

This is a repository copy of *Bicyclic picomolar OGA inhibitors enable chemoproteomic mapping of its endogenous post-translational modifications*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/182086/>

Version: Accepted Version

---

**Article:**

González-Cuest, Manuel, Sidhu, Peter, Ashmus, Roger A. et al. (10 more authors) (2022) Bicyclic picomolar OGA inhibitors enable chemoproteomic mapping of its endogenous post-translational modifications. *Journal of the American Chemical Society*. 832–844. ISSN 1520-5126

<https://doi.org/10.1021/jacs.1c10504>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

1 **Bicyclic picomolar OGA inhibitors enable chemoproteomic mapping of its endogenous post-**  
2 **translational modifications.**

3 Manuel González-Cuesta,<sup>a</sup> Peter Sidhu,<sup>b,c</sup> Roger A. Ashmus,<sup>b</sup> Alexandra Males,<sup>d</sup> Cameron Proceviat,<sup>b</sup> Zarina  
4 Madden,<sup>b</sup> Jason C. Rogalski,<sup>c</sup> Jil A. Busmann,<sup>e</sup> Leonard J. Foster,<sup>c</sup> José M. García Fernández,<sup>f</sup> Gideon J. Davies,<sup>d,\*</sup>  
5 Carmen Ortiz Mellet,<sup>a,\*</sup> David J. Vocadlo<sup>b,e,\*</sup>  
6

7 <sup>a</sup> Departamento de Química Orgánica, Facultad de Química, Universidad de Sevilla, 41012, Sevilla, Spain.

8 <sup>b</sup> Department of Chemistry, Simon Fraser University, Burnaby V5A 1S6, British Columbia, Canada.

9 <sup>c</sup> Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia,  
10 V6T 1Z4, Canada.

11 <sup>d</sup> Department of Chemistry, University of York, York YO10 5DD, United Kingdom.

12 <sup>e</sup> Department of Molecular Biology and Biochemistry, Simon Fraser University, V5A 1S6, Burnaby, British Columbia,  
13 Canada.

14 <sup>f</sup> Instituto de Investigaciones Químicas (IIQ), CSIC - Universidad de Sevilla, 41092 Sevilla, Spain.  
15  
16

17 **ABSTRACT:**

18 Owing to its roles in human health and disease, the modification of nuclear, cytoplasmic, and  
19 mitochondrial proteins with O-linked N-acetylglucosamine residues (O-GlcNAc) has emerged as a topic of  
20 great interest. Despite the presence of O-GlcNAc on hundreds of proteins within cells, only two enzymes  
21 regulate this modification. One of these enzymes is O-GlcNAcase (OGA), a dimeric glycoside hydrolase  
22 that has a deep active site cleft in which diverse substrates are accommodated. Chemical tools to control  
23 OGA are emerging as essential resources for helping to decode the biochemical and cellular functions of  
24 the O-GlcNAc pathway. Here we describe rationally designed bicyclic thiazolidine inhibitors that exhibit  
25 superb selectivity and picomolar inhibition of human OGA. Structures of these inhibitors in complex with  
26 human OGA reveal the basis for their exceptional potency and show that they extend out of the enzyme  
27 active site cleft. Leveraging this structure, we create a high affinity chemoproteomic probe that enables  
28 simple one-step purification of endogenous OGA from brain and targeted proteomic mapping of its post-  
29 translational modifications. These data uncover a range of new modifications including some that are less-  
30 known, such as O-ubiquitination and N-formylation. We expect that these inhibitors and  
31 chemoproteomics probes will prove useful as fundamental tools to decipher the mechanisms by which  
32 OGA is regulated and directed to its diverse cellular substrates. Moreover, the inhibitors and structures  
33 described here lay out a blueprint that will enable the creation of chemical probes and tools to interrogate  
34 OGA and other carbohydrate active enzymes.  
35

36 **KEYWORDS:**

37 O-GlcNAc, glycoside hydrolase, inhibitor, affinity probe, chemoproteomics, post-translational  
38 modifications,  
39

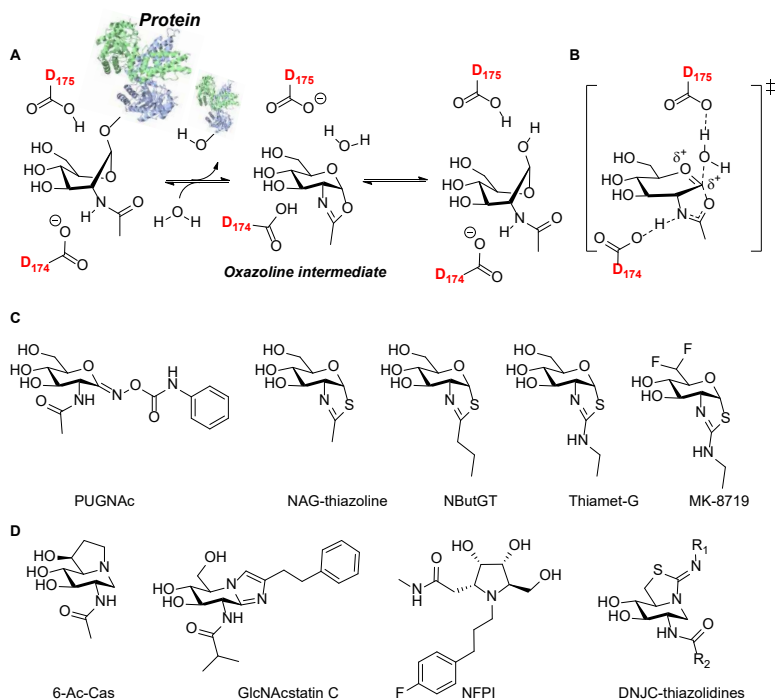
## 40 INTRODUCTION:

41 The incorporation of single N-acetylglucosamine (GlcNAc) residues O-linked to serine and  
42 threonine residues of nuclear and cytoplasmic proteins (O-GlcNAcylation)<sup>1</sup> is a widespread modification  
43 conserved among all multicellular eukaryotes. Only two enzymes regulate this process; the  
44 glycosyltransferase O-GlcNAc transferase (OGT)<sup>2</sup> acts to install O-GlcNAc and the glycoside hydrolase  
45 (OGA)<sup>3</sup> acts to cleave this modification off from proteins.<sup>4</sup> Levels of O-GlcNAc vary depending on nutrient  
46 availability, and even at basal levels O-GlcNAc is notably abundant, with well over a thousand proteins  
47 modified.<sup>5</sup> In line with its extensive distribution, O-GlcNAc has been shown to play fundamental roles on  
48 proteins including, for example, blocking ubiquitin-mediated degradation<sup>6</sup> and controlling protein  
49 subcellular localization<sup>7</sup>. Within cells, O-GlcNAc coordinates adaptive responses including helping to cope  
50 with diverse stresses<sup>8</sup> and regulating transcription.<sup>9</sup> In accord with its biochemical and physiological roles,  
51 enhancing O-GlcNAc levels has been shown to be protective in various animal models of disease. Within  
52 brain in particular, increased O-GlcNAc levels are protective against ischemic injuries<sup>10,11</sup> and various  
53 neurodegenerative disorders.<sup>12-15</sup>

54 Given the importance of O-GlcNAc in the vital processes of organisms, the molecular mechanisms  
55 by which O-GlcNAc is regulated by its cycling enzymes and how it exerts its molecular roles in controlling  
56 cellular processes are of great interest.<sup>4</sup> Considerable effort is being directed toward the development of  
57 chemical biology tools<sup>16,17</sup> to aid in understanding the O-GlcNAc cycling pathway from the molecular to  
58 the organismal level. Prominent among these chemical tools are inhibitors that have afforded structural  
59 insights into the molecular details of the O-GlcNAc regulatory enzymes.<sup>18-24</sup> OGA is a two-domain enzyme  
60 comprising an inactive acetyltransferase domain and a glycoside hydrolase CAZy family 84 (GH84) domain.  
61 Inhibitors binding to the active site of this GH84 domain of human OGA enable control over the levels of  
62 O-GlcNAc in cells and tissues and have gained significant attention. Some of these have been advanced as  
63 tool compounds<sup>20,25</sup> and even entered into the clinic.<sup>15</sup> Among known OGA inhibitors are polyhydroxylated  
64 compounds that have the basic carbon skeleton shared by GlcNAc. These carbohydrate-inspired inhibitors  
65 are generally thought to derive their potency against OGA by virtue of their resemblance to the transition  
66 state structures, or tightly bound intermediate,<sup>20,26,27</sup> found in the reaction mechanism of the OGA-  
67 catalyzed hydrolysis of  $\beta$ -N-acetylglucosaminides.

68 The catalytic mechanism used by OGA involves two distinct chemical steps (Figure 1) leading to  
69 hydrolysis with retention of configuration at the anomeric center. Experimental<sup>28,29</sup> and modelling<sup>30</sup>  
70 studies show that this reaction mechanism involves nucleophilic participation of the 2-acetamido group  
71 of the substrate to form a transient and tightly bound oxazoline intermediate. Such studies have also  
72 shown that the transition states flanking this bicyclic intermediate bear significant delocalized positive  
73 charge at both the anomeric center and endocyclic oxygen. This charge delocalization is enabled by a  
74 flattening of the pyranose ring to adopt a <sup>4</sup>H<sub>3</sub>/<sup>4</sup>E conformation (Figure 1). Accordingly, potent  
75 carbohydrate-based inhibitors of OGA typically manifest one or more of these features. Among  
76 carbohydrate-based OGA inhibitors is 1,2-dideoxy-2'-methyl- $\alpha$ -D-glucopyranoso[2,1-d]- $\Delta$ 2'-thiazoline  
77 (NAG-thiazoline, Figure 1,  $K_i = 70$  nM)<sup>28</sup>, which resembles the oxazoline intermediate found in the reaction  
78 pathway. Building on these thiazolines, aminothiazolines that are more hydrolytically stable have been  
79 generated, including 1,2-dideoxy-2'-ethylamino- $\alpha$ -D-glucopyranoso-[2,1-d]- $\Delta$ 2'-thiazoline (Thiamet-G,  $K_i =$   
80 2.1 nM). The increased basicity of this aminothiazoline as compared to the analogous thiazoline, favors  
81 its protonation at physiological pH values and enables formation of a favorable ionic interaction with the  
82 key catalytic residue Asp<sup>174, 26</sup>.

83 Other potent carbohydrate-based inhibitors of OGA are known, including *O*-(2-acetamido-2-  
 84 deoxy-D-glucopyranosylidene)-amino-*N*-phenylcarbamate (PUGNAc,  $K_i = 46$  nM, Figure 1).<sup>3</sup> PUGNAc has a  
 85  $sp^2$ -hybridized carbon at the pseudo-anomeric center and a  ${}^4E$  conformation resembling the OGA  
 86 transition states (Figure 1). GlcNAcstatin C ( $K_i = 4$  nM, Figure 1) is a potent analog of the natural product  
 87 NAGstatin that manifests this same  ${}^4E$  conformation and bears a pendent group that mimics the aglycone  
 88 portion of a glycoside. Notably, such compounds have been modified to make them of value as  
 89 fluorescence polarization probes for use in screening for new inhibitors of OGA.<sup>24</sup> Iminosugars that bear  
 90 a formal positive charge at the nitrogen atom that occupies the position of the endocyclic ring oxygen in  
 91 GlcNAc have also demonstrated activity against OGA, albeit with notably lower potency. 2-Acetamido-  
 92 1,2-dideoxy-nojirimycin (DNJNAc), for example, is moderately active ( $K_i = 23$   $\mu$ M)<sup>31</sup> whereas the related,  
 93 but more rigid, bicyclic 6-acetamido-6-deoxy-castanospermine (6-Ac-Cas,  $K_i = 300$  nM) (Figure 1) is a much  
 94 better inhibitor of OGA.<sup>32</sup> Less intuitively, tailoring the configurational and substitution pattern about a  
 95 pyrrolidine core, coupled with introduction of a mimic of the aglycone, led to a potent cationic OGA  
 96 inhibitor *N*-(*p*-fluorophenylpropyl) iminocyclitol derivative NFPI ( $K_i = 9$  nM, Figure 1).<sup>21</sup>



97  
 98 **Figure 1.** The substrate-assisted catalytic mechanism used by OGA to cleave *O*-GlcNAc from protein substrates and carbohydrate-  
 99 based inhibitors of this enzyme. A) The catalytic mechanism of OGA involves two carboxyl residues within the enzyme active site  
 100 (D174 and D175) that catalyze a two-step process proceeding through an oxazoline intermediate. B) The proposed  ${}^4E$  envelope  
 101 or  ${}^4H_3$  half-chair conformation for the transition states for formation and breakdown of the oxazoline intermediate. C) Structures  
 102 of carbohydrate-based hOGA inhibitors PUGNAc, NAG-thiazoline, NButGT, Thiamet-G, MK-8719. D) Structures of iminosugar-type  
 103 inhibitors of hOGA including 6-Ac-Cas, GlcNAcstatin C, NFPI and the general structure of the DNJC-thiazolidines reported in this  
 104 work.

105  
 106 One challenge associated with creating rationally designed carbohydrate-based inhibitors of OGA  
 107 as tool compounds has been their selectivity toward OGA over other functionally related glycoside  
 108 hydrolases. In humans, the lysosomal hexosaminidases HexA and HexB from family GH20 and *N*-acetyl- $\alpha$ -



109 glucosaminidase NAGLU from family GH89 play important roles in regulating levels of various  
110 glycoconjugates and their concomitant inhibition is accordingly undesirable. Underscoring this problem,  
111 NAG-Thiazoline, PUGNAc, and 6Ac-CAS all lack selectivity for OGA.<sup>28,32</sup> Selectivity has been realized by  
112 incorporating a slightly larger N-acyl substituent at the position typically occupied by the N-acetyl group  
113 of the natural substrate, which is tolerated by a pocket found within the active site of OGA but that is not  
114 present in these other functionally related enzymes.<sup>21-23,33</sup> This approach was demonstrated for 1,2-  
115 dideoxy-2'-propyl- $\alpha$ -D-glucopyranoso[2,1-d]- $\Delta$ 2'-thiazoline (NButGT, Figure 1) to provide 700-fold  
116 selectivity for OGA. A similar strategy applied to GlcNAcstatin-G afforded 150-fold selectivity, however,  
117 when applied to PUGNAc, this approach yielded only modest selectivity of 10-fold. Thiamet-G exploits this  
118 approach to yield excellent 37,000-fold selectivity, which led to a search for improved analogues that  
119 could be used clinically. The result of a major chemistry effort led to MK-8719<sup>15</sup> ( $K_i$  = 7.9 nM, Figure 1), a  
120 less polar difluoromethyl derivative of Thiamet-G having excellent properties as a tool compound with no  
121 significant measurable activity against lysosomal hexosaminidases and NAGLU. This work stresses the  
122 need for potent carbohydrate-based OGA inhibitors that can be readily tailored to tune their molecular  
123 properties. The subsequent advance of MK-8719 to first-in-human phase I clinical trials for potential  
124 treatment of neurodegeneration highlights the interest and potential of such inhibitors, not only as tool  
125 compounds for basic biology, but also as potential first-in-class therapeutics.<sup>15</sup>

126 Here we describe rationally conceived iminosugar inhibitors of OGA that combine advantageous  
127 structural features seen in various known OGA inhibitors. These inhibitors show exceptional potency and  
128 selectivity toward OGA over functionally related human enzymes. In particular, we incorporated a rigid  
129 bicyclic core with an embedded basic isothioureia moiety that confers mimicry of the cationic nature of  
130 the transition state. Appended to this core we link an aryl moiety to mimic the aglycone leaving group.  
131 These 6*S*,5*N*-alkyliminomethylidene-2-carboxamido-1,2-dideoxynojirimycin (DNJC-thiazolidine)  
132 derivatives are new representatives of the class of sp<sup>2</sup>-iminosugar glycomimetics that are characterized  
133 by the presence of a pseudoamide functionality in their core structure.<sup>34-38</sup> We further define the  
134 molecular basis of this inhibition through structural studies with human OGA and with a bacterial  
135 orthologue of hOGA. We then take advantage of this information to optimize the structure in terms of  
136 inhibition potency and selectivity towards hOGA. We additionally show that these compounds exhibit  
137 superb cell-based potency and are also brain permeable. Using this new structural class of inhibitors, we  
138 create a chemoproteomic probe that enables convenient isolation of endogenous OGA from brain tissue  
139 with high efficiency. Finally, using this tool we have identified a diverse set of native post-translational  
140 modifications on endogenous OGA from brain, most of which have been identified for the first time.

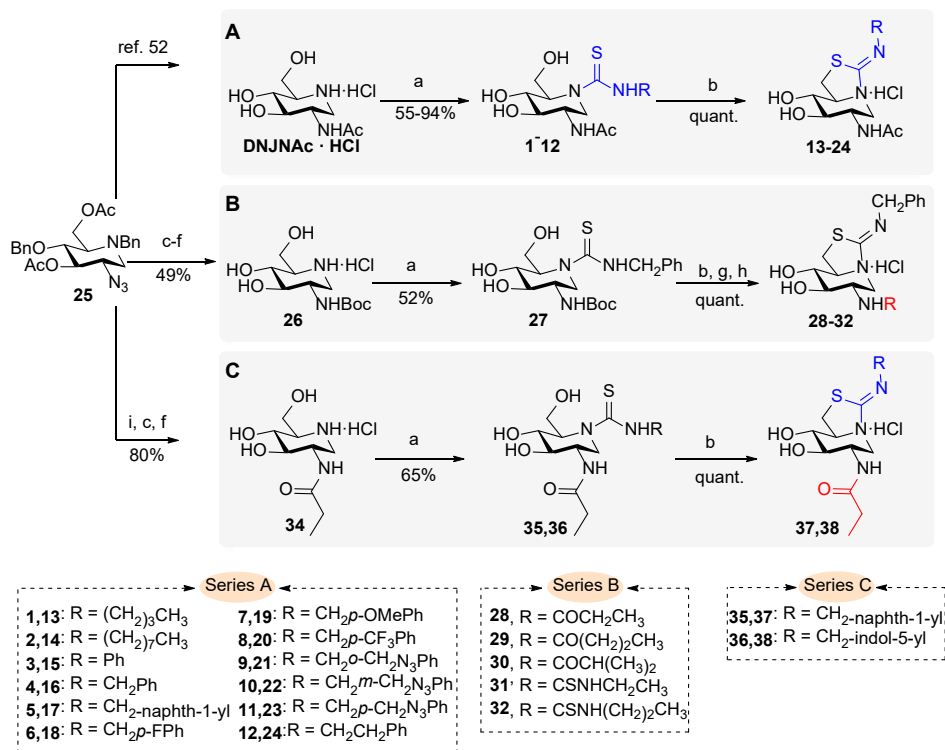
141

## 142 RESULTS AND DISCUSSION:

143 We were motivated by interest in OGA inhibitors and chemical probes to identify a carbohydrate-  
144 based inhibitor that could have its molecular properties readily tuned. An underexplored area are bicyclic  
145 sp<sup>2</sup>-iminosugars that bear an isothioureia group to which various substituents can be appended.<sup>36,39-41</sup> The  
146 dimeric nature of hOGA, in which the active site is located at the base of a cleft found at the dimer  
147 interface, has suggested that compounds having a pendant group can reach the complementary monomer  
148 and may be able to derive binding potency in this manner.<sup>21,22</sup> We therefore reasoned that tuning the  
149 substituent on the exocyclic nitrogen could provide a convenient way to tune these inhibitors in a way  
150 that cannot readily be accomplished for Thiamet-G<sup>25</sup> or the GlcNAcstatins<sup>20</sup> (Figure 1). Modification of

151 these later two compounds requires complex synthetic manipulations as seen, for example, in the  
 152 generation of MK-8719.<sup>15</sup> With regard to selectivity for OGA over functionally related enzymes, we  
 153 reasoned that modifying the 2-acetamido group might afford selectivity for hOGA over functionally  
 154 related enzymes such as HexA and HexB, which have more constrained active site pockets.<sup>33</sup>

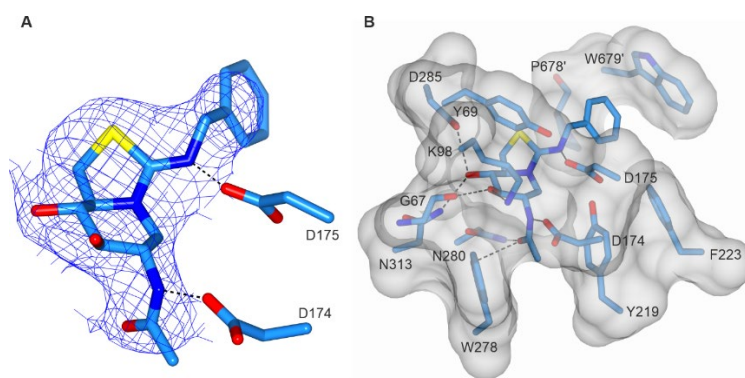
155 With the above considerations in mind, we set out to synthesize new sp<sup>2</sup>-iminosugars having a  
 156 common 2-acetamido-1,2-dideoxynojirimycin (DNJNac) core with the piperidine nitrogen engaged in a  
 157 five-membered cyclic isothiourea moiety. We therefore implemented a simple two-step synthetic  
 158 strategy to generate 6S,5N-alkyliminomethylidene-2-carboxamido-1,2-dideoxynojirimycin (DNJC-  
 159 thiazolidine) using as an advanced intermediate the iminosugar DNJNac. We accessed DNJNac using a  
 160 conveniently scalable procedure from commercially available D-glucuronolactone.<sup>42</sup> We first focused on  
 161 using a series of simple alkyl and aryl isothiocyanates that could help test the potential value of extending  
 162 the pseudo-aglycone moiety by measuring the resulting effects on inhibition of hOGA. Straightforward  
 163 coupling of DNJNac with a series of isothiocyanates afforded, with total chemoselectivity, the DNJNac-  
 164 thioureas **1-12**. These intermediates were in turn smoothly transformed through acid-catalyzed  
 165 intramolecular nucleophilic cyclization to furnish us with the desired DNJNac-thiazolidines **13-24** (Scheme  
 166 1A). In agreement with data for analogous systems,<sup>35</sup> we observed only a single set of signals in the NMR  
 167 spectra of the 2-imino-1,3-thiazolidine derivatives. These spectra therefore indicated that only the more  
 168 stable Z-diastereomer at the C=N double bond, devoid of unfavorable 1,3-parallel steric interactions, is  
 169 formed.<sup>35,43</sup>



170  
 171 **Scheme 1.** Synthesis of DNJNac-thioureas **1-12** and DNJNac-thiazolidines **13-38**. a. Et<sub>3</sub>N, R-NCS, DMF, RT, 18 h, 55-94%. b.  
 172 HCl(conc), MeOH, RT, 18 h, quant. c. NaOMe, MeOH, RT, 18 h; d. PPh<sub>3</sub>, THF-NH<sub>4</sub>OH, 60 °C, 18 h; e. Boc<sub>2</sub>O, Et<sub>3</sub>N, dioxane, RT, 18 h;  
 173 f. H<sub>2</sub>, Pd/C, MeOH, RT, 18 h; g. TFA, 1:1 DCM-H<sub>2</sub>O; h. RCOCl or RNCS, Et<sub>3</sub>N, MeOH, RT, 18 h, quant; i. Zn, CuSO<sub>4</sub> (aq), THF-PrOH-  
 174 Pr<sub>2</sub>O, RT, 20 min.

175 We next evaluated these compounds (**13-24**) as inhibitors of hOGA using 4-methylumbelliferyl 2-  
176 acetamido-2-deoxy-D-glucopyranoside (MU-GlcNAc)<sup>3</sup> as a substrate. We found, however, that the  
177 potency of these compounds was such that the  $K_i$  values were lower than the enzyme concentration  
178 needed to perform the assay. We therefore turned to using the more sensitive substrate resorufin 2-  
179 acetamido-2-deoxy-D-glucopyranoside (Res-GlcNAc) in combination with analysis using the Morrison  
180 equation.<sup>44</sup> These data (Table 1) revealed that the identity of the group appended to the exocyclic  $N'$ -  
181 nitrogen had a major influence on the inhibitory potency of these molecules. We found that while the  
182 aliphatic  $N'$ -alkyl derivatives **13** ( $K_i = 24 \pm 8$  nM) and **14** ( $K_i = 20 \pm 7$  nM) were more potent than their  $N'$ -  
183 phenyl counterpart **15** ( $K_i = 400 \pm 130$  nM), positioning the aromatic ring further from the endocyclic  
184 nitrogen as in the  $N'$ -benzyl derivatives **16** ( $K_i = 3 \pm 1$  nM),  $N'$ -(1-naphthylmethyl) **17** ( $K_i = 0.3 \pm 0.1$  nM)  
185 and  $N'$ -phenethyl **24** ( $K_i = 0.9 \pm 0.3$  nM) markedly enhanced their potency against hOGA. Finally, various  
186 groups at the aromatic moiety in compounds **18-20** were detrimental ( $K_i = 30-50$  nM), whereas electron  
187 donating azidomethyl substituents in **21-23** enhanced the inhibition potency ( $K_i = 0.22-0.40$  nM).

188 Given these observations we sought to understand the molecular basis of inhibition of OGA by  
189 these bicyclic inhibitors. We initially obtained structures of the complex with the bacterial orthologue of  
190 hOGA from *Bacteroides thetaiotamicron* (BtGH84),<sup>45</sup> at 1.50 Å resolution, then subsequently with the  
191 recently described dimeric catalytic domain of hOGA,<sup>21-23</sup> at 2.41 Å resolution. Analysis of the diffraction  
192 data obtained from crystals of the catalytic domain of hOGA soaked with **16** showed unambiguous  
193 electron density within the active site of hOGA (Figure 2A). The catalytic residues, D174 and D175, engage  
194 compound **16** with D174 hydrogen bonding to the 2-acetamido group and D175 hydrogen bonding to the  
195 exocyclic nitrogen. As seen in complexes with substrates and transition state analogues,<sup>26,29,32,45</sup> the 2-  
196 acetamido group is oriented and stabilised by interactions with W278, Y219, and N280 (Figure 2B). The  
197 hydroxyl groups at the C3 and C4 position of the piperidine ring form hydrogen bonds with the backbone  
198 oxygen of G67 and the sidechains of K98, D285, and N313. A hydrophobic pocket formed by the residues  
199 of F223, from one monomer, and P678' and W679', from the second monomer, interacts with the phenyl  
200 group of **16** through  $\pi$ -stacking interactions.



202 **Figure 2.** Compound **16** in complex with hOGA. Active site residues of hOGA are shown in blue. (A) **16** is shown in blue with the  
203 corresponding REFMAC maximum-likelihood/ $\sigma$ A-weighted  $2F_o - F_c$  map, shown in dark blue, contoured at 0.11 electrons Å<sup>-3</sup>. (B)  
204 Surface representation of the active site pocket of hOGA with neighbouring residues and hydrogen bond interactions highlighted  
205 by dashed lines.

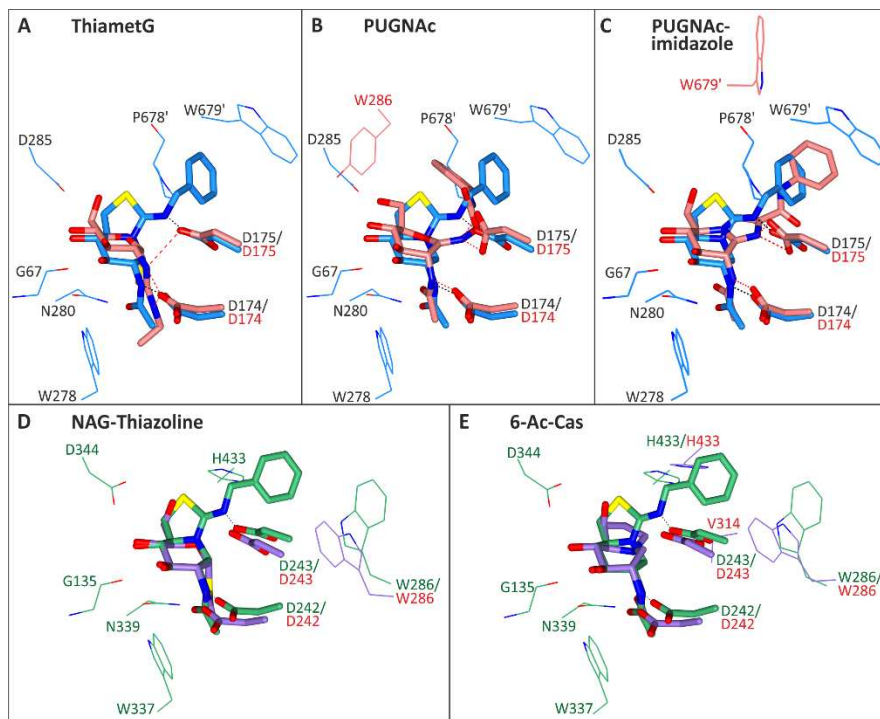
206 A higher resolution structure of compound **16** in complex with BtGH84 was sought to gain more  
207 precise insights into the interactions of **16** with GH84 enzymes. Compound **16** also shows good inhibition  
208 of BtGH84 ( $K_i = 733$  nM), consistent with its structurally conserved active site making it a useful model of

209 hOGA. The piperidine ring of **16** was observed to bind in an unambiguous  ${}^4C_1$  conformation in complex  
210 with *BtGH84* and the thiazolidine ring was observed in a  ${}^1E$  conformation. The conformation of the  
211 piperidine ring corresponds to that of the pyranose ring seen for the GalNAc-oxazoline intermediate in  
212 the reaction coordinate of  $\beta$ -*N*-acetylglucosaminide hydrolysis catalyzed by GH84 enzymes (Figure 1A).  
213 Similarly, when bound in the active site of hOGA, the pyranose ring of the **16** was also bound in a  ${}^4C_1$   
214 conformation and engages in a similar set of protein contacts. The high resolution *BtGH84* structure also  
215 clearly shows the *N*-alkyl imine of **16** in the *Z*-configuration, which is consistent with the weaker electron  
216 density observed for this substituent in the complex of **16** with hOGA. Within the *BtGH84* structure we  
217 note, however, some features that are distinct from those seen in the hOGA complex. The position of the  
218 phenyl group of the *N*-alkyl substituent in *BtGH84* has shifted to be pointing out of the active site pocket,  
219 possibly due to a steric clash with H433 and the absence of an equivalent residue in place of W679' of  
220 hOGA. Also, the sulfur atom of the thiazolidine ring is in close proximity to and interacts with H433 and  
221 two water molecules, whereas the residue in the equivalent position of hOGA, P678', is not within bonding  
222 distance (Supplemental Figure 13). Finally, a common feature in these two structures that likely confers  
223 affinity is interaction of the isothioureia group through a hydrogen bond between the exocyclic nitrogen  
224 and the catalytic residue D175 in hOGA and D243 *BtGH84*, respectively (Supplemental Figure 13). This  
225 bond would not be present if the moiety adopted an *E*-configuration, where the endocyclic sulfur atom  
226 and the exocyclic benzylic methyl are in a quasi-*trans* disposition.

227 Comparing the structure of hOGA in complex with **16** to complexes with Thiamet-G,<sup>21,23</sup> PUGNAc,<sup>23</sup>  
228 and PUGNAc-imidazole hybrid inhibitors<sup>21</sup> shows that the six-membered ring and 2-acetamido group  
229 adopt the same position in all cases (Figure 3). The phenyl moiety of the PUGNAc inhibitors shows a degree  
230 of movement compared to **16** that is tolerated within the active site (Figure 3). The phenyl group of **16**  
231 adopts a more compact position than that of phenyl group of the PUGNAc-imidazole type inhibitor. This  
232 movement allows the protein backbone to lie closer to the active site as can be seen for W679' where the  
233 C $\alpha$  has moved 5.4 Å closer to the active site. The position of the backbone seen in the complex with **16**  
234 would cause a clash with the benzyl group of PUGNAc, however, the residues between R664-R682 are not  
235 observed in this structure, suggesting they adopt a flexible disordered state further away from the active  
236 site.<sup>23</sup> Thiamet-G binds in the same  ${}^4C_1$  conformation, whereas PUGNAc binds in a  ${}^4E$  conformation and  
237 PUGNAc-imidazole binds in a  ${}^4E/{}^4H_3$  conformation.

238 The 5,6-fused ring structure of **16** bears resemblance to the bicyclic iminosugar 6-Ac-Cas and, to  
239 a lesser extent, that of NAG-thiazoline (Figure 3D and E). Known complexes of 6-Ac-Cas and NAG-  
240 thiazoline in complex with *BtGH84*<sup>32</sup> allowed us to compare the orientations of these molecules with that  
241 seen for compound **16**. The general position of the inhibitors 6-Ac-Cas and NAG-thiazoline in the active  
242 site of *BtGH84* are the same. However, in comparison to 6-Ac-Cas where the five-membered ring lies in  
243 the same plane as the six-membered ring, the five-membered ring of **16** is pointing  $\sim 70^\circ$  away from the  
244 plane aligning to the C6 hydroxyl of NAG-thiazoline. This may arise due to introduction of the phenyl  
245 moiety appended to the five-membered ring needing to avoid a steric clash with the sidechain of V314.  
246 Strikingly, the puckering of the 5-membered ring of **16** allows it to access a different ring conformation.  
247 6-Ac-Cas adopts a  ${}^1,4B/{}^1S_3$  conformation, whereas Thiamet-G and NAG-thiazoline adopt a  ${}^4C_1$   
248 conformation. In summary, we find that bicyclic compound **16** binds to both hOGA and *BtGH84* with its 5-  
249 membered ring adopting a similar position as seen for the analogous ring of 6-Ac-Cas and the 6-membered  
250 ring adopting an undistorted  ${}^4C_1$  conformation like NAG-thiazoline. These two features of **16**, coupled with  
251 the additional bonding to the exocyclic nitrogen and the optimal positioning of the pendent aryl ring

252 within the active site, likely account for its picomolar inhibition constant as compared to the ~300 nM  $K_i$   
253 values seen for NAG-thiazoline and 6-Ac-Cas with hOGA.<sup>32</sup>



255 **Figure 3.** Superimposition of **16** bound to hOGA and *BtGH84* with various characterised OGA inhibitors. The general acid/general  
256 base residues, D174/D175 from hOGA and D242/D243 from *BtGH84*, are represented with thicker bond widths compared to the  
257 surrounding active site residues. Residues in the overlaid structures that occupy different positions and that make important  
258 interactions are highlighted. Panels A-C compare compound **16** bound to hOGA and panels D-E compare **16** bound to *BtGH84*.  
259 (A) Thiamet-G, shown in orange, in complex with hOGA (PDB ID: 5M7S)<sup>21</sup>. (B) PUGNAc, shown in orange, in complex with hOGA  
260 (PDB ID: 5UHO)<sup>23</sup>. (C) PUGNAc-imidazole hybrid inhibitor, shown in orange, in complex with hOGA (PDB ID: 5M7T)<sup>21</sup>. (D) NAG-  
261 thiazoline, shown in purple, in complex with *BtGH84* (PDB ID: 2CHN)<sup>45</sup>. (E) 6-Ac-Cas, shown in purple, in complex with *BtGH84*  
262 (PDB ID: 2XJ7)<sup>32</sup>.

263 The structure of **16** in complex with hOGA and its comparison to that of known selective and non-selective  
264 inhibitors of hOGA suggested that these compounds may also be effective inhibitors of human GH20  
265 hexosaminidases, which as noted above use a similar catalytic mechanism and also processes GalNAc-  
266 configured substrates. We therefore assessed the inhibition of HexA by DNJNAc-thiazolidines **13-24** (Table  
267 1) and, in parallel, commercial hexosaminidases (Supplemental Table 1). While these arylalkylimino  
268 DNJNAc-thiazolidines, which all contain a simple 2-acetamido group as seen for the natural substrate, rank  
269 among the most potent hOGA inhibitors reported to date, their potency toward HexA was good but not  
270 exceptional. The resulting selectivity of these inhibitors for hOGA over HexA ( $K_i$  hOGA/ $K_i$  Hex) was quite  
271 promising, being up to ~30000-fold for some of the picomolar hOGA inhibitors including **17**, **21**, and **22**  
272 (Table 1). Nevertheless, we judged that improving on this selectivity would be valuable to deliver highly  
273 selective tool compounds that would be convenient to use and avoid potential off-target effects.

274

275 **Table 1.** Inhibition constants ( $K_i$ ) of inhibitors **13-24** for hHexA  $K_i$  ( $\mu\text{M}$ ) and hOGA  $K_i$  (nM).<sup>a</sup>

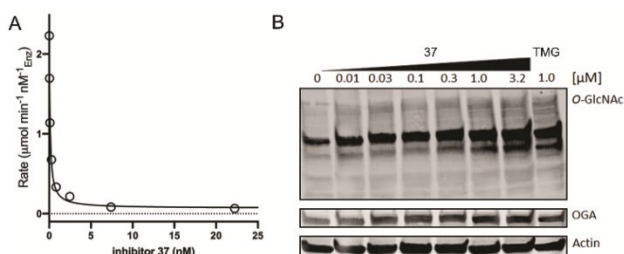
Compound	NHR group	=N-R group	hHexA $K_i$ ( $\mu\text{M}$ )	hOGA $K_i$ (nM)	Selectivity ratio $K_i\text{Hex}/ K_i\text{hOGA}$
<b>Series A - Potency</b>					
<b>13</b>	Ac	(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	4.2 ± 1	24 ± 8	175
<b>14</b>	Ac	(CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub>	3.8 ± 1	20 ± 7	190
<b>15</b>	Ac	Ph	14 ± 4	400 ± 130	35
<b>16</b>	Ac	CH <sub>2</sub> Ph	9.9 ± 4	3 ± 1	3300
<b>17</b>	Ac	CH <sub>2</sub> -1-naphthyl	9.0 ± 4	0.3 ± 0.1	30000
<b>18</b>	Ac	CH <sub>2</sub> p-FPh	4.2 ± 1	50 ± 20	84
<b>19</b>	Ac	CH <sub>2</sub> p-OMePh	8.8 ± 4	30 ± 10	293
<b>20</b>	Ac	CH <sub>2</sub> p-CF <sub>3</sub> Ph	3.0 ± 1	50 ± 20	60
<b>21</b>	Ac	CH <sub>2</sub> o-CH <sub>2</sub> N <sub>3</sub> Ph	5.0 ± 3	0.27 ± 0.09	18518
<b>22</b>	Ac	CH <sub>2</sub> m-CH <sub>2</sub> N <sub>3</sub> Ph	5.0 ± 3	0.22 ± 0.07	22727
<b>23</b>	Ac	CH <sub>2</sub> p-CH <sub>2</sub> N <sub>3</sub> Ph	3.7 ± 1	0.4 ± 0.2	9250
<b>24</b>	Ac	CH <sub>2</sub> CH <sub>2</sub> Ph	21 ± 7	0.9 ± 0.3	23333
<b>Series B - Selectivity</b>					
<b>28</b>	COCH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> Ph	>100	8 ± 3	>10000
<b>29</b>	CO(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> Ph	>100	900 ± 300	>100
<b>30</b>	COCH(CH <sub>3</sub> ) <sub>2</sub>	CH <sub>2</sub> Ph	>100	300 ± 100	>300
<b>31</b>	CSNHCH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> Ph	>100	1000 ± 400	>100
<b>32</b>	CSNH(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> Ph	>100	800 ± 300	>100
<b>Series C - Combined</b>					
<b>37</b>	COCH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> -1-naphthyl	>100	0.8 ± 0.2	>120000
<b>38</b>	COCH <sub>2</sub> CH <sub>3</sub>	CH <sub>2</sub> -5-indolyl	>100	2.4 ± 0.9	>40000

276 Inhibition was competitive in all cases. <sup>a</sup> Human HexA and human OGA enzyme inhibition were  
 277 measured as described (SI) and  $K_i$  values are from  $\geq 2$  assays with standard deviation of  $\leq 35\%$  of the  
 278 reported mean.  
 279

280 Inspired by work showing the improvement on selectivity against hOGA for inhibitors bearing  
 281 extended N-acyl groups,<sup>20,25,28</sup> we therefore devised a second series of DNJC-thiazolidines that retained a  
 282 benzyl substituent at the imino center as in compound **16**, but the 2-acetamido group was replaced by  
 283 propionamido **28**, butyramido **29**, isobutyramido **30**, *N'*-ethylthioureido **31**, or *N'*-propylthioureido **32**  
 284 functionalities (Scheme 1B). We made the choice to retain the benzyl group based on the data showing  
 285 favorable contribution of an aromatic substituent located several bonds away from the sugar-like core for  
 286 hOGA inhibition. We implemented a divergent synthetic strategy starting from the known 2-azido-2-deoxy  
 287 DNJ derivative **25**, a synthetic intermediate in the route leading to DNJNAc (Scheme 1B).<sup>42</sup> Deacetylation  
 288 using Zemplen conditions followed by Staudinger reduction of the azido group with *in situ* Boc-protection  
 289 of the resulting amine, followed by Pd-catalyzed hydrogenolysis, delivered iminosugar **26** in 60% yield  
 290 over three steps. Reaction of **26** with benzyl isothiocyanate afforded us the pivotal thiourea **27**, which was  
 291 transformed into our target 2-benzyliminothiazolidines through parallel three-step, one-purification  
 292 reaction sequences. This series of reactions involved: (i) acid-catalyzed cyclization to form the five-  
 293 membered isothiourea ring, (ii) TFA-induced hydrolysis of the carbamate, and (iii) N-acylation (**28-30**) or  
 294 isothiocyanation reaction (**31** and **32**) of the free amine with the corresponding acyl chloride or alkyl  
 295 isothiocyanate (Scheme 1B). We assayed these candidate inhibitors and found that modifications of this  
 296 group completely eliminated inhibition of HexA at concentrations up to 100  $\mu\text{M}$ . In contrast, hOGA was  
 297 relatively tolerant to these structural variations. Nevertheless, the scope of changes that were well  
 298 tolerated was limited. Replacement of the acetamido group of **16** ( $K_i = 3.0 \pm 1$  nM) by butyramido,  
 299 isobutyramido, *N'*-ethylthioureido, or *N'*-propylthioureido, as incorporated into compounds **29-32**,  
 300 resulted in a drop of potency of approximately two orders of magnitude ( $K_i = 800\text{-}1000$  nM). A subtle  
 301 change to a propionamido group as in **28**, however, only slightly affected inhibitory potency ( $K_i = 8 \pm 3$

302 nM) with an over 10,000-fold selectivity for hOGA (Table 1). These collective inhibition data advised  
 303 generating an hOGA inhibitor that would exhibit high selectivity by virtue of incorporating a 2-  
 304 propionamido group as in **28**, as compared to the 2-acetamido group of **16**. We also reasoned that it  
 305 would be beneficial to install a pendent polycyclic aromatic group, since for **17** we see a 10-fold  
 306 improvement in potency over compound **16** that has a simple pendant phenyl group. Towards this end,  
 307 we prepared 2-deoxy-2-propionamido-DNJ derivative **34** from the common intermediate **25** in a three-  
 308 step procedure (Scheme 1C). We then N-acylated **34** using 1-naphthylmethyl isothiocyanate or 5-  
 309 indolylmethyl isothiocyanate and converted the thiourea adducts, **35** and **36**, under acidic conditions to  
 310 the desired bicyclic sp<sup>2</sup>-iminosugars **37** and **38** (Scheme 1C). We found these compounds to be potent,  
 311 showing high pM (**37**) to low nM (**38**) K<sub>i</sub> values (Table 1, Figure 4), as well as being highly selective for  
 312 hOGA (>40,000 to >120,000 fold). Given their properties, we judged that both **37** and **38** could serve as  
 313 useful tool compounds for studying O-GlcNAc. We therefore examined the dose-dependent effect of  
 314 these compounds on protein O-GlcNAcylation using the convenient neuronal cell model SK-N-SH. Using  
 315 as a benchmark the compound Thiamet-G, which has similar potency (K<sub>i</sub> = 2.1 nM),<sup>26</sup> we found **37** was  
 316 equally effective at increasing cellular O-GlcNAc levels. **37** Was active at nM concentrations as assessed in  
 317 immunoblot assays both by using pan-specific anti-O-GlcNAc and anti-OGA antibodies to monitor levels  
 318 of O-GlcNAc and OGA (Figure 4).

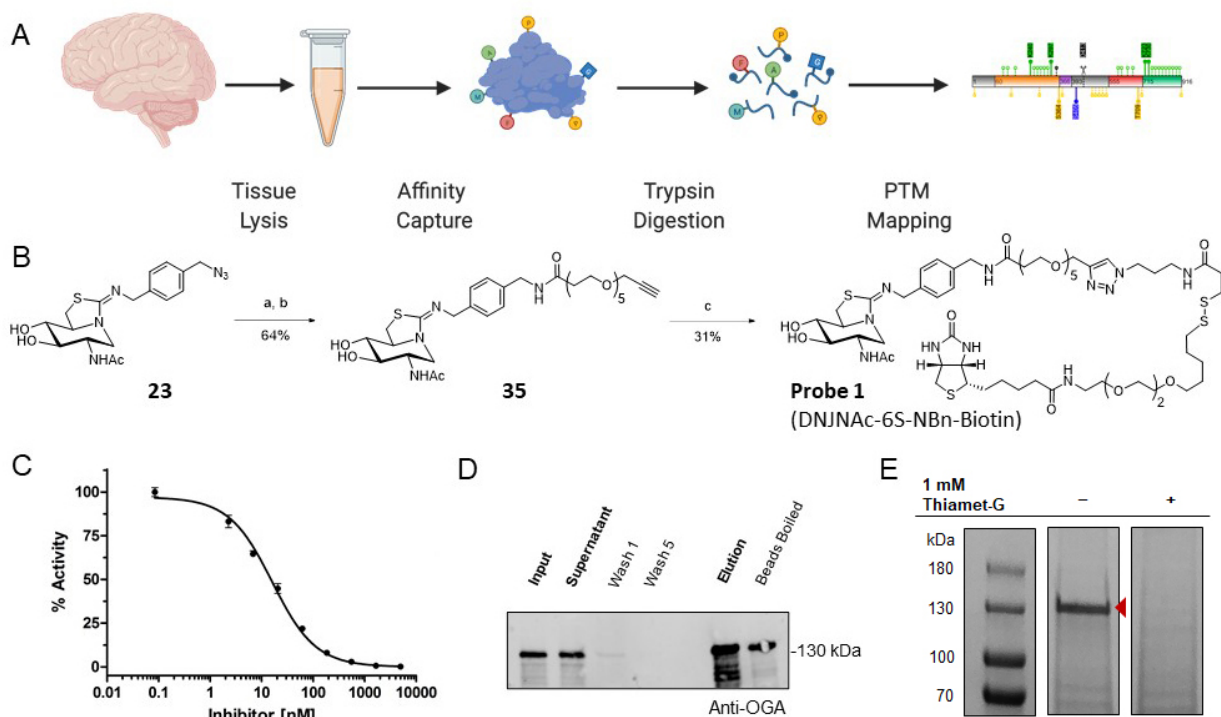
319 The high potency of these compounds, coupled with tolerance to various pendent aryl groups  
 320 that can extend out of the immediate active site cleft (Figure 2), suggested to us that modification of these  
 321 inhibitors could yield useful affinity-based chemoproteomics probes. Previous non-covalent  
 322 chemoproteomic probes have been reported for other enzyme classes including, for example, Src  
 323 multidomain kinases,<sup>46</sup> casein kinases,<sup>47</sup> heat shock protein (HSP) chaperones,<sup>48</sup> protease complexes,<sup>49</sup>  
 324 histone deacetylases,<sup>50</sup> as well as the non-lysosomal glycosidase β-glucosidase 2 (GBA2)<sup>51</sup>. Few, however,  
 325 have been applied to identify the post-translational modification state of specific proteins of interest. We  
 326 felt that such a tool could prove useful for enriching OGA from various tissues to allow a detailed analysis  
 327 of the properties of endogenous OGA, including identifying sites of post-translational modifications. We  
 328 envisioned that we could tether a biotin affinity tag through a long linker to the phenyl group of inhibitor  
 329 **23**, which exhibits pM potency against hOGA, such that the biotin moiety could still reach a streptavidin-  
 330 presenting matrix even with the inhibitor bound within the deep active site cleft of OGA. We elected to  
 331 use a disulphide-containing linker that could be cleaved using gentle conditions. We were inspired in this  
 332 regard by the work of Teranishi *et al.* on proteases<sup>49</sup>, who used a 70 Å polyethylene glycol (PEG) linker  
 333 within an affinity probe.



335 **Figure 4:** In vitro and in cell potency of **37**. (A) K<sub>i</sub> value determined using the Morrison equation and (B) cell-based assays obtained  
 336 using SK-N-SH neuroblastoma cells treated with **37** for 24 h using immunoblot of cell lysates for dose-dependent effects of cell-  
 337 based inhibition of OGA on global O-GlcNAc levels and OGA levels. A comparison of the effects of **37** compared to Thiamet-G  
 338 (TMG) is made.



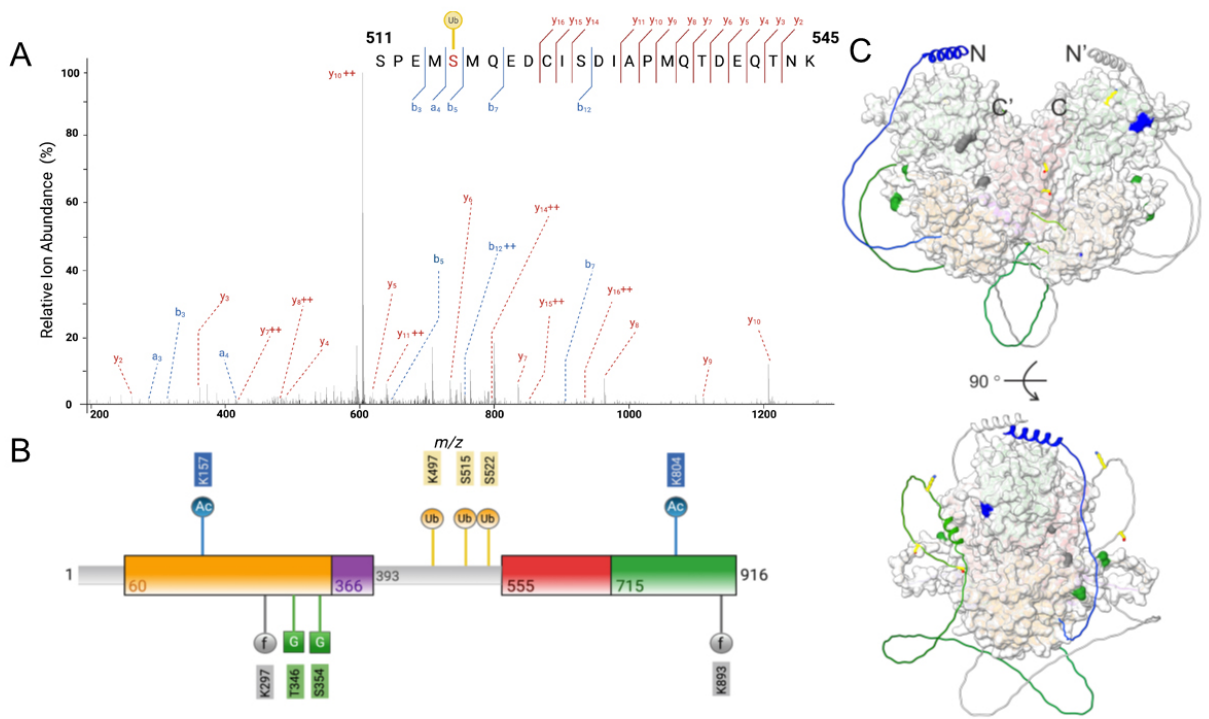
339 During the synthesis of our target probe, we found that using large quantities of PEG linkers, which  
 340 are hygroscopic, proved troublesome during amide coupling reactions. To circumvent this problem, we  
 341 elected to couple together suitably functionalized inhibitor and biotin components using a copper azide-  
 342 alkyne cycloaddition (CuAAC) reaction. We therefore reduced the azide of inhibitor **23** and coupled it with  
 343 alkyne-functionalized PEG<sub>4</sub>-acid (Figure 5A). The resulting alkyne intermediate **39** was then coupled to  
 344 disulfide biotin azide using CuAAC to give **40** (DNJNAc-6S-NBn-Biotin). We then determined the inhibitory  
 345 potency of DNJNAc-6S-NBn-Biotin toward OGA and found it showed an IC<sub>50</sub> value of 7.6 ± 0.5 nM (Figure  
 346 5C), which we reasoned was suitably potent binding to enable its intended use as a chemoproteomic  
 347 probe. We envisioned a workflow (Figure 5B) in which the probe would be first bound to OGA within  
 348 clarified tissue lysates, after which the binary OGA-Probe complex would be precipitated using magnetic  
 349 streptavidin beads. After washing of the beads, elution under mild reducing conditions with dithiothreitol  
 350 (DTT) would release the enriched OGA. We first used recombinant hOGA to determine how effective our  
 351 proposed workflow would be in precipitating OGA. Success of the probe in capturing of hOGA, using this  
 352 protocol, was verified by detection of hOGA in the eluate by anti-OGA immunoblot (Supplemental Figure  
 353 14). Furthermore, we found that five rounds of washing of the beads still resulted in retention of hOGA  
 354 bound to immobilized DNJNAc-6S-NBn-Biotin on the beads. These data collectively show that DNJNAc-6S-  
 355 NBn-Biotin has sufficient affinity to bind hOGA and enable its high-level enrichment.



357 **Figure 5:** Chemoproteomic enrichment workflow with probe design, synthesis, in vitro evaluation, and use for precipitation of  
 358 bovine OGA from brain tissue. (A) Scheme showing the chemoproteomic strategy used to enrich bovine OGA from brain. (B)  
 359 Scheme showing the chemical synthesis of cleavable DNJNAc-6S-NBn-Biotin using **23** as a starting material. Reagents and  
 360 conditions: a) PPh<sub>3</sub>, THF/H<sub>2</sub>O, RT, 21 h. b) alkyne-PEG<sub>4</sub>-acid, HBTU, DIPEA, DMF, RT, 22 h; c) biotin-azide, sodium ascorbate,  
 361 CuSO<sub>4</sub>·5H<sub>2</sub>O, DMF/H<sub>2</sub>O, RT, 17 h. (C) In vitro IC<sub>50</sub> value determined for DNJNAc-6S-NBn-Biotin toward hOGA reveals an IC<sub>50</sub> value  
 362 of 7.6 nM. (D) Small scale validation experiment showing DNJNAc-6S-NBn-Biotin enables enrichment of bovine OGA from brain  
 363 as determined using immunoblot assay. (E) Large scale chemoproteomic enrichment of bovine OGA from brain tissue followed  
 364 by analysis of the elution by SDS-PAGE with detection using Coomassie stain in parallel with a Thiamet-G blocking control.



365 We next used DNJNAc-6S-NBn-Biotin to precipitate OGA from bovine brain tissue (bOGA). We  
 366 selected bovine brain because OGA was first isolated from this tissue,<sup>52</sup> where the enzyme is particularly  
 367 abundant, in view of the therapeutic potential of OGA inhibition in various neurodegenerative diseases.<sup>12-  
 368 15</sup> We used DNJNAc-6S-NBn-Biotin concentrations that would be maximized but not exceed available sites  
 369 of streptavidin on a convenient volume of beads (100  $\mu$ L). Following the precipitation experiment, for  
 370 which we performed a parallel control involving blocking of bOGA binding by adding the potent inhibitor  
 371 Thiamet-G, analysis of the OGA immunoreactivity by immunoblot (Figure 5D) showed that DTT elution  
 372 resulted in release of OGA from the resin. Released OGA was observed as having an approximate mass of  
 373 130 kDa, consistent with the initial report on its biochemical purification.<sup>3</sup> Analysis of the eluates showed  
 374 efficient and high-level enrichment of OGA (Figure 5E) that was blocked by Thiamet-G. Given the efficiency  
 375 of the pull-down, we focused on excising the band corresponding to OGA and identifying OGA-derived  
 376 peptides that were post-translationally modified. Tryptic in-gel digestion of the isolated OGA and analysis  
 377 of the resulting tryptic fragments by LC-MS/MS sequencing and analysis using Byonic revealed 436 unique  
 378 peptides that covered 62% of full-length bOGA. Several of these peptides were found to bear various post-  
 379 translational modifications. We performed a targeted search of the data using Byonic for the 30 most  
 380 common protein modifications and, using this strategy, identified 1676 peptide spectral matches, with  
 381 339 peptides bearing modifications. We elected to focus on those PTMs that are known to be  
 382 enzymatically installed as we felt these would be of wider interest and more likely to have regulatory  
 383 functions. To focus on candidate enzymatic modifications, we therefore tabulated such modifications that  
 384 also met our false discovery rate (FDR) criterion to be considered a modification on a peptide of having a  
 385  $|\text{Log Prob}| > 1.52$  (3% FDR). We assigned high-confidence modifications as those with  $|\text{Log Prob}| > 2.0$  (1%  
 386 FDR), which corresponded to a total of nine modifications including O-GlcNAcylation (2), lysine acetylation  
 387 (2), lysine N-formylation (2), and both O- and N-ubiquitination (3) (Table 2).



389 **Figure 6:** Mapping of post-translational modifications on endogenous OGA from bovine brain. A) Representative MS/MS peptide  
 390 sequencing data showing B and Y ion-series that enable mapping of O-ubiquitination to Ser515 of OGA. B) Domain structure of  
 391 bovine OGA with sites of identified post-translational modifications noted. Domains are coloured orange (GH84 domain), red  
 392 (stalk domain), purple (hinge domain), green (acetyltransferase-like domain), grey (intrinsically disordered regions). Modifications  
 393 are colour coded (Grey = Formylation, Blue = Acetylation, Green = O-GlcNAcylation, Yellow = Ubiquitination). C) Identified sites  
 394 of post-translational modifications marked on a homodimeric bovine OGA model (sequence according to UNIPROT entry  
 395 E1BQ16), built using an AlphaFold<sup>55</sup> predicted monomeric structure of bovine OGA with the second monomer guided into place  
 396 using the crystal structure of hOGA (PDB ID: 7OU6). Post-translational modification sites are colour coded the same as Panel B on  
 397 a surface representation of the structured regions. Domains are represented by ribbons with the colour from panel B applied;  
 398 the disordered regions of one monomer are coloured by Jones' rainbow (N-terminus: blue, C-terminus: red). Although AlphaFold  
 399 predicts the lowest energy monomeric 3D fold, regions 1-29 and 715-916 were manually moved in coot<sup>56</sup> due to steric clashes of  
 400 the disordered regions with the second monomer.

401 The MS/MS sequencing data for these modified tryptic peptides (Figure 6A, Supplemental Figures  
 402 15-23, Supplemental Tables 3-11) allowed us to identify the sites of these modifications, which we  
 403 mapped onto the domain structure and a three-dimensional model of bOGA (Figure 6B and C). Most of  
 404 these modification sites were previously unknown (Table 2) and almost all are found on the surface of the  
 405 protein. They include a ubiquitination site at K497 within the intervening disordered region between the  
 406 catalytic domain and the stalk domain, as well as the much less studied O-ubiquitination modification<sup>53</sup> at  
 407 S515 and S522. In addition, we observed new O-GlcNAcylation sites at T346 and S354. We also found  
 408 acetyl groups at K157 and K804 within the C-terminal putative acetyltransferase domain between residues  
 409 716-916, as well as the relatively little understood formyl modification<sup>54</sup> at K297 and K893. Interestingly,  
 410 K157, which is acetylated, is not surface exposed but instead is buried in the catalytic domain where it is  
 411 locked in place by three hydrogen bond interactions by the backbone oxygen atoms of K102, W107, and  
 412 E109. The human and bovine OGA 3-D coordinate sets do not contain any post-translational modifications  
 413 due to the production of hOGA in *Escherichia coli* and the use of structure prediction software for the  
 414 available bovine OGA model. Perhaps, when produced endogenously in mammals, acetylation of the  
 415 lysine could cause it to flip to be surface exposed or, alternatively, this region of the protein may have  
 416 some flexibility. Collectively, these targeted sequencing data confirm several sites of modification  
 417 previously identified using high-throughput proteomic analyses and uncovers multiple new sites of post-  
 418 translational modification on OGA from bovine brain.

419 **Table 2.** Post-translational modification data acquired for analysis of endogenous OGA.

Modification	Site	Residues	Peptide Sequence	Log Probability	Score	Site Annotated
Acetyl	K157	156 – 167	R(KAc)LDQVSQFGCR	3.3	359	*[Ub] <sup>57</sup>
Acetyl	K804	795 - 804	ISWIPFMQE(KAc)	2.99	299.7	
Formyl	K297	291 - 299	LFLGPY(Kf)GR	1.7	286.7	*[Ub] <sup>57</sup>
Formyl	K893	887 - 893	ILEFYS(Kf)	1.67	237.7	*[Ub] <sup>57</sup>
O-GlcNAc	T346	343 - 359	DVVM(Tg)DSEDSTVSIQIK	1.8	177	
O-GlcNAc	S354	343 - 359	DVVM(TSg)DSEDSTVSIQIK	6.38	304.3	
Ubiquitin	K497	492 - 510	MAEEL(KUb)PMDTDKESIAESK	2.24	237	
Ubiquitin	S515	510 - 545	SPEM(SUb)MQEDCISDIAPMQTDEQTNK	5.39	411.1	*[P] <sup>58</sup>
Ubiquitin	S522	510 - 545	SPEM(SMUb)DIAPMQTDEQTNK	3.44	200.2	*[P] <sup>57</sup>

420 Modifications are denoted on peptide sequence: (Ac) – Acetylation; (f) – Formylation; (g) – O-GlcNAcylation;  
 421 (Ub) – Ubiquitination; (P) - Phosphorylation. \* Site has been previously annotated for a modification.

422

423 **CONCLUSION:**

424 Small molecule inhibitors of the O-GlcNAc cycling enzyme OGA are proving widely useful in  
425 evaluating its functional roles in a range of physiological and pathophysiological processes. Tuning the  
426 properties of these inhibitors to enable their ready use in vivo has emerged as an important consideration.  
427 One challenge for carbohydrate-based inhibitors of OGA such as Thiamet-G, however, has been the  
428 synthetic difficulty associated with modifying the structure of the inhibitor.<sup>15</sup> An emerging approach<sup>21</sup> has  
429 been to exploit OGA inhibitors having features that extend outside of the active site pocket. Here we  
430 report a class of picomolar OGA inhibitors that we show, through structural analyses of their binding to  
431 human OGA, exploit the presence of such a pendent group to drive binding. These inhibitors are highly  
432 active in cell systems with cell potencies that are equivalent to that of Thiamet-G. Notably, these inhibitors  
433 and are also brain permeable (Supplemental Figures 24 and 25). Accordingly, we anticipate that such  
434 compounds should prove useful for evaluating the neuroprotective effects of OGA inhibition in different  
435 preclinical animal models. Such compounds may also serve as valuable leads for eventual clinical  
436 development.

437 The presence of only two O-GlcNAc cycling enzymes and their apparent constitutive activity makes  
438 the regulation of O-GlcNAc on target proteins a major puzzle for the field. Emerging data suggests that  
439 targeting of OGA and OGT to specific complexes may be one mechanism that dictates their activities. In  
440 this regard, the post-translational modifications of these proteins are topics of considerable interest since  
441 targeting may be controlled through such modifications. Indeed, this has been shown in the case of OGT,  
442 where phosphorylation was shown to affect nuclear import in myotubes, altering binding partners,  
443 increased activity, and stability from proteasomal degradation, as well O-GlcNAcylation affecting cellular  
444 localization.<sup>4,7</sup> The available information regarding the repertoire of post-translational modifications on  
445 OGT is, however, limited and even less is known for OGA (Table 2). Accordingly, deeper knowledge  
446 regarding endogenous modifications found on these proteins is needed to understand their regulation  
447 and the control of cellular O-GlcNAcylation.

448 One often used approach to mapping post-translational modifications is to overexpress proteins  
449 of interest fused to various peptide epitope tags. Alternatively, to avoid potential artifacts associated with  
450 overexpression, high affinity antibodies are sometimes used to immunoprecipitate proteins from tissues.  
451 However, the limited availability of high-quality antibodies validated for immunoprecipitation, and their  
452 cost, can deter the pursuit of such experiments. High affinity chemoproteomic probes offer a less  
453 explored, yet powerful, strategy to accomplish this goal. We exploited the high affinity bicyclic inhibitors  
454 reported here to develop a chemoproteomic affinity probe that could be used to enrich endogenous OGA  
455 from homogenized tissues. The initial report on purification of OGA from bovine brain was a heroic effort  
456 involving eight steps of purification from one kg of bovine brain that enabled detection of OGA in  
457 polyacrylamide gel by silver stain. In contrast, DNJNAc-6S-NBn-Biotin enabled highly efficient one-step  
458 purification of OGA, allowing ready detection of OGA by simple Coomassie staining within a  
459 polyacrylamide gel. The ability to rapidly isolate  $\mu\text{g}$  quantities of OGA has allowed us to map for the first  
460 time, in a targeted manner, post-translational modifications on this enzyme. Previous mapping of post-  
461 translational modifications to OGA in various high-throughput proteomic studies used cancer tissues  
462 rather than healthy tissues.<sup>57,59-61</sup> Furthermore, most of the modifications identified have only been  
463 observed once and such assignments are often not unambiguous. In addition, these previous studies have  
464 principally focused on specific modifications, such as phosphorylation and ubiquitination, rather than  
465 adopting a modification agnostic strategy. Here we identified a range of different modifications on OGA

466 from endogenous brain tissue including, for example, sites of formylation<sup>54</sup> (Table 2). The roles of this  
467 modification on cytoplasmic proteins remain poorly understood.

468 We expect that application of this OGA chemoproteomic probe, or other suitably functionalized  
469 high affinity OGA ligands, should allow comparative assessment of post-translational modifications from  
470 different tissues and species. Such data are expected to enable understanding which modification sites  
471 are most relevant to regulation of OGA localization and activity. Moreover, though O-GlcNAc is ubiquitous  
472 in tissues its abundance and that of the cycling enzymes that control this modification vary widely,  
473 suggesting that tissue-specific post-translational modification of OGA may contribute to its regulation.  
474 Targeted cell and biochemical studies will aid our understanding of the physiological roles of these  
475 modification sites on OGA function. Finally, though applied here to OGA, we also expect that this approach  
476 should be readily amenable to the study of other glycan processing enzymes that are likely to be regulated  
477 through various post-translational modifications.

478

## 479 **ASSOCIATED CONTENT**

### 480 **Supporting Information**

481 The supporting Information is available free of charge on the ACS Publications website:

- 482 • Methods for chemical syntheses, proteomic analyses, enzyme kinetic analyses, and X-ray structure  
483 determinations.
- 484 • Supplementary figures for proteomic data, characterization data for compounds, brain permeability  
485 data, and  $K_i$  determinations.
- 486 • Tables summarizing inhibition data, fragment ion series of modified peptides, and data collection and  
487 refinement statistics for X-ray structures.
- 488 • An .xls file with source data for MS-proteomics experiments.

489

## 490 **AUTHOR INFORMATION**

### 491 **Corresponding Authors**

492 \*Gideon J. Davies: gideon.davies@york.ac.uk \*Carmen Ortiz Mellet: mellet@us.es \*David J. Vocadlo:  
493 dvocadlo@sfu.ca

### 494 **Acknowledgements**

495 The authors are grateful for support from the Ministerio de Ciencia, Innovación y Universidades, the  
496 Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación, the European Regional  
497 Development Funds (projects RTI2018-097609-B-C21 and PID2019-105858RB-I00), the Junta de  
498 Andalucía, the Canada Foundation for Innovation, the British Columbia Knowledge Development  
499 Fund, Genome Canada/Genome BC (264PRO), and the Canadian Institutes of Health Research (CIHR;  
500 PJT-148732, PJT-156202). DJV thanks the Canada Research Chairs program for support as a Tier I Canada  
501 Research Chair in Chemical Biology. RAA is supported by a trainee award from the Michael Smith

502 Foundation Health Research (MSFHR). MG-C is a FPI fellow. Technical assistance from the research  
503 support services of the University of Seville (CITIUS) is also acknowledged.

504 **Notes**

505 The authors declare a patent covering inhibitors of OGA related to this work has been filed.

## 506 REFERENCES:

- 507 (1) Carmen-Rosa, T.; Hart, G. W. Topography and polypeptide distribution of terminal N-  
508 Acetylglucosamine residues on the surfaces of intact lymphocytes. *J. Biol. Chem.* **1984**, *259*, 3308-3317.
- 509 (2) Kreppel, L. K.; Blomberg, M. A.; Hart, G. W. Dynamic glycosylation of nuclear and cytosolic  
510 proteins cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide  
511 repeats. *J. Biol. Chem.* **1997**, *272*, 9308-9315.
- 512 (3) Gao, Y.; Wells, L.; Comer, F. I.; Parker, G. J.; Hart, G. W. Dynamic O-glycosylation of nuclear and  
513 cytosolic proteins Cloning and characterization of a neutral, cytosolic  $\beta$ -N-acetylglucosaminidase from  
514 human brain. *J. Biol. Chem.* **2001**, *276*, 9838-9845.
- 515 (4) King, D. T.; Males, A.; Davies, G. J.; Vocadlo, D. J. Molecular mechanisms regulating O-linked N-  
516 acetylglucosamine (O-GlcNAc)-processing enzymes. *Curr. Opin. Chem. Biol.* **2019**, *53*, 131-144.
- 517 (5) Wulff-Fuentes, E.; Berendt, R. R.; Massman, L.; Danner, L.; Malard, F.; Vora, J.; Kahsay, R.;  
518 Olivier-Van Stichelen, S. The human O-GlcNAc database and meta-analysis. *Sci. Data* **2021**, *8*, 25.
- 519 (6) Ruan, H. B.; Nie, Y.; Yang, X. Regulation of protein degradation by O-GlcNAcylation: crosstalk  
520 with ubiquitination. *Mol. Cel. Proteom.* **2013**, *12*, 3489-3497.
- 521 (7) Yang, X.; Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev.*  
522 *Molec. Cell Biol.* **2017**, *18*, 452-465.
- 523 (8) Martinez, M. R.; Dias, T. B.; Natov, P. S.; Zachara, N. E. Stress-induced O-GlcNAcylation: an  
524 adaptive process of injured cells. *Biochem. Soc. Trans.* **2017**, *45*, 237-249.
- 525 (9) Hart, G. W. Nutrient regulation of signaling and transcription. *J. Biol. Chem.* **2019**, *294*, 2211-  
526 2231.
- 527 (10) Jiang, M.; Yu, S.; Yu, Z.; Sheng, H.; Li, Y.; Liu, S.; Warner, D. S.; Paschen, W.; Yang, W. XBP1 (X-  
528 box-binding protein-1)-dependent O-GlcNAcylation Is neuroprotective in ischemic stroke in young mice  
529 and its impairment in aged mice Is rescued by thiamet-G. *Stroke* **2017**, *48*, 1646-1654.
- 530 (11) Jensen, R. V.; Andreadou, I.; Hausenloy, D. J.; Botker, H. E. The Role of O-GlcNAcylation for  
531 Protection against Ischemia-Reperfusion Injury. *Int. J. Mol. Sci.* **2019**, *20*, 404.
- 532 (12) Park, J.; Lai, M. K. P.; Arumugam, T. V.; Jo, D. G. O-GlcNAcylation as a Therapeutic Target for  
533 Alzheimer's Disease. *Neuromol. Med.* **2020**, *22*, 171-193.
- 534 (13) Zhao, M. J.; Yao, X.; Wei, P.; Zhao, C.; Cheng, M.; Zhang, D.; Xue, W.; He, W. T.; Xue, W.; Zuo, X.;  
535 Jiang, L. L.; Luo, Z.; Song, J.; Shu, W. J.; Yuan, H. Y.; Liang, Y.; Sun, H.; Zhou, Y.; Zhou, Y.; Zheng, L.; Hu, H.  
536 Y.; Wang, J.; Du, H. N. O-GlcNAcylation of TDP-43 suppresses proteinopathies and promotes TDP-43's  
537 mRNA splicing activity. *EMBO Rep.* **2021**, *22*, e51649.
- 538 (14) Wang, X.; Li, W.; Marcus, J.; Pearson, M.; Song, L.; Smith, K.; Terracina, G.; Lee, J.; Hong, K.-L. K.;  
539 Lu, S. X. MK-8719, a novel and selective O-GlcNAcase inhibitor that reduces the formation of  
540 pathological tau and ameliorates neurodegeneration in a mouse model of tauopathy. *J. Pharmacol. Exp.*  
541 *Ther.* **2020**, *374*, 252-263.
- 542 (15) Selnick, H. G.; Hess, J. F.; Tang, C.; Liu, K.; Schachter, J. B.; Ballard, J. E.; Marcus, J.; Klein, D. J.;  
543 Wang, X.; Pearson, M. Discovery of MK-8719, a potent O-GlcNAcase inhibitor as a potential treatment  
544 for tauopathies. *J. Med. Chem.* **2019**, *62*, 10062-10097.
- 545 (16) Groenevelt, J. M.; Corey, D. J.; FehI, C. Chemical Synthesis and Biological Applications of O-  
546 GlcNAcylated Peptides and Proteins. *Chembiochem* **2021**, *22*, 1854-1870.
- 547 (17) Ju Kim, E. O-GlcNAc Transferase: Structural Characteristics, Catalytic Mechanism and Small-  
548 Molecule Inhibitors. *Chembiochem* **2020**, *21*, 3026-3035.
- 549 (18) Lazarus, M. B.; Jiang, J.; Gloster, T. M.; Zandberg, W. F.; Whitworth, G. E.; Vocadlo, D. J.; Walker,  
550 S. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat. Chem. Biol.* **2012**, *8*,  
551 966-968.

552 (19) Lazarus, M. B.; Jiang, J.; Kapuria, V.; Bhuiyan, T.; Janetzko, J.; Zandberg, W. F.; Vocadlo, D. J.;  
553 Herr, W.; Walker, S. HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **2013**, *342*, 1235-  
554 1239.

555 (20) Dorfmüller, H. C.; Borodkin, V. S.; Schimpl, M.; Zheng, X.; Kime, R.; Read, K. D.; van Aalten, D.  
556 M. Cell-penetrant, nanomolar O-GlcNAcase inhibitors selective against lysosomal hexosaminidases.  
557 *Chem. Biol.* **2010**, *17*, 1250-1255.

558 (21) Roth, C.; Chan, S.; Offen, W. A.; Hemsworth, G. R.; Willems, L. I.; King, D. T.; Varghese, V.;  
559 Britton, R.; Vocadlo, D. J.; Davies, G. J. Structural and functional insight into human O-GlcNAcase. *Nat.*  
560 *Chem. Biol.* **2017**, *13*, 610-612.

561 (22) Li, B.; Li, H.; Lu, L.; Jiang, J. Structures of human O-GlcNAcase and its complexes reveal a new  
562 substrate recognition mode. *Nat. Struct. Mol. Biol.* **2017**, *24*, 362-369.

563 (23) Elsen, N. L.; Patel, S. B.; Ford, R. E.; Hall, D. L.; Hess, F.; Kandula, H.; Kornienko, M.; Reid, J.;  
564 Selnick, H.; Shipman, J. M.; Sharma, S.; Lumb, K. J.; Soisson, S. M.; Klein, D. J. Insights into activity and  
565 inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* **2017**, *13*, 613-615.

566 (24) Borodkin, V. S.; Rafie, K.; Selvan, N.; Aristotelous, T.; Navratilova, I.; Ferenbach, A. T.; Van Aalten,  
567 D. M. O-GlcNAcase fragment discovery with fluorescence polarimetry. *ACS Chem. Biol.* **2018**, *13*, 1353-  
568 1360.

569 (25) Yuzwa, S. A.; Macauley, M. S.; Heinonen, J. E.; Shan, X.; Dennis, R. J.; He, Y.; Whitworth, G. E.;  
570 Stubbs, K. A.; McEachern, E. J.; Davies, G. J. A potent mechanism-inspired O-GlcNAcase inhibitor that  
571 blocks phosphorylation of tau in vivo. *Nat. Chem. Biol.* **2008**, *4*, 483-490.

572 (26) Cekic, N.; Heinonen, J.; Stubbs, K.; Roth, C.; He, Y.; Bennet, A.; McEachern, E.; Davies, G.;  
573 Vocadlo, D. Analysis of transition state mimicry by tight binding aminothiazoline inhibitors provides  
574 insight into catalysis by human O-GlcNAcase. *Chem. Sci.* **2016**, *7*, 3742-3750.

575 (27) Bergeron-Brelek, M.; Meanwell, M.; Britton, R. Direct synthesis of imino-C-nucleoside analogues  
576 and other biologically active iminosugars. *Nat. Commun.* **2015**, *6*, 6903.

577 (28) Macauley, M. S.; Whitworth, G. E.; Debowski, A.; Chin, D.; Vocadlo, D. J. O-GlcNAcase uses  
578 substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired  
579 inhibitors. *J. Biol. Chem.* **2005**, *280*, 25313-25322.

580 (29) He, Y.; Macauley, M. S.; Stubbs, K. A.; Vocadlo, D. J.; Davies, G. J. Visualizing the reaction  
581 coordinate of an O-GlcNAc hydrolase. *J. Am. Chem. Soc.* **2010**, *132*, 1807-1809.

582 (30) Xiong, J.; Xu, D. Mechanistic Insights into the Hydrolysis of O-GlcNAcylation Catalyzed by Human  
583 O-GlcNAcase. *J. Phys. Chem. B* **2020**, *124*, 9310-9322.

584 (31) Stubbs, K. A.; Bacik, J. P.; Perley-Robertson, G. E.; Whitworth, G. E.; Gloster, T. M.; Vocadlo, D. J.;  
585 Mark, B. L. The development of selective inhibitors of NagZ: increased susceptibility of Gram-negative  
586 bacteria to beta-lactams. *Chembiochem* **2013**, *14*, 1973-1981.

587 (32) Macauley, M. S.; He, Y.; Gloster, T. M.; Stubbs, K. A.; Davies, G. J.; Vocadlo, D. J. Inhibition of O-  
588 GlcNAcase using a potent and cell-permeable inhibitor does not induce insulin resistance in 3T3-L1  
589 adipocytes. *Chem. Biol.* **2010**, *17*, 937-948.

590 (33) Mark, B. L.; Mahuran, D. J.; Cherney, M. M.; Zhao, D.; Knapp, S.; James, M. N. Crystal structure of  
591 human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J.*  
592 *Mol. Biol.* **2003**, *327*, 1093-1109.

593 (34) Tiscornia, G.; Vivas, E. L.; Matalonga, L.; Berniakovich, I.; Barragan Monasterio, M.; Eguizabal, C.;  
594 Gort, L.; Gonzalez, F.; Ortiz Mellet, C.; Garcia Fernandez, J. M.; Ribes, A.; Veiga, A.; Izpisua Belmonte, J. C.  
595 Neuronopathic Gaucher's disease: induced pluripotent stem cells for disease modelling and testing  
596 chaperone activity of small compounds. *Hum. Mol. Genet.* **2013**, *22*, 633-645.

597 (35) Suzuki, H.; Ohto, U.; Higaki, K.; Mena-Barragan, T.; Aguilar-Moncayo, M.; Ortiz Mellet, C.; Nanba,  
598 E.; Garcia Fernandez, J. M.; Suzuki, Y.; Shimizu, T. Structural basis of pharmacological chaperoning for  
599 human beta-galactosidase. *J. Biol. Chem.* **2014**, *289*, 14560-14568.

600 (36) Yu, Y.; Mena-Barragan, T.; Higaki, K.; Johnson, J. L.; Drury, J. E.; Lieberman, R. L.; Nakasone, N.;  
601 Ninomiya, H.; Tsukimura, T.; Sakuraba, H.; Suzuki, Y.; Nanba, E.; Mellet, C. O.; Garcia Fernandez, J. M.;  
602 Ohno, K. Molecular basis of 1-deoxygalactonojirimycin arylthiourea binding to human alpha-  
603 galactosidase a: pharmacological chaperoning efficacy on Fabry disease mutants. *ACS Chem. Biol.* **2014**,  
604 *9*, 1460-1469.

605 (37) de la Mata, M.; Cotan, D.; Oropesa-Avila, M.; Garrido-Maraver, J.; Cordero, M. D.; Villanueva  
606 Paz, M.; Delgado Pavon, A.; Alcocer-Gomez, E.; de Lavera, I.; Ybot-Gonzalez, P.; Paula Zaderenko, A.;  
607 Ortiz Mellet, C.; Garcia Fernandez, J. M.; Sanchez-Alcazar, J. A. Pharmacological Chaperones and  
608 Coenzyme Q10 Treatment Improves Mutant beta-Glucocerebrosidase Activity and Mitochondrial  
609 Function in Neuronopathic Forms of Gaucher Disease. *Sci. Rep.* **2015**, *5*, 10903.

610 (38) Risquez-Cuadro, R.; Matsumoto, R.; Ortega-Caballero, F.; Nanba, E.; Higaki, K.; Garcia Fernandez,  
611 J. M.; Ortiz Mellet, C. Pharmacological Chaperones for the Treatment of alpha-Mannosidosis. *J. Med.*  
612 *Chem.* **2019**, *62*, 5832-5843.

613 (39) Aguilar-Moncayo, M.; Takai, T.; Higaki, K.; Mena-Barragan, T.; Hirano, Y.; Yura, K.; Li, L.; Yu, Y.;  
614 Ninomiya, H.; Garcia-Moreno, M. I.; Ishii, S.; Sakakibara, Y.; Ohno, K.; Nanba, E.; Ortiz Mellet, C.; Garcia  
615 Fernandez, J. M.; Suzuki, Y. Tuning glycosidase inhibition through aglycone interactions: pharmacological  
616 chaperones for Fabry disease and GM1 gangliosidosis. *Chem. Commun.* **2012**, *48*, 6514-6516.

617 (40) Mena-Barragan, T.; Narita, A.; Matias, D.; Tiscornia, G.; Nanba, E.; Ohno, K.; Suzuki, Y.; Higaki, K.;  
618 Garcia Fernandez, J. M.; Ortiz Mellet, C. pH-Responsive Pharmacological Chaperones for Rescuing  
619 Mutant Glycosidases. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 11696-11700.

620 (41) Mena-Barragan, T.; Garcia-Moreno, M. I.; Sevsek, A.; Okazaki, T.; Nanba, E.; Higaki, K.; Martin, N.  
621 I.; Pieters, R. J.; Fernandez, J. M. G.; Mellet, C. O. Probing the Inhibitor versus Chaperone Properties of  
622 sp(2)-Iminosugars towards Human beta-Glucocerebrosidase: A Picomolar Chaperone for Gaucher  
623 Disease. *Molecules* **2018**, *23*, 927.

624 (42) Glawar, A. F.; Best, D.; Ayers, B. J.; Miyauchi, S.; Nakagawa, S.; Aguilar-Moncayo, M.; Garcia  
625 Fernandez, J. M.; Ortiz Mellet, C.; Crabtree, E. V.; Butters, T. D.; Wilson, F. X.; Kato, A.; Fleet, G. W.  
626 Scalable syntheses of both enantiomers of DNJNAc and DGJNAc from glucuronolactone: the effect of N-  
627 alkylation on hexosaminidase inhibition. *Eur. J. Chem.* **2012**, *18*, 9341-9359.

628 (43) Anitha, M.; Swamy, K. C. K. Synthesis of thiazolidine-thiones, imino-thiazolidines and  
629 oxazolidines via the base promoted cyclisation of epoxy-sulfonamides and heterocumulenes. *Org.*  
630 *Biomol. Chem.* **2018**, *16*, 402-413.

631 (44) Murphy, D. J. Determination of accurate KI values for tight-binding enzyme inhibitors: an in silico  
632 study of experimental error and assay design. *Anal. Biochem.* **2004**, *327*, 61-67.

633 (45) Dennis, R. J.; Taylor, E. J.; Macauley, M. S.; Stubbs, K. A.; Turkenburg, J. P.; Hart, S. J.; Black, G.  
634 N.; Voadlo, D. J.; Davies, G. J. Structure and mechanism of a bacterial beta-glucosaminidase having O-  
635 GlcNAcase activity. *Nat. Struct. Mol. Biol.* **2006**, *13*, 365-371.

636 (46) Fang, L.; Chakraborty, S.; Dieter, E. M.; Potter, Z. E.; Lombard, C. K.; Maly, D. J. Chemoproteomic  
637 Method for Profiling Inhibitor-Bound Kinase Complexes. *J. Am. Chem. Soc.* **2019**, *141*, 11912-11922.

638 (47) Duncan, J. S.; Gyenis, L.; Lenahan, J.; Bretner, M.; Graves, L. M.; Haystead, T. A.; Litchfield, D. W.  
639 An unbiased evaluation of CK2 inhibitors by chemoproteomics: characterization of inhibitor effects on  
640 CK2 and identification of novel inhibitor targets. *Mol. Cell Proteom.* **2008**, *7*, 1077-1088.

641 (48) Moulick, K.; Ahn, J. H.; Zong, H.; Rodina, A.; Cerchiatti, L.; Gomes DaGama, E. M.; Caldas-Lopes,  
642 E.; Beebe, K.; Perna, F.; Hatzi, K.; Vu, L. P.; Zhao, X.; Zatorska, D.; Taldone, T.; Smith-Jones, P.; Alpaugh,  
643 M.; Gross, S. S.; Pillarsetty, N.; Ku, T.; Lewis, J. S.; Larson, S. M.; Levine, R.; Erdjument-Bromage, H.;



644 Guzman, M. L.; Nimer, S. D.; Melnick, A.; Neckers, L.; Chiosis, G. Affinity-based proteomics reveal cancer-  
645 specific networks coordinated by Hsp90. *Nat. Chem. Biol.* **2011**, *7*, 818-826.

646 (49) Teranishi, Y.; Hur, J. Y.; Welander, H.; Frånberg, J.; Aoki, M.; Winblad, B.; Frykman, S.; Tjernberg,  
647 L. O. Affinity pulldown of  $\gamma$ -secretase and associated proteins from human and rat brain. *J. Cell Mol.*  
648 *Med.* **2010**, *14*, 2675-2686.

649 (50) Becher, I.; Dittmann, A.; Savitski, M. M.; Hopf, C.; Drewes, G.; Bantscheff, M. Chemoproteomics  
650 reveals time-dependent binding of histone deacetylase inhibitors to endogenous repressor complexes.  
651 *ACS Chem. Biol.* **2014**, *9*, 1736-1746.

652 (51) Cruz, I. N.; Barry, C. S.; Kramer, H. B.; Chuang, C. C.; Lloyd, S.; van der Spoel, A. C.; Platt, F. M.;  
653 Yang, M.; Davis, B. G. Glycomimetic affinity-enrichment proteomics identifies partners for a clinically-  
654 utilized iminosugar. *Chem. Sci.* **2013**, *4*, 3442-3446.

655 (52) Dong, D. L.; Hart, G. W. Purification and characterization of an O-GlcNAc selective N-acetyl- $\beta$ -D-  
656 glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **1994**, *269*, 19321-19330.

657 (53) Kelsall, I. R.; Zhang, J.; Knebel, A.; Arthur, J. S. C.; Cohen, P. The E3 ligase HOIL-1 catalyses ester  
658 bond formation between ubiquitin and components of the Myddosome in mammalian cells. *Proc. Natl.*  
659 *Acad. U.S.A.* **2019**, *116*, 13293-13298.

660 (54) Wisniewski, J. R.; Zougman, A.; Mann, M. Nepsilon-formylation of lysine is a widespread post-  
661 translational modification of nuclear proteins occurring at residues involved in regulation of chromatin  
662 function. *Nucl. Acid. Res.* **2008**, *36*, 570-577.

663 (55) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.;  
664 Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.;  
665 Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski,  
666 M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.;  
667 Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold.  
668 *Nature* **2021**, *596*, 583-589.

669 (56) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Cryst.*  
670 *D* **2010**, *66*, 486-501.

671 (57) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.;  
672 Gillette, M. A.; Jaffe, J. D.; Carr, S. A. Integrated proteomic analysis of post-translational modifications by  
673 serial enrichment. *Nat. Meth.* **2013**, *10*, 634-637.

674 (58) Wu, X.; Tian, L.; Li, J.; Zhang, Y.; Han, V.; Li, Y.; Xu, X.; Li, H.; Chen, X.; Chen, J.; Jin, W.; Xie, Y.;  
675 Han, J.; Zhong, C. Q. Investigation of receptor interacting protein (RIP3)-dependent protein  
676 phosphorylation by quantitative phosphoproteomics. *Mol. Cell Proteom.* **2012**, *11*, 1640-1651.

677 (59) Yi, T.; Zhai, B.; Yu, Y.; Kiyotsugu, Y.; Raschle, T.; Etzkorn, M.; Seo, H. C.; Nagiec, M.; Luna, R. E.;  
678 Reinherz, E. L.; Blenis, J.; Gygi, S. P.; Wagner, G. Quantitative phosphoproteomic analysis reveals system-  
679 wide signaling pathways downstream of SDF-1/CXCR4 in breast cancer stem cells. *Proc. Natl. Acad.*  
680 *U.S.A.* **2014**, *111*, E2182-E2190.

681 (60) Elia, A. E.; Boardman, A. P.; Wang, D. C.; Huttlin, E. L.; Everley, R. A.; Dephoure, N.; Zhou, C.;  
682 Koren, I.; Gygi, S. P.; Elledge, S. J. Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the  
683 DNA Damage Response. *Mol. Cell* **2015**, *59*, 867-881.

684 (61) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.;  
685 Comb, M. J.; Harper, J. W.; Gygi, S. P. Systematic and quantitative assessment of the ubiquitin-modified  
686 proteome. *Mol. Cell* **2011**, *44*, 325-340.

687

688

– Supporting Information –

**Bicyclic picomolar OGA inhibitors enable chemoproteomic mapping of its endogenous post-translational modifications.**

Manuel González-Cuesta,<sup>a</sup> Peter Sidhu,<sup>b,c</sup> Roger A. Ashmus,<sup>b</sup> Alexandra Males,<sup>d</sup> Cameron Proceviat,<sup>b</sup> Zarina Madden,<sup>b</sup> Jason C. Rogalski,<sup>c</sup> Jil A. Busmann,<sup>e</sup> Leonard J. Foster,<sup>c</sup> José M. García Fernández,<sup>f</sup> Gideon J. Davies,<sup>d,\*</sup> Carmen Ortiz Mellet,<sup>a,\*</sup> David J. Vocadlo<sup>b,e,\*</sup>

<sup>a</sup> Departamento de Química Orgánica, Facultad de Química, Universidad de Sevilla, 41012, Sevilla, Spain.

<sup>b</sup> Department of Chemistry, Simon Fraser University, Burnaby V5A 1S6, British Columbia, Canada.

<sup>c</sup> Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada.

<sup>d</sup> Department of Chemistry, University of York, York YO10 5DD, United Kingdom.

<sup>e</sup> Department of Molecular Biology and Biochemistry, Simon Fraser University, V5A 1S6, Burnaby, British Columbia, Canada.

<sup>f</sup> Instituto de Investigaciones Químicas (IIQ), CSIC - Universidad de Sevilla, 41092 Sevilla, Spain.

## List of Contents

<b>General Methods</b>	<b>S3-S8</b>
<b>Materials</b>	<b>S8</b>
<b>Synthetic Procedures and characterization data</b>	<b>S9-S31</b>
<b>Morrison plots for <math>K_i</math> determinations</b>	<b>S32-S35</b>
<b>X-Ray structure solution methods</b>	<b>S36-S38</b>
<b>Chemoproteomic and mapping methods and data</b>	<b>S39-S52</b>
<b>Brain permeability data</b>	<b>S53</b>
<b>Copies of <math>^1\text{H}</math> and <math>^{13}\text{C}</math> NMR and MS spectra</b>	<b>S54-S96</b>
<b>References</b>	<b>S97-S99</b>

## General methods

All reagents and solvents were purchased from commercial sources and used without further purification unless otherwise stated. Thin-layer chromatography (TLC) was carried out on aluminium sheets coated with *Silica gel 60 F<sub>254</sub> Merck* with visualization by UV light ( $\lambda$  254 nm) and by charring with 10% ethanolic H<sub>2</sub>SO<sub>4</sub>, 0.1% ethanolic ninhydrin and heating at 100 °C. With preparative purposes, column chromatography was carried out on *Silice 60 A.C.C. Chromagel* (SDS 70-200 and 35-70  $\mu$ m). CombiFlash was performed on CombiFlash Rf 200 from Teledyne ISCO. In reference to 'dry load', the residue was dissolved in appropriate solvent then silica gel was added; this solvent was then evaporated to produce the residue adhered to silica which was loaded onto the column in dry form. HPLC was performed on Agilent 1100 series equipped with a variable wavelength UV-Vis detector using either an Eclipse XDB-C18 column (3.5  $\mu$ m particle size, 3.0  $\times$  150 mm for analytical runs and 5.0  $\mu$ m, 9.4  $\times$  250 mm for semi-preparative scale purifications) or ZORBAX 300SB C8 column (5.0  $\mu$ m particle size, 9.4  $\times$  250 mm for analytical runs and semi-preparative scale purifications) using HPLC grade solvents. Optical rotations were measured at 20  $\pm$  2 °C in 1 cm tubes on a Jasco P-2000 polarimeter using a sodium lamp ( $\lambda$  589 nm). UV spectra were recorded on JASCO V-630 instrument; unit for  $\epsilon$  values: mm<sup>-1</sup>cm<sup>-1</sup>. Elemental analyses were carried out at the Instituto de Investigaciones Químicas (Sevilla, Spain) using an elemental analyser *Leco CHNS-932* o *Leco TruSpec CHN*. NMR experiments were performed at 300 (75.5), 500 (125.7) and 600 (150) MHz with Bruker 300 ADVANCE, 500 DRX and Bruker Avance 600 equipped with a QNP or TCI cryoprobe (600 MHz). 1D TOCSY, 2D COSY, HMQC and HSQC experiments were used to assist on NMR assignments. The chemical shift values are given in ppm (part per million), using the solvent as internal standard, tetramethylsilane (for CDCl<sub>3</sub>). The values of the coupling constant ( $J$ ) are measured in Hz. Abbreviations to indicate the multiplicities of the signals are: s (singlet), bs (broad singlet), d (doublet), t (triplet), q (quartet) and m (multiplet). Mass spectra were carried out on Micromass AutoSpecQ (for chemical ionization, CI) or Bruker Daltonics Esquire6000<sup>TM</sup> (for electrospray ionization, ESI). For CI mass spectrometry, samples were introduced via solid probe heated from 30 to 280 °C. In the case of ESI, methanol was used as the solvent. High resolution mass spectrometer (HRMS) spectra were recorded on a Bruker MaXis Impact or Agilent Technologies 6210 Time-of-Flight LC/MS spectrometers using positive or negative ESI.

**Inhibition assays against commercial  $\beta$ -N-acetylglucosaminidases.** Inhibition constant ( $K_i$ ) values were determined by spectrophotometrically measuring the residual hydrolytic activities of the glycosidases against *p*-nitrophenyl *N*-acetyl- $\beta$ -D-glucosaminide. The  $K_m$  values for the glycosidases used in the tests and the corresponding working pHs are listed herein:  $\beta$ -N-

acetylglucosaminidase (from human placenta),  $K_m = 0.34$  mM (pH 5.5);  $\beta$ -N-acetylglucosaminidase (from bovine kidney),  $K_m = 0.48$  mM (pH 5.5). Each assay was performed in phosphate-citrate buffer and the reactions were initiated by addition of the enzyme to a solution of the substrate in the absence or presence of various inhibitor concentrations. The mixture was incubated for 10-30 min at 37 °C and the reaction was quenched by addition of 1 M  $\text{Na}_2\text{CO}_3$ . Reaction times were appropriate to obtain 10-20% conversion of the substrate in order to achieve linear rates. The absorbance of the resulting mixture was determined at 405 nm. Approximate values of  $K_i$  were determined using a fixed concentration of substrate (around the  $K_m$  value for the different glycosidases) and various inhibitor concentrations. Full  $K_i$  determinations and enzyme inhibition mode were determined from the slope of Lineweaver-Burk plots and double reciprocal analysis using a Microsoft Office Excel 2007 program.

**Inhibition assays against hOGA.** Inhibition constant ( $K_i$ ) values for compounds against hOGA enzyme are determined by measuring the change in fluorescent signal corresponding to the rate of hydrolytic activity against the artificial substrate, Resorufin-N-acetyl- $\beta$ -D-glucosaminide. hOGA activity assays were performed in a buffer of 20 mM HEPES, 5 mM EDTA, 150 mM KCl, pH 7.1 and 0.2 nM [hOGA].  $K_m = 25\mu\text{M}$  was measured with these conditions. Inhibition assays were run at 0.2 nM [hOGA] in the same buffer, in the presence or absence of various concentrations of inhibitors and at a fixed substrate concentration of 25  $\mu\text{M}$  and 1% DMSO. First, inhibitor was serially diluted to the desired range of concentrations in 2% DMSO buffer. Next, 50  $\mu\text{L}$  of inhibitor solutions at various concentrations in 2% DMSO buffer was added to 50  $\mu\text{L}$  of 0.8 nM enzyme and allowed to incubate at 25 °C for 5 minutes. 100  $\mu\text{L}$  of 25  $\mu\text{M}$  substrate was then added and reaction mixture was immediately mixed and aliquoted in 45  $\mu\text{L}$  triplicates to a CORNING 384 well black plate. Fluorescence signal was measured continuously for 20 minutes at 37 °C in a BioTek Neo 2 Plate reader set at excitation and emission wavelengths of 572 and 610 nm, respectively. Maximal reaction rates for all inhibitor concentrations were calculated within Gen5 BioTek reader software. % Activity was subsequently calculated for each inhibitor concentration against the maximal reaction rate of uninhibited hOGA. GraphPad Prism 2016 was used to approximate the Morrison  $K_i$  values for each inhibitor.

**Inhibition assays against rhHexA.** Inhibition constant ( $K_i$ ) values for compounds against recombinant human Hexosaminidase A (HexA) enzyme, purchased from R&D Systems (cat# 6237-GH-020), were determined by measuring the fluorescent signal corresponding to the rate of hydrolytic activity against the commercially available artificial substrate, 4-methylumbelliferone-N-acetyl- $\beta$ -D-glucosaminide. HexA activity assays were performed in a buffer of 100 mM sodium citrate, 250 NaCl, pH 4.5 and 1.0 nM [HexA], then stopped after 20 minutes with a

solution of 1.0 M Tris, pH 9.5 to enhance fluorescent signal. Preliminary experiments have shown that reaction rates are linear for 20 min after substrate addition, and that 1.0 M Tris at pH 9.5 is sufficient for stopping activity, as signal is stable after stopping. Inhibition assays were run in the presence or absence of various concentrations of inhibitors and at a fixed substrate concentration of 100  $\mu$ M and 1% DMSO. First, inhibitor was serially diluted to the desired range of concentrations in 2% DMSO buffer. Next, 25  $\mu$ L of inhibitor solutions at various concentrations in 2% DMSO buffer was added to 25  $\mu$ L of 4.0 nM enzyme and allowed to incubate at 25  $^{\circ}$ C for 5 min. 50  $\mu$ L of 200  $\mu$ M and 1% DMSO substrate was then added and allowed to react for 20 min. 100  $\mu$ L of stop solution was added to reaction mixture, and immediately mixed and aliquoted in 45  $\mu$ L triplicates to a CORNING 384 well black plate. Fluorescence signal was measured in a BioTek Neo 2 Plate reader set at excitation and emission wavelengths of 355 and 450 nm, respectively. % Activity was subsequently calculated for each inhibitor concentration against the fluorescence signal of uninhibited HexA. GraphPad Prism 2016 was used to approximate the Morrison  $K_i$  values for each inhibitor.

**Inhibition assays against rhHexB.** Inhibition constant ( $K_i$ ) values against recombinant human Hexosaminidase B (HexB), purchased from R&D Systems (cat# 8907-GH-020) are determined by measuring the change in fluorescent signal corresponding to the rate of hydrolytic activity against the commercially available artificial substrate, 4-methylumbelliferone N-acetyl- $\beta$ -D-galactosaminide. HexB activity assays were performed in 100 mM MES buffer at pH 5.5 and 5nM [HexB]. Inhibition assays were in the presence or absence of various concentrations of inhibitors and at a fixed substrate concentration of 150  $\mu$ M and 1% DMSO. First, inhibitor was serially diluted to the desired range of concentrations in 2% DMSO buffer. Next, 50  $\mu$ L of inhibitor solutions at various concentrations in 2% DMSO Buffer was added to 50  $\mu$ L of 20 nM enzyme and allowed to incubate at 25  $^{\circ}$ C for 5 min. 100  $\mu$ L of 300  $\mu$ M and 1% DMSO substrate was then added and reaction mixture was immediately mixed and aliquoted in 45  $\mu$ L triplicates to a CORNING 384 well black plate. Fluorescence signal was measured continuously for 20 min at 25  $^{\circ}$ C in a BioTek Neo 2 Plate reader set at excitation and emission wavelengths of 355 and 450 nm, respectively. Max reaction rates for all inhibitor concentrations were calculated within Gen5 BioTek reader software. % Activity was subsequently calculated for each inhibitor concentration against the max reaction rate of uninhibited HexB. GraphPad Prism 2016 was used to approximate the Morrison  $K_i$  values for each inhibitor.

**Inhibition assays against BtGH84.** The inhibition constant ( $K_i$ ) value for compound **16** against recombinant OGA from *Bacteriodes thetaiotaocron* (BtGH84), which was produced and purified according to Dennis *et al*, 2006, was determined spectrophotometrically by measuring the change

in absorbance at 405 nm upon hydrolysis of *p*-nitrophenyl *N*-acetyl- $\beta$ -D-glucosaminide. A Biotek Epoch Microplate Spectrophotometer was used to monitor the enzyme assays that were performed in 384-well clear bottomed plates. The  $K_M$  value used in the tests was 1.09 mM at pH 6.5 (50 mM MES pH 6.5, 200 mM NaCl). The enzyme was incubated for 5 mins at 25 °C in the presence of various inhibitor concentrations after which the reaction was initiated by the addition of the substrate. The reaction was observed over 10 mins to achieve linear rates. An  $IC_{50}$  was conducted using a fixed concentration of substrate (around the  $K_M$  value) to approximate the  $K_i$  value. Full  $K_i$  determinations were obtained from the slope of Lineweaver-Burk plots and taking the reciprocal of the slope on a plot of  $K_{app}$  vs the inhibitor concentration.

**Cell culture.** All cells were cultured at 37°C in a humidified incubator with 5% CO<sub>2</sub> using standard procedures. Briefly, SK-N-SH cells (ATCC) were cultured in EMEM (M5650, Sigma) supplemented with 10% FBS (Gibco) and Pen/Strep antibiotics (Bioshop). Cells were expanded in T175 flasks and media was changed every 3 to 4 days until cells reached 80-90 % confluency. For each passage, cells were washed with warm PBS, treated with the minimal volume of Trypsin/EDTA, incubated for 3 to 5 minutes until cells were fully detached, and re-suspended in media containing FBS. Cells were counted using BioRad Automated Cell Counter TC20 with Trypan blue (Gibco), and either split into a new flask for sub-culturing (1:4 to 1:5 dilution) or plated for treatments. All experiments were carried out with cells having a passage number between P13 and P18.

**Cell plating and treatment.** After counting, the concentration of cells was adjusted and  $2.5 \times 10^6$  cells were plated in 10 cm petri dishes (TC Dish-100, Sarstedt). Dishes were incubated for 2 days before medium was exchanged and inhibitors diluted in medium (final concentrations of 0–3.2  $\mu$ M for **33** and 1  $\mu$ M for Thiamet-G; <0.1% DMSO content) were added. The treated cells were incubated for 2 days before harvesting.

**Cell harvesting.** Culture plates were put on ice, the medium was aspirated, and cold PBS (4 °C, LonzaBioWhittaker) was added to all wells. The cells were scraped off and spun down at 300 g for 5 min at 4 °C (Sorvall Legend Micro 17R, Thermo Scientific). Cell pellets were resuspended in cold PBS buffer (3x the volume of the cell pellet) containing complete protease inhibitor cocktail tablets (Roche) prior to freeze-thaw-lysis. Samples were 3x plunged into liquid nitrogen and thawed in a water bath at 25 °C (280 series, Precision). Following, samples were spun down at 20,800 g at 4 °C for 30 min (centrifuge 5417 C, Eppendorf) and the supernatant was collected in fresh microcentrifuge tubes. The protein concentration of the samples was measured using the Quick Start Bradford 1x Dye Reagent (Biorad) and the Synergy neo2 plate reader (BioTek).

**Immunoblotting.** Samples were diluted in 5x SDS-PAGE loading buffer and heated at 90 °C for 3 min. Following, samples were quickly spun down and separated in Mini-PROTEAN TGX gels (4-15%, 10-well comb, 50 ul, Bio-Rad) at 70 V (PowerPac Basic, BioRad). Proteins were transferred to a 0.45 µm nitrocellulose membrane (Bio-Rad) by wet transfer at 110 V at 4 °C for 70 min. The membrane was incubated with 4% BSA (BioShop) in PBS buffer at 4 °C for 1 hour prior to incubation overnight (4 °C) on a rocking platform (Scilogex) with primary antibodies for fluorescent immunoblotting. Following primary antibodies were used: anti-OGA (1:5,000, rabbit, SAB4200311, Sigma-Aldrich), anti-*O*-GlcNAc clone CTD110.6 (1:3,000, mouse, MABS1254, EMD Millipore), and β-anti-Actin (1:10,000, rabbit, 926-42210, Li-Cor). On the following day, the membrane was washed 3x with PBS-T (0.1% tween 20, BioShop) for each 2 min. The membrane was then incubated with the secondary antibodies in 2% BSA in PBS-T at RT for 1 hour. Following secondary antibodies were used: Goat-anti-rabbit IRDye 800CW (1:20,000, 926-32211, Li-Cor) and Goat-anti-mouse IRDye 680LT (1:20,000, 926-68020, Li-Cor). Membranes were then washed 3x with PBS-T for each 2 min prior to imaging (Odyssey, Li-Cor).

**Expression and purification of recombinant hOGA.** *E. coli* cells that were transformed with a plasmid expressing hexahistidine tagged hOGA (full-length) were inoculated in lysogeny broth (LB) medium containing kanamycin (50 µg/mL), and were incubated aerobically at 37 °C overnight. The following day, 10 L of LB containing kanamycin (50 µg/mL), were inoculated with the overnight culture (1:100) and incubated aerobically at 37 °C to an OD600 of ~1.0. Once reached, cells were cold shocked at 4 °C for 30 min, then had IPTG added (0.1 mM) to induce protein synthesis, and expression induced at 21°C overnight. The cells were harvested by centrifugation (high-speed) at 20,000 RPM at 4°C for 20 min, and cells were flash frozen in liquid nitrogen, and stored at -80°C until needed.

For purification of hOGA, cells were thawed and resuspended in 25 mM HEPES, 500 mM NaCl, 1 mM DTT, pH 7.0, 5 mM imidazole (resuspension buffer), and rocked with lysozyme (1 mg/mL) at 4°C for 30 min. Cells were lysed using a sonic dismembrator at amplitude 60%, for 3 min (20 sec on, 40 sec off). The lysate was cleared by centrifugation (high-speed) at 20,000 RPM at 4 °C for 30 min, and again (ultra-speed) at 45,000 RPM at 4 °C for 40 min. The supernatant was passed through a pre-equilibrated HisTrap FF column (GE Healthcare) in suspension buffer using a peristaltic pump (flow rate 2mL/min). The column loaded with hOGA was purified by a gradient of 0% to 50% of elution buffer (25 mM HEPES, 500 mM NaCl, 1 mM DTT, pH 7.0, and 500 mM imidazole) over 50 mL of volume. Fractions containing hOGA were pooled and dialyzed using a molecular weight cut-off (MWCO) of 30 kDa in final buffer (10 mM HEPES, 250 mM NaCl, 1 mM DTT, pH 7.0). Dialyzed hOGA was concentrated using a Vivaspin of 70 kDa



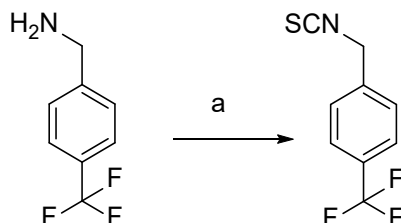
MWCO, and had concentrations determined by Nanodrop 2000 UV-Vis spectrophotometer. Purified hOGA was aliquoted, flash frozen in liquid nitrogen, and stored at -80 °C until required.

**Western blotting.** Samples were diluted in 5x SDS-PAGE loading buffer and heated at 90°C for 3 min. Following, samples were quickly spun down and separated in Mini-PROTEAN TGX gels (4-15%, 10-well comb, 50 µl, Bio-Rad) at 70 V (PowerPac Basic, BioRad). Proteins were transferred to a 0.45 µm nitrocellulose membrane (Bio-Rad) by wet transfer at 110 V at 4 °C for 70 min. The membrane was incubated with 4% BSA (BioShop) in PBS buffer at 4 °C for 1 hour prior to incubation overnight (4 °C) on a rocking platform (Scilogex) with primary antibodies for fluorescent immunoblotting. Following primary antibodies were used: anti-OGA (1:5,000, rabbit, SAB4200311, Sigma-Aldrich), anti-*O*-GlcNAc clone CTD110.6 (1:3,000, mouse, MABS1254, EMD Millipore), and β-anti-Actin (1:10,000, rabbit, 926-42210, Li-Cor). On the following day, the membrane was washed 3x with PBS-T (0.1% tween 20, BioShop) for each 2 min. The membrane was then incubated with the secondary antibodies in 2% BSA in PBS-T at RT for 1 hour. Following secondary antibodies were used: Goat-anti-rabbit IRDye 800CW (1:20,000, 926-32211, Li-Cor) and Goat-anti-mouse IRDye 680LT (1:20,000, 926-68020, Li-Cor). Membranes were then washed 3x with PBS-T for each 2 min prior to imaging (Odyssey, Li-Cor).

## Materials

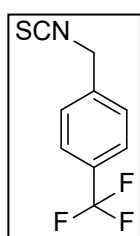
2-Acetamido-1,2-dideoxynojirimicin hydrochloride (DNJNAc HCl), 2-azido-3,6-di-*O*-acetyl-4-*O*-benzyl-1,2,5-trideoxy-1,5-imino-D-glucitol (**25**),<sup>1</sup> 1,4-bis(azidomethyl)benzene<sup>2</sup> and 1-naphthylmethyl isothiocyanate<sup>3</sup> were prepared according to literature procedures. 1,5-Imino-2-propamido-1,2,5-trideoxy-D-glucitol (**34**)<sup>4</sup> was previously characterized. *o*-, *m*- and *p*-Azidomethylbenzylamines were obtained by controlled reduction of the corresponding *o*-, *m*- and *p*-bis(azidomethyl)benzene derivatives with triphenylphosphine, following the method reported by Lau *et al.*<sup>5</sup> They were used in the next isothiocyanation reaction without further purification. The non-commercial *p*-trifluoromethylbenzyl, *o*-, *m*- and *p*-azidomethylbenzyl and 5-indolyl isothiocyanates were synthesized by isothiocyanation of the corresponding amines as detailed hereinafter:

## Synthesis of *p*-Trifluoromethylbenzyl isothiocyanate.



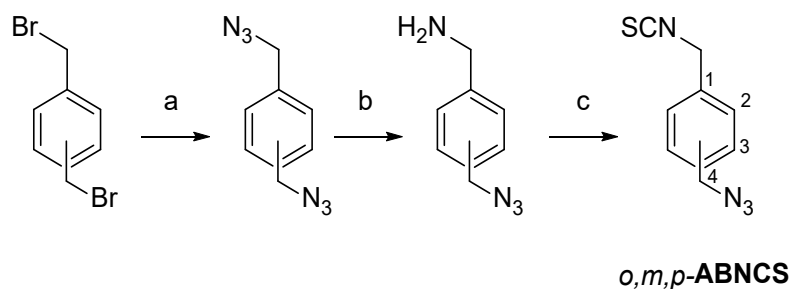
**Supplemental scheme 1.** Reagents and conditions: a. CS<sub>2</sub>, Et<sub>3</sub>N, Boc<sub>2</sub>O, DMAP, EtOH, RT, 18 h.

To a solution of *p*-(trifluoromethyl)benzylamine (213  $\mu$ L, 1.494 mM) in EtOH (8 mL), CS<sub>2</sub> (928  $\mu$ L, 14.940 mM) and Et<sub>3</sub>N (414  $\mu$ L, 2.99 mM) were added. The mixture was stirred at RT for 30 min. Then di-*tert*-butyl dicarbonate (326 mg, 1.49 mmol) and 4-dimethylaminopyridine (3.7 mg, 0.029 mM) were added at 0 °C and the reaction mixture was stirred for 18 h at RT. The solvent was evaporated under reduced pressure and the residue was dissolved with DCM (20 mL), washed with water (20 mL), brine (20 mL), dried (MgSO<sub>4</sub>) and concentrated. The crude product was purified by column chromatography (1:8  $\rightarrow$  1:3 EtOAc- cyclohexane). Yield: 128 mg (60%).  $R_f$  0.66 (1:3 EtOAc-cyclohexane).  $[\alpha]_D +5.6$  ( $c$  1.0 in DCM).



**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>):  $\delta$  = 7.68 (d, 1 H, <sup>3</sup> $J_{H,H}$  = 8.1 Hz, CH<sub>arom</sub>), 7.47 (d, 1H, CH<sub>arom</sub>), 4.81 (s, 2 H, CH<sub>2</sub>NCS). **<sup>13</sup>C NMR** (100.6 MHz, CDCl<sub>3</sub>):  $\delta$  = 138.2 (C-1), 130.9 (C-4), 127.0 (C-2), 126.0 ( $J_{C,F}$  = 3.7 Hz, CF<sub>3</sub>, C-3), 122.0 (NCS). **ESIMS:**  $m/z$  215.8 [M - H]<sup>-</sup>. Anal. Calcd. for C<sub>9</sub>H<sub>6</sub>F<sub>3</sub>NS: C 49.77, H 2.78, F 26.24, N 6.45, S 14.76. Found C 49.63, H 2.56, N 6.27, S 14.49.

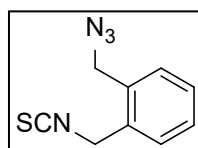
## Synthesis of *o*-, *m*- and *p*-Azidomethylbenzyl isothiocyanates.



**Supplemental scheme 2.** Reagents and conditions: a. NaN<sub>3</sub>, DMF, 60 °C, 18 h; b. PPh<sub>3</sub>, HCl 1M Et<sub>2</sub>O, RT, 24 h; c. CS<sub>2</sub>, Cu<sub>2</sub>SO<sub>4</sub>·5H<sub>2</sub>O, Et<sub>3</sub>N, 2:1 EtOAc-H<sub>2</sub>O, RT, 18 h.

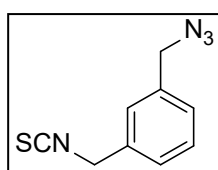
**General procedure for *o,m,p*-ABNCS synthesis.** To a solution of the corresponding *o*-, *m*- or *p*-azidomethylbenzylamine (1.0 g, 6.5 mM) in EtOAc-H<sub>2</sub>O (2:1, 18 mL), CS<sub>2</sub> (3.9 mL, 65 mM) and Et<sub>3</sub>N (1.8 mL, 13 mM) were added and the mixture was stirred at r.t. for 1 h. Then, Cu<sub>2</sub>SO<sub>4</sub>·5H<sub>2</sub>O (413 mg, 1.65 mM) was added and the mixture was further stirred at RT for 16 h. The crude product was filtered, diluted with EtOAc (20 mL), washed with H<sub>2</sub>O (20 mL), dried (MgSO<sub>4</sub>) and concentrated. The crude was purified by column chromatography (1:10 EtOAc-cyclohexane).

***o*-(Azidomethyl)benzyl isothiocyanate (*o*-ABNCS).** Yield: 0.97 g (72%). R<sub>f</sub> 0.51 (1:8 EtOAc-cyclohexane). [α]<sub>D</sub> -1.62 (*c* 1.00 in DCM).



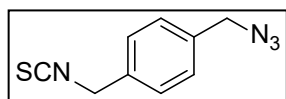
**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>): δ 7.49-7.33 (m, 4 H, CH<sub>arom</sub>), 4.82 (s, 2 H, CH<sub>2</sub>NCS), 4.42 (s, 2 H, CH<sub>2</sub>N<sub>3</sub>). **<sup>13</sup>C NMR** (75.5 MHz, CDCl<sub>3</sub>): δ = 133.2 (NCS), 132.9 (C-1), 132.8 (C-6), 130.2 (C-5), 129.4 (C-2), 129.0 (C-3), 128.8 (C-4), 52.4 (CH<sub>2</sub>NCS), 46.3 (CH<sub>2</sub>N<sub>3</sub>). **ESIMS:** *m/z* 215.8 [M - H]<sup>-</sup>. Anal. Calcd. for C<sub>9</sub>H<sub>8</sub>N<sub>4</sub>S: C 52.92, H 3.95, N 27.43, S 15.70. Found C 53.23, H 4.21, N 27.31, S 15.48.

***m*-(Azidomethyl)benzyl isothiocyanate.** Yield: 0.92 g (68%). R<sub>f</sub> 0.51 (1:8 EtOAc-cyclohexane). [α]<sub>D</sub> -0.71 (*c* 1.00 in DCM).



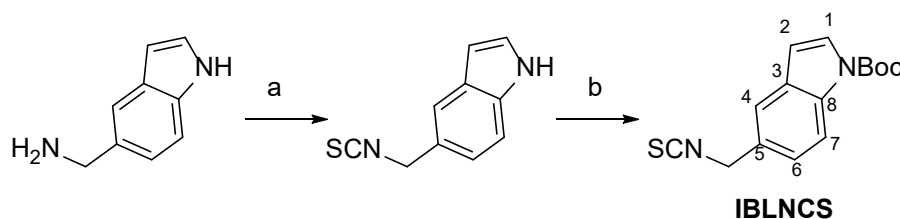
**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>): δ 7.46-7.28 (m, 4 H, CH<sub>arom</sub>), 4.76 (s, 2 H, CH<sub>2</sub>NCS), 4.40 (s, 2 H, CH<sub>2</sub>N<sub>3</sub>). **<sup>13</sup>C NMR** (75.5 MHz, CDCl<sub>3</sub>): δ = 136.4 (C-1), 135.0 (C-5), 132.8 (NCS), 129.5 (C-2), 128.1 (C-3), 126.7 (C-4), 126.4 (C-6), 54.4 (CH<sub>2</sub>NCS), 48.4 (CH<sub>2</sub>N<sub>3</sub>). **ESIMS:** *m/z* = 215.8 [M - H]<sup>-</sup>. Anal. Calcd. for C<sub>9</sub>H<sub>8</sub>N<sub>4</sub>S: C 52.92, H 3.95, N 27.43, S 15.70. Found C 53.18, H 4.18, N 27.26, S 15.44.

***p*-(Azidomethyl)benzyl isothiocyanate.** Yield: 1.0 g (74%);  $R_f$  0.51 (1:8 EtOAc-cyclohexane).  $[\alpha]_D -2.51$  ( $c$  1.00 in DCM).



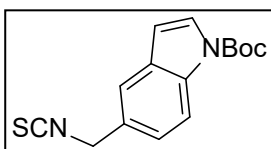
$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.36 (s, 4 H,  $\text{CH}_{\text{arom}}$ ), 4.74 (s, 2 H,  $\text{CH}_2\text{NCS}$ ), 4.37 (s, 2 H,  $\text{CH}_2\text{N}_3$ ).  $^{13}\text{C NMR}$  (100.6 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 135.7 (C-1), 134.3 (C-4), 129.7 (NCS), 128.7 (C-2), 127.3 (C-3), 54.3 ( $\text{CH}_2\text{NCS}$ ), 48.3 ( $\text{CH}_2\text{N}_3$ ). **ESIMS:**  $m/z$  = 215.8 [ $\text{M} - \text{H}$ ] $^-$ . Anal. calcd. for  $\text{C}_9\text{H}_8\text{N}_4\text{S}$ : C 52.92, H 3.95, N 27.43, S 15.70. Found C 53.04, H 3.99, N 27.17, S 15.39.

### Synthesis of *tert*-butyl 5-(isothiocyanatomethyl)-1H-indole-1-carboxylate.



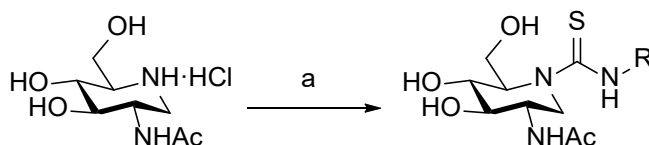
**Supplemental scheme 3.** Reagents and conditions: a.  $\text{CS}_2$ ,  $\text{Et}_3\text{N}$ ,  $\text{Boc}_2\text{O}$ , DMAP, EtOH, RT, 18 h; b.  $\text{Boc}_2\text{O}$ ,  $\text{Et}_3\text{N}$ , DCM, RT, 18 h.

To a solution of (1H-indol-5-yl)methanamine (600 mg, 2 mM) in EtOH (22 mL),  $\text{CS}_2$  (2.4 mL, 20 mM) and  $\text{Et}_3\text{N}$  (1 mL, 4 mM) were added. The mixture was stirred at RT for 30 min. Then di-*tert*-butyl dicarbonate (875 mg, 2 mM) and 4-dimethylaminopyridine (10 mg, 0.04 mM) were added at 0 °C and the reaction mixture was stirred for 18 h at RT. The solvent was evaporated under reduced pressure and the residue was dissolved with DCM (30 mL), washed with water (30 mL), brine (30 mL), dried ( $\text{MgSO}_4$ ) and concentrated. The crude product was directly dissolved in DCM (15 mL) and  $\text{Et}_3\text{N}$  (0.5 mL) and  $\text{Boc}_2\text{O}$  (800 mg) were added and the mixture was stirred for 18 h at RT. The solvent was evaporated and the crude product purified by column chromatography (1:4 EtOAc-cyclohexane). Yield: 300 mg (50%, over 2 steps).  $R_f$  0.50 (1:4 EtOAc-cyclohexane).



$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  8.08 (d, 1 H, H-4), 7.55 (d, 1 H, H-7), 7.43 (s, 1 H, H-6), 7.10 (dd, 1 H, H-1), 6.40 (d, 1 H, H-2), 4.71 (s, 2 H,  $\text{CH}_2\text{NCS}$ ), 1.60 (s, 9 H,  $\text{C}(\text{CH}_3)_3$ ).  $^{13}\text{C NMR}$  (75.5 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 149.5 (CO), 134.9, 130.9, 128.5, 126.9, 123.1, 119.4, 115.6, 107.0 (NCS,  $\text{C}_{\text{arom}}$ ), 84.0 ( $\text{C}(\text{CH}_3)_3$ ), 48.9 ( $\text{CH}_2\text{NCS}$ ), 28.1 ( $\text{C}(\text{CH}_3)_3$ ). **ESIMS:**  $m/z$  = 287.40 [ $\text{M} - \text{H}$ ] $^-$ .

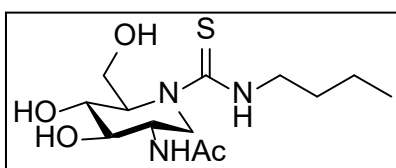
### Synthesis of DNJNAc-thioureas 1-12



**Supplemental scheme 4.** Reagents and conditions: a. Et<sub>3</sub>N, R-NCS, DMF, RT, 18 h.

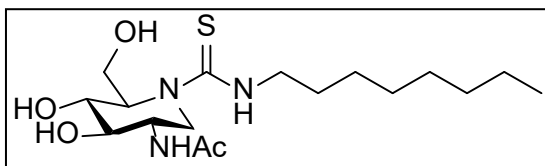
**General procedure for the preparation of the DNJNAc-thioureas 1-12.** To a solution of 2-*N*-acetamido-1,2-dideoxy-5-*NH*·HCl (100 mg, 0.415 mM) in DMF (3.7 mL), Et<sub>3</sub>N (0.115 mL, 0.830 mM) and the corresponding isothiocyanate (0.498 mM, 1.2 eq) were added. The mixture was stirred at RT for 18 h. Then, the solvent was coevaporated with toluene and the resulting residue was purified by column chromatography using the eluent indicated in each case.

**2-Acetamido-5-*N*-(*N*'-butylthiocarbamoyl)-1,2-dideoxy-5-*NH*·HCl (1).** Column chromatography, eluent 80:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 116 mg (86%). [α]<sub>D</sub> -114.7 (*c* 1.04, MeOH). R<sub>f</sub> 0.35 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 249 nm (ε<sub>mM</sub> 12.5).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 4.83 (bd, 1 H, *J*<sub>1a,1b</sub> = 14.4 Hz, H-1a), 4.25 (m, 1 H, H-5), 3.92 (dd, 1 H, *J*<sub>5,6a</sub> = 8.6 Hz, *J*<sub>6a,6b</sub> = 11.4 Hz, H-6a), 3.90 (m, 1 H, H-2), 3.84 (dd, 1 H, *J*<sub>5,6b</sub> = 3.6 Hz, H-6b), 3.67 (dd, 1 H, *J*<sub>3,4</sub> = 6.1 Hz, *J*<sub>4,5</sub> = 4.7 Hz, H-4), 3.57 (m, 2 H, CH<sub>2</sub>NH, H-3), 3.47 (dd, 1 H, *J*<sub>1b,2</sub> = 4.0 Hz, H-1b), 1.95 (s, 1 H, COCH<sub>3</sub>), 1.59 (m, 2 H, CH<sub>2</sub>), 1.38 (m, 2 H, CH<sub>2</sub>CH<sub>3</sub>), 0.95 (t, 3 H, *J*<sub>H,H</sub> = 7.4 Hz, CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 186.6 (CS), 173.0 (CO), 72.8 (C-3), 70.8 (C-4), 66.2 (C-5), 62.2 (C-6), 53.3 (C-2), 46.8 (CH<sub>2</sub>N), 45.2 (C-1), 32.3 (CH<sub>2</sub>), 23.0 (COCH<sub>3</sub>), 21.1 (CH<sub>2</sub>CH<sub>3</sub>), 14.1 (CH<sub>3</sub>). **CIMS:** *m/z* 320 (10, [M + H]<sup>+</sup>). Anal. Calcd for C<sub>13</sub>H<sub>25</sub>N<sub>3</sub>O<sub>4</sub>S: C, 48.88; H, 7.89; N, 13.16; S, 10.04. Found: C, 48.72; H, 7.69; N, 12.83; S, 9.69.

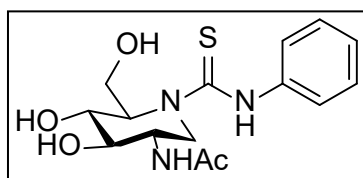
**2-Acetamido-1,2-dideoxy-5-*N*-(*N*'-octylthiocarbamoyl)-5-*NH*·HCl (2).** Column chromatography, eluent 80:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 124 mg (78%). [α]<sub>D</sub> -73.8 (*c* 1.00, MeOH). R<sub>f</sub> 0.39 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 248 nm (ε<sub>mM</sub> 11.4).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 4.83 (bd, 1 H, *J*<sub>1a,1b</sub> = 14.6 Hz, H-1a), 4.23 (m, 1 H, H-5), 3.92 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.4 Hz, *J*<sub>5,6a</sub> = 8.8 Hz, H-6a), 3.88 (m, 1 H, H-2), 3.83 (dd, 1 H, *J*<sub>5,6b</sub> = 3.5 Hz, H-6b), 3.66 (t, 1 H, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 6.0 Hz, H-4), 3.56 (m, 2 H, CH<sub>2</sub>NH, H-3), 3.46 (dd, 1 H, *J*<sub>1b,2</sub> = 4.1 Hz, H-1b), 1.94 (s, 1 H, COCH<sub>3</sub>), 1.60 (m, 2 H, CH<sub>2</sub>), 1.31 (m, 10 H, CH<sub>2</sub>), 0.90 (t, 3 H, <sup>3</sup>*J*<sub>H,H</sub> = 7.0 Hz, CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 186.6 (CS), 173.0 (CO), 72.8 (C-3), 70.8 (C-4), 66.2 (C-5), 62.1 (C-6), 53.2 (C-2), 47.1 (CH<sub>2</sub>N), 45.2 (C-1), 32.9, 30.4, 30.3, 30.1, 28.0, 23.6 (CH<sub>2</sub>), 23.0 (COCH<sub>3</sub>), 14.3 (CH<sub>3</sub>). **ESIMS:** *m/z* 376

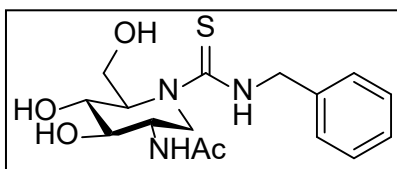
$[M + H]^+$ , 398  $[M + Na]^+$ . Anal. Calcd for  $C_{17}H_{33}N_3O_4S$ : C, 54.37; H, 8.86; N, 11.19; S, 8.54. Found: C, 54.51; H, 8.91; N, 10.94; S 8.22.

**2-Acetamido-1,2-dideoxy-5-*N*-(*N'*-phenylthiocarbamoyl)nojirimycin (3).** Column chromatography, eluent 80:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 118 mg (82%).  $[\alpha]_D -132.8$  (*c* 1.12, MeOH).  $R_f$  0.29 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 258 nm ( $\epsilon_{mM}$  14.9).



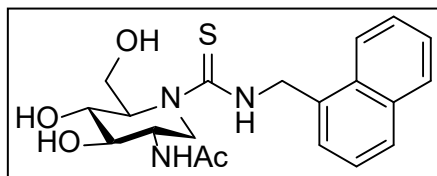
**<sup>1</sup>H NMR** (400 MHz, acetone-*d*<sub>6</sub>):  $\delta$  9.83 (bs, 1 H, NH), 7.88 (bd, 1 H,  $J_{NH,2} = 6.4$  Hz, NH), 7.67-7.08 (m, 5 H, CH<sub>arom</sub>), 5.45 (bs, 1 H, OH), 5.03 (bs, 1 H, OH), 4.91 (d, 1 H,  $J_{1a,1b} = 14.4$  Hz, H-1a), 4.62 (bs, 1 H, OH), 4.40 (m, 1 H, H-5), 4.13 (t, 1 H,  $J_{5,6a} = J_{6a,6b} = 11.1$  Hz, H-6a), 4.04 (dd, 1 H,  $J_{5,6b} = 2.9$  Hz, H-6b), 3.92 (m, 1 H, H-2), 3.75 (t, 1 H,  $J_{3,4} = J_{4,5} = 6.8$  Hz, H-4), 3.63 (dd, 1 H,  $J_{2,3} = 4.7$  Hz, H-3), 3.57 (dd, 1 H,  $J_{1b,2} = 4.4$  Hz, H-1b), 1.93 (s, 1 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (100.6 MHz, acetone-*d*<sub>6</sub>):  $\delta$  185.8 (CS), 171.2 (CO), 141.8, 129.0, 124.9, 124.2 (C<sub>arom</sub>), 74.6 (C-3), 70.5 (C-4), 66.6 (C-5), 62.7 (C-6), 54.4 (C-2), 44.5 (C-1), 23.2 (COCH<sub>3</sub>). **ESIMS**:  $m/z$  362  $[M + Na]^+$ . Anal. Calcd for  $C_{15}H_{21}N_3O_4S$ : C, 53.08; H, 6.24; N, 12.38; S, 9.45. Found: C, 53.20; H, 6.28; N, 12.41; S 9.60.

**2-Acetamido-5-*N*-(*N'*-benzylthiocarbamoyl)-1,2-dideoxynojirimycin (4).** Column chromatography, eluent 80:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 117 mg (78%).  $[\alpha]_D -48.7$  (*c* 0.46, MeOH).  $R_f$  0.33 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 208 nm ( $\epsilon_{mM}$  18.0).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  7.37-7.21 (CH<sub>arom</sub>), 4.89 (m, 2 H, H-1a, CH<sub>2</sub>Ph), 4.78 (d, 1 H,  $^2J_{H,H} = 15$  Hz, CH<sub>2</sub>Ph), 4.38 (m, 1 H, H-5), 3.95 (dd, 1 H,  $J_{6a,6b} = 11.4$  Hz,  $J_{5,6a} = 8.7$  Hz, H-6a), 3.93 (m, 1 H, H-2), 3.83 (dd, 1 H,  $J_{5,6b} = 3.6$  Hz, H-6b), 3.70 (t, 1 H,  $J_{3,4} = J_{4,5} = 5.1$  Hz, H-4), 3.60 (t, 1 H,  $J_{2,3} = 5.1$  Hz, H-3), 3.46 (dd, 1 H,  $J_{1a,1b} = 14.3$  Hz,  $J_{1b,2} = 3.8$  Hz, H-1b), 1.93 (s, 1 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD):  $\delta$  186.9 (CS), 173.0 (CO), 140.0, 129.4, 128.6, 128.0 (C<sub>arom</sub>), 72.3 (C-3), 70.6 (C-4), 66.0 (C-5), 61.8 (C-6), 53.0 (C-2), 50.5 (CH<sub>2</sub>Ph), 45.3 (C-1), 23.0 (COCH<sub>3</sub>). **ESIMS**:  $m/z$  376  $[M + Na]^+$ . Anal. Calcd for  $C_{16}H_{23}N_3O_4S$ : C, 54.37; H, 6.56; N, 11.89; S, 9.07. Found: C, 54.45; H, 6.69; N, 11.76; S 8.84.

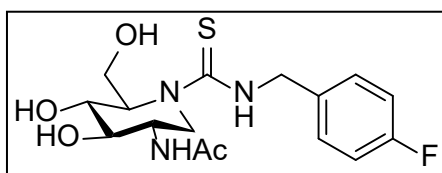
**2-Acetamido-5-*N*-(*N'*-1-naphthylmethylthiocarbamoyl)-1,2-dideoxynojirimycin (5).** Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 90 mg (60%).  $[\alpha]_D -54.6$  (*c* 1.00, MeOH).  $R_f$  0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD):  $\delta$  8.10-7.41 (CH<sub>arom</sub>), 5.27 (bs, 2 H, CH<sub>2</sub>Naph), 4.95 (m, 1 H, H-1a), 4.30 (m, 1 H,

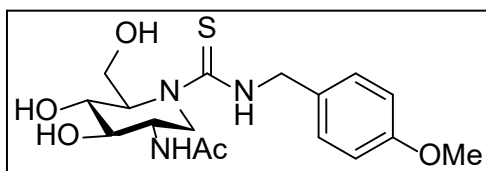
H-5), 3.93 (dd, 1 H,  $J_{6a,6b} = 11.2$  Hz,  $J_{5,6a} = 8.8$  Hz, H-6a), 3.92 (m, 1 H, H-2), 3.82 (dd, 1 H,  $J_{5,6b} = 3.5$  Hz, H-6b), 3.66 (t, 1 H,  $J_{3,4} = J_{4,5} = 5.7$  Hz, H-4), 3.61 (t, 1 H,  $J_{2,3} = 5.4$  Hz, H-3), 3.53 (dd, 1 H,  $J_{1a,1b} = 14.1$  Hz,  $J_{1b,2} = 3.8$  Hz, H-1b), 1.94 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (100.6 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.3 (CS), 171.6 (CO), 133.9-123.2 ( $\text{C}_{\text{arom}}$ ), 70.9 (C-3), 69.2 (C-4), 64.8 (C-5), 60.4 (C-6), 51.6 (C-2), 48.4 ( $\text{CH}_2\text{Naph}$ ), 43.9 (C-1), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  401.9  $[\text{M} - \text{H}]^-$ . Anal. Calcd for  $\text{C}_{20}\text{H}_{25}\text{N}_3\text{O}_4\text{S}$ : C 59.53, H 6.25, N 10.41, S 7.95. Found: C, 59.42; H, 6.13; N, 10.29; S 7.80.

**2-Acetamido-5-*N*-(*N'*-*p*-fluorobenzylthiocarbamoyl)-1,2-dideoxynojirimycin (6).** Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 128 mg (83%).  $[\alpha]_{\text{D}} -97.5$  ( $c$  1.00, MeOH).  $R_f$  0.33 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 227 nm ( $\epsilon_{\text{mM}}$  4.8).



$^1\text{H NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.36, 7.04 ( $\text{CH}_{\text{arom}}$ ), 4.92 (m, 1 H, H-1a), 4.76 (d, 2 H,  $^3J_{\text{H,H}} = 15$  Hz,  $\text{CH}_2\text{Ph}$ ), 4.38 (m, 1 H, H-5), 3.95 (dd, 1 H,  $J_{6a,6b} = 11.1$  Hz,  $J_{5,6a} = 8.9$  Hz, H-6a), 3.93 (m, 1 H, H-2), 3.83 (dd, 1 H,  $J_{5,6b} = 3.7$  Hz, H-6b), 3.70 (t, 1 H,  $J_{3,4} = J_{4,5} = 5.2$  Hz, H-4), 3.60 (t, 1 H,  $J_{2,3} = J_{3,4} = 5.2$  Hz, H-3), 3.46 (dd, 1 H,  $J_{1a,1b} = 14.3$  Hz,  $J_{1b,2} = 3.8$  Hz, H-1b), 1.95 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (125.7 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.4 (CS), 171.6 (CO), 162.2, 134.8, 129.0, 114.6 ( $\text{C}_{\text{arom}}$ ), 70.9 (C-3), 69.2 (C-4), 64.5 (C-5), 60.4 (C-6), 51.6 (C-2), 48.2 ( $\text{CH}_2\text{Ph}$ ), 43.9 (C-1), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  394.1  $[\text{M} + \text{Na}]^+$ . Anal. Calcd for  $\text{C}_{16}\text{H}_{22}\text{FN}_3\text{O}_4\text{S}$ : C 51.74, H 5.97, N 11.31, S 8.63. Found: C, 51.57; H, 6.10; N, 11.19; S 8.36.

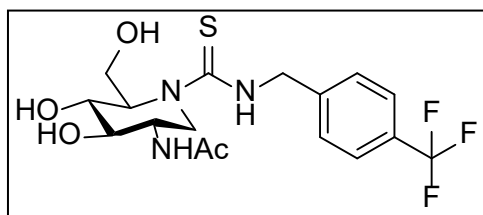
**2-Acetamido-5-*N*-(*N'*-*p*-methoxybenzylthiocarbamoyl)-1,2-dideoxynojirimycin (7).** Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 150 mg (94%).  $[\alpha]_{\text{D}} -72.2$  ( $c$  1.00, MeOH).  $R_f$  0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 230 nm ( $\epsilon_{\text{mM}}$  31.4).



$^1\text{H NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.27 (m, 2 H,  $\text{CH}_{\text{arom}}$ ), 6.8 (m, 2 H,  $\text{CH}_{\text{arom}}$ ), 4.91 (m, 1 H, H-1a), 4.76 (d, 2 H,  $^3J_{\text{H,H}} = 15$  Hz,  $\text{CH}_2\text{Ph}$ ), 4.34 (m, 1 H, H-5), 3.95 (dd, 1 H,  $J_{6a,6b} = 11.2$  Hz,  $J_{5,6a} = 8.7$  Hz, H-6a), 3.93 (m, 1 H, H-2), 3.82 (dd, 1 H,  $J_{5,6b} = 3.7$  Hz, H-6b), 3.78 (s, 3 H,  $\text{OCH}_3$ ), 3.69 (t, 1 H,  $J_{3,4} = J_{4,5} = 5.4$  Hz, H-4), 3.60 (t, 1 H,  $J_{2,3} = J_{3,4} = 5.4$  Hz, H-3), 3.46 (dd, 1 H,  $J_{1a,1b} = 14.4$  Hz,  $J_{1b,2} = 3.7$  Hz, H-1b), 1.95 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (125.7 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.2 (CS), 171.6 (CO), 158.9, 130.5, 128.5, 113.4 ( $\text{C}_{\text{arom}}$ ), 71.0 (C-3), 69.2 (C-4), 64.6 (C-5), 60.4 (C-6), 54.2 ( $\text{OCH}_3$ ), 51.6 (C-2), 48.7 ( $\text{CH}_2\text{Ph}$ ), 43.9 (C-1), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  406.2  $[\text{M} + \text{Na}]^+$ . Anal. Calcd for  $\text{C}_{17}\text{H}_{25}\text{N}_3\text{O}_5\text{S}$ : C 53.25, H 6.57, N 10.96, S 8.36. Found: C, 53.35; H, 6.72; N, 10.80; S 8.09.

**2-Acetamido-5-*N*-(*N'*-*p*-trifluoromethylbenzylthiocarbamoyl)-1,2-dideoxynojirimycin (8).**

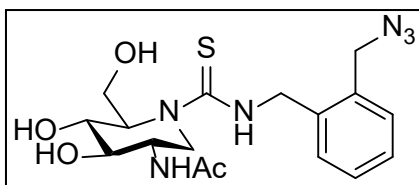
Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 87 mg (55%). [ $\alpha$ ]<sub>D</sub> +4.2 (*c* 1.00, MeOH). *R*<sub>f</sub> 0.31 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 223 nm ( $\epsilon_{mM}$  15.4).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  7.61 (d, 2 H,  $J_{H,H}$  = 7.9 Hz, CH<sub>arom</sub>), 7.52 (d, 2 H, CH<sub>arom</sub>), 5.01 (d, 2 H,  $^3J_{H,H}$  = 15 Hz, CH<sub>2</sub>Ph), 4.91 (m, 1 H, H-1a), 4.47 (m, 1 H, H-5), 4.12 (m, 1 H, H-2), 3.99 (dd, 1 H,  $J_{6a,6b}$  = 11.2 Hz,  $J_{5,6a}$  = 8.7 Hz, H-6a), 3.86 (dd, 1 H,  $J_{5,6b}$  = 3.8 Hz, H-6b), 3.75 (t, 1 H,  $J_{3,4} = J_{4,5} = 5.2$  Hz, H-4), 3.63 (dd, 1 H,  $J_{2,3} = 7.2$  Hz,  $J_{3,4} = 5.2$  Hz, H-3), 3.46 (dd, 1 H,  $J_{1a,1b} = 14.1$  Hz,  $J_{1b,2} = 3.4$  Hz, H-1b), 1.93 (s, 3 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD):  $\delta$  185.7 (CS), 171.6 (CO), 143.7, 127.5 (C<sub>arom</sub>), 124.7 (q, CF<sub>3</sub>), 113.4 (C<sub>arom</sub>), 70.7 (C-3), 69.2 (C-4), 64.4 (C-5), 60.3 (C-6), 51.5 (C-2), 48.3 (CH<sub>2</sub>Ph), 43.9 (C-1), 21.6 (COCH<sub>3</sub>). **ESIMS**: *m/z* 444.1 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>22</sub>F<sub>3</sub>N<sub>3</sub>O<sub>4</sub>S: C 48.45, H 5.26, N 9.97, S 7.61. Found: C, 48.36; H, 5.17; N, 9.72; S 7.36.

**2-Acetamido-5-*N*-(*N'*-*o*-azidomethylbenzylthiocarbamoyl)-1,2-dideoxynojirimycin (9).**

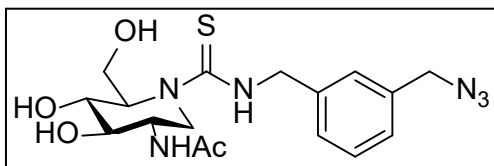
Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 114 mg (75%). [ $\alpha$ ]<sub>D</sub> -44.0 (*c* 1.00, MeOH). *R*<sub>f</sub> 0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 230 nm ( $\epsilon_{mM}$  20.3).



**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD):  $\delta$  7.43-7.28 (d, 4 H, CH<sub>arom</sub>), 4.96 (bd, 1 H, H-1a), 4.91 (d, 2 H, CH<sub>2</sub>Ph), 4.53 (s, 2 H, CH<sub>2</sub>N<sub>3</sub>), 4.37 (m, 1 H, H-5), 3.97 (dd, 1 H,  $J_{6a,6b} = 11.3$  Hz,  $J_{5,6a} = 8.5$  Hz, H-6a), 3.94 (m, 1 H, H-2), 3.82 (dd, 1 H,  $J_{5,6b} = 3.8$  Hz, H-6b), 3.70 (bdd, 1 H,  $J_{3,4} = J_{4,5} = 5.4$  Hz, H-4), 3.61 (t, 1 H,  $J_{2,3} = 5.4$  Hz, H-3), 3.52 (dd, 1 H,  $J_{1a,1b} = 14.2$  Hz,  $J_{1b,2} = 3.7$  Hz, H-1b), 1.95 (s, 3 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (100.6 MHz, CD<sub>3</sub>OD):  $\delta$  185.2 (CS), 171.6 (CO), 137.1, 133.5, 129.4, 128.6, 128.3, 127.1 (C<sub>arom</sub>), 70.8 (C-3), 69.2 (C-4), 64.5 (C-5), 60.4 (C-6), 51.9 (CH<sub>2</sub>N<sub>3</sub>), 51.5 (C-2), 46.4 (CH<sub>2</sub>Ph), 43.9 (C-1), 21.6 (COCH<sub>3</sub>). **ESIMS**: *m/z* 431.2 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>24</sub>N<sub>6</sub>O<sub>4</sub>S: C 49.99, H 5.92, N 20.57, S 7.85. Found: C, 49.81; H, 5.80; N, 20.36; S 7.64.

**2-Acetamido-5-*N*-(*N'*-*m*-azidomethylbenzylthiocarbamoyl)-1,2-dideoxynojirimycin (10).**

Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 94 mg (65%). [ $\alpha$ ]<sub>D</sub> -64.3 (*c* 1.00, MeOH). *R*<sub>f</sub> 0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O). UV (MeOH) 230 nm ( $\epsilon_{mM}$  31.4).



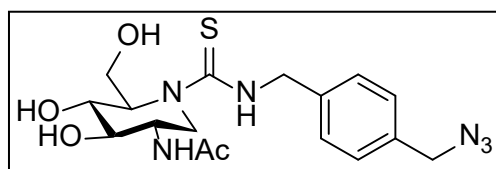
**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD):  $\delta$  7.27-7.11 (m, 4 H, CH<sub>arom</sub>), 4.83 (bd, 1 H, H-1a), 4.78 (bd, 2 H, CH<sub>2</sub>Ph), 4.29 (m, 1 H, H-5), 4.24 (s, 2 H, CH<sub>2</sub>N<sub>3</sub>), 3.86 (dd, 1



H,  $J_{6a,6b} = 11.3$  Hz,  $J_{5,6a} = 8.6$  Hz, H-6a), 3.83 (m, 1 H, H-2), 3.72 (dd, 1 H,  $J_{5,6b} = 3.8$  Hz, H-6b), 3.61 (bdd, 1 H,  $J_{3,4} = J_{4,5} = 5.3$  Hz, H-4), 3.50 (t, 1 H,  $J_{2,3} = 5.3$  Hz, H-3), 3.42 (dd, 1 H,  $J_{1a,1b} = 14.2$  Hz,  $J_{1b,2} = 3.8$  Hz, H-1b), 1.82 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C}$  NMR (100.6 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.5 (CS), 171.6 (CO), 139.5, 135.9, 128.4, 127.1, 126.9, 126.6 ( $\text{C}_{\text{arom}}$ ), 70.8 (C-3), 69.2 (C-4), 64.5 (C-5), 60.3 (C-6), 54.1 ( $\text{CH}_2\text{N}_3$ ), 51.5 (C-2), 48.8 ( $\text{CH}_2\text{Ph}$ ), 43.8 (C-1), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  431.2  $[\text{M} + \text{Na}]^+$ . Anal. Calcd for  $\text{C}_{17}\text{H}_{24}\text{N}_6\text{O}_4\text{S}$ : C 49.99, H 5.92, N 20.57, S 7.85. Found: C, 49.75; H, 5.67; N, 20.20; S 7.51.

**2-Acetamido-5-*N*-(*N'*-*p*-azidomethylbenzylthiocarbamoyl)-1,2-dideoxynojirimycin (11).**

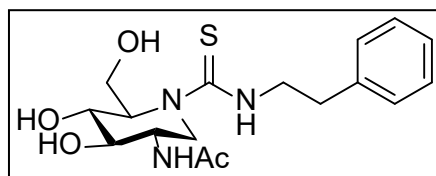
Column chromatography, eluent 100:10:1 DCM-MeOH- $\text{H}_2\text{O}$ . Yield: 100 mg (60%).  $[\alpha]_{\text{D}} -54.0$  ( $c$  1.00, MeOH).  $R_f$  0.40 (70:10:1 DCM-MeOH- $\text{H}_2\text{O}$ ). UV (MeOH) 230 nm ( $\epsilon_{\text{mM}}$  31.4).



$^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.39 (d, 2 H,  $J_{\text{H,H}} = 8.2$  Hz,  $\text{CH}_{\text{arom}}$ ), 7.32 (d, 2 H,  $\text{CH}_{\text{arom}}$ ), 4.94 (bd, 1 H, H-1a), 4.82 (d, 2 H,  $^2J_{\text{H,H}} = 15.0$  Hz,  $\text{CH}_2\text{Ph}$ ), 4.43 (m, 1 H, H-5), 4.35 (s, 2 H,  $\text{CH}_2\text{N}_3$ ), 3.98 (dd, 1 H,  $J_{6a,6b} = 11.3$  Hz,  $J_{5,6a} = 8.5$  Hz, H-6a), 3.94 (m, 1 H, H-2), 3.85 (dd, 1 H,  $J_{5,6b} = 3.7$  Hz, H-6b), 3.73 (bdd, 1 H,  $J_{3,4} = J_{4,5} = 5.4$  Hz, H-4), 3.62 (t, 1 H,  $J_{2,3} = 5.4$  Hz, H-3), 3.54 (dd, 1 H,  $J_{1a,1b} = 14.2$  Hz,  $J_{1b,2} = 3.7$  Hz, H-1b), 1.95 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C}$  NMR (100.6 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.4 (CS), 171.6 (CO), 139.0, 134.4, 128.1, 127.1 ( $\text{C}_{\text{arom}}$ ), 70.8 (C-3), 69.2 (C-4), 64.5 (C-5), 60.4 (C-6), 53.9 ( $\text{CH}_2\text{N}_3$ ), 51.6 (C-2), 48.7 ( $\text{CH}_2\text{Ph}$ ), 43.9 (C-1), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  431.2  $[\text{M} + \text{Na}]^+$ . Anal. Calcd for  $\text{C}_{17}\text{H}_{24}\text{N}_6\text{O}_4\text{S}$ : C 49.99, H 5.92, N 20.57, S 7.85. Found: C, 49.75; H, 5.67; N, 20.20; S 7.51.

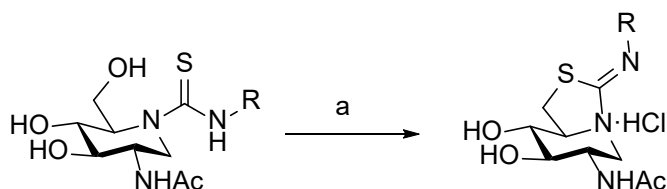
**2-Acetamido-5-*N*-(*N'*-2-phenylethylthiocarbamoyl)-1,2-dideoxynojirimycin (12).**

Column chromatography, eluent 100:10:1 DCM-MeOH- $\text{H}_2\text{O}$ . Yield: 120 mg (78%).  $[\alpha]_{\text{D}} +4.29$  ( $c$  1.00, MeOH).  $R_f$  0.42 (70:10:1 DCM-MeOH- $\text{H}_2\text{O}$ ). UV (MeOH) 223 nm ( $\epsilon_{\text{mM}}$  15.4).



$^1\text{H}$  NMR (300 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.33-7.17 (m, 5 H,  $\text{CH}_{\text{arom}}$ ), 4.86 (bd, 1 H, H-1a), 4.25 (m, 1 H, H-5), 3.92 (m, 1 H, H-2), 3.91 (dd, 1 H,  $J_{6a,6b} = 11.2$  Hz,  $J_{5,6a} = 8.3$  Hz, H-6a), 3.81 (m, 3 H, H-6b,  $\text{CH}_2\text{CH}_2\text{Ph}$ ), 3.69 (dd, 1 H,  $J_{3,4} = 5.9$  Hz,  $J_{4,5} = 4.6$  Hz, H-4), 3.57 (bdd, 1 H, H-3), 3.48 (dd, 1 H,  $J_{1a,1b} = 14.2$  Hz,  $J_{1b,2} = 3.9$  Hz, H-1b), 2.92 (m, 2 H,  $\text{CH}_2\text{CH}_2\text{Ph}$ ), 1.97 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C}$  NMR (75.5 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  185.0 (CS), 171.6 (CO), 139.2, 128.4, 128.0, 125.9 ( $\text{C}_{\text{arom}}$ ), 71.1 (C-3), 69.2 (C-4), 64.4 (C-5), 60.4 (C-6), 51.6 (C-2), 47.1 ( $\text{CH}_2\text{CH}_2\text{Ph}$ ), 43.8 (C-1), 34.9 ( $\text{CH}_2\text{CH}_2\text{Ph}$ ), 21.6 ( $\text{COCH}_3$ ). **ESIMS**:  $m/z$  390.2  $[\text{M} + \text{Na}]^+$ . Anal. Calcd for  $\text{C}_{17}\text{H}_{25}\text{N}_3\text{O}_4\text{S}$ : C 55.57, H 68.6, N 11.44, S 8.72. Found: C, 48.36; H, 5.17; N, 9.72; S 7.36.

## Synthesis of DNJC-thiazolidines 13-24

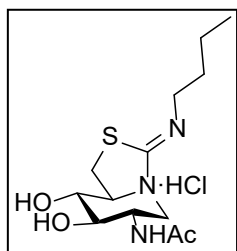


**Supplemental scheme 5.** Reagents and conditions: a.  $\text{HCl}_{(\text{conc})}$ , MeOH, RT, 18 h, quant.

**General procedure for the preparation of the DNJNAc-thiazolidines 13-24.** The corresponding DNJNAc thiourea **1-12** (0.27 mM) was dissolved in MeOH (2 mL) and concentrated HCl was dropwise added until pH 1. The solution was stirred at RT until complete disappearance of the starting material (18-24 h). The solvent was removed under reduced pressure, the residue was coevaporated with MeOH (3 x 10 mL) and the product was purified by column chromatography using the eluent indicated in each case.

### **(Z)-2-Acetamido-5-N,6-S-(N'-butyliminomethylidene)-1,2-dideoxy-6-thionojirimycin**

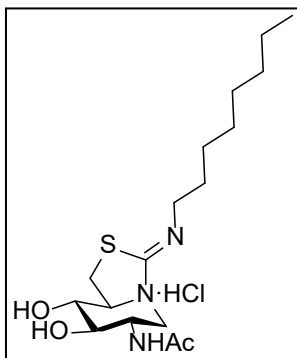
**Hydrochloride (13).** Column chromatography, eluent: 60:10:1  $\rightarrow$  50:10:1  $\rightarrow$  40:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 81 mg (quantitative).  $[\alpha]_{\text{D}} +41.6$  (*c* 0.88, MeOH).  $R_f$  0.25 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  4.11 (dd, 1 H,  $J_{1a,1b} = 13.2$  Hz,  $J_{1a,2} = 5.3$  Hz, H-1a), 3.99 (m, 1 H, H-5), 3.83 (m, 1 H, H-2), 3.76 (dd, 1 H,  $J_{6a,6b} = 11.4$  Hz,  $J_{5,6a} = 7.7$  Hz, H-6a), 3.51 (m, 2 H, H-3, H-6b), 3.43 (t, 1 H,  $J_{3,4} = J_{4,5} = 9.4$ , H-4), 3.37 (t, 2 H, CH<sub>2</sub>NH), 3.01 (dd, 1 H,  $J_{1b,2} = 11.7$  Hz, H-1b), 2.01 (s, 1 H, COCH<sub>3</sub>), 1.66 (m, 2 H, CH<sub>2</sub>), 1.41 (m, 2 H, CH<sub>2</sub>), 0.97 (t, 3 H,  $^3J_{\text{H,H}} = 7.4$  Hz, CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD):  $\delta$  173.9 (CO), 171.8 (CN), 75.4 (C-3), 74.7 (C-4), 69.1 (C-5), 50.9 (CH<sub>2</sub>N), 50.4 (C-2), 46.9 (C-1), 32.2 (C-6), 32.1 (CH<sub>2</sub>), 22.7 (COCH<sub>3</sub>), 20.8 (CH<sub>2</sub>), 13.9 (CH<sub>3</sub>). **ESIMS:**  $m/z$  302 [M + H]<sup>+</sup>, 324 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>13</sub>H<sub>23</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C, 46.21; H, 7.16; N, 12.44; S, 9.49. Found: C, 46.09; H, 7.22; N, 12.15; S, 9.13.

**(Z)-2-Acetamido-1,2-dideoxy-5-N,6-S-(N'-octyliminomethylidene)-6-thionojirimycin**

**Hydrochloride (14).** Column chromatography, eluent: 70:10:1 → 60:10:1 → 50:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 96 mg (quantitative).  $[\alpha]_D^{25} +36.3$  (*c* 0.98, MeOH). *R<sub>f</sub>* 0.53 (40:10:1 DCM-MeOH-H<sub>2</sub>O).

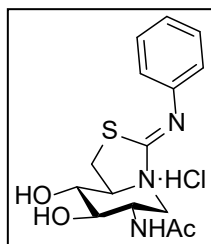


**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  4.10 (dd, 1 H,  $J_{1a,1b} = 13.2$  Hz,  $J_{1a,2} = 5.3$  Hz, H-1a), 3.96 (m, 1 H, H-5), 3.82 (m, 1 H, H-2), 3.73 (dd, 1 H,  $J_{6a,6b} = 11.4$  Hz,  $J_{5,6a} = 7.7$  Hz, H-6a), 3.49 (m, 2 H, H-6b, H-3), 3.42 (t, 1 H,  $J_{3,4} = J_{4,5} = 7.4$  Hz, H-4), 3.35 (t, 2 H, CH<sub>2</sub>NH), 2.98 (t, 1 H,  $J_{1b,2} = J_{1a,1b} = 13.2$  Hz, H-1b), 2.00 (s, 1 H, COCH<sub>3</sub>), 1.67 (m, 2 H, CH<sub>2</sub>), 1.33 (m, 10 H, CH<sub>2</sub>), 0.91 (t, 3 H,  $^3J_{H,H} = 6.8$  Hz, CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD):  $\delta$  173.9 (CO), 171.4 (CN), 75.5 (C-3),

74.8 (C-4), 67.0 (C-5), 50.9 (CH<sub>2</sub>N), 50.9 (C-2), 46.9 (C-1), 32.9 (C-6), 32.0, 30.3, 30.3, 30.2, 27.7, 23.7 (CH<sub>2</sub>), 22.7 (COCH<sub>3</sub>), 14.4 (CH<sub>3</sub>). **ESIMS:** *m/z* 358 (100, [M + H]<sup>+</sup>). Anal. Calcd for C<sub>17</sub>H<sub>31</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C, 51.83; H, 8.19; N, 10.67; S, 8.14. Found: C, 52.00; H, 8.31; N, 10.44; S, 7.97.

**(Z)-2-Acetamido-1,2-dideoxy-5-N,6-S-(N'-phenyliminomethylidene)-6-thionojirimycin**

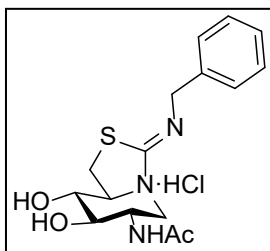
**Hydrochloride (15).** Column chromatography, eluent: 60:10:1 → 50:10:1 → 40:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 86 mg (quantitative).  $[\alpha]_D^{25} +12.2$  (*c* 0.89, MeOH). *R<sub>f</sub>* 0.23 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (500 MHz, 10:1 CD<sub>3</sub>CN-D<sub>2</sub>O):  $\delta$  7.32 (t, 2 H,  $^3J_{H,H} = 7.6$  Hz, CH<sub>arom</sub>), 7.14 (t, 1 H, CH<sub>arom</sub>), 7.0 (d, 2 H, CH<sub>arom</sub>), 4.21 (dd, 1 H,  $J_{1a,1b} = 13.1$  Hz,  $J_{1a,2} = 5.3$  Hz, H-1a), 3.81 (m, 1 H, H-2), 3.65 (m, 1 H, H-5), 3.47 (t, 1 H,  $J_{2,3} = J_{3,4} = 9.2$  Hz, H-3), 3.43 (dd, 1 H,  $J_{6a,6b} = 11.3$  Hz,  $J_{5,6a} = 7.2$  Hz, H-6a), 3.39 (t, 1 H,  $J_{4,5} = 9.2$ , H-4), 3.18 (dd, 1 H,  $J_{5,6b} = 6.9$  Hz, H-6b), 2.79 (dd, 1 H,  $J_{1b,2} = 11.5$  Hz, H-1b), 1.94 (s, 1 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, 10:1 CD<sub>3</sub>CN-D<sub>2</sub>O):  $\delta$  173.4 (CO), 164.1 (CN), 130.2-118.7 (C<sub>arom</sub>), 75.8 (C-3), 74.3 (C-4), 65.9 (C-5), 50.8 (C-2), 46.7 (C-1), 30.9 (C-6), 23.1 (COCH<sub>3</sub>). **ESIMS:** *m/z* 322 [M + H]<sup>+</sup>, 344 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>15</sub>H<sub>19</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C, 50.35; H, 5.63; N, 11.74; S, 8.96. Found: C, 50.67; H, 5.81; N, 11.42; S, 8.35.

**(Z)-2-Acetamido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin**

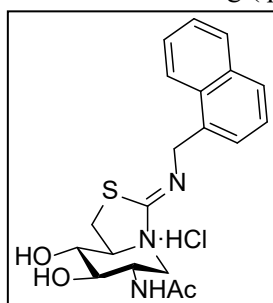
**Hydrochloride (16).** Column chromatography, eluent: 60:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 90 mg (quantitative).  $[\alpha]_D^{25} +36.2$  (*c* 1.04, MeOH). *R<sub>f</sub>* 0.39 (40:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  7.37-7.27 (m, 5 H, CH<sub>arom</sub>), 4.49 (m, 2 H, CH<sub>2</sub>Ph), 4.16 (dd, 1 H, *J*<sub>1a,1b</sub> = 13.1 Hz, *J*<sub>1a,2</sub> = 5.3 Hz, H-1a), 3.85 (m, 1 H, H-2), 3.77 (m, 1 H, H-5), 3.63 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.2 Hz, *J*<sub>5,6a</sub> = 7.3 Hz, H-6a), 3.48 (t, 1 H, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> = 9.6, H-3), 3.38 (m, 1 H, H-4, H-6b), 2.87 (dd, 1 H, *J*<sub>1b,2</sub> = 11.7 Hz, H-1b), 1.99 (s, 1 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD):  $\delta$  173.8 (CO), 171.4 (CN), 138.5, 129.7, 128.8 (C<sub>arom</sub>), 76.0 (C-3), 74.0 (C-4), 68.0 (C-5), 55.5 (CH<sub>2</sub>Ph), 51.1 (C-2), 47.1 (C-1), 31.9 (C-6), 22.7 (COCH<sub>3</sub>). **ESIMS:** *m/z* 336 [M + H]<sup>+</sup>, 358 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>16</sub>H<sub>22</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C, 51.68; H, 5.96; N, 11.30; S, 8.62. Found: C, 51.82; H, 6.14; N, 11.09; S, 8.37.

**(Z)-2-Acetamido-5-N,6-S-(N'-1-naphthymethyliminomethylidene)-1,2-dideoxy-6-**

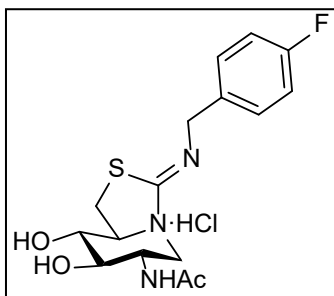
**thionojirimycin Hydrochloride (17).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 104 mg (quantitative).  $[\alpha]_D^{25} +4.8$  (*c* 1.00, 1:1 DCM-MeOH). *R<sub>f</sub>* 0.62 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (300 MHz, 8:1 CD<sub>3</sub>OD-CDCl<sub>3</sub>):  $\delta$  8.03-7.43 (m, 2 H, CH<sub>arom</sub>), 4.89 (bs, 2H, CH<sub>2</sub>Naph), 4.20 (dd, 1 H, *J*<sub>1a,1b</sub> = 13.1 Hz, *J*<sub>1a,2</sub> = 5.2 Hz, H-1a), 3.86 (m, 1 H, H-2), 3.65 (m, 1 H, H-5), 3.58 (m, 1 H, H-6a), 3.46 (t, 1 H, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> = 9.7, H-3), 3.38 (m, 2 H, H-4, H-6b), 2.78 (dd, 1 H, *J*<sub>1b,2</sub> = 11.5 Hz, H-1b), 1.98 (s, 3 H, NHCOCH<sub>3</sub>). **<sup>13</sup>C NMR** (100.6 MHz, 8:1 CD<sub>3</sub>OD-CDCl<sub>3</sub>):  $\delta$  172.4 (CO), 165.5 (CN), 133.8-122.9 (C<sub>arom</sub>), 75.0 (C-3), 73.7 (C-4), 65.9 (C-5), 53.3 (CH<sub>2</sub>Naph), 49.8 (C-2), 45.8 (C-1), 30.4 (C-6), 21.5 (NHCOCH<sub>3</sub>). **ESIMS:** *m/z* 386.2 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>20</sub>H<sub>23</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C 56.93, H 5.73, N 9.96, S 7.60. Found: C, 56.65; H, 5.48; N, 9.68; S 7.29.

**(Z)-2-Acetamido-5-N,6-S-(N'-p-fluorobenzyliminomethylidene)-1,2-dideoxy-6-**

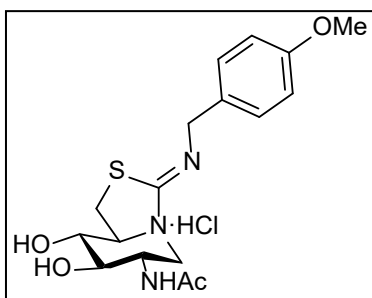
**thionojirimycin Hydrochloride (18).** Column chromatography, eluent: 60:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 95 mg (quantitative).  $[\alpha]_D^{25} +34.4$  (*c* 1.00, MeOH). *R<sub>f</sub>* 0.53 (40:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD):  $\delta$  7.42 (m, 2 H, CH<sub>arom</sub>), 7.12 (bt, 2 H, *J*<sub>H,H</sub> = 8.7 Hz, CH<sub>arom</sub>), 4.54 (bs, 2H, CH<sub>2</sub>Ph), 4.17 (dd, 1 H, *J*<sub>1a,1b</sub> = 13.1 Hz, *J*<sub>1a,2</sub> = 5.2 Hz, H-1a), 3.94 (m, 1 H, H-2), 3.86 (m, 1 H, H-5), 3.72 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.1 Hz, *J*<sub>5,6a</sub> = 7.5 Hz, H-6a), 3.53 (t,

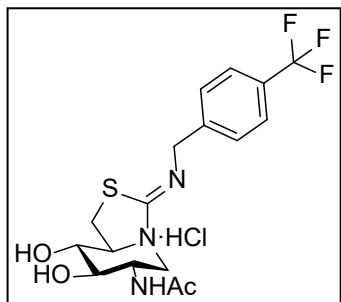
1 H,  $J_{2,3} = J_{3,4} = 9.4$  Hz, H-3), 3.40 (m, 2 H, H-4, H-6b), 3.00 (dd, 1 H,  $J_{1b,2} = 11.7$  Hz, H-1b), 2.02 (s, 1 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (125.7 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  172.3 (CO), 169.5 (CN), 162, 128.6, 113.5 ( $\text{C}_{\text{arom}}$ ), 74.2 (C-3), 73.4 (C-4), 67.2 (C-5), 52.1 ( $\text{CH}_2\text{Ph}$ ), 49.5 (C-2), 45.6 (C-1), 30.7 (C-6), 21.3 ( $\text{NHCOCH}_3$ ). **ESIMS**:  $m/z$  354 [ $\text{M} + \text{H}$ ] $^+$ , 376 [ $\text{M} + \text{Na}$ ] $^+$ . Anal. Calcd for  $\text{C}_{16}\text{H}_{20}\text{FN}_3\text{O}_3\text{S}\cdot\text{HCl}$ : C 49.29, H 5.43, N 10.78, S 8.22. Found: C, 49.14; H, 5.53; N, 10.62; S, 7.99.

**(Z)-2-Acetamido-5-N,6-S-(N'-p-methoxybenzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (19)**. Column chromatography, eluent: 60:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 98 mg (quantitative).  $[\alpha]_{\text{D}} +36.2$  ( $c$  1.00, MeOH).  $R_f$  0.32 (40:10:1 DCM-MeOH-H<sub>2</sub>O).



$^1\text{H NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.26 (d, 2 H,  $J_{\text{H,H}} = 8.5$  Hz,  $\text{CH}_{\text{arom}}$ ), 6.91 (d, 2 H,  $J_{\text{H,H}} = 8.5$  Hz, Ph), 4.41 (d, 2H,  $^3J_{\text{H,H}} = 15$  Hz,  $\text{CH}_2\text{Ph}$ ), 4.15 (dd, 1 H,  $J_{1a,1b} = 13.1$  Hz,  $J_{1a,2} = 5.4$  Hz, H-1a), 3.84 (m, 1 H, H-2), 3.79 (s, 3 H,  $\text{OCH}_3$ ), 3.69 (m, 1 H, H-5), 3.63 (dd, 1 H,  $J_{6a,6b} = 11.1$  Hz,  $J_{5,6a} = 7.2$  Hz, H-6a), 3.46 (t, 1 H,  $J_{2,3} = J_{3,4} = 9.7$ , H-3), 3.38 (m, 2 H, H-4, H-6b), 2.87 (dd, 1 H,  $J_{1b,2} = 11.7$  Hz, H-1b), 2.00 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (125.7 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  172.3 (CO), 168.1 (CN), 159.2, 128.6, 113.5 ( $\text{C}_{\text{arom}}$ ), 74.8 (C-3), 73.6 (C-4), 66.2 (C-5), 54.3 ( $\text{CH}_2\text{Ph}$ ,  $\text{CH}_3\text{O}$ ), 49.7 (C-2), 45.7 (C-1), 30.3 (C-6), 21.3 ( $\text{NHCOCH}_3$ ). **ESIMS**:  $m/z$  366.2 [ $\text{M} + \text{H}$ ] $^+$ , 388.1 [ $\text{M} + \text{Na}$ ] $^+$ . Anal. Calcd for  $\text{C}_{17}\text{H}_{23}\text{N}_3\text{O}_4\text{S}\cdot\text{HCl}$ : C 50.80, H 6.02, N 10.46, S 7.98. Found: C, 50.53; H, 6.14; N, 10.33; S, 7.75.

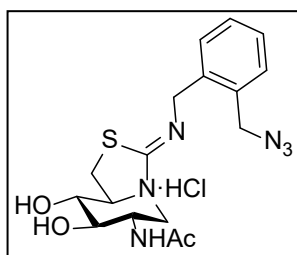
**(Z)-2-Acetamido-5-N,6-S-(N'-p-trifluoromethylbenzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (20)**. Column chromatography, eluent: 60:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 95 mg (quantitative).  $[\alpha]_{\text{D}} +34.4$  ( $c$  1.00, MeOH).  $R_f$  0.50 (40:10:1 DCM-MeOH-H<sub>2</sub>O).



$^1\text{H NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  7.42 (m, 2 H,  $\text{CH}_{\text{arom}}$ ), 7.12 (bt, 2 H,  $J_{\text{H,H}} = 8.7$  Hz,  $\text{CH}_{\text{arom}}$ ), 4.54 (bs, 2H,  $\text{CH}_2\text{Ph}$ ), 4.17 (dd, 1 H,  $J_{1a,1b} = 13.1$  Hz,  $J_{1a,2} = 5.2$  Hz, H-1a), 3.94 (m, 1 H, H-2), 3.86 (m, 1 H, H-5), 3.72 (dd, 1 H,  $J_{6a,6b} = 11.1$  Hz,  $J_{5,6a} = 7.5$  Hz, H-6a), 3.53 (t, 1 H,  $J_{2,3} = J_{3,4} = 9.4$ , H-3), 3.40 (m, 2 H, H-4, H-6b), 3.00 (dd, 1 H,  $J_{1b,2} = 11.7$  Hz, H-1b), 2.02 (s, 3 H,  $\text{NHCOCH}_3$ ).  $^{13}\text{C NMR}$  (125.7 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  172.34 (CO), 163.7 (CN), 128.6, 113.5 ( $\text{C}_{\text{arom}}$ ), 74.2 (C-3), 73.4 (C-4), 67.2 (C-5), 52.1 ( $\text{CH}_2\text{Ph}$ ), 49.5 (C-2), 45.6 (C-1), 30.7 (C-6), 21.3 ( $\text{NHCOCH}_3$ ). **ESIMS**:  $m/z$

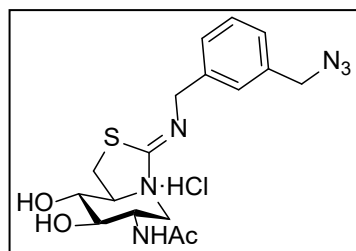
354.1 [M + H]<sup>+</sup>, 376.1 [M + Na]<sup>+</sup>. Anal. Calcd for C<sub>16</sub>H<sub>20</sub>FN<sub>3</sub>O<sub>3</sub>S.HCl: C 49.29, H 5.43, N 10.78, S 8.22. Found: C, 49.14; H, 5.53; N, 10.62; S, 7.99.

**(Z)-2-Acetamido-5-N,6-S-(N'-o-azidomethylbenzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (21).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 105 mg (quantitative). [α]<sub>D</sub> +26.4 (c 1.00, MeOH). R<sub>f</sub> 0.75 (50:10:1 DCM-MeOH-H<sub>2</sub>O).



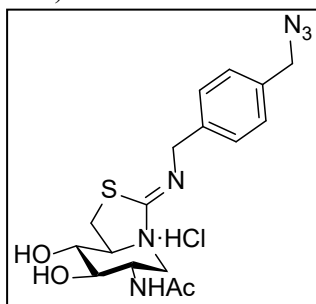
**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD): δ 7.31-7.13 (m, 4 H, CH<sub>arom</sub>), 4.36 (s, 2H, NCH<sub>2</sub>Ph), 4.33 (bd, 2H, CH<sub>2</sub>N<sub>3</sub>), 4.07 (dd, 1 H, J<sub>1a,1b</sub> = 12.9 Hz, J<sub>1a,2</sub> = 5.2 Hz, H-1a), 3.73 (m, 1 H, H-2), 3.37 (bdd, 1 H, J<sub>6a,6b</sub> = 11.1 Hz, J<sub>5,6a</sub> = 6.2 Hz, H-6a), 3.33 (m, 1 H, H-5), 3.26 (m, 2 H, H-3, H-4), 3.10 (dd, 1 H, H-6b), 2.50 (dd, 1 H, J<sub>1b,2</sub> = 11.3 Hz, H-1b), 1.86 (s, 3 H, NHCOCH<sub>3</sub>). **<sup>13</sup>C NMR** (75.5 MHz, CD<sub>3</sub>OD): δ 172.2 (CO), 162.9 (CN), 138.5, 133.4, 129.3, 128.5, 128.2, 126.8 (C<sub>arom</sub>), 75.5 (C-3), 73.8 (C-4), 64.7 (C-5), 54.8 (CH<sub>2</sub>Ph), 51.9 (CH<sub>2</sub>N<sub>3</sub>), 49.9 (C-2), 45.7 (C-1), 29.9 (C-6), 21.3 (NHCOCH<sub>3</sub>). **ESIMS:** m/z 391.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>22</sub>N<sub>6</sub>O<sub>3</sub>S.HCl: C 47.83, H 5.43, N 19.69, S 7.51. Found: C, 47.51; H, 5.19; N, 19.33; S, 7.14.

**(Z)-2-Acetamido-5-N,6-S-(N'-m-azidomethylbenzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (22).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 105 mg (quantitative). [α]<sub>D</sub> +37.4 (c 1.00, MeOH). R<sub>f</sub> 0.75 (50:10:1 DCM-MeOH-H<sub>2</sub>O).



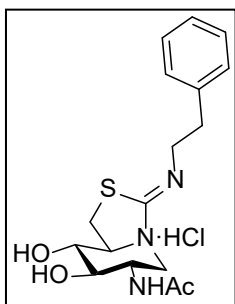
**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD): δ 7.38-7.22 (m, 4 H, CH<sub>arom</sub>), 4.42 (bd, 2H, NCH<sub>2</sub>Ph), 4.36 (s, 2H, CH<sub>2</sub>N<sub>3</sub>), 4.18 (dd, 1 H, J<sub>1a,1b</sub> = 12.9 Hz, J<sub>1a,2</sub> = 5.2 Hz, H-1a), 3.87 (m, 1 H, H-2), 3.50 (bdd, 1 H, J<sub>6a,6b</sub> = 11.1 Hz, J<sub>5,6a</sub> = J<sub>5,6b</sub> = 6.2 Hz, H-6a), 3.45 (m, 1 H, H-5), 3.39 (m, 2 H, H-3, H-4), 3.22 (dd, 1 H, H-6b), 2.65 (dd, 1 H, J<sub>1b,2</sub> = 11.3 Hz, H-1b), 1.99 (s, 3 H, NHCOCH<sub>3</sub>). **<sup>13</sup>C NMR** (75.5 MHz, CD<sub>3</sub>OD): δ 172.2 (CO), 162.9 (CN), 140.4, 135.8, 128.4, 127.0, 126.4 (C<sub>arom</sub>), 75.5 (C-3), 73.9 (C-4), 64.9 (C-5), 56.8 (CH<sub>2</sub>Ph), 54.1 (CH<sub>2</sub>N<sub>3</sub>), 49.9 (C-2), 45.8 (C-1), 30.0 (C-6), 21.3 (NHCOCH<sub>3</sub>). **ESIMS:** m/z 391.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>22</sub>N<sub>6</sub>O<sub>3</sub>S.HCl: C 47.83, H 5.43, N 19.69, S 7.51. Found: C, 47.69; H, 5.24; N, 19.39; S, 7.20.

**(Z)-2-Acetamido-5-N,6-S-(N'-p-azidomethylbenzylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (23).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 105 mg (quantitative).  $[\alpha]_D +1.07$  (c 1.00, MeOH).  $R_f$  0.75 (50:10:1 DCM-MeOH-H<sub>2</sub>O).



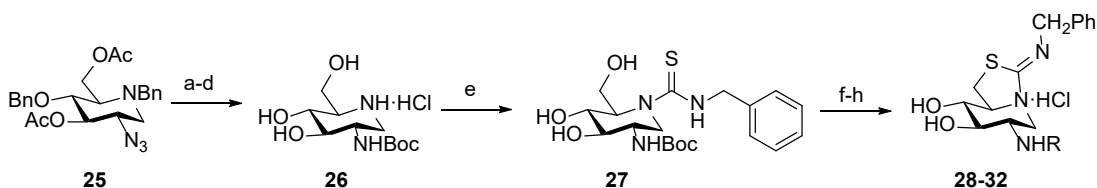
**<sup>1</sup>H NMR** (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.35 (d, 2 H,  $J_{H,H} = 8.3$  Hz, CH<sub>arom</sub>), 7.30 (d, 2 H, CH<sub>arom</sub>), 4.40 (bt, 2 H,  ${}^2J_{H,H} = 15.0$  Hz, NCH<sub>2</sub>Ph), 4.34 (s, 2 H, CH<sub>2</sub>N<sub>3</sub>), 4.18 (dd, 1 H,  $J_{1a,1b} = 12.9$  Hz,  $J_{1a,2} = 5.2$  Hz, H-1a), 3.88 (m, 1 H, H-2), 3.47 (m, 1 H, H-5), 3.39 (m, 2 H, H-3, H-4), 3.34 (m, 1 H, H-6a), 3.19 (dd, 1 H,  $J_{6a,6b} = 11.1$  Hz,  $J_{5,6b} = 6.8$  Hz, H-6b), 2.61 (dd, 1 H,  $J_{1b,2} = 11.3$  Hz, H-1b), 2.00 (s, 3 H, NHCOCH<sub>3</sub>). **<sup>13</sup>C NMR** (100.6 MHz, CD<sub>3</sub>OD):  $\delta$  172.2 (CO), 161.9 (CN), 140.4, 134.0, 128.0, 127.5 (C<sub>arom</sub>), 75.6 (C-3), 74.0 (C-4), 64.6 (C-5), 54.2 (CH<sub>2</sub>Ph), 53.9 (CH<sub>2</sub>N<sub>3</sub>), 49.9 (C-2), 45.8 (C-1), 29.8 (C-6), 21.3 (NHCOCH<sub>3</sub>). **ESIMS:**  $m/z$  391.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>22</sub>N<sub>6</sub>O<sub>3</sub>S.HCl: C 47.83, H 5.43, N 19.69, S 7.51. Found: C, 47.49; H, 5.19; N, 19.43; S, 7.19.

**(Z)-2-Acetamido-5-N,6-S-(N'-2-phenylethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (24).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 94 mg (quantitative).  $[\alpha]_D +2.84$  (c 1.00, MeOH).  $R_f$  0.30 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



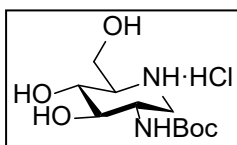
**<sup>1</sup>H NMR** (300 MHz, CD<sub>3</sub>OD):  $\delta$  7.24-7.09 (m, 4 H, CH<sub>arom</sub>), 3.97 (dd, 1 H,  $J_{1a,1b} = 12.9$  Hz,  $J_{1a,2} = 5.2$  Hz, H-1a), 3.73 (m, 1 H, H-2), 3.67 (m, 1 H, H-5), 3.48 (dd, 1 H,  $J_{6a,6b} = 11.4$  Hz,  $J_{5,6a} = 7.3$  Hz, H-6a), 3.39 (m, 3 H, H-3, CH<sub>2</sub>CH<sub>2</sub>Ph), 3.20 (m, 1 H, H-4, H-6b), 2.80 (m, 3 H, CH<sub>2</sub>CH<sub>2</sub>Ph, H-1b), 1.90 (s, 3 H, NHCOCH<sub>3</sub>). **<sup>13</sup>C NMR** (75.5 MHz, CD<sub>3</sub>OD):  $\delta$  172.4 (CO), 168.1 (CN), 138.0, 128.7, 128.3, 126.3 (C<sub>arom</sub>), 74.4 (C-3), 73.1 (C-4), 66.8 (C-5), 52.2 (CH<sub>2</sub>CH<sub>2</sub>Ph), 49.6 (C-2), 45.5 (C-1), 35.5 (C-6), 30.1 (CH<sub>2</sub>CH<sub>2</sub>Ph), 21.3 (NHCOCH<sub>3</sub>). **ESIMS:**  $m/z$  372.1 [M + Na]<sup>+</sup>, 350.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>23</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C 52.91, H 6.27, N 10.89, S 8.31. Found: C, 52.56; H, 6.08; N, 10.55; S, 7.99.

### Synthesis of DNJC-thiazolidines 28-32



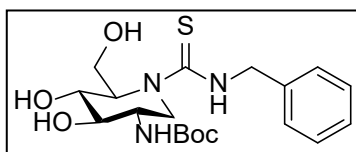
**Supplemental scheme 6.** Reagents and conditions: a. NaOMe, MeOH, RT, 18 h; b. PPh<sub>3</sub>, THF-NH<sub>4</sub>OH, 60 °C, 18 h; c. Boc<sub>2</sub>O, Et<sub>3</sub>N, dioxane, RT, 18 h; d. H<sub>2</sub>, Pd/C, MeOH, RT, 18 h; e. BnNCS, Et<sub>3</sub>N, MeCN, RT, 18 h; f. HCl, MeOH, RT, 18 h; g. TFA, 1:1 DCM-H<sub>2</sub>O, RT, 18 h. RCOCl or RNCS, Et<sub>3</sub>N, MeOH, RT, 18 h, quant.

**2-tert-Butoxycarbonylamino-1,2-dideoxynojirimycin (26).** 2-Azido-3,6-di-*O*-acetyl-4-*O*-benzyl-1,2,5-trideoxy-1,5-imino-D-glucitol<sup>1</sup> **25** (207 mg, 0.45 mM) was dissolved in MeOH (4.5 mL) and MeONa (5 mg, 0.09 mM) was added. The mixture was stirred overnight at RT. Then, dry ice was added until neutral pH and the solvent was evaporated. The deacetylated product was dissolved in THF-NH<sub>4</sub>OH (3:1, 4 mL) and reduced with PPh<sub>3</sub> (192 mg, 0.73 mM) overnight at 60 °C. The solvent was evaporated and the crude amine was dissolved in dioxane (4.8 mL). Boc<sub>2</sub>O (213 mg, 0.978 mM) and Et<sub>3</sub>N (131 μL, 0.978 mM) were added and the mixture stirred overnight. The reaction mixture was concentrated to dryness, the crude Boc-protected derivative was dissolved in MeOH (4.5 mL) and debenzylated by treatment with Pd/C 10% (50 mg) overnight under H<sub>2</sub> atmosphere. The mixture was filtered over celite, evaporated and suspended in H<sub>2</sub>O (10 mL). The insoluble PPh<sub>3</sub>O precipitate was filtered and the aqueous solution was lyophilized to give **26**. Yield: 58 mg (49%, 4 steps). R<sub>f</sub> 0.30 (6:3:1 MeCN-H<sub>2</sub>O-NH<sub>4</sub>OH).



<sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD): δ 3.93 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.8 Hz, *J*<sub>5,6a</sub> = 2.8 Hz, H-6a), 3.86 (dd, 1 H, *J*<sub>5,6b</sub> = 5.3 Hz, H-6b), 3.69 (m, 1 H, H-2), 3.55 (bt, 1 H, *J*<sub>3,4</sub> = 9.6 Hz, H-4), 3.45 (bt, 1 H, H-3), 3.39 (dd, 1 H, *J*<sub>1a,1b</sub> = 12.2, *J*<sub>1a,2</sub> = 4.7 Hz, H-1a), 3.07 (m, 1 H, H-5), 2.89 (bt, 1 H, *J*<sub>1b,2</sub> = 12 Hz, H-1b), 1.46 (s, 9 H, COCMe<sub>3</sub>). <sup>13</sup>C NMR (125.7 MHz, CD<sub>3</sub>OD): δ 156.5 (CO), 79.3 (CMe<sub>3</sub>), 73.8 (C-3), 68.9 (C-4), 60.5 (C-5), 57.7 (C-6), 49.7 (C-2), 44.8 (C-1), 27.2 (CMe<sub>3</sub>). HRMS (ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>11</sub>H<sub>22</sub>N<sub>2</sub>O<sub>5</sub>]<sup>+</sup> 263.1601; found 263.1605.

**2-tert-Butoxycarbonylamino-5-*N*-(*N*'-methylbenzylthiocarbamoyl)-1,2-dideoxynojirimycin (27).** Compound **26** (58 mg, 0.22 mM) in MeCN (2.2 mL) was treated with Et<sub>3</sub>N (62 μL, 0.44 mmol) and benzyl isothiocyanate (44 μL, 0.331 mM). The mixture was stirred at RT overnight and concentrated to dryness. The crude product was purified by column chromatography using 100:10:1 DCM-MeOH-H<sub>2</sub>O as eluent. Yield: 47 mg (52%). R<sub>f</sub> 0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



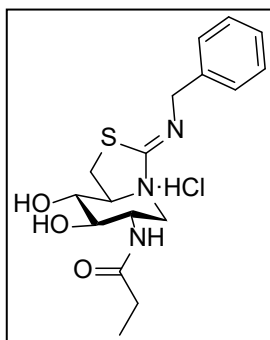
<sup>1</sup>H NMR (600 MHz, CD<sub>3</sub>OD): δ 7.36-7.22 (CH<sub>arom</sub>), 4.90 (m, 2 H, H-1a, CH<sub>2</sub>Ph), 4.80 (d, 1 H, <sup>2</sup>*J*<sub>H,H</sub> = 15 Hz, CHPh), 4.49 (m, 1 H, H-5), 3.97 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.4 Hz, *J*<sub>5,6a</sub> = 8.7 Hz, H-6a), 3.82 (dd, 1 H, *J*<sub>5,6b</sub> = 3.6 Hz, H-6b), 3.73 (t, 1 H, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 5.1 Hz, H-4), 3.66 (m, 1 H, H-2), 3.61 (t, 1 H, *J*<sub>2,3</sub> = 5.1 Hz, H-3), 3.52 (dd, 1 H, *J*<sub>1a,1b</sub> = 14.3 Hz, *J*<sub>1b,2</sub> = 3.8 Hz, H-1b), 1.45 (s, 9 H, CO<sub>2</sub>Me<sub>3</sub>). <sup>13</sup>C NMR (150 MHz, CD<sub>3</sub>OD): δ 185.5 (CS), 156.3 (CO), 138.7, 128.0, 127.2, 126.6 (C<sub>arom</sub>), 79.0



(CMe<sub>3</sub>), 70.8 (C-3), 69.2 (C-4), 64.3 (C-5), 60.4 (C-6), 52.0 (C-2), 49.2 (CH<sub>2</sub>Ph), 43.9 (C-1), 27.3 (CMe<sub>3</sub>). **HRMS** (ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>19</sub>H<sub>29</sub>N<sub>3</sub>O<sub>5</sub>S]<sup>+</sup> 412.1901; found 412.1904.

**General procedure for the preparation of the DNJC-thiazolidines 28-32.** Compound **27** (47 mg, 0.11 mM) was dissolved in MeOH (2 mL) and concentrated HCl was dropwise added until pH 1. The mixture was stirred at rt overnight and monitored by ESIMS to confirm total conversion into the corresponding cyclic isourea. The reaction mixture was concentrated and the crude product was dissolved in 1:1 DCM-H<sub>2</sub>O (2 mL). TFA (30 μL, 0.38 mM, 3 eq) was then added and the solution was stirred overnight. The mixture was evaporated, co-evaporated with toluene (3 x 5 mL) and lyophilized. The free amine derivative thus obtained was engaged in the next reactions without further purification. Thus, 10 mg (0.034 mM) of this material was dissolved in MeOH (3 mL) and reacted with either propionyl, butyryl or isobutyryl chloride or ethyl or propyl isothiocyanate (1.2 eq) in the presence of Et<sub>3</sub>N (3 eq). The reaction mixtures were stirred at RT overnight, the solvent was evaporated and the crude product was purified as indicated for each compound.

**(Z)-2-Propionamido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin Hydrochloride (28).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 94 mg (quantitative). [α]<sub>D</sub> +31.9 (c 1.00, MeOH). R<sub>f</sub> 0.30 (70:10:1 DCM-MeOH-H<sub>2</sub>O).

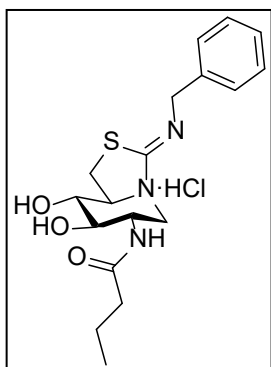


**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 7.18-7.08 (m, 5 H, CH<sub>arom</sub>), 4.26 (d, 2 H, <sup>3</sup>J<sub>H,H</sub> = 15 Hz, CH<sub>2</sub>Ph), 4.05 (dd, 1 H, J<sub>1a,1b</sub> = 12.6 Hz, J<sub>1a,2</sub> = 5.3 Hz, H-1a), 3.75 (m, 1 H, H-2), 3.36 (dd, 1 H, J<sub>6a,6b</sub> = 11.2 Hz, J<sub>5,6a</sub> = 7.3 Hz, H-6a), 3.29 (m, 1 H, H-5), 3.23 (m, 2 H, H-3, H-4), 3.08 (dd, 1 H, H-6b), 2.51 (dd, 1 H, J<sub>1b,2</sub> = 11.7 Hz, H-1b), 2.13 (q, 2 H, COCH<sub>2</sub>CH<sub>3</sub>), 1.03 (t, 3 H, COCH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 177.4 (CO), 163.7 (CN), 141.3, 129.3, 128.5, 127.8 (C<sub>arom</sub>), 77.0 (C-3), 75.5 (C-4), 66.2 (C-

5), 58.8 (CH<sub>2</sub>Ph), 51.2 (C-2), 47.3 (C-1), 31.4 (C-6) 30.3 (COCH<sub>2</sub>CH<sub>3</sub>), 10.3 (COCH<sub>2</sub>CH<sub>3</sub>). **ESIMS:** *m/z* 350.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>17</sub>H<sub>23</sub>N<sub>3</sub>O<sub>3</sub>S.HCl: C 52.91, H 6.27, N 10.89, S 8.31. Found: C, 52.98; H, 6.40; N, 10.71; S, 8.52.

**(Z)-2-Butyramido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin**

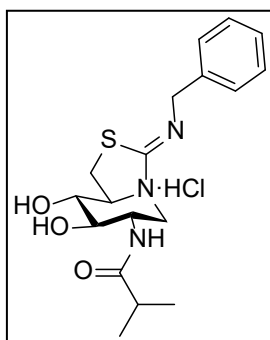
**(29).** Column chromatography, eluent: 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 98 mg (quantitative).  
R<sub>f</sub> 0.50 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 7.31-7.20 (m, 5 H, CH<sub>arom</sub>), 4.37 (d, 2 H, <sup>3</sup>J<sub>H,H</sub> = 15.0 Hz, CH<sub>2</sub>Ph), 4.16 (dd, 1 H, J<sub>1a,1b</sub> = 12.6 Hz, J<sub>1a,2</sub> = 5.3 Hz, H-1a), 3.88 (m, 1 H, H-2), 3.48 (dd, 1 H, J<sub>6a,6b</sub> = 11.2 Hz, J<sub>5,6a</sub> = 7.3 Hz, H-6a), 3.40 (m, 1 H, H-5), 3.37 (m, 2 H, H-3, H-4), 3.19 (dd, 1 H, H-6b), 2.61 (dd, 1 H, J<sub>1b,2</sub> = 11.2 Hz, H-1b), 2.49, 1.66 (m, 4 H, CH<sub>2</sub>CH<sub>3</sub>), 0.97 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (150 MHz, CD<sub>3</sub>OD): δ 175.0 (CO), 162.1 (CN), 140.0, 127.8, 127.1, 126.3 (C<sub>arom</sub>), 75.5 (C-3), 74.2 (C-4), 64.7 (C-5), 57.6 (CH<sub>2</sub>Ph), 49.7 (C-2), 45.8 (C-1), 37.7 (C-6), 29.9, 18.9 (CH<sub>2</sub>CH<sub>3</sub>), 12.5 (CH<sub>2</sub>CH<sub>3</sub>). **HRMS** (ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>18</sub>H<sub>25</sub>N<sub>3</sub>O<sub>3</sub>S]<sup>+</sup> 364.1689; found 364.1696.

**2-Isobutyramido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin**

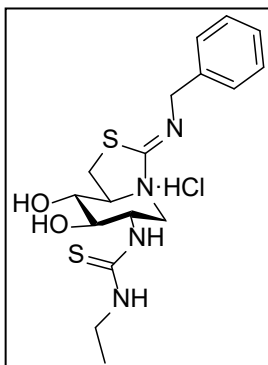
**Hydrochloride (30).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 98 mg (quantitative). [α]<sub>D</sub> +36.2 (c 1.04, MeOH). R<sub>f</sub> 0.50 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 7.31-7.20 (m, 5 H, CH<sub>arom</sub>), 4.39 (d, 2 H, <sup>3</sup>J<sub>H,H</sub> = 15.0 Hz, CH<sub>2</sub>Ph), 4.17 (dd, 1 H, J<sub>1a,1b</sub> = 12.6 Hz, J<sub>1a,2</sub> = 5.3 Hz, H-1a), 3.87 (m, 1 H, H-2), 3.49 (dd, 1 H, J<sub>6a,6b</sub> = 11.2 Hz, J<sub>5,6a</sub> = 7.3 Hz, H-6a), 3.47 (m, 1 H, H-5), 3.39 (m, 2 H, H-3, H-4), 3.20 (dd, 1 H, H-6b), 2.64 (dd, 1 H, J<sub>1b,2</sub> = 11.7 Hz, H-1b), 2.49 (q, 2 H, COCH<sub>2</sub>Me<sub>2</sub>), 1.15 (t, 6 H, COCH<sub>2</sub>Me<sub>2</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 180.6 (CO), 163.8 (CN), 141.2, 129.3, 128.6, 127.8 (C<sub>arom</sub>), 76.9 (C-3), 75.7 (C-4), 66.2 (C-5), 58.8 (CH<sub>2</sub>Ph), 51.1 (C-2), 47.1 (C-1), 36.4 (C-6) 31.4 (COCHMe<sub>2</sub>), 20.0, 19.7 (COCHMe<sub>2</sub>). **ESIMS**: *m/z* 364.1 [M + H]<sup>+</sup>. Anal. Calcd for C<sub>18</sub>H<sub>25</sub>N<sub>3</sub>O<sub>3</sub>S·HCl: C, 54.06; H, 6.55; N, 10.51; S, 8.02. Found: C, 53.79; H, 6.47; N, 10.32; S, 7.72.

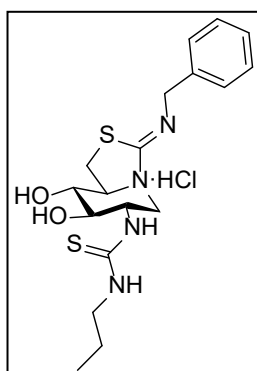
**(Z)-2-N'-Ethylthioureido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-**

**thionojirimycin (31).** Column chromatography, eluent: 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 102 mg (quantitative). R<sub>f</sub> 0.50 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 7.33-7.21 (m, 5 H, CH<sub>arom</sub>), 4.37 (d, 3 H, <sup>3</sup>J<sub>H,H</sub> = 15.0 Hz, CH<sub>2</sub>Ph, H-1a), 3.47 (m, 3 H, H-6a, CH<sub>2</sub>N), 3.41 (m, 1 H, H-2), 3.37 (m, 3 H, H-3, H-4, H-5), 3.20 (dd, 1 H, H-6b), 2.60 (dd, 1 H, J<sub>1b,2</sub> = 11.2 Hz, H-1b), 1.16 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (150 MHz, CD<sub>3</sub>OD): δ 161.7 (CN), 140.2, 127.8, 127.1, 126.2 (C<sub>arom</sub>), 76.3 (C-3), 73.8 (C-4), 64.5 (C-5), 57.7 (CH<sub>2</sub>Ph), 48.1 (CH<sub>2</sub>N, C-2), 45.7 (C-1), 29.7 (C-6), 13.2 (CH<sub>2</sub>CH<sub>3</sub>). **HRMS** (ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>17</sub>H<sub>24</sub>N<sub>4</sub>O<sub>2</sub>S<sub>2</sub>]<sup>+</sup> 381.1413; found 381.1424.

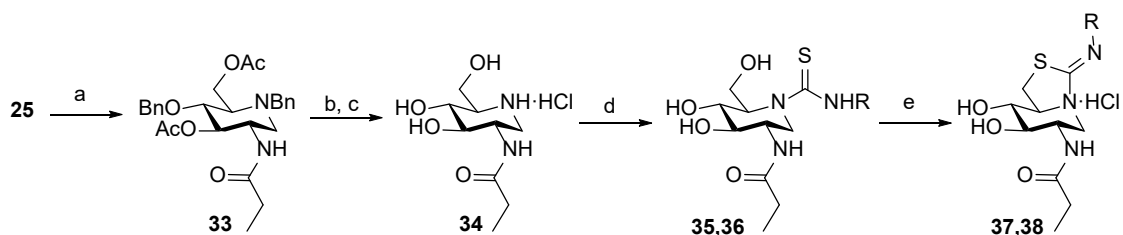
**(Z)-2-Propylthioureido-5-N,6-S-(N'-benzyliminomethylidene)-1,2-dideoxy-6-thionojirimycin (32)**. Column chromatography, eluent: 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 106 mg (quantitative). R<sub>f</sub> 0.51 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 7.33-7.20 (m, 5 H, CH<sub>arom</sub>), 4.38 (d, 3 H, <sup>3</sup>J<sub>H,H</sub> = 15.0 Hz, CH<sub>2</sub>Ph, H-1a), 3.48 (m, 3 H, H-6a, CH<sub>2</sub>N), 3.42 (m, 1 H, H-2), 3.37 (m, 3 H, H-3, H-4, H-5), 3.21 (dd, 1 H, H-6b), 2.62 (dd, 1 H, J<sub>1b,2</sub> = 11.2 Hz, H-1b), 1.58 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>), 0.94 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (150 MHz, CD<sub>3</sub>OD): δ 162.0 (CN), 140.0, 127.8, 127.1, 126.3 (C<sub>arom</sub>), 76.3 (C-3), 73.8 (C-4), 64.5 (C-5), 57.8 (CH<sub>2</sub>Ph), 48.1 (CH<sub>2</sub>N, C-2), 45.7 (C-1), 29.7 (C-6), 21.9 (CH<sub>2</sub>CH<sub>3</sub>), 10.2 (CH<sub>2</sub>CH<sub>3</sub>). **HRMS**

(ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>18</sub>H<sub>26</sub>N<sub>4</sub>O<sub>2</sub>S<sub>2</sub>]<sup>+</sup> 395.1570; found 395.1567.

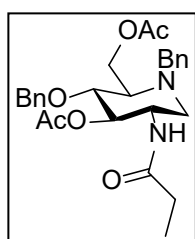
## Synthesis of DNJC-thiazolidines **37** and **38**.



**Supplemental scheme 7.** Reagents and conditions: a. Zn, CuSO<sub>4(aq)</sub>, THF-PrOH-Pr<sub>2</sub>O, RT, 20 min; b. NaOMe, MeOH, RT, 18 h; c. H<sub>2</sub>, Pd/C, MeOH, RT, 18 h; d. R-NCS, Et<sub>3</sub>N, MeCN, RT, 18 h; e. HCl, MeOH, RT, 18 h.

### 2-Propionamide-3,6-di-O-acetyl-1-N,4-O-dibenzyl-1,2-dideoxynojirimycin (**33**).

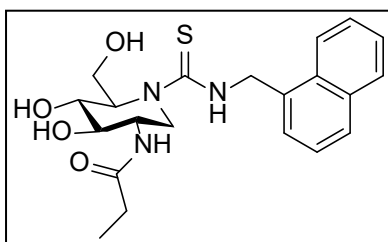
Powdered zinc (373 mg, 5.74 mM) was added to a solution of the azide **25**<sup>1</sup> (130 mg, 0.28 mM) in 3:2:1 THF-propionic acid-propionic anhydride (3.5 mL) and stirred vigorously before the dropwise addition of a saturated copper (II) sulfate solution (0.9 mL) to initiate the reaction. After 20 min, t.l.c. analysis (1:2 EtOAc-CyHex) indicated the complete consumption of the starting material (*R<sub>f</sub>* 0.50) and the formation of a major product (*R<sub>f</sub>* 0.10). The reaction mixture was filtered through Celite<sup>®</sup>, concentrated *in vacuo* and the crude residue purified by flash chromatography (50:1, DCM-MeOH) to afford the propanamide **33** (109 mg, 79%). [ $\alpha$ ]<sub>D</sub> +8.80 (*c* 1, MeOH). *R<sub>f</sub>* 0.50 (20:1 DCM-MeOH).



<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>):  $\delta$  7.36-7.23 (m, 10 H, CH<sub>arom</sub>), 6.07 (s, 1 H,  $J_{\text{NH},2} = 8.7$  Hz, NH), 4.90 (dd, 1 H,  $J_{2,3} = J_{3,4} = 8.6$  Hz, H-3), 4.65 (dd, 2 H,  $^2J_{\text{H,H}} = 11.1$  Hz, OCH<sub>2</sub>Ph), 4.51 (dd, 1 H,  $J_{6a,6b} = 12.1$  Hz,  $J_{5,6a} = J_{5,6b} = 4.1$  Hz, H-6a), 4.37 (dd, 1 H, H-6b), 4.15 (m, 1 H, H-2), 3.92 (dd, 2 H,  $J_{\text{gem}} = 13.6$  Hz, NCH<sub>2</sub>Ph), 3.73 (t, 1 H,  $J_{4,3} = J_{4,5} = 6.7$  Hz, H-4), 3.05 (dd, 1 H,  $J_{1a,1b} = 12.0$  Hz,  $J_{1a,2} = 3.7$  Hz, H-1a), 2.86 (m, 1 H, H-5), 2.18 (dd, 1 H,  $J_{1b,2} = 7.8$  Hz, H-1b), 2.05 (m, 2 H, COCH<sub>2</sub>CH<sub>3</sub>), 2.04 (s, 6 H, COCH<sub>3</sub>), 1.02 (t, 3 H, COCH<sub>2</sub>CH<sub>3</sub>). <sup>13</sup>C NMR (75.5 MHz, CDCl<sub>3</sub>):  $\delta$  173.1, 170.9, 170.7 (COCH<sub>3</sub>, NHCOCH<sub>2</sub>), 138.1-127.3 (C<sub>arom</sub>), 75.5 (C-4), 74.4 (C-3), 73.9 (OCH<sub>2</sub>Ph), 61.9 (C-5), 59.7 (C-6), 57.3 (NCH<sub>2</sub>Ph), 51.6 (C-1), 47.9 (C-2), 29.7 (COCH<sub>2</sub>CH<sub>3</sub>), 20.9, 20.8 (COCH<sub>3</sub>), 9.5 (COCH<sub>2</sub>CH<sub>3</sub>). **ESIMS**: 483.3 [M + H]<sup>+</sup>, 505.3 [M+Na]<sup>+</sup>. Anal. Calcd for C<sub>27</sub>H<sub>34</sub>N<sub>2</sub>O<sub>6</sub>: C 67.20, H 7.10, N 5.81. Found: C, 67.33; H, 7.26; N, 5.74.

**2-Propionamido-5-*N*-(*N'*-1-naphthylmethylthiocarbamoyl)-1,2-dideoxynojirimycin (35).**

Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 90 mg (60%). *R<sub>f</sub>* 0.40 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 8.10-7.42 (m, 8 H, CH<sub>arom</sub>), 5.27 (bs, 2 H, CH<sub>2</sub>Naph), 4.98 (m, 1 H, H-1a), 4.31 (m, 1 H, H-5), 3.95 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.2 Hz, *J*<sub>5,6a</sub> = 8.8 Hz, H-6a), 3.95 (m, 1 H, H-2), 3.75 (dd, 1 H, *J*<sub>5,6b</sub> = 3.5 Hz, H-6b), 3.67 (t, 1 H, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 5.7 Hz, H-4), 3.62 (t, 1 H, *J*<sub>2,3</sub> = 5.4 Hz, H-3), 3.54

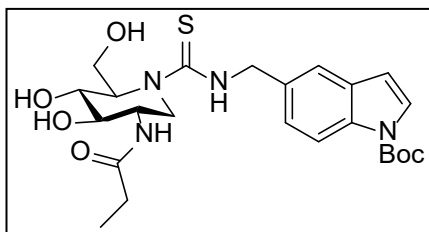
(dd, 1 H, *J*<sub>1a,1b</sub> = 14.1 Hz, *J*<sub>1b,2</sub> = 3.8 Hz, H-1b), 2.22 (q, 2 H, CH<sub>2</sub>CH<sub>3</sub>), 1.14 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>).

**<sup>13</sup>C NMR** (150 MHz, CD<sub>3</sub>OD): δ 185.4 (CS), 175.1 (CO), 133.9-123.2 (C<sub>arom</sub>), 70.7 (C-3), 69.1 (C-4), 64.8 (C-5), 60.3 (C-6), 51.3 (C-2, CH<sub>2</sub>Naph), 43.6 (C-1), 29.1 (CH<sub>2</sub>CH<sub>3</sub>), 8.6 (CH<sub>2</sub>CH<sub>3</sub>).

**HRMS** (ESI) *m/z* [M + Na]<sup>+</sup> calcd for [C<sub>21</sub>H<sub>27</sub>N<sub>3</sub>NaO<sub>4</sub>S]<sup>+</sup> 440.1614; found 440.1607

**2-Propionamido-5-*N*-(*N'*-(1-Boc-indol-5-ylmethyl)thiocarbamoyl)-1,2-dideoxynojirimycin (36).**

Column chromatography, eluent 100:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 90 mg (65%). [α]<sub>D</sub> -54.6 (*c* 1.0, MeOH). *R<sub>f</sub>* 0.60 (70:10:1 DCM-MeOH-H<sub>2</sub>O).



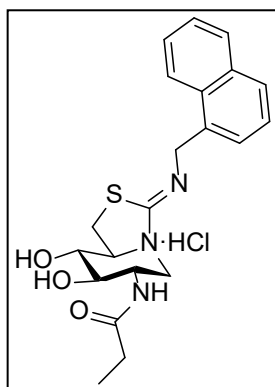
**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 8.04-6.55 (m, 5 H, CH<sub>arom</sub>), 4.88 (d, 3 H, <sup>3</sup>*J*<sub>H,H</sub> = 15.0 Hz, CH<sub>2</sub>Ph), 4.84 (m, 1 H, H-1a), 4.40 (m, 1 H, H-5), 3.95 (dd, 1 H, *J*<sub>6a,6b</sub> = 11.2 Hz, *J*<sub>5,6a</sub> = 8.8 Hz, H-6a), 3.92 (m, 1 H, H-2), 3.80 (dd, 1 H, *J*<sub>5,6b</sub> = 3.6 Hz, H-6b), 3.69 (t, 1 H, *J*<sub>3,4</sub> = *J*<sub>4,5</sub> = 5.2 Hz, H-4),

3.60 (t, 1 H, *J*<sub>2,3</sub> = 5.4 Hz, H-3), 3.51 (dd, 1 H, *J*<sub>1a,1b</sub> = 14.3 Hz, *J*<sub>1b,2</sub> = 3.6 Hz, H-1b), 2.17 (q, 2 H, CH<sub>2</sub>CH<sub>3</sub>), 1.68 (s, 9 H, CO<sub>2</sub>Me<sub>3</sub>), 1.07 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 185.4 (CS), 175.1 (CO<sub>amide</sub>), 149.6 (CO<sub>carbamate</sub>), 149.6-106.9 (C<sub>arom</sub>), 83.5 (CCH<sub>3</sub>), 70.7 (C-3), 69.2 (C-4), 64.6 (C-5), 60.2 (C-6), 51.3 (C-2), 49.2 (CH<sub>2</sub>Ph), 43.6 (C-1), 29.1 (CH<sub>2</sub>CH<sub>3</sub>), 26.9 (CO<sub>2</sub>Me<sub>3</sub>), 8.6 (CH<sub>2</sub>CH<sub>3</sub>). **HRMS** (ESI) *m/z* [M + Na]<sup>+</sup> calcd for [C<sub>24</sub>H<sub>34</sub>N<sub>4</sub>NaO<sub>6</sub>S]<sup>+</sup> 529.2091; found 529.2082

**(*Z*)-2-Propionamido-5-*N*,6-*S*-(*N'*-1-naphthylmethyliminomethylidene)-1,2-dideoxy-6-**

**thionojirimycin Hydrochloride (37).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 104 mg (quantitative). *R<sub>f</sub>* 0.62 (50:10:1 DCM-MeOH-H<sub>2</sub>O).

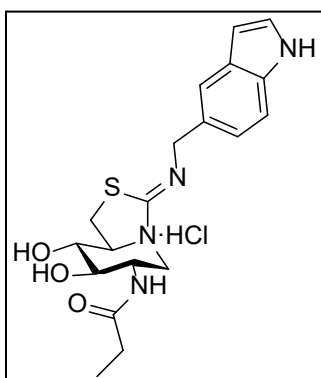
**<sup>1</sup>H NMR** (500 MHz, 1:1 CD<sub>3</sub>OD-CDCl<sub>3</sub>): δ 8.03-7.43 (m, 8 H, CH<sub>arom</sub>), 4.90 (bs, 2H, CH<sub>2</sub>Naph), 4.22 (dd, 1 H, *J*<sub>1a,1b</sub> = 13.1 Hz, *J*<sub>1a,2</sub> = 5.2 Hz, H-1a), 3.86 (m, 1 H, H-2), 3.65 (m, 1 H, H-5), 3.58



H, 5.82; N, 9.30; S, 7.03.

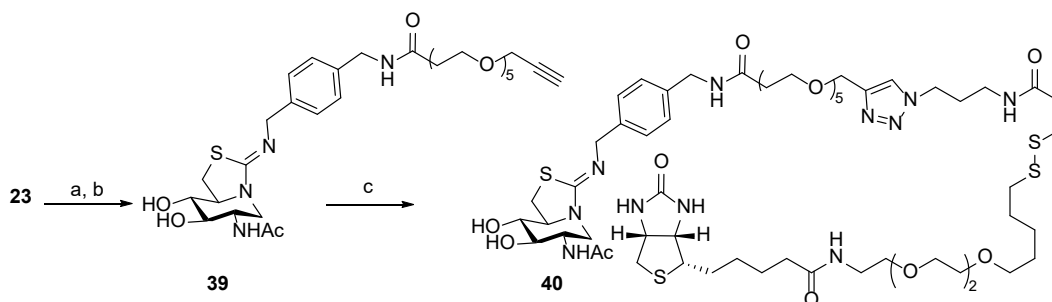
**(Z)-2-Propionamido-5-N,6-S-(N'-5-indolylmethyliminomethylidene)-1,2-dideoxy-6-**

**thionojirimycin Hydrochloride (38).** Column chromatography, eluent: 70:10:1 DCM-MeOH-H<sub>2</sub>O. Yield: 104 mg (quantitative). *R<sub>f</sub>* 0.40 (50:10:1 DCM-MeOH-H<sub>2</sub>O).



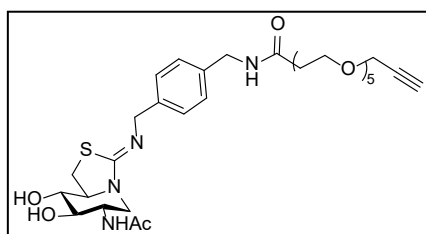
**<sup>1</sup>H NMR** (500 MHz, CD<sub>3</sub>OD): δ 7.57-6.46 (m, 5 H, CH<sub>arom</sub>), 4.60 (bs, 2H, CH<sub>2</sub>Ph), 4.18 (dd, 1 H, *J*<sub>1a,1b</sub> = 13.3 Hz, *J*<sub>1a,2</sub> = 5.3 Hz, H-1a), 3.88 (m, 1 H, H-2), 3.83 (m, 1 H, H-5), 3.68 (m, 1 H, *J*<sub>6a,6b</sub> = 11.3 Hz, *J*<sub>5,6a</sub> = 7.4 Hz, H-6a), 3.52 (t, 1 H, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> = 9.0, H-3), 3.42 (m, 2 H, H-4, H-6b), 2.93 (dd, 1 H, *J*<sub>1b,2</sub> = 11.5 Hz, H-1b), 2.29 (q, 2 H, CH<sub>2</sub>CH<sub>3</sub>), 1.16 (t, 3 H, CH<sub>2</sub>CH<sub>3</sub>). **<sup>13</sup>C NMR** (125.7 MHz, CD<sub>3</sub>OD): δ 177.7 (CO), 169.9 (CN), 137.4-102.5 (C<sub>arom</sub>), 75.9 (C-3), 75.1 (C-4), 68.4 (C-5), 55.8 (CH<sub>2</sub>Ph), 51.0 (C-2), 47.2 (C-1), 32.2 (C-6), 30.3 (CH<sub>2</sub>CH<sub>3</sub>), 10.4 (CH<sub>2</sub>CH<sub>3</sub>). **HRMS** (ESI) *m/z* [M + H]<sup>+</sup> calcd for [C<sub>19</sub>H<sub>25</sub>N<sub>4</sub>O<sub>3</sub>S]<sup>+</sup> 389.1642; found 389.1637. Anal. Calcd for C<sub>19</sub>H<sub>25</sub>N<sub>4</sub>O<sub>3</sub>S.HCl: C 53.70, H 5.93, N 13.18, S 7.54. Found: C, 53.48; H, 5.70; N, 12.86; S, 7.21.

**Synthesis of DNJNAc-6S-NBn-Biotin probe**



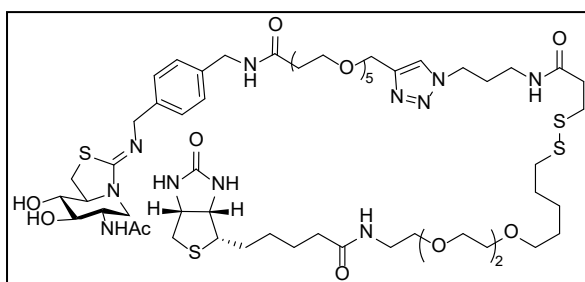
**Supplemental scheme 8.** Reagents and conditions: a. PPh<sub>3</sub>, THF-H<sub>2</sub>O, RT, 18 h; b. alkyne-PEG4-acid, HBTU, DIPEA, DMF, RT, 24 h; c. biotin-azide, sodium ascorbate, CuSO<sub>4</sub>·5H<sub>2</sub>O, DMF-H<sub>2</sub>O, RT, 18 h.

**Alkyne intermediate (39).** A mixture of inhibitor **23** (13.5 mg, 0.0316 mM) and PPh<sub>3</sub> (9.1 mg, 0.0348 mM) in 4:1 THF-H<sub>2</sub>O (2.5 mL) was stirred overnight and then concentrated with co-evaporation with 1:1 toluene-*i*PrOH. The residue was dried under vacuum for 2.5 h. To the residue was added HBTU (16.8 mg, 0.0442 mM) and alkyne-PEG4-acid (12.5 mg, 0.0411 mM) followed by anhydrous DMF (1.5 mL) and DIPEA (16.6 μL, 0.0948 mM). The reaction mixture was stirred overnight and then concentrated with co-evaporation with toluene. The resulting residue was purified by CombiFlash (dry load, 4 g RediSep Gold Column, 18 mL/min) using a gradient of 1:0:0→6:3:1 EtOAc-MeOH-H<sub>2</sub>O to give intermediate **39** (12.9 mg, 64%) as a clear viscous oil. *R*<sub>f</sub> 0.41 (9:3:1 EtOAc-MeOH-H<sub>2</sub>O).



**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 7.27 (d, 2 H, *J* = 8.3 Hz, CH-arom), 7.25 (d, 2 H, *J* = 8.6 Hz, CH<sub>arom</sub>), 4.39 (d, 1 H, *J* = 14.8 Hz, PhCH<sub>2</sub>N), 4.38 (s, 2 H, PhCH<sub>2</sub>N), 4.33 (d, 1 H, *J* = 14.9 Hz, ArCH<sub>2</sub>N), 4.19 (s, 2 H, OCH<sub>2</sub>C), 4.16 (dd, 1 H, *J* = 12.7, 5.2 Hz, H-1), 3.86 (ddd, 1 H, *J* = 11.2, 10.2, 5.2 Hz, H-2), 3.77 (t, 2 H, *J* = 6.0 Hz, COCH<sub>2</sub>CH<sub>2</sub>O), 3.71–3.62 (m, 16 H, 8 × OCH<sub>2</sub>), 3.47 (dd, 1 H, *J* = 10.9, 5.9 Hz, H-6), 3.40 (dd, 1 H, *J* = 10.2, 8.4 Hz, H-3), 3.36–3.32 (m, 2 H, H-4, H-5), 3.18 (dd, 1 H, *J* = 10.8, 7.1 Hz, H-6'), 2.89 (t, 1 H, *J* = 2.4 Hz, C<sub>q</sub>CH), 2.59 (dd, 1 H, *J* = 12.8, 11.3 Hz, H-1'), 2.51 (t, 2 H, *J* = 6.0 Hz, COCH<sub>2</sub>CH<sub>2</sub>O), 1.99 (s, 3 H, COCH<sub>3</sub>). **<sup>13</sup>C NMR** (150 MHz, CD<sub>3</sub>OD): δ 172.5, 172.2 (CO), 161.8 (CN), 139.2, 136.9, 127.2, 127.1 (C<sub>arom</sub>), 79.2 (C<sub>q</sub>CH), 75.6 (C3), 74.1 (C4), 70.1 (OCH<sub>2</sub>), 70.0 (OCH<sub>2</sub>), 69.9 (OCH<sub>2</sub>), 68.7 (OCH<sub>2</sub>), 66.9 (COCH<sub>2</sub>CH<sub>2</sub>O), 64.6 (C5), 57.6 (OCH<sub>2</sub>C<sub>q</sub>, NCH<sub>2</sub>Ar), 57.5 (NCH<sub>2</sub>Ar), 49.9 (C2), 45.8 (C1), 36.3 (COCH<sub>2</sub>CH<sub>2</sub>), 29.9 (C6), 21.4 (COCH<sub>3</sub>). **HRMS (ESI)** *m/z* [M+H]<sup>+</sup> calcd for C<sub>31</sub>H<sub>47</sub>N<sub>4</sub>O<sub>9</sub>S: 651.3064, found 651.3631.

**DNJNAc-6S-NBn-Biotin (40).** To a mixture of intermediate **39** (4.7 mg, 0.0072 mmol) and biotin-SS-azide (5.5 mg, 0.0079 mM) obtained from ClickChemistryTools (1168-25.) in DMF (250 μL) was added a solution of sodium ascorbate (1.4 mg, 0.00722 mM) and CuSO<sub>4</sub>·5H<sub>2</sub>O (0.9 mg, 0.00361 mM) in H<sub>2</sub>O (250 μL). The reaction mixture was stirred overnight at RT. The mixture was concentrated and co-evaporated with a 1:1 toluene-*i*PrOH solution, then dried under vacuum for 1 h. The resulting residue was purified by HPLC (XDB C18 semi-prep column) to



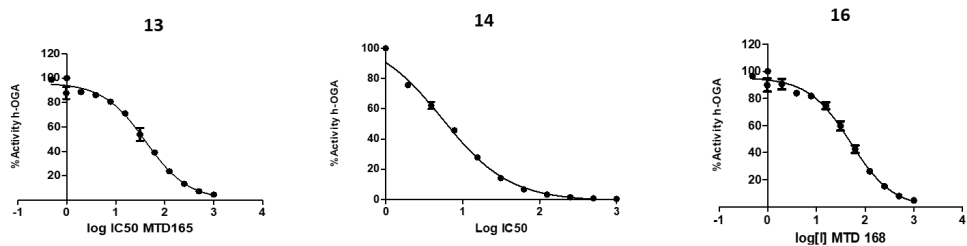
give the final DNJNAc-6S-NBn-Biotin (**40**) (3.0 mg, 31%) as a colorless oil. *R*<sub>f</sub> 0.20 (4:3:1 EtOAc-MeOH-H<sub>2</sub>O).

**<sup>1</sup>H NMR** (600 MHz, CD<sub>3</sub>OD): δ 8.05 (s, 1 H, CH<sub>arom</sub>), 7.39–7.32 (m, 4 H, CH<sub>arom</sub>), 4.64 (s,

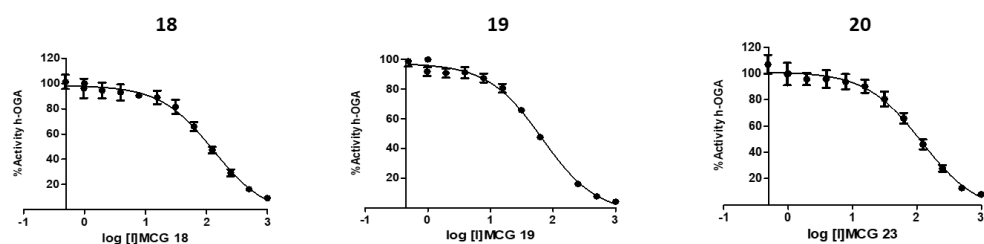
2 H,  $\text{OCH}_2\text{Ph}$ ), 4.60 (s, 2 H,  $\text{NCH}_2\text{Ph}$ ), 4.52 (dd, 1H,  $J = 13.3, 5.1$  Hz,  $\text{CHN}_{\text{biotin}}$ ), 4.47 (t, 2 H,  $J = 6.8$  Hz,  $\text{N}_{\text{Ar}}\text{CH}_2\text{CH}_2\text{CH}_2$ ), 4.42 (s, 2 H,  $\text{ArCH}_2\text{N}$ ), 4.33 (dd, 1 H,  $J = 7.9, 4.4$  Hz,  $\text{CHN}_{\text{biotin}}$ ), 4.18 (dd, 1 H,  $J = 12.4, 5.2$  Hz, H-1), 4.09 (app. q, 1 H,  $J = 8.3$  Hz, H-5), 3.86 (ddd, 1 H,  $J = 11.9, 10.7, 5.1$  Hz, H-2), 3.81–3.75 (m, 3 H, H-6,  $\text{COCH}_2\text{CH}_2\text{O}_{\text{L1}}$ ), 3.73–3.45 (m, 28 H, H-3, H-4,  $\text{COCH}_2\text{CH}_2\text{O}_{\text{L2}}$ ,  $13 \times \text{OCH}_2$ ), 3.42–3.37 (m, 4H,  $\text{OCH}_2\text{CH}_2\text{N}_{\text{L2}}$ ,  $\text{SCH}_2\text{CH}_2\text{N}_{\text{L2}}$ ), 3.26–3.21 (m, 3 H,  $\text{CHS}_{\text{biotin}}$ ,  $\text{CH}_2\text{CH}_2\text{CH}_2\text{N}_{\text{(L2)}}$ ), 3.08 (dd, 1 H,  $J = 12.2, 12.2$  Hz, H-1'), 3.03–2.91 (m, 5 H,  $\text{CH}_2\text{S}_{\text{biotin}}$ ,  $\text{CH}_2\text{CH}_2\text{S}_{\text{L2}}$ ,  $\text{SCH}_2\text{CH}_2\text{(L2)}$ ), 2.73 (d, 1 H,  $J = 12.7$  Hz,  $\text{CH}_2'\text{S}_{\text{biotin}}$ ), 2.66–2.61 (m, 4 H,  $\text{COCH}_2\text{CH}_2\text{S}_{\text{L2}}$ ,  $\text{COCH}_2\text{CH}_2\text{O}_{\text{L2}}$ ), 2.52 (t, 2 H,  $J = 5.8$  Hz,  $\text{COCH}_2\text{CH}_2\text{O}_{\text{L1}}$ ), 2.24 (t, 2 H,  $J = 7.3$  Hz,  $\text{NCOCH}_2\text{CH}_2\text{(biotin)}$ ), 2.13 (app. t, 2 H,  $J = 6.8$  Hz,  $\text{NCH}_2\text{CH}_2\text{CH}_2\text{N}_{\text{L2}}$ ), 2.02 (s, 3 H,  $\text{COCH}_3$ ), 1.83–1.55 (m, 4 H,  $\text{CH}_2\text{(biotin)}$ ,  $\text{CH}_2\text{(biotin)}$ ), 1.49–1.44 (m, 2 H,  $\text{CH}_2\text{(biotin)}$ ).  $^{13}\text{C}$  NMR (150 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  174.7, 172.7, 172.5, 172.3, 172.0, 139.4, 133.5, 127.7, 127.6, 73.8, 73.2, 70.2, 70.1 (4 $\times$ ), 70.0 (2 $\times$ ), 69.8 (3 $\times$ ), 69.4, 69.2, 69.1, 68.1, 66.9, 62.0, 60.2, 55.6, 51.1, 49.5, 45.6, 42.2, 39.7, 39.1, 39.0, 36.4, 36.1, 35.4, 35.1, 35.0 (2 $\times$ ), 33.7, 33.6, 31.1, 29.7, 28.4, 28.1, 25.5, 21.3. **HRMS (ESI)**  $m/z$   $[\text{M}+\text{NH}_4]^+$  calcd for  $\text{C}_{59}\text{H}_{101}\text{N}_{12}\text{O}_{15}\text{S}_4$ : 1345.5828, found 1345.5766.



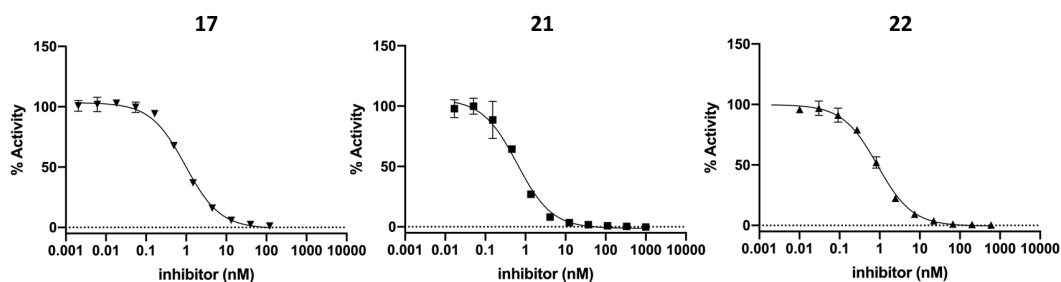
## Morrison $K_i$ determination against human OGA, HexA and HexB



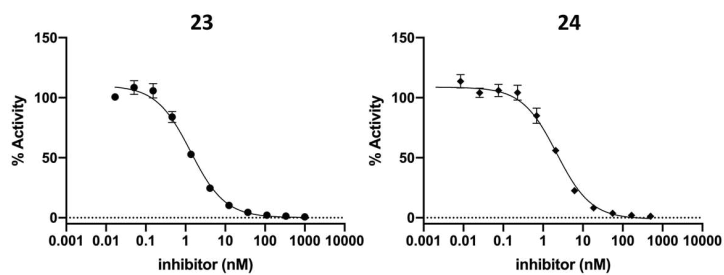
**Supplemental Figure 1.**  $K_i$  determination of **13** ( $24 \pm 8$  nM), **14** ( $20 \pm 7$  nM) and **16** ( $3 \pm 1$  nM) against hOGA.



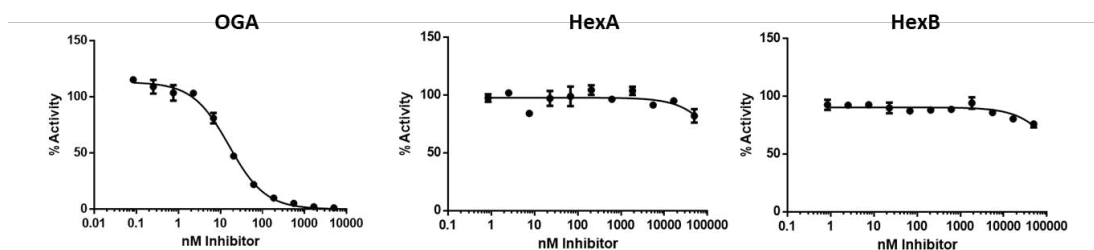
**Supplemental Figure 2.**  $K_i$  determination of **18** ( $50 \pm 20$  nM), **19** ( $30 \pm 10$  nM), and **20** ( $50 \pm 20$  nM), against hOGA.



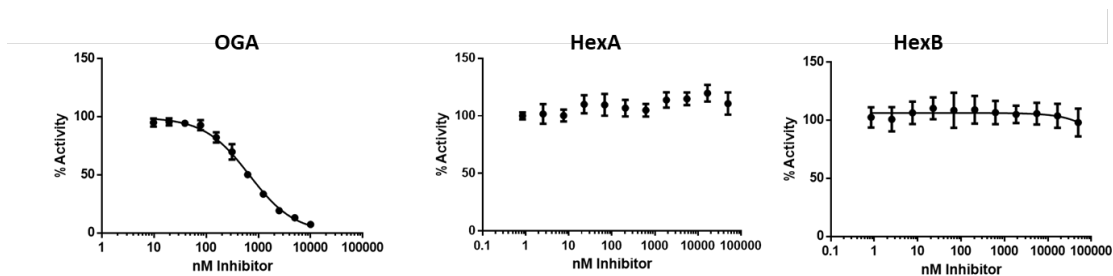
**Supplemental Figure 3.**  $K_i$  determination of **17** ( $0.3 \pm 0.1$  nM), **21** ( $0.27 \pm 0.09$  nM), and **22** ( $0.22 \pm 0.07$  nM), against hOGA.



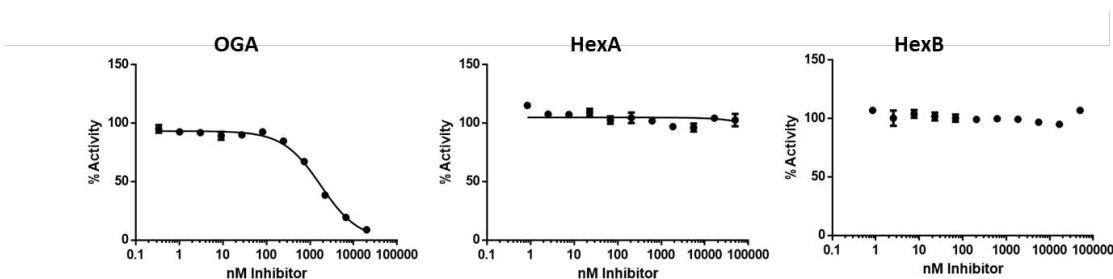
**Supplemental Figure 4.**  $K_i$  determination of **23** ( $0.4 \pm 0.2$  nM), and **24** ( $0.9 \pm 0.3$  nM), against hOGA.



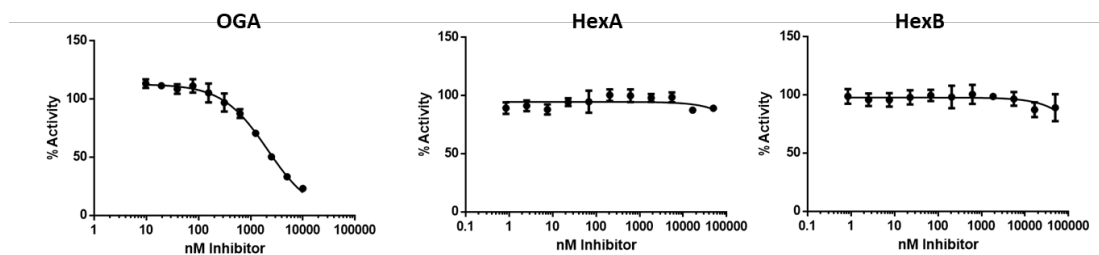
**Supplemental Figure 5.**  $K_i$  determination of **28** against hOGA ( $8 \pm 3$  nM), HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



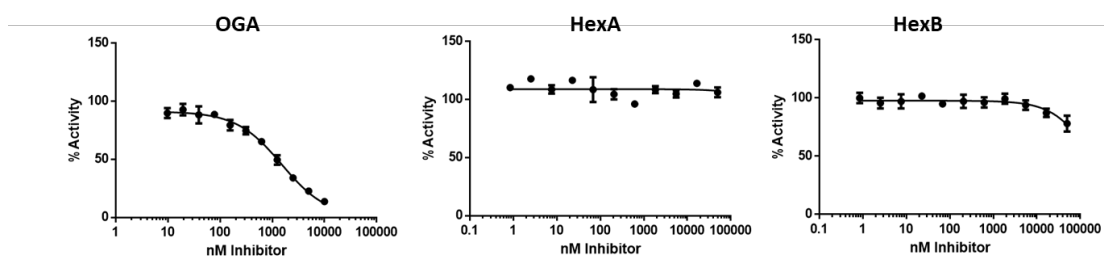
**Supplemental Figure 6.**  $K_i$  determination of **29** ( $900 \pm 300$  nM) against hOGA, HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



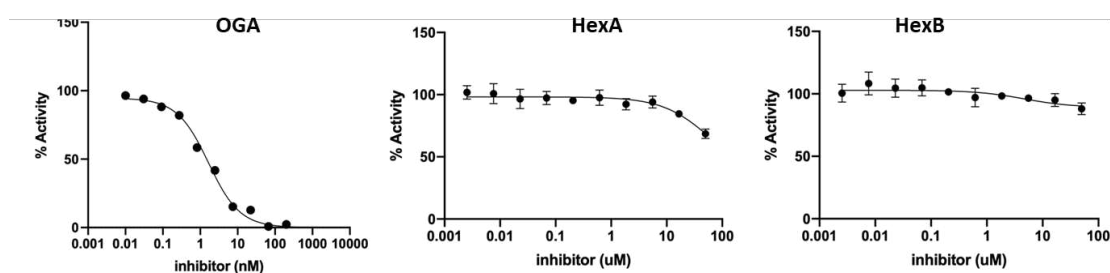
**Supplemental Figure 7.**  $K_i$  determination of **30** ( $300 \pm 100$  nM) against hOGA, HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



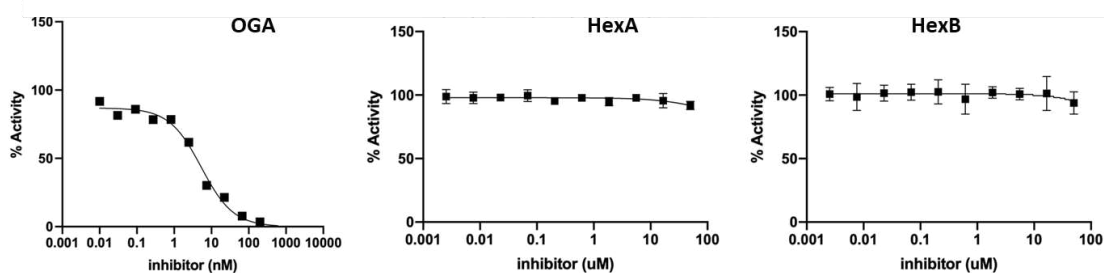
**Supplemental Figure 8.**  $K_i$  determination of **31** against hOGA ( $1000 \pm 400$  nM), HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



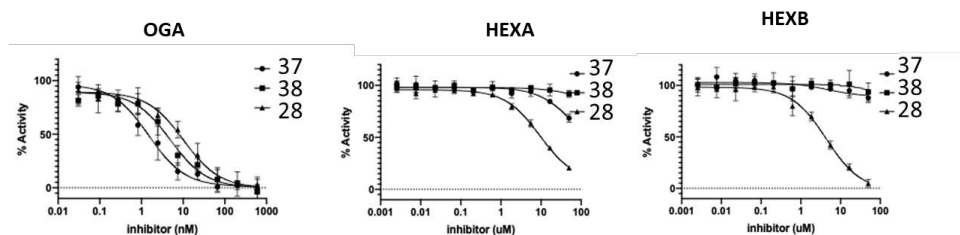
**Supplemental Figure 9.**  $K_i$  determination of **32** against hOGA ( $800 \pm 300$  nM), HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



**Supplemental Figure 10.**  $K_i$  determination of **37** against hOGA ( $0.8 \pm 0.2$  nM), HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).

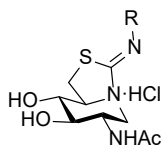


**Supplemental Figure 11.**  $K_i$  determination of **38** against hOGA ( $2.4 \pm 0.9$  nM), HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).



**Supplemental Figure 12.** Comparative  $K_i$  determination of **28**, **37** and **38** against hOGA, HexA ( $<100$   $\mu$ M) and HexB ( $<100$   $\mu$ M).

### Inhibition of compounds 1-24 against commercial hexosaminidases.



**Supplemental Table 1.** Inhibition constants ( $K_i$ ) for DNJNAc-thiazolidines **13-24** against commercial hexosaminidases

Compound	Structure (R)	Bovine kidney Hex	Human placenta Hex
		$K_i$ ( $\mu$ M)	$K_i$ ( $\mu$ M)
<b>13</b>	(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	$0.57 \pm 0.04$	$0.24 \pm 0.02$
<b>14</b>	(CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub>	$0.19 \pm 0.02$	$0.24 \pm 0.02$
<b>15</b>	Ph	$0.70 \pm 0.05$	$0.58 \pm 0.04$
<b>16</b>	CH <sub>2</sub> Ph	$0.46 \pm 0.03$	$0.42 \pm 0.03$
<b>17</b>	CH <sub>2</sub> -1-naphthyl	$0.89 \pm 0.8$	n.d. <sup>a</sup>
<b>18</b>	CH <sub>2</sub> p-FPh	$1.5 \pm 0.2$	$0.28 \pm 0.02$
<b>19</b>	CH <sub>2</sub> p-OMePh	$1.2 \pm 0.1$	$0.73 \pm 0.04$
<b>20</b>	CH <sub>2</sub> p-CF <sub>3</sub> Ph	$0.43 \pm 0.04$	$0.40 \pm 0.03$
<b>21</b>	CH <sub>2</sub> o-CH <sub>2</sub> N <sub>3</sub> Ph	$0.91 \pm 0.09$	n.d. <sup>a</sup>
<b>22</b>	CH <sub>2</sub> m-CH <sub>2</sub> N <sub>3</sub> Ph	$0.83 \pm 0.06$	n.d. <sup>a</sup>
<b>23</b>	CH <sub>2</sub> p-CH <sub>2</sub> N <sub>3</sub> Ph	$0.72 \pm 0.05$	n.d. <sup>a</sup>
<b>24</b>	CH <sub>2</sub> CH <sub>2</sub> Ph	$1.4 \pm 0.12$	n.d. <sup>a</sup>

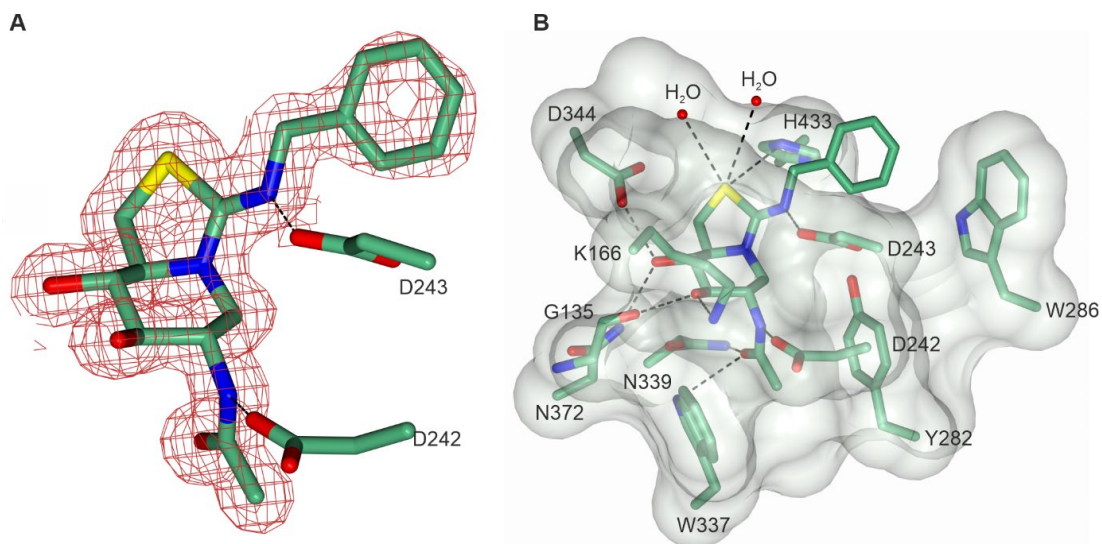
## **X-Ray structure solution methods for hOGA and *BtGH84* in complex with compound 16**

hOGA was produced and purified as a split construct, DNA encoding residues 11-396 and 535-715 was inserted in two separate plasmids, according to the protocol published by Roth, *et al.* 2017<sup>6</sup>. hOGA was crystallised by hanging drop vapour diffusion method at a concentration of 30 mg mL<sup>-1</sup> under the conditions of 200 mM Triammonium citrate pH 7.5 and 24% Polyethylene glycol 3350. A seed stock was added at 20x diluted concentration in a 1 : 0.2 : 0.8  $\mu$ L ratio of protein : seed stock : reservoir. *BtGH84* was produced and purified as described previously by Dennis *et al.*, 2006<sup>7</sup>. *BtGH84* was crystallised by sitting drop vapour diffusion method at a concentration of 12 mg mL<sup>-1</sup> under the conditions of 125 mM imidazole pH8, 8% PEG 8000, 3% trimethylamine N-oxide dehydrate, 15% ethylene glycol in a 1:1 ratio of protein to reservoir solution. Crystals were soaked with compound **16** at a final concentration of 5 mM for either 2 hours or 7 days, *BtGH84* and hOGA respectively. Diffraction data were collected on the I03 and I04-1 beamlines at Diamond Light Source (DLS). The data were indexed, integrated and scaled using Xia2<sup>8</sup> and reintegrated using Aimless<sup>9,10</sup> from the CCP4 software<sup>11</sup>; the data collection statistics are shown in table 2. Molecular replacement was conducted using Molrep<sup>12</sup> with coordinates 5M7R and 5ABE for hOGA and *BtGH84*, respectively. The model was adjusted and refined using alternating cycles of manual model building and real space refinement in Coot<sup>13</sup> and reciprocal space refinement in Refmac<sup>14-17</sup>. Data quality was assessed by the multimeric model geometry validation programme<sup>18</sup> in CCP4 and Privateer<sup>19</sup> was used to assess the confidence of the carbohydrate ligand conformation. Figures were produced using CCP4mg<sup>20</sup>.

**Supplemental Table 2.** Data collection and refinement statistics for hOGA and BtGH84 enzymes in complex with compound **16**.

	<i>hOGA_16</i>	<i>BtGH84_16</i>
PDB code	<b>7OU6</b>	<b>7OU8</b>
<b>Data collection</b>		
Beamline	Diamond I03	Diamond I04-1
Wavelength (Å)	0.9763	0.9159
Space group	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 2 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions: <i>a</i> , <i>b</i> , <i>c</i> (Å)	101.29, 101.29, 284.52	51.53, 160.62, 224.51
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	58.21-2.41 (2.48-2.41)	75.62-1.50 (1.53-1.50)
<i>R</i> <sub>merge</sub>	0.05 (2.28)	0.07 (0.81)
<i>R</i> <sub>pim</sub>	0.01 (0.59)	0.03 (0.37)
<i>CC</i> (1/2)	1.00 (0.56)	1.00 (0.65)
<i>I</i> / $\sigma$ <i>I</i>	26.5 (1.5)	12.3 (1.5)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Redundancy	15.8 (15.6)	7.5 (5.8)
<b>Refinement</b>		
Resolution (Å)	58.21-2.41 (2.48-2.41)	75.62-1.50 (1.53-1.50)
No. reflections	58168	298007
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.22/0.27	0.21/0.23
No. atoms		
Protein	6704	11529
Ligands/Ions	46	50/1
Water	18	903
<i>B</i> -factors (Å <sup>2</sup> )		
Protein	89	28
Ligand/Ions	82	19/23
Water	65	31
R.m.s. deviations		

Bond lengths (Å)	0.007	0.011
Bond angles (°)	1.52	1.74
Ramachandran		
Favoured (%)	94.2	95.8
Allowed (%)	4.7	3.6
Outliers (%)	1.1	0.6



**Supplemental Figure 13.** Structure of *BtGH84* in complex with compound **16**. Active site residues of *BtGH84* are shown in green. (A) **16** is shown in dark green with the corresponding REFMAC maximum-likelihood/ $\sigma$ A-weighted  $2F_o - F_c$  map, shown in red, contoured at 0.46 electrons  $\text{\AA}^{-3}$ . (B) Surface representation of the active site pocket of *BtGH84* with neighbouring residues, waters and hydrogen bond interactions highlighted.

## **Chemoproteomic and mapping of post-transcriptional modifications**

**Preliminary pull-down experiments with recombinant hOGA.** Recombinant hOGA was added to lysis buffer (25 mM HEPES, 150 mM NaCl, 5 mM EDTA, 5% glycerol, pH 7.1) containing protease inhibitor cocktail (Roche cOmplete™, EDTA free) at a concentration of 100 nM at 4 °C. DNJNAc-6S-NBn-Biotin was added at concentration of 500 nM and incubated with rotation for 1 h at 4 °C. The sample was then incubated with the streptavidin beads (Tri-Link Biotechnologies) for 1 h at 4 °C. The beads were separated and washed with lysis buffer 5 times at 4 °C. The beads were then eluted with 100 mM DTT in lysis buffer.

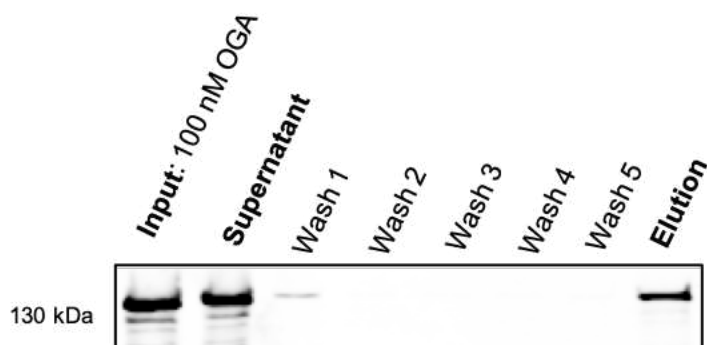
**Large scale precipitation of bovine OGA.** Bovine brains were obtained from Animal Technologies, Inc. The lysate was prepared using a 4:1 lysis buffer: dry cell extractions. The entire brain was homogenized by cryomilling under liquid nitrogen. The brain tissue was processed using a dounce homogenizer in 4:1 lysis buffer (25 mM HEPES, 133 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM EDTA, 10% Glycerol, pH 7.1, SIGMAFAST™ Protease Inhibitor Cocktail) to dry tissue, at 4 °C. The lysate was cleared by centrifugation (Sorvall RC-6 Plus) at 20,000 RPM for 30 min at 4 °C. The supernatant was collected and the remaining debris was removed by ultracentrifugation (Beckman Optima Ultracentrifuge L-80 XP) at 45,000 RPM for 45 min at 4 °C. The supernatant was collected and separated into fresh tubes. The samples were incubated for 30 min at 4 °C with streptavidin beads to deplete endogenously biotinylated proteins. The sample was centrifuged and the supernatant collected to remove the beads from the sample after which 128 nM of DNJNAc-6S-NBn-Biotin was added and the sample was mixed by rotation for 1 hr at 4 °C. A parallel control experiment was performed in which we added a high concentration of Thiamet-G (1 mM) to block binding to the DNJNAc-6S-NBn-Biotin ligand. The sample and control were then incubated with the beads (200 µL) for 1 hr at 4 °C. The beads were separated and washed with cold lysis buffer 5 times at 4 °C. The affinity purified protein and control were then eluted by addition of DTT in lysis buffer to provide a final concentration of 60 mM and the samples were mixed using a thermomixer (Eppendorf) at 37 °C for 10 min.

**LC-MS/MS sample preparation and analysis.** The samples were loaded onto 12% SDS-PAGE minigels (Bio-Rad), electrophoresed, and stained with Coomassie blue G-250 (Thermo Fisher Scientific). The band corresponding to OGA at the 130 kDa range was excised, destained, protein disulfide bonds were reduced (10 mM DTT in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, 45 min, 56 °C), after which thiols were alkylated (55 mM iodoacetamide in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, dark, 30 min, room temperature). Protein was subsequently digested (trypsin, Promega, in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, overnight, 37 °C) and the resulting peptides were extracted with a gradient of acetonitrile solutions starting with aqueous 0.1% formic acid. Peptide mixtures were purified by solid-phase extraction

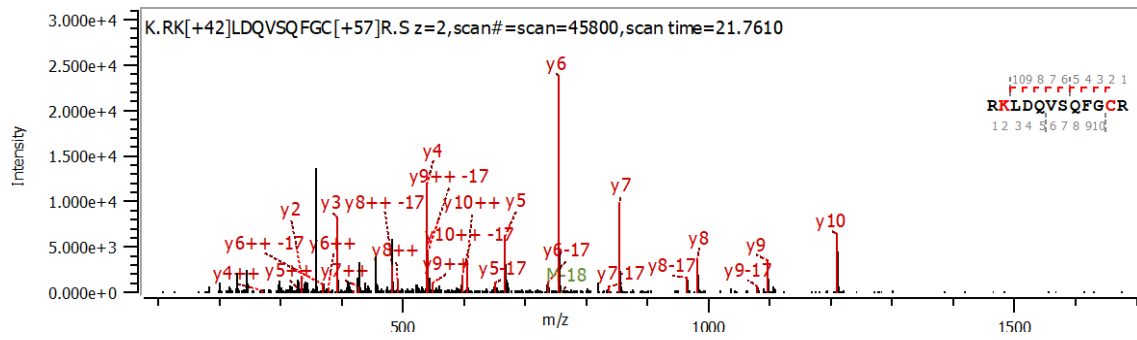


using C18 StageTips (Thermo Fisher Scientific). 200 ng of sample was injected into LC-MS/MS, analysis was performed using a Bruker timfTOF coupled to a Bruker nanoElute UHPLC system (Thermo Fisher Scientific).

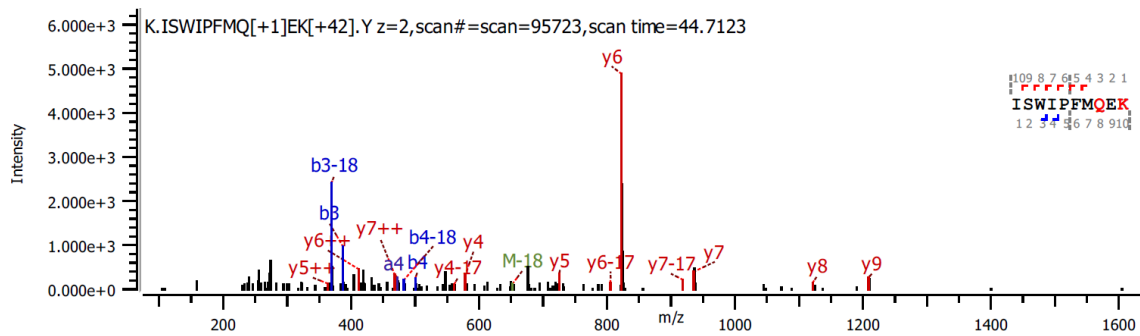
Acquired data was then searched against the Uniprot protein database for *Bos taurus* using the Byonic™ (v4.0.12 – ProteinMetrics Inc.) search algorithm from Protein Metrics Inc, with 15 ppm and 40 ppm mass accuracies for precursor and product ion masses respectively, and a 1% false discovery rate cut-off. Assigning of peptide and protein identifications included carbamidomethyl @ C (fixed), oxidation @ M (common1) and deamidation @ N,Q (common1) modifications, RK fully specific digestion (maximum two missed cleavages), 15 and 40 ppm precursor and fragment mass tolerance. Several variable modifications were selected for a second pass search, these can be found in Supplemental Spreadsheet. Focusing on the enzymatic modifications, an FDR criterion was defined for the modifications identified on a peptide hits as those with a  $|\text{Log Prob}| > 1.52$  (3%FDR), and high-confidence modifications were assigned as  $|\text{Log Prob}| > 2.0$  (1% FDR). (Additional information regarding search parameters selected can be found in Supplemental Spreadsheet)



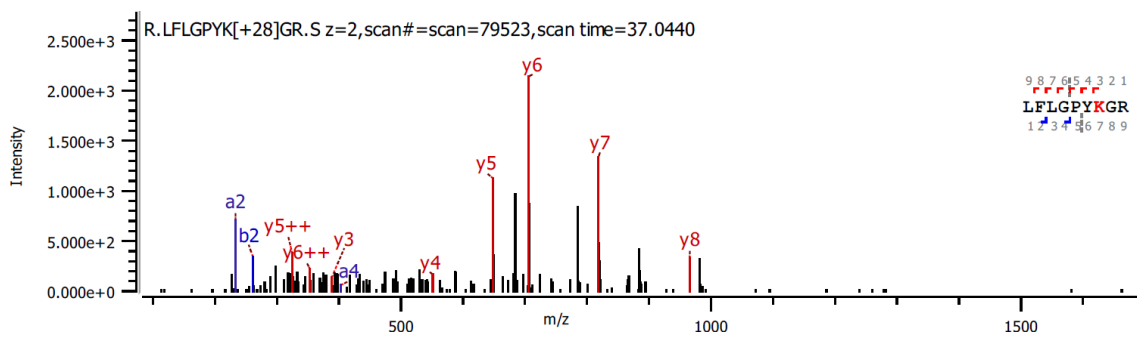
**Supplemental Figure 14.** Small scale validation experiment showing **DNJNAc-6S-NBn-Biotin** enables enrichment of recombinant hOGA as determined using immunoblot assay.



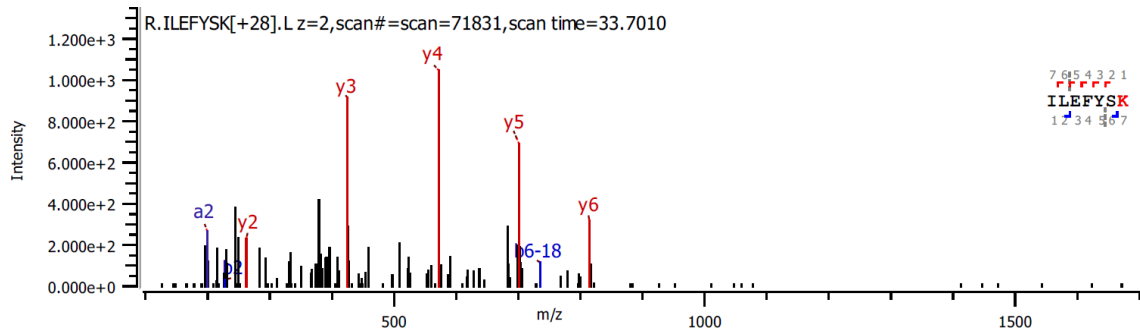
Supplemental Figure 15. MS/MS spectrum for acetylation identified at K157.



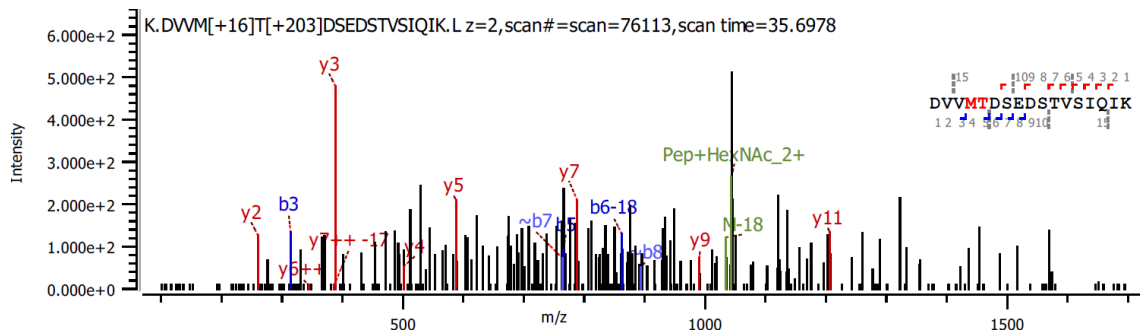
Supplemental Figure 16. MS/MS spectrum for acetylation identified at K805.



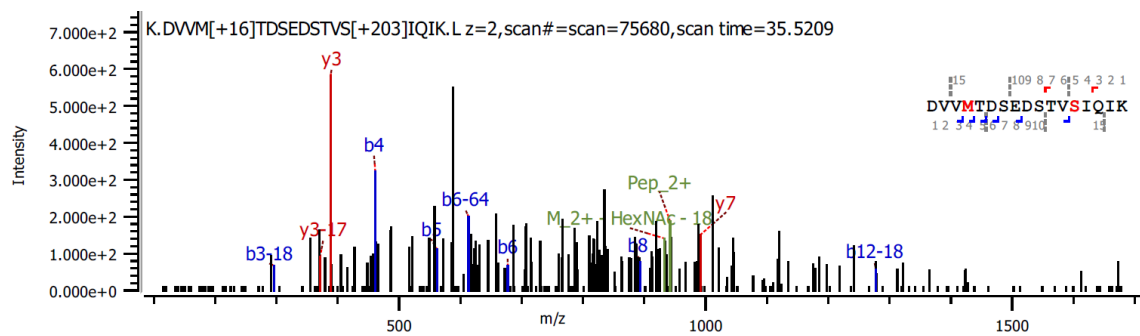
Supplemental Figure 17. MS/MS spectrum for formylation identified at K297.



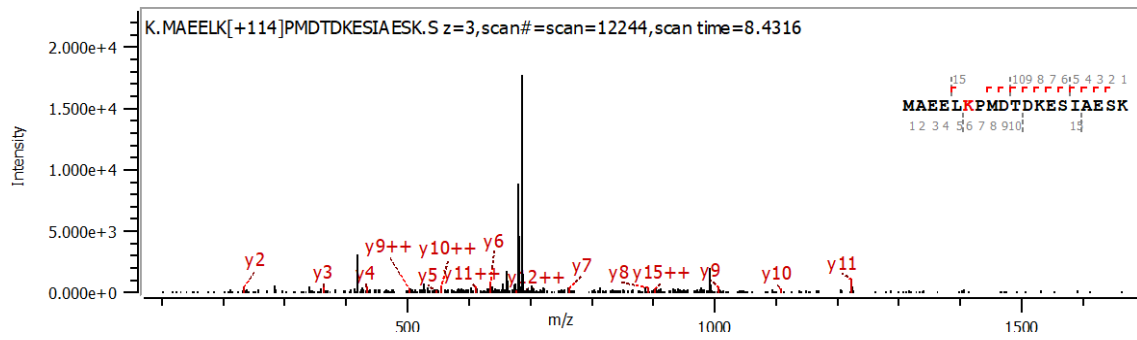
Supplemental Figure 18. MS/MS spectrum for formylation identified at K893.



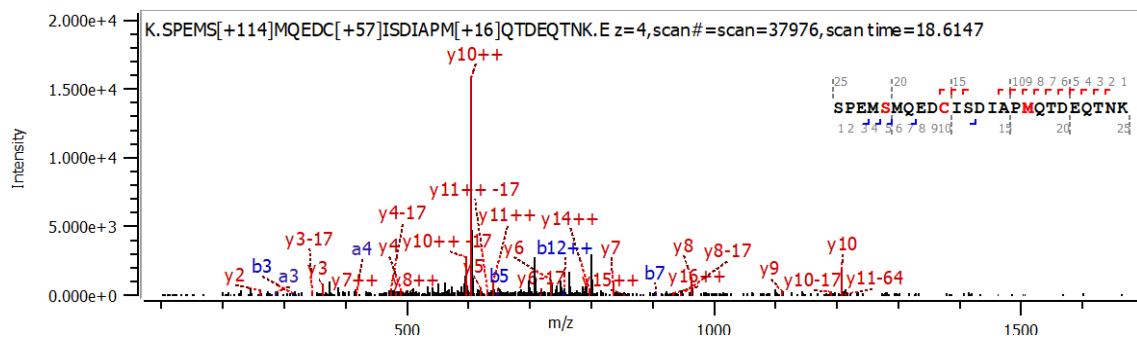
Supplemental Figure 19. MS/MS spectrum for O-GlcNAcylation identified at T346.



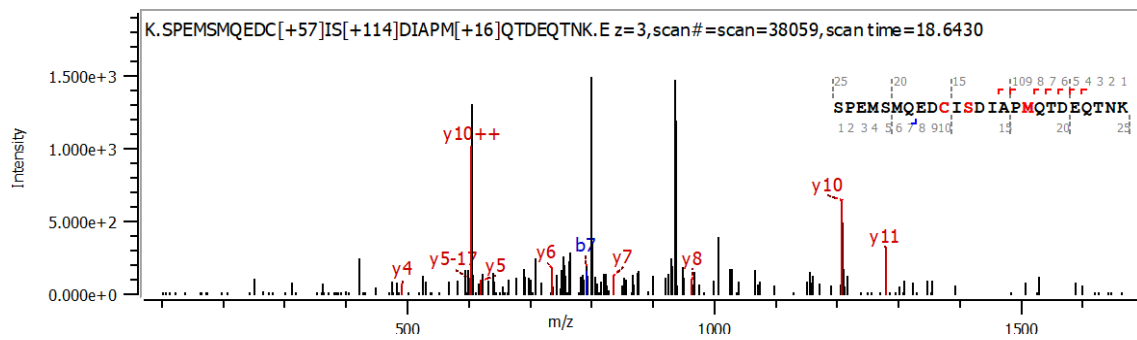
Supplemental Figure 20. MS/MS spectrum for O-GlcNAcylation identified at S354.



Supplemental Figure 21. MS/MS spectrum for N-ubiquitination identified at K497.



Supplemental Figure 22. MS/MS spectrum for O-ubiquitination identified at S515.



Supplemental Figure 23. MS/MS spectrum for O-ubiquitination identified at S522.

**Supplemental Table 3.** Fragment ion series of mass spectra from acetylated Lys157 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	129.1135			157.1084			139.0978			R							12
2	299.219			327.2139			309.2033			K	1379.6736			690.3404			11
3	412.3031			440.298			422.2874			L	1209.5681	1209.5676	-0.0005	605.2877	605.2874	-0.0002	10
4	527.33			555.3249			537.3143			D	1096.484	1096.4829	-0.0011	548.7456	548.743	-0.0027	9
5	655.3886			683.3835			665.3729			Q	981.4571	981.4567	-0.0004	491.2322	491.2281	-0.0041	8
6	754.457			782.4519			764.4413			V	853.3985	853.3932	-0.0053	427.2029	427.1967	-0.0062	7
7	841.489			869.4839			851.4734			S	754.3301	754.328	-0.002	377.6687	377.664	-0.0047	6
8	969.5476			997.5425			979.5319			Q	667.298	667.2953	-0.0027	334.1527	334.1502	-0.0025	5
9	1116.616			1144.6109			1126.6004			F	539.2395	539.2367	-0.0028	270.1234	270.1241	0.0007	4
10	1173.6375			1201.6324			1183.6218			G	392.1711	392.1692	-0.0019	196.5892			3
11	1333.6681			1361.663			1343.6525			C	335.1496	335.1479	-0.0017	168.0784			2
12										R	175.119			88.0631			1

**Supplemental Table 4.** Fragment ion series of mass spectra from acetylated Lys805 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	86.0964			114.0913			96.0808			I							10
2	173.1285			201.1234			183.1128			S	1208.5656	1208.5888	0.0232	604.7864			9
3	359.2078	359.2091	0.0013	387.2027	387.1984	-0.0042	369.1921	369.1891	-0.0031	W	1121.5336	1121.5592	0.0256	561.2704			8
4	472.2918	472.2934	0.0016	500.2867	500.2866	-0.0002	482.2762			I	935.4543	935.4616	0.0073	468.2308			7
5	569.3446			597.3395			579.3289			P	822.3702	822.3776	0.0074	411.6887	411.6924	0.0036	6
6	716.413			744.4079			726.3973			F	725.3174	725.3282	0.0107	363.1624			5
7	847.4535			875.4484			857.4378			M	578.249	578.2581	0.009	289.6282			4
8	976.4961			1004.491			986.4804			Q	447.2086			224.1079			3
9	1105.5387			1133.5336			1115.523			E	318.166			159.5866			2
10										K	189.1234			95.0653			1

**Supplemental Table 5.** Fragment ion series of mass spectra from formylated Lys297 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	86.0964			114.0913			96.0808			L							9
2	233.1648	233.1635	-0.0014	261.1597	261.1567	-0.0031	243.1492			F	965.5203	965.5163	-0.004	483.2638			8
3	346.2489			374.2438			356.2332			L	818.4519	818.4503	-0.0016	409.7296			7
4	403.2704	403.2642	-0.0061	431.2653			413.2547			G	705.3678	705.3676	-0.0002	353.1876	353.1834	-0.0042	6
5	500.3231			528.318			510.3075			P	648.3464	648.3451	-0.0013	324.6768	324.6788	0.002	5
6	663.3864			691.3814			673.3708			Y	551.2936	551.2972	0.0036	276.1504			4
7	819.4763			847.4712			829.4607			K	388.2303	388.2242	-0.0061	194.6188			3
8	876.4978			904.4927			886.4821			G	232.1404			116.5738			2
9										R	175.119			88.0631			1

**Supplemental Table 6.** Fragment ion series of mass spectra from formylated Lys893 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	86.0964			114.0913			96.0808			I							7
2	199.1805	199.183	0.0025	227.1754	227.1735	-0.0019	209.1648			L	814.3981	814.3996	0.0015	407.7027			6
3	328.2231			356.218			338.2074			E	701.3141	701.3176	0.0035	351.1607			5
4	475.2915			503.2864			485.2758			F	572.2715	572.265	-0.0064	286.6394			4
5	638.3548			666.3497			648.3392			Y	425.2031	425.2002	-0.0029	213.1052			3
6	725.3868			753.3818			735.3712	735.3912	0.02	S	262.1397	262.14	0.0003	131.5735			2
7										K	175.1077			88.0575			1



**Supplemental Table 7.** Fragment ion series of mass spectra from O-GlcNAcylated Thr346 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	88.0393			116.0342			98.0237			D							17
2	187.1077			215.1026			197.0921			V	1970.9474			985.9773			16
3	286.1761			314.171	314.1609	-0.0101	296.1605			V	1871.879			936.4431			15
4	433.2115			461.2064			443.1959			M	1772.8106			886.9089			14
5	737.3386			765.3335	765.3178	-0.0157	747.3229			T	1625.7752			813.3912			13
6	852.3655			880.3604			862.3499	862.3284	-0.0214	D	1321.6482			661.3277			12
7	939.3975			967.3925			949.3819			S	1206.6212	1206.6213	0.0001	603.8142			11
8	1068.4401			1096.435			1078.4245			E	1119.5892			560.2982			10
9	1183.4671			1211.462			1193.4514			D	990.5466	990.5421	-0.0045	495.7769			9
10	1270.4991			1298.494			1280.4835			S	875.5197			438.2635			8
11	1371.5468			1399.5417			1381.5311			T	788.4876	788.4909	0.0033	394.7474			7
12	1470.6152			1498.6101			1480.5995			V	687.4399			344.2236	344.237	0.0134	6
13	1557.6472			1585.6421			1567.6316			S	588.3715	588.3682	-0.0033	294.6894			5
14	1670.7313			1698.7262			1680.7156			I	501.3395	501.3279	-0.0116	251.1734			4
15	1798.7899			1826.7848			1808.7742			Q	388.2554	388.2561	0.0006	194.6314			3
16	1911.8739			1939.8688			1921.8583			I	260.1969	260.1967	-0.0002	130.6021			2
17										K	147.1128			74.06			1

**Supplemental Table 8.** Fragment ion series of mass spectra from O-GlcNAcylated Ser354 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b-18 calc.	b-18 obs.	b-18 delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	88.0393			116.0342			98.0237			D							17
2	187.1077			215.1026			197.0921			V	1970.9474			985.9773			16
3	286.1761			314.171			296.1605	296.1571	-0.0034	V	1871.879			936.4431			15
4	433.2115			461.2064	461.1902	-0.0162	443.1959			M	1772.8106			886.9089			14
5	534.2592			562.2541	562.2681	0.014	544.2436			T	1625.7752			813.3912			13
6	649.2861			677.2811	677.2795	-0.0016	659.2705			D	1524.7275			762.8674			12
7	736.3182			764.3131			746.3025			S	1409.7006			705.3539			11
8	865.3608			893.3557	893.3476	-0.0081	875.3451			E	1322.6686			661.8379			10
9	980.3877			1008.3826			990.3721			D	1193.626			597.3166			9
10	1067.4197			1095.4146			1077.4041			S	1078.599			539.8031			8
11	1168.4674			1196.4623			1178.4518			T	991.567	991.5449	-0.0221	496.2871			7
12	1267.5358			1295.5307			1277.5202	1277.5149	-0.0052	V	890.5193			445.7633			6
13	1557.6472			1585.6421			1567.6316			S	791.4509			396.2291			5
14	1670.7313			1698.7262			1680.7156			I	501.3395			251.1734			4
15	1798.7899			1826.7848			1808.7742			Q	388.2554	388.2532	-0.0022	194.6314			3
16	1911.8739			1939.8688			1921.8583			I	260.1969			130.6021			2
17										K	147.1128			74.06			1

**Supplemental Table 9.** Fragment ion series of mass spectra from N-ubiquitinated Lys497 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b++ calc.	b++ obs.	b++ delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	104.0528			132.0478			66.5275			M							19
2	175.09			203.0849			102.0461			A	2135.0172			1068.0122			18
3	304.1325			332.1275			166.5674			E	2063.9801			1032.4937			17
4	433.1751			461.17			231.0887			E	1934.9375			967.9724			16
5	546.2592			574.2541			287.6307			L	1805.8949			903.4511	903.4276	-0.0235	15
6	788.3971			816.392			408.6996			K	1692.8109			846.9091			14
7	885.4498			913.4448			457.226			P	1450.673			725.8401			13
8	1016.4903			1044.4852			522.7463			M	1353.6202			677.3137	677.3246	0.0109	12
9	1131.5173			1159.5122			580.2597			D	1222.5797	1222.5854	0.0057	611.7935	611.7944	0.0009	11
10	1232.5649			1260.5599			630.7836			T	1107.5528	1107.5622	0.0094	554.28	554.2744	-0.0056	10
11	1347.5919			1375.5868			688.297			D	1006.5051	1006.5058	0.0007	503.7562	503.7531	-0.0031	9
12	1475.6868			1503.6818			752.3445			K	891.4782	891.4658	-0.0123	446.2427			8
13	1604.7294			1632.7243			816.8658			E	763.3832	763.3694	-0.0138	382.1952			7
14	1691.7615			1719.7564			860.3818			S	634.3406	634.3441	0.0034	317.6739			6
15	1804.8455			1832.8404			916.9239			I	547.3086	547.3073	-0.0013	274.1579			5
16	1875.8826			1903.8775			952.4424			A	434.2245	434.2194	-0.0052	217.6159			4
17	2004.9252			2032.9201			1016.9637			E	363.1874	363.183	-0.0045	182.0973			3
18	2091.9573			2119.9522			1060.4797			S	234.1448	234.1478	0.003	117.5761			2
19										K	147.1128			74.06			1

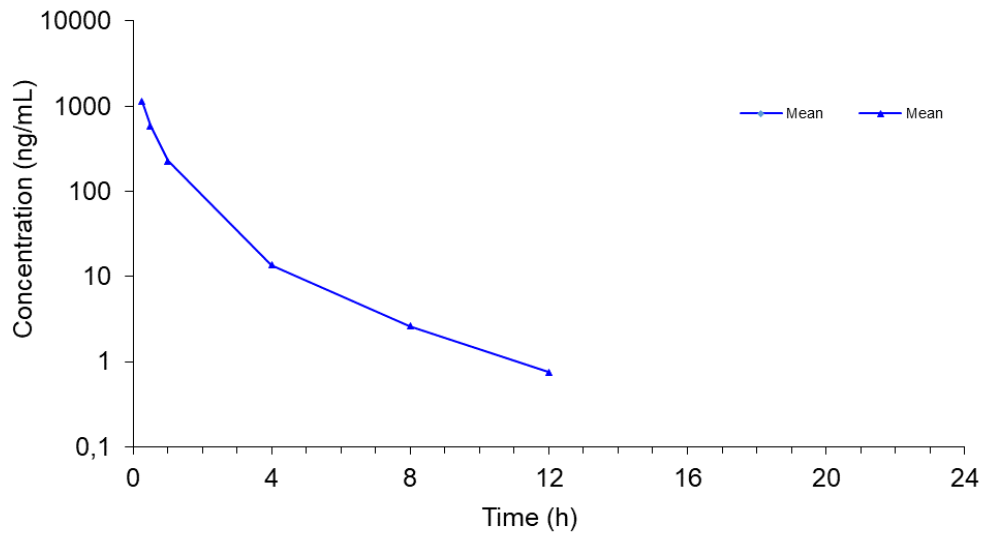
Supplemental Table 10. Fragment ion series of mass spectra from O-ubiquitinated Ser515 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b++ calc.	b++ obs.	b++ delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	60.0444			88.0393			44.5233			S							25
2	157.0972			185.0921			93.0497			P	2928.2002			1464.6037			24
3	286.1397	286.1367	-0.0031	314.1347	314.1301	-0.0046	157.571			E	2831.1474			1416.0773			23
4	417.1802	417.1868	0.0066	445.1751			223.0912			M	2702.1048			1351.556			22
5	618.2552			646.2501	646.2548	0.0047	323.6287			S	2571.0643			1286.0358			21
6	749.2957			777.2906			389.1489			M	2369.9894			1185.4983			20
7	877.3542			905.3492	905.3553	0.0062	453.1782			Q	2238.9489			1119.9781			19
8	1006.3968			1034.3917			517.6995			E	2110.8903			1055.9488			18
9	1121.4238			1149.4187			575.213			D	1981.8477			991.4275			17
10	1281.4544			1309.4493			655.2283			C	1866.8208			933.914	933.9101	-0.004	16
11	1394.5385			1422.5334			711.7703			I	1706.7901			853.8987	853.9253	0.0265	15
12	1481.5705			1509.5654			755.2863	755.3037	0.0173	S	1593.7061			797.3567	797.3519	-0.0047	14
13	1596.5974			1624.5924			812.7998			D	1506.6741			753.8407			13
14	1709.6815			1737.6764			869.3418			I	1391.6471			696.3272			12
15	1780.7186			1808.7135			904.8604			A	1278.5631			639.7852	639.7825	-0.0026	11
16	1877.7714			1905.7663			953.3868			P	1207.5259	1207.5258	-0.0002	604.2666	604.2647	-0.0019	10
17	2024.8068			2052.8017			1026.9045			M	1110.4732	1110.4827	0.0095	555.7402			9
18	2152.8654			2180.8603			1090.9338			Q	963.4378	963.439	0.0013	482.2225	482.2222	-0.0003	8
19	2253.913			2281.9079			1141.4576			T	835.3792	835.3736	-0.0056	418.1932	418.1956	0.0024	7
20	2368.94			2396.9349			1198.9711			D	734.3315	734.3222	-0.0093	367.6694			6
21	2497.9826			2525.9775			1263.4924			E	619.3046	619.2951	-0.0095	310.1559			5
22	2626.0411			2654.0361			1327.5217			Q	490.262	490.2578	-0.0042	245.6346			4
23	2727.0888			2755.0837			1378.0455			T	362.2034	362.2053	0.0019	181.6053			3
24	2841.1318			2869.1267			1435.067			N	261.1557	261.1527	-0.0031	131.0815			2
25										K	147.1128			74.06			1

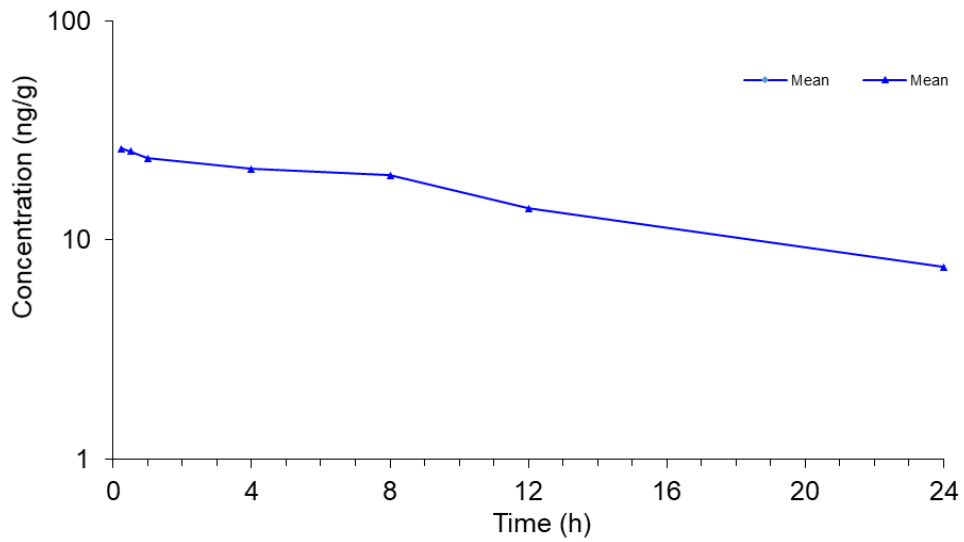
Supplemental Table 11. Fragment ion series of mass spectra from O-ubiquitinated Ser522 peptide.

#	a calc.	a obs.	a delta	b calc.	b obs.	b delta	b++ calc.	b++ obs.	b++ delta	Seq.	y calc.	y obs.	y delta	y++ calc.	y++ obs.	y++ delta	#
1	60.0444			88.0393			44.5233			S							25
2	157.0972			185.0921			93.0497			P	2928.2002			1464.6037			24
3	286.1397			314.1347			157.571			E	2831.1474			1416.0773			23
4	417.1802			445.1751			223.0912			M	2702.1048			1351.556			22
5	504.2123			532.2072			266.6072			S	2571.0643			1286.0358			21
6	635.2527			663.2476			332.1275			M	2484.0323			1242.5198			20
7	763.3113			791.3062	791.3231	0.0169	396.1568			Q	2352.9918			1176.9995			19
8	892.3539			920.3488			460.678			E	2224.9332			1112.9703			18
9	1007.3808			1035.3758			518.1915			D	2095.8907			1048.449			17
10	1167.4115			1195.4064			598.2068			C	1980.8637			990.9355			16
11	1280.4955			1308.4905			654.7489			I	1820.8331			910.9202			15
12	1481.5705			1509.5654			755.2863			S	1707.749			854.3781			14
13	1596.5974			1624.5924			812.7998			D	1506.6741			753.8407			13
14	1709.6815			1737.6764			869.3418			I	1391.6471			696.3272			12
15	1780.7186			1808.7135			904.8604			A	1278.5631	1278.5622	-0.0008	639.7852			11
16	1877.7714			1905.7663			953.3868			P	1207.5259	1207.5212	-0.0047	604.2666	604.2566	-0.01	10
17	2024.8068			2052.8017			1026.9045			M	1110.4732			555.7402			9
18	2152.8654			2180.8603			1090.9338			Q	963.4378	963.4367	-0.0011	482.2225			8
19	2253.913			2281.9079			1141.4576			T	835.3792	835.3679	-0.0113	418.1932			7
20	2368.94			2396.9349			1198.9711			D	734.3315	734.3225	-0.009	367.6694			6
21	2497.9826			2525.9775			1263.4924			E	619.3046	619.2927	-0.0119	310.1559			5
22	2626.0411			2654.0361			1327.5217			Q	490.262	490.2635	0.0015	245.6346			4
23	2727.0888			2755.0837			1378.0455			T	362.2034			181.6053			3
24	2841.1318			2869.1267			1435.067			N	261.1557			131.0815			2
25										K	147.1128			74.06			1

### Brain permeability in vivo of compound 16

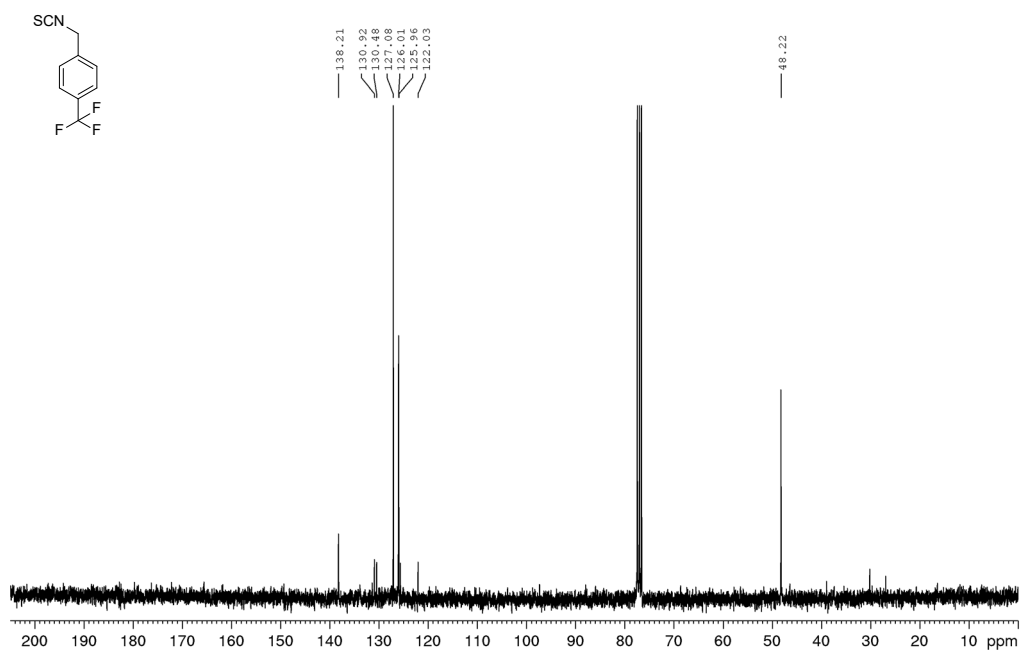
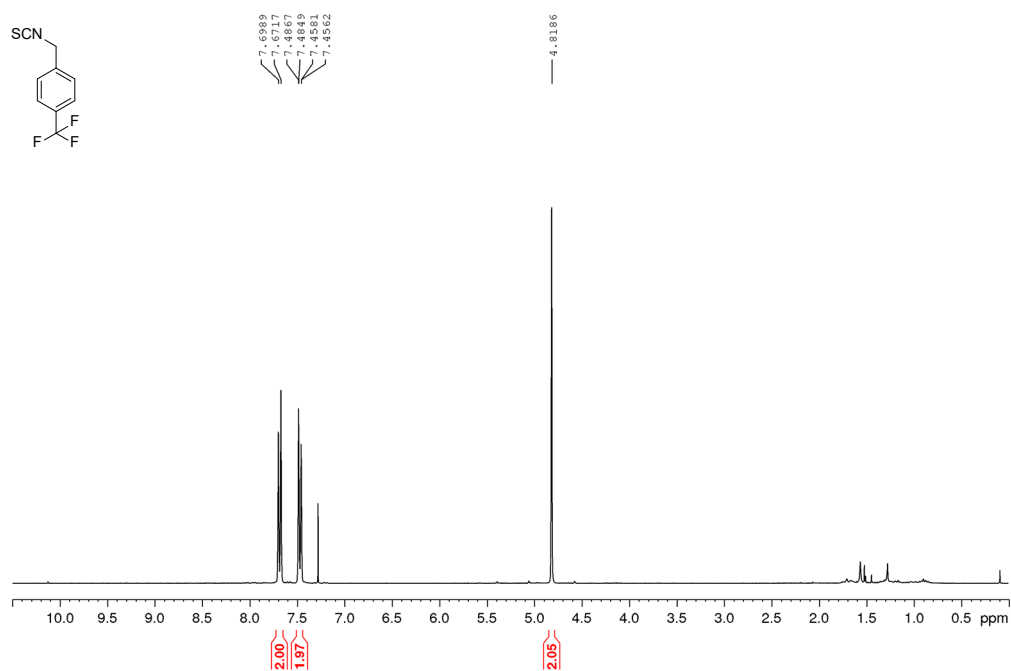


Supplemental Figure 24. Rat exposure in plasma from a single oral dose of 16 (3 mg/kg).

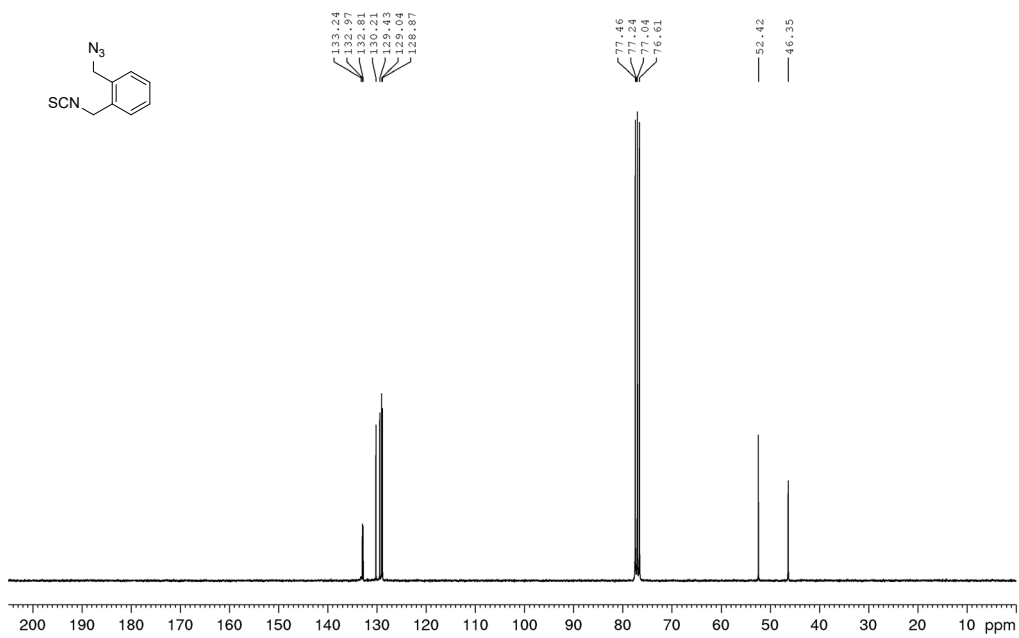
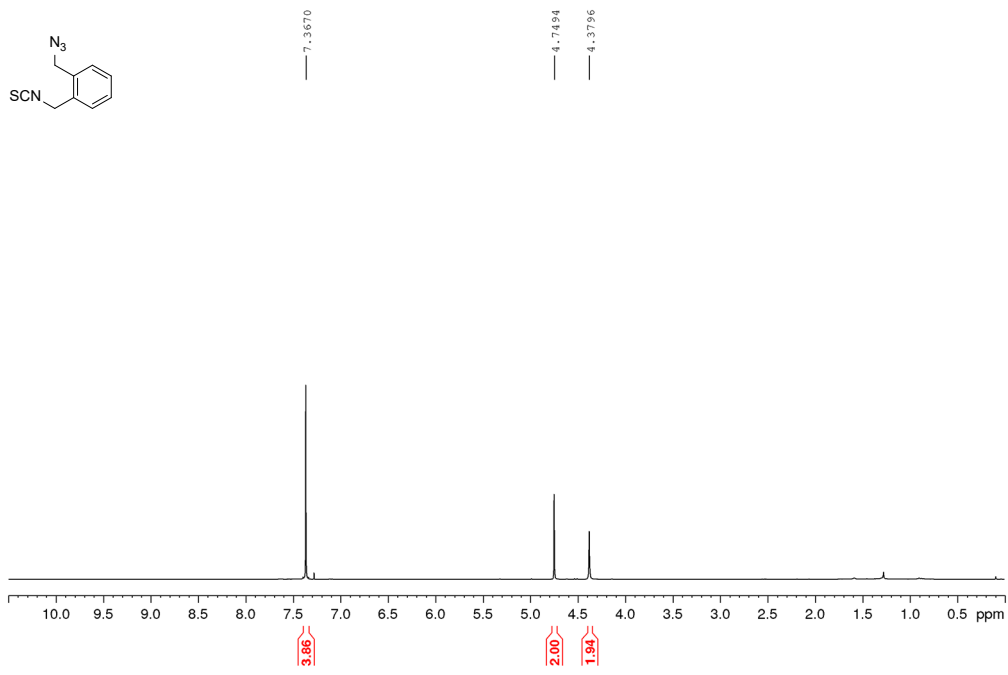


Supplemental Figure 25. Rat exposure in homogenized brain from a single oral dose of 16 (3 mg/kg).

# Copies of $^1\text{H}$ and $^{13}\text{C}$ NMR and MS spectra

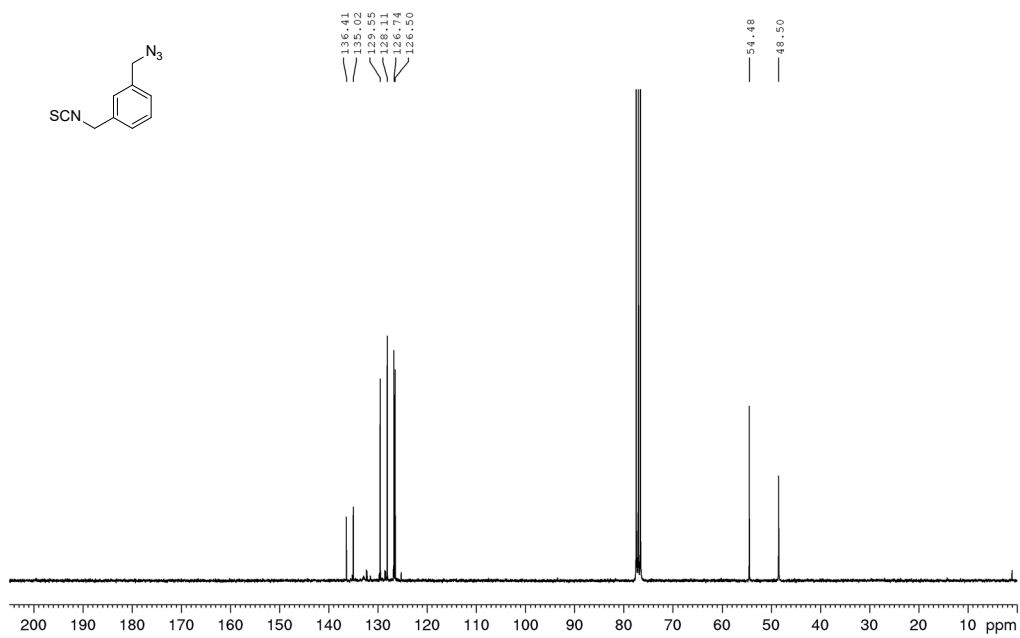
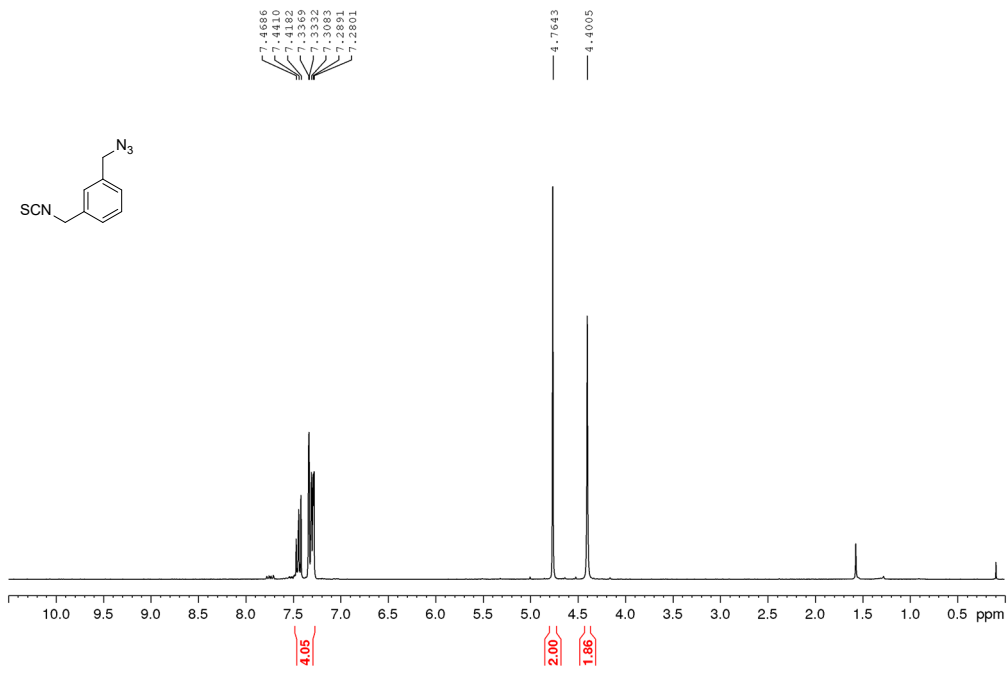


Supplemental Figure 26.  $^1\text{H}$  NMR and  $^{13}\text{C}$  NMR (300 MHz, 75.5 MHz  $\text{CDCl}_3$ ) of TFBNCS

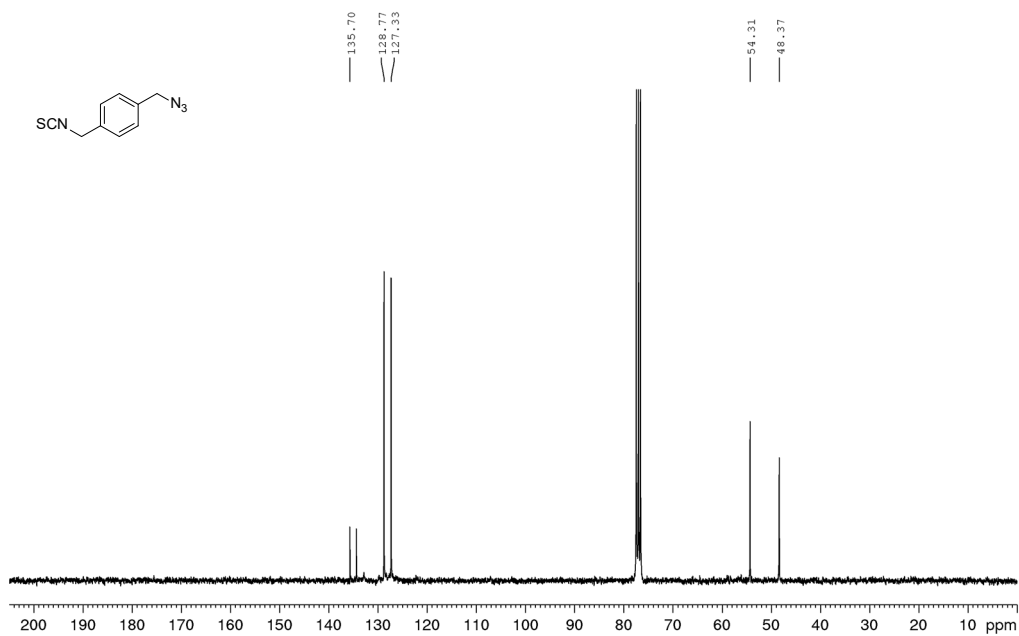
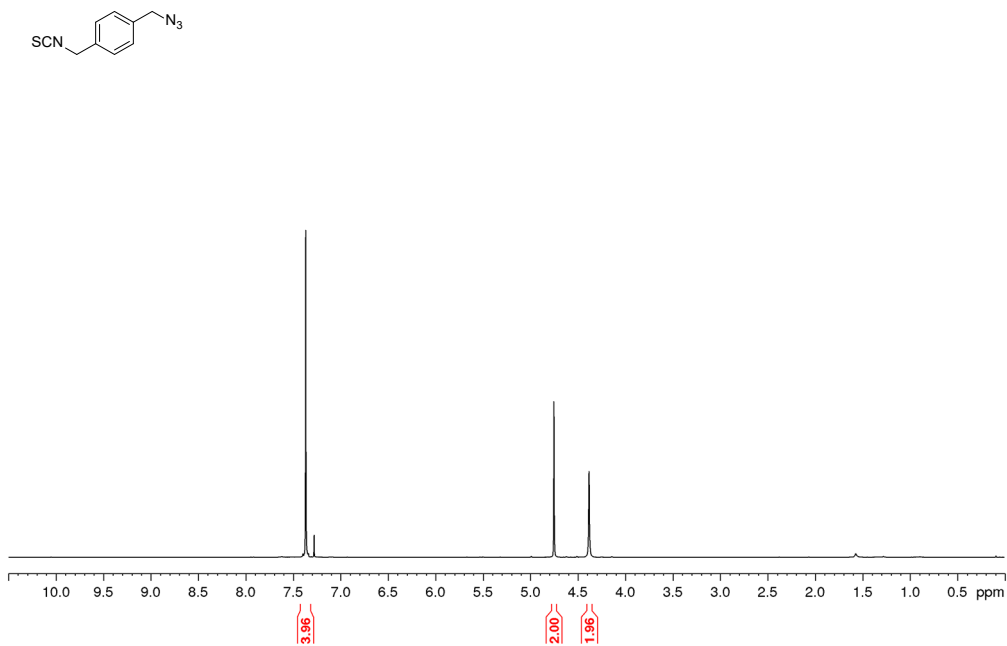


Supplemental Figure 27. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CDCl<sub>3</sub>) of *o*-ABNCS

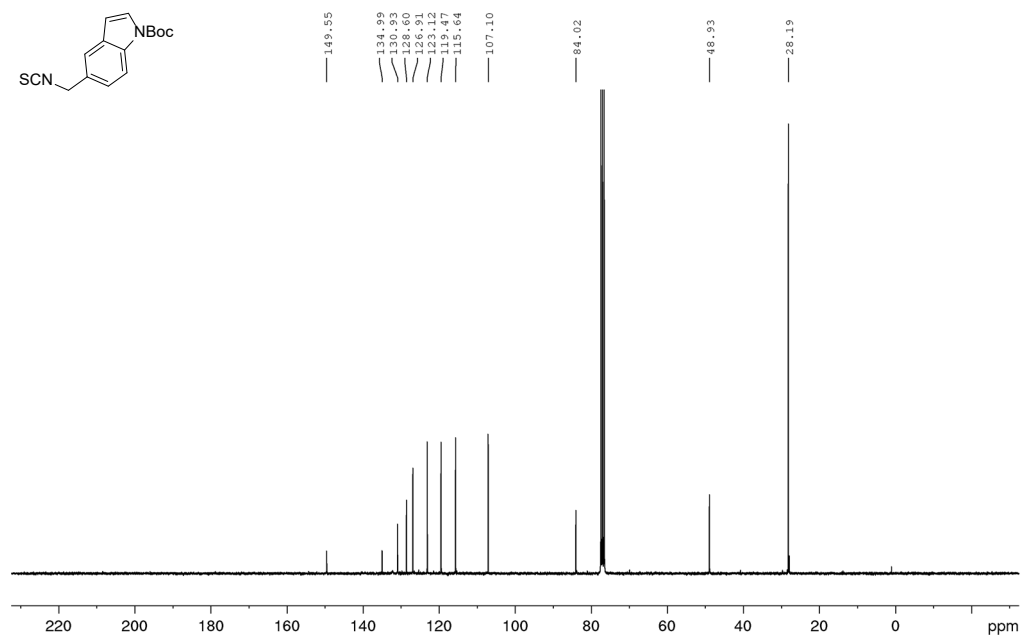
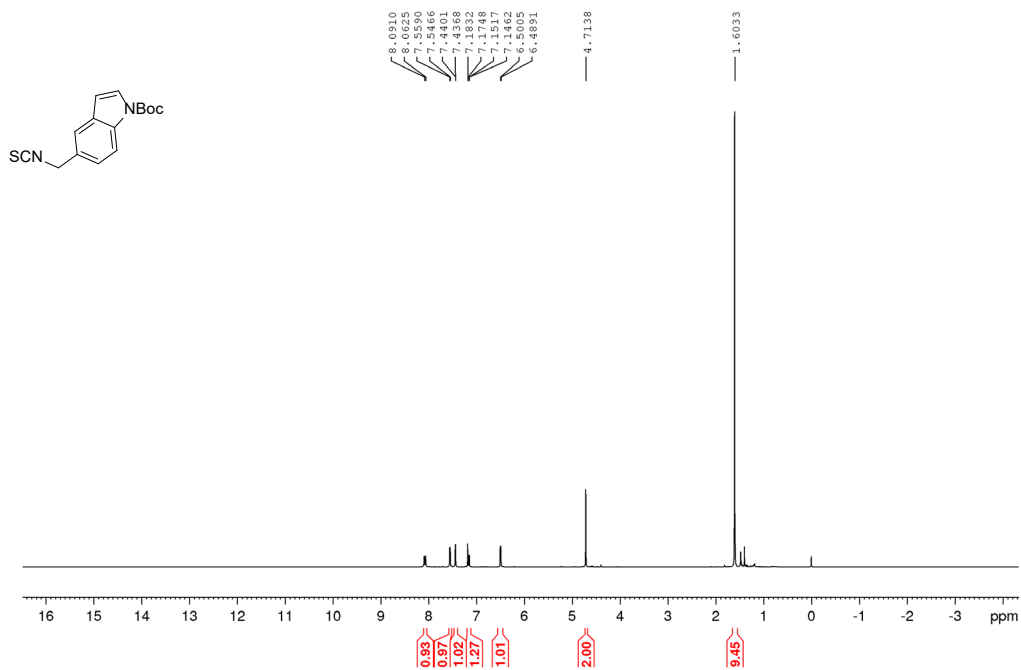




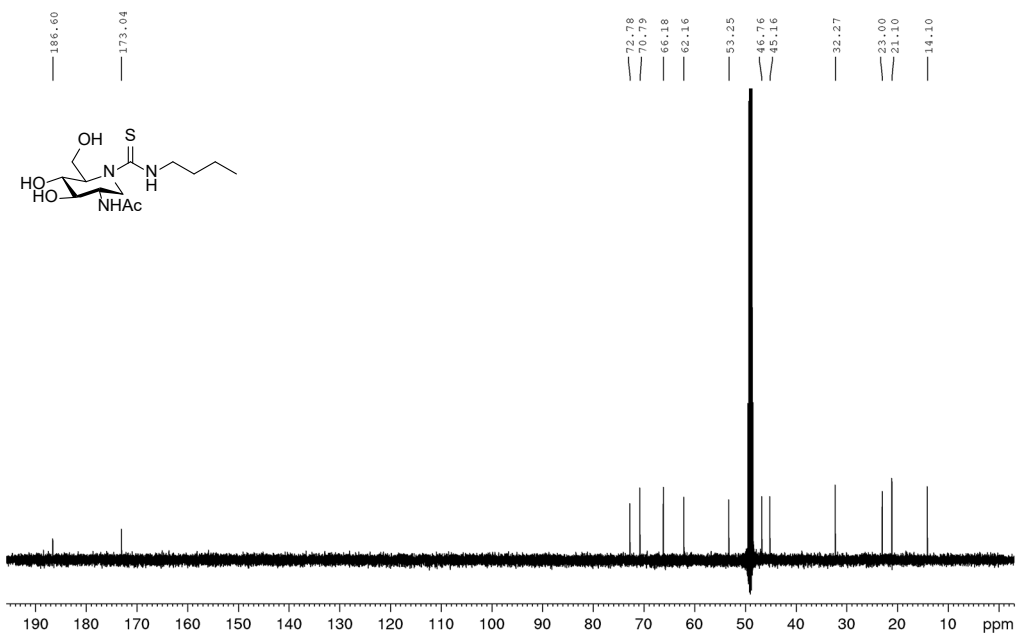
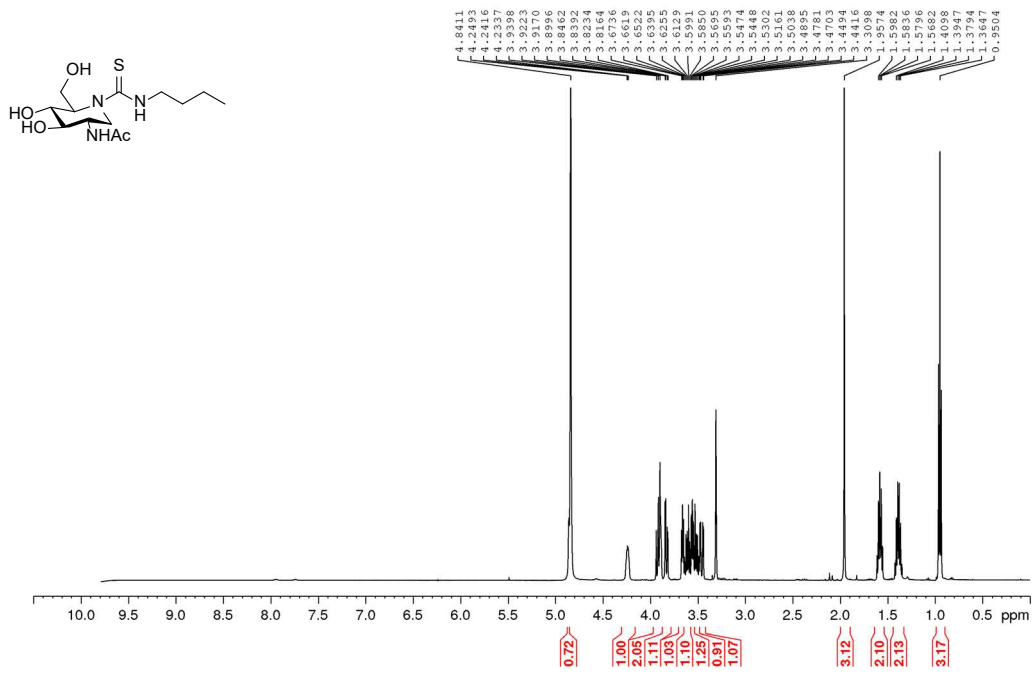
Supplemental Figure 28. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CDCl<sub>3</sub>) of *m*-ABNCS



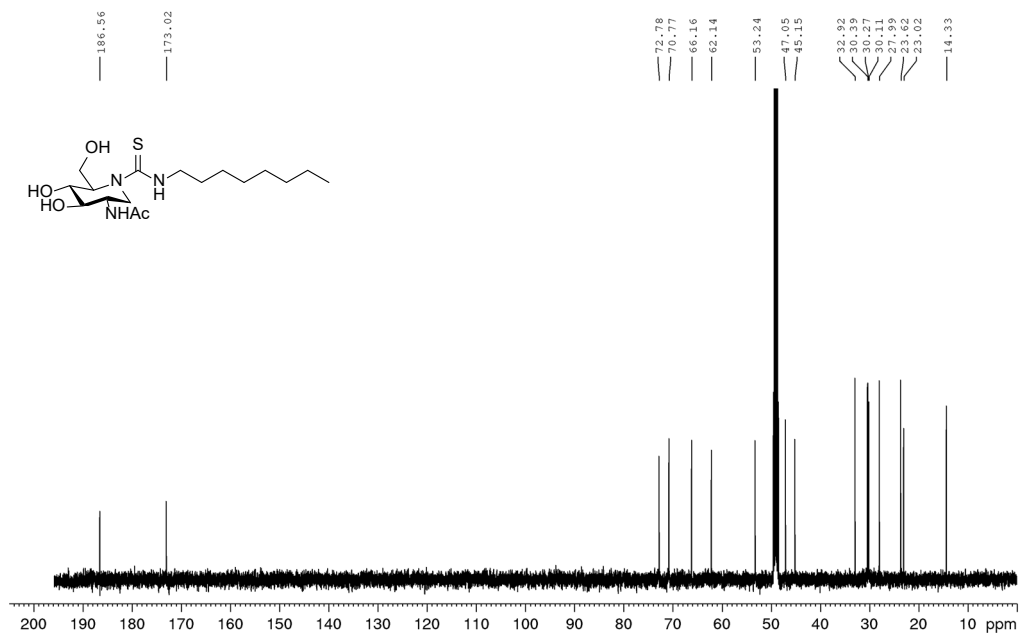
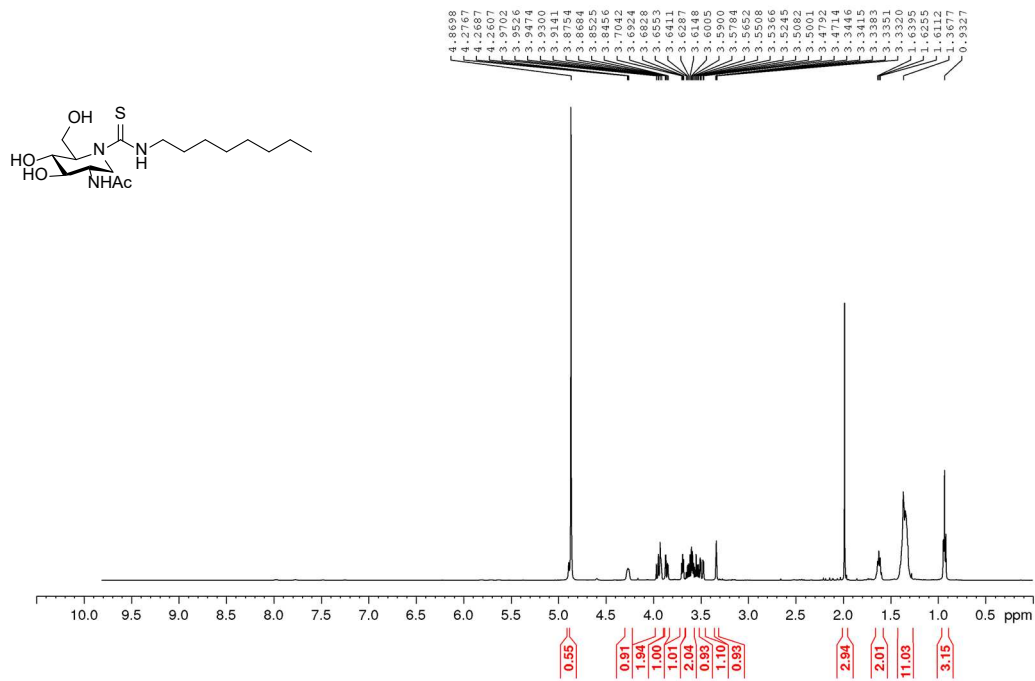
**Supplemental Figure 29.** <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CDCl<sub>3</sub>) of *p*-ABNCS



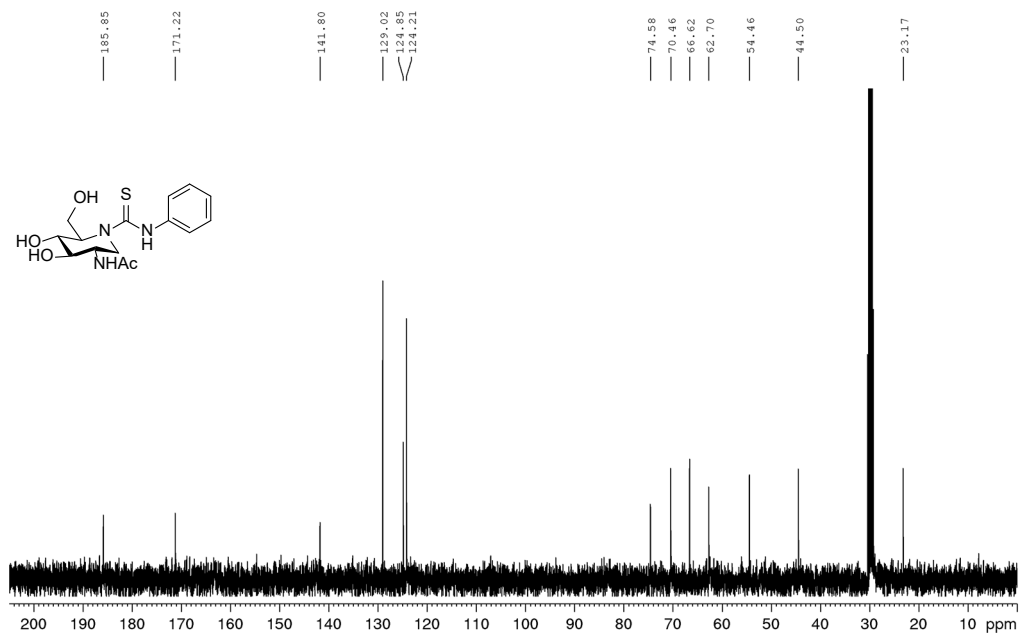
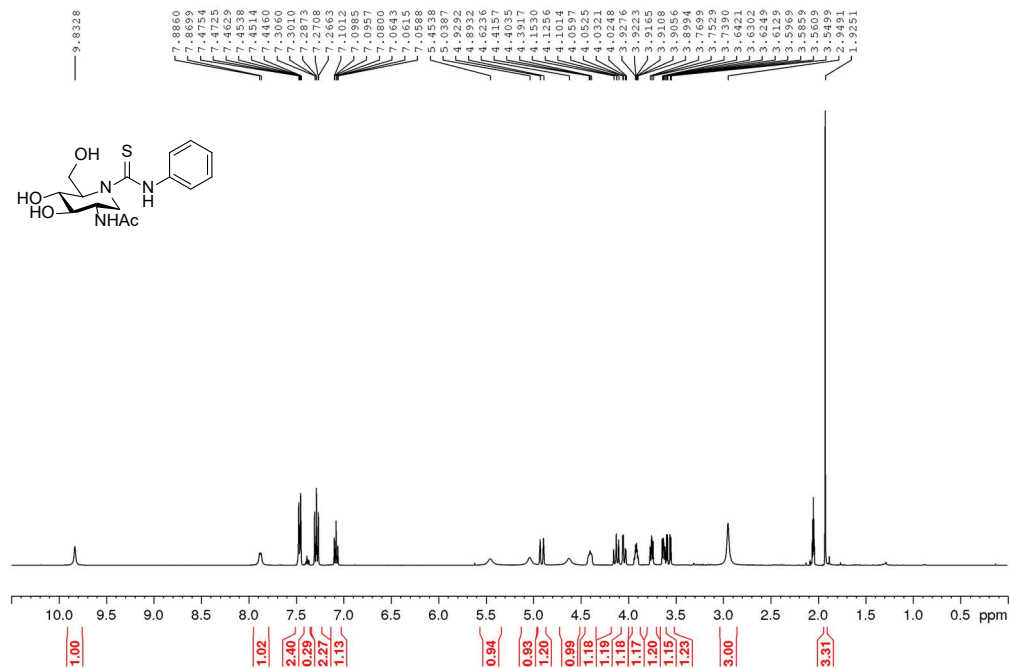
Supplemental Figure 30. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CD<sub>3</sub>OD) of IBLCS  
S58



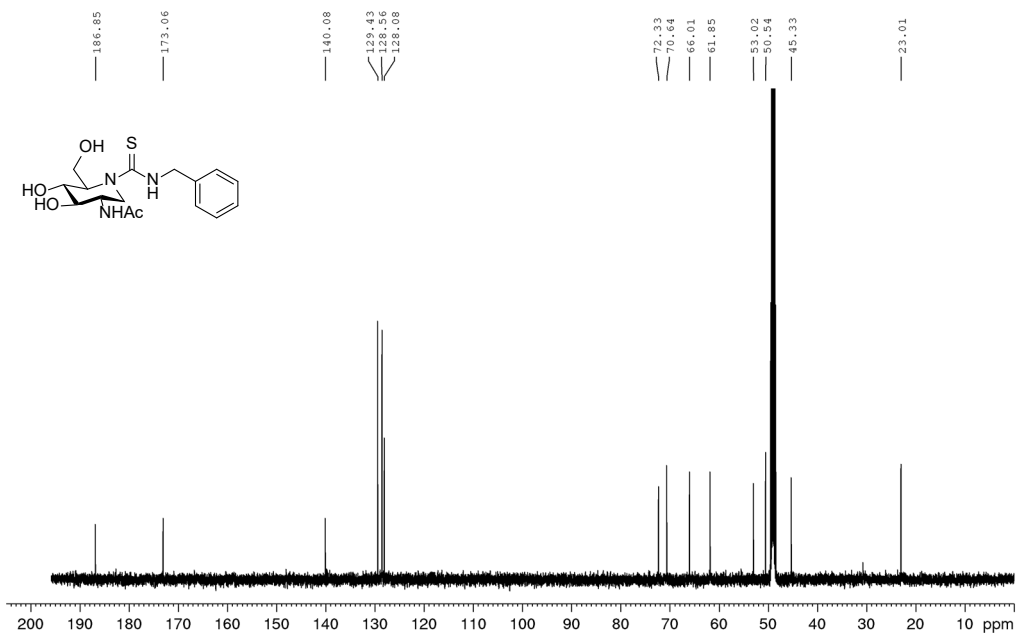
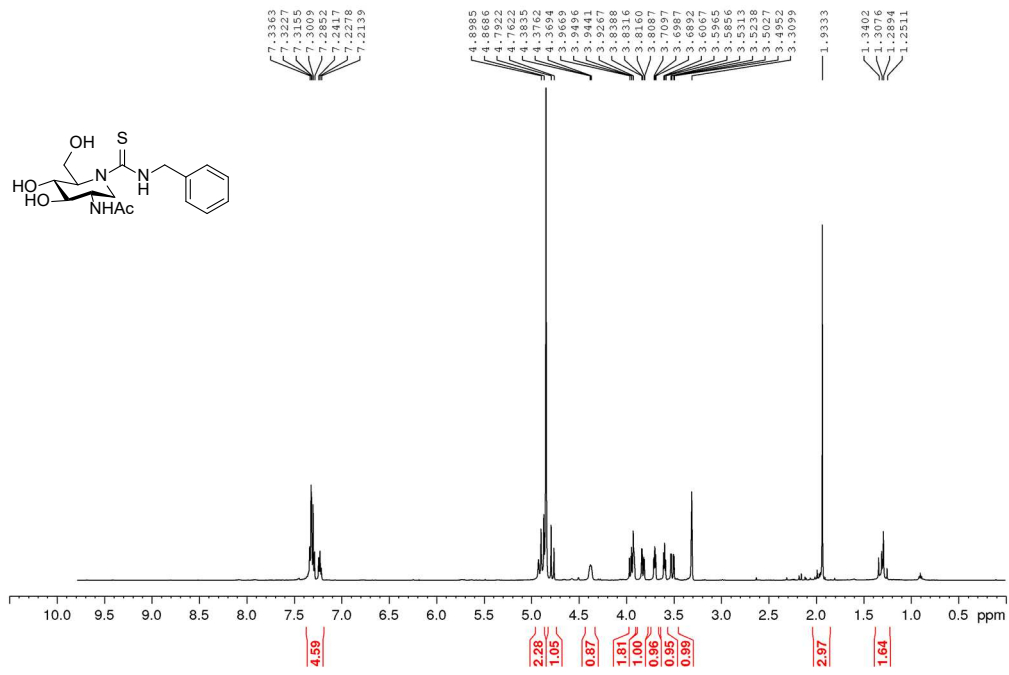
Supplemental Figure 31. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 1



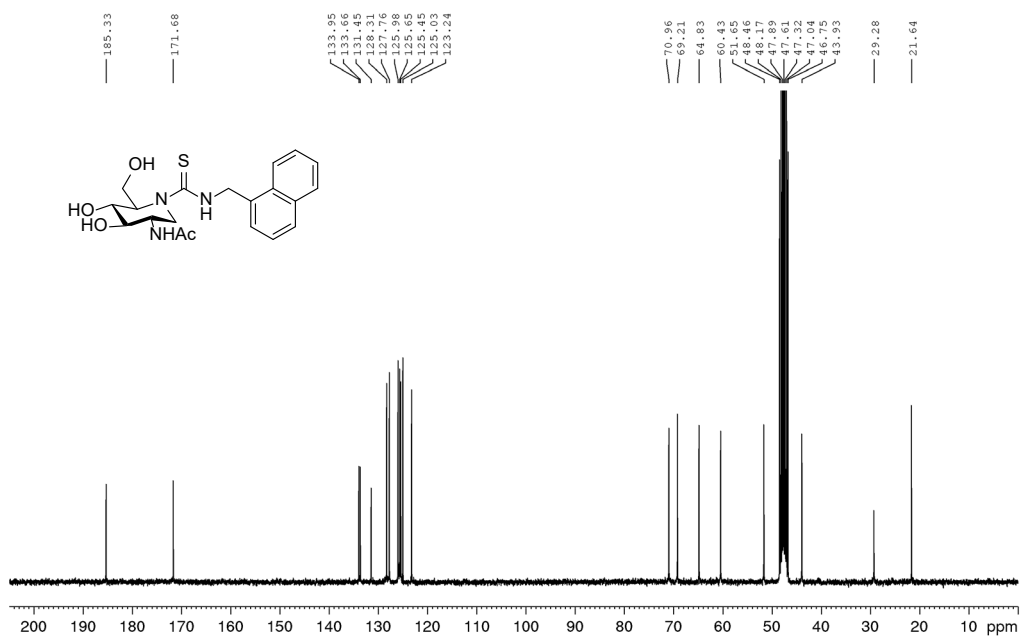
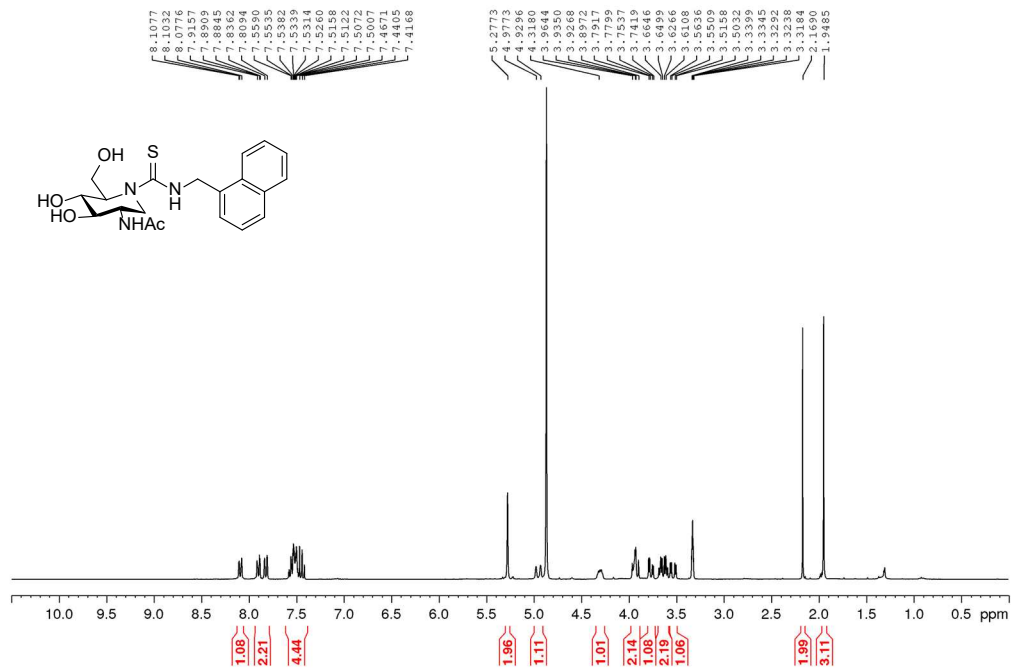
Supplemental Figure 32. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 2



Supplemental Figure 33. <sup>1</sup>H NMR and <sup>13</sup>C NMR (400 MHz, 100.6 MHz acetone-*d*<sub>6</sub>) of 3

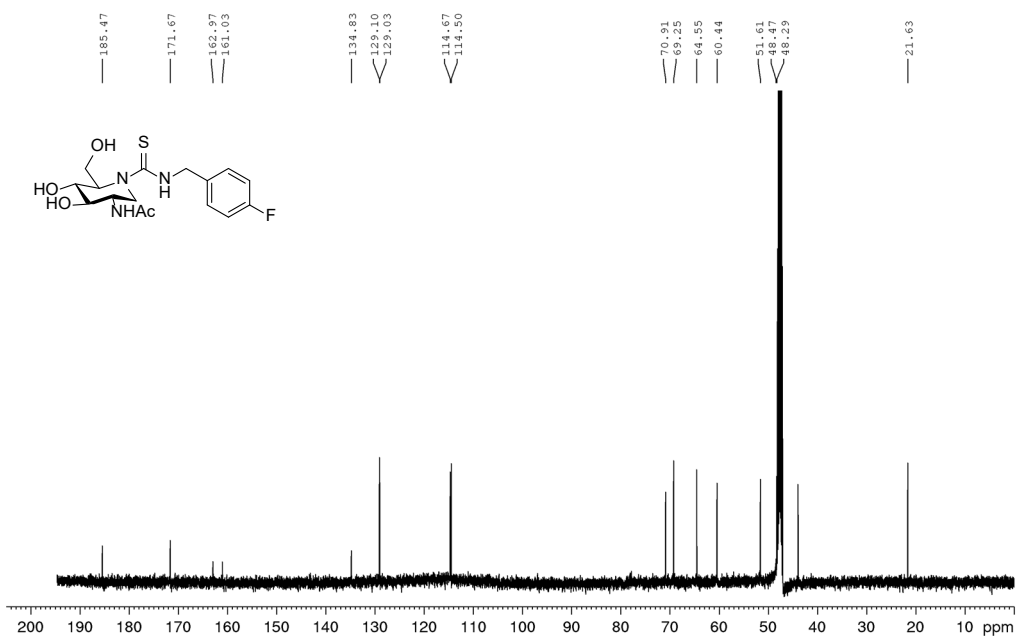
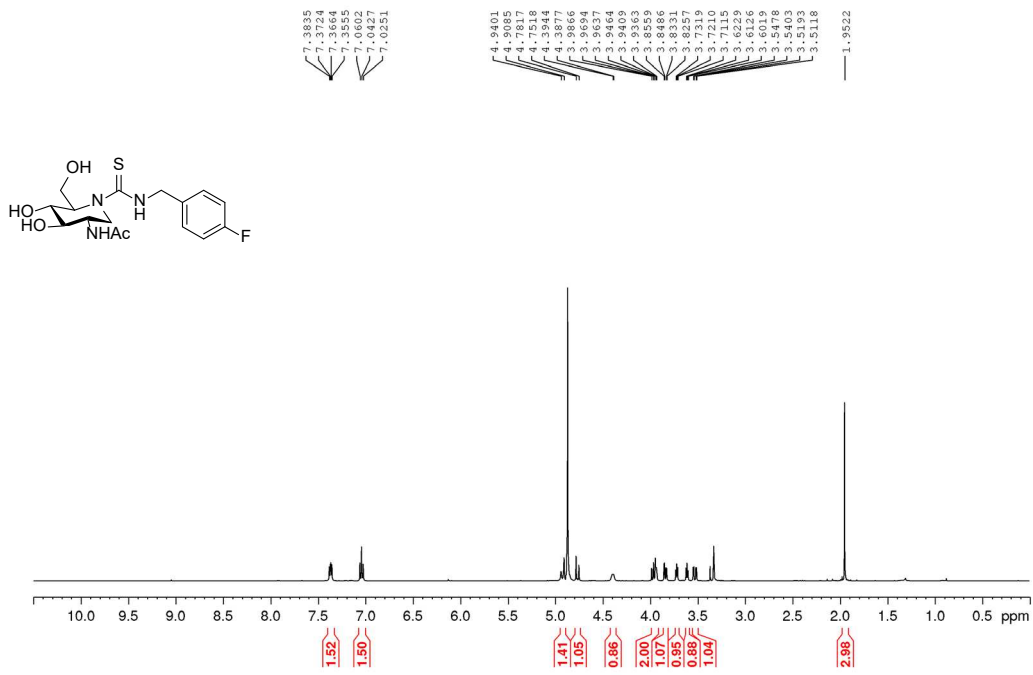


Supplemental Figure 34. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 4

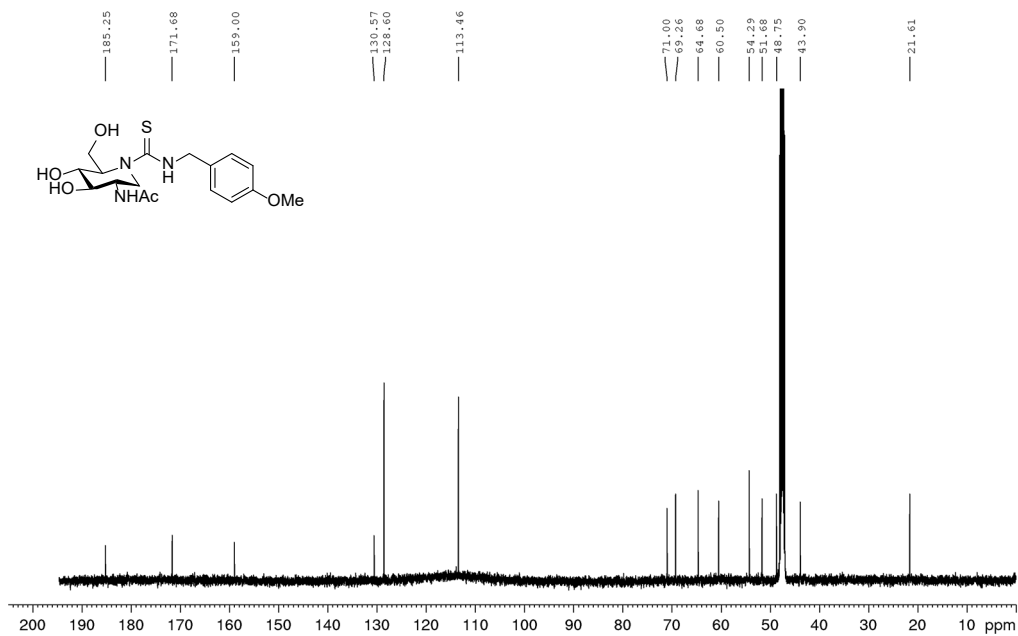
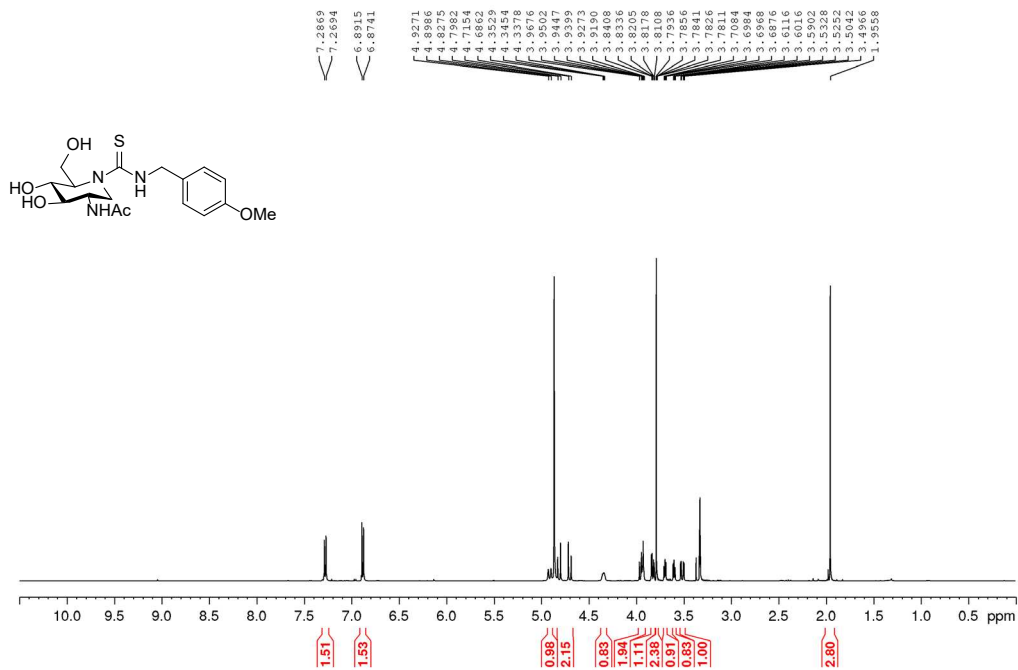


Supplemental Figure 35. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 100.6 MHz CD<sub>3</sub>OD) of 5

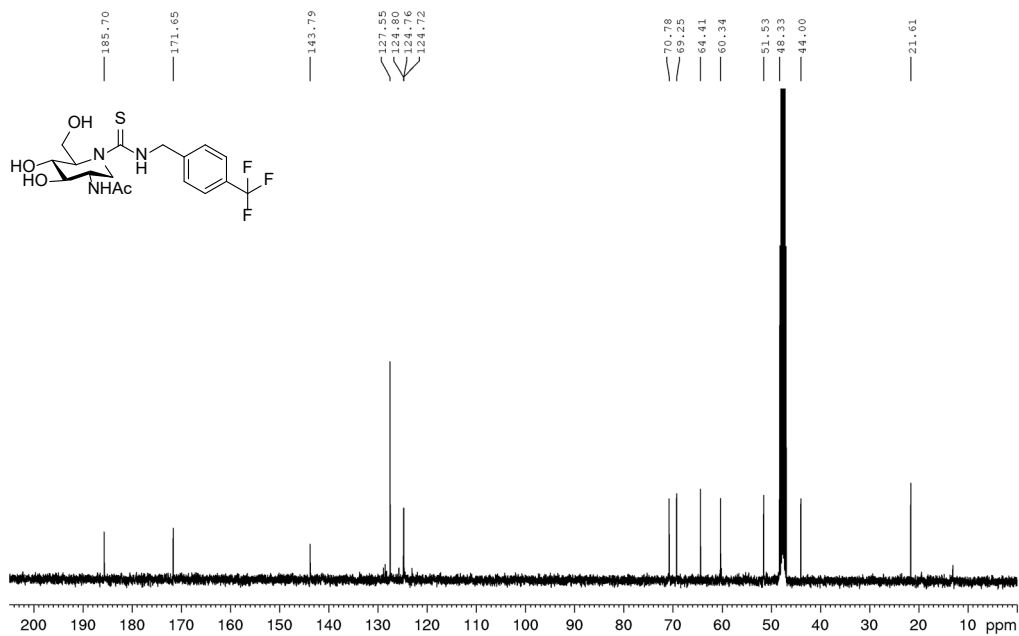
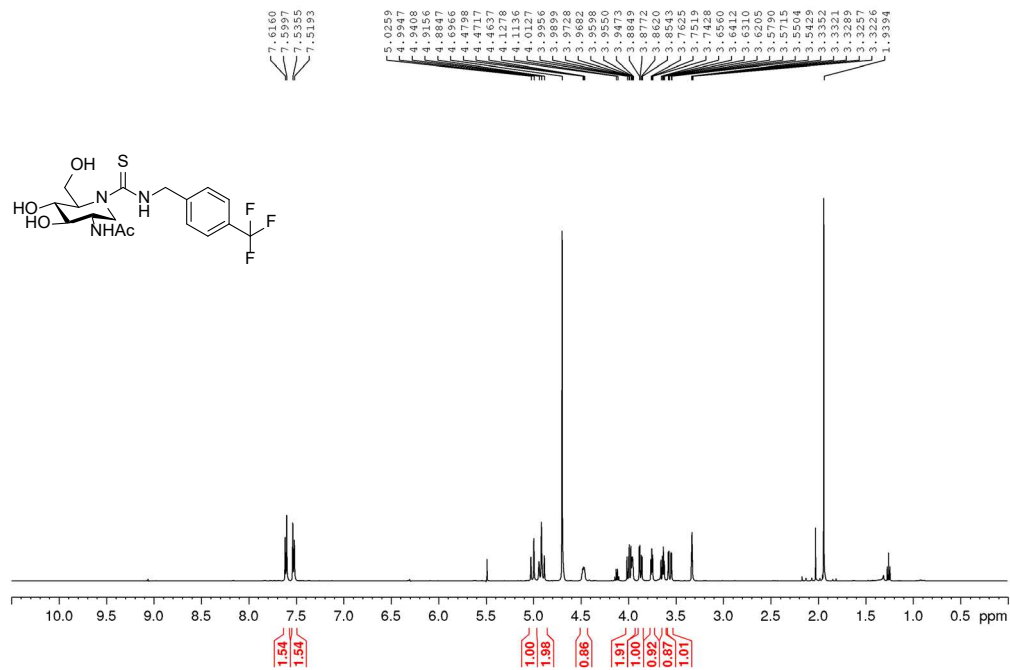




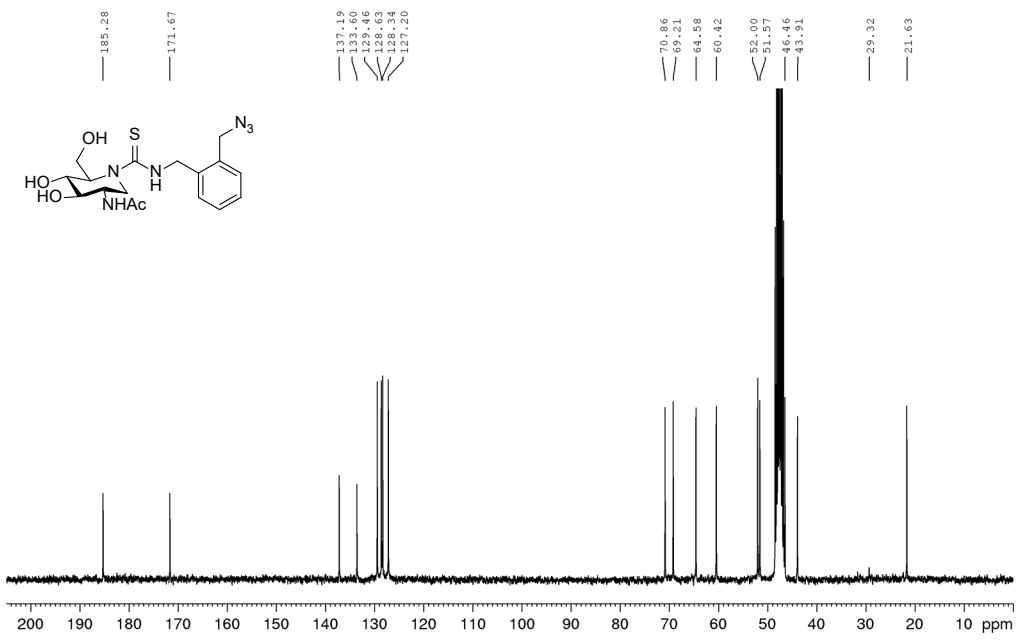
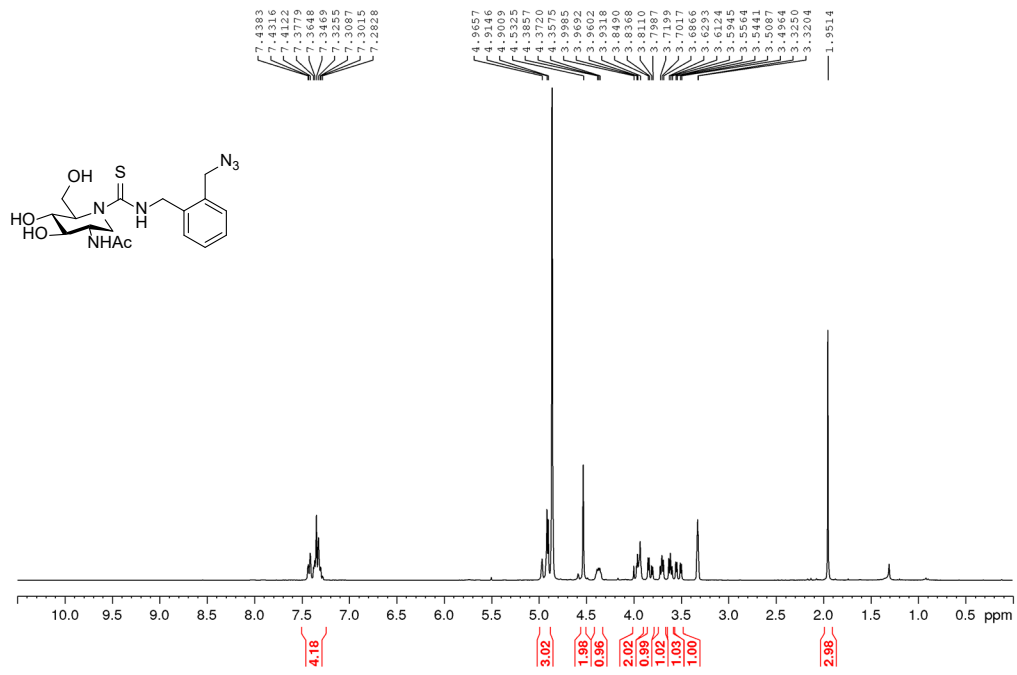
Supplemental Figure 36. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 6



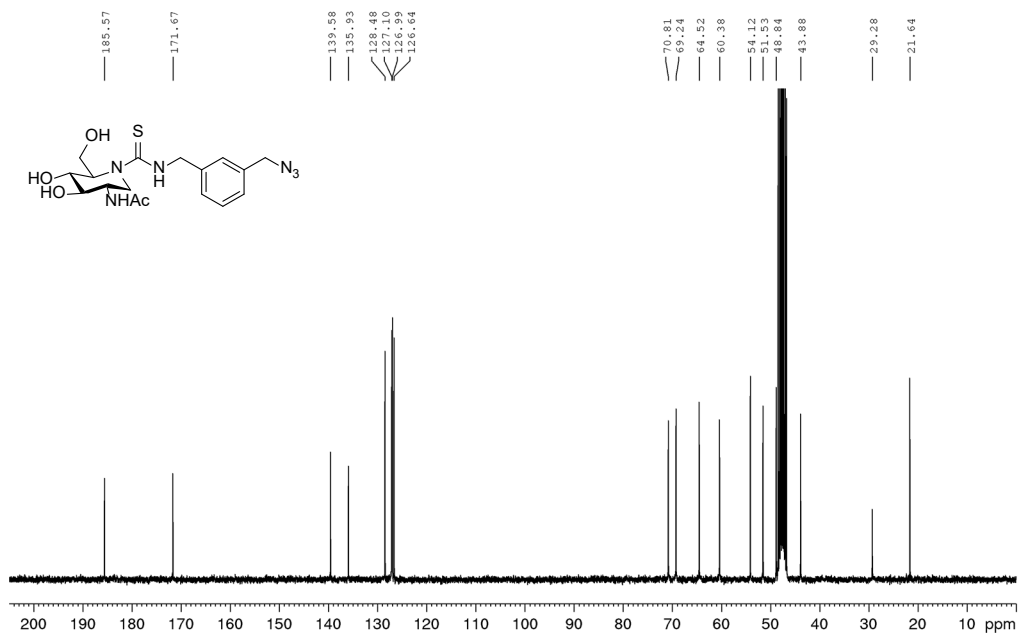
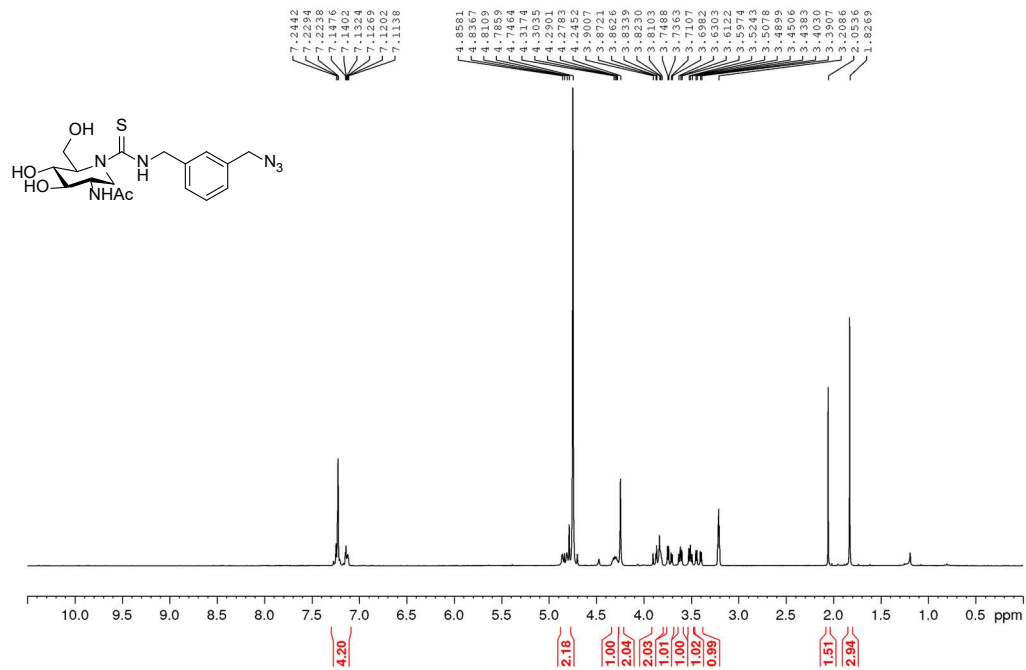
Supplemental Figure 37. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 7



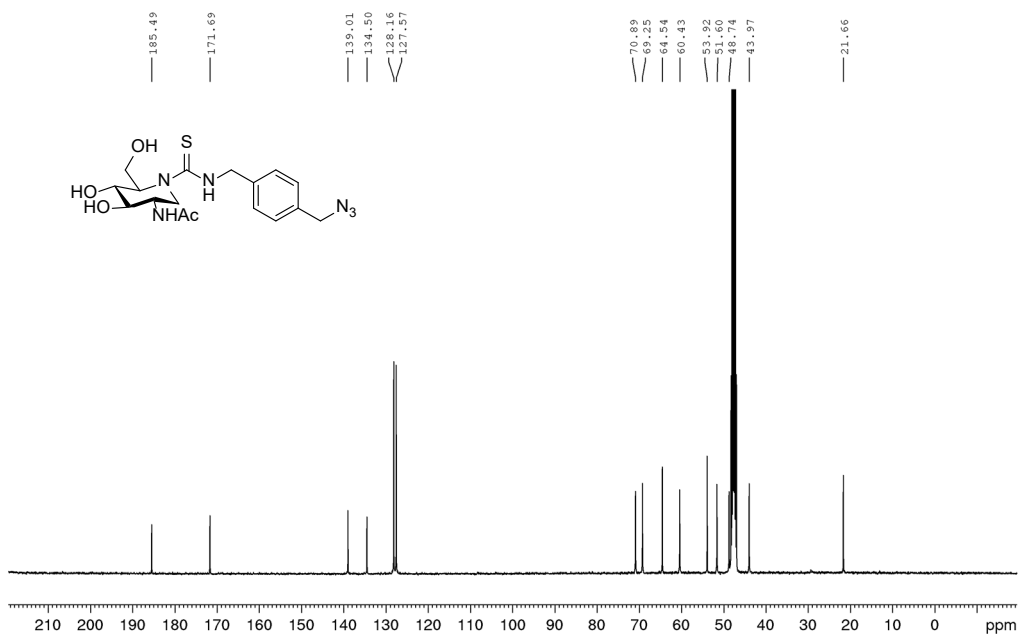
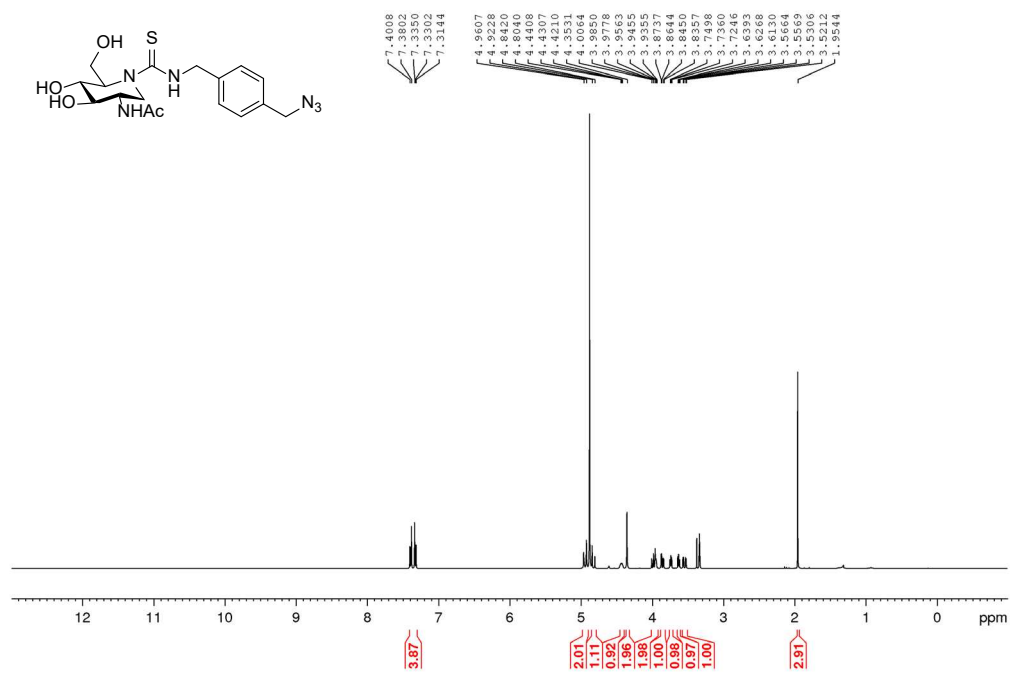
Supplemental Figure 38. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of **8**



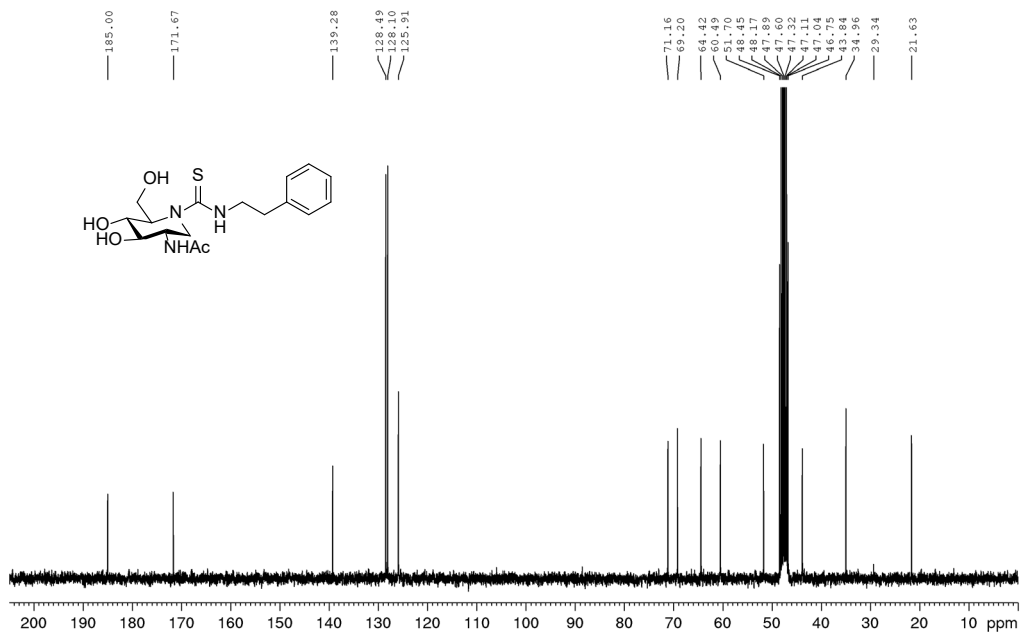
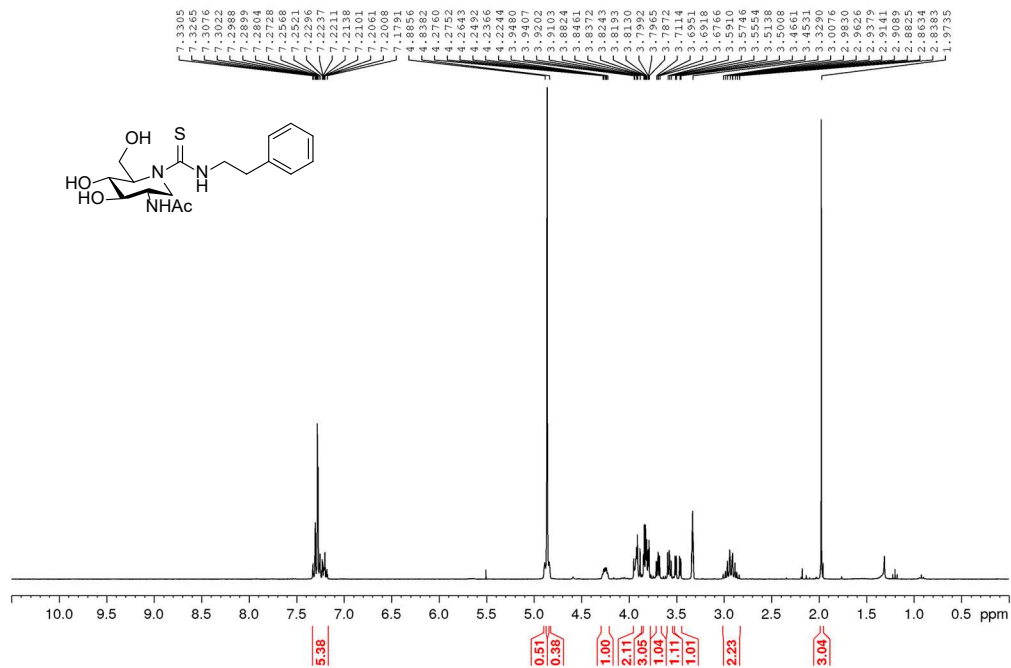
Supplemental Figure 39. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 100.6 MHz CD<sub>3</sub>OD) of 9



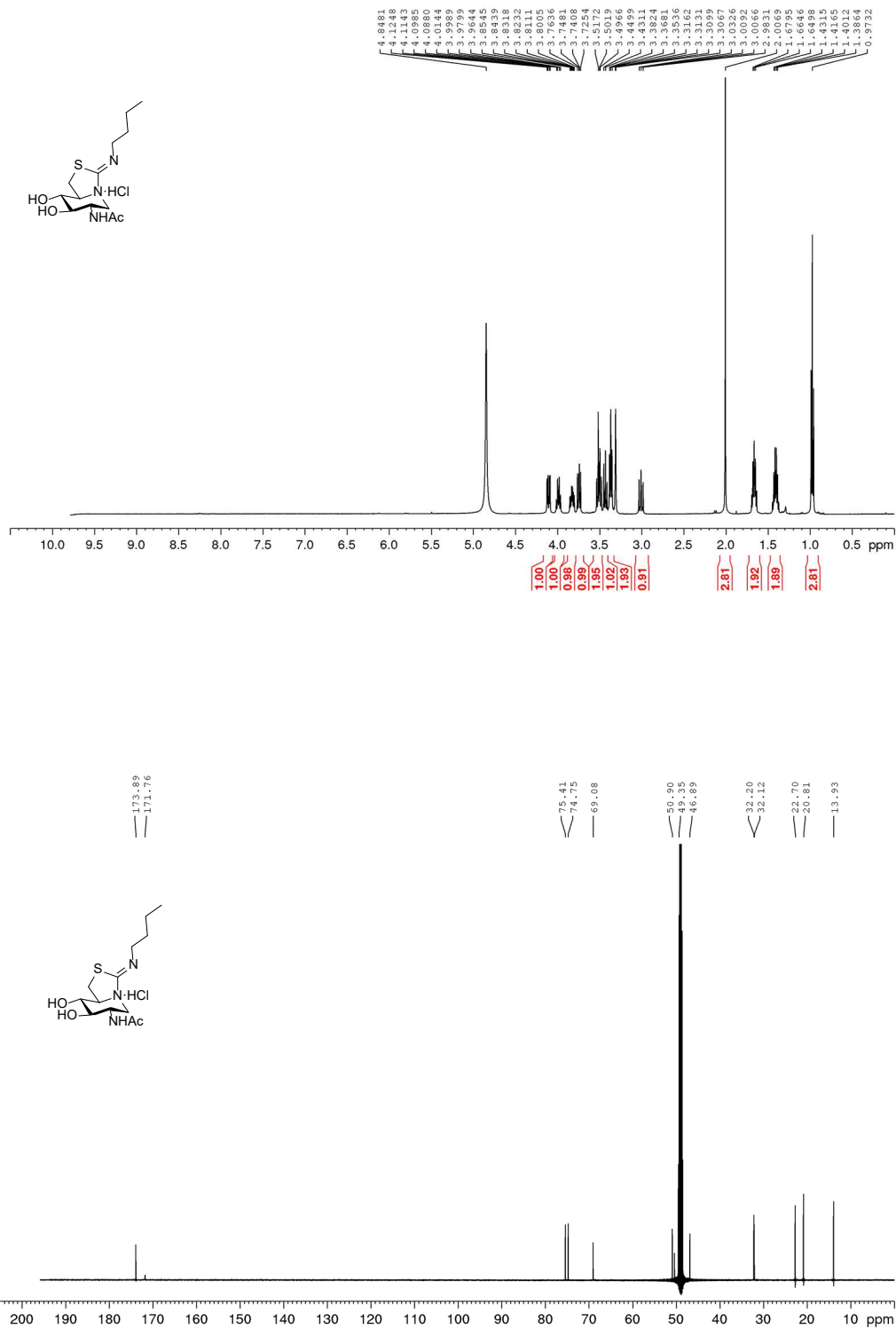
Supplemental Figure 40. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 100.6 MHz CD<sub>3</sub>OD) of 10



Supplemental Figure 41. <sup>1</sup>H NMR and <sup>13</sup>C NMR (400 MHz, 100.6 MHz CD<sub>3</sub>OD) of 11

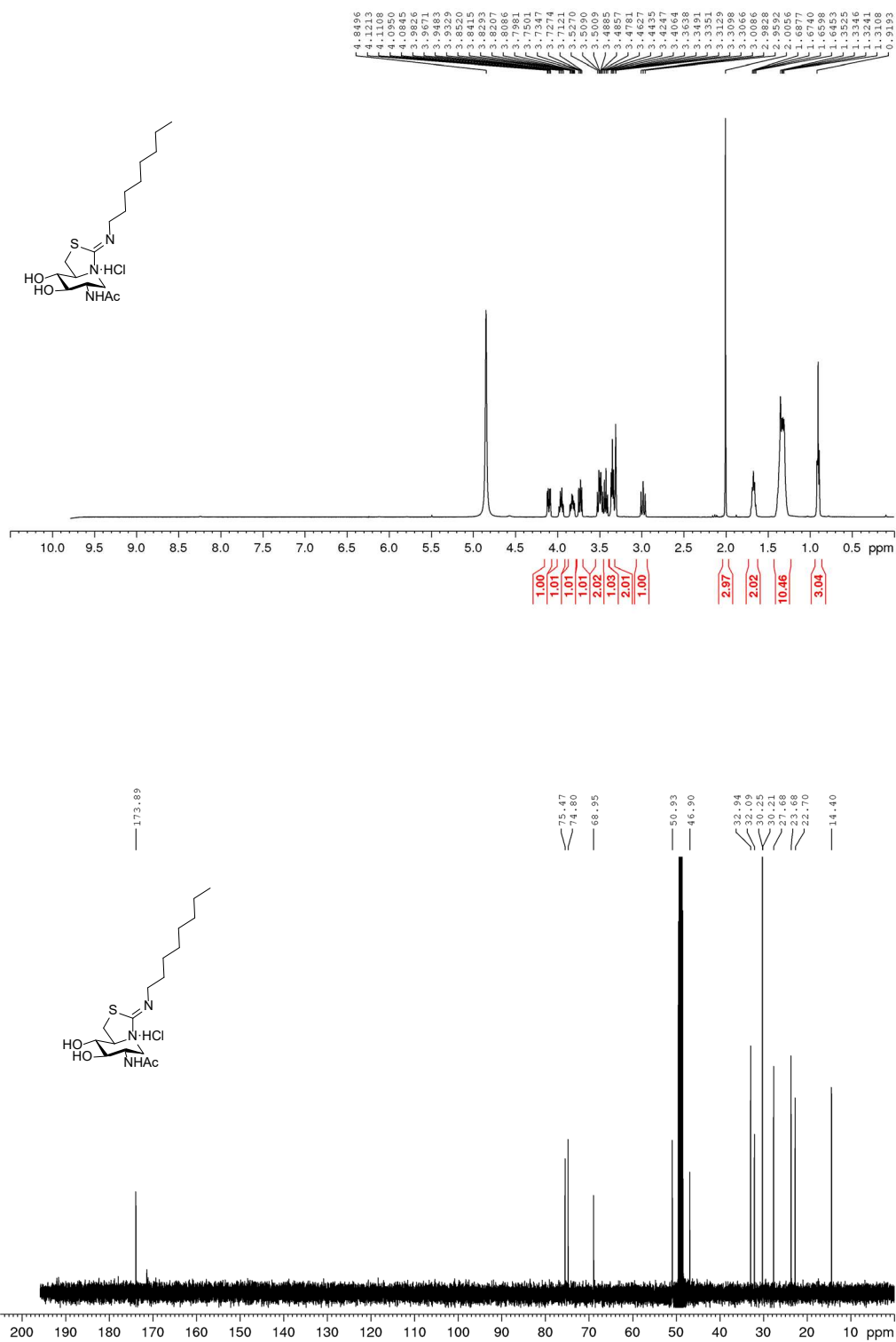


Supplemental Figure 42 <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CD<sub>3</sub>OD) of 12

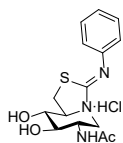
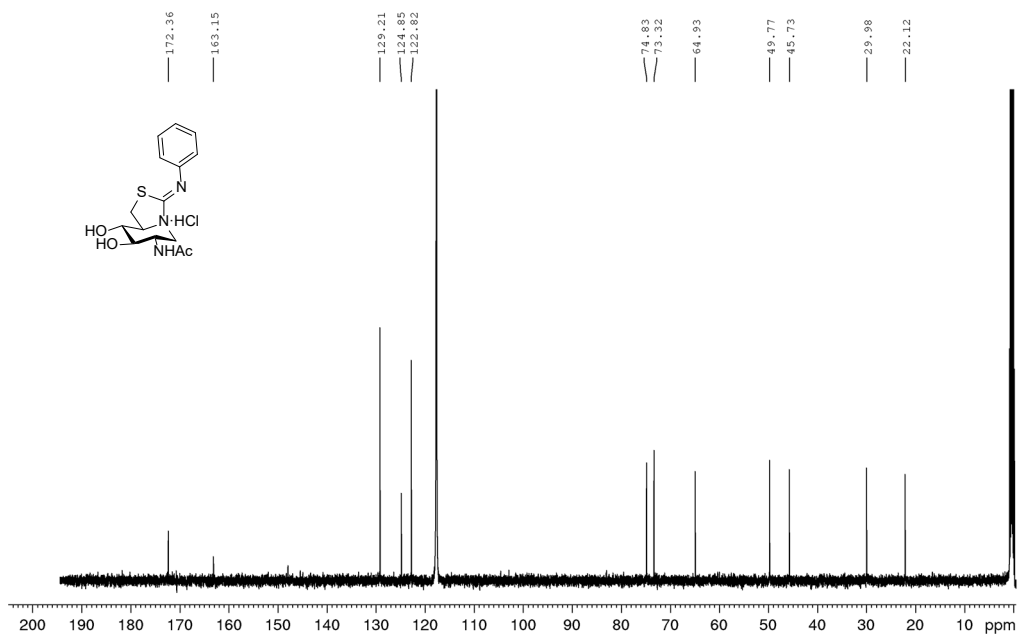
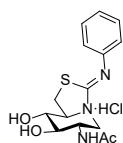
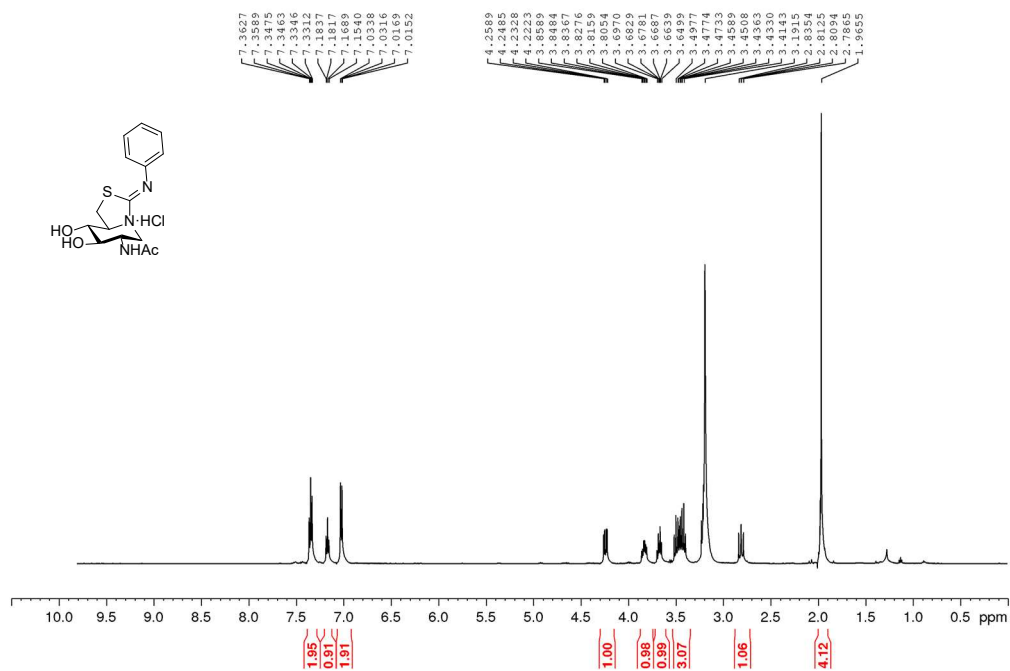


Supplemental Figure 43. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 13

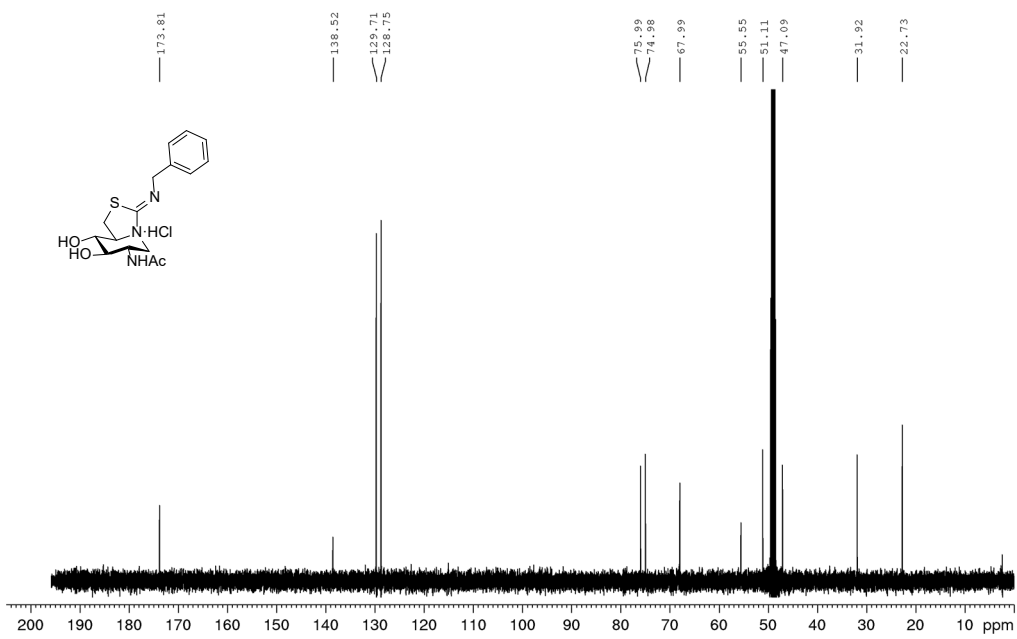
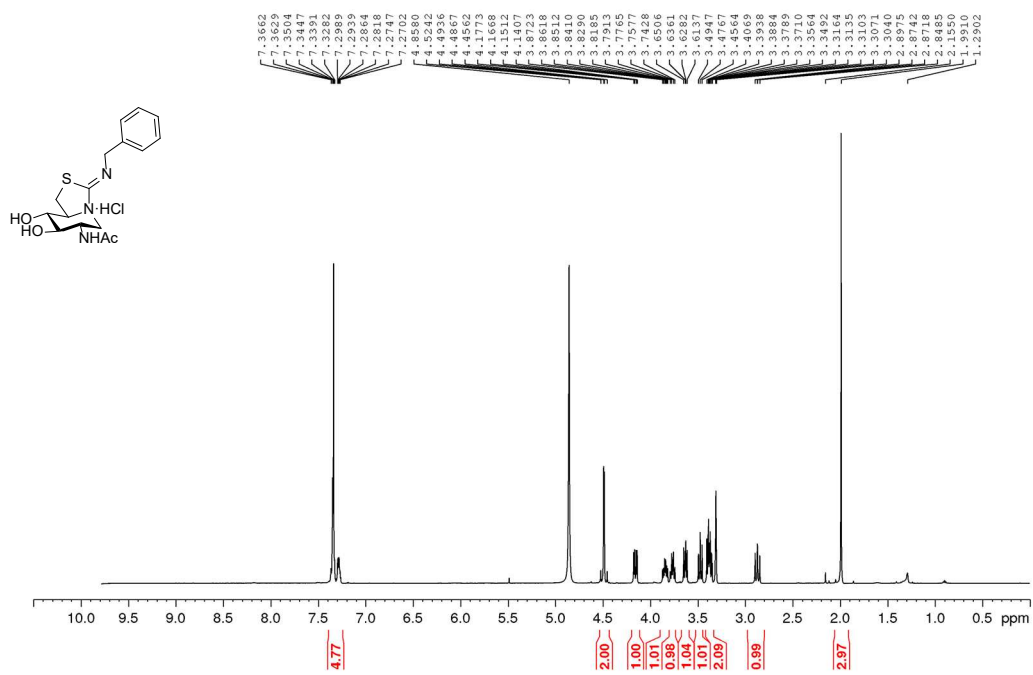




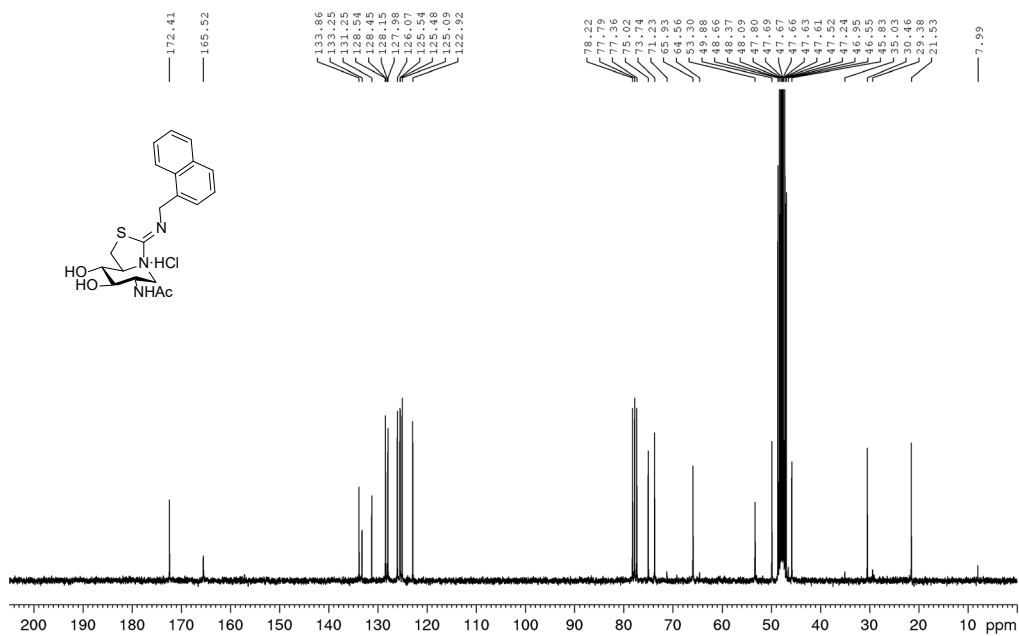
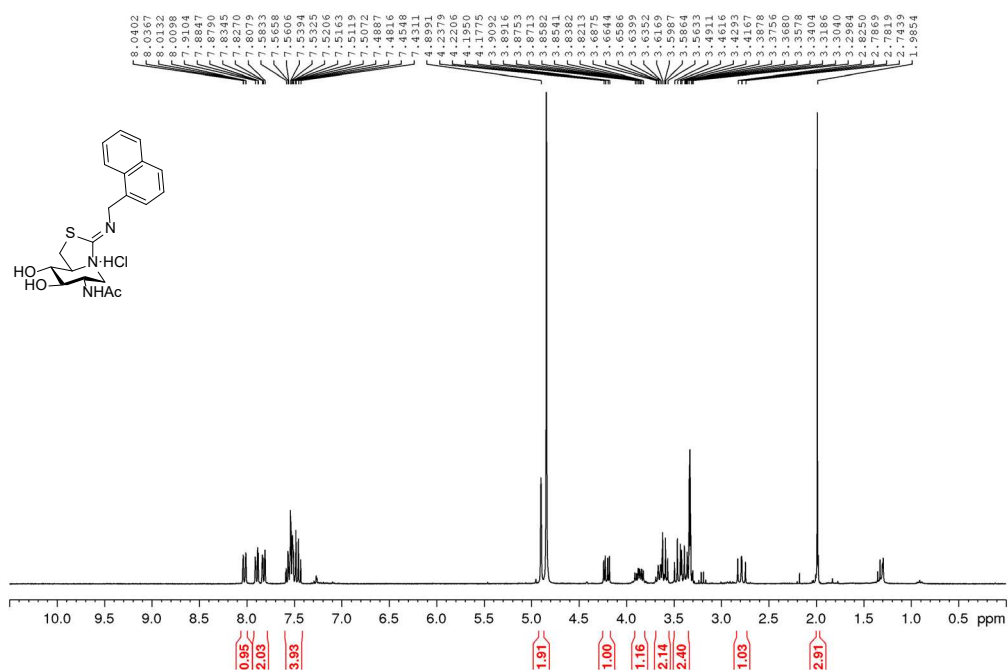
Supplemental Figure 44. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 14



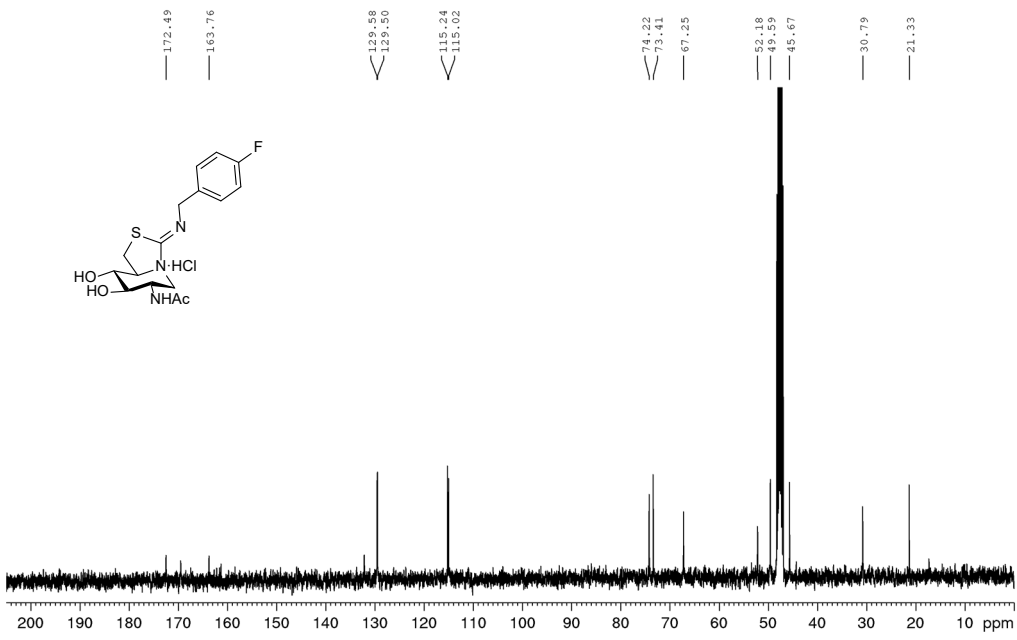
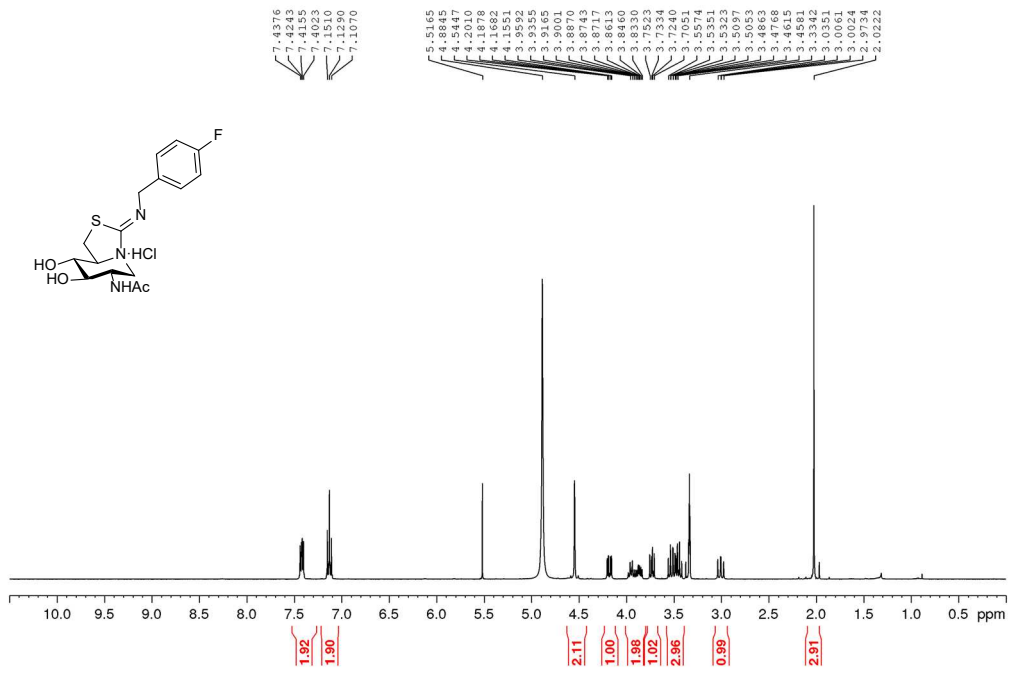
**Supplemental Figure 45.** <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz 1:10 CD<sub>3</sub>CN-D<sub>2</sub>O) of 15



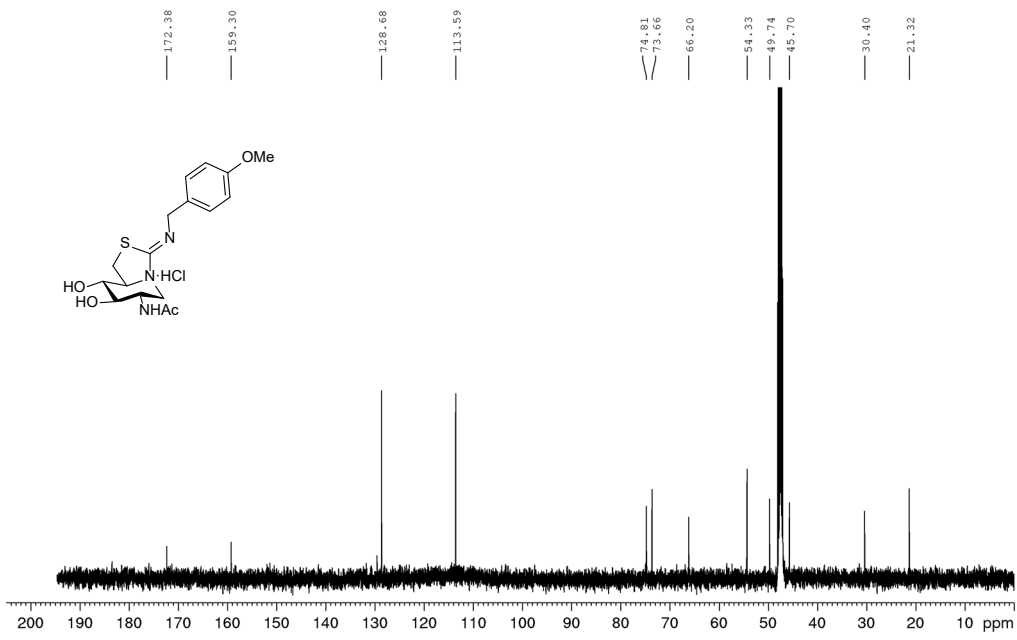
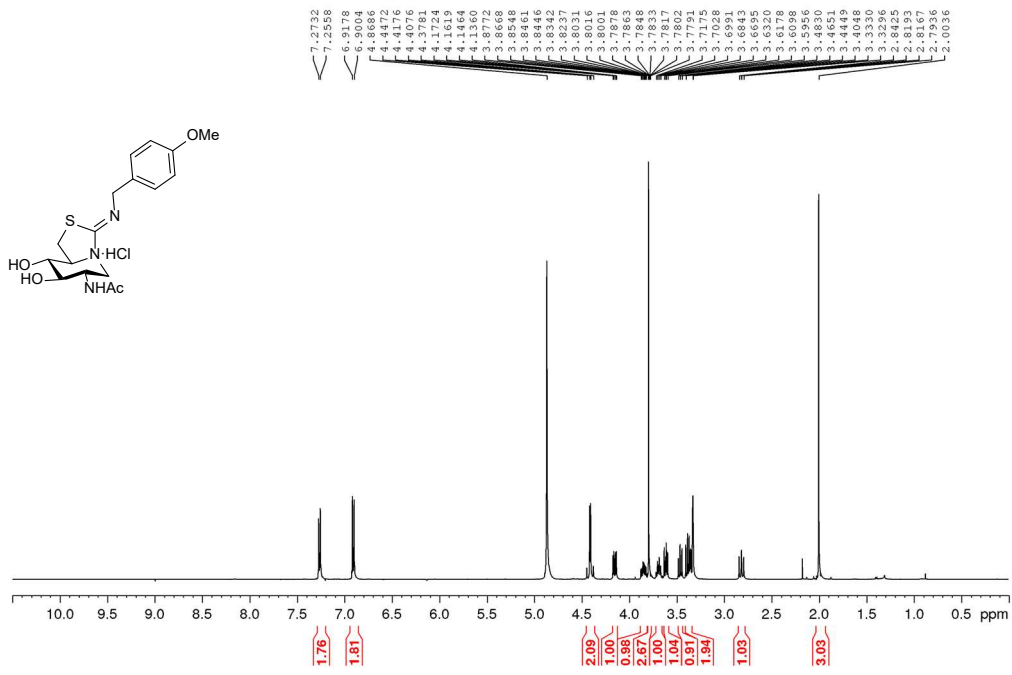
Supplemental Figure 46. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 16



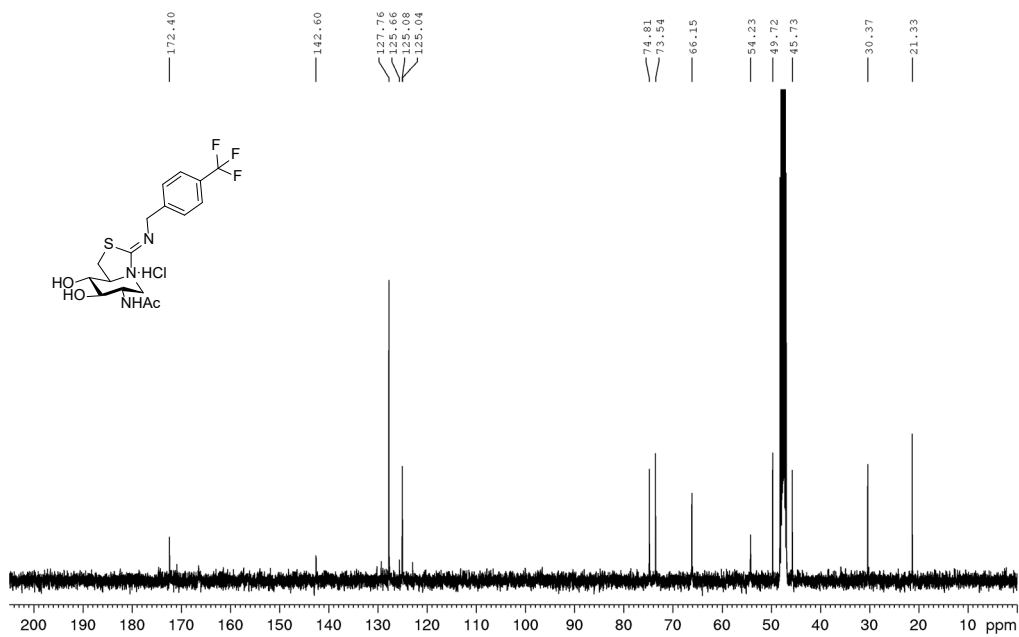
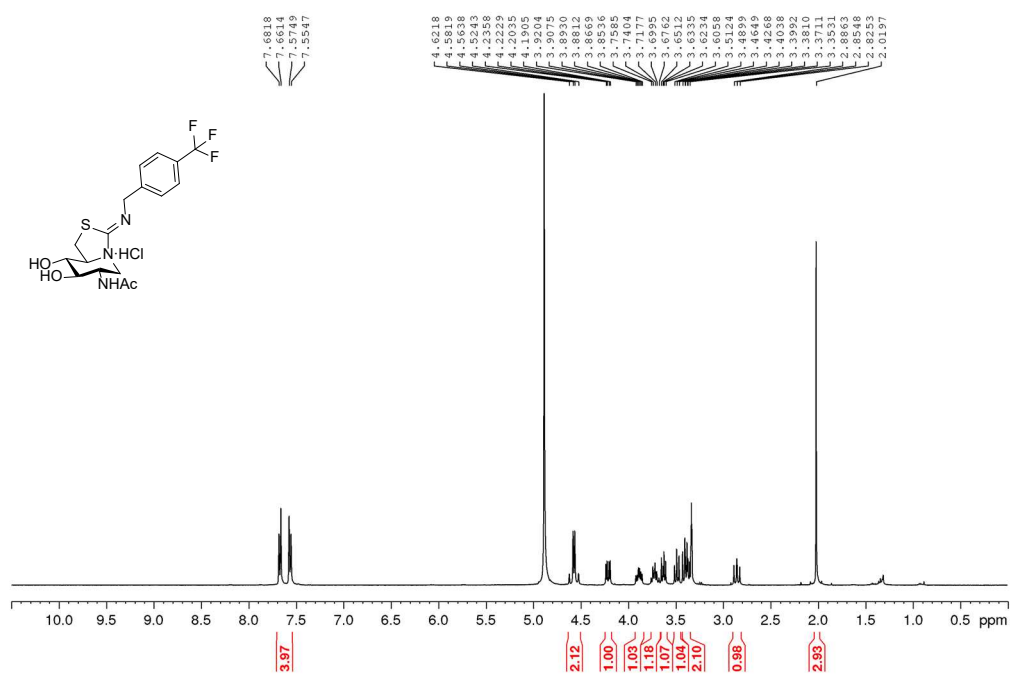
**Supplemental Figure 47.** <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 100.6 MHz 8:1 CD<sub>3</sub>OD-CDCl<sub>3</sub>) of 17



Supplemental Figure 48. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 18



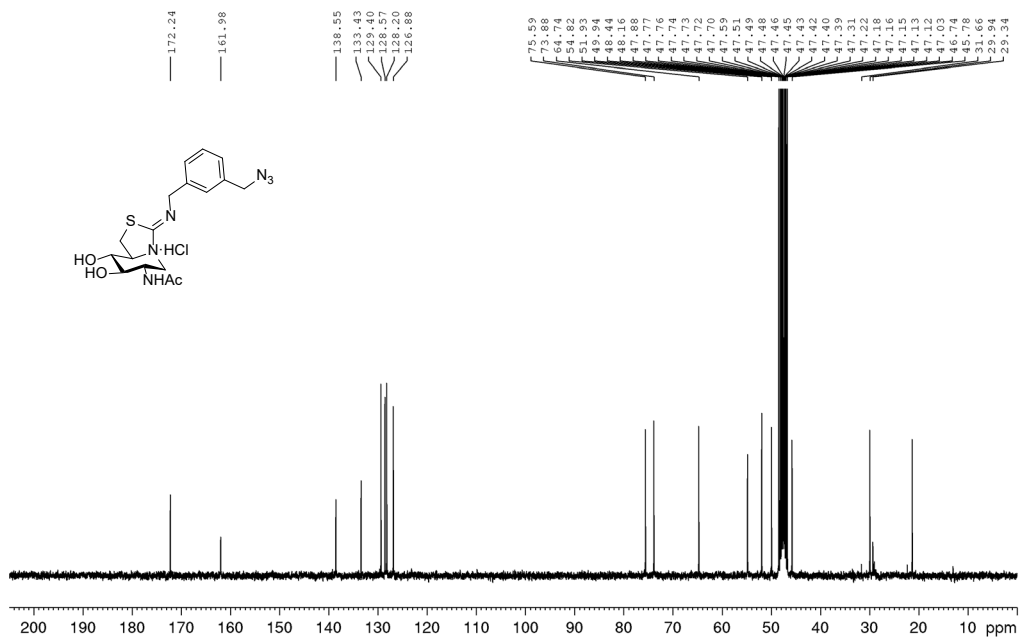
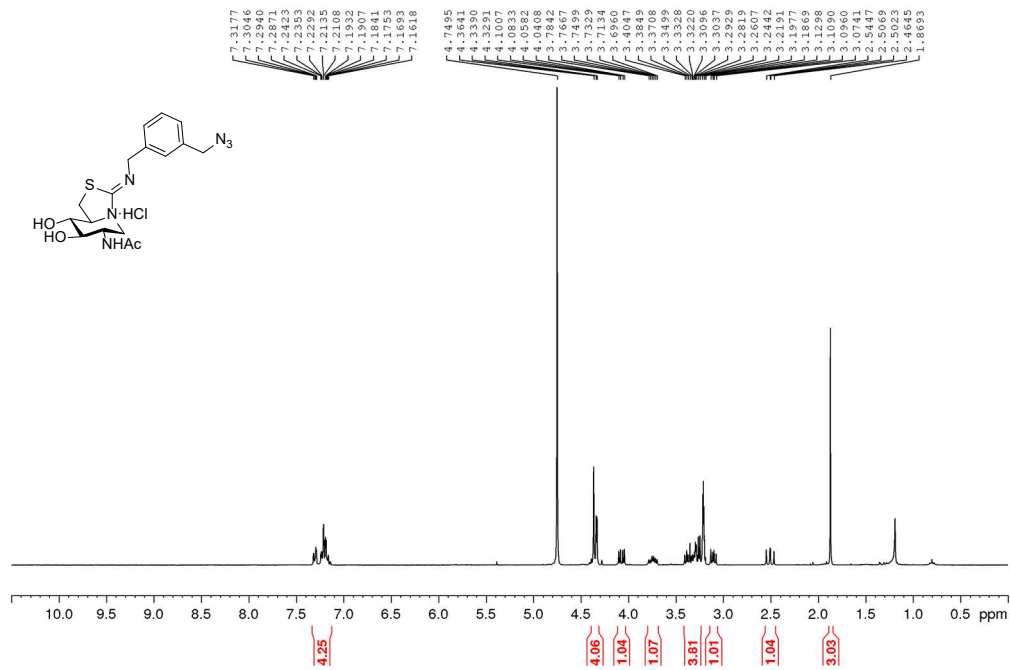
Supplemental Figure 49. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 19



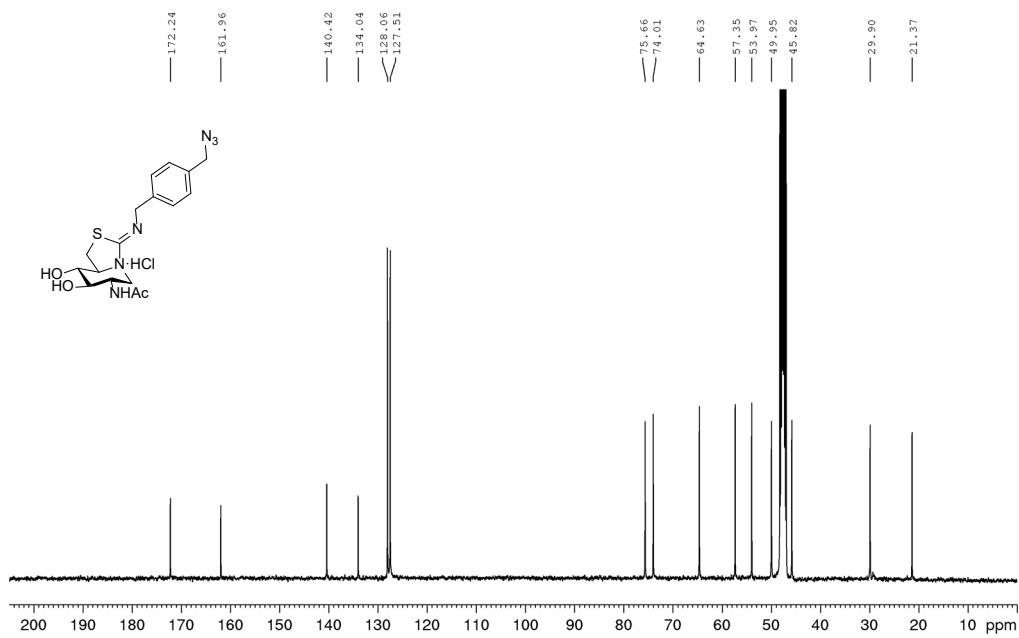
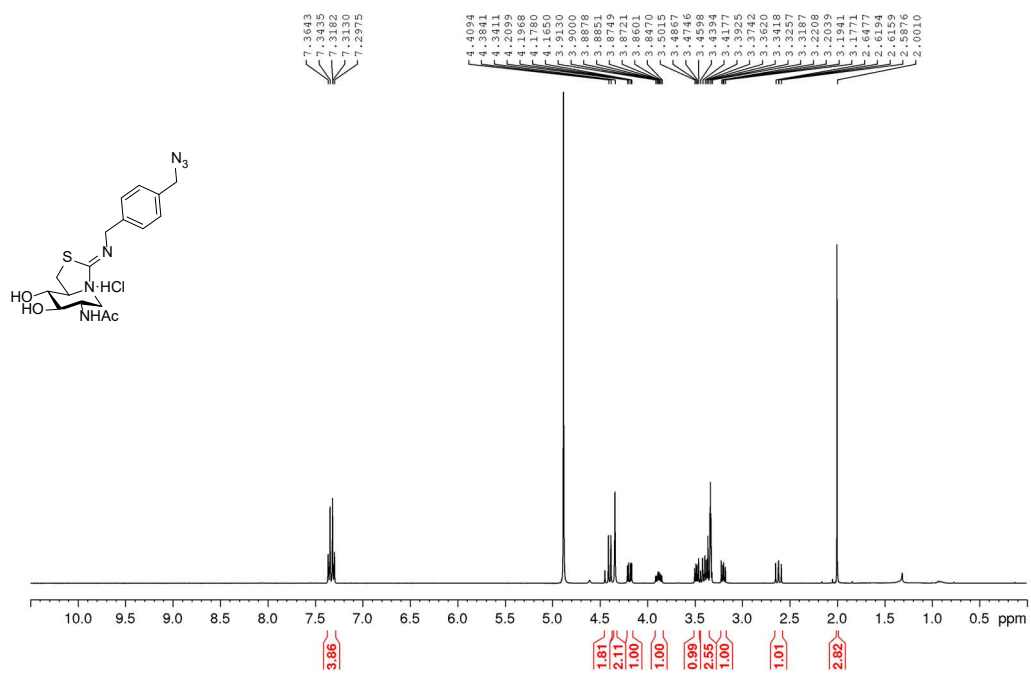
Supplemental Figure 50. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 20



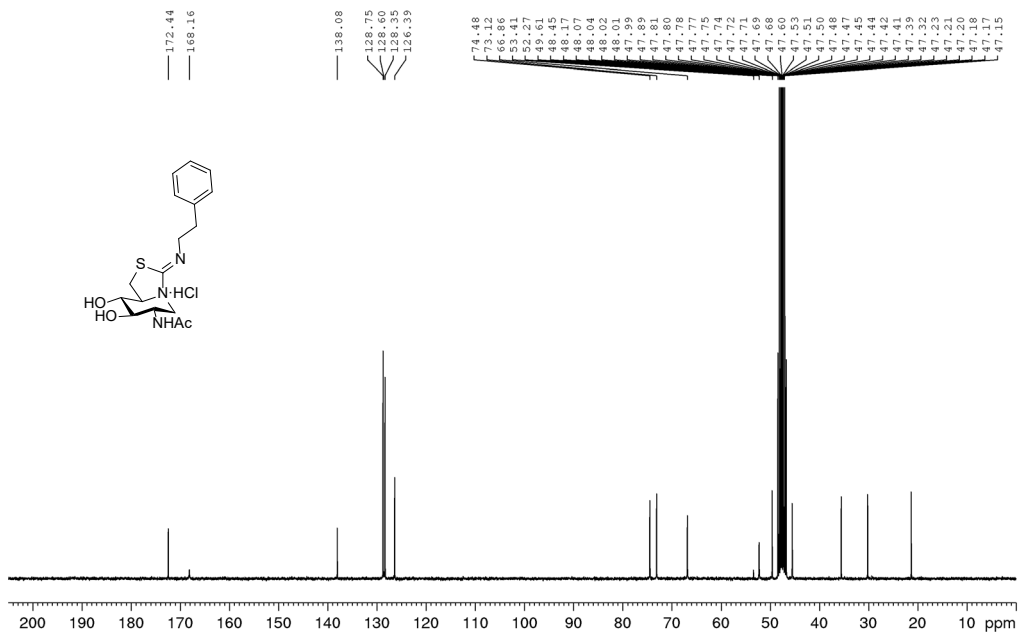
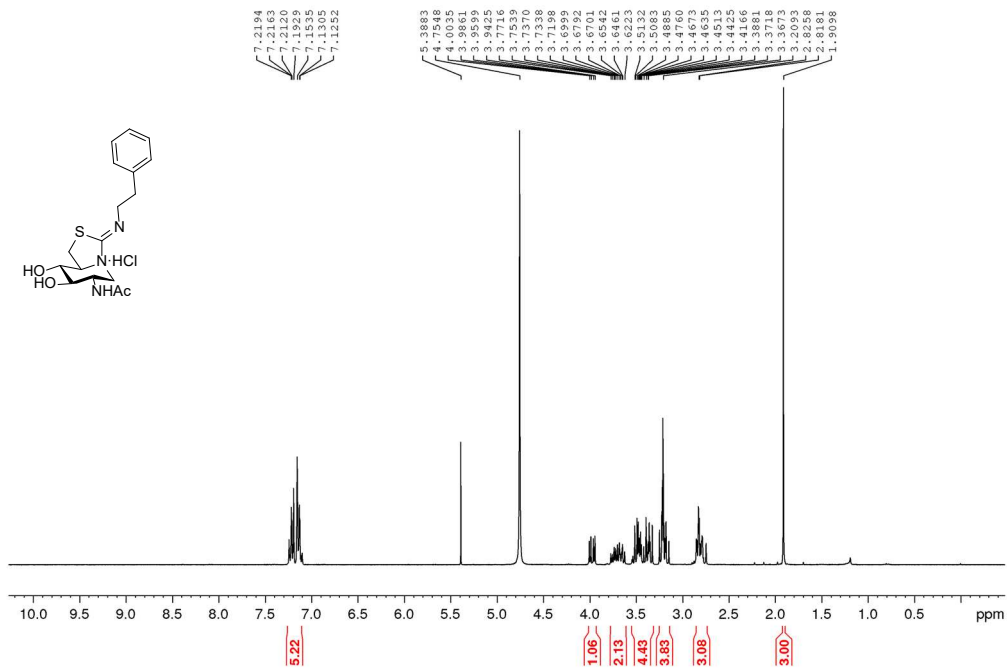




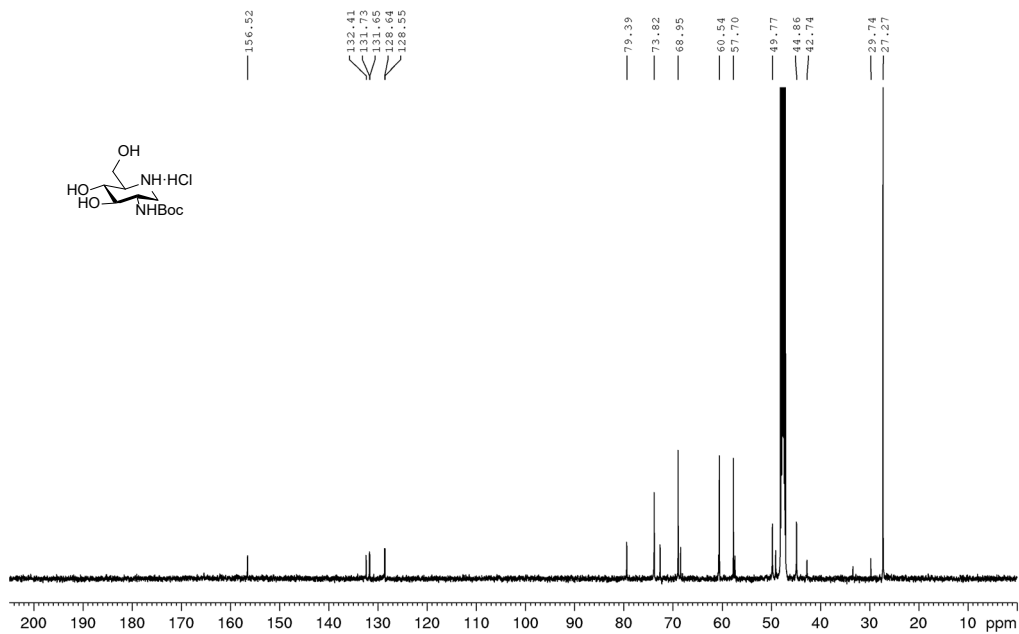
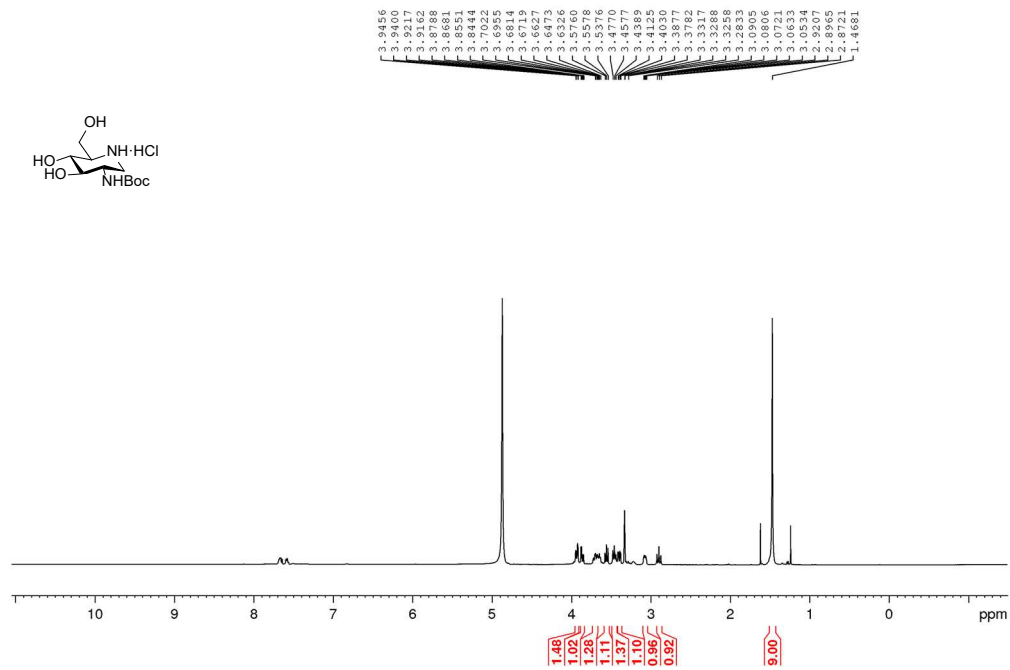
Supplemental Figure 52. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CD<sub>3</sub>OD) of 22



Supplemental Figure 53. <sup>1</sup>H NMR and <sup>13</sup>C NMR (400 MHz, 100.6 MHz CD<sub>3</sub>OD) of 23

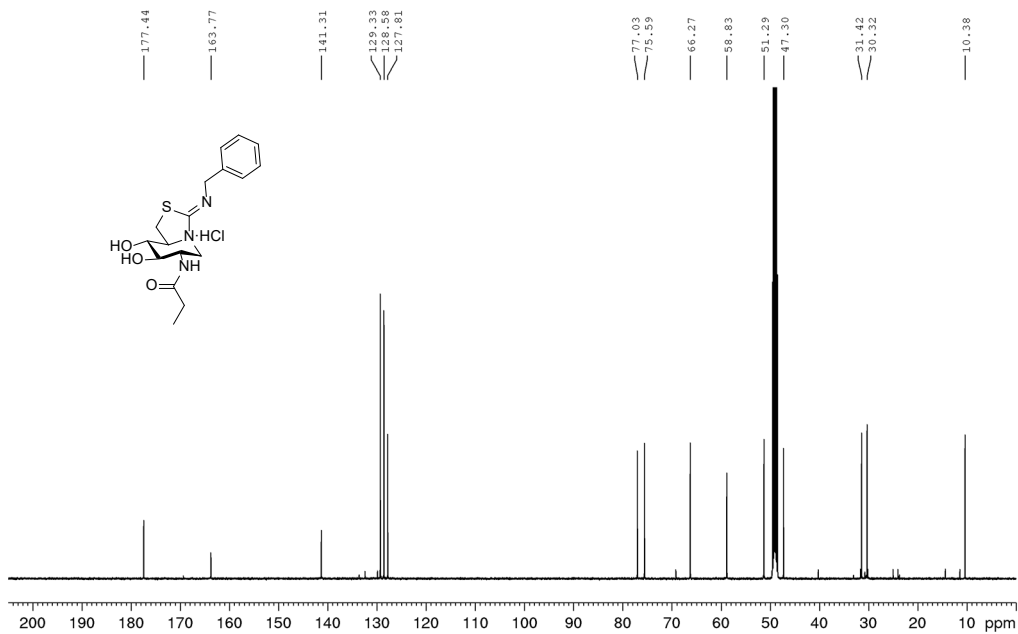
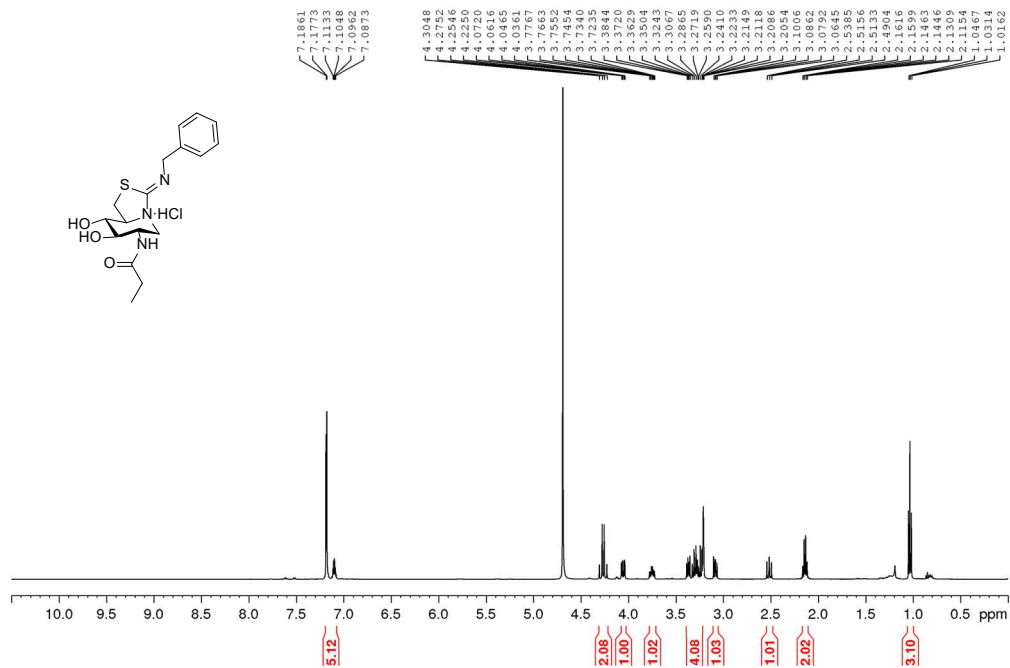


Supplemental Figure 54. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CD<sub>3</sub>OD) of 24

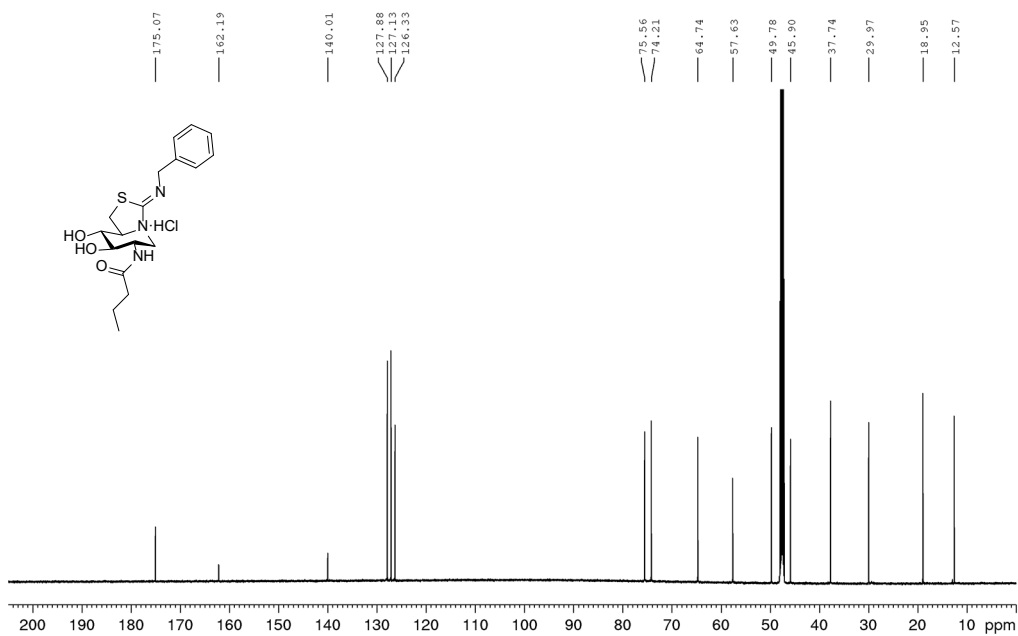
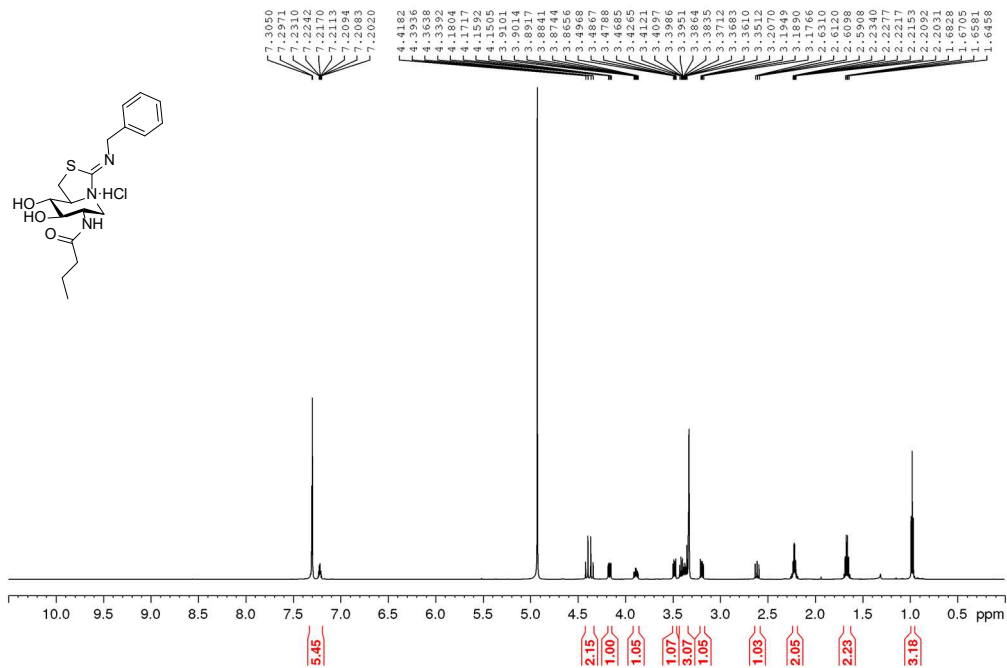


Supplemental Figure 55. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 26

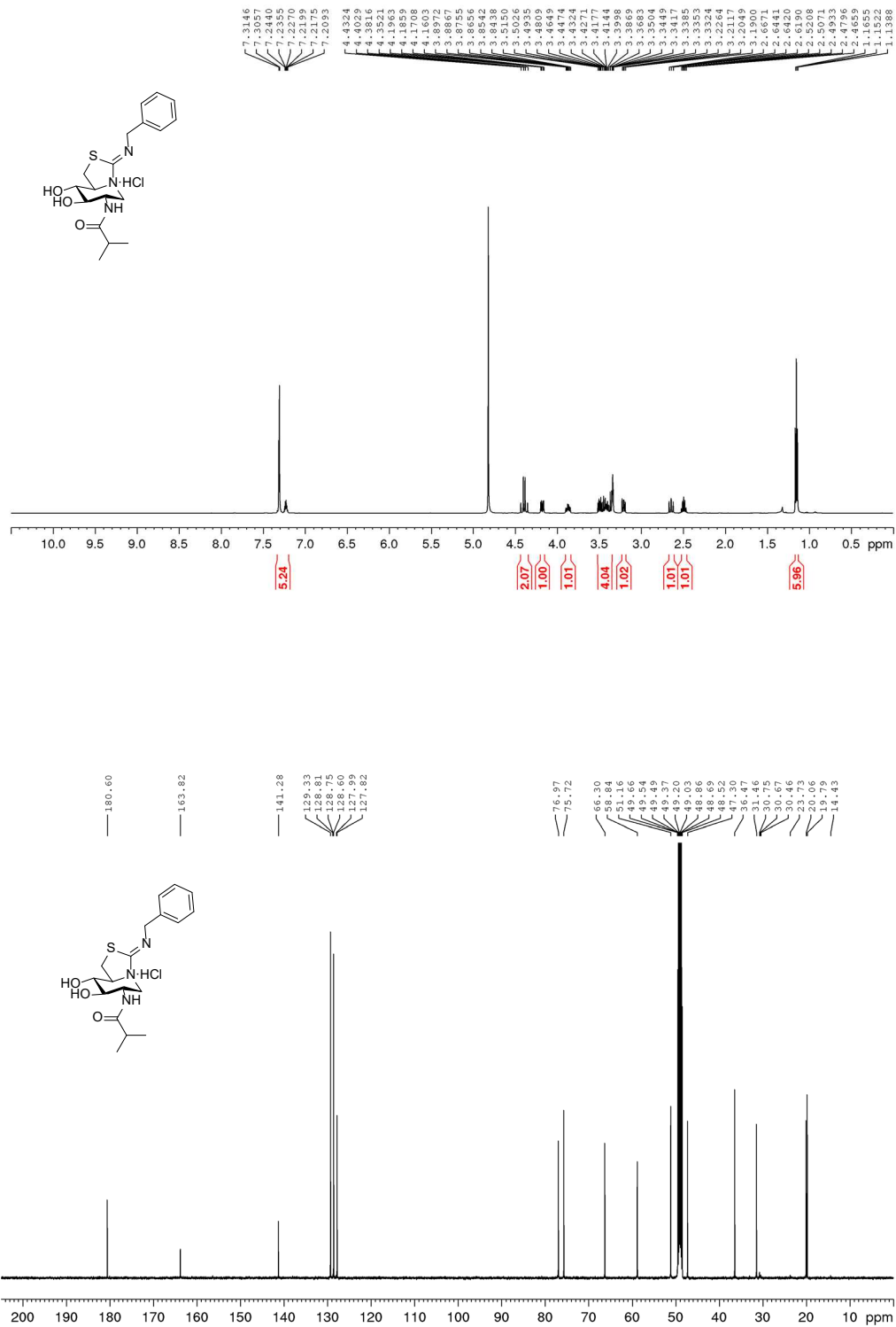




Supplemental Figure 57. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 28

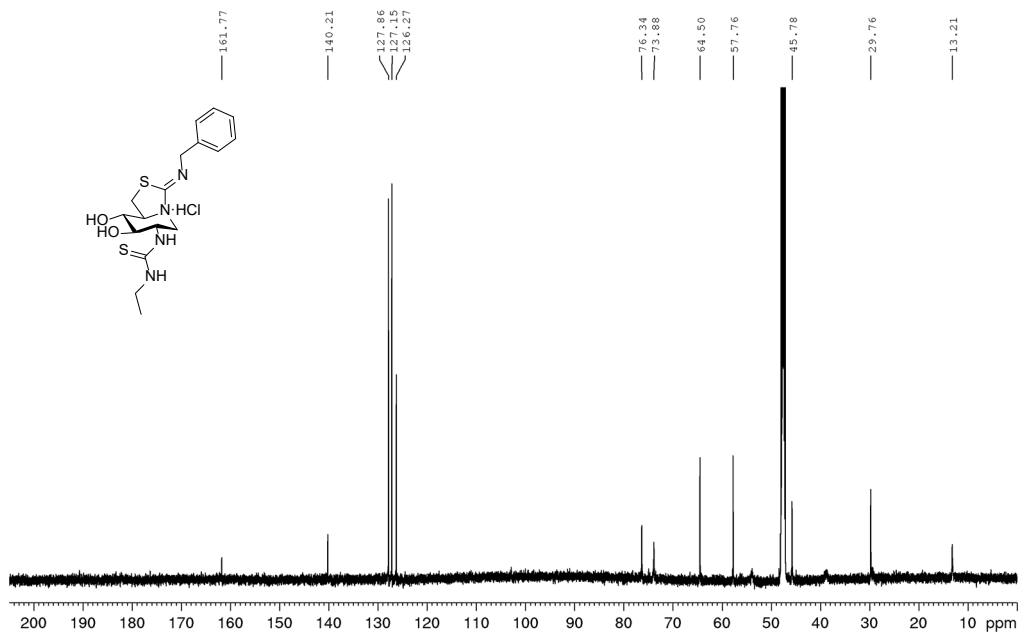
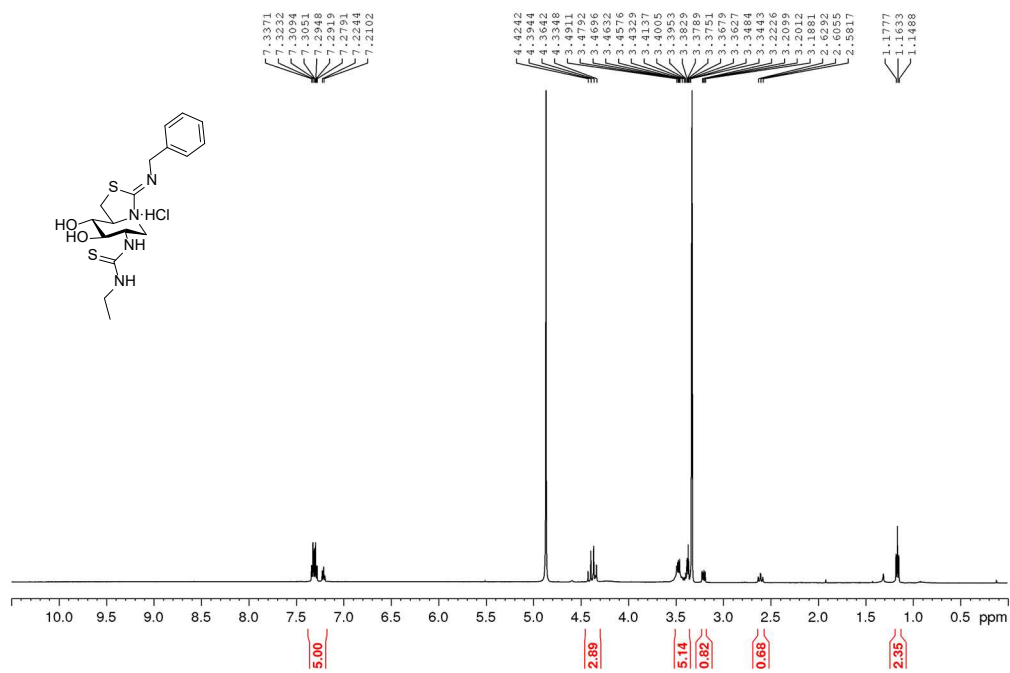


Supplemental Figure 58. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 29

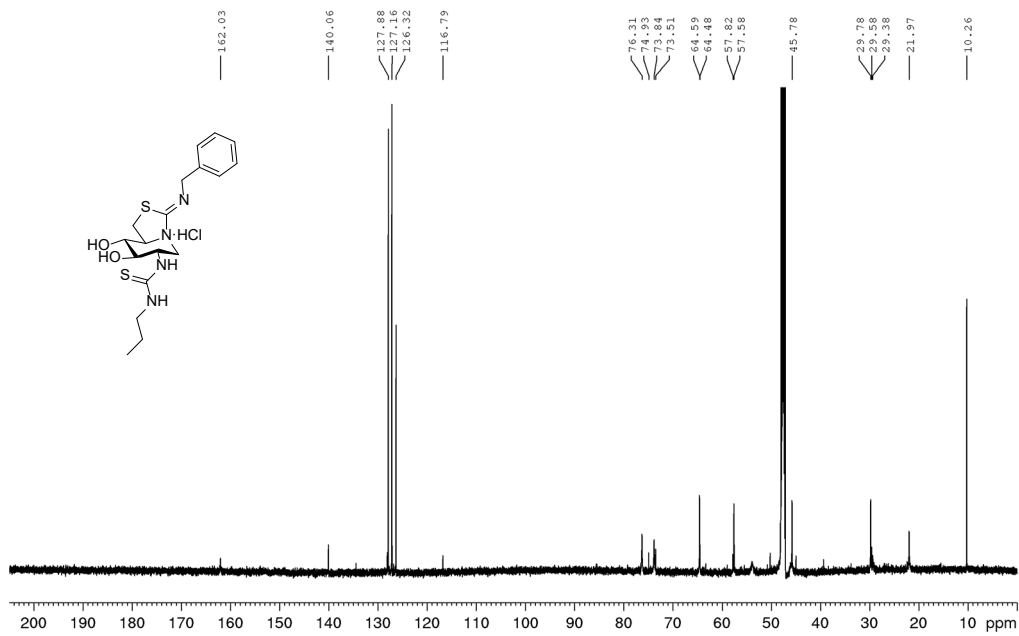
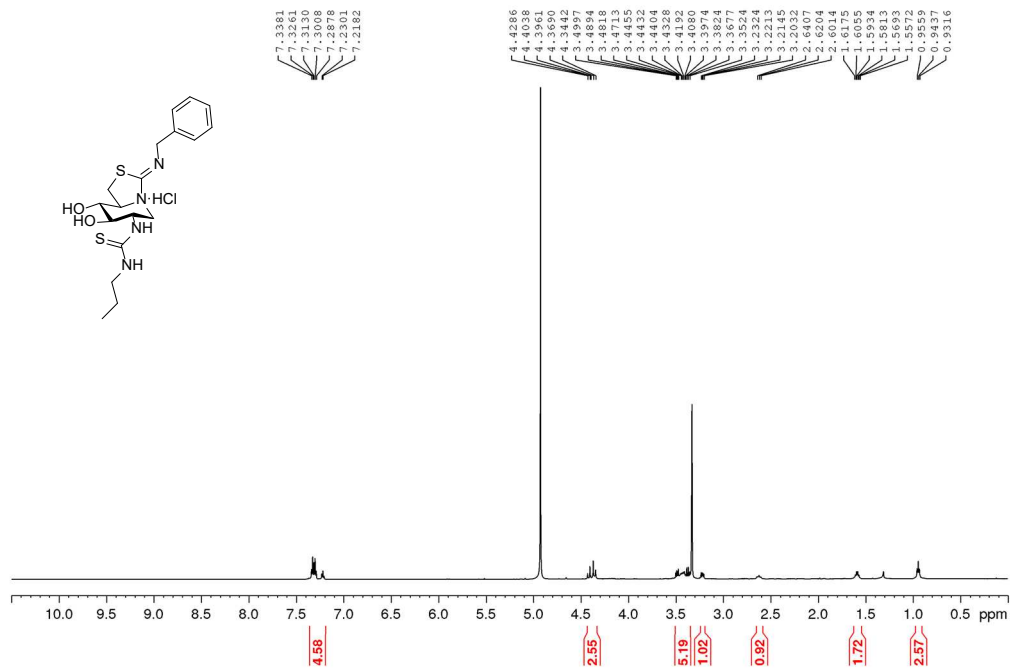


Supplemental Figure 59. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 30

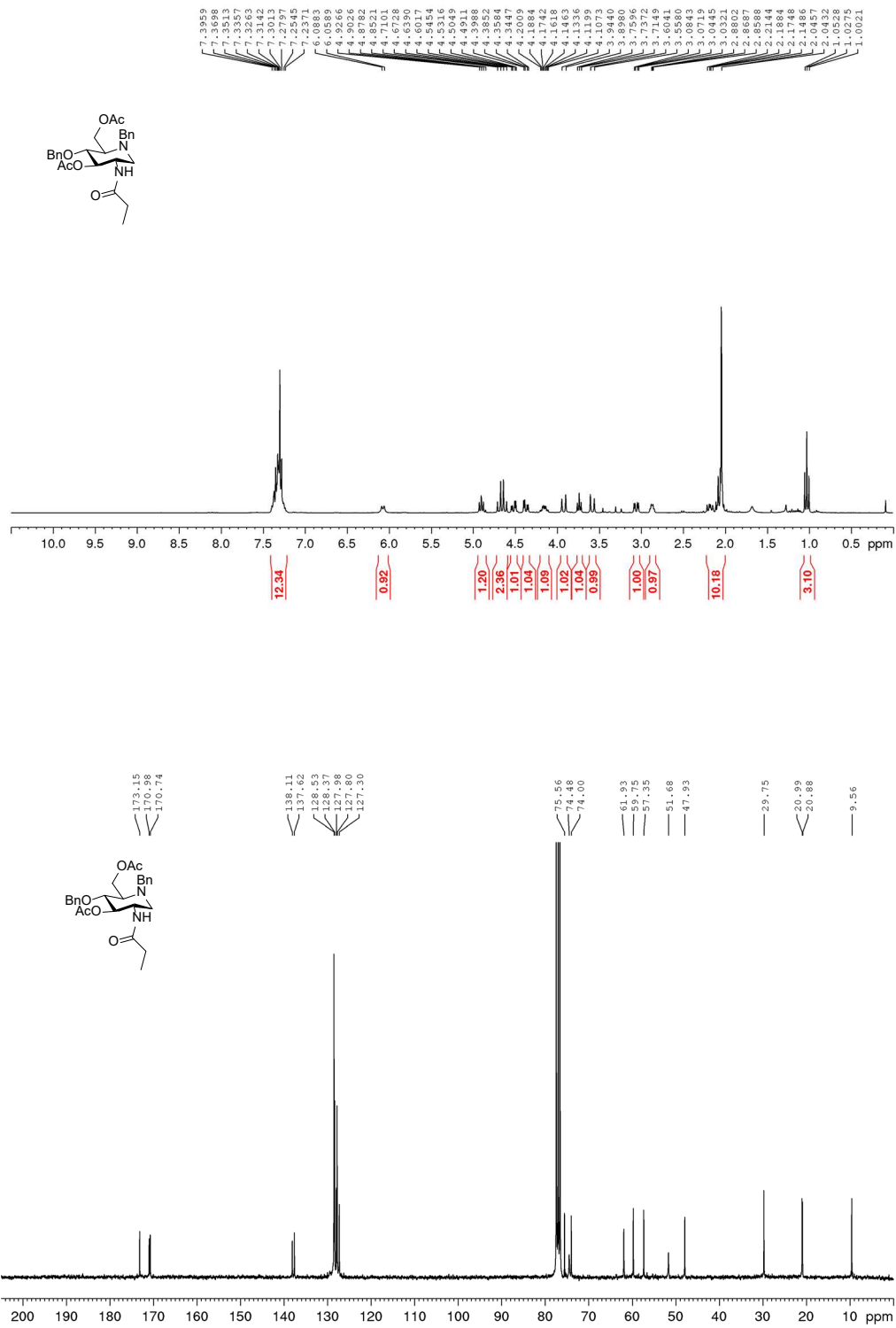




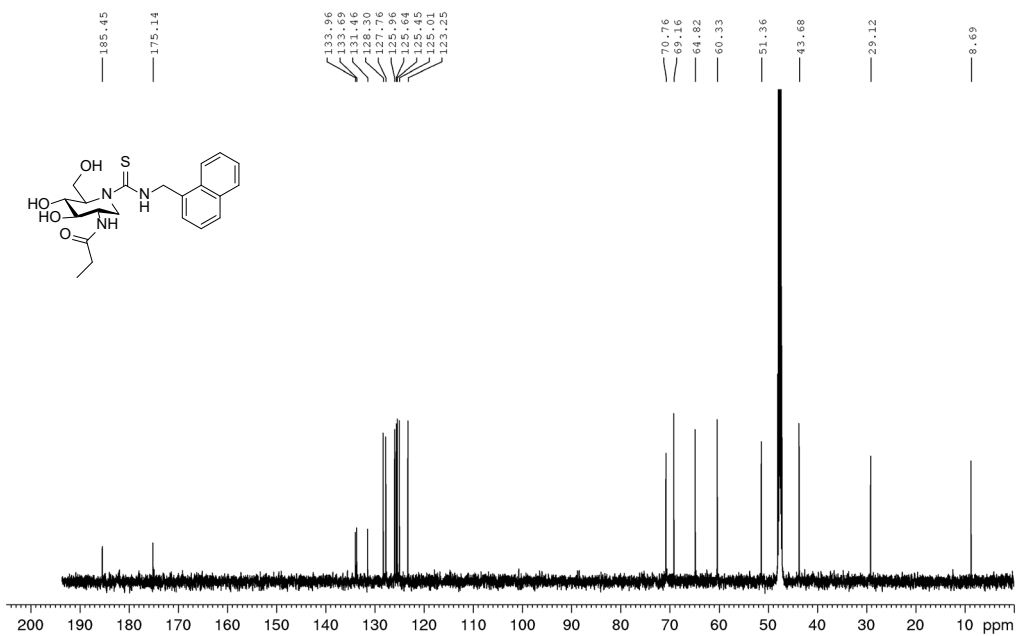
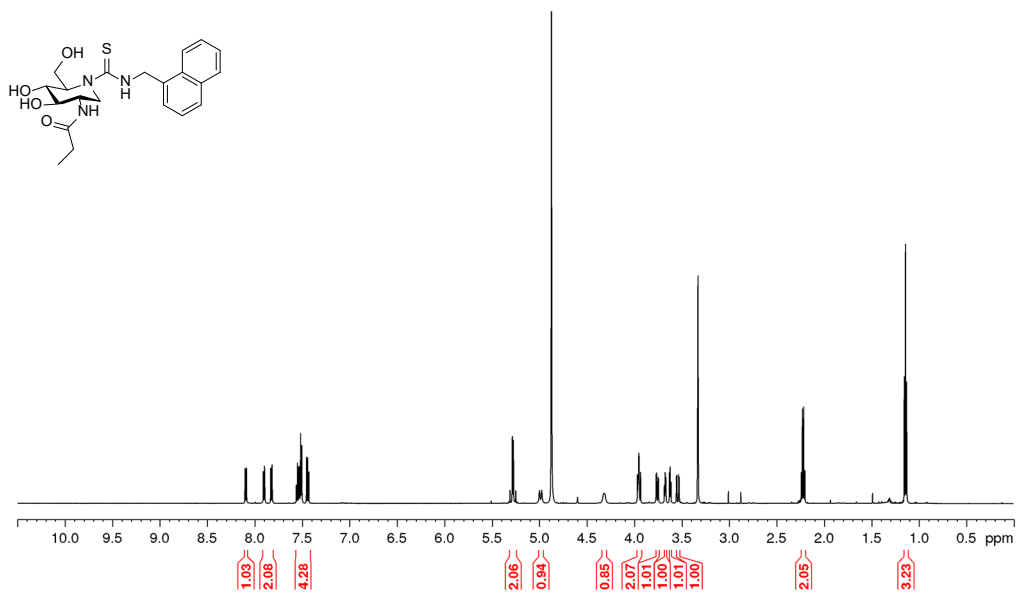
Supplemental Figure 60. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 31



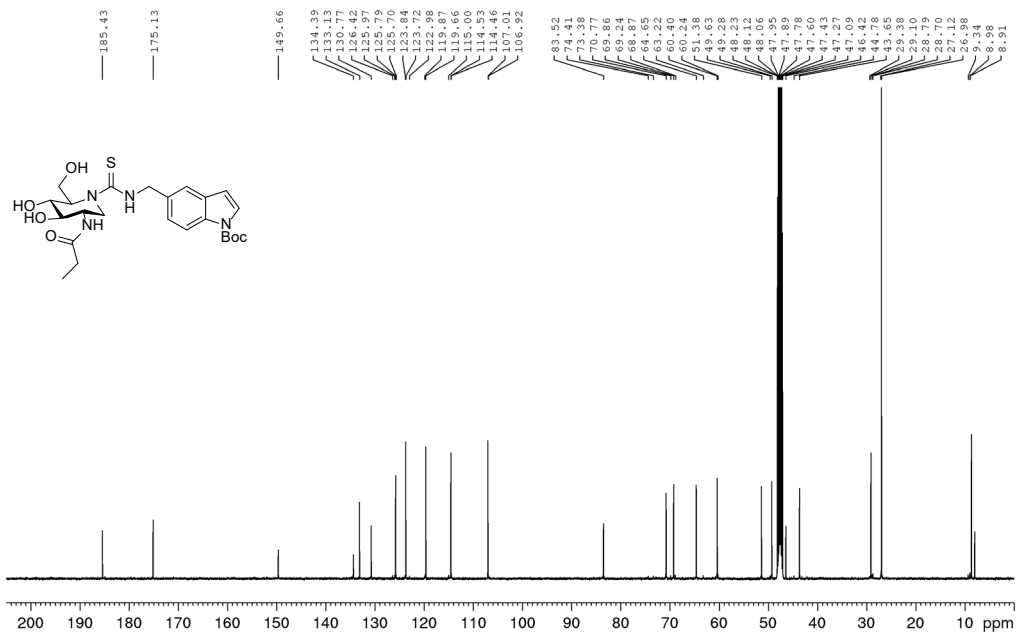
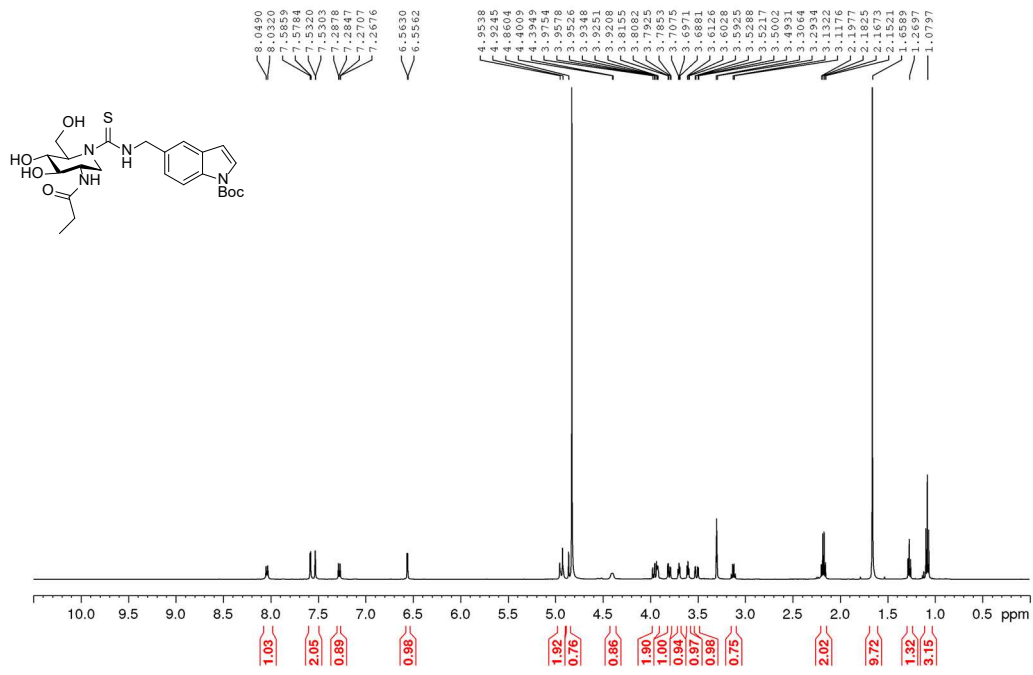
Supplemental Figure 61. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 32



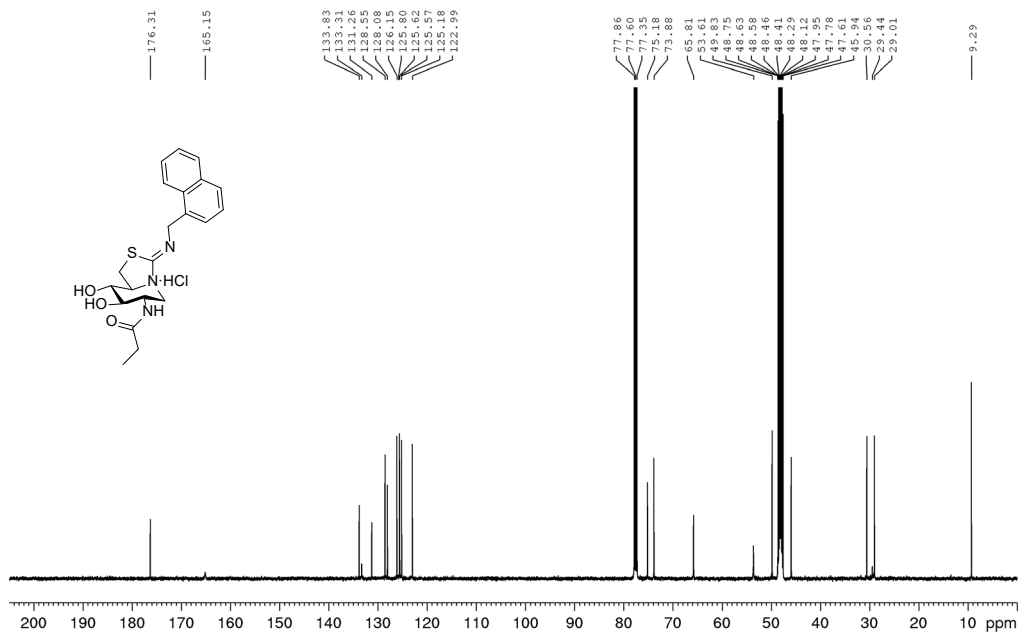
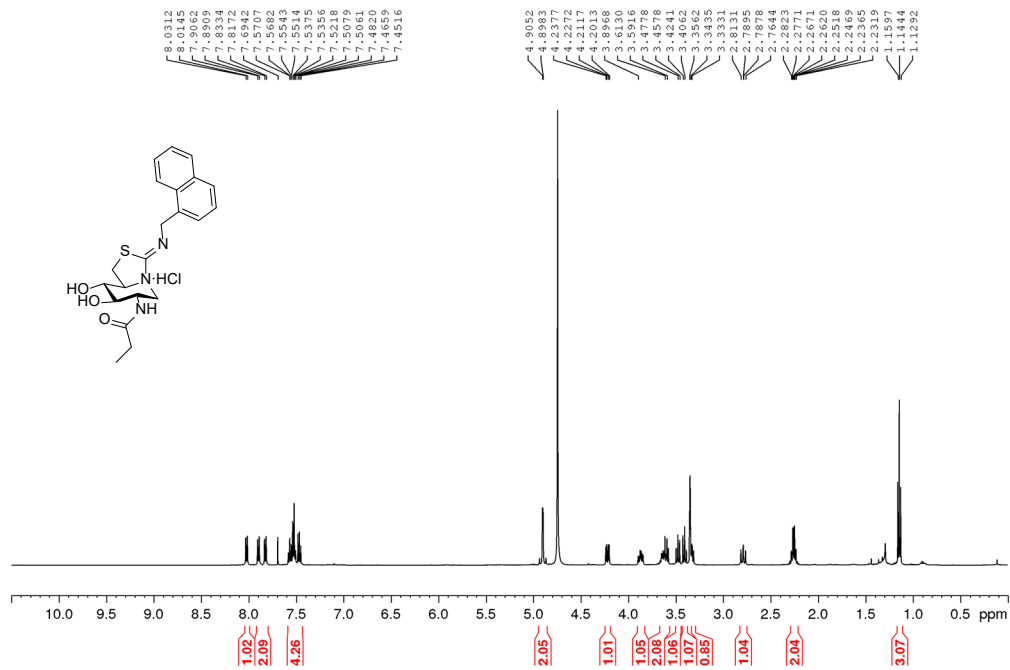
Supplemental Figure 62. <sup>1</sup>H NMR and <sup>13</sup>C NMR (300 MHz, 75.5 MHz CD<sub>3</sub>OD) of 33



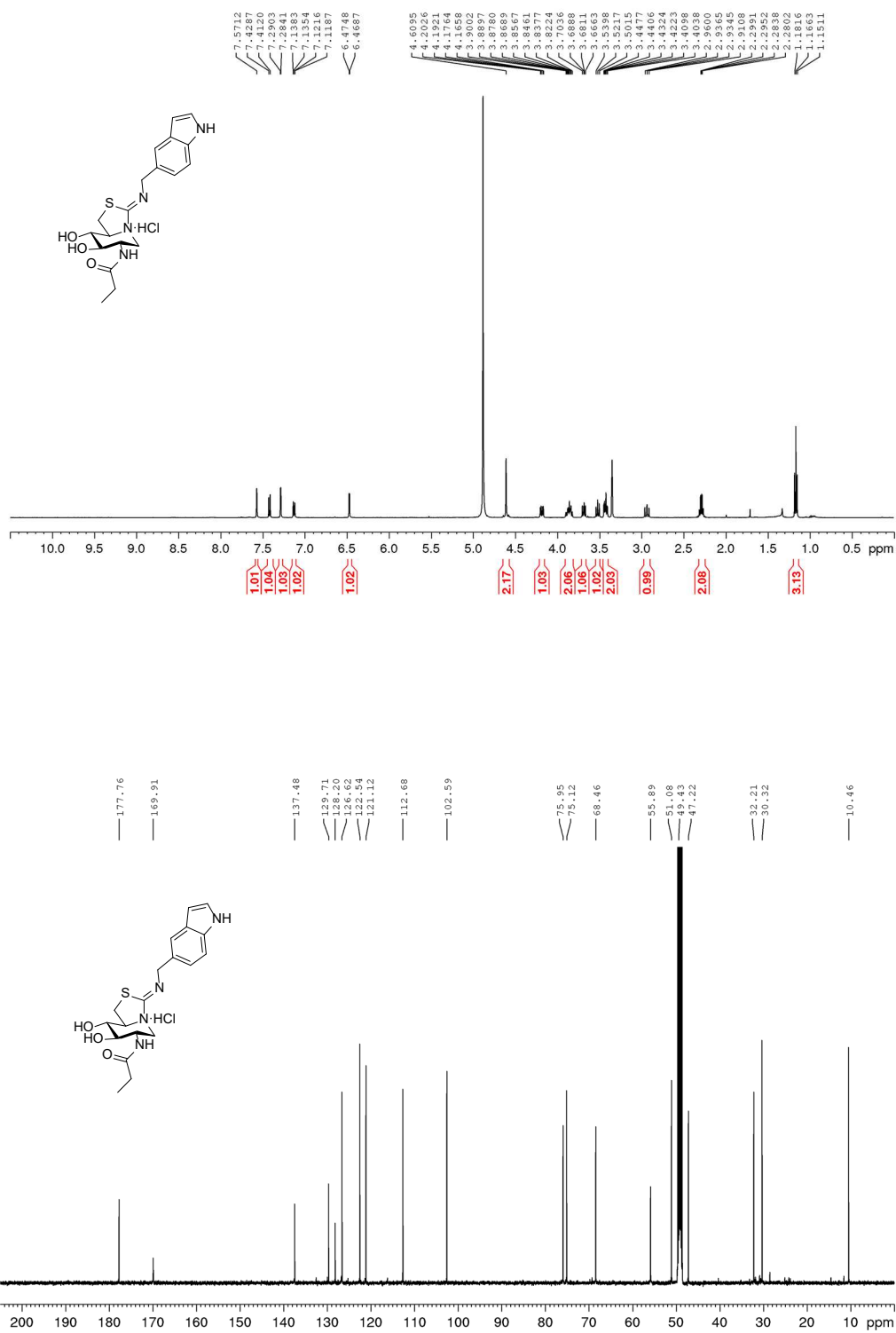
Supplemental Figure 63. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 35



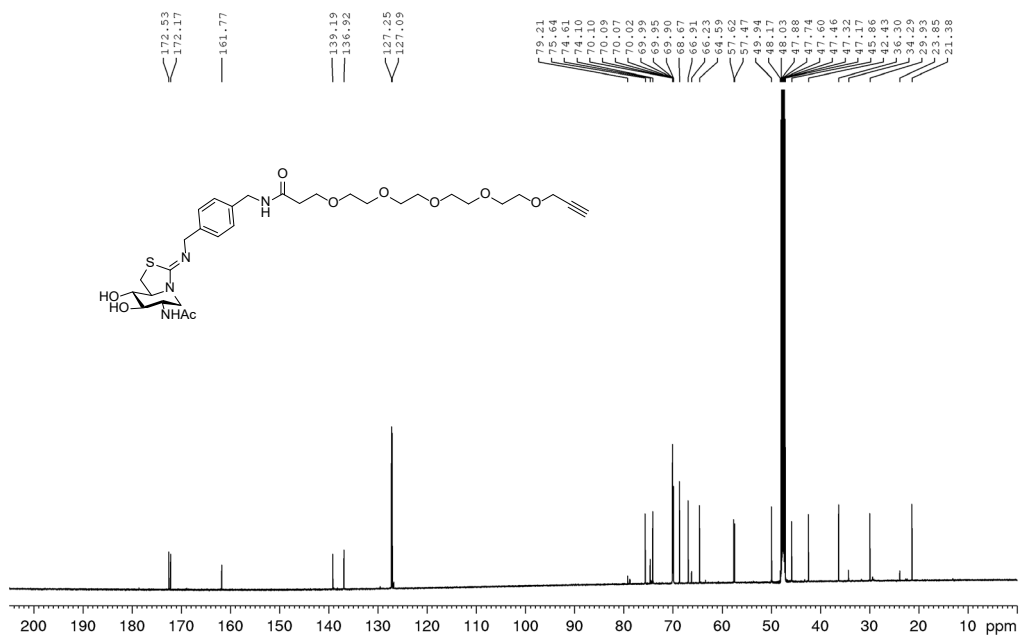
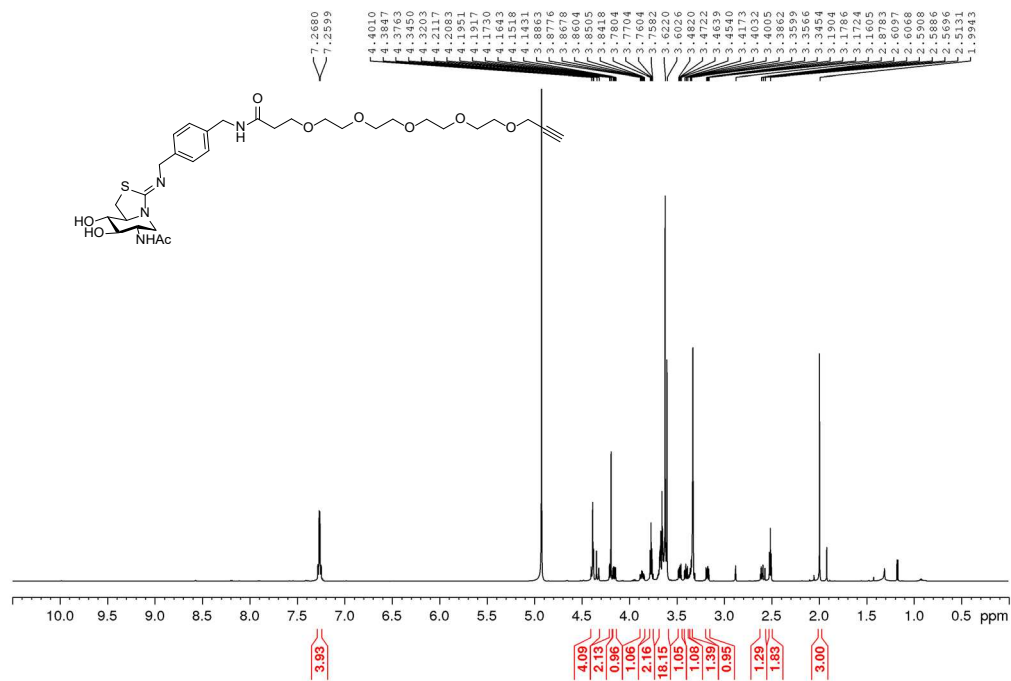
Supplemental Figure 64. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 36



Supplemental Figure 65. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz 1:1 CD<sub>3</sub>OD-CDCl<sub>3</sub>) of 37

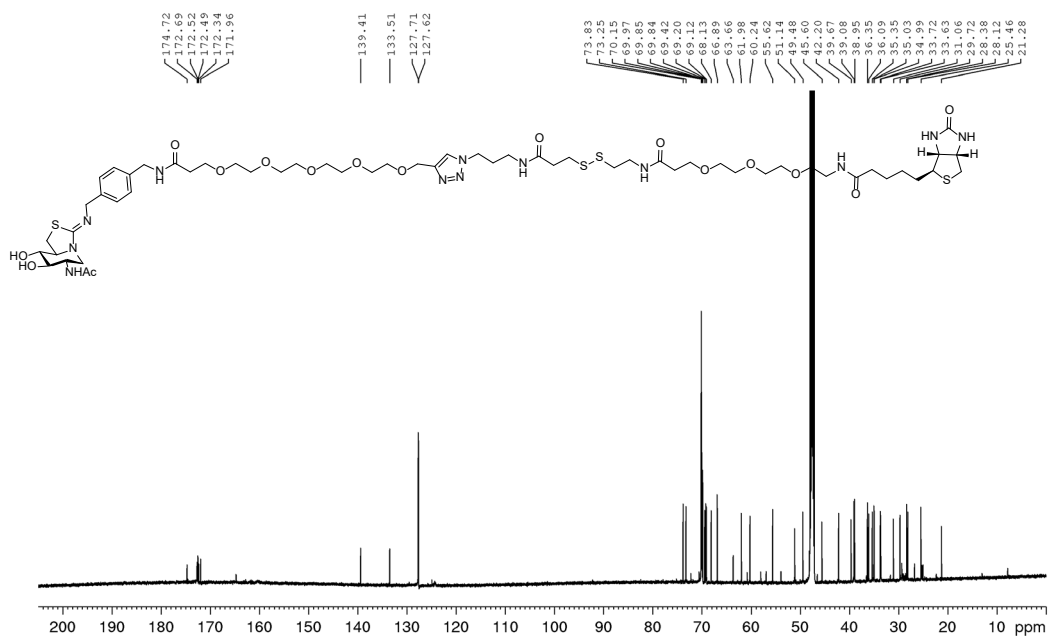
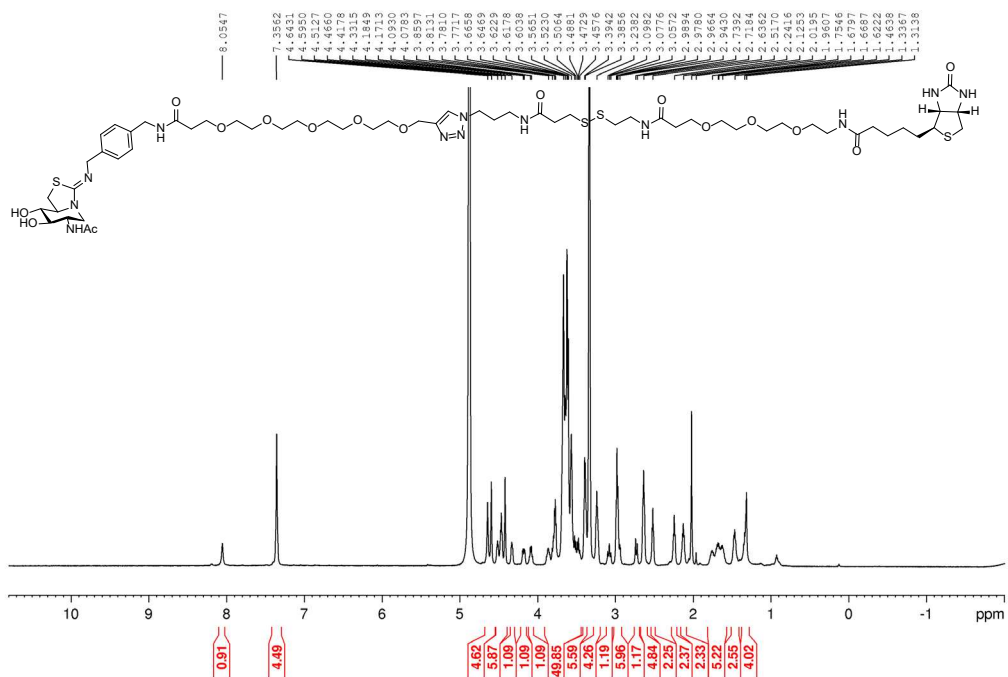


Supplemental Figure 66. <sup>1</sup>H NMR and <sup>13</sup>C NMR (500 MHz, 125.7 MHz CD<sub>3</sub>OD) of 38



Supplemental Figure 67. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 39





Supplemental Figure 68. <sup>1</sup>H NMR and <sup>13</sup>C NMR (600 MHz, 150 MHz CD<sub>3</sub>OD) of 40

## References

- (1) A. F. G. Glawar, D. Best, B. J. Ayers, S. Miyauchi, S. Nakagawa, M. Aguilar-Moncayo, J. M. García Fernández, C. Ortiz Mellet, E. V. Crabtree, T. D. Butters, F. X. Wilson, A. Kato, G. W. J. Fleet. Scalable Syntheses of Both Enantiomers of DNJNAc and DGJNAc from Glucuronolactone: The Effect of N-Alkylation on Hexosaminidase Inhibition. *Chem. Eur. J.* **2012**, *18*, 9341–9359.
- (2) J. R. Thomas, X. Liu, P. J. Hergenrother. Size-Specific Ligands for RNA Hairpin Loops. *J. Am. Chem. Soc.* **2005**, *127*, 12434–12435.
- (3) K. N. Farrugia, D. Makuc, A. Podborska, K. Szaciłowski, J. Plavec, D. C. Magri. Colorimetric Naphthalene-Based Thiosemicarbazide Anion Chemosensors with an Internal Charge Transfer Mechanism. *Eur. J. Org. Chem.* **2016**, *16*, 4415–4422.
- (4) K. A. Stubbs, J. P. Bacik, G. E. Perley-Robertson, G. E. Whitworth, T. M. Gloster, D. J. Vocadlo, B. L. Mark. The Development of Selective Inhibitors of Nagz: Increased Susceptibility of Gram-Negative Bacteria to  $\beta$ -Lactams. *ChemBioChem* **2013**, *14*, 1973–1981.
- (5) K. N. Lau, H. F. Chow, M. C. Chan, K. W. Wong. Dendronized Polymer Organogels from Click Chemistry: A Remarkable Gelation Property Owing to Synergistic Functional-Group Binding and Dendritic Size Effects. *Angew. Chem. Int. Ed.* **2008**, *47*, 6912–6916.
- (6) C. Roth, S. Chan, W. A. Offen, G. R. Hemsworth, L. I. Willems, D. T. King, V. Varghese, R. Britton, D. J. Vocadlo, G. J. Davies. Structural and Functional Insight into Human O-GlcNAcase. *Nat. Chem. Biol.* **2017**, *13*, 610-612.
- (7) R. J. Dennis, E. J. Taylor, M. S. Macauley, K. A. Stubbs, J. P. Turkenburg, S. J. Hart, G. N. Black, D. J. Vocadlo, G. J. Davies. Structure and Mechanism of a Bacterial  $\beta$ -Glucosaminidase Having O-GlcNAcase Activity. *Nat. Struct. Mol. Biol.* **2006**, *13*, 365-371.
- (8) G. Winter, J. Appl. Xia2 : An Expert System for Macromolecular Crystallography Data Reduction. *Crystallogr.* **2010**, *43*, 186-190.

- (9) P. Evans. Scaling and Assessment of Data Quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62*, 72-82.
- (10) P. R. Evans, G. N. Murshudov. How Good Are My Data and What Is the Resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1204-1214.
- (11) M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson. Overview of the CCP 4 Suite and Current Developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 235-242.
- (12) A. Vagin, A. Teplyakov. Molecular Replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 22-25.
- (13) P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486-501.
- (14) G. N. Murshudov, A. A. Vagin, E. J. Dodson. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53*, 240-255.
- (15) G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin. REFMAC 5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355-367.
- (16) N. S. Pannu, G. N. Murshudov, E. J. Dodson, R. J. Read. Incorporation of Prior Phase Information Strengthens Maximum-Likelihood Structure Refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 1285-1292.
- (17) A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, G. N. Murshudov. REFMAC 5 Dictionary: Organization of Prior Chemical Knowledge and Guidelines for Its Use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2184-2195.
- (18) C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B.

Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* **2018**, *27*, 293-315.

(19) J. Agirre, J. Iglesias-Fernández, C. Rovira, G. J. Davies, K. S. Wilson, K. D. Cowtan. Privateer: Software for the Conformational Validation of Carbohydrate Structures. *Nat. Struct. Mol. Biol.* **2015**, *22*, 833-834.

(20) S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble. Presenting Your Structures: The CCP 4 Mg Molecular-Graphics Software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 386-394.