
Article:
Speciale, Gaetano, Jin, Y. orcid.org/0000-0002-6927-4371, Davies, Gideon J. orcid.org/0000-0002-7343-776X et al. (2 more authors) (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

Title

YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids

Authors

Gaetano Speciale[1,†], Yi Jin[2,†], Gideon J. Davies[2], Spencer J. Williams[1], Ethan D. Goddard-Borger[3,4,*]

Affiliations

[1] School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010 (Australia)

[2] Department of Chemistry, University of York, Heslington, York, YO10 5DD (UK)

[3] ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052 (Australia)

[4] 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).

Abstract

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

Main Text

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

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

While SQDG is the primary source of SQ supplying these metabolic pathways, it is unknown how SQ is liberated from SQDG. Aside from an early report ascribing weak SQDG hydrolytic activity to *E. coli* β-galactosidase,[11] no glycoside hydrolases (GHs) dedicated to processing sulfoquinovosides have been reported. Putative SQases are located within the *E. coli* sulfoglycolysis and *P. putida* SQ Entner-Doudoroff gene clusters (YihQ and *Pp*SQ1_00094, respectively).[3,4] They are members of GH family 31 within the CAZy[12] sequence-based classification (http://www.cazy.org), a family that contains enzymes with α-glucosidase, α-glucan lyase, and α-xylanase activity. We cloned and expressed YihQ in *E. coli* using a similar approach to that reported.[13] Incubation of YihQ with SQDG isolated from spinach, and analysis of the reaction mixture by liquid chromatography/mass spectrometry (LC/MS), revealed complete conversion to SQ (Fig. 1b). A similar experiment on 1-sulfoquinovosylglycerol (SQGro), a metabolite of SQDG generated by the action of lipases,[14] indicated YihQ also hydrolysed simple sulfoquinovosides (Supplementary Results, Supplementary Fig. 1). To facilitate the convenient and accurate determination of kinetic parameters for YihQ we prepared a highly soluble, chromogenic substrate, *para*-nitrophenyl α-sulfoquinovoside (PNPSQ, Supplementary Notes). PNPSQ allowed the continuous acquisition of YihQ reaction rate data and the calculation of Michaelis-Menten parameters, revealing robust catalysis with k_{cat} = 14.3±0.4 s^{-1}, K_M = 0.22±0.03 mM and k_{cat}/K_M = (6.4±1.0)×10^4 M^{-1} s^{-1} (Supplementary Table 1). Under comparable conditions, we could not detect any activity against PNP α-D-glucopyranoside. Since CAZy GH family 31 contains both retaining glycosidases and α-glucan lyases,[15] we performed ^{1}H NMR spectroscopic analysis of the YihQ-catalyzed cleavage of PNPSQ to demonstrate rapid hydrolysis to the α-anomer of SQ, which after further time underwent mutarotation, confirming that YihQ is a retaining GH (Fig. 1c).

The X-ray structure of YihQ reveals an $(\alpha\beta)_8$ barrel appended with a small β-sheet domain (Fig. 2a, Supplementary Table 2). By soaking with the mechanism-based inactivator 5-fluoro-β-L-idopyranosyl fluoride (5FIdoF),[16] a covalent glycosyl-enzyme complex (in a $^1S_3$ pyranose conformation) was obtained, supporting assignment of D405 as the catalytic nucleophile (Fig. 2b).[17]

Located appropriately to protonate the glycosidic oxygen is D472, assigned as the catalytic acid/base residue.[18] A pseudo Michaelis complex was obtained by construction of a catalytically-inactive variant by mutation of the acid/base residue. The complex of PNPSQ with YihQ D472N revealed binding of the intact substrate in a $^4C_1$ conformation (Fig. 2c, Supplementary Fig. 2). Overall, the architectural features of the YihQ active site, formation of a glycosyl-enzyme intermediate with 5FIdoF, and the observation of retention of stereochemistry upon substrate hydrolysis are consistent with a classical Koshland retaining mechanism in which D405 fulfils the role of catalytic nucleophile and D472 acts as a general acid/base. Consistent with these assignments, the D405A, D405N, D472A and D472N YihQ variants were each catalytically inactive (Supplementary Table 1). Our data are consistent with a conformational itinerary of $^4C_1 \rightarrow {}^4H_3^{\ddagger} \rightarrow {}^1S_3$ for the YihQ glycosylation half-reaction (Supplementary Fig. 3).[19,20]

The PNPSQ complex with YihQ D472N reveals a detailed picture of the structural features required to recognise the distinguishing sulfonate group of SQ (Fig. 2d, Supplementary Fig. 2). The positively-charged R301 residue forms a salt-bridge with one oxygen of the anionic sulfonate group, a hydrogen-bond is formed between the indole N-H of W304 and a second sulfonate oxygen, and a well-ordered water molecule hydrogen bonded to O4 of SQ and Y508 forms a hydrogen bond to the third sulfonate oxygen. Comparison of the complex with that of a related GH31 α-glucosidase, SBG (from sugar beet) in complex with acarbose,[21] highlights key differences in the active-site residues that may explain the specificity of YihQ for SQ over D-glucose (Fig. 2e). Most notably, while there exists an equivalent residue to YihQ H537 in SBG (H626) that interacts with the 3-hydroxyl in both complexes, significant differences are seen in the residues around the 4- and 6-positions. Within YihQ, clearly defined roles in sulfonate recognition can be ascribed to W304, R301 and Y508 (the latter through a bridging water molecule). However, within SBG the sugar hydroxymethyl group adopts a different geometry leading to significantly different interactions. SBG W432 adopts a roughly similar position to YihQ W304, but does not appear to be involved in 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 α-glucosidase activity.

122 Sequence alignment of a range of GH31 members with α-glucosidase, α-xylanase and α-
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 *Pp*SQ1_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

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

Accession Codes

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

Acknowledgements

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

Author Contributions

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

References for Main Text

1. Harwood, J.L. & Nicholls, R.G. The plant sulpholipid – a major component of the sulphur cycle. *Biochem. Soc. Trans.* **7**, 440–447 (1979).

2. Benning, C. Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 53–75 (1998).

3. Denger, K. et al. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **507**, 114–117 (2014).

4. Felux, A.K., Spiteller, D., Klebensberger, J. & Schleheck, D. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. USA* **112**, E4298–305 (2015).

5. Shimojima, M. Biosynthesis and functions of the plant sulfolipid. *Prog. Lipid Res.* **50**, 234–239 (2011).

6. Martelli, H.L. & Benson, A.A. Sulfocarbohydrate metabolism. I. Bacterial production and utilization of sulfoacetate. *Biochim. Biophys. Acta* **93**, 169–171 (1964).

7. Roy, A.B., Hewlins, M.J., Ellis, A.J., Harwood, J.L. & White, G.F. Glycolytic breakdown of sulfoquinovose in bacteria: a missing link in the sulfur cycle. *Appl. Environ. Microbiol.* **69**, 6434–6441 (2003).

8. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A.M. Sulfoquinovose degraded by pure cultures of bacteria with release of C_3-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39–45 (2012).

9. Sugimoto, K., Sato, N. & Tsuzuki, M. Utilization of a chloroplast membrane sulfolipid as a major internal sulfur source for protein synthesis in the early phase of sulfur starvation in *Chlamydomonas reinhardtii*. *FEBS Lett.* **581**, 4519–4522 (2007).

10. Durham, B.P. et al. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. USA* **112**, 453–457 (2015).

11. Shibuya, I. & Benson, A.A. Hydrolysis of a-sulphoquinovosides by β-galactosidase. *Nature* **192**, 1186–1187 (1961).

12. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–5 (2014).

13. Okuyama, M., Mori, H., Chiba, S. & Kimura, A. Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from *Escherichia coli*. *Protein Expr. Purif.* **37**, 170–179 (2004).

14. Andersson, L., Carriere, F., Lowe, M.E., Nilsson, A. & Verger, R. Pancreatic lipase-related protein 2 but not classical pancreatic lipase hydrolyzes galactolipids. *Biochim. Biophys. Acta* **1302**, 236–40 (1996).

15. Lee, S.S., Yu, S. & Withers, S.G. α-1,4-Glucan lyase performs a trans-elimination via a nucleophilic displacement followed by a syn-elimination. *J. Am. Chem. Soc.* **124**, 4948–9 (2002).

16. McCarter, J.D., and Withers, S.G. 5-Fluoro glycosides: A new class of mechanism based inhibitors of both α- and β-glucosidases. *J. Am. Chem. Soc.* **118**, 241 (1996).

17. Quaroni, A. & Semenza, G. Partial amino acid sequences around the essential carboxylate in the active sites of the intestinal sucrase-isomaltase complex. *J. Biol. Chem.* **251**, 3250–3253 (1976).

18. Okuyama, M. et al. Carboxyl group of residue Asp647 as possible proton donor in catalytic reaction of α-glucosidase from *Schizosaccharomyces pombe*. *Eur. J. Biochem.* **268**, 2270–2280 (2001).

19. Davies, G.J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316 (2012).

20. Speciale, G., Thompson, A.J., Davies, G.J. & Williams, S.J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13 (2014).

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

Figure Legends for Main Text

Figure 1: *E. coli* YihQ is a sulfoquinovosidase that hydrolyzes SQDG to SQ. (**a**) Biosynthesis and catabolism of SQ via SQDG. Biosynthesis occurs in photosynthetic organisms and catabolism occurs through sulfoglycolytic processes including the SQ Embden–Meyerhof–Parnas (SQ-EMP) sulfoglycolysis and SQ Entner–Doudoroff (SQ-ED) pathways, with the resulting 2,3-dihydroxypropane-1-sulfonate (DHPS) and sulfolactate (SL) undergoing biomineralization to sulfate in members of the bacterial community. (**b**) SQDG from spinach 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 SQDG (*m/z* 815.5) before (blue) and after (red) incubation with YihQ. (**c**) Time-course ^{1}H NMR spectra showing that YihQ hydrolyzes PNPSQ to release SQ with retention of anomeric configuration.

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

Online Methods

Cloning the *yihQ* gene

The *yihQ* gene (Supplementary Fig. 6) was amplified by colony PCR from *E. coli* K12 strain DH5α using the oligonucleotides YihQ-f and YihQ-r as primers (Supplementary Table 3). The 2053 base-pair product was digested with the restriction enzymes *Nde*I and *Xho*I and ligated into pET29b(+) (Novagen) that had been digested with the same enzymes and treated with alkaline phosphatase. The sequence of the gene in the resulting plasmid (pET29-YihQ) was confirmed by Sanger sequencing using the oligonucleotide primers T7, T7-term and YihQ-seq (Supplementary Table 3).

YihQ expression and purification

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

Mutagenesis of YihQ

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

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

LC/MS analysis of YihQ-digested SQDG and SQGro

SQDG (>95%, Indofine, 20 μl, 1 mg ml^{-1} in $CHCl_3$:MeOH, 1:1) in 200 mM NH_4CO_3, 5% DMSO, pH 6.2 (90 μl) was treated with YihQ (10 μl, 1.36 μM) and incubated overnight at 23 °C. SQGro was prepared by incubating SQDG (20 μl, 1 mg ml^{-1} in $CHCl_3$:MeOH) with NaOMe in MeOH (1 μl, 100 mM) at 23 °C for 3 h prior to the addition of buffer (90 μl, 200 mM NH_4CO_3, pH 6.2) and YihQ (10 μl, 1.36 μM) and incubation of the mixture at 23 °C overnight. Both reaction mixtures were lyophilised three times from H_2O, resuspended in $CHCl_3$:MeOH:H_2O (4:6:1) and centrifuged at 14,000 rpm for 10 min. The supernatant was analysed by LC/MS (Agilent electrospray ionization-time of flight mass spectrometer, negative mode), using an Agilent HILIC plus (3.5 μm, 2.1 x 100 mm) column. Control reactions without the addition of YihQ had no detectable quantities of SQ, indicating that the rate of non-enzymatic hydrolysis is negligible under these conditions.

Stereochemical outcome of YihQ-catalyzed hydrolysis

The YihQ catalyzed hydrolysis of PNPSQ (for synthesis see Supplementary Notes) was monitored by ^{1}H NMR spectroscopy using a 500 MHz instrument. A solution of YihQ in buffered D_2O (0.25 ml, 0.1 mM in 50 mM sodium phosphate, 250 mM NaCl, pH 7.2) was added to a solution of PNPSQ (4.0 mg, 9.9 mmol) in buffered D_2O (0.75 ml, 50 mM sodium phosphate, 250 mM NaCl, pH 7.2) at 22 °C. ^{1}H NMR spectra were acquired at different time points (t = 0, 7 min, 24 min and 24 h).

pH dependence of YihQ activity

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

Enzyme kinetics

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

X-ray crystallography and structure solution

Well-diffracting crystals of YihQ were obtained by mixing 25 mg ml^{-1} protein stock with equal volume of precipitant composed of 50–60% (v/v) 2-methyl-2,4-pentanediol, 0.10–0.15 M $CaCl_2$, and bis-tris, pH 6.5 after 3–4 days at 20 °C using the sitting drop vapor diffusion method.

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

Graphics

The amino acid conservation scores for YihQ in Fig. 2a were calculated using the Consurf Server. [25,26] The figures for the 3D protein structures were produced using Pymol[27] and CCP4MG.[28]

Sequence alignments

Amino acid sequences for putative SQases were identified through a series of protein-protein BLAST (blastp) searches of the NCBI non-redundant protein sequences database targeting the different kingdoms/phyla of life using *E. coli* YihQ as the query sequence. Representative protein sequences from different phyla were selected for alignment using Clustal Omega. The genes encoding putative eukaryotic SQases possessed multiple intronic regions, providing confidence that they were not artefacts arising from bacterial contamination of their respective genome assemblies. Global phylogenies of the putative SQases were computed using Phylogeny.fr, which performed initial tree generation with BIONJ and optimization with PhyML 3.0 using the LG substitution

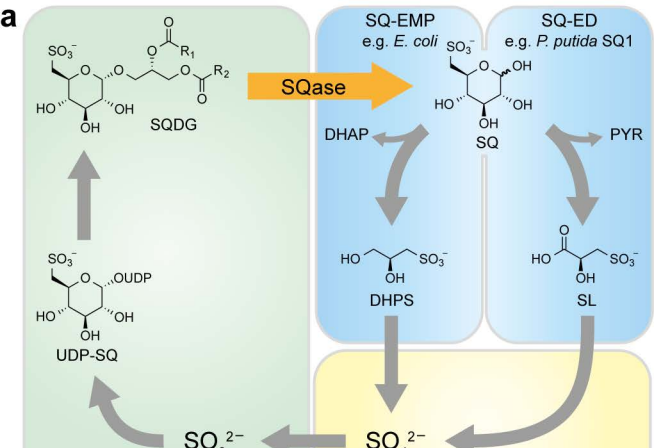
model (γ 8, NNI) and branch support by the aLRT-SH method with bootstrap (100) analyses. [29] Phylogenetic tree diagrams were constructed using TreeDyn 198.3, collapsing branches with bootstrap values less than 0.9.

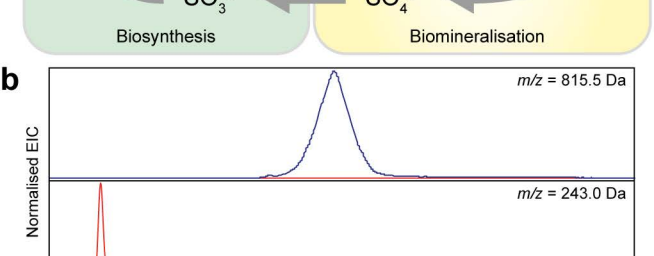
Methods-only References

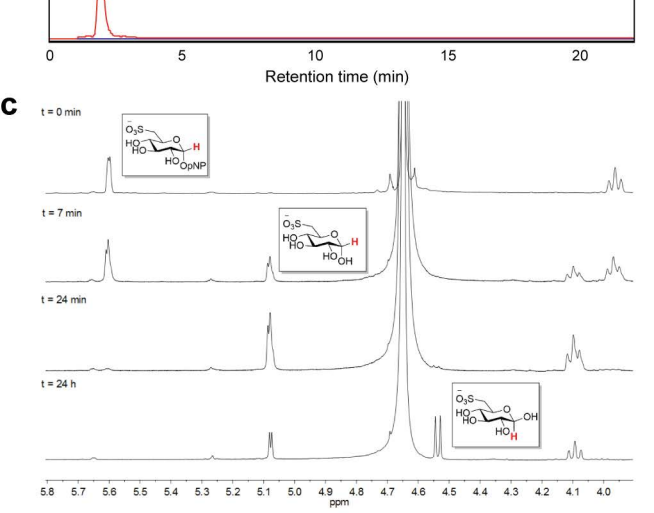
22. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–34 (2005).
23. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
24. The CCP4 suite: programs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

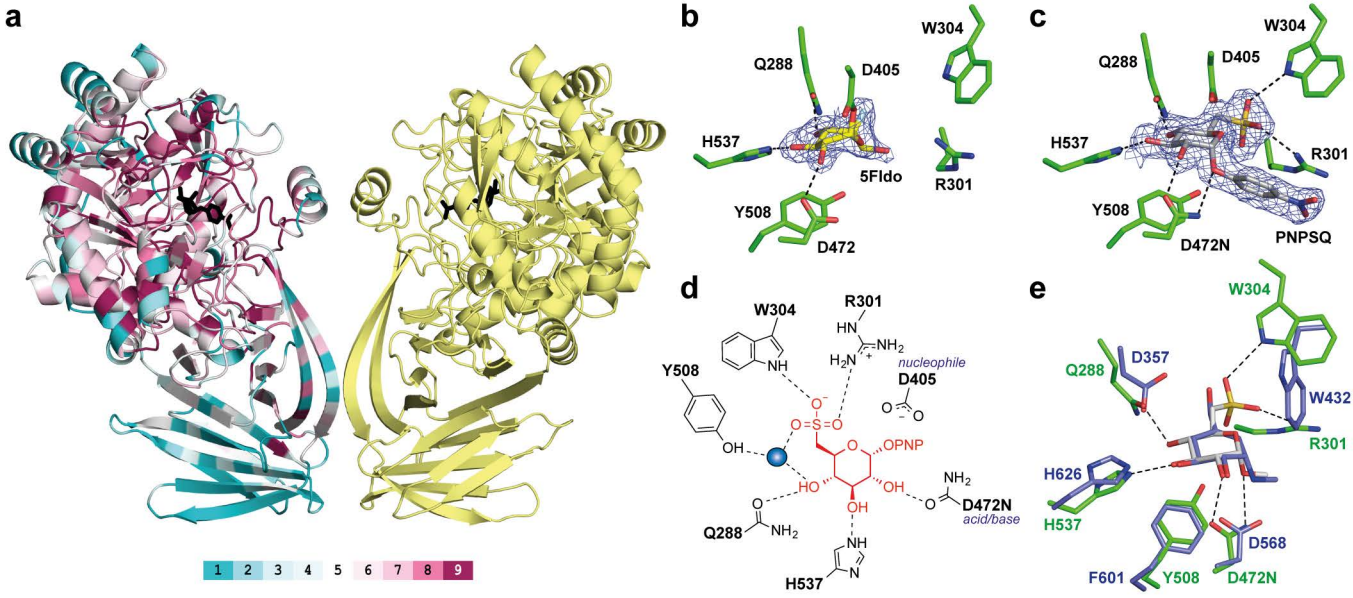
Competing financial interests

The authors declare no competing financial interests.


a
SQDG
SQase
SQ-EMP
e.g. E. coli
SQ-ED
e.g. P. putida SQ1
SQ
DHAP
PYR
DHPS
SL
UDP-SQ
Biosynthesis
Biomineralisation


b
m/z = 815.5 Da
m/z = 243.0 Da
Normalised EIC
Retention time (min)


c
t = 0 min
t = 7 min
t = 24 min
t = 24 h
ppm


a
1 2 3 4 5 6 7 8 9
b
W304
Q288
D405
H537
5FIdo
R301
Y508
D472
c
W304
Q288
D405
H537
R301
Y508
D472N
PNPSQ
d
W304
R301
Y508
nucleophile
D405
OPNP
D472N
acid/base
Q288
H537
e
W304
Q288
D357
W432
R301
H626
H537
D568
F601
Y508
D472N

SUPPLEMENTARY INFORMATION

YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids

Gaetano Speciale[1,†], Yi Jin[2,†], Gideon J. Davies[2], Spencer J. Williams[1] and Ethan D. Goddard-Borger[3,4,*]

[1] School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010 (Australia)

[2] Department of Chemistry, University of York, Heslington, York, YO10 5DD (UK)

[3] ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052 (Australia)

[4] 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^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
YihQ (WT)	pNPSQ	0.22±0.03	14.3±0.4	$(6.4\pm1.0)\times10^{4}$
	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[a]
	pNPGlc	n/d	n/d	n/d
YihQ (R301Q)	pNPSQ	2.9±0.4	$(3.1\pm0.1)\times10^{-3}$	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	$(7.2\pm0.5)\times10^{-5}$	$(1.1\pm0.4)\times10^{-2}$
YihQ (W304F)	pNPSQ	n/d	n/d	n/d
	pNPGlc	n/d	n/d	n/d

[a] saturation not reached

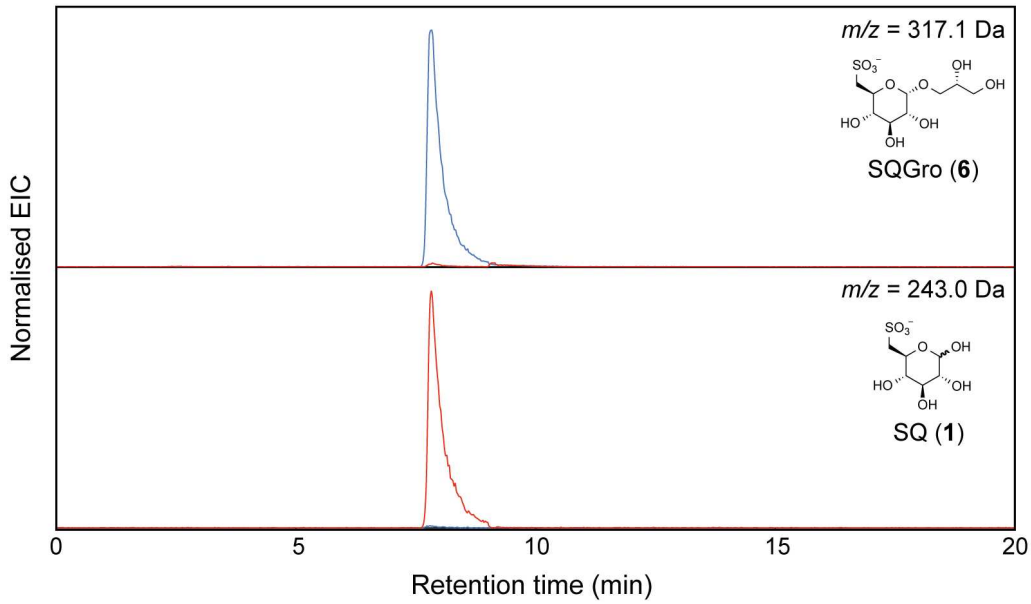
Supplementary Table 2: X-ray data collection, processing and refinement statistics.

	wt native	wt-5FIdo	D472N-pNPSQ
Data collection			
Space group	P21	P21	P21
Cell dimensions			
a, *b*, *c* (Å)	78.7, 113.2, 111.7	78.7, 112.6, 111.9	78.0, 107.6, 103.3
a, b, g (°)	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)
R_{sym} or R_{merge}	0.073 (0.41)	0.054 (0.621)	0.050 (0.878)
I / sI	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)
Refinement			
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)
R_{work} / R_{free}	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
B-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

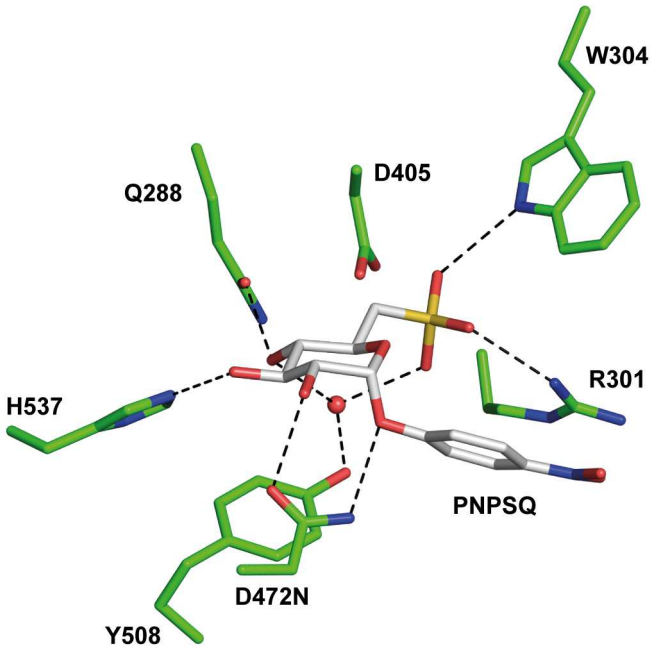
Supplementary Table 3: Primers for cloning and mutagenesis.

primer	sequence
YihQ-f	ATATACATATGGATACGCCACGTCCAC
YihQ-r	TGGTGCTCGAGGATGCTTTTTAACGACGCGAAC
YihQ-seq	ACTTCGACTTTAGTGCCC
T7	TAATACGACTCACTATAGGG
T7-term	GCTAGTTATTGCTCAGCGG
YihQ-D405A-f	GGCTGCGGCGGCTGGATGGCT**GCC**TTCGGCGAGTATCTGCCCACC
YihQ-D405A-r	GGTGGGCAGATACTCGCCGAA**GGC**AGCCATCCAGCCGCCGCAGCC
YihQ-D405N-f	GGCTGCGGCGGCTGGATGGCT**AAC**TTCGGCGAGTATCTGCCCACC
YihQ-D405N-r	GGTGGGCAGATACTCGCCGAA**GTT**AGCCATCCAGCCGCCGCAGCC
YihQ-D472A-f	TCCACCATGATGTGGGCGGGC**GCC**CAGAACGTCGACTGGAGTCTCGAC
YihQ-D472A-r	GTCGAGACTCCAGTCGACGTTCTG**GGC**GCCCGCCCACATCATGGTGGA
YihQ-D472N-f	TCCACCATGATGTGGGCGGGC**AAC**CAGAACGTCGACTGGAGTCTCGAC
YihQ-D472N-r	GTCGAGACTCCAGTCGACGTTCTG**GTT**GCCCGCCCACATCATGGTGGA
YihQ-R301A-f	CGTATGACCTCTTTTGGCAAA**GCC**GTGATGTGGAACTGGAAGTGG
YihQ-R301A-r	CCACTTCCAGTTCCACATCAC**GGC**TTTGCCAAAAGAGGTCATACG
YihQ-R301K-f	CGTATGACCTCTTTTGGCAAA**AAG**GTGATGTGGAACTGGAAGTGG
YihQ-R301K-r	CCACTTCCAGTTCCACATCAC**CTT**TTTGCCAAAAGAGGTCATACG
YihQ-R301E-f	CGTATGACCTCTTTTGGCAAA**GAA**GTGATGTGGAACTGGAAGTGG
YihQ-R301E-r	CCACTTCCAGTTCCACATCAC**TTC**TTTGCCAAAAGAGGTCATACG
YihQ-R301Q-f	CGTATGACCTCTTTTGGCAAA**CAG**GTGATGTGGAACTGGAAGTGG
YihQ-R301Q-r	CCACTTCCAGTTCCACATCAC**CTG**TTTGCCAAAAGAGGTCATACG
YihQ-Q288E-f	GAAGGTCAACGGCATCTGGGCG**GAG**GACTGGTCCGGTATTCGTATGACCTC
YihQ-Q288E-r	GAGGTCATACGAATACCGGACCAGTC**CTC**CGCCCAGATGCCGTTGACCTTC
YihQ-W304F-f	CCTCTTTTGGCAAACGCGTGATG**TTT**AACTGGAAGTGGAACAGCGAAAAC
YihQ-W304F-r	GTTTTCGCTGTTCCACTTCCAGTT**AAA**CATCACGCGTTTGCCAAAAGAGG

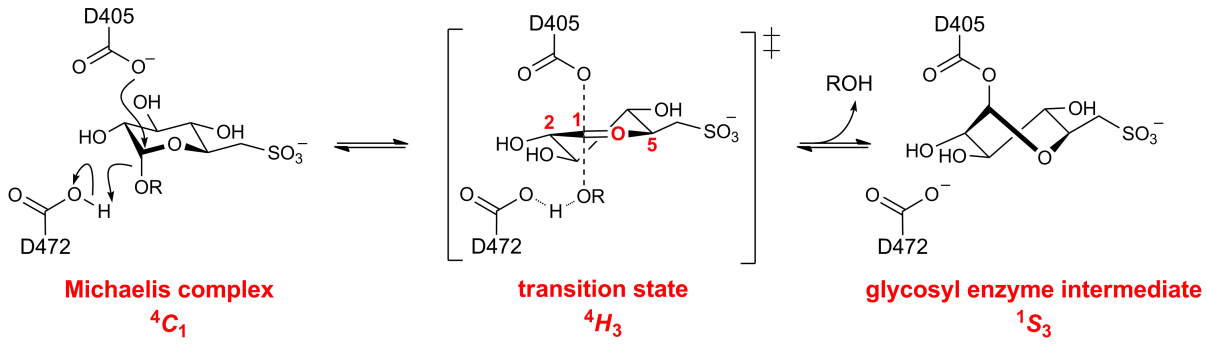
Supplementary Figures



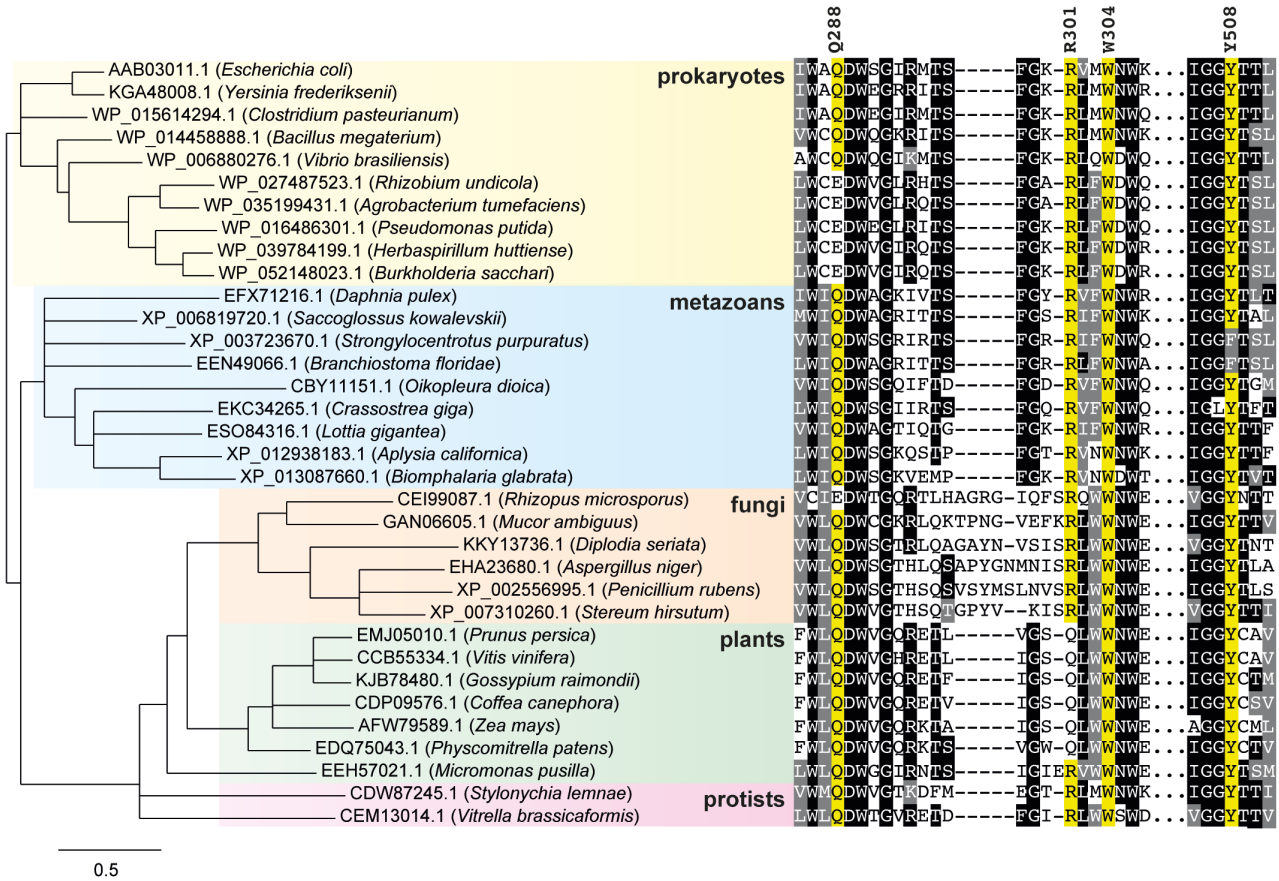
Supplementary Figure 1: YihQ acts on 1-sulfoquinovosylglycerol (SQGro). SQGro, prepared by Zemplén transesterification on SQDG, was incubated with YihQ and analyzed by LC/MS. Extracted ion chromatogram mass spectra showing normalized ion count of SQ (*m/z* = 243.0) and SQGro (*m/z* 317.1) before (blue)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



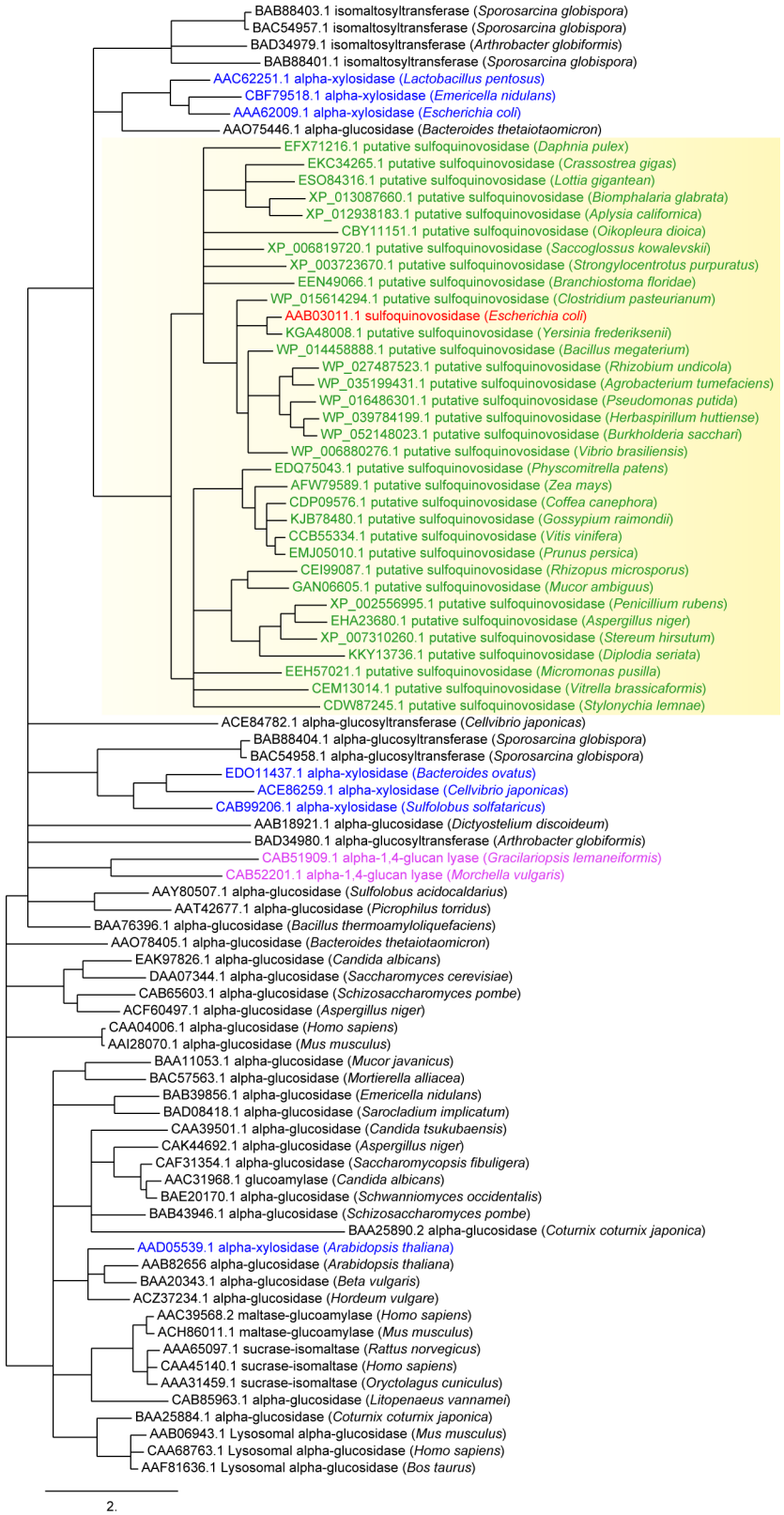
Supplementary Figure 2: Detailed view of active site of YihQ D472N in complex with PNPSQ.



Supplementary Figure 3: Glycosylation half reaction for retaining GH31 sulfoquinovosidase.



Supplementary Figure 4: Identification of signature sulfoquinovosidase residues allows prediction of distribution of SQase activity. Phylogenetic tree and sequence alignment illustrating the wide distribution of putative SQases across the tree of life.



Supplementary Figure 5: Phylogentic 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).

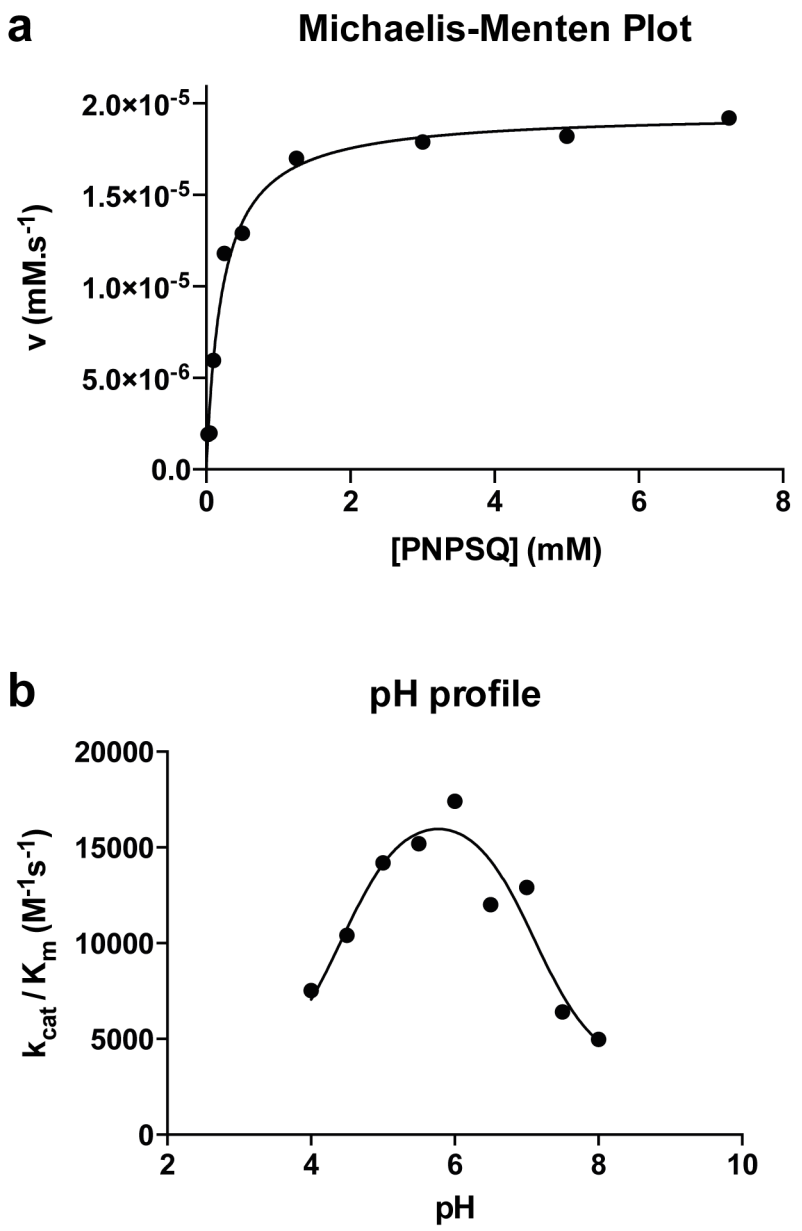
a

```
>YihQ-His6
ATGGATACGCCACGTCCACAGTTATTAGATTTTCAATTTCATCAGAATAACGACAGTTTTACCCTACATTTTCAACAACG
TCTTATTTTAACCCATAGCAAAGATAATCCTTGTTTATGGATTGGCTCAGGTATAGCGGATATCGATATGTTCCGCGGTA
ATTTCAGCATTAAAGATAAACTACAGGAGAAAATTGCGCTTACCGACGCCATCGTCAGCCAGTCACCGGATGGTTGGTTA
ATTCATTTCAGCCGTGGTTCTGACATTAGCGCCACGCTGAATATCTCTGCCGACGATCAGGGGCGTTTATTGCTGGAACT
ACAAAACGACAACCTTAACCACAACCGTATCTGGCTGCGCCTTGCCGCTCAACCAGAGGACCATATCTACGGCTGCGGCG
AACAGTTTTCCTACTTCGATCTGCGTGGCAAACCGTTCCCGCTATGGACCAGTGAACAAGGCGTTGGTCGCAACAAACAA
ACCTATGTCACCTGGCAGGCCGACTGCAAAGAAAATGCGGGCGGCGACTATTACTGGACTTTCTTCCCACAGCCTACGTT
TGTCAGCACGCAGAAGTATTACTGCCATGTTGATAACAGTTGCTATATGAACTTCGACTTTAGTGCCCCGGAATACCATG
AACTGGCGCTGTGGGAAGACAAAGCAACGCTGCGTTTTGAATGTGCTGACACATACATTTCCCTGCTGGAAAAATTAACC
GCCCTGCTGGGACGCCAGCCAGAACTGCCCGACTGGATTTATGACGGAGTAACGCTCGGCATTCAGGGCGGGACGGAAGT
GTGCCAGAAGAAACTGGACACCATGCGTAACGCGGGCGTGAAGGTCAACGGCATCTGGGCGCAGGACTGGTCCGGTATTC
GTATGACCTCTTTTGGCAAACGCGTGATGTGGAACTGGAAGTGGAACAGCGAAAACTACCCGCAACTGGATTCACGCATT
AAGCAGTGGAATCAGGAGGGCGTGCAGTTCCTGGCCTATATCAACCCGTATGTTGCCAGCGATAAAGATCTCTGCGAAGA
AGCGGCACAACACGGCTATCTGGCAAAAGATGCCTCTGGCGGTGACTATCTGGTGGAGTTTGGCGAGTTTTACGGCGGCG
TTGTCGATCTCACTAATCCAGAAGCCTACGCCTGGTTCAAGGAAGTGATCAAAAAGAACATGATTGAACTCGGCTGCGGC
GGCTGGATGGCTGACTTCGGCGAGTATCTGCCCACCGACACGTACTTGCATAACGGCGTCAGTGCCGAAATTATGCATAA
CGCCTGGCCTGCGCTGTGGGCGAAGTGTAACTACGAAGCCCTTGAAGAAACGGGCAAGCTCGGCGAGATCCTTTTCTTTA
TGCGCGCCGGTTCTACCGGTAGCCAGAAATACTCCACCATGATGTGGGCGGGCGACCAGAACGTCGACTGGAGTCTCGAC
GATGGCCTGGCGTCGGTTGTCCCGGCGGCGCTGTCGCTGGCAATGACCGGACATGGCCTGCACCACAGCGACATTGGCGG
TTACACCACCCTGTTTGAGATGAAGCGCAGCAAAGAGCTGCTGCTGCGCTGGTGCGATTTCAGCGCCTTCACGCCGATGA
TGCGCACCCACGAAGGTAACCGTCCTGGCGACAACTGGCAGTTTGACGGCGACGCAGAAACCATCGCCCATTTCGCCCGT
ATGACCACCGTCTTCACCACCCTGAAACCTTACCTGAAAGAGGCCGTCGCGCTGAATGCGAAGTCCGGCCTGCCGGTTAT
GCGCCCGCTGTTCCTGCATTACGAAGACGATGCGCACACTTACACCCTGAAATATCAGTACCTGTTAGGTCGCGACATTC
TGGTCGCTCCGGTGCATGAAGAAGGCCGTAGCGACTGGACGCTCTATCTGCCGGAGGATAACTGGGTCCACGCCTGGACG
GGTGAAGCGTTCCGGGGCGGGGAAGTTACCGTTAATGCGCCCATCGGCAAGCCGCCGGTCTTTTATCGCGCCGATAGCGA
ATGGGCGGCACTGTTCGCGTCGTTAAAAAGCATCCTCGAGCACCACCACCACCACCACTGA
```

b

```
>YihQ-His6
MDTPRPQLLDFQFHQNNDSFTLHFQQRLILTHSKDNPCLWIGSGIADIDMFRGNFSIKDK
LQEKIALTDAIVSQSPDGWLIHFSRGSDISATLNISADDQGRLLLELQNDNLNHNRIWLR
LAAQPEDHIYGCGEQFSYFDLRGKPFPLWTSEQGVGRNKQTYVTWQADCKENAGGDYYWT
FFPQPTFVSTQKYYCHVDNSCYMNFDFSAPEYHELALWEDKATLRFECADTYISLLEKLT
ALLGRQPELPDWIYDGVTLGIQGGTEVCQKKLDTMRNAGVKVNGIWAQDWSGIRMTSFGK
RVMWNWKWNSENYPQLDSRIKQWNQEGVQFLAYINPYVASDKDLCEEAAQHGYLAKDASG
GDYLVEFGEFYGGVVDLTNPEAYAWFKEVIKKNMIELGCGGWMADFGEYLPTDTYLHNGV
SAEIMHNAWPALWAKCNYEALEETGKLGEILFFMRAGSTGSQKYSTMMWAGDQNVDWSLD
DGLASVVPAALSLAMTGHGLHHSDIGGYTTLFEMKRSKELLLRWCDFSAFTPMMRTHEGN
RPGDNWQFDGDAETIAHFARMTTVFTTLKPYLKEAVALNAKSGLPVMRPLFLHYEDDAHT
YTLKYQYLLGRDILVAPVHEEGRSDWTLYLPEDNWVHAWTGEAFRGGEVTVNAPIGKPPV
FYRADSEWAALFASLKSILEHHHHHH*
```

Supplementary Figure 6: YihQ sequence details. (**a**) The nucleotide sequence encoding His_6-tagged YihQ and (**b**) the amino acid sequence of His_6-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.

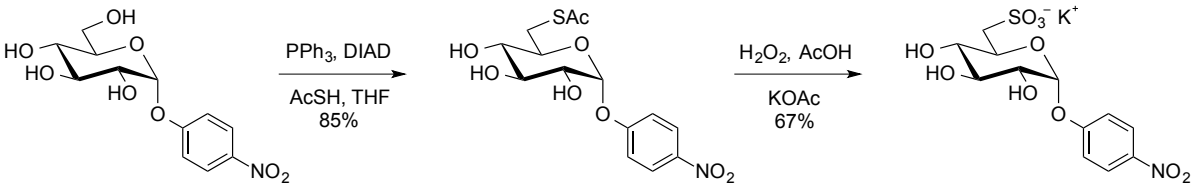
Supplementary Notes

Chemical synthesis of PNPSQ

General

All chemical reagents were purchased from Sigma-Aldrich at >95% purity unless otherwise stated. ^{1}H and ^{13}C NMR spectra were recorded using a 400 MHz instrument. All signals were referenced to solvent peaks (d_4-MeOH: δ 3.49 ppm for ^{1}H or 49.0 ppm for ^{13}C). TLC analysis used aluminium backed Merck Silica Gel 60 F_{254} sheets, detection was achieved using UV light, 5% H_2SO_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*.[1] Dry THF was obtained from a dry solvent apparatus (Glass Contour of SG Water, Nashua, U.S.A.).[2] Melting points were obtained using a hot-stage microscope. $[\alpha]_D$ values are given in deg 10^{-1} cm^2 g^{-1}.

Synthetic scheme



*4-Nitrophenyl 6-*S*-acetyl-6-thio-α-D-glucopyranoside*

A solution of 4-nitrophenyl α-D-glucopyranoside (0.10 g, 0.33 mmol) and thioacetic acid (28 μl, 0.40 mmol) in dry THF (2.0 ml) at 0 °C was added to a mixture of PPh_3 (0.10 g, 0.40 mmol) and DIAD (80 μl, 0.40 mmol) in THF (1.0 ml) at 0 °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°C, EtOAc); $[\alpha]_D^{23}$ +1.2° (*c* 0.55, CH_3OH); ^{1}H NMR (400 MHz, CD_3OD) δ 2.21 (3 H, s, 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}C NMR (101 MHz, CD_3OD) δ 30.25 (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 $(ESI)^+$ *m/z* 360.0760 [$C_{14}H_{17}NO_8S$ $(M + H)^+$ requires 360.0753].

Potassium 4-nitrophenyl 6-deoxy-6-sulfonato-α-D-glucopyranoside (PNPSQ)

30% H_2O_2 (0.70 ml) was added to a solution of 4-nitrophenyl 6-*S*-acetyl-6-thio-α-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_3 in Et_2O (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_2O, 19:2:1→7:2:1) and C_{18} reversed phase chromatography (H_2O/CH_3CN, 95:5), affording PNPSQ (38 mg, 0.094 mmol, 67%) as a white solid. $[\alpha]_D^{23}$ +1.3° (*c* 0.91, CH_3OH); ^{1}H NMR (400 MHz, CD_3OD) δ 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); ^{13}C NMR (101 MHz, CD_3OD) δ 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 $(ESI)^-$ *m/z* 364.0340 [$C_{12}H_{14}NO_{10}S$ $(M - H)^-$ requires 364.0343].

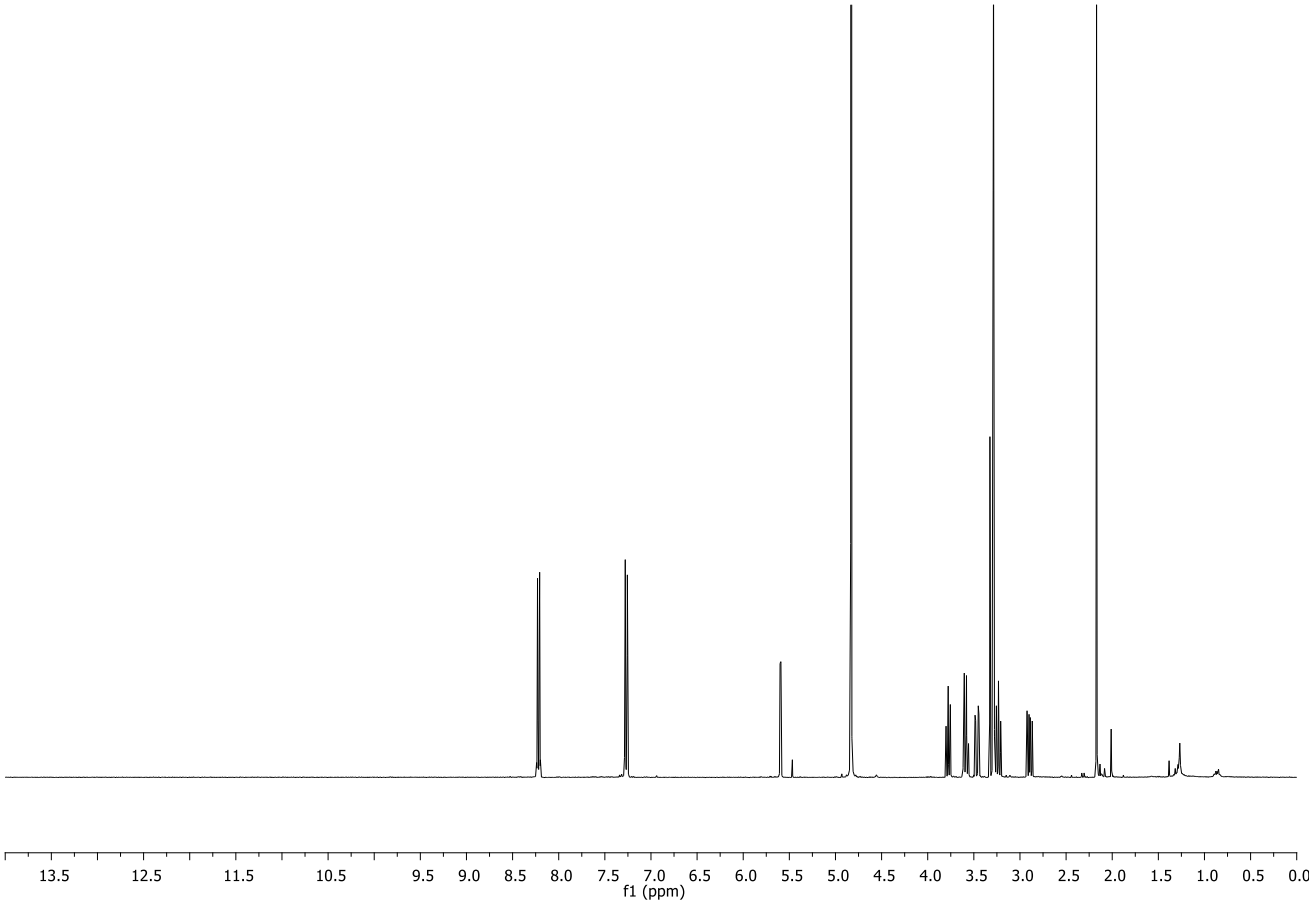
References

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).

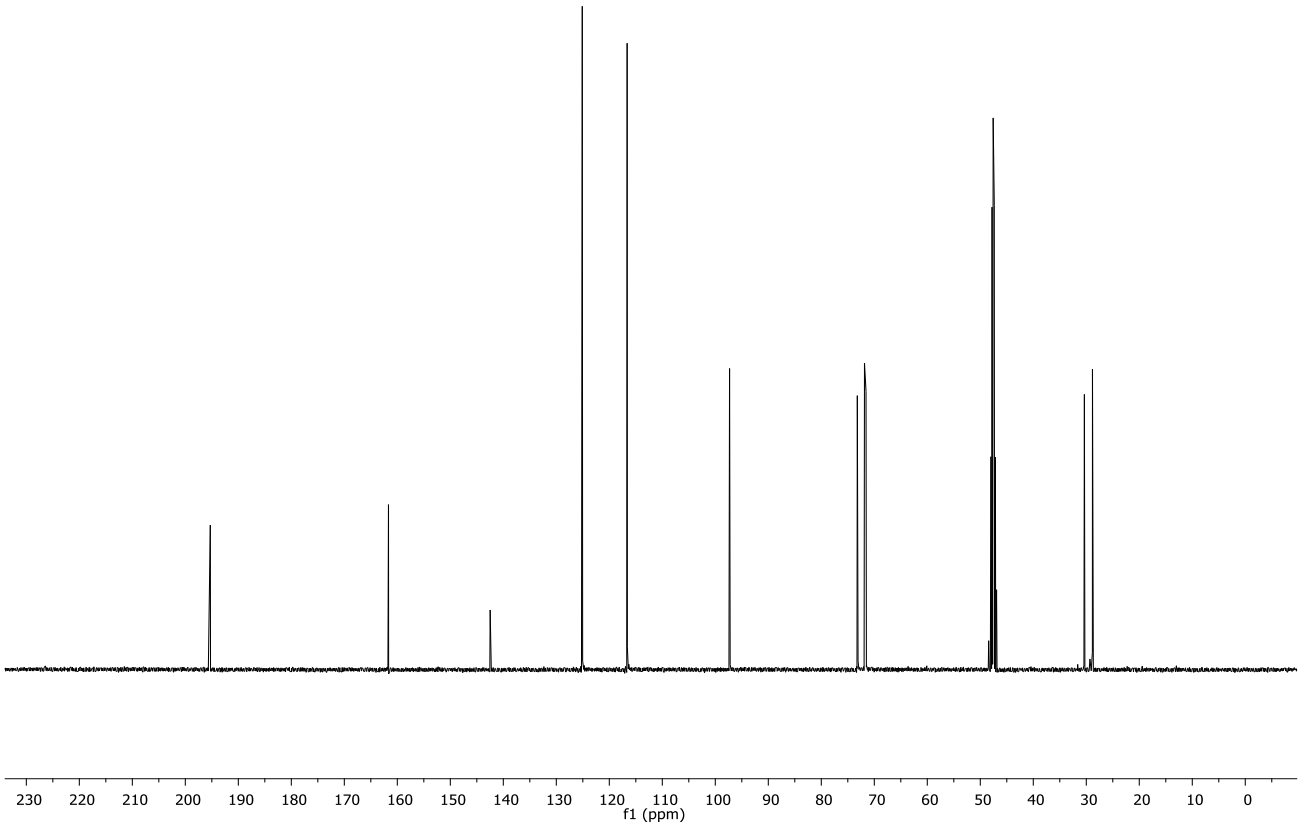
2 Pangborn, A. B., Giardello, M. A., Grubbs, R. H., Rosen, R. K. & Timmers, F. J. Safe and convenient procedure for solvent purification. *Organometallics* **15**, 1518-1520, (1996).

4-Nitrophenyl 6-S-acetyl-6-thio-α-D-glucopyranoside

^{1}H NMR

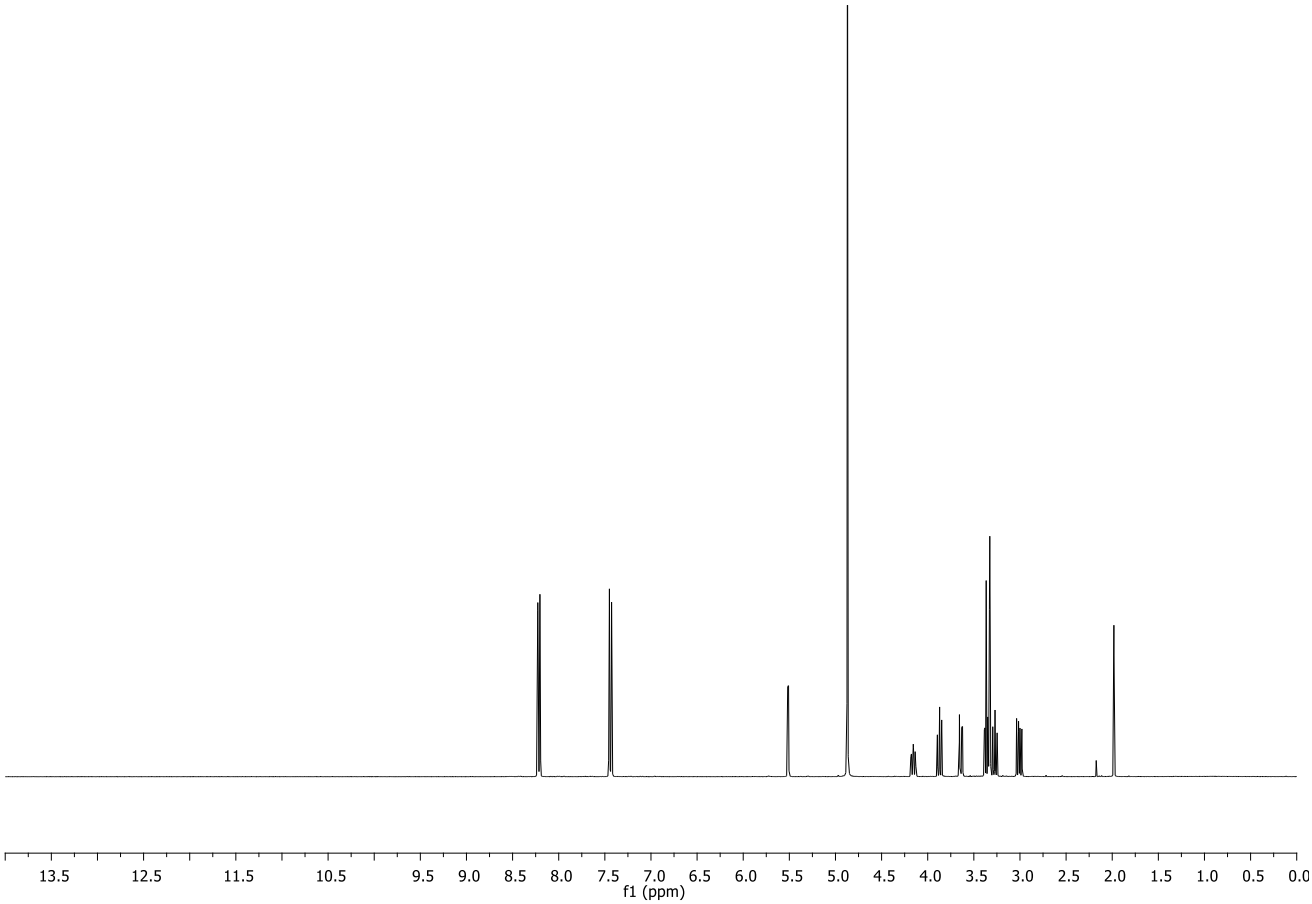


^{13}C NMR



Potassium 4-nitrophenyl 6-deoxy-6-sulfonato-α-D-glucopyranoside (PNPSQ)

^{1}H NMR



^{13}C NMR