracts was diluted with McIlvaine buffer (750 mM, pH 4.5) in a total volume of 500 μL. Three samples were prepared, the first incubated with DMSO only, the second firstly with 5 μM ABP **5** and secondly with 10 μM *in situ* generated ABP **6** (1 mM **4** + 2 mM **9**, O/N incubation at RT in DMSO), and the third with firstly DMSO and secondly the 10 μM *in situ* generated ABP **6**. All incubation step was performed at 37 °C for 1 h. Samples were denatured and subjected to pull-down, on-bead tryptic digest, and LC-MS protein identification as described in Chapter 6.

7.4.9 ABPP in mouse tissue extracts with ABP **5**

20 μg total protein from extracts of mouse kidney, brain, epididymis, testis, duodenum, and intestine were diluted in 10 μL McIlvaine buffer (150 mM, various pH from 3.0 to 7.0). Samples were then pre-incubated with 2.5 μL ABP JJB70[58] (diluted in McIlvaine buffer pH 3.0 – 7.0, 200 nM during incubation) for 30 min at 37 °C, and then 2.5 μL ABP **5** (diluted in McIlvaine buffer pH 3.0 – 7.0, 1 μM during incubation) for 30 min at 37 °C, before subjected to SDS-PAGE-based fluorescent detection. For deglycosylation analysis, ABP labeled samples (60 μg protein diluted consecutively in 20 μL McIlvaine buffer pH 6.0 (5 min on ice), 5 μL ABP JJB70 (30 min, 37 °C) and 5 μL ABP **5** (30 min, 37 °C), at identical ABP concentrations as described) were firstly desalted using Pierce 7K polyacrylamide spin column (Thermo Fisher), and a 10 μL aliquot was treated with PNGase F according to the manufacturer's protocol (New England BioLabs). Non-treated samples (20 μg protein diluted in 10 μL McIlvaine buffer pH 6.0) were similarly labeled with ABPs. Both the non-treated and PNGase F-treated samples were subjected to SDS-PAGE and fluorescence detection. For labeling on recombinant LPH, wild-type of catalytic mutant LPH constructs[59] were expressed in HEK293T cells using the PEI method (Chapter 6). Cells were lysed in lysis buffer, and 20 μg total protein from the resulting lysates were labeled with 10 μM ABP **5** at pH 6.0 for 30 min at 37°C, before subjecting to SDS-PAGE and fluorescence detection.

7.4.10 Bioinformatics

Amino acid sequences (human GLB1: NP_000395.3, GLB1L: NP_001273352.1, GLB1L2: NP_612351.2, GLB1L3: NP_001073876.2; mouse (*Mus musculus*) GLB1: NP_033882.1, GLB1L: NP_083286.1, GLB1L2: NP_722498.1, GLB1L3: NP_001106794.1) were retrieved from NCBI. Multiple sequence alignment was performed with CLUSTALW.[60] Signal peptide prediction was based on SignalP 5.0[61] (DTU Bioinformatics). Transmembrane domain prediction was performed with XtalPred-RF.[62] N-glycosylation sites were predicted using NetNGlyc 1.0[63] (DTU Bioinformatics). Cellular localization was predicted using WoLF PSORT.[64]

7.5 References

1 Hennet T (2002) The galactosyltransferase family. *Cell Mol Life Sci* **59**, 1081–1095.

2 Suzuki Y & Nanba E (2010) β-Galactosidase deficiency (β-galactosidosis) GM1-gangliosidosis and Morquio B disease. In *The Metabolic and Molecular Basis of Inherited Disease*. New York, NY: McGraw-Hill.

3 Suzuki K & Suzuki Y (1970) Globoid cell leucodystrophy (Krabbe's disease): deficiency of galactocerebroside beta-galactosidase. *Proc Natl Acad Sci USA* **66**, 302–309.

4 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2014) The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490–D495.

5 D'Azzo A, Hoogeveen A, Reuser AJ, Robinson D & Galjaard H (1982) Molecular defect in combined beta-galactosidase and neuraminidase deficiency in man. *Proc Natl Acad Sci USA* **79**, 4535–4539.

6 van der Spoel A, Bonten E & d'Azzo A (2000) Processing of lysosomal beta-galactosidase. The C-terminal precursor fragment is an essential domain of the mature enzyme. *J Biol Chem* **275**, 10035–10040.

7 Kreutzer R, Kreutzer M, Pröpsting MJ, Sewell AC, Leeb T, Naim HY & Baumgärtner W (2008) Insights into post-translational processing of beta-galactosidase in an animal model resembling late infantile human G-gangliosidosis. *J Cell Mol Med* **12**, 1661–1671.

8 Ohto U, Usui K, Ochi T, Yuki K, Satow Y & Shimizu T (2012) Crystal structure of human β-galactosidase: structural basis of Gm1 gangliosidosis and morquio B diseases. *J Biol Chem* **287**, 1801–1812.

9 Wilkening G, Linke T, Uhlhorn-Dierks G & Sandhoff K (2000) Degradation of membrane-bound ganglioside GM1. Stimulation by bis(monoacylglycero)phosphate and the activator proteins SAP-B and GM2-AP. *J Biol Chem* **275**, 35814–35819.

10 Galjart NJ, Gillemans N, Harris A, van der Horst GT, Verheijen FW, Galjaard H & d'Azzo A (1988) Expression of cDNA encoding the human "protective protein" associated with lysosomal beta-galactosidase and neuraminidase: homology to yeast proteases. *Cell* **54**, 755–764.

11 Morreau H, Galjart NJ, Gillemans N, Willemsen R, van der Horst GT & d'Azzo A (1989) Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein. *J Biol Chem* **264**, 20655–20663.

12 Caciotti A, Donati MA, Boneh A, d'Azzo A, Federico A, Parini R, Antuzzi D, Bardelli T, Nosi D, Kimonis V, Zammarchi E & Morrone A (2005) Role of beta-galactosidase and elastin binding protein in lysosomal and nonlysosomal complexes of patients with GM1-gangliosidosis. *Hum Mutat* **25**, 285–292.

13 Nagano S, Yamada T, Shinnoh N, Furuya H, Taniwaki T & Kira J (1998) Expression and processing of recombinant human galactosylceramidase. *Clin Chim Acta* **276**, 53–61.

14 Wenger DA, Rafi MA & Luzi P (1997) Molecular genetics of Krabbe disease (globoid cell leukodystrophy): diagnostic and clinical implications. *Hum Mutat* **10**, 268–279.

15 Deane JE, Graham SC, Kim NN, Stein PE, McNair R, Cachón-González MB, Cox TM & Read RJ (2011) Insights into Krabbe disease from structures of galactocerebrosidase. *Proc Natl Acad Sci USA* **108**, 15169–15173.

16 Hill CH, Graham SC, Read RJ & Deane JE (2013) Structural snapshots illustrate the catalytic cycle of β-galactocerebrosidase, the defective enzyme in Krabbe disease. *Proc Natl Acad Sci USA* **110**, 20479–20484.

17 Matsuda J, Vanier MT, Saito Y, Tohyama J, Suzuki K & Suzuki K (2010) A mutation in the saposin

A domain of the sphingolipid activator protein (prosaposin) gene results in a late-onset, chronic form of globoid cell leukodystrophy in the mouse. *Hum Mol Genet* **10**, 1191–1199.

18 Hill CH, Cook GM, Spratley SJ, Fawke S, Graham SC & Deane JE (2018) The mechanism of glycosphingolipid degradation revealed by a GALC-SapA complex structure. *Nat Commun* **9**, 151.

19 Chen YQ, Rafi MA, de Gala G & Wenger DA (1993) Cloning and expression of cDNA encoding human galactocerebrosidase, the enzyme deficient in globoid cell leukodystrophy. *Hum Mol Genet* **2**, 1841–1845.

20 Brunetti-Pierri N & Scaglia F (2008) GM1 gangliosidosis: review of clinical, molecular, and therapeutic aspects. *Mol Genet Metab* **94**, 391–396.

21 Suzuki Y, Nakamura N & Fukuoka K (1978) GM1-gangliosidosis: accumulation of ganglioside GM1 in cultured skin fibroblasts and correlation with clinical types. *Hum Genet* **43**, 127–131.

22 Tessitore A, dPM M, Sano R, Ma Y, Mann L, Ingrassia A, Laywell ED, Steindler DA, Hendershot LM & d'Azzo A (2004) GM1-ganglioside-mediated activation of the unfolded protein response causes neuronal death in a neurodegenerative gangliosidosis. *Mol Cell* **15**, 753–766.

23 Takamura A, Higaki K, Kajimaki K, Otsuka S, Ninomiya H, Matsuda J, Ohno K, Suzuki Y & Nanba E (2008) Enhanced autophagy and mitochondrial aberrations in murine G(M1)-gangliosidosis. *Biochem Biophys Res Commun* **367**, 616–622.

24 Sano R, Annunziata I, Patterson A, Moshiach S, Gomero E, Opferman J, Forte M & d'Azzo A (2009) GM1-ganglioside accumulation at the mitochondria-associated ER membranes links ER stress to Ca(2+)-dependent mitochondrial apoptosis. *Mol Cell* **36**, 500–511.

25 Spiegel R, Bach G, Sury V, Mengistu G, Meidan B, Shalev S, Shneor Y, Mandel H & Zeigler M (2005) A mutation in the saposin A coding region of the prosaposin gene in an infant presenting as Krabbe disease: first report of saposin A deficiency in humans. *Mol Genet Metab* **84**, 160–166.

26 Won JS, Singh AK & Singh I (2016) Biochemical, cell biological, pathological, and therapeutic aspects of Krabbe's disease. *J Neurosci Res* **94**, 990–1006.

27 Won JS, Kim J, Paintlia MK, Singh I & Singh AK (2013) Role of endogenous psychosine accumulation in oligodendrocyte differentiation and survival: implication for Krabbe disease. *Brain Res* **1508**, 44–52.

28 Graziano AC, Parenti R, Avola R & Cardile V (2016) Krabbe disease: involvement of connexin43 in the apoptotic effects of sphingolipid psychosine on mouse oligodendrocyte precursors. *Apoptosis* **21**, 25–35.

29 Deodato F, Procopio E, Rampazzo A, Taurisano R, Donati MA, Dionisi-Vici C, Caciotti A, Morrone A & Scarpa M (2017) The treatment of juvenile/adult GM1-gangliosidosis with Miglustat may reverse disease progression. *Metab Brain Dis* **32**, 1529–1536.

30 Condori J, Acosta W, Ayala J, Katta V, Flory A, Martin R, Radin J, Cramer CL & Radin DN (2016) Enzyme replacement for GM1-gangliosidosis: Uptake, lysosomal activation, and cellular disease correction using a novel β-galactosidase:RTB lectin fusion. *Mol Genet Metab* **117**, 199–209.

31 Gupta M, Pandey H & Sivakumar S (2017) Intracellular Delivery of β-Galactosidase Enzyme Using Arginase-Responsive Dextran Sulfate/Poly-l-arginine Capsule for Lysosomal Storage Disorder. *ACS Omega* **2**, 9002–9012.

32 Schalli M, Weber P, Tysoe C, Pabst BM, Thonhofer M, Paschke E, Stütz AE, Tschernutter M, Windischhofer W & Withers SG (2017) A new type of pharmacological chaperone for G(M1)-gangliosidosis related human lysosomal β-galactosidase: N-Substituted 5-amino-1-hydroxymethyl-cyclopentanetriols. *Bioorg Med Chem Lett* **27**, 3431–3435.

33 Hossain MA, Higaki K, Saito S, Ohno K, Sakuraba H, Nanba E, Suzuki Y, Ozono K & Sakai N (2015) Chaperone therapy for Krabbe disease: potential for late-onset GALC mutations. *J Hum Genet* **60**, 539–545.

34 Baek RC, Broekman ML, Leroy SG, Tierney LA, Sandberg MA, d'Azzo A, Seyfried TN & Sena-Esteves M (2010) AAV-mediated gene delivery in adult GM1-gangliosidosis mice corrects lysosomal storage in CNS and improves survival. *PLoS One* **5**, e13468.

35 Przybilla MJ, Ou L, Tăbăran AF, Jiang X, Sidhu R, Kell PJ, Ory DS, O'Sullivan MG & Whitley CB (2019) Comprehensive behavioral and biochemical outcomes of novel murine models of GM1-gangliosidosis and Morquio syndrome type B. *Mol Genet Metab* **126**, 139–150.

36 Rafi MA, Rao HZ, Luzi P, Luddi A, Curtis MT & Wenger DA (2015) Intravenous injection of AAVrh10-GALC after the neonatal period in twitcher mice results in significant expression in the central and peripheral nervous systems and improvement of clinical features. *Mol Genet Metab* **114**, 459–466.

37 Folts CJ, Scott-Hewitt N, Pröschel C, Mayer-Pröschel M & Noble M (2016) Lysosomal Re-acidification Prevents Lysosphingolipid-Induced Lysosomal Impairment and Cellular Toxicity. *PLoS Biol* **14**, e1002583.

38 Marques AR, Willems LI, Herrera Moro D, Florea BI, Scheij S, Ottenhoff R, van Roomen CP, Verhoek M, Nelson JK, Kallemeijn WW, Biela-Banas A, Martin OR, Cachón-González MB, Kim NN, Cox TM, Boot RG, Overkleeft HS & Aerts JM (2017) A Specific Activity-Based Probe to Monitor Family GH59 Galactosylceramidase, the Enzyme Deficient in Krabbe Disease. *Chembiochem* **18**, 402–412.

39 Schröder SP, van de Sande JW, Kallemeijn WW, Kuo CL, Artola M, van Rooden EJ, Jiang J, Beenakker TJM, Florea BI, Offen WA, Davies GJ, Minnaard AJ, Aerts JMFG, Codée JDC, van der Marel GA & Overkleeft HS (2017) Towards broad spectrum activity-based glycosidase probes: synthesis and evaluation of deoxygenated cyclophellitol aziridines. *Chem Commun (Camb)* **53**, 12528–12531.

40 Willems LI, Beenakker TJM., Murray B, Gagestein B, van den Elst H, van Rijssel ER, Codée JDC, Kallemeijn WW, Aerts JMFG, van der Marel GA & Overkleeft HS (2014) Synthesis of alpha- and beta-Galactopyranose-Configured Isomers of Cyclophellitol and Cyclophellitol Aziridine, *Eur J Org Chem* **2014**, 6044–6056.

41 Beenakker TJM (2017) *Design and development of conformational inhibitors and activity-based probes for retaining glycosidases*. (Doctoral dissertation). Retrieved from Leiden University Repository.

42 Devaraj NK, Weissleder R & Hilderbrand SA (2008) Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug Chem* **19**, 2297–2299.

43 Martino S, Tiribuzi R, Tortori A, Conti D, Visigalli I, Lattanzi A, Biffi A, Gritti A & Orlacchio A (2009) Specific determination of beta-galactocerebrosidase activity via competitive inhibition of beta-galactosidase. *Clin Chem* **55**, 541–548.

44 Marques ARA (2016) *Lysosomal glycosidases and glycosphingolipids: New avenues for research*. (Doctoral dissertation). Retrieved from UvA-DARE.

45 Larsbrink J, Thompson AJ, Lundqvist M, Gardner JG, Davies GJ & Brumer H (2014) A complex gene locus enables xyloglucan utilization in the model saprophyte Cellvibrio japonicus. *Mol Microbiol* **94**, 418–433.

46 Witte MD, Kallemeijn WW, Aten J, Li KY, Strijland A, Donker-Koopman WE, Van Den Nieuwendijk AMCH, Bleijlevens B, Kramer G, Florea BI, Hooibrink B, Hollak CEM, Ottenhoff R, Boot RG, Van Der Marel GA, Overkleeft HS & Aerts JMFG (2010) Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat. Chem. Biol.* **6**, 907–913.

47 Wacker H, Keller P, Falchetto R, Legler G & Semenza G (1992) Location of the two catalytic sites in intestinal lactase-phlorizin hydrolase. Comparison with sucrase-isomaltase and with other glycosidases, the membrane anchor of lactase-phlorizin hydrolase. *J Biol Chem* **267**, 18744–18752.

48 Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW 3rd & Su AI (2009) BioGPS: an extensible and customizable portal for querying and organizing gene

annotation resources. *Genome Biol* **10**, R130.

49 Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, Saijo K, Glass CK, Hume DA, Kellie S & Sweet MJ (2008) Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res* **4**, 5.

50 Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I & Pereira-Smith O (1995) A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci USA* **92**, 9363–9367.

51 Lee BY, Han JA, Im JS, Morrone A, Johung K, Goodwin EC, Kleijer WJ, DiMaio D & Hwang ES (2006). Senescence-associated beta-galactosidase is lysosomal beta-galactosidase. *Aging Cell* **5**, 187–195.

52 Hildebrand DG, Lehle S, Borst A, Haferkamp S, Essmann F & Schulze-Osthoff K (2013) α-Fucosidase as a novel convenient biomarker for cellular senescence. *Cell Cycle* **12**, 1922–1927.

53 Lumba MA, Willis LM, Santra S, Rana R, Schito L, Rey S, Wouters BG & Nitz M (2017) A β-galactosidase probe for the detection of cellular senescence by mass cytometry. *Org Biomol Chem* **15**, 6388–6392.

54 Rings EHHM, van Beers EH, Krasinski SD, Verhave M, Montgomery RK, Grand RJ, Dekker J & Büller HA (1994) Lactase; Origin, gene expression, localization, and function. *Nutri Res* **14**, 775–797.

55 Lloyd M, Mevissen G, Fischer M, Olsen W, Goodspeed D, Genini M, Boll W, Semenza G & Mantei N (1992) Regulation of intestinal lactase in adult hypolactasia. *J Clin Invest* **89**, 524–529.

56 Kuo CL, Kallemeijn WW, Lelieveld LT, Mirzaian M, Zoutendijk I, Vardi A, Futerman AH, Meijer AH, Spaink HP, Overkleeft HS, Aerts JMFG & Artola M (2019) In vivo inactivation of glycosidases by conduritol B epoxide and cyclophellitol as revealed by activity-based protein profiling. *FEBS J* **286**, 584–600.

57 Artola M, Kuo CL, Lelieveld LT, Rowland RJ, van der Marel GA, Codée JDC, Boot RG, Davies GJ, Aerts JMFG & Overkleeft HS (2019) Functionalized Cyclophellitols Are Selective Glucocerebrosidase Inhibitors and Induce a Bona Fide Neuropathic Gaucher Model in Zebrafish. *J Am Chem Soc* **141**, 4214–4218.

58 Jiang J, Beenakker TJ, Kallemeijn WW, van der Marel GA, van den Elst H, Codée JD, Aerts JM & Overkleeft HS (2015) Comparing Cyclophellitol N-Alkyl and N-Acyl Cyclophellitol Aziridines as Activity-Based Glycosidase Probes. *Chemistry* **21**, 10861–10869.

59 Kallemeijn WW, Witte MD, Voorn-Brouwer TM, Walvoort MT, Li KY, Codée JD, van der Marel GA, Boot RG, Overkleeft HS & Aerts JM (2014) A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining β-glucosidases. *J Biol Chem* **289**, 35351–35362.

60 Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

61 Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G & Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423

62 Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA & Godzik A (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405.

63 Gupta R & Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*, 310–322.

64 Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ & Nakai K (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, W585–587.

APPENDIX

7.S1. Supplementary Tables and Figures

Table 7.S1. Apparent IC_{50} values for compounds towards CjGH35A, GBA, and GBA2. Error range = ± SD, *n* = 2 biological replicates.

Compound	CjGH35A
2	21.55 ± 2.14
3	1.22 ± 0.66
4	27.16 ± 1.57
5	33.9 ± 2.15

Compound	GBA	GBA2
2	1,560	3,290
3	14.6	164
4	5,370	> 10,000
5	85.8	752
6	119	1,830
SYD215	430	34.9

Table 7.S2. Structural determination and refinement statistics for CjGH35A in complex with 2 (University of York).

	CjGH35A in complex with **5** PDB code: 5JAW
Data collection	
Space group	P1
Cell dimensions	
a, *b*, *c* (Å)	98.9, 115.8, 116.0
α, β, γ (°)	90.2, 90.2, 90.4
Resolution (Å)	81.99-1.6 (1.63-1.60)
R_{merge}	0.080 (0.588)
R_{pim}	0.080 (0.588)
CC(1/2)	0.979 (0.482)
$I/\sigma I$	6.0 (1.2)
Completeness (%)	95.7 (94.1)
Redundancy	1.8 (1.8)
Refinement	
Resolution (Å)	1.60
No. reflections	650665
R_{work} / R_{free}	0.20/0.21
No. atoms	
Protein	33482
Ligand/ion	125
Water	1637
B-factors (Å^2)	
Protein	25
Ligand/ion	20
Water	28
R.m.s. deviations	
Bond lengths (Å)	0.018
Bond angles (°)	1.927

*Values in parentheses are for highest-resolution shell.

Table 7.S3. Comparison between GLB1 and GLB1-like proteins in A) mouse and B) human.

A

Protein	**GLB1**	**GLB1-like 1**	**GLB1-like 2**	**GLB1-like 3**
Gene	*Glb1*	*Glb1l*	*Glb1l2*	*Glb1l3*
Animo acid	647 (isoform 1)	646	652 (isoform 1)	662
Molecular weight	73.12	73.28	73.96	75.59
N-terminal signal sequence	1-24	1-23	No	No
Transmembrane domain	No	No	13-35	No
N-glycan sites	Asn27, Asn248, Asn500, Asn504, Asn510	Asn93, Asn239	No	No
Cellular localization	Lysosomal	Extracellular	Plasma membrane (outward-facing)	Mitochondrial / peroxisomal

B

Protein	**GLB1**	**GLB1-like 1**	**GLB1-like 2**	**GLB1-like 3**
Gene	*GLB1*	*GLB1L*	*GLB1L2*	*GLB1L3*
Animo acid	677 (isoform a)	654 (isoform 1)	636	653
Molecular weight	76.075	74.158	72.079	74.823
N-terminal signal peptide	1-23	1-27	No	No
Transmembrane domain	No	No	13-32	No
N-glycan sites	Asn26, Asn247, Asn464, Asn498, Asn542, Asn545, Asn555	Asn97, Asn243, Asn625	No	No
Cellular localization	Lysosomal	Mitochondrial / extracellular	Plasma membrane (outward-facing)	Mitochondrial / cytosolic

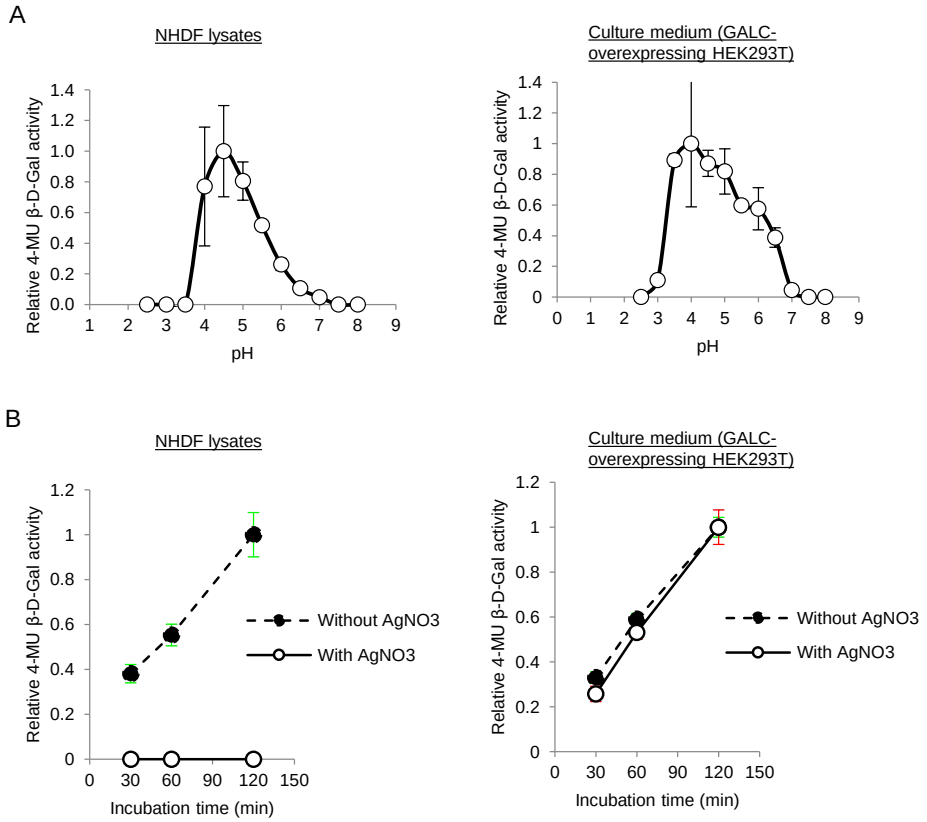


Figure 7.S1. Discriminating activities from GLB1 and GALC in different samples by enzymatic assay. A) pH-dependent β-galactosidase activity (relative to the highest measured activity). B) Effect of $AgNO_3$ (11 μM). Error range = ± SD, *n* = 3 technical replicates.

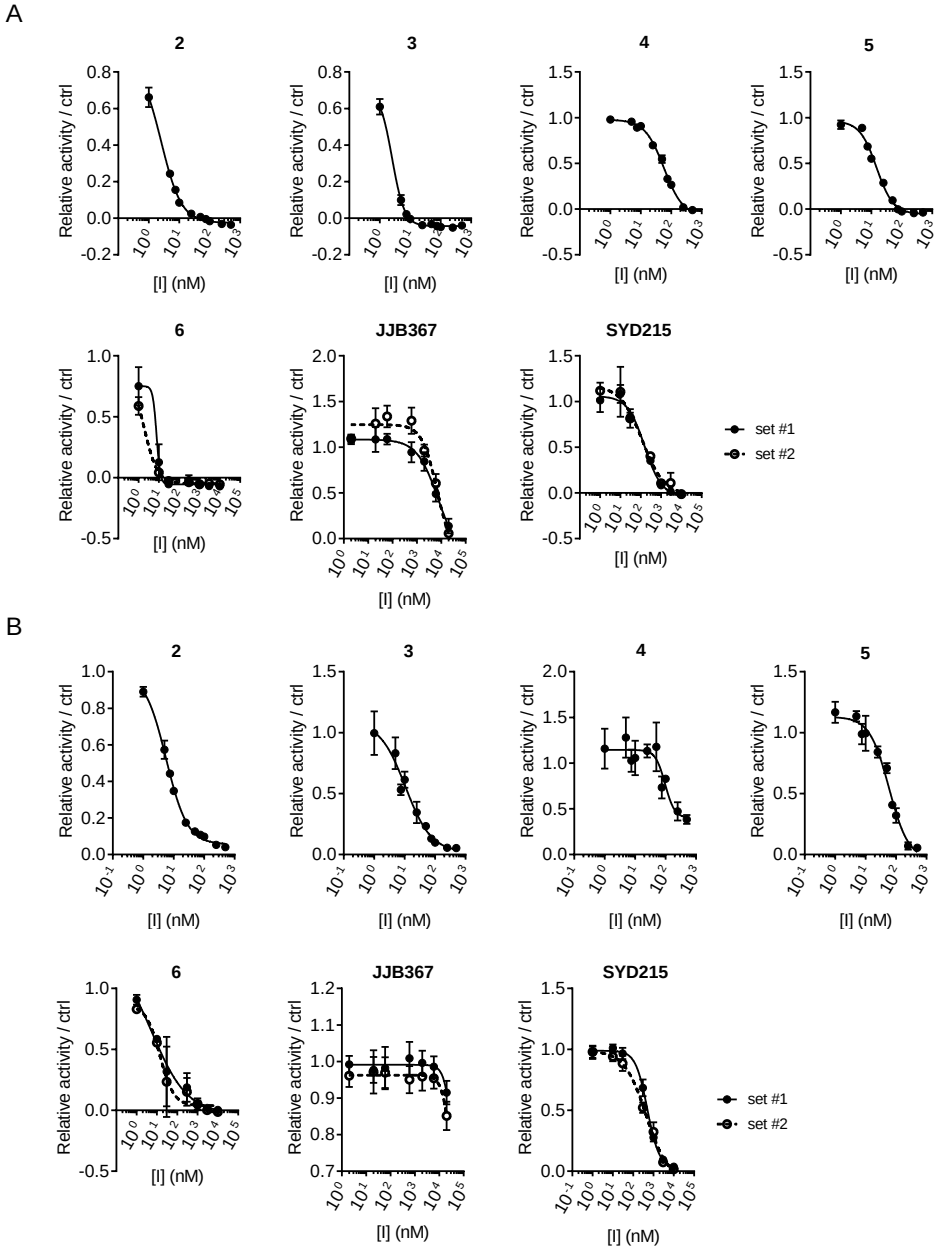


Figure 7.S2. (Continued, 1 of 2).

C

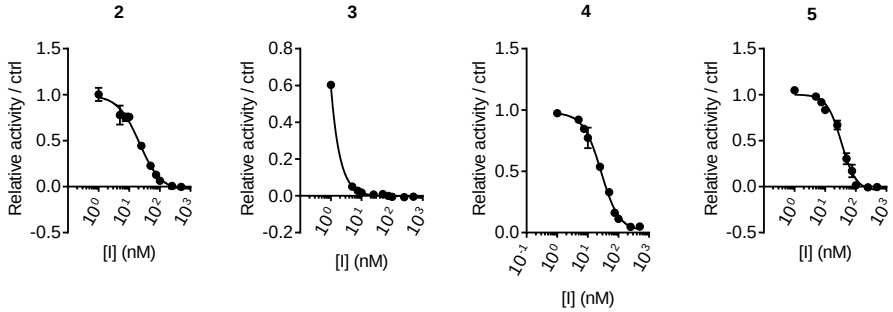


Figure 7.S2. Inhibition curves by compounds towards A) human GLB1 (hGLB1, NHDF lysates), B) mouse GALC (mGALC, from culture medium of HEK293T cells overexpressing the protein), and C) CjGH79. Error range = ± SD (*n* = 3 technical replicates).

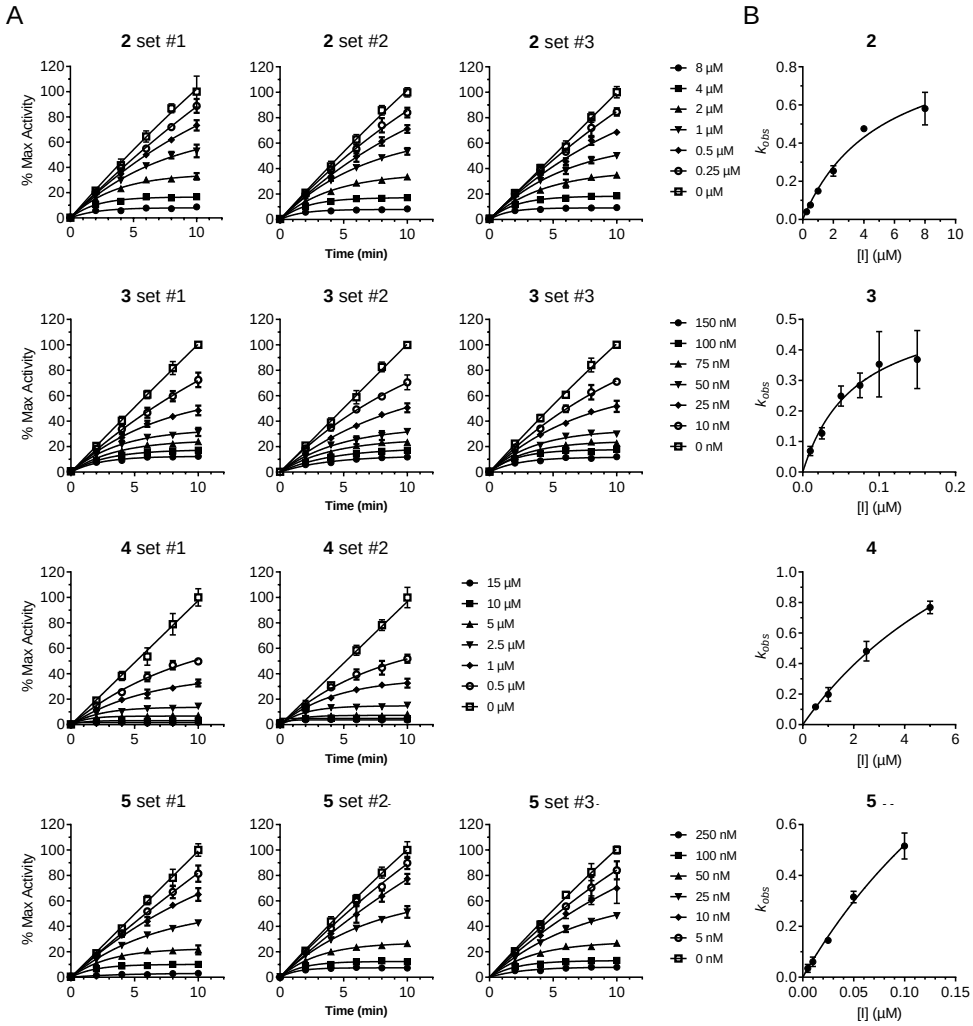


Figure 7.S3. Measurement of inhibition kinetic parameters by compounds 4-5 towards CjGH79. A) Processing curves. Error range = ± SD, *n* = 3 technical replicates. B) k_{obs} vs [I] plots, fitted with Michaelis-Menten equation. Error range = ± SD, *n* = 3 biological replicates.

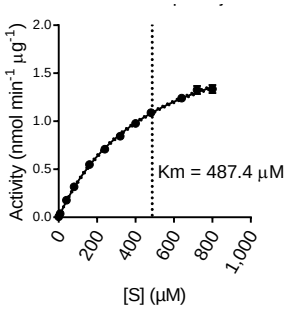


Figure 7.S4. Michaelis-Menten plot of 4-MU-β-gal towards CjGH35. Error range = ± SD, *n* = 3 technical replicates.

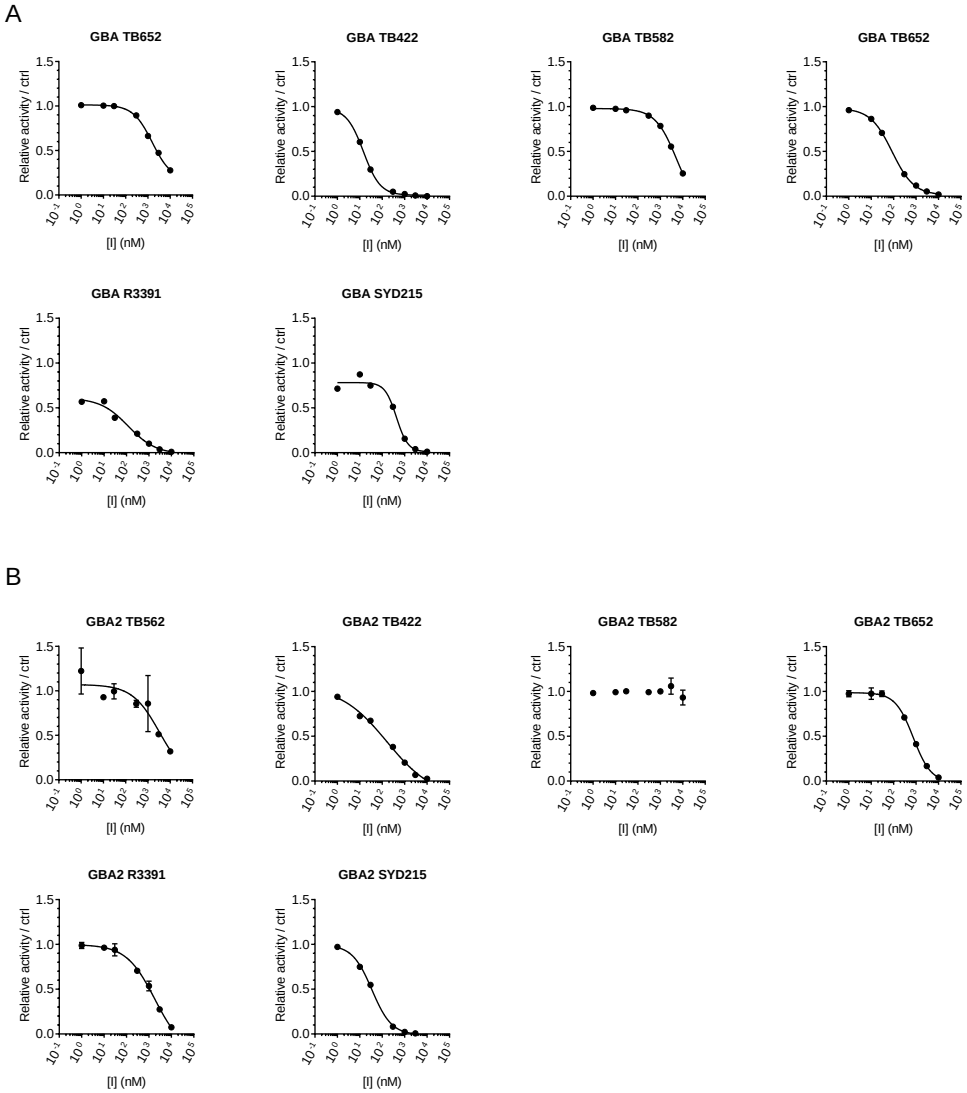


Figure 7.S5. Inhibition curves by compounds towards A) GBA (imiglucerase) and B) GBA2 (lysates of HEK293T cells overexpressing GBA2). Error range = ± SD, *n* = 3 technical replicates.

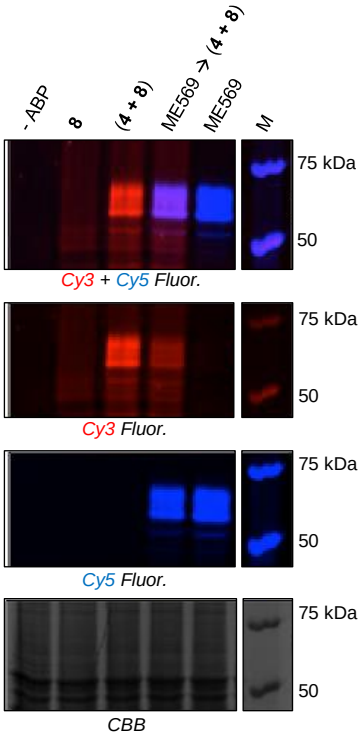


Figure 7.S6. Labeling by the in situ generated probe (4 + 8) in human fibroblast (NHDF) lysates with or without pre-incubation with the GBA-specific ABP ME569. *CBB*, Coomassie Brilliant Blue staining for assessing total protein loading.

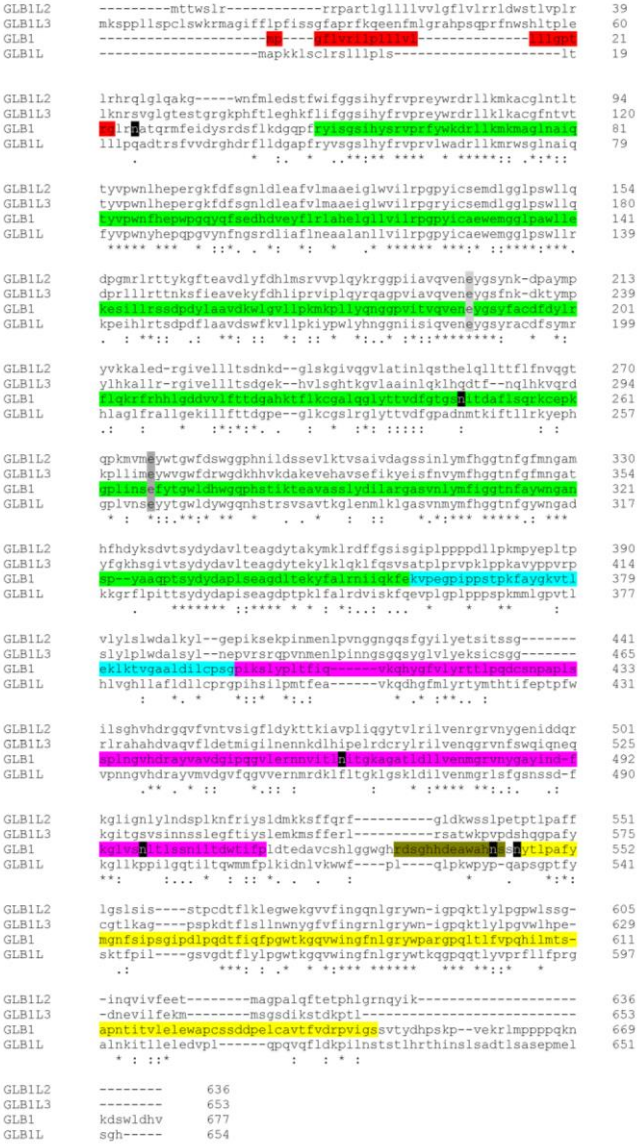


Figure 7.S7. Comparing amino acid sequences between GLB1 and GLB1-like proteins. Red, signal peptide of GLB1. Green, TIM barrel domain of GLB1. Cyan, TIM- β1 loop. Magenta, β-sheet domain 1 of GLB1. Dark yellow, cleaved peptide in mature GLB1. Yellow, β-sheet domain 2 of GLB1. Black, N-linked glycan sites of GLB1. Light gray, catalytic acid/base. Dark gray, catalytic nucleophile. Domain structures assigned based on Ohto *et al.*[main text ref 8]

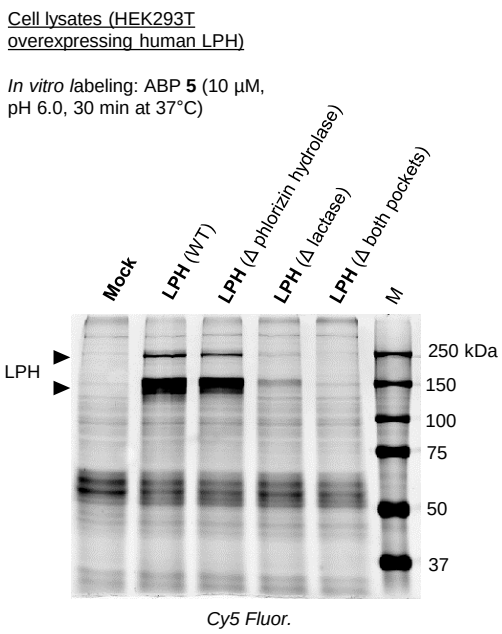
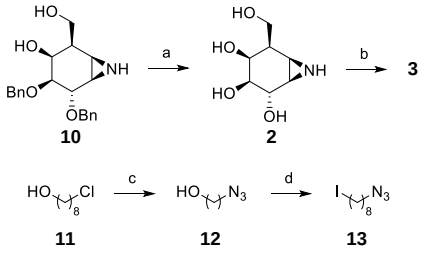
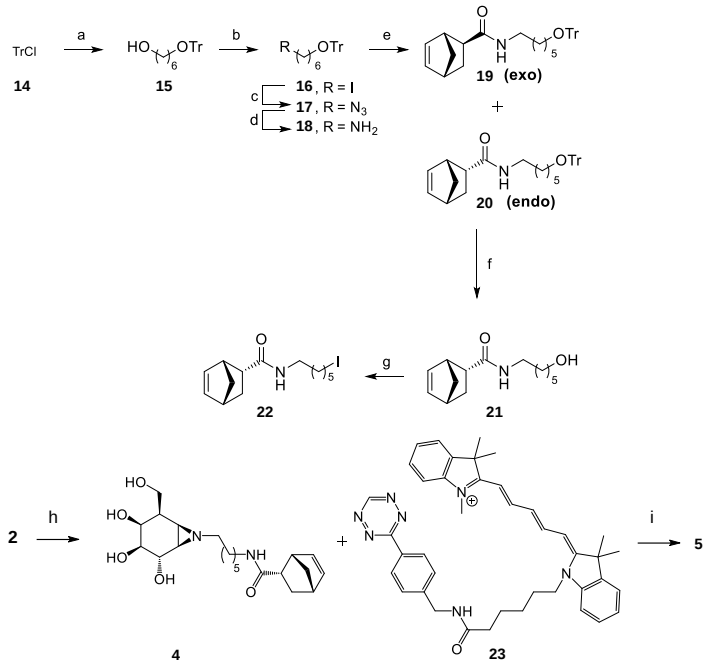


Figure 7.S8. ABP 5 labeling in HEK293T lysates containing expressed human wild-type or mutant LPH.

7.S2. Synthetic strategies for compounds used in this chapter (Department of Bio-Organic Synthesis, Leiden University)



Scheme 7.S1. Synthesis of compounds 2 and 3. Reagents and conditions: a) Li, NH_3, -60 °C, b) **10**, K_2CO_3, DMF, 80 °C, overnight, 25 %; c) NaN_3, DMF, 80 °C, quant; d) PPh_3, imidazole, I_2, THF, -20 °C to RT, 15 min, 69 %.



Scheme 7.S2. Synthesis of compounds 4 and 5. Reagents and conditions: a) 1,6-hexanediol, pyridine, CH_2Cl_2, RT, 90 min, 95 %; b) imidazole, PPh_3, I_2, Et2O, CH_3CN, RT, overnight, 80 %; c) NaN_3, DMF, 80 °C, overnight, quant; d) PPh_3 on beads, H_2O, THF, 48 h, quant; e) norbornene-OSu[1], DIPEA, DCE, RT, overnight, **19** 28 %, **20** 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) **22**, K_2CO_3, DMF, 75 °C, overnight, 12 %. i) Cy5 tetrazine **23**, MeOH, overnight, 87 %.

7.S3 Protein crystallography (University of York)

The gene for CjGH35A, cloned into a pET28a vector modified for Ligation Independent Cloning, was expressed, and the protein purified and crystallized as described previously[main text ref 45], in 2.7 M sodium acetate pH 7.2 (protein at 30 mg/ml, drop 1.2:1 μL over well). A crystal was soaked in the presence of a speck of powder of compound **2** for 70 h. The crystal was fished directly into liquid nitrogen without the need for additional cryoprotectant. Data were collected on beamline IO2 at the Diamond Light Source at wavelength 0.97950 Å, and were processed using DIALS[2] and scaled with AIMLESS[3] to 1.6 Å (Table 7.S2). The space group was P1 and the unit cell dimensions, 98.9, 115.8, 116.0 Å, and angles, 90.2, 90.2, 90.4°. The structure was solved using programs from the CCP4I2 suite. Molecular replacement was performed using Phaser[4], with the native structure, PDB entry 4d1i, as the model. The model was built manually

in Coot[5], followed by repeated cycles of REFMAC[6] employing twin refinement using observed intensities. The structure was annotated in PDB (5JAW).

7.S4 Supplementary References

1 Willems LI, Li N, Florea BI, Ruben M, van der Marel GA & Overkleeft HS (2012) Triple bioorthogonal ligation strategy for simultaneous labeling of multiple enzymatic activities. *Angew Chem Int Ed Engl* **51**, 4431–4.

2 Gildea RJ, Waterman DG, Parkhurst JM, Axford D, Sutton G, Stuart DI, Sauter NK, Evans G & Winter G (2014) New methods for indexing multi-lattice diffraction data. *Acta Crystallogr D Biol Crystallogr* **70**, 2652–2666.

3 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214.

4 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.

5 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.

6 Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-–367.