A microbial sulfoquinovose monooxygenase pathway 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 A. Järvå,[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 desulfurization

*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

Catabolism of sulfoquinovose (SQ, 6-deoxy-6-sulfoglucose), the ubiquitous sulfosugar produced by photosynthetic organisms, is an important component of the biogeochemical carbon and sulfur cycles. Here, we describe a new pathway for SQ degradation that involves oxidative desulfurization to release sulfite and the intact carbon skeleton of the sugar to support the growth of the plant pathogen *Agrobacterium tumefaciens*. SQ or its glycoside sulfoquinovosyl glycerol (SQGro) are imported into the cell by an ABC transporter system with an associated SQ binding protein. A sulfoquinovosidase hydrolyses the SQ glycoside and the liberated SQ is acted on by a flavin mononucleotide-dependent sulfoquinovose monooxygenase, in concert with an NADH-dependent flavin reductase, to release sulfite and 6-oxo-glucose. An NADPH-dependent oxidoreductase reduces the 6-oxo-glucose to glucose, enabling entry into primary metabolic pathways. Structural and biochemical studies provide detailed insights into the recognition of key metabolites by proteins in this pathway. Bioinformatic analyses reveal that the sulfoquinovose monooxygenase (smo) pathway is distributed across Alpha- and Betaproteobacteria and is especially prevalent within the Rhizobiales order. This strategy for SQ catabolism is distinct from previously described pathways as it enables the complete utilization of all carbons within SQ by a single organism with concomitant production of inorganic sulfite.

Introduction

Sulfoquinovose (SQ; 6-deoxy-6-sulfoglucose) is an anionic sulfosugar found in plant and cyanobacterial sulfolipids, 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 so its cycling is a significant component of the biogeochemical sulfur cycle[2]. Microbial communities play a dominant role in SQ cycling and usually more than one organism is required to completely assimilate this source of carbon and sulfur. Organisms with a tier 1 pathway, termed sulfoglycolysis, perform scission of the C3-C4 bond of SQ to give two three-carbon fragments; carbons 1-3 enter central metabolism, while carbons 4-6 bearing the sulfonate are excreted as dihydroxypropanesulfonate (DHPS) or sulfolactate (SL). Organisms with a tier 2 pathway are those that process DHPS and SL to access the remaining three carbon fragment and release inorganic sulfur. To date, 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' to release inorganic sulfite, which under aerobic conditions is readily oxidized to sulfate[1]. While many of the steps in the three tier 1 sulfoglycolysis pathways differ, all three pathways share the presence of a specialized glycoside hydrolase, a sulfoquinovosidase (SQase), which catalyzes the hydrolysis of SQ glycosides, such as SQGro, to release SQ[8,9].

While the tier 1 and 2 pathways described to date require two or more organisms to complete the 'biomineralization' of SQ, there is some evidence that this can also be accomplished by a single organism. Roy and co-workers have reported that an *Agrobacterium* strain from soil can completely consume SQ, with release of sulfate, although the genetic and biochemical details behind this process were not investigated[10]. We previously reported that *A. tumefaciens* C58 contains a functional SQase, with the ability to hydrolyze SQGro[8]. However, analysis of its genome did not reveal any genes homologous to those expected for known tier 1 sulfoglycolysis pathways.

Here, we investigate the 'biomineralization' of SQ by *Agrobacterium tumefaciens* (*Agrobacterium fabrum*) strain C58 and show that this organism effects the oxidoreductive desulfurization of SQ to release inorganic sulfite and glucose, which can feed into primary metabolism. We show that this pathway involves: a novel SQ/SQGro solute binding protein and associated ATP-binding cassette (ABC) transporter; an SQase to release SQ from its glycosides; a flavin-dependent SQ monooxygenase with paired flavin-reductase to effect oxidative desulfurization of SQ to sulfite and 6-oxo-glucose; and a NADPH-dependent oxidoreductase to reduce 6-oxo-glucose to glucose. X-ray

structures determined for each of these proteins in complex with relevant metabolites reveal the molecular basis of substrate binding and catalysis. We go on to show through bioinformatics analyses that this pathway – the first to enable the complete assimilation of SQ – is distributed across Alpha- and Betaproteobacteria and is particularly well-represented within the Rhizobiales order.

Results

Differential expression of a gene cluster in the presence sulfoquinovose

To determine if *A. tumefaciens* C58 can utilize SQ as a carbon source, we attempted to grow this organism in M9 minimal media containing SQ as the sole carbon source. *A. tumefaciens* C58 exhibited robust growth in this media and analysis of spent culture supernatant failed to detect DHPS or SL. Instead, the culture 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 the supernatant 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* grew on other sulfoquinovosides, including SQGro and methyl α-sulfoquinovoside (MeSQ), 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 concomitant release of sulfite.

We performed comparative proteomic experiments to identify changes associated with the growth of *A. tumefaciens* on SQ compared to glucose at mid-log phase (**Fig. 1b**). The largest and most significant change we observed was an increase in the abundance of proteins encoded by a single cluster of genes (*Atu3277-Atu3285*) for cells grown on SQ. Proteins encoded by *Atu3283* and *Atu3284* were not observed; however, they are predicted to be integral membrane proteins that can be difficult to detect using conventional proteomic workflows[13]. Thus, the gene cluster *Atu3277-Atu3285*, which was subsequently renamed *smoA-smoI*, appeared to be important for growth on SQ (**Fig. 1c**). While the protein encoded by *Atu3285* was previously identified as an SQase[8], the proteins encoded by other genes in the cluster were not annotated with functions that were consistent with any tier 1 pathway, suggesting that *A. tumefaciens* uses a different approach for the catabolism of SQ. The automated annotations ascribed to the respective gene products in the cluster, which included a putative ABC transporter system, sulfonate monooxygenase, SDR oxidoreductase, flavin reductase and exporters, enabled development of a hypothetical biochemical pathway that could explain the complete

assimilation of SQ by *A. tumefaciens* (**Fig. 1d**). We proceeded to biochemically validate this hypothesis and gain structural insights into the proteins involved.

Atu3282 (smoF) encodes an ABC transporter solute-binding protein that binds SQGro

Within the gene cluster identified through proteomics, *Atu3281* (*smoE*), *Atu3283* (*smoG*), and *Atu3284* (*smoH*) were annotated as an ABC transporter system, with *Atu3282* (*smoF*) encoding an associated periplasmic solute-binding protein. The substrate preferences of solute binding proteins are useful for assigning functions to their associated ABC transporters[14]. Accordingly, we produced recombinant SmoF (**Supplementary Fig. 3)** and demonstrated that it binds SQGro with K_d = 290 nM (ΔH = -11 kcal mol^{-1}, ΔS = -7 cal mol^{-1} deg^{-1}) (**Fig. 2a**, **Supplementary Table 3**). No binding was observed for the stereochemically-related monosaccharides D-glucose and D-glucuronic acid.

To delineate how SmoF recognizes its ligand, we used X-ray diffraction methods to obtain a high-resolution 3D structure of SmoF in its ligand-free apo state and in complex with SQGro (**Fig. 2b, Supplementary Table 4**). Like most ABC transporter solute-binding proteins, SmoF possesses two globular domains with a similar α/β fold forming a deep cleft lined with aromatic and polar residues to capture the ligand. Comparisons of the structures for ligand-free SmoF and the SQGro complex revealed a large conformational change in the protein resulting from inter-domain rotation upon SQGro binding. The relative movement of domains was assessed using the DynDom server, which indicated a hinge rotation of 31° about four linker regions connecting the two domains (**Supplementary Fig. 4**). SQGro is buried deep within the inter-domain cleft and residues from both domains accommodate this ligand through a network of hydrogen-bonding interactions (**Fig. 2c,d**). The sulfonate of SQGro, which is the defining feature of this sulfosugar, is accommodated by hydrogen-bonds to the side-chain of Thr220 (2.6 Å), backbone amides of Gly166 (3 Å) and Ser43 (2.8 Å), and an ordered water molecule that in turn hydrogen-bonds to the sidechain of His13 (3 Å) and Gln46 (3.2 Å) (**Fig. 2c,d**). These and the other interactions in the SQGro-bound ‘closed’ state stabilized SmoF substantially, as evidenced by a 15 °C increase in the protein melting temperature (**Supplementary Fig. 5**).

The structural basis of SQGro recognition by the SQase Atu3285 (SmoI)

We previously reported that *Atu3285 (smoI)* encodes an SQase that preferentially hydrolyses 2'*R*-SQGro, the natural stereoisomer of this glycoside[8]. To understand the molecular basis of the preference SmoI has for this stereoisomer, we determined the 3D structure of a pseudo-Michaelis complex: the inactive acid/base mutant SmoI-D455N in complex with 2'*R*-SQGro (**Fig. 2e,f**). SmoI-D455N•SQGro crystallized with four protomers in the asymmetric unit, each showing unambiguous

density of the substrate bound at the active site. As described previously, 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]. Arg438 and Glu135 make hydrogen-bonding interactions with the glyceryl aglycone of 2'*R*-SQGro. Only Arg438 interacts with the C2-hydroxyl group of the glyceryl aglycone and thus this residue appears to drive selectivity for the 2'*R*-SQGro stereoisomer.

Atu3277 (smoA) encodes a flavin mononucleotide (FMN) reductase

SmoA, annotated as a flavin reductase, was recombinantly expressed in *E. coli* and maintained a yellow color throughout purification, suggesting that it had co-purified with a flavin co-factor. A sample of this protein was heat-denatured to release the co-factor and the supernatant analyzed by LC-MS to reveal that FMN was the sole detectable flavin (**Supplementary Fig. 6**). Michaelis-Menten kinetics were conducted for SmoA with saturating FMN and NADH or NADPH to determine which of these reductants was preferred by the enzyme. 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 for SmoA (**Fig. 3a, Supplementary Fig. 7)**. Owing to difficulties in obtaining structural data for this enzyme, we also studied a close homologue from *Rhizobium oryzae* (*Ro*SmoA, UniProt accession number: A0A1X7D6Q3), which possesses a syntenic gene cluster to *Atu3277-Atu3285*. Recombinant *Ro*SmoA also co-purified with FMN (**Supplementary Fig. 6**) and utilized the NADH cofactor 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)**.

Atu3279 (smoC) encodes an SQ monooxygenase that desulfurizes SQ

SmoC is annotated as an alkanesulfonate monooxygenase, though it possesses only 30% sequence identity with the well-characterized alkanesulfonate monooxygenase SsuD, from *E. coli*. SsuD catalyzes the $FMNH_2$- and O_2-dependent oxidation of alkanesulfonates to produce the corresponding aldehyde and sulfite, with a preference for pentanesulfonate.[15] The mechanism of this and related enzymes have been intensively studied yet remain enigmatic. The transformation is thought to involve initial formation of a C4a-peroxy or N5-peroxy flavin species on-enzyme. One mechanism posits that the terminal peroxide oxygen attacks the sulfonate sulfur of the substrate before undergoing a rearrangement to effect C-S bond fissure and release of the aldehyde and sulfite products (**Supplementary Fig. 8a**)[16]. An alternative mechanism suggests the peroxide deprotonates C6, which is then oxidized to an α-hydroxysulfonate that undergoes elimination to produce sulfite and the aldehyde[17]. To demonstrate activity for recombinant SmoC (**Supplementary Fig. 3**), we adapted assays developed for SsuD that use Ellman's reagent to detect sulfite released by the enzyme[18]. Direct detection of the putative sugar product, 6-oxo-glucose (6-OG), is not trivial as this molecule exists as

a complex equilibrium of (hemi)acetals and hydrates that have poor stability. Thus, SmoC was incubated with SQ in the presence of SmoA, FMN and NADH, which generate $FMNH_2$ *in situ*, and the concentration of sulfite determined periodically using Ellman's reagent (**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 when SQ was replaced with other sulfonates, including SQGro (the precursor to SQ) or HEPES (an unrelated sulfonate) demonstrating that, unlike the promiscuous SsuD, SmoC has high specificity for SQ. As such, the hydrolysis of SQGro by SmoI necessarily precedes oxidative desulfurization by SmoC. This observation is further supported by ITC, where SQ was found to bind SmoC with K_d = 3 µM in the absence of any flavin-based cofactors, whereas no binding was detected for SQGro (**Fig. 3c**, **Supplementary Table 3**). The unique SQ monooxygenase activity of SmoC defines this pathway: it is the enzyme that effects fissure of the C–S bond in SQ, and so it was chosen as the namesake for this gene cluster and *Atu3277-Atu3285* were renamed the **S**Q **M**ono**O**xygenase cluster (*smoA-I*).

While we could readily crystallize SmoC, these crystals only diffracted to a maximum resolution of 3.4 Å. The corresponding low-resolution map suggested that SmoC exists as a dimer, which was confirmed in solution by SEC-MALS (**Supplementary Fig. 9**). To obtain structural information for an SQ monooxygenase, we turned to the homolog from *R. oryzae* (*Ro*SmoC). Recombinant *Ro*SmoC exhibited similar activity and substrate selectivity for SQ to SmoC (**Supplementary Fig. 8**) and provided crystals that diffracted to 1.9 Å. Importantly, the low-resolution structure of *A. tumefaciens* SmoC superimposed with the high-resolution *Ro*SmoC structures with a peptide backbone rmsd of 0.4 Å across the entire structure, providing confidence that both enzymes shared a common structure and function (**Supplementary Fig. 10**). Both SQ monooxygenases consist of a core $(\alpha/\beta)_8$ TIM barrel with three additional insertion regions, analogous to monooxygenases from the bacterial luciferase family. The protomers exist as a homodimer that buries 4697 $Å^2$ of surface area, amounting to 18% of total accessible surface area for each protomer (**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, a long-chain alkane monooxygenase LadA (3B9O.pdb, rmsd 2.6/312 residues, Z-score of 31.0), and $FMNH_2$-dependent alkanesulfonate monooxygenase SsuD (1M41.pdb, rmsd 1.8/317 residues, Z-score of 41.2).

Comparisons of the *Ro*SmoC structure with LadA (3B9O.pdb) in complex with FMN (a pseudo co-factor) enabled identification of the FMN binding for site *Ro*SmoC: a deep hydrophobic pocket that accommodates the isoalloxazine ring system and extends out to the protein-solvent interface, which

is gated by conserved phosphate-binding residues Tyr136 and Ser189 (**Fig. 3e**)[19]. The close structural and functional relationship of *Ro*SmoC to alkanesulfonate monooxygenase SsuD (1M41.pdb) was evident from the conservation of a putative sulfonate binding site comprised of the side-chains Trp206, Arg236, His238, Tyr341 and His343 (**Fig. 3f**).[18,19] Aside from conferring these enzymes with an ability to bind sulfonates, these conserved active-site residues have been suggested to contribute to the stabilization of a peroxyflavin intermediate in SsuD[18,19]. Efforts to obtain crystals of a *Ro*SmoC–SQ complex were unsuccessful, limiting further insights into the origin of enzyme specificity towards SQ over other sulfonates.

Atu3278 (smoB) encodes an oxidoreductase that converts 6-oxo-glucose to glucose

SmoB is annotated as a short-chain dehydrogenase/reductase (SDR) and we had hypothesized that it was responsible for reduction of 6-OG to glucose (**Fig. 1d**). Since 6-OG is difficult to study directly, we tested our hypothesis by looking for SmoB-mediated isotope incorporation into glucose at equilibrium (**Fig. 4a**). Assuming our hypothesis to be true, and as a consequence of microscopic reversibility, incubation of SmoB with a nicotinamide co-factor and glucose in $H_2{}^{18}O$ should result in transient formation of 6-OG, rapid and reversible hydration/dehydration with $H_2{}^{18}O$ to compete-out ^{16}O at C6 for ^{18}O, and reduction to give 6-^{18}O-glucose. In parallel to this process, ^{18}O incorporation will occur at C1 of glucose through a similar series of hydration/dehydration reactions. Before proceeding with these experiments, we used ITC to establish which nicotamide cofactor was suitable for SmoB: NADPH bound to SmoB with K_d ~2 μM, while no binding was observed for NADH (**Supplementary Fig. 11, Supplementary Table S3**). Thus, glucose pre-equilibrated in $H_2{}^{18}O$ was incubated with SmoB and $NADP^+$ then analyzed by mass spectrometry to reveal the formation of a product 4 Da greater in mass than glucose, presumably due to the incorporation of two ^{18}O atoms into glucose. The crude reaction mixture was subjected to peracetylation (Ac_2O/pyridine) then LC-MS analysis to confirm that the +4 Da product co-eluted with authentic D-glucose-pentaacetate (**Supplementary Fig. 12**). To determine that the ^{18}O label was being incorporated at C6 of glucose, we used electron-impact GC-MS, which required conversion of the reaction product to the acyclic pentapropionate aldonitrile (**Supplementary Fig. 13**)[20]. This approach provided diagnostic C1-C5 and C5-C6 fragment 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 is incorporated at C6. Only enzymatic reactions conducted in the presence of $NADP^+$ produced product labelled with ^{18}O at C6: NAD^+ failed to produce any product, supporting our observations by ITC and defining the cofactor specificity of SmoB.

We determined the 3D structure of SmoB using X-ray diffraction methods. This initial structure revealed that SmoB exists as a compact trimer, however the C-terminal His_6-tag in this construct occupied the putative active site of adjoining subunits, making co-crystallization with cofactors difficult (**Supplementary Fig. 14**). To overcome this issue, SmoB was subcloned into a different vector and expressed with a cleavable N-terminal purification tag. This protein maintained the same catalytic activity and SEC-MALS confirmed it remained a trimer in solution (**Supplementary Fig. 15**). This SmoB construct was co-crystallized with NADPH and a ternary product complex obtained by soaking crystals with D-glucose (**Fig. 4b**). These crystals diffracted to a resolution of 1.5 Å and the resulting model revealed that SmoB is an $(\alpha/\beta)_8$ TIM barrel fold with a C-terminal cofactor binding site. The overall fold has high structural conservation with members of the 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 a π-π stacking interaction with the nicotinamide ring that positions the reactive center (C4) at a distance of 3 Å from C-6 of glucose, appropriate 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 that is common to the AKR superfamily (**Fig. 4e**)[21].

SMO pathways occur in the Alphaproteobacteria and Betaproteobacteria

To ascertain how widespread this pathway for SQ utilization might be, a Multigene BLAST search was conducted of the non-redundant protein set of the NCBI for gene clusters that contain homologous SQases and SQ monoxygenases. This identified many putative *smo* gene clusters across the *Agrobacterium* and *Rhizobium* genus within the *Rhizobiales* order and evidence of some broader expansion into the Alphaproteobacteria and Betaproteobacteria classes (**Fig. 5**). Amongst these putative *smo* gene clusters, some were syntenic while others were substantially rearranged (non-syntenic) or modified to make use of other (non-ABC) transporter systems. The use of diverse transport systems is not surprising: a similar phenomenon has been observed for the tier-1 sulfo-ED pathway[4,5]. Indeed, sulfo-ED gene clusters have been identified in several *Rhizobiales*[4,5], suggesting that there has been ample opportunity for genetic exchanges between these pathways during their evolution.

Discussion

This work builds on a 20 year-old observation that an unclassified *Agrobacterium* strain can grow on SQ and release inorganic sulfur in the process[10]. We have confirmed this observation and provide a detailed account of the discovery and biochemical/structural characterisation of the first pathway to facilitate the complete assimilation of SQ and its glycosides within a single organism. We have also used this information to provide broader context to the discovery by illustrating this pathway's distribution amongst other microorganisms, predominantly those of the Alphaproteobacteria class. The pathway features several proteins with hitherto undescribed activities, including: an SQGro-binding protein; an $FMNH_2$- and O_2-dependent SQ monooxygenase that defines this 'SMO' pathway by catalyzing scission of the C–S bond in SQ, and an oxidoreductase dedicated to the NADPH-dependent reduction of 6-OG to glucose. Like all other sulfoglycolytic pathways studied to date, the SMO pathway also possesses a conserved SQase, which is essential for liberating SQ from its precursor glycoside SQGro[8,9].

The SMO pathway shares similarities with other sugar-metabolizing pathways in bacteria. For example, the presence of SmoI (SQase), SmoF (SQGro binding protein) and SmoE/G/H (ABC transporter) encoded by the *smo* cluster, is analogous to MalP (maltodextrin phosphorylase), MalE (maltose binding protein) and MalF/G/K (ABC transporter) encoded by the *mal* operon of *E. coli* that imports and degrades maltose[22]. Meanwhile, SmoC (SQ monooxygenase) and SmoA (flavin reductase) of the SMO pathway are reminiscent of the SsuD ($FMNH_2$-dependent alkylsulfonate monooxygenase) and SsuE (NADPH-dependent FMN reductase) pair encoded by the *ssu* operon of *E. coli* that degrades alkanesulfonates[15]. Indeed, it is feasible, perhaps even likely, that the SMO pathway arose through the recombination of analogous sugar- and sulfonate-metabolising pathways.

Through our structural studies, we have identified key residues involved in sulfosugar recognition and processing, in order to provide greater confidence to bioinformatic analyses of putative *smo* gene clusters: an approach that has proven valuable for the identification of tier 1 sulfoglycolytic pathways[8,23,24]. This includes the Thr220-Gly166-Ser43-H_2O(His13-Gln46) cluster of SmoF for the recognition of SQGro, the Arg283-Tryp286-H_2O(Tyr491) triad of SmoI for the recognition of SQGro; and the Trp206-Arg236-His238-Tyr341-His343 constellation of SmoC for the recognition of SQ. Given the importance of the SQ monooxygenase SmoC to the SMO pathway, further empirical and computational work is warranted to understand what interactions drive its selectivity for SQ, which lies in contrast with the promiscuity exhibited by alkanesulfonate monoxygenases like SsuD.

The prevalence of the SMO pathway in bacteria of the *Rhizobiales* order is intriguing, since many members of this order are plant symbionts or pathogens. Indeed, those that do not possess an SMO pathway often possess a complementary tier 1 sulfo-ED pathway[4]. Accordingly, it appears that plant sulfolipid catabolism is important for bacteria in this order, whether they be plant pathogens/symbionts or free-living organisms adopting an oligotrophic saprophytic lifestyle in substrate replete with decaying plant tissues. Symbiotic bacteria of the *Rhizobiales* order reside within the root nodules of their plant host, where they harness four-carbon substrates from the host for energy and central metabolism[25]. Sugawara and co-workers showed that sulfonate utilization gene clusters were expressed by the plant symbiont *Bradyrhizobium diazoefficiens* USDA 110 within these nodules and that this may be important for utilizing diverse sulfur sources to support symbiotic and possibly free-living lifestyles[26]. With sulfolipid representing a large and accessible pool of sulfur in plants, one possible purpose of the SMO pathway may be to salvage sulfur for these bacteria. This is an important distinction between the SMO pathway and the tier 1 sulfoglycolytic pathways: the latter supports two-member microbial communities containing a second member with a tier 2 pathway to provide access to the sulfur of SQ[27]. In this sense, use of the SMO pathway, which enables the complete utilization of the carbon skeleton and access to the sulfur of the monosaccharide can be considered a ‘selfish’ metabolic strategy, and could provide an advantage in the highly competitive soil environment or in the absence of other bacterial species within colonized plant tissues. These ideas, combined with the pathway’s requirement for molecular oxygen to effect C–S bond fissure, likely explain why the SMO pathway occurs within those bacteria that are commonly associated with plants. Understanding how the SMO and tier 1 pathways impact fitness within the different environmental niches discussed here remains an important question, with answers that have significant implications for understanding plant diseases and symbioses, as well as soil chemistry.

Methods

Growth studies

Cultures of *A. tumefaciens* C58 were grown in a phosphate-buffered mineral salts media (M9, pH 7.2), with glucose 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 for culture supernatant

The reducing sugar assay was performed according to the procedure of Blakeney and Mutton[28]. This assay uses pre-prepared alkaline diluent and PAHBAH working solution. Alkaline diluent was prepared by the addition of sodium hydroxide (20 g, 0.50 mol) to a solution of 0.10 M trisodium citrate (50 mmol, 500 ml) and 0.02 M calcium chloride (13 mmol, 500 ml). PAHBAH working solution was prepared by dissolving 4-hydroxybenzhydrazide (PAHBAH) (0.25 g, 1.6 mmol) in alkaline diluent (50 ml). The PAHBAH working solution should be made fresh shortly before use. To determine reducing sugar concentration, 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 then 0.5 ml of the mixture was diluted into 1.0 ml of deionized water and the absorbance read at 415 nm using a Varian Cary50 UV/visible spectrophotometer. Concentrations of SQ were determined with reference to a standard curve constructed using SQ.

Turbidometric sulfate assay for culture supernatant

The sulfate assay was performed according to the procedure of Sörbo[29]. This assay uses a Ba-PEG reagent, which contains PEG to stabilize $BaSO_4$ crystals and a small amount of pre-formed $BaSO_4$ seed crystals to improve the reproducibility and linearity of the assay. The Ba-PEG reagent should be prepared fresh before use. Ba-PEG reagent was prepared by dissolving $BaCl_2$ (42 mg, 0.20 mmol) and polyethylene glycol 6000 (0.75 g) in deionized water (5.0 ml). A small amount of Na_2SO_4 (10 µl, 50 mM) was added to this solution, with efficient magnetic stirring to generate preformed $BaSO_4$ seed crystals. Individual sulfate assays were conducted as follows. A sample (typically 100 µl, containing a maximum of 2.5 µmol of Na_2SO_4) was diluted to 0.1 ml with deionized water before the addition of 0.5 M HCl (0.1 ml) followed by Ba-PEG reagent (0.1 ml). The mixture was mixed vigorously and the absorbance of the sample at 400 nm determined using a Varian Cary50 UV/visible

spectrophotometer. Concentrations of sulfate were determined by reference to a standard curve constructed using Na_2SO_4.

Colorimetric fuchsin sulfite assay for culture supernatant

The fuchsin sulfite assay was performed according to the procedures of Brychkova *et al*.[30] and Kurmanbayeva *et al*.[31]. This procedure requires three pre-prepared solutions, Reagents A, B and C. Reagent A was prepared by dissolution of basic fuchsin (4.0 mg, 12 µmol) in deionized water (8.25 ml) at 0 °C, prior to the addition of 98% H_2SO_4 (1.25 ml). Reagent B was prepared by diluting formaldehyde (36% in H_2O, 0.32 ml) in deionized water (9.68 mL) at 0°C. Reagent C was prepared by dilution of Reagent A (1 ml) in deionized water (7 ml), prior to the addition of solution reagent B (1 ml). Individual sulfite assays were performed by addition of Reagent C (516 µl) to a mixture of sample (72 µl) and 0.5 mM Na_2SO_3 (12 µl), with the latter providing a stable background signal for reference. The sample was incubated at r.t. for 10 min and the absorbance of the sample at 570 nm determined using a Varian Cary50 UV/visible spectrophotometer. Concentrations of sulfite were determined by reference to a standard curve constructed using Na_2SO_3.

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

M9 minimal media (5 ml) containing 10 mM glucose was inoculated with *A. tumefaciens* C58 and grown to stationary phase at 30 °C (250 rpm). A 50 µl aliquot of this 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). At OD_{600} 0.27 and OD_{600} 0.49, 950 µl samples of culture supernatant were diluted with 100 µl of D_2O and ^{13}C-NMR spectra acquired using a 400 MHz spectrophotomer (100 MHz for ^{13}C).

Growth of A. tumefaciens *C58 on diverse alkanesulfonates*

M9 minimal media (5 ml) containing 10 mM glucose 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 (MeSQ), glycer-1-yl α-sulfoquinovoside (SQGro), 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), MeSQ, and SQGro, but not on any other sulfonate. Control experiments established that *A. tumefaciens* grows on glucose in the presence and absence of cyclohexylamine or dicyclohexylamine, and does not grow on cyclohexylamine or dicyclohexylamine alone.

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 and 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[32]. Digested samples were acidified to a final concentration of 0.5% formic acid and desalted using C18 stage tips[33] 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^{-1} 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^{-1}. 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)[34] 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[35] was enabled in addition to the re-quantification module. The resulting protein group output was processed within the Perseus (v1.4.0.6)[36] 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[37] partner repository with the dataset identifier PXD014115.

Cloning

Oligonucleotides encoding Atu3277 (SmoA), Atu3278 (SmoB), Atu3279 (SmoC) and Atu3282 (SmoF) were amplified by PCR using Phusion polymerase HF master mix (NEB), the appropriate primers listed in **Supplementary Table 1** and *A. tumefaciens* C58 gDNA as template. Oligonucleotides encoding *Ro*SmoA and *Ro*SmoC were synthesized (IDT) to provide the sequences listed in **Supplementary Table 1**. These were cloned into the pET29b(+) vector at the *NdeI* and *XhoI* sites and sequence-verified by Sanger sequencing to give expression vectors for SmoA, SmoB, SmoC, SmoF, *Ro*SmoA and *Ro*SmoC. Due to interference from the SmoB C-terminal His_6-tag during structural studies, the *smoB* (*Atu3278*) gene was sub-cloned into the pET-YSBLIC3C vector[38] by PCR amplification with the relevant primers in **Supplementary Table 1** and In-Fusion® cloning (Clontech Laboratories, Inc.) into linearized YSBLIC3C vector according to the manufacturer's protocol. The expression plasmid was sequence-verified by Sanger sequencing.

Protein expression and purification

All vectors were transformed into 'T7 Express' *E. coli* (NEB), except for the vector encoding SmoF (Atu3282), which 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 starter cultures were 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 400 µM, and incubation with shaking (200 rpm) continued at 18 °C for 19 h. Cells were harvested by centrifugation at 8,000 g for 20 min at 4

°C then resuspended in 40 ml 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 (**Supplementary Fig. 2**). SmoI (Atu3285 or *At*SQase) was prepared as previously described[8].

SEC-MALS analyses

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 > 2 CV 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 of protein at 6 mg ml^{-1} 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 min shorter than the LC solutions run to maintain synchronization. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data were analyzed using the Astra V software. Molecular weights 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 $\mu Cal.s^{-1}$. Proteins and substrates were equilibrated into degassed and filter-sterilized buffer (50 mM NaPi, 200 mM NaCl, pH 7.4 for SmoC/F and 25 mM NaPi, pH 7.5 for Smo B). Protein concentration was determined by BCA assay (Thermo Fisher) before intiating experiments. For SmoC–SQ binding, 600 µM 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 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, 1 mM NADH was titrated into the ITC cell containing 40 µM SmoB as a series of 19 × 3 µl injections with a pre-injection of 1 × 4 µl. The delay between injections was set at 150 s, with an initial injection delay of 180 s. All data analysis was performed in MicroCal ITC Origin Analysis software (Malvern).

Nano Differential Scanning Fluorescence analysis of SmoF

Thermal stability analysis for SmoF in the presence and absence of SQGro ligand 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 a concentration of 1 mg ml^{-1} in 50 mM citrate, 150 mM NaCl at pH 5.5, with a 10 µl capillary load per sample. Data acquisition and analysis was performed with PR.ThermControl (NanoTemper) software.

Identification of the flavin co-factor that co-purified with SmoA

100 µl of recombinant flavin reductase (SmoA or *Ro*SmoA) at a concentration of 20 mg ml^{-1} 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 analyzed 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% FA; solvent B: MeCN + 0.1% FA); flow rate: 0.6 ml/min; DAD – 254 and 214 nm.

Michaelis-Menten kinetic analyses of SmoA and Ro*SmoA*

Reactions were conducted at 25 °C in 96-well plate format and involved the addition of SmoA or *Ro*SmoA (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-catalyzed 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 (Perkin Elmer). 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

This SQ monooxygenase activity assay is based on a previously described alkanesulfonate monooxygenase activity assays[18] and uses Ellman's reagent to quantify sulfite released by these enzymes. 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 (SmoC or *Ro*SmoC) 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 Multimodal Plate Reader (Perkin Elmer). 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 monooxygenases 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, glucose was incubated in 98% $H_2{}^{18}O$ with heating at 80 °C for 2 days, then evaporated to dryness to give C1-^{18}O-labelled glucose. Labelling was determined to be 95% by mass spectrometry based on intensities of the M and M+2 peaks. Using $H_2{}^{18}O$ buffer (100 mM potassium phosphate, pH 7.0), NAD^+ and $NADP^+$ were each added at 0.05 molar equivalent to C1-^{18}O-glucose and SmoB. Four control experiments were conducted: one without enzyme, one without NAD^+ and $NADP^+$, one in $H_2{}^{16}O$, and one in $H_2{}^{16}O$ with unlabeled glucose. Reactions were monitored by mass spectrometry. Only in the experimental sample containing enzyme, $H_2{}^{18}O$ and $NAD^+/NADP^+$ was an M+4 signal observed and this reached a maximum intensity after 72 h. Two additional reactions were performed using SmoB, glucose and either $NADP^+$ or NAD^+ in $H_2{}^{18}O$ and only the reaction containing $NADP^+$ generated the M+4 species. To confirm that the M+4 species was 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 at the anomeric position during HPLC analysis. The reaction mixture from above was evaporated under reduced pressure. 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 remove pyridine. The organic solution containing peracetylated glucose was analyzed 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: 0 to 65% B over 20 min (solvent A: water + 0.1% FA; solvent B: MeCN + 0.1% FA); flow rate: 0.6 ml/min. 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 of SmoB-glucose reaction mixture (containing ≈2.5 nmol glucose) was transferred to a GC vial insert (deactivated) together with 1 nmol *scyllo*-inositol as an internal standard. Samples were derivatized as described in Antoniewicz *et al.*[20], with minor modifications. Briefly, samples were dried (*in vacuo*, 35 °C with a 40 µl methanol wash), followed by addition of hydroxylamine hydrochloride (20 mg ml^{-1} in 25 µl pyridine) and incubation at 90 °C for 1 h. Vials were cooled briefly at r.t. before the addition of propionic anhydride (50 µl) and incubation at 60 °C for 30 min. Samples were evaporated to dryness under a stream of nitrogen at 60 °C and resuspended in EtOAc (40 µl). 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 a 2.5 nmol scale in the assay buffer mixture. Samples were blinded for analysis. The derivatized labelled glucose samples (**Supplementary Fig. 13** and **Supplementary Table 7**) were analyzed 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 °C 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^{-1}; 295 °C to 320 °C at 25 °C min^{-1}; 320 °C for 2 min. Glucose and *scyllo*-inositol were identified by reference to authentic standards. A calibration curve was generated using glucose standard in assay buffer (starting concentration 50 nmol, 2-fold dilution series). **Supplementary Fig. 12** shows the fraction of labelled fragments, corrected for isotope natural abundance by DExSI analysis[39].

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. Unless otherwise stated, all crystals were grown at 20 °C.

Crystals of apo-SmoF were obtained by mixing 0.15 µl of protein stock (50 mg ml^{-1} protein in 50 mM citrate, 150 mM NaCl, pH 5.5) with 0.15 µl mother liquor (0.3 M ammonium acetate, 0.1 M Bis-Tris, 25% w/v PEG 3350, pH 5.5) housed in a Rigaku Xtaltrak plate hotel to enable consistent growth and monitoring at 6 °C. Crystals were harvested with nylon CryoLoops™ (Hampton Research) and cryopreserved in liquid nitrogen without additional cryoprotectants.

Crystals of SmoF were initially obtained by mixing 0.15 µl of protein stock (3.5 mg ml^{-1} protein with 2'*R*-SQGro at a 1:10 molar ratio in 50 mM citrate, 150 mM NaCl, pH 5.5) with 0.15 µl mother liquor (30% (w/v) polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M tris chloride, pH 8.5). The resulting crystals were used to prepare a seed stock by mixing the crystallization drop with 100 µl mother liquor and vortexing for 60 s with one teflon bead. An optimisation plate was setup with drops comprised of 0.1 µl of various mother liquors (28-36% (w/v) polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M tris chloride, pH 7.1-9.1), 50 nl seed stock solution, and 0.15 µl protein stock (4 mg ml^{-1} protein with 2'*R*-SQGro at a 1:10 molar ratio in 50 mM citrate, 150 mM NaCl, pH 5.5). 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 harvested with a nylon CryoLoop™ (Hampton Research) and cryopreserved in liquid nitrogen with 25% (v/v) ethylene glycol as cryoprotectant.

Crystals of SmoI-D455N-E370A-E371A were obtained by mixing 0.4 µl of protein stock (35 mg ml^{-1} protein in 50 mM NaPi, 300 mM NaCl, pH 7.4) with 0.5 µl mother liquor (26% PEG 3350 w/v, 0.2 M KSCN, 0.1 M Bis-Tris propane, pH 6.5). Crystals were soaked with solid SQGro in mother liquor for 2 min prior to harvesting with nylon CryoLoops™ (Hampton Research) and cryopreserved without additional cryoprotectants.

Crystals of apo-SmoC were obtained by mixing 0.6 µl of protein stock (60 mg ml^{-1} protein in 50 mM Tris, 300 mM NaCl, pH 7.5) with 0.5 µl mother liquor (0.2 M NaCl, 0.1 M MES pH 6, 26% PEG 6000 w/v and 10 mM SQ-glucitol). Crystals of apo-*Ro*SmoC were obtained by mixing 0.1 µl of protein stock (11.7 mg ml^{-1} protein in 50 mM Tris, 300 mM NaCl, pH 7.5) with 0.2 µl mother liquor (0.2M $NaNO_3$, 20% PEG 3350 w/v and 10 mM SQ). Crystals were harvested with nylon CryoLoops™ (Hampton Research) and cryopreserved in liquid nitrogen without additional cryoprotectants.

Crystals of SmoB-apo (YSBLIC3C construct) were obtained by mixing 0.15 µl of protein stock (20 mg ml^{-1} protein in 50 mM NaPi, 150 mM NaCl, pH 7.4) with 0.15 µl mother liquor (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 complex, crystals were obtained by mixing 0.15 µl of protein stock (20 mg ml^{-1} protein in 50 mM NaPi, 150 mM NaCl, 2 mM NADPH, pH 7.4) with 0.15 µl mother liquor (0.1 M succinic acid, sodium dihydrogen phosphate, glycine buffer (SPG buffer, Qiagen), 25% w/v PEG 1500 at pH 6.0). For the SmoB•NADPH•Glc complex, crystals were obtained in a hanging drop by mixing 1 µl of protein stock (13 mg ml^{-1} protein in 50 mM NaPi 150 mM NaCl, pH 7.4) with 1 µl of mother liquor (2 mM NADPH, 0.1 M SPG (Qiagen), 25% w/v PEG 1500 at pH 6). Crystals were soaked with solid glucose in mother liquor for 1 min prior to harvesting with nylon CryoLoops™ (Hampton Research) and cryopreserved without additional cryoprotectants.

X-ray data collection, processing and refinement

The data were processed and integrated using XDS[40] and scaled using SCALA[41] included in the Xia2 processing system[42]. 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[43] with a search model created from PDB ID: 6DTQ[44]. The structure of *Ro*SmoC was solved by molecular replacement using the ensemble based on PDB ID: 1M41[18] as an initial search model. The structure of SmoB was determined using molecular replacement with the monomer of an aldo-keto reductase from *S. enterica* (PDB ID: 4R9O) as the initial model. The apo-SmoF structure was solved using a dissected C-terminal domain of the SmoF•SQGro structure. Structures were built and refined by iterative cycles using Coot[45] and REFMAC[46] or Phenix,[47] 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[48]. The final structures gave R_{cryst} and R_{free} values along with data and refinement statistics that are presented in **Supplementary Table 4-6.** 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 at the Australian Synchrotron using the MX2 beamline (At3282•SQGro complex, to 1.7 Å). 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 analyzed using the protein interactions, surfaces, and assemblies (PISA) server.[49] Structural comparisons and structure-based sequence alignments were conducted using PDB25 search on DALI server against a representative subset of the Protein Data Bank[50]. All structure figures were generated using ccp4mg[51].

Bioinformatic analysis SMO pathway prevalence

Each gene within the *A. tumefaciens* C58 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[52] and queried using the *A. tumefaciens* C58 SMO gene cluster. Clusters identified by this workflow with both an SQ monoxygenase and SQase homolog were regarded as putative SMO gene clusters. Clusters representative of the observed diversity were visualized using Clinker[53]. A phylogenetic tree of species possessing a putative SMO gene cluster was generated by pruning the All-Species Living Tree Project's 16s rRNA release 132[54] using iTOL[55].

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[56] (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**, E4298-305 (2015).
5. Li, J. et al. A Sulfoglycolytic Entner-Doudoroff Pathway in *Rhizobium leguminosarum* bv. *trifolii* SRDI565. *Appl. Environ. Microbiol.* **86**, e00750-20 (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**, 29358-65 (1999).
16. Thakur, A. et al. Substrate-Dependent Mobile Loop Conformational Changes in Alkanesulfonate Monooxygenase from Accelerated Molecular Dynamics. *Biochemistry* **59**, 3582-3593 (2020).
17. Liew, J.J.M., El Saudi, I.M., Nguyen, S.V., Wicht, D.K. & Dowling, D.P. Structures of the alkanesulfonate monooxygenase MsuD provide insight into C-S bond cleavage, substrate scope, and an unexpected role for the tetramer. *J. Biol. Chem.* **297**(2021).
18. 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).
19. 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**, 453-65 (2008).
20. Antoniewicz, M.R., Kelleher, J.K. & Stephanopoulos, G. Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal. Chem.* **83**, 3211-6 (2011).
21. Penning, T.M. The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234**, 236-46 (2015).
22. Dippel, R. & Boos, W. The Maltodextrin System of *Escherichia coli*: Metabolism and Transport. *J. Bacteriol.* **187**, 8322 (2005).
23. Sharma, M. et al. Dynamic Structural Changes Accompany the Production of Dihydroxypropanesulfonate by Sulfolactaldehyde Reductase. *ACS Catalysis* **10**, 2826-2836 (2020).
24. Sharma, M. et al. Molecular Basis of Sulfosugar Selectivity in Sulfoglycolysis. *ACS Cent. Sci.* **7**, 476-487 (2021).
25. Udvardi, M. & Poole, P.S. Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781-805 (2013).
26. 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).

27. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A.M. Sulfoquinovose degraded by pure cultures of bacteria with release of C_3-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39-45 (2012).
28. 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).
29. Sörbo, B. Sulfate: Turbidimetric and nephelometric methods. in *Methods Enzymol.*, Vol. 143 3-6 (Academic Press, 1987).
30. 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).
31. Kurmanbayeva, A. et al. Determination of Total Sulfur, Sulfate, Sulfite, Thiosulfate, and Sulfolipids in Plants. in *Plant Stress Tolerance: Methods and Protocols* (ed. Sunkar, R.) 253-271 (Springer New York, New York, NY, 2017).
32. 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-mcp201 (2011).
33. 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).
34. 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).
35. 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).
36. Tyanova, S. et al. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453-6 (2015).
37. 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).
38. Fogg, Mark J. & Wilkinson, Anthony J. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.* **36**, 771-775 (2008).
39. 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).
40. Kabsch, W. Xds. *Acta Crystallogr., Section D: Biol. Crystallogr.* **66**, 125-132 (2010).

41. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D* **62**, 72-82 (2006).
42. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186-190 (2010).
43. Storoni, L.C., McCoy, A.J. & Read, R.J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432-8 (2004).
44. Shukla, S. et al. Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864-5876 (2018).
45. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126-2132 (2004).
46. 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).
47. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-21 (2010).
48. Long, F. et al. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. Sect. D* **73**, 112-122 (2017).
49. 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).
50. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-W549 (2010).
51. 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).
52. Medema, M.H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218-23 (2013).
53. Gilchrist, C.L.M. & Chooi, Y.H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* (2021).
54. Yilmaz, P. et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucl. Acids Res.* **42**, D643-D648 (2013).
55. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* **47**, W256-W259 (2019).
56. 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 the SMO gene cluster; 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.

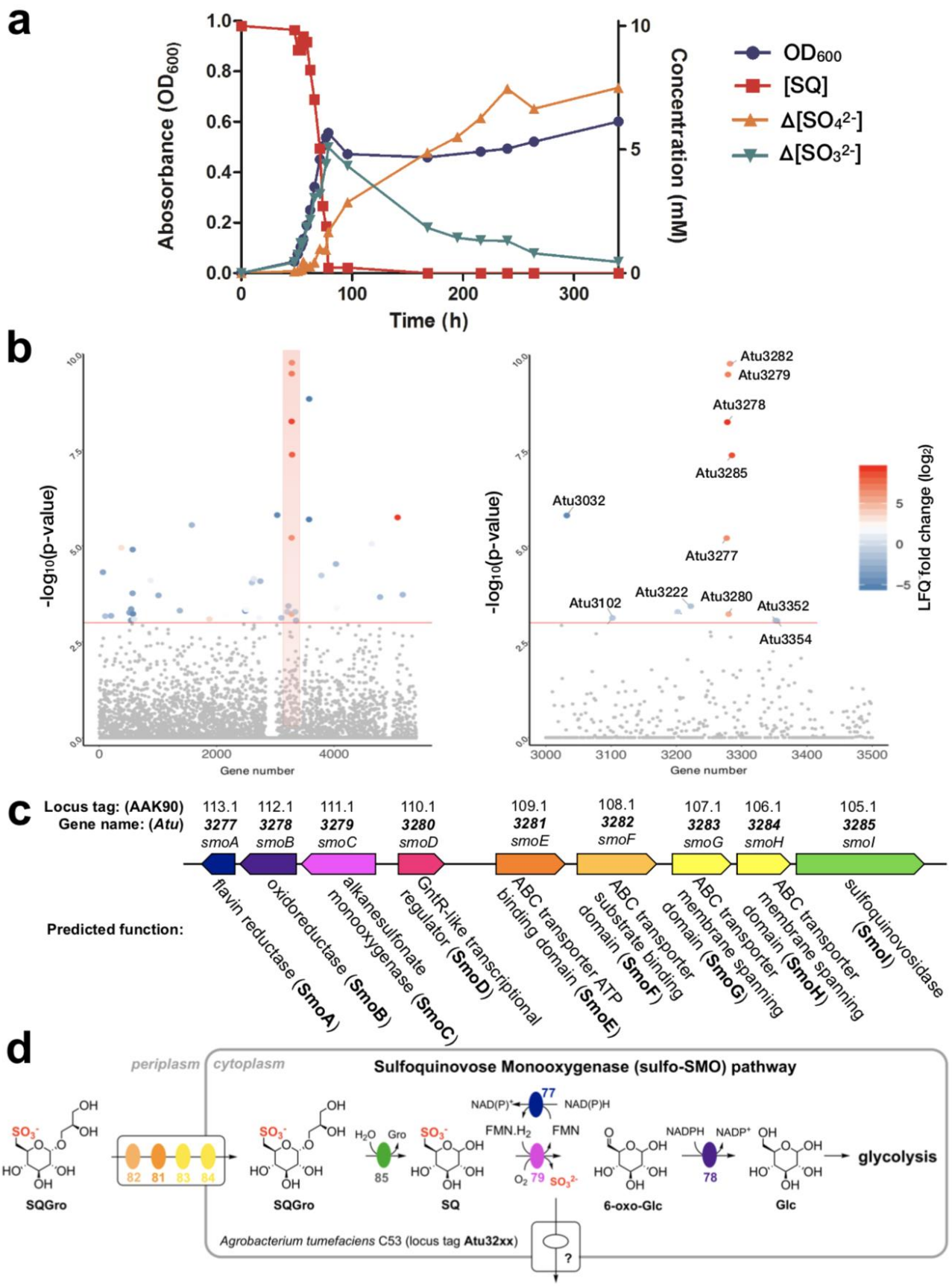

Figure 1: *A. tumefaciens* utilizes SQ and its glycosides as a carbon source. (a) Optical density of *A. tumefaciens* C58 culture (blue) and [SQ] (red), change in [sulfite] (green) and change in [sulfate] (yellow), with respect to time. (b) Manhattan plot of comparative proteomics data for *A. tumefaciens*

C58 grown on SQ vs glucose, demonstrating that the most heavily upregulated proteins belong to a single gene cluster. (c) A cartoon of the upregulated cluster with automated annotations for each of the gene products. These would later be renamed *smoABCDEFGHI*, to reflect the importance of the sulfoquinovose monooxygenase enzyme activity to this new biochemical pathway. (d) A cartoon illustrating the hypothetical roles played by the gene products of this pathway to complete the catabolism of SQGro.

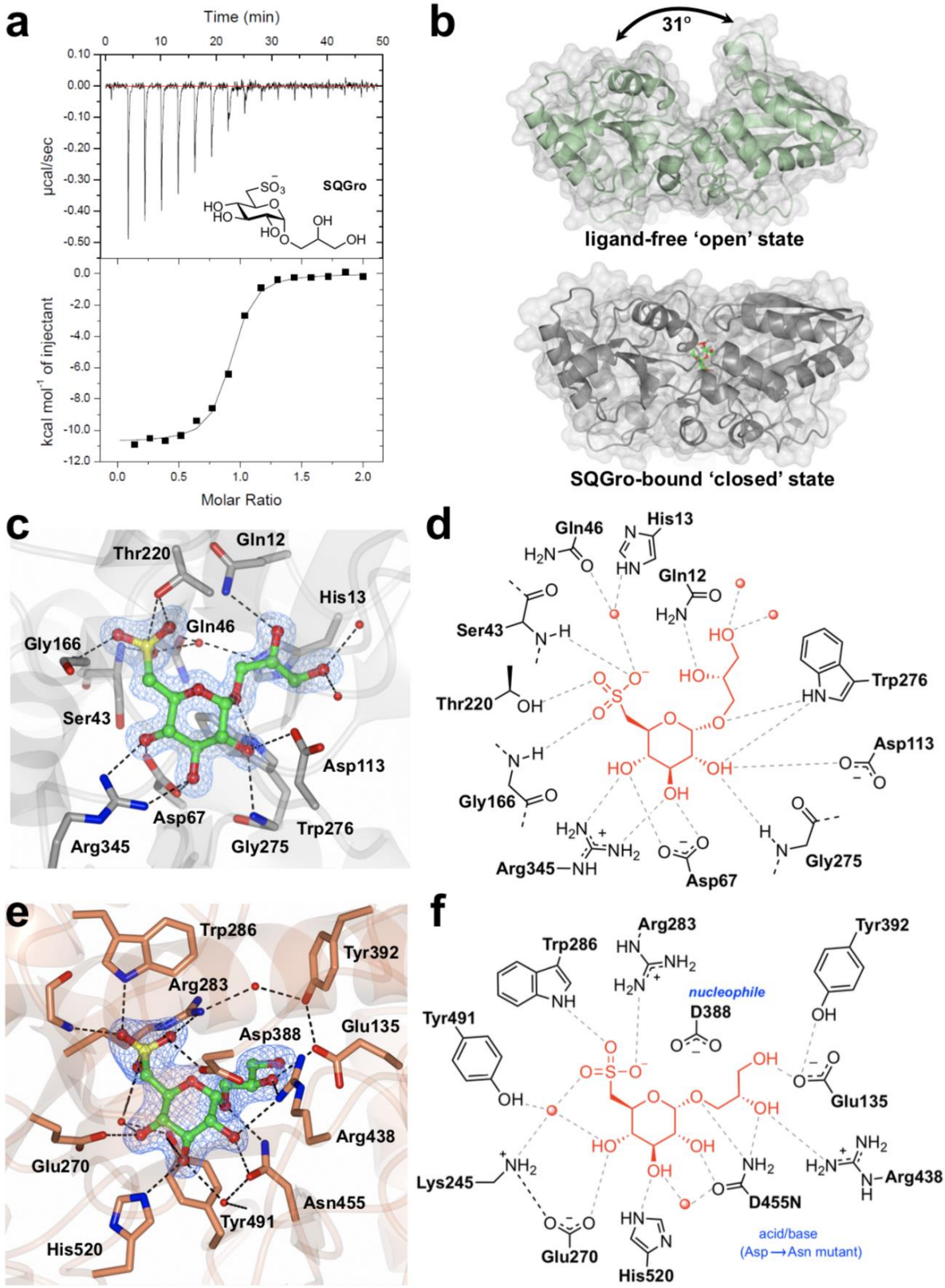

Figure 2: Biochemical and structural analyses of the SQGro-binding protein SmoF (Atu3282) and SQase SmoI (Atu3285). (a) A representative ITC isotherm for SmoF titrated against its cognate ligand 2'*R*-SQGro. (b) Ribbon diagrams (with transparent surface) for the open and closed (liganded) conformations 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) Interactions between protein and ligand within the

SmoF•2'*R*-SQGro complex: SmoF is in grey, 2'*R*-SQGro is in green, and the 2Fo − Fc map at 1.5σ is in blue. (d) A cartoon highlighting key interactions from c. (e) Interactions between protein and ligand within the complex pf SmoI-D455N SQase and 2'*R*-SQGro: SmoI is in gold, 2'*R*-SQGro is in green, and the 2Fo − Fc map at 1.5σ is in blue. (f) A cartoon highlighting key interactions from e: red spheres represent ordered water molecules; dotted lines represent proposed hydrogen bonds.

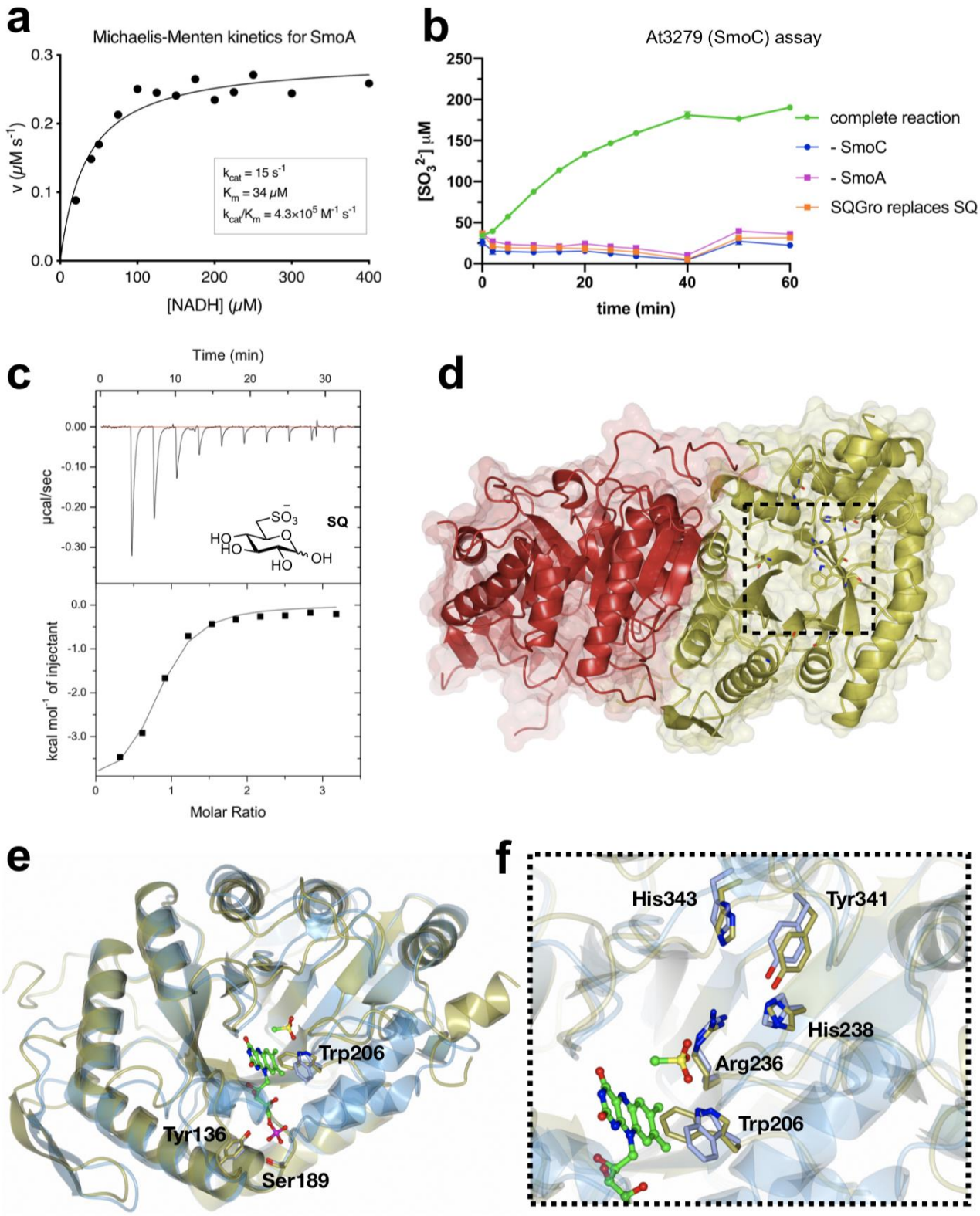

Figure 3: Biochemical and structural analyses of the flavin reductase SmoA and SQ monooxygenase SmoC. (a) Michaelis-Menten kinetics for SmoA-catalysed reduction of FMN by NADH (b) SmoC activity assessed using sulfite release assay with Ellman's reagent in the presence of FMN, flavin reductase, NADH and 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.

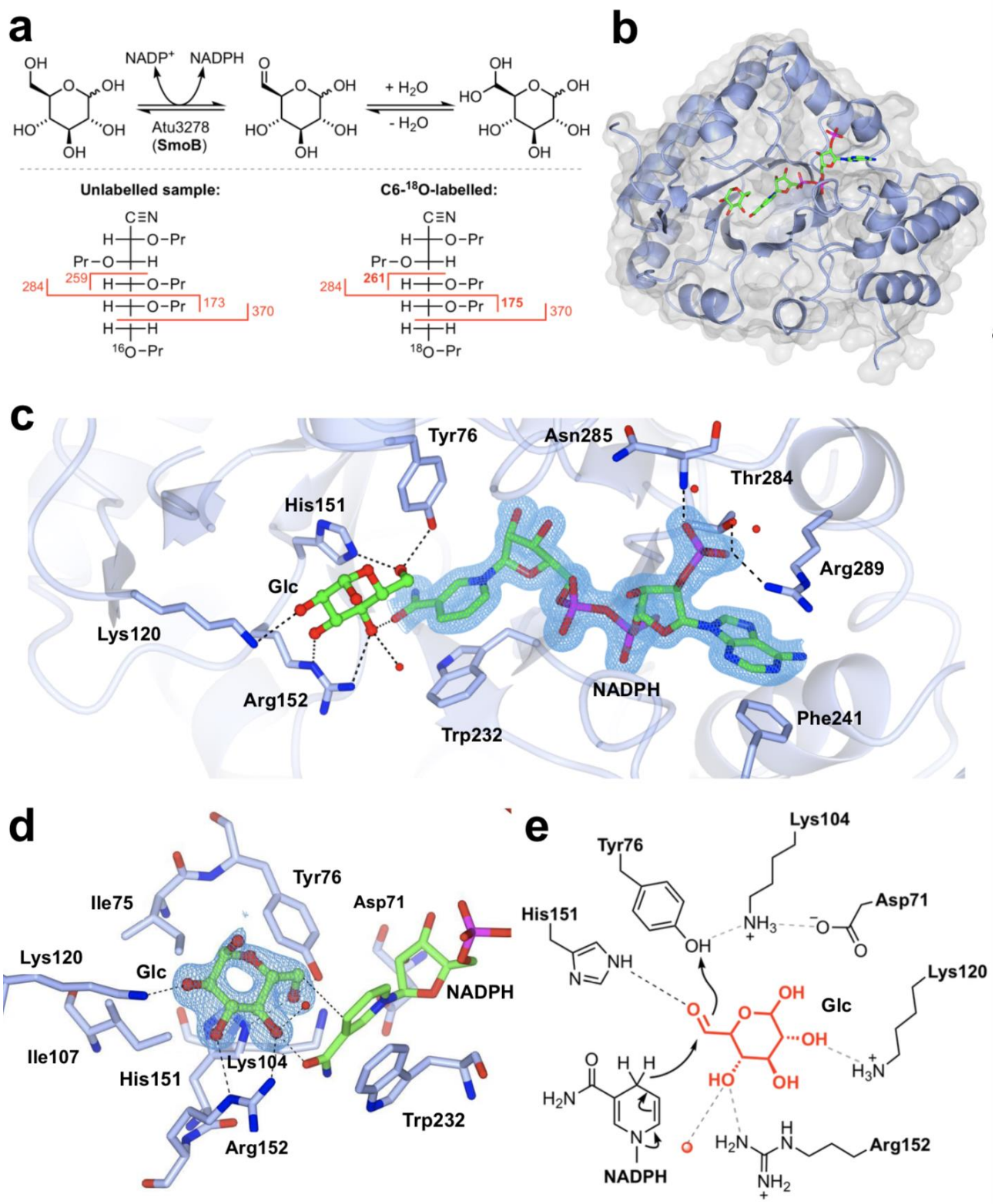

Figure 4: Biochemical and structural analyses of 6-oxo-glucose reductase SmoB. (a) Top: Equilibrium oxygen exchange at C-6 of Glc via 6-OG facilitated by SmoB when incubated with $NADP^+$ in $H_2{}^{18}O$. Bottom: Derivatization and MS fragmentation allows localization of ^{18}O to C6 of Glc. (b) Transparent molecular surface and ribbon diagram of SmoB in complex with NADPH and Glc. (c) Closeup view of SmoB•NADPH•Glc ternary complex. Backbone and carbon atoms of SmoB are shown in ice blue and NADPH and glucose are shown in cylinder format. Electron density for

NADPH corresponds to the 2Fo − Fc map in blue at levels of 1σ. (d) Substrate binding pocket of 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.

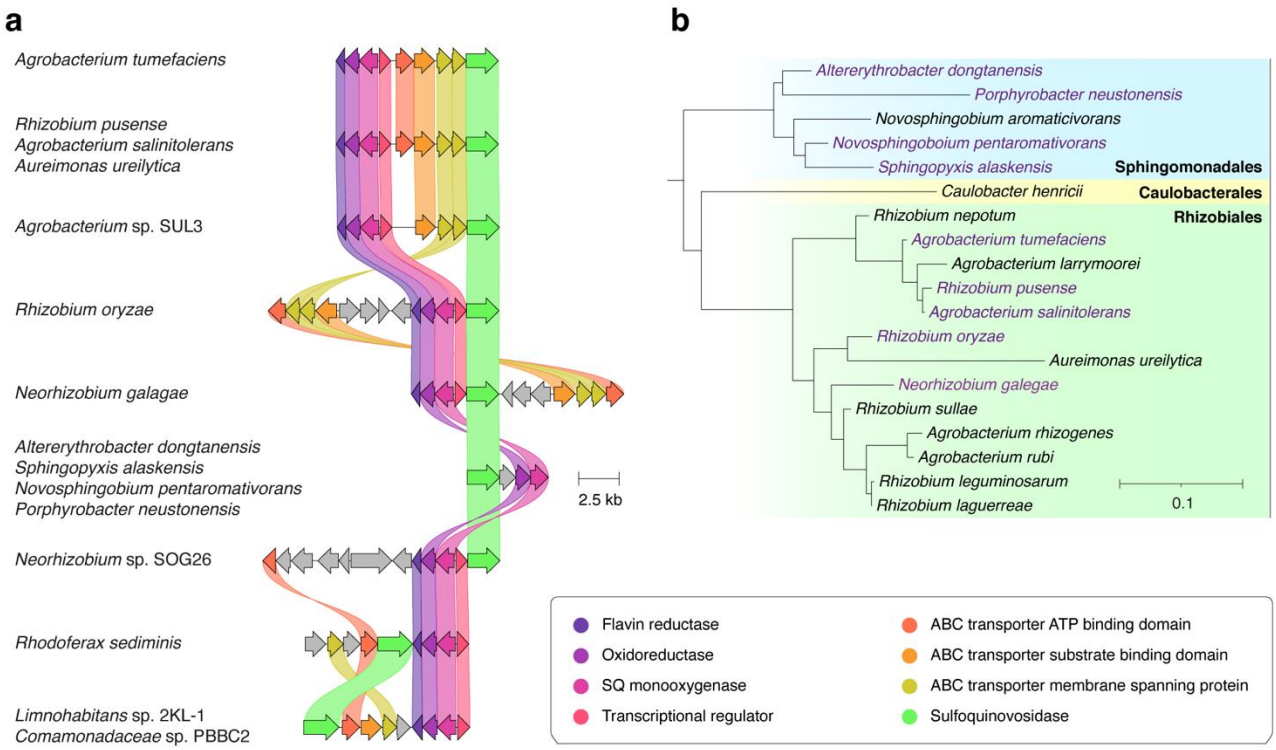

Figure 5: Prevalence of the SMO pathway. (a) Architecture of the SMO gene cluster in *A. tumefaciens* and homologous gene clusters in other organisms. Colored links indicate ≥ 30% protein sequence similarity. Only those clusters encoding putative SQ monoxygenases and SQases were annotated as putative SMO gene clusters. (b) A phylogenetic tree demonstrating the diversity of organisms possessing putative 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/).