A microbial sulfoquinovose monooxygenase pathway that enables sulfosugar assimilation

Mahima Sharma,[1] James P. Lingford,[2,3] Marija Petricevic,[4,5] Alexander J.D. Snow,[1] Yunyang Zhang,[4,5] Michael Jarva,[2,3] Janice W.-Y. Mui,[4,5] Nichollas E. Scott,[6] Eleanor C. Saunders,[7] Runyu Mao,[2,3] Ruwan Epa,[4,5] Bruna M. da Silva,[7,8] Douglas E. V. Pires,[7,8] David B. Ascher,[5,7] Malcolm J. McConville,[7] Gideon J. Davies,[1]* Spencer J. Williams,[4,5]* Ethan D. Goddard-Borger[2,3]*

[1] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, YO10 5DD, U.K.

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

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

[4] School of Chemistry, University of Melbourne, Parkville, Victoria 3010, Australia.

[5] Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

[6] Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Parkville, Victoria 3010, Australia.

[7] Department of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

[8] School of Computing and Information Systems, University of Melbourne, Melbourne, Victoria 3010, Australia

Keywords: carbohydrate metabolism, sulfur cycle, oxidative desulfurisation

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

Abstract

Breakdown of the sulfosugar sulfoquinovose (SQ, 6-deoxy-6-sulfoglucose), produced by photosynthetic organisms, is an important component of the biogeochemical carbon and sulfur cycles. Here, we reveal a new pathway for SQ degradation involving oxidative desulfurization to release sulfite and complete breakdown of the carbon skeleton of this sugar to support the growth of the plant pathogen *Agrobacterium tumefaciens*. SQ or its glycoside sulfoquinovosyl glycerol are imported by an ABC transporter system with associated SQ binding protein. A sulfoquinovosidase cleaves the SQ glycoside and a flavin mononucleotide-dependent sulfoquinovose monooxygenase acts in concert with an NADH-dependent flavin reductase to release sulfite and form 6-oxo-glucose. A short-chain dehydrogenase/reductase oxidoreductase reduces 6-oxo-glucose to glucose, allowing it to enter primary metabolism. Structural and biochemical studies provide detailed insights into the binding and recognition of key species along the reaction coordinate. This sulfoquinovose monooxygenase pathway is distributed across alphaproteobacteria and especially within the rhizobiales. This metabolic strategy for SQ catabolism is distinct from previously described pathways as it allows the complete utilization of all carbons with SQ by a single organism and release of inorganic sulfite.

Sulfoquinovose (SQ; 6-deoxy-6-sulfoglucose) is an anionic sulfosugar found in plant and cyanobacterial sulfolipid, and in S-layer proteins in archaea[1]. It is estimated that SQ holds around half of all sulfur in the biosphere, with 10 billion tonnes produced each year in Nature, and therefore its cycling is important for the biogeochemical sulfur cycle[2]. SQ is a carbon substrate for bacterial growth, and the release of sulfur is essential for its environmental cycling. Previously described SQ metabolic pathways involve two tiers. Tier 1 pathways, termed sulfoglycolysis, involve scission of the C3-C4 bond of SQ to give two C_3 fragments, with carbons 1-3 entering central metabolism, while carbons 4-6 and the sulfonate group are excreted as dihydroxypropanesulfonate (DHPS) or sulfolactate (SL). In the second tier SL- and DHPS-utilizing bacteria release inorganic sulfur as sulfite. Three Tier 1 pathways have been described, the sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP)[3], Entner-Doudoroff (sulfo-ED)[4, 5] and sulfofructose transaldolase (sulfo-SFT) pathways[6, 7]. Tier 2 metabolism has been described for various specialized bacteria that utilize SL or DHPS and perform 'biomineralization', releasing sulfur as sulfite, which under aerobic conditions can readily oxidize to sulfate[1]. While many of the steps in the three described Tier 1 sulfoglycolysis pathways differ, all three pathways usually share the presence of a specialized glycoside hydrolase, a sulfoquinovosidase (SQase), which catalyzes the hydrolysis of SQ glycosides such as SQDG and SQGro to release SQ[8, 9].

While extant sulfoglycolytic pathways release C3-sulfonates, there is evidence for an additional pathway: Roy and co-workers reported that an *Agrobacterium* strain from soil could accomplish the complete metabolism of SQ, releasing sulfate, but the details of this pathway were not identified[10]. We previously reported that *A. tumefaciens* C58 contains a functional SQase, with the ability to hydrolyze SQGro[8]. However, analysis of the genome of this organism did not reveal any genes homologous to those expected for known Tier 1 sulfoglycolysis pathways. Here, we identify a new pathway in *Agrobacterium tumefaciens* strain C58 that effects the oxidoreductive desulfurization of SQ to release sulfite and the complete degradation of the carbon skeleton of SQ. We show that this pathway involves a novel SQ/SQGro solute binding protein and an ATP-binding cassette (ABC) transporter, an SQase to release SQ from its glycoside, a flavin-dependent SQ monooxygenase and an NADPH-dependent reductase that collaborate to oxidoreductively desulfurize SQ to produce glucose and sulfite. X-ray structures reveal the molecular basis of substrate binding and catalysis. This pathway is distributed across alphaproteobacteria and is especially represented within the Rhizobiales.

Results

A metabolic gene cluster is highly expressed when Agrobacterium tumefaciens C58 grows on sulfoquinovose

To establish whether *A. tumefaciens* C58 can utilize SQ, we inoculated M9 media containing SQ as sole carbon source. *A. tumefaciens* C58 exhibited robust growth, but analysis of spent culture media did not reveal DHPS or SL. Instead, the supernatant accumulated sulfate, but with a lag between consumption of SQ and sulfate release (**Fig. 1a**), as was previously reported by Roy and co-workers for *Agrobacterium* sp. strain ABR2[10]. Noting that sulfite is generally released from organosulfonate degradation pathways[1, 11], we analyzed for sulfite (SO_3^{2-}), and observed that SQ consumption is coincident with production of sulfite, which slowly undergoes autooxidation to sulfate. To investigate the metabolism of the carbon skeleton of SQ, we cultured *A. tumefaciens* on $^{13}C_6$-SQ[12] and analyzed the culture supernatant using ^{13}C NMR spectroscopy (**Supplementary Fig. 1**). The only significant ^{13}C-labelled product we could detect was ^{13}C-bicarbonate, which formed transiently during exponential phase growth, and the ^{13}C-labelled bicarbonate signal disappeared at stationary phase, presumably through exchange with atmospheric CO_2. *A. tumefaciens* also grew on SQGro but did not grow on other alkylsulfonates including DHPS, SL, sulfoacetic acid, taurine, pentanesulfonate, MES, MOPS, HEPES, PIPES, cysteic acid or methanesulfonic acid (**Supplementary Fig. 2**). Collectively, this data demonstrates that *A. tumefaciens* effects the complete metabolism of the carbon backbone of SQ, with release of sulfite into the media.

We undertook comparative proteomics to identify the protein changes associated with *A. tumefaciens* growth on SQ compared to glucose, at mid-log phase (**Fig. 1b**). Growth on SQ resulted in multiple proteome changes with the largest alteration corresponding to the increased abundance of proteins within a single region of the genome, *Atu3277-Atu3285*. Proteins corresponding to genes within this cluster were robustly and statistically more abundant during SQ growth with the unobserved proteins Atu3283 and Atu3284 corresponding to transmembrane proteins, a class of proteins known to be difficult to detect using standard proteomics approaches[13]. Therefore, we assign the operon encoding genes *Atu3277-Atu3285* (*smoA-smoI*) as responsible for the ability of *A. tumefaciens* to grow on SQ (**Fig. 1c**). This operon encodes Atu3285 (SmoI), previously identified as an SQase[8], but the remaining predicted proteins were not annotated with functions consistent with either a sulfo-EMP or sulfo-ED pathway, suggesting that *A. tumefaciens* uses an undescribed pathway to metabolize SQ. The automated bioinformatic annotations of the respective genes that highlight the presence of an ABC transporter system, and putative sulfonate monooxygenase, SDR oxidoreductase, flavin reductase and exporters. Structural and biochemical studies outlined below demonstrate that these proteins encode a novel SQ import and oxidoreductive desulfurization system that releases sulfite (**Fig. 1d**).

SmoF is an ABC transporter solute-binding protein that binds SQGro

The operon encodes a predicted ABC transporter system, Atu3281 (SmoE), Atu3283 (SmoG), and Atu3284 (SmoH) with associated periplasmic solute binding protein, Atu3282 (SmoF). Solute binding proteins can provide insights into the function of their associated ABC transporters[14]. Recombinantly-expressed SmoF bound SQGro with K_d = 200 nM (ΔH = -11 kcal mol^{-1}, ΔS = 5 cal mol^{-1} deg^{-1}) indicating a preference for the former (**Fig. 2a**, **Supplementary Table S3**). No binding was observed for the stereochemically-related monosaccharides D-glucose and D-glucuronic acid.

To delineate how sulfosugars bind to SmoF, we obtained high-resolution 3D crystal structures in its ligand-free apo and SQGro-bound complex (**Fig. 2b**). SmoF (MW 43,000 Da) contains two globular domains with similar secondary α/β fold and a deep cleft lined with aromatic and polar residues to capture its cognate solute molecules. A DALI search against PDB25 entries, a representative subset of all structures in the Protein Data Bank, suggested high structural similarity to solute binding proteins associated with ATP-binding cassette importers, such as periplasmic sugar-binding protein from *Thermus thermophilus* HB8 (PDB ID: 6JAL with DALI z score of 48.0, rmsd 2.2 and 36% sequence identity), D-mannitol bound solute-binding protein from *Agrobacterium vitis* S4 (PDB ID: 4RYA with DALI z score of 43.8, rmsd 2.7 and 24% sequence identity) and glycerol-3-phosphate binding periplasmic binding protein from *E. coli* (PDB ID: 4AQ4 with z score of 37.5 and 16% sequence identity). Comparison of the structures of ligand-free SmoF and the SQGro bound complex reveals a large conformational change of the protein from inter-domain rotation upon binding the sulfosugar. The relative movement of domains was assessed using DynDom server, which indicated a hinge rotation of 31º about four linker regions connecting the two domains (**Supplementary Fig. 4**). SQGro is buried deep within the inter-domain cleft and several residues from both domains interact with the sugar hydroxyls and make electrostatic interactions with the sulfonate group. The sulfonate oxygens are recognized by hydrogen bonding to side-chain hydroxyl of Thr220 (at a distance of 2.6 Å), backbone amides of Gly166 (3 Å) and Ser43 (2.8 Å), and binds an ordered water molecule that in turn makes a hydrogen bonding interaction with the sidechain of His13 (3 Å) and Gln46 (3.2 Å) (**Fig. 2c,d**). Thermodynamic stability of the 'closed' state relative to the 'open' form was confirmed by a significant shift in melting temperature of SmoF by 15 ºC upon binding the sulfonate ligand (**Supplementary Fig. 5**).

The structural basis of substrate recognition by SQase SmoI (Atu3285)

We previously reported that SmoI is an SQase that preferentially hydrolyses 2'*R*-SQGro, the natural stereoisomer of this glycoside[8]. To understand the molecular basis of binding of this compound to

SmoI, we determined a 3D structure of the 'Michaelis' complex with 2'*R*-SQGro using an inactive acid/base mutant D455N (**Fig. 2e,f**). SmoI-D455N•SQGro crystallized in $P2_1$ space group with four protomers in the asymmetric unit, each showing unambiguous density of the substrate bound at the active site. The overall fold is an $(\alpha/\beta)_8$ barrel appended with small β sheet domain and the sulfonate group is recognized by Arg283/Trp286/Tyr491 triad[8]. Interactions with the substrate glyceryl moiety involves hydrogen-bonds to Arg438 and Glu135. Phe280 contained within the flexible loop 276-281 at the substrate entry site, previously reported to make π-π interactions with artificial substrate PNPSQ,[8] was observed in different orientations in the different protomers within this structure.

SmoA (Atu3277) is a flavin mononucleotide (FMN) reductase

Recombinant SmoA was expressed in *E. coli* and purified by sequential immobilized metal affinity, size exclusion and ion-exchange chromatography. Throughout this process the protein solution was yellow in color, suggesting that it co-purified with a flavin co-factor. A sample of protein was denatured using 1 M hydrochloric acid to release the co-factor and analysis of the supernatant by LC-MS revealed FMN as the sole detectable flavin (**Supplementary Fig. 6**). Michaelis-Menten kinetics were conducted for SmoA with saturating FMN and NADH and NADPH to determine which of these cofactors was preferred by the reductase. With NADH the kinetic parameters were K_M = 35±5 μM, k_{cat} = 14.5±0.5 s^{-1} and k_{cat}/K_M = 4.1×10^5 M^{-1} s^{-1}; while for NADPH saturation was not observed and k_{cat}/K_M = 6.8×10^2 M^{-1} s^{-1}, indicating that NADH is the preferred cofactor (**Fig. 3a, Supplementary Fig. 7)**. Owing to difficulties in obtaining a high-resolution 3D structure by X-ray diffraction (*vide infra*) we also studied a close homologue from *Rhizobium oryzae*. *R. oryzae* possesses a syntenic operon that contains putative sulfo-SMO operon including genes homologous to *smoA* and *smoC* (**Fig. 5a**). Recombinantly expressed *Ro*SmoA from *R. oryzae* also released FMN upon acid treatment (**Supplementary Fig. 6**) and was also NADH-dependent with K_M = 16 μM, k_{cat} = 33 s^{-1} and k_{cat}/K_M = 2.1 × 10^6 M^{-1} s^{-1} (**Supplementary Fig. 7)**.

SmoC (Atu3279) is an SQ monooxygenase that catalyses the oxidative desulfurization of SQ

SmoC shares low sequence similarity with SsuD, an alkanesulfonate monooxygenase that catalyzes the $FMNH_2$-dependent oxidation of the α-carbon of alkylsulfonates to form an α-hydroxysulfonate that eliminates sulfite to produce an aldehyde[15]. To demonstrate activity for SmoC, we initially focused on the detection of sulfite release from SQ, since the putative sugar product, 6-oxo-glucose (6-OG), exists in a complex equilibrium of (hemi)acetals and hydrates that are difficult to detect and characterise directly. A sulfite detection assay based on Ellman's reagent enabled monitoring of sulfite release by the combination of SmoA and SmoC in the presence of FMN, NADH and SQ, which were necessary and sufficient for enzyme-catalyzed sulfite release (**Fig. 3b**). Maximal

substrate conversion was approximately 200 µM, which is commensurate with the solubility of molecular oxygen in water under standard conditions, with peak activity observed at pH 8.5 (**Supplementary Fig. 8**). No activity was observed for SQGro or HEPES, an unrelated sulfonate, demonstrating the specificity of the monooxygenase for SQ and that SQGro hydrolysis by SQase necessarily precedes oxidative desulfurisation. SQ binds to SmoC with a K_d 3 µM in the absence of any flavin-based cofactors (**Fig. 3c**, **Supplementary Table S3**). FMN demonstrated no detectable affinity for SmoC by isothermal titration calorimetry, commensurate with $FMNH_2$ being the co-substrate for this enzyme (**Supplementary Table S3**). The homolog from *R. oryzae*, *Ro*SmoC, exhibited similar activity and substrate selectivity to SmoC (**Supplementary Fig. 8**).

Size exclusion chromatography-multiangle light scattering (SEC-MALS) shows that SmoC exists as a dimer in solution (**Supplementary Fig. 9**). While we could crystallize SmoC, only poor crystals were obtained that diffracted to a maximum of 3.4 Å. Better results were obtained with a homologue from *Rhizobium oryzae* (hereafter *Ro*SmoC), which diffracted to 1.9 Å; the low-resolution *A. tumefaciens* SmoC apo structure superposes with *Ro*SmoC with rmsd of 0.4 over the entire structure indicating identical structural and functional features (**Supplementary Fig. 10**). The monooxygenases consist of a core $(\alpha/\beta)_8$ TIM barrel with three additional insertion regions, similar to monooxygenases from the bacterial luciferase family. The protomers exist as a homodimer with an extended, flat subunit interface with a buried surface area of 4697 $Å^2$ (amounting to 18% of total accessible surface area) (**Fig. 3d**). Pairwise structural analysis using the DALI server identified close relationships to a putative luciferase-like monooxygenase (3RAO.pdb) with an rmsd of 2.4 over 314 residues and a Z score of 34.3, long-chain alkane monooxygenase LadA (3B9O.pdb, rmsd 2.6/312 residues, Z-score of 31.0), and $FMNH_2$-dependent alkane sulfonate monooxygenase SsuD (1M41.pdb, rmsd 1.8/317 residues, Z-score of 41.2).

Comparison of the *Ro*SmoC structure with LadA (3B9O.pdb) in complex with coenzyme allowed identification of the FMN binding site as a deep hydrophobic pocket that accommodates the modelled isoalloxazine ring system extending up to surface guarded by the conserved phosphate binding residues Tyr136 and Ser189 (**Fig. 3e**)[16]. A close structural and functional relationship to alkane sulfonate monooxygenase SsuD (1M41.pdb) was evident from the presence of a putative sulfonate substrate binding site lined with positively charged side-chains: Trp206, Arg236, His238, Tyr341 and His343 (**Fig. 3f**). Aside from conferring specificity for binding sulfonate, these conserved active-site residues have previously been suggested to contribute to the stabilization of a peroxyflavin intermediate[16, 17].

SmoB (Atu3278) is an NADPH-dependent reductase that converts 6-oxo-glucose to glucose

As 6-OG is a poorly behaved species that confounded direct analysis, we studied the reaction of SmoB using equilibrium isotope incorporation (**Fig. 4a**). Incubation of SmoB and co-factor with D-glucose in ^{18}O-water should allow transient formation of 6-OG, which will undergo hydration to make an ^{18}O-labelled hydrate. The reverse of this reaction involving dehydration and reduction should lead to 50% incorporation of the label to give C6-^{18}O-glucose. Repeated cycles of oxidation/reduction will lead to increased levels of ^{18}O incorporation. However, a side-reaction will involve exchange of the ^{18}O-label at C1 through similar hydration/dehydration reactions. To avoid the complicating effect of this side-reaction on mass spectrometric analysis, we used C1-^{18}O-glucose as substrate. Mass spectrometric analysis of mixtures of C1-^{18}O-glucose in ^{18}O-water with SmoB with NAD^+ and $NADP^+$ revealed the formation of an M+4 product, arising from the incorporation of two atoms of ^{18}O. To confirm that the product is in fact glucose, we acetylated the crude reaction mixture (Ac_2O/pyridine) and subjected the product to LC-MS analysis. This revealed the formation of material that co-eluted with authentic D-glucose-pentaacetate, and gave signals corresponding to the sodium adduct of glucose pentaacetate plus 2 and 4 Da, consistent with incorporation of one and two ^{18}O atoms (**Supplementary Fig. 11**). We next conducted electron-impact GC-MS to locate the ^{18}O label, by converting the labelled glucose to the acyclic pentapropionate aldonitrile (**Supplementary Fig. 12**)[18]. This method gives rise to diagnostic fragment ions, namely C1-C5 and C5-C6 ions. The ^{18}O-labelled product gave a C5-C6 fragment that was 2 mass units higher (*m*/*z* 173 versus 175), whereas the C1-C5 fragment was the same as unlabelled glucose reference (*m*/*z* 370), demonstrating that the ^{18}O label is located at C6. Only enzymatic reactions conducted in the presence of $NADP^+$ produced product labelled with ^{18}O at C6, defining the cofactor specificity of SmoB; the solution with NAD^+ was inert. ITC confirmed high-affinity binding to NADPH with a K_d of ~2 μM with no observable binding of NADH, ruling out dual cofactor specificity for SmoB (**Supplementary Fig. 13, Supplementary Table S3**).

The 3D X-ray structure of SmoB revealed a compact trimer with the C-terminal His_6-tag from the adjoining subunit blocking the putative active site, thereby precluding cofactor binding (**Supplementary Fig. 14**). By altering the location to an N-terminal cleavable purification tag, an SmoB-LIC3C construct was generated, which yielded an active enzyme and SEC-MALS confirmed a trimeric biological assembly in solution (**Supplementary Fig. 15**). SmoB was co-crystallized with NADPH and a ternary product complex was obtained by soaking crystals with D-glucose (**Fig. 4b**). SmoB is an $(\alpha/\beta)_8$ TIM barrel fold with a C-terminal cofactor binding site. The overall fold shows high structural conservation with members of aldo-keto reductase (AKR) superfamily.

SmoB binds NADPH with the 2'-phosphate oxygens hydrogen-bonded to Thr284, Arg289 and backbone amide of Asn285 and the adenine ring stacked between Arg289 and Phe241 at the C-terminus (**Fig. 4c**). NADPH binds in an extended *anti*-conformation and the nicotinamide ring is located at the base of the substrate binding pocket. Trp232 makes π-π stacking interaction that positions the C-4 of the nicotinamide ring at an appropriate distance of 3 Å from carbon C-6 of glucose, poised for hydride transfer (**Fig. 4d**). Within the SmoB•$NADP^+$•glucose complex glucose interacts with Arg152 (2.9 Å) and Lys120 (3 Å), as well as His151 (2.8 Å) and Tyr76 (2.7 Å) within the conserved catalytic tetrad His/Tyr/Lys/Asp common to the AKR superfamily (**Fig. 4e**)[19].

SQ oxidative desulfurization pathway is distributed across the alphaproteobacteria

A Multigene BLAST search of the non-redundant protein set of the NCBI for gene clusters that contain homologous SQases and SQ monoxygenases identified putative Smo gene clusters across the Agrobacterium and Rhizobium class within the *Rhizobiales*, and evidence for limited expansion into other alphaproteobacteria and betaproteobacteria (**Fig 5**). These gene clusters were both similarly and differently assembled (ie non-syntenic) and some lacked the ABC-transporter system identified in *A. tumefaciens*. However, different SQ transporter systems have been identified in other sulfoglycolytic gene clusters, and ABC-transporter systems have been noted in sulfo-ED pathways[4, 5], suggesting that these transporters may be interchanged among different sulfoglycolytic pathways. The sulfo-ED pathway has been identified in a range of Rhizobiales[4, 5], suggesting that different pathways can support sulfoglycolysis in this bacterial order.

Discussion

This work highlights a new sulfoglycolytic 'oxidoreductive' pathway that involves scission of the C–S bond of SQ, and therefore allows the complete catabolism of SQ. Unlike previously reported pathways, this occurs within a single organism, *A. tumefaciens*, representing similar operons found in other Rhizobiales and alphaproteobacteria. The pathway involves several novel proteins: an SQ binding protein, a flavin mononucleotide-dependent SQ monooxygenase that releases sulfite, and an NADPH-dependent 6-OG dehydrogenase. The pathway appears to be the same as that described for an unclassified *Agrobacterium* sp. strain isolated almost 20 years ago[10]. Like most other sulfoglycolytic pathways, the sulfo-SMO pathway contains a highly conserved SQase, which allows use of SQ glycosides such as SQDG and SQGro[8, 9].

The sulfo-SMO pathway shares similarity with other known metabolic pathways. The presence of an SQase, SQGro binding protein and ABC cassette is reminiscent of MalP (maltodextrin phosphorylase), MalE (maltose binding protein), MalF/MalG (intrinsic membrane proteins) and

MalK (transport ATPase) expressed by the *mal* operon that is used by *E. coli* to import and cleave maltose[20], and the SQ monooxygenase SmoC and associated flavin recycling enzyme SmoA are related to SsuD ($FMNH_2$-dependent alkylsulfonate monooxygenase) and SsuE (NADPH-dependent FMN reductase) expressed by the *ssu* operon that allows *E. coli* to degrade alkylsulfonates[15]. The structural studies highlight key residues in the sulfosugar processing enzymes that have evolved to bind this distinguishing group: an Arg283-Tryp286-H_2O(Tyr491) triad for recognition of the sulfonate in the SQase; a Thr220-Gly166-Ser43-H_2O(His13-Gln46) cluster for recognition of the sulfonate in the SQGro binding protein; and a proposed binding pocket of Trp206-Arg236-His238-Tyr341-His343 for recognition of sulfonate in the SQ monooxygenase. Well-defined sulfonate binding pockets have been highlighted for the enzymes of the sulfo-EMP pathway[8, 21, 22] and are useful sequence signatures for bioinformatic studies and assignment of the pathways in unstudied organisms.

The occurrence of the sulfo-SMO pathway in bacteria of the order Rhizobiales is interesting as many members of this grouping are plant symbionts or pathogens. However, the sulfo-ED pathway has also been identified in other Rhizobiales[4]. Sulfoglycolytic pathways may contribute to growth of pathogenic Rhizobiales on plants, allowing use of SQDG sourced from photosynthetic tissues. Rhizobiales are also free-living organisms and the sulfo-SMO pathway may be useful in the adoption of an oligotrophic saprophytic lifestyle in soil and on decaying plant matter. Symbiotic Rhizobiales form bacteroides that utilize C4-substrates for energy and central metabolism[23]. Sugawara and co-workers showed that sulfonate utilization gene clusters were expressed by *Bradyrhizobium diazoefficiens* USDA 110 within symbiotic nodules and may be important for utilizing diverse sulfur sources to support free-living and symbiotic lifestyles[24]. The requirement of oxygen for the SQ monooxygenase is suggestive that this pathway may be limited to organisms capable of aerobic growth. Previously described pathways for SQ degradation result in release of C3-sulfonates that support the growth of Tier 2 sulfonate degrading bacteria, and thus support the growth of communities of bacteria[1]. The sulfo-SMO pathway reported here results in complete consumption of SQ and thus represents a 'selfish' mechanism for utilization of this monosaccharide, and may provide an advantage in the highly competitive environment of soil as well as supporting the development of symbiosis on photosynthetic, SQ-producing organisms.

Methods

Growth studies

Cultures of *A. tumefaciens* C58 were grown in a phosphate-buffered mineral salts media (M9, pH 7.2), with Glc or SQ (10 mM) as the sole carbon source. Cultures were incubated at 30 °C (250 rpm), with adaptation and robust growth observed within 2–3 days. These were sub-cultured (1% inoculum) into the same media (10 ml) and grown at 30 °C (250 rpm). Bacterial growth was quantitated using a Varian Cary50 UV/visible spectrophotometer to measure OD_{600}. Growth experiments were replicated twice.

Reducing sugar assay

The reducing sugar assay was performed according to the procedure of Blakeney and Mutton[25]. This assay uses pre-prepared alkaline diluent and PAHBAH working solution. *Alkaline diluent:* sodium hydroxide (20 g, 0.5 mol) was added to a solution of 0.10 M trisodium citrate (0.05 mol, 0.5 L) and 0.02 M calcium chloride (0.013 mol, 0.5 L). *PAHBAH working solution:* This was made freshly immediately before use by dissolving 4-hydroxybenzhydrazide (PAHBAH) (0.25 g, 1.64 mmol) in alkaline diluent (50 mL). The PAHBAH working solution should be freshly made shortly before use. *Procedure for reducing sugar assay:* 0.90 mL of PAHBAH working solution was added to 0.10 mL of sample. The mixture was heated at 98 °C for 4 min. 0.5 mL of the mixture was diluted into 1 mL of deionized water and the absorbance read at 415 nm using a UV/visible spectrophotometer. Concentrations of SQ were determined by reference to a standard curve constructed using SQ.

Turbidometric sulfate assay

The sulfate assay was followed according to the procedure of Sörbo[26], with reference to a standard curve constructed using known concentrations of sodium sulfate. The Ba-PEG reagent contains PEG to stabilize barium sulfate crystals, and a small amount of pre-formed $BaSO_4$ seed crystals to improve reproducibility and linearity of the assay. The Ba-PEG reagent should be freshly prepared before use. *Ba-PEG reagent*: $BaCl_2$ (41.7 mg, 0.20 mmol) and polyethylene glycol 6000 (750 mg) were dissolved in deionized water (5 mL). Na_2SO_4 (10 µL, 50 mM) was added to this solution, with efficient magnetic stirring. *Procedure for sulfate assay:* Samples (typically 100 µl, containing a maximum of 2.5 µmol of Na_2SO_4) was diluted to 0.1 mL with deionized water. To this solution, 0.5 M HCl (0.1 mL) was added followed by Ba-PEG reagent (0.1 mL). The mixture was vigorously mixed and the absorbance of the sample was measured at 400 nm. Concentrations of sulfate were determined by reference to a standard curve constructed using sodium sulfate.

Colorimetric fuchsin sulfite assay

The sulfite assay was followed according to the procedures of Brychkova *et al*.[27] and Kurmanbayeva *et al*.[28]. This procedure requires three pre-prepared solutions, Reagents A, B and C. *Reagent A:* Basic fuchsin (4.0 mg, 0.012 mmol) was dissolved in deionized water (8.25 mL) at 0 °C. To this solution was added 98% H_2SO_4 (1.25 mL). *Reagent B:* Formaldehyde (36% in H_2O, 0.32 mL) was added to deionized water (9.68 mL) at 0°C. *Reagent C:* Reagent A (1 mL) was added to deionized water (7 mL); to this solution reagent B (1 mL) was added.

Procedure for sulfite assay: Reagent C (516 µL) was added to a mixture of sample (72 µL) and 0.5 mM Na_2SO_3 (12 µL), as a background. The sample was incubated at r.t. for 10 min and the absorbance of the sample was measured at 570 nm using a UV/vis spectrophotometer. Concentrations of sulfite were determined by reference to a standard curve constructed using sodium sulfite.

NMR analysis of metabolites produced from ($^{13}C_6$)SQ

M9 minimal media (5 mL) containing with 10 mM Glc was inoculated with *A. tumefaciens* C58 and grown to stationary phase at 30 °C (250 rpm). A 50 µl aliquot of this starter culture was used to inoculate 2 ml of M9 minimal media containing 10 mM ($^{13}C_6$)SQ and the culture incubated at 30 °C (250 rpm). Samples were collected at OD_{600} 0.27 and OD_{600} 0.49. 950 µl of the culture was diluted with 100 µl of D_2O and ^{13}C-NMR spectra were acquired.

Growth of A. tumefaciens *C58 on various alkanesulfonates*

M9 minimal media (5 mL) containing 10 mM Glc was inoculated with *A. tumefaciens* C58 and grown to stationary phase at 30 °C (250 rpm). A 50 µl aliquot of this starter culture was used to inoculate 2 ml of M9 minimal media containing 10 mM of the alternative alkanesulfonate substrate: SQ (positive control), methyl α-sulfoquinovoside, glycer-1-yl α-sulfoquinovoside, dicyclohexylammonium sulfolactate, cyclohexylammonium dihydroxypropanesulfonate, sulfoacetic acid, taurine, sodium pentanesulfonate, cysteic acid, MOPS, HEPES, PIPES, MES and methanesulfonic acid. Cultures were incubated for 30 days at 30 °C (250 rpm) with daily observations of optical density at 600 nm. Each experiment was performed in duplicate. Growth was observed on SQ (positive control), methyl α-sulfoquinovoside, and glycer-1-yl α-sulfoquinovoside, but not on any other sulfonate. Control experiments established that *A. tumefaciens* grows on Glc and on Glc+cyclohexylamine and does not grow on cyclohexylamine or dicyclohexylamine.

Digestion of samples for quantitative proteomics

Freeze dried *A. tumefaciens* whole-cell pellets were resuspend in 500 µl lysis buffer (4% SDS, 50 mM Tris pH 8.5, 10 mM DTT) and boiled at 95 ˚C for 10 min with shaking at 2000 rpm to shear DNA and inactivate protease activity. Lysates were cooled to room temperature, protein

concentration determined using a BCA assay. Each sample (200 µg of protein) was acetone precipitated by mixing 4 volumes of ice-cold acetone with one volume of sample. Samples were precipitated overnight at -20 °C and then centrifuged at 4000 × g for 10 min at 4 ˚C. The precipitated protein pellets were resuspended with 80% ice-cold acetone and precipitated for an additional 4 h at -20 °C. Samples were centrifuged at 17000 × g for 10 min at 4 ˚C to collect precipitated protein, the supernatant was discarded and excess acetone driven off at 65 °C for 5 min. Dried protein pellets were resuspended in 6 M urea, 2 M thiourea, 40 mM NH_4HCO_3 and reduced / alkylated prior to digestion with Lys-C (1/200 w/w) then trypsin (1/50 w/w) overnight as previously described[29]. Digested samples were acidified to a final concentration of 0.5% formic acid and desalted using C18 stage tips[30] before analysis by LC-MS.

Quantitative proteomics using reversed phase LC-MS

Purified peptides were resuspended in Buffer A* (2% MeCN, 0.1% TFA) and separated using a Proflow-equipped Dionex Ultimate 3000 Ultra-Performance Liquid Chromatography system (Thermo Fisher Scientific) with a two-column chromatography set up composed of a PepMap100 C18 20 mm × 75 µm trap and a PepMap C18 500 mm × 75 µm analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at 5 µL/min with Buffer A (2% MeCN, 0.1% FA) for 6 min and then infused into an Orbitrap Q-Exactive HF Mass Spectrometer (Thermo Fisher Scientific) at 250 nl/min. Peptides were separated using 124-min gradients altering the buffer composition from 2% Buffer B (80% MeCN, 0.1% FA) to 8% B over 14 min, then from 8% B to 30% B over 80 min, 30% B to 45% B over 10 min, 45% B to 95% B over 2 min, holding at 95% B for 10, then dropped to 2% B over 1 min and holding at 2% B for the remaining 7 min. The Q-Exactive HF™ Mass Spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (120,000 resolution) and a maximum of 20 MS-MS scans (HCD NCE 28, maximum fill time 40 ms, AGC 2×10^5 with a resolution of 15,000).

Mass spectrometry data analysis

Proteomics datasets were searched using MaxQuant (v1.5.3.3)[31] against the *A. tumefaciens* C58 proteome (Uniprot proteome id UP000000813, downloaded 27/01/2018, 5344 entries). Searches were performed with carbamidomethylation of cysteine set as a fixed modification and oxidation of methionine as well as acetylation of protein N-termini allowed as variable modifications. The protease specificity was set to trypsin allowing 2 miscleavage events with a maximum false discovery rate (FDR) of 1.0% set for protein and peptide identifications. To enhance the identification of peptides between samples the Match Between Runs option was enabled with a precursor match window set to 2 min and an alignment window of 10 min. For label-free quantitation, the MaxLFQ

option within Maxquant[32] was enabled in addition to the re-quantification module. The resulting protein group output was processed within the Perseus (v1.4.0.6)[33] analysis environment to remove reverse matches and common protein contaminates prior. For LFQ comparisons missing values were imputed using Perseus and Pearson correlations visualized using R. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[34] partner repository with the dataset identifier PXD014115.

Cloning

pET29 vector: Nucleotide sequences for *Atu3277* (*smoA*), *Atu3278* (*smoB*), *Atu3279* (*smoC*) and *Atu3282* (*smoF*) were amplified by PCR using Phusion polymerase HF master mix (NEB), the primers listed in **Supplementary Table 1** and *A. tumefaciens* C58 gDNA as template. The amplicons were cloned into the pET29b(+) vector at the *NdeI* and *XhoI* sites to give pET29-Atu3277, pET29-Atu3278, pET29-Atu3279 and pET29-Atu3282. All plasmids were sequence-verified using Sanger sequencing.

pET-YSBLIC3C vector: Atu3277, *Atu3278* and *Atu3279* were sub-cloned into the pET-YSBLIC3C vector[35] for crystallization using the In-Fusion® HD Cloning kit (Clontech Laboratories, Inc.). The gene of interest was amplified by PCR using relevant primers (Table S1) and purified using a QIAquick® Gel Extraction Kit. The In-Fusion reaction was performed using linearized YSBLIC3C vector and purified genes according to manufacturer's protocol. Insertion of target genes into the final constructs was verified by colony PCR and DNA sequencing.

Protein expression from pET29 vectors and purification

The pET29-Atu3277, pET29-Atu3278 and pET29-Atu3279 plasmids were transformed into 'T7 Express' *E. coli* (NEB), while the pET29-Atu3282 plasmid was transformed into 'Shuffle® T7 *E. coli* (NEB), and all were plated onto LB-agar (50 µg/ml kanamycin) and incubated at 37 °C for 16 h. A single colony was used to inoculate 10 ml of LB media containing 50 µg/ml kanamycin and the cultures incubated at 37 °C for 16 h. These cultures was used to inoculate 1000 ml of S-broth (35 g tryptone, 20 g yeast extract, 5 g NaCl, pH 7.4) containing 50 µg/ml kanamycin, which was incubated with shaking (250 rpm) at 37 °C until it reached an OD_{600} of 0.8. Each culture was cooled to room temperature, isopropyl thiogalactoside (IPTG) added to a final concentration of 100 µM, and then incubated with shaking (200 rpm) at 18 °C for 19 h. Cells were harvested by centrifugation at 8,000 g for 20 min at 4 °C. Each cell pellet was resuspended in 40 ml of binding buffer (50 mM NaP_i, 300 mM NaCl, 5 mM imidazole, pH 7.5) containing protease inhibitor (Roche cOmplete EDTA-free protease inhibitor cocktail) and lysozyme (0.1 mg/ml) by nutating at 4 °C for 30 min. Benzonase (1

µl) was added to the mixture then lysis was effected by sonication [10× (15 s on / 45 s off) at 45% amplitude]. The lysate was centrifuged at 18,000 g for 20 min at 4 °C and the supernatant collected. The supernatants were filtered (0.45 µm) and loaded onto a 1 ml HiTrap TALON IMAC column (GE). The column was washed with 3 × 10 ml of binding buffer, then the protein was eluted using elution buffer (50 mM NaP_i, 300 mM NaCl, 400 mM imidazole, pH 7.5). Fractions containing product, as judged by SDS-PAGE, were further purified by size exclusion chromatography on a HiPrep 16/60 Sephacryl S-200 HR column (GE) using 50 mM NaP_i, 150 mM NaCl, pH 7.5 (Atu3277 SmoA; Atu3278, SmoB; Atu3279, SmoC) or 50 mM sodium citrate, 150 mM NaCl, pH 5.5 (Atu3282, SmoF) as buffer. Atu3285 (SmoI, *At*SQase) was prepared as previously described[8].

Protein expression from pET-YSBLIC3C vectors and purification

The plasmid containing the gene for target enzyme was used to transform competent *E. coli* BL21(DE3) cells for expression. Initial cultures were grown in LB-medium (5 mL) containing 30 µg mL^{-1} kanamycin for 18 h at 37 °C with shaking at 220 r.p.m. 1 L expression cultures were inoculated with the initial culture (5 mL) and incubated at 37 °C, with shaking at 200 r.p.m. until an OD_{600} of 0.6-0.8 was reached. Gene expression was induced by addition of IPTG (0.5-1 mM) and shaking was continued overnight at 18 °C at 220 r.p.m. Cells were harvested by centrifugation at 5000 g for 20 min and resuspended in 50 mM NaPi buffer pH 7.4, containing 500 mM NaCl and 30 mM imidazole. Cells were disrupted by ultrasonication for 3 × 5 min, 30 s on, 30 s off cycles, and the suspension was centrifuged at 50,000 g for 30 min to yield a clear lysate. The C-terminal His_6-tagged protein was purified using immobilised-metal affinity chromatography (IMAC) using Ni-NTA column, followed by size exclusion chromatography (SEC) (**Supplementary Figure 2**). For IMAC, the lysate was loaded onto a pre-equilibrated Ni-NTA column, followed by washing with a load buffer (50 mM NaPi, 500 mM NaCl, 30 mM imidazole pH 7.4). The bound protein was eluted using a linear gradient with buffer containing up to 500 mM imidazole. Protein fractions were pooled, concentrated and loaded onto a HiLoad 16/600 Superdex 200 gel filtration column pre-equilibrated with 50 mM NaPi, 300 mM NaCl pH 7.4 buffer. The protein was concentrated to a final concentration of 65 mg mL^{-1} using a Vivaspin® 6 with a 300 kDa MW cut-off membrane.

Size exclusion chromatography – multi angle laser scattering analysis

Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu LC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Experiments were conducted at room temperature (20 ± 2°C). Solvents were filtered through a 0.2 µm filter prior to use and a 0.1 µm filter was present in the flow path. The column was equilibrated

with at least 2 column volumes of buffer (50 mM NaPi, 300 mM NaCl pH 7.4) before use and buffer was infused at the working flow rate until baselines for UV, light scattering and refractive index detectors were all stable. The sample injection volume was 100 µL SmoB at 6 mg/mL in 50 mM NaPi buffer, 300 mM NaCl pH 7.4. Shimadzu LC Solutions software was used to control the LC and Astra V software for the HELEOS-II and rEX detectors. The Astra data collection was 1 minute shorter than the LC solutions run to maintain synchronisation. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data were analysed using the Astra V software. MWs were estimated using the Zimm fit method with degree 1. A value of 0.158 was used for protein refractive index increment (dn/dc).

Isothermal Titration Calorimetry

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 µCal.s^{-1}. Proteins and substrates were equilibrated into degassed and filter sterilised buffer (50 mM NaPi, 200 mM NaCl, pH 7.4). For SmoC–SQ binding, 600 µM of SQ was titrated into the ITC cell containing 40 µM SmoC as a series of 10 × 3.94 µl injections with a pre-injection of 1 × 0.4 µl. For SmoF–SQGro binding, 200 µM of SQGro was titrated into the ITC cell containing 20 µM SmoF as a series of 15 × 2.94 µl injections with a pre-injection of 1 × 0.4 µl. The delay between injections was set at 120 s, with an initial injection delay of 60 s. For SmoB-NAD(P)H binding, protein and cofactors were equilibrated into degassed and filter sterilised buffer containing 25mM NaPi at pH 7.5. NADH at 1 mM was injected into an ITC cell containing 40 µM SmoB as a series of 19 × 3 µl injections with a pre-injection of 1 × 4 µl. Delay between injections were set as 150 s with an initial injection delay of 180 s. SmoB concentration was confirmed by BCA protein concentration assay (Thermo Fisher) prior to all runs. All data analysis was performed in MicroCal ITC Origin Analysis software (Malvern).

Nano Differential Scanning Fluorescence of SmoF

Thermal stability analysis by was performed on a Prometheus NT.48 (NanoTemper), at 15% excitation, scanning from 20 °C to 65 °C at 0.5 °C min^{-1}. All protein samples were at 1 mg ml^{-1} in 50 mM citrate, 150 mM NaCl at pH 5.5, with a 10 µl capillary load per sample. Data recording and initial analysis was performed with PR.ThermControl (NanoTemper) software.

Identification of SmoA flavin co-factor

100 µl of recombinant flavin reductase (SmoA, *Ro*SmoA) at a concentration of 20 mg/ml in 50 mM Tris, 150 mM NaCl pH 8.5 was heated at 90 °C for 10 min. The sample was clarified by centrifugation (16,000×g, 10 min, 4 °C) and the supernatant filtered (0.2 µm). Samples were analysed by LCMS on

an Agilent LCMS system (G6125B mass detector, 1290 Infinity G7120A high speed pump, 1290 Infinity G7129B autosampler, and 1290 Infinity G7117B diode array detector). Conditions for LC were as follows: column: Phenomenex 00B-4752-AN Luna Omega 1.6 µm PS C_{18} 100Å (50 × 2.1 mm); injection volume: 1 µl; gradient: 3 to 100% B over 20 min (solvent A: water + 0.1% formic acid; solvent B: MeCN + 0.1% formic acid); flow rate: 0.6 ml/min; DAD – 254 and 214 nm.

Michaelis-Menten kinetic analysis of SmoC and RoSmoC

Reactions were conducted at 25 °C in 96-well plate format and involved the addition of SmoC or *Ro*SmoC (final concentration of 20 nM for NADH and 500 nM for NADPH) to 20–800 µM NAD(P)H in 50 mM NaPi, 150 mM NaCl, 30 µM FMN, 0.01% BSA, pH 7.4 at a total volume of 100 µl. The progress of the enzyme-catalysed conversion of NAD(P)H to $NAD(P)^+$ was monitored by measuring loss of absorbance at 340 nM over time using an Envision Multimodal Plate Reader (GE Healthcare). Initial rates for each reaction were calculated after first subtracting the rate of spontaneous NAD(P)H oxidation (determined using an enzyme-free control) and an empirically determined extinction coefficient for NAD(P)H under these conditions. Each initial rate was determined in triplicate and fit to a Michaelis-Menten equation using Prism 8 (GraphPad).

Sulfoquinovose monooxygenase assay

The SQ monooxygenase activity assay is based on previously described alkanesulfonate monooxygenase activity assays[17] and uses Ellman's reagent to quantify sulfite release. A 2 ml reaction containing 1 mM SQ, 1 mM NADH, 3 µM FMN, 0.01% (w/v) BSA, 100 nM SmoA or *Ro*SmoA and 300 nM SQ monooxygenase in buffer (25 mM Tris pH 9.1, 25 mM NaCl) was incubated at 30°C, along with controls lacking reaction components or using alternate sulfonate substrates. Reactions were initiated by the addition of SmoA or *Ro*SmoA to the mixture. Sulfite concentration in the samples was determined at discrete time points by quenching 40 µl of the reaction in 160 µl of Ellman's reagent (0.125 $mg.ml^{-1}$ in 25 mM NaPi pH 7.0, prepared fresh) within a 96-well plate. After 60 s, the absorbance of the sample at 405 nm was determined using an Envision (Perkin Elmer) plate reader. The sulfite concentration was interpolated using a calibration curve generated under these conditions: a linear relationship between sulfite concentration and absorbance at 405 nm was observed for 5–1000 µM Na_2SO_3. The activity of SQ monooxygenase at different pH was determined by modifying the buffer in the above reactions (MES: pH 6.0, 6.5 and Tris: pH 7.0, 7.5, 8.0, 8.5, 9.1) using an endpoint of t = 30 min.

Equilibrium isotope labelling using SmoB

In order to pre-label the anomeric position, D-glucose was incubated with 98% ^{18}O water by heating

at 80°C for 2 days, then evaporated to dryness, to give C1-^{18}O-labelled glucose with labelling determined to be 95%, by analysis of the M and M+2 peaks by mass spectrometry. Using ^{18}O-water buffer (100 mM potassium phosphate, pH 7), NAD^+ and $NADP^+$ were each added at 0.05 equivalent to C1-^{18}O-glucose and SmoB. Four control experiments were conducted: (2) no enzyme, (3) no NAD^+ and $NADP^+$, (4) ^{16}O-water, and (5) ^{16}O-water + unlabelled glucose. Reactions were monitored by mass spectrometry. Only in the experimental sample containing enzyme, ^{18}O-water and NAD^+/$NADP^+$ was an M+4 signal observed that reached a maximum intensity after 72 hours. Subsequently, two additional reactions were performed using SmoB, D-glucose and either $NADP^+$ or NAD^+ in ^{18}O-water. Only the reaction containing $NADP^+$ generated the M+4 signal.

To confirm that the M+4 signal observed in the mass spectra is D-glucose with two ^{18}O labels, we studied the product by HPLC. However, under aqueous HPLC conditions the ^{18}O-label at C1 is lost through chemical exchange with solvent. Therefore, we acetylated the product to form the pentaacetate to ensure no exchange during HPLC analysis. The reaction mixture from above was evaporated under reduced pressure and dried. The crude residue was treated with acetic anhydride in pyridine (1:2, 1 ml) overnight. The product was extracted with EtOAc and washed with sat. $CuSO_4$ to give a solution of acetylated products. These were analysed by LC-MS with a C18 column from 100% water to acetonitrile/water 65:35. Peaks with *m*/*z* 413 $[M+Na]^+$, *m*/*z* 415 $[M+2+Na]^+$, and *m*/*z* 417 $[M+4+Na]^+$ had the same retention time as an authentic glucose pentaacetate standard.

GC-MS analysis of isotopically-labelled carbohydrates

A 0.1 uL aliquot (containing approx 2.5 nmol glucose) was transferred to GC vial inserts (deactivated) in addition to 1 nmol *scyllo*-inositol as an internal standard. Samples were derivatised as described in Antoniewicz *et al.*[18], with minor modifications. Briefly, samples were dried (*in vacuo*, 35 °C, 40 uL methanol wash), followed by addition of hydroxylamine hydrochloride (Sigma, 20 mg/ml, 25 uL pyridine) with incubation (90 °C, 1 hr). Vials were cooled briefly at room temperature followed by the addition of propionic anhydride (50 uL, 60 °C, 30 min). Samples were evaporated to dryness under a stream of nitrogen (60 °C) and resuspended in ethyl acetate (40 uL). Control samples of U-^{12}C-glucose, U-^{13}C-glucose, 1,2-$^{13}C_2$-glucose and 6,6-$^{2}H_2$-glucose were also prepared at 2.5 nmol in the assay buffer mixture. Samples were randomised for analysis.

Carbohydrate sample analysis

Labelled glucose (**Supplementary Figure 12** and **Supplementary Table 7**) was detected by GC-MS using a DB5 capillary column (J&W Scientific, 30 m, 250 μm inner diameter, 0.25-μm film thickness), with a 10-m inert duraguard. The injector insert and GC-MS transfer line temperatures

were 270 and 250 °C, respectively. The oven temperature gradient was programmed as follows: 70 °C (1 min); 70 °C to 295 °C at 12.5 °C/min; 295 °C to 320 °C at 25 °C/min; 320 °C for 2 min. D-Glucose and *scyllo*-inositol were identified by reference to authentic standards. A calibration curve was generated using D-glucose standard in assay buffer (starting concentration 50 nmol, 2-fold dilution series). **Supplementary Figure 12** shows fraction of labelled fragments, corrected for isotope natural abundance by DExSI analysis[36].

Protein crystallization

Initial crystallization screening was performed using commercially available INDEX (Hampton Research), PACT premier and CSSI/II (Molecular Dimensions) screens in 96-well sitting drop trays. Further optimization was carried out in a 48-well sitting drop or 24-well hanging-drop format to obtain optimal crystals for X-ray diffraction. For co-crystallization experiments, 0.2 M stock solution of cofactor NADPH and 0.5 M glucose stock solution was prepared in water. All crystals were grown at 20 °C, unless otherwise mentioned.

For the SmoI-D455N•SQGro structure, crystals were grown from 35 mg mL^{-1} enzyme in 50 mM NaPi buffer pH 7.4 containing 300 mM NaCl in a drop with 0.4 µL protein: 0.5 µL mother liquor, with the reservoir solution containing 26% PEG 3350 w/v, 0.2 M KSCN, 0.1 M Bis-Tris propane pH 6.5 mother liquor. The crystal was soaked with solid SQGro in the mother liquor for 2 min prior to fishing.

A crystal of SmoF-apo was grown using a 50 mg mL^{-1} protein solution in 50 mM citrate, 150 mM NaCl at pH 5.5 in a drop containing 0.15 µL protein: 0.15 µL mother liquor, the latter comprising 0.3 M ammonium acetate, 0.1 M Bis-Tris, 25% w/v PEG 3350 at pH 5.5; housed in a Rigaku Xtaltrak plate hotel to enable consistent growth and monitoring at 6 °C.

Initial crystals of SmoF was grown over several days in sitting drops at 20°C by mixing 150 nL well solution containing 30% (w/v) polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M tris chloride, pH 8.5, with 150 nL SmoF at 3.5 mg/mL with 2'*R*-SQGro at a 1:10 molar ratio. The resulting crystals were used to prepare a seed stock by mixing the crystallization drop with 100 uL reservoir solution and vortexing it for 60 sec with one teflon seed bead. An optimisation plate was setup by mixing 100 nL well solution containing 28-36% (w/v) polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M tris chloride, pH 7.1-9.1, with 50 nL seed stock solution, and 150 nL SmoF at 4 mg/mL with 2'*R*-SQGro at a 1:10 molar ratio. A single crystal grown at 31.8% (w/v) polyethylene glycol 4000, 0.2 M

sodium acetate, 0.1 M tris chloride, pH 8.95, was flash frozen in mother liquor supplemented with 25% (v/v) ethylene glycol using liquid nitrogen.

For SQ monooxygenases, a crystal of apo-SmoC was grown using a 60 mg mL^{-1} in 50 mM Tris pH 7.5, 300 mM NaCl in a drop containing 0.6 µL protein: 0.5 µL mother liquor, the latter containing 0.2 M NaCl, 0.1 M MES pH 6, 26% PEG 6000 w/v and 10 mM SQ-glucitol. A crystal of apo-*Ro*SmoC was grown using a 11.7 mg mL^{-1} in 50 mM Tris pH 7.5, 300 mM NaCl in a drop containing 0.1 µL protein: 0.2 µL mother liquor, the latter containing 0.2M $NaNO_3$, 20% PEG 3350 w/v and 10 mM SQ.

A crystal of SmoB-apo (YSBLIC3C construct) was grown using a 20 mg mL^{-1} protein solution in 50 mM NaPi buffer pH 7.4, 150 mM NaCl in a drop containing 0.15 µL protein: 0.15 µL mother liquor, the latter comprising 0.2 M sodium malonate dibasic monohydrate, 0.1 M Bis-Tris propane pH 8.5, 20% w/v PEG 3350. For the SmoB•NADPH structure, crystals were grown with SmoB at 20 mg mL^{-1} in 50 mM NaPi buffer pH 7.4, 150 mM NaCl containing 2 mM NADPH using a drop containing 0.15 µL protein: 0.15 µL mother liquor, the latter comprising 0.1 M succinic acid, sodium dihydrogen phosphate, and glycine buffer (SPG buffer, Qiagen), 25% w/v PEG 1500 at pH 6.0. For the SmoB•NADPH•Glc structure, crystals were grown from 13 mg mL^{-1} enzyme in 50 mM NaPi buffer pH 7.4 containing 150 mM NaCl in a hanging drop with 1 µL protein: 1 µL mother liquor, with the reservoir solution containing 2 mM NADPH, in 0.1 M SPG (Qiagen), 25% w/v PEG 1500 at pH 6. The crystal was then soaked with solid glucose in the mother liquor for 1 min prior to fishing. The crystals were harvested into liquid nitrogen, using nylon CryoLoops™ (Hampton Research) *via* mother liquor without any cryoprotectant.

X-ray data collection, processing and refinement

The data were processed and integrated using XDS[37] and scaled using SCALA[38] included in the Xia2 processing system[39]. Data reduction was performed with AIMLESS, and resolution was cut until CC1/2 = 0.5. The structure of the SmoI•SQGro complex was determined using molecular replacement using 5OHS[8] as the initial model. For SmoF, the structure was solved by molecular replacement using PHASER[40] using a search model created using PDB ID: 6DTQ[41]. The structure of *Ro*SmoC was solved using a molecular replacement, using the ensemble based on 1M41[17] as initial search model. Molecular replacement of SmoB was achieved using the monomer of an aldo-keto reductase from *S. enterica* (PDB = 4R9O.pdb) as initial MR model. The apo SmoF structure was solved using a dissected C-terminal domain of the SmoF•SQGro structure.

Structures were built and refined using iterative cycles using Coot[42] and REFMAC[43] or Phenix,[44] the latter employing local NCS restraints. Following building and refinement of the protein and water molecules, clear residual density was observed in the omit maps for co-complex structures, respective ligands were modelled into these. The coordinate and refinement library files were prepared using ACEDRG[45]. The final structures gave R_{cryst} and R_{free} values along with data and refinement statistics that are presented in **Supplementary Table S4-6.** Data were collected at Diamond light source, Didcot, Oxfordshire, U.K., on beamlines I24 (SmoI-D455N•SQGro, to 2.15 Å; SmoF-apo, to 1.88 Å), I04 (*Ro*SmoC to 1.75 Å) and I04-1 (SmoC-apo, to 3.2 Å; SmoB-apo_YSBLIC3C, to 1.5 Å; SmoB-apo; pET29a; SmoB•NADPH and SmoB•NADPH•Glc) and using MX2 beamline (At3282•SQGro complex, to 1.7 Å) at the Australian Synchrotron. The coordinate files and structure factors have been deposited in the Protein DataBank (PDB) with the coordinate accession numbers **7OFX** (SmoI-D455N•SQGro), **7NBZ** (SmoF-apo), **7OFY** (SmoF•SQGro)**, 7OH2** (*Ro*SmoC), **7OLF** (SmoC-apo), **7BBY** (SmoB-apo; pET29a), **7BBZ** (SmoB-apo; YSBLIC3C), **7BC0** (SmoB•NADPH) and **7BC1** (SmoB•NADPH•Glc).

Structure-based analyses

Crystal packing interactions were analysed using the protein interactions, surfaces, and assemblies (PISA) server.[46] Structural comparisons and structure-based sequence alignments were conducted using PDB25 search on DALI server against a representative subset of the Protein Data Bank[47]. All structure figures were generated using ccp4mg[48].

Determining the prevalence of the sulfo-SMO pathway

Each gene within the *A. tumefaciens* C58 sulfo-SMO gene cluster (*Atu3277-Atu3285*) was submitted as a query to the NCBI BLASTp algorithm to search a database comprised of non-redundant protein sequences with *A. tumefaciens* (taxid: 358) sequences excluded. Standard algorithm parameters were used, except the maximum target sequences was set to 10,000. Results were filtered to only retain protein sequences with E-value $\leq 1.19\times10^{-51}$. The corresponding nucleotide accession numbers for each protein from all nine searches were extracted, combined and duplicates removed to provide a list of candidate genome sequences. This was converted into a reference library for MultiGeneBLAST[49] and queried using the *A. tumefaciens* C58 sulfo-SMO gene cluster. Clusters identified by this workflow with both an SQ monoxygenase and SQase homolog were regarded as putative sulfo-SMO gene clusters. Clusters representative of the observed diversity were visualized using Clinker[50]. A phylogenetic tree of species possessing a putative sulfo-SMO gene cluster was generated by pruning the All-Species Living Tree Project's 16s rRNA release 132[51] using iTOL[52].

Data Availability Statement

All relevant data are available from the authors upon request. Structure coordinates have been deposited in the Protein Data Bank (https://www.rcsb.org/) under accession codes **7OFX**, **7OFY**, **7NBZ**, **7OH2**, **7OLF**, **7BBZ**, **7BC0**, **7BC1** and **7BBY**. Proteomics data are available via ProteomeXchange[53] (http://www.proteomexchange.org/) with the identifier PXD014115.

References

1. Goddard-Borger, E.D. & Williams, S.J. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem. J.* **474**, 827–849 (2017).

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

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

4. Felux, A.K., Spiteller, D., Klebensberger, J. & Schleheck, D. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. USA* **112**, E4298-4305 (2015).

5. Li, J. *et al.* A Sulfoglycolytic Entner-Doudoroff Pathway in *Rhizobium leguminosarum* bv. *trifolii* SRDI565. *Appl. Environ. Microbiol.* **86**, e00750-00720 (2020).

6. Frommeyer, B. *et al.* Environmental and Intestinal Phylum Firmicutes Bacteria Metabolize the Plant Sugar Sulfoquinovose via a 6-Deoxy-6-sulfofructose Transaldolase Pathway. *iScience* **23**, 101510 (2020).

7. Liu, Y. *et al.* A transaldolase-dependent sulfoglycolysis pathway in *Bacillus megaterium* DSM 1804. *Biochem. Biophys. Res. Commun.* **533**, 1109-1114 (2020).

8. Abayakoon, P. *et al.* Structural and Biochemical Insights into the Function and Evolution of Sulfoquinovosidases. *ACS Cent. Sci.* **4**, 1266-1273 (2018).

9. Speciale, G., Jin, Y., Davies, G.J., Williams, S.J. & Goddard-Borger, E.D. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **12**, 215-217 (2016).

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

11. Kertesz, M.A. Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiol. Rev.* **24**, 135-175 (2000).

12. Abayakoon, P. *et al.* Comprehensive synthesis of substrates, intermediates and products of the sulfoglycolytic Embden-Meyerhoff-Parnas pathway. *J. Org. Chem.* **84**, 2910-2910 (2019).

13. Speers, A.E. & Wu, C.C. Proteomics of Integral Membrane Proteins: Theory and Application. *Chem. Rev.* **107**, 3687-3714 (2007).

14. Davidson, A.L., Dassa, E., Orelle, C. & Chen, J. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiol. Mol. Biol. Rev.* **72**, 317-364 (2008).

15. van Der Ploeg, J.R., Iwanicka-Nowicka, R., Bykowski, T., Hryniewicz, M.M. & Leisinger, T. The *Escherichia coli* ssuEADCB gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. *J. Biol. Chem.* **274**, 29358-29365 (1999).

16. Li, L. *et al.* Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. *J. Mol. Biol.* **376**, 453-465 (2008).

17. Eichhorn, E., Davey, C.A., Sargent, D.F., Leisinger, T. & Richmond, T.J. Crystal Structure of *Escherichia coli* Alkanesulfonate Monooxygenase SsuD. *J. Mol. Biol.* **324**, 457-468 (2002).

18. Antoniewicz, M.R., Kelleher, J.K. & Stephanopoulos, G. Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal. Chem.* **83**, 3211-3216 (2011).

19. Penning, T.M. The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234**, 236-246 (2015).

20. Dippel, R. & Boos, W. The Maltodextrin System of *Escherichia coli*: Metabolism and Transport. *J. Bacteriol.* **187**, 8322 (2005).

21. Sharma, M. *et al.* Dynamic Structural Changes Accompany the Production of Dihydroxypropanesulfonate by Sulfolactaldehyde Reductase. *ACS Catalysis* **10**, 2826-2836 (2020).

22. Sharma, M. *et al.* Molecular Basis of Sulfosugar Selectivity in Sulfoglycolysis. *ACS Cent. Sci.* **7**, 476-487 (2021).

23. Udvardi, M. & Poole, P.S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781-805 (2013).

24. Speck, J.J., James, E.K., Sugawara, M., Sadowsky, M.J. & Gyaneshwar, P. An Alkane Sulfonate Monooxygenase Is Required for Symbiotic Nitrogen Fixation by *Bradyrhizobium diazoefficiens* (syn. *Bradyrhizobium japonicum*) USDA110(T). *Appl. Environ. Microbiol.* **85** (2019).

25. Blakeney, A.B. & Mutton, L.L. A simple colorimetric method for the determination of sugars in fruit and vegetables. *J. Sci. Food Agric.* **31**, 889-897 (1980).

26. Sörbo, B. Sulfate: Turbidimetric and nephelometric methods. *Methods Enzymol.*, vol. 143. Academic Press, 1987, pp 3-6.

27. Brychkova, G., Yarmolinsky, D., Fluhr, R. & Sagi, M. The determination of sulfite levels and its oxidation in plant leaves. *Plant Sci.* **190**, 123-130 (2012).

28. Kurmanbayeva, A. *et al.* Determination of Total Sulfur, Sulfate, Sulfite, Thiosulfate, and Sulfolipids in Plants. In: Sunkar, R. (ed). *Plant Stress Tolerance: Methods and Protocols*. Springer New York: New York, NY, 2017, pp 253-271.

29. Scott, N.E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni. *Mol. Cell. Proteomics* **10**, M000031-mcp000201 (2011).

30. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896-1906 (2007).

31. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

32. Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513 (2014).

33. Tyanova, S. *et al.* Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453-1456 (2015).

34. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442-d450 (2019).

35. Fogg, Mark J. & Wilkinson, Anthony J. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.* **36**, 771-775 (2008).

36. Dagley, M.J. & McConville, M.J. DExSI: a new tool for the rapid quantitation of ^{13}C-labelled metabolites detected by GC-MS. *Bioinformatics* **34**, 1957-1958 (2018).

37. Kabsch, W. Xds. *Acta Crystallogr., Section D: Biol. Crystallogr.* **66**, 125-132 (2010).

38. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D* **62**, 72-82 (2006).

39. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186-190 (2010).

40. Storoni, L.C., McCoy, A.J. & Read, R.J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432-438 (2004).

41. Shukla, S. *et al.* Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876 (2018).

42. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126-2132 (2004).

43. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D* **53**, 240-255 (1997).

44. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).

45. Long, F. *et al.* AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. Sect. D* **73**, 112-122 (2017).

46. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256-2268 (2004).

47. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-W549 (2010).

48. McNicholas, S., Potterton, E., Wilson, K.S. & Noble, M.E.M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **67**, 386-394 (2011).

49. Medema, M.H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218-1223 (2013).

50. Gilchrist, C.L.M. & Chooi, Y.H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* (2021).

51. Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucl. Acids Res.* **42**, D643-D648 (2013).

52. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* **47**, W256-W259 (2019).

53. Deutsch, E.W. *et al.* The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucl. Acids Res.* **45**, D1100-D1106 (2016).

Acknowledgements

Dr Monica Doblin is thanked for the provision of *Agrobacterium tumefaciens* strain C58. This work was supported in part by National Health and Medical Research Council of Australia (NHMRC) project grants GNT1100164 (N.E.S), GNT1174405 (D.B.A.), GNT1139546 and GNT1139549 (E.D.G.-B); the Leverhulme Trust grant RPG-2017-190 (G.J.D.); Australian Research Council grant DP180101957 and DP210100233 (S.J.W.), and DP210100362 (N.E.S.); and 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 (E.D.G.-B). G.J.D is supported by the Royal Society Ken Murray Research Professorship, E.D.G.-B. is supported by the Brian M. Davis Charitable Foundation Centenary Fellowship, M.J.M. is an NHMRC Principal Research Fellow, N.E.S. is supported by and Australian Research Council Future Fellowship (FT200100270), B.M. was supported by Melbourne Research Scholarship, J.M. by a Sir John and Lady Higgins Scholarship, M.P. by an Australian Postgraduate Award. We acknowledge Dr. Johan P. Turkenburg and Sam Hart for assistance with X-ray data collection; the Diamond Light Source for access to beamlines I04, i24 and I04-1 under proposal number mx-18598; and the Australian Synchrotron, part of ANSTO, for access to the MX-2 beamline, which made use of the Australian Cancer Research Foundation (ACRF) detector. We thank the 'Melbourne Mass Spectrometry and Proteomics Facility' of the Bio21 Molecular Science and Biotechnology Institute at The University of Melbourne for the support of mass spectrometry analysis and the 'Bioscience Technology Facility' (University of York) for assistance with SEC-MALS analyses.

Author Contributions

EDG-B discovered sulfo-SMO operon; SJW, EDG-B, GJD conceived project; MP and JW-YM conducted microbial growth experiments; NES conducted proteomics; JPL, MS, AS, MJ performed molecular biology, protein expression and structural and biophysical characterization; YZ, JPL, AS, MS, RM performed biochemical assays; ECS and MJM conducted carbohydrate analysis; YZ, JW-YM, BM and DA performed bioinformatics analysis; SJW, MS, EGB wrote the paper with input from all authors.

Competing Financial Interests Statement

The authors declare no competing interests.

Additional information

Supplementary information

Correspondence and requests for materials should be addressed to S.J.W, G.J.D or E.D.G.-B.

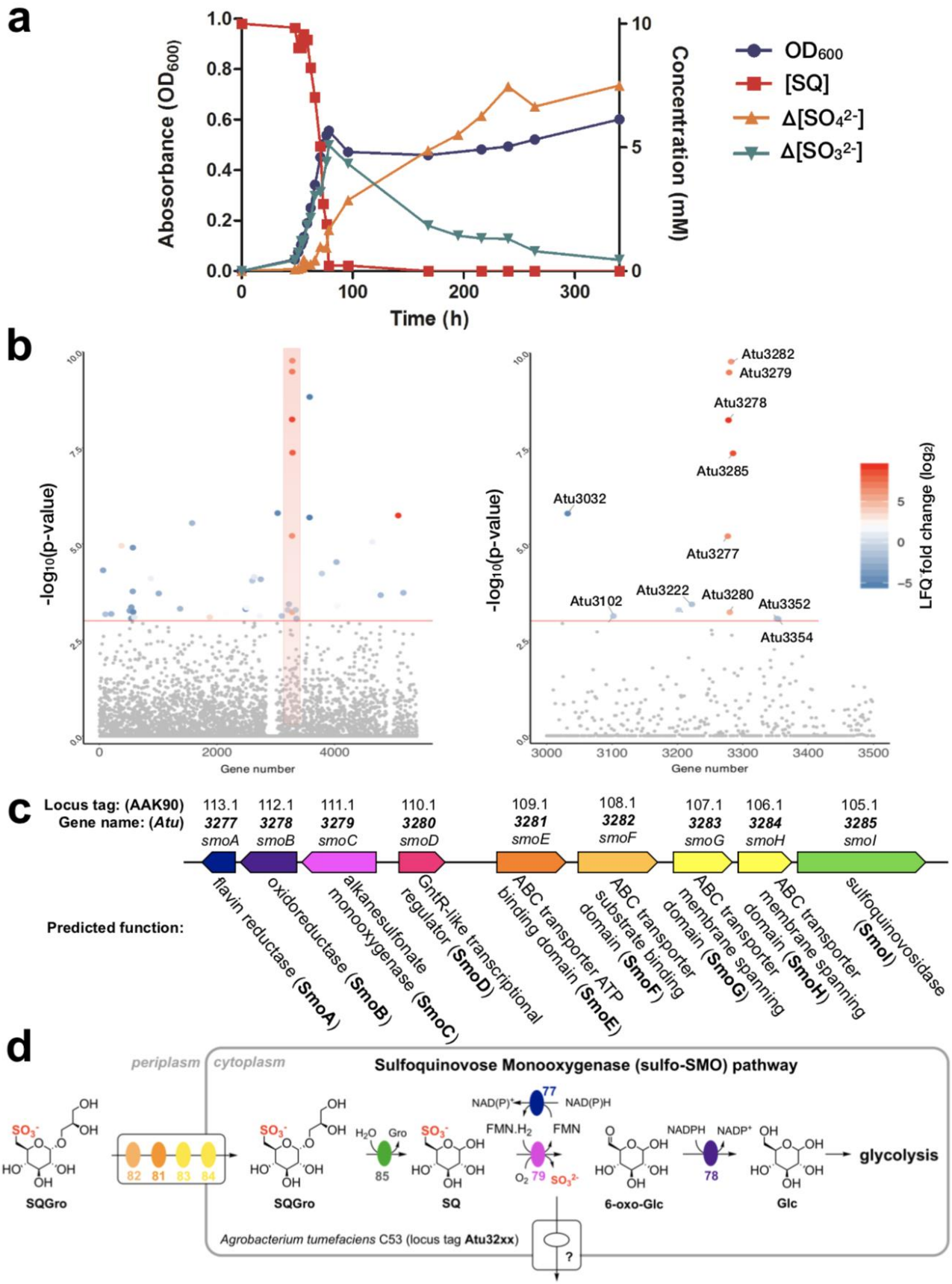


Figure 1: *A. tumefaciens* utilizes SQ glycosides as a carbon source. (a) Optical density of *A. tumefaciens* C58 culture (blue) and [SQ] (red), change in [sulfite] (green) and change in [sulfate] (yellow), with respect to time. (b) Manhattan plot of comparative proteomics data for *A. tumefaciens*

grown on SQ vs Glc demonstrating that the upregulated proteins belong to a single gene cluster. (c) Cartoon of the upregulated cluster with annotations for each of the gene products. These genes were termed *smoABCDEFGHI* (sulfoquinovose monooxygenase pathway gene cluster). (d) Proposed sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway for the metabolism of SQ in *A. tumefaciens*.

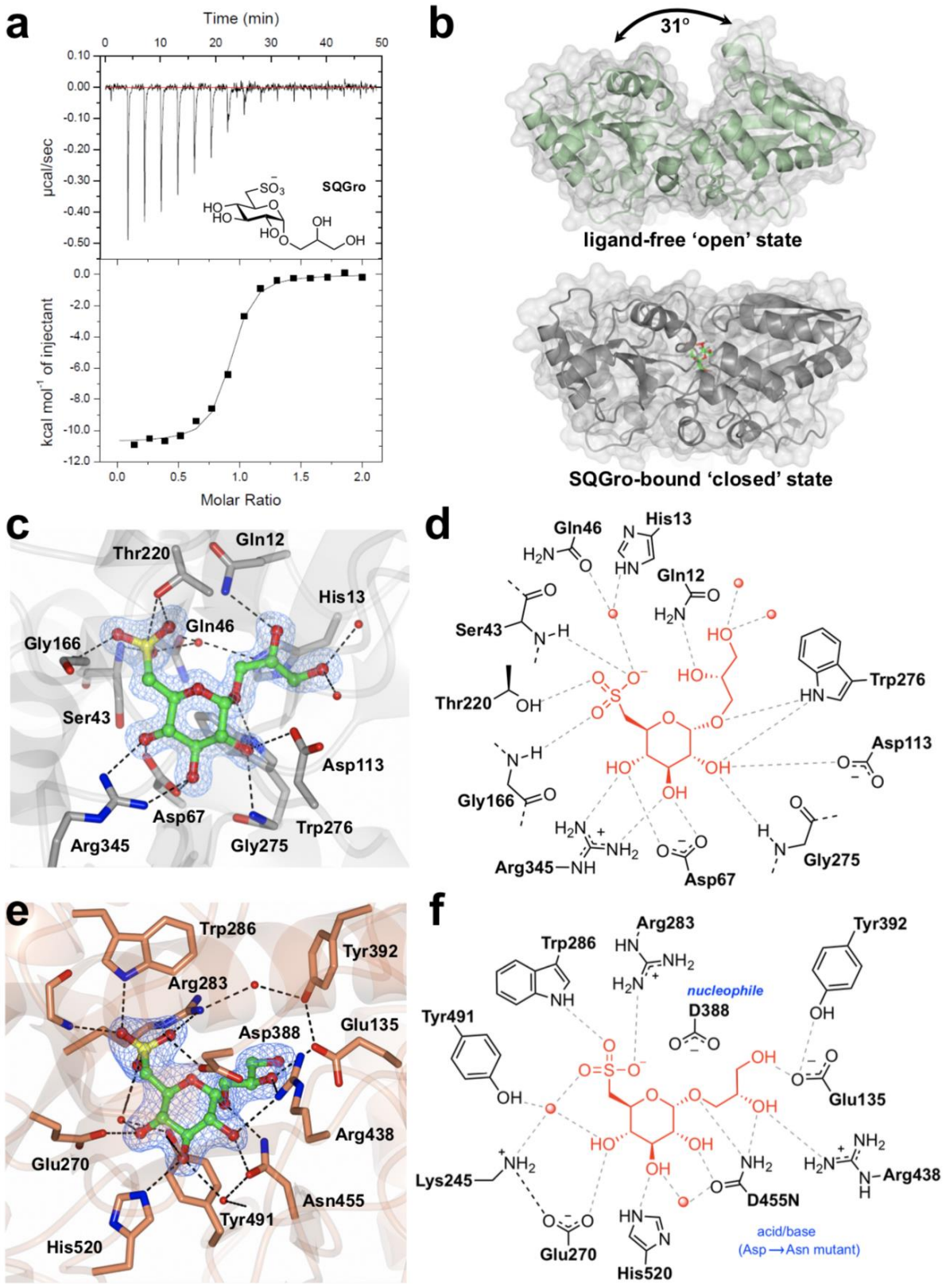


Figure 2: Biochemical and structural analyses of the SQGro-binding protein SmoF (Atu3282) and SQase SmoI (Atu3285). (a) Thermodynamic parameters for interaction of binding protein SmoF with its cognate ligand 2'*R*-SQGro determined by ITC. (b) Transparent molecular surface and ribbon diagram of open and closed conformations of SmoF. 2'*R*-SQGro is bound tightly in the inter-domain cleft and is inaccessible to the bulk solvent in the closed conformation. (c) SmoF•2'*R*-SQGro complex

showing detailed interactions of the sulfonate ligand bound in the cleft. Backbone and side-chains are shown in grey and SQGro depicted in cylinder format. Electron density corresponds to the 2Fo − Fc map (in blue) at levels of 1.5σ. (d) Cartoon highlighting key interactions from c. (e) Detailed view of active site interactions of complex of SmoI-D455N SQase with 2'*R*-SQGro. Backbone and carbon atoms of SmoI are shown in gold and 2'*R*-SQGro is shown in cylinder format. Electron density for 2'*R*-SQGro corresponds to the 2Fo − Fc and in blue at levels of 1.5σ. (f) Cartoon highlighting key interactions from e. Red spheres are bound water molecules; dotted lines are proposed hydrogen bonds.

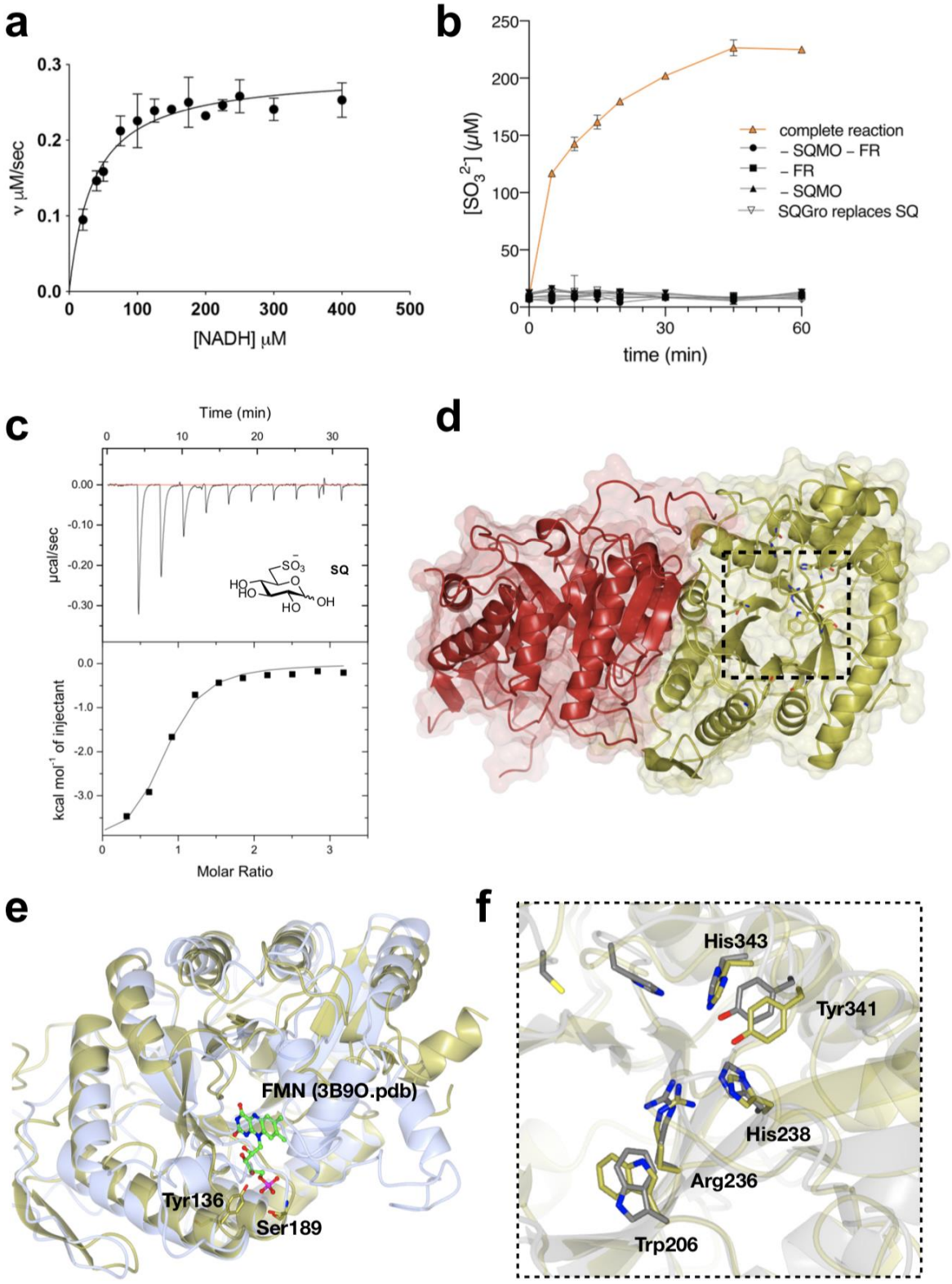
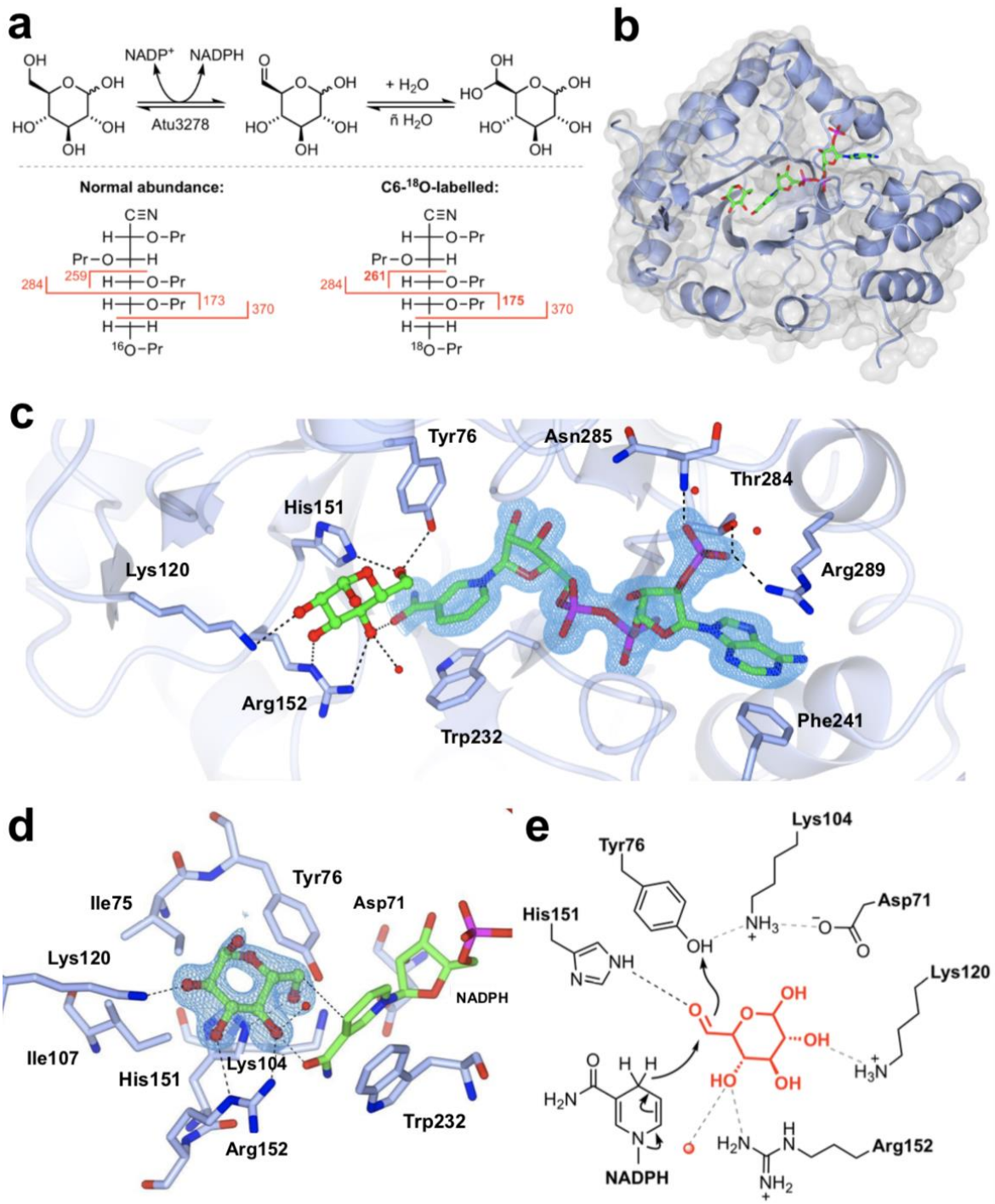


Figure 3: Biochemical and structural analyses of the flavin reductase SmoA and SQ monooxygenase SmoC. (a) Michaelis-Menten kinetics for SmoA-catalysed reduction of FMN by NADH (b) SmoC activity assessed using sulfite release assay with Ellman's reagent in the presence of FMN, flavin reductase, NADH and SQ. (c) Isothermal titration calorimogram of interaction of SmoC with SQ as determined by ITC. (d) Transparent molecular surface and ribbon diagram of *Ro*SmoC homodimer showing cofactor binding pocket and active site. (e) Overlay of *Ro*SmoC (in

gold) and LadA·FMN complex (3B9O.pdb in ice blue) showing location of FMN pocket. (f) Overlay of *Ro*SmoC (in gold) and SsuD (1M41.pdb in grey) showing detailed view of proposed substrate-binding pocket and conserved residues lining the active site of *Ro*SmoC.



958

959 **Figure 4: Biochemical and structural analyses of 6-oxo-glucose reductase SmoB.** (a) Top:
960 Equilibrium oxygen exchange at C-6 of Glc via 6-OG facilitated by SmoB when incubated with
961 $NADP^+$ in $H_2{}^{18}O$. Bottom: Derivatization and MS fragmentation allows localization of ^{18}O to C6 of
962 Glc. (b) Transparent molecular surface and ribbon diagram of SmoB in complex with NADPH and
963 Glc. (c) Closeup view of SmoB•NADPH•Glc ternary complex. Backbone and carbon atoms of SmoB
964 are shown in ice blue and NADPH and glucose are shown in cylinder format. Electron density for
965 NADPH corresponds to the 2Fo − Fc map in blue at levels of 1σ. (d) Substrate binding pocket of
966 SmoB depicting hydrogen bonding interactions of glucose with the active site residues including the

conserved catalytic residues Asp71, Lys 104, His151 and Tyr76. Electron density corresponds to the 2Fo − Fc map (in blue) at levels of 1σ. The geometry of the SmoB-Glc complex indicates the likely trajectory of hydride addition to 6-OG. (e) Proposed mechanism of SmoB catalyzed reduction of 6-OG by NADPH showing hydride transfer from C4 of nicotinamide ring of NADPH to C6 carbonyl and Y76 (within the catalytic tetrad) as the proton donor. The red sphere is a bound water molecule; dotted lines are proposed hydrogen bonds.

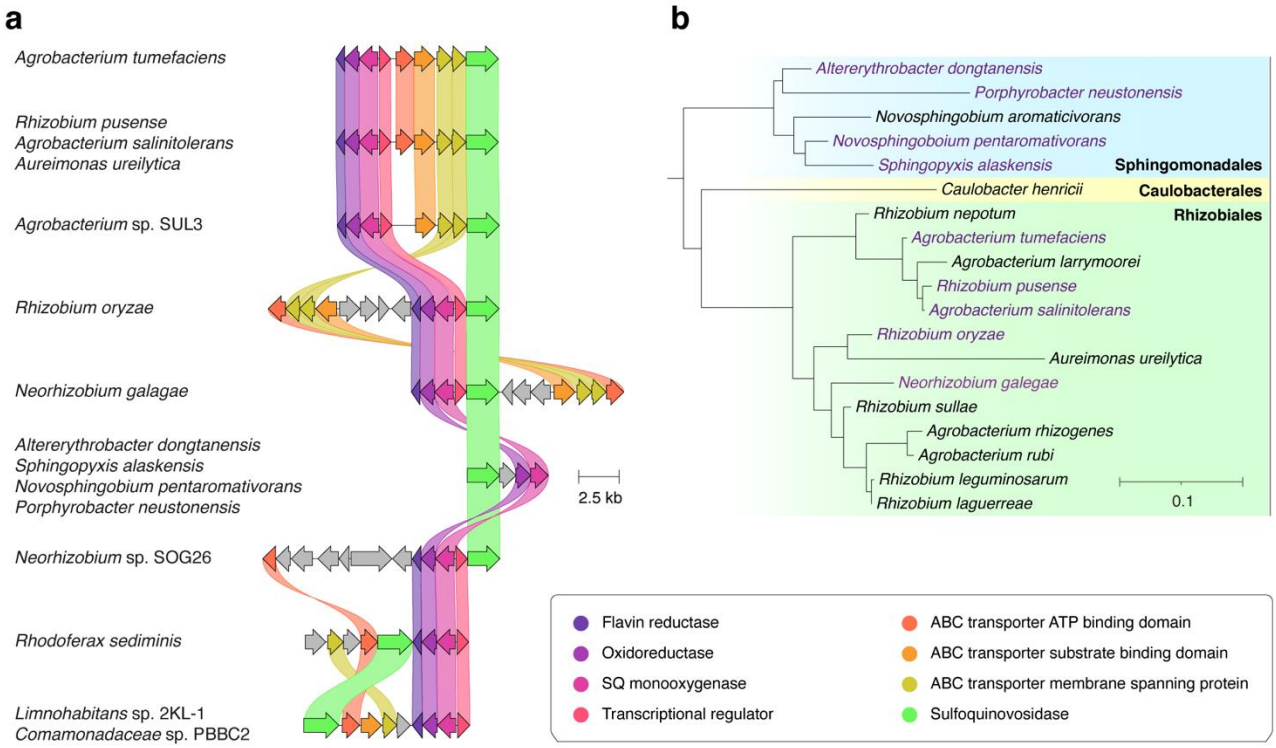


Figure 5: Prevalence of the sulfo-SMO pathway. (a) Architecture of the sulfo-SMO gene cluster in *A. tumefaciens* and homologous gene clusters in other organisms. Coloured links indicate ≥ 30% protein sequence similarity. Only those clusters encoding putative SQ monoxygenases and SQases were annotated as putative sulfo-SMO gene clusters. (b) A phylogenetic tree demonstrating the diversity of organisms possessing putative sulfo-SMO gene clusters. The tree was constructed by pruning of the All-Species Living Tree Project's 16s rRNA-based LTP release 132 (https://www.arb-silva.de/projects/living-tree/).