CHAPTER 4

Biological characterization of activity-based probes for α-glucosidase and β-glucuronidase

Manuscript published as:

(1) Jiang J, **Kuo CL**, Wu L, Franke C, Kallemeijn WW, Florea BI, van Meel E, van der Marel GA, Codée JDC, Boot RG, Davies GJ, Overkleeft HS & Aerts JMFG (2016) Detection of active mammalian GH31 α-glucosidases in health and disease using in-class, broad-spectrum activity-based probes. *ACS Central Science* **2**, 351–358.

(2) Wu L, Jiang J, Jin Y, Kallemeijn WW, **Kuo CL**, Artola M, Dai W, van Elk C, van Eijk M, van der Marel GA, Codée JDC, Florea BI, Aerts JMFG, Overkleeft HS & Davies GJ (2017) Activity-based probes for functional interrogation of retainingβ-glucuronidases. *Nat Chem Biol* **8**, 867–873.

ABSTRACT

The lysosomal glycosidases α-glucosidase (GAA) and β-glucuronidase (GUSB) catalyze fragmentation of glycogen and glycosaminoglycans, respectively.. Their deficiency due to genetic aberrations leads to the lysosomal storage disorder (LSD) Pompe disease (GAA deficiency) and mucopolysaccharidosis type VII (MPSVII; Sly Syndrome), both severe disease conditions with still unmet clinical needs. To generate tools to facilitate biochemical and biomedical research on these two enzymes, activity-based probes targeting either enzyme (GH31 α-glucosidases or GH2 β-glucuronidase) were developed and evaluated for their mechanism-of-action, inhibition potency, and target selectivity. Both sets of ABPs successfully label their respective target enzyme class in activity-based manner. They are nanomolar irreversible inhibitors with fast inhibition kinetics. At acidic labeling pH and in cells, the lysosomal acid α-glucosidase GAA or β-glucuronidase GUSB are labeled by their respective ABPs, as revealed by both in-gel fluorescence and chemical proteomics. In the case of both GAA and GUSB ABPs, some minor out-of-class targets are observed. By altering the labeling pH or pre-incubating samples with previously generated inhibitors for the off-targets, GAA or GUSB can be visualized with their corresponding ABPs simultaneously with the off-target glycosidases. The new ABPs for GH31 α-glucosidases and GH2 β-glucuronidase successfully expand the chemical toolbox for lysosomal glycosidases and should find future applications in studies of fundamental and applied LSD research.

4.1 Introduction

Deficiency in lysosomal proteins due to genetic mutations underlies lysosomal storage disorders (LSDs) in which undegraded substrate accumulate in the lysosomes of particular cell types in the affected patients, and lead to a complex cascade of downstream pathological events[1]. An important class of deficient proteins in LSDs are the lysosomal glycosidases (E.C. 3.2.1), accounting for a third of known LSD enzymopathies. Glycosidases can be broadly classified by their substrate preference on the glycon part (e.g. sugar configuration, exo-acting *vs* endo-acting), mechanism of hydrolysis (e.g. retaining, inverting, substrate assisted), or evolutionary similarity by amino acid sequence and folding by the glycoside hydrolase (GH) family system.[1]

Man expresses at least four exo-acting retaining α-glucosidases that belong to the glycoside hydrolase family 31 (GH31): the α-subunit of ER α-glucosidase II (GANAB) (E.C. 3.2.1.20 and 3.2.1.84) which plays a role in quality control of N-glycoprotein synthesis, the two intestinal enzymes maltase-glucoamylase (MGAM) (E.C. 3.2.1.20) and sucrase-isomaltase (SI) (E.C. 3.2.1.20 and 3.2.1.48) degrading dietary glycosides, and the lysosomal acid α-glucosidase (GAA) (E.C. 3.2.1.20) degrading glycogen.[2] The human GAA is formed as a 110 kDa proenzyme, that upon mannose-6-phosphate dependent transport to the lysosome is processed into the 76 kDa and 70 kDa active forms.[3, 4] Mutations in GAA gene may lead to Glycogen storage disease type II, also known as Pompe disease.[5] This relatively common LSD (1:40,000 live birth worldwide)[6] is characterized by the intra-lysosomal accumulation of undegraded glycogen, which causes progressive muscle weakness in the heart and skeletal muscles and eventually affects the liver and central nervous system.[7] Classification of the disease is based on the age of onset, and patients classified into the infantile form (4 to 8 months of age) usually do not survive before one year of age.[8] Later onset forms (juvenile and adult forms) are characterized by a slower progressive decrease in muscle strength, firstly in the legs and then the trunks and the arms, and finally to fatality through respiratory failure.[9] Severity of Pompe disease correlates with the extent of lost GAA activity.[10] The disease is currently treated by chronic intravenous injection of recombinant GAA (rGAA, alglucosidase alpha, Myozyme®), which impressively delays the fatal

symptoms in infantile patients.[11] However, due to poor correction in muscle cell pathology, it only has limited response in adult patients.[12]

Another LSD-relevant lysosomal glycosidase, the retaining exo-β-glucuronidase (GUSB), belongs to the GH2 family.[2] It catalyzes the hydrolysis of terminal β-glucuronic acid from heparan sulfates, dermatan sulfates, and chondroitin-4,6-sulfate—polysaccharides presented in glycosaminoglycans (GAGs) which are found in the extracellular matrix that serve structural and signaling functions. Deficiency of GUSB activity due to hereditary genetic mutations forms the basis of another LSD known as mucopolysaccharidosis type VII (MPSVII), or Sly syndrome.[13] Affected individual are characterized by lysosomal accumulation of GAGs in many tissues, which ultimately affects the central nervous system and causes hepatosplenomegaly and bone dysplasia.[14] No targeted therapy is currently available for this rare disease. Notably, there exists another lysosomal enzyme in human that has exhibits β-glucuronidase activity—heparanase (HPSE). This is an endo-acting retaining β-glucuronidase, and belongs to the GH79 family, that functions in the turnover of heparan sulfate.[2] Interestingly, the bacterial GH79 homologue from *Acidobacterium capsulatum,* that exhibits similar protein fold, is an exo-acting retaining β-glucuronidase.[15] HPSE activity is implicated in disease-related processes such as inflammation, tumor metastasis and angiogenesis.[16, 17]

In the past, the mechanism-based irreversible β-glucosidase inhibitor cyclophellitol (CP) has been used as a scaffold to generate potent and specific inhibitors and ABPs towards the lysosomal glucocerebrosidase (GBA).[1] The configurational and functional isomers of CP-aziridine have been found to inactivate and label other classes of retaining exo-glycosidases that follow the formal Koshland double-displacement mechanism.[1] Both GAA and GUSB are retaining exo-glycosidases (**Fig. 4.1A**),[18] which makes them possible candidates for activity-based probe labeling (**Fig. 4.1B**). It is therefore hypothesized that the α-glucose and β-glucuronic acid configured cyclophellitol epoxides and aziridines may render inhibitors and ABPs targeting GH31 retaining exo-α-glucosidases (GAA and GANAB) and GH2 retaining exo-β-glucuronidase (GUSB), respectively. In this chapter the characterization of these newly generated compounds is presented in detail, which provides a basis for future laboratory, pre-clinical, and

clinical applications.

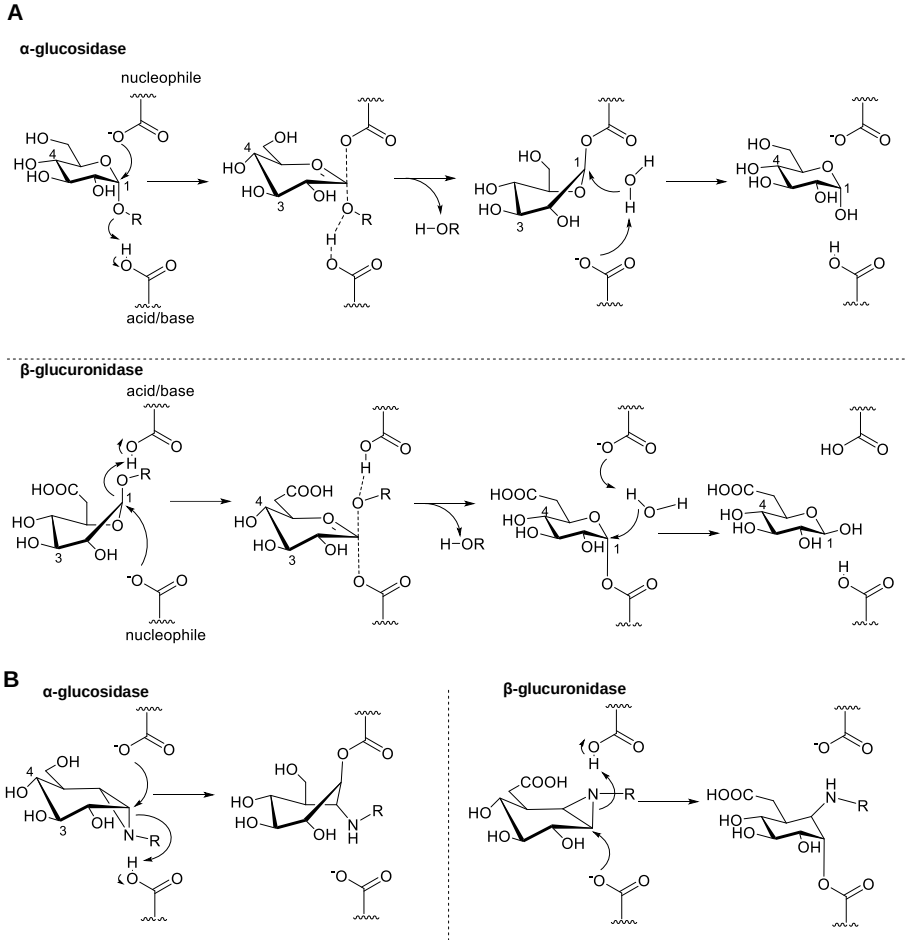


Figure 4.1. Reaction mechanisms of compounds used in this chapter. A) Koshland double-displacement mechanism employed by retaining α-glucosidase and β-glucuronidase. B) Proposed inhibition mechanism of cyclophellitol aziridine-based α-glucosidase inhibitor (left) and β-glucuronidase inhibitor. The corresponding ABPs may be obtained by grafting a detection group (fluorophore or biotin) at the R position.

4.2 Results and Discussion

4.2.1 Synthesis of α-glucose- and β-glucuronic acid-configured inhibitors and ABPs

Compounds were synthesized at the at the Department of Bio-organic Synthesis, Leiden University. The α-glucose-configured-configured cyclophellitol (**1**) and cyclophellitol aziridines (**2**) were synthesized (**Fig. 4.2**) from a cyclohexane diol (**17**) (see **Scheme 4.S1** for synthetic strategy, and reference Jiang et al.[19] for synthetic and characterization details of compounds). Compound **2** was alkylated to give compound **3**, which was further reacted with different substituted alkynes by copper (I)-catalyzed [2+3] azide-alkyne click reaction to afford ABPs **4**–**7** (**Fig. 4.2**, **Scheme 4.S1**).

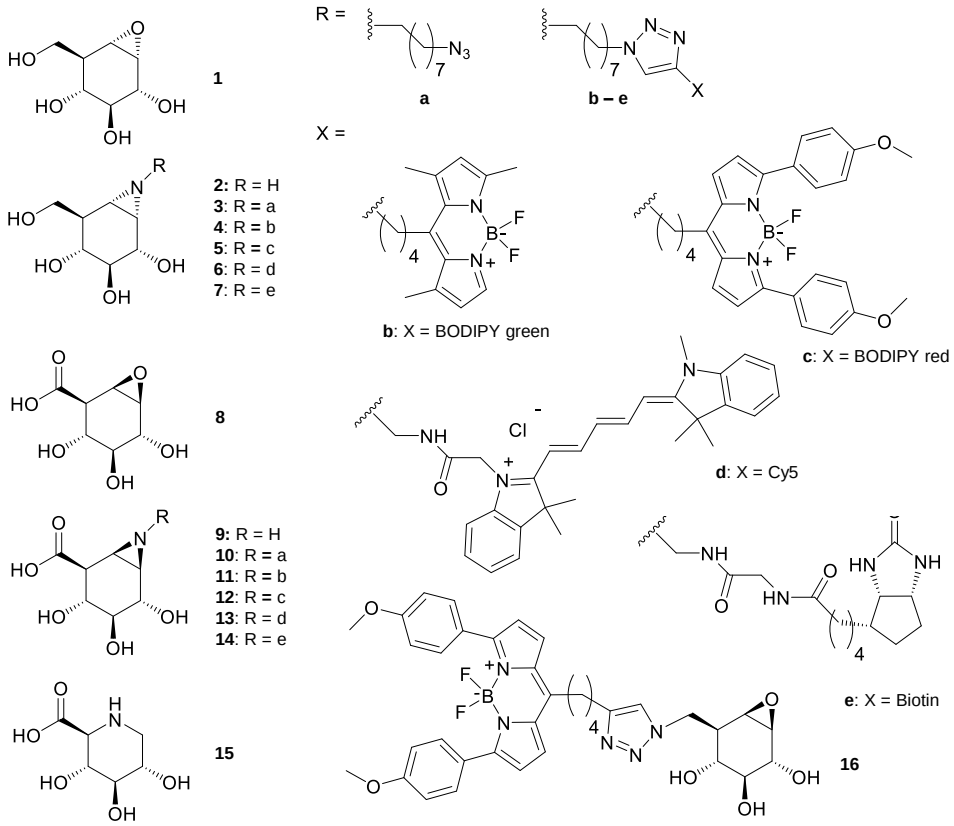


Figure 4.2. Structures of compounds used in this chapter.

The β-glucuronic acid-configured cyclophellitol **8** and cyclophellitol aziridine **9** (**Fig. 4.2**) were also synthesized from **17** (**Scheme 4.S2**). The synthesis of *N*-alkyl-substituted β-glucuronic acid-configured cyclophellitol aziridines utilized the *N*-alkyl-substituted β-glucose-configured cyclophellitol aziridine[20] as a precursor, and upon selective oxidation at the primary alcohol, the β-glucuronic acid-configured compound **10** was obtained (**Fig. 4.2**, **Scheme 4.S3**). Further click reaction afforded the *N*-alkyl-substituted β-glucuronic acid-configured ABPs **11–14** (**Fig. 4.2**). The iminosugar **15** (**Fig. 4.2**) was synthesized as a putative GUSB inhibitor, according to the strategy depicted in **Scheme 4.S4**.

4.2.2 *In vitro* inhibition and labeling of enzyme targets by compounds **1–14**

At the start of the evaluation of the newly generated compounds, the therapeutic recombinant human GAA (rGAA) and recombinant GH79 retaining exo-β-glucuronidase (AcGH79) from the bacteria *Acidobacterium capsulatum* were used as enzyme sources. The inhibitory activity of compounds **1–7** towards rGAA, and compounds **8–14** on AcGH79, were determined by measuring the apparent IC_{50} value of each compound at 30 min incubation time using fluorogenic 4-methylumblliferyl (4-MU)-glycoside substrate assays. As can be seen in **Fig. 4.3A** (top panel), the value for the α-glucose configured cyclophellitol epoxide **1** was at a low-micromolar apparent range (14.6 ± 1.6 μM) towards rGAA; the values for cyclophellitol aziridine **2** and its alkyl derivative **3** were at mid-nanomolar range (apparent IC_{50} = 30–50 nM), and that the values for ABPs **4–7** were at high-nanomolar range (apparent IC_{50} = 200–800 nM). For the β-glucuronic-acid-configured compounds **8–14**, all exhibited nanomolar apparent IC_{50} values towards AcGH79. The *N*-alkyl-substituted compound **10** and ABPs **11–14** were the most potent, reaching low- to sub-nanomolar potency (**Fig. 4.3A**, lower panel). The aziridine compound **9** and the epoxide compound **8** were also nanomolar inhibitors towards AcGH79 (**Fig. 4.3A**, lower panel), but slightly less reactive compared to the N-alkylated compounds.

Next, the formation of covalent enzyme-ABP complex was examined by fluorescent gel-based ABPP assay. ABP **4** or ABP **11**, both containing a green BODIPY fluorophore, were incubated with their respective target enzyme (rGAA or AcGH79) at 1 μM concentration for 30

minutes. Afterwards, samples were denatured and subjected to SDS-PAGE. The subsequent fluorescent scanning on the wet slab-gels containing the rGAA samples (**Fig. 4.3B**, left panel) revealed a clear green band corresponds to the expected molecular weight for rGAA (~100 kDa) in the ABP **4**-treated lane (second lane of the gel). This band is absent in samples without ABP treatment or pre-incubated with compounds **1–3** and **7**, the GAA substrate 4-MU-α-glc and maltose, and the protein-denaturing agent SDS. In samples pre-incubated with ABP **5** or **6**, red

A

Compounds	IC_{50} (nM) (Myozyme)
1 (JJB307)	14,600 ± 1,600
2 (CF21)	37.8 ± 3.40
3 (CF22)	51.7 ± 2.30
4 (JJB382)	459 ± 38.4
5 (JJB347)	383 ± 22.10
6 (JJB383)	269 ± 7.10
7 (JJB384)	767 ± 3.50

Compounds	IC_{50} (nM) (AcGH79)
8 (JJB400)	33.4 ± 3.1
9 DW222	4.6 ± 0.03
10 (JJB355)	0.1 ± 0.01
11 (JJB395)	0.6 ± 0.2
12 (JJB391)	0.8 ± 0.2
13 (JJB392)	1.1 ± 0.1
14 (JJB397)	0.4 ± 0.02

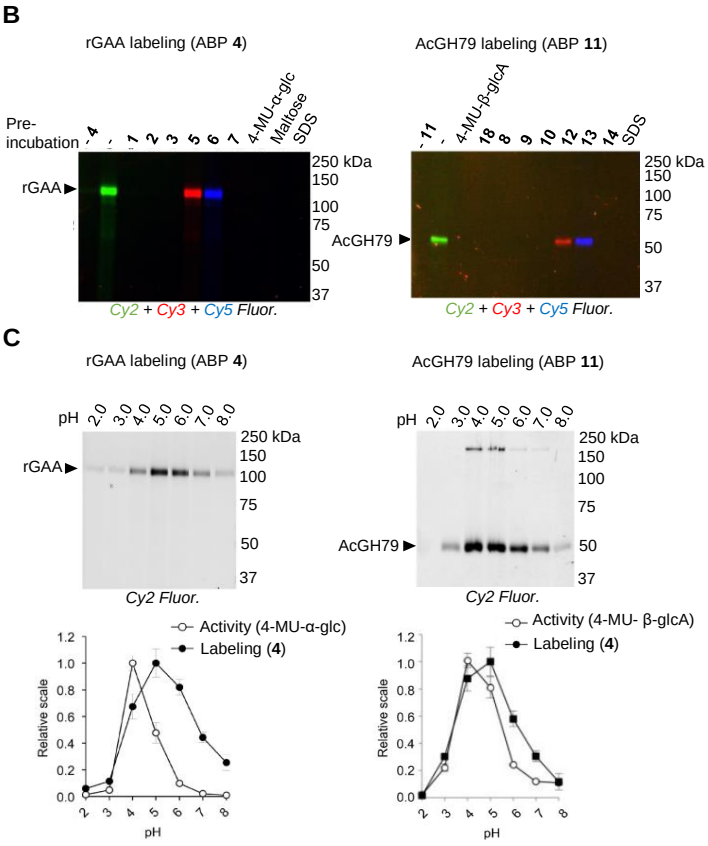


Figure 4.3. In vitro inhibition and labeling on recombinant enzymes by compounds 1–14. A) Apparent IC_{50} values for compounds **1–7** towards rGAA (top) and for compounds **8–14** towards AcGH79 (bottom). B) Fluorescent gel-based cABPP by compound **1–7** towards rGAA (left) and by compounds **8–14** towards AcGH79 (right). Enzymes were pre-incubated with inhibitors shown on the top of gels, and labeled with ABP **4** or **11**. C) pH-dependent labeling of ABP **4** (left) or ABP **11** (right) towards corresponding enzymes. The intensity of the labeled bands was quantified, and compared with the activity measured by fluorogenic substrates at each concentration (lower panels).

or Cy5 (shown in blue) fluorescence replaced the green fluorescence, suggesting the labeling of these ABP on rGAA competed away for the one from ABP **4**. Similarly, the gel containing the AcGH79 samples (**Fig. 4.3B**, right panel) showed a comparable labeling pattern. A green fluorescent band at the expected molecular weight of AcGH79 (~50 kDa) was observed in the ABP **11**-treated sample (second lane), and absent (or replaced by fluorescence of other colors) in samples pre-incubated with compounds **8–10** and **12–14** for AcGH79), known substrate (4-MU-β-glcA), putative competitive enzyme inhibitor (compound **15**), or 2 % (w/v) SDS. These results suggested that the labeling of ABP **4** or **11** towards their respective enzyme target occurred at the active site. To further evaluate whether the labeling by these ABPs occurred in an activity-based manner, ABP labeling on both enzymes was examined for the pH-dependence. For this, enzymes were labeled by ABPs across a range of pH (**Fig. 4.3C**, top panels), and the intensity of ABP labeling were compared to the measured enzymatic activity across the same pH range. The comparison revealed that the pH optimum of 5.0 for ABP labeling differed only slightly with that for substrate hydrolysis at pH 4.0 (**Fig. 4.3C**, lower panels).

Next, the kinetics for ABP labeling was examined. Both ABP **4** and **11** showed time-dependent labeling on their respective enzyme target, and reached saturation during 10–15 minutes of incubation time (**Fig. 4.S1**). The inhibition kinetic parameters were measured for the β-glucuronic acid-configured compound **8–14** on AcGH79 by fluorogenic substrate assay in a continuous setup (see **4.S2 Supporting note**), and the results revealed that beside the epoxide **8**, all compounds had pseudo-first order rate constant (k_i/K_I) in the rage of 3.5–25 μM^{-1} min^{-1} (**Table 4.1**; see **Fig. 4.S2** for processing curves and k_{obs} vs [I] data, **Fig 4.S3** for Michaelis-Menton plot). These values were comparably high (i.e. fast in labeling) to the reported values for the previously generated cyclophellitol ABPs towards rGBA (0.8–25 μM^{-1} min^{-1}).[21]

Table 4.1. Kinetic parameters for compounds 8–14 towards AcGH79. Error ranges = SD from triplicate sets of experiment.

Compounds	8 JJB400	9 DW222	10 JJB355	11 JJB395	12 JJB391	13 JJB392	14 JJB397
k_i/K_I (μM^{-1} min^{-1}) (AcGH79)	0.49 ± 0.05	3.5 ± 0.2	18.8 ± 0.7	25.0 ± 0.7	18.2 ± 0.9	14.0 ± 0.8	5.5 ± 0.2

4.2.3 Structural analysis of compounds in complex with recombinant enzymes

The mechanistic aspect of inhibition/labeling by the compounds were further examined by using protein crystallography studies (University of York), which could reveal structural insights how compounds bind in the active-site pocket of an enzyme, such as the formation of covalent bond, or the conformation that compounds adopt. For this purpose, crystals of the bacterial GAA homologue CjAgd31B from *Cellvibrio japonicus* or the bacterial GUSB homologue AcGH79 (wild-type or the E287Q nucleophile mutant) were soaked with either compound **3** or compound **10**, before subjecting to X-ray crystallography and structural determination. The results revealed that compound **3** adopted a $^{1}S_{3}$ conformation in the active site of CjAgd31B (**Fig. 4.4A**), identical with the known conformation adopted by other GAA substrates[22, 23] (**Fig. 4.1A**, third structure in the top panel). It also showed a covalent glycosidic bond between anomeric carbon (C1) of **3** and the nucleophile Asp412 of the enzyme, confirming the predicted mechanism-based inhibition/labeling (**Fig. 4.1**).

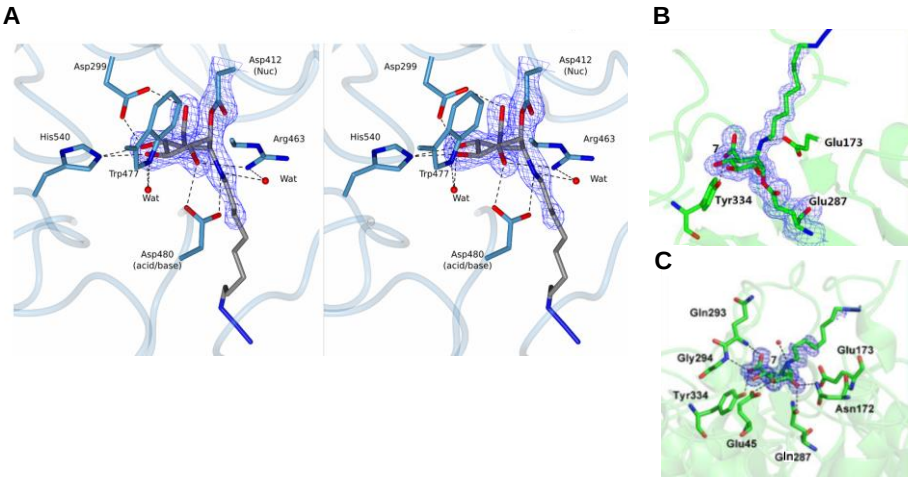


Figure 4.4. Structures of inhibitor-enzyme complex determined by protein crystallography. A) Spectroscopic view of compound **3** in the active site of CjAgd31B. B) Structure of compound **10** in the active site of AcGH79. C) Structure of compound **10** in the active site of the E287Q catalytic nucleophile mutant of AcGH79.

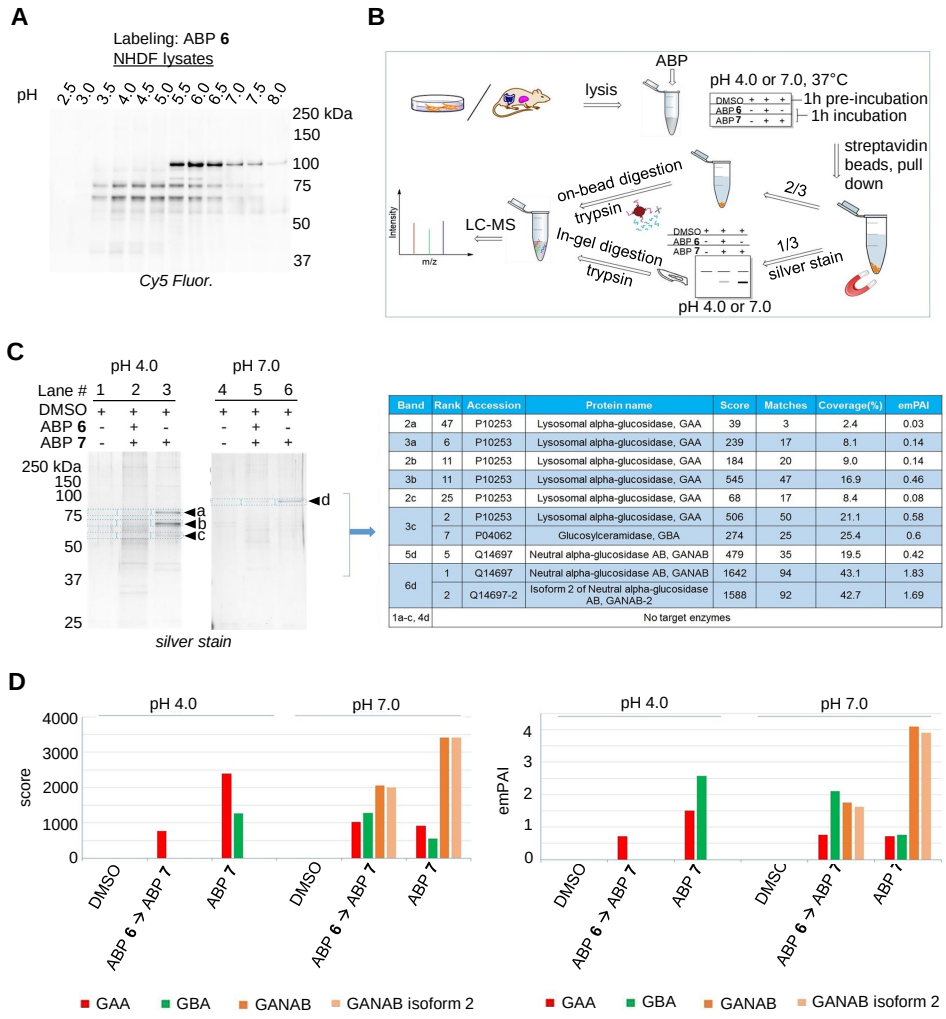
The structure of compound **10** coupled to AcGH79 revealed a $^{4}C_{1}$ conformation adopted by the compound, as well as the presence of the trans-diaxial C1-nucleophile glycosidic linkage.

On the other hand, the structure of **10** in complex with the E287Q catalytic nucleophile mutant of AcGH79 revealed a $^4H_3$ conformation adopted by the compound, and the absence of the glycosidic linkage between the compound and the enzyme (**Fig. 4.4C**). This demonstrated that the compound is a transition state-mimic of the substrates of β-glucuronidase, and that the inhibition/labeling followed the expected $^4H_3 \rightarrow {}^4C_1$ itinerary from the transition-state to the covalent substrate-enzyme intermediate (**Fig. 4.1**).[15]

4.2.4 Glycosidase targets of ABPs **4**–**7** in complex biological samples

Whether the generated ABPs are applicable to detecting GAA or GUSB in complex biological samples was tested next. For GAA, lysates of human fibroblasts were incubated with 1 μM of the Cy5 ABP **6** for 30 minutes at 37°C at pH range from 2.5 to 8.0, and subjected to gel-based fluorescence detection. The results clearly showed two lower bands (around 75 kDa) prominently presented at pH from 3.5 to 6.5, and a higher band (around 100 kDa) seen at pH from 5.5 to 7.5 (**Fig. 4.5A**). These two lower bands had acidic pH optimum and molecular weights characteristic of the two mature lysosomal GAA forms (70 and 76 kDa), while the higher band has a neutral pH optimum despite havening similar molecular weight of pro-GAA (110 kDa). The amounts of pro-form of GAA should be low in cells, and it is—as well as the mature forms—active at acidic pH. Hence, this higher running band is likely to be another target, likely the α-subunit of the ER-residing α-glucosidase II (GANAB), which possesses a similar molecular weight and a more neutral pH optimum. To identify these ABP targets, chemical proteomic method was employed. In this method, the biotin-containing ABP **7** was incubated with fibroblasts lysates at 5 μM concentration at either pH 4.0 or 7.0, with or without pre-incubation of ABP **4** (5 μM , as a negative control). Samples were subsequently denatured, affinity-enriched for biotin-containing molecules using streptavidin beads, subjected to either on-bead tryptic digestion or in-gel tryptic digestion, and peptides were analyzed by nano-LC-MS-based protein identification (**Fig. 4.5B**). For the in-gel digestion protocol, biotin-enriched samples were subjected to SDS-PAGE and silver stained. The resulting bands were excised out and subjected to in-gel tryptic digestion and proteomic identification. The silver stain results (**Fig. 4.5C**) showed similar band patterns with those by the labeling of Cy5 ABP **6** (**Fig. 4.5A**): two prominent bands at around 75 kDa were observed in the sample labeled with the biotin ABP **7** at pH 4.0, and one 100 kDa band were detected in the sample labeled with the biotin

ABP **7** at pH 7.0 (**Fig. 4.5C**). These band were absent from samples not treated with ABP **7**, and from samples pre-incubated with the green BODIPY ABP **4** (**Fig. 4.5C**), suggesting they were truly protein targets that were labeled with the biotin-containing ABP **7** and affinity-enriched. Subsequent proteomic identification revealed that GAA was the



Band	Rank	Accession	Protein name	Score	Matches	Coverage(%)	emPAI
2a	47	P10253	Lysosomal alpha-glucosidase, GAA	39	3	2.4	0.03
3a	6	P10253	Lysosomal alpha-glucosidase, GAA	239	17	8.1	0.14
2b	11	P10253	Lysosomal alpha-glucosidase, GAA	184	20	9.0	0.14
3b	11	P10253	Lysosomal alpha-glucosidase, GAA	545	47	16.9	0.46
2c	25	P10253	Lysosomal alpha-glucosidase, GAA	68	17	8.4	0.08
3c	2	P10253	Lysosomal alpha-glucosidase, GAA	506	50	21.1	0.58
	7	P04062	Glucosylceramidase, GBA	274	25	25.4	0.6
5d	5	Q14697	Neutral alpha-glucosidase AB, GANAB	479	35	19.5	0.42
6d	1	Q14697	Neutral alpha-glucosidase AB, GANAB	1642	94	43.1	1.83
	2	Q14697-2	Isoform 2 of Neutral alpha-glucosidase AB, GANAB-2	1588	92	42.7	1.69
1a-c, 4d	No target enzymes						

Figure 4.5. Identification of ABP targets in human fibroblast lysates by chemical proteomics. A) Gel-based fluorescent labeling of ABP **6** in lysates of normal human dermal fibroblasts (NHDF) at different pHs. B) Schematic representation of the chemical proteomic workflow. C) Silver stain of gels containing samples enriched for biotin-containing molecules (left), and the resulting list of identified glycosidases at the indicated positions (lane 1–6, band a–d). D) Comparison of relative abundance of glycosidases by Mascot search engine peptide score (left) and exponentially modified Protein Abundance Index (emPAI) (right).

prominent glycosidase in position a–c (around 75–50 kDa) in the gel containing samples labeled at pH 4.0, while GANAB was exclusively identified in position d (around 100 kDa) in the gel containing samples labeled at pH 7.0 (**Fig. 4.5C**, right). Interestingly, the lysosomal glucocerebrosidase (GBA) was also identified as a target at position c of the pH 4.0 gel (**Fig. 4.5C**). To verify the activity of compound **1–7** towards GBA, compounds were subjected to enzymatic assay with the therapeutic recombinant human GBA (rGBA, Imiglucerase, Cerezyme®). The results demonstrated that, besides the epoxide compound **2** and the aziridine compound **3**, all other N-alkylated compounds inhibited GBA at high nanomolar to low micromolar range. They were, however, less reactive towards rGBA compared to rGAA (**Table. 4.S1**).

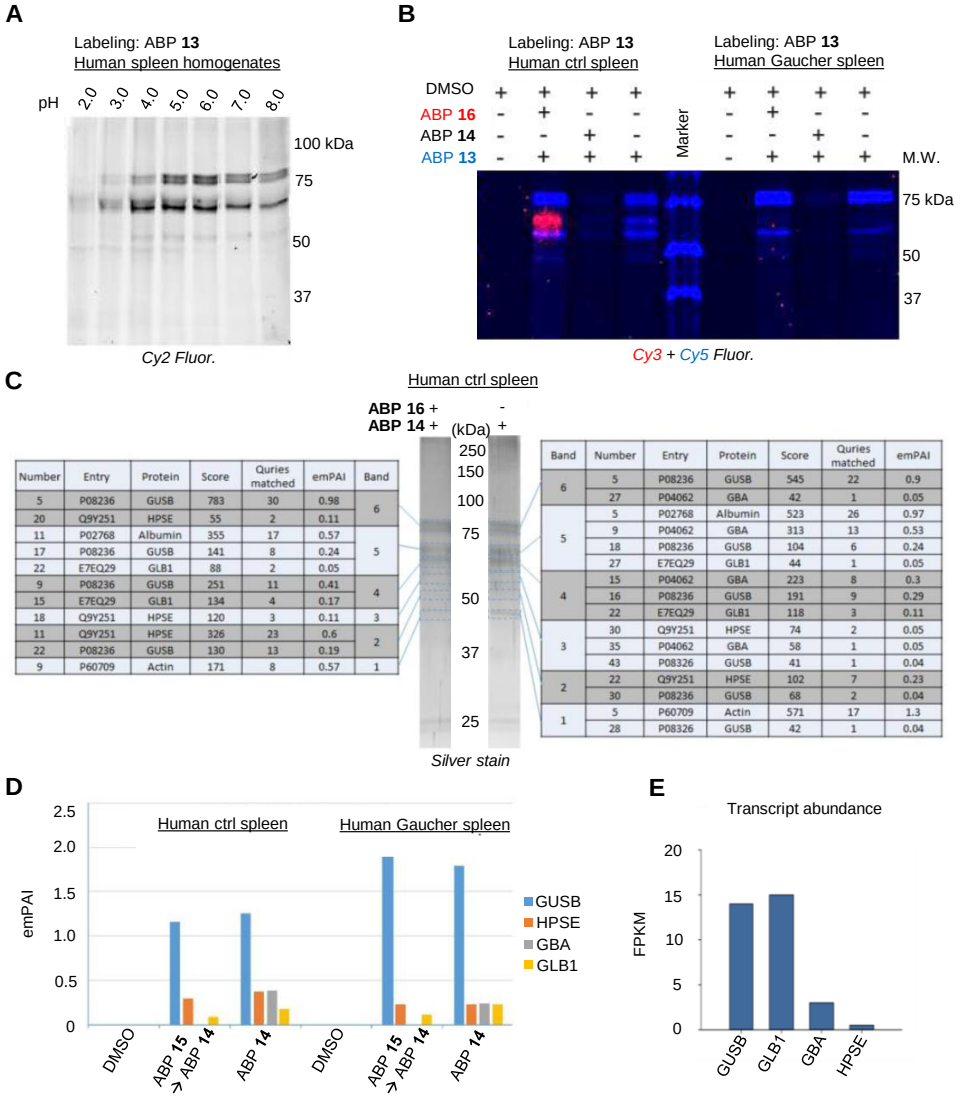
In a parallel experiment, the affinity enriched samples were trypsin-digested in the solution (i.e. on-bead digest), and the resulting peptides were comparably analyzed by nano-LC-MS. The identified glycosidases in each sample were compared for relative abundance using the Mascot search score (**Fig. 4.5D**, left) and the exponentially modified Protein Abundance Index (emPAI) (**Fig. 4.5D**, right), which is based on comparing the coverage of matched peptides for each protein against the search database. The results were in excellent agreement with the previous finding, in which GAA and GBA were the only two ABP glycosidase targets identified at pH 4.0, and that GANAB (and its isoform 2) were additionally identified at pH 7.0.

The chemical proteomic protocol was also applied to tissue extracts from mouse. Labeling by ABP **7** in mouse liver homogenates also enabled identification of GAA and GBA at pH 4.0, and GANAB at pH 7.0 (**Fig. 4.S4**), while labeling by ABP **7** in mouse intestine homogenates identified three additional enzymes: maltase-glucoamylase (MGAM, predominantly at pH 4.0), sucrase-isomaltase (SI, at both pH 4.0 and 7.0), and lactase-phlorizin hydrolase (LPH, or protein Lct, at pH 7.0) (**Fig. 4.S5**). Thus, compound **7** labeled all four mammalian GH31 α-glucosidases, while also reactive towards the retaining exo-β-glucosidases GBA and LPH.

4.2.5 Glycosidase targets of ABPs **11–14** in complex human proteome

Identification of targets by the β-glucuronic acid configured ABPs was performed in

homogenates of human spleen, a material which is predicted to express high GUSB level,



Number	Entry	Protein	Score	Quries matched	emPAI	Band
5	P08236	GUSB	783	30	0.98	6
20	Q9Y251	HPSE	55	2	0.11	
11	P02768	Albumin	355	17	0.57	5
17	P08236	GUSB	141	8	0.24	
22	E7EQ29	GLB1	88	2	0.05	
9	P08236	GUSB	251	11	0.41	4
15	E7EQ29	GLB1	134	4	0.17	
18	Q9Y251	HPSE	120	3	0.11	3
11	Q9Y251	HPSE	326	23	0.6	2
22	P08236	GUSB	130	13	0.19	
9	P60709	Actin	171	8	0.57	1

Band	Number	Entry	Protein	Score	Quries matched	emPAI
6	5	P08236	GUSB	545	22	0.9
	27	P04062	GBA	42	1	0.05
5	5	P02768	Albumin	523	26	0.97
	9	P04062	GBA	313	13	0.53
	18	P08236	GUSB	104	6	0.24
	27	E7EQ29	GLB1	44	1	0.05
4	15	P04062	GBA	223	8	0.3
	16	P08236	GUSB	191	9	0.29
	22	E7EQ29	GLB1	118	3	0.11
3	30	Q9Y251	HPSE	74	2	0.05
	35	P04062	GBA	58	1	0.05
	43	P08326	GUSB	41	1	0.04
2	22	Q9Y251	HPSE	102	7	0.23
	30	P08236	GUSB	68	2	0.04
1	5	P60709	Actin	571	17	1.3
	28	P08326	GUSB	42	1	0.04

Figure 4.6. Identification of ABP targets in human spleen homogenates. A) Gel-based fluorescent labeling of ABP **6** at different pHs. B) Competitive ABPP of GBA-specific ABP **14** or biotin ABP **16** with the Cy5 ABP **13** in control (left) or Gaucher (right) spleen homogenates. C) List of glycosidases identified at each gel position by chemical proteomics. Albumin and actin were included based on their absence in the control samples. D) emPAI values of identified glycosidases. E) Transcriptomic abundance of the identified glycosidases in human spleen, based on data from FANTOM5 consortium experiment E-MTAB-3358. FPKM, Fragments sequenced per kilobase transcript per million reads.

according to the Expression Atlas transcriptome data base.[24] Gel-based fluorescence detection

in samples treated with 100 nM Cy5 ABP **13** at various pH showed four prominent bands between around 80–60 kDa, all of which have labeling optimum at around pH 5.0 (**Fig. 4.6A**). Human GUSB is processed from the 78 kDa pro-form into a 60 kDa and a 18 kDa fragment in the lysosomes[25, 26], while two other isoforms of 70 (isoform 2) and 58 kDa (isoform 3) were known. It is likely that three of the bands were GUSB isoforms, and one was another protein. Thus, the two higher bands were likely the pro-form of GUSB and the isoform 2, and the lowest band could be a mixture of mature GUSB and isoform 3. The identity of the second-lowest band was hypothesized as the mature form of GUSB, or GBA, which has a similar molecular weight as appeared on SDS-PAGE.[21] To verify the identity of this band, gel-based competitive ABPP was employed, in which the biotin ABP **14** or the GBA-specific red BODIPY ABP **16**[21] (**Fig. 4.2**) were pre-incubated with the sample at pH 5.0, prior to the addition of the Cy5 ABP **13**. Indeed, in sample pre-treated with ABP **16**, a red band was observed, in place of the second-lowest band labeled by ABP **13**, suggesting that the second-lowest band was GBA (**Fig. 4.6B**, left). In support of this, ABP **13** labeling was also performed in homogenates of human Gaucher spleen that is deficient in GBA, where this band was absent from samples labeled with either ABP **13** or ABP **16** (**Fig. 4.6B**, right). In samples pre-treated with ABP **14,** all labeling were abrogated, which indicated that both ABP **13** and **14** could label at least GBA, and possibly on GUSB.

Similar to the target identification for the α-glucose configured cyclophellitol aziridine compound **7**, chemical proteomics were applied to further identify the protein targets of the biotin ABP **14**. Human spleen homogenates were pre-incubated with or without the GBA-specific ABP **16**, and labeled with ABP **14** at 10 μM concentration. Upon denaturation and affinity-purification of biotinylated molecules, samples were again divided for in-gel tryptic digestion and on-bead tryptic digestion. As expected, in-gel digestion protocol resulted in the identification of GUSB and GBA at the indicated positions (**Fig. 4.6C**) in samples treated with ABP **14**, while pre-incubation of ABP **16** abrogated the identification of GBA by both in-gel (**Fig. 4.6C**, left) and on-bead digestion protocol (**Fig. 4.6D**). Also identified were the human GH79 retaining endo-β-glucuronidase heparanase (HPSE, at around 50 kDa) and the lysosomal

acid β-galactosidase (GLB1, at around 60 kDa). However, the abundance of GBA, HPSE, and GLB1 were all much lower compared to the level of GUSB, based on emPAI value (**Fig. 4.6D**). Analysis of transcript abundance of the identified proteins in human spleen (based on data from the FANTOM5 consortium experiment E-MTAB-3358) showed that GUSB and GLB1 transcripts are much more abundant compared to the GBA and HPSE transcripts. In combination, these results suggested that GUSB is the major target by ABP **14**, and that GLB1 is a minor target. GBA and HPSE could also be major targets of ABP **14**, given sufficient expression levels in the labeled tissue.

4.2.5 GAA and GUSB labeling in intact cells

The applicability of the ABPs to inhibit and label their target glycosidases in intact cells were next investigated. For inhibition, confluent human fibroblasts were treated with the ABPs in a range of different concentrations for 2 hours, washed and lysed, and the lysates were measured for the relative loss of GAA or GUSB activity. For most of the α-glucose-configured compounds, it was observed that their *in situ* apparent IC_{50} values towards the endogenous GAA in intact cells were similar to their *in vitro* apparent IC_{50} values towards rGAA (mid- to high nanomolar) (**Fig. 4.7A**, left). The epoxide compound **1** and the biotin ABP **7** were the only two exceptions, showing no inhibition at the highest applied concentration (50 or 10 μM). This was in contrast to the β-glucuronic acid configured compounds, in which only the red BODIPY ABP **12** and the Cy5 ABP **13** showed moderate activity towards GUSB (apparent *in situ* IC_{50} value = 1.7 and 1.8 μM, respectively), and other compounds did not show inhibition up to the highest applied concentration (15 μM) (**Fig. 4.7A**, right). This suggested that installation of the biotin moiety on the α-glucose-configured cyclophellitol aziridine, as well as the presence of carboxylic acid at the C8 position of the cyclophellitol aziridine, are detrimental for *in situ* inhibition of the two lysosomal glycosidases.

To visualize the *in situ* labeling, the α-glucose-configured Cy5 ABP **6** was treated to cells at 100 nM concentration for various time periods (10 minutes to 4 hours), and the fluorescence in cell lysates was visualized by gel-based ABPP. Another set of cells were identically treated, but was lysed in lysis buffer containing an excess of green BODIPY ABP **4** (10 μM), to exclude the

possible labeling of GAA by ABP **6** during cell lysis. The gel-based ABPP analysis showed three prominent bands corresponding to the molecular weight of GANAB (~100 kDa) and GAA (76 and 70 kDa), as identified previously, and that the labeling intensity increased with incubation time (**Fig. 4.7B**, left). Addition of the green BODIPY ABP **4** resulted in a band with overlapping molecular weight as the putative GANAB, and another lower band at around 55 kDa that was not labeled by the Cy5 ABP **6**—likely to be GBA according to its appeared molecular weight.[21] Because the addition of ABP **4** in the lysis buffer did not change the labeling pattern of the *in situ* treated Cy5 ABP **6**, it can be concluded that ABP **6** labeling on GANAB and GAA was most likely taken place in intact cells. The higher band labeled by ABP **4** in the lysis buffer was also

A

Compounds	IC_{50} (nM) GAA in NHDF	Compounds	IC_{50} (nM) GUSB in NHDF
1 (JJB307)	> 50,000	8 (JJB400)	> 15,000
2 (CF21)	71.2 ± 9.0	9 DW222	> 15,000
3 (CF22)	75.8 ± 10.9	10 (JJB355)	> 15,000
4 (JJB382)	502 ± 4.1	11 (JJB395)	> 15,000
5 (JJB347)	562 ± 352	12 (JJB391)	1,700 ± 600
6 (JJB383)	42.2 ± 4.9	13 (JJB392)	1,800 ± 400
7 (JJB384)	> 10,000	14 (JJB397)	> 15,000

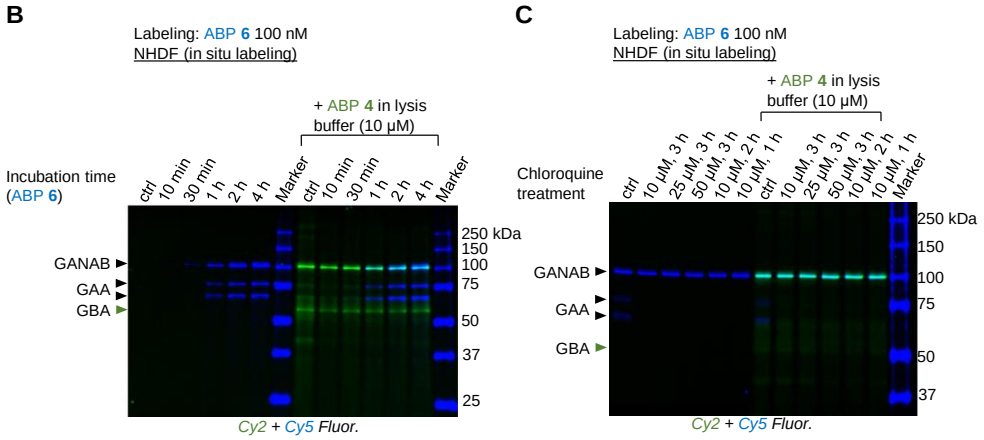


Figure 4.7. Glycosidase inhibition and labeling in intact cells by compounds 1–14. A) *In situ* IC_{50} values towards GAA by compounds **1**–**7** (left) or GUSB by compounds **8**–**14** (right). B) *In situ* labeling by ABP **6**, with or without an excess of ABP 4 in the lysis buffer. C) Effect of chloroquine pre-incubation on

ABP **6** labeling in cells.

putatively assigned to GANAB, as the *in situ* treated ABP **6** at 100 nM seemed not to reach saturate labeling of that protein at 4 h incubation time. This is in contrast to the labeling on the GAA bands, which were saturated at around 2 h incubation. In another experimental setup, chloroquine was added at various dosage and incubation time to the cells, prior to ABP **6** incubation and the subsequent lysis with or without an excess of ABP **4**. Chloroquine is a chemical agent that is known to raise lysosomal pH and thereby disrupt lysosomal functions[27], and its presence could inhibit GAA activity in the lysosome and therefore abrogate ABP labeling. As predicted, it was clearly observed that the labeling by ABP **6** on the putative GAA bands was abrogated in the presence of chloroquine, while that the labeling on GANAB was not affected (**Fig 4.7C**). This result confirmed the ability of ABP **6** to inhibit and label the lysosomal GAA in intact cells in an activity-based manner, and demonstrated the applicability of this ABP in monitoring GAA activity in cells.

4.2.6 Visualizing GAA deficiency in patient fibroblasts

As ABP **6** was shown to allow specific detection of active GAA in gel-based assays, it's applicability for GAA labeling in a diagnostic setup was examined. Fibroblasts from healthy donors or Pompe disease patients were cultured, and cell lysates were labeled with 1 μM ABP **6** for 30 minutes, at either pH 4.0 or 7.0. The results showed the marked loss of GAA labeling by ABP **6** at pH 4.0 in samples of Pompe patients (**Fig. 4.8A**), but not on those of GANAB at pH 7.0 (**Fig. 4.8B**). This loss of active GAA in patient samples correlated with the loss of GAA proteins, as determined by the subsequent Western blot detection of GAA from the same gel (**Fig. 4.8A** and **B**, lower panels). This result therefore demonstrated the potential application of ABP **6** in laboratory diagnosis of Pompe disease.

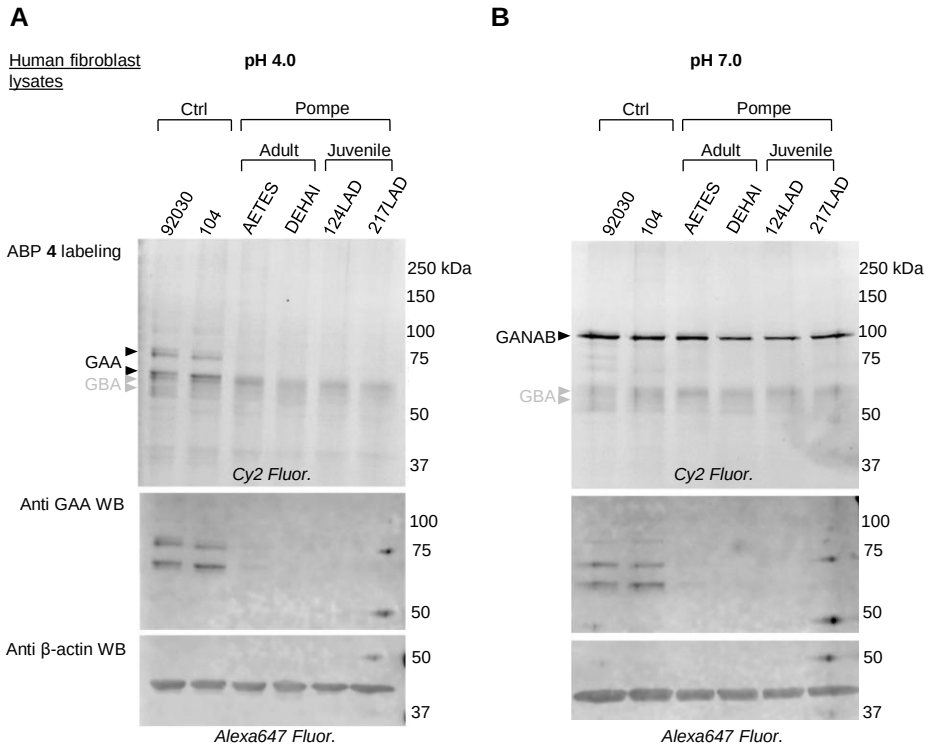


Figure 4.8. ABP 6 labeling in fibroblasts lysates of control or Pompe patients. A) Labeling at pH 4.0. B) Labeling at pH 7.0. The same gels were subjected to subsequent Western blot detection of GAA and β-actin (protein loading control).

4.3 Discussion

Activity-based protein profiling has emerged in the past decade allowing unprecedented studies on active enzymes such as quantitative visualization and enrichment. Glycosidases possessing the Koshland double-displacement mechanism have been shown to be applicable for activity-based protein profiling through the appropriate designing of small molecules grafted with an electrophilic trap and exhibit sufficient substrate mimicry. Functionalization and tuning the configuration of cyclophellitol epoxide and cyclophellitol aziridine was found to offer a viable strategy for generating ABPs against a number of retaining glycosidases.[20, 21, 28–30] In this chapter, the α-glucose- and β-glucuronic acid-configured cyclophellitol and (*N*-tagged) cyclophellitol aziridines are examined for their biological activities, including mechanism of

action in recombinant enzymes and selectivity in complex proteomes and in intact cells.

Both sets of compounds inhibit recombinant enzymes in a pH- and time-dependent manner, and generally with high (nanomolar) potency. Protein crystallographic studies reveals that they react with the nucleophile of their respective target enzyme, and adopting identical conformation as the true substrates. Targets of these ABPs have been identified qualitatively by gel-based fluorescent detection, and quantitatively by chemical proteomics. For α-glucose-configured cyclophellitol aziridine ABPs in human and mouse samples these encompass four GH31 retaining exo-α-glucosidases, including GAA, GANAB, and the intestinal dietary enzymes MGAM and SI. Intriguingly, the retaining β-glucosidases GBA and LPH are also identified as targets by chemical proteomics. As these two enzymes exhibit no α-glucosidase activity towards artificial 4-MU substrates, the labeling mechanism of ABP **14** towards the retaining β-glucosidases remains to be explored in the future. The β-glucuronic acid-configured cyclophellitol aziridine ABP labels the intended target GUSB with exceptionally high potency. It nevertheless shows additional minor targets in GBA, HPSE, and GLB1. Labeling of the GUSB ABPs of HPSE, as well as pro-HPSE, has been further demonstrated by chemical proteomics and protein crystallography in the published article based on the results from this chapter.[31] Because GUSB and HPSE have distinct molecular weights, the β-glucuronic acid configured cyclophellitol aziridine ABPs should also be viable tools to specifically detect HPSE activity in complex biological samples using an SDS-PAGE setup. Altogether, the *in vitro* labeling results suggest that tuning the glycon configuration of the cyclophellitol aziridine ABP is a viable strategy to target the intended glycosidase class, but labeling of other out-of-class glycosidases can in some cases still occur. Nevertheless, by identifying the off-target glycosidases, specific visualization of target glycosidases, such as GAA or GUSB, can be achieved by varying the labeling pH, ABP concentration, and/or pre-incubation of known ABPs/inhibitors for the off-target glycosidases, as demonstrated in this chapter.

It is shown that the α-glucose-configured cyclophellitol aziridine compounds are cell-permeable, and that the Cy5 ABP **6** at 100 nM specifically labeled GAA over GBA in cells. The β-glucuronic acid configured cyclophellitol aziridines are less effective in intact cells, as they

exhibit at least 1000-fold reduction in apparent potency in cells when compared to *in vitro* assay. Nevertheless, the red BODIPY ABP **12** and Cy5 ABP **13** should be applicable for *in situ* GUSB labeling, given enough incubation time and prior incubation with a GBA inhibitor or ABP. In the future, the ability of the ABPs to label active lysosomal enzymes in lysates/homogenates and in intact cells should assist important studies such as those on processing and distribution of therapeutic enzymes, and the efficacy of other LSD therapeutic approaches that aim to promote the lysosomal activity of deficient enzymes at target tissues/organs like gene therapy and chaperone therapy.

Finally, in this chapter the applicability of the GAA ABP in the laboratory diagnosis of Pompe disease is demonstrated. Common biochemical diagnosis methods for Pompe disease rely on GAA ativity measurement using the 4-MU-α-glc substrate assay from lysates of cultured fibroblasts, muscle biopsy, urine, or blood samples.[32] The biggest drawbacks for the substrate assay is the low detection dynamic range (typically in the order of two)[33], due to background activity from 4-MU, and the presence of MGAM in blood samples that requires the addition of acarbose to relatively block MGAM activity in the assay[34]. Mass spectrometry-based activity assays have improved the detection sensitivity[33, 35], but suffer from the associated laborious procedures. Therefore, it is envisioned that the gel-based ABPP described in this chapter could circumvent these shortcomings, offering rapid and sensitive visualization of active GAA on a simple and rapid SDS-PAGE-based assay. Other possible GAA activity readouts using ABPs are fluorescent microscopy (Chapter 1, this thesis) and fluorescence-assisted cell sorting (FACS).[21] However, in both cases, the concomitant labeling on GANAB in intact cells would introduce undesirable background fluorescence, which could complicate data interpretation. This issue can be overcome by pre-incubation with a GANAB-selective or GAA-selective active-site blocker. Although no such inhibitors are currently known, using the ABPs in a gel-based competitive ABPP assay (cABPP) or in a fluorescent polarization assay[36] would assist identification of such inhibitors. Studies are currently pursued in this direction (Daniel Lahav, Department of Bio-organic Synthesis, Leiden University).

In conclusion, true mechanism-based inhibitors and ABPs for GAA and GUSB have been

successfully developed and extensively characterized. They enable specific detection of their target enzymes in complex biological samples. The applicability of the ABP in GAA detection has been demonstrated in intact cells, and its value for diagnosis of Pompe disease has been demonstrated. The novel ABPs should find future applications in the study of lysosomal glycosidases in both fundamental and applied research in the context of Pompe disease, Sly Syndrome, and even pathologies involving heparanase activity.

4.4 Experimental procedures

4.4.1 Materials

Chemicals were obtained from Sigma-Aldrich (St. Louis, MO, USA), if not otherwise indicated. Trypsin and was commercially available from Promega (Madison, WI, USA). Recombinant GAA was obtained from Sanofi Genzyme (Cambridge, MA, USA). Fibroblasts were obtained with consent from donors. Pompe patients were diagnosed on the basis of reduced GAA activity. Fibroblast cell lines were cultured in HAMF12-DMEM medium (ThermoFisher Invitrogen™, Waltham, MA, USA) supplied with 10% (v/v) FCS. Mouse tissue were isolated according to guidelines approved by the ethical committee of Leiden University (DEC#13191). All the cell or tissue lysates were prepared in potassium phosphate (KPi) lysis buffer (25 mM K_2HPO_4/KH_2PO_4, pH 6.5, supplemented with protease inhibitor cocktail (EDTA-free, Roche, Basel, Switzerland) and 0.1 % (v/v) Triton X-100) via homogenization with silent crusher S equipped with Typ 7 F/S head (30 rpm x 1000, 3 × 7 sec) on ice. Protein concentration in lysates was determined with BCA Protein Assay Kit (ThermoFisher Pierce™). Lysates were stored in small aliquots at –80 °C until use.

4.4.2 Cloning, expression, and purification of bacterial enzymes (University of York)

CjAgd31B expression and purification was carried out as previously described.[37] For AcGH79, the coding sequence of the AcGH79 gene was cloned into the pET28a (Novagen, Madison, WI, USA) expression vector with an N-terminal His6 tag. The E287N mutant was produced by polymerase chain reaction (PCR) using the primer 5'-CCTGACCCAAACGAATTC-3' (forward primer) and 5'-GAATTCGTTTGGGTCAGG-3' (reverse primer). Both the wild-type and mutant proteins were overexpressed in *Escherichia coli* strain BL21 (DE3) GOLD using LB medium. The transformed cells were grown at 37 °C in LB media containing 50 μg mL^{-1} kanamycin until the A600 nm reached 0.8. Expression of the recombinant proteins was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 12 h at 25 °C. The cells were harvested by centrifugation at 8000 g for 30 min and resuspended in 50 mL HEPES lysis buffer (20 mM HEPES, NaCl 200 mM, imidazole 5 mM, pH 7.0). After 20 min of

sonication and 30 min of centrifugation at 12000 g, the filtered supernatant containing His6-AcGH79 was loaded onto a His Trap column (GE Healthcare, Chicago, IL, USA), equilibrated with the lysis buffer. The column was washed with HEPES lysis buffer and the His6-AcGH79 protein was eluted with the same buffer with supplement of 400 mM imidazole over a gradient of 100 mL. The fractions containing the His6-AcGH79 were then loaded onto a Hiload 16/60 Superdex 75 column (GE Healthcare). The fractions containing the His6-AcGH79 were pooled and concentrated to the final concentration of 14.5 mg mL^{-1}.

4.4.3 Enzyme activity assays and *in vitro* IC_{50} measurements

The α-D-glucosidase activity of lysosomal α-D-glucosidase GAA was assayed in individual wells of medium-binding flat-bottomed black 96-well plates (Greiner, Kremsmünster, Austria) at 37 °C by incubating samples with 3.0 mM 4- methylumbelliferyl-α-D-glucopyranoside (4-MU-α-glc) as substrate in McIlvaine buffer (150 mM sodium citrate-Na_2HPO_4, supplemented with 0.01 % (w/v) NaN_3 as bacteriostatic)[38], supplemented with 0.1 % (w/v) BSA, at pH 5.0 for rGAA and at pH 4.0 for GAA in cell lysates or tissue homogenates. Activity of rGBA was measured using similar conditions but with 3.75 mM 4-methylumbelliferyl-β-D-glucopyranoside (4-MU-β-glc) as substrate at pH 5.2, supplemented with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate. Activity of AcGH79 and GUSB was performed with 2.5 mM 4-methylumbelliferyl-β-D-glucuronic acid (4-MU-β-glc) as substrate at pH 5.0, supplemented with 0.1 % (w/v) BSA. To determine the apparent *in vitro* IC_{50} value, recombinant enzymes (12.5 µL) were firstly pre-incubated with inhibitor dilutions (12.5 µL) for 30 min at 37 °C, prior to incubation with substrates (100 µL) for a further 30 min at 37 °C. The enzymatic reaction was quenched by adding 200 µL of Glycine-NaOH (1 M, pH 10.3), after which fluorescence of liberated 4-methylumbelliferyl was measured with a fluorimeter LS55 (Perkin Elmer, Waltham, MA, USA) at λ_{EX} = 366 nm and λ_{EM} = 445 nm. For experiments with varying pH, enzymes, inhibitors, and substrates were prepared in McIlvaine buffer at the indicated pH. All apparent IC_{50} values were determined from biological duplicates. Data was corrected for background fluorescence, then normalized to the untreated control condition and finally curve-fitted via one phase exponential decay function using Prism 5.0 (GraphPad Software, San Diego, CA, USA).

4.4.4 *In vitro* ABP labeling for recombinant enzymes

All enzymes were pre-incubated in McIlvaine buffer at pH 5.0 (or other indicated pH) on ice for 5 min. *In vitro* labeling on rGAA was performed by incubating 100 fmol rGAA with(out) prior incubation with inhibitors in McIlvaine buffer (pH 5.0 if not otherwise indicated) for 30 min at 37 °C, and ABP **4** labeling at 1 μM probe concentration for 30 min at 37 °C. For *In vitro* labeling on AcGH79, 200 fmol of the enzyme was incubated with(out) inhibitors with the identical conditions as those for rGAA, and ABP incubation was performed with 1 μM ABP **11** for 30 min at 37 °C. Inhibitor concentrations were: 10 μM for compounds **3**, **5**–**7**, **10**, and **12**–**14**; 100 μM for compound **1**–**2**, **8**, **9**, and **15**; 2.5 mM for maltose, 10 mM for 4-MU substrates, and 2 % (w/v) for sodium dodecyl sulfate (SDS). After ABP incubation, samples were denatured with 5x Laemmli buffer (50 % (v/v) 1.0 M Tris−HCl, pH 6.8, 50% (v/v) 100 % glycerol, 10 % (w/v) DTT, 10 % (w/v) SDS, 0.01 % (w/v) bromophenol blue) by boiling for 5 min at 98 °C. Denatured samples were separated by SDS-PAGE using 7.5 % or 10% polyacrylamide gels, at 90 V for 30 min and 200 V for 50–70 min. Wet slab-gels were scanned for ABP-emitted fluorescence using a Bio-Rad ChemiDoc MP imager (Bio-Rad, Hercules, CA, USA) using Cy2 (λ_{EX} = 470 nm, bandpass 30 nm; λ_{EM} = 530 nm, bandpass 28), Cy3 (λ_{EX} = 530 nm, bandpass 28 nm; λ_{EM} = 605 nm, bandpass 50), and Cy5 (λ_{EX} = 625 nm, bandpass 30 nm; λ_{EM} = 695 nm, bandpass 55) channels. Detected fluorescence was quantified by ImageLab Software (Bio-Rad) and plotted by Prism 5.0 Software (GraphPad). For labeling kinetics, enzymes and ABPs were prepared as earlier described, and incubated at either 4°C or 37 °C for 2 to 60 min. Samples without ABP addition were used as the 0 min controls.

4.4.5 Measurement of inhibition kinetic parameters

Kinetic parameters for inhibition of AcGH79 were determined using a continuous method involving simultaneous exposure of the enzyme to substrates and inhibitor (see **Supporting Note 4.S2**; ref. [39]), using fluorogenic substrate assay. During the enzymatic reaction, the concentration of enzyme (AcGH79) and substrates were 0.26 nM and 2 mM, respectively. For each inhibitor, triplicate sets of eight 2 mL Eppendorf tubes were prepared. These were added

firstly with 8.13 µL inhibitor at various concentrations (diluted in DMSO, 200x of reaction concentration, control = DMSO) and secondly with 154.4 µL McIlvaine buffer (150 mM, pH 5.0) and 1300 µL substrate mixture (2.5 mM 4-MU-β-GlcA, pH 5.0, 0.1 % (w/v) BSA). The reaction tubes were pre-warmed on a thermoshaker at 37 °C for 10 min. For each inhibitor, the enzyme (AcGH79) was prepared by diluting with McIlvaine buffer in an Eppendorf tube to a volume of 1250 µL and 2.6 nM enzyme concentration. A negative control (enzyme blank) was prepared in a separate Eppendorf tube by substituting the enzyme with denatured enzyme (boiled at 98 ºC for 5 min). Both were pre-warmed on a thermoshaker at 37 °C for 10 min. Prior to the reaction, a black 96-well plate was loaded with 200 µL stop buffer (1 M glycine-NaOH, pH 10.3) in each well. The t = 0 samples were prepared by filling the first two columns of the plate with 12.5 µL enzyme (or enzyme blank, in the last row) and 112.6 µL of each from the substrate-inhibitor mixture (1 concentration per row, in duplicate wells). To start the reaction, 137.6 µL aliquots of the pre-warmed 26 nM enzyme was transferred to each but the last (containing DMSO dilution and substrate mixture) of the pre-warmed 2 mL tube containing the substrate-inhibitor mixture, with a time interval of 15 seconds between samples. 137.6 µL aliquots from the pre-warmed enzyme blank were similarly transferred to the last 2 mL tube. To ensure proper mixing of the components, during the addition of enzyme aliquots the 2 mL tubes were continuously incubated in the thermal shaker at 37 °C under constant shaking at 850 rpm. At t = 5, 10, 15, 30 and 60 min, duplicates of 125 µL aliquots from each of the reaction tubes were transferred to the 96-well plate that contained stop buffer, with a 15-second interval between each tube. The plate was vortexed shortly and 4-MU fluorescence was measured immediately. The observed pseudo-first order inactivation rate (k_{obs}) was calculated for each concentration of each inhibitor by fitting the data with the one-phase exponential association function: 4-MU fluorescence = A*(1–e^(−k_{obs}*t)) using GraphPad Prism. The obtained k_{obs} values for each concentration of each inhibitor were then plotted against the inhibitor concentration, and the resulting plots were fitted using a linear function, which gives the combined apparent inhibition parameter k_{inact}/K'_I as the slope. k_{inact}/K_I was derived from k_{inact}/K'_I by correcting for the presence of competing 4-MU-β-GlcA substrate, using the relationship $K'_I = K_I (1+[S]/K_M)$, where [S] = 2 mM and K_M = 18.2 µM (**Fig. 4.S3**). The k_{inact}/K_I

values from three independent sets measured for each inhibitor were used to calculate the average and SD of the k_{inact}/K_I value of each inhibitor.

4.4.6 Protein crystallography (University of York)

Protein crystals of CjAgd31B were obtained using 1.8 M ammonium sulfate, 0.10 M HEPES (pH = 7.0), 2% PEG 400 at 20 °C by the sitting drop vapor diffusion method. Crystal complexes with compound **3** was obtained by soaking in mother liquor containing 5.0 mM compound for 2 h, before cryoprotecting in 2.0 M lithium sulfate, 0.10 M HEPES (pH = 7.0), 2 % (w/v) PEG 400, and flash freezing in liquid N_2 for data collection. AcGH79 was tested against a range of commercial crystallization screens. Well diffracting crystals of wild-type AcGH79 were obtained by the sitting-drop vapor-diffusion method at 20 °C using 0.8–1.2 M of 0.5:9.5 NaH_2PO_4:K_2HPO_4 (v/v) at a protein:well ratio of 700:500 nL. Crystals of AcGH79 E287Q mutant were obtained using 1.2–1.5 M of 1.0:9.0 NaH_2PO_4:K_2HPO_4 (v/v) at a protein:well ratio of 500:500 nL. Crystals typically appeared after 1 day. Protein complexes with compound **10** were prepared by supplying the drops containing crystals with 1.0 µL of 5.0 mM compound, in the precipitant solution freshly prepared, before soaking procedure for 1 h at 20 °C before fishing. Crystals of AcGH79 were cryoprotected using 2 M lithium sulfate before flash freezing in liquid N_2 for data collection. All data were collected at 100 K on beamline I04 of the Diamond Light Source UK. Data were processed manually using XDS31[40] and reduced using Aimless[41] (CjAgd31B), or with the xia2 pipeline[42] of the CCP4 software suite (AcGH79). Complex structures were solved by molecular replacement using MolRep,[43] before subsequent rounds of manual model building and refinement using Coot[44] and REFMAC5[45] respectively. Refinements were carried out using TLS determination of molecular motions.[46] Ligand coordinates were built using jLigand.[47] Ribbon and protein surface diagrams were generated using PyMOL. Crystal structure figures were generated using CCP4mg.[48] The data processing statistics and structure refinement are listed in **Table 4.S2**.

4.4.7 *In vitro* ABP labeling in cell lysates and tissue homogenates

All ABP incubation were performed at 37 °C for 30 min. For ABP **6** labeling in human fibroblast lysates, 10 µg total protein from the lysates was incubated with McIlvaine buffer at various pH in 10 µL of volume for 5 min on ice, followed by incubation with 5 µL ABP **6** (diluted with McIlvaine buffer at the matching pH) at a final ABP concentration of 1 µM. For ABP **13** labeling in human spleen homogenates, 20 µg total protein from the samples was incubated firstly with McIlvaine buffer at various pH and secondly with 5 µL ABP **13** (diluted with McIlvaine buffer at the matching pH) at a final ABP concentration of 100 nM, in the same volume and incubation time described for ABP **6**. For competitive ABPP in human spleen homogenates, samples (40 µg total protein) were diluted in 10 µL McIlvaine buffer (pH 5.0), pre-incubated with 2.5 µL ABP **14**, ABP **16,** or DMSO (diluted in McIlvaine buffer) at 100 nM ABP concentration during reaction. These were followed by incubation with 2.5 µL ABP **13** or DMSO (diluted in McIlvaine buffer), at 100 nM ABP **13** during incubation.

4.4.8 Chemical proteomics

For target identification of ABP **7**, human fibroblast lysates (3 mg total protein), mouse liver homogenates (10 mg total protein) or mouse intestine homogenates (6 mg total protein) were incubated at pH 4.0 or 7.0 in McIlvaine buffer, at 37 °C in 0.5 mL volume, with either (1) 0.1 % (v/v) DMSO for 2 h, (2) firstly with 5.0 µM ABP **4** for 1 h, followed by ABP **7** (5 µM for fibroblasts lysates and mouse intestine homogenates; 10 µM for mouse liver homogenates) for 1 h, or (3) ABP **7** (5 µM for fibroblasts lysates and mouse intestine homogenates; 10 µM for mouse liver homogenates) for 1 h. After incubation, samples were denatured by the addition of 125 µL 10 % (w/v) SDS, and boiling for 5 min at 100 °C. From here on, samples were prepared for pull-down with streptavidin beads as published earlier.[49] After pull-down procedure, the samples were divided, two-thirds for on-bead digestion and one-thirds for in-gel digestion. On-bead digestion samples were incubated with 400 µL of trypsin digestion buffer (100 mM Tris-HCl pH 7.8, 100 mM NaCl, 1.0 mM $CaCl_2$, 2% (v/v) acetonitrile (ACN) and 10 ng/µL trypsin) at 37 °C overnight under constant shaking. The supernatant was desalted using stage tips, followed by evaporation of ACN and dilution in 70 µL sample solution (H_2O/ACN/TFA,

95:3:0.1, v:v:v) for LC-MS analysis following previously described procedures.[49] For the in-gel digestion samples, peptides were eluted from the streptavidin resin by boiling at 100 °C with 30 µL of 1 x Laemmli buffer (supplemented with 10 µL biotin). The eluted proteins were separated on 0.75 mm 10 % protein gels by SDS-PAGE at 200 V for 1 h, and silver-stained using the SilverQuest kit (Thermo Fisher). The resulting protein bands were excised and cut into 1 mm^3 cubes with a sterile surgery knife, transferred into 1.5 mL Eppendorf tubes and treated with 30 µL of in-gel digestion buffer (10 mM NH_4HCO_3, 5 % (v/v) ACN, 1 mM $CaCl_2$, 10 ng/µL trypsin). The supernatant containing the trypsin-digested peptides was extracted, and desalted for LC-MS measurement following the published protocol.[49] For LC-MS measurement, all peptide samples were analyzed with a 2 h gradient of 5 % → 25 % ACN on nano-LC, hyphenated to an LTQ-Orbitrap and identified via the Mascot protein search engine, and the Raw data was calculated by MaxQuant program against the Uniprot human or mouse proteome database to present the protein identification list.[49] Mascot identifications were manually validated. The identification results were exported as Excel file including protein accession numbers, Mascot peptide scores, mass of the protein, % coverage of the protein by amino acids identified by LC-MS, peptide matches, miss cleavages, C-terminal peptides and protein emPAI values.[49] For target identification of ABP **14**, 3 mg total protein from human spleen or Gaucher spleen homogenates was incubated with either (1) 0.1% (v/v) DMSO, (2) firstly with 10 µM ABP **16** followed by 10 µM **14**, or (3) 10 µM **14**, each step taking 30 min at 37 °C, in a total volume of 0.5 mL McIlvaine buffer at pH 5.0. Subsequent steps were identical to the method describes for target identification of ABP **7**.

4.4.9 Determination of *in situ* apparent IC_{50} values

The *in situ* IC_{50} value was determined by incubating fibroblast cell lines expressing wild-type GAA and GUSB, grown to confluence, with a range of inhibitor dilutions for 2 h. Hereafter, cells were washed three times with PBS and subsequently harvested by scraping in KPi lysis buffer. *In situ* apparent IC_{50} values for each compound were determined by using enzymatic assay method described earlier, with GAA activity measured at pH 4.0 to avoid possible readout of GANAB activity.

4.4.10 *In situ* ABP labeling

Cultured fibroblasts were cultured in 6-well plates to confluency, and treated with 100 nM ABP **6** diluted in culture medium, at 1 mL medium per well, for 10 to 240 min. For chloroquine treatment experiment, cultured fibroblasts were incubated with chloroquine (10–50 μM, for 1 h up to 3 h) and subsequently with ABP **6** (100 nM, for 3 h) in medium. After washing 3 times with PBS, cells were scraped from culture dishes in the presence of KPi lysis buffer (50 μL per well, with or without 10 μM ABP **4**), collected in Eppendorf tubes, incubated on ice for 30 min for lysis, and stored at –80 °C until use. Lysates containing 5–10 μg total protein were subjected to SDS-PAGE and fluorescence scanning by methods described earlier. Labeling were visualized by fluorescent detection using Bio-Rad ChemiDoc MP Imager under the channels Cy5 for ABP **6** and Cy2 for ABP **4** with the above described settings. For Chloroquine treatment experiment, cultured fibroblasts were incubated with chloroquine (10-50 μM, for up to 3 h) and subsequently with ABP **6** (100 nM, for 3 h) in medium. Cells were harvested with or without 10 mM ABP **4** with the above describe methods. Homogenates were denatured, resolved on SDS-PAGE and detected for ABP **6** labeling by fluorescent scanning.

4.4.11 GAA detection in fibroblast lysates of control and Pompe patients

10 μg protein from each lysate were labeled with 1 μM ABP **4** for 30 min at 37 °C at either pH 4.0 or 7.0, before subsequent SDS-PAGE and fluorescent scanning using a Typhoon FLA9500 Imager (GE Healthcare) using λ_{EX} = 635 nm laser and $\lambda_{EM} \geq$ 665 nm filter, and 100 mm as pixel size. After scanning, proteins on wet slab gels were transferred to PVDF membranes and blocked with 5 % (w/v) BSA in TBST. For GAA detection, the membranes were incubated firstly with mouse polyclonal anti GAA (kind gift from the Kornfeld lab, Washington University of St. Louis, MO, USA) and subsequently with goat anti mouse Alexa647 (Life Technologies, Carlsbad, CA, USA). Blots were scanned on the Typhoon FLA 9500 Imager using 633 nm laser and LPR filter, and 100 mm as pixel size. Rabbit anti β-actin (Cell Signaling, Danvers, MA, USA) and goat anti rabbit Alexa647 (Invitrogen) were used for loading control.

4.5 References

1 Kuo CL (2019) General introduction, this thesis.

2 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2008) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**(Database issue), D233–238.

3 Moreland RJ, Jin X, Zhang XK, Decker RW, Albee KL, Lee KL, Cauthron RD, Brewer K, Edmunds T & Canfield WM (2005) Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor. *J Biol Chem* **280**, 6780–6791.

4 Wisselaar HA, Kroos MA, Hermans MM, van Beeumen J & Reuser AJ (1993) Structural and functional changes of lysosomal acid alpha-glucosidase during intracellular transport and maturation. *J Biol Chem* **268**, 2223–2231.

5 Hers, HG (1963) alpha-Glucosidase deficiency in generalized glycogenstorage disease (Pompe's disease). *Biochem J* **86**, 11−16.

6 Ausems MG, Verbiest J, Hermans MP, Kroos MA, Beemer FA, Wokke JH, Sandkuijl LA, Reuser AJ & van der Ploeg AT (1999) Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling. *Eur J Hum Genet* **7**, 713−716.

7 Reuser AJ, Kroos MA, Hermans MM, Bijvoet AG, Verbeet MP, Van Diggelen OP, Kleijer WJ & Van der Ploeg AT (1995) Glycogenosis type II (acid maltase deficiency). *Muscle Nerve Suppl* **3**, S61−69.

8 Kishnani PS, Hwu WL, Mandel H, Nicolino M, Yong F & Corzo D (2006) A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease. *J Pediatr* **148**, 671−676.

9 Winkel LP, Hagemans ML, van Doorn PA, Loonen MC, Hop WJ, Reuser AJ & van der Ploeg AT (2005) The natural course of non-classic Pompe's disease; a review of 225 published cases. *J Neurol* **252**, 875−884.

10 Umapathysivam K, Hopwood JJ & Meikle PJ (2005) Correlation of acid alpha-glucosidase and glycogen content in skin fibroblasts with age of onset in Pompe disease. Clin Chim Acta **361**, 191–198.

11 Kishnani PS, Corzo D, Nicolino M, Byrne B, Mandel H, Hwu WL, Leslie N, Levine J, Spencer C, McDonald M, Li J, Dumontier J, Halberthal M, Chien YH, Hopkin R, Vijayaraghavan S, Gruskin D, Bartholomew D, van der Ploeg A, Clancy JP, Parini R, Morin G, Beck M, De la Gastine GS, Jokic M, Thurberg B, Richards S, Bali D, Davison M, Worden M A, Chen YT & Wraith JE (2007) Recombinant human acid α-glucosidase: major clinical benefits in infantile-onset Pompe disease. *Neurology* **68**, 99−109.

12 van der Ploeg AT, Clemens PR, Corzo D, Escolar DM, Florence J, Groeneveld GJ, Herson S, Kishnani P S, Laforet P, Lake SL, Lange DJ, Leshner RT, Mayhew JE, Morgan C, Nozaki K, Park DJ, Pestronk A, Rosenbloom B, Skrinar A, van Capelle CI, van der Beek N A, Wasserstein M & Zivkovic S (2010) A randomized study of alglucosidase alfa in late-onset Pompe's disease. *N Engl J Med* **362**, 1396−1406.

13 Sly WS Quinton BA McAlister WH & Rimoin DL (1973) Beta glucuronidase deficiency: report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. *J Pediatr* **82**, 249–257.

14 Tomatsu S, Montaño AM, Dung VC, Grubb JH & Sly WS (2009) Mutations and polymorphisms in GUSB gene in mucopolysaccharidosis VII (Sly Syndrome). *Hum Mutat* **30**, 511–519.

15 Michikawa M, Ichinose H, Momma M, Biely P, Jongkees S, Yoshida M, Kotake T, Tsumuraya Y, Withers SG, Fujimoto Z & Kaneko S (2012) Structural and biochemical characterization of glycoside hydrolase family 79 β-glucuronidase from Acidobacterium capsulatum. *J Biol Chem* **287**, 14069–14077.

16 Rivara S, Milazzo FM & Giannini G (2016) Heparanase: a rainbow pharmacological target associated to multiple pathologies including rare diseases. *Future Med Chem* **8**, 647–680.

17 Vlodavsky I, Beckhove P, Lerner I, Pisano C, Meirovitz A, Ilan N & Elkin M (2012) Significance of heparanase in cancer and inflammation. *Cancer Microenviron* **5**, 115–132.

18 Davies GJ & Williams SJ (2016) Carbohydrate-active enzymes: sequences, shapes, contortions and cells. *Biochem Soc Trans* **44**, 79−87.

19 Jiang J, Kuo CL, Wu L, Franke C, Kallemeijn WW, Florea BI, van Meel E, van der Marel GA, Codée JD, Boot RG, Davies GJ, Overkleeft HS & Aerts JM (2016) Detection of Active Mammalian GH31 α-Glucosidases in Health and Disease Using In-Class, Broad-Spectrum Activity-Based Probes. *ACS Cent Sci* **2**, 351−358.

20 Jiang J, Kallemeijn WW, Wright DW, van den Nieuwendijk AMCH, Rohde VC, Folch EC, van den Elst H, Florea BI, Scheij S, Donker-Koopman WE, Verhoek M, Li N, Schürmann M, Mink D, Boot RG, Codée JDC, van der

Marel GA, Davies GJ, Aerts JMFG & Overkleeft HS (2015) In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 α-l-fucosidases. *Chem Sci* **6**, 2782−2789.

21 Witte MD, Kallemeijn WW, Aten J, Li KY, Strijland A, Donker-Koopman WE, van den Nieuwendijk AM, Bleijlevens B, Kramer G, Florea BI, Hooibrink B, Hollak CE, Ottenhoff R, Boot RG, van der Marel GA, Overkleeft HS, Aerts JM (2010) Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat Chem Biol* **6**, 907-913.

22 Davies GJ, Planas A & Rovira C (2012) Conformational Analyses of the Reaction Coordinate of Glycosidases. *Acc Chem Res* **45**, 308−316.

23 Speciale G, Thompson AJ, Davies GJ & Williams SJ (2014) Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr Opin Struct Biol* **28**, 1−13.

24 Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastings E, Huber W, Jupp S, Keays M, Kryvych N, McMurry J, Marioni JC, Malone J, Megy K, Rustici G, Tang AY, Taubert J, Williams E, Mannion O, Parkinson HE & Brazma A (2014) Expression Atlas update--a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res* **42**(Database issue), D926-32.

25 Brot FE, Bell CE, Jr, & Sly WS (1978) Purification and properties of beta-glucuronidase from human placenta. *Biochemistry* **17**, 385–391.

26 Oshima A, Kyle JW, Miller RD, Hoffmann JW, Powell PP, Grubb JH, Sly WS, Tropak M, Guise KS & Gravel RA (1987) Cloning, sequencing, and expression of cDNA for human beta-glucuronidase. *Proc Natl Acad Sci USA* **84**, 685–689.

27 Poole B & Ohkuma S (1981) Effect of weak bases on the intralysosomal pH in mouse peritoneal macrophages. *J Cell Biol* **90**, 665−669.

28 Kallemeijn WW, Li KY, Witte MD, Marques AR, Aten J, Scheij S, Jiang J, Willems LI, Voorn-Brouwer TM, van Roomen CP, Ottenhoff R, Boot RG, van den Elst H, Walvoort MT, Florea BI, Codée JD, van der Marel GA, Aerts JM & Overkleeft HS (2012) Novel activity-based probes for broad-spectrum profiling of retaining β-exoglucosidases in situ and in vivo. *Angew Chem Int Ed Engl* **51**, 12529−12533.

29 Willems LI, Beenakker TJ, Murray B, Scheij S, Kallemeijn WW, Boot RG, Verhoek M, Donker-Koopman WE, Ferraz MJ, van Rijssel ER, Florea BI, Codée JD, van der Marel GA, Aerts JM & Overkleeft HS (2014) Potent and selective activity-based probes for GH27 human retaining α-galactosidases. *J Am Chem Soc* **136**, 11622−11625.

30 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.

31 Wu L, Jiang J, Jin Y, Kallemeijn WW, Kuo CL, Artola M, Dai W, van Elk C, van Eijk M, van der Marel GA, Codée JDC, Florea BI, Aerts JMFG, Overkleeft HS & Davies GJ (2017) Activity-based probes for functional interrogation of retaining β-glucuronidases. *Nat Chem Biol* **13**, 867−873.

32 Kishnani PS, Steiner RD, Bali D, Berger K Byrne BJ, Case LE, Crowley JF, Downs S, Howell RR, Kravitz RM, Mackey J, Marsden D, Martins AM, Millington DS, Nicolino M, O'Grady G, Patterson MC, Rapoport DM, Slonim A, Spencer CT, Tifft CJ & Watson MS (2006) Pompe disease diagnosis and management guideline. *Genet Med* **8**, 267−88.

33 Kumar AB, Masi S, Ghomashchi F, Chennamaneni NK, Ito M, Scott CR, Turecek F, Gelb MH & Spacil Z (2015) Tandem Mass Spectrometry Has a Larger Analytical Range than Fluorescence Assays of Lysosomal Enzymes: Application to Newborn Screening and Diagnosis of Mucopolysaccharidoses Types II, IVA, and VI. *Clin Chem* **61**, 1363−71.

34 Zhang H, Kallwass H, Young SP, Carr C, Dai J, Kishnani PS, Millington DS, Keutzer J, Chen YT & Bali D (2006) Comparison of maltose and acarbose as inhibitors of maltase-glucoamylase activity in assaying acid alpha-glucosidase activity in dried blood spots for the diagnosis of infantile Pompe disease. *Genet Med* **8**, 302−306.

35 Lin, N, Huang, J, Violante, S, Orsini, JJ, Caggana, M, Hughes, EE, Stevens, C, DiAntonio, L, Chieh Liao, H, Hong, X, Ghomashchi, F, Babu Kumar, A, Zhou, H, Kornreich, R, Wasserstein, M, Gelb, M H & Yu, C (2017) Liquid Chromatography-Tandem Mass Spectrometry Assay of Leukocyte Acid α-Glucosidase for Post-Newborn Screening Evaluation of Pompe Disease. *Clin Chem* **63**, 842−851.

36 Lahav D, Liu B, van den Berg RJBHN, van den Nieuwendijk AMCH, Wennekes T, Ghisaidoobe AT, Breen I, Ferraz MJ, Kuo CL, Wu L, Geurink PP, Ovaa H, van der Marel GA, van der Stelt M, Boot RG, Davies GJ, Aerts JMFG & Overkleeft HS (2017) A Fluorescence Polarization Activity-Based Protein Profiling Assay in the

Discovery of Potent, Selective Inhibitors for Human Nonlysosomal Glucosylceramidase. *J Am Chem Soc* **139**, 14192−14197.

37 Larsbrink J, Izumi A, Hemsworth GR, Davies GJ & Brumer H (2012) *J Biol Chem* **287**, 43288−43299.

38 Kuo CL, van Meel E, Kytidou K, Kallemeijn WW, Witte M, Overkleeft HS, Artola ME & Aerts JM (2018) Activity-Based Probes for Glycosidases: Profiling and Other Applications. *Methods Enzymol* **598**, 217−235.

39 Baici A, Schenker P, Wächter M & Rüedi P (2009) 3-Fluoro-2,4-dioxa-3-phosphadecalins as inhibitors of acetylcholinesterase. A reappraisal of kinetic mechanisms and diagnostic methods. *Chem Biodivers* **6**, 261–282.

40 Kabsch W (2010) XDS. *Acta Crystallogr. D Biol Crystallogr* **66**, 125–132.

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

42 Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190.

43 Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25.

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

45 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.

46 Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439–450.

47 Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA & Murshudov GN (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68**, 431–40.

48 McNicholas S, Potterton E, Wilson KS & Noble MEM (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**, 386–394.

49 Li N, Kuo CL, Paniagua G, van den Elst H, Verdoes M, Willems LI, van der Linden WA, Ruben M, van Genderen E, Gubbens J, van Wezel GP, Overkleeft HS & Florea BI (2013) Relative quantification of proteasome activity by activity-based protein profiling and LC-MS/MS. *Nat Protoc* **8**, 1155–68.

APPENDIX

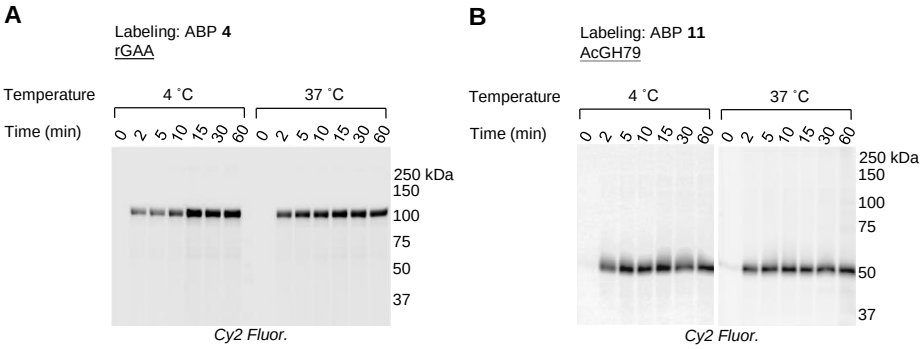


Figure 4.S1. Labeling kinetics of ABPs towards recombinant enzymes at 4 °C or at 37 °C. A) ABP **4** labeling towards rGAA. B) ABP **11** labeling towards AcGH79.

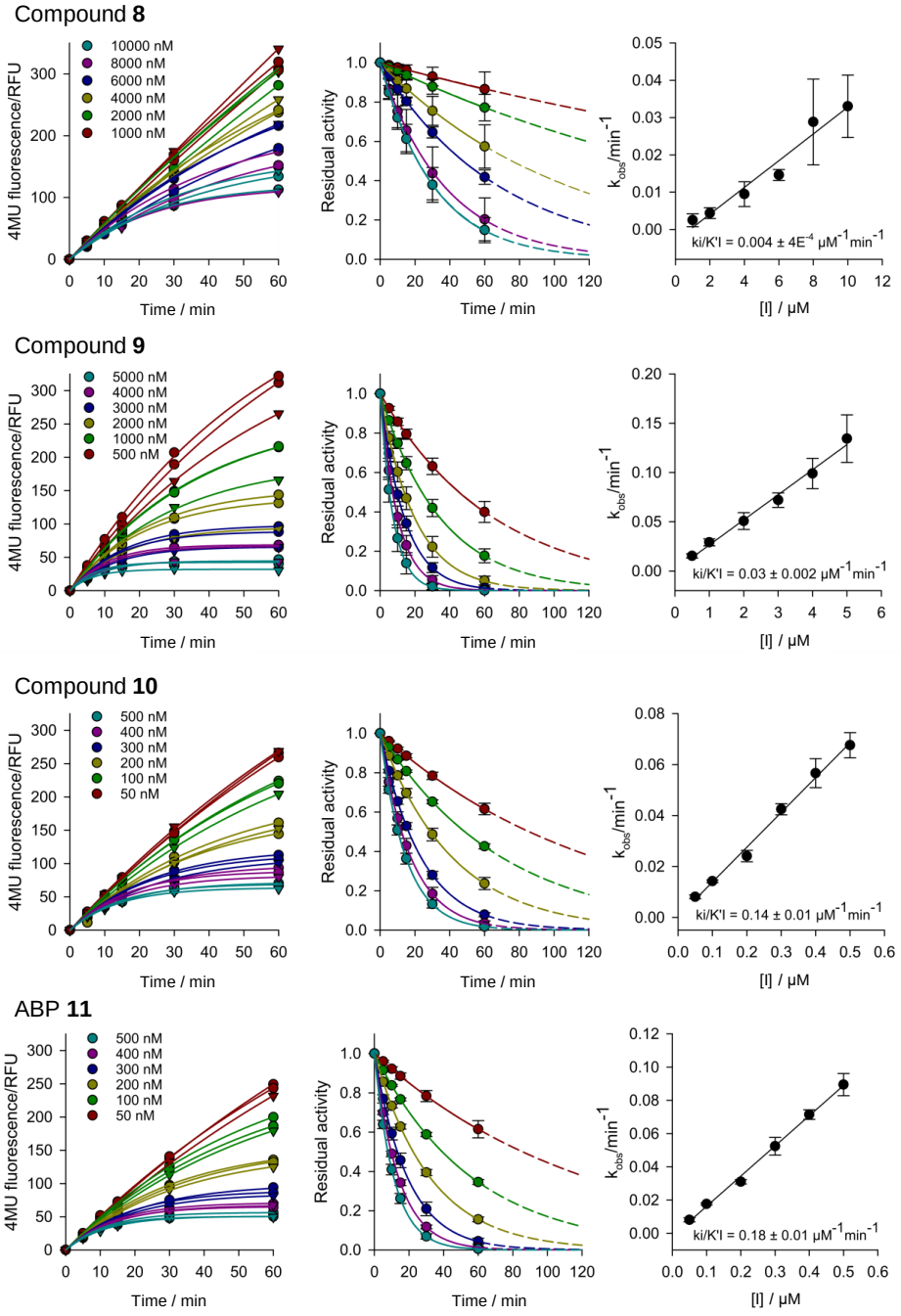


Figure 4.S2 (continued, 1 of 2).

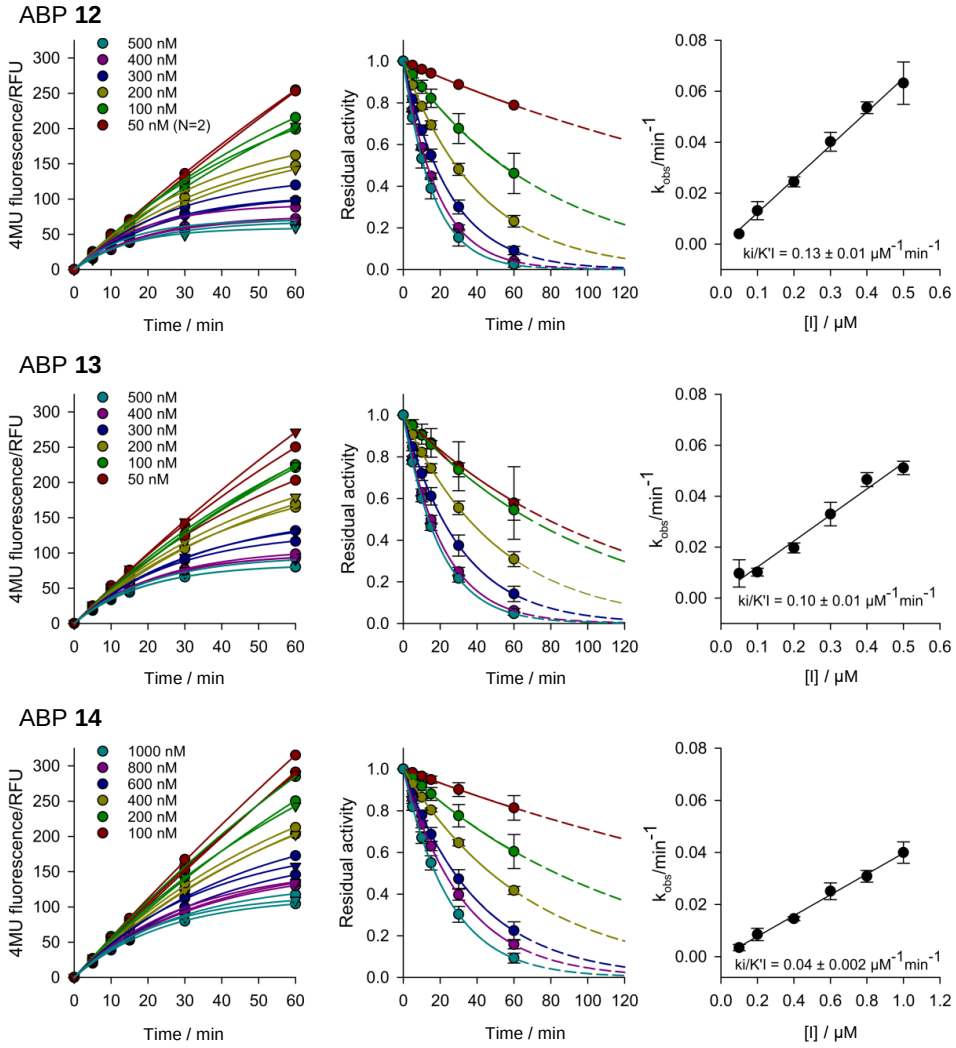


Figure 4.S2. Kinetic parameter measurements for compound 8–14 towards AcGH79. Left panels: progressing curves of relative fluorescent unit (RFU) for each sample incubated with the indicated inhibitor concentration for different incubation time. Middle panels: inhibition curves for each sample incubated with the indicated inhibitor concentration through the indicated incubation time, relative to the activity from the untreated sample at 60 min incubation time. Right panel: k_{obs} vs inhibitor concentration ([I]) plots for each compound.

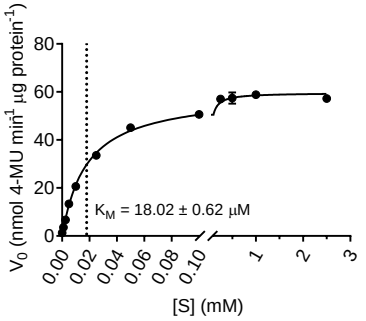
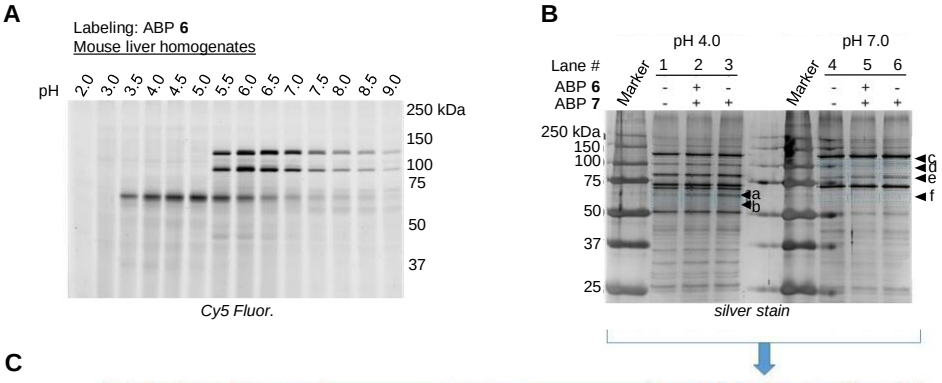


Figure 4.S3. Michaelis-Menten plot for AcGH79 using 4-MU-β-glc as substrate.

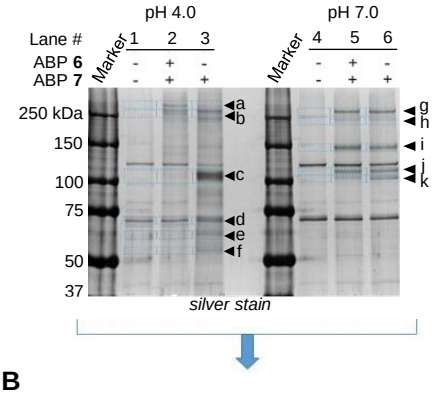


C

Band	Rank	Accession	Protein name	Score	Matches	Coverage(%)	emPAI
2a	2	P70699	Lysosomal alpha-glucosidase, GAA	471	33	18.2	0.44
	4	P17439	Glucosylceramidase, GBA	253	20	19.6	0.47
3a	2	P70699	Lysosomal alpha-glucosidase, GAA	760	40	19.9	0.72
	4	P17439	Glucosylceramidase, GBA	238	21	19.8	0.39
2b	17	P17439	Glucosylceramidase, GBA	139	13	13.4	0.32
3b	13	P17439	Glucosylceramidase, GBA	125	11	15.3	0.18
6c	2	Q8BHN3-2	Isoform 2 of Neutral alpha-glucosidase AB, GANAB-2	909	79	46.8	1.64
6d	2	Q8BHN3	Neutral alpha-glucosidase AB, GANAB	762	62	30.0	1.06
6e	3	Q8BHN3	Neutral alpha-glucosidase AB, GANAB	641	47	23.7	0.82
5f	18	P17439	Glucosylceramidase, GBA	125	11	12.2	0.25
	20	P70699	Lysosomal alpha-glucosidase, GAA	120	8	6.3	0.06
6f	26	P70699	Lysosomal alpha-glucosidase, GAA	67	4	3.8	0.09
1a, 1b, 4c-f, 5c-e	No target enzymes						

Figure 4.S4. Target identification for ABP 6 and 7 using chemical proteomics in mouse liver homogenates. A) ABP **6** labeling at various pHs. B) Silver stain of gels containing samples labeled with(out) ABPs at pH 4.0 or 7.0, and affinity-enriched for biotin. C) List of identified glycosidases by LC-MS measurement by LC-MS-based proteomics at positions shown in B.

A



C

pH 4.0	pH 7.0
a = MGAM, Sis	**g** = Sis
b = Sis, MGAM	**h** = Sis
c = Sis, MGAM	**i** = Lct
d = Sis, MGAM, GAA	**j** = Sis, MGAM
e = Sis, MGAM	**k** = GANAB, Sis
f = MGAM, Sis, GBA	

B

Band	Rank	Accession	Protein name	Score	Matches	Coverage(%)	emPAI
2a	4	B5THE2	Maltase-glucoamylase, MGAM	312	35	12.3	0.18
	30	B5THE3	Sucrase-isomaltase, Sis	43	5	2.5	0.01
3a	19	B5THE2	Maltase-glucoamylase, MGAM	160	12	4.5	0.07
	20	B5THE3	Sucrase-isomaltase, Sis	151	11	4.2	0.06
3b	16	B5THE3	Sucrase-isomaltase, Sis	292	29	12.3	0.13
	31	B5THE2	Maltase-glucoamylase, MGAM	125	11	15.3	0.18
3c	1	B5THE3	Sucrase-isomaltase, Sis	1260	51	14.4	0.37
	33	B5THE2	Maltase-glucoamylase, MGAM	129	9	3.4	0.04
3d	12	B5THE3	Sucrase-isomaltase, Sis	211	19	10.3	0.09
	23	B5THE2	Maltase-glucoamylase, MGAM	124	9	4.9	0.04
	29	A2AFL5	Lysosomal alpha-glucosidase, GAA	55	1	6.2	0.09
3e	12	B5THE3	Sucrase-isomaltase, Sis	309	20	6.5	0.1
	17	B5THE2	Maltase-glucoamylase, MGAM	241	12	4.5	0.07
3f	7	B5THE2	Maltase-glucoamylase, MGAM	505	30	12.8	0.2
	20	B5THE3	Sucrase-isomaltase, Sis	265	19	8	0.1
	72	P17439	Glucosylceramidase, GBA	28	2	2.9	0.05
5g	1	B5THE3	Sucrase-isomaltase, Sis	468	45	13.9	0.28
6g	4	B5THE3	Sucrase-isomaltase, Sis	421	47	15.3	0.3
5h	18	B5THE3	Sucrase-isomaltase, Sis	190	9	3.7	0.06
6h	1	B5THE3	Sucrase-isomaltase, Sis	302	28	10.3	0.18
5i	6	F8VPT3	Protein Lct, Lct	567	54	13.3	0.27
6i	1	F8VPT3	Protein Lct, Lct	549	65	14.1	0.32
5j	11	B5THE3	Sucrase-isomaltase, Sis	553	36	11.8	0.23
	43	B5THE2	Maltase-glucoamylase, MGAM	53	8	2.9	0.03
6j	2	B5THE3	Sucrase-isomaltase, Sis	605	52	16.1	0.3
5k	2	B5THE3	Sucrase-isomaltase, Sis	551	33	11.2	0.24
6k	1	Q8BHN3	Neutral alpha-glucosidase AB, GANAB	683	65	34.4	1.0
	2	B5THE3	Sucrase-isomaltase, Sis	614	39	12.3	0.23
1a-f, 2b-f, 4g-k	No target enzymes						

Figure 4.S5. Target identification for ABP 6 and 7 using chemical proteomics in mouse intestine homogenates. A) Silver stain of gels containing samples labeled with(out) ABPs at pH 4.0 or 7.0, and affinity-enriched for biotin. B) List of identified glycosidases by LC-MS measurement by LC-MS-based proteomics at positions shown in A. C) Table of the most abundant glycosidase identified at each gel position shown in A.

Table 4.S1. Apparent IC_{50} values for compounds 1–7 towards recombinant human GBA (rGBA).

Compounds	IC_{50} (nM) (rGBA)
1 (JJB307)	155,000 ± 1,770
2 (CF21)	41,900 ± 490
3 (CF22)	603 ± 28.0
4 (JJB382)	593 ± 214
5 (JJB347)	1,080 ± 63.6
6 (JJB383)	816 ± 526
7 (JJB384)	2,060 ± 219

Table 4.S2. Crystallographic data collection and refinement statistics.

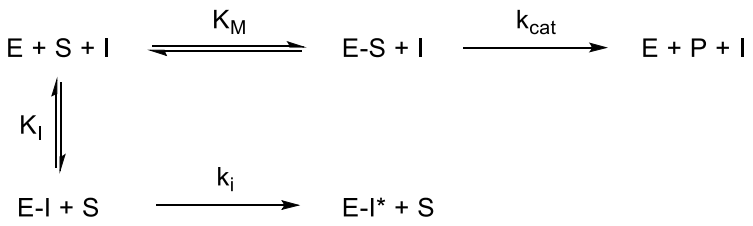
	CjAgd31B–**3**	AcGH79(wt)–**10**	AcGH79(E287Q)–**10**
PDB	5123	5G0Q	5L77
Data collection			
Space group	P622	C2	C2
Cell dimensions			
a, b, c (Å)	198.0, 198.0, 103.0	148.5, 44.9, 82.2	150.2, 45.1, 82.0
α, β, γ (°)	90, 90, 120	90, 114.8, 90	90, 115.1, 90
Resolution (Å)	85.74-1.95	42.60-1.60	28.90-1.24
R_{merge}	0.13 (1.64)	0.058 (0.664)	0.045 (0.489)
$I/\sigma I$	19.6 (2.3)	12.1 (1.8)	13.7 (1.9)
Completeness (%)	100 (100)	99.5 (99.7)	89.8 (49.7)
Redundancy	20.0 (20.1)	4.0 (3.7)	3.9 (2.7)
Refinement			
Resolution (Å)	85.74-1.95	42.60-1.60	28.90-1.24
No. reflections	82110	56355	120177
R_{work}/ R_{free}	0.17/0.19	0.12/0.17	0.13/0.16
No. atoms			
Protein	6299	3434	3477
Ligand/ion	102	36	56
Water	616	294	486
B-factors			
Protein	33.8	19.82	17.87
Ligand/ion	64.8	28.07	33.05
Water	39.7	30.29	30.95
R.m.s. deviations			
Bond lengths (Å)	0.012	0.020	0.016
Bond angles (°)	1.57	1.86	1.73

Value in parenthesis is for highest-resolution shell.

4.S2. Supporting note

Modeling irreversible enzyme inactivation by ABPs in presence of substrate.

Inactivation of enzyme by ABP is assumed to follow the following model (**Main text** Ref. 39):



K_M is the equilibrium constant for E-S formation (i.e. Michaelis constant; the concentration of S in which product formation is at 1/2 V_{max}), k_{cat} is the rate constant for P formation, K_I is the equilibrium constant for E-I formation (i.e. inhibition constant), k_i is the rate constant for E-I* formation (i.e. inactivation rate constant).

In the experiments I is the ABP, S is the substrate 4-MU-GlcA, E is AcGH79 and P is the measured product 4-MU. [I] and [S] are both >> $[E_{tot}]$ which is fixed at 260 pM ($[E_{tot}]$ = [E]+[EI]+[ES]+[EI*]). [S] is 2.5 mM.

In this model, the substrate and ABP compete reversibly for the enzyme active site, followed by irreversible reactions to give either P, or the inactivated enzyme complex E-I*.

Under these conditions, initial enzyme reaction rate V_0 is given by:

$$V_0 = {dP}/{dt} = \frac{k_{cat}[S][E_{tot}]}{K_M\left(1+{[I]}/{K_I}\right)+[S]} \qquad (1)$$

At time t:

$$V_t = \frac{k_{cat}[S][E_{tot} - EI^*]}{K_M\left(1+{[I]}/{K_I}\right)+[S]} \qquad (2)$$

Analogously, the rate of enzyme inactivation at time t is given by:

$$dEI^*/dt = \frac{k_i[I][E_{tot} - EI^*]}{K_I\left(1+{[S]}/{K_M}\right)+[I]} = k_{obs}[E_{tot} - EI^*] \quad (3)$$

Where k_{obs} is defined as the observed inactivation rate constant:

$$k_{obs} = \frac{k_i[I]}{K_I\left(1+{[S]}/{K_M}\right)+[I]} \quad (4)$$

Here, the appeared equilibrium constant for I is dependent on the type and the amount of S applied in the assay (K_M and [S]), and is defined as K'_I:

$$K_I\left(1+{[S]}/{K_M}\right) = K'_I \quad (5)$$

$$k_{obs} = \frac{k_i[I]}{K'_I+[I]} \quad (6)$$

For deriving k_{obs} experimentally, [I] and [S] can be both set at >> [E_{tot}]. Under this condition, k_{obs} is a pseudo-first order inactivation rate constant (dependent on [I]). The extent of enzyme inactivation at time t is quantified by the residual activity V_t/V_0, where V_t is the enzyme reaction rate at time t. Thus, V_t/V_0 for each value of [I] can be described by the first-order rate equation:

$$V_t/V_0 = e^{-k_{obs}t} \quad (7)$$

$$V_t = V_0 e^{-k_{obs}t} \quad (8)$$

The amount of product [P] formed at time t is the integral of V_t, with respect to t, between the interval [0, t]:

$$[P] = \int_0^t [V_t]dt = {V_0}/{k_{obs}}\,(1 - e^{-k_{obs}t}) \quad (9)$$

Fitting 4-MU fluorescence (which is $\propto$ [P]) *vs.* time using (9) allows k_{obs} to be derived for each value of [I]. Next, the derived k_{obs} values at each [I] can be plotted against each [I], and the resulting plot of k_{obs} vs [I] can be fitted using (6), which is an analogue of Michaelis-Menton equation, to derive K'_I and k_i.

When [I] << K'_I, as observed from the data in this chapter, K'_I and k_i cannot be accurately calculated. Instead, (4) is well approximated by:

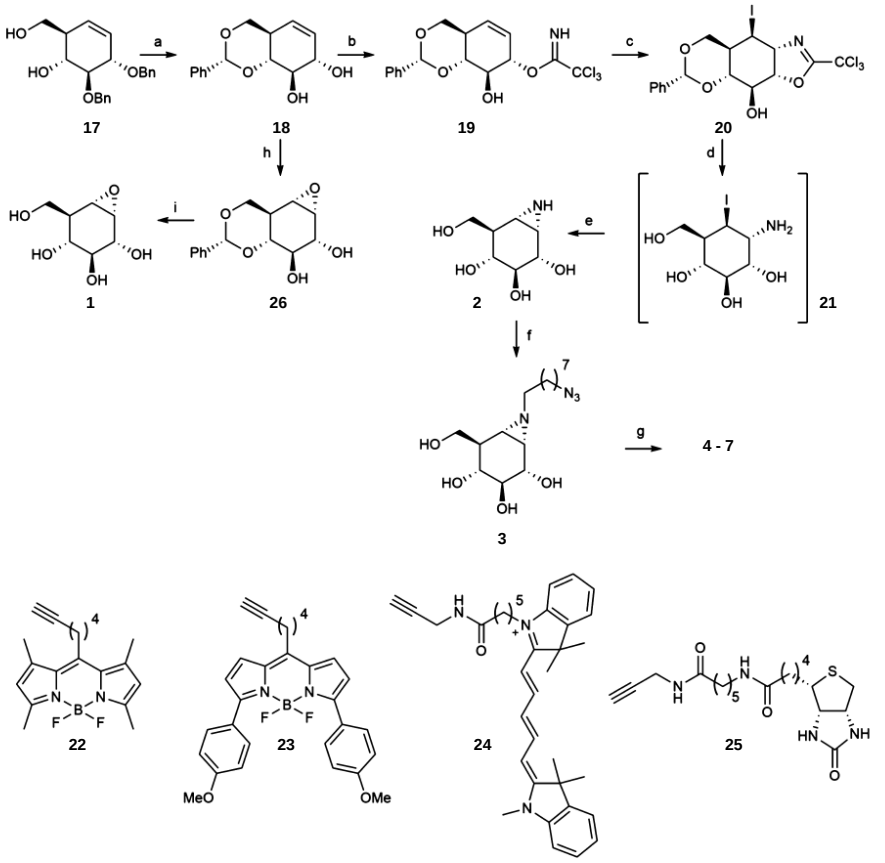
$$k_{obs} \approx {k_i}/{K'_I}\,[I] \qquad (10)$$

Thus, a subsaturated plot of k_{obs} against [I] is better fitted using linear function (10), allowing an estimate of k_i/K'_I to be derived from the gradient of the slope.

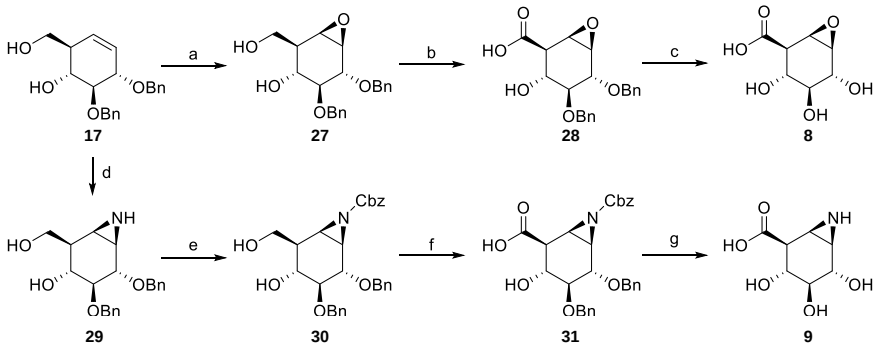
k_i/K_I is derived from k_i/K'_I by applying (5), where [S] is known and K_M can be derived by a separate Michaelis-Menton experiment for S on the given enzyme.

4.S3 Synthetic schemes for compounds 1-15 (Department of Bio-organic synthesis, Leiden University)

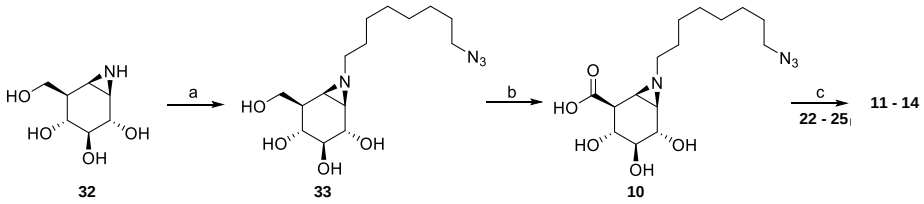
For complete synthetic method and characterization data, the reader is directed to the supporting information in the published articles.[Main text refs 19, 31]



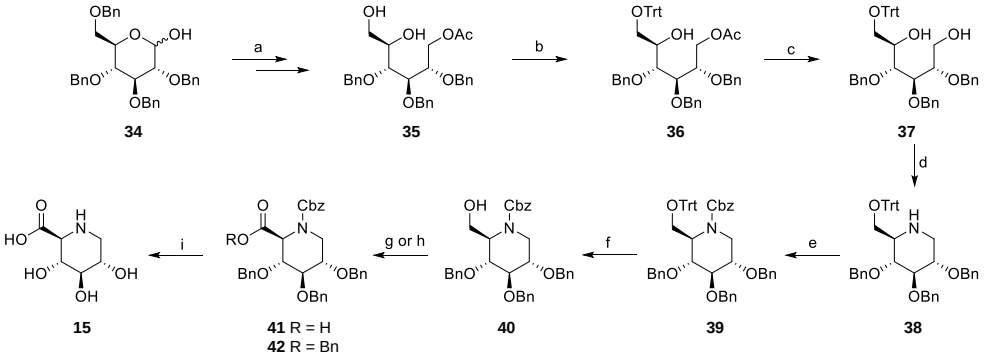
Scheme 4.S1. Synthesis of 1,6-epi-cyclophellitol 1, the cyclophellitol aziridine inhibitors 2, 3, and ABPs 4−7. Reagents and conditions: (a) (i) Li, NH_3, THF, −60 °C, 57 %; (ii) $PhCH(OMe)_2$, CSA, DMF, 61 %. (b) CCl_3CN, DBU, DCM, 0 °C. (c) $NaHCO_3$, I_2, H_2O, two step yield 41 %. (d) 37 % HCl aq, dioxane. (e) $NaHCO_3$, MeOH, two step yield 63 %. (f) 1-azido-8-iodooctane, K_2CO_3, DMF, 80 °C, 39 %. (g) **4**, **5**, **6**, or **7**, $CuSO_4$, sodium ascorbate, DMF, 38 % 3, 11 % 4, 24 % 5, 23 % 6. (h) *m*-CPBA, DCM, 40 °C, 44 %. (i) $Pd(OH)_2/C$, H_2, MeOH, 68 %.



Scheme 4.S2. Synthesis of uronic cyclophellitol epoxide 8 and aziridine 9. Reagents and conditions: (a) *m*-CPBA, Na_2HPO_4 (aq., 1 M), NaH_2PO_4 (aq., 1.0 M), DCM, 50 °C, 55 %; (b) BAIB, TEMPO, DCM/^{t}BuOH/H_2O (4:4:1), 0 °C, 79 %; (c) H_2, $Pd(OH)_2/C$, MeOH, 98 %; (d) i) CCl_3CN, DBU, DCM, 0 °C; ii) I_2, $NaHCO_3$, H_2O; iii) 37 % HCl aq., dioxane, 60 °C; iv) $NaHCO_3$, MeOH, 45 %; (e) Cbz-Cl, Pyridine, THF, 75 %; (f) BAIB, TEMPO, DCM/^{t}BuOH/H_2O (4:4:1), 0 °C, 60 %; (g) NH_3, Li, THF, -60 °C, 60 %.



Scheme 4.S3. Synthesis of compounds 10-14. Reagents and conditions**:** (a) K_2CO_3, DMF, 1-azido-8-iodooctane, 80 °C, 50 %; (b) TEMPO, NaClO, NaBr, NaOH, H_2O, 15 %; (c) $CuSO_4$, sodium ascorbate, DMF, 22 % **11**, 13 % **12**, 17 % **13**, 13 % **14.**



Scheme 4.S4. Synthesis of 2S, 3R, 4R, 5S-trihydroxypipecolic acid 15. Reagents and conditions: (a) Supplementary Ref. [1] (b) DMAP, TEA, TrtCl, DCM, 3 h, 75 %; (c) Na, MeOH, 3 h, 91 %; (d) i): $(COCl)_2$, DMSO, DCM, –65 °C, 2 h; ii): TEA, -65 to 5 °C, 2 h; iii): NH_4HCO_3, $NaBH_3CN$, 3 Å mol. sieves, MeOH, 0 °C, 1 h, then 0 °C to RT, 20 h, 85 %; (e) CbzCl, THF, TEA, RT, 20 h, 72 %; (f) *p*-TsOH, DCM, 0 °C to RT, 3 h, 83%; (g) TEMPO, BAIB, DCM/H_2O (2/1), 0 °C, 5 h, **41**: 71 %; (h) i): TEMPO, BAIB, DCM/H_2O (2/1), 0 °C, 5 h; ii): BnBr, Cs_2CO_3, DMF, RT, 3 h, **42**: 75 %; (i) H_2, $Pd(OH)_2$/C, AcOH/H_2O(4/1), RT, 16 h, 57 %.

6.S4 Supplementary Reference

1 La Ferla B, Bugada P & Nicotra F (2006) Synthesis of the dimethyl ester of 1-deoxy-L-idonojirimycin-1-methylenphosphonate: A new approach to iminosugar phosphonates. *J Carb Chem* **25**, 151–162.