Activity-based protein profiling of glucosidases, fucosidases and glucuronidases

Jiang, J.

Activity-based protein profiling of glucosidases, fucosidases and glucuronidases

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,

volgens het besluit van het College voor Promoties

te verdedigen op 23 juni 2016

klokke 12:30 uur

door

江建冰

Jianbing Jiang

Geboren te Liaocheng, China in 1985

Promotiecommissie

Promotor:	Prof. dr. H. S. Overkleeft
	Prof. dr. J. M. F. G. Aerts
Co-promotor:	Dr. J. D. C. Codée
Overige leden:	Prof. dr. G. J. Davies (University of York)
	Prof. dr. G. A. van der Marel
	Prof. dr. J. Brouwer
	Dr. L. I. Willems (Simon Fraser University)
	Dr. L. Wu (University of York)

Doctoral Thesis, Leiden University, 2016

Cover design: Jianbing Jiang

Printed by Gildeprint

ISBN: 978-94-6233-3147

To my family

Table of Contents

List of abbreviations

ABP	activity-based probe
ABPP	activity-based protein profiling
Ac	acetyl
ACN	acetonitrile
AMP-DNM	*N*-[5-(adamantan-1-yl-methoxy)-pentyl]-1-deoxynojirimycin
aq.	aqueous
Asp	aspartic acid
Bn	benzyl
Boc	*tert*-butyloxycarbonyl
BODIPY	Boron-dipyrromethane
br	broad
BSA	bovine serum albumin
C	chair
CAZy	carbohydrate-active enzymes
CBE	conduritol β-epoxide
Cbz	benzyloxycarbonyl
COS-7	African green monkey kidney fibroblast cell line
d	doublet
δ	chemical shift
DBU	1,8-diazabicycloundec-7-ene
DCM	dichlormethane
dd	double doublet
ddd	double double doublet
DIC	*N*.*N*-diisopropyl carbodiimide
DMAP	4-(dimethylamino)pyridine
HPLC	high performance liquid chromatography
HPSE	heparanase
HRMS	high resolution mass spectrometry
DMF	*N.N*-dimethylformamide
DMSO	dimethylsulfoxide
dt	double triplet
DTT	dithiothreitol
EDC	1-ethyl-3-(3-dimethyl-aminopropyl)-carbodiimide
EDTA	ethlenediaminetetraacetate
EEDQ	2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline
emPAI	Exponentially modified protein abundance index
eq.	molar equivalents
Et	ethyl
et al.	et alii (and others)
Fmoc	(9*H*-fluoren-9-yl)methoxycarbonyl
FUCA	α-L-fucosidase
GAA	lysosomal α-glucosidase
GBA	glucocerebrosidase
GBA2	nonlysosomal glucocerebrosidase
GBA3	braod specifity β-glucosidase
GLA	lysosomal α-galacosidase
GLB	lysosomal β-galactosidase
GUSB	lysosomal β-glucuronidase
h	hour(s)
HEK293	Human embryonic kidney cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
SDS	sodium dodecyl sulfate
sp.	species
Su	succinimidyl
t	triplet
TBAF	tetrabutylammonium fluoride

HRP	horseradish peroxidase
Hz	Hertz
IC_{50}	half maximal inhibitory
IR	infrared
J	coupling constant
LCMS	liquid chromatography mass spectrometry
m	multiplet
M	molar
m/z	mass to charge ratio
*m*CPBA	*meta*-chloroperoxybenzoic acid
Me	methyl
min	minutes
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NIS	*N*-iodosuccinimide
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
Pd/C	palladium on charcoal
PEG	polyethyleneglycol
Ph	phenyl
ppm	parts per million
PVDF	polyvinylidene difluoride
q	quartet
quant.	quantitive
R_f	retention factor
rt	room temperature
R_t	retention time
rpm	revolutions per minute
s	singlet
sat.	saturated
TBAI	tetrabutylammonium iodide
TBS	Tris-buffered saline
*t*Bu	*tert*-butyl
TEA	triethyl amine
tert	teritiary
Tf	triflate
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	trimethylsilane
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
Trt	triphenylmethane
UV	ultraviolet
WT	wild type

1

General introduction

Glycoside hydrolases (GHs), enzymes that catalyze the hydrolytic cleavage of glycosidic bonds, receive continuing interest both in fundamental and applied biology and biomedicine. Lysosomal storage disorders (LSDs) are caused by inborn metabolic errors due to deficiency in specific lysosomal enzymes, most commonly GHs. Diagnosis and treatment of LSDs require regular quantification of the active lysosomal enzymes in patient tissues. Activity-based protein profiling (ABPP) has emerged in the past decades as a powerful technique to study enzyme families in cell extracts and living tissues. Originally developed for serine hydrolases and cysteine proteases, various enzyme classes can be studied by means of ABPP today, including retaining GHs. The research described in this thesis focused on expanding the field of activity-based glycosidase profiling through the development and application of activity-based probes (ABPs) for several retaining GHs, namely β-glucosidases, α-L-fucosidases, α-glucosidases and β-glucuronidases. The first part of this chapter provides an overview of retaining and inverting β-glucosidases, including the catalytic mechanisms they employ in processing their substrates and the mechanism-based inhibitors of retaining β-glucosidases. The final part of this chapter presents a historical perspective on ABPs of retaining β-glucosidases as well as an outline of the contents of the following chapters.

1.1 β-Glucosidases

Glycoside hydrolases (GHs) are found in all kingdoms of life, and play important roles in a broad range of biological processes.[1-5] β-Glucosidases catalyze the hydrolysis of β-D-glucosidic linkages, with release of glucose (*exo*-glucosidase) or oligosaccharides featuring a glucose moiety at the reducing end (*endo*-glucosidases) as the result,[6] this could also be expanded to other *exo*- and *endo*-glycosidases classification. Based on the stereochemical outcome of the hydrolysis reaction, β-glucosidases can be divided into retaining and inverting enzyme groups. The two catalytic mechanisms employed by β-glucosidases to achieve glycosidic bond cleavage are shown in Figure 1.

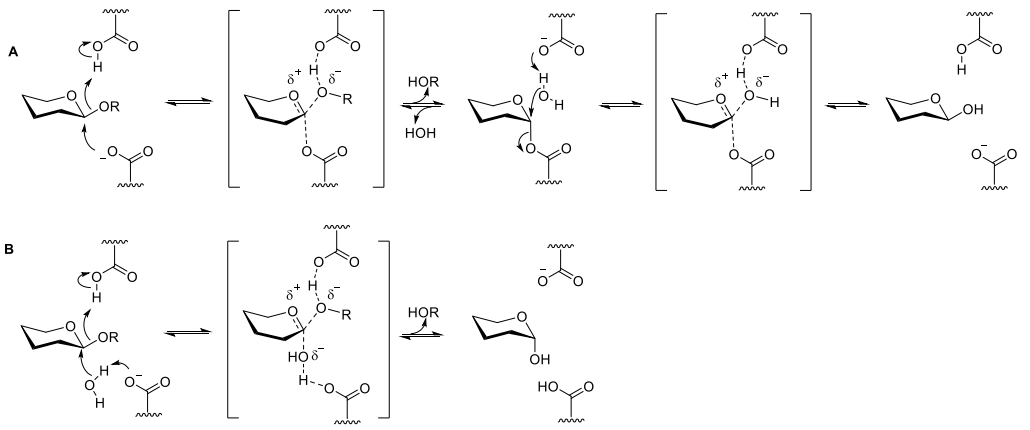


Figure 1. Generalized mechanisms for β-D-glucosides hydrolysis: A) retaining β-glucosidase and B) inverting β-glucosidase.

Retaining GHs process their substrates with overall retention of anomeric configuration whereas inverting GHs do so with inversion of configuration.[7] In 1953, Daniel Koshland proposed that retaining glycosidases act by a double-displacement mechanism involving an enzyme or substrate nucleophile (two inversions resulting in net retention of the anomeric configuration).[8] The catalytic machinery of a retaining β-glucosidase involves two catalytic carboxylates (Figure 1A): one acting as an acid-base catalyst and the other acting as a nucleophile. In the first step (glycosylation), the aglycon is protonated and expelled in a formal S_N2 process by the nucleophilic residue to yield a covalent glycosyl-enzyme intermediate. In the second step (deglycosylation), water enters the enzyme active site, gets activated by (partial) deprotonation by the catalytic acid/base carboxylate and substitutes in another formal S_N2 process the active site nucleophile in the glycosyl-enzyme adduct to yield with overall retention of configuration at the anomeric center of the hydrolyzed sugar. The active sites of inverting β-glucosidases are also composed of two carboxyl groups, with one acting as a general acid catalyst and the other as a general base catalyst (Figure 1B). Hydrolysis proceeds in a single step, in which the catalytic acid activates the aglycon while the catalytic

base activates the nucleophilic water molecule which displaces the aglycon via a single transition state with substantial oxocarbenium ion character.[9]

1.2 Mechanism-based inhibitors and activity-based probes of retaining β-glucosidases

Mechanism-based inhibitors are molecules that react with an enzyme, normally within the active site and following the mechanism employed by the enzyme in substrate processing. When a mechanism-based inhibitor is brought into contact with a target enzyme, a covalent bond between the reactants (enzyme and inhibitor) is formed that is stable over time, thus leading to permanent inactivation of the enzyme. Mechanism-based enzyme inhibition can be achieved most easily when an enzyme processes its substrate through the intermediate formation of a covalent intermediate. From this point of view, retaining β-glucosidases appear more susceptible to mechanism-based inhibition than inverting β-glucosidases.[10] Cyclophellitol (**1**)[10] and cyclophellitol aziridine (**2**)[11] are typical examples of mechanism-based inhibitors against retaining β-glucosidases (Figure 2A). They are activated by protonation by the general acid/base catalytic residue and next opened by the nucleophilic carboxylic acid to yield a covalent enzyme-substrate adduct. Compared to the acylal (O-C-O-C=O) linkage that emerges during β-glucoside processing (see Figure 1A), the resulting ester linkage is considerably more stable and the retaining β-glucosidase is effectively and irreversibly inhibited.

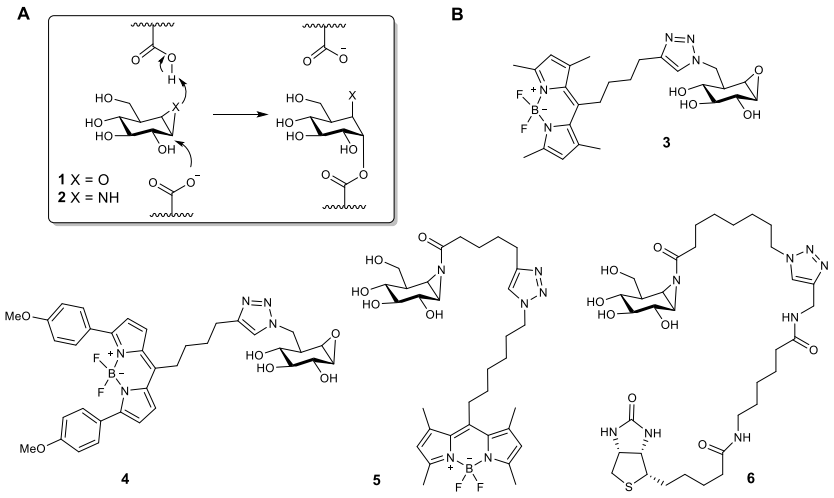


Figure 2. A) Proposed mechanism of retaining β-glucosidases inhibition by cyclophellitol (**1**) and cyclophellitol aziridine (**2**). B) Examples of ABPs based on cyclophellitol and cyclophellitol aziridine scaffolds.

Activity-based protein profiling (ABPP) has emerged as a useful technology to study GHs activities in various surroundings. ABPP, pioneered by the Cravatt laboratory for the study of

serine hydrolase families,[12] utilizes activity-based probes (ABPs), compounds designed to specifically react in a covalent and irreversible fashion with an enzyme or a class of enzymes and that are equipped with a reporter molecule (fluorophore, biotin, bioorthogonal group) for detection and/or identification of the covalently captured enzymes (Figure 3). An ABP normally contains a reactive moiety (or 'warhead') that can form a covalent bond with the enzyme (family) of interest, but is sufficiently inert to survive in cell extracts or living cells when not in contact with the enzyme target(s). A spacer links the reactive moiety to a reporter group (tag), so that the latter does not interfere with binding the enzyme. The third essential structural element, the tag group, commonly exists of a fluorescent group (for instance, BODIPY, rhodamine or fluorescein) for visualization in gel or in living cells or a biotin group for affinity enrichment, purification and subsequent mass spectrometry detection.

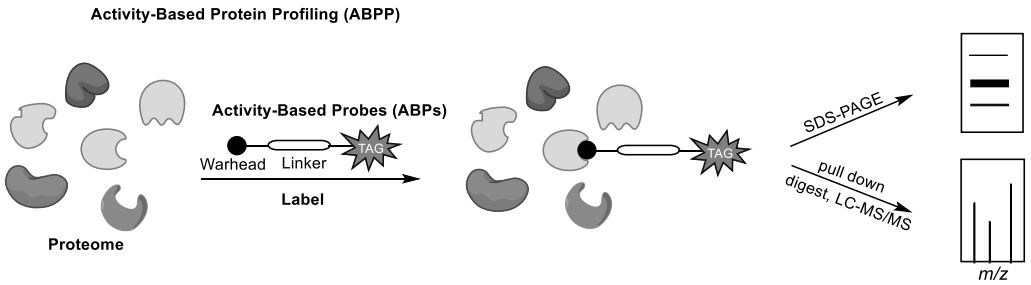


Figure 3. In a typical ABPP experiment, a proteome is treated with the ABP. Key to the success of the probe is the covalent attachment to the enzyme(s) of interest. Depending on the tag (reporter) group, the labeled enzyme(s) can then be either directly visualized with SDS–PAGE by fluorescent scanning or can be purified by streptavidin pull down, digested with trypsin followed by analysis of the resulting peptides by mass spectrometry

ABPs are generally designed with most ease when dealing with enzymes that, during processing, form a covalent intermediate with their substrate. Based on this theory, mechanism-based retaining β-glucosidase inhibitors, cyclophellitol (**1**) and cyclophellitol aziridine (**2**), have been used as scaffolds for ABP development. These ABPs include cyclophellitol derivatives and cyclophellitol aziridines derivatives modified with either a BODIPY or a biotin.[13] The first-generation of cyclophellitol-inspired retaining β-glucosidase ABPs are compounds **3** and **4** (Figure 2B). Both ABPs **3** and **4**, though having a large BODIPY moiety appeared to be quite potent and efficient inhibitors of glucocerebrosidase (GBA), inhibiting this enzyme much more potently than the parent compound, cyclophellitol **1**.[14] These probes were applied to visualize active GBA molecules in various mouse tissue extracts and in living cells. More recently, cyclophellitol aziridine **2** was employed for the design of second generation retaining β-glucosidase ABPs, such as fluorescent probe **5** and biotin probe **6**.[15] These probes, with the epoxide substituted for aziridine allowing the reporter groups to be introduced on the aziridine nitrogen proved to be more potent towards a range of retaining

β-glucosidases. They were shown to label in murine tissue extracts, apart from GBA, the non-lysosomal glycosylceramidase (GBA2), the cytosolic β-glucosidase (GBA3) and intestinal lactase/phorizin hydrolase (LPH), as well as numerous plant retaining GHs such as myrosinases, β-glucosidases, β-galacosidases and β-xylosidases in *Nicotiana benthamiana*.[16]

1.3 Activity-based retaining glycosidase probes based on cyclophellitol aziridine

Retaining β-glucosidases are one class of glycoside hydrolases (GHs), which are often classified on the basis of their amino acid sequence according to the CAZy system (www.cazy.org), a database of Carbohydrate-Active enZYmes (CAZymes).[17] GHs in the same CAZy family often share similarities in their catalytic mechanism.[18] For instance, glycosyl hydrolase family 1 (GH1) contains enzymes that possess a classical $(\alpha/\beta)_8$ triosephosphate isomerase (TIM) barrel fold and employ a Koshland double-displacement mechanism in their substrate turnover.[19] Cyclophellitol aziridine is a rather potent and selective mechanism-based retaining β-glucosidase inhibitor, and is readily modified into a retaining β-glucosidase ABP through acylation or alkylation of the aziridine nitrogen with functional tag groups.[13] The research described in this thesis aimed to demonstrate that the cyclitol aziridine scaffold allows for the development of ABPs targeting other retaining glycoside hydrolases, specifically, α-L-fucosidases, α-glucosidases and β-glucuronidases. The following paragraph outlines the contents of this thesis, which essentially describes the feasibility of configurational and functional analogues of cyclophellitol aziridine as effective and selective in-class retaining glycosidase activity-based probes.

1.4 Aim and outline of thesis

The design of configurational cyclophellitol aziridine isomers as starting points for the development of ABPs targeting different retaining GHs requires suitable routes of synthesis for their preparation. Existing literature on cyclophellitol aziridine synthesis is scarce, but has grown in recent years. **Chapter 2** provides a concise overview of the existing routes for synthesis of cyclophellitol aziridine isomers. **Chapter 3** describes the synthesis of cyclophellitol aziridine based α-L-fucosidase ABPs and L-fuconojirimycin inhibitors, as well as *in vitro* and *in vivo* profiling of active GH29 α-L-fucosidases in mammalian tissue. A comparative study between *N*-acyl aziridine and *N*-alkyl aziridine ABPs for β-glucosidases and α-L-fucosidases is described in **Chapter 4**. **Chapter 5** provides the development of α-glucoside cyclophellitol aziridine ABPs for GH31 α-glucosidases. Cyclophellitol aziridine ABPs for β-glucuronidases are the subject of the studies described in **Chapter 6** (synthesis) and **Chapter 7** (biological studies). In **Chapter 6**, the preparation of both *N*-acyl and *N*-alkyl aziridine isomers of β-glucuronide and related ABPs for β-glucuronidases are described. **Chapter 7** shows the probes from **Chapter 6** to be able to modify both GH2 lysosomal β-glucuronidase (an *exo*-glycosidase) and GH79 heparanase (an *endo*-glycosidases). **Chapter 8** gives a summary of the research described in this thesis and suggests some future prospects.

1.5 References

[1] R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683-720.

[2] J. Lee, *J. Biotech.*, **1997**, *56*, 1-24.

[3] J. W. Dennis, S. Laferte, C. Waghorne, M. L. Breitman and R. S. Kerbel, *Science* **1987**, *236*, 582-585.

[4] Y. Nagai, *Pure Appl. Chem.* **1997**, *69*, 1893-1896.

[5] R. G. Spiro, *J. Biol. Chem.* **2000**, *275*, 35657-35660.

[6] M. M. Cox, *Lehninger principles of biochemistry,* Worth Publishers New York, **2000**.

[7] J. E. Barnett, *Biochem. J.* **1971**, *123*, 607-611.

[8] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416-436.

[9] M. L. Sinnott, *Chem. Rev.* **1990**, *90*, 1171-1202.

[10] S. Atsumi, K. Umezawa, H. Iinuma, H. Naganawa, H. Nakamura, Y. Iitaka and T. Takeuchi, *J. Antibiot.* **1990**, *43*, 49-53.

[11] K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima and M. Nakata, *J. Antibiot.* **1991**, *44*, 912-914.

[12] B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383-414.

[13] L. I. Willems, J. Jiang, K. Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. Beenakker, S. P. Schroder, J. M. Aerts, G. A. van der Marel, J. D. Codee and H. S. Overkleeft, *Chem. Eur. J.* **2014**, *20*, 10864-10872.

[14] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft and J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907-913.

[15] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529-12533.

[16] B. Chandrasekar, T. Colby, A. Emran Khan Emon, J. Jiang, T. N. Hong, J. G. Villamor, A. Harzen, H. S. Overkleeft and R. A. van der Hoorn, *Mol. Cell. Proteomics* **2014**, *13*, 2787-2800.

[17] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490-495.

[18] B. Henrissat and G. Davies, *Curr. Opin. Struct.Biol.* **1997**, *7*, 637-644.

[19] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon and G. Davies, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7090-7094.

2

The synthesis of cyclophellitol aziridine and its configurational and functional isomers

Jianbing Jiang, Marta Artola, Thomas J. M. Beenakker, Sybrin P. Schröder, Rita Petracca, Casper de Boer, Gijsbert A. van der Marel, Jeroen D. C. Codée, Johannes M. F. G. Aerts and Herman S. Overkleeft, *European Journal of Organic Chemistry*, **2016**, DOI: 10.1021/ejoc.201600472 (micro-review).

2.1 Introduction

The natural product cyclophellitol[1] (**1**, Figure 1A) and its synthetic analogue, cyclophellitol aziridine (**2**)[2] are potent, mechanism-based and irreversible retaining β-glucosidase inhibitors. Retaining β-glucosidases employ a Koshland double displacement mechanism[3] in substrate hydrolysis (Figure 1B), and this process proceeds through a covalent enzyme-glucoside intermediate. Both cyclophellitol (**1**) and cyclophellitol aziridine (**2**) capitalize on this mechanism. They are configurational β-glucopyranose analogues, and outperform the structurally related (and much wider studied) conduritol B epoxide (**3**, CBE – lacking the C8 methylene compared to **1** in potency and selectivity as retaining β-glucosidase inhibitors.[4] Compounds **1** and **2** adopt a $^4H_3$ half chair conformation, thereby emulating the oxocarbenium ion-like transition state.[5] This oxocarbenium ion is trapped to yield the glycosyl-enzyme intermediate and the acylal linkage is then hydrolyzed to release β-glucose with double inversion – thus net retention – of the anomeric carbon configuration.[6] Due to their preferred conformation, cyclitols **1** and **2** fit well

in the retaining β-glucosidase active site and present the epoxide (**1**) or aziridine (**2**) heteroatom for protonation by the acid-base residue in the binding pocket. In an acid-catalyzed nucleophilic attack, the epoxide/aziridine ring opens to form a covalent enzyme-inhibitor adduct. Compared to the acylal linkage featuring during β-glucose hydrolysis, the formed ester is considerably more stable, and the β-glucosidase is irreversibly disabled.

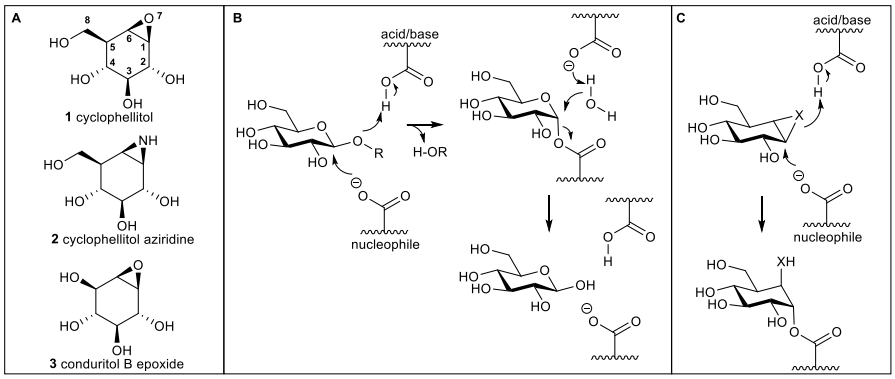


Figure 1. A) Structure of cyclophellitol (**1**), cyclophellitol aziridine (**2**) and conduritol B epoxide (**3**). B) Mechanism employed by retaining β-glucosidases. C) Mechanism-based, irreversible retaining β-glucosidase inhibition by compounds **1** and **2** (X = O or NH).

By virtue of their mechanism of action, cyclophellitol (**1**) and cyclophellitol aziridine (**2**) are selective for retaining β-glucosidases over inverting β-glucosidases. This property distinguishes cyclophellitols from another class of glycosidase inhibitors: the iminosugars (amongst which the archetypal competitive glucosidase inhibitor is deoxynojirimycin).[7] In recent years, covalent and irreversible inhibitors have received growing attention, as they are ideal starting points for the development of activity-based probes (ABPs).[8] In the field of activity-based protein profiling (ABPP), covalent and irreversible inhibitors of an enzyme, or enzyme family, are equipped with a reporter entity – either a fluorophore, a biotin or a bioorthogonal tag – and used to profile their target enzymes in complex biological samples. Cyclophellitol (**1**) is on paper suited for this purpose, and has indeed been transformed[9] into a highly selective ABP highly selective for the human retaining β-glucosidases.

Cyclophellitol aziridine is the more attractive lead for ABP development.[10] It is at least as potent an inhibitor as cyclophellitol, and the aziridine nitrogen can be modified with a range of functional groups, including reporter functionalities, without interfering with recognition by the target enzyme (at least, *exo*-glycosidases to which the target enzymes of **1** and **2** belong are often largely indiscriminate to the nature of the aglycon – the general space also occupied by an aziridine substituent). For this reason, as well as the finding that about half of the

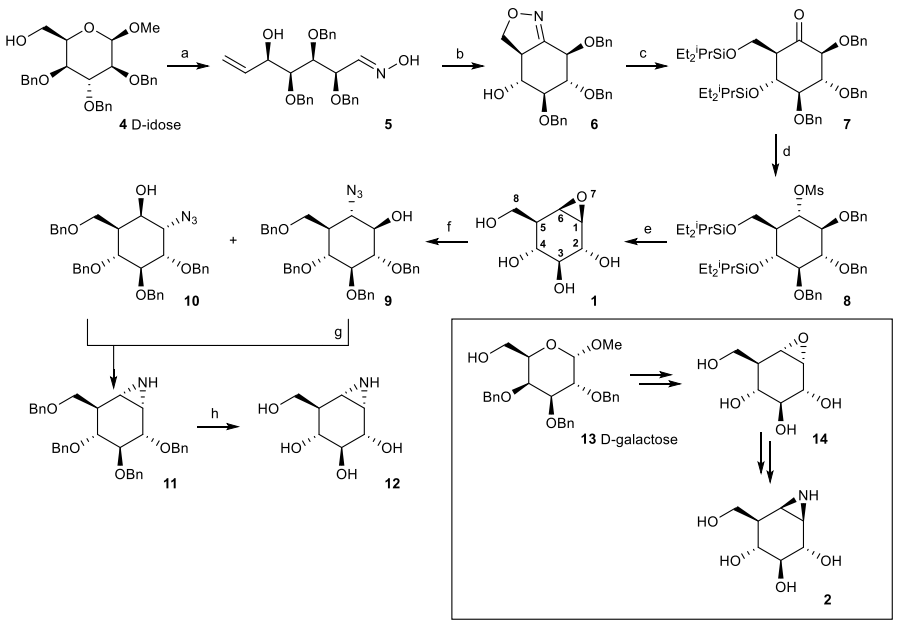
glycosidases known employ a Koshland double replacement mechanism,[11] interest in cyclophellitol aziridines has grown considerably in recent years. Their use as inhibitors and ABPs to monitor glycosidases recognizing and processing configurational and functional isomers of glucose requires however effective synthetic routes to cyclophellitol aziridine synthesis. Herein, the synthesis strategies reported to date are reviewed, starting with the known syntheses of cyclophellitol aziridine (**2**), followed by strategies towards configurational and functional analogues and overall with a focus on methodology towards cyclohexitol aziridines mimicking pyranose sugars.

2.2 Synthetic strategies

2.2.1 Tatsuta's synthesis of cyclophellitol, cyclophellitol aziridine and their 1,6-*epi* isomers

The group of Tatsuta was the first to study the synthesis of cyclophellitol derivatives.[12] Closely following the discovery, by Umezawa, Takeuchi and colleagues,[1] of cyclophellitol (**1**) and its annotation as a mechanism-based retaining β-glucosidase inhibitor,[13] Tatsuta and co-workers disclosed the first synthesis of cyclophellitol (**1**) and cyclophellitol aziridine (**2**), as well as their configurational analogues 1,6-*epi*-cyclophellitol (**14**, Scheme 1) and 1,6-*epi*-cyclophellitol aziridine **12** (atom numbering as indicated in Figure 1, compound **1**). As with most literature syntheses[14] of cyclophellitol, the Tatsuta scheme[14a] starts from a chiral building block, here partially protected D-idopyranose derivative **4**, which is prepared from L-glucose following established procedures.[15] Swern oxidation and Wittig olefination on the primary alcohol is followed by hydrolysis of the methyl acetal. Subsequent reaction of the liberated hemi-acetal with hydroxylamine provided oxime **5** in a series of standard transformations. *In situ* oxidation of the oxime in **5** to the corresponding nitrile oxide (treatment with sodium hypochlorite in methylene chloride) led to a [2+3] cycloaddition to give the key intermediate, isoxazoline **6** as the single stereoisomer. The isoxazoline N-O bond was reduced by Raney nickel catalyzed hydrogenation, after which the free alcohols were protected as the diethylisopropylsilyl ethers to give cyclohexanone **7**. Reduction of the carbonyl provided the desired alcohol in a 3:1 diastereomeric ratio, which was then transformed into mesylate **8**. Catalytic hydrogenation followed by treatment with base led to removal of all protective groups. Finally, intramolecular S_N2 substitution of the methanesulfonyl group yielded cyclophellitol (**1**). Perbenzylation of **1** followed by opening of the epoxide gave the mixture of *trans*-1,2-azido-alcohols **9** and **10**. Treatment of this mixture with triphenylphosphine in a mixture of THF and water gave, after formation of the phosphazene and expulsion of triphenylphosphine oxide in a Staudinger type reaction[16], tetra-*O*-benzyl-1,6-*epi*-cyclophellitol aziridine **11**. Finallly, removal of the benzyl groups under Birch conditions afforded 1,6-*epi*-cyclophellitol aziridine **12**.

Scheme 1. Synthesis of cyclophellitol (**1**), 1,6-*epi*-cyclophellitol (**14**), cyclophellitol aziridine (**2**) and 1,6-*epi*-cyclophellitol aziridine (**12**) by Tatsuta and co-workers.



Reagents and conditions: (a) i) $(COCl)_2$, DMSO, Et_3N, -78 °C; ii) $Ph_3P{=}CH_2$, benzene, 75% over two steps; iii) HCl (aq.), dioxane; iv) HO-NH_2 (HCl salt), pyridine, 80% over two steps; (b) NaOCl, DCM, 70%; (c) i) H_2 (1.0 atm.), Raney Ni, 80%; ii) diethylisopropyl triflate, 2,6-lutidine, DCM; d) i) $BH_3{\cdot}Me_2S$, 45% over three steps); e) i) H_2 (1 atm.), $Pd(OH)_2$, MeOH; ii) NaOH (1.0 M, aq.), 75% over two steps); (f) i) NaH, BnBr, DMF, 90%; ii) NaN_3, DMF, 110 °C, **9**: 27%, **10**: 41%; (g) i) Ph_3P, THF/H_2O; ii) NaOMe, MeOH, 40% over two steps; h) Li, NH_3 (liq.), THF, -78 °C, 60%.

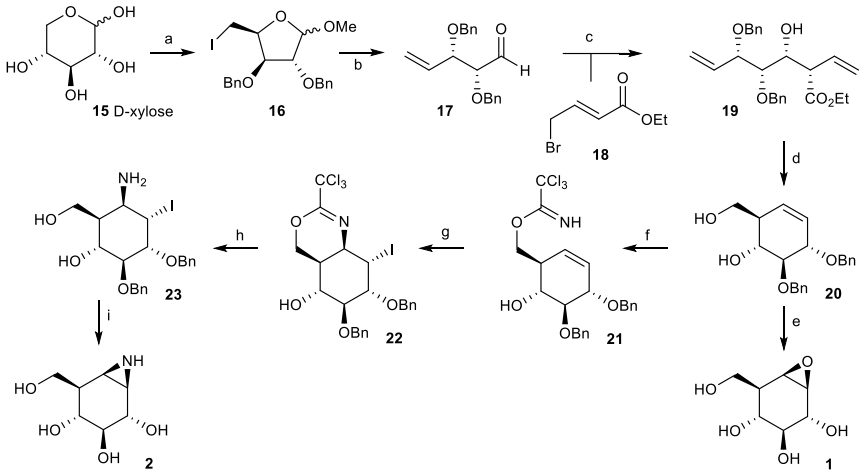
In a similar fashion (though with altered conditions at various stages), Tatsuta and co-workers synthesized 1,6-*epi*-cyclophellitol **14**[17] starting from D-galactopyranose derivative **13** (a configurational isomer of D-idose **3** with the same protective group pattern). The strategy to open the epoxide with sodium azide yielded a mixture of azido-alcohols, which were both transformed into the epimeric (with respect to the epoxide) aziridine using Staudinger conditions, proved to also be effective in the synthesis of cyclophellitol aziridine **2** from **14** in comparable yields to that of the preparation of aziridine **12**. The general strategy – installation of an aziridine in a two-step sequence (epoxide opening with nucleophilic azide followed by Staudinger reduction/cyclisation with inversion of configuration) – features in a number of subsequent syntheses as is described further on in this review.

2.2.2 Synthesis of cyclophellitol aziridines 2 and 12: Ring-closing metathesis and intramolecular iodo-imination as the key steps

The years following the pioneering synthesis studies of Tatsuta's group witnessed the rising

impact of ring closing metathesis (RCM)[18] – the transformation of two terminal alkenes of an (acyclic) substrate into an internal alkene in a cyclic product – in organic chemistry. Suitable transition metal catalysts became available,[19] which is capable to produce small to medium-sized rings from dienes featuring numerous functional groups. RCM evolved to become a key step in the synthesis of functionalized heterocycles[20] and carbacycles[21], and is compatible with carbohydrate chemistry.[22] This holds true as well for the synthesis of cyclophellitol/cyclophellitol aziridine analogues, most of which are prepared nowadays through synthetic routes involving a RCM step.

Scheme 2. Ring-closing metathesis as a key step in the synthesis of compounds **1** and **2**.



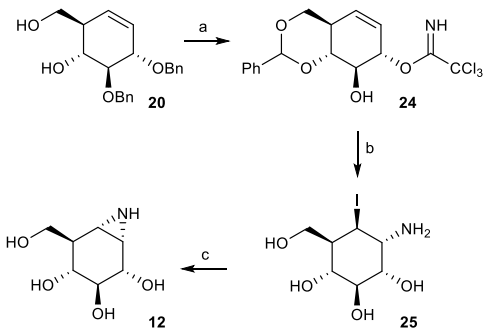
Reagents and conditions: (a) i) HCl, MeOH, 5 °C; ii) I_2, Ph_3P, imidazole, THF, 74%; iii) $BnOC(=NH)CCl_3$, TfOH, dioxane, 90%; (b) Zn, THF/H_2O, ultrasound, 78%; (c) indium powder, $La(OTf)_3$, H_2O, ultrasound, 80%; (d) i) Grubbs 2nd generation catalyst, DCM, 40 °C, 89%; ii) DIBAL-H, THF, 0 °C→RT; iii) $NaBH_4$, H_2O, EtOAc, 94%; (e) i) *m*CPBA, Na_2HPO_4 (aq., 1.0 M), NaH_2PO_4 (aq., 1.0 M), DCE, 50 °C, 55%; ii) H_2, $Pd(OH)_2$, MeOH, 88%; (f) CCl_3CN, DBU, DCM, 0 °C; (g) I_2, $NaHCO_3$, H_2O; (h) HCl (37%, aq.), dioxane, 60 °C; (i) i) $NaHCO_3$, MeOH, 60% over four steps; ii) Li, NH_3 (liq.), THF, -60 °C, 70%.

Madsen and co-workers[23] prepared cyclophellitol (**1**) via cyclohexene intermediate **20**, itself prepared through RCM on the appropriate diene **19** (Scheme 2). In this route (optimized by us recently),[24] D-xylose is converted in three steps into iodofuranoside **16**. First, kinnetic Fischer glycosylation of D-xylose in methanol provided methyl xylofuranoside. Next the iodine was installed using conditions developed by Garegg and Samuelsson[25] after which benzylation under acidic conditions of the two remaining secondary hydroxyls) gave **16**. Vasella fragmentation of **16** with zinc dust under sonication gave aldehyde **17**. Indium-mediated Barbier reaction of **17** with ethyl 4-bromocrotonate **18** yielded diene **19** in good yield and excellent stereochemistry – much better, as stated by the authors[23], than those found when performing the Barbier

allylation with 4-bromobut-2-en-1-ol instead of crotonate **18** (which, if successful, would have obviated the reduction of the methyl ester to the corresponding primary hydroxyl in a later stage). RCM of diene **19** with Grubbs second-generation ruthenium alkylidene catalyst[26] and ensuing reduction of the methyl ester to the primary alcohol (treatment with DIBAL-H, followed by sodium borohydride reduction of the intermediate aldehyde) gave cyclohexene **20** in good yield. Madsen and colleagues continued their synthesis to cyclophellitol (**1**) by capitalizing on the homoallylic alcohol in **20** for stereospecific introduction of the epoxide. Therefore, treatment of **20** with *meta*-chloroperbenzoic acid (*m*CPBA) in methylene chloride followed by catalytic hydrogenation yielded cyclophellitol (**1**) in ten steps starting from D-xylose (**15**).

The homoallylic alcohol embedded in partially protected cyclohexene **20** proved also ideal for the stereospecific introduction of an amine functionality, as demonstrated in previous synthesis of cyclophellitol aziridine (**2**).[24] Reaction of **20** with trichloroacetonitrile and 1,8-diazabicycloundec-7-ene (DBU) as base gave trichloroacetimidate **21**. Iodocyclisation (with iodine and sodium bicarbonate) followed by acidolysis of the resulting cyclic imidate **22**, yielded stereospecifically *trans*-1-iodo-2-amine **23**. Under mild basic conditions the iodine in **23** is displaced to form the aziridine – again in a stereospecific fashion – after which Birch reduction (lithium in liquid ammonia) gave cyclophellitol aziridine in five steps from the Madsen cyclohexene **20**.

Scheme 3. Transformation of common intermediate **19** into 1,6-*epi*-cyclophellitol aziridine **11**.



Reagents and conditions: (a) i) Li, NH_3 (liq.), THF, -60 °C, 57%; ii) $PhCH(OMe)_2$, CSA, DMF, 61%; iii) CCl_3CN, DBU, DCM, 0 °C; (b) i) $NaHCO_3$, I_2, H_2O, 41% over two steps; ii) HCl (37%, aq.), dioxane; (c) $NaHCO_3$, MeOH, 63% over two steps.

The accessible synthesis of cyclohexene **20**, which can be performed to yield multi-gram quantities, combined with the control in stereochemical outcome exerted by the intramolecular iodo-imidation/iodine displacement sequence, led to the adaptation of the synthetic route of cyclophellitol aziridine (**2**) (Scheme 2) towards 1,6-*epi*-cyclophellitol aziridine (**12**).[27] As outlined in Scheme 3, removal of the two benzyl ethers in **20** under Birch conditions was followed by

selective installation of the benzylidene and selective transformation of the allylic alcohol (as opposed to the homo-allylic alcohol) into the trichloroacetimidate, yielding (labile) intermediate **24**. Treatment of **24** with iodine now results in iodo-imidation with delivery of the nitrogen at C1 from the 'alpha' face (in terms of the parent D-glucopyranoside), as compared to C6-beta-delivery of the nitrogen seen in the transformation of **21** into **22** (Scheme 2). Following the same sequence of events, acidolysis of the cyclic imidate yielded *trans*-iodo-amine **25**, which under mild basic conditions gave the desired 1,6-*epi*-cyclophellitol aziridine **12**.

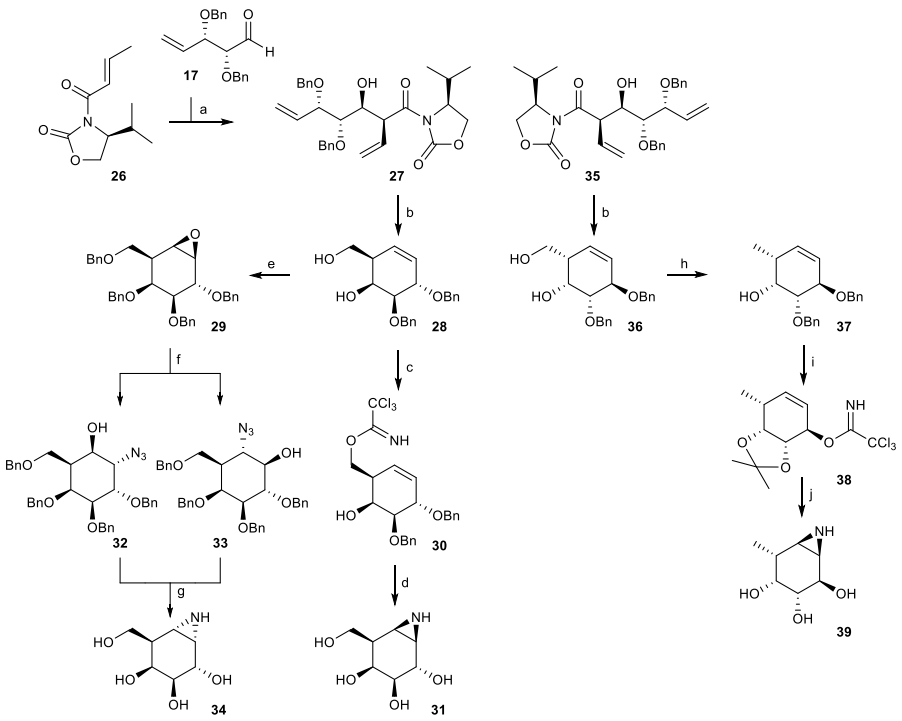
2.2.3 Synthesis of configurational and functional cyclophellitol aziridine isomers: iodo-imination/substitution versus epoxide opening/Staudinger cyclisation

In recent years syntheses of a number of cyclophellitol aziridines differing in configuration and/or substitution pattern from the glucopyranose configured compounds **1** and **14** have appeared in the literature. These syntheses share a number of features with the strategies outlined above. They all use chiral pool building blocks as starting material, RCM may feature as a key step and the aziridine moiety is introduced from either an epoxide precursor (with full stereocontrol and retention of configuration – stereocontrol in epoxide formation is not always complete though) or through intramolecular iodocyclisation from a partially protected (homo)-allylic alcohol precursor.

The synthesis of 4-*epi*-cyclophellitol aziridine **31** (a configurational analogue of β-galactopyranose) and 1,4,6-*epi*-cyclophellitol aziridine **34** (a configurational analogue of α-galactopyranose) starts with a dibutylboryl triflate-catalyzed stereoselective aldol condensation of aldehyde **17** and Evans' oxazolidinone **26** (Scheme 4), in a procedure developed by Llebaria and co-workers.[28] Reduction of the amide in **27** to the primary alcohol (with concomitant removal of the chiral Evans' auxiliary) followed by RCM yielded partially protected cyclohexene **28**, which can be regarded as the galactopyranose equivalent of common building block **20** that was used to synthesize both cyclophellitol aziridine **2** (Scheme 2) and 1,6-*epi*-cyclophellitol aziridine (**12**) (Scheme 1). Following the synthetic schemes as outlined for compounds **2** and **12** indeed yielded 5-*epi*-cyclophellitol aziridine (**31**) and 1,2,5-*epi*-cyclophellitol (**34**).[29] The synthesis of compound **31** proceeded with an efficiency equal to that observed in the synthesis of cyclophellitol aziridine (**2**) and with absolute stereocontrol as offered by the intramolecular iodo-imination/intramolecular substitution protocol. The synthesis of compound **34** required the preparation of 4-*epi*-cyclophellitol **29**, which was accomplished after perbenzylation of **28** followed by epoxidation with *m*CPBA, which proceeded with remarkable stereoselectivity given that no (homo)-allylic alcohol is present in the precursor to guide the epoxidation. In their original publication[28] on the synthesis of **28**, Llebaria and co-workers produced the perbenzylated, *galacto*-configured cyclophellitol **29** through stereoselective dihydroxylation of cyclohexene **28** followed by reaction with excess of the Mattocks-Moffatt reagent,[30] 2-acetoxyisobutyryl bromide. The resulting crude mixture of *trans*-cyclohexane bromoacetates

was treated with potassium carbonate to give compound **29** in good yield as well.

Scheme 4. Synthesis of *galacto*se and *fucose*-configured cyclophellitol aziridines.

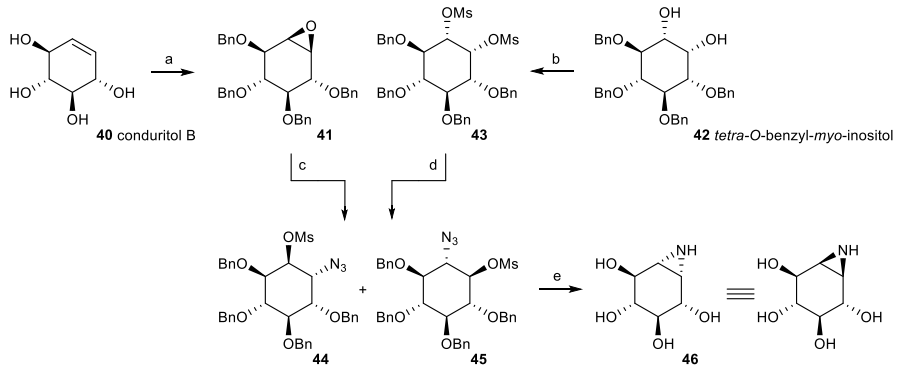


Reagents and conditions: (a) $Bu_2BSO_3CF_3$, Et_3N, DCM, -78 °C to -20 °C, 80%; (b) i) $LiBH_4$, THF/H_2O, 0 °C→RT, 85%; ii) second-generation Grubbs catalyst, DCM, 40 °C, 78%; (c) CCl_3CN, DBU, DCM, 0 °C; (d) i) I_2, $NaHCO_3$, H_2O; ii) HCl (37% aq.), MeOH, iii) HCl (37% aq.), dioxane, 60 °C; iv) $NaHCO_3$, MeOH, 60% over four steps), (d) Li, NH_3 (liq.), THF, -60 °C; (e) i) NaH, BnBr, TBAI, DMF, 85%; ii) *m*CPBA, DCM, 63%; (f) NaN_3, $LiClO_4$, MeCN, 80 °C, 73% (**32**:**33** = 1:1.3); (g) i) Ph_3P, MeCN, 80 °C (26%); ii) Li, NH_3 (liq.), THF, -60 °C; (h) i) TsCl, Et_3N, DCM, 87%; ii) $LiAlH_4$, THF, 0 °C→RT, 87%; (i) i) Li, NH_3 (liq.), THF, -60 °C, 73%; ii) 2,2-dimethoxypropane, CSA, 60%; iii) CCl_3CN, DBU, DCM; (j) i) CCl_3CN, DBU, DCM; ii) $NaHCO_3$, I_2, H_2O, 46%; iii) HCl (37% aq.), MeOH, 60 °C; iv) $NaHCO_3$, MeOH, 65% over four steps.

One attractive feature of chiral pool material in synthesis is that – providing that the enantiomer is available and affordable – subjecting this enantiomer to the same sequence of events will yield the mirror image products. Both aldehyde **17** and oxazolidinone **26** are readily available in enantiomeric form and their condensation following the Mukayama-aldol procedure developed by Llebaria and co-workers[28] gave access to diene **35**, being the mirror image of **27**. Processing **35** similar to **27** would yield a set of L-galactose-configured cyclophellitol aziridines. The 6-deoxy analogue of L-fucose is often encountered – α-linked – in naturally occurring glycoconjugates. With the aim to develop covalent and irreversible α-fucosidase inhibitors, we synthesized 8-

deoxy-2,3,5-*epi*-cyclophellitol aziridine (**39**) following a route related to the one depicted in Scheme 3.[29] Following RCM (**35** to **36**), the primary alcohol was reduced to the methyl group (**36** to **37**) after which the debenzylation, protection of the *cis*-diol and reaction of the remaining allylic alcohol with trichloroacetonitrile yielded imidate **38**. Iodo-imination, acidolysis and base-catalyzed intramolecular iodide substitution provided aziridine (**39**) – a configurational analogue of α-L-fucopyranose. The above schemes represent all configurational and functional cyclophellitol aziridines (namely, **1**, **12**, **31**, **34** and **39**) whose synthesis has been reported to date. Studies on the synthesis of related compounds exists, however, both targeting cyclopentitol aziridines[32] (not shown here) and in particular conduritol aziridines.

Scheme 5. Syntheses of conduritol B aziridine.



Reagents and conditions: (a) i) *m*CPBA, DCM; ii) NaH, BnBr, DMF (on racemic conduritol, for enantiopure **40** see[33]); (b) MsCl, pyridine; (c) i) NaN_3, MeCN, 2 N $LiClO_4$, 91%; ii) MsCl, Et_3N, THF, only **45**: 88%; (d) NaN_3, DMF, 42% over two steps; (e) i) $LiAlH_4$, Et_2O, 0 °C, 60%; ii) Na, NH_3 (liq.), THF, -78 °C, 86%.[34]

Scheme 5 depicts two recent and representative syntheses of the broad-spectrum retaining glucosidase inhibitor, conduritol B aziridine (**46**) (both α- and β-glucosidase).[35] In the first route,[36] the natural product, conduritol B (**40**) is perbenzylated followed by epoxidation to afford fully protected conduritol B epoxide **41**. Opening of the epoxide with azide, mesylation of the resulting secondary hydroxyl, Staudinger ring-closure and final debenzylation provided conduritol B aziridine (**46**). In an alternative route, the mixture of azides **44** and **45** was prepared from another natural product, *myo*-inositol (**42**)[37] protected as the *tetra-O*-benzyl derivative, via a number of standard protective group and functional group manipulations.

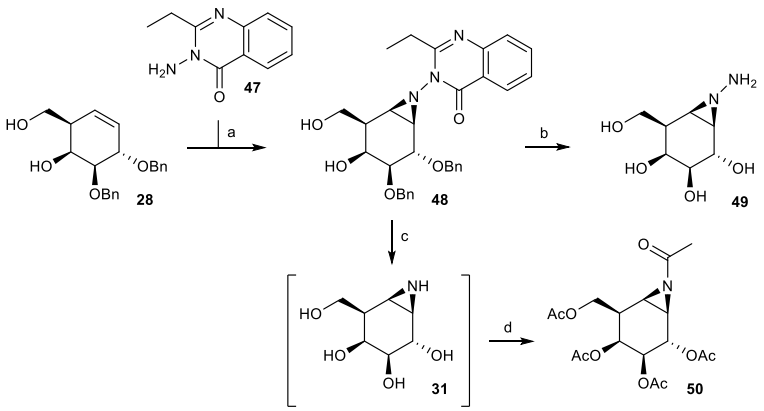
2.3 Discussion

Cyclitol epoxides were widely used in the sixties of the past century as mechanism-based, irreversible glycosidase inhibitors. Conduritol B epoxide (**3**, CBE), discovered and exploited by Legler and co-workers[4], was the glucosidase inhibitor of choice up until the discovery in 1967 of

the polyhydroxylated alkaloid, deoxynojirimycin.[7] In the following decades, cyclitol epoxides and cyclitol aziridines received relatively little attention. In 1989, the Withers group[35] proposed that a CBE homologue bearing an extra methylene would fit better within a glucosidase active site and suggested that such a compound should be synthesized. Shortly thereafter, cyclophellitol (**1**) was discovered[1] as a natural product and shown to be a potent and highly selective inhibitor of retaining β-glycosidases. These discoveries led to some renewed interest in the cyclitol epoxide/aziridine compound class, and several studies on the synthesis of **1** and evaluation as enzyme inhibitors and its functional and configurational analogues were reported.

In recent years, and in conjunction with a general rise in interest in covalent, irreversible inhibitors as starting point for the development of ABPP methodology,[8] cyclophellitol (**1**), cyclophellitol aziridine (**2**) and analogous structures are receiving renewed attention. To date, ABPs have been reported that enable selective profiling of retaining glycosidases such as β-glucosidases,[37] α-glucosidases,[27] α-galactosidases[29b] and α-fucosidases,[31] which are designed based on compound **2** bearing *N*-substituents featuring a fluorophore or biotin as a reporter entity. Cyclophellitol aziridine (**2**) has proven to be a superior scaffold compared to 2 (or 5)-deoxy-5-fluoroglycosides[37] for *in vitro* and *in situ* ABPP of retaining β-glucosidase activities[40] and adaptation of the configuration and substitution pattern will likely yield selective ABPs for retaining glycosidase families evolved to recognize and hydrolyze the underlying configurational carbohydrates – next to monosaccharides (*exo*-glycosidases[41]) likely also oligosaccharides (*endo*-glycosidases[42]). To fulfill this promise, though, synthetic methodology needs to be expanded to enable easy access to an array of configurational and functional cyclophellitol/cyclophellitol aziridine analogues. The aziridine moiety in all the syntheses described in this review is installed either by modification of an epoxide precursor or through iodocyclization starting from a (homo) allylic alcohol precursor. Expansion of methodology that enables the introduction of an aziridine[43] at various stages of the synthesis is important to access to a wide array of cyclophellitol aziridine analogues. Llebaria and co-workers recently reported[44] direct aziridination of *galacto*-configured cyclohexene **28** (Scheme 6). Although the objective of this synthetic study was to obtain *N*-amino-aziridine **49**, the authors showed that reduction of the N-N bond yields 5-*epi*-cyclophellitol aziridine (**31**) (the structure of which was established after peracetylation to **50**). Direct aziridination – either by reaction with aminoquinazolinones[44] or by means of other recently published[45] methodology – will likely evolve to become a complementary method for the synthesis of cyclophellitol aziridine analogues, thus expanding the chemical toolbox of covalent, irreversible glycosidase inhibitors and activity-based glycosidase probes derived thereof.

Scheme 6. Direct aziridination of protected cyclohexene **28**.



Reagents and conditions: (a) $PhI(OAc)_2$, K_2CO_3, DCM, 54%; (b) i) $H_2N\text{-}NH_2$, 120 °C, 78%; ii) Na, NH_3 (liq.), THF, -78 °C, 91%; (c) i) Na, NH_3 (liq.), THF, -78 °C; ii) Ac_2O, pyridine, 28% over two steps).

2.4 References

[1] S. Atsumi, K. Umezawa, H. Iinuma, H. Naganawa, H. Nakamura, Y. Iitaka, T. Takeuchi, *J. Antibiotics* **1990**, *43*, 49-53.

[2] S. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima, M. Nakata, *J. Antibiotics* **1991**, *44*, 912-914.

[3] D. Koshland, *Biol. Rev.* **1953**, *28*, 416-436.

[4] a) G. Legler, *Hoppe-Seyler's Z. Physiol. Chem.* **1966**, *345*, 197-214; b) G. Legler, *Hoppe-Seyler's Z. Physiol. Chem.* **1968**, *349*, 767-774.

[5] G. J. Davies, A. Planas, C. Rovira, *Acc. Chem. Res.* **2012**, *45*, 308-316.

[6] See for some reviews on the mechanism of glycosyl hydrolases: a) D. J. Vocadlo, G. J. Davies, *Curr. Opin. Chem. Biol*. **2008**, *12*, 539-555; b) G. Davies, B. Henrissat, *Structure* **1995**, *15*, 853-859.

[7] a) H. Paulsen, *Angew. Chem. Int. Ed.* **1966**, *5*, 495-510; b) M. Yagi, T. Koumo, Y. Aoyagi, H. Murai, *Nippon Nogei Kagaku Kaishi* **1976**, *50*, 571-572; c) Iminosugars as glycosidase inhibitors: nojirimycin and beyond, A. E. Stütz, editor, Wiley VCH, **1999**.

[8] a) Y. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14694-14699; b) D. Greenbaum, K. F. Medzihradsky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, *7*, 569-581; c) A. Borodovsky, B. M. Kessler, R. Casagrande, H. S. Overkleeft, K. D. Wilkinson, H. L. Ploegh, *EMBO J.* **2001**, *20*, 5187-5196.

[9] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, B. Blijlevens, G. Kramer, A. M. C. H. van den Nieuwendijk, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft, J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907-913.

[10] L. I. Willems, J. Jiang, K.-Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. N. Beenakker, S. P. Schröder, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée, H. S. Overkleeft, *Chem. Eur. J.* **2014**, *20*, 10864-10872.

[11] a) See the CAZypedia website on carbohydrate-active enzymes: http://www.cazypedia.org/index.php/Glycoside_hydrolases; b) See for a description of the CAZypedia website: V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014,** *42,* D490-495.

[12] K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima, M. Nakata, *Tetrahedron Lett.* **1990**, *31*, 1171-1172.

[13] K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima, M. Nakata, *J. Antibiotics* **1991**, *44*, 456-458.

[14] a) K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima, M. Nakata, *Carbohydr. Res.* **1991**, *222*, 189-203; b) See for a review on cyclophellitol synthesis: b) J. Marco-Contelles, *Eur. J. Org. Chem*. **2001**, 1607-1618.

[15] a) D. Semeria, M. Philippe, J.-M. Delaumeny, A.-M. Sepulchre, S. D. Gero, *Synthesis* **1983**, 710-713; b) A. Lipták, I. Jodal, P. Násáni, *Carbohydr. Res.* **1975**, *44*, 1-11.

[16] a) P. Pöchlauer, E. P. Müller, *Helv. Chim. Acta* **1984**, *67*, 1238-1247; b) B. Ritzen, M. C. M. van Oers, F. L. van Delft, F. P. J. T. Rutjes, *J. Org. Chem.* **2009**, *74*, 7548-7551.

[17] a) See for a review of the Tatsuta laboratory on the synthesis of cyclophellitols: K. Tsatuta, *Pure Appl. Chem.* **1996**, *68*, 1341-1446; b) see also references 2 and 13.

[18] a) R. R. Schrock, J. S. Murdzek, G. C. Bazan, J. Robbins, M. DiMare, M. O'Regan, *J. Am. Chem. Soc.* **1990**, *112*, 3875-3886; b) G. C. Fu, R. H. Grubbs, *J. Am. Chem. Soc.* **1992**, *114*, 7324-7325.

[19] a) S. Nguyen, R. H. Grubbs, *J. Am. Chem. Soc*. **1993**, *115*, 3800-3801; b) P. Schwab, R. H. Grubbs, J. W. Ziller, *J. Am. Chem. Soc.* **1996**, *118*, 100-110.

[20] See for example a) M. F. Schneider, H. Junga, S. Blechert, *Tetrahedron* **1995**, *51*, 13003-13014; b) H. S. Overkleeft, U. K. Pandit, *Tetrahedron Lett*. **1996**, *37*, 547-550; c) see for a review: R. H. Grubs, S. Chang, *Tetrahedron* **1998**, *54*, 4413-4450.

[21] See for example a) H. Ovaa, J. D. C. Codée, B. Lastdrager, H. S. Overkleeft, G. A. van der Marel, J. H. van Boom, *Tetrahedron Lett.* **1999**, *40*, 5063-5066; b) P. Kapferer, F. Sarabia, A. Vasella, *Helv. Chim. Acta* **1999**, *82*, 645-656.

[22] M. Jorgensen, P. Hadwiger, R. Madsen, A. E. Stütz, T. M. Wrodnigg, *Curr. Org. Chem.* **2000**, *4*, 565-588.

[23] F. G. Hansen, E. Bundgaard, R. Madsen, *J. Org. Chem.* **2005**, *70*, 10139-10142.

[24] K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel, H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, 6030-6043.

[25] P. J. Garegg, B. Samuelsson, *J. Chem. Soc., Perkin Trans.* **1980**, *1*,2866-2869.

[26] M. Scholl, S. Ding, C. W. Lee, R. H. Grubbs, *Org. Lett.* **1999**, *1*, 953-956.

[27] J. Jiang, C.-L. Kuo, L. Wu, C. Franke, W. W. Kallemeijn, B. I. Florea, E. van Meel, G. A. van der Marel, J. D. C. Codée, R. G. Boot, G. J. Davies, H. S. Overkleeft, J. M. F. G. Aerts, *ACS Central Sci.* **2016**, doi:10.1021/acscentsci.6b00057.

[28] Y. Harrak, C. M. Barra, A. Delgado, A. R. Castano, A. Llebaria, *J. Am. Chem. Soc.* **2011**, *133*, 12079-12084.

[29] a) L. I. Willems, T. J. M. Beenakker, B. Murray, B. Gadestein, H. van den Elst, E. R. van Rijssel, J. D. C. Codée, W. W. Kallemeijn, J. M. F. G. Aerts, G. A. van der Marel, H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, 6044-6056; b) L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts, H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, *136*, 11622-11625.

[30] a) A. R. Mattocks, *J. Chem. Soc.* **1964**, 4840-4845; b) A. F. Russell, S. Greenberg, J. G. Moffatt, *J. Am. Chem. Soc.* **1973**, *95*, 4025-4030.

[31] J. Jiang, W. W. Kallemeijn, D. W. Wright, A. M. C. H. van den Nieuwendijk, V. C. Rohde, E. C. Folch, H. van den Elst, B. I. Florea, S. Scheij, W. E. Donker-Koopman, M. Verhoek, N. Li, M. Schurmann, D. Mink, R. G. Boot, J. D. C. Codée, G. A. van der Marel, G. J. Davies, J. M. F. G. Aerts, H. S. Overkleeft, *Chem. Sci*. **2015**, *6*, 2782-2789.

[32] O. L. Lopez, J. G. Fernandez-Bolanos, V. H. Lillelund and M. Bols, *Org. Biomol. Chem.* **2003**, *1*, 478-482.

[33]a) C. Jaramillo, J.-L. Chiara, M. Martín-Lomas, *J. Org. Chem.* **1994**, *59*, 3135-3141; b) P. Serrano, A. Llebaria, A. Delgado, *J. Org. Chem.* **2002**, *67*, 7165-7167.

[34] Though the final Birch reduction is not described in the referenced paper, this procedure does yield aziridine **46** as we have found (not published) in the stated yield. Alternatively, compound **45** features also as an intermediate in the first synthesis of **46** (see ref 35), where it is stated that **46** can be produced from **45** by catalytic hydrogenation.

[35] G. Caron, S. G. Withers, *Biochem. Biophys. Res. Commun.* **1989**, *163*, 495-499.

[36] P. Serrano, A. Llebaria, A. Delgado, *J. Org. Chem.* **2005**, *70*, 7829-7840.

[37] B. T. Adams, S.Niccoli, M. A. Chowdhury, A. N. K. Esarik, S. J. Lees, B. P. Rempel, C. P. Phenix, *Chem. Commun.* **2015**, *51*, 11390-11393.

[38] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts, H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529-12533.

[39] a) S. G. Withers, I. P. Street, P. Bird, D. H. Dolphin, *J. Am. Chem. Soc.* **1987**, *109*, 7530-7531; b) S. G. Withers, K. Rupitz, I. P. Street, *J. Biol. Chem.* **1988**, *263*, 7929-7932; c) see for a review on mechanism-based glycosidase inhibitors, including fluoroglycosides: B. P. Rempel, S. G. Withers, *Glycobiology* **2008**, *18*, 570-586.

[40] a) M. T. C. Walvoort, W. W. Kallemeijn, L. I. Willems, M. D. Witte, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée, H. S. Overkleeft, *Chem. Commun.* **2012**, *48*, 10386-10388; See for studies on the use of fluoroglysosides in activity-based glycosidase profiling studies: b) D. J. Vocadlo, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2004**, *43* 5338-5342; c) O. Hekmat, Y.-W. Kim, S. J. Williams, S. He, S. G. Withers, *J. Biol. Chem.* **2005**, *280*, 35126-35135; d) K. A. Stubbs, A. Scaffidi, A. W. Debowski, B. L. Mark, R. V. Stick, D. J. Vocadlo, *J. Am. Chem. Soc.* **2008**, *130*, 327-335

[41] See for an example of a 6-phospho-cyclophellitol-sensitive enzyme that would be amenable for 6-phospho-cyclophellitol-aziridine based ABPP experiments: D. H. Kwan, Y. Jin, J. Jiang, H. M. Chen, M. P. Kötzler, H. S. Overkleeft, G. Davies, S. G. Withers, *FEBS Lett.* **2016**, *590*, 461-468.

[42] See for an example of a glucosyl 1,6-*epi*-cyclophellitol-sensitive endoglucosidase that would be amenable for glucosyl 1,6-*epi*-cyclophellitol aziridine based ABPP experiments: S. Caner, X. Zhang, J. Jiang, H.-M. Chen, N. T. Nguyen, H. Overkleeft, G. D. Brayer, S. G. Withers, *FEBS Lett.* **2016**, *590*, 1143-1151.

[43] See for recent reviews on aziridine synthesis: a) H. Pellissier, *Tetrahedron* **2010**, *66*, 1509-1555; b) L. Degennaro, P. Trinchera, R. Luisi, *Chem. Rev.* **2014**, *114*, 7881-7929.

[44] A. Alcaide, A. Trapero, Y. Pérez, A. Llebaria, *Org. Biomol. Chem.* **2015**, *13*, 5690-5697.

[45] a) J. L. Jat, M. P. Paudyal, H. Gao, Q.-L. Xu, M. Youssufuddin, D. Devarajan, D. H. Ess, L. Kürti, J. R. Falck, *Science* **2014**, *343*, 61-65; b) S. O. Scholz, E. P. Farney, S. Kim, D. M. Bates, T. P. Yoon, *Angew. Chem. Int. Ed.* **2016**, *55*, 2239-2242.

3

In vitro and *in vivo* comparative and competitiveactivity-based protein profiling of GH29 α-L-Fucosidases

Jianbing Jiang, Wouter W. Kallemeijn, Daniel W. Wright, Adrianus M. C. H. van den Nieuwendijk, Veronica Coco Rohde, Elisa Colomina Folch, Hans van den Elst, Bogdan I. Florea, Saskia Scheij, Wilma E. Donker-Koopman, Marri Verhoek, Nan Li, Martin Schürmann, Daniel Mink, Rolf G. Boot, Jeroen D. C. Codée, Gijsbert A. van der Marel, Gideon J. Davies, Johannes M. F. G. Aerts and Herman S. Overkleeft, *Chemical Science*, **2015**, *6*, 2782-2789.

3.1 Introduction

GH29 α-L-fucosidases catalyze the hydrolysis of terminal α-L-fucosidic linkages.[1] The GH29[2] glycoside hydrolase family of retaining α-L-fucosidases contains members from various kingdoms of life including eukaryota[3] and bacteria (for example, *Bacteroides thetaiotaomicron*,[4] *Sulfolobus solfataricus*,[5] and *Thermotoga maritima*[6]). GH29 α-L-fucosidases process their substrate with overall retention of configuration at the anomeric center of the cleaved fucopyranose and do so through a double-displacement mechanism. In this mechanism, first proposed by Koshland,[7] (Figure 1A), S_N2 displacement of the aglycon (activated through protonation by the general acid/base residue) by nucleophilic attack of the catalytic nucleophile yields a fucosyl-enzyme intermediate, which is subsequently hydrolyzed to yield α-L-fucopyranose together with the released aglycon.

Two GH29 α-L-fucosidases are expressed in man. Of these, FUCA1 is found in lysosomes whereas FUCA2 is secreted into the plasma.[8] Deficiency of FUCA1 α-L-fucosidase activity causes fucosidosis,[9] a rare autosomal recessive lysosomal storage disorder. Next to its role in lysosomal turnover of fucosylated substrates, FUCA1 is also involved in sperm transport and sperm-egg interactions.[10] FUCA1 activity levels are considered to be a biomarker for cellular senescence[11] as well as for the diagnosis of hepatocellular cancers.[12] Deficiency of human FUCA2 has been shown to protect against *Helicobacter pylori* adhesion to gastric cancer cells.[8]

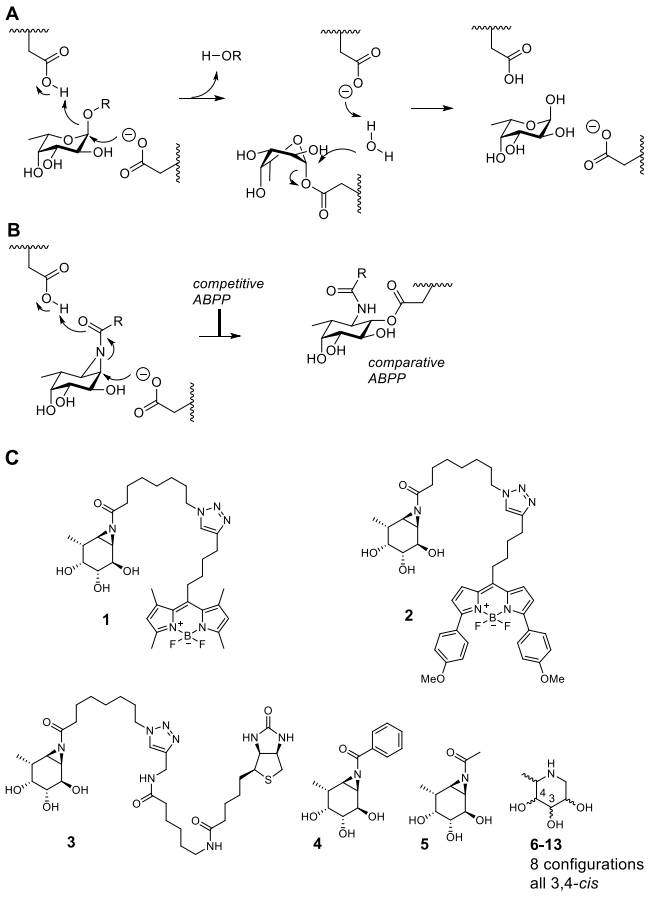


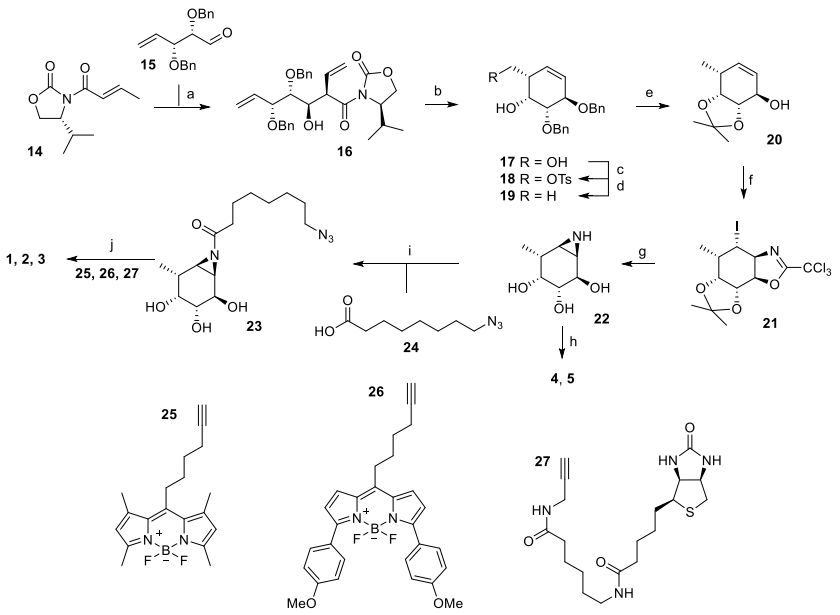
Figure 1. A) Double-displacement mechanism of retaining α-L-fucosidases. B) Comparative and competitive activity-based profiling of GH29 α-L-fucosidases presented here. C) Inhibitors and probes subject of this chapter.

The biological and biomedical relevance of GH29 retaining α-L-fucosidases warrant the development of efficient methods to monitor their functional state and activity *in vitro*, *in situ* and *in vivo*. In this respect, activity-based probes (ABPs) have shown their merit as tools to detect active enzyme molecules in their native environment (Figure1B).[13] It was previously

shown that cyclophellitol aziridines are useful scaffolds for the design of *in situ* and *in vivo* active ABPs directed at retaining β-glucosidases[14] and retaining α-galactosidases.[15] The specificity of these probes appeared due to their configuration, with the β-glucopyranose configured cyclitol aziridine being highly selective towards retaining β-glucosidases and their α-galacto-configured counterparts selective towards α-galactosidases.

In this chapter the development of retaining GH29 α-L-fucosidase ABPs is described. The ABPs are based on the cyclophellitol aziridine structure having the α-L-fucoside configuration and are equipped with a green (**1**, JJB256) or red (**2**, JJB244) BODIPY fluorophore or a biotin tag (**3**, JJB243) (Figure 1C). The probes turned out to be highly sensitive and selective and can be used for *in situ* and *in vivo* monitoring of mammalian and bacterial GH29 retaining α-L-fucosidases. ABPs **1** and **2** can also be used in a competitive activity-based protein profiling (ABPP) assay[16] to rapidly identify retaining α-L-fucosidase inhibitors from a library of eight configurational isomers of deoxy-L-fuconojirimycin (**6**-**13**); a library prepared specifically for this purpose. Finally the validity of the cyclophellitol aziridine design platform, for ABP development of retaining glycosidases, is established unambiguously by solving the crystal structure of a retaining α-L-fucosidase from *Bacteroides thetaiotaomicron* 2970, covalently bound to *N*-acyl cyclophellitol aziridines **4** and **5**.

Scheme 1. Synthesis of aziridine ABPs **1**, **2**, **3** and inhibitors **4**, **5**.



Reagents and conditions: (a) Bu_2BOTf (1.0 M in DCM), Et_3N, DCM, -78 °C, 71%; (b) i) $LiBH_4$, THF, 83%; ii) Grubbs 2nd generation, DCM, 95%; (c) *p*-TsCl, Et_3N, DCM, 87%; (d) $LiAlH_4$, THF, 0 °C to rt, 87%; (e) i) Li, NH_3 (liq.), THF, -60 °C, 73%, ii) 2,2-dimethoxypropane, CSA, 60%; (f) i) CCl_3CN, DBU, DCM; ii) $NaHCO_3$, I_2, H_2O, 46%; (g) i) 37% HCl (aq.), MeOH,

60 °C; ii) $NaHCO_3$, MeOH, 65%; (h) EEDQ, benzoic acid or acetic acid, DMF, 0 °C, **4**: 7% , **5**: 28%; (i) EEDQ, **24**, DMF, 0 °C 25%; (j) $CuSO_4$ (1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), DMF, **25** , **26** or **27**, **1**: 19% , **2**: 12%**, 3**: 13% .

3.2 Results and discussion

The synthesis of α-L-fucopyranose-configured cyclophellitol aziridine-based target compounds started with aldol condensation of aldehyde **15** and chiral acrylamide **14**, following the procedure reported[17] by Llebaria and co-workers for the enantiomer of **16** (Scheme 1). Reductive removal of the Evans template in **16** followed by ring-closing metathesis yielded, according to the Llebaria procedure,[17] partially protected L-galactopyranose-configured cyclohexene **17** in good yield. Tosylation of the primary alcohol in **17** was followed by hydride displacement of the tosylate to afford L-fucopyranose-configured cyclohexene **19**. The benzyl groups in **19** were reduced under Birch conditions, after which the *cis*-diol was protected as the isopropylidene acetal to give **20**. The secondary alcohol in **20** was transformed into the corresponding trichloroacetimidate, after which iodocyclisation yielded in a stereospecific fashion intermediate **21** analogous to the procedure reported for the synthesis of retaining β-glucosidase ABPs.[18] Acidic hydrolysis of both acetal and imino protections in **21** was followed by base-induced intramolecular nucleophilic substitution of the iodine to yield aziridine **22**. Acetylation or benzoylation of the aziridine in **22** under the agency of 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (EEDQ), following conditions developed previously,[18] yielded compounds **4** and **5**, respectively. EEDQ-induced acylation of **22** with 8-azidooctanoic acid[19] **24** provided azide **23**, which was transformed into target ABPs **1**, **2** and **3** by conjugation to BODIPY-alkynes **25**, **26** and biotin-alkyne **27**, respectively, via copper(I)-catalyzed Huisgen [2+3] cycloaddition. The final compounds were purified by reverse phase HPLC.

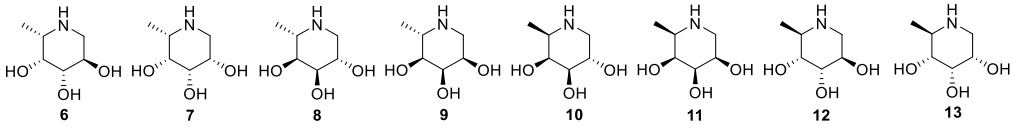
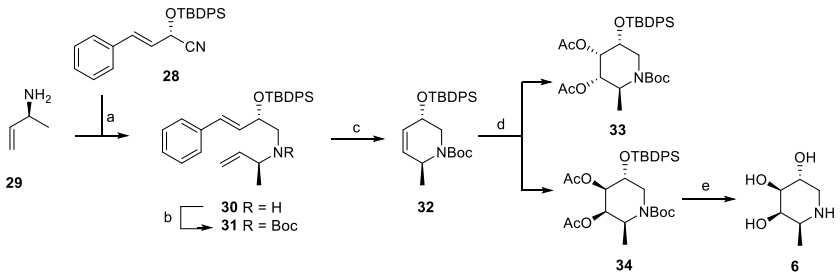


Figure 2. Structures of 1-deoxy-L-fuconojirimycin **6** and configurational isomers **7**-**13**.

The configurational fuconojirimycin isomers **6**-**13** (Figure 2) were synthesized following the strategy exemplified for 1-deoxy-L-fuconojirimycin **6** (Scheme 2). Key steps in the synthetic scheme include a DIBAL-H reduction-transimination-sodium borohydride reduction cascade of reactions involving enantiomerically pure cyanohydrin **28**, prepared employing (S)-hydroxynitrile lyase (*S*-*Hb*HNL) from *Hevea brasiliensis* rubber tree,[20] and the allylic amine **29** prepared by reported strategy,[21] to give secondary amine **30**. *N*-Boc protection (**30** to **31**), ring-closing metathesis (**31** to **32**) and Upjohn dihydroxylation afforded a mixture of *syn*-diols,

which were acetylated and separated by silica gel purification to yield diastereomers **33** and **34** in a 1:3 ratio.

Scheme 2. Synthesis of 1-deoxy-L-fuconojirimycin **6**.



Reagents and conditions: (a) i) Et_2O, DIBAL-H, -80 °C, ii) MeOH, -90 °C, iii) amine **29**, NaOMe, iv) $NaBH_4$,-15 °C to rt, 88%; (b) Boc_2O, 50 °C, 100%; (c) Grubbs 1st generation, DCM, 88%; (d) i) $K_2OSO_4{\cdot}2H_2O$, 4-methylmorpholine-N-oxide (NMO), acetone/H_2O (1:1, v/v), -10 °C, ii) Ac_2O, pyridine, DMAP, 0 °C, 88% (**33**:**34** = 1:3); (e) i) K_2CO_3, MeOH, ii) TBAF, THF, iii) HCl (6 M in H_2O), MeOH, 71%.

Global deprotection of **34** afforded 1-deoxy-L-fuconoijirimycin **6**, of which the analytical and spectroscopical data were in full agreement with those reported in the literature.[22] The seven configurational isomers **7**-**13** were prepared by alteration of the building blocks and/or the chemical transformations. It should be noted that four out of the eight isomers of fuconojirimycin were derived from **28**. To enable the synthesis of the four enantiomers, the enantiomer of cyanohydrin **28** was required, and this was prepared using almond (*R*)-hydroxynitrile lyase (*R*-*Pa*HNL).[23]

***In vitro* GH29 α-L-fucosidase activity assays**

Having the cyclophellitol aziridine inhibitors and probes in hand, their inhibitory potency towards the human lysosomal α-L-fucosidase, FUCA1, was determined. Inhibition potency was determined by measuring the residual enzyme activity using the fluorogenic substrate 4-methylumbelliferyl-α-L-fucopyranoside after pre-incubation of lysates of COS-7 cells over-expressing recombinant human FUCA1, with varying concentrations of the non-fluorescent, irreversible cyclophellitol aziridine inhibitors **4**, **5** and **23**; the ABPs **1** (JJB256), **2** (JJB244), **3** (JJB243), 1-deoxy-L-fuconojirimycin 6, and the seven fuconojirimycin isomers **7**-**13**. All *N*-acyl aziridines inhibited human FUCA1 activity with nanomolar IC_{50} activities (Table 1).

Table 1. *In vitro* and *in situ* inhibition of recombinant human GH29 α-L-fucosidase, given as half-maximal inhibitory concentration (IC_{50}). Data were averaged values of two separate experiments measured in duplicate, error ranges depict stand deviation.

Compounds	*In vitro* IC_{50} (nM)	*In situ* IC_{50} (nM)
1 (JJB256)	6.9 ± 0.8	28.9 ± 4.4
2 (JJB244)	8.7 ± 1.2	157.8 ± 22
3 (JJB243)	7.2 ± 1.1	25,512 ± 2,278
4	371.6 ± 21.2	515.4 ± 72.9
5	46.8 ± 3.4	77.9 ± 9.1
23	8.7 ± 1.1	65.4 ± 7.5
6	3,979 ± 257	ND
7 - 13	> 100,000	ND

ND: not determined. It should be noted that IC_{50} values on competitive inhibitors do not compare well to those obtained for mechanism-based inhibitors. The values given above allow for a comparison of inhibitory potency within the two classes of compounds studied.

Perusal of the inhibitory data does reveal some trends that allow some tentative conclusions. *N*-benzoyl aziridine **4**, bearing a bulky aromatic *N*-benzoate group, is about eight-fold less active compared to its *N*-acetylated counterpart, **5**. In contrast, comparing **5** with **23** reveals that a bulky aliphatic *N*-acyl substituent is in fact favored for enzyme inhibition (5-fold increase, Table 1). ABPs **1** and **2**, bearing a BODIPY fluorophore attached to the alkyl chain, inhibit FUCA1 in the same order as their azide precursor **23**. *In situ* inhibition of FUCA1 in living fibroblasts by ABP **1** and **2** occurs with similar efficacy as **23** and ABP **3**, the latter carrying a biotin attached to the alkyl chain. Exposure of cells to ABP **3** revealed in a dramatically decreased inhibitory potency, suggesting a reduced ability to penetrate into cells to reach the lysosomal FUCA1. The known competitive fucosidase inhibitor, 1-deoxy-L-fuconojirimycin **6** inhibits FUCA1 with an IC_{50} of 3.9 µM, in accordance with the literature values.[24] The seven configurational isomers **7**-**13** do not significantly inhibit FUCA1 activity up to 100 µM, a result that corroborates previous findings on some of the configurational analogues, which were reported as poor fucosidase inhibitors.[25]

As the next research objective, activity-based profiling of GH29 α-L-fucosidases from varying sources with green-fluorescent aziridine ABP **1** was examined, in the presence or absence of excess concentrations of, either the mechanism-based inhibitor **4** or the competitive inhibitor **6**. As is shown in Figure 3A, ABP **1** efficiently labels purified α-L-fucosidase from *Bacteroides thetaiotaomicron*. In lysates from an *E. coli* culture overexpressing recombinant α-L-fucosidase from *Bacteroides thetaiotaomicron* gene 2970, several fluorescent protein bands are visible upon labeling with **1** (Figure 3B), with the most prominent band at around 50 kDa,

corresponding to the predicted molecular weight of the enzyme. Labeling of the major band at 50 kD could moreover be blocked following pre-incubation with either 100 µM **4** or with 5 mM **6** (Figure 4B). Red fluorescent ABP **2** labels α-(1-2, 3, 4) and α-(1-6)-fucosidases from various bacterial sources in a similar fashion (Figure 3C).

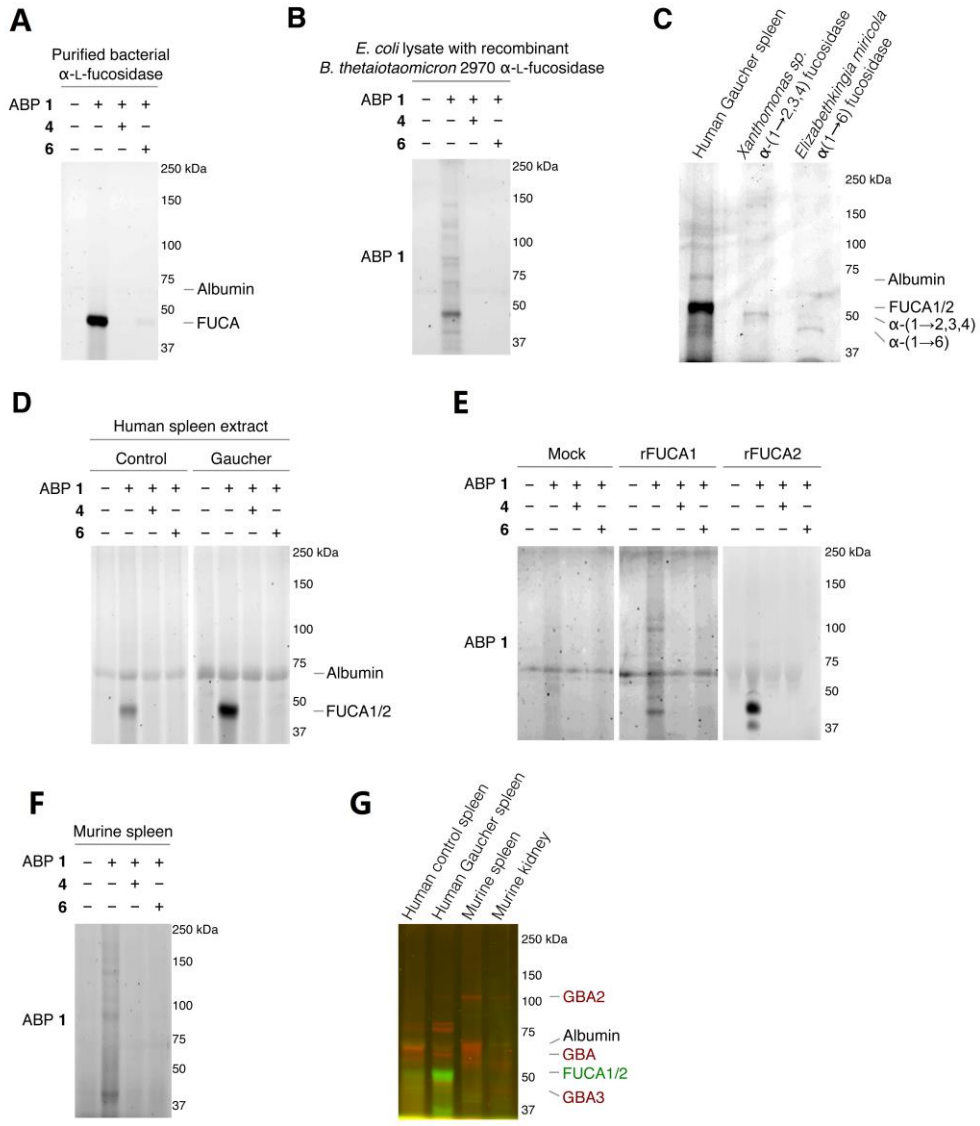


Figure 3. *In vitro* activity-based protein profiling of GH29 α-L-fucosidases. A), B) Labeling with ABP **1** of recombinant α-L-fucosidase and lysate of *E. coli* expressing recombinant α-L-fucosidase from *Bacteroides thetaiotaomicron* 2970. C) *In vitro* labeling of lysate of spleen from a Gaucher disease patient, α-(1-2,3,4)-fucosidase from *Xanthomonas sp.* and α-(1-6)-fucosidase from *Elizabethkingia miricola* with ABP **2**. D) *In vitro* labeling of human healthy and Gaucher disease spleen. E) ABP **1** labeling of α-L-fucosidases present in lysate of COS-7 cells transfected with empty plasmid (Mock) or

plasmid encoding human FUCA1 or FUCA2. F) Labeling of tissue homogenate of wild-type murine spleen with ABP **1**. G) Direct labeling of GH29 α-L-fucosidases with green-fluorescent ABP **1** and retaining β-glucosidases GBA, GBA2 and GBA3 with red-fluorescent JJB75.[18] The location of albumin autofluorescence is designated on each gel.

Subsequently, lysates of spleens from a healthy human individual and a patient suffering from Gaucher disease were exposed to ABP **1**. As can be seen in Figure 3D, a single fluorescent protein migrating slightly below 50 kDa was fluorescently labeled. The labeled band from human Gaucher spleen lysate is considerably more intense than the corresponding band in healthy human spleen, reflecting elevated α-L-fucosidase activity in the former. This result corroborates earlier observations that mRNA encoding lysosomal glycosidases are upregulated in Gaucher tissue.[26] As before, labeling with ABP **1** was suppressed after pre-treatment with either *N*-benzoyl aziridine **4** or L-fuconojirimycin **6**. To further ascertain that ABP **2** labels mammalian GH29 retaining α-L-fucosidases a number of control experiments were performed (Figure 3E). Compound **1** labeled both FUCA1 and FUCA2 that were overexpressed in COS cells. Treatment of murine spleen lysates with **1** yielded a result essentially as observed for healthy human tissue (Figure 3F).

Finally, simultaneous labeling of retaining β-glucosidases and retaining α-L-fucosidase was evaluated. To this end tissue lysates were incubated with **1** and the previously reported broad-spectrum activity-based retaining β-glucosidase probe, JJB75.[18] As can be seen (Figure 3G) the applied ABPs label a distinct set of proteins. Since they bear complementary fluorophores they can be used jointly to profile both retaining glycosidase families in a single experiment.

The FUCA1 enzymatic activity is maximal at around pH 5.0 (Figure 4A, B), consistent with the acidic pH of its natural lysosomal environment. Labeling efficiency with **1** largely reflects the pH dependence of FUCA1 at pH below 7. Of note, the pH dependence of labeling of FUCA1 does not follow that of its enzymatic activity. At alkaline conditions, where enzymatic activity is low, labeling still proceeds. Similar observations were made in the past for ABPP of retaining *exo*-β-glucosidases using cyclophellitol β-aziridine ABPs.[14] This result reveals the high reactivity of the aziridine probes in the initial displacement step employed by retaining glycosidases, a feature which, in conjunction with their selectivity in binding and relative stability in physiological environment, explains their efficacy as activity-based glycosidase probes.[27] Assessment of fluorescent labeling kinetics by employing 10 nM **1** labeling at stoichiometric concentrations of rhFUCA1 at 4 °C and 37 °C revealed that, during time-course experiments, examined on SDS-PAGE gels, labeling is near complete within the first minute at 4 °C (Figure 4C), impairing accurate determination of kinetic constants.

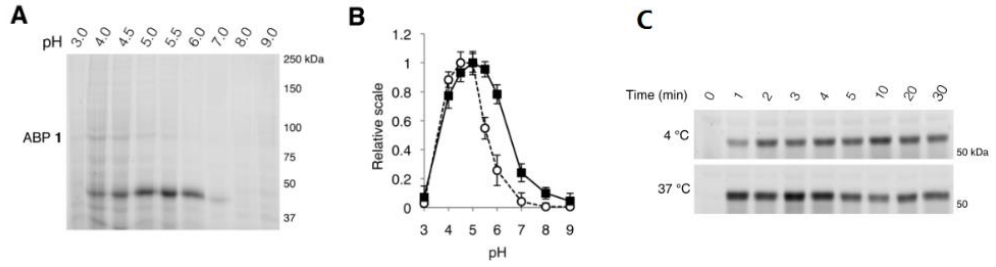


Figure 4. *In vitro* pH profile and labeling kinetics of recombinant FUCA1. A) *In vitro* labeling of COS-7 cell lysate containing over-expressed human FUCA1 at various pH with ABP **1**. B) Relative *in vitro* labeling of FUCA1 with ABP **1** (closed squares) compared to relative enzymatic activity towards artificial 4-methylumbelliferyl-α-L-fucopyranoside substrate (open circles) at various pH values. Data average of *n*=3 experiments, ± standard deviation. C) Stoichiometric *in situ* labeling of COS-7 lysate containing over-expressed human FUCA1 with 10 nM ABP **1** at 4 °C and 37 °C.

In vivo GH29 α-L-fucosidase assays

The ability of ABP **1** to label α-L-fucosidase in living mice was investigated next. Four wild-type C57Bl/6J male mice were injected with 100 µL phosphate-buffered saline (PBS) or PBS containing 10, 100 or 1000 pmol ABP **1**. After two hours, the mice were anesthetized, perfused with PBS and then brain, spleen, liver and kidney tissues were isolated. Tissue homogenates were prepared and each lysate was labeled prior to gel electrophoresis with red-fluorescent cyclophellitol β-aziridine JJB75, which labels *exo*-β-glucosidases as loading control (Figure 5). Furthermore, tissue homogenates of vehicle-treated animals were labeled with excess ABP **1** to visualize the maximal α-L-fucosidase labeling achievable in each tissue.

After treatment of mice with ABP **1**, a dose-dependent labeling of retaining α-L-fucosidases is observed in spleen, liver and kidney (Figure 5). Injection of 1000 pmol ABP **1** results in substantial labeling of α-fucosidase in spleen, liver and kidney. Detected fluorescence levels are comparable to that in matching samples from vehicle-treated mice incubated *in vitro* with excess **1**. In contrast, no *in vivo* brain FUCA1 labeling was observed after any of the administered doses of ABP **1**, and it can be concluded that ABP **1**, just as its β-glucose and α-galactose congeners,[14, 15] does not penetrate the brain.

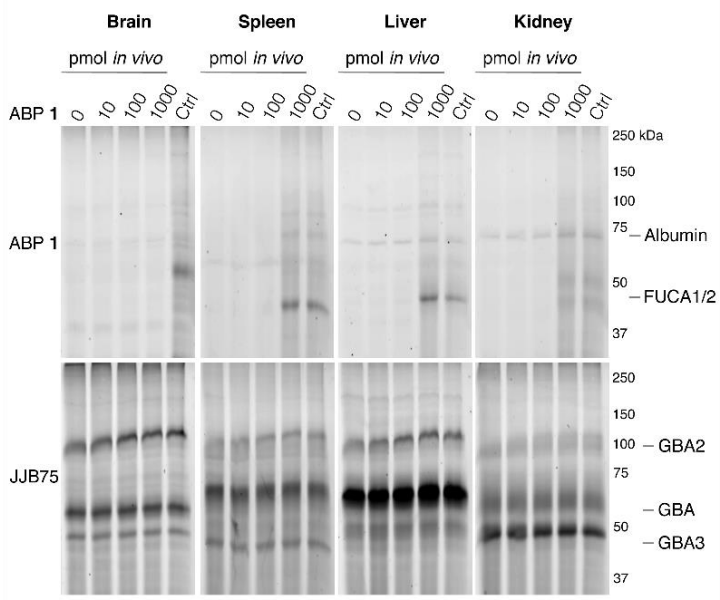


Figure 5. *In vivo* labeling of α-L-fucosidases in mice with various concentrations of ABP **1** during 2 hours. Top: *in vivo* labeling compared to maximal *in situ* labeling with excess ABP **1** of matched homogenates of untreated animals (Ctrl). Bottom: *in vitro* labeling of retaining β-glucosidases GBA, GBA2 and GBA3 with JJB75 in all homogenates as loading control.

Competitive activity-based GH29 α-L-fucosidase profiling

Having established the efficacy of ABP **2** to selectively label GH29 retaining α-L-fucosidases from various sources in an activity-based manner, the inhibition potential of deoxyfuconojirimycin **6** and its **7** stereoisomers **7**-**13** was determined in a competitive ABPP format (Figure 6). In contrast to **6**, none of the seven configurational isomers **7**-**13** were capable of blocking ABP **1** labeling of α-L-fucosidase, a result that matches with the data on inhibition of recombinant FUCA1 in the fluorogenic activity assay (Table 1).

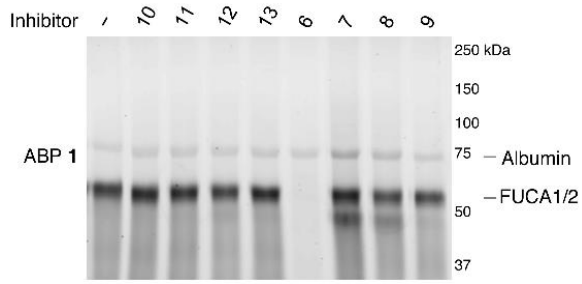


Figure 6. Competitive ABPP on recombinant FUCA1 with deoxyfuconojirimycin **6** and configurational analogues **7**-**13** towards α-L-fucosidases, with ABP **1** labeling as readout.

Identification of biotin labeled GH29 α-L-fucosidases

To further determine the specificity of the developed ABPs, ABP labeling of proteins was analyzed in complex tissue homogenates. For this purpose, Gaucher spleen lysate was incubated with biotinylated ABP **3**; with DMSO (control) or competitive ABP **3**-labeling by first incubating with fluorescent ABP **1**. Glycosylated proteins were then enriched via concanavilin A (ConA), followed by affinity purification with Streptavidin-coated paramagnetic beads. The identity of biotinylated and ABP **3**-labeled proteins was determined by on-bead digestion with trypsin, peptide analysis by LC-MS/MS and matching against the human UniProt database, using the Mascot search engine as previously reported.[14] FUCA1 was identified after ABP **3** pull-down as one of the top identified proteins (SI Table S1A) in the supplementary data), but was not found in the competitive (SI Table S1B) or untreated control (SI Table S1C). Proteins with higher scores were background proteins such as abundant endogenously biotinylated propionyl-CoA carboxylase alpha chain (PCCA), pyruvate carboxylase (PC), and keratin contaminations. FUCA1 was selectively found in the ABP **3**-pull down experiment only. Table 2 shows the analysis parameters of the identified peptides from FUCA1, with accuracy below 5 ppm, and Mascot ion scores above 40 (indicating reliable MS/MS fragment annotation and match) and manually curated fragmentation patterns. These results show FUCA1 can be undisputedly affinity purified and identified via biotinylated ABP **3**. Moreover the binding of ABP **3** can be completely blocked by pre-incubation with ABP **1**, which may indicate that both ABPs bind at the same site of the enzyme.

Table 2. Analysis parameters of peptides derived from the P04066 Tissue alpha-L-fucosidase, FUCA1 protein after affinity purification, on bead digest and LC-MS/MS analysis.

start-end	*m/z* obs.	z	ppm	Ion score	Sequence
114-130	982.4723	2	2	75	FFHPEEWADLFQAAGAK
163-173	572.3213	2	-1	58	DLVGELGTALR
392-420	1049.211	3	0	84	GSAVYAIFLHWPENGVLNLESPITTSTTK
439-461	828.1304	3	1	49	GLFISLPQLPPSAVPAEFAWTIK

Start-end gives the position of the identified peptide in the protein sequence, *m/z* observed is the measured *m/z* of the peptide, z is the charge, ppm is the measurement accuracy between the calculated and the observed *m/z*, ion score is the Mascot search engine score calculated for the match of the MS/MS fragmentation to the human protein database, sequence is the identified peptide sequence.

3-D crystal structure analysis of bacterial FUCA1 complexed to 4 and 5

In order to obtain experimental evidence for the formation of a covalent adduct between α-L-fucosidase and ABPs **1**, **2** and **3**, a crystal structure of *Bt*Fuc2970 (often used as a surrogate for the mammalian enzyme) with mechanism-based inhibitor **4** was obtained (PDB code: 4WSK). While the resulting crystal structure clearly demonstrates the formation of an

enzyme-**4** complex, electron density around the aryl group of the inhibitor "aglycon" moiety was close to the side-chain of the catalytic acid/base of the enzyme, residue E288 (Figure 7A), resulting in disorder. This likely reflects steric clashing and considerable conformational flexibility in the **4** aglycon when bound to *Bt*Fuc2970. “Aglycon” here reflects the aziridine *N*-acyl moiety to which the BODIPY tags are grafted in ABPs **1** and **2** and which is designed to occupy the space normally occupied by the substrate fucoside.

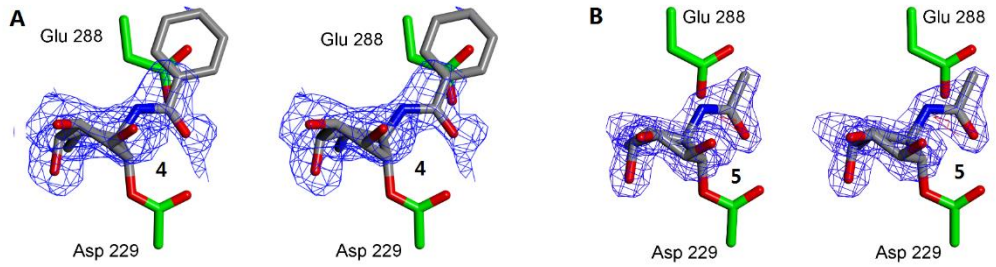


Figure 7. Crystal structures of α-L-fucosidase from *Bacteroides thetaiotaomicron* in complex with A) **4** and B) **5**. Catalytic residues are annotated: Asp 229 (nucleophile) and Glu 288 (acid/base). Electron density displayed is F_o-F_c density from phases calculated prior to the inclusion of **4** and **5** in refinement, contoured at 2σ and 3σ, respectively. Figure was prepared using CCP4MG[29]. Note that there is no electron density, at this level, for the aryl group of **4** presumably reflecting considerable disorder and/or steric clashes.

In the here-presented design of aziridine-based retaining *exo*-glycosidase ABPs it was assumed that such aglycon-like moieties would not interfere with enzyme binding (this also based on the numerous fluorogenic substrates that are in use to study *exo*-glycosidases and in which the aglycon moiety can take on made shapes and sizes). The 3-D fold of *Bt*Fuc2970, however, appears not to provide sufficient space to accommodate an extended aryl aglycon pendant to the atom, which would equate to the ring oxygen in fucose. In order to minimize steric clashes, *Bt*Fuc2970 was subsequently incubated with **5**. The crystal structure of the resulting complex (PDB code: 4WSJ) revealed unambiguous electron density for the presence of a covalent enzyme-**5** complex (Figure 7B). The complex has C-O bond lengths between the *Bt*Fuc2970 catalytic nucleophile and **5** of *ca*. 1.43 Å as would be expected for a C-O ester bond. As expected from the reduction in aglycon size, **5** is better ordered than **4** when bound to the bacterial enzyme and provides a clearer definition of the resulting conformation and interactions. Upon trans-diaxial opening of the acylaziridine the covalently bound and substituted cyclohexane adopts a slightly distorted $^3H_4$ conformation (between $^3H_4$ and $^3S_1$; consistent with the expected catalytic itinerary.[28]

3.3 Conclusions

This chapter reports on the development of potent and selective aziridine-based ABPs **1**, **2** and **3** for selective profiling of active GH29 α-L-fucosidases in cell extracts from bacteria, mice and man as well as *in vivo* in mice. Labeling of GH29 retaining α-L-fucosidases with the L-fucopyranose-configured, cyclophellitol aziridine-based ABPs proceeds with good potency and high selectively both *in vitro* and *in vivo* with the single caveat that the probes do not penetrate brain tissue in mice. The covalent irreversible aziridine inhibitors proved much more potent than their iminosugar counterparts, of which L-fuconojirimycin **6** appeared to be the single compound from this set of configurational isomers that is able to inhibit FUCA1 in a competitive ABPP setting. The crystal structures of α-L-fucosidase from *Bacteroides thetaiotaomicron* in complex with compound **4** and **5** provide strong evidence for the covalent binding of cyclophellitol aziridine to active α-L-fucosidases and by this virtue the validity of the cyclophellitol aziridine design for activity-based profiling of retaining glycosidases that employ the Koshland double displacement mechanism. Whereas aziridines and epoxides that are annulated to cyclohexane rings preferably open in a trans-diaxial fashion through a chair-like transition state, reaction with the α-L-fucosidase nucleophile takes place at the aziridine carbon corresponding to the anomeric center of a substrate α-L-fucoside. This corresponds to ring opening to yield a skew boat, as is observed in the trapped enzyme active site in the co-crystal.

In conclusion, ABPs **1**, **2** and **3** are useful reagents for the discovery and annotation of new members of the GH29 family of α-L-fucosidase in comparative ABPP experiments, for monitoring retaining α-L-fucosidase activities in health and disease, and for the discovery of inhibitors able to interfere with specific α-L-fucosidases in competitive ABPP experiments. ABPs **1**, **2** and **3** add to the growing series of *in situ* and *in vivo* active retaining glycosidase ABPs and moreover their design hold promise for the design of ABPs targeting retaining glycosidases recognizing and processing differently configured and substituted carbohydrates.

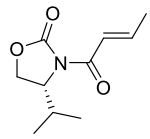
3.4 Experimental section

Synthesis:

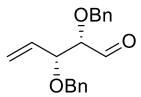
General synthetic methods. All reagents were of a commercial grade and were used as received unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and *N,N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an argon atmosphere unless stated otherwise. Solvents used for flash column chromatography were of pro analysis quality. Reactions were monitored by TLC analysis using Merck aluminium sheets pre-coated with silica gel 60 with detection by UV absorbtion (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}\cdot H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4\cdot H_2O$ (10 g/L) in 10% sulfuric acid followed by charring at ~150 °C or by spraying with 20% sulfuric acid in ethanol followed by charring at ~150 °C. Column chromatography was performed using either Baker - or Screening Device silica gel 60 (0.04-0.063 mm) in the indicated solvents. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a Bruker DMX-600 (600/150 MHz) and

a Bruker AV-400 (400/100 MHz) spectrometer in the given solvent. Chemical shifts are given in ppm relative to the chloroform residual solvent peak or tetramethylsilane (TMS) as internal standard. Coupling constants are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. High-resolution mass spectra were recorded with a LTQ Orbitrap (Thermo Finnigan). Optical rotations were measured on Propol automatic polarimeter (Sodium D-line, λ = 589 nm). IR spectra were recorded on a Shimadzu FT-IR 83000 spectrometer. LC-MS analysis was performed on a Jasco HPLC-system (detection simultaneously at 214 nm and 254 nm) equipped with buffers A: H_2O, B: acetonitrile (MeCN) and C: 10% 0.5 M NH_4OAc, and coupled to a Perkin Elmer Sciex API 165 mass instrument. For reverse phase HPLC purifications an Agilent Technologies 1200 series instrument equipped with a semi preparative Gemini C18 column (10 x 250 mm) was used. The applied buffers were A: H_2O, and B: MeCN.

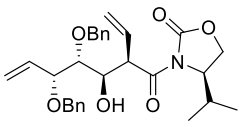
Synthesis of compounds 1-5:



(*R,E*)-3-(But-2-enoyl)-4-isopropyloxazolidin-2-one (14): Compound **14** was prepared from Boc-D-Valine via the strategy reported by Evans[30] *et al.* for its enantiomer, giving compound **14** (3.6 g, 18 mmol, 38% over three steps) as a yellow oil. TLC: R_f 0.59 (pentane/EtOAc, 1/1, v/v); $[\alpha]_D^{20}$ -104 (*c* = 1, $CHCl_3$); lit.[1] for enantiomer: $[\alpha]_D^{20}$ -105 (*c* = 1.97, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.28 (d, *J* = 15.8 Hz, 1H), 7.20-7.11 (m, 1H), 4.51-4.47 (m, 1H), 4.34 – 4.18 (m, 2H), 2.48 – 2.27 (m, 1H), 1.96 (d, *J* = 6.7 Hz, 4H), 0.93 (d, *J* = 7.2 Hz, 3H); 0.89 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 165.01, 154.13, 146.71, 121.92, 63.39, 58.55, 28.51, 18.56, 18.06, 14.72; IR (neat, cm^{-1}): 2965, 1773, 1684, 1638, 1389, 1364, 1233, 1202, 1119, 1061, 1036, 970, 926, 754, 714; HRMS: calculated for $C_{10}H_{15}NO_3$ [M+H$^+$] 198.11247, found: 198.11224;

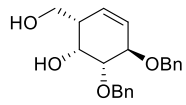


(2*R*,3*S*)-2,3-Bis(benzyloxy)pent-4-enal (14): Building block **15** was prepared from L-(-)-xylose by the reported strategy[31] of Hansen *et. al.* for its enantiomer. Compound **15** was obtained as a clear oil (6.1 g, 21 mmol, overall yield 47%). TLC: R_f 0.45 (pentane/EtOAc, 5/1, v/v); $[\alpha]_D^{20}$ -88 (*c* = 1, $CHCl_3$); lit.[31] for enantiomer: $[\alpha]_D^{20}$ +68.7 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 9.66 (d, *J* = 7.6 Hz, 1H), 7.41 – 7.20 (m, 10H), 5.97 – 5.87 (m, 1H), 5.44 – 5.22 (m, 2H), 4.73 (d, *J* = 12.1 Hz, 1H), 4.60 (t, *J* = 12.0 Hz, 2H), 4.33 (d, *J* = 12.1 Hz, 1H), 4.17 – 4.13 (dd, *J* = 7.7, 4.1 Hz, 1H), 3.81 (d, *J* = 4.1 Hz, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 202.38, 137.53, 137.06, 133.79, 128.46, 128.35, 128.12, 128.08, 127.90, 127.75, 119.84, 85.11, 79.86, 73.39, 70.61; IR (neat, cm^{-1}): 2862, 1732, 1454, 1207, 1069, 1028, 934, 737, 696; HRMS: calculated for $C_{19}H_{20}O_3$ [M+H$^+$] 297.14852, found 297.14922.

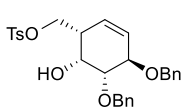


(*R*)-3-((2*R*,3*R*,4*R*,5*R*)-4,5-Bis(benzyloxy)-3-hydroxy-2-vinylhept-6-enoyl)-4-isopropyloxazolidin-2-one (16): The oxazolidinone **14** (2.5 g, 13 mmol, 1.2 eq.) was dissolved in anhydrous DCM (20 mL). After addition of a solution of 1.0 M dibutylboryl trifluoromethanesulfonate (Bu_2BOTf) in anhydrous DCM (13 mL, 13 mmol, 1.2 eq.) at -78 °C, the resulting dark green mixture was removed from the cold bath to dissolve any frozen triflate and cooled again to -78 °C. Triethylamine (2.1 mL, 15 mmol, 1.3 eq.) was added subsequently, causing the dark green color to fade. The solution was stirred for 50 minutes at -78 °C and then at 0 °C for 15 minutes (the solution turned yellow). While the reaction mixture was being cooled back down to -78 °C, a solution of aldehyde **15** (3.4 g, 11 mmol, 1.0 eq.) in anhydrous DCM (20 mL) was added to the reaction mixture via a syringe. The temperature was slowly raised to

-20 °C for one hour and then maintained at this temperature for an additional hour. The resulting yellow solution was then stirred at -15 °C for another hour and then warmed to -5 °C and quenched with a phosphate buffer (pH 7) solution (25 mL). A 30% H_2O_2 solution was then added dropwise while maintaining the internal temperature below 5 °C. Addition of the peroxide was continued until the internal temperature remained constant. The mixture was stirred for an additional 45 minutes while slowly warming to room temperature. The reaction was then poured into aq. sat. $NaHCO_3$ (100 mL) and the aq. layer extracted with DCM (3 x 50 mL). The combined organic layers were dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (2%→20% EtOAc in pentane) giving product **16** as colorless oil (4.4 g, 8.9mmol, 71%). TLC: R_f 0.47 (pentane/EtOAc, 3/1, v/v); $[\alpha]_D^{20}$ +24 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.37 – 7.20 (m, 10H), 6.08-6.00 (m, 1H), 5.96 – 5.82 (m, 1H), 5.42 (m, 2H), 5.37 (m, 1H), 5.28 (d, *J* = 9.9 Hz, 1H), 5.02 – 4.97 (m, 1H), 4.68 (d, *J* = 11.5 Hz, 2H), 4.55 – 4.33 (m, 3H), 4.30 – 4.26 (m, 1H), 4.07 – 3.95 (m, 1H), 3.83 (dd, *J* = 8.9, 3.1 Hz, 1H), 3.57 (dd, *J* = 8.3, 3.9 Hz, 1H), 3.35 (d, *J* = 2.1 Hz, 1H), 3.24 (t, *J* = 8.8 Hz, 1H), 2.21 (m, 1H), 0.77 (d, *J* = 7.0 Hz, 3H), 0.72 (d, *J* = 6.9 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 172.58, 153.62, 138.06, 137.76, 134.48, 133.61, 128.53, 128.34, 128.00, 127.91, 127.66, 127.42, 120.58, 119.37, 81.75, 79.88, 73.04, 71.23, 70.73, 62.45, 58.04, 50.21, 28.02, 17.88, 14.49; IR (neat, cm^{-1}): 3503, 2963, 1776, 1697, 1385, 1371, 1300, 1202, 1099, 1061, 928, 739, 698; HRMS: calculated for $C_{29}H_{35}NO_6$ [M+H$^+$] 494.25371, found: 494.25344.

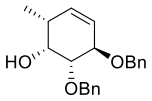


(1*R*,2*S*,5*R*,6*R*)-5,6-Bis(benzyloxy)-2-(hydroxymethyl)cyclohex-3-en-1-ol (17): The product **16** (4.41 g, 8.90 mmol, 1.0 eq.) was dissolved in a mixture of THF (65 mL) and H_2O (3.3 mL). Next, $LiBH_4$ (2 M solution in THF, 26 mL, 52 mmol) was added at 0 °C. After stirring at 0 °C for one hour, the reaction mixture was warmed to room temperature and stirring was continued for one hour. The reaction was quenched with aq. 2 M NaOH (50 mL) and diluted with Et_2O (50 mL). After stirring for five minutes the reaction mixture was extracted with Et_2O (100 mL), and the separated organic phase was washed with aq. sat. $NaHCO_3$ (20 mL) and brine (100 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure to give the crude alcohol that was purified by silica gel column chromatography (10-50%, EtOAc in pentane) giving the intermediate primary alcohol as a white solid (2.72 g, 7.38 mmol, 0.83 eq.) that was dissolved in DCM (260 mL). After addition of the second generation Grubbs catalyst (313 mg, 0.37 mmol, 0.040 eq.), the mixture was stirred at 40 °C in the dark for 24 h. DMSO (0.50 mL) was next added, and the solution was stirred at room temperature for another 3 h. The solvent was evaporated under reduced pressure to give a crude mixture, which was purified by silica gel column chromatography (20-50% EtOAc in pentane) giving product **17** as a black solid (2.4 g, 6.9 mmol, 78%). TLC: R_f 0.41 (pentane/EtOAc, 3/2, v/v); $[\alpha]_D^{20}$ -147 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.41 – 7.25 (m, 10H), 5.89 – 5.84 (m, 1H), 5.60 – 5.56 (m, 1H), 4.80 – 4.66 (m, 4H), 4.38 (br, 1H), 4.35 – 4.27 (m, 1H), 3.92 – 3.76 (m, 2H), 3.68 (dd, *J* = 7.8, 2.2 Hz, 1H), 2.65 (m, 1H), 2,60 (s, 1H), 2.50 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 138.64, 138.14, 128.66, 128.53, 128.07, 128.01, 127.94, 127.78, 127.64, 126.97, 81.88, 76.69, 72.40, 72.36, 70.47, 63.88, 41.97; IR (neat, cm^{-1}) 3422, 2872, 1454, 1206, 1090, 1051, 1026, 860, 801, 733, 696; HRMS: calculated for $C_{21}H_{24}O_4$ [M+H$^+$] 341.17474, found: 341.17486.

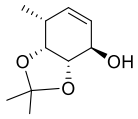


((1*S*,4*R*,5*R*,6*R*)-4,5-Bis(benzyloxy)-6-hydroxycyclohex-2-en-1-yl)methyl 4-methylbenzenesulfonate (18): A solution of **17** (1.73 g, 5.09 mmol, 1.0 eq.) in anhydrous DCM (40 mL), containing Et_3N (1.8 mL, 13 mmol, 2.5 eq.) was cooled to 0 °C and treated with

p-TsCl (2.2 g, 11 mmol, 2.2 eq.). The reaction mixture was stirred at room temperature for 3 h, followed by extra addition of Et_3N (0.80 mL, 5.7 mmol, 1.1 eq.) and *p*-TsCl (1.0 g, 5.0 mmol, 1.0 eq.). After TLC confirmed full conversion of the starting material, the reaction was diluted with water and extracted with DCM. The organic layer was washed with brine and dried over $MgSO_4$. The solvent was removed under reduced pressure and the residue was purified by silica gel column chromatography (5%→35% EtOAc in pentane) giving product **18** as a pale yellow solid (2.18 g, 4.42 mmol, 87%). TLC: R_f 0.30 (pentane/EtOAc, 3/1, v/v); $[\alpha]_D^{20}$ -133 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.77 (d, *J* = 8.0 Hz, 2H), 7.39 - 7.23 (m, 12H), 5.80 – 5.73 (m, 1H), 5.34 (d, *J* = 9.5 Hz, 1H), 4.69 - 4.62 (m, 4H), 4.24 (m, 1H), 4.16 (m, 2H), 4.04 – 3.98 (m, 1H), 3.57 (dd, *J* = 7.6, 2.0 Hz, 1H), 2.73 – 2.63 (m, 1H), 2.51 (br, 1H), 2.39 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 144.89, 138.43, 137.93, 132.65, 129.88, 128.47, 128.45, 128.36, 127.92, 127.86, 127.80, 127.62, 123.64, 81.72, 76.57, 72.31, 72.17, 69.84, 66.83, 40.25, 21.64; IR (neat, cm^{-1}): 3032, 2872, 1597, 1497, 1454, 1358, 1175, 1096, 964, 787, 698, 664; HRMS: calculated for $C_{28}H_{30}O_6S$ [M+H$^+$] 495.18359, found: 495.18300;

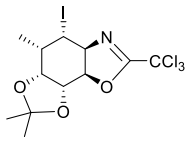


(1*R*,2*R*,5*R*,6*R*)-5,6-Bis(benzyloxy)-2-methylcyclohex-3-en-1-ol (19): Compound **18** (2.2 g, 4.4 mmol, 1.0 eq.) was dissolved in dry THF (18 mL) at 0 °C. A solution of $LiAlH_4$ (2 M in THF) (3.3 mL, 6.6 mmol, 1.5 eq.) was added dropwise. The reaction mixture was stirred at room temperature for 3h, diluted with Et_2O and quenched with dropwise addition of sat. aq. NaCl. The solid material was removed by filtration and the residue washed thoroughly 3 times with hot EtOAc. The filtrate was dried over $MgSO_4$, filtered again and the solvents removed under reduced pressure. The crude product was purified by silica gel column chromatography (10-20% EtOAc in pentane) giving product **19** as yellow oil (1.2 g, 3.8 mmol, 87%). TLC: R_f 0.53 (pentane/EtOAc, 3/1, v/v); $[\alpha]_D^{20}$ -121 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.41 – 7.20 (m, 10H), 5.73 – 5.68 (m, 1H), 5.47 – 5.42 (m, 1H), 4.76 – 4.62 (m, 4H), 4.33 – 4.28 (m, 1H), 4.06 (br, 1H), 3.67 (dd, *J* = 7.7, 2.2 Hz, 1H), 2.45-2.39 (m, 1H), 2.25 (br, 1H), 1.12 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 138.77, 138.37, 131.27, 128.54, 128.44, 127.91, 127.90, 127.86, 127.64, 125.26, 82.73, 76.99, 72.28, 72.09, 70.90, 35.00, 16.36; IR (neat, cm^{-1}) : 3456, 2872, 1497, 1454, 1207, 1090,1057, 980, 785, 735, 696; HRMS: calculated for $C_{21}H_{24}O_3$ [M+H$^+$] 325.17982, found: 325.17995.

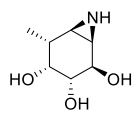


(3*S*,4*R*,7*R*,7*R*)-2,2,7-trimethyl-3a,4,7,7a-tetrahydrobenzo[d][1,3]dioxol-4-ol (20): Ammonia (50 mL) was condensed at -60 °C. Lithium (525 mg, 76 mmol, 10 eq.) was added and the mixture was stirred until the lithium was completely dissolved. To this solution was added a solution of cyclohexene **19** (2.5 g, 7.6 mmol, 1.0 eq.) in dry THF (60 mL). The reaction mixture was stirred for 30 minutes at -60 °C and subsequently quenched with water (10 mL). The resulting solution was allowed to come to room temperature and stirred until all ammonia had evolved. Then the solution was concentrated under reduced pressure, re-dissolved in water and neutralized with Amberlite H$^+$. The resin was removed by filtration, and the filtrate was concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (5%→20% MeOH in DCM) giving a white crystaline product (1.0 g, 7.0 mmol, 0.92 eq.) which was dissolved in 2,2-dimethoxypropane (70 mL) and cooled to 0 °C. A catalytic amount of D-(+)-10-camphorsulfonic acid (CSA) (162 mg, 0.70 mmol, 0.10 eq.) was added and the mixture was stirred at 0 °C for 2 h. TLC analysis showed complete conversion and the mixture was diluted by MeOH/H_2O (50 mL, 9/1, v/v) and stirred at room temperature for 30 minutes. The reaction mixture was neutralized with Et_3N, concentrated under reduced pressure, extracted with DCM, washed with

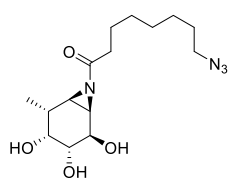
brine, dried over $MgSO_4$, filtered and concentrated. After purification by silica gel column chromatography (0%→8% MeOH in DCM) product **20** was obtained as a pale yellow oil (870 mg, 4.73 mmol, 63%). TLC: R_f 0.40 (DCM/MeOH, 9/1, v/v); $[\alpha]_D^{20}$ -126 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 5.93-5.89 (m, 1H), 5.76 (dd, *J* = 10.0, 3.6 Hz, 1H), 4.42 - 4.39 (m, 1H), 4.28-4.24 (m, 2H), 2.69-2.66 (m, 1H), 1.79 (br, 1H), 1.40 (s, 3H), 1.35 (s, 3H), 1.15 (d, *J* = 7.4 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 135.50, 128.09, 108.20, 80.17, 76.46, 67.80, 31.11, 26.85, 24.91, 16.12; IR(neat, cm^{-1}): 3345, 2913, 2699, 1161, 1084, 1053, 986, 858, 783; HRMS: calculated for $C_{10}H_{16}O_3$ [M+H$^+$] 185.11722, found: 185.11714.



(3*R*,4*S*,5*S*,5*S*,8*R*,8*R*)-5-iodo-2,2,4-trimethyl-7-(trichloromethyl)-3,4,5,5a,8a,8b-hexahydro-[1, 3]dioxolo[4',5':3,4]benzo[1,2-d]oxazole (21): Compound **20** (870 mg, 4.7 mmol, 1.0 eq.) was dissolved in anhydrous DCM (70 mL). The solution was cooled to 0 °C and treated with trichloroacetonitrile (946 µL, 9.5 mmol, 2.0 eq.) and 1,8-diazobicyclo[5.4.0]undec-7-ene (68 µL, 0.47 mmol, 0.10 eq.). After 2h stirring at 0 °C, TLC analysis revealed complete conversion to a higher running product. To the resulting solution was added water (18 mL), $NaHCO_3$ (3.9 g, 47 mmol, 10 eq.) and iodine (4.3 g, 17 mmol, 3.5 eq.). The reaction mixture was stirred overnight at room temperature before being quenched with aq. 10% $Na_2S_2O_3$ solution and extracted three times with EtOAc. The organic layer was dried over $MgSO_4$, filtered, concentrated under reduced pressure and the residue purified by silica gel column chromatography (0%→8% EtOAc in pentane) giving product **21** as brown oil (980 mg, 2.16 mmol, 46%). TLC: R_f 0.49 (pentane/EtOAc, 9/1, v/v); $[\alpha]_D^{20}$ +34 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 5.12 (dd, *J* = 10.2, 4.1 Hz, 1H), 4.89 (dd, *J* = 10.2, 7.0 Hz, 1H), 4.47 – 4.43 (m, 1H), 4.35 (dd, *J* = 8.1, 4.0 Hz, 1H), 4.12 (dd, *J* = 7.0, 3.0 Hz, 1H), 2.11-2.07 (m, 1H), 1.58 (s, 3H), 1.32 (s, 3H), 1.17 (d, *J* = 7.0 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 162.43, 109.70, 84.10, 74.88, 73.75, 73.30, 35.47, 26.11, 25.09, 23.93, 14.98; IR (neat, cm^{-1}): 2982, 1661, 1381, 1207, 1067, 1045, 988, 953, 837, 791, 665, 652; HRMS: calculated for $C_{12}H_{15}Cl_3INO_3$ [M+H$^+$] 453.92350, found: 453.92347.

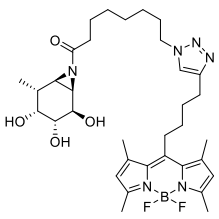


(1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-5-methyl-7-azabicyclo[4.1.0]heptane-2, 3, 4-triol (22): Compound **21** (980 mg, 2.2 mmol, 1.0 eq.) was dissolved in MeOH (32 mL). The solution was treated with concentrated HCl (8.0 mL) at 60 °C overnight. LC-MS analysis showed complete conversion. The solution was concentrated under reduced pressure and re-dissolved in MeOH (30 mL), $NaHCO_3$ (3.9 g, 47 mmol, 22 eq.) was added. After stirring at room temperature for 4 days, the reaction mixture was filtered and concentrated under reduced pressure. After purification by silica gel column chromatography (5%→20% MeOH in DCM) product **22** was obtained as a colorless oil (225 mg, 1.4 mmol, 65%). TLC: R_f 0.26 (DCM/MeOH, 5/1, v/v); $[\alpha]_D^{20}$ -97 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 4.07 (dd, J = 8.8, 4.2 Hz, 1H), 3.57-3.56 (m, 1H), 3.36 – 3.32 (m, 2H), 2.51 (dd, *J* = 6.3, 4.2 Hz, 1H), 1.95 (d, *J* = 6.3 Hz, 1H), 1.92 – 1.84 (m, 1H), 1.17 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 75.90, 74.39, 70.23, 37.32, 36.92, 36.38, 16.82; IR (neat, cm^{-1}): 3283, 1456, 1090, 1065, 995, 914, 874, 752; HRMS: calculated for $C_7H_{13}NO_3$ [M+H$^+$] 160.09682, found: 160.09711.

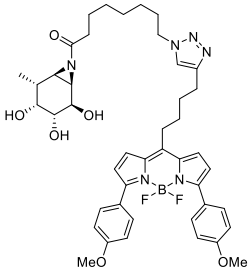


8-Azido-1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)-octan-1-one (23): 8-azido-octanoic acid **24**[19] (207 mg, 1.1 mmol, 1.3 eq.) and EEDQ (277 mg, 1.1 mmol, 1.3 eq.) were dissolved in anhydrous DMF (1.1 mL) and stirred at

room temperature for 2h. This pre-activated mixed anhydride solution (600 µL, 0.61 eq.) was added to a solution of aziridine **22** (137 mg, 0.86 mmol, 1.0 eq.) in DMF (5.0 mL) at 0 °C and stirred for 30 minutes after which the remaining portion of the pre-activated mixed anhydride solution (500 µL, 0.51 eq.) was added. The resulting mixture was stirred at 0 °C for 3h. The reaction was quenched by 2 mL MeOH and the mixture was concentrated *in vacuo*. Then the crude product was purified by silica gel column chromatography (1-10% MeOH in DCM) giving **23** as a colorless oil (162 mg, 0.50 mmol, 58% yield)（or by semi-preparative reversed HPLC (linear gradient: 27-33% B in A, 12 min, solutions used A: H_2O, B: MeCN) followed by lyophilization yielding white powder (2.14mg, 6.6 µmol, 2%). TLC: R_f 0.31 (DCM/MeOH, 10/1, v/v); $[\alpha]_D^{20}$ -29 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.06 (dd, *J* = 8.7, 3.9 Hz, 1H), 3.67 – 3.55 (m, 1H), 3.38 (dd, *J* = 8.0, 1.6 Hz, 1H), 3.28 (t, *J* = 6.9 Hz, 2H), 2.96 (dd, *J* = 6.0, 3.6 Hz, 1H), 2.59 – 2.49 (m, 1H), 2.45 (t, *J* = 7.6 Hz, 1H), 2.42 – 2.36 (m, 1H), 2.03 – 1.95 (m, 1H), 1.64 – 1.55 (m, 4H), 1.41-1.33 (m, 6H), 1.20 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 188.56, 75.71, 74.42, 69.20, 52.36, 43.48, 42.60, 36.94, 36.57, 30.10, 29.90, 29.79, 27.60, 25.83, 16.17; IR (neat, cm^{-1}): 3402, 2932, 2959, 2093, 1674, 1425, 1258, 1167, 1063, 997, 816; LC-MS: R_t 5.35 min, linear gradient 10-90% B in 15 min; ESI-MS: *m/z* = 327.4 $(M+H)^+$; HRMS: calculated for $C_{15}H_{26}N_4O_4$ $[M+H^+]$ 327.20268, found: 327.20266.

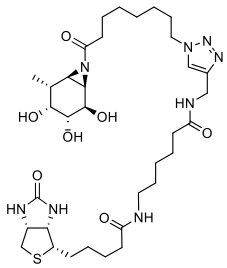


8-(4-(4-(5,5-Difluoro-1,3,7,9-tetramethyl-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diaza-borinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)-1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octan-1-one (1): Azide **23** (40 mg, 0.12 mmol) was dissolved in DMF (3 mL), Bodipy compound **25**[32] (44 mg, 0.13 mmol, 1.1 eq.) and aq. solutions of $CuSO_4$ (1.0 M 24 µL, 0.024 mmol, 0.2 eq.) and aq. solutions of sodium ascorbate (1.0 M, 25 µL, 0.025 mmol, 0.20 eq) were added to the solution under argon atmosphere. The mixture was stirred at room temperature for 2h. The volatiles were removed under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 44%→46% B in A, 12 min, solutions used A: H_2O, B: MeCN) and the pure product **1** was obtained as orange powder after lyophilization (9.5 mg, 0.0145 mmol, 12% yield). ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 7.73 (s, 1H), 6.11 (s, 2H), 4.34 (t, *J* = 6.9 Hz, 2H), 4.05 (dd, *J* = 8.7, 3.9 Hz, 1H), 3.65 – 3.55 (m, 1H), 3.37 (dd, *J* = 8.7, 1.8 Hz, 1H), 3.02 – 2.94 (m, 2H), 2.92 (dd, *J* = 6.0, 3.9 Hz, 1H), 2.77 (t, *J* = 7.3 Hz, 2H), 2.54 – 2.47 (m, 1H), 2.43 (s, 6H), 2.39 – 2.33 (m, 1H), 2.36 (s, 6H), 2.22 (t, *J* = 7.5 Hz, 1H), 2.00-1.93 (m, 1H), 1.92-1.82 (m, 4H), 1.69 – 1.50 (m, 4H), 1.33-1.26 (m, 6H), 1.18 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 188.59, 154.91, 148.51, 147.89, 142.20, 132.58, 123.38, 122.61, 75.83, 74.49, 69.27, 51.17, 43.53, 42.66, 36.89, 36.76, 36.67, 32.22, 32.18, 31.16, 30.83, 29.98, 29.70, 29.65, 29.04, 27.28, 27.18, 26.84, 25.86, 25.75, 16.48, 16.18, 14.45; LC-MS: R_t 8.58 min, linear gradient 10%→90% B in 15 min; ESI-MS: *m/z* = 655.5 $(M+H)^+$; HRMS: calculated for $C_{34}H_{49}BF_2N_6O_4$ $[M+H^+]$ 655.39552, found: 655.39549.

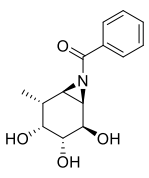


8-(4-(4-(5, 5-Difluoro-3, 7-bis(4-methoxyphenyl)-5H-4l4, 5l4-dipyrrolo[1, 2-c:2',1'-f]-[1, 3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)-1-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octan-1-one (2): Azide **23** (31 mg, 0.099 mmol, 1.0 eq) was dissolved in DMF (3 mL), Bodipy compound **26**[32] (53 mg, 0.11 mmol, 1.1 eq), and aq. solutions of $CuSO_4$ (1.0 M, 20 µL, 0.019 mmol, 0.20 eq.) and aq. solutions of sodium ascorbate (1.0 M, 21 µL, 0.021 mmol, 0.20 eq) were added to the

solution under argon atmosphere and the mixture was stirred at room temperature for 2 h. The reaction was checked with LC-MS within the elution system of 10% NH_4OAc. The volatiles were removed under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 52%→58% B in A, 12 min, solutions used A: H_2O, B: MeCN) resulting a dark blue powder as the product **2** after lyophilization (15.32 mg, 0.019 mmol, 19%). ^{1}H-NMR (400 MHz, CD_3CN): δ ppm 7.83 – 7.74 (m, 4H), 7.51 – 7.49 (s, 1H), 7.46 (d, *J* = 4.4 Hz, 2H), 7.03 – 6.94 (m, 4H), 6.69 (d, *J* = 4.4 Hz, 2H), 4.27 (t, *J* = 7.0 Hz, 2H), 3.91 (dd, *J* = 8.5, 3.9 Hz, 1H), 3.84 (s, 6H), 3.56 – 3.54 (m, 1H), 3.25 (dd, *J* = 8.6, 1.8 Hz, 1H), 3.09 – 3.00 (m, 2H), 2.82 (dd, *J* = 6.1, 3.9 Hz, 1H), 2.78 – 2.69 (m, 2H), 2.44 – 2.23 (m, 3H), 1.88 – 1.74 (m, 6H), 1.55 – 1.45 (m, 2H), 1.31-1.20 (m, 7H), 1.11 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3CN): δ ppm 161.85, 158.30, 148.12, 147.71, 137.18, 132.13, 132.09, 132.05, 128.79, 126.09, 122.37, 121.27, 118.40, 114.62, 101.03, 74.97, 74.36, 69.23, 56.15, 50.66, 42.70, 41.99, 36.77, 36.05, 34.07, 31.01, 30.89, 30.34, 29.64, 29.28, 26.88, 25.72, 25.52, 16.24; LC-MS: R_t 9.15 min, linear gradient 0-90% B in 15 min; ESI-MS: *m/z* = 811.8 $(M+H)^+$; HRMS: calculated for $C_{44}H_{53}BF_2N_6O_6$ [$M+H^+$] 811.41681, found: 811.41690.

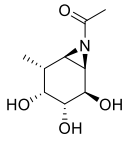


N-((1-(8-oxo-8-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)-6-(5-((3S,4S,6R)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamide (3): Azide compound **23** (31 mg, 0.099 mmol, 1.0 eq) was dissolved in DMF(3.0 mL), biotin-ahx-alkyne **27**[33] (38 mg,0.099 mmol, 1.0 eq), aq. solutions of $CuSO_4$ (1.0 M, 20 µL, 0.019 mmol, 0.20 eq) and aq. solutions of sodium ascorbate (1.0 M, 21 µL,0.021 mmol, 0.22 eq) was added to the solution under argon atmosphere, and the mixture was stirred at 80 °C overnight, the reaction was followed by LC-MS. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 18%→24% B in A, 12min, solutions used A: H_2O, B: actonitrile) and the fraction were freeze-dried without concentration resulting white powder product **3** (9.631 mg, 0.013 mmol, 13%). ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 7.85 (s, 1H), 4.51 (dd, *J* = 7.6, 4.8 Hz, 1H), 4.42 (s, 2H), 4.38 (t, *J* = 3.6 Hz, 2H), 4.32 (dd, *J* = 8.0, 4.4 Hz, 1H), 4.08 (dd, *J* = 8.8, 4.0 Hz, 1H), 3.61 (t, *J* = 1.6 Hz, 1H), 3.39 (dd, *J* = 8.8, 1.6 Hz, 1H), 3.22 - 3.14 (m, 3H), 2.95 - 2.90 (m, 2H), 2.72 (d, *J* = 12.8 Hz, 1H), 2.58 - 2.50 (m, 1H), 2.47 - 2.38 (m, 2H), 2.25 - 2.17 (m, 4H), 2.01 - 1.96 (m, 1H), 1.95 - 1.87 (m, 2H), 1.79 - 1.58 (m, 8H), 1.53 - 1.41 (m, 4H), 1.37 - 1.30 (m, 8H), 1.21 (d, *J* =2.4Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 188.7, 177.0, 176.0, 166.1, 146.3, 124.2, 75.8, 74.5, 69.3, 63.55, 63.4, 61.6, 57.0, 51.3, 43.6, 42.7, 41.1, 40.2, 36.9, 36.8, 36.7, 36.7, 35.6, 31.2, 30.1, 30.0, 29.9, 29.8, 29.7, 29.5, 27.5, 27.3, 27.2, 26.9, 26.8, 26.5, 25.8, 16.2; LC-MS: R_t 4.31min, linear gradient 10→90% B in 15 min; ESI-MS: *m/z* = 721.7 $(M+H)^+$; HRMS: calculated for $C_{34}H_{56}N_8O_7S$ [$M+H^+$] 721.40655, found: 721.40661.



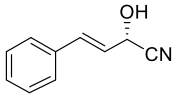
Phenyl((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)methanone (4): Benzoic acid (49 mg, 0.40 mmol, 2.0 eq.) and EEDQ (99 mg, 0.40 mmol, 2.0 eq.) were dissolved in anhydrous DMF (0.40 mL) and stirred at room temperature for 2 h. This pre-activated mixed anhydride solution (200 µL) was added to aziridine **22** (32 mg, 0.20 mmol, 1.0 eq.) in dry DMF (1.0 mL) at 0 °C and stirred for 30 minutes. The remaining half of the pre-activated mixed anhydride solution (200 µL) was added and the resulting mixture was stirred at 0 °C for 2 h. The reaction was quenched with MeOH (1.0 mL) and the mixture was concentrated *in vacuo*. The crude product was purified by HPLC

(linear gradient: 15%→21% B in A, 12 min, solutions used A: H_2O, B: actonitrile) giving compound **4** as white powder (3.8 mg, 15 µmol, 7% yield). TLC: R_f 0.52 (DCM/MeOH, 5/1, v/v); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 8.21 – 8.11 (m, 2H), 7.60 (m, 1H), 7.48 (dd, *J* = 7.6 Hz, 2H), 4.18 (dd, *J* = 8.8, 4.0 Hz, 1H), 3.73-3.71(m, 1H) , 3.55 (dd, *J* = 8.8, 1.8 Hz, 1H), 3.14 (dd, *J* = 6.2, 4.0 Hz, 1H), 2.43 (d, *J* = 6.2 Hz, 1H), 2.31 – 2.23 (m, 1H), 1.90 (s, 1H), 1.21 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 181.40, 134.11, 134.06, 130.50, 129.50, 75.78, 74.65, 69.40, 44.43, 44.34, 36.94, 16.30; LC-MS: R_t 5.35min; linear gradient 0%→90% B in 15 min; ESI-MS: *m/z* = 264.3 $(M+H)^+$; HRMS: calculated for $C_{14}H_{17}NO_4$ $[M+H^+]$ 264.12304, found: 264.12308.



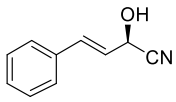
1-((1*R*,2*R*,3*R*,4*R*,5*R*,6*R*)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)ethan-1-one (5): Acetic acid (16 µL, 0.28 mmol, 2.0 eq.) and EEDQ (68 mg, 0.28 mmol, 2.0 eq.) were dissolved in anhydrous DMF (0.30 mL) and stirred at room temperature for 2h. This pre-activated mixed anhydride solution (150 µL) was added to aziridine **22** (22 mg, 0.14 mmol, 1.0 eq.) in dry DMF (0.70 mL) at 0 °C and stirred for 30 min. The remaining half of the pre-activated mixed anhydride solution (150 µL) was added and the resulting mixture was stirred at 0 °C for 2 h. The reaction was quenched by adding MeOH (0.50 mL) and the mixture was concentrated *in vacuo*. The crude product was purified by semi-preparative reversed HPLC (linear gradient: 0%→10% B in A, 12 min, solutions used A: H_2O, B: MeCN) giving compound **5** as white powder after lyophilization (6.9 mg, 34 µmol, 25% yield). TLC: R_f = 0.38 (DCM/MeOH, 10/3, v/v); ^{1}H-NMR (400 MHz, D_2O): δ ppm 4.06 (dd, *J* = 8.9, 4.0 Hz, 1H), 3.76 – 3.63 (m, 1H), 3.46 (dd, *J* = 8.9, 1.8 Hz, 1H), 3.12 (dd, *J* = 6.0, 4.0 Hz, 1H), 2.56 (d, *J* = 6.1 Hz, 1H), 2.16 (s, 3H), 2.13-2.07 (m, 1H) 1.16 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 187.12, 74.33, 72.46, 67.67, 42.08, 42.03, 34.32, 22.58, 14.83; LC-MS: R_t 2.13 min, linear gradient 0%→90% B in 15 min; ESI-MS: *m/z* = 202.2 $(M+H)^+$; HRMS: calculated for $C_8H_{15}NO_4$ $[M+H^+]$ 202.10738, found: 202.10740.

Synthesis of compounds 28 and 29:

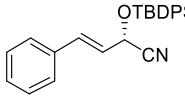


(*S,E*)-2-Hydroxy-4-phenylbut-3-enenitrile:[20] (**Caution!!!** Toxic gas (HCN) may evolve! Work in a well ventilated hood!) In an erlenmeyer flask KCN (26 g, 406 mmol, 3.6 eq.) was dissolved in H_2O (90 mL). On top a layer of methyl *tert*-butyl ether (MTBE, 80 mL) was placed and the mixture was magnetically stirred at such a rate that the two layer system remains. Under slight ice-cooling an aq. 20% (w/w) citric acid solution was added in portions until a pH of 5.45 was reached (pH meter control). At that time the mixture was transferred into a separation funnel, shaken firmly and separated. The water layer was extracted once more with MTBE (80 mL) and the combined MTBE layers were combined and kept on ice. In the meantime a 500 mL three necked flask, equipped with a magnetic stirrer and a thermometer, was charged with a citrate buffer (75 mL, 0.10 M, pH = 5.45), MTBE (20 mL) and cinnamon aldehyde (14.7 g, 111 mmol, 1.0 eq.). The mixture was cooled on an ice bath and *Hb*HNL extract (4.5 g) was added. Under argon and vigorous stirring the ice cold HCN buffer was added drop wise in 15 minutes at 8 °C. The reaction was stirred at this temperature for one hour and for 24 hours at room temperature. At this time TLC showed almost complete conversion and the reaction was stopped. The layers were separated, the water layer extracted once more with MTBE (50 mL). The combined MTBE layers were dried ($MgSO_4$), filtered and evaporated to afford the crude product as a yellow oil (20.6 g, 93% *e.e.* as determined by chiral HPLC). After two crystallizations from DCM/pentane the target cyanohydrin was obtained (9.71 g, 55 %, *e.e.* = 99%) as colorless crystals. $[\alpha]_D^{20}$ -30 (*c* = 1, $CHCl_3$), lit.[34a] $[\alpha]_D^{20}$ -21.8 (*c* = 0.97, $CHCl_3$), lit.[34b] $[\alpha]_D^{20}$ -24.1 (*c* =

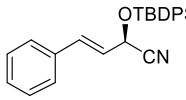
0.8, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.41 - 7.37 (m, 2H), 7.37 - 7.27 (m, 3H), 6.88 (d, *J* = 16.1 Hz, 1H), 6.23 (dd, *J* = 16.1, 6.2 Hz, 1H), 5.14 (dd, *J* = 6.2, 6.2, Hz, 1H), 3.28 (d, *J* = 7.2 Hz, 1H); Chiral HPLC: R_t 14.8 min (Daicel Chiralcel OD, UV 254 nm, Hexane/2-propanol/acetic acid = 85/15/0.1, v/v/v; 1.0 mL/min).



(*R,E*)-2-Hydroxy-4-phenylbut-3-enenitrile:[23] (**Caution!!!** Toxic gas (HCN) may evolve! Work in a well ventilated hood!) In an erlenmeyer flask KCN (28.3 g, 435 mmol, 2.6 eq.) was dissolved in water (100 mL). On top a layer of MTBE (100 mL) was placed and mixture was magnetically stirred at such a rate that the two layer system remains. Under slight ice-cooling an aq. 20% (w/w) citric acid solution was added in portions until a pH of 5.45 was reached (pH meter control). At that time the mixture was transferred into a separation funnel, shaken firmly and separated. The water layer is extracted once more with MTBE (100 mL) and the combined MTBE layers were combined and kept on ice. In the meantime, a 500 mL three necked flask, equipped with a magnetic stirrer and a thermometer, was charged with a citrate buffer (50 mL, 0.1 M, pH = 5.45), MTBE (60 mL) and cinnamon aldehyde (21.5 g, 163 mmol, 1.0 eq.). The mixture was cooled on an ice bath and *pa*HNL (142 mg) was added. Under vigorous stirring the ice cold HCN buffer was added drop wise in 15 minutes. After 64 hours the reaction was stopped, the layers separated and the water layer extracted once more with MTBE (50 mL). The combined MTBE layers were dried ($MgSO_4$), filtered and evaporated to afford a yellow oil (28.8 g) as the crude product. The oil was dissolved in DCM (150 mL) and pentane (200 mL) was added. After standing at room temperature for 2 hours and 2 hours at 4 ºC the formed crystals were collected by filtration and washed with cold pentane twice. Drying afforded the title compound as colorless crystals (11.8 g, 74 mmol, 46 %, *e.e.* = 99%). $[\alpha]_D^{20}$ +30 (*c* = 1, $CHCl_3$), lit.[35] $[\alpha]_D^{20}$ +28.8 (*e.e*. = 92%; *c* = 1.02, $CHCl_3$), lit.[23] $[\alpha]_D^{20}$ +30.5 (*c* = 1.0, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.45 - 7.40 (m, 2H), 7.40 - 7.30 (m, 3H), 6.92 (d, *J* = 15.8 Hz, 1H), 6.26 (dd, *J* = 15.8, 6.0 Hz, 1H), 5.17 (dd, *J* = 6.2, 6.2 Hz, 1H), 2.86 (d, *J* = 7.1 Hz).

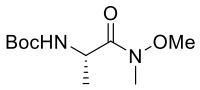


(*S,E*)-2-((*tert*-Butyldiphenylsilyl)oxy)-4-phenylbut-3-enenitrile((*S*)-28): *tert*-Butylchlorodiphenyl- silane (7.2 g, 26 mmol, 1.3 eq.) was dissolved in DMF (80 mL) and imidazole (2.7 g, 40.0 mmol) was added. The mixture was stirred at room temperature for 15 min. Then it was cooled on ice and (*S,E*)-2-hydroxy-4-phenylbut-3-enenitrile (3.2 g, 20 mmol, 1.0 eq.) was added and the reaction stirred for 24 h. TLC analysis revealed complete conversion and the reaction was quenched with H_2O (250 ml), extracted with Et_2O (3 x 100 mL). The combined organic layers were washed with H_2O (50 mL) and brine (50 mL), dried with $MgSO_4$ and evaporated. The mixture was purified by silicagel column chromatography (1%→2%, EtOAc in pentane) to afford the title compound as a colorless oil (7.8 g, 19 mmol, 98 %). $[\alpha]_D^{20}$ +75 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.75 (d, *J* = 7.6, 2H), 7.67 (d, *J* = 7.6, 2H), 7.52 - 7.23 (m, 11H), 6.56 (d, *J* = 15.8 Hz, 1H), 6.13 (d, *J* = 15.8, 6.4 Hz, 1H), 4.97 (d, *J* = 6.4 Hz, 1H), 1.12 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 135.92, 135.87, 135.13, 131.97, 131.59, 130.57, 130.47, 128.85, 128.78, 128.13, 128.05, 127.05, 134.47, 123.38, 118.28, 63.61, 26.77, 19.40; IR (neat, cm^{-1}): 3024, 2933, 2860, 1472, 1428, 1116, 1112, 1075, 1060, 965, 753, 741, 700, 613, 504.

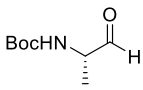


(*R,E*)-2-((*tert*-Butyldiphenylsilyl)oxy)-4-phenylbut-3-enenitrile((*R*)-28): Prepared as described for compound **28**, obtained as pale yellow oil (54 mmol scale, yield 21 g, 99%). $[\alpha]_D^{20}$ -75 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.75 (d, *J* = 7.7, 2H), 7.67 (d, *J* = 7.7,

2H), 7.51 - 7.22 (m, 11H), 6.56 (d, *J* = 15.8 Hz, 1H), 6.13 (dd, *J* = 15.8, 6.4 Hz, 1H), 4.97 (d, *J* = 6.4 Hz, 1H), 1.18 (s, 9H). ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 135.93, 135.88, 135.75, 135.14, 134.89, 131.99, 131.60, 130.67, 130.58, 130.47, 129.68, 128.86, 128.79, 128.39, 128.14, 128.05, 127.77, 127.06, 126.66, 134.49, 123.40, 118.28, 63.62, 26.78, 19.41.

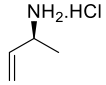


***tert*-Butyl(*S*)-(1-(methoxy(methyl)amino)-1-oxopropan-2-yl)carbamate:**[21] A solution of Boc-L-Alanine (19.2 g, 102 mmol, 1.0 eq.) in DCM (400 mL) was cooled to -15 °C followed by addition of *N,O*-dimethylhydroxylamine hydrochloride (10.1 g, 103 mmol, 1.0 eq.) and then NMM (11.3 ml, 103 mmol). *N*-(3-Dimethylaminopropyl)-*N*-ethylcarbodiimide hydrochloride (19.8 g, 103 mmol, 1.0 eq.) was added portion wise as a solid over 30 minutes. The reaction was stirred at room temperature for 24 hours. After cooling on an ice-bath, 1M HCl was added (30 mL).The aq. layer was extracted twice with DCM (150 mL) and the combined organic layers were washed with a sat. aq. $NaHCO_3$ solution (60 mL), dried ($MgSO_4$), filtered and the solvent was evaporated under vacuum to give the Weinreb amide as a white solid (22.1 g, 96 mmol, 94%) that was used without purification. ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 5.28 (d, *J* = 7.0 Hz), 4.75 - 4.60 (m, 1H), 3.77 (s, 3H), 3.21 (s, 3H), 1.44 (s, 9H), 1.31 (d, *J* = 6.9 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.36, 79.58, 61.71, 46.62, 32.25, 28.47, 18.78.

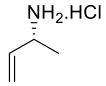


***tert*-Butyl(*S*)-(1-oxopropan-2-yl)carbamate:**[21] Weinreb amide from above (22 g, 95 mmol, 1.0 eq.) was dissolved in anhydrous THF (300 mL) and cooled to 0 °C. A solution of 2.0 M $LiAlH_4$ in THF (47 ml, 95 mmol, 1.0 eq.) was added dropwise and the mixture was stirred for another 30 minutes. The reaction was cooled to -15 °C and a sat. aq. $KHSO_4$ solution (250 mL) was added carefully. The solution was diluted with Et_2O (500 mL) and stirred vigorously for 30 min. The organic layer was separated, dried with $MgSO_4$, filtered and the solvent was evaporated under vacuum to give the aldehyde as a white solid (16 g, 95 mmol, quant.) that was used crude. ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 9.58 (s, 1H), 5.36 – 4.91 (m, 1H), 4.30 - 4.07 (m, 1H), 1.46 (s, 9H), 1.33 (d, *J* = 7.3 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 199.93, 155.41, 80.06, 55.61, 28.39, 14.92.

***tert*-Butyl(*S*)-but-3-en-2-ylcarbamate:**[21] Methyltriphenylphosphonium bromide (57 g, 159 mmol, 17 eq.) was suspended in THF (750 mL) at room temperature and KHMDS (30 g, 152 mmol, 16 eq.) was added. The resultant yellow suspension was stirred at room temperature for 1 hour and then cooled to -78 °C and a solution of the aldehyde from above (16 g, 95 mmol, 1.0 eq.) dissolved in THF (150 mL) was added dropwise. The cooling bath was removed and the mixture was stirred for another 2 hours. The reaction was quenched with MeOH (100 mL) and the resulting mixture was poured into sat. ammonium chloride solution (500 mL). Extraction with Et_2O (3 x 200 mL), dry $MgSO_4$ and evaporation of the solvent *in vacuo* afforded an orange semi-solid that was treated several times with pentane. The combined pentane fractions were filtered and the solvent was evaporated to give the target compound as a white solid that was purified by silica gel column chromatography (2%→5%, EtOAc in pentane) to afford the title compound as a white solid (15 g, 86 mmol, 91 %). $[\alpha]_D^{20}$ -4.2 (*c* = 1, $CHCl_3$), lit.[36] $[\alpha]_D^{20}$ -6.33 (*c* = 1.2, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 5.82 (ddd, *J* = 17.2, 10.4, 5.0 Hz), 5.24 (d, *J* = 7.7 Hz, 1H), 5.12 (d, *J* = 17.2 Hz, 1H), 5.01 (d, *J* = 10.4 Hz, 1H), 4.31 - 4.14 (m, 1H), 1.43 (s, 1H), 1.20 (d, *J* = 7.0 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 154.70, 139.88, 112.84, 78.14, 47.57, 27.93, 20.04. IR (neat, cm^{-1}): 3459, 3020, 1703, 1498, 1215, 1170, 748; HRMS: calculated for $C_9H_{17}NO_2$ [M+H$^+$] 173.13321, found: 173.13307.



(*S*)-But-3-en-2-amine hydrochloride ((*S*)-29): The Boc-protected amine (14 g, 83 mmol, 1.0 eq.) from above was dissolved in MeOH (110 mL), aq. 6.0 M HCl (100 mL) was added and the mixture stirred overnight at room temperature. The solvents were evaporated using a water aspirator affording the title salt as an off white solid (10 g, 76 mmol, quant., 92% overall, *e.e.* > 99%). $[\alpha]_D^{20}$ +2.4 (*c* = 1, MeOH), lit.[37] $[\alpha]_D^{20}$ -3.5 (*c* = 1, EtOH); ^{1}H-NMR (400 MHz, D_2O): δ ppm 5.90 (ddd, *J* = 17.2, 10.6, 6.7 Hz, 1H), 5.37 (d, *J* = 17.2 Hz, 1H), 5.33 (d, *J* = 10.6 Hz, 1H), 3.92 (c, *J* = 6.7 Hz, 1H), 2.20 (s, 2H), 1.37 (d, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 134.76, 118.62, 49.34, 17.90. IR (neat, cm^{-1}): 3400 - 3200, 1649, 1610, 1483, 1425, 1385, 1019, 995, 927, 659.

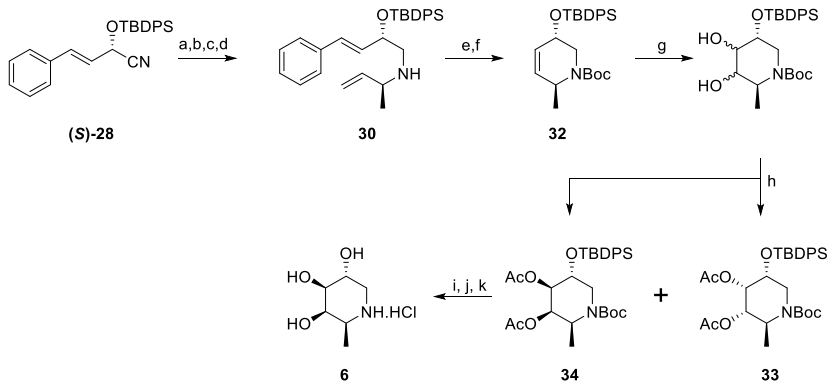


(*R*)-But-3-en-2-amine hydrochloride ((*R*)-29): Prepared from Boc-D-Alanine in the same manner as described above for the (*S*)-enantiomer, *e.e.* = 95%. $[\alpha]_D^{20}$ -3.6 (*c* = 1, MeOH). All spectral data were identical.

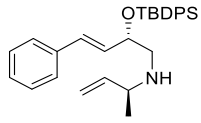
Determination of enantiomeric excess: Analytical samples of both obtained amine hydrochlorides were treated with benzoyl chloride in DCM in the presence of triethyl amine. After work up and purification these samples were subjected to Chiral HPLC analysis on a Daicel Chiralcel OD column (250 x 4.5 mm) using Hexane/2-propanol = 90/10 (v/v), 1.0 mL/min, UV detection (254 nm). (*R*)-isomer, *e.e.* = 95% (left chromatogram); (*S*)-isomer, *e.e.* > 99% (left chromatogram); See chromatograms below.

Preparation of fuconojirimycin (6) and the configurational isomers 7 – 13:

Scheme 3: Preparation of fuconojirimycin **(6)** from cyanohydrin **(S)-28**.

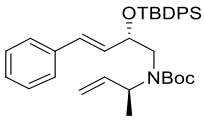


Reagents and conditions: (a) DIBAL-H, -78→5 °C; (b) MeOH, -90 °C; (c) amine **(*S*)-29**, NaOMe, RT, 18 h; (d) $NaBH_4$, 5 °C→RT, 4 h; (e) Boc_2O, 50 °C; (f) Grubbs 1st generation, DCM, reflux; (g) $K_2OsO_4 \cdot 2H_2O$, NMO, acetone/H_2O; (h) Ac_2O, pyridine, DMAP, 0 °C→RT; (i) K_2CO_3, MeOH; (j) TBAF, THF; (k) HCl (6 M in H_2O), MeOH.

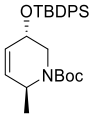


(*S,E*)-*N*-((*S*)-But-3-en-2-yl)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-amine (30): In a flame dried flask and under argon atmosphere, a solution of ((*S,E*)-2-((tert-butyldiphenylsilyl)oxy)-4-phenylbut-3-enenitrile (3.6 g, 9.0 mmol, 1.0 eq.) in

dry Et_2O (80 mL) was cooled to -78 °C. A 1 M solution of DIBAL-H in toluene (14 mL, 14 mmol, 1.5 eq.) was added dropwise and the reaction was allowed to warm up slowly to 10 °C. After cooling to -90 °C, absolute MeOH (14 mL) was added at once, followed by a solution (*S*)-but-3-en-2-amine hydrochloride (3.2 g, 29 mmol, 3.3 eq.) in MeOH (20 mL). Subsequently dry sodium methoxide (2.4 g, 45 mmol, 5.0 eq.) was added to deprotonate the (*S*)-but-3-en-2-amine hydrochloride *in situ*. The cooling bath was removed and the mixture stirred overnight at room temperature under a light flow of argon to reduce the volume of the reaction by half. The mixture was cooled on an ice bath and $NaBH_4$ (1.2 g, 33 mmol, 3.6 eq.) was added in three portions. After stirring for 30 min on the ice bath and two hours at room temperature, the reaction was poured in to an aq. 0.50 M NaOH (90 mL) solution and extracted with diethyl ether (3 x 80 mL). The combined organic layers were washed with a cold aq. 1.0 M HCl solution (100 mL). Evaporation of this acidic aq. layer afforded recovered (*S*)-but-3-en-2-amine hydrochloride (2.07 g, 19.2 mmol, 2.1 eq.). The organic layer was washed subsequently with aq. 0.50 M NaOH (60 mL) solution and brine (30 mL). Drying on $MgSO_4$, filtering and evaporation of the solvent afforded the crude product that was purified by silica gel column chromatography (3%→10%, EtOAc in pentane) to afford the target compound as a yellow oil (3.6 g, 7.9 mmol, 88%). $[\alpha]_D^{20}$ +128 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400MHz, $CDCl_3$): δ ppm 7.72 - 7.62 (m, 4H), 7.46 - 7.13 (m, 11H), 6.19 (d, *J* = 16.0 Hz, 1H), 6.10 (dd, *J* = 16.0, 7 Hz, 1H), 5.60 (ddd, *J* = 17.4, 10.0, 7.7 Hz, 1H), 5.01 (d, *J* = 17.4 Hz, 1H), 4.98 (d, *J* = 10.0 Hz, 1H), 4.45 (dt, *J* = 12.5, 6.3 Hz, 1H), 3.11 (m, 1H), 2.72 (dd, *J* = 11.8, 6.3 Hz, 1H), 2.68 (dd, *J* = 11.8, 5.8 Hz, 1H), 1.08 (s, 9H), 1.05 (d, *J* = 3.4 Hz, 3H); ^{13}C-NMR (101MHz, $CDCl_3$): δ ppm 142.24, 136.80, 136.12, 136.03, 135.68, 134.97, 134.15, 134.06, 131.46, 130.68, 129.80, 129.69, 128.50, 127.71, 127.58, 126.60, 114.88, 74.34, 56.73, 53.64, 27.22, 21.50, 19.50; IR (neat, cm^{-1}): 3071, 2958, 2930, 2856, 1471, 1427, 1109, 740; HRMS: calculated for $C_{30}H_{37}NOSi$ [M+H$^+$] 456.27172, found: 456.27122.

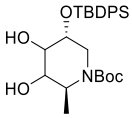


***tert*-Butyl((*S*)-but-3-en-2-yl)((*S,E*)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-yl)carba-mate (31):** To compound **30** (8.3 g, 15 mmol, 1.0 eq.) was added Boc_2O (5.1 g, 23 mmol, 1.5 eq.) and the mixture was stirred overnight at 50 °C. TLC analysis showed complete conversion and after evaporation of the solvent the mixture was purified by silica gel column chromatography (2%→5%, EtOAc in pentane) to afford the title compound **31** as a colorless oil (8.7 g, 15 mmol, 100%). $[\alpha]_D^{20}$ +27 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.78 – 7.62 (m, 4H), 7.47 – 7.12 (m, 11H), 6.15 – 6.00 (m, 2H), 5.77 (ddd, *J* = 16.3, 10.6, 5.2 Hz, 1H), 5.00 – 4.86 (m, 2H), 4.53 – 4.37 (m, 1H), 3.60 – 3.37 (m, 1H), 3.37 – 3.19 (m, 1H), 3.19 - 3.01 (m, 1H), 1.47 - 1.18 (m, 9H), 1.07 (s, 9H), 1.06 (d, *J* = 3.4 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 139.54, 136.15, 136.05, 135.52, 131.12, 130.82, 129.78, 129.69, 128.43, 127.93, 129.69, 128.43, 127.93, 127.69, 127.56, 126.56, 115.13, 73.55, 50.13, 45.54, 28.45, 27.20, 19.44, 17.69; IR (neat, cm^{-1}): 3072, 2933, 2858, 1690, 1428, 1391, 1365, 1164, 1111, 736; HRMS: calculated for $C_{35}H_{45}NO_3Si$ [M+H$^+$] 556.32415, found: 556.32387.

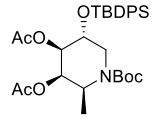


***tert*-Butyl (3*S*,6*S*)-3-((*tert*-butyldiphenylsilyl)oxy)-6-methyl-3,6-dihydropyridine-1(2*H*)-carboxylate (32):** Boc-protected diene **31** (8.7 g, 15 mmol, 1.0 eq.) was dissolved in DCM and argon was bubbled through the solution for five minutes. Grubbs 1st generation (260 mg, 0.32 mmol, 0.020 eq.) was added and the reaction refluxed under argon for 48 hours after which TLC analysis revealed complete conversion. Evaporation of the solvent and silica gel column chromatography (3%→5%, EtOAc in pentane) afforded the compound **32** as a colorless oil (6.85 g, 14 mmol, 97%). $[\alpha]_D^{20}$ +158 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.72 (d, *J* = 6.8

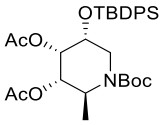
Hz, 2H), 7.66 (d, *J* = 6.8 Hz, 2H), 7.46 – 7.32 (m, 6H), 5.74 – 5.43 (m, 1H), 5.60 - 5.43 (m, 1H), 4.67 – 4.46 (m, 1H), 4.27 – 4.08 (m, 1H), 4.08 – 3.97 (m, 1H), 2.96 – 2.75 (m, 1H), 1.50 (s, 9H), 1.08 (d, *J* = 7.0 Hz, 3H), 1.05 (s, 9H);. ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 154.80, 135.93, 134.93, 134.12, 133.58, 129.80, 127.78, 125.39, 79.46, 64.20, 49.91, 41.94, 28.63, 27.02, 19.34, 17.18. IR (neat, cm^{-1}): 2965, 2931, 2858, 1695, 1416, 1384, 1175, 1131, 1106, 1073, 702; HRMS: calculated for $C_{27}H_{37}NO_3Si$ [M+H$^+$] 452.2615, found: 452.26159.



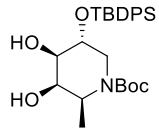
Upjohn dihydroxylation of compound 32: Compound **32** (8.3 g, 18 mmol, 1.0 eq.) was dissolved in a mixture of acetone (70 mL) and water (70 mL) and cooled to -10 °C. *N*-Methylmorpholine-*N*-oxide monohydrate (6.7 g, 50 mmol, 16 eq.) and $K_2OsO_4{\cdot}2H_2O$ (72 mg, 0.19 mmol, 1.0 mol %) were added subsequently. After 24 - 48 hours TLC analysis showed complete conversion of the starting material **32**. The reaction was quenched with an aq. sat. Na_2SO_3 solution (100 mL) and stirred for 30 min. The mixture was diluted with water (100 mL) and extracted with EtOAc (3 x 60 mL). The combined organic layers were washed with aq. 0.6 M HCl, sat. aq. $NaHCO_3$ and brine. After drying (Na_2SO_4), filtering and evaporation of the solvent, afforded a 3:1 mixture of diastereoisomers (5.4 g, 11 mmol, 61%) that could not be separated by silica gel column chromatography.



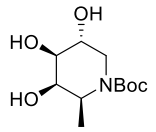
(2*S*,3*R*,4*R*,5*R*)-1-(*tert*-butoxycarbonyl)-5-((*tert*-butyldiphenylsilyl)oxy)-2-methylpiperidine-3,4-diyl diacetate (34): Mixture from above (5.4 g, 11 mmol, 1.0 eq.) was dissolved in pyridine (25 mL) and cooled to 0 °C. Acetic anhydride (6.0 mL, 63 mmol, 5.7 eq.) and a few crystals of DMAP were added and the reaction was stirred for 24 – 48 hours at room temperature. TLC analysis showed complete conversion of the starting material. The reaction was diluted with toluene (100 mL) and the solvents evaporated. The resulting mixture was diluted with EtOAc (100 mL) and washed with H_2O (50 mL), 1 M HCl (50 mL), sat. aq. $NaHCO_3$ solution (50 mL) and brine (50 mL). After drying (Na_2SO_4), filtering and evaporation of the solvent, the silica gel column chromatography (3%→7%, EtOAc in pentane) afforded compound **32** as the first eluting isomer, pale yellow oil (4.0 g, 6.9 mmol, 63%). $[\alpha]_D^{20}$ +12 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.72 – 7.66 (m, 4H), 7.47 – 7.34 (m, 6H), 5.44 (dd, *J* = 6.9, 3.1 Hz, 1H), 5.12 (dd, *J* = 3.9, 3.9 Hz, 1H), 4.54 (qd, *J* = 6.9, 6.9 Hz, 1H), 3.94 (d, *J* = 14.2 Hz, 1H), 3.76 (ddd, *J* = 3.9, 1.5, 1.5 Hz, 1H), 3.07 (dd, *J* = 14.2, 1.5 Hz, 1H) 2.03 (s, 3H), 1.93 (s, 3H), 1.46 (s, 9H), 1.20 (d, *J* = 6.9 Hz, 3H), 1.10 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 169.79, 169.37, 154.89, 135.97, 135.94, 133.13, 133.05, 130.03, 129.94, 127.85, 127.80, 80.03, 70.99, 68.68, 67.41, 48.08, 40.83, 28.48, 26.94, 21.00, 20.96, 19.27, 12.50; IR (neat, cm^{-1}): 2933, 2859, 1752, 1697, 1418, 1366, 1284, 1162, 703; HRMS: calculated for $C_{31}H_{43}NO_7Si$ [M+H$^+$] 570.28816, found: 570.28780.



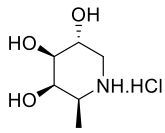
(2*S*,3*S*,4*S*,5*R*)-1-(*tert*-butoxycarbonyl)-5-((*tert*-butyldiphenylsilyl)oxy)-2-methylpiperidine-3,4-diyl di- acetate (33): Obtained as the later eluting isomer, pale yellow oil (1.2 g, 2.0 mmol, 19%). ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.78 – 7.69 (m, 4H), 7.45 – 7.32 (m, 6H), 5.03 (m, 1H), 4.93 (t, *J* = 3.2 Hz, 1H), 4.60 (q, *J* = 7.2 Hz, 1H), 4.06 – 3.97 (m, 2H), 2.89 (d, *J* = 11.2 Hz, 1H), 2.15 (s, 3H), 1.83 (s, 3H), 1.41 (s, 9H), 1.19 (d, *J* = 7.4 Hz, 3H), 1.11 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 170.89, 170.16, 154.99, 136.10, 136.01, 133.69, 133.23, 129.75, 129.60, 127.63, 127.29, 79.90, 70.90, 68.12, 67.33, 51.01, 44.19, 28.32, 26.76, 21.35, 20.77, 19.52, 14.61.



***tert*-Butyl(2*S*,3*R*,4*R*,5*R*)-5-((*tert*-butyldiphenylsilyl)oxy)-3,4-dihydroxy-2-methylpiperidine-1-carboxy- late:** Compound **34** (4.0 g, 7.0 mmol, 1.0 eq.) was dissolved in MeOH (100 mL) and K_2CO_3 (1.3 g, 9.4 mmol, 1.3 eq,) was added. The reaction was stirred for 24 hours after which TLC analysis showed complete conversion of the material **34**. The reaction was acidified with AcOH until pH 5, subsequently diluted with EtOAc (80 mL) and washed with brine (80 mL). Drying (Na_2SO_4), filtering, evaporation of the solvent and silica gel column chromatography (10%→50%, EtOAc in pentane) afforded the title compound as a pale yellow oil (3.4 g, quant.). $[\alpha]_D^{20}$ -4.2 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.71 (dd, *J* = 7.8, 1.2 Hz, 2H), 7.68 (dd, *J* = 7.8, 1.2 Hz, 2H), 7.50 - 7.37 (m, 6H), 4.51 (qd, *J* = 6.9, 6.9 Hz, 1H), 4.14 (dd, *J* = 6.4, 3.1 Hz, 1H), 3.93 - 3.84 (m, 2H), 3.77 (dd, *J* = 3.1, 3.1 Hz, 1H), 3.16 (dd, *J* = 14.5, 1.8 Hz, 1H), 2.80 - 2.17 (m, 2H), 1.48 (s, 9H), 1.23 (d, *J* = 7.1 Hz, 3H), 1.10 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.44, 135.98, 135.82, 133.86, 133.38, 130.00, 128.22, 128.22, 128.01, 127.86, 127.85, 127.09, 127.09, 79.83, 72.46, 70.70, 66.59, 49.83, 39.67, 28.59, 27.09, 19.34, 12.04. IR (neat, cm^{-1}): 3412, 2832, 2858, 1662, 1426, 1365, 1158, 1092, 1017, 755, 740, 700; HRMS: calculated for $C_{27}H_{39}NO_5Si$ [M+H$^+$] 486.26703, found: 486.26669.

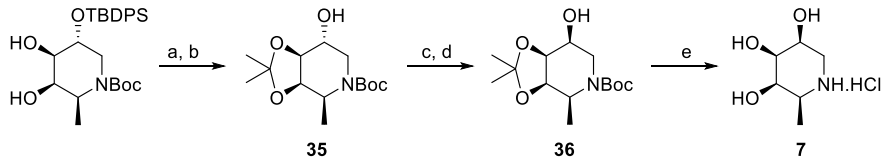


***tert*-Butyl (2*S*,3*R*,4*S*,5*R*)-3,4,5-trihydroxy-2-methylpiperidine-1-carboxylate:** The TBDPS-ether from above (1.5 g, 3.1 mmol, 1.0 eq.) was dissolved in THF (30 mL) and TBAF·$3H_2O$ (2.8 g, 8.8 mmol, 2.7 eq.) was added at room temperature. The reaction was stirred at ambient temperature overnight. TLC indicated complete conversion and the mixture was concentrated. The crude compound was purified by silica gel column chromatography (50%→100%, EtOAc in pentane) to afford the title compound as a colorless oil (727 mg, 2.9 mmol, 94%). $[\alpha]_D^{20}$ +19 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 4.28 (dq, *J* = 6.9, 6.9 Hz, 1H), 3.86 (dd, *J* = 6.9, 2.9 Hz, 1H), 3.85 – 3.75 (m, 3H), 3.27 (dd, *J* = 14.2, 1.7 Hz, 1H), 1.46 (s, 9H), 1.25 (d, *J* = 6.9 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 157.62, 81.17, 73.43, 71.01, 67.57, 52.28, 40.98, 28.90, 12.71; IR (neat, cm^{-1}): 3400 - 3200, 2977, 2931, 1659, 1420, 1365, 1315, 1252, 1158, 1069, 1044, 1015, 732; HRMS: calculated for $C_{11}H_{21}NO_5$ [M+H$^+$] 248.14925, found: 248.14920.



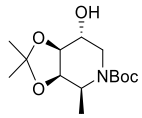
(2*S*,3*R*,4*S*,5*R*)-2-methylpiperidine-3,4,5-triol hydrochloride (6): The Boc-protected-imino sugar from above (645 mg, 2.6 mmol, 1.0 eq.) was dissolved in a mixture of MeOH (20 mL) and aq. 6.0 M HCl (3.0 mL) and stirred overnight at room temperature. The mixture was concentrated to afford the title compound **6** as a white foam (366 mg, 2.0 mmol, 76%). $[\alpha]_D^{20}$ -36 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.04 – 3.92 (m, 2H), 3.60 (dd, *J* = 9.9, 2.8 Hz, 1H), 3.45 – 3.37 (m, 2H), 2.81 (t, *J* = 12.0 Hz, 1H), 1.30 (d, *J* = 6.7 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 73.04, 69.75, 64.29, 54.92, 46.01, 14.01. IR (neat, cm^{-1}): 3400 - 3200, 2942, 2816, 2464, 1457, 1388, 1076, 1003; HRMS: calculated for $C_6H_{13}NO_3$ [M+H$^+$] 148.09682, found: 148.09658.

Scheme 4: Preparation of iminosugar **7** from intermediate **34**.

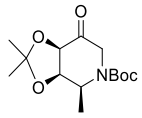


35 **36** **7**

Reagents and conditions: (a) Acetone/2,2-dimethoxypropane, $BF_3{\cdot}EtO_2$, 5 °C; (b) TBAF, THF; (c) Dess-Martin periodinane, DCM; (d) $NaBH_4$, EtOH, -78 °C→RT; (e) HCl (6.0 M in H_2O), MeOH.

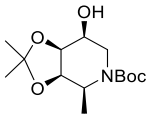


***tert*-Butyl(3*R*,4*S*,7*R*,7*S*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carb-oxylate (35):** The diol derived from diacetate **32** (2.2 g, 4.5 mmol, 1.3 eq.) was dissolved in a mixture of acetone (40 mL) and 2,2-dimethoxypropane (10 ml) and cooled to 5 °C. Boron trifluoride diethyl etherate (200 µL) was added and the reaction stirred on an ice bath during 30 minutes and at room temperature for 18 hours. The reaction was quenched with TEA (2.0 mL) and diluted with EtOAc (125 mL). The mixture was washed with brine (60 mL), dried with $MgSO_4$, filtered and evaporated. The crude product was purified by silica gel column chromatography (1%→5%, EtOAc in pentane) to afford the title compound as a yellow oil (2.0 g, 3.8 mmol, 85%). The TBDPS-protected compound (1.8 g, 3.5 mmol, 1.0 eq.) was dissolved in THF (40 mL) and TBAF·3H_2O (3.4 g, 10 mmol, 3.0 eq.) was added and the reaction stirred at room temperature for 24 hours. TLC analysis confirmed complete conversion. The mixture was diluted with EtOAc (150 mL) and washed with water (20 mL) and brine (20 mL), dried ($MgSO_4$), filtered and concentrated. The crude mixture was purified by silica gel column chromatography (5%→25%, EtOAc in pentane) to afford compound **35** (0.78 g, 2.7 mmol, 77%). $[\alpha]_D^{20}$ +2.0 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.33 (dd, *J* = 6.8, 5.8 Hz, 1H), 4.16 – 4.10 (m, 1H), 4.08 (dd, *J* = 6.8, 3.1 Hz, 1H), 3.90 (dd, *J* = 6.6, 3.1 Hz, 1H), 3.64 – 3.54 (m, 1H), 3.44 (dd, *J* = 13.5, 3.1 Hz,1H), 2.82 – 2.45 (m 1H), 1.48 (s, 3H), 1.46 (s, 9H), 1.34 (s , 3H), 1.28 (d, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 156.58, 108.79, 80.26, 77.33, 74.00, 68.39, 47.71, 42.98, 28.52, 26.65, 24.63, 17.56; IR (neat, cm^{-1}): 3500 – 3200, 2929, 1672, 1405, 1381, 1367, 1253, 1215, 1166, 1049, 751; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18055, found: 288.18061.

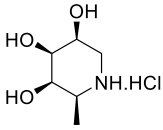


***tert*-Butyl(3*R*,4*S*,7*R*)-2,2,4-trimethyl-7-oxotetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylate:** The alcohol **35** (690 mg, 2.4 mmol, 1.0 eq.) was dissolved in DCM (30 mL) and at 0 °C Dess-Martin periodinane (1.8 g, 4.3 mmol, 1.8 eq.) was added. The reaction mixture was allowed to warm up to room temperature and stirred overnight. TLC indicated complete conversion and the reaction was quenched with a mixture of sat. aq. $NaHCO_3$ (30 mL) and sat. aq. $Na_2S_2O_3$ (30 mL) and stirred for 5 minutes. The mixture was extracted with EtOAc (2 x 50 mL), dried with $MgSO_4$, filtered and concentrated. The crude product was purified by silica gel column chromatography (5%→10%, EtOAc in pentane) to afford the title compound (613 mg, 2.1 mmol, 89%). $[\alpha]_D^{20}$ -13 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.92 – 4.75 (m, 2H), 4.55 (d, *J* = 9.0 Hz, 1H), 4.51 – 4.29 (m, 1H), 3.74 (d, *J* = 19.2 Hz, 1H), 1.53 (s, 3H), 1.48 (s, 9H), 1.39 (s, 3H), 1.16 (d, *J* = 7.0 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 204.12, 154.05, 110.92, 81.12, 77.39, 74.47, 50.20, 28.20, 26.05, 24.33, 12.70. IR (neat, cm^{-1}):

2979, 2932, 1699, 1369, 1247, 1219, 1159, 1016, 773, 745; HRMS: calculated for $C_{14}H_{23}NO_5$ [M+H$^+$] 286.16490, found: 286.16488.

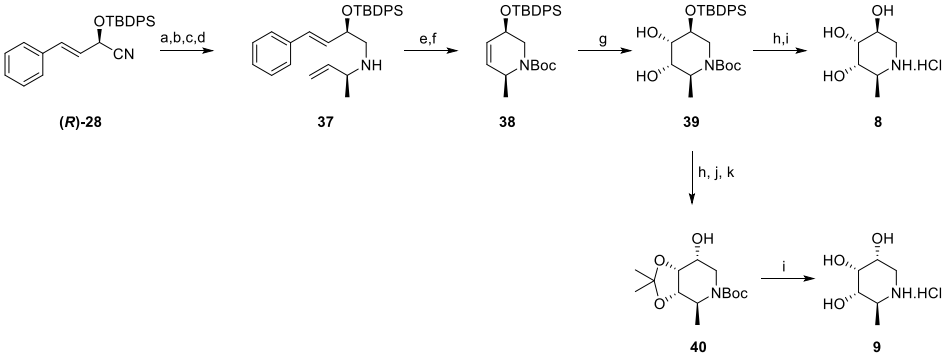


***tert*-Butyl(3*R*,4*S*,7*S*,7*S*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylate (36):** The ketone from above (456 mg, 1.6 mmol, 1.0 eq.) was dissolved in EtOH (20 mL) and at -78 °C $NaBH_4$ (48 mg, 1.3 mmol, 0.80 eq.) was added and the reaction was allowed to warm up slowly over night. TLC indicated complete conversion and the mixture was diluted with EtOAc (50 mL), washed subsequently with water (30 mL) and brine (20 mL), dried with $MgSO_4$, filtered and evaporated to give a crude product that was purified by silica gel column chromatography (25%→100%, EtOAc in pentane) to afford compound **36** (184 mg, 0.64 mmol, 40%). $[\alpha]_D^{20}$ +2.0 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 4.43 – 4.33 (m, 2H), 4.01 (dq, *J* = 6.6, 6.6 Hz, 1H), 3.88 (dd, *J* = 12.0, 3.8 Hz, 1H), 3.61 (ddd, *J* = 10.2, 4.3, 4.3 Hz, 1H), 2.99 (t, *J* = 12.0 Hz, 1H), 2.92 – 2.60 (m, 1H), 1.53 (s, 3H), 1.46 (s, 9H), 1.39 (s, 3H), 1.34 (d, *J* = 6.6 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 155.13, 108.78, 80.12, 76.03, 72.85, 66.35, 47.44, 41.89, 28.48, 26.42, 24.80, 16.77; IR (neat, cm^{-1}): 3500 – 3200, 2980, 2934, 1688, 1393, 1366, 1251, 1212, 1159, 1023, 867, 773, 734; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18056, found: 288.18057.

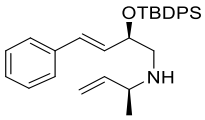


(2*S*,3*R*,4*S*,5*S*)-2-methylpiperidine-3,4,5-triol hydrochloride (7): The Boc-protected-iminosugar **36** (144 mg, 0.50 mmol, 1.0 eq.) was dissolved in a mixture of MeOH (20 mL) and aq. 6.0 M HCl (3.0 mL) and stirred over night at room temperature. The mixture was concentrated to afford the title compound **7** as a white foam (92 mg, quant.). $[\alpha]_D^{20}$ +13 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): δ ppm 4.20 – 4.16 (m, 1H), 4.00 – 3.97 (m, 1H), 3.83 (t, *J* = 3.3 Hz, 1H), 3.40 (dd, *J* = 13.8, 2.8 Hz, 1H), 3.41 – 3.36 (m, 1H), 3.22 (dd, *J* = 13.8, 1.2 Hz, 1H), 1.36 (d, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 70.46, 67.34, 66.21, 55.03, 48.21, 14.12; IR (neat, cm^{-1}): 3676, 3400 – 3200, 2971, 2925, 1724, 1568, 1148, 1407, 1394, 1250, 1118, 1075, 1066; HRMS: calculated for $C_6H_{13}NO_3$ [M+H$^+$] 148.09682, found: 148.09675.

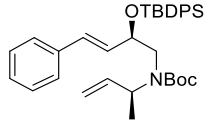
Scheme 5: Preparation of iminosugars **8** and **9**.



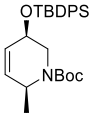
Reagents and conditions: (a) DIBAL-H, -78→5 °C; (b) MeOH, -90 °C; (c) amine **(*S*)-29**, NaOMe, 18 h; (d) $NaBH_4$, 5 °C→RT, 4 h; (e) Boc_2O, 50 °C; (f) Grubbs 1st generation, DCM, reflux; (h) i) $K_2OsO_4{\cdot}2H_2O$, NMO, acetone/H_2O; ii) TBAF, THF; (i) HCl (6.0 M in H_2O), MeOH; (j) Dess-Martin periodinane, DCM; (k) $NaBH_4$, EtOH, -78 °C→RT.



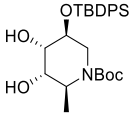
(*R, E*)-*N*-((*S*)-But-3-en-2-yl)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-amine (37): Prepared as described for compound **30** (18 mmol scale, yield 6.8 g, 83%). $[\alpha]_D^{20}$ -92 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.71 (d, *J* = 7.1 Hz, 2H), 7.68 (d, *J* = 7.1 Hz, 2H), 7.48 - 7.09 (m, 11H), 6.21 (d, *J* = 16.0 Hz, 1H), 6.10 (dd, *J* = 16.0, 6.9 Hz, 1H), 5.57 (ddd, *J* = 17.4, 10.3, 7.8 Hz, 1H), 4.98 (d, *J* = 10.3 Hz, 1H), 4.95 (d, *J* = 7.8 Hz, 1H), 4.46 (dd, *J* = 12.0, 6.3 Hz, 1H), 3.04 (c, *J* = 6.7 Hz, 1H), 2.78 (dd, *J* = 12.0, 6.7 Hz, 1H), 2.59 (dd, *J* = 12.0, 5.2 Hz, 1H), 1.08 (s, 9H), 1.02 (d, *J* = 6.4Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 142.45, 136.76, 136.03, 135.92, 135.36, 134.10, 134.07, 133.96, 131.20, 130.79, 129.72, 129.16, 128.40, 127.65, 127.50, 126.48, 114.51, 74.07, 56.25, 53.64, 27.16, 21.77, 19.43; IR (neat, cm^{-1}): 2931, 2858, 1219, 1112, 772, 702; HRMS: calculated for $C_{30}H_{37}NOSi$ [$M+H^+$] 456.27172, found: 456.27144.



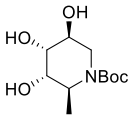
***tert*-Butyl((*S*)-but-3-en-2-yl)((*R,E*)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-yl) carba- mate:** Prepared as described for **31** (3.9 mmol scale, yield 2.25 g, quant.). $[\alpha]_D^{20}$ -68 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.69 (d, *J* = 6.8 Hz, 2H), 7.65 (d, *J* = 6.8 Hz, 2H), 7.48 – 7.09 (m, 11H), 6.17 – 6.00 (m, 2H), 5.71 (ddd, *J* = 16.1, 10.3, 4.8 Hz, 1H), 5.08 – 4.78 (m, 2H), 4.47 – 4.34 (m, 1H), 3.59 – 3.40 (m, 1H), 3.37 – 3.18 (m, 1H), 3.15 – 3.00 (m, 1H), 1.42 – 1.22 (m, 9H), 1.07 (s, 9H), 1.06 - 1.04 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.61, 143.50, 136.15, 136.05, 135.68, 131.05, 130.89, 129.81, 129.71, 128.02, 127.94, 127.73, 127.58, 126.58, 114.89, 79.64, 73.93, 50.39, 41.47, 28.44, 27.21, 19.44, 17.40; IR (neat, cm^{-1}): 3072, 2932, 2858, 1693, 1266, 1167, 1113, 1070, 741, 702; HRMS: calculated for [$C_{35}H_{45}NO_3Si$ [$M+H^+$] 556.32415, found: 556.32436.



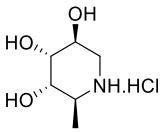
***tert*-Butyl (3*R*, 6*S*)-3-((*tert*-butyldiphenylsilyl)oxy)-6-methyl-3, 6-dihydropyridine-1(2*H*)-carboxy-late (38):** Prepared as described for **32**. Compound **38** was obtained as a clear oil (3.8 mmol scale, yield 1.6 g, 96%). $[\alpha]_D^{20}$ +22 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$) *δ* ppm 7.75 - 7.60 (m, 4H), 7.47 – 7.32 (m, 6H), 5.76 – 5.37 (m, 2H), 4.44 – 4.28 (m, 1H), 4.28 – 4.15 (m, 1H), 4.09 – 3.88 (m, 1H), 2.80 – 2.67 (m, 1H), 1.34 (s, 9H), 1.16 (d, *J* = 6.6 Hz, 3H), 1.08 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$) *δ* ppm 153.93, 135.91, 135.80, 135.46, 130.86, 130.57, 129.88, 129.78, 128.31, 127.79, 127.57, 126.52, 79.61, 65.63, 47.42, 43.99, 28.48, 27.06, 19.31, 17.85. IR (neat, cm^{-1}): 3072, 2932, 2858, 1697, 1453, 1366, 1162, 1112, 763, 741, 702; HRMS: calculated for $C_{27}H_{37}NO_3Si$ [$M+H^+$] 452.26155, found: 452.26185.



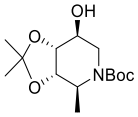
***tert*-Butyl(2*S*,3*S*,4*S*,5*S*)-5-((*tert*-butyldiphenylsilyl)oxy)-3,4-dihydroxy-2-methylpiperidine-1-carbox ylate (39):** Prepared as described in the Upjohn procedure concerning compound **32**. Compound **37** was obtained as a clear oil (1.1 mmol scale, yield 430 mg, 79%). $[\alpha]_D^{20}$ +21 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.70 (dd, *J* = 7.9, 1.5 Hz, 2H), 7.68 (dd, *J* = 7.9, 1.5 Hz, 2H), 7.48 - 7.36 (m, 6H), 4.55 - 4.29 (m, 1H), 4.29 - 3.96 (m, 1H), 3.90 (td, J = 10.3, 5.4 Hz, 1H), 3.84 - 3.74 (m, 1H), 3.71 (dd, *J* = 8.8, 2.9 Hz, 1H), 2.77 (dd, *J* = 13.2, 10.7 Hz, 1H), 2.38 - 2.00 (m, 2H), 1.34 (s, 9H), 1.13 (d, *J* = 7.3 Hz, 3H), 1.09 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.32, 135.92, 135.81, 133.75, 130.16, 130.13, 128.03, 127.97, 80.03, 73.46, 72.52, 70.24, 48.03, 41.78, 28.42, 27.14, 19.46, 14.18; IR (neat, cm^{-1}): 3020, 1215, 1111, 770, 748, 668; HRMS: calculated for $C_{27}H_{39}NO_5Si$ [$M+H^+$] 486.26703, found: 486.26723.



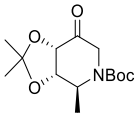
***tert*-Butyl (2*S*, 3*S*, 4*R*, 5*S*)-3,4,5-trihydroxy-2-methylpiperidine-1-carboxylate:** The TBDPS-ether **39** (1.9 g, 3.9 mmol, 1.0 eq.) was dissolved in THF (40 mL) and TBAF·$3H_2O$ (3.5 g, 11 mmol, 2.8 eq.) was added at room temperature. The reaction was stirred at ambient temperature overnight. TLC indicated complete conversion and the mixture was concentrated. The crude compound was purified by silica gel column chromatography (50%→100%, EtOAc in pentane) to afford the title compound as a colorless oil (957 mg, 3.8 mmol, 98%). $[\alpha]_D^{20}$ +38 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 4.35 (m, 1H), 4.08 (dd, *J* = 13.0, 5.4 Hz, 1H), 3.83 – 3.69 (m, 2H), 3.55 (dd, *J* = 9.5, 3.1 Hz, 1H), 2.74 – 2.61 (m, 1H), 1.46 (s, 9H), 1.15 (d, *J* = 7.3 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 157.27, 81.38, 73.99, 73.38, 68.19, 56.32, 45.39, 28.84, 14.45; IR (neat, cm^{-1}): 3020, 1215, 770, 747; HRMS: calculated for $C_{11}H_{21}NO_5$ [M+H$^+$] 248.14925, found: 248.14924.



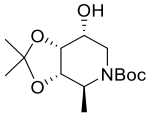
(2*S*,3*S*,4*R*,5*S*)-2-methylpiperidine-3,4,5-triol hydrochloride (8): Prepared as described for iminosugar **6**. Boc-iminosugar from above was used and **8** was obtained as a white foam (2.7 mmol scale, yield 445 mg, 90%). $[\alpha]_D^{20}$ -10 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.14 (ddd, *J* = 4.9, 3.3, 3.3 Hz, 1H), 4.00 (dd, *J* = 4.1, 3.1 Hz, 1H), 3.85 (dd, *J* = 9.9, 3.1 Hz, 1H), 3.39 (dq, *J* = 9.9, 6.6 Hz, 1H), 3.33 (dd, *J* = 13.6, 1.9 Hz, 1H), 3.17 (dd, *J* = 13.6, 3.1 Hz, 1H), 1.39 (d, *J* = 6.6 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): 68.45, 68.09, 66.32, 51.07, 43.88, 12.72; IR (neat, cm^{-1}): 3369, 3271, 3022, 2952, 1727, 1583, 1437, 1260, 1086, 1069, 1051, 965, 701; HRMS: calculated for $C_6H_{13}NO_3$ [M+H$^+$] 148.09682, found: 148.09672.



***tert*-Butyl(3*S*,4*S*,7*S*,7*R*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carb-oxylate:** Prepared as described en route towards **35** (scheme 2, steps b, c). The title compound was obtained as a clear oil (4.8 mmol scale, yield 1.25 g, 91%). $[\alpha]_D^{20}$ +48 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.68 – 4.52 (m, 1H), 4.05 (d, *J* = 5.4 Hz, 1H), 4.00 (t, *J* = 6.3 Hz, 1H), 3.93 (dd, *J* = 13.5, 4.7 Hz, 1H), 3.76 (m, 1H), 2.92 (br, 1H), 2.82 (dd, *J* = 13.5, 10.2 Hz, 1H), 1.48 (s, 3H), 1.47 (s, 9H), 1.35 (s, 3H), 1.20 (d, *J* = 7.3Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.43, 108.88, 80.31, 78.10, 77.85, 69.50, 48.18, 41.46, 28.48, 28.22, 26.20, 16.86; IR (neat, cm^{-1}): 3020, 1215, 748, 668; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18055, found: 288.18053.

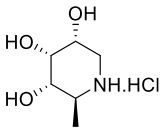


***tert*-Butyl(3*S*,4*S*,7*S*)-2,2,4-trimethyl-7-oxotetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylat-e:** Prepared as described en route towards **36** (scheme 2, step d). The title ketone was obtained as a white solid (0.50 mmol scale, yield 108 mg, 76%). $[\alpha]_D^{20}$ -2.4 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.76 – 4.57 (m, 2H), 4.36 (d, *J* = 6.6 Hz, 1H), 4.30 (d, *J* = 6.6 Hz, 1H), 3.66 (d, *J* = 18.5 Hz, 1H), 1.48 (s, 9H), 1.46 (s, 3H), 1.35 (s, 3H), 1.16 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 203.20, 157.70, 111.69, 80.86, 80.25, 75.54, 50.94, 49.23, 26.67, 26.18, 24.98, 15.47; IR (neat, cm^{-1}): 2980, 2935, 1737, 1693, 1408, 1367, 1221, 1157, 1049, 867; HRMS: calculated for $C_{14}H_{23}NO_5$ [M+H$^+$] 286.16490, found: 286.16478.



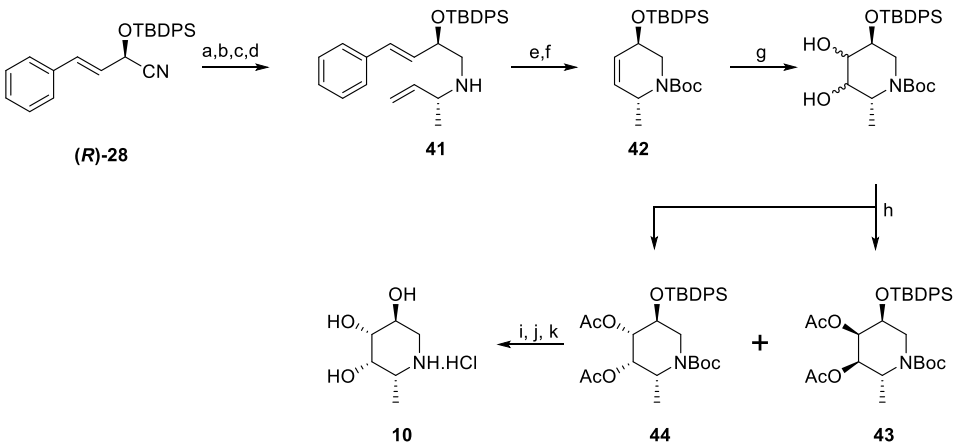
***tert*-Butyl(3*S*,4*S*,7*R*,7*R*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-c-arboxylate (40):** Prepared as described for alcohol **36** from the ketone described above. Alcohol **40** was obtained as a mixture of two diastereoisomers (ratio 93:7, 0.42 mmol scale, yield 104 mg, 86%). $[\alpha]_D^{20}$ +42 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.43 (dd, *J* = 6.8, 4.3 Hz, 1H), 4.33

(dd, *J* = 7.4, 2.0 Hz, 1H), 4.27 – 4.19 (m, 1H), 3.96 (ddd, *J* = 11.3, 4.5, 4.5 Hz, 1H), 3.61 (dd, *J* = 11.7, 4.4 Hz, 1H), 3.08 (t, *J* = 11.7 Hz, 1H), 2.68 – 2.48 (m, 1H), 1.50 (s, 3H), 1.47 (s, 9H), 1.38 (s, 3H), 1.17 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.14, 108.85, 79.82, 77.35, 71.79, 65.03, 47.68, 42.46, 28.49, 26.23, 24.37, 19.10; IR (neat, cm^{-1}): 3437, 2978, 2935, 1683, 1401, 1369, 1255, 1212, 1169, 1049, 877, 731; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18056, found: 288.18063.

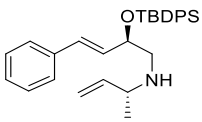


(2*S*,3*S*,4*R*,5*R*)-2-methylpiperidine-3,4,5-triol hydrochloride (9): Prepared as described for iminosugar **6** from Boc-protected-iminosugar **38**. Iminosugar **8** was obtained as a colorless foam (1.7 mmol scale, yield 311 mg, quant.) in a *d.r.* of 93:7. *N*-Boc-protection, column chromatography and subsequent deprotection (HCl/MeOH) afforded the diastereomerically pure compound **9**. $[\alpha]_D^{20}$ -20 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.19 – 4.12 (m, 1H), 4.00 (ddd, *J* = 11.5, 4.9, 2.6 Hz, 1H) 3.62 (dd, *J* =10.4, 2.5 Hz, 1H), 3.33 (dq, *J* = 4.9, 4.9 Hz, 1H), 3.25 (dd, *J* = 12.0, 4.9 Hz, 1H), 3.09 (t, *J* = 12.0 Hz, 1H), 1.37 (d, *J* = 6.6 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 69.95, 69.61, 64.72, 49.92, 41.65, 14.16; IR (neat, cm^{-1}): 3400 - 3200, 2939, 2804, 1456, 1158, 1106, 1072, 1043, 1018, 996, 962, 816, 707; HRMS: calculated for $C_6H_{13}NO_3$ [M+Na$^+$] 170.07876, found: 170.07865.

Scheme 6: Preparation of iminosugar **10**.

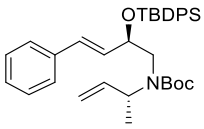


Reagents and conditions: (a) DIBAL-H, -78→5 °C; (b) MeOH, -90 °C; (c) amine **(*R*)-29**, NaOMe, 18 h; (d) $NaBH_4$, 5 °C→RT, 4 h; (e) Boc_2O, 50 °C; (f) Grubbs 1st generation, DCM, reflux; (g) $K_2OsO_4{\cdot}2H_2O$, NMO, acetone/H_2O; (h) Ac_2O, pyridine, DMAP, 0 °C→RT; (i) K_2CO_3, MeOH; (j) TBAF, THF; (k) HCl (6.0 M in H_2O), MeOH.

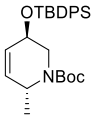


(*R,E*)-*N*-((*R*)-But-3-en-2-yl)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-amine (41): Prepared as described for compound **30**. (35 mmol scale, yield 13 g, 83%). $[\alpha]_D^{20}$ -109 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.72 – 7.62 (m, 4H), 7.43 – 7.15 (m, 11H), 6.20 (d, *J* = 16.0 Hz, 1H), 6.10 (dd, *J* = 16.0, 7.0 Hz, 1H), 5.60 (ddd, *J* = 16.8, 10.2, 7.0 Hz, 1H), 5.01 (d, *J*

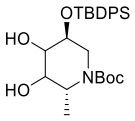
= 16.8 Hz, 1H), 4.97 (d, *J* = 10.2 Hz, 1H), 4.44 (dt, *J* = 12.6, 6.2 Hz, 1H), 3.10 (m, 1H), 2.72 (dd, *J* = 11.8, 6.3 Hz, 1H), 2.69 (dd, *J* = 11.8, 5.6 Hz, 1H), 1.08 (s, 9H), 1.05 (t, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 142.46, 136.88, 136.14, 136.06, 135.49, 134.24, 134.14, 131.40, 130.83, 129.79, 129.69, 128.51, 127.71, 127.59, 126.61, 114.69, 74.45, 56.72, 53.77, 27.25, 21.58, 19.52; IR (neat, cm^{-1}): 3500 - 3200, 2967, 1653, 1111, 1055, 1033, 1015, 741, 700; HRMS: calculated for $C_{30}H_{37}NOSi$ [M+H$^+$] 456.27172, found: 456.27139.



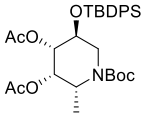
***tert*-Butyl((*R*)-but-3-en-2-yl)((*R,E*)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-yl)carba-mate:** Prepared as described for **29** (29 mmol scale, yield 16 g, 99%). $[\alpha]_D^{20}$ -25 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.78 – 7.62 (m, 4H), 7.47 – 7.10 (m, 11H), 6.15 - 6.05 (m, 2H), 5.78 (ddd, *J* = 16.6, 10.6, 5.2 Hz, 1H), 5.00 – 4.86 (m, 2H), 4.55 – 4.38 (m, 1H), 3.59 – 3.37 (m, 1H), 3.37 – 3.20 (m, 1H), 3.20 – 3.04 (m, 1H), 1.40 - 1.20 (m, 9H), 1.08 (s, 9H), 1.08 – 1.04 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 155.57, 146.87, 136.14, 135.12, 134.92, 131.13, 130.84, 129.78, 128.42, 127.68, 126.56, 79.64, 73.60, 50.16, 43.52, 28.45, 27.54, 19.43, 17.62; IR (neat, cm^{-1}): 2977, 2933, 2858, 1808, 1757, 1691, 1396, 1370, 1212, 1166, 1113, 1065, 739, 701; HRMS: calculated for $C_{35}H_{45}NO_3Si$ [M+H$^+$] 556.32415, found: 556.32416.



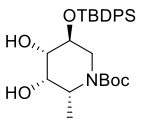
***tert*-Butyl (3*R*, 6*R*)-3-((*tert*-butyldiphenylsilyl)oxy)-6-methyl-3,6-dihydropyridine-1(2*H*)-carboxylate(42):** Prepared as described for compound **32** from the Boc-protected diene mentioned above. Compound **42** was obtained as a colorless oil (8.4 mmol scale, yield 3.6 g, 95%). $[\alpha]_D^{20}$ -153 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.72 (d, *J* = 7.7 Hz), 7.66 (d, *J* = 7.7 Hz, 2H), 7.46 – 7.32 (m, 6H), 5.71 – 5.62 (m, 1H), 5.58 – 5.47 (m, 1H), 4.64 – 4.48 (m, 1H), 4.28 – 4.09 (m, 1H), 4.07 – 4.00 (m, 1H), 2.96 – 2.76 (m, 1H), 1.50 (s, 9H), 1.08 (d, *J* = 6.9 Hz, 3H**),** 1.05 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 154.80, 135.94, 134.12, 129.80, 129.65, 127.78, 127.58, 79.45, 64.21, 46.44, 28.63, 27.03, 19.35, 17.27; IR (neat, cm^{-1}): 2978, 2933, 2858, 1808, 1757, 1691, 1470, 1212, 1166, 1113, 1065, 739, 701; HRMS: calculated for $C_{27}H_{37}NO_3Si$ [M+H$^+$] 452.26155, found: 452.26155.



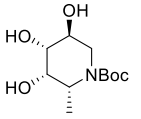
***tert*-Butyl(2*R*, 5*S*)-5-((tert-butyldiphenylsilyl)oxy)-3, 4-dihydroxy-2-methylpiperidine-1-carboxylate:** The procedure described for the Upjohn dihydroxylation of compound **32** afforded a 3 : 1 mixture of inseparable diastereoisomers (27 mmol scale, yield 10 g, 78%). To separate these diastereoisomers the mixture was directly converted into the diacetates.



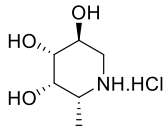
(2*R*,3*S*,4*S*,5*S*)-1-(*tert*-butoxycarbonyl)-5-((*tert*-butyldiphenylsilyl)oxy)-2-methylpiperidine-3,4-diyl di-acetate (44): Prepared as described for **34**. Compound **44** was obtained as a pale yellow oil (21 mmol scale, yield 7.2 g, 61%). $[\alpha]_D^{20}$ -10 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.73 – 7.67 (m, 4H), 7.46 – 7.35 (m, 6H), 5.44 (dd, *J* = 6.9, 3.2 Hz, 1H), 5.12 (dd, *J* = 3.2, 3.6 Hz, 1H), 4.54 (qd, *J* = 6.9, 6.9 Hz, 1H), 3.94 (d, *J* = 14.3 Hz, 1H), 3.77 (dd, *J* = 3.6, 1.5 Hz, 1H), 3.08 (dd, *J* = 14.3, 1.5 Hz, 1H) 2.03 (s, 3H), 1.92 (s, 3H), 1.47 (s, 9H), 1.20 (d, *J* = 6.9 Hz, 3H**),** 1.09 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 169.75, 169.33, 154.87, 135.95, 135.91, 133.11, 133.04, 130.01, 129.92, 127.83, 127.78, 80.00, 70.98, 68.67, 67.39, 48.07, 40.83, 28.46, 26.92, 20.97, 20.92, 19.25, 12.48; IR (neat, cm^{-1}): 3073, 2933, 2859, 1750, 1694, 1417, 1365, 1237, 1218, 1160, 1026, 753, 740, 701; HRMS: calculated for $C_{31}H_{43}NO_7Si$ [M+H$^+$] 570.28816, found: 570.28791.



***t*ert-Butyl(2*R*,3*S*,4*S*,5*S*)-5-((*tert*-butyldiphenylsilyl)oxy)-3,4-dihydroxy-2-methylpiperidine-1-carbo xy-late:** Prepared as described en route towards fucononojirimycin (**6**) and was obtained as a colorless oil (12 mmol scale, yield 5.5 g, 85%). $[\alpha]_D^{20}$ +5.4 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.68 (dd, *J* = 7.8, 1.4 Hz, 2H), 7.64 (dd, *J* = 7.8, 1.4 Hz, 2H), 7.45 – 7.32 (m, 6H), 4.48 (qd, *J* = 6.9, 6.9 Hz, 1H), 4.11 (dd, *J* = 6.4, 3.2 Hz, 1H), 3.89 – 3.80 (m, 2H), 3.74 (t, *J* = 3.1 Hz, 1H), 3.12 (dd, *J* = 14.5, 1.7 Hz, 1H), 2.81 (br, 1H), 2.44 (br, 1H), 1.44 (s, 9H), 1.19 (d, *J* = 7.1 Hz, 3H), 1.07 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.45, 135.95, 135.80, 133.81, 133.36, 129.97, 127.83, 79.84, 72.36, 70.61, 66.48, 49.81, 39.62, 28.57, 27.06, 19.32, 12.03; IR (neat, cm^{-1}): 3500 – 3200, 2932, 2859, 1663, 1426, 1365, 1156, 1093, 1017, 753, 701; HRMS: calculated for $C_{27}H_{39}NO_5Si$ [M+H$^+$] 486.26703, found: 486.26708.

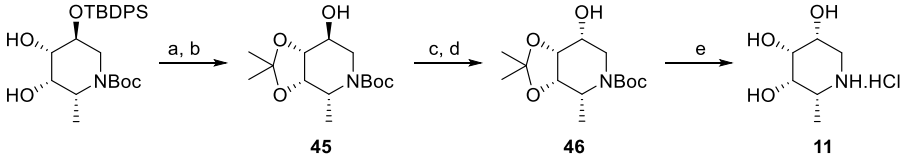


***tert*-Butyl(2*S*,3*R*,4*S*,5*R*)-3,4,5-trihydroxy-2-methylpiperidine-1-carboxylate:** Prepared as described for alcohol **35** and was obtained as a colorless oil (3.8 mmol scale, yield 860 mg, 92%). $[\alpha]_D^{20}$ -20 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 4.28 (dq, *J* = 6.9, 6.9 Hz, 1H), 3.86 (dd, *J* = 6.9, 2.3 Hz, 1H), 3.85 – 3.75 (m, 3H), 3.27 (dd, *J* = 14.2, 1.5 Hz, 1H), 1.46 (s, 9H), 1.24 (d, *J* = 6.9 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 157.57, 81.14, 73.37, 70.95, 67.53, 52.23, 40.96, 28.90, 12.72; IR (neat, cm^{-1}): 3400 – 3200, 2978, 2931, 1660, 1421, 1366, 1157, 1068, 907, 729; HRMS: calculated for $C_{11}H_{21}NO_5$ [M+H$^+$] 248.14925, found: 248.14933.

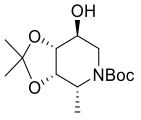


(2*R*,3*S*,4*R*,5*S*)-2-methylpiperidine-3,4,5-triol hydrochloride (10): Prepared as described for **5** and was obtained as a white foam (3.0 mmol scale, yield 549 mg, quant.). $[\alpha]_D^{20}$ +31 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.07 – 3.98 (m, 2H), 3.64 (dd, *J* = 9.9, 3.0 Hz, 1H), 3.50 – 3.40 (m, 2H), 2.85 (t, *J* = 12.0 Hz, 1H), 1.35 (d, *J* = 6.7 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 73.07, 69.79, 64.33, 54.96, 46.05, 14.05; IR (neat, cm^{-1}): 3400 – 3200, 2939, 2805, 1583, 1444, 1386, 1159, 1074, 1022, 999, 708; HRMS: calculated for $C_6H_{13}NO_3$ [M+H$^+$] 148.09682, found: 148.09658.

Scheme 7: Preparation of iminosugar **11**.

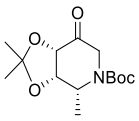


Reagents and conditions: (a) Acetone/2,2-dimethoxypropane, $BF_3{\cdot}EtO_2$, 5 °C; (b) TBAF, THF; (c) Dess-Martin periodinane, DCM; (d) $NaBH_4$, EtOH, -78 °C→RT; (e) HCl (6.0 M in H_2O), MeOH.

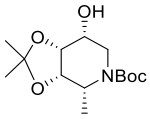


***tert*-Butyl(3*S*,4*R*,7*S*,7*R*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-car b- oxylate (45):** Prepared as described for alcohol **33** from the silyl ether mentioned above (4.6 mmol scale, yield 1.2 g, 85% for two steps). $[\alpha]_D^{20}$ -3.4 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.34 (dd, *J* = 6.9, 5.7 Hz, 1H), 4.13 (qd, *J* = 6.6, 6.6 Hz, 1H), 4.08 (dd, *J* = 6.5, 3.3 Hz, 1H), 3.96

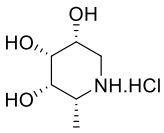
(ddd, *J* = 6.5, 6.5, 3.3 Hz, 1H), 3.60 (dd, *J* = 13.2, 6.5 Hz, 1H), 3.44 (dd, *J* = 13.2, 3.2 Hz, 1H), 2.55 – 2.36 (m 1H), 1.48 (s, 3H), 1.47 (s, 9H), 1.35 (s, 3H), 1.28 (d, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 158.48, 108.84, 80.26, 77.38, 73.99, 68.53, 47.74, 42.97, 28.54, 26.67, 24.66, 17.56; IR (neat, cm^{-1}): 3500 – 3200, 2980, 2934, 1670, 1403, 1367, 1253, 1212, 1166, 1056, 868, 773; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18055, found: 288.18059.



***tert*-Butyl(3*S*,4*R*,7*S*)-2,2,4-trimethyl-7-oxotetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylate:** Prepared as described en route towards compound **36** (scheme 2, step d). The title compound was obtained as a white solid (3.9 mmol scale, yield 919 mg, 82 %). $[\alpha]_D^{20}$ +13 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.89 – 4.71 (m, 2H), 4.52 (d, *J* = 9.0 Hz, 1H), 4.49 – 4.27 (m, 1H), 3.75 (d, *J* = 19.2 Hz, 1H), 1.51 (s, 3H), 1.49 (s, 9H), 1.39 (s , 3H), 1.16 (d, *J* = 7.0 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 204.28, 154.21, 111.07, 81.28, 80.25, 75.55, 74.62, 50.35, 48.12, 28.35, 26.20, 24.49, 12.86; IR (neat, cm^{-1}): 2979, 2935, 1741, 1696, 1409, 1368, 1381, 1162, 1081, 1031, 875; HRMS: calculated for [$C_{14}H_{23}NO_5$ + H]$^+$ 286.16490, found: 286.16488.

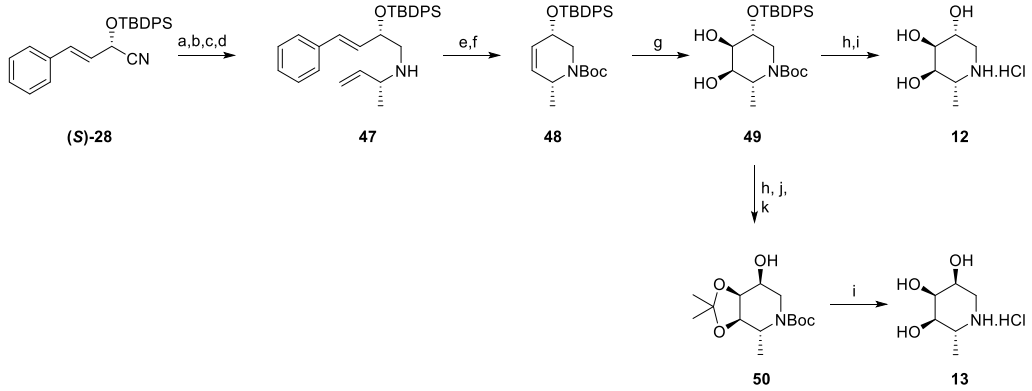


***tert*-Butyl(3*S*,4*R*,7*R*,7*R*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridi-ne-5(4*H*)-carboxylate (46):** Prepared as described for **36** from the ketone mentioned above. Alcohol **46** was obtained (3.8 mmol scale, yield 610 mg, 56%). $[\alpha]_D^{20}$ -4.0 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.44 – 4.34 (m, 2H), 4.06 – 3.96 (m, 1H), 3.88 (dd, *J* = 12.0, 3.8 Hz, 1H), 3.68 – 3.56 (m, 1H), 2.99 (t, *J* = 12.0 Hz, 1H), 2.79 – 2.61 (m 1H), 1.53 (s, 3H), 1.46 (s, 9H), 1.39 (s, 3H), 1.34 (d, *J* = 6.6 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.14, 108.79, 80.13, 76.06, 72.83, 66.36, 47.44, 41.92, 28.49, 26.43, 24.81, 16.78; IR (neat, cm^{-1}): 3500 – 3200, 2979, 2934, 1688, 1406, 1393, 1367, 1250, 1221, 1156, 1061, 1046, 1033, 867; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+ H$^+$] 288.18056, found: 288.18058.

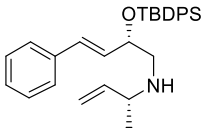


(2*R*,3*S*,4*R*,5*R*)-2-methylpiperidine-3,4,5-triol hydrochloride (11): Prepared as described for **6** from the protected iminosugar **46**. The title compound **10** was obtained as a white foam (1.5 mmol scale, yield 275 mg, quant.). $[\alpha]_D^{20}$ -16 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.2 (ddd, *J* = 4.5, 2.8, 1.4 Hz, 1H), 4.00 – 3.97 (m, 1H), 3.84 (t, *J* = 3.5 Hz, 1H), 3.44 (dd, *J* = 13.7, 2.8 Hz, 1H), 3.42 – 3.37 (m, 1H), 3.23 (dd, *J* = 13.7, 1.4 Hz, 1H), 1.36 (d, *J* = 6.8 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 70.48, 67.33, 66.22, 55.05, 48.26, 14.20; IR (neat, cm^{-1}): 3676, 3348, 3130, 2971, 1559, 1443, 1312, 1277, 1249, 1121, 1056, 1007, 988; HRMS: calculated for [$C_6H_{13}NO_3$ + Na]$^+$ 170.07876, found: 170.07864.

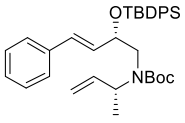
Scheme 8: Preparation of iminosugars **12** and **13**



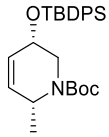
Reagents and conditions: (a) DIBAL-H, -78→5 °C; (b) MeOH, -90 °C; (c) amine **(*R*)-29**, NaOMe, 18 h; (d) $NaBH_4$, 5 °C→RT, 4 h; (e) Boc_2O, 50 °C; (f) Grubbs 1st generation, DCM, reflux; (h) $K_2OsO_4 \cdot 2H_2O$, NMO, acetone/H_2O; (h) TBAF, THF; (i) HCl (6.0 M in H_2O), MeOH; (j) Dess-Martin periodinane, DCM; (k) $NaBH_4$, EtOH, -78 °C→RT.



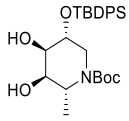
(*S,E*)-*N*-((*R*)-But-3-en-2-yl)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-amine (47): Preparation as described for compound **30** from **(*S*)-27** (3.6 g, 9.1 mmol) and **(*R*)-29·HCl** (3.2 g, 29 mmol, 3.3 eq.) afforded the target compound (3.6 g, 87 %). $[\alpha]_D^{23}$= +95 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.74 – 7.62 (m, 4H) 7.44 – 7.13 (m, 11H), 6.20 (d, *J* = 16.0 Hz, 1H), 6.11 (dd, *J* = 16.0, 6.9 Hz, 1H), 5.57 (ddd, *J* = 17.4, 10.5, 7.5 Hz, 1H), 4.98 (d, *J* = 17.4 Hz, 1H), 4.96 (d, *J* = 10.5 Hz, 1H), 4.43 (td, *J* = 6.8, 5.1 Hz, 1H), 3.04 (m, 1H), 2.78 (dd, *J* = 11.9, 6.7 Hz, 1H), 2.59 (dd, *J* = 11.9, 5.2 Hz, 1H), 1.65, (br, 1H), 1.08 (s, 9H), 1.02 (d, *J* = 6.4Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 142.57, 136.86, 136.10, 136.00, 134.99, 134.19, 134.06, 131.24, 130.91, 129.77, 129.66, 128.47, 127.70, 127.55, 126.56, 114.55, 74.13, 56.32, 53.72, 27.21, 21.84, 19.49. IR (neat, cm^{-1}): 3053, 2957, 2930, 2857, 1471, 1428, 1111, 772, 701; HRMS: calculated for $C_{30}H_{37}NOSi$ $[M+H^+]$ 456.27172, found: 456.27147.



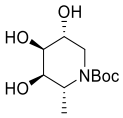
***tert*-Butyl((*R*)-but-3-en-2-yl)((*S,E*)-2-((*tert*-butyldiphenylsilyl)oxy)-4-phenylbut-3-en-1-yl)carbamate:** Prepared as described for compound **31** (3.7 mmol scale, 2.0 g in 98% yield). $[\alpha]_D^{20}$ +76 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.71 (d, *J* = 6.8 Hz, 2H), 7.65 (d, *J* = 6.8 Hz, 2H), 7.48 - 7.16 (m, 11H), 6.19 - 6.00 (m, 2H), 5.78 - 5.64 (m, 1H), 5.06 - 4.80 (m, 2H), 4.47 - 4.32 (m, 1H), 3.59 - 3.39 (m, 1H), 3.39 - 3.20 (m, 1H), 3.14 - 3.00 (m, 1H), 1.43 - 1.19 (m, 9H), 1.07 (s, 9H), 1.05 (d, *J* = 2.5 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 139.08, 136.15, 136.05, 135.51, 131.02, 130.87, 129.81, 129.71, 128.44, 128.36 127.73, 127.58, 126.56, 114.92, 79.63, 73.92, 50.33, 47.00, 28.43, 27.55, 19.43, 17.40; IR (neat, cm^{-1}): 3057, 2931, 2858, 1691, 1391, 1365, 1165, 1110, 1069, 740, 700; HRMS: calculated for $C_{35}H_{45}NO_3Si$ $[M+H^+]$ 556.32415, found: 556.32427.



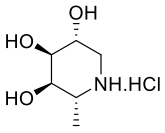
***tert*-Butyl(3*S*,6*R*)-3-((*tert*-butyldiphenylsilyl)oxy)-6-methyl-3,6-dihydropyridine-1(2*H*)-carboxylate (48):** Prepared as described for compound **32** (3.6 mmol scale, 1.6 g, 98%). $[\alpha]_D^{20}$ -12 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.74 - 7.63 (m, 4H), 7.46 - 7.32 (m, 6H), 5.72 - 5.42 (m, 2H), 4.44 - 4.28 (m, 1H), 4.28 - 4.16 (m, 1H), 3.98 (dd, *J* = 12.5, 5.6 Hz, 1H), 2.79 - 2.68 (m, 1H), 1.35 (s, 9H), 1.16 (d, *J* = 6.7 Hz, 3H**),** 1.08 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 153.92, 135.79, 134.91, 130.84, 130.55, 129.87, 128.77, 127.78, 126.59, 79.65, 65.60, 47.40, 43.96, 28.44, 27.21, 19.44, 17.82; IR (neat, cm^{-1}): 2963, 2930, 2858, 1697, 1427, 1410, 1365, 1157, 1109, 1059, 740; HRMS: calculated for $C_{27}H_{37}NO_3Si$ [M+H$^+$] 452.26155, found: 452.26158.



***tert*-Butyl(2*R*,3*R*,4*R*,5*R*)-5-((*tert*-butyldiphenylsilyl)oxy)-3,4-dihydroxy-2-methylpiperidine-1-carbo xy-late (49):** Compound **48** (6.9 g, 15 mmol, 1.0 eq.) was dissolved in a mixture of acetone (70 mL) and water (70 mL) and cooled to -10 °C. *N*-Methylmorpholine-*N*-oxide monohydrate (6.7 g, 50 mmol, 3.3 eq.) and $K_2OsO_4.2H_2O$ (60 mg, 0.16 mmol, 1.1 mol %) were added subsequently. After 24 - 48 hours TLC analysis showed complete conversion of the starting material **48**. The reaction was quenched with an aq. sat. Na_2SO_3 solution (100 mL) and stirred for 30 min. The mixture was diluted with water (100 mL) and extracted with EtOAc (3 x 60 mL). The combined organic layers were washed with aq. 0.60 M HCl, sat. $NaHCO_3$ and brine. After drying (Na_2SO_4), filtering and evaporation of the solvent, the silicagel column chromatography (10%→50%, EtOAc in pentane) afforded the compound **49** as a colorless oil (6.40 g, 1.3 mmol, 86%). $[\alpha]_D^{20}$ -34 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.71 (dd, *J* = 7.9, 1.5 Hz, 2H), 7.66 (dd, *J* = 7.9, 1.5 Hz, 2H), 7.48 - 7.34 (m, 6H), 4.56 - 4.26 (m, 1H), 4.26 - 3.95 (m, 1H), 3.90 (td, *J* = 10.3, 5.2 Hz, 1H), 3.83 - 3.73 (m, 1H), 3.71 (dd, *J* = 8.8, 3.0 Hz, 1H), 2.77 (dd, *J* = 13.3, 10.3 Hz, 1H), 2.40 - 2.09 (m, 2H), 1.49 - 1.19 (s, 9H), 1.13 (d, *J* = 7.3 Hz, 3H), 1.09 (s, 9H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.32, 135.90, 135.81, 133.73, 130.13, 130.10, 128.00, 127.95, 80.03, 73.41, 72.50, 70.21, 48.65, 43.18, 28.40, 27.13, 19.45 , 14.15; IR (neat, cm^{-1}): 3300, 3020, 2254, 1683, 1112, 904, 725; HRMS: calculated for $C_{27}H_{39}NO_5Si$ [M+H$^+$] 486.26703, found: 486.26736.

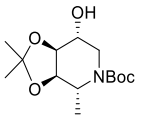


***tert*-Butyl(2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxy-2-methylpiperidine-1-carboxylate:** The silyl ether **49** (1.7 g, 3.5 mmol, 1.0 eq.) was dissolved in THF (30 mL) and TBAF·$3H_2O$ (3.5 g, 11 mmol, 3.1 eq.) was added at room temperature. The reaction was stirred at ambient temperature overnight. TLC indicated complete conversion and the mixture was concentrated. The crude compound was purified by silica gel column chromatography (50%→100%, EtOAc in pentane) to afford the title compound as a clear oil (794 mg, 3.2 mmol, 91%). $[\alpha]_D^{20}$ -40 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 4.36 (m, 1H), 4.09 (dd, *J* = 13.2, 5.6 Hz, 1H), 3.78 – 3.71 (m, 2H), 3.55 (dd, *J* = 9.6, 2.4 Hz, 1H), 2.67 (t, *J* = 12.5 Hz, 1H), 1.46 (s, 9H), 1.16 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 157.28, 81.41, 74.02, 73.41, 68.22, 54.77, 46.40, 28.85, 14.45; IR (neat, cm^{-1}): 3400 - 3200, 2977, 2932, 1665, 1419, 1366, 1167, 1074, 731; HRMS: calculated for $C_{11}H_{21}NO_5$ [M+H$^+$] 248.14925, found: 248.14920.

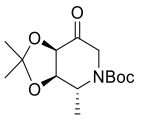


(2*R*,3*R*,4*S*,5*R*)-2-methylpiperidine-3,4,5-triol hydrochloride (12): *N*-Boc-protected-iminosugar from above (502 mg, 2.0 mmol, 1.0 eq.) was dissolved in a mixture of MeOH (20 mL) and aq. 6.0 M HCl (3.0 mL) and the reaction was stirred overnight at room temperature. TLC indicated complete conversion and the mixture was concentrated to afford the title iminosugar **12** as a white foam (317 mg, 1.7 mmol, 86%). $[\alpha]_D^{20}$ +13 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.16 (dd, *J* = 6.4, 4.8

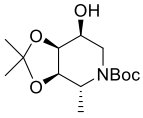
Hz, 1H), 4.02 (dd, *J* = 3.2, 3.2 Hz, 1H), 3.88 (dd, *J* = 9.8, 3.2 Hz, 1H), 3.42 (dq, *J* = 9.8, 6.7 Hz, 1H), 3.36 (m, 1H), 3.20 (dd, *J* = 13.4, 3.2 Hz, 1H), 1.47 (d, *J* = 6.7 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 68.50, 68.15, 66.36, 51.16, 43.95, 14.26; IR (neat, cm^{-1}): 3271, 3022, 2951, 2914, 1583, 1438, 1260, 1086, 1069, 1051, 965; HRMS: calculated for $C_6H_{13}NO_3$ [M+H$^+$] 148.09682, found: 148.09674.



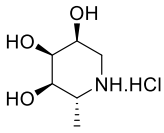
***tert*-Butyl(3*R*,4*R*,7*R*,7*S*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylate:** Preparation from TBDPS-ether **49** as described en route towards compound **40** (9.4 mmol scale), afforded the title compound (2.3 g, 7.9 mmol, 84% over two steps). $[\alpha]_D^{20}$ -38 (*c* = 0.3, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.64 – 4.54 (m, 1H), 4.05 (d, *J* = 5.8 Hz, 1H) 4.00 (t, *J* = 5.8 Hz, 1H), 3.93 (dd, *J* = 13.6, 4.9 Hz, 1H), 3.76 (ddd, *J* = 10.2, 6.4, 4.9 Hz, 1H), 2.53 – 2.19 (br, 1H), 2.82 (dd, J = 13.6, 10.2 Hz, 1H), 1.49 – 1.45 (m, 12H), 1.35 (s, 3H), 1.26 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.42, 108.92, 80.31, 78.13, 77.90, 69.67, 48.21, 41.48, 28.52, 28.24, 26.22, 16.91; IR (neat, cm^{-1}): 3500 – 3200, 2979, 2929, 1694, 1671, 1413, 1367, 1166, 1059, 873.



***tert*-Butyl(3*R*,4*R*,7*R*)-2,2,4-trimethyl-7-oxotetrahydro-[1,3]dioxolo[4,5-c]pyridine-5(4*H*)-carboxylate:** Preparation from TBDPS-ether **49** as described en route towards compound **40** (6.9 mmol scale), afforded the title compound (1.49 g, 5.2 mmol, 76%). $[\alpha]_D^{20}$ +2.2 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.75 – 4.55 (m, 2H), 4.36 (dd, *J* = 6.8, 1.4 Hz, 1H), 4.30 (d, *J* = 6.8 Hz, 1H), 3.65 (d, *J* = 18.5 Hz, 1H), 1.48 (s, 9H), 1.46 (s, 3H), 1.35 (s, 3H), 1.16 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 203.12, 154.74, 111.62, 80.78, 80.22, 75.51, 54.68, 49.15, 28.49, 26.64, 24.95, 15.43; IR (neat, cm^{-1}): 3426, 2981, 2837, 1773, 1728, 1695, 1369, 1252, 1214, 1150, 1050, 870, 753.



***tert*-Butyl(3*R*,4*R*,7*S*,7*S*)-7-hydroxy-2,2,4-trimethyltetrahydro-[1,3]dioxolo[4,5-c]-pyridine-5(4*H*)-carboxylate (50):** Prepared as described for alcohol **40** from the ketone mentioned above (4.3 mmol scale), yielded title compound (1.0 g, 3.6 mmol, 83%, *d.r.* = 93:7). $[\alpha]_D^{20}$ -35 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 4.42 (dd, *J* = 7.3, 4.4 Hz, 1H), 4.33 (dd, *J* = 7.3, 1.8 Hz, 1H), 4.28 – 4.19 (m, 1H), 3.95 (ddd, *J* = 10.7, 4.4, 4.4 Hz, 1H), 3.61 (dd, *J* = 11.9, 4.4 Hz, 1H), 3.07 (t, *J* = 11.9 Hz, 1H), 2.55 – 2.36 (m 1H), 1.50 (s, 3H), 1.47 (s, 9H), 1.38 (s, 3H), 1.17 (d, *J* = 7.2 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 155.21, 108.95, 79.91, 77.42, 71.80, 65.15, 47.70, 42.56, 28.56, 26.30, 24.45, 19.16; IR (neat, cm^{-1}): 3432, 2979, 2936, 1739, 1687, 1401, 1366, 1167, 1049, 877, 750; HRMS: calculated for $C_{14}H_{25}NO_5$ [M+H$^+$] 288.18056, found: 288.18056.



(2*R*,3*R*,4*S*,5*S*)-2-methylpiperidine-3,4,5-triol hydrochloride (13): Prepared as described for **6** from protected iminosugar **50**, to afford **13** as a colorless foam (3.4 mmol scale, yield 613 mg, 98%) in a *d.r.* of 93:7. $[\alpha]_D^{20}$ +21 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 4.13 – 4.17 (m, 1H), 3.96 (ddd, *J* = 11.5, 4.9, 2.6 Hz, 1H) 3.57 (dd, *J* =10.4, 2.5 Hz, 1H), 3.28 (dq, *J* = 6.5, 2.5 Hz, 1H), 3.20 (dd, *J* = 12.0, 4.9 Hz, 1H), 3.05 (t, *J* = 12.0 Hz, 1H), 1.33 (d, *J* = 6.5 Hz, 3H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 69.93, 69.61, 64.70, 49.08, 41.22, 14.16; IR (neat, cm^{-1}): 3400 - 3200, 2942, 2810, 2512, 1457, 1044, 1016, 962, 818, 707; HRMS: calculated for $C_6H_{13}NO_3$ [M+Na$^+$] 148.09682, found: 148.09658.

Biological assays:

Materials. Cyclophellitol β-aziridine ABP JJB75 was synthesized as described earlier.[18] Gaucher patients were diagnosed on the basis of reduced GBA activity and demonstration of an abnormal genotype.[38] Cell lines were cultured in HAMF12-DMEM medium (Invitrogen) supplied with 10% (v/v). Spleens from a normal subject and a patient suffering from type 1 Gaucher were collected after splenectomy and immediately stored frozen.

Molecular cloning and recombinant expression. The coding sequences of *H. sapiens* FUCA1 (NCBI reference sequence XM_005245821.1, using forward primer 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCACCATGCGGGCTCCGGGGATG-3' and reverse primer 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACTTCACTCCTGTCAGCTTTAT-3'), and of *H. sapiens* FUCA2 (NCBI reference sequence NM_032020.4, using forward primer 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCACCACCATGCGGCCCCAGGAGCTC-3' and reverse primer 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGATCACATTAGTCAGGGCTA-3') were amplified via PCR and cloned into pDNOR-221 and thereafter sub-cloned in pcDNA™-DEST40 vector using the Gateway system (Invitrogen). Correctness of all constructs was verified by sequencing. Confluent COS-7 cells were transfected with pcDNA3.1 empty vector (Mock) or the vector with described insert in conjunction with FuGENE (Roche). After 72 hours, medium isolated and frozen at -80 °C and cells were harvested by scraping in 25 mM potassium phosphate buffer (pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)). After determination of the protein concentration (BCA kit, Pierce), lysates were aliquoted and frozen at -80 °C.

Enzyme activity assays. The enzymatic activity of α-L-fucosidase was assayed at 37 °C by incubating with 1.5 mM 4-methylumbelliferyl-α-L-fucopyranoside as substrate in 150 mM McIlvaine buffer, pH 4.5, supplemented with 0.1% (w/v) BSA. To determine the apparent IC_{50} value, COS-7 cell lysate containing over-expressed human recombinant FUCA1 was pre-incubated with a range of inhibitor dilutions for 30 min at 37 °C. The reaction was quenched by adding excess NaOH-glycine (pH 10.6), after which fluorescence was measured with a fluorimeter LS55 (Perkin Elmer) using λ_{EX} 366 nm and λ_{EM} 445 nm.

***In vitro* labeling and SDS-PAGE analysis.** All pre-incubations and ABP labeling-reactions occurred for 30 min at 37 °C. Total lysates (50 µg), medium (500 µg) or purified protein preparations (5.0 µg) were labeled with 1.0 µM JJB256 **1** or JJB244 **2**, dissolved in 150 mM McIlvaine buffer, pH 4.5, incubating for 30 min at 37 °C. For ABPP, protein preparations were pre-incubated with compounds **4** (100 µM), **6** (100 µM), **7**-**13** (1 or 5 mM, specified in the main text) prior to the addition of 100 nM fluorescent ABPs. Influence of pH on ABP labeling involved pre-incubation at pH 3-10 prior to addition of 100 nM ABP **1**. Direct labeling of retaining β-glucosidases occurred at pH 5.0 in conjunction with 100 nM ABP JJB75. Samples were then denatured with 5× 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), boiled for 4 min at 100 °C, and separated by gel electrophoresis on 10% (w/v) SDS-PAGE gels running continuously at 90 V[13, 14]. Wet slab-gels were scanned on fluorescence using a Typhoon Variable Mode Imager (Amersham Biosciences) using λ_{EX} 488 nm and λ_{EM} 520 nm (band pass filter 40 nm) for green fluorescent JJB256 **1** and λ_{EX} 532 nm and λ_{EM} 610 nM (band pass filter 30 nm) for red fluorescent JJB244 **2** and JJB75.

***In vivo* labeling.** The appropriate ethics committee for animal experiments approved all experimental procedures. Wild-type C57Bl/6J male mice were obtained from Harlan and fed a commercially available lab diet (RMH-B; Hope Farms). Four C57BL/6J mice were injected intraperitoneally with 100 µL sterile PBS (vehicle) or PBS containing 10, 100, or 1000 pmol of ABP **1** (about 0.20 µg kg−1, 2 µg kg−1, and 20 µg kg−1, respectively). At 2 h post-administration, urine was collected and the mice were anesthetized with FFM mix (25/25/50 fentanylcitrate/ 67idazolam/H_2O), blood was collected and perfused with PBS at 3.0 mL min^{-1}. Then urine and organs were collected and directly frozen in liquid nitrogen. Homogenates were made in 25 mM potassium phosphate buffer, pH 6.5, supplemented with 0.10% (v/v) Triton X-100 and protease inhibitor cocktail (Roche). After determination of protein concentration (BCA kit, Pierce), 50 µg total protein was incubated with 100 nM red fluorescent JJB75 and analyzed by SDS-PAGE. As controls, matching tissue homogenates of vehicle-treated animals were concomitantly labeled with 1.0 µM ABP **1** and 100 nM JJB75 prior to SDS-PAGE.

ABP pulldown and LC-MS/MS analysis. Gaucher spleen lysate (1.0 g) was cut with a sterile scalpel, mixed with 4.0 mL 25 mM potassium phosphate buffer (pH 6.5, supplemented with 0.10% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)) and sonicated three times for 10 seconds at 100% strength and kept on ice. After 2 min, procedure repeated twice. Lysate was then centrifuged for 10 min at 10,000 *g*, supernatant collected carefully, protein concentration determined and 200 mg total protein was then incubated with either 0.10% (v/v) DMSO, ABP **1** or ABP **3**, or firstly with 10 µM ABP **1** followed by 10 µM of ABP **3**, each step taking 30 min at 37 °C, in a total volume of 1.0 mL McIlvaine buffer, pH 4.5. Glycosylated biomolecules were enriched by using 1.0 mL ConA-Sepharose per sample, according to manufacturer's instructions (Amersham Pharmacia Biotech AB, Sweden) and subsequently denatured by the presence of 2.0% (w/v) SDS and boiling for 5 min at 100 °C. From here on, samples were prepared for MS as published earlier.[14] After desalting on StageTips, peptides were analyzed with a 2 h gradient of 5-25% CAN on nano-LC, hyphenated to an LTQ-Orbitrap and identified via the Mascot protein search engine.[14]

X-ray Crystallography. Recombinant *Bacteroides thetaiotaomicron* 2970 α-L-fucosidase (*Bt*Fuc2970) was prepared as described previously.[28] Protein crystals were obtained through hanging (Crystals soaked with compound **5**) or sitting (crystals soaked with compound **4**) drop vapor diffusion (for further details see PDB file headers). Compounds **4** and **5** were dissolved in crystallization mother liquor at a concentration of **5** or 20 mM respectively and added to crystallization drops containing crystals of *Bt*Fuc2970 in a 1:1 ratio. After *ca*. 1 h soaking with ligands, crystals were fished into cryo-protectant solutions (mother liquor supplemented with 20% v/v glycerol) and cryo-cooled in liquid nitrogen. Diffraction data were collected at Diamond Light Source. Diffraction images were indexed and integrated using MOSFLM[29] (**4**) or XDS[39] (**5**) and scaled and merged using AIMLESS.[40] Crystals grew in an almost isomorphous space group to PDB entry 4JFV, and coordinates from this entry were used directly to obtain a starting model for refinement. Iterative stages of model-building (COOT[41]) and maximum-likelihood refinement (REFMAC5[42]) were conducted to yield final models. Maximum-likelihood restraints for compounds **4** and **5** were created using the PRODRG online server[43] and link restraints generated using JLIGAND.[44] X-ray crystallographic data statistics are available in supplemental Table 3.

Table 3. X-ray crystallographic data table

	*Bt*Fuc2970-**5** PDB: 4WSJ	*Bt*Fuc2970-**4** PDB:4WSK
data collection		
Beamline/Date	Diamond i03 18/10/14	Diamond i03 02/02/14
Wavelength (Å)	0.97625	0.97625
Cell dimensions		
a, *b*, *c* (Å)	55.5, 187.0, 98.2	55.6, 186.5, 98.2
α, β, γ (°)	90, 94.2, 90	90, 94.2, 90
Resolution (Å)	93.5-1.64	62.17-1.92
R_{merge}	0.058(0.62)	0.095(0.86)*
$I / \sigma I$	11.2(2.0)	7.0(1.8)
Completeness (%)	98.1(98.5)	96.9(96.7)
Redundancy	4.1(4.3)	4.0(3.7)
Wilson B value	20.4	26.1
Refinement		
Resolution (Å)	97.9-1.64	97.9-1.71
No. reflections	328440	292076
R_{work} / R_{free}	0.16 / 0.19	0.18 / 0.23
R.m.s. deviations		
Bond lengths (Å)	0.019	0.019
Bond angles (°)	1.78	1.81
Ramachandran Statistics (%)		
Preferred	96.7	95.8
Allowed	2.4	3.1
Outliers	0.9	1.1

Values in parentheses are for highest-resolution shell.

3.5 References and notes

[1] J. Intra, M.-E. Perotti, G. Pavesi and D. Horner, *Gene* **2007**, 392, 34-46.

[2] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.* **2014**, 42, D490-D495.

[3] S. W. Liu, C. S. Chen, S. S. Chang, K. K. T. Mong, C. H. Lin, C. W. Chang, C. Y. Tang and Y. K. Li, *Biochemistry* **2009**, 48, 110-120.

[4] L. Guillotin, P. Lafite and R. Daniellou, *Biochemistry* **2014**, 53, 1447-1455.

[5] B. Cobucci-Ponzano, M. Mazzone, M. Rossi and M. Moracci, *Biochemistry* **2005**, 44, 6331-6342.

[6] F. A. Shaikh, A. L. van Bueren, G. J. Davies and S. G. Withers, *Biochemistry* **2013**, 52, 5857-5864.

[7] D. E. Koshland, *Biol. Rev.* **1953**, 28, 416-436.

[8] T. W. Liu, C. W. Ho, H. H. Huang, S. M. Chang, S. D. Popat, Y. T. Wang, M. S. Wu, Y. J. Chen and C. H. Lin, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 14581-14586.

[9] J. S. O'Brien, P. J. Willems, H. Fukushima, J. R. de Wet, J. K. Darby, R. D Cioccio, M. L. Fowler and T. B. Shows, *Enzyme* **1987**, 38, 45-53.

[10] K. Phopin, W. Nimlamool, L. J. Lowe-Krentz, E. W. Douglass, J. N. Taroni and B. S. Bean, *Mol. Reprod. Dev*. **2013**, 80, 273-285.

[11] D. G. Hildebrand, S. Lehle, A. Borst, S. Haferkamp, F. Essmann and K. Schulze-Osthoff, *Cell Cycle* **2013**, 12, 1922-1927.

[12] M. G. Giardina, M. Matarazzo, R. Morante, A. Lucariello, A. Varriale, V. Guardasole and G. De Marco, *Cancer* **1998**, 83, 2468-2474.

[13] B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, 77, 383-414.

[14] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, 51, 12529-12533.

[15] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, 136, 11622-11625.

[16] M. J. Niphakis and B. F. Cravatt, *Ann. Rev. Biochem.* **2014**, 83, 341-377.

[17] Y. Harrak, C. M. Barra, A. Delgado, A. R. Castano and A. Llebaria, *J. Am. Chem. Soc.* **2011**, 133, 12079-12084.

[18] K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, 2014, 6030-6043.

[19] O. David, W. J. N. Meester, H. Bieräugel, H. E. Schoemaker, H. Hiemstra and J. H. van Maarseveen, *Angew. Chem. Int. Ed.* **2003**, 42, 4373-4375.

[20] H. Griengl, N. Klempier, P. Pöchlauer, M. Schmidt, N. Shi and A. A. Zabelinskaja-Mackova, *Tetrahedron* **1998**, 54, 14477-14486.

[21] A. D. Campbell, T. M. Raynham and R. J. K. Taylor, *Synthesis* **1998**, 1707-1709.

[22] G. W. J. Fleet, A. N. Shaw, S. V. Evans and L. E. Fellows, *J. Chem. Soc., Chem. Commun.* **1985**, 841-842.

[23] E. G. J. C. Warmerdam, A. M. C. H. van den Nieuwendijk, J. Brussee, A. van der Gen and C. G. Kruse, *Recl. Trav. Chim. Pays-Bas* **1996**, 115, 20-24.

[24] E. Moreno-Clavijo, A. T. Carmona, A. J. Moreno-Vargas, L. Molina and I. Robina, *Curr. Org. Synth.* **2011**, 8, 102-133.

[25] a) Y. Bleriot, D. Gretzke, T. M. Krulle, T. D. Butters, R. A. Dwek, R. J. Nash, N. Asano and G. W. Fleet, *Carbohydr. Res.* **2005**, 340, 2713-2718. b) A. Bordier, P. Compain, O. R. Martin, K. Ikeda and N. Asano, *Tetrahedron: Asymm.* **2003**, 47-51. c) P. Jakobsen, J. M. undbeck, M. Kristiansen, J. Breinholt, H. Demuth, J. Pawlas, M. P.Torres Candela, B. Andersen, N. Westergaard, K. Lundgren and N. Asano, *Bioorg. Med. Chem*. **2001**, 9, 733-744. d) J. Streith, A. Boiron, J.-L. Paillaud, E.-M. Rodriguez-Perez, C. Strehler, T. Tschamber and M. Zehnder, *Helv. Chim. Acta* **1995**, 78, 61-72.

[26] M. T. Moran, J. P. Schofield, A. R. Hayman, G. P. Shi, E. Young and T. M. Cox, *Blood* **2000**, 96, 1969-1978.

[27] L. I. Willems, J. Jiang, K.-Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. N. Beenakker, S. P. Schröder, J. M. F. G. Aerts, G. A. van der Marel, J. D. C. Codée and H. S. Overkleeft, *Chem. Eur. J.* **2014**, 20, 10864-10872.

[28] A. Lammerts van Bueren, A. Ardèvol, J. Fayers-Kerr, B. Luo, Y. Zhang, M. Sollogoub, Y. Blériot, C. Rovira and G. J. Davies, *J. Am. Chem. Soc.* **2010**, 132, 1804-1806.

[29] S. Mcnicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallogr. Sect. D: Biol.* **2011**, 67, 386-394.

[30] D. A. Evans, K. T. Chapman and J. Bisaha, *J. Am. Chem. Soc.* **1988**, 110, 1238-1256.

[31] F. G. Hansen, E. Bundgaard and R. Madsen, *J. Org. Chem.* **2005**, 70, 10139-10142.

[32] M. Verdoes, U. Hillaert, B. I. Florea, M. Sae-Heng, M. D. P. Risseeuw, D. V. Filippov, G. A. van der Marel, H. S. Overkleeft, *Bioorg. Med. Chem. Lett.* **2007**, 17, 6169–6171

[33] M. Verdoes, B. I. Florea, U. Hillaert,L. I. Willems, W. A. van der Linden, M. Sae-Heng, D. V. Filippov, A. F. Kisselev, G. A. van der Marel, H. S. Overkleeft, *ChemBioChem* **2008**, 9, 1735–1738

[34] a) E. Blocka, M. J. Bosiak, M. Welniak, A. Ludwiczak, A. Wojtczak *Tetrahedron: Asymm.* **2014**, 25, 554-562. b) Y. Hamashima, D. Sawada, H. Nogami, M. Kanai, M. Shiwasaki *Tetrahedron* **2001**, 57, 805-814.

[35] N. Kurono, T. Yoshikawa, M. Yamasaki, T. Ohkuma, *Org. Lett.*, **2011** , 13, 1254-1257.

[36] T. Moriwake, S. Hamano, S. Saito, S. Torii, *Chem. Lett.* **1987**, 2085-2088.

[37] M. Pallavicini, E. Valoti, L. Villa, O. Piccolo *Tetrahedron: Asymm.* **2000**, 11, 4017-4025.

[38] I. M. Aerts and S. van Weely, *Human Mutation* **1997**, 10, 348-358.

[39] A. G. W. Leslie, *Acta Crystallogr., Sect. D: Biol.* **2006**, 62, 48-57.

[40] W. Kabsch, *Acta Crystallogr., Sect. D: Biol.* **2010**, 66, 125-132.

[41] P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol.* **2010**, 66, 486-501.

[42] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr., Sect. D: Biol.* **2011**, 67, 355-367.

[43] A. W. Schuttelkopf and D. M. F. van Aalten, *Acta Crystallogr., Sect. D: Biol.* **2004**, 60, 1355-1363.

[44] A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, *Acta Crystallogr. Sect. D: Biol.* **2012**, 68, 431-440.

3.6 Supporting Information

Table S1 (A): Proteins identified by affinity purification with the biotinylated Probe **3**(JJB243, on bead digest, LC-MS/MS analysis and Mascot search engine match to the human Uniprot database (Jan. 2015).

nr	accession	protein description	score	cover	emPAI
1	P04264	Keratin, type II cytoskeletal 1, KRT1	677	18	0,62
2	P13645	Keratin, type I cytoskeletal 10, KRT10	580	22	0,92
3	P05165	Propionyl-CoA carboxylase, mitochondrial, PCCA	541	17	0,38
4	P35527	Keratin, type I cytoskeletal 9, KRT9	374	23	0,44
5	P35908	Keratin, type II cytoskeletal 2 epidermal, KRT2	366	10	0,34
6	P11498	Pyruvate carboxylase, mitochondrial, PC	262	8	0,19
7	P04066	Tissue alpha-L-fucosidase, FUCA1	210	17	0,35
8	H0YA55	Serum albumin, ALB	192	5	0,2
9	U3KQK0	Histone H2B, HIST1H2BN	191	9	0,39
10	H7C469	Uncharacterized protein	144	12	0,25
11	H0YFX9	Histone H2A, HISTH2A	142	21	0,82
12	P19013	Keratin, type II cytoskeletal 4, KRT4	132	4	0,12
13	Q14956	Transmembrane glycoprotein NMB, GPNMB	131	3	0,11
14	P81605	Dermcidin OS, DCD	96	13	0,70
15	F8W6P5	LVV-hemorphin-7, HBB	93	26	0,85
16	P31025	Lipocalin-1, LCN1	71	6	0,18
17	P69905	Hemoglobin subunit alpha, HBA1	71	37	0,83
18	A5A3E0	POTE ankyrin domain family member F, POTEF	70	2	0,03
19	F5GWP8	Keratin, type I cytoskeletal 17, KRT17	69	5	0,17

Table S1 (B): Proteins identified after competitive experiment between the fluorescent fucosidase probe **1** and the biotin probe **3**, followed by affinity purification, on bead digest, LC-MS/MS analysis and Mascot search engine match to the human UniProt database (Jan. 2015)

nr	accession	protein description	score	cover	emPAI
1	P05165	Propionyl-CoA carboxylase, mitochondrial, PCCA	426	22	0,69
2	P13645	Keratin, type I cytoskeletal 10, KRT10	418	21	1,03
3	E9PS68	Pyruvate carboxylase, mitochondrial, PC	276	16	0,47
4	Q14956	Transmembrane glycoprotein NMB, GPNMB	236	3	0,11
5	A0A087WY73	Proline-rich protein 4, PRR4	227	14	0,44
6	P35908	Keratin, type II cytoskeletal 2 epidermal, KRT2	218	12	0,48
7	P04264	Keratin, type II cytoskeletal 1, KRT1	218	12	0,4
8	P35527	Keratin, type I cytoskeletal 9, KRT9	213	11	0,29
9	U3KQK0	Histone H2B, HIST1H2BN	147	9	0,39
10	H0YI76	Keratin, type II cytoskeletal 5, KRT5	126	12	0,31
11	H7C469	Uncharacterized protein	97	5	0,12
12	H0YFX9	Histone H2A, H2A	84	21	0,35
13	P81605	Dermcidin, DCD	64	10	0,3
14	A8MUF7	Hemoglobin subunit epsilon, HBE1	48	12	0,37
15	P69905	Hemoglobin subunit alpha, HBA1	47	11	0,22
16	F8W0V3	Extracellular glycoprotein lacritin, LACRT	43	6	0,26

Table S1 (C): Proteins identified after no probe control experiment, aspecific background stickiness of proteins to the paramagnetic beads, followed by affinity purification, on bead digest, LC-MS/MS analysis and Mascot search engine match to the human UniProt database (Jan. 2015)

nr	accession	protein description	score	cover	emPAI
1	P04264	Keratin, type II cytoskeletal 1, KRT1	1166	35	1,91
2	P35527	Keratin, type I cytoskeletal 9, KRT9	972	39	1,67
3	P05165	Propionyl-CoA carboxylase, mitochondrial, PCCA	701	28	0,98
4	P35908	Keratin, type II cytoskeletal 2 epidermal, KRT2	427	20	0,71
5	P13645	Keratin, type I cytoskeletal 10, KRT10	342	12	0,46
6	P02533	Keratin, type I cytoskeletal 14, KRT14	341	18	0,54
7	P13647	Keratin, type II cytoskeletal 5, KRT5	283	13	0,43
8	P11498	Pyruvate carboxylase, mitochondrial, PC	282	7	0,16
9	P02538	Keratin, type II cytoskeletal 6A, KRT6A	270	12	0,38
10	P48668	Keratin, type II cytoskeletal 6C, KRT6C	257	12	0,38
11	P08779	Keratin, type I cytoskeletal 16, KRT16	252	15	0,55
12	P04259	Keratin, type II cytoskeletal 6B, KRT6B	222	9	0,31
13	Q7Z794	Keratin, type II cytoskeletal 1b, KRT77	179	4	0,11
14	P81605	Dermcidin, DCD	136	23	1,22
16	F8W6P5	LVV-hemorphin-7, HBB	82	36	1,52
17	Q14956	Transmembrane glycoprotein NMB, GPNMB	73	3	0,11
18	P31025	Lipocalin-1, LCN1	48	6	0,18
19	P69905	Hemoglobin subunit alpha, HBA1	47	11	0,22
20	A0A075B6N7	Ig alpha-2 chain C region, IGHA2	46	3	0,09
21	Q5T8M7	Actin, alpha skeletal muscle, ACTA1	43	3	0,09

Protein score is the Mascot score calculated for the peptide matches of the protein to the human protein database, protein mass is given in kDa, protein coverage is the percentage of the amino acid sequence that has been identified, emPAI value gives an indication of the abundancy or relative concentration of the protein in the LC-MS run.

4

Comparing *N*-alkyl and *N*-acyl cyclophellitol aziridine and its isomer as activity-based glycosidase probes

Jianbing Jiang, Thomas J. M. Beenakker, Wouter W. Kallemeijn, Gijsbert A. van der Marel, Hans van den Elst, Jeroen D. C. Codée, Johannes M. F. G. Aerts and Herman S. Overkleeft, *Chemistry–A European Journal,* **2015**, *21*, 10861-10869.

4.1 Introduction

Glycosidases constitute a large family of hydrolytic enzymes expressed throughout all kingdoms of life and essential in a myriad of biological processes. From a substrate point of view glycosidases can be roughly classified as *endo*-glycosidases (cleaving within an oligo/polysaccharide to yield an oligosaccharide) and *exo*-glycosidases (recognizing a single monosaccharide at the non-reducing end of an oligosaccharide/glycoconjugate). All glycosidases hydrolyze their substrate glycosidic bonds with either retention or inversion of configuration with respect to the anomeric configuration of the released glycoside.[1] This difference in enzymatic hydrolysis, which is perhaps irrelevant with respect to the produced hemi-acetals that will undergo post-hydrolysis anomerization, is caused by the distinct mechanisms employed by the two enzyme families. Inverting glycosidases directly substitute, in an S_N2 fashion, the aglycon of the glycosidic bond and do so by simultaneously activating the anomeric leaving group (protonation by a catalytic acid/base residue present in the enzyme active site) and activation of a water molecule residing in the active site by a catalytic

base. Retaining glycosidases in contrast employ a two-step double displacement mechanism. In this mechanism, first proposed by Koshland,[2] S_N2 displacement of the aglycon (activated through protonation by the general acid/base residue) via attack of the enzyme catalytic nucleophile yields a covalent glycosyl-enzyme intermediate and concomitant expulsion of the aglycon. In a second step and after entry of a water molecule in the enzyme active site this enzyme-glycosyl intermediate is hydrolyzed to yield the carbohydrate hemi-acetal.

The occurrence of a covalent intermediate in the catalytic cycle of retaining glycosidases invites the development of tagged, covalent inhibitors, termed activity-based probes (ABPs), and thereby monitoring these enzymes in cell extracts, *in situ* and *in vivo* by means of activity-based protein profiling (ABPP) experiments. Indeed, the development of retaining glycosidase ABPs has met with considerable more success than identification of related probes for inverting glycosidases.[3] The first reported conceptual design for retaining glycosidase ABPs is from the laboratories of Withers, Vocadlo and Bertozzi, who employed electron-deficient, fluorine-modified carbohydrates (either substitution of the 2-OH or the 5-H for fluorine).[4] The work described in this chapter focused on the natural product mechanism-based β-glucosidase inhibitor, cyclophellitol.[5]

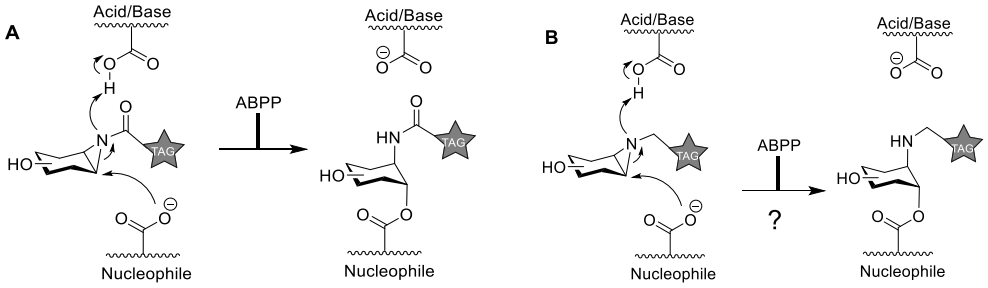


Figure 1. A) Covalent and irreversible inhibition of retaining glycosidases by *N*-acyl cyclophellitol aziridine ABPs. B) Proposed new retaining glycosidase ABP design based on *N*-alkyl cyclophellitol aziridines.

The first inroads into activity-based glycosidase profiling concerned installment of fluorophores at C6 (glucopyranose numbering) of cyclophellitol.[6] The resulting compounds proved highly potent and very specific for GH30 (CAZypedia nomenclature[7]) human acid glucosylceramidase (GBA) over the other three human retaining β-glucosidases (GBA2, GBA3 and LPH). They also proved more effective than the corresponding 2-deoxy-2-fluoroglucosides in subsequent studies, underscoring the usefulness of the cyclophellitol scaffold for ABP design for retaining glycosidases.[8] With the aim to arrive at in-class (with respect to the retaining glycosidase family at hand) broad spectrum ABPs, the epoxide in cyclophellitol was in subsequent studies substituted for aziridine and the reporter moiety was installed at the aziridine nitrogen through *N*-acylation. In this fashion, effective ABPs were obtained for GH30

β-glucosidases[9], GH27 α-galactosidases (GLA)[10] and GH29 α-fucosidases (FUCA).[11] In each case, as depicted in Figure 1A, after attack from the catalytic nucleophile present in the glycosidase active site of enzyme, the detectable probe-enzyme complex is formed. Probe specificity appears configuration dependent, with the configuration of cyclitol aziridine conferring selectively towards the corresponding retaining glycosidases.

In the studies on *N*-acyl aziridine-based glycosidase probes, it was observed that the *N*-acyl moiety are relatively hydrolysis-prone, which puts some strain on their synthesis, purification and handling. With the aim to establish whether *N*-alkyl cyclophellitol aziridines would be valid scaffolds for retaining glycosidase ABP design and following the literature precedent set by Tatsuma and co-workers[12], a set of cyclophellitol aziridine analogues were recently evaluated varying in *N*-substitution as inhibitors of human retaining β-glucosidases. It was found that *N*-pentanyl cyclophellitol aziridine inhibits GBA, GBA2 and GBA3 at least equally potent as the corresponding *N*-pentanoyl cyclophellitol aziridine.[13] Based on these observations the effectiveness, both in synthesis and in activity-based glycosidase profiling, of a set of *N*-alkyl aziridine and *N*-acyl aziridine probes targeting GH30 β-glucosidases and GH29 α-fucosidases (Figure 1B) was evaluated, The results of these studies are presented in this chapter and the structures of the compounds studied here are depicted in Figure 2.

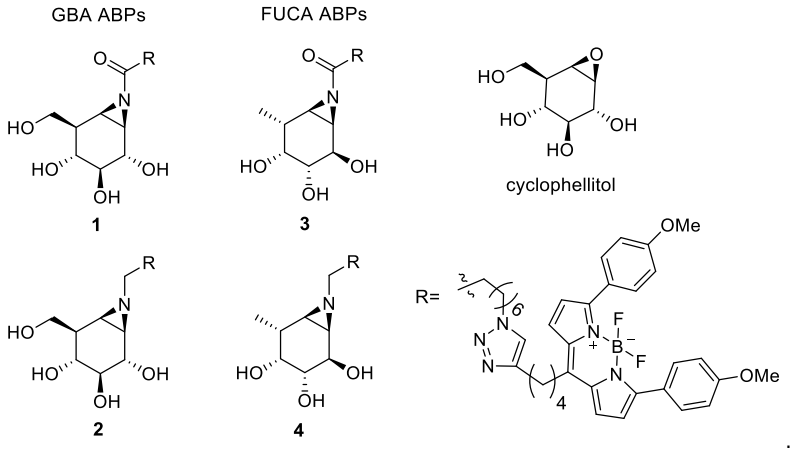
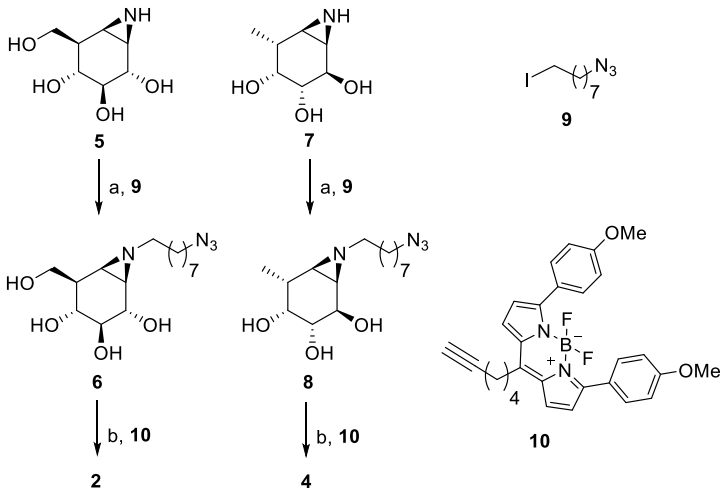


Figure 2. Structures of the natural product, cyclophellitol, and *N*-alkyl/acyl aziridines derived thereof and that are subject of the here presented studies.

4.2 Results and Discussion

The preparation of *N*-acyl cyclophellitol aziridine ABPs **1** and **3** (Figure 2) is described in previous literature reports on the development of these ABPs.[11,14] In order to produce the corresponding *N*-alkyl cyclophellitol aziridines **2** and **4** (Scheme 1), sufficient quantities of the unmodified aziridines **5** and **7** were prepared. These were condensed with

1-azido-8-iodooctane **9** in DMF with K_2CO_3 as base, giving azido-aziridine **6** and **8**. The target ABPs **2** and **4** were synthesized by conjugation of alkyl-BODIPY **10** followed by HPLC purification and lyophilization in moderate yields.



Scheme 1. Synthesis of alkyl aziridine ABPs **2** and **4**. (a) K_2CO_3, DMF, 80 °C, **6**: 49%, **8**: 36%; (b) $CuSO_4$ (1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), DMF, **2**: 30%, **4**: 9%.

At a first glance there appears not much difference between the efficiency (quantity, yield) of the synthesis of *N*-acyl aziridine **1** (5.8 mg, 6%)[14a] and its alkyl counterpart **2** (26 mg, 15%). The same holds true when comparing *N*-acyl aziridine **3** (15 mg, 4%)[11] and *N*-alkyl aziridine **4** (1.8 mg, 3%). However, the HPLC purification of cyclophellitol *N*-acyl aziridines has to be performed at neutral (H_2O, MeCN) or slightly basic (25 mM NH_4HCO_3 in H_2O/MeCN) pH, and hydrolysis of the *N*-acyl aziridine is always at risk during lyophilization and use in ABPP. The *N*-alkyl aziridines in contrast can be purified by standard silica gel column chromatography, and are stable during standard HPLC purification conditions (50 mM aqueous NH_4HCO_3) and LC-MS detection in the presence of 1.0 % TFA, as well as during evaporation or lyophilization.

After having the cyclophellitol aziridine ABPs **1**-**4** in hand, their inhibitory potency was determined towards GBA (**1**, **2**) and FUCA (**3**, **4**) as well as their potential as ABPs for these three retaining glycosidases in a set of head-to-head comparative experiments, focusing in each case on the difference between *N*-acyl- and *N*-alkyl substitution on the cyclophellitol aziridines of the same configuration. For this purpose, recombinant GBA and FUCA enzymes were expressed in COS-7 cells.

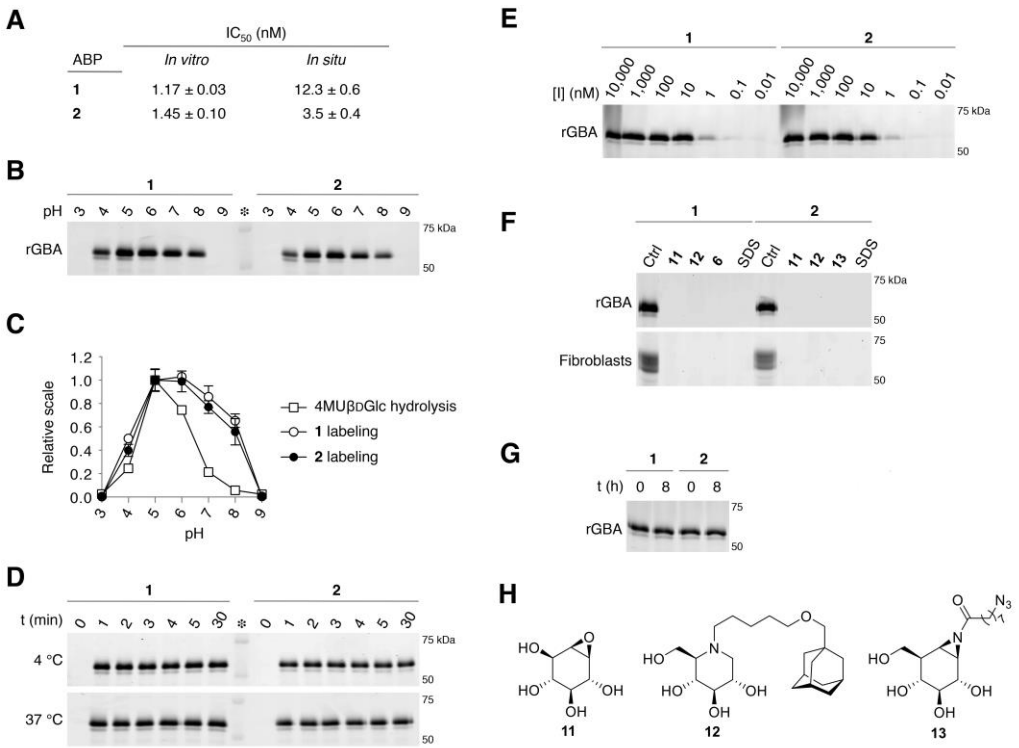


Figure 3. A) Apparent IC_{50} values of β-D-glucopyranose ABPs **1** and **2**, determined *in vitro* against rGBA and *in situ* in wild-type GBA fibroblasts. Data are averages of two separate experiments in duplo, and error bars depict standard deviation. B) ABP-labeling of rGBA at various pH. C) Quantification of ABP-labeling in (B) compared to hydrolysis of 4-methylumbelliferyl-β-D-glucopyranoside at varying pH. Data are averages of two separate experiments in duplo, and error bars depict standard deviation. D) Labeling of 100 fmol rGBA with 100 fmol ABP at 4 °C (top) and 37 °C (bottom). E) Detection limit of ABP-emitted fluorescence. rGBA (10 pmol) was labeled with varying concentrations of ABP **1** or **2**. F) ABP labeling of rGBA (top) or wild-type GBA fibroblast lysate (bottom) competed with inhibitors **6** (Scheme 1)**, 11**, **12**, **13** or SDS. G) Stability of ABP-rGBA nucleophile adducts after a 0 and 8 hour chase in the dark, with hourly extensive washing. H) Structure of compounds CBE **11**, **12** and **13**.

The first comparative study was performed on acyl aziridine ABP **1** and alkyl aziridine ABP **2** as inhibitors and ABPs of recombinant human acid β-glucoceramidase (GBA). Both aziridines inhibit GBA activity in the nanomolar range, both *in vitro* and *in situ* (IC_{50} values in Figure 3A). The observed values for the known *N*-acyl aziridine **1** correspond with those reported previously.[9] Both compounds exhibit a similar pH-dependent activity (Figure 3B) and effective labeling of recombinant purified GBA is observed at concentrations as low as 1 nanomolar (Figure 3E). Figure 3C depicts a quantification curve of GBA labeling with **1** or **2** as derived from the data presented in Figure 3B offset against GBA-mediated hydrolysis at various pH of the fluorogenic substrate, 4-methylumbelliferyl-β-D-glucopyranoside. These data show that

N-alkyl-aziridine **2**, as is *N*-acyl-aziridine **1**, is able to covalently modify GBA up to slightly basic conditions, whereas the pH optimum of GBA as reflected by substrate hydrolysis is at around pH 5. Labeling efficiency of *N*-acyl aziridine **1** and *N*-alkyl aziridine **2** at 100 femtomole were determined at both 4 °C and 37 °C (Figure 3D). Labeling of GBA was almost complete within one minute with both compounds, and thus the effective GBA labeling is too fast to allow accurate determination of kinetic constants. In competitive activity-based protein profiling experiment, recombinant GBA and wild type fibroblast were pre-incubated with the mechanism-based inhibitor, conduritol B epoxide (CBE, **11**), the competitive inhibitor, *N*-(adamantanemethyloxypentyl)- deoxynojirimycin **12,** as well as acyl aziridine (JJB103, **13**) and alkyl aziridine (JJB339, **6**), the latter two compounds being the non-fluorescent precursors towards the synthesis of compounds **1** and **2**, respectively. All compounds proved able to completely block the GBA activity in the conditions applied (Figure 3F). Finally, labeling intensity of GBA treated with either **1** or **2** did not change after an 8 hour chase (Figure 3G), which was performed in darkness with hourly continuous extensive washing and the control adducts were denatured and frozen at -20 °C. It was suggested that both **1** and **2** formed stable ABP-GBA nucleophile adducts. Based on these results, it can be concluded that *N*-alkyl aziridine **2** performs equal to *N*-acyl aziridine **1** in labeling GBA and is, based on the fact that it is easier to prepare and handle, perhaps the retaining β-glucosidase ABP of choice.

In a next set of experiments, recombinant human α-fucosidase (FUCA) was subjected to a similar analysis, now using *N*-acyl aziridine **3** and *N*-alkyl aziridine **4**. As can be seen in Figure 4A, *N*-alkyl aziridine **4** inhibits FUCA about 500-fold less potently for than *N*-acyl aziridine **3**. This result is consistent with the detection limit in FUCA labeling with these probes (Figure 4E). FUCA hydrolyses the fluorogenic substrate, 4-methylumbelliferyl-α-L-fucopyranoside (4-MUαFuc, Figure 4C) optimally at pH 5.0, at which pH also optimal labeling with **3** and (though less effective) **4** is observed (Figure 4B). Time-dependent FUCA labeling with ABP **3** and **4** were also difficult to analyze similar as GBA labeling due to the very fast inhibition rates, this result is backed up by the differences in labeling seen in Figure 4D. The *N*-alkyl aziridine inhibitor **8** (JJB349) blocked FUCA labeling with either **5** or **6** equally effective as non-tagged *N*-acyl aziridines **14** and **15**. As before, in-gel labeling intensity appeared unchanged between 0 and 8 hours of chase (Figure 4G). These results indicate that *N*-acyl aziridine ABP **5** is by far the more effective FUCA activity-based probe and the reagent of choice, even though *N*-alkyl aziridine **6** is the most user-friendly in terms of stability and handling.

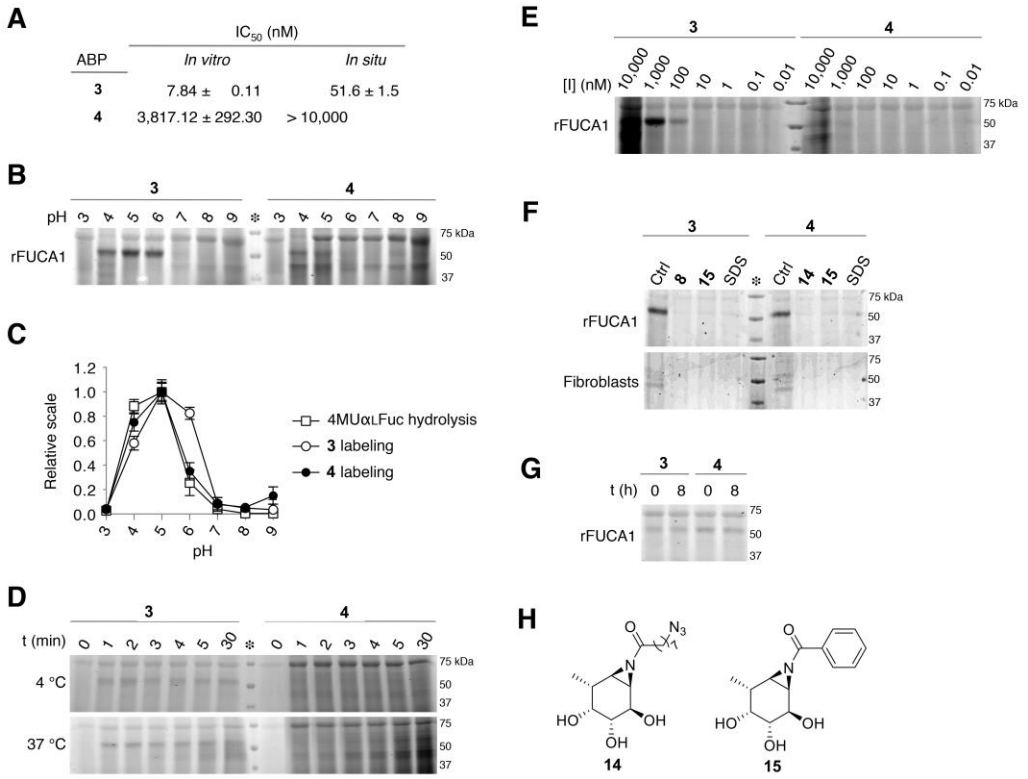


Figure 4. A) Apparent IC_{50} values of α-L-fucopyranose ABPs **3** and **4**, determined *in vitro* against rFUCA and *in situ* in wild-type FUCA fibroblasts. Data are averages of two separate experiments in duplo, and error bars depict standard deviation. B) ABP-labeling of rFUCA at varying pH. C) Quantification of ABP-labeling in (B) compared to hydrolysis of 4-methylumbelliferyl-α-L-fucopyranoside at varying pH. Data are averages of two separate experiments in duplo, and error bars depict standard deviation. D) Labeling of 100 fmol rFUCA with 100 fmol ABP at 4 °C (top) and 37 °C (bottom). E) Detection limit of ABP-emitted fluorescence. rFUCA (10 pmol) was labeled with varying concentrations of ABP. F) ABP labeling of rFUCA (top) or wild-type FUCA fibroblast lysate (bottom) competed with the irreversible inhibitors **8** (Scheme 1), **14** or **15** or SDS. G) Stability of ABP-rFUCA nucleophile adducts after 0 and 8 hour chase in the dark, with hourly extensive washing. (H) Structure of compounds **14** and **15**.

4.3 Conclusions

In conclusion, an in-depth study has been conducted on the efficacy of a new series of *N*-alkyl aziridine-based probes **2** and **4** as activity-based probes for retaining glycosidases. The efficacy of these probes were compared – relatively easy to synthesize and handle – with the previously reported set of *N*-acyl aziridine probes **1** and **3**, which were developed for the profiling of GBA and FUCA, respectively. Preparation of alkyl aziridine compounds is easier because of the intrinsically more stable (compared to *N*-acyl aziridines) *N*-alkyl aziridine electrophilic trap. Activity-based labeling of GBA proved equally effective with *N*-alkyl aziridine **2** as with *N*-acyl aziridine acyl **1**, but the corresponding *N*-alkyl analogues for FUCA proved to

perform less well. These intriguing results warrant future investigations using crystals of the various retaining glycosidases soaked with corresponding ABPs to render an explanation. It can be concluded that *N*-alkyl aziridines can be considered as scaffold to design activity-based probes directed at retaining glycosidases other than targeted to date, but when not active the corresponding *N*-acyl derivatives need be assessed as well before disregarding the cyclophellitol scaffold for ABPP targeting of the glycosidase at hand.

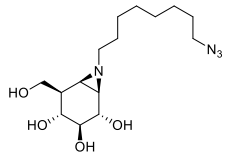
4.4 Experimental section

General synthesis

All reagents were of a commercial grade and were used as received unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and *N, N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an argon atmosphere unless stated otherwise. Solvents used for flash column chromatography were of pro analysis quality. Reactions were monitored by TLC analysis using Merck aluminium sheets pre-coated with silica gel 60 with detection by UV absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}.H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4.H_2O$ (10 g/L) in 10% sulfuric acid followed by charring at ~150 ˚C or by spraying with an aqueous solution of $KMnO_4$ (7%) and K_2CO_3 (2%) followed by charring at ~150 °C. Column chromatography was performed using either Baker or Screening Device silica gel 60 (0.04 - 0.063 mm) in the indicated solvents. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-850 (850/214 MHz), Bruker DMX-600 (600/150 MHz) and Bruker AV-400 (400/100 MHz) spectrometer in the given solvent. Chemical shifts are given in ppm relative to the deuterated chloroform or methanol residual solvent peak or tetramethylsilane (TMS) as internal standard. Coupling constants are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. High-resolution mass spectra were recorded with a LTQ Orbitrap (Thermo Finnigan). Optical rotations were measured on Propol automatic polarimeter (Sodium D-line, λ = 589 nm). IR spectra were recorded on a Shimadzu FT-IR 83000 spectrometer. LC-MS analysis was performed on an LCQ Adventage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C_{18} column (Gemini, 4.6 mm x 50 mm, 3.0 µm particle size, Phenomenex) equipped with buffers A: H_2O, B: acetonitrile (MeCN) and C: 1% aqueous TFA. For reversed-phase HPLC purifications an Agilent Technologies 1200 series instrument equipped with a semi preparative Gemini C_{18} column (10 x 250 mm) was used. The applied buffers were A: 50 mM NH_4HCO_3 in H_2O, B: MeCN.

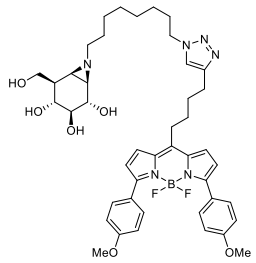
Synthesis of compounds 1 - 15

The β-D-glucoside acyl β-aziridine ABPs JJB75 **1** and JJB103 **13** were synthesized as described earlier.[9] CBE **11** was bought from Sigma and iminosugar **12** was synthesized as reported before.[15] The α-L-fucoside acyl α-aziridines JJB244 **3**, JJB237 **14** and JJB261 **15** were synthesized as described earlier.[11]

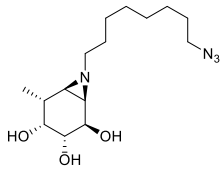


(1R,2S,3S,4R,5R,6R)-7-(8-azidooctyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (6): Unprotected cyclophellitol aziridine **7**[14a] (163 mg, 0.93 mmol, 1.0 eq.) was dissolved in DMF (4.0 mL). Compound 1-azido-8-iodooctane **9** (360 mg, 1.3 mmol, 1.4 eq.) and K_2CO_3 (448 mg, 4.0 mmol, 4.3 eq.) were added to the solution and the mixture

was stirred at 80 °C for 24 h. After the full conversion of the starting material, the resulting solution was filtered over a pad of celite in syringe and the volatiles were concentrated under reduced pressure. Then the crude product was purified by silica gel column chromatograph (linear gradient: 2%→20%, MeOH in DCM) resulting as a colorless oil **6** (150 mg, 0.46 mmol, 49%). TLC: R_f 0.41 (15%, v/v, MeOH in DCM); $[\alpha]_D^{20}$ +49.6 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 3.99 (dd, *J* = 11.3, 4.5 Hz, 1H), 3.66 – 3.59 (m, 2H), 3.28 (t, *J* = 8.0 Hz, 2H), 3.15 – 3.02 (m, 2H), 2.40 – 2.33 (m, 1H), 2.18 – 2.11 (m, 1H), 2.01 – 1.98 (m, 1H), 1.95 – 1.85 (m, 1H), 1.66 (d, *J* = 6.2 Hz, 1H), 1.62 – 1.52 (m, 4H), 1.44 – 1.27 (d, *J* = 13.6 Hz, 8H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 78.92, 73.79, 70.15, 63.76, 62.05, 52.40, 45.36, 45.34, 42.98, 30.45, 30.24, 30.14, 29.85, 28.25, 27.72; IR (neat, cm^{-1}): 3316, 2926, 2855, 2095, 1454, 1348, 1256, 1096, 1020, 818; LC-MS: R_t 4.41 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 329.20 $(M+H)^+$; HRMS: calculated for $C_{15}H_{28}N_4O_4$ $[M+H^+]$ 329.21833, found: 329.21809.

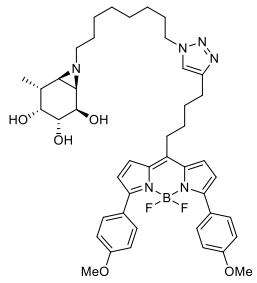


(1R,2S,3S,4R,5R,6R)-7-(8-(4-(4-(5,5-difluoro-3,7-bis(4-methoxyphenyl)-5H-4λ^4,5 λ^4-dipyrrolo[1,2-c:2',1'-*f*][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (2): Azide compound **6** (36 mg, 0.11 mmol, 1 eq.) was dissolved in DMF (3.0 mL), red BODIPY compound **10** (53 mg, 0.11 mmol, 1.1 eq), $CuSO_4$ (1.0 M in H_2O, 60 µL, 0.060 mmol, 0.55 eq.) and sodium ascorbate (1.0 M in H_2O, 70 µL, 0.070 mmol, 0.64 eq.) was added to the solution under argon atmosphere. After stirring at room temperature overnight, the resulting mixture was concentrated under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 50%→56% B in A, 12 min, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting in product **2** as a purple powder (26.5 mg, 0.032 mmol, 30%). ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 7.86 – 7.81 (m, 4H), 7.63 (s, 1H), 7.41 (d, *J* = 4.4 Hz, 2H), 7.00 – 6.91 (m, 4H), 6.68 (d, *J* = 4.4 Hz, 2H), 4.29 (t, *J* = 7.0 Hz, 2H), 3.99 (dd, *J* = 10.0, 4.4 Hz, 1H), 3.83 (s, 6H), 3.62 – 3.57 (m, 2H), 3.13 – 3.08 (m, 1H), 3.03 – 2.98 (m, 3H), 2.75 (t, *J* = 6.7 Hz, 2H), 2.32 – 2.25 (m, 1H), 2.10 – 2.01 (m, 1H), 1.97 – 1.94 (m, 1H), 1.90 – 1.78 (m, 7H), 1.62 (d, *J* = 6.3 Hz, 1H), 1.55 – 1.45 (m, 2H), 1.35 – 1.15 (m, 8H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 162.18, 158.75, 148.58, 146.75, 137.48, 132.15, 128.44, 126.49, 123.23, 121.05, 114.64, 79.02, 73.89, 70.12, 63.78, 62.09, 55.83, 51.22, 45.50, 45.42, 43.01, 34.14, 31.22, 30.97, 30.34, 30.23, 29.87, 28.16, 27.34, 25.77; LC-MS: R_t 6.85 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 813.27 $(M+H)^+$; HRMS: calculated for $C_{44}H_{55}BF_2N_6O_6$ $[M+H^+]$ 813.43245, found: 813.43125.



(1R,2R,3R,4R,5R,6R)-7-(8-azidooctyl)-5-methyl-7-azabicyclo[4.1.0]heptane-2,3,4-triol (8): Unprotected aziridine compound **7**[11] (44 mg, 0.28 mmol, 1 eq.) was dissolved in DMF (2 mL). Compound 1 -azido-8-iodooctane **9** (118 mg, 0.42 mmol, 1.5 eq.) and K_2CO_3 (174 mg, 1.3 mmol, 1.4 eq.) were added to the solution and the mixture was stirred at 80 °C for 24 h. After the full conversion of the starting material, the resulting solution was filtered over a pad of celite in syringe and the volatiles were concentrated under reduced pressure. Then the crude product was purified by silica gel column chromatography (linear gradient: 0%→10%, MeOH in DCM) resulting in **8** as colorless oil (31 mg, 0.10 mmol, 36%). TLC: R_f = 0.39 (10%, v/v, MeOH in DCM); $[\alpha]_D^{20}$ - 47.6 (*c* = 0.5, MeOH); ^{1}H-NMR (850 MHz, CD_3OD): *δ* ppm 3.99 (dd, *J* = 8.7, 4.3 Hz, 1H), 3.54 – 3.53 (m, 1H), 3.34 (dd, *J* = 9.4, 7.5 Hz,

1H), 3.28 (t, *J* = 6.9 Hz, 2H), 2.37 – 2.34 (m, 1H), 2.15 – 2.12 (m, 1H), 1.91 – 1.89 (m, 1H), 1.87 – 1.86 (m, 1H), 1.61 – 1.57 (m, 4H), 1.40 – 1.33 (m, 9H), 1.16 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 76.05, 75.13, 70.18, 62.33, 52.45, 46.45, 45.76, 36.95, 30.60, 30.57, 30.18, 29.89, 28.32, 27.75, 16.75; IR (cm^{-1}): 3370, 2927, 2854, 2092, 1456, 1348, 1253, 1105,1057, 814, 748; LC-MS: R_t 5.06 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 313.20 $(M+H)^+$; HRMS: calculated for $C_{15}H_{28}N_4O_3$ [$M+H^+$] 313.22342, found: 313.22387.



(1R,2R,3R,4R,5R,6R)-7-(8-(4-(4-(5,5-difluoro-3,7-bis(4-methoxyphenyl)-5H-4λ^4,5λ^4-dipyrrolo[1,2-c:2',1'-*f*][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-methyl-7-azabicyclo[4.1.0]heptane-2,3,4-triol (4): Azide compound **8** (8.3 mg, 0.0266 mmol, 1 eq.) was dissolved in DMF (0.8 mL). Red BODIPY **10** (13 mg, 0.033 mmol, 1.24 eq.), $CuSO_4$ (1.0 M in H_2O, 12 µL, 0.012 mmol, 0.45 eq.) and sodium ascorbate (1.0 M in H_2O, 13 µL, 0.013 mmol, 0.48 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 58%→68% B in A, 12 min, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting in product **4** as a purple powder (1.84 mg, 2.31 µmol, 9%). ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 7.86 – 7.83 (m, 4H), 7.69 (s, 1H), 7.43 (d, *J* = 4.4 Hz, 2H), 6.98 – 6.96 (m, 4H), 6.69 (d, *J* = 4.3 Hz, 2H), 4.33 (t, *J* = 7.0 Hz, 2H), 3.99 (dd, *J* = 8.7, 4.3 Hz, 1H), 3.85 (s, 6H), 3.52 – 3.51 (m, 1H), 3.34 – 3.32 (m, 1H) 3.06 (t, *J* = 7.3 Hz, 2H), 2.78 (t, *J* = 6.8 Hz, 2H), 2.28 – 2.24 (m, 1H), 2.10 – 2.05 (m, 1H), 1.89 – 1.82 (m, 7H), 1.52 – 1.48 (m, 2H), 1.31 – 1.21 (m, 9H), 1.13 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 162.18, 158.76, 146.75, 137.48, 132.15, 128.42, 126.49, 123.25, 121.00, 114.61, 76.03, 75.08, 70.11, 62.26, 55.81, 51.24, 46.44, 45.69, 36.91, 34.13, 31.24, 30.96, 30.53, 30.45, 30.32, 29.91, 28.20, 27.34, 25.75, 16.79; LC-MS: R_t 6.96 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 797.07 $(M+H)^+$; HRMS: calculated for $C_{44}H_{55}BF_2N_6O_5$ [$M+H^+$] 796.44041, found: 796.44218.

Materials of biological assays.

Recombinant GBA was obtained from Genzyme (Cambridge, MA, USA). Fibroblast cell lines containing wild-type GBA and FUCA1 were cultured in HAMF12-DMEM medium (Invitrogen) supplied with 10% (v/v).

Molecular cloning and recombinant expression.

Confluent COS-7 cells were transfected with pcDNA3.1 empty vector (mock) or vector containing the coding sequence of *H. sapiens* FUCA1 (NCBI reference sequence XM_005245821.1; cloning described in conjunction with FuGENE (Roche).[11] After 72 hours, cells were harvested by scraping in potassium phosphate buffer (25 mM K_2HPO_4-KH_2PO_4, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)). After determination of the protein concentration (BCA kit, Pierce), lysates were aliquoted and frozen at -80 °C.

Enzyme activity assays and IC_{50} measurements.

The β-D-glucosidase activity of rGBA was assayed at 37 °C by incubating with 3.8 mM 4-methylumbelliferyl-β-D-glucopyranoside as substrate in 150 mM McIlvaine buffer, pH 5.2, supplemented with 0.2% (w/v) sodium taurocholate, 0.1% (v/v) Triton X-100 and 0.1% (w/v) BSA. The α-L-fucosidase activity of rFUCA1 was

determined by incubating with 1.5 mM 4-methylumbelliferyl-α-L-fucopyranoside in 150 mM McIlvaine buffer, pH 5.0, supplemented with 0.1% (w/v) BSA. The values obtained correspond to net α-L-fucosidase activity left after subtracting endogenous α-L-fucosidase activity present in lysate of mock-transfected COS-7 cells. To determine the apparent *in vitro* IC_{50} value, recombinant GBA or lysate of COS-7 cells, either mock or over-expressing FUCA1, was firstly pre-incubated with a range of inhibitor dilutions for 30 min at 37 °C, prior to addition of the substrate. To determine the influence of pH on the enzymatic activity, enzyme mixtures were firstly pre-incubated for 30 min on ice with McIlvaine buffers of pH 3–9 whereafter substrate was added, dissolved in Nanopure H_2O. The enzymatic reaction was quenched by adding excess NaOH-glycine (pH 10.6), after which fluorescence of liberated 4-methylumbelliferyl was measured with a fluorimeter LS55 (Perkin Elmer) using λ_{EX} 366 nm and λ_{EM} 445 nm. The *in situ* IC_{50} value was determined by incubating fibroblast cell lines, grown to confluency, with a range of inhibitor dilutions for 2 h. Hereafter, cells were washed three times with PBS and subsequently harvested by scraping in potassium phosphate buffer (25 mM K_2HPO_4-KH_2PO_4, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)). After determination of the protein concentration (BCA kit, Pierce), lysates were aliquoted and frozen at -80 °C. All IC_{50} values were determined by replicating each assay twice in duplo in two separate cell lines. Data was corrected for background fluorescence, then normalized to the untreated control condition and finally curve-fitted via one phase exponential decay function (GraphPad Prism 5.0). [8a,9,10,11]

***In vitro* labeling and SDS-PAGE analysis.**

All pre-incubations and ABP labeling-reactions occurred for 30 min at 37 °C, unless stated otherwise. The detection limit of each ABP was analyzed by labeling rGBA (10 pmol) or rFUCA (100 µg total protein in lysate of COS-7 cells over-expressing rFUCA1) with 10,000−0.01 nM ABP (**1**/**2**, **3**/**4**, respectively) for 30 min at 37 °C. Influence of pH on ABP labeling involved pre-incubation of the aforementioned enzyme/lysate at pH 3−10 for 30 min on ice, prior to addition of 100 nM ABP **1**/**2**, 1.0 µM ABP **5** or 10 µM ABP **6**, dissolved in Nanopure H_2O and incubating for 30 min at 37 °C. For ABPP, rGBA (10 pmol) or rFUCA (100 µg total protein in lysate of COS-7 cells over-expressing rFUCA1), or 100 µg total protein in lysate of human, wild-type GBA/ FUCA1 fibroblasts, was pre-incubated with compounds 10 mM AMP-DNM, 1.0 mM CBE, 100 µM JJB339, JJB103, JJB349, JJB261 or boiled for 4 min in 2% (w/v) SDS, prior to labeling with 100 nM ABP **1**/**2**, 1.0 µM ABP **5** or 10 µM ABP **6** (all dissolved in Nanopure H_2O) for 30 min at 37 °C. Stability of the ABP-nucleophile adduct was analyzed by firstly labeling the various enzyme mixtures with the corresponding ABPs whilst at appropriate McIlvaine conditions, where-after the mixture was washed over Zeba Spin Desalting Columns with 40K MWCO resin, according to the manufacturer's instructions (Thermo Scientific). The eluted sample was separated, with 50% being snap-frozen in liquid nitrogen and stored at -20 °C, whereas the remaining 50% was chased for 8 h, including hourly washing with the appropriate McIlvaine buffer, over a new Zeba column. Samples were denatured with 5 × 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), boiled for 4 min at 100 °C, and separated by gel electrophoresis on 10% (w/v) SDS-PAGE gels running continuously at 90 V[8a, 9, 10, 11]. Wet slab-gels were then scanned for ABP-emitted fluorescence using a Typhoon TRIO Variable Mode Imager (Amersham Biosciences) using λ_{EX} 532 nm and λ_{EM} 610 nM (band pass filter 30 nm) for red fluorescent ABPs **1**−**6**.

4.5 References

[1] a) D. J. Vocadlo, G. J. Davies, R. Laine and S. G. Withers, *Nature* **2001**, 412, 835-838; b) D. L. Zechel and S. G. Withers, *Acc. Chem. Res.* **2000**, *33*, 11-18; c) G. Davies and B. Henrissat, *Structure* **1995**, 3, 853-859.

[2] D. E. Koshland, *Biol. Rev.* **1953**, 28, 416-436.

[3] a) K. A. Stubbs, *Carbohydr. Res.* **2014**, 390, 9-19; b) L. I. Willems, J. Jiang, K. Y. Li, M. D. Witte, W. W. Kallemeijn, T. J. Beenakker, S. P. Schroder, J. M. Aerts, G. A. van der Marel, J. D. C. Codée and H. S. Overkleeft, *Chem. Eur. J.* **2014**, 20, 10864-10872; c) B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, 77, 383-414.

[4] a) D. J. Vocadlo and C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2004**, *43*, 5452-5456; b) K. A. Stubbs, A. Scaffidi, A. W. Debowski, B. L. Mark, R. V. Stick and D. J. Vocadlo, *J. Am. Chem. Soc.* **2008**, 130, 327-335; c) O. Hekmat, Y.-W. Kim, S. J. Williams, S. He and S. G. Withers, *J. Biol. Chem.* **2005**, 280, 35126-35135; d) J. D. McCarter and S. G. Withers, *J. Am. Chem. Soc.* **1996**, 118, 241-242; e) B. P. Rempel and S. G. Withers, *Glycobiology* **2008**, 18, 570-586.

[5] S. Atsumi, K. Umezawa, H. Iinuma, H. Naganawa, H. Nakamura, Y. Iitaka and T. Takeuchi, *J. Antibiot.* **1990**, *43*, 49-53; b) S. Atsumi, H. Iinuma, C. Nosaka and K. Umezawa, *J. Antibiot.* **1990**, *43*, 1579-1585.

[6] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft and J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, 6, 907-913.

[7] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.* **2014**, 42, D490-D495.

[8] a) M. D. Witte, M. T. Walvoort, K. Y. Li, W. W. Kallemeijn, W. E. Donker-Koopman, R. G. Boot, J. M. Aerts, J. D. C. Codée, G. A. van der Marel and H. S. Overkleeft, *ChemBioChem* **2011**, *12*, 1263-1269; b) M. T. Walvoort, W. W. Kallemeijn, L. I. Willems, M. D. Witte, J. M. Aerts, G. A. van der Marel, J. D. C. Codée and H. S. Overkleeft, *Chem. Commun.* **2012**, 48, 10386-10388.

[9] a) W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529-12533; b) B. Chandrasekar, T. Colby, A. Emran Khan Emon, J. Jiang, T. N. Hong, J. G. Villamor, A. Harzen, H. S. Overkleeft and R. A. van der Hoorn, *Mol. Cell. Proteomics* **2014**, *13*, 2787-2800.

[10] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codée, G. A. van der Marel, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, *136*, 11622-11625.

[11] J. Jiang, W. W. Kallemeijn, D. W. Wright, A. M. C. H. van den Nieuwendijk, V. C. Rohde, E. C. Folch, H. van den Elst, B. I. Florea, S. Scheij, W. E. Donker-Koopman, M. Verhoek, N. Li, M. Schurmann, D. Mink, R. G. Boot, J. D. C. Codée, G. A. van der Marel, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. Sci.* **2015**, *6*, 2782-2789.

[12] a) K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima and M. Nakata, *J. Antibiot.* **1991**, *44*, 912-914; b) M. Nakata, C. Chong, Y. Niwata, K. Toshima and K. Tatsuta, *J. Antibiot.* **1993**, *46*, 1919-1922; c) K. Tatsuta, *Pure Appl. Chem.* **1996**, *68*, 1341-1346.

[13] K. Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, W. E. Donker-Koopman, R. G. Boot, J. M. F. G. Aerts, J. D. C. Codée, G. A. van der Marel and H. S. Overkleeft, *Org. Biomol. Chem.* **2014**, 12, 7786-7791.

[14] a) K. Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C. S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, 6030-6043; b) L. I. Willems, T. J. M. Beenakker, B. Murray, B. Gagestein, H. van den Elst, E. R. van Rijssel, J. D. C. Codée, W. W. Kallemeijn, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, 6044-6056.

[15] T. Wennekes, R. J. van den Berg, W. Donker, G. A. van der Marel, A. Strijland, J. M. Aerts and H. S. Overkleeft, *J. Org. Chem*. **2007**, 72, 1088-1097.

5

Detection of active mammalian GH31 α-glucosidases in health and disease using in-class, broad-spectrum activity-based probes

Jianbing Jiang, Chi-Lin Kuo, Liang Wu, Christian Franke, Wouter W. Kallemeijn, Bogdan I. Florea, Eline van Meel, Gijsbert A. van der Marel, Jeroen D. C. Codée, Rolf G. Boot, Gideon J. Davies, Herman S. Overkleeft and Johannes M. F. G. Aerts, *ACS Central Science*, **2016**, in press, DOI: org/10.1021/acscentsci.6b00057.

5.1 Introduction

Lysosomal α-glucosidase (GAA, α-1,4-glucosidase, acid maltase) (E.C.3.2.1.20) is a retaining α-glucosidase, which has been classified into CAZy family GH31.[1-4] Following processing within endoplasmic reticulum the 110 kDa (952 AA) precursor is transported to the lysosomes where it is modified to active 76 and 70 kDa isoforms.[5,6] Within the lysosomes, GAA catalyzes the degradation of glycogen via a general acid/base catalyzed double displacement mechanism,[7,8] releasing a molecule of α-glucose with net retention of stereochemistry at its anomeric center (Figure 1A). Deficiency in GAA leads to the glycogen storage disease type II, known as Pompe disease.[9] In Pompe patients, intra-lysosomal glycogen accumulation causes progressive muscle weakness in heart and skeletal muscles and also affects the liver and nervous system.[10] Different clinical forms of Pompe disease are usually discerned based on age of onset. The infantile-onset form manifests at 4 to 8 months and, when untreated, results in death in the first years of life.[11] Later onset forms generally progress more slowly and are characterized by progressive decrease in muscle strength in the legs followed by smaller muscles in the trunk

and arms and ultimately to fatality through respiratory failure.[12] The severity of the Pompe disease and its progress correlates with the extent of enzyme loss. Pompe disease is currently treated by chronic intravenous administration of recombinant human GAA (rGAA, alglucosidase alfa; Myozyme). Impressive delay of fatal symptoms by enzyme replacement therapy (ERT) is observed in infantile Pompe disease patients,[13] but treatment of late onset disease requires very large amounts of therapeutic enzyme due to poor correction of muscle cell pathology.[14]

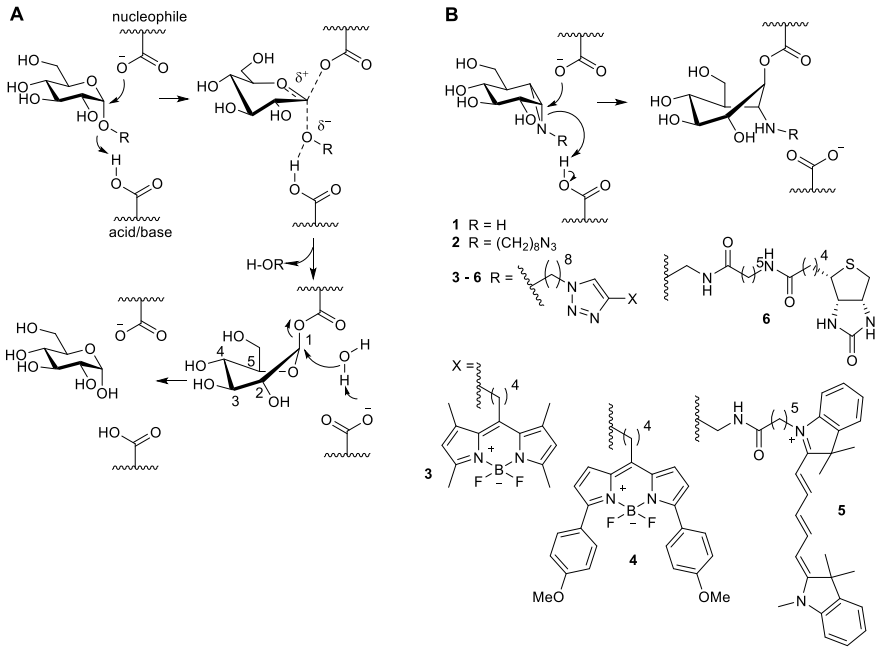


Figure 1. The mechanism of action of retaining α-glucosidases allows the development of activity-based probes. (A) Koshland double-displacement mechanism employed by retaining α-glucosidases. (B) α-Glucose-configured *N*-alkyl cyclophellitol aziridines as mechanism-based irreversible retaining α-glucosidase inhibitors (**1**, **2**) and probes (**3**-**6**).

The identification of GAA activity in living cells or tissues is challenging. It has been previously shown that activity-based protein profiling (ABPP) is an effective methodology to quantify β-glucosidases responsible for the lysosomal storage disorder, Gaucher disease.[15,16] Linking "reporter moieties" to the aziridine nitrogen in the retaining β-glucosidase inhibitor, cyclophellitol aziridine, yielded in-class broad-spectrum retaining β-glucosidase probes. These probes label, besides lysosomal glucosylceramidase (GBA-the enzyme deficient in Gaucher disease) also neutral glucosylceramidase (GBA2), broad-specificity β-glucosidase (GBA3) and lactase phloridzin hydrolase (LPH).[16-21] Based on these findings, it was hypothesized that the corresponding α-glucopyranose configured, nitrogen-substituted cyclophellitol aziridines (Figure 1B), would be effective ABPs for specific labeling of GH31 retaining α-glucosidases,

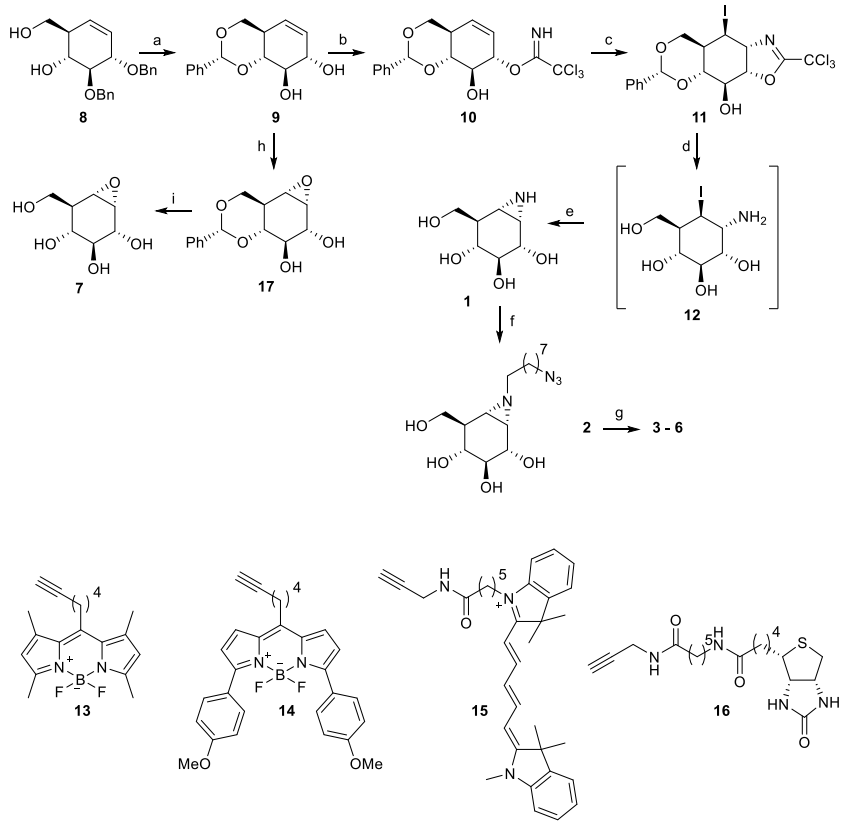
besides GAA also ER α-glucosidase II (GANAB, a crucial enzyme in the quality control of newly formed glycoproteins in the ER) and intestinal α-glucosidases involved in food processing. Here, the successful development of such in-class GH31 α-glucosidase ABPs is reported. Both mass spectrometry and X-ray crystallography demonstrate the covalent labeling of the probe on a representative bacterial GH31 enzyme, whch is homologous to GAA. GH31 α-glucosidases are labeled *in situ* by designed ABPs and that the absence of GAA in Pompe disease cells is readily demonstrated in a diagnostic manner.

5.2 Results

Synthesis of mechanism-based inhibitors and activity-based probes

As the first research objective, to develop an efficient route of synthesis for the preparation of α-configured cyclophellitol aziridine **1** was set out (Scheme 1). The synthesis route for cyclohexene **8**, the precursor in the total synthesis of cyclophellitol (a natural product and a retaining β-glucosidase inhibitor) reported by Madsen and co-workers,[22] has been developed in previous works on retaining β-glucosidase probes development.[15,16] It was reasoned that an iodocyclization scheme, with an appropriate nitrogen nucleophile delivered to the alkene-derived iodonium ion from the allylic alcohol position, would yield an appropriately configured 2-amino-1-iodo species for subsequent intramolecular cyclisation in a stereospecific fashion. For this purpose, the benzyl protective groups in **8** were removed by Birch reduction, after which the 4,6-benzylidene (glucopyranose numbering) was installed to give **9**. The allylic alcohol in **9** proved the most reactive of the two secondary alcohols towards trichloroacetonitrile in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) as the catalytic base, providing imidate **10** as the major product in the presence of small amounts of the bis-imidate. Subsequent key iodocyclization afforded in a stereospecific fashion cyclic imidate **11**. Both cyclic imidate and benzylidene acetal were hydrolysed under acidic conditions and the resulting crude primary amine **12** was treated with sodium bicarbonate giving aziridine **1**. The aziridine nitrogen in **1** was alkylated with 1-iodo-7-azidoheptane to yield **2**, onto which and by means of copper (I)-catalyzed azide-alkyne [2+3] cycloaddition either a BODIPY-green dye, a BODIPY-red dye, a Cy5 dye or a biotin was conjugated, to give ABPs **3**, **4**, **5** and **6**, respectively (Scheme 1). 1,6-*epi*-cyclophellitol **7** was prepared by stereoselective epoxidation of **9** with 3-chloroperbenzoic acid (*m*CPBA) and deprotection of benzylidene acetal group of the resulting **17** by Pearlman's catalyst hydrogenation. As reference competitive inhibitors, deoxynojirimycin derivatives **18**-**22** and maltose **23** were used for screening (SI, Figure S1).[23]

Scheme 1. Synthesis of the cyclophellitol aziridine inhibitors **1**, **2**, probes **3**-**6** and 1,6-*ep*i-cyclophellitol **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 steps yield: 41%. (d) 37% HCl aq., dioxane. (e) $NaHCO_3$, MeOH, two steps yield: 63%. (f) 1-azido-8-iodooctane, K_2CO_3, DMF, 80 °C, 39%. (g) **13**, **14**, **15** or **16**, $CuSO_4$ (1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), 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%;

In vitro inhibition and labeling of recombinant human GAA

The inhibitory properties of compounds **1**-**7** (Figure 2) were firstly tested toward rGAA (Myozyme, Genzyme) at pH 4.0 using 4-methylumbelliferyl-α-glucose as substrate (Figure 2A). All compounds proved to be potent rGAA inhibitors with apparent IC_{50} values in the nanomolar range. The most optimized potent inhibitor of the series is cyclophellitol aziridine **1**, with epoxide **7** and *N*-alkyl aziridine **2** within the same range, albeit slightly weaker (Figure 2A). Fluorescent ABPs **3**-**5** all showed about 10-fold increased apparent IC_{50} values compared to **1**, and biotin-conjugated ABP **6** a further 2-fold increase. *In situ* enzyme inhibition within human fibroblasts exhibited similar potency, except for biotin compound **6** (IC_{50} > 10 µM), which apparently is the least cell-permeable of the series. The pH optimum of 5.0 found for labeling of 110 kDa rGAA differed slightly from that of enzymatic activity towards 4MU-α-glucose pH

4.0 (Figure 2B). When rGAA was pre-incubated for 30 min with 10 µM of compounds **1**-**2** and **4**-**7**, followed by labeling with compound **3**, green fluorescent labeling of rGAA by compound **3** was abrogated (Figure 2C). Likewise, the presence of high concentrations of the substrates 4MU-α-glucose and maltose reduced labeling by compound **3** as does prior denaturation of rGAA by incubation with 2% SDS. Inhibitors **1**-**7** reacted too fast for kinetics analysis: for instance, time dependent labeling of rGAA by 100 nM ABP **3** at 4 °C and 37 °C demonstrated most of the enzyme could be labeled within 2 min (SI, Figure S2A).

A

Compounds	IC_{50} (nM)	
	In vitro	*In situ*
1, CF021	37.8 ± 3.4	71.2 ± 9.0
2, CF022	51.7 ± 2.3	75.8 ± 10.9
3, JJB382	458.9 ± 38.4	502.2 ± 4.1
4, JJB347	383.1 ± 22.1	561.6 ± 352.2
5, JJB383	268.7 ± 7.1	42.2 ± 4.9
6, JJB384	767.1 ± 3.5	> 10,000
7, JJB307	54.1 ± 4.9	97.6 ± 14.5

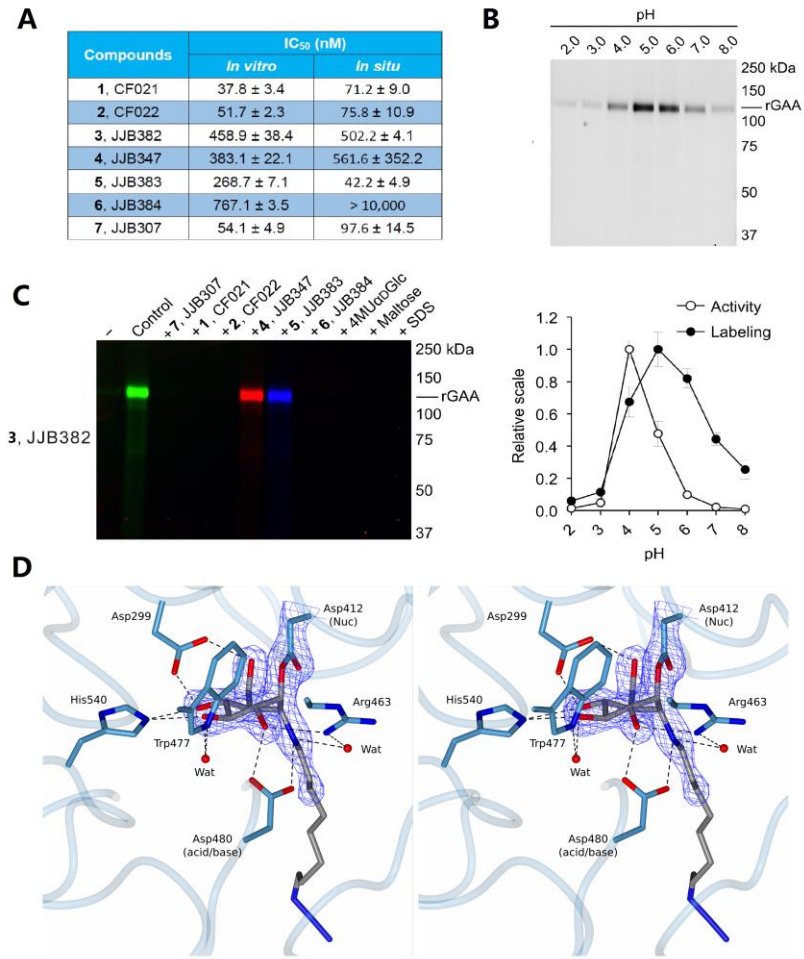


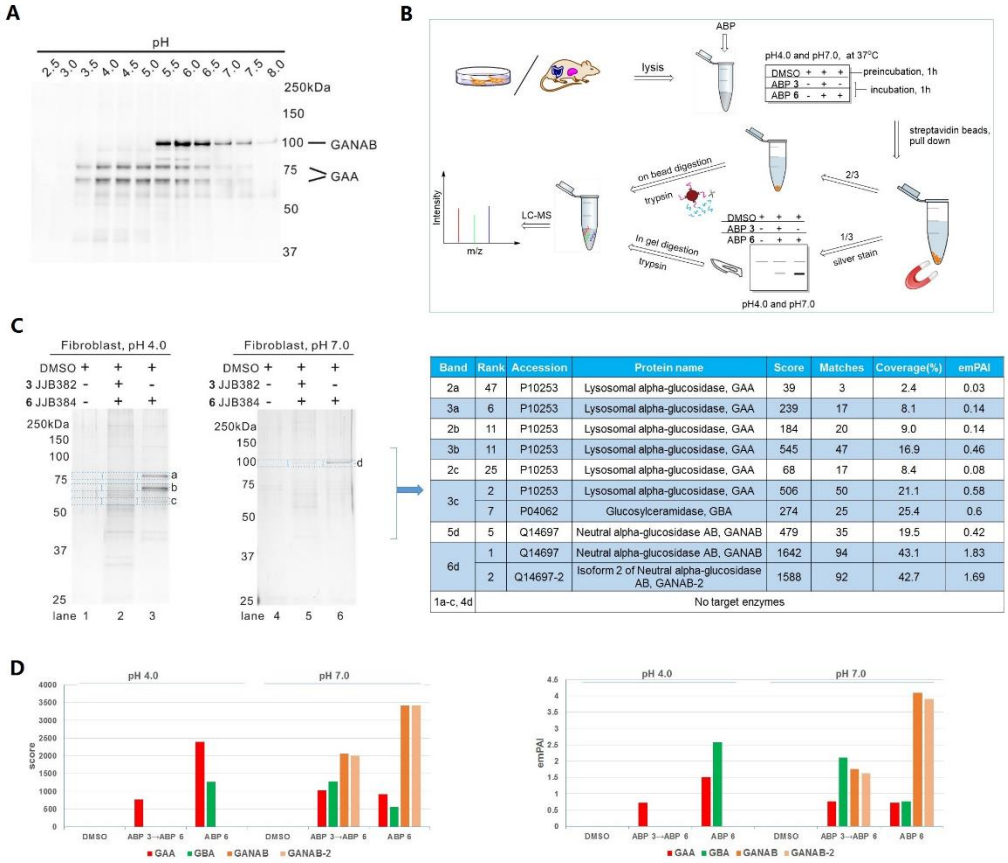
Figure 2. *In vitro* inhibition and labeling of α-glucosidases. A) Inhibition of recombinant GAA. Apparent IC_{50} values (extrapolated with one phase exponential association) are the average values of two separate experiments measured in duplicate, error ranges depict standard deviation B) Labeling of rGAA with compound **3** at various pH as compared to activity towards 4-MU-α-D-glucopyranose. Error bars represent standard deviation. C) ABP labeling of rGAA with **3** competed with compounds **1**-**2** and **4**-**7**. D) Stereoscopic view of the *Cj*Agd31B active site in complex with compound **2**, showing covalent link to C*j*Agd31B enzymatic nucleophile Asp412, and H-bonding interactions to neighboring residues. Electron density is REFMAC maximum-likelihood/σ_A-weighted 2 F_o–F_c synthesis contoured at 0.49 electrons per Å^3.

To establish the mechanism-based mode of action of the probes, a bacterial homologue of GAA from CAZy family GH31, the α-glucosidase from Cellvibrio japonicus, C*j*Agd31B, was used, which is readily amenable to structure studies of GH31 ligand complexes and displays 27% identity to the human GAA enzyme over 615 amino acids including absolute conservation of the active centre "-1" glucosyl site.[24] The 1.95 Å resolution structure of *Cj*Agd31B (Figure 2D) soaked with cyclophellitol aziridine **2** revealed unambiguous electron density showing covalent binding of the ring opened cyclophellitol aziridine β-linked to the active site nucleophile residue (Asp412; equating to Asp518 of GAA and Asp542 of GANAB). Apart from demonstrating irrevocably the mechanism-based mode of action of the probes, the GH31 complex with **2** is consistent with the reaction pathway employed by GH31 in processing α-glucosidic linkages. Substrates are proposed to bind in a $^4C_1$ conformation with catalysis occurring via a $^4H_3$ oxacarbenium ion-like transition state to a covalent adduct in $^1S_3$ skew-boat conformation.[25,26] This $^1S_3$ conformation of the enzyme-inhibitor adduct is well captured in the complex shown in Figure 2D. Similarly, the $^1S_3$ conformation of the enzyme-inhibitor **1** adduct was also well captured in 1.85 Å resolution (SI, Figure S2B).

ABP labeling of multiple α-glucosidases in homogenates of fibroblasts

Having established the expected covalent inactivation, ABP labeling of α-glucosidases was subsequently examined in a more complex and physiologically relevant biological mixture of proteins. For this, homogenates of cultured fibroblasts were prepared and exposed to 1.0 μM ABP **5** at various pH values. Incubation with ABP **5** at pH 4.0 gave clean labeling of what appeared to be the mature lysosomal forms of GAA (two bands at around 70 kDa). At higher pH values an additional protein with a molecular weight of around 100 kDa was observed (Figure 3A). GAA precursor amounts are low in cells and are – as with the mature form – optimally active at low pH. Therefore, it was thought that the protein labeled at 100kDa would unlikely be the GAA pro-form. Rather, it was envisaged that this band would correspond to another α-glucosidase, possibly ER α-glucosidase II. Unambiguously, to establish the nature of the 70 kDa proteins identified at acidic pH as well as the 100 kDa protein seen at neutral pH, affinity purification was performed and chemical proteomics of fibroblast homogenates were treated with biotin-conjugated ABP **6**. A lysate of fibroblasts was incubated at pH 4.0 and pH 7.0 with ABP **6**, in either the presence or absence of pre-treated ABP **3**. Next, biotinylated proteins were enriched by pull down using streptavidin-coated magnetic beads. Loaded streptavidin beads were split for in-gel digestion and on-bead digestion. Captured proteins for subsequent in-gel digestion were released in Laemmli buffer, resolved on SDS-PAGE and visualized by silver staining (Figure 3B). Three distinct bands were obtained that all could be competed for by inclusion of α-cyclophellitol aziridine **3**. Two bands with apparent molecular masses of around 70 kDa are most prominent in the pull-down at pH 4.0 and one band at approximately 100 kDa was the predominant species labeled at pH 7.0. Tryptic digestion was performed on these three protein bands and the resulting tryptic peptides were analyzed by

nanoscale liquid chromatography coupled to tandem mass spectrometry (nano-LCMS/MS). The proteins were identified via matching of the obtained peptide sequences against the Mascot database. In this manner, the ~70 kDa proteins were identified as the two mature forms (70 and 76 kDa) of GAA (Figure 3C). The labeled 100 kDa protein proved to be GANAB and isoform 2 of GANAB; the known retaining ER α-glucosidase II.



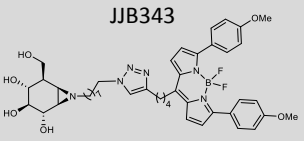
Band	Rank	Accession	Protein name	Score	Matches	Coverage(%)	emPAI
2a	47	P10253	Lysosomal alpha-glucosidase, GAA	39	3	2.4	0.03
3a	6	P10253	Lysosomal alpha-glucosidase, GAA	239	17	8.1	0.14
2b	11	P10253	Lysosomal alpha-glucosidase, GAA	184	20	9.0	0.14
3b	11	P10253	Lysosomal alpha-glucosidase, GAA	545	47	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 3. Labeling of multiple α-glucosidases and their identification by proteomics. A) Labeling of proteins in fibroblast lysate with compound **3**. B) Work flow of pull down proteomics experiments. C) In-gel digestion silver staining and identification of target proteins modified by biotin-ABP **6**. D) Glycosidases identification with Mascot emPAI and score values after on-bead pull-down and processing of fibroblast lysate with biotin-ABP **6**.

Using on-bead trypsin digestion very similar results were obtained. GAA in the pH 4.0 incubation and GANAB in the pH 7.0 incubation topped the exponentially modified Protein Abundance Index (emPAI), a relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result, and protein scores from Mascot search report (Figure 3D and SI, Figure S3). Comparable experiments with mouse liver lysates (SI, Figure S4, S5) resulted in identification of GAA and GANAB as the major targets of biotin-

aziridine ABP **6**, matching the fluorescent bands on SDS-PAGE. Using a final concentration of 10 μM of biotin-aziridine **6** in the affinity purification experiments, enrichment of a protein with apparent molecular mass of 60-65 kDa was also observed (SI, Figure S4). Proteomics analysis identified this protein as GBA, the lysosomal acid β-glucosidase deficient in Gaucher disease. To corroborate this observation, the possible inhibition of enzymatic activity of pure GBA by compounds **1**-**7** was determined. Indeed, GBA is also inhibited by all of the prepared compounds with apparent IC_{50} values in the range of 592.8 nM – 155.3 μM (Table 1). For comparison, the corresponding β-configured cyclophellitol aziridine JJB343[15] was a far more potent inhibitor towards GBA with 500-fold lower apparent IC_{50} values than α-aziridine **4** JJB347 in the same test.

Table 1. Apparent IC_{50} values of α-configured cyclophellitol azridrines **1-6** epoxide **7** JJB307 and on rGBA, determined *in vitro*. JJB343 (cyclophellitol-β-aziridine-BODIPY) was included for comparison. Data were average values of two separate experiments measured in duplicate, error ranges depict standard deviation.

Compounds	*In vitro* IC_{50} (nM)
1 CF021	41,943.5 ± 490.0
2 CF 022	603.2 ± 28.0
3 JJB382	592.8 ± 214.3
4 JJB347	1,076.0 ± 63.6
5 JJB383	815.8 ± 526.4
6 JJB384	2,060.0 ± 219.2
7 JJB307	155,333.0 ± 1,773.4
JJB343 	2.05 ± 0.93

Labeling of α-glucosidases present in murine gastrointestinal tracts

Given the observed high affinity labeling with the ABPs of both GAA and GANAB, labeling of proteins was examined in extracts from intestines, tissue known to contain additional retaining GH31 α-glucosidases. For this purpose, mouse intestine was freshly collected, food remains and bacteria removed by rinsing and the tissue lysed in KPi-Triton buffer using sonication. The lysate was incubated with biotinylated compound **6** and bound proteins were analyzed by gel electrophoresis and chemical proteomics as before (SI, Figure S6, S7). In this way, biotin-ABP **6** could label GAA and GANAB, as well as sucrase-isomaltase (Sis) and maltase-glycoamylase (MGAM), two retaining α-glucosidases expressed specifically in intestinal tissue. In addition, the retaining β-glucosidases, lactase (Lct) and GBA were identified. The findings were recapitulated by labeling lysates of pull-down and supernatant samples with ABPs **3** and

6, protein separation by gel electrophoresis and fluorescence imaging of the wet gel slabs, followed by transferring protein to polyvinylidene fluoride (PVDF) membrane for Western blot analysis with HRP-streptavidin. Fluorescent bands, chemiluminescence bands and silver stain protein bands matched well, and the expected molecular masses for GAA, GANAB, Sis, MGAM, Lct and GBA were observed (SI, Figure S6).

Labeling and concomitant inhibition of α-glucosidases in intact cells.

Prompted by the effective *in vitro* labeling of GH31 α-glucosidases in cell and tissue extracts, the labeling of enzymes in intact cultured fibroblasts was studied with ABPs **3** and **5**. Fibroblasts were exposed to 100 nM ABP **5** for various times (from 10 min to 4 h) by including these in the culture medium. Subsequently, cells were extensively washed and harvested, lysates were prepared at 4 °C and then separated by SDS-PAGE followed by fluorescence scanning of the wet gel slabs (Figure 4A left). Fluorescent labeling of GAA and GANAB was prominent in blue bands. Of note, no GBA was concomitantly labeled *in situ*, illustrating the far lower affinity of the α-glucopyranose ABP to label this β-glucosidase, but GBA was labeled by excess of ABP **3** under *in vitro* conditions (Figure 4A right). During the procedure an excess of ABP **3** was added to exclude artificial labeling of enzymes by non-reacted free ABP **5** during the process of cell lysis. This addition did not change the result, demonstrating that labeling of GAA and GANAB truly occurs *in situ*.

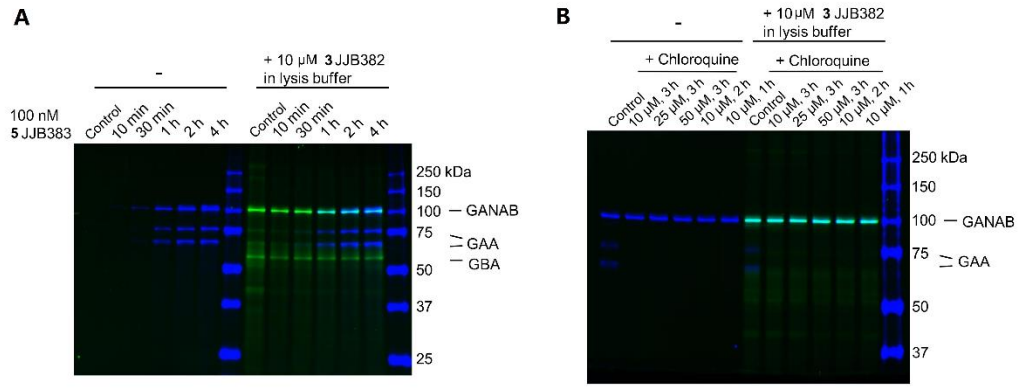


Figure 4. *In situ* labeling of GAA and GANAB in living fibroblasts. A) Time-dependent labeling of GAA and GANAB in fibroblasts by ABP **5** *in situ* and ABP **3** *in vitro*. B) Chloroquine blocks *in situ* GAA labeling.

GAA requires an acid pH for labeling by the α-glucopyranose ABP (Figure 2A). The effect of chloroquine, an agent that raises intralysosomal pH, was therefore investigated.[27] Fibroblasts were incubated with either increasing amount of chloroquine (10, 25 and 50 μM) for three h or increasing treatment time (1, 2 and 3 h) after which ABP **5** was added to the medium. After the time points cells were washed, harvested and analyzed on labeled protein. As shown in Figure 4B, the presence of chloroquine abolished *in situ* labeling of GAA but not GANAB.

Diagnostic application: probing for GAA activity in Pompe disease patient material

Finally, the value of the ABPs for diagnosis of Pompe disease was studied. Fibroblasts of normal individuals and Pompe disease patients suffering from the infantile and adult variants of disease were cultured. Cell lysates were labeled with 0.5 µM of ABP **3** for 30 min at pH 4.0 and pH 7.0. Gel electrophoresis of the denatured protein content and subsequent fluorescence imaging and immunoblots revealed a prominent distinction between material from normal persons and Pompe disease patients (Figure 5). Marked absence of fluorescently labeled mature 70-76 kDa GAA with concomitant normal levels of GANAB was only observed for the patient fibroblasts. This finding demonstrates the potential of α-glucosidase ABPs in laboratory diagnosis of Pompe disease.

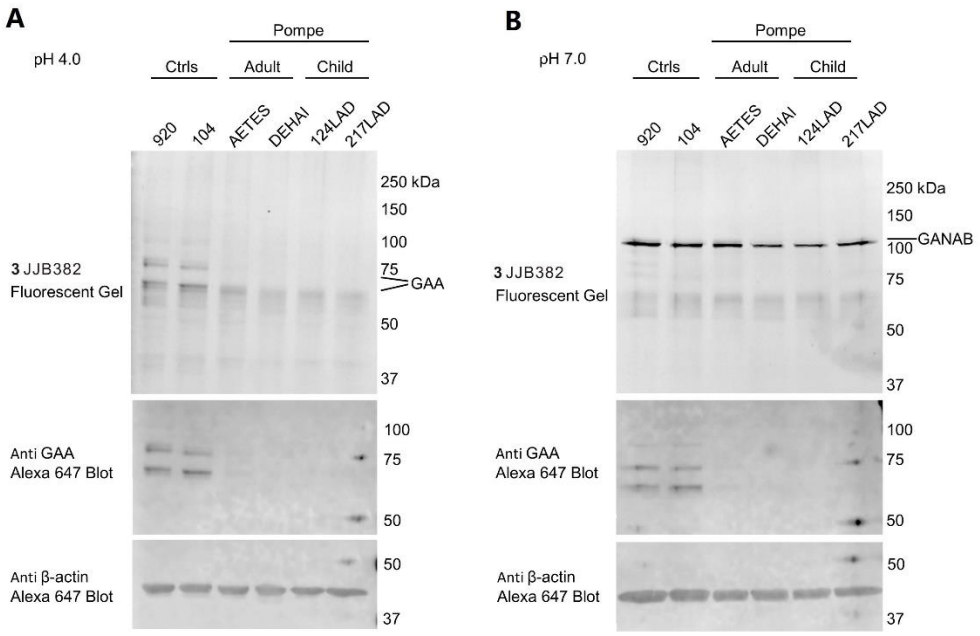


Figure 5. *In vitro* ABP labeling and Western blot detection of α-glucosidases in wild type and Pompe fibroblasts. A) *In vitro* GAA labeling at pH 4.0 with **3**, followed by Western blot detection of GAA in various fibroblast lysates, containing wild type or mutant (Pompe) GAA. B) *In vitro* GANAB labeling at pH 7.0 with **3** followed by Western blot detection of GAA in various fibroblast lysates, containing wild type or mutant (Pompe) GAA**.**

5.3 Discussion

Activity-based protein profiling (ABPP) has emerged in the past fifteen years as a powerful technique to identify enzymes and to study their activity in the context of the physiological processes they partake in, both *in vitro* and *in situ*. In the first instance developed for serine hydrolases and cysteine proteases,[28,29] ABPs can on paper be designed for any enzyme, provided that a covalent enzyme-substrate adduct emerges during enzyme action. Retaining

glycosidases that employ a Koshland two-step double displacement mechanism fulfill this requirement, and it has been reported previously on the versatility of the natural product, cyclophellitol, as a scaffold for activity-based glycosidase probe design. In this work, the generic design principle is adopted by the development of a set of α-configured cyclophellitol aziridines as in-class GH31 retaining α-glucosidase ABPs. The probes label GH31 retaining α-glucosidases in a tissue-dependent fashion and allow detection and identification of these by in-gel fluorescence and by chemical proteomics. They are highly selective towards GH31 retaining α-glucosidases when applied in the appropriate concentration and even at higher concentrations show little to no cross-reactivity, apart from labeling the retaining β-glucosidase, GBA. The ABPs label subcellular α-glucosidases optimally at the pH at which the α-glucosidases display maximal enzymatic activity. Thus, whereas GAA in fibroblasts is maximally labeled at pH 4.0-5.0, GANAB, or ER-α-glucosidase II, is optimally labeled at neutral pH. This finding underscores that the probes report on functional enzymes, rather than protein levels, an observation that is corroborated by the X-ray structure of C*j*Agd31B rGAA, in which ABP precursor **2** has reacted with the active site nucleophile, Asp412.

The tissue-specific and pH-dependent mode of action of the ABPs allow for probing each of the targeted α-glucosidases independently and within their physiological context, as well as in health and disease. The diagnostic value of the probes is demonstrated in the assessment of the lack of GAA activity in infantile and adult Pompe disease tissue (Figure 5). The in-class broad-spectrum GH31 α-glucosidase probes may also find use in the discovery of inhibitors specific for either GAA or GANAB in a competitive ABPP assay. Selective inhibitors for GAA or GANAB would be of interest: GAA inhibitors for pharmacological chaperone discovery in the context of Pompe disease and GANAB inhibitors for antiviral or anticancer drug discovery (GANAB activity being a crucial factor in the quality control of nascent *N*-linked glycoproteins). An initial competitive ABPP was performed in which fibroblasts expressing GAA and GANAB were first treated with varying concentrations of the iminosugars **18**-**22** and maltose **23** followed by incubation at both pH 4.0 and pH 7.0 with ABP **5**. As can be seen in the Figure 6 these compounds efficiently inhibit both enzymes. Indeed no selective inhibitors for either enzyme are known and the ABPs may be of help in identifying such compounds. The ABPs finally also efficiently target and identify intestinal α-glucosidases, which are key players in glucose assimilation and the targets of the anti-diabetic drugs, miglitol and acarbose. The ability to also study these intestinal enzymes in detail in their physiological surroundings holds promise for the future identification of more effective (and selective with respect to GAA/GANAB inhibition) inhibitors, which may come from natural sources and become part of nutraceutical regimes.

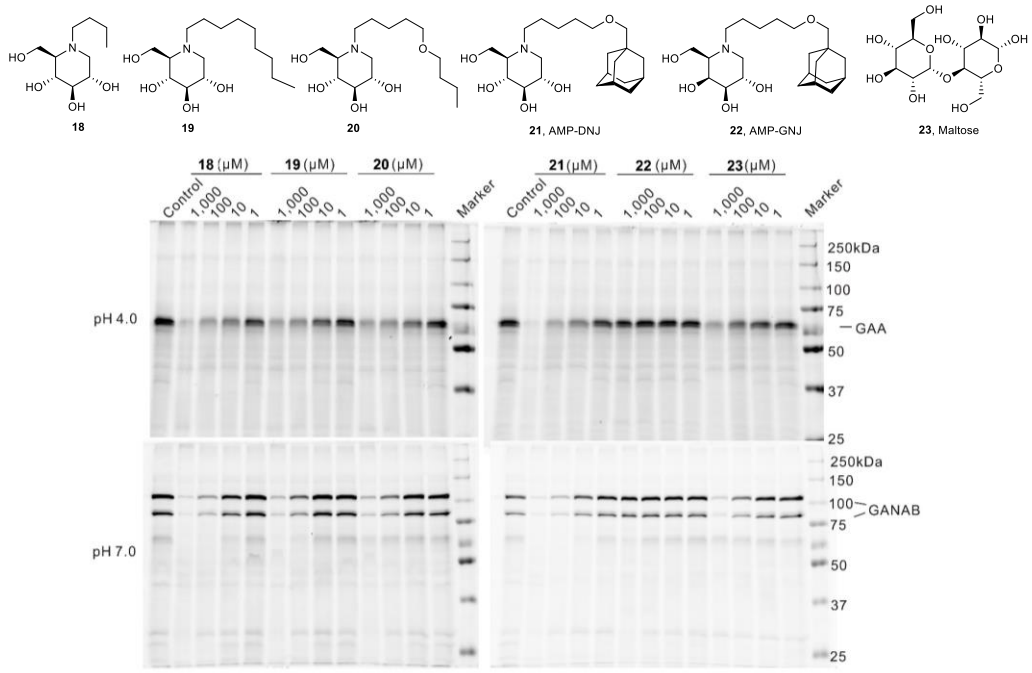


Figure 6. ABP **5** competitive labeling of α-glucosidases in mouse liver extract at pH 4.0 and pH 7.0 with pre-incubation of inhibitors **18-23**.

5.4 Experimental methods

Biological assays:

Materials

Chemicals were obtained from Sigma-Aldrich, if not otherwise indicated. Trypsin and Endoproteinase GluC were commercially available from Promega. Recombinant GAA was obtained from Genzyme (Cambridge, MA, USA). Fibroblasts were obtained with consent from donors. Pompe patients were diagnosed on the basis of reduced GAA activity. Cell lines were cultured in HAMF12-DMEM medium (Invitrogen) 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 lysis buffer(25 mM in pH 6.5, supplemented with protease inhibitor 1x cocktail (Roche)) via homogenization with silent crusher S equipped with Typ 7 F/S head (30 rpm x 1000, 3 × 7 sec) on ice and lysate concentration was determined with BCA Protein Assay Kit (Pierce™). The protein fractions were stored in small aliquots at -80 °C until use.

Enzyme activity assays and IC_{50} measurements.

The α-D-glucosidase activity of lysosomal α-D-glucosidase GAA was assayed at 37 °C by incubating with 3.0 mM 4-methylumbelliferyl-α-D-glucopyranoside as substrate in 150 mM McIlvaine buffer, pH 5.0, supplemented with 0.1% (w/v) BSA. Activity of rGBA was measured using similar conditions but with 3.8 mM 4-methylumbelliferyl-β-D-glucopyranoside as substrate at pH 5.2, supplemented with 0.1% (v/v) Triton X-100 and 0.2% (w/v) sodium taurocholate. To determine the apparent *in vitro* IC_{50} value, recombinant GAA or rGBA was firstly pre-incubated with a

range of inhibitor dilutions for 30 min at 37 °C, prior to addition of the substrate. To determine the influence of pH on the enzymatic activity, enzyme mixtures were firstly pre-incubated for 30 min on ice with McIlvaine buffers of pH 2.0–8.0 whereafter substrate was added, dissolved in nanopure H_2O. The enzymatic reaction was quenched by adding excess NaOH-glycine (pH 10.3), after which fluorescence of liberated 4-methylumbelliferyl was measured with a fluorimeter LS55 (Perkin Elmer) using λ_{EX} 366 nm and λ_{EM} 445 nm. The *in situ* IC_{50} value was determined by incubating fibroblast cell lines expressing wild-type GAA, 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 potassium phosphate buffer (25 mM K_2HPO_4-KH_2PO_4, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)). Residual GAA activity was measured using the aforementioned substrate assay. All *in situ* IC_{50} values were determined by replicating each assay twice in duplo in two separate cell lines. Data was corrected for background fluorescence, then normalized to the untreated control condition and finally curve-fitted via one phase exponential decay function (GraphPad Prism 5.0).[15,16]

ABP 7 pull-down and LC-MS/MS analysis

3mg total protein from human fibroblast lysate or 6 mg total protein from mouse intestines lysate was incubated with either 0.1% (v/v) DMSO, 5.0 µM **7** JJB384, or firstly with 5.0 µM **3** JJB382 followed by 5.0 µM **6** JJB384. For mouse liver lysate, 6.0 mg total protein and 10 µM ABPs **3** and **6** were used. Each step taking 30 min at 37 °C, in a total volume of 0.50 mL McIlvaine buffer of pH 4.0 and pH 7.0, subsequently denatured through the addition of 10% (w/v) SDS 125 µL and boiling for 5 min at 100 °C. From here on, samples were prepared for pull-down with streptavidin beads as published earlier[2]. After pull-down procedure, the samples were divided, 2/3 for on-bead digestion and 1/3 for in-gel digestion. On-bead digestion samples including beads were treated by the trypsin digestion buffer (a mixture containing 100 mM Tris-HCl pH 7.8, 100 mM NaCl, 1.0 mM $CaCl_2$, 2% MeCN and 10 ng/µL trypsin) and the bead suspension was incubated in a shaker at 37 °C overnight. The supernatant containing the trypsin-digested peptides was desalted using stage tips, followed by evaporation of MeCN and dilution in 70 µL sample solution (H_2O/MeCN/TFA, 95/3/0.1, v/v/v) for LC-MS analysis. The beads containing active-site peptides were treated with endoproteinase Glu-C digestion buffer (100ng/µL in PBS solution); incubated in a shaker at 37 °C overnight after which the supernatant was desalted using stage tips and for LC-MS. In-gel digestion samples were eluted by boiling the beads at 100 °C in 30 µL of 1 x Laemmli buffer. The eluted proteins were separated on 10% protein gels at 200 V for 1 h, and the protein gels were silver stained using the Invitrogen kit,[30] and visualized by Bio-Rad Chemi-Doc MP Imager using the silver stain channel. Bands were excised with a surgery knife by hand and treated with in-gel digestion buffer (10 mM NH_4HCO_3, 5% ACN, 1mM $CaCl_2$, 10 ng/µL trypsin). The supernatant containing the trypsin-digested peptides was desalted using Stage Tips and prepared for LC-MS. All the 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.[2] 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. [30]

SDS-PAGE analysis and fluorescence scanning

For labeling procedures see below. All the labeling samples were pre-incubated in 150 mM McIlvaine buffer on ice for 5 min. Electrophoresis was performed with sodium dodecylsulfate containing 10% polyacrylamide gels. Wet slab-gels were then scanned for ABP-emitted fluorescence using a Bio-Rad ChemiDoc MP imager using green Cy2 (λ_{EX} 470 nm, bandpass 30 nm; λ_{EM} 530 nm, bandpass 28) for **3** JJB382, red Cy3 (λ_{EX} 530 nm, bandpass 28 nm; λ_{EM} 605 nm, bandpass 50) for **4** JJB347, and blue Cy5 (λ_{EX} 625 nm, bandpass 30 nm; λ_{EM} 695 nm, bandpass 55) for **5** JJB383. All samples were denatured with 5× 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), boiled for 5 min at 100 °C, and separated by gel electrophoresis on 10% (w/v) SDS-PAGE gels running continuously at 90 V for 30 min and 200 V for 50 min.

***In vitro* labeling of GAA**

The detection limit of each ABP was analyzed by labeling 1 pmol rGAA with 1,000−0.01 fmol **3** JJB382 in 150 mM McIlvaine buffer, pH 5.0, for 30 min at 37 °C. Influence of pH on ABP labeling involved pre-incubation of either 100 fmol rGAA or 10 µg murine liver lysate at pH 2−8 for 30 min on ice, prior to addition of 1.0 µM **3** JJB382, dissolved in Nanopure H_2O and incubating for 30 min at 37 °C. Assessment of 3 JJB382 labeling kinetics involved pre-cooling of 100 fmol rGAA on ice for 15 min, followed by addition of similarly cooled 100 nM **3** JJB382 solution. After mixing, **3** JJB382 labeling was chased for 0−60 min at either 4 °C or 37 °C, whereafter labeling was stopped by denaturation. For competitive ABPP on rGAA, 1 pmol rGAA was pre-incubated with inhibitors (100 µM **7** JJB307, **1** Chris021, **2** Chris022, 10 µM **4** JJB347, **5** JJB383 or **6** JJB384, 10 mM 4MU-α-D-glucopyranoside, 2.5 M maltose) for 30 min at 37 °C, or boiled for 4 min in 2% (w/v) SDS, prior to labeling with 1.0 µM **3** JJB382 for 30 min at 37 °C. Labeling of human fibroblast lysates was performed on 100 µg total protein, using 1.0 µM **3** JJB382 dissolved in either 150 mM McIlvaine buffer, pH 4.0 for labeling of GAA, or pH 7.0 for labeling of GANAB. For pH optimum labeling on lysate of C57Bl6/J mouse liver or human fibrlblast, 50 µg total protein was incubated with 1.0 µM **3** JJB382 in McIlvaine buffers of pH 2.0−8.0 for 30 min at 37 °C. For competitive ABPP on C57Bl6/J mouse liver, 50 µg of total protein was pre-incubated with compounds 18-23 with decreasing concentrations (1000 µM, 100 µM, 10 µM, 1.0 µM) in 150 mM McIlvaine buffer pH 4.0 and pH 7.0 for 30 min at 37 °C, prior to labeling with 1.0 µM **5** JJB383 for 30 min at 37 °C.

Visualizing pulled-down proteins by fluorescent detection and Western Blot.

After pull-down of mouse intestinal enzymes using **6** JJB384, the supernatants and eluents (20 µL from both) of all treatment conditions (DMSO, competition, and **6** JJB384, under pH 4.0 or pH 7.0) were subjected to SDS-PAGE and fluorescent detection (Bio-Rad ChemiDoc MP Imager; Cy2 channel). Proteins were subsequently transferred to PVDF membranes using a Trans-Blot® Turbo system (BioRad), blocked in 4% (w/v) BSA in TBST, and detected for biotin **6** JJB384 binding with Streptavidin-HRP antibody (Dako). The blot was developed in the dark using a 10 mL luminal solution, 100 µL ECL enhancer and 3.0 µL 30% H_2O_2 solution. Chemiluminescence was visualized using the same ChemiDoc imager (BioRad).

***In situ* labeling of GAA**

Fibroblasts were treated with **5** JJB383 (100 nM, for 10 to 240 min) in culture medium. After washing 3 times with PBS, cells were harvested in potassium phosphate buffer (25 mM K_2HPO_4-KH_2PO_4, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)), prepared with or without 10 µM **3** JJB382. 20 µg total protein

from each cell homogenate of the lysate was subjected to SDS-PAGE. Labeling were visualized by fluorescent detection using Bio-Rad ChemiDoc MP Imager under the channels Cy5 for **5** JJB383 and Cy2 for **3** JJB382 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 **5** JJB383 (100 nM, for 3 h) in medium. Cells were harvested with or without 10 mM **3** JJB382 with the above describe methods. Homogenates were denatured, resolved on SDS-PAGE and detected for **5** JJB383 labeling by fluorescent scanning.

Pompe GAA detection by Western blot

Following fluorescent scanning of SDS-PAGE, proteins on wet slab gel were transferred to PVDF membrane and blocked as described in the previous section. For GAA detection, the membrane was incubated firstly with mouse polyclonal anti GAA and subsequently with goat anti mouse Alexa647 (Life Technologies). Blot was scanned on a Typhoon FLA 9500 Imager (GE Healthcare) using 633 nm laser and LPR filter, and 100 mm as pixel size. Rabbit anti β-actin (Cell Signaling) and goat anti rabbit Alexa647 (Invitrogen) were used for loading control.

***Agd31B* expression and 3-D crystallography**

*A*gd31B expression and purification was carried out as previously described.[24] Protein crystals 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 **1** CF021 and **2** CF022 were obtained by soaking in mother liquor containing 5.0 mM probe for 2 h, before cyroprotecting in 2.0 M lithium sulfate, 0.10 M HEPES (pH = 7.0), 2% PEG 400, and flash freezing in liquid N_2 for data collection (Table 2).

All data were collected at beamline I04 of the Diamond Light Source, processed using XDS[31] and reduced using Aimless.[32] Complex structures were solved by molecular replacement using MolRep,[33] before subsequent rounds of manual model building and refinement using Coot[34] and REFMAC5[35] respectively. Refinements were carried out using TLS determination of molecular motions.[36] Ligand coordinates were built using jLigand.[37] Crystal structure figures were generated using ccp4mg.[38]

Table 2. Crystal data collection and refinement statistics

	C*j*Agd31B-**2** CF022-complex PDB: 5123	C*j*Agd31B- **1** Chris021-complex PDB: 5124
Data collection		
Space group	P622	P622
Cell dimensions		
a, b, c (Å)	198.0, 198.0, 103.0	197.3, 197.3, 102.9
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	85.74-1.95 (1.99-1.95)	49.83-1.85 (1.88-1.85)
R_{merge}	0.13 (1.64)	0.10 (1.21)
$I/\sigma I$	19.6 (2.3)	26.8 (3.0)
Completeness (%)	100 (100)	100 (100)
Redundancy	20.0 (20.1)	24.5 (25.0)
Refinement		
Resolution (Å)	85.74-1.95	49.83-1.85
No. reflections	82110	95234
R_{work}/ R_{free}	0.17/0.19	0.17/0.19
Protein	6299	6295
Ligand/ion	102	81
Water	616	625
B-factors (TLS refinement)		
Protein	33.8	31.99
Ligand/ion	64.8	54.42
Water	39.7	38.0
R.m.s deviations		
Bond lengths (Å)	0.012	0.013
Bond angles (°)	1.57	1.57

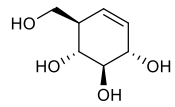
Values in parentheses are for highest-resolution shell.

Synthesis:

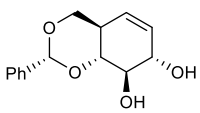
General synthesis

All reagents were of a commercial grade and were used as received unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and *N, N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an argon atmosphere unless stated otherwise. Solvents used for flash column chromatography were of pro analysis quality. Reactions were monitored by TLC analysis using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}.H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4.H_2O$ (10 g/L) in 10% sulfuric acid followed by charring at

~150 °C or by spraying with an aq. solution of $KMnO_4$ (7%) and K_2CO_3 (2%) followed by charring at ~150 °C. Column chromatography was performed using either Baker or Screening Device silica gel 60 (0.04 - 0.063 mm) in the indicated solvents. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker AV-850 (850/214 MHz), Bruker DMX-600 (600/150 MHz) and Bruker AV-400 (400/100 MHz) spectrometer in the given solvent. Chemical shifts are given in ppm relative to the deuterated chloroform or methanol residual solvent peak or tetramethylsilane (TMS) as internal standard. Coupling constants are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. High-resolution mass spectra were recorded with a LTQ Orbitrap (Thermo Finnigan). Optical rotations were measured on Propol automatic polarimeter (Sodium D-line, λ = 589 nm). LC-MS analysis was performed on an LCQ Adventage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C_{18} column (Gemini, 4.6 mm x 50 mm, 3.0 µm particle size, Phenomenex) equipped with buffers A: H_2O, B: MeCN (MeCN) and C: 1% aq. TFA, For reversed-phase HPLC purifications an Agilent Technologies 1200 series instrument equipped with a semi preparative Gemini C_{18} column (10 x 250 mm) was used. The applied buffers were A: 50 mM NH_4HCO_3 in H_2O, B: MeCN.

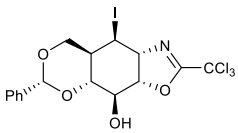


(1R,2R,3S,6R)-6-(Hydroxymethyl)-cyclohex-4-ene-1,2,3-triol The starting material diol **8** was prepared according to previously reported procedure.[22] Diol compound **8** (680 mg, 2.0 mmol, 1.0 eq.) was co-evaporated 3 times with toluene and then dissolved in dry THF (25 mL) and cooled to -60°C under argon atmosphere. Ammonia (20 mL) was condensed at -60°C under argon atmosphere. Lithium (207 mg, 30.0 mmol, 15 eq.) was added and the mixture was stirred until lithium was completely dissolved. To this solution was slowly added the solution of **8** in THF. The reaction mixture was stirred for 30 min at -60 °C and then quenched by adding of H_2O (30 mL). The resulting solution was allowed to come to room temperature and stirred until all ammonia had evolved. Next, the solution was concentrated *in vacuo*, redissolved in H_2O, and neutralized with Amberlite IR-120 H^+. Then, the filtration mixture was concentrated *in vacuo* and purified by silica gel column chromatography (10%→20%, MeOH in DCM) to afford product as colorless oil (180 mg, 1.1 mmol, 57%). TLC: R_f 0.48 (DCM/MeOH, 4/1, v/v); $[\alpha]_D^{20}$ +105.2 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 5.66 – 5.57 (m, 2H), 4.06 – 4.03 (m, 1H), 3.81 (dd, *J* = 10.6, 4.1 Hz, 1H), 3.66 - 3.59 (m, 1H), 3.49 - 3.41 (m, 2H), 2.31 - 2.26 (br, 1H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 130.97, 128.57, 78.81, 73.62, 72.03, 63.45, 47.65; HRMS: calculated for $C_7H_{12}O_4$ $[M+H^+]$ 161.08084, found: 161.08087.

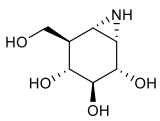


(1R,3R,6R,9S,10S)-9,10-Dihydroxy-3-phenyl-2,4-dioxabicyclo[4.4.0]dec-7-ene (9) (1R,2R,3S,6R)-6-(Hydroxymethyl)-cyclohex-4-ene-1,2,3-triol (180 mg, 1.1 mmol, 1.0 eq.) was dissolved in dry DMF (2.0 mL) and dry MeCN (6.0 mL) in an inert atmosphere. 10-Camphorsulfonic acid (52 mg, 0.25 mmol, 0.20 eq.) was added to the solution, followed by $PhCH(OMe)_2$ (253 µL, 1.7 mmol, 1.5 eq.). After 48 h, the reaction was quenched with Et_3N (31.0 µL, 0.23 mmol, 0.2 eq.) and concentrated *in vacuo*. The reaction mixture was separated out with EtOAc and H_2O and the aq. layer was further extracted with EtOAc. The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (30%→70%, EtOAc in pentane) to afford **9** (170 mg, 0.69 mmol, 61%). TLC: R_f 0.45 (pentane/EtOAc, 7/3, v/v); $[\alpha]_D^{20}$ +30.8 (10 mg/mL in $CDCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.52 – 7.47 (m, 2H), 7.37 – 7.31 (m, 3H), 5.57 – 5.52 (m, 2H), 5.28 – 5.25 (m, 1H), 4.24 - 4.18 (m, 2H), 3.86 (dd, *J* = 10.1, 7.5 Hz, 1H), 3.58 – 3.51 (m, 2H), 2.84 (d, *J* = 8.0 Hz,1H), 2.60 – 2.54 (m, 1H); ^{13}C-NMR

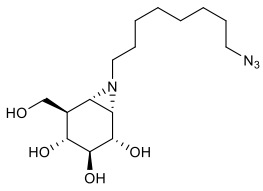
(100 MHz, $CDCl_3$): δ ppm 137.79, 130.49, 130.64, 129.16, 128.34, 126.39, 124.05, 102.16, 80.72, 75.42, 73.74, 69.96, 38.53; HRMS: calculated for $C_{14}H_{16}O_4$ [M+H$^+$] 249.11214, found: 249.11224.



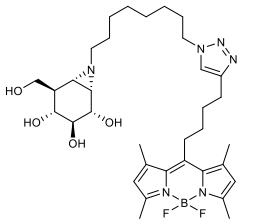
(1R,6R,7R,8R,9S,10R)-10-Hydroxy-7-iodo-3-phenyl-12-trichloromethyl-13-aza-2,4,11-trioxatricyclo[7.4.4.0-8.4.3.0]tridec-12-ene (11) A solution of **9** (535 mg, 2.1 mmol, 1.0 eq.) in dry DCM (60 mL) was made under argon atmosphere and cooled to 0 °C. DBU (64 µL, 0.43 mmol, 0.20 eq) and trichloroacetonitrile (216 µL, 11 mmol, 5 eq.) was added to the solution. After 3 h, more trichloroacetonitrile (108 µL, 5.4 mmol, 2.5 eq.) was added to the reaction mixture. Starting material was fully conversed to imidate **10** with higher running spot on TLC after 21 h stirring at room temperature. Then iodine (1.7 g, 6.7 mmol, 3.1 eq), $NaHCO_3$ (1.8 g, 21 mmol, 10 eq.) and H_2O (3.8 mL) was added to the reaction mixture. After 26 h of the first addition of the iodine and $NaHCO_3$, iodine (1.70 g, 6.68 mmol, 3.1 eq.) and $NaHCO_3$ (1.81 g, 21.5 mmol, 10 eq.) was added to the reaction mixture. After approximately 96 h after the first addition of iodine and $NaHCO_3$, the reaction was quenched by adding $Na_2S_2O_3$ (10% aq. solution) until the purple color had disappeared. The reaction mixture was separated out and the DCM layer was concentrated *in vacuo*, redissolved in EtOAc and washed with H_2O and brine. The initial aq. layer was extracted with EtOAc and the combined organic layers were dried over $MgSO_4$, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (0%→16%, EtOAc in pentane) to afford **11** (455 mg, 0.88 mmol, 41%) was produced. TLC: R_f 0.57 (PE/EtOAc, 4/1, v/v); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.49 – 7.44 (m, 2H), 7.49 – 7.33 (m, 3H), 5.62 (s, 1H), 5.22 (t, J = 7.5 Hz, 1H), 4.85 – 4.83 (m, 1H), 4.72 (s, 1H), 4.21 (dd, J = 11.3, 4.7 Hz, 1H), 4.03 (t, J = 9.8 Hz, 1H), 3.92 (t, J = 10.6 Hz, 1H), 3.83 – 3.79 (m, 1H), 3.08 (s, 1H), 1.20 – 1.13 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 163.72, 137.29, 129.48, 128.51, 126.26, 101.64, 86.69, 77.77, 75.19, 74.93, 72.57, 34.90, 23.69; HRMS: calculated for $C_{16}H_{15}Cl_3NO_4$ [M+H$^+$] 519.91567, found: 519.91504.



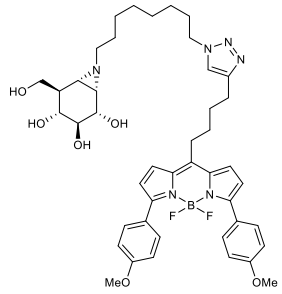
(1S,2R,3S,4R,5R,6R)-5-(Hydroxymethyl)-7-aza-bicyclo-[4.1.0]heptane-2,3,4-triol (1) A solution of **11** (455 mg, 0.88 mmol, 1.0 eq.) was dissolved in 1, 4-dioxane (9.2 mL) and heated to 60 °C. Then aq. HCl (37%, 2.60 mL) was added to the solution. The reaction mixture was stirred at 60 °C overnight. The reaction mixture was concentrated *in vacuo* and then separated out with EtOAc and H_2O. The aq. layer was washed with EtOAc and concentrated *in vacuo* and co-evaporated with toluene. The crude product of the free amine intermediate **12** was dissolved in MeOH (30 mL) and $NaHCO_3$ (2.9 g, 35 mmol, 40 eq.) was added to the solution. The reaction mixture was stirred at room temperature for 4 days. The reaction mixture filtered and then concentrated *in vacuo*. The residue was redissolved in H_2O and filtered over a pad of Amberlite IR-120 H$^+$ resin, washed with H_2O and followed by 1.0 M aq. NH_4OH. The filtrate was concentrated *in vacuo* to afford the aziridine product **1** (97 mg, 0.56 mmol, 63%) as light brown oil. ^{1}H-NMR (400 MHz, D_2O): δ ppm 4.11 – 4.04 (m, 2H), 3.97 – 3.90 (m, 1H), 3.55 – 3.39 (m, 2H), 2.82 – 2.80 (m, 1H), 2.57 (d, J = 6.4 Hz, 1H), 2.11 – 2.01 (m, 1H). ^{13}C-NMR (100 MHz, D_2O): δ ppm 73.46, 71.27, 70.07, 61.50, 44.37, 35.66, 31.69; HRMS: calculated for $C_7H_{13}NO4$ [M+H$^+$] 176.09173, found: 176.09163.



(1S,2R,3S,4R,5R,6R)-2,3,4-trihyrdoxy-5-(hydroxymethyl)-7-(8-azidooctyl)-7-aza-bicyclo[4.1.0]heptane (2) A solution of **1** (97 mg, 0.56 mmol, 1.0 eq.) in dry DMF (2.2 mL) was made under argon atmosphere and heated to 80 °C. 1-azido-8-iodooctane (234 mg, 0.83 mmol, 1.5 eq) and K_2CO_3 (345 mg, 2.5 mmol, 4.5 eq.) were added to the solution. After 21 h stirring, the reaction was quenched with MeOH (0.15 mL) and filtered over a pad of celite. The filtrate was concentrated *in vacuo*. The product was purified by silica gel column chromatography (5%→18% MeOH in DCM) to afford product **2** CF022 as yellow oil (72 mg, 0.22 mmol, 39%). TLC: R_f 0.39 (DCM/MeOH, 3/1, v/v); $[\alpha]_D^{21}$ +25.4 (*c* = 1, MeOH); ^{1}H-NMR (850 MHz, CD_3OD): *δ* ppm 3.87 (dd, *J* = 10.7, 4.0 Hz, 1H), 3.66 (dd, *J* = 8.6, 3.7 Hz, 1H), 3.63 (dd, *J* = 10.8, 7.1 Hz, 1H), 3.34 – 3.31 (m, 1H), 3.29 (t, *J* = 6.8 Hz, 2H), 3.05 (t, *J* = 10.0 Hz, 1H), 2.36 – 2.33 (m, 1H), 2.17 – 2.14 (m, 1H), 1.86 – 1.82 (m, 2H), 1.68 (d, *J* = 6.5 Hz, 1H), 1.60 – 1.57 (m, 4H), 1.40 – 1.32 (m, 8H); ^{13}C-NMR (214 MHz, CD_3OD): *δ* ppm 75.76, 73.37, 72.45, 63.52, 62.25, 52.46, 46.89, 46.00, 41.95, 30.58, 30.48, 30.19, 29.90, 28.35, 27.77; LC/MS: R_t 4.42 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 329.20 $(M+H)^+$; HRMS: calculated for $C_{15}H_{28}N_4O_5$ $[M+H^+]$ 329.21833, found: 329.21828.

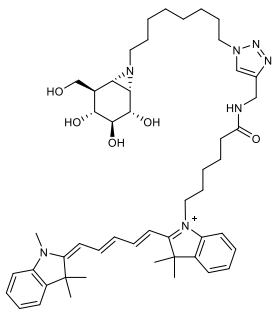


(1S,2S,3S,4R,5R,6S)-7-(8-(4-(4-(5,5-difluoro-1,3,7,9-tetramethyl-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (3) A solution of azide compound **2** (17 mg, 0.052 mmol, 1.0 eq.) in DMF (1.5 mL) was made under argon atmosphere, green BODIPY compound **13** (19 mg, 0.058 mmol, 1.1 eq), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.010 mmol, 0.20 eq.) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.011 mmol, 0.21 eq) were added to the solution and the mixture was stirred at room temperature for 12 h. The resulting mixture was concentrated under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 45%→53% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized to afford product **3** as orange solid (13 mg, 20 µmol, 39%). ^{1}H-NMR (850 MHz, CD_3OD): *δ* ppm 7.73 (s, 1H), 6.11 (s, 2H), 4.34 (t, *J* = 7.0 Hz, 2H), 3.87 (dd, *J* = 8.7, 3.9 Hz, 1H), 3.65 (dd, *J* = 8.5, 3.7 Hz, 1H), 3.61 (dd, *J* = 10.7, 7.2 Hz, 1H), 3.31 – 3.30 (m, 1H), 3.05 (t, *J* = 10.0 Hz, 1H), 3.02 – 2.97 (m, 2H), 2.77 (t, *J* = 7.3 Hz, 2H), 2.43 (s, 6H), 2.37 (s, 6H), 2.34 – 2.26 (m, 1H), 2.15 – 2.11 (m, 1H), 1.91 – 1.81 (m, 6H), 1.67 – 1.62 (m, 3H), 1.57 – 1.53 (m, 2H), 1.36 – 1.22 (m, 8H).; ^{13}C-NMR (214 MHz, CD_3OD) *δ* ppm 154.90, 148.49, 147.88, 142.18, 132.57, 123.36, 122.60, 75.74, 73.34, 72.41, 63.50, 62.19, 51.20, 46.86, 45.99, 41.93, 32.23, 31.24, 30.83, 30.42, 29.88, 29.05, 28.22, 27.32, 25.87, 16.48, 14.44; LC/MS: R_t 5.88 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 657.33 $(M+H)^+$, 679.53 $[M+Na]^+$; HRMS: calculated for $C_{34}H_{51}BF_2N_6O_4$ $[M+H^+]$ 657.41117, found: 657.41033.

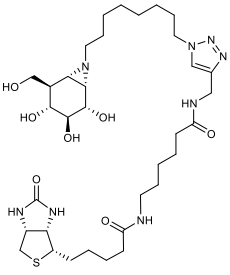


(1S,2S,3S,4R,5R,6S)-7-(8-(4-(4-(5,5-difluoro-3,7-bis(4-methoxyphenyl)-5H-4 ^{4},5 ^{4}-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (4) A solution of azide compound **2** (18 mg, 0.055 mmol, 1.0eq.) in DMF (1.5 mL) was made under argon atmosphere, red BODIPY compound **14** (29 mg, 0.060 mmol, 1.1 eq), $CuSO_4$ (1.0 M in H_2O, 11 µL, 0.011 mmol, 0.20 eq.) and sodium ascorbate (1.0 M in H_2O, 12 µL, 0.012 mmol, 0.21 eq) were added to the solution and the mixture was stirred at

room temperature for 14 h. The resulting mixture was concentrated under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 51%→57% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized to afford product **4** as purple solid (5.1 mg, 6.3 µmol, 12%). ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 7.88 – 7.79 (m, 4H), 7.68 (s, 2H), 7.43 (d, *J* = 4.3 Hz, 3H), 7.02 – 6.94 (m, 4H), 6.69 (d, *J* = 4.3 Hz, 2H), 4.32 (t, *J* = 7.1 Hz, 2H), 3.86 – 3.84(m, 7H), 3.64 (dd, *J* = 8.5, 3.7 Hz, 1H), 3.60 (dd, *J* = 10.7, 7.1 Hz, 1H), 3.32 – 3.31 (m, 1H), 3.07 – 3.03 (m, 3H), 2.78 (t, *J* = 6.9 Hz, 2H), 2.29 – 2.26 (m, 1H), 2.12 – 2.09 (m, 1H), 1.88 – 1.79 (m, 6H), 1.64 (d, *J* = 6.5 Hz, 1H), 1.55 – 1.49 (m, 2H), 1.30 – 1.20 (m, 8H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 162.20, 158.79, 148.60, 146.78, 137.49, 132.16, 132.14, 132.12, 128.42, 126.51, 123.24, 121.03, 114.85, 114.63, 75.75, 73.36, 72.46, 63.53, 62.18, 55.83, 51.24, 46.86, 45.97, 41.92, 34.13, 31.23, 30.98, 30.40, 30.33, 29.87, 28.20, 27.33, 25.78; LC/MS: R_t 6.82 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 813.33 $(M+H)^+$; HRMS: calculated for $C_{44}H_{55}BF_2N_6O_6$ $[M+H^+]$ 813.43245, found: 813.43137.

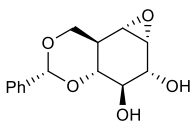


3,3-dimethyl-1-(6-oxo-6-(((1-(8-((1S,2S,3S,4R,5R,6S)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)amino)hexyl)-2-((1E,3E)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-indol-1-ium (5) A solution of azide compound **2** (17 mg, 0.052 mmol, 1.0 eq.) in DMF (1.5 mL) was made under argon atmosphere, Cy5 compound **15** (29 mg, 0.057 mmol, 1.1 eq), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.010 mmol, 0.20 eq.) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.011 mmol, 0.21 eq) were added to the solution and the mixture was stirred at room temperature for 12h. The resulting mixture was concentrated under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 40%→70% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized to afford product **5** as dark blue powder (10 mg, 12 µmol, 23%). ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 8.32 – 8.16 (m, 2H), 7.84 (s, 1H), 7.55 – 7.48 (m, 2H), 7.41 (m, 2H), 7.35 – 7.21 (m, 4H), 6.62 (t, *J* = 12.4 Hz, 1H), 6.29 – 6.27 (m, 2H), 4.41 (s, 2H), 4.36 (t, *J* = 7.1 Hz, 2H), 4.09 (t, *J* = 7.5 Hz, 2H), 3.86 (dd, *J* = 10.7, 3.9 Hz, 1H), 3.69 – 3.57 (m, 5H), 3.33 – 3.30 (m, 1H), 3.10 – 3.00 (m, 1H), 2.34 – 2.29 (m, 1H), 2.25 (t, *J* = 7.3 Hz, 2H), 2.15 – 2.11 (m, 1H), 1.93 (s, 6H), 1.90 – 1.77 (m, 6H), 1.73 – 1.57 (m, 14H), 1.60 – 1.51 (m, 2H), 1.50 – 1.42 (m, 2H), 1.33 – 1.27 (m, 8H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 178.47, 175.73, 175.38, 174.59, 155.54, 155.47, 146.13, 144.23, 143.53, 142.61, 142.51, 129.77, 129.73, 126.60, 126.27, 126.21, 124.15, 123.41, 123.28, 112.01, 111.84, 104.41, 104.23, 75.72, 73.31, 72.39, 63.48, 62.17, 51.33, 50.53, 50.50, 46.87, 46.00, 44.74, 41.92, 36.46, 35.56, 31.50, 31.30, 30.44, 30.41, 29.94, 28.25, 28.12, 27.93, 27.78, 27.37, 27.30, 26.39; LC/MS: R_t 5.18 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 848.60 $(M)^+$; HRMS: calculated for $C_{50}H_{70}N_7O_5^+$ $[M+H^+]$ 849.54653, found: 849.55112.

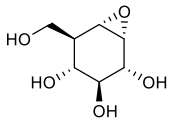


6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)-N-((1-(8-((1S,2S,3S,4R,5R,6S)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)hexanamide (6) A solution of azide compound **2** (15 mg, 0.0457 mmol, 1.0 eq.) in DMF (1.5 mL) was made under argon atmosphere, biotin compound **16** (19.8 mg, 0.0502 mmol, 1.1 eq), $CuSO_4$ (1.0 M in H_2O, 9.1 µL, 0.0091

mmol, 0.20 eq.) and sodium ascorbate (1.0 M in H_2O, 9.6 µL, 0.0096 mmol, 0.21 eq) were added to the solution and the mixture was stirred at room temperature for 12 h. The resulting mixture was concentrated under reduced pressure and the crude product was purified by semi-preparative reversed HPLC (linear gradient: 19%→28% B in A, 12 min, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized to afford product **6** as white solid (7.6 mg, 11 µmol, 23%). ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 7.85 (s, 1H), 4.49 (dd, *J* = 7.8, 4.8 Hz, 1H), 4.44 – 4.34 (m, 4H), 4.30 (dd, *J* = 7.9, 4.5 Hz, 1H), 3.86 (dd, *J* = 10.7, 3.9 Hz, 1H), 3.69 – 3.58 (m, 2H), 3.34 – 3.30 (m, 1H), 3.25 – 3.10 (m, 3H), 3.05 (t, *J* = 10.0 Hz, 1H), 2.93 (dd, *J* = 12.7, 5.0 Hz, 1H), 2.74 – 2.64 (m, 1H), 2.37 – 2.29 (m, 1H), 2.26 – 2.10 (m, 5H), 1.95 – 1.79 (m, 4H), 1.79 – 1.40 (m, 14H), 1.37 – 1.30 (m, 10H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 176.01, 166.11, 161.52, 146.22, 124.16, 75.74, 73.34, 72.41, 63.50, 63.38, 62.20, 61.62, 57.02, 51.35, 46.89, 46.01, 41.95, 41.05, 40.18, 36.81, 36.75, 35.58, 31.29, 30.45, 30.12, 29.95, 29.78, 29.50, 28.26, 27.54, 27.37, 26.93, 26.52; LC/MS: R_t 4.37 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 723.47 $(M+H)^+$; HRMS: calculated for $C_{34}H_{58}N_8O_7S$ $[M+H^+]$ 723.42245, found: 723.42219.



(1aR,2R,3R,3aR,7aS,7bS)-5-phenylhexahydro-2H-oxireno[2',3':3,4]benzo[1,2-d][1,3]dioxine-2,3-diol (17) A solution of compound **9** (100 mg, 0.40 mmol, 1.0 eq.) and 3-chloroperbenzoic acid (*m*CPBA, 75%) (135 mg, 0.60 mmol, 1.5 eq) in dry DCM (10 mL) were made under argon atmosphere and refluxed at 40°C for 24h. The reaction mixture was washed by sat. aq. $NaHCO_3$ and brine. The organic layer was dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified via silica gel column chromatography (50%→100%, EtOAc in pentane) to afford the title compound **17** (47 mg, 0.18 mmol, 44%). TLC: R_f 0.30 (pentane/EtOAc, 3/1, v/v); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.49 – 7.46 (m, 2H), 7.37 – 7.33 (m, 3H), 5.48 (s, 1H), 4.38 (dd, *J* = 12.0, 4.0 Hz, 1H), 3.97 – 3.95 (m, 1H), 3.74 – 3.68 (m, 1H), 3.64 – 3.60 (m, 1H), 3.32 – 3.25 (m, 2H), 2.78 (d, *J* = 4.0 Hz, 1H), 2.22 – 2.15 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 137.46, 129.38, 128.45, 126.44, 102.14, 79.79, 73.32, 72.76, 68.37, 55.68, 51.71, 38.00; HRMS: calculated for $C_{14}H_{16}O_5$ $[M+H^+]$ 265.10705, found: 265.10720.



(1R,2R,3S,4R,5R,6S)-5-(hydroxymethyl)-7-oxabicyclo[4.1.0]heptane-2,3,4-triol (7) A mixture of product **17** (20 mg, 0.076 mmol, 1.0 eq.) and $Pd(OH)_2/C$ (15 mg, 20% wt. loading(dry basis)) in MeOH (2.0 mL) was stirred at room temperature under hydrogen atmosphere overnight. The catalyst was then filtered off and washed with MeOH. The filtrate and washings were combined and concentrated under reduced pressure. Crude product was purified by semi-preparative reversed phase HPLC (linear gradient: 0%→20%, 3 CV, solutions used A: 50mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilization to afford the title compound *epi*-cyclophellitol **7** (9.0 mg, 0.051 mmol, 68%) as white solid. TLC: R_f 0.26 (MeOH:$CDCl_3$, 1/3, v/v); ^{1}H-NMR (400 MHz, D_2O): *δ* ppm 3.91 – 3.86 (m, 2H), 3.78 – 3.74 (m, 1H), 3.44 – 3.43 (m, 1H), 3.41 - 3.37 (m, 1H), 3.34 – 3.29 (m, 2H), 2.04 – 1.99 (m, 1H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 73.05, 71.23, 69.35, 60.32, 57.47, 55.09, 44.12; HRMS-: calculated for $C_7H_{12}O_5$ $[M+H^+]$ 177.07575, found: 177.07571.

5.5 References

[1] L. H. Hoefsloot, M. Hoogeveen-Westerveld, A. J. Reuser and B. A. Oostra, *Biochem. J.* **1990**, 272, 493-497.

[2] S. Chiba, K. Hiromi, N. Minamiura, M. Ohnishi, T. Shimomura, K. Suga, T. Suganuma, A. Tanaka, S. Tomioka and T. Yamamoto, *J. Biochem*. **1979**, 85, 1135-1141.

[3] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic acids Res.* **2014**, 42, D490-495.

[4] G. J. Davies and S. J. Williams, *Biochem. Soc. Trans.* **201**6, 44, 79-87.

[5] R. J. Moreland, X. Jin, X. K. Zhang, R. W. Decker, K. L. Albee, K. L. Lee, R. D. Cauthron, K. Brewer, T. Edmunds and W. M. Canfield, *J. Biol. Chem.* **2005**, 280, 6780-6791.

[6] H. A. Wisselaar, M. A. Kroos, M. M. Hermans, J. van Beeumen and A. J. Reuser, *J. Biol. Chem.* **1993**, 268, 2223-2231.

[7] S. S. Lee, S. He and S. G. Withers, *Biochem. J*. **2001**, 359, 381-386.

[8] D. E. Koshland, *Biol. Rev.* **1953**, 28, 416-436.

[9] H. G. Hers, *Biochem. J.***1963**, 86, 11-16.

[10] A. J. Reuser, M. A. Kroos, M. M. Hermans, A. G. Bijvoet, M. P. Verbeet, O. P. Van Diggelen, W. J. Kleijer and A. T. Van der Ploeg, *Muscle Nerve. Suppl.* **1995**, 3, S61-69.

[11] P. S. Kishnani, W. L. Hwu, H. Mandel, M. Nicolino, F. Yong and D. Corzo, *J. Pediatr.* **2006**, 148, 671-676.

[12] L. P. Winkel, M. L. Hagemans, P. A. van Doorn, M. C. Loonen, W. J. Hop, A. J. Reuser and A. T. van der Ploeg, *J. Neurol.* **2005**, 252, 875-884.

[13] P. S. Kishnani, D. Corzo, M. Nicolino, B. Byrne, H. Mandel, W. L. Hwu, N. Leslie, J. Levine, C. Spencer, M. McDonald, J. Li, J. Dumontier, M. Halberthal, Y. H. Chien, R. Hopkin, S. Vijayaraghavan, D. Gruskin, D. Bartholomew, A. van der Ploeg, J. P. Clancy, R. Parini, G. Morin, M. Beck, G. S. De la Gastine, M. Jokic, B. Thurberg, S. Richards, D. Bali, M. Davison, M. A. Worden, Y. T. Chen and J. E. Wraith, *Neurology* **2007**, 68, 99-109.

[14] A. T. van der Ploeg, P. R. Clemens, D. Corzo, D. M. Escolar, J. Florence, G. J. Groeneveld, S. Herson, P. S. Kishnani, P. Laforet, S. L. Lake, D. J. Lange, R. T. Leshner, J. E. Mayhew, C. Morgan, K. Nozaki, D. J. Park, A. Pestronk, B. Rosenbloom, A. Skrinar, C. I. van Capelle, N. A. van der Beek, M. Wasserstein and S. A. Zivkovic, *N. Engl. J. Med.* **2010**, 362, 1396-1406.

[15] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft and J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, 6, 907-913.

[16] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, 51, 12529-12533.

[17] J. Jiang, T. J. M. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codee, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. Eur. J.* **2015**, 21, 10861-10869.

[18] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codee, G. A. van der Marel, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc*. **2014**, 136, 11622-11625.

[19] J. Jiang, W. W. Kallemeijn, D. W. Wright, A. M. C. H. van den Nieuwendijk, V. C. Rohde, E. C. Folch, H. van den Elst, B. I. Florea, S. Scheij, W. E. Donker-Koopman, M. Verhoek, N. Li, M. Schurmann, D. Mink, R. G. Boot, J. D. C. Codee, G. A. van der Marel, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. Sci.* **2015**, 6, 2782-2789.

[20] B. T. Adams, S. Niccoli, M. A. Chowdhury, A. N. Esarik, S. J. Lees, B. P. Rempel and C. P. Phenix, *Chem. Commun.* **2015**, 51, 11390-11393.

[21] A. Alcaide, A. Trapero, Y. Perez and A. Llebaria, *Org. Biomol. Chem.* **2015**, 13, 5690-5697.

[22] F. G. Hansen, E.. Bundgaard, R. Madsen, A Short Synthesis of (+)-Cyclophellitol. *J. Org. Chem.* **2005**, 70, 10139-10142..

[23] A. Ghisaidoobe, P. Bikker, A. C. de Bruijn, F. D. Godschalk, E. Rogaar, M. C. Guijt, P. Hagens, J. M. Halma, S. M. Van't Hart, S. B. Luitjens, V. H. van Rixel, M. Wijzenbroek, T. Zweegers, W. E. Donker-Koopman, A. Strijland, R. Boot,. G. van der Marel, H. S. Overkleeft, J. M. Aerts, R. J. van den Berg. Identification of Potent and Selective Glucosylceramide Synthase Inhibitors from a Library of N-Alkylated Iminosugars. *ACS Med. Chem. Lett.* **2011,** 2, 119-123.

[24] J. Larsbrink, A. Izumi, G. R. Hemsworth, G. J. Davies and H. Brumer, *J. Biol. Chem.* **2012**, 287, 43288-43299.

[25] G. J. Davies, A. Planas and C. Rovira, *Acc. Chem. Res.* **2012**, 45, 308-316.

[26] G. Speciale, A. J. Thompson, G. J. Davies and S. J. Williams, *Curr. Opin. Struct. Biol.* **2014**, 28, 1-13.

[27] B. Poole and S. Ohkuma, *J. Cell Biol.* **1981**, 90, 665-669.

[28] Y. Liu, M. P. Patricelli and B. F. Cravatt, *Proc. Natl. Acad. Sci. U S A* **1999**, 96, 14694-14699.

[29] D. Greenbaum, K. F. Medzihradszky, A. Burlingame and M. Bogyo, *Chem. Biol.* **2000**, 7, 569-581.

[30] N. Li, C.-L. Kuo, G. Paniagua, H. van den Elst, M. Verdoes, L. I. Willems, W. A. van der Linden, M. Ruben, E. van Genderen, J. Gubbens, G. P. van Wezel, H. S. Overkleeft and B. I. Florea, *Nat. Protocols* **2013**, 8, 1155-1168.

[31] W. Kabsch, *Acta Crystallogr. D* **2010**, 66, 125-132.

[32] P. R. Evans and G. N. Murshudov, *Acta Crystallogr. D* **2013**, 69, 1204-1214.

[33] A. Vagin and A. Teplyakov, MOLREP: an automated program for molecular replacement. *J Appl. Crystallogr.* **1997**, 30, 1022-1025.

[34] P. Emsley and K. Cowtan, *Acta Crystallogr D* **2004**, 60, 2126-2132.

[35] G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr. D* **1997**, 53, 240-255.

[36] J. Painter and E. A. Merritt, *Acta Crystallogr. D* **2006**, 62, 439-450.

[37] A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov, J. Painter and E. A. Merritt, *Acta Crystallogr. D* **2012**, 68, 431-440.

[38] S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallogr. D* **2011**, 67, 386-394.

5.6 Supporting Information

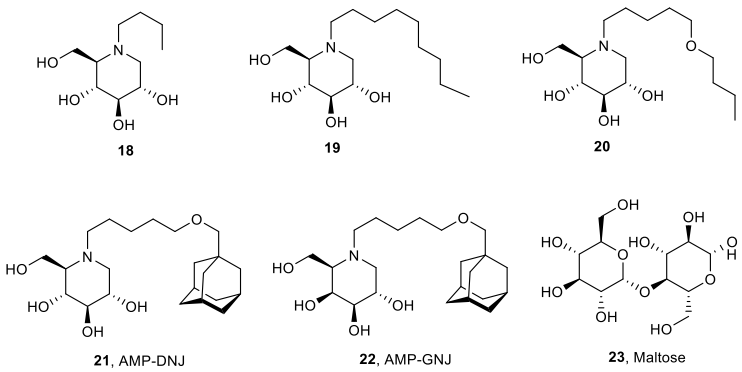


Figure S1. Chemical structures of iminosugar compounds **18**-**22** and maltose **23**.

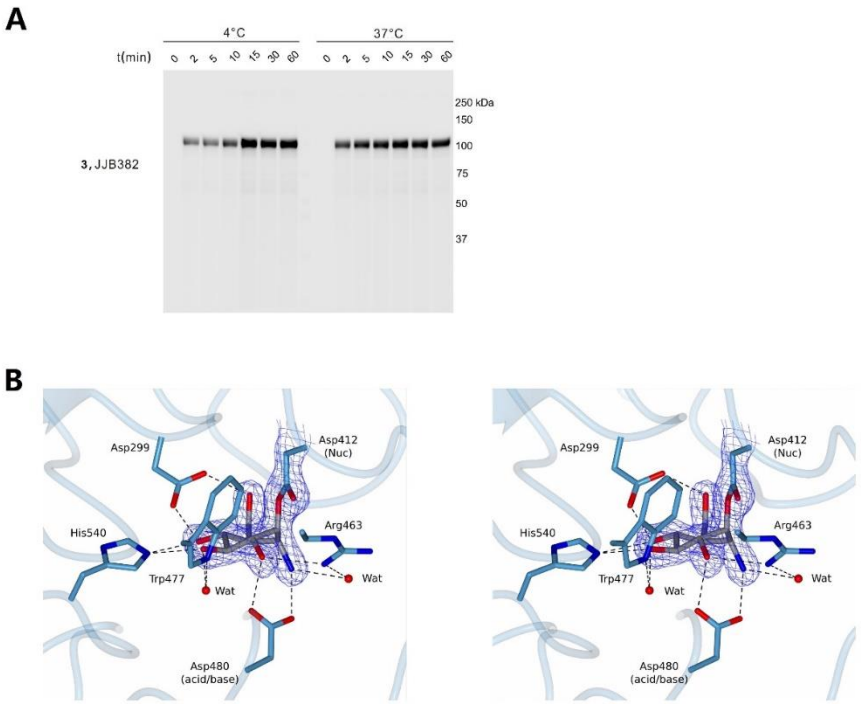


Figure S2. A) Rates determined through direct 4 °C and 37 °C different temperature labeling with **3** JJB382; B) 3-D view of crystal structure of C*j*Agd31B GAA-**1** CF021-complex.

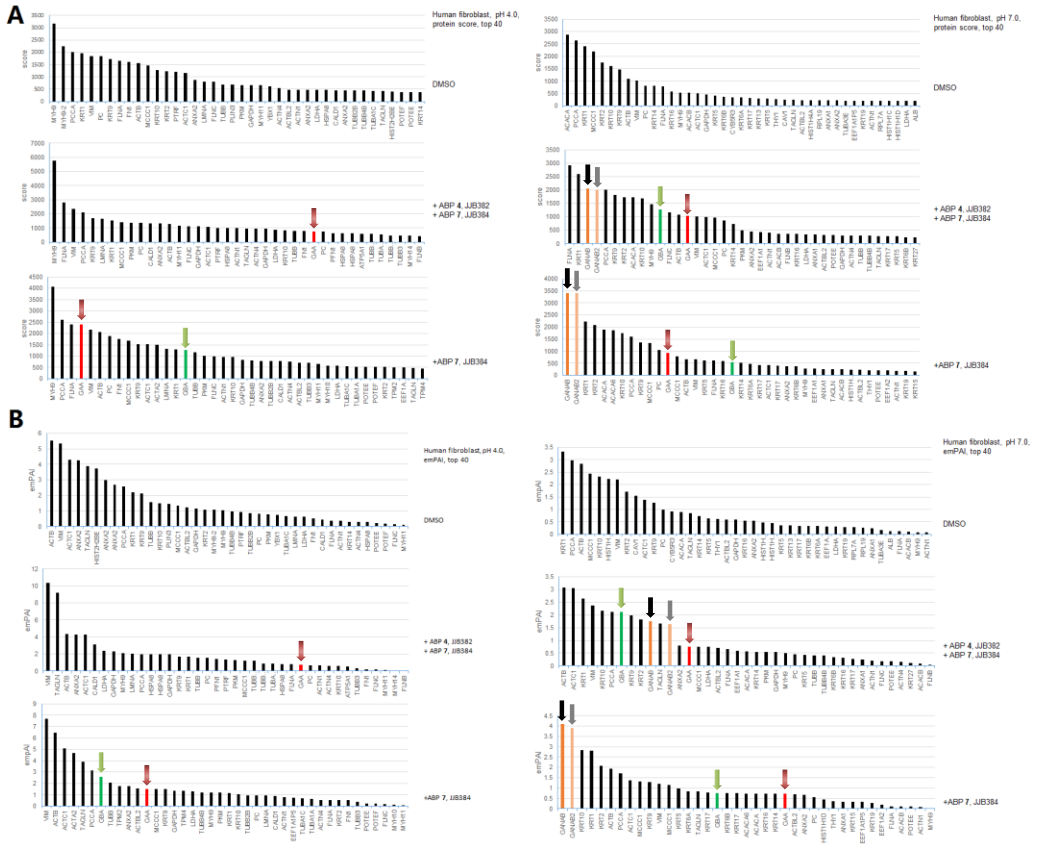
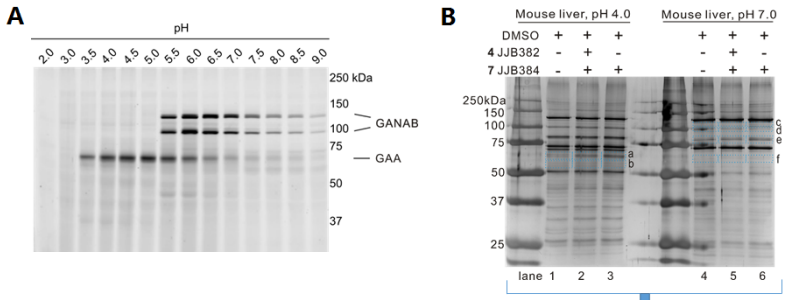


Figure S3. On-bead digest chemical proteomics data of human fibroblast lysate labeling of α-glucosidases. Protein score A) and emPAI value B) of top 40 proteins on- bead digetion analysis using samples of DMSO, ABP **3** competitive inhibition and ABP **6** direct labeling in pH 4.0 and pH 7.0.



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 S4. A) Mouse liver lysate labeling of α-glucosidases in various pH *in vitro* with ABP **3**. B) In-gel digestion silver staining gel and identification of target proteins modified by biotin-ABP **6** with mass spectrometry and mascot search analysis, comparing with DMSO and ABP **3** competitive inhibition samples.

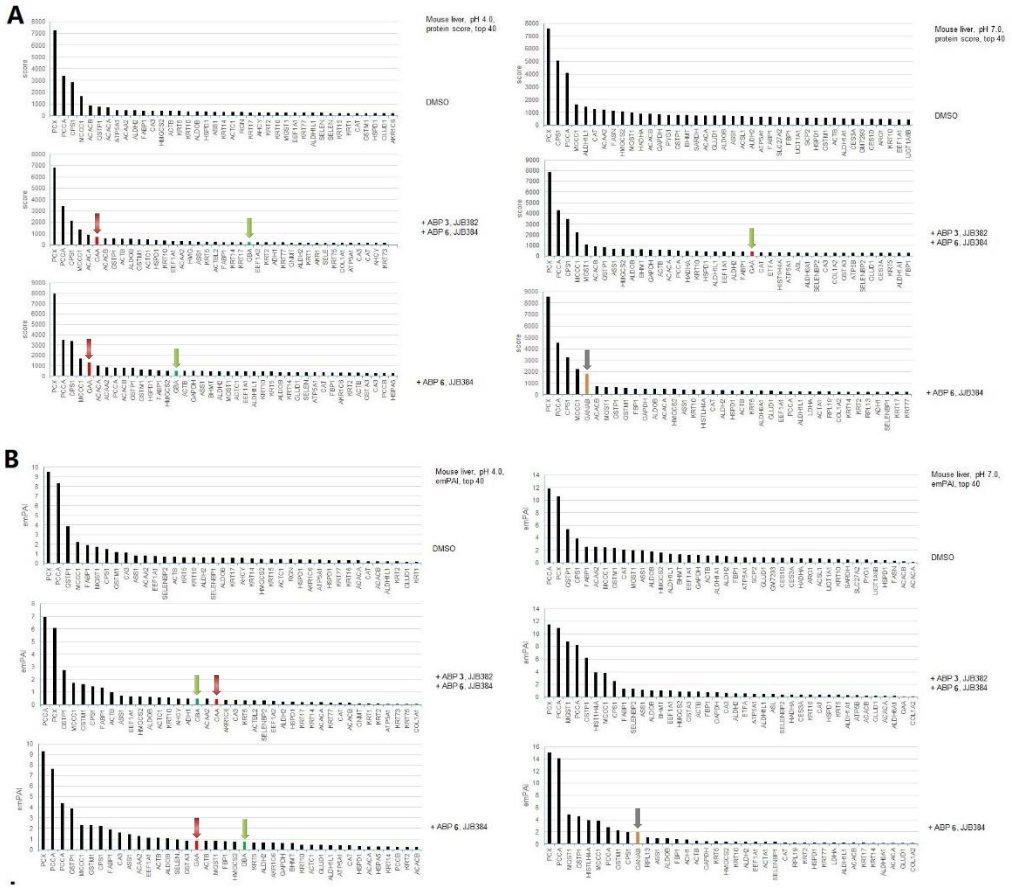
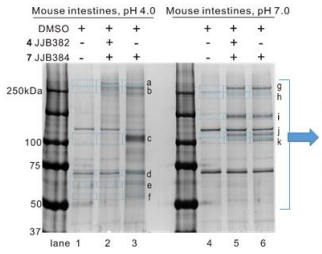


Figure S5. On-bead digest chemical proteomics data of mouse liver lysate labeling of α-glucosidases. Protein score A) and emPAI value B) of top 40 proteins on-bead digestion analysis using samples of DMSO, ABP **3** competitive inhibition and ABP **6** direct labeling in pH 4.0 and pH 7.0.

A



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						

B

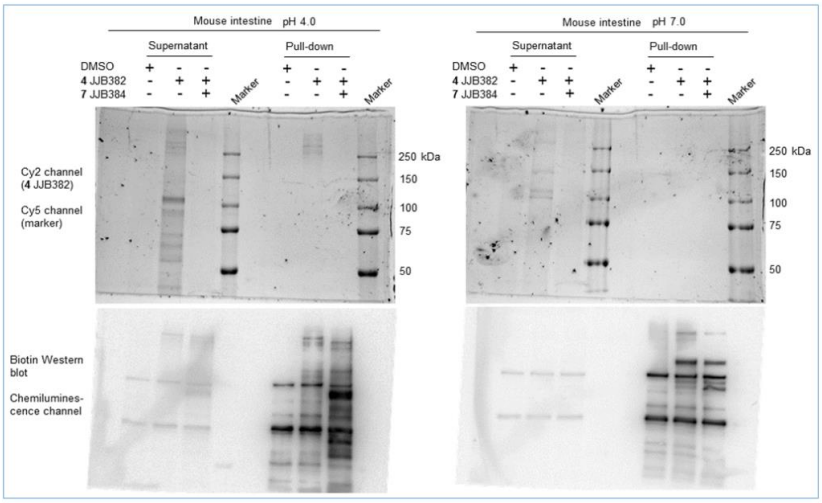


Figure S6. Mouse intestine lysate labeling of α-glucosidases and their identification by proteomics. A) In-gel digestion silver staining gel and identification of target proteins modified by biotin-ABP **7** with mass spectrometry and mascot search analysis, comparing with DMSO and ABP **3** competitive inhibition samples. B) Pull-down and supernatant samples labeling of α-glucosidases in pH 4.0 and pH 7.0 by fluorescent SDS-PAGE scanning for ABP **3** and Western blotting for ABP **6** analysis

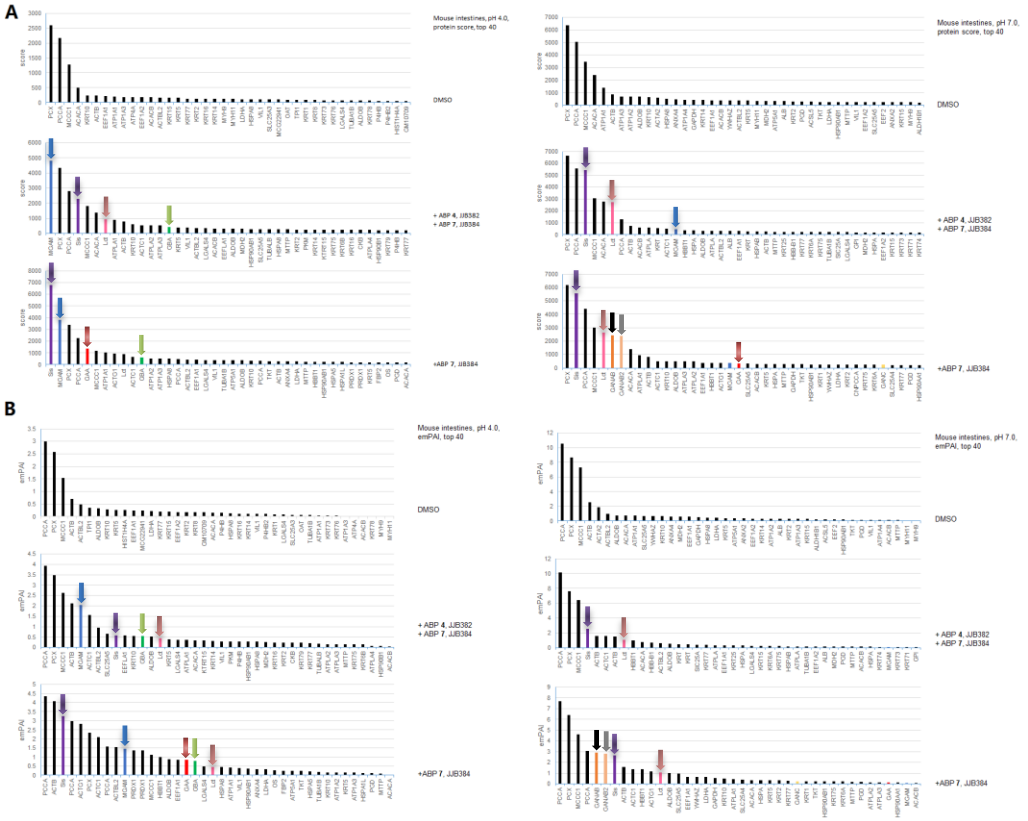


Figure S7. On-bead digest chemical proteomics data of mouse intestine lysate labeling of α-glucosidases. Protein score A) and emPAI value B) of top 40 proteins on-bead digestion analysis using samples of DMSO, ABP **3** competitive inhibition and ABP **6** direct labeling in pH 4.0 and pH 7.0.

6

Synthesis and biological evaluation of 8-carboxy-cyclophellitol aziridine derivatives as β-glucuronidase inhibitors and activity-based probes

Jianbing Jiang, Yi Jin, Wouter W. Kallemeijn, Wei Dai, Cas van Elk, Gijsbert A. van der Marel, Jeroen D. C. Codée, Johannes M. F. G. Aerts, Gideon J. Davies and Herman S. Overkleeft, **2016**, manuscript in preparation.

6.1 Introduction

β-Glucuronidases hydrolyze β-glucuronic acid (GlcUA)-containing carbohydrates to release GlcUA and are found in all living organisms. β-Glucuronidases are classified into two glycoside hydrolase (GH) families in human, namely, GH2 and GH79, in the Carbohydrate-Active enZymes (CAZy) database, according to their amino acid sequences and tertiary structure.[1,2] β-Glucuronidases hydrolyze β-glucuronide substrates through a formal two-step double displacement mechanism, first proposed by Koshland.[3] Human GH2 lysosomal β-glucuronidase cleaves GlcUA residues from the non-reducing termini of glycosaminoglycans incuding dermatan sulphate and heparan sulfate, and its deficiency in man causes the lysosomal storage disorder, mucopolysaccharidosis type VII (MPSVII), also known as Sly syndrome.[4] The GH79 family contains a number of β-glucuronidases from mammalian, fungi and bacterial,[5] such as heparanase that is responsible for hydrolysis of heparin sulfate proteoglycans. β-glucuronidase activities are often associated with inflammation processes and have also been found responsible for the promotion of cancers.[6]

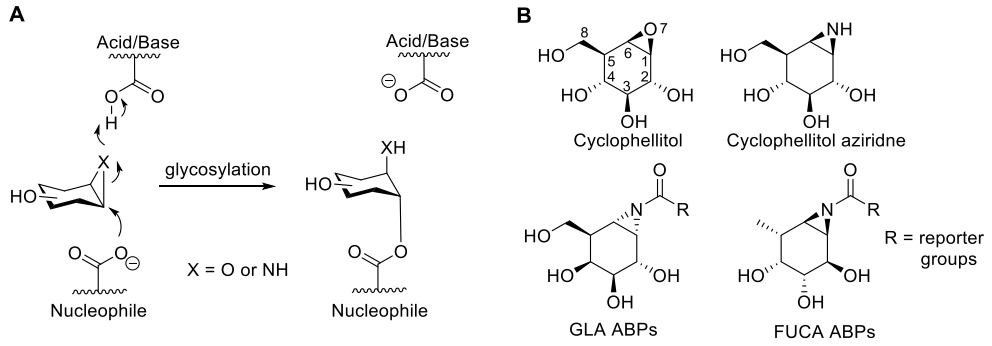


Figure 1. A) Mechanism-based inhibition of retaining glycosidases by cyclophellitol and cyclophellitol aziridine derivatives. B) Structure of the β-glucosidase inhibitors, cyclophellitol and cyclophellitol aziridine as well as activity-based probes for α-galactosidases (GLA) and α-L-fucosidases (FUCA).

The research described in the previous chapters focused on the development of configurational and functional isomers of the natural product, cyclophellitol, as an activity-based retaining *exo*-glycosidase probe scaffold. Cyclophellitol is a potent and selective mechanism-based irreversible inhibitor of retaining β-glucosidases.[7,8] Substitution of the epoxide moiety in cyclophellitol for an aziridine yields cyclophellitol aziridine,[9] which inhibits retaining β-glucosidases in a similar fashion. Structurally, cyclophellitol and its aziridine analog are close to β-glucopyranose in configuration. They fit well in the active site pocket of retaining β-glucosidases, and undergo acid-catalyzed ring-opening following nucleophilic attack on C1 by the enzyme active site nucleophilic residue (Figure 1A) to yield a stable, covalent adduct. Both cyclophellitol and cyclophellitol aziridine have been modified to yield retaining β-glucosidase-selective activity-based probes (ABPs) by grafting reporter moieties on either the cyclophellitol C8 carbon or the cyclophellitol aziridine nitrogen.[10]

The high potency and selectivity of cyclophellitol (aziridine), together with conserved catalytic mechanism of retaining GHs, opens up the potential to exploit selective inhibitors and ABPs for other class of retaining GHs. Indeed, by adapting the configuration of cyclophellitol aziridine to the natural substrate of target GHs, potent mechanism-based inhibitors of α-glucosidases, α/β-galactosidases, and α/β-mannosidases have been prepared in the past.[9,11,12] More recently, cyclophellitol aziridine-based ABPs have been developed for profiling α-galactosidases (GLA) and α-L-fucosidases (FUCA) (Figure 1B).[13-15] These ABPs were created by *N*-acylation of the aziridine nitrogen with 8-azidooctanoic acid, followed by introduction of reporter groups (biotin, fluorophores) through azide-alkyne [2+3] cycloaddition 'click' ligation. In line with these studies, this chapter investigates the potential to develop 8-carboxy-cyclophellitol aziridine inhibitors and probes aimed at inhibiting and labeling β-D-glucuronidases. Specifically, the synthesis and enzyme inhibition properties of compounds

2, **3, 12** and ABPs **4-11** (Figure 2), emulating in configuration and substitution pattern glucuronic acid (**1**) are presented. The chapter ends with describing the elucidation of the crystal structure of selected cyclophellitol aziridine derivatives in complex to the bacterial β-glucuronase from *Acidobacterium capsulatum*, AcaGH79. [5c]

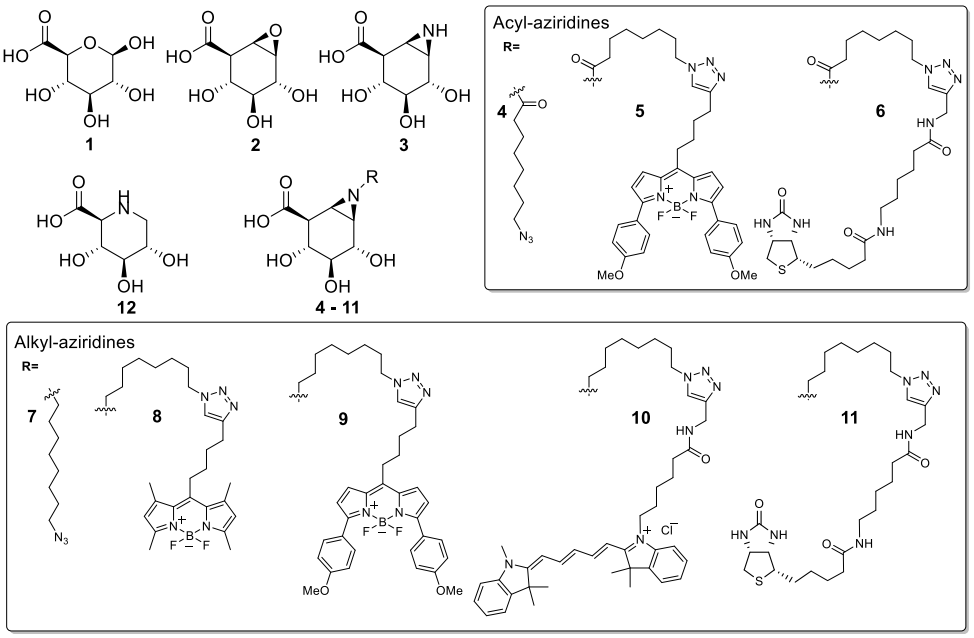


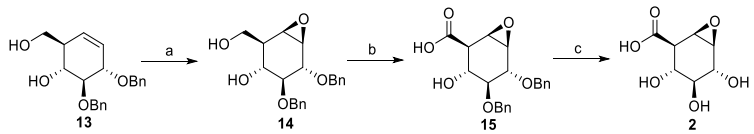
Figure 2. Structures of β-D-glucuronic acid, β-glucuronidase inhibitors and activity-based probes subject of the research described in this chapter.

6.2 Results and Discussion

Synthesis

The synthesis procedure of 8-carboxy-cyclophellitol **1** is depicted in Scheme 1. Cyclohexene diol **13**, prepared as reported before,[16, 17] was treated with *meta*-chloroperbenzoic acid in a biphasic buffer to yield epoxide **14** in a stereoselective fashion as described in the literature.[17] Next, the primary alcohol in **14** was selectively oxidized by 2,2,6,6-tetramethylpiperidine-1-oxyl and (diacetoxyiodo)benzene (TEMPO/BAIB) oxidation yielding carboxylate **15**. Addition of *t*BuOH to homogenize the biphasic (DCM/H_2O) reaction mixture enhanced the rate of hydration of the intermediate aldehyde, thus ensuring efficient oxidation of the lipophilic substrate. Purification of **15** with silica gel column chromatography caused 20-40% epoxide hydrolysis. To circumvent this, **15** was purified by HPLC using 50 mM aq. NH_4HCO_3 as the mobile phase. Hydrogenolysis with hydrogen gas and Pearlman's catalyst followed by HPLC purification yielded 8-carboxy-cyclophellitol **2**, in 4.4% overall yield based on **13**.

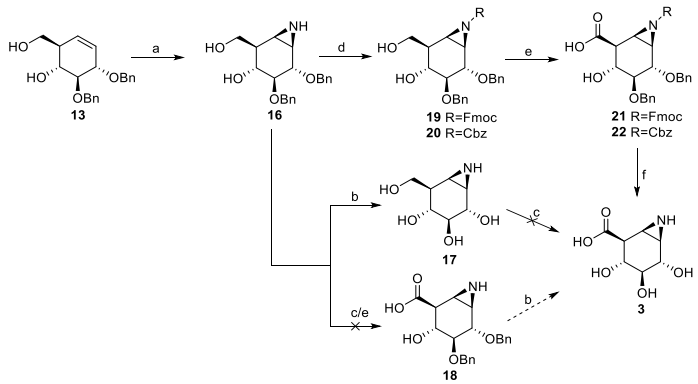
Scheme 1. Synthesis of 8-carboxy-cyclophellitol **2**.



Reagents and conditions: (a) 3-chloroperoxybenzoic acid (*m*CPBA), Na_2HPO_4 (aq., 1.0 M), 1,2-dichloroethane, 50 °C, 55%; (b) BAIB, TEMPO, DCM/*t*BuOH/H_2O (4:4:1, v/v/v), 0 °C, 35%; (c) H_2, $Pd(OH)_2$/C, MeOH, 23%.

The preparation of 8-carboxy-cyclophellitol aziridine **3** started from cyclohexane diol **13** as well (Scheme 2). Partially protected aziridine **16** was prepared following the three-step procedure for the installation described in Chapter 5 for the construction of α-*gluco*-configured cyclophellitol aziridines (first installation of the trichloroacetimidate, next iodo-imination, then hydrolysis of the formed cyclic iminal and final intramolecular iodine displacement to give the aziridine) in 45% yield. Removal of the benzyl groups was accomplished by Birch reduction, as reported previously to yield cyclophellitol aziridine (**17**).[17] In the first instance, oxidation of the primary alcohol in **17** with catalytic amounts of TEMPO and NaClO as co-oxidant at pH 10.5 was attempted, but the desired 8-carboxy-cyclophellitol aziridine (**3**) was not detected in the resulting reaction mixtures. Application of TEMPO/BAIB combination led to the formation of corresponding aldehyde (from **17**) as witnessed by LC-MS analysis. However, further oxidation of this aldehyde to **3** proceeded very sluggishly and only a small amount of the target product was isolated after silica gel chromatography. Using a biphasic solution (DCM/*t*BuOH/H_2O) and varying the concentration of the starting material in solvents also proved ineffective (Table 1). The aziridine moiety appeared to be sensitive to the oxidation conditions used. To avoid this sensitivity the aziridine nitrogen was protected as, either the fluorenylmethyloxycarbonyl (Fmoc) group (**19**) or the carboxybenzyl (Cbz) group (**20**). Subjection of **19** or **20** to the TEMPO/BAIB conditions yielded **21** or **22**, respectively, in good yields (Table 1). Birch reduction on either **21** or **22** led to the efficient and simultaneous removal of all protective groups. Cation exchange chromatography (Amberlite IR120 H^+ resin) on the crude product obtained from subjecting Fmoc aziridine **21** to the Birch conditions did not result in the isolation of pure aziridine **3**, this because of concomitant elution from the resin of Fmoc protective group remnants. However, subjecting Cbz aziridine **22** to the Birch reduction/cation exchange purification sequence gave homogeneous aziridine **3** in 79% yield (21 mg) in the final deprotection/purification sequence (16% yield over the four steps starting from **13**).

Scheme 2. Synthesis of 8-carboxy -cyclophellitol aziridine **3**.



Reagents and conditions: (a) 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%; (b) i) NH_3 (liq.), Li, THF, -60 °C; ii) Amberlite IR120 H^+/NH_4^+, **17**: 65%; (c) NaOCl, NaBr, TEMPO, H_2O; (d) Fmoc-OSu, pyridine, THF, **19**: 75%, or CbzCl, pyridine, THF, **20**: 75%; (e) BAIB, TEMPO, DCM/*t*BuOH/H_2O (4:4:1, v/v/v), 0˚C, **21**: 92% or **22**: 60%; (f) i) NH_3 (liq.), Li, THF, -60˚C; ii) Amberlite IR120 H^+, **3**: 79%;

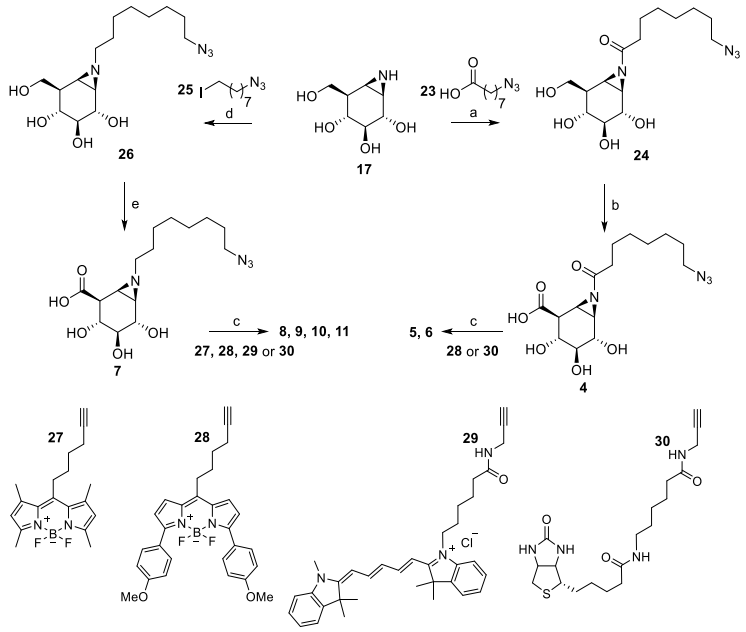
Table 1. TEMPO oxidations of cyclophellitol aziridines **16**, **17**, **19** and **20**

	Regent	Solvent	Temperature	Reaction time	Yield
16	TEMPO/NaClO NaBr, NaOH	H_2O	20 °C	6h	0
16	TEMPO/BAIB	ACN/H_2O (1/1, v/v)	0 °C	24h	0
16	TEMPO/BAIB	DCM/H_2O (2/1, v/v)	0 °C, 20 °C	5h	<1%
16	TEMPO/BAIB	DCM/*t*BuOH/H_2O (4/4/1, v/v/v)	0 °C	22h	<1%
16	TEMPO/BAIB	DCM/*t*BuOH/H_2O (8/8/1, v/v/v)	0 °C	3h	<1%
17	TEMPO/NaClO NaBr, NaOH	H_2O	20 °C	6h	0
19	TEMPO/BAIB	DCM /H_2O (2/1, v/v)	0 °C	5h	48%
19	TEMPO/BAIB	DCM/*t*BuOH/H_2O 4/4/1, v/v/v)	0 °C	5h	92%
20	TEMPO/BAIB	DCM/*t*BuOH/H_2O 4/4/1, v/v/v)	0 °C	5h	60%

With 8-carboxy-cyclophellitol (**2**) and 8-carboxy-cyclophellitol aziridine (**3**) in hand, synthetic routes of *N*-alkylated and *N*-acylated derivatives of **3** were explored. Direct acylation or alkylation of aziridine **3** proved complicated, but oxidation of the primary alcohol in partially protected, *N*-alkyl/acyl aziridine **17** proved more tractable (Scheme 3). Because compound **3** was hard to be obtained and compound **24** and **26** have already been synthesized in previous report.[14] In the first instance, *N*-acyl aziridine **24** was prepared by acylation of the aziridine

nitrogen in **17** with 8-azidooctanoic acid **23** using 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (EEDQ) as the activating agent. Selective oxidation of the primary alcohol in **24** proved to be challenging, as **24** appeared to be highly unstable in both acidic and (strong) basic media. Following the optimized procedure developed for the synthesis of 8-carboxy-cyclophellitol aziridine (**3**), treatment of **24** with TEMPO/BAIB at 0 ˚C under mild basic conditions yielded crude **4**. The crude reaction mixture was lyophilized and purified by HPLC using neutral (milliQ H_2O/ACN) or slightly basic (25mM aq. NH_4OAc) eluents, followed again by lyophilization.

Scheme 3. Synthesis of *N*-substituted 8-carboxy-cyclophellitol aziridine derivatives **4**-**11**.



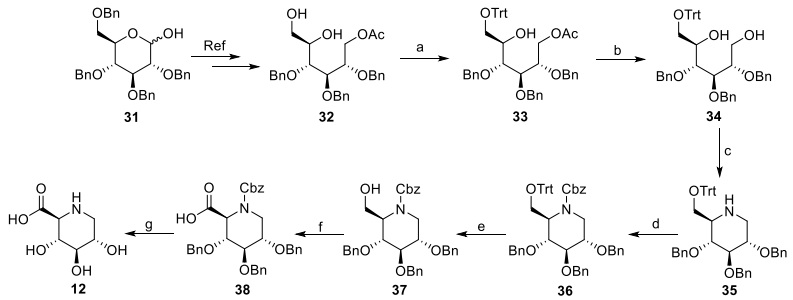
Reagents and conditions: (a) EEDQ, DMF, 57%; (b) TEMPO, BAIB, H_2O/MeCN (1/1), $NaHCO_3$, 0 **°**C, 5%; (c) $CuSO_4$ (1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), DMF, **5**: 5%, **6**: 16% , **8**: 7%, **9**: 8%, **10**: 21%, **11**: 8%; (d) K_2CO_3, DMF, 50%; (e) TEMPO, NaClO, NaBr, NaOH, H_2O, 15%.

Hydrolysis of the aziridine moiety proved to complicate the synthesis (copper (I)-catalyzed [2+3] azide-alkyne click ligation to BODIPY-alkyne **28** and biotin-alkyne **30** respectively) and purification of compounds **5** and **6** as well. Although these compounds could be obtained in crude form (as judged by LC-MS and NMR spectrum), no analytically pure material could be harvested from the crude products (the compounds proved to be active though as can be seen from the biological evaluation below).

The synthesis of *N*-alkylated 8-carboxy-cyclophellitol aziridine derivatives proved to be much easier than the synthesis of their *N*-acyl counterparts. Thus, reaction of **17** with

1-azido-8-iodooctane and K_2CO_3 in DMF afforded compound **26** in 50% yield. Selective oxidation of the primary alcohol in **26** was accomplished using TEMPO/NaClO as the oxidant at pH 10.5 to afford after HPLC purification compound **7** in 15% yield. Finally, ABPs **8-11** were prepared in moderate yields but good purity by Cu(I)-catalyzed azide alkyne click ligation with BODIPY-alkynes **27** and **28**, Cy5-alkyne **29** and biotin-alkyne **30**, respectively, followed by HPLC purification and lyophilization. The final products proved to be stable during lyophilization and HPLC purification conditions (50 mM NH_4CO_3 in H_2O), and even during LC/MS detection in the presence of 1% TFA.

Scheme 4. Synthesis of trihydroxypipecolic acid 3.



Reagents and conditions: (a) DMAP, Et_3N, TrtCl, DCM, 3 h, 75%; (b) Na, MeOH, 3 h, 91%; (c) i: $(COCl)_2$, DMSO, DCM, -65 °C, 2 h; ii: Et_3N, -65 to 5 °C, 2 h; iii: NH_4HCO_3, $NaBH_3CN$, 3 Å molecular sieves, MeOH, 0 °C, 1 h, then 0 °C to rt, 20 h, 85%; (d) CbzCl, THF, Et_3N, rt, 20 h, 72%; (e) *p*-TsOH, DCM, 0 °C to rt, 3 h, 83%; (f) TEMPO, BAIB, DCM/H_2O (2/1, v/v), 0 °C, 5 h, 71%; (g) i) BnBr, Cs_2CO_3, DMF, 3 h, 82%; ii: H_2, $Pd(OH)_2$/C, AcOH/H_2O (4/1, v/v) 16 h, 57%.

(2S,3R,4R,5S)-2,3,4-trihydroxypipecolic acid **12** was synthesized as depicted in Scheme 4. Starting from tetrabenzyl glucose **31**, intermediate diol **32** was prepared in 6 steps in 83% yield following the literature procedure.[18] Protection of the primary hydroxyl group in **32** as the trityl ether afforded **33**. After deacetylation of the C-1 hydroxyl group, Swern oxidation of the two free alcohols followed by double reductive amination gave orthogonally protected deoxynjirimycin **35** following a procedure previously developed for differently protected deoxynojirimycin derivatives. Subsequently, the amine in **35** was protected as the *tert*-butyloxycarbamate, after which the trityl ester was removed under acidic (TsOH) conditions. Subsequently the primary alcohol of **37** was oxidized (TEMPO/BAIB) to carboxylic acid **38** and hydrogenation over palladium(II) hydroxide on carbon in aq. acetic acid gave inhibitor **12**, the analytical and spectroscopical data of which matched those reported.[19]

Comparative and competitive ABPP on GH79 β-glucuronidase AcaGH79

Having the panel of inhibitors and probes **2**-**12** available, their potency (as reflected by apparent IC_{50} values) in inhibiting 4MU-β-glucuronide hydrolysis by recombinant

Acidobacterium capsulatum β-glucuronidase (AcaGH79)[15] was established firstly. All compounds proved to inhibit AcaGH79, although their potency (IC_{50} values) varied considerably (Table 1). Cy5-aziridine **10** proved to be the most potent inhibitor of the set of ABPs **5**, **6**, **8**-**11** (IC_{50} 51.8 ± 7.4 nM) with the two biotin-aziridines **6, 11** being the weakest of the series (though with an IC_{50} of around 9.5 μM and 8.1 μM, respectively, still quite potent). Of the non-tagged covalent inhibitors, acyl aziridine **4** proved to be the most potent inhibitor, slightly less outperforming than the competitive inhibitor **12**. Epoxide **2** and alkyl aziridine **7** turned out to be considerably less potent AcaGH79 inhibitors. Importantly, all ABPs excepting (at the concentrations measured) biotin aziridines **6** and **11** are able to reach subcellular β-glucuronidase at nanomolar concentrations. Additionally, the potency of alkyl aziridine **7**-**11** is roughly equal to the potency of acyl aziridines **4**-**6**. From this observation combined with the observation that alkyl aziridines are much more stable and easily accessible, one may conclude that *N*-alkylated cyclophellitol aziridine derivatives are the reagents of choice in the development of retaining β-glucuronidase activity-based probes. The cyclophellitol and cyclophellitol aziridine derivatives bind AcaGH79 covalently and irreversibly, as evident from their second order kinetic constants (measured for compounds **2**, **3**, **7**, **8**, see Figure 3A). As well, ABP **8** effectively labels AcaGH79 both at 4 °C and 37 °C (Figure 3B). Labeling of AcaGH79 with ABP **8** can be abolished by competition with either of the compounds **2**, **3, 7**, **12**, **9**-**11** or with the fluorogenic substrate, 4MU-β-glucuronide as well as by denaturing AcaGH79 prior to inclusion of the probe (Figure 3C). Competitive labeling was also obtained with acyl aziridine inhibitor **4** and ABP **5** (Figure 3D). Maximal AcaGH79 labeling with ABP **8** was observed at pH 4-5, in line with the observed maximal hydrolysis of fluorogenic substrate by AcaGH79 (Figure 3E, F).

Table 1. Potency (IC_{50}) of compounds **2**-**12** as inhibitors of recombinant AcaGH79 from *Acidobacterium capsulatum in vitro* using 4MU-β-glucuronide as the fluorogenic substrate. Data were average values of two separate experiments measured in duplicate, error ranges depict standard deviation.

Compounds	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
IC_{50}(nM)	4,520 ± 315.9	159.4 ± 19.6	95.2 ± 11.2	111.1 ± 8.4	9,532.5 ± 61.5	21,555 ± 1193.8	435.6 ± 28.9	294.9 ± 19.0	51.8 ± 7.4	8,084.4 ± 553.2	44.5 ± 3.3

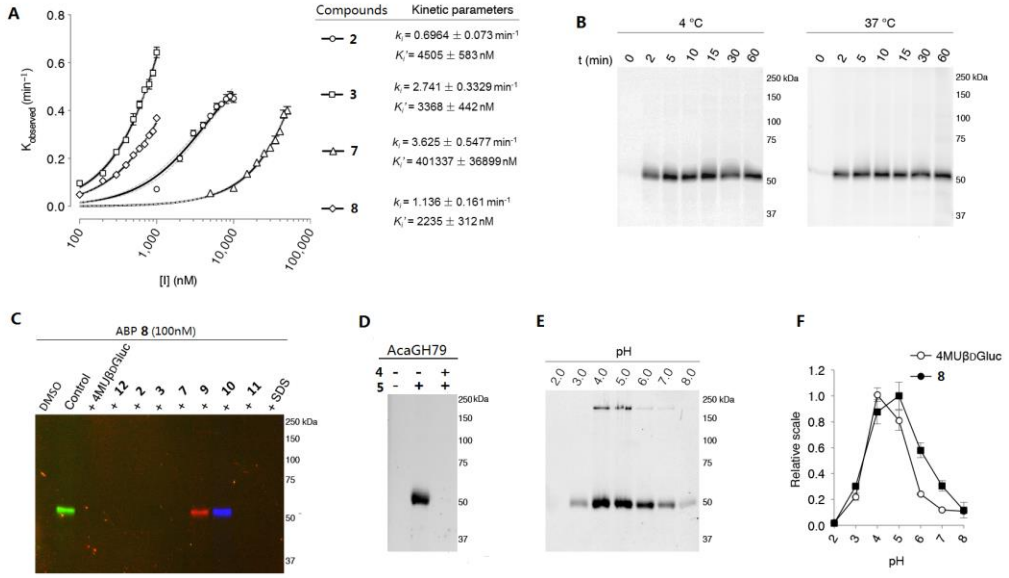


Figure 3 A) Inhibition kinetic parameters for compounds **2**, **3**, **7**, **8**. The graph depicts the obtained $K_{observed}$ and the table shows the kinetic parameters k_i and K_i' estimated with one-phase exponential association. Data shown represents average of two separate experiments in duplicate, and the error bars depict standard deviation. B) Time-dependent labeling of recombinant AcaGH79 with green-BODIPY ABP **8** at 4 °C and 37 °C, respectively. C) Competitive labeling of AcaGH79 with ABP **8** by pre-incubation with 10% SDS, substrate, inhibitors **2**, **3**, **12**, **7** or probes **9**, **10**, **11**. D) Competitive labeling of AcaGH79 with ABP **5** by pre-incubation with acyl aziridine inhibitor **4**. E) Detection of AcaGH79 labeling efficiency at various pH with ABP **8**. F) AcaGH79 labeling with ABP **8** (closed squares) compared to enzymatic activity of substrate 4-methylumbelliferyl β-D-glucuronide (4MUβDGluc) hydrolysis (open circles) at various pHs.

X-ray crystallography of AcaGH79 complexed to inhibitors 3, 4 and 7

In order to provide further evidence of the covalent attachment of the β-glucuronic acid-configured ABPs to β-glucuronidase, crystal structures of AcaGH79 in complex with aziridines **4** and **7** were obtained. These structures clearly showed covalent binding of the compounds, with the aziridine ring opened by the AcaGH79 active site nucleophile, Glu287 (Figure 4A, B). Both acyl aziridine **4** and alkyl aziridine **7** adopt a $^4C_1$ conformation with the aziridine rings opened across the C1-C6 bond in trans-diaxial fashion. The structure of the AcaGH79 mutant in its catalytic nucleophile (E287Q) in complex with **7** revealed (Figure 4C) unhydrolyzed **7** present in the active site and adopting a $^4H_3$ conformation. This conformation is also seen for competitive inhibitor **12** in wild type AcaGH79 (Figure 4D), and is consistent with the expected $^1S_3$-$^4H_3$-$^4C_1$ catalytic itinerary employed by β-glucuronidases.[20]

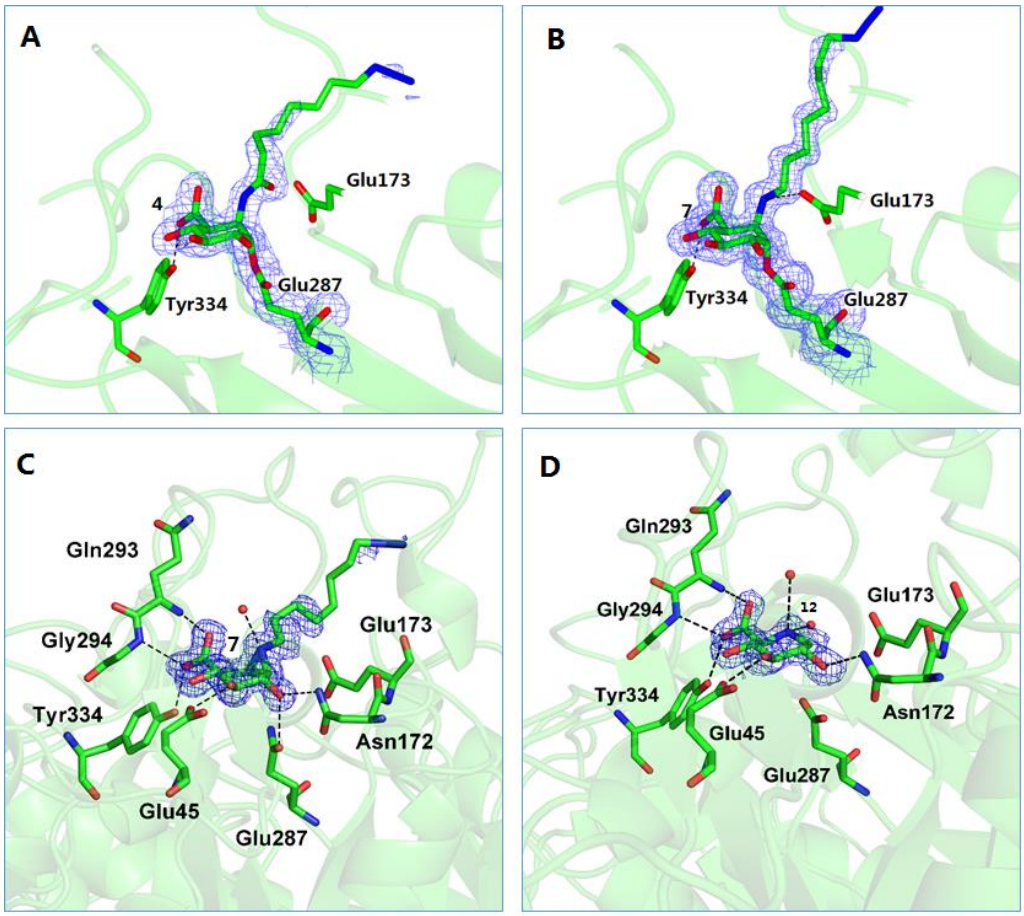


Figure 3. Crystal structures of AcaGH79 from *Acidobacterium capsulatum* in complex with **4**, **7** and **12**. A) AcaGH79-acyl aziridine **4**; B) AcaGH79-alkyl aziridine **7**; C) Mutant E287N AcaGH79--alkyl aziridine **7**; D) AcaGH79-trihydroxypipecolic acid **12**.

6.3 Conclusion

In conclusion, this chapter describes the development of potent, mechanism-based β-glucuronidase inhibitors and activity-based probes. The synthesis of *N*-acyl-8-carboxy-cyclophellitol aziridine proved challenging and no homogeneous pure material could be obtained. The corresponding *N*-alkyl derivatives however were accessible and turned out to be at least as effective in inhibiting and labeling the bacterial β-glucuronidase, AcaGH79. These results match those obtained for *N*-alkyl-cyclophellitol aziridine and *N*-acyl-cyclophellitol aziridine, the reduced counterparts of the molecules described in chapter 4 of this thesis in inhibiting and labeling retaining β-glucosidases. The crystal structure of AcaGH79 complexed to **4** and **7** together with the structure of AcaGH79 complexed to the competitive inhibitor **12** (a structure matched by that of the mutant enzyme complexed with **7**) gives a clear view on the conformational trajectory by which a substrate

glucuronide is processed by retaining β-glucuronidases. The activity-based β-glucuronidase probes described here will be further evaluated, both in activity-based protein profiling and in structural studies, in the next chapter.

6.4 Experimental sections

Gene cloning, expression and protein purification

The coding sequence of the AcGH79 gene was cloned into the pET28a (Novagen) 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 unitl the A600 nm reached 0.8. Expression of the recombinant proteins was induced by the addition of 1 mM isopropyl β−D-1-thiogalactopyranoside for 12 h at 25 °C. The cells were harvested by centrifugation at 8000 g for 30 min and resuspended in 50 mL 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 His_6-AcGH79 was loaded onto a His Trap column (GE Healthcare), equilibrated with the lysis buffer. The column was washed with lysis buffer and the His_6-AcGH79 protein was eluted with the same buffer with supplement of 400 mM imidazole over a gradient of 100 mL. The fractions containing the His_6-AcGH79 were then loaded onto a Hiload 16/60 Superdex 75 column (GE Healthcare). The fractions containing the His_6-AcGH79 were pooled and concentrated to the final concentration of 14.5 mg mL^{-1}.

Enzyme activity assays and IC_{50} measurements for AcaGH79.

The β-D-glucuronidase activity of AcaGH79 was assayed at 37 °C by incubating with 2.5 mM 4-methylumbelliferyl-β-D-glucuronide as substrate in 150 mM McIlvaine buffer, pH 5.0, supplemented with 0.1% (w/v) BSA. To determine the apparent *in vitro* IC_{50} value, recombinant AcaGH79 was firstly pre-incubated with a range of inhibitor dilutions for 30 min at 37 °C, prior to addition of the substrate. The enzymatic reaction was quenched by adding excess NaOH-glycine (pH 10.6), after which fluorescence of liberated 4-methylumbelliferyl was measured with a fluorimeter LS55 (Perkin Elmer) using λ_{EX} 366 nm and λ_{EM} 445 nm. The *in situ* IC_{50} value was determined by incubating fibroblast cell lines expressing wild-type AcaGH79, 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 potassium phosphate buffer (25 mM K_2HPO_4-KH_2PO_4, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail (Roche)). After determination of the protein concentration (BCA kit, Pierce), lysates were aliquoted and frozen at -80 °C. Residual AcaGH79 activity was measured using the aforementioned substrate assay. Data was corrected for background fluorescence, then normalized to the untreated control condition and finally curve-fitted via one phase exponential decay function (GraphPad Prism 5.0).[9]

Determination of inhibition constants

The kinetic parameters of inhibition were determined by adding AcaGH79 to the appropriate McIlvaine buffer, containing both substrate and various concentrations of the corresponding inhibitor. The mixtures were incubated for 0 to 60 min at 37 °C and the reactions were quenched at intervals of 5 min with excess NaOH-glycine (pH 10.6).

Inhibitory constants k_i and K'_i were determined by firstly calculating the $K_{observed}$ per inhibitor concentration, followed by curve-fitting the data to a one phase exponential association function (GraphPad Prism 5.0).[9]

***In vitro* labeling and SDS-PAGE analysis and fluorescence scanning**

All the labeling samples were samples resolved on 10% SDS-PAGE. Electrophoresis in sodium dodecylsulfate containing 10% polyacrylamide gels was performed as earlier described.[21] Wet slab-gels were then scanned for ABP-emitted fluorescence using a Bio-rad ChemiDoc MP imager using green Cy2 (λ_{EX} 470 nm, bandpass 30 nm; λ_{EM} 530 nm, bandpass 28 nm) for **8**, red Cy3 (λ_{EX} 530 nm, bandpass 28 nm; λ_{EM} 605 nm, bandpass 50 nm) for **5**, **9**, and blue Cy5 (λ_{EX} 625 nm, bandpass 30 nm; λ_{EM} 695 nm, bandpass 55) for **10**. All samples were denatured with 5× 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), boiled for 4 min at 100 °C, and separated by gel electrophoresis on 10% (w/v) SDS-PAGE gels running continuously at 90 V for 30 min and 200 V for 50min.

Influence of pH on ABP labeling involved pre-incubation of AcaGH79 at pH 2−8 for 30 min on ice, prior to addition of 100 nM **8**, dissolved in Nanopure H_2O and incubating for 30 min at 37 °C. Assessment of **2**, **3**, **7**, **8** labeling kinetics involved pre-cooling of 10 ng AGH79 on ice for 15 min, followed by addition of similarly cooled 100 nM **8** solution. After mixing, **8** labeling was chased for 0−60 min at either 4 °C or 37 °C, whereafter labeling was stopped by denaturation. For alkyl aziridines competitive labeling on AcaGH79, 100 ng AcaGH79 was pre-incubated with inhibitors (25 mM 4MU-β-D-glucuronide, 100 µM **2**, **3**, 1.0 mM **12**, or 1.0 µM **7**, **9**, **10**, **11**, or boiled for 4 min in 2% (w/v) SDS, prior to labeling with 100 nM **8**, dissolved in Nanopure H_2O, for 30 min at 37 °C. For acyl aziridines competitive labeling, 100 ng wild-type AcaGH79 was pre-incubated with 1 µM inhibitor **4**, prior to labeling with 100 nM **5** in 150 mM McIlvaine buffer, pH 5.0, for 30 min at 37 °C.

X-ray crystallography and structure solution

Initial crystallization screening of native AcGH79 was conducted by sitting drop vapor diffusion method at 20 °C, by mixing 0.5 µL of the protein solution (4.5 mg mL^{-1}) and an equal volume of precipitant against 100 µL of reservoir solution composed of 0.8 M to 1.5 M sodium phosphate monobasic/postassium phosphate dibasic in the ratio of NaH_2PO_4/K_2HPO_4 = 0.5/9.5 (v/v, pH not adjusted). Some preliminary hits were obtained in several conditions and after refinement of the crystallization conditions, well diffracting crystals were obtained by mixing 0.7-0.8 µL 4.5 mg mL^{-1} protein stock with 0.5 µL precipitant composed of 0.8 M to 1.2 M NaH_2PO_4/K_2HPO_4 = 0.5/9.5 (v/v, pH not adjusted) overnight at 20 °C, using the sitting drop vapour diffusion method. The E287Q mutant of AcGH79 was crystallized by mixing 0.3-0.5 µL of the protein solution (12 mg mL^{-1}) with 0.50 µL precipitant composed of 1.2 M to 1.5 M NaH_2PO_4:K_2HPO_4 = 1.0 : 9.0 (v/v, pH not adjusted).

All four complexes, namely wt-**12**, wt-**4**, wt-**7** and E287Q-**7** were produced by soaking method. The drops containing the crystals of the corresponding wt AcGH79 or E287Q were supplemented with 1.0 µL of 5.0 mM **12**, **4** or **7** in the precipitant solution freshly prepared before soaking procedure for 1 h at 20 °C before fishing. 2 M lithium sulphate was used as cryoprotectant for all the crystals before being flash frozen in liquid nitrogen. Diffraction data for all the crystals were collected at 100 K on beamline i04-1 the Diamond Light Source and were processed using the *xia2*

implementation of XDS and programs from CCP4 suite.[22] Parts of the data processing statistics and structure refinement are listed in Table 1.

Table 1. The statistics of the data processing and structure refinement.

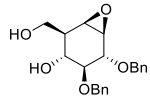
	Native	wt-**4**	wt-**7**	wt-**12**	E287Q-**7**
Data collection					
Space group	C2	C2	C2	C2	C2
Cell dimensions					
a, *b*, *c* (Å)	150.7, 44.9, 83.4	150.7, 44.8, 83.3	148.5, 44.9, 82.2	150.6, 44.7, 83.2	150.2, 45.1, 82.0
α, β, γ (°)	90°, 115.6°, 90°	90°, 115.5°, 90°	90°, 114.8°, 90°	90°, 115.5°, 90°	90°, 115.1°, 90°
Resolution (Å)	42.1 (1.19)	66.7 (1.80)	42.6 (1.60)	42.4 (1.60)	28.9 (1.24)
R_{sym} or R_{merge}	0.033 (0.443)	0.068 (0.540)	0.058 (0.664)	0.054 (0.594)	0.045 (0.489)
$I/\sigma I$	16.9 (2.2)	13.6 (2.1)	12.1 (1.8)	14.2 (1.9)	13.7 (1.9)
Completeness (%)	92.0 (55.4)	98.0 (98.0)	99.5 (99.7)	95.8 (96.3)	89.8 (49.7)
Redundancy	3.8 (2.7)	4.1 (4.1)	4.0 (3.7)	4.2 (4.3)	3.9 (2.7)
Refinement					
Resolution (Å)	1.22 (1.19)	1.85 (1.80)	1.63 (1.60)	1.63 (1.60)	1.27 (1.24)
No. reflections	147896 (6537)	45856 (3332)	64976 (3170)	63367 (3085)	126454 (5143)
R_{work}/ R_{free}	0.1595	0.1967	0.1675	0.1631	0.1699

Highest resolution shell is shown in parenthesis.

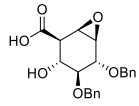
General synthesis

All reagents were of a commercial grade and were used as received unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and *N*, *N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an argon atmosphere unless stated otherwise. Solvents used for flash column chromatography were of pro analysis quality. Reactions were monitored by TLC analysis using Merck aluminium sheets pre-coated with silica gel 60 with detection by UV absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}.H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4.H_2O$ (10 g/L) in 10% sulfuric acid followed by charring at ~150 ˚C or by spraying with an aq. solution of $KMnO_4$ (7%) and K_2CO_3 (2%) followed by charring at ~150 °C. Column chromatography was performed using either Baker or Screening Device silica gel 60 (0.04 - 0.063 mm) in the indicated solvents. 1H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-850 (850/214 MHz), Bruker DMX-600 (600/150

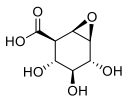
MHz) and Bruker AV-400 (400/100 MHz) spectrometer in the given solvent. Chemical shifts are given in ppm relative to the deuterated chloroform or methanol residual solvent peak or tetramethylsilane (TMS) as internal standard. Coupling constants are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. High-resolution mass spectra were recorded with a LTQ Orbitrap (Thermo Finnigan). Optical rotations were measured on Propol automatic polarimeter (Sodium D-line, λ = 589 nm). IR spectra were recorded on a Shimadzu FT-IR 83000 spectrometer. LC-MS analysis was performed on an LCQ Adventage Max (Thermo Finnigan) ion-trap spectrometer (ESI$^+$) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C_{18} column (Gemini, 4.6 mm x 50 mm, 3 µm particle size, Phenomenex) equipped with buffers A: H_2O, B: MeCN (MeCN) and C: 1% aq. TFA or 50 mM NH_4HCO_3 in H_2O, For reversed-phase HPLC purifications an Agilent Technologies 1200 series instrument equipped with a semi preparative Gemini C_{18} column (10 x 250 mm) was used. The applied buffers were A: 25mM NH_4OAc or 50 mM NH_4HCO_3 in H_2O, B: MeCN.



(1R,2R,3R,4S,5R,6R)-4,5-bis(benzyloxy)-2-(hydroxymethyl)-7-oxabicyclo[4.1.0]heptan-3-ol (14): A mixture of diol **13**[15] (3.4 g, 10 mmol, 1.0 eq.) and *m*CPBA (4.4 g, 18 mmol, 1.8 eq.) in 1, 2-dichloroethane (166 mL), NaH_2PO_4 (1.0 M in H_2O, 100 mL), and Na_2HPO_4 (1.0 M in H_2O, 100 mL) was vigorously stirred at 50 °C for 18 h. The layers were separated, and the water layer was extracted with EtOAc. The combined organic layers were dried with $MgSO_4$ and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (88%→92%, Et_2O in pentane) to provide β-epoxide **14** (1.9 g, 5.4 mmol, 55%). ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.40 – 7.28 (m, 10H), 4.97 (d, *J* = 11.6 Hz, 2H), 4.83 (d, *J* = 11.2 Hz, 2H), 4.70 – 4.66 (m, 2H), 4.03 (dd, *J* = 6.4, 10.8Hz, 1H), 3.91 (dd, *J* = 5.6, 10.8Hz, 1H), 3.83(d, *J* = 7.6Hz, 1H), 3.51 – 3.46 (m, 1H), 3.42 – 3.38 (m, 1H), 3.29 – 3.28 (m, 1H), 3.17 (d, *J* = 4.0Hz, 1H), 3.01 – 2.77 (br, 1H), 2.19 – 2.13 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 138.27, 137.46, 128.70, 128.64, 128.51, 128.18, 128.09, 128.01, 127.98, 83.60, 79.45, 74.94, 72.69, 68.64, 63.99, 55.01, 53.07, 43.44; HRMS: calculated for $C_{21}H_{25}O_5$ [M+H$^+$] 357.16965, found 357.16987.

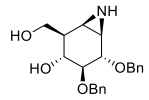


(1R,2R,3R,4S,5R,6R)-4,5-bis(benzyloxy)-3-hydroxy-7-oxabicyclo[4.1.0]heptane-2-carboxylic acid (15): 2, 3-Di-O-benzylcyclophellitol **14** (33 mg, 0.10 mmol, 1.0 eq.) was dissolved in 650 µL DCM/*t*BuOH/H_2O (4/4/1, v/v/v) mixture and was cooled down to 0˚C. TEMPO (75 mg, 19 µmol, 0.20 eq.) and BAIB (75 mg, 0.27 mmol, 2.5 eq.) were added to the solution and it was stirred at 0˚C. The reaction was monitored by TLC. The reaction was quenched after 3 h with 1.0 mL sat. NaS_2O_3 solution and extracted with EtOAc. The organic layer with crude product was dried with $MgSO_4$ and concentrated *in vacuo*. Crude product was purified via HPLC (linear gradient: 28%→38% B, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and concentrated *in vacuo* resulting in **15** (12 mg, 0,032 mmol, 35%). ^{1}H-NMR (400 MHz, CD_3OD): *δ* ppm 7.36 – 7.23 (m, 10H), 4.87 – 4.73 (m, 3H), 4.66 (d, *J* = 11.5 Hz, 1H), 3.81 – 3.74 (m, 2H), 3.54 – 3.53 (m, 1H), 3.41 – 3.36 (m, 1H), 3.19 (d, *J* = 3.7 Hz, 1H), 2.68 – 2.65 (M, 1H); ^{13}C-NMR (100 MHz, CD_3OD): *δ* ppm 140.36, 139.42, 129.38, 129.15, 129.13, 129.02, 128.77, 128.41, 85.29, 80.46, 75.79, 73.62, 69.24, 56.52, 54.39, 51.36; HRMS: calculated for $C_{21}H_{22}O_6$ [M+Na$^+$] 393.13086, found 393.13077.

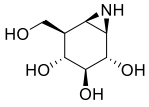


(1R,2R,3R,4S,5R,6S)-3,4,5-trihydroxy-7-oxabicyclo[4.1.0]heptane-2-carboxylic acid (2): A mixture of Product **15** (12 mg, 0.032 mmol, 1.0 eq) and $Pd(OH)_2$/C (3.0 mg, 20% wt. loading(dry basis)) in MeOH

(1.5 mL) was stirred at room temperature under hydrogen atmosphere for 24 h. The catalyst was then filtered off through celite and washed with MeOH. The filtrate was concentrated under reduced pressure and the crude product was purified by semi preparative reversed phase HPLC (linear gradient: 0%→20% B, 3 CV, solutions sed A: 50mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilization gave the title compound **2** (1.4 mg, 7.4 µmol, 23 %) as white solid. ^{1}H-NMR (400 MHz, D_2O): δ ppm 3.76 (d, J = 8.5 Hz, 1H), 3.61 – 3.56 (m, 1H), 3.51 - 3.50 (m, 1H), 3.33 – 3.28 (m, 1H), 3.17 (d, J = 3.8 Hz, 1H), 2.71 – 2.68 (m, 1H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 75.59, 70.74, 67.65, 56.13, 55.79, 50.77; HRMS: calculated for $C_7H_{10}O_6$ [M+Na$^+$] 213.03696, found 213.03260.

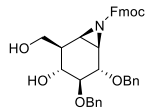


(1R,2R,3R,4S,5S,6R)-4,5-bis(benzyloxy)-2-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-3-ol (16): Building block product **13** (1.4 g, 4.11 mmol, 1 eq.) was co-evaporated thrice in toluene and was subsequently dissolved in dry DCM (100 mL). The solution was cooled down to 0 ˚C. DBU (62 µL, 0.42 mmol, 0.1 eq.) and trichloroacetonitrile (0.495 ml, 4.94 mmol, 1.2 eq.) were added to the solution. The reaction was stirred at 0 ˚C and monitored with TLC. After 4 h, extra DBU (0.1 eq.) and trichloroacetonitrile (1.2 eq.) were added to the solution. After 5 h, TLC showed that the starting material was completely consumed. Subsequently, H_2O (13 mL), $NaHCO_3$ (3.6 g, 43 mmol, 10 eq.) and iodine (3.2 g, 13 mmol, 3.1 eq.) were added to the solution. And the reaction mixture was stirred overnight at room temperature. Afterwards, it was quenched with 10% aq. NaS_2O_3, concentrated *in vacuo* and extracted with EtOAc. The organic layer with the intermediate product was dried over $MgSO_4$ and concentrated *in vacuo* again. In the following step the intermediate was dissolved in dioxane (37 mL) and 37% HCl (aq., 12 mL). The reaction mixture was refluxed at 60 ˚C for 4 h. After which it was concentrated *in vacuo* and redissolved in 100 mL MeOH. $NaHCO_3$ (14 g, 165 mmol, 40 eq.) was added until pH = 7. The reaction mixture was stirred at room temperature for 3 days. The reaction mixture was concentrated *in vacuo* and extracted with EtOAc. The organic layer was dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified via silica gel column chromatography (4%→10%, MeOH in DCM) yielding aziridine **16** (0.66 g, 1.8 mmol, 45%) as white solid. TLC: R_f 0.5 (EtOAc/EtOH, 3/2, v/v); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.39 – 7.27 (m, 10H), 4.98 (d, J = 9.7 Hz, 1H), 4.78 (d, J = 9.3 Hz, 1H), 4.67 (dd, J = 11.4, 4.5 Hz, 2H), 3.99 (m, 2H), 3.75 (d, J = 9.4 Hz, 1H), 3.53 (t, J = 9.9 Hz, 1H), 3.39 – 3.34 (dd, J = 10.0, 8.1 Hz, 1H), 2.43 – 2.41 (m, 1H), 2.27 (d, J = 6.1 Hz, 1H), 2.08 – 2.02 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 138.54, 137.92, 128.66, 128.61, 128.05, 128.03, 128.01, 127.92, 84.38, 81.23, 74.89, 72.37, 68.36, 64.49, 42.57, 33.22, 31.72; HRMS: calculated for $C_{21}H_{26}NO_4$ [M+H$^+$] 356.1852, found 356.1856.

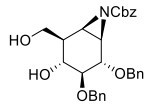


(1R,2S,3S,4R,5R,6R)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (17): Ammonia (8 mL) was condensed at -60 ˚C and lithium (75 mg, 11 mmol, 24 eq.) was added and stirred until all the lithium was dissolved. Then a solution of aziridine **16** (160mg, 0.45 mmol, 1 eq.) in dry THF (10 mL) was added to the solution above. The reaction mixture was stirred at -60 ˚C. The reaction was quenched with milliQ-H_2O (3.0 mL) after 1 h. The solution was allowed to come to room temperature and stirred until all ammonia had evaporated. Next, the solution was concentrated *in vacuo*, redissolved in milliQ-H_2O, and neutralized with Amberlite IR-120 H$^+$. Product bound to the resin was washed with water and subsequently eluted with NH_4OH solution (1.0 M) and evaporated under reduced pressure. The resulting solid was purified by Amberlite IR-120 NH_4^+ resin using milliQ-H_2O as eluens until the eluate was neutral. Concentration of the combined eluate under reduced pressure gave the fully deprotected aziridine **17** (51 mg, 0.30 mmol, 65%). ^{1}H-NMR (400 MHz, D_2O): δ ppm 3.81 – 3.74 (m, 1H), 3.54 –

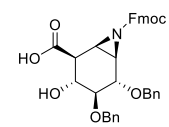
3.44 (m, 2H), 3.11 (t, *J* = 10.1 Hz, 1H), 2.88 (t, *J* = 10.1 Hz, 1H), 2.52 – 2.45 (m, 1H), 2.21 (d, *J* = 6.3 Hz, 1H), 2.01 – 1.95 (m, 1H); ^{13}C-NMR (100 MHz, D_2O): *δ* ppm 76.70, 71.67, 67.38, 61.47, 42.83, 34.78, 33.05.



(9H-fluoren-9-yl)methyl (1R,2S,3S,4R,5R,6R)-2,3-bis(benzyloxy)-4-hydroxy-5-(hydroxymethyl) -7-azabicyclo[4.1.0]heptane-7-carboxylate (19): Aziridine **16** (220 mg, 0.61 mmol, 1.0 eq.) was dissolved in THF (6.2 mL). Fmoc-OSU (160 mg, 0.62 mmol, 1.0 eq.) and pyridine (55 µL, 0.68 mmol, 1.1 eq.) were added. The reaction was stirred at room temperature for 3 h, after which TLC showed the completion of the reaction. The reaction mixture was concentrated *in vacuo*, redissolved in EtOAc. It was washed with water and brine. The organic layer was dried with $MgSO_4$ and concentrated *in vacuo* again. The crude product was purified via silica gel column chromatography (40%→60%, EtOAc in pentane) yielding aziridine **19** (269 mg, 0.47 mmol, 75%) a white solid. TLC: R_f 0.25 (EtOAc/Pentane, 1/1, v/v); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.72 (t, *J* = 7.2 Hz, 2H), 7.58 – 7.53 (m, 2H), 7.43 – 7.19 (m, 14H), 4.87 (d, *J* = 11.3 Hz, 1H), 4.70 – 4.51 (m, 4H), 4.44 (d, *J* = 11.5 Hz, 1H), 4.19 (t, *J* = 5.3 Hz, 1H), 3.82 (dd, *J* = 11.1, 5.9 Hz, 1H), 3.75 – 3.61 (m, 2H), 3.35 – 3.19 (m, 2H), 2.49 – 2.41 (m, 2H), 2.00 – 1.90 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 162.64, 143.42, 143.26, 141.50, 141.44, 138.22, 137.31, 128.60, 128.56, 128.20, 128.08, 128.06, 128.02, 127.94, 127.32, 127.27, 124.78, 124.75, 120.13, 83.80, 79.45, 74.86, 71.87, 68.41, 67.44, 63.88, 47.28, 42.33, 40.28, 38.55, 29.79; HRMS: calculated for $C_{36}H_{35}NO_6$ [M+H$^+$] 578.25371, found 578.25321.

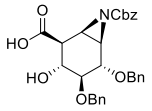


Benzyl(1R,2S,3S,4R,5R,6R)-2,3-bis(benzyloxy)-4-hydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0] heptane-7-carboxylate (20): Aziridine **16** (82 mg, 0.23 mmol, 1 eq.) was dissolved in THF (2.3 mL). CbzCl (39 mg, 0.23 mmol, 1.0 eq.) and pyridine (22.3 µL, 0.28 mmol, 1.2 eq.) were added to the solution. The reaction was stirred at room temperature for 1.5 h and was monitored by TLC. It was shown that all the starting material was converted to a higher running product. The reaction mixture was extracted with EtOAc and concentrated *in vacuo*. The crude product was purified via silica gel column chromatography (1%→9% MeOH in DCM) yielding aziridine **20** (85 mg, 0.17 mmol, 75%). TLC: R_f 0.6(MeOH/DCM, 1/19, v/v); $[\alpha]_D^{20}$ +49.2 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): *δ* ppm 7.40 – 7.26 (m, 15H), 5.14 – 5.13 (m, 2H), 4.96 (d, *J* = 11.3 Hz, 1H), 4.83 (d, *J* = 11.5 Hz, 1H), 4.66 – 4.61 (m, 2H), 4.02 – 3.98 (m, 1H), 3.90 (br, 1H), 3.81 (d, 1H), 3.45 (t, *J* = 9.7 Hz, 1H), 3.35 (m, 1H), 2.86 – 2.79 (m, 2H), 2.12 – 2.07 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): *δ* ppm 162.55, 138.26, 137.31, 135.45, 128.82, 128.67, 128.63, 128.22, 128.13, 128.00, 127.97, 83.84, 79.53, 74.93, 72.07, 68.66, 68.64, 64.20, 42.40, 40.35, 38.83; HRMS: calculated for $C_{29}H_{31}NO_6$ [M+H$^+$] 490,22241, found 490,22215.

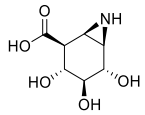


(1R,2S,3R,4S,5S,6R)-7-(((9H-fluoren-9-yl)methoxy)carbonyl)-4,5-bis(benzyloxy)-3-hydroxy-7-az abicyclo[4.1.0]heptane-2-carboxylic acid (21): Fmoc-protected aziridine **19** (172 mg, 0.30 mmol, 1 eq.) was dissolved in 1.5 mL DCM/*t*BuOH/H_2O (v/v/v, 4:4:1) mixture and cooled down to 0 ˚C. TEMPO (9.3 mg, 60 µmol, 0.20 eq.) and BAIB (240 mg, 0.74 mmol, 2.5 eq.) were added to the solution and it was stirred at 0 ˚C. The reaction was monitored by TLC. Extra TEMPO (0.10 eq.) and BAIB (1.2 eq.) were added after 3.5 h. The reaction was quenched with sat. aq. NaS_2O_3 solution (5.0 mL), extracted with EtOAc. The organic layer was dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography (40%→90% EtOAc with 0.15% AcOH in Pentane) yielding aziridine **21** (162 mg, 0.27 mmol, 92%) as white solid. TLC: R_f 0.45 (EtOH/EtOAc/AcOH, 0.98/9/0.02, v/v); $[\alpha]_D^{20}$ +18 (*c* = 0.5, $CHCl_3$); IR (neat, cm^{-1}): 2922, 2852,

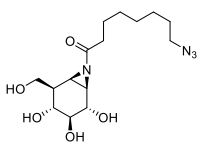
1724, 1450, 1433, 1381, 1278, 1251, 1215, 1111, 1074, 1055, 1082, 987, 756, 698; ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.65 (dd, *J* = 7.2, 3.2 Hz, 2H), 7.55 – 7.49 (m, 2H), 7.41 – 7.24 (m, 14H), 4.82 (d, *J* = 11.3 Hz, 1H), 4.64 (d, *J* = 10.9 Hz, 2H), 4.52 – 4.40 (m, 2H), 4.33 (dd, *J* = 10.7, 6.2 Hz, 1H), 4.13 (t, *J* = 6.1 Hz, 1H), 3.81 (t, *J* = 10.0 Hz, 1H), 3.74 (d, *J* = 8.1 Hz, 1H), 3.30 (dd, *J* = 9.8, 8.4 Hz, 1H), 2.96 (dd, *J* = 5.7, 3.0 Hz, 1H), 2.80 (dd, *J* = 9.9, 2.8 Hz, 1H), 2.52 (d, *J* = 6.0 Hz, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 174.24, 162.21, 143.63, 143.39, 141.38, 141.35, 138.23, 137.28, 128.62, 128.54, 128.24, 128.12, 127.99, 127.92, 127.32, 127.28, 125.17, 125.02, 120.09, 120.04, 83.00, 79.12, 74.98, 72.22, 68.16, 66.93, 47.22, 47.00, 39.40, 39.13; HRMS: calculated for $C_{36}H_{33}NO_7$ [M+H$^+$], 592.23298; found 592.23273.



(1R,2S,3R,4S,5S,6R)-4,5-bis(benzyloxy)-7-((benzyloxy)carbonyl)-3-hydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (22): Aziridine **20** (85 mg, 0.17 mmol, 1.0 eq.) was dissolved in DCM/*t*BuOH/H_2O mixture (1.2 mL , 4:4:1, v/v/v) and cooled down to 0 ˚C. TEMPO (5.3 mg, 34 µmol, 0.20 eq.) and BAIB (137 mg, 42 mmol, 2.5 eq.) were added to the solution and it was stirred at 0˚C for 3 h. The reaction was monitored with TLC. The reaction was quenched by sat. aq. NaS_2O_3 solution (3.0 mL), extracted with EtOAc. The organic layer was dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography (0.5%→10% MeOH in DCM) yielding aziridine **22** (50 mg, 0.10 mmol, 60%). TLC: R_f 0.3 (MeOH/DCM, 1/19, v/v); $[\alpha]_D^{20}$ +56 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.46 – 6.26 (m, 15H), 5.11 – 5.02 (m, 2H), 4.89 (d, *J* = 10.7 Hz, 1H), 4.79 – 4.62 (m, 3H), 3.86 – 3.80 (m, 2H), 3.42 – 3.34 (m, 1H), 3.15 (br, 1H), 2.87 – 2.80 (m, 2H; ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 162.24, 138.30, 137.30, 135.55, 128.70, 128.63, 128.54, 128.42, 128.19, 128.10, 127.98, 127.94, 127.89, 83.13, 79.11, 74.99, 72.32, 68.55, 67.05, 39.58, 39.40; HRMS: calculated for $C_{29}H_{29}NO_7$ [M+H$^+$] 504.20168, found 504.20148.

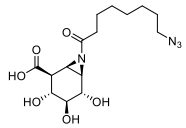


(1R,2S,3R,4S,5S,6R)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (3): Ammonia 10 mL was condensed at -60 ˚C and lithium (23 mg, 3.4 mmol, 24 eq.) was dissolved in the liquid. After 30 min, aziridine **22** (50 mg, 0.14 mmol, 1.0 eq.) was dissolved in dry THF (6.0 mL) and added to the solution. The reaction mixture was stirred at -60 ˚C for 0.5 h, then 15 mg extra lithium was added to the solution for stirring one more hour. The reaction was quenched with 4.0 mL milliQ-H_2O. The solution was allowed to come to room temperature and stirred until all ammonia had extracted. Next, the solution was concentrated *in vacuo*, redissolved in milliQ-H_2O, and neutralized with Amberlite IR-120 H$^+$. Product bound to the resin was washed with water and subsequently eluted with NH_4OH solution (1.0 M) and evaporated under reduced pressure. Concentration of the combined eluate fractions under reduced pressure gave the fully deprotected aziridine **3** (21 mg, 0.11 mmol, 79%). ^{1}H-NMR (400 MHz, D_2O): δ ppm 3.61 (d, *J* = 8.5 Hz, 1H), 3.51 – 3.46 (m, 1H), 3.23 – 3.19 (m, 1H), 2.62 – 2.50 (m, 2H), 2.20 – 2.18 (m, 1H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 179.42, 76.65, 72.21, 67.95, 49.86, 34.26, 33.41; HRMS: calculated for $C_7H_{11}NO_5$ [M+H$^+$] 190.07100, found 190.07105.

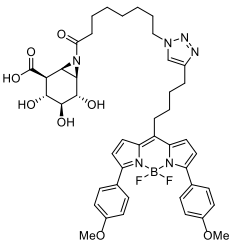


8-Azido-1-((1R,2S,3S,4R,5R,6R)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)octan-1-one (24): 8-azidooctanoic acid **23** (93 mg, 0.50 mmol, 1.2 eq.) and EEDQ (123 mg, 0.50 mmol, 1.2 eq.) were dissolved in DMF (0.5 mL) and stirred at room temperature for 2 h. Pre-activated mixed anhydride solution (300 µL, 0.60 eq.) was added to the deprotected aziridine **17** solution at 0 °C and stirred for 30 min. Additional pre-activated

solution (200 µL, 0.40 eq) was added. The resulting mixture was stirred at 0 °C for 3 h. The reaction was quenched by MeOH (2.0 mL) and the mixture was concentrated *in vacuo*. Then the crude product was purified by HPLC (linear gradient: 25%→31% B in A, 3 CV, solutions used: A: H_2O, B: MeCN). After lyophilization, pure product **24** was obtained as colorless oil (27 mg, 0.079 mmol, 18.9%). ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.05 (dd, J = 10.4, 4.4 Hz, 1H), 3.69 - 3.65 (m, 2H), 3.30 - 3.25 (m, 2H, and MeOD solvent signals), 3.19 (dd, J = 10.0, 8.4,1H), 3.06 (t, J = 9.6 Hz, 1H), 3.01 (dd, J = 6.0, 2.8Hz, 1H), 2.72 (d, J = 6.0 Hz, 1H), 2.48 (t, J = 7.2Hz, 2H), 1.99 - 1.93 (m, 1H), 1.63–1.54 (m, 4H), 1.36 (s, 6H); ^{13}C-NMR (101MHz, CD_3OD): δ ppm 188.54, 79.02, 73.37, 69.34, 63.69, 52.38, 45.2, 42.40, 41.00, 36.83, 30.09, 29.80 27.61 25.93; IR (neat, cm^{-1}): 3341.9, 2934.6, 2860.2, 2363.2, 2343.2, 2095.8, 1671.7, 1559.8, 1418.4, 1371.6, 1278.3, 1180.9, 1088.7, 1027.3, 976.6, 668.2; HRMS: calculated for $C_{15}H_{27}N_4O_5$ [M+H$^+$] 343.19760, found 343.19761.

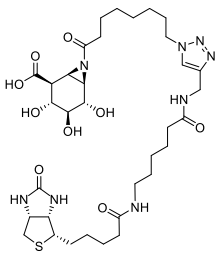


(1R,2S,3R,4S,5S,6R)-7-(8-azidooctanoyl)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (4): The primary alcohol compound **24** (89 mg, 0.26 mmol, 1.0 eq.) was dissolved in MeCN/H_2O (3.0 mL, 1/1, v/v). The solution was cooled to 0 °C, then BAIB (183 mg, 0.57 mmol, 2.2 eq), the solution of TEMPO (8.1 mg, 0.052 mmol, 0.20 eq) in MeCN/H_2O (200 µL, 1/1, v/v) and $NaHCO_3$ (88 mg, 1.0 mmol, 4.0 eq.) was added, the resulting mixture was stirred for 2 h at 0 °C. More addition of BAIB (183 mg, 0.57 mmol, 2.2 eq), the solution of TEMPO (8.1 mg, 0.052 mmol, 0.20 eq.) in MeCN/H_2O (100 µL, 1/1, v/v) and $NaHCO_3$ (50 mg, 0.60 mmol, 2.4 eq) was added and the reaction mixture stirred overnight from 0 °C to room temperature, then the reaction was quenched with 10% $Na_2S_2O_3$. After lyophilization, the crude product was purified by semi-preparative reversed HPLC (linear gradient: 16%→19% B in A, 3 CV, solutions used A: 25 mM NH_4OAc in H_2O, B: MeCN) and the fraction were freeze-dried to afford the desired product as a white powder **4** (4.2 mg, 0.012 mmol, 5%). ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 3.67 (dd, J = 13.1, 9.0 Hz, 2H), 3.27 (t, J = 7.0 Hz, 2H), 3.13 (dd, J = 5.8, 2.5 Hz, 1H), 2.66 (d, J = 5.8 Hz, 1H), 2.61 – 2.49 (m, 2H), 2.49 – 2.38 (m, 1H), 2.28 – 2.17 (m, 1H), 1.68 – 1.52 (m, 4H), 1.43 – 1.28 (m, 6H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 78.88, 73.55, 69.40, 52.47, 50.68, 41.71, 41.54, 36.76, 30.10, 30.02, 29.85, 27.68, 26.03; ^{1}H-NMR (400 MHz, D_2O): δ ppm 3.78 (d, J = 8.4 Hz, 1H), 3.64 (t, J = 10.2 Hz, 1H), 3.32 – 3.25 (m, 3H), 3.17 (dd, J = 5.7, 3.3 Hz, 1H), 2.75 (d, J = 5.8 Hz, 1H), 2.64 (dd, J = 9.9, 3.3 Hz, 1H), 2.52 – 2.32 (m, 2H), 2.24 (d, J = 7.0 Hz, 1H), 1.61 – 1.48 (m, 4H), 1.29 - 1.32 (m, 6H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 190.68, 77.04, 72.20, 68.74, 51.93, 51.22, 41.13, 41.00, 36.32, 28.75, 28.64, 28.60, 26.45, 25.01; LC-MS: R_t 3.12 min, linear gradient 10%→90% B with 10% NH_4OAc in 15 min; ESI-MS: m/z = 357.2 (M+H)$^+$; HRMS: calculated for $C_{15}H_{24}N_4O_6$ [M+H$^+$] 357.17686, found: 357.17687.

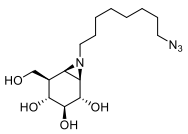


(1R,2S,3R,4S,5S,6R)-7-(8-(4-(4-(5,5-difluoro-3,7-bis(4-methoxyphenyl)-5H-4λ^4,5λ^4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octanoyl)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (5): Acyl Azido-aziridine compound **4** (5.2 mg, 0.015 mmol, 1.0 eq.) was dissolved in DMF (0.60 mL), red BODIPY-alkyne (8.5 mg, 0.017 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.010 mmol, 0.60 eq.) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.011 mmol, 0.70 eq.) were added to the solution under argon atmosphere, and the mixture was stirred at room temperature for 2 h, the reaction was checked with LC-MS within the elution system of 10% NH_4OAc. The reaction mixture was lyophilization and then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 42%→48% B in A, 3 CV, solutions used

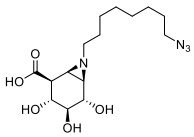
A: 25 mM NH_4OAc in H_2O, B: MeCN) and the fractions were freeze-dried to afford the desired product **5** as a purple powder (1.2 mg, 0.0014 mmol, 9.5%); ^{1}H-NMR (500 MHz, CD_3OD): δ ppm 7.87 – 7.81 (m, 4H), 7.74 – 7.70 (m, 2H), 7.67 – 7.61 (m, 2H), 7.45 (d, *J* = 4.4 Hz, 2H), 6.99 – 6.95 (m, 4H), 6.69 (d, *J* = 4.4 Hz, 2H), 4.33 (t, *J* = 7.1 Hz, 2H), 3.85 (s, 7H), 3.81 (d, *J* = 13.3 Hz, 4H), 3.08 (t, *J* = 5.1 Hz, 2H), 2.82 – 2.73 (m, 3H), 2.24 – 2.09 (m, 5H), 2.04 (s, 4H), 1.91 – 1.83 (m, 8H), 1.66 – 1.50 (m, 6H), 1.39 – 1.22 (m, 20H); LC-MS: R_t 7.51 min, linear gradient 0%→90% B with 10% NH_4OAc in 15 min; ESI-MS: *m*/*z* = 841.6 ($M+H^+$); HRMS: calculated for $C_{44}H_{51}BF_2N_6O_{48}$ $[M+H]^+$ 841.39023, found: 841.38982.



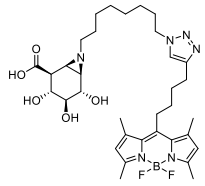
(1R,2S,3R,4S,5S,6R)-3,4,5-trihydroxy-7-(8-(4-((6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamido)methyl)-1H-1,2,3-triazol-1-yl)octanoyl)-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (6): Acyl azido-aziridine compound **4** (9.5 mg, 0.027 mmol, 1 eq.) was dissolved in DMF (1.0 mL) in eppi tube, Biotin-Ahx-alkyne (11 mg, 0.029 mmol, 1.1 eq.), $CuSO_4$ (1.0 M in H_2O) (5.3 µL, 0.0053 mmol, 0.20 eq.) and sodium ascorbate (1.0 M in H_2O) (5.6 µL, 0.0056 mmol, 0.21 eq.) was added to the solution under argon atmosphere, and the mixture was stirred at room temperature overnight. The reaction was checked by LC-MS within the elution system of 10% aq. NH_4OAc. The volatiles were removed by lyophilisation and then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 12%→27% B in A, 3 CV, solutions used A: 25mM NH_4OAc in H_2O, B: MeCN), the fractions were lyophilized directly to afford product **6** as a white powder (3.3 mg, 0.0044 mmol, 16.4%). ^{1}H-NMR (600 MHz, D_2O): δ ppm 7.84 (s, 1H), 4.58 (dd, *J* = 7.8, 4.8 Hz, 1H), 4.42 (s, 2H), 4.39 – 4.37 (m, 3H), 3.78 (d, *J*=8.4 Hz, 1H), 3.66 (t, *J* = 10.2 Hz, 1H), 3.34- 3.28 (m, 2H), 3.17 – 3.16 (m, 1H), 3.13-3.10 (m, 2H), 2.96 (dd, *J* = 12.6, 4.8Hz, 1H), 2.76-2.72(m, 2H), 2.70 (s, 2H), 2.66 (dd, *J* = 9.6, 3 Hz, 1H), 2.27 (t, *J* = 7.2 Hz, 3H), 2.22 (t, *J* = 7.2 Hz, 3H), 1.87 – 1.84 (m, 2H), 1.63-1.42 (m, 13H), 1.38-1.34 (m, 2H), 1.38 - 1.30 (m, 8H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 190.7, 182.4, 177.7, 177.5, 166.2, 124.7, 77.3, 72.4, 68.9, 63.0, 61.2, 56.3, 56.2, 51.4, 51.3, 41.3, 41.2, 40.6, 40.0, 39.6, 36.5, 36.4, 36.3, 35.2, 30.1, 28.9, 28.9, 28.8, 28.6, 28.6, 28.5, 26.3, 26.2, 26.1, 25.8, 25.1, 24.2; LC-MS: R_t 5.06 min, linear gradient 0%→90% B with 10% NH_4OAc in 15 min; ESI-MS: *m*/*z* = 751.8 ($M+H^+$); HRMS: calculated for $C_{34}H_{54}N_8O_9S$ $[M+H^+]$ 751.38027, found: 751.38028.



(1R,2S,3S,4R,5R,6R)-7-(8-azidooctyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (26): Unprotected aziridine **17**[14] (163 mg, 0.93 mmol, 1.0 eq.) was dissolved in DMF (4.0 mL). 1-Azido-8-iodooctane **25** (379 mg, 1.3 mmol, 1.5 eq.) and K_2CO_3 (448 mg, 4.0 mmol, 4.3 eq.) were added to the solution and the reaction mixture was stirred at 80 °C under refluxing conditions for 24 h, the volatiles were filtered via a pad of celite and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography (MeOH in DCM, 2%→20%) yielding **26** as light yellow oil (150 mg, 0.46 mmol, 49%). TLC: R_f 0.39 (DCM/MeOH, 5/1, v/v); $[\alpha]_D^{20}$ +50 (*c* = 1, MeOH); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.00 (dd, *J* = 10.1, 4.6 Hz, 1H), 3.66 – 3.59 (m, 2H), 3.28 (t, *J* = 6.8 Hz, 2H), 3.15 – 3.02 (m, 2 H), 2.40 – 2.33 (m, 1H), 2.18 – 2.11 (m, 1H), 2.01 – 1.98 (m, 1H), 1.93 – 1.86 (m, 1H), 1.66 (d, *J* = 6.3 Hz, 1H), 1.62 – 1.54 (m, 4H), 1.42 – 1.29 (m, 8H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 78.92, 73.79, 70.15, 63.76, 62.05, 52.40, 45.36, 45.34, 42.98, 30.45, 30.24, 30.14, 29.85, 28.25, 27.72; IR (neat, cm^{-1}): 3315, 2926, 2095, 1464, 1348, 1256, 1096, 1020, 818; LC-MS: R_t 4.52 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m*/*z* = 329.2 $(M+H)^+$; HRMS: calculated for $C_{15}H_{28}N_4O_4$ $[M+H^+]$ 329.21833, found: 329.21809.

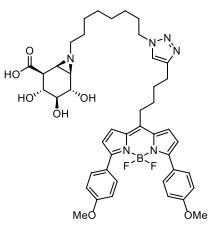


(1R,2S,3R,4S,5S,6R)-7-(8-azidooctyl)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (7): Compound **26** (77 mg, 0.23 mmol, 1.0 eq.), TEMPO (0.23 mg, 0.0015 mmol, 0.0065 eq.), and NaBr (9.5 mg, 0.092 mmol, 0.40 eq.) were dissolved in water (5.0 mL) at 2 °C. A 13% sodium hypochlorite solution (0.24 mL, 0.51 mmol, 2.2 eq.) was added dropwise to the mixture, which was adjusted to pH 10.5 by adding aq. NaOH (0.50 M). The reaction was conducted at 2 °C and checked with LC-MS with an elution system of 10% aq. NH_4OAc. Upon completion, the reaction was quenched by adding 96% EtOH (1.0 mL) and the pH was adjusted to 7 by adding aq. 0.5 M HCl. The volatiles were removed by lyophilization and then the crude product was purified by semi-preparative reverse phase HPLC (linear gradient: 19%→26% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN), the fractions were concentrated and lyophilized to afford product **7** as a white powder (12.0 mg, 0.035 mmol, 15%). ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 3.63 (d, J = 8.2 Hz, 1H), 3.62 – 3.57 (m, 1H), 3.27 (t, J = 6.9 Hz, 2H), 3.18 – 3.16 (m, 1H), 2.41 – 2.38 (m, 1H), 2.35 – 2.30 (m, 1H), 2.18 – 2.10 (m, 2H), 1.62 – 1.55 (m, 4H), 1.53 (d, J = 6.2 Hz, 1H), 1.43 – 1.27 (m, 8H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 78.89, 74.25, 70.09, 62.24, 52.47, 50.87, 44.64, 43.67, 30.53, 30.28, 30.21, 29.91, 28.39, 27.80; ^{1}H-NMR (400 MHz, D_2O): δ 3.62 (d, J = 8.7 Hz, 1H), 3.50 (t, J = 10.2 Hz, 1H), 3.24 (t, J = 6.9 Hz, 2H), 3.20 – 3.15 (m, 1H), 2.52 (dd, J = 9.9, 3.7 Hz, 1H), 2.32 - 2.24 (m, 1H), 2.15 - 2.04 (m, 2H), 1.68 (d, J = 6.3 Hz, 1H), 1.56 – 1.49 (m, 2H), 1.46 - 1.35 (m, 2H), 1.34 - 1.13 (m, 8H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 179.81, 76.44, 71.91, 68.53, 59.87, 51.23, 50.17, 42.90, 42.35, 28.42, 28.34, 28.11, 27.91, 26.31, 25.86; LC-MS: R_t 4.41 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: m/z = 343.2 $(M+H)^+$; HRMS: calculated for $C_{15}H_{26}N_4O_5$ $[M+H^+]$ 343.19760, found: 343.19777.

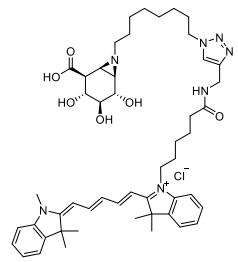


(1R,2S,3R,4S,5S,6R)-7-(8-(4-(4-(5,5-difluoro-1,3,7,9-tetramethyl-5H-4λ^4,5λ^4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (8): Azide compound **7** (5.0 mg, 0.015 mmol, 1.0 eq.) was dissolved in DMF (0.5 mL) in an 1 mL Eppendorf tube, green BODIPY-alkyne **27** (6.9 mg, 0.017 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 8.8 µL, 0.0088 mmol, 0.60 eq) and sodium ascorbate (1.0 M in H_2O, 9.5 µL, 0.0095 mmol, 0.65 eq.) were added to the solution under argon atmosphere, and the mixture was stirred at room temperature overnight. The reaction was checked by LC-MS with an elution system of 10% aq. NH_4OAc. The volatiles were concentrated *in vacuo* and then the crude product was purified by semi-preparative reverse phase HPLC (linear gradient: 35%→44% B in A, 3 CV, solutions used A: 25 mM NH_4HCO_3 in H_2O, B: MeCN) and the fractions were concentrated and lyophilized giving **8** as an orange powder (2.2 mg, 0.0032 mmol, 22%). ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 7.90 (s, 2H), 7.75 (s, 1H), 6.11 (s, 2H), 4.34 (t, J = 7.0 Hz, 2H), 3.63 – 3.58 (m, 2H), 3.18 – 3.12 (m, 1H), 3.05 – 3.01 (m, 2H), 2.78 (t, J = 7.3 Hz, 2H), 2.49 - 2.45 (m, 1H), 2.43 (s, 6H), 2.38 (s, 6H), 2.22 – 2.18 (m, 2H), 2.11 – 2.08 m, 1H), 1.92 – 1.88 (m, 5H), 1.88 – 1.83 (m, 2H), 1.69 – 1.63 (m, 2H), 1.57 – 1.48 (m, 2H), 1.36 – 1.21 (m, 8H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 154.92, 148.51, 147.88, 142.17, 132.57, 123.41, 122.59, 78.66, 74.02, 69.83, 62.02, 51.23, 44.79, 43.39, 32.31, 31.24, 30.87, 30.36, 29.87, 29.09, 28.19, 27.34, 25.91, 16.50, 14.43; LC-MS: R_t 6.24 min, linear gradient 10%→90% B, 12.5 min; ESI-MS: m/z = 671.2 $(M+H)^+$; HRMS: calculated for $C_{34}H_{49}BF_2N_6O_5$ $[M+H^+]$ 671.39043, found: 671.39001.

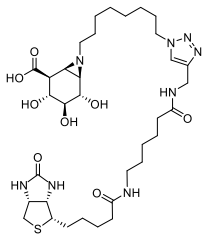
(1R,2S,3R,4S,5S,6R)-7-(8-(4-(4-(5,5-difluoro-3,7-bis(4-methoxyphenyl)-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (9):



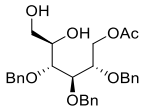
Azide compound **7** (6.0 mg,0.018 mmol, 1.0 eq.) was dissolved in DMF (0.6 mL), red BODIPY-alkyne **28** (10 mg, 0.021 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.011 mmol, 0.60 eq) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.012 mmol, 0.65 eq.) were added to the solution under argon atmosphere, and the mixture was stirred at room temperature overnight, the reaction was checked by LC-MS with an elution system of 10% aq. NH_4OAc. The volatiles were concentrated *in vacuo* and then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 43%→50% B in A, 12 min, solutions used A: 25 mM NH_4HCO_3 in H_2O, B: MeCN) and the fractions were concentrated and lyophilized to afford product **9** as a purple powder (1.9 mg, 0.0023 mmol, 13%).^{1}H-NMR (850 MHz, CD_3OD): δ ppm 7.87 – 7.83 (m, 4H), 7.71 (s, 1H), 7.44 (d, J = 4.3 Hz, 2H), 6.99 – 6.95 (m, 4H), 6.69 (d, J = 4.3 Hz, 2H), 4.33 (t, J = 7.1 Hz, 2H), 3.85 (s, 6H), 3.65 – 3.58 (m, 2H), 3.18 – 3.16 (t, J = 9.2 Hz, 1H), 3.08 – 3.06 (m, 2H), 2.80 – 2.78 (m, 2H), 2.39 – 2.38 (m, 1H), 2.30 – 2.25 (m, 1H), 2.13 – 2.11 (m, 1H), 1.89 – 1.82 (m, 6H), 1.51 (d, J = 6.2 Hz, 1H), 1.34 – 1.20 (m, 11H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 162.18, 158.78, 148.59, 146.79, 137.49, 132.16, 128.44, 126.52, 123.26, 121.03, 114.61, 78.88, 74.21, 70.08, 62.19, 55.82, 51.27, 50.83, 44.62, 43.65, 34.22, 31.28, 30.99, 30.42, 30.38, 30.22, 29.91, 28.26, 27.40, 25.82; LC-MS: R_t 7.00 min; linear gradient 10%→90% B in 12.5 min; ESI-MS: m/z = 827.3 $(M+H)^+$; HRMS: calculated for $C_{44}H_{53}BF_2N_6O_7$ $[M+H^+]$ 827.41172, found: 827.41098.



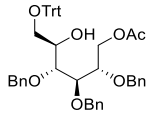
1-(6-(((1-(8-((1R,2S,3R,4S,5S,6R)-2-carboxy-3,4,5-trihydroxy-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)amino)-6-oxohexyl)-3,3-dimethyl-2-((1E,3E)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-indol-1-ium chloride (10): Azide compound **7** (6.0 mg, 0.017 mmol, 1.0 eq.) was dissolved in DMF (0.6 mL), Cy-5 allkyne **29** (12 mg, 0.021 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.011 mmol, 0.6 eq) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.012 mmol, 0.65 eq.) were added to the solution under argon atmosphere, and the mixture was stirred at room temperature overnight, the reaction was checked by LC-MS with an elution system of 10% aq. NH_4OAc. The volatiles were concentrated *in vacuo* and then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 40%→50% B in A, 12 min, solutions used A: 25 mM NH_4HCO_3 in H_2O, B: MeCN) and the fractions were concentrated and lyophilizedto afford product **10** as a dark blue powder (2.5 mg, 0.0029 mmol, 17%). ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 8.26 – 8.22 (m, 2H), 7.85 (s, 1H), 7.49 – 7.48 (m, 1.3 Hz, 2H), 7.42 – 7.39 (m, 2H), 7.31 – 7.28 (m, 1H), 7.27 – 7.25 (m, 1H), 6.62 (t, J = 12.4 Hz, 1H), 6.29 – 6.27 (m, 2H), 4.41 (s, 2H), 4.36 (t, J = 7.0 Hz, 2H), 4.08 (t, J = 7.6 Hz, 2H), 3.63 – 3.58 (m, 3H), 3.16 (t, J = 9.2 Hz, 1H), 2.40 – 2.35 (m, 1H), 2.25 (t, J = 7.3 Hz, 2H), 2.16 – 2.09 (m, 2H), 1.89 – 1.83 (m, 8H), 1.84 – 1.78 (m, 2H), 1.72 – 1.68 (m, 10H), 1.56 – 1.42 (m, 7H), 1.34 – 1.22 (m, 10H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 175.75, 175.36, 174.62, 155.53, 155.48, 146.14, 144.26, 143.53, 142.63, 142.52, 129.79, 129.73, 126.65, 126.25, 126.22, 124.18, 123.41, 123.27, 112.05, 111.85, 104.43, 104.28, 78.91, 74.23, 70.11, 62.14, 51.35, 50.81, 50.54, 50.50, 44.77, 44.61, 43.66, 36.47, 35.59, 31.53, 31.25, 30.30, 30.19, 29.86, 28.20, 28.12, 27.95, 27.81, 27.33, 27.30, 26.40, 24.18; LC-MS: R_t 6.29 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: m/z = 862.5 $(M+H)^+$; HRMS: calculated for $C_{50}H_{68}BF_2N_7O_6$ $[M+H^+]$ 863.53038, found: 863.52618.



(1R,2S,3R,4S,5S,6R)-3,4,5-trihydroxy-7-(8-(4-((6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)hexanamido)methyl)-1H-1,2,3-triazol-1-yl)octyl)-7-azabicyclo[4.1.0]heptane-2-carboxylic acid (11): Azide compound **7** (9.2 mg, 0.027 mmol, 1.0 eq.) was dissolved in DMF (0.8 mL) in eppi tube, Biotin-alkyne **30** (12.6 mg, 0.032 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 10 µL, 0.011 mmol, 0.60 eq.) and sodium ascorbate (1.0 M in H_2O, 11 µL, 0.012 mmol, 0.65 eq.) were added to the solution under argon atmosphere, and the mixture was stirred at room temperature overnight, the reaction was checked with LC-MS with an elution system of 10% aq. NH_4OAc. The volatiles were concentrated *in vacuo* and then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 40%→50% B in A, 12 min, solutions used A: 25 mM NH_4HCO_3 in H_2O, B: MeCN) and the fractions were concentrated and lyophilized to afford product **11** as a white powder (2.6 mg, 0.0035 mmol, 13%). ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 7.87 (s, 1H), 4.49 (dd, *J* = 7.9, 4.9 Hz, 1H), 4.46 – 4.34 (m, 4H), 4.31 (dd, *J* = 7.9, 4.4 Hz, 1H), 3.70 - 3.54 (m, 2H), 3.24 – 3.12 (m, 4H), 2.96 – 2.89 (m, 1H), 2.70 (d, *J* = 12.7 Hz, 1H), 2.43 (br, 1H), 2.34 - 21 (m, 6H), 1.98 – 1.81 (m, 3H), 1.78 – 1.47 (m, 11H), 1.44 (q, *J* = 7.5 Hz, 2H), 1.39 – 1.24 (m, 10H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 176.00, 175.98, 166.14, 124.20, 79.21, 78.79, 74.15, 70.06, 69.31, 63.68, 63.38, 62.13, 61.62, 57.03, 51.39, 44.75, 43.60, 41.06, 40.20, 36.82, 36.75, 35.62, 31.30, 30.40, 30.25, 30.12, 29.96, 29.79, 29.50, 28.25, 27.56, 27.40, 26.94, 26.70, 26.54; LC-MS: R_t 4.06 min, linear gradient 10-90% B in 12.5 min; ESI-MS: *m*/*z* = 737.3 $(M+H)^+$; HRMS: calculated for $C_{34}H_{56}N_8O_8S$ $[M+H^+]$ 737.40146, found: 737.40167.

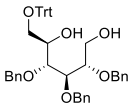


(2S,3R,4R,5R)-2,3,4-tris(benzyloxy)-5,6-dihydroxyhexyl acetate (32): The Intermediate **32** (20.6 g, 41.7 mmol) was prepared via previous reported synthesis,[17] in 6 steps from compound **31** in 83% total yield. $[\alpha]_D^{20}$ +11.8 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$) δ ppm 7.35 – 7.24 (m, 15H), 4.70 – 4.55 (m, 6H), 4.31 – 4.26 (m, 1H), 4.23 – 4.18 (m, 1H), 3.96 – 3.92 (m, 1H), 3.84 – 3.81 (m, 1H), 3.76 – 3.71 (m, 3H), 3.67 – 3.63 (m, 1H), 1.97 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$) δ ppm 170.75, 137.80, 137.66, 137.49, 128.80, 128.60, 128.58, 128.56, 128.22, 128.20, 128.07, 128.06, 78.04, 77.08, 76.73, 76.42, 73.94, 73.77, 73.36, 71.81, 63.63, 63.53, 20.96; IR (neat, cm^{-1}): 3447, 2876, 1738, 1497, 1454, 1368, 1231, 1026, 733, 696; HRMS: calculated for $C_{29}H_{34}NO_7$ $[M+Na^+]$ 517.21967, found: 517.21919.

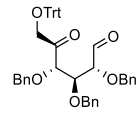


(2S,3R,4R,5R)-2,3,4-tris(benzyloxy)-5-hydroxy-6-(trityloxy)hexyl acetate (33): Product **32** (20 g, 41 mmol, 1.0 eq) was dissolved in DCM (500 mL). Then DMAP (0.25 g, 2.1 mmol, 0.05 eq.), Et_3N (10 mL, 72 mmol, 1.7 eq) and triphenylmethyl chloride (15 g, 46 mmol, 1.3 eq.) were added. The reaction was then stirred for 3 h at room temperature. Afterwards aq. sat. $NaHCO_3$ solution (600 mL) was added to quench the reaction. The product was extracted from the water layer with EtOAc. The combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude mixture was purified by silica gel column chromatography (10%→20%, EtOAc in pentane) providing product **33** (23 g, 31 mmol, 75 %). $[\alpha]_D^{20}$ +6 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.45 – 7.41 (m, 6H), 7.29 – 7.19 (m, 22H), 7.13 – 7.10 (m, 2H), 4.65 – 4.56 (m, 3H), 4.48 – 4.44 (m, 3H), 4.38 (dd, *J* = 12.0, 3.6 Hz, 1H), 4.15 – 4.04 (m, 2H), 3.94 – 3.85 (m, 2H), 3.77 – 3.74 (m, 1H), 3.36 – 3.29 (m, 2H), 3.01 (d, *J* = 5.6 Hz, 1H), 1.90 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 170.77, 143.86, 138.09, 137.86, 137.82, 128.78, 128.51, 128.46, 128.36, 128.25, 128.13, 127.95, 127.83, 127.74, 127.18, 86.75,

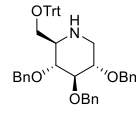
78.35, 77.11, 76.81, 74.31, 73.15, 72.86, 70.69, 64.74, 64.12, 20.95; IR (neat, cm^{-1}): 3030, 2876, 2347, 1738, 1449, 1231, 1067, 737, 696, 633; HRMS: calculated for $C_{48}H_{48}O_7$ [M+Na^+] 759.32922, found: 759.32905.



(2S,3R,4R,5R)-2,3,4-tris(benzyloxy)-6-(trityloxy)hexane-1,5-diol (34): Product **33** (21 g, 29 mmol, 1.0 eq.) was co-evaporated with toluene and then dissolved in dry MeOH (500 mL). Thereafter catalytic amount of Na (0.60 g) was added. The mixture was stirred for 3 h at room temperature and then concentrated at reduced pressure. The residue was dissolved in H_2O (300 mL). The product was extracted from the water layer with EtOAc. The combined organic layers were washed with brine (300 mL), dried over $MgSO_4$, filtered and concentrated under reduced pressure. Finally the reaction mixture was purified by silica gel column chromatography (12%→50%, EtOAc in pentane) giving product **34** (18.56 g, 26.71 mmol, 91 %). $[\alpha]_D^{20}$ -4 (*c*=1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.72 – 7.67 (m, 6H), 7.51 – 7.35 (m, 22H), 7.32 – 7.27 (m, 2H), 4.95 – 4.71 (m, 4H), 4.69 – 4.63 (m, 2H), 4.33 (br, 1H), 4.17 – 4.07 (m, 2H), 4.06 – 3.99 (m, 1H), 3.98 – 3.89 (m, 1H), 3.85 – 3.75 (m, 1H), 3.69 – 3.62 (m, 1H), 3.61 – 3.45 (m, 2H), 2.76 (br, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 143.71, 138.09, 137.75, 137.60, 129.06, 128.75, 128.61, 128.31, 128.22, 128.17, 128.04, 127.80, 127.74, 127.62, 127.59, 126.97, 125.97, 86.53, 79.39, 79.01, 76.91, 74.30, 72.81, 72.75, 70.63, 64.58, 61.64; IR (neat, cm^{-1}): 2932, 1736, 1449, 1371, 1240, 1043, 745, 696; HRMS: calculated for $C_{46}H_{46}NO_6$ [M+Na^+] 717.31866, found: 717.31864.

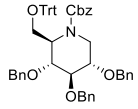


(2R,3R,4S)-2,3,4-tris(benzyloxy)-5-oxo-6-(trityloxy)hexanal: $COCl_2$ (8.7 mL, 101 mmol, 4.0 eq.) was dissolved in DCM (92 mL) and cooled to -78 °C. Then DMSO (9.0 mL, 126 mmol, 5.0 eq.) in DCM (54 mL) was added dropwise over 10 min. The mixture was stirred for 40 min while kept below -70 °C. Then a solution of **34** (17 g, 25 mmol, 1.0 eq.) in DCM (50 mL) was added dropwise over 15 min. After this the mixture was kept below -65 °C for 2 h while stirring, followed by dropwise addition of Et_3N (42 mL, 302 mmol, 12 eq.) over 10 min. The mixture was then warmed to -5 °C over 2 h and concentrated under reduced pressure.

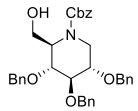


(2R,3R,4R,5S)-3,4,5-tris(benzyloxy)-2-((trityloxy)methyl)piperidine (35): Crude product aldehyde (1.0 eq.) was co-evaporated with toluene thrice after which it was dissolved in MeOH (500 mL). Subsequently NH_4HCO_2 (32 g, 504 mmol, 20 eq.) was added. The solution was then cooled to 0 °C and stirred for 20 min. Activated molecular sieves (250 g, 4 Å) were added and the mixture was stirred for another 20 min. Then $NaBH_3CN$ (6.3 g, 101 mmol, 4.0 eq.) was added. The mixture was cooled to 0 °C and stirred for 1 hour, after which the cooling was removed and the mixture was stirred for another 20 h. It was then filtered and concentrated under reduced pressure. The residue was dissolved in EtOAc (300 mL). The solution was then washed with aq., sat. $NaHCO_3$ (300 mL). The product was extracted from the water layer with EtOAc. Thereafter the combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. At last the mixture was purified by silica gel column chromatography (11%→25%, EtOAc in pentane) giving product **35** (14 g, 21 mmol, 85 % over 2 steps). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.51 – 7.20 (m, 28H), 6.90 – 6.88 (m, 2H), 4.97 (d, *J* = 10.8 Hz, 1H), 4.81 – 4.67 (m, 4H), 4.22 (d, *J* = 10.8 Hz, 1H), 3.58 – 3.50 (m, 3H), 3.41 – 3.28 (m, 2H), 3.24 – 3.15 (m, 1H), 2.79 – 2.71 (m, 1H), 2.62 – 2.50 (m, 1H), 1.98 (br, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 143.79, 138.93, 138.67, 138.21, 128.83, 128.51, 128.47, 128.26, 128.11, 128.05, 127.92, 127.85, 127.76, 127.66, 127.58, 127.16, 87.51, 86.65, 75.89, 80.93, 80.38, 75.10, 72.90,

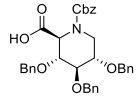
63.55, 60.41, 48.45; IR (neat, cm^{-1}): 3030, 2874, 1736, 1495, 1449, 1086, 1047, 731, 694; HRMS: calculated for $C_{46}H_{45}NO_4$ [M+H$^+$] 676.34214, found: 676.34168.



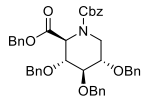
Benzyl (2R,3R,4R,5S)-3,4,5-tris(benzyloxy)-2-((trityloxy)methyl)piperidine-1-carboxylate (36): Product **35** (160 mg, 0.24 mmol, 1.0 eq.) was dissolved in 3mL THF, Et_3N (33 μL, 0.24 mmol, 1.0 eq.) and CbzCl (82 mg, 0.48 mmol, 2.0 eq.) were added into the solution, then the mixture was stirred at room temperature for 20 h. The reaction was diluted by EtOAc, washed by sat. aq. $NaHCO_3$ and brine. The organic layer was dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography (5%→30%, EtOAc in pentane) giving in product **36** (140 mg, 0.17 mmol, 72%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.65 – 7.13(m, 35H), 5.20 (br, 2H), 4.87 (d, J = 11.1 Hz, 2H), 4.78 – 4.67 (m, 2H), 4.63 – 4.42 (m, 3H), 4.42 – 3.99 (m, 2H), 3.91 – 3.80 (m, 2H), 3.72 – 3.40 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 143.90, 141.13, 138.17, 138.05, 136.56, 128.96, 128.88, 128.68, 128.42, 128.31, 127.97, 127.92, 127.88, 127.82, 127.72, 127.67, 127.64, 127.51, 127.08, 126.94, 126.11, 86.70, 82.09, 78.31, 74.77, 73.90, 73.07, 70.53, 67.35, 65.18, 61.89, 56.42, 29.75; HRMS: calculated for $C_{54}H_{51}NO_6$ [M+H$^+$] 810.37891, found: 810.37956.



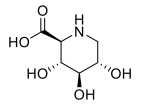
Benzyl (2R,3R,4R,5S)-3,4,5-tris(benzyloxy)-2-(hydroxymethyl)piperidine-1-carboxylate (37): Product **36** (140 mg, 0.17 mmol, 1.0 eq.) was dissolved in DCM/MeOH (2.0 mL, 1/1, v/v), followed by addition of *p*-toluenesulfonic acid monohydrate (6.6 mg, 0.035mmol, 0.20 eq.). After stirring overnight at room temperature, the reaction was quenched with Et_3N. The reaction was diluted by EtOAc, washed by brine. The combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The mixture was then purified by silica gel column chromatography (10%→50%, EtOAc in pentane) to afford product **37** (80 mg, 0.14 mmol, 83%). $[\alpha]_D^{20}$ + 2.2 (c = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.37 – 7.24 (m, 20H), 5.14 (s, 2H), 4.86 – 4.58 (m, 5H), 4.53 (d, J = 11.8 Hz, 1H), 4.02 – 3.85 (m, 3H), 3.78 – 3.67 (m, 5H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 156.37, 138.13, 138.03, 136.38, 128.63, 128.56, 128.52, 128.49, 128.20, 128.06, 128.03, 127.92, 127.88, 127.80, 77.53, 75.52, 73.37, 71.35, 67.63, 61.48; HRMS: calculated for $C_{35}H_{37}NO_6$ [M+H$^+$] 568.26936, found: 568.26952.



(2S,3R,4R,5S)-3,4,5-tris(benzyloxy)-1-((benzyloxy)carbonyl)piperidine-2-carboxylic acid (38): The primary alcohol compound **37** (80 mg, 0.14 mmol, 1.0 eq.) was dissolved in DCM/H_2O (1.5 mL, 2/1, v/v), then BAIB (112.7 mg, 0.35 mmol, 2.5 eq.) and TEMPO (4.4 mg, 0.028 mmol, 0.20 eq.) were added in the solution, the resulting mixture was stirred for 5 h at room temperature. The reaction was quenched with 10% $Na_2S_2O_3$, was extracted by EtOAc and washed by brine. The combined organic layers were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The mixture was then purified by silica gel column chromatography (25%→50%, EtOAc in pentane with 1% AcOH) to afford product **38** (58 mg, 0.10 mmol, 71%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 10.85 (br, 1H), 7.44 – 7.21 (m, 20H), 5.47 – 5.07 (m, 3H), 4.93 – 4.51 (m, 5H), 4.46 – 4.14 (m, 3H), 3.90 – 3.56 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 174.90, 174.74, 156.91, 156.17, 138.06, 137.58, 137.38, 136.29, 128.91, 128.67, 128.41, 128.38, 128.26, 128.00, 127.90, 127.77, 127.71, 127.39, 126.09, 77.48, 77.16, 76.84, 74.17, 73.75, 73.38, 71.64, 71.51, 70.72, 67.71, 55.95, 55.05, 40.74, 39.97; HRMS: calculated for $C_{35}H_{35}NO_7$ [M+H$^+$] 582.24863, found: 582.24883.



Dibenzyl (2S,3R,4R,5S)-3,4,5-tris(benzyloxy)piperidine-1,2-dicarboxylate: Product **38** (180 mg, 0.32 mmol, 1.0 eq.) was dissolved in DCM/H_2O (3.0 mL, 2/1, v/v) and cooled to 0 °C. TEMPO (10 mg, 0.064 mmol, 0.20 eq.) and BAIB (255 mg, 0.79 mmol, 2.5 eq.) were added at 0 °C. After 6 h stirring, the reaction mixture was quenched by 10% aq. $Na_2S_2O_3$ solution and extracted by DCM, being successively washed with water and brine, dried over $MgSO_4$ and concentrated under reduced pressure. The residue was co-evaporated twice with toluene. To a stirred solution of the residual syrup and benzyl bromide (76 µL, 0.63 mmol, 2.0 eq.) in dry DMF (5.0 mL) was added cesium carbonate (103 mg, 0.32 mmol, 1.0 eq.) at room temperature. The mixture was stirred for 3 h, and poured into water, before the mixture was extracted with Et_2O. The extract was washed by water and brine, dried over $MgSO_4$ and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (10 % → 30 %, EtOAc in pentane) to give Bn protected **38** (160 mg, 0.24 mmol, 75% over two steps). $[\alpha]_D^{20}$ +19 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, DMSO-d_6): δ ppm 7.34 – 7.22 (m, 25H), 5.14 – 5.07 (m, 3H), 4.99 – 4.95 (m, 2H), 4.62 – 4.47 (m, 6H), 4.25 – 4.24 (m, 1H), 4.03 – 3.99 (m, 1H), 3.87 – 3.85 (m, 1H), 3.76 – 3.72 (m, 1H), 3.64 – 3.60 (m, 1H), 3.09 (s, 1H); ^{13}C-NMR (100 MHz, DMSO-d_6): δ ppm 168.46, 155.21, 138.01, 137.44, 136.24, 135.07, 128.20, 128.04, 127.87, 127.81, 127.71, 127.63, 127.57, 127.54, 127.28, 127.20, 127.12, 127.06, 126.94, 126.90, 126.73, 74.51, 73.45, 70.89, 70.56, 69.90, 66.21, 66.03, 55.94, 40.26.



(2S,3R,4R,5S)-3,4,5-trihydroxypiperidine-2-carboxylic acid (12): A mixture of Product Bn-protected **38** (160 mg, 0.24 mmol, 1.0 eq.) and $Pd(OH)_2$/C (50 mg, 20% wt. loading(dry basis)) in 80% AcOH in H_2O (5.0 mL) was stirred at room temperature under a hydrogen atmosphere for 24 h. The catalyst was then filtered off and washed with MeOH. The filtrate and washings were combined and concentrated under reduced pressure to give semi-crystalline solid. Recrystallization from water ethanol gave final product **12** (24 mg, 0.14 mmol, 57%) as crystalline solid.^{1}H-NMR (400 MHz, D_2O): δ ppm 3.75 – 3.67 (m, 1H), 3.62 (t, *J* = 9.6 Hz, 1H), 3.47 (t, *J* = 9.0 Hz, 1H), 3.41 – 3.34 (m, 2H), 2.81 (t, *J* = 11.8 Hz, 1H), 1.88 (s, 1H); ^{13}C-NMR (100 MHz, D_2O): δ ppm 173.10, 76.05, 70.77, 67.73, 61.71, 45.70; HRMS: calculated for $C_6H_{11}NO_5$ [M+H$^+$] 178.07100, found: 178.07105.

6.5 References

[1] V. Lombard, H. G. Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490-495.

[2] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416-436.

[3] O. Tohyama, A. Imura, A. Iwano, J. N. Freund, B. Henrissat, T. Fujimori and Y. Nabeshima, *J. Biol. Chem.* **2004**, *279*, 9777–9784.

[4] W. S. Sly and C. Vogler, *Nat. Med.* **1997**, *3*, 719-720.

[5] a) C. R. Parish, C. Freeman and M. D. Hulett, *Biochim. Biophys. Acta* **2001**, *1471*, M99-M108; b) T. Konishi, T. Kotake, D. Soraya, K. Matsuoka, T. Koyama, S. Kaneko, K. Igarashi, M. Samejima and Y. Tsumuraya, *Carbohydr. Res.* **2008**, *343*, 1191–1201; c) M. Michikawa, H. Ichinose, M. Momma, P. Biely, S. Jongkees, M. Yoshida, T. Kotake, Y. Tsumuraya, S. G. Withers, Z. Fujimoto and S. Kaneko, *J. Bio. Chem.* **2012**, *287*, 14069-14077.

[6] a) D. Beaud, P. Tailliez and J. Anba-Mondoloni, *Microbiology* **2005**, *151*, 2323-2330; b) H. Takada, T. Hirooka, Y. Hiramatsu and M. Yamamoto, *Cancer Res.* **1982**, *42*, 331-334; c) H. Naz, A. Islam, A. Waheed, W. S. Sly, F. Ahmad and M. I. Hassan, *Rejuvenation Res.* **2013**, *16*, 352-363; d) N. Ilan, M. Elkin and I. Vlodavsky, *Int. J. Biochem. Cell Biol.* **2006**, *38*, 2018-2039.

[7] a) S. Atsumi, K. Umezawa, H. Iinuma, H. Naganawa, H. Nakamura, Y. Iitaka and T. Takeuchi, *J Antibiot.* **1990**, *43*, 49-53; b) K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima and M. Nakata, *Tetrahedron lett.* **1990**, *31*, 1171-1172.

[8] S. G. Withers and K. Umezawa, *Biochem. Biophys. Res. Comm.* **1991**, *177*, 532-537.

[9] K. Tatsuta, Y. Niwata, K. Umezawa, K. Toshima and M. Nakata, *J Antibiot (Tokyo)* **1991**, *44*, 912-914.

[10] a) M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft and J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907-913. b) W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529-12533.

[11] T. K. M. Shing and V. W. F. Tai, *J. Chem. Soc. Chem. Commun.* **1993**, 995-997.

[12] L. I. Willems, T. J. M. Beenakker, B. Murray, B. Gagestein, H. van den Elst, E. R. van Rijssel, J. D. C. Codee, W. W. Kallemeijn, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, *2014*, 6044-6056.

[13] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codee, G. A. van der Marel, J. M. F. G. Aerts and H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, *136*, 11622-11625.

[14] J. Jiang, W. W. Kallemeijn, D. W. Wright, A. M. C. H. van den Nieuwendijk, V. C. Rohde, E. C. Folch, H. van den Elst, B. I. Florea, S. Scheij, W. E. Donker-Koopman, M. Verhoek, N. Li, M. Schurmann, D. Mink, R. G. Boot, J. D. C. Codee, G. A. van der Marel, G. J. Davies, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. Sci.* **2015**, *6*, 2782-2789.

[15] J. Jiang, T. J. M. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codee, J. M. F. G. Aerts and H. S. Overkleeft, *Chem. Eur. J.* **2015**, *21*, 10861-10869.

[16] F. G. Hansen, E. Bundgaard and R. Madsen, *J. Org. Chem.* **2005**, *70*, 10139-10142.

[17] K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, *2014*, 6030-6043.

[18] B. La Ferla, P. Bugada and F. Nicotra, *J. Carbohyd. Chem.* **2006**, *25*, 151-162.

[19] S. Takahashi and H. Kuzuhara, *Biosci. Biotech. Biochem.* **1995**, *59*, 762-764.

[20] M. Michikawa, H. Ichinose, M. Momma, P. Biely, S. Jongkees, M. Yoshida, T. Kotake, Y. Tsumuraya, S. G. Withers, Z. Fujimoto and S. Kaneko, *J. Bio. Chem.* **2012**, *287*, 14069-14077.

[21] G. J. Davies, A. Planas and C. Rovira, *Acc. Chem. Res.* **2012**, *45*, 308-316.

[22] J. M. Aerts, A. W. Schram, A. Strijland, S. van Weely, L. M. Jonsson, J. M. Tager, S. H. Sorrell, E. I. Ginns, J. A. Barranger, G. J. Murray, *Biochim. Biophys. acta* **1988**, *964*, 303-308.

[23] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. D* **2011**, *67*, 235-242.

7

Exo- and *endo*-retaining β-glucuronidase activities and mechanisms revealed by cyclophellitol aziridine-based inhibitors and probes

Jianbing Jiang, Liang Wu, Chi-Lin Kuo, Wouter W. Kallemeijn, Gijsbert A. van der Marel, Jeroen D. C. Codée, Marco C. van Eijk, Bogdan I. Florea, Johannes M. F. G. Aerts, Herman S. Overkleeft and Gideon J. Davies, **2016**, manuscript in preparation.

7.1 Introduction

GH2 and GH79 β-glucuronidases catalyze the hydrolysis of β-glucuronic acid linkages from a widespread number of substrates and are expressed in numerous different species including bacteria and eukaryotes.[1,2] GH2 and GH79 β-glucuronidases are retaining enzymes that process their substrate to yield β-glucuronic acid residues. β-Glucuronidase mediated hydrolysis proceeds through a Koshland two-step double displacement mechanism (Figure 1A).[3] In the first step the aglycon is protonated through the action of a general acid-base residue positioned at the β-face of the substrate. Concomitantly, in a formal S_N2 nucleophilic displacement the protonated aglycon is expelled via nucleophilic attack by a nucleophilic residue residing at the α-face of the substrate, forming an intermediate enzyme-α-glucuronide adduct. Next, water enters the enzyme active site and in a reversal of steps the glucuronide is released from the enzyme with overall retention of anomeric configuration. This mechanism of action is adopted by both *exo*- and *endo*-β-glucuronidases, and the differences between these enzymes are in their substrate recognition and physiological roles.

Both GH2 and GH79 belong to the GH-A clan.[4] The GH2 family consists of only *exo*-acting enzymes, such as GH2 human β-glucuronidase (GUSB), which is responsible for cleaving β-linked D-glucuronides from the non-reducing end of glycosaminoglycans (GAGs). Deficiency of GUSB causes the autosomal recessive disease: mucopolysaccharidosis type VII (MPSVII), also known as Sly syndrome.[5] The GH79 family consists of both *exo*-acting and *endo*-acting enzymes, such as bacterial β-Glucuronidase from *Acidobacterium capsulatum* (AcaGH79) and human heparanase (HPSE) respectively.[6,7] HPSE is an important factor in the processing and degradation of heparan sulfate, and has been implicated in a variety of processes underlying human pathologies such as inflammation, tumor metastasis and angiogenesis.[8]

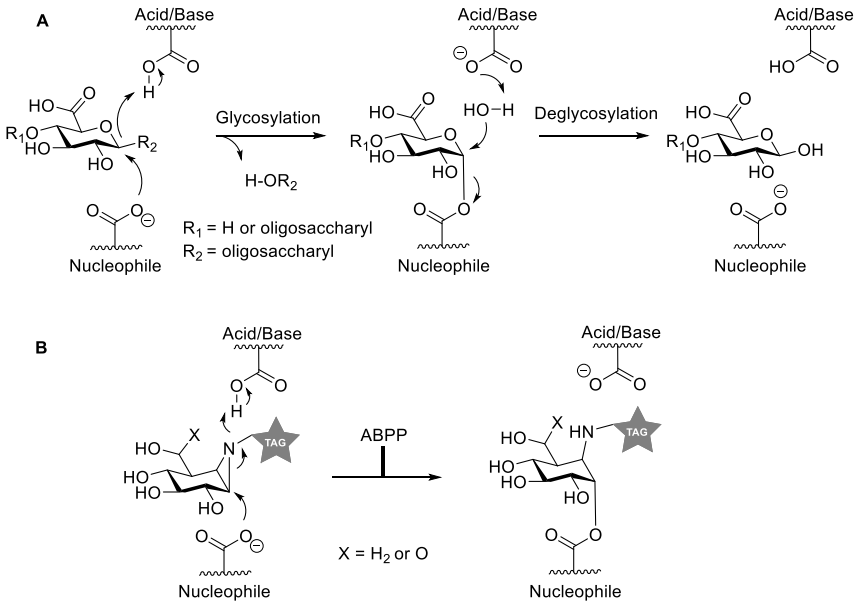


Figure 1. A) Koshland two-step double displacement mechanism employed by GH2 and GH79 retaining β-glucuronidases. B) Cyclophellitol aziridine-derived mechanism-based inhibitors and activity-based probes for retaining β-glucosidases (**1**, X = H_2) and retaining β-glucuronidases (**2**, **3** X = O).

Retaining glycosidases that employ a Koshland double displacement mechanism and that form during substrate processing a covalent enzyme-glycoside adduct are amenable to activity-based protein profiling (ABPP).[9] It has been shown in the past that fluorescent or biotin-conjugated derivatives of the broad-spectrum retaining β-glucosidase inhibitor, *N*-alkyl cyclophellitol aziridine (Figure 1B, **1**, X = H_2) are suitable activity-based probes (ABPs) for profiling retaining β-glucosidases *in vitro*, *in situ* and *in vivo*.[10,11] The efficiency of the ABPs is based on their tight initial binding to the enzyme active site, subsequent protonation of the aziridine nitrogen, and finally S_N2 substitution of the aziridinium ion by the active site nucleophile. The resulting enzyme-cyclitol adduct is comparatively more stable than the acylal

linkage that results from the natural process (as in Figure 1A) and the retaining β-glucosidase is thus inhibited in a mechanism-based manner. The covalent enzyme-inhibitor adduct is stable after protein denaturation, allowing for biochemical (gel electrophoresis) and analytical (mass spectrometry-based proteomics) study of the captured enzyme(s).

Altering the configuration of the cyclophellitol aziridine scaffold yielded probes comparably effective and selective for retaining GH27 α-galactosidases[12] and GH29 α-L-fucosidases,[13] showing the general applicability of the activity-based glycosidase profiling methodology. Modification of the cyclophellitol aziridine core to emulate a glucuronic acid moiety (Figure 1B, **2**,**3**, X = O) yielding ABPs capable of modifying and visualizing β-glucuronidase AcaGH79 have been described in Chapter 6, as well as the design, synthesis and evaluation of a suite of retaining β-glucuronidase inhibitors and ABPs. Here, an in-depth study on the use of the activity-based β-glucuronidase probes in labeling and identification of both *exo*- and *endo*-glucuronidases in various research settings and from various tissues is described. As well, crystal structures of compound **4** in complex with wild type and (E to Q) nucleophile-mutant human HPSE (*endo*) GH79 retaining β-glucuronidases, comparing the complex structure of **4** with bacterial (*exo*) AcaGH79 (Chapter 6) are described in this chapter.

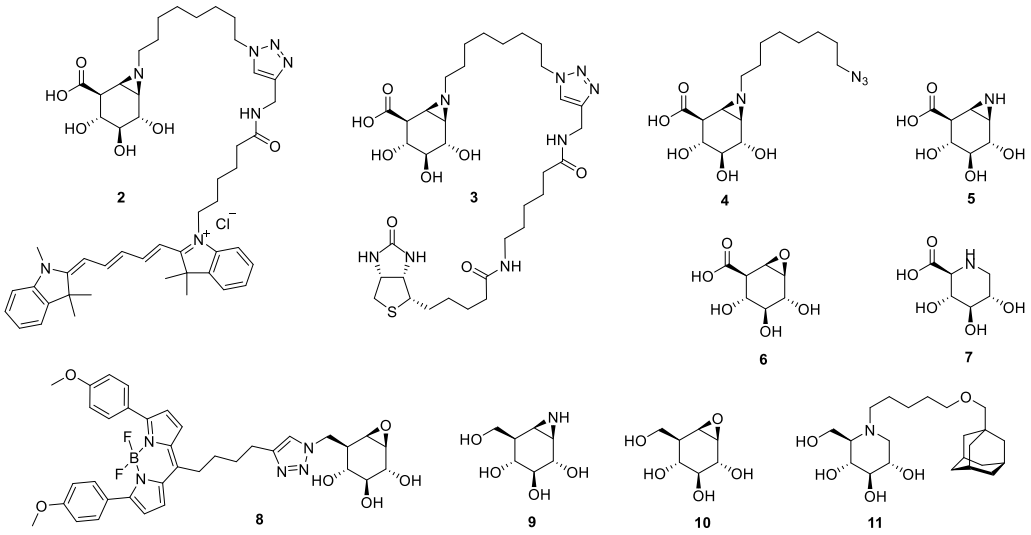


Figure 2. Inhibitors and ABPs used in the here-reported study. **2** and **3** are β-glucuronidase ABPs; **4**, **5**, **6** and **7** are β-glucuronidases inhibitors; **8** is β-glucosidase ABP; **9**, **10** and **11** are β-glucosidase inhibitors.

7.2 Results

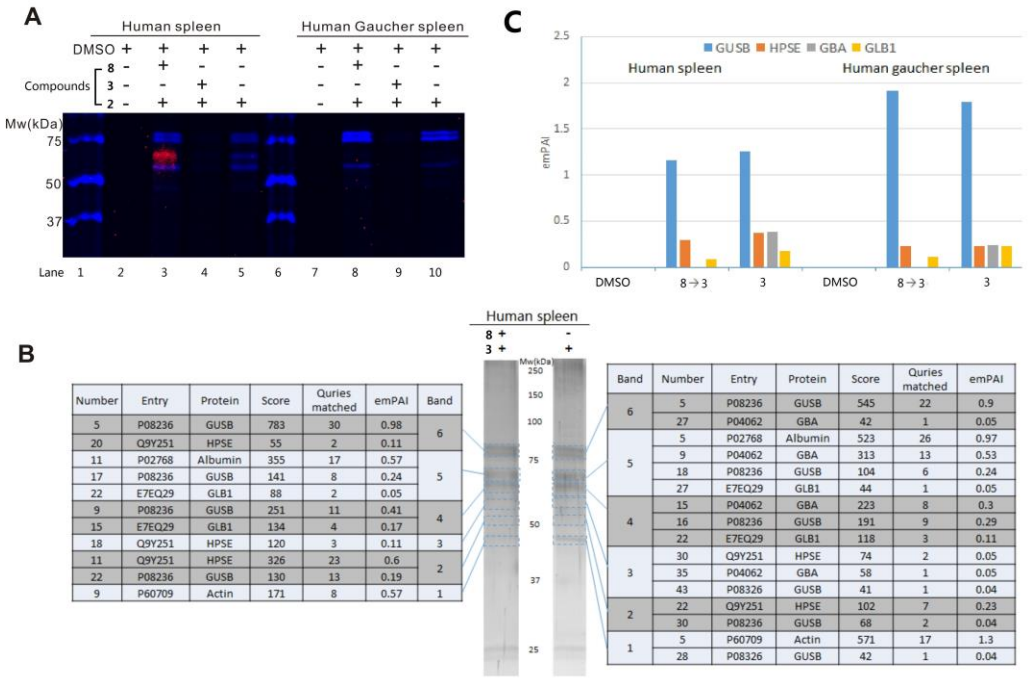
Inhibitor and probe design

The β-glucuronidase inhibitors and probes used in this study are depicted in Figure 2 (see Chapter 6 for their synthesis). ABPs **2** and **3** are composed of β-glucuronic acid-configured

cyclophellitol aziridine, bearing at the aziridine-nitrogen an alkyl spacer that terminates in Cy5 and biotin respectively. Fluorescent ABP **2** is used for in-gel β-glucuronidase detection (fluorescence scanning of gel slabs) and compound **3** for biotin-Streptavidin pull down enrichment-identification through chemical proteomics strategies with mass spectrometry. The set of ABPs are accompanied by azide-modified cyclophellitol aziridine **4** (the precursor in the preparation of ABPs **2**, **3**), unsubstituted cyclophellitol aziridine **5**, cyclophellitol-6-carboxylate **6** as well as trihydroxypipecolic acid **7** as a competitive β-glucuronidases inhibitor. In addition to the β-glucuronidase inhibitors and ABPs, the acid glucosylceramidase (GBA) ABP **8**, mechanism-based GBA inhibitors **9** (cyclophellitol aziridine), **10** (cyclophellitol) and competitive GBA inhibitor **11** are also employed in the studies as described below.

Chemical proteomics reveals GH2 GUSB and GH79 HPSE as targets of ABPs 2 and 3

In a first set of experiments to assess the activity and selectivity of our β-glucuronidase ABPs human spleen tissue extracts were treated with Cy5-modified cyclophellitol aziridine **2**. Human spleen is expected to express high GUSB levels and low HPSE levels, according to the Expression Atlas transcriptome database.[14] Following denaturing of the samples and resolving of their protein content on SDS-PAGE, labeled proteins were visualized by fluorescence scanning of the wet gel slab at 605 and 695 nm (overlay of Cy3 and Cy5 channel, Figure 3A left panel, for separate channel images see the supporting information Figure S1). Several distinct bands at molecular weights roughly between 50 and 80 kD became apparent, all of which were absent in the (DMSO) mock-treated sample (lane 2). The lower and second to highest band was putatively assigned as GUSB and the highest molecular weight band as its pro-form, based on literature information on GUSB expression forms in human spleen.[15-16] The band at 58-62 kD could not be assigned to a known β-glucuronidase, but it was previously reported that labeling human spleen with the close (with respect to the configuration of the cyclitol aziridine) structural analogue, ABP **8**, yielded exclusive labeling of glucocerebroside (GBA) at exactly this position.[17] Indeed, pre-incubation with **8** prior to treatment with **2** yielded a labeling pattern (lane 3) reminiscent of that observed in lane 5, but lacking the band at 58-62 kD. The identity of GUSB and pro-GUSB as labeling-target of ABP **2** was further confirmed by repeating the exact same series of experiments, but in extracts derived from human Gaucher spleen tissue containing less activity of GBA[18] (Figure 3A right panel, note the absence of the bands at 50-80 kD in both lanes 8 and 10). ABP **2** labeling in both tissue extracts could be completely blocked by pre-incubation with biotin-aziridine **3**, revealing that this pull-down chemical proteomics probe targets at least the same set of proteins as fluorescent probe **2**.



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 3. A) Labeling of human spleen and human Gaucher spleen with ABP **2** and **8** (lanes 1 and 6 are protein marker). B) Silver staining of in-gel digestion and identification of target proteins modified by biotin-ABP **3** with Mascot data parameters table. C) Mascot parameter emPAI value of glycosidases identified after on-bead pull-down and processing of spleen lysate with biotin-ABP **3**.

With the aim to unambiguously establish the targets of the β-glucuronidase ABPs, a number of chemical proteomics experiments were performed. As described above, biotin-aziridine ABP **3** is able to compete with fluorescent ABP **2** (Figure 3A, lane 4 and lane 9) and thus targets at least the set of proteins as seen in the gels in Figure 3A, lane 5 and lane 10. In a first chemical proteomics experiment, human spleen and Gaucher spleen extracts were treated with biotin-aziridine **3**, DMSO (mock treatment), or treatment with **3** after pre-incubating with ABP **8**. Biotinylated proteins were captured on magnetic streptavidin-loaded beads and non-biotinylated proteins were washed off. The samples were separated in two parts: one-third for in-gel digestion and two-thirds for on-bead digestion. The captured proteins for in-gel digestion were liberated from the beads by treatment with Laemmli buffer containing excess biotin. Thus, protein pools were obtained that besides endogenous biotinylated proteins should also contain those proteins able to react covalently and irreversibly with biotin-aziridine **3**. The eluted proteins were resolved on SDS-PAGE, visualized by silver staining (Figure 3B and SI, Figure S1B), stained bands excised from the gel and treated with trypsin. Tryptic peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) and the resolved peptide sequences matched against the Mascot database.[19] As can

be seen (Figure 3B), GUSB was identified in this manner, alongside GBA as the main off-target protein alongside retaining β-galactosidase (GLB1), keratin, and a number of endogenously biotinylated CoA carboxylases (SI, Figure S1C). Interestingly, HPSE in band 2 and 3 of both lanes were also identified in both wild type and Gaucher tissues (SI, Figure S1D), despite low HPSE expression level according to the Expression Atlas database. Essentially, the same results were obtained when performing the on-bead trypsin digestion instead of in-gel (Figure 3C, the Exponentially Modified Protein Abundance Index (emPAI)[20] value offers approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result). In both experimental set-ups, GBA-specific ABP **8** could compete for GBA labeling/pull down. Identification of the labeled proteins in the chemical proteomics experiments shown in Figure 3B, 3C is based on sequence identification using the Mascot search engine. The software does not recognize the active site fragments of the ABP-modified retaining glycosidases because of their altered molecular structure. Therefore, a manual search was performed for the molecular weight of the putative tryptic fragment containing the active site nucleophile modified by biotin aziridine **3**, both for GUSB and HPSE, in the LC-MS spectra. Neither of the expected masses (*m/z*=4494.19 for GUSB Tyr532-Lys563 and *m/z*=3882.90 for HPSE Val339-Lys368) could be identified and thus an alternative digestion protocol was used for protein digestion. Following this protocol, the trypsinolysis-derived peptide pool was further treated with endoproteinase Glu-C,[21] a protease that cleaves specifically after acidic residues. Although not successful for HPSE, due to the small amount of total protein in the samples, this protocol proved successful in identifying the oligopeptide containing the active site nucleophile of human GUSB in Gaucher spleen extract. As can be seen (SI, Figure S1E, F), a peak corresponding to the predicted mass of this modified peptide from Gaucher spleen extract was observed in the LC-MS trace, and LC-MS/MS fragmentation of this peptide delivered the expected sequence (*m/z*=2261.14) Tyr532-Glu544 with nucleophile active-site Glu540 modified by ABP **3**.

In-depth comparative and competitive ABPP on GH2 human GUSB

As evidenced from the proteomics data presented above, the main off-target identified for ABPs **2** and **3** proved to be the retaining β-glucosidase, GBA. GBA was also identified after labeling several mouse tissue extracts with ABP **2** (Figure 4B), and the labeling intensity with ABP **2** in mouse liver extracts appeared optimal at pH 5-6, corresponding with the observation made using human spleen extract (Figure 4C). Previously, it was reported that 6-modified (glucopyranose numbering) cyclophellitol derivative **8**,[17] with a BODIPY-TMR tag positioned at the 6 position is a highly selective and sensitive probe for GBA, which when deficient is at the basis of the lysosomal storage disorder Gaucher disease. The activity of GBA and GUSB as well as the impact of putative inhibitors on these enzymes can be dissected, as shown in Figure 4, by making use of Cy5-aziridine **2** and BODIPY-TMR epoxide **8**. Human spleen extracts were treated with Cy5-aziridine **2**, BODIPY-TMR-epoxide **8**, or a combination thereof, and a series of

competitive and covalent inhibitors **5**-**7** (GUSB inhibitors) and **9**-**11** (GBA inhibitors) was included in a competitive ABPP format. Lanes 1-3 recapitulate the results shown in Figure 4A, with red fluorescence (Cy3) corresponding to GBA labeling and blue fluorescence to GUSB (including its pro-form) labeling. The covalent (**5**, **6**) and competitive (**7**) GUSB inhibitors selectively outcompete ABP **2** (lanes 3-5), whereas the corresponding covalent (**9**, **10**) and competitive (**11**) inhibitors abolish to a large extend labeling with ABP **8**.

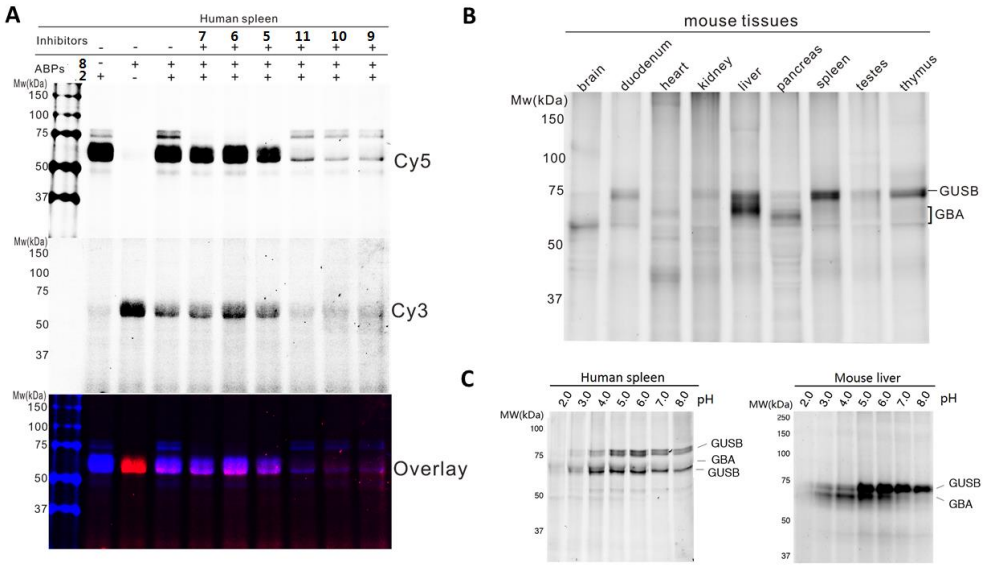


Figure 4. A) *In vitro* competitive ABPP on human spleen lysate after probing for both GBA (with ABP **8**) and GUSB (with ABP **2**) in Cy3, Cy5 and overlay channels. B) *In vitro* labeling of several mouse tissue extracts with ABP **2** for determination of GUSB expression and activity in different tissues. C) Detection of GUSB from human spleen and mouse liver extracts labeling efficiency at various pH with ABP **2**.

In-depth comparative and competitive ABPP on GH79 human HPSE

With the aim to explore Cy5 aziridine ABP **2** to study HSPE – the second β-glucuronidase activity identified in our chemical proteomics experiments (see Figure 3B), we turned our attention to tissues expressing this enzyme in relative high abundance. As HPSE could not be identified from fluorescence scanning on SDS-PAGE gel slabs derived from human spleen extract labeling (see Figure 3A and SI, Figure S2A), we brought HPSE to overexpression in the first instance in HEK293T cells. Labeling extracts of this tissue with Cy5 aziridine **2** showed time-dependent labeling of HPSE both as the mature enzyme and as the pro-enzyme, which is originally synthesized in an inactive 65 kDa form.[22] The identity of these bands was confirmed with immunostaining and Western blotting with an anti-HPSE 1 antibody (Figure 5A). In an independent experiment, primary human fibroblasts were treated with recombinant pro-HSPE, and its internalization was subsequently demonstrated after labeling cell extracts with ABP **2**

(SI, Figure S2B). Labeling of pro-HSPE was somewhat surprising, as literature data indicates that this immature enzyme is not active towards its natural substrate, heparan sulfate.[23] However, comparison of fluorescence images with those obtained from anti-HPSE western blotting shows significantly more pro-HPSE than mature HPSE in the cell lysates, indicating that labeling of the mature protein is more effective than that of the proenzyme.

Human platelets are known to express high HSPE levels and were therefore used to study labeling of this *endo*-glycosidases in more depth.[24] In platelet extracts prepared in the presence of EDTA elevated HPSE activity was observed, as indicated by labeling with ABP **2** and immunostaining (SI, Figure S2C). As can be seen (Figure 5B and SI, Figure S2E), ABP **2** clearly and in a pH-dependent manner labels HSPE. Labeling optimum is at pH 4-5, thus the pH corresponding with lysosomal pH – the natural environment of the enzyme. In contrast, labeling of GUSB in its various forms occurs at comparatively higher pH 5.5-6.5 (SI, Figure S2F). Therefore, GUSB and HPSE labeling can be done selectively in neutral and acidic conditions respectively. In these experiments using endogenous platelet samples no pro-HSPE labeling was observed. HSPE labeling in platelet extracts is optimal at 100 nanomolar probe and 60 minutes incubation (SI, Figure S2D) and can be partially competed by pre-incubation with heparin (Figure 5C), which had no effect on labeling of GUSB. It was observed that compounds **5**, **6**, **7** and known inhibitor Siastatin B[25] in millimolar concentration could completely block both GUSB and HPSE activity in platelet, following by ABP **2** incubation and readout (SI, Figure S2G,H). These results suggest that the GUSB catalytic pocket can only accommodate small monosugar-like inhibitors, whereas the HPSE binding site can accommodate both polysaccharide heparin or monosugar inhibitors.

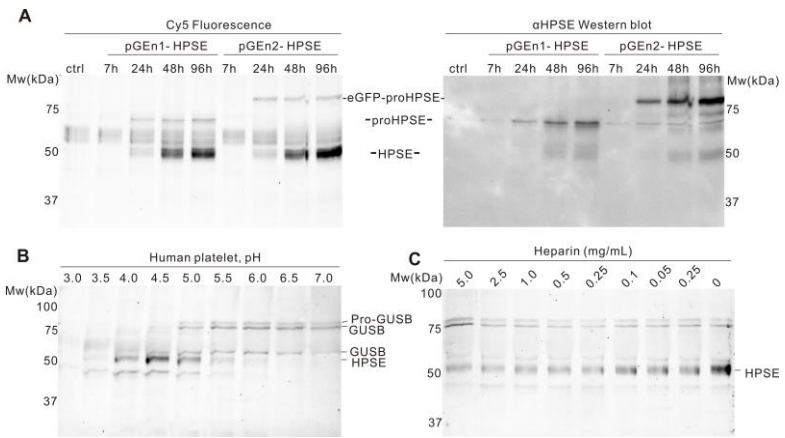


Figure 5. A) Overexpressed HPSE in HEK293 cells visualized with ABP **2** at different time points, by scanning the slab gel in Cy5 fluorescent channel (left) and Western blot analysis with HPSE specific antibody (right). B) Detection of endogenous HSPE in human platelets with ABP **2** at various pH values. C) HSPE labeling in platelets can be partially abolished by incubation with its natural substrate, heparin.

Structural analysis reveals mechanistic similarities between two GH79 β-glucuronidases: AcaGH79 (*exo*) and HSPE (*endo*)

To further prove the covalent labeling of both *exo*- and *endo*- GH79 β-glucuronidases by alkyl aziridine ABPs and analyze the catalytic mechanism employed by the enzymes, the crystal structures of AcaGH79 (see Chapter 6) and HPSE with mechanism-based inhibitor **4** were obtained. The resulting crystal structures clearly showed the formation of enzyme-**4** adducts, with a covalent O-C bond between the enzyme active-site nucleophile (E287 for AcaGH79, E343 for HPSE) and the carbon equivalent to the anomeric carbon of a bona fide substrate in **4** (Figure 6A). The alkylated linker on aziridine nitrogen shows considerable flexibility and accommodates the space normally occupied by enzyme substrates. During catalytic hydrolysis of both enzymes, aziridine rings on **4** open in trans-diaxial fashion and the substituted cyclohexane adopts $^{4}C_{1}$ conformation in both *exo*-AcaGH79 and *endo* HPSE catalytic pocket. The corresponding mutant enzymes (E287Q for AcaGH79, E343Q for HPSE) crystal structures with **4** were also obtained, revealing the umambiguous electron density for unhydrolyzed **4** in $^{4}H_{3}$ conformation (Figure 6B and Chapter 6 Figure 4). These results are consistent with the expected $^{1}S_{3}$-$^{4}H_{3}$-$^{4}C_{1}$ catalytic itinerary.[26,27] It is also observed that overlay of AcaGH79-**4** and HPSE-**4** complex exhibited the similar position and configuration of **4** or residues in catalytic pockets (Figure 6C). Moreover, although the *exo*-GH2 GUSB-**4** complex crystal structure (which would provide structural insights into covalent labeling of GH2 family β-glucuronidase) is not obtained yet, the overlay of active site residues of a GH2 *E. coli* β-glucuronidase[28] with two GH79 β-glucuronidases revealed the structural similarity between their active sites. Therefore, **4** could also fit in the catalytic cleft of GH2 GUSB well with similar hydrogen bond between amino acid residues and hydroxyl groups of cyclohexane for complex stability as GH79 β-glucuronidases-**4** complexes.

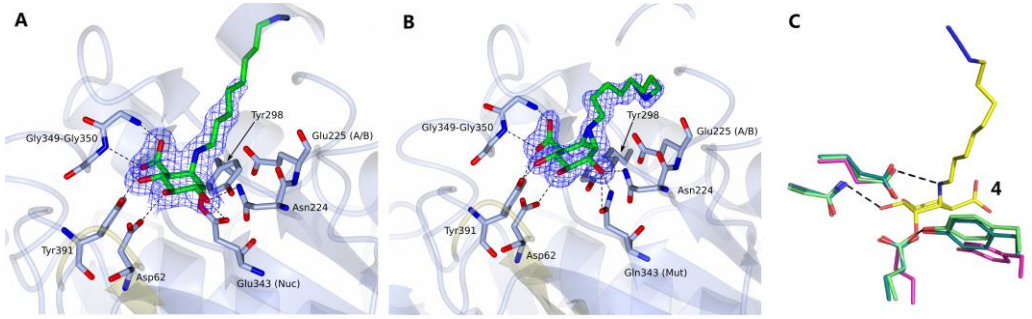


Figure 6. Crystal structures of A) HPSE-**4** complex and B) mutant (E343Q) HPSE-**4** complex. C) The overlay of HPSE (dark green)-**4** (yellow), AcaGH79 (light green)-**4** complexes and *E.coli* GH2 β-glucuronidase active site residues (purple).

7.3 Discussion

Activity-based probes (ABPs) have been proven to be a powerful tool to investigate retaining GHs activity in biological complex systems. GH2 and GH79 β-glucuronidases can be labeled

with high efficiency using β-glucuronic acid-configured cyclophellitol aziridine ABPs **2** and **3**, which is consistent with the efficacy in labeling of GH30 β-glucosidases, GH27 α-galactosidases and GH29 α-L-fucosidases with *N*-alkyl aziridine ABPs corresponding to the configuration and substitution pattern of their substrate glycosides.[11] Chemical proteomics applied by biotin ABP **3** in affinity enrichment of human spleen extracts for mass spectrometry analysis identified not only the main target enzyme GUSB, but also GBA as the main off target. Fluorescent ABP **2** labeled GBA in the same position as the specific GBA ABP **8** when analyzed by SDS-PAGE. GBA recognition of substrate with sugar C5 position modification has much more tolerance, comparing and former GBA inhibitors or probes.[10,18] Similarly, α-glucoside-configurated aziridine ABPs of GH31 α-glucosidases also have ability to label GBA, which has been described in Chapter 5. Based on the broad spectrum labeling of ABP **2**, more advanced application of this probe such as selective inhibitor discovery can be performed, Figure 4 shows an example to selectively distinguish between GBA and GUSB labeling with suitable concentration of the inhibitors **5**-**7** (used for GUSB inhibition) and **9**-**11** (prefer to GBA inhibition) in competitive ABPP strategy.

Gel based labeling of human platelet by ABP **2** (Figure 5 and SI, Figure S2) and chemical proteomics identification of human spleen by ABP **3** gave the first evidence of both *exo*- and *endo*-β-glucuronidases (GUSB and HPSE) labeling by ABPs presented in Figure 3. Both enzyme labeling are pH dependent. Interestingly, their maximum labeling pH is different, pH 4.0-5.0 for HPSE and pH 5.5-6.5 for GUSB (Figure 4). Thus, a pH-selective way could be used to label either GUSB or HPSE in a mixture of protein lysates. Based on these results, it is proposed that the monosaccharide-based ABPs can also label *endo*-glycosidase which cleaves interally within oligosaccharide chains. This point was also proved by competitive ABPP labeling on human platelet by ABP **4**, the monosugar inhibitors **5**, **6**, **7** and Sistatin B can inhibit both *exo*-GUSB and *endo*-HPSE activity, but oligosaccharide heparin can only block HPSE activity (Figure 5C and SI, Figure S2G,H). Therefore, although *exo*- and *endo*-β-glucuronidases share the same two-step double displacement mechanism, the catalytic pocket of *exo*-β-glucuronidase could only be occupied by monosugar from the terminal end of glycoconjugates, whereas for *endo*-β-glucuronidase, the catalytic cleft is suitable for both monosaccharide and oligosaccharides.

Although a structure of a glucuronic cyclophellitol aziridine bound to a GH2 family enzyme was not obtained during this project, the active site peptide identification data after ABP **3** pull down clearly shows the covalent modification of GH2 GUSB nucleophile Glu540 by ABP **3**. Active site comparisons between GH79 and GH2 enzymes allow us to infer the likely configuration of aziridine probes within a GH2 active site. In the CAZy classification[1,4], both GH79 and GH2 enzymes belong to the wider GH-A clan, enzymes which are characterized by the presence of a $(\beta/\alpha)_8$ barrel , which is a conserved protein fold consisting of eight α-helices and eight parallel β-strands that alternate along the peptide backbone. The active site of both

GH2 and GH79 enzymes are located in this $(\beta/\alpha)_8$ barrel. Furthermore, clan GH-A enzymes show functional conservation of several key amino acids within their active sites, which are involved in the catalytic process.[29]

Crystallographic study of GH79 *exo*-β-glucuronidase (AcaGH79)[6] complex with aziridine **4** has been viewed in Chapter 6. Active site overlays of **4** labelled AcaGH79 and HPSE (GH79) clearly show the conservation of four residues within their active sites, which appear invariant between GH2 and GH79 family enzymes: two glutamates which comprise the catalytic nucleophile and acid/base, an asparagine which lies –1 from the acid/base, and a tyrosine which lies adjacent to the nucleophile (Figure 6C). These are the key amino acids which form the actual catalytic machinery of these enzymes, and are involved in attacking the anomeric centre of glucuronide substrates, or the aziridine ABPs. In contrast, the main structural difference between the active sites of GH2 and GH79 enzymes appears to be in their method of recognising the C8 carboxylate of their glucuronide substrates. GH79 enzymes employ a conserved Tyrosine backbone amide motif to hold the C8 carboxylate, via primarily H-bonding interactions (Figure 6A). GH2 enzymes have been suggested to interact with C6 carboxylates in a more electrostatic fashion, using a charged Asn-X-Lys motif (where X is typically His, Lys or Arg)[30] (Figure 6C). Whilst these differences towards the C8 position may affect the initial non-covalent interaction between ABPs and GH79 vs GH2 enzymes, they are unlikely to play a significant role in the subsequent irreversible covalent bond formation step. Thus given the similarities between the active site machinery of GH2 and GH79 enzymes, it is postulated that ABPs bound to GH2 GUSB should show a similar configuration in the active as observed for AcaGH79 and HPSE.

Four crystal structures of wild type and mutant human HPSE in complex with compound **4** provide strong evidence for the covalent binding of cyclophellitol aziridine to active *endo*-β-glucuronidases and therefore the validity of the cyclophellitol aziridine design for activity-based profiling of retaining glycosidases that employ the Koshland double displacement mechanism. HPSE-**4** complex formation also provides evidence that monosugar type inhibitors or probes are able to effectively bind with *endo*-acting glycosidases. In all cases, aziridines on cyclohexane rings preferably open in a *trans*-diaxial fashion through a $^4H_3$ transition state, and reaction with the enzymes nucleophile takes place at the cyclohexane anomeric equivalent center. This corresponds to ring opening to yield a $^4C_1$ chair conformation, as is observed in the trapped enzyme catalytic cleft in the co-crystals. ABPs bound to AcaGH79 and HPSE in a similar configuration in the active site involving the typical $^1S_3$-$^4H_3$-$^4C_1$ catalytic itinerary of β-glycosidases, and it is infered that GH2 GUSB complex with aziridine ABPs will show the same $^4C_1$ configuration. Last but not least, comparative and competitive ABPP experiments, for monitoring retaining *exo*- and *endo*-β-glucuronidases activities, and for the discovery of

selective inhibitors provide promising tools to study enzyme function *in vivo*, which may also contribute to opening avenues for treating glycosaminoglycan metobolism diseases.

7.4 Experimental section

Biological assays:

Materials

Cyclophellitol ABP **8** and inhibitors **9**, **10** and **11** were prepared as described earlier.[31] Siastain B was bought from Sigma. Human spleen, human Gaucher spleen and fibroblast lysates were prepared same as previous report.[10,19] Cell lines were cultured in HAMF12-DMEM medium (Invitrogen) supplied with 10% (v/v). Rabbit anti-heparanase 1 antibody (ab59787) was obtained from Abcam. Human platelet was from healthy blood donors, mixed with the anti-coagulent (EDTA or citrate) and centrifuged for 15 min at 230 g at 37 °C, transferred plasma containing the platelet to a 15 ml tube and centrifuged for 10 min at 2200 g at 37 °C, collected the enriched platelet pellet fraction and frozen them in small aliquots at -80 °C. All the tissue/cell lysates were prepared in lysis buffer (25 mM potassium phosphate buffer in pH 6.5, supplemented with protease inhibitor 1x cocktail (Roche)) via homogenization with silent crusher S equipped with Typ 7 F/S head (30,000 rpm, 3 × 7 sec) on ice and lysate concentration was determined with Qubit 2.0 Fluorometer assay (Invitrogen). The protein fractions were stored in small aliquots at -80 °C until use.

ABP pull-down and LC-MS/MS analysis

2.0 mg total protein from human fibroblast lysate or 3 mg total protein from human spleen or Gaucher spleen lysate was incubated with either 0.1% (v/v) DMSO, 10 µM **3**, or firstly with 10 µM **8** followed by 10 µM **3**, each step taking 30 min at 37 °C, in a total volume of 0.5 mL McIlvaine buffer, pH 5.0, subsequently denatured through the addition of 10% (w/v) SDS 125 µL and boiling for 5 min at 100 °C. From here on, samples were prepared for pull-down with streptavidin beads as published earlier.[20] After pull-down the samples were divided, 2/3 for on-bead digestion and 1/3 for in-gel digestion. On-bead digestion samples were treated by the trypsin digestion buffer (a mixture containing 100 mM Tris-HCl pH 7.8, 100 mM NaCl, 1.0 mM $CaCl_2$, 2% acetonitrile and 10 ng/µL trypsin) and the bead suspension was incubated in a shaker at 37 °C overnight. The supernatant containing the trypsin-digested peptides was desalted using stage tips, followed by evaporation of MeCN and dilution in 70 µL sample solution (H_2O/MeCN/TFA, 95/3/0.1, v/v/v) for LC-MS analysis. The beads containing active-site peptides were treated with endoproteinase Glu-C digestion buffer (100 ng/µL in PBS solution); incubated in a shaker at 37 °C overnight after which the supernatant was desalted using stage tips and for LC-MS. In-gel digestion samples were eluted by boiling the beads at 100 °C in 30 µL of 1x Laemmli buffer containing 10 µM biotin. The eluted proteins were separated on 10% protein gels at 200 V for 1 h, and the protein gels were silver stained using the Invitrogen kit,[4] and visualized by Bio-rad Chemi-Doc MP Imager using the silver stain channel. Bands were excised with a surgery knife by hand and treated with gel digestion buffer (10 mM NH_4HCO_3, 5% MeCN, 1.0 mM $CaCl_2$, 10 ng/µL trypsin). The supernatant containing the trypsin-digested peptides was desalted using stage tips and prepared for LC-MS. All the 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.[20]

Human HPSE gene cloning, expression and protein purification

Mature recombinant HPSE expression, purification was carried out as previously described strategy.[7]

Bacterial AcaGH79 gene cloning, expression and protein purification

The coding sequence of the AcGH79 gene was cloned into the pET28a (Novagen) 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 unitl the A600 nm reached 0.8. Expression of the recombinant proteins was induced by the addition of 1.0 mM isopropyl β-D-1-thiogalactopyranoside for 12 hr at 25 °C. The cells were harvested by centrifugation at 8000 ×g for 30 min and resuspended in 50 mL lysis buffer (20 mM HEPES, NaCl 200 mM, imidazole 5.0 mM, pH 7.0). After 20 min of sonication and 30 min of centrifugation at 12000 g, the filtered supernatant containing His_6-AcGH79 was loaded onto a His Trap column (GE Healthcare), equilibrated with the lysis buffer. The column was washed with lysis buffer and the His_6-AcGH79 protein was eluted with the same buffer with supplement of 400 mM imidazole over a gradient of 100 mL. The fractions containing the His_6-AcGH79 were then loaded onto a Hiload 16/60 Superdex 75 column (GE Healthcare). The fractions containing the His_6-AcGH79 were pooled and concentrated to the final concentration of 14.5 mg/mL.

***In vitro* labeling and SDS-PAGE analysis and fluorescence scanning**

For labeling procedures see below. All the labeling samples were samples resolved on 10% SDS-PAGE. Electrophoresis in sodium dodecylsulfate containing 10% polyacrylamide gels was performed as earlier described.[32] Wet slab-gels were then scanned for ABP-emitted fluorescence using a Bio-Rad ChemiDoc MP imager using green Cy2 (λ_{EX} 470 nm, bandpass 30 nm; λ_{EM} 530 nm, bandpass 28) for **2** and blue Cy5 (λ_{EX} 625 nm, bandpass 30 nm; λ_{EM} 695 nm, bandpass 55) for **8**. All samples were denatured with 5× 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), boiled for 4 min at 100 °C, and separated by gel electrophoresis on 10% (w/v) SDS-PAGE gels running continuously at 90 V for 30 min and 200 V for 50 min.

A) Human/Gaucher spleen extracts labeling: Human spleen lysates from control or Gaucher patient (20 μg total protein per sample) were accommodated in McIlvaine buffer pH 5.0 (10 μL total volume) for 5 min on ice. To selectively inhibit GBA or beta-glucuronide, samples were pre-incubated with 1.0 μM **8**, 1.0 μM **3**, or DMSO for 30 min at 37 °C, followed by 30 min incubation with 100 nM probe **2** or DMSO for another 30 min. For competitive ABPP using ctrl spleen to screen inhibitor selectivity, the samples were pre-incubated with 1.0 mM **7**, 100 μM **6**, **5**, 1.0 mM **11**, 100 μM **10** and 100 μM **9** for 30 min at 37°C, followed by 30 min incubation of 100 nM **2** or 1.0 μM **8** or probe mixture of 100 nM **2** and 1.0 μM **8** for another 30 min at 37 °C.

B) Mouse tissues extracts labeling: For direct labeling in mouse tissue extracts, 20 μg total protein per sample was accommodated in McIlvaine buffer pH 5.0 (10 μL total volume) for 5 min on ice and treated with 1.0 μM **2** (10 μL 2x solution in McIlvaine buffer). For ABPP on lysate of C57Bl6/J mouse liver, 50 μg in total was incubated with 1.0 μM **2** for 30 min in 150 mM McIlvaine buffer, pH 5.0, whilst at 37 °C.

C) Human recombinant HPSE labeling: All the recombinant HPSE labeling was performed by incubation of 200 ng HPSE with 100 nM **2** for 30 min at 37 °C. The influence of pH on ABP labeling was established by pre-incubation of 200 ng HPSE at pH 3.0–7.0 citric acid – Na_2HPO_4 buffer solutions for 5 min on ice, and incubation with 100 nM **2** for 30 min at 37 °C.

D) Human cell lines labeling: Direct labeling in overexpressed HEK293 cell lysate from different timepoints or internalized human fibroblast cell lysate (20 µg total protein per sample) were accommodated in 150 mM McIlvaine buffer pH 5.0 (10 µL total volume) for 5 min on ice and treated with 100 nM **2**, incubated for 30 min at 37 °C, followed by Western blot.

E) Human platelet extracts labeling: Direct labeling with ABP **2** of human platelet-EDTA or platelet-citric lysate (20 µg total protein per sample) was accommodated in 150 mM McIlvaine buffer pH 5.0 (10 µL total volume) for 5 min on ice and treated with 100 nM **2** for 30 min at 37 °C. ABP concentration and time limit labeling was analyzed by incubation of 20 µg total protein with 0–100 nM **22** in 150 mM McIlvaine buffer, pH 5.0, for 10 min, 30 min and 60min at 37 °C respectively. Influence of pH on ABP labeling was established by pre-incubation of 20 µg total protein at pH 3.0–7.0 for 5 min on ice and incubation with 100 nM **2** for 30 min at 37 °C. For competitive ABPP on platelet, 100 ng total protein was pre-incubated with decreasing amount of heparin (5.0 – 0.25 mg/mL) and then treated with 100 nM **2** for 30 min at 37 °C.

Western blot analysis

After SDS-PAGE analysis and fluorescence scanning, proteins were transferred from gel to a PVDF membrane using a Trans-Blot® Turbo system (Bio-Rad). Human HPSE was stained using rabbit anti-heparanase 1 antibody as primary antibody, and goat-anti-rabbit HRP as secondary antibody. The blot was developed in the dark using a 10 mL luminal solution, 100 µL ECL enhancer and 3.0 µL 30% H_2O_2 solution. Chemiluminescence was visualized using a ChemiDoc XRS (Bio-Rad).

Overexpression of HPSE in HEK293T cells

pGEn1-HPSE and pGEn2-HPSE plasmids were obtained from the DNASU repository.[33] HEK293T cells were grown in DMEM media supplemented with 10% newborn calf serum (NBCS). 20 µg DNA was transfected into low passage HEK293T cells at ~80% confluency using PEI at a ratio of 3:1 (PEI : DNA). After 7 h, 24 h, 48 h and 72 h cells were washed with ice cold PBS, harvested using a cell scraper, and pelleted. Cell pellets were stored at -80 °C prior to use in labeling experiments.

proHPSE fibroblast internalized experiment

Primary human fibroblasts were grown in F12 media supplemented with 10% NBCS and penicillin/streptomycin. Cells at 70-80% confluency were washed with PBS, before addition of F12 media supplemented with proHPSE at 10 µg/mL final concentration. At relevant time points, cells were washed with ice cold PBS twice, harvested using a cell scraper, and pelleted. Cell pellets were stored at -80 °C prior to use in labeling experiments.

proHPSE bacmid preparation and protein production

Baculovirus encoding for N-6xHis-TEV-proHPSE was sub-cloned from the pGEn1-HPSE plasmid. cDNA encoding for the mellitin signal peptide and TEV-proHPSE were extracted by PCR, along with synthetic DNA encoding for 6xHis-TEV. DNA fragments were ligated using T4 DNA ligase (NEB), before cloning into a pOMNI vector (Geneva biotech) using SLIC.[34] All PCR steps were carried out using Phusion polymerase (Thermo). Primers and synthetic DNA sequences are listed in (Table 1). Subsequent bacmid and expression steps were carried out as described for HPSE.

Table 1 Primers used for cloning proHPSE construct.

pOMNI-Mellitin F primer	ccatcgggcgcggatccatgaaatttttggtg
Mellitin-XmaI R primer	gatcgcccgggtccgcataaatgtagctaatg
XmaI-6xHis-KpnI F sequence	ctagcccgggccatcatcaccaccatcatggtaccgatc
XmaI-6xHis-KpnI R sequence	gatcggtaccatgatggtggtgatgatggcccgggctag
KpnI-TEV-proHPSE F primer	gatcggtaccgcagaaaacttgtactttcaaggccagg
proHPSE-pOMNI R primer	gtacttctcgacaagctttcagatgcaagcagcaactttg

Purification of pro-HPSE

3.0 L of conditioned media was cleared of cells by centrifugation at 400 g for 15 min at 4 °C, followed by further clearing of debris by centrifugation at 4000 g for 60 min at 4 °C. DTT (1.0 mM) and AEBSF (0.10 mM) were added to cleared media, which was then loaded onto a pre-equilibrated HiTrap Sepharose SP FF 5.0 mL column (GE healthcare). The SP column was washed with 10 CV of IEX buffer A (20 mM HEPES (pH = 7.4), 100 mM NaCl, 1.0 mM DTT), and eluted with a linear gradient over 30 CV using IEX buffer B (20 mM HEPES (pH = 7.4), 1.5 mM NaCl, 1.0 mM DTT). proHPSE containing fractions were pooled and diluted 10 fold into HisTrap buffer A (20 mM HEPES (pH = 7.4), 500 mM NaCl, 20 mM imidazole, 1.0 mM DTT), before loading onto a pre-equilibrated HisTrap FF 1.0 mL column (GE healthcare). The HisTrap column was washed with 10 CV HisTrap buffer A, and eluted with a linear gradient over 30 CV using HisTrap buffer B (20 mM HEPES (pH = 7.4), 500 mM NaCl, 1.0 M imidazole, 1.0 mM DTT). proHPSE containing fractions were pooled and concentrated to ~2.0 mL using a 30 kDa cutoff Vivaspin concentrator (GE Healthcare), and treated with 5.0 µL EndoH (NEB) and 5.0 uL AcTEV protease (Invitrogen) for 4 h. Digested protein was purified by size exclusion chromatography (SEC) using a Superdex S75 16/600 column (GE Healthcare) in SEC buffer (20 mM HEPES (pH = 7.4), 200 mM NaCl, 1.0 mM DTT). proHPSE containing fractions were concentrated to 1.0 mg/mL using a 30 kDa Vivaspin concentrator, and flash frozen for use in further experiments.

HPSE-4 complex crystallography and structure solution

HPSE complex with **4** was generated by soaking the *apo* crystal in 250 µM **4** in mother liquor (100 mM MES (pH = 5.5), 100 mM $MgCl_2$, 17% PEG3350). Crystals were soaked overnight at room temperature, transferred to cryoprotectant solution (100 mM MES (pH = 5.5), 100 mM $MgCl_2$, 17% PEG3350, 25% DMSO), and subsequently flash frozen in liquid N_2 for data collection. Complexes were solved by refining directly against the *apo* structure using REFMAC5,[35] before further rounds of manual model building and refinement using Coot[36] and REFMAC5. Ligand coordinates were built using jLigand.[37] Crystal structure figures were generated using ccp4mg.[38]

7.5 References

[1] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490-D495.

[2] L. Arul, G. Benita, D. Sudhakar, B. Thayumanavan, P. Balasubramanian, *Bioinformation* **2008**, *3*, 194-197.

[3] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416-436.

[4] B. Henrissat, G. J. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637-644.

[5] W. S. Sly, Vogler, C. *Nat. Med.* **1997**, *3*, 719-720.

[6] M. Michikawa, H. Ichinose, M. Momma, P. Biely, S. Jongkees, M. Yoshida, T. Kotake, Y. Tsumuraya, S. G. Withers, Z. Fujimoto, S. Kaneko, *J. Bio. Chem.* **2012**, *287*, 14069-14077.

[7] L. Wu, C. M. Viola, A. M. Brzozowski, G. J. Davies, *Nat. Struct. Mol. Biol.* **2015**, 22, 1016–1022

[8] I. Vlodavsky, P. Beckhove, I. Lerner, C. Pisano, A. Meirovitz, N. Ilan, M. Elkin, *Cancer Microenvironment* **2012**, *5*, 115-132.

[9] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383-414.

[10] W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts, H. S. Overkleeft, *Angew. Chem. Int. Ed.* **2012**, *51*, 12529-12533.

[11] J. Jiang, T. J. M. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codee, J. M. F. G. Aerts, H. S. Overkleeft, *Chem. Eur. J.* **2015**, *21*, 10861-10869.

[12] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, B. I. Florea, J. D. C. Codee, G. A. van der Marel, J. M. F. G. Aerts, H. S. Overkleeft, *J. Am. Chem. Soc.* **2014**, *136*, 11622-11625.

[13] J. Jiang, W. W. Kallemeijn, D. W. Wright, A. M. C. H. van den Nieuwendijk, V. C. Rohde, E. C. Folch, H. van den Elst, B. I. Florea, S. Scheij, W. E. Donker-Koopman, M. Verhoek, N. Li, M. Schurmann, D. Mink, R. G. Boot, J. D. C. Codee, G. A. van der Marel, G. J. Davies, J. M. F. G. Aerts, H. S. Overkleeft, *Chem. Sci.* **2015**, *6*, 2782-2789.

[14] R. Petryszak, T. Burdett, B. Fiorelli, N. A. Fonseca, M. Gonzalez-Porta, E. Hastings, W. Huber, S. Jupp, M. Keays, N. Kryvych, J. McMurry, J. C. Marioni, J. Malone, K. Megy, G. Rustici, A. Y. Tang, J. Taubert, E. Williams, O. Mannion, H. E. Parkinson, A. Brazma, *Nucleic Acids Res.* **2014**, *42*, D926-D932.

[15] M. Ono, N. Taniguchi, A. Makita, M. Fujita, C. Sekiya, M. Namiki, *J. Biol. Chem.* **1988**, *263*, 5884-5889.

[16] M. R. Islam, J. H. Grubb, W. S. Sly, *J. Biol. Chem.* **1993**, *268*, 22627-22633.

[17] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, B. Hooibrink, C. E. M. Hollak, R. Ottenhoff, R. G. Boot, G. A. van der Marel, H. S. Overkleeft, J. M. F. G. Aerts, *Nat. Chem. Biol.* **2010**, *6*, 907-913.

[18] R. G. Boot, C. E. M. Hollak, M. Verhoek, P. Sloof, B. J. H. M. Poorthuis, W. J. Kleijer, R. A. Wevers, M. H. J. van Oers, M. M. A. M. Mannens, J. M. F. G. Aerts, S. van Weely, *Human Mutant.* **1997**, *10*, 348-358.

[19] N. Li, C.-L. Kuo, G. Paniagua, H. van den Elst, M. Verdoesm, L. I. Willems, W. A. van der Linden, M. Ruben, E. van Genderen, J. Gubbens, G. P. van Wezel, H. S. Overkleeft, B. I. Florea, *Nat. Protocols* **2013**, *8*, 1155-1168.

[20] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, *Mol. Cell Proteomics* **2005**, *4*, 1265-1272.

[21] J. J. Birktoft, K. Breddam, *Methods Enzymol.* Academic Press: **1994**, *244*, 114-126.

[22] M. B. Fairbanks, A. M. Mildner, J. W. Leone, G. S. Cavey, W. R. Mathews, R. F. Drong, J. L. Slightom, M. J. Bienkowski, C. W. Smith, C. A. Bannow and R. L. Heinrikson, *J.Biol. Chem.* **1999**, *274*, 29587-29590.

[23] L. Nadav, A. Eldor, O. Yacoby-Zeevi, E. Zamir, I. Pecker, N. Ilan, B. Geiger, I. Vlodavsky, B. Z. Katz, *J. Cell Sci.***2002**, *115*, 2179-2187.

[24] I. Shafat, A. B. Barak, S. Postovsky, R. Elhasid, N. Ilan, I. Vlodavsky and M. W. Arush, Neoplasia **2007,** *9*, 909-916.

[25] Y. Nishimura, *J. Antibiot.* **2009**, *62*, 407-423.

[26] G. J. Davies, A. Planas, C. Rovira, *Accounts Chem. Res.* **2012**, 45, 308-316.

[27] A. Ardevol, C. Rovira, *J. Am. Chem. Soc.* **2015**, *137*, 7528-7547.

[28] B. D. Wallace, H. Wang, K. T. Lane, J. E. Scott, J. Orans, J. S. Koo, M. Venkatesh, C. Jobin, L. A. Yeh, S. Mani, M. R. Redinbo, *Science* **2010**, *330*, 831-835.

[29] P. Durand, P. Lehn, I. Callebaut, S. Fabrega, B. Henrissat, J. P. Mornon, *Glycobiology* **1997**, 7, 277-284.

[30] B. D. Wallace, A. B. Roberts, R. M. Pollet, J. D. Ingle, K. A. Biernat, S. J. Pellock, M. K. Venkatesh, L. Guthrie, S. K. O'Neal, S. J. Robinson, M. Dollinger, E. Figueroa, S. R. McShane, R. D. Cohen, J. Jin, S. V. Frye, W. C. Zamboni, C. Pepe-Ranney, S. Mani, L. Kelly, M. R. Redinbo, *Chem. Biol.* **2015**, 22, 1238-1249.

[31] K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel, H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, *2014*, 6030-6043.

[32] J. M. Aerts, A. W. Schram, A. Strijland, S. van Weely, L. M. Jonsson, J. M. Tager, S. H. Sorrell, E. I. Ginns, J. A. Barranger, G. J. Murray, *Biochim. Biophys. Acta* **1988**, *964*, 303-308.

[33] C. Y. Seiler, J. G. Park, A. Sharma, P. Hunter, P. Surapaneni, C. Sedillo, J. Field, R. Algar, A. Price, J. Steel, A. Throop, M. Fiacco, J. LaBaer, *Nucleic Acids Res.* **2014**, *42*, D1253-D1260.

[34] M. Z. Li, S. J. Elledge, *Methods Mol. Biol.* **2012**, *852*, 51-59.

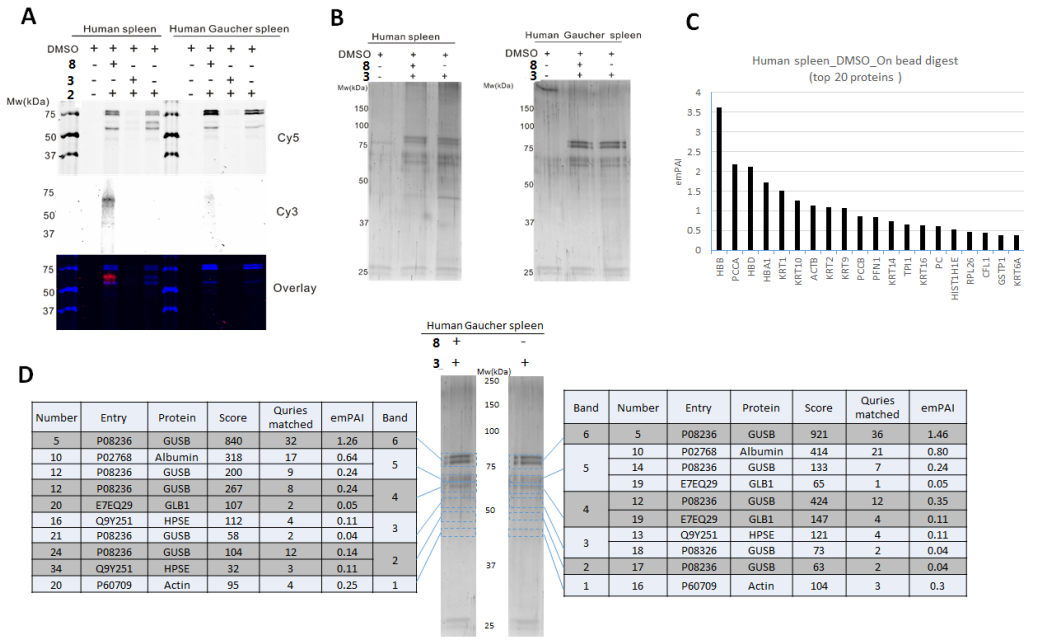
[35] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D* **2011** *67*, 355-367.

[36] P. Emsley, K. Cowtan, *Acta Crystallogr. D* **2004**, *60*, 2126-2132.

[37] A. A. Lebedev, P. Young, M. N. Isupov, O.V. Moroz, A. A. Vagin, G. N. Murshudov, *Acta Crystallogr. D* **2012**, *68*, 431-440.

[38] S. McNicholas, E. Potterton, K. S. Wilson, M. E. Noble, *Acta Crystallogr. D* **2011**, *67*, 386-394.

7.7 Supporting Information

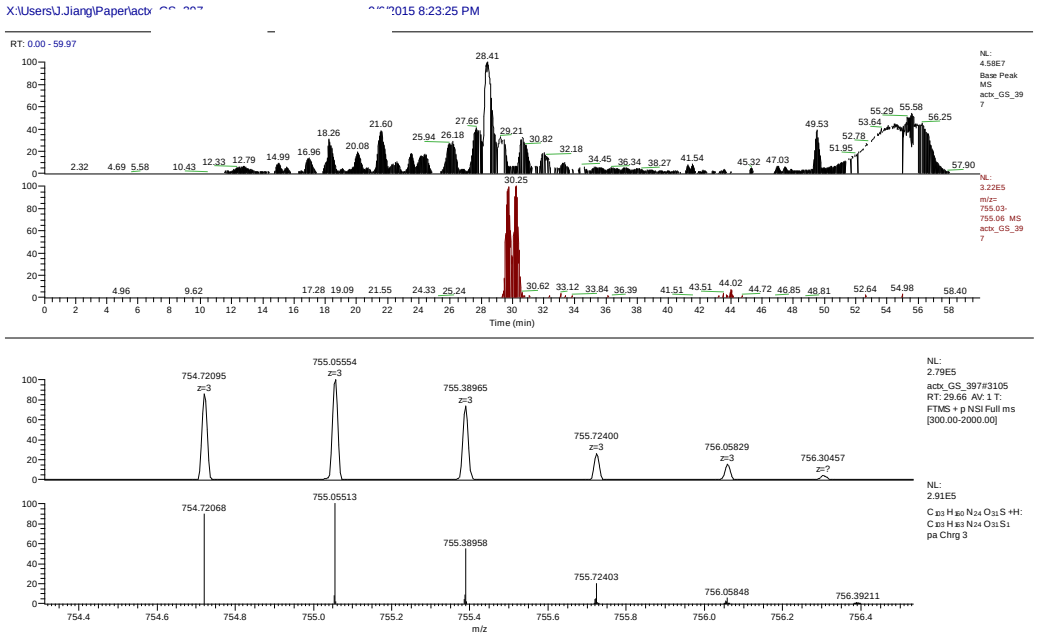


D

Number	Entry	Protein	Score	Quries matched	emPAI	Band
5	P08236	GUSB	840	32	1.26	6
10	P02768	Albumin	318	17	0.64	5
12	P08236	GUSB	200	9	0.24	
12	P08236	GUSB	267	8	0.24	4
20	E7EQ29	GLB1	107	2	0.05	
16	Q9Y251	HPSE	112	4	0.11	3
21	P08236	GUSB	58	2	0.04	
24	P08236	GUSB	104	12	0.14	2
34	Q9Y251	HPSE	32	3	0.11	
20	P60709	Actin	95	4	0.25	1

Band	Number	Entry	Protein	Score	Quries matched	emPAI
6	5	P08236	GUSB	921	36	1.46
5	10	P02768	Albumin	414	21	0.80
	14	P08236	GUSB	133	7	0.24
	19	E7EQ29	GLB1	65	1	0.05
4	12	P08236	GUSB	424	12	0.35
	19	E7EQ29	GLB1	147	4	0.11
3	13	Q9Y251	HPSE	121	4	0.11
	18	P08326	GUSB	73	2	0.04
2	17	P08236	GUSB	63	2	0.04
1	16	P60709	Actin	104	3	0.3

E



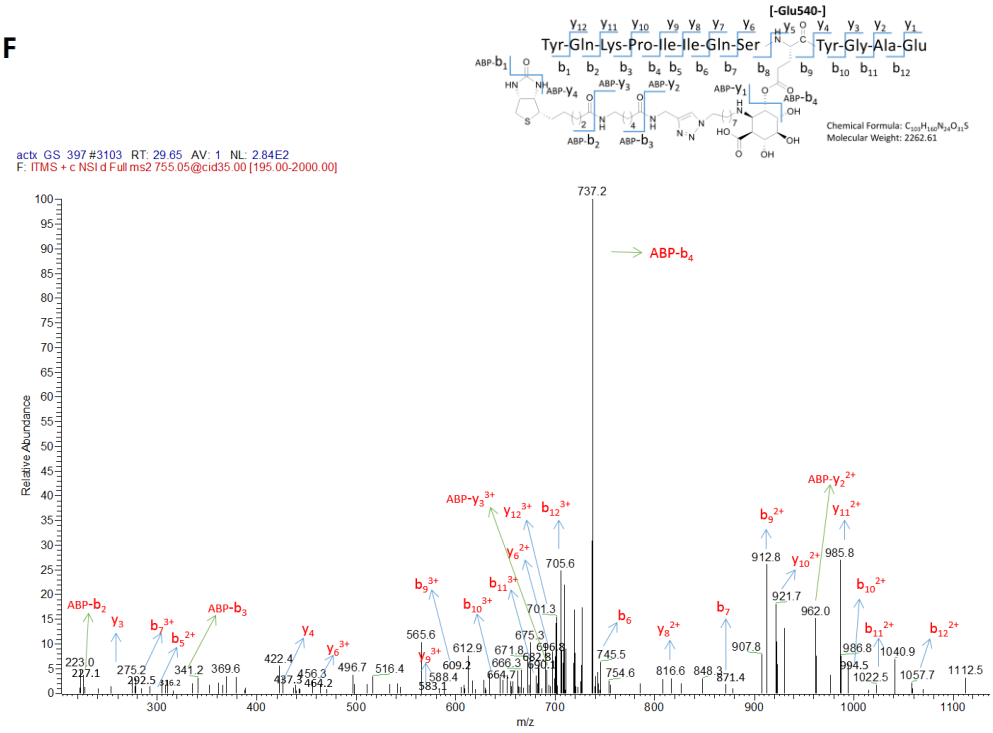


Figure S1. A) Human spleen and human Gaucher spleen extracts labeled with ABP **3** and with or without pre-incubation with **8**, independently shown in Cy5 and Cy3 fluorescent channels. B) Silver stain gel of human spleen and human Gaucher spleen extracts after pull-down with ABP **3** without pre-incubation with **8**; these gels were used for in-gel digestion. C) In-gel trypsin digestion proteins table from human Gaucher spleen extracts involving target glycosidases and off-target proteins. D) Human spleen incubation with DMSO sample after pull-down on bead digest results (emPAI values) of captured proteins. E) Electrospray ionization spectra of ABP **3** modified GUSB nucleophile active-site peptide in 2 charge statement. F) Identification of the GUSB active site fragment modified with biotin-ABP **3** after on-bead capture, trypsinolysis and LC-MS/MS analysis of the tryptic fragment containing the modified active site nucleophile.

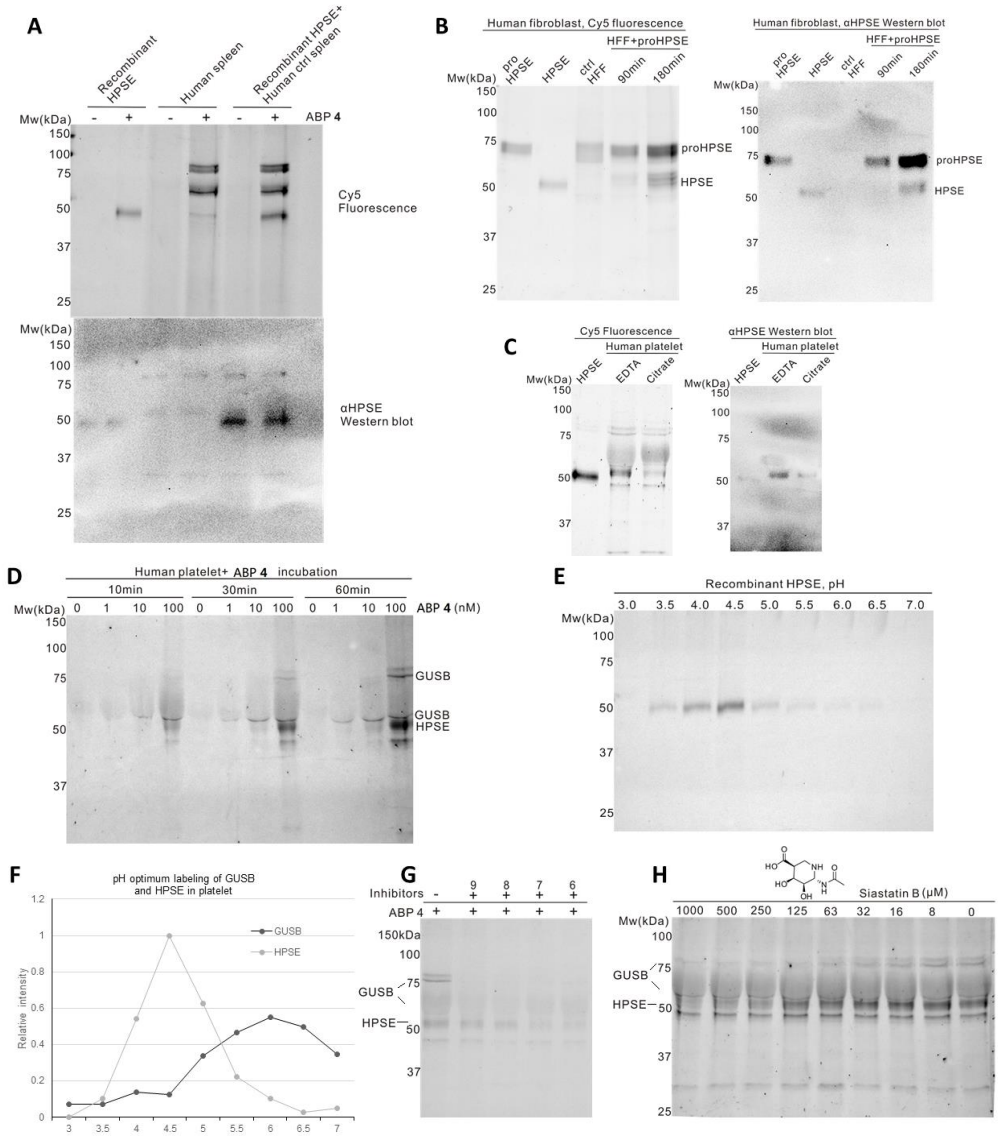


Figure S2. A) Recombinant HPSE, human spleen extracts and their mixture labeled with ABP **2**. SDS-PAGE of Cy5 fluorescent channel is shown on the left and Western blot of PVDF membrane scanning of chemiluminescence channel is shown on the right. B) ABP **2** labels internalized HPSE in human fibroblast (HF). Fibroblast cells are able to internalize and process proHPSE in culture media. ABP labeling intensity reflects intracellular accumulation of HPSE over time. C) Endogenous HPSE labeling from human platelet EDTA and citrate buffer by ABP **2**. SDS-PAGE of Cy5 fluorescent channel is shown on the left and Western blot of PVDF membrane scanning of chemiluminescence channel is shown on the right. D) Time- and concentration-dependent labeling of human platelet extract by ABP **2** as visualized by SDS-PAGE and Western blot. E) pH optimum labeling of human platelet GUSB and HPSE, with maximum labeling for GUSB at pH 5.5-6.5, and that of HPSE at pH 4.0-5.0. F) Detection of recombinant HPSE labeling efficiency at various pH. G) HSPE labeling in platelets is abolished by pre-incubation with known inhibitor Siastatin B.

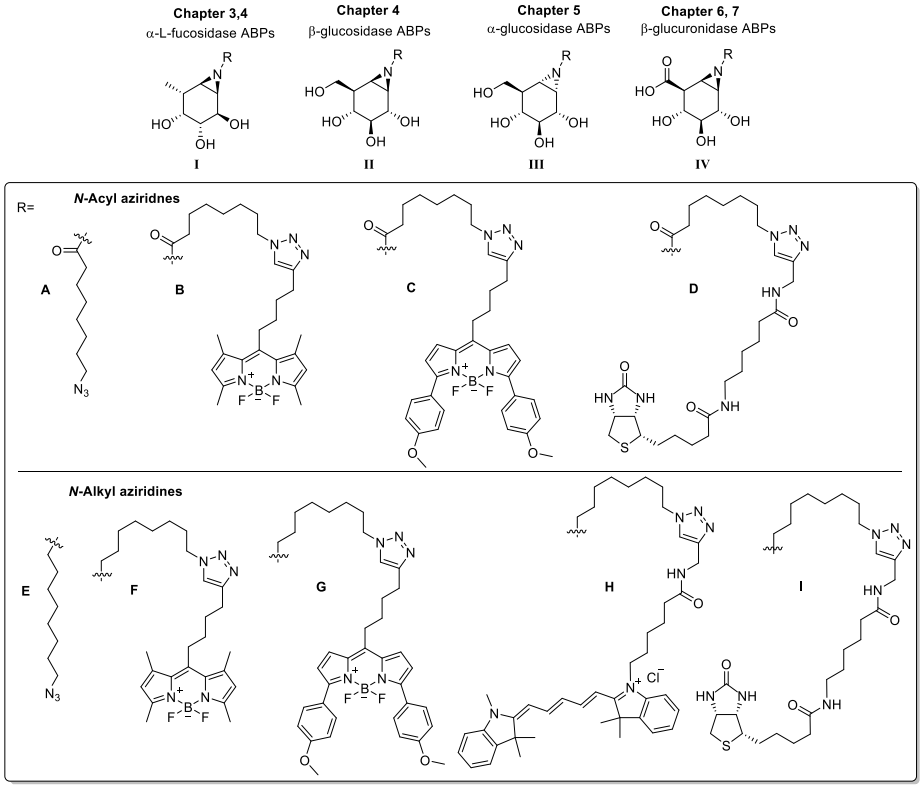
8

Summary and future prospects

8.1 Summary

This thesis describes the development of functional and configurational analogues of cyclophellitol aziridine as activity-based probes (ABPs) for various retaining glycoside hydrolases (GHs), namely α-L-fucosidases, β-glucosidases, α-glucosidases and β-glucuronidases (Figure 1). Attention is focused on the design and synthesis of the cyclophellitol aziridine derivatives and their application in chemical biology studies of various retaining GHs.

Chapter 1 introduces the research subject described in this thesis. Subjects introduced in this chapter include β-glucosidases classification, a description of the molecular mechanisms employed by these enzymes in the hydrolysis of their substrates, as well as the cyclophellitol aziridine-based ABPs targeting retaining β-glucosidases. The retaining glycosidases employ Koshland double displacement mechanism,[1] and on the basis of which the research of this thesis was formulated. **Chapter 2** reviews the literature on the synthesis of cyclitol aziridines as the scaffold of choice for the development of retaining GHs ABPs.



ABPs	A	B	C	D	E	F	G	H	I
I	JJB237	JJB256	JJB244	JJB243	JJB349	JJB350	JJB380	JJB381	JJB385
II	JJB103	JJB70	JJB75	JJB111	JJB339	JJB376	JJB343	JJB367	JJB377
III					CF022	JJB382	JJB347	JJB383	JJB384
IV	JJB133		JJB144	JJB249	JJB355	JJB395	JJB391	JJB392	JJB397

Figure 1. Activity-based glycosidase probes (ABPs) discussed in this thesis. **I**-**IV** are configurational and functional cyclophellitol aziridine isomers emulating the structure of the parent monosaccharides as indicated; **A**-**I** are acyl and alkyl functional tags.

Retaining α-L-fucosidase (FUCA) is a member of the glycoside hydrolases family 29 (GH29). FUCA deficiency is at the basis of the rare lysosomal storage disorder, fucosidosis. **Chapter 3** describes an activity-based protein profiling (ABPP) study on GH29 α-L-fucosidases. L-Fucopyranose-configured cyclophellitol aziridines (JJB237, JJB243, JJB244 and JJB256), are applied as ABPs for selective *in vitro* and *in vivo* labeling of GH29 α-L-fucosidases from bacteria, mice and man. The chapter also reports on the synthesis of eight configurational L-fuconojirimycin isomers, which were screened as potential α-L-fucosidase inhibitors in a competitive ABPP setting. The covalent aziridine-enzyme adduct between the carbon

corresponding to the anomeric center in the substrate fucoside and the FUCA active site nucleophile was trapped in a co-crystal of *Bacteroides thetaiotaomicron* α-L-fucosidase treated with *N*-acetyl-L-*fuco*-cyclophellitol aziridine. The *trans*-diaxial skew boat conformation adopted by the covalently attached inhibitor also provides insight in the conformational changes substrate α-L-fucosides undergo during FUCA-mediated hydrolysis.

In order to improve the stability and availability of retaining GHs ABPs, a set of next generation probes, namely *N*-alkyl cyclophellitol aziridine ABPs, were investigated. The synthesis and evaluation of fluorescent *N*-alkyl probes directed at GH30 β-glucosidases (JJB339, JJB343) and GH29 α-L-fucosidases (JJB349, JJB380) are described in **Chapter 4**. In comparison with the corresponding acyl aziridine ABPs reported previously, the alkyl aziridine ABPs proved relatively easy to synthesize and are more stable in mildly acidic and basic media. The β-*gluco*-configured alkyl aziridine ABPs proved equally effective in inhibiting and labeling the lysosomal β-glucosidase (GBA) as its *N*-acyl counterparts. In contrast, the *N*-acyl aziridines targeting α-L-fucosidase outperform their *N*-alkyl counterparts. Therefore, *N*-alkyl cyclophellitol aziridines can be an attractive alternative in retaining GHs ABP design, but in targeting a new retaining glycosidase both *N*-alkyl and *N*-acyl aziridines are best considered at the onset of a new study.

Chapter 5 describes the development of ABPP technology to study GH31 α-glucosidases *in vitro* and *in situ*. To this end a comprehensive set of α-*gluco*-cyclophellitol aziridines bearing either a fluorescent group (JJB347, JJB382, JJB383) or a biotin (JJB384) was synthesized. α-Glucosidases are involved in diverse physiological processes in the human body, including carbohydrate assimilation in the gastrointestinal tract, glycoprotein processing in the endoplasmic reticulum (ER), and intra-lysosomal glycogen catabolism. Inherited deficiency of the lysosomal acid α-glucosidase (GAA) causes Pompe disease, a relatively common lysosomal glycogen storage disorder. The developed ABPs proved to be highly potent and irreversible inhibitors towards recombinant α-glucosidase as established by enzyme inhibition assays and X-ray crystallography analysis. Moreover the ABPs can specifically label distinct retaining GH31 α-glucosidases, notably, the lysosomal GAA and the ER α-glucosidase II and this labelling can be tuned by pH. The chapter further describes a direct diagnostic application in Pompe disease patient fibroblast cells, and reports on the analysis of intestinal dietary α-glucosidases, such as sucrase-isomaltase (Sis) and maltase-glycoamylase (MGAM).

Chapter 6 describes a synthesis strategy towards β-glucuronide-configured cyclophellitol, cyclophellitol aziridine and its derivatives (JJB133, JJB144, JJB249, JJB355, JJB391, JJB392, JJB395 and JJB397) as ABPs for GH2 and GH79 β-glucuronidases. The former enzyme is related to the glycosaminoglycan (GAG) storage disorder, Sly disease, and the latter enzyme is related to inflammation, tumour angiogenesis and cell migration. Uronic *N*-alkyl cyclophellitol

aziridines are easier to be prepared than uronic *N*-acyl cyclophellitol aziridines, and both alkyl and acyl inhibitors and probes proved equally effective in inhibition and labelling of β-glucuronidases, which is in line with the findings in **Chapter 4**. Crystallographical analysis on bacterial β-glucuronidase (AcaGH79) complexed with alkyl and acyl aziridines further confirmed the covalent modification of the enzyme nucleophile active site and showed a $^4C_1$ chair conformation for both enzyme-inhibitor adducts. Applications of the β-glucuronide-configured *N*-alkyl cyclophellitol aziridine ABPs are described in **Chapter 7** and include GH2 human lysosomal β-glucuronidase (GUSB) and GH79 human heparanase (HPSE) identification by activity-based proteomics. An interesting finding is that the *endo*-β-glucuronidase HPSE is effectively labeled and thus inhibited by cyclophellitol aziridine derivatives, which in essence are monosaccharide mimetics (thus expected to bind efficiently to *exo*-glycosidases, but not necessarily *endo*-glycosidases). Crystallographical analysis of the structures of recombinant *exo*-GUSB, *exo*-AcaGH79 and *endo*-HPSE as well as their corresponding nucleophile mutant (Glu to Ala) complexed with cyclophellitol aziridine JJB355, reveals the Koshland double displacement mechanism employed by these enzymes and $^1S_3$-$^4H_3$-$^4C_1$ itineraries substrate β-glucuronides undergo when processed.

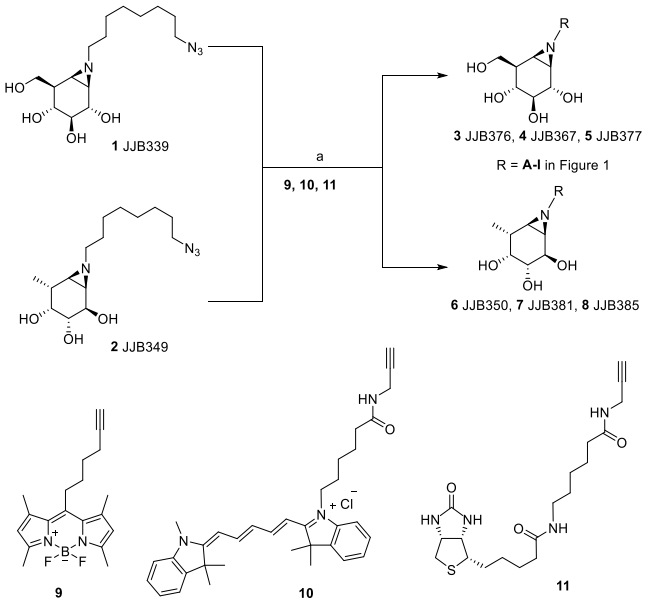
8.2 Future prospects

The research described in this thesis entails the design, synthesis and application of functional configurational cyclophellitol aziridine derivatives as ABPs selective for in-class GH family retaining glycosidases. Besides red BODIPY probes of β-glucosidases and α-L-fucosidases shown in Chapter 3, also green BODIPY probes **3** JJB376, **6** JJB350, blue Cy5 probes **4** JJB367, **7** JJB381 and biotin probes **5** JJB377, **8** JJB385 ABPs (Figure 1) can be prepared following a similar strategy (Scheme 1).[2]

By altering either configuration or substitution pattern, or both, selectivity of the resulting probes changes in a highly predictable manner. This bodes well for the development of ABPs directed at other retaining glycosidases. For instance, ABPs **12**-**14** (Figure 2) may target with good efficiency their respective underlying glycosidases: β-glucosidases (already a proven fact), β-galactosidases and β-mannosidases, which are of interest in the context of specific lysosomal storage disorders: Gaucher disease, GM1 gangliosidosis and mannosidosis respectively.[3] In comparison with ABPP technology developed for serine hydrolases and cysteine proteases, however, the in-class GH family selectivity displayed by the ABPs described in this thesis may also be considered a disadvantage. Cysteine protease probes and especially serine hydrolase probes have been reported that efficiently label a large number – sometimes hundreds of – related enzymes, which may be of advantage in for instance competitive ABPP experiments aimed at the discovery of selective inhibitors. Arguably, the nature (substitution pattern, configuration) of the highly functionalized cyclitol aziridines (Figure 1) determines their glycosidase specificity, and deleting some substituents may yield probes that, though less

potent, are also less selective, in other words, more broad-spectrum. For instance, diol cyclohexane aziridine **15** (Figure 2) may be developed as broad spectrum ABP for retaining β-glucosidases, β-galactosidases and β-mannosidases based on the idea that the hydroxyl at C2 (distinguishing β-glucose from β-mannose) and C4 (distinguishing β-glucose from β-galactose) are removed.

Scheme 1. Synthesis of alkyl aziridine ABPs for β-glucosidases and α-L-fucosidases.



Reagent and conditions: (a) $CuSO_4$(1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), DMF, **3** JJB376: 30%, **4** JJB367: 13%, **5** JJB377: 34%, **6** JJB350: 8%, **7** JJB381: 13%, **8** JJB385: 20%.

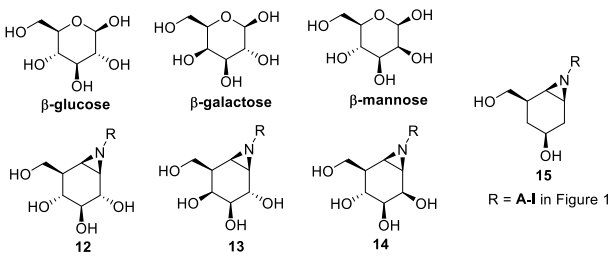
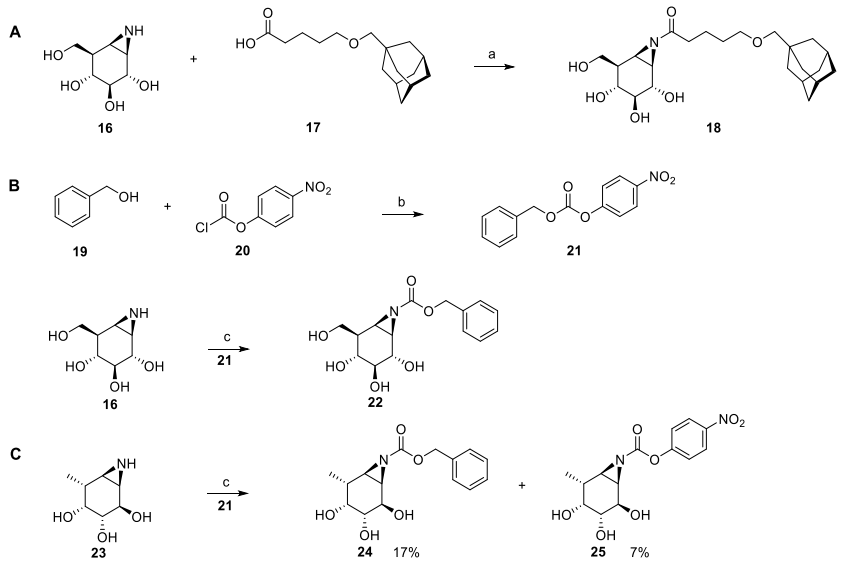


Figure 2. Chemical structures of β -glucose, β-galactose, β-mannose, ABPs **12-14** for β-glucosidases, β-galactosidase, β-mannosidase and broad spectrum ABP **15**.

Activity-based protein profiling is a powerful technique both to discover new enzyme activities (comparative ABPP) and to study their expression, activity and sensitivity in a tissue- and condition-dependent setting (competitive ABPP). The research described in this thesis focused predominantly on the latter, with probes designed with a specific GH glycosidase family in mind. Research in Chapter 5, however also demonstrates the potential of biotin-modified cyclophellitol aziridines to identify, by means of chemical proteomics, enzyme activities in a tissue-dependent fashion. In line with this, and underscored in a literature study on the application of β-glucopyranose-configured cyclophellitol aziridines to annotate *Arabidopsis* retaining β-glucosidases,[4] a comparative ABPP investigation can be envisaged in which all biotinylated cyclophellitol aziridine probes (Figure 1) are employed to screen tissue from various kingdoms and map their respective retaining glycosidase activities.

The research described in this thesis reveals that cyclophellitol aziridines are highly potent irreversible retaining glycosidase inhibitors, and that the inhibitory potency can be tuned by varying the nature of the aziridine *N*-substituent. For instance (see **Chapter 3**), *N*-acetyl-cyclophellitol aziridine (IC_{50} = 46.8 nM) inhibits recombinant FUCA1 about 8 times more potently than the corresponding *N*-benzoyl-cyclophellitol aziridine (IC_{50} = 371.6 nM). However, previously described in the literature, that *N*-alkylated derivatives of the natural product deoxynojirimycin (DNM) and in which the nitrogen substituent is large and hydrophobic, are considerably more potent GBA inhibitors than analogous derivatives that bear a small nitrogen substituent. For instance, *N*-(adamantanemethyloxypentyl)-deoxynojirimycin (AMP-DNM) is a considerably more potent GBA inhibitor than DNM as well as *N*-butyl-DNM.[5] In this light, it is of interest to explore cyclophellitol aziridine derivatives bearing as nitrogen substituent a variety of alkyl and acyl substituents. Scheme 2 describes the synthesis of several of these compounds, namely, *N*-adamantane-methoxypentanoyl aziridine **18** and *N*-carboxybenzyl aziridine **22** as potential GBA inhibitors and two *N*-substituted L-*fuco*-cyclophellitol aziridine inhibitors (Cbz-aziridine **24** and *p*-nitrophenyloxycarbonyl aziridine **25**) as potential FUCA inhibitors. Compound **18** was prepared with the appropriate acid under the agency of EEDQ as the condensation agent following a procedure as presented in **Chapter 3** to yield JJB237. Unprotected aziridines **16** and **23** were transformed in the corresponding carboxybenzyl derivatives via activated anhydride **21** yielding **22** and **24**. Interestingly, during the synthesis of fucose analogue **24**, compound **25** was also isolated after HPLC purification using a neutral eluent in a yield of 7%.

Scheme 2. Synthesis of acyl cyclophellitol aziridines **18**, **22, 24** and **25**.



Reagent and conditions: (a) EEDQ, DMF, 0 °C, 22%; (b) pyridine, DCM; (c) Et_3N, DMF, **22**: 42%, **24**: 17%, **25**: 7%.

The potency and selectivity of cyclophellitol and cyclophellitol aziridine as retaining β-glucosidase inhibitors can be attributed to their tight initial binding to the enzyme active site. Upon binding, the electrophilic carbon corresponding to C1 of the substrate glucoside is ideally positioned for reaction with the enzyme active site nucleophile, leading to covalent and irreversible modification of the enzyme active site. Glycosidase specificity is guided by the configuration and substitution pattern of the cyclophellitol derivative, as is amply demonstrated in this thesis. Looking at the conformation of cyclophellitol in comparison with that of a substrate β-glucoside, one could argue that initial binding happens, not so much as a substrate analogue, but as a transition analogue. Whereas β-glucopyranosides adopt a preferential $^4C_1$ chair conformation of Michaelis complex (Figure 3A), cyclophellitol prefers to adopt $^4H_3$ half-chair conformation (Figure 3B).

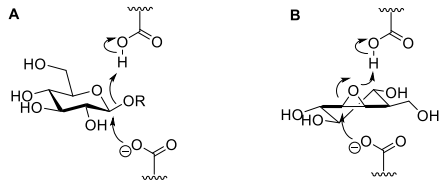
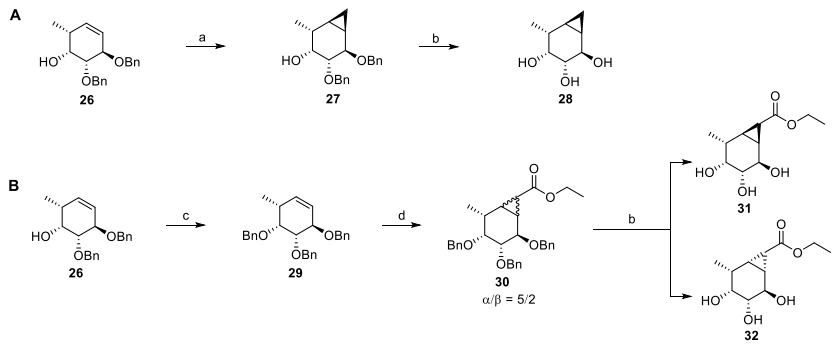


Figure 3. A) β-Glucopyranosides complex and B) cyclophellitol Michaelis complex conformations in enzyme catalysis.

Arguably, this conformation emulates that of an emerging oxocarbenium ion that may be formed as a transient intermediate in the enzyme active site and onto which the active site nucleophile will add. In a similar way, the configurational and substitutional isomers of cyclophellitol aziridine may resemble in conformation transition state oxocarbenium ions more than substrate glycosides of their underlying corresponding carbohydrates. Following this reason, cyclophellitol derivatives featuring the same substitution pattern and configuration, that are able to adopt a similar conformation but without an appropriate leaving group may turn out to be effective competitive inhibitors. With this reasoning in mind, *carba*-cyclophellitol analogues, with the epoxide oxygen substituted for methylene, can be proposed as a new class of glycosidase inhibitors.[6] Scheme 3 represents the synthesis of compounds **28**, **31** and **32**, being *carba*-cyclophellitol analogues featuring α-L-fucopyranose (**28**, **31**) and β-L-fucopyranose (**32**) configurations. Cyclopropanation of alkene **26** with diethylzinc/diiodomethane yielded cyclopropane **27** as the major stereoisomer, which was converted to α-cyclitol-cyclopropane **28** by palladium hydroxide catalyzed hydrogenolysis of the benzyl protective groups.

Scheme 3. Synthesis of L-*fuco*-carba-cyclophellitol compounds **28**, **31** and **32**

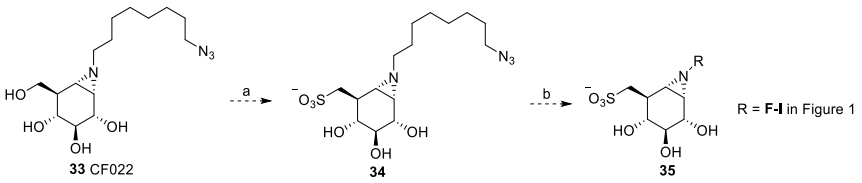


Reagent and conditions: (a) Et_2Zn, $BF_3{\cdot}O\,Et_2$, CH_2I_2, Et_2O, DCM, 83%; (b) $Pd(OH)_2/C$, H_2, MeOH, **28**: 99%, **31** 26%, **32** 11% over two steps; (c) BnBr, NaH, TBAI, DMF, 0 °C→20 °C, 95%; (d) $Cu(AcAc)_2$, Ethyl diazoacetate, EtOAc.

Substituted cyclopropane derivatives **31** and **32** were prepared as follows. Benzylation of C4-OH in **26** yielded **29**, which was treated with copper (II) acetylacetonate ($Cu(AcAc)_2$) and diazoacetate giving **30** as a stereomeric mixture of cyclopropane products in a 5:2 ratio. Following debenzylation using Pearlman's catalyst and dihydrogen, cyclitol-cyclopropane **31** and **32** were isolated following silica gel column chromatography. Future studies will reveal whether these *carba*-cyclophellitol derivatives are able to inhibit FUCA1/2, and if so whether they do so by keeping a half-chair conformation within the enzyme active site.

Recently, Speciale *et al.* have identified a new sulfoquinovosidase in *Escherichia coli*, YihQ.[7] This enzyme is a GH31 α-glycosidase that cleaves the modified glucose derivative named sulfoquinovose (SQ) from sulfoquiovosyl diacylglyceride (SQDG) sulfolipids. It would be of interest to develop YihQ activity-based probes to identify its activity in SQDG sulfolipids metabolism in bacteria. In **Chapter 5**, *epi*-cyclophellitol aziridine CF022, JJB347, JJB382-384 have already been successfully synthesized and applied in ABPP of GH31 α-glucosidases. Introduction of a sulfite moiety on to this cyclitol aziridine scaffold would yield potential YihQ ABPs. A proposed synthesis approach for YihQ ABPs **35** is depicted in Scheme 4. O-sulfonation of primary alcohol with sulfur trioxide trimethylamine complex ($SO_3{\cdot}Et_3N$) would lead to the corresponding sulfate derivative **34**, following a sulfonaton protocol as previously reported.[8]

Scheme 4. Synthesis of sulfo-aziridine ABPs **35**.



Reagent and conditions: (a) $SO_3{\cdot}Et_3N$, DMF, 60 °C; (b) $CuSO_4$(1.0 M in H_2O), sodium ascorbate (1.0 M in H_2O), DMF, **9** or **10** or **11**.

Lysosomal β-glucuronidase (GUSB) is an *exo*-glucuronidase that removes D-glucuronic acid residues from the reducing end of glycosaminoglycans (GAGs), whereas heparanase (HPSE) is an *endo*-glucuronidase able to hydrolyse glycosidic linkages within heparin sulfate (HS) chains. Surprisingly, both of GUSB and HPSE are labeled by glucuronic acid emulating cyclophellitol aziridines JJB355, JJB392, JJB395 or JJB397 (compounds in **Chapter 6, 7**). In order to selectively modify HPSE *in vitro* and *in vivo*, it would be useful to have specific HPSE ABPs. Based on the chemical structure of HS and also the HPSE catalytic cleft spatial structure analysis recently performed by Wu *et al*.,[9] HPSE would prefer to recognize oligosaccharide substrate mimetics. Moreover, *exo*-glucuronidases would not cleave the internal linkage of oligosaccharide. Based on these considerations disaccharide derivatives **36**-**43** may turn out to be selective HPSE inhibitors and probes (Figure 4), also considering that the *exo*-glycosidase, GUSB, would not be able to fit these disaccharide-like compounds.

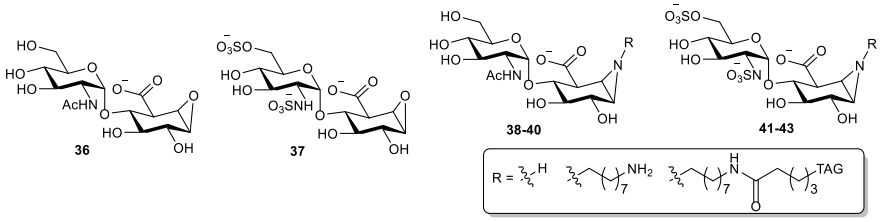
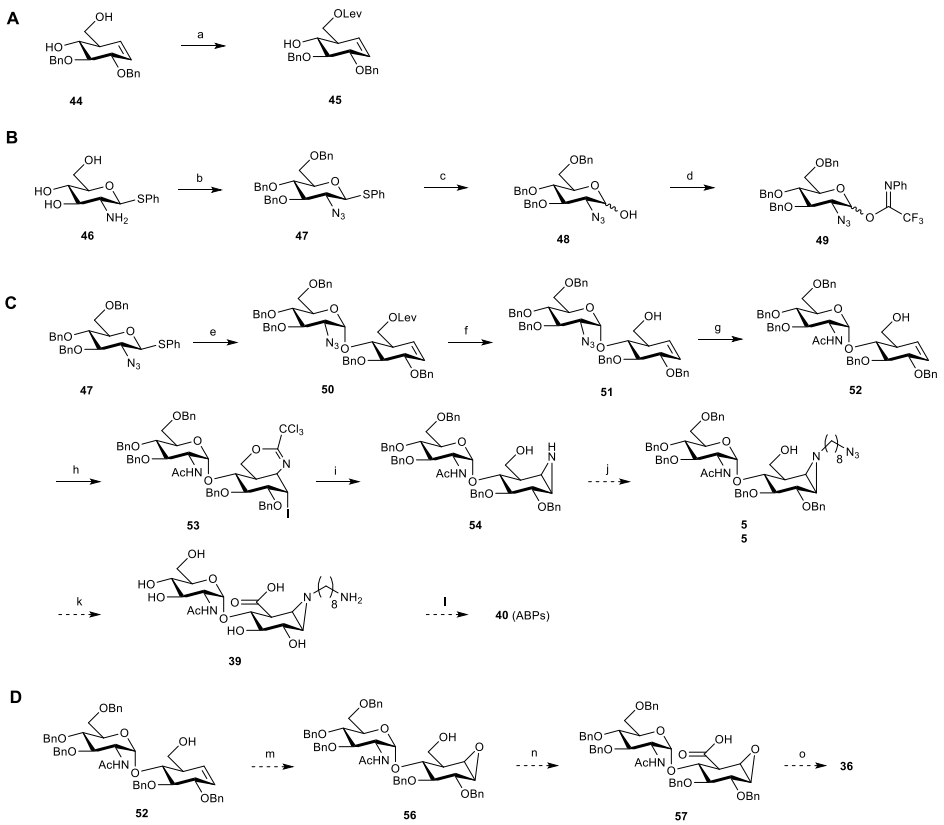


Figure 4. Proposed disaccharide cyclophellitol epoxide and aziridine inhibitors and probes for HPSE.

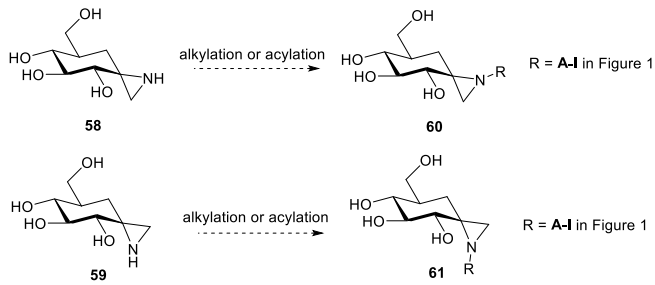
Scheme 5. Synthesis of HPSE inhibitors and probes.



Reagent and conditions: (a) Levulinic acid, DIC, DMAP, DCM, 78%; (b) i) imidazole-1-sulfonyl azide, H_2O, $CuSO_4{\cdot}5H_2O$; ii) BnBr, TBAI, NaH, DMF, 0 °C-20 °C, 43%; (c) *N*-iodosuccinimide (NIS), TFA. DCM/H_2O (v/v, 10/1), 94%; (d) $CF_3C(NPh)Cl$, Cs_2CO_3, acetone/H_2O (20/1), 68%; (e) *N*-formylmorpholine (NFM), NIS, TMSOTf, DCM, molecular sieves 3Å, **45**, -20 °C, 28%; (f) NH_2NH_2, AcOH, pyridine/AcOH (4/1), 64%; (g) AcSH, pyridine, $CHCl_3$, 70%; (h) i) Trichloroacetonitrile, DCM; ii) I_2, H_2O, $NaHCO_3$, 62%; (i) i) 1.0 M aq. HCl, MeOH; ii) $NaHCO_3$, MeOH, 58%; (j) 1-azido-8-iodooctane, K_2CO_3, DMF, 50 °C; (k) i) TEMPO/BAIB, DCM/H_2O (v/v, 2/1), 0 °C; ii) Li, NH_3, THF, -60 °C; (l) BODIPY/Cy5/Biotin-OSu; (m) *m*CPBA, DCM, 40 °C; (n) TEMPO/BAIB, DCM/H_2O (v/v, 2/1), 0 °C; (o) Pd/C, H_2, MeOH.

Compounds **36**, **40** can be synthesized via the strategy proposed in Scheme 5. Firstly, building blocks **45** ('acceptor') and **47**, **49** ('donor') are prepared from diol **44** and thiophenyl **46**, respectively. The primary alcohol in **44**[10] is selectively protected as the levulinoyl ester yielding acceptor **45**. The free amine in **46** is converted into the azide by treatment with Stick's reagent, followed by benzylation of the remaining hydroxyl groups. The glycosylation step proved a major limiting step during these initial synthesis studies, because it involves selective α-1,4-glycosylation linkage formation. Literature procedures towards the installation of related 1,2-*cis*-glycosidic linkages make use of activation strategies including nucleophile modulators, such as dimethylacetamide (DMA),[11] *N*-formylmorpholine (NFM)[12] and diphenyl sulfoxide (DPSO).[13] Herein, NFM was used as the modulator for α-selectivity of glycosylation from **47** to **50**, and after the reaction, 28% α-glycosyl product **50** was isolated. The use of *N*-phenyl-trifluoroacetimidate **49** as donor could be an alternative option to increase the efficiency and α-selectivity in the desired glycosylation event. Disaccharide aziridine **54** was successfully synthesized using the intramolecular aziridination protocol described in Chapter 6. *N*-alkylation of the aziridine nitrogen in **54** followed by global deprotection and oxidation of the cyclophellitol aziridine primary hydroxyl would require some carefully designed synthesis strategies, but seems feasible based on methodology described in this thesis. The disaccharide cyclophellitol epoxide **36** could be prepared from intermediate **52** by *m*CPBA epoxidation, TEMPO/BAIB oxidation and debenzylation depicted in Scheme 5D.

Scheme 6. Synthesis of proposed new type aziridine ABPs **60** and **61** for inverting GHs.

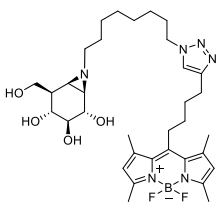


Finally, carbasugar-derived *spiro*-aziridines compounds **58** and **59** (Scheme 6) were reported by Vasella and co-workers group in 2003.[14] These aziridines, structurally related to the cyclophellitol aziridines described in this thesis, proved to be moderately weak inhibitors of *Caldocellum saccharrolyticum* β-glucosidase and yeast α-glucosidase. One might consider investigating whether *spiro*-aziridines **58** and **59** would be useful scaffolds for the development of a new generation of glycosidase activity-based probes. Possibly, the exocyclic aziridine would be able to expel water from the active site of inverting glycosidases to next react with the catalytic base. In case valid, this would lead to the discovery of inverting glycosidase ABPs, although it can not be excluded that compounds **58** and **59** also irreversibly

inhibit retaining glycosidases. The proposed ABPs **60** and **61** could be developed from *spiro*-aziridines **58** and **59** respectively via alkylation or acylation of the nitrogen on the aziridine ring and conjugation of reporter groups.

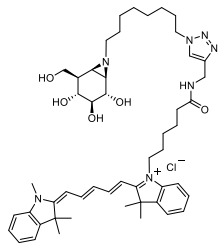
8.3 Experimental section

All reagents were of a commercial grade and were used as received unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and *N*, *N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an argon atmosphere unless stated otherwise. Solvents used for flash column chromatography were of pro analysis quality. Reactions were monitored by TLC analysis using Merck aluminium sheets pre-coated with silica gel 60 with detection by UV absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}.H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4.H_2O$ (10 g/L) in 10% sulfuric acid followed by charring at ~150 ˚C or by spraying with an aqueous solution of $KMnO_4$ (7%) and K_2CO_3 (2%) followed by charring at ~150 °C. Column chromatography was performed using either Baker or Screening Device silica gel 60 (0.04 - 0.063 mm) in the indicated solvents. ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-850 (850/214 MHz), Bruker DMX-600 (600/150 MHz) and Bruker AV-400 (400/100 MHz) spectrometer in the given solvent. Chemical shifts are given in ppm relative to the chloroform residual solvent peak or tetramethylsilane (TMS) as internal standard. Coupling constants are given in Hz. All given ^{13}C spectra are proton decoupled. High-resolution mass spectra were recorded with a LTQ Orbitrap (Thermo Finnigan). Optical rotations were measured on Propol automatic polarimeter (Sodium D-line, λ = 589 nm). IR spectra were recorded on a Shimadzu FT-IR 83000 spectrometer. LC-MS analysis was performed on an LCQ Advantage Max (Thermo Finnigan) ion-trap spectrometer (ESI+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C_{18} column (Gemini, 4.6 mm x 50 mm, 3 µm particle size, Phenomenex) equipped with buffers A: H_2O, B: acetonitrile (MeCN) and C: 1% aqueous TFA or 50 mM NH_4HCO_3 in H_2O, For reversed-phase HPLC purifications an Agilent Technologies 1200 series instrument equipped with a semi preparative Gemini C_{18} column (10 x 250 mm) was used. The applied buffers were A: 25mM NH_4OAc or 50 mM NH_4HCO_3 in H_2O, B: MeCN.

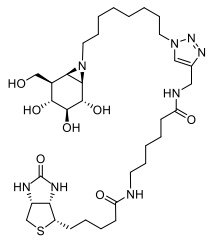


(1R,2S,3S,4R,5R,6R)-7-(8-(4-(4-(5,5-difluoro-1,3,7,9-tetramethyl-5H-4λ⁴,5λ⁴-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-2,3,4-triol (3): Azide compound **1** (13 mg, 0.039 mmol, 1.0 eq.) was dissolved in DMF (0.80 mL). Green BODIPY **9** (15 mg, 0.045 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 15 µL, 0.015 mmol, 0.38 eq.) and sodium ascorbate (1.0 M in H_2O, 16 µL, 0.016 mmol, 0.40 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 40%→65% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting as orange product **3** JJB376 (7.7 mg, 11.7 µmol, 30%). ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 7.73 (s, 1H), 6.11 (s, 2H), 4.35 (t, J = 7.0 Hz, 2H), 3.99 (dd, J = 10.1, 4.4 Hz, 1H), 3.63 – 3.58 (m, 2H), 3.13 – 3.08 (m, 1H), 3.05 – 2.98 (m, 3H), 2.78 (t, J = 7.2 Hz, 2H), 2.43 (s, 6H), 2.37 (s, 6H), 2.36 – 2.29 (m, 1H), 2.16 – 2.07 (m, 1H), 2.01 – 1.95 (m, 1H), 1.94 – 1.82 (m, 5H), 1.68 – 1.59 (m, 3H), 1.58 – 1.50 (m, 2H), 1.37 – 1.19 (m, 8H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 154.91, 148.50, 147.87, 142.19, 132.57, 123.38, 122.61, 79.01, 73.88, 70.08,

63.73, 62.09, 51.19, 45.51, 45.45, 43.02, 32.23, 31.22, 30.81, 30.33, 30.23, 29.84, 29.06, 28.16, 27.30, 25.86, 16.49, 14.45; LC-MS: R_t 6.05 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 657.27 $(M+H)^+$, 637.47 $(M-F)^+$; HRMS: calculated for $C_{34}H_{51}BF_2N_6O_4$ [$M+H^+$] 657.41117, found: 657.41122.

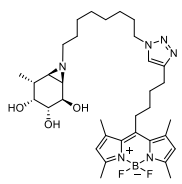


6-(3,3-dimethyl-2-((1E,3E)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-1l4-indol-1-yl)-N-((1-(8-((1R,2S,3S,4R,5R,6R)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)hexanamide (4): Azide compound **1** (20 mg, 0.061 mmol, 1 eq.) was dissolved in DMF (1.0 mL). Blue Cy5 **10** (41 mg, 0.074 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 25 µL, 0.025 mmol, 0.40 eq.) and sodium ascorbate (1.0 M in H_2O, 28 µL, 0.028 mmol, 0.46 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 41%→49% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting as deep blue product **4** JJB367 (6.9 mg, 7.8 µmol, 13%). ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 8.28 – 8.22 (m, 2H), 7.84 (s, 1H), 7.50 (d, *J* = 7.5 Hz, 2H), 7.44 – 7.39 (m, 2H), 7.32 – 7.25 (m, 4H), 6.62 (t, *J* = 12.4 Hz, 1H), 6.28 (dd, *J* = 13.7, 2.6 Hz, 2H), 4.45 – 4.32 (m, 4H), 4.09 (t, *J* = 7.5 Hz, 2H), 3.99 (dd, *J* = 10.1, 4.4 Hz, 1H), 3.69 – 3.54 (m, 5H), 3.12 – 3.09 (m, 1H), 3.00 (t, *J* = 9.8 Hz, 1H), 2.35 – 2.29 (m, 1H), 2.25 (t, *J* = 7.3 Hz, 2H), 2.14 – 2.09 (m, 1H), 1.99 – 1.97 (m, 1H), 1.93 (s, 6H), 1.90 – 1.79 (m, 4H), 1.73 – 1.68 (m, 10H), 1.63 (d, *J* = 6.3 Hz, 1H), 1.59 – 1.42 (m, 4H), 1.39 – 1.23 (m, 8H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 175.74, 175.39, 174.59, 155.55, 155.47, 144.24, 143.54, 142.62, 142.51, 129.78, 129.74, 126.61, 126.28, 126.21, 124.17, 123.42, 123.28, 112.02, 111.85, 104.42, 104.23, 79.03, 73.91, 70.05, 63.72, 62.09, 51.33, 50.54, 45.54, 45.47, 44.75, 43.04, 36.46, 35.57, 31.51, 31.29, 30.37, 30.25, 29.94, 28.21, 28.13, 27.94, 27.79, 27.37, 27.31, 26.39; LC-MS: R_t 5.97 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 848.60 $(M)^+$; HRMS: calculated for $C_{50}H_{70}N_7O_5^+$ [M^+] 848.54329, found: 848.54304.

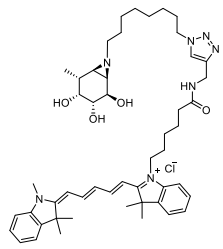


6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)-N-((1-(8-((1R,2S,3S,4R,5R,6R)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)hexanamide (5): Azide compound **1** (13 mg, 0.039 mmol, 1.0 eq.) was dissolved in DMF (0.8 mL). Biotin compound **11** (18.6 mg, 0.047 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 15 µL, 0.015 mmol, 0.38 eq.) and sodium ascorbate (1.0 M in H_2O, 16 µL, 0.016 mmol, 0.40 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 15%→35% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting in white powder product **5** JJB377 (9.6 mg, 13.4 µmol, 34%). ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 7.84 (s, 1H), 4.49 (dd, *J* = 7.9, 4.9 Hz, 1H), 4.42 (s, 2H), 4.38 (t, *J* = 7.1 Hz, 2H), 4.30 (dd, *J* = 7.9, 4.5 Hz, 1H), 3.99 (dd, *J* = 10.1, 4.4 Hz, 1H), 3.67 – 3.55 (m, 2H), 3.23 – 3.19 (m, 1H), 3.18 – 3.14 (m, 2H), 3.13 – 3.09 (m, 1H), 3.02 (t, *J* = 9.8 Hz, 1H), 2.93 (dd, *J* = 12.7, 5.0 Hz, 1H), 2.70 (d, *J* = 12.7 Hz, 1H), 2.36 – 2.31 (m, 1H), 2.27 – 2.10 (m, 5H), 1.99 (dd, *J* = 6.3, 3.5 Hz, 1H), 1.94 (s, 1H), 1.92 – 1.86 (m, 3H), 1.79 – 1.40 (m, 13H), 1.38 – 1.28 (m, 10H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 176.01, 175.98, 166.11, 146.26, 124.17, 79.05, 73.93, 70.11, 63.77, 63.39, 62.10, 61.63, 57.01, 51.35, 45.54, 45.46, 43.04, 41.05, 40.19, 36.83, 36.76, 35.59, 31.27, 30.37,

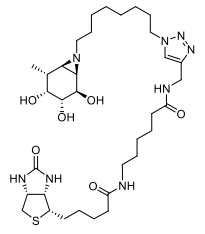
30.28, 30.12, 29.93, 29.78, 29.50, 28.22, 27.54, 27.37, 26.93, 26.51; LC-MS: R_t 3.94 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 723.40 $(M+H)^+$; HRMS: calculated for $C_{34}H_{58}N_8O_7S$ $[M+H^+]$ 723.42219, found: 723.42259.



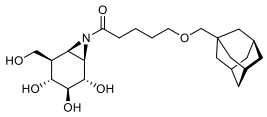
(1R,2R,3R,4R,5R,6R)-7-(8-(4-(4-(5,5-difluoro-1,3,7,9-tetramethyl-5H-4λ⁴,5λ⁴-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)butyl)-1H-1,2,3-triazol-1-yl)octyl)-5-methyl-7-azabicyclo[4.1.0]heptane-2,3,4-triol (6): Azide compound **2** (14 mg, 0.045 mmol, 1.0 eq.) was dissolved in DMF (1.0 mL). Green BODIPY **9** (20 mg, 0.054 mmol, 1.2 eq.), $CuSO_4$ (1.0 M in H_2O, 18 µL, 0.018 mmol, 0.4 eq.) and sodium ascorbate (1.0 M in H_2O, 19 µL, 0.019 mmol, 0.42 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 47%→53% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting as orange product **6** JJB350 (2.3 mg, 3.5 µmol, 8%). ^{1}H-NMR (850 MHz, CD_3OD): *δ* ppm 7.73 (s, 1H), 6.12 (s, 2H), 4.35 (t, *J* = 7.0 Hz, 2H), 3.99 (dd, *J* = 8.7, 4.3 Hz, 1H), 3.54 – 3.52 (m, 1H), 3.34 – 3.32 (m, 1H), 3.06 – 3.00 (m, 2H), 2.79 (t, *J* = 7.3 Hz, 2H), 2.44 (s, 6H), 2.39 (s, 6H), 2.31 – 2.27 (m, 1H), 2.13 – 2.10 (m, 1H), 1.96 – 1.80 (m, 6H), 1.69 – 1.62 (m, 2H), 1.56 – 1.52 (m, 2H), 1.36 – 1.21 (m, 8H), 1.14 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (214 MHz, CD_3OD): *δ* ppm 154.93, 148.50, 147.89, 142.17, 132.59, 123.37, 122.60, 76.04, 75.14, 70.18, 62.29, 51.21, 46.47, 45.73, 36.95, 32.27, 31.22, 30.82, 30.56, 30.42, 29.89, 29.09, 28.21, 27.32, 25.90, 16.77, 16.50, 14.43; LC-MS: R_t 6.26 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 641.33 $(M+H)^+$; HRMS: calculated for $C_{34}H_{51}BF_2N_6O_3$ $[M+H^+]$ 641.41265, found: 641.41541.



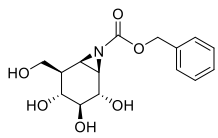
3,3-dimethyl-1-(6-oxo-6-(((1-(8-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl)amino)hexyl)-2-((1E,3E)-5-((E)-1,3,3-trimethylindolin-2-ylidene)penta-1,3-dien-1-yl)-3H-indol-1-ium (7): Azide compound **2** (8.3 mg, 0.027 mmol, 1.0 eq.) was dissolved in DMF (0.80 mL). Blue Cy5 **10** (16 mg, 0.030 mmol, 1.1 eq.), $CuSO_4$ (1.0 M in H_2O, 12 µL, 0.012 mmol, 0.44 eq.) and sodium ascorbate (1.0 M in H_2O, 13 µL, 0.013 mmol, 0.48 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 45%→55% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting as a purple powder product **7** JJB381 (6.9 mg, 7.8 µmol, 13%). ^{1}H-NMR (850 MHz, CD_3OD): *δ* ppm 8.28 – 8.22 (m, 2H), 7.84 (s, 1H), 7.50 (d, *J* = 7.4 Hz, 2H), 7.44 – 7.38 (m, 2H), 7.34 – 7.22 (m, 4H), 6.62 (t, *J* = 12.4 Hz, 1H), 6.28 (dd, *J* = 13.7, 3.3 Hz, 2H), 4.41 (s, 2H), 4.36 (t, *J* = 7.1 Hz, 2H), 4.09 (t, *J* = 7.3 Hz, 2H), 3.99 (dd, *J* = 8.7, 4.3 Hz, 1H), 3.63 (s, 2H), 3.53 – 3.52 (m, 1H), 2.66 (s, 1H), 2.31 – 2.27 (m, 1H), 2.25 (t, *J* = 7.3 Hz, 2H), 2.15 – 2.10 (m, 1H), 1.91 – 1.79 (m, 5H), 1.73 – 1.58 (m, 16H), 1.56 – 1.50 (m, 2H), 1.50 – 1.44 (m, 2H), 1.35 – 1.27 (m, 8H), 1.14 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (214 MHz, CD_3OD): *δ* ppm 175.72, 175.41, 174.61, 155.56, 155.48, 146.13, 144.24, 143.54, 142.63, 142.52, 129.78, 129.74, 126.61, 126.29, 126.22, 124.12, 123.42, 112.02, 111.85, 104.41, 104.23, 76.04, 75.11, 70.13, 62.29, 51.34, 50.54, 50.52, 46.47, 45.71, 44.75, 36.94, 36.46, 35.57, 31.50, 31.30, 30.56, 30.48, 29.99, 28.26, 28.12, 27.94, 27.79, 27.40, 27.31, 26.38; LC-MS: R_t 4.83 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 833.60 $(M)^+$; HRMS: calculated for $C_{50}H_{70}N_7O_4^+$ $[M^+]$ 833.55621, found: 833.55124.



6-(5-((3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamido)-N-((1-(8-((1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5-methyl-7-azabicyclo[4.1.0]heptan-7-yl)octyl)-1H-1,2,3-triazol-4-yl)methyl) hexanamide (8): Azide compound **2** (8.3 mg, 0.027 mmol, 1.0 eq.) was dissolved in DMF (0.80 mL). Biotin compound **11** (12 mg, 0.030 mmol, 1.1 eq.), $CuSO_4$ (1.0 M in H_2O, 12 µL, 0.012 mmol, 0.44 eq.) and sodium ascorbate (1.0 M in H_2O, 13 µL, 0.013 mmol, 0.48 eq) were added to the solution under argon atmosphere. After stirring at room temperature for 12 h, the reaction volatiles were removed under reduced pressure. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 5%→40% B in A, 3 CV, solutions used A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) and lyophilized resulting as white product **8** JJB385 (3.8 mg, 5.3 µmol, 20%). ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 7.84 (s, 2H), 4.50 – 4.48 (m, 1H), 4.42 (s, 2H), 4.38 (t, *J* = 7.1 Hz, 2H), 4.30 (dd, *J* = 7.9, 4.5 Hz, 1H), 4.00 (dd, *J* = 8.7, 4.3 Hz, 1H), 3.54 – 3.52 (m, 1H), 3.36 – 3.32 (m, 1H), 3.23 – 3.18 (m, 1H), 3.17 – 3.13 (m, 2H), 2.92 (dd, *J* = 12.8, 5.0 Hz, 1H), 2.70 (d, *J* = 12.7 Hz, 1H), 2.66 (s, 1H), 2.34 – 2.30 (m, 1H), 2.23 (t, *J* = 7.5 Hz, 2H), 2.21 – 2.18 (m, 2H), 2.17 – 2.12 (m, 1H), 1.92 – 1.83 (m, 3H), 1.78 – 1.71 (m, 1H), 1.71 – 1.53 (m, 7H), 1.53 – 1.47 (m, 2H), 1.45 – 1.40 (m, 2H), 1.40 – 1.26 (m, 10H), 1.15 (d, *J* = 7.5 Hz, 3H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 176.00, 175.96, 166.11, 146.22, 124.14, 76.06, 75.10, 70.14, 63.38, 62.30, 61.62, 57.02, 51.35, 46.48, 45.73, 41.05, 40.18, 36.94, 36.81, 36.75, 35.59, 31.29, 30.57, 30.12, 29.79, 29.50, 27.55, 27.39, 26.93, 26.52, 16.79; LC-MS: R_t 4.48 min, linear gradient 10%→90% B in 12.5 min; ESI-MS: *m/z* = 707.47 $(M+H)^+$; HRMS: calculated for $C_{34}H_{58}N_8O_6S$ $[M+H^+]$ 707.42728, found: 707.42759.

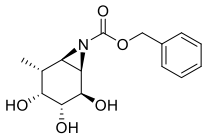
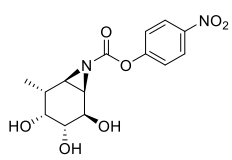


5-(((3R,5R,7R)-adamantan-1-yl)methoxy)-1-((1R,2S,3S,4R,5R,6R)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-7-yl)pentan-1-one (18): *N*-Ethoxycarbonyl-2-ethoxy-1, 2-dihydroquinoline (EEDQ) (79 mg, 0.32 mmol, 1.6 eq.) and adamantane-acid **17** (85 mg, 0.32 mmol, 1.6 eq.) were dissolved in anhydrous DMF (0.32 mL) and stirred at room temperature for 2 h. Pre-activated mixed anhydride solution (160 µL, 0.80 eq.) was added to deprotected aziridine **16** (35 mg, 0.20 mmol, 1.0 eq.) in dry DMF (1.2 mL) at 0 °C and stirred for 30 min. Additional pre-activated solution (160 µL, 0.80 eq) was added. The resulting mixture was stirred at 0 °C for 2 h. The reaction was quenched by MeOH (1.0 mL) and the mixture was concentrated *in vacuo*. Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 38%→47% B in A, 12min, solutions used A: H_2O, B: MeCN) and the fractions were lyophilized directly yielding **18** as white powder (21 mg, 0.05 mmol, 25%).^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.08 (d, *J* = 10.4, 2H), 3.72 – 3.66 (m, 2H), 3.41 (t, *J* = 6.4 Hz, 2H), 3.23 (dd, *J* = 10.0, 8.0 Hz, 1H), 3.10 – 3.02 (m, 2H), 2.74 (d, *J* = 5.6 Hz, 1H), 2.55 – 2.51 (m, 2H), 2.01-1.95 (m, 4H), 1.77 – 1.66 (m, 8H); 1.64 – 1.56 (m, 8H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 188.5, 83.0, 79.1, 73.4, 72.2, 69.4, 63.6, 45.3, 42.4, 41.1, 40.8, 38.3, 36.6, 35.2, 30.1, 29.9, 29.8, 22.9; LC-MS: R_t 7.87min; linear gradient 10→90% B in 15 min, ESI-MS: *m/z* = 424.4 $(M+H)^+$; HRMS: calculated for $C_{14}H_{17}NO_4$ $[M+H^+]$ 424.26936, found: 424.26921.

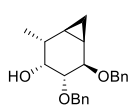


Benzyl (1R,2S,3S,4R,5R,6R)-2,3,4-trihydroxy-5-(hydroxymethyl)-7-azabicyclo[4.1.0]heptane-7-carboxylate (22): Benzyl alcohol (28 µL, 0.28 mmol, 1.1 eq) and *p*-nitrophenol chloroformate (105mg, 0.32 mmol, 2.0 eq.) were dissolved in DCM (6.0 mL). Pyridine (104 µL, 1.3 mmol, 5.0 eq.) was added into the solution, the resulting mixture was stirred

at room temperature for 4 h. After quenching the reaction with brine (1.5 mL), the mixture was transferred into a separatory funnel, washed the organic layer with brine, dried over $MgSO_4$, concentrated *in vacuo*, redissolved in dry DMF 1mL, followed by addition of Et_3N (0.2 mL). Aziridine compound (45.6 mg, 0.26 mmol, 1.0 eq.) in DMF (1.0 mL) was added dropwise into active ester solution under argon atmosphere at room temperature overnight. The resulting solution was concentrated *in vacuo* and Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 17%→23% B in A, 3 CV, solutions used A: H_2O, B: MeCN) and the fractions were lyophilized directly yielding **22** as white powder product (30 mg, 0.11 mmol, 42%). TLC: R_f 0.31 (DCM/MeOH, 5/1, v/v); ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 7.39 – 7.29 (m, 5H), 5.15 – 5.09 (m, 2H), 4.04 (dd, *J* = 10.4, 4.4 Hz, 1H), 3.78 (dd, *J* = 10.4, 8.4 Hz, 1H), 3.71 (d, *J* = 8.0 Hz, 1H), 3.21 (dd, *J* = 10.0, 8.0 Hz, 1H), 3.12 – 3.02 (m, 2H), 2.75 (d, *J* = 6.0 Hz, 1H); 1.98 – 1.90 (m, 1H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 164.6, 137.4, 129.6, 129.3, 129.1, 129.0, 79.0, 73.3, 69.3, 69.3, 63.6, 45.1.3, 43.4.7, 42.2; LC-MS: R_t 5.16 min, linear gradient 00→90% B in 15 min; ESI-MS: *m/z* = 310.3 $(M+H)^+$; HRMS: calculated for $C_{15}H_{19}NO_5$ $[M+H^+]$ 310.12851, found: 310.12876.

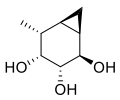



Benzyl (1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5- methyl-7-azabicyclo[4.1.0]heptane-7-carboxylate (24) and 4-nitrophenyl(1R,2R,3R,4R,5R,6R)-2,3,4-trihydroxy-5- methyl-7-azabicyclo[4.1.0]heptane-7-carboxylate (25): Benzyl alcohol (23 µL, 0.23 mmol, 1.1 eq) and p-nitrophenol chloroformate (80.6 mg, 0.40 mmol, 2.0 eq.) were dissolved in DCM (5 mL). Pyridine (80 µL, 1 mmol, 5.0 eq.) was added into the solution, the resulting mixture was stirred at room temperature for 4 h. After quenching the reaction with brine (1.2 mL), the mixture was transferred into a separatory funnel, washed the organic layer with brine, dried over $MgSO_4$, concentrated *in vacuo*, redissolved in dry DMF (1.0 mL), followed by addition of Et_3N (0.20 mL). Aziridine compound (32 mg, 0.20 mmol, 1.0 eq.) in DMF (1.5mL) was added dropwise under argon atmosphere at room temperature overnight. The resulting solution was concentrated *in vacuo* and Then the crude product was purified by semi-preparative reversed HPLC (linear gradient: 19%→23% B in A, 3 CV, solutions used A: H_2O, B: MeCN) giving product **24** (10 mg, 34 µmol, 17% yield) and **25** (4.3 mg, 13 µmol, 7% yield) as white powder. **24** TLC: R_f 0.51 (DCM/MeOH, 8/1, v/v),: ^{1}H-NMR (600 MHz, CD_3OD): δ ppm 7.39 – 7.32 (m, 4H), 5.15 (dd, *J* = 21.6, 12 Hz, 2H), 4.03 (dd, *J* = 9.0, 4.2 Hz, 1H), 3.60 – 3.59 (m, 1H), 3.37 (dd, *J* = 9.0, 1.8 Hz, 1H), 2.98 (dd, *J* = 6.0, 3.6 Hz, 1H), 2.41 (d, *J* = 6.6 Hz, 1H), 2.04 – 2.02 (m, 1H); 1.89 (s, 1H), 1.19 (d, *J* = 7.8Hz, 3H); ^{13}C-NMR (150 MHz, CD_3OD): δ ppm 165.0, 137.5, 129.6, 129.3, 129.2, 129.0, 127.0, 75.7, 74.4, 69.2, 69.2, 44.6, 44.3, 36.7, 16.2; LC-MS: R_t 4.28 min, linear gradient 10→90% B in 15 min; ESI-MS: *m/z* = 294.3 $(M+H)^+$; HRMS: Calculated for $C_{15}H_{19}NO_5$ $[M+H^+]$ 294.13360. Found: 264.13364. **25** ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 8.24 – 8.20 (m, 2H), 7.18 – 7.15 (m, 2H), 4.78 (t, J = 7.8 Hz, 1H), 4.66 (t, J = 6.4 Hz, 1H), 4.26 (t, *J* = 6.4 Hz, 1H), 3.98 – 3.93 (m, 2H), 2.56 – 2.52 (m, 1H), 1.04 (d, *J* = 7.6Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 164.7, 161.6, 143.2, 126.9, 116.6, 81.3, 78.9, 73.4, 73.0, 55.6, 35.4, 11.3; **25** LC-MS: R_t 4.15min, linear gradient 10→90% B in 15 min; ESI-MS: *m/z* = 342.1 $(M+NH_4^+)$.

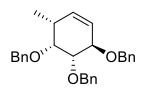


(1R,2R,3R,4R,5R,6R)-4,5-bis(benzyloxy)-2-methylbicyclo[4.1.0]heptan-3-ol (27): Diethyl zinc solution (1.0 M in Et_2O, 930 µL, 930 µmol, 10.0 eq.), $BF_3{\cdot}OEt_2$ liquid (56.7 µL, 463 µmol, 5.0 eq.) and Et_2O (97 µL, 930 µmol, 10.0 eq.) were dissolved in 0.5 mL dry DCM. Diiodomethane (149 µL, 1.8

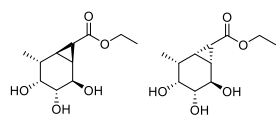
mmol, 20 eq.) was added to the solution and stirred at room temperature for 5 min. A solution of compound **26** (30 mg, 0.093 mmol 1.0 eq.) in dry DCM (0.50 mL) was added to the mixture dropwise. After stirring overnight, the reaction was quenched by saturated aqueous NH_4Cl, the mixture was poured into H_2O and extracted by EtOAc. The organic layer was washed with brine and dried over $MgSO_4$. The solvent was removed under reduced pressure and residue was purified by silica gel column chromatography (5%→20% EtOAc in pentane) yielding cyclopropane compound **27** (26 mg, 0.077 mmol, 83%). TLC: R_f 0.45 (EtOAc/pentane, 1/5, v/v); $[\alpha]_D^{20}$ - 101 (*c* = 0.5, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.46 – 7.21 (m, 10H), 4.82 (d, *J* = 11.8 Hz, 1H), 4.77 – 4.59 (m, 3H), 4.25 (t, *J* = 7.3 Hz, 1H), 3.75 (s, 1H), 3.17 – 3.13 (m, 1H), 2.08 (s, 1H), 1.65 – 1.59 (m, 1H), 1.51 – 1.41 (m, 1H), 1.22 (d, *J* = 12.0 Hz, 3H), 0.89 – 1.74 (m, 2H), 0.27 – 0.20 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 139.24, 138.69, 128.51, 128.41, 127.91, 127.86, 127.76, 127.52, 82.61, 77.48, 77.16, 76.84, 76.47, 73.01, 72.59, 71.00, 36.12, 18.42, 16.94, 16.18, 11.63; HRMS: calculated for $C_{22}H_{26}O_3$ [M+H$^+$] 339.19547, found: 339.19599.



(1R,2R,3R,4R,5R,6R)-5-methylbicyclo[4.1.0]heptane-2,3,4-triol (28): A mixture of product **27** (26 mg, 0.077 mmol, 1.0 eq) and $Pd(OH)_2/C$ (20 wt.% loading(dry basis), 5.0 mg) in MeOH (5.0 mL) was stirred at room temperature under hydrogen atmosphere for 24 h. The catalyst was then filtered off and washed with MeOH. The filtrate were combined and concentrated under reduced pressure. Crude product was purified by silica gel column chromatography (5%→10%, MeOH in DCM) giving the title compound **28** (12 mg, 0.076 mmol, 99%). TLC: R_f 0.60 (1/5, MeOH/DCM, v/v); $[\alpha]_D^{20}$ - 102 (*c* = 0.2, MeOH); ^{1}H-NMR (850 MHz, CD_3OD): δ ppm 4.19 (dd, *J* = 8.9, 6.5 Hz, 1H), 3.53 – 3.48 (m, 1H), 3.05 (dd, *J* = 8.9, 1.8 Hz, 1H), 1.70 – 1.66 (m, 1H), 1.35 – 1.31 (m, 1H), 1.17 (d, *J* = 7.3 Hz, 3H), 0.74 – 0.70 (m, 1H), 0.69 – 0.65 (m, 1H), 0.22 – 0.17 (m, 1H); ^{13}C-NMR (214 MHz, CD_3OD): δ ppm 76.96, 75.85, 70.40, 38.05, 19.89, 18.77, 17.94, 11.82; IR (neat, cm^{-1}): 3321, 2965, 2905, 1366, 1250, 1045, 1022, 995, 929, 812, 682; HRMS: calculated for $C_8H_{14}O_3$ [M+H$^+$] 159.10157, found: 159.10149.

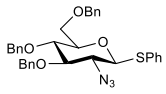


((((1R,2S,3R,6R)-6-methylcyclohex-4-ene-1,2,3-triyl)tris(oxy))tris(methylene)) tribenzene (29): To a solution of **26** (65 mg, 0.2 mmol, 1.0 eq.) in dry DMF (1 mL), benzyl bromide (47.6 µL, 0.4 mmol, 2.0 eq.) and tetrabutylammonium iodide (1.5 mg, 2.0 µmol, 0.02 eq.) were added. The mixture was cooled down to 0 °C, and NaH (60% (w/w) in mineral oil, 16 mg, 0.40 mmol, 2.0 eq.) was added slowly and keep the low temperature for 2 h. After stirring at room temperature overnight, the reaction was quenched with H_2O (0.5 mL), and the mixture was extracted with EtOAc. The organic layer was washed with brine and dried over $MgSO_4$. The solvent was removed under reduced pressure and residue was purified by silica gel column chromatography (0%→20% EtOAc in pentane) yielding cyclohexane **29** (80 mg, 0.19 mmol, 95%). TLC: R_f 0.62 (EtOAc/pentane, 1/8, v/v); $[\alpha]_D^{20}$ - 103 (*c* = 1, $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.54 – 7.33 (m, 15H), 5.85 – 5.80 (m, 1H), 5.58 (d, *J* = 10.0 Hz, 1H), 5.03 (d, *J* = 11.8 Hz, 1H), 4.95 – 4.76 (m, 5H), 4.58 – 4.53 (m, 1H), 4.02 – 3.98 (m, 1H), 3.90 – 3.87 (m, 1H), 2.63 – 2.50 (m, 1H), 1.19 (d, *J* = 7.3 Hz, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 138.93, 132.05, 128.41, 128.37, 128.24, 128.10, 127.85, 127.58, 127.52, 127.48, 127.48, 125.52, 83.30, 77.95, 77.58, 73.82, 72.39, 72.27, 35.84, 16.69; IR (neat, cm^{-1}): 3028, 2872, 1497, 1454, 1089, 1068, 734, 696; HRMS: calculated for $C_{28}H_{30}O_3$ [M+H$^+$] 415.22677, found: 415.22683.

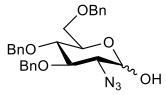


Ethyl (1S,2R,3R,4R,5R,6R,7S)-2,3,4-trihydroxy-5-methylbicyclo [4.1.0]heptane -7-carboxylate (31) and ethyl (1R,2R,3R,4R,5R,6S,7R)-2,3,4-trihydroxy-5-

methylbicyclo[4.1.0]heptane-7-carboxylate (32): A 2-necked pear flask was charged with **29** (80 mg, 0.19 mmol, 1.0 eq.), Copper(II) acetylacetonate ($Cu(AcAc)_2$) (5.0 mg, 0.019 mmol, 0.10 eq.) and dry EtOAc (0.2 mL). The mixture was refluxed at 90 °C and ethyl diazoacetate (13% wt. in DCM, 46 µL, 0.38 mmol, 2.0 eq.) was added in dry EtOAc (0.80 mL) by syringe pump over 6 h. After full conversion of starting material, the mixture was concentrated under reduced pressure, and silica gel column chromatography (1%→10%, EtOAc in pentane) produced the mixture of α and β cyclopropane ester mixture **30**, which was treated by 20 mg $Pd(OH)_2/C$ (20% wt. loading (dry basis)) in MeOH (2.0 mL) at room temperature under hydrogen atmosphere overnight. The reaction mixture was filtered with celite and washed with MeOH. The solvent was removed under reduced pressure and residue was purified by silica gel column chromatography (0%→10% MeOH in DCM) giving target α-cyclopropane **31** (11 mg, 0.049 mmol, 26%) and byproduct β-cyclopropane **32** (4.6 mg, 0.020 mmol, 11%). **31** TLC: R_f 0.34 (MeOH/DCM, 1/9, v/v); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.20 (dd, J = 8.9, 6.2 Hz, 1H), 4.10 (q, J = 7.3 Hz, 2H), 3.55 – 3.51 (m, 1H), 3.08 – 3.03 (m, 1H), 1.98 – 1.92 (m, 1H), 1.84 – 1.76 (m, 1H), 1.58 (t, J = 4.6 Hz, 1H), 1.29 – 1.22 (m, 4H), 1.19 (d, J = 7.3 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 175.52, 76.17, 75.59, 68.90, 61.70, 36.92, 30.95, 29.08, 26.80, 18.47, 14.53; HRMS: calculated for $C_{11}H_{18}O_5$ [M+H$^+$] 231.11261, found: 231.11039. **32** TLC: R_f 0.29 (1/9, MeOH/DCM, v/v); ^{1}H-NMR (400 MHz, CD_3OD): δ ppm 4.15 – 4.06 (m, 2H), 3.82 (d, J = 8.7 Hz, 1H), 3.68 – 3.65 (m, 1H), 3.28 – 3.24 (m, 1H), 2.20 – 2.10 (m, 1H), 2.08 (t, J = 4.6 Hz, 1H), 1.57 – 1.51 (m, 2H), 1.27 – 1.21 (m, 3H), 1.12 (d, J = 6.9 Hz, 3H); ^{13}C-NMR (100 MHz, CD_3OD): δ ppm 78.07, 74.08, 70.34, 61.58, 33.45, 30.32, 29.37, 23.83, 16.82, 14.55; HRMS: calculated for $C_{11}H_{18}O_5$ [M+H$^+$] 231.11261, found: 231.11058.

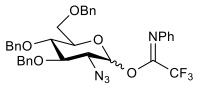


(2S,3R,4R,5S,6R)-3-azido-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)-2-(phenylthio)tetrahydro-2H-pyran (47): A mixture of compound **46** (5.7 g, 21 mmol, 1.0 eq.), 'Stick' reagent, imidazole-1-sulfonyl azide (5.5 g, 26 mmol, 1.2 eq.) and $CuSO_4{\cdot}5H_2O$ (52 mg, 0.21 mmol, 0.010 eq.) in 100 mL EtOH was stirred at room temperature overnight. The resulting solution was concentrated to remove most of EtOH *in vacuo*, diluted by 1.0 M aqueous HCl (100 mL), extracted by EtOAc. The organic layer washed by sat. aq. $NaHCO_3$ and brine and dried over $MgSO_4$ and concentrated *in vacuo*. After three times co-evaporation with toluene, the residue was dissolved in dry DMF (60 mL), followed by addition of BnBr (6.5 mL, 55 mmol, 2.6 eq.), tetrabutylammonium iodide (77 mg, 0.21 mmol, 0.010 eq) and NaH (60%(w/w) in mineral oil, 1.9 g, 55 mmol, 2.6 eq.). The reaction mixture was stirred at 0 °C to room temperature overnight, quenched with MeOH (20 mL) and extracted by Et_2O. The crude product was purified by silica gel column chromatography (0%→15%, EtOAc in pentane) to afford product **47** as a white powder (5.2 g, 8.9 mmol, 43%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.60 (d, J = 6.9 Hz, 2H), 7.38 – 7.15 (m, 18H), 4.84 (d, J = 3.6 Hz, 2H), 4.82 – 4.76 (m, 1H), 4.65 – 4.58 (m, 2H), 4.58 – 4.51 (m, 1H), 4.41 (d, J = 10.1 Hz, 1H), 3.81 – 3.70 (m, 2H), 3.61 (t, J = 9.3 Hz, 1H), 3.48 – 3.52 (m, 2H), 3.34 (t, J = 9.7 Hz, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 149.09, 138.32, 137.96, 137.72, 133.77, 131.27, 129.12, 128.66, 128.61, 128.51, 128.35, 128.16, 128.04, 127.98, 127.75, 86.06, 85.20, 79.45, 77.65, 76.04, 75.19, 73.56, 68.85, 65.16; HRMS: calculated for $C_{33}H_{33}N_3O_4S$ [M+Na$^+$] 590.20840, found: 590.20796.

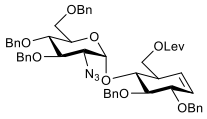


(3R,4R,5S,6R)-3-azido-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-2-ol (48): Compound **47** (568 mg, 1.0 mmol, 1.0 eq.) was dissolved in DCM/H_2O (11 mL, 10/1, v/v) and cooled to 0°C. *N*-iodosuccinimide (225 mg, 1.0 mmol, 1.0 eq.) and TFA (0.74 mL, 1.0 mmol, 1.0 eq.) were added into the solution and stirred for 3 h, and quenched by sat. aq. $Na_2S_2O_3$ (25 mL). The resulting mixture

was diluted by EtOAc, washed by sat. aq. $NaHCO_3$ twice and dried over $MgSO_4$. After concentration *in vacuo*, the residue was purified by silica gel column chromatography (10%→40%, EtOAc in pentane) to afford mixture product **48** as white crystals (449 mg, 0.94 mmol, 94%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.43 – 7.08 (m, 22H), 5.32 (d, J = 3.5 Hz, 1H), 4.88 (s, 2H), 4.84 – 4.76 (m, 2H), 4.61 – 4.47 (m, 5H), 4.11 – 4.05 (m, 1H), 4.04 – 3.96 (m, 1H), 3.71 – 3.53 (m, 4H), 3.52 – 3.34 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 137.94, 137.92, 128.62, 128.59, 128.27, 128.22, 128.14, 128.12, 128.05, 127.99, 127.96, 127.93, 96.30, 92.21, 83.19, 80.25, 78.58, 77.79, 75.73, 75.69, 75.18, 75.15, 75.00, 73.68, 73.63, 70.76, 68.70, 68.64, 67.56, 64.13.

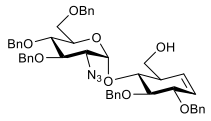


(3R,4R,5S,6R)-3-azido-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-2-yl (E)-2,2,2-trifluoro-N-phenylacetimidate (49): Compound **48** (316 mg, 0.66 mmol, 1.0 eq.) was dissolved in 7 mL actone/H_2O (20/1, v/v) and cooled to 0 °C. *N*-phenyl trifluoroacetimidoyl chloride (280 mg, 1.3 mmol, 2.0 eq.) and Cs_2CO_3 (261 mg, 0.80 mmol, 1.2 eq.) were added into the solution and stirred for 24 h, quenched by adding Et_3N dropwise. The resulting mixture was diluted by EtOAc, washed by H_2O and brine aqueous and dried over $MgSO_4$. After concentration *in vacuo*, the residue was purified by silica gel column chromatography (0%→10%, EtOAc in pentane) yielding product **49** as light yellow oil (293 mg, 0.45 mmol, 68%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.42 – 7.23 (m, 18H), 7.17 (d, J = 6.2 Hz, 3H), 7.10 (t, J = 7.4 Hz, 1H), 6.82 (d, J = 7.7 Hz, 2H), 4.92 (s, 2H), 4.88 – 4.76 (m, 1H), 4.67 – 4.45 (m, 4H), 4.07 – 3.87 (m, 2H), 3.86 – 3.77 (m, 2H), 3.76 – 3.60 (m, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 177.75, 143.39, 137.85, 137.78, 137.76, 137.72, 128.87, 128.63, 128.56, 128.51, 128.17, 128.11, 128.02, 128.00, 127.95, 124.51, 93.89, 93.83, 80.37, 77.58, 75.79, 75.39, 75.18, 73.65, 73.51, 67.92, 63.03.

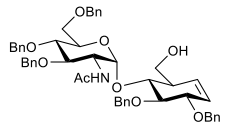


((1R,4S,5R,6R)-6-(((2S,3R,4R,5S,6R)-3-azido-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-2-yl)oxy)-4,5-bis(benzyloxy) cyclohex-2-en-1-yl)methyl 4-oxopentanoate (50): Mixture of thioglycoside donor **47** (108 mg, 0.19 mmol, 1.2 eq.) and flame-dried molecular sieve (3Å) was suspended in dry DCM (1.0 mL), the final concentration of the thioglycoside donor was 60 mM. Then, *N*-formylmorpholine (306 µL, 3.0 mmol, 16 eq.) was added to the mixture. The resulting mixture was stirred at room temperature for 10 min and at reaction temperature specified for particular reaction for additional 10 min. Subsequently, NIS (43 mg, 0.19 mmol, 1.2 eq.) and trimethylsilyltriflate (62 µL, 0.34 mmol, 1.8 eq.) were added, the reaction progress was monitored by TLC (pentane/EtOAc, 5/2, v/v). Upon completion activation of the glycosyl donor, acceptor **45** (70 mg, 0.16 mmol, 1.0 eq.) was added to the reaction mixture. The progress of glycosylation was monitored by TLC. Upon completion of reaction sat. aq. $NaHCO_3$, and $Na_2S_2O_3$ were added to the mixture, followed by vigorous stirring until the brown color of the reaction mixture faded away. The resulting mixture was filtered and extracted by DCM, dried over $MgSO_4$, filtered and concentrated before silica gel column chromatography (EtOAc in pentane, 10%→30%) yielding pure product **50** as light yellow oil (40 mg, 0.045 mmol, 28%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.39 – 7.18 (m, 23H), 7.19 – 7.10 (m, 2H), 5.82 – 5.73 (m, 2H), 5.59 – 5.51 (m, 1H), 5.08 (d, J = 10.9 Hz, 1H), 4.95 – 4.80 (m, 3H), 4.76 (d, J = 10.8 Hz, 1H), 4.71 – 4.57 (m, 3H), 4.55 – 4.44 (m, 2H), 4.30 (dd, J = 11.0, 3.4 Hz, 1H), 4.24 – 4.18 (m, 1H), 4.06 (dd, J = 11.0, 5.3 Hz, 1H), 3.98 – 3.86 (m, 3H), 3.85 – 3.67 (m, 3H), 3.61 – 3.55 (m, 1H), 3.28 (dd, J = 10.4, 4.0 Hz, 1H), 2.69 – 2.60 (m, 3H), 2.52 – 2.47 (m, 2H), 2.15 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 206.35, 172.63, 138.99, 138.19, 138.03, 138.00, 137.98, 128.56, 128.56, 128.54,

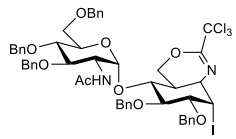
128.49, 128.40, 128.12, 128.04, 128.02, 127.99, 127.94, 127.93, 127.85, 127.80, 127.63, 127.40, 127.36, 98.35, 84.17, 80.90, 80.25, 78.19, 75.50, 75.14, 74.76, 74.52, 73.64, 71.77, 71.59, 68.05, 64.52, 63.30, 42.90, 37.91, 29.96, 27.89; HRMS: calculated for $C_{53}H_{57}N_3O_{10}$ [M+Na$^+$] 918.39362, found: 918.39371.



((1R,4S,5R,6R)-6-(((2S,3R,4R,5S,6R)-3-azido-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-2-yl)oxy)-4,5-bis(benzyloxy) cyclohex-2-en-1-yl)methanol (51): Compound **50** (40 mg, 0.045 mmol, 1.0 eq.) was dissolved in pyridine/AcOH mixture (0.5 mL, 4/1, v/v) at room temperature, followed by addition of hydrazine acetate, NH_2NH_2·AcOH (20.5 mg, 2.23 mmol, 5.0 eq.). The resulting mixture was stirred at room temperature for 1 h and quenched by acetone. The reaction mixture was diluted by EtOAc, washed by H_2O and brine aqueous and dried over $MgSO_4$. After concentration *in vacuo*, the residue was purified by silica gel column chromatography (EtOAc in pentane, 20%→40%) yielding colorless oil product **51** (30 mg, 0.038 mmol, 84%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.38 – 7.24 (m, 23H), 7.11 – 7.06 (m, 2H), 5.84 – 5.75 (m, 2H), 5.66 – 5.59 (m, 1H), 5.10 (d, *J* = 10.9 Hz, 1H), 4.95 – 4.81 (m, 3H), 4.75 (d, *J* = 10.8 Hz, 1H), 4.70 – 4.63 (m, 1H), 4.64 – 4.55 (m, 2H), 4.52 – 4.46 (m, 1H), 4.41 (d, *J* = 10.9 Hz, 1H), 4.30 – 4.25 (m, 1H), 4.13 (t, *J* = 9.6 Hz, 1H), 3.99 – 3.89 (m, 4H), 3.67 – 3.61 (d, *J* = 10.3 Hz, 2H), 3.41 – 3.20 (m, 3H), 2.50 – 2.44 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 139.16, 138.29, 137.84, 137.42, 137.10, 129.66, 128.66, 128.64, 128.61, 128.54, 128.39, 128.27, 128.23, 128.19, 128.16, 128.05, 127.81, 127.76, 127.33, 97.94, 84.79, 81.54, 80.64, 78.79, 75.62, 75.31, 74.64, 73.65, 73.01, 71.75, 71.70, 68.73, 63.27, 61.42, 45.60; HRMS: calculated for $C_{48}H_{51}N_3O_8$ [M+Na$^+$] 820.35684, found: 820.35694.

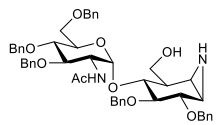


N-((2S,3R,4R,5S,6R)-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)-2-(((1R,2R,5S,6R)-5,6-bis(benzyloxy)-2-(hydroxymethyl)cyclohex-3-en-1-yl)oxy)tetrahydro-2H-pyran-3-yl)acetamide (52): Compound **51** (22 mg, 0.027 mmol, 1.0 eq.) was dissolved in $CHCl_3$ (350 µL), followed by addition of pyridine (300 µL) and thioacetic acid (300µL). The resulting mixture was stirred at room temperature overnight. The reaction mixture was diluted by EtOAc, washed by H_2O and brine aqueous and dried over $MgSO_4$. After concentration *in vacuo*, the residue was purified by silica gel column chromatography (EtOAc in pentane, 50%→90%) yielding product **52** as colorless oil (15.4 mg, 0.019 mmol, 70%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.37 – 7.24 (m, 23H), 7.16 – 7.11 (m, 2H), 6.33 (d, *J* = 9.7 Hz, 1H), 5.88 (dt, *J* = 10.3, 2.7 Hz, 1H), 5.72 – 5.67 (m, 1H), 5.05 (d, *J* = 3.7 Hz, 1H), 4.90 (d, *J* = 11.3 Hz, 1H), 4.79 – 4.62 (m, 4H), 4.60 – 4.50 (m, 4H), 4.44 (d, *J* = 10.8 Hz, 1H), 4.34 (td, *J* = 9.9, 3.6 Hz, 1H), 4.19 – 4.14 (m, 1H), 4.03 – 3.86 (m, 3H), 3.77 (dd, *J* = 8.3, 5.9 Hz, 1H), 3.73 – 3.65 (m, 3H), 3.64 – 3.51 (m, 3H), 2.47 – 2.39 (m, 1H), 1.50 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 170.10, 138.46, 137.97, 137.88, 137.69, 129.62, 128.73, 128.67, 128.61, 128.60, 128.49, 128.28, 128.16, 128.10, 128.06, 128.03, 127.98, 127.88, 127.85, 127.70, 126.83, 126.45, 99.55, 81.53, 81.41, 78.67, 78.12, 76.51, 75.11, 75.06, 74.36, 73.57, 72.39, 71.55, 69.03, 62.56, 52.70, 46.10, 23.01; HRMS: calculated for $C_{50}H_{55}NO_9$ [M+Na$^+$] 837.37690, found: 837.37663.



***N*-((2S,3R,4R,5S,6R)-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)-2-(((4aR,5R,6S,7R,8S,8aR)-6,7-bis(benzyloxy)-8-iodo-2-(trichloromethyl)-4a,5,6,7,8,8a-hexahydro-4H-benzo[d][1,3]oxazin-5-yl)oxy)tetrahydro-2H-pyran-3-yl)acetamide (53):** Product **52** (11 mg,

0.013 mmol, 1.0 eq.) was thrice co-evaporated with toluene and was subsequently dissolved in dry DCM (100 mL). The solution was cooled down to 0 ˚C. DBU (2.7 µL, 0.027 mmol, 2.0 eq.) and trichloroacetonitrile (0.95 µL, 1.4 µmol, 0.10 eq.) were added to the solution. The reaction was stirred at 0˚C for 2 h, TLC showed that the starting material was completely converted into a higher running product. Subsequently, H_2O (42 µL), $NaHCO_3$ (15 mg, 0.18 mmol, 13 eq.) and iodine (11 mg, 0.044 mmol, 3.1 eq.) were added to the solution. It was stirred overnight at room temperature. The reaction mixture was quenched with NaS_2O_3 (10% aqueous), concentrated *in vacuo* and extracted with EtOAc. The organic layer with the intermediate product was dried over $MgSO_4$ and concentrated for purification by silica gel column chromatography (EtOAc in pentane, 10%→40%) yielding product **53** as colorless oil (9.0 mg, 8.3 µmol, 62%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.42 – 7.04 (m, 25H), 6.46 (d, *J* = 9.8 Hz, 1H), 5.14 (d, *J* = 11.6 Hz, 1H), 5.04 (d, *J* = 10.3 Hz, 1H), 4.94 (d, *J* = 3.5 Hz, 1H), 4.85 (t, *J* = 3.5 Hz, 1H), 4.82 – 4.69 (m, 3H), 4.59 (d, *J* = 11.0 Hz, 3H), 4.56 – 4.47 (m, 3H), 4.43 – 4.34 (m, 1H), 4.21 – 4.13 (m, 1H), 4.09 – 4.04 (m, 1H), 4.03 – 3.99 (m, 1H), 3.90 (t, *J* = 9.3 Hz, 1H), 3.77 – 3.70 (m, 4H), 3.53 (t, *J* = 10.0 Hz, 1H), 2.75 (dd, *J* = 9.3, 3.8 Hz, 1H), 2.69 – 2.62 (m, 1H), 1.36 (s, 3H); HRMS: calculated for $C_{52}H_{54}Cl_3N_2O_9$ [M+H$^+$] 1083.20123, found: 1083.20164.



***N*-((2S,3R,4R,5S,6R)-4,5-bis(benzyloxy)-6-((benzyloxy)methyl)-2-(((1R,2R,3R,4S,5S,6R)-4,5-bis(benzyloxy)-2-(hydroxymethyl)-7-azabicyclo[4.1.0]heptan-3-yl)oxy)tetrahydro-2H-pyran-3-yl)acetamide (54):** The intermediate **53** (9.0 mg, 8.3 µmol, 1.0 eq.) was dissolved in dioxane (0.5 mL) and 2 drops of HCl 37% in water. The reaction mixture was refluxed at 60 ˚C for 4 h. The reaction progress was monitored by TLC with pentane/EtOAc mixture (2/1, v/v). After which it was concentrated *in vacuo* and redissolved in MeOH (0.50 mL), and $NaHCO_3$ (14 mg, 0.17 mmol, 20 eq.) was added making the pH > 7. The reaction mixture was stirred at room temperature for three days. The reaction mixture was concentrated *in vacuo*, redissolved in H_2O (5.0 mL) and extracted with EtOAc. The organic layer was dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified via silica gel column chromatography (2%→6%, MeOH in DCM) yielding aziridine **54** as white solid (4.0 mg, 4.8 µmol, 58%). TLC: R_f 0.43 (MeOH/DCM, 1/9, v/v); ^{1}H-NMR (400 MHz, $CDCl_3$): δ ppm 7.37 – 7.24 (m, 20H), 7.24 – 7.13 (m, 5H), 4.96 (d, *J* = 11.1 Hz, 1H), 4.90 (d, *J* = 3.6 Hz, 1H), 4.77 (d, *J* = 11.2 Hz, 2H), 4.66 (d, *J* = 10.2 Hz, 3H), 4.63 – 4.52 (m, 3H), 4.52 – 4.44 (m, 3H), 4.32 (td, *J* = 9.8, 3.6 Hz, 1H), 4.26 – 4.19 (m, 1H), 4.09 – 4.00 (m, 3H), 3.82 (d, *J* = 7.7 Hz, 1H), 3.73 – 3.59 (m, 6H), 3.58 – 3.50 (m, 1H), 2.59 – 2.55 (m, 1H), 2.35 (d, *J* = 6.2 Hz, 1H), 2.07 – 1.98 (m, 1H), 1.42 (s, 3H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ ppm 170.25, 138.56, 138.11, 138.05, 137.70, 128.94, 128.73, 128.68, 128.58, 128.55, 128.44, 128.23, 128.17, 128.12, 127.98, 127.96, 127.85, 127.75, 127.61, 100.62, 81.63, 78.03, 75.25, 75.10, 74.85, 73.57, 72.47, 72.41, 69.18, 68.30, 62.80, 53.28, 44.01, 31.52, 31.50, 30.49, 29.85, 29.07, 23.87, 23.14, 22.82; HRMS: calculated for $C_{50}H_{56}N_2O_9$ [M+H$^+$] 829.40586, found: 829.40593.

8.4 References

[1] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416-436.

[2] J. Jiang, T. J. Beenakker, W. W. Kallemeijn, G. A. van der Marel, H. van den Elst, J. D. C. Codée, J. M. Aerts and H. S. Overkleeft, *Chem. Eur. J.* **2015**, *21*, 10861-10869.

[3] A. H. Futerman and G. van Meer, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 554-565.

[4] B. Chandrasekar, T. Colby, A. Emran Khan Emon, J. Jiang, T. N. Hong, J. G. Villamor, A. Harzen, H. S. Overkleeft and R. A. van der Hoorn, *Mol. Cell Proteomics* **2014**, *13*, 2787-2800.

[5] H. S. Overkleeft, G. H. Renkema, J. Neele, P. Vianello, I. O. Hung, A. Strijland, A. M. van der Burg, G. J. Koomen, U. K. Pandit and J. M. F. G. Aerts, *J. Biol. Chem.* **1998**, *273*, 26522-26527.

[6] a) K. S. E. Tanaka, G. C. Winters, R. J. Batchelor, F. W. B. Einstein and A. J. Bennet, *J. Am. Chem. Soc.* **2001**, *123*, 998-999. b) C. Bluchel, C. V. Ramana and A. Vasella, *Helv. Chim. Acta* **2003**, *86*, 2998-3036. c) N. Akiyama, S. Noguchi and M. Hashimoto, *Biosci. Biotech. Bioch.* **2011**, *75*, 1380-1382.

[7] G. Speciale, Y. Jin, G. J. Davies, S. J. Williams and E. D. Goddard-Borger, *Nat. Chem. Biol.*, **2016**, *4*, 215-217.

[8] Y.-P. Hu, S.-Y. Lin, C.-Y. Huang, M. M. L. Zulueta, J.-Y. Liu, W. Chang and S.-C. Hung, *Nat. Chem.* **2011**, 3, 557-563.

[9] L. Wu, C. M. Viola, A. M. Brzozowski and G. J. Davies, *Nat. Struct. Mol. Biol.* **2015**, *22*, 1016-1022.

[10] K.-Y. Li, J. Jiang, M. D. Witte, W. W. Kallemeijn, H. van den Elst, C.-S. Wong, S. D. Chander, S. Hoogendoorn, T. J. M. Beenakker, J. D. C. Codée, J. M. F. G. Aerts, G. A. van der Marel and H. S. Overkleeft, *Eur. J. Org. Chem.* **2014**, *2014*, 6030-6043.

[11] S. Koto, N. Morishima, M. Owa, S. Zen, *Carbohydr. Res.* **1984**, *130*, 73-83.

[12] A. B. Ingle, C. S. Chao, W. C. Hung and K. K. Mong, *Org. Lett.* **2013**, *15*, 5290-5293.

[13] D. Crich, W. Li, *Org. Lett.* **2006**, *8*, 959-962.

[14] P. Kapferer, V. Birault, J.-F. Poisson and A. Vasella, *Helv. Chim. Acta* **2003**, *86*, 2210-2227.

Chinese summary
小结

Activity-based protein profiling of glucosidases, fucosidases and glucuronidases
活性蛋白表达谱技术在葡萄糖苷酶、岩藻糖苷酶和葡糖醛酸酶研究中的应用

本书描述了六元环多醇氮丙啶 (cyclophellitol aziridine) 及以构象类似物为骨架的活性分子探针(activity-based probes, ABPs)的设计与合成，探讨了相应的 ABPs 应用于葡萄糖苷酶、岩藻糖苷酶和葡糖醛酸酶等保留型糖苷酶的化学生物学研究。

第一章简要介绍了 β-葡萄糖苷酶的一些性质，包括保留型/反转型酶分类的方法以及水解相应催化底物所采用的分子机制。其中，保留型葡萄糖苷酶是采用 Koshland 两步双取代机制。此外，本章节还简单描述了六元环多醇氮丙啶为基础的 ABPs 在保留型 β-葡萄糖苷酶活性蛋白表达谱(Activity-based protein profiling, ABPP)中的应用，此应用也是本研究的理论与实验依据。

第二章是关于六元环多醇氮丙啶及其类似物合成方法的文献综述。本章节首先介绍文献报道过的两种不同的六元环多醇氮丙啶合成策略，一种是以反转式环氧环为起始底物；另一种则是以高烯丙基醇分子内的碘环化反应为关键步骤的合成。随后介绍了这两种策略在其他六元环多醇氮丙啶构型与功能类似物合成中的应用。

第三章描述了将 ABPP 应用于保留型 α-L-岩藻糖苷酶 (FUCA) 的研究。保留型 FUCA 属于糖苷水解酶家族 29 (GH29)，人体内 FUCA 的缺失将会导致一种罕见性疾病——岩藻糖苷贮积症(fucosidosis)，它是众多溶酶体贮积病中的一种。在本章研究内容中，L-吡喃岩藻糖苷构型的六元环多醇氮丙啶化合物 (JJB237、JJB244、JJB256 和 JJB243) 在诸多体内和体外试验中都能够选择性的标记细菌、老鼠或人类的 GH29 FUCA。这一章还报道了八种 L-fuconojirimycin 构型异构体的合成方法，以及利用竞争性 ABPP 技术筛选这些化合物以便作为 FUCA 的潜在抑制剂。最后，透过多形拟杆菌 2970 的岩藻糖苷酶与 N-乙酰化-L-吡喃岩藻糖苷-环多醇氮丙啶化合物共结晶和 X-射线衍射分析，我们捕获了岩藻糖苷型异头碳与 FUCA 亲和活性位点共价连接的氮丙啶-糖苷酶复合物晶体及其反向扭船式结构，这也揭示了 FUCA 介导的 L-吡喃岩藻糖苷底物水解构象变化。

第四章着重介绍了基于 N-烷基化六元环多醇氮丙啶结构的保留型糖苷酶的 ABPs。本章节报道了两种 *N*-烷基化六元环多醇氮丙啶 ABPs 的合成以及它们在对于 β-葡萄糖苷酶、和 α-L-岩藻糖苷酶选择性标记中的应用。相较之前报道过的酰化氮丙啶 ABPs，烷基化氮丙啶 ABPs 合成策略更为简便且在弱酸弱碱环境下具有更高的稳定性，从而增强了其实用性。另外，β-葡萄糖苷构象的烷基化氮丙啶 ABPs 与相应的酰化 ABPs 具有相似的标记效力，而 α-L-岩藻糖苷构象的烷基化氮丙啶 ABPs 则比其对应的酰化 ABPs 标记活性略低。因此，烷基化氮丙啶 ABPs 在保留型糖苷酶 ABPs 设计中可以是一个更具竞争力的选择，对于

新型保留型糖苷酶 ABPs 的设计，烷基化和酰化氮丙啶 ABPs 都可作为理想的探针类型。

第五章阐述了一系列α-D-葡萄糖苷烷基化氮丙啶ABPs的合成，包括荧光探针JJB347、JJB382 和 JJB383 以及生物素探针 JJB384，介绍了这些探针结合 APP 技术在糖苷水解酶家族 31 (GH31) α-葡萄糖苷酶研究中的应用。α-葡糖苷酶参与人体多种生理过程，如胃肠中碳水化合物的分解，内质网 (ER) 中的糖蛋白加工和溶酶体中糖原分解代谢等。溶酶体 α-葡萄糖苷酶 (GAA) 的缺乏会导致庞贝氏(Pompe)症，一种比较常见的溶酶体糖原贮积症。酶活抑制实验和 X 射线晶体衍射分析的结果证明了新合成的 ABPs 是 GH31 α-葡萄糖苷酶的共价抑制剂。另外，这些 ABPs 可以特异性的标记保留型 α-葡萄糖苷酶，包括 GAA 和内质网 II 型 α-葡萄糖苷酶，而且标记效力与专一性依赖于 pH 值。此外，此章节还介绍了用以上 ABPs 来对庞贝氏症病人成纤维细胞进行诊断，并报道了其在肠道消化型α-葡萄糖苷酶，如蔗糖酶(Sis)和麦芽糖酶(MGAM)等的蛋白质组学分析中的应用。

第六章介绍了 β-葡糖醛酸构型的六元环多醇氮丙啶及其衍生物(JJB133、JJB144、JJB249、JJB355、JJB391、JJB392 和 JJB395)的合成策略。它们被用来作为糖苷水解酶家族 2 和 79 (GH2 和 GH79) α-葡糖醛酸酶的抑制剂与 ABPs。GH2 β-葡糖醛酸酶 (GUSB) 与糖胺聚糖 (GAG) 贮积症-Sly 疾病直接相关，GH79 β-葡糖醛酸酶与炎症，肿瘤血管生成和细胞迁移等有关。糖醛酸-*N*-烷基化六元环多醇氮丙啶比糖醛酸-*N*-酰化-六元环多醇氮丙啶较易合成，但以二者为基础制备的 ABPs 在抑制和标记 β-葡糖醛酸酶方面具有同等效力，这与第四章介绍的有关结论相一致。此外，通过晶体学分析荚膜醋杆菌(Acidobacterium capsulatum) β-葡糖醛酸酶(AcaGH79)与烷基化或酰化的糖醛酸氮丙啶的复合物进一步验证了酶亲和活性位点的共价修饰以及复合物的 $^{4}C_{1}$ 椅式构象。糖醛酸-*N*-烷基化六元环多醇氮丙啶 ABPs 的应用在**第七章**展开，包括使用 LC-MS/MS 方法准确地鉴定人类 GH2 溶酶体 β-葡糖醛酸酶 (GUSB) 和 GH79 肝素酶 (HPSE) 以及富集的 GUSB 的亲和活性位点。有趣的是，HPSE 作为一种内切糖苷酶(*endo*-GHs)也能被糖醛酸-*N*-烷基化六元环多醇氮丙啶衍生物抑制或标记，虽然单糖类似物一般是用来设计外切糖苷酶 (*exo*-GHs) 底物的。最后，此章节还比较分析了氮丙啶化合物 JJB355 与外切型 GUSB、外切型 AcaGH79 和内切型 HPSE 形成的复合物的共晶体结构，揭示了这些保留型糖苷酶的 Kosland 双取代催化机制以及 $^{1}S_{3}$-$^{4}H_{3}$-$^{4}C_{1}$ 类型的 β-葡糖醛酸水解过程。

第八章对本论文的实验工作进行了归纳和总结，并对未来的研究工作进行了展望。其中，以氮丙啶类化合物为基本骨架为其他保留型糖苷水解酶设计了一系列 ABPs，为内切型 HPSE 设计了二糖构型的氮丙啶 ABPs，最后引入螺环氮丙啶 (spiro-aziridine)为设计反转型糖苷酶 ABPs 提供新的思路。

Curriculum vitae

Jianbing Jiang was born on 1st August 1985 in Liaocheng, Shaodong, China. When he was in high school, he joined the Chinese chemistry and biology Olympiad, which stimulated his interest in subjects of chemistry and biology. After taking the National College Entrance Examination in 2004, he entered Yantai University and obtained his bachelor degree of biotechnology in 2008. Thereafter, he moved to Lanzhou University for his master study in the field of biochemistry. In 2011, he obtained his master degree with the thesis 'Collagen model peptides labeling and their triple helices folding/unfolding analysis' under the supervision of Prof. Dr. Shouliang Dong.

From September 2011 to June 2016, he studied in the group of Bio-organic Synthesis in Leiden University as a PhD student under supervision of Prof. Dr. Herman S. Overkleeft and Prof. Dr. Johannes M. F. G. Aerts. The author conducted research in the field of chemical glycobiology and was supported by the China Scholarship Council (CSC). During his PhD study, he developed new generations of activity-based probes for the study of several retaining glycosidases. Parts of his work was orally presented at the CHEMPROBE-COST Meeting (Cambridge, 2014), the CHAINS-Dutch chemistry conference (Veldhoven, 2014) and the 18th European Carbohydrate Symposium (Moscow, 2015), and as poster prtersentations at the 4th and 5th EMBO Chemical Biology meeting (Heidelberg, 2012 and 2014). The author also participated in the AIMMS Post-Graduate Course of chemical biology/medicinal chemistry (Amsterdam, 2011) and the HRSMC summer school 'New vistas for organic synthesis' (Maastricht, 2013).

List of publications

1. *Exo*- and *endo*-retaining β-glucuronidases activities and mechanisms revealed by cyclophellitol aziridine-based inhibitors and probes.
   Jiang, J., Wu, L., Jin, Y., Kallemeijn, W. W., Kuo, C. L., Dai, W., van Elk, C., van der Marel, G. A., Codée, J. D. C., Florea, B. I., van den Elst, H., van Eijk, M. C., Aerts, J. M. F. G., Overkleeft, H. S., and Davies, G. J. **2016**, manuscript in preparation.

2. Synthesis and biological evaluation of glucuronide-configured cyclophellitol and cyclophellitol aziridine based β-glucuronidase inhibitors and probes
   Jiang, J., Jin, Y., Kallemeijn, W. W., van Elk, C., van der Marel, G. A., Codée, J. D. C., Aerts, J. M. F. G., Davies, G. J., and Overkleeft, H. S. **2016**, manuscript in preparation.

3. Detection of active mammalian GH31 α-glucosidases in health and disease using in-class, broad spectrum activity-based probes.
   Jiang, J., Kuo, C.-L., Wu, L., Franke, C., Kallemeijn, W. W., Florea, B.I., van Meel, E., van der Marel, G. A., Codée, J. D. C., Boot, R. G., Gideon J. D., Overkleeft, H, S., and Aerts, J. M. F. G. *ACS Central Science*, **2016**, DOI:10.1021/acscentsci.6b00057.

4. Pharmacological chaperones stabilize glycosylceramidase through hydrophobic interactions.
   Bdira, B. F., **Jiang, J.**, Kallemeijn, W. W., Haan, A., Florea, B. I., Bleijlevens, B., Boot, R. G., Overkleeft, H. S., Aerts, J. M. F. G., and Ubbink, M, *Biochemistry,* **2016**, under review.

5. The synthesis of cyclophellitol aziridine and its configurational and functional isomers: a literature survey.
   Jiang, J., Artola, M., Beenakker, T. J. M., Schröder, S. P., Petracca, R., De Boer, C., Aerts, J. M. F. G., van der Marel, G. A., Codée, J. D. C., and Overkleeft, H. S., *European Journal of Organic Chemistry,* **2016**, DOI:10.1021/ejoc.201600472.

6. Glucosyl epi-cyclophellitol allows mechanism-based inactivation and structural analysis of human pancreatic alpha-amylase.
   Caner, S., Zhang, X., **Jiang, J**., Chen, H.-M., Nguyen, N. T., Overkleeft, H. S., Brayer, G. D., and Withers, S. G. *FEBS Letters*, **2016**, 590(8): 1143-1151.

7. Chemoenzymatic synthesis of 6-phospho-cyclophellitol as a novel probe of 6-phospho-β-glucosidases.
   Kwan, D. H., Jin, Y. **Jiang, J.**, Chen, H.-M., Kötzler, M. P., Overkleeft, H. S. Davies, G. J., and Withers, S.G. *FEBS Letters,* **2016**, 590(4): 461-468.

8. Comparing cyclophellitol N-alkyl and N-acyl cyclophellitol aziridines as activity-based glycosidase probes.
Jiang, J., Beenakker, T. J. M., Kallemeijn, W. W., van der Marel, G. A., van den Elst, H., Codée, J. D. C., Aerts J. M. F. G., and Overkleeft, H. S. *Chemistry – A European Journal,* **2015**. 6(5): 2782-2789.

9. *In vitro* and *in vivo* comparative and competitive activity-based protein profiling of GH29 alpha-L-fucosidases.
Jiang, J., Kallemeijn, W. W., Wright, D. W., van den Nieuwendijk, A. M. C. H., Rohde, V. C., Folch, E. C., van den Elst, H., Florea, B. I., Scheij, S., Donker-Koopman, W. E., Verhoek, M., Li, N., Schurmann, M., Mink, D., Boot, R. G., Codée, J. D. C., van der Marel, G. A., Davies, G. J., Aerts, J. M. F. G., and Overkleeft, H. S. *Chemical Science,* **2015**, 21(30): 10861-10869.

10. Synthesis of cyclophellitol, cyclophellitol aziridine, and their tagged derivatives.
Li, K.-Y., **Jiang, J.**, Witte, M. D., Kallemeijn, W. W., van den Elst, H., Wong, C. S., Chander, S. D., Hoogendoorn, S., Beenakker, T. J. M., Codée, J. D. C., Aerts, J. M. F. G., van der Marel, G. A., and Overkleeft, H. S. *European Journal of Organic Chemistry,* **2014**, 2014(27): 6030-6043.

11. Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors.
Li, K. Y., **Jiang, J.**, Witte, M. D., Kallemeijn, W. W., Donker-Koopman, W. E., Boot, R. G., Aerts, J. M. F. G., Codée, J. D. C., van der Marel, G. A., and Overkleeft, H. S. *Organic & Biomolecular Chemistry,* **2014**, 12(39): 7786-7791.

12. From covalent glycosidases to activity-based glycosidase probes.
Willems, L. I., **Jiang, J**. Li, K. Y., Witte, M. D., Kallemeijn, W. W., Beenakker, T. J., Schroder, S. P., Aerts, J. M., van der Marel, G. A., Codée, J. D., and Overkleeft, H. S. *Chemistry – A European Journal,* **2014**, 20(35): 10864-10872.

13. Rational design of activity-based retaining β-exoglucosidase probes. In: concepts and case studies in chemical biology.
Li, K.-Y., Kallemeijn, W. W., **Jiang, J.**, Walvoort, M., Willems, L. I., Beenakker, T., van den Elst, H., van der Marel, G., Codée, J., Aerts, H., Florea, B., Boot, R., Witte, M., and Overkleeft, H. S. *Wiley-VCH Verlag GmbH & Co. KGaA,* **2014**, 191-206.

14. Broad-range glycosidase activity profiling.
Chandrasekar, B., Colby, T., Emon, A. E. K., **Jiang, J**., Hong, T. N., Villamor, J. G., Harzen, A., Overkleeft, H. S., and van der Hoorn, R. A. L. *Molecular & Cellular Proteomics,* **2014**, 13(10): 2787-2800.

15. Site-directed spin labeling of a collagen mimetic peptide.
Jiang, J., L. Yang, Jin, Q., Ma, W., Moroder, L., and Dong, S. *Chemistry – A European Journal*, **2013**, 19(52): 17679-17682.

16. Novel activity-based probes for broad-spectrum profiling of retaining beta-exoglucosidases *in situ* and *in vivo*.
Kallemeijn, W. W., Li, K. Y., Witte, M. D., Marques, A. R., Aten, J., Scheij, S., **Jiang, J.**, Willems, L. I., Voorn-Brouwer, T. M., van Roomen, C. P., Ottenhoff, R., Boot, R. G., van den Elst, H., Walvoort, M. T., Florea, B. I., Codée, J. D., van der Marel, G. A., Aerts, J. M., and Overkleeft, H. S. *Angewandte Chemie International Edition,* **2012**, 51(50): 12529-12533.