Article:
Abayakoon, Palika, Jin, Yi orcid.org/0000-0002-6927-4371, Lingford, James P. et al. (8 more authors) (2018) Structural and biochemical insights into the function and evolution of sulfoquinovosidases. ACS Central Science. 1266–1273. ISSN 2374-7943

https://doi.org/10.1021/acscentsci.8b00453

Structural and biochemical insights into the function and evolution of sulfoquinovosidases

Palika Abayakoon[1],† Yi Jin[2],† James P. Lingford[3,4], Marija Petricevic[1], Alan John[3,4], Eileen Ryan[1], Janice Wai-Ying Mui[1], Douglas E.V. Pires[5], David B. Ascher[5], Gideon J. Davies[2,*], Ethan D. Goddard-Borger[3,4,*], Spencer J. Williams[1]*

Affiliations:

1. School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Vic 3010 (Australia)

2. York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, YO10 5DD (UK)

3. ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Vic 3010 (Australia)

4. Department of Medical Biology, University of Melbourne, Parkville, Vic 3010 (Australia)

5. Department of Biochemistry and Molecular Biology, and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Vic 3010 (Australia)

† These authors contributed equally to this work

* Correspondence should be addressed to S.J.W (sjwill@unimelb.edu.au), E.D.G.-B. (goddard-borger.e@wehi.edu.au) and G.J.D. (gideon.davies@york.ac.uk)

Keywords

Sulfoglycolysis, sulfoquinovose, glycosidase, sulfur cycle, enzyme evolution

Abstract

An estimated 10 billion tonnes of sulfoquinovose (SQ) are produced and degraded each year. Prokaryotic sulfoglycolytic pathways catabolise sulfoquinovose (SQ) liberated from plant sulfolipid, or its delipidated form α-D-sulfoquinovosyl glycerol (SQGro), through the action of a sulfoquinovosidase (SQase) but little is known about the capacity of SQ glycosides to support growth. Structural studies of the first reported SQase (*E. coli* YihQ) have identified three conserved residues that are essential for substrate recognition but cross-over mutations exploring active site residues of predicted SQases from other organisms have yielded inactive mutants casting doubt on bioinformatic functional assignment. Here, we show that SQGro can support the growth of *E. coli* on par with D-glucose, and that the *E. coli* SQase prefers the naturally occurring diastereomer of SQGro. A predicted, but divergent, SQase from *Agrobacterium tumefaciens* proved to have highly specific activity towards SQ glycosides, and structural, mutagenic and bioinformatic analyses revealed the molecular co-evolution of catalytically-important amino acid pairs directly involved in substrate recognition, as well as structurally-important pairs distal to the active site. Understanding the defining features of SQases empowers bioinformatic approaches for mapping sulfur metabolism in diverse microbial communities and sheds light on this poorly-understood arm of the biosulfur cycle.

Introduction

Sulfoquinovose is found in the ubiquitous plant sulfolipid α-D-sulfoquinovosyl diacylglyceride (SQDG), which is one of the most abundant organic sulfur compounds in nature.[1] SQDG is produced by most photosynthetic organisms and forms an integral part of the thylakoid membrane of the chloroplast, maintaining membrane charge and modulating the function of photosynthetic proteins.[2] The rapid turnover of photosynthetic cells, on land and in the oceans, makes the biosynthesis and degradation of SQDG a very significant arm of the global sulfur cycle. While the biosynthesis of SQDG is well-understood, details of its catabolism have only recently been elucidated. Two sulfoglycolytic processes have been identified, termed the sulfo-Embden-Meyerhof-Parnas (sulfo-EMP)[3] and sulfo-Entner-Doudoroff (sulfo-ED)[4] pathways (Fig. 1a). These prokaryotic pathways involve the catabolism of SQ to dihydroxypropanesulfonate (DHPS) or sulfolactate (SL), respectively. The sulfo-EMP and sulfo-ED pathways within bacteria are found in a single gene cluster that encodes a suite of sulfoglycolytic enzymes and includes a glycoside hydrolase (GH) from family 31 of the carbohydrate-active enzyme (CAZy) classification system.[3-4] The corresponding enzyme from the *E. coli* sulfo-EMP pathway (*Ec*YihQ) was recently shown to be a dedicated sulfoquinovosidase (SQase) capable of hydrolyzing SQDG or α-D-sulfoquinovosyl glycerol (SQGro).[5] SQases are thought to be important to sulfoglycolytic organisms because SQ is seldom found as the free sugar in nature; it must be liberated from ubiquitous SQ glycosides like SQDG, lyso-SQDG, or SQGro.[6-7] Enteric organisms, like *E. coli*, are most likely to encounter SQGro because SQDG is rapidly delipidated by lipases in the mammalian GI tract.[8] A sole report has described the ability of a soil-derived Flavobacterium species to utilize the methyl α-glycoside of SQ (MeSQ) as sole carbon source;[9] the ability of *E. coli*, or all other sulfoglycolytic organisms, to use SQ glycosides as a carbon source for growth has not been studied.

Structural and biochemical studies of the *E. coli* SQase (*Ec*YihQ) revealed that it is a stereochemically-retaining glycoside hydrolase that utilizes a classical Koshland retaining mechanism involving a catalytic nucleophilic carboxylate (D405) and acid/base (D472) residue.[5] The protein adopts a fold similar to other members of family GH31 but possesses unique active site residues that recognize the characteristic sulfonate of the substrate (Fig. 1b). In particular, all three oxygens of the SQ sulfonate moiety were involved in polar interactions with either R301, W304 or Y508 (RWY; Tyr through a well-ordered water molecule). Collectively these residues constrain the anionic sulfonate group so as to not

impede the approach of the negatively-charged catalytic nucleophile to the anomeric center of the substrate. This sulfonate-binding triad is not strictly conserved among predicted SQases: for example, predicted plant SQases possess a QWY motif and mutagenesis of the *Ec*YihQ RWY sulfonate-binding motif to a QWY motif provides a competent SQase, supporting the annotation of these plant proteins as SQases.[5] Beyond the sulfonate-binding motif, many putative SQases also possess a Gln residue (Q288 in *Ec*YihQ; QRWY motif) that interacts with the 4-hydroxyl group of SQ (Fig. 1b), while others have an Glu residue at this position (ERWY motif). The Q288E mutant of *Ec*YihQ possesses little SQase activity, alluding to an incomplete understanding of substrate recognition by SQases and casting doubt on the assignment of ERWY-motif enzymes as SQases. Defining the essential features of SQases will facilitate the confident identification of sulfoglycolytic pathways/organisms, and their place in the sulfur cycle, using (meta)genomic approaches.

Here, we explore the substrate preferences of SQases with a QRWY substrate-binding motif and an ERWY motif, using both natural and unnatural derivatives of SQ. Both enzymes have selectivity for SQ glycosides and demonstrate a preference for the natural diastereomer of SQGro, which proved to be superior to SQ as a carbon source for *E. coli* growth. By solving the structure of these different enzymes bound to an aza-sugar (IFGSQ), we identified a fifth residue that defines the substrate-binding motif. A thorough mutagenic, structural and bioinformatic analysis revealed the co-evolutionary relationships between SQ-recognizing residues and revealed the presence of other co-evolutionarily related residues, distal to the active site, that have played a role in the evolution of SQases within the GH31 enzyme family.

<Figure 1>

Results

SQases enable sulfoglycolytic utilization of SQGro

Pioneering work demonstrated that *E. coli* K-12 can utilise SQ as its sole carbon source,[3] yet the preponderance of SQGro in the lower gastrointestinal tract and the conserved presence of SQases in sulfoglycolytic gene clusters suggests that *E. coli* is probably better adapted to using SQGro as sole carbon source, though this remains unproven. Indeed, *E. coli* may exhibit better growth with SQGro because SQase-mediated hydrolysis provides equimolar glycerol, which is also a viable substrate. To test this hypothesis we synthesized SQGro, from allyl α-D-glucopyranoside,[10] as a 11:9 mixture of 2'*R* and 2'*S* diastereoisomers and confirmed

that *Ec*YihQ could cleave both diastereoisomers by monitoring hydrolysis by ^{1}H NMR spectroscopy (Fig. 2a). While both diastereoisomers were hydrolysed by the enzyme, the 2'*R* stereoisomer, which corresponds to the natural stereochemistry of SQDG, hydrolyzed 6-fold faster than the 2'*S* stereoisomer (Fig. 2b); notably, there are no SQase homologs of YihQ within *E. coli*, and the upregulation of YihQ expression upon growth on SQ[3] strongly suggests that all sulfoquinovosidase activity can be ascribed to YihQ. The growth curves of *E. coli* strain BW25113 adapted for minimal media with 4 mM SQ, SQGro, D-glucose (Glc) or glycerol (Gro) as sole carbon source were determined and compared. Cultures grown on SQ grew to a similar optical density (OD_{580}) as cultures grown on Gro, and to approximately half the OD_{580} of cultures grown on Glc or SQGro (Fig. 2c). The similar cell densities obtained for Glc and SQGro, and Gro and SQ suggest that these pairs provide similar amounts of carbon to *E. coli*, commensurate with SQ and Gro yielding one three-carbon metabolite per molecule and Glc providing two. Quantitative analysis for the sulfo-EMP by-product DHPS in the spent culture media of *E. coli* grown on 4 mM SQGro revealed the by-product concentration to be 3.96 mM and that complete hydrolysis and catabolism of SQGro had occurred (Fig. 2d, Supplementary Fig. 1). Furthermore, relative growth rates on Glc, Gro, SQ, and SQGro were 0.11, 0.045, 0.034, and 0.086 h^{-1}, respectively, demonstrating that SQGro is preferred in this medium to both SQ and Gro. Collectively, these data reveal that *E. coli* can utilize SQGro as a sole carbon source, enabled by endogenous SQase activity, and that it metabolizes the liberated Gro and SQ fragments faster than if they are individually present in the medium. Interestingly, SQMe also supported growth of *E. coli*, demonstrating tolerance for this simple aglycon (data not shown).

<Figure 2>

Variations in the SQase substrate-binding motif

Previous structural and mutagenic studies of *E. coli* YihQ identified the RWY (or in the case of plants, QWY) sulfonate-binding motif as being crucial for SQase activity and enabled reclassification of some proteins within GH family 31 as putative SQases. Beyond this sulfonate-binding motif, a fourth residue attracted our attention: *Ec*YihQ Q288. While many putative SQases possess a Gln at this position, others have a Glu residue and, intriguingly, the *Ec*YihQ Q288E mutant has little SQase activity.[5] We sought to validate that enzymes with the ERWY substrate-binding motif were *bona fide* SQases like those with the QRWY motif, and elucidate the sequence or structural context behind this discrepancy. The putative SQase

PpSQ1_00094 from *P. putida* SQ1, which possess a characterized sulfo-ED pathway, has an ERWY motif but it failed to yield useful amounts of soluble protein in an *E. coli* expression system. *Agrobacterium tumifaciens* has been reported to possess sulfoglycolytic capacity, and is able to grow on SQ as sole sulfur source.[11] WP_035199431 (hereafter *At*SQase), a putative SQase with an ERWY motif from *A. tumefaciens*, expressed well in *E. coli* to provide useful quantities of high-quality protein (Supplementary Fig. 2). *At*SQase exhibited high specificity for 4-nitrophenyl α-D-sulfoquinovoside (PNPSQ) (k_{cat} = 22.3 ± 0.6 s^{-1}, K_M = 0.21 ± 0.03 mM, k_{cat}/K_M = (1.1 ± 0.1) × 10^5 M^{-1} s^{-1}), with no detectable activity towards 4-nitrophenyl α-D-glucopyranoside (PNPGlc) under comparable conditions (Supplementary Table 2), and had a pH optimum of 8.0 (Supplementary Fig. 3), similar to *Ec*YihQ. *At*SQase hydrolyzed both epimers of SQGro, with a preference for the natural 2'*R* isomer (Supplementary Fig. 3) as had been observed for *Ec*YihQ. The substrate specificity of both *Ec*YihQ and *At*SQase was further explored using a panel of modified substrates. We synthesized substrate analogues to explore the importance of stereochemistry and the nature of the charged group: 4-nitrophenyl α-D-sulfofucoside (PNPSFuc) and 4-nitrophenyl α-D-sulforhamnoside (PNPSRha) are epimers of PNPSQ, while 4-nitrophenyl α-D-glucuronoside (PNPGlcA) has a carboxylate moiety instead of the sulfonate of SQ (Fig. 3). No activity was detected for either enzyme on these substrates, revealing a high specificity for the correct D-*gluco* configuration and sulfonate of SQ. The lack of activity on PNPGlcA is noteworthy considering that various plants produce α-glucuronosyl diglycerides under conditions of phosphate starvation that together with SQDG appear to compensate for reduced levels of phosphatidyl glycerols.[12] Likewise, sulfofucose has been detected in a cell surface glycoprotein from *Thermoplasma acidophilum.*[13] Collectively these data illustrate the many functional similarities between *At*SQase and *Ec*YihQ and confirm the reliability of SQase annotation based solely on the presence of the Q/RWY sulfonate-binding motif.

<Figure 3>

Second-shell amino acid variations in *E. coli* and *A. tumefaciens* SQases

Structural studies were performed on *At*SQase and *Ec*YihQ to gain insights into how the Q288E mutation disables the activity of *Ec*YihQ while a Glu residue at the corresponding position in *At*SQase provides for excellent catalytic activity. Despite extensive crystallisation screening, no crystallization conditions could be identified. Guided by the Surface Entropy

Reduction prediction (SERp) server,[14] *At*SQase was mutated to E370A/E371A double surface mutant. This mutant yielded several crystal conditions with better diffraction quality and higher resolutions. The 3D structure of *At*SQase was determined by molecular replacement using the previously determined structure of *Ec*YihQ, and revealed a fold essentially identical to *Ec*YihQ (Fig. 4a). In order to illuminate the molecular basis of substrate binding, we determined structures of *At*SQase with ligands bound in the active site. To obtain a complex with substrate, we mutated the acid/base carboxylate D455 to obtain a catalytically inactive variant, *At*SQase-D455N, which was determined in complex with PNPSQ at 1.97 Å (Fig. 4b, Supplementary Fig. 5a). To ensure that the active site structure had not been appreciably perturbed by the mutation, we also sought a ligand complex with wild-type enzyme. To this end, we synthesized the aza-sugar IFGSQ. IFGSQ bound to *Ec*YihQ with $K_D = 0.96 \pm 0.12$ μM and to *At*SQase with K_D of 6.8 ± 0.2 μM (Supplementary Fig. 4). Structures of IFGSQ bound to *At*SQase and *Ec*YihQ were determined to resolutions of 1.77 and 1.87 Å, respectively (Fig. 4c,d, Supplementary Fig. 5b,c). Both complexes revealed binding of the sulfonate residue with RWY motifs in essentially identical manners to that seen for PNPSQ in the pseudo-Michaelis complexes with the acid/base mutants of *Ec*YihQ and *At*SQase, involving direct hydrogen bonding by Arg and Trp, and a bridging water molecule with Tyr. Both the Trp and Tyr residues are involved in multiple π interactions within the protein and with the substrate, while the Arg residue participated in ionic interactions with the sulfonate group of the substrate (Supplementary Fig. 6). Previously we showed that substitutions at the Trp and Tyr caused a dramatic loss in enzyme activity;[5] *in silico* analysis supported these observations with substitutions at these positions predicted to be energetically unfavourable, due to protein destabilisation and reduction in ligand affinity.[15] Computational docking using Autodock of 2'*R*-SQGro into the structures of each enzyme yielded poses in which binding of the sugar ring and the sulfonate group was conserved compared with that seen for PNPSQ and IFGSQ, and identified possible binding poses of the glyceryl moiety (Supplementary Fig. 7).

<Figure 4>

Comparison of complexes of *Ec*YihQ and *At*SQase wild-type with IFGSQ and complexes of their acid/base mutants with PNPSQ reveal that for *At*SQase, E270 interacts with the 4-hydroxyl of the substrate/IFGSQ in a similar fashion to the equivalent residue

Q288 in *Ec*YihQ. A key difference between structures lay in the second shell of residues that surround the active site residues: in *Ec*YihQ the active site Q288 is in contact with Q262 in the second shell, whereas in *At*SQase, E270 is in contact with K245 of the second shell; in each case these comprise an overall neutral pair. To explore whether the 'neutral' Q288/Q262 pairing in *Ec*YihQ and the E270/K245 pairing in *At*SQase are required for catalysis, we undertook a series of stepwise mutational studies in which we interconverted the KE and QQ pairings in the two enzymes (Fig. 5b). In the case of *Ec*YihQ enzyme, the active-site Q288E variant resulted in ≈1000-fold loss of activity in terms of k_{cat}/K_M relative to wildtype. A similar loss in activity was observed for the second shell Q262K mutant. Remarkably, the double mutant Q288E/Q262K exhibited a recovery of activity relative to the individual mutants of around 10-fold, being only ≈100-fold less active than wildtype. While this recovery of activity is imperfect, it demonstrates the importance of this neutral pair, and second shell residues, for SQase activity. The equivalent series of mutations were conducted for *At*SQase. The E270Q and K245Q mutants suffered >1000-fold reductions in k_{cat}/K_M values, whereas the K245Q/E270Q double mutant, recovered greater than 10-fold activity relative to the single mutants, again demonstrating the importance of pairing these residues and the role that second shell residues play in facilitating catalysis.

<Figure 5>

Pairwise co-evolutionary relationships of residues in sulfoquinovosidases

Multiple sequence alignments provide information regarding residue conservation and variation over evolution, giving information on inter-relationships between residues. A multiple sequence alignment of putative SQases identified using the RWY sulfonate-binding motif (Fig. 7) was constructed and revealed that most sequences possessed either the QQ and KE pairs identified by our structural and mutagenesis studies (Figure 5a). The mutual co-evolutionary relationship between two positions can be quantified using mutual information (MI) theory.[16] We applied the average product correction method to identify co-evolving pairs in SQases. The QQ and KE pairs, with MI scores of 12.1 and 12.1, respectively, have a strong co-evolutionary relationship (Fig. 6).

<Figure 6>

Moving outwards from the protein active site, two other co-evolving residue pairs were identified with MI scores >8 for both *E. coli* and *A. tumefaciens* SQases. The M468-F181 and L451-N164 pairing exhibit very strong co-evolution signals with MI scores of 13.7 and 15.0, respectively, while the A46-F51 and E36-Y41 pairing exhibit slightly weaker MI scores of 11.0 and 8.6, respectively. In these two pairing cases the residues are located > 10 Å away from one another and do not directly interact with substrate (Fig. 6).

Co-evolutionary relationships between amino acid residues within proteins may arise from selective pressures on functional and physicochemical factors. In order to understand how the identified co-evolving pairs may influence the function and properties of SQases, we mapped their physical locations onto the X-ray structure of the two SQases with PNPSQ bound and used *in silico* methods to evaluate the structural effects of their mutation. The energetic penalties for modelling of individual amino acid mutations within each pairing was obtained by applying the mCSM-lig,[17] and DynaMut[18] and SDM[19] methods, which predict effects of mutations upon ligand affinity and protein stability, respectively. Mutation of individual residues at each position with the co-evolutionarily linked pairs distal to the active site lead to energetic penalties that are compensated for by mutation at the paired site (Supplementary Table 4). On the other hand, application of this analysis to the Q262-Q288/K245-E270 pairing proximal to the active site reveals that individual mutation at each position results in an energetic penalty for ligand binding. Consistent with our mutagenesis results, the effects oppose each other, such that mutations at each site compensate for ligand binding affinity and protein stability.

Discussion and conclusions

The prokaryotic sulfo-EMP and sulfo-ED pathways play a significant role in the global sulfur cycle as the first sequence of events in the biomineralization of SQ, a major reservoir of organic sulfur. To date these pathways have only been studied in the context of their ability to degrade SQ, yet bacteria more commonly encounter SQDG, or its delipidated forms lyso-SQDG and SQGro, and rely on SQases to liberate SQ from these substrates. Our data reveal that SQases preferentially act on the natural 2'*R*-diastereomer of SQGro and that *E. coli*, which possesses a sulfo-EMP pathway, actually prefers to grow on SQGro rather than SQ. Growth on SQGro occurs at comparable rates to that on Glc. Together with the release of a stoichiometric quantity of DHPS, these data suggest that each molecule of SQGro yields two three-carbon metabolites for primary metabolism: one from SQ and one from Gro. In this regard, SQGro should be broadly equivalent to Glc as a source of carbon and energy for *E.*

coli, in line with predictions made on pyruvate, ATP and NADH yields (Supplementary Table 5).[2] However, it should be noted that growth on SQ or SQGro likely involves gluconeogenesis, whereas this is not required for growth on Glc. Furthermore, the superior growth rate for SQGro relative to SQ and Gro suggests that SQGro is a preferred substrate for the transporter that imports these substrates into *E. coli* and highlights the need for future studies of SQ catabolism to appreciate that SQGro is the substrate that microbes encounter and utilize.

Because SQase mediated hydrolysis of SQ glycosides is the indispensable first step in sulfoglycolytic pathways, these enzymes are promising markers for probing which organisms in a given environmental niche are responsible for processing the biosulfur assimilated into SQDG, a significant arm of the biosulfur cycle. Our early studies of SQases identified the RWY motif as important for structural recognition of the sulfonate group of SQ, and a potentially useful signature for identifying SQases. However, variations in other substrate-binding residues, combined with conflicting biochemical mutagenesis data, limited the certainty of predictions based solely on the RWY motif. To address this limitation, we expressed *At*SQase, a putative SQase with a different substrate-binding motif to *Ec*YihQ, and demonstrated that its properties are essentially identical to *Ec*YihQ: both are highly specific for the stereochemistry and charge of SQ glycosides.

Structural analyses of *Ec*YihQ and *At*SQase bound to substrate analogues and an iminosugar (IFGSQ – the first aza-sugar targeting SQases to be reported) were conducted to determine why the Q288E mutation in *Ec*YihQ greatly attenuated SQase activity when the corresponding residue in *At*SQase, E270, is a Glu. The structures revealed that *Ec*YihQ Q288 and *At*SQase E270 occupy identical positions in the active site, both hydrogen bonding to O4 of the SQ substrate. An important difference was noted in the second shell of residues are the active site: *Ec*YihQ Q288 hydrogen bonded to Q262, while the charge of *At*SQase E270 was paired with K245, leading to the hypothesis that these residue pairs were important to defining SQase activity. An extensive kinetic analysis of single- and double-mutant enzymes revealed that the Q288/Q262 and E270/K245 pairings are essential for the activity of these two SQases.

In order to understand whether the requirement of the Q288/Q262 and E270/K245 pairings applies more widely to all SQases, we constructed an alignment of putative SQases based on the presence of the RWY sulfonate-binding motif (Fig. 7) and quantified the prevalence of the QQ and KE pairs (Figure 5a). This alignment revealed a strong conservation of the aromatic residues of the motif (Trp, Tyr), with slightly less stringency for

the Arg residue. While greater variation is seen at the first and second shell positions corresponding to the QQ and KE pairs, the majority of sequences possessed one pair or the other, alluding to a strong co-evolutionary relationship between residues throughout SQase evolution (Fig. 5a, Fig. 7).

Mutual information analyses confirmed the strong co-evolutionary relationship between these residues in these pairs, and predicted that the co-evolution of these residues is important for ligand binding. Other strongly correlated co-evolutionary pairings were identified in the SQases at locations distal to the active site; these are predicted to play a role in maintaining protein stability.

The essential features of SQases reported here (a well-conserved sulfonate-binding Q/RWY motif and the presence of co-evolved residue pairs, one of which is essential for SQase activity) provide the means to confidently annotate SQases, and because of the role of these enzymes in SQ glycoside catabolism, provide a means to identify sulfoglycolytic organisms and perhaps even discover new catabolic pathways.

Acknowledgements

Australian Research Council (FT130100103, DP180101957), the European Research Council (ERC-2012-AdG-322942), the Ramaciotti Foundation and VESKI with additional support from the Australian Cancer Research Foundation and Victorian State Government Operational Infrastructure Support, NHMRC IRIISS grant 9000220. GJD is supported by the Royal Society and this work underpins recent Leverhulme Trust funding (RPG-2017-190). We acknowledge the staff of the Diamond Light Source (UK) for provision of I02 beamline facilities (proposal number mx-9948), and Amicus Therapeutics for a gift.

Author contributions

P.A. and M.P. conducted kinetic assays. J.P.L. and A.J. cloned, expressed and purified proteins. E.R. analyzed DHPS production. M.P. and J.P.L. conducted microbial growth assays. E.D.G.-B., D.B.A. and D.E.V.P. performed bioinformatic analysis. Y.J. conducted structural studies. M.P., J.W.Y.M. and P.A. synthesized chemical reagents. Experiments were designed and interpreted by D.A., G.J.D., E.D.G.-B. and S.J.W. All authors contributed to preparing this manuscript. Mr Christopher Bengt is thanked for technical contributions.

Supporting information

Supplementary figures (1-7), Supplementary tables (1-5), all experimental details, X-ray data collection, processing and refinement statistics (PDF)

References

1. Harwood, J. L.; Nicholls, R. G., The plant sulpholipid - a major component of the sulphur cycle. *Biochem. Soc. Trans.* **1979,** *7*, 440-447.
2. Goddard-Borger, E. D.; Williams, S. J., Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem. J.* **2017,** *474*, 827–849.
3. Denger, K.; Weiss, M.; Felux, A. K.; Schneider, A.; Mayer, C.; Spiteller, D.; Huhn, T.; Cook, A. M.; Schleheck, D., Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **2014,** *507*, 114-117.
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* **2015,** *112*, E4298-305.
5. Speciale, G.; Jin, Y.; Davies, G. J.; Williams, S. J.; Goddard-Borger, E. D., YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **2016,** *12*, 215-217.
6. Shibuya, I.; Hase, E., Degradation and formation of sulfolipid occurring concurrently with de- and re-generation of chloroplasts in the cells of *Chlorella protothecoides*. *Plant Cell Physiol.* **1965,** *6*, 267-283.
7. Yagi, T.; Benson, A. A., Plant sulfolipid. V. Lysosulfolipid formation. *Biochim. Biophys. Acta* **1962,** *57*, 601-3.
8. Andersson, L.; Bratt, C.; Arnoldsson, K. C.; Herslof, B.; Olsson, N. U.; Sternby, B.; Nilsson, A., Hydrolysis of galactolipids by human pancreatic lipolytic enzymes and duodenal contents. *J. Lipid Res.* **1995,** *36*, 1392-400.
9. Martelli, H. L.; Benson, A. A., Sulfocarbohydrate metabolism. I. Bacterial production and utilization of sulfoacetate. *Biochim. Biophys. Acta* **1964,** *93*, 169-171.
10. Miyano, M.; Benson, A. A., The plant sulfolipid. VII. Synthesis of 6-sulfo- : -D-quinovopyranosyl-(1⊐1⍰)-glycerol and radiochemical syntheses of sulfolipids. *J. Am. Chem. Soc.* **1962,** *84*, 59-62.
11. 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.* **2003,** *69*, 6434-6441.
12. Okazaki, Y.; Otsuki, H.; Narisawa, T.; Kobayashi, M.; Sawai, S.; Kamide, Y.; Kusano, M.; Aoki, T.; Hirai, M. Y.; Saito, K., A new class of plant lipid is essential for protection against phosphorus depletion. *Nat. Commun.* **2013,** *4*, 1510.
13. Vinogradov, E.; Deschatelets, L.; Lamoureux, M.; Patel, G. B.; Tremblay, T. L.; Robotham, A.; Goneau, M. F.; Cummings-Lorbetskie, C.; Watson, D. C.; Brisson, J. R.; Kelly, J. F.; Gilbert, M., Cell surface glycoproteins from *Thermoplasma acidophilum* are modified with an N-linked glycan containing 6-C-sulfofucose. *Glycobiology* **2012,** *22*, 1256-67.
14. Goldschmidt, L.; Cooper, D. R.; Derewenda, Z. S.; Eisenberg, D., Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* **2007,** *16*, 1569-76.
15. Pires, D. E.; Chen, J.; Blundell, T. L.; Ascher, D. B., In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci. Rep.* **2016,** *6*, 19848.
16. Buslje, C. M.; Santos, J.; Delfino, J. M.; Nielsen, M., Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. *Bioinformatics* **2009,** *25*, 1125-1131.

17. Pires, D. E.; Blundell, T. L.; Ascher, D. B., mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.* **2016,** *6*, 29575.

18. Rodrigues, C. H.; Pires, D. E.; Ascher, D. B., DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* **2018,** *46*, W350-W355.

19. Pandurangan, A. P.; Ochoa-Montano, B.; Ascher, D. B.; Blundell, T. L., SDM: a server for predicting effects of mutations on protein stability. *Nucl. Acids Res.* **2017,** *45*, W229-W235.

Fig. 1 Role of sulfoquinovosidases (SQases) in allowing sulfoglycolytic utilization of sulfoquinovose glycosides. **a** Sulfoglycolysis pathways in bacteria highlighting proposed role of sulfoquinovidases. **b** Cartoon of active site residues involved in binding PNP-SQ, from the X-ray structure of *E. coli* YihQ D472N.

Fig. 2 Sulfoquinovosyl glycerol (SQGro) is a superior substrate to sulfoquinovose (SQ) for growth of *E. coli*. **a** NMR time course of hydrolysis of SQGro hydrolysis by *E. coli* YihQ. **b** Rates of consumption of individual SQGro diastereoisomers by YihQ. **c** Growth of *E. coli* BW25113 in M9 minimal media containing 4 mM Glc, Gro, SQGro or SQ as sole carbon source at 30 °C. **d** MS/MS spectrum of DHPS produced in culture media of *E. coli* grown on SQGro.

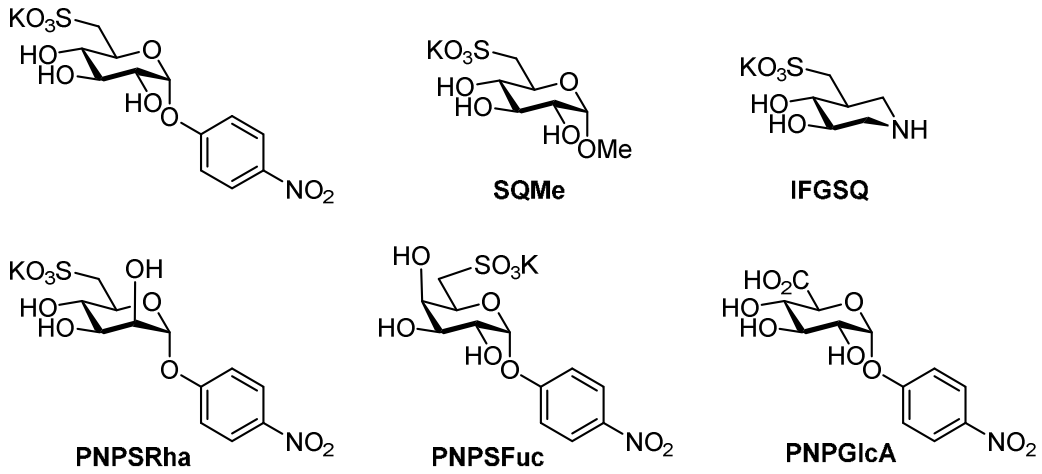


Fig. 3 Structures of SQ-derived substrates, ligands and analogues.

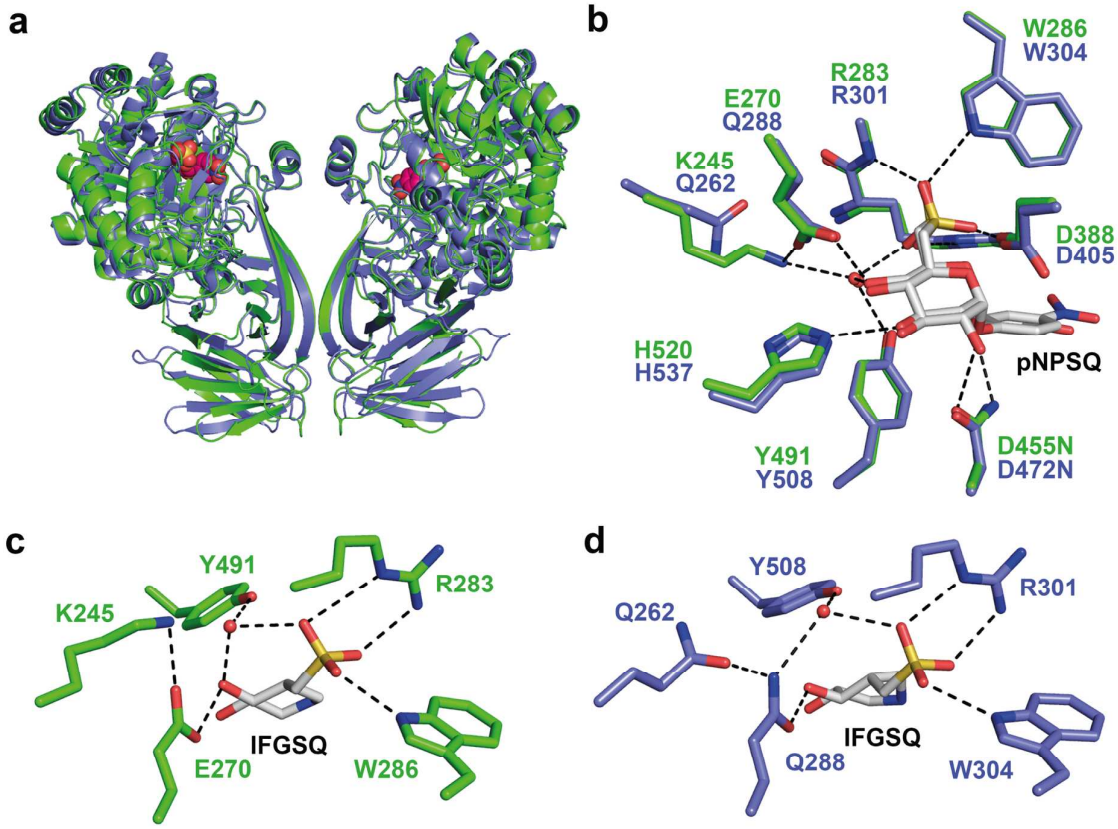
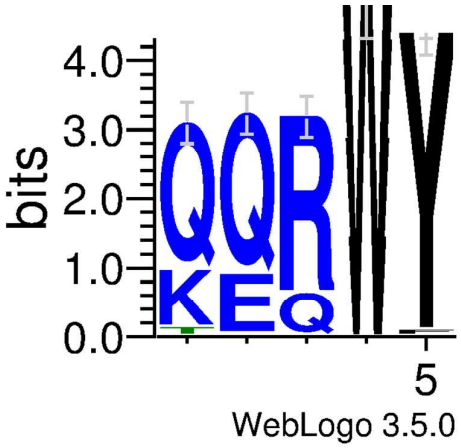


Fig. 4 Structural basis of SQ recognition by SQases. **a** Overlay of *Ec*YihQ and *At*SQase. **b** comparison of Michaelis complexes of acid/base mutants of *Ec*YihQ and *At*SQase. **c** IFGSQ bound to AtSQase. **d** IFGSQ bound to *Ec*YihQ. For electron density maps see Supplementary Figure 5.

a



b

Enzyme	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
*At*SQase	0.212 ± 0.025	22.3 ± 0.61	(1.05 ± 0.10) × 10^5
K245Q	1.64 ± 0.23	(2.57 ± 0.14) × 10^{-2}	(1.57 ± 0.16) × 10^1
E270Q	1.29 ± 0.22	(2.28 ± 0.16) × 10^{-2}	(1.76 ± 0.20) × 10^1
K245Q/E270Q	4.45 ± 0.39	1.89 ± 0.08	(4.25 ± 0.22) × 10^2
YihQ	0.150 ± 0.014	32.7 ± 0.59	(2.18 ± 0.18) × 10^5
Q262K	2.07 ± 0.14	(2.12 ± 0.06) × 10^{-1}	(1.02 ± 0.04) × 10^2
Q288E	- [a]	- [a]	(4.81 ± 0.29) × 10^2
Q288E/Q262K	4.28 ± 0.32	11.2 ± 0.38	(2.61 ± 0.12) × 10^3

Fig. 5 a Sequence logo highlighting relative proportions of different residues found at each position within the QQRWY/KERWY motif of SQases, using the 84 sequences of Fig. 7. **b** Kinetic analysis of mutants investigating the effect of stepwise variation of QQ/KE sequence of *Ec*YihQ and *At*SQase.

[a] Saturation was not reached.

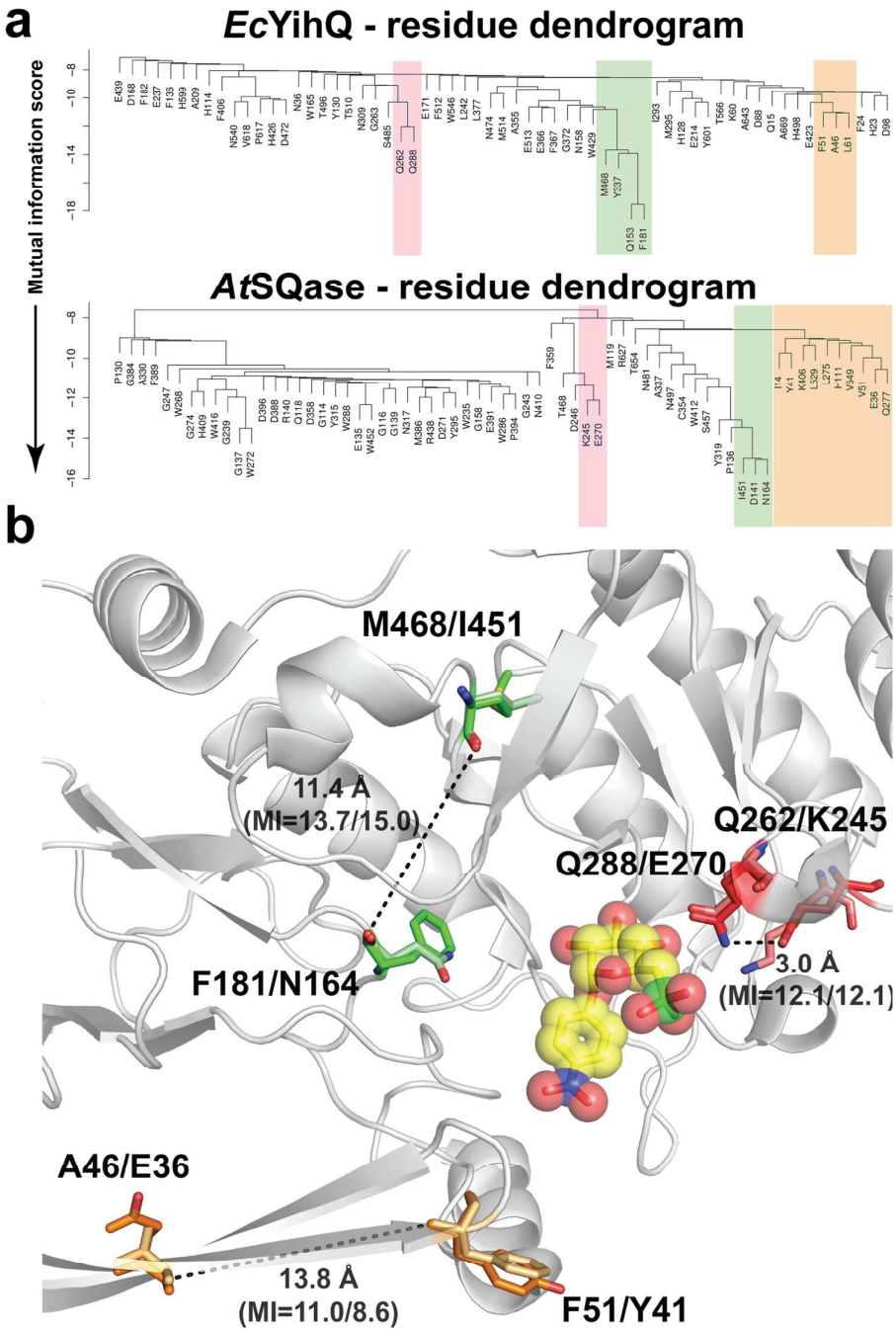


Fig. 6 a Dendrogram of interrelationships between sequence positions of *Ec*YihQ and *At*SQase. Co-evolving groups are highlighted in colored boxes. **b** Spatial distribution of three pairs of co-evolving residues on the 3D structures. Residues identified by MISTIC based on mutual information are presented in similar colors. Residue 451 exhibits natural variation in NCBI/RefSeq entries WP_010972911.1 (Ile; used for the X-ray structure herein) and WP_035199431.1 (Leu).

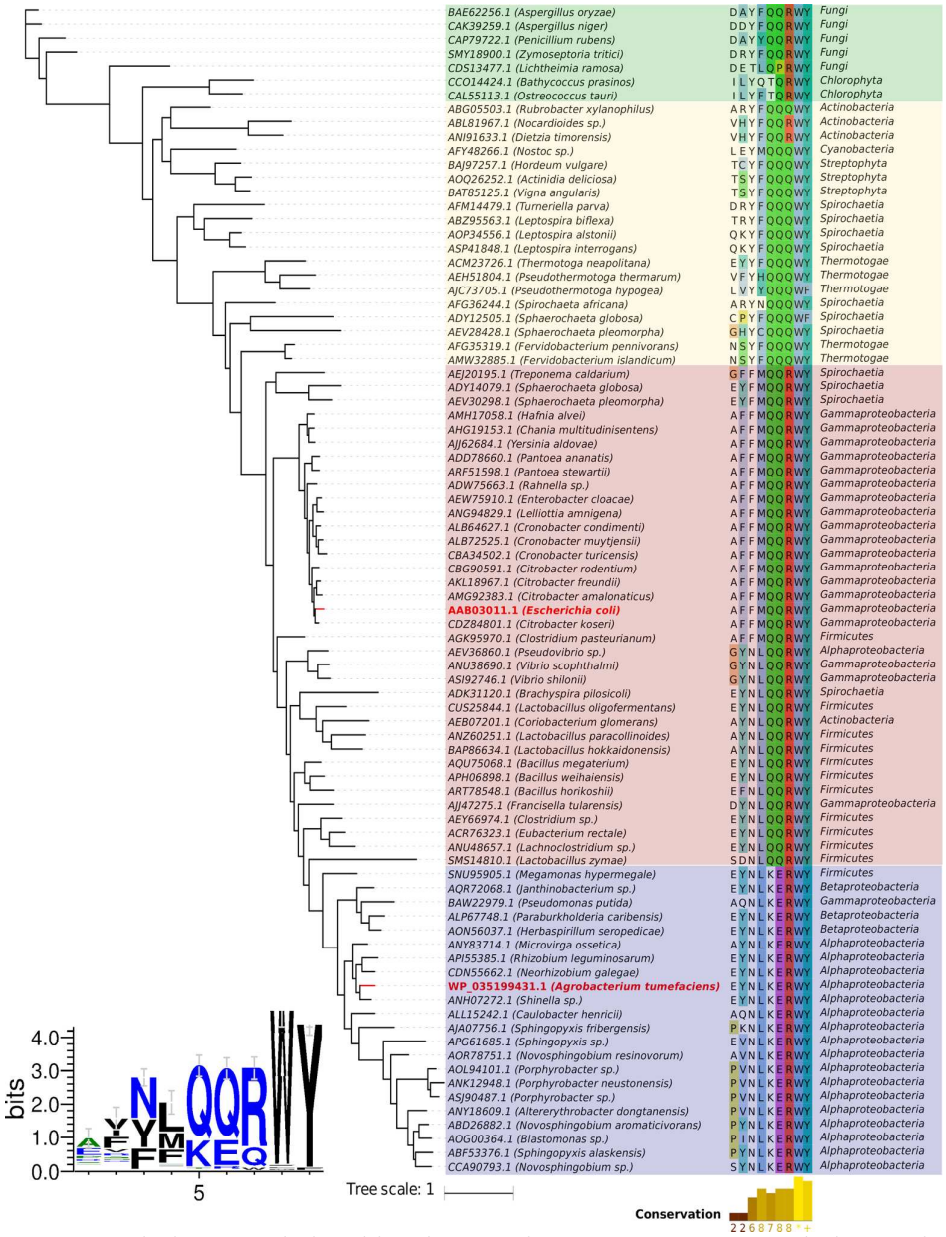
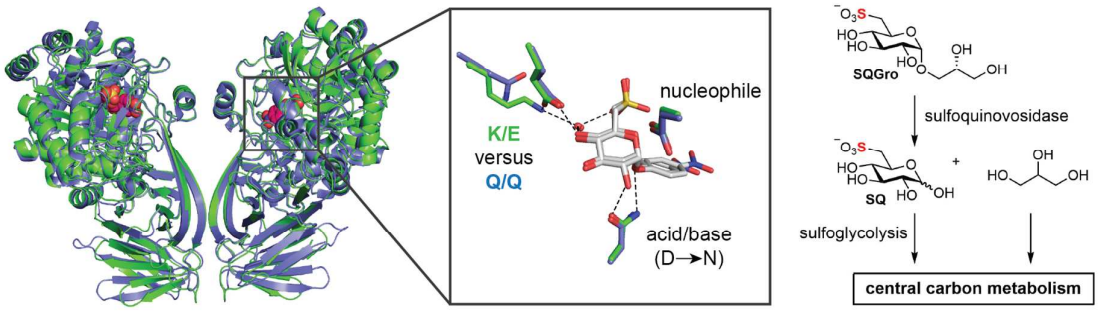


Fig. 7 Evolutionary relationships for putative SQases. **Right** A phylogenetic tree of putative SQases obtained via multiple sequence alignment presenting a conserved KERWY/QQRWY motif. The alignment of the motif region is depicted together with the positions of the other two co-evolving residue pairs identified. Organism taxonomy (class level) is also depicted. Sequences were highlighted by coloured boxes based on motif conservation in two main groups: in blue those that presented the KERWY motif and in red those that presented the QQRWY motif. The yellow box groups sequences that in general do not present the arginine conserved and the remaining sequences from plants and fungi were grouped in green. **Left** A sequence logo of the KERWY/QQRWY motif supplemented with the aforementioned co-evolving pairs. Figure generated with WebLogo 3.5.0.

Graphical abstract



Synopsis

We identify evolutionarily-conserved active-site residues within sulfoquinovosidases, gateway enzymes that cleave glycosides of sulfoquinovose and provide entry into sulfoglycolytic metabolism.