Structural and Biochemical Insights into the Function and Evolution of Sulfoquinovosidases

Palika Abayakoon,† Yi Jin,‡ James P. Lingford,§∥ Marija Petricevic,† Alan John,§∥ Eileen Ryan,† Janice Wai-Ying Mui,† Douglas E.V. Pires,* David B. Ascher,* Gideon J. Davies,* Ethan D. Goddard-Borger,*§∥ and Spencer J. Williams*†

† School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia
‡ York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington YO10 5DD, United Kingdom
§ ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3010, Australia
∥ Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia

Supporting Information

ABSTRACT: An estimated 10 billion tonnes of sulfoquinovose (SQ) are produced and degraded each year. Prokaryotic sulfglycolytic pathways catabolize 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 (Escherichia coli YihQ) have identified three conserved residues that are essential for substrate recognition, but crossover 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 toward SQ glycosides, and structural, mutagenic, and bioinformatic analyses revealed the molecular coevolution 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-sulfoquinovosyldiacylglycerol (SQDG), which is one of the most abundant organic sulfur compounds in nature. 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. 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 sulfglycolytic processes have been identified, termed the sulfo-Embden–Meyerhof–Parnas (sulfo-EMP) and sulfo-Entner–Doudoroff (sulfo-ED) pathways (Figure 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 sulphoglycolytic enzymes and includes a glycoside hydrolase (GH) from family 31 of the carbohydrate-active enzyme.
The corresponding enzyme from the Escherichia coli sulfo-EMP pathway (EcYihQ) was recently shown to be a dedicated sulfoquinovosidase (SQase) capable of hydrolyzing SQDG or α-D-sulfoquinovosyl glycerol (SQGro). 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. Enteric organisms, like E. coli, are most likely to encounter SQGro because SQDG is rapidly delipidated by lipases in the mammalian GI tract. 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; 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 (EcYihQ) 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. 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 (Figure 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 EcYihQ RWY sulfonate-binding motif to a QWY motif provides a competent SQase, supporting the annotation of these plant proteins as SQases. Other than the sulfonate-binding motif, many putative SQases also possess a Glu residue at this position (Q288 in EcYihQ; QRWY motif) that interacts with the 4-hydroxyl group of SQ (Figure 1b), while others have a Gln residue at this position (ERWY motif). The Q288E mutant of EcYihQ possesses little SQase activity, alluding to an incomplete understanding of substrate recognition by SQases and casting doubt on the assignment of ERWY-motif proteins as SQases. Defining the essential features of SQases will facilitate the confident identification of sulfoglucolytic 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 corresponds to the natural stereochemistry of SQDG, and that complete hydrolysis and catabolism of SQGro had occurred (Figure 2d; Supplementary Information, Figure S1). 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).

**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: EcYihQ Q288. While many putative SQases possess a Glu at this position, others have a Gln residue, and intriguingly, the EcYihQ Q288E mutant has little SQase activity. We sought to validate that enzymes with the ERWY substrate-binding motif were bona fide SQases like those with the QWY motif, and elucidate the sequence or structural context behind this discrepancy. The putative SQase PpSQ1_00094 from Pseudomonas putida SQ1, which possesses 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 tumefaciens, expressed well in E. coli to provide useful quantities of high-quality protein (Supporting Information, Figure S2). AtSQase exhibited high specificity for 4-nitrophenyl α-d-sulfofucoside (PNPSFuc) and 4-nitrophenyl α-d-glucofucoside (PNPGlcA) (Figure 2a), a putative SQase with an ERWY motif from A. tumefaciens, expressed well in E. coli to provide useful quantities of high-quality protein (Supporting Information, Table S2), and had a pH optimum of 8.0 (Supporting Information, Figure S3), similar to EcYihQ. AtSQase hydrolyzed both epimers of SQGro, with a preference for the natural 2’R isomer (Supporting Information, Figure S3) as had been observed for EcYihQ. The substrate specificity of both EcYihQ and AtSQase 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-sulfofuranoside (PNPSRha) are epimers of PNPSQ, while 4-nitrophenyl α-d-glucuronoside (PNPGlcA) has a carboxylate moiety instead of the sulfonate of SQ (Figure 3). No activity was detected for either enzyme on these substrates, which have a high specificity for the correct α-glucosidic 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. Likewise, sulfoglucose has been detected in a cell surface glycoprotein from Thermoplasma acidophilum. Collectively these data illustrate the many functional similarities between AtSQase and EcYihQ and confirm the reliability of SQase
annotation based solely on the presence of the Q/RWY sulfonate-binding motif.

**Second-Shell Amino Acid Variations in *E. coli* and *A. tumefaciens* SQases.** Structural studies were performed on AtSQase and EcYihQ to gain insights into how the Q288E mutation disables the activity of EcYihQ while a Glu residue at the corresponding position in AtSQase provides for excellent catalytic activity. Despite extensive crystallization screening, no crystallization conditions could be identified. Guided by the Surface Entropy Reduction prediction (SERp) server, AtSQase was mutated to E370A/E371A double surface mutant. This mutant yielded several crystal conditions with good diffraction quality and higher resolutions. The 3D structure of AtSQase was determined by molecular replacement using the previously determined structure of EcYihQ, and revealed a fold essentially identical to EcYihQ (Figure 4a). To illuminate the molecular basis of substrate binding, we determined structures of AtSQase 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, AtSQase-D455N, which was determined in complex with PNPSQ at 1.97 Å (Figure 4b; Supporting Information, Figure S5a). 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 EcYihQ with $K_D = 0.96 \pm 0.12 \mu M$ and to AtSQase with $K_D$ of 6.8 ± 0.2 μM (Supporting Information, Figure S4). Structures of IFGSQ bound to AtSQase and EcYihQ were determined to resolutions of 1.77 and 1.87 Å, respectively (Figure 4c,d; Supporting Information, Figure S5b,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 EcYihQ and AtSQase, 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 $\pi$ interactions within the protein and with the substrate, while the Arg residue participated in ionic interactions with the sulfonate group of the substrate (Supporting Information, Figure S6). Previously we showed that substitutions at the Trp and Tyr residues caused a dramatic loss in enzyme activity; in silico analysis supported these observations with substitutions at these positions predicted to be energetically unfavorable, due to protein destabilization and reduction in ligand affinity. 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 (Supporting Information, Figure S7).

Comparison of complexes of EcYihQ and AtSQase wild-type with IFGSQ and complexes of their acid/base mutants with PNPSQ reveal that, for AtSQase, E270 interacts with the 4-hydroxyl of the substrate/IFGSQ in a similar fashion to the equivalent residue Q288 in EcYihQ. A key difference between structures lay in the second shell of residues that surround the active site residues: in EcYihQ the active site Q288 is in contact with Q262 in the second shell, whereas in AtSQase,
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 EcYihQ and the E270/K245 pairing in AtSQase 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 (Figure 5b). In the case of EcYihQ enzyme, the active-site Q288E variant resulted in \( \approx 1000 \)-fold loss of activity in terms of \( k_{\text{cat}}/K_{M} \) relative to wild-type. 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 \( \approx 100 \)-fold less active than wild-type. 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 was conducted for AtSQase. The E270Q and K245Q mutants suffered >1000-fold reductions in \( k_{\text{cat}}/K_{M} \) 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.

**Pairwise Coevolutionary Relationships of Residues in Sulfoquinovosidases.** Multiple sequence alignments provide information regarding residue conservation and variation over evolution, giving information on interrelationships between residues. A multiple sequence alignment of 84 putative SQases identified using the RWY sulfonate-binding motif (Figure 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 coevolutionary relationship between two positions can be quantified using mutual information (MI) theory.\(^{16}\) We applied the average product correction method to identify coevolving pairs in SQases. The QQ and KE pairs, with MI scores of 12.1 and 12.1, respectively, have a strong coevolutionary relationship (Figure 6).

**Figure 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 Figure 7. (b) Kinetic analysis of mutants investigating the effect of stepwise variation of QQ/KE sequence of EcYihQ and AtSQase. Footnote a: saturation was not reached.

**Figure 6.** (a) Dendrogram of interrelationships between sequence positions of EcYihQ and AtSQase. Coevolving groups are highlighted in colored boxes. (b) Spatial distribution of three pairs of coevolving 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).

Moving outward from the protein active site, two other coevolving residue pairs were identified with MI scores >8 for both E. coli and A. tumefaciens SQases. The M468-F181 and L451-N164 pairing exhibits very strong coevolution signals with MI scores of 13.7 and 15.0, respectively, while the A46-F51 and E36-Y41 pairing exhibits 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 (Figure 6).
Figure 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 coevolving residue pairs identified. Organism taxonomy (class level) is also depicted. Sequences were highlighted by colored 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 a conserved arginine, and the remaining sequences from plants and fungi were grouped in

ACS Cent. Sci. 2018, 4, 1266−1273
Coevolutionary relationships between amino acid residues within proteins may arise from selective pressures on functional and physicochemical factors. To understand how the identified coevolving 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 modeling of individual amino acid mutations within each pairing were obtained by applying the mCSM-lig,17 and DynaMut18 and SDM19 methods, which predict effects of mutations upon ligand affinity and protein stability, respectively. Mutation of individual residues at each position with the coevolutionarily linked pairs distal to the active site leads to energetic penalties that are compensated for by mutation at the paired site (Supporting Information, Table S4). 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 biosynthesis 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 lysosQDG 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 and a similar final density 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 (Supporting Information, Table S5).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 is the indispensable first step in sulfoglycolysis of SQ glycosides, 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 RWWY 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 RWWY motif. To address this limitation, we expressed AtSQase, a putative SQase with a different substrate-binding motif to EcYihQ, and demonstrated that its properties are essentially identical to EcYihQ: both are highly specific for the stereochemistry and charge of SQ glycosides.

Structural analyses of EcYihQ and AtSQase bound to substrate analogues and an azasugar (IFGSQ—the first inhibitor targeting SQases to be reported) were conducted to determine why the Q288E mutation in EcYihQ greatly attenuated SQase activity whereas the corresponding residue in AtSQase, E270, is a Glu. The structures revealed that EcYihQ Q288 and AtSQase 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 is the active site: EcYihQ Q288 hydrogen bonded to Q262, while the charge of AtSQase 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.

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 RWWY sulfonate-binding motif (Figure 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 (Tyr, Trp), 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 coevolutionary relationship between residues throughout SQase evolution (Figures 5a and 7).

Mutual information analyses confirmed the strong coevolutionary relationship between these residues in these pairs, and predicted that the coevolution of these residues is important for ligand binding. Other strongly correlated coevolutionary 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/RWWY motif and the presence of coevolved 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.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentscic8b00453.
Additional experimental details and figures including LC/MS chromatogram, polynucleotide sequence, amino acid sequence, pH dependence, Michaelis–Menten plot, NMR spectra, consumption rates, ITC calorimograms, and schematics (PDF)

■ AUTHOR INFORMATION
Corresponding Authors
*E-mail: gideon.davies@york.ac.uk.
*E-mail: goddard-borger.e@wehi.edu.au.
*E-mail: sjwill@unimelb.edu.au.

ORCID
David B. Ascher: 0000-0003-2948-2413
Gideon J. Davies: 0000-0002-7343-776X
Spencer J. Williams: 0000-0001-6341-4364

Author Contributions

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS
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, are acknowledged. G.J.D. is supported by the Royal Society, and this work underpins recent Leverhulme Trust funding (RPG-2017-190). G.J.D. is supported by the Royal Society, and this work underpins recent Leverhulme Trust funding (RPG-2017-190). G.J.D. is supported by the Royal Society, and this work underpins recent Leverhulme Trust funding (RPG-2017-190).

■ REFERENCES
(16) Busilje, C. M.; Santos, J.; Dellino, 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.