
Article:
Snow, Alexander J. D., Sharma, Mahima orcid.org/0000-0003-3960-2212, Lingford, James P. et al. (6 more authors) (2022) The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids. Current Research in Structural Biology. pp. 51-58. ISSN 2665-928X

https://doi.org/10.1016/j.crstbi.2022.03.001

The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids

Alexander J. D. Snow,[a] Mahima Sharma,[a] James P. Lingford,[b,c] Yunyang Zhang,[d] Janice W.-Y. Mui,[d] Ruwan Epa,[d] Ethan D. Goddard-Borger,[b,c] Spencer J. Williams,[d] and Gideon J. Davies,*[a]

[a] York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD

[b] The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia.

[c] Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia.

[d] School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.

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 sulfoglycolytic pathways. Exogenous SQDG metabolites are imported into bacteria through membrane spanning transporter proteins. The recently discovered sulfoglycolytic sulfoquinovose monooxygenase (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 sulfoglycolytic pathways, can provide their host organisms direct access to most of the SQ metabolites known to be produced by phototrophs.

Keywords: X-ray crystallography, substrate-binding protein, isothermal-titration calorimetry, sulfoglycolysis

1. Introduction

Sulfoquinovose (6-deoxy-6-sulfoglucose, SQ) is a sulfosugar that occurs primarily as the anionic headgroup of the plant sulfolipids collectively termed sulfoquinovosyl diacylglycerol (SQDG) (Benson, et al. 1959). SQDG is a constituent of the thylakoid membranes of photosynthetic organisms (Mizusawa and Wada 2012; Sato et al. 2016) and associates with membrane proteins such as photosystem II (Loll et al. 2005). SQDG is a major global reservoir of organosulfur with an estimated 10 petagrams produced annually (Goddard-Borger et al. 2017; Harwood and Nicholls 1979). The catabolism of SQDG occurs in a wide range of gram-positive and gram-negative bacteria through one of five sulfoglycolytic pathways (Snow et al. 2021, J. Liu et al. 2021).

The sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP) (Denger et al. 2012, 2014; Sharma et al. 2021), sulfoglycolytic Entner-Doudoroff (sulfo-ED) (Felux et al. 2015; Li et al. 2020) and sulfoglycolytic sulfofructose transaldolase (sulfo-SFT) (Frommeyer et al. 2020; Y. Liu et al. 2020) pathways involve scission of the six-carbon SQ backbone into two three-carbon fragments: carbons 1-3 are metabolized, while carbons 4-6 are excreted as a three-carbon sulfonate after subsequent reduction to 2,3-dihydroxypropanesulfonate DHPS by NADH-dependent SLA reductase (YihU) (Sharma et al. 2020) or oxidation to sulfolactate SL by the NAD^+-dependent SLA dehydrogenase). The sulfoglycolytic transketolase (sulfo-TK pathway) involves stepwise scission of two car-bons (carbons 1 and 2, and carbons 3 and 4) from sulfofructose and transfer to G3P, while carbons 5 and 6 form sulfoacetaldehyde, which is reduced and excreted as isethionate (J. Liu et al. 2021). In contrast to all other known pathways, the sulfoglycolytic SQ monooxygenase (sulfo-SMO) pathway involves the cleavage of the sulfur-carbon bond of SQ with excretion of inorganic sulfur (predominantly as sulfite) and enables the utilization of all six carbons of SQ (Sharma et al. 2022, J. Liu et al. 2021). For sulfoglycolytic pathways to catabolise SQ, it must first be 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 sulfoquinovose 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), (Figure 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) (Figure 1). There are seven classes of ABC transporters (Thomas et al. 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; Spurlino, 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 (Duan et al. 2001). The SQ binding protein SmoF, like other solute binding proteins, has a structural fold comprised of two globular lobes (interconnected by polypeptide chains), which undergo conformational changes upon ligand binding (Berntsson 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 SQGro 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 immobilised-metal affinity chromatography (IMAC) on a nickel-nitrilotriacetic acid (NiNTA) column using 50 mM TRIS 300 mM NaCl pH 7.4 containing 30m M 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 1000, PEG 3350 at 10 °C, with a 5:6 mother liquor: protein ratio. SmoF•SQMe 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 1:1 protein:mother liquor ratio in-drop. In both cases 2 mM ligand was added to protein 10 min prior to drop formation. SmoF•SQDG-(C4:0/C16:0) crystals were grown in sitting drop at 6 °C, with 25 $mg.ml^{-1}$ protein and 2.5 mM SQDG-(C4:0/C16:0) 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 4 (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 et al., 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 et al. 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 REFMAC5 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 (Potterton et al. 2018). Diagram preparation for molecular models was performed using CCP4MG, Pymol or UCSF Chimera, depending on the desired outcome (McNicholas 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.ml^{-1} in 50 mM Tris, 300 mM NaCl pH 7.5. SQ, SQMe were dissolved in and diluted with ultrapure

water with the exception of SQDG analogs (SQDG-(C4:0/C16:0) and SQDG-(C18:1/C16:0) which were dissolved in DMSO to give a 250 mM stock which was further diluted to final concentration of 2 mM with 50 mM Tris, 300 mM NaCl pH 7.5 for binding studies (with final amount of DMSO co-solvent ranging between 0.4-1%). All samples were centrifuged at 13,000 rpm for 5 min prior to loading.

2.5. Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using a MicroCal iTC200 (GE Healthcare) at 25 °C, with a 750 r.p.m. stirring speed and a reference power of 10 $\mu Cal.s^{-1}$. SmoF was equilibrated into degassed and filter sterilised ITC buffer (50 mM NaPi, 200 mM NaCl, pH 7.4) by dialysis using 3 kDa MWCO Visking tubing. All ligands were dissolved directly into the same buffer. For SmoF/SQ, 2000 µM SQ was titrated into a cell containing 160 µM SmoF, and for SmoF/SQMe 2000 µM SQMe was titrated into 160 µM SmoF. Both were injected as a series of 15 × 2.94 µl injections with a pre-injection of 1 × 0.4 µl. Delays between injection were set at 120 s, with an initial injection delay of 60 s. All data was processed using PEAQ-ITC (MicroCal).

2.6. Bioinformatics

To find sulfo-SMO and sulfo-ED clusters containing SmoF homologs, the protein sequence of *A. tumefaciens* C58 SmoF was submitted to the NCBI psiBLAST algorithm, searching a non-redundant protein sequence (nr) database. Browsing the outputs allowed identification of homologues in sulfo-SMO and sulfo-ED clusters. To find sulfo-EMP clusters containing SmoF homologs, the *E. coli* sulfoquinovosidase (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 queries to the NCBI BLASTp tool. The database searched was the non-redundant protein sequence (nr) database, with *E. coli* (taxid: 562) sequences excluded. Standard algorithm parameters were used, except the maximum target sequences was set to 10,000. The results were filtered, with only 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/Gene-Cluster-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 Figure S1). To explore the ability of SmoF to bind to different glycosides, we synthesized methyl α-sulfoquinovoside (SQMe), and a naturally occurring SQDG, α-sulfoquinovosyl 1-oleoyl-2-palmitoylglycerol (SQDG-(C18:1/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-sulfoquinovosyl 1-butanoyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)), which bears a shorter butanoyl 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 had Tm of 43.9 °C, which was raised to 54.2 °C in the presence of 2 mM SQ (ΔTm = 10.3 °C). Using 2 mM SQMe gave a Tm of 58.5 °C (ΔTm = 14.6 °C), and in the presence of 2 mM SQDG-(C4:0/C16:0) the Tm of SmoF increased to 51.8 °C (ΔTm = 7.7 °C). In contrast, 2 mM SQDG-(C18:1/C16:0) did not result in a significant change in Tm (Figure 2c, Supplementary Figure 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 Figure S1). We next studied the direct binding of these ligands to SmoF by isothermal titration calorimetry (ITC). SQ bound with a K_d value of 2.4 μM, and SQMe bound with a K_d value of 11.5 μM, which is 10-fold and 40-fold weaker affinity compared to SQGro, respectively (Figure 2a, b, Supplementary Figure S3). ITC was attempted with SQDG-(C4:0/C16:0) but was unsuccessful, possibly due to the formation of lipid micelles leading to phase separation (Figure 2c; Supplementary Table S1).

Crystals of complexes of SmoF with SQ, SQMe and SQDG-(C4:0/C16:0) were obtained by co-crystallization with SmoF and diffracted to 1.8, 1.59 and 2.14 Å resolution, respectively (Table 1). Crystal structure of SmoF with SQDG-(C4:0/C16:0) was obtained in

$P2_1$ space group and contains two copies in the asymmetric unit with no significant differences between the copies. Crystals of SmoF with SQ are in space-group, $P3_121$, and SmoF•SQMe was obtained in $P2_12_12_1$, each with one copy in the asymmetric unit. Of the 394 residues present in the protein, 386 were pre-sent in SmoF•SQ structure, 387 in the SmoF•SQMe structure, and 392 in the SmoF•SQDG-(C4:0/C16:0) structure. In all three cases, 95% or more of the amino acids in the protein were observed in the electron density, with exceptions occurring primarily in flexible loops and the affinity tag. In all cases SmoF maintains the globular, primarily α-helical structure with the two-domain fold observed previously (Sharma et al. 2022). Within the SmoF•SQ complex, SQ is present as the α-anomer and adopts a $^4C_1$ (chair) conformation. Its binding interactions involve the C6 sulfonate and sugar C2, C3 and C4 hydroxyls (Figure 3a). The sulfonate occupies a binding pocket comprised of Gln_{12}-Ser_{43}-Gly_{166}-Thr_{220}. Within this pocket, the side-chain nitrogen of Gln_{12} and side-chain hydroxyl of Thr_{220} engage in a hydrogen bond with one sulfonate oxygen (2.6 Å, 2.9 Å), the second sulfonate oxygen forms a hydrogen bond to the backbone secondary amine of Gly_{166} (2.8 Å), and the third sulfonate oxyanion forms a hydrogen bond to the backbone secondary amide of Ser_{43} (2.9 Å) and an ordered water (3.0 Å). The C1 hydroxyl forms a hydrogen bond with a side-chain secondary amine on His_{13} (2.7 Å). The C2 hydroxyl group forms hydrogen bonds to the side chain carboxyl of Asp_{113} (2.7 Å) and the indole nitrogen of Trp_{276} (2.8 Å). The C3 and C4 hydroxyls each bind one nitrogen of Arg_{345} and the carboxyl oxygen of Asp_{67} (2.9, 2.7, 3.0, 2.5 Å) (Figure 3a).

The SmoF•SQMe complex contains small differences in binding recognition compared to SQ. The sulfonate pocket is identical with the exception of Gln_{12}, which is too distant (3.6 Å) from the sulfonate oxygen to form a hydrogen bond. His_{13}, 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 SQ (Figure 3b). The SmoF•SQDG-(C4:0/C16:0) complex involves identical interactions as for SQMe with the C2-4 hydroxyl groups, and within the sulfonate binding pocket. The palmitic acid chain protrudes through the top of the binding pocket and forms a crystal contact with another SmoF molecule in the asymmetric unit (Supplementary Figure S4).

SQ and SQMe reside within an internal cavity that entirely encloses the ligand (Figures 3d,e). For SQ the volume is 297 $Å^3$ and for SQMe is 447 $Å^3$, 66% larger. The cavity for SQGro is 476 $Å^3$, 6% larger than SQMe (Supplementary Figure S5). As noted above, SQDG-(C4:0/C16:0) is not fully enclosed by the protein and the cavity features three openings. One of these is occupied by the protruding palmitoyl chain, while the other two are near the sulfonate. The internal volume of the cavity at 1283 $Å^3$ 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) (Figure 3f,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 (Figure 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 Figure 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 Gln_{12} and His_{13} away from SQDG-(C4:0/C16:0) allowing its large lipid groups to bind (Supplementary Figure S7, 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 galagae* 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 (Figure 4a). 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 (Figure 4b).

We next studied whether sulfonate binding pockets were conserved across SmoF homologs 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 assorted alkanesulfonates. 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 Figure 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), which have been used to identify new sulfoglycolysis gene clusters (Speciale et al. 2016). 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 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. 2015) gene clusters also contained TauE 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 μm) with similar affinities.

5. 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 for helpful discussions.

6. Author Contributions

Conceptualization, Administration, Supervision and Funding: GJD, SJW and EDG-B conceived the study, acquired funding and designed the supervised experiments. AJDS, MS, YZ, JM and RE developed Methodology, performed Investigations, analyzed, curated and interpreted the data. All authors contributed to manuscript production at both original draft and revision stages.

7. 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). 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 (NHMRC) 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.

8. Conflict of Interest Statement

The authors declare no conflict of interest.

9. References

Abayakoon, P., Jin, Y., Lingford, J.P., Petricevic, 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, 1266–73. https://doi.org/10.1021/acscentsci.8b00453

Beale, J., Lee, S.Y., Iwata, S., Beis, K., 2010. Structure of the aliphatic sulfonate-binding protein SsuA from Escherichia coli. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66, 391–96. https://doi.org/10.1107/S1744309110006226

Beilsten-Edmands, 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, 385–99. https://doi.org/10.1107/S2059798320003198

Benson, A.A., Daniel, H., Wiser, R., 1959. A SULFOLIPID IN PLANTS. Proc. Natl. Acad. Sci. 45, 1582–87. https://doi.org/10.1073/pnas.45.11.1582

Berntsson, R.P.A., Smits, S.H.J., Schmitt, L., Slotboom, D.J., Poolman, B., 2010. A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–17. 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, 1002–11. https://doi.org/10.1107/S0907444906022116

Davidson, A.L., Dassa, E., Orelle, C., Chen, J., 2008. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. Microbiol. Mol. Biol. Rev. 72, 317–64. https://doi.org/10.1128/mmbr.00031-07

Denger, K., Huhn, T., Hollemeyer, K., Schleheck, 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/10.1111/j.1574-6968.2011.02477.x

Denger, K., Weiss, M., Felux, A.K., Schneider, A., Mayer, C., Spiteller, D., Huhn, T., Cook, A.M., Schleheck, D., 2014. Sulphoglycolysis in Escherichia coli K-12 closes a gap in the

biogeochemical sulphur cycle. Nature 507, 114–17. https://doi.org/10.1038/nature12947

Duan, X., Hall, J.A., Nikaido, H., Quiocho, F.A., 2001. Crystal structures of the maltodextrin/maltose-binding protein complexed with reduced oligosaccharides: Flexibility of tertiary structure and ligand binding. J. Mol. Biol. 306, 1115–26. https://doi.org/10.1006/jmbi.2001.4456

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 1204–14. https://doi.org/10.1107/S0907444913000061

Felux, A.K., Franchini, P., Schleheck, D., 2015a. Permanent draft genome sequence of sulfoquinovose-degrading Pseudomonas putida strain SQ1. Stand. Genomic Sci. 10, 1–6. https://doi.org/10.1186/s40793-015-0033-x

Felux, A.K., Spiteller, D., Klebensberger, J., Schleheck, D., 2015b. Entner-Doudoroff pathway for sulfoquinovose degradation in Pseudomonas putida SQ1. Proc. Natl. Acad. Sci. U. S. A. 112, E4298–305. https://doi.org/10.1073/pnas.1507049112

Frommeyer, B., Fiedler, A.W., Oehler, S.R., Hanson, B.T., Loy, A., Franchini, P., Spiteller, D., Schleheck, D., Hanson, S.R., Loy, B.T., 2020. Environmental and intestinal phylum Firmicutes bacteria metabolize the plant sugar sulfoquinovose via a 6-deoxy-6-sulfofructose transaldolase pathway. ISCIENCE 23, 101510. https://doi.org/10.1016/j.isci.2020.101510

Girdlestone, C., Hayward, S., 2016. The DynDom3D webserver for the analysis of domain movements in multimeric proteins. J. Comput. Biol. 23, 21–26. https://doi.org/10.1089/cmb.2015.0143

Goddard-Borger, E.D., Williams, S.J., 2017. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. Biochem. J. 474, 827–49. https://doi.org/10.1042/BCJ20160508

Harwood, J.L., Nicholls, R.G., 1979. The plant sulpholipid-- a major component of the sulphur cycle. Biochem. Soc. Trans. 7, 440–47. https://doi.org/10.1042/bst0070440

Li, J., Epa, R., Scott, N.E., Skoneczny, D., Sharma, M., Snow, A.J.D., Lingford, J.P., Goddard-Borger, E.D., Davies, G.J., Mcconville, M.J., Williams, S.J., 2020. A sulfoglycolytic entner-doudoroff pathway in rhizobium leguminosarum bv. trifolii SRDI565. Appl. Environ. Microbiol. 86. https://doi.org/10.1128/AEM.00750-20

Liu, J., Wei, Y., Ma, K., An, J., Liu, X., Liu, Y., Ang, E.L., Zhao, H., Zhang, Y., 2021. Mechanistically Diverse Pathways for Sulfoquinovose Degradation in Bacteria. ACS Catal. 11, 14740–50. https://doi.org/10.1021/acscatal.1c04321

Liu, Y., Wei, Y., Zhou, Y., Ang, E.L., Zhao, H., Zhang, Y., 2020. A transaldolase-dependent sulfoglycolysis pathway in Bacillus megaterium DSM 1804. Biochem. Biophys. Res. Commun. 533, 1109–14. https://doi.org/10.1016/j.bbrc.2020.09.124

Loll, B., Kern, J., Saenger, W., Zouni, A., Biesiadka, J., 2005. Towards complete cofactor

arrangement in the 3.0 Å resolution structure of photosystem II. Nature 438, 1040–44. https://doi.org/10.1038/nature04224

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. Journal of Applied Crystallography, 40(4), 658–674. https://doi.org/10.1107/S0021889807021206

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: The CCP4mg molecular-graphics software. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 386–94. https://doi.org/10.1107/S0907444911007281

Mizusawa, N., Wada, H., 2012. The role of lipids in photosystem II. Biochim. Biophys. Acta - Bioenerg. https://doi.org/10.1016/j.bbabio.2011.04.008

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 355–67. https://doi.org/10.1107/S0907444911001314

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–12. https://doi.org/10.1002/jcc.20084

Porebski, B.T., Ho, B.K., Buckle, A.M., 2013. Interactive visualization tools for the structural biologist. J. Appl. Crystallogr. 46, 1518–20. https://doi.org/10.1107/S0021889813017858

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S.J., Nicholls, R.A., Noble, M., Pannu, N.S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., Von Delft, F., Waterman, D., Wilson, K., Winn, M., Wojdyr, M., 2018. CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. Acta Crystallogr. Sect. D Struct. Biol. 74, 68–84. https://doi.org/10.1107/S2059798317016035

Qu, F., ElOmari, K., Wagner, A., De Simone, A., Beis, K., 2019. Desolvation of the substrate-binding protein TauA dictates ligand specificity for the alkanesulfonate ABC importer TauABC. Biochem. J. 476, 3649–60. https://doi.org/10.1042/BCJ20190779

Quiocho, F.A., Spurlino, J.C., Rodseth, L.E., 1997. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. Structure 5, 997–1015. https://doi.org/10.1016/S0969-2126(97)00253-0

Sato, N., Kobayashi, S., Aoki, M., Umemura, T., Kobayashi, I., Tsuzuki, M., 2016. Identification of genes for sulfolipid synthesis in primitive red alga Cyanidioschyzon merolae. Biochem. Biophys. Res. Commun. 470, 123–29. https://doi.org/10.1016/j.bbrc.2016.01.006

Schrödinger, L., 2015. The PyMol Molecular Graphics System.

Sharma, M., Abayakoon, P., Epa, R., Jin, Y., Lingford, J.P., Shimada, T., Nakano, M., Mui,

J.W.Y.W.-Y., Ishihama, A., Goddard-Borger, E.D., Davies, G.J., Williams, S.J., 2021. Molecular Basis of Sulfosugar Selectivity in Sulfoglycolysis. ACS Cent. Sci. 7, 476–87. https://doi.org/10.1021/acscentsci.0c01285

Sharma, M., Abayakoon, P., Lingford, J.P., Epa, R., John, A., Jin, Y., Goddard-Borger, E.D., Davies, G.J., Williams, S.J., Epa, R., Goddard-Borger, E.D., Davies, G.J., Williams, S.J., 2020. Dynamic structural changes accompany production of Dihydroxypropanesulfonate by Sulfolactaldehyde reductase. ACS Catal. 10, 2826–36. https://doi.org/10.26434/chemrxiv.9725120.v2

Sharma, M., Lingford, J.P., Petricevic, M., Snow, A.J.D., Zhang, Y., Järvå, M.A., Mui, J.W.-Y., Scott, N.E., Saunders, E.C., Mao, R., Epa, R., Silva, B.M. da, Pires, D.E.V., Ascher, D.B., McConville, M.J., Davies, G.J., Williams, S.J., Goddard-Borger, E.D., 2022. Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. Proc. Natl. Acad. Sci. 119, e2116022119. https://doi.org/10.2185/jjrm.23.202

Shlykov, M.A., Zheng, W.H., Chen, J.S., Saier, M.H., 2012. Bioinformatic characterization of the 4-Toluene Sulfonate Uptake Permease (TSUP) family of transmembrane proteins. Biochim. Biophys. Acta - Biomembr. 1818, 703–17. https://doi.org/10.1016/j.bbamem.2011.12.005

Snow, A.J.D., Burchill, L., Sharma, M., Davies, G.J., Williams, S.J., 2021. Sulfoglycolysis: catabolic pathways for metabolism of sulfoquinovose. Chem. Soc. Rev. 50, 13628–45.

Speciale, G., Jin, Y., Davies, G.J., Williams, S.J., Goddard-Borger, E.D., 2016. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. Nat. Chem. Biol. 12, 215–17. https://doi.org/10.1038/nchembio.2023

Spurlino, J.C., Lu, G.Y., Quiocho, F.A., 1991. The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J. Biol. Chem. 266, 5202–19. https://doi.org/10.1016/s0021-9258(19)67774-4

Tang, C., Schwieters, C.D., Clore, G.M., 2007. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 449, 1078–82. https://doi.org/10.1038/nature06232

Thomas, C., Tampé, R., 2020. Structural and Mechanistic Principles of ABC Transporters. Annu. Rev. Biochem. 89, 605–36. https://doi.org/10.1146/annurev-biochem-011520-105201

Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., 2018. CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res. 46, W363–67. https://doi.org/10.1093/nar/gky473

Tófoli De Araújo, F., Bolanos-Garcia, V.M., Pereira, C.T., Sanches, M., Oshiro, E.E., Ferreira, R.C.C., Chigardze, D.Y., Gonçalves Barbosa, J.A., Ferreira, L.C.D.S., Benedetti, C.E.,

Blundell, T.L., Balan, A., 2013. Structural and physiological analyses of the alkanesulphonate-binding protein (SsuA) of the citrus pathogen Xanthomonas citri. PLoS One 8, e80083. https://doi.org/10.1371/journal.pone.0080083

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallographica Section D: Biological Crystallography, 66(1), 22–25. https://doi.org/10.1107/S0907444909042589

Winter, G. (2010). Xia2: An expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography, 43(1), 186–190. https://doi.org/10.1107/S0021889809045701

Zhang, Y., Mui, J.W.Y., Arumaperuma, T., Lingford, J.P., Goddard-Borger, E.D., White, J.M., Williams, S.J., 2020. Concise synthesis of sulfoquinovose and sulfoquinovosyl diacylglycerides, and development of a fluorogenic substrate for sulfoquinovosidases. Org. Biomol. Chem. 18, 675–86. https://doi.org/10.1039/c9ob02540e

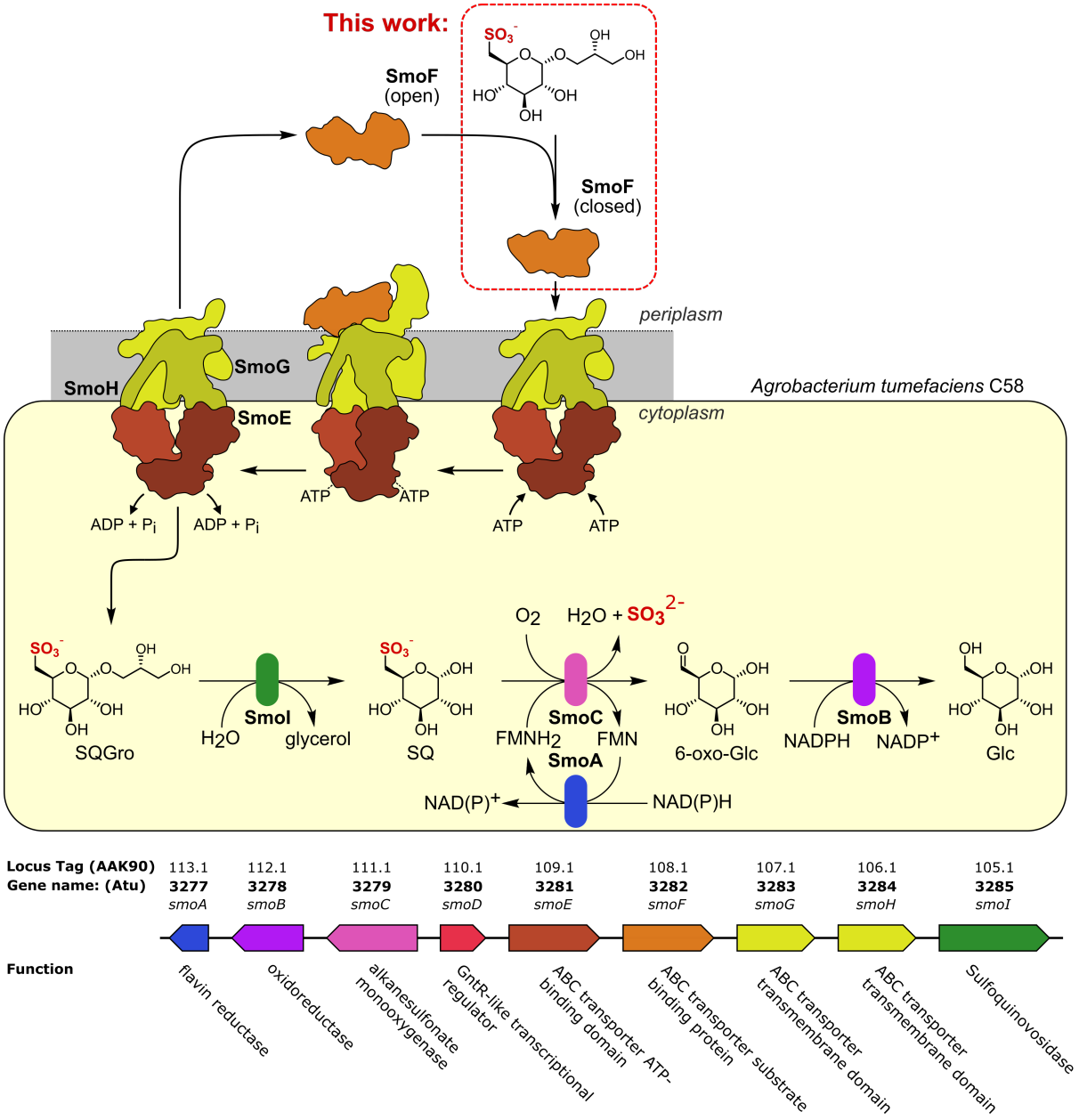
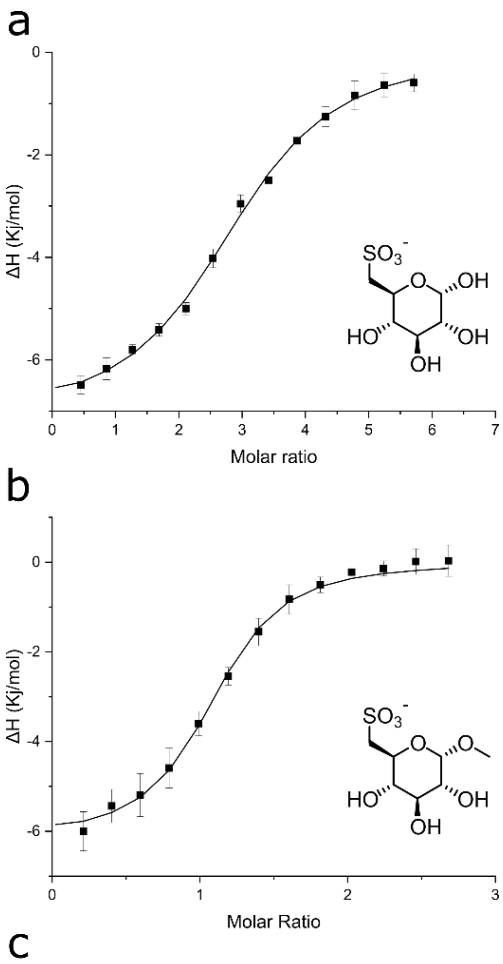


Figure 1. The sulfoquinovose monooxygenase (sulfo-SMO) pathway of *Agrobacterium tumefaciens*. The SQ-Gro binding protein SmoF works in concert with the ABC transporter SmoE-SmoG-SmoH to transport SQ metabolites into the cytoplasm. SQ glycosides are cleaved by sulfoquinovosidase SmoI, and SQ is converted to glucose and sulfite by $FMNH_2$-dependent monooxygenase SmoC and 6-oxoglucose reductase SmoB.



c

Condition	Tm (°C)	Tm shift (°C)	K_D (μM)
Ligand-free	43.9±0.1	-	-
SQGro	59.2±0.8	15.3	0.29±0.17
SQ	54.2±0.8	10.3	8.65±2.4
SQMe	58.3±2.0	14.4	11.5±2.3
SQDG (C4:0/C16:0)	51.8±0.8	7.9	-
SQDG (C18:1/C16:0)	42.9±0.2	-1.0	-

Figure 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 (T_m) of SmoF, as determined by differential scanning fluorimetry, the T_m shift relative to apo-SmoF, and K_d 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.

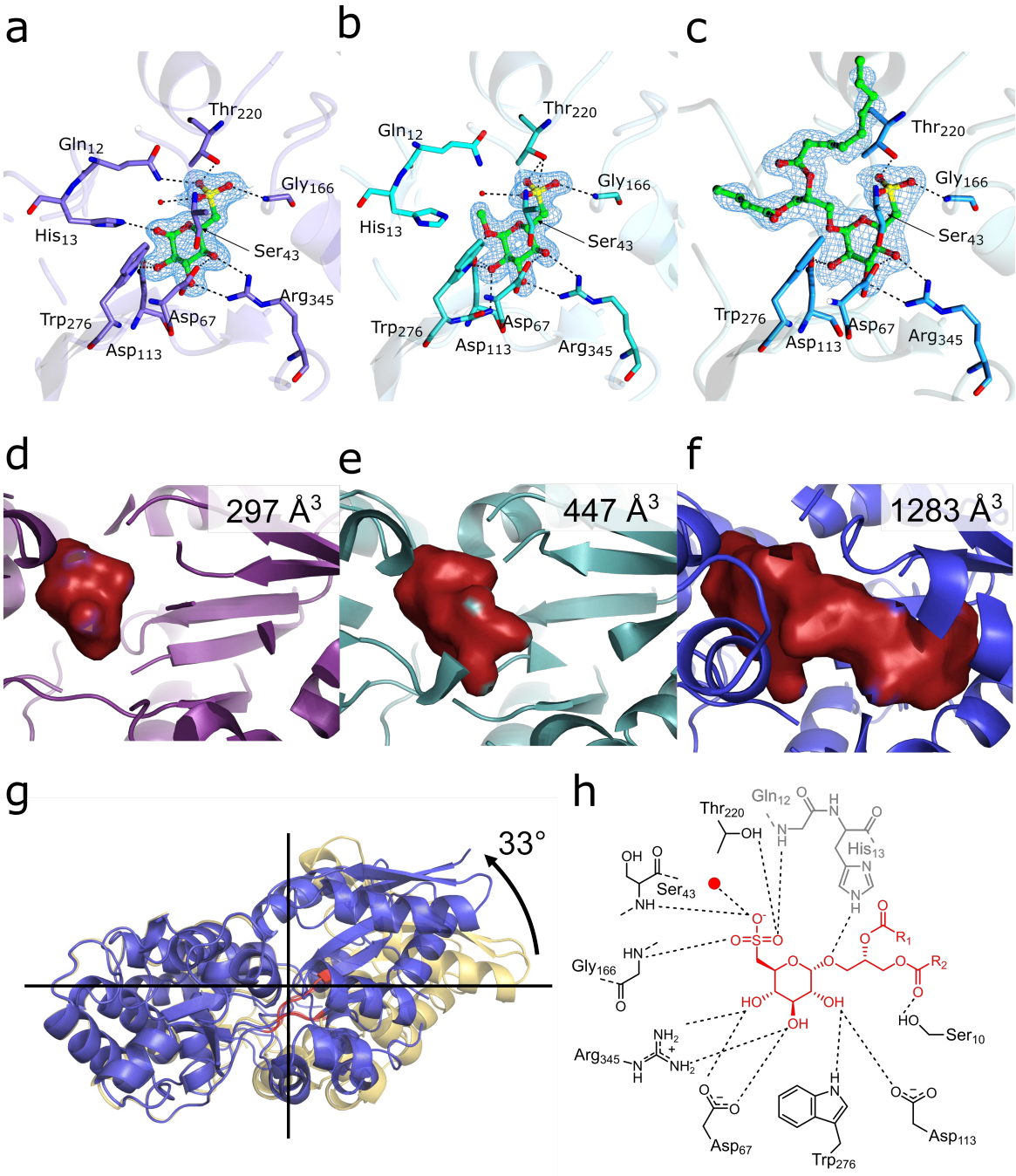


Figure 3. Structural basis and induced conformational changes for binding of SQ and its derivatives to SmoF. SmoF-ligand complex formation with a) SQ, b) SQMe, and c) SQDG-(C4:0/C16:0). d-f) Internal cavities filled by SQ, SQMe and SQDG-(C4:0/C16:0), as detected using the CASTp server. g) Superposition of ligand-free SmoF (yellow) and complex with SQDG-(C4:0/C16:0) (blue). Hinge angle and domain selection performed using the Dyn-Dom web server, and hinge residues highlighted in red. h) Binding interactions of SQ and glycosides. Interactions present in SQ complex but not SQMe or SQDG-(C4:0/C16:0) in grey. In all cases electron density (2Fo-Fc) has been contoured to 1.0 σ or 0.44 e/Å^3 for SmoF•SQ, 0.61 e/Å^3 for SmoF•SQMe and 0.44 e/Å^3 for SmoF•SQDG-(C16:0/C4:0).

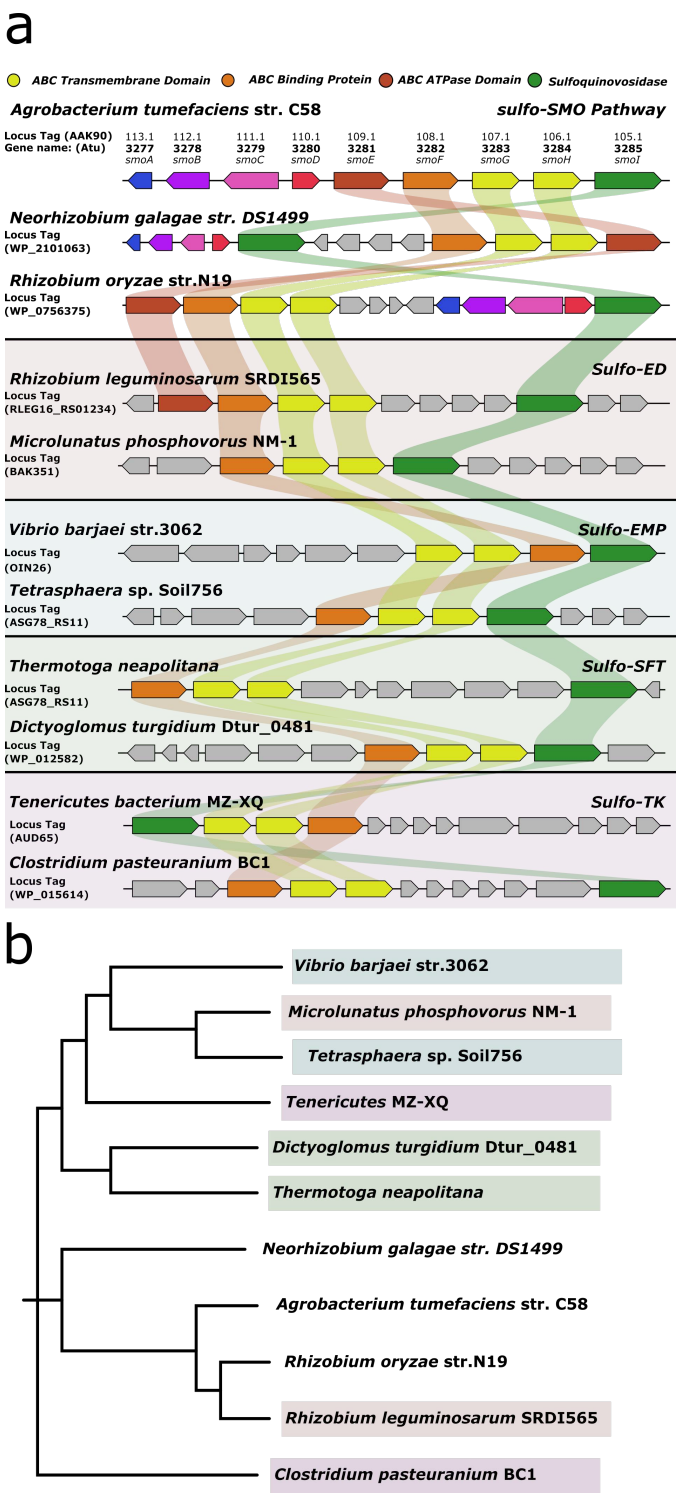


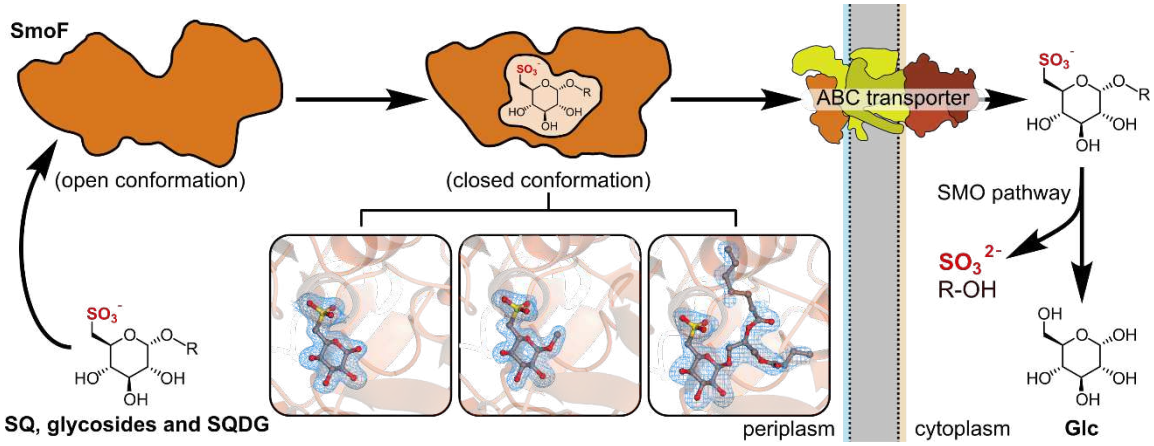
Figure 4. Comparison of *A. tumefaciens* sulfo-SMO gene cluster with other proposed sulfoglycolytic gene clusters containing SmoF homologues. a) Gene clusters for sulfoglycolytic sulfo-SMO, sulfo-ED, sulfo-EMP, sulfo-TAL and sulfo-TK pathways containing ABC transporters featuring a sulfoquinovose binding protein. b) Cladogram of SmoF and homologues found in different organisms featuring sulfoglycolysis pathways in their core genome.

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
α, β, ɣ (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 91.54, 90.0
Resolution (Å)	88.5-1.80 (1.84-1.80)	49.6-1.59(1.62-1.59)	69.6-2.14 (2.20-2.14)
R_{merge}	0.179 (2.77)	0.280 (1.38)	0.093 (0.305)
R_{pim}	0.06 (0.93)	0.148 (0.909)	0.081 (0.265)
I / σI	10.1 (1.3)	7.7 (1.3)	8.9 (3.6)
CC1/2	1.0 (0.65)	0.98(0.68)	0.99(0.94)
Completeness (%)	100 (100)	99.4 (96.1)	99.9 (100)
Redundancy	19.2 (19.1)	6.9 (5.6)	4.1 (4.1)
Refinement			
Resolution (Å)	1.8	1.59	2.14
No. unique reflections	38847	48233	24477
R_{work} / R_{free}	0.20/0.23	0.24/0.27	0.20/0.26
No. atoms			
Protein	5740	5796	11648
Ligand/ion	27	30	142
Water	138	299	239
B-factors (Å^2)			
Protein	32	20	27
Ligand/ion	23	18	26
Water	34	27	29
R.M.S. deviations			
Bond lengths (Å)	0.0146	0.0143	0.0075
Bond angles (°)	1.85	1.79	1.50
Ramachandran Plot Residues			
In most favourable regions (%)	98.4	98.1	97.8
In allowed regions (%)	1.6	1.6	2.2

Outliers	0.0	0.3	0.0
PDB code	**7YZS**	**7YZU**	**7QHV**

Graphical abstract



The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids

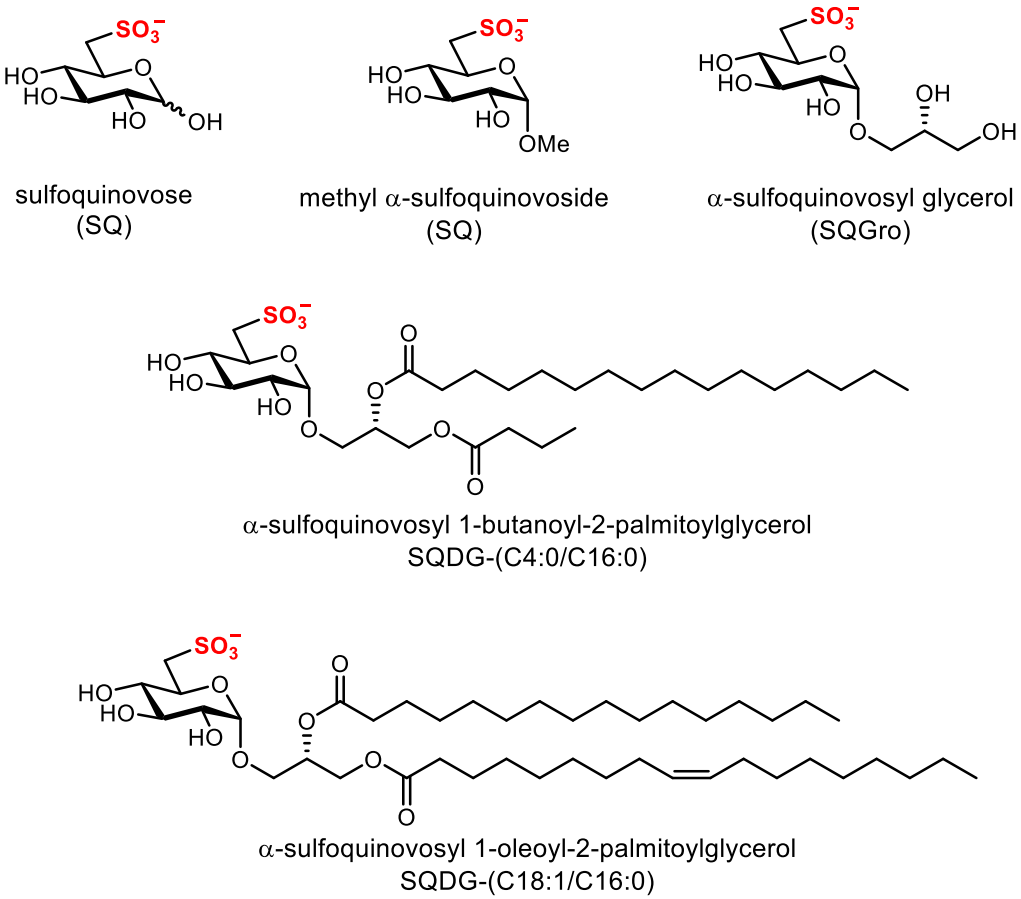
Alexander J. D. Snow,[a] Mahima Sharma,[a] James P. Lingford,[b,c] Yunyang Zhang,[d] Janice W.-Y. Mui,[d] Ruwan Epa,[d] Ethan D. Goddard-Borger,[b,c] Spencer J. Williams,[d] and Gideon J. Davies,*[a]

[a]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD

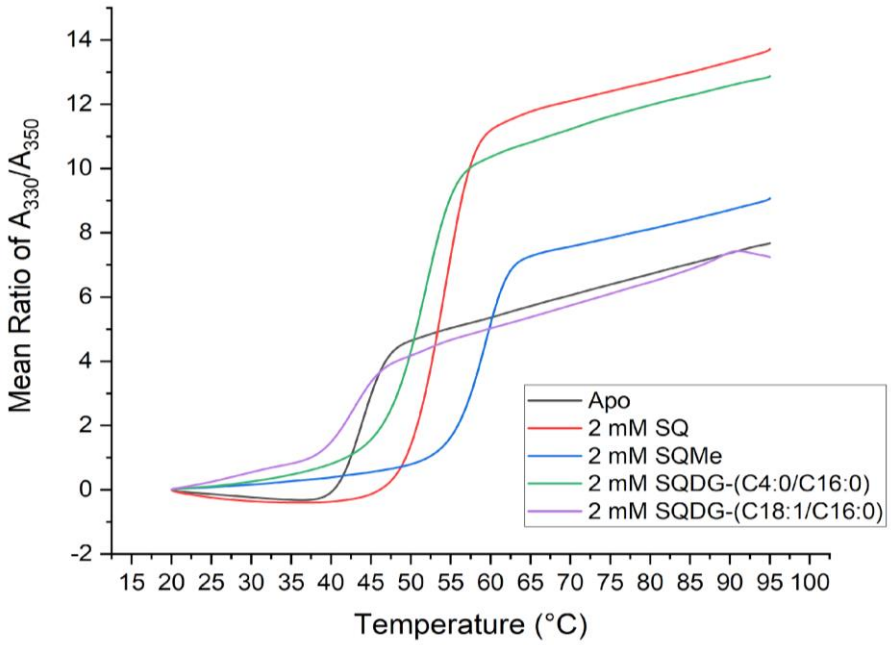
[b]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia.

[c]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3010, Australia.

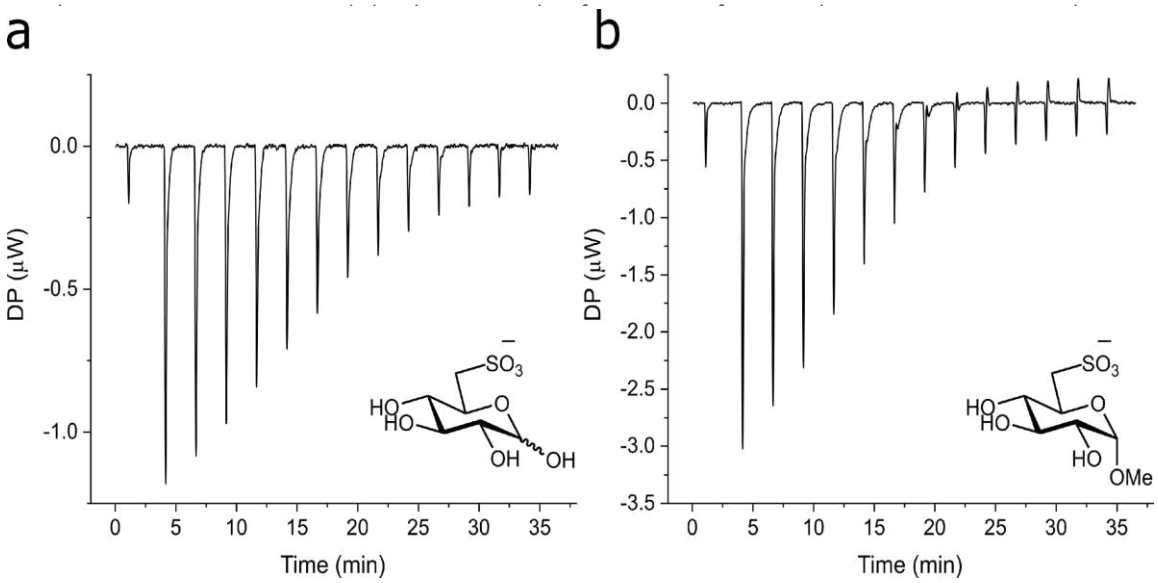
[d]School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia.



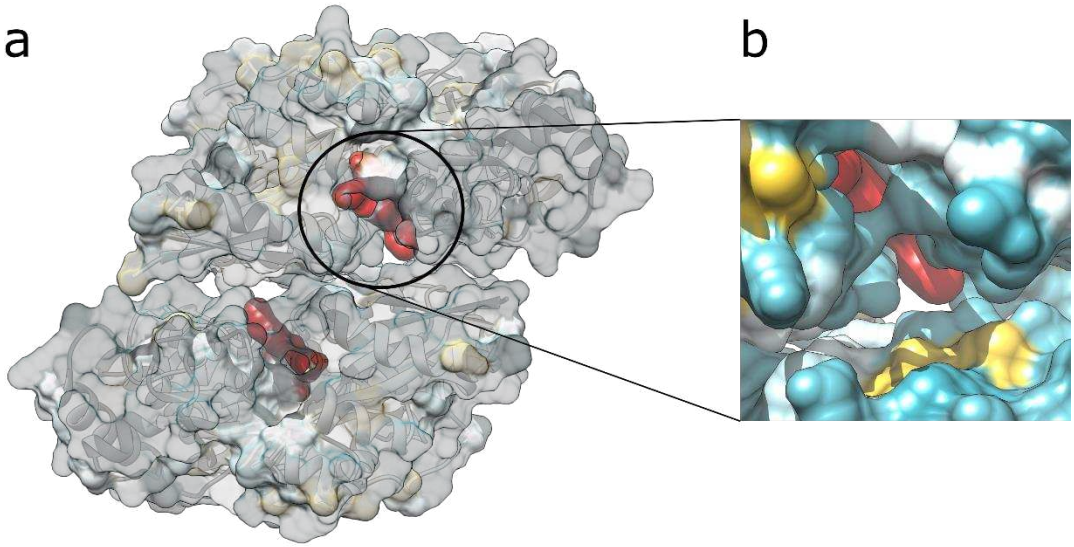
Supplementary Figure S1: Chemical structures of sulfoquinovose (SQ), SQMe, sulfoquinovosyl glycerol and sulfoquinovosyl diacylglycerols SQDG-(C4:0/C16:0) and SQDG-(C18:1/C16:0).



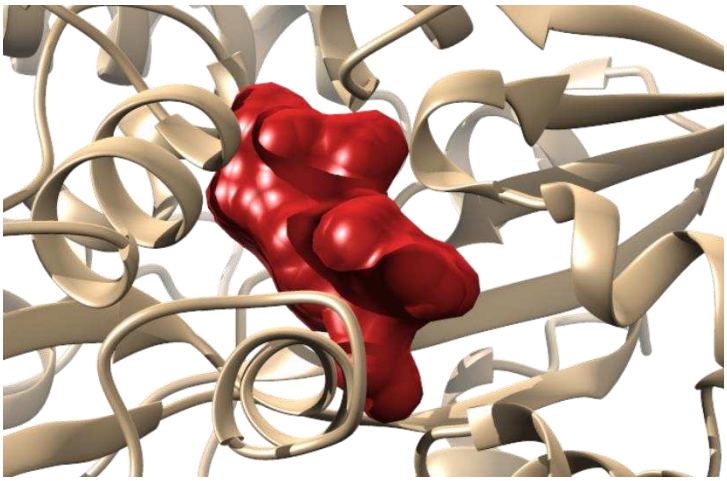
Supplementary Figure S2: Thermal unfolding profile of ligand free SmoF (black), or when incubated with 2 mM SQ (red), SQMe (blue), SQDG-(C4:0/C16:0) (green) or SQDG-(C18:1/C16:0) (purple).



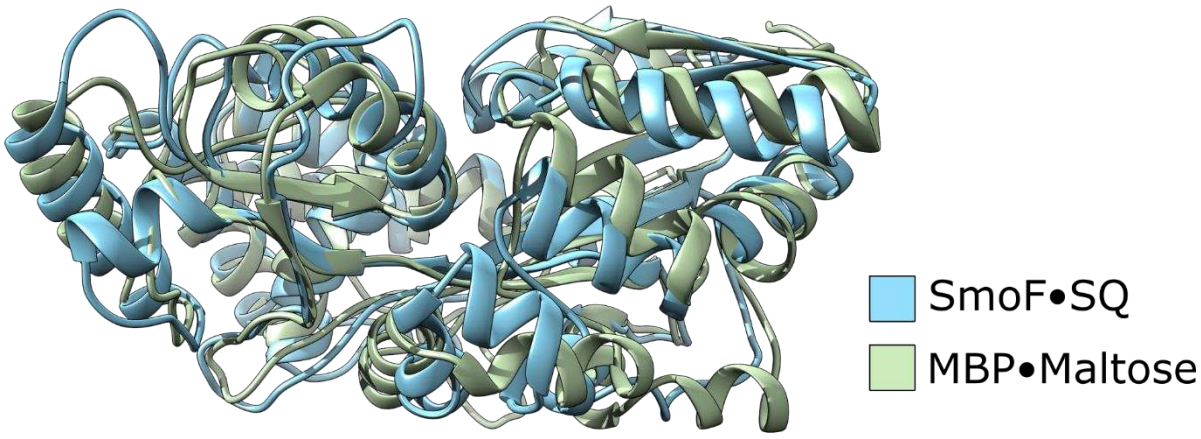
Supplementary Figure S3: Isothermal titration calorimograms for titration of SmoF with (a) SQ or (b) SQMe.



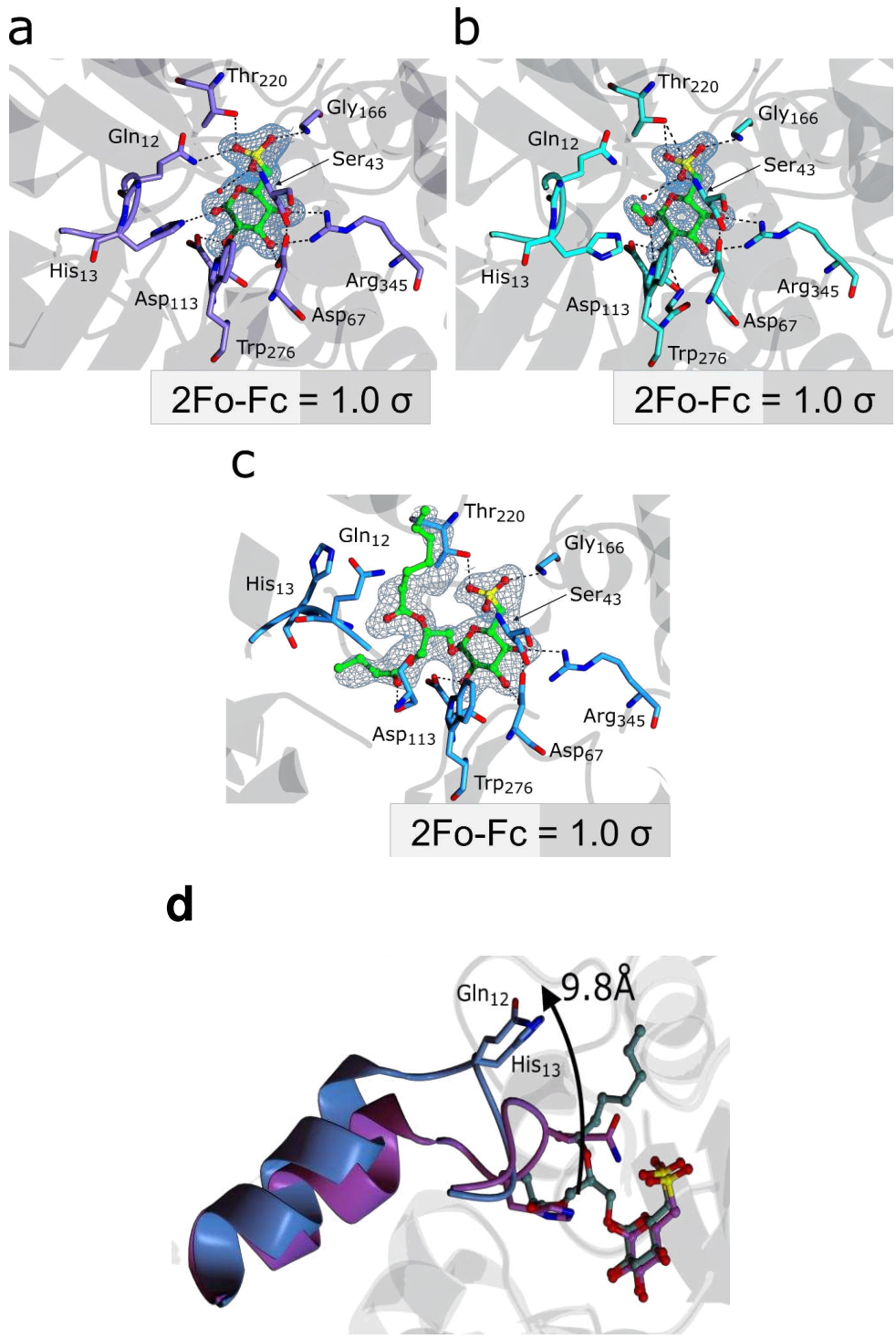
Supplementary Figure S4: Crystal contact of SmoF with SQDG-(C4:0/C16:0) showing mutual cross-chain solvent protection of acyl chain. a) Asymmetric unit of SmoF•SQDG-(C4:0/C16:0) with ligand highlighted in red. Protein has been coloured by residue hydrophobicity with blue denoting polar and yellow hydrophobic. b) inset detail of SQDG-(C4:0/C16:0) occlusion within crystal contact.



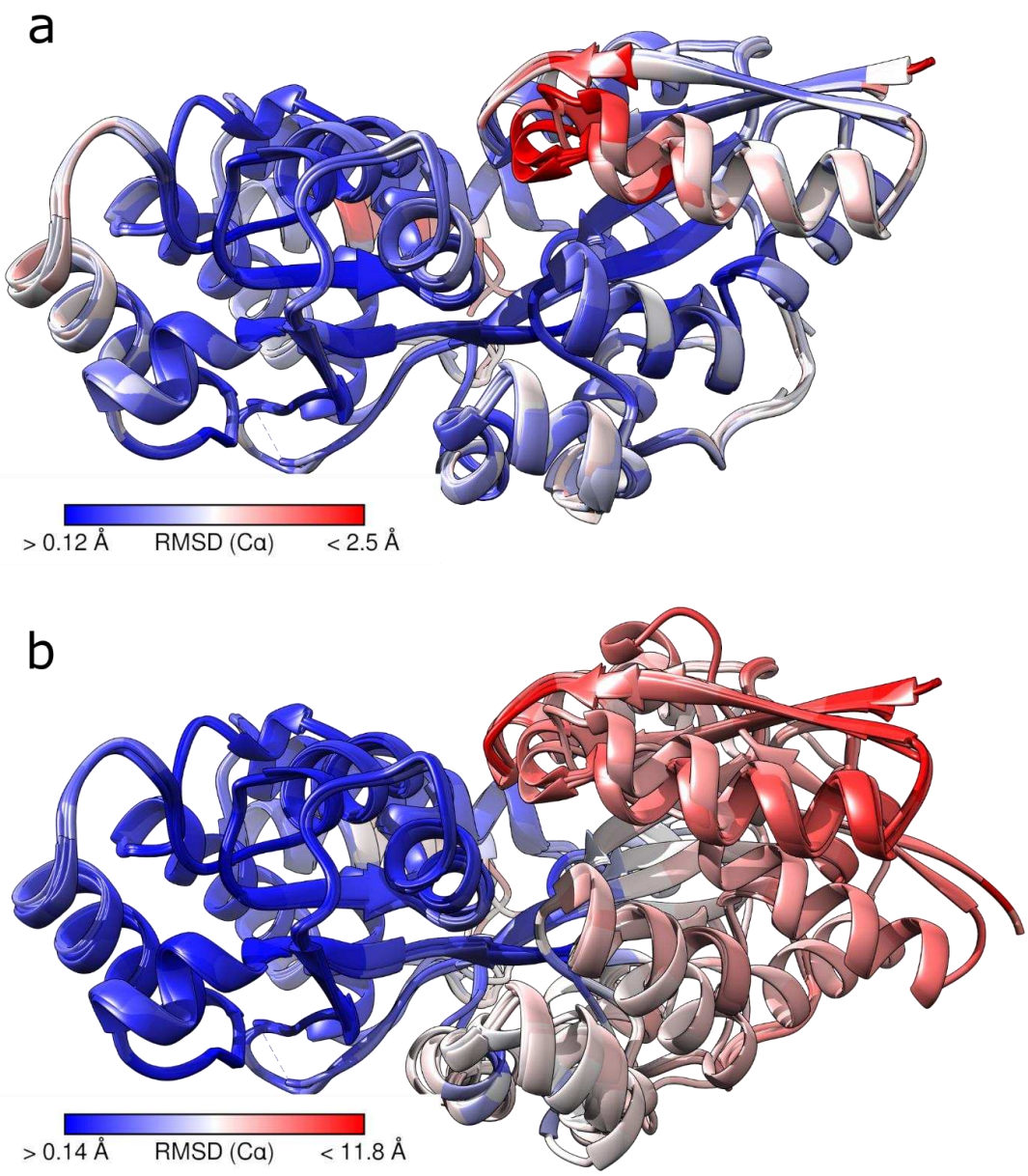
Supplementary Figure S5: Cavity detected using CASTp within the SmoF•SQGro complex. The volume of the pocket (red) is 476 Å^3.



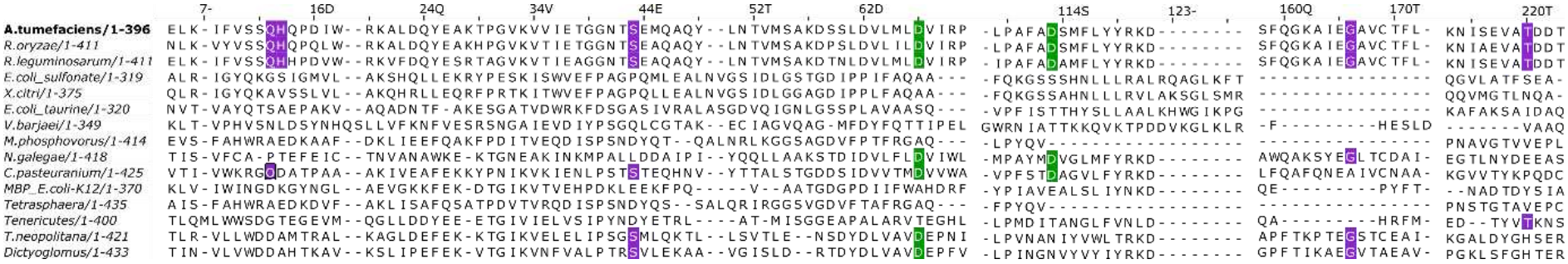
Supplementary Figure S6: Superposition of SmoF•SQ, and MBP•maltose (1MPD.pdb). SmoF is in blue, and MBP is in green. The proteins share a sequence identity of 18.9% and a global RMSD of 2.6 Å. The ligands have been omitted.



Supplementary Figure S7: Ligand binding site of SmoF including movement of the Gln_{12}-His_{13}. Structure is of complexes of SmoF with **a)** SQ **b)** SQMe **c)** SQDG-(C16:0/C4:0). In all cases electron density has been contoured identically to Figure 3 of the main work (1.0 σ or 0.44, 0.61 and 0.44 e/Å^3 respectively). **(d)** Deflection of α-helix 1, Gln12 and His13 by 9.8 Å by SQDG-(C4:0/C16:0). SQ-bound structure in purple, SQDG-(C4:0/C16:0) bound structure in blue.



Supplementary Figure S8: Variations in ligand-induced conformation changes in SmoF. In both cases residues with high RMSD (and thus large differences between conformers) between each structure are in red, low in blue. Superposition shows a global RMSD of closed conformations of <0.4 Å for SmoF•SQMe and SmoF•SQDG relative to SmoF•SQ. **A**. Conformation variety in all ligand-bound SmoF structures. **B**. Conformation variety in all ligand-bound SmoF structures versus the ligand-free 'open' conformation a rotation of 30-33°.



Supplementary Figure S9: Multiple sequence alignment of SmoF and homologues in (top to bottom) *Agrobacterium tumefaciens, Rhizobium oryzae, Rhizobium leguminosarum, E. coli* K-12 (SsuA, alkylsulfonate binding), *Xanthomonas citri, E. coli* K-12 (taurine binding), *Vibrio barjaei, Microlunatus phosphovorus, Neorhizobium galagae, Tetrasphaera, Clostridium pastueranium, E. coli* K-12 (MalE), *Thermotoga neopolitana, Tenericutes* and *Dictyoglomus tugidium.* Conserved residues featured in the sulfonate binding pocket of SmoF are highlighted in purple, and residues involved in C2-4 hydroxyl binding in green.

Supplementary Table S1. Thermodynamic values for SmoF binding to SQ, SQMe, SQDG-(C4:0/C16:0) as measured by ITC.

Condition	K_D (μM)	N (sites)	ΔH (kJ/mol)	ΔG (kJ/mol)	-TΔS (kJ/mol)
Ligand-free	-	-	-	-	-
SQ	8.65±2.41	2.89±0.08	-7.72±0.621	-28.9	-21.2
SQme	11.5±2.25	1.08±0.02	-6.88±0.282	-28.2	-21.3
SQDG (C4:0/C16:0)	-	-	-	-	-
SQDG (C18:1/C16:0)	-	-	-	-	-