Exploring chemical space in covalent and competitive glycosidase inhibitor design

Chen, Y.

Exploring chemical space in covalent and competitive glycosidase inhibitor design

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 13 oktober 2022
klokke 10:00 uur

door

Yurong Chen
geboren te Yunnan, China
in 1992

Promotores

Prof. dr. H. S. Overkleeft

Prof. dr. J. D. C. Codée

Co-promotor

Dr. M. E. Artola Perez de Azanza

Promotiecommissie

Prof. dr. G. A. van der Marel

Prof. dr. J. M. F. G. Aerts

Prof. dr. M. van der Stelt

Dr. Z. W. B. Armstrong

Prof. dr. G. J. Davies (University of York)

Dr. L. I. Willems (University of York)

Doctoral Thesis, Leiden University, 2022

Table of Contents

Chapter 1 **1**

General introduction

Chapter 2 **17**

Activity-based protein profiling of retaining α-amylases in complex biological samples

Chapter 3 **75**

A new generation of fluorescent activity-based probes targeting starch-degrading glycosidases

Chapter 4 **107**

Synthesis and biochemical evaluation of bifunctional cyclophellitol aziridines as selective GBA inhibitors and activity-based probes

Chapter 5 **135**

Novel uronic acid-type 1-*N*-iminosugars derived from siastatin B as competitive heparanase inhibitors

Chapter 6 **179**

Summary and future prospects

Nederlandse Samenvatting **215**

Summary in Chinese **219**

List of Publications **221**

Curriculum vitae **223**

1

General Introduction

Glycoside hydrolases (glycosidases/GHs) are widely abundant enzymes in all kingdoms of life and are important biocatalysts that catalyze the hydrolysis of glycosidic linkages in oligo/polysaccharides, glycoproteins and glycolipids with tremendous efficiency.[1] Abnormal glycosidase activity is intimately associated with a variety of human diseases. Overexpression of heparanase, for example, is implicated in almost all cancers examined, and correlates with increased tumor size, tumor angiogenesis, enhanced metastasis and poor prognosis.[2] Specific inhibitors of glycosidases are of great value, not only because they can serve as useful biological tools to study the catalytic machinery, mechanism and itinerary of target enzymes by crystal structure analysis of (covalent) inhibitor-enzyme complexes,[3] but also because they may act as starting points for the development of therapeutic drugs for the treatment of glycosidase-mediated diseases.[4] Additionally, covalent mechanism-based inhibitors have been used as scaffolds for the development of activity-based probes (ABPs) which allow profiling of glycosidases in complex biological systems.[5] This dissertation describes the synthesis and biochemical evaluation of covalent inhibitors and ABPs for retaining α-amylases and lysosomal β-glucocerebrosidase (GBA), as well as the development of a panel of uronic acid-type 1-*N*-iminosugars as potential competitive heparanase inhibitors. This chapter introduces some mechanistic aspects of glycosidases, including the catalytic mechanisms employed by retaining and inverting glycosidases, the design of mechanism-based enzyme inhibitors and an overview of activity-based protein profiling workflows.

1.1 Classification of glycosidases

Glycosidases can be classified in several ways. The simplest classification is the one based on substrate specificity and is captured in the Enzyme Commission (EC) number for a given enzyme.[6] This classification, however, does not reflect the structural features of the enzymes and is not appropriate for enzymes that act on several substrates. In 1991, a new classification system based on amino acid sequence similarities was introduced,[7] which proved to be a highly valuable tool in understanding the structure/function relationships of glycosidases. To date, over 160 Glycoside Hydrolase (GH) families are categorized and are available in the Carbohydrate Active enZyme (CAZy) database (www.cazy.org).[8] Since sequence and protein fold similarities are directly related, glycosidases within the same GH family usually possess similar structural features and catalytic mechanisms.[9,10]

Glycosidases typically hydrolyze their substrates by employing either a retaining or inverting mechanism, as firstly described by Daniel E. Koshland, Jr in 1953.[11] Two major catalytic mechanisms employed by most glycosidases involve oxocarbenium ion-like transition

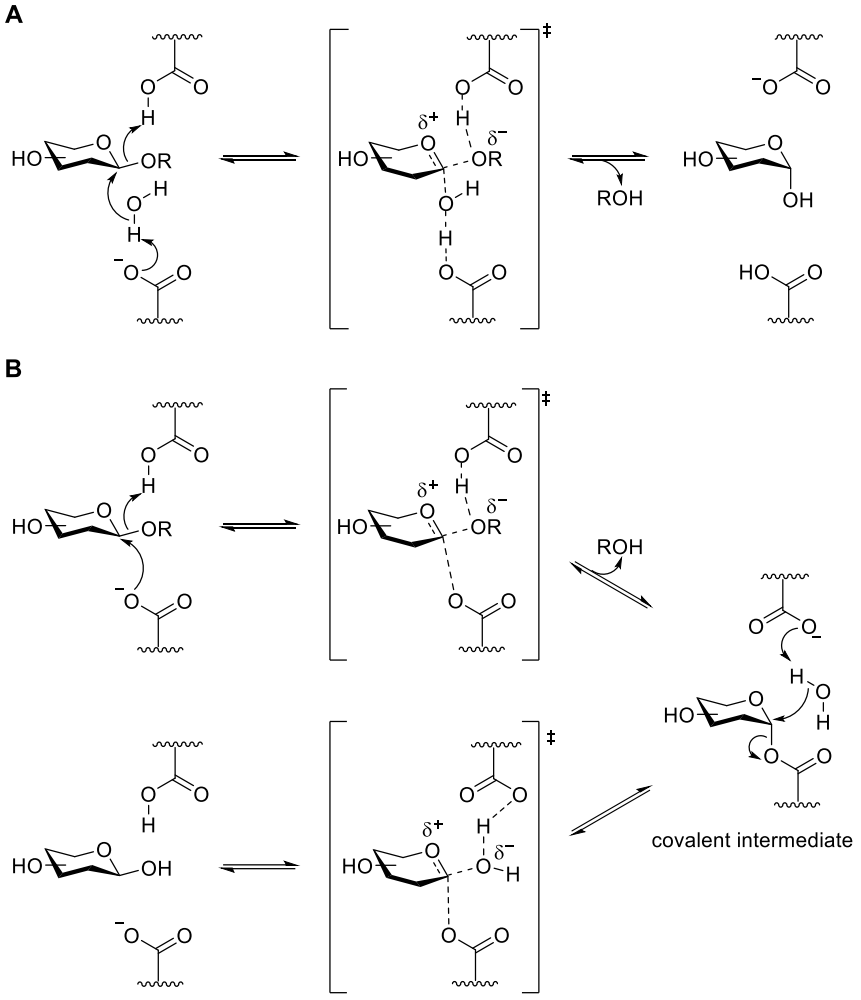


Figure 1.1. General mechanism of the hydrolysis of a β-glycoside employed by A) inverting glycosidases and B) retaining glycosidases.

states and a pair of carboxylic acids/carboxylates (aspartic acid/aspartate or glutamic acid/glutamate) as key residues within the glycosidase active site (Figure 1.1). In inverting glycosidases, one carboxylate residue acts as the catalytic base to deprotonate the incoming water molecule during its attack at the anomeric carbon. The other carboxylic acid then acts as the catalytic acid to protonate the aglycon oxygen atom, thereby assisting in its departure from the anomeric center (Figure 1.1A). The reaction completes in one step where the hydroxyl displaces the aglycon following a S_N2-type mechanism, releasing the hydrolyzed product with inversion of stereochemistry at the anomeric carbon.[12] In retaining glycosidases, S_N2-type displacement occurs twice involving two oxocarbenium ion-like transition states (classical Koshland double-displacement mechanism). During the first step, the catalytic nucleophile attacks the anomeric center while the catalytic acid/base residue protonates the leaving aglycon to form a covalent substrate-enzyme intermediate with inversion of stereochemistry at the anomeric position (Figure 1.1B). During the second step, the catalytic acid/base deprotonates a

water molecule, which subsequently displaces the anomeric carboxylate in a similar reaction sequence adopted by inverting glycosidases, releasing the enzyme and the hydrolyzed product with net retention of anomeric stereochemistry.[12]

Glycosidases are further classified into two groups based on the position of hydrolysis in the substrate (Figure 1.2). Exo-glycosidases usually hydrolyze their substrates at the non-reducing end of a polysaccharide chain. The active site of exo-glycosidases is typically pocket-shaped, providing space to accommodate only the terminal monosaccharide.[13] In contrast, endo-glycosidases cleave their substrates at internal positions of the polysaccharide chain. They have a wider cleft- or tunnel-shaped active site to accommodate multiple sugar residues, often at both sites of the scissile glycosidic linkage.[13]

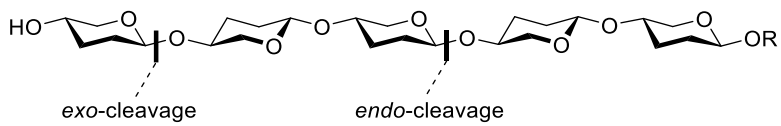


Figure 1.2. Exo- and endo-glycosidase cleavage site in the polysaccharide chain.

1.2 Covalent, irreversible, cyclophellitol-based glycosidase inhibitors

Glycosidases distort their sugar substrates during the catalytic process, and analyzing the conformational changes of substrates during the reaction itineraries they undergo can help considerably in the design of enzyme inhibitors.[3,14] For example, the glucopyranose in a β-glucoside preferably adopts a lowest energy chair ($^4C_1$) conformation in solution, but upon accommodation in the active site of a retaining β-glucosidase, it is distorted into a skew boat ($^1S_3$) conformation forming a Michaelis complex with the enzyme where the aglycon is pseudo-axially positioned to facilitate in-line attack by the catalytic nucleophile (Figure 1.3A).[15,16] During transition state formation the glucopyranose moiety of the substrate (also termed the glycon) is further distorted into a higher energy half-chair ($^4H_3$) conformation where the C5, O5, C1 and C2 atoms are coplanar, to accommodate the developing partial oxocarbenium double bond between O5 and C1. Following substitution of the aglycon, the substrate forms a covalent intermediate with the catalytic nucleophile, adopting a relaxed $^4C_1$ conformation. During the second half of the reaction the glycon conformation is further distorted in a reverse order, releasing the product with same conformation as the initial substrate. The same $^1S_3 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$ itinerary is also proposed for other retaining β-glycosidases including GH79 β-glucuronidases[17], GH10 β-xylanases[18] and GH2 β-galactosidases.[19] In contrast, many retaining α-glycosidases, such as GH31 α-glucosidases[20] and GH27 α-galactosidases,[21] employ a reverse $^4C_1 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^1S_3$ conformational itinerary (Figure 3B). It is less clear which conformational

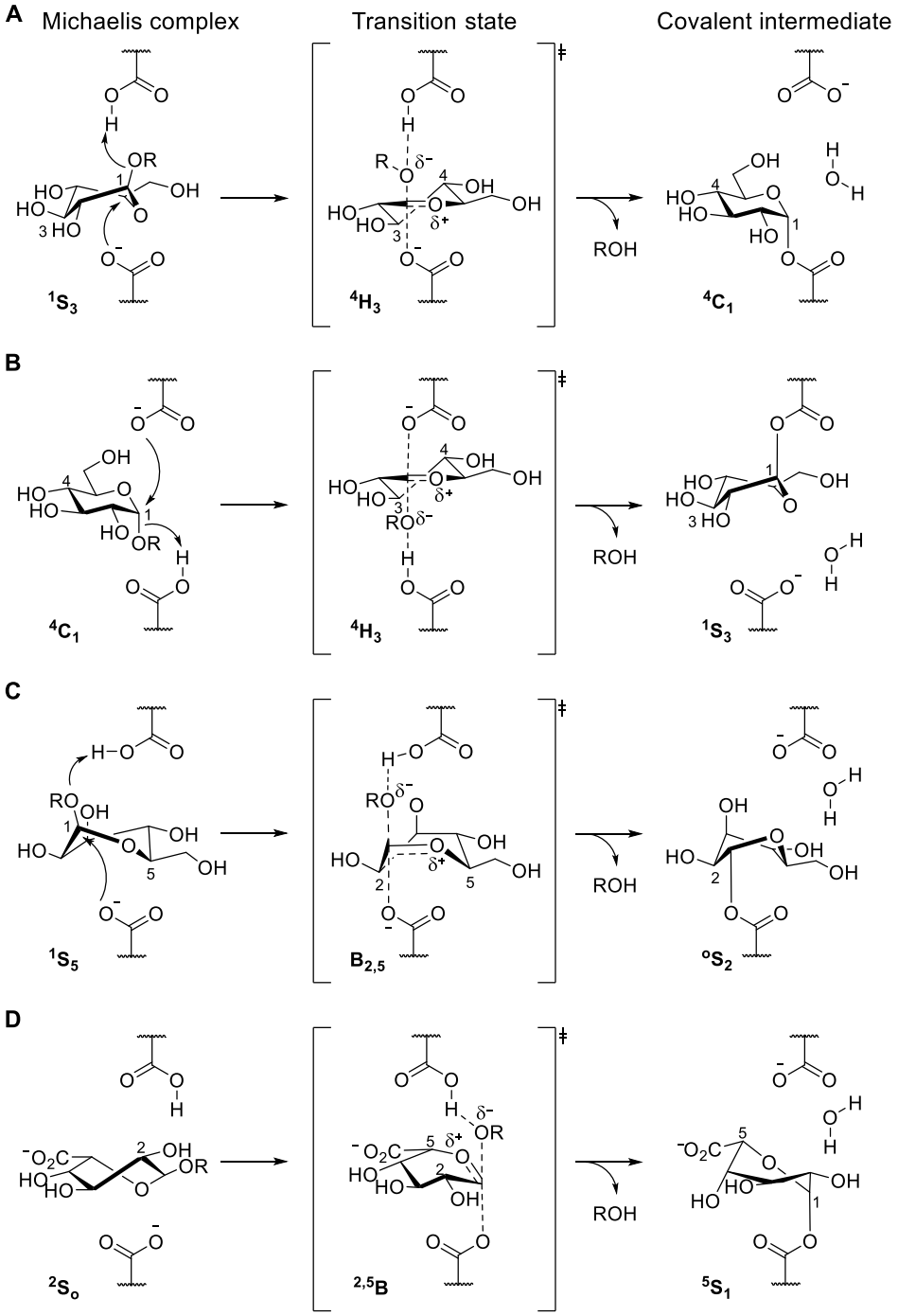


Figure 1.3. Different conformational itineraries of the substrate during hydrolysis (from initial substrate binding to covalent intermediate formation) by retaining β-D-glucosidase (A), α-D-glucosidase (B), β-D-mannosidase (C) and α-L-iduronidase (D).

itinerary the GH13 α-amylases use. Structural characterization of the covalent intermediate complexes of human pancreatic α-amylase (HPA) trapped using two different compounds – an *in-situ* generated 5-fluoro-idosyl trisaccharide (MeG2-5Fido) and a chemoenzymatically synthesized maltobiose 1,6-*epi*-cyclophellitol (both by the Withers' group) – revealed that the sugar in the -1 subsite adopts a $^4C_1$ chair conformation.[22,23] Although a computational analysis

of the conformational itineraries for a GH13 α-amylase has yet to be done, a detailed analysis of a homologous GH13 amylosucrase from *Neisseria polysaccharea* proposes a $^{4}C_{1} \rightarrow [^{4}H_{3}]^{\ddagger} \rightarrow {}^{4}C_{1}$ trajectory for the glycosylation step,[24] consistent with the observed conformation of the inhibitors bound in the active site of HPA.

There are also other glycosidase families which process their substrates following a completely different reaction itinerary.[3,14] For example, hydrolysis of a β-mannopyranoside by retaining β-mannosidases involves the formation of the Michaelis complex in an $^{1}S_{5}$ conformation (Figure 1.3C). Following a boat conformation ($B_{2,5}$) in the transition state, the covalent β-mannoside-enzyme complex is formed in a $^{o}S_{2}$ conformation. Similarly, GH38 α-mannosidases follow a reverse $^{o}S_{2} \rightarrow [B_{2,5}]^{\ddagger} \rightarrow {}^{1}S_{5}$ conformational pathway for the glycosylation step. For GH39 α-L-iduronidases the conformational change from Michaelis complex to transition state and covalent intermediate has been reported to follow a $^{2}S_{o} \rightarrow [^{2,5}B]^{\ddagger} \rightarrow {}^{5}S_{1}$ itinerary (Figure 1.3D).

Covalent, irreversible glycosidase inactivators abrogate the enzyme activity through the formation of a covalent bond between the enzyme and a reactive functionality on the inactivator.[25,26] This covalent attachment of the inactivator either blocks access to the enzyme active site or modifies an active site residue that is crucial for catalysis, thus leading to loss of enzyme activity. An important class of covalent retaining glycosidase inactivators are cyclitol epoxides such as conduritol B epoxide[27] (CBE, **1**, Figure 1.4A). CBE irreversibly inhibits β-glucosidases and (with lower potency) α-glucosidases in a mechanism-based manner.[28,29] The reactivity of CBE towards both β- and α-glucosidases can be attributed to its inherent C_{2}-symmetry in its structure, making it both a β- and an α-glucopyranose mimetic.[30] Cyclophellitol (**2**, Figure 1.4A), a natural product isolated from *Phellinus sp.*[31] and now synthetically available,[32,33] is, by virtue of its hydroxymethyl substituent, a better β-glucopyranose mimic and inhibits retaining β-glucosidases with much higher potency than CBE (lacking the C6 methylene compared to **2**).[34-36] It exhibits considerably improved selectivity towards β-glucosidases over α-glucosidases. Substitution of the epoxide oxygen with a nitrogen atom gives cyclophellitol aziridine (**3**, Figure 1.4A), which is an even more potent inhibitor of β-glucosidases.[37,38] Both compounds **2** and **3** adopt a $^{4}H_{3}$ half chair conformation, thereby mimicking the transition state conformation adopted by retaining β-glucosidase substrates during their hydrolysis.[39] Upon binding in the enzyme active site, the strained epoxide/aziridine ring is opened by nucleophilic attack at the 'anomeric carbon' by the catalytic nucleophile, forming a covalent inactivator-enzyme adduct (Figure 1.4B). Compared to the acylal linkage

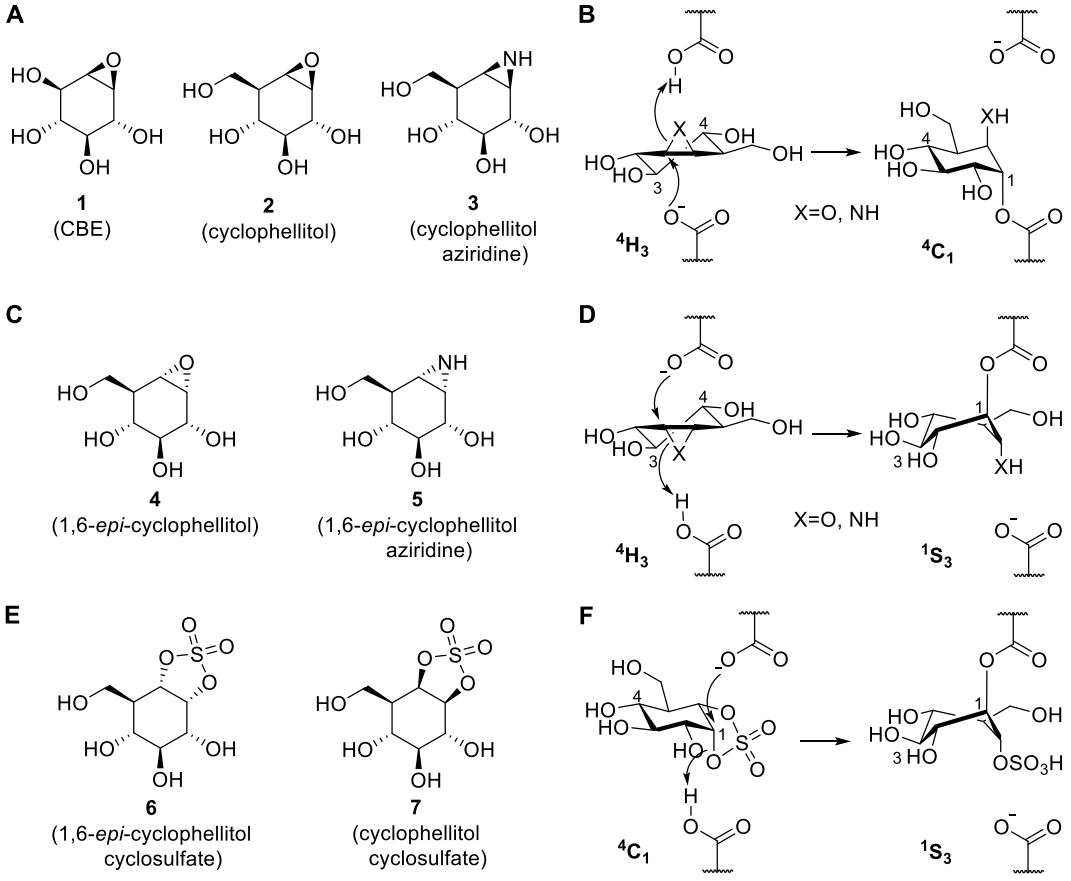


Figure 1.4. Covalent irreversible inhibitors of retaining β- and α-glucosidases and mechanisms of glucosidase inactivation. A) Structure of conduritol B epoxide (**1**), cyclophellitol (**2**), and cyclophellitol aziridine (**3**); B) Mechanism of retaining β-glucosidase inactivation by **2** and **3**; C) Structure of 1,6-*epi*-cyclophellitol (**4**) and 1,6-*epi*-cyclophellitol aziridine (**5**); D) Mechanism of retaining α-glucosidase inactivation by **4** and **5**; E) Structure of 1,6-*epi*-cyclophellitol cyclosulfate (**6**) and cyclophellitol cyclosulfate (**7**); F) Mechanism of retaining α-glucosidase inactivation by **6**.

formed during β-glucoside hydrolysis, the resulting ester linkage cannot undergo normal hydrolysis due to the absence of the endocyclic oxygen, and the enzyme is irreversibly inhibited. In analogy, α-glucopyranose-configured cyclitol epoxide and aziridine, termed 1,6-*epi*-cyclophellitol and 1,6-*epi*-cyclophellitol aziridine (**4** and **5**, Figure 1.4C)[40,41] are potent inhibitors of GH31 α-glucosidases by again mimicking the substrate $^4H_3$ transition state conformation (Figure 1.4D).[42] The scope of cyclophellitol-based covalent inhibitors was further expanded by the development of 1,6-*epi*-cyclophellitol cyclosulfate (**6**, Figure 1.4E) which is a rapid and potent irreversible inhibitor of α-glucosidases.[43] Compound **6** binds to the active site of α-glucosidases adopting a favored $^4C_1$ 'Michaelis complex like' conformation, which is perfectly poised for nucleophilic attack by the enzyme (Figure 1.4F). In contrast, the β-glucose-configured cyclophellitol cyclosulfate (**7**, Figure 1.4E) which also adopts a $^4C_1$ conformation,

is a slow and weak inactivator of β-glucosidases because the $^4C_1$ conformation does not match that of a typical β-glucosidase Michaelis complex.[43] These results illustrate that the cyclophellitol-based inhibitors' conformations substantially impact their inhibitory potency. So far, the conformational mimicry strategy based on the cyclophellitol template has been widely applied for other retaining glycosidases, including those that follow similar $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ or $^4C_1 \rightarrow {}^4H_3 \rightarrow {}^1S_3$ reaction itineraries such as GH27 α-galactosidases[44-46], GH79 β-glucuronidases[17] and GH10 β-xylanases[47], and also those following other conformational trajectories.[48-52]

1.3 Competitive glycosidase inhibitors based on transition state mimicry

The design of most competitive glucosidase inhibitors is based on mimicking either the positive charge developed along the bond between the anomeric carbon and the endocyclic oxygen or the canonical half chair conformation of the transition state (Figure 1.5A).[53,54] Iminosugars that feature a nitrogen atom in the carbohydrate-mimetic ring are perhaps the best known of the competitive glycosidase inhibitor class. Nojirimycin (NJ, **8**, Figure 1.5B), the first iminosugar isolated from a *Streptomyces* strain in 1966,[55] is a close glucopyranose analogue that has a nitrogen in place of the endocyclic oxygen. NJ is a potent inhibitor of various α- and β-glucosidases, however, the presence of the hemiaminal moiety in its structure renders this compound unstable in physiological conditions.[56] In 1967, Paulsen and co-workers reported the synthesis of 1-deoxynojirimycin (DNJ, **9**)[57] which was also later found to exist as a natural compound.[58,59] DNJ lacks the hydroxyl group at C1 and is hence much more stable than NJ while possessing similar biological properties. On the basis of DNJ, a lot of configurational analogues[60] and *N*-substituted derivatives[61] have been developed and some of them are employed for therapeutic applications. For example, the α-glucosidase inhibitor miglitol (**10**), which contains a hydroxyethyl group at the nitrogen atom, is used for the treatment of type II diabetes.[62] The synthetic isofagomine (**11**), another type of iminosugar in which the anomeric carbon is replaced by a nitrogen atom, was developed by Bols *et al.* in 1994 and was shown to be a more potent inhibitor of β-glucosidases than α-glucosidases.[63] A previous study on the thermodynamics of binding of **9** or **11** to a β-glucosidase enzyme suggested that the 2-OH of **9** contributed significantly to binding enthalpy.[64] Therefore noeuromycin **12**, an analogue of **11** where the 2-hydroxyl group is present, also synthesized by Bols and co-workers, proved to be a strong inhibitor of both α- and β-glucosidases.[65] The scope of the isofagomine-type 1-*N* iminosugars was further expanded by Ichikawa and co-workers, who showed that configurationally isosteric compounds are potent inhibitors of their corresponding β-

glycosidases.[53] It has been proposed that the protonated form of the iminosugar inhibitors to some extent mimic the positive charge developed at the endocyclic oxygen or the anomeric carbon of the glycosidase transition state.

A different approach for the design of competitive inhibitors is to mimic the conformation of the transition state, which can be achieved by introduction of bicyclic structures. Nojiritetrazole **13**, for example, has a tetrazole moiety fused onto the pseudo-glycoside ring which causes distortion of the ring into a $^4H_3$ conformation. This compound was shown to be a potent retaining β-glucosidase inhibitor.[66] Following this, a series of cyclitol-azole type derivatives, such as glucotriazole (**14**) and glucoimidazole (**15**, **16**), with varying inhibitory potencies for β-glucosidases were developed by different research groups.[53,67] Based on the $^4H_3$ conformation adopted by cyclophellitol **2** (Figure 1.4A), *carba*-cyclophellitol **17**, with the epoxide oxygen replaced by a carbon atom, proved to be a potent reversible inhibitor (K_i = 8.2 nM) of *Tm*GH1 β-glucosidase.[68] Distortion of the pseudo-glycoside ring to emulate the transition state conformation can also be done by incorporation of an unsaturated double bond. Acarbose **18**, for example, is a reversible α-glucosidase inhibitor which has been employed for clinical treatment of type II diabetes.[69] Constrained by the double bond between C5 and C7 (corresponding to the endocyclic O5 of glucopyranose), the valienamine unit in the compound may adopts a half chair conformation, which was also observed in X-ray crystallographic studies of acarbose-glucoamylase complex.[70]

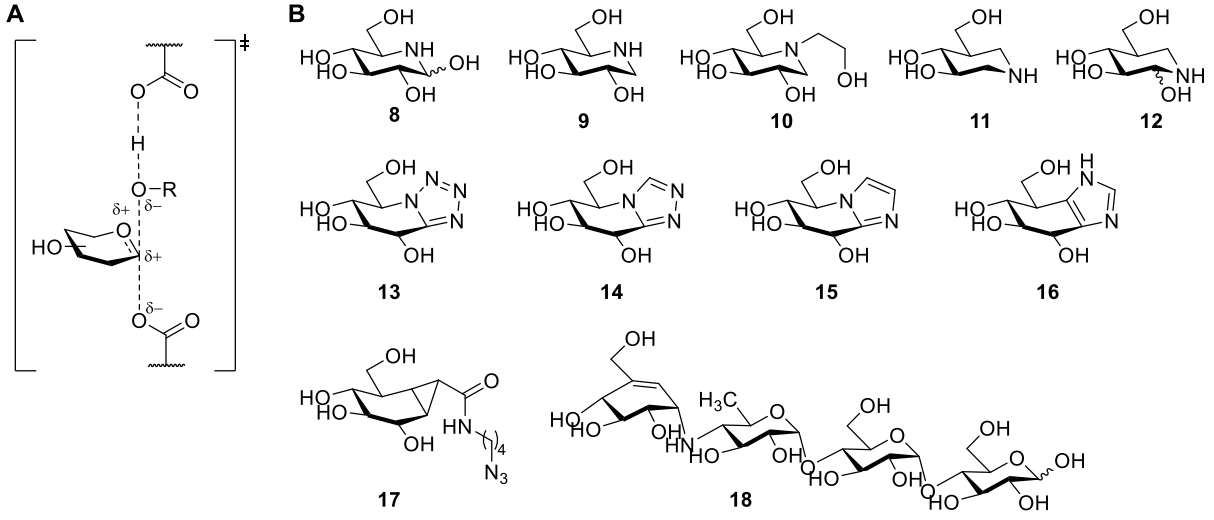


Figure 1.5. A) Transition state for the glycosylation step of a retaining glycosidase; B) Structure of nojirimycin (**8**), 1-deoxynojirimycin (**9**), miglitol (**10**), isofagomine (**11**), noeuromycin (**12**), nojiritetrazole (**13**), glucotriazole (**14**), glucoimidazoles (**15** and **16**), *carba*-cyclophellitol (**17**), and acarbose (**18**).

1.4 Activity-based protein profiling

Activity-based protein profiling (ABPP), first described by Benjamin F. Cravatt in 1999,[71] has become a powerful tool to study the functional state of enzymes in complex biological systems.[72] This technique relies on the development of suitable activity-based probes (ABPs) that irreversibly inhibit an enzyme or a class of enzymes through covalent attachment, allowing direct detection and/or identification of target enzymes. An ABP generally includes three major structural elements: a reactive moiety (or warhead) sometimes embedded in a recognition motif, a reporter group, and a linker (Figure 1.6A). Reactive moieties can be derived from known mechanism-based inhibitors that react covalently and irreversibly with the catalytic amino acids in the active site of the enzyme (class) of interest. Reporter groups are usually fluorescent dyes (for instance a BODIPY, cyanine or rhodamine fluorophore) or enrichment/capture agents (such as biotin or a bio-orthogonal group) that enable ABP-labeled proteins to be visualized and enriched for subsequent studies. A linker can be a hydrophilic or lipophilic chain to provide enough space between the reactive moiety and the reporter group. The covalent and irreversible cyclophellitol-based glycosidase inhibitors discussed in the above section have proven to be suitable warheads and introduction of a reporter entity on (generally) the aziridine nitrogen generates ABPs targeting different classes of retaining glycosidases.

Since the methods for performing ABPP experiments are diverse, this technology has been applied in various formats. For example, fluorescent probes can be used for the detection of enzyme activities. After incubation of a proteome with a fluorescent probe, the protein mixture is denatured and resolved by SDS-PAGE, and the ABP-modified proteins are then visualized by in-gel fluorescence scanning (Figure 1.6B). Since precursor enzymes, inactive or malfunctioning enzymes are generally not labeled, the band intensity can be correlated with the expression level of active enzymes. In this way, ABPs can be used as diagnostic tools to detect different enzyme activities between patients and healthy individuals.[42,73] Alternatively, when the proteome is incubated with a biotin-tagged ABP, the labeled proteins can be enriched by affinity purification using streptavidin magnetic beads (Figure 1.6C). Enriched proteins are then digested on beads by a proteolytic enzyme (usually the serine protease, trypsin) and the resulting peptide mixture analyzed by LC-MS/MS. The obtained protein sequences are matched against protein sequence databases to identify the target protein.[74] Another major application of ABPP is in the discovery of enzyme inhibitors, using methodologies termed competitive ABPP.[75,76] Compared to conventional inhibitor screening methods, competitive ABPP allows enzymes to be tested in native proteomes, without the need for purification and can be done

even if the identity of an active enzyme is unknown. In this experiment, a biological sample is pre-incubated with an inhibitor followed by labelling with a fluorescent ABP (Figure 1.6D). After denaturation and resolving by SDS-PAGE, the inhibitor potency can be read out by quantification of the reduction in probe labeling intensity. When the ABP used targets multiple enzymes in the same sample, a panel of inhibitors can be tested in parallel and the selectivity of inhibitors can be assessed.

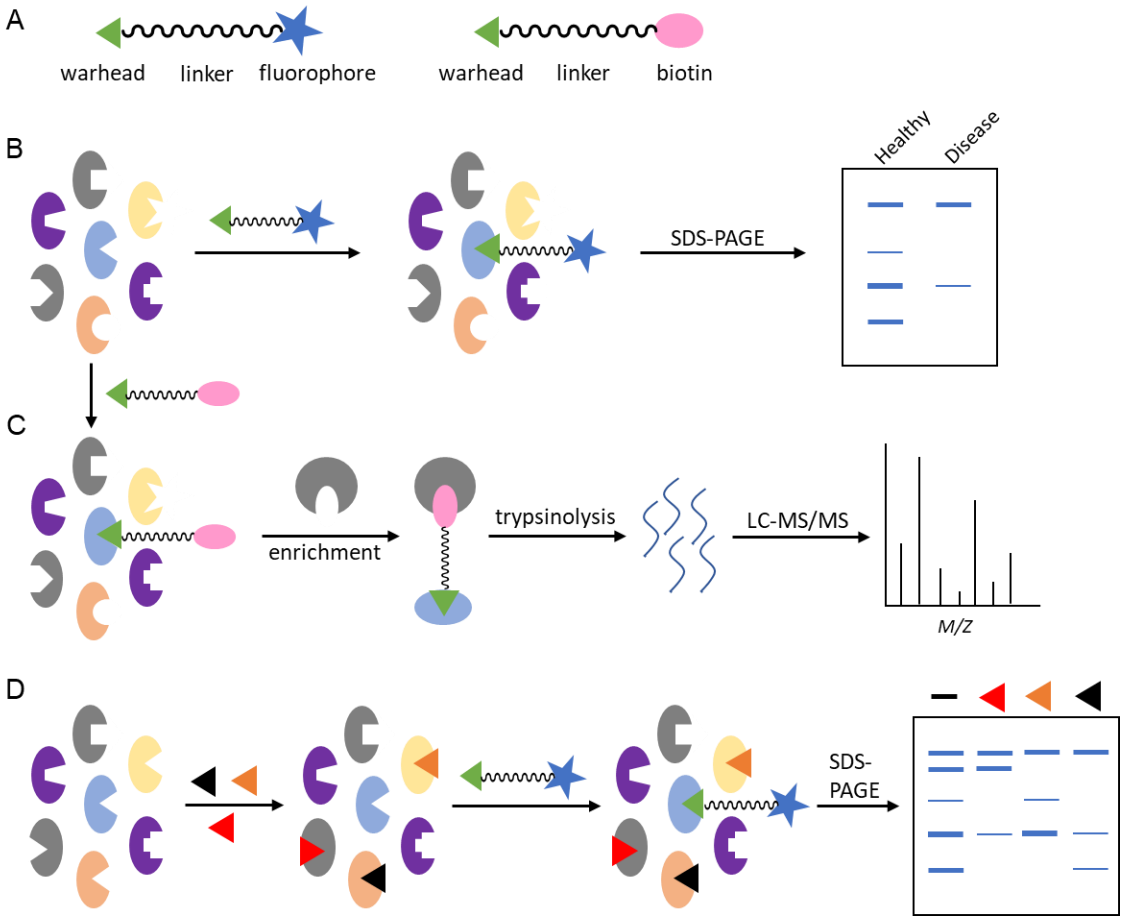


Figure 1.6. Activity-based protein profiling (ABPP). A) General structure of an activity-based probe (ABP); B) Fluorescent ABPs can be used to detect and analyze enzyme activities by SDS-PAGE in a comparative ABPP experiment; C) Identification of the target enzymes by proteomics; D) Inhibitor discovery by competitive ABPP.

1.5 Outline of this thesis

To date, cyclitol-epoxide/azirine/cyclosulfate based inhibitors and ABPs have been applied to the study of various retaining *exo*- and *endo*-acting glycosidases, and the work described in this Thesis was in part focused on expanding this to capture – with ABPP – additional endoglycosidases of biomedical and biotechnological importance. The first two chapters of this Thesis focus on the development of covalent inhibitors and ABPs targeting starch-degrading

enzymes. **Chapter 2** describes the synthesis of a set of maltobiose *epi*-cyclophellitol derived retaining α-amylase inhibitors and probes, and their labeling efficiency is biochemically evaluated on complex biological samples. **Chapter 3** reports on the synthesis of *epi*-cyclophellitol-based pseudotrisaccharides equipped with a suite of reporter entities. The construction of the branched pseudotrisaccharide probes involves selective α-1,4- and α-1,6-glycosylation of proper glucose donors on cyclohexene acceptors followed by further chemical elaboration, giving access to the desired epoxides. **Chapter 4** describes the synthesis of a set of bifunctional cyclophellitol aziridines which are functionalized at both C6 and aziridine nitrogen, as well as the biochemical evaluation of their activity and selectivity towards human lysosomal β-glucocerebrosidase. **Chapter 5** reports on the synthesis of a panel of *gluco*- and *galacto*-configured uronic acid-type 1-*N*-iminosugars as competitive heparanase inhibitors, and how the structural basis for enzyme inhibition by the synthetic iminosugars is studied by X-ray crystallography. **Chapter 6** gives a summary of the Thesis and provides some directions for future work based on the here presented results.

1.6 References

[1] Wolfenden, R.; Lu, X.; Young, G., *J. Am. Chem. Soc.* **1998**, *120* (27), 6814-6815.

[2] Sanderson, R. D.; Elkin, M.; Rapraeger, A. C.; Ilan, N.; Vlodavsky, I., *FEBS J.* **2017**, **284** (1), 42-55.

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

[4] Asano, N., *Glycobiology* **2003**, *13* (10), 93R-104R.

[5] Wu, L.; Armstrong, Z.; Schroder, S. P.; de Boer, C.; Artola, M.; Aerts, J. M.; Overkleeft, H. S.; Davies, G. J., *Curr. Opin. Chem. Biol.* **2019**, *53*, 25-36.

[6] IUBMB: *Enzyme Nomenclature. Recommendations.* San Diego: Academic Press; 1992.

[7] Henrissat, B., *Biochem. J.* **1991**, *280*, 309-316.

[8] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., *Nucleic Acids Res.* **2014**, *42* (D1), D490-D495.

[9] Henrissat, B.; Davies, G. J., *Curr. Opin. Struct. Biol.* **1997**, *7*, 637-344.

[10] Davies, G. J.; Sinnott, M. L., *Biochem. J.* **2008**, *30* (4), 26-32.

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

[12] Zechel, D. L.; Withers, S. G., *Acc. Chem. Res.* **2000**, *33* (1), 11-18.

[13] Davies, G. J.; Henrissat, B., *Structure* **1995**, *3*, 853-859.

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

[15] Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J., *Biochemistry* **1996**, *35*, 15280-15287.

[16] Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G., *Biochemistry* **1998**, *37*, 11707-11713.

[17] Wu, L.; Jiang, J.; Jin, Y.; Kallemeijn, W. W.; Kuo, C.-L.; Artola, M.; Dai, W.; van Elk, C.; van Eijk, M.; van der Marel, G. A.; Codée, J. D. C.; Florea, B. I.; Aerts, J. M. F. G.; Overkleeft, H. S.; Davies, G. J., *Nat. Chem. Biol.* **2017**, *13*, 867-873.

[18] Suzuki, R.; Fujimoto, Z.; Ito, S.; Kawahara, S.; Kaneko, S.; Taira, K.; Hasegawa, T.; Kuno, A., *J. Biochem.* **2009**, *146* (1), 61-70.

[19] Juers, D. H.; Heightman, T. D.; Vasella, A.; McCarter, J. D.; Mackenzie, L.; Withers, S. G.; Matthews, B. W., *Biochemistry* **2001**, *40*, 14781-14794.

[20] Lovering, A. L.; Lee, S. S.; Kim, Y.-W.; Withers, S. G.; Strynadka, N. C. J., *J. Biol. Chem.* **2005**, *280* (3), 2105-2115.

[21] Guce, A. I.; Clark, N. E.; Salgado, E. N.; Ivanen, D. R.; Kulminskaya, A. A.; Brumer, H.; Garman, S. C., *J. Biol. Chem.* **2010**, *285* (6), 3625-3632.

[22] Zhang, R.; Li, C.; Williams, L. K.; Rempel, B. P.; Brayer, G. D.; Withers, S. G., *Biochemistry* **2009**, *48* (45), 10752-10764.

[23] Caner, S.; Zhang, X.; Jiang, J.; Chen, H. M.; Nguyen, N. T.; Overkleeft, H.; Brayer, G. D.; Withers, S. G., *FEBS Lett.* **2016**, *590* (8), 1143-1151.

[24] Alonso-Gil, S.; Coines, J.; Andre, I.; Rovira, C., *Front. Chem.* **2019**, *7*, 1-10.

[25] Withers, S. G.; Aebersold, R., *Protein Sci.* **1995**, *4*, 361-372.

[26] Rempel, B. P.; Withers, S. G., *Glycobiology* **2008**, *18* (8), 570-586.

[27] Legler, G., *Hoppe. Seylers. Z. Physiol. Chem.* **1966**, *345*(4), 197-214.

[28] Premkumar, L.; Sawkar, A. R.; Boldin-Adamsky, S.; Toker, L.; Silman, I.; Kelly, J. W.; Futerman, A. H.; Sussman, J. L., *J. Biol. Chem.* **2005**, *280* (25), 23815-23819.

[29] Yang, S.-J.; Ge, S.-G.; Zeng, Y.-C.; Zhang, S.-Z., *Biochimica. et. Biophysica. Acta.* **1985**, *828*, 236-240.

[30] Braun, H.; Legler, G.; Deshusses, J.; Semenza, G., *Biochimica. et. Biophysica. Acta.* **1977**, *483*, 135-140.

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

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

[33] Li, K.-Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; van den Elst, H.; Wong, C.-S.; Chander, S. D.; Hoogendoorn, S.; Beenakker, T. J. M.; Codée, J. D. C.; Aerts, J. M. F. G.; van der Marel, G. A.; Overkleeft, H. S., *Eur. J. Org. Chem.* **2014**, *2014* (27), 6030-6043.

[34] Atsumi, S.; Iinuma, H,; Nosaka, C.; Umezawa, K., *J. Antibiot.* **1990**, *43*, 1579-1585.

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

[36] Withers, S. G.; Umezawa, K., *Biochem. Biophys. Res. Commun.* **1991**, *177*, 532-537.

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

[38] Li, K. Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; Donker-Koopman, W. E.; Boot, R. G.; Aerts, J. M.; Codée, J. D.; van der Marel, G. A.; Overkleeft, H. S., *Org. Biomol. Chem.* **2014**, *12* (39), 7786-7791.

[39] Gloster, T. M.; Madsen, R.; Davies, G. J., *Org. Biomol. Chem.* **2007**, 5 (3), 444-446.

[40] Tatsuta, K.; Niwata, Y.; Umezawa, K.; Toshima, K.; Nakata, M., *Carbohydr. Res.* **1991**, *222*, 189-203.

[41] 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.; Overkleeft, H. S., *Eur. J. Org. Chem.* **2016**, *2016* (22), 3671-3678.

[42] 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.; Boot, R. G.; Davies, G. J.; Overkleeft, H. S.; Aerts, J. M., *ACS Cent. Sci.* **2016**, *2* (5), 351-358.

[43] Artola, M.; Wu, L.; Ferraz, M. J.; Kuo, C. L.; Raich, L.; Breen, I. Z.; Offen, W. A.; Codée, J. D. C.; van der Marel, G. A.; Rovira, C.; Aerts, J.; Davies, G. J.; Overkleeft, H. S., *ACS Cent. Sci.* **2017**, *3* (7), 784-793.

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

[45] Willems, L. I.; Beenakker, T. J.; Murray, B.; Scheij, S.; Kallemeijn, W. W.; Boot, R. G.; Verhoek, M.; Donker-Koopman, W. E.; Ferraz, M. J.; van Rijssel, E. R.; Florea, B. I.; Codée, J. D.; van der Marel, G. A.; Aerts, J. M.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2014**, *136* (33), 11622-11625.

[46] Artola, M.; Hedberg, C.; Rowland, R. J.; Raich, L.; Kytidou, K.; Wu, L.; Schaaf, A.; Ferraz, M. J.; van der Marel, G. A.;

Codée, J. D. C.; Rovira, C.; Aerts, J. M. F. G.; Davies, G. J.; Overkleeft, H. S., *Chem. Sci.* **2019**, *10* (40), 9233-9243.

[47] Schröder, S. P.; de Boer, C.; McGregor, N. G. S.; Rowland, R. J.; Moroz, O.; Blagova, E.; Reijngoud, J.; Arentshorst, M.; Osborn, D.; Morant, M. D.; Abbate, E.; Stringer, M. A.; Krogh, K.; Raich, L.; Rovira, C.; Berrin, J. G.; van Wezel, G. P.; Ram, A. F. J.; Florea, B. I.; van der Marel, G. A.; Codée, J. D. C.; Wilson, K. S.; Wu, L.; Davies, G. J.; Overkleeft, H. S., *ACS Cent. Sci.* **2019**, *5* (6), 1067-1078.

[48] Jiang, J.; Kallemeijn, W. W.; Wright, D. W.; van den Nieuwendijk, A.; 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.; Overkleeft, H. S., *Chem. Sci.* **2015**, *6* (5), 2782-2789.

[49] Artola, M.; Kuo, C. L.; McMahon, S. A.; Oehler, V.; Hansen, T.; van der Lienden, M.; He, X.; van den Elst, H.; Florea, B. I.; Kermode, A. R.; van der Marel, G. A.; Gloster, T. M.; Codée, J. D. C.; Overkleeft, H. S.; Aerts, J., *Chem. Eur. J.* **2018**, *24* (71), 19081-19088.

[50] McGregor, N. G. S.; Artola, M.; Nin-Hill, A.; Linzel, D.; Haon, M.; Reijngoud, J.; Ram, A.; Rosso, M. N.; van der Marel, G. A.; Codée, J. D. C.; van Wezel, G. P.; Berrin, J. G.; Rovira, C.; Overkleeft, H. S.; Davies, G. J., *J. Am. Chem. Soc.* **2020**, *142* (10), 4648-4662.

[51] Armstrong, Z.; Kuo, C. L.; Lahav, D.; Liu, B.; Johnson, R.; Beenakker, T. J. M.; de Boer, C.; Wong, C. S.; van Rijssel, E. R.; Debets, M. F.; Florea, B. I.; Hissink, C.; Boot, R. G.; Geurink, P. P.; Ovaa, H.; van der Stelt, M.; van der Marel, G. M.; Codée, J. D. C.; Aerts, J.; Wu, L.; Overkleeft, H. S.; Davies, G. J., *J. Am. Chem. Soc.* **2020**, *142* (30), 13021-13029.

[52] McGregor, N. G. S.; Coines, J.; Borlandelli, V.; Amaki, S.; Artola, M.; Nin-Hill, A.; Linzel, D.; Yamada, C.; Arakawa, T.; Ishiwata, A.; Ito, Y.; van der Marel, G. A.; Codée, J. D. C.; Fushinobu, S.; Overkleeft, H. S.; Rovira, C.; Davies, G. J., *Angew. Chem. Int. Ed. Engl.* **2021**, *60* (11), 5754-5758.

[53] Lillelund, V. H.; Jensen, H. H.; Liang, X.; Bols, M., *Chem. Rev.*, 2002, **102**, 515-553.

[54] Gloster, T. M.; Davies, G. J., *Org. Biomol. Chem.* **2010**, *8* (2), 305-320.

[55] Inouye, S.; Tsuroka, T.; Nida, T.; *J. Antibiot.*, **1966**, *19*, 288-292.

[56] Inouye, S.; Tsuroka, T.; Ito, T.; Nida, T., *Tetrahedron* **1968**, *24*, 2124-2144.

[57] Paulsen, H.; Sangster, I.; Heyns, K., *Chem. Ber.*, **1967**, *100*, 802-815.

[58] Yagi, M.; Kouno, T.; Aoyagi, Y.; Murai, H., *Nippon Nogei Kagaku Kaishi* **1976**, *50*, 571-572.

[59] Schmidt, D. D.; Frommer, W.; Müller, L.; Truscheit, E., *Naturwissenschaften* **1979***, 66,* 584-585.

[60] Legler, G., *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 319-384.

[61] Afarinkia, K.; Bahar, A., Tetrahedron: Asymmetry **2005**, *16* (7), 1239-1287.

[62] Scott, L. J.; Spencer, C. M., *Drugs* **2000**, *59*, 521-549.

[63] Jespersen, T. M.; Bols, M., *Tetrahedron* **1994**, *50*, 13449-13460.

[64] Bülow, A.; Plesner, I. W.; Bols, M., *J. Am. Chem. Soc.* **2000**, *122*, 8567-8568.

[65] Liu, H.; Liang, X.; Søhoel, H.; Bülow, A.; Bols, M., *J. Am. Chem. Soc.* **2001**, *123,* 5116-5117.

[66] Ermert, P.; Vasella, A.; Weber, M.; Rupitz, K.; Withers, S. G., *Carbohydr. Res.*, **1993**, *250*, 113-128.

[67] Schröder, S. P.; Wu, L.; Artola, M.; Hansen, T.; Offen, W. A.; Ferraz, M. J.; Li, K. Y.; Aerts, J.; van der Marel, G. A.; Codée, J. D. C.; Davies, G. J.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2018**, *140* (15), 5045-5048.

[68] Beenakker, T. J. M.; Wander, D. P. A.; Offen, W. A.; Artola, M.; Raich, L.; Ferraz, M. J.; Li, K. Y.; Houben, J.; van Rijssel, E. R.; Hansen, T.; van der Marel, G. A.; Codée, J. D. C.; Aerts, J.; Rovira, C.; Davies, G. J.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2017**, *139* (19), 6534-6537.

[69] Campbell, L. K.; White, J. R.; Campbell, R. K., *Ann. Pharmacother.* **1996**, *30* (11), 1255-1262.

[70] Aleshin, A. E.; Stoffer, B.; Firsov, L. M.; Svensson, B.; Honzatko, R. B., *Biochemistry* **1996**, *35*, 8319-8328.

[71] Liu, Y.; Patricelli, M. P.; Cravatt, B. F., *Proc. Natl. Acad. Sci.* **1999**, *96*, 14694-14699.

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

[73] Yang, Y.; Yang, X.; Verhelst, S. H., *Molecules* **2013**, *18* (10), 12599-12608.

[74] Huber, K. V. M.; Superti-Furga, G., *Methods Mol. Biol.* **2016**, *1394*, 211-218.

[75] Niphakis, M. J.; Cravatt, B. F., *Annu. Rev. Biochem.* **2014**, *83*, 341-377.

[76] Deu, E.; Yang, Z.; Wang, F.; Klemba, M.; Bogyo, M., *PLoS One* **2010**, *5* (8), e11985.

2

Activity-Based Protein Profiling of Retaining α-Amylases in Complex Biological Samples

Manuscript published as:

Yurong Chen,[*] Zachary Armstrong,[*] Marta Artola, Bogdan I. Florea, Chi-Lin Kuo, Casper de Boer, Mikkel S. Rasmussen, Maher Abou Hachem, Gijsbert A. van der Marel, Jeroen D. C. Codée, Johannes M. F. G. Aerts, Gideon J. Davies, and Herman S. Overkleeft, *J. Am. Chem. Soc.* **2021**, *143*, 2423-2432.

2.1 Introduction

Retaining *α*-amylases, belonging to the glycoside hydrolase (GH) 13 family (www.cazy.org),[1] are starch-processing enzymes present in plants, animals and microorganisms.[2-4] They catalyze the hydrolysis of internal α-1,4-glycosidic linkages in starch and related polysaccharides through a two-step Koshland double displacement mechanism (Figure 2.1). In this mechanism, the aglycon is protonated by the general acid/base residue, concomitant with an S_N2 nucleophilic attack by the catalytic nucleophile at the anomeric center to form a covalent enzyme-substrate intermediate with inversion of anomeric configuration. The covalent enzyme-substrate adduct is subsequently hydrolyzed to yield oligosaccharide products (for instance linear and alpha-1,6-branched *malto*-oligosaccharides) with net retention of α-anomeric configuration.[5]

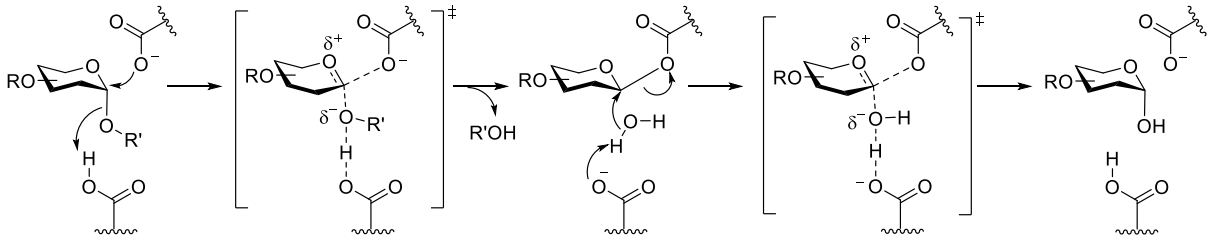


Figure 2.1. General Koshland double displacement mechanism employed by retaining α-amylases.

Starch is the main product of major human and animal feedstock crops including wheat, rice, maize, tapioca and potato. Traditionally, large-scale starch processing, an industry established over a hundred years ago, relied on the degradation of starch by acid hydrolysis.[6,7] Amylases have gained considerable traction as alternatives to conventional chemical starch hydrolysis, as they are more environmentally benign, are less energy intensive and moreover may produce more defined oligosaccharide products from starch polymers.[8,9] Today, a large number of α-amylases from microbial sources are commercially available and are widely used for the production of high glucose and fructose syrups.[10,11] Another major application of α-amylases is found in the detergent industry, where they act at low temperatures and alkaline pH to remove tough starch stains from clothes and porcelains under mild conditions.[12] α-Amylases are furthermore extensively employed in textile, food, paper and biofuel industries. The α-amylases currently applied in industrial biotechnology originate from a limited number of fungal and bacterial strains, including *Aspergillus oryzae* and *Bacillus licheniformis*.[11,13] The microbial world however may harbor many additional α-amylases including ones that differ functionally in terms of activity, substrate specificity and resistance to (harsh) biotechnological conditions.

In addition to their potential in biotechnological applications, α-amylases have received increased interest in the past decades in the context of human health. In mammals, digestion of starch by salivary and pancreatic α-amylases leads to linear and branched *malto*-oligosaccharides. These are largely hydrolyzed to glucose by α-glucosidases located in the intestinal mucosa, which then enters the blood stream by means of facilitated diffusion. Control of postprandial glucose levels, which has proven to be a key factor for the treatment of diabetes and obesity,[14] might be achieved by activity modulation of starch-processing enzymes. Anti-diabetic drugs currently used in the clinic include the α-glucosidase inhibitors,[15,16] miglitol, voglibose and acarbose, the latter of which also inhibits human pancreatic α-amylase (HPA).[17,18] However, their effectiveness is offset by undesirable side effects, including diarrhea, flatulence, and abdominal pain, secondary effects that arise predominantly from inhibition of intestinal α-glucosidases. Hence, selective inhibitors targeting exclusively HPA, the first enzyme in the digestion sequence, would be of great interest.[19-24]

Activity-based protein profiling (ABPP) has emerged as a powerful strategy to obtain qualitative and quantitative activity information on enzymes of interest in complex mixtures.[25,26] ABPP is especially useful for retaining glycosidases that form a covalent enzyme–substrate intermediate during hydrolysis. It relies on the availability of suitable activity-based probes (ABPs) that are generally generated from covalent and irreversible inhibitors by equipping these with a reporter entity (a fluorophore or biotin) or a bio-orthogonal tag. By means of ABPP, enzyme activities can be rapidly identified and quantified (comparative ABPP) and the effect of inhibitors on target enzymes studied in cell extracts and living cells (competitive ABPP). Previous work on retaining glycosidase ABPP is rooted in the known mechanism-based *β*-glucosidase inhibitor cyclophellitol, isolated from *Phellinus sp.*[27] In the first instance cyclophellitol-derived ABPs were applied to the study of retaining *exo*-acting *α*- and *β*-glycosidases.[28-32] More recently, this methodology was expanded towards retaining endoglycosidases in work on xylobiose-configured ABPs to investigate *endo*-acting xylanases in *Aspergillus* secretomes.[33,34] This chapter addresses the efficacy of retaining glycosidase ABPP by the design and chemical synthesis of a panel of maltobiose *epi*-cyclophellitol derived retaining α-amylase inhibitors and ABPs, the kinetic and structural analysis of their interactions with microbial α-amylases, and the application of these probes for rapid detection and identification of α-amylases in fungal secretomes, mouse tissue and human saliva. The described work complements a recent study by Withers and colleagues,[24] who reported a chemo-enzymatic synthesis of maltobiose *epi*-cyclophellitol, to which the chemically derived

material and as well the structural work on α-amylase-inhibitor complexes could be matched. The results described in this chapter moreover illustrate the potential of retaining α-amylase ABPP in the annotation of α-amylases in complex biological samples.

2.2 Results and discussion

2.2.1 Design and synthesis of α-amylase inhibitors and activity-based probes

1,6-*Epi*-cyclophellitol, as well as its aziridine and cyclic sulfate analogues (Figure 2.2A) are potent mechanism-based α-glucosidase inhibitors, and it has recently become clear that their tagged versions are effective activity-based retaining α-exoglucosidase probes.[30,35] These monosaccharidic cyclitols however do not inhibit *endo*-acting α-amylases, which arguably is because of the absence of productive binding with the larger *endo*-glycosidase active sites (Chapter 1). In line with this consideration, Withers and co-workers subsequently reported that the maltose disaccharide mimetic α-1,4-glucopyranosyl *epi*-cyclophellitol (compound **1a**, Figure 2.2B), which they prepared by enzymatic α-glucosylation of 1,6-*epi*-cyclophellitol, is an effective mechanism-based, covalent and irreversible HPA inhibitor.[24] Crystallographic analysis of an HPA-inhibitor complex points to the O4' of the nonreducing end sugar,[24] which may be alkylated to block potential exoglycosidase action, as a promising site for orthogonal functionalization (as in **1b**) to allow introduction of a reporter entity (a fluorophore or biotin, as in **1c** and **1d**). Alternatively, and in line with previous[33] work on xylanase ABPs maltose-configured cyclophellitol aziridines can be designed bearing the reporter tag on the aziridine nitrogen, either with or without an O4' cap (compounds **2a**-**2c**, Figure 2.2B). Finally, and in line with previous results on 1,6-*epi*-cyclophellitol cyclosulfate being a potent and selective retaining α-exoglucosidase inhibitor,[30,35] it was decided to synthesize the corresponding maltose cyclosulfate (compound **3a**, Figure 2.2B) as well, with the aim to investigate whether this Michaelis complex emulating inhibitor would also act on amylases.

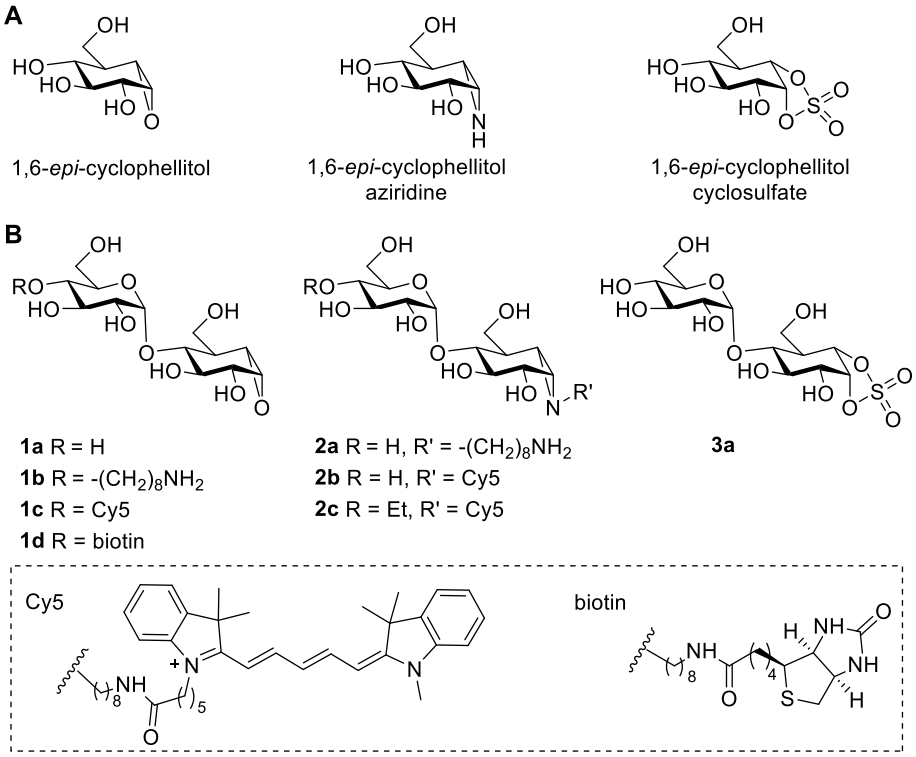
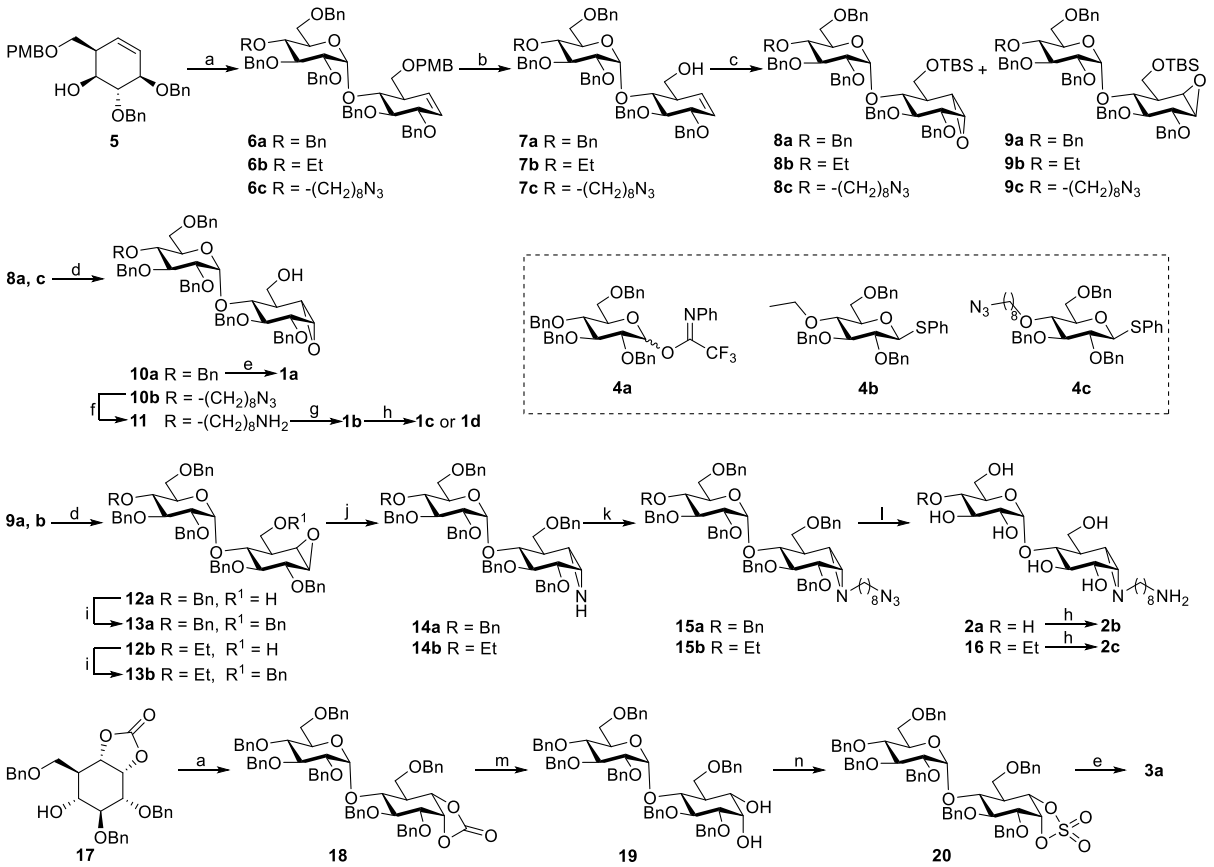


Figure 2.2. Design of mechanism-based retaining *α*-glucosidase inhibitors and ABPs. A) Structures of retaining α-exoglucosidase inhibitors; B) Structures of retaining α-endoglucosidase inhibitors and ABPs, which are the subject of the described studies in this chapter.

The recently reported synthesis of xylobiose-cyclophellitols was accomplished via the direct glycosylation of xylo-cyclophellitol (aziridine) acceptors.[33] This route however appeared not feasible for the construction of compounds **1a-d** since the epoxide functionality in partially protected 1,6-*epi*-cyclophellitols proved unstable towards stoichiometric amounts of Lewis acid under the conditions required to effect formation of α-1,4-glycosidic bonds using contemporary glycosylation conditions. Alternatively, cyclohexene **5** was found to readily undergo glycosylation with imidate donor **4a** under *N*,*N*-dimethyl formamide (DMF) and trifluoromethanesulfonic acid (TfOH) activating conditions, affording pseudodisaccharide **6a** in good yield.[36] Selective deprotection of the 4-methoxybenzyl group in **6a** followed by epoxidation afforded an inseparable mixture of *α/β*- epoxides in a 1:3 ratio, which were silylated and separated by silica gel column chromatography to yield pure **8a** and **9a**. Desilylation and global debenzylation of **8a** by Pearlman's catalyst hydrogenation gave maltobiose *epi*-cyclophellitol **1a**, of which all analytical and spectroscopical data matched those of the chemo-enzymatically produced compound.[24] The epoxide in **9a** was converted to aziridine **14a** by treatment with sodium azide followed by a Staudinger-type ring closure. Alkylation and global

deprotection of **14a** under Birch conditions afforded alkyl-aziridine **2a**, which was equipped with a Cy5 fluorophore to produce ABP **2b** (Scheme 2.1).



Scheme 2.1. Synthesis of the inhibitors and ABPs subject of the here-described studies. Reagents and conditions: a) for **6a** and **18**: **4a**, TfOH (1.5 eq), DMF, DCM, 4 Å MS, -20 °C to 0 °C, **6a** 96%, **18** 57%; for **6b**: **4b**, NIS, TfOH (1.5 eq), DMF, DCM, 4 Å MS, 0 °C, 88%; for **6c**: **4c**, NIS, TfOH (1.5 eq), DMF, DCM, 4 Å MS, 0 °C, 87%; b) DDQ, DCM/H_2O (19/1), rt, **7a** 80%, **7b** 74%, **7c** 71%; c) *i*) *m*CPBA, DCM, 0 °C to rt; *ii*) TBSCl, DMAP, imidazole, DCM, rt, **8a** 20%, **9a** 58%, **8b** 21%, **9b** 65%, **8c** 22%, **9c** 63%; d) TBAF, THF, rt, **10a** 90%, **10b** 87%, **12a** 91%, **12b** 88%; e) $Pd(OH)_2$/C, H_2, MeOH/H_2O/dioxane (2/1/2), rt, **1a** 100%, **3a** 89%; f) polymer-bound PPh_3, MeCN, H_2O, 70 °C, 79%; g) $Pd(OH)_2$/C, H_2, HOAc, ^{t}BuOH/H_2O/dioxane (1/2/1), rt, 91%; h) Cy5-OSu or biotin-OSu, DIPEA, DMF, rt, **1c** 33%, **1d** 16%, **2b** 10%, **2c** 17%; i) BnBr, NaH, TBAI, DMF, 0 °C to rt, **13a** 82%, **13b** 75%; j) *i*) NaN_3, $LiClO_4$, DMF, 100 °C; *ii*) polymer-bound PPh_3, MeCN, 60 °C, **14a** 48%, **14b** 42%; k) 8-azidooctyltrifluoromethanesulfonate, DIPEA, DCM, 0 °C to rt, **15a** 90%, **15b** 88%; l) Na, NH_3 (liq.), ^{t}BuOH, THF, -60 °C, **2a** 79%, **16** 94%; m) NaOMe, DCM/MeOH (1/1), rt, 93%; n) *i*) $SOCl_2$, TEA, DCM, 0 °C; *ii*) $RuCl_3.3H_2O$, $NaIO_4$, EtOAc/ACN/H_2O (2/2/1), 0 °C, 58%.

Previous studies on *endo*-xylanases revealed that the *xylobiose*-cyclophellitol aziridine probe was susceptible to hydrolysis by *exo*-xylosidases present in *Aspergillus* secretomes.[33] Thus, it was hypothesized that "capping" the O4' position of maltobiose *epi*-cyclophellitol azirines with

an ethyl group would prevent cleavage by *exo*-acting α-glucosidases in complex biological samples. This was achieved by using O4 ethyl capped thiophenyl donor **4b** for glycosylation with acceptor **5** under *N*-iodosuccinimide (NIS)/DMF/TfOH catalysis, giving 4'-ethyl pseudodisaccharide **6b** in good yield. Following the same route as for **2a**, 4'-ethyl capped maltobiose *epi*-cyclophellitol aziridine **16** could be readily obtained from intermediate **9b**, which was equipped with a Cy5 fluorophore to produce ABP **2c**. We constructed epoxide-based ABPs by using thiophenyl donor **4c**, where the azidooctyl group was introduced in an early stage for later installation of reporter entities and also to prevent *exo*-glycosidase processing. Following the established route, epoxide **1b** was obtained after desilylation, azide reduction and debenzylation of **8c**. Following coupling with a Cy5 fluorophore or biotin moiety, ABPs **1c** and **1d** were obtained respectively.

The same strategy was also applied to construct maltobiose *epi*-cyclophellitol cyclosulfate **3a**. Carbonate acceptor **17** was readily glycosylated with donor **4a** under TfOH/DMF conditions, giving pseudodisaccharide **18** in moderate yield. Treatment of **18** with sodium methoxide gave diol **19**. Generation of the cyclosulfite using thionyl chloride and subsequent oxidation gave perbenzylated cyclosulfate **20**, the benzyl ethers of which were removed by Pearlman's catalyst hydrogenation to afford inhibitor **3a**.

2.2.2 Kinetic studies for time-dependent inhibition of *Aspergillus oryzae* α-amylase by maltobiose inhibitors

To assess the potency of the synthesized maltobiose *epi*-cyclophellitols, the time-dependent inhibition of *Aspergillus oryzae* α-amylase (hereafter termed as Taka-amylase) was monitored. Taka-amylase, one of the earliest amylases to be characterized and used commercially,[37,38] is a retaining α-amylase belonging to GH13 subfamily 1. This enzyme uses a pair of catalytic active site residues – Asp206 as a nucleophile and Glu230 as the acid/base – to cleave amylose chains and requires several subsites to be occupied to catalyse hydrolysis at meaningful rates.[39]

To assess time-dependent inhibition, the hydrolysis of 2-chloro-4-nitrophenyl maltotrioside (CNP-M3) by Taka-amylase was measured after pre-incubation with 2 mM of inhibitors **1a**, **1b, 2a** and **3a**. All four inhibitors displayed a time-dependent inhibition of Taka-amylase, however the rates of inhibition varied greatly (Figure 2.3). Both the *epi*-cyclophellitol epoxide **1a** and aziridine **2a** had slow rates of inactivation, with rate constants of 0.0061 ± 0.0002 min^{-1} and 0.005 ± 0.002 min^{-1} respectively. The rate constant for the inactivation by **1a** is somewhat higher than was previously observed for its inhibition of HPA, determined to be 0.001 min^{-1} at

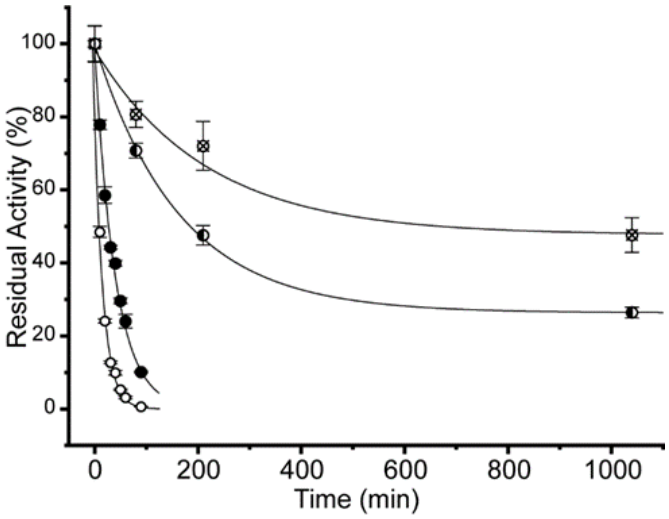


Figure 2.3. Time-dependent inactivation of Taka-amylase. Residual activity of Taka-amylase incubated with 2 mM of either **1a** (◐), **1b** (○), **2a** (⊗) or **3a** (●) was assayed over time and fit to an exponential decay curve.

a higher concentration of inhibitor (3.6 mM).[24] It is somewhat surprising that the aziridine had a rate of inhibition slightly worse than that of the epoxide, as previous work has demonstrated aziridine based inhibitors to be more potent for α-glucosidases.[30] It was speculated that *N*-alkylation of the aziridine may reduce the initial binding of inhibitor **2a**, consequently reducing inactivation rates. Inhibition by the cyclosulfate inhibitor **3a** was 4-5 fold faster (rate constant = 0.0250 ± 0.0009 min^{-1}) when compared to **1a** or **2a**. This observation agrees well with our previous work, which showed that the monosaccharide *epi*-cyclophellitol cyclosulfate was indeed the most potent cyclophellitol type inhibitor of recombinant lysosomal acid α-glucosidase.[35] Surprisingly, 4'-alkylamine **1b** proved to be the most effective inhibitor tested, with an inhibition rate constant that was nearly 10 times that of the untagged **1a** (rate constant = 0.057 ± 0.003 min^{-1}). This suggests that initial binding of **1b** is aided by interactions between the -3 subsite and the 4'-alkylamine present, which in turn increases the rate of inhibition. Previous kinetic analysis of Taka-amylase catalytic rates has shown dramatic rate increases for the hydrolysis of a maltotetrasaccharide over a maltotrioside,[39] highlighting the importance of binding in the -3 position with a natural substrate.

2.2.3 Structural analysis of inhibitor-enzyme complexes

X-ray crystallography studies were conducted next to further understand the binding of inhibitors **1a**, **1b**, **2a** and **3a** to the active site of Taka-amylase. All four inhibitors formed covalent complexes with Taka-amylase, solved with resolution between 1.67 and 1.35 Å, and were covalently linked to the enzyme through the catalytic nucleophile Asp206 (Figure 2.4).

The slow reaction rates necessitated prolonged incubation of the inhibitors with pre-formed crystals. The *epi*-cyclophellitol epoxides were soaked for 3 days. At this stage **1a** reached 80% active site occupancy, while **1b** had complete occupancy, reflective of its increased rate constant of inhibition when compared to **1a**. After a week-long soak with **2a** Taka-amylase crystals had approximately 70% inhibitor occupancy, thus soak length for this compound was further extended to 1 month after which complete occupancy was observed. The soak with **3a** was also extended to 1 month to allow for complete active site occupancy.

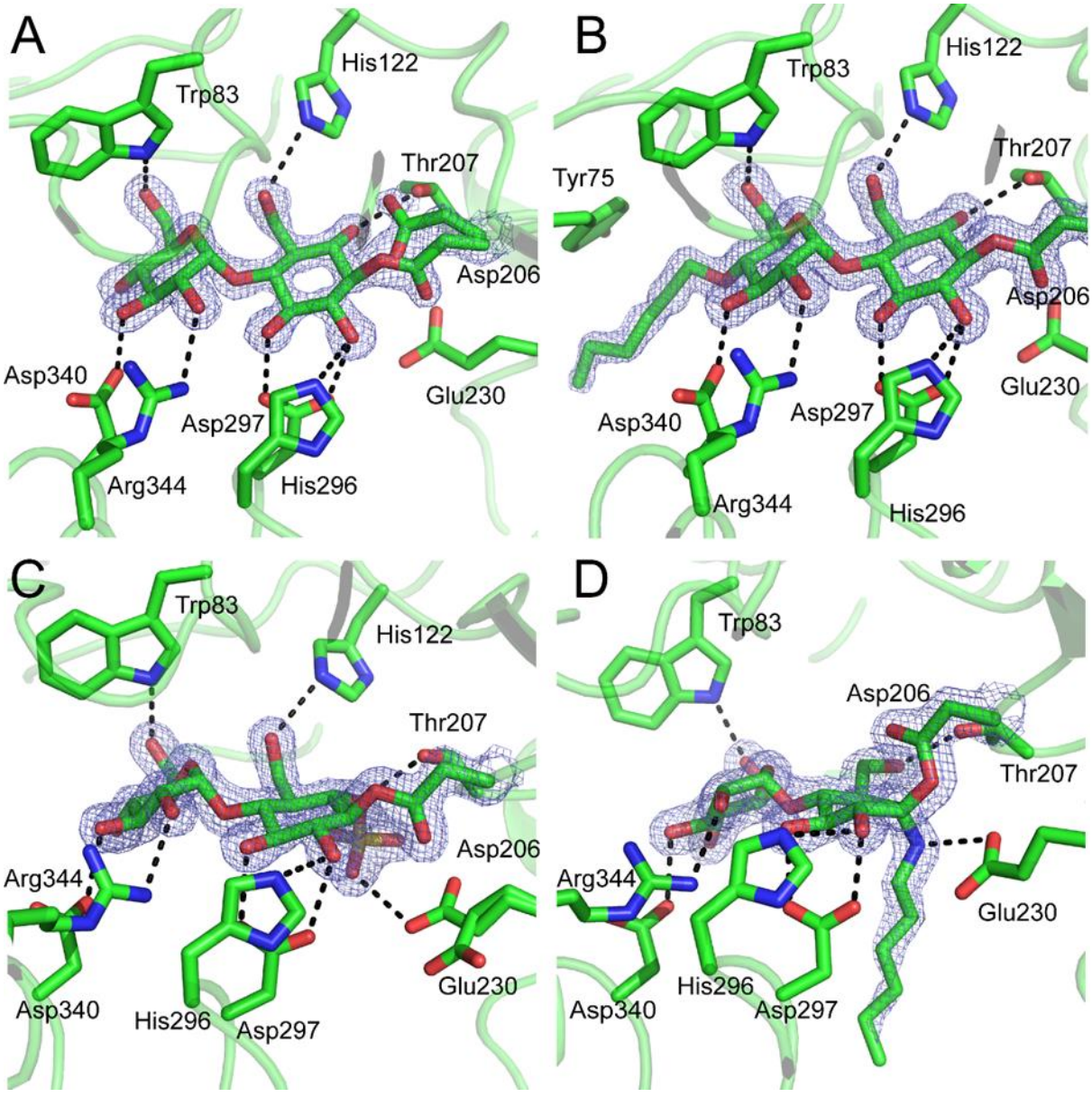


Figure 2.4. Covalent inhibition of Taka-amylase with glycosylated cyclophellitols. Crystal structures are shown between Taka-amylase and each of **1a** (A), **1b** (B), **3a** (C), and **2a** (D). Electron density ($2F_o–F_c$) is shown for both the ligand and the catalytic nucleophile (Asp206) as a blue mesh contoured at 1.5σ (**1a** = 0.59 e^-/ Å^3, **1b** = 0.61 e^-/ Å^3, **2a** = 0.46 e^-/ Å^3, **3a** = 0.54 e^-/ Å^3). The polypeptide is shown

in cartoon form with active site residues shown as sticks. Apparent hydrogen bonding interactions are shown as dotted black lines. Active-site residue Arg204 is omitted for clarity.

The crystal structure of **1a** bound to the active site nucleophile (Asp206) showed the cyclitol moiety to adopt a $^4C_1$ chair conformation, which is also observed for **1b** and **3a**. This conformation matches that observed for the -1 sugar by Caner *et al.*[24] for the binding of enzymatically synthesized **1a** to HPA and the conformation of an *in-situ* generated 5-fluoro-idose trisaccharide, also to HPA.[23] Although a detailed computational analysis of the conformational itineraries for a GH13 α-amylase has yet to be done, a detailed analysis of a homologous GH13 amylosucrase from *Neisseria polysaccharea* has been published.[40] This analysis proposes a $^4C_1 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^4C_1$ trajectory for the glycosylation step, consistent with the observed conformation of the inhibitor bound in the -1 position. The nucleophile, Asp206, in the **1a** complex structure is present in two conformations, one minor unbound conformation which aligns with the uncomplexed form (PDB: 6YQ7), and the other rotated approximately 120 degrees from the uncomplexed orientation that we ascribe to the conformation present in the covalently inhibited enzyme (Figure 2.4, panel A, PDB: 6YQ9). This second conformation is also seen for the **1b** complex which is at full occupancy. Bound **1a** also forms 9 direct hydrogen bonds with Taka-amylase – with His122, Arg204, Thr207, His296 and Asp297 in the -1 subsite and Trp83, Asp340 and Arg344 in the -2 subsite – replicating those seen in the structure of a complex between Taka-amylase and acarbose (PDB: 7TAA), with the exception of the hydrogen bond between the 6-OH of **1a** and Thr207, and the lack of an H-bond with the catalytic nucleophile that is observed for acarbose.[41]

The structure of 4'-alkylamine *epi*-cyclophellitol **1b** bound in the active site (Figure 2.4, panel B, PDB: 6YQC) maintains an identical hydrogen bonding network to that seen in the complex with **1a**, the only difference being the complete occupancy of the inhibitor, resulting in the absence of an unbound Asp206 conformer, and the presence of an alkyl tail. The 4' tail of **1b** can only be modelled to the sixth carbon (away from the maltose core) indicating that the terminal amine may form multiple interactions in the -3 subsite. The portion of the alkyl tail that could be modelled with confidence is appropriately positioned to form a C-H/π hydrogen bond through the second carbon which is 3.9 Å away from Tyr75. This interaction may in part be responsible for better binding of **1b** to Taka-amylase and the resulting increased rate of inhibition.

The structure of the *epi*-cyclophellitol cyclosulfate **3a** complex has many similarities with complexes of the epoxide inhibitors **1a** and **1b** (Figure 2.4, panel C, PDB: 6YQB). The cyclitol

is covalently attached to Asp206, again rotated approximately 120° from the unreacted enzyme, and is in a $^4C_1$ conformation. The -2 sugar is positioned exactly as for **1a**, however the presence of an exocyclic sulfur at the 6 position has resulted in Thr207 being positioned within hydrogen bonding distance to the O6 that forms the sulfate ester. The sulfate group is also positioned within hydrogen bonding distance to the acid base residue, Glu230, however this residue appears in two different conformations in the structure.

Binding of **2a** to Taka-amylase is strikingly dissimilar to the other three inhibitor complexes (Figure 2.4, panel D, PDB: 6YQA). The inhibitor is again covalently bound to Asp206, however the nucleophile is rotated nearly 180° when compared to the other three complexes, such that the non-bonded δO is positioned above the "β" face of the bound cyclitol ring rather than below it. The hydrogen bonds between the inhibitor and the enzyme -1 subsite are also altered, with the cyclitol methoxy alcohol rotated to form a hydrogen bond with Thr207 – instead of His122 – and the C6 nitrogen forms a hydrogen bond with the acid/base residue (Glu230). Furthermore, the reacted cyclitol, rather than $^4C_1$, is in an unprecedented E_3 conformation, the emergence of which we attribute to the presence of the alkyl-chain. Previous observations underscore the notion that cyclophellitol aziridines with *N*-alkylation are generally well accepted in both *exo*-acting enzymes as these often have open + subsites[28-32] and *endo*-acting enzymes that target *β*-linkages,[33] which typically contain long, tunnel or groove-like active sites. The active site of Taka-amylase, conversely, is a V-shaped surface depression with a severe angle between the -1 and +1 sugar binding sites, this in turn constrains the conformation of the aziridine alkylation and the conformation of the cyclitol ring.

2.2.4 *In vitro* labeling of recombinant human saliva α-amylase with ABPs 1c, 2b and 2c

With the fluorescent ABPs **1c**, **2b** and **2c** in hand, their labeling efficiency and specificity towards recombinant human saliva α-amylase (Type XIII-A, HSA) was investigated next. Kinetic analysis for inactivation of Taka-amylase revealed the parent inhibitors **1a**, **1b** and **2a** as slow enzyme binders. Thus, recombinant α-amylase was incubated with ABPs **1c**, **2b** or **2c** at two quite high concentrations (100 μM or 10 μM, 1 h, 37 °C, pH 7). As shown in Figure 2.5A, HSA was effectively labeled by epoxide-based ABP **1c**, while the labeling potency of aziridine-based ABPs **2b** and **2c** was dramatically reduced, and only very weak bands were observed even at 100 μM probe concentrations. In accordance with the observations from the kinetic studies, epoxide inhibitor **1b** is much more potent than aziridine inhibitor **2a**. This observation is supported by the X-ray structure analysis of inhibitor-enzyme complexes, which indicates that interactions between the -3 subsite in α-amylase active sites and the O4'

substituent may contribute to the initial binding of **1b** and the resulting increased inhibition potency while the alkylation of aziridine warhead in **2a** is not preferred during enzyme binding.

2.2.5 ABPP of retaining α-amylases in complex biological samples

As the next research objective, the visualization of α-amylases was attempted in concentrated human saliva (untreated human saliva and saliva supernatant were also tested in an initial screening, and the results were shown in Figure 2.S2) and lysates from mouse pancreas and salivary gland with ABP **1c**. As shown in Figure 2.5B, ABP **1c** labeled a distinct band at ~57 kDa, corresponding to the molecular weight of HSA.[42] The same band was also observed in mouse pancreas lysates where significant labeling appeared at 25 μM (2 h, 37 °C, pH 7), while at higher probe concentrations background fluorescence was observed. Interestingly, labeling in mouse salivary gland lysates with **1c** showed two clear bands, not only the one at ~57 kDa, but also a slightly stronger band at ~70 kDa. Next, the pH dependence of labeling was determined (Figure 2.5C). ABP **1c** exhibits a similar pH-dependent activity towards α-amylase (at the time unknown ~57 kDa band) in these three biological samples, and the optimal pH for labeling is determined at 5.0, different from the reported neutral pH optimum in HPA[21] and HSA[42] activity. Of note, in mouse salivary gland lysates, the band at ~70 kDa only appeared following labeling at pH ranging from 6.0-8.0 with optimal labeling observed at pH 7.5. Additionally, the efficiency and the rate of labeling of α-amylase (at the time unknown ~57 kDa band) were further investigated under optimal pH 5.0 (Figure 2.5D and 2.5E). In these three samples, α-amylase (the ~57 kDa band) was effectively labeled by **1c** in a concentration- and time-dependent manner, consistent with the irreversible inhibition mechanism one would expect for the maltose cyclophellitols.

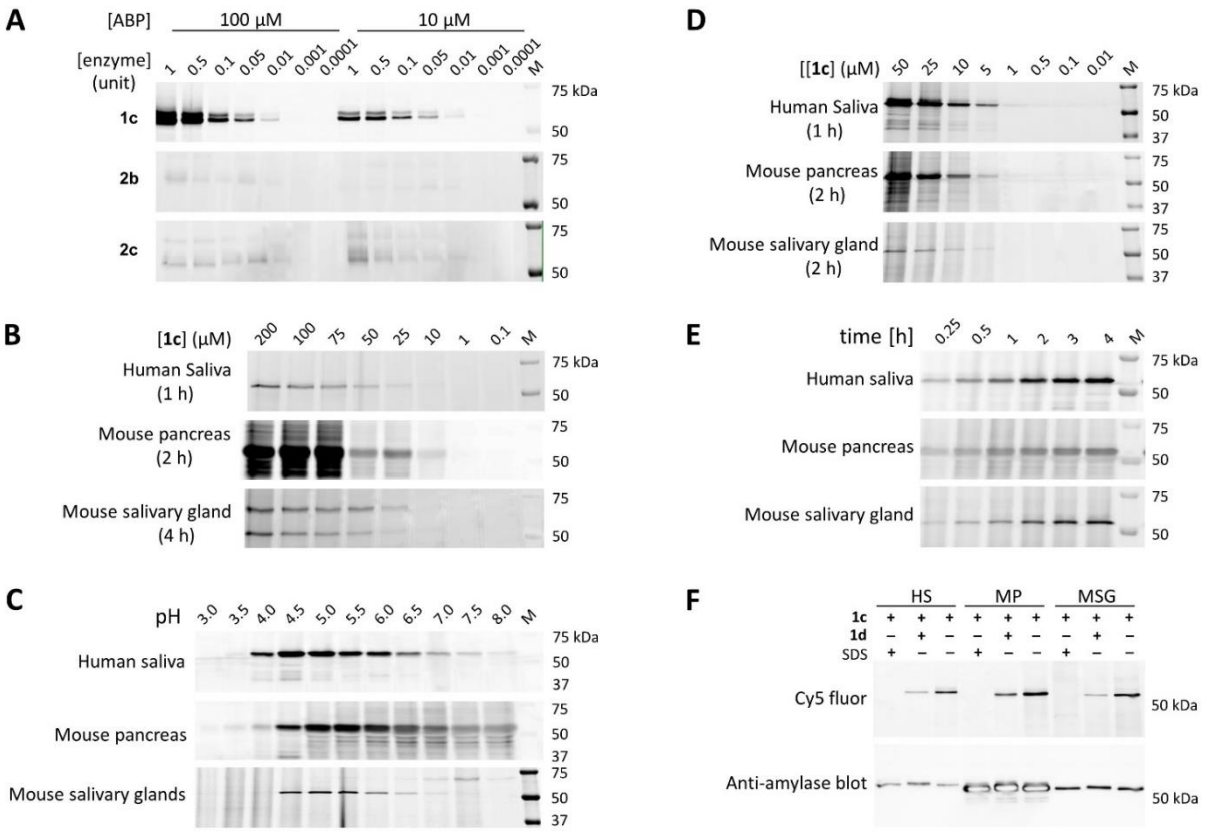


Figure 2.5. A) Fluorescent labeling of recombinant human saliva α-amylase (Type XIII-A) with ABPs **1c**, **2b** and **2c** at pH 7 after 1 h incubation at 37 °C; B) Fluorescent labeling of concentrated human saliva, mouse pancreas and salivary gland lysates (55 µg total protein) with different concentrations of ABP **1c** (pH 7.0, 37 °C); C) pH-dependent labeling with ABP **1c** in complex biological samples. The optimal pH is approximately 5.0 in these three samples; D) Detection limit of α-amylase in complex biological samples labeled with ABP **1c** (pH 5.0, 37 °C); E) Time-dependent labeling of α-amylase in complex biological samples with ABP **1c** (pH 5.0, 37 °C); F) Competitive ABPP on amylases in extracts from human saliva (HS), mouse pancreas (MP) and mouse salivary gland (MSG), offset against Western blot detection of amylase abundance using an anti-amylase antibody.

Having established that amylase activities could be visualized in extracts from human saliva, murine pancreas and murine salivary glands, competitive ABPP (cABPP) experiments were performed to see whether these activities could be negated by inclusion of a non-tagged inhibitor. Pre-incubation of biological samples with inhibitor **1d** followed by labeling with Cy5 ABP **1c** resulted in decrease of fluorescent signals similar to denaturation of the samples with 2% SDS (Figure 2.5F). After fluorescent imaging, proteins were transferred to a polyvinylidene fluoride (PVDF) membrane for Western blot analysis with an anti-amylase antibody. The observed fluorescent bands and chemiluminescence bands match well where no inhibitor is included, with the chemiluminescence bands remaining equal in intensity where the ABP signal

decreases in an inhibitor-dependent fashion, demonstrating that the labeled bands by ABP **1c** can be α-amylase. cABPP assays were also performed with inhibitors **1a** and **1b** in concentrated human saliva (Figure 2.S4). Compared with **1d**, the inhibition potency of these two was reduced and only minor competition was observed with "untagged" **1a** at the highest concentration applied (100 μM), further indicating that alkylation of the O4' position of the epoxide-based inhibitor may improve its inhibition potency.

2.2.6 ABPP of fungal secretomes

The probes developed herein can also be applied to monitor the production of amylases by biotechnologically relevant fungi. Secretomes from the saprophytic ascomycete *Aspergillus nidulans* grown on various starches have previously been shown to contain a complex mixture of amylases, glucosidases and oxidative enzymes.[43] ABPP was performed on secretomes from a time course of *A. nidulans* strain FGSC A4 grown in minimal media supplemented with 1% wheat starch as a carbon source. Secretome samples were taken daily over a 5-day period and the fluorescent Cy5 ABP **1c** was incubated with these secretome samples to reveal a complex pattern of fluorescent bands that could first be detected starting on day 3. The intensity of amylase bands increased after day 3 with a maximum intensity at day 5 (Figure 2.6C).

Conventional industrial starch hydrolysis relies on enzymes that are both thermostable and acid tolerant.[44,45] Both of these metrics can be measured directly from the secretome using ABPP, capturing these parameters in the native enzyme environment. Labeling of the day 5 secretome in the presence of McIlvaine buffers at pH between 3.0 and 8.0 revealed pH dependent labeling patterns (Figure 2.6A). The most intense band – observed at ~66 kDa – was labeled between pH 4.0–7.0 while the bands appearing at ~52 and ~48 kDa were constrained to labeling between 4.0-5.5 and 5.0-7.0 respectively. The temperature-dependent labeling of the day 5 *A. nidulans* secretome at pH 5.0 and at temperatures between 25 and 55 °C was assessed next (Figure 2.6B). Within this range the bands at ~66, ~52 and ~48 kDa all showed temperature dependent labeling, with labeling efficiency decreased above 37 °C, presumably due to denaturation of the corresponding proteins at this temperature. Conversely the bands at ~58 kDa and ~44 kDa remained present at the highest temperature tested.

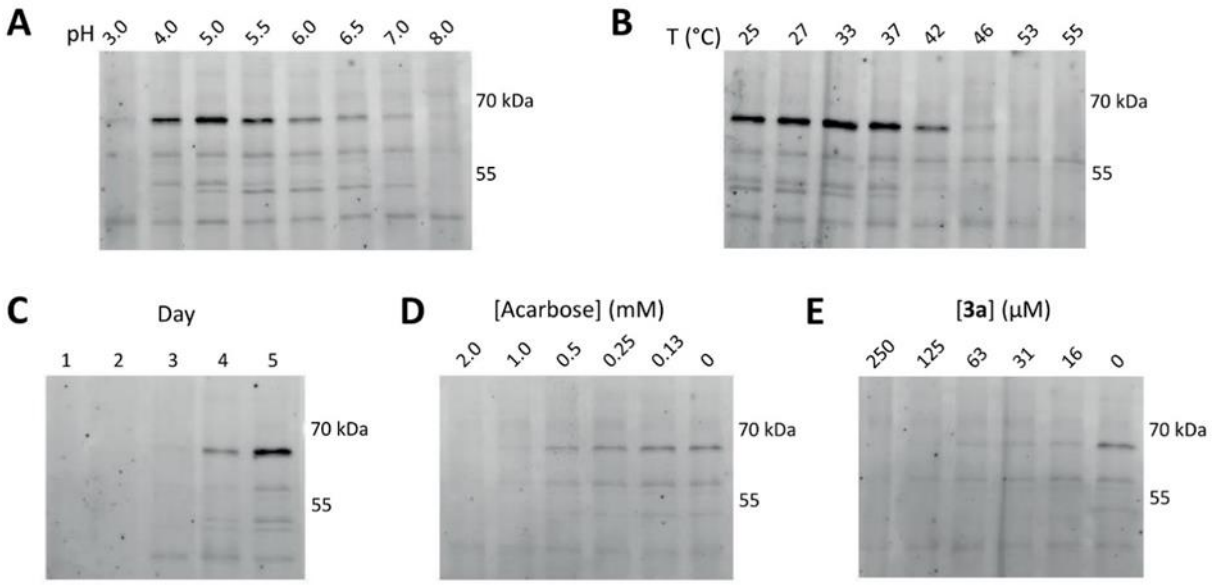


Figure 2.6. A) Fluorescent labeling of day 5 *A. nidulans* secretome with ABP **1c** at pH 3-8 after incubation for 1 h at 37 °C; B) Temperature dependent labeling of day 5 *A. nidulans* secretome with ABP **1c** after incubation at pH 5.0 for 1 h at 25-55 °C; C) Fluorescent labeling of day 1-5 *A. nidulans* secretomes with ABP **1c** at pH 5.0 after incubation for 1 h at 37 °C; D) Competitive ABPP of day 5 *A. nidulans* secretome with acarbose. The secretome was pre-incubated with acarbose at varying concentrations (30 min, pH 5.0, 37 °C) prior to labeling with ABP **1c** (67.5 μM, 1 h, pH 5.0, 37 °C); E) Competitive ABPP of day 5 *A. nidulans* secretome with **3a**. The secretome was pre-incubated with **3a** at varying concentrations (30 min, pH 5.0, 37 °C) prior to labeling with ABP **1c** (67.5 μM, 1 h, pH 5.0, 37 °C).

cABPP was applied next to show the sensitivity of enzyme labeling to both competitive and covalent inhibitors (Figure 2.6D and 2.6E). As mentioned above, cABPP allows for the identification of active enzyme inhibitors, without the need for purification and even if the identity of the active enzyme is unknown.[46,47] Pre-incubation of secretomes for 30 minutes with either acarbose, or **3a** prior to labeling with ABP **1c** resulted in decreased labeling of specific bands in a concentration dependent manner. Enzyme bands at ~66, ~58, ~52 and ~48 kDa were all inhibited by either acarbose or **3a**, while the band at 44 kDa appeared to be insensitive to these α-amylase inhibitors.

2.2.7 Identification of ABP labeled bands

To further determine the specificity of the developed ABPs, ABP-labeled proteins in human saliva samples were analyzed. For this purpose, a human saliva sample was incubated with biotinylated ABP **1d**, a no probe control (DMSO only), or a competitor-inhibited control (first treatment with **1b**). ABP **1d**-labeled proteins were enriched by using magnetic streptavidin

beads. After on-bead tryptic digestion, the resulting peptides were analyzed by LC-MS/MS and matched against the human UniProt database, using the Mascot search engine as previously described.[48] As expected, AMY1A (P04745) was clearly abundant in the ABP **1d** pulldown experiment (Table 2.S2). Other proteins with high scores were background proteins such as enolase standard and keratin contaminations. Pretreatment of the saliva sample with inhibitor **1b** caused a significant loss of AMY1A peptide abundance, and almost no AMY1A was found in the untreated control. These results demonstrate AMY1A in human saliva samples can be selectively labeled and identified via biotinylated ABP **1d**.

Based on the known content of *A. nidulans* secretome,[43] the most prominent ~66 kDa band labeled with **1c** was tentatively assigned as the mature form of α-amylase AmyB (ANIA_03402), see Table 2.S4 for list of mature α-amylase molecular masses. The two minor bands at ~52 and ~48 kDa were also tentatively assigned as AmyA (ANIA_03388) and AmyF (ANIA_0201) α-amylases, respectively. As the wheat starch grown secretome of *A. nidulans* only contains three amylases,[43] the final two minor bands could not be confidently assigned. It can however be speculated that these are proteins that are highly abundant in the secretome.

To identify the labeled enzymes in the *A. nidulans* secretome an activity-based protein pulldown was performed as well. The day 5 secretome was treated with either the biotinylated ABP **1d** or an equivalent volume of DMSO, then a pulldown was performed using streptavidin beads, followed by an on bead tryptic digest and LC-MS/MS analysis. All three amylases from the *A. nidulans* secretome (AmyA, AmyF and AmyB) were detected at elevated levels in the probe **1d**-treated samples relative to the negative control and total secretome, confirming their initial assignment of the major and two minor bands observed by in-gel fluorescence (Table 2.S3). In addition, ChiB, an endochitanase and AgdB, an α-glucosidase were also detected, both of which are highly abundant in the secretome (Table 2.S3). The molecular weights of the detected amylases match well with the labeled bands seen from ABPP with probe **1c**. Furthermore, these results suggest the identities of the minor bands seen at 44 and 58 kDa may belong to ChiB and AgdB respectively – the latter, which is proteolytically processed from its 108 kDa form to a heterodimer with the catalytic protomer having a 58 kDa molecular weight.[49]

2.3 Conclusion

This Chapter describes the synthesis of a panel of *maltobiose*-configured *epi*-cyclophellitol derivatives as mechanism-based covalent inhibitors and ABPs for retaining α-amylases. Kinetic studies of the inhibition of Taka-amylase revealed that cyclosulfate warhead (**3a**) is the most

potent, compared with its epoxide and aziridine counterparts (**1a** and **2a**). As well, alkylation of O4' of the nonreducing end sugar in **1a** led to compound **1b** that inhibits α-amylase up to 10-fold more potently than the parent compound. Compound conformations can strongly influence the inhibitory potencies and selectivities of cyclophellitol analogues. X-ray studies have shown that both epoxide and cyclosulfate-based inhibitors bind to the active site nucleophilic residue of Taka-amylase in a $^4C_1$ chair conformation, whilst the structure of aziridine **2a** complex was found in an unprecedented E_3 conformation, likely due to the presence of the aziridine alkyl-chain, which might account for its lower potency. The inability of fluorescent ABPs **2b** and **2c** to modify recombinant human saliva α-amylase further indicated that substituted aziridines are not accepted by the tested amylases. In contrast, epoxide-based ABP **1c** effectively labels α-amylases in human saliva and mice tissue in a concentration- and time-dependent manner, with optimum labeling at pH 5.0. In addition, it can also detect α-amylases in fungal secretomes and allow easy read out of pH and temperature dependence and susceptibility to inhibitors. Looking ahead, we speculate that installation of reporter entities at O4' of the nonreducing end sugar of cyclosulfate-based inhibitor **3a** would yield more potent ABPs based on kinetic and crystallographic studies. In conclusion, the described maltobiose *epi*-cyclophellitol ABPs allow activity-based profiling of α-amylases in complex biological samples, and may find use in the discovery of new and effective human amylase inhibitors that may have potential as therapeutic reagents for the treatment of type 2 diabetes, and in the discovery of new microbial amylases with potentially advantageous properties for biotechnological application.

2.4 Acknowledgements

Zachary Armstrong and Gideon Davies from University of York, UK are kindly acknowledged for the kinetic, X-ray and fungal secretome labeling experiments, and valuable discussion. Mikkel Rasmussen and Maher Abou Hachem from Technical University of Denmark are kindly acknowledged for providing the *Aspergillus nidulans* secretomes.

2.5 Experimental methods

2.5.1 Biochemical experiments

Materials

Recombinant human saliva α-amylase (Type XIII-A) was obtained as a lyophilized power from Sigma-Aldrich (Product Number: A1031). Taka-amylase was purchased as a solid power from Sigma-Aldrich (Product Number: 86247). Anti-Pancreatic alpha amylase antibody was purchased as a liquid from Abcam (Product Number: ab199132). Concanavalin A beads was purchased from Sigma-Aldrich

(Product Number: C9017). Mouse tissue were isolated according to guidelines approved by the ethical committee of Leiden University (DEC#13191). All the tissue lysates were prepared in potassium phosphate lysis buffer (25 mM in pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and 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 Bicinchoninic acid (BCA) Protein Assay Kit (PierceTM).[50] The protein fractions were stored in small aliquots at -80 °C until use. BCA protein assay kit was acquired from Thermo Fisher Scientific. Trypsin was commercially available from Promega. Dynabeads™ MyOne™ Streptavidin T1 was obtained from Invitrogen. Dithiothreitol (DTT) was obtained from Biochemica and Iodoacetamide (IAA) was obtained from Sigma Aldrich. All other chemicals were obtained from standard commercial sources.

Preparation of human saliva sample

Unstimulated whole saliva sample (20 mL) from a normal subject was expectorated into a 50-mL polypropylene tube. The tube was centrifuged at 4 °C for 10 min at 4000 rpm. Supernatant was collected into a new, ice cold 50-mL tube and stored at -20 °C. The pellets were resuspended into 1 mL ice-cold KPi lysis buffer (25 mM phosphate buffer pH 6.5 supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor 1x cocktail (Roche)). To homogenize, the suspension was sonicated for 60 seconds at 60% amplitude strength on ice. Obtained lysates were stored in aliquots at -20 °C. Protein concentration of lysate was determined by using the BCA method.[50]

Purification of Taka-amylase

To purify Taka-amylase for crystallization, 10 g of powder was dissolved in 200 mL of buffer A (20 mM Tris pH 7.5, 5 mM $CaCl_2$). This suspension was concentrated and washed extensively with buffer A using 10 kDa cutoff Vivaspin centrifugal filter columns (Sigma-Aldrich). Concentrated protein was further purified using a gel-filtration column (HiLoad 16/600 Superdex 75 pg; GE Healthcare), which had been pre-equilibrated with buffer B (20 mM Tris pH 7.5, 5 mM $CaCl_2$, 200 mM NaCl). Peak fractions were pooled, concentrated and washed with buffer A. Protein stocks were diluted to 30 mg/ml and flash frozen in liquid nitrogen.

Inhibition Kinetics

To assess the time dependent inactivation of the synthesized inhibitors samples, Taka-amylase (60 µM) was incubated with 2 mM of either **1a**, **1b**, **2a**, **3a** or buffer alone in assay buffer (20 mM MES, 20 mM NaCl, 5 mM $CaCl_2$ pH 5.0) at 30 °C. Aliquots of these inactivation mixtures were removed at time intervals and diluted 100-fold into assay cells containing assay buffer and 2-chloro-4-nitrophenyl α-maltotrioside (5 mM) pre-incubated at 30 °C, and absorbance change was monitored at 400 nm. Measured residual rates as a function of time for each inactivator were fit to an equation describing first order decay with offset in OriginPro 2019 (OriginLab).

Crystallization of Taka-amylase

Taka-amylase at 30 mg/mL was tested against a range of commercial crystallization screens. Large split crystals were found in 0.1 M sodium citrate pH 5.6, 20% (v/v) 2-Propanol, 20% (w/v) PEG 4,000, a condition that was used for further optimization. The optimized crystals were grown in MRC maxi 48-well plates (Swissci) using the sitting-drop vapor-diffusion method at 20 °C using a well-solution composed of: 0.1 M sodium citrate pH 5.6, 18% (v/v) 2-propanol. Droplets contained a mixture of 500 nL protein stock and 500 nL of well-solution. Crystals appeared within one week. Crystals were soaked in 0.1 M sodium citrate pH 5.6, 25 % (v/v) ethylene glycol, 20 % (w/v) PEG 4,000 prior to flash cooling in liquid nitrogen. Inhibitor complexes for **1a**, **1b**, **2a** and **3a** were obtained by soaking crystals in well solution containing 10 mM of the inhibitor for several days (3 days for **1a** and **1b** and 1 month for **2a** and **3a**) before flash cooling in liquid nitrogen.

All diffraction data were collected at Diamond beamlines (un-liganded and **1a** complex: I03, complexes with **1b**, **2a** and **3a**: I04), and data were processed with the CCP4i2 suite.[51] All structures of Taka-amylase were solved by molecular replacement using the PDB structure 7TAA as the search model in Phaser.[52,53] Refinement was performed using cycles of maximum-likelihood refinement using REFMAC5[54] were interspersed with manual corrections of the models using COOT.[55] Structural figures were drawn with PyMol (DeLano Scientific LLC, http://pymol.sourceforge.net/).

Labeling and SDS-PAGE of recombinant human saliva α-amylase

To prepare for labeling, recombinant human saliva α-amylase stock (1 unit/µL) was diluted with assay buffer (150 mM McIlvaine buffer supplemented with 10 mM $CaCl_2$ and 10 mM NaCl, pH 7.0, if not otherwise specified) to final concentrations of 1, 0.5, 0.1, 0.05, 0.01, 0.001, 0.0001 unit in 10 µL volume. ABPs **1c**, **2b** or **2c** stock (10 mM in DMSO) was diluted with assay buffer (2% (v/v) DMSO) to 3.5x of intended assay concentrations. For labeling, 10 µL enzyme dilution was incubated with 4 µL ABP (**1c**, **2b** or **2c**) dilution at 37 °C for 1 h (Figure 2.5A). To determine the detection limit, 0.2 unit pure enzyme in 10 µL assay buffer (pH 5.0) was incubated with 4 µL ABP **1c** at varying concentrations at pH 5.0 for 4 h, at 37 °C (Figure 2.S5A). The time-dependent labeling was assessed by incubating 0.2 unit pure enzyme in 10 µL assay buffer (pH 5.0) with 4 µL of 175 µM ABP **1c** dissolved in assay buffer pH 5.0 at 37 °C for 10, 30, 60, 120, 180, 240 or 360 min (Figure 2.S5B). Samples were then denatured with 4 µL sample buffer (5x Laemmli buffer, containing 50% (v/v) 1M Tris-HCl pH 6.8, 50% (v/v) glycerol, 10% (w/v) Dithiothreitol (DTT), 10% (w/v) sodium dodecyl sulphate (SDS), 0.01% bromophenol blue) and heated at 98 °C for 5 minutes. Proteins were resolved by electrophoresis in sodium dodecylsulfate (SDS-PAGE) 10% polyacrylamide gels, running at a constant of 90V for 30 minutes followed by 140V for approximately 60 minutes. Wet slab gels were scanned on fluorescence using a Typhoon FLA9500 Imager (GE Healthcare) using λ_{EX} 635 nm; λ_{EM} > 665 nm and images were

processed using ImageLab 5.2.1 (BioRad). Gels were subsequently extensively stained with Coomassie Brilliant Blue solution for assessing total protein amounts in each lane of sample.

Labeling of untreated human saliva, saliva supernatant and concentrated human saliva

To prepare for labeling, concentrated human saliva pellet (16.9 μg/μL in KPi buffer) was diluted with assay buffer pH 7.0 (2% (v/v) DMSO) to final concentration of 55 μg in 24 μL volume. Untreated saliva and saliva supernatant were thaw on ice and 24 μL of each was directly used for labeling. ABPs **1c** (10 mM in DMSO) was diluted with assay buffer pH 7.0 (2% (v/v) DMSO) to 5x of intended assay concentrations. For labeling, 6 μL of ABP **1c** dilution was incubated with 24 μL enzyme dilution at 37 °C for 1 h. Samples were then denatured with 8 μL sample buffer (5x Laemmli buffer, containing 50% (v/v) 1M Tris-HCl pH 6.8, 50% (v/v) glycerol, 10% (w/v) Dithiothreitol (DTT), 10% (w/v) sodium dodecyl sulphate (SDS), 0.01% bromophenol blue) and heated at 98 °C for 5 minutes, and subjected to SDS-PAGE and fluorescence scan as described above. Results were shown in Figure 2.S2.

Labeling of lysates and SDS-PAGE analysis

Concentrated human saliva pellet (55 μg total protein per sample), mouse pancreas lysates (55 μg total protein per sample), or mouse salivary gland lysates (55 μg total protein per sample) were diluted in assay buffer (pH 7.0, if not otherwise specified) and incubated with ABP **1c** at intended concentrations (2% (v/v) DMSO) at 37 °C for the intended time periods (Figure 2.5B). Influence of pH on ABP labeling was analyzed using 55 μg total protein incubated with 25 μM **1c** (for concentrated human saliva and mouse pancreas) or 50 μM **1c** (for mouse salivary gland), dissolved in assay buffer (2% (v/v) DMSO), pH 3.0−8.0, for 4 h at 37 °C (Figure 2.5C). To determine the detection limit under optimal pH, 55 μg protein was labeled with 0.01−50 μM ABP **1c** in assay buffer pH 5.0 (2% (v/v) DMSO) at 37 °C for the intended time periods (Figure 2.5D). The time-dependent labeling under optimal pH was assessed by incubating 55 μg protein with 10 μM **1c** (for concentrated human saliva and mouse pancreas) or 50 μM **1c** (for mouse salivary gland), dissolved in assay buffer pH 5.0 (2% (v/v) DMSO) at 37 °C for 0.25, 0.5, 1, 2, 3 or 4 h (Figure 2.5E). Samples were then denatured with sample buffer (5x Laemmli buffer, containing 50% (v/v) 1M Tris-HCl pH 6.8, 50% (v/v) glycerol, 10% (w/v) Dithiothreitol (DTT), 10% (w/v) sodium dodecyl sulphate (SDS), 0.01% bromophenol blue) and heated at 98 °C for 5 minutes and subjected to SDS-PAGE and fluorescence scan as described above.

DMSO control experiment

To prepare for labeling, assay buffer pH 5.0 was supplemented with varying concentrations of DMSO (0%, 0.5%, 1% and 2% (v/v)). Concentrated human saliva (16.9 μg/μL in KPi buffer) was diluted with assay buffer containing 0–2% (v/v) DMSO to a final concentration of 55 μg protein in 24 μL volume. ABPs **1c** (10 mM stock in DMSO) was diluted with assay buffer containing 0–2% (v/v) DMSO to 5x of intended assay concentrations. For labeling, 6 μL of ABP **1c** dilution was incubated with 24 μL

enzyme dilution at 37 °C for 1 h. Samples were then denatured with 8 µL Laemmli (5x) sample buffer and heated at 98 °C for 5 minutes, and subjected to SDS-PAGE and fluorescence scan as described above. Results were shown in Figure 2.S3.

Labeling of *A. nidulans* secretome and SDS-PAGE analysis

A. nidulans strain FGSC A4 was grown in minimal medium containing 1 % (w/v) wheat starch (Sigma-Aldrich, S5127) as described previously.[56] Secretome samples were removed daily over a period of five days, flash frozen in liquid nitrogen and stored at -20 ˚C until use in labeling experiments. Secretome samples (12 µL) were diluted with 150 mM McIlvaine buffer of appropriate pH to a final volume of 15 µL, before addition of 3 µL ABP **1c** to a final concentration of 100 µM. Labeling reactions were carried out for 1 h at 37 °C with shaking at 400 rpm, then stopped by addition of 6 µL Laemmli (4X) sample buffer and boiling at 95°C for 5 minutes. 10 µL of each reaction was separated on a 10% SDS-PAGE gel prior to imaging using the Cy5 laser/filter settings on a Typhoon 5 scanner (GE Healthcare). PageRuler Plus Prestained protein ladder (Thermo Fisher Scientific) was used as marker.

Competitive ABPP experiments

Concentrated human saliva (5 µg total protein), mouse pancreas (5 µg total protein) and mouse salivary gland (25 µg total protein) were diluted with assay buffer (pH 5.0, 2% (v/v) DMSO) to a final 10 µL volume, pre-incubated with 2.5 µL **1d** (final 10 µM for human saliva sample; final 50 µM for mouse pancreas and salivary gland samples) at 37 °C for 4 h. Then, 2.5 µL of ABP **1c** was added to a final concentration of 10 µM (for human saliva and mouse pancreas) or 50 µM (for mouse salivary gland). The labeling reaction was incubated at 37 °C for 2 h. Additionally a positive control was performed: the protein was pre-incubated with 2.5 µL assay buffer (pH 5.0, 2% (v/v) DMSO) for 4 h at 37 °C, and subsequently labelled with ABP **1c** at the concentrations described above. A negative control was also performed: 2.5 µL SDS (10 % (w/v)) was added to the 10 µL protein dilution, boiling at 98 °C for 5 min, and incubated with ABP **1c** at 37 °C for 2 h. Samples were then denatured with 4 µL Laemmli (5x) sample buffer and heated at 98 °C for 5 minutes. Proteins were resolved by electrophoresis in 10% SDS-PAGE gels, running at a constant of 90V for 30 minutes followed by 140V for approximately 60 minutes. Wet slab gels were scanned on fluorescence using a Typhoon FLA9500 Imager (GE Healthcare) using λ_{EX} 635 nm; λ_{EM} > 665 nm and images were processed using ImageLab 5.2.1 (BioRad).

Pre-incubation of secretomes with inhibitors was carried out in 18 µL reactions which contained, 12 µL of secretome, 3 µL McIlvaine buffer pH 5.0 and 3 µL of inhibitor at varying concentrations. After pre-incubation for 30 min at 37°C, ABP **1c** was added to a final concentration of 67.5 µM. This labeling reaction was incubated at 37°C for 1 h with shaking at 400 rpm, then stopped by addition of 5 µL Laemmli (4X) sample buffer and boiling at 95°C for 5 minutes. 10 µL of each reaction was separated on a 10% SDS-PAGE gel prior to imaging using the Cy5 laser/filter settings on a Typhoon 5 scanner

(GE Healthcare). PageRuler Plus Prestained protein ladder (Thermo Fisher Scientific) was used as marker.

Western blotting

Proteins resolved by SDS–PAGE were transferred to a PVDF membrane using a Trans-Blot Turbo system (Bio-Rad). Membrane was blocked in 10 mL of 5% (w/v) powder milk in TBST (50 mM Tris (pH 7.5), 150 mM NaCl, 0.1% Tween 20) for 1 h at room temperature, then incubated with rabbit polyclonal anti-Pancreatic alpha amylase (Abcam; ab199132) at 1:2,000 dilution in TBST (5% (w/v) powder milk) at 4 °C overnight. Membrane was washed 3× with TBST and blocked in 5% (w/v) powder milk in TBST for 30 min at room temperature, then incubated with HRP-conjugated mouse anti-rabbit IgG at 1:5,000 dilution in 5% (w/v) powder milk in TBST for 1 h at room temperature. Membrane was washed again 3× with TBST, 3× with TBS and 2× with Demi-H_2O and blots visualized using Amersham prime ECL reagent (GE Healthcare) and recorded using a Bio-Rad ChemiDoc system.

Detection sensitivity of ABP 1c towards α-amylase in human plasma

To prepare for labeling, recombinant human saliva α-amylase stock (1 unit/μL) was diluted with assay buffer pH 5.0 to final concentrations of 0, 0.0005, 0.005, 0.025, 0.05, 0.5 unit/μL. ABPs **1c** stock (10 mM in DMSO) was diluted with assay buffer pH 5.0 to two different concentrations: 200 μM and 50 μM.

Pre-wash of Concanavalin A (ConA) beads: 940 μL of ConA bead slurry was put in a 15 mL centrifuge tube and washed with 4700 μL of ConA buffer (0.1 M NaOAc, 0.1 M NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM $MnCl_2$, pH 6.0. Keep on ice all the time when using). Centrifuging at 4 °C for 1 min and the supernatant (4700 μL) was carefully pipette out. After washing 4x more, 564 μL of ConA buffer was added to the bead slurry and 100 μL of the bead slurry was pipette into a series of 1.5 mL Eppendorf tubes.

For labeling:

Set **A**: 2 μL of enzyme dilution + 22 μL of assay buffer + 8 μL of 200 μM ABP **1c** dilution was incubated at 37 °C for 4 h. Samples were then denatured with 8 μL Laemmli (5x) sample buffer and heated at 98 °C for 5 minutes, and subjected to SDS-PAGE as described above.

Set **B**: 2 μL of enzyme dilution + 22 μL of assay buffer + 8 μL of 200 μM ABP **1c** dilution was incubated at 37 °C for 4 h. During this period, 25 μL of human plasma was added to each of the Eppendorf tubes that contain 100 μL pre-washed ConA bead slurry, binding at 4 °C on a roller for 1 h. After binding, the beads were washed with ConA buffer (5 x 1000 μL) and all the supernatant was carefully pipette out upon the fifth wash. Samples were then denatured with 8 μL Laemmli (5x) sample buffer, vortexed slightly and the final 40 μL mixture was immediately added to the ConA purified

plasma and heated at 98 °C for 5 minutes. After short spin down, the supernatants were collected in new Eppendorf tubes and subjected to SDS-PAGE as described above.

Set **C**: 2 µL of enzyme dilution was spiked into 25 µL of human plasma, which was added to each of the Eppendorf tubes that contain 100 µL pre-washed ConA bead slurry, incubating at 4 °C on a roller for 1 h. After washing, 32 µL of 50 µM ABP **1c** dilution was immediately added to the ConA purified plasma and the mixture was incubated at 37 °C for 4 h. Samples were then denatured with 8 µL Laemmli (5x) sample buffer and heated at 98 °C for 5 minutes. After short spin down, the supernatants were collected in new Eppendorf tubes and subjected to SDS-PAGE as described above. Results were shown in Figure 2.S1.

Pull-down and LC-MS/MS analysis

Proteomics experiment was performed with human saliva in triplicate for the DMSO control, **1b**-inhibited control and **1d** pulldown. Human saliva sample (300 µg total protein) was incubated with either 0.5% (v/v) DMSO or ABP **1d** (5 µM), or firstly with 25 µM inhibitor **1b** followed by 5 µM of ABP **1d**, each step taking 4 h at 37 °C, in a total volume of 200 µL McIlvaine buffer pH 5.0 supplemented with 10 mM $CaCl_2$ and 10 mM NaCl, followed by denaturation by addition of 10% (w/v) SDS 20 µL and heating for 5 minutes at 98 °C. The solutions were centrifuged at 10,000 rpm for 10 min at room temperature and clear supernatants were transferred carefully into new 2-mL Eppendorf tubes, followed by methanol-chloroform precipitation: for each sample, 800 µL methanol was added into the 220 µL protein solution followed by 200 µL chloroform and vortexed vigorously upon each addition. Then 600 µL Milli-Q was added and the suspension was vortexed vigorously prior to centrifuging at 10,000 rpm for 10 minutes at room temperature. The aqueous top layer was removed carefully and 600 µL methanol was added, then the sample was vortexed slightly and centrifuged as above. The supernatant was then removed, and the pellet was left to air-day for 5 min.

The protein pellet was resuspended in 100 µL PBS buffer with 2% (w/v) SDS, vortexed vigorously and heated to 45 °C for 20 minutes with shaking at 600 rpm. 14 µL of 50 mM DTT was added and the samples were incubated at 65 °C for 30 minutes to reduce disulfide bonds, then 200 µL PBS buffer was added and the suspensions were sonicated for 30 seconds at 20% amplitude at 0 °C. Then, 14 µL of freshly-prepared 150 mM iodoacetamide (IAA) was added and the samples were incubated in the dark for 30 minutes at room temperature. The excessive IAA was subsequently quenched by adding 7 µL of 50 mM DTT, incubating for 30 minutes at room temperature. The 335 µL protein suspensions were stepwise diluted with PBS buffer (200 µL/1000 µL/500 µL, vortex vigorously upon each addition) to a final 2-mL volume. Then the resulting 2-mL suspensions were centrifuged at 10,000 rpm for 10 minutes at room temperature and supernatants were transferred carefully into new 2-mL Eppendorf tubes.

The samples were then ready for pull-down with prewashed Dynabeads™ MyOne™ Streptavidin T1 (30 µL) as published previously.[57] After pull-down, all samples were used for on-bead digestion.

The samples were treated with 100 µL on-bead digestion buffer (50 mM NH_4HCO_3 (pH 8), 2% ACN, 1 mM $CaCl_2$, 2.5 ng/µL of trypsin) and incubated overnight at 37 °C. The supernatants, containing tryptic-digested peptides were then transferred to new 2-mL low-binding tubes and desalted using StageTips. Consequently, the acetonitrile was evaporated using a SpeedVac at 45 °C followed by addition of 50 µL of LC-MS sample solution (97:3:0.1, H_2O:acetonitrile:formic acid) for LC-MS analysis. All peptide samples were analyzed with a two-hour gradient of 5% to 25% acetonitrile on nano-LC, hyphenated to an LTQ-Orbitrap and identified using the Mascot protein search engine.[58] Raw data was calculated using MaxQuant against UniProt of human saliva proteome database to obtain an identification list of found proteins. The abundance of the protein hits was quantified as previously described,[48] in unsupervised mode using the default settings of the PLGS (Waters) and IsoQuant software. Raw and quantified data can be found in Table 2.S2.

Fungal Secretome Proteomics

Proteomic analysis was performed in triplicate for the DMSO control, **1d** pulldown and total secretome. *A. nidulans* day 5 wheat secretome was concentrated 12.5-fold using a Vivaspin centrifugal concentrator (3 kDa cut-off). Concentrated secretome (40 µl) was buffered with 30 mM McIlvane buffer pH 5.0 and transferred to a Lo-bind 0.5 ml tube (Eppendorf). The buffered secretome was then treated with either 300 µM **1d** or DMSO control. Reactions were incubated 4 hours at 30 °C with shaking at 450 rpm. Labeling was terminated by the addition of DTT and SDS to a final concentration of 4 mM and 0.2 % (v/v) respectively, and heating to 95 °C for 5 minutes. The denatured proteins were then cooled to 25 °C, IAA was added to a final concentration of 24 mM then samples were incubated in the dark for 30 minutes at 25°C. IAA, DTT, SDS, buffer and excess probe were removed by acetone precipitation; 4 volume equivalents of acetone (-20 °C) was added to the sample and the mixture was vortexed. The sample was then placed at -20 °C for one hour, then proteins were pelleted by centrifugation at 14,000 x *g* for 1 minute. The supernatant was removed, then the pellet was washed with the same volume of -20 °C acetone then centrifuged again at 14,000 x *g* for 1 minute. The supernatant was then removed, and the pellet was left to briefly air dry.

Samples were then resuspended 50 µL of 4 M urea, then diluted with 450 µL of 0.05% SDS in phosphate buffered saline. Strep Mag Sepharose beads (GE healthcare; 20 µL; washed twice with phosphate buffered saline (PBS) were added to each sample, followed by incubation at 25 °C with constant agitation to prevent bead settling for 3 hours. Beads were collected using a magnetic rack and the supernatant was discarded. Beads were washed with 1 mL of 2 % SDS at 25 °C then with 1 mL of 2 % SDS at 65 °C to eliminate non-specific binding, followed by 1 ml of 2 M urea for 5 minutes to remove SDS and then 1 mL of water two times to remove residual SDS and urea. The washed beads were then resuspended in 20 µL of 0.05 M triethylamonium bicarbonate and 1 µL of 0.5 µg/µL trypsin (Promega V5280) in 50 mM acetic acid was added. The on-bead digest was incubated overnight at 37 °C. Following the tryptic digest, the beads were immobilized and released peptides were removed. The total

secretome samples were treated as above, except the acetone pellet was resuspended directly in 0.05 M triethylamonium bicarbonate and digested with trypsin in solution.

The samples were desalted via a C18 ZipTip then resolubilised with aqueous 0.1% TFA, before injection onto a 50 cm EasyNano C18 PepMap column. Peptides were eluted at a flow rate of 300 nL/min over 1 hour acquisitions onto an Orbitrap Fusion Tribrid mass spectrometer operated in DDA mode. Tandem mass spectra were searched using both Mascot and X!Tandem search programs against the *Emericella nidulans* strain FGSC A4 subset of the Uniprot database and typical contaminants. Carboxymethylation of Cysteine was set as a fixed modification and oxidation of Met was set as variable. Resulting spectral matches were combined in Scaffold and filtered to require a global protein and PSM false discovery rate of 1%, and a minimum of two unique peptides per protein group identification. Label-free quantification was performed using Progenesis QI (Waters). Following chromatographic alignment, peaks were integrated and assigned. Protein abundance was estimated using the integrated intensity of non-conflicting peptides. Results of this analysis for all identified proteins can be found in Table 2.S3.

2.5.2 Chemical synthesis

General experimental details

All reagents were of experimental grade and were used without further purification unless stated otherwise. Dichloromethane (DCM) and tetrahydrofuran (THF) were stored over 3 Å molecular sieves and *N*,*N*-dimethylformamide (DMF) was stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an Argon or N_2 atmosphere unless stated otherwise. Reactions were monitored by analytical thin layer chromatography (TLC) using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV-absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}\cdot H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4\cdot H_2O$ (10 g/mL) 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 manually using Screening Device b.v. silica gel 60 (0.04-0.063 mm) in the indicated solvents. LC-MS analysis was performed on a LCQ Advantage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C18 column (Gemini, 4.6 mm x 50 mm, 5 µM particle size, Phenomenex). The applied buffers were H_2O, acetonitrile (MeCN) and 1% aqueous trifluoroacetic acid (TFA). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-400 (400/101 MHz), Bruker AV-500 (500/126 MHz), and Bruker AV-850 (850/214 MHz) spectrometers in the given solvent. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (TMS) as internal standard (^{1}H NMR in $CDCl_3$) or the residual signal of the deuterated solvent. Coupling constants (*J*) are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. The following abbreviations are used to describe peak patterns when appropriate: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), Ar (aromatic), C_q (quarternary

carbon). 2D NMR experiments (COSY, HSQC) were carried out to assign protons and carbons of the new structures. High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electronspray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL/min, capillary temperature 250 °C) with resolution R = 60000 at m/z 400 (mass range m/z = 150 – 2000) and dioctyl phthalate (m/z = 391.28428) as a "lock mass". The high-resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan).

Experimental Procedures and Characterization Data of Products

Known compounds **4a**[36], **5**[33], **S1**[36], **S2**[33], Cy5-OSu[59], biotin-OSu[60] were synthesized following procedures previously described and their spectroscopic data are in agreement with those previously reported.

Standard procedure A: glycosylation with imidate donor (4a)

The donor (1.5 equiv.) was co-evaporated with toluene (3 x) and dissolved in dry DCM (0.05 M) under nitrogen and stirred over fresh flame-dried molecular sieves 3 Å, after which DMF (16 equiv. of donor) was added to the solution. The solution was cooled to -20 °C and TfOH (1.5 equiv.) was added. After 1 h, the pre-activation was complete as indicated by TLC-analysis. Acceptor (1.0 equiv.) was added to the solution and the mixture was placed in an ice bath. The reaction was stirred at 0 °C until TLC-analysis showed complete conversion of the acceptor. The reaction was quenched with Et_3N, filtered and concentrated *in vacuo*. The products were purified by size exclusion (eluent MeOH/DCM: 1/1) and silica gel column chromatography (See experimental description below for eluent system). *Note: for the glycosylation of donor **4a** with acceptor **5**, the reaction was first kept at -20 °C for 24 h then warmed to 0 °C and stirred until TLC-analysis showed complete conversion of the acceptor.*

Standard procedure B: glycosylation with thio-donors (4b and 4c)

The donor (1.5 equiv.) was co-evaporated with toluene (3 x) and dissolved in dry DCM (0.05 M) under nitrogen and stirred over fresh flame-dried molecular sieves 3 Å, after which DMF (16 equiv. of donor) was added to the solution. The solution was cooled to -20 °C, NIS (1.5 equiv.) and TfOH (1.5 equiv.) were added successively. The mixture was pre-activated at -20 °C for 2 h. Then acceptor (1.0 equiv.) was added to the solution and the mixture was placed in an ice bath. The reaction was stirred at 0 °C until TLC-analysis showed complete conversion of the acceptor. The reaction mixture was diluted with DCM and was quenched with sat. aq. $Na_2S_2O_3$. The organic phase was washed with water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The products were purified by size exclusion (eluent MeOH/DCM: 1/1) and silica gel column chromatography (See experimental description below for eluent system).

General procedure C: Deprotection of PMB protecting group

The starting material (1.0 equiv.) was dissolved in DCM/H_2O (19/1, 0.05 M). DDQ (1.2 equiv.) was added to the mixture. The reaction was stirred at rt until TLC-analysis indicated full consumption of the starting material (± 4 h). Then the mixture was diluted with DCM and washed with sat. aq. $NaHCO_3$ (2 x), water (1 x) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (See experimental description below for eluent system).

General procedure D: Epoxidation with *m*-CPBA

The starting material (1.0 equiv.) was dissolved in dry DCM (0.05 M) and cooled 0 °C, *m*-CPBA (6.0 equiv.) was added and the reaction was stirred at rt overnight. The mixture was diluted with DCM and washed with sat. aq. $Na_2S_2O_3$ (1 x), sat. aq. $NaHCO_3$ (1 x), water and brine successively, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The crude was directly used for next step (silylation) without further purification.

General procedure E: Protection with TBS group

The starting material (1.0 equiv.) was dissolved in dry DCM (0.1 M) and cooled 0 °C, imidazole (3.0 equiv.), TBSCl (2.0 equiv.) and DMAP (0.1 equiv.) were added successively and the reaction was stirred at rt overnight. The mixture was diluted with DCM, washed with water (3 x) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (See experimental description below for eluent system).

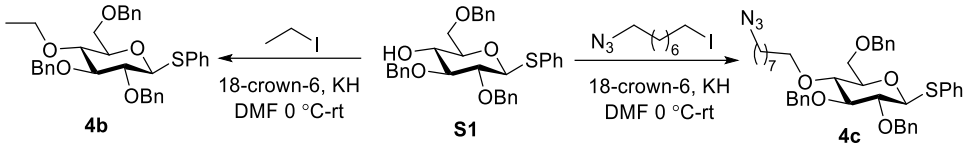
General procedure F: Deprotection of TBS group

The starting material (1.0 equiv.) was dissolved in dry THF (0.1 M), TBAF (1.0 M in THF, 6.0 equiv.) was added. The reaction was stirred at rt until TLC-analysis indicated full consumption of the starting material (± 2h). The reaction was quenched sat. aq. NH_4Cl, extracted with DCM (2 x), the combined organic layers were washed with sat. aq. $NaHCO_3$, water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (See experimental description below for eluent system).

General procedure G: Protection with Bn group

The starting material (1.0 equiv.) was dissolved in dry DMF (0.1 M) and cooled to 0 °C. To the mixture was added NaH (60% in mineral, 2.0 equiv.) and stirred at 0 °C for 30 min. Then BnBr (1.5 equiv.) and TBAI (0.05 equiv.) were added successively and the mixture was warmed to rt and stirred overnight. The reaction was quenched with H_2O at 0 °C, extracted with EtOAc (2 x), the combined organic layers were washed with water (2 x) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (See experimental description below for eluent system).

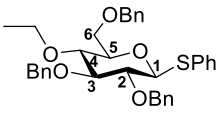
Scheme 2.2. Preparation of donor **4b** and **4c**



For **4b**: KH (30 wt%, 0.36 g, 2.7 mmol) was dispersed in dry DMF (18 mL), purged with N_2 and cooled to 0 °C. Compound **S1** (1.0 g, 1.8 mmol) was added to the mixture in several portions over a period of 15 min, then the reaction was warmed to rt and stirred for 30 min. After which the mixture was cooled to 0 °C again, 18-crown-6 (97 mg, 0.37 mmol) and iodoethane (0.44 mL, 5.5 mmol) were added. After stirring at rt overnight, the reaction was carefully quenched with H_2O at 0 °C, diluted with water (80 mL), extracted with EtOAc (2 x 100 mL), the combined organic layers were washed with water (2 x 150 mL), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by flash column chromatography (pentane/EtOAc 20:1→12:1) affording compound **4b** (906 mg, 1.59 mmol, 86%) as a white solid.

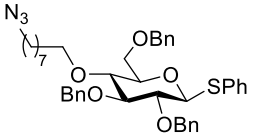
For **4c**: Starting from compound **S1** (1.50 g, 2.76 mmol) and 1-azido-8-iodooctane (1.17 g, 4.15 mmol), the reaction was carried out following the same procedure described for **4b**. The product was purified by flash column chromatography (pentane/EtOAc 20:1→15:1) affording compound **4c** (1.27 g, 1.82 mmol, 66%) as a yellow oil.

Compound 4b



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.63 – 7.53 (m, 2H, CH Ar), 7.41 – 7.24 (m, 15H, CH Ar), 7.23 – 7.17 (m, 3H, CH Ar), 4.90 – 4.81 (m, 3H, CH*H* Bn), 4.71 (d, *J* = 10.3 Hz, 1H, CH*H* Bn), 4.65 (d, *J* = 9.6 Hz, 1H, H1), 4.56 (dd, *J* = 29.6, 11.9 Hz, 2H, CH*H* Bn), 3.86 – 3.76 (m, 2H, H6a and CH_3CH*H*O), 3.75 – 3.69 (m, 1H, H6b), 3.60 (m, 2H, H4 and CH_3CH*H*O), 3.49 – 3.41 (m, 3H, H2, H3 and H5), 1.13 (t, *J* = 7.0 Hz, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.5, 138.4, 138.2, 133.9 ($4C_q$ Ar), 132.0, 129.0, 128.6, 128.5, 128.4, 128.3, 128.0, 127.8, 127.7, 127.6, 127.5 (20CH Ar), 87.5 (C1), 86.7 (C4), 80.7 (C2), 79.3 (C3), 78.1 (C5), 75.9, 75.5, 73.5 ($3CH_2$ Bn), 69.1 (C6), 68.5 (O*C*H_2CH_3), 15.9 (CH_3). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{35}H_{38}O_5SNa$ 593.2332, found 593.2330.

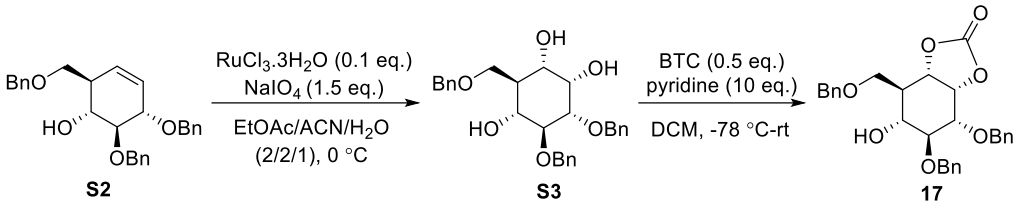
Compound 4c



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.58 (dd, *J* = 6.6, 3.0 Hz, 2H, CH Ar), 7.44 – 7.11 (m, 18H, CH Ar), 4.90 – 4.78 (m, 3H, CH*H* Bn), 4.70 (d, *J* = 10.3 Hz, 1H, CH*H* Bn), 4.65 (d, *J* = 9.6 Hz, 1H, H1), 4.56 (dd, *J* = 29.6, 11.9 Hz, 2H, CH*H* Bn), 3.82 – 3.66 (m, 3H, H6a, H6b and O*CH*H linker), 3.61 (t, *J* = 8.4 Hz, 1H, H4), 3.55 – 3.37 (m, 4H, O*CH*H linker, H2, H3 and H5), 3.22 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 1.66 – 1.40 (m, 4H, $2CH_2$ linker), 1.39 – 1.12 (m, 8H, $4CH_2$ linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.6, 138.5,

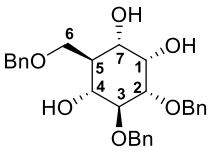
138.2, 134.0 (4C_q Ar), 132.0, 129.0, 128.5, 128.4, 128.3, 128.0, 127.9, 127.8, 127.7, 127.6, 127.5 (20CH Ar), 87.5 (C1), 86.8 (C4), 80.8 (C2), 79.4 (C3), 78.1 (C5), 75.9, 75.5, 73.5 (3CH_2 Bn), 73.3 (OCH_2 linker), 69.2 (C6), 51.5 (CH_2N_3), 30.4, 29.5, 29.2, 28.9, 26.8, 26.2 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{41}H_{49}N_3O_5SNa$ 718.3285, found 718.3281.

Scheme 2.3. Preparation of acceptor **17**



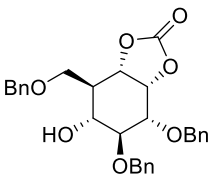
Cyclohexane **S2** (1.94 g, 4.51 mmol) was dissolved in EtOAc (25 mL) and MeCN (25 mL) and cooled to 0 °C. A solution of $RuCl_3.3H_2O$ (94 mg, 0.45 mmol) and $NaIO_4$ (1.45 g, 6.78 mmol) in H_2O (9 mL) was added and the mixture was stirred at 0 °C vigorously for 2 h. The reaction was quenched by addition of sat. aq. $Na_2S_2O_3$ (10 mL) and the mixture was stirred for 15 min. Then the mixture was diluted with H_2O (100 mL) and extracted with EtOAc (3 x 80 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by flash column chromatography (pentane/EtOAc, 5:1 → 2:1) affording compound **S3** (750 mg, 1.61 mmol, 36%) as a clean oil. Compound **S3** (565 mg, 1.21 mmol) was dissolved in dry DCM (5 mL) and pyridine (1.0 mL, 12 mmol) was added. The mixture was purged with N_2 and cooled to -78 °C. A solution of triphosgene (0.18 g, 0.61 mmol) in dry DCM (2 mL) was added to the mixture at -78 °C. Then the reaction was stirred at rt for 1 h and treated with sat. aq. NH_4Cl (60 mL), extracted with DCM (2 x 80 mL). The combined organic layers were washed successively with water, sat. aq. $NaHCO_3$, water and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by flash column chromatography (pentane/acetone 11:1→7:1) affording compound **17** (0.46 g, 0.94 mmol, 78%) as a yellow oil.

Compound S3



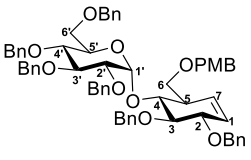
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.49 – 7.11 (m, 15H, CH Ar), 4.99 (d, J = 11.4 Hz, 1H, CH*H* Bn), 4.75 – 4.62 (m, 3H, CH*H* Bn), 4.56 – 4.45 (m, 2H, CH*H* Bn), 4.13 (t, J = 2.7 Hz, 1H, H1), 3.88 (ddd, J = 9.2, 4.0, 1.2 Hz, 1H, H6a), 3.80 – 3.67 (m, 2H, H3 and H6b), 3.53 (dt, J = 10.8, 1.7 Hz, 1H, H7), 3.42 – 3.29 (m, 2H, H2 and H4), 2.89 – 2.63 (br s, 2H, 2OH), 2.16 (tdd, J = 10.5, 6.1, 3.9 Hz, 1H, H5). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.8, 137.9, 137.9 (3C_q Ar), 128.6, 128.6, 128.5, 128.0, 128.0, 127.9, 127.9, 127.7 (15CH Ar), 82.4 (C3), 79.9 (C2), 75.5, 73.7, 72.1 (3CH_2 Bn), 70.3 (C4), 70.2 (C1), 70.17 (C7), 70.0 (C6), 42.5 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{28}H_{32}O_6Na$ 487.2091, found 487.2091.

Compound 17



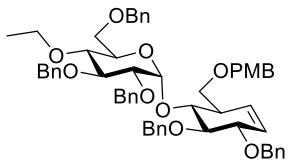
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.20 (m, 15H, CH Ar), 4.82 – 4.72 (m, 2H, H1 and H7), 4.70 – 4.54 (m, 4H, CH*H* Bn), 4.52 (s, 2H, CH*H* Bn), 3.89 – 3.78 (m, 2H, H6a and H2), 3.73 – 3.64 (m, 2H, H3 and H4), 3.61 (dd, *J* = 9.5, 3.9 Hz, 1H, H6b), 2.78 (s, 1H, OH), 2.44 – 2.30 (m, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 154.6 (C=O), 137.7, 137.7, 137.1 ($3C_q$ Ar), 128.7, 128.6, 128.6, 128.3, 128.1, 128.0, 128.0, 127.8, 127.8 (15CH Ar), 81.3 (C3), 76.0 (C2), 75.0 (C1), 74.4 (C7), 73.6, 73.4, 73.2 ($3CH_2$ Bn), 69.6 (C4), 66.7 (C6), 43.2 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{29}H_{30}O_7Na$ 513.1884, found 513.1887.

Compound 6a



Starting from compound **5** (0.46 g, 1.0 mmol), using donor **4a** (1.1 g, 1.5 mmol) and following **Standard procedure A**. The crude was divided into two portions and purified by size exclusion (DCM/MeOH = 1/1) giving the target product as a single isomer (970 mg, α/β > 20/1). Then the compound was further purified by flash column chromatography (pentane/EtOAc 13:1→9:1) to obtain compound **6a** (0.95 g, 0.96 mmol, 96%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.03 (m, 32H, CH Ar), 6.83 – 6.78 (m, 2H, CH Ar), 5.74 (tt, *J* = 4.8, 2.4 Hz, 2H), 5.63 (dt, *J* = 10.1, 2.1 Hz, 1H), 5.00 (d, *J* = 12.1 Hz, 1H), 4.95 – 4.87 (m, 2H), 4.86 – 4.74 (m, 3H), 4.72 – 4.22 (m, 12H), 4.10 (t, *J* = 9.1 Hz, 1H), 4.03 – 3.86 (m, 2H), 3.79 (d, *J* = 16.4 Hz, 5H), 3.71 – 3.65 (m, 1H), 3.65 – 3.58 (m, 2H), 3.56 – 3.45 (m, 3H), 2.67 (ddt, *J* = 8.2, 5.3, 2.7 Hz, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 139.3, 139.0, 138.6, 138.4, 138.2, 138.0, 130.5 ($8C_q$ Ar), 129.8 (C7), 129.2, 129.0, 128.4, 128.4, 128.4, 128.3, 128.3, 128.3, 128.2, 128.1, 128.0, 128.0, 127.9, 127.9, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.6, 127.6, 127.3, 127.1, 127.0, 126.8 (32CH Ar), 126.6 (C1), 113.9, 113.8 (2CH Ar), 97.0 (C1'), 85.0, 82.0, 80.8, 79.6, 77.9, 75.6, 75.1, 74.1, 73.6, 73.2, 72.8, 72.5, 71.8, 70.9, 69.6 (C6), 68.4 (C6'), 55.4 (CH_3 PMB), 43.9 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{63}H_{66}O_{10}Na$ 1005.4548, found 1005.4550.

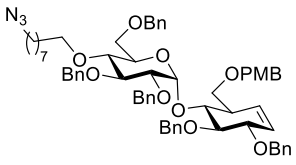
Compound 6b



Starting from compound **5** (0.46 g, 1.0 mmol), using donor **4b** (0.86 g, 1.5 mmol) and following **Standard procedure B**. The crude was divided into two portions and purified by size exclusion (DCM/MeOH = 1/1) giving the target product as a single isomer (845 mg, α/β > 20/1). Then the compound was further purified by flash column chromatography (pentane/EtOAc 13:1→9:1) to obtain compound **6b** (0.81 g, 0.88 mmol, 88%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 – 7.06 (m, 27H, CH Ar), 6.86 (d, *J* = 8.3 Hz, 2H, CH Ar), 5.78 – 5.67 (m, 2H, H1 and H1'), 5.66 – 5.60 (m, 1H, H7), 4.98 (d, *J* = 12.0 Hz, 1H, CH*H* Bn), 4.92 – 4.84 (m, 2H, CH*H* Bn), 4.77 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.59 (td, *J* = 12.7, 12.1, 10.5 Hz, 3H, CH*H* Bn), 4.49 – 4.28 (m, 5H, CH*H* Bn), 4.28 – 4.22 (m, 1H, H2), 4.08 (t, *J* = 9.1 Hz, 1H, H4), 3.94 – 3.85 (m, 2H, H3 and H3'), 3.78 (s, 5H, 1CH*H* Et, CH_3 PMB and H5'), 3.66 –

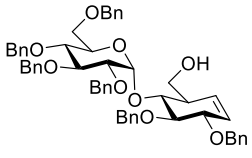
3.39 (m, 7H, H6ab, H6'ab, H4', H2' and 1CH*H* Et), 2.67 (dq, *J* = 6.8, 3.5 Hz, 1H, H5), 1.07 (q, *J* = 6.7 Hz, 3H, CH_3 Et). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 139.4, 139.0, 138.4, 138.3, 138.1, 130.5 (7C_q Ar), 129.8 (C7), 129.1, 128.4, 128.4, 128.4, 128.3, 128.3, 128.2, 128.2, 128.18, 128.0, 128.0, 128.0, 127.9, 127.9, 127.8, 127.8, 127.7, 127.7, 127.6, 127.5, 127.5, 127.1, 126.8 (27CH Ar), 126.6 (C1), 113.8 (2CH Ar), 97.1 (C1'), 85.0 (C3), 81.9 (C3'), 80.8 (C2), 79.4, 77.8, 75.6 , 74.1, 73.6, 73.4 (C4), 72.9, 72.5, 71.8, 71.0 (C5'), 69.7 (C6), 68.4 (CH_2 Et), 68.3 (C6'), 55.4 (CH_3 PMB), 43.9 (C5), 16.0 (CH_3 Et). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{58}H_{64}O_{10}Na$ 943.4392, found 943.4385.

Compound 6c



Starting from compound **5** (560 mg, 1.20 mmol), using donor **4c** (1.27 g, 1.80 mmol) and following **Standard procedure B**. The crude was divided into two portions and purified by size exclusion (DCM/MeOH = 1/1) giving the target product as a single isomer (1.21 g, α/β > 20/1). Then the compound was further purified by flash column chromatography (pentane/EtOAc 13:1→9:1) to obtain compound **6c** (1.10 g, 1.05 mmol, 87%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.02 (m, 27H, CH Ar), 6.90 – 6.81 (m, 2H, CH Ar), 5.77 – 5.68 (m, 2H, H1 and H1'), 5.63 (dt, *J* = 10.1, 2.1 Hz, 1H, H7), 4.98 (d, *J* = 12.0 Hz, 1H, CH*H* Bn), 4.94 – 4.82 (m, 2H, CH*H* Bn), 4.75 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.65 – 4.28 (m, 8H, CH*H* Bn), 4.25 (dt, *J* = 7.4, 2.4 Hz, 1H, H2), 4.08 (t, *J* = 9.1 Hz, 1H, H4)), 3.95 – 3.83 (m, 2H, H3 and H3'), 3.81 – 3.67 (m, 5H, 1CH*H*O linker, CH_3 PMB and H5'), 3.66 – 3.29 (m, 7H, H6ab, H6'ab, H4', H2', 1CH*H*O linker), 3.20 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.67 (dp, *J* = 8.3, 2.8 Hz, 1H, H5), 1.54 (p, *J* = 7.0 Hz, 2H, 2CH*H* linker), 1.43 (h, *J* = 7.8, 6.9 Hz, 2H, 2CH*H* linker), 1.26 (ddd, *J* = 18.1, 12.9, 5.5 Hz, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 139.3, 139.1, 138.4, 138.2, 138.1, 130.5 (7C_q Ar), 129.8 (C7), 129.1, 128.4, 128.4, 128.4, 128.3, 128.2, 128.2, 128.1, 128.0, 127.9, 127.8, 127.8, 127.7, 127.7, 127.6, 127.5, 127.5, 127.1, 126.8 (27CH Ar), 126.6 (C1), 113.8 (2CH Ar), 97.1 (C1'), 85.0 (C3), 81.9 (C3'), 80.8 (C2), 79.4, 77.8, 75.6, 74.1, 73.6, 73.3 (C4), 73.2 (CH_2O linker), 72.9, 72.5, 71.8, 71.0 (C5'), 69.6 (C6), 68.4 (C6'), 55.4 (CH_3 PMB), 51.5 (CH_2N_3), 43.9 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{64}H_{75}N_3O_{10}Na$ 1068.5345, found 1068.5342.

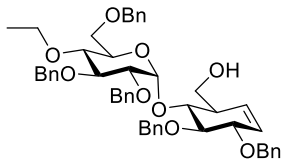
Compound 7a



Starting from compound **6a** (1.04 g, 1.06 mmol), and following **General procedure C**, the reaction was purified by flash column chromatography (pentane/EtOAc 5:1→2:1) to obtain compound **7a** (0.73 g, 0.85 mmol, 80%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 – 7.12 (m, 26H, CH Ar), 7.11 – 6.98 (m, 4H, CH Ar), 5.86 – 5.75 (m, 2H, H1 and H1'), 5.65 (dt, *J* = 10.1, 2.1 Hz, 1H, H7), 5.05 (d, *J* = 12.2 Hz, 1H, CH*H* Bn), 4.88 (dd, *J* = 11.6, 8.1 Hz, 2H, CH*H* Bn), 4.75 (dd, *J* = 10.9, 3.1 Hz, 2H, CH*H* Bn), 4.65 – 4.43 (m, 6H, CH*H* Bn), 4.38 (d, *J* = 10.8 Hz, 1H, CH*H* Bn), 4.27 (ddt, *J* = 7.3, 3.8, 1.9 Hz,

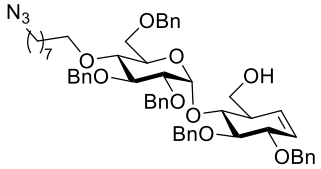
1H, H2), 4.18 (t, *J* = 9.2 Hz, 1H, H4), 4.03 – 3.87 (m, 4H, H3, H3', H5' and H6a), 3.71 – 3.61 (m, 2H, H6'a and H6b), 3.47 (dd, *J* = 9.8, 4.0 Hz, 1H, H2'), 3.40 (dd, *J* = 9.8, 7.1 Hz, 1H, H6'b), 3.29 (dd, *J* = 10.2, 8.7 Hz, 1H, H4'), 3.10 (br s, 1H, OH), 2.57 (dp, *J* = 8.8, 3.0 Hz, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 139.4, 138.7, 138.3, 138.1, 137.8, 137.3 ($6C_q$ Ar), 129.8 (C7), 128.6, 128.6, 128.5, 128.5, 128.3, 128.3, 128.3, 128.3, 128.1, 128.1, 128.0, 127.9, 127.7, 127.7, 127.6, 127.5 (C1), 127.0, 126.5, 97.3 (C1'), 84.5 (C3), 82.0 (C3'), 80.9 (C2), 79.2 (C2'), 78.4 (C4'), 75.6, 75.2, 73.8, 73.6, 73.1(C4), 73.0, 71.7, 71.5 (C5'), 69.1 (C6'), 61.9 (C6), 45.8 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{55}H_{58}O_9Na$ 885.3973, found 885.3971.

Compound 7b



Starting from compound **6b** (0.78 g, 0.85 mmol), and following **General procedure C**, the reaction was purified by flash column chromatography (pentane/EtOAc 7:1→3:1) to obtain compound **7b** (0.50 g, 0.62 mmol, 74%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) *δ* 7.45 – 7.02 (m, 25H, CH Ar), 5.81 (dt, *J* = 10.2, 2.4 Hz, 1H, H1), 5.75 (d, *J* = 4.0 Hz, 1H, H1'), 5.65 (dt, *J* = 10.2, 2.1 Hz, 1H, H7), 5.02 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.94 – 4.78 (m, 2H, CH*H* Bn), 4.73 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.63 – 4.58 (m, 3H, CH*H* Bn), 4.54 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 4.46 (s, 2H, CH*H* Bn), 4.26 (ddt, *J* = 7.3, 3.1, 2.0 Hz, 1H, H2), 4.16 (dd, *J* = 9.6, 8.7 Hz, 1H, H4), 4.02 – 3.81 (m, 4H, H6a, H3, H5' and H3'), 3.76 (dq, *J* = 9.2, 7.0 Hz, 1H, CH*H* Et), 3.67 (td, *J* = 11.1, 10.5, 2.3 Hz, 2H, H6'a and H6b), 3.50 (dd, *J* = 9.9, 6.9 Hz, 1H, H6'b), 3.45 – 3.36 (m, 2H, H2' and CH*H* Et), 3.10 (dd, *J* = 10.1, 8.7 Hz, 1H, H4'), 3.02 (t, *J* = 6.3 Hz, 1H, OH), 2.58 (dh, *J* = 8.8, 2.9 Hz, 1H, H5), 1.04 (t, *J* = 7.0 Hz, 3H, CH_3 Et). ^{13}C NMR (126 MHz, $CDCl_3$) *δ* 139.5, 138.8, 138.3, 138.2, 137.4 ($5C_q$ Ar), 129.8 (C7), 128.6, 128.5, 128.4, 128.3, 128.2, 128.2, 128.1, 128.0, 127.9, 127.7, 127.7, 127.6, 127.6, 127.5, 127.0 (C1), 126.6, 97.4 (C1'), 84.5 (C3), 81.8 (C3'), 80.8 (C2), 79.1 (C2'), 78.8 (C4'), 75.6, 73.9, 73.7 ($3CH_2$ Bn), 73.3 (C4), 73.1, 71.7 ($2CH_2$ Bn), 71.6 (C5'), 69.3 (C6'), 68.6 (CH_2 Et), 62.1 (C6), 45.8 (C5), 15.8 (CH_3 Et). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{50}H_{56}O_9Na$ 823.3817, found 823.3821.

Compound 7c

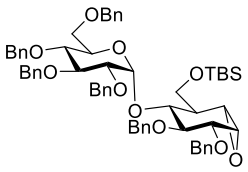


Starting from compound **6c** (1.10 g, 1.05 mmol) and following **General procedure C**, The reaction was purified by flash column chromatography (pentane/EtOAc 7:1→3:1) to obtain compound **7c** (0.70 g, 0.75 mmol, 71%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.38 – 7.11 (m, 23H, CH Ar), 7.10 – 7.02 (m, 2H, CH Ar), 5.81 (dt, *J* = 10.2, 2.4 Hz, 1H, H1), 5.76 (d, *J* = 4.0 Hz, 1H, H1'), 5.65 (dt, *J* = 10.1, 2.1 Hz, 1H, H7), 5.03 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.92 – 4.81 (m, 2H, CH*H* Bn), 4.71 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.64 – 4.50 (m, 4H, CH*H* Bn), 4.46 (s, 2H, CH*H* Bn), 4.26 (ddt, *J* = 7.3, 3.4, 2.0 Hz, 1H, H2), 4.17 (t, *J* = 9.1 Hz, 1H, H4), 4.01 – 3.81 (m, 4H, H6a, H3, H5' and H3'), 3.69 (tdd, *J* = 10.4, 7.6, 2.6 Hz, 3H, CH*H*O linker, H6'a and H6b), 3.50 (dd, *J* = 9.9, 6.8 Hz, 1H, H6'b),

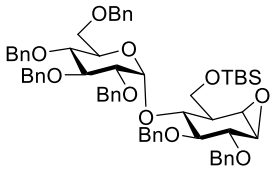
3.43 (dd, *J* = 9.8, 4.0 Hz, 1H, H2'), 3.31 (dt, *J* = 9.0, 6.8 Hz, 1H, CH*H*O linker), 3.21 (t, *J* = 6.9 Hz, 2H, CH_2N_3), 3.10 (dd, *J* = 10.1, 8.7 Hz, 1H, H4'), 2.58 (dh, *J* = 8.7, 2.8 Hz, 1H, H5), 1.54 (p, *J* = 7.0 Hz, 2H, 2CH*H* linker), 1.40 (dq, *J* = 13.4, 6.7 Hz, 2H, 2CH*H* linker), 1.33 – 1.05 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.3, 138.7, 138.2, 138.1, 137.3 ($5C_q$ Ar), 129.7 (C7), 128.5, 128.4, 128.3, 128.2, 128.1, 128.1, 127.9, 127.9, 127.7, 127.6, 127.6, 127.5, 127.5, 127.4, 126.9 (C1), 126.4, 97.3 (C1'), 84.4 (C3), 81.7 (C3'), 80.7 (C2), 79.0 (C2'), 78.8 (C4'), 75.4, 73.8, 73.6 ($3CH_2$ Bn), 73.3 (CH_2O linker), 73.2 (C4), 72.9, 71.6 ($2CH_2$ Bn), 71.5 (C5'), 69.2 (C6'), 62.0 (C6), 51.4 (CH_2N_3), 45.7 (C5), 30.2, 29.3, 29.0, 28.8, 26.7, 26.0 ($6CH_2$ linker). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{56}H_{67}N_3O_9Na$ 948.4770, found 948.4769.

Compound 8a and 9a

Starting from compound **7a** (0.70 mg, 0.81 mmol) and following **General procedure D and E**, the product was purified by flash column chromatography (pentane/Et_2O 13:1→9:1) to obtain compound **8a** (0.16 g, 0.16 mmol, 20%) as a clean oil and compound **9a** (0.47 g, 0.47 mmol, 58%) as a clean oil.



8a: ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 – 7.10 (m, 30H, CH Ar), 5.34 (d, *J* = 3.6 Hz, 1H, H1'), 4.96 (dd, *J* = 27.9, 11.2 Hz, 2H, CH*H* Bn), 4.89 – 4.79 (m, 3H, CH*H* Bn), 4.78 – 4.67 (m, 2H, CH*H* Bn), 4.59 (dd, *J* = 17.7, 12.2 Hz, 2H, CH*H* Bn), 4.51 – 4.43 (m, 3H, CH*H* Bn), 4.02 (t, *J* = 9.4 Hz, 1H, H3'), 3.95 (t, *J* = 9.1 Hz, 1H, H3), 3.88 (dt, *J* = 9.8, 3.0 Hz, 1H, H5'), 3.85 – 3.75 (m, 3H, H2, H4 and H6a), 3.75 – 3.70 (m, 2H, H6b and H6'a), 3.66 (dd, *J* = 10.0, 9.0 Hz, 1H, H4'), 3.58 (dd, *J* = 10.4, 2.0 Hz, 1H, H6'b), 3.48 (dd, *J* = 9.9, 3.6 Hz, 1H, H2'), 3.31 (d, *J* = 4.2 Hz, 1H, epoxide), 3.08 (d, *J* = 4.2 Hz, 1H, epoxide), 2.39 (dq, *J* = 8.4, 4.8, 4.1 Hz, 1H, H5), 0.87 (s, 9H, $(CH_3)_3CSi$), 0.02 (d, *J* = 5.2 Hz, 6H, $2CH_3Si$). ^{13}C NMR (126 MHz, $CDCl_3$) δ 139.6, 138.8, 138.5, 138.4, 138.4, 138.0 ($6C_q$ Ar), 128.5, 128.5, 128.3, 128.2, 128.1, 128.1, 127.9, 127.9, 127.8, 127.7, 127.6, 127.5, 127.0, 127.0, 98.7 (C1'), 81.9 (C3'), 80.9 (C3), 79.6 (C2'), 79.0 (C2), 78.6 (C4), 78.0 (C4'), 75.7, 75.1, 74.7, 73.7, 73.1, 72.9 ($6CH_2$ Bn), 71.1 (C5'), 68.6 (C6'), 63.0 (C6), 55.7 (epoxide), 55.2 (epoxide), 44.1 (C5), 26.0 ($(CH_3)_3CSi$), 18.2 ($(CH_3)_3CSi$), -5.2 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{61}H_{72}O_{10}SiNa$ 1015.4787, found 1015.4792.

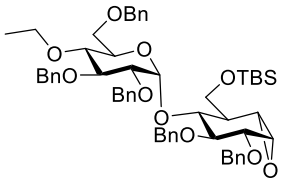


9a:^{1}H NMR (400 MHz, $CDCl_3$) δ 7.55 – 6.83 (m, 30H, CH Ar), 5.70 (d, *J* = 3.6 Hz, 1H, H1'), 4.93 – 4.67 (m, 6H, CH*H* Bn), 4.64 – 4.54 (m, 2H, CH*H* Bn), 4.49 – 4.39 (m, 4H, CH*H* Bn), 4.15 (dd, *J* = 9.1, 4.3 Hz, 1H, H6a), 3.94 (dd, *J* = 9.9, 8.7 Hz, 1H, H3'), 3.90 – 3.85 (m, 1H, H3), 3.76 (dt, *J* = 10.1, 2.6 Hz, 1H, H5'), 3.74 – 3.59 (m, 6H, H6'ab, H6b, H2, H4 and H4'), 3.53 (dt, *J* = 4.1, 1.0 Hz, 1H, epoxide), 3.45 (dd, *J* = 9.9, 3.7 Hz, 1H, H2'), 3.22 (d, *J* = 3.8 Hz, 1H, epoxide), 2.36 (dtd, *J* = 10.5, 5.4, 2.5 Hz, 1H, H5), 0.87 (s, 9H, $(CH_3)_3CSi$), 0.04 (d, *J* = 2.1 Hz, 6H, $2CH_3Si$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.1, 138.8, 138.7, 138.2, 138.0, 137.6 ($6C_q$ Ar), 128.6, 128.5, 128.4, 128.3, 128.3, 128.1, 128.1, 128.0, 128.0, 127.8, 127.7, 127.6, 127.6, 127.5, 127.1, 126.5, 97.0 (C1'), 85.0, 81.9 (C3'), 80.3 (C3), 79.3 (C2'),

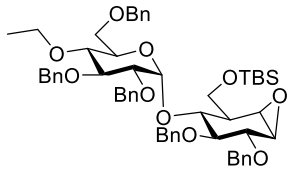
77.7 (C4'), 75.7, 74.8, 73.8, 73.7, 73.0, 72.9 (6CH_2 Bn), 71.5 (C5'), 70.0, 68.4 (C6'), 62.8 (C6), 55.0 (epoxide), 53.8 (epoxide), 43.8 (C5), 26.1 ((*C*H_3)$_3$CSi), 18.5 ((CH$_3$)$_3$*C*Si), -5.3 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{61}H_{72}O_{10}SiNa$ 1015.4787, found 1015.4789.

Compound 8b and 9b

Starting from compound **7b** (447 mg, 0.558 mmol) and following **General procedure D and E**, the product was purified by flash column chromatography (pentane/Et_2O 11:1→5:1) to obtain compound **8b** (108 mg, 0.116 mmol, 21%) as a light yellow oil and compound **9b** (338 mg, 0.363 mmol, 65%) as a light yellow oil.



8b:^{1}H NMR (400 MHz, $CDCl_3$) δ 7.51 – 7.01 (m, 25H, CH Ar), 5.34 (d, *J* = 3.6 Hz, 1H, H1'), 5.02 (d, *J* = 11.6 Hz, 1H, CH*H* Bn), 4.94 – 4.81 (m, 3H, CH*H* Bn), 4.80 – 4.70 (m, 2H, CH*H* Bn), 4.67 (d, *J* = 12.0 Hz, 1H, CH*H* Bn), 4.60 (d, *J* = 12.3 Hz, 1H, CH*H* Bn), 4.49 (dd, *J* = 14.5, 12.1 Hz, 2H, CH*H* Bn), 4.01 – 3.91 (m, 2H), 3.88 – 3.65 (m, 7H), 3.61 (dd, *J* = 10.3, 2.1 Hz, 1H), 3.57 – 3.41 (m, 3H), 3.36 – 3.31 (m, 1H, H1), 3.10 (dd, *J* = 4.3, 0.8 Hz, 1H, H7), 2.43 (dt, *J* = 7.1, 3.4 Hz, 1H, H5), 1.11 (t, *J* = 7.0 Hz, 3H, CH_3 Et), 0.90 (s, 9H, (CH$_3$)$_3$CSi), 0.05 (d, *J* = 4.2 Hz, 6H, 2CH_3Si). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.6, 138.9, 138.4, 138.4, 138.0 (5C_q Ar), 128.5, 128.4, 128.3, 128.2, 128.1, 127.8, 127.8, 127.6, 127.6, 127.4, 127.0, 127.0, 98.7 (C1'), 81.8, 80.9, 79.4, 79.0, 78.5, 78.0, 75.7 (CH_2 Bn), 74.8 (CH_2 Bn), 73.7 (CH_2 Bn), 73.1 (CH_2 Bn), 72.9 (CH_2 Bn), 71.21 (C5'), 68.6 (C6'), 68.5 (CH_2 Et), 63.0 (C6), 55.7 (C7), 55.2 (C1), 44.1 (C5), 26.0 ((*C*H_3)$_3$CSi), 18.2 ((CH$_3$)$_3$*C*Si), 15.9 (CH_3 Et), -5.2 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{56}H_{70}O_{10}SiNa$ 953.4631, found 953.4635.

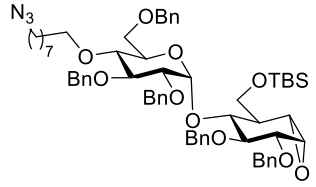


9b:^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 6.93 (m, 25H, CH Ar), 5.67 (d, *J* = 3.6 Hz, 1H, H1'), 4.98 – 4.36 (m, 10H, CH*H* Bn), 4.14 (dd, *J* = 9.2, 4.3 Hz, 1H, H6a), 3.90 – 3.82 (m, 2H), 3.80 – 3.56 (m, 7H), 3.52 (dt, *J* = 3.8, 1.2 Hz, 1H, epoxide), 3.49 – 3.35 (m, 3H), 3.21 (d, *J* = 3.8 Hz, 1H, epoxide), 2.35 (tdd, *J* = 9.4, 4.3, 1.5 Hz, 1H, H5), 1.03 (t, *J* = 7.0 Hz, 3H, CH_3 Et), 0.91 (s, 9H, (CH$_3$)$_3$CSi), 0.07 (d, *J* = 1.9 Hz, 6H, 2CH_3Si). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.1, 139.0, 138.3, 138.1, 137.6 (5C_q Ar), 128.6, 128.4, 128.4, 128.3, 128.2, 128.1, 128.0, 127.9, 127.8, 127.6, 127.6, 127.4, 127.1, 126.5, 97.0 (C1'), 85.0, 81.8, 80.3, 79.2, 77.7, 75.6 (CH_2 Bn), 73.8 (CH_2 Bn), 73.7 (CH_2 Bn), 73.1 (CH_2 Bn), 72.8 (CH_2 Bn), 71.6, 70.0, 68.4 (C6'), 68.2 (CH_2 Et), 62.8 (C6), 55.0 (epoxide), 53.7 (epoxide), 43.7 (C5), 26.1 ((*C*H_3)$_3$CSi), 18.5 ((CH$_3$)$_3$*C*Si), 15.9 (CH_3 Et), -5.3 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{56}H_{70}O_{10}SiNa$ 953.4631, found 953.4632.

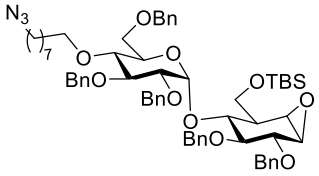
Compound 8c and 9c

Starting from compound **7c** (666 mg, 0.719 mmol) and following **General procedure D and E**, the product was purified by flash column chromatography (pentane/Et_2O 11:1→7:1) to obtain compound

8c (165 mg, 0.156 mmol, 22%) as a yellow oil and compound **9c** (478 mg, 0.453 mmol, 63%)as a yellow oil.

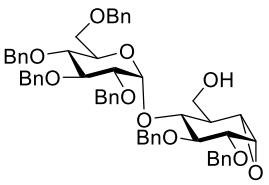


8c:^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.13 (m, 25H, CH Ar), 5.33 (d, *J* = 3.6 Hz, 1H, H1'), 5.12 – 4.34 (m, 10H, CH*H* Bn), 4.02 – 3.89 (m, 2H), 3.86 – 3.65 (m, 7H), 3.60 (dd, *J* = 10.3, 2.1 Hz, 1H), 3.50 – 3.37 (m, 3H), 3.36 – 3.31 (m, 1H, epoxide), 3.28 (t, *J* = 6.9 Hz, 2H, CH_2N_3), 3.11 (dd, *J* = 4.2, 0.9 Hz, 1H, epoxide), 2.43 (dt, *J* = 6.7, 3.4 Hz, 1H, H5), 1.69 – 1.55 (m, 2H, 2CH*H* linker), 1.48 (dt, *J* = 8.7, 2.8 Hz, 2H, 2CH*H* linker), 1.39 – 1.24 (m, 8H, 8CH*H* linker), 0.90 (s, 9H, $(CH_3)_3CSi$), 0.09 – 0.03 (m, 6H, $2CH_3Si$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.6, 138.9, 138.4, 138.4, 138.0 ($5C_q$ Ar), 128.4, 128.4, 128.3, 128.2, 128.1, 128.0, 128.0, 127.8, 127.7, 127.6, 127.4, 127.0, 98.7 (C1'), 81.8, 80.9, 79.4, 79.0, 78.6, 78.0, 75.6 (CH_2 Bn), 74.8 (CH_2 Bn), 73.7 (CH_2 Bn), 73.2 (CH_2O linker), 73.1 (CH_2 Bn), 72.9 (CH_2 Bn), 71.2, 68.7 (C6'), 63.0 (C6), 55.7 (epoxide), 55.2 (epoxide), 51.6 (CH_2N_3), 44.1 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 ($6CH_2$ linker), 26.0 ((CH_3)$_3$CSi), 18.1 ((CH_3)$_3$*C*Si), -5.2 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{81}N_3O_{10}SiNa$ 1078.5583, found 1078.5581.



9c:^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 – 6.88 (m, 25H, CH Ar), 5.69 (d, *J* = 3.7 Hz, 1H, H1'), 5.00 – 4.28 (m, 10H, CH*H* Bn), 4.15 (dd, *J* = 9.1, 4.3 Hz, 1H, H6a), 3.90 – 3.80 (m, 2H), 3.78 – 3.56 (m, 7H), 3.56 – 3.50 (m, 1H, epoxide), 3.46 – 3.29 (m, 3H), 3.25 – 3.17 (m, 3H, 1H epoxide and CH_2N_3), 2.36 (tdd, *J* = 9.4, 4.4, 2.4 Hz, 1H, H5), 1.55 (dq, *J* = 8.3, 6.8 Hz, 2H, 2CH*H* linker), 1.46 – 1.33 (m, 2H, 2CH*H* linker), 1.32 – 1.08 (m, 8H, 8CH*H* linker), 0.91 (s, 9H, $(CH_3)_3CSi$), 0.08 (d, *J* = 2.3 Hz, 6H, $2CH_3Si$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.1, 138.9, 138.2, 138.0, 137.6 ($5C_q$ Ar), 128.5, 128.4, 128.3, 128.2, 128.2, 128.0, 128.0, 127.8, 127.7, 127.6, 127.5, 127.4, 127.0, 126.4, 96.9 (C1'), 85.1, 81.8, 80.2, 79.1, 77.7, 75.5 (CH_2 Bn), 73.7 (CH_2 Bn), 73.6 (CH_2 Bn), 73.0 (CH_2 Bn), 72.9 (CH_2O linker), 72.7 (CH_2 Bn), 71.6, 69.7, 68.4 (C6'), 62.8 (C6), 54.9 (epoxide), 53.7 (epoxide), 51.5 (CH_2N_3), 43.7 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 ($6CH_2$ linker), 26.1 ((CH_3)$_3$CSi), 18.4 ((CH_3)$_3$*C*Si), -5.3 (CH_3Si), -5.4 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{81}N_3O_{10}SiNa$ 1078.5583, found 1078.5581.

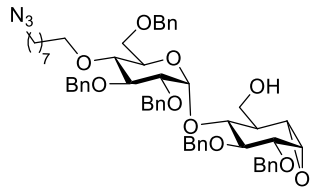
Compound 10a



Starting from compound **8a** (0.10 g, 0.10 mmol) and following **General procedure F**, the product was purified by flash column chromatography (pentane/EtOAc 4:1→2:1) to obtain compound **10a** (80 mg, 0.09 mmol, 90%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 6.99 (m, 30H, CH Ar), 5.58 (d, *J* = 4.0 Hz, 1H, H1'), 4.97 – 4.83 (m, 3H, CH*H* Bn), 4.76 – 4.63 (m, 4H, CH*H* Bn), 4.58 (d, *J* = 12.3 Hz, 1H, CH*H* Bn), 4.55 – 4.45 (m, 2H, CH*H* Bn), 4.39 (dd, *J* = 11.5, 7.7 Hz, 2H, CH*H* Bn), 4.01 – 3.75 (m, 7H, H3', H5', H6ab, H2, H3 and H4), 3.66 (dd, *J* = 9.9, 1.7 Hz, 1H, H6'a), 3.47 – 3.31 (m,

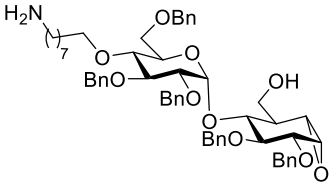
3H, H2', H6'b and epoxide), 3.29 – 3.08 (m, 2H, H4' and epoxide), 2.34 – 2.25 (m, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.5, 138.5, 138.2, 138.1, 137.7, 137.1 (6C_q Ar), 128.7, 128.6, 128.5, 128.4, 128.3, 128.3, 128.2, 128.1, 128.1, 128.0, 127.8, 127.7, 127.5, 127.0, 126.5, 97.8 (C1'), 81.8, 81.4, 79.8, 79.0 (C2'), 78.5 (C4'), 75.6 (CH_2 Bn), 75.1 (CH_2 Bn), 74.6 (CH_2 Bn), 74.5, 73.7 (CH_2 Bn), 73.2 (CH_2 Bn), 72.8 (CH_2 Bn), 71.5 (C5'), 69.5 (C6'), 61.6 (C6), 55.5 (epoxide), 54.9 (epoxide), 44.2 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{55}H_{58}O_{10}Na$ 901.3922, found 901.3920.

Compound 10b



Starting from compound **8c** (0.15 g, 0.14 mmol) and following **General procedure F**, the product was purified by flash column chromatography (pentane/EtOAc 3:1→1:1) to obtain compound **10c** (117 mg, 0.124 mmol, 86%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 – 7.00 (m, 25H, CH Ar), 5.53 (d, *J* = 3.9 Hz, 1H, H1'), 4.91 (s, 2H, CH*H* Bn), 4.82 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.71 (d, *J* = 7.7 Hz, 3H, CH*H* Bn), 4.64 – 4.53 (m, 2H, CH*H* Bn), 4.50 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.39 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 3.97 – 3.76 (m, 7H, H3', H5', H6ab, H2, H3 and H4), 3.76 – 3.66 (m, 2H, H6'a and CH*H*O linker), 3.49 (dd, *J* = 9.8, 7.6 Hz, 1H, H6'b), 3.43 – 3.26 (m, 3H, H2', epoxide and CH*H*O linker), 3.26 – 3.19 (m, 3H, epoxide and CH_2N_3), 3.03 (dd, *J* = 10.1, 8.5 Hz, 1H, H4'), 2.31 (ddd, *J* = 7.7, 4.4, 2.6 Hz, 1H, H5), 1.56 (p, *J* = 7.0 Hz, 2H, 2CH*H* linker), 1.41 (p, *J* = 6.9, 6.5 Hz, 2H, 2CH*H* linker), 1.36 – 1.11 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.5, 138.6, 138.2, 138.1, 137.2 (5C_q Ar), 128.6, 128.4, 128.4, 128.2, 128.2, 128.2, 128.1, 128.1, 127.8, 127.8, 127.6, 127.5, 127.4, 126.9, 126.5, 97.9 (C1'), 81.5, 81.4, 79.7, 79.0 (C4'), 78.9 (C2'), 75.4 (CH_2 Bn), 74.9, 74.6 (CH_2 Bn), 73.7 (CH_2 Bn), 73.3 (CH_2O linker), 73.2 (CH_2 Bn), 72.7 (CH_2 Bn), 71.7, 69.7 (C6'), 61.6 (C6), 55.4 (epoxide), 54.8 (epoxide), 51.5 (CH_2N_3), 44.2 (C5), 30.3, 29.4, 29.1, 28.9, 26.7, 26.0 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{56}H_{67}N_3O_{10}Na$ 964.4719, found 964.4717.

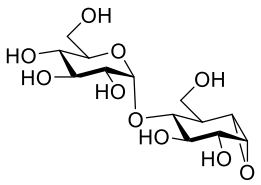
Compound 11



Compound **10b** (110 mg, 0.117 mmol) was dissolved in MeCN (2.4 mL), polymer-bound triphenylphosphine (3 mmol/g loading, 973 mg, 0.292 mmol) and H_2O (21.1 µL, 1.17 mmol) were added and the mixture was stirred overnight at 70 °C. TLC-analysis indicated total consumption of the starting material, and additional H_2O (500 µL) was added and the mixture was stirred for 5 h at 70 °C. The mixture was cooled to rt, filtered, washed with MeCN (3 x 5 mL). The combined filtrate was diluted with MeCN (50 mL), dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by flash column chromatography (DCM/MeOH 19:1→15:1) affording compound **11** (84 mg, 92 µmol, 79%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 – 6.99 (m, 25H, CH Ar), 5.55 (d, *J* = 3.9 Hz, 1H, H1'), 4.90 (s, 2H, CH*H* Bn), 4.82 (d, *J* = 11.0 Hz, 1H, CH*H* Bn), 4.69 (q, *J* = 6.3, 5.6 Hz, 3H, CH*H*

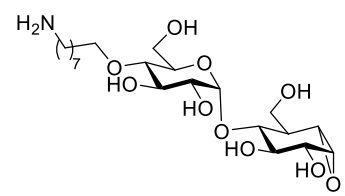
Bn), 4.57 (s, 2H, CH*H* Bn), 4.49 (d, *J* = 12.2 Hz, 1H, CH*H* Bn), 4.39 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 3.85 (ddt, *J* = 22.1, 12.8, 9.4 Hz, 7H, H3', H5', H6ab, H2, H3 and H4), 3.75 – 3.65 (m, 2H, H6'a and CH*H*O linker), 3.54 – 3.43 (m, 1H, H6'b), 3.40 – 3.24 (m, 3H, H2', epoxide and CH*H*O linker), 3.22 (d, *J* = 4.1 Hz, 1H, epoxide), 2.86 (t, *J* = 7.7 Hz, 2H, C*H*$_2$NH$_2$), 2.31 (dt, *J* = 7.7, 3.4 Hz, 1H, H5), 1.66 (dq, *J* = 15.4, 7.7, 7.0 Hz, 2H, 2CH*H* linker), 1.45 – 1.35 (m, 2H, 2CH*H* linker), 1.36 – 1.10 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 139.5, 138.7, 138.2, 138.1, 137.3 (5C$_q$ Ar), 128.6, 128.5, 128.4, 128.4, 128.2, 128.2, 128.2, 128.1, 127.9, 127.8, 127.7, 127.5, 127.5, 127.0, 126.5, 97.9 (C1'), 81.6, 81.5, 79.8, 79.0 (C4'), 78.9 (C2'), 75.4, 74.8, 74.6, 73.8, 73.4 (CH_2O linker), 73.2, 72.8, 71.7, 69.7 (C6'), 61.7 (C6), 55.4 (epoxide), 54.9 (epoxide), 44.3 (C5), 40.4 (CH_2NH_2), 30.3, 29.4, 29.0, 28.8, 26.6, 26.1 (6CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{56}H_{70}NO_{10}$ 916.4994, found 916.4993.

Compound 1a



Compound **10a** (21 mg, 24 μmol) was dissolved in a mixture of MeOH/H_2O/dioxane (2/1/2, 1 mL) under Argon and $Pd(OH)_2$/C (20 wt%, 25 mg, 36 μmol) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 3 h under H_2 atmosphere, the mixture was filtered over a small celite pad and evaporate to afford the product in high purity as a white powder (7.8 mg, quant) after lyophilization. ^{1}H NMR (400 MHz, MeOD) *δ* 5.02 (d, *J* = 3.9 Hz, 1H, H1'), 3.90 (dd, *J* = 11.1, 3.0 Hz, 1H, H6a), 3.85 – 3.76 (m, 3H, H6b, H6'a and H4'), 3.74 – 3.66 (m, 2H, H6'b and H2), 3.66 – 3.54 (m, 2H, H3 and H3'), 3.44 (dd, *J* = 9.7, 3.9 Hz, 1H, H2'), 3.41 – 3.33 (m, 1H, H4), 3.30 – 3.25 (m, 2H, epoxide and H5'), 3.21 (d, *J* = 4.0 Hz, 1H, epoxide), 2.05 (ddd, *J* = 9.0, 5.6, 3.0 Hz, 1H, H5). ^{13}C NMR (101 MHz, MeOD) *δ* 101.1 (C1'), 80.8 (C4), 72.9 (C3'), 72.7 (C3), 72.3 (C2), 71.8 (C2'), 70.5 (C4'), 69.3 (C5'), 60.5 (C6'), 59.5 (C6), 55.8 (epoxide), 53.4 (epoxide), 43.1 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{22}O_{10}Na$ 361.1105, found 361.1102.

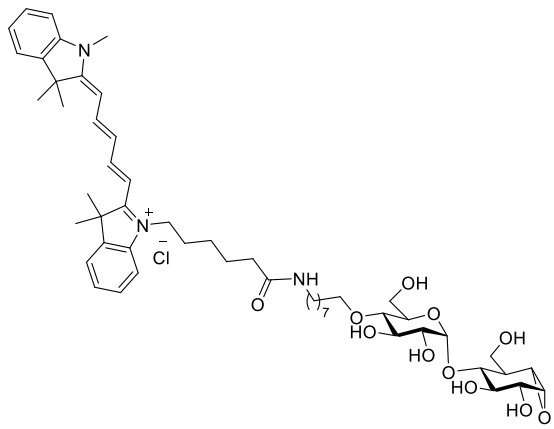
Compound 1b



Compound **11** (30 mg, 33 μmol) was dissolved in a mixture of ^{t}BuOH/H_2O/dioxane (1/2/1, 1.6 mL) under Argon, then HOAc (19 μL, 0.33 mmol) and $Pd(OH)_2$/C (20 wt%, 35 mg, 50 μmol) were added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 6.5 h under H_2 atmosphere, the mixture was filtered over a small celite pad. The filtrate was concentrated and purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The product was obtained as a white powder (13.9 mg, 29.9 μmol, 91%) after lyophilization. ^{1}H NMR (850 MHz, D_2O) *δ* 5.11 (d, *J* = 4.0 Hz, 1H, H1'), 3.92 (dd, *J* = 8.9, 1.8 Hz, 1H, H2), 3.87 (dd, *J* = 11.4, 2.9 Hz, 1H, H6a), 3.82 – 3.77 (m, 2H, CH*H*O linker and H6'a), 3.76 – 3.67 (m, 4H, H3', H5', H6b and H6'b), 3.65 – 3.59 (m, 2H, H3 and CH*H*O linker), 3.54 (ddd, *J* = 9.9, 4.5 Hz, 1H, H2'), 3.46 – 3.42 (m, 2H, H4 and epoxide),

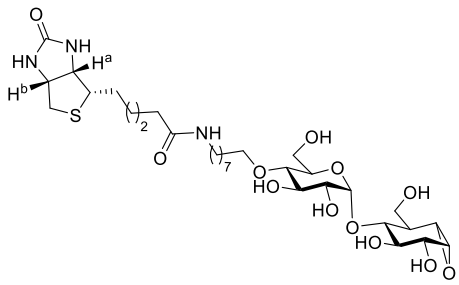
3.32 (d, *J* = 4.2 Hz, 1H, epoxide), 3.24 (t, *J* = 9.6 Hz, 1H, H4'), 2.94 (t, *J* = 7.6 Hz, 2H, C*H*$_2$NH$_2$), 2.19 (ddd, *J* = 9.0, 5.7, 2.9 Hz, 1H, H5), 1.64 – 1.58 (m, 2H, 2CH*H* linker), 1.58 – 1.51 (m, 2H, 2CH*H* linker), 1.35 – 1.26 (m, 8H, 8CH*H* linker). ^{13}C NMR (214 MHz, D$_2$O) *δ* 100.6 (C1'), 79.9 (C4), 77.6 (C4'), 73.4 (C3), 73.4 (CH$_2$O linker), 72.9 (C3'), 71.8 (C2'), 71.6 (C5'), 70.6 (C2), 60.3 (C6), 60.2 (C6'), 57.3 (epoxide), 55.2 (epoxide), 42.8 (C5), 39.5 (CH$_2$NH$_2$), 29.2, 28.2, 28.1, 26.7, 25.5, 25.1 (6CH$_2$ linker). HRMS (ESI) m/z: [M+H]$^+$ calc for $C_{21}H_{40}NO_{10}$ 466.2647, found 466.2644.

Compound 1c



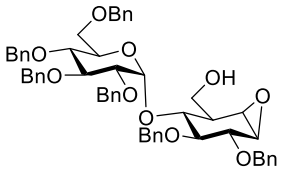
Compound **1b** (5.4 mg, 12 μmol) was dissolved in dry DMF (0.5 mL), then DIPEA (6.0 μL, 34 μmol) and Cy5-OSu (7.9 mg, 13 μmol) were added and the mixture was stirred at rt for 40 h. Full conversion was observed by LC-MS and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain the title compound (3.7 mg, 3.9 μmol, 33%) as a blue powder. ^{1}H NMR (500 MHz, MeOD) *δ* 8.30 – 8.19 (m, 2H), 7.49 (d, *J* = 7.4 Hz, 2H), 7.41 (tdd, *J* = 7.7, 3.8, 1.2 Hz, 2H), 7.33 – 7.23 (m, 4H), 6.63 (t, *J* = 12.4 Hz, 1H), 6.28 (dd, *J* = 13.7, 5.9 Hz, 2H), 5.00 (d, *J* = 3.9 Hz, 1H, H1'), 4.11 (t, *J* = 7.4 Hz, 2H, CH$_2$N$^+$), 3.92 – 3.83 (m, 2H, H6a and CH*H*O linker), 3.82 – 3.72 (m, 3H, H2, H6b and H6'a), 3.72 – 3.65 (m, 3H, H3', H5' and H6'b), 3.63 (s, 3H, CH$_3$N), 3.62 – 3.52 (m, 2H, H3 and CH*H*O linker), 3.43 (dd, *J* = 9.7, 3.9 Hz, 1H, H2'), 3.36 – 3.33 (m, 1H, H4), 3.27 (dd, *J* = 4.0, 1.8 Hz, 1H, epoxide), 3.22 – 3.18 (m, 1H, epoxide), 3.19 – 3.07 (m, 3H, H4' and C*H*$_2$NH), 2.20 (t, *J* = 7.3 Hz, 2H, CH$_2$C=O), 2.04 (ddt, *J* = 11.5, 6.0, 2.6 Hz, 1H, H5), 1.83 (dt, *J* = 15.0, 7.4 Hz, 2H, 2CH*H* linker), 1.73 (s, 14H, 4CH$_3$ and 2CH*H* linker), 1.63 – 1.50 (m, 2H, 2CH*H* linker), 1.50 – 1.41 (m, 4H, 4CH*H* linker), 1.32 (dq, *J* = 5.6, 2.9, 2.3 Hz, 8H, 8CH*H* linker). ^{13}C NMR (126 MHz, MeOD) *δ* 175.7, 175.4, 174.7, 155.5, 155.5, 144.3, 143.6, 142.7, 142.5, 129.7, 129.7, 126.7, 126.2, 126.2, 123.4, 123.3, 112.1, 111.8, 104.4, 104.3, 103.2 (C1'), 83.1 (C4), 79.2 (C4'), 75.3 (C3'), 74.9 (C3), 74.1 (C2'), 74.0 (CH$_2$O linker), 73.7 (C5'), 72.7 (C2), 62.3 (C6'), 61.6 (C6), 58.0 (epoxide), 55.5 (epoxide), 50.6 (C$_q$), 50.5 (C$_q$), 45.2 (C5), 44.8 (CH$_2$N$^+$), 40.4 (CH$_2$NH), 36.7 (CH$_2$C=O), 31.5 (CH$_3$N), 31.4, 30.5, 30.4, 30.3, 28.2 (5CH$_2$ linker), 28.0 (2CH$_3$), 27.9 (CH$_2$ linker), 27.8 (2CH$_3$), 27.4, 27.2, 26.6 (3CH$_2$ linker). HRMS (ESI) m/z: [M]$^+$ calc for $C_{53}H_{76}N_3O_{11}$ 930.5474, found 930.5475.

Compound 1d



Compound **1b** (5.7 mg, 12 μmol) was dissolved in dry DMF (0.5 mL), then DIPEA (4.3 μL, 24 μmol) and biotin-OSu (4.6 mg, 13 μmol) were added and the mixture was stirred at rt for 22 h. Full conversion was observed by LC-MS and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain the title compound (1.4 mg, 2.0 μmol, 16%) as a white solid. ^{1}H NMR (850 MHz, MeOD) δ 5.00 (d, J = 3.9 Hz, 1H, H1'), 4.51 – 4.48 (m, 1H, H^b), 4.31 (dd, J = 7.9, 4.5 Hz, 1H, H^a), 3.90 – 3.85 (m, 2H, H6a and CH*H*O linker), 3.82 – 3.75 (m, 3H, H2, H6b and H6'a), 3.72 – 3.65 (m, 3H, H3', H5' and H6'b), 3.61 – 3.55 (m, 2H, H3 and CH*H*O linker), 3.43 (dd, J = 9.7, 3.9 Hz, 1H, H2'), 3.33 (d, J = 9.6 Hz, 1H, H4), 3.27 (dd, J = 4.1, 1.8 Hz, 1H, epoxide), 3.23 – 3.19 (m, 2H, epoxide and CHS), 3.18 – 3.13 (m, 3H, H4' and CH_2NH), 2.93 (dd, J = 12.8, 5.0 Hz, 1H, CH*H*S), 2.71 (d, J = 12.7 Hz, 1H, CH*H*S), 2.23 – 2.14 (m, 2H, CH_2C=O), 2.05 (ddd, J = 9.2, 5.8, 3.0 Hz, 1H, H5), 1.78 – 1.53 (m, 6H, 6CH*H* linker), 1.50 (t, J = 7.0 Hz, 2H, 2CH*H* linker), 1.45 (q, J = 7.6 Hz, 2H, 2CH*H* linker), 1.41 – 1.30 (m, 8H, 8CH*H* linker). ^{13}C NMR (214 MHz, MeOD) δ 176.0, 166.1, 103.2 (C1'), 83.0 (C4), 79.3 (C4'), 75.3 (C3'), 74.9 (C3), 74.1 (C2'), 74.0 (CH_2O linker), 73.7 (C5'), 72.7 (C2), 63.4 (CH^a), 62.3 (C6'), 61.7 (C6), 61.6 (CH^b), 57.9 (epoxide), 57.0 (CHS), 55.5 (epoxide), 45.2 (C5), 41.1 (CH_2S), 40.4 (CH_2NH), 36.9 (CH_2C=O), 31.4, 30.5, 30.4, 30.3, 29.8, 29.5, 27.9, 27.1, 27.0 (9CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{31}H_{54}N_3O_{12}S$ 692.3422, found 692.3412.

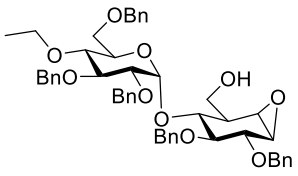
Compound 12a



Starting from compound **9a** (322 mg, 0.324 mmol) and following **General procedure F**, the product was purified by flash column chromatography (pentane/EtOAc 4:1→2:1) to obtain compound **12a** (0.26 g, 0.29 mmol, 91%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.47 – 6.91 (m, 30H, CH Ar), 5.66 (d, J = 3.9 Hz, 1H, H1'), 4.87 (t, J = 11.5 Hz, 2H, CH*H* Bn), 4.75 (ddd, J = 26.1, 19.7, 11.7 Hz, 4H, CH*H* Bn), 4.61 – 4.55 (m, 2H, CH*H* Bn), 4.55 – 4.38 (m, 4H, CH*H* Bn), 4.03 (dd, J = 11.3, 5.0 Hz, 1H, H6a), 4.01 – 3.85 (m, 5H, H6b, H3, H3' and H5'), 3.73 – 3.63 (m, 2H, H6'a), 3.58 (dd, J = 10.3, 5.0 Hz, 1H, H6'b), 3.50 – 3.41 (m, 2H, H2'), 3.39 (dd, J = 4.1, 1.7 Hz, 1H, epoxide), 3.16 (d, J = 3.8 Hz, 1H, epoxide), 2.28 – 2.20 (m, 1H, H5). ^{13}C NMR (126 MHz, $CDCl_3$) δ 139.2, 138.8, 138.2, 138.2, 137.8, 137.6 (6C_q Ar), 128.6, 128.6, 128.5, 128.5, 128.3, 128.3, 128.3, 128.2, 128.1, 128.1, 128.0, 127.9, 127.7, 127.7, 127.5, 127.1, 126.5, 97.4 (C1'), 84.8, 81.9, 80.2 (C3'), 79.3 (C2'), 78.1, 75.5 (CH_2 Bn), 75.1 (CH_2 Bn), 73.8 (CH_2 Bn), 73.7 (CH_2 Bn), 73.2 (CH_2 Bn), 72.9 (CH_2 Bn), 71.5 (C5'), 70.3, 68.9 (C6'), 61.9

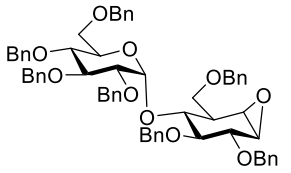
(C6), 56.5 (epoxide), 52.5 (epoxide), 43.3 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{55}H_{58}O_{10}Na$ 901.3922, found 901.3921.

Compound 12b



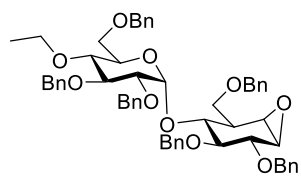
Starting from compound **9b** (338 mg, 0.36 mmol) and following **General procedure F**, the product was purified by flash column chromatography (pentane/EtOAc 3:1→2:1) to obtain compound **12b** (261 mg, 0.32 mmol, 88%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 – 6.99 (m, 25H, CH Ar), 5.66 (d, *J* = 3.9 Hz, 1H, H1'), 4.95 – 4.35 (m, 10H, CH*H* Bn), 4.08 – 3.91 (m, 3H), 3.91 – 3.64 (m, 6H), 3.59 (dd, *J* = 10.3, 4.9 Hz, 1H, H6'b), 3.46 – 3.37 (m, 3H, H2', CH*H* Et and epoxide), 3.26 (dd, *J* = 10.1, 8.8 Hz, 1H), 3.17 (d, *J* = 3.8 Hz, 1H, epoxide), 2.31 – 2.20 (m, 1H, H5), 1.04 (t, *J* = 7.0 Hz, 3H, CH_3 Et). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.2, 138.8, 138.2, 137.8, 137.5 (5C_q Ar), 128.6, 128.5, 128.4, 128.3, 128.2, 128.2, 128.1, 128.1, 127.9, 127.9, 127.7, 127.6, 127.5, 127.1, 126.5 (25CH Ar), 97.4 (C1'), 84.9, 81.8, 80.1, 79.1, 78.3, 75.5, 73.8, 73.7, 73.2, 72.9, 71.6, 70.0, 68.8 (C6'), 68.5 (CH_2 Et), 62.0 (C6), 56.5 (epoxide), 52.5 (epoxide), 43.3 (C5), 15.8 (CH_3 Et). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{50}H_{56}O_{10}Na$ 839.3766, found 839.3771.

Compound 13a



Starting from compound **12a** (0.26 g, 0.30 mmol) and following **General procedure G**, the product was purified by flash column chromatography (pentane/EtOAc 11:1→7:1) to obtain compound **13a** (234 mg, 0.24 mmol, 81%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 – 6.96 (m, 35H, CH Ar), 5.71 (d, *J* = 3.7 Hz, 1H, H1'), 4.88 (dd, *J* = 13.8, 11.4 Hz, 2H, CH*H* Bn), 4.82 – 4.66 (m, 4H, CH*H* Bn), 4.61 – 4.31 (m, 8H, CH*H* Bn), 3.94 – 3.79 (m, 4H), 3.73 – 3.58 (m, 4H), 3.55 (dd, *J* = 10.6, 2.8 Hz, 1H, H6'a), 3.51 – 3.40 (m, 3H, H6'b, H2' and epoxide), 3.19 (d, *J* = 3.8 Hz, 1H, epoxide), 2.43 (dddd, *J* = 9.5, 7.7, 3.8, 1.6 Hz, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.0, 138.87, 138.6, 138.3, 138.1, 138.0, 137.5 (7C_q Ar), 128.6, 128.5, 128.4, 128.4, 128.3, 128.3, 128.2, 128.1, 128.1, 127.9, 127.9, 127.8, 127.7, 127.7, 127.6, 127.5, 127.5, 127.1, 126.5, 96.9 (C1'), 85.2, 82.0, 80.2 (C3'), 79.3 (C2'), 77.7, 75.6, 75.1, 73.8, 73.6, 73.1, 73.0, 72.8, 71.1, 69.5, 69.1 (C6), 68.2 (C6'), 55.7 (epoxide), 53.2 (epoxide), 41.7 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{64}O_{10}Na$ 991.4392, found 991.4394.

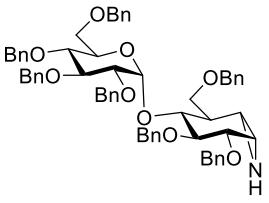
Compound 13b



Starting from compound **12b** (255 mg, 0.312 mmol) and following **General procedure G**, the product was purified by flash column chromatography (pentane/EtOAc 11:1→7:1) to obtain compound **13b** (213 mg, 0.235 mmol, 75%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 – 6.98 (m, 30H, CH Ar), 5.67 (d, *J* = 3.7 Hz, 1H, H1'), 4.92 – 4.66 (m, 5H, CH*H* Bn), 4.66 – 4.26 (m, 7H, CH*H* Bn), 3.92 – 3.58 (m, 8H), 3.54 – 3.36 (m, 6H), 3.19 (d, *J* = 3.8 Hz, 1H, epoxide), 2.44 (dddd,

J = 9.6, 7.9, 3.8, 1.6 Hz, 1H, H5), 1.06 (t, *J* = 7.0 Hz, 3H, CH_3 Et). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 139.1, 139.0, 138.3, 138.2, 138.0, 137.6 ($7C_q$ Ar), 128.6, 128.5, 128.4, 128.4, 128.3, 128.2, 128.2, 128.1, 128.0, 127.9, 127.7, 127.7, 127.7, 127.6, 127.5, 127.5, 127.1, 126.5, 97.0 (C1'), 85.1, 81.9, 80.2, 79.2, 77.8, 75.5, 73.7, 73.6, 73.2, 73.1, 72.8, 71.3, 69.8, 69.2 (C6), 68.4 (CH_2 Et), 68.2 (C6'), 55.6 (epoxide), 53.3 (epoxide), 41.7 (C5), 16.0 (CH_3 Et). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{57}H_{62}O_{10}Na$ 929.4235, found 929.4234.

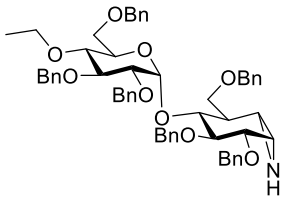
Compound 14a



Compound **13a** (234 mg, 0.241 mmol) was dissolved in dry DMF (2.4 mL) and purged with N_2. Then NaN_3 (158 mg, 2.41 mmol) and $LiClO_4$ (514 mg, 4.82 mmol) were added and the reaction was stirred at 100 ^{0}C for 18 h under inert atmosphere. The reaction mixture was diluted with H_2O (50 mL) and extracted with EtOAc (2 x 40 mL), the combined organic layers were washed with water (2 x 50 mL) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a mixture of azido-alcohols (0.21 g). The crude (0.21 mmol) was dissolved in dry MeCN (2 mL) and polymer-bound PPh_3 (1.6 mmol/g loading, 0.26 g, 0.42 mmol) was added to the solution. The reaction was stirred at 60 ^{0}C for 16 h under inert atmosphere. Then the beads were removed by filtration, the organic fitrate was concentrated *in vacuo* and purified by silica gel column chromatography (pentane/acetone 7:1→4:1) to obtain **14a** (96 mg, 99 µmol, 48%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.46 – 6.97 (m, 35H, CH Ar), 5.48 (s, 1H, H1'), 4.94 – 4.73 (m, 5H, CH*H* Bn), 4.69 (s, 2H, CH*H* Bn), 4.60 – 4.31 (m, 7H, CH*H* Bn), 3.97 (t, *J* = 9.3 Hz, 1H, H3'), 3.93 – 3.72 (m, 3H, H3, H2 and H6a), 3.61 (dddd, *J* = 14.3, 10.2, 6.3, 3.3 Hz, 5H, H6b, H4, H4', H5' and H6'a), 3.49 – 3.41 (m, 2H, H6'b and H2'), 2.49 (dd, *J* = 6.3, 2.8 Hz, 1H, aziridine), 2.37 (d, *J* = 7.2 Hz, 2H, H5), 2.30 (d, *J* = 6.2 Hz, 1H, aziridine). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 138.9, 138.6, 138.4, 138.3, 138.0 ($7C_q$ Ar), 128.8, 128.5, 128.5, 128.5, 128.4, 128.4, 128.3, 128.25, 128.2, 128.2, 128.1, 128.0, 128.0, 127.9, 127.8, 127.8, 127.7, 127.7, 127.6, 127.6, 127.5, 127.4, 127.1, 127.0, 126.8 (35CH Ar), 82.0, 82.0, 79.5, 79.5, 77.8, 77.8, 75.7 (CH_2 Bn), 75.1, 73.7, 73.6, 73.0, 72.8 ($5CH_2$ Bn), 72.2 (CH_2 Bn, assigned by HSQC), 70.9, 70.7 (C6), 68.4 (C6'), 43.4 (C5, assigned by HSQC), 33.8 (CH aziridine, assigned by HSQC), 33.1 (CH aziridine, assigned by HSQC). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{65}NO_9Na$ 990.4552, found 990.4548.

Compound 14b

compound **13b** (0.21 g, 0.23 mmol) was dissolved in dry DMF (2.4 mL) and purged with N_2. Then NaN_3 (0.15 g, 2.3 mmol) and $LiClO_4$ (0.49 g, 4.6 mmol) were added and the reaction was stirred at 100 ^{0}C for 18 h under inert atmosphere. After cooling to rt, the reaction mixture was diluted with H_2O (50 mL) and extracted with EtOAc (2 x 40 mL). The combined organic layers were washed with water (2 x 50 mL) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a mixture of azido-

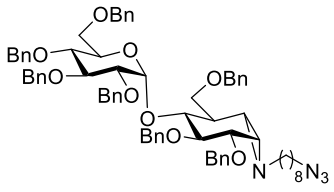


alcohols (204 mg). Starting from the azido-alcohol mixture (0.21 mmol), the reaction was carried out following the same procedure described for **14a**. The product was purified by flash column chromatography (pentane/EtOAc 7:1→1:1) to obtain **14b** (81 mg, 89 µmol, 42%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.38 – 7.02 (m, 30H, CH Ar), 5.48 – 5.36 (m, 1H, H1'), 4.90 – 4.35 (m, 12H, CH*H* Bn), 3.94 – 3.68 (m, 5H), 3.67 – 3.54 (m, 3H), 3.50 – 3.35 (m, 5H), 2.48 (dd, *J* = 6.3, 2.6 Hz, 1H, aziridine), 2.38 (q, *J* = 6.0, 5.5 Hz, 1H, H5), 2.30 (dd, *J* = 6.1, 1.1 Hz, 1H, aziridine), 1.05 (t, *J* = 7.0 Hz, 3H, CH_3 Et). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.4, 138.9, 138.7, 138.4, 138.4, 138.0 ($6C_q$ Ar), 128.5, 128.4, 128.2, 128.2, 128.1, 128.1, 128.0, 128.0, 127.7, 127.7, 127.6, 127.6, 127.6, 127.4, 127.0, 126.8 (30CH Ar), 97.4 (C1'), 81.8, 81.8, 79.4, 79.4, 77.8, 77.8, 75.6 (CH_2 Bn), 74.1, 73.6, 73.0, 72.9, 72.1 ($5CH_2$ Bn), 71.1, 70.8 (C6), 68.4 (CH_2 Et), 68.4 (C6'), 42.8 (C5, assigned by HSQC), 33.0 (C1 and C7, overlap), 15.9 (CH_3 Et). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{57}H_{63}NO_9Na$ 928.4395, found 928.4391.

Synthesis of 1-azido-8-trifluoromethylsulfonyloctane

To a dry DCM (0.1 M, 5.8 mL) was added 8-azidooctan-1-ol (100 mg, 0.58 mmol) and pyridine (57 µL, 0.70 mmol) and the mixture was cooled to -20 ^{0}C. Triflic anhydride (118 µL, 0.70 mmol) was added and the mixture was stirred for 15 min. Then the mixture was diluted with DCM (50 mL), washed with cold water (3 x 30 mL). The organic layer was dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo* at rt. The crude was used directly for the alkylation of the aziridine.

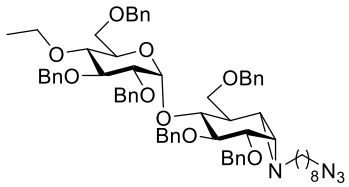
Compound 15a



Aziridine **14a** (70 mg, 72 µmol) was dissolved in dry $CHCl_3$ (0.8 mL) and cooled 0 ^{0}C. Then DIPEA (15 µL, 87 µmol) and freshly-made 8-azidooctyl trifluoromethanesulfonate (1.0 M in $CHCl_3$, 216 µL) were added and the mixture was stirred at rt for 2 h until TLC-analysis showed complete conversion of **14a**. Then another portion of DIPEA (38 µL, 216 µmol) and MeOH (220 µL) were added and the mixture was stirred at rt overnight. The reaction mixture was diluted with EtOAc (60 mL), washed with sat. aq. $NaHCO_3$ (30 mL), water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/acetone 20:1→18:1) to obtain **16a** (73 mg, 65 µmol, 90%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 6.96 (m, 35H, CH Ar), 5.56 (d, *J* = 3.6 Hz, 1H, H1'), 4.97 – 4.85 (m, 3H, CH*H* Bn), 4.78 (t, *J* = 10.8 Hz, 2H, CH*H* Bn), 4.74 – 4.64 (m, 2H, CH*H* Bn), 4.59 – 4.28 (m, 7H, CH*H* Bn), 3.98 (t, *J* = 9.3 Hz, 1H, H3'), 3.88 – 3.72 (m, 3H, H3, H2 and H6a), 3.67 – 3.52 (m, 5H, H6b, H4, H4', H5' and H6'a), 3.44 (dt, *J* = 10.4, 2.8 Hz, 2H, H6'b and H2'), 3.23 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.31 (td, *J* = 7.0, 6.6, 3.3 Hz, 2H, H5 and CH*H*N), 2.07 (dt, *J* = 10.5, 6.9 Hz, 1H, CH*H*N), 1.80 (dd, *J* = 6.5, 3.0 Hz, 1H, H1), 1.56 (dq, *J* = 14.9, 7.6, 6.9 Hz, 5H, H7 and 4CH*H* linker), 1.39 – 1.20 (m,

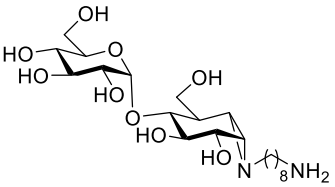
8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.8, 139.0, 138.9, 138.6, 138.5, 138.3, 138.0 (7C_q Ar), 128.4, 128.4, 128.3, 128.2, 128.2, 128.1, 128.0, 128.0, 127.9, 127.8, 127.7, 127.6, 127.5, 127.4, 127.3, 126.8, 126.6 (35CH Ar), 97.7 (C1'), 83.1 (C3), 82.0 (C3'), 80.3 (C2), 79.5 (C2'), 77.9 (C4'), 75.9 (C4), 75.6, 75.1, 74.5, 73.6, 72.9, 72.9, 72.2 (7CH_2 Bn), 70.9 (C5'), 70.8 (C6), 68.3 (C6'), 61.2 (CH_2N), 51.6 (CH_2N_3), 43.4 (C5), 42.1 (C1), 41.6 (C7), 29.9, 29.6, 29.2, 28.9, 27.4, 26.8 (6CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{70}H_{81}N_4O_9$ 1121.5998, found 1121.5994.

Compound 15b



Starting from aziridine **14b** (63 mg, 70 µmol), the reaction was carried out following the same procedure described for **15a**. The product was purified by flash column chromatography (pentane/EtOAc 13:1→11:1) to obtain **15b** (65 mg, 61 µmol, 88%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.49 – 6.95 (m, 30H, CH Ar), 5.51 (d, J = 3.6 Hz, 1H, H1'), 4.89 (d, J = 1.4 Hz, 2H, CH*H* Bn), 4.84 (d, J = 10.8 Hz, 1H, CH*H* Bn), 4.79 – 4.62 (m, 3H, CH*H* Bn), 4.57 (d, J = 12.1 Hz, 1H, CH*H* Bn), 4.52 – 4.43 (m, 3H, CH*H* Bn), 4.43 – 4.32 (m, 2H, CH*H* Bn), 3.92 – 3.68 (m, 5H), 3.67 – 3.58 (m, 3H), 3.53 (dd, J = 10.5, 3.1 Hz, 1H), 3.48 – 3.34 (m, 4H), 3.23 (t, J = 7.0 Hz, 2H, CH_2N_3), 2.36 – 2.25 (m, 2H, H5 and CH*H*N), 2.09 (dt, J = 11.5, 7.4 Hz, 1H, CH*H*N), 1.80 (dd, J = 6.5, 3.0 Hz, 1H, H1), 1.62 – 1.48 (m, 5H, H7 and 4CH*H* linker), 1.38 – 1.24 (m, 8H, 8CH*H* linker), 1.05 (t, J = 7.0 Hz, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.8, 139.0, 139.0, 138.5, 138.4, 138.0 (6C_q Ar), 128.4, 128.4, 128.3, 128.2, 128.1, 128.0, 127.9, 127.7, 127.6, 127.5, 127.5, 127.4, 127.3, 126.7, 126.6 (30CH Ar), 97.8 (C1'), 83.0, 81.8, 80.3, 79.4, 77.9, 76.4, 75.5 (CH_2 Bn), 74.4, 73.6, 72.9, 72.9, 72.2 (5CH_2 Bn), 71.0, 71.0 (C6), 68.4 (CH_2 Et), 68.3 (C6'), 61.2 (CH_2N), 51.6 (CH_2N_3), 43.3 (C5), 42.1 (C1), 41.6 (C7), 29.9, 29.6, 29.2, 28.9, 27.4, 26.8 (6CH_2 linker), 15.9 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{65}H_{79}N_4O_9$ 1059.5842, found 1059.5838.

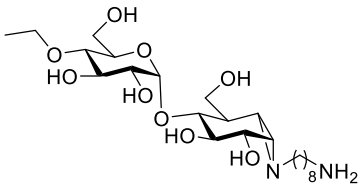
Compound 2a



Ammonium (3.0 mL) was condensed in a dry flask at -60 ^{0}C, and sodium (34 mg, 1.4 mmol) was added. The resulting deep-blue solution was stirred for 15 min to dissolve all sodium. Aziridine **15a** (23 mg, 20 µmol) and ^{t}BuOH (20 µL, 200 µmol) were taken up in dry THF (1.0 mL) and slowly added to the reaction mixture. After stirring for 1 h, the reaction was carefully quenched with H_2O. The mixture was slowly warmed to rt and evaporated. The crude was dissolved in H_2O and eluted over a column packed with Amberlite CG-50 (MH_4^+) with 0.5 M NH_4OH as eluent, concentrated *in vacuo*, affording the title compound (7.5 mg, 16 µmol, 79%) as a light-yellow powder after lyophilization. ^{1}H NMR (400 MHz, D_2O) δ 5.11 (d, J = 4.0 Hz, 1H, H1'), 3.94 – 3.64 (m, 7H), 3.60 – 3.47 (m, 2H), 3.35 (dt, J = 26.7, 9.7 Hz, 2H), 2.94 (t, J = 7.5 Hz, 2H, C*H*$_2$NH$_2$), 2.27 (t, J = 7.5 Hz, 2H, CH_2N), 2.11 – 1.93 (m, 2H, H5 and H1), 1.82 (d, J = 6.6 Hz, 1H, H7), 1.66 – 1.57 (m, 2H,

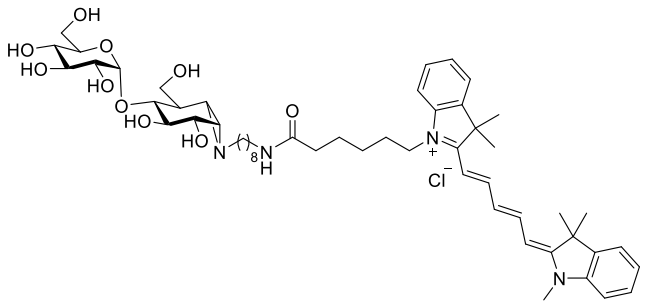
2CH*H* linker), 1.51 (s, 2H, 2CH*H* linker), 1.36 – 1.22 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, D_2O) δ 101.0 (C1'), 81.3, 74.9, 73.3, 72.9, 72.1, 71.3, 69.7, 61.9 (C6'), 60.8 (C6), 60.4 (CH_2N), 44.5 (C1), 43.6 (C5), 41.1 (C7), 39.9 (CH_2NH_2), 29.0, 28.8, 28.4, 27.2, 26.7, 25.9 (6CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{21}H_{41}N_2O_9$ 465.2807, found 465.2806.

Compound 16



Ammonium (8.0 mL) was condensed in a dry flask at -60 ^{0}C, and sodium (94 mg, 4.1 mmol) was added. The resulting deep-blue solution was stirred for 15 min to dissolve all sodium. Aziridine **15b** (62 mg, 59 µmol) and ^{t}BuOH (56 µL, 0.59 mmol) were taken up in dry THF (1.5 mL) and slowly added to the reaction mixture. After stirring for 1 h, the reaction was quenched with H_2O. The mixture was slowly warmed to rt and evaporated. The crude was dissolved in H_2O and eluted over a column packed with Amberlite CG-50 (MH_4^+) with 0.5 M NH_4OH as eluent, concentrated *in vacuo*, affording the title compound (27 mg, 55 µmol, 94%) as a white powder after lyophilization. ^{1}H NMR (400 MHz, D_2O) δ 5.09 (d, J = 4.0 Hz, 1H, H1'), 3.89 – 3.64 (m, 9H), 3.61 – 3.48 (m, 2H), 3.34 – 3.21 (m, 2H), 2.93 (dt, J = 15.2, 7.2 Hz, 2H, C*H*$_2$NH$_2$), 2.26 (t, J = 7.3 Hz, 2H, CH_2N), 2.06 – 1.94 (m, 2H, H1 and H5), 1.81 (d, J = 6.6 Hz, 1H, H7), 1.66 – 1.41 (m, 4H, 4CH*H* linker), 1.29 (s, 8H, 8CH*H* linker), 1.17 (t, J = 7.0 Hz, 3H, CH_3). ^{13}C NMR (101 MHz, D_2O) δ 100.5 (C1'), 80.9, 77.5, 74.6, 72.8, 71.7, 71.6, 70.9, 68.9 (CH_2 Et), 61.5 (C6), 60.1 (C6'), 60.0 (CH_2N), 44.2 (C1), 43.1 (C5), 40.7 (C7), 39.6 (CH_2NH_2), 28.6, 28.4, 28.1, 27.2, 26.4, 25.5 (6CH_2 linker), 14.7 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{23}H_{45}N_2O_9$ 493.3120, found 493.3120.

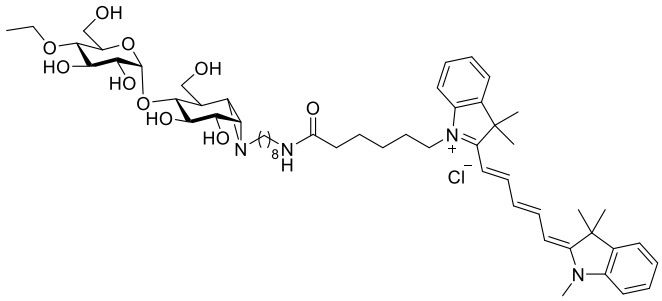
Compound 2b



Compound **2a** (7.0 mg, 15 µmol) was dissolved in dry DMF (0.5 mL), then DIPEA (5.7 µL, 33 µmol) and Cy5-OSu (10 mg, 16 µmol) were added and the mixture was stirred at rt for 16 h. Full conversion was observed by LC-MS and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain the title compound (1.4 mg, 1.5 µmol, 10%) as a blue powder. ^{1}H NMR (600 MHz, MeOD) δ 8.29 – 8.20 (m, 2H), 7.50 (dt, J = 7.4, 1.4 Hz, 2H), 7.42 (tdd, J = 7.6, 4.9, 1.2 Hz, 2H), 7.33 – 7.23 (m, 4H), 6.63 (t, J = 12.5 Hz, 1H), 6.28 (dd, J = 13.7, 7.4 Hz, 2H), 4.98 (d, J = 3.9 Hz, 1H, H1'), 4.11 (t, J = 7.4 Hz, 2H, CH_2N^+), 3.88 (dd, J = 11.0, 3.1 Hz, 1H, H6a), 3.85 – 3.78 (m, 1H, H6a'), 3.76 – 3.66 (m, 4H, H2, H5', H6b and H6'b), 3.62 (d, J = 15.3 Hz, 4H, CH_3N and H3'), 3.55 (dd, J = 9.9,

8.8 Hz, 1H, H3), 3.45 – 3.40 (m, 1H, H2'), 3.27 (d, *J* = 9.2 Hz, 1H, H4'), 3.21 – 3.15 (m, 1H, H4), 3.12 (t, *J* = 7.2 Hz, 2H, C*H*$_2$NH), 2.31 (dt, *J* = 11.6, 7.1 Hz, 1H, CH*H*N aziridine), 2.20 (t, *J* = 7.3 Hz, 2H, CH$_2$C=O), 2.13 (t, *J* = 5.8 Hz, 1H, CH*H*N aziridine), 1.95 – 1.88 (m, 1H, H5), 1.87 – 1.79 (m, 3H, H1 and 2CH*H* linker), 1.73 (s, 15H, H7, 2CH*H* linker and 4CH$_3$), 1.56 (q, *J* = 7.5 Hz, 2H, 2CH*H* linker), 1.50 – 1.42 (m, 4H, 4CH*H* linker), 1.36 – 1.27 (m, 8H, 8CH*H* linker). ^{13}C NMR (214 MHz, MeOD) *δ* 175.7, 175.4, 174.7, 155.5, 155.5, 144.3, 143.6, 142.7, 142.5, 129.8, 129.8, 126.6, 126.3, 126.3, 123.4, 123.3, 112.1, 111.9, 104.4, 104.3, 103.2 (C1'), 84.1 (C4), 75.9 (C3), 75.0 (C3'), 74.4, 74.0 (C2'), 72.7, 71.4 (C4'), 62.9 (C6), 62.7 (C6'), 62.2 (CH$_2$N aziridine), 50.6 (C$_q$), 50.5 (C$_q$), 45.9 (C1), 45.8 (C5), 44.8 (CH$_2$N$^+$), 42.1 (C7), 40.4 (CH$_2$NH), 36.7 (*C*H$_2$C=O), 31.5 (CH$_3$N), 30.7, 30.5, 30.4, 30.3, 28.4, 28.2, 28.0 (7CH$_2$ linker), 27.9 (2CH$_3$), 27.8 (2CH$_3$), 27.4, 26.6 (2CH$_2$ linker). HRMS (ESI) m/z: [M]$^+$ calc for C$_{53}$H$_{77}$N$_4$O$_{10}$ 929.5634, found 929.5627.

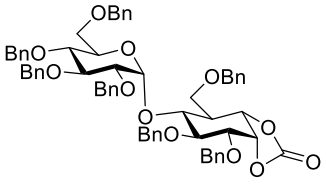
Compound 2c



Compound **16** (7.0 mg, 14 µmol) was dissolved in dry DMF (0.5 mL), then DIPEA (5.5 µL, 31 µmol) and Cy5-OSu (10 mg, 17 µmol) were added and the mixture was stirred at rt overnight. Full conversion was observed by LC-MS and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH$_4$HCO$_3$ in H$_2$O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain the title compound (2.4 mg, 2.4 µmol, 17%) as a blue powder. ^{1}H NMR (850 MHz, MeOD) *δ* 8.30 – 8.21 (m, 2H), 7.50 (dt, *J* = 7.3, 1.5 Hz, 2H), 7.44 – 7.39 (m, 2H), 7.32 – 7.25 (m, 4H), 6.63 (t, *J* = 12.4 Hz, 1H), 6.28 (dd, *J* = 13.7, 11.2 Hz, 2H), 4.97 (d, *J* = 3.9 Hz, 1H, H1'), 4.11 (t, *J* = 7.5 Hz, 2H, CH$_2$N$^+$), 3.92 – 3.84 (m, 2H, H6a and CH*H*O Et), 3.77 (dd, *J* = 11.7, 2.1 Hz, 1H, H6'a), 3.74 – 3.61 (m, 9H, H2, H3', H5', H6b, H6'b, CH*H*O Et and CH$_3$N), 3.56 – 3.52 (m, 1H, H3), 3.47 – 3.40 (m, 1H, H2'), 3.20 – 3.09 (m, 4H, H4, H4' and C*H*$_2$NH), 2.31 (ddt, *J* = 14.7, 7.7, 3.6 Hz, 1H, CH*H*N aziridine), 2.20 (t, *J* = 7.2 Hz, 2H, CH$_2$C=O), 2.15 (ddd, *J* = 11.7, 8.7, 7.2 Hz, 1H, CH*H*N aziridine), 1.90 (ddd, *J* = 9.9, 6.7, 3.1 Hz, 1H, H5), 1.87 – 1.75 (m, 3H, H1 and 2CH*H* linker), 1.73 (s, 15H, H7, 2CH*H* linker and 4CH$_3$), 1.64 – 1.52 (m, 2H, 2CH*H* linker), 1.46 (dq, *J* = 14.2, 6.9, 6.4 Hz, 4H, 4CH*H* linker), 1.39 – 1.27 (m, 8H, 8CH*H* linker), 1.20 – 1.15 (m, 3H, CH$_3$ Et). ^{13}C NMR (214 MHz, MeOD) *δ* 175.7, 175.4, 174.7, 155.5, 155.5, 144.3, 143.6, 142.7, 142.5, 129.78, 129.8, 126.7, 126.3, 126.3, 123.4, 123.3, 112.1, 111.9, 104.4, 104.3, 103.1 (C1'), 84.1, 79.2, 75.8 (C3), 75.2, 74.1 (C2'), 73.6, 72.7, 69.3 (CH$_2$ Et), 62.9 (C6), 62.1 (C6'), 62.1 (CH$_2$N aziridine), 50.6 (C$_q$), 50.5 (C$_q$), 45.9 (C1), 45.8 (C5), 44.8 (CH$_2$N$^+$), 42.2 (C7), 40.4 (CH$_2$NH), 36.7 (*C*H$_2$C=O), 31.51 (CH$_3$N), 30.7,

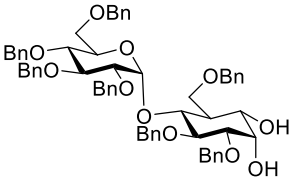
30.4, 30.4, 30.3, 28.4, 28.2, 28.0 ($7CH_2$ linker), 28.0 ($2CH_3$), 27.8 ($2CH_3$), 27.4, 26.6 ($2CH_2$ linker), 16.0 (CH_3 Et). HRMS (ESI) m/z: $[M]^+$ calc for $C_{55}H_{81}N_4O_{10}$ 957.5947, found 957.5945.

Compound 18



Starting from compound **17** (98 mg, 0.20 mmol), using donor **4a** (214 mg, 0.30 mmol) and following **Standard procedure A**. The reaction mixture was purified by size exclusion (DCM/MeOH = 1/1) to give the product as a mixture of two isomers (130 mg, α/β = 13/1, determined by ^{1}H NMR). Then the mixture was further purified by flash column chromatography (pentane/acetone 13:1→9:1) to obtain compound **18** (115 mg, 114 µmol, 57%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.03 (m, 35H, CH Ar), 5.14 (d, *J* = 3.5 Hz, 1H, H1'), 5.03 (t, *J* = 9.0 Hz, 1H, H7), 4.92 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.82 (dd, *J* = 10.9, 6.3 Hz, 2H, CH*H* Bn), 4.79 – 4.72 (m, 2H, 1CH*H* Bn and H1), 4.61 (s, 2H, CH*H* Bn), 4.55 (d, *J* = 11.9 Hz, 1H, CH*H* Bn), 4.44 (td, *J* = 21.8, 20.8, 12.0 Hz, 4H, CH*H* Bn), 4.25 (d, *J* = 12.2 Hz, 2H, CH*H* Bn), 4.17 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 3.99 – 3.83 (m, 5H, H2, H3', H3, H4 and H6a), 3.76 (dt, *J* = 10.1, 2.4 Hz, 1H, H5'), 3.73 – 3.66 (m, 1H, H4'), 3.62 (dd, *J* = 9.7, 2.4 Hz, 1H, H6b), 3.57 (dd, *J* = 9.8, 3.5 Hz, 1H, H2'), 3.42 (dd, *J* = 10.9, 2.8 Hz, 1H, H6'a), 3.29 (dd, *J* = 10.9, 1.9 Hz, 1H, H6'b), 2.80 (ddd, *J* = 11.9, 9.4, 2.3 Hz, 1H, H5). ^{13}C NMR (126 MHz, $CDCl_3$) δ 154.8 (C=O), 138.8, 138.6, 138.4, 138.1, 138.0, 137.2, 137.1 ($7C_q$ Ar), 128.7, 128.6, 128.5, 128.4, 128.4, 128.3, 128.2, 128.1, 128.1, 128.0, 128.0, 128.0, 127.9, 127.8, 127.7, 127.6, 127.6, 127.3 (35CH Ar), 95.1 (C1'), 82.3, 81.0, 79.8 (C2'), 77.7 (C4'), 75.7, 75.1 ($2CH_2$ Bn), 74.1, 74.2 (C1), 73.9 (C7), 73.8, 73.5, 73.3, 73.0 ($4CH_2$ Bn), 71.9, 71.6 (CH_2 Bn), 71.0 (C5'), 68.1 (C6'), 65.6 (C6), 41.3 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{63}H_{64}O_{12}Na$ 1035.4290, found 1035.4286.

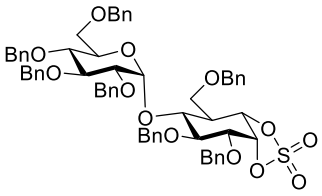
Compound 19



Compound **18** (90 mg, 89 µmol) was dissolved in DCM/MeOH (1/1, 2.0 mL), then 30 wt% NaOMe in MeOH solution (20 µL, 0.11 mmol) was added and the reaction mixture was stirred at rt for 2 h. After which the mixture was diluted with DCM (2.0 mL) and neutralized with washed Amberlite IR-120 H^+ resin until pH ≈ 7, filtered, washed with DCM (3 x 3 mL) and concentrated *in vacuo*. The product was purified by flash column chromatography (pentane/acetone 7:1→3:1) to obtain compound **19** (82 mg, 83 µmol, 93%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.04 (m, 35H, CH Ar), 5.81 (d, *J* = 3.7 Hz, 1H, H1'), 4.99 (d, *J* = 11.9 Hz, 1H, CH*H* Bn), 4.89 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.83 – 4.72 (m, 3H, CH*H* Bn), 4.65 – 4.56 (m, 3H, CH*H* Bn), 4.56 – 4.48 (m, 2H, CH*H* Bn), 4.47 – 4.36 (m, 3H, CH*H* Bn), 4.27 (d, *J* = 12.2 Hz, 1H, CH*H* Bn), 4.15 (t, *J* = 2.8 Hz, 1H), 4.05 (t, *J* = 9.0 Hz, 1H), 3.97 (dd, *J* = 9.9, 8.8 Hz, 1H), 3.91 – 3.68 (m, 5H), 3.65 (dd, *J* = 10.1, 8.8 Hz, 1H), 3.54 – 3.41 (m, 3H), 3.38 (dd, *J* = 10.7, 1.9 Hz, 1H), 2.49 (brs, 3H), 2.30 (tt, *J* = 10.6, 3.8 Hz, 1H). ^{13}C NMR

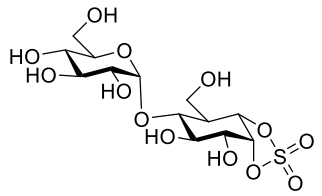
(101 MHz, $CDCl_3$) δ 139.0, 138.9, 138.6, 138.2, 138.1, 138.0, 137.8 ($7C_q$ Ar), 128.6, 128.4, 128.4, 128.3, 128.2, 128.1, 128.0, 127.9, 127.8, 127.8, 127.7, 127.6, 127.6, 127.3, 127.1, 126.5, 96.4 (C1'), 82.8, 82.1, 80.5, 79.5, 77.8, 75.6 (CH_2 Bn), 75.0, 74.0, 73.6, 73.4, 72.9, 72.5 ($6CH_2$ Bn), 71.0, 69.8, 69.0, 68.3 (C6'), 67.6 (C6), 42.9 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{66}O_{11}Na$ 1009.4497, found 1009.4495.

Compound 20



Compound **19** (82 mg, 83 µmol) was dissolved in dry DCM (1.5 mL) and purged with N_2. After cooling to 0 ^{0}C, TEA (92 µL, 0.66 mmol) and thionyl chloride (24 µL, 0.33 mmol) were added successively. After stirring at 0 ^{0}C for 15 min, the reaction was diluted with DCM (50 mL), washed with sat. aq. $NaHCO_3$ (30 mL), H_2O (2 x 30 mL) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo* to give the crude sulfite as a pale-yellow oil. The crude was directly dissolved in a mixture of EtOAc and MeCN (1/1, 2.0 mL) and cooled to 0 ^{0}C. A solution of $RuCl_3.3H_2O$ (2 mg, 8.3 µmol) and $NaIO_4$ (35 mg, 0.16 mmol) in H_2O (0.4 mL) was added and the mixture was stirred vigorously at 0 ^{0}C for 2 h. The reaction was quenched by addition of sat. aq. $Na_2S_2O_3$ (1 mL) and the mixture was stirred vigorously for 15 min. Then the mixture was diluted with H_2O (30 mL) and extracted with EtOAc (2 x 40 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by flash column chromatography (pentane/acetone 13:1→11:1) affording compound **20** (50 mg, 48 µmol, 58%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 – 6.99 (m, 35H, CH Ar), 5.28 (d, *J* = 3.7 Hz, 1H, H1'), 5.22 (dd, *J* = 9.8, 6.5 Hz, 1H, H7), 4.99 (dd, *J* = 6.9, 3.4 Hz, 1H, H1), 4.90 (d, *J* = 10.4 Hz, 1H, CH*H* Bn), 4.81 (dd, *J* = 10.9, 4.0 Hz, 2H, CH*H* Bn), 4.76 – 4.58 (m, 3H, CH*H* Bn), 4.56 – 4.20 (m, 8H, CH*H* Bn), 3.99 – 3.85 (m, 5H, H2, H3', H3, H4 and H6a), 3.78 – 3.63 (m, 2H, H5' and H4'), 3.62 – 3.51 (m, 2H, H6b and H2'), 3.49 – 3.40 (m, 1H, H6'a), 3.31 (d, *J* = 10.8 Hz, 1H, H6'b), 2.87 (t, *J* = 10.1 Hz, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.7, 138.5, 138.2, 137.9, 137.8, 137.3, 137.2 ($7C_q$ Ar), 128.7, 128.6, 128.6, 128.5, 128.4, 128.4, 128.3, 128.1, 128.0, 128.0, 127.9, 127.8, 127.8, 127.7, 127.3 (35CH Ar), 95.8 (C1'), 82.1, 81.1, 79.7 (C2'), 79.4 (C1), 78.3 (C7), 77.7 (C4'), 75.7, 75.1 ($2CH_2$ Bn), 74.3, 73.7 (CH_2 Bn), 73.6, 73.5, 73.0, 72.4 ($4CH_2$ Bn), 71.2, 71.1 (C5'), 68.1 (C6'), 64.8 (C6), 41.8 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{62}H_{64}O_{13}SNa$ 1071.3960, found 1071.3965.

Compound 3a



Compound **20** (20 mg, 19 µmol) was dissolved in a mixture of MeOH/H_2O/dioxane (2/1/2, 2 mL) under Argon and $Pd(OH)_2$/C (20 wt%, 20 mg, 29 µmol) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 3 h under H_2 atmosphere, the mixture was filtered over a small celite pad and evaporate to afford the product **3a** in high purity as a white solid (7.1 mg, 17 µmol, 89%) after lyophilization. ^{1}H NMR (400

MHz, D_2O) δ 5.47 (dd, J = 4.6, 3.4 Hz, 1H, H1), 5.35 – 5.23 (m, 2H, H1' and H7), 4.02 – 3.87 (m, 3H, H2, H3 and H6a), 3.86 – 3.55 (m, 7H), 3.40 (t, J = 9.2 Hz, 1H), 2.43 – 2.32 (m, 1H, H5). ^{13}C NMR (101 MHz, D_2O) δ 100.3 (C1'), 84.4 (C1), 82.1 (C7), 76.2, 73.2, 72.9, 72.8, 71.7, 69.2, 68.0, 60.4 (C6'), 56.5 (C6), 43.7 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{22}O_{13}SNa$ 441.0673, found 441.0670.

2.6 References

[1] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., *Nucleic Acids Res.* **2014**, *42* (D1), D490-D495.

[2] Huang, N.; Stebbins, G. L.; Rodriguez, R. L., *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7526-7530.

[3] Brayer, G. D.; Luo, Y.; Withers, S. G., *Protein Sci.* **1995**, *4*, 1730-1742.

[4] Gupta, R.; Gigras, P.; Mohapatra, H.; Goswami, V. K.; Chauhan, B., *Process Biochem.* **2003**, *38* (11), 1599-1616.

[5] Zechel, D. L.; Withers, S. G., *Acc. Chem. Res.* **2000**, *33* (1), 11-18.

[6] Nakazawa, Y.; Wang, Y. J., *Carbohydr. Res.* **2003**, *338* (24), 2871-2882.

[7] Wang, S.; Copeland, L., *Crit. Rev. Food Sci. Nutr.* **2015**, *55* (8), 1081-1097.

[8] Kirk, O.; Borchert, T. V.; Fuglsang, C. C., *Curr. Opin. Biotech.* **2002**, *13* (4), 345-351.

[9] van der Maarel, M. J. E. C.; van der Veen, B.; Uitdehaag, J. C. M.; Leemhuis, H.; Dijkhuizen, L., *J. Biotech.* **2001**, *94*, 137-155.

[10] Souza, P. M. d.; Magalhães, P. d. O. e., *Braz. J. Microbiol.* **2010**, *41*, 850-861.

[11] Sundarram, A.; Murthy, T. P. K., *J. Appl. Environ. Microbiol.* **2014**, *2*, 166-175.

[12] Hmidet, N.; El-Hadj Ali, N.; Haddar, A.; Kanoun, S.; Alya, S.-K.; Nasri, M., *Biochem. Eng. J.* **2009**, *47* (1-3), 71-79.

[13] Rana, N.; Walia, A.; Gaur, A., *Natl. Acad. Sci. Lett.* **2013**, *36* (1), 9-17.

[14] Mooradian, A. D.; Thurman, J. E., *Drugs* **1999**, *57* (1), 19-29.

[15] Scott, L. J.; Spencer, C. M., *Drugs* **2000**, *59*, 521-549.

[16] Chiasson, J.-L.; Josse, R. G.; Gomis, R.; Hanefeld, M.; Karasik, A.; Laakso, M., *The Lancet* **2002**, *359* (9323), 2072-2077.

[17] Li, C.; Begum, A.; Numao, S.; Park, K. H.; Withers, S. G.; Brayer, G. D., *Biochemistry* **2005**, *44*, 3347-3357.

[18] Nahoum, V.; Roux, G.; Anton, V.; RougÉ, P.; Puigserver, A.; Bischoff, H.; Henrissat, B.; Payan, F., *Biochem. J.* **2000**, *346* (1), 210-208.

[19] Tarling, C. A.; Woods, K.; Zhang, R.; Brastianos, H. C.; Brayer, G. D.; Andersen, R. J.; Withers, S. G., *Chembiochem* **2008**, *9* (3), 433-438.

[20] Williams, L. K.; Zhang, X.; Caner, S.; Tysoe, C.; Nguyen, N. T.; Wicki, J.; Williams, D. E.; Coleman, J.; McNeill, J. H.; Yuen, V.; Andersen, R. J.; Withers, S. G.; Brayer, G. D., *Nat. Chem. Biol.* **2015**, *11* (9), 691-696.

[21] Braun, C.; Brayer, G. D.; Withers, S. G., *J. Biol. Chem.* **1995**, *270*, 26778-26781.

[22] Zhang, R.; McCarter, J. D.; Braun, C.; Yeung, W.; Brayer, G. D.; Withers, S. G., *J. Org. Chem.* **2007**, *73*, 3070-3077.

[23] Zhang, R.; Li, C.; Williams, L. K.; Rempel, B. P.; Brayer, G. D.; Withers, S. G., *Biochemistry* **2009**, *48* (45), 10752-10764.

[24] Caner, S.; Zhang, X.; Jiang, J.; Chen, H. M.; Nguyen, N. T.; Overkleeft, H.; Brayer, G. D.; Withers, S. G., *FEBS Lett.* **2016**, *590* (8), 1143-1151.

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

[26] Willems, L. I.; Overkleeft, H. S.; van Kasteren, S. I., *Bioconjugate Chem.* **2014**, *25* (7), 1181-1191.

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

[28] Jiang, J.; Kallemeijn, W. W.; Wright, D. W.; van den Nieuwendijk, A.; 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.; Overkleeft, H. S., *Chem. Sci.* **2015**, *6* (5), 2782-2789.

[29] Willems, L. I.; Beenakker, T. J.; Murray, B.; Scheij, S.; Kallemeijn, W. W.; Boot, R. G.; Verhoek, M.; Donker-Koopman, W. E.; Ferraz, M. J.; van Rijssel, E. R.; Florea, B. I.; Codée, J. D.; van der Marel, G. A.; Aerts, J. M. F. G.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2014**, *136* (33), 11622-11625.

[30] 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.; Boot, R. G.; Davies, G. J.; Overkleeft, H. S.; Aerts, J. M., *ACS Cent. Sci.* **2016**, *2* (5), 351-358.

[31] Wu, L.; Jiang, J.; Jin, Y.; Kallemeijn, W. W.; Kuo, C.-L.; Artola, M.; Dai, W.; van Elk, C.; van Eijk, M.; van der Marel, G. A.; Codée, J. D. C.; Florea, B. I.; Aerts, J. M. F. G.; Overkleeft, H. S.; Davies, G. J., *Nat. Chem. Biol.* **2017**, *13*, 867-873.

[32] McGregor, N. G. S.; Artola, M.; Nin-Hill, A.; Linzel, D.; Haon, M.; Reijngoud, J.; Ram, A.; Rosso, M. N.; van der Marel, G. A.; Codée, J. D. C.; van Wezel, G. P.; Berrin, J. G.; Rovira, C.; Overkleeft, H. S.; Davies, G. J., *J. Am. Chem. Soc.* **2020**, *142* (10), 4648-4662.

[33] Schröder, S. P.; de Boer, C.; McGregor, N. G. S.; Rowland, R. J.; Moroz, O.; Blagova, E.; Reijngoud, J.; Arentshorst, M.; Osborn, D.; Morant, M. D.; Abbate, E.; Stringer, M. A.; Krogh, K.; Raich, L.; Rovira, C.; Berrin, J. G.; van Wezel, G. P.; Ram, A. F. J.; Florea, B. I.; van der Marel, G. A.; Codée, J. D. C.; Wilson, K. S.; Wu, L.; Davies, G. J.; Overkleeft, H. S., *ACS Cent. Sci.* **2019**, *5* (6), 1067-1078.

[34] de Boer, C.; McGregor, N. G. S.; Peterse, E.; Schröder, S. P.; Florea, B. I.; Jiang, J.; Reijngoud, J.; Ram, A. F. J.; van Wezel, G. P.; van der Marel, G. A.; Codée, J. D. C.; Overkleeft, H. S.; Davies, G. J., *RSC Chemical Biology* **2020**, *1* (3), 148-155.

[35] Artola, M.; Wu, L.; Ferraz, M. J.; Kuo, C. L.; Raich, L.; Breen, I. Z.; Offen, W. A.; Codée, J. D. C.; van der Marel, G. A.; Rovira, C.; Aerts, J.; Davies, G. J.; Overkleeft, H. S., *ACS Cent. Sci.* **2017**, *3* (7), 784-793.

[36] Wang, L.; Overkleeft, H. S.; van der Marel, G. A.; Codée, J. D. C., *J. Am. Chem. Soc.* **2018**, *140* (13), 4632-4638.

[37] Sherman, H. C.; Kendall, E. C.; Clark, E. D., *J. Am. Chem. Soc.* **1910**, *32* (9), 1073-1086.

[38] Takamine, J., *J. Soc. Chem. Ind.* **1898**, *17* (2), 118-120.

[39] Kato, S.; Shimizu-Ibuka, A.; Mura, K.; Takeuchi, A.; Tokue, C.; Arai, S., *Bioscience, Biotechnology, and Biochemistry* **2014**, *71* (12), 3007-3013.

[40] Alonso-Gil, S.; Coines, J.; André, I.; Rovira, C., *Front. Chem.* **2019**, *7*, 1-10.

[41] Brzozowski, A. M.; Davies, G. J., *Biochemistry* **1997**, *36* (36), 10837-10845.

[42] Fisher, S. Z.; Govindasamy, L.; Tu, C.; Agbandje-McKenna, M.; Silverman, D. N.; Rajaniemi, H. J.; McKenna, R., *Sect. F: Struct. Biol. Crystallogr. Commun.* **2006**, *62* (2), 88-93.

[43] Nekiunaite, L.; Arntzen, M. O.; Svensson, B.; Vaaje-Kolstad, G.; Abou Hachem, M., *Biotechnol. Biofuels* **2016**, *9* (1), 187-202.

[44] Parashar, D.; Satyanarayana, T., *Front. Bioeng. Biotechnol.* **2018**, *6*, 1-14.

[45] Crabb, W.D., Mitchinson, C., *Trends Biotechnol.* **1997**, *15*, 349-352.

[46] Wu, L.; Armstrong, Z.; Schröder, S. P.; de Boer, C.; Artola, M.; Aerts, J. M.; Overkleeft, H. S.; Davies, G. J., *Curr. Opin. Chem. Biol.* **2019**, *53*, 25-36.

[47] Armstrong, Z.; Kuo, C. L.; Lahav, D.; Liu, B.; Johnson, R.; Beenakker, T. J. M.; de Boer, C.; Wong, C. S.; van Rijssel, E. R.; Debets, M. F.; Florea, B. I.; Hissink, C.; Boot, R. G.; Geurink, P. P.; Ovaa, H.; van der Stelt, M.; van der Marel, G. M.; Codée, J. D. C.; Aerts, J.; Wu, L.; Overkleeft, H. S.; Davies, G. J., *J. Am. Chem. Soc.* **2020,** *142* (30), 13021-13029.

[48] Kuharev, J.; Navarro, P.; Distler, U.; Jahn, O.; Tenzer, S., *Proteomics* **2015**, *15* (18), 3140-3151.

[49] Kato, N.; Suyama, S.; Shirokane, M.; Kato, M.; Kobayashi, T.; Tsukagoshi, N., *Appl. Environ. Microbiol.* **2002**, *68* (3), 1250-1256.

[50] Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C., *Anal. Biochem.* **1985**, *150* (1), 76-85.

[51] Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M., *Acta. Crystallogr., Sect. D: Struct. Biol.* **2018**, *74* (2), 68-84.

[52] Skubák, P.; Pannu, N. S., *Nat. Commun.* **2013,** *4* (1), 1-6.

[53] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *J. Appl. Crystallogr.* **2007,** *40* (4), 658-674.

[54] Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011,** *67* (4), 355-367.

[55] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta. Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66* (4), 486-501.

[56] Nekiunaite, L.; Arntzen, M. O.; Svensson, B.; Vaaje-Kolstad, G.; Abou Hachem, M., *Biotechnol. Biofuels* **2016**, *9* (1), 187-202.

[57] Li, N.; Kuo, C.-L.; Paniagua, G.; van den Elst, H.; Verdoes, M.; Willems, L. I.; van der Linden, W. A.; Ruben, M.; van Genderen, E.; Gubbens, J.; van Wezel, G. P.; Overkleeft, H. S.; Florea, B. I., *Nat. Protoc.* **2013**, *8* (6), 1155-1168.

[58] Ishihama, Y., *Mol. Cell. Proteomics* **2005**, *4*, 1265-1272.

[59] Kvach, M. V.; Ustinov, A. V.; Stepanova, I. A.; Malakhov, A. D.; Skorobogatyi, M. V.; Shmanai, V. V.; Korshun, V. A., *Eur. J. Org. Chem.* **2008**, *2008* (12), 2107-2117.

[60] Susumu, K.; Uyeda, H. T.; Medintz, I. L.; Pons, T.; Delehanty, J. B.; Mattoussi, H., *J. Am. Chem. Soc.* **2007**, *129*, 13987-13996.

APPENDIX

2.S1 Supplemental Results and Discussion

2.S1.1 Detection sensitivity of ABP 1c towards α-amylase in human plasma

From the medical perspective, it would be very attractive to measure α-amylase in human samples that allow to discriminate between disease and healthy patients. For instance, serum amylase level is often used as a biomarker of acute pancreatitis.[1-3] The normal range of serum amylase level for adults is usually between 15 international units (IU)/L to 110 IU/L (may vary between laboratories).[4] However, patients with acute pancreatitis present elevated α-amylase level in plasma, which can rise up to 4 to 6 times the upper limit of normal range. Here, we explored a potential application of our ABP in biomedical sector by analysing the detection of α-amylase in human plasma.

To determine the sensitivity of detection, the amount of enzyme detectable with the Cy5 ABP **1c** was first explored by spiking recombinant human salivary amylase at varying concentrations (0–1 unit) into 25 μL of normal human plasma (Figure 2.S1, set C), followed by Concanavalin A (ConA) beads purification to clean the sample. Then the purified glycoproteins were labeled with 50 μM ABP **1c** at pH 5.0 for 4 h. Unfortunately, recombinant human salivary amylase can only be detected at the maximum concentration applied here (1 unit/25 μL plasma), which is much higher than the upper limit of amylase level in normal (0.003 unit/25 μL) or pancreatitis (0.017 unit/25 μL) plasma. Besides, all other purified glycoproteins in plasma were also labeled under this condition, which can further decrease the sensitivity of detection. In addition, two positive control experiments were included as well by first incubating pure enzyme with 50 μM ABP **1c** at pH 5.0 for 4 h, with (Figure 2.S1, set B) or without (Figure 2.S1, set A) adding the final reaction mixture into ConA purified plasma. The detection limit of pure recombinant amylase with ABP **1c** (no plasma control, set A) is about 0.01 unit, which is just above the level of 0.017 unit/25 μL plasma of acute pancreatitis patients, but still too insensitive to the level in normal person.

In conclusion, the nonspecific labeling of all other purified glycoproteins together with the low probe potency points to a detection limit of 1 unit/25 μL plasma, meaning that these biochemical probes are not optimal for the detection of α-amylase in human plasma of patients suffering from pancreatitis.

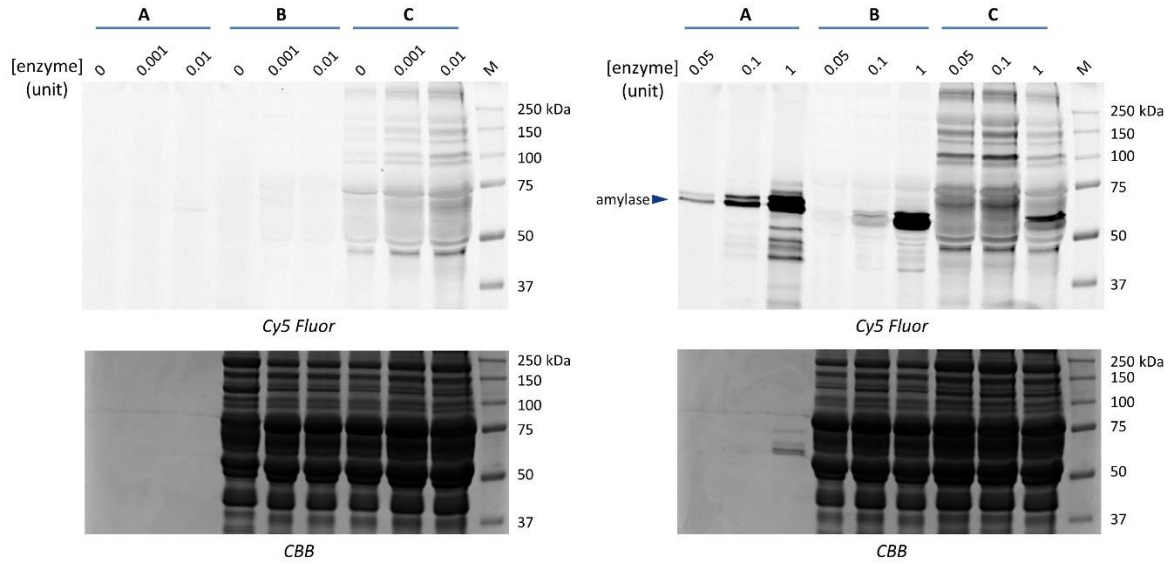


Figure 2.S1. Detection limit of α-amylase in human plasma with ABP **1c**. Set **A**: no plasma control, recombinant human saliva α-amylase (0–1 unit) was incubated with 50 μM ABP **1c** at pH 5.0 for 4 h; Set **B**: recombinant human saliva α-amylase (0–1 unit) was incubated with 50 μM ABP **1c** at pH 5.0 for 4 h, then the reaction mixture was added to ConA purified plasma; Set **C**: recombinant human saliva α-amylase (0–1 unit) was first spiked into 25 μL normal human plasma, followed by ConA purification. The purified glycoproteins were then incubated with 50 μM ABP **1c** at pH 5.0 for 4 h.

2.S1.2 Fluorescent labeling of different saliva samples with ABP 1c

In initial studies on human saliva, labeling of different saliva samples (untreated saliva, saliva supernatant and concentrated saliva pellet) with ABP **1c** were investigated. For labeling, untreated saliva (24 µL per sample), saliva supernatant (24 µL per sample), or concentrated saliva pellet (55 µg total protein per sample, diluted with assay buffer pH 7.0 to a final 24 µL volume) was incubated with 6 µL of ABP **1c** at intended concentrations in assay buffer pH 7.0 at 37 °C for 1 h. As is shown in Figure 2.S2, the detection limit of α-amylase in all these three samples is very similar (10 µM of **1c**) although slightly stronger labeling was observed in untreated saliva and saliva supernatant. However, the protein concentration of human saliva varies among different individuals, to facilitate the quantification of the protein amount needed, and to study the optimal labeling pH, the concentrated saliva pellet (16.9 µg/µL in Kpi buffer) was considered for further studies.

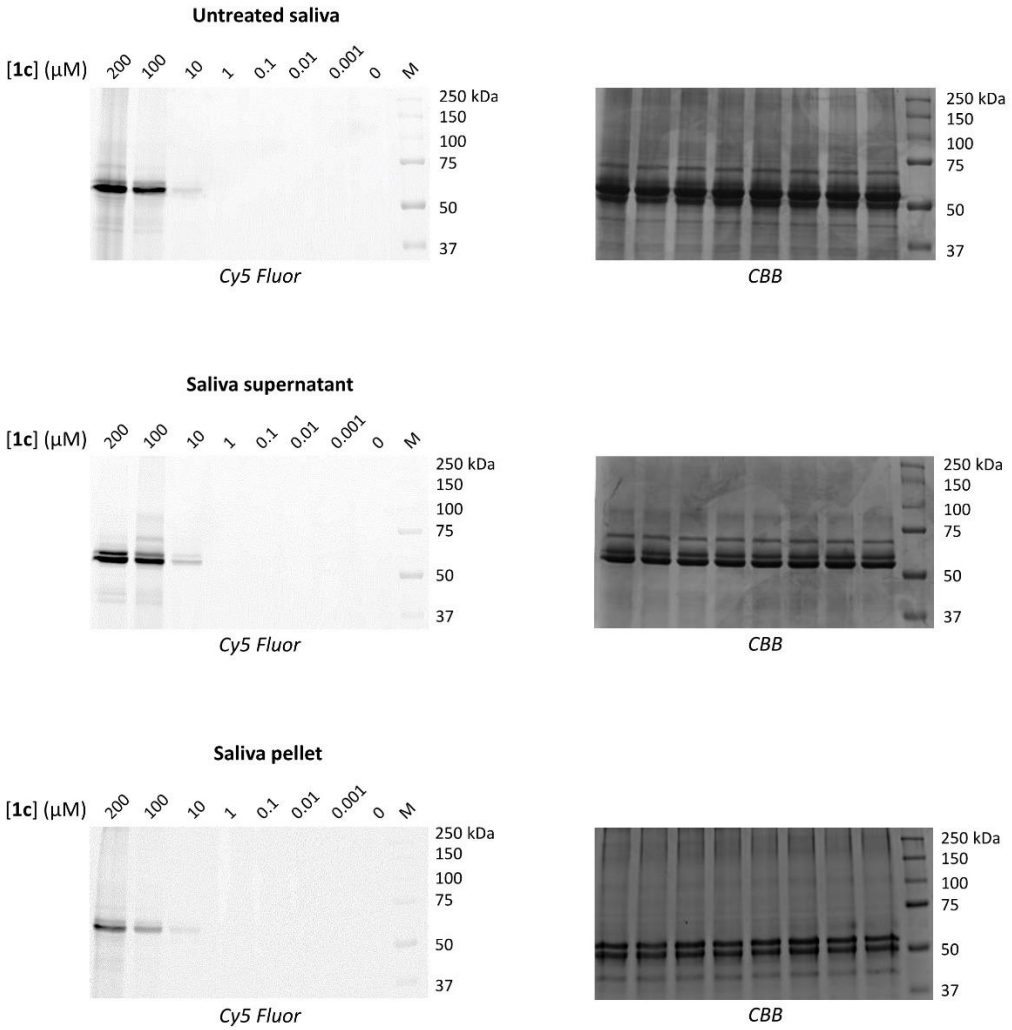


Figure 2.S2. Fluorescent labeling of untreated human saliva (24 µL), saliva supernatant (24 µL) and concentrated human saliva pellet (55 µg total protein) with different concentrations of ABP **1c** (1 h, 37 °C, pH 7.0).

2.S2 Supplemental Figures and Tables

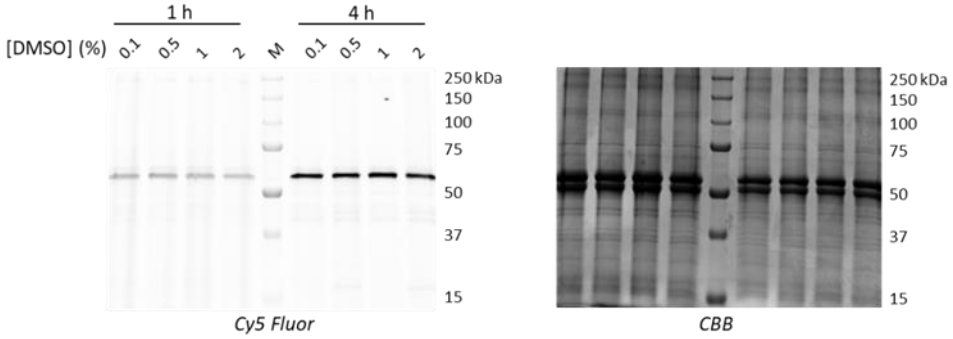


Figure 2.S3. Fluorescent labeling of concentrated human saliva (55 µg total protein) with ABP **1c** (10 µM, 37 °C, pH 5.0, 1 h or 4 h) in the presence of varying concentrations of DMSO (v/v).

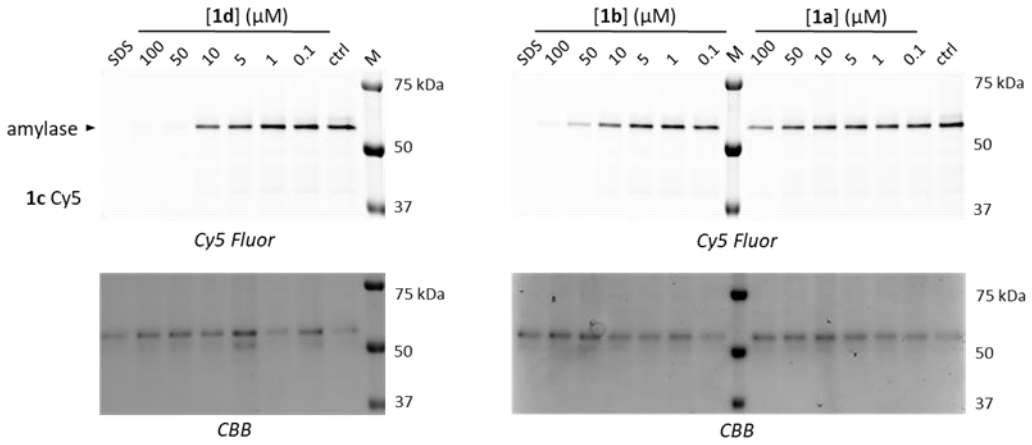


Figure 2.S4. Competitive assay of concentrated human saliva. The sample (5 µg total protein/10 µL) was pre-incubated with 2.5 µL **1a**, **1b** or **1d** at different concentrations (4 h, 37 °C, pH 5.0), followed by labeling with 2.5 µL of 10 µM ABP **1c** (2 h, 37 °C, pH 5.0).

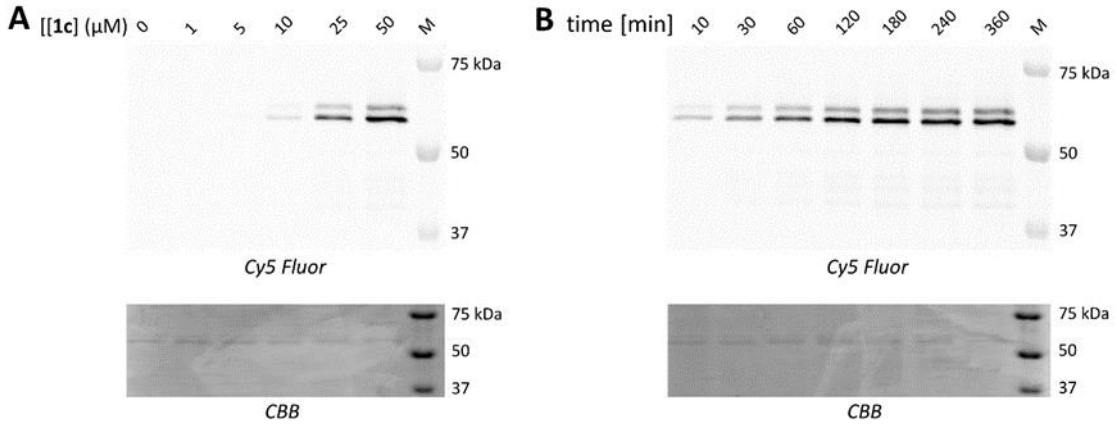


Figure 2.S5. A) Concentration-dependent labeling of recombinant human saliva α-amylase (0.2 unit) with ABP **1c** under optimal pH 5.0 after incubation 4 h at 37 °C; B) Time-dependent labeling of recombinant human saliva α-amylase with 50 μM of ABP **1c** at pH 5.0 and 37 °C.

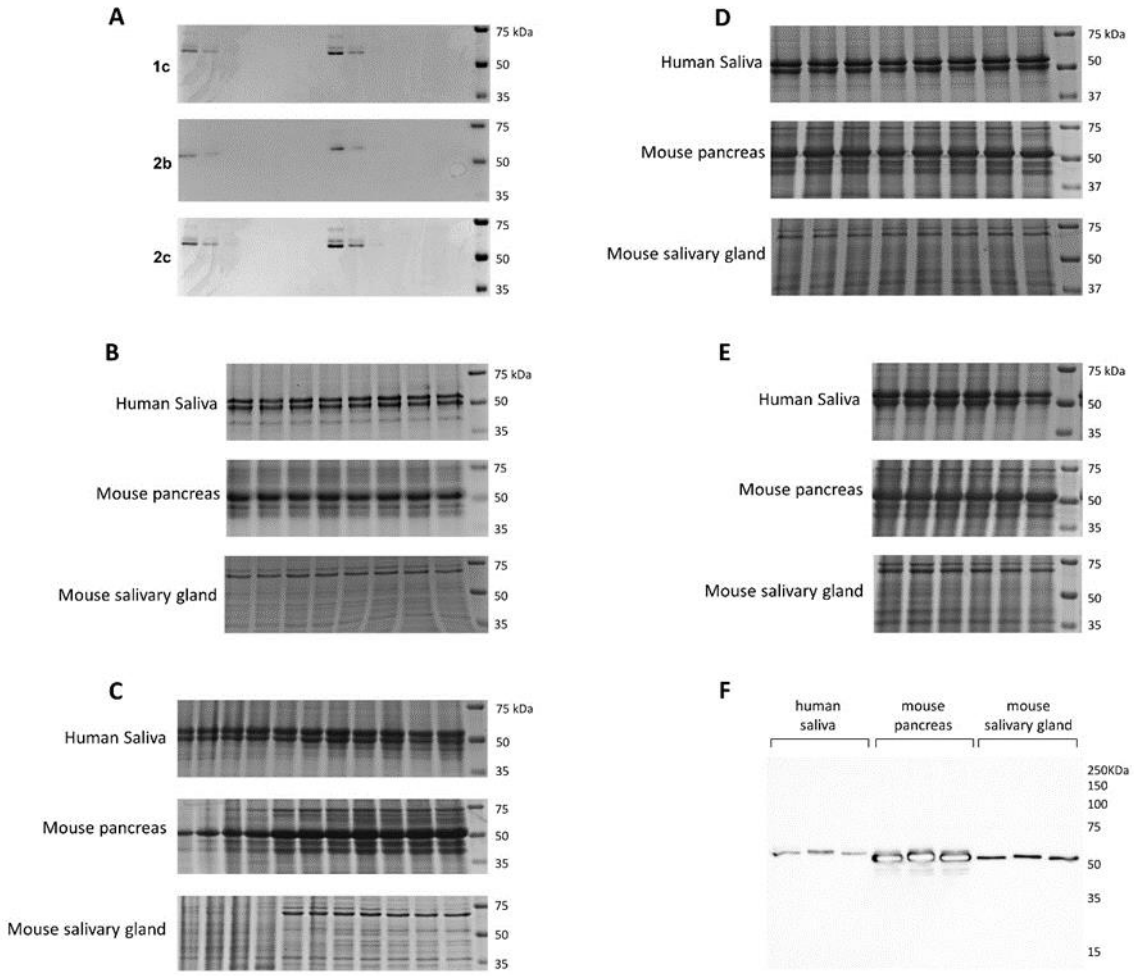


Figure 2.S6. Coomassie Brilliant Blue staining and full-length image of western blot of gels presented in main text Figure 2.5.

Table 2.S1. Data collection and refinement statistics (molecular replacement)

	Taka-amylase	**Taka-amylase + 1a**	**Taka-amylase + 2a**	**Taka-amylase + 3a**	**Taka-amylase + 1b**
PDB ID	6YQ7	6YQ9	6YQA	6YQB	6YQC
Data collection					
Space group	$P12_11$	$P12_11$	$P12_11$	$P12_11$	$P12_11$
Cell dimensions					
a, b, c (Å)	65.2, 103.1, 75.2	65.2, 103.1, 75.4	65.7, 103.3, 75.5	65.6, 102.9, 75.6	64.8, 99.0, 75.5
α, β, γ (°)	90.0, 103.6, 90.0	90.0, 103.6, 90.0	90.0, 103.8, 90.0	90.0, 103.8, 90.0	90.0, 103.6, 90.0
Resolution (Å)	59.65-1.58 (1.61-1.58)	103.10-1.55 (1.58-1.55)	103.32-1.67 (1.70-1.67)	59.78 -1.50 (1.53-1.50)	73.33-1.35 (1.37-1.35)
$CC_{1/2}$	0.99 (0.73)	0.99 (0.87)	0.99 (0.72)	0.99 (0.92)	0.99 (0.76)
I / σI	5.7 (1.7)	11.8 (2.1)	9.6 (1.1)	10.3 (1.6)	7.5 (1.0)
Completeness (%)	98.4 (97.0)	99.9 (99.8)	100 (100)	98.7 (97.0)	95.5 (73.0)
Redundancy	4.2 (4.2)	4.0 (3.8)	4.2 (4.2)	6.8 (6.8)	3.9 (2.7)
Refinement					
Resolution (Å)	1.58	1.55	1.67	1.50	1.35
No. reflections	129,847	139,928	113,215	153,221	192,746
R_{work} / R_{free}	0.16/0.19	0.15/0.17	0.17/0.20	0.16 / 0.19	0.18/0.20
No. atoms					
Protein	7,441	7,480	7,404	7,464	7,468
Ligand/ion	40/2	90/2	105/2	106/2	151/2
Water	518	815	519	785	687
B-factors					
Protein	21	18	28	22	18
Ligand/ion	36/16	23/13	34/23	27/17	22/14
Water	28	28	34	31	28
R.m.s. deviations					
Bond lengths (Å)	0.01	0.01	0.01	0.01	0.01
Bond angles (°)	1.7	1.8	1.7	1.7	1.7
*ligands in the uncomplexed structure are derived from ethylene glycol/Calcium					

Table 2.S2. Proteins identified in concentrated human saliva after biotinylated probe **1d**-treatment, competitor-inhibited control, and no probe control pull down experiments

					Intensity		
Accession	**Protein description**	**Abbreviation**	**cover [%]**	**Score**	**no probe**	**COMP (1b+1d)**	**Probe (1d)**
P60709	Actin, cytoplasmic 1	ACTB	23.7	62.07	3.37E+07	7.06E+07	6.05E+07
P13647	Keratin, type II cytoskeletal 5	KRT5	21	56.28	6.44E+07	4.16E+07	5.04E+07
P04745	Alpha-amylase 1	AMY1A	56.4	323.31	8.55E+06	2.12E+09	1.05E+10
P02769	BOVIN Serum albumin	ALB	42.7	320.79	2.37E+10	1.99E+10	3.18E+10
P04264	Keratin, type II cytoskeletal 1	KRT1	41	283.71	2.48E+09	1.37E+09	2.26E+09
P13645	Keratin, type I cytoskeletal 10	KRT10	34.1	251.8	1.14E+09	6.89E+08	1.14E+09
P00761	Trypsin	TRYP	25.1	231.57	4.89E+09	4.09E+09	4.28E+09
P22629	Streptavidin	SAV	41.7	185.2	2.57E+10	2.42E+10	2.84E+10
P00924	Enolase 1	ENO1	37.1	159.13	2.90E+09	3.04E+09	3.56E+09
P35527	Keratin, type I cytoskeletal 9	KRT9	33.4	154.79	5.52E+08	2.96E+08	3.79E+08
P04083	Annexin A1	ANXA1	40.2	151.37	2.93E+08	2.84E+06	1.25E+07
P35908	Keratin, type II cytoskeletal 2 epidermal	KRT2	25.5	118.83	2.20E+08	1.19E+08	1.62E+08
P07477	Trypsin-1	PRSS1	8.1	11.06	9.14E+07	9.35E+05	6.30E+06
P48668	Keratin, type II cytoskeletal 6C	KRT6C	26.8	106.72	1.05E+08	3.04E+07	4.65E+07
P01591	Immunoglobulin J chain	IGJ	7.5	94,959	0.00E+00	3.77E+06	3.09E+06
P19013	Keratin, type II cytoskeletal 4	KRT4	26.5	94,148	6.13E+07	2.78E+05	1.19E+06
P02533	Keratin, type I cytoskeletal 14	KRT14	28.8	93,395	1.78E+08	6.80E+07	9.36E+07
P31025	Lipocalin-1	LCN1	6.2	89,124	1.68E+06	4.39E+05	2.73E+05
P04792	Heat shock protein beta-1	HSPB1	8.3	87,612	2.20E+06	0.00E+00	7.36E+04
Q96DA0	Zymogen granule protein 16 homolog B	ZG16B	7.7	83,808	1.82E+06	1.43E+06	7.95E+06
P04080	Cystatin-B	CSTB	12.2	79,422	6.11E+06	1.43E+06	4.49E+06
P15924	Desmoplakin	DSP	4.2	78,899	3.28E+07	3.52E+06	5.99E+06
Q92817	Envoplakin	EVPL	5.9	74,045	1.44E+07	5.74E+05	3.02E+05
P25311	Zinc-alpha-2-glycoprotein	AZGP1	3.4	71,511	1.40E+06	2.16E+06	4.82E+06
Q96FQ6	Protein S100-A16	S100A16	11.7	71,024	6.61E+06	3.85E+05	3.86E+05
P07355	Annexin A2	ANXA2	2.9	70,778	3.52E+06	8.60E+05	1.51E+06
P19957	Elafin	PI3	8.5	66,777	1.84E+07	2.14E+07	2.42E+07
Q13835	Plakophilin-1	PKP1	5	36,197	1.57E+07	1.02E+06	1.39E+06
P14923	Junction plakoglobin	JUP	7	35,763	1.13E+07	7.10E+05	1.43E+06
P13646	Keratin, type I cytoskeletal 13	KRT13	19.2	32,911	3.34E+07	9.50E+05	2.61E+06
P08779	Keratin, type I cytoskeletal 16	KRT16	23.9	32,806	1.50E+07	7.95E+06	9.48E+06
P02788	Lactotransferrin	LTF	6.2	31,411	2.70E+06	7.97E+06	2.03E+07
P00450	Ceruloplasmin	CP	3.8	25,213	1.37E+07	8.43E+06	1.53E+07
P01876	Ig alpha-1 chain C region	IGHA1	11	25,078	2.44E+06	1.34E+07	2.22E+07
Q9UBC9	Small proline-rich protein 3	SPRR3	43.2	25,045	3.32E+07	4.00E+05	1.20E+06
Q9UBG3	Cornulin	CRNN	17.8	23,478	2.75E+07	1.18E+06	1.31E+06
P05109	Protein S100-A8	S100A8	23.7	21,429	1.73E+07	1.18E+06	5.84E+06
P62805	Histone H4	HIST1H4A	33	20,959	2.71E+07	3.14E+06	7.53E+06
P10412	Histone H1.4	HIST1H1E	15.5	19,715	1.05E+07	5.54E+05	3.31E+06
P35325	Small proline-rich protein 2B	SPRR2B	36.1	19,356	2.88E+07	5.94E+05	5.64E+05
Q9Y646	Carboxypeptidase Q	CPQ	4.7	18,965	2.16E+07	1.61E+07	2.35E+07
Q08188	Protein-glutamine gamma-glutamyltransferase	TGM3	3.5	14,731	4.26E+06	3.20E+06	5.58E+06
Q06830	Peroxiredoxin-1;Peroxiredoxin-2	PRDX1;PRDX2	10.6	14,399	4.28E+06	2.38E+04	8.18E+05
A8K2U0	Alpha-2-macroglobulin-like protein 1	A2ML1	1.7	14,234	9.20E+05	4.67E+06	6.00E+06
P01036	Cystatin-S;Cystatin-SN;Cystatin-SA	CST4;CST1;CST2	20.6	13,111	8.36E+05	2.72E+06	1.72E+07
P01834	Ig kappa chain C region	IGKC	34.6	12,985	2.55E+06	5.83E+06	1.38E+07
P68104	Elongation factor 1-alpha 1	EEF1A1	5	12,822	5.09E+06	3.79E+05	1.16E+05
P01857	Ig gamma-1 chain C region	IGHG1	6.7	12,503	1.02E+06	7.25E+05	4.58E+06
P09211	Glutathione S-transferase P	GSTP1	10	11,362	1.17E+06	1.50E+06	3.73E+06
P12273	Prolactin-inducible protein	PIP	17.8	10,784	6.18E+05	4.43E+05	4.67E+06
Q9Y577	E3 ubiquitin-protein ligase TRIM17	TRIM17	1.7	-2	1.66E+05	1.93E+05	1.40E+05
Q96NH3	Protein broad-minded	TBC1D32	0.8	-2	5.81E+06	4.98E+06	8.47E+06
Q9UL36	Zinc finger protein 236	ZNF236	0.5	-2	1.46E+06	5.61E+06	8.81E+06

Table 2.S3. Analysis of proteins identified in the *Aspergillus nidulans* secretomes after biotinylated probe **1d**-treatment, no probe control, and total secretome pull down experiments

					Raw abundance								
					Probe	Probe	Probe	No_Probe	No_Probe	No_Probe	Secretome	Secretome	Secretome
Accession	Abbreviation	Predicted Function	Unique peptides	Mass Da	D412_1	D412_2	D412_3	D412_4	D412_5	D412_6	D412_7	D412_8	D412_9
G5EAT0	AmyB	Amylase, GH13	14	69125	3.07E+07	5.08E+07	5.85E+07	1.69E+05	1.01E+05	1.30E+05	2.55E+06	2.57E+06	5.12E+06
Q5B7U2	AmyF	Amylase, GH13	11	50626	2.39E+06	4.75E+06	5.77E+06	9.93E+04	6.43E+04	5.48E+04	9.55E+05	1.09E+06	1.99E+06
G5EAZ3	ChiB	Endochitinase, GH18	24	44293	1.02E+06	1.42E+06	1.85E+06	1.48E+06	1.26E+06	9.78E+05	1.03E+08	1.29E+08	2.52E+08
G5EB45	AmyA	Amylase, GH13	5	54729	4.44E+05	8.52E+05	1.09E+06	1.34E+04	1.14E+04	3.66E+03	4.57E+04	3.33E+04	1.42E+05
G5EB11	AgdB	alpha-glucosidase, GH31	23	108005	4.27E+05	6.12E+05	8.62E+05	2.62E+05	1.70E+05	1.16E+05	1.21E+07	1.42E+07	2.72E+07
Q5B3Q5	-	Glycoside Hydrolase Family 55	25	97260	1.43E+05	2.61E+05	2.77E+05	1.58E+05	8.70E+04	6.12E+04	8.09E+06	1.18E+07	2.14E+07
Q5ATD5	-	Peptide hydrolase	8	54200	3.52E+04	5.16E+04	9.64E+04	3.95E+04	2.84E+04	2.24E+04	3.87E+06	5.30E+06	1.27E+07
Q5AUM3	abnC	Arabinan endo-1,5-alpha-L-arabinosidase, GH43	4	34580	3.65E+04	3.36E+04	5.03E+04	1.62E+04	2.81E+03	2.49E+04	1.63E+06	4.48E+06	3.50E+06
Q5AUI4	-	Extracellular lipase	2	31613	4.97E+03	3.86E+03	3.76E+03	7.22E+03	1.77E+03	1.29E+03	9.26E+04	1.44E+05	2.09E+05

					Spectral counts								
					Probe	Probe	Probe	No_Probe	No_Probe	No_Probe	Secretome	Secretome	Secretome
Accession	Abbreviation	Predicted Function	Unique peptides	Mass Da	D412_1_SC	D412_2_SC	D412_3_SC	D412_4_SC	D412_5_SC	D412_6_SC	D412_7_SC	D412_8_SC	D412_9_SC
G5EAT0	AmyB	Amylase, GH13	14	69125	34	40	41	0	0	0	16	13	17
Q5B7U2	AmyF	Amylase, GH13	11	50626	10	18	21	0	0	0	6	8	9
G5EAZ3	ChiB	Endochitinase, GH18	24	44293	1	1	4	0	1	0	71	87	102
G5EB45	AmyA	Amylase, GH13	5	54729	5	4	5	0	0	0	3	2	5
G5EB11	AgdB	alpha-glucosidase, GH31	23	108005	9	9	9	0	0	0	42	65	63
Q5B3Q5	-	Glycoside Hydrolase Family 55	25	97260	2	3	3	0	0	0	27	42	56
Q5ATD5	-	Peptide hydrolase	8	54200	2	2	1	0	0	0	10	17	21
Q5AUM3	abnC	Arabinan endo-1,5-alpha-L-arabinosidase, GH43	4	34580	2	1	2	0	0	0	5	8	5
Q5AUI4	-	Extracellular lipase	2	31613	1	1	2	0	0	0	4	2	4

Table 2.S4. α-Amylases previously identified in the starch secretomes of *Aspergillus nidulans*

Protein (name)	**CBM**	**Protein family**	**Uniprot**	**Molecular mass (kDa)***
α-amylase (AmyA)		GH13_1	G5EB45	52.6
α-amylase (AmyB)	CBM20	GH13_1	G5EAT0	66.7
α-amylase (AmyF)		GH13_1	Q5B7U2	48.3
*The mass is of the mature protein lacking the predicted signal peptide using Signal P (http://www.cbs.dtu.dk/services/SignalP/)				

2.S3 Supplemental References

[1] Moridani, M. Y.; Bromberg, I. L., *Clin. Biochem.* **2003,** *36* (1), 31-33.

[2] Ismail, O. Z.; Bhayana, V., *Clin. Biochem.* **2017,** *50* (18), 1275-1280.

[3] Rompianesi, G.; Hann, A.; Komolafe, O.; Pereira, S. P.; Davidson, B. R.; Gurusamy, K. S., *Cochrane Database Syst. Rev.* **2017,** *4*, CD012010.

[4] Ali Esmaili, H., *Biomed. & Pharmacol. J.* **2017,** *10* (1), 389-394.

3

A New Generation of Fluorescent Activity-Based Probes Targeting Starch-Degrading Glycosidases

3.1 Introduction

Starch is a main carbohydrate storage product of many plants and is an important constituent of the human diet. Starch polysaccharides are composed of two types of molecules – amylose and amylopectin (Figure 3.1).[1,2] Amylose, constituting about 20-25% of the starch content depending on the source, is a linear polymer chain consisting of up to 6,000 glucopyranose units linked through α-1,4-glucosidic bonds (Figure 3.1A). Amylopectin, constituting 75-80% of common starch, is composed of short α-1,4-linked linear chains of 10–60 glucose units and is highly branched at the α-1,6 positions (Figure 3.1B). Due to its intrinsic advantages of renewability, biodegradability, high abundance and low cost, starch has attracted significant attention from academic and industrial researchers alike and is currently used in numerous industrial applications. Besides its predominant use in food industry, starch, either in its native form or in the chemically or enzymatically modified form, has been widely applied as well in textile industries as wrap sizing agents, in paper industries as adhesives, in pharmaceutical industries as a drug carrier, and also in bioplastic producing and construction industries.[3-5]

Starch can be biodegraded by a cooperative set of enzymes, which based on their mode of actions, are divided into three groups: endoamylases (α-amylases), exoamylases (β-amylases, γ-amylases and α-glucosidases) and debranching enzymes (isoamylases and pullulanases).[6] Most of the enzymes belong to one single glycoside hydrolase (GH) family – the GH13 family, also known as the α-amylase family – based on the Carbohydrate-Active enZyme (CAZy, www.cazy.org) classification.[7] Starch-degrading enzymes are widely present in animals, plants and microorganisms. However, enzymes from microbial sources are the most preferred one for large-scale production mainly due to the advantages of cost effectiveness, easy availability, and ease of process manipulation.[8] So far, a large number of microbial starch-processing enzymes, especially retaining α-amylases which initiate the hydrolysis of starch by hydrolytic cleavage of internal α-1,4-glucosidic linkages randomly and destroy the whole starch structure quickly, are commercially available and have been used as catalysts in various industry sectors including food processing, the production of detergents, textile production, and also in paper and biofuel industries.[9-13]

The prevailing conditions in many industrial applications where the enzymes are used are rather extreme, especially with respect to temperature and pH. Considerable efforts have been made to improve the activity and stability of the enzymes in order to meet the requirements set by specific applications. One approach is to screen for novel microbial strains that have adapted to extreme environmental conditions, which has led to the identification of many extremophilic

enzymes with potential starch-degrading activities.[14-17] Another approach involves the use of protein engineering methods to create enzyme variants with increased activity/stability, altered pH optimum, or improved product specificity at given industrial process conditions.[18-20] Activity-based protein profiling (ABPP) has proven to be a powerful method for rapid detection and identification of enzyme activities of interest in their native environment. The work described in the previous chapter has revealed the efficacy of this methodology by developing maltobiose-configured activity-based probes (ABPs) (compound **1**, Figure 3.1C) to investigate retaining α-amylases in complex biological samples (Chapter 2). However, complete and efficient degradation of starch requires synergistic action of a set of enzymes, especially for amylopectin, of which the branching points containing α-1,6-glucosidic linkages are resistant to attack by α-1,4-specific amylases. In this chapter, the synthesis of two types of glucose-isomaltose (GIM) and isomaltose-glucose (IMG) configured pseudotrisaccharide ABPs is described (compound **2** and **3**, Figure 3.1C). The new probes are based on the reported pseudodisaccharide ABP **1**, which is substituted with a glucopyranose residue at the O6 position of either *epi*-cyclophellitol or non-reducing end sugar to mimic branched parts of the amylopectin structure. These probes may serve as tools for the discovery of industrially relevant starch-degrading enzymes that can accommodate α-1,6 branches in the -1 or -2 subsites.[21,22]

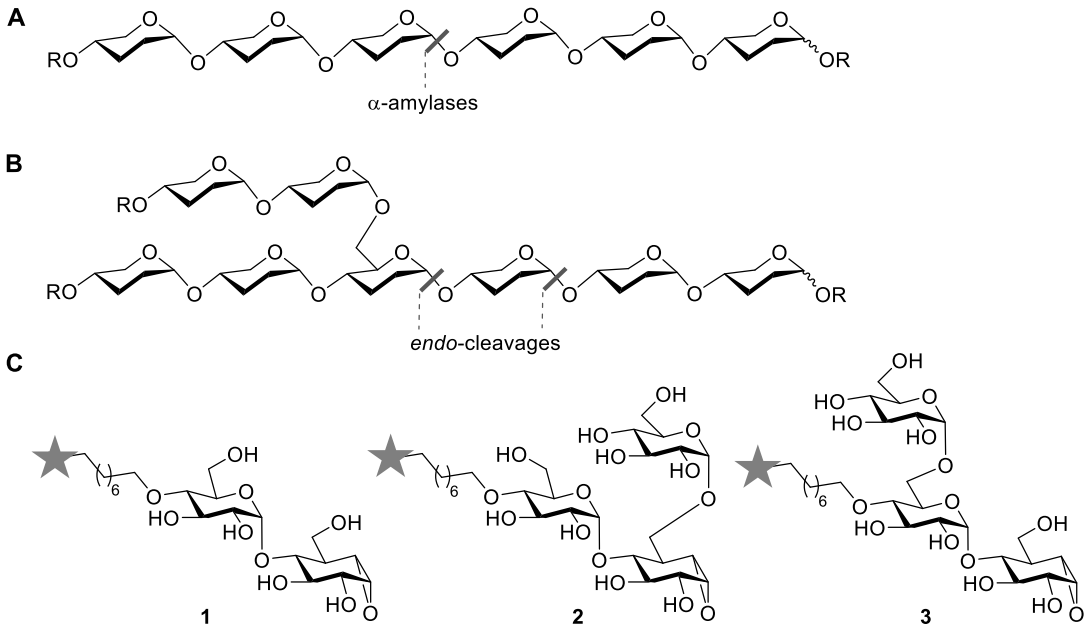
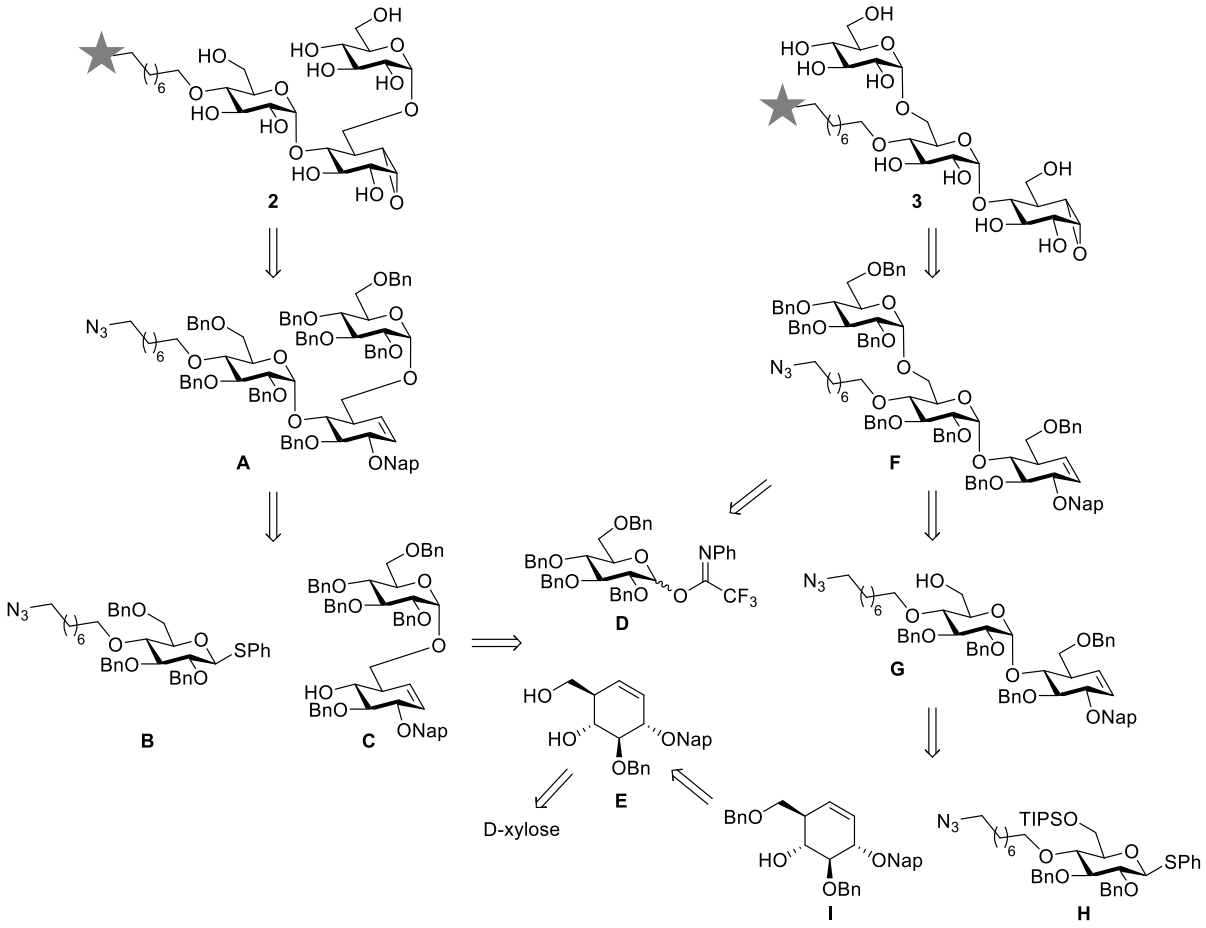


Figure 3.1. A) Chemical structure of linear amylose; B) Chemical structure of branched amylopectin; C) Structures of activity-based probes (ABPs) mimicking parts of the starch structure. Previously synthesized ABPs that target α-amylases (**1**) and new ABPs described in this chapter (**2** and **3**) that potentially target starch-degrading enzymes with potential preference for branched amylopectin-type polysaccharides.

For the preparation of the branched GIM (**2**) and IMG (**3**) configured *epi*-cyclophellitol based ABPs a synthesis strategy based on the use of three building blocks was devised (Scheme 3.1). It was envisioned that selective attachment of a reporter entity (a fluorophore or biotin) at the non-reducing end via amide coupling could be achieved by introduction of an azidooctyl group at the O4' position in an early stage. The epoxide functionality is introduced after the construction of pseudotrisaccharidic backbone to allow flexibility in the glycosylation conditions. Orthogonal protection at O2 of cyclohexene **E** with a naphthyl (Nap) ether group would allow stereoselective epoxidation after its removal following elaboration towards trisaccharide **A** or **F**. Benzyl ether protection of the remaining alcohols would allow global deprotection in one step at the end of the synthesis, and these non-participating protecting groups promote 1,2-*cis* glycosylation under the right glycosylation conditions.



Scheme 3.1. Retrosynthetic analysis of GIM (**2**) and IMG (**3**) configured probes.

Trisaccharide **A** is first disconnected into O4-alkylated thioglucoside donor **B** and pseudodisaccharide acceptor **C**. Compound **C** would be accessible from cyclohexene building block **E** by regio- and stereoselective glucosylation of the primary alcohol with per-benzylated glucosyl imidate donor **D**. Cyclohexene **E** is accessible from D-xylose based on chemistry

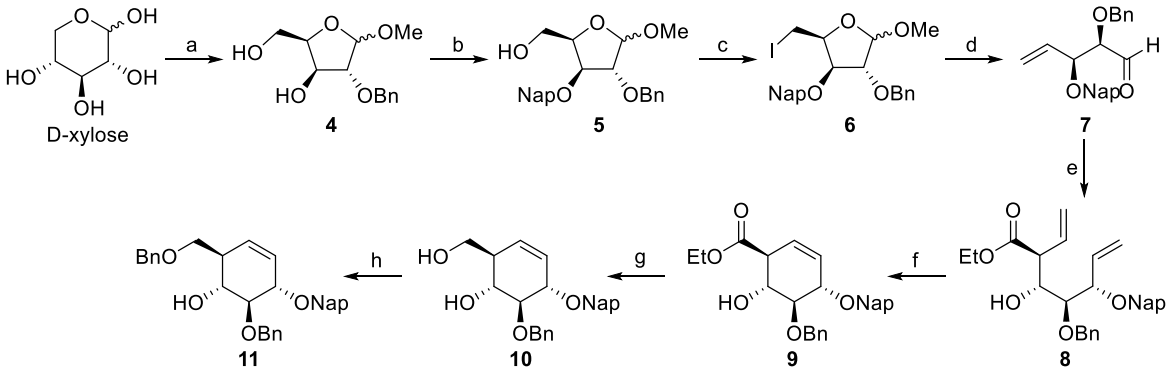
developed by Madsen and co-workers.[23,24] Similarly, trisaccharide **F** can be accessed by 1,2-*cis* glucosylation with imidate donor **D** and pseudodisaccharide acceptor **G**. Compound **G** can be synthesized by glycosylation of cyclohexene **I** with thioglucoside donor **H**, followed by desilylation. Acceptor **I** would be available from **E** by regioselective benzylation.

3.2 Results and discussion

3.2.1 Synthesis of IMG configured ABPs

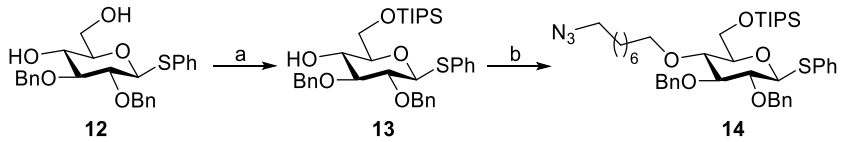
In previous work on the synthesis of maltobiose 1,6-*epi*-cyclophellitol ABPs (Chapter 2), a cyclohexene acceptor protected with a 4-methoxybenzyl (PMB) group at O6 was used for α-1,4 glycosylation reactions with different glucoside donors. Selective deprotection of the PMB ether in the resulting pseudodisaccharide followed by epoxidation with *meta*-chloroperoxybenzoic acid (*m*-CPBA) afforded an α/β mixture of epoxides (α/β as in the anomeric configuration of 'real' carbohydrates; which in terms of cyclophellitol nomenclature would translate to 'cyclophellitol/*epi*-cyclophellitol') which could be separated after silylation of O6. This key silylation step proved however not compatible with the target pseudotrisaccharidic epoxides, therefore a new cyclohexene acceptor **11** was synthesized (Scheme 3.2), in which the Nap ether can be selectively removed in a later stage to release the homoallylic alcohol for diastereoselective introduction of the desired α-epoxide.

The synthesis commenced with commercially available D-xylose, which was transformed into **4** via a four-step procedure. First, kinetic Fischer glycosylation of D-xylose afforded methyl xylofuranoside, which was then protected with a 3,5-benzylidene acetal. Benzylation of the remaining secondary alcohol followed by deprotection of the benzylidene acetal using a catalytic amount of *p*-toluenesulfonic acid (TsOH) yielded intermediate **4** in 71% over four steps. The primary alcohol in **4** was temporarily protected with a trityl group, after which the secondary alcohol was protected as a Nap ether followed by detritylation to afford intermediate **5**. Transformation of compound **5** to **10** follows procedures developed by Madsen and co-workers.[23,24] Iodination of the primary alcohol in **5** provided iodide **6**, which after Vasella fragmentation (**6** to **7**) and Barbier addition yielded diene **8**. Ring-closing metathesis of **8** was achieved by treatment with 3.5 mol-% Grubb's II catalyst, resulting in compound **9** in good yield. The ethyl ester was first reduced to the aldehyde by treatment with DIBAL-H, followed by a second reduction with $NaBH_4$ to afford diol **10**. Selective benzylation of the primary alcohol in **10** using borinate catalysis[25] yielded cylcohexene **11** in excellent yield.



Scheme 3.2. Synthesis of cyclohexene acceptor **11**. Reagents and conditions: a) *i*) acetyl chloride, MeOH, rt; *ii*) $PhCH(OMe)_2$, CSA, DMF, rt; *iii*) BnBr, NaH, TBAI, DMF, 0 °C to rt; *iv*) TsOH, DCM/MeOH, rt, 71% over 4 steps; b) *i*) TrCl, Et_3N, DMAP, DMF, rt; *ii*) NapBr, NaH, TBAI, DMF, 0 °C to rt; *iii*) TsOH, DCM/MeOH, rt, 81% over 3 steps; c) I_2, PPh_3, imidazole, THF, reflux, 95%; d) Zn, THF/H_2O, ultrasound, 40 °C, 85%; e) ethyl 4-bromocrotonate, indium powder, $La(OTf)_3$, H_2O, rt, 61%; f) Grubb's II catalyst, DCM, 40 °C, 85%; g) *i*) DIBAL-H, THF, 0 °C to rt; *ii*) $NaBH_4$, H_2O, EtOAc, rt, 87% over 2 steps; h) BnBr, 2-aminoethyl diphenylborinate, KI, K_2CO_3, MeCN, 60 °C, 95%.

To allow modification of trisaccharide inhibitors at the O4' position with reporter entities, a O4-alkylated donor **14** was synthesized (Scheme 3.3). Attempts to selectively protect the primary alcohol of diol **12**[26] as a PMB ether using borinate catalysis led to the formation of multiple side products. As an alternative, a silyl ether was chosen for selective protection of the primary alcohol. Reaction of **12** with triisopropylsilyl chloride (TIPSCl) proceeded smoothly, resulting in **13** in excellent yield. Subsequent alkylation of the remaining secondary alcohol was performed with an excess (4.0 eq.) of 1-azido-8-iodooctane in the presence of sodium hydride (NaH), giving product **14** in high yield. The TIPS ether was chosen over a *tert*-butyldiphenylsilyl (TBDPS) ether as the latter was found to be unstable during alkylation in the presence of strong base.



Scheme 3.3. Synthesis of thioglucoside donor **14**. Reagents and conditions: a) TIPSCl, imidazole, DCM, 0 °C to rt, 97%; b) 1-azido-8-iodooctane, NaH, 18-crown-6, DMF, 0 °C to rt, 94%.

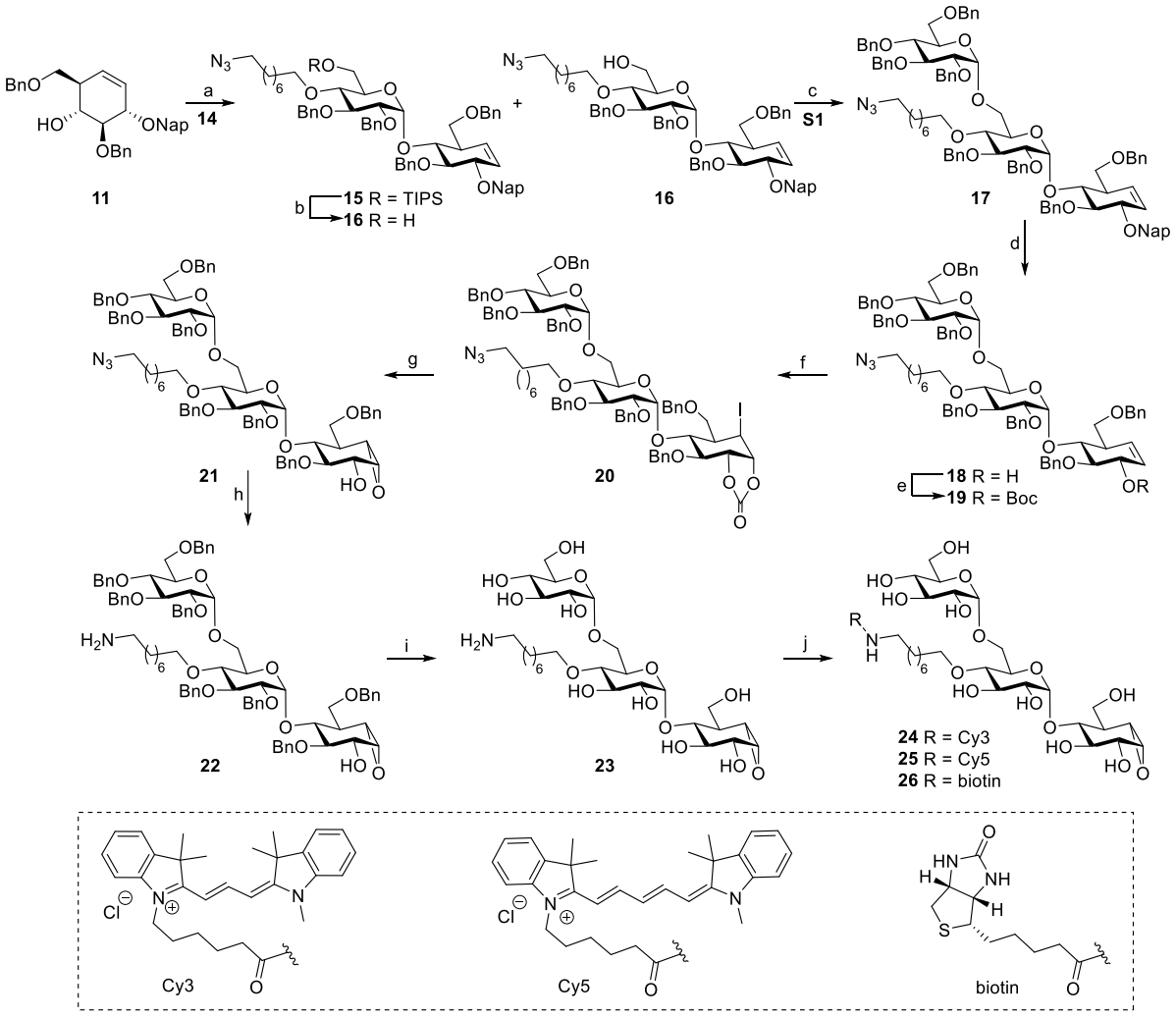
With donor **14** and acceptor **11** in hand, the synthesis of IMG probes was investigated (Scheme 3.4). Acceptor **11** was glycosylated with donor **14** under *N*-iodosuccinimide (NIS), trifluoromethanesulfonic acid (TMSOTf) and *N*,*N*-dimethyl formamide (DMF) activating conditions, affording pseudodisaccharide **15** in good α-selectivity along with the desilylated

product **16** which was formed due to prolonged reaction times (45 h) in the presence of stoichiometric amounts of Lewis acid (TMSOTf). The TIPS ether in **15** was then removed by TBAF and the so-obtained **16** was combined with that formed in the previous step. The second α-selective glycosylation on the primary alcohol of **16** was achieved by reacting with per-benzylated glucosyl imidate donor **S1** (Chapter 2) under TMSI/$OPPh_3$ conditions, yielding pseudotrisaccharide **17** in good yield. The Nap ether was removed with DDQ to provide the allylic alcohol **18**. Direct epoxidation of **18** with *m*-CPBA at 0 °C afforded an inseparable mixture of α/β epoxides in a 3/1 ratio. To improve the stereoselectivity of the reaction, an iodocarbonylation approach was employed.[27-30] *t*-Butyloxylcarbonyl (Boc) protection of the secondary alcohol afforded **19**. It was found that unreacted starting material **18** can readily react with product **19** by attacking the carbonyl group of the Boc in **19** to form a dimer side product, and an excess (15 eq.) of di-*tert*-butyl dicarbonate was needed to suppress this side reaction. Compound **19** was then reacted with NIS to yield iodocarbonate **20** with complete stereospecificity. Hydrolysis of the carbonate by sodium methoxide in methanol and subsequent intramolecular iodine displacement afforded α-epoxide **21**. The azido group was reduced by Staudinger reduction resulting in amine **22**, which was then debenzylated using Pearlman's catalyst ($Pd(OH)_2/C$) under hydrogen atmosphere, affording **23** in quantitative yield.

To ensure that the latter $Pd(OH)_2/C$ catalyzed hydrogenation reaction proceeded smoothly, several key factors need to be taken into account, including the use of a suitable solvent system, reactant concentration, catalyst poisoning by the amine, reaction time and a way to monitor the reaction process. In this case, a solvent mixture of H_2O/dioxane (1/3) was employed. The higher proportion of dioxane was needed to dissolve the lipophilic starting material and partially deprotected intermediates. At the same time, a low reactant concentration, of around 0.015 M, ensured that there was enough water to dissolve all hydrophilic products. For different substrates, the ratio of solvents or the reaction concentration may need to be adapted accordingly. To avoid catalyst poisoning by the free amine, an equivalent (relative to the reactant) amount of trifluoroacetic acid (TFA) was added to capture the amine by protonation. Of note, oxiranes are known to undergo reductive opening of the strained three-membered ring under palladium catalysis,[31] and they are prone to hydrolysis in the presence of strong acid, however the epoxide functionality remained unaffected by employing stoichiometric amounts of Pearlman's catalyst to minimize (around 3 h) reaction times. Direct transformation from azide **21** to **23** using the same hydrogenation conditions as for the conversion of **22** proved to be sluggish, and TLC-analysis indicated no further conversion of intermediates after 2 h, which

was presumably due to the emerging amine that poisoned palladium catalyst prior to being captured by TFA.

Returning to the synthesis of the target IMG probes, Cy3COOH, Cy5COOH or biotin were pre-activated as their pentafluorophenyl esters and directly coupled with the primary amine **23** to yield ABPs **24-26** after semi-preparative HPLC purification.

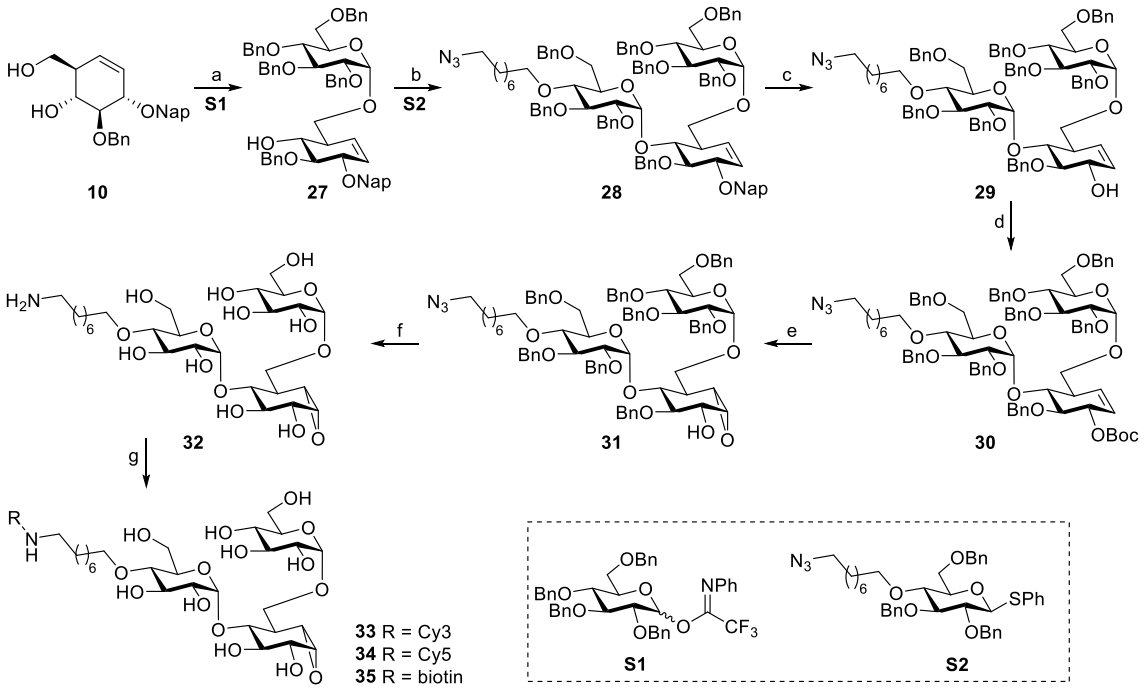


Scheme 3.4. Synthesis of ABPs **24-26**. Reagents and conditions: a) **14**, NIS, TMSOTf, DMF, DCM, 3 Å MS, 0 °C, **15** 52%, **16** 32%; b) TBAF, THF, rt, 87%; c) **S1**, TMSI, $OPPh_3$, 3 Å MS, DCM, rt, 82%; d) DDQ, DCM/H_2O (10/1), rt; e) Boc_2O, DMAP, THF, 0 °C to rt, 75% over 2 steps; f) NIS, AcOH, DCM, rt, 61%; g) NaOMe, MeOH, DCM, rt, 81%; h) polymer-bound PPh_3, MeCN, H_2O, 65 °C, 88%; i) $Pd(OH)_2/C$, TFA, H_2, dioxane, H_2O, rt, quant; j) Cy3COOH or Cy5COOH or biotin, pentafluorophenyl trifluoroacetate, DIPEA, DMF, Cy3 40%, Cy5 55%, biotin 38%.

3.2.2 Synthesis of GIM configured ABPs

The synthesis of GIM configured probes follows a strategy similar to that employed for the synthesis of the IGM configured probes as outlined in the previous section. Acceptor **10** was

glycosylated with imidate donor **S1** in a TMSI/$OPPh_3$ mediated reaction, leading to the regio- and stereoselective generation of **27** in 65% yield (Scheme 3.5). The second glycosylation of acceptor **27** with O4-alkylated thioglucoside donor **S2** (Chapter 2) was accomplished employing the same pre-activation protocol as for the synthesis of **15**, affording pseudotrisaccharide **28** in good yield and selectivity. Elaboration of **28** to epoxide **31** follows the same procedures as those developed for **21**. Deprotection of the Nap ether with DDQ (**28** to **29**) and subsequent Boc protection resulted in product **30**. Conversion of **30** into cyclic iodocarbonate with NIS was not complete after 22 h and prolonged reaction times were not tried because it was reasoned that these would lead to deprotection of the Boc group in the presence of HOAc. The carbonate intermediate and the starting material **30** were difficult to separate by standard column chromatography, therefore after work-up the resulting crude mixture was directly used for base hydrolysis, stereospecifically affording epoxide **31** in 46% over two steps, while unreacted **30** was recovered as well. Azide reduction and global debenzylation of **31** were performed under Birch conditions, affording compound **32** after desalting by size exclusion chromatography over HW40. Sodium hydroxide, formed while quenching the Birch reaction, was neutralized with HOAc. Omission of this neutralization step can lead to hydrolysis of the epoxide during concentration of the reaction mixture in the presence of aqueous base. Fully deprotected amine **32** was then coupled with the corresponding pentafluorophenyl activated esters of Cy3COOH, Cy5COOH or biotin, yielding ABPs **33**-**35** after HPLC purification.



Scheme 3.5. Synthesis of ABPs **33**-**35**. Reagents and conditions: a) **S1**, TMSI, $OPPh_3$, 3 Å MS, DCM, rt, 65%; b) **S2**, NIS, TMSOTf, DMF, DCM, 3 Å MS, 0 °C, 78%; c) DDQ, DCM/H_2O (10/1), rt, 76%; d) Boc_2O, DMAP, THF, 0 °C to rt, 85%; e) *i*) NIS, AcOH, DCM, rt; *iii*) NaOMe, MeOH, DCM, rt, 46% over 2 steps; f) Na, NH_3 (l), ^{t}BuOH, THF, -60 °C, 57%; g) Cy3COOH or Cy5COOH or biotin, pentafluorophenyl trifluoroacetate, DIPEA, DMF, Cy3 28%, Cy5 53%, biotin 52%.

3.3 Conclusion

In this chapter the synthesis of two types of branched pseudotrisacchride ABPs is presented. NIS/TMSOTf/DMF mediated α-1,4-glycosylations of thioglucoside donors and TMSI/$OPPh_3$ mediated α-1,6-glycosylations of anomeric imidate donors on cyclohexene acceptors afforded access to the desired pseudotrisacchride backbones. A key step involved orthogonal protection of the O2 position of cyclohexene with a Nap ether, which allowed selective deprotection post-glycosylation and the resulting allylic alcohol gave access to the desired α-epoxide functionality by stereoselective epoxidation.

Labeling patterns and efficiencies of the newly synthesised ABPs in complex biological samples containing starch-degrading enzymes are projected to be investigated in the near future. Once these are established, the data can be compared with that of maltobiose-configured ABPs reported previously. By mimicking parts of the branched amylopectin structure, the IMG- and GIM probes are expected to show preference or specificity towards starch-degrading enzymes that like branched substrates in -1 or -2 subsites. Furthermore, the set of branched ABPs can be used for the screening of microbial species with the aim to discover new industrially relevant enzymes with beneficial characteristics.

3.4 Experimental methods

General experimental details

All reagents were of experimental grade and were used without further purification unless stated otherwise. Dichloromethane (DCM), acetonitrile (MeCN) and tetrahydrofuran (THF) were stored over 3 Å molecular sieves and *N*,*N*-dimethylformamide (DMF) was stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an Argon or N_2 atmosphere unless stated otherwise. Reactions were monitored by analytical thin layer chromatography (TLC) using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV-absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}\cdot H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4\cdot H_2O$ (10 g/mL) in 10% sulfuric acid followed by charring at ~150 °C. Column chromatography was performed manually using Screening Device b.v. silica gel 60 (0.04-0.063 mm) in the indicated solvents. LC-MS analysis was performed on a LCQ Advantage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a

Surveyor HPLC system (Thermo Finnigan) equipped with a C18 column (Gemini, 4.6 mm x 50 mm, 5 µM particle size, Phenomenex). The applied buffers were H_2O, acetonitrile (MeCN) and 1% aqueous trifluoroacetic acid (TFA). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-400 (400/101 MHz), and Bruker AV-500 (500/126 MHz) spectrometers in the given solvent. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (TMS) as internal standard (^{1}H NMR in $CDCl_3$) or the residual signal of the deuterated solvent. Coupling constants (J) are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. The following abbreviations are used to describe peak patterns when appropriate: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), Ar (aromatic), C_q (quarternary carbon). 2D NMR experiments (COSY, HSQC) were carried out to assign protons and carbons of the new structures. High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electronspray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL/min, capillary temperature 250 °C) with resolution R = 60000 at m/z 400 (mass range m/z = 150 – 2000) and dioctyl phthalate (m/z = 391.28428) as a "lock mass". The high-resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan).

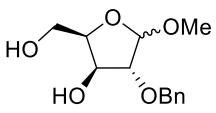
Experimental Procedures and Characterization Data of Products

Known compounds **12**[26], Cy3COOH[32], and Cy5COOH[32] were synthesized following procedures as previously described and their spectroscopic data are in agreement with those reported.

General procedure A

The appropriate carboxylic acid (25 µmol) was dissolved in DMF (200 µL). DIPEA (24 µL, 138 µmol, 5.5 eq) and pentafluorophenyl trifluoroacetate (8.6 µL, 50 µmol, 2.0 eq) were added and the mixture was stirred at rt for 2 – 3 h until LC-MS indicated almost full conversion of the free acid. Water (2.5 µL) was then added and part of the stock solution (for Cy3 and Cy5: 1.05 eq acid to amine; for biotin: 1.2 eq acid to amine) was added a solution of amine in DMF (300 µL). The reaction was stirred overnight and the product was purified on semi-preparative HPLC eluting with a linear gradient of solution A (MeCN) in solution B (50 mM NH_4OAc in H_2O). The fractions were concentrated under reduced pressure, co-evaporated with MilliQ/MeCN (1/1, at least 3 times), dissolved with MilliQ/^{t}BuOH and lyophilized to yield the product. *Note: NH_4OAc from HPLC solution is hard to remove, and repeated co-evaporation and lyophilization are necessary to remove final traces of NH_4OAc. For all final compounds, decrease of the amount is less than 0.1 – 0.3 mg after the last lyophilization.*

Compound 4



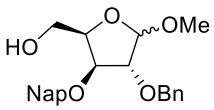
D-Xylose (15.0 g, 100 mmol) was added to a solution of acetyl chloride (3.0 mL, 42 mmol) in MeOH (300 mL). The reaction mixture was stirred at rt for 5 h, after which it was neutralized with TEA. The solution was concentrated *in vacuo* and co-evaporated with dioxane (3 x). The resulting residue was taken up in DMF (300 mL), benzaldehyde

dimethyl acetal (22.0 mL, 150 mmol) and CSA (7.0 g, 30 mmol) were added. The reaction was set up on a rotavap at rt under reduced pressure (≈16 mbar) overnight. TLC analysis indicated the presence of starting material so more benzaldehyde dimethyl acetal (7.0 mL, 50 mmol) was added and the reaction was warmed up to 40 °C under reduced pressure (≈140 mbar). After stirring for 6 h, the reaction mixture was diluted with H_2O (600 mL), extracted with EtOAc (2 x 700 mL). The combined organic layers were washed with H_2O (3 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*.

The resulting residue was co-evaporated with toluene (3 x) and directly dissolved in dry DMF (300 mL). BnBr (18.0 mL, 150 mmol) and a catalytic amount of TBAI (1.85 g, 5.00 mmol) were added, after which the solution was cooled to 0 °C. Sodium hydride (60% dispersion in mineral oil, 10.0 g, 250 mmol) was added in three portions, and the reaction mixture was stirred at rt for 20 h. Then the reaction was cooled to 0 °C and quenched carefully with H_2O. The solution was further diluted with water, extracted with Et_2O (2 x). The combined organic layers were washed with H_2O (3 x), brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo*.

The resulting light-yellow solid was re-dissolved in CH_2Cl_2 (100 mL) and MeOH (300 mL), and *p*-toluenesulfonic acid (1.9 g, 10 mmol) was added. The reaction mixture was stirred at rt for 48 h, after which it was quenched with Et_3N (1.4 mL, 10 mmol). The solvent was evaporated under reduced pressure and the crude was purified with silica column chromatography (Pentane/EtOAc 3:1 → 1:2) to obtain compound **4** (18.2 g, 71.7 mmol, 71%) as an anomeric mixture (α/β ratio ≈ 2:1). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.53 – 7.15 (m, 17H), 4.92 (s, 2H, H-1α), 4.77 (d, J = 4.2 Hz, 1H, H-1β), 4.67 (d, J = 6.7 Hz, 3H), 4.64 – 4.57 (m, 4H), 4.56 – 4.51 (m, 1H), 4.40 – 4.34 (m, 2H), 4.34 – 4.29 (m, 3H), 4.17 (dt, J = 7.2, 3.3 Hz, 1H), 3.95 (d, J = 2.0 Hz, 2H), 3.91 – 3.84 (m, 6H), 3.81 (d, J = 3.3 Hz, 2H), 3.39 (s, 7H), 3.37 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 137.7, 137.4, 128.6, 128.5, 128.3, 128.1, 127.9, 107.3, 100.5, 88.2, 85.7, 82.3, 76.5, 75.8, 75.8, 72.8, 72.1, 62.2, 62.0, 55.5, 55.3 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{18}O_5Na$ 277.1046, found 277.1043.

Compound 5



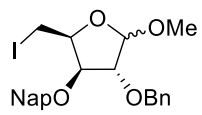
Compound **4** (18.2 g, 71.7 mmol) was dissolved in dry DMF (350 mL). Et_3N (25.0 mL, 180 mmol), trityl chloride (40.0 g, 144 mmol), and DMAP (438 mg, 3.59 mmol) were added successively and the reaction mixture was stirred at rt for 25 h. The mixture was diluted with H_2O (650 mL), extracted with EtOAc (2 x 700 mL). The combined organic layers were washed with H_2O (3 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*.

The resulting oil was again taken up in dry DMF (300 mL), and NapBr (23.8 g, 108 mmol) and a catalytic amount of TBAI (1.32 g, 3.59 mmol) were added, after which the solution was cooled to 0 °C. Sodium hydride (60% dispersion in mineral oil, 7.20 g, 180 mmol) was added in 2 portions, and the reaction mixture was stirred at ambient temperature for 15 h. Then the reaction was cooled to 0 °C and quenched carefully with H_2O. The solution was further diluted with water, extracted with Et_2O (2 x).

The combined organic layers were washed with H_2O (3 x), brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo*.

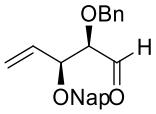
The resulting residue was redissolved in a mixture of CH_2Cl_2/MeOH (150 mL/150 mL), and *p*-toluenesulfonic acid was added until pH 2 was reached. The reaction mixture was stirred at rt for 21 h, after which it was neutralized with Et_3N. The solvent was evaporated under reduced pressure and the crude was purified with silica column chromatography (Pentane/EtOAc 5:1 → 1:1) to obtain compound **5** (23.2 g, 58.8 mmol, 81%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.85 – 7.75 (m, 9H), 7.71 (s, 3H), 7.51 – 7.25 (m, 23H), 4.92 – 4.83 (m, 2H), 4.83 – 4.70 (m, 3H), 4.68 – 4.61 (m, 3H), 4.58 (d, *J* = 2.0 Hz, 1H), 4.53 (d, *J* = 14.0 Hz, 3H), 4.50 – 4.44 (m, 2H), 4.32 (dt, *J* = 6.9, 4.7 Hz, 2H), 4.25 – 4.18 (m, 3H), 4.13 (dd, *J* = 3.8, 1.9 Hz, 2H), 4.09 (dd, *J* = 6.5, 4.2 Hz, 1H), 3.88 – 3.74 (m, 6H), 3.41 (s, 6H), 3.37 (s, 3H), 2.63 (t, *J* = 6.6 Hz, 2H), 2.47 (dd, *J* = 8.6, 5.0 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 137.6, 137.5, 135.0, 134.9, 133.3, 133.2, 133.1, 133.1, 128.5, 128.5, 128.4, 128.3, 128.1, 129.0, 128.0, 127.9, 127.8, 127.8, 126.9, 126.8, 126.3, 126.3, 126.2, 126.2, 125.7, 125.7, 108.0, 100.2, 87.3, 84.6, 82.8, 82.2, 80.7, 76.3, 72.9, 72.7, 72.6, 72.3, 62.4, 62.3, 55.7, 55.2 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{24}H_{26}O_5Na$ 417.1673, found 417.1670.

Compound 6



Alcohol **5** (23.2 g, 58.8 mmol) was co-evaporated with toluene (3 x), after which it was dissolved in anhydrous THF (250 mL). Triphenylphosphine (23.1 g, 88.2 mmol) and imidazole (8.02 g, 118 mmol) were added to the solution, which was heated to reflux. A solution of iodine (22.4 g, 88.2 mmol) in anhydrous THF (60 mL) was added, and the reaction mixture was heated at reflux until TLC-analysis revealed full conversion of the starting material (around 30 min). The reaction mixture was then cooled to rt and quenched with sat. aq. $Na_2S_2O_3$, after which the mixture was concentrated under reduced pressure to remove most of the THF. The residue was re-dissolved in EtOAc, washed with H_2O and brine, dried with $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 20:1 → 4:1) to obtain compound **6** (28.2 g, 56.0 mmol, 95%) as a yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.89 – 7.72 (m, 13H), 7.55 – 7.41 (m, 9H), 7.39 – 7.26 (m, 11H), 7.26 – 7.19 (m, 5H), 4.94 (d, *J* = 1.3 Hz, 2H), 4.88 (d, *J* = 4.2 Hz, 1H), 4.83 – 4.61 (m, 7H), 4.58 – 4.38 (m, 9H), 4.25 (dd, *J* = 6.4, 4.8 Hz, 1H), 4.10 (dd, *J* = 5.8, 2.7 Hz, 2H), 4.07 – 4.01 (m, 3H), 3.48 (d, *J* = 6.5 Hz, 1H), 3.45 (d, *J* = 6.5 Hz, 1H), 3.44 (s, 6H), 3.41 (s, 3H), 3.41 – 3.35 (m, 3H), 3.24 (dd, *J* = 10.2, 7.4 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 137.6, 137.4, 135.2, 135.0, 133.3, 133.3, 133.2, 133.1, 128.6, 128.4, 128.3, 128.2, 128.1, 128.0, 127.9, 127.8, 127.1, 126.8, 126.4, 126.3, 126.2, 126.1, 126.1, 125.9, 108.5, 101.0, 86.7, 84.0, 82.1, 81.9, 81.8, 77.7, 73.0, 72.8, 72.8, 72.2, 56.2, 55.7, 4.5, 3.1 ppm. HRMS (ESI) m/z: $[M+NH_4]^+$ calc for $C_{24}H_{29}IO_4N$ 522.1136, found 522.1134.

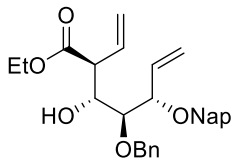
Compound 7



Zinc dust was activated and dried immediately before use: it (around 100 g) was added slowly to a vigorously stirred HCl solution (3 M, 600 mL). The mixture was stirred for 5-10 min at ambient temperature, filtered, and subsequently rinsed with water, dioxane Et_2O, after which it was dried under high vacuum at 40 °C.

Iodide **6** (28.2 g, 56.0 mmol) was divided into 3 batches (A: 9.21 g, 18.2 mmol; B: 9.44 g, 18.7 mmol; C: 9.54 g, 18.9 mmol). Each batch was dissolved in a mixture of THF/H_2O (9:1, v/v, 200 mL). Then the freshly activated zinc dust (15.0 equiv. for each) was added. The reaction mixture was flushed with argon and sonicated under argon stream at 40 °C until TLC analysis revealed full conversion of the starting material (around 1.5 h). The reaction mixture was filtrated and concentrated under reduced pressure to remove most of the THF. The combined resulting residue was diluted with Et_2O, washed with H_2O and brine, dried with $MgSO_4$, filtered, and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/Et_2O 30:1 → 10:1) to obtain compound **7** (16.4 g, 47.4 mmol, 85%) as a clear oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.69 (d, J = 1.5 Hz, 1H), 7.85 – 7.72 (m, 3H), 7.69 (d, J = 1.6 Hz, 1H), 7.51 – 7.43 (m, 2H), 7.39 – 7.23 (m, 6H), 5.96 (ddd, J = 17.3, 10.4, 7.7 Hz, 1H), 5.42 – 5.29 (m, 2H), 4.77 (dd, J = 15.3, 12.2 Hz, 2H), 4.62 (d, J = 12.1 Hz, 1H), 4.51 (d, J = 12.3 Hz, 1H), 4.20 (ddt, J = 7.6, 4.2, 1.0 Hz, 1H), 3.84 (dd, J = 4.1, 1.5 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 202.6, 137.2, 135.1, 133.9, 133.3, 133.1, 128.6, 128.3, 128.3, 128.2, 128.0, 127.8, 126.9, 126.2, 126.1, 126.0, 120.1, 85.3, 79.9, 73.6, 70.8 ppm. HRMS (ESI) m/z: $[M+NH_4]^+$ calc for $C_{23}H_{26}O_3N$ 364.1907, found 364.1903.

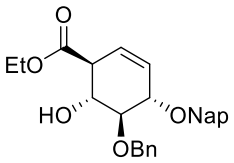
Compound 8



Ethyl 4-bromocrotonate (27.0 mL, 145 mmol), $La(OTf)_3$ (57.3 g, 96.8 mmol), and indium powder (12.2 g, 106 mmol) were added to a solution of aldehyde **7** (15.4 g, 44.5 mmol) in H_2O (220 mL). After stiring vigorously at rt for 3 days, the reaction mixture was filtered through a plug of celite, washed with EtOAc. The layers were separated and the organic layer was washed with H_2O, brine, dried with $MgSO_4$, filtered, and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 13:1 → 5:1) to obtain compound **8** as a clear oil (9.21 g, 20.0 mmol, 61% based on recovered starting material **7** (4.1 g, 12 mmol)). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.92 – 7.67 (m, 4H), 7.44 (ddd, J = 14.1, 7.4, 2.5 Hz, 3H), 7.37 – 7.23 (m, 5H), 5.92 – 5.79 (m, 1H), 5.78 – 5.65 (m, 1H), 5.48 – 5.37 (m, 2H), 5.17 (dd, J = 10.2, 1.3 Hz, 1H), 5.11 – 4.99 (m, 2H), 4.78 (d, J = 11.7 Hz, 1H), 4.57 (dd, J = 19.3, 11.5 Hz, 2H), 4.23 (t, J = 7.7 Hz, 1H), 4.10 (q, J = 7.1 Hz, 2H), 3.99 (td, J = 9.5, 8.8, 1.2 Hz, 1H), 3.57 (dd, J = 7.7, 1.1 Hz, 1H), 3.29 (t, J = 9.2 Hz, 1H), 2.71 (dd, J = 10.0, 1.9 Hz, 1H), 1.21 (t, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.6, 138.5, 135.9, 135.0, 133.4, 133.0, 133.0, 128.5, 128.2, 128.1,

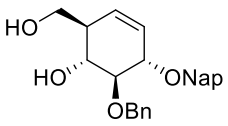
128.0, 127.8, 127.8, 126.6, 126.2, 126.1, 126.0, 120.2, 120.1, 83.0, 79.5, 74.7, 72.2, 71.0, 61.0, 55.3, 14.2 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{29}H_{32}O_5Na$ 483.2142, found 483.2139.

Compound 9



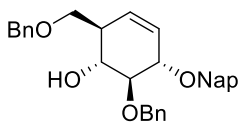
Compound **8** (9.21 g, 20.0 mmol) was dissolved in DCM (100 mL), after which the solution was sonicated under argon stream for 30 min. The second-generation Grubbs catalyst (600 mg, 0.70 mmol, 3.5 mol-%) was added and the reaction mixture was heated at reflux in the dark for 17 h under argon atmosphere. The reaction mixture was concentrated and purified with silica column chromatography (Pentane/EtOAc 9:1 → 5:1) to obtain compound **9** (7.38 g, 17.0 mmol, 85%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.93 – 7.67 (m, 4H), 7.52 – 7.42 (m, 3H), 7.39 – 7.25 (m, 5H), 5.82 (ddd, J = 10.2, 2.9, 2.2 Hz, 1H), 5.67 (dt, J = 10.2, 2.2 Hz, 1H), 4.98 (d, J = 11.4 Hz, 1H), 4.88 – 4.77 (m, 3H), 4.26 – 4.10 (m, 4H), 3.68 (dd, J = 9.8, 7.5 Hz, 1H), 3.26 (dq, J = 8.8, 2.9 Hz, 1H), 2.99 (d, J = 2.4 Hz, 1H), 1.26 (t, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 172.0, 138.5, 135.6, 133.4, 133.1, 128.6, 128.4, 128.4, 128.0, 128.0, 127.9, 127.8, 126.8, 126.3, 126.1, 126.0, 124.2, 82.6, 79.3, 75.0, 72.1, 70.5, 61.4, 50.2, 14.3 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{27}H_{28}O_5Na$ 455.1829, found 455.1824.

Compound 10



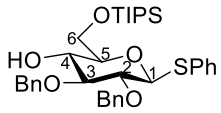
Ester **9** (4.0 g, 9.2 mmol) was co-evaporated with toluene (3 x), after which it was dissolved in dry THF (45 mL). DIBAL-H (1 M in toluene, 46 mL, 46 mmol) was added to the solution at 0 °C. The reaction mixture was stirred for 30 min at 0 °C and TLC analysis revealed full conversion of the starting material. It was then quenched with EtOAc (18 mL) at 0 °C, followed by slow addition of H_2O (9.2 mL) and sodium borohydride (2.3 g, 60 mmol). The reaction mixture was stirred for 18 h at ambient temperature, after which it was concentrated *in vacuo*. The crude gelatinous product was dispersed with EtOAc (250 mL), to which 1 M HCl solution was added at 0 °C under stirring until the mixture became clear solution (200 mL HCl in total). The layers were separated and the water layer was extracted with EtOAc (150 mL). The combined organic layers were washed with H_2O and brine, dried with $MgSO_4$, and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 2:1 → 1:2) to obtain compound **10** (3.14 g, 8.05 mmol, 87%) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.86 – 7.76 (m, 4H), 7.50 – 7.44 (m, 3H), 7.37 – 7.27 (m, 5H), 5.80 (ddd, J = 10.2, 2.9, 2.1 Hz, 1H), 5.51 (dt, J = 10.2, 2.0 Hz, 1H), 5.05 (d, J = 11.4 Hz, 1H), 4.86 (d, J = 11.7 Hz, 1H), 4.77 (dd, J = 15.1, 11.5 Hz, 2H), 4.30 – 4.19 (m, 1H), 3.77 (ddd, J = 11.5, 7.4, 4.2 Hz, 1H), 3.73 – 3.64 (m, 3H), 2.95 (d, J = 1.3 Hz, 1H), 2.52 (tq, J = 7.2, 2.7 Hz, 1H), 2.48 – 2.38 (m, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.5, 135.7, 133.4, 133.2, 128.7, 128.4, 128.1, 128.0, 127.8, 127.6, 127.5, 126.8, 126.3, 126.1, 126.0, 83.5, 80.3, 75.0, 72.9, 71.7, 65.6, 45.4 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{25}H_{26}O_4Na$ 413.1723, found 413.1719.

Compound 11



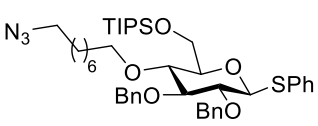
Compound **10** (781 mg, 2.00 mmol) was dissolved in dry MeCN (10 mL), K_2CO_3 (304 mg, 2.20 mmol), KI (332 mg, 2.00 mmol), 2-aminoethyl diphenylborinate (45 mg, 0.20 mmol) and BnBr (0.36 mL, 3.0 mmol) were added and the mixture was stirred for 21 hours at 60 °C. The reaction was quenched with sat. aq. $NaHCO_3$ (150 mL) and the mixture was extracted with EtOAc (2 x). The combined organic layers were washed with H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 11:1 → 7:1) to obtain compound **11** (913 mg, 1.90 mmol, 95%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.83 – 7.72 (m, 4H), 7.44 (ddt, J = 8.4, 6.8, 3.2 Hz, 3H), 7.38 – 7.19 (m, 10H), 5.75 (dt, J = 10.2, 2.4 Hz, 1H), 5.62 (dt, J = 10.2, 2.0 Hz, 1H), 5.01 (d, J = 11.3 Hz, 1H), 4.87 – 4.73 (m, 3H), 4.50 (d, J = 1.1 Hz, 2H), 4.27 – 4.19 (m, 1H), 3.76 – 3.62 (m, 2H), 3.57 (qd, J = 9.0, 5.4 Hz, 2H), 2.99 (d, J = 1.4 Hz, 1H), 2.57 – 2.50 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.7, 138.2, 135.8, 133.3, 133.0 ($5C_q$ Ar), 129.0, 128.5, 128.5, 128.5, 128.5, 128.3, 128.3, 128.0, 128.0, 127.8, 127.7, 127.7, 127.6, 127.3, 126.8, 126.648, 126.2, 126.0, 126.0, 83.9, 80.2, 75.0, 73.4, 71.7, 71.2, 71.1, 44.1 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{32}H_{32}O_4Na$ 503.2193, found 503.2189.

Compound 13



Compound **12** (2.32 g, 5.13 mmol) was dissolved in dry DCM (50 mL), imidazole (874 mg, 12.8 mmol) was added and the mixture was stirred at rt until imidazole fully dissolved. After which, the mixture was cooled to 0 °C and stirred for 1 hour. TIPSCl (1.43 mL, 6.67 mmol) was added at 0 °C and the reaction was warmed to rt slowly and stirred overnight. The mixture was diluted with DCM, washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 50:1 → 20:1) to obtain compound **13** (3.03 g, 4.98 mmol, 97%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.59 – 7.49 (m, 2H, CH Ar), 7.43 – 7.21 (m, 13H, CH Ar), 4.89 – 4.82 (m, 3H, 3CH*H* Bn), 4.71 (dd, J = 17.0, 10.0 Hz, 2H, 1CH*H* Bn and H1), 4.02 (dd, J = 10.3, 4.9 Hz, 1H, H6a), 3.94 (dd, J = 10.4, 5.5 Hz, 1H, H6b), 3.73 (td, J = 9.2, 1.7 Hz, 1H, H4), 3.58 (t, J = 8.8 Hz, 1H, H4), 3.49 – 3.44 (m, 1H, H2), 3.44 – 3.36 (m, 1H, H5), 3.05 (d, J = 1.7 Hz, 1H, OH), 1.19 – 1.02 (m, 21H, TIPS). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.7, 138.2, 134.0 ($3C_q$ Ar), 131.9, 129.0, 128.7, 128.5, 128.4, 128.1, 128.0, 128.0, 127.6 (15CH Ar), 87.7 (C1), 86.4 (C3), 80.2 (C2), 78.5 (C5), 75.7, 75.5 ($2CH_2$ Bn), 73.2 (C4), 65.1 (C6), 18.1 ($6CH_3$ TIPS), 11.9 (3CH TIPS) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{35}H_{48}O_5SSiNa$ 631.2884, found 631.2879.

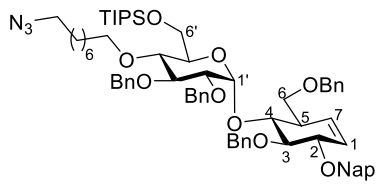
Compound 14



Compound **13** (3.03 mg, 4.98 mmol) was co-evaporated with toluene (2 x) and dissolved in dry DMF (25 mL). After cooling to 0 °C, NaH (300 mg, 7.47 mmol), 18-crown-6 (263 mg, 0.996 mmol) and 1-azido-8-iodooctane

(2.80 g, 9.96 mmol) were added successively. The mixture was stirred at 0 °C for 10 min, then warmed to rt and stirred for 7 h. TLC-analysis indicated the presence of starting material so another portion of NaH (300 mg, 7.47 mmol) and 1-azido-8-iodooctane (2.80 g, 9.96 mmol) were added at 0 °C and the mixture was stirred at rt overnight. The reaction was quenched with H_2O at 0 °C, diluted with Et_2O, washed with H_2O (3 x) and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 100:1 → 30:1) to obtain compound **14** (3.56 g, 4.67 mmol, 94%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.63 – 7.52 (m, 2H, CH Ar), 7.42 – 7.20 (m, 13H, CH Ar), 4.89 – 4.77 (m, 3H, 3CH*H* Bn), 4.66 (dd, *J* = 18.9, 10.0 Hz, 2H, 1CH*H* Bn and H1), 4.05 – 3.96 (m, 1H, H6a), 3.91 (dd, *J* = 11.1, 3.9 Hz, 1H, H6b), 3.82 (dt, *J* = 8.7, 6.5 Hz, 1H, 1CH*H*O linker), 3.69 – 3.57 (m, 2H, 1CH*H*O linker and H3), 3.46 (dt, *J* = 12.8, 9.2 Hz, 2H, H2/H4 and H5), 3.31 – 3.19 (m, 3H, H2/H4 and CH_2N_3), 1.63 – 1.49 (m, 4H, $2CH_2$ linker), 1.40 – 1.21 (m, 8H, $4CH_2$ linker), 1.21 – 1.03 (m, 21H, TIPS). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.6, 138.3, 134.2 ($3C_q$ Ar), 131.9, 128.9, 128.6, 128.5, 128.3, 128.0, 127.9, 127.8, 127.4 (15CH Ar), 87.5 (H1), 87.0 (H3), 80.7, 80.5 (H2 and H4), 77.5 (H5), 76.0, 75.5 ($2CH_2$ Bn), 73.2 (CH_2O linker), 62.5 (C6), 51.6 (CH_2N_3), 30.6, 29.5, 29.2, 28.9, 26.8, 26.3 ($6CH_2$ linker), 18.2 ($6CH_3$ TIPS), 12.1 (3CH TIPS). HRMS (ESI) m/z: $[M+NH_4]^+$ calc for $C_{43}H_{67}N_4O_5SSi$ 779.4596, found 779.4589.

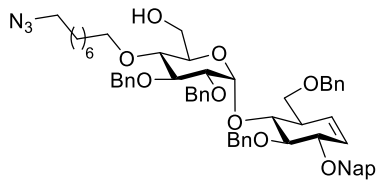
Compound 15



Donor **14** (2.08 g, 2.73 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (33 mL) under nitrogen and stirred over fresh flame-dried 3 Å molecular sieves, after which DMF (3.36 mL, 43.7 mmol, 16 equiv. of donor) was added to the solution. The solution was cooled to 0 °C, NIS (615 mg, 2.73 mmol) and TMSOTf (494 µL, 2.73 mmol) were added successively. The mixture was pre-activated at 0 °C for 1 h. Then acceptor **11** (875 mg, 1.82 mmol) was dissolved with dry DCM (3 mL in total) and added to the solution and the reaction was stirred at 0 °C for 45 h. The reaction was then quenched with a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$, stirred vigorously at rt until the brown color faded. The mixture was filtered and diluted with DCM. The layers were separated and the organic layer was washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/EtOAc 20:1 → 11:1) to afford product **15** (*dr* > 20/1, 1.08 g, 0.953 mmol, 52%) as a clean oil and with eluent Pantane/EtOAc = 3:1, a TIPS fall-off side product (here after compound **16**, 576 mg, 0.590 mmol, 32%) was also obtained as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.78 (dd, *J* = 6.1, 3.4 Hz, 1H, CH Ar), 7.75 – 7.68 (m, 3H, CH Ar), 7.48 – 7.18 (m, 19H, CH Ar), 7.15 (s, 4H, CH Ar), 5.79 (dt, *J* = 10.2, 2.5 Hz, 1H, H1/H7), 5.67 (dt, *J* = 10.1, 2.2 Hz, 1H, H1/H7), 5.57 (d, *J* = 3.6 Hz, 1H, anomeric), 5.01 – 4.92 (m, 2H), 4.86 (d, *J* = 10.8 Hz, 1H), 4.80 – 4.69 (m, 3H), 4.55 – 4.44 (m, 3H), 4.42 (d, *J* = 12.2 Hz, 1H), 4.30 (dq, *J* = 7.4, 2.4 Hz, 1H), 4.11 (t, *J* = 9.1 Hz, 1H), 3.91 (q, *J* = 9.3, 8.3 Hz, 2H), 3.85 – 3.69 (m, 4H), 3.59 (td, *J* = 9.9, 8.8, 4.9 Hz, 3H), 3.44 (q, *J* = 9.5, 8.1 Hz, 1H), 3.36 (dd, *J* = 9.8, 3.5 Hz, 1H),

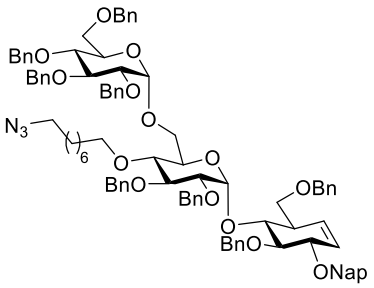
3.20 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.70 (ddp, *J* = 8.4, 5.8, 2.9 Hz, 1H, H5), 1.55 (h, *J* = 7.2 Hz, 4H, $2CH_2$ linker), 1.36 – 1.20 (m, 8H, $4CH_2$ linker), 1.09 – 0.94 (m, 21H, TIPS). ^{13}C NMR (126 MHz, $CDCl_3$) *δ* 139.5, 139.1, 138.5, 138.3, 136.0, 133.4, 133.0 ($7C_q$ Ar), 129.9, 128.4, 128.4, 128.3, 128.3, 128.2, 128.0, 127.8, 127.7, 127.6, 127.5, 127.5, 127.4, 127.0, 126.8, 126.7, 126.6, 126.1, 125.9, 96.7 (anomeric), 85.1, 81.9, 80.5, 80.1, 77.6, 75.6, 74.0, 73.5, 73.1, 72.9, 72.5, 71.9, 70.0, 62.4, 51.5 (CH_2N_3), 44.0 (C5), 30.6, 29.5, 29.3, 28.9, 26.8, 26.3 ($6CH_2$ linker), 18.1 ($3CH_3$ TIPS), 18.1 ($3CH_3$ TIPS), 12.1 (3CH TIPS) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{69}H_{89}N_3O_9SiNa$ 1154.6260, found 1154.6259.

Compound 16



Compound **15** (1.04 g, 0.918 mmol) was co-evaporated with toluene (3 x) and dissolved in dry THF (9.2 mL). TBAF (1.0 M in THF, 2.3 mL, 2.3 mmol) was added and the reaction was stirred at rt for 3 h. Then the mixture was quenched with sat. aq. NH_4Cl at 0 °C, diluted with EtOAc, washed with H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/EtOAc 7:1 → 3:1) to afford product **16** (787 mg, 0.806 mmol, 87%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.84 – 7.64 (m, 4H, CH Ar), 7.52 – 7.04 (m, 23H, CH Ar), 5.79 (dt, *J* = 10.1, 2.7 Hz, 1H, H1/H7), 5.66 (dd, *J* = 12.4, 3.1 Hz, 2H, H1/H7 and anomeric H1'), 5.03 (d, *J* = 11.8 Hz, 1H), 4.98 – 4.84 (m, 2H), 4.84 – 4.65 (m, 3H), 4.62 – 4.39 (m, 4H), 4.37 – 4.28 (m, 1H), 4.10 (t, *J* = 9.1 Hz, 1H), 4.00 – 3.86 (m, 2H), 3.81 (dt, *J* = 8.9, 6.4 Hz, 1H), 3.77 – 3.48 (m, 6H), 3.43 – 3.17 (m, 4H), 2.70 (dp, *J* = 10.8, 3.6 Hz, 1H, H5), 1.93 – 1.78 (m, 1H, OH), 1.55 (p, *J* = 6.7 Hz, 4H, $2CH_2$ linker), 1.30 (dd, *J* = 15.5, 7.8 Hz, 8H, $4CH_2$ linker). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 139.3, 138.9, 138.2, 138.1, 135.8, 133.3, 133.0 ($7C_q$ Ar), 129.7, 128.5, 128.4, 128.4, 128.3, 128.2, 128.0, 127.8, 127.8, 127.8, 127.7, 127.6, 127.6, 127.2, 126.8, 126.8, 126.7, 126.1, 125.9 (27CH Ar, C1 and C7), 96.9 (anomeric), 85.1, 81.6, 80.7, 79.5, 78.2, 75.5, 74.2, 73.4, 73.1, 73.0, 72.0, 71.7, 69.6, 61.9, 51.5 (CH_2N_3), 44.0 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 ($6CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{60}H_{69}N_3O_9Na$ 998.4926, found 998.4923.

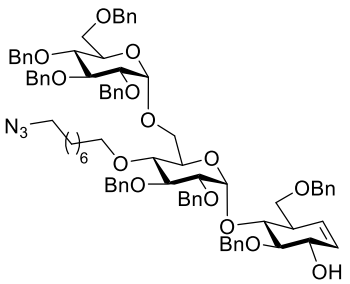
Compound 17



Imidate donor **S1** (1.49 g, 2.09 mmol) and acceptor **16** (1.36 g, 1.39 mmol) were co-evaporated with toluene (3 x), then $OPPh_3$ (2.32 g, 8.34 mmol) and activated 3 Å molecular sieves were added under nitrogen. The mixture was dissolved with dry DCM (28 mL) and stirred at rt for 30 min. Then TMSI (297 µL, 2.09 mmol) was added slowly and the reaction was stirred at rt for 43 h. The reaction was then quenched with a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$, stirred vigorously at rt until the deep red color faded. The mixture was filtered over celite and diluted with DCM. The layers were separated and the organic layer was washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and

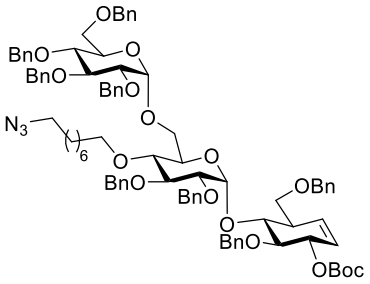
concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/EtOAc 11:1 → 6:1) to afford product **17** (*dr* > 20/1, 1.74 g, 1.16 mmol, 82%) as a light-yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.81 – 7.75 (m, 1H, CH Ar), 7.74 – 7.66 (m, 3H, CH Ar), 7.49 – 6.96 (m, 43H, CH Ar), 5.76 (dt, *J* = 10.2, 2.4 Hz, 1H, H1/H7), 5.66 (dt, *J* = 10.2, 2.0 Hz, 1H, H1/H7), 5.64 (d, *J* = 3.7 Hz, 1H, anomeric α-1,4), 5.09 (d, *J* = 3.5 Hz, 1H, anomeric α-1,6), 5.02 – 4.88 (m, 3H, 3CH*H* Bn), 4.84 (dd, *J* = 10.9, 6.0 Hz, 2H, 2CH*H* Bn), 4.78 – 4.70 (m, 3H, 3CH*H* Bn), 4.69 – 4.43 (m, 7H, 7CH*H* Bn), 4.40 – 4.34 (m, 3H, 3CH*H* Bn), 4.33 – 4.29 (m, 1H), 4.12 (t, *J* = 9.2 Hz, 1H), 3.97 – 3.90 (m, 2H), 3.90 – 3.85 (m, 1H), 3.82 (dt, *J* = 9.2, 6.7 Hz, 1H), 3.79 – 3.46 (m, 12H), 3.25 (dd, *J* = 9.8, 3.7 Hz, 1H), 3.15 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.71 (ddp, *J* = 8.5, 5.8, 2.9 Hz, 1H, H5), 1.60 – 1.45 (m, 4H, $2CH_2$ linker), 1.26 (td, *J* = 16.6, 13.7, 5.0 Hz, 8H, $4CH_2$ linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 139.4, 139.1, 139.0, 138.6, 138.5, 138.4, 138.2, 138.1, 136.0, 133.4, 133.0 ($11C_q$ Ar), 129.8, 128.5, 128.5, 128.5, 128.4, 128.3, 128.3, 128.3, 128.2, 128.1, 128.1, 128.0, 128.0, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.7, 127.6, 127.6, 127.6, 127.5, 127.5, 127.5, 127.1, 126.7, 126.7, 126.1, 126.1, 125.9, 97.3 (anomeric α-1,6), 96.8 (anomeric α-1,4), 85.4, 81.9, 80.9, 80.1, 79.9, 77.8, 77.7, 75.7, 75.5, 75.2, 74.1, 73.6, 73.2, 73.1, 73.0, 72.9, 72.0, 71.7, 70.4, 70.0, 68.6, 65.6, 51.5 (CH_2N_3), 43.9 (C5), 30.6, 29.6, 29.2, 28.9, 26.8, 26.2 ($6CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{94}H_{103}N_3O_{14}Na$ 1520.7332, found 1520.7339.

Compound 18



Compound **17** (1.74 g, 1.16 mmol) was dissolved in a mixture of DCM/H_2O (10/1, 22 mL). Then DDQ (277 mg, 1.22 mmol) was added and the reaction mixture was stirred for 2 h at rt. The mixture was diluted with DCM, washed with sat. aq. $NaHCO_3$ (3 x), H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 9:1 → 4:1) to obtain compound **18** (1.25 g) as a colourless oil contaminated with a small amount of unknown impurities which may come from dirty tubes, thus a yield over two steps is provided after the next step.

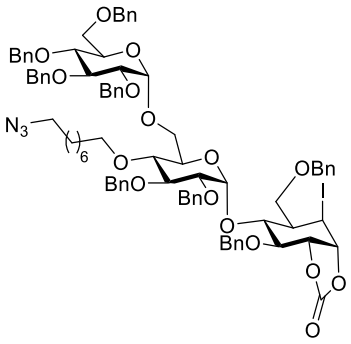
Compound 19



Compound **18** (1.25 g, 0.920 mmol) was co-evaporated with toluene (3 x) and dissolved in dry THF (20 mL). Boc_2O (3.01 g, 13.8 mmol) and DMAP (11 mg, 0.092 mmol) were added at 0 °C. Then the reaction mixture was stirred at rt for 5 h. The reaction was quenched with sat. aq. NH_4Cl at 0 °C and concentrated *in vacuo* to remove most of the THF. Then the residue was diluted with Et_2O, washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 15:1 → 7:1) and size

exclusion (eluent MeOH/DCM: 1/1) to obtain pure compound **19** (1.27 g, 0.871 mmol, 75% over two steps) as a colourless oil. *Note: size exclusion was used to separate the unknow contaminations from the previous step.* ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 – 6.98 (m, 40H, CH Ar), 5.69 (dt, J = 10.1, 2.0 Hz, 1H, H7), 5.63 (dt, J = 10.1, 2.4 Hz, 1H, H1), 5.45 (d, J = 3.6 Hz, 1H, anomeric α-1,4), 5.43 – 5.38 (m, 1H, H2), 5.05 (d, J = 3.5 Hz, 1H, anomeric α-1,6), 4.96 – 4.81 (m, 4H, 4CH*H* Bn), 4.79 – 4.71 (m, 3H, 3CH*H* Bn), 4.65 – 4.52 (m, 3H, 3CH*H* Bn), 4.51 – 4.43 (m, 3H, 3CH*H* Bn), 4.41 – 4.34 (m, 3H, 3CH*H* Bn), 4.12 (dd, J = 9.5, 8.4 Hz, 1H, H4), 3.96 – 3.78 (m, 4H), 3.77 – 3.45 (m, 12H), 3.25 (dd, J = 9.8, 3.7 Hz, 1H), 3.16 (t, J = 7.0 Hz, 2H, CH_2N_3), 2.74 – 2.65 (m, 1H, H5), 1.58 – 1.46 (m, 4H, $2CH_2$ linker), 1.41 (s, 9H, $3CH_3$ Boc), 1.32 – 1.11 (m, 8H, $4CH_2$ linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 153.2 (C=O Boc), 139.1, 139.0, 139.0, 138.6, 138.6, 138.4, 138.3, 138.2 ($8C_q$ Ar), 131.1 (C1), 128.5, 128.5, 128.4, 128.4, 128.4, 128.1, 128.3, 128.0, 128.0, 128.0, 127.9, 127.8, 127.8, 127.7, 127.6, 127.6, 127.6, 127.6, 127.5, 127.5, 127.2, 127.0 (40CH Ar), 125.3 (C7), 97.4 (anomeric α-1,6), 97.2 (anomeric α-1,4), 83.4, 82.4 (C_q Boc), 81.9, 81.9, 80.1, 79.9, 77.7, 77.7, 77.6, 75.7, 75.5, 75.2, 74.1, 73.9, 73.6, 73.3, 73.1, 72.9, 72.0, 71.8, 70.4, 69.9, 68.6, 65.7, 51.5 (CH_2N_3), 43.8 (C5), 30.6, 29.6, 29.3, 28.9 ($4CH_2$ linker), 27.8 ($3CH_3$ Boc), 26.8, 26.2 ($2CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{88}H_{103}N_3O_{16}Na$ 1480.7231, found 1480.7232.

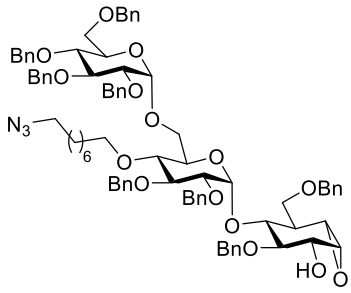
Compound 20



Compound **19** (635 mg, 0.436 mmol) was dissolved in a mixture of DCM/HOAc (0.15 M, 1/1 (v/v)), NIS (196 mg, 0.872 mmol) was added and the reaction mixture was stirred at rt for 22 h. The mixture was diluted with Et_2O and quenched with Et_3N (4.0 mL). Then the Et_2O layer was washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, sat. aq. $Na_2S_2O_3$, H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 11:1 → 6:1) to obtain compound **20** (402 mg, 0.263 mmol, 61%) as a colourless oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.42 – 7.01 (m, 40H, CH Ar), 5.32 (d, *J* = 3.7 Hz, 1H, anomeric α-1,4), 5.20 – 5.12 (m, 1H), 5.01 (d, *J* = 3.6 Hz, 1H, anomeric α-1,6), 4.93 (d, *J* = 10.9 Hz, 1H), 4.88 – 4.68 (m, 7H), 4.67 – 4.57 (m, 4H), 4.54 – 4.40 (m, 5H), 4.34 (d, *J* = 11.8 Hz, 1H), 4.08 – 3.94 (m, 3H), 3.89 (t, *J* = 8.0 Hz, 1H), 3.87 – 3.72 (m, 5H), 3.69 – 3.52 (m, 6H), 3.48 (dd, *J* = 9.3, 6.0 Hz, 1H), 3.42 (t, *J* = 9.4 Hz, 1H), 3.26 (dd, *J* = 9.9, 3.8 Hz, 1H), 3.17 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.03 (s, 1H, H5), 1.54 (ddt, *J* = 32.1, 14.4, 7.1 Hz, 4H, $2CH_2$ linker), 1.32 – 1.18 (m, 8H, $4CH_2$ linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 153.0 (C=O), 138.9, 138.8, 138.5, 138.4, 138.0, 137.9, 137.8, 137.3 ($8C_q$ Ar), 128.7, 128.5, 128.5, 128.5, 128.4, 128.2, 128.2, 128.1, 128.0, 128.0, 127.9, 127.9, 127.8, 127.8, 127.7, 127.7, 127.6, 127.6, 127.6, 127.1 (40CH Ar), 98.1 (anomeric α-1,4), 97.8 (anomeric α-1,6), 81.9, 81.6, 81.3, 80.3, 80.2, 79.4, 77.9, 77.7, 76.2, 75.6, 75.6, 75.3, 74.0, 73.6, 73.6, 73.4, 73.3, 72.4, 72.3, 71.5, 70.4, 68.7, 66.7, 51.5 (CH_2N_3), 43.9 (C5, assign by HSQC), 30.6, 29.6, 29.2, 28.9, 26.8,

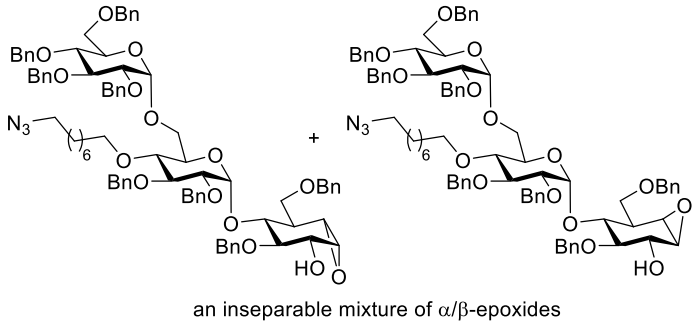
26.2 (6CH_2 linker), 25.0 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{84}H_{94}IN_3O_{16}Na$ 1550.5571, found 1550.5573.

Compound 21



Compound **20** (402 mg, 0.263 mmol) was dissolved in a mixture of DCM/MeOH (1.9 mL/3.3 mL), NaOMe (25 wt% in MeOH, 178 µL, 0.789 mmol) was added and the mixture was stirred at rt overnight. The reaction was quenched with $Et_3N{\cdot}HCl$ (131 mg, 0.95 mmol) and concentrated *in vacuo*. The residue was diluted with EtOAc, washed with H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 7:1 → 3:1) to obtain compound **21** (0.29 g, 0.21 mmol, 81%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.41 – 7.00 (m, 40H, CH Ar), 4.99 – 4.91 (m, 4H, anomeric α-1,4 and α-1,6 and 2CH*H* Bn), 4.84 (dd, *J* = 12.1, 10.8 Hz, 2H, 2CH*H* Bn), 4.76 (dd, *J* = 10.9, 6.0 Hz, 2H, 2CH*H* Bn), 4.67 – 4.54 (m, 4H, 4CH*H* Bn), 4.49 – 4.34 (m, 6H, 6CH*H* Bn), 4.02 – 3.71 (m, 10H), 3.69 – 3.60 (m, 5H), 3.59 – 3.52 (m, 2H), 3.44 – 3.37 (m, 1H), 3.33 (dd, *J* = 4.0, 2.0 Hz, 1H, epoxide), 3.28 (dd, *J* = 9.8, 3.6 Hz, 1H), 3.18 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 3.13 (dd, *J* = 4.1, 0.8 Hz, 1H, epoxide), 2.52 (dt, *J* = 7.9, 4.1 Hz, 1H, H5), 2.40 (d, *J* = 5.0 Hz, 1H, OH), 1.53 (dq, *J* = 14.4, 7.7, 7.2 Hz, 4H, 2CH_2 linker), 1.36 – 1.21 (m, 8H, 4CH_2 linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.9, 138.9, 138.9, 138.5, 138.5, 138.3, 138.2, 138.1 (8C_q Ar), 128.6, 128.5, 128.5, 128.5, 128.4, 128.4, 128.1, 128.1, 128.0, 128.0, 128.0, 127.8, 127.8, 127.8, 127.7, 127.7, 127.7, 127.7, 127.6, 127.6 (40CH Ar), 98.0 (anomeric), 97.5 (anomeric), 82.0, 81.7, 80.9, 80.1, 80.1, 78.1, 77.7, 77.4, 75.7, 75.6, 75.3, 74.8, 73.6, 73.4, 73.4, 73.1, 72.3, 71.9, 70.5, 70.4, 69.6, 68.6, 66.3, 55.5 (epoxide), 55.4 (epoxide), 51.5 (CH_2N_3), 42.4 (C5), 30.6, 29.6, 29.2, 28.9, 26.8, 26.2 (6CH_2 linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{83}H_{95}N_3O_{15}Na$ 1396.6655, found 1396.6656.

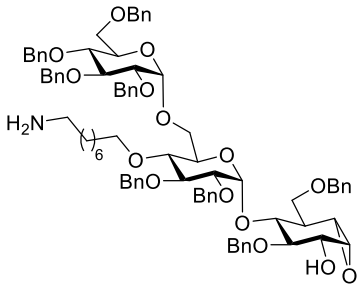
Direct epoxidation of compound 18 with *m*-CPBA



Compound **18** (84 mg, 62 µmol) was co-evaporated with toluene (3x) and dissolved in dry DCM (1.2 mL). After cooling to 0 ^{0}C, *m*-CPBA (71%, 28 mg, 0.12 mmol) was added and the mixture was stirred at 0 ^{0}C for 22 h. TLC-analysis indicated the presence of starting material so another portion of *m*-CPBA (28 mg, 0.12 mmol) was added and the mixture was stirred at 0 ^{0}C for 24 h. The reaction mixture was then diluted with EtOAc, washed with a mixture of sat. aq. $NaHCO_3$ and sat. aq. $Na_2S_2O_3$ (3x), H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude

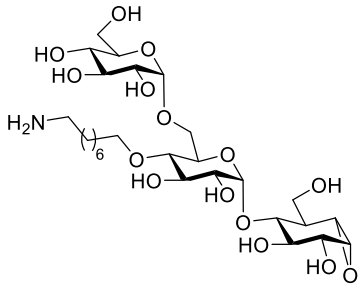
was purified with silica column chromatography (Pentane/EtOAc 10:1 → 3:1) to obtain product (66 mg, 48 µmol, 77%) as an inseparable mixture of α/β epoxides (3/1, determined by ^{1}H NMR).

Compound 22



Compound **21** (107 mg, 78.0 µmol) was dissolved in MeCN (1.6 mL). H_2O (100 µL) and polymer-bound PPh_3 (3 mmol/g loading, 130 mg, 390 µmol) were added and the mixture was stirred at 65 ^{0}C for 23 h. After which, another 150 µL H_2O was added and the mixture was stirred at 65 ^{0}C for another 2 h. The reaction was then cooled to rt, filtered and the solvent was concentrated *in vacuo*. The crude was purified with silica column chromatography (DCM/MeOH 50:1 → 19:1, with 0.1% Et_3N) to obtain product **22** (92 mg, 68 µmol, 88%) which was co-evaporated with toluene (3x) to remove Et_3N residue. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33 – 7.08 (m, 40H, CH Ar), 4.99 – 4.91 (m, 4H), 4.84 (t, *J* = 10.7 Hz, 2H), 4.76 (dd, *J* = 10.9, 4.8 Hz, 2H), 4.64 – 4.56 (m, 4H), 4.51 – 4.35 (m, 6H), 4.02 – 3.92 (m, 2H), 3.92 – 3.71 (m, 8H), 3.69 – 3.59 (m, 5H), 3.55 (dt, *J* = 9.7, 2.7 Hz, 2H), 3.40 (dd, *J* = 10.0, 9.0 Hz, 1H), 3.33 (dd, *J* = 4.1, 1.9 Hz, 1H), 3.30 – 3.16 (br s, 2H), 3.28 (dd, *J* = 9.8, 3.6 Hz, 1H), 3.12 (d, *J* = 4.1 Hz, 1H), 2.70 – 2.63 (m, 2H, CH_2NH_2), 2.52 (dt, *J* = 7.4, 4.0 Hz, 1H, H5), 2.34 (s, CH_3 toluene residue), 1.59 – 1.51 (m, 2H, CH_2 linker), 1.45 (q, *J* = 7.1 Hz, 2H, CH_2 linker), 1.33 – 1.20 (m, 8H, 4CH_2 linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.9, 138.9, 138.9, 138.5, 138.5, 138.3, 138.2, 138.0 (8C_q Ar), 129.1, 128.6, 128.5, 128.4, 128.4, 128.4, 128.3, 128.1, 128.1, 128.0, 128.0, 128.0, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.7, 127.7, 127.6, 127.6, 127.6, 127.5, 125.4 (40CH Ar), 98.0 (anomeric), 97.5 (anomeric), 82.0, 81.6, 81.0, 80.1, 80.0, 78.1, 77.7, 77.4, 75.7, 75.6, 75.2, 74.8, 73.6, 73.4, 73.1, 72.3, 71.9, 70.5, 70.3, 69.6, 68.6, 66.4, 55.6 (epoxide), 55.4 (epoxide), 42.3 (C5), 41.5 (CH_2NH_2), 32.0, 30.6, 29.6, 29.4, 26.8, 26.3 (6CH_2 linker) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{83}H_{98}NO_{15}$ 1348.6931, found 1348.6934.

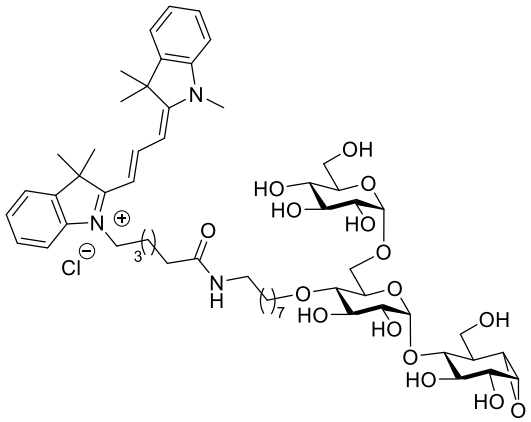
Compound 23



Compound **22** (43 mg, 32 µmol) was dissolved in dioxane (0.75 mL), then H_2O (0.5mL) was added under stirring. Upon addition of H_2O, the clear solution turned into white emulsion, so more dioxane (in total 0.75 mL) was added until the solution became clear again. The solution was purged with nitrogen for 2 minutes under vigorous stirring. TFA (2.5 µL, 32 µmol) and $Pd(OH)_2$/C (20 wt%, 22 mg, 32 µmol) were added under nitrogen atmosphere. While stirring vigorously, the N_2 balloon was replaced with a H_2 balloon and flushed with H_2 for 2 minutes. After stirring for 3 h under H_2 atmosphere, the mixture was filtered over a small celite pad and concentrated under reduced pressure. Product **23** including minor impurities was obtained as a white powder (quant) after lyophilization which can be directly used for

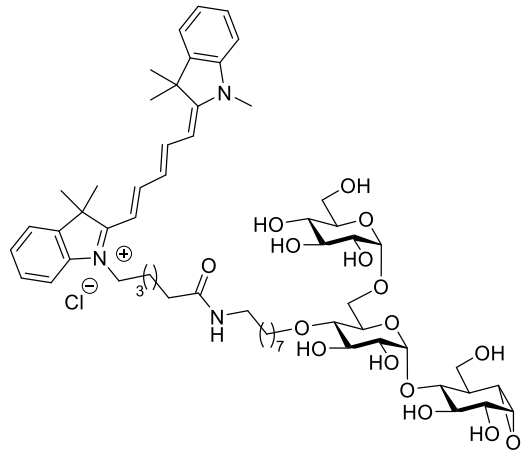
coupling with fluorescent tags. *Note: the hydrogenation reaction was always performed on small scales (25.0 – 33.0 µmol) and it usually took around 3 h for full deprotection of the benzyl groups. It's better to monitor the reaction hourly by TLC-analysis to avoid further hydrolysis or reductive opening of the epoxide moiety.* ^{1}H NMR (500 MHz, D_2O) δ 5.14 (d, J = 4.1 Hz, 1H, anomeric α-1,4), 4.95 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 3.95 (tq, J = 8.6, 3.0, 2.5 Hz, 4H), 3.91 – 3.51 (m, 12H), 3.50 – 3.42 (m, 3H), 3.38 – 3.30 (m, 2H), 3.01 – 2.94 (m, 2H, CH_2NH_2), 2.28 – 2.19 (m, 1H, H5), 1.69 – 1.53 (m, 4H, $2CH_2$ linker), 1.34 (d, J = 7.3 Hz, 8H, $4CH_2$ linker). ^{13}C NMR (126 MHz, D_2O) δ 100.7 (anomeric α-1,4), 98.9 (anomeric α-1,6), 80.4, 77.7, 73.4, 73.2, 73.1, 72.9, 72.0, 71.8, 71.6, 70.7, 70.4, 69.4, 66.8, 60.4, 57.3 (epoxide), 55.2 (epoxide), 42.8 (C5), 39.6 (CH_2NH_2), 29.3, 28.4, 28.1, 26.7, 25.5, 25.2 ($6CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{27}H_{50}NO_{15}$ 628.3175, found 628.3172.

Compound 24



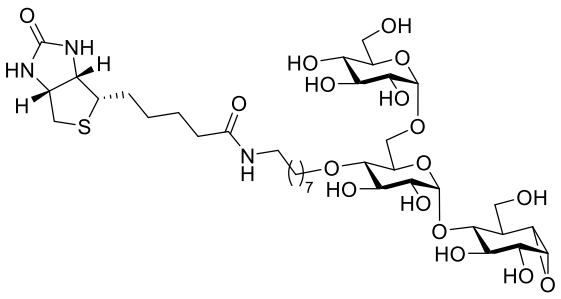
Amine **23** (14.0 mg, 22.3 µmol) was reacted with Cy3 acid (1.05 eq) for 22 h according to general procedure A. This afforded product **24** (9.9 mg, 9.0 µmol, 40%) as a red powder after lyophilization (3x). ^{1}H NMR (500 MHz, MeOD) δ 8.55 (t, J = 13.5 Hz, 1H), 7.55 (ddd, J = 7.5, 1.2, 0.6 Hz, 2H), 7.46 (dddd, J = 8.0, 7.5, 3.9, 1.2 Hz, 2H), 7.40 – 7.28 (m, 4H), 6.44 (dd, J = 13.5, 6.3 Hz, 2H), 4.94 (1H, anomeric α-1,4, obscured by H_2O peak), 4.82 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 4.16 (t, J = 7.5 Hz, 2H, CH_2N^+), 4.06 – 4.02 (m, 1H), 3.96 – 3.63 (m, 14H), 3.59 (dd, J = 9.8, 8.7 Hz, 1H), 3.52 (dt, J = 9.0, 6.6 Hz, 1H), 3.43 (ddd, J = 16.3, 9.7, 3.8 Hz, 2H), 3.34 (dd, J = 9.7, 8.9 Hz, 1H), 3.28 – 3.18 (m, 3H), 3.12 (t, J = 7.1 Hz, 2H, $NHCH_2$), 3.09 – 3.02 (m, 1H), 2.21 (t, J = 7.3 Hz, 2H, CH_2C=O), 2.10 – 2.04 (m, 1H, H5), 1.86 (p, J = 7.7 Hz, 2H), 1.77 (s, 12H), 1.71 (p, J = 7.4 Hz, 2H), 1.59 – 1.40 (m, 6H), 1.40 – 1.21 (m, 8H). ^{13}C NMR (126 MHz, MeOD) δ 176.7, 176.0, 175.7, 152.1, 144.1, 143.4, 142.2, 142.1, 130.0, 130.0, 126.8, 126.8, 123.6, 123.4, 112.5, 112.3, 103.8, 103.7, 103.1 (anomeric α-1,4), 100.5 (anomeric α-1,6), 83.7, 80.0, 75.3, 75.0, 74.8, 74.0, 74.0, 73.7, 73.5, 72.6, 72.3, 71.7, 68.8, 62.6, 61.8, 58.0 (epoxide), 55.3 (epoxide), 50.7 (C_q), 50.6 (C_q), 45.2 (C5), 45.1 (CH_2N^+), 40.4 ($NHCH_2$), 36.7 (CH_2C=O), 31.8 (NCH_3), 31.4, 30.5, 30.4, 30.3 ($4CH_2$ linker), 28.3 ($2CH_3$), 28.2 (CH_2 linker), 28.2 ($2CH_3$), 27.9, 27.3, 27.1, 26.6 ($4CH_2$ linker) ppm. *Note: peaks at 23.3 and 179.0 ppm in ^{13}C NMR are belong to NH_4OAc residue which can be removed after repeated lyophilization.* HRMS (ESI) m/z: $[M]^+$ calc for $C_{57}H_{84}N_3O_{16}$ 1066.5846, found 1066.5844.

Compound 25



Amine **23** (20.0 mg, 31.8 µmol) was reacted with Cy5 acid (1.05 eq) for 28 h according to general procedure A. This afforded product **25** (19.8 mg, 17.5 µmol, 55%) as a blue powder after lyophilization (3x). ^{1}H NMR (500 MHz, MeOD) δ 8.28 – 8.19 (m, 2H), 7.49 (dd, J = 7.5, 1.3 Hz, 2H), 7.41 (tdd, J = 7.7, 3.3, 1.2 Hz, 2H), 7.33 – 7.23 (m, 4H), 6.64 (t, J = 12.4 Hz, 1H), 6.28 (dd, J = 13.7, 4.0 Hz, 2H), 4.94 (1H, anomeric α-1,4, obscured by H_2O peak), 4.82 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 4.10 (d, J = 7.4 Hz, 2H, CH_2N^+), 4.04 (dd, J = 10.9, 3.0 Hz, 1H), 3.99 – 3.55 (m, 15H), 3.52 (dt, J = 9.0, 6.7 Hz, 1H), 3.48 – 3.39 (m, 2H), 3.38 – 3.32 (m, 1H), 3.30 – 3.19 (m, 3H), 3.12 (t, J = 7.2 Hz, 2H, $NHCH_2$), 3.06 (t, J = 9.5 Hz, 1H), 2.20 (t, J = 7.3 Hz, 2H, CH_2C=O), 2.08 (ddd, J = 9.7, 6.9, 3.1 Hz, 1H, H5), 1.86 – 1.78 (m, 2H), 1.72 (s, 14H), 1.59 – 1.50 (m, 2H), 1.50 – 1.40 (m, 4H), 1.40 – 1.24 (m, 8H). ^{13}C NMR (126 MHz, MeOD) δ 175.7, 175.3, 174.6, 155.5, 155.4, 144.2, 143.5, 142.6, 142.5, 129.8, 129.7, 126.6, 126.3, 126.2, 123.4, 123.3, 112.0, 111.9, 104.4, 104.3, 103.1 (anomeric α-1,4), 100.5 (anomeric α-1,6), 83.6, 80.0, 75.3, 74.9, 74.9, 74.8, 74.0, 73.9, 73.7, 73.6, 73.6, 73.5, 72.6, 72.2, 71.6, 68.7, 62.5, 61.8, 58.0 (epoxide), 55.3 (epoxide), 50.5 (C_q), 50.5 (C_q), 45.1 (C5), 44.8 (CH_2N^+), 40.4 ($NHCH_2$), 36.7 (CH_2C=O), 31.6 (NCH_3), 31.4, 30.5, 30.4, 30.3 (4CH_2 linker), 28.2 (CH_2 linker), 28.0 (2CH_3), 27.9 (CH_2 linker), 27.8 (2CH_3), 27.4, 27.1, 26.6 (3CH_2 linker) ppm. HRMS (ESI) m/z: $[M]^+$ calc for $C_{59}H_{86}N_3O_{16}$ 1092.6003, found 1092.6000.

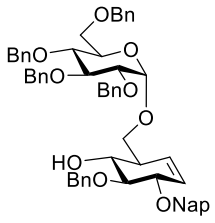
Compound 26



Amine **23** (13.4 mg, 21.4 µmol) was reacted with biotin (1.2 eq) for 25 h according to general procedure A. This afforded product **26** (7.1 mg, 8.3 µmol, 38%) as a white powder after lyophilization (3x). ^{1}H NMR (500 MHz, MeOD) δ 4.95 (d, J = 3.9 Hz, 1H, anomeric α-1,4), 4.83 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 4.50 (ddd, J = 7.9, 5.0, 0.9 Hz, 1H, NHC*H* biotin), 4.31 (dd, J = 7.9, 4.5 Hz, 1H, NHC*H* biotin), 4.04 (dd, J = 10.9, 3.0 Hz, 1H), 3.96 – 3.83 (m, 3H), 3.83 – 3.63 (m, 8H), 3.59 (dd, J = 9.8, 8.7 Hz, 1H), 3.56 – 3.50 (m, 1H), 3.43 (ddd, J = 17.0, 9.7, 3.8 Hz, 2H), 3.35 (dd, J = 9.7, 8.8 Hz, 1H), 3.29 – 3.25 (m, 2H), 3.24 – 3.12 (m, 4H), 3.07 (dd, J = 10.0, 8.9 Hz, 1H), 2.93 (dd, J = 12.8, 5.0 Hz, 1H, SC*H*H), 2.71 (d, J = 12.7 Hz, 1H, SC*H*H), 2.20 (t, J = 7.3 Hz, 2H, CH_2C=O), 2.13 – 2.04 (m, 1H, H5), 1.77 – 1.41 (m, 10H), 1.38 – 1.28 (m, 8H), 1.22 (s, CH_3, ^{t}BuOH residue). ^{13}C NMR (126 MHz, MeOD) δ 176.0 (NHCONH), 166.1 (CONH), 103.1 (anomeric α-1,4), 100.5 (anomeric α-1,6), 83.6, 80.0, 75.3, 75.0, 74.8, 74.0, 73.7, 73.5, 72.6, 72.3, 71.7, 68.8, 63.4,

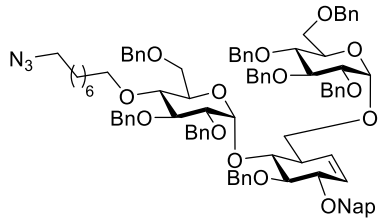
62.5, 61.8, 61.6, 58.0 (epoxide), 57.0 (SCH), 55.3 (epoxide), 45.1 (C5), 41.1 (SCH_2), 40.4 ($NHCH_2$), 36.8 ($CH_2C=O$), 31.4 (CH_2 linker), 31.1 (CH_3, ^{t}BuOH residue), 30.5, 30.4, 30.4, 29.8, 29.5, 27.9, 27.1, 27.0 ($8CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{37}H_{64}N_3O_{17}S$ 854.3951, found 854.3950.

Compound 27



The imidate donor **S1** (1.5 g, 2.1 mmol) and cyclohexane acceptor **10** (0.78 g, 2.0 mmol) were co-evaporated with toluene (3 x), then $OPPh_3$ (3.34 g, 12.0 mmol) and activated 3 Å molecular sieves were added under nitrogen. The mixture was dissolved with dry DCM (40 mL) and stirred at rt for 30 min. Then TMSI (342 µL, 2.4 mmol) was added slowly and the reaction was stirred at rt for 24 h. The reaction was then quenched with a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$, stirred vigorously at rt until the deep red color faded. The mixture was filtered over celite and diluted with DCM. The layers were separated and the organic layer was washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was first purified by silica gel column chromatography (Pentane/EtOAc 9:1 → 3:1) to give an inseparable mixture of product and hydrolyzed donor. Then the mixture was further purified with size exclusion (eluent MeOH/DCM: 1/1) to afford product **27** as a mixture of two isomers (α-1,6 linkage/β-1,6 linkage ≈ 14/1, 1.2 g, 1.3 mmol, 65%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 – 7.74 (m, 4H, CH Ar), 7.50 – 7.41 (m, 3H, CH Ar), 7.41 – 7.22 (m, 23H, CH Ar), 7.19 – 7.09 (m, 2H, CH Ar), 5.77 – 5.71 (m, 1H, H7), 5.59 (dt, *J* = 10.2, 2.0 Hz, 1H, H1), 5.01 – 4.71 (m, 9H, 8CH*H* Bn and H1'), 4.64 (d, *J* = 12.0 Hz, 1H, 1CH*H* Bn), 4.58 (d, *J* = 12.1 Hz, 1H, 1CH*H* Bn), 4.46 (dd, *J* = 11.5, 5.9 Hz, 2H, 2CH*H* Bn), 4.27 – 4.20 (m, 1H, H2), 3.94 (t, *J* = 9.3 Hz, 1H, H3'), 3.80 (dd, *J* = 9.5, 6.1 Hz, 1H, H6a), 3.76 – 3.66 (m, 4H, H3, H4, H5' and H6'a), 3.65 – 3.59 (m, 2H, H6'b and H4'), 3.56 (dd, *J* = 9.6, 3.7 Hz, 1H, H2'), 3.45 (dd, *J* = 9.5, 6.4 Hz, 1H, H6b), 3.28 (d, *J* = 1.2 Hz, 1H, OH), 2.66 – 2.58 (m, 1H, H5). ^{13}C NMR (126 MHz, $CDCl_3$) δ 139.0, 138.9, 138.3, 138.3, 138.0, 136.0, 133.4, 133.1 ($8C_q$ Ar), 128.6, 128.6, 128.5, 128.3, 128.1, 128.1, 128.0, 128.0, 128.0, 127.8, 127.7, 127.7, 127.5, 127.4, 126.7, 126.2, 126.0, 126.0 (32CH Ar, C1 and C7), 97.6 (C1'), 83.9 (C3), 82.2 (C3'), 80.1, 80.0 (C2 and C2'), 77.7 (C4'), 75.8, 75.2, 75.2, 73.6, 73.3 ($5CH_2$ Bn/Nap), 72.6 (C4), 72.0 (CH_2 Bn/Nap), 70.5 (C5'), 70.3 (C6), 68.5 (C6'), 43.9 (C5) ppm. HRMS (ESI) m/z: $[M+NH_4]^+$ calc for $C_{59}H_{64}O_9N$ 930.4576, found 930.4569.

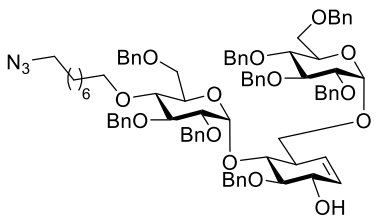
Compound 28



Donor **S2** (1.37 g, 1.97 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (22.0 mL) under nitrogen and stirred over fresh flame-dried 3 Å molecular sieves, after which DMF (2.4 mL, 31.5 mmol, 16.0 eq. of donor) was added to the solution. The solution was cooled to 0 °C, NIS (442 mg, 1.97 mmol) and TMSOTf (380 µL, 1.97 mmol) were added successively. The mixture was pre-activated at 0 °C for 1 h. Then

acceptor **27** (1.20 g, 1.31 mmol) was dissolved with dry DCM (4 mL in total) and added to the solution and the reaction was stirred at 0 °C for 61 h. The reaction was then quenched with a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$, stirred vigorously at rt until the brown color faded. The mixture was filtered over celite and diluted with DCM. The layers were separated and the organic layer was washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was first purified by silica gel column chromatography (Pentane/EtOAc 11:1 → 5:1), then the impure fractions were collected and further purified with size exclusion (eluent MeOH/DCM: 1/1) to afford product **28** (*dr* > 20/1, 1.55 g, 1.03 mmol, 78%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.80 – 7.73 (m, 1H, CH Ar), 7.70 (d, *J* = 8.4 Hz, 1H, CH Ar), 7.68 – 7.63 (m, 2H, CH Ar), 7.46 – 7.38 (m, 2H, CH Ar), 7.36 – 7.06 (m, 41H, CH Ar), 5.80 (ddd, *J* = 10.2, 2.7, 1.8 Hz, 1H), 5.75 (dt, *J* = 10.2, 2.1 Hz, 1H), 5.61 (d, *J* = 3.7 Hz, 1H, anomeric α-1,4), 4.98 – 4.62 (m, 12H, 11CH*H* Bn and anomeric α-1,6), 4.57 (dd, *J* = 15.4, 12.1 Hz, 2H), 4.51 – 4.45 (m, 3H), 4.38 (dd, *J* = 20.7, 12.1 Hz, 2H), 4.20 (ddt, *J* = 6.7, 4.4, 2.0 Hz, 1H), 4.02 – 3.85 (m, 4H), 3.82 – 3.63 (m, 9H), 3.62 – 3.52 (m, 2H), 3.49 – 3.33 (m, 3H), 3.19 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.83 (tdq, *J* = 7.4, 4.6, 2.5 Hz, 1H, H5), 1.51 (p, *J* = 7.1 Hz, 2H, CH_2 linker), 1.40 (p, *J* = 6.8 Hz, 2H, CH_2 linker), 1.30 – 1.10 (m, 8H, 4CH_2 linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 139.4, 139.0, 138.9, 138.5, 138.4, 138.4, 138.2, 138.1, 135.9, 133.4, 133.0 (11C_q Ar), 128.6, 128.5, 128.5, 128.5, 128.5, 128.4, 128.4, 128.3, 128.2, 128.2, 128.1, 128.0, 127.9, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.7, 127.7, 127.6, 127.5, 127.4, 127.1, 126.8, 126.6, 126.1, 126.0, 125.9 (47CH Ar and C1 and C7), 97.7 (anomeric α-1,4), 97.1 (anomeric α-1,6), 84.0, 82.1, 81.9, 80.1, 80.0, 79.5, 77.9, 77.7, 75.8, 75.6, 75.5, 75.2, 74.0, 73.6, 73.5, 73.3, 72.9, 72.8, 71.9, 71.5, 70.6, 68.8, 68.6, 68.5, 51.5 (CH_2N_3), 43.5 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.1 (6CH_2 linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{94}H_{103}N_3O_{14}Na$ 1520.7332, found 1520.7339.

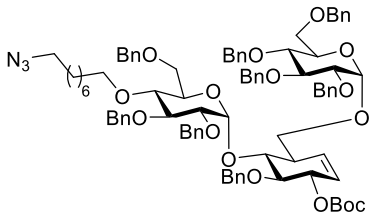
Compound 29



Compound **28** (1.52 g, 1.01 mmol) was dissolved in a mixture of DCM/H_2O (10/1, 20 mL). Then DDQ (252 mg, 1.11 mmol) was added and the reaction mixture was stirred for 1.5 h at rt. The reaction was diluted with DCM, washed with sat. aq. $NaHCO_3$ (3 x), H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 9:1 → 4:1) to obtain compound **29** (1.05 g, 0.773 mmol, 76%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 – 7.09 (m, 40H, CH Ar), 5.98 (ddt, *J* = 10.3, 4.6, 1.5 Hz, 1H), 5.81 (ddt, *J* = 10.3, 4.6, 1.1 Hz, 1H), 4.96 (d, *J* = 10.7 Hz, 1H), 4.87 – 4.74 (m, 6H), 4.66 – 4.52 (m, 7H), 4.49 – 4.39 (m, 3H), 4.37 – 4.32 (m, 1H), 4.20 (d, *J* = 12.0 Hz, 1H), 4.07 (dd, *J* = 11.6, 4.5 Hz, 1H), 3.95 (t, *J* = 9.2 Hz, 1H), 3.88 – 3.77 (m, 2H), 3.76 – 3.61 (m, 9H), 3.59 – 3.42 (m, 5H), 3.37 (dt, *J* = 9.0, 6.8 Hz, 1H), 3.19 (t, *J* = 7.0 Hz, 2H, CH_2N_3), 2.67 (q, *J* = 7.1, 6.7 Hz, 1H, H5), 1.51 (p, *J* = 6.9 Hz, 2H, CH_2 linker), 1.46 – 1.37 (m, 2H, CH_2 linker), 1.29 – 1.09 (m, 8H, 4CH_2 linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.9, 138.8, 138.5, 138.4, 138.1, 138.1, 138.0,

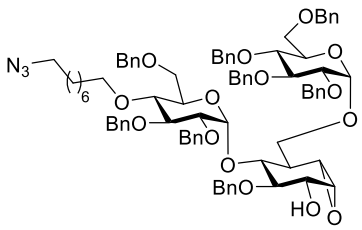
137.7 ($8C_q$ Ar), 128.7, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.1, 128.0, 128.0, 127.9, 127.9, 127.9, 127.8, 127.8, 127.7, 127.7, 127.7, 127.6, 127.6, 127.4, 126.5 (40CH Ar and C1 and C7), 97.1 (anomeric), 95.9 (anomeric), 82.4, 82.1, 79.9, 78.6, 77.8, 77.5, 76.3, 75.9, 75.6, 75.2, 73.9, 73.6, 73.4, 73.3, 72.6, 72.1, 71.4, 70.3, 70.2, 69.2, 68.5, 68.4, 65.6, 51.5 (CH_2N_3), 41.7 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 ($6CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{83}H_{95}N_3O_{14}Na$ 1380.6706, found 1380.6706.

Compound 30



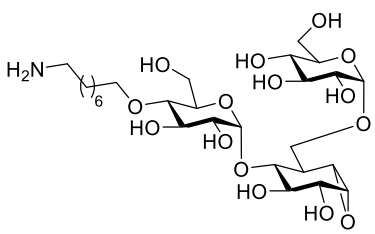
Compound **29** (1.04 g, 0.766 mmol) was co-evaporated with toluene (3 x) and dissolved in dry THF (15.3 mL) and cooled to 0 °C. Boc_2O (2.51 g, 11.5 mmol) and DMAP (9.3 mg, 76.6 µmol) were added at 0 °C. Then the reaction mixture was stirred at rt for 4 h. The reaction was quenched with sat. aq. NH_4Cl at 0 °C and concentrated *in vacuo* to remove most of the THF. Then the residue was diluted with Et_2O, washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 15:1 → 6:1) to obtain compound **30** (944 mg, 0.648 mmol, 85%) as a clean oil. *Note: the starting material can react with the product by attacking the carbonyl group of Boc to form a dimer side product, thus far more excessive Boc_2O was needed to compete the side reaction.* ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.39 – 7.08 (m, 40H, CH Ar), 5.86 (ddd, J = 10.2, 3.1, 2.0 Hz, 1H), 5.65 (dt, J = 10.3, 2.4 Hz, 1H), 5.44 (d, J = 3.7 Hz, 1H, anomeric α-1,4), 5.34 (dq, J = 6.9, 2.4 Hz, 1H, H2), 4.95 (d, J = 10.8 Hz, 1H, 1CH*H* Bn), 4.91 – 4.86 (m, 2H, 2CH*H* Bn), 4.85 – 4.71 (m, 5H, anomeric α-1,6 and 4CH*H* Bn), 4.64 (d, J = 2.1 Hz, 2H, 2CH*H* Bn), 4.60 – 4.50 (m, 4H, 4CH*H* Bn), 4.47 (d, J = 10.8 Hz, 1H, 1CH*H* Bn), 4.38 (dd, J = 25.5, 12.1 Hz, 2H, 2CH*H* Bn), 4.01 – 3.85 (m, 4H), 3.81 – 3.61 (m, 9H), 3.57 (ddd, J = 19.4, 10.0, 2.3 Hz, 2H), 3.49 – 3.34 (m, 3H), 3.19 (t, J = 7.0 Hz, 2H, CH_2N_3), 2.83 (dq, J = 7.8, 2.7 Hz, 1H, H5), 1.56 – 1.47 (m, 2H, CH_2 linker), 1.40 (s, 11H, CH_2 linker and $3CH_3$ Boc), 1.31 – 1.09 (m, 8H, $4CH_2$ linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 153.2 (C=O, Boc), 139.0, 138.9, 138.9, 138.4, 138.3, 138.3, 138.2, 138.0 ($8C_q$ Ar), 129.8 (C1/C7), 128.5, 128.4, 128.4, 128.4, 128.3, 128.2, 128.0, 128.0, 127.8, 127.8, 127.7, 127.7, 127.6, 127.6, 127.5, 127.5, 127.2, 127.0 (40CH Ar), 125.3 (C1/C7), 97.8 (anomeric α-1,4), 97.1 (anomeric α-1,6), 82.3 (C_q, Boc), 82.1, 81.9, 81.8, 79.9, 79.4, 77.8, 77.6, 76.3, 75.8, 75.7, 75.5, 75.2, 74.0, 73.6, 73.4, 73.2, 72.8, 71.5, 70.6, 68.7, 68.6, 68.4, 51.5 (CH_2N_3), 43.1 (C5), 30.5, 29.5, 29.2, 28.9 ($4CH_2$ linker), 27.8 ($3CH_3$ Boc), 26.8, 26.1 ($2CH_2$ linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{88}H_{103}N_3O_{16}Na$ 1480.7231, found 1480.7229.

Compound 31



Compound **30** (936 mg, 0.642 mmol) was dissolved in a mixture of DCM/HOAc (0.2 M, 2/1 (v/v)), NIS (288 mg, 1.28 mmol) was added and the reaction was stirred at rt for 22 h. The mixture was diluted with Et_2O and quenched with Et_3N (2.6 mL). Then the Et_2O layer was washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, sat. aq. $Na_2S_2O_3$, H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. As the conversion was not complete, the crude (896 mg, yellow oil) was obtained as a mixture of an iodocarbonate intermediate and starting material **30** (ratio ≈ 1.5:1, determined by ^{1}H NMR) and it's difficult to separate them by standard column chromatography. Therefore the crude was directly dissolved in a mixture of DCM/MeOH (2.7 mL/4.3 mL), NaOMe (25 wt% in MeOH, 0.24 mL, 1.05 mmol) was added and the mixture was stirred at rt overnight. The reaction was quenched with $Et_3N{\cdot}HCl$ (173 mg, 1.26 mmol) and concentrated *in vacuo*. The residue was diluted with EtOAc, washed with H_2O and brine, dried over Na_2SO_4, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 12:1 → 3:1) to obtain compound **31** (404 mg, 0.294 mmol, 46% over two steps) as a clean oil, and unreacted starting material **30** (292 mg, 0.20 mmol) from the first step was recovered as well. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.52 – 6.95 (m, 40H, CH Ar), 4.97 – 4.74 (m, 8H), 4.68 (d, J = 11.9 Hz, 1H), 4.62 – 4.52 (m, 6H), 4.49 (d, J = 10.8 Hz, 1H), 4.39 (dd, J = 20.7, 12.0 Hz, 2H), 3.99 (ddd, J = 10.0, 5.7, 2.4 Hz, 2H), 3.90 (td, J = 9.3, 7.5 Hz, 2H), 3.82 (ddd, J = 10.1, 3.7, 2.3 Hz, 1H), 3.79 – 3.52 (m, 10H), 3.48 – 3.34 (m, 4H), 3.29 (dd, J = 4.1, 2.3 Hz, 1H, epoxide), 3.20 (t, J = 7.0 Hz, 2H, CH_2N_3), 3.17 (dd, J = 4.1, 1.0 Hz, 1H, epoxide), 2.68 (d, J = 5.8 Hz, 1H, OH), 2.62 (q, J = 4.8 Hz, 1H, H5), 1.54 (p, J = 7.0 Hz, 2H, CH_2 linker), 1.44 (p, J = 6.7 Hz, 2H, CH_2 linker), 1.36 – 1.10 (m, 8H, 4CH_2 linker). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.8, 138.7, 138.7, 138.3, 138.3, 138.1, 138.1, 138.0 (8C_q Ar), 128.6, 128.5, 128.4, 128.4, 128.4, 128.3, 128.0, 128.0, 128.0, 127.9, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.7, 127.6 (40CH Ar), 98.6 (anomeric α-1,4), 97.7 (anomeric α-1,6), 82.0, 81.6, 80.0, 79.5, 79.3, 79.0, 78.0, 77.6, 75.7, 75.6, 75.2, 74.4, 73.6, 73.4, 73.3, 73.3, 73.0, 71.5, 70.9, 69.1, 68.8, 68.5, 68.4, 54.9 (epoxide), 54.8 (epoxide), 51.5 (CH_2N_3), 41.6 (C5), 30.4, 29.4, 29.1, 28.9, 26.7, 26.1 (6CH_2 linker) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{83}H_{95}N_3O_{15}Na$ 1396.6655, found 1396.6655.

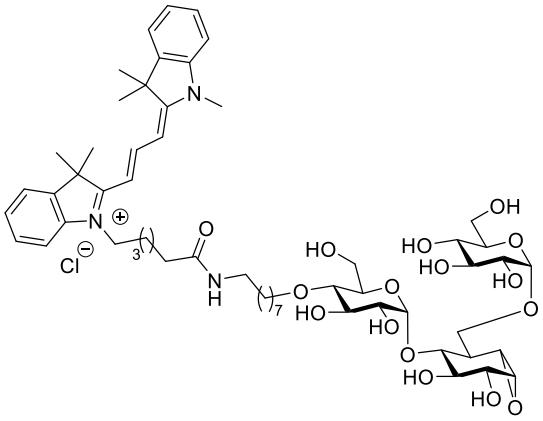
Compound 32



Ammonium (6.0 mL) was condensed in a dry flask at -60 ^{0}C. Sodium (81 mg, 3.5 mmol) was added and the resulting deep-blue solution was stirred for 15 min to dissolve all sodium. Compound **31** (60 mg, 44 μmol) and ^{t}BuOH (340 μL, 3.5 mmol) were taken up in anhydrous THF (1.5 mL) and slowly added into the reaction mixture under a flow of argon. After stirring for 1 h, the reaction was carefully quenched with aq. HOAc (225 μL HOAc in 0.5 mL H_2O, 3.92 mmol), slowly warmed to rt and stirred for 1 h. The reaction mixture was concentrated

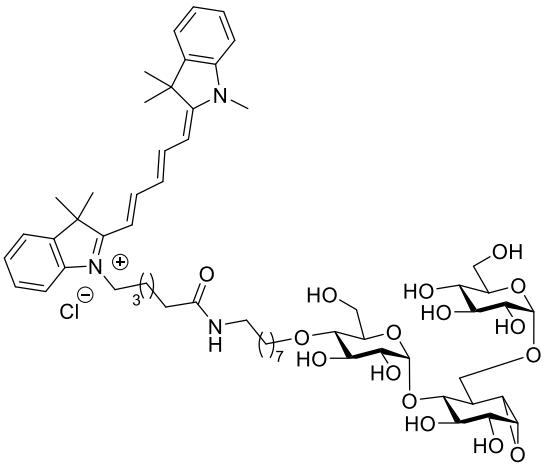
and desalted by size exclusion over HW-40 (1% HOAc in water). The compound **32** was obtained as a white powder (16.1 mg, 25.5 µmol, 57%) after lyophilization. ^{1}H NMR (500 MHz, D_2O) δ 5.10 (d, J = 4.0 Hz, 1H, anomeric α-1,4), 4.95 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 3.98 (ddd, J = 16.1, 9.4, 3.9 Hz, 2H), 3.89 – 3.62 (m, 12H), 3.60 – 3.52 (m, 2H), 3.51 – 3.40 (m, 4H), 3.28 (t, J = 9.5 Hz, 1H), 2.98 (td, J = 7.5, 1.2 Hz, 2H, CH_2NH$_2$), 2.41 (ddd, J = 9.4, 6.2, 2.9 Hz, 1H, H5), 1.62 (dp, J = 29.4, 6.9 Hz, 4H, 2CH$_2$ linker), 1.35 (dt, J = 10.3, 5.6 Hz, 8H, 4CH$_2$ linker). ^{13}C NMR (126 MHz, D_2O) δ 100.9 (anomeric α-1,4), 98.8 (anomeric α-1,6), 81.1, 77.6, 73.3 (CH$_2$O linker), 73.3, 72.9, 72.2, 71.9, 71.6, 71.4, 70.6, 69.5, 67.2 (C6), 60.4, 60.3 (C6' and C6''), 57.4 (epoxide), 55.2 (epoxide), 41.1 (C5), 39.6 (CH$_2$NH$_2$), 29.2, 28.3, 28.1, 26.8, 25.5, 25.1 (6CH$_2$ linker) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{27}H_{50}NO_{15}$ 628.3175, found 628.3178.

Compound 33



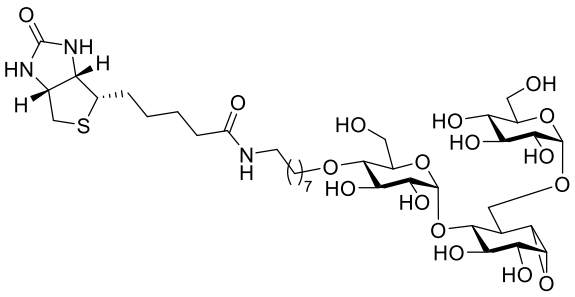
Amine **32** (12.3 mg, 19.6 µmol) was reacted with Cy3 acid (1.05 eq) for 24 h according to general procedure A. This afforded product **33** (6.1 mg, 5.5 µmol, 28%) as a red powder after lyophilization (2x). ^{1}H NMR (500 MHz, MeOD) δ 8.55 (t, J = 13.5 Hz, 1H), 7.60 – 7.52 (m, 2H), 7.46 (tdd, J = 7.6, 4.1, 1.2 Hz, 2H), 7.39 – 7.28 (m, 4H), 6.45 (d, J = 6.1 Hz, 1H), 6.43 (d, J = 6.0 Hz, 1H), 4.99 (d, J = 3.9 Hz, 1H, anomeric α-1,4), 4.83 (1H, anomeric α-1,6, obscured by H_2O peak), 4.16 (t, J = 7.5 Hz, 2H, CH$_2$N$^+$), 4.01 (dd, J = 9.9, 6.4 Hz, 1H), 3.88 – 3.74 (m, 5H), 3.73 – 3.52 (m, 11H), 3.42 (ddd, J = 12.7, 9.8, 3.8 Hz, 2H), 3.36 – 3.32 (m, 1H), 3.29 – 3.26 (m, 3H), 3.15 – 3.07 (m, 3H), 2.28 – 2.16 (m, 3H, H5 and CH$_2$C=O), 1.92 – 1.80 (m, 2H), 1.77 (s, 12H, 4CH$_3$), 1.71 (p, J = 7.4 Hz, 2H), 1.60 – 1.41 (m, 6H), 1.39 – 1.26 (m, 8H). ^{13}C NMR (126 MHz, MeOD) δ 176.7, 176.0, 175.7, 152.1, 144.1, 143.4, 142.2, 142.1, 130.0, 130.0, 126.8, 126.8, 123.6, 123.4, 112.5, 112.3, 103.8, 103.7, 103.3 (anomeric α-1,4), 100.8 (anomeric α-1,6), 84.0, 79.4, 75.3, 74.8, 74.8, 74.2, 74.0, 73.8, 73.5, 72.6, 71.9, 68.7, 62.8, 62.4, 58.1 (epoxide), 55.4 (epoxide), 50.7 (C$_q$), 50.6 (C$_q$), 45.1 (CH$_2$N$^+$), 43.3 (C5), 40.4 (NHCH$_2$), 36.7 (*C*H$_2$C=O), 31.8 (NCH$_3$), 31.4, 30.5, 30.4, 30.3 (4CH$_2$ linker), 28.3 (2CH$_3$), 28.2 (CH$_2$ linker), 28.2 (2CH$_3$), 27.9, 27.3, 27.1, 26.6 (4CH$_2$ linker) ppm. HRMS (ESI) m/z: $[M]^+$ calc for $C_{57}H_{84}N_3O_{16}$ 1066.5846, found 1066.5849.

Compound 34



Amine **32** (13.8 mg, 22.0 µmol) was reacted with Cy5 acid (1.05 eq) for 17 h according to general procedure A. This afforded the product **34** (13.2 mg, 11.7 µmol, 53%) as a blue powder after lyophilization (3x). ^{1}H NMR (500 MHz, MeOD) δ 8.29 – 8.20 (m, 2H), 7.52 – 7.47 (m, 2H), 7.42 (tdd, J = 7.7, 3.6, 1.2 Hz, 2H), 7.33 – 7.23 (m, 4H), 6.63 (t, J = 12.4 Hz, 1H), 6.28 (dd, J = 13.7, 4.6 Hz, 2H), 4.99 (d, J = 3.9 Hz, 1H, anomeric α-1,4), 4.83 (1H, anomeric α-1,6, obscured by H_2O peak), 4.11 (t, J = 7.4 Hz, 2H, CH_2N^+), 4.01 (dd, J = 9.9, 6.4 Hz, 1H), 3.89 – 3.74 (m, 5H), 3.72 – 3.50 (m, 11H), 3.42 (ddd, J = 12.3, 9.7, 3.8 Hz, 2H), 3.34 (d, J = 8.9 Hz, 1H), 3.31 – 3.25 (m, 3H), 3.15 – 3.07 (m, 3H), 2.27 – 2.17 (m, 3H, H5 and CH_2C=O), 1.87 – 1.79 (m, 2H), 1.73 (s, 14H), 1.59 – 1.50 (m, 2H), 1.50 – 1.39 (m, 4H), 1.39 – 1.25 (m, 8H). ^{13}C NMR (126 MHz, MeOD) δ 175.7, 175.4, 174.7, 155.5, 144.2, 143.6, 142.6, 142.5, 129.8, 129.8, 126.6, 126.3, 126.3, 123.4, 123.3, 112.1, 111.9, 104.4, 104.3, 103.3 (anomeric α-1,4), 100.8 (anomeric α-1,6), 84.0, 79.4, 75.3, 74.8, 74.8, 74.1, 74.0, 73.8, 73.5, 72.6, 71.9, 68.7, 62.8, 62.4, 58.1 (epoxide), 55.4 (epoxide), 50.5 (C_q), 50.5 (C_q), 44.8 (CH_2N^+), 43.3 (C5), 40.4 ($NHCH_2$), 36.7 (CH_2C=O), 31.5 (NCH_3), 31.4, 30.5, 30.4, 30.3 (4CH_2 linker), 28.2 (CH_2 linker), 28.0 (2CH_3), 27.9 (CH_2 linker), 27.8 (2CH_3), 27.4, 27.1, 26.6 (3CH_2 linker) ppm. HRMS (ESI) m/z: $[M]^+$ calc for $C_{59}H_{86}N_3O_{16}$ 1092.6003, found 1092.6005.

Compound 35



Amine **32** (10.1 mg, 16.1 µmol) was reacted with biotin (1.2 eq) for 19 h according to general procedure A. This afforded the product **35** (7.1 mg, 8.3 µmol, 52%) as a white powder after lyophilization (2x). ^{1}H NMR (500 MHz, D_2O) δ 5.10 (d, J = 4.0 Hz, 1H, anomeric α-1,4), 4.96 (d, J = 3.7 Hz, 1H, anomeric α-1,6), 4.64 – 4.58 (m, 1H, NHC*H*), 4.42 (dd, J = 8.0, 4.5 Hz, 1H, NHC*H*), 4.03 – 3.94 (m, 2H), 3.89 – 3.62 (m, 12H), 3.57 (td, J = 9.6, 3.8 Hz, 2H), 3.52 – 3.39 (m, 4H), 3.33 (ddd, J = 9.1, 5.8, 4.5 Hz, 1H, SCH), 3.28 (t, J = 9.4 Hz, 1H), 3.17 (hept, J = 6.7 Hz, 2H, NHCH_2), 3.00 (dd, J = 13.0, 5.0 Hz, 1H, SC*H*H), 2.79 (d, J = 13.1 Hz, 1H, SCH*H*), 2.41 (ddd, J = 9.2, 6.1, 3.1 Hz, 1H, H5), 2.25 (t, J = 7.1 Hz, 2H, CH_2C=O), 1.78 – 1.54 (m, 6H), 1.50 (t, J = 6.6 Hz, 2H), 1.45 – 1.26 (m, 10H), 1.24 (s, CH_3, ^{t}BuOH residue). ^{13}C NMR (126 MHz, D_2O) δ 176.6 (NHCONH), 165.4 (CONH), 101.0 (anomeric α-1,4), 98.8 (anomeric α-1,6), 81.1, 77.5, 73.3, 73.3, 72.9, 72.9, 72.2, 71.9, 71.7, 71.4, 70.6, 69.5, 67.2, 62.1, 60.4, 60.3, 60.3, 57.4 (epoxide), 55.5

(SCH), 55.2 (epoxide), 41.1 (C5), 39.8 (SCH_2), 39.3 ($NHCH_2$), 35.6 ($CH_2C=O$), 29.7 (CH_3, a small peak, ^{t}BuOH residue), 29.3, 28.5, 28.3, 28.3, 27.9, 27.7, 26.0, 25.3, 25.2 ($9CH_2$ linker), 22.6 (CH_3COONH_4 residue *which was removed after repeated lyophilization*) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{37}H_{64}N_3O_{17}S$ 854.3951, found 854.3948.

3.5 References

[1] Bertoft, E., *Agronomy* **2017**, *7* (3), 56-84.

[2] Tester, R. F.; Karkalas, J.; Qi, X., *J. Cer. Sci.* **2004**, *39* (2), 151-165.

[3] Miyazaki, M.; Van Hung, P.; Maeda, T.; Morita, N., *Trends in Food Sci. & Technol.* **2006**, *17* (11), 591-599.

[4] Ogunsona, E.; Ojogbo, E.; Mekonnen, T., *Eur. Polymer J.* **2018**, *108*, 570-581.

[5] Chen, Q.; Yu, H.; Wang, L.; ul Abdin, Z.; Chen, Y.; Wang, J.; Zhou, W.; Yang, X.; Khan, R. U.; Zhang, H.; Chen, X., *RSC Advances* **2015**, *5* (83), 67459-67474.

[6] van der Maarel, M. J. E. C.; van der Veen, B.; Uitdehaag, J. C. M.; Leemhuis, H.; Dijkhuizen, L., *J. Biotech.* **2001**, *94*, 137–155.

[7] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., *Nucleic Acids Res.* **2014**, *42* (D1), D490-D495.

[8] Sun, H.; Zhao, P.; Ge, X.; Xia, Y.; Hao, Z.; Liu, J.; Peng, M., *Appl. Biochem. Biotechnol.* **2010**, *160* (4), 988-1003.

[9] Rana, N.; Walia, A.; Gaur, A., *Natl. Acad. Sci. Lett.* **2013**, *36* (1), 9-17.

[10] Gupta, R.; Gigras, P.; Mohapatra, H.; Goswami, V. K.; Chauhan, B., *Process Biochem.* **2003**, *38* (11), 1599-1616.

[11] de Souza, P. M.; de Oliveira e Magalhães, P., *Braz. J. Microbiol.* **2010**, *41*, 850-861.

[12] Sundarram, A.; Murthy, T. P. K., *J. Appl. Environ. Microbiol.* **2014**, *2*, 166-175.

[13] Saini, R.; Saini, H. S.; Dahiya, A., *Journal. Pharmacogn. Phytochem.* **2017**, *6*, 1865-1871.

[14] Kang, J.; Park, K. M.; Choi, K. H.; Park, C. S.; Kim, G. E.; Kim, D.; Cha, J., *Enzyme Microb. Technol.* **2011**, *48* (3), 260-266.

[15] Lee, Y. S.; Seo, S. H.; Yoon, S. H.; Kim, S. Y.; Hahn, B. S.; Sim, J. S.; Koo, B. S.; Lee, C. M., *Int. J. Biol. Macromol.* **2016,** *82*, 514-521.

[16] Choi, K. H.; Hwang, S.; Lee, H. S.; Cha, J., *J. Microbiol.* **2011**, *49* (4), 628-634.

[17] Msarah, M. J.; Ibrahim, I.; Hamid, A. A.; Aqma, W. S., *Heliyon* **2020**, *6* (6), e04183.

[18] Shaw, A.; Bott, R.; Day, A. G., *Curr. Opin. Biotechnol.* **1999**, *10*, 349-352.

[19] Nielsen, J. E.; Borchert, T. V., *Biochim. Biophys. Acta.* **2000**, *1543*, 253-274.

[20] Priyadharshini, R.; Gunasekaran, P., *Biotechnol. Lett.* **2007**, *29* (10), 1493-1499.

[21] Arnal, G.; Cockburn, D. W.; Brumer, H.; Koropatkin, N. M., *Protein Sci.* **2018**, *27* (6), 1093-1101.

[22] Agirre, J.; Moroz, O.; Meier, S.; Brask, J.; Munch, A.; Hoff, T.; Andersen, C.; Wilson, K. S.; Davies, G. J., *Acta. Crystallogr. D: Struct. Biol.* **2019**, *75*, 1-7.

[23] Hansen, F. G.; Bundgaard, E.; Madsen, R., *J. Org. Chem.* **2005**, *70* (24), 10139-10142.

[24] Li, K.-Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; van den Elst, H.; Wong, C.-S.; Chander, S. D.; Hoogendoorn, S.; Beenakker, T. J. M.; Codée, J. D. C.; Aerts, J. M. F. G.; van der Marel, G. A.; Overkleeft, H. S., *Eur. J. Org. Chem.* **2014**, *2014* (27), 6030-6043.

[25] Lee, D.; Williamson, C. L.; Chan, L.; Taylor, M. S., *J. Am. Chem. Soc.* **2012**, *134* (19), 8260-8267.

[26] Wang, L.; Overkleeft, H. S.; van der Marel, G. A.; Codée, J. D. C., *J. Am. Chem. Soc.* **2018**, *140* (13), 4632-4638.

[27] Bongini, A.; Cardillo, G.; Orena, M.; Porzi, G.; Sandri, S., *J. Org. Chem.* **1982**, *47*, 4626-4633.

[28] Bartlett, P. A.; Meadows, J. D.; Brown, E. G.; Morimoto, A.; Jernstedt, K. K., *J. Org. Chem.* **1982**, *47*, 4013-4018.

[29] Duan, J. J.-W.; Smith, A. B., *J. Org. Chem.* **1993**, *58*, 3703-3711.

[30] Shan, M.; Sharif, E. U.; O'Doherty, G. A., *Angew. Chem., Int. Ed.* **2010**, *49* (49), 9492-9495.

[31] Thiery, E.; Le Bras, J.; Muzart, J., *Eur. J. Org. Chem.* **2008**, *2009* (7), 961-985.

[32] Kvach, M. V.; Ustinov, A. V.; Stepanova, I. A.; Malakhov, A. D.; Skorobogatyi, M. V.; Shmanai, V. V.; Korshun, V. A., *Eur. J. Org. Chem.* **2008**, *2008* (12), 2107-2117.

4

Synthesis and Biochemical Evaluation of Bifunctional Cyclophellitol Aziridines as Selective GBA Inhibitors and Activity-Based Probes

Part of this chapter is published as:

Rhianna J. Rowland,[*] Yurong Chen,[*] Imogen Breen, Liang Wu, Wendy A. Offen, Thomas J. Beenakker, Qin Su, Adrianus M. C. H. van den Nieuwendijk, Johannes M. F. G. Aerts, Marta Artola, Herman S. Overkleeft and Gideon J. Davies, *Chem. Eur. J.* **2021**, *27*, 1–13.

4.1 Introduction

Glucocerebrosidase (acid glucosylceramidase, GCase, GBA, EC 3.2.1.45), belonging to the glycoside hydrolase (GH) 30 family (www.cazy.org),[1] is a lysosomal retaining β-glucosidase. GBA is primarily responsible for catalyzing the degradation, in lysosomes, of glucosylceramide (GlcCer) by hydrolytic cleavage of the β-glucose moiety from the aglycon to yield free ceramide and glucose with net retention of β-anomeric configuration,[2–4] through a two-step Koshland double displacement mechanism (Figure 4.1). Inherited deficiency in GBA causes the most common lysosomal storage disorder, Gaucher disease (GD), which is primarily characterized by the cellular accumulation of GlcCer, and its deacylated derivative glucosylsphingosine (GlcSph).[2,5,6] The multisystemic storage of these glycolipids leads to the clinical symptoms of GD, which can vary considerably in frequency and severity. Clinical manifestation of GD type 1 and GD type 2 commonly include skeletal disease and visceral disease affecting the spleen, kidneys, liver and heart.[7–10] In more severe cases (GD type 3), neurological disorders also arise due to GlcCer deposition in the brain.[11,12] Moreover, mutations in the gene coding for GBA have recently been identified as the highest known genetic risk factor for Parkinson's disease (PD).[13–15] Thanks to the clinical importance of GBA in both GD and PD, it is one of the most widely studied human glucosidase with relentless interest in developing novel chaperones,[16–19] inhibitors,[20–22] and activity-based probes (ABPs)[23–25] to study this enzyme in disease pathogenesis, diagnosis and treatment.

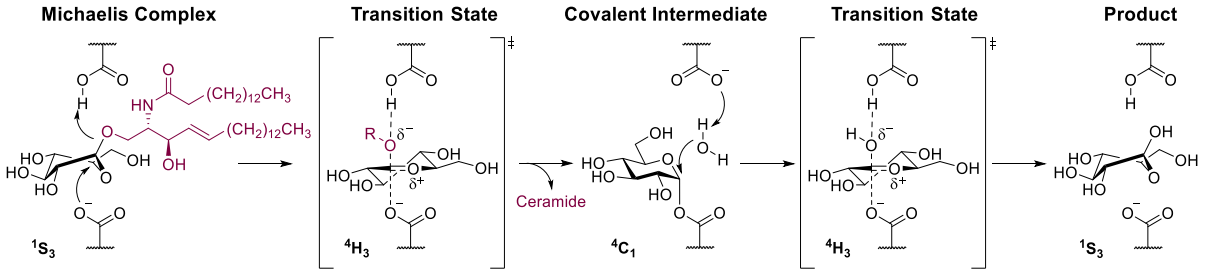


Figure 4.1. Glucocerebrosidase (GBA) catalyzes the hydrolysis of glucosylceramide in a two-step double displacement mechanism to yield glucose and ceramide.

Cyclophellitol **1**, a natural product originally isolated from *Phellinus sp.*, is a potent irreversible inhibitor of retaining β-glucosidases, and inactivates GBA and GBA2 with merely equal efficiency.[26] Previous studies have shown that functionalization at the C6 position of cyclophellitol (cyclophellitol numbering is depicted in Figure 4.2) with a BODIPY reporter group yielded highly potent and very specific activity-based probes (ABPs) to monitor GBA activity *in vitro*, *in situ*, and *in vivo*,[23,27] with potential applications in diagnostics and

therapeutic evaluation. More recently, this methodology was further extended with the design of a C6-Cy5 tagged cyclophellitol ABP **3** and C6-biphenyl/adamantanyl substituted cyclophellitol derivatives **4** and **5**.[22] Both compounds **4** and **5** proved to be very potent and selective GBA inhibitors *in vitro* and *in vivo*, and are suitable for generating neuropathic Gaucher zebrafish models. In the studies on cyclophellitol analogues, substitution of the epoxide oxygen for nitrogen led to the synthesis of cyclophellitol aziridine **2**,[28] which inhibits retaining β-glucosidases with equal or slightly higher potency than the parent cyclophellitol **1**. One route to modify cyclophellitol aziridines and facilitate their conversion to ABPs is functionalization of the aziridine nitrogen. Consequently, installation of reporter moieties (BODIPY, Cy5 or biotin) at the aziridine nitrogen through *N*-acylation[24] or *N*-alkylation[25] yielded broad spectrum retaining β-glucosidase ABPs, which can efficiently label the four human retaining β-glucosidases, GBA, GBA2, GBA3 and lactase-phlorizin hydrolase (LPH).

GBA is notable for its tolerance, indeed preference, for C6-substituted reagents which exhibit increased specificity for GBA over other β-glucosidases (GBA2, GBA3 and LPH), [22,23] while it also favors iminosugar inhibitors[29,30] and cyclophellitol aziridines[24,25] extended at the aziridine nitrogen position. These preferences of GBA suggested that binding of the C6-substituent and aziridine *N*-functionalization may reflect the enzyme's specificity for a lipid substrate with two alkyl tails. It would be of interest to investigate whether a new generation of bi-functional cyclophellitol aziridines, which are functionalized at both the C6-position and the aziridine nitrogen, may exhibit further improvements in potency and selectivity for GBA. In this chapter, the synthesis of a C6-Cy5 tagged *N*-octyl-cyclophellitol aziridine **7** is described, and its GBA activity and selectivity are compared to Cy5 tagged cyclophellitol epoxide **3** and Cy5 cyclophellitol aziridine **6**. Moreover, and in line with previous findings that C6-biphenyl/adamantanyl substituted cyclophellitols are very potent and selective GBA inhibitors, the corresponding bifunctional cyclophellitol aziridine inhibitors **8** and **9** were synthesized, and their *in vitro* activity and selectivity toward GBA evaluated as well.

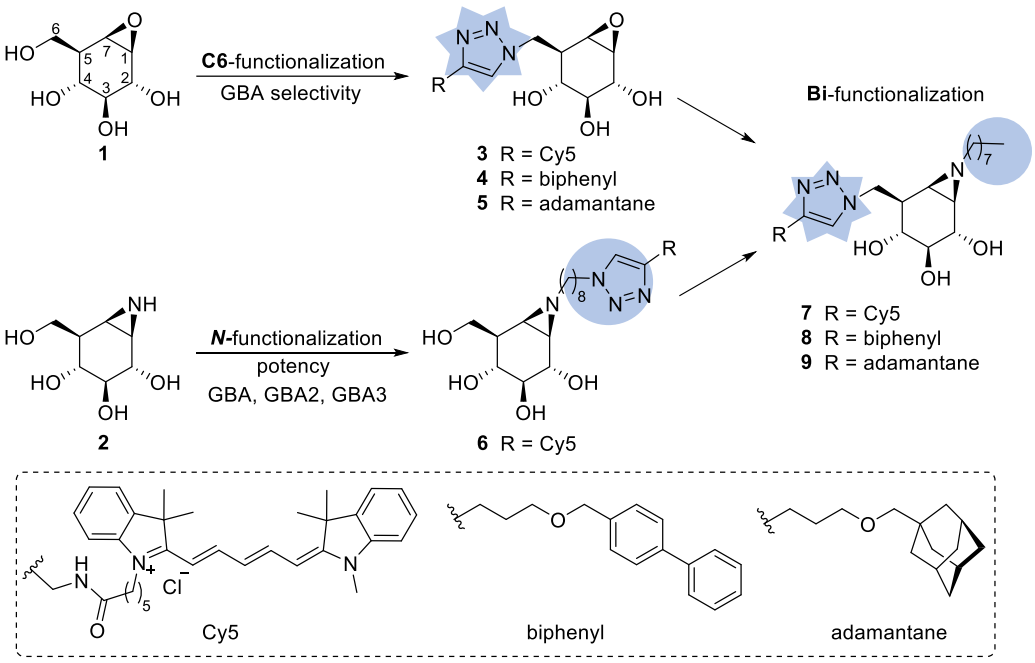


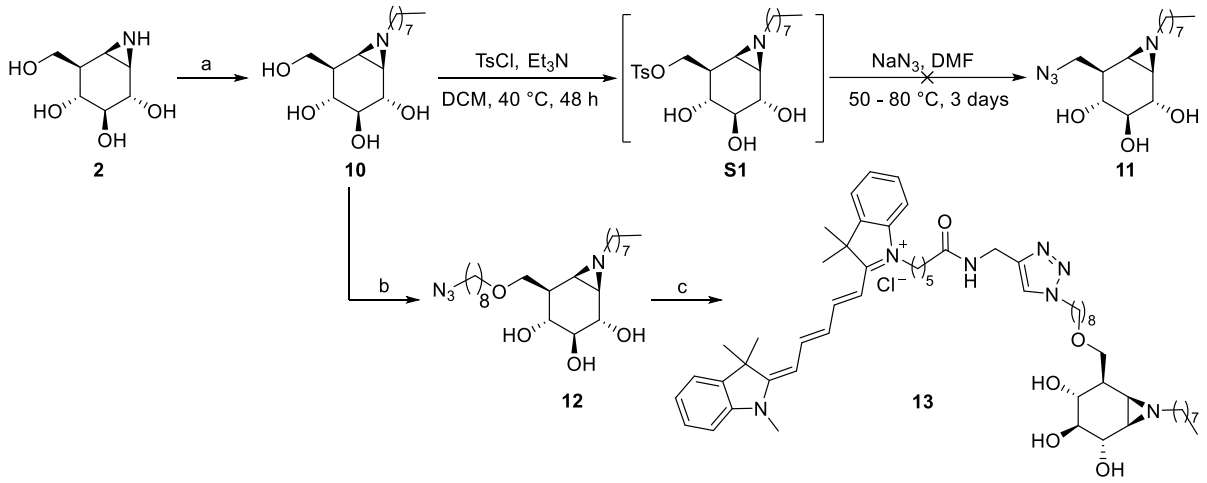
Figure 4.2. Structures of covalent and irreversible retaining β-glucosidase inhibitors and ABPs described in this chapter: cyclophellitol **1** and its C6-functionalized derivatives **3-5**; cyclophellitol aziridine **2** and its *N*-functionalized ABP **6** and bi-functionalized derivatives **7-9**.

4.2 Results and discussion

4.2.1 Synthesis of bi-functionalized cyclophellitol aziridine ABPs and inhibitors

It was envisioned that target cyclophellitol aziridines **7-9** could be obtained from azide **11** via click-chemistry. To gain access to azide **11** (Scheme 4.1), compound **2** was synthesized according to the literature procedures.[28] Reaction of aziridine **2** with 1-iodooctane at 55 °C for prolonged reaction times resulted in isolation of **10** in 39% yield. Introduction of an azido functionality at C6 position was first performed via a two-step procedure. While tosylation of the primary alcohol with an excess of tosyl chloride (6.0 eq.) in the presence of Et_3N (7.2 eq.) gave no reaction at room temperature, heating the reaction to 40 °C resulted in the formation of intermediate **S1** after prolonged reaction times. After work-up, tosylate **S1** was directly reacted with excess sodium azide (20 eq.). However, no conversion of the starting material was observed at 50 °C even after 2 days and increasing the temperature to 80 °C mainly resulted in nucleophilic opening of the aziridine. Other attempts to directly introduce the azido group in one step using diphenylphosphoryl azide (DPPA) in the presence diisopropylazodicarboxylate (DIAD) and triphenylphosphine (Ph_3P)[31–33] were unproductive as well, possibly due to the low reactivity of the unprotected aziridine. Alternatively, the primary alcohol in **10** could be alkylated with an 8-azidooctyl trifluoromethanesulfonate linker, giving compound **12** in low

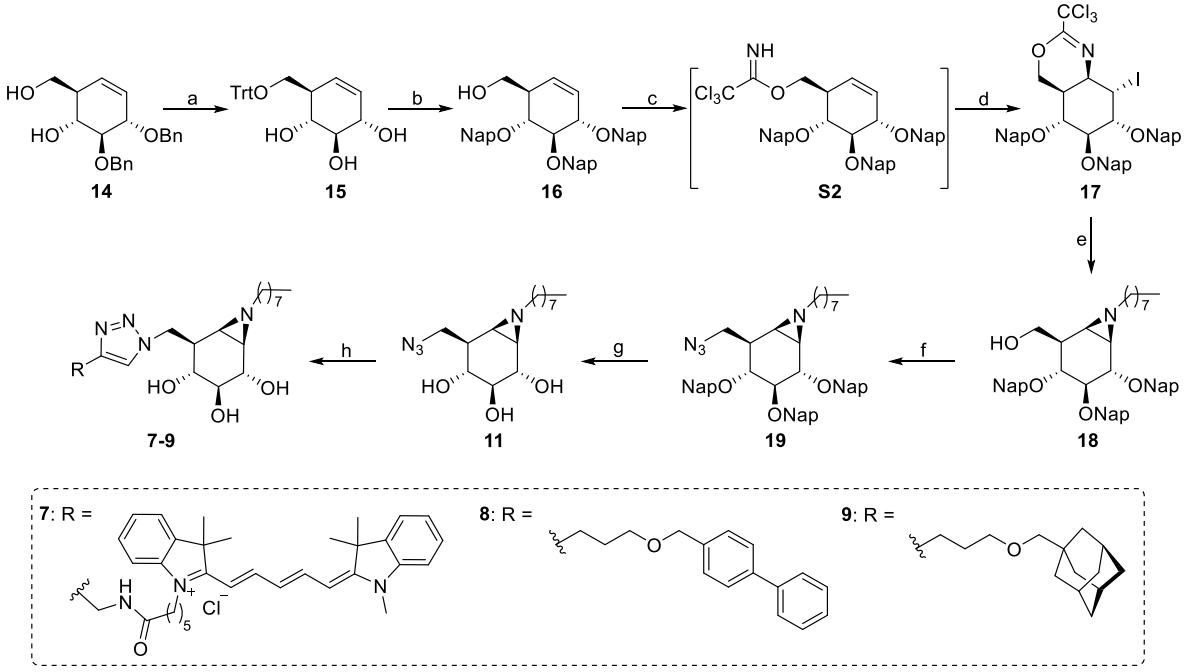
yield. The azide handle could then be functionalized with a Cy5 dye by click-chemistry to afford ABP **13** after HPLC purification. Although not the final target, ABP **13** can also be studied as a bifunctional probe of GBA.



Scheme 4.1. Synthesis of ABP **13**. Reagents and conditions: a) 1-iodooctane, K_2CO_3, DMF, 55 °C, 48 h, 39%; b) 8-azidooctyl trifluoromethanesulfonate, DIPEA, DCM, rt, 4 days, 11%; c) Cy5-alkyne, $CuSO_4$, NaAsc, DMF, rt, 24 h, 25%.

In another attempt to obtain the target C6-azido *N*-octyl aziridine **11**, a new synthetic route following a key 2-naphthylmethyl ether (Nap) protecting group strategy was developed (Scheme 4.2). Briefly, starting intermediate **14** was synthesized in nine steps from D-xylose based on chemistry developed by Madsen and co-workers.[34,35] Standard benzyl deprotection conditions (i.e. Birch reduction or palladium catalyzed hydrogenation) were not compatible with the required azide functionality and electrophilic aziridine, therefore, debenzylation of **14** with boron trichloride (BCl_3) was performed at an early stage, followed by selective tritylation of the primary alcohol affording intermediate **15**. The remaining secondary alcohols were protected as Nap ethers followed by detritylation of the primary alcohol to afford **16**. Treatment of **16** with trichloroacetonitrile yielded a primary imidate intermediate **S2**, which was found to be unstable and partly decomposed during a purification attempt by column chromatography, forming a less polar impurity which was eluted out together with the product. The mixture containing **S2** was directly treated with *N*- iodosuccinimide (NIS) which resulted in the isolation of cyclic imidate **17** in a moderate yield (40%). Hydrolysis of **17** under acidic conditions and subsequent treatment of the resulting ammonium salt with excess base resulted in intramolecular iodine displacement to form a free aziridine, which was then alkylated with an octyl linker to give compound **18**. Tosylation of the primary alcohol using Et_3N as the single base proved to be sluggish, therefore the combination of Et_3N (2.0 eq.) and *N*-methyl imidazole

(7.0 eq.) was employed, allowing complete tosylation of **18** after 28 hours as indicated by TLC-analysis. After work-up, the resulting crude tosylate was reacted with an excess of sodium azide (10.0 eq.) under gentle heating for prolonged reaction times (40 h), resulting in **19** in 68% yield after two steps. The naphthylmethyl ethers were then removed by DDQ to afford compound **11**, which was finally functionalized with appropriate alkynes using click-chemistry to give ABP **7** and inhibitors **8**-**9** after HPLC purification (Scheme 4.2).



Scheme 4.2. Synthesis of ABP **7** and inhibitors **8**-**9**. Reagents and conditions: a) *i*) BCl_3, DCM, -78 °C, 2 h; *ii*) TrCl, Et_3N, DMAP, DMF, rt, 19 h, 30% over two steps; b) *i*) NapBr, NaH, TBAI, DMF, 0 °C – rt, 5 h; *ii*) TsOH, DCM/MeOH (1/1), rt, overnight, 73% over two steps; c) Cl_3CCN, DBU, DCM, rt, overnight; d) NIS, $CHCl_3$, rt, 17 h, 40% over two steps; e) *i*) HCl, DCM/MeOH (1/1), rt, overnight, then Amberlite IRA-67, 20 h; *ii*) 1-iodooctane, K_2CO_3, DMF, 80 °C, 16 h, 35% over two steps; f) *i*) TsCl, Et_3N, *N*-methyl imidazole, DCM, rt, 28 h; *ii*) NaN_3, DMF, 50 °C, 40 h, 68% over two steps; g) DDQ, DCM/H_2O (10/1), rt, 24 h, 66%; h) Cy5-alkyne, $CuSO_4$, NaAsc, DMF, rt, overnight, 30%.

4.2.2 Structural analysis of the ABP 7-enzyme complex

To investigate the accommodation of the two functionalities of ABP **7** by recombinant human GBA (rhGBA, produced in an insect-baculovirus expression vector system (BEVS)[37]), a co-crystal structure in complex with bi-functional ABP **7** was obtained at 1.80 Å resolution, demonstrating covalent binding of the cyclophellitol aziridine to the catalytic nucleophile of GBA (Figure 4.3). Specifically, the reacted cyclophellitol adopts the expected $^4C_1$ chair conformation, with a covalent bond length of 1.47 Å to Glu340. Furthermore, unambiguous electron density for the ring opened *N*-alkyl aziridine warhead was observed, allowing the first

5 carbons of the *N*-octyl chain to be modelled. This was sufficient to establish binding of the *N*-alkyl chain to the narrow active site channel formed by Gln284, Tyr313, Lys346 and Trp348. In fact, the *N*-alkyl chain of ABP **7** extends through this pocket towards the surface of the protein, which may provide some indication into the binding of the fatty acid portion of the natural GlcCer substrate which is thought to project out from the protein and interact with the lipid bilayer.[36] Unfortunately, whilst sufficient electron density for the O6-triazole linker and subsequent amide group was observed, the Cy5 tag could not be modelled. Nevertheless, the O6-triazole linker was found to bind in the hydrophobic cavity formed by Trp348, Phe246 and Tyr313, which was reported previously to accommodate the triazole linker of ABP **3**.[22] Additionally, this binding cavity extends towards the broader hydrophobic allosteric site at the dimer interface where a previously reported C6-BODIPY tag was shown to bind.[23]

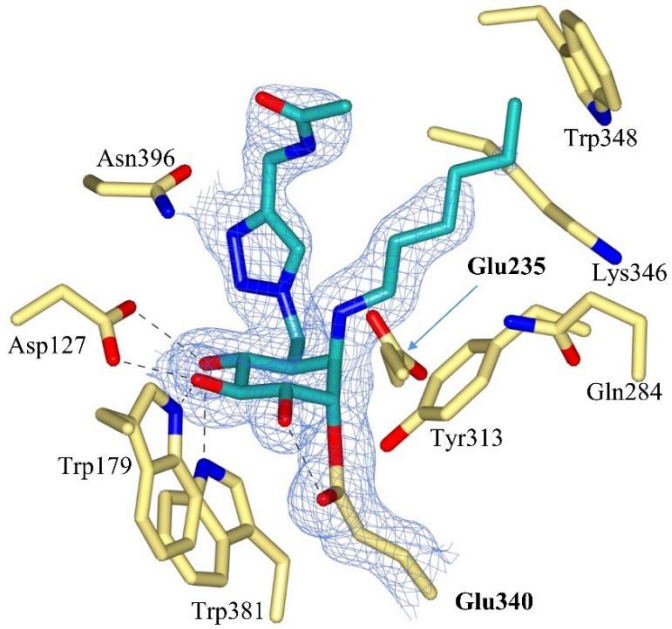


Figure 4.3. Observed electron density for ABP **7** bound covalently to the catalytic nucleophile (Glu340) of rhGBA by trans-diaxial ring opening of the *N*-alkyl aziridine warhead. Maximum-likelihood/σA weighted ($2F_o$-F_c) electron density map contoured to 1.0 σ (a = 1.31 $e^-/Å^3$).

4.2.3 Fluorescent labeling of rhGBA with ABP 3 and ABP 7

Activity-based labeling of rhGBA (produced in BEVS) with bi-functional ABP **7** was performed and compared to labeling by its C6-mono-functionalized epoxide derivative ABP **3**. Firstly, in solution labeling of excess rhGBA (200 nM) was performed in the presence of decreasing ABP concentrations (150-0.001 nM), demonstrating clear concentration dependent labeling with a gel-detection limit of 1 nM for ABP **7** and 0.1 nM for ABP **3** (Figure 4.4A). Secondly, labeling assays in which ABP **7** and ABP **3** (150 nM) were incubated with decreasing

rhGBA concentrations (500-0.01 nM) were performed to further demonstrate the concentration dependent labeling down to 1 nM rhGBA with ABP **7** and 0.1 nM rhGBA with ABP **3** (Figure 4.4B). Additionally, denaturing the enzyme prior to probe incubation totally abrogated labeling in all assays, indicating that labeling is activity-based. These in gel detection limits are concordant across both assays and demonstrate that ABP **3** exhibits ~ 10-fold higher potency than **7**.

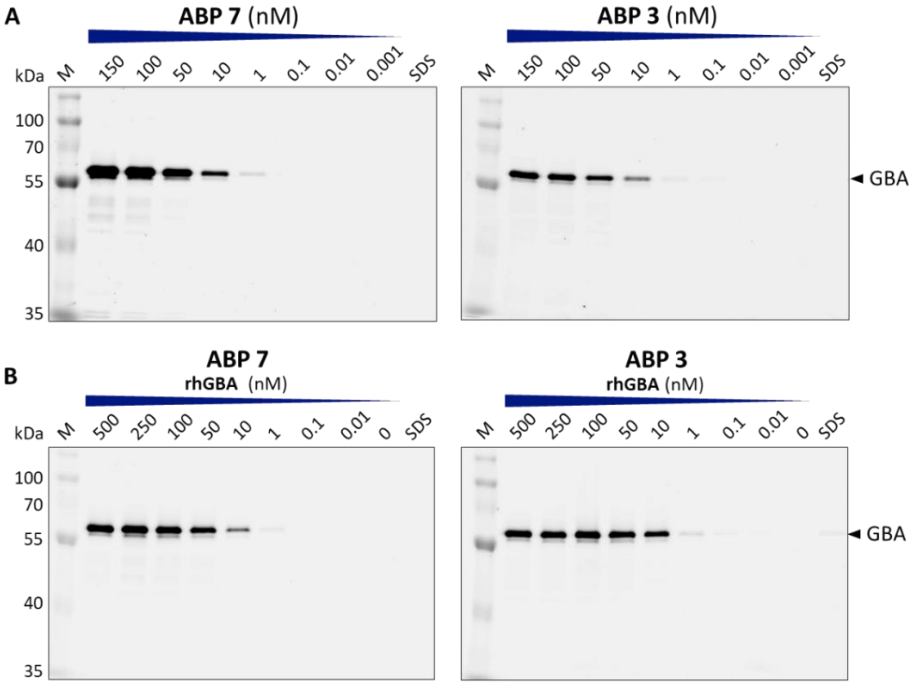


Figure 4.4. A) Labeling of rhGBA (200 nM) with decreasing concentrations (150-0.001 nM) of ABP **3** or ABP **7** at 37 °C for 30 mins followed by SDS-PAGE separation. B) Incubation of ABP **3** or ABP **7** (150 nM) with decreasing concentrations of rhGBA (500-0.01 nM) followed by SDS-PAGE analysis. Fluorescently labeled rhGBA visualized by Cy5 fluorescent readout. SDS = denatured protein sample.

4.2.4 *In vitro* activity and selectivity of bi-functionalized cyclophellitol aziridines

To further investigate the potency and selectivity of the newly synthesized ABPs (**7** and **13**) and inhibitors (**8** and **9**), *in vitro* activity assays against GBA and two related β- and α-glucosidases (GBA2 and GAA) were performed and compared to known C6-mono-functionalized epoxide derivatives **3**-**5**.

Compounds **7**, **8**, **9** and **13** were pre-incubated with recombinant human GBA (rhGBA, Imiglucerase), human GBA2 (from lysates of GBA2 overexpressed cells) and recombinant human GAA (rhGAA, Myozyme) for 3 hours followed by enzymatic activity measurement using 4-methylumbelliferyl-β- and α-glucosides as fluorogenic substrates. As shown in Table

4.1, ABP **13** proved to be a nanomolar inhibitor of rhGBA (with an apparent IC_{50} value of 3.4 nM), which was 15-fold more potent than ABP **7** (with an apparent IC_{50} value of 53.1 nM). Both compounds **7** and **13** were rather inactive towards GBA2 and GAA (apparent IC_{50} values >10 μM), thus exhibiting comparable selective inhibition of GBA (IC_{50} ratio $>10^3$ for both GBA2/GBA and GAA/GBA) as reported previously for ABP **3**.[22] Inhibitors **8** and **9** had *in vitro* IC_{50} values of 10 to 13 nM towards rhGBA, and are thus 10- to 13-times less potent than their epoxide analogues **4** or **5** (apparent IC_{50} values around 1.0 nM). Similar to ABP **7** and **13**, both compounds **8** and **9** proved inactive to GBA2 and GAA, exhibiting IC_{50} ratio of > 10^4 for both GBA2/GBA and GAA/GBA.

Table 4.1. Apparent IC_{50} values for *in vitro* inhibition of rhGBA, rhGAA and GBA2 from overexpressed cell lysates by compounds 7, 8, 9 and 13. Error ranges depict standard deviations from technical duplicates.

***In vitro* IC_{50} (nM)**	**ABP 3**	**4**	**5**	**ABP 13**	**ABP 7**	**8**	**9**
rhGBA	3.20 ± 0.17[a]	1.06 ± 0.19 [a]	0.96 ± 0.17 [a]	3.35 ± 0.75	53.06 ± 2.65	13.35 ± 1.24	9.63 ± 0.01
GBA2 (HEK293T lysate)	412 x 10^3 ± 10.1 x 10^3 [a]	> 10^5 [a]	> 10^5 [a]	23.3 x 10^3 ± 0.13 x 10^3	> 10^5	> 10^5	> 10^5
rhGAA	> 10^5 [a]	> 10^5 [a]	> 10^5 [a]	> 10^5	> 10^5	> 10^5	> 10^5

[a]IC_{50} values from ref 22.

Next, the labeling efficiency and selectivity of ABPs **3**, **7** and **13** towards GBA in mouse brain lysate were evaluated at pH 5.2 (containing 0.2% taurocholate and 0.1% Triton-100) and pH 5.8 as these are the respective optimal conditions for GBA and GBA2 activities (Figure 4.5). As expected, ABPs **3**, **7** and **13** selectively labeled GBA in a concentration-dependent manner under pH 5.2 (upper panels), all with significant labeling observed at 10 nM of ABP. Under pH 5.8 (lower panels), the labeling efficiency of all three ABPs towards GBA decreased and significant labeling can only be observed at 1 μM (ABPs **7** and **13**) and 300 nM (ABP **3**). Importantly, no labeling of GBA2 was observed up to 1 μM, showing good GBA selectivity. For comparison, broad-spectrum β-glucosidase ABP **6** efficiently labeled both GBA and GBA2 at 100 nM under both pH conditions.

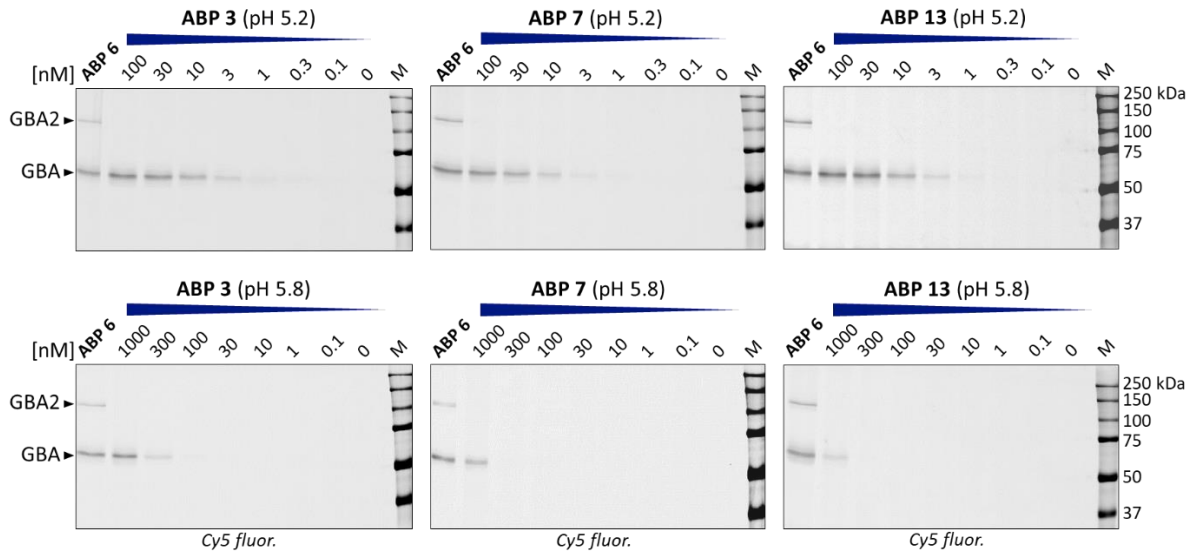


Figure 4.5. Fluorescent labeling of mouse brain lysate (25 µg total protein) with different concentrations of ABPs **3**, **7** and **13** at pH 5.2 (upper panels) or pH 5.8 (lower panels) after incubation for 30 min at 37 °C. Labeling by broad-spectrum β-glucosidase ABP **6** is shown for comparison (100 nM, pH 5.2 or pH 5.8, 30 min, 37 °C).

In addition to these studies, competitive ABPP experiments were performed (Figure 4.6). In this experiment, mouse brain lysates were pre-incubated with different concentrations of inhibitors **8** or **9**, followed by labeling of the residual active enzymes with 50 nM of broad-spectrum ABP **6**. Both compounds **8** and **9** could selectively inhibit GBA in a concentration-dependent manner in the range of 10–1 µM, although full labeling competition was not achieved below 10 µM. In line with the ABP 7 analogue, neither **8** nor **9** showed competition of GBA2 labeling even at the highest concentration applied (10 µM), demonstrating again that functionalization at C6 is detrimental for GBA2 binding.

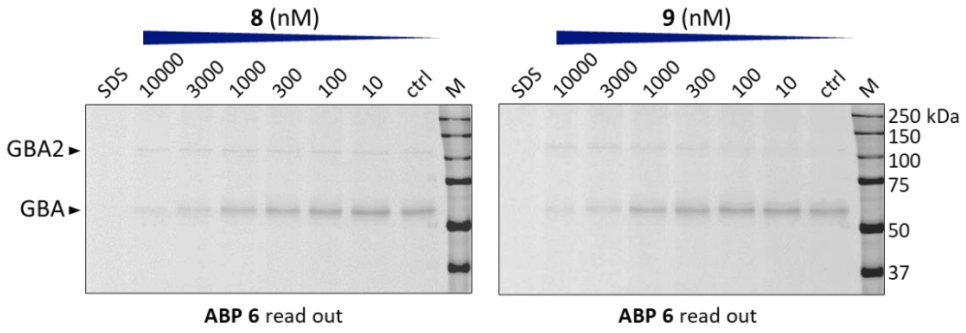


Figure 4.6. Competitive ABPP of mouse brain lysates with **8** and **9**. The sample (25 µg total protein) was pre-incubated with **8** or **9** at different concentrations (60 min, 37 °C, pH 5.2), followed by labeling with 50 nM of ABP **6** (30 min, 37 °C, pH 5.2).

4.2.5 *In situ* labeling of GBA and GBA2 in living cells by ABPs 3, 6 and 7

The selectivity of bi-functionalized ABP **7** towards GBA over GBA2 was further evaluated by *in situ* labeling of living cells and compared to that of broad-spectrum ABP **6** and C6-mono-functionalized epoxide ABP **3**. HEK293T cells containing endogenous GBA and overexpressed GBA2 were treated with the ABPs at different concentrations (1-1,000 nM) for 24 hr. Cells were then washed, lysed and visualized by SDS-PAGE gel-based fluorescence (Figure 4.7). Treatment with broad spectrum ABP **6** resulted in unbiased labeling of GBA and GBA2 at 10 nM, with labeling of both enzymes reaching saturation at 100 nM after 24 hours incubation. In comparison, selective labeling of GBA in ABP **7** treated cells was observed at 10 nM, with some GBA2 labeling observed at higher probe concentrations (100 nM). More selective labeling of GBA was achieved with ABP **3**, which did not label GBA2 even at the highest concentration of probe applied (1 µM).

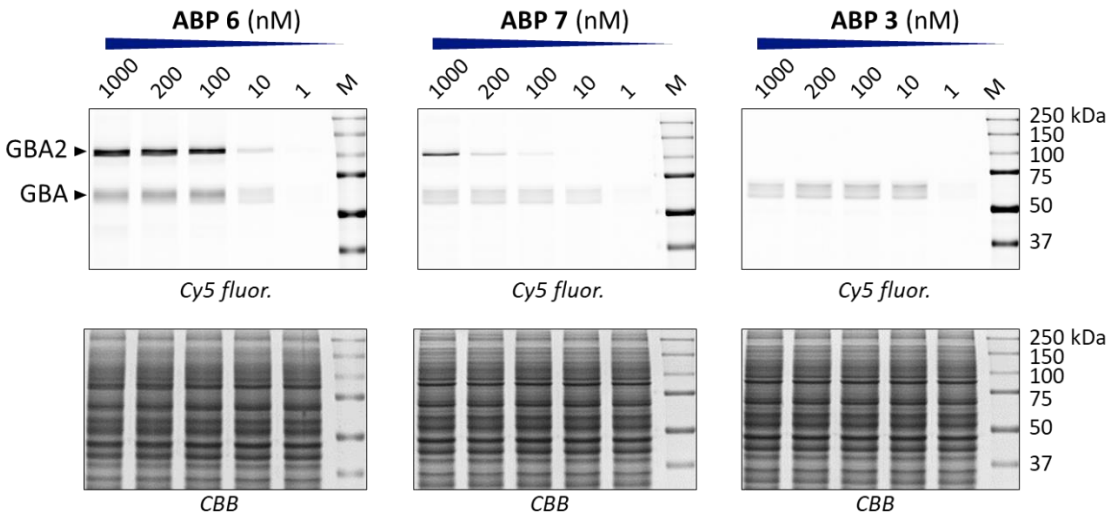


Figure 4.7. *In situ* labeling of GBA and GBA2 in HEK293T cells with ABPs **3**, **6** and **7** at varying concentrations at 37 °C for 24 h, followed by SDS-PAGE separation (top panels). *CBB*, Coomassie Brilliant Blue staining (bottom panels).

4.3 Conclusion

To summarize, this chapter describes the synthesis of a panel of bi-functional cyclophellitol aziridine ABPs and inhibitors for human β-glucocerebrosidase (GBA) with supportive structural analysis and ABPP studies. The X-ray structure of rhGBA in complex with ABP **7** not only demonstrates the mechanism-based mode of action of the compound as a covalent inactivator, but also highlights the binding of C6 and aziridine nitrogen substituents in two distinct active site clefts. Whilst it was structurally validated that the C6- and aziridine

functionalities are well accommodated in two binding cavities, these bi-functional ABPs and inhibitors showed no further improvement in potency or selectivity over their C6-mono-functionalized epoxide counterparts. Nevertheless, this work exemplifies the complexity of ABP development and these bi-functional compounds remain nanomolar inhibitors of GBA which may serve useful in the study of GBA in relation to Gaucher disease and inform the design of next-generation inhibitors and probes through modification of both the C6- and aziridine nitrogen substituents.

4.4 Acknowledgements

Rhianna Rowland and Gideon Davies from University of York, UK are kindly acknowledged for rhGBA (produced in BEVS) labeling experiments and crystallographic studies. Qin Su is acknowledged for his help with IC_{50} and *in situ* labeling experiments.

4.5 Experimental methods

4.5.1 Biochemical experiments

Materials

Recombinant human GBA (rhGBA, imiglucerase, Cerezyme®) and GAA (rhGAA, alglucosidase alfa, Myozyme) were obtained from Sanofi Genzyme (Cambridge, MA, USA). HEK293T (CRL-3216) cell lines were purchased from ATCC (Manassas, VA, USA). Cell lines were cultured in DMEM medium (Sigma-Aldrich, St. Lois, MO, USA), supplied with 10% (v/v) FCS, 0.1% (w/v) penicillin/streptomycin and 1% (v/v) Glutamax, under 5% CO_2 at 37°C. The generation of HEK293T cells overexpressing GBA2[38] and the preparation of cellular homogenates[22] were performed as previously described. Mouse tissue were isolated according to guidelines approved by the ethical committee of Leiden University (DEC#13191). All the tissue lysates were prepared in potassium phosphate lysis buffer (25 mM in pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and 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 Bicinchoninic acid (BCA) Protein Assay Kit (PierceTM).[39] The protein fractions were stored in small aliquots at -80 °C until use.

Production and Crystallisation of Recombinant GBA from BEVS

Recombinant human GBA (rhGBA) was also produced in an insect-baculovirus expression vector system (BEVS) and purified according to previously published procedures.[37] rhGBA was subsequently crystallised in a 48-well MRC sitting-drop vapour-diffusion format using previously reported 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) containing conditions.[37]

Co-crystal Complex with rhGBA from BEVS

Co-crystal complex of **7** was obtained by soaking unliganded rhGBA crystals (produced in BEVS) overnight in mother liquor [0.2 M sodium sulfate, 0.25 M HEPES pH 7.0, 14% (v/v) PEG 3350] spiked with 2 mM ABP **7** and 10% DMSO. Soaked crystals were transferred to a cryoprotectant solution containing 15% ethylene glycol before flash freezing in liquid nitrogen for data collection.

Data Collection, Structure Solution and Refinement

Data were collected at the i04 beamline of the Diamond Light Source (DLS) UK, and processed using XIA2[40] and AIMLESS[41,42] data reduction pipelines in the CCP4i2 suite.[43] Ligand complex of **7** with rhGBA (produced in BEVS) was solved by molecular replacement using PDB 6TJK[37] as the homologous search model.

Refinement was performed using REFMAC[44] followed by several rounds of manual model building with COOT.[45,46] Idealized coordinate sets and refinement dictionaries for the ligand were generated using JLIGAND.[47] Sugar conformations were validated using Privateer[48] and all structures were validated using MolProbity[49] and the wwPDB Validation service (validate-rcsb-1.wwpdb.org/) prior to deposition. Data collection and refinement statistics are summarized in Table 4.S1 and Table 4.S2 (Appendix). Crystal structure figures were generated in CCP4mg.[50]

***In vitro* activity of inhibitors on glycosidases measured by 4-MU substrates**

In vitro apparent IC_{50} measurements with compounds (**7**, **8**, **9** and **13**) in rhGBA and rhGAA were determined using the fluorogenic substrate methods described[22] previously at 3 h incubation time and 37°C. For *in vitro* apparent IC_{50} measurements of GBA2, 8 volumes of cell lysates (4 μg total protein/μL) containing overexpressed human GBA2 were firstly pre-incubated with 1 volume of MDW941 (100 nM final concentration, 0.5% (v/v DMSO)) for 30 min at 37°C to selectively inhibit GBA activity. Lysates were then incubated with 1 volume of compounds at various concentrations for 3 h at 37°C, before subsequent enzymatic assay for GBA2 activity as described earlier.[22,38] The substrate mixtures used for each enzyme are listed as follows: rhGBA, 3.75 mM 4-MU-β-D-glucopyranoside (Glycosynth, Warrington Cheshire, UK) at pH 5.2 (150 mM McIlvaine buffer), supplemented with 0.2% (w/v) sodium taurocholate and 0.1% (v/v) Triton X-100; GBA2, 3.75 mM 4-MU-β-D-glucopyranoside at pH 5.8; rhGAA, 3 mM 4-MU-α-D-glucopyranoside at pH 4.0. All assays were performed in duplicate sets, each with 3 technical replicates at each inhibitor concentration. DMSO concentration was kept at 0.5% - 1% (v/v) in all assays during incubation with compounds. *In vitro* apparent IC_{50} values were calculated by fitting data with [inhibitor] vs response—various slope (four parameters) function using Graphpad Prism 7.0 software. Average values and standard deviations were calculated from the two sets.

Titration of ABP 3 and ABP 7 with rhGBA

rhGBA produced in BEVS[37] was diluted to 200 nM in 150 mM McIlvaine buffer pH 5.2 (containing 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) and ABP **3** or ABP **7** were added to 150, 100, 50, 10, 1, 0.1, 0.01 or 0.001 nM in a final reaction volume of 10 µL. The reactions were incubated at 37 °C for 30 mins and then denatured with Laemmli (x3) sample buffer at 95 °C for 5 minutes. The samples were resolved by electrophoresis in 10% SDS-PAGE gels, running at 200 V for approximately 50 minutes. Wet slab gels were scanned on fluorescence using an Amersham Typhoon 5 Imager (GE Healthcare) with λ_{EX} 635 nm; λ_{EM} > 665 nm.

rhGBA Titration with ABP 3 and ABP 7

rhGBA produced in BEVS[37] was prepared at 500, 250, 100, 50, 10, 1, 0.1 and 0.01 nM in 150 mM McIlvaine buffer pH 5.2 (supplemented with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate). ABP **3** or ABP **7** were added to 150 nM final concentration and the reactions were incubated at 37 °C for 30 mins. The samples were denatured with Laemmli (x3) sample buffer at 95 °C for 5 minutes and resolved by electrophoresis in 10% SDS-PAGE gels, running at 200 V for approximately 50 minutes. Wet slab gels were scanned on fluorescence using an Amersham Typhoon 5 Imager (GE Healthcare) with λ_{EX} 635 nm; λ_{EM} > 665 nm.

Fluorescent Labeling of lysates and SDS-PAGE analysis

Mouse brain lysate (25 µg total protein per sample) was diluted with 150 mM McIlvaine buffer pH 5.2 (with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) or pH 5.8 to a final 10 µL volume and labeled with different concentrations of ABPs **3**, **7** or **13** (diluted with McIlvaine buffer at matching pH to a final 5 µL volume) at 37 °C for 30 min. Fluorescent labeling with broad-spectrum ABP **6** was performed at 100 nM ABP concentration at 37 °C for 30 min at pH 5.2 (with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) or pH 5.8 respectively. Samples were then denatured with 4 µL Laemmli (5x) sample buffer and heated at 98 °C for 5 minutes. Proteins were resolved by electrophoresis in 10% SDS-PAGE gels, running at a constant of 90V for 30 minutes followed by 120V for approximately 60 minutes. Wet slab gels were scanned on fluorescence using a Typhoon FLA9500 Imager (GE Healthcare) using λ_{EX} 635 nm; λ_{EM} > 665 nm and images were processed using ImageLab 5.2.1 (BioRad).

Competitive ABPP experiments

Mouse brain lysate (25 µg total protein per sample) was diluted with 150 mM McIlvaine buffer pH 5.2 (with 0.1 % (v/v) Triton X-100 and 0.2 % (w/v) sodium taurocholate) to a final 10 µL volume, pre-incubated with 2.5 µL of inhibitor (**8** or **9**) at varying concentrations at 37 °C for 60 min. Then, 2.5 µL of ABP **6** was added to a final concentration of 50 nM and the labeling reaction was incubated at 37 °C for 30 min. Additionally a negative control was also performed. 2.5 µL SDS (10 % (w/v)) was added to the 10 µL protein, boiling at 98 °C for 5 min, and incubated with ABP **6** at 37 °C for 30 min. Samples

were then denatured with 4 µL Laemmli (5x) sample buffer at 98 °C for 5 minutes, and proceeded to SDS-PAGE and fluorescent detection as described above.

***In situ* Labeling of HEK 293T cells and SDS-PAGE analysis**

Experiment was conducted based on the previously described methods.[27] The HEK293T cells which contain endogenous GBA and overexpressed GBA2 were cultured in 6-well plates and let them grow to at least 80% confluency before experiment. ABP **3**, ABP **6** and ABP **7** were diluted with DMSO into various concentrations (200x of the final concentration), and 5 µL of ABP at different concentrations were added into 1 mL fresh medium containing cells and incubated at 37°C for 24 h. After incubation, cells were washed 3 times with PBS, detached from culture dishes by scraping and lysed in 90 µL of 25 mM KPi buffer (pH 6.5 supplemented with 0.1% (v/v) Triton X-100 and protease inhibitor cocktail) by sonication. The protein concentrations were determined using BCA kit. For SDS-PAGE, each sample containing the same amount of protein (25 µg total protein) was diluted with 25 mM KPi buffer pH 6.5 (+0.1% (v/v) Triton X-100 and protease inhibitor cocktail) into a total volume of 15 µL, denatured by incubation with 4 µL Laemmli (5x) sample buffer at 98 °C for 5 minutes, and proceeded to SDS-PAGE and fluorescent detection as described above.

4.5.2 Chemical synthesis

General experimental details

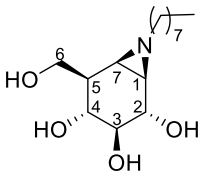
All reagents were of a commercial grade and were used as received unless stated otherwise. Polymer-bound PPh_3 (100-200 mesh, 3.0 mmol/g loading) was purchased from Sigma-Aldrich. Dichloromethane (DCM), chloroform ($CHCl_3$), toluene, tetrahydrofuran (THF) and *N*,*N*-dimethylformamide (DMF) were stored over 4 Å molecular sieves, which were dried *in vacuo* before use. Triethylamine was dried over KOH and distilled before using. 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 analytical thin-layer chromatography (TLC) using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}\cdot H_2O$ (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 an aqueous solution of $KMnO_4$ (7%) and K_2CO_3 (2%) followed by charring at ~150 °C. Column chromatography was performed manually using either Baker or Screening Device silica gel 60 (0.04 - 0.063 mm) or a Biotage Isolera™ flash purification system using silica gel cartridges (Screening devices SiliaSep HP, particle size 15-40 µm, 60A) in the indicated solvents. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker DMX-600 (600/150 MHz), Bruker AV-400 (400/101 MHz), Bruker AV-500 (500/126 MHz) and Bruker AV-850 (850/214 MHz) spectrometers 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. The following abbreviations are used to describe peak patterns

when appropriate: s (singlet), d (doublet), t (triplet), qt (quintet), m (multiplet), br (broad), Ar (aromatic), app (apparent). 2D NMR experiments (HSQC, COSY and NOESY) were carried out to assign protons and carbons of the new structures and assignation follows the general numbering shown in compound **10**. High-resolution mass spectra (HRMS) of intermediates were recorded with a LTQ Orbitrap (Thermo Finnigan) and final compounds were recorded with an apex-QE instrument (Bruker). 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 C18 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 an Agilent Technologies 1260 Infinity LCMS with a 6120 Quadrupole MS system equipped with buffers A: H_2O, B: acetonitrile (MeCN) and C: 100 mM NH_4OAc. For reversed-phase HPLC-MS purifications an Agilent Technologies 1200 series prepLCMS with a 6130 Qudropole MS system was used equipped with buffers A: 50 mM NH_4HCO_3 in H_2O and B: MeCN.

Experimental Procedures and Characterization Data of Products

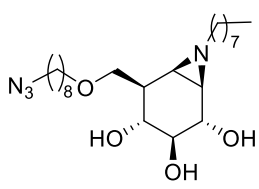
Known compounds **2**[28] and **14**[35] were synthesized following procedures previously described and their spectroscopic data are in agreement with those previously reported.

Compound 10



Compound **2** (475 mg, 2.71 mmol) was dissolved in anhydrous DMF (20 mL). 1-iodooctane (0.75 mL, 4.16 mmol) and K_2CO_3 (1.69 g, 12.2 mmol) were added to the solution and the mixture was stirred at 55 °C for 48 h. The reaction mixture was filtered over a pad of celite and concentrated under reduced pressure. The product was purified by silica gel column chromatography (DCM/MeOH, 100:0→90:10) affording compound **10** (300 mg, 1.05 mmol, 39%) as a yellowish oil. ^{1}H NMR (400 MHz, MeOD) δ 3.99 (dd, J = 10.1, 4.4 Hz, 1H, H6a), 3.67 – 3.56 (m, 2H, H6b and H2), 3.11 (dd, J = 9.9, 8.1 Hz, 1H, H3), 3.02 (t, J = 9.8 Hz, 1H, H4), 2.36 (dt, J = 11.7, 7.7 Hz, 1H, NCH*H*), 2.16 (dt, J = 11.5, 7.3 Hz, 1H, NCH*H*), 2.01 (dd, J = 6.4, 3.5 Hz, 1H, H7), 1.94 – 1.84 (m, 1H, H5), 1.66 (d, J = 6.3 Hz, 1H, H1), 1.63 – 1.51 (m, 2H, CH_2 octyl), 1.39 – 1.22 (m, 10H, 5CH_2 octyl), 0.93 – 0.87 (m, 3H, CH_3). ^{13}C NMR (101 MHz, MeOD) δ 79.0 (C3), 73.9 (C2), 70.1 (C4), 63.7 (C6), 62.2 (NCH_2), 45.5, 45.5 (C1 and C5), 43.1 (C7), 33.0, 30.7, 30.4, 30.3, 28.4, 23.7 (6CH_2 octyl), 14.4 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{15}H_{30}NO_4$ 288.21693, found 288.21674.

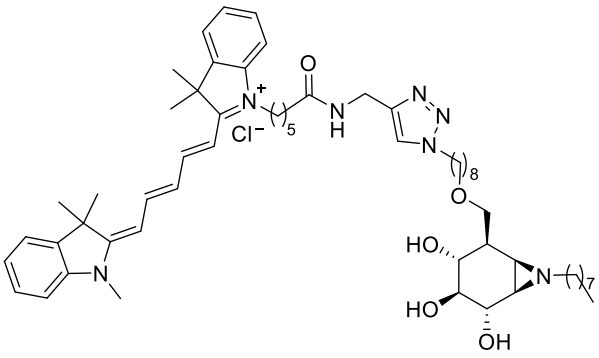
Compound 12



Compound **10** (16 mg, 56 µmol) was dissolved in anhydrous DCM (2 mL) and cooled to -25 °C. DIPEA (19.6 µL, 112 µmol) and freshly prepared 8-azidooctyl trifluoromethanesulfonate (1.0 M in anhydrous DCM, 112 µmol) were added, and the reaction mixture was stirred at -25 °C for 1 h and warmed to rt and stirred over-weekend. LC-MS indicated the presence of starting material so more DIPEA (19.6

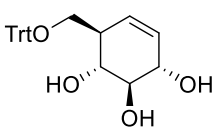
µL, 112 µmol) and 8-azidooctyl trifluoromethanesulfonate (1.0 M in anhydrous DCM, 112 µmol) were added. After stirring at rt for 30 h, the reaction mixture was diluted with EtOAc, washed with sat. aq. $NaHCO_3$, water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (DCM/MeOH, 100:0→90:10) affording product **12** (2.7 mg, 6.1 µmol, 11%) as a clean oil and unreacted starting material **10** (6.5 mg, 23 µmol) was recovered as well. ^{1}H NMR (600 MHz, MeOD) δ 3.80 (dd, J = 8.7, 3.9 Hz, 1H, H6a), 3.58 (d, J = 8.3 Hz, 1H, H2), 3.53 (dt, J = 9.3, 6.4 Hz, 1H, OCH*H* azido-octyl), 3.49 – 3.41 (m, 2H, H6b and OCH*H* azido-octyl), 3.28 (td, J = 6.5, 1.9 Hz, 2H, CH_2N_3), 3.10 (dd, J = 10.0, 8.3 Hz, 1H, H3), 2.97 (t, J = 9.8 Hz, 1H, H4), 2.57 (ddd, J = 11.6, 9.3, 6.5 Hz, 1H, NCH*H*), 2.03 – 1.94 (m, 2H, H5 and H7), 1.94 – 1.86 (m, 1H, NCH*H*), 1.67 (d, J = 6.2 Hz, 1H, H1), 1.64 – 1.51 (m, 6H, $3CH_2$), 1.45 – 1.24 (m, 18H, $9CH_2$), 0.91 (t, J = 7.0 Hz, 3H, CH_3). ^{13}C NMR (151 MHz, MeOD) δ 78.9 (C3), 73.8 (C2), 72.4, 72.3 (C6 and OCH_2), 69.8 (C4), 62.3 (CH_2NH_2), 52.4 (CH_2N_3), 45.8 (C1), 43.6 (C5), 43.3 (C7), 33.1, 30.9, 30.8, 30.6, 30.4, 30.3, 30.3, 30.0, 28.5, 27.8, 27.4, 23.8 ($12CH_2$), 14.5 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{23}H_{45}N_4O_4$ 441.34353, found 441.34388.

Compound 13



Compound **12** (2.7 mg, 6.1 µmol) was dissolved in DMF (0.5 mL). Then Cy5-alkyne (4.4 mg, 7.9 µmol), $CuSO_4$ (0.1 M in H_2O, 18.3 uL, 1.83 µmol, 0.3 eq.) and sodium ascorbate (0.1 M in H_2O, 18.3 uL, 1.83 µmol, 0.3 eq.) were added and the mixture was stirred for 24 h at rt until full conversion was observed by LC-MS. The solvent was evaporated and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) affording ABP **13** (1.5 mg, 1.5 µmol, 25%) as a blue powder after lyophilization. ^{1}H NMR (600 MHz, MeOD) δ 8.28 (t, J = 13.0 Hz, 2H), 7.87 (s, 1H), 7.53 (d, J = 7.4 Hz, 2H), 7.45 (td, J = 7.2, 3.2 Hz, 2H), 7.32 (dt, J = 8.6, 6.7 Hz, 4H), 6.65 (t, J = 12.4 Hz, 1H), 6.31 (d, J = 13.7 Hz, 2H), 4.47 – 4.35 (m, 4H), 4.12 (t, J = 7.4 Hz, 2H), 3.82 (dd, J = 8.5, 3.9 Hz, 1H), 3.67 (s, 3H), 3.62 (d, J = 8.2 Hz, 1H), 3.55 – 3.41 (m, 3H), 3.16 – 3.08 (m, 1H), 3.00 (t, J = 9.6 Hz, 1H), 2.64 – 2.48 (m, 1H), 2.28 (t, J = 7.2 Hz, 2H), 2.04 – 1.81 (m, 10H), 1.76 (s, 14H), 1.55 (dd, J = 24.4, 7.3 Hz, 4H), 1.34 (d, J = 9.7 Hz, 18H), 0.92 (t, J = 5.6 Hz, 3H). HRMS (ESI) m/z: $[M]^+$ calc for $C_{58}H_{86}N_7O_5$ 960.66850, found 960.66832.

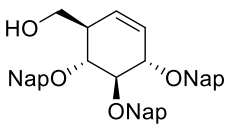
Compound 15



Compound **14** (1.0 g, 2.9 mmol) was dissolved in dry DCM (20 mL) and cooled to -78 °C. BCl_3 (1.0 M in DCM, 15.0 mL, 15.0 mmol) was added slowly and the mixture was stirred at -78 °C for 2 h. After quenching with MeOH, the solvent was

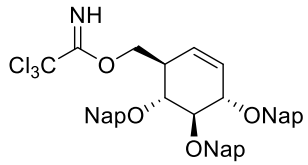
evaporated and the residue was co-evaporated with toluene (3 x) and directly dissolved in dry DMF (15 mL). Then Et_3N (1.0 mL, 7.5 mmol), trityl chloride (1.7 g, 6.0 mmol), and DMAP (18 mg, 0.15 mmol) were added and the reaction mixture was stirred at rt for 19 h. The mixture was diluted with H_2O, extracted with EtOAc (2 x), the combined organic layers were washed with water (2 x) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (DCM/MeOH, 50:1→9:1) affording compound **15** (0.36 g, 0.89 mmol, 30% over two steps) as a pale-yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.47 – 7.35 (m, 5H, 5CH Ar), 7.31 – 7.14 (m, 10H, 10CH Ar), 5.53 (dt, *J* = 10.1, 2.4 Hz, 1H, H7), 5.40 (dt, *J* = 10.2, 2.1 Hz, 1H, H1), 4.18 – 4.06 (m, 1H, H2), 3.61 – 3.49 (m, 2H, H3 and H4), 3.25 (dd, *J* = 8.8, 5.4 Hz, 1H, H6a), 3.21 – 3.13 (m, 1H, H6b), 2.50 – 2.42 (m, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$) δ 143.9 ($3C_q$ Ar), 129.3 (C1), 128.7 (6CH Ar), 128.0 (6CH Ar), 127.4 (C7), 127.2 (3CH Ar), 87.1 (C_q), 77.7 (C3/C4), 72.3 (C2), 72.3 (C3/C4), 65.1 (C6), 44.4 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{26}H_{26}O_4Na$ 425.17233, found 425.17274.

Compound 16



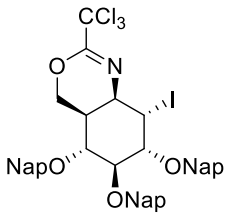
Compound **15** (328 mg, 0.815 mmol) was dissolved in dry DMF (6 mL) and cooled to 0 °C. NaH (60% in mineral, 261 mg, 6.52 mmol) was added to the mixture and stirred at 0 °C for 30 min. Then NapBr (1.08 g, 4.86 mmol) and TBAI (30 mg, 0.082 mmol) were added successively and stirred at 0 °C for 10 min. Then the mixture was warmed to rt and stirred for 5 h. The reaction was quenched with H_2O at 0 °C, diluted with H_2O and extracted with EtOAc (2 x), the combined organic layers were washed with water (2 x) and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The crude was then dissolved in a mixture of DCM/MeOH (1/1, 4 mL/4mL). TsOH was added until pH ≈ 2 and the mixture was stirred at rt overnight. Et_3N was added to quench the reaction and the solvent was concentrated *in vacuo*. EtOAc was added and the solution was washed with sat. aq. NH_4Cl, sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 6:1→3:1) affording compound **16** (345 mg, 0.595 mmol, 73%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.88 – 7.63 (m, 11H, 11CH Ar), 7.53 – 7.34 (m, 10H, 10CH Ar), 5.80 (dt, *J* = 10.1, 2.5 Hz, 1H, H7), 5.57 (dt, *J* = 10.1, 2.0 Hz, 1H, H1), 5.17 (d, *J* = 11.5 Hz, 1H, CH*H* Nap), 5.12 (d, *J* = 4.4 Hz, 2H, CH_2 Nap), 4.87 (s, 2H, CH_2 Nap), 4.83 (d, *J* = 11.3 Hz, 1H, CH*H* Nap), 4.35 (ddd, *J* = 7.8, 3.4, 1.9 Hz, 1H, H2), 3.96 (dd, *J* = 10.1, 7.7 Hz, 1H, H6a), 3.81 – 3.65 (m, 3H, H6b, H3 and H4), 2.59 – 2.53 (m, *J* = 9.4, 3.3 Hz, 1H, H5), 1.78 – 1.49 (br s, 1H, OH). ^{13}C NMR (101 MHz, $CDCl_3$) δ 136.4, 136.0, 135.8, 133.4, 133.4, 133.4, 133.1, 133.1, 133.0 ($9C_q$ Ar), 128.4, 128.4, 128.3, 128.3, 128.2, 128.1, 128.0, 128.0, 127.8, 127.0, 126.9, 126.7, 126.6, 126.3, 126.2, 126.2, 126.1, 126.1, 126.0, 126.0, 125.9 (21CH Ar, C1 and C7), 85.2 (C4), 80.8, 79.0 (C2 and C3), 75.4, 75.3, 72.2 ($3CH_2$ Nap), 63.4 (C6), 45.9 (C5). HRMS (ESI) m/z: $[M+NH_4]^+$ calc for $C_{40}H_{40}O_4N$ 598.29519, found 598.29479.

Compound S2



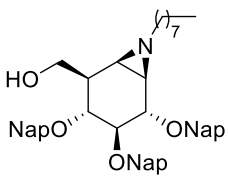
Compound **16** (0.34 g, 0.59 mmol) was dissolved in dry DCM (5.0 mL). Then trichloroacetonitrile (117 µL, 1.18 mmol) and DBU (9 µL, 59 µmol) were added and the mixture was stirred at rt overnight. The mixture was concentrated and the product was purified by silica gel column chromatography (pentane/EtOAc, 13:1→7:1) affording intermediate **S2** (302 mg) as a clean oil which was not completely pure and a yield over two steps is provided after the next step. *Note: TLC-analysis showed clean conversion of the reaction, while the product partly decomposed during a purification attempt by column chromatography, forming a less polar impurity which was eluted out together with the product.*

Compound 17



Compound **S2** (0.30 g) was co-evaporated with toluene, dissolved in dry $CHCl_3$ (4.5 mL) and cooled down to 0 °C. Then NIS (140 mg, 0.62 mmol) was added and the mixture was stirred at rt for 17 h. The reaction was quenched with sat. aq. $Na_2S_2O_3$ and stirred vigorously for 10 minutes. The mixture was diluted with $CHCl_3$, washed with sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 50:1→20:1) affording compound **17** (199 mg, 0.234 mmol, 40% over two steps) as a white solid. ^{1}H NMR (850 MHz, $CDCl_3$) δ 7.86 – 7.78 (m, 6H, 6CH Ar), 7.76 (dd, *J* = 8.1, 1.4 Hz, 1H, CH Ar), 7.74 – 7.70 (m, 3H, 3CH Ar), 7.68 (d, *J* = 1.7 Hz, 1H, CH Ar), 7.63 (dd, *J* = 7.9, 1.3 Hz, 1H, CH Ar), 7.58 (dd, *J* = 8.4, 1.7 Hz, 1H, CH Ar), 7.51 – 7.39 (m, 8H, 8CH Ar), 5.19 – 5.16 (m, 2H, CH_2 Nap), 4.98 (d, *J* = 10.7 Hz, 1H, CH*H* Nap), 4.93 (d, *J* = 11.3 Hz, 1H, CH*H* Nap), 4.85 (t, *J* = 3.4 Hz, 1H, H2), 4.80 (dd, *J* = 11.3, 4.2 Hz, 2H, CH_2 Nap), 4.66 (dd, *J* = 11.3, 1.4 Hz, 1H, H6a), 4.27 (dd, *J* = 11.2, 2.9 Hz, 1H, H6b), 4.10 – 4.06 (m, 2H, H4 and H1), 3.48 (dd, *J* = 11.2, 9.2 Hz, 1H, H7), 2.84 (dd, *J* = 9.3, 3.8 Hz, 1H, H3), 2.76 (dddd, *J* = 11.2, 4.4, 2.9, 1.4 Hz, 1H, H5). ^{13}C NMR (214 MHz, $CDCl_3$) δ 153.5 (C=N), 135.9, 135.4, 135.0, 133.4, 133.4, 133.4, 133.3, 133.2, 133.1 (9C_q Ar), 128.5, 128.5, 128.3, 128.2, 128.1, 128.1, 127.9, 127.8, 127.8, 127.5, 127.2, 127.1, 126.5, 126.5, 126.3, 126.3, 126.2, 126.2, 126.2, 126.0, 126.0 (21CH Ar), 91.4 (CCl_3), 85.2 (C4), 77.2 (C3), 76.4, 76.2 (2CH_2 Nap), 76.1 (C7), 72.4 (CH_2 Nap), 68.3 (C6), 58.7 (C1), 36.3 (C2), 33.8 (C5). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{42}H_{36}Cl_3INO_4$ 850.07491, found 850.07431.

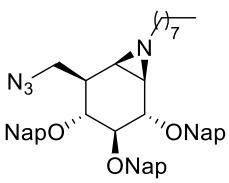
Compound 18



Compound **17** (0.19 g, 0.22 mmol) was dissolved in a mixture of DCM/MeOH (1/1, 1 mL/1 mL) and HCl (1.25 M in MeOH, 0.53 mL, 0.66 mmol) was added. The mixture was stirred overnight and subsequently neutralized by addition of Amberlite IRA-67. After stirring at rt for 20 h, the reaction mixture was filtered

and the resin was washed with DCM/MeOH (3 x). The filtrate was concentrated to afford an oil which was co-evaporated with toluene (3 x) and directly taken up in dry DMF (2 mL). Then K_2CO_3 (39 mg, 0.28 mmol) and 1-iodooctane (86 µL, 0.48 mmol) were added. The mixture was heated up to 80 °C and stirred for 16 h. After cooling to rt, the mixture was diluted with H_2O, extracted with EtOAc (2 x), the combined organic layers were washed with water (2 x) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 7:1→4:1) affording compound **18** (55 mg, 78 µmol, 35% over two steps) as a pale yellow solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.83 – 7.69 (m, 9H, 9CH Ar), 7.69 – 7.60 (m, 3H, 3CH Ar), 7.52 – 7.35 (m, 9H, 9CH Ar), 5.15 – 4.85 (m, 5H, 5CH*H* Nap), 4.74 (d, *J* = 10.9 Hz, 1H, 1CH*H* Nap), 4.03 – 3.90 (m, 2H, H6ab), 3.90 – 3.82 (m, 1H, H2), 3.66 – 3.57 (m, 2H, H3 and H4), 2.95 – 2.80 (br s, 1H, OH), 2.23 (dt, *J* = 11.5, 7.2 Hz, 1H, NCH*H*), 2.17 – 2.06 (m, 2H, H5 and NCH*H*), 1.84 (dd, *J* = 6.3, 3.5 Hz, 1H, H7), 1.60 (d, *J* = 6.3 Hz, 1H, H1), 1.46 (p, *J* = 7.7 Hz, 2H, CH_2), 1.32 – 1.17 (m, 10H, 5CH_2), 0.89 (t, *J* = 7.0 Hz, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 136.6, 136.1, 135.7, 133.5, 133.4, 133.4, 133.2, 133.1, 133.0 (9C_q Ar), 128.4, 128.2, 128.1, 128.1, 127.8, 127.8, 127.7, 126.9, 126.7, 126.4, 126.3, 126.1, 126.1, 126.1, 126.1, 125.9, 125.8 (21CH Ar), 85.6 (C3), 81.5 (C2), 77.1 (C4), 75.8, 75.4, 73.0 (3CH_2 Nap), 63.9 (C6), 61.1 (NCH_2), 43.3 (C5), 42.6 (C7), 40.7 (C1), 32.0, 29.6, 29.6, 29.4, 27.5, 22.8 (6CH_2), 14.3 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{48}H_{54}NO_4$ 708.40474, found 708.40436.

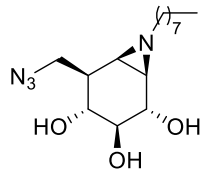
Compound 19



Compound **18** (50 mg, 70 µmol) was dissolved in dry DCM (0.7 mL) in a microwave tube. Et_3N (19 µL, 0.14 mmol) and *N*-methyl imidazole (40 µL, 0.50 mmol) were added and the mixture was cooled to 0 °C. Tosylchloride (40 mg, 0.21 mmol) was added at 0 °C. The tube was then sealed and the mixture was stirred at rt for 28 h. The reaction was quenched with H_2O at 0 °C, diluted with EtOAc, washed with diluted aq. 1.0 M HCl, sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. After co-evaporation with toluene (2 x), the crude intermediate was dissolved in dry DMF (0.7 mL). NaN_3 (45 mg, 0.70 mmol) was added and the mixture was heated up to 50 °C and stirred for 40 h. After cooling to rt, the mixture was diluted with H_2O, extracted with EtOAc (2 x), the combined organic layers were washed with H_2O (2 x) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 18:1→10:1) affording compound **19** (35 mg, 48 µmol, 68% over two steps) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.84 – 7.58 (m, 12H, 12CH Ar), 7.53 – 7.31 (m, 9H, 9CH Ar), 5.10 – 4.85 (m, 5H, 5CH*H* Nap), 4.60 (d, *J* = 11.2 Hz, 1H, 1CH*H* Nap), 3.89 (d, *J* = 8.2 Hz, 1H, H2), 3.76 (dd, *J* = 11.6, 3.5 Hz, 1H, H6a), 3.56 (dd, *J* = 10.0, 8.1 Hz, 1H, H3), 3.31 (dd, *J* = 11.6, 10.0 Hz, 1H, H6b), 3.23 (t, *J* = 10.0 Hz, 1H, H4), 2.22 – 2.09 (m, 3H, H5 and NCH_2), 1.88 (dd, *J* = 6.1, 3.1 Hz, 1H, H7), 1.70 (d, *J* = 6.1 Hz, 1H, H1), 1.44 (tt, *J* = 9.2, 6.9, 3.1 Hz, 2H, CH_2), 1.26 (d, *J* = 6.3 Hz, 10H, 5CH_2), 0.89 (t, *J* = 6.7 Hz, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 136.5, 135.9, 135.6, 133.4, 133.4, 133.4, 133.2, 133.1, 133.0 (9C_q Ar),

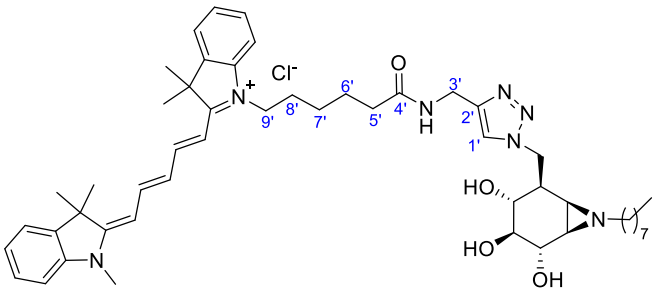
128.4, 128.3, 128.1, 128.1, 128.0, 127.8, 127.8, 126.9, 126.7, 126.4, 126.3, 126.2, 126.1, 126.1, 126.1, 126.0, 126.0, 125.8 (21CH Ar), 85.7 (C3), 81.5 (C2), 77.3 (C4), 75.6, 75.4, 72.9 ($3CH_2$ Nap), 61.1 (NCH_2), 52.3 (C6), 42.3 (C5), 42.1 (C7), 42.0 (C1), 32.0, 29.6, 29.6, 29.4, 27.5, 22.8 ($6CH_2$), 14.4 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{48}H_{53}N_4O_3$ 733.41122, found 733.41092.

Compound 11



Compound **19** (31 mg, 42 µmol) was dissolved in a mixture of DCM/H_2O (10/1, 0.8 mL/0.08 mL) and DDQ (39 mg, 0.17 mmol) was added. Then the mixture was stirred at rt for 24 h until full conversion was observed by LC-MS. The reaction was quenched with a mixture of sat. aq. $NaHCO_3$ and sat. aq. $Na_2S_2O_3$, extracted with EtOAc (3 x), 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 (DCM/MeOH, 100:1→19:1) and by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) affording compound **11** (8.7 mg, 28 µmol, 66%) as a white powder after lyophilization. ^{1}H NMR (850 MHz, MeOD) δ 3.83 (dd, J = 11.8, 3.6 Hz, 1H, H6a), 3.58 (d, J = 8.3 Hz, 1H, H2), 3.34 – 3.31 (m, 1H, H6b), 3.09 (dd, J = 10.0, 8.3 Hz, 1H, H3), 2.98 (t, J = 9.8 Hz, 1H, H4), 2.41 (ddd, J = 11.7, 8.9, 6.5 Hz, 1H, NCH*H*), 2.11 (ddd, J = 11.8, 8.8, 5.7 Hz, 1H, NCH*H*), 1.95 – 1.92 (m, 1H, H7), 1.90 (dt, J = 9.9, 3.5 Hz, 1H, H5), 1.67 (d, J = 6.2 Hz, 1H, H1), 1.60 – 1.53 (m, 2H, CH_2), 1.41 – 1.27 (m, 10H, $5CH_2$), 0.91 (t, J = 7.2 Hz, 3H, CH_3). ^{13}C NMR (214 MHz, MeOD) δ 78.9 (C3), 73.9 (C2), 69.9 (C4), 62.1 (NCH_2), 53.5 (C6), 45.8 (C1), 43.6 (C5), 43.1 (C7), 33.0, 30.7, 30.5, 30.4, 28.4, 23.7 ($6CH_2$), 14.4 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{15}H_{29}N_4O_3$ 313.22342, found 313.22312.

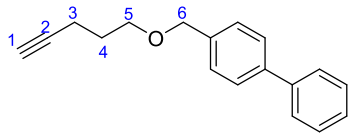
ABP 7



Compound **11** (4.0 mg, 13 µmol) was dissolved in DMF (0.5 mL). Then Cy5-alkyne (10.6 mg, 19 µmol), $CuSO_4$ (0.4 mg, 2.5 µmol) and sodium ascorbate (1.0 mg, 5.1 µmol) were added and the mixture was stirred overnight at rt until full conversion was observed by LC-MS. The solvent was evaporated and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) affording ABP **7** (3.4 mg, 3.9 µmol, 30%) as a blue powder after lyophilization. ^{1}H NMR (850 MHz, MeOD) δ 8.24 (td, J = 13.0, 3.9 Hz, 2H, 2CH=CH), 7.89 (s, 1H, H1'), 7.49 (dd, J = 7.4, 1.2 Hz, 2H, 2CH Ar), 7.41 (qd, J = 7.5, 7.1, 1.2 Hz, 2H, 2CH Ar), 7.32 – 7.24 (m, 4H, 4CH Ar), 6.61 (t, J = 12.3 Hz, 1H, CH=CH), 6.28 (d, J = 13.6 Hz, 2H, 2CH=CH), 4.74 (dd, J = 13.5, 3.6 Hz, 1H, H6a), 4.51 (dd, J = 13.5, 10.0 Hz, 1H, H6b), 4.43 (s, 2H, H3'ab), 4.09 (t,

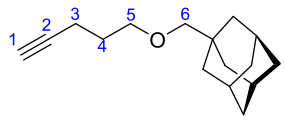
J = 7.6 Hz, 2H, H9'ab), 3.63 (s, 3H, NCH_3 Cy5), 3.59 (d, *J* = 7.8 Hz, 1H, H2), 3.15 – 3.08 (m, 2H, H4 and H3), 2.51 (ddd, *J* = 11.7, 8.4, 6.3 Hz, 1H, NCH*H* octyl), 2.28 – 2.21 (m, 3H, H5 and H5'ab), 1.87 – 1.80 (m, 3H, NCH*H* octyl and H8'ab), 1.74 – 1.68 (m, 14H, H6'ab and $4CH_3$ Cy5), 1.57 – 1.43 (m, 6H, H1, H7, H7'ab and CH_2 octyl), 1.35 – 1.23 (m, 10H, $5CH_2$ octyl), 0.88 (t, *J* = 7.1 Hz, 3H, CH_3 octyl). ^{13}C NMR (214 MHz, MeOD) *δ* 175.7, 175.5, 174.6 (C4' and C=N and *C*=CH), 155.6, 155.5 (2CH=CH), 146.4 (C2'), 144.3, 143.6, 142.6, 142.5 ($4C_q$ Ar), 129.8, 129.7, 126.6, 126.3, 126.2 (CH Ar and CH=CH), 124.6 (C1'), 123.4, 123.3, 112.0, 111.9, 104.4, 104.2 (CH Ar and CH=CH), 79.0 (C3), 74.0 (C2), 70.1 (C4), 62.1 (NCH_2 octyl), 52.0 (C6), 50.5 (C_q), 50.4 (C_q), 45.8 (C1), 44.9 (C5), 44.8 (C9'), 42.4 (C7), 36.5 (C5'), 35.7 (C3'), 33.0 (CH_2 octyl), 31.5 (NCH_3 Cy5), 30.7, 30.6, 30.4, 28.5 ($4CH_2$ octyl), 28.1 (C8'), 28.0 ($2CH_3$ Cy5), 27.8 ($2CH_3$ Cy5), 27.3 (C7'), 26.4 (C6'), 23.7 (CH_2 octyl), 14.5 (CH_3 octyl). HRMS (ESI) m/z: $[M]^+$ calc for $C_{50}H_{70}N_7O_4$ 832.54838, found 832.54775.

Compound S3



[1,1'-biphenyl]-4-ylmethanol (191 mg, 1.04 mmol) was dissolved in dry DMF (5 mL) and cooled to 0 °C. NaH (24 mg, 0.62 mmol) was added and the mixture was stirred for 10 min at 0 °C, followed by 30 min at rt. After cooling to 0 °C again, 5-bromopent-1-yne (0.10 g, 0.52 mmol) was added and the mixture was stirred at rt for 2 h. The reaction was quenched with H_2O at 0 °C and further diluted with EtOAc and H_2O. The layers were separated and the aqueous layer extracted with EtOAc. The combined organic layers were washed with H_2O (2 x) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 100:0→100:1) affording compound **S3** (44 mg, 0.18 mmol, 35%) as a light yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.62 – 7.54 (m, 4H, 4CH Ar), 7.48 – 7.38 (m, 4H, 4CH Ar), 7.38 – 7.29 (m, 1H, CH Ar), 4.55 (s, 2H, H6ab), 3.61 (t, *J* = 6.2 Hz, 2H, H5ab), 2.34 (td, *J* = 7.1, 2.7 Hz, 2H, H3ab), 1.94 (t, *J* = 2.7 Hz, 1H, H1), 1.85 (tt, *J* = 7.2, 6.2 Hz, 2H, H4ab). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 141.1, 140.7, 137.6 ($3C_q$ Ar), 128.9, 128.2, 127.4, 127.3, 127.2 (CH Ar), 84.1 (C1), 72.9 (C6), 68.9 (C5), 68.6 (C2), 28.8 (C4), 15.5 (C3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{18}H_{19}O$ 251.14304, found 251.14282.

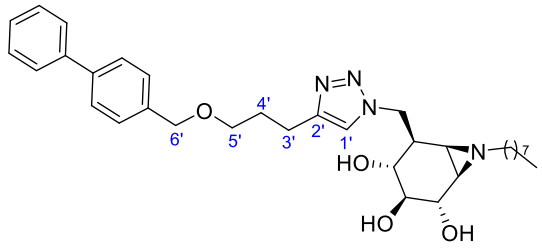
Compound S4



1-adamantanemethanol (173 mg, 1.04 mmol) was dissolved in dry DMF (5 mL) and cooled to 0 °C. NaH (24 mg, 0.62 mmol) was added and the mixture was stirred for 10 min at 0 °C, followed by 30 min at rt. After cooling to 0 °C again, 5-bromopent-1-yne (0.10 g, 0.52 mmol) was added and the mixture was stirred at rt for 6 h. The reaction was quenched with H_2O at 0 °C and further diluted with EtOAc and H_2O. The layers were separated and the aqueous layer extracted with EtOAc. The combined organic layers were washed with H_2O (2 x) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc, 100:0→50:1) affording compound **S4** (25 mg,

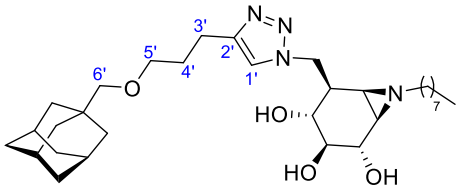
0.11 mmol, 21%) as a light yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 3.46 (t, J = 6.1 Hz, 2H, H5ab), 2.97 (s, 2H, H6ab), 2.29 (td, J = 7.2, 2.7 Hz, 2H, H3ab), 1.98 – 1.91 (m, 4H, H1 and 3CH), 1.81 – 1.74 (m, 2H, H4ab), 1.74 – 1.68 (m, 3H, 3CH*H*), 1.67 – 1.61 (m, 3H, 3CH*H*), 1.53 (d, J = 2.9 Hz, 6H, 6CH*H*). ^{13}C NMR (126 MHz, $CDCl_3$) δ 84.5 (C1), 82.1 (C6), 69.9 (C5), 68.3 (C2), 39.9 ($3CH_2$), 37.4 ($3CH_2$), 34.3 (C_q), 28.8 (C4), 28.5 (3CH), 15.4 (C3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{16}H_{25}O$ 233.1899, found 233.1897.

Compound 8



Compound **11** (2.1 mg, 6.7 µmol) was dissolved in DMF (0.3 mL) and biphenyl alkyne (2.1 mg, 8.1 µmol, dissolved with DMF, 0.1 mL in total) was added. Then $CuSO_4$ (0.1 M in H_2O, 13.4 uL, 1.34 µmol, 0.2 eq.) and sodium ascorbate (0.1 M in H_2O, 27 uL, 2.7 µmol, 0.4 eq.) were added and the mixture was stirred at rt for 16 h. Starting material was still observed by LC-MS, thus another portion of $CuSO_4$ (13.5 uL) and sodium ascorbate (27 uL) were added and the mixture was stirred for another 7 h until full conversion was observed by LC-MS. The mixture was diluted with H_2O (10 mL), extracted with EtOAc (2 x 15 mL), the combined organic layers were washed with H_2O (2 x 15 mL), brine (15 mL), dried over $MgSO_4$, filtered and concentrated in vacuo. The product was purified by silica gel column chromatography (DCM/MeOH, 100:1→20:1) affording compound **8** (3.0 mg, 5.3 µmol, 79%) as a white powder after lyophilization. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.62 – 7.55 (m, 4H, 4CH Ar), 7.46 – 7.38 (m, 4H, 4CH Ar), 7.36 – 7.32 (m, 1H, CH Ar), 7.31 – 7.29 (m, 1H, H1'), 4.59 (dd, J = 13.6, 5.3 Hz, 1H, H6a), 4.54 (s, 2H, H6'ab), 4.48 (dd, J = 13.6, 8.6 Hz, 1H, H6b), 3.97 (brs, 1H, OH), 3.83 (s, 1H, H2), 3.55 (t, J = 6.2 Hz, 2H, H5'ab), 3.44 (brs, 1H, OH), 3.36 (dd, J = 5.4, 2.2 Hz, 2H, H3 and H4), 3.27 (brs, 1H, OH), 2.84 (t, J = 7.6 Hz, 2H, H3'ab), 2.49-2.35 (m, 2H, H5 and NCH*H*), 2.06 – 1.92 (m, 3H, H4'ab and NCH*H*), 1.72 (d, J = 6.0 Hz, 1H, H1), 1.61 (dd, J = 6.0, 4.0 Hz, 1H, H7), 1.55-1.44 (m, 2H, CH_2), 1.34-1.16 (m, 10H, $5CH_2$), 0.87 (t, J = 6.5 Hz,, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 147.8 (C2'), 141.0, 140.7, 137.7 ($3C_q$ Ar), 128.9, 128.4, 127.5, 127.3, 127.2 (9CH Ar), 122.0 (C1'), 76.4 (C3), 72.8 (C6'), 71.6 (C2), 69.5 (C5'), 69.3 (C4), 61.0 (NCH_2), 51.4 (C6), 43.7 (C1), 43.0 (C5), 40.5 (C7), 32.0, 29.7, 29.7, 29.5, 29.4, 27.5, 22.8, 22.5 (C4', C3' and $6CH_2$), 14.3 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{33}H_{47}N_4O_4$ 563.35918, found 563.35902.

Compound 9



Compound **11** (2.1 mg, 6.7 µmol) was dissolved in DMF (0.3 mL) and adamantane alkyne (1.9 mg, 8.1 µmol, dissolved with DMF, 0.1 mL in total) was added. Then $CuSO_4$ (0.1 M in H_2O, 13.4 uL, 1.34 µmol, 0.2 eq.) and sodium ascorbate (0.1 M in

H_2O, 27 uL, 2.7 μmol, 0.4 eq.) were added and the mixture was stirred at rt for 16 h. Starting material was still observed by LC-MS, thus another portion of $CuSO_4$ (13.5 uL) and sodium ascorbate (27 uL) were added and the mixture was stirred for another 9 h until full conversion was observed by LC-MS. The mixture was diluted with H_2O (10 mL), extracted with EtOAc (2 x 15 mL), the combined organic layers were washed with H_2O (2 x 15 mL), brine (15 mL), dried over $MgSO_4$, filtered and concentrated in vacuo. The product was purified by silica gel column chromatography (DCM/MeOH, 100:1→20:1) and by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN) affording compound **9** (1.4 mg, 2.6 μmol, 39%) as a white powder after lyophilization. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (s, 1H, H1'), 4.63 (dd, J = 13.6, 5.3 Hz, 1H, H6a), 4.53 (dd, J = 13.6, 8.6 Hz, 1H, H6b), 3.93 – 3.63 (m, 2H, H2 and OH), 3.45-3.38 (m, 4H, H3, H4 and H5'ab), 3.38-3.06 (brs, 2H, 2OH), 2.97 (s, 2H, H6'ab), 2.80 (t, J = 7.7 Hz, 2H, H3'ab), 2.55-2.48 (m, 1H, H5), 2.45 (dt, J = 11.5, 7.3 Hz, 1H, NCH*H* octyl), 2.05 – 1.86 (m, 4H, NCH*H* octyl and 3CH), 1.79 – 1.56 (m, 10H, H1, H7, H4'ab and 3CH_2), 1.56-1.47 (m, 8H, 4CH_2), 1.34 – 1.19 (m, 10H, 5CH_2), 0.92 – 0.85 (t, J = 5.0 Hz, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 148.1 (C2'), 122.0 (C1'), 82.1 (C6'), 76.3, 71.6 (C2), 70.6 (C5'), 69.4, 61.0 (NCH_2 octyl), 51.3 (C6), 43.7 (C1), 43.0 (C5), 40.5 (C7), 39.9 (3CH_2), 37.4 (3CH_2), 34.3 (C_q), 32.0, 29.7, 29.7, 29.4, 29.4 (4CH_2 octyl and C4'), 28.4 (3CH), 27.5 (CH_2 octyl), 22.8 (C3'), 22.5 (CH_2 octyl), 14.3 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{31}H_{53}N_4O_4$ 545.40613, found 545.40594.

4.6 References

[1] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., *Nucleic Acids Res.* **2014**, *42* (D1), D490-D495.

[2] Brady, R. O.; Kanfer, J. N.; Shapiro, D., *J. Biol. Chem.* **1965**, *240*, 39–43.

[3] Beutler, E., *Science* **1992**, *256*, 794–799.

[4] Futerman, A. H.; Platt, F. M., *Mol. Genet. Metab.* **2017**, *120*, 22–26.

[5] Brady, R. O.; Kanfer, J. N.; Shapiro, D., *Biochem. Biophys. Res. Commun.* **1965**, *18*, 221–225.

[6] Brady, R. O.; Kanfer, J. N.; Bradley, R. M.; Shapiro, D., *J. Clin. Invest.* **1966**, *45*, 1112–1115.

[7] Stirnemann, J.; Belmatoug, N.; Camou, F.; Serratrice, C.; Froissart, R.; Caillaud, C.; Levade, T.; Astudillo, L.; Serratrice, J.; Brassier, A.; Rose, C.; Billette de Villemeur, T.; Berger, M., *Int. J. Mol. Sci.* **2017**, *18*, 441–470.

[8] Baris, H. N.; Cohen, I. J.; Mistry, P. K., *Pediatr Endocrinol Rev.* **2014**, *12*, 72–81.

[9] Sidransky, E., *Mol. Genet. Metab.* **2004**, *83*, 6–15.

[10] Kolodny, E. H.; Ullman, M. D.; Mankin, H. J.; Raghavan, S. S.; Topol, J.; Sullivan, J. L., *Prog. Clin. Biol. Res.* **1982**, *95*, 33–65.

[11] Erikson, A.; Bembi, B.; Schiffmann, R., *Baillieres. Clin. Haematol.* **1997**, *10*, 711–723.

[12] Nilsson, O.; Svennerholm, L., *J. Neurochem.* **1982**, *39*, 709–718.

[13] Aharon-Peretz, J.; Badarny, S.; Rosenbaum, H.; Gershoni-Baruch, R., *Neurology* **2005**, *65*, 1460–1461.

[14] Sato, C.; Morgan, A.; Lang, A. E.; Salehi-Rad, S.; Kawarai, T.; Meng, Y.; Ray, P. N.; Farrer, L. A.; St George-Hyslop, P.; Rogaeva, E., *Mov. Disord.* **2005**, *20*, 367–370.

[15] Stojkovska, I.; Krainc, D.; Mazzulli, J. R., *Cell Tissue Res.* **2018**, *373*, 51–60.

[16] Marugan, J. J.; Zheng, W.; Motabar, O.; Southall, N.; Goldin, E.; Westbroek, W.; Stubblefield, B. K.; Sidransky, E.; Aungst, R. A.; Lea, W. A.; Simeonov, A.; Leister, W.; Austin, C. P., *J. Med. Chem.* **2011**, *54*, 1033–1058.

[17] Hill, T.; M. Tropak, B.; Mahuran, D.; Withers, S. G., *ChemBioChem* **2011**, *12*, 2151–2154.

[18] Goddard-Borger, E. D.; Tropak, M. B.; Yonekawa, S.; Tysoe, C.; Mahuran, D. J.; Withers, S. G., *J. Med. Chem.* **2012**, *55*, 2737–2745.

[19] Martínez-Bailén, M.; Carmona, A. T.; Patterson-Orazem, A. C.; Lieberman, R. L.; Ide, D.; Kubo, M.; Kato, A.; Robina, I.; Moreno-Vargas, A. J., *Bioorg. Chem.* **2019**, *86*, 652–664.

[20] Street, I. P.; Kempton, J. B.; Withers, S. G., *Biochemistry* **1992**, *31*, 9970–9978.

[21] Schröder, S. P.; Wu, L.; Artola, M.; Hansen, T.; Offen, W. A.; Ferraz, M. J.; Li, K.-Y.; Aerts, J. M. F. G.; Van Der Marel, G. A.; Codée, J. D. C.; Davies, G. J.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2018**, *140*, 5045–5048.

[22] Artola, M.; Kuo, C.-L.; Lelieveld, L. T.; Rowland, R. J.; van der Marel, G. A.; Codée, J. D. C.; Boot, R. G.; Davies, G. J.; Aerts, J. M. F. G.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2019**, *141*, 4214–4218.

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

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

[25] 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.; Overkleeft, H. S., *Chem. - A Eur. J.* **2015**, *21*, 10861–10869.

[26] Withers, S. G.; Umezawa, K., *Biochem. Biophys. Res. Commun.* **1991**, *177*, 532–537.

[27] Kuo, C.-L.; van Meel, E.; Kytidou, K.; Kallemeijn, W. W.; Witte, M.; Overkleeft, H. S.; Artola, M.; Aerts, J. M. F. G., *Methods Enzymol.* **2018**, *598*, 217–235.

[28] 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.; Overkleeft, H. S., *Org. Biomol. Chem.* **2014**, *12*, 7786–7791.

[29] Butters, T. D.; Dwek, R. A.; Platt, F. M., *Curr. Top. Med. Chem.* **2003**, *3*, 561–574.

[30] Wennekes, T.; van den Berg, R. J. B. H. N.; Boot, R. G.; van der Marel, G. A.; Overkleeft, H. S.; Aerts, J. M. F. G., *Angew. Chem., Int. Ed.* **2009**, *48*, 8848–8869.

[31] Cordero, F. M.; Bonanno, P.; Chioccioli, M.; Gratteri, P.; Robina, I.; Vargas, A. J. M.; Brandi, A., *Tetrahedron* **2011**, *67*, 9555–9564.

[32] Sorrentino, E.; Connon, S. J., *Org. Lett.* **2016**, *18*, 5204−5207.

[33] Aliouane, L.; Chao, S.; Brizuela, L.; Pfund, E.; Cuvillier, O.; Jean, L.; Renard, P.-Y.; Lequeux, T., *Bioorg. Med. Chem.* **2014**, *22*, 4955–4960.

[34] Gundorph Hansen, F.; Bundgaard, E.; Madsen, R., *Biochem. Biophys. Res. Commun.* **1990**, *43*, 10139–10142.

[35] Li, K.-Y.; Jiang, J.; Witte, M. D.; Kallemeijn, W. W.; Van Den Elst, H.; Wong, C. S.; Chander, S. D.; Hoogendoorn, S.; Beenakker, T. J. M.; Codée, J. D. C.; Aerts, J. M. F. G.; Van Der Marel, G. A.; Overkleeft, H. S., *Eur. J. Org. Chem.* **2014**, *2014*, 6030–6043.

[36] Dvir, H.; Harel, M.; McCarthy, A. A.; Toker, L.; Silman, I.; Futerman, A. H.; Sussman, J. L., *EMBO Rep.* **2003**, *4*, 704–709.

[37] Rowland, R. J.; Wu, L.; Liu, F.; Davies, G. J., *Acta Crystallogr. Sect. D Struct. Biol.* **2020**, *76*, 565–580.

[38] Lahav, D.; Liu, B.; Van Den Berg, R. J. B. H. N.; Van Den Nieuwendijk, A. M. C. H.; T. Wennekes,; Ghisaidoobe, A. T.; Breen, I.; Ferraz, M. J.; Kuo, C.-L.; Wu, L.; Geurink, P. P.; Ovaa, H.; Van Der Marel, G. A.; Van Der Stelt, M.; Boot, R. G.; Davies, G. J.; Aerts, J. M. F. G.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2017**, *139*, 14192–14197.

[39] Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C., *Anal. Biochem.* **1985**, *150*, 76–85.

[40] Winter, G., *J. Appl. Crystallogr.* **2010**, *43*, 186–190.

[41] Evans, P., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 72–82.

[42] Evans, P. R.; Murshudov, G. N., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1204–1214.

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

[44] Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53*, 240–255.

[45] Emsley, P.; Cowtan, K., *Acta Crystallogr D* **2004**, *60*, 2126–2132.

[46] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.

[47] Lebedev, A. A.; Young, P.; Isupov, M. N.; Moroz, O. V.; Vagin, A. A.; Murshudov, G. N., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 431–440.

[48] Agirre, J.; Iglesias-Fernández, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D., *Nat. Struct. Mol. Biol.* **2015**, *22*, 833–834.

[49] Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 12–21.

[50] McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 386–394.

APPENDIX

Table 4.S1. Data collection and Processing

Compound	ABP **7**
PDB Entry	6Z3I
Diffraction source	Diamond Beamline i04
Wavelength (Å)	0.979507
Temperature (K)	100
Detector	Eiger2 XE 16M
Rotation range per image (°)	0.1
Total rotation range (°)	360
Space group	P 2_1
a, *b*, *c* (Å)	53.1, 76.7, 68.0
α, β, γ (°)	90, 102, 90
Resolution range (Å)	88.49-1.80 (1.84-1.80)
Total No. of reflections	348887 (20199)
No. of unique reflections	49541 (2966)
Completeness (%)	100 (100)
Redundancy	7.0 (6.8)
$\langle I/\sigma(I) \rangle$	6.7 (0.8)
$R_{meas.}$	0.22 (2.55)
$CC_{1/2}$	0.99 (0.50)
Overall *B* factor from Wilson plot ($Å^2$)	27

Table 4.S2. Structure solution and refinement

Compound	ABP **7**
PDB Entry	6Z3I
Resolution range (Å)	88.49-1.80 (1.84-1.80)
Completeness (%)	100 (100)
No. of reflections	49519
working set	
No. of reflections	2509
test set	
Final R_{cryst}	0.19
Final R_{free}	0.20
No. of non-H atoms	
Protein	3972
Ligand	193
Water	340
Total	4505
R.m.s. deviations	
Bonds (Å)	0.013
Angles (°)	1.70
Average *B* factors (Å^2)	
Protein	26
Ligand	48
Water	36
Ramachandran plot	
Most favoured (%)	94.7
Allowed (%)	4.2

5

Novel Uronic Acid-Type 1-*N*-Iminosugars Derived from Siastatin B as Competitive Heparanase Inhibitors

5.1 Introduction

Heparan sulfate (HS) is a long linear polysaccharide consisting of repeating disaccharide units formed by a uronic acid (either D-glucuronic acid, GlcA or L-iduronic acid, IdoA) and a D-glucosamine (either *N*-acetylated, GlcNAc or *N*-sulfated, GlcNS) that is modified at various positions by sulfation, yielding domains with high sulfation along the HS chain separated by low/minimal sulfation segments.[1] The structure of heparan sulfate proteoglycans (HSPGs) consists of a core protein to which variable HS chains are covalently attached.[2, 3] HSPGs are mainly localized in the extracellular matrix (ECM), on the cell surface and also in the basement membrane (BM), and contribute significantly to the structural integrity, self-assembly and insolubility of the ECM and BM.[4, 5] Moreover, HSPGs provide a reservoir for the binding of numerous bioactive molecules such as growth factors, cytokines, chemokines and enzymes, with specific binding events mediated by local features (e.g. sulfation patterns or anionic charges)[6, 7]of HS chains.

Heparanase (HPSE) is the only known mammalian *endo*-β-D-glucuronidase capable of cleaving within HS chains. It catalyses the hydrolysis of internal β-1,4-linked glycosidic bonds between GlcA and GlcNS (Figure 5.1) at a limited number of sites. HPSE processes its substrates through a Koshland double displacement mechanism, generating smaller HS fragments with net retention of the anomeric configuration.[8, 9] Under normal physiological conditions, HPSE is expressed at high levels in placenta and a few cell types such as platelets, neutrophils and mast cells, and is poorly expressed in other human tissues. However, upregulated expression and increased enzymatic activity of HPSE have been reported in multiple pathological conditions,[10-12] particularly in cancer progression and inflammation disorders where HPSE-mediated degradation of HS and the consequent release of HS-sequestered molecules enables remodelling of the ECM network and facilitates cell migration, invasion and proliferation.

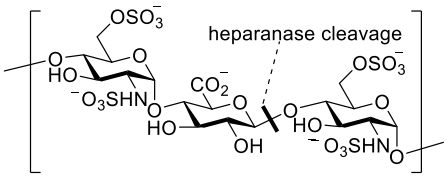


Figure 5.1. Heparanase cleavage site within the structure of heparan sulfate.

The pivotal role of HPSE in the establishment and development of numerous diseases has made it an attractive pharmacological target and several different classes of inhibitors have been developed over the past two decades, ranging from polysulfated oligo- and polysaccharides,

nucleic acids and monoclonal antibodies to small-molecule inhibitors.[13-16] Amongst these, heparin/HS mimetics are the most intensively studied ones and four compounds (PI-88, SST0001, M402, and PG545)[17-20] of this class have been tested in clinical trials for various types of cancer. However, these compounds share some limitations related to their high-molecular weight nature, heterogeneous structure and parenteral administration, which may limit their potential therapeutic application. The discovery of small molecules endowed with anti-HPSE activity would provide another alternative to the development of novel therapeutic drugs because of their more conducive pharmacokinetic properties and ease of optimization for oral administration.[15, 21]

Iminosugars are monosaccharide analogues which contain a nitrogen atom instead of an oxygen in the ring.[22] Many compounds of this type, such as 1-deoxynojirimycin, are potent and selective inhibitors of enzymes associated with carbohydrate metabolism,[23, 24] and have been investigated as potential therapeutics for tumor metastasis, diabetes, lysosomal storage disorders and other diseases.[25-27] Siastatin B, a natural product originally isolated from a *Streptomyces* culture by Umezewa *et al*.,[28] is a *gem*-diamine 1-*N*-iminosugar in which the anomeric carbon is replaced by nitrogen[29] (Figure 5.2A). It effectively inhibits neuraminidases as well as β-D-glucuronidases and *N*-acetyl-β-D-glucosaminidases, presumably because it structurally resembles *N*-acetylneuraminic acid, β-D-glucuronic acid and *N*-acetyl-β-D-glucosamine. Upon achievement of the total synthesis[30] of siastatin B, Nishimura and co-workers further reported the synthesis of a series of siastatin B derivatives characterized by a trifluoroacetamido substituent at the 2 position of the piperidine scaffold (the ring numbering is depicted in siastatin B).[31, 32] These 2-trifluoroacetamido derivatives **2**-**4** (Figure 5.2A) were all shown to be very potent inhibitors of bovine liver β-D-glucuronidase and also showed moderate inhibition of recombinant human HPSE. NMR analyses of compound **2** in the media of enzyme assays have suggested that it can undergo pH-dependent Amadori rearrangement in solution, yielding a hemiaminal/hydrated ketone that may act as the true inhibitor (Figure 5.2B).[33] This rearrangement starts by elimination of the 2-trifluoroacetamido group that can be followed by hydration of the resulting iminium to form hemiaminal **5**. Enolization of **6** can give enol **7**, which can further tautomerize to ketone **8**, followed by addition of a H_2O molecule to give hydrate **9**. This solvent-mediated rearrangement has not been demonstrated for siastatin B, which appears to be relatively stable in aqueous solutions.

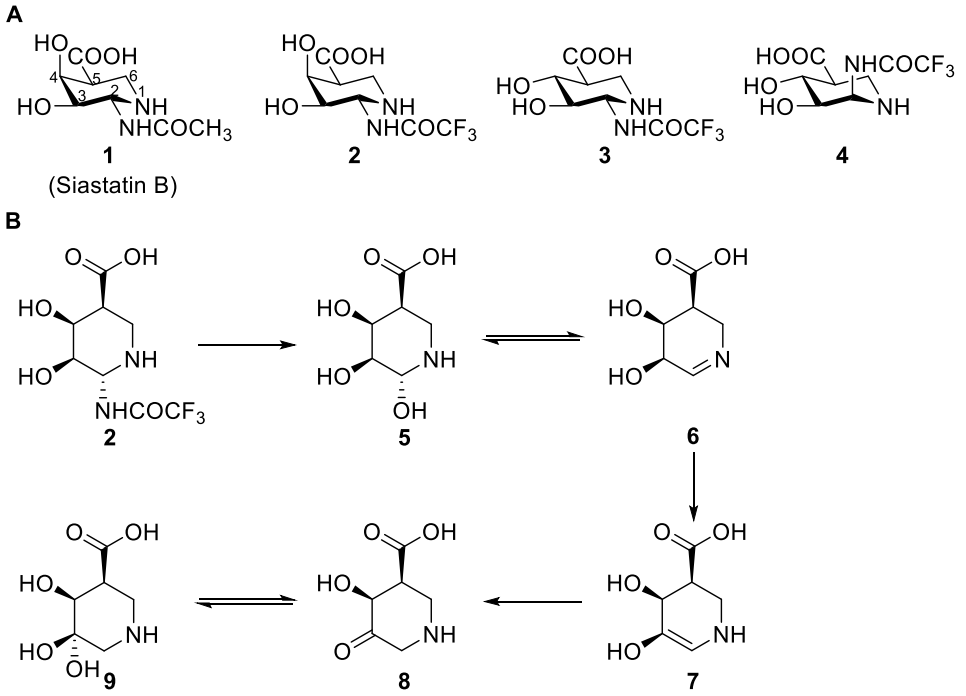


Figure 5.2. A) Chemical structures of siastatin B (**1**) and its 2-trifluoroacetamido derivatives **2**-**4**; B) Transformation of D-galacturonic acid-type gem-diamine 1-*N*-iminosugar **2** into hydrate **9** in aqueous acetate buffer (pH 5.0) as proposed by Nishimura and co-workers.

The crystal structural of human HPSE has recently been resolved by Wu *et al.*,[34] enabling structure-based design of HPSE inhibitors. Surprisingly, preliminary X-ray structures of human HPSE treated with siastatin B, as conducted by Liang Wu (University of York), has indicated that it is hydrate **9**, and not siastatin B, that occupies the enzyme active site. To establish the mode of action by which siastatin B inhibits heparanases and β-glucuronidases, X-ray crystallographic analysis of co-complexes between siastatin B and different enzymes are presented in this chapter. Furthermore, in order to understand the action of the breakdown products and their HPSE inhibitory properties, a panel of uronic acid-type 1-*N*-iminosugars **8**-**14** was synthesized (Figure 5.3). Besides the *galacto*-configured ketone **8** that may be formed by degradation of siastatin B, the *gluco*-configured analogue **12** was considered as well since it more closely mimics the glucose configuration of the natural HPSE substrates: glucuronic acid. Previous work on endo-glycosidase inhibitors as described in this thesis (Chapters 2 and 3) and elsewhere[35-37] have demonstrated that endo-glycosidases have larger active sites to accommodate elongation at the non-reducing end sugars where the polysaccharide chain continues – here the O4 position. In line with this observation, O4-methylated compounds **10** and **13** were included in the panel as well, with the aim to achieve selectivity for HPSE over *exo*-β-D-glucuronidases. Additionally, uronic acid-type isofagomine **14** has previously been

reported as a potent inhibitor of β-D-glucuronidase,[38-40] and comprises the reduced (keto/hydrate to alcohol) derivative of compound **2**. Since its inhibitory activity towards HPSE has yet to be explored, compound **14** as well as its *galacto*-configured isomer **11** were included in the panel.

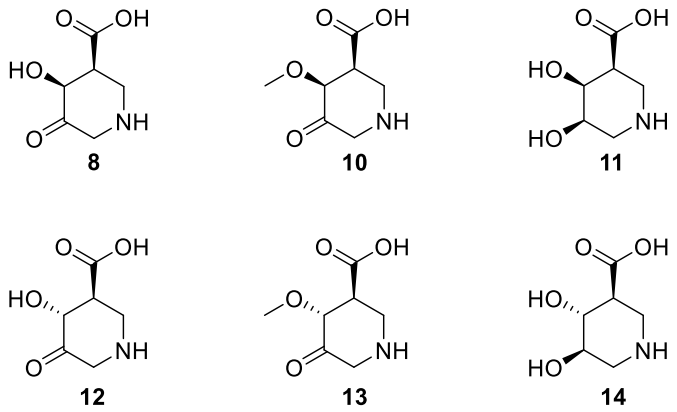


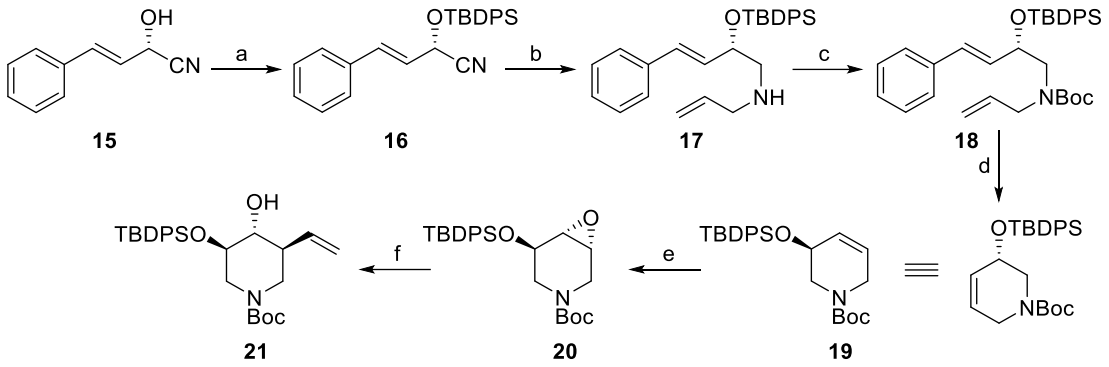
Figure 5.3. Structures of D-*galacto*- and D-*gluco*-uronic acid-type 1-*N*-iminosugars which are subject of the studies described in this chapter.

5.2 Results and discussion

5.2.1 Synthesis of D-glucuronic acid-type 1-*N*-imingugars

The synthesis of glucuronic acid-type 1-*N*-imingugars commenced with the preparation of key intermediate **21** (Scheme 5.1). Transformation of **15** into **19** was achieved using adaptations of the procedures described by Jiang *et al*. for the construction of 1-deoxy-L-fuconojirimycin.[41] Starting from enantiomerically pure cyanohydrin **15**, prepared employing (*S*)-hydroxynitrile lyase from the *Hevea brasiliensis* rubber tree,[42] the secondary alcohol was silylated to give **16** in excellent yield. Conversion of **16** via a one-pot DIBAL-H reduction-transimination-sodium borohydride reduction cascade sequence, using commercially available allylamine, followed by subsequent *N*-Boc protection, afforded compound **18** in 81% yield over two steps. Ring-closing metathesis of **18** using Grubbs' catalyst provided heterocyclic alkene **19**. The yield (45%) is however relatively low compared to those (88%–97%) of previously reported analogues,[41] possibly due to the higher reaction concentration (0.2 M instead of 0.1 M) applied here, which led to the formation of undesired intermolecular metathesis side products. The last two steps in the synthesis route followed strategies as reported by Takahata *et al*.[43] Epoxidation of **19** was performed using methyl(trifluoromethyl) dioxirane (generated *in situ* from 1,1,1-trifluroacetone and oxone[44]), and epoxide **20** was formed preferably due to a favored attack of the dioxirane on the less hindered *anti*-side of the large TBDPS group. Nucleophilic opening of epoxide **20**

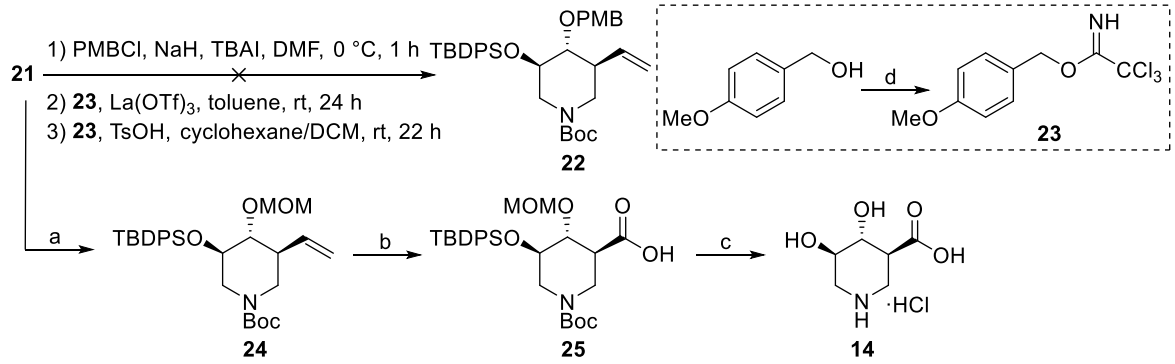
was carried out with *in situ* generated cuprate[45] $(CH_2=CH)_2CuCNLi_2$ in the presence of Lewis acid at low temperature for 2 h, affording product **21** in 64% yield.



Scheme 5.1. Synthesis of key intermediate **21**. Reagents and conditions: a) TBDPSCl, imidazole, DMF, 0 °C to rt, 98%; b) *i*) DIBAL-H, Et_2O, -78 °C to 10 °C; *ii*) MeOH, -90 °C; *iii*) allylamine, NaOMe, rt; *iv*) $NaBH_4$, 0 °C to rt; c) Boc_2O, Et_3N, DCM, rt, 81% over two steps; d) Grubbs 1st generation, DCM, reflux, 45%; e) oxone, CH_3COCF_3, $NaHCO_3$, EDTA, MeCN, H_2O, 0 °C, 60%; f) *n*-butyllithium, tetravinyltin, $BF_3{\cdot}Et_2O$, Et_2O, -78 °C, 2 h, 64%.

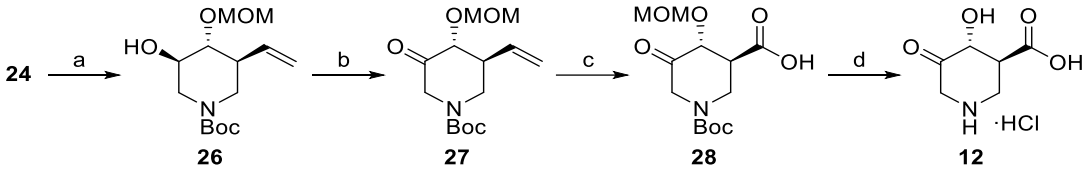
Glucuronic acid-type isofagomine **14** was synthesized by Ichikawa *et al.* starting from commercially available D-arabinose in 13 steps.[38] Here, an alternative synthesis route starting from the vinyl intermediate **21** is described (Scheme 5.2). It was envisioned that protecting the secondary alcohol in **21** with an acid-labile group would facilitate its removal together with the Boc and TBDPS groups in the final stage of the synthesis, therefore a 4-methoxybenzyl (PMB) group was first chosen for this purpose. However, treatment of **21** with PMBCl in the presence of sodium hydride (NaH) at 0 °C for 1 h only led to the formation of a desilylated side product. Other attempts by reacting **21** with freshly prepared PMB-imidate **23** using catalytic amounts of acid[46-48] proved unproductive as well. As an alternative, protection of the hydroxyl group as the methoxymethyl (MOM) ether was investigated. While treatment of **21** with an excess of MOMCl (7.0 eq.) in the presence of DIPEA (8.0 eq.) proved unproductive when conducted at room temperature, heating the reaction mixture to reflux (40 °C) overnight afforded product **24** in good yield. It was found that performing the reaction in a sealed microwave tube, which could prevent the evaporation of solvent (CH_2Cl_2) and the low-boiling MOMCl during the refluxing process, allowed to achieve full conversion of the starting material. Oxidative cleavage of the terminal alkene to the corresponding carboxylic acid was best achieved in one-pot by using ruthenium tetraoxide (generated *in situ* from ruthenium chloride and sodium periodate) in a solvent system of $CCl_4/CH_3CN/H_2O$ as developed by Sharpless and co-workers,[49] resulting in the isolation of **25** in 64% yield. Attempts to replace (toxic) CCl_4 by

EtOAc or CH_2Cl_2 led to incomplete conversion even after prolonged reaction times (15 h), and the corresponding vicinal diol and aldehyde intermediates (from **24**) were observed as the main products by TLC analysis. Ultimately, all of the protecting groups in **25** were removed by treatment with 3 M HCl under reflux for 4 h, yielding product **14** in excellent yield, and all spectroscopic data obtained for **14** proved to be in agreement with those reported in the literature.[38-40]



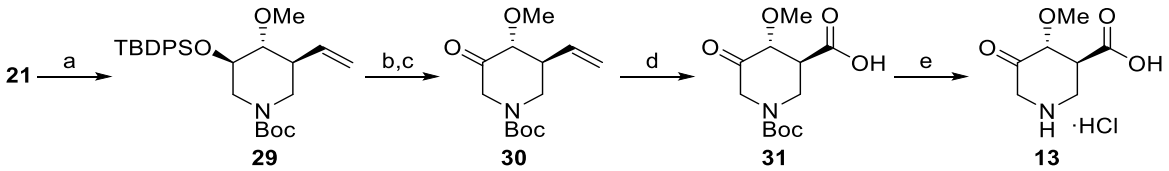
Scheme 5.2. Synthesis of glucuronic acid-type isofagomine **14**. Reagents and conditions: a) MOMCl, DIPEA, DCM, reflux, 87%; b) $RuCl_3 \cdot 3H_2O$ (10 mol%), $NaIO_4$ (5.0 eq.), $CCl_4/CH_3CN/H_2O$, rt, 3 h, 64%; c) 3 M HCl in H_2O, dioxane, 100 °C, 4 h, 95%; d) CCl_3CN, NaH, Et_2O, 0 °C to rt, 2 h, 98%.

Next, the synthesis of **12** from intermediate **24** was investigated (Scheme 5.3). The TBDPS group in **24** could be readily removed by tetrabutylammonium fluoride (TBAF) in THF, resulting in the formation of alcohol **26**, which was then oxidized to ketone **27** with Dess-Martin periodinane (DMP). Oxidative cleavage of the terminal alkene proceeded smoothly using the same conditions as described for the preparation of **25**. Complete conversion of the starting material was achieved after 2 h and carboxylic acid **28** was observed as the single product by TLC and LC-MS analyses. However, extraction of the product from the reaction mixture proved to be troublesome after quenching the reaction with saturated aqueous sodium thiosulfate ($Na_2S_2O_3$). It was found that the reaction mixture had an original pH of around 2 to 3, which after treatment with $Na_2S_2O_3$ solution increased to ~8, pH at which the product became water-soluble. This problem was circumvented by quenching the reaction with isopropanol[32] and the product could be easily extracted into the organic phase. After purification by standard silica gel column chromatography, carboxylic acid **28** was isolated in good yield. Removal of the protecting groups could be accomplished by treatment with 4 M HCl in dioxane/H_2O for 5 h, providing compound **12** as the single product after precipitation from MeOH/Et_2O. Of note, deprotection of **28** under acidic conditions for prolonged reaction times (18 h) resulted in the formation of a complex mixture of several compounds.



Scheme 5.3. Synthesis of compound **12**. Reagents and conditions: a) TBAF, THF, rt, 94%; b) Dess-Martin periodinane, DCM, 0 °C to rt, 93%; c) $RuCl_3 \cdot 3H_2O$ (10 mol%), $NaIO_4$ (5 eq.), $CCl_4/CH_3CN/H_2O$, rt, 2 h, 83%; d) 4 M HCl in dioxane/H_2O, rt, 70%.

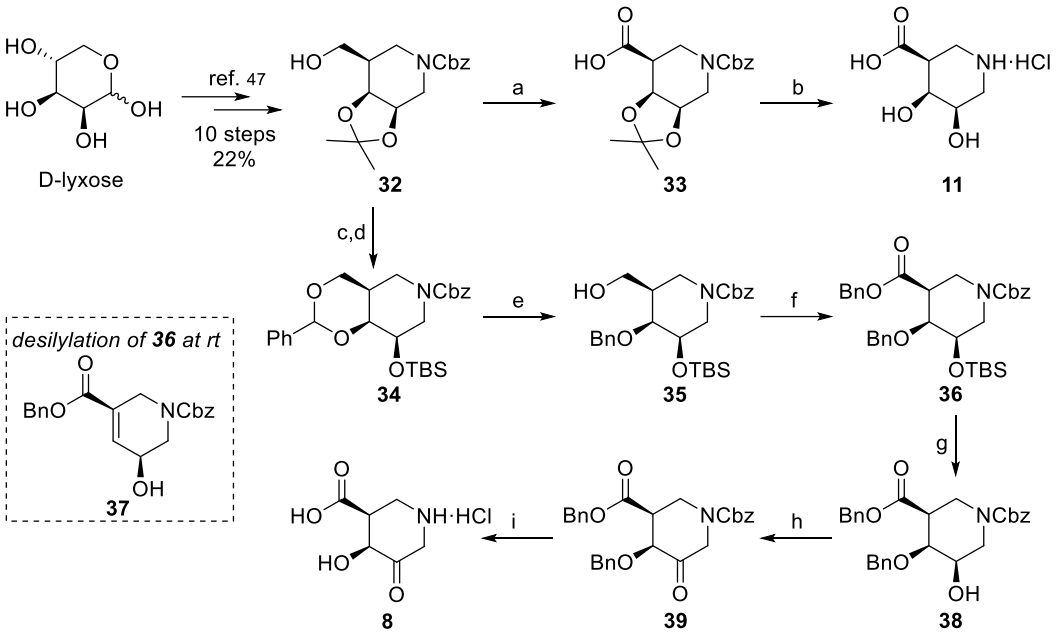
As the next objective, the versatility of alkene **21** as starting material for the synthesis of **13** was investigated (Scheme 5.4). The secondary alcohol in **21** was methylated by reaction with an excess (8.0 eq.) of trimethyloxonium tetrafluoroborate, affording compound **29** in good yield. Desilylation of **29** followed by oxidation resulted in the formation of ketone **30**. Oxidative cleavage of the terminal double bond in **30** had to be performed carefully. Treatment of **30** with 10 mol% of $RuCl_3$ and 5.0 equivalent of $NaIO_4$ at room temperature for 2 h (the same condition as used for the preparation of **28**) resulted in the isolation of carboxylic acid **31** in a low yield (28%). NMR analysis of the crude product revealed the presence of several impurities which were presumably intermediates resulting from incomplete oxidation. However, after prolonging the reaction time to 17 h, compound **31** could not be detected in the crude while the impurities were observed by NMR, which indicates that these impurities might be a result of over-oxidation since RuO_4 is known as a strong oxidant. To confirm this speculation, the reaction was performed with decreased amounts of $RuCl_3$ (5 mol%) and $NaIO_4$ (3.5 eq.) at 0 °C, and the oxidation process was monitored every 15 minutes by TLC analysis and terminated as soon as no intermediates (diol and aldehyde) remained. In this way carboxylic acid **31** was obtained in 75% yield and no over-oxidation side products were detected. The Boc group in **31** was then removed under acidic conditions to give the 4-methylated iminosugar **13** in around 80% purity.



Scheme 5.4. Synthesis of compound **13**. Reagents and conditions: a) trimethyloxonium tetrafluoroborate, proton sponge, 3 Å MS powder, DCM, 0 °C to rt, 75%; b) TBAF, THF, rt, 90%; c) Dess-Martin periodinane, DCM, 0 °C to rt, 83%; d) $RuCl_3 \cdot 3H_2O$ (5 mol%), $NaIO_4$ (3.6 eq.), $CCl_4/CH_3CN/H_2O$, 0 °C for 1 h then rt for 30 min, 75%; e) 0.2 M HCl/HFIP, H_2O, rt, 80% in purity.

5.2.2 Synthesis of D-galacturonic acid-type 1-*N*-iminosugars

The synthesis towards galacturonic acid-type 1-*N*-iminosugars commenced with commercially available D-lyxose that was smoothly converted into key intermediate **32** (Scheme 5.5) in ten steps following procedures described by Ichikawa *et al*. for the preparation of the *N*-Boc protected version of compound **32**.[50] Oxidation of the primary alcohol in **32** with Dess-Martin periodinane gave an aldehyde intermediate that was further oxidized by means of Pinnick oxidation,[51] providing carboxylic acid **33** in 65% yield over two steps. Deprotection of the isopropylidene acetal and the carboxybenzyl (Cbz) group was achieved by palladium catalyzed hydrogenolysis in the presence of acid, affording galacturonic acid-type isofagomine **11** in quantitative yield. Next, the preparation of **8** from common intermediate **32** was investigated. The 3,4-isopropylidene acetal in **32** was removed smoothly under acidic conditions followed by a thermodynamically controlled installation of the benzylidene acetal and subsequent silylation of the remaining 3-hydroxyl group, affording fully protected **34** in 68% yield over three steps. Regioselective cleavage of the benzylidene acetal using BH_3·THF/TMSOTf[52] resulted in the isolation of compound **35** in excellent yield. Jones oxidation of the primary alcohol followed by benzyl protection of the resulting carboxylic acid gave benzyl ester **36** in moderate yield. Removal of the silyl group with TBAF needed to be performed carefully. Treatment of **36** with TBAF at room temperature for prolonged reaction times resulted in the isolation of unsaturated ester **37** which was formed by tetrabutylammonium hydroxide catalyzed β-elimination. This problem could be circumvented by lowering the temperature and shortening the reaction time, providing compound **38** in 94% yield. Oxidation of the secondary alcohol to the ketone followed by global deprotection using palladium catalyst under hydrogen atmosphere finally afforded target 1-*N*-iminosugar **8**.

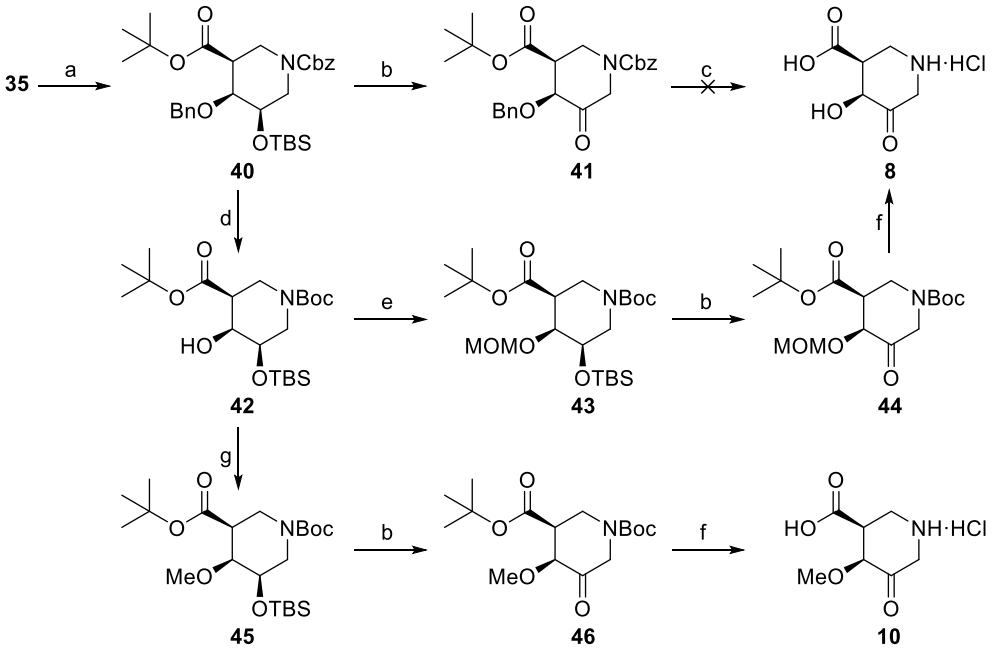


Scheme 5.5. Synthesis of compounds **8** and **11**. Reagents and conditions: a) *i*) Dess-Martin periodinane, DCM, rt; *ii*) $NaClO_2$, NaH_2PO_4, 30% H_2O_2, CH_3CN, H_2O, 0 °C to rt, 65% over two steps; b) H_2, 10% Pd/C, H_3O^+, THF, rt, quant; c) *i*) 8 M HCl, MeOH, rt; *ii*) $PhCH(OMe)_2$, CSA, DMF, 60 °C; d) TBSCl, imidazole, DCM, rt, 68% over three steps; e) $THF{\cdot}BH_3$, TMSOTf, DCM, 0 °C to rt, 97%; f) *i*) Jones reagent, acetone, rt; *ii*) benzyl alcohol, DIC, DMAP, DCM, rt, 46% over two steps; g) TBAF (75 wt% in H_2O), THF, **38**: 0 °C, 5 h, 94%, **37**: rt, overnight, 98%; h) Dess-Martin periodinane, DCM, rt, 71%; i) H_2, 10% Pd/C, THF, H_3O^+, rt, quant.

It was anticipated that installation of a *tert*-butyl ester instead of the benzyl ester in **36** would provide more possibilities for the preparation of the targeted keto-iminosugars. For this purpose, the primary alcohol in **35** was oxidized to a carboxylic acid via a two-step oxidation sequence (Scheme 5.6). Direct esterification of the carboxylic acid with *tert*-butanol under the activation of *N*,*N'*-diisopropylcarbodiimide (DIC) only afforded *tert*-butyl ester **40** in poor yield (30%). LC-MS analysis indicated the presence of a product with a mass corresponding to the acylurea rearrangement product as the major product. Alternatively, esterification of the carboxylic acid with the commercially available *tert*-butyl *N*,*N*'-diisopropylcarbamimidate was investigated and this transformation proceeded smoothly in toluene at 60 °C, affording ester **40** in 71% yield over three steps. Desilylation and subsequent oxidation gave protected ketone **41** in good yield. Unexpectedly, reductive deprotection of **41** by palladium catalyzed hydrogenation under acidic conditions resulted in a complex mixture of compounds.

Inspired by the preparation of the *gluco*-configured keto-iminosugars, a precursor that is fully protected with acid-labile groups was investigated (Scheme 5.6). Treatment of compound

40 with catalytic palladium under hydrogen atmosphere in methanol for a short reaction time (2 h) resulted in selective removal of the Cbz group to afford a free amine, which was subsequently protected with a Boc group. After work-up, the resulting Boc-protected intermediate was directly subjected to a second palladium catalyzed hydrogenolysis for prolonged reaction times (48 h) to remove the benzyl group at O4, affording alcohol **42** in 69% yield over three steps. Protection of the secondary alcohol in **42** with MOMCl was performed in a sealed microwave tube at 100 °C, providing compound **43** in good yield. Subsequent desilylation and oxidation provided ketone **44** which was deprotected under acidic conditions to afford compound **8**, of which the spectroscopic data are in accordance with those prepared by hydrogenolysis of **39** (Scheme 5.5). On the other hand, methylation of **42** with trimethyloxonium tetrafluoroborate as used for the preparation of **29** proved to be troublesome, and compound **45** could only be obtained in very low yield (13%). As an alternative, compound **42** was reacted with iodomethane in the presence of silver oxide and dimethyl sulfide,[53] giving **45** in satisfying yield (69%), which after desilylation, oxidation and final deprotection provided iminosuguar **10**.



Scheme 5.6. Alternative synthetic routes toward the preparation of compound **8** and synthesis of compound **10**. Reagents and conditions: a) *i*) Dess-Martin periodinane, DCM, rt; *ii*) $NaClO_2$, NaH_2PO_4, 30% H_2O_2, CH_3CN, H_2O, 0 °C to rt; *iii*) *tert*-butyl *N*,*N*-diisopropylcarbamimidate, toluene, 60 °C, 5 h, 71% over three steps; b) *i*) TBAF (75 wt% in H_2O), THF, 0 °C, 5 h; *ii*) Dess-Martin periodinane, DCM, rt, **41** 85%, **44** 76%, **46** 73% over two steps; c) H_2, 10% Pd/C, H_3O^+, rt; d) *i*) H_2, 10% Pd/C, MeOH, rt, 2 h; *ii*) Boc_2O, DIPEA, DCM, rt; *iii*) H_2, 10% Pd/C, MeOH, 48 h, 69% over three steps; e) MOMCl,

DIPEA, DCM, 100 °C in microwave tube, 88%; f) concentrated HCl (36%-38%), HFIP, H_2O, rt, **8** quant, **10** quant; g) CH_3I, Ag_2O, $(CH_3)_2S$, THF, rt, 69%.

5.2.3 Structural analysis of siastatin B-enzyme complexes

To determine how siastatin B inhibits both heparanases and β-glucuronidases, the structure of co-complexes between siastatin B and each of human heparanase (HPSE), *A. capsulatum* β-glucuronidase (AcGH79), *Burkholderia pseudomallei* heparanase (BpHep) and *E. coli* β-glucuronidase (EcGusB) were determined (see Table S5.1 for data collection and refinement statistics). Each of the structures contains a ligand present in the active site with full occupancy (Figure 5.4). Three of the four enzymes contain the same compound bound to the enzyme active site: AcGH79, BpHep and EcGusB all have the hemiaminal breakdown product **5** bound in the active site, as opposed to siastatin B (Figure 5.4A-C). Although all three enzymes normally accommodate glucuronic acid at the position where the inhibitor is located, compound **5** appears to bind productively to the active site. The 4-hydroxyl – with an epimeric configuration with respect to glucuronic acid – forms hydrogen bond interactions with the enzyme active site, for both AcGH79 and EcGusB. The 4-hydroxyl of **5** complexed to BpHep does not form a hydrogen bond, however it is exposed to the bulk solvent, and forms a hydrogen bond with an active site water molecule.

The co-complex structure of HPSE and siastatin B, unlike the three other co-complex structures, contains the 3-geminial diol derivative **9** bound to the enzyme active site which is a further breakdown product of **5** (Figure 5.4D). Despite the additional hydroxyl present at the 3-position of **9** this compound appears to be well accommodated in the active site of HPSE. Both hydroxyls present at the 3-position form hydrogen bond interactions with the active site: the hydroxyl on the β-face forms a hydrogen bond with Asp27 and the backbone nitrogen of Thr62 while the hydroxyl on the α-face is positioned to form hydrogen bonds with the catalytic nucleophile Glu343, Gln383 and Asp27. The hydroxyl at the 4-position of inhibitor **9** lacks the hydrogen bond with Trp391 which is observed for *gluco*-configured inhibitors,[54] but it does form hydrogen bonds with water molecules present in the crystal structure.

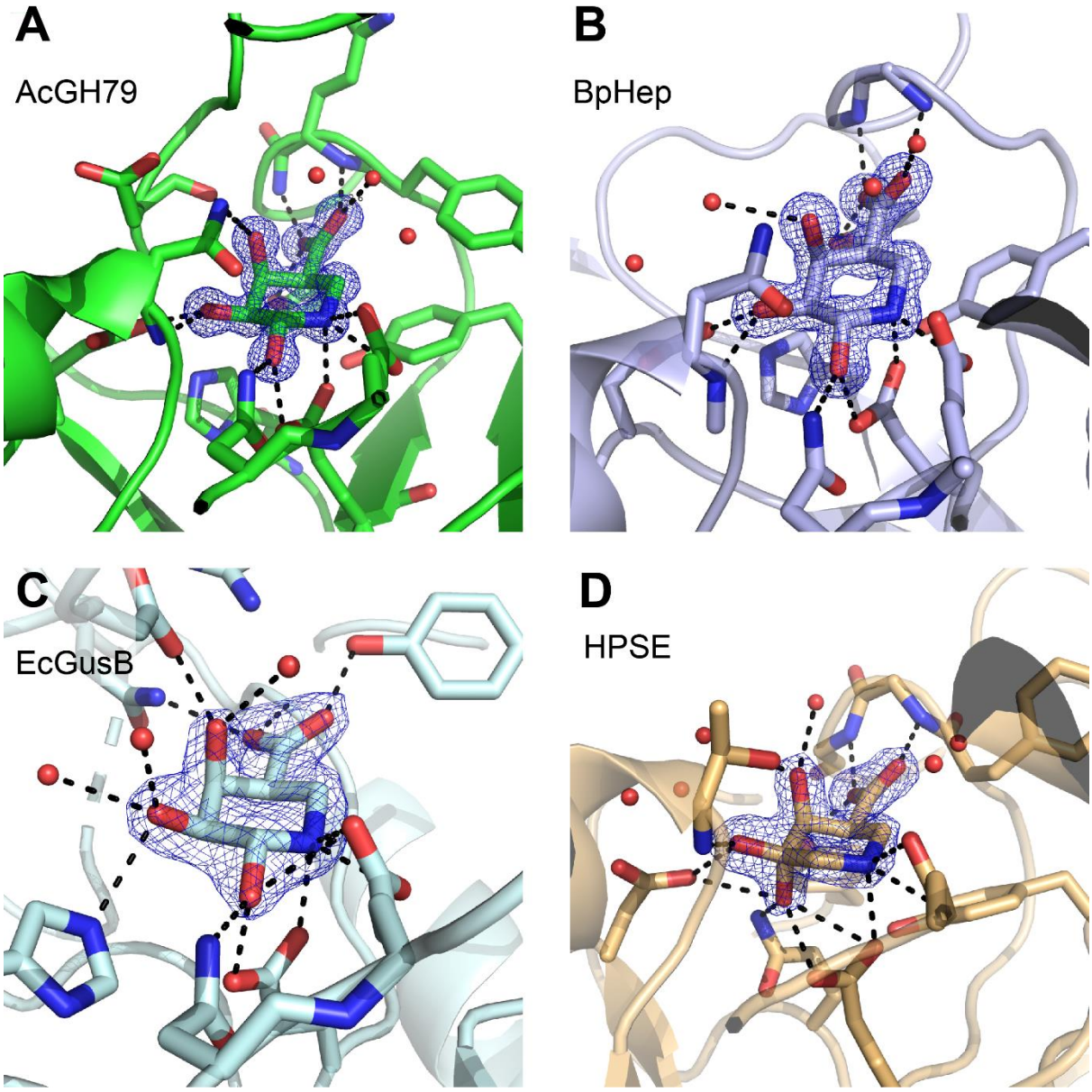


Figure 5.4. Structures of heparanases and β-glucuronidases in complex with siastatin B breakdown products. Siastatin B breakdown product **5** is shown bound to AcGH79 (A), BpHep (B), EcGusB (C) and further breakdown product **9** is shown in HPSE (D). Electron density ($2F_o - F_c$) is shown for the ligand as a blue mesh contoured at 2 σ (AcGH79 = 0.92 $e^-/Å^3$, BpHep = 0.77 $e^-/Å^3$, EcGusB = 0.41 $e^-/Å^3$, HPSE = 0.61 $e^-/Å^3$). The polypeptide is shown in cartoon form with active site residues shown as sticks. Apparent hydrogen bonding interactions are shown as dotted black lines.

5.2.4 Structural basis for enzyme inhibition by synthetic iminosugars

To examine the structural basis for inhibition by the panel of synthetic *gluco*- and *galacto*-configured iminosugars, the co-crystal structures between two of these inhibitors and both HPSE and AcGH79 were determined (Figure 5.5). The uronic acid-type isofagomine **14** was well accommodated in the active sites of both AcGH79 and HPSE. For the co-complex between

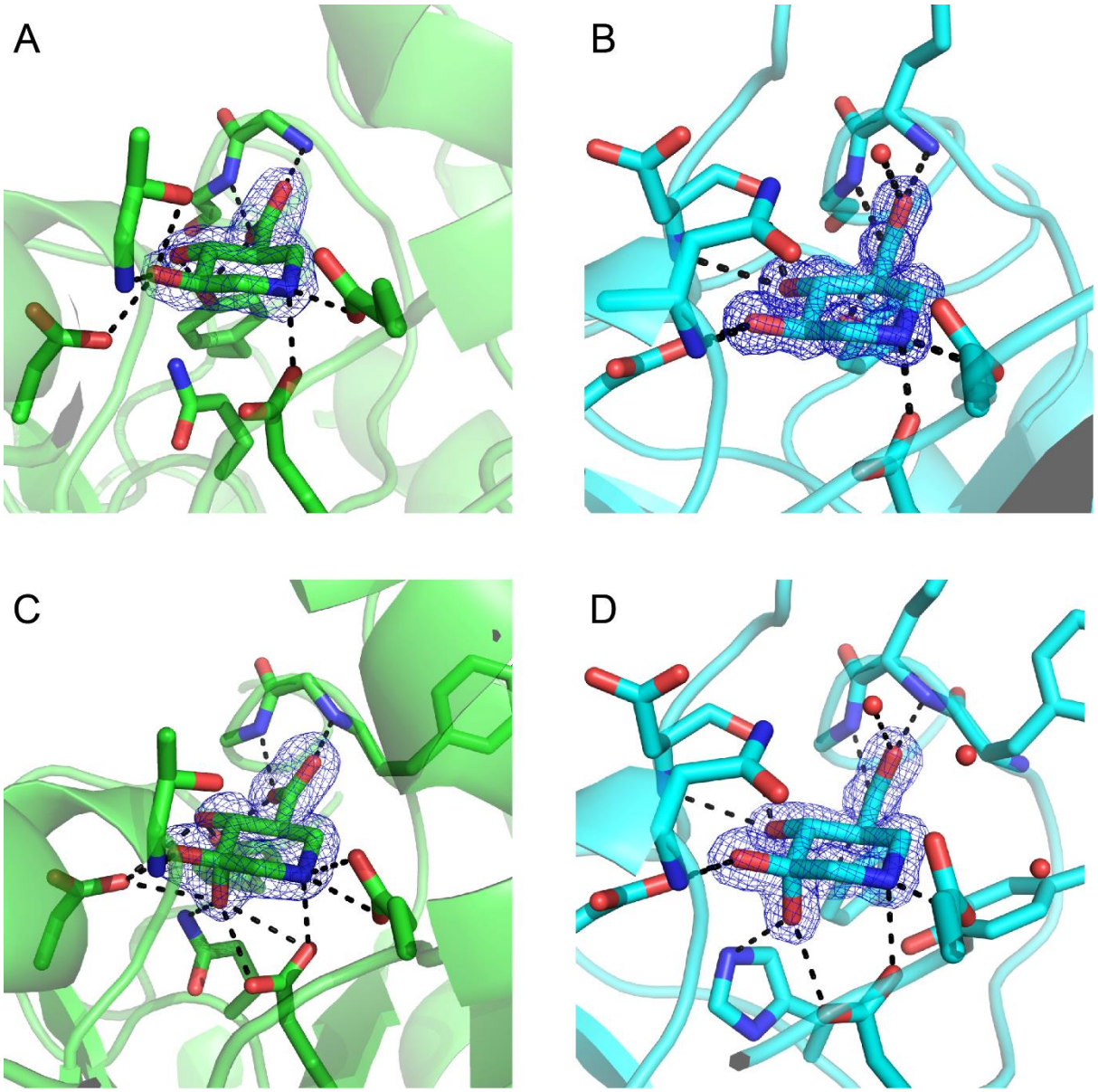


Figure 5.5. Structures of synthetic iminosugars **12** and **14** bound to HPSE and AcGH79. A) Complex between **14** and HPSE; B) Complex between **14** and AcGH79; C) Complex between **12** and HPSE; D) Complex between **12** and AcGH79. Electron density ($2F_o - F_c$) is shown for the ligand as a blue mesh contoured at 2 σ (A = 0.53 $e^-/Å^3$, B = 0.85 $e^-/Å^3$, C = 0.60 $e^-/Å^3$, D = 0.69 $e^-/Å^3$). The polypeptide is shown in cartoon form with active site residues shown as sticks. Apparent hydrogen bonding interactions are shown as dotted black lines.

AcGH79 and **14** interactions between the endocyclic nitrogen, 3-hydroxyl and the carboxylic acid moiety are the same as for the structure with compound **5** (Figure 5.5B and 5.4A). However, as the 4-hydroxyl is in the 'gluco' configuration it is now able to form hydrogen bonds with both the backbone carbonyl and the nitrogen of Asp105, instead of Asn80. The co-complex of **14** with HPSE also shows the anticipated switch in hydrogen bonding interactions for a *gluco*-

configured compound, with the 4-hydroxyl now hydrogen bonding to Tyr391 instead of an active site water (Figure 5.5A).

Ketone **12** was also soaked into pre-formed crystals of HPSE and AcGH79. In both structures the inhibitor is present as the geminal diol (Figure 5.5C and D). This mirrors compound **9** in the active site of HPSE when it was soaked with siastatin B. The positioning of the **12**-derived geminal diol is nearly identical to the structure with **9**, with the only exception being the hydrogen bonding of the 4-hydroxyl to Tyr391 (Figure 5.5C). The **12**-derived geminal diol in the active site of AcGH79 has the same hydrogen bonding network as the structure containing **14** and additional hydrogen bonds between the hydroxyl on the 'α-face' of 3-position with the nucleophile (Glu287) and His327 (Figure 5.5D).

5.3 Conclusion

This chapter describes the in-depth structural analysis of co-complexes between siastatin B and four kinds of β-D-glucuronidases including HPSE, AcGH79, BpHep and EcGusB, showing that siastatin B, in contrast to previous reports describing this to be a stable inhibitor, can provide hemiaminal **5** and geminal diol **9**, that are responsible for enzyme inhibition. In particular, binding of geminal diol **9** in the active site of HPSE suggests that this new 1-*N*-iminosugar may serve as a potent inhibitor of HPSE. In line with this consideration, the synthesis of a panel of *galacto*- and *gluco*-configured 1-*N*-iminosugar derivatives of compound **9** is presented in this chapter. The synthesis of the *gluco*-configured iminosugars comprises the preparation of a common intermediate that is protected with acid-labile groups, followed by oxidation and global deprotection under acidic conditions. The key acid-labile protecting group strategy can also be applied to the synthesis of the *galacto*-configured iminosugars **8** and **10**. In addition, an alternative benzyl-protection route is also described for the preparation of compound **8**, demonstrating that these ketone iminosugars are compatible with hydrogenolysis conditions as well.

Structural analysis for enzyme inhibition by the synthetic *gluco*-configured iminosugars showed that both compounds **12** and **14** are well accommodated in the active sites of HPSE and AcGH79. Compound **12** bound in the active sites of both enzymes in a hydrated form similar to compound **9** that was derived from siastatin B when it was soaked with HPSE. In the future, the inhibition potency of this panel of synthetic iminosugars toward HPSE can be evaluated, which will provide useful information for the development of potent HPSE inhibitors as potential therapeutics.

5.4 Acknowledgements

Zachary Armstrong, Liang Wu and Gideon Davies from University of York, UK are kindly acknowledged for crystallographic studies and valuable discussion. Rian van den Nieuwendijk is acknowledged for the synthesis of D-galacturonic acid-type 1-*N*-iminosugars and valuable discussion.

5.5 Experimental methods

5.5.1 Biochemical experiments

Recombinant protein production and purification

HPSE – Human HPSE was expressed and purified according to previously reported procedures.[34]

AcGH79 – AcGH79 was expressed and purified according to previously reported procedures.[54]

BpHep – The coding sequence of BpHep was cloned into the pET28a vector (Novagen), behind an N-terminal 6xHis tag and thrombin cleavage site and used to transform *E. coli* strain BL21 Gold (DE3) (Agilent). Transformants were grown in TB media supplemented with 50 mg/mL kanamycin at 37 °C until cultures reached an OD600 of 0.8-1.0, whereupon expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma). Induced cultures were grown at 16 °C overnight, then harvested by 4,000 *g* centrifugation at 4 °C for 15 min.

Harvested cells were resuspended in ~50 mL Histrap buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM DTT), supplemented with DNAse I (Sigma; bovine pancrease), and cOmplete™ EDTA protease inhibitors (Roche). Cells were lysed using a cell disruptor (Constant systems) at 40 kPSI operating pressure, and lysate clarified by centrifugation at 40,000 *g* at 4 °C for 30 min. Clarified supernatant was loaded onto a 5 mL HisTrap FF crude column (Cytiva) pre-equilibrated with HisTrap buffer A, washed with 10 column volumes (CV) of HisTrap buffer A, before eluting with HisTrap buffer B (20 mM Tris pH 8.0, 500 mM NaCl, 1000 mM imidazole, 1 mM DTT) over a 20 CV linear gradient. BpHep containing fractions were pooled, and buffer exchanged into 20 mM HEPES pH 7.4, 100 mM NaCl by at least 3 rounds of sequential concentration/dilution using a 30 kDa molecular weight cut-off (MWCO) Vivaspin centrifugal concentrator (Cytiva). Buffer exchanged BpHep was digested overnight at ambient temperature with thrombin (Sigma; bovine plasma) at 1:100 mass ratio thrombin:BpHep.

Digested BpHep was rerun over a 5 mL HisTrap FF crude column pre-equilibrated with HisTrap buffer A, which was further washed with 3 CV of Histrap buffer A. Combined flowthrough and wash fractions were concentrated to ~2 mL volume using a 30 kDa MWCO Vivaspin centrifugal concentrator, then loaded onto a Superdex S75 16/600 pg size exclusion chromatography (SEC) column (Cytiva) pre-equilibrated in SEC buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 1 mM DTT). BpHep containing fractions were pooled and concentrated using a 30 kDa MWCO Vivaspin centrifugal concentrator to a

final concentration of ~20 mg/ml. Purified protein was flash frozen in liquid nitrogen and stored at –80 °C for use in further experiments.

EcGUSB – The coding sequence of EcGUSB was cloned into the pET28a vector, behind an N-terminal 6xHis tag, and used to transform *E. coli* strain BL21 Gold (DE3). Transformants were grown in TB media supplemented with 50 µg/mL kanamycin at 37 °C until cultures reached an OD600 of 0.8-1.0, whereupon gene expression was induced by the addition of 0.5 mM IPTG. Induced cultures were grown at 16 °C overnight, then harvested by 4,000 *g* centrifugation at 4 °C for 15 min.

Harvested cells were resuspended in ~50 mL Histrap buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM DTT), supplemented with DNAse I, and cOmplete™ EDTA protease inhibitors. Cells were lysed using a cell disruptor at 40 kPSI operating pressure, then lysate clarified by centrifugation at 40,000 *g* at 4 °C for 30 min. Clarified supernatant was loaded onto a 5 mL HisTrap FF crude column pre-equilibrated with HisTrap buffer A, washed with 10 CV of HisTrap buffer A, before eluting with HisTrap buffer B (20 mM Tris pH 8.0, 500 mM NaCl, 1000 mM imidazole, 1 mM DTT) over a 20 CV linear gradient. BpHep containing fractions were pooled, and concentrated using a 30 kDa MWCO Vivaspin centrifugal concentrator to a volume of ~ 2 mL. Concentrated protein was loaded onto a Superdex S200 16/600 pg SEC column pre-equilibrated in SEC buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 1 mM DTT). EcGUSB containing fractions were pooled and concentrated using a 30 kDa MWCO Vivaspin centrifugal concentrator to a final concentration of ~24.5 mg/ml. Purified protein was flash frozen in liquid nitrogen and stored at -80 °C for use in further experiments.

Crystallization

AcGH79 – Well diffracting crystals of AcGH79 were obtained by the sitting-drop vapor-diffusion method at 20 °C using a well solution containing 0.5 M ammonium sulfate, 1 M lithium sulfate, 0.1 M trisodium citrate, and a protein to well solution ratio of 500 nl: 500 nl. Crystals typically appeared after 1 week.

Inhibitor ligand complexes were obtained by transferring AcGH79 crystals to drops containing 2 M lithium sulfate and 1 mM inhibitor. Crystals were incubated with ligand for ~0.5 – 1 h, then directly harvested and flash-cooled in liquid nitrogen for data collection.

BpHep – Well diffracting crystals of BpHep were obtained by the sitting-drop vapor-diffusion method at 20 °C using a well solution containing 0.1 M sodium citrate pH 5.0, 14% (w/v) PEG 6000, and a protein to well solution ratio of 300 nl: 500nl. Crystals typically appeared after 3 days.

Inhibitor ligand complexes were obtained by transferring BpHep crystals to drops of well solution supplemented with 25 % (v/v) ethylene glycol and 1–5 mM inhibitor. Crystals were incubated with ligand for ~24 h, then directly harvested and flash-cooled in liquid nitrogen for data collection.

EcGusB – Initial crystals of EcGUSB were obtained by the sitting-drop vapor-diffusion method at 20 °C using a well solution containing 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350, 0.2 M $NaNO_3$. These initial crystals were used to prepare a microseed stock using Seed Beads (Hampton), then used to seed well diffracting crystals of EcGUSB in the same well conditions, at a protein to seed to well solution ratio of 500 nL: 200 nL: 1000 nL.

Inhibitor ligand complexes were obtained by transferring EcGUSB crystals to drops of well solution supplemented with 25 % (v/v) ethylene glycol and 1–5 mM inhibitor. Crystals were incubated with ligand for ~2 – 4 h, then directly harvested and flash-cooled in liquid nitrogen for data collection.

HPSE – Well diffracting crystals of HPSE were obtained by the sitting-drop vapor-diffusion method at 20 °C using a well solution containing 0.1 M MES pH 5.5, 0.1 M $MgCl_2$, 17 % (w/v) PEG 3350, and a protein to well solution ratio of 200 nL : 500 nL. Crystals typically appeared after 1 week.

Inhibitor ligand complexes were obtained by transferring HPSE crystals to drops of well solution supplemented with 25 % (v/v) ethylene glycol and 1 mM inhibitor. Crystals were incubated with ligand for ~0.5-1h, then directly harvested and flash-cooled in liquid nitrogen for data collection.

X-ray data collection and structure solution

X-ray diffraction data were collected at 100 K at beamlines i03, i04 and i04-1 of the Diamond Light Source UK. Reflections were autoprocessed using the XIA2 pipeline.[55] Complexes were solved by directly refining against their unliganded structures where solved using molecular replacement with PHASER4 (search model PDB accessions 5E98 (HPSE), 3VNY (AcGH79), 3K46 (EcGUSB)). Solved structures were iteratively improved by rounds of manual model building and maximum-likelihood refinement using COOT[56] and REFMAC5[57] respectively. Ligand coordinates were built using jLigand.[58] Diagrams were generated using PyMOL.

5.5.2 Chemical synthesis

General experimental details

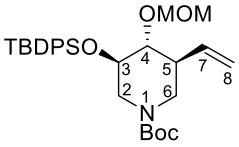
All reagents were of experimental grade and were used without further purification unless stated otherwise. Dichloromethane (DCM) and tetrahydrofuran (THF) were stored over 3 Å molecular sieves and *N,N*-dimethylformamide (DMF) was stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an Argon or N_2 atmosphere unless stated otherwise. Reactions were monitored by analytical thin layer chromatography (TLC) using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV-absorption (254 nm) and by spraying with a solution of $(NH_4)_6Mo_7O_{24}{\cdot}H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4{\cdot}H_2O$ (10 g/mL) 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 manually using Screening Device b.v. silica gel 60 (0.04-0.063 mm) in the indicated solvents. LC-MS analysis was performed on a LCQ

Advantage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C18 column (Gemini, 4.6 mm x 50 mm, 5 µM particle size, Phenomenex). The applied buffers were H_2O, acetonitrile (MeCN) and 1% aqueous trifluoroacetic acid (TFA). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-400 (400/101 MHz), Bruker AV-500 (500/126 MHz), and Bruker AV-850 (850/214 MHz) spectrometers in the given solvent. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (TMS) as internal standard (^{1}H NMR in $CDCl_3$) or the residual signal of the deuterated solvent. Coupling constants (*J*) are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. The following abbreviations are used to describe peak patterns when appropriate: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), Ar (aromatic), C_q (quarternary carbon). 2D NMR experiments (COSY, HSQC) were carried out to assign protons and carbons of the new structures and assignation follows the general numbering shown in compounds **24** . High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electronspray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL/min, capillary temperature 250 °C) with resolution R = 60000 at m/z 400 (mass range m/z = 150 – 2000) and dioctyl phthalate (m/z = 391.28428) as a "lock mass". The high-resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan).

Experimental Procedures and Characterization Data of Products

The spectroscopic data of known compounds **15-16**[41], **19-21**[43] and **S3**[50] are in agreement with those previously reported.

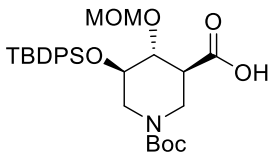
Compound 24



Compound **21** (218 mg, 0.454 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (3.0 mL) in a microwave tube. After cooling to 0 °C, DIPEA (632 µL, 3.63 mmol) and MOMCl (241 µL, 3.18 mmol) were added successively. The tube was then sealed and the mixture was heated to reflux at 40 °C overnight. TLC-analysis indicated full conversion of the starting material and the reaction was quenched with sat. aq. $NaHCO_3$ at 0 °C and diluted with DCM. The layers were separated and the aqueous layer was extracted with DCM (3 x). The combined organic layers were washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 50:1→20:1) affording compound **24** (208 mg, 0.396 mmol, 87%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$ at 293K) δ 7.71 (ddt, J = 13.7, 6.7, 1.6 Hz, 4H, CH Ar), 7.50 – 7.31 (m, 6H, CH Ar), 5.88 (s, 1H, H7), 5.25 – 5.05 (m, 2H, H8), 4.88 (s, 1H, CH*H* MOM), 4.58 – 4.40 (m, 1H, CH*H* MOM), 3.75 (brs, 2H, H3 and H6a), 3.63 (d, J = 13.2 Hz, 1H, H2a), 3.44 (dd, J = 8.3, 6.9 Hz, 1H, H4), 3.28 (s, 3H, CH_3 MOM), 2.85 (brs, 2H, H2b and H6b), 2.26 (qd, J = 8.6, 4.4 Hz, 1H, H5), 1.28 (s, 9H, $(CH_3)_3$CSi), 1.06 (s, 9H, $(CH_3)_3$CO). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) δ 154.6 (C=O), 137.5 (C7), 135.6 (4CH Ar), 134.1 (C_q Ar), 133.3 (C_q Ar), 129.9 (2CH Ar), 127.7 (4CH Ar), 117.1 (C8), 98.0 (CH_2

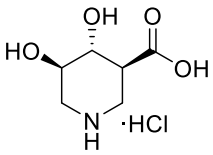
MOM), 82.5 (C4), 79.8 (C_q Boc), 73.0 (C3), 56.0 (CH_3 MOM), 48.4 (C2), 46.1 (C6), 45.2 (C5), 28.4 ((CH_3)$_3$CSi), 27.1 ((CH_3)$_3$CO), 19.3 ((CH_3)$_3$CSi) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{30}H_{43}NO_5SiNa$ 548.2803, found 548.2800.

Compound 25



To a solution of compound **24** (26 mg, 50 µmol) in a mixture of CCl_4 (0.3 mL) and MeCN (0.3mL) was added a solution of $NaIO_4$ (53 mg, 0.25 mmol) and $RuCl_3 \cdot 3H_2O$ (0.1 M in H_2O, 50 µL, 5.0 µmol) in water (0.45 mL) at 0 °C. The reaction mixture was then stirred vigorously at rt for 3 h. After which, sat. aq. $Na_2S_2O_3$ was added to quench the reaction and the resulting mixture was stirred for another 15 min at rt. The layers were separated and the aqueous layer was extracted with EtOAc (3 x). The combined organic layers were washed with H_2O, brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 15:1→5:1, with 0.1% HOAc) affording compound **25** (17.5 mg, 32.0 µmol, 64%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 7.74 – 7.66 (m, 4H, CH Ar), 7.48 – 7.34 (m, 6H, CH Ar), 4.67 (s, 1H, CH*H* MOM), 4.49 (s, 1H, CH*H* MOM), 3.99 (t, J = 6.3 Hz, 1H, H4), 3.84 – 3.64 (brs, 1H, H6a), 3.73 (q, J = 5.7 Hz, 1H, H3), 3.60 – 3.29 (brs, 2H, H6b and H2a), 3.22 (s, 3H, CH_3 MOM), 2.59 (brs, 1H, H2b), 1.33 (s, 9H, (CH_3)$_3$CSi), 1.05 (s, 9H, (CH_3)$_3$CO). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 136.0 (4CH Ar), 130.0 (2CH Ar), 127.8 (4CH Ar), 97.4 (CH_2 MOM, assigned by HSQC), 78.8 (C4, assigned by HSQC), 70.9 (C3, assigned by HSQC), 56.0 (CH_3 MOM), 47.4 (C2, assigned by HSQC), 46.0 (C5, assigned by HSQC), 41.9 (C6), 28.4 ((CH_3)$_3$CSi), 27.1 ((CH_3)$_3$CO), 19.3 ((CH_3)$_3$CSi) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{29}H_{41}NO_7SiNa$ 566.2545, found 566.2544.

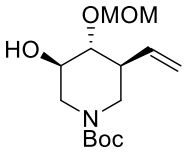
Compound 14



A solution of compound **25** (16 mg, 29 µmol) in dioxane (0.25 mL) in a microwave tube was added aqueous 3 M HCl (1.5 mL). The tube was then sealed and the mixture was stirred at 100 °C for 4 h until LC-MS indicated full deprotection. After cooling to rt, Et_2O (5 mL) was added and the mixture was stirred vigorously at rt for 5 min. The phases were then separated and the organic layer was taken out carefully. The remaining H_2O layer was washed two more times with Et_2O (2 x 5 mL) and concentrated. The residue was purified by silica gel column chromatography (isopropanol/H_2O/28-30% NH_4OH 12:1:1→10:2:1) to give chromatographically pure **14**. Milli-Q water (1 mL) and aq. 1 M HCl (2 mL) were added to the residue, and the solution was evaporated to form a hydrochloride salt of **14**, which was further purified by a column of Sephadex G-25 with Milli-Q as eluent, affording the title compound **14** (5.5 mg, 28 µmol, 95%) as a white solid after lyophilization. ^{1}H NMR (500 MHz, D_2O at 293K) δ 4.04 – 3.96 (m, 1H, H4), 3.90 – 3.82 (m, 1H, H3), 3.49 (dp, J = 12.9, 3.6, 2.7 Hz, 2H, H6a and H2a), 3.30 (tt, J = 10.8, 2.6 Hz, 1H, H6b), 3.08 – 2.99 (m, 1H, H2b), 2.81 (td, J = 7.9, 3.6 Hz, 1H, H5). ^{13}C NMR (126 MHz, D_2O at

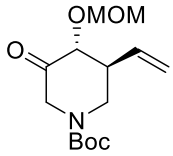
293K) δ 174.3 (C=O), 70.2 (C4), 66.9 (C3), 45.7 (C2), 45.0 (C5), 42.5 (C6). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_6H_{12}NO_4$ 162.0761, found 162.0763.

Compound 26



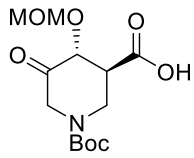
Compound **24** (187 mg, 0.35 mmol) was dissolved in dry THF (3.5 mL). TBAF (1 M in THF, 2.1 mL, 2.1 mmol) was added and the mixture was stirred at rt for 1.5 h. The reaction was quenched with sat. aq. NH_4Cl and concentrated *in vacuo* to remove THF. The resulting residue was diluted with DCM, washed with H_2O (2 x), brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 9:1→3:1) affording compound **26** (96 mg, 0.33 mmol, 94%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 5.74 – 5.63 (m, 1H, H7), 5.19 (dd, *J* = 12.1, 1.3 Hz, 1H, H8a), 5.19 – 5.14 (m, 1H, H8b), 4.77 (dd, *J* = 6.9, 0.9 Hz, 1H, CH*H* MOM), 4.65 (dd, *J* = 7.0, 1.1 Hz, 1H, CH*H* MOM), 4.53 – 4.17 (m, 2H, H6a and H2a), 3.50 – 3.40 (H3), 3.46 (s, 3H, CH_3 MOM), 3.08 (dd, *J* = 10.2, 8.3 Hz, 1H, H4), 2.64 – 2.47 (m, 2H, H6b and H2b), 2.35 – 2.23 (m, 1H, H5), 1.46 (d, *J* = 1.3 Hz, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 154.6 (C=O), 135.9 (C7), 117.9 (C8), 98.5 (CH_2 MOM), 89.2 (C4), 80.2 ($(CH_3)_3\mathit{C}$), 70.0 (C3), 56.0 (CH_3 MOM), 48.5 (C2, assigned by HSQC), 46.3 (C6, assigned by HSQC), 44.9 (C5, assigned by HSQC), 28.5 ($(\mathit{C}H_3)_3C$). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{14}H_{25}NO_5Na$ 310.1625, found 310.1623.

Compound 27



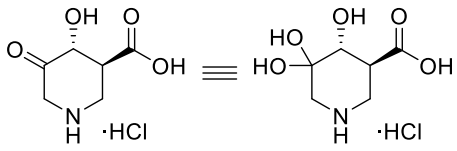
To a solution of compound **26** (96 mg, 0.33 mmol) in dry DCM (5 mL) at 0 °C was added Dess-Martin periodinane (283 mg, 0.66 mmol). The mixture was stirred at rt for 1 h until TLC-analysis indicated total consumption of the starting material. A mixture of sat. aq. $Na_2S_2O_3$ (3 mL) and sat. aq. $NaHCO_3$ (3 mL) was added to quench the reaction and the mixture was stirred at rt for another 15 min until the white emulsion became clear solution. The layers were separated and the aqueous layer was extracted with DCM (2 x). The combined organic layers were washed with H_2O, brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 9:1→3:1) affording compound **27** (88 mg, 0.31 mmol, 93%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 5.77 (ddd, *J* = 17.6, 10.4, 7.6 Hz, 1H, H7), 5.30 – 5.20 (m, 2H, H8), 4.79 (dd, *J* = 7.1, 1.0 Hz, 1H, CH*H* MOM), 4.67 (dd, *J* = 7.0, 1.0 Hz, 1H, CH*H* MOM), 4.28 (dd, *J* = 16.7, 1.5 Hz, 1H, H2a), 4.10 (d, *J* = 10.2 Hz, 1H, H4), 4.07 – 3.99 (brs, 1H, H6a), 3.84 (brs, 1H, H2b), 3.39 (d, *J* = 1.0 Hz, 3H, CH_3 MOM), 3.37 – 3.17 (m, 1H, H6b), 2.71 (qd, *J* = 9.7, 5.0 Hz, 1H, H5), 1.46 (d, *J* = 1.0 Hz, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 202.5 (C3), 154.3 (C=O Boc), 135.3 (C7), 118.5 (C8), 96.5 (CH_2 MOM), 81.1 ($(CH_3)_3\mathit{C}$), 80.0 (C4), 56.3 (CH_3 MOM), 54.2 (C2, assigned by HSQC), 46.6 (C5), 46.0 (C6, assigned by HSQC), 28.4 ($(\mathit{C}H_3)_3C$) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{14}H_{23}NO_5Na$ 308.1468, found 308.1465.

Compound 28



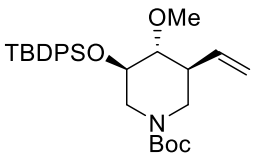
To a solution of compound **27** (35 mg, 0.12 mmol) in a mixture of CCl_4 (1.4 mL) and MeCN (1.4 mL) was added a solution of $NaIO_4$ (130 mg, 0.60 mmol, 5.0 eq.) and $RuCl_3 \cdot 3H_2O$ (2.5 mg, 0.012 mmol, 0.1 eq.) in water (2.1 mL) at 0 °C. The reaction mixture was then stirred vigorously at rt for 2 h until TLC and LC-MS analysis indicated complete conversion. The phases were separated and the aqueous phase was extracted with EtOAc (3 x 15 mL). To the combined organic extracts was added isopropanol (0.2 mL) and the mixture was stirred at rt for an additional 1 hour. The mixture was then washed with H_2O (15 mL), brine (15 mL), dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (DCM/MeOH 100:0→100:4) affording compound **28** (31 mg, 0.10 mmol, 83%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.79 (d, J = 6.7 Hz, 1H, CH*H* MOM), 4.71 (d, J = 6.7 Hz, 1H, CH*H* MOM), 4.49 (d, J = 8.5 Hz, 1H, H4), 4.09 (d, J = 16.8 Hz, 1H, H2a), 3.99 (d, J = 16.7 Hz, 2H, H2b and H6a), 3.79 (dd, J = 13.7, 7.4 Hz, 1H, H6b), 3.39 (s, 3H, CH_3 MOM), 3.05 – 2.97 (m, 1H, H5), 1.46 (s, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 174.0 (COOH), 154.4 (C=O Boc), 96.9 (CH_2 MOM), 81.8 ($(CH_3)_3C$), 77.2 (C4), 56.4 (CH_3 MOM), 53.4 (C2, assigned by HSQC), 48.1 (C5), 43.4 (C6), 28.4 ($(CH_3)_3C$) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{21}NO_7Na$ 326.1210, found 326.1210; $[M+H_2O+Na]^+$ calc for $C_{13}H_{23}NO_8Na$ 344.1316, found 344.1316.

Compound 12



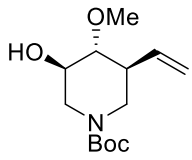
Compound **28** (24 mg, 79 µmol) was dissolved in 4 M dioxane/H_2O (1.5 mL) and stirred at rt for 5 h. After which, water (1.0 mL) and Et_2O (~8.0 mL) were added. The mixture was stirred vigorously for a while and stood for stratification. The Et_2O layer was taken out carefully and the remaining water layer was washed two more times with Et_2O (2 x 8.0 mL) and concentrated *in vacuo*. The resulting dry residue was re-dissolved in MeOH (1.0 mL) and Et_2O (~8.0 mL) was added slowly under vigorous stirring. Upon addition of Et_2O, a lot of light yellow solid appeared. After stirring vigorously for a while, the mixture was stood for solid precipitation and the Et_2O supernatant was taken out carefully. The solid was washed repeatedly with Et_2O (3 x 6.0 mL), re-dissolved in Milli-Q water (~1.5 mL) and filtered over a Whatman filter paper. After lyophilization, the target product (10.8 mg, 55 µmol, 70%) was obtained as a light yellow solid (HCl salt). ^{1}H NMR after lyophilization (500 MHz, D_2O at 293K) [hydrate form] δ 4.07 (d, J = 7.5 Hz, 1H, H4), 3.48 (ddd, J = 13.2, 4.5, 1.1 Hz, 1H, H6a equatorial), 3.38 (dd, J = 13.2, 8.0 Hz, 1H, H6b axial), 3.35 (dd, J = 12.9, 1.0 Hz, 1H, H2a), 3.14 (d, J = 12.9 Hz, 1H, H2b), 3.04 (td, J = 8.0, 7.5, 4.5 Hz, 1H, H5). ^{13}C NMR after lyophilization (126 MHz, D_2O at 293K) [hydrate form] δ 173.6 (COOH), 90.7 (C3 hydrate), 71.1 (C4), 49.0 (C2), 44.4 (C5), 41.9 (C6) ppm. HRMS (ESI) m/z: $[M_{ketone}+H]^+$ calc for $C_6H_{10}NO_4$ 160.06043, found 160.06059; $[M_{hydrate}+H]^+$ calc for $C_6H_{12}NO_5$ 178.07100, found 178.07119.

Compound 29



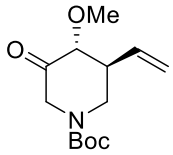
Compound **21** (0.15 g, 0.31 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (3 mL) under nitrogen in a microwave tube and stirred over fresh flame-dried 3 Å molecular sieve powder. After cooling to 0 °C, proton sponge (669 mg, 3.1 mmol) and trimethyloxonium tetrafluoroborate (0.37 mg, 2.5 mmol) were added successively. The tube was then sealed and protected from light. After stirring at rt for 22 h, the reaction slurry was diluted with EtOAc and filtered over a small pad of celite. The filter pad was washed with EtOAc (2 x) and the combined filtrates were washed with sat. aq. NH_4Cl (2 x), sat. aq. $NaHCO_3$, H_2O and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 100:1→30:1) affording compound **29** (116 mg, 0.23 mmol, 75%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.78 – 7.65 (m, 4H, CH Ar), 7.45 – 7.28(m, 6H, CH Ar), 5.92 – 5.81 (m, 1H, H7), 5.20 – 5.05 (m, 2H, H8), 3.85 – 3.61 (m, 3H, H3, H6a and H2a), 3.30 (s, 3H, CH_3O), 3.01 (ddt, J = 8.2, 6.6, 1.8 Hz, 1H, H4), 2.96 – 2.65 (m, 2H, H6b and H2b), 2.31 – 2.10 (m, 1H, H5), 1.32 (s, 9H, $(CH_3)_3CSi$), 1.09 (q, J = 2.5, 1.4 Hz, 9H, $(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 154.7 (C=O), 137.4 (C7), 136.2 (2CH Ar), 136.1 (2CH Ar), 134.5 (C_q Ar), 134.0 (C_q Ar), 129.9 (CH Ar), 129.8 (CH Ar), 127.8 (2CH Ar), 127.7 (2CH Ar), 116.6 (C8), 86.6 (C4), 79.8 ($(CH_3)_3\mathit{C}O$), 71.8 (C3), 59.5 (CH_3O), 48.6 (C2, assign by HSQC), 46.2 (C6), 45.2 (C5), 28.5 ($(\mathit{C}H_3)_3CSi$), 27.3 ($(\mathit{C}H_3)_3CO$), 19.5 ($(CH_3)_3\mathit{C}Si$) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{29}H_{41}NO_4SiNa$ 518.2697, found 518.2693.

Compound S1



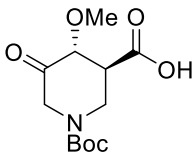
Compound **29** (153 mg, 0.31 mmol) was dissolved in dry THF (3.1 mL). TBAF (1 M in THF, 1.85 mL, 1.85 mmol) was added and the mixture was stirred at rt for 2 h. The reaction was quenched with sat. aq. NH_4Cl and concentrated *in vacuo* to remove THF. The resulting residue was diluted with DCM, washed with H_2O, brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 11:1→3:1) affording compound **S1** (72 mg, 0.28 mmol, 90%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 5.76 (ddd, J = 17.8, 10.4, 7.9 Hz, 1H, H7), 5.28 – 5.12 (m, 2H, H8), 4.20 (brs, 1H, H2a), 4.11 – 3.77 (m, 1H, H6a), 3.54 (brs, 1H, H3), 3.51 (s, 3H, CH_3O), 2.95 (dd, J = 9.9, 8.4 Hz, 1H, H4), 2.81 (brs, 1H, OH), 2.74 – 2.50 (m, 2H, H2b and H6b), 2.28 (brs, 1H, H5), 1.45 (s, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 154.7 (C=O), 136.3 (C7), 117.7 (C8), 87.3 (C4), 80.3 (C_q Boc), 70.3 (C3), 59.8 (CH_3O), 48.1 (C2, assign by HSQC), 47.1 (C6, assign by HSQC), 45.5 (C5), 28.5 ($(\mathit{C}H_3)_3C$) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{23}NO_4Na$ 280.1519, found 280.1517.

Compound 30



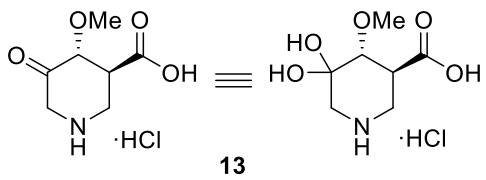
To a solution of compound **S1** (79 mg, 0.30 mmol) in dry DCM (4.5 mL) at 0 °C was added Dess-Martin periodinane (260 mg, 0.60 mmol). The mixture was stirred at rt for 3.5 h until TLC-analysis indicated total consumption of the starting material. A mixture of sat. aq. $Na_2S_2O_3$ (3 mL) and sat. aq. $NaHCO_3$ (3 mL) was added to quench the reaction and the mixture was stirred at rt for another 15 min until the white emulsion became clear solution. The layers were separated and the aqueous layer was extracted with DCM (2 x). The combined organic layers were washed with H_2O, brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 9:1→4:1) affording compound **30** (65 mg, 0.25 mmol, 83%) as a colourless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 5.77 (ddd, J = 17.4, 10.7, 6.8 Hz, 1H, H7), 5.24 – 5.16 (m, 2H, H8), 4.13 (dd, J = 16.8, 1.2 Hz, 1H, H2a), 3.90 (dd, J = 17.6, 13.2 Hz, 2H, H6a and H2b), 3.57 (d, J = 8.6 Hz, 1H, H4), 3.45 (s, 3H, CH_3O), 3.44 – 3.37 (m, 1H, H6b), 2.72 (dddt, J = 12.8, 8.0, 4.7, 1.3 Hz, 1H, H5), 1.46 (s, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 202.8 (C3), 154.4 (C=O Boc), 135.3 (C7), 117.8 (C8), 85.1 (C4), 80.9 (C_q Boc), 58.9 (CH_3O), 53.6 (C2), 45.6 (C5), 45.4 (C6), 28.4 ((CH_3)$_3$C). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{13}H_{21}NO_4Na$ 278.1363, found 278.1361.

Compound 31

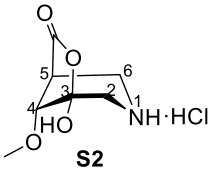


To a stirred solution of compound **30** (20 mg, 0.078 mmol) in CCl_4 (0.45 mL) and MeCN (0.45 mL) was added a solution of $NaIO_4$ (60 mg, 0.28 mmol, 3.6 eq.) and $RuCl_3 \cdot 3H_2O$ (0.1 M in H_2O, 40 µL, 4.0 µmol, 0.05 eq.) in water (0.67 mL) at 0 °C. The reaction mixture was stirred vigorously at 0 °C for 1 h. TLC-analysis indicated the presence of the aldehyde intermediate, so the mixture was warmed to rt and stirred for another 30 min. The phases were separated and the aqueous phase was extracted with EtOAc (3 x 10 mL). To the combined organic extracts was added isopropanol (0.2 mL) and the mixture was stirred at rt for an additional 1 h. The mixture was then washed with brine (10 mL), dried over Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (DCM/MeOH 80:1→20:1) affording compound **31** (16 mg, 59 µmol, 75%) as a pale yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.08 – 3.97 (m, 3H, H4 and H2ab), 3.97 – 3.85 (m, 1H, H6a), 3.84 – 3.76 (m, 1H, H6b), 3.49 (s, 3H, CH_3O), 3.03 – 2.93 (m, 1H, H5), 1.45 (s, 9H, $(CH_3)_3C$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 154.5 (C=O Boc), 81.7 (C_q Boc), 81.5 (C4), 59.3 (CH_3O), 53.1 (C2, assigned by HSQC), 47.8 (C5), 43.0 (C6), 28.4 ((CH_3)$_3$C) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{12}H_{19}NO_6Na$ 296.1105, found 296.1102.

Major product 13 and minor product S2

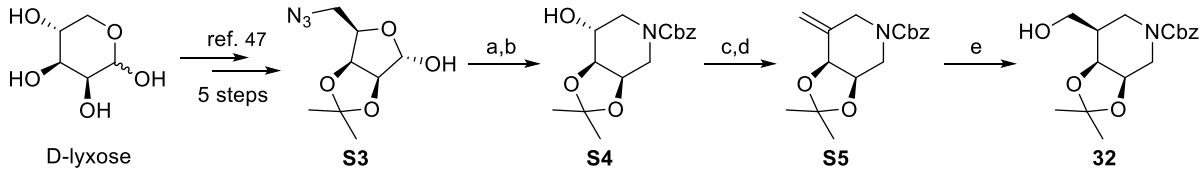


To a stirred white suspension of compound **31** (17 mg, 62 µmol) in the mixture of HFIP (600 µL) and H_2O (300 µL) was added 0.2 M HCl/HFIP (340 µL, 68 µmol, 1.1 eq.) and the resulting clear solution was stirred at rt for 1 h. TLC-analysis indicated the presence of the starting material, so more 0.2 M HCl/HFIP (340 µL, 68 µmol) was added and the mixture was stirred for another 1 h. After which, 37% HCl (12 M, 5.7 µL, 68 µmol) was added and the mixture was stirred for additional 1.5 h until TLC-analysis indicated full deprotection. The solvent was concentrated under reduced pressure and the resulting residue was re-dissolved in Milli-Q water (~1.5 mL), washed with Et_2O (3 x 8.0 mL) and filtered over a Whatman filter paper. After lyophilization, the product was obtained as a mixture of two compounds (10.9 mg, 52 µmol, 84%, ratio ≈ 4:1) as a hygroscopic light yellow solid. Major product **13**: ^{1}H NMR (500 MHz, D_2O at 328K) [hydrate form] δ 4.13 (d, J = 6.0 Hz, 1H, H4), 3.88 (s, 3H, CH_3O), 3.81 (dd, J = 13.3, 6.1 Hz, 1H, H6a), 3.70 (dd, J = 13.3, 4.3 Hz, 1H, H6b), 3.62 (d, J = 12.9 Hz, 1H, H2a), 3.58 – 3.52 (m, 1H, H5), 3.49 (d, J = 12.9 Hz, 1H, H2b). ^{13}C NMR (126 MHz, D_2O at 328K) [hydrate form] δ 173.9 (COOH), 91.2 (C3 hydrate), 80.6 (C4), 59.6 (CH_3O), 49.1 (C2), 41.8 (C5), 41.4 (C6) ppm. HRMS (ESI) m/z: $[M_{ketone}+H]^+$ calc for $C_7H_{12}NO_4$ 174.07608, found 174.07607; $[M_{hydrate}+H]^+$ calc for $C_7H_{14}NO_5$ 192.08665, found 192.08655.



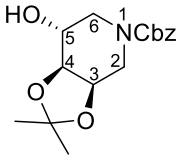
Proposed structure of minor product **S2** which may be formed by dehydration of **13**. The protons of **S2** could not be assigned confidently due to the peaks partially overlap with those of major product **13**. The carbon peaks were observed at δ 91.9 (C3), 79.9 (C4), 61.4 (CH_3O), 47.2 (C2), 42.0 (C5), 39.4 (C6) ppm. Of note the carbonyl carbon was not found on the NMR spectrum.

Synthesis of key intermediate 32



Scheme 5.6. Synthesis of key intermediate **32** following reported procedures for the preparation of its *N*-Boc protected analogue. Reagents and conditions: a) H_2, 10% Pd/C, MeOH, rt; b) CbzCl, sat. aq. $NaHCO_3$, THF, 0 °C to rt, 74% over two steps; c) Dess-Martin periodinane, DCM, 0 °C to rt; d) Ph_3PCH_3Br, ^{t}BuOK, THF, -20 °C to rt, 84% over two steps; (e) *i*) 9-BBN, THF, 0 °C to rt; *ii*) 30% H_2O_2, 1 M NaOH in H_2O, rt, 73%.

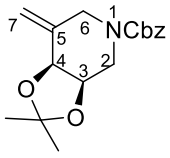
Compound S4



Azide **S3** (10.7 g, 49.8 mmol) was dissolved in MeOH (1000 mL) and nitrogen was bubbled through the solution before 20% $Pd(OH)_2/C$ (2.0 g) was added. While stirring vigorously, the mixture was flushed with two H_2 balloons. After stirring for 20 h under H_2 atmosphere, NMR analysis of a small amount of the reaction mixture revealed almost no conversion. Therefor 10% Pd/C (2.06 g) was added and the reaction was stirred for another 20 h until NMR analysis indicated full conversion of the starting material. The reaction mixture was filtered and concentrated to give the iminosugar intermediate (8.2 g, 47 mmol) as a white solid. The crude amine was directly dissolved in a mixture of THF (350 mL) and sat. aq. $NaHCO_3$ (250 mL). After cooling to 0 °C, CbzCl (95%, 10.5 mL, 69.8 mmol) was added dropwise and the reaction was stirred at rt for 40 h until LC-MS analysis indicated full conversion. The layers were separated and the water layer was extracted with EtOAc (2 x 200 mL). The combined organic layers were washed with brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The product was purified by silica gel column chromatography (pentane/EtOAc 9:1→1:1) affording compound **S4** (11.4 g, 37.1 mmol, 74% over two steps) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.43 – 7.14 (m, 5H, CH Ar), 5.12 (s, 2H, CH_2 Cbz), 4.33 – 4.27 (m, 1H, H5), 4.07 (dd, J = 6.5, 4.1 Hz, 1H, H4), 3.97 – 3.83 (m, 2H, H3, H6b), 3.63 (dd, J = 13.5, 3.2 Hz, 1H, H2b), 3.51 (dd, J = 14.2, 3.3 Hz, 1H, H6a), 3.36 (dd, J = 13.5, 5.2 Hz, 1H, H2a), 3.11 (brs, 1H, OH), 1.41 (s, 3H, CH_3), 1.31 (s, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 156.5 (C=O), 136.9 (C_q Ar), 128.5, 128.0, 127.8 (CH Ar), 109.2 (OCO), 75.6 (C4), 72.0 (C5), 67.5 (C3), 67.3 (CH_2 Cbz), 44.9 (C2), 42.8 (C6), 27.3 (CH_3), 25.0 (CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 2:1] δ 7.37 – 7.23 (m, 5H, CH Ar), 5.15 – 5.06 (m, 2H, CH_2 Cbz), 4.38 – 4.27 (m, 1H, H5), 4.12 – 4.04 (m, 1H, H4), 4.03 – 3.85 (m, 2H, H3, H6b), 3.83 – 3.56 (m, 2H, OH and H2b), 3.55 – 3.43 (m, 1H, H6a), 3.43 – 3.31 (m, 1H, H2a), 1.43 (s, 2H, CH_3), 1.41 (s, 1H, CH_3), 1.32 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 2:1] δ 156.5 (C=O), 136.5 (C_q Ar), 128.5, 128.0, 127.807, 127.7 (Ar), 109.1 (OCO), 75.1 (C4), 71.9 & 71.8 (C5), 67.2 (CH_2 Cbz), 67.2 (C3), 44.6 & 44.4 (C2), 42.5 & 42.2 (C6), 27.2 & 27.1 (CH_3), 24.9 & 24.8 (CH_3) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{16}H_{22}NO_5$ 308.14925, found 308.14907.

Compound S5

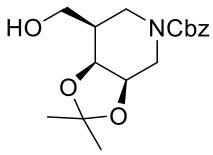


Alcohol **S4** (11.3 g, 36.8 mmol) was dissolved in dry DCM (400 mL) under argon. After cooling to 0 °C, Dess-Martin periodinane (27.3 g, 64.4 mmol) was added in several portions. After stirring for 30 minutes at 0 °C and for 2 h at room temperature, the reaction was diluted with EtOAc (600 mL) and washed successively with 2 M $Na_2S_2O_3$ solution (2 x 250 mL), sat. aq. $NaHCO_3$ (2 x 200 mL) and brine (150 mL). The organic layer was dried over $MgSO_4$, filtrated and concentrated *in vacuo* to afford the crude ketone. In the meantime methyltriphenylphosphonium bromide (55.3 g, 155 mmol) was suspended in dry THF (250 mL) and

cooled to 0 ºC, ^{t}BuOK (15.4 g, 138 mmol) was added. The resulting yellow suspension was stirred for 1 h at 0 ºC and for 2 h at rt. After which, the suspension was cooled to -20 ºC and a solution of the crude ketone in THF (150 mL) was added dropwise over 20 minutes. The reaction was then stirred overnight at room temperature and quenched by addition of sat. aq. NH_4Cl (250 mL). The layers were separated and the water layer extracted with EtOAc (2 x 100 mL). The combined organic layers were washed with brine (100 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 95:5→3:1) affording compound **S5** (9.4 g, 31 mmol, 84% over two steps) as a slightly orange oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.53 – 7.08 (m, 5H, CH Ar), 5.28 – 5.14 (m, 4H, H7ab and CH_2 Cbz), 4.61 (d, J = 7.4 Hz, 1H, H4), 4.32 (m, 2H, H3 and H6b), 3.97 – 3.76 (m, 2H, H6a and H2b), 3.11 (dd, J = 14.3, 2.8 Hz, 1H, H2a), 1.38 (s, 3H, CH_3), 1.33 (s, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 155.9 (C=O), 139.7 (C5), 137.0 (C_q Ar), 128.4, 127.9, 127.8 (CH Ar), 116.4 (C7), 109.7 (OCO), 76.3 (C4), 74.6 (C3), 67.1 (CH_2 Cbz), 46.6 (C6), 44.2 (C2), 26.6 (CH_3), 24.8 (CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.40 – 7.24 (m, 5H, CH Ar), 5.29 – 5.07 (m, 4H, H7ab and CH_2 Cbz), 4.65 (dd, J = 7.4, 3.8 Hz, 1H, H4), 4.44 – 4.30 (m, 2H, H3 and H6b), 3.97 (dd, J = 14.4, 2.7 Hz, 1H, H2b), 3.87 (dd, J = 14.1, 8.5 Hz, 1H, H6a), 3.07 (ddd, J = 14.1, 7.7, 2.3 Hz, 1H, H2a), 1.40 (s, 1.5H, CH_3), 1.37 (s, 1.5H, CH_3), 1.35 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 156.0 & 155.8 (C=O), 139.3 (C5), 136.8 (C_q Ar), 128.4, 128.4, 128.0, 127.9, 127.9, 127.8 (CH Ar), 117.1 & 116.8 (C7), 109.6 (OCO), 76.2 & 76.1 (C4), 74.5 (C3), 67.1 & 67.0 (CH_2 Cbz), 46.5 & 46.3 (C6), 44.1 & 43.8 (C2), 26.5 (CH_3), 24.6 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{17}H_{22}NO_4$ 304.15433, found 304.15419.

Compound 32

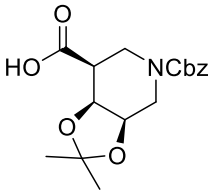


Alkene **S5** (4.70 g, 15.5 mmol) was dissolved in dry THF (150 mL) under argon. After cooling to 0 ºC, 9-BBN (0.5 M in THF, 140 mL, 70 mmol) was added dropwise over 30 minutes and the mixture was stirred for 20 h at rt until TLC-analysis indicated full conversion of the starting material. Then the mixture was cooled to 0 ºC again, and water (10 mL), 1 M NaOH in H_2O (10 mL) and 30% H_2O_2 (10 mL) were added successively. After stirring overnight at rt, the mixture was diluted with EtOAc (250 mL) and washed with brine (2 x 200 mL). The combined water layers were extracted with EtOAc (100 mL) and the combined organic layers were dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 1:1→3:7) to afford compound **32** (3.66 g, 11.4 mmol, 73%) as a thick oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.37 – 7.22 (m, 5H, Ar), 5.13 (s, 2H, CH_2 Cbz), 4.42 (dd, J = 7.1, 2.7 Hz, 1H, H4), 4.29 (s, 1H, H3), 3.80 (d, J = 10.9 Hz, 1H, H2b), 3.68 (m, 2H, H7ab), 3.56 (dd, J = 12.3, 5.0 Hz, 1H, H6b), 3.29 (d, J = 13.9 Hz, 1H, H2a), 3.22 (app t, J = 12.3 Hz, 1H, H6a), 2.50 (brs, 1H, OH), 1.99 (m, 1H, H5), 1.39 (s, 3H, CH_3), 1.31 (s, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 156.1 (C=O), 137.0 (C_q Ar), 128.5, 127.9, 127.8 (CH Ar), 108.9

(OCO), 72.6 (C3), 72.4 (C4), 67.1 (CH_2 Cbz), 62.5 (C7), 43.2 (C2), 40.3 (C6), 38.3 (C5), 26.7 (CH_3), 24.7 (CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.38 – 7.24 (m, 5H, Ar), 5.13 (s, 2H, CH_2 Cbz), 4.44 (dd, J = 7.3, 2.2 Hz, 1H, H4), 4.39 – 4.26 (m, 1H, H3), 3.93 – 3.75 (m, 1H, H2b), 3.69 (s, 2H, H7ab), 3.55 (m, 1H, H6b), 3.24 (m, 2H, H6a and H2a), 2.99 (s, 0.5H, OH), 2.89 (s, 0.5H, OH), 2.00 (s, 1H, H5), 1.40 (s, 3H, CH_3), 1.32 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 156.3 & 155.9 (C=O), 136.7 (C_q Ar), 128.4, 127.9, 127.8, 127.7 (CH Ar), 108.7 (OCO), 72.4 (C4), 72.1 & 72.0 (C3), 67.0 (CH_2 Cbz), 62.3 & 62.2 (C7), 43.0 & 42.6 (C2), 40.2 & 39.9 (C6), 38.0 (C5), 26.6 (CH_3), 24.6 (CH_3) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{17}H_{23}NO_5Na$ 344.1468, found 344.1477.

Compound 33

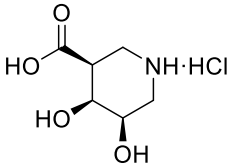


Alcohol **32** (347 mg, 1.08 mmol) was dissolved in dry DCM (14 mL) under argon and Dess-Martin periodinane (720 mg, 1.70 mmol) was added. After stirring at rt for 3 h, TLC-analysis confirmed complete conversion of the starting material. The mixture was diluted with EtOAc (60 mL) and washed successively with aqueous 2 M $Na_2S_2O_3$ (2 x 15 mL), sat. aq. $NaHCO_3$ (2 x 15 mL) and brine (20 mL), dried over $MgSO_4$, filtrated and concentrated to afford the crude aldehyde as a colorless oil (350 mg). The aldehyde was directly dissolved in a mixture of MeCN (18 mL) and water (3.5 mL) and $NaH_2PO_4 \cdot 2H_2O$ (1.31 g, 8.40 mmol) was added. The mixture was cooled to 0 °C and 30% H_2O_2 (142 µL, 1.39 mmol) and a solution of $NaClO_2$ (80% purity, 157 mg, 1.14 mmol) in water (8.5 mL) were added subsequently. The slightly yellow solution was stirred at 0 °C for 15 minutes and at rt for 1 h until TLC-analysis confirmed full conversion of the aldehyde. The reaction was quenched by addition of Na_2SO_3 (0.5 g) and stirred for another 15 minutes. After which, the mixture was diluted with EtOAc (60 mL) and the layers were separated. The aqueous layer was extracted with EtOAc (2 x 10 mL) and the combined organic layers were washed with brine (20 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. Purification by silica gel column chromatography (pentane/EtOAc 1:1→0:1) afforded the target compound **33** (236 mg, 0.703 mmol, 65%) as a glassy solid. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 8.14 (s, 1H, OH), 7.37 – 7.23 (m, 5H, CH Ar), 5.15 (s, 2H, CH_2 Cbz), 4.72 (dd, J = 6.9, 2.9 Hz, 1H, H4), 4.35 (s, 1H, H3), 3.94 – 3.72 (m, 2H, H6b and H2b), 3.50 (app t, J = 12.7 Hz, 1H, H6a), 3.34 (d, J = 13.6 Hz, 1H, H2a), 2.80 (ddd, J = 12.6, 4.8, 3.2 Hz, 1H, H5), 1.41 (s, 3H, CH_3), 1.33 (s, 3H, CH_3). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 173.8 (C=O acid), 156.1 (C=O Cbz), 136.8 (C_q Ar), 128.6, 128.1, 128.0 (CH Ar), 109.5 (OCO), 72.3 (C3), 71.9 (C4), 67.5 (CH_2 Cbz), 43.3 (C2), 42.0 (C5), 38.6 (C6), 26.7 (CH_3), 24.8 (CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 9.41 (brs, 1H, OH), 7.82 – 7.12 (m, 5H, CH Ar), 5.15 (s, 2H, CH_2 Cbz), 4.74 (dd, J = 7.0, 2.7 Hz, 1H, H4), 4.56 – 4.25 (m, 1H, H3), 4.09 –

3.67 (m, 2H, H6b and H2b), 3.49 (m, 1H, H6a), 3.32 (m, 1H, H2a), 2.87 – 2.75 (m, 1H, H5), 1.45 – 1.39 (m, 3H, CH_3), 1.33 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 174.3 (C=O acid), 156.3 & 155.9 (C=O Cbz), 136.5 (C_q Ar), 128.5, 128.1, 128.0 (CH Ar), 109.2 (OCO), 72.1 (C3), 71.6 (C4), 67.4 (CH_2 Cbz), 43.1 & 42.7 (C2), 41.6 (C5), 38.3 & 38.1 (C6), 26.6 & 26.5 (CH_3), 24.7 & 24.6 (CH_3) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{17}H_{21}NO_6Na$ 358.12611, found 358.12580.

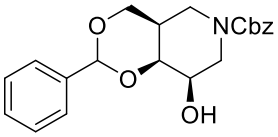
Compound 11



Method A (deprotection of compound **33**): Carboxylic acid **33** (135 mg, 0.403 mmol) was dissolved in a mixture of dioxane (9 mL) and aq. 4 M HCl (1 mL), 10% Pd/C (100 mg) was added subsequently. The mixture was stirred vigorously under a balloon of hydrogen for 16 h, filtered over a Whatman filter and concentrated to give a white residue (130 mg). NMR analysis revealed removal of the isopropylidene acetal while the Cbz-group was still present. Therefore the residue was re-dissolved in a mixture of THF (10 mL) and water (500 µL), aq. 8 M HCl (100 µL) and 10% Pd/C (139 mg) were added. After stirring vigorously overnight under hydrogen atmosphere, the mixture was filtered over a Whatman filter and concentrated to give a crude product, of which NMR analysis proved removal of the Cbz-group. The residue was then dissolved in absolute ethanol (1.2 mL) and added dropwise to a stirred solution of dry diethyl ether (20 mL) affording a white precipitate. The precipitate was collected on a filter, washed with dry diethyl ether (2 x 5 mL) and dried *in vacuo* to give the target compound **11** (85 mg, quant) as an off-white powder.

Method B (deprotection of compound **S8**): To a stirred solution of diol **S8** (40 mg, 0.126 mmol) in dry diethyl ether (1 mL) at 0 °C was added 4 M HCl/dioxane (4 mL). The mixture was stirred at 0 °C for 10 min and at rt for 4 h. At which LC-MS analysis revealed complete removal of the Boc group while the *tert*-butyl ester was still present in large amount. Water (200 µL) was added and the mixture was left stirring overnight until LC-MS confirmed full conversion. The solvents were removed *in vacuo* and the residue was re-dissolved in anhydrous 2-propanol (1 mL). Upon addition of dry diethyl ether a white precipitate appeared. The supernatant was taken out carefully and the residue washed with dry diethyl ether (3 x 2 mL). Evaporation of the volatiles afforded the target compound (19 mg, 0.096 mmol, 76%) as a white foam. ^{1}H NMR (400 MHz, D_2O at 293K) δ 4.42 (s, 1H, H4), 3.99 (dt, J = 11.2, 3.6 Hz, 1H, H3), 3.40 (dd, J = 12.9, 4.4 Hz, 1H, H6b), 3.29 – 3.18 (m, 2H, H2b and H6a), 3.09 – 2.99 (m, 2H, H2a and H5). ^{13}C NMR (101 MHz, D_2O at 293K) δ 173.3 (C=O), 66.8 (C4), 65.3 (C3), 42.7 (C5), 41.9 (C2), 38.5 (C6). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_6H_{12}NO_4$ 162.07608, found 162.07572.

Compound S6

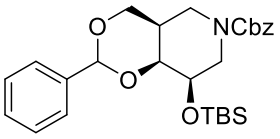


To a stirred solution of alcohol **32** (3.65 g, 11.4 mmol) in MeOH (100 mL) was added 8 M HCl solution (2.24 mL) and the mixture stirred at room temperature overnight until TLC-analysis confirmed full conversion of the

starting material. The reaction was quenched by addition of Et_3N (5 mL). The solvent was evaporated and the residue treated with EtOAc (100 mL) and filtered. The filter cake was washed with EtOAc (3 x 10 mL) and the combined filtrates were concentrated to afford the crude triol that was directly dissolved in DMF (80 mL). After addition of benzaldehyde dimethyl acetal (3.40 mL, 22.6 mmol) and (+)-camphorsulfonic acid (1.57 g, 6.76 mmol), the mixture was stirred at 60 °C for 20 h until LC-MS analysis showed full conversion. After cooling to rt, sat. aq. $NaHCO_3$ (300 mL) was added and the mixture was extracted with EtOAc (3 x 130 mL). The combined organic layers were washed with brine (2 x 50 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 4:1→1:1) affording compound **S6** (3.03 g, 8.20 mmol, 72% over two steps) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.49 – 7.39 (m, 2H, CH Ar), 7.40 – 7.26 (m, 8H, CH Ar), 5.53 (s, 1H, CHPh), 5.13 (dd, J = 12.0 Hz, 2H, CH_2 Cbz), 4.24 (s, 1H, H4), 4.17 (d, J = 9.1 Hz, 1H, H2b), 4.09 – 3.96 (m, 3H, H6b and H7ab), 3.63 (ddd, J = 10.9, 5.2, 3.3 Hz, 1H, H3), 3.41 (t, J = 12.7 Hz, 1H, H6a), 2.89 (t, J = 11.9 Hz, 1H, H2a), 2.35 (s, 1H, OH), 1.68 (d, J = 9.3 Hz, 1H, H5). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 155.6 (C=O), 138.2 (C_q Ar), 136.9 (C_q Ar), 129.3, 128.7, 128.5, 128.2, 128.1, 126.3 (CH Ar), 102.0 (OCHO), 76.1 (C4), 68.7 (C7), 67.8 (C3), 67.5 (CH_2 Cbz), 45.2 (C2), 41.5 (C6), 34.3 (C5) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) δ 7.45 (dd, J = 6.6, 3.1 Hz, 2H, CH Ar), 7.42 – 7.26 (m, 8H, CH Ar), 5.53 (s, 1H, CHPh), 5.12 (s, 2H, CH_2 Cbz), 4.29 – 3.89 (m, 5H, H4, H2b, H6b and H7ab), 3.62 (s, 1H, H3), 3.38 (m, 1H, H6a), 2.95 – 2.82 (m, 1H, H2a), 2.60 (d, J = 10.5 Hz, 1H, OH), 1.79 – 1.59 (m, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) δ 155.4 (C=O), 137.9 (C_q Ar), 137.2 (C_q Ar), 129.4, 128.6, 128.4, 128.2, 128.0, 126.3 (CH Ar), 101.8 (OCHO), 75.9 (C4), 68.6 (C7), 67.7 (C3), 67.4 (CH_2 Cbz), 44.9 (C2), 41.2 (C6), 33.8 (C5) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{21}H_{24}NO_5$ 370.1649, found 344.1644.

Compound 34

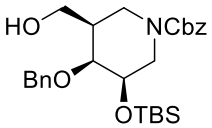


Alcohol **S6** (2.35 g, 6.36 mmol) was dissolved in DCM (35 mL) and imidazole (1.96 g, 28.8 mmol) and TBSCl (2.3 g, 15 mmol) were added subsequently. The resulting suspension was stirred for 20 h at room temperature. The mixture was diluted with EtOAc (150 mL), washed with water and brine, dried over $MgSO_4$, filtered and concentrated in *vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 95:5→9:1) to afford the target compound **34** (2.88 g, 5.95 mmol, 94%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.48 (d, J = 7.4 Hz, 2H, CH Ar), 7.38 – 7.23 (m, 8H, CH Ar), 5.51 (s, 1H, CHPh), 5.18 (d, J = 12.4 Hz, 1H, CH*H* Cbz), 5.10 (d, J = 12.4 Hz, 1H, CH*H* Cbz), 4.13 (s, 1H, H4), 4.08 – 3.89 (m, 4H, H7ab, H6b and H2b), 3.68 (d, J = 9.9 Hz, 1H, H3), 3.45 (t, J = 12.5 Hz, 1H, H6a), 3.08 (t, J = 11.5 Hz, 1H, H2a), 1.64 (d, J = 10.1 Hz, 1H, H5), 0.88 (s, 9H, $(CH_3)_3C$), 0.06 (s, 6H, $2CH_3$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 155.7 (C=O), 138.8, 137.1 ($2C_q$ Ar), 128.9, 128.7, 128.3, 128.2, 128.1, 126.2 (CH Ar), 101.5 (CHPh), 77.2 (C4), 69.9 (C3), 69.0

(C7), 67.4 (CH_2 Cbz), 45.1 (C2), 41.8 (C6), 34.7 (C5), 25.9 ((CH_3)$_3$C), 18.3 ((CH_3)$_3$*C*), -4.4 (CH_3), -4.3 (CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.49 (m, 2H, CH Ar), 7.37 – 7.26 (m, 8H, CH Ar), 5.51 (s, 1H, CHPh), 5.22 – 5.14 (m, 1H, CH*H* Cbz), 5.09 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 4.11 (s, 1H, H4), 4.10 – 3.84 (m, 4H, H7ab, H6b and H2b), 3.68 (m, 1H, H3), 3.50 – 3.36 (m, 1H, H6a), 3.09 (m, 1H, H2a), 1.62 (s, 1H, H5), 0.89 (s, 9H, (CH_3)$_3$C), 0.09 (s, 3H, CH_3), 0.06 (s, 3H, CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 155.4 (C=O), 138.4, 136.7 (2C_q Ar), 128.7, 128.5, 128.1, 128.0, 127.9, 126.0 (CH Ar), 101.2 (OCHO), 76.8 (C4), 69.6 & 69.2 (C3), 68.7 (C7), 67.2 (CH_2 Cbz), 44.7 & 44.5 (C2), 41.3 (C6), 34.4 & 34.1 (C5), 25.8 ((CH_3)$_3$C), 18.2 ((CH_3)$_3$*C*), -4.6 (2CH_3) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{27}H_{37}NO_5SiNa$ 506.2333, found 506.2338.

Compound 35

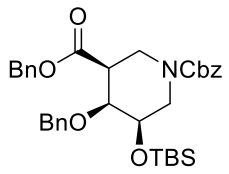


Compound **34** (2.59 g, 5.36 mmol) was dissolved in dry DCM (60 mL) and cooled on an ice-bath. BH_3·THF (1.0 M in THF, 22 mL, 22 mmol) was added dropwise in five minutes and followed by addition of TMSOTf (0.16 mL, 0.88 mmol). The resulting solution was stirred for 10 minutes on the ice-bath and for 3 h at room temperature. TLC-analysis confirmed full conversion and the reaction was quenched with TEA (2.5 mL) at 0 ºC followed by careful addition of MeOH (25 mL). After stirring for another 1 h, the reaction was concentrated, co-evaporated twice with MeOH to give the crude product. Purification by silica gel column chromatography (pentane/EtOAc 4:1→1:1) afforded the target compound **35** (2.53 g, 5.21 mmol, 97%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.35 – 7.22 (m, 10H, CH Ar), 5.18 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 5.05 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 4.94 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 4.59 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 3.85 – 3.75 (m, 2H, H4 and H2b), 3.74 – 3.65 (m, 2H, H3 and H6b), 3.62 (m, 2H, H7ab), 3.38 – 3.30 (m, 1H, H2a), 3.13 (dd, *J* = 12.9, 10.7 Hz, 1H, H6a), 2.05 (s, 1H, OH), 1.84 (s, 1H, H5), 0.91 (s, 9H, (CH_3)$_3$C), 0.10 (s, 6H, 2CH_3). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 155.7 (C=O), 139.1, 137.0 (2C_q Ar), 128.5, 128.5, 128.0, 128.0, 127.9, 127.7 (CH Ar), 77.7 (C4), 74.1 (CH_2 Bn), 71.7 (C3), 67.3 (CH_2 Cbz), 61.9 (C7), 46.1 (C2), 42.4 (C5), 41.7 (C6), 25.9 ((CH_3)$_3$C), 18.1 ((CH_3)$_3$*C*), -4.7, -4.7 (2CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.45 – 7.16 (m, 10H, CH Ar), 5.19 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 5.08 – 4.89 (m, 2H, CH*H* Cbz and CH*H* Bn), 4.60 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 3.97 – 3.51 (m, 6H, H4, H2b, H3, H6b and H7ab), 3.46 – 3.22 (m, 1H, H2a), 3.20 – 3.06 (m, 1H, H6a), 2.36 (s, 0.5H, OH), 2.18 (s, 0.5H, OH), 1.94 – 1.76 (m, 1H, H5), 0.92 (s, 4.5H, (CH_3)$_3$C), 0.89 (s, 4.5H, (CH_3)$_3$C), 0.22 – 0.01 (m, 6H, 2CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 155.6 (C=O), 138.9, 136.7 (2C_q Ar), 128.5, 128.5, 128.1, 127.9, 127.9, 127.8 (CH Ar), 77.4 (C4), 74.1 & 73.9 (CH_2 Bn), 71.4 (C3), 67.3 (CH_2 Cbz), 62.1 & 61.6 (C7), 46.1 & 45.6 (C2), 42.1

& 41.9 (C5), 41.6 & 41.3 (C6), 25.8 ((CH_3)$_3$C), 18.1 ((CH_3)$_3$*C*), -4.8 (2CH_3) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{27}H_{40}NO_5Si$ 486.26703, found 486.26696.

Compound 36

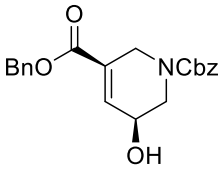


To a stirred solution of alcohol **35** (930 mg, 1.92 mmol) in acetone (40 mL) was added 2.4 M Jones reagent dropwise until an orange/brown color persisted. The reaction was stirred for an additional 10 minutes and then quenched by the addition of 2-propanol (3.0 mL). Subsequently sat. aq. $NaHCO_3$ (50 mL) and water (50 mL) were added and the acetone was carefully (foaming) evaporated from the mixture. Next acetic acid was added drop by drop until pH = 5. EtOAc (50 mL) was added and the mixture stirred vigorously for a few minutes after which the mixture was filtered over a pad of celite. The layers were separated and the water layer extracted with EtOAc (50 mL). The combined organic layers were dried over $MgSO_4$, filtered and concentrated to give the crude carboxylic acid (0.95 g) as a colorless oil. The crude acid was dissolved in DCM (20 mL) and benzyl alcohol (540 µL, 5.1 mmol) and DMAP (33 mg, 0.27 mmol) were added. After cooling on an ice-bath DIC (320 µL, 2.05 mmol) was added dropwise and the mixture stirred for one more hour on the ice-bath and for 6 h at rt. At that time TLC indicated circa 50% conversion, so the reaction was left overnight. However, the situation did not change. Extra portions of benzyl alcohol (540 µL) and DIC (320 µL) were added and the reaction stirred for an extra 6 h upon which TLC did not show any further progress of the reaction. The mixture was concentrated and purified by column chromatography (pentane/EtOAc 95:5→4:1) to afford the benzyl ester **36** as a colorless oil (520 mg, 0.882 mmol) in 46% yield. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.34 – 7.16 (m, 15H, CH Ar), 5.17 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 5.11 (d, *J* = 12.3 Hz, 1H, CH*H* Bn_{ester}), 5.07 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 4.99 (d, *J* = 12.3 Hz, 1H, CH*H* Bn_{ester}), 4.95 (d, *J* = 11.3 Hz, 1H, CH*H* Bn), 4.47 (d, *J* = 11.3 Hz, 1H, CH*H* Bn), 4.26 (s, 1H, H6b), 4.23 (s, 1H, H4), 3.98 (s, 1H, H2b), 3.68 (d, *J* = 7.6 Hz, 1H, H3), 3.29 (app t, *J* = 12.7 Hz, 1H, H6a), 3.13 (app t, *J* = 11.6 Hz, 1H, H2a), 2.61 (d, *J* = 9.7 Hz, 1H, H5), 0.91 (s, 9H, (CH_3)$_3$C), 0.10 (s, 6H, 2CH_3) ppm. ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 170.3 (OC=O), 155.4 (NC=O), 139.3, 136.9, 135.9 (3C_q Ar), 128.6, 128.6, 128.4, 128.2, 128.1, 128.0, 127.4, 127.4 (CH Ar), 77.9 (C4), 75.2 (CH_2 Bn), 71.5 (C3), 67.4 (CH_2 Cbz), 66.6 (CH_2 Bn_{ester}), 46.3 (C5), 45.0 (C2), 39.9 (C6), 25.9 ((CH_3)$_3$C), 18.1 ((CH_3)$_3$*C*), -4.6, -4.7 (2CH_3) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.41 – 7.14 (m, 15H, CH Ar), 5.18 (d, *J* = 12.3 Hz, 1H, CH*H* Cbz), 5.12 (d, *J* = 12.2 Hz, 1H, CH*H* Bn_{ester}), 5.06 (d, *J* = 12.3 Hz, 1H, CH*H* Cbz), 4.98 (m, 2H, CH*H* Bn_{ester} and CH*H* Bn), 4.46 (d, *J* = 11.3 Hz, 1H, CH*H* Bn), 4.37 – 4.14 (m, 3H, H6b and H4), 4.12 – 3.84 (m, 1H, H2b), 3.77 – 3.58 (m, 1H, H3), 3.39 – 3.21 (m, 1H, H6a), 3.21 – 3.00 (m, 1H, H2a), 2.63 (s, 1H, H5), 0.90 (s, 9H, (CH_3)$_3$C), 0.24 – 0.02 (m, 6H, 2CH_3). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 170.4 (OC=O), 155.3 (NC=O), 139.1, 136.6, 135.6 (3C_q Ar), 128.6, 128.6, 128.4, 128.2, 128.1, 128.0, 127.4 (CH Ar), 77.6 (C4), 75.1 (CH_2 Bn), 71.4 & 71.0 (C3), 67.4 (CH_2 Cbz), 66.6 (CH_2 Bn_{ester}), 46.1 & 45.9 (C5), 44.8 (C2), 39.6 (C6), 25.8 ((CH_3)$_3$C), 18.1

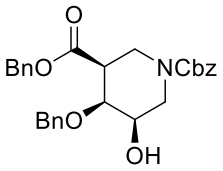
((CH_3)$_3$*C*), -4.8 (2CH_3) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{34}H_{44}NO_6Si$ 590.29324, found 590.29347.

Compound 37



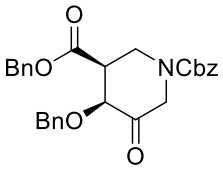
To a stirred solution of TBS-ether **36** (0.52 g, 0.88 mmol) in THF (20 mL) was added TBAF (75 wt% in H_2O, 680 mg, 1.95 mmol) and the mixture was left stirring overnight. The mixture was concentrated under reduced pressure to remove most of the solvent, then diluted with EtOAc, washed with brine, dried over $MgSO_4$, filtered and concentrated to give the crude product. Purification by silica gel column chromatography (pentane/EtOAc 9:1→1:1) afforded a colorless oil which after NMR and LC-MS analyses turned out to be the title unsaturated ester **37** (0.32 mg, 0.87 mmol, 98%). ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.36 – 7.22 (m, 10H, CH Ar), 6.98 (d, *J* = 2.6 Hz, 1H, H4), 5.17 (s, 2H, CH_2 Bn_{ester}), 5.12 (s, 2H, CH_2 Cbz), 4.28 (s, 1H, H3), 4.23 (d, *J* = 18.6 Hz, 1H, H6b), 4.13 (d, *J* = 18.6 Hz, 1H, H6a), 3.84 (dd, *J* = 12.8, 3.8 Hz, 1H, H2b), 3.25 (dd, *J* = 12.8, 6.2 Hz, 1H, H2a), 2.84 (br. s, 1H, OH). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 164.8 (OC=O), 155.6 (NC=O), 139.4 (C4), 136.5 (C_q Ar), 135.8 (C_q Ar), 128.7, 128.6, 128.4, 128.2, 128.2, 128.0 (CH Ar), 67.6 (CH_2 Cbz), 66.7 (CH_2 Bn_{ester}), 63.4 (C3), 47.0 (C2), 42.8 (C6).

Compound 38



To an ice cold solution of TBS-ether **36** (325 mg, 0.551 mmol) in THF (8 mL) was added TBAF (75 wt% in H_2O, 420 mg, 1.20 mmol) and the mixture was left stirring on the ice bath. After 5 h TLC-analysis showed complete conversion of the TBS-ether. The mixture was diluted with EtOAc (40 mL), washed with brine (2 x 15 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by column chromatography (pentane/EtOAc 95:5→7:3) to afford the title alcohol **38** (247 mg, 0.519 mmol, 94%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$ at 333K) δ 7.36 – 7.12 (m, 15H, CH Ar), 5.07 (m, 4H, CH_2 Cbz and CH_2 Bn_{ester}), 4.60 (m, 1H, CH*H* Bn), 4.48 (d, *J* = 11.6 Hz, 1H, CH*H* Bn), 4.26 – 3.83 (m, 3H, H4, H6b and H2b), 3.61 (brs, 1H, H3), 3.35 (app. t, *J* = 12.3 Hz, 1H, H6a), 3.12 – 2.78 (m, 2H, H2a and OH), 2.61 (brs, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$ at 333K) δ 170.7 (OC=O), 155.2 (NC=O), 138.3, 136.4, 135.2 (3C_q Ar), 128.6, 128.5, 128.5, 128.5, 128.4, 128.1, 127.9, 127.7, 127.4 (CH Ar), 76.9 (C4), 74.3 (CH_2 Bn), 68.5 (C3), 67.4 (CH_2 Cbz), 66.8 (CH_2 Bn_{ester}), 45.6 & 45.3 (C5), 45.2 & 44.9 (C2), 39.7 (C6). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{28}H_{29}NO_6Na$ 498.18871, found 498.18891.

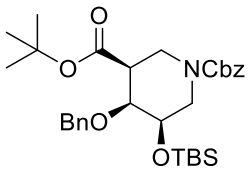
Compound 39



To a stirred solution of compound **38** (225 mg, 0.474 mmol) in DCM (9 mL) was added Dess-Martin periodinane (60%, 682 mg, 0.965 mmol) and the mixture was left stirring at rt for 3 h. The reaction mixture was then diluted with EtOAc (50 mL), washed with aq. 2 M $Na_2S_2O_3$ (2 x 15 mL), sat. aq. $NaHCO_3$ (2 x 15 mL) and brine (15 mL), dried over $MgSO_4$, filtered and concentration *in vacuo*. The crude was purified by silica

gel column chromatography (pentane/EtOAc 4:1→7:3) to afford the ketone **39** (160 mg, 0.338 mmol, 71%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$ at 293K) δ 7.37 – 7.14 (m, 15H, CH Ar), 5.19 – 4.94 (m, 4H, CH_2 Cbz and CH_2 Bn_{ester}), 4.69 (d, *J* = 12.0 Hz, 1H, CH*H* Bn), 4.52 – 4.41 (m, 1H, CH*H* Bn), 4.35 – 3.77 (m, 5H, H2ab, H6ab and H4), 3.16 (brs, 1H, H5). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) δ 200.3 (C=O), 169.5 (OC=O), 154.9 (NC=O), 136.8, 136.0, 135.2 ($3C_q$ Ar), 128.5, 128.5, 128.4, 128.3, 128.3, 128.1, 127.8, 127.4 (CH Ar), 78.31 (C4), 72.4 (CH_2 Bn), 67.8 (CH_2 Cbz), 67.1 (CH_2 Bn_{ester}), 52.7 (C2), 46.4 (C5), 41.8 (C6) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{28}H_{29}NO_6Na$ 496.17306, found 496.17304.

Compound 40

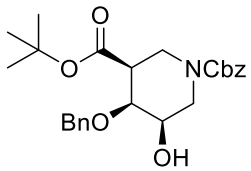


Alcohol **35** (1.25 g, 2.58 mmol) was dissolved in dry DCM (25 mL). Dess-Martin periodinane (60% purity, 2.65 g, 3.75 mmol) was added and the reaction was stirred at room temperature for 2 h. The reaction mixture was diluted with EtOAc (100 mL) and washed successively with sat. aq. $Na_2S_2O_3$ (2 x 50 mL), sat. aq. $NaHCO_3$ (2 x 50 mL) and brine (50 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford the crude aldehyde (1.47 g). To the stirred solution of the crude aldehyde in MeCN (45 mL) and water (8.5 mL), $NaH_2PO_4 \cdot 2H_2O$ (3.12 g, 20.0 mmol) was added and the mixture was stirred at rt until it turned homogeneous. After which, the reaction was cooled to 0 °C and 30% H_2O_2 (340 µL, 4.35 mmol) and a solution of $NaClO_2$ (80% purity, 375 mg, 3.31 mmol) in water (8.5 mL) were added. The mixture was stirred at 0 °C for 15 minutes and 1 h at rt until TLC-analysis confirmed full conversion. The reaction was quenched by addition of Na_2SO_3 (0.95 g) and stirred for 10 more minutes. The mixture was diluted with EtOAc (60 mL) and the layers were separated. The water layer was extracted with EtOAc (2 x 30 mL) and the combined organic layers were washed with brine (30 mL), dried over $MgSO_4$, filtered and concentrated to afford the crude carboxylic acid (1.47 g). The crude acid was co-evaporated twice with dry toluene and then dissolved in dry toluene (10 mL). *tert*-Butyl *N,N'*-diisopropylcarbaimidate (1.94 g, 9.70 mmol) was added and the reaction stirred at 60 °C for 5 h. LC-MS analysis confirmed complete conversion of the acid and the mixture was concentrated and filtered over a plug of silica gel eluting with pentane/Et_2O (9:1) to afford the crude ester. Purification by silica gel column chromatography (pentane/Et_2O 98:2→9:1) afforded target compound **40** (1.01g, 1.82 mmol, 71% over three steps) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.33 – 7.24 (m, 9H, CH Ar), 7.22 – 7.16 (m, 1H, CH Ar), 5.18 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 5.10 – 5.01 (m, 2H, CH*H* Cbz and CH*H* Bn), 4.59 (d, *J* = 11.1 Hz, 1H, CH*H* Bn), 4.26 – 4.16 (m, 2H, H4 and H6b), 3.98 (s, 1H, H3), 3.67 (d, *J* = 7.3 Hz, 1H, H2b), 3.20 (t, *J* = 12.7 Hz, 1H, H6a), 3.13 (t, *J* = 11.6 Hz, 1H, H2a), 2.52 – 2.44 (m, 1H, H5), 1.40 (s, 9H, $(CH_3)_3CO$), 0.91 (s, 9H, $(CH_3)_3CSi$), 0.11 (s, 3H, CH_3Si), 0.10 (s, 3H, CH_3Si). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 169.5 (OC=O), 155.3 (NC=O), 139.3, 136.9 ($2C_q$ Ar), 128.5, 128.1, 128.0, 127.9, 127.2 (CH Ar), 81.2 ($(CH_3)_3CO$), 78.1 (C4), 75.2 (CH_2 Bn), 71.6 (C3),

67.2 (CH_2 Cbz), 46.9 (C5), 44.9 (C2), 40.0 (C6), 28.1 ((CH_3)$_3$CO), 25.8 ((CH_3)$_3$CSi), 18.0 ((CH_3)$_3$CSi), -4.7, -4.8 (2CH_3Si) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.37 – 7.18 (m, 10H, CH Ar), 5.23 – 5.15 (d, *J* = 12.4 Hz, 1H, CH*H* Cbz), 5.06 (m, 2H, CH*H* Cbz and CH*H* Bn), 4.58 (d, *J* = 11.1 Hz, 1H, CH*H* Bn), 4.31 – 3.85 (m, 3H, H4, H6b and H2b), 3.78 – 3.57 (m, 1H, H3), 3.30 – 3.01 (m, 2H, H2a and H6a), 2.51 (s, 1H, H5), 1.41 (s, 9H, (CH_3)$_3$CO), 0.91 (s, 9H, (CH_3)$_3$CSi), 0.16 (s, 3H, CH_3Si), 0.09 & 0.06 (s, 3H, CH_3Si). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 169.6 (OC=O), 155.1 (NC=O), 139.1, 136.6 (2C_q Ar), 128.4, 128.0, 128.0, 127.8, 127.2 (CH Ar), 81.2 ((CH_3)$_3$*C*O), 77.9 (C4), 75.2 (CH_2 Bn), 71.4 & 71.1 (C3), 67.2 (CH_2 Cbz), 46.5 (C5), 44.7 (C2), 39.8 (C6), 28.0 ((CH_3)$_3$CO), 25.7 ((CH_3)$_3$CSi), 18.0 ((CH_3)$_3$*C*Si), -4.9 (2CH_3Si) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{31}H_{45}NO_6SiNa$ 578.29084, found 578.29100.

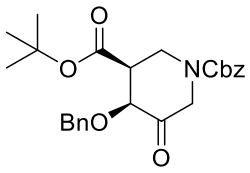
Compound S7



To an ice cold solution of TBS-ether **40** (0.35 g, 0.63 mmol) in THF (8 mL) was added TBAF (75 wt% in H_2O, 430 mg, 1.24 mmol) and the mixture was left stirring on the ice bath. After 5 h TLC-analysis showed complete conversion of the TBS-ether. The mixture was diluted with EtOAc (40 mL), washed with brine (2 x 10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 9:1→4:1) to afford the title alcohol **S7** (267 mg, 0.605 mmol, 96%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 7.34 – 7.19 (m, 10H, CH Ar), 5.11 (s, 2H, CH_2 Cbz), 4.73 (dd, *J* = 11.4 Hz, 2H, CH_2 Bn), 4.20 (s, 1H, H4), 4.06 (dd, *J* = 13.5, 4.4 Hz, 1H, H6b), 3.92 (d, *J* = 8.9 Hz, 1H, H2b), 3.61 (s, 1H, H3), 3.32 (dd, *J* = 13.5, 11.0 Hz, 1H, H6a), 3.11 – 3.03 (m, 1H, H2a), 2.78 (d, *J* = 7.6 Hz, 1H, OH), 2.54 – 2.46 (m, 1H, H5), 1.43 (s, 9H, (CH_3)$_3$C). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 170.1 (OC=O), 155.4 (NC=O), 138.6, 136.8 (2C_q Ar), 128.5, 128.4, 128.0, 127.8, 127.7, 127.4 (CH Ar), 81.7 ((CH_3)$_3$*C*), 77.3 (C4), 74.4 (CH_2 Bn), 68.8 (C3), 67.3 (CH_2 Cbz), 46.5 (C5), 45.5 (C2), 40.2 (C6), 28.1 ((CH_3)$_3$C) ppm.

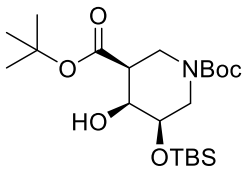
^{1}H NMR (500 MHz, $CDCl_3$ at 293K) [mixture of rotamers 1:1] δ 7.48 – 6.99 (m, 10H, CH Ar), 5.09 (s, 2H, CH_2 Cbz), 4.88 – 4.62 (m, 2H, CH_2 Bn), 4.20 (s, 1H, H4), 4.13 – 3.80 (m, 2H, H6b and H2b), 3.71 – 3.54 (m, 1H, H3), 3.35 – 3.27 (m, 1H, H6a), 3.22 – 2.98 (m, 2H, H2a and OH), 2.50 (s, 1H, H5), 1.43 (s, 9H, (CH_3)$_3$C). ^{13}C NMR (126 MHz, $CDCl_3$) [mixture of rotamers 1:1] δ 170.1 (OC=O), 155.3 (NC=O), 138.3, 136.5 (2C_q Ar), 128.4, 128.3, 128.0, 127.8, 127.6, 127.3 (CH Ar), 81.7 ((CH_3)$_3$*C*), 77.0 (C4), 74.3 (CH_2 Bn), 68.6 & 68.5 (C3), 67.3 (CH_2 Cbz), 46.4 & 46.2 (C5), 45.3 & 45.0 (C2), 40.0 (C6), 27.9 ((CH_3)$_3$C) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{25}H_{31}NO_6Na$ 464.20436, found 464.20419.

Compound 41



Compound **S7** (135 mg, 0.306 mmol) was dissolved in dry DCM (6 mL) and Dess-Martin periodinane (252 mg, 0.594 mmol) was added. The reaction was stirred at rt for 2 h until TLC-analysis indicated complete conversion of the alcohol. The mixture was diluted with EtOAc (25 mL) and washed subsequently with aq. 2 M $Na_2S_2O_3$ (2 x 10 mL), sat. aq. $NaHCO_3$ (2 x 10 mL) and brine (10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 4:1→7:3) to afford the target ketone **41** (119 mg, 0.271 mmol, 89%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$ at 333K) δ 7.39 – 7.21 (m, 10H, CH Ar), 5.10 (s, 2H, CH_2 Cbz), 4.76 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.57 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.30 (d, *J* = 17.0 Hz, 1H, H2b), 4.08 (br. s, 1H, H6b), 4.00 (d, *J* = 4.5 Hz, 1H, H4), 3.96 (br. s, 1H, H2a), 3.79 (d, *J* = 11.4 Hz, 1H, H6a), 3.05 (s, 1H, H5), 1.38 (s, 9H, $(CH_3)_3C$). ^{13}C NMR (101 MHz, $CDCl_3$ at 333K) δ 200.6 (C=O), 168.8 (OC=O), 154.8 (NC=O), 137.0, 136.0 ($2C_q$ Ar), 128.6, 128.5, 128.3, 128.1, 128.0, 127.8 (CH Ar), 82.4 ($(CH_3)_3\mathit{C}$), 78.8 (C4), 72.4 (CH_2 Bn), 67.8 (CH_2 Cbz), 52.9 (C2), 47.4 (C5), 42.3 (C6), 27.8 ($(\mathit{C}H_3)_3$C) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{25}H_{31}NO_6Na$ 462.18871, found 462.18832.

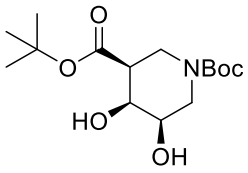
Compound 42 and S8



Compound **40** (700 mg, 1.26 mmol) was dissolved in methanol (25 mL) under argon and 10% Pd/C (140 mg) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 2 h under H_2 atmosphere TLC and LC-MS analyses confirmed complete removal of the Cbz-group. The mixture was filtered over a Whatman filter and concentrated. The crude residue was directly taken up in DCM (10 mL) and DIPEA (250 µL, 1.44 mmol) and Boc_2O (700 mg, 3.21 mmol) were added. The mixture was stirred at rt overnight until LC-MS analysis confirmed complete *N*-protection. The mixture was concentrated, and eluted over a small plug of silica gel (pentane / EtOAc = 7 / 3) and the eluate was concentrated again. The material was redissolved in methanol (20 mL) under argon and 10% Pd/C (140 mg) was added. After stirring for 16 h under H_2 atmosphere, LC-MS indicated only circa 20% conversion of the starting material, so more 10% Pd/C (280 mg) and 20% $Pd(OH)_2$/C (300 mg) were added and the mixture was stirred over the weekend until LC-MS analysis indicated full conversion. The mixture was filtered over a Whatman filter and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 9:1→4:1) to afford compound **42** (372 mg, 0.862 mmol, 69% over three steps) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.20 (s, 1H, H4), 4.02 (d, *J* = 11.3 Hz, 1H, H6b), 3.85 (d, *J* = 7.4 Hz, 1H, H2b), 3.60 (ddd, *J* = 10.6, 5.2, 2.8 Hz, 1H, H3), 3.11 (t, *J* = 12.7 Hz, 1H, H6a), 2.88 (t, *J* = 10.9 Hz, 1H, H2a), 2.44 (m, 2H, H5 and OH), 1.47 (s, 9H, $(CH_3)_3CO$), 1.46 (s, 9H, $(CH_3)_3CO$), 0.91 (s, 9H, $(CH_3)_3CSi$), 0.13 (s, 6H, $2CH_3Si$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 169.9 (OC=O), 154.7 (NC=O), 81.2, 79.9 ($(CH_3)_3\mathit{C}O$), 69.5 (C4), 69.4 (C3), 46.1

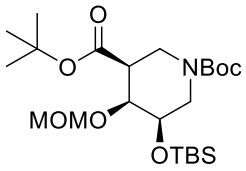
(C5), 44.1 (C2), 39.1 (C6), 28.4, 28.1 ((CH_3)$_3$CO), 25.8 ((CH_3)$_3$CSi), 18.1 ((CH_3)$_3$CSi), -4.6 (CH_3Si), -4.8 (CH_3Si) ppm.

^{1}H NMR (400 MHz, $CDCl_3$ at 293K) δ 4.21 (s, 1H, H4), 4.16 – 3.71 (m, 2H, H6b and H2b), 3.60 (s, 1H, H3), 3.10 (s, 1H, H6a), 3.01 – 2.67 (m, 1H, H2a), 2.58 – 2.37 (m, 2H, H5 and OH), 1.48 (s, 9H, (CH_3)$_3$CO), 1.46 (s, 9H, (CH_3)$_3$CO), 0.91 (s, 9H, (CH_3)$_3$CSi), 0.13 (s, 3H, CH_3Si), 0.12 (s, 3H, CH_3Si). ^{13}C NMR (101 MHz, $CDCl_3$ at 293K) δ 170.2 (OC=O), 154.7 (NC=O), 81.4, 80.1 ((CH_3)$_3$CO), 69.4 (C4), 69.2 (C3), 45.9 (C5), 44.5 (C2), 38.7 (C6), 28.5, 28.1 ((CH_3)$_3$CO), 25.8 ((CH_3)$_3$CSi), 18.1 ((CH_3)$_3$CSi), -4.6 (2CH_3Si) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{21}H_{41}NO_6SiNa$ 454.25954, found 454.25934.



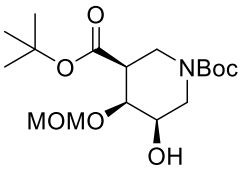
Further elution from the column with pentane/EtOAc = 1:1 afforded a byproduct (90 mg, 0.28 mmol, 22%) that was identified as diol **S8**. ^{1}H NMR (400 MHz, MeOD at 293K) δ 4.26 (app t, *J* = 2.3 Hz, 1H, H4), 4.03 – 3.91 (m, 1H, H6b), 3.90 – 3.78 (m, 1H, H2b), 3.50 (ddd, *J* = 10.9, 5.1, 2.7 Hz, 1H, H3), 3.19 – 3.00 (m, 1H, H6a), 2.99 – 2.75 (m, 1H, H2a), 2.49 (ddd, *J* = 11.8, 4.7, 2.5 Hz, 1H, H5), 1.48 (s, 9H, (CH_3)$_3$CO), 1.46 (s, 9H, (CH_3)$_3$CO). ^{13}C NMR (101 MHz, MeOD at 293K) δ 172.0 (OC=O), 156.4 (NC=O), 82.2 ((CH_3)$_3$CO), 81.3 ((CH_3)$_3$CO), 69.9 (C4), 69.3 (C3), 47.4 (C5), 44.6 (C2, assign by HSQC), 39.8 (C6, assign by HSQC), 28.6 ((CH_3)$_3$CO), 28.1 ((CH_3)$_3$CO) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{15}H_{27}NO_6Na$ 340.17306, found 340.17280.

Compound 43



In a microwave tube compound **42** (0.18 g, 0.42 mmol) was dissolved in a mixture of dry DCM (4 mL), DIPEA (1.5 mL) and MOMCl (0.5 mL). The tube was sealed and heated for 1 h at 100 °C in a microwave (Biotage initiator+). After which the mixture was diluted with EtOAc (25 mL), washed with water (10 mL), sat. aq. $NaHCO_3$ (10 mL) and brine (10 mL), dried over $MgSO_4$, filtered and concentrated. The crude was purified by silica gel column chromatography (pentane/EtOAc 98:2→95:5) to give the target compound **43** (176 mg, 0.370 mmol, 88%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.81 (d, *J* = 5.8 Hz, 1H, CH*H*O MOM), 4.77 (d, *J* = 5.8 Hz, 1H, CH*H*O MOM), 4.19 (s, 1H, H4), 4.10 (d, *J* = 9.3 Hz, 1H, H6b), 3.86 (d, *J* = 9.1 Hz, 1H, H2b), 3.56 (ddd, *J* = 10.8, 4.9, 2.2 Hz, 1H, H3), 3.35 (s, 3H, OCH_3), 3.09 (t, *J* = 12.8 Hz, 1H, H6a), 2.99 (t, *J* = 11.3 Hz, 1H, H2a), 2.43 (ddd, *J* = 12.0, 4.5, 2.0 Hz, 1H, H5), 1.47 (s, 9H, (CH_3)$_3$CO), 1.46 (s, 9H, (CH_3)$_3$CO), 0.92 (s, 9H, (CH_3)$_3$CSi), 0.12 (s, 6H, 2CH_3Si). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 169.6 (OC=O), 154.8 (NC=O), 98.2 (CH_2 MOM), 81.3 ((CH_3)$_3$CO), 79.9 ((CH_3)$_3$CO), 76.0 (C4), 71.3 (C3), 56.2 (CH_3 MOM), 46.9 (C5), 44.8 (C2), 39.5 (C6), 28.5 ((CH_3)$_3$CO), 28.2 ((CH_3)$_3$CO), 25.9 ((CH_3)$_3$CSi), 18.2 ((CH_3)$_3$CSi), -4.6 (CH_3Si), -4.8 (CH_3Si) ppm. HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{23}H_{46}NO_7Si$ 476.30381, found 476.30378.

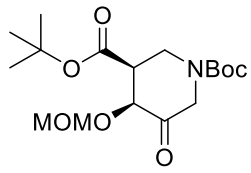
Compound S9



Compound **43** (175 mg, 0.368 mmol) was dissolved in THF (5 mL) and cooled on an ice-bath. TBAF (75 wt% in H_2O, 230 mg, 0.66 mmol) was added and the mixture was stirred on the ice bath for 5 h. The mixture was then diluted with EtOAc (30 mL), washed with brine (2 x 10 mL), dried over Na_2SO_4, filtered and concentrated. The crude was purified by silica gel column chromatography (pentane/EtOAc 95:5→ 7:3) to afford compound **S9** (127 mg, 0.351 mmol, 95%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.72 (d, J = 6.5 Hz, 1H, CH*H* MOM), 4.68 (d, J = 6.5 Hz, 1H, CH*H* MOM), 4.19 (t, J = 2.5 Hz, 1H, H4), 4.10 (d, J = 10.6 Hz, 1H, H6b), 4.00 – 3.86 (m, 2H, OH and H2b), 3.52 – 3.45 (m, 1H, H3), 3.44 (s, 3H, CH_3 MOM), 3.08 (dd, J = 13.5, 11.6 Hz, 1H, H6a), 2.83 (app t, J = 11.7 Hz, 1H, H2a), 2.54 (ddd, J = 11.4, 4.8, 2.5 Hz, 1H, H5), 1.46 (s, 18H, 2$(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 169.6 (OC=O), 154.9 (NC=O), 98.9 (CH_2 MOM), 81.4 (($CH_3)_3C$O), 80.8 (C4), 80.1 (($CH_3)_3C$O), 67.8 (CH_3 MOM), 56.2 (C3), 46.7 (C5), 45.5 (C2), 39.7 (C6), 28.5, 28.2 (($CH_3)_3$CO) ppm.

^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 4.73 (d, J = 6.7 Hz, 1H, CH*H* MOM), 4.69 (d, J = 6.7 Hz, 1H, CH*H* MOM), 4.27 – 3.88 (m, 4H, H4, OH, H2b and H6b), 3.54 – 3.47 (m, 1H, H3), 3.45 (s, 3H, CH_3 MOM), 3.12 – 2.99 (m, 1H, H6a), 2.89 – 2.70 (m, 1H, H2a), 2.57 (d, J = 9.3 Hz, 1H, H5), 1.46 (s, 18H, 2$(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 169.5 (OC=O), 154.8 (NC=O), 98.7 (CH_2 MOM), 81.4 (($CH_3)_3C$O), 81.1 (C4), 80.1 (($CH_3)_3C$O), 67.6 (CH_3 MOM), 56.2 (C3), 46.4 (C5), 28.4, 28.1 (($CH_3)_3$CO) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{17}H_{31}NO_7Na$ 384.19927, found 384.19904.

Compound 44

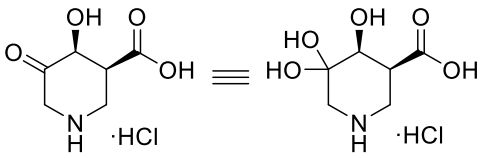


To a stirred solution of alcohol **S9** (126 mg, 0.349 mmol) in dry DCM (5 mL) was added Dess-Martin periodinane (248 mg, 0.585 mmol) and the mixture was stirred at rt for 3 h until complete conversion was confirmed by TLC and LC-MS analyses. The mixture was diluted with EtOAc (30 mL) and washed subsequently with aq. 2 M $Na_2S_2O_3$ (2 x 10 mL), sat. aq. $NaHCO_3$ (2 x 10 mL) and brine (10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/acetone 98:2→85:15) to afford ketone **44** (100 mg, 0.278 mmol, 80%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.76 (d, J = 6.8 Hz, 1H, CH*H* MOM), 4.73 (d, J = 6.8 Hz, 1H, CH*H* MOM), 4.31 (dd, J = 17.4, 1.2 Hz, 1H, H2b), 4.22 – 4.13 (m, 2H, H4 and H6b), 3.87 (d, J = 17.4 Hz, 1H, H2a), 3.68 (dd, J = 14.1, 4.2 Hz, 1H, H6a), 3.38 (s, 3H, CH_3 MOM), 3.13 (dd, J = 9.6, 4.8 Hz, 1H, H5), 1.45 (s, 9H, $(CH_3)_3CO$), 1.43 (s, 9H, $(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 200.3 (C=O), 169.2 (OC=O), 154.1 (NC=O), 96.6 (CH_2 MOM), 82.4, 81.0 (($CH_3)_3C$O), 77.5 (C4), 56.1 (CH_3 MOM), 53.6 (C2), 47.7 (C5), 43.1 (C6), 28.4, 28.0 (($CH_3)_3$CO) ppm.

^{1}H NMR (500 MHz, $CDCl_3$ at 293K) δ 4.80 (d, J = 6.9 Hz, 1H, CH*H* MOM), 4.75 (d, J = 6.9 Hz, 1H, CH*H* MOM), 4.43 – 4.11 (m, 3H, H2b, H4 and H6b), 3.96 – 3.75 (m, 1H, H2a), 3.75 – 3.52 (m, 1H,

H6a), 3.40 (s, 3H, CH_3 MOM), 3.19 (br. s, 1H, H5), 1.45 (s, 9H, $(CH_3)_3CO$), 1.44 (s, 9H, $(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 293K) δ 200.5 (C=O), 169.3 (OC=O), 154.0 (NC=O), 96.3 (CH_2 MOM), 82.4, 80.9 ($(CH_3)_3CO$), 77.3 (C4), 56.0 (CH_3 MOM), 54.0 (C2), 47.7 (C5), 43.0 (C6), 28.3, 27.9 ($(CH_3)_3CO$). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{17}H_{31}NO_7Na$ 382.18362, found 382.18337.

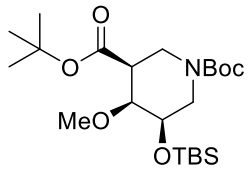
Compound 8



Method A (deprotection of ketone **39**): Ketone **39** (68 mg, 0.14 mmol) was dissolved in a mixture of THF (10 mL) and aqueous 3 M HCl (300 µL). Subsequently, 10% Pd/C (72 mg) was added and the mixture was stirred under hydrogen atmosphere for 18 h. The mixture was filtered over a Whatman® filter and concentrated *in vacuo* to afford the target compound (31 mg) as a white foam in quantitative yield.

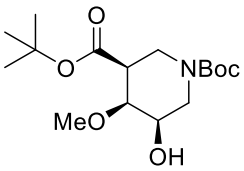
Method B (deprotection of ketone **44**): Ketone **44** (23 mg, 64 µmol) was dissolved in a mixture of hexafluoro-2-propanol (960 µL) and water (200 µL). Concentrated HCl (40 µL) was added and the mixture stirred at rt for 5 h after which the mixture was concentrated *in vacuo* and co-evaporated with water at 35 ºC (3 x) to afford the target compound (14 mg) in quantitative yield. ^1H NMR (500 MHz, D_2O at 293K) [hydrate form] δ 4.20 – 4.15 (m, 1H, H4), 3.44 (dd, J = 12.1, 3.8 Hz, 1H, H6b), 3.32 (ddd, J = 12.7, 3.8, 2.5 Hz, 1H, H5), 3.26 (d, J = 12.4 Hz, 1H, H6a), 3.21 (d, J = 12.7 Hz, 1H, H2b), 3.16 (d, J = 12.7 Hz, 1H, H2a). ^{13}C NMR (126 MHz, D_2O) δ 173.3 (C=O), 91.1 (C3), 69.4 (C4), 46.1 (C2), 41.9 (C5), 38.4 (C6). HRMS (ESI) m/z: $[M_{hydrate}+H]^+$ calc for $C_6H_{12}NO_5$ 178.07100, found 178.07068.

Compound 45



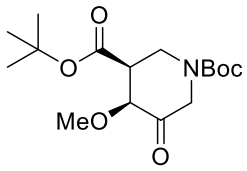
To a solution of alcohol **42** (0.35 g, 0.81 mmol) in THF (8 mL) were added Ag_2O (2.44 g, 10.5 mmol), CH_3I (3.0 mL, 48.2 mmol) and $(CH_3)_2S$ (100 µL, 1.36 mmol). The flask was protected from light by packing in alumina foil and the mixture was stirred vigorously at rt overnight. The mixture was filtered over a pad of celite and the filter cake washed with EtOAc (3 x). The filtrate was concentrated *in vacuo* to give a thick oil (377 mg) that was purified by silica gel column chromatography (pentane/EtOAc 98:2→95:5) to afford the target product (0.25 g, 0.56 mmol) in 69% yield. ^1H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.01 (d, J = 10.3 Hz, 1H, H6b), 3.86 – 3.74 (m, 2H, H4, H2b), 3.56 – 3.47 (m, 4H, H3, OCH_3), 2.99 – 2.84 (m, 2H, H6a and H2a), 2.35 (ddd, J = 11.9, 4.6, 2.3 Hz, 1H, H5), 1.42 (s, 9H, $(CH_3)_3CO$), 1.40 (s, 9H, $(CH_3)_3CO$), 0.88 (s, 9H, $(CH_3)_3CSi$), 0.07 (s, 6H, $2CH_3Si$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 169.9 (OC=O), 154.8 (NC=O), 80.9 ($(CH_3)_3CO$), 79.8 (C4), 79.8 ($(CH_3)_3CO$), 71.6 (C3), 61.5 (OCH_3), 46.7 (C5), 44.8 (C2), 39.6 (C6), 28.5 ($(CH_3)_3CO$), 28.2 ($(CH_3)_3CO$), 25.9 ($(CH_3)_3CSi$), 18.1 ($(CH_3)_3CSi$), -4.6 (CH_3Si), -4.8 (CH_3Si). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{22}H_{43}NO_6SiNa$ 468.27519, found 468.27500.

Compound S10



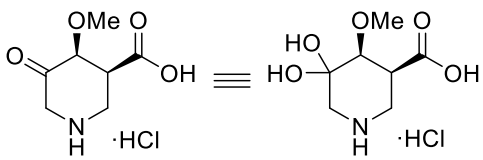
To an ice cold solution of TBS-ether **45** (0.25 g, 0.56 mmol) in THF (10 mL) was added TBAF (75 wt% in H_2O, 410 mg, 1.17 mmol). The mixture was allowed to warm up slowly to room temperature and stirred for 5 h until TLC analysis indicated complete conversion of the starting material. The mixture was diluted with EtOAc (30 mL) and washed with brine (2 x 10 mL), dried over Na_2SO_4, filtered and concentrated. The crude was purified by silica gel column chromatography (pentane/EtOAc 95:5→2:1) to give the target alcohol (178 mg, 0.537 mmol, 96%) as a colorless solid. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 3.95 – 3.87 (m, 2H, H6b and H4), 3.82 (d, J = 9.9 Hz, 1H, H2b), 3.61 (s, 1H, H3), 3.53 (s, 3H, OCH_3), 3.19 (dd, J = 13.7, 10.7 Hz, 1H, H6a), 3.03 – 2.95 (m, 1H, H2a), 2.71 (s, 1H, OH), 2.52 – 2.45 (m, 1H, H5), 1.48 (s, 9H, $(CH_3)_3CO$), 1.45 (s, 9H, $(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 170.5 (OC=O), 154.9 (NC=O), 81.6 ($(CH_3)_3\mathit{C}O$), 80.1 ($(CH_3)_3\mathit{C}O$), 79.2 (C4), 68.7 (C3), 60.6 (OCH_3), 46.5 (C5), 45.7 (C2), 40.1 (C6), 28.5 ($(\mathit{C}H_3)_3CO$), 28.2 ($(\mathit{C}H_3)_3CO$). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{16}H_{29}NO_6Na$ 354.18871, found 354.18850.

Compound 46



To a stirred solution of alcohol **S10** (176 mg, 0.53 mmol) in dry DCM (10 mL) was added Dess-Martin periodinane (488 mg, 1.15 mmol) and the mixture was stirred at rt for 4 h until complete conversion was confirmed by TLC analysis. The mixture was diluted with EtOAc (30 mL) and washed subsequently with aq. 2 M $Na_2S_2O_3$ (2 x 10 mL), sat. aq. $NaHCO_3$ (2 x 10 mL) and brine (10 mL), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/acetone 9:1→7:3) to afford ketone **46** (135 mg, 0.41 mmol, 76%) as a colorless oil. ^{1}H NMR (500 MHz, $CDCl_3$ at 333K) δ 4.10 (d, J = 17.1 Hz, 1H, H2b), 4.02 (d, J = 17.1 Hz, 1H, H2a), 3.87 (d, J = 5.9 Hz, 2H, H6), 3.80 (d, J = 4.6 Hz, 1H, H4), 3.45 (s, 3H, OCH_3), 3.04 (q, J = 5.9 Hz, 1H, H5), 1.45 (s, 9H, $(CH_3)_3CO$), 1.44 (s, 9H, $(CH_3)_3CO$). ^{13}C NMR (126 MHz, $CDCl_3$ at 333K) δ 200.7 (C=O), 168.9 (OC=O), 154.3 (NC=O), 82.2 ($(CH_3)_3\mathit{C}O$), 81.8 (C4), 81.0 ($(CH_3)_3\mathit{C}O$), 58.7 (OCH_3), 52.9 (C2), 47.2 (C5), 41.9 (C6), 28.5 ($(\mathit{C}H_3)_3CO$), 28.1 ($(\mathit{C}H_3)_3CO$). HRMS (ESI) m/z: $[M+Na]^+$ calc for $C_{16}H_{27}NO_6Na$ 352.17306, found 352.17283.

Compound 10



Ketone **46** (47 mg, 0.14 mmol) was dissolved in a mixture of hexafluoro-2-propanol (1.5 mL) and water (0.4 mL). After cooling on an ice bath, concentrated hydrochloric acid (0.1 mL) was added. After five minutes the cooling bath was removed and the mixture stirred at room temperature for 4 h. The solvents were removed *in vacuo* and the residue co-evaporated with water (2 x) to afford the target product (33 mg) as a foam in quantitative

yield. ^{1}H NMR (400 MHz, D_2O at 293K) [hydrate form] δ 3.82 (d, J = 2.4 Hz, 1H, H4), 3.48 (s, 3H, OCH_3), 3.38 (dd, J = 12.6, 4.3 Hz, 1H, H6b), 3.27 (ddd, J = 12.6, 4.3, 2.8 Hz, 1H, H5), 3.19 – 3.05 (m, 3H, H6a and H2ab). ^{13}C NMR (101 MHz, D_2O at 293K) [hydrate form] δ 173.3 (COOH), 91.5 (C3), 79.3 (C4), 61.0 (OCH_3), 46.5 (C2), 41.4 (C5), 38.8 (C6). HRMS (ESI) m/z: $[M_{ketone}+H]^+$ calc for $C_7H_{12}NO_4$ 174.07608, found 174.07624; $[M_{hydrate}+H]^+$ calc for $C_7H_{14}NO_5$ 192.08665, found 192.08675; $[M_{hydrate}+Na]^+$ calc for $C_7H_{13}NO_5Na$ 214.06859, found 214.06866.

5.6 References

[1] Li, J. P.; Kusche-Gullberg, M., *Academic Press* **2016**, *325*, 215-273.

[2] Sarrazin, S.; Lamanna, W. C.; Esko, J. D., Cold Spring Harb Perspect Biol. 2011, 3, 1-33.

[3] Iozzo, R. V.; Schaefer, L., Matrix Biol. 2015, 42, 11-55.

[4] Iozzo, R. V., *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 646-656.

[5] Masola, V.; Bellin, G.; Gambaro, G.; Onisto, M., *Cells* **2018**, *7*, 236-246.

[6] Billings, P. C.; Pacifici, M., *Connect Tissue Res.* **2015**, *56*, 272-280.

[7] Matsuo, I.; Kimura-Yoshida, C., *Phil. Trans. R. Soc. B* **2014**, *369*, 1-9.

[8] Koshland, D.E., *Biol. Rev. Camb. Philos. Soc.* **1953**, *28*, 416–436.

[9] Wilson, J. C.; Laloo, A. E.; Singh, S.; Ferro, V., *Biochem. Biophys. Res. Commun.* **2014**, *443*, 185-188.

[10] Goldberg, R.; Meirovitz, A.; Hirshoren, N.; Bulvik, R.; Binder, A.; Rubinstein, A. M.; Elkin, M., *Matrix Biol.* **2013**, *32*, 234-240.

[11] Sanderson, R. D.; Elkin, M.; Rapraeger, A. C.; Ilan, N.; Vlodavsky, I., *FEBS J.* **2017**, *284*, 42-55.

[12] Vlodavsky, I.; Singh, P.; Boyango, I.; Gutter-Kapon, L.; Elkin, M.; Sanderson, R. D.; Ilan, N., *Drug Resist. Updat.* **2016**, *29*, 54-75.

[13] Rivara, S.; Milazzo, F. M.; Giannini, G., *Future Med. Chem.* **2016**, *8*, 647-680,

[14] Jia, L.; Ma, S., *Eur. J. Med. Chem.* **2016**, *121*, 209-220.

[15] Mohan, C. D.; Hari, S.; Preetham, H. D.; Rangappa, S.; Barash, U.; Ilan, N.; Nayak, S. C.; Gupta, V. K.; Basappa; Vlodavsky, I.; Rangappa, K. S., *iScience* **2019**, *15*, 360-390.

[16] Vlodavsky, I.; Sanderson, R. D.; Ilan, N., *Springer Cham.* **2020**, 567-603.

[17] Karoli, T.; Liu, L.; Fairweather, J. K.; Hammond, E.; Li, C. P.; Cochran, S.; Bergefall, K.; Trybala, E.; Addison, R. S.; Ferro, V., J. Med. Chem. 2005, 48, 8229-8236.

[18] Ritchie, J. P.; Ramani, V. C.; Ren, Y.; Naggi, A.; Torri, G.; Casu, B.; Penco, S.; Pisano, C.; Carminati, P.; Tortoreto, M.; Zunino, F.; Vlodavsky, I.; Sanderson, R. D.; Yang, Y., *Clin. Cancer Res.* **2011**, *17*, 1382-1393.

[19] O'Reilly, E. M.; Roach, J.; Miller, P.; Yu, K. H.; Tjan, C.; Rosano, M.; Krause, S.; Avery, W.; Wolf, J.; Flaherty, K.; Nix, D.; Ryan, D. P., *Oncologist* **2017**, *22*, 1429-e139.

[20] Dredge, K.; Brennan, T. V.; Hammond, E.; Lickliter, J. D.; Lin, L.; Bampton, D.; Handley, P.; Lankesheer, F.; Morrish, G.; Yang, Y.; Brown, M. P.; Millward, M., *Br. J. Cancer* **2018**, *118*, 1035-1041.

[21] Madia, V. N.; Messore, A.; Pescatori, L.; Saccoliti, F.; Tudino, V.; De Leo, A.; Bortolami, M.; Scipione, L.; Costi, R.; Rivara, S.; Scalvini, L.; Mor, M.; Ferrara, F. F.; Pavoni, E.; Roscilli, G.; Cassinelli, G.; Milazzo, F. M.; Battistuzzi, G.; Di Santo, R.; Giannini, G., *J. Med. Chem.* **2018**, *61*, 6918-6936.

[22] Watson, A. A.; Fleet, G. W. J.; Asano, N.; Molyneux, R. J.; Nash, R. J., *Phytochemistry* **2001**, *56*, 265–295.

[23] Gloster, T. M.; Davies, G. J., *Org. Biomol. Chem.* **2010**, *8*, 305-320.

[24] Conforti, I.; Marra, A., *Org. Biomol. Chem.* **2021**, *19*, 5439-5475.

[25] Platt, F. M.; Neises, G. R.; Dwek, R. A.; Butters, T. D., *J. Bio. Chem.* **1994**, *269*, 8362-8365.

[26] Gross,P. E.; Baker, M. A.; Carver, J. P.; Dennis, J. W., *Clin. Cancer Res*. **1995**, *1*, 935-944.

[27] Scott, L. J.; Spencer, C. M., *Drugs* **2000**, *59*, 521-549.

[28] Umezawa, H.; Aoyagi, T.; Komiyama, H.; Morishima, H.; Hamada, M.; Takeuchi, T., *J. Antibiot.* **1974**, *27*, 963-969.

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

[30] Nishimura, Y.; Wang, W.; Kondo, S.; Aoyagi, T.; Umezawa, H., *J. Am. Chem. Soc.* **1988**, *110*, 7249-7250.

[31] Nishimura, Y.; Kudo, T.; Kondo, S.; Takeuchi, T., *J. Antibiot.* **1994**, *47*, 101-107.

[32] Nishimura, Y.; Shitara, E.; Adachi, H.; Toyoshima, M.; Nakajima, M.; Okami, Y.; Takeuchi, T., *J. Org. Chem.* **2000**, *65*, 2-11.

[33] Kondo, K., Adachi, H., Shitara, E., Kojima, F.; Nishimura, Y., *Bioorg. Med. Chem.* **2001**, *9*, 1091–1095.

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

[35] Caner, S.; Zhang, X.; Jiang, J.; Chen, H. M.; Nguyen, N. T.; Overkleeft, H.; Brayer, G. D.; Withers, S. G., *FEBS Lett.* **2016**, *590*, 1143-1151.

[36] Schroder, S. P.; de Boer, C.; McGregor, N. G. S.; Rowland, R. J.; Moroz, O.; Blagova, E.; Reijngoud, J.; Arentshorst, M.; Osborn, D.; Morant, M. D.; Abbate, E.; Stringer, M. A.; Krogh, K.; Raich, L.; Rovira, C.; Berrin, J. G.; van Wezel, G. P.; Ram, A. F. J.; Florea, B. I.; van der Marel, G. A.; Codee, J. D. C.; Wilson, K. S.; Wu, L.; Davies, G. J.; Overkleeft, H. S., *ACS Cent. Sci.* **2019**, *5*, 1067-1078.

[37] de Boer, C.; McGregor, N. G. S.; Peterse, E.; Schroder, S. P.; Florea, B. I.; Jiang, J.; Reijngoud, J.; Ram, A. F. J.; van Wezel, G. P.; van der Marel, G. A.; Codee, J. D. C.; Overkleeft, H. S.; Davies, G. J., *RSC Chem. Biol.* **2020**, *1*, 148-155.

[38] Igarashi, Y.; Ichikawa, M.; Ichikawa, Y., *Tetrahedron Lett*. **1996**, *37*, 2707-2708.

[39] Ichikawa, Y.; Igarashi, Y.; Ichikawa, M.; Suhara, Y., *J. Am. Chem. Soc.* **1998**, *120*, 3007-3018.

[40] Jip Kim, Y.; Ichikawa, M.; Ichikawa, Y., *J. Org. Chem.* **2000**, *65*, 2599-2602.

[41] Jiang, J.; Kallemeijn, W. W.; Wright, D. W.; van den Nieuwendijk, A.; 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.; Overkleeft, H. S., *Chem. Sci.* **2015**, *6*, 2782-2789.

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

[43] Ouchi, H.; Mihara, Y.; Takahata, H., *J. Org. Chem.* **2005**, *70*, 5207-5214.

[44] Yang, D.; Wong, M.-K.; Yip, Y.-C., *J. Org. Chem.* **1995**, *60*, 3887-3889.

[45] Lipshutz, B. H.; Wilhelm, R. S.; Kozlowski, J. A., *Tetrahedron* **1984**, *40*, 5005-5038.

[46] Joly, G. D.; Jacobsen, E. N., *Org. Lett.,* **2002**, *4*, 1795-1798.

[47] Custar, D. W.; Zabawa, T. P.; Hines, J.; Crews, C. M.; Scheidt, K. A., *J. Am. Chem. Soc.* **2009**, *131*, 12406-12414.

[48] Kato, T.; Sato, T.; Kashiwagi, Y.; Hosokawa, S., *Org. Lett.* **2015**, 17, 2274-2277.

[49] Carlson, P. H. J.; Katsuki, T.; Martin, V. S.; Sharpless, K. B., *J. Org. Chem.* **1981**, *46*, 3936-3938.

[50] Ichikawa, Y.; Igarashi, Y., *Tetrahedron Lett.* **1995**, *36*, 4585-4586.

[51] Bal, B. S.; Childers, W. E.; Pinnick, H. W., *Tetrahedron* **1981**, *37*, 2091-2096.

[52] Daragics, K.; Fügedi, P., *Tetrahedron Lett.* **2009**, *50*, 2914-2916.

[53] Werz, D. B.; Seeberger, P. H., *Angew. Chem. Int. Ed. Engl.* **2005**, *44* (39), 6315-6318.

[54] Wu, L.; Jiang, J.; Jin, Y.; Kallemeijn, W. W.; Kuo, C.-L.; Artola, M.; Dai, W.; van Elk, C.; van Eijk, M.; van der Marel, G. A.; Codée, J. D. C.; Florea, B. I.; Aerts, J. M. F. G.; Overkleeft, H. S.; Davies, G. J., *Nat. Chem. Biol.* **2017**, *13*, 867-873.

[55] Winter, G., *J. Appl. Crystallogr.* **2010**, *43*, 186-190.

[56] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta. Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486-501.

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

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

APPENDIX

Table 5.S1. Data collection and refinement statistics

	HPSE + SiasB	HPSE + **14**	HPSE + **12**	AcGH79 + SiasB
Data collection				
Space group	$P2_1$	$P2_1$	$P2_1$	$I2_1$
Cell dimensions				
a, *b*, *c* (Å)	46.11, 70.92, 78.56	44.51, 71.57, 78.36	44.45, 71.16, 78.29	82.75 44.67 136.27
α, β, γ (°)	90.0, 95.5, 90.0	90.0 98.34 90.0	90.0, 98.48, 90.0	90.0, 97.46, 90.0
Resolution (Å)	45.90-1.70 (1.73-1.70)	52.64-1.90 (1.94-1.90)	44.01-1.80 (1.84-1.80)	67.56-1.05 (1.07-1.05)
$I / \sigma I$	10.8 (1.6)	10.6 (1.2)	9.2 (1.3)	15.1 (1.4)
Completeness (%)	99.6 (99.9)	100 (100)	99.2 (98.3)	96.5 (88.8)
Redundancy	1.9 (1.9)	2.0 (2.0)	6.9 (6.5)	6.3 (4.5)
Refinement				
Resolution (Å)	1.70	1.90	1.80	1.05
No. reflections	104,973	75,449	306,166	1,385,607
R_{work} / R_{free}	0.17/0.21	0.19/0.23	0.20/0.23	0.17/0.18
No. atoms				
Protein	3,674	3,654	3,643	3,613
Ligand/ion	104/3	51/1	34/1	17/0
Water	249	100	183	585
B-factors				
Protein	29.83	37.63	30.97	13.12
Ligand/ion	61.33/ 35.52	53.34/ 35.99	42.17/35.28	14.48
Water	35.04	36.17	35.92	24.68
R.m.s. deviations				
Bond lengths (Å)	0.01	0.01	0.01	0.02
Bond angles (°)	1.7	1.6	1.6	2.0

*Number of xtals for each structure should be noted in footnote. *Values in parentheses are for highest-resolution shell.

Table S1, continued

	AcGH79 + **14**	AcGH79 + **12**	BpHep + SiasB	EcGusB + SiasB
Data collection				
Space group	$I2_1$	$I2_1$	$P2_12_12_1$	C2
Cell dimensions				
a, *b*, *c* (Å)	83.05, 44.67, 137.17	82.04, 42.61, 140.14	76.37, 104.61, 113.71	126.28, 76.71, 141.02
α, β, γ (°)	90.0, 97.6, 90.0	90.0, 99.42, 90.0	90.0, 90.0, 90.0	90.0, 102.01, 90.0
Resolution (Å)	42.44-1.25 (1.27-1.25)	40.72-1.50 (1.53-1.50)	61.68-1.27 (1.29-1.27)	67.04-1.95 (1.98-1.95)
$I / \sigma I$	10.2 (1.6)	11.0 (1.0)	12.6 (1.1)	8.1 (1.2)
Completeness (%)	99.0 (87.9)	100.0 (100.0)	99.9 (98.5)	98.4 (97.4)
Redundancy	6.2 (3.8)	6.4 (6.3)	8.2 (7.4)	4.1 (4.2)
Refinement				
Resolution (Å)	1.25	1.50	1.27	1.95
No. reflections	840,749	489,904	1,960,048	384,016
R_{work} / R_{free}	0.16/0.18	0.20/0.22	0.16/0.18	0.22/0.28
No. atoms				
Protein	3,487	3,435	6,441	9,574
Ligand/ion	11/0	17/0	56/0	24/0
Water	357	185	1037	316
B-factors				
Protein	15.07	24.02	18.44	54.8
Ligand/ion	10.47	22.25	25.01	42.05
Water	22.38	26.77	33.14	43.74
R.m.s. deviations				
Bond lengths (Å)	0.01	0.01	0.02	0.01
Bond angles (°)	1.9	1.8	1.9	1.6

*Number of xtals for each structure should be noted in footnote. *Values in parentheses are for highest-resolution shell.

6

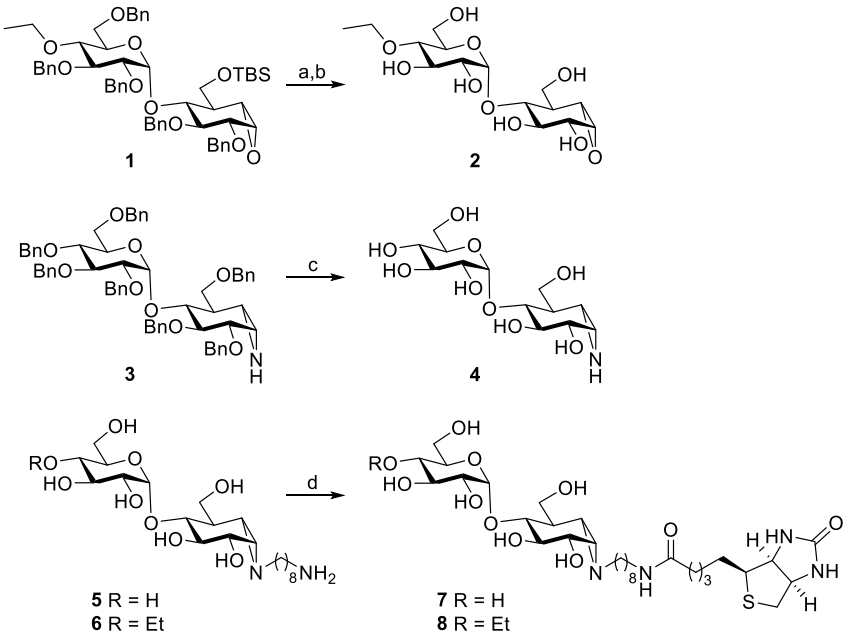
Summary and Future Prospects

The research described in this dissertation focus on the development and biochemical evaluation of covalent inhibitors and activity-based probes (ABPs) for retaining *endo*- and *exo*-glycosidases including starch-degrading enzymes and human lysosomal β-glucocerebrosidase (GBA), as well as the synthesis of a panel of uronic acid-type 1-*N*-iminosugars as potential competitive heparanase and β-glucuronidase inhibitors. The design of such covalent and competitive glycosidase inhibitors relies on the understanding of enzyme's function and mechanistic aspects. **Chapter 1** introduces the general mechanisms employed by retaining and inverting glycosidases during the hydrolysis of their substrates. The reaction itineraries of retaining α- and β-glucosidases are described and the design of covalent cyclophellitol-based inhibitors based on these itineraries discussed. Additionally, the design of competitive inhibitors based on transition state mimicry is described. Finally, an overview of the activity-based protein profiling (ABPP) workflow is presented.

The biodegradation of starch requires the synergistic action of a set of enzymes.[1] Amongst these, retaining α-amylases, which catalyze the hydrolysis of internal α-1,4-glycosidic linkages, are most extensively studied both in biomedicine[2] and biotechnology.[3] In **Chapter 2**, the synthesis of a set of *maltobiose*-configured *epi*-cyclophellitol derivatives as retaining α-amylase inhibitors and ABPs is presented. The key step in the synthesis route involves stereoselective α-1,4-glycosylation under proper pre-activation conditions, using acid-tolerant cyclohexene or *epi*-cyclophellitol cyclocarbonate acceptors. Following a series of chemical transformations the desired epoxide, aziridine and cyclosulfate warheads were obtained. Activities and efficiencies of the synthesized inhibitors and probes were substantiated by extensive biochemical analysis. Fluorescent labeling of recombinant human saliva α-amylase showed that the aziridine-based probe is less potent than the epoxide-based probe. This observation is in accordance with the kinetic studies on Taka-amylase where the aziridine inhibitor has the slowest rate of inhibition. X-ray crystallographic analysis of inhibitor-enzyme complexes showed that both epoxide and cyclosulfate inhibitors bind to the active site of Taka-amylase in a $^4C_1$ chair conformation, whilst the aziridine inhibitor adopts an unprecedented E_3 conformation, which might account for its lower potency. The epoxide probe was shown to be able to effectively label α-amylases in human saliva, murine tissue and fungal secretomes in a concentration-, pH-, time-, and temperature-dependent fashion, and the labeling could be competed with the non-tagged inhibitors, as well as the commercially available competitive inhibitor, acarbose. In addition, pull-down analysis with the biotinylated epoxide probe identified α-amylase in human saliva and fungal secretomes.

Starch polysaccharides are composed of linear amylose and branched amylopectin, the latter of which constitutes a major component of common starch.[4] In amylopectin, the branch points containing α-1,6-glucosidic linkages are resistant to the hydrolysis of α-1,4-specific amylases. As an extension of the maltobiose *epi*-cyclophellitol epoxide probes developed in Chapter 2, **Chapter 3** describes the synthesis of a panel of glucose-isomaltose (GIM) and isomaltose-glucose (IMG) configured *epi*-cyclophellitol probes, which mimic the branched parts of the amylopectin structure. Construction of the pseudotrisaccharide backbones was accomplished by stereoselective α-1,4- and α-1,6-glycosylations of cyclohexene acceptors under proper pre-activation conditions. After this, the epoxide functionality was introduced via a stereoselective epoxidation sequence, which was followed by global deprotection and amide-coupling with a reporter tag to afford the final probes after HPLC purification. The set of branched probes will be useful for the detection of industrially relevant starch-degrading enzymes that show preference for branched amylopectin-type polysaccharides.

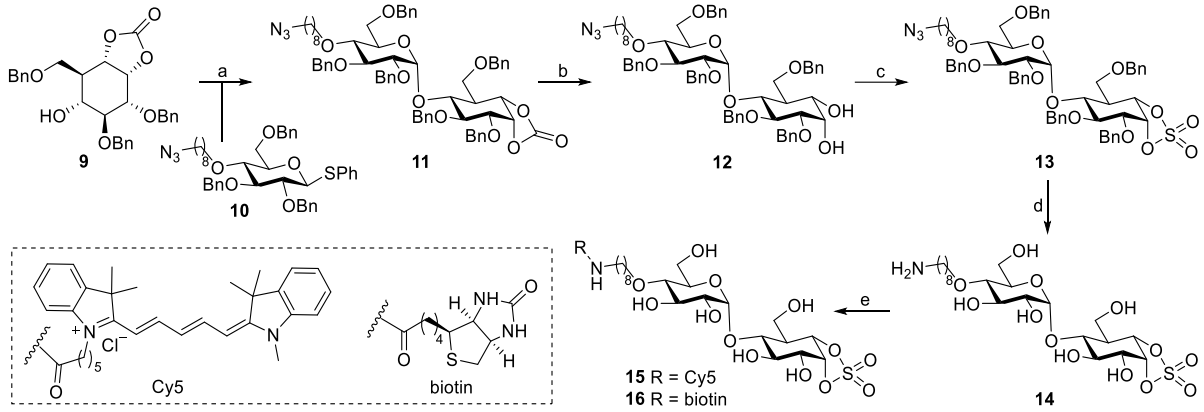
Besides the compounds presented in Chapter 2, several other pseudodisaccharide inhibitors and biotin probes were synthesized following similar strategies (Scheme 6.1). Desilylation of intermediate **1** (Chapter 2) followed by global debenzylation using Pearlman's catalyst under hydrogen atmosphere afforded O4' 'capped' inhibitor **2** in good yield. Preparation of aziridine inhibitor **4** was accomplished by deprotection of compound **3** (Chapter 2) under Birch debenzylation conditions. Amide coupling of the primary amine in **5** or **6** (Chapter 2) with biotin succinate ester (biotin-OSu) followed by HPLC purification provided maltobiose *epi*-cyclophellitol aziridine probes **7** and **8**.



Scheme 6.1. Synthesis of inhibitors **2** and **4**, and biotin probes **7** and **8**. Reagents and conditions: a) TBAF, THF, rt, 90%; b) $Pd(OH)_2/C$, H_2, $MeOH/H_2O$/dioxane (2/1/2), rt, 89%; c) Na, NH_3 (liq.), ^tBuOH, THF, -60 °C, 96%; d) biotin-OSu, DIPEA, DMF, rt, **7** 20%, **8** 44%.

Kinetic studies for the inhibition of Taka-amylase by the *maltobiose*-configured *epi*-cyclophellitols revealed that the cyclosulfate containing compound is 4-5 times more potent than its epoxide or aziridine counterparts while elongation at the nonreducing end O4' position of the maltobiose *epi*-cyclophellitol epoxide with an alkyl chain can increase its rate of inhibition nearly 10-fold (Chapter 2). Thus, it was envisioned that elongation at O4' of the maltobiose *epi*-cyclophellitol cyclosulfate inhibitor would yield more potent species. For this purpose, the corresponding maltobiose cyclosulfate probes bearing reporter entities at O4' position of the nonreducing end pyranose were synthesized (Scheme 6.2). The synthesis commenced with carbonate acceptor **9** (Chapter 2), which was glycosylated with O4-alkylated thiophenyl donor **10** (Chapter 2) under NIS/TfOH/DMF conditions, giving pseudodisaccharide **11** in good yield and stereoselectivity. The cyclic carbonate was then hydrolyzed under basic conditions affording diol **12**. Treatment of **12** with thionyl chloride resulted in the generation of a mixture of cyclosulfites, which was further oxidized to give cyclosulfate **13** in 88% yield over two steps. In the first instance, global debenzylation and concomitant azide reduction of **13** was attempted under Birch conditions which have been successfully applied for the deprotection of diverse cyclophellitol and cyclophellitol aziridine derivatives. However, in this case a complex mixture of products was obtained as indicated by NMR analysis (Table 6.1, entry 1). As an alternative, $Pd(OH)_2/C$ catalyzed hydrogenolysis was investigated. Treatment of **13** with an excess of Pearlman's catalyst in the presence of HOAc (10 eq.) for 16 h led to incomplete deprotection (entry 2 and 3). The azido group could be smoothly reduced to amine under these conditions while deprotection of the benzyl groups proved to be sluggish. Prolonged reaction time indeed led to the formation of desired product **14**, however an obvious side product of which the molecular weight is 94 Dalton less than that of **14** was detected by LC-MS analysis, indicating that the cyclosulfate ring may be hydrogenated as well to form a cyclohexane (entry 4). The slow rate of hydrogenolysis is attributed to catalyst poisoning by the emerging amine that has not been captured by the weak acetic acid (HOAc). This problem was circumvented by replacing HOAc with HCl to effectively capture the amine prior to catalyst poisoning. In this way the deprotection proceeded smoothly. Complete debenzylation was achieved after 5 h and the cyclosulfate functionality remained unaffected during the short reaction time, giving product **14** in good yield (entry 5). Amide coupling of the primary amine with the corresponding reporter tags gave ABPs **15** and **16** after HPLC purification eluting with

50 mM AcOH (note: HPLC purification by 50 mM NH_4HCO_3 resulted in nucleophilic opening of the cyclosulfate. See the experimental section for proposed structures and characterization of the rearranged products). Unfortunately, the fluorescent probe **15** was found unstable and decomposed during storage at -20 °C for months as revealed by LC-MS and NMR analyses, and thus could not be used for further ABPP experiments. In contrast, the biotin probe and 'untagged' inhibitors remain unaffected even after storing at -20 °C for more than one year.



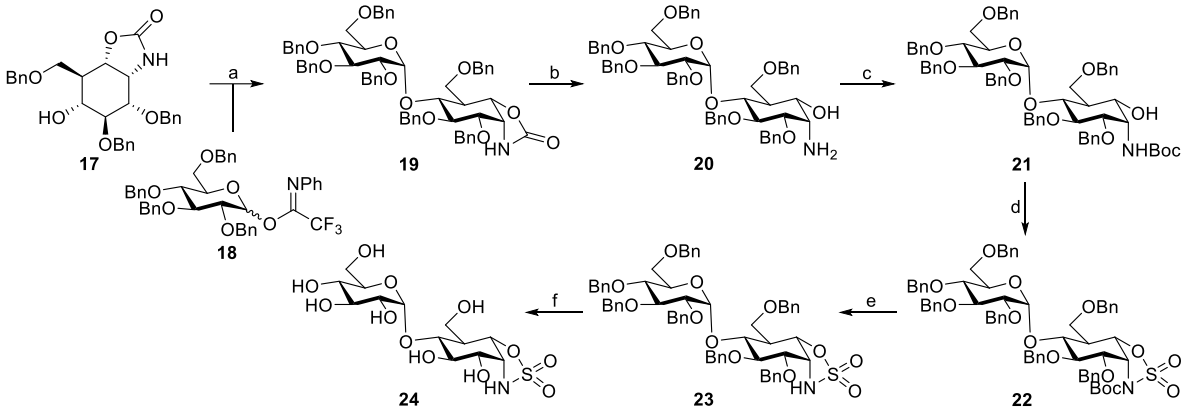
Scheme 6.2. Synthesis of maltobiose *epi*-cyclophellitol cyclosulfate probes. Reagents and conditions: a) NIS, TfOH, DMF, DCM, 4 Å MS, -20 °C to 0 °C, 90%; b) NaOMe, DCM/MeOH (1/1), rt, 94%; c) *i*) $SOCl_2$, TEA, DCM, 0 °C; *ii*) $RuCl_3.3H_2O$, $NaIO_4$, EtOAc/ACN/H_2O (2/2/1), 0 °C, 88%; d) $Pd(OH)_2/C$, H_2, HOAc, ^{t}BuOH/H_2O/dioxane (1/2/1), rt, 88%; e) for **15**: Cy5COOH, DIC, pentafluorophenol, Et_3N, DMF, rt, 11%; for **16**: biotin-OSu, DIPEA, DMF, rt, 17%.

Table 6.1. Azide reduction and global debenzylation of compound **13**.

Entry	Reagent	Solvent/Temperature	Time	Result
1	Na, ^{t}BuOH	NH_3 (liq.), THF, -60 °C	1 h	messy
2	$Pd(OH)_2/C$ (1.5 eq.), H_2 HOAc (10 eq.)	^{t}BuOH/H_2O/dioxane, rt (2/1/2, v/v/v)	16 h	azide→amine no benzyl deprotection
3	$Pd(OH)_2/C$ (1.5 eq.), H_2 HOAc (10 eq.)	^{t}BuOH/H_2O/dioxane, rt (1/2/1, v/v/v)	16 h	azide→amine partial debenzylatiom
4	$Pd(OH)_2/C$ (1.5 eq.), H_2 HOAc (10 eq.)	^{t}BuOH/H_2O/dioxane, rt (1/2/1, v/v/v)	2 days	product **14** (minor) cyclohexane side product (major)
5	$Pd(OH)_2/C$ (1.0 eq.), H_2 HCl (6 eq.)	^{t}BuOH/H_2O/dioxane, rt (1/2/1, v/v/v)	5 h	product **14** (88%)

Recently, α-D-Gal-cyclophellitol cyclosulfamidate has been developed as a Michaelis complex analog that stabilizes recombinant α-galactosidase A (α-GalA) both *in vitro* and *in*

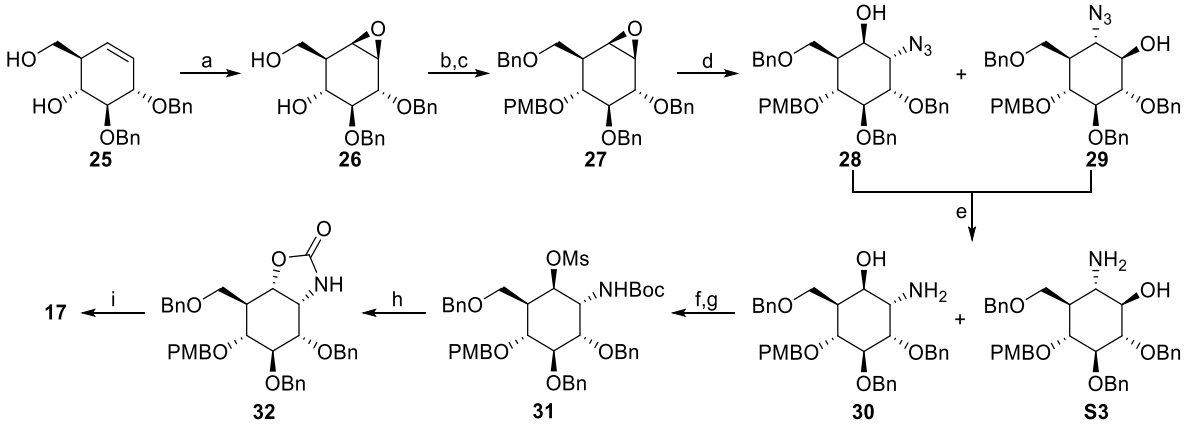
cellulo.[5] The design of this inhibitor is based on the replacement of the 'anomeric' oxygen of the parent cyclosulfate with a nitrogen atom. This replacement decreases its leaving group capacity, making the cyclosulfamidate a potent reversible inhibitor of α-GalA. It would be of interest to transfer the structural characteristic to *maltobiose*-configured *epi*-cyclophellitol cyclosulfate, which may provide a selective competitive inhibitor of retaining α-amylases. For this purpose, the maltobiose *epi*-cyclophellitol cyclosulfamidate was synthesized. The synthesis commenced with carbamate acceptor **17**, which was glycosylated with imidate donor **18** (Chapter 2) under TfOH/DMF pre-activation conditions, giving pseudodisaccharide **19**. Hydrolysis of the cyclic carbamate was accomplished by treatment with aqueous 1 M NaOH at elevated temperature to afford amino-alcohol **20**, of which the primary amine was protected with a *tert*-butyloxycarbonyl (Boc) group to provide **21**. Treatment of **21** with thionyl chloride followed by oxidation afforded fully protected cyclosulfamidate **22**. Removal of the Boc group with trifluoroacetic acid and subsequent global debenzylation using Pearlman's catalyst under hydrogen atmosphere provided final compound **24**.



Scheme 6.3. Synthesis of reversible maltobiose *epi*-cyclophellitol cyclosulfamidate inhibitor **24**. Reagents and conditions: a) TfOH, DMF, DCM, 4 Å MS, -20 °C to rt, 70%; b) 1 M NaOH, EtOH/THF (5/1), 80 °C, 95%; c) Boc_2O, Et_3N, DCM, rt, 95%; d) *i*) $SOCl_2$, TEA, imidazole, DCM, 0 °C; *ii*) $RuCl_3.3H_2O$, $NaIO_4$, EtOAc/ACN/H_2O (2/2/1), 0 °C, 90%; e) TFA, DCM, rt, 86%; f) $Pd(OH)_2/C$, H_2, MeOH/H_2O/dioxane (2/1/2), rt, quant.

Acceptor **17** was synthesized from published diol **25**[6] following adapted procedures from a paper previously reported by Artola *et al.* (Scheme 6.4).[5] Epoxidation of **25** with *meta*-chloroperoxybenzoic acid (*m*-CPBA) provided partially protected cyclophellitol **26** in 72% yield. The primary alcohol was selectively benzylated using 2-aminoethyl diphenylborinate as catalyst, followed by PMB protection of the remaining secondary alcohol to afford **27**. Nucleophilic opening of the epoxide with sodium azide in the presence of lithium perchlorate

at elevated temperature afforded an inseparable mixture of azido-alcohols **28** and **29**, which were directly reduced via platinum catalyzed hydrogenolysis resulting in the isolation of **30** in 46% yield. *N*-bocylation and subsequent mesylation of the secondary alcohol afforded **31**, which was heated to 120 °C in DMF to provide the cyclic carbamate **32** via intramolecular substitution of the mesylate group. The PMB ether was then removed by 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) to give acceptor **17** in good yield.



Scheme 6.4. Synthesis of acceptor **17**. Reagents and conditions: a) *m*-CPBA, DCM, 0 °C, 72%; b) BnBr, KI, K_2CO_3, 2-aminoethyl diphenylborinate, MeCN, 60 °C, 86%; c) PMBCl, NaH, TBAI, DMF, 0 °C to rt, 75%; d) NaN_3, $LiClO_4$, DMF, 80 °C; e) PtO_2, H_2, THF, rt, **30** 46%, **S3** 44% over 2 steps; f) Et_3N, Boc_2O, DCM, rt, 93%; g) MsCl, *N*-methyl imidazole, Et_3N, $CHCl_3$, rt; h) DMF, 120 °C, 71% over 2 steps; i) DDQ, DCM/H_2O (19/1), rt, 85%.

Next, the inhibitory potency of cyclosufamidate **24** and parent cyclosulfate **33** (Chapter 2) toward α-amylases in human saliva were investigated by competitive ABPP. Pre-incubation of human saliva samples with competitors for 30 min at 37 °C prior to labeling with 5 µM ABP **34** (Chapter 2) for 1 h resulted in a concentration-dependent competition of fluorescent labeling in the range of 0 to 500 µM (Figure 6.1A). While pre-incubation with **24** did not fully abrogate labeling, pre-incubation with **33** was able to fully compete the signal at concentrations higher than 10 µM. In order to evaluate whether the inhibition by **24** is reversible, human saliva samples were pre-incubated with inhibitors for different time periods (15, 30, 60, 120, 180 and 240 min), followed by labeling with probe **34** for 1 h (Figure 6.1B). Pre-incubation with covalent inhibitor **33** resulted in a decrease of fluorescent signal with longer inhibition time, whereas an almost constant intensity of the signal was observed by pre-incubation with compound **24** for extended inhibition times, indicating that cyclosulfamidate **24** may be a reversible inhibitor. To further confirm that **24** is indeed a non-covalent, active-center-directed,

inhibitor of α-amylase, kinetic studies and X-ray crystallographic analysis could be performed in the future.

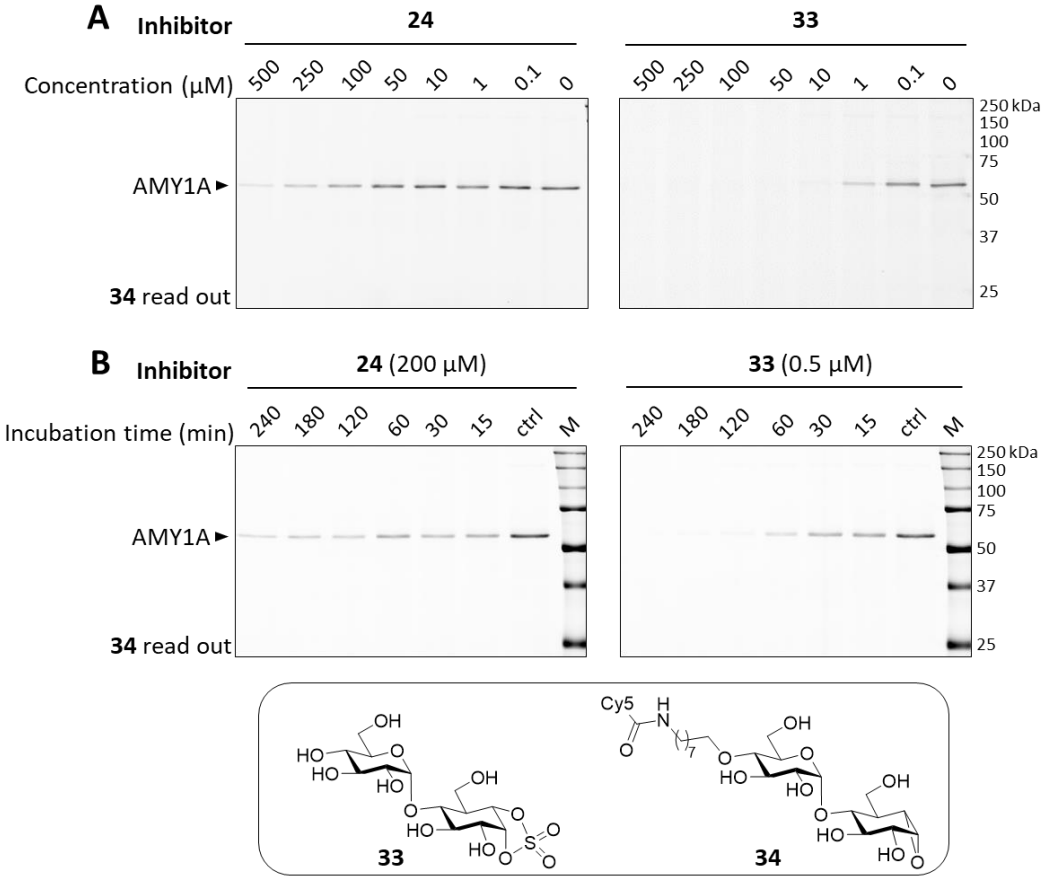
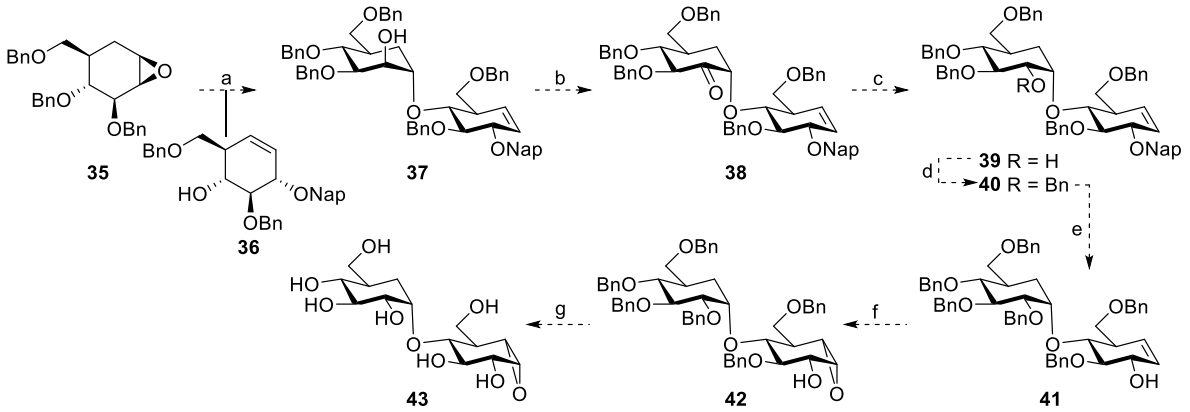


Figure 6.1. Competitive ABPP of α-amylase in human saliva with inhibitors **24** and **33**. A) The human saliva sample (10 μg total protein) was pre-incubated with **24** or **33** at varying concentrations (30 min, pH 5.0, 37 °C) prior to labeling with ABP **34** (5 μM, 1 h, pH 5.0, 37 °C); B) The human saliva sample (10 μM total protein) was pre-incubated with 200 μM of **24** or 0.5 μM of **33** for different time periods (pH 5.0, 37 °C) prior to labeling with ABP **34** (5 μM, 1 h, pH 5.0, 37 °C), "ctrl" means no pre-incubation with competitors.

The glycosidic linkage in an *endo*-acting 'disaccharide' inhibitor may be susceptible to hydrolysis by *exo*-glycosidases present in complex biological samples. One strategy to prevent the potential *exo*-glycosidase cleavage is to 'cap' O4 of the non-reducing end pyranose with an alkyl group (Chapter 2 and Chapter 3). An alternative strategy would be the substitution of the endocyclic oxygen of the non-reducing end pyranose with a methylene group to create stabilized *endo*-amylase inhibitors. For example, regioselective opening of epoxide **35**[7] with cyclohexene **36** (Chapter 3) in a *trans*-diaxial fashion under Lewis acid catalysis may provide alcohol **37** (Scheme 6.5). Oxidation of the hydroxy group with Dess-Martin periodate would afford C2-ulose derivative **38**. Sodium borohydride ($NaBH_4$) mediated reduction[8-10] of the

ketone may provide pseudodisaccharide **39** with a 1,2-*cis* linkage. Benzyl protection of the secondary hydroxy group in **39** followed by removal of the naphthyl group in **40** would afford allylic alcohol **41**. Stereoselective epoxidation of **41** could be achieved via an iodocarbonate intermediate to provide **42**, which after global debenzylation using Pearlman's catalyst under hydrogen atmosphere would provide stabilized *carba*-pseudodisaccharide **43**.



Scheme 6.5. Proposed synthetic scheme towards stabilized α-amylase inhibitor **43**. Reagents and conditions: a) $Cu(OTf)_2$, DCM; b) Dess-Martin periodinane, DCM; c) $NaBH_4$, MeOH/THF; d) BnBr, NaH, TBAI, DMF; e) DDQ, DCM/H_2O; f) *i*) Boc_2O, DMAP, THF; *ii*) NIS, AcOH, DCM; *iii*) NaOMe, MeOH, DCM; g) $Pd(OH)_2$/C, H_2, MeOH/H_2O/dioxane.

Chapter 4 describes the synthesis of a set of β-D-*gluco*-cyclophellitol aziridine inhibitors and probes, which are functionalized at both position C6 and aziridine nitrogen. X-ray crystallographic analysis revealed the structurally exclusive accommodation of the two functionalities by recombinant human β-glucocerebrosidase (rhGBA). Selectivity and potency of these bifunctional aziridines towards GBA were investigated both *in vitro* and *in situ*. The IC_{50} values for the inhibition of rhGBA showed that the new bifunctional cyclophellitol aziridine inactivators are around 10-15 times less potent than the corresponding C6-mono-functionalized cyclophellitol epoxides that were reported previously.[11] In addition, incubation of the bifunctional probes with mouse brain lysates indeed exhibited a selective labeling of GBA over GBA2, whilst a decreased selectivity was observed by *in situ* labeling of HEK293T cells containing endogenous GBA and overexpressed GBA2. While the bifunctional aziridines are less selective and potent than their mono-functionalized epoxide counterparts, they remain potent inactivators of GBA which can be used for the study of GBA in relation to Gaucher disease.

Siastatin B, a natural product originally isolated from a *Streptomyces* culture,[12] has previously been reported as a potent and effective inhibitor of β-D-glucuronidases.[13] However,

the molecular structure of siastatin B appears too bulky to be accommodated within a typical retaining β-glucuronidase active site pocket.[14-16] NMR analyses of trifluoroacetamide containing siastatin B derivatives have suggested that solvent mediated breakdown of these molecules can rapidly occur, releasing a hemiaminal/hydrated ketone that acts as the true inhibitor.[17,18] However, this rearrangement has not been demonstrated for siastatin B. In **Chapter 5**, X-ray crystallographic analysis of co-complexes between siastatin B and different enzymes including human heparanase (HPSE), *A. capsulatum* β-glucuronidase (AcGH79), *Burkholderia pseudomallei* (BpHEP), and *E. coli* β-glucuronidase (EcGusB) is presented, confirming that it is hemiaminal breakdown product **44** or hydrated ketone **45**, and not siastatin B, that binds to the enzyme active site (Figure 6.2). In order to understand the action of the breakdown products, a panel of *galacto*- and *gluco*-configured 1-*N*-iminosugar derivatives was synthesized. Preliminary structural analysis for enzyme inhibition by the synthetic *gluco*-configured iminosugars revealed that compound **49** binds to the active site of both HPSE and AcGH79 in a hydrated form, similar to **45** derived from siastatin B when soaked with HPSE. This result indicates that these synthetic iminosugars may be potential inhibitors of HPSE and β-D-glucuronidases. To evaluate the inhibitory properties of these compounds in depth, kinetic studies and competitive ABPP experiments can be performed in future studies.

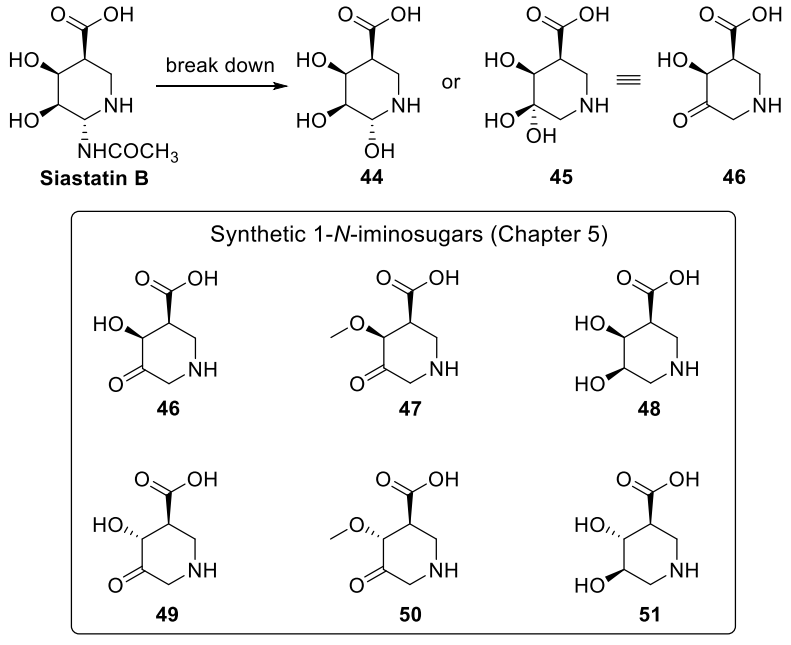


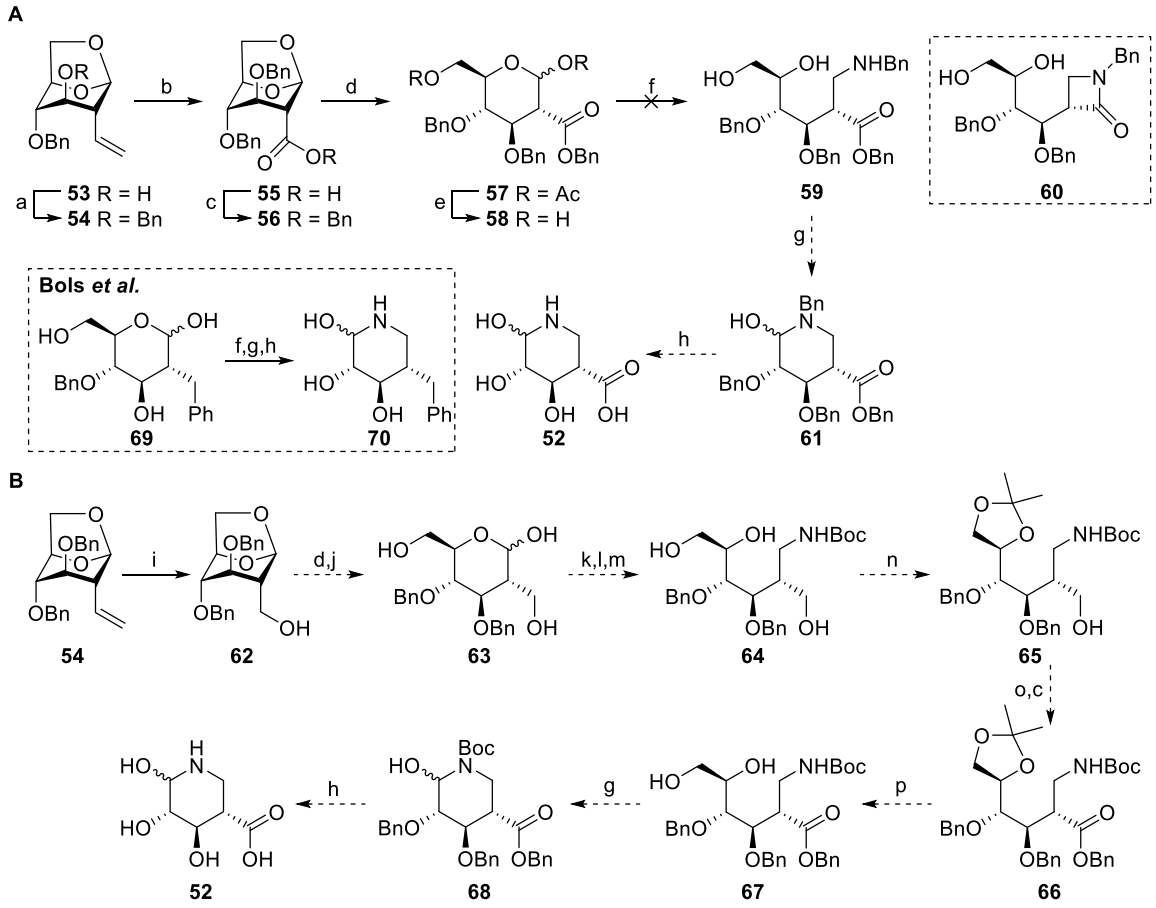
Figure 6.2. Chemical structures of siastatin B and its breakdown products as well as the synthetic *galacto*- and *gluco*-configured 1-*N*-iminosugar derivatives as described in chapter 5.

Since AcGH79, BpHEP and EcGusB all have hemiaminal **44** bound in the active site when treated with siastatin B, it would be of interest to synthesize compound **44** and its *gluco*-

configured analogue **52**. The synthesis of glucuronic acid-type noeuromycin **52** has been partially conducted (Scheme 6.6A). Compound **53** was prepared on a large scale from commercially available tri-*O*-acetyl-D-glucal according to reported procedures.[19,20] Benzyl protection of the secondary alcohol in **53** afforded compound **54**. Oxidative cleavage of the terminal alkene in **54** resulted in the formation of carboxylic acid **55**, which was condensed with benzyl alcohol under the activation of *N*,*N'*-diisopropylcarbodiimide (DIC), providing ester **56** in good yield. Direct hydrolysis of the 1,6-anhydride with aqueous 2 M HCl/dioxane (1/1, v/v) at elevated temperature (100 °C) for 4 h gave a complex mixture of products, presumably due to elimination of the 3-substituent.[21] In contrast, acetolysis[20] of the 1,6-anhydro bond in **56** was readily accomplished using trifluoracetic acid in acetic anhydride, providing diacetate **57** as a mixture of anomers (α/β = 1:2). Deacetylation of **57** could be promoted by *p*-toluenesulfonic acid (*p*-TsOH), affording lactol **58** in good yield. Transformation from **58** to product **52** was attempted by following the procedures reported by Bols and co-workers for the preparation of 6-deoxy-6-phenylnoeuromycin **70**.[21] The three-step sequence involves reductive amination of lactol **58** to provide vicinal diol **59**, followed by periodate cleavage to afford hemiaminal **61** which can be deprotected using Pearlman's catalyst under hydrogen atmosphere to provide final product **52**. However, reductive amination of **58** with benzylamine proved to be sluggish and an intractable mixture of products was obtained. This is attributed to the presence of the neighboring benzyl ester group which might be attacked intramolecularly by the amine to form lactam **60**, of which the molecular weight was detected by TLC-MS and LC-MS analyses.

Based on these results, it was envisioned that reductive amination of the lactol should be performed at an early stage. For this purpose, an alternative synthesis route was proposed (Scheme 6.6B). Oxidation of **54** with osmium tetroxide in the presence of sodium periodate gave the labile aldehyde, which was directly reduced with sodium borohydride to provide alcohol **62** in 84% yield over two steps. Acetolysis of the 1,6-anhydride followed by deacetylation would provide lactol **63**. Reductive amination of **63** with allylamine followed by deallylation using Wilkinson's catalyst[22] would provide a free amine intermediate which could be protected with *tert*-butyloxycarbonyl (Boc) group to afford **64**. The vicinal diol in **64** may then be regioselectively protected as an isopropylidene acetal. The remaining primary alcohol in **65** could be oxidized to a carboxylic acid, which is condensed with benzyl alcohol to afford ester **66**. Acidic hydrolysis of the isopropylidene acetal followed by periodate cleavage of the

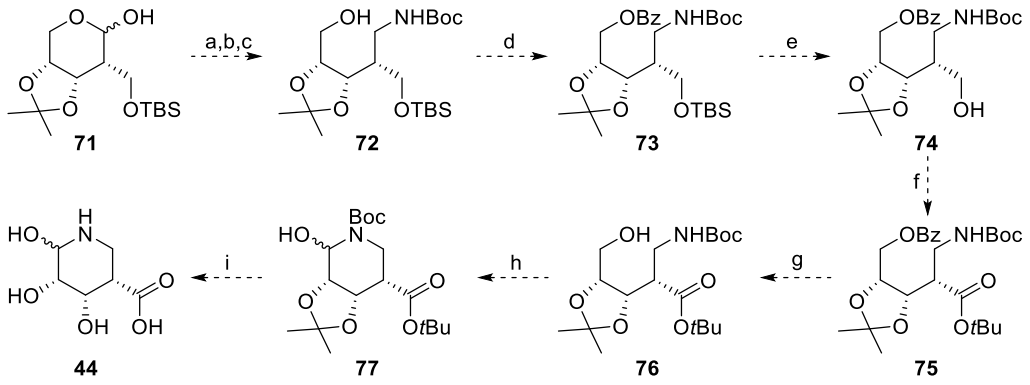
liberated vicinal diol in **67** would provide hemiaminal **68**, which after global deprotection under palladium catalyzed hydrogenolysis in the presence of HCl may provide product **52**.



Scheme 6.6. Proposed synthetic schemes for the preparation of glucuronic acid-type noeuromycin **52**. Regents and conditions: a) BnBr, NaH, TBAI, DMF, 0 °C to rt, 79%; b) $RuCl_3.3H_2O$, $NaIO_4$, EtOAc/ACN/H_2O (2/2/3), rt, 57%; c) BnOH, DIC, DMAP, DCM, 0 °C, 92%; d) TFA, Ac_2O, 0 °C, 93%; e) *p*-TsOH, THF/H_2O (9/1), 70 °C, 82%; f) $BnNH_2$, HOAc, $NaCNBH_3$, MeOH, 50 °C; g) $NaIO_4$, THF/H_2O, rt; h) Pd/C, H_2, HCl, EtOH/H_2O, rt; i) *i*) OsO_4, $NaIO_4$, dioxane/H_2O, rt; *ii*) $NaBH_4$, THF/isopropanol, 0 ^{0}C, 84% over two steps; j) NaOMe, MeOH; k) allylamine, HOAc, $NaCNBH_3$; l) $(Ph_3P)_3RhCl$, MeCN/H_2O, reflux; m) Boc_2O, TEA, DCM; n) *p*-TsOH, 2,2-dimethoxypropane; o) Jones reagent, acetone; p) 80% HOAc.

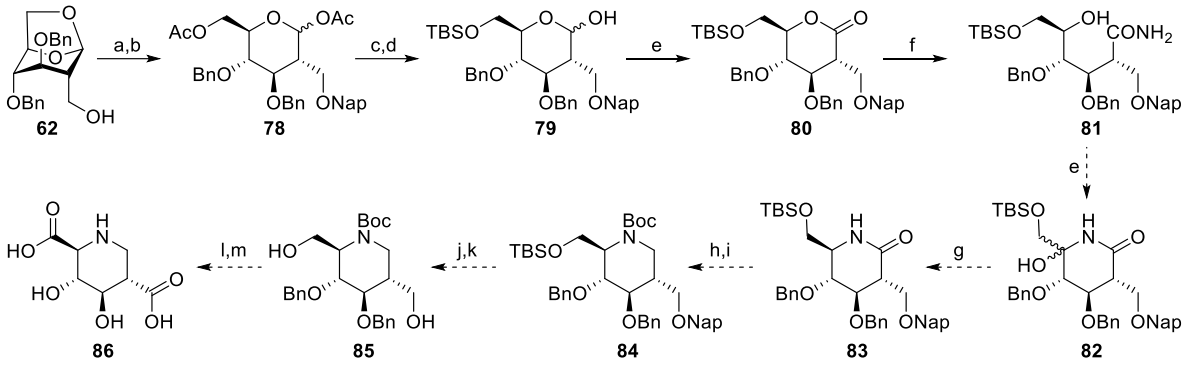
D-Galacturonic acid-type noeuromycin **44** can be prepared from known compound **71**[23] (Scheme 6.7). Reductive amination of **71** with allylamine followed by deallylation using Wilkinson's catalyst[22] and Boc-protection of the resulting free amine would afford alcohol **72**. Benzoylation of the primary hydroxy group in **72** followed by desilylation using tetrabutylammonium fluoride (TBAF) would provide compound **74**. Jones oxidation of the primary alcohol to the corresponding carboxylic acid followed by esterification with *tert*-butyl *N,N'*-diisopropylcarbamimidate in toluene may afford ester **75**. Debenzoylation of **75** under

Zémplen conditions would provide **76**. Alcohol **76** can then be oxidized with 2,2,6,6-tetramethylpiperidinyloxyl (TEMPO) and *meta*-chloroperbenzoic acid (*m*-CPBA) to provide hemiaminal **77**, which after global deprotection under acidic conditions may finally afford product **44**.



Scheme 6.7. Proposed synthetic scheme for the preparation of galacturonic acid-type noeuromycin **44**. Regents and conditions: a) allylamine, HOAc, $NaCNBH_3$; b) $(Ph_3P)_3RhCl$, $MeCN/H_2O$, reflux; c) Boc_2O, TEA, DCM; d) BzCl, Et_3N, DMAP, DCM; e) TBAF, THF; f) *i*) Jones reagent, acetone; *ii*) *tert*-butyl *N*,*N'*-diisopropylcarbamimidate, toluene; g) NaOMe, MeOH; h) TEMPO, *m*-CPBA, DCM; i) HCl, dioxane, H_2O.

Both compounds **44** and **52** may not be stable due to the presence of the 2-hydroxy group, which can undergo pH-dependent Amadori rearrangement[24,25] in aqueous solutions (Chapter 5). As a replacement, a 2-carboxy group can be introduced to increase the stability. The synthesis of the *gluco*-configured 2-carboxy derivative has been started (Scheme 6.8). The primary alcohol in **62** was protected as a naphthyl ether followed by acetolysis of the 1,6-anhydro bond to afford diacetate **78** as a mixture of anomers ($\alpha/\beta \approx 2:1$). Deacetylation under Zemplén conditions followed by selective protection of the primary alcohol with a *tert*-butyldimethylsilyl (TBS) group afforded lactol **79** in moderate yield. Oxidation of **79** under Albright-Goldman conditions[26] provided lactone **80**, which was treated with methanolic ammonium resulting in the isolation of hydroxy-amide **81** in 55% yield. To complete the synthesis several more steps could be executed. Dimethyl sulfoxide (DMSO)-acetic anhydride (Ac_2O) mediated oxidation of **81** may provide a diastereomeric mixture of hemiaminal **82**. Treatment of **82** with sodium cyanoborohydride under acidic conditions may allow the formation of **83**. Reduction of lactam **83** could be achieved by using borane-methyl sulfide complex[27] and the resulting amine may be protected with a Boc group to afford **84**. Removal of the naphthyl and silyl ether groups would provide diol **85**. Jones oxidation of the primary alcohols in **85** followed by global deprotection under palladium catalyzed hydrogenolysis conditions in the presence of HCl may finally provide inhibitor **86**.



Scheme 6.8. Proposed synthetic scheme for the preparation of **86**. Regents and conditions: a) NapBr, NaH, DMF, 0 °C to rt, 92%; b) TFA, Ac_2O, 0 °C, 80%; c) NaOMe, MeOH/DCM, rt; d) TBSCl, imidazole, DMAP, DCM, 0 °C to rt, 57% over two steps; e) DMSO, Ac_2O, 30 °C, 79%; f) 7 M NH_3 in MeOH, rt, 55%; g) $NaCNBH_3$, HCOOH, CH_3CN; h) BH_3Me_2S, DCM; i) Boc_2O, TEA, DCM; j) TBAF, THF; k) DDQ, DCM/H_2O; l) Jones reagent, acetone; m) Pd/C, H_2, HCl, dioxane/H_2O.

6.4 Experimental methods

6.4.1 Competitive ABPP experiments

Concentrated human saliva (10 µg total protein) was diluted with 150 mM McIlvaine buffer (pH 5.0, supplemented with 10 mM $CaCl_2$ and 10 mM NaCl) to a final 10 µL volume, pre-incubated with 2.5 µL inhibitor **24** or inhibitor **33** (diluted in McIlvaine buffer) at the indicated concentration at 37 °C for the indicated time period. These were followed by labeling with 2.5 µL ABP **34** (diluted in McIlvaine buffer) at a final ABP concentration of 5 µM at 37 °C for 1 h. Samples were then denatured with 4 µL sample buffer (5x Laemmli buffer, containing 50% (v/v) 1M Tris-HCl pH 6.8, 50% (v/v) glycerol, 10% (w/v) Dithiothreitol (DTT), 10% (w/v) sodium dodecyl sulphate (SDS), 0.01% bromophenol blue) and heated at 98 °C for 5 minutes. Proteins were resolved by electrophoresis in 10% SDS-PAGE gels, running at a constant of 90V for 30 minutes followed by 120V for 60 minutes. Wet slab gels were scanned on fluorescence using a Typhoon FLA9500 Imager (GE Healthcare) using λ_{EX} 635 nm; λ_{EM} > 665 nm and images were processed using ImageLab 5.2.1 (BioRad). Gels were subsequently extensively stained with Coomassie Brilliant Blue solution for assessing total protein amounts in each lane of sample.

6.4.2 Chemical synthesis

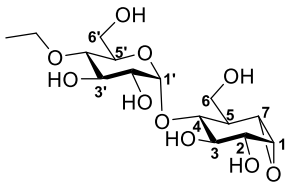
General experimental details

All reagents were of experimental grade and were used without further purification unless stated otherwise. Dichloromethane (DCM), tetrahydrofuran (THF) and toluene were stored over 3 Å molecular sieves and *N*,*N*-dimethylformamide (DMF) was stored over 4 Å molecular sieves, which were dried *in vacuo* before use. All reactions were performed under an Argon or N_2 atmosphere unless stated otherwise. Reactions were monitored by analytical thin layer chromatography (TLC) using Merck aluminum sheets pre-coated with silica gel 60 with detection by UV-absorption (254 nm) and by

spraying with a solution of $(NH_4)_6Mo_7O_{24}\cdot H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4\cdot H_2O$ (10 g/mL) 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 manually using Screening Device b.v. silica gel 60 (0.04-0.063 mm) in the indicated solvents. LC-MS analysis was performed on a LCQ Advantage Max (Thermo Finnigan) ion-trap spectrometer (ESI^+) coupled to a Surveyor HPLC system (Thermo Finnigan) equipped with a C18 column (Gemini, 4.6 mm x 50 mm, 5 µM particle size, Phenomenex). The applied buffers were A: H_2O, B: acetonitrile (MeCN) and C: 1% aqueous trifluoroacetic acid (TFA). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker AV-400 (400/101 MHz), Bruker AV-500 (500/126 MHz), and Bruker AV-850 (850/214 MHz) spectrometers in the given solvent. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (TMS) as internal standard (^{1}H NMR in $CDCl_3$) or the residual signal of the deuterated solvent. Coupling constants (*J*) are given in Hz. All given ^{13}C-NMR spectra are proton decoupled. The following abbreviations are used to describe peak patterns when appropriate: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), Ar (aromatic), C_q (quaternary carbon). 2D NMR experiments (COSY, HSQC) were carried out to assign protons and carbons of the new structures. High-resolution mass spectrometry (HRMS) analysis was performed with a LTQ Orbitrap mass spectrometer (Thermo Finnigan), equipped with an electronspray ion source in positive mode (source voltage 3.5 kV, sheath gas flow 10 mL/min, capillary temperature 250 °C) with resolution R = 60000 at m/z 400 (mass range m/z = 150 – 2000) and dioctyl phthalate (m/z = 391.28428) as a "lock mass". The high-resolution mass spectrometer was calibrated prior to measurements with a calibration mixture (Thermo Finnigan).

Experimental Procedures and Characterization Data of Products

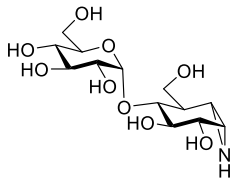
Compound 2



Compound **1** (108 mg, 0.116 mmol) was dissolved in dry THF (1.5 mL), TBAF (1.0 M in THF, 0.70 mL, 0.70 mmol) was added and the reaction was stirred at rt for 1 h. The reaction was quenched with sat. aq. NH_4Cl, extracted with DCM (2 x). The combined organic layers were washed with sat. aq. $NaHCO_3$, water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 4:1→2:1) to afford the alcohol intermediate (85 mg, 0.10 mmol, 90%) as a white solid. The alcohol (29 mg, 35 µmol) was dissolved in a mixture of MeOH/H_2O/dioxane (2/1/2, 1 mL) under Argon and $Pd(OH)_2$/C (20 wt%, 38 mg, 53 µmol) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 3 h under H_2 atmosphere, the mixture was filtered over a small celite pad and evaporated to afford product **2** in high purity as a white solid (11.4 mg, 31.1 µmol, 89%) after lyophilization. ^{1}H NMR (400 MHz, MeOD) δ 5.01 (d, J = 3.9 Hz, 1H, H1'), 3.96 – 3.84 (m, 2H, H6a and 1CH*H* Et), 3.83 – 3.73 (m, 3H), 3.73 – 3.62 (m, 4H), 3.58 (dd, J = 9.8, 8.7 Hz, 1H), 3.44 (dd, J = 9.7, 3.9 Hz, 1H, H2'), 3.35 (d, J = 9.7 Hz, 1H, H4), 3.27 (dd, J = 4.0, 1.8 Hz, 1H, epoxide), 3.21 (d, J = 4.0 Hz, 1H, epoxide), 3.16 (t, J = 9.4

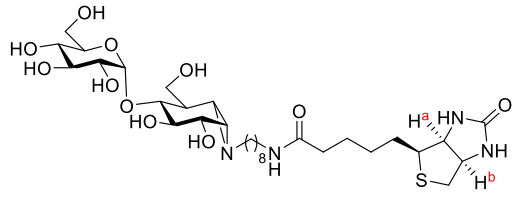
Hz, 1H), 2.04 (ddd, *J* = 9.1, 5.7, 3.0 Hz, 1H, H5), 1.18 (t, *J* = 7.0 Hz, 3H, CH_3). ^{13}C NMR (101 MHz, MeOD) *δ* 103.1 (C1'), 83.0, 79.2, 75.1, 74.9, 74.0, 73.6, 72.7, 69.3 (CH_2 Et), 62.2 (C6'), 61.6 (C6), 57.9 (epoxide), 55.5 (epoxide), 45.2 (C5), 16.0 (CH_3). HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{15}H_{26}O_{10}Na$ 389.1418, found 389.1422.

Compound 4



Ammonium (3.0 mL) was condensed in a dry flask at -60 °C under inert atmosphere. Sodium (50 mg, 2.2 mmol) was added and the resulting deep-blue solution was stirred for 15 min to dissolve all sodium completely. Aziridine **3** (30 mg, 31 µmol) and ^{t}BuOH (30 µL, 310 µmol) were taken up in dry THF (1 mL) and slowly added to the reaction mixture. After stirring for 1 h at -60 °C, the reaction was quenched carefully with H_2O. The mixture was allowed to warmed to rt and stirred until all ammonium had evolved. After sovent evaporation under reduced pressure the crude was re-dissolved in MilliQ-H_2O and eluted over a column packed with Amberlite CG-50 (NH_4^+) with 0.5M NH_4OH as eluent. The fractions containing product was concentrated *in vacuo*, affording compound **4** (10.1 mg, 29.7 µmol, 96%) as a white powder after lyophilization. ^{1}H NMR (500 MHz, D_2O) *δ* 5.12 (d, *J* = 4.0 Hz, 1H, H1'), 3.91 – 3.79 (m, 3H, H2, H6a and H6a'), 3.78 – 3.69 (m, 3H, H6b, H6b' and H5'), 3.69 – 3.63 (m, 1H, H3'), 3.59 – 3.52 (m, 2H, H3 and H2'), 3.42 – 3.33 (m, 2H, H4 and H4'), 2.57 (dd, *J* = 6.4, 3.5 Hz, 1H, H1), 2.36 (d, *J* = 6.4 Hz, 1H, H7), 2.02 (ddd, *J* = 9.4, 6.3, 3.0 Hz, 1H, H5). ^{13}C NMR (126 MHz, D_2O) *δ* 100.6 (C1'), 80.8, 74.0, 72.9 (C3'), 72.4 (C5'), 71.7, 70.9 (C2), 69.3, 61.6 (C6), 60.4 (C6'), 43.2 (C5), 35.4 (C1), 31.6 (C7). HRMS (ESI) m/z: $[M+H]^+$ calc for $C_{13}H_{23}NO_9$ 360.1265, found 360.1267.

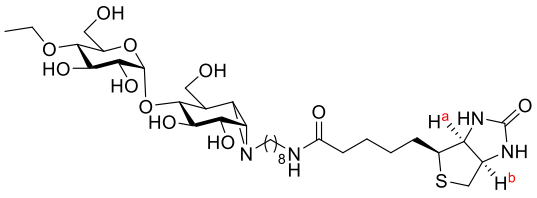
Compound 7



Compound **5** (4.6 mg, 10 µmol) was dissolved in dry DMF (0.5 mL). DIPEA (3.8 µL, 22 µmol) and biotin-OSu (3.8 mg, 11 µmol) were added and the mixture was stirred overnight at rt. Full conversion was observed by LC-MS analysis and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (3/1) again and lyophilized to obtain product **7** (1.4 mg, 2.0 µmol, 20%) as a white powder. ^{1}H NMR (500 MHz, MeOD) *δ* 4.99 (d, *J* = 3.9 Hz, 1H, H1'), 4.50 (ddd, *J* = 8.0, 5.0, 0.9 Hz, 1H, H^b), 4.30 (dd, *J* = 7.9, 4.5 Hz, 1H, H^a), 3.89 (dd, *J* = 11.0, 3.0 Hz, 1H, H6a), 3.86 – 3.78 (m, 1H, H6a'), 3.70 (tdd, *J* = 7.3, 5.4, 3.6 Hz, 4H, H2, H5', H6b and H6'b), 3.61 (t, *J* = 9.3 Hz, 1H, H3'), 3.56 (dd, *J* = 9.9, 8.8 Hz, 1H, H3), 3.45 – 3.40 (m, 1H, H2'), 3.31-3.27 (m, 1H, H4'), 3.23 – 3.12 (m, 4H, H4, CHS and C*H*$_2$NHC=O), 2.93 (dd, *J* = 12.8, 5.0 Hz, 1H, CH*H*S), 2.71 (d, *J* = 12.7 Hz, 1H, CH*H*S), 2.35 (dt, *J* = 11.6, 7.2 Hz, 1H, CH*H*-N aziridine), 2.19 (t, *J* = 7.3 Hz, 2H, CH_2C=O), 2.19-

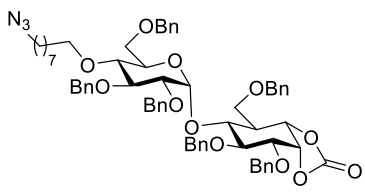
2.14 (m, 1H, CH*H*-N aziridine), 1.92 (ddd, *J* = 9.7, 6.6, 3.1 Hz, 1H, H5), 1.86 (dd, *J* = 6.5, 3.6 Hz, 1H, H1), 1.79 – 1.54 (m, 7H, H7 and 6CH*H* linker), 1.46 (dp, *J* = 23.1, 7.4 Hz, 4H, 4CH*H* linker), 1.34 (s, 8H, 8CH*H* linker). ^{13}C NMR (126 MHz, MeOD) δ 176.0, 103.2 (C1'), 84.1 (C4), 75.9 (C3), 75.0 (C3'), 74.4, 74.0 (C2'), 72.8, 71.5 (C4'), 63.4 (CH^{a}), 62.9 (C6), 62.7 (C6'), 62.2 (CH_2-N aziridine), 61.6 (CH^{b}), 57.0 (CHS), 45.9 (C1), 45.8 (C5), 42.2 (C7), 41.1 (CH_2S), 40.4 (CH_2NHC=O), 36.8 (CH_2NHC=O), 30.6, 30.5, 30.4, 30.3, 29.8, 29.5, 28.4, 28.0, 27.0 (9CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{31}H_{55}N_4O_{11}S$ 691.3583, found 691.3581.

Compound 8



Compound **6** (6.5 mg, 13 µmol) was dissolved in dry DMF (0.5 mL). DIPEA (5.1 µL, 29 µmol) and biotin-OSu (5.4 mg, 16 µmol) were added and the mixture was stirred overnight at rt. Full conversion was observed by LC-MS and the product was purified by semi-preparative reversed phase HPLC (linear gradient. Solution used: A: 50 mM NH_4HCO_3 in H_2O, B: MeCN). The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (3/1) again and lyophilized to obtain compound **8** (4.2 mg, 5.8 µmol, 44%) as a white powder. ^{1}H NMR (850 MHz, MeOD) δ 4.97 (d, *J* = 3.9 Hz, 1H, H1'), 4.50 (ddd, *J* = 7.9, 5.0, 0.9 Hz, 1H, H^{b}), 4.30 (dd, *J* = 7.9, 4.5 Hz, 1H, H^{a}), 3.92 – 3.85 (m, 2H, H6a and CH*H* Et), 3.77 (dd, *J* = 11.7, 2.0 Hz, 1H, H6'a), 3.73 – 3.62 (m, 6H, H2, H3', H5', H6b, H6'b and CH*H* Et), 3.55 (dd, *J* = 9.9, 8.8 Hz, 1H, H3), 3.42 (dd, *J* = 9.7, 3.9 Hz, 1H, H2'), 3.23 – 3.18 (m, 1H, CHS), 3.18 – 3.12 (m, 4H, H4, H4' and C*H*$_2$NHC=O), 2.93 (dd, *J* = 12.8, 5.0 Hz, 1H, CH*H*S), 2.71 (d, *J* = 12.7 Hz, 1H, CH*H*S), 2.34 (ddd, *J* = 11.6, 8.1, 6.3 Hz, 1H, CH*H*-N aziridine), 2.22 – 2.15 (m, 3H, CH_2C=O and CH*H*-N aziridine), 1.91 (ddd, *J* = 10.0, 6.9, 3.2 Hz, 1H, H5), 1.86 (dd, *J* = 6.4, 3.7 Hz, 1H, H1), 1.77 – 1.54 (m, 7H, H7 and 6CH*H* linker), 1.49 (h, *J* = 7.1 Hz, 2H, 2CH*H* linker), 1.46 – 1.41 (m, 2H, 2CH*H* linker), 1.38 – 1.30 (m, 8H, 8CH*H* linker), 1.18 (t, *J* = 7.0 Hz, 3H, CH_3). ^{13}C NMR (214 MHz, MeOD) δ 176.0, 166.1, 103.1 (C1'), 84.1, 79.1, 75.8 (C3), 75.1, 74.1 (C2'), 73.6, 72.7, 69.3 (CH_2 Et), 63.4 (CH^{a}), 62.9 (C6), 62.2 (C6'), 62.1 (CH_2-N aziridine), 61.6 (CH^{b}), 57.1 (CHS), 45.9 (C1), 45.8 (C5), 42.1 (C7), 41.1 (CH_2S), 40.4 (CH_2NHC=O), 36.8 (CH_2C=O), 30.6, 30.5, 30.4, 30.4, 29.8, 29.5, 28.4, 28.0, 27.0 (9CH_2 linker), 16.0 (CH_3). HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{33}H_{59}N_4O_{11}S$ 719.3896, found 719.3892.

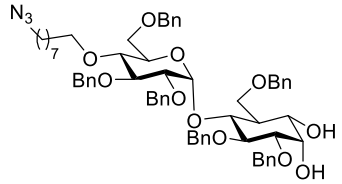
Compound 11



Donor **10** (313 mg, 0.45 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (9 mL) under nitrogen and stirred over fresh flame-dried 3 Å molecular sieves, after which DMF (0.55 mL, 7.2 mmol, 16 equiv. of donor) was added. After cooling to -20 °C, NIS (101 mg, 0.45 mmol) and TfOH (40 µL, 0.45 mmol) were added

successively. The reaction was pre-activated at -20 °C for 2 h. Then acceptor **9** (147 mg, 0.30 mmol) was added and the reaction mixture was placed in an ice bath. The mixture was stirred at 0 °C until TLC-analysis showed complete conversion of the acceptor. The reaction was then quenched with a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$, stirred vigorously at rt until the brown color faded. The mixture was filtered over celite and diluted with DCM. The layers were separated and the organic layer was washed with water and brine, dried with anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by size exclusion (DCM/MeOH = 1/1) giving the product as a single isomer (*dr* > 20/1, 300 mg) which was further purified by flash column chromatography (pentane/EtOAc 11:1 → 7:1) to obtain compound **11** (292 mg, 0.27 mmol, 90%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 7.16 (m, 28H, CH Ar), 7.15 – 7.06 (m, 2H, CH Ar), 5.11 (d, *J* = 3.6 Hz, 1H, H1'), 5.07 – 4.98 (t, *J* = 8.8 Hz, 1H, H7), 4.89 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.80 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.78 – 4.70 (m, 2H, CH*H* Bn and H1), 4.69 – 4.54 (m, 2H, CH*H* Bn), 4.56 – 4.37 (m, 4H, CH*H* Bn), 4.32 – 4.21 (m, 2H, CH*H* Bn), 4.17 (d, *J* = 11.5 Hz, 1H, CH*H* Bn), 3.98 – 3.81 (m, 5H, H2, H3, H4, H3' and H6a), 3.80 – 3.65 (m, 2H, CH*H*O linker and H5'), 3.63 (dd, *J* = 9.6, 2.5 Hz, 1H, H6b), 3.55 – 3.28 (m, 5H, H2', H4′, H6′ab and CH*H*O linker), 3.21 (t, *J* = 6.9 Hz, 2H, CH_2N_3), 2.79 (ddt, *J* = 11.8, 9.4, 2.4 Hz, 1H, H5), 1.61 – 1.50 (m, 2H, 2CH*H* linker), 1.50 – 1.38 (m, 2H, 2CH*H* linker), 1.39 – 1.11 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 154.8 (C=O), 138.9, 138.4, 138.1, 138.0, 137.2, 137.1 (6C_q Ar), 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.3, 128.2, 128.2, 128.2, 128.1, 128.0, 128.0, 127.9, 127.9, 127.8, 127.7, 127.7, 127.6, 127.5, 127.5, 127.4 (30CH Ar), 95.2 (C1'), 82.2, 80.8, 79.6 (C2'), 77.7 (C4'), 75.6 (Bn), 74.0, 74.0 (C1), 73.8 (C7), 73.7 (Bn), 73.5 (Bn), 73.2 (Bn), 73.1 (CH_2O linker), 73.0 (Bn), 72.0, 71.5 (Bn), 71.1(C5'), 68.1 (C6'), 65.6 (C6), 51.5 (CH_2N_3), 41.3 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{64}H_{73}N_3O_{12}Na$ 1098.5087, found 1098.5083.

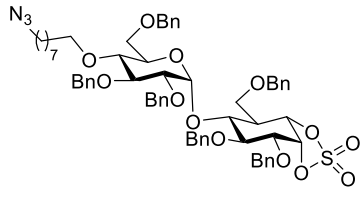
Compound 12



Compound **11** (284 mg, 0.26 mmol) was dissolved in a mixture of DCM/MeOH (1/1, 5.2 mL). NaOMe (30 wt% in MeOH, 30 µL, 0.16 mmol) was added and the reaction mixture was stirred at rt for 2 h. After which the reaction was diluted with DCM (5 mL) and neutralized with washed Amberlite IR-120 H^+ resin until pH ≈ 7. The mixture was filtered and the resin was washed with DCM (3 x 5 mL). The combined filtrates were concentrated *in vacuo* and the product was purified by flash column chromatography (pentane/acetone 9:1 → 6:1) to obtain compound **12** (260 mg, 0.244 mmol, 94%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.37 – 7.09 (m, 30H, CH Ar), 5.78 (d, *J* = 3.7 Hz, 1H, H1'), 4.97 (d, *J* = 11.9 Hz, 1H, CH*H* Bn), 4.85 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.76 (dd, *J* = 11.3, 2.0 Hz, 2H, CH*H* Bn), 4.63 – 4.39 (m, 7H, CH*H* Bn), 4.30 (d, *J* = 12.1 Hz, 1H, CH*H* Bn), 4.15 (t, *J* = 2.9 Hz, 1H), 4.04 (t, *J* = 9.0 Hz, 1H), 3.91 – 3.73 (m, 4H), 3.70 (ddt, *J* = 13.4, 10.0, 2.6 Hz, 3H), 3.50 – 3.37 (m, 5H), 3.36 – 3.30 (m, 1H, CH*H*O linker), 3.21 (t, *J* = 6.9 Hz, 2H, CH_2N_3), 3.00 (brs, 1H, OH), 2.64 (brs, 1H, OH), 2.37 – 2.25 (m, 1H, H5), 1.60 – 1.49

(m, 2H, 2CH*H* linker), 1.48 – 1.36 (m, 2H, 2CH*H* linker), 1.36 – 1.06 (m, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.0, 139.0, 138.2, 138.1, 138.1, 137.8 (6C_q Ar), 128.5, 128.4, 128.4, 128.3, 128.3, 128.1, 128.1, 128.0, 127.8, 127.8, 127.7, 127.6, 127.5, 127.5, 127.1, 126.5 (30CH Ar), 96.5 (C1'), 82.7, 82.0, 80.5, 79.4, 77.8, 75.4 (Bn), 73.9 (Bn), 73.6 (Bn), 73.4 (Bn), 73.1 (Bn), 73.0 (CH_2O linker), 72.5 (Bn), 71.2, 71.2, 69.7, 69.1, 68.4 (C6'), 67.7 (C6), 51.5 (CH_2N_3), 42.9 (C5), 30.5, 29.5, 29.2, 28.9, 26.8, 26.2 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{63}H_{75}N_3O_{11}Na$ 1072.5294, found 1072.5294.

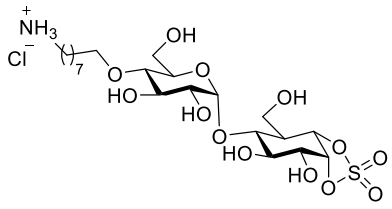
Compound 13



Compound **12** (0.28 g, 0.26 mmol) was dissolved in dry DCM (5 mL) and purged with nitrogen. After cooling to 0 °C, Et_3N (0.29 mL, 2.1 mmol) and thionyl chloride (57 µL, 0.78 mmol) were added successively. After stirring at 0 °C for 10 min, the reaction was diluted with DCM (100 mL), washed with sat. aq. $NaHCO_3$ (50 mL), H_2O (2 x 50 mL) and brine, dried over Na_2SO_4, filtered and concentrated *in vacuo* to give the crude cyclic sulfite as a yellow oil. The crude was directly dissolved in a mixture of EtOAc/MeCN (2.5 mL/2.5 mL) and cooled to 0 °C. A solution of $RuCl_3.3H_2O$ (0.5 M in H_2O, 5.2 µL, 2.6 µmol) and $NaIO_4$ (85 mg, 0.39 mmol) in H_2O (1.25 mL) was added and the mixture was stirred vigorously at 0 °C for 40 min. The reaction was quenched by addition of sat. aq. $Na_2S_2O_3$ (2 mL) and stirred vigorously at rt for 15 min. Then the mixture was diluted with H_2O (50 mL) and extracted with EtOAc (2 x 80 mL). The combined organic layers were washed with brine, dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was purified by flash column chromatography (pentane/acetone 15:1 → 13:1) affording compound **13** (0.26 g, 0.23 mmol, 88%) as a clean oil. (*Note: oxidation with $RuCl_3$ and $NaIO_4$ should be monitored every 10 min by TLC analysis. The reaction should be quenched as soon as the conversion of the sulfite is complete. Otherwise an over-oxidized side product would be formed due to prolonged reaction times*). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.38 – 7.16 (m, 28H, CH Ar), 7.14 – 7.04 (m, 2H, CH Ar), 5.24 (d, *J* = 3.5 Hz, 1H, H1'), 5.22 – 5.18 (m, 1H, H7), 4.99 (dd, *J* = 6.9, 3.5 Hz, 1H, H1), 4.85 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.77 (d, *J* = 10.9 Hz, 1H, CH*H* Bn), 4.72 – 4.59 (m, 3H, CH*H* Bn), 4.59 – 4.43 (m, 3H, CH*H* Bn), 4.40 (d, *J* = 11.5 Hz, 2H, CH*H* Bn), 4.32 (dd, *J* = 13.7, 11.8 Hz, 2H, CH*H* Bn), 3.96 – 3.83 (m, 5H, H2, H3, H4, H3' and H6a), 3.74 (dt, *J* = 9.2, 6.5 Hz, 1H, CH*H*O linker), 3.67 (dt, *J* = 9.8, 2.7 Hz, 1H, H5'), 3.57 (dd, J = 9.7, 2.4 Hz, 1H, H6b), 3.52 – 3.41 (m, 3H, H2', H4' and H6'a), 3.41 – 3.30 (m, 2H, H6'b and CH*H*O linker), 3.22 (t, *J* = 6.9 Hz, 2H, CH_2N_3), 2.86 (td, *J* = 12.1, 2.3 Hz, 1H, H5), 1.63 – 1.50 (m, 2H, 2CH*H* linker), 1.47 – 1.38 (m, 2H, 2CH*H* linker), 1.27 (ddt, *J* = 21.3, 11.0, 7.1 Hz, 8H, 8CH*H* linker). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.9, 138.3, 138.0, 137.9, 137.3, 137.2 (6C_q Ar), 128.7, 128.6, 128.6, 128.6, 128.5, 128.5, 128.4, 128.3, 128.1, 128.0, 128.0, 128.0, 127.9, 127.9, 127.7, 127.5, 127.4 (30CH Ar), 96.0 (C1'), 82.1, 81.0, 79.6 (C2'), 79.4 (C1), 78.3 (C7), 77.7 (C4'), 75.6 (Bn), 74.3, 73.8 (Bn), 73.6 (Bn), 73.6 (Bn), 73.3 (Bn), 73.1 (CH_2O linker), 72.5 (Bn), 71.5, 71.3 (C5'), 68.3 (C6'),

64.9 (C6), 51.6 (CH_2N_3), 41.9(C5), 30.6, 29.6, 29.3, 29.0, 26.8, 26.2 (6CH_2 linker). HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{63}H_{73}N_3O_{13}SNa$ 1134.4756, found 1134.4761.

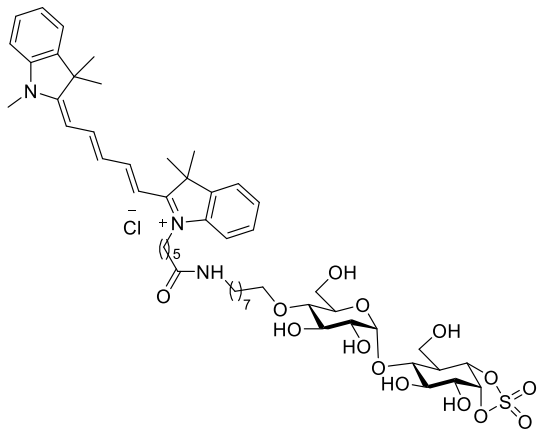
Compound 14



Compound **13** (64 mg, 58 µmol) was dissolved in a mixture of ^{t}BuOH/H_2O/dioxane (1/2/1, 2.4 mL). The solution was purged with nitrogen for 2 minutes under vigorous stirring. Then concentrated HCl (12 M, 28 µL, 0.34 mmol) and $Pd(OH)_2$/C (20 wt%, 64 mg) were added under nitrogen atmosphere. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 5 h under H_2 atmosphere, the mixture was filtered over a small celite pad and evaporated to afford product **14** as a white solid (29.6 mg, 51 µmol, 88%) after lyophilization. ^{1}H NMR (500 MHz, D_2O) δ 5.48 (t, J = 3.8 Hz, 1H, H1), 5.31 – 5.28 (m, 2H, H1' and H7), 4.01 – 3.89 (m, 3H, H2, H3 and H6a), 3.86 – 3.72 (m, 6H), 3.70 – 3.55 (m, 3H, H2′, H4 and CH*H*O linker), 3.34 – 3.23 (m, 1H), 2.96 (t, J = 7.6 Hz, 2H, C*H*$_2$NH$_2$), 2.41 – 2.33 (m, 1H, H5), 1.60 (dt, J = 30.6, 7.2 Hz, 4H, 4CH*H* linker), 1.32 (d, J = 6.7 Hz, 8H, 8CH*H* linker). ^{13}C NMR (126 MHz, D_2O) δ 100.3 (C1'), 84.5 (C1), 82.2 (C7), 77.5, 76.4, 73.3 (CH_2O linker), 73.2, 72.9, 72.0, 71.9, 68.1, 60.2 (C6'), 56.5 (C6), 43.7 (C5), 39.6 (CH_2NH_2), 29.2, 28.2, 28.1, 26.7, 25.5, 25.1 (6CH_2 linker). HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{21}H_{40}NO_{13}S$ 546.2215, found 546.2219.

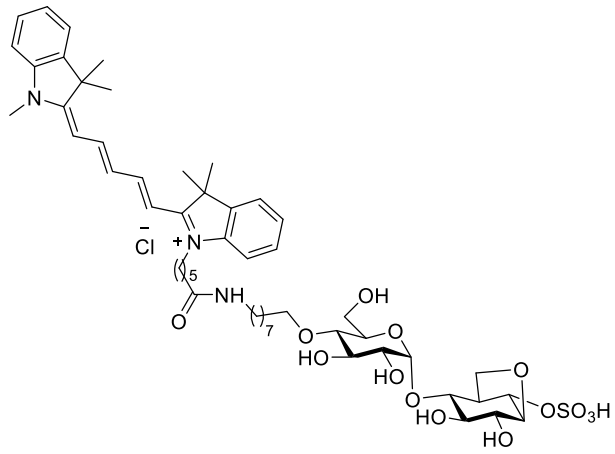
ABP 15 and proposed structure of the rearranged product

Cy5COOH (18 mg, 35 µmol) was dissolved in dry DMF (0.5 mL). 2,3,4,5,6-Pentaflurophenol (32 mg, 175 µmol), Et_3N (24 µL, 175 µmol) and DIC (8.1 µL, 52.5 µmol) were added successively and the mixture was stirred at rt for 3 h. Part of the stock solution (0.38 mL, 26.6 µmol) was added to the amine **14** (15.5 mg, 26.6 µmol) in dry DMF (0.2 mL) and Et_3N (3.7 µL, 26.6 µmol) was added. The reaction mixture was stirred at rt for 26 h until full conversion was observed by LC-MS analysis.



The crude was purified by semi-preparative reversed phase HPLC with linear gradient, solutions A: 50 mM HOAc in H_2O and B: CH_3CN. The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain compound **15** (3.3 mg, 3.2 µmol, 11%) as a blue powder. ^{1}H NMR (500 MHz, MeOD) δ 8.24 (t, J = 13.0 Hz, 2H), 7.49 (d, J = 7.4 Hz, 2H), 7.45 – 7.38 (m, 2H), 7.33 – 7.23 (m, 4H), 6.63 (t, J = 12.4 Hz, 1H), 6.28 (dd, J = 13.7, 4.5 Hz, 2H), 5.28 (t, J = 4.0 Hz, 1H, H1), 5.21 (dd, J = 10.1, 4.4 Hz, 1H, H7), 5.13 (d, J = 3.9 Hz, 1H, H1'), 4.10 (t, J = 7.5 Hz, 2H, CH_2N^+), 4.03 (dd, J = 11.5, 2.4 Hz, 1H, H6a), 3.91 – 3.78 (m, 3H, H3, H6'a and CH*H*O linker), 3.78 – 3.69 (m,

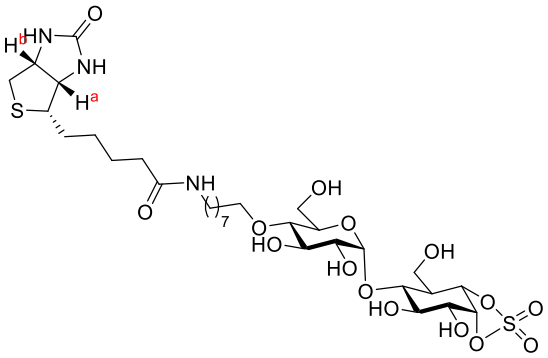
3H, H2, H3' and H5'), 3.68 – 3.60 (m, 5H, H6b, H6'b and CH_3N), 3.59 – 3.49 (m, 2H, CH*H*O linker and H4), 3.47 (dd, *J* = 9.8, 3.9 Hz, 1H, H2'), 3.14 – 3.10 (m, 3H, H4' and C*H*$_2$NHC=O), 2.23 – 2.14 (m, 3H, CH_2C=O and H5), 1.83 (q, *J* = 7.8 Hz, 2H, 2CH*H* linker), 1.73 (s, 14H, 4CH_3 and 2CH*H* linker), 1.61 – 1.41 (m, 4H, 4CH*H* linker), 1.40 – 1.20 (m, 10H, 10CH*H* linker). ^{13}C NMR (214 MHz, MeOD) δ 175.7, 175.4, 174.7, 155.5, 155.5, 144.3, 143.6, 142.7, 142.5, 129.8, 129.8, 126.6, 126.3, 126.3, 123.4, 123.3, 112.1, 111.9, 104.4, 104.3, 103.1 (C1'), 85.8 (C1), 82.8 (C7), 80.0 (C4), 79.4 (C4'), 75.2 (C3'), 74.7 (C3), 74.3 (C2'), 74.2 (C5'), 74.0 (CH_2O linker), 69.9 (C2), 62.4 (C6'), 56.8 (C6), 50.6 (C_q), 50.5 (C_q), 45.9 (C5), 44.8 (CH_2N^+), 40.4 (*C*H_2NHC=O), 36.4 (*C*H_2C=O), 31.4 (CH_2 linker), 30.7 (CH_3N), 30.5, 30.4, 30.3, 28.2 (4CH_2 linker), 28.0 (2CH_3), 27.9 (CH_2 linker), 27.8 (2CH_3), 27.4, 27.2, 26.6 (3CH_2 linker). LC-MS (ESI): R_t 6.03 min, linear gradient 30%→70% B in 10 min; m/z: $[M]^+$ calculated for $C_{53}H_{76}N_3O_{14}S$ 1010.50, observed 1010.83. HRMS (ESI) m/z: $[M]^+$ calculated for $C_{53}H_{76}N_3O_{14}S$ 1010.5043, found 1010.5049.



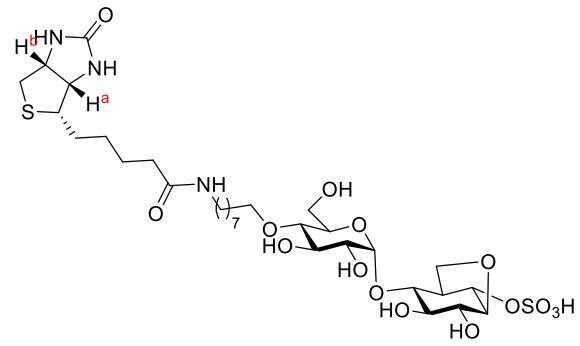
When the crude was purified by HPLC with linear gradient, solutions A: 50 mM NH_4HCO_3 in H_2O and B: CH_3CN, a rearranged product was obtained (2.1 mg, 2.0 µmol, 7%). ^{1}H NMR (500 MHz, MeOD) δ 8.24 (t, *J* = 13.0 Hz, 2H), 7.49 (dt, *J* = 7.4, 1.7 Hz, 2H), 7.45 – 7.38 (m, 2H), 7.32 – 7.23 (m, 4H), 6.65 (t, *J* = 12.4 Hz, 1H), 6.29 (dd, *J* = 13.7, 8.9 Hz, 2H), 4.91 (d, *J* = 3.8 Hz, 1H, H1'), 4.64 (dd, *J* = 8.9, 5.3 Hz, 1H, H1), 4.50 (d, *J* = 5.3 Hz, 1H, H7), 4.42 (t, *J* = 4.8 Hz, 1H, H3), 4.27 – 4.22 (m, 1H, H2,), 4.19 (t, *J* = 4.2 Hz, 1H, H4), 4.11 (t, *J* = 7.4 Hz, 2H, CH_2N^+), 3.88 – 3.78 (m, 2H, CH*H*O linker and H6'a), 3.77 – 3.68 (m, 2H, H5' and H3'), 3.67 – 3.58 (m, 5H, H6'b, NCH_3 and H6a), 3.56 – 3.49 (m, 2H, CH*H*O linker and H6b), 3.37 – 3.33 (m, 1H, H2'), 3.15 – 3.10 (m, 3H, H4' and C*H*$_2$NHC=O), 2.71 (td, *J* = 7.6, 3.7 Hz, 1H, H5), 2.21 (t, *J* = 7.2 Hz, 2H, CH_2C=O), 1.83 (t, *J* = 7.7 Hz, 2H, 2CH*H* linker), 1.73 (d, *J* = 0.7 Hz, 14H, 4CH_3 and 2CH*H* linker), 1.57 – 1.42 (m, 4H, 4CH*H* linker), 1.39 – 1.27 (m, 10H, 10CH*H* linker). ^{13}C NMR (214 MHz, MeOD) δ 175.8, 175.3, 174.7, 155.5, 144.3, 143.6, 142.7, 142.537, 129.8, 126.7, 126.2, 123.4, 123.3, 112.1, 111.9, 104.4, 100.0 (C1'), 83.9 (C7), 82.9 (C4), 79.2 (C4'), 76.5 (C3), 75.1 (C3'), 73.7 (C2'), 73.5 (CH_2O linker), 73.2 (C5'), 72.5 (C1), 72.3 (C2), 64.1 (C6), 62.3 (C6'), 50.6 (C_q), 50.5 (C_q), 45.4 (C5), 44.8 (CH_2N^+), 40.4 (*C*H_2NHC=O), 36.7 (*C*H_2C=O), 31.6 (NCH_3), 31.4, 30.5, 30.3, 30.3, 28.2 (5CH_2 linker), 28.0 (2CH_3), 27.9 (CH_2 linker), 27.8 (2CH_3), 27.3, 27.2, 26.6 (3CH_2 linker) ppm. LC-MS (ESI): R_t 5.31 min, linear gradient 30%→70% B in 10 min; m/z: $[M]^+$ calculated for $C_{53}H_{76}N_3O_{14}S$ 1010.50, observed 1010.17.

ABP 16 and proposed structure of the rearranged product

Compound **14** (9.6 mg, 16.5 µmol) was dissolved in dry DMF (0.5 mL). DIPEA (5.8 µL, 33 µmol) and biotin-OSu (8.5 mg, 25 µmol) were added and the reaction mixture was stirred at rt for 6 h until full conversion was observed by LC-MS analysis.



The crude was purified by semi-preparative reversed phase HPLC with linear gradient, solution A: 50 mM HOAc in H_2O and B: CH_3CN. The fractions were concentrated under reduced pressure, co-evaporated with Milli-Q/MeCN (1/1, 3 x), dissolved in Milli-Q/^{t}BuOH (1/1) again and lyophilized to obtain compound **16** (2.3 mg, 2.9 µmol, 17%) as a white powder. ^{1}H NMR (500 MHz, MeOD) δ 5.28 (dd, J = 4.4, 3.4 Hz, 1H, H1), 5.21 (dd, J = 10.1, 4.4 Hz, 1H, H7), 5.14 (d, J = 3.9 Hz, 1H, H1'), 4.49 (ddd, J = 7.9, 5.0, 0.9 Hz, 1H, H^b), 4.30 (dd, J = 7.9, 4.4 Hz, 1H, H^a), 4.04 (dd, J = 11.4, 2.3 Hz, 1H, H6a), 3.91 – 3.79 (m, 3H, H3, H6'a and CH*H*O linker), 3.78 – 3.70 (m, 3H, H2, H3' and H5'), 3.69 – 3.63 (m, 2H, H6b and H6'b), 3.61 – 3.49 (m, 2H, CH*H*O linker and H4), 3.47 (dd, J = 9.8, 3.9 Hz, 1H, H2'), 3.24 – 3.09 (m, 4H, CHS, H4' and CH_2NHC=O), 2.93 (dd, J = 12.8, 5.0 Hz, 1H, CH*H*S), 2.71 (d, J = 12.8 Hz, 1H, CH*H*S), 2.19 (dd, J = 9.1, 5.6 Hz, 3H, CH_2C=O and H5), 1.79 – 1.50 (m, 6H, 6CH*H* linker), 1.51 – 1.30 (m, 12H, 12CH*H* linker). ^{13}C NMR (126 MHz, MeOD) δ 176.0, 166.1 (2C=O), 103.0 (C1'), 85.8 (C1), 82.8 (C7), 79.8 (C4), 79.4 (C4'), 75.2 (C3'), 74.7 (C3), 74.2 (C2'), 74.2 (C5'), 74.0 (CH_2O linker), 70.0 (C2), 63.4 (CH^a), 62.4 (C6'), 61.6 (CH^b), 57.0 (CHS), 56.8 (C6), 45.9 (C5), 41.1 (CH_2S), 40.3 (*C*H_2NHC=O), 36.8 (CH_2C=O), 31.4, 30.5, 30.4, 30.3, 29.7, 29.5, 27.1, 27.1, 27.0 (9CH_2 linker). LC-MS (ESI): R_t 6.75 min, linear gradient 00%→50% B in 10 min; m/z: $[M+H]^+$ calculated for $C_{31}H_{54}N_3O_{15}S_2$ 772.30, observed 772.17. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{31}H_{54}N_3O_{15}S_2$ 772.2991, found 772.2995.

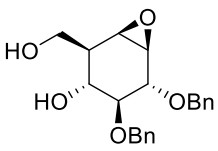


When the crude was purified by HPLC with linear gradient, solutions A: 50 mM NH_4HCO_3 in H_2O and B: CH_3CN, a rearranged product was obtained (1.8 mg, 2.3 µmol, 13%). ^{1}H NMR (850 MHz, MeOD) δ 4.68 – 4.60 (dd, J = 8.5, 5.1 Hz, 1H, H1), 4.52 – 4.48 (m, 2H, H^b and H7), 4.43 (t, J = 4.8 Hz, 1H, H3), 4.31 (dd, J = 7.9, 4.4 Hz, 1H, H^a), 4.25 (ddd, J = 9.0, 5.0, 1.5 Hz, 1H, H2), 4.20 (ddd, J = 5.0, 3.7, 1.5 Hz, 1H, H4), 3.85 (dt, J = 9.0, 6.4 Hz, 1H, CH*H*O linker), 3.81 (dd, J = 11.9, 2.2 Hz, 1H, H6'a), 3.75 – 3.70 (m, 2H, H3' and H5'), 3.66 (dd, J = 11.9, 4.9 Hz, 1H, H6'b), 3.62 (dd, J = 10.9, 7.4 Hz, 1H, H6a), 3.56 (dt, J = 9.1, 6.7 Hz, 1H, CH*H*O linker), 3.51 (dd, J = 11.0, 7.9 Hz, 1H, H6b), 3.37 (dd, J = 9.7, 3.9 Hz,

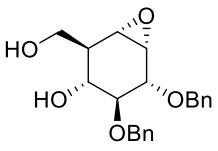
1H, H2'), 3.24 – 3.19 (m, 1H, CHS), 3.19 – 3.12 (m, 3H, C*H*$_2$NHC=O and H4'), 2.93 (dd, *J* = 12.8, 5.0 Hz, 1H, CH*H*S), 2.74 – 2.66 (m, 2H, CH*H*S and H5), 2.20 (td, *J* = 7.3, 1.7 Hz, 2H, CH$_2$C=O), 1.78 – 1.52 (m, 6H, 6CH*H* linker), 1.50 (t, *J* = 6.9 Hz, 2H, 2CH*H* linker), 1.46 – 1.42 (m, 2H, 2CH*H* linker), 1.42 – 1.32 (m, 8H, 8CH*H* linker). ^{13}C NMR (214 MHz, MeOD) δ 176.0, 166.2 (2C=O), 99.8 (C1'), 83.8 (C7), 82.5 (C4), 79.2 (C4'), 76.4 (C3), 75.1 (C3'), 73.6 (CH$_2$O linker), 73.6 (C2'), 73.2 (C5'), 72.5 (C1), 72.2 (C2), 64.1 (C6), 63.4 (CH^a), 62.2 (C6'), 61.6 (CH^b), 57.0 (CHS), 45.4 (C5), 41.1 (CH$_2$S), 40.4 (*C*H$_2$NHC=O), 36.8 (*C*H$_2$C=O), 31.4, 30.5, 30.4, 30.4, 29.8, 29.5, 27.9, 27.2, 27.0 (9CH$_2$ linker). LC-MS (ESI): R$_t$ 5.48 min, linear gradient 00%→50% B in 10 min; m/z: [M+H]$^+$ calculated for $C_{31}H_{54}N_3O_{15}S_2$ 772.30, observed 772.00.

Compound 26 and S1

Cyclohexene **25** (1.02 g, 3.00 mmol) was dissolved in dry DCM (30 mL) and cooled to 0 °C. *m*-CPBA (77 wt%, 4.03 g, 18.0 mmol) was added and the mixture was stirred at 0 °C for 2 days. The reaction was quenched by a mixture of sat. aq. $Na_2S_2O_3$ and sat. aq. $NaHCO_3$. The layers were separated and the aqueous layer was extracted with DCM (2 x). The combined organic layers were washed with sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The ratio of two isomers was determined as 1:8 (α/β) by crude ^{1}H NMR spectrum. Then the crude was purified by silica column chromatography (Pentane/EtOAc 7:1 → 3:1) to obtain β-epoxide **26** (770 mg, 2.16 mmol, 72%) and α-epoxide **S1** (86 mg, 0.24 mmol, 8%) as a white solid.

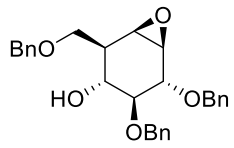


β-epoxide (**26**): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 – 7.18 (m, 10H), 4.94 (d, *J* = 11.3 Hz, 1H), 4.81 (d, *J* = 11.4 Hz, 1H), 4.66 (dd, *J* = 11.3, 6.9 Hz, 2H), 4.00 (dd, *J* = 10.9, 6.7 Hz, 1H), 3.92 – 3.84 (m, 1H), 3.81 (dd, *J* = 8.0, 0.8 Hz, 1H), 3.47 (td, *J* = 9.8, 1.3 Hz, 1H), 3.38 (dd, *J* = 10.0, 7.9 Hz, 1H), 3.27 (dd, *J* = 3.9, 1.7 Hz, 1H), 3.15 (d, *J* = 3.7 Hz, 1H), 2.88 (d, *J* = 1.7 Hz, 1H), 2.72 (s, 1H), 2.19 – 2.10 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.3, 137.5, 128.7, 128.7, 128.6, 128.6, 128.2, 128.1, 128.1, 128.0, 128.0, 128.0, 83.6, 79.5, 75.0, 72.7, 68.7, 64.1, 55.0, 53.1, 43.4. HRMS (ESI) m/z: [M+Na]$^+$ calculated for $C_{21}H_{24}O_5Na$ 379.1516, found 379.1510.



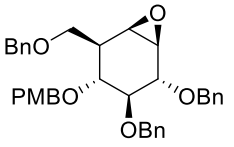
α-epoxide (**S1**): ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.49 – 7.18 (m, 10H), 5.00 (d, *J* = 11.2 Hz, 1H), 4.89 – 4.75 (m, 2H), 4.65 (d, *J* = 11.2 Hz, 1H), 3.99 – 3.77 (m, 3H), 3.58 (dd, *J* = 9.9, 8.0 Hz, 1H), 3.44 (t, *J* = 9.7 Hz, 1H), 3.37 (ddd, *J* = 3.9, 1.9, 0.6 Hz, 1H), 3.13 (d, *J* = 3.9 Hz, 1H), 2.76 (s, 1H), 2.19 (d, *J* = 4.0 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.4, 138.1, 128.8, 128.7, 128.6, 128.1, 128.1, 128.0 , 80.9, 79.6, 75.6, 72.2, 70.6, 63.0, 54.6, 54.2, 43.6. HRMS (ESI) m/z: [M+Na]$^+$ calculated for $C_{21}H_{24}O_5Na$ 379.1516, found 379.1514.

Compound S2



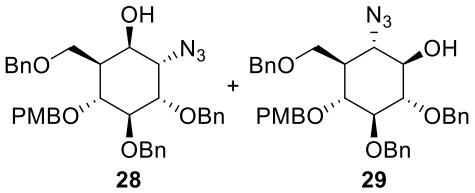
Compound **26** (693 mg, 1.94 mmol), 2-animoethyl diphenylborinate (44 mg, 0.19 mmol), KI (322.7 mg, 1.94 mmol) and K_2CO_3 (295 mg, 2.13 mmol) were dissolved in dry MeCN (20 mL). Then BnBr (346 µL, 2.91 mmol) was added. The reaction mixture was heated to 60 °C and stirred overnight. Most of the solvent was evaporated *in vacuo* and the residue was diluted with EtOAc, washed H_2O (2 x) and brine, dried over $MgSO_4$, filtrated, and concentrated *in vacuo*. The crude was purified by silica column chromatography (Pentane/EtOAc 11:1 → 7:1) to yield compound **S2** (744 mg, 1.67 mmol, 86%) as a yellow-white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.43 – 7.20 (m, 15H), 4.92 (d, J = 11.3 Hz, 1H), 4.81 (d, J = 11.4 Hz, 1H), 4.70 (d, J = 4.8 Hz, 1H), 4.68 (d, J = 4.8 Hz, 1H), 4.57 (d, J = 2.3 Hz, 2H), 3.89 (dd, J = 9.0, 5.0 Hz, 1H), 3.84 – 3.79 (m, 1H), 3.68 (t, J = 8.7 Hz, 1H), 3.43 (dd, J = 4.1, 1.7 Hz, 1H), 3.41 – 3.30 (m, 2H), 3.20 (d, J = 3.7 Hz, 1H), 2.72 (s, 1H), 2.26 (tdd, J = 8.7, 5.0, 1.8 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 138.5, 138.2, 137.6, 128.7, 128.7, 128.6, 128.6, 128.5, 128.5, 128.2, 128.1, 128.1, 128.0, 128.0, 128.0, 127.8, 127.7, 127.7, 83.8, 79.4, 75.0, 73.6, 72.8, 70.0, 67.4, 54.9, 54.0, 42.2. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{28}H_{30}O_5Na$ 469.1986, found 469.1982.

Compound 27



Compound **S2** (744 mg, 1.67 mmol) was dissolved in dry DMF (10 mL) and cooled to 0 °C. Then NaH (60% in mineral oil, 160 mg, 6.68 mmol) was added and the reaction mixture was stirred at 0 °C for 10 min and at rt for 20 min. After cooling to 0 °C again, 4-methoxybenzyl chloride (452 µL, 3.34 mmol) and TBAI (31 mg, 0.084 mmol) were added and the reaction mixture was stirred at rt for 4 h. The reaction was carefully quenched with H_2O at 0 °C, diluted with water, extracted with EtOAc (2 x). The combined organic layers were washed with water (2 x), brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 18:1 → 9:1) to obtain compound **27** (710 mg, 1.25 mmol, 75%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.20 (m, 15H), 7.09 – 7.01 (m, 2H), 6.82 – 6.74 (m, 2H), 4.84 (s, 2H), 4.80 – 4.66 (m, 3H), 4.51 (d, J = 2.6 Hz, 2H), 4.31 (d, J = 10.5 Hz, 1H), 3.85 (dd, J = 8.2, 0.8 Hz, 1H), 3.73 (s, 3H), 3.71 (d, J = 3.6 Hz, 1H), 3.59 – 3.49 (m, 2H), 3.45 (dt, J = 3.9, 1.0 Hz, 1H), 3.24 (t, J = 10.0 Hz, 1H), 3.17 (d, J = 3.7 Hz, 1H), 2.27 (dddd, J = 10.3, 8.7, 3.5, 1.6 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 138.8, 138.3, 137.7, 130.4, 129.8, 128.5, 128.4, 128.3, 127.9, 127.9, 127.7, 127.7, 127.6, 127.6, 127.5, 113.8, 85.0, 79.8, 75.3, 75.0, 74.7, 73.2, 73.0, 68.6, 55.6, 55.2, 53.9, 42.5. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{36}H_{38}O_6Na$ 589.2561, found 589.2558.

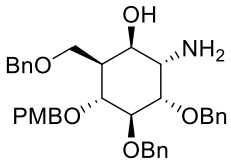
Compound 28 and 29



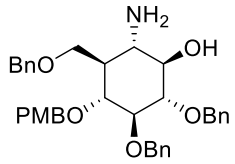
Compound **27** (671 mg, 1.18 mmol) was dissolved in dry DMF (15 mL) and cooled to 0 °C. NaN_3 (769 mg, 11.8 mmol) and $LiClO_4$ (2.50 g, 23.6 mmol) were added, then the mixture was heated to 80 °C and stirred overnight. After cooling to rt, the reaction was diluted with H_2O and extracted with EtOAc (2 x). The combined organic layers were washed with H_2O (3 x), brine, dried over $MgSO_4$, filtrated, concentrated *in vacuo*. Products **28** and **29** (680 mg) was obtained as an inseparable mixture and the crude was directly used for next step without further purification.

Compound 30 and **S3**

The inseparable mixture of compounds **28** and **29** (1.07 g, 1.76 mmol) was dissolved in anhydrous THF (35 mL) and purged with N_2 for 2 min under vigorous stirring. Then PtO_2 (0.16 g, 0.70 mmol) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 4 h under H_2 atmosphere, the reaction mixture was filtrated over celite and concentrated *in vacuo*. The crude residue was purified with silica column chromatography (DCM/MeOH 80:1 → 20:1) to obtain compound **30** (471 mg, 0.807 mmol, 46%) as a clean oil and compound **S3** (452 mg, 0.774 mmol, 44%) as a yellow-white solid.

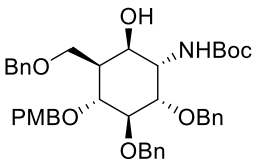


Amino alcohol **30**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.23 (m, 15H), 7.15 – 7.07 (m, 2H), 6.84 – 6.77 (m, 2H), 4.94 (d, J = 10.9 Hz, 1H), 4.84 (dd, J = 10.7, 8.7 Hz, 2H), 4.74 – 4.62 (m, 2H), 4.53 – 4.38 (m, 3H), 4.09 (dd, J = 9.1, 3.1 Hz, 1H), 4.02 (dd, J = 3.5, 2.1 Hz, 1H), 4.01 – 3.85 (m, 3H), 3.75 (s, 3H), 3.69 (dd, J = 9.1, 2.4 Hz, 1H), 3.47 (t, J = 3.6 Hz, 1H), 2.21 (dq, J = 10.9, 2.6 Hz, 1H), 1.69 – 1.07 (br s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 139.3, 138.8, 137.4, 131.1, 129.6, 128.7, 128.4, 128.4, 128.1, 127.9, 127.8, 127.8, 127.6, 127.5, 113.8, 83.8, 80.7, 76.9, 75.6, 75.3, 75.0, 73.7, 72.6, 70.6, 55.3, 52.7, 40.9. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{36}H_{42}NO_6$ 584.3007, found 584.3004.



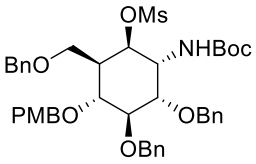
Amino alcohol **S3**: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.23 (m, 15H), 7.15 – 7.08 (m, 2H), 6.87 – 6.77 (m, 2H), 4.97 – 4.84 (m, 3H), 4.80 (t, J = 10.5 Hz, 2H), 4.51 – 4.39 (m, 3H), 3.77 (s, 3H), 3.80 – 3.72 (m, 1H), 3.69 – 3.59 (m, 2H), 3.56 (t, J = 9.2 Hz, 1H), 3.38 (t, J = 9.1 Hz, 1H), 3.23 (t, J = 9.5 Hz, 1H), 2.79 (t, J = 10.4 Hz, 1H), 1.94 (brs, 3H), 1.38 (tt, J = 10.8, 2.5 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.3, 138.8, 128.7, 138.4, 130.8, 129.7, 128.7, 128.6, 128.6, 128.1, 127.9, 127.9, 127.8, 127.7, 113.9, 86.3, 83.9, 78.4, 76.8, 75.6, 75.5, 75.3, 73.1, 65.0, 55.4, 51.0, 46.9. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{36}H_{42}NO_6$ 584.3007, found 584.3002.

Compound S4



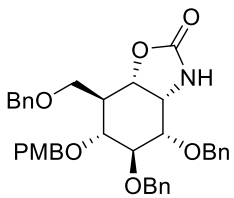
Compound **30** (471 mg, 0.807 mmol) was dissolved in anhydrous DCM (8 mL) and cooled to 0 °C. Then, Et_3N (563 µL, 4.04 mmol) and Boc_2O (212 mg, 0.968 mmol) were added at 0 °C and the reaction mixture was stirred overnight at rt. Next, the reaction was quenched with sat. aq. NH_4Cl and diluted with water. The aqueous layer was extracted with DCM (3 x) and the combined organic layers were washed with H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 7:1 → 4:1) to obtain compound **S4** (512 mg, 0.749 mmol, 93%) as a light yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.23 (m, 15H), 7.12 – 7.05 (m, 2H), 6.84 – 6.77 (m, 2H), 4.95 (d, J = 10.8 Hz, 1H), 4.86 – 4.73 (m, 3H), 4.66 (d, J = 11.1 Hz, 1H), 4.60 – 4.38 (m, 5H), 4.20 – 4.09 (m 2H), 4.03 (dd, J = 9.1, 3.2 Hz, 1H), 3.97 (dd, J = 11.3, 8.8 Hz, 1H), 3.78 (d, J = 3.7 Hz, 1H), 3.75 (s, 3H), 3.70 (dd, J = 9.2, 2.4 Hz, 1H), 3.54 (t, J = 9.0 Hz, 1H), 1.86 (d, J = 11.0 Hz, 1H), 1.43 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.2, 156.1, 139.0, 138.2, 137.3, 130.9, 129.6, 128.7, 128.5, 128.4, 128.2, 128.1, 127.9, 127.8, 127.8, 127.6, 113.9, 84.4, 79.8, 78.1, 76.1, 75.7, 75.2, 73.7, 72.1, 71.9, 70.1, 55.3, 51.9, 41.2, 28.4. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{41}H_{50}NO_8$ 684.3531, found 684.3526.

Compound 31



Compound **S4** (511 mg, 0.75 mmol) was dissolved in anhydrous $CHCl_3$ (7.5 mL) and cooled to 0 °C. Et_3N (521 µL, 3.75 mmol), *N*-methyl-imidazole (597 µL, 7.50 mmol) and MsCl (290 µL, 3.75 mmol) were added and the mixture was stirred at rt for 2 h. Then the reaction mixture was diluted with EtOAc, washed successively with diluted 1 M HCl solution, sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was directly used for next step without further purification.

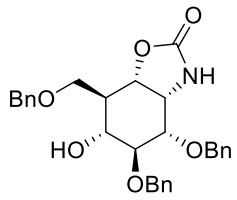
Compound 32



Crude compound **31** was dissolved in anhydrous DMF (37.5 mL) and the reaction was heated to 120 °C for 5 h until full conversion was observed by TLC-analysis. After cooling to rt, the reaction mixture was diluted with H_2O and extracted with EtOAc (2 x). The combined organic layers were washed H_2O (3 x), bine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/Acetone 9:1 → 4:1) to obtain carbamate **32** (329 mg, 0.53 mmol, 71% over 2 steps) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.38 – 7.26 (m, 15H), 7.13 – 7.08 (m, 2H), 6.84 – 6.78 (m, 2H), 5.22 (t, J = 5.5 Hz, 1H), 4.81 – 4.70 (m, 5H), 4.57 (d, J = 11.9 Hz, 1H), 4.53 – 4.41 (m, 3H), 4.05 (dd, J = 7.4, 4.4 Hz, 1H), 3.85 (dd, J = 9.2, 2.3 Hz, 1H), 3.81 (dd, J = 8.3, 7.4 Hz, 1H), 3.77 (s, 3H), 3.61 (dd, J = 9.2, 2.2 Hz, 1H), 3.56 (dd, J = 7.4, 4.3 Hz, 1H), 3.51 (dd, J = 11.6, 8.3 Hz, 1H), 2.14 (ddt, J = 11.5, 9.2, 2.2 Hz, 1H).^{13}C NMR (101 MHz, $CDCl_3$) δ 159.4, 158.9,

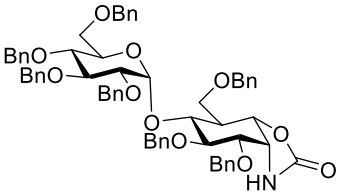
138.4, 138.1, 137.7, 130.6, 129.6, 128.7, 128.5, 128.5, 128.2, 128.1, 127.9, 127.9, 127.8, 113.9, 82.6, 77.4, 75.9, 74.7, 74.7, 74.2, 73.6, 73.5, 65.4, 55.4, 54.5, 44.6. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{37}H_{39}NO_7Na$ 632.2619, found 632.2614.

Compound 17



Compound **32** (320 mg, 0.52 mmol) was dissolved in a mixture of DCM/H_2O (19/1, 10.6 mL). Then DDQ (144 mg, 0.64 mmol) was added and the reaction mixture was stirred at rt for 1.5 h. The reaction was diluted with DCM, washed with sat. aq. $NaHCO_3$ (3 x), H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/Acetone 8:1 → 4:1) to obtain compound **17** (221 mg, 0.45 mmol, 85%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.24 (m, 15H), 5.61 (s, 1H), 4.88 (d, J = 11.3 Hz, 1H), 4.75 – 4.67 (m, 2H), 4.66 – 4.55 (m, 2H), 4.50 (d, J = 1.9 Hz, 2H), 4.08 (dd, J = 7.1, 4.2 Hz, 1H), 3.86 (dd, J = 9.3, 2.5 Hz, 1H), 3.69 – 3.50 (m, 4H), 2.78 (d, J = 2.4 Hz, 1H), 2.05 (ddt, J = 11.8, 9.2, 2.9 Hz, 1H. ^{13}C NMR (101 MHz, $CDCl_3$) δ 159.0, 138.2, 138.1, 137.5, 128.8, 128.7, 128.5, 128.3, 128.1, 128.1, 128.0, 127.9, 127.7, 81.9, 77.8, 75.0, 74.0, 73.5, 73.3, 68.3, 66.3, 54.6, 44.7. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{29}H_{31}NO_6Na$ 512.2044, found 512.2041.

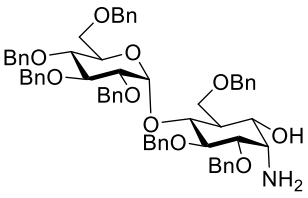
Compound 19



Imidate donor **18** (187 mg, 0.26 mmol) was co-evaporated with toluene (3 x) and dissolved in dry DCM (2.5 mL) under nitrogen and stirred over fresh flame-dried 3 Å molecular sieves, after which DMF (323 µL, 4.2 mmol) was added. The solution was cooled to -20 °C and TfOH (23 µL, 0.26 mmol) was added. After stirring at -20 °C for 1 h, the pre-activation was complete as indicated by TLC-analysis. Acceptor **17** (86 mg, 0.17 mmol) was dissolved with dry DCM (1.0 mL in total) and added to the solution. The reaction mixture was slowly warmed to rt and stirred at rt over-weekend. The reaction was then quenched with Et_3N, filtered and concentrated *in vacuo*. The crude was first purified by silica gel column chromatography (Pentane/EtOAc 7:1 → 2:1) to give an inseparable mixture of the product and unreacted acceptor. Then the mixture was further purified with size exclusion (MeOH/DCM = 1/1) to afford compound **19** (126 mg, 0.12 mmol, 70%) as a clean oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.36 – 7.19 (m, 29H), 7.16 (ddd, J = 7.6, 5.5, 2.7 Hz, 4H), 7.12 – 7.08 (m, 2H), 5.61 (s, 1H), 5.31 (d, J = 3.5 Hz, 1H), 4.95 – 4.85 (m, 2H), 4.81 (dd, J = 10.9, 4.5 Hz, 2H), 4.70 (d, J = 11.9 Hz, 1H), 4.65 (d, J = 12.0 Hz, 1H), 4.56 (d, J = 12.0 Hz, 1H), 4.48 (d, J = 12.1 Hz, 1H), 4.46 – 4.34 (m, 6H), 4.24 (d, J = 12.1 Hz, 1H), 4.01 – 3.91 (m, 3H), 3.89 – 3.83 (m, 2H), 3.79 (dt, J = 10.1, 2.5 Hz, 1H), 3.68 (dd, J = 10.1, 8.9 Hz, 1H), 3.65 – 3.59 (m, 2H), 3.55 (dd, J = 9.8, 3.5 Hz, 1H), 3.42 (dd, J = 10.8, 2.9 Hz, 1H), 3.30 (dd, J = 10.7, 2.0 Hz, 1H), 2.64 (ddt, J = 11.1, 8.9, 2.4 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 159.6, 138.8, 138.6, 138.4, 138.2, 138.0, 137.8, 137.6, 128.7, 128.5, 128.5,

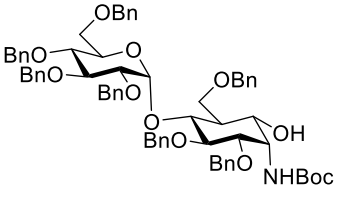
128.5, 128.4, 128.4, 128.4, 128.2, 128.2, 128.0, 127.9, 127.9, 127.8, 127.8, 127.8, 127.7, 127.7, 127.6, 127.3, 95.8, 82.2, 81.4, 79.8, 77.8, 75.7, 75.5, 75.1, 74.6, 73.6, 73.5, 73.0, 72.8, 72.7, 72.2, 70.9, 68.2, 66.2, 52.6, 42.2. HRMS (ESI) m/z: $[M+NH_4]^+$ calculated for $C_{63}H_{69}N_2O_{11}$ 1029.4896, found 1029.4890.

Compound 20



Compound **19** (125 mg, 0.124 mmol) was dissolved in a mixture of EtOH/THF (5/1, 6 mL/1.2 mL). NaOH (1 M in H_2O, 1.8 mL, 1.8 mmol) was added and the reaction mixture was stirred at 70 °C for 8 h. TLC-analysis indicated the presence of starting material so more NaOH (1 M in H_2O, 1 mL) was added and the reaction mixture was further heated to 80 °C and stirred for another 7 h until full conversion was observed by TLC-analysis. The solvent was then evaporated *in vacuo* and the residue was diluted with EtOAc, washed with H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (DCM/MeOH 100:0 → 50:1) to give compound **20** (116 mg, 0.118 mmol, 95%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 – 7.13 (m, 33H), 7.10 (dd, J = 6.9, 2.7 Hz, 2H), 5.76 (d, J = 3.6 Hz, 1H), 4.95 (d, J = 11.8 Hz, 1H), 4.87 (d, J = 10.8 Hz, 1H), 4.78 (dd, J = 10.9, 6.0 Hz, 2H), 4.72 (d, J = 11.8 Hz, 1H), 4.61 (d, J = 12.0 Hz, 1H), 4.57 – 4.48 (m, 4H), 4.41 (d, J = 11.4 Hz, 3H), 4.29 (d, J = 12.2 Hz, 1H), 4.02 (t, J = 8.5 Hz, 1H), 3.95 (t, J = 9.3 Hz, 1H), 3.84 (ddt, J = 8.3, 6.4, 3.2 Hz, 2H), 3.75 (dtd, J = 13.8, 10.1, 2.8 Hz, 3H), 3.65 (t, J = 9.4 Hz, 1H), 3.54 – 3.45 (m, 4H), 3.39 (dd, J = 10.7, 2.0 Hz, 1H), 2.29 (dp, J = 13.6, 4.3 Hz, 1H), 2.19 – 1.87 (brs, 2H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 139.0, 138.9, 138.6, 138.2, 138.2, 138.1, 138.0, 128.5, 128.4, 128.4, 128.3, 128.2, 127.9, 127.9, 127.9, 127.8, 127.8, 127.7, 127.6, 127.6, 127.4, 127.1, 126.6, 96.5, 82.0, 82.1, 81.1, 79.6, 77.9, 75.6, 75.0, 73.6, 73.6, 73.5, 72.9, 72.2, 71.9, 71.0, 69.3, 68.3, 68.2, 52.0, 42.8. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{62}H_{68}NO_{10}$ 986.4838, found 986.4830.

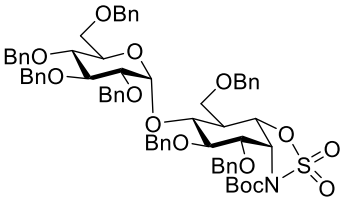
Compound 21



Compound **20** (116 mg, 0.118 mmol) was dissolved in dry DCM (2 mL) and cooled to 0 °C. Et_3N (82 µL, 0.59 mmol) and Boc_2O (31 mg, 0.14 mmol) were added at 0 °C and the mixture was stirred overnight at rt. Then the reaction was quenched with sat. aq. NH_4Cl, extracted with DCM (3 x). The combined organic layers were washed with sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/EtOAc 7:1 → 4:1) to give compound **21** (121 mg, 0.111 mmol, 95%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 – 7.12 (m, 33H), 7.12 – 7.04 (m, 2H), 5.47 – 5.37 (m, 1H), 5.23 (d, J = 6.8 Hz, 1H), 4.83 – 4.67 (m, 4H), 4.65 – 4.49 (m, 6H), 4.41 (d, J = 10.2 Hz, 4H), 4.27 (d, J = 12.2 Hz, 1H), 4.18 (s, 1H), 4.08 (t, J = 6.5 Hz, 1H), 4.00 (s, 1H, OH), 3.93 – 3.76 (m, 5H), 3.75 – 3.64 (m, 3H), 3.53 (ddd, J = 13.4, 10.2, 3.1 Hz, 2H), 3.42 (dd, J = 10.8, 1.9 Hz, 1H), 2.3

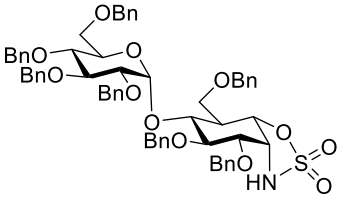
– 2.24 (m, 1H), 1.45 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.9, 138.7, 138.6, 138.2, 138.2, 138.0, 137.5, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.3, 128.1, 128.0, 128.0, 127.9, 127.9, 127.8, 127.7, 127.6, 127.6, 127.4, 127.3, 127.2, 97.2, 82.6, 80.2, 79.8, 79.8, 78.3, 77.8, 75.6, 75.0, 73.5, 73.5, 72.9, 72.8, 72.2, 71.8, 70.9, 68.9, 68.2, 67.1, 50.5, 45.2, 28.5. HRMS (ESI) m/z: $[M+H]^+$ calculated for $C_{67}H_{76}NO_{12}$ 1086.5362, found 1086.5355.

Compound 22



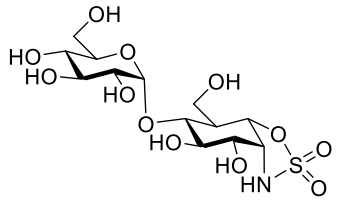
Compound **21** (121 mg, 0.111 mmol) was dissolved in dry DCM (2.0 mL) and purged with N_2. After cooling to 0 °C, Et_3N (163 µL, 1.17 mmol), imidazole (42 mg, 0.61 mmol) and thionyl chloride (49 µL, 0.66 mmol) were added successively. After stirring at 0 °C for 15 min, the reaction mixture was diluted with DCM, washed with diluted 1 M HCl solution, sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo* to give the crude sulfite as a pale-yellow oil. Then the crude was dissolved in a mixture of EtOAc/MeCN (1 mL/1 mL) and cooled to 0 °C. A solution of $RuCl_3.3H_2O$ (4 mg/mL stock solution in H_2O, 100 µL, 0.002 mmol) and $NaIO_4$ (35 mg, 0.16 mmol) in H_2O (0.4 mL) was added and the mixture stirred vigorously at 0 °C for 10 min. The reaction was quenched by addition of sat. aq. $Na_2S_2O_3$ (1 mL) and the mixture was stirred vigorously at rt for 15 min. Then the mixture was diluted with H_2O and extracted with EtOAc (2 x). The combined organic layers were washed with H_2O, brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/Acetone, 15:1 → 11:1) to afford compound **22** (115 mg, 100 µmol, 90%) as a clean oil. (*Note: oxidation with $RuCl_3$ and $NaIO_4$ should be monitored every 10 min by TLC analysis. The reaction should be quenched as soon as the conversion of the sulfite was complete. Otherwise an over-oxidized side product would be formed due to prolonged reaction times*). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.07 (m, 35H), 5.24 (dd, *J* = 9.2, 7.4 Hz, 1H), 5.19 (d, *J* = 3.5 Hz, 1H), 4.88 (d, *J* = 10.9 Hz, 1H), 4.79 (dd, *J* = 10.9, 5.5 Hz, 2H), 4.72 (d, *J* = 11.8 Hz, 1H), 4.66 – 4.37 (m, 8H), 4.35 (t, *J* = 4.8 Hz, 1H), 4.24 (t, *J* = 12.6 Hz, 2H), 4.13 (d, *J* = 11.3 Hz, 1H), 4.01 – 3.81 (m, 4H), 3.78 – 3.63 (m, 2H), 3.57 (ddd, *J* = 15.3, 9.7, 3.0 Hz, 2H), 3.37 (dd, *J* = 10.8, 2.6 Hz, 1H), 3.25 (dd, *J* = 11.0, 1.9 Hz, 1H), 2.93 (ddt, *J* = 11.6, 9.0, 2.3 Hz, 1H), 1.48 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 148.6, 138.8, 138.5, 138.3, 138.0, 137.8, 137.4, 137.0, 128.6, 128.5, 128.5, 128.4, 128.3, 128.2, 128.1, 128.0, 128.0, 128.0, 127.9, 127.9, 127.7, 127.7, 127.6, 127.6, 127.2, 95.0, 85.4, 82.2, 81.8, 79.8, 77.6, 75.6, 75.6, 75.0, 73.6, 73.4, 72.8, 72.8, 72.7, 71.2, 71.0, 70.9, 67.9, 64.8, 56.4, 40.4, 28.0. HRMS (ESI) m/z: $[M+NH_4]^+$ calculated for $C_{67}H_{77}N_2O_{14}S$ 1165.5090, found 1165.5089.

Compound 23



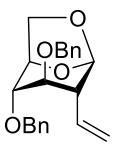
Compound **22** (105 mg, 91.5 µmol) was dissolved in dry DCM (4 mL). TFA (0.4 mL) was added and the mixture was stirred at rt for 2 h. Then the reaction was diluted with DCM, washed with sat. aq. $NaHCO_3$ (2 x), H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (Pentane/EtOAc, 11:1 → 7:1) to afford compound **23** (82 mg, 78 µmol, 86%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.16 (m, 31H), 7.12 (d, J = 6.7 Hz, 4H), 5.07 – 4.94 (m, 3H), 4.83 (dd, J = 22.3, 10.9 Hz, 2H), 4.69 (d, J = 11.1 Hz, 1H), 4.65 – 4.51 (m, 4H), 4.51 – 4.33 (m, 6H), 4.24 (dt, J = 11.4, 6.2 Hz, 1H), 4.17 – 3.98 (m, 4H), 3.87 (t, J = 10.0 Hz, 2H), 3.75 – 3.62 (m, 4H), 3.57 (dd, J = 9.9, 3.3 Hz, 1H), 3.49 (d, J = 10.5 Hz, 1H), 2.52 (t, J = 10.3 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.6, 138.3, 138.3, 137.9, 137.8, 137.1, 136.9, 128.8, 128.7, 128.6, 128.5, 128.5, 128.5, 128.5, 128.4, 128.3, 128.2, 128.1, 127.9, 127.9, 127.8, 127.7, 127.7, 127.6, 127.5, 127.4, 97.7, 82.0, 79.7, 79.1, 77.8, 75.4, 75.3, 75.1, 75.1, 74.5, 73.7, 73.3, 72.2, 71.9, 71.8, 71.5, 68.3, 65.6, 53.4, 42.1. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{62}H_{65}NO_{12}SNa$ 1070.4120, found 1070.4118.

Compound 24



Compound **23** (41 mg, 40 µmol) was dissolved in a mixture of MeOH/H_2O/dioxane (2/1/2, 4 mL) under Argon and $Pd(OH)_2$/C (20 wt%, 41 mg, 60 µmol) was added. While stirring vigorously, the mixture was flushed with a H_2 balloon. After stirring for 5 h under H_2 atmosphere, the mixture was filtered over a small celite pad and evaporated to afford the product in high purity as a white powder (15.8 mg, quant) after lyophilization. ^{1}H NMR (400 MHz, MeOD) δ 5.13 (d, J = 3.8 Hz, 1H, H1'), 4.95 (ddd, J = 10.3, 5.3, 1.2 Hz, 1H, H7), 4.36 (t, J = 5.0 Hz, 1H, H1), 4.06 (dd, J = 11.4, 2.3 Hz, 1H, H6a), 3.94 – 3.83 (m, 2H, H3 and H6'a), 3.78 (ddd, J = 9.9, 6.1, 2.2 Hz, 1H, H5'), 3.72 – 3.60 (m, 4H, H2, H6b, H3' and H6'b), 3.55 – 3.43 (m, 2H, H4 and H2'), 3.26 (t, J = 9.4 Hz, 1H, H4'), 2.27 – 2.16 (m, 1H, H5). ^{13}C NMR (101 MHz, MeOD) δ 103.0 (C1'), 81.5 (C7), 80.4 (C4), 75.0, 75.0, 74.8 (C3, C3' and C5'), 74.1 (C2'), 71.6 (C4'), 70.5 (C2), 62.8 (C6'), 60.8 (C1), 57.1 (C6), 45.7 (C5). HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{13}H_{23}NO_{12}SNa$ 440.0833, found 440.0832.

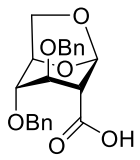
Compound 54



Compound **53** (2.49 g, 9.49 mmol) was dissolved in dry DMF (35 mL) and cooled to 0 °C. Then NaH (60% in mineral oil, 948 mg, 23.7 mmol) was added and the reaction mixture was stirred at 0 °C for 10 min and at rt for 20 min. After cooling to 0 °C again, BnBr (1.7 mL, 14.2 mmol) and TBAI (175 mg, 0.47 mmol) were added and the reaction mixture was stirred at rt for 2 h. The reaction was carefully quenched with H_2O at 0 °C, diluted with water and

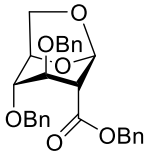
extracted with EtOAc (2 x). The combined organic layers were washed with water (2 x), brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 30:1 → 15:1) to obtain compound **54** (2.64 g, 7.50 mmol, 79%) as a light yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.39 – 7.24 (m, 10H), 5.95 (dt, J = 17.2, 9.9 Hz, 1H), 5.37 (s, 1H), 5.20 – 5.09 (m, 2H), 4.63 – 4.51 (m, 4H), 4.47 (t, J = 11.7 Hz, 1H), 4.13 (d, J = 7.0 Hz, 1H), 3.73 (t, J = 6.4 Hz, 1H), 3.51 (t, J = 1.5 Hz, 1H), 3.39 (d, J = 1.9 Hz, 1H), 2.59 (d, J = 9.4 Hz, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.1, 137.9, 136.6, 128.6, 128.5, 127.9, 127.8, 127.8, 127.7, 117.4, 103.2, 76.8, 76.5, 74.2, 71.7, 71.1, 65.0, 48.7 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{22}H_{24}O_4Na$ 375.1567, found 375.1565.

Compound 55



Compound **54** (0.56 g, 1.6 mmol) was dissolved in a mixture of EtOAc/MeCN (10 mL/10 mL) and cooled to 0 °C. A solution of $RuCl_3.3H_2O$ (33 mg, 0.16 mmol) and $NaIO_4$ (1.7 g, 8.0 mmol) in H_2O (15 mL) was added at 0 °C and the mixture was stirred vigorously at rt for 1.5 h. The phases were separated and the organic phase was extracted with EtOAc (2 x 15 mL). To the combined organic layers was added isopropanol (1.0 mL) and the mixture was stirred at rt for 1 h. After which the mixture was washed with H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (DCM/EtOAc 12:1 → 4:1) to afford compound **55** (336 mg, 0.908 mmol, 57%) as a pale yellow oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.47 (brs, 1H), 7.38 – 7.23 (m, 10H), 5.93 (s, 1H), 4.66 – 4.55 (m, 3H), 4.44 (dd, J = 12.2, 9.8 Hz, 2H), 4.23 (s, 1H), 4.13 (d, J = 7.1 Hz, 1H), 3.76 – 3.68 (m, 1H), 3.40 (s, 1H), 2.91 (s, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 174.6, 137.6, 137.4, 128.6, 128.0, 128.0, 127.8, 99.7, 75.5, 74.8, 72.8, 72.0, 70.9, 65.2, 48.6 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{21}H_{22}O_6Na$ 393.1309, found 393.1307.

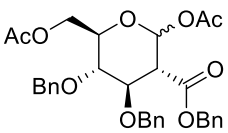
Compound 56



Compound **55** (336 mg, 0.908 mmol) was dissolved in dry DCM (9 mL) and cooled to 0 °C. Benzyl alcohol (281 µL, 2.72 mmol), *N*,*N*'-diisopropylcarbodiimide (168 µL, 1.09 mmol) and DMAP (33 mg, 0.27 mmol) were added at 0 °C and the mixture was stirred overnight at rt. The solvent was evaporated under reduced pressure and the resulting residue was dissolved with EtOAc, washed with diluted 1 M HCl solution, sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 13:1 → 7:1) to afford compound **56** (384 mg, 0.833 mmol, 92%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 – 7.22 (m, 15H), 5.96 (s, 1H), 5.23 – 5.09 (m, 2H), 4.55 (dd, J = 12.3, 3.3 Hz, 3H), 4.46 (d, J = 12.0 Hz, 1H), 4.38 (d, J = 12.3 Hz, 1H), 4.29 (s, 1H), 4.10 (d, J = 7.1 Hz, 1H), 3.71 (t, J = 6.4 Hz, 1H), 3.38 (s, 1H), 2.87 (s, 1H). ^{13}C NMR (101 MHz, $CDCl_3$)

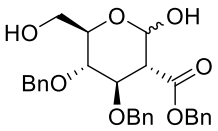
δ 169.6, 137.8, 137.7, 135.6, 128.6, 128.6, 128.5, 128.4, 128.3, 128.2, 127.9, 127.8, 127.8, 99.9, 76.0, 74.9, 73.0, 71.9, 70.8, 66.9, 65.3, 48.7 ppm.

Compound 57



TFA (230 µL, 3.01 mmol) was added to a cooled (0 °C) and stirred solution of **56** (252 mg, 0.547 mmol) in acetic anhydride (7.0 mL). After 5 h, the mixture was concentrated and the residue was co-evaporated with toluene (3 x 30 mL).The crude was purified with silica column chromatography (Pentane/acetone 15:1 → 7:1) to obtain compound **57** (α/β = 1:2, 286 mg, 0.508 mmol, 93%) as a colorless oil. ^{1}H NMR [α,β-anomers] (500 MHz, $CDCl_3$) *δ* 7.40 – 7.10 (m, 45H, Ar CH (α/β)), 6.38 (d, *J* = 3.7 Hz, 1H, H1 (α)), 5.80 (d, *J* = 9.0 Hz, 2H, H1 (β)), 5.18 – 5.03 (m, 6H, CH_2 ester (α,β)), 4.92 – 4.85 (m, 3H, CH*H* Bn (α/β)), 4.82 (d, *J* = 10.8 Hz, 2H, CH*H* Bn (α/β)), 4.73 (d, *J* = 10.9 Hz, 2H, CH*H* Bn (α/β)), 4.56 (td, *J* = 12.9, 10.8 Hz, 5H, CH*H* Bn (α/β)), 4.35 – 4.20 (m, 7H, H3 (α) and H6ab (α,β)), 4.06 (dd, *J* = 10.8, 8.7 Hz, 2H, H3 (β)), 3.94 (dt, *J* = 10.2, 3.2 Hz, 1H, H5 (α)), 3.69 (ddd, *J* = 9.8, 4.6, 2.2 Hz, 2H, H5 (β)), 3.61 (t, *J* = 9.6 Hz, 1H, H4 (α)), 3.55 (t, *J* = 9.3 Hz, 2H, H4 (β)), 3.07 (dd, *J* = 11.0, 3.7 Hz, 1H, H2 (α)), 2.93 (dd, *J* = 10.8, 9.0 Hz, 2H, H2 (β)), 2.03 (d, *J* = 3.5 Hz, 9H, CH_3 (α,β)), 1.90 (d, *J* = 5.8 Hz, 9H, CH_3 (α,β)). ^{13}C NMR [α,β-anomers] (126 MHz, $CDCl_3$) *δ* 170.7, 170.7, 169.2, 168.8, 168.7, 168.1 (6C=O (α,β)), 138.4, 137.8, 137.5, 137.5, 135.5, 135.4 (6C_q Ar (α,β)), 128.7, 128.7, 128.7, 128.7, 128.6, 128.6, 128.3, 128.3, 128.2, 128.1, 128.0, 127.9, 127.9 (CH Ar (α,β)), 92.3 (C1 (β)), 90.8 (C1 (α)), 81.3 (C3 (β)), 78.2 (C4 (α)), 77.8 (C3 (α)), 77.5 (C4 (β)), 75.7 (Bn (α)), 75.4 (Bn (α)), 75.2 (Bn (β)), 75.1 (Bn (β)), 74.1 (C5 (β)), 71.5 (C5 (α)), 67.2 (Bn ester (α)), 67.1 (Bn ester (β)), 62.7 (C6 (β)), 62.6 (C6 (α)), 54.0 (C2 (β)), 52.0 (C2 (α)), 21.0 (CH_3 (β)), 20.9 (CH_3 (α)), 20.7 (CH_3 (β)), 20.7 (CH_3 (α)) ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{32}H_{34}O_9Na$ 585.2095, found 585.2099.

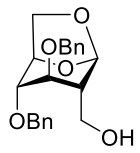
Compound 58



To a stirred solution of **57** (153 mg, 0.272 mmol) in THF/H_2O (9/1, 2.7 mL/0.3 mL) was added *p*-TsOH·H_2O (103 mg, 0.544 mmol) and the mixture was heated to 70 °C. After 7 h, the reaction mixture was cooled to rt and diluted with EtOAc, washed with sat. aq. $NaHCO_3$, H_2O and brine, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The crude was purified by silica gel column chromatography (pentane/EtOAc 3:1 → 1:1) to afford compound **58** (α/β ≈ 1:1,107 mg, 0.223 mmol, 82%) as a colorless oil. ^{1}H NMR [α,β-anomers] (500 MHz, $CDCl_3$) *δ* 7.38 – 7.20 (m, 26H), 7.20 – 7.08 (m, 4H, CH_2 ester (α,β)), 5.48 (t, *J* = 2.9 Hz, 1H, H1 (α)), 5.25 – 5.04 (m, 4H), 4.93 (d, *J* = 8.6 Hz, 1H, H1 (β)), 4.87 – 4.75 (m, 4H, CH*H* Bn (α/β)), 4.70 (d, *J* = 11.0 Hz, 1H, CH*H* Bn (α/β)), 4.62 (t, *J* = 11.2 Hz, 2H, CH*H* Bn (α/β)), 4.47 (d, *J* = 11.0 Hz, 1H, CH*H* Bn (α/β)), 4.34 (dd, *J* = 10.9, 9.1 Hz, 1H, H3 (α)), 4.00 (ddd, *J* = 16.5, 9.3, 5.6 Hz, 2H, H5 (α) and H3 (β)), 3.82 (ddd, *J* = 14.5, 12.0, 2.4 Hz, 2H, H6a (α,β)), 3.66 (ddd, *J* = 15.1, 7.6, 4.3 Hz, 2H, H6b (α,β)), 3.54 (t, *J* = 9.3 Hz, 1H, H4 (β)), 3.51 – 3.46 (m, 1H, H4 (α)), 3.43 (ddd, *J* = 9.8, 4.7, 2.3 Hz, 1H,

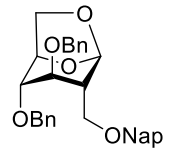
H5 (β)), 2.92 (dd, *J* = 10.9, 3.7 Hz, 1H, H2 (α)), 2.80 (ddd, *J* = 10.7, 8.5, 1.3 Hz, 1H, H2 (β)). ^{13}C NMR [α,β-anomers] (126 MHz, $CDCl_3$) *δ* 170.8, 169.8, 138.6, 138.0, 138.0, 137.9, 135.7, 135.4, 128.7, 128.7, 128.6, 128.6, 128.6, 128.5, 128.4, 128.4, 128.4, 128.4, 128.3, 128.2, 128.1, 127.9, 127.8, 127.7, 95.3, 91.8, 81.3, 79.3, 77.9, 77.7, 75.9, 75.5, 75.1, 75.1, 74.9, 71.6, 67.1, 67.0, 61.9, 61.6, 56.7, 53.5, 29.8 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{28}H_{30}O_7Na$ 501.1884, found 501.1883.

Compound 62



Compound **54** (1.36 g, 3.86 mmol) was dissolved in a mixture of dioxane/H_2O (5/2, 50 mL/20 mL). Osmium tetraoxide (0.1 M in H_2O, 2 mL, 0.2 mmol) and $NaIO_4$ (1.82 g, 8.50 mmol) were added and the reaction mixture was stirred at rt for 5 h. TLC-analysis indicated the presence of intermediate vicinal diol so more $NaIO_4$ (826 mg, 3.86 mmol) was added and the mixture was stirred at rt overnight. The mixture was then diluted with H_2O (50 mL) and extracted with EtOAc (3 x 80 mL). The combined organic layers were washed with brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The resulting residue was dissolved in a mixture of THF/isopropanol (9/1, 45 mL/5 mL). The solution was cooled to 0 °C and $NaBH_4$ (294 mg, 7.72 mmol) was added. After 1 h, the reaction was quenched with sat. aq. NH_4Cl (50 mL) and extracted with EtOAc (3 x 80 mL). The combined organic layers were dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 4:1 → 2:1) to obtain compound **62** (1.16 g, 3.26 mmol, 84%) as a light yellow oil. ^{1}H NMR (500 MHz, $CDCl_3$) *δ* 7.40 – 7.26 (m, 10H), 5.55 (d, *J* = 1.3 Hz, 1H), 4.61 – 4.49 (m, 4H), 4.46 (d, *J* = 12.2 Hz, 1H), 4.15 – 4.09 (m, 1H), 3.71 (pd, *J* = 7.7, 7.2, 4.0 Hz, 3H), 3.52 (p, *J* = 1.5 Hz, 1H), 3.42 (dq, *J* = 2.4, 1.1 Hz, 1H), 2.24 (d, *J* = 8.1 Hz, 1H), 2.11 – 2.04 (m, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) *δ* 138.1, 137.7, 128.6, 128.6, 128.1, 127.9, 127.9, 127.7, 101.6, 76.6, 74.4, 74.1, 71.5, 71.2, 64.7, 62.3, 45.6 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{21}H_{24}O_5Na$ 379.1516, found 379.1515.

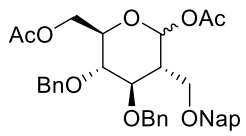
Compound S5



NaH (60% in mineral oil, 326 mg, 8.15 mmol) was added to a cooled (0 °C) and stirred solution of compound **62** (1.16 g, 3.26 mmol) in dry DMF (30 mL). NapBr (1.08 g, 4.89 mmol) was added and the reaction mixture was stirred at rt for 2.5 h. After which the reaction was carefully quenched with H_2O at 0 °C, diluted with water (60 mL) and extracted with EtOAc (3 x 50 mL). The combined organic layers were washed with water (2 x 100 mL), brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 15:1 → 7:1) to obtain compound **S5** (1.49 g, 3.00 mmol, 92%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.80 (dd, *J* = 9.1, 6.7 Hz, 3H), 7.72 (d, *J* = 1.7 Hz, 1H), 7.51 – 7.36 (m, 3H), 7.35 – 7.15 (m, 10H), 5.58 (s, 1H), 4.67 (d, *J* = 12.1 Hz, 1H), 4.60 – 4.49 (m, 3H), 4.48 – 4.38 (m, 3H), 4.16 – 4.07 (m, 1H), 3.71 (t, *J* = 6.5 Hz, 1H), 3.65 (dd, *J* = 9.3, 7.2 Hz, 1H), 3.56 – 3.47 (m, 2H), 3.35 (d, *J* = 1.9 Hz, 1H), 2.30 – 2.22 (m, 1H). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 138.2, 137.9,

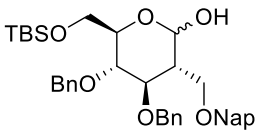
135.8, 133.3, 133.1, 128.6, 128.5, 128.2, 128.0, 127.9, 127.8, 127.8, 127.8, 127.7, 126.5, 126.2, 126.0, 125.9, 101.2, 76.9, 74.5, 73.4, 73.4, 71.3, 71.0, 69.2, 64.8, 43.8 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{32}H_{32}O_5Na$ 519.2142, found 519.2145.

Compound 78



TFA (1.26 mL, 16.4 mmol) was added to a cooled (0 °C) and stirred solution of **S5** (1.48 g, 2.98 mmol) in acetic anhydride (42 mL). After 45 min, the mixture was concentrated and the residue was co-evaporated with toluene (3 x 50 mL). The crude was purified with silica column chromatography (Pentane/EtOAc 11:1 → 4:1) to obtain compound **78** (α/β ≈ 2:1, 1.43 g, 2.39 mmol, 80%) as a white solid. ^{1}H NMR [α, β-anomers] (500 MHz, $CDCl_3$) δ 7.81 (ddd, *J* = 10.7, 8.6, 5.0 Hz, CH Ar (α,β)), 7.70 (d, *J* = 3.9 Hz, CH Ar (α,β)), 7.52 – 7.44 (m, CH Ar (α,β)), 7.41 – 7.16 (m, CH Ar (α,β)), 6.37 (d, *J* = 3.0 Hz, H1 (α)), 5.87 (d, *J* = 9.2 Hz, H1 (β)), 4.92 – 4.80 (m, CH*H* Bn (α,β)), 4.69 – 4.55 (m, CH*H* Bn (α,β)), 4.47 (dd, *J* = 20.0, 12.2 Hz, CH*H* Bn (α,β)), 4.34 – 4.24 (m, H6ab (α,β)), 3.98 (dd, *J* = 10.9, 8.8 Hz, H3 (β)), 3.90 (d, *J* = 9.8 Hz, H5 (α)), 3.79 (dd, *J* = 11.2, 8.8 Hz, H3 (α)), 3.73 – 3.59 (m, H7a (α), H7a (β), H5 (β) and H4 (α)), 3.56 (t, *J* = 9.3 Hz, H4 (β)), 3.47 – 3.37 (m, 2H, H7b (α) and H7b (β)), 2.40 – 2.31 (m, H2 (α)), 2.04 (d, *J* = 6.2 Hz, CH_3 (α,β)), 1.88 (d, *J* = 1.5 Hz, H2 (β) and CH_3 (α,β)). ^{13}C NMR [α,β-anomers] (126 MHz, $CDCl_3$) δ 170.9, 170.9, 169.2, 169.2, 138.3, 138.0, 137.8, 137.7, 135.5, 135.4, 133.3, 133.3, 133.1, 133.1, 128.7, 128.7, 128.6, 128.6, 128.4, 128.3, 128.2, 128.2, 128.1, 128.1, 128.0, 128.0, 127.9, 127.9, 127.8, 127.8, 126.8, 126.6, 126.3, 126.3, 126.1, 126.1, 126.0, 125.8, 92.2, 91.9, 79.2, 78.9, 78.8, 78.4, 75.5, 75.3, 75.3, 75.0, 73.7, 73.4, 73.3, 71.5, 66.5, 63.7, 63.2, 63.0, 47.3, 45.8, 21.0, 21.0, 21.0, 21.0 ppm. HRMS (ESI) m/z: $[M+Na]^+$ calculated for $C_{36}H_{38}O_8Na$ 621.2459, found 621.2457.

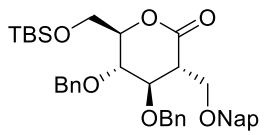
Compound 79



Sodium methoxide (4.373 M in MeOH, 165 µL, 0.72 mmol) was added to a solution of compound **78** (1.43 g, 2.39 mmol) in DCM/MeOH (1/1, 24 mL/24 mL). After 45 min, the reaction mixture was neutralized with washed Amberlite IR120 (H^+), filtered and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 3:1 → 1:1) to obtain the target product which was contaminated with an unknown byproduct. The mixture was directly dissolved in dry DCM (24 mL) and imidazole (0.41 g, 6.0 mmol) was added. After stirring at rt until imidazole completely dissolved, the reaction mixture was further stirred at 0 °C for another 1 h. Then *tert*-butyldimethylsilyl chloride (488 mg, 3.24 mmol) and DMAP (29 mg, 0.24 mmol) were added at 0 °C and the mixture was stirred at rt for 4 h. After which the mixture was diluted with DCM, washed with sat. aq. NH_4Cl, H_2O and brine, dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 15:1 → 5:1) to obtain compound **79** (α/β ≈ 2.5:1, 874 mg, 1.39 mmol, 57% over two steps) as a clean oil. ^{1}H NMR [α,β-anomers] (500 MHz, $CDCl_3$) δ 7.84 – 7.76 (m), 7.75 – 7.71 (m), 7.51

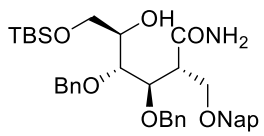
– 7.44 (m), 7.42 (ddd, *J* = 8.4, 3.8, 1.7 Hz), 7.36 – 7.25 (m), 7.24 – 7.15 (m), 7.12 (ddd, *J* = 6.6, 3.5, 1.6 Hz), 5.39 (t, *J* = 2.7 Hz, H1 (α)), 4.89 (dd, *J* = 8.5, 5.3 Hz, H1 (β)), 4.85 (dt, *J* = 10.9, 3.2 Hz), 4.75 (dd, *J* = 18.1, 10.9 Hz), 4.69 – 4.48 (m), 4.00 – 3.86 (m), 3.85 – 3.76 (m), 3.76 – 3.61 (m), 3.50 (d, *J* = 5.3 Hz), 3.36 (ddd, *J* = 9.5, 3.3, 2.4 Hz), 2.05 (ddt, *J* = 11.2, 6.1, 3.0 Hz, H2 (α)), 1.77 – 1.67 (m, H2 (β)), 0.90 (s, *t*-Bu (α,β)), 0.10 – 0.03 (m, CH_3 (α,β)). ^{13}C NMR [α,β-anomers] (126 MHz, $CDCl_3$) δ 138.8, 138.6, 138.5, 138.4, 135.6, 135.0, 133.3, 133.2, 133.1, 128.6, 128.6, 128.5, 128.5, 128.3, 128.1, 128.0, 128.0, 128.0, 128.0, 127.9, 127.9, 127.8, 127.8, 127.8, 126.9, 126.6, 126.4, 126.3, 126.2, 126.0, 125.9, 95.6, 93.9, 79.7, 79.4, 79.3, 77.4, 76.3, 75.4, 75.4, 74.9, 74.8, 73.7, 73.5, 72.3, 67.2, 65.9, 62.6, 62.3, 49.8, 46.8, 26.1, 18.5, 18.5, -4.8, -5.2 ppm. HRMS (ESI) m/z: $[M+NH_4]^+$ calculated for $C_{38}H_{52}O_6SiN$ 646.3558, found 646.3558.

Compound 80



Acetic anhydride (1.83 mL, 19.4 mmol) was added to a solution of compound **79** (553 mg, 0.881 mmol) in anhydrous DMSO (8.0 mL). The mixture was stirred at 30 °C for 16 h, after which it was diluted with Et_2O (150 mL), washed with aq. $NaHCO_3$ (10%, 3 x 15 mL) and brine (40 mL). The organic layer was dried over $MgSO_4$, filtrated and concentrated *in vacuo*. The crude was purified with silica column chromatography (Pentane/EtOAc 25:1 → 9:1) to obtain compound **80** (440 mg, 0.703 mmol, 79%) as a clean oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 (td, *J* = 5.4, 4.9, 2.6 Hz, 3H), 7.72 (s, 1H), 7.50 – 7.20 (m, 11H), 7.13 (dt, *J* = 7.6, 1.8 Hz, 2H), 4.91 – 4.75 (m, 3H), 4.68 (d, *J* = 12.1 Hz, 1H), 4.53 (dt, *J* = 11.3, 3.0 Hz, 2H), 4.20 – 3.95 (m, 4H), 3.90 (d, *J* = 1.8 Hz, 2H), 3.71 (dt, *J* = 8.9, 2.2 Hz, 1H), 2.65 (dt, *J* = 9.3, 2.4 Hz, 1H), 0.89 (d, *J* = 1.8 Hz, 9H), 0.08 (d, *J* = 1.8 Hz, 6H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.6, 138.0, 138.0, 135.4, 133.3, 133.1, 128.7, 128.6, 128.6, 128.3, 128.1, 128.1, 128.0, 128.0, 127.9, 127.8, 126.7, 126.2, 126.0, 126.0, 79.3, 76.8, 76.7, 74.9, 74.9, 73.6, 66.6, 61.4, 48.8, 26.0, 18.3, -5.1, -5.3 ppm. $[M+NH_4]^+$ calculated for $C_{38}H_{50}O_6SiN$ 644.3402, found 644.3400.

Compound 81



Compound **80** (432 mg, 0.690 mmol) was dissolved in methanolic ammonia (7 M, 7.0 mL). After stirring at rt for 1.5 h, the reaction mixture was concentrated and traces of ammonia were removed by co-evaporation with toluene (15 mL). The crude was purified with silica column chromatography (Pentane/EtOAc 4:1 → 1:1) to obtain product **81** (245 mg, 0.381 mmol, 55%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.89 – 7.67 (m, 4H), 7.56 – 7.36 (m, 3H), 7.37 – 7.21 (m, 10H), 6.43 (s, 1H), 5.36 (s, 1H), 4.74 – 4.52 (m, 6H), 4.38 (dd, *J* = 7.0, 3.0 Hz, 1H), 3.90 (tt, *J* = 7.4, 4.4 Hz, 3H), 3.77 (dd, *J* = 10.2, 3.5 Hz, 1H), 3.72 – 3.59 (m, 2H), 3.06 (dd, *J* = 9.3, 3.8 Hz, 1H), 2.75 (d, *J* = 5.9 Hz, 1H), 0.89 (s, 9H), 0.05 (s, 6H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 175.2, 138.4, 138.1, 135.3, 133.3, 133.1, 128.5, 128.4, 128.2,

128.2, 128.0, 127.9, 127.8, 127.8, 126.8, 126.3, 126.1, 125.9, 79.2, 77.6, 74.4, 74.0, 73.7, 71.4, 68.4, 64.1, 47.7, 26.0, 18.4, -5.2, -5.2 ppm. $[M+H]^+$ calculated for $C_{38}H_{50}NO_6Si$ 644.3402, found 644.3402.

6.5 References

[1] van der Maarel, M. J. E. C.; van der Veen, B.; Uitdehaag, J. C. M.; Leemhuis, H.; Dijkhuizen, L., *J. Biotech.* **2001**, *94*, 137-155.

[2] Caner, S.; Zhang, X.; Jiang, J.; Chen, H. M.; Nguyen, N. T.; Overkleeft, H.; Brayer, G. D.; Withers, S. G., *FEBS Lett.* **2016**, *590* (8), 1143-1151.

[3] Sundarram, A.; Murthy, T. P. K., *J. Appl. Environ. Microbiol.* **2014**, *2*, 166-175.

[4] Bertoft, E., *Agronomy* **2017**, *7* (3), 56-84.

[5] Artola, M.; Hedberg, C.; Rowland, R. J.; Raich, L.; Kytidou, K.; Wu, L.; Schaaf, A.; Ferraz, M. J.; van der Marel, G. A.; Codée, J. D. C.; Rovira, C.; Aerts, J. M. F. G.; Davies, G. J.; Overkleeft, H. S., *Chem. Sci.* **2019**, *10*, 9233-9243.

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

[7] Bordoni, V.; Porkolab, V.; Sattin, S.; Thépaut, M.; Frau, I.; Favero, L.; Crotti, P.; Bernardi, A.; Fieschi, F.; Di Bussolo, V., *RSC Adv.* **2016**, *6* (92), 89578-89584.

[8] Thadke, S. A.; Mishra, B.; Hotha, S., *J. Org. Chem.* **2014**, *79* (16), 7358-7371.

[9] Cai, J.; Hu, J.; Qin, C.; Li, L.; Shen, D.; Tian, G.; Zou, X.; Seeberger, P. H.; Yin, J., *Angew. Chem. Int. Ed. Engl.* **2020**, *59* (46), 20529-20537.

[10] Lichtenthaler, F. W.; Metz, T. W., *Tetrahcdron Lett.* **1997**, *38* (31), 5477-5480.

[11] Artola, M.; Kuo, C. L.; Lelieveld, L. T.; Rowland, R. J.; van der Marel, G. A.; Codee, J. D. C.; Boot, R. G.; Davies, G. J.; Aerts, J.; Overkleeft, H. S., *J. Am. Chem. Soc.* **2019**, *141* (10), 4214-4218.

[12] Umezawa, H.; Aoyagi, T.; Komiyama, H.; Morishima, H.; Hamada, M.; Takeuchi, T., *J. Antibiot.* **1974**, *27*, 963-969.

[13] Nishimura, Y.; Wang, W.; Kondo, S.; Aoyagi, T.; Umezawa, H., *J. Am. Chem. Soc.* **1988**, *110*, 7249-7250..

[14] Michikawa, M.; Ichinose, H.; Momma, M.; Biely, P.; Jongkees, S.; Yoshida, M.; Kotake, T.; Tsumuraya, Y.; Withers, S. G.; Fujimoto, Z.; Kaneko, S., *J. Biol. Chem.* **2012**, *287* (17), 14069-14077.

[15] Bohlmann, L.; Tredwell, G. D.; Yu, X.; Chang, C. W.; Haselhorst, T.; Winger, M.; Dyason, J. C.; Thomson, R. J.; Tiralongo, J.; Beacham, I. R.; Blanchard, H.; von Itzstein, M., *Nat. Chem. Biol.* **2015**, *11* (12), 955-957.

[16] Wu, L.; Viola, C. M.; Brzozowski, A. M.; Davies, G. J., *Nat. Struct. Mol. Biol.* **2015**, *22* (12), 1016-1022.

[17] Kondo, K., Adachi, H., Shitara, E., Kojima, F.; Nishimura, Y., *Bioorg. Med. Chem.* **2001**, *9*, 1091-1095.

[18] Nishimura, Y., *J. Antibiot.* **2009**, *62* (8), 407-423.

[19] Katavic, P. L.; Yong, K. W. L.; Herring, J. N.; Deseo, M. A.; Blanchfield, J. T.; Ferro, V.; Garson, M. J., Tetrahedron **2013**, *69* (37), 8074-8079.

[20] van den Berg, R. J.; Donker-Koopman, W.; van Boom, J. H.; Aerts, H. M.; Noort, D., *Bioorg. Med. Chem.* **2004**, *12* (5), 891-902.

[21] Lohse, A.; Bols, M., *Tetrahedron* **1997**, *53*, 6917-6924.

[22] Liu, H.; Liang, X.; Søhoel, H.; Bülow, A.; Bols, M., *J. Am. Chem. Soc.* **2001**, *123*, 5116-5117.

[23] Mohan, C. D.; Hari, S.; Preetham, H. D.; Rangappa, S.; Barash, U.; Ilan, N.; Nayak, S. C.; Gupta, V. K.; Basappa; Vlodavsky, I.; Rangappa, K. S., *iScience* **2019**, *15*, 360-390.

[24] Paulsen, H., *Liebigs Ann. Chem.* **1965**, *683*, 187-198.

[25] Liu, H.; Lillelund, V. H.; Andersch, J.; Liang, X.; Bols, M., *J. Carbohydr. Chem.* **2004**, *23* (4), 223-238.

[26] Albright, J. D.; Goldman, L., *J. Am. Chem. Soc.* **1967**, *89*, 2416-2423.

[27] Hembre, E. J.; Pearson, W. H., *Tetrahedron* **1997**, *53*, 11021-11032.

Nederlandse Samenvatting

Glycosidasen zijn wijdverspreide enzymen in alle uithoeken van de natuur en zijn belangrijke biokatalysatoren die de hydrolyse van glycoside-bindingen in oligo/polysacchariden, glycoproteïnen en glycolipiden met een enorme efficiëntie katalyseren. Abnormale glycosidase-activiteit is nauw verbonden met een verscheidenheid aan menselijke ziekten. Neem bijvoorbeeld de overexpressie van heparanase die betrokken is bij bijna alle onderzochte soorten kanker. Deze overexpressie correleert met verhoogde tumorgrootte, tumor angiogenese, een toename in uitzaaiingen en een slechte prognose voor de patiënt. Specifieke remmers van glycosidasen zijn van grote waarde. Niet alleen omdat ze kunnen dienen als nuttige biologische hulpmiddelen om de katalytische machinerie, het mechanisme en zijn verschillende stadia te bestuderen door analyse van de kristalstructuur van (covalente) remmer-enzym complexen, maar ook omdat ze kunnen dienen als aanknopingspunten voor de ontwikkeling van geneesmiddelen voor de behandeling van glycosidase-gemedieerde ziekten. Bovendien is aangetoond dat covalent bindende remmers gebruikt kunnen worden als basis voor de ontwikkeling van activity-based probes (ABPs) die het indentificeren van glycosidasen in complexe biologische systemen mogelijk maken.

Dit proefschrift beschrijft de ontwikkeling en biochemische evaluatie van covalente remmers en ABPs voor stereochemie behoudende *endo*- en *exo*-glycosidasen waaronder zetmeel afbrekende enzymen en humaan lysosomaal β-glucocerebrosidase (GBA), als ook de synthese van een groep uronzuur achtige 1-*N*-iminosuikers als potentiële competitieve heparanase remmers. Het ontwerp van dergelijke covalente en competitieve glycosidase remmers is afhankelijk van het begrijpen van de functie en het mechanisme van deze enzymen. **Hoofdstuk 1** introduceert de algemene mechanismen die gebruikt worden door stereochemie behoudende en inverterende β-glycosidasen voor de hydrolyse van hun substraten. De verschillende stappen en tussenproducten van stereochemie behoudende α- en β-glucosidasen worden beschreven en het ontwerp van covalente cyclophellitol-achtige remmers op basis van deze routes wordt besproken. Daarnaast wordt het ontwerp van competitieve remmers, die de mechanistische overgangstoestanden nabootsen, beschreven. Tenslotte wordt een overzicht gegeven van de activity-based protein profiling (ABPP) workflow.

De biologische afbraak van zetmeel vereist de synergetische werking van een reeks enzymen. Van deze enzymen zijn de α-amylasen, die de hydrolyse van de interne α-1,4-glucosidebindingen katalyseren het meest uitgebreid bestudeerd, zowel binnen de geneeskunde als in de biotechnologie. **Hoofdstuk 2** rapporteert over het ontwerp en de synthese van een selectie 1,6-*epi*-cyclophellitol-gebaseerde pseudo-disachariden uitgerust met een reeks van reporter entiteiten en hun gebruik in ABPP van stereochemie behoudende amylasen in complexe biologische monsters. De activiteit en efficiëntie van de gesynthetiseerde remmers

en sondes werden aangetoond door uitgebreide biochemische analyse. De selectiviteit voor amylasen boven verwante endoglycosidasen werd gevalideerd door structurele studies. De maltobiose *epi*-cyclophellitol ABPs bleken effectief α-amylasen te kunnen labelen in menselijk speeksel, muizenweefsel en schimmel secretomen op een concentratie-, pH-, tijd- en temperatuurafhankelijke wijze. De labeling kon beconcurreerd worden met de niet-gemerkte remmers, evenals met de commercieel verkrijgbare concurrerende remmer acarbose. Deze ABPs kunnen worden gebruikt bij de ontdekking van nieuwe en doeltreffende humane amylase remmers. Deze remmers zouden kunnen bijdragen aan de behandeling van type 2-diabetes en bij de ontdekking van nieuwe microbiële amylasen met potentieel voordelige eigenschappen voor biotechnologische toepassingen.

Zetmeelpolysachariden zijn samengesteld uit lineaire amylose en vertakte amylopectine, waarvan de laatstgenoemde een belangrijk bestanddeel van standaard zetmeel vormt. In amylopectine zijn de vertakkingspunten die α-1,6-glucosidebindingen bevatten resistent tegen hydrolyse door α-1,4-specifieke amylasen. Als een uitbreiding van de maltobiose *epi*-cyclophellitol sondes behandeld in Hoofdstuk 2, beschrijft **Hoofdstuk 3** de synthese van een set glucose-isomaltose (GIM) en isomaltose-glucose (IMG) geconfigureerde *epi*-cyclophellitol sondes die de vertakte delen van de amylopectine structuur nabootsen. De centraal staande pseudo-trisachariden werden geconstrueerd door stereoselectieve α-1,4- en α-1,6-glycosyleringen van cyclohexeen acceptoren onder de juiste preactivatie condities. Hierna werd de epoxide functionaliteit aangebracht via een stereoselectieve epoxidatie, die werd gevolgd door globale ontscherming en een amide-koppeling met een reportertag om de uiteindelijke chemische sondes te verkrijgen na HPLC-zuivering. De reeks vertakte sondes zal bruikbaar zijn voor de detectie van industrieel relevante zetmeel-afbrekende enzymen die een voorkeur vertonen voor vertakte amylopectine achtige polysachariden.

Hoofdstuk 4 beschrijft de synthese van een set β-D-*gluco*-cyclophellitol aziridine remmers en ABPs, die gefunctionaliseerd zijn op zowel de C6 positie als de endocyclische stikstof van de aziridine. Röntgen kristallografische analyse toonde aan dat recombinant humaan β-glucocerebrosidase (rhGBA) de twee functionaliteiten kon accommoderen. De selectiviteit en potentie van deze bifunctionele aziridines voor GBA werden zowel *in vitro* als *in situ* onderzocht. De IC_{50} waarden voor de remming van rhGBA toonden aan dat de nieuwe bifunctionele cyclophellitol aziridine verbindingen ongeveer 10-15 keer minder potent waren dan de eerder gerapporteerde C6-mono-functionele cyclophellitol epoxides. Bovendien vertoonde incubatie van muis hersenlysaten met de bifunctionele verbindingen inderdaad een selectief markeringspatroon voor GBA boven GBA2, terwijl een verminderde selectiviteit werd waargenomen bij *in situ* markering van HEK293T cellen die endogeen GBA en tot overexpressie gebracht GBA2 bevatten. Hoewel de bifunctionele aziridines minder selectief en potent zijn dan hun C6-monogefunctionaliseerde epoxide tegenhangers, blijven ze nanomolair

inactiverende stoffen voor GBA die gebruikt kunnen worden voor de studie van GBA in relatie tot de ziekte van Gaucher.

Siastatine B is eerder gerapporteerd als een krachtige en effectieve remmer van β-D-glucuronidasen. De moleculaire structuur van siastatine B blijkt echter te volumineus om te passen in het katalytische centrum van een glucuronidase. NMR-analyse van 2-trifluoraceetamide bevattende siastatine B derivaten suggereerde dat oplosmiddel gemedieerde afbraak van deze moleculen snel kan plaatsvinden, waarbij een hemiaminal/hydraat keton vrijkomt dat als de echte remmer fungeert. Deze omlegging is echter niet aangetoond voor siastatine B. In **Hoofdstuk 5** wordt een röntgen-kristallografische analyse gepresenteerd van co-complexen tussen siastatine B en verschillende *exo*- en *endo*-β-glucuronidases, die bevestigen dat het de hemiaminale of het gehydrateerde keton is en niet siastatine B, die zich bindt in het katalytische centrum van het enzym. Om de werking van de afbraakproducten te begrijpen, werd een selectie van *galacto*- en *gluco*-geconfigureerde 1-*N*-iminosuikerderivaten gesynthetiseerd en werd een voorlopige structurele analyse voor enzym inhibitie door de synthetische *gluco*-geconfigureerde iminosuikers onderzocht. Om de remmende eigenschappen van deze verbindingen grondig te evalueren, kunnen in de toekomst kinetische studies en competitieve ABPP-experimenten worden uitgevoerd.

Tenslotte wordt in **Hoofdstuk 6** een meer gedetailleerde samenvatting gegeven van de, in dit proefschrift beschreven bevindingen, in combinatie met voorstellen voor vervolgonderzoek.

Summary in Chinese

中文总结

本论文对包含淀粉降解酶和溶酶体葡糖神经酰胺酶（GBA）在内的保留型内切/外切糖苷酶共价抑制剂和活性分子探针（ABPs）的设计，合成与生物评价进行了描述，同时还介绍了一系列糖醛酸型 1-*N*-亚氨糖的合成并对其作为潜在的竞争性乙酰肝素酶抑制剂进行了研究。这些共价和竞争性糖苷酶抑制剂的设计依赖于对相应酶功能和作用机制的理解。**第一章**简要介绍了保留型/反转型糖苷酶催化水解相应底物所采用的分子机制，同时描述了保留型 α-和 β-葡萄糖苷酶的反应路径，并探讨了基于这些路径的六元环多醇类共价抑制剂的设计。此外，本章节还介绍了基于底物过渡态模拟的竞争性抑制剂的设计并对活性蛋白表达谱（ABPP）的作用流程进行了概述。

淀粉的生物降解需要一系列酶的协同作用，其中，催化糖链内部α-1,4-糖苷键水解的保留型α-淀粉酶在生物医学和生物工程中都得到了最广泛的研究。**第二章**介绍了麦芽二糖构型的六元环多醇类α-淀粉酶抑制剂和 ABPs 的合成。该合成的关键步骤在于使用对酸耐受的环己烯或差像六元环多醇环碳酸酯作为受体，在适当的预活化条件下实现立体选择性的α-1,4-糖苷化，随后通过一系列化学转化获得所需的环氧，氮丙啶和环状硫酸酯官能团。随后，本章通过大量的生物化学实验对所合成抑制剂和 ABPs 的活性和有效性进行了研究。与环氧类 ABP 相比，氮丙啶类 ABP 在对重组人类唾液α-淀粉酶的标记中表现出较低的活性。这一结果与对米曲霉淀粉酶（Taka-amylase）的动力学研究中氮丙啶类抑制剂表现出最慢的抑制速率结果一致—而构象限制为可能的原因。进一步研究结果显示，环氧类 ABP 能够以浓度、温度、时间和 pH 依赖型的方式有效地标记人类唾液、小鼠组织和真菌分泌组中的α-淀粉酶，并且这一标记可以被同源的共价抑制剂以及市售的竞争性抑制剂阿卡波糖所抑制。此外，本章还利用生物素化环氧 ABP 对人类唾液和真菌分泌组中的α-淀粉酶进行了蛋白质组学分析。

淀粉多糖由直链淀粉和支链淀粉组成，其中后者是常规淀粉的主要组成成分。支链淀粉中α-1,6-糖苷键分支点附近的α-1,4-糖苷键对特异性的α-1,4-淀粉酶的水解具有抗性。因此，在第二章所研究的麦芽二糖构型六元环多醇环氧化物 ABPs 的基础上，**第三章**进一步描述了一系列葡萄糖-异麦芽糖 (GIM) 和异麦芽糖-葡萄糖 (IMG) 构型的六元环多醇环氧化物 ABPs 的合成。其中，假三糖骨架的构建是通过在适当的预活化条件下对环己烯受体进行立体选择性的α-1,4-和α-1,6-糖苷化来实现的。随后，通过立体选择性环氧化，苄基脱保护以及与相应的标记基团进行酰胺偶联，最后经过 HPLC 纯化即可获得最终探针。未来的研究将对这类新合成的 ABPs 在复杂生物样品中对淀粉降解酶的标记模式和效率进行探索并将所获得的数据与之前报道的麦芽二糖构型的 ABPs 数据进行比较。通过模拟支链淀粉的分支结构，这些 IMG 和 GIM 构型的 ABPs 有望对偏

爱支链底物的淀粉降解酶展现出特异性,进而将可用于与工业相关的降解支链淀粉型多糖的淀粉降解酶的检测与发现。

第四章描述了一系列β-D-葡萄糖构型的六元环多醇氮丙啶抑制剂和 ABPs 的合成，其特点是在它们的 C6 位和氮丙啶氮上都进行了官能团化。X 射线晶体衍射结果显示重组人类 GBA（rhGBA）能够在结构上有效地容纳这两个官能团。随后，本章节通过一系列体外和原位实验研究了这些双功能氮丙啶对 GBA 的选择性和有效性.IC_{50} 测试结果表明，这类新的双功能六元环多醇氮丙啶抑制剂对 rhGBA 的抑制效力比先前所报道的 C6-单官能团化六元环多醇环氧化物的效力低约 10-15 倍。此外，这类双功能探针虽然能够在体外对小鼠大脑裂解物中的 GBA 进行选择性标记，而在对含有内源性 GBA 和过表达非溶酶体葡糖神经酰胺酶（GBA2）的 HEK293T 细胞的原位标记中，其 GBA 选择性则有所下降。虽然双功能氮丙啶的选择性和有效性均低于相应的单功能环氧化物，但它们仍然是极具效力的 GBA 失活剂，可用于与高雪氏病（Gaucher disease）相关的 GBA 代谢研究。

Siastatin B 是一种从链霉菌培养物中分离而来的天然产物，此前已有报道它是一种有效的β-D-葡萄糖醛酸酶抑制剂。然而，siastatin B 的分子结构对于经典的保留型糖苷酶活性位点来说似乎过于巨大。Nishimura 课题组通过对含有三氟乙酰胺基的 siastatin B 衍生物进行 NMR 分析，发现这类三氟乙酰衍生物在溶剂中可迅速发生分解，从而释放出一类半胺醛/水合酮化合物作为真正的糖醛酸酶抑制剂。然而，这一重排现象尚未在 siastatin B 中得到证实。**第五章**通过对 siastatin B 和一系列糖醛酸酶，包括人类乙酰肝素酶(HPSE)、荚膜菌β-葡萄糖醛酸酶(AcGH79)、假单胞菌乙酰肝素酶(BpHEP)和大肠杆菌β-葡萄糖醛酸酶(EcGusB)的共结晶和 X 射线晶体衍射研究，证实了与这些酶活性位点进行共价结合的不是 siastatin B，而是其半胺醛或水合酮降解物。为了了解这类降解产物的作用模式，本章节合成了一系列半乳糖醛酸和葡萄糖醛酸构型的 1-*N*-亚氨糖衍生物。随后，对合成的葡萄糖醛酸构型亚氨糖与酶相互作用的初步结构研究表明，这些合成亚氨糖可能是潜在的 HPSE 和β-D-葡萄糖醛酸酶抑制剂。在未来的研究中，可通过动力学和竞争性 ABPP 实验对这些化合物的抑制特性进行深入评估。

第六章对本论文的研究成果进行了归纳和总结，并对未来的研究工作进行了展望。其中，描述了麦芽二糖构型的六元环多醇环状硫酸酯 ABPs 的合成，为α-淀粉酶设计/合成了麦芽二糖构型的竞争性抑制剂和耐受外切酶水解的稳定性共价抑制剂，最后为半胺醛类亚胺糖和双羧基类亚胺糖的合成提供了思路。

List of publications

Synthesis and biochemical evaluation of novel uronic acid-type 1-*N*-iminosugars derived from siastatin B as competitive heparanase inhibitors

Chen, Y.; van den Nieuwendijk, A. M. C. H.; Armstrong, Z.; Wu, L.; Overkleeft, H. S. and Davies, G. J., *manuscript in preparation.*

Design, synthesis and structural analysis of glucocerebrosidase imaging agents

Rowland, R. J.;* **Chen, Y.**;* Breen, I.; Wu, L.; Offen, W. A.; Beenakker, T. J. M.; Su, Q.; van den Nieuwendijk, A. M. C. H.; Aerts, J. M. F. G.; Artola, M.; Overkleeft, H. S. and Davies, G. J., *Chem. Eur. J.* **2021**, *27*, 1-13.

Activity-based protein profiling of retaining α-amylases in complex biological samples

Chen, Y.;* Armstrong, Z.;* Artola, M.; Florea, B. I.; Kuo, C.-L.; de Boer, C.; Rasmussen, M. S.; Abou Hachem, M.; van der Marel, G. A.; Codée, J. D. C.; Aerts, J. M. F. G.; Davies, G. J.; and Overkleeft, H. S., *J. Am. Chem. Soc.* **2021**, *143*, 2423-2432.

Direct stereoselective aziridination of cyclohexenols with 3-amino-2-(trifluoromethyl)quinazolin-4(3*H*)-one in the synthesis of cyclitol aziridine glycosidase inhibitors

Artola, M.; Wouters, S.; Schröder, S. P.; de Boer C.; **Chen, Y.**; Petracca, R.; van den Nieuwendijk, A. M. C. H.; Aerts, J. M. F. G.; van der Marel, G. A.; Codée, J. D. C. and Overkleeft, H. S., *Eur. J. Org. Chem.* **2019**, *6*, 1397-1404.

Interrupted Morita–Baylis–Hillman-type reaction of α-substituted activated olefins

Gu, J.; Xiao, B.-X.; **Chen, Y.-R.**; Li, Q.-Z.; Ouyang, Q.; Du, W.; Chen, Y.-C., *Org. Lett.* **2018**, *20*, 2088-2091.

Regioselective asymmetric formal (3+2) cycloadditions of nitrone ylides from isatins and enals

Chen, Y.-R.; Zhan, G.; Du, W.; Chen, Y.-C., *Adv. Synth. Catal.* **2016**, *358*, 3759-3764.

Asymmetric Diels-Alder and cascade reaction of quinone imine ketals and 2,4-dienals: construction of chiral benzo[*de*]quinolone derivatives

Gu, J.; Xiao, B.-X.; **Chen, Y.-R.**; Du, W.; Chen, Y.-C., *Adv. Synth. Catal.* **2016**, *358*, 296-302.

** Shared first-authorship*

Curriculum Vitae

Yurong Chen was born on 7th November 1992 in Kunming, Yunnan, China. In 2010 she graduated from Kunming No.1 High School and commenced her Bachelor studies in Pharmacy at West China School of Pharmacy at Sichuan University, Sichuan, China. During this period she became interested in Medicinal Chemistry and performed a six-month research internship on the chemical synthesis of indole derivatives and α,β-unsaturated aldehydes in Prof. dr. Ying-Chun Chen's group. In 2014, she obtained her BSc degree and was awarded "Outstanding Graduate of Sichuan University".

In 2014, she started her Master studies in Medicinal Chemistry at Sichuan University. Here she studied asymmetric catalysis for the construction of biologically active molecules under the supervision of Qing-Zhu Li, Dr. Wei Du and Prof. dr. Ying-Chun Chen. Her master thesis was entitled: "Regioselective asymmetric formal (3+2) cycloadditions of nitrone ylides from isatins and enals" and this work was published in *Advanced Synthesis & Catalysis*. During her Master's education, she was awarded the "Second Class Academic Scholarship" of Sichuan University for three consecutive years.

After obtaining her MSc degree in 2017, she was awarded a scholarship from the Chinese Scholarship Council (CSC) to start her doctoral research in the Bio-organic Synthesis group at Leiden University under the supervision of Prof. dr. Hermen Overkleeft, Prof. dr. Jeroen Codée and Dr. Marta Artola. Her PhD study focused on the development of activity-based probes and inhibitors for the study of several retaining glycosidases. Parts of the research described in this Thesis were presented on posters at CHAINS (2018 and 2019, Veldhoven, the Netherlands), the Annual ABPP conference (2019, Leuven, Belgium), the European Carbohydrate Symposium (2019, Leiden, the Netherlands) and Reedijk Symposium (2019, Leiden, the Netherlands).