The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids

Alexander J.D. Snow, Mahima Sharma, James P. Lingford, Yunyang Zhang, Janice W.-Y. Mui, Ruwan Epa, Ethan D. Goddard-Borger, Spencer J. Williams, Gideon J. Davies

Handling editor: Natalie Strynadka

Keywords: X-ray crystallography, Substrate-binding protein, Isothermal-titration calorimetry, Sulfofatty acid

ARTICLE INFO

1. Introduction

Sulfoquinovose (SQ) is the anionic headgroup of the ubiquitous plant sulfolipid, sulfoquinovosyl diacylglycerol (SQDG). SQDG can undergo delipidation to give sulfoquinovosyl glycerol (SQGro) and further glycoside cleavage to give SQ, which can be metabolized through microbial sulfofolic acid pathways. Exogenous SQDG metabolites are imported into bacteria through membrane spanning transporter proteins. The recently discovered sulfofolic acid sulfoquinovosyl monoxygenase (sulfo-SMO) pathway in Agrobacterium tumefaciens features a periplasmic sulfoquinovosyl glycerol binding protein, SmoF, and an ATP-binding cassette (ABC) transporter. Here, we use X-ray crystallography, differential scanning fluorimetry and isothermal titration calorimetry to study SQ glycoside recognition by SmoF. This work reveals that in addition to SQGro, SmoF can also bind SQ, a simple methyl glycoside and even a short-chain SQDG analogue. Molecular recognition of these substrates is achieved through conserved interactions with the SQ-headgroup together with more plastic interactions with the aglycones. This suggests that the solute binding protein of A. tumefaciens, and related SQ-binding proteins from other sulfofolic acid pathways, can provide their host organisms direct access to most of the SQ metabolites known to be produced by phototrophs.
liberated from imported SQDG or its delipidated forms sulfoquinovosyl monacylglycerol (SQMG) and sulfoquinovosyl glycerol (SQGro). SQ is hydrolyzed from these molecules by ‘gateway’ sulfoquinovosidases, which belong to glycoside hydrolase family GH31 (www.cazy.org) (Abayakoon et al., 2018; Speciale et al., 2016), while import of SQ glycosides is mediated by specialized permeases or transport systems.

The sulfo-SMO pathway of Agrobacterium tumefaciens utilizes a two-component system, comprised of an FMNH$_2$-dependent SQ monooxygenase and a flavin reductase, to cleave the carbon-sulfur bond of SQ to form 6-oxo-glucose (6-OG) and sulfite (Sharma et al., 2022), (Fig. 1). Reduction of 6-OG to glucose is catalyzed by an NADPH-dependent 6-OG reductase, enabling the product, glucose, to enter central metabolism. The smo gene cluster encodes an ATP-binding cassette (ABC) transport system consisting of a pair of identical ATPase domains (SmoE) and two distinct transmembrane domains (SmoG, SmoH). The ABC transporter engages with the periplasmic solute binding protein SmoF, which binds SQGro with sub-micromolar affinity and recruits it for import into the cell (Sharma et al., 2022). The sulfo-ED pathway gene cluster in Rhizobium leguminosarum also contains an ABC transporter and putative SQGro binding protein, suggesting that ABC transporters may be utilized in other sulfoglycolytic pathways in different organisms (Li et al., 2020).

Solute binding proteins, such as maltose binding protein (MalE) and SmoF, are associated with ABC transporters and are involved in the recruitment of the substrate ligand to the transmembrane domains to enable ATP-dependent transport across the membrane (Davidson et al., 2008) (Fig. 1). There are seven classes of ABC transporters (Thomas and Tampé, 2020), with the specificity and mechanism of the type 1 ABC transporter maltose transporter MalEFGK$_2$ perhaps the best characterized. Maltose transporter operates in conjunction with a periplasmic substrate binding protein MalE, with maltooligosaccharide loaded-MalE docking with the membrane components MalFGK$_2$ (Quiocho et al., 1997; Sparling et al., 1991). In free (apo) form MalE adopts an open conformation, and upon ligand binding MalE undergoes a hinge bending motion to a closed conformation. The adoption of the closed conformation is essential for productive interaction of MalE with the cytoplasmic-membrane components of the ABC transporter complex and importation of maltooligosaccharides across the membrane (Bermisson et al., 2010; Sharma et al., 2022). Studies of solute binding proteins show that the ligand-free form undergoes equilibration between open and semi-closed states (Tang et al., 2007). In the case of SmoF, once the open ligand-free form binds SQGro (Sharma et al., 2022), it undergoes a domain rotation to a closed conformation that encapsulates the ligand.

Here, we study the ligand specificity of SmoF, showing that in addition to SQGro, it can bind SQ, the simple glycoside SQMe and, unexpectedly, a short-chain derivative of SQDG. The thermodynamic and structural basis of binding for these ligands is explored. This work suggests that SmoF may facilitate the delivery of both SQ, SQGro and even plant sulfolipids to the ABC transporter and that this transporter system allows acquisition of a range of SQ glycosides by the host organism. We
demonstrate that the residues involved in sulfonate binding are conserved within several other organisms in the taxon *Rhizobiales* but not in putative solute-binding proteins associated with ABC transporters in other sulfoglycolytic bacteria, suggesting that their sulfonate-binding proteins may have evolved other modes of SQ recognition.

2. Materials and Methods

2.1. Gene expression and protein purification

Gene expression and purification of SmoF was performed as previously documented (Sharma et al., 2022). Briefly, expression of SmoF was achieved using pET29b(+) vector using BL21(DE3) competent E. coli. The native gene sequence for SmoF encodes a signal peptide, but this was deleted for recombinant expression. His6-tagged binding protein was purified by rebinding metal affinity chromatography (IMAC) on a nickel-nitrilotriacetic acid (NiNTA) column using 50 mM TRIS 300 mM NaCl pH 7.4 containing 30 mM imidazole and the bound protein was eluted using a linear gradient with buffer containing 300 mM imidazole. SmoF fractions were pooled, concentrated and loaded onto a HiLoad 16/600 Superdex 75 gel filtration column pre-equilibrated with 50 mM NaPi, 300 mM NaCl pH 7.4 buffer. The pooled fractions were concentrated to 40 mg/ml for crystallization trials.

2.2. Protein crystallization and optimization

SmoF-SQ crystals were grown in a sitting drop using 20 mg ml⁻¹, in 0.1 M NaBr, NaI, 0.1 M imidazole, MES pH 6.9, 13.5% MPD, PEG 3350. SmoF crystals were grown using 50 mg ml⁻¹ protein in 0.3 M sodium acetate, 0.1 M Bis-Tris (pH 5.5) and 35% PEG 2000 MME in a sitting drop, with a 5:6 mother liquor: protein ratio. SmoF/SQ crystals were grown using 50 mg ml⁻¹ protein in 0.3 M sodium acetate, 0.1 M Bis-Tris (pH 5.5) and 35% PEG 2000 MME in a sitting drop, with a 5:6 mother liquor: protein ratio. SmoF/SQ crystals were grown in a sitting drop at 6 °C with 25 mg ml⁻¹ protein and 25 mM SQDG-(C4:0/C15:0) dissolved in DMSO, in 25 mM NaPi, 150 mM NaCl pH 7.0, incubated at room temperature with 2.5 mM SQDG-(C4:0/C16:0) for 10 min prior to crystallization. Diffraction quality crystals were collected from a direct scale up of the Morpheus screen (Molecular Dimensions), condition H12. This contains 0.1 M amino acids (0.2M L-Na-glutamate, 0.2 M alanine, 0.2 M glycine, 0.2 M lysine hydrochloride, 0.2 M serine), 0.1 M buffer system 3 pH 8.5 (1 M TRIS, 1 M bicine) and 50% v/v precipitant mix (25% MPD, 25% PEG 1000, 25% PEG 3350). Crystals only formed in premade mother liquor. No cryoprotectant was used on the resultant crystals due to the presence of cryoprotecting PEG solutions in the mother liquor. Crystals were harvested then flash cooled in liquid nitrogen, using nylon Cryoloops (Hampton).

2.3. Data collection and structure determination

All crystals were tested using a Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV (Porebski et al., 2013). X-ray data collection occurred at the Diamond Light Source, using beamline I-03 during collection mx18598-51. Data collection statistics are available in Table 1. Data indexing and initial processing for SmoF-SQ and SmoF-SQMe was performed at Diamond, using either DIALS or 3dii automated pipelines from the Xia2 package (Beilsten-Edmands et al., 2020; Winter, 2010). For SmoF with SQDG-(C4:0/C16:0), indexing was performed manually using DUI (Beilsten-Edmands et al., 2020). In all cases AIMLESS was used for data reduction and quality assessment (Evans and Murshudov, 2013). Resolution was cut to CC1/2 = 0.5, or to the highest resolution possible while maintaining an outer shell completeness of 80% or higher. Molecular replacement used either MOLREP or PHASER (Vagin and Teplyakov, 2010; McCoy et al., 2007). The model used for the SmoF-SQMe structure was 70FY, and the SmoF-SQMe structure was then used for the other two datasets. Early model building was automated using BUCCANEER (Cowtan, 2006). Model refinement was performed using REFMACS employing local NCS restraints in the refinement cycles, and all interactive modelling and validation performed in COOT (Emsley & Cowtan, 2004; Murshudov et al., 2011). All steps excluding manual data integration were performed from within the CCP4i2 system (Pötterton et al., 2018). Diagram preparation for molecular models was performed using CCP4MG, Pymol or UCSF Chimera, depending on the desired outcome (Nicholas et al., 2011; Pettersen et al., 2004; Schrödinger, 2015). Analyses of conformational changes and internal cavities were performed using the DynDom web server and the CASTp V.3.0 Pymol plugin, respectively (Girdlestone and Hayward, 2016; Tian et al., 2018). We detect anisotropy in SmoF-SQ and SmoF-SQMe datasets as evident from a much higher anisotropic B value for data along the c* axis, affecting the respective data processing statistics. The resolution cut-off for these datasets was chosen based on higher quality of maps and better refinement statistics.

2.4. NanoDSF

NanoDSF was performed in 10 µl sample capillaries on a Prometheus NT.48 (NanoTemper). Excitation was 15% for ligand-free, SQ and SQMe samples, and 20% for SQDG-(C4:0/C16:0) and SQDG-(C18:1/C16:0). The 330/350 nm ratio of fluorescence was recorded between 15 °C and 95 °C, at 1 °C min⁻¹. Data collection and preliminary analysis performed using ThermalControl (NanoTemper). All SmoF samples were at 1 mg

| Table 1 | Data collection and refinement statistics for SmoF structures complexed with SQ, SQMe, SQDG-(C4:0/C16:0). |
| --- | --- |
| SmoF-SQ | SmoF-SQMe | SmoF-SQDG-(C4:0/C16:0) |
| Data collection | Space group P 3 1 2 1 | P 2 1; 2 1; 2 1 | P 2 1 |
| Cell dimensions a, b, c (Å) | 102.2, 102.2, 67.96 | 53.76, 66.27, 99.38 | 53.22, 69.59, 104.57 |
| Resolution (Å) | 88.5–1.80 | 94.6–1.59 | 69.6–2.14 |
| Rmerge | 0.179 (2.77) | 0.280 (1.38) | 0.003 (0.305) |
| Rfree | 0.06 (0.93) | 0.148 (0.909) | 0.081 (0.265) |
| I/σI | 101.3 (1.3) | 7.7 (1.3) | 8.9 (3.6) |
| Completeness (%) | 100 (100) | 99.4 (96.1) | 99.9 (100) |
| Redundancy | 19.2 (19.1) | 6.9 (6.6) | 4.1 (4.1) |
| Refinement | Resolution (Å) 1.8 | 1.59 | 2.14 |
| No. unique reflections | 38847 | 48233 | 24477 |
| No. atoms | 5740 | 5796 | 11468 |
| Protein | 5740 | 5796 | 11468 |
| Lipid/ion | 27 | 30 | 142 |
| Water | 138 | 299 | 239 |
| R factors (Å²) | 0.20 | 0.20 | 0.20 |
| Protein | 32 | 20 | 27 |
| Lipid/ion | 23 | 18 | 26 |
| Water | 34 | 27 | 29 |
| Ramachandran Plot Residues | 98.4 | 98.1 | 97.8 |
| In most favourable regions (%) | 1.6 | 1.6 | 2.2 |
| In allowed regions (%) | 0.0 | 0.3 | 0.0 |
| PDB code | 7Y9S | 7YZU | 7QHV |
3. Results and discussion

were submitted to Clustal 2.1 for multiple sequence alignment, and re-

Gene clusters possessing a putative SQase, putative

corresponding nucleotide accession numbers for each protein from all

maximum target sequences was set to 10,000. The results were

available on GitHub (https://github.com/jmui-unimelb/Gene-Clust

crystallization with SmoF and diffracted to

2.5. Isothermal titration calorimetry (ITC)

ITC experiments were performed using a MicroCal iTC200 (GE

To find sulfo-SMO and sulfo-ED clusters containing SmoF homo-

the protein sequence of A. tumefaciens C58 SmoF was submitted to

The corresponding nucleotide accession numbers for each protein from all

The C3 and C4 hydroxyl groups, and

The SmoF

2.6. Bioinformatics

To find sulfo-SMO and sulfo-ED clusters containing SmoF homo-

sulfo-EMP gene cluster as a query. Scripts for this pipeline are

the third sulfonate oxyanion forms a hydrogen bond with the backbone

The SQ moiety exists in nature as the free sugar SQ, and as glycosides

SQ and SQMe reside within an internal cavity that entirely encloses

The ligand-free SmoF had a Tm of 43.9 °C, which was raised to

5.41e-44 retained. The

effect the poor solubility of this glycolipid and the formation of

The SmoF sulfo-EMP gene cluster as a query. Scripts for this pipeline are available on GitHub (https://github.com/jmui-unimelb/GenecLust-er-Search-Pipeline). Gene clusters possessing a putative SQase, putative SQ isomerase, putative SF kinase and putative SFP aldolase were deemed putative sulfo-EMP operons. These putative sulfo-EMP operons were manually searched to identify their transporter types. Candidate SQBPs were submitted to Clustal 2.1 for multiple sequence alignment, and results were used to generate a cladogram.

3. Results and discussion

The SQ moiety exists in nature as the free sugar SQ, and as glycosides including SQDG and SQGro (Supplementary Fig. S1). To explore the ability of SmoF to bind to different glycosides, we synthesized methyl α-sulfooquinovoside (SQMe), and a naturally occurring SQDG, α-sulfoo-quinovosyl 1-oleoyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)) (Zhang et al., 2020). Because the full-length lipids endow this lipoform with poor aqueous solubility we also synthesized a more water-soluble analogue, α-sulfooquinovosyl 1-butanolyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)), which bears a shorter butanoloyl lipid.

Initially, we assessed binding of the analogues to SmoF using nano
differential scanning fluorimetry (nanoDSF). NanoDSF uses tryptophan or tyrosine fluorescence to monitor protein unfolding as a function of temperature and allows calculation of a melting temperature (Tm) that describes the thermodynamic stability of the protein or protein-ligand complex. Ligand-free SmoF gave a Tm of 58.5 °C (∆Tm = 14.6 °C), and in the presence of 2 mM SQDG-(C4:0/C16:0) the Tm of SmoF increased to 51.8 °C (∆Tm = 7.7 °C). In contrast, 2 mM SQMe-(C4:0/C16:0) did not result in a significant change in Tm (Fig. 2c, Supplementary Fig. S2). This may indicate this long-chain SQDG does not bind, that it binds with no change, or may simply reflect the poor solubility of this glycolipid and the formation of micelles unable to bind SQBP (Supplementary Fig. S1). We next studied the direct binding of these ligands to SmoF by isothermal titration calorimetry (ITC). SQ bound with a Kd value of 2.4 μM, and SQMe bound with a Kd value of 11.5 μM, which are 10-fold and 40-fold weaker affinity compared to SQGro, respectively (Fig. 2a and b, Supplementary Fig. S3). It was attempted with SQDG-(C4:0/C16:0) but was unsuccessful, possibly due to the formation of lipid micelles leading to phase separation (Fig. 2c, Supplementary Table S1).

Crystals of complexes of SmoF with SQ, SQMe and SQDG-(C4:0/

To test if sulfo-SMO and sulfo-ED clusters contain SmoF homologues in sulfo-SMO and sulfo-ED clusters. To test if sulfo-SMO and sulfo-ED clusters contain SmoF homologues, the E. coli sulfooquinovosidase (NP.418314.1, locus tag b3878), SQ mutarotase (NP.418315.3, locus tag b3879), SQ isomerase (NP.418316.4, locus tag b3880), SF kinase (NP.418319.2, locus tag b3883), SFP aldolase (NP.418317.1, locus tag b3881), SLA reductase (NP.418318.1, locus tag b3882) and sulfo-EMP regulator (NP.418320.2, locus tag b3884) were submitted separately as

redundant protein sequence (nr) database. Browsing the outputs allowed identi-

The C6 sulfonate and sugar C2, C3 and C4 hydroxyls (Fig. 3a). The

α-sulfoquinovoside in SQMe. All other interactions are identical to those observed

Asp67 (2.9, 2.7, 3.0, 2.5 Å) (Fig. 3a). The C3 and C4 hydroxyls each bind one nitrogen of Arg345 and the

The SmoF sulfo-EMP gene cluster contains small differences in binding recognition compared to SQ. The sulfoate pocket is identical with the exception of Gin12, which is too distant (3.6 Å) from the sulfoate oxygen to form a hydrogen bond. His13, which is on the same loop, is unable to form a hydrogen bond with the C1 oxygen as it is now present as a glycoside in SQMe. All other interactions are identical to those observed

and the carbonyl oxygen of Asp67 (2.9, 2.7, 3.0, 2.5 Å) (Fig. 3a).

The SmoF-SQMe complex contains small differences in binding recognition compared to SQ. The sulfoate pocket is identical with the exception of Gin12, which is too distant (3.6 Å) from the sulfoate oxygen to form a hydrogen bond. His13, which is on the same loop, is unable to form a hydrogen bond with the C1 oxygen as it is now present as a glycoside in SQMe. All other interactions are identical to those observed

with the two-domain fold observed previously (Sharma et al., 2022). Within the SmoF-SQ complex, SQ is present as the

the cavity features three openings. One of

2.6. Bioinformatics

To find sulfo-SMO and sulfo-ED clusters containing SmoF homologues, the protein sequence of A. tumefaciens C58 SmoF was submitted to the NCBI psiBLAST algorithm, searching a non-redundant protein sequence (nr) database, with

the cavity features three openings. One of

The SmoF

The C3 and C4 hydroxyls each bind one nitrogen of Arg345 and the

within the sulfonate binding pocket. The palmitic acid chain protrudes

identical interactions as for SQMe with the C2-4 hydroxyl groups, and

All other interactions are identical to those observed

sulfo-EMP regulator (NP.418320.2, locus tag b3884) were submitted separately as queries to the NCBI BLASTp tool. The database searched was the non-

protein sequences with E-value ≤ 5.41e-44 retained. The corresponding nucleotide accession numbers for each protein from all seven searches were extracted, and the seven lists combined and duplicates removed to give a list of candidate genome sequences. This list was converted into a MultiGeneBLAST reference library and searched using the E. coli sulfo-EMP gene cluster as a query. Scripts for this pipeline are available on Github (https://github.com/jmui-unimelb/GenecLust-er-Search-Pipeline). Gene clusters possessing a putative SQase, putative SQ isomerase, putative SF kinase and putative SFP aldolase were deemed putative sulfo-EMP operons. These putative sulfo-EMP operons were manually searched to identify their transporter types. Candidate SQBPs were submitted to Clustal 2.1 for multiple sequence alignment, and results were used to generate a cladogram.

3. Results and discussion

The SQ moiety exists in nature as the free sugar SQ, and as glycosides including SQDG and SQGro (Supplementary Fig. S1). To explore the ability of SmoF to bind to different glycosides, we synthesized methyl α-sulfooquinovoside (SQMe), and a naturally occurring SQDG, α-sulfoo-quinovosyl 1-oleoyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)) (Zhang et al., 2020). Because the full-length lipids endow this lipoform with poor aqueous solubility we also synthesized a more water-soluble analogue, α-sulfooquinovosyl 1-butanolyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)), which bears a shorter butanoloyl lipid.

Initially, we assessed binding of the analogues to SmoF using nano
C16:0). Dissociation constants for SQDG-(C4:0/C16:0) and SQDG-(C18:1/C15) could not be measured (noted by a dash). Data for SQGro (in blue) was determined by ITC for SQ, SQMe, SQDG-(C4:0/C16:0) and SQDG-(C18:1/C15).

The protein-ligand interactions are almost identical in all cases, and these ligands result in a large conformational change in the protein versus the apo form, which are found in other SBPs. MalE features a comparable closure angle upon ligand binding (37°) (Tang et al., 2007) (Supplementary Fig. S6). In the SmoF•SQDG-(C4:0/C16:0) complex, ligand binding is accompanied by an upwards deflection in α-helix 1 by 9.8 Å. This moves Gin12 and His13 away from SQDG-(C4:0/C16:0) allowing its large lipid groups to bind (Supplementary Figs. S7 and S8). Collectively, this data shows that SmoF retains its interactions around the sulfosugar yet has sufficient conformational plasticity to accommodate larger aglycones. For the diacyl glycerol substituent this enables binding even though the entire lipid chain cannot be contained within the binding pocket.

The ligand complexes described above identify a set of residues that are involved in binding a range of SQ analogues, and thus could potentially serve to identify SQ-binding proteins. Previous work has identified several other sulfoglycolytic clusters containing likely SQ-binding proteins in association with ABC transporters: the sulfo-ED gene cluster of *R. leguminosarum* SRDI565 contains a SmoF homologue with 80% sequence identity (Li et al., 2020), and a SmoF homologue was identified in the SMO gene cluster of *Rhizobium oryzae* with 78% identity (Sharma et al., 2022). To identify other candidate SQ-binding proteins, we performed a search for sulfoglycolytic operons that contained putative SQ binding proteins and ABC cassettes. We identified a candidate sulfo-SMO gene cluster in *Neorhizobium galache* str. DS1499; a candidate sulfo-ED gene cluster in *Microlunatus phosphovorus* NM-1; and candidate sulfo-EMP gene clusters in *Vibrio barjaei* str. 3062 and *Tetrasphaera* sp. Soil756, all of which contained genes encoding SmoF homologues and ABC transporters (Fig. 4a). Sulfo-TK clusters containing a putative SQ binding protein and ABC cassette were reported by Liu and co-workers (Liu et al., 2021). A cladogram of these putative SQ binding proteins shows close homology between the proteins in *Rhizobiales* but otherwise no relationship between sequence identity and the sulfoglycolysis pathway (Fig. 4b).

We next studied whether sulfonate binding pockets were conserved across SmoF homologues as well as other sulfonate-targeting solute binding proteins. Thus, we included SsuA from *E. coli* and *Xanthomonas citri* (Beale et al., 2010; Tófoli De Araújo et al., 2013) and the taurine-binding protein TauA from *E. coli* (Qu et al., 2019), which are solute-binding proteins associated with ABC transporters that bind assolated sulfonates. We also included MalE as a well-characterized SBP that binds a non-sulfonated ligand. Multiple sequence alignment of the SmoF homologues, SsuA, TauA and MalE revealed conservation of the *A. tumefaciens* SmoF sulfonate binding pocket with only *R. leguminosarum* and *R. oryzae* putative SQ-binding proteins (Supplementary Fig. S9). The SQ hydroxyl-binding arginine and aspartic acid residues are conserved in *Neorhizobium* but not among other putative SQ-binding proteins. The poor conservation of binding residues across putative SQ-binding proteins stands in contrast to the strongly conserved sulfonate binding residues present in SQases (Abayakoon et al., 2018; Speciale et al., 2016), which have been used to identify new sulfoglycolysis gene clusters (Liu et al., 2021). There was no conservation of sulfonate binding residues in SsuA or TauA, or in MalE.

**4. Conclusions and future work**

We show that the solute binding protein SmoF can bind SQ and SQMe, in addition to SQGro as previously reported (Sharma et al., 2022). The protein-ligand interactions are almost identical in all cases, and these ligands result in a large conformational change in the protein versus the apo form, and complete enclosure of the ligand. We also show that SmoF can bind a simplified SQDG. Despite the large lipid groups, binding occurs through largely conserved interactions with the SQ headgroup but involves plasticity in its binding site to partially accommodate the lipid groups. Minor conformational changes in the protein result in an opening from which the lipids protrude. These results suggest that SmoF may

---

**Fig. 2. Binding affinity of SQ and its glycosides for SmoF.** a) Isothermal titration calorimogram showing titration of SQ into SmoF. b) Calorimogram of SQMe into SmoF. c) Melting temperature (Tm) of SmoF, as determined by differential scanning fluorimetry, the Tm shift relative to apo-SmoF, and **K**ₙ values determined by ITC for SQ, SQMe, SQDG-(C4:0/C16:0) and SQDG-(C18:1/C16:0). Dissociation constants for SQDG-(C4:0/C16:0) and SQDG-(C18:1/C16:0) could not be measured (noted by a dash). Data for SQGro (in blue) was reported in (Sharma et al., 2022) and has been included for comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are near the sulfonate. The internal volume of the cavity at 1283 Å³ is > 4 times larger than that of the SmoF•SQ complex, with a large non-polar region occupied by the butanoyl chain of SQDG-(C4:0/C16:0) (Fig. 3f and g).

The structures of the SmoF complexes with SQ, SQMe and SQDG-(C4:0/C16:0) show large conformational changes relative to the ligand-free form, as observed previously with SQGro (Sharma et al., 2022), and undergo interdomain rotations of up to 33° compared to the unliganded state (Fig. 3g). This movement centers around a pair of hinges, which are found in other SBPs. MalE features a comparable closure angle upon ligand binding (37°) (Tang et al., 2007) (Supplementary Fig. S6).
allow capture of free SQ, SQGro and even lipidated SQ glycosides such as SQMG and SQDG, allowing metabolism of the lipidic part in addition to the SQ and the glycerol. Previously, there has been no evidence that SQDG can be completely metabolized by sulfoglycolytic organisms. Instead, various non-specific lipases have been reported that can cleave the lipid chains (Snow et al., 2021), suggesting that the sulfo-SMO pathway is used in partnership with non-sulfoglycolytic organisms (possibly including plants) that excise and metabolize the energy-rich lipid chains of SQDG, releasing the sulfosugar SQGro. However, the ability of SmoF to bind SQDG suggests that A. tumefaciens can on its own achieve the import of SQDG and SQMG. Within this scenario, SmoF, working in concert with A. tumefaciens sulfoquinovosidase SmoI, which is expressed with a signal peptide that will direct expression to the periplasm, enables capture of the full carbon-content of SQDG/SQMG. Possibly, this could allow A. tumefaciens to utilize intact sulfolipids as a nutrient upon infecting a plant host.

A search for other ABC transporters and associated solute binding proteins in sulfoglycolytic gene clusters led to identification of ABC transporter systems similar to that of A. tumefaciens in organisms with gene clusters encoding sulfo-SMO, sulfo-ED, sulfo-EMP, sulfo-SFT and sulfo-TK pathways. This complements earlier reports showing that sulfo-EMP and sulfo-ED (Denger et al., 2014; Felux et al. 2015a,b) gene clusters also contained TauF transporters of the 4-toluene sulfonate uptake permease (TSUP) family (Shlykov et al., 2012) to import SQ and its glycosides. The occurrence of TSUP family or ABC transporter systems in various sulfoglycolysis gene clusters suggests that the specific transporter used to import the sulfosugar substrate is not restricted to a particular pathway. Sequence alignment of putative SQ-binding proteins from this range of organisms revealed that SQ binding residues identified in A. tumefaciens SmoF are not well conserved, and thus that acquisition of SQ-binding function may have arisen through independent evolutionary events. Thus, sequence-based searches for new SQ-binding proteins may have poor predictive power, and will require consideration of genetic context and whether the solute binding protein and ABC transporter are associated with a sulfoglycolytic gene cluster. Finally, the ability of SmoF to bind SQ glycosides bearing extended lipid chains means it may be possible to exploit this SQ-binding protein to bind to SQ-linked structures for affinity-based protein capture and purification applications, in a way analogous to the use of maltose-binding protein that binds its cognate ligands ($K_d = 0.5–2 \mu \text{m}$) with similar affinities.

**Funding**

GJD is supported by the Royal Society Ken Murray Research Professorship. MS and AS were funded by the Leverhulme Trust (grant RPG-2017-190) and MS subsequently by the Biotechnology and Biological Sciences Research Council (BB/W003805/1). SJW is supported by the Australian Research Council (DP180101957 and DP210100233). EDGB is supported by the Brian M. Davis Charitable Foundation Centenary Fellowship, National Health and Medical Research Council of Australia.
Fig. 4. Comparison of *A. tumefaciens* sulfo-SMO gene cluster with other proposed sulfo-glycolytic gene clusters containing SmoF homologues. a) Gene clusters for sulfo-glycolytic sulfo-SMO, sulfo-EMP, sulfo-TAL and sulfo-TK pathways containing ABC transporters featuring a sulfoquinovose binding protein. b) Cladogram of SmoF and homologues found in different organisms featuring sulfo-glycolysis pathways in their core genome.

(NHMB) project grants GNT1139546, GNT1139549 and GNT2000517, and acknowledges support from The Walter and Eliza Hall Institute of Medical Research, the Australian Cancer Research Fund and a Victorian State Government Operational Infrastructure support grant.

CRediT authorship contribution statement

Alexander J.D. Snow: Methodology, Investigation. Mahima Sharma: Methodology, Investigation. Yunyang Zhang: Methodology, Investigation. Janice W.-Y. Mui: Methodology, Investigation. Ruwan Epa: Methodology, Investigation. Ethan D. Goddard-Borger: Conceptualization, Project administration, Supervision, Funding acquisition. Spencer J. Williams: Conceptualization, Project administration, Supervision, Funding acquisition. Gideon J. Davies: Conceptualization, Project administration, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Johan Turkenburg and Sam Hart from maintaining X-ray data collection facilities in York and coordinating Diamond data collection, whose staff are also thanked for provision of beamline facilities (project mx18598). We also thank Prof. Eleanor Dodson FRS for helpful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crsb.2022.03.001.

References

Abayakoon, P., Jin, Y., Lingford, J.P., Petricervic, M., John, A., Ryan, E., Wai-Ying Mui, J., Pires, D.E.V., Ascher, D.B., Davies, G.J., Goddard-Borger, E.D., Williams, S.J., 2018. Structural and biochemical insights into the function and evolution of sulfoquinovosidases. ACS Cent. Sci. 4, 1266–1273. https://doi.org/10.1021/acscentsi.8b00453.

Beale, J., Lee, S.Y., Iwata, S., Beis, K., 2010. Structure of the aliphatic sulfolipid-binding protein SmuA from Escherichia coli. Acta Crystallogr. Sect. F Struct. Biol. Crystallogr. Commun. 66, 391–396. https://doi.org/10.1107/S1744309110006226.

Bilstein-Edmonds, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D., Evans, G., 2020. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. Acta Crystallogr. Sect. D Struct. Biol. 76, 385–399. https://doi.org/10.1107/S1600536820003198.

Benson, A.A., Daniel, H., Wiser, R., 1959. A sulfolipid IN plants. Proc. Natl. Acad. Sci. Unit. States Am. 45, 1582–1587. https://doi.org/10.1073/pnas.45.11.1582.

Bermisson, R.P.A., Smits, S.H.J., Schmitt, L., Slobooboom, D.J., Poolman, B., 2010. A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–2617. https://doi.org/10.1016/j.febslet.2010.04.043.

Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. Sect. D Biol. Crystallogr. 62, 1002–1011. https://doi.org/10.1107/S0907444906022116.

Davidson, A.L., Dassa, E., Oettle, C., Chen, J., 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol. Mol. Biol. Rev. 72, 317–364. https://doi.org/10.1128/MMBR.00031-07.

Denger, K., Huhn, T., Hollemeyer, K., Schlebeck, D., Cook, A.M., 2012. Sulfoquinovose degraded by pure cultures of bacteria with release of C3-organosulfonates: complete degradation in two-member communities. FEMS Microbiol. Lett. 328, 39–45. https://doi.org/10.1016/j.femsle.2012.03.011.

Denger, K., Weiss, M., Felux, A.K., Schneider, A., Mayer, C., Spiteller, D., Huhn, T., Cook, A.M., Schlebeck, D., 2014. Sulphoglycolysis in Escherichia coli K-12 closes a gap in the biogeochemical sulphur cycle. Nature 507, 114–124. https://doi.org/10.1038/nature13097.

Duran, X., Hall, J.A., Nikaido, H., Quichoio, P.A., 2001. Crystal structures of the maltohexidin/maltose-binding protein complexed with reduced oligosaccharides: flexibility of tertiary structure and ligand binding. J. Mol. Biol. 306, 1115–1126. https://doi.org/10.1006/jmbi.2001.4556.

Evans, P.R., Marshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 1204–1214. https://doi.org/10.1107/S0907444913000601.

Felux, A.K., Chiesa, M., Avis, N., Schlebeck, D., 2009b. Proposal for a revised nomenclature system for sulfoquinovosidases. Biosci. Rep. 29, 427–430. https://doi.org/10.1042/BSR20090004.

Felux, A.K., Chiesa, M., Avis, N., Schlebeck, D., 2009a. Proposal for a revised nomenclature system for sulfoquinovosidases. Biosci. Rep. 29, 427–430. https://doi.org/10.1042/BSR20090004.

Harwood, J.L., Nicholls, R.G., 1979. The plant sulpholipid. Nature 282, 584–586. https://doi.org/10.1038/282584a0.
