



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/98363/>

Version: Accepted Version

---

**Article:**

Speciale, Gaetano, Jin, Y., Davies, Gideon J. et al. (2016) YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. NATURE CHEMICAL BIOLOGY. 215–217. ISSN: 1552-4450

<https://doi.org/10.1038/nchembio.2023>

---

**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 ***Title***

2 **YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids**

3

4 ***Authors***

5 Gaetano Speciale<sup>1,†</sup>, Yi Jin<sup>2,†</sup>, Gideon J. Davies<sup>2</sup>, Spencer J. Williams<sup>1</sup>, Ethan D. Goddard-  
6 Borger<sup>3,4,\*</sup>

7

8 ***Affiliations***

9 <sup>1</sup> School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of  
10 Melbourne, Parkville, Victoria 3010 (Australia)

11 <sup>2</sup> Department of Chemistry, University of York, Heslington, York, YO10 5DD (UK)

12 <sup>3</sup> ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research,  
13 Parkville, Victoria 3052 (Australia)

14 <sup>4</sup> Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010 (Australia)

15

16 † These authors contributed equally to this work

17 \* Correspondence should be addressed to E.D.G.-B. ([goddard-borger.e@wehi.edu.au](mailto:goddard-borger.e@wehi.edu.au)).

18

19 ***Abstract***

20 Sulfoquinovose is produced by photosynthetic organisms at a rate of 10<sup>10</sup> tons per annum and is  
21 degraded by bacteria as a source of carbon and sulfur. We have identified *Escherichia coli* YihQ as  
22 the first dedicated sulfoquinovosidase and the gateway enzyme to sulfoglycolytic pathways.  
23 Structural and mutagenesis studies unveiled the sequence signatures for binding the distinguishing  
24 sulfonate residue, and revealed that sulfoquinovoside degradation is widespread across the tree of  
25 life.

26

27

28 **Main Text**

29 Photosynthetic organisms synthesize the anionic sugar sulfoquinovose (SQ) in quantities estimated  
30 at 10<sup>10</sup> tons per annum.<sup>1</sup> The principal form of SQ is the plant glycolipid sulfoquinovosyl  
31 diacylglyceride (SQDG), which represents a significant component of the thylakoid membrane of  
32 the chloroplast.<sup>2</sup> SQ is present in such abundance, globally, that it comprises a major reservoir of  
33 organosulfur, approximately equal to that present as cysteine and methionine in proteins.<sup>1</sup> Recent  
34 discoveries have identified two sulfoglycolytic pathways enabling SQ metabolism in bacteria,<sup>3,4</sup>  
35 however, the enzymes responsible for cleaving sulfoquinovosides have remained obscure.

36 The biogenesis of SQDG involves the assembly of uridine 5'-diphospho (UDP)-SQ from  
37 UDP-glucose, and glycosyltransferase-catalyzed conjugation to diacylglycerol (Fig. 1a).<sup>5</sup> While it  
38 has been recognized for some time that certain prokaryotic<sup>6–8</sup> and eukaryotic<sup>9</sup> microorganisms are  
39 capable of metabolizing SQ to access its carbon and sulfur, only recently have the first biochemical  
40 pathways responsible for SQ catabolism been identified and characterized.<sup>3,4</sup> Sulfoglycolysis,  
41 named after the Embden–Meyerhof–Parnas glycolysis pathway, was first defined within  
42 *Escherichia coli* and involves the conversion of SQ, via sulfofructose-1-phosphate, to (S)-2,3-  
43 dihydroxypropane-1-sulfonate (DHPS) and dihydroxyacetone phosphate (DHAP).<sup>3</sup> DHAP supports  
44 primary metabolism, whereas DHPS is transported out of the cell and is degraded by other  
45 bacteria.<sup>10</sup> The sulfoglycolysis gene cluster is a feature of the core-genome of all sequenced *E. coli*  
46 strains and is present in a wide range of *Gammaproteobacteria*, revealing widespread utilization of  
47 SQ as a carbon source, and a source of DHPS for the greater bacterial community.<sup>3</sup> The Entner–  
48 Doudoroff pathway for SQ degradation (hereafter SQ Entner–Doudoroff pathway), as recently  
49 identified in *Pseudomonas putida* SQ1, converts SQ, via 6-deoxy-6-sulfogluconate, to (S)-  
50 sulfolactate (SL) and pyruvate (PYR).<sup>4</sup> Pyruvate enters the tricarboxylic acid cycle, whereas SL is  
51 exported and utilized by other bacteria.<sup>8</sup> This pathway appears to be widespread in *Alpha*-, *Beta*-  
52 and *Gammaproteobacteria*.<sup>4</sup>

53 While SQDG is the primary source of SQ supplying these metabolic pathways, it is  
54 unknown how SQ is liberated from SQDG. Aside from an early report ascribing weak SQDG  
55 hydrolytic activity to *E. coli*  $\beta$ -galactosidase,<sup>11</sup> no glycoside hydrolases (GHs) dedicated to  
56 processing sulfoquinovosides have been reported. Putative SQases are located within the *E. coli*  
57 sulfoglycolysis and *P. putida* SQ Entner-Doudoroff gene clusters (YihQ and PpSQ1\_00094,  
58 respectively).<sup>3,4</sup> They are members of GH family 31 within the CAZy<sup>12</sup> sequence-based  
59 classification (<http://www.cazy.org>), a family that contains enzymes with  $\alpha$ -glucosidase,  $\alpha$ -glucan  
60 lyase, and  $\alpha$ -xylanase activity. We cloned and expressed YihQ in *E. coli* using a similar approach to  
61 that reported.<sup>13</sup> Incubation of YihQ with SQDG isolated from spinach, and analysis of the reaction  
62 mixture by liquid chromatography/mass spectrometry (LC/MS), revealed complete conversion to  
63 SQ (Fig. 1b). A similar experiment on 1-sulfoquinovosylglycerol (SQGro), a metabolite of SQDG  
64 generated by the action of lipases,<sup>14</sup> indicated YihQ also hydrolysed simple sulfoquinovosides  
65 (Supplementary Results, Supplementary Fig. 1). To facilitate the convenient and accurate  
66 determination of kinetic parameters for YihQ we prepared a highly soluble, chromogenic substrate,  
67 *para*-nitrophenyl  $\alpha$ -sulfoquinovoside (PNPSQ, Supplementary Notes). PNPSQ allowed the  
68 continuous acquisition of YihQ reaction rate data and the calculation of Michaelis-Menten  
69 parameters, revealing robust catalysis with  $k_{\text{cat}} = 14.3 \pm 0.4 \text{ s}^{-1}$ ,  $K_{\text{M}} = 0.22 \pm 0.03 \text{ mM}$  and  $k_{\text{cat}}/K_{\text{M}} =$   
70  $(6.4 \pm 1.0) \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  (Supplementary Table 1). Under comparable conditions, we could not detect  
71 any activity against PNP  $\alpha$ -D-glucopyranoside. Since CAZy GH family 31 contains both retaining  
72 glycosidases and  $\alpha$ -glucan lyases,<sup>15</sup> we performed <sup>1</sup>H NMR spectroscopic analysis of the YihQ-  
73 catalyzed cleavage of PNPSQ to demonstrate rapid hydrolysis to the  $\alpha$ -anomer of SQ, which after  
74 further time underwent mutarotation, confirming that YihQ is a retaining GH (Fig. 1c).

75 The X-ray structure of YihQ reveals an  $(\alpha\beta)_8$  barrel appended with a small  $\beta$ -sheet domain  
76 (Fig. 2a, Supplementary Table 2). By soaking with the mechanism-based inactivator 5-fluoro- $\beta$ -L-  
77 idopyranosyl fluoride (5FIdoF),<sup>16</sup> a covalent glycosyl-enzyme complex (in a <sup>1</sup>S<sub>3</sub> pyranose  
78 conformation) was obtained, supporting assignment of D405 as the catalytic nucleophile (Fig. 2b).<sup>17</sup>

79 Located appropriately to protonate the glycosidic oxygen is D472, assigned as the catalytic  
80 acid/base residue.<sup>18</sup> A pseudo Michaelis complex was obtained by construction of a catalytically-  
81 inactive variant by mutation of the acid/base residue. The complex of PNPSQ with YihQ D472N  
82 revealed binding of the intact substrate in a <sup>4</sup>C<sub>1</sub> conformation (Fig. 2c, Supplementary Fig. 2).  
83 Overall, the architectural features of the YihQ active site, formation of a glycosyl-enzyme  
84 intermediate with 5FIdoF, and the observation of retention of stereochemistry upon substrate  
85 hydrolysis are consistent with a classical Koshland retaining mechanism in which D405 fulfils the  
86 role of catalytic nucleophile and D472 acts as a general acid/base. Consistent with these  
87 assignments, the D405A, D405N, D472A and D472N YihQ variants were each catalytically  
88 inactive (Supplementary Table 1). Our data are consistent with a conformational itinerary of <sup>4</sup>C<sub>1</sub> →  
89 <sup>4</sup>H<sub>3</sub><sup>‡</sup> → <sup>1</sup>S<sub>3</sub> for the YihQ glycosylation half-reaction (Supplementary Fig. 3).<sup>19,20</sup>

90 The PNPSQ complex with YihQ D472N reveals a detailed picture of the structural features  
91 required to recognise the distinguishing sulfonate group of SQ (Fig. 2d, Supplementary Fig. 2). The  
92 positively-charged R301 residue forms a salt-bridge with one oxygen of the anionic sulfonate group,  
93 a hydrogen-bond is formed between the indole N-H of W304 and a second sulfonate oxygen, and a  
94 well-ordered water molecule hydrogen bonded to O4 of SQ and Y508 forms a hydrogen bond to the  
95 third sulfonate oxygen. Comparison of the complex with that of a related GH31  $\alpha$ -glucosidase,  
96 SBG (from sugar beet) in complex with acarbose,<sup>21</sup> highlights key differences in the active-site  
97 residues that may explain the specificity of YihQ for SQ over D-glucose (Fig. 2e). Most notably,  
98 while there exists an equivalent residue to YihQ H537 in SBG (H626) that interacts with the 3-  
99 hydroxyl in both complexes, significant differences are seen in the residues around the 4- and 6-  
100 positions. Within YihQ, clearly defined roles in sulfonate recognition can be ascribed to W304,  
101 R301 and Y508 (the latter through a bridging water molecule). However, within SBG the sugar  
102 hydroxymethyl group adopts a different geometry leading to significantly different interactions.  
103 SBG W432 adopts a roughly similar position to YihQ W304, but does not appear to be involved in  
104 hydrogen-bonding interactions with the ligand. SBG F601 sits in essentially the same place as YihQ

105 Y508, yet no ordered water molecule is present. No residue in SBG is located in an equivalent  
106 position to YihQ R301, suggesting that this residue is critical for sulfonate recognition. The major  
107 contributor to binding and recognition of the 6-hydroxyl group in SBG is D357, which makes a  
108 bidentate hydrogen-bonding interaction with the 4- and 6-hydroxyls of the acarbose valienamine;  
109 within YihQ the equivalent residue is Q288, which is rotated relative to SBG D357 and makes a  
110 single hydrogen bonding interaction with the 4-hydroxyl group of SQ.

111 Site-directed mutagenesis of YihQ R301 with a neutral Ala or negatively-charged Glu  
112 residue resulted in mutant proteins with no detectable activity against PNPSQ (Supplementary  
113 Table 1). In contrast, mutation to a basic Lys residue resulted in a variant with residual, but  
114 compromised activity. The raised  $K_M$  value of this mutant suggests that optimal binding of the  
115 sulfonate group requires not only a nearby positive charge, but may also involve a specific  
116 hydrogen-bonding interaction with arginine. The residual activity of the R301Q mutant may result  
117 from the ability of this residue to satisfy this hydrogen bond requirement of the sulfonate group.  
118 Mutagenesis of the W304 residue to Phe resulted in a complete loss of activity, demonstrating the  
119 critical importance of this residue in catalysis. None of these mutants displayed any detectable  
120 activity against PNPGlc. The YihQ Q288E variant was inactive towards PNPSQ, but possessed  
121 very weak, but detectable  $\alpha$ -glucosidase activity.

122 Sequence alignment of a range of GH31 members with  $\alpha$ -glucosidase,  $\alpha$ -xy lanase and  $\alpha$ -  
123 glucan lyase activities revealed that Q288, R301, W304 and Y508 are found only in YihQ  
124 homologues from predicted sulfoglycolysis gene clusters. Other putative SQases including  
125 *PpSQ1\_00094* and its homologues in predicted SQ Entner-Doudoroff gene clusters reveal  
126 conservation of R301, W304 and Y508, while Q288 was replaced by Glu. This supports the notion  
127 that these residues constitute a hallmark of SQase activity. Using YihQ as a query in a series of  
128 blastp searches targeted at diverse phyla, we identified putative GH family 31 members possessing  
129 these hallmark residues in diverse organisms across the tree of life (Supplementary Fig. 4). A  
130 phylogenetic comparison of these putative SQases with all functionally characterised GH family 31

131 enzymes revealed that the SQases comprise their own clade (Supplementary Fig. 5). These findings  
132 suggest that degradation of sulfoquinovosides is far more widespread than previously anticipated  
133 and alludes to widespread utilization of this ubiquitous source of carbon and sulfur.

134

#### 135 *Accession Codes*

136 Atomic coordinates and structure factors for the reported crystal structures have been deposited  
137 with the Protein Data Bank under accession codes 5AED (apo YihQ), 5AEG (YihQ-5Fido), and  
138 5AEE (D472N YihQ-PNPSQ).

139

#### 140 *Acknowledgements*

141 We thank S. G. Withers for the gift of 5-fluoro- $\beta$ -L-idopyranosyl fluoride and N. A. Williamson for  
142 technical assistance. This work was supported by grants from the UK Biotechnology and Biological  
143 Sciences Research Council, and the European Research Council (to G.J.D.), the Australian  
144 Research Council (to S.J.W.), the Ramaciotti Foundation and the Victorian Endowment for Science  
145 Knowledge and Innovation with additional support from the Australian Cancer Research  
146 Foundation and Victorian State Government Operational Infrastructure Support, NHMRC IRIISS  
147 grant 9000220 (to E.D.G.-B.). We thank the Diamond Light Source (Didcot (UK)) for access to  
148 beamlines IO4, IO4-1 and IO2 (proposal number mx-9948).

149

#### 150 *Author Contributions*

151 G.S. synthesized substrate and performed LC/MS analysis. G.S. and E.D.G.-B. cloned, expressed,  
152 mutagenized and purified enzyme, and performed kinetic analyses. Y.J. performed crystallographic  
153 studies and prepared the accompanying figures. Experiments were designed by G.J.D, S.J.W., and  
154 E.D.G.-B., who collectively wrote the paper.

155

#### 156 *References for Main Text*

- 157 1. Harwood, J.L. & Nicholls, R.G. The plant sulfolipid – a major component of the sulphur  
158 cycle. *Biochem. Soc. Trans.* **7**, 440–447 (1979).
- 159 2. Benning, C. Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol.  
160 *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 53–75 (1998).
- 161 3. Denger, K. et al. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the  
162 biogeochemical sulphur cycle. *Nature* **507**, 114–117 (2014).
- 163 4. Felux, A.K., Spitteller, D., Klebensberger, J. & Schleheck, D. Entner-Doudoroff pathway for  
164 sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. USA* **112**,  
165 E4298–305 (2015).
- 166 5. Shimojima, M. Biosynthesis and functions of the plant sulfolipid. *Prog. Lipid Res.* **50**, 234–  
167 239 (2011).
- 168 6. Martelli, H.L. & Benson, A.A. Sulfocarbohydrate metabolism. I. Bacterial production and  
169 utilization of sulfoacetate. *Biochim. Biophys. Acta* **93**, 169–171 (1964).
- 170 7. Roy, A.B., Hewlins, M.J., Ellis, A.J., Harwood, J.L. & White, G.F. Glycolytic breakdown of  
171 sulfoquinovose in bacteria: a missing link in the sulfur cycle. *Appl. Environ. Microbiol.* **69**,  
172 6434–6441 (2003).
- 173 8. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A.M. Sulfoquinovose  
174 degraded by pure cultures of bacteria with release of C<sub>3</sub>-organosulfonates: complete  
175 degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39–45 (2012).
- 176 9. Sugimoto, K., Sato, N. & Tsuzuki, M. Utilization of a chloroplast membrane sulfolipid as a  
177 major internal sulfur source for protein synthesis in the early phase of sulfur starvation in  
178 *Chlamydomonas reinhardtii*. *FEBS Lett.* **581**, 4519–4522 (2007).
- 179 10. Durham, B.P. et al. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc.*  
180 *Natl. Acad. Sci. USA* **112**, 453–457 (2015).
- 181 11. Shibuya, I. & Benson, A.A. Hydrolysis of α-sulphoquinovosides by β-galactosidase. *Nature*  
182 **192**, 1186–1187 (1961).

- 183 12. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The  
184 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–5  
185 (2014).
- 186 13. Okuyama, M., Mori, H., Chiba, S. & Kimura, A. Overexpression and characterization of  
187 two unknown proteins, YicI and YihQ, originated from *Escherichia coli*. *Protein Expr.*  
188 *Purif.* **37**, 170–179 (2004).
- 189 14. Andersson, L., Carriere, F., Lowe, M.E., Nilsson, A. & Verger, R. Pancreatic lipase-related  
190 protein 2 but not classical pancreatic lipase hydrolyzes galactolipids. *Biochim. Biophys. Acta*  
191 **1302**, 236–40 (1996).
- 192 15. Lee, S.S., Yu, S. & Withers, S.G.  $\alpha$ -1,4-Glucan lyase performs a trans-elimination via a  
193 nucleophilic displacement followed by a syn-elimination. *J. Am. Chem. Soc.* **124**, 4948–9  
194 (2002).
- 195 16. McCarter, J.D., and Withers, S.G. 5-Fluoro glycosides: A new class of mechanism based  
196 inhibitors of both  $\alpha$ - and  $\beta$ -glucosidases. *J. Am. Chem. Soc.* **118**, 241 (1996).
- 197 17. Quaroni, A. & Semenza, G. Partial amino acid sequences around the essential carboxylate in  
198 the active sites of the intestinal sucrase-isomaltase complex. *J. Biol. Chem.* **251**, 3250–3253  
199 (1976).
- 200 18. Okuyama, M. et al. Carboxyl group of residue Asp647 as possible proton donor in catalytic  
201 reaction of  $\alpha$ -glucosidase from *Schizosaccharomyces pombe*. *Eur. J. Biochem.* **268**, 2270–  
202 2280 (2001).
- 203 19. Davies, G.J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of  
204 glycosidases. *Acc. Chem. Res.* **45**, 308–316 (2012).
- 205 20. Speciale, G., Thompson, A.J., Davies, G.J. & Williams, S.J. Dissecting conformational  
206 contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13  
207 (2014).

208 21. Tagami, T. et al. Molecular basis for the recognition of long-chain substrates by plant a-  
209 glucosidases. *J. Biol. Chem.* **288**, 19296–303 (2013).

210

211 **Figure Legends for Main Text**

212 **Figure 1: *E. coli* YihQ is a sulfoquinovosidase that hydrolyzes SQDG to SQ.** (a) Biosynthesis  
213 and catabolism of SQ via SQDG. Biosynthesis occurs in photosynthetic organisms and catabolism  
214 occurs through sulfoglycolytic processes including the SQ Embden–Meyerhof–Parnas (SQ-EMP)  
215 sulfoglycolysis and SQ Entner–Doudoroff (SQ-ED) pathways, with the resulting 2,3-  
216 dihydroxypropane-1-sulfonate (DHPS) and sulfolactate (SL) undergoing biomineralization to  
217 sulfate in members of the bacterial community. (b) SQDG from spinach was incubated with YihQ  
218 and analyzed by LC/MS. Extracted ion chromatogram mass spectra showing normalized ion count  
219 of SQ ( $m/z = 243.0$ ) and SQDG ( $m/z 815.5$ ) before (blue) and after (red) incubation with YihQ. (c)  
220 Time-course  $^1\text{H}$  NMR spectra showing that YihQ hydrolyzes PNPSQ to release SQ with retention  
221 of anomeric configuration.

222

223 **Figure 2: Structural identification of SQ binding residues.** (a) Overview of the structure of *E.*  
224 *coli* YihQ D472N in complex with PNPSQ (in black). Shown are two molecules within the unit cell.  
225 The molecule at left is colored by sequence conservation using the putative GH31 SQases within  
226 the sequence alignment shown in Supplementary Fig. 4. (b) Active site of glycosyl-enzyme  
227 complex formed by wildtype YihQ and 5-fluoro- $\beta$ -L-idopyranosyl fluoride. (b) Active site of YihQ  
228 D472N in complex with PNPSQ. Electron density maps shown in blue mesh are  $\sigma^A$ -weighted  $2F_o -$   
229  $F_c$  contoured at  $1\sigma$  (0.27 and 0.25 electrons per  $\text{\AA}^3$ , respectively). (d) Cartoon showing major  
230 hydrogen-bonding and electrostatic interactions of PNPSQ with YihQ D472N, highlighting  
231 interaction with the sulfonate group. (e) Overlay of YihQ D472N complex with PNPSQ and sugar  
232 beet  $\alpha$ -glucosidase (SBG) complex with acarbose (PDB: 3W37). For clarity the catalytic  
233 nucleophiles, and the aglycons (except for the first carbon atom), have been omitted.

## 234 **Online Methods**

### 235 **Cloning the *yihQ* gene**

236 The *yihQ* gene (Supplementary Fig. 6) was amplified by colony PCR from *E. coli* K12 strain DH5 $\alpha$   
237 using the oligonucleotides YihQ-f and YihQ-r as primers (Supplementary Table 3). The 2053 base-  
238 pair product was digested with the restriction enzymes *NdeI* and *XhoI* and ligated into pET29b(+)  
239 (Novagen) that had been digested with the same enzymes and treated with alkaline phosphatase.  
240 The sequence of the gene in the resulting plasmid (pET29-YihQ) was confirmed by Sanger  
241 sequencing using the oligonucleotide primers T7, T7-term and YihQ-seq (Supplementary Table 3).

242

### 243 **YihQ expression and purification**

244 *Escherichia coli* BL21 (DE3) transformed with pET29-YihQ was grown in LB media with shaking  
245 (200 rpm) at 37 °C (50  $\mu\text{g ml}^{-1}$  kanamycin) until the culture reached an OD<sub>600</sub> of 0.8. The culture  
246 was cooled to 22 °C and isopropyl  $\beta$ -D-thiogalactopyranoside was added to a final concentration of  
247 100  $\mu\text{M}$  and shaking (200 rpm) was continued at this temperature for 16 h. Cells were harvested by  
248 centrifugation (17,000 g, 20 min, 4 °C), resuspended in PBS with a cocktail of protease inhibitors,  
249 lysed by sonication, and clarified by centrifugation (17,000 g, 20 min, 4 °C). The supernatant was  
250 filtered (0.22  $\mu\text{m}$ ) and then subjected to immobilized metal affinity chromatography. Fractions  
251 containing product (as determined by SDS-PAGE) were combined and further purified by size  
252 exclusion chromatography (GE Superdex 200 16/600) using 50 mM sodium phosphate, 150 mM  
253 NaCl, pH 7.5 buffer. The protein obtained was estimated to be >95% pure by Coomassie-stained  
254 SDS-PAGE. Protein concentration was determined by Bradford assay. The yield of YihQ was  
255 estimated to be 50 mg  $\text{l}^{-1}$ .

256

### 257 **Mutagenesis of YihQ**

258 Site directed mutagenesis of YihQ was accomplished using a PCR approach with the  
259 oligonucleotide primers listed in Supplementary Table 3. As an example, the YihQ-D405A

260 expression construct was assembled using two rounds of PCR. First, two amplification reactions  
261 were conducted using pET29-YihQ as template and oligonucleotide primer pairs: T7 / YihQ-  
262 D405A-r and T7-Term / YihQ-D405A-f. The products of these reactions were purified by agarose  
263 gel electrophoresis and mixed in equimolar amounts to serve as a template for a third amplification  
264 reaction using T7 and T7-term as primers. The PCR product was cloned into pET29b, the sequence  
265 verified, and expressed and purified as for wildtype YihQ.

266

### 267 **LC/MS analysis of YihQ-digested SQDG and SQGro**

268 SQDG (>95%, Indofine, 20  $\mu$ l, 1 mg ml<sup>-1</sup> in CHCl<sub>3</sub>:MeOH, 1:1) in 200 mM NH<sub>4</sub>CO<sub>3</sub>, 5% DMSO,  
269 pH 6.2 (90  $\mu$ l) was treated with YihQ (10  $\mu$ l, 1.36  $\mu$ M) and incubated overnight at 23 °C. SQGro  
270 was prepared by incubating SQDG (20  $\mu$ l, 1 mg ml<sup>-1</sup> in CHCl<sub>3</sub>:MeOH) with NaOMe in MeOH (1  
271  $\mu$ l, 100 mM) at 23 °C for 3 h prior to the addition of buffer (90  $\mu$ l, 200 mM NH<sub>4</sub>CO<sub>3</sub>, pH 6.2) and  
272 YihQ (10  $\mu$ l, 1.36  $\mu$ M) and incubation of the mixture at 23 °C overnight. Both reaction mixtures  
273 were lyophilised three times from H<sub>2</sub>O, resuspended in CHCl<sub>3</sub>:MeOH:H<sub>2</sub>O (4:6:1) and centrifuged  
274 at 14,000 rpm for 10 min. The supernatant was analysed by LC/MS (Agilent electrospray  
275 ionization-time of flight mass spectrometer, negative mode), using an Agilent HILIC plus (3.5  $\mu$ m,  
276 2.1 x 100 mm) column. Control reactions without the addition of YihQ had no detectable quantities  
277 of SQ, indicating that the rate of non-enzymatic hydrolysis is negligible under these conditions.

278

### 279 **Stereochemical outcome of YihQ-catalyzed hydrolysis**

280 The YihQ catalyzed hydrolysis of PNPSQ (for synthesis see Supplementary Notes) was monitored  
281 by <sup>1</sup>H NMR spectroscopy using a 500 MHz instrument. A solution of YihQ in buffered D<sub>2</sub>O (0.25  
282 ml, 0.1 mM in 50 mM sodium phosphate, 250 mM NaCl, pH 7.2) was added to a solution of  
283 PNPSQ (4.0 mg, 9.9  $\mu$ mol) in buffered D<sub>2</sub>O (0.75 ml, 50 mM sodium phosphate, 250 mM NaCl,  
284 pH 7.2) at 22 °C. <sup>1</sup>H NMR spectra were acquired at different time points (t = 0, 7 min, 24 min and  
285 24 h).

286

### 287 **pH dependence of YihQ activity**

288 The  $k_{\text{cat}}/K_M$  values for YihQ were measured for PNPSQ hydrolysis using the substrate depletion  
289 method in 50 mM citrate/phosphate buffer, 150 mM NaCl at a range of pH values (4.0, 4.5, 5.0, 5.5,  
290 6.0, 7.5, 8.0) at 23 °C (Supplementary Fig. 7b). Reactions were initiated by the addition of 13.6 nM  
291 YihQ to PNPSQ (10  $\mu\text{M}$ ) in buffer, and aliquots quenched at different time points into glycine  
292 buffer (1 M, pH 10.0) and absorbance measured using a UV/vis spectrophotometer ( $\lambda = 405 \text{ nm}$ ).  
293 The extinction coefficient for 4-nitrophenolate under the assay conditions was determined to be  
294  $11500 \text{ M}^{-1} \text{ cm}^{-1}$ .  $k_{\text{cat}}/K_M$  and  $\text{p}K_a$  values were calculated using the Prism 6 software package  
295 (Graphpad Scientific Software). All assays were repeated in triplicate.

296

### 297 **Enzyme kinetics**

298 Kinetic analysis of wildtype YihQ and the mutants (D405A, D405N, D472A, D472N, R301A,  
299 R301E, R301K, R301Q, Q288E, W304F) was performed using PNPSQ and PNPGlc, using a  
300 UV/visible spectrophotometer to measure the release of the 4-nitrophenolate ( $\lambda = 405 \text{ nm}$ ). Assays  
301 were carried out in 50 mM sodium phosphate, 150 mM NaCl, pH 7.2 at 23 °C using 13.6 nM YihQ-  
302 wt or 17.4–26.2 nM of the mutant YihQ (D405A, D405N, D472A, D472N, R301A, R301E, R301K,  
303 R301Q, Q288E, W304F) enzyme at substrate concentrations ranging from 0.4  $\mu\text{M}$  to 10 mM. The  
304 extinction coefficient for 4-nitrophenolate under the assay conditions was determined to be  $9026 \text{ M}^{-1}$   
305  $\text{cm}^{-1}$ . Kinetic parameters were calculated using the Prism 6 software package (Graphpad Scientific  
306 Software). All assays were repeated in triplicate.

307

### 308 **X-ray crystallography and structure solution**

309 Well-diffracting crystals of YihQ were obtained by mixing 25  $\text{mg ml}^{-1}$  protein stock with equal  
310 volume of precipitant composed of 50–60% (v/v) 2-methyl-2,4-pentanediol, 0.10–0.15 M  $\text{CaCl}_2$ ,  
311 and bis-tris, pH 6.5 after 3–4 days at 20 °C using the sitting drop vapor diffusion method.

312 Selenomethionine (Se-Met) labeled YihQ was produced in *E. coli* BL21(DE3), a methionine  
313 prototroph, following the PASM-5052 auto-induction protocol.<sup>22</sup> Se-Met YihQ and D472N YihQ  
314 were crystallized under the same conditions as for the native enzyme. The D472N-PNPSQ and wt-  
315 5FIdo complexes were produced using the soaking method. Drops containing crystals of wildtype  
316 or D472N mutant were supplemented with 10 mM PNPSQ or 5FIdoF in the same precipitant  
317 solution for 10 min at 20 °C before collecting and freezing crystals. No cryoprotectant was used for  
318 the crystals before they were flash frozen in liquid nitrogen. Diffraction data were collected at 100  
319 K on beamline I02, I04 or I04-I of the Diamond Light Source and were processed using the *xia2*  
320 implementation of XDS<sup>23</sup> and programs from CCP4 suite.<sup>24</sup> Experimental phasing was performed  
321 by single wavelength anomalous diffraction methods at a wavelength of 0.91742 Å for crystals  
322 from Se-Met labeled YihQ. The statistics of the data processing and structure refinement are listed  
323 in Supplementary Table 2.

324

## 325 **Graphics**

326 The amino acid conservation scores for YihQ in Fig. 2a were calculated using the ConSurf Server.

327 <sup>25,26</sup> The figures for the 3D protein structures were produced using Pymol<sup>27</sup> and CCP4MG.<sup>28</sup>

328

## 329 **Sequence alignments**

330 Amino acid sequences for putative SQases were identified through a series of protein-protein

331 BLAST (blastp) searches of the NCBI non-redundant protein sequences database targeting the

332 different kingdoms/phyla of life using *E. coli* YihQ as the query sequence. Representative protein

333 sequences from different phyla were selected for alignment using Clustal Omega. The genes

334 encoding putative eukaryotic SQases possessed multiple intronic regions, providing confidence that

335 they were not artefacts arising from bacterial contamination of their respective genome assemblies.

336 Global phylogenies of the putative SQases were computed using Phylogeny.fr, which performed

337 initial tree generation with BIONJ and optimization with PhyML 3.0 using the LG substitution

338 model ( $\gamma$  8, NNI) and branch support by the aLRT-SH method with bootstrap (100) analyses.<sup>29</sup>  
339 Phylogenetic tree diagrams were constructed using TreeDyn 198.3, collapsing branches with  
340 bootstrap values less than 0.9.

341

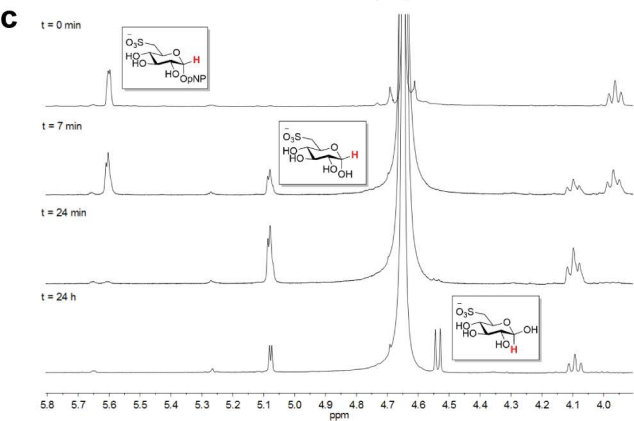
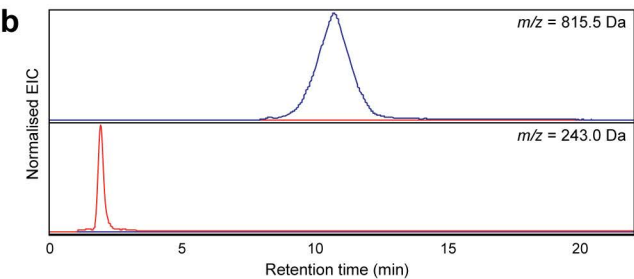
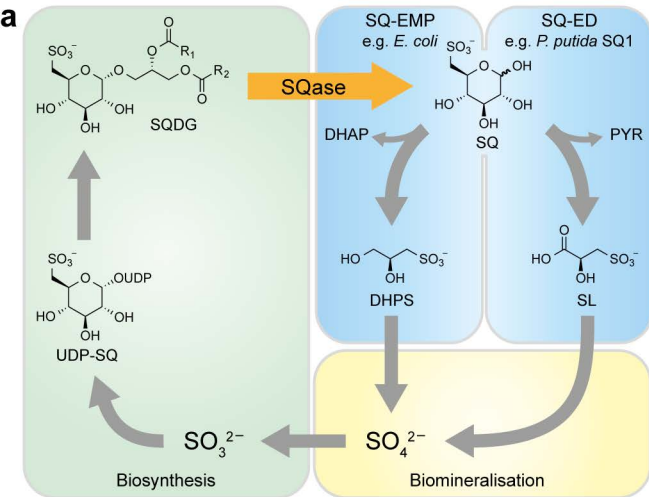
#### 342 ***Methods-only References***

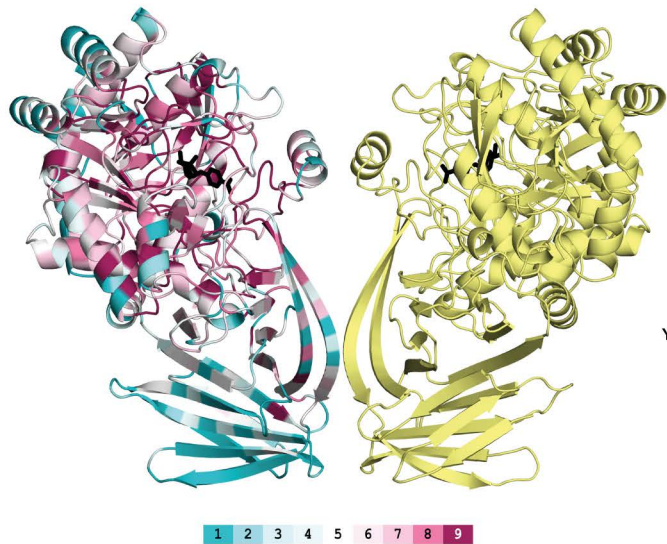
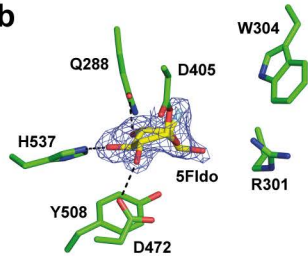
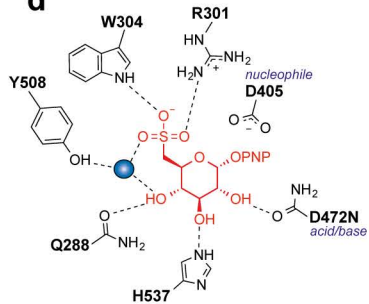
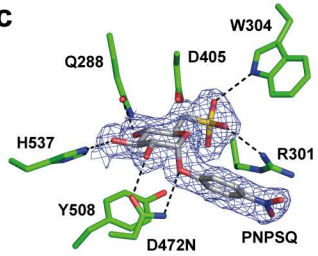
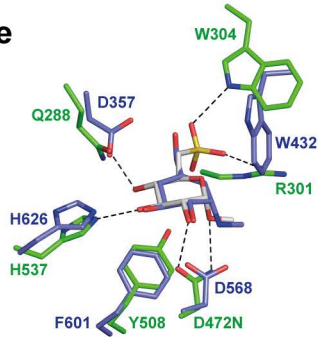
- 343 22. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein*  
344 *Expr. Purif.* **41**, 207–34 (2005).
- 345 23. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J.*  
346 *Appl. Crystallogr.* **43**, 186–190 (2010).
- 347 24. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*  
348 **50**, 760–3 (1994).
- 349 25. Celniker, G. et al. ConSurf: Using evolutionary data to raise testable hypotheses about  
350 protein function. *Isr. J. Chem.* **53**, 199–206 (2013).
- 351 26. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating  
352 evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic*  
353 *Acids Res.* **38**, W529–33 (2010).
- 354 27. The PyMOL Molecular Graphics System. Version 1.7.4 edn (Schrödinger, LLC).
- 355 28. McNicholas, S., Potterton, E., Wilson, K.S. & Noble, M.E.M. Presenting your structures:  
356 the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **67**, 386–394 (2011).
- 357 29. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic*  
358 *Acids Res.* **36**, W465–9 (2008).

359

#### 360 ***Competing financial interests***

361 The authors declare no competing financial interests.



**a****b****d****c****e**

## **SUPPLEMENTARY INFORMATION**

### **YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids**

Gaetano Speciale<sup>1,†</sup>, Yi Jin<sup>2,†</sup>, Gideon J. Davies<sup>2</sup>, Spencer J. Williams<sup>1</sup>  
and Ethan D. Goddard-Borger<sup>3,4,\*</sup>

<sup>1</sup> School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010 (Australia)

<sup>2</sup> Department of Chemistry, University of York, Heslington, York, YO10 5DD (UK)

<sup>3</sup> ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052 (Australia)

<sup>4</sup> Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010 (Australia)

† These authors contributed equally to this work

\* Correspondence should be addressed to E.D.G.-B. (goddard-borger.e@wehi.edu.au).

## SUPPLEMENTARY RESULTS

### Supplementary Tables

**Supplementary Table 1: Kinetic parameters for the hydrolysis of PNPSQ and PNPGlc catalyzed by YihQ and mutated variants.**

Enzyme	Substrate	$K_M$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )
YihQ (WT)	pNPSQ	0.22±0.03	14.3±0.4	(6.4±1.0)×10 <sup>4</sup>
	pNPGlc	n/d	n/d	n/d
YihQ (D405A)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (D405N)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (D472A)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (D472N)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (R301A)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (R301E)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d
YihQ (R301K)	pNPSQ	-	-	0.71±0.02 <sup>a</sup>
	pNPGlc	n/d	n/d	n/d
YihQ (R301Q)	pNPSQ	2.9±0.4	(3.1±0.1)×10 <sup>-3</sup>	1.1±0.2
	pNPGlc	n/d	n/d	n/d
YihQ (Q288E)	pNPSQ	n/d	n/d	n/d
	pNPGlc	6.7±2.0	(7.2±0.5)×10 <sup>-5</sup>	(1.1±0.4)×10 <sup>-2</sup>
YihQ (W304F)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d

<sup>a</sup> saturation not reached

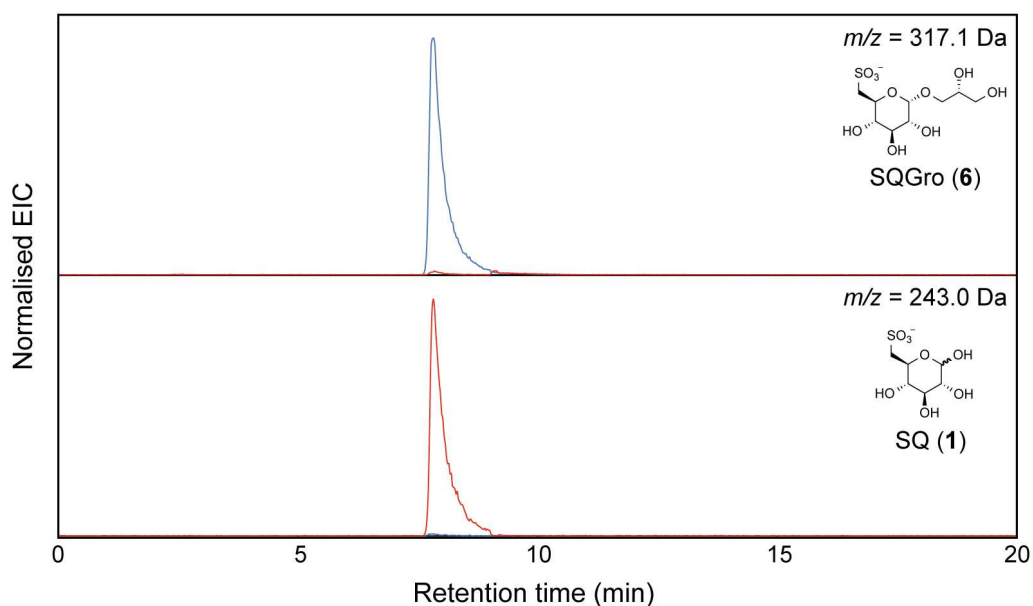
**Supplementary Table 2: X-ray data collection, processing and refinement statistics.**

	wt native	wt-5FIdo	D472N-pNPSQ
<b>Data collection</b>			
Space group	P21	P21	P21
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.7, 113.2, 111.7	78.7, 112.6, 111.9	78.0, 107.6, 103.3
<i>a</i> , <i>b</i> , <i>g</i> (°)	90.0, 109.2, 90.0	90.0, 109.5, 90.0	90.0, 107.9, 90.0
Resolution (Å)	48.1-1.91 (1.94-1.91)	49.7-1.85 (1.88- 1.85)	53.8-1.85 (1.90-1.85)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.073 (0.41)	0.054 (0.621)	0.050 (0.878)
<i>I</i> / <i>sI</i>	7.9 (2.0)	13.5 (2.1)	13.5 (1.2)
Completeness (%)	96.8 (99.2)	98.6 (95.7)	87.7 (72.9)
Redundancy	2.8 (2.9)	4.0 (3.7)	3.4 (2.7)
<b>Refinement</b>			
Resolution (Å)	48.1-1.91 (1.94-1.91)	49.7 - 1.85 (1.88-1.85)	53.8-1.85 (1.90-1.85)
No. reflections	391505 (20452)	617752 (27747)	416164 (19733)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.1681 / 0.2013	0.1601 (0.1935)	0.1615 (0.2002)
No. atoms			
Protein	10851	10769	10741
Ligand/ion	27	41	60
Water	655	724	602
<i>B</i> -factors			
Protein	33.5	35.5	32.3
Ligand/ion	n/a	40.3	30.1
Water	40.7	38.0	37.7
R.m.s. deviations			
Bond lengths (Å)	0.019	0.019	0.019
Bond angles (°)	1.81	1.86	1.87

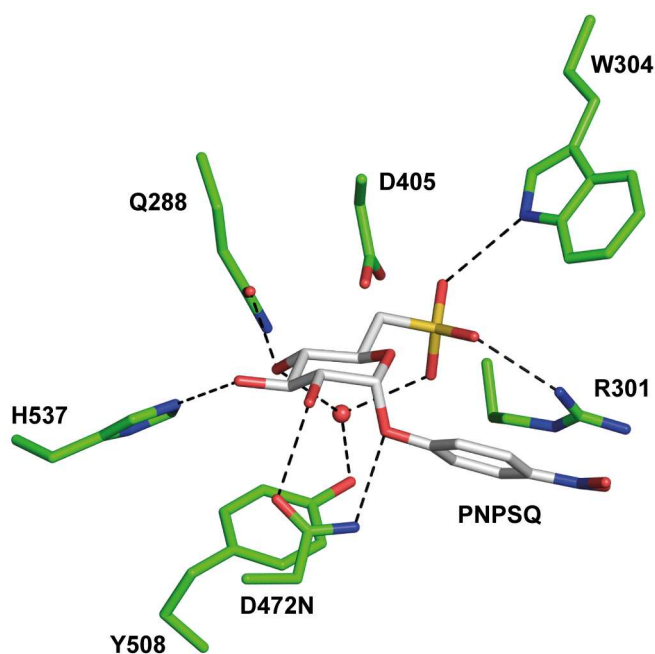
**Supplementary Table 3: Primers for cloning and mutagenesis.**

<b>primer</b>	<b>sequence</b>
YihQ-f	ATATACATATGGATACGCCACGTCCAC
YihQ-r	TGGTGCCTCGAGGATGCTTTTTAACGACGCGAAC
YihQ-seq	ACTTCGACTTTAGTGCCC
T7	TAATACGACTCACTATAGGG
T7-term	GCTAGTTATTGCTCAGCGG
YihQ-D405A-f	GGCTGCGGCGGCTGGATGGCTGCCTTCGGCGAGTATCTGCCACC
YihQ-D405A-r	GGTGGGCAGATACTCGCCGAAGGCAGCCATCCAGCCGCCGCAGCC
YihQ-D405N-f	GGCTGCGGCGGCTGGATGGCTAACTTCGGCGAGTATCTGCCACC
YihQ-D405N-r	GGTGGGCAGATACTCGCCGAAGTTAGCCATCCAGCCGCCGCAGCC
YihQ-D472A-f	TCCACCATGATGTGGGCGGGCGCCAGAACGTCGACTGGAGTCTCGAC
YihQ-D472A-r	GTCGAGACTCCAGTCGACGTTCTGGGCGCCCGCCACATCATGGTGGA
YihQ-D472N-f	TCCACCATGATGTGGGCGGGCAACCAGAACGTCGACTGGAGTCTCGAC
YihQ-D472N-r	GTCGAGACTCCAGTCGACGTTCTGGTTGCCCGCCACATCATGGTGGA
YihQ-R301A-f	CGTATGACCTCTTTTGGCAAAGCCGTGATGTGGAAGTGGAAAGTGG
YihQ-R301A-r	CCACTTCCAGTTCACATCACGGCTTTGCCAAAAGAGGTCATACG
YihQ-R301K-f	CGTATGACCTCTTTTGGCAAAGGTTGATGTGGAAGTGGAAAGTGG
YihQ-R301K-r	CCACTTCCAGTTCACATCACCTTTTGGCAAAGAGGTCATACG
YihQ-R301E-f	CGTATGACCTCTTTTGGCAAAGAAGTGTGATGTGGAAGTGGAAAGTGG
YihQ-R301E-r	CCACTTCCAGTTCACATCACTTCTTTGCCAAAAGAGGTCATACG
YihQ-R301Q-f	CGTATGACCTCTTTTGGCAAACAGGTGATGTGGAAGTGGAAAGTGG
YihQ-R301Q-r	CCACTTCCAGTTCACATCACCTGTTTGGCAAAGAGGTCATACG
YihQ-Q288E-f	GAAGGTCAACGGCATCTGGGCGGAGGACTGGTCCGGTATTCGTATGACCTC
YihQ-Q288E-r	GAGGTCATACGAATACCGGACCAGTCTCCGCCAGATGCCGTTGACCTTC
YihQ-W304F-f	CCTCTTTTGGCAAACGCGTGATGTTAACTGGAAGTGGAAACAGCGAAAAC
YihQ-W304F-r	GTTTTGCTGTTCCACTTCCAGTTAAACATCACGCGTTTGGCAAAGAGG

## Supplementary Figures

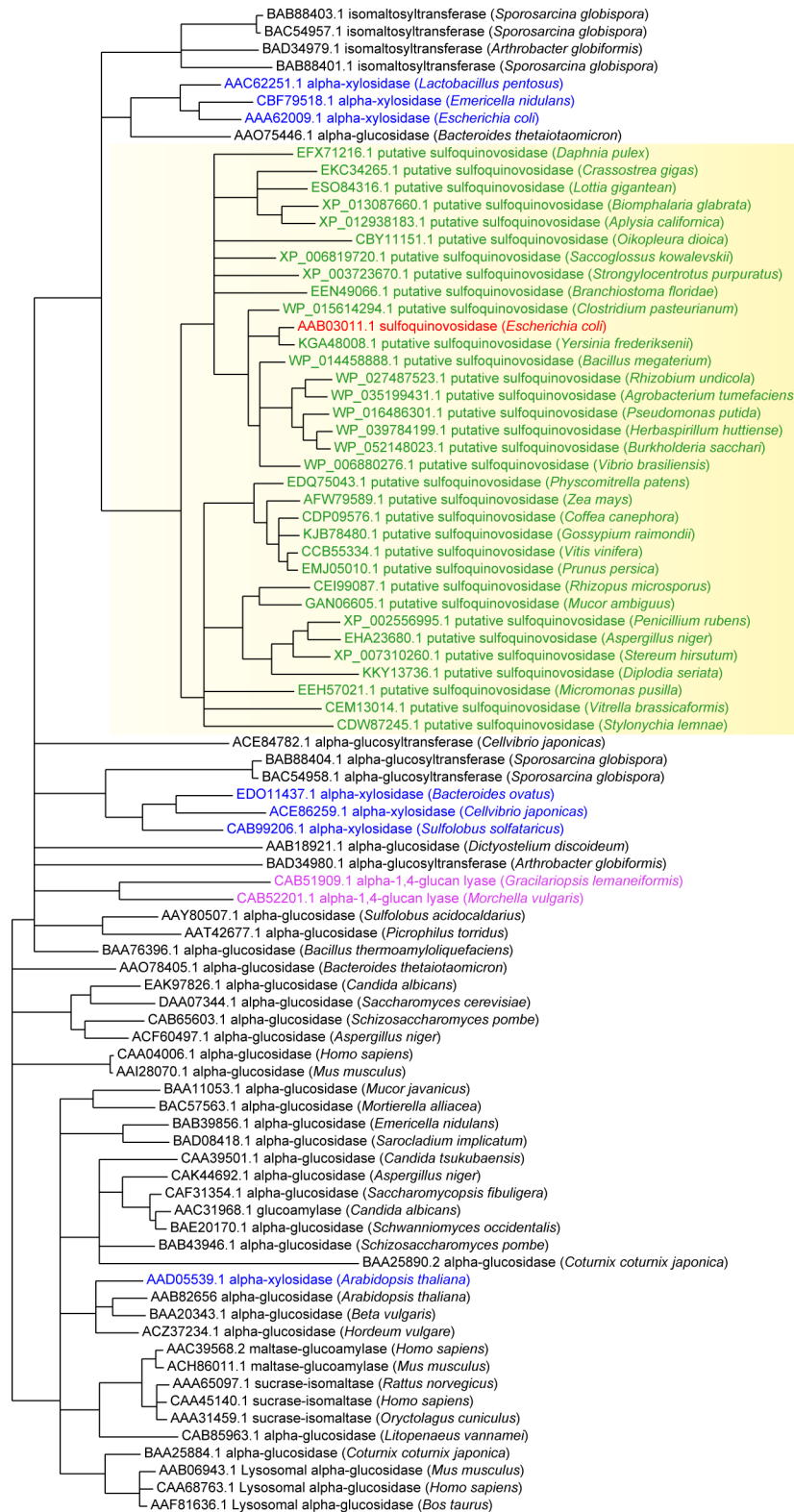


**Supplementary Figure 1: YihQ acts on 1-sulfoquinovosylglycerol (SQGro).** SQGro, prepared by Zemplén transesterification on SQDG, was incubated with YihQ and analyzed by LC/MS. Extracted ion chromatogram mass spectra showing normalized ion count of SQ ( $m/z = 243.0$ ) and SQGro ( $m/z 317.1$ ) before (blue) and after (red) incubation with YihQ.



**Supplementary Figure 2: Detailed view of active site of YihQ D472N in complex with PNPSQ.**





2.

**Supplementary Figure 5: Phylogenetic analysis of GH family 31 enzymes reveal SQases constitute a clade.** The phylogenies of functionally-characterized GH family 31 enzymes and putative SQases from a diverse array of species were computed using PhyML 3.0. Enzymes are colored according to activity (glucosidases in black, xylosidases in blue, glucan lyases in pink, sulfoquinovosidases in red, and putative sulfoquinovosidases in green).

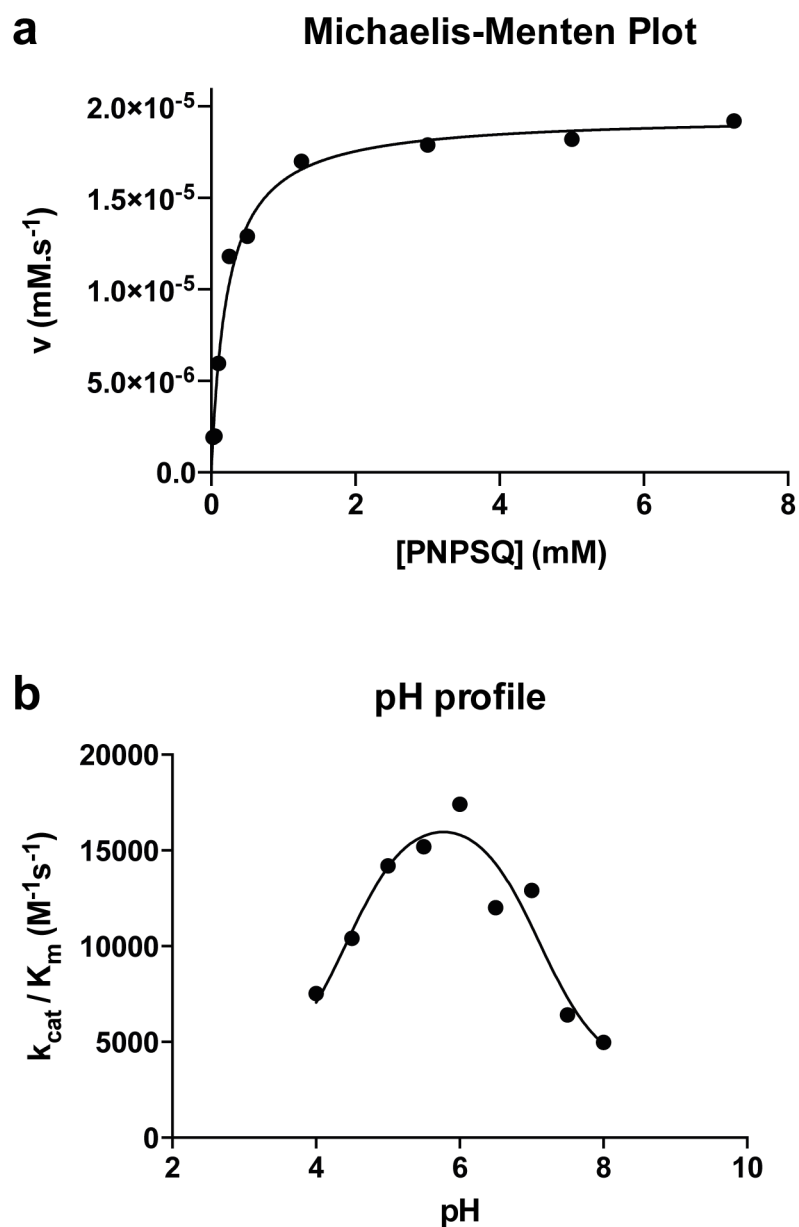
**a**

```
>YihQ-His6
ATGGATACGCCACGTCCACAGTTATTAGATTTTCAATTCATCAGAATAACGACAGTTTACCCTACATTTTCAACAACG
TCTTATTTTAAACCCATAGCAAAGATAATCCTTGTGTTATGGATTGGCTCAGGTATAGCGGATATCGATATGTTCCGCGGTA
ATTTTCAGCATTAAAGATAAATACAGGAGAAAATGCGCTTACCGACGCATCGTCAGCCAGTCACCGGATGGTTGGTTA
ATTCAATTCAGCCGTGGTTCTGACATTAGCGCCACGCTGAATATCTCTGCCGACGATCAGGGGCGTTTATTGCTGGAAC
ACAAAACGACAACCTTAAACACAACCGTATCTGGCTGCGCTTGGCCGCTCAACCAGAGGACCATATCTACGGCTGCGGG
AACAGTTTTCTACTTCGATCTGCGTGGCAAACCGTTCCCGCTATGGACCAGTGAACAAGGCGTTGGTCGCAACAAACAA
ACCTATGTACCTGGCAGGCCGACTGCAAAGAAAATGCGGGCGGCGACTATTACTGGACTTTCCTCCACAGCCTACGTT
TGTCAGCACGCAGAAGTATTACTGCCATGTTGATAACAGTTGCTATATGAACCTCGACTTTAGTGCCCCGGAATACCATG
AACTGGCGCTGTGGGAAGACAAGCAACGCTGCGTTTTGAATGTGCTGACACATACATTTCCCTGCTGGAAAAATTAACC
GCCCTGCTGGGACGCCAGCCAGAAGTCCCGACTGGATTATGACGGAGTAACGCTCGGCATTCAGGGCGGACGGAAGT
GTCCAGAAGAAAACCTGGACACCATGCGTAACGCGGGCGTGAAGTCAACGGCATCTGGGCGCAGGACTGGTCCGGTATTC
GTATGACCTCTTTTGGCAAACGCGTGTGTTGGAAGTGAAGTGAACAGCGAAAACCTACCCGCAACTGGATTACGCAAT
AAGCAGTGGAAATCAGGAGGGCGTGCAGTTCCTGGCCATATCAACCCGATGTTGCCAGCGATAAAGATCTCTCGAAGA
AGCGGCACAACACGGCTATCTGGCAAAGATGCTCTGGCGGTGACTATCTGGTGGAGTTTGGCGAGTTTACGGCGGG
TTGTGATCTCCTACTAATCCAGAAGCCTACGCCCTGGTTCAAGGAAGTGTCAAAAAGAACATGATTGAACTCGGCTGCGGC
GGCTGGATGGCTGACTTCGGCGAGTATCTGCCACCGACACGTAATGTCATAACGGCGTCAGTGCCGAAATTAATGCATA
CGCCTGGCCTGCGCTGTGGGCGAAGTGAACACGAAGCCCTTGAAGAAAACGGCAAGCTCGGCGAGATCCTTTTCTTTA
TGCGCGCCGGTCTTACCGGTAGCCAGAAATATCTCCACCATGATGTGGCGGGCGACCAGAACGTCGACTGGAGTCTCGAC
GATGGCCTGGCGTGGTGTGTCGCCGGCGCGTGTGCTGGCAATGACCGGACATGGCCTGCACCACAGCGACATTTGGCGG
TTACACCACCCGTGTTTGAATGAAGCGCAGCAAAGAGTGTGCTGCGCTGGTGCATTTACAGCCCTTACCGCGATGA
TGCGCACCCACGAAGGTAACCGTCTGGCGACAACCTGGCAGTTTACGGCGACGCAGAAACCATCGCCCATTTCCCGCGT
ATGACCCCGCTTACCACCCCTGAAACCTTACCTGAAAGAGGCGCTGCGCGTGAATGCGAAGTCCCGGCTGCGGTTAT
GCGCCCGCTGTTTCTGCATTACGAAGACGATGCGCACACTTACCCCTGAAATATCAGTACCTGTTAGGTGCGGACATTC
TGGTTCGCTCCGCTGCATGAAGAGGCCGTAGCGACTGACGCTCTATCTGCCGAGGATAAATGGGTCCACGCTGGACG
GGTGAAGCGTCCGGGGCGGGGAGTTACCGTAAATGCGCCCATCGCAAGCCCGCGGCTTTTATCGCGCCGATAGCGA
ATGGGCGGCACGTTCCTCGCTCGTTAAAAAGCATCCTCGAGCACCAACCACCACCACCTGA
```

**b**

```
>YihQ-His6
MDTPRPQLLDFQFHQNDSFTLHFQORLILTHSKDNPLWIGSGIADIDMFRGNFSIKDK
LQEKIALTDIVSQSPDGWLIFHSRGSDISATLNISSADDQGRLLLELQNDNLNHNRIWLR
LAAQPEDHIYCGEQFSYFDLRGKPFPLWTSEQGVGRNKQTYVTWQADCKENAGGDYIYWT
FFPQPTFVSTQKYCHVDNSCYMNFDFSAPEYHELALWEDKATLRFECADTYISLLEKLT
ALLGRQPELPDWIYDGVTLGIQGGTEVCQKKLDTMRNAGVKVNGIWAQDWSGIRMTSFGK
RVMWNWKNSENYPQLDSRIKQWNQEVQFLAYINPYVASDKDLCEEAQHGYLAKDASG
GDYLVFEGEFYGGVVDLTNPEAYAWFKEVIKKNMIELGCGWMADFGEYLPDFTYLHNGV
SAEIMHNAWPALWAKCNYEAELETGKLGELFFMRAGSTGSQKYSTMMWAGDQNVDSLD
DGLASVVPAAALSLAMTGHGLHSDIGYTTLFFEMKRSKELLLRWCDFAFTPMRTHGEN
RPGDNWQFDGDAETIAHFARMTTVFTTLKPYLKEAVALNAKSGLPVMRPLFLHYEDDAHT
YTLKYQYLLGRDILVAPVHEEGRSDWTLYLPEDNWWHAWTGEAFRGGEVTVNAPIGKPPV
FYRADSEWAALFASLKSILEHHHHHH*
```

**Supplementary Figure 6: YihQ sequence details. (a)** The nucleotide sequence encoding His<sub>6</sub>-tagged YihQ and **(b)** the amino acid sequence of His<sub>6</sub>-tagged YihQ used in this study.



**Supplementary Figure 7: Enzymology plots for YihQ-wt against PNPSQ. (a)** Michaelis-Menten plot and **(b)** pH profile for YihQ-wt activity on PNPSQ.

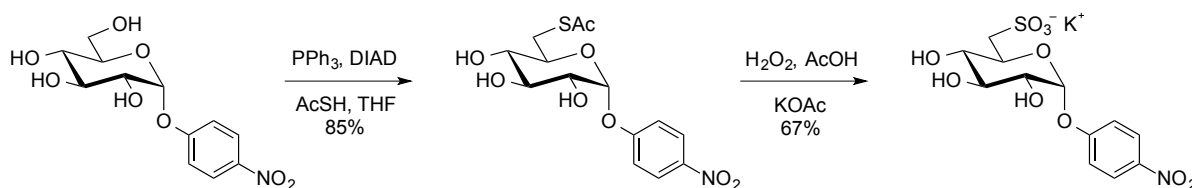
## Supplementary Notes

### Chemical synthesis of PNPSQ

#### General

All chemical reagents were purchased from Sigma-Aldrich at >95% purity unless otherwise stated.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded using a 400 MHz instrument. All signals were referenced to solvent peaks ( $d_4$ -MeOH:  $\delta$  3.49 ppm for  $^1\text{H}$  or 49.0 ppm for  $^{13}\text{C}$ ). TLC analysis used aluminium backed Merck Silica Gel 60 F<sub>254</sub> sheets, detection was achieved using UV light, 5%  $\text{H}_2\text{SO}_4$  in MeOH, or ceric ammonium molybdate solution with heating as necessary. Flash chromatography was performed using Geduran silica gel according to the method of Still *et al.*<sup>1</sup> Dry THF was obtained from a dry solvent apparatus (Glass Contour of SG Water, Nashua, U.S.A.).<sup>2</sup> Melting points were obtained using a hot-stage microscope.  $[\alpha]_D$  values are given in deg  $10^{-1} \text{ cm}^2 \text{ g}^{-1}$ .

#### Synthetic scheme



#### 4-Nitrophenyl 6-S-acetyl-6-thio- $\alpha$ -D-glucopyranoside

A solution of 4-nitrophenyl  $\alpha$ -D-glucopyranoside (0.10 g, 0.33 mmol) and thioacetic acid (28  $\mu\text{l}$ , 0.40 mmol) in dry THF (2.0 ml) at 0  $^\circ\text{C}$  was added to a mixture of  $\text{PPh}_3$  (0.10 g, 0.40 mmol) and DIAD (80  $\mu\text{l}$ , 0.40 mmol) in THF (1.0 ml) at 0  $^\circ\text{C}$ . The reaction mixture was allowed to warm to r.t. and stirred overnight. The mixture was concentrated and the residue was purified by flash chromatography (EtOAc/hexane, 40-100%), to afford the thioacetate (0.10 g, 0.28 mmol, 85%) as a white crystalline solid. A small portion was recrystallized (m.p. 122-123 $^\circ\text{C}$ , EtOAc);  $[\alpha]_D^{23} +1.2^\circ$  (c 0.55,  $\text{CH}_3\text{OH}$ );  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  2.21 (3 H, s,  $\text{CH}_3$ ), 2.93 (1 H, dd,  $J_{6,6} = 13.9$ ,  $J_{5,6} = 8.6$  Hz, H6a), 3.27 (1 H, dd,  $J_{3,4} = 9.3$ ,  $J_{4,5} = 9.3$  Hz, H4), 3.51 (1 H, dd,  $J_{5,6} = 2.7$  Hz, H6b), 3.66–3.58 (2 H, m, H2,5), 3.86–3.77 (1 H, dd,  $J_{2,3} = 9.3$  Hz, H3), 5.64 (1 H, d,  $J_{1,2} = 3.7$  Hz, H1), 7.34–7.28 (2 H, m, Ar), 8.29–8.23 (2 H, m, Ar);  $^{13}\text{C}$  NMR (101 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  30.25 ( $\text{CH}_3$ ), 31.80 (C6), 73.03 (C2), 73.31 (C5), 74.56 (C3), 74.67 (C4), 98.76 (C1), 118.08, 126.54, 143.91, 163.12 (4C, Ar), 196.68 (C=O); HRMS (ESI)<sup>+</sup>  $m/z$  360.0760 [ $\text{C}_{14}\text{H}_{17}\text{NO}_8\text{S}$  (M + H)<sup>+</sup> requires 360.0753].

### **Potassium 4-nitrophenyl 6-deoxy-6-sulfonato- $\alpha$ -D-glucopyranoside (PNPSQ)**

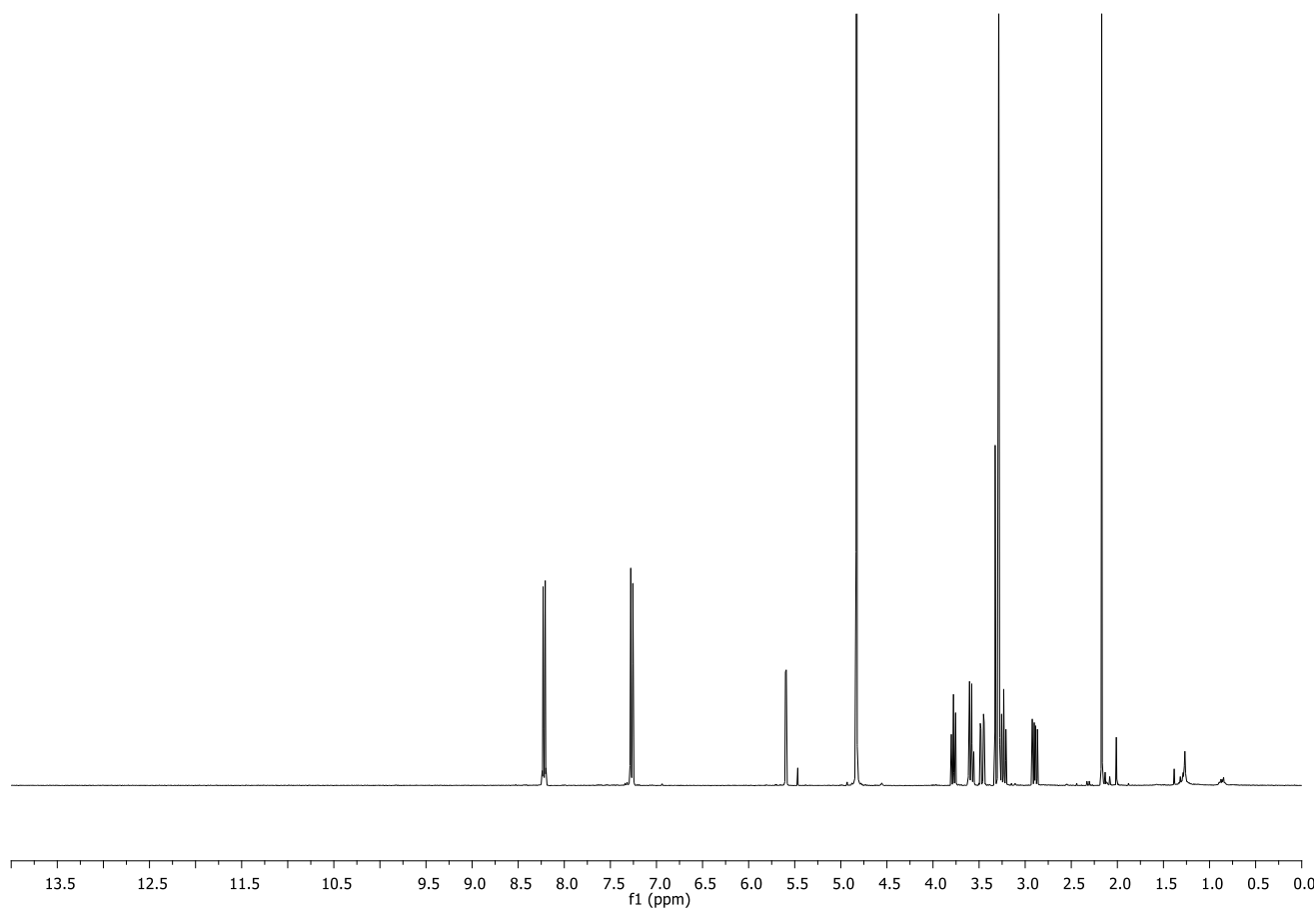
30% H<sub>2</sub>O<sub>2</sub> (0.70 ml) was added to a solution of 4-nitrophenyl 6-S-acetyl-6-thio- $\alpha$ -D-glucopyranoside (50 mg, 0.14 mmol) and KOAc (15 mg, 0.16 mmol) in glacial AcOH (0.70 ml). The mixture was stirred at 50 °C for 24 h then diluted with water and quenched by the addition of PPh<sub>3</sub> in Et<sub>2</sub>O (2.0 M, 3.0 ml). The aqueous phase was separated, and the organic phase was extracted twice with water. The combined aqueous phases were concentrated and the residue purified by flash chromatography (EtOAc/MeOH/H<sub>2</sub>O, 19:2:1→7:2:1) and C<sub>18</sub> reversed phase chromatography (H<sub>2</sub>O/CH<sub>3</sub>CN, 95:5), affording PNPSQ (38 mg, 0.094 mmol, 67%) as a white solid.  $[\alpha]_D^{23} +1.3^\circ$  (*c* 0.91, CH<sub>3</sub>OH); <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  3.01 (1 H, dd,  $J_{6,6} = 14.4$ ,  $J_{5,6} = 8.5$  Hz, H6a), 3.27 (1 H, dd,  $J_{3,4} = 9.7$ ,  $J_{4,5} = 9.0$  Hz, H4), 3.36 (1 H, dd,  $J_{5,6b} = 2.5$  Hz, H6b), 3.64 (1 H, dd,  $J_{2,3} = 9.8$ ,  $J_{1,2} = 3.7$  Hz, H2), 3.86 (1 H, dd, H3), 4.16 (1 H, ddd, H5), 5.51 (1 H, d, H1), 7.47–7.39 (2 H, m, Ar), 8.26–8.18 (2 H, m, Ar); <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD)  $\delta$  53.09 (C6), 69.78 (C5), 71.59 (C2), 73.02 (C3), 73.43 (C4), 98.52 (C1), 117.51, 125.04, 142.61, 162.65 (4C, Ar); HRMS (ESI)<sup>-</sup> *m/z* 364.0340 [C<sub>12</sub>H<sub>14</sub>NO<sub>10</sub>S (M – H)<sup>-</sup> requires 364.0343].

### **References**

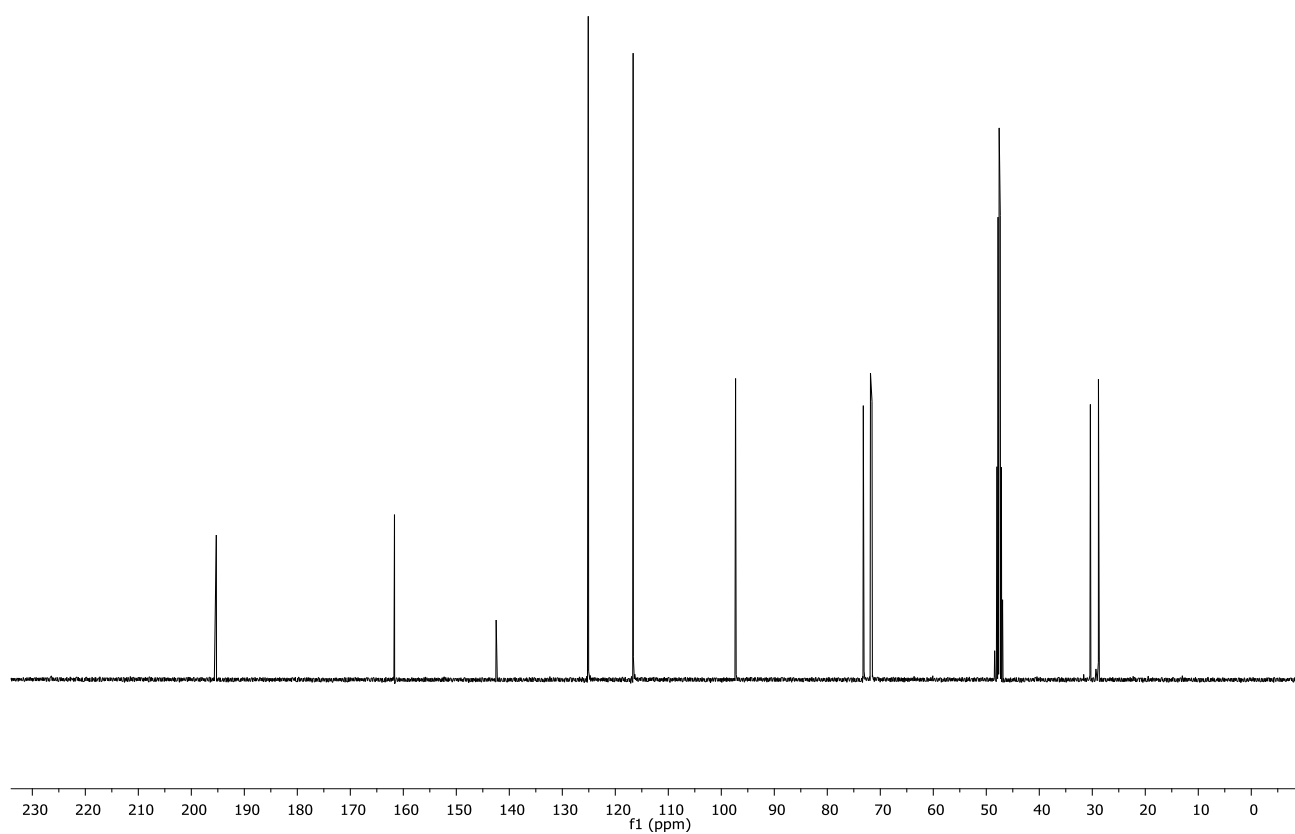
- 1 Still, W. C., Kahn, M. & Mitra, A. M. Rapid chromatographic technique for preparative separations with moderate resolution. *J. Org. Chem.* **43**, 2923-2925, (1978).
- 2 Pangborn, A. B., Giardello, M. A., Grubbs, R. H., Rosen, R. K. & Timmers, F. J. Safe and convenient procedure for solvent purification. *Organometallics* **15**, 1518-1520, (1996).

*4-Nitrophenyl 6-S-acetyl-6-thio- $\alpha$ -D-glucopyranoside*

$^1\text{H}$  NMR

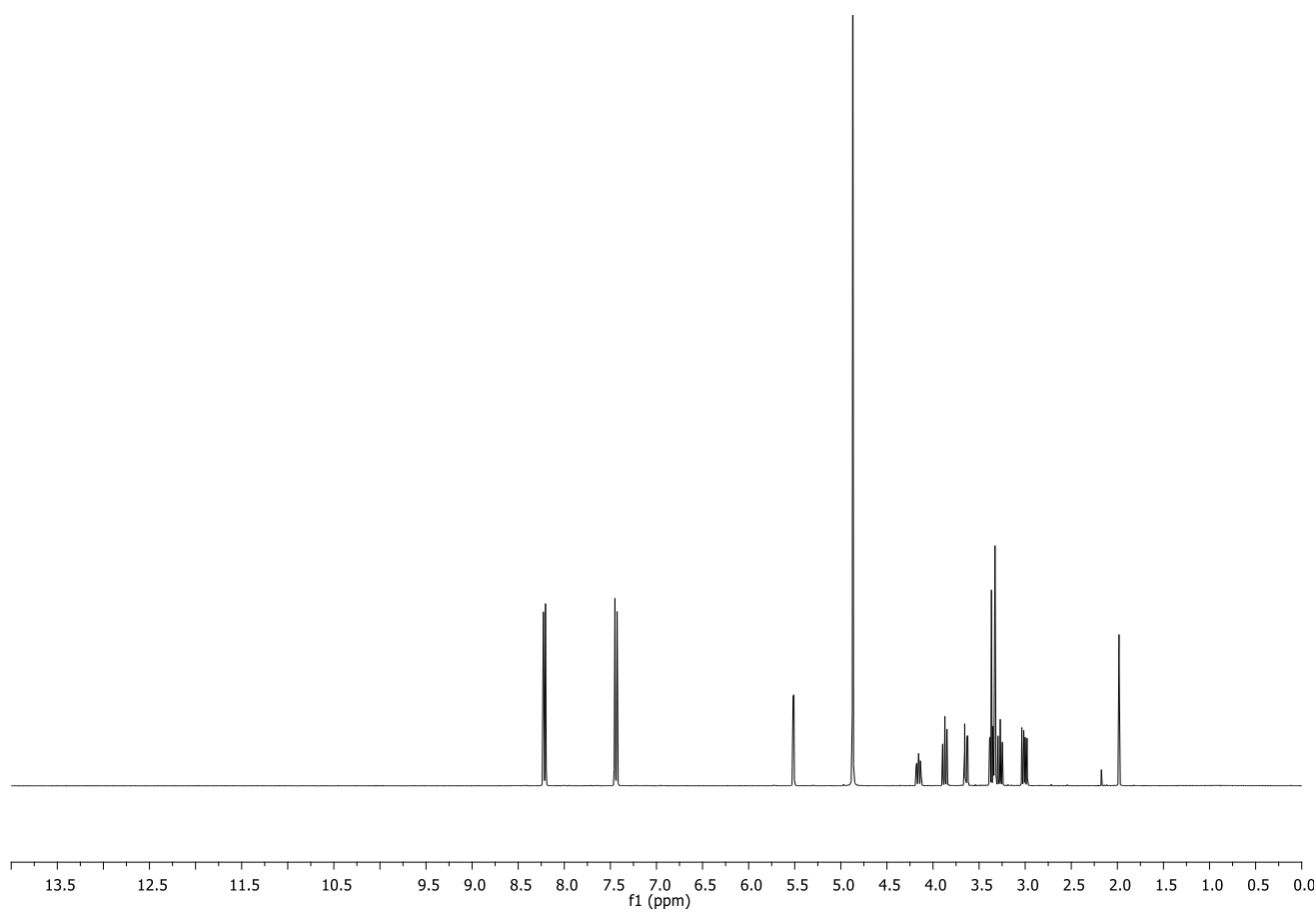


$^{13}\text{C}$  NMR



*Potassium 4-nitrophenyl 6-deoxy-6-sulfonato- $\alpha$ -D-glucopyranoside (PNPSQ)*

$^1\text{H}$  NMR



$^{13}\text{C}$  NMR

