
A Structural and Biophysical Dissection of Bacterial Sulfoquinovose Catabolism

Alexander James Dakota Snow



Doctor of Philosophy

University of York

Chemistry

March 2022

I. Acknowledgements

As I sit down, coffee and a biscuit in-hand in a café in Leeds I find myself reflecting on the last three-and-a-half years. While writing this, I'm doing the final pre-submission edits on the document that has dominated my last year and represents the most important single undertaking of my life thus far. *Thesis*. A word suggestive of sweeping statements and grand theories. I feel more like what I've written is some strange mix between literature review, series of publications and a small book. Not so much a loud "here is the answer!" as an "I chipped away at a mountain for a few years, here are the prettier shavings". Pretty shavings they are too, I like to think. Certainly, I've seen the scientific process more clearly and made my own contributions which I am immensely proud of.

I didn't make these contributions alone though. I particularly want to thank my supervisor, Professor Gideon Davies. Gideon, you've been a wonderful mentor who's really allowed me to stretch my independence and tackle the project on my own terms. You've also been very patient with me when that independence led me to make mistakes. Overall, thank you for being far-and-away the best supervisor I could ask for during my Ph.D. I also want to thank the woman who essentially acted as co-supervisor to my project, Dr Mahima Sharma. Mahima, there really aren't the words but I'll try. I owe almost everything I know about how to conduct effective practical research to you, and I would not be in the position I am today without your help. It's been a privilege to learn from you, and I look forward immensely to seeing the directions your own research takes. Thank you! The SQ project wouldn't be what it is without Spencer Williams, Ethan Goddard-Borger and groups. Thank you for setting such a high standard for collaborations. Also, thanks to the Leverhulme Trust for funding this work, and to the University of York for the continued support and guidance.

Thanks to my family. Even with many of you halfway across the world, I've felt your support and love. Special thanks to Rex and Jess.

I also want to thank the technicians, Louise Haigh, Simon Grist, Jules Borgia, Dr Johan Turkenburg and Sam Hart. I've never been so pampered in a lab as I have at YSBL – it's so clear the pride you all take in keeping everything running smoothly and it really shows in the quality of the work outputted. I'm certain that you are all largely the reason YSBL has the reputation it does for quality, collaboration, friendliness, and support. I haven't forgotten you either, Tim

Kirk. Not only an incredible IT specialist without whom the lab would grind to a halt, but a good friend. Additionally, Dr Andrew Leech of TF has been a great help and supportive tutor. Thanks also to Drs Jamie Blaza, Zach Armstrong, and Richard Meek for the assistance with learning cryo-EM. Interesting how a side-project ended up influencing my future career so heavily. Also, many thanks to Professor Tony Wilkinson for always being a kind source of personal and career advice. I hope you'll be as kind in my Viva! Finally, anyone who knows YSBL knows Eleanor Dodson, and I also want to thank you, too. I'll admit to being a little starstruck when I first met you! You've taught me a huge amount directly and indirectly and had a look at more than a few of my nastier X-ray datasets. Also, now I can say I've had my baking complimented by *the* Eleanor Dodson, which might end up being my largest accomplishment in science.

Last but not least, I want to thank Dr Rhianna Rowland, Ben Rowlinson, Wendy Robinson, and Ethan Redmond. You've all been like family to me through good and bad, and while you certainly aren't getting rid of me just like that, I will miss working with and seeing you all every day.

The very last person I want to thank is the reader, whether you are one of my examiners, or someone looking to learn more about sulfosugar catalysis. I hope this document provides the answers you're after. If it doesn't, feel free to track me down and contact me and I'll see if I can help.

Have a lovely day,

Alex

II. Table of Contents

Chapter 1: Background and Introduction	1
2.1. The biogeochemical sulfur cycle.....	1
2.2. The plant sulfolipid, SQDG	2
2.3. SQDG biosynthesis	3
2.4. Role of SQDG in photosynthetic membranes.....	4
2.5. SQDG catabolism: SQDG to SQ.....	6
2.5.1 SQDG lipases	6
2.5.2 Glycoside Hydrolases.....	6
2.6. Sulfoglycolytic pathways	10
2.6.1 The Sulfo-EMP pathway	10
2.6.1.1 SQ isomerase EcYihS.....	11
2.6.1.2 SF kinase EcYihV	13
2.6.1.3 SFP aldolases EcYihT, SeYihT.....	14
2.6.1.4 Transcription factor CsqR	14
2.6.2 The Sulfo-ED pathway	15
2.6.2.1 Sulfo-ED in <i>Pseudomonas putida</i>	15
2.6.2.2 Sulfo-ED in <i>Rhizobium leguminosarum</i>	16
2.6.2.3 SQ mutarotase HsSQM	18
2.6.2.4 SLA reductase PpYihU.....	19
2.6.3 The sulfo-EMP2 pathway.....	19
2.6.4 SQ catabolism in anaerobic conditions	20
2.6.5 The sulfo-TAL pathway	21
2.6.6 The sulfo-TK pathway	23
2.6.6.1 SQ isomerase Sqwl	24
2.6.6.2 SF transketolase SqwGH	24
2.6.6.3 SLA reductase SqwF.....	24
2.6.7 The sulfo-SMO pathway in <i>Agrobacterium tumefaciens</i>	24
2.6.7.1 Flavin mononucleotide reductase SmoA.....	26
2.6.7.2 Flavin mononucleotide-dependent sulfoquinovose monooxygenase SmoC	26
2.7. Carbon retention in sulfoglycolysis.....	27
2.8. Project Aims.....	27
Chapter 2: SQ binding protein SmoF	30

2.9.	Abstract.....	30
2.10.	Introduction.....	31
2.10.1	Substrate binding proteins in nature	31
2.10.2	Classes of ABC transporter	32
2.10.3	Transport cycle of an ABC transporter	32
2.10.3.1	SBP recognition and substrate movement	32
2.10.3.2	Coupling ATP hydrolysis to substrate transport.....	34
2.10.4	Substrate binding proteins	36
2.10.4.1	Substrate recognition and binding.....	36
2.10.4.2	Conformation changes upon SBP-ligand binding.....	38
2.10.5	SBPs and sulfoglycolysis: Aims of this chapter.....	39
2.11.	Materials and Methods	40
2.11.1	Expression and purification of SmoF.....	40
2.11.2	Intact ESI mass spectroscopy.....	41
2.11.3	NanoDSF	41
2.11.4	X-ray crystallography and structure determination.....	42
2.11.5	Isothermal Titration Calorimetry	43
2.11.6	Bioinformatics	44
2.12.	Results	44
2.12.1	Expression, purification and early characterization	44
2.12.2	SmoF binds SQGro, SQDG, SQ and C1 SQ derivatives	45
2.12.3	Structural characterization of SmoF.....	49
2.12.4	SmoF binds C1-methylated and octylated SQ derivatives.....	52
2.12.5	SmoF and SQDG	53
2.12.6	SmoF undergoes conformational changes to accommodate C1 SQ glycosides	53
2.12.7	SmoF bioinformatics.....	56
2.13.	Conclusion and future directions.....	57
Chapter 3: 6-oxo-glucose reductase SmoB.....		63
3.1.	Abstract.....	63
3.2.	Introduction.....	64
3.2.1	Oxidoreductases.....	64
3.2.2	Redox reactions in sulfoglycolysis	65
3.2.3	Aldo-keto reductases.....	66
3.2.4	AKR mechanisms	71
3.2.5	Chapter Aims.....	71
3.3.	Materials and Methods.....	72

3.3.1	Expression and purification of Atu3278	72
3.3.2	SEC-MALLS	73
3.3.3	NanoDSF	73
3.3.4	X-ray crystallography and structure determination	74
3.3.5	Molecular Cloning	75
3.3.6	Isothermal Titration Calorimetry	75
3.3.7	Steady-state kinetics	75
3.4.	Results	76
3.4.1	Gene expression, protein purification and early characterization	76
3.4.2	Binding assays and cofactor specificity	79
3.4.3	SmoB and cofactor recognition.....	81
3.4.4	Structural basis of trimer formation	84
3.4.5	Activity of SmoB and sugar binding.....	87
3.5.	Conclusion and future directions.....	90
Chapter 4: SF transaldolase SqvA		93
4.1.	Abstract	93
4.2.	Introduction.....	94
4.2.1	Aldolases, Transaldolases and Transketolases	94
4.2.2	TALs in sulfoglycolysis.....	101
4.2.3	Substrate recognition and TAL activity	102
4.2.4	Goals of this chapter	102
4.3.	Materials and Methods	103
4.3.1	Gene Expression and protein purification	103
4.3.2	Intact ESI mass spectroscopy	104
4.3.3	SEC-MALLS	104
4.3.4	NanoDSF	105
4.3.5	X-ray crystallography and structure determination.....	105
4.3.6	Cryo-Electron Microscopy and structure determination	105
4.4.	Results	106
4.4.1	Expression, purification, and early characterization	106
4.4.2	A two-pronged approach to structural characterization.....	109
4.4.3	SqvA exhibits the same general fold and extended CTH seen in other decameric TALs	112
4.4.4	Structural basis of SF recognition and insights into SqvA activity.....	114
4.4.5	Structural solutions by cryo-EM and XRD show variations in SqvA fold and conformation	115
4.4.6	Structural basis of decamer formation.....	117

4.4.7	SqvA informatics	120
4.5.	Conclusion and future directions.....	124
5.1.	Impetus and development of this research	127
5.2.	SmoF	127
5.3.	SmoB section	128
5.4.	SqvA	129
5.5.	The contribution of AlphaFold to structural biology	129
5.6.	Future perspectives	131
5.7.	Impact on the wider field	132
5.8.	Closing thoughts	133
6.1	Appendix abstract.....	135
6.2	Materials and Methods.....	136
5.10.1	Gene expression and protein purification	136
5.10.2	Atu3280ct	136
5.10.3	SlaB.....	137
5.10.4	Molecular cloning of BmSlaB	138
5.10.5	ESI peptide-ID mass spectroscopy	138
5.10.6	X-ray crystallography and structure solution.....	139
6.3	Transcription Factor Atu3280.....	140
5.11.1	Introduction	140
5.11.2	Results and discussion.....	143
5.11.3	Conclusions and future perspectives	145
6.4	Aldehyde dehydrogenase SlaB	147
5.12.1	Introduction	147
5.12.2	Results and discussion.....	149
5.12.3	Conclusions and future perspectives	152
7.0.	Amino acid sequences used in this thesis	155
8.0	References.....	157
9.0	Publications Released as part of this work	170

III. List of Figures

Chapter 1

- Figure 1-1 Schematic of the biogeochemical sulfur cycle
- Figure 1-2 Sulfoquinovosyl compounds and photosynthetic membrane lipids
- Figure 1-3 Biosynthesis of SQDG
- Figure 1-4 Crystal structure of PSII from *C.reinhardtii* featuring SQDG
- Figure 1-5 Action of sulfoquinovosidase YihQ
- Figure 1-6 The sulfo-EMP pathway as seen in *E.coli*-K12; a dissection of key enzymes.
- Figure 1-7 The sulfo-ED pathway as seen in *P.putida* SQ1
- Figure 1-8 The sulfo-TAL pathway as seen in *B.megaterium* DSM184
- Figure 1-9 The sulfo-TK pathway as seen in *Clostridium* Sp. MSTE9
- Figure 1-10 The sulfo-SMO pathway as seen in *A.tumefaciens* C58
- Figure 1-11 FMN binding cleft in MsuD and comparison to SmoC

Chapter 2

- Figure 2-1 Transporters and systems employing SBPs
- Figure 2-2 Transport cycle of the MalFGK₂ ABC transporter
- Figure 2-3 Conformation changes and ligand binding pattern of OppA and MalE
- Figure 2-4 Role of SmoF in the sulfo-SMO pathway
- Figure 2-5 Expression and purification of SmoF
- Figure 2-6 Sugars and lipids used in this study
- Figure 2-7 Thermal stability assays of SmoF with ligands
- Figure 2-8 Thermodynamic analysis of SmoF with SQ, SQMe
- Figure 2-9 Crystals of SmoF with ligands
- Figure 2-10 Crystal structure of SmoF•SQ, superposed with MBP•Maltodextrose
- Figure 2-11 Binding mode and accommodation of SmoF
- Figure 2-12 Crystal packing observed in SmoF•SQDG-(C18:1/C16:0)
- Figure 2-13 Ligand-induced conformation changes in SmoF
- Figure 2-14 Large-scale “hinge” motion of SmoF
- Figure 2-15 Sulfoglycolytic pathways containing putative SQBPs
- Figure 2-16 Multiple sequence alignment of SmoF and putative SQBPs

Chapter 3

- Figure 3-1 Oxidoreductases across nature and sulfoglycolysis
- Figure 3-2 Fold and structural basis of cofactor dual-specificity of xylose reductase
- Figure 3-3 Generalised reaction mechanism based on 3 α -HSD for AKRs
- Figure 3-4 Dendrogram of bacterial AKRs, with SmoB included
- Figure 3-5 Expression and purification of SmoB in the pET29a construct
- Figure 3-6 Crystal structure of trimeric SmoB in pET29a construct
- Figure 3-7 Molecular cloning of *Smob* into pET-YSBLiC3C
- Figure 3-8 Expression and purification of SmoB in the pET-YSBLiC3C construct
- Figure 3-9 SEC-MALLS of SmoB
- Figure 3-10 Cofactor and glucose binding assays for SmoB
- Figure 3-11 Determination of cofactor specificity in SmoB
- Figure 3-12 Micrographs of SmoB crystals in the pET-YSBLiC3C construct
- Figure 3-13 Structural overview of SmoB – ligand complexes
- Figure 3-14 Structural basis of NADPH selectivity

Figure 3-15	Structural basis of SmoB trimer formation
Figure 3-16	Structural and kinetic basis for SmoB activity
Chapter 4	
Figure 4-1	Aldolases, TALs, TKs, and the non-oxidative pentose-5-phosphate pathway
Figure 4-2	Active site conservation and multimer formation in classical TALs
Figure 4-3	Sulfoglycolytic pathways featuring TALs
Figure 4-4	Proposed activity for SqvA
Figure 4-5	Expression, purification and solution state of SqvA
Figure 4-6	T _m shift assays of SqvA
Figure 4-7	Crystallography of SqvA
Figure 4-8	3D reconstruction of ligand-free SqvA through <i>Relion</i>
Figure 4-9	Fold and binding site of SqvA by cryo-EM
Figure 4-10	Crystal structure of ligand-free SqvA, and differences relative to EM
Figure 4-11	Structural basis of SqvA decamer formation
Figure 4-12	Bioinformatic analysis of SqvA active triad and sulfonate binding pocket
Figure 4-13	FSC curves for EM structures
Chapter 6	
Figure 6-1	Structure of a representative GntR TF and effector recognition mechanism
Figure 6-2	Conformation changes allowing DNA binding in DasR
Figure 6-3	Construct design, expression, and purification of Atu3280-ct
Figure 6-4	X-ray crystallography of Atu3280-ct
Figure 6-5	Structure, cofactor recognition and mechanism of GapN
Figure 6-6	Initial expression and sub-cloning of SlaB in pET29a
Figure 6-7	Expression and purification for SlaB in pET-YSBLiC3C
Figure 6-8	Crystal structure of SlaB

IV. List of Tables

Table 2-1	Data collection and refinement statistics for SmoF (ligand-free, SQ, SQ-Me)
Table 2-2	Data collection and refinement statistics for SmoF (SQ-Oct, SQDG-(C16:0/C4:0))
Table 3-1	Data collection and refinement statistics for SmoB
Table 3-2	DNA sequences used for amplification of <i>SmoB gene</i> and pET-YSBLiC3C
Table 3-3	Amino acid sequences of <i>SmoB gene</i> in pET29a and pET-YSBLiC3C
Table 4-1	Micrograph collection parameters, reconstruction, and model validation statistics for EM structures of SqvA
Table 4-2	XRD data collection and refinement statistics for ligand-free SqvA
Table 6-1	Data collection and refinement statistics for SlaB

IV. List of Abbreviations

Sugars

6-OG	6-Oxo-Glucose
F6P	Fructose-1-Phosphate
FBP	Fructose-1,6-Bisphosphate
GlcNAc	N-Acetyl Glucosamine
KDSG	2-Keto-3,6-Dideoxy- 6-Sulfoluconate
PNP-SQ	4-Nitrophenyl A-D-6-Sulfoquinovoside
SF	Sulfofructose
SFP	Sulfofructose-1-Phosphate
SL	Sulfolactone
SLA	Sulfolactaldehyde
SQ	Sulfoquinovose
SQDG	Sulfoquinovosyl Diacylglycerol
SQGro	Sulfoquinovosyl Glycerol
SQ-Me	1-Methyl Sulfoquinovose
SQ-Oct	1-Octyl Sulfoquinovose
SR	Sulforhamnose

Cofactors /Small molecules

ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
FMNH	Flavin Mononucleotide
FMNH₂	Flavin Mononucleotide (Reduced)
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
NAD⁺	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide (Reduced)
NADP⁺	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (Reduced)
PNP	4-Nitrophenyl

Pathways

ED	Entner-Doudoroff
EMP	Embden-Meyerhoff Parnas
SMO	Sulfo-Monooxygenase

Enzymes

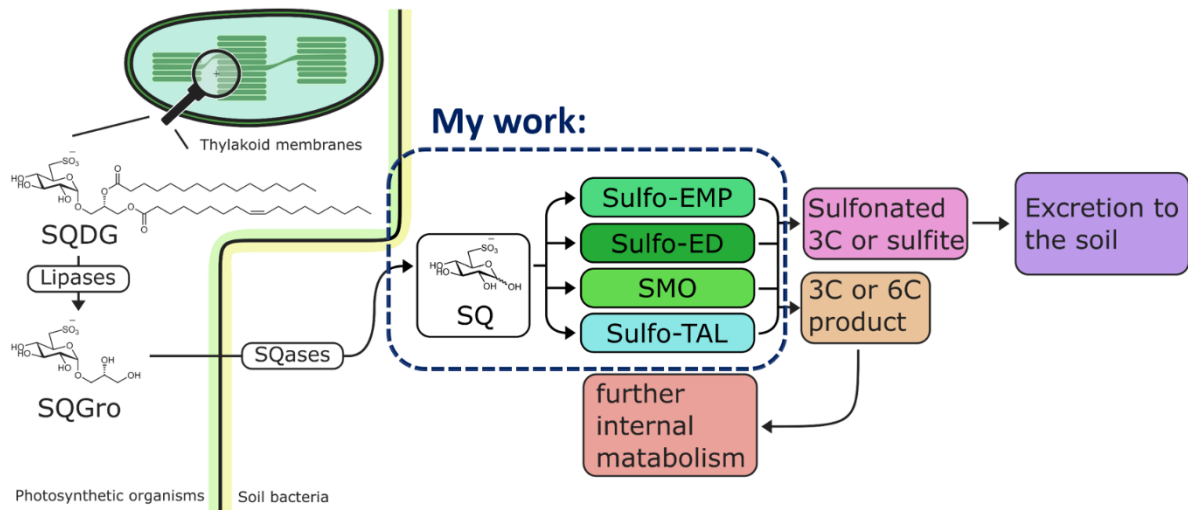
ABC	ATP-Binding Casette
AKR	Aldo-Keto Reductase
ALDH	Aldehyde Dehydrogenase
GH	Glycoside Hydrolase
SBP	Substrate Binding Protein
SQase	Sulfoquinovosidase
SQBP	Sulfoquinovosyl Binding Protein
TAL	Transaldolase

TF	Transcription Factor
TK	Transketolase

Techniques/Technical Concepts

2D	2-Dimensional
3D	3-Dimensional
ASU	Asymmetric Unit
Cryo-EM	Cryo-Electron Microscopy
ESI	Electrospray Ionisation
FSC	Fourier-Shell Correlation
ITC	Isothermal Titration Calorimetry
K_{cat}	Turnover Rate
K_d	Dissociation Constant
K_M	Michaelis Constant
MALLS	Multi-Angle Laser Light Scattering
MS	Mass Spectroscopy
N	Binding Stoichiometry
NanoDSF	Nanoscale Differential Scanning Fluorimetry
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Databank
SDS	Sodium Dodecyl Sulfate
SEC	Size-Exclusion Chromatography
V_{max}	Theoretical Maximum Enzyme Velocity
ΔG	Gibbs Free Energy
ΔH	Enthalpy
ΔS	Entropy

V. Graphical Abstract



VI. Abstract

After the amino acids cysteine and methionine, the largest population of organosulfur in the biosphere is as part of the sugar headgroup of sulfoquinovosyl diacylglycerol (SQDG), which is found ubiquitously in photosynthetic membranes in a poorly-understood role. SQDG is composed of a diacylglycerol moiety linked to the sulfosugar sulfoquinovose (SQ). SQ is liberated from SQDG by stepwise removal of each acyl chain, then glycosidase action to cleave the C1 glycerol moiety. Following this, SQ is broken down through a range of pathways collectively known as “sulfoglycolysis”, of which 5 distinct variants are known. The first two SQ degradation pathways to be characterized mirror the Entner-Doudoroff and Embden-Meyerhof-Parnas glycolytic pathways in *E.coli* and *P.putida* respectively. Both cases they use specific enzymes in an SQ-inducible operon. The third, sulfo-SMO, is found in *A.tumefaciens* and uses an SQ monooxygenase to enact direct desulfurization of SQ, producing glucose as a final product. The fourth and fifth make use of a central transaldolase (sulfo-TAL) and transketolase (sulfo-TK) respectively.

The goal of this PhD is to characterize, both structurally and biophysically, proteins in the sulfo-SMO pathway in *A.tumefaciens* and the sulfo-TAL pathway in *B.megaterium* in order to confirm and analyse their role in SQDG catabolism.

Statement of Authenticity

Except where stated otherwise, all work presented in this thesis was completed by the author. All work not completed by the author was done so within a collaborative environment and is clearly stated as such within the body or declared below. I declare that this thesis is a collection of original work, and that I am the sole author. This work has not previously been presented at this or any other university for any other award.

Work overseen or completed by others:

- The SmoF construct in the pET29a expression vector was kindly provided by the lab of Spencer Williams, University of Melbourne.
- All C1 SQ derivatives were synthesized by the lab of Spencer Williams, University of Melbourne.
- Dr. Mahima Sharma, University of York created constructs for and ordered genes for SmoB, Atu3280ct, Atu3280fl and SlaB.
- Discovery of sulfoglycolytic operons containing SmoF-like binding proteins was performed by the lab of Spencer Williams, University of Melbourne.
- Confirmation of sulfo-SMO enzyme activities by HPLC-MS was performed by the lab of Ethan Goddard-Borger, University of Melbourne.
- The SEC-MALLS experiments were performed and analysed by Dr. Andrew Leech in the York Technology Facility, University of York.
- The protein mass-spectroscopy experiments were performed and analysed by Dr. Adam Dowle in the York Technology Facility, University of York.

1. Chapter 1: Background and Introduction

1.1. The biogeochemical sulfur cycle

Sulfur is one of the most common elements found in the biosphere, comprising $\sim 0.5\%$ of plant and microbial dry weight ¹. It is mostly found in the soil, either as “available sulfur”-sulfonates, sulfate esters (95%), or as part of a bulk mass of organic matter ². Of these only the first is readily accessible to plants. The latter is instead processed into available sulfur by bacteria (fig 1-1). This can then be used by plants, either returning the sulfur to the bulk organic portion or having it directly re-processed by bacteria to produce more available sulfur.

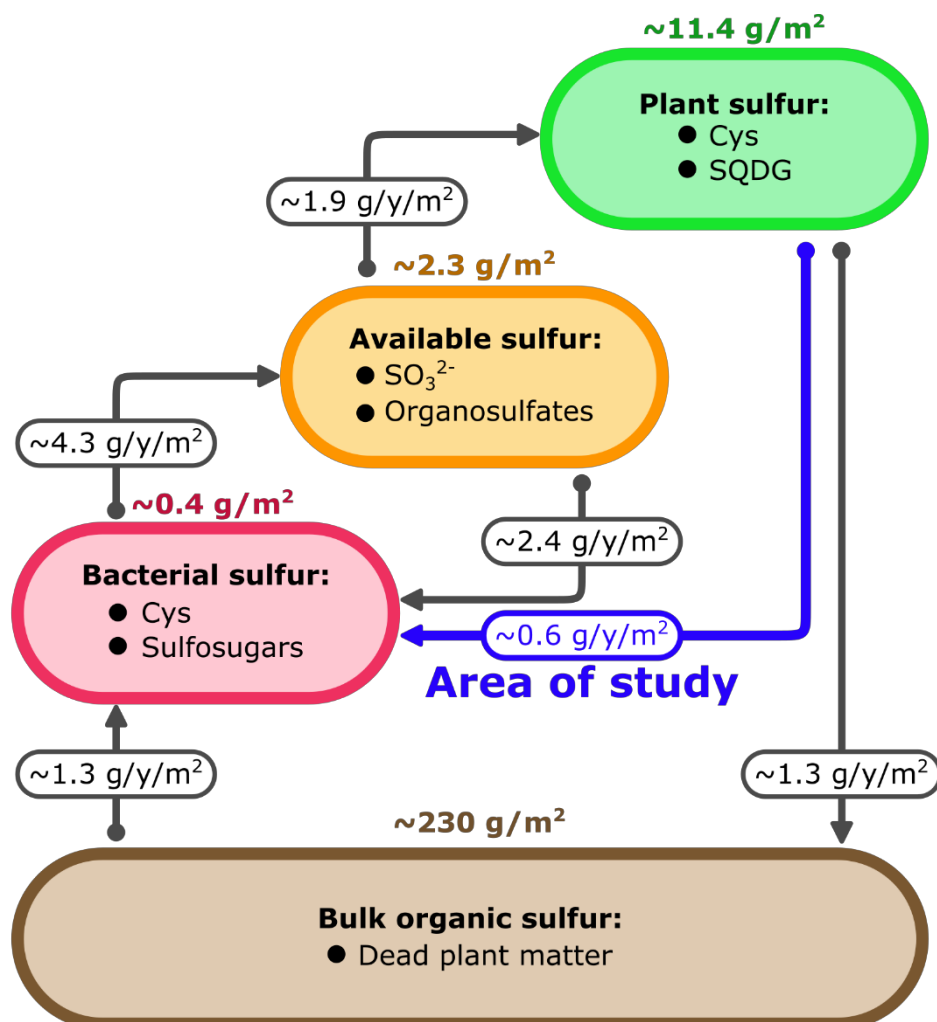


Fig 1-1: Schematic of the biogeochemical sulfur cycle. Values on sulfur pools are estimated sulfur contents of average soil, and values on arrows are estimated sulfur movement per year (Harwood *et al* 1979). The arrow and values shown in blue denote the portion of the biogeochemical sulfur cycle of interest in this work.

This forms the basis of the biogeochemical sulfur cycle. Around 20% of all plant sulfur is cycled either to the bulk or bacterial pools per year³. While all of this eventually returns to the available pool via bacteria, the majority does so indirectly. The majority of plant sulfur is found either in the amino acid cysteine, or as sulfoquinovose (SQ), which forms the headgroup of the sulfolipid sulfoquinovosyl diacylglycerol (SQDG)⁴.

1.2. The plant sulfolipid, SQDG

SQDG is one of the four main lipids found in photosynthetic membranes, with the others being monogalactosyl diacylglycerol (MGDG), diagalactosyl diacylglycerol (DGDG) and phosphatidylglycerol (PG) as well as traces of phosphatidylcholine and 1,2-diacylglycerol-3-O-4'-(*N,N,N*-trimethyl)homoserine (fig 1-2)^{5,6}. While only a minority of these membranes in terrestrial plant leaves is made up of it, at ~4-7%, some marine plants and algae have photosynthetic membranes with SQDG contents as high as 69%^{7,8}. Structurally SQDG can be described as a 6C-sulfonated α -D-hexose coupled to a diacylglycerol (DAG) moiety by a C1 ester linkage. The acyl chains are non-polar and the headgroup polar allowing usage in a lipid bilayer⁴. Lipofoms of SQDG are species-dependent due to the differing methods used to

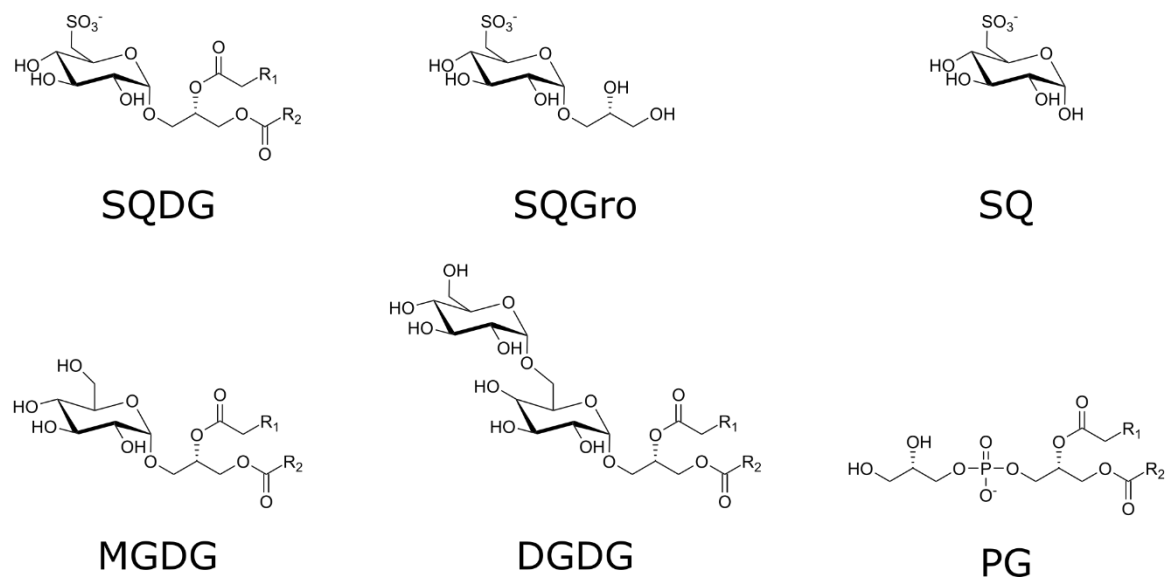


Fig 1-2: Sulfoquinovosyl compounds and photosynthetic membrane lipids. Chemical structures of (left to right, top to bottom) sulfoquinovosyl diacylglycerol (SQDG), sulfoquinovosyl glycerol (SQGro), sulfoquinovose (SQ), monogalactosyl diacylglycerol (MGDG), diagalactosyl diacylglycerol (DGDG) and phosphatidylglycerol (PG).

produce DAG. In *A. variabilis* and many other algae, SQDG and other lipids are made by headgroup modification of a precursor DAG, meaning all lipids have the same lipofoms ⁹. In another algae, *H. carterae*, the lipofoms of SQDG include a palmitoyl and three isomers of hexadeconoyl and eicopentenoyl chains, with varying desaturation locations as confirmed by MS ¹⁰. Plants use two pathways for DAG biosynthesis, resulting in differences in distribution of C16 and C18 chains. The first pathway is found in the plastid and produces mostly C18:1/C16:0 and C16:0/C16:0 lipofoms, reminiscent of those found in cyanobacteria. The second pathway, in the ER produces C16:0/C18:1 and C18:1/C18:1 ¹¹.

1.3. SQDG biosynthesis

While synthesis of DAG is well understood, the specific steps required to produce SQDG were not elucidated until later, with the discovery of a pathway involving the sulfonation of UDP-

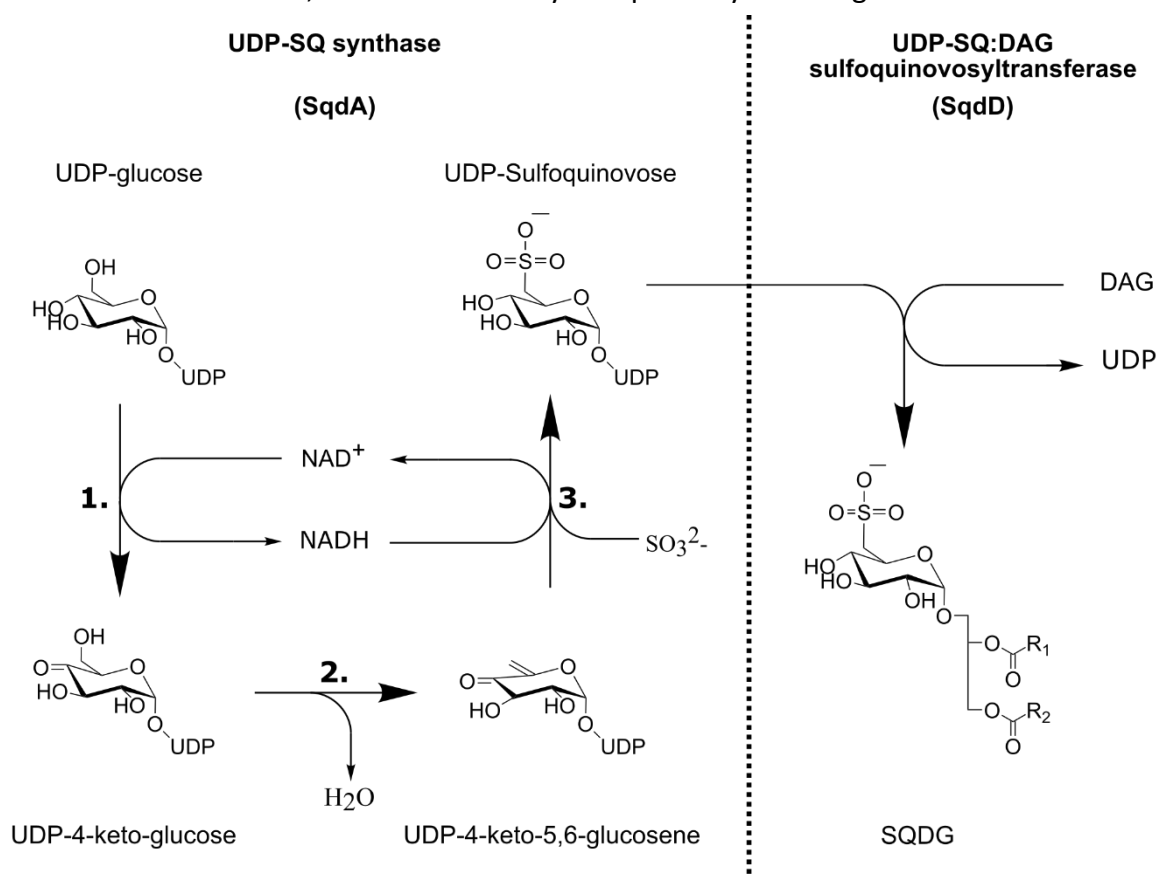


Fig 1-3: Biosynthesis of SQDG. Schematic is split by a dashed line between the actions of SqaA and SqaD. SqaA features a 3-part mechanism wherein UDP glucose is oxidized to UDP-4-keto-glucose (1) then dehydrated to form a UDP-4-keto-5,6-glucosene intermediate (2) to which sulfite can be added, forming UDP-SQ (3). SqaD then transfers SQ from UDP to DAG completing SQDG.

glucose producing UDP-SQ. SQ is then transferred from UDP to DAG, producing SQDG ^{12,13}. The genes encoding the proteins responsible for this mechanism were discovered by knocking out an operon containing the genes *sqdA-D* in *Rhodobacter sphaeroides*, which caused an SQDG deficiency ¹⁴. Knocking out *SqdB* exclusively produced bacteria that were still equally capable of photosynthesis as the wildtype, but had a lowered growth rate ^{15,16}. *SqdA*, the protein product of *sqdA* is a UDP-SQ-synthase, also known as SQD1 in plants such as *Arabidopsis thaliana* and forms the first half of the pathway ¹⁷.

Knockouts of the *SqdD* gene in *R.sphaeroides* were unable to produce any SQDG. This, combined with the sequence-derived function for the gene product, *SqdD* as a GT8 glycosyltransferase served to complete the SQDG biosynthetic route ¹⁸: *SqdB* enacts the sulfonation of UDP-glucose to form UDP-SQ, which is then transferred to DAG using *SqdD*, forming SQDG. The exact function of *SqdC* remains unknown. Knockouts of it result in a 90% reduction in SQDG content, though labelling assays have been used to determine it must be used later than the formation of UDP-SQ. It has been proposed to act in conjunction with *SqdD*, though the nature of this interaction is unclear ¹⁹.

A second family of SQD enzymes was found in cyanobacteria strains *Synchoccus* PCC9742 and *Synechocystis* PCC6803. These encode *SqdX*, the gene for which caused SQDG deficiency when knocked out ^{20,21}. Similarly, a plant homologue, SQD2 has been discovered ²². *SqdX* and SQD2 both fall into glycosyltransferase family 4 and appear to have analogous functions, as either restores SQDG biosynthetic capacity to an *SqdX* knockout of *Synechococcus* ²³.

1.4. Role of SQDG in photosynthetic membranes

Knockouts of the *SqdA-D*, *X* and *2* genes all showed that organisms unable to produce SQDG were still viable, raising the question of what role the sulfolipid plays exactly in these

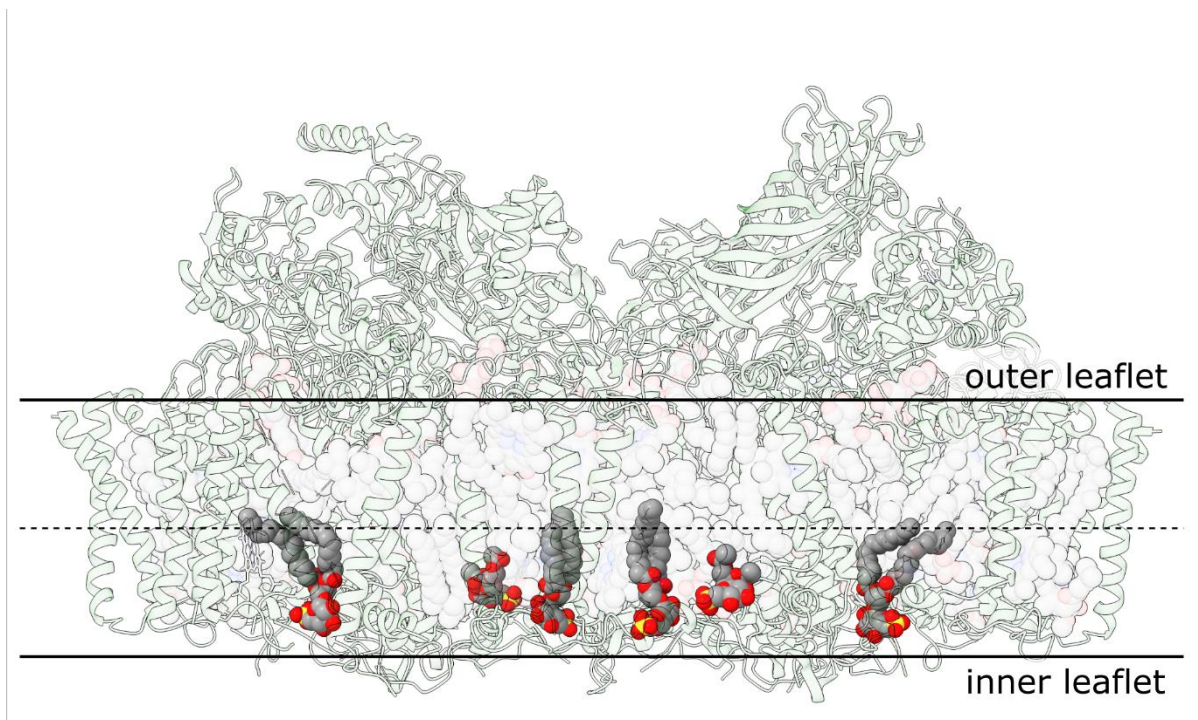


Fig 1-4: Crystal structure of PSII from *C.reinhardtii* featuring SQDG on the inner leaflet. PSII is shown in faded green, and bulk lipids in white. SQDG instances are shaded in grey to highlight them. PDB: 7EDA

membranes. Since SQDG and PG are both anionic it has been suggested that the two share close or even identical roles, distinct from MGDG and DGDG which are thought to have a mainly structural or “bulk” lipid role ²⁴.

PG is known to associate with photosystem II, having been observed in crystal structures of the light-harvesting complex around the trimer interface, though only on the inner leaflet ^{25,26}. MGDG and DGDG were found on both leaflets further consolidating their role as bulk membrane ²⁶. Mutants of *C.reinhardtii* unable to produce PG were less able to form these trimers ²⁷, though this observation is not consistent between all species assessed suggesting a less universal role for PG ²⁸.

Similarly to PG, SQDG deficiency has varying effects. While SQDG-deficient mutants of *R.sphaeroides* showed little change in growth rate or PSII activity, *C.reinhardtii* mutants unable to produce it showed a 40% reduction in PSII activity, which could be restored by addition of the sulfolipid ^{20,29,30}. SQDG may be stabilizing PSII in this case by inhabiting a binding site near the reaction center, as it is difficult to remove from PSII with organic solvents ^{29,31}. These binding sites were also observed in a crystal structure of PSII in *C.reinhardtii* along

with 26 instances of SQDG on the inner leaflet, suggesting a reliance on the sulfolipid even if it is not critical to photosynthetic function ^{32,33}.

The lack of life-critical function for SQDG and chemical similarity between it and PG lead to the idea of SQDG and PG being interchangeable with one acting as a replacement for the other in phosphate or sulfur-limited conditions. Both typically represent a minority of average lipid content; however, the large surface area of these membranes means this interchangeability would allow much less dependence on having both present in the environment in large concentrations. This has been observed experimentally both before and after the discovery of its uses in *C. reinhardtii*, as bacterial mutants lacking PG biosynthetic routes have been able to entirely replace PG with SQDG ^{8,34,35}.

1.5. SQDG catabolism: SQDG to SQ

1.5.1. SQDG lipases

The catabolism of SQDG and subsequent return of sulfur to the available pool is a crucial step in the biogeochemical sulfur cycle. This process begins with the deacylation of SQDG to form sulfoquinovosyl glycerol (SQGro) via a sulfolipase ^{36,37}. Acyl chain removal occurs in two stages, with the first producing a monoacyl intermediate, lyso-SQ. then subsequent action producing SQGro ³⁷. Similar activity was also discovered in plants such as *Phaseolus multiflorus* though attributed to a pair of acyl hydrolases, I and II, with I able to deacylate phosphatidylcholine and oleoglycerol and II able to do the same for oleoglycerol and glycosylglycerides including MGDG and SQDG ³⁸. Glycolipid deacylation in humans occurs in the pancreas using a range of lipases for SQDG, DGDG and MGDG ³⁹. Thus far, no lipase has been discovered in the same operon as any of the downstream sulfoglycolytic machinery suggesting the use of promiscuous bacterial lipases for this purpose ^{40,41}. The product, SQGro is the sulfoquinovose headgroup coupled by a C1 ester linkage to a glycerol moiety. This glycerol can then be cleaved by a sulfoquinovosidase (SQase) ⁴².

1.5.2. Glycoside Hydrolases

Due to the extreme variety in oligo and polysaccharide isomers (as many as 10^{12} are possible just from a single reducing hexasaccharide ⁴³) and the diverse functions these can have within a cell, GHs are generally highly specific to not cause aberrant cleavage. This specificity is

achieved both through their tertiary structure, and residue-level active site topology ⁴⁴. GHs can cleave only the last glycosidic bond in a chain (exo-acting), any position within the chain (endo-acting) or be active on either. Specificity for exo-action is often partially achieved by positioning the active site within a pocket, preventing more than a single monosaccharide from entering. Endo-action or dual-specificity can be achieved with an active site groove or tunnel, which also allows for cleavage of every non-branched glycosidic bond in a chain ⁴⁵.

Anomeric and stereo-specificity is generally achieved through the active site. GHs employ a variety of mechanisms, specific to either α or β anomeric substrates, and with specification about whether the product will share the anomeric state of the substrate or have the opposite one (retaining and inverting GHs respectively) ⁴⁶. Mechanistically these differ considerably. Inverting GHs typically work by a one-step single-displacement mechanism, requiring a general acid and base. Here, nucleophilic attack by a water molecule is aided by general base catalysis. This hydrolysis of the aglycone causes an oxycarbenium ion-like transition state to form, after which the hydrolysed sugar (now the opposite anomer) and leaving group will depart ⁴⁵.

A retaining GH usually uses a classical Koshland retaining mechanism ⁴⁷. This is a two-step double-displacement and forms a glucosyl-enzyme intermediate state. In the first step, a nucleophilic residue (often a glutamate or aspartate) will attack the anomeric centre of the substrate, causing in turn displacement of the aglycone. An oxycarbenium ion-like transition results, then resolves to form the glucosyl-enzyme intermediate. This is known as the glycosylation step. In the second step, deglycosylation, a basic residue deprotonates an attacking water by nucleophilic attack. The resulting hydroxyl group does the same to the anomeric centre causing hydrolysis of the glucosyl-enzyme intermediate. This also occurs with an oxycarbenium ion-like transition state, and results in a cleaved product. This mechanism is identical for both α and β GHs, however the positions of the nucleophilic and acid/base residues are “swapped” to facilitate attack of the anomeric centre in the correct direction ^{44,46}.

Multiple classification systems are relevant for GHs, though the most common is by sequence conservation in the Carbohydrate Active enZyme database (CAZy) ^{45,48–52}. There are nearly

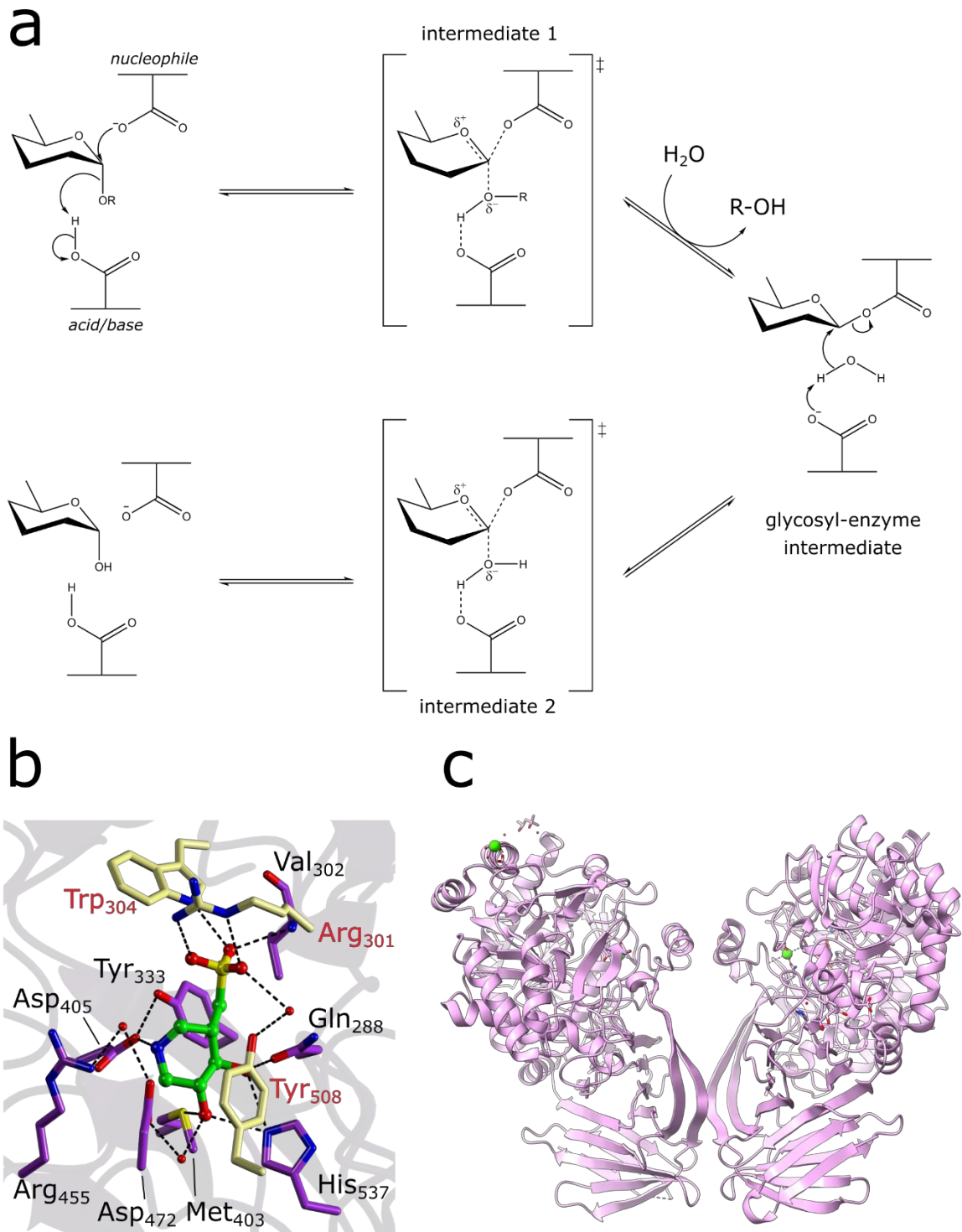


Fig 1-5: Action of sulfoquinovosidase YihQ. A. Generalised retaining glycosidase mechanism. **B.** Active site detail of EcYihQ with mechanistic aza-inhibitor IFGSQ in green. Conserved RWY motif is shown in cream. **C.** Crystal structure of EcYihQ complexed with IFGSQ showing dimeric arrangement. PDB: 5OHT

300 families within this system, though it also includes glycosyltransferases, polysaccharide lyases, carbohydrate esterases and others. In all cases classification into a family is achieved

at 40% sequence conservation. This has not only served to group GHs by their fold, but their activity, and has led to predictions of putative GH activity based on the GH family it inhabits. Other classification systems include the Enzyme Commission (EC) number. Here, enzymes are grouped depending on experimentally confirmed activity. The system makes no distinction of enzyme type or homology, and enzymes with multiple activities can have more than one EC number ⁵¹.

SQGro can be cleaved to form SQ and glycerol by a retaining glycosidase, YihQ (fig 1-5) ⁴². This sulfoquinovosidase (SQase) can be found in family GH31, which also contains other α -glucosidases as well as α -glucan lysases and α -xyalanses ⁵³. The activity of these enzymes can be calculated through usage of 4-nitrophenyl α -D-6-sulfoquinovoside (pNPSQ), producing a K_M of 0.22 mM ⁴².

YihQ is specific to α -SQ, and as a retaining GH exhibits a classical Koshland retaining mechanism (fig 1-5, a) ^{42,47}. Structural characterization with the mechanistic inhibitor 5-fluoro- β -L-idopyranosyl fluoride (5FIdoF) identified D₄₀₅ as a catalytic nucleophile and D₄₇₂ as a likely general acid/base in the *E.coli* protein, EcYihQ (fig 1-5, b). The same structure also revealed a $\alpha_8\beta_8$ barrel fold with a small β domain. YihQ forms a solution dimer, which was also observed in the structure (fig 1-5, c) ⁴².

While YihQ maintains a general fold consistent with other GH31 enzymes ⁵⁴, through the method for substrate binding contains notable differences. In EcYihQ the 6C sulfonate is bound through a salt bridge with Arg₃₀₁, and hydrogen bonds to the indole amine of W₃₀₄ and side-chain hydroxyl of Tyr₅₀₈ via an ordered water (fig 1-5, b). These residues form a “sulfonate binding pocket” and characteristic RWY motif. This triad is heavily conserved across all SQases, and is sometimes accompanied by Gln₂₈₈. This motif was used to identify other SQases in *P.putida* and *A.tumefaciens*. PpYihQ and AtYihQ ^{42,55}. While AtYihQ features the same RWY motif, IFGSQ bound to Glu₂₇₀ in place of Gln, making it an example of the ERWY motif, a common variation seen through bioinformatic analysis ⁵⁵. Otherwise, the ERWY motif forms the same interactions as (Q)RWY. The enzymes further differ through a series of second-shell contacts made to the E or Q residue. EcYihQ forms a neutral pair between Gln₂₈₈ in the active site and Gln₂₆₂ in the second shell, while AtYihQ a similar arrangement forms between Glu₂₇₀ and Lys₂₄₅. Mutagenesis of these QQ and EK motifs revealed residual activity can be preserved only when the entire motif is swapped: QK or EQ mutants were inactive ⁵⁵.

These pairs can be seen in the sequences of many putative SQases, suggesting a strong co-evolutionary history in the pairings ⁵⁵.

1.6. Sulfoglycolytic pathways

In order to complete the conversion of plant sulfur back to the available pool, SQ must undergo further catalysis. Concepts for sulfoglycolysis have existed since initial observations that various *flavobacterium* could grow on methyl- α -sulfoquinovoside, producing 3-sulfolactate in the process ⁵⁶. SL was later observed being excreted by bacteria known to contain SQDG ⁵⁷. These two studies presented the idea that bacteria may be able to process either SQDG or smaller sulfoquinovosides and therefore must have dedicated pathways for this purpose. This was later expanded on, as strains of *Pseudomonas*, *Agrobacterium* and *Klebsiella* extracted from forest leaf mold or sewage sludge were successfully grown on media including SQ as the sole carbon source. This produced SO_3^{2-} for the latter two, but not *Pseudomonas*. *Klebsiella* produced SO_3^{2-} at the same rate as SQ uptake, with sulfopropanediol being observed as an intermediate through ^{13}C -NMR ^{58,59}. *Agrobacterium* also produced SO_3^{2-} in direct response to SQ uptake though no sulfopropanediol was observed. Assaying for glycolytic enzymes in *Klebsiella* cell extract revealed the presence of an enzyme with “SQ kinase” activity, which along with the presence of an NADP^+ dependent dehydrogenase provided early evidence that sulfoglycolysis could occur in a system similar to the traditional Embden-Meyerhof-Parnas (EMP) pathway ^{59,60}.

1.6.1. The Sulfo-EMP pathway

The first conclusive evidence of a sulfoglycolytic EMP pathway (sulfo-EMP) was found in *E.coli* K-12. In combination with *C.pinatubonensis* full biomineralization of sulfur was achieved when grown on media containing SQ as a sole carbon source ⁴⁰. By comparing this to the same system grown on glucose, an operon in K12 was identified as SQ-inducible. This operon contained a predicted α -glucosidase (yihQ), epimerase (yihR), isomerase (yihS), NADH-linked dehydrogenase/reductase (yihU) and a sugar kinase (yihV). When knocked out the capacity of *E.coli* K-12 to grow on SQ was suspended, with no effect on glucose processing, providing

evidence that the sulfo-EMP pathway is entirely separate from traditional glycolysis. To determine the exact role of each enzyme, yihS,T,U and V were all expressed recombinantly and added sequentially to a reaction and the contents assayed by HPLC-MS. YihS converted SQ to a product with an identical Mw but different retention time, indicative of an isomerization to form sulfofructose (SF) as previously predicted⁵⁹. Addition of YihV and ATP produced sulfofructose-1-phosphate (SFP), confirming YihV as an SF kinase. YihT consumed SFP and produced 2 3D products, dihydroxyacetone phosphate (DHAP) and sulfolactaldehyde (SLA), consistent with the predicted aldolase role. This reaction was not as efficient as the others, however the addition of yihU caused a further reduction in the SQ and SFP peaks, as well as the disappearance of the SLA peak and increase in the DHAP peak. A new peak also formed and was identified as DHPS. During this reaction NADH was used, producing NAD⁺ This not only identified yihU as a dehydrogenase but also showed product clearance having an effect on yihT activity, something observed with other aldolases^{40,61}. Overall, these data definitively proved sulfoglycolysis could occur in a manner analogous to EMP glycolysis (fig 1-6, a). Other enzymes in the operon include an SQase, an epimerase and an MFS transporter system⁴⁰. The core enzymes EcYihS, EcYihV and YihT, along with the repressor thought to regulate the operon, CsqR were then subjected to further characterization in order to more fully understand the pathway^{40,62}.

The gene cluster described here was found in >91% of the 1,110 commensal *E.coli* genomes as well as some pathogenic strains available in 2013, found in the integrated microbial genomes and human microbiome project databases. This cluster is therefore thought to be a core-genome feature of *E.coli* in general⁴⁰.

1.6.1.1. SQ isomerase EcYihS

After glycoside hydrolysis of SQGro to form SQ, and mutarotation to ensure formation of α -SQ, isomerization to form SF can commence. Beyond early characterization by HPLC-MS⁴⁰, confirmation of isomerase activity for YihS was performed by proton NMR. This did identify SF as compared to an authentic standard; however, it also identified a second, unknown species. Given the capacity for YihS to interconvert fructose, mannose, and glucose, it was suggested the second product was sulforhamnose (SR), the C2 epimer of SQ. Subsequent synthesis of an authentic SR standard proved this correct. An equilibrium ratio of 30:21:49

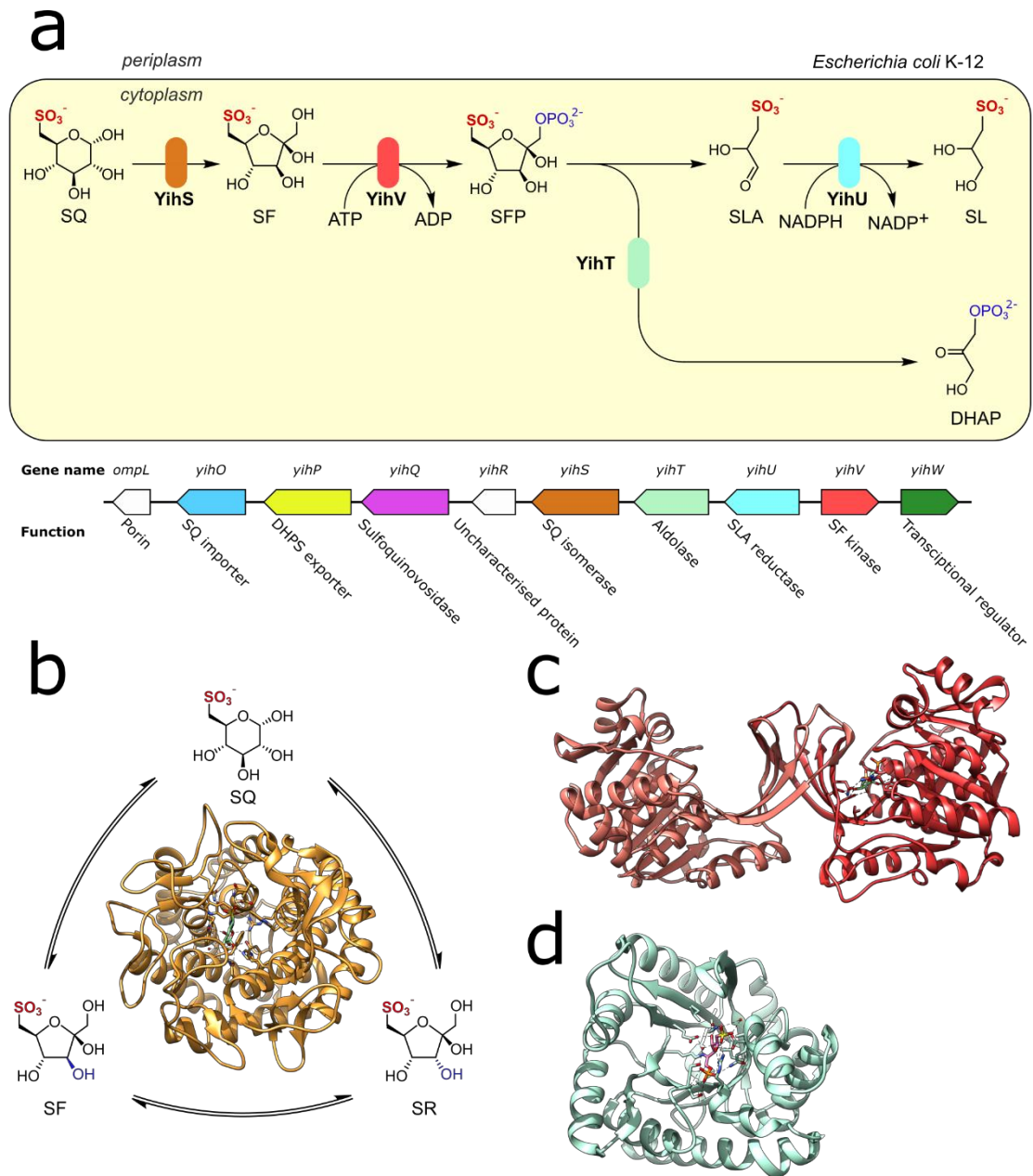


Fig 1-6: The sulfo-EMP pathway as seen in *E.coli*-K12; a dissection of some of the key enzymes.
A. Schematic of the sulfo-EMP pathway wherein SQ is isomerized to SF (orange, YihS), which is then phosphorylated to form SFP (red, YihV). The aldolase YihT (teal) then cleaves this into DHAP and SLA. SLA is finally reduced to SL (YihU, cyan) and excreted. Sulfo-ED operon for *E.coli* is shown below in consistent colours. **B.** Structure and SQ-SF-SR equilibrium of the SQ isomerase YihS. **C.** Crystal structure of dimeric YihV. Deep red monomer is SFP-bound and in the closed conformation, lighter monomer is ligand-free and open. **D.** Crystal structure of SFP aldolase YihT. The structure features a Schiff base with SFP in accordance with its action as a Class I aldolase. PDBs: 7AG4 (YihS), 7AGK (YihV), 7NE2 (YihT)

SF/SR/SQ was then found, showing SR to be a minor product compared to SF, and HPLC-MS was used to determine the subsequent production of SF, SR and use of SQ (fig 1-6, b). Mechanistically this is thought to work by forming a 1,2-enediol intermediate by first forming

acyclic SQ, then protonating C2. From here, C1 protonation makes SF while protonation of C2 produces either SR or SQ, depending on the orientation of the protonation relative to the sugar⁶². YihS activity is 178x more efficient on SQ than mannose showing a clear substrate preference, and proton NMR confirmed the preference of the β anomer of SQ^{62,63}. Activity on the α -anomer cannot be ruled out, though activity seen on α -SQ may be the result of spontaneous mutarotation to form the β anomer. YihS forms a solution hexamer, and X-ray crystallography revealed a TIM barrel fold like that found in GlcNAc epimerases. The structure contained β -SF, which was visible in the active site (fig 1-6, b)⁶².

1.6.1.2. SF kinase *EcYihV*

After isomerization SF is phosphorylated to form SFP by the pfkB-family ribokinase, YihV. As is common in the family, structures of YihV show a characteristic nucleotide-binding domain and a β -domain forming a lid to enclose the substrate binding site^{62,64}. YihV was observed both in the open and closed conformations, which form depending on substrate binding (fig 1-6, c). The closed conformation allows activity as it positions the γ -phosphate of ATP (AMPPNP in the structure) close enough to allow phosphoryl transfer. The nucleotide also coordinates an Mg^{2+} using the α -phosphate. Substrate binding occurs in the cleft between the nucleotide binding domain and the lid, and this site features a sulfonate binding pocket of Arg₁₃₈-Asn₁₀₉-Lys₂₇. The pocket is formed of residues from both domains. SF also binds to Asp₂₄₄ on the C2 hydroxyl, and the C3-4 hydroxyls form hydrogen bonds to Asp₁₃. That only the closed conformation is active is evidence that YihV follows a sequential bi-bi model. Enzyme activity against AF was measured as $k_{cat} = 3.1 \pm 0.2 \text{ s}^{-1}$, $K_M = 1.0 \pm 0.2 \text{ mM}$, and $k_{cat} / K_M = 3.2 \pm 0.8 \text{ mM}^{-1} \text{ s}^{-1}$. The enzyme features comparable levels of allosteric control to other PfkB ribokinases, with sulfo-EMP metabolites SQ, SLA and DHAP all enhancing activity, a trait also shared by traditional EMP metabolites F6P, fructose bisphosphate (FBP) and phosphoenolpyruvate (PEP). From this it seems likely that YihV acts as a central modulation point for the sulfo-EMP pathway similarly to the actions of many other ribokinases^{62,65}.

1.6.1.3. SFP aldolases *EcYihT*, *SeYihT*

Following phosphorylation SFP is cleaved to form SLA and DHAP by YihT. Activity with SFP could not be confirmed in *EcYihT* by HPLC/MS, however an 87% homologue from *Salmonella enterica*, *SeYihT* was active under the same conditions, producing SLA and DHAP ($k_{\text{cat}} = 47.7 \pm 2.4 \text{ s}^{-1}$, $K_M = 3.57 \pm 0.42 \text{ mM}$, and $k_{\text{cat}}/K_M = 13 \pm 2 \text{ mM s}^{-1}$). *SeYihT* belongs to the fructose bisphosphate aldolase family, which can fall into two distinct catalytic mechanisms: Class I aldolases use an active site Lysine to form Schiff base intermediate and can be further classified into class Ia if they form large multimers. Class 2 aldolases are metalloenzymes found in prokaryotes and fungi, and universally use a divalent cation though the metal itself varies.

Structures of *EcYihT* and *SeYihT* were solved by X-ray crystallography. Both enzymes are class I and have a TIM-barrel fold (fig 1-6, d). They both also share structural homology with other bacterial class I aldolases. Structures of *SeYihT* feature a sulfate ion near Lys₁₉₃, thought to be the catalytic base. SFP soaks were then used to produce structures of catalytic intermediates. Here, two monomers in the ASU contained hexoses, thought to be open-chain forms of SFP. The hexoses are covalently bound to Lys₁₉₃, which is acting as a Schiff base (fig 5-6, d). Others showed density for the same, though cleaved at the C3-4 bond, suggesting in-crystal cleavage. These structures feature a sulfonate binding pocket, in which one sulfonate oxygen forms a hydrogen bond to Arg253, and the other two to bound waters. This interaction is identical to that found between the ligand-free protein and sulfate ions. Ser226, Ser227 and Arg253 also form a phosphate binding pocket. The sulfonate binding pocket residues are all conserved, and may serve as a method of classification for these enzymes⁶².

1.6.1.4. Transcription factor *CsqR*

CsqR is a transcription factor, originally found on the same operon as the core sulfo-EMP enzymes⁴⁰. Given the SQ-inducible nature of the operon, it was unsurprising that gel shift assays identified SQ and SQGro as derepressors⁶⁶. Later, SR was found capable of the same action. This answered the questions raised by the action of *YihS*, wherein SQ, SF and SR were all formed in equilibrium⁶².

1.6.2. The Sulfo-ED pathway

Like *E. coli* K-12, a range of *Pseudomonas* strains can grow using SQ as a sole carbon source⁵⁹. In *P. putida* SQ1, SQ consumption was accompanied by the excretion of SLA, which can in turn be used by *P. pantotrophus* NKNCYSA^{41,67}. A comparative proteomics assessment of the *P. putida* genome when grown on SQ versus glucose revealed the upregulation of four genes, all within a single gene cluster^{41,68}. The same cluster contained two more genes, leading to the identification of an operon: A succinate semialdehyde dehydrogenase (PpSQ1_00090), a short chain alcohol dehydrogenase (PpSQ1_00088), a candidate lactonase (PpSQ1_00091), phosphogluconate dehydratase (PpSQ1_00089) and an aldolase (PpSQ1_00100). A mutant lacking the gene for the dehydrogenase PpSQ1_00090 was able to grow on glucose but not SQ, indicating the pathway is required for growth on the sulfosugar and not only induced by it. The composition of the gene cluster led to the idea that the pathway may be related to the Entner-Doudoroff glycolytic pathway, and it was labelled the sulfo-ED pathway. ED glycolytic enzymes were upregulated when the cells were grown on glucose, but not SQ.

1.6.3. Sulfo-ED in *Pseudomonas putida*

From there, genes for the succinate semialdehyde dehydrogenase (PpSQ1_00090) short chain alcohol dehydrogenase (PpSQ1_00088), lactonase (PpSQ1_00091), phosphogluconate dehydratase (PpSQ1_00089) and aldolase (PpSQ1_00100) were all recombinantly expressed in order to better characterize the pathway.

In vitro kinetics from recombinantly expressed PpSQ1_00090 confirmed the enzyme could form SGL using SQ and an NAD(P)⁺ cofactor, though the k_{cat}/K_M was higher with NAD⁺ indicating a possible cofactor preference. The enzyme was also tested with glucose-6-phosphate but this was not active with either cofactor, showing a high degree of specificity to SQ.

PpSQ1_00088 was found active with SLA and NAD(P)⁺ as the substrate and cofactor, with SL forming as a product. Reverse kinetics with PpSQ1_00088 were not possible, an observation seen in other aldehyde dehydrogenases⁶¹.

The other genes of the core pathway (lactonase PpSQ1_00091, dehydratase PpSQ1_00089, aldolase PpSQ1_00100) were also recombinantly expressed, and added sequentially to a reaction to fully reconstitute the pathway. The SQ dehydrogenase PpSQ1_00090 was added first due to the earlier observations, and successfully produced SGL from SQ, confirmed by HPLC-MS. A peak was still present for SG, suggesting spontaneous hydrolysis of SGL to form SG at a low rate. Addition of the SGL lactonase PpSQ1_00091 caused a disappearance of the SGL peak showing activity. The SG dehydratase 00089 gave a product with an identical MS trace as SG but a different retention time, identified as KDSG. The MS trace for SG disappeared. Addition of the aldolase PpSQ1_00100 produced a new peak, identified as SLA and a decrease in KDSG peak intensity. A positive lactate dehydrogenase activity assay suggested pyruvate as a second product, which was confirmed against an authentic pyruvate standard by HPLC co-chromatography. The SLA dehydrogenase, 00088 caused the SLA peak to disappear and showed formation of a new one, identified as SL. Overall this confirmed the *P.putida* SQ degradation pathway as a parallel to classical Entner-Doudoroff glycolysis. The SG dehydratase and KDSG aldolase (PpSQ1_00089, PpSQ1_00100) were inactive with classical ED glycolytic intermediates suggesting a total specificity of these enzymes for the novel pathway (fig 1-7) ⁴¹.

Sulfo-ED pathways also often employ mutarotases at the same stage as those found in sulfo-EMP. These enzymes have a common mechanism though lack sequence homology; A mutarotase from the sulfo-ED pathway in *H. seropedicaea* has only a 50% sequence similarity with that of *P.putida*, and a 37% similarity to the equivalent in *E.coli*, YihR ⁶⁹.

1.6.4. Sulfo-ED in *Rhizobium leguminosarum*

A similar pathway has also been found in *Rhizobium leguminosarum* SRDI565. The operon responsible was previously discovered, though no evidence for its function or up-regulation had been found at the time ². The pathway was revisited, and though it lacks synteny with the equivalent in *P.putida*, it contains the same core genes; a putative SQase, SQ dehydrogenase, SL lactonase, SG dehydratase, KDSG aldolase and SLA dehydrogenase. It also contains an SL exporter, though not an equivalent mutarotase. It also features an ABC transporter system complete with substrate-binding protein. *R.leguminosarum* SRDI565 can grow using SQ as a

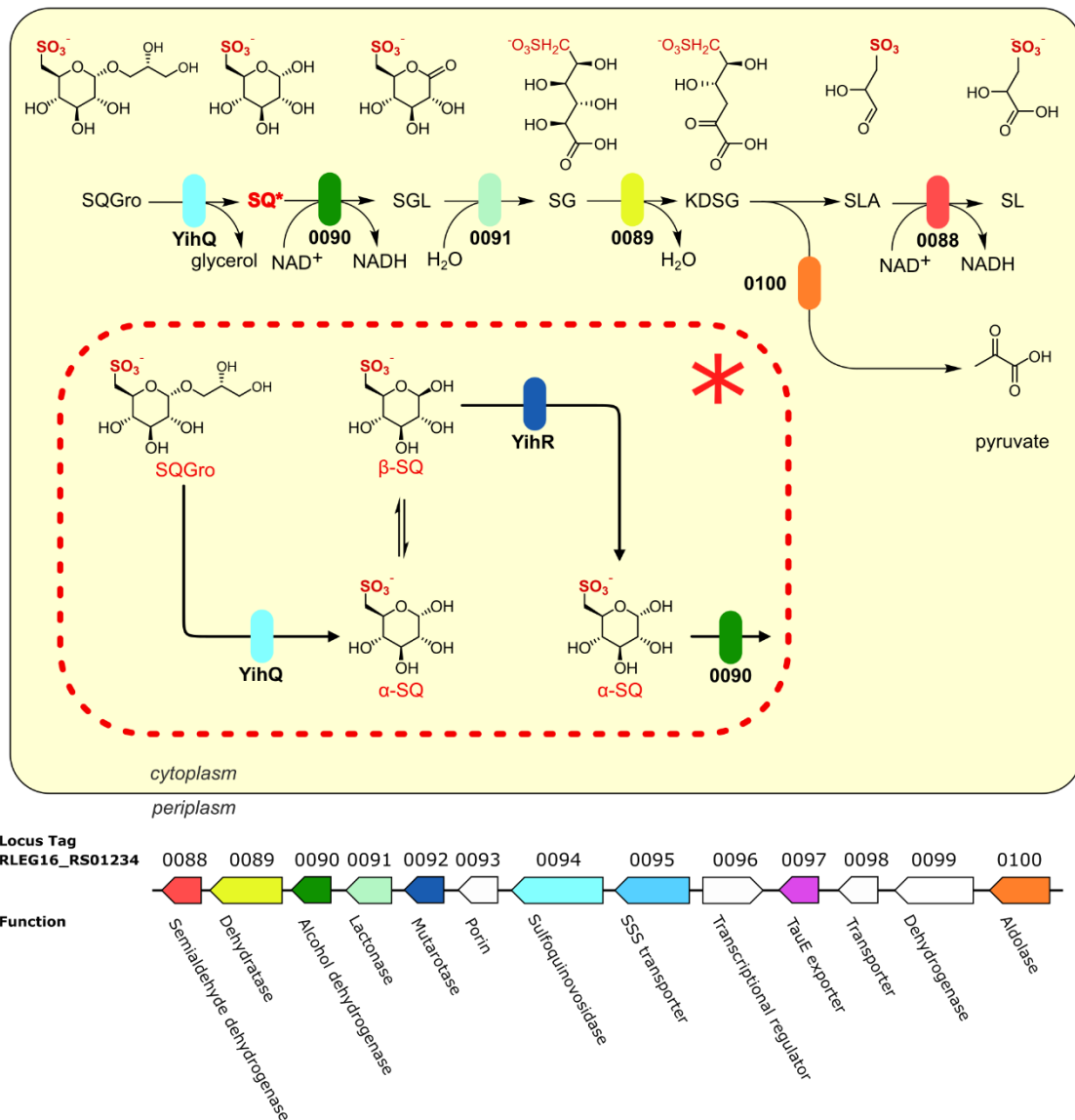


Fig 1-7: The sulfo-ED pathway as seen in *P.putida* SQ1. Here, glycosidase activity of SQGro by PpYihQ (cyan) produces SQ. α -SQ is then reduced by dehydrogenase PpSQ1_0090 (green). Hydration by lactonase PpSQ1_0091 (teal) produces sulfogluconate (SG), which is then dehydrated by dehydratase PpSQ1_0089 (yellow) to form 2-keto-3,6-dideoxy- 6-sulfogluconate (KDSG). Aldolase activity by PpSQ1_0100 (orange) cleaved KDSG into pyruvate and SLA, which is then converted to SL by reductase PpSQ_0088. Sulfo-EMP operon for *P.putida* is shown below in consistent colours. **Inset ***. Action of SQ mutarotase YihR. Glycoside hydrolysis by yihQ yields α -SQ, which can spontaneously mutarotate into β -SQ which cannot be used by dehydrogenase PpSQ1_0090. YihR converts β -SQ to α -SQ to circumvent the issue.

sole carbon source, excreting SL in the process. A proteomic analysis indicated upregulation of the aldolase, SQase and dehydrogenase in the operon with increasing environmental [SQ]. Additionally, an NAD(P)-dependent oxidoreductase, NADH-quinone oxidoreductase subunit

(NuoH, an NADH-dependent succinate-semialdehyde dehydrogenase and a citrate synthase/methylcitrate synthase were also upregulated, suggesting a possible change to the TCA cycle and oxidative phosphorylation with SQ growth. A comparison of the proteome with SQ-grown and mannitol-grown *R.leguminosarum* showed that many of the proteins within the pathway were not specifically upregulated due to the SQ, with only the SQase and KDSG aldolase being specific to the sulfosugar. This may indicate a constitutive upregulation of much of the pathway, which would fall in-line with the oligotrophic nature of the bacteria. The activity of this pathway was further investigated through a similar LC/MS pathway detection assay to those used previously on other pathways. This detected SQ, SF, SG, SL and DHPS as compared to authentic standards, though no SFP or SLA. The SQase was recombinantly expressed, and labelled *R/SQase*. As with other SQases this is a solution dimer, and activity was measured using pNPSQ both as pure protein and in the lysate, with both showing similar activity ⁷⁰.

1.6.4.1. SQ mutarotase HsSQM

A mutarotase is an enzyme that produces a single epimer of a species, allowing downstream enzymes to be specific to it ⁷¹. While the α -SQ produced by YihQ is more stable than the equivalent in traditional ED glycolysis, F6P, it does still undergo spontaneous mutarotation to form the β epimer ^{69,72}. As with *EcYihR* in sulfo-EMP, the *P.putida* mutarotase YihR could not be expressed in sufficient quantities for study ^{40,41}. However, a homologue from *Herbaspirillum seropedicaea* could be ⁶⁹. *H.seropedicaea* is a nitrogen-fixing endophyte found mostly in the intracellular spaces of grasses, and contains a sulfo-ED pathway with the same core enzymes as *P.putida*, though they lack synteny. NMR exchange spectroscopy was used to determine activity of *HsYihR* against SQ, as well as other aldohexoses with an equatorial 2' hydroxyl group. The efficiency however was 17,000x higher with SQ than any other sugar, showing a clear preference, though the enzyme is still more promiscuous than most others in the pathway (fig 1-7, inset) ⁶⁹. Mutarotases often follow a primarily β fold, with two layers of antiparallel β -sheets flanked by smaller α -helices. HsSQM is predicted to follow this fold though structural studies have not been completed ⁶⁹.

1.6.4.2. SLA reductase PpYihU

The final stage of the sulfo-ED pathway is the reduction of SLA to form DHPS, which can then be excreted. This is performed by an NADH-specific reductase, YihU, from the β -hydroxyacid dehydrogenase (β -HAD) family. PpYihU was confirmed active against a racemic mix of SLA with NADH, producing a K_M of 0.082mM, k_{cat} of $5.48 \times 10^2 \text{ s}^{-1}$ with respect to the cofactor. Despite using racemic SLA, it was assumed that PpYihU is specific to D-SLA. PpYihU is also specific to SLA in a similar fashion to other sulfo-ED enzymes as it does not interact with GAP, the equivalent metabolite in classical ED glycolysis. PpYihU follows a similar fold to other β -HAD family enzymes: an N-terminal nucleotide binding domain following a Rossmann fold, and a C-terminal α -helical bundle linked by an α -helix. It also follows the same C-terminal domain-swap dimerization method. A structure with DHPS revealed a dedicated sulfonate binding pocket consisting of Arg-Asn-Ser-Ala as well as an ordered water. The central Arginine in this pocket forms part of the conserved GRT binding motif, which is conserved in the family ^{73,74}.

1.6.5. The sulfo-EMP2 pathway

In *Bacillus urumquiensis*, a bioinformatic approach uncovered a pathway generally analogous to the sulfo-EMP pathway as previously described ^{5,75}. However, *B. urumquiensis* is rare in organisms featuring sulfoglycolytic pathways as it is Gram positive. A pathway assay was assembled in which the identified enzymes were sequentially added to a reaction and the products identified by LC-MS. Upon incubation of SqiD with SQ, a peak was produced corresponding to SF, confirming SqiD as an SQ isomerase. Addition of ATP and SqiK produced a peak matching sulfofructose-6-phosphate, confirming kinase activity. Addition of the putative aldolase SqiA produced a pair of peaks, matching two isomers of DHAP, and SLA. Addition of SlaB from *B. megaterium* and NADH produced SL. Notably none of these enzymes share high sequence identity with known enzymes featured in sulfo-EMP pathways despite identical action ⁷⁵.

1.6.6. SQ catabolism in anaerobic conditions

Thus far all SQ degradative pathways had been described in aerobic conditions, however the same two-member, total degradation model has also been observed as an anaerobic process. *E.coli* K-12 was determined capable of employing the sulfo-EMP pathway this way through mutational studies of the core enzymes and proteomic analyses between SQ and glucose-grown cultures. The second party system, discovered in a *Desulfovibrio* strain, DF1, was extracted from anaerobic sewage sludge. Differential proteomics uncovered a pathway where DHPS is oxidized to 3-sulfolactaldehyde, then 3-sulfoloactate by a pair of NAD^+ - dependent dehydrogenases, DhpA and SlaB. The SL is then cleaved by an SL-sulfate lyase (SuyAB), forming pyruvate and sulfite. The pyruvate is oxidized to form acetate, and the sulfite is used as an electron acceptor in respiration and ultimately reduced to form H_2S ⁷⁶.

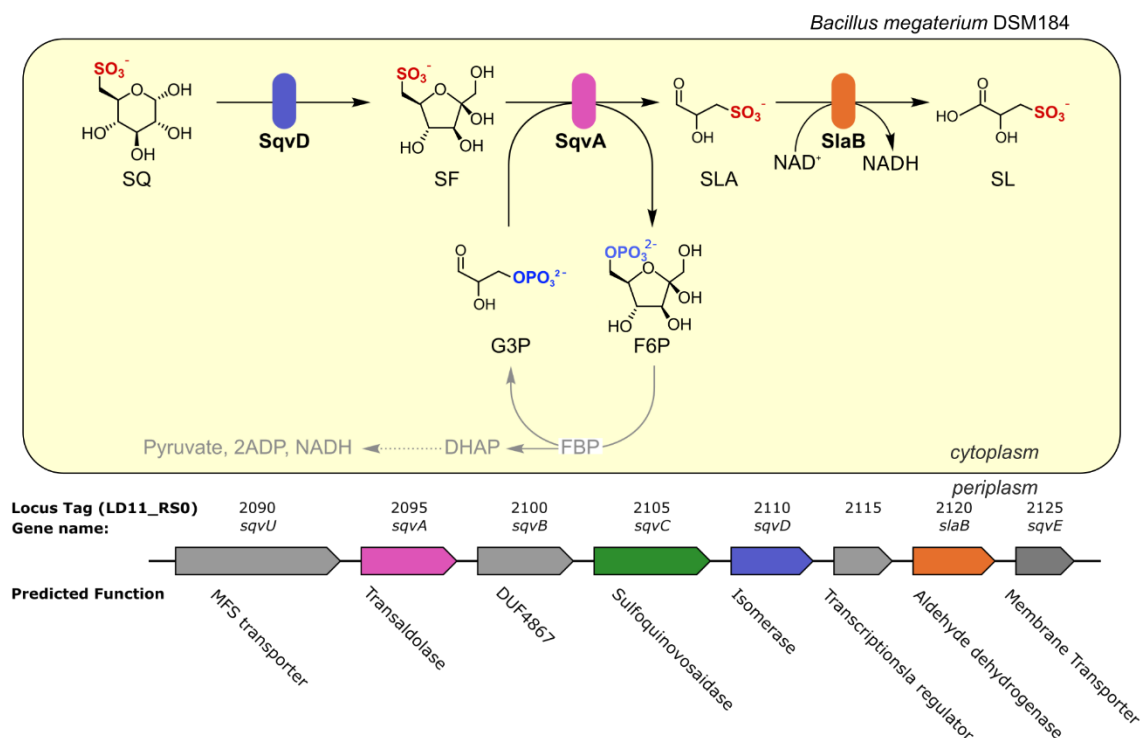


Fig 1-8: The sulfo-TAL pathway as seen in *B.megaterium* DSM184. Here, SQ is isomerized by SqvD (blue) to SF before the sulfo-transaldolase SqvA (purple) transfers dihydroxyacetone from SF to G3P, producing SLA and F6P. F6P can then be cycled through traditional glycolysis to regenerate G3P. SLA is reduced to SL by SlaB (orange). Sulfo-TAL operon for *B.megaterium* is shown below with consistent colours.

1.6.7. The sulfo-TAL pathway

An SQ catabolic pathway was first found in a Gram-positive bacterium in 2020, in *Bacillus aryabhatai*. The pathway was uncovered by differential proteomics of aerobic enrichment cultures grown with SQ as the sole carbon source, and inoculated with soil samples, pond water and plant leaves. The surviving bacteria were mostly identified as strains of *Pseudomonas*, *Rhanella* and *Aeromas*, which were all expected to be using a sulfo-ED or sulfo-EMP pathway. A strain of *Bacillus* however was uncovered from a maple leaf sample that consumed and excreted stoichiometric amounts of SQ and SL respectively. The strain was labelled SOS1 and cell extracts were used to assay for sulfo-ED or EMP enzyme activity. SF was produced and identified by HPLC-MS when SQ was added indicating an SQ isomerase but no ATP was required for any detected step, and no peaks were found for SG or KDSG. This suggested a novel SQ degradative pathway. Sequencing of SOS1 followed by differential proteomics and 2D-PAGE between SQ and glucose grown bacteria found 6 SQ-inducible proteins. When sequenced by MS/MS and located in the genome these formed a cluster, containing Ga0111075_1003_1319 (MFS-type symporter), 1320 (transaldolase), 1321 (unknown function), 1322 (glucoside hydrolase/ α -glucosidase), 1323 (aldose/ketose isomerase), 1324 (GntR-like transcription factor), 1325 (aldehyde dehydrogenase) and 1326 (TauE-type sulfite/organosulfonate exporter). The transaldolase has a 41.7% sequence identity to a characterized fructose-6-phosphate transaldolase from *T.acidophilum*⁷⁷, and a lower homology to a characterized enzyme with purely aldolase activity. This suggests the use of either a glyceraldehyde-3-phosphate (GAP) or erythrose-4-phosphate (E4P) cosubstrate. To better assay this, the cell extract was purified by gel filtration to remove all possible cosubstrates and acceptors. Addition of SQ allowed formation of SF (as confirmed through HPLC-MS) but no further reaction. Subsequent addition of either GAP or E4P caused production of SLA, and SL was produced when NAD⁺ was added. The pathway was expected, therefore, to center around an SF transaldolase which would produce SL from SF, as well as a 6C or 7C product depending on the cosubstrate used. Use of ¹³C-labelled SQ as the carbon source in the extract produced the same products and the products confirmed to be labelled through NMR had the same retention times as those originally seen in HPLC-MS, which in turn had the same as authentic standards. The pathway was then described as follows: An isomerase converts SQ to SF, which is then cleaved to form SLA, with the remaining 3C added

to a cosubstrate to produce a 6 or 7C product. The SLA is then oxidised to form SL using an NAD⁺ dependent dehydrogenase (fig 1-8).

The transaldolase (1320), isomerase (1323) and dehydrogenase (1325) were expressed and purified and added sequentially to a reaction containing ¹³C₆-SQ, with samples taken for HPLC-MS. Addition of 1323 produced a peak with the same mass as SQ but a different retention time and an MS fragmentation pattern matching ¹³C₆-SF. This confirmed the isomerase activity, and the name was changed to SftI. The transaldolase and unlabelled GAP were then added. From this ¹³C₃-SLA and [1,2,3-¹³C₃]-F6P were identified, indicating the non-sulfonated moiety from SF cleavage is added to GAP. When the same reaction is performed with E4P [1,2,3-¹³C₃]S7P is formed showing the same moiety is used. Overall this confirmed 1323 as a transaldolase, which was renamed SftT. Addition of the dehydrogenase (1325) and NAD⁺ showed conversion of ¹³C₃-SLA to ¹³C₃-SL as confirmed through comparison to an authentic SL standard. This would make it another SLA dehydrogenase dependent on NAD⁺, not unlike the one used in *P.putida* sulfo-ED ⁴¹. The enzyme could also oxidise GAP and E4P but not SL or DHPS. Specific activities were unavailable due to a lack of sufficient SLA as this was generated through a reconstituted pathway rather than synthesized. The enzyme was renamed SftD.

Using this pathway as a template 189 candidate clusters were then obtained, mostly in *Bacilli* and *Clostridia*. Overall about 1% of *Bacillus* genomes contained clusters with genes for SftITD, and an α-glucosidase or similar for SQ-glyceride cleavage. This may indicate a use in specific ecologies but that SQ catabolism genes are not a part of the core *Bacillus* genome. All *Clostridia* genomes (a strictly anaerobic bacteria) encoded an NADH-dependent DHPS-forming SLA reductase rather than an SLA dehydrogenase like SftD. This is suggested to allow these bacteria to form DHPS to recover NAD⁺ as seen in the sulfo-EMP pathway ⁷⁸. Additionally, an almost identical pathway was concurrently discovered in *B.megaterium*, containing the genes sqvA (transaldolase), SqvB, SqvC (α-glucosidase). SqvD (isomerase), SqvE (membrane transporter), SqvU (MFS-type importer) and SlaB (aldehyde dehydrogenase), for which a similar amount of characterization was performed ⁷⁹.

1.6.8. The sulfo-TK pathway

In 2021 a bioinformatic approach involving the use of YihQ as a marker of sulfoglycolytic activity revealed a novel pathway using a central transketolase, SqwGH. The same operon also contains a homologue of ribose-5-phosphate isomerase (SqwI), a metal-dependent alcohol dehydrogenase (SqwF), a sulfonate exporter (SqwE) and enzymes SqwB and SqwD. The proposed pathway would have SQ isomerized to SF which is then cleaved by SqwGH. The sulfonated aldehyde product would then be reduced by SqwF and excreted by SqwE. These enzymes were then recombinantly expressed and assembled in a pathway assay to further elucidate their contributions to the pathway (fig 1-89. This pathway is found in diverse anaerobic bacteria, including *Firmicutes*, *Spirochaetes* and *Thermotogae* ⁷⁵.

Clostridium Sp. MSTE9

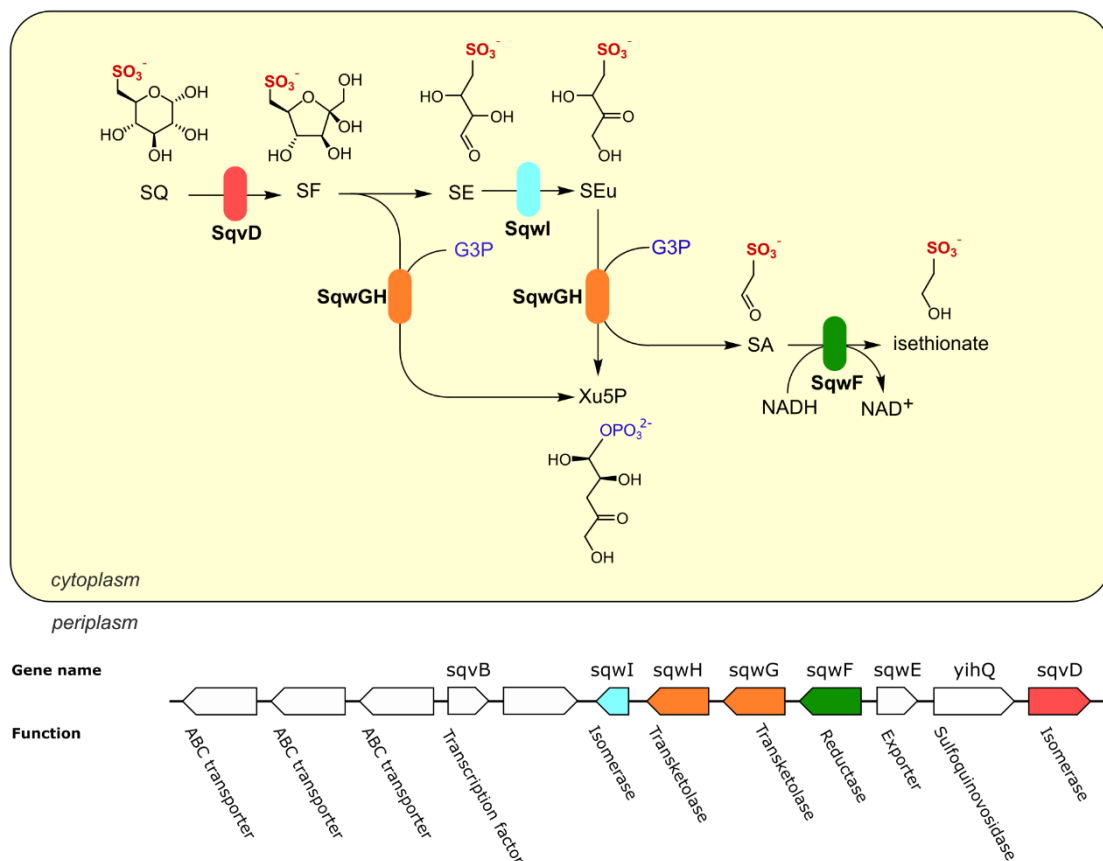


Fig 1-9: The sulfo-TK pathway as seen in *Clostridium* Sp. MSTE9. SQ is isomerized to SF through the SQ isomerase SqvD (red). The sulfo-transketolase SqwGH (orange) then uses a G3P donor to produce xylose-5-phosphate (Xu5P) and sulfoerythrose (SE). The isomerase SqwI (cyan) then converts this to sulfoerythulose (SEu), which SqwGH can also use with G3P to make more Xu5P, as well as sulfoacetaldehyde (SA) which can be reduced by reductase SqwF (green) to isethionate. Below is the gene cluster for the sulfo-TK pathway in *Clostridium*.

1.6.8.1. *SQ isomerase SqwI*

Incubation of SQ with SqwI produced a peak in HPLC-MS matching that of SF as seen in other structures ⁷⁹. SqwI is a part of the ribose-5-phosphate isomerase family, and is thought to convert SE to SEu, which was also seen in the assay ⁷⁵.

1.6.8.2. *SF transketolase SqwGH*

The second experiment was with SqvD, SqwGH, and added the ketol acceptor G3P. The expectation was that SqwGH would convert SF into 4-deoxy-4-sulfoerythrose (SE) and xyulose-5-phosphate with G3P. The formation of xyulose was confirmed. It is possible that SqwGH transfers a pair of C2 ketol groups rather than one, acting successively. This was suggested as a peak corresponding to sulfolactaldehyde was also found. This may work mechanistically with production of SE as before, which spontaneously isomerizes to 4-deoxy-4-sulfoerythulose (SEu) before having a second ketol added forming sulfolactaldehyde ⁷⁵.

1.6.8.3. *SLA reductase SqwF*

SqwF is a metal-dependent alcohol dehydrogenase and is related to the NADH SLA reductase TauF. Inclusion of the enzyme and NADH in the pathway assay led to the disappearance of sulfolactaldehyde and the appearance of isethionate. Therefore, SqwI is an NADH-dependent sulfolactaldehyde reductase. The reverse reaction is also possible ⁷⁵.

1.6.9. *The sulfo-SMO pathway in Agrobacterium tumefaciens*

The pathways described until now all result in production of a C3 sulfonate, namely DHPS or SL, which is excreted in addition to the desulfurised C3 product used by the host cell. In this way SQ catabolism has always been a two-party process requiring downstream organisms for complete catalysis. The sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway, as seen in *Agrobacterium tumefaciens* C58, is unlike the others as it achieves desulfurization while retaining a 6C produce, glucose ⁸⁰. Therefore, there is no need for a two-party system. C58 can grow on a minimal media containing SQ as a sole carbon source, in the process releasing sulfite and bicarbonate. Comparative proteomics showed SQ-dependent upregulation of a gene cassette containing the previously characterized AtSQase, which was

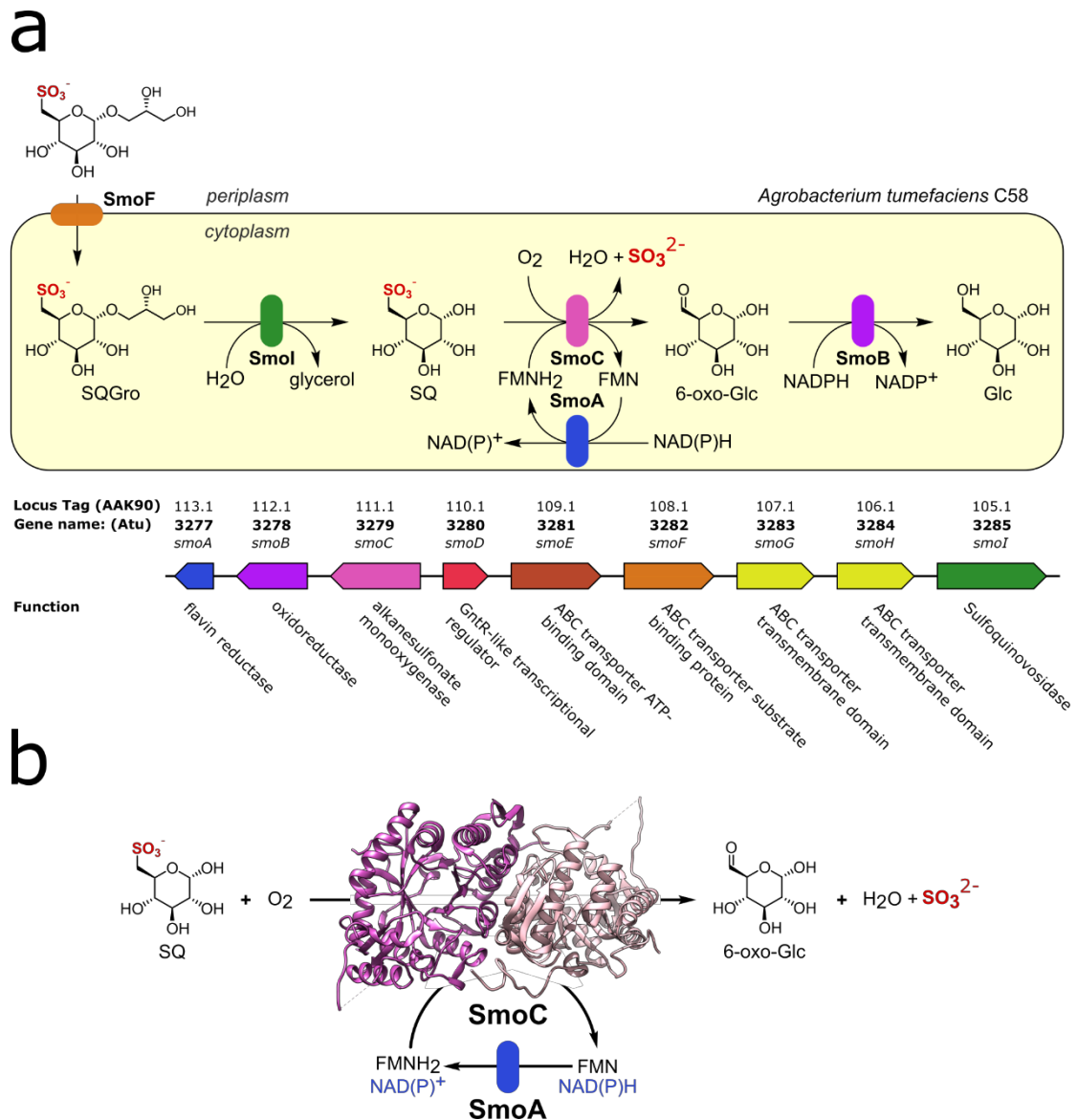


Fig 1-10: The sulfo-SMO pathway as seen in *A.tumefaciens* C58. A. Schematic of the sulfo-SMO pathway. SQGro is imported through ABC transporter SmoGHE₂ and substrate-binding protein SmoF (orange). SQase Smol (green) then removes glycerol forming SQ. Oxidative desulfurization of SQ then commences through monooxygenase SmoC (purple), using oxygen and producing 6-oxo-glucose, sulfate and water, and converting FMNH₂ to FMN. The cofactor can then be regenerated by SmoA (blue) using NAD(P)H. 6-oxo-glucose is then reduced to glucose by aldo-keto reductase SmoB (purple). Below is the gene cluster for the sulfo-SMO pathway in *A.tumefaciens* C58. **B.** Reaction and structural overview of SQ monooxygenase SmoC. RoSmoC dimer shown in light and dark purples. PDB: 7OH2

then re-labelled as SmoI. It also contained genes for an ATP binding cassette (ABC) transporter with associated substrate binding protein (SmoEGH, and SmoF), an NAD(P)H-dependent SQ monooxygenase and flavin mononucleotide reductase (SmoC and SmoA) and an aldo-keto

reductase (SmoB) (fig 1-10, a) ⁸⁰. The same pathway was uncovered through a similar bioinformatic method simultaneously by Liu *et al.* ⁷⁵.

1.6.9.1. *Flavin mononucleotide reductase SmoA*

SmoA and SmoC are both present in gene cassettes in *Agrobacterium* sp., *Rhizobium oryzae* and *Auremonas flava*, which combined with the predicted function of both enzymes suggested a two-component system in which the desulfurization of SQ is achieved by its oxidation using a flavin cofactor, which is subsequently reduced to allow further activity. The flavin reductase SmoA is similar to SsuE, as found in the alkylsulfonate gene cluster *ssuEADCB* ⁸¹. Crystal structures of SsuE exist both ligand-free and FMN bound and combined with solution state characterization show an equilibrium between dimeric and tetrameric states forms with FMN binding. This may allow association with the monooxygenase of that system, SsuD allowing more efficient flavin transfer ^{82,83}. The ability of the SmoA-SmoC system to undergo the same action is currently unknown. SmoA uses NADH as a cofactor to enable hydride ion transfer ^{80,84}.

1.6.9.2. *Flavin mononucleotide-dependent sulfoquinovose monooxygenase SmoC*

SmoC is a sulfoquinovose monooxygenase that binds SQ specifically, with a K_D of 3 μ M. Upon binding SQ is oxidised producing 6-oxo-glucose and sulfite using an FMNH cofactor (fig 1-10, b). Based on the activity of similar monooxygenases SsuD, LadA and MsuD it is likely this mechanism involves either the production of a C4a-(hydro)peroxyflavin or N5-peroxyflavin intermediate ⁸⁵⁻⁸⁷. As a category II two-component flavoprotein monooxygenase SmoC must be reduced by SmoA prior to activity. While structures of *AtSmoC* exist, they are of low resolution. They do however share high structural similarity to a homologue in *R.oryzae* which has a higher resolution structure. Superposition to a crystal structure of MsuD complexed with FMNH₂ revealed the two share a common cofactor and substrate recognition mechanism (fig 1-11). A hydrophobic cleft present in both structures accommodates the isoalloxazine ring, and a sulfonate binding pocket comprised of Trp₂₀₆, Arg₂₃₆, His₂₃₈ and His₃₄₃. In SsuD, substrate binding induces a conformational change in pairs of conserved arginines and glutamates distal to the active site. Possible mechanisms for SmoC activity include the formation of either a C4a-peroxy or N5-peroxyflavin intermediate, followed by deprotonation

of the substrate C6 by the peroxide. The carbanion formed then is oxidised to α -hydroxysulfonate which undergoes elimination, forming sulfite and 6-oxo-glucose. Another possible mechanism is that the sulfonate sulfur is attacked by the terminal peroxide oxygen of the flavin intermediate. This then causes C-S bond cleavage resulting in release of sulfite and 6-oxo-glucose⁸⁰.

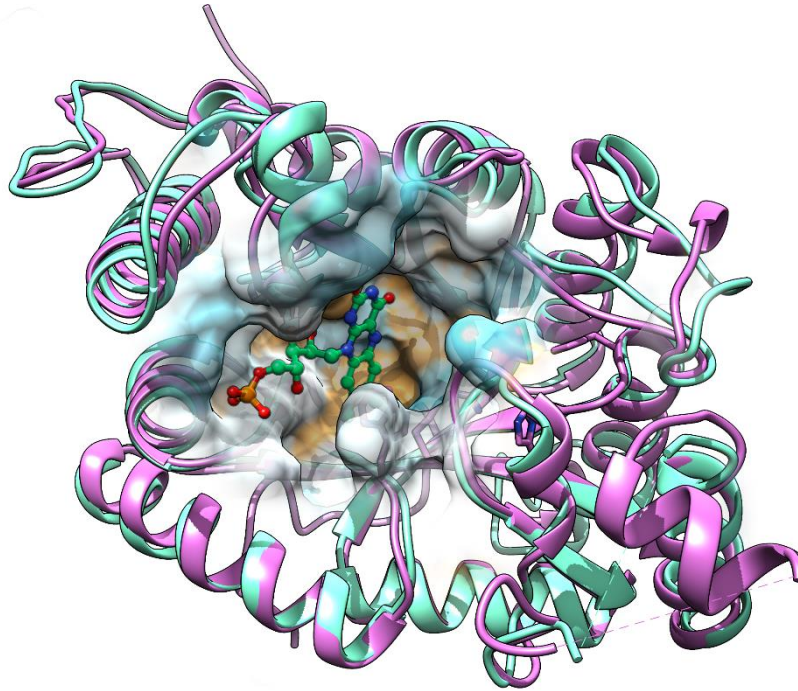


Fig 1-11: FMN binding cleft in MsuD and comparison to SmoC. Superposition of SmoC (purple) and MsuD (cyan). FMN from MsuD structure is depicted within a surface view of the cleft, coloured by residue hydrophobicity (yellow = non-polar, blue = polar). PDBs: 7OH2 (SmoC), 7JYB (MsuD).

1.7. Carbon retention in sulfoglycolysis

The sulfoglycolytic pathways so far characterized can be grouped based on their method of desulfurization. The most common approach is the cleavage of 6C SQ into a 3C species to be used by the cell, and a 3C sulfonate which is then typically reduced and excreted^{40,41,70,75,78,79}. Compared to glucose, this provides approximately half the energy per molecule making these pathways about 50% as efficient as traditional glycolysis. This can be observed as many of the organisms containing a sulfo-ED operon grow using SQ as a sole carbon source at about half the rate observed using glucose⁶⁷. The 3C sulfonate can then be used by other bacteria such as *C. pinatubonensis* allowing total biomineralization of sulfur^{67,84}. The second group are

those pathways that produce sulfate through desulfurization, retaining the entire 6C for internal use. Thus far the only example of this is the sulfo-SMO pathway, which produces glucose as an end product ^{75,80}.

1.8. Project Aims

Structural characterization of enzymes necessarily occurs typically after early biophysical characterization. Therefore, many enzymes that have been described thus far lack structures and detailed biophysical characterization. The primary goal of this project was to create a more detailed profile of the key enzymes in these pathways, with a focus on structural biology. Obtaining structures with and without key substrates and cofactors and supplementing these observations with biophysical data such as binding affinities, enzyme kinetics, substrate specificities and multimeric states could then allow for a detailed examination of how the enzyme in question functions and the greater role it plays in the pathway and organism it comes from. These analyses were performed on SmoF and SmoB from the *A.tumefaciens* sulfo-SMO pathway and SqvA from the *B.megaterium* sulfo-TAL pathway, with less detailed analyses performed on Atu3280 from *A.tumefaciens* (sulfo-SMO) and SlaB from *B.megaterium* (sulfo-TAL). These are laid out in the following fashion:

Chapter 2. SQ binding protein SmoF

Chapter 3. 6-oxo-glucose reductase SmoB

Chapter 4. Sulfofructose transaldolase SqvA

Chapter 5. Conclusions and future perspectives

Chapter 6. Appendix: GntR-like transcription factor Atu3280, SLA dehydrogenase SlaB

Chapter 7. References

This is then further supplemented by the publications released as a part of this project:

- Li, J., Epa, R., Scott, N. E., Skoneczny, D., Sharma, M., Snow, A. J. D., ... Williams, S. J. (2020). A sulfoglycolytic entner-doudoroff pathway in *Rhizobium*

leguminosarum bv. trifolii SRDI565. *Applied and Environmental Microbiology*, 86(15). <https://doi.org/10.1128/AEM.00750-20>

- Snow, A. J. D., Burchill, L., Sharma, M., Davies, G. J., & Williams, S. J. (2021). Sulfoglycolysis: catabolic pathways for metabolism of sulfoquinovose. *Chemical Society Reviews*, 50(24), 13628–13645. Retrieved from <https://pubs.rsc.org/en/content/articlehtml/2021/cs/d1cs00846c>
- Sharma, M., Lingford, J. P., Petricevic, M., Snow, A. J. D., Zhang, Y., Järvå, M. A., ... Goddard-Borger, E. D. (2022). Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proceedings of the National Academy of Sciences*, 119(4), e2116022119. <https://doi.org/10.2185/jjrm.23.202>
- Snow, A. J. D., Sharma, M., Lingford, J. P., Zhang, Y., W.-Y.Mui, J., Epa, R., ... Davies, G. J. (2022). The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids. *Current Research in Structural Biology*. <https://doi.org/10.1016/j.crstbi.2022.03.001>

2. Chapter 2: SQ binding protein SmoF

2.1. Abstract

Substrate-binding proteins are commonly found across nature acting as periplasmic solute scavengers and paired either with active transport apparatus for uptake of the substrate, ligand-gated ion channels or transcription factors. Their varying binding modes confer a wide range of recognition and binding specificities for their substrates, with some binding few molecules and some able to sequester with high promiscuity. In all cases binding is associated with a large-scale conformation change to enclose the ligand and exclude solvent. This conformation change is often modulated in the action of downstream transporters. The first known sulfoquinovosyl binding protein, SmoF, has been discovered in *A.tumefaciens* C58. Here SmoF is structurally characterized by X-ray crystallography and isothermal titration calorimetry, culminating in a comparison of the ligand free and ligand-bound states. Structures with a wide range of sulfoquinovosyl derivatives and glycosides are also obtained allowing for a detailed analysis of the binding mode. This allows identification of a dedicated sulfonate binding pocket as well as consistent recognition of the C2-4 hydroxyl groups. C1 and derivative recognition varies more with a flexible Gln-His containing “QH” loop deflecting with larger derivative binding, showing a degree of small-scale conformation as the basis of substrate promiscuity in SmoF. A bioinformatic analysis was also used to identify other putative SQBPs both using the ABC and TRAP transport archetypes, though a lack of consistent sulfonate recognition motifs among these sequences suggests SQBP action is the result of convergent evolution, as opposed to the strong conservation seen in sulfoquinovosidases.

2.2. Introduction

2.2.1. Substrate binding proteins in nature

Substrate-binding proteins (SBPs) are a class of protein commonly used for the specific sequestration of a solute or range of solutes, which can then be acted on in a through a downstream process. In prokaryotes, they can often be found as part of periplasmic scavenging systems that are typically paired to a transporter. When covalently bound, they are reclassified as substrate-binding domains (SBDs). SBDs can also be paired to a DNA-binding domain, working as an effector binding domain⁸⁸. They have also been found acting with ligand-gated ion channels⁸⁹, peptide receptors, G-protein coupled receptors and sensor kinases⁹⁰. While the most common pairing is the ATP-binding cassette (ABC) transporter class, Tripartite ATP-independent periplasmic (TRAP) and tripartite tricarboxylate transporters (TTTs) also employ them^{90,91}.

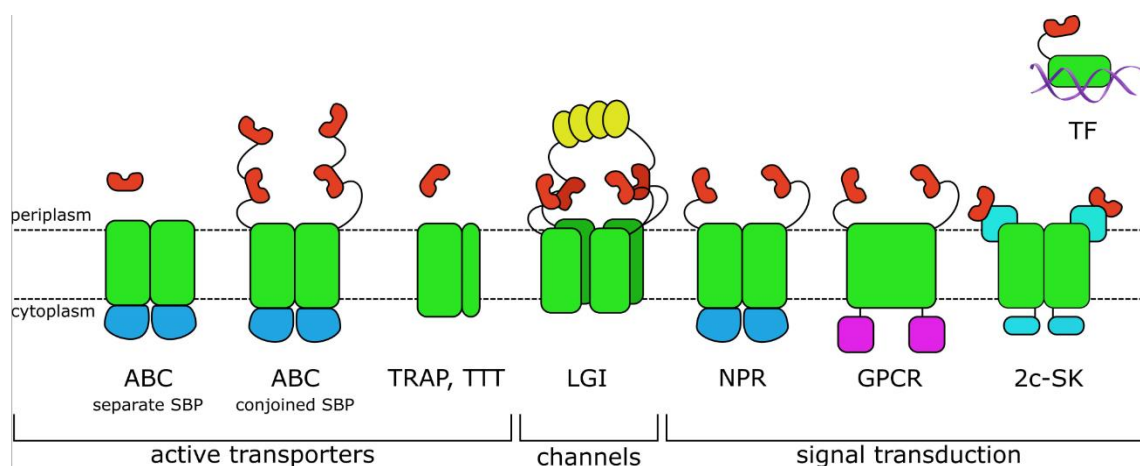


Fig 2-1: Transporters and systems employing SBPs. ABC transporters. ABC transporters rely on SBPs (orange) as separate proteins or as tethered domains for ligand capture, and in some cases recognition. TRAP/TTT is an ATP-independent tricarboxylate transporter which uses a periplasmic SBP. **LGI** Ligand-gated ion channels often have tethered SBPs. **NPR** The natriuretic peptide receptor uses a pair of tethered SBPs. **GPCR** G-protein coupled receptors including the glutamate receptor often employ SBPs. **2c-SK** 2-component kinase receptors can also use SBPs. **TF** Transcription factors can include SBPs to control transcription based on cellular substrate concentration.

2.2.2. *Classes of ABC transporter*

ATP binding cassette (ABC) transporters are a ubiquitous class of transmembrane transport complex, found in prokaryotes and eukaryotes with roles as both a nutrient import and metabolite export transporter. All ABC transporters have a conserved, 2-fold symmetrically similar arrangement of two identical transmembrane subunits, and two nucleotide binding domains (NBDs) in the cytoplasm. Some, such as MalFGK₂ (fig 2-2) also have an independent substrate-binding protein (SBP), which handles the specificity and substrate delivery for the transporter. ABC transporters can be grouped in two classes, dependent on how they recognise cargo. Type I transporters have no substrate specificity intrinsic to the transporter itself, and instead have a binding site for a dedicated SBD. Examples of such proteins are maltose binding protein (MBP) and the periplasmic oligopeptide binding protein (OppA)⁹²⁻⁹⁴. Type II instead have that specificity within the transporter itself. Energy-coupling factor (ECF) transporters have a comparable design to type II transporters, with a series of transporters fuelled by a transiently binding ATPase. The transporters are each specific to a different vitamin or ion. This similarity has led some to consider these prokaryotic systems as a third type of ABC transporter⁹⁵.

2.2.3. *Transport cycle of an ABC transporter*

2.2.3.1. *SBP recognition and substrate movement*

Substrate recognition by an SBP, and subsequent binding of the SBP to the transporter is the major influence to ATP hydrolysis and active transport. This is demonstrated in MBP through mutation studies of residues known to form part of the binding site between it and the transporter. These residues all lie on the cleft, closed in the conformation change of MBP with maltose bound, indicating they would be in a very different position apo, therefore suggesting a sensitivity to conformation in the binding mode^{96,97}. MBP can bind the transporter with or without a ligand with similar affinity causing inhibition of transport. However, spin labelling of MBP on each lobe and use of electron paramagnetic resonance (EPR) on an MBP-MalFGK₂ complex trapped in an intermediate state determined the open conformation of MBP exclusively in the complex⁹⁸. Cross-linking experiments with the histidine binding protein and transporter system suggests a difference in interaction between the ligand-free and bound states of HBP with the transporter despite similar overall affinity⁹⁹. The presence of the

binding protein is required for ATP hydrolysis. ATP is hydrolysed in an MBP-MalFGK₂ complex with or without maltose, though the turnover is far higher with maltose present¹⁰⁰. MBP closure is important to transport as ligands capable of binding but not inducing closure of MBP are not transported despite high affinity for MBP. Stabilisation of an intermediate state using vanadate to trap ADP in a transition state immediately after γ -phosphate hydrolysis reveals a lower-affinity maltose-MBP interaction consistent with the open state of the binding protein as well as an increased affinity for between the transporter and maltose. MgATP non-hydrolysable analogues have also been used to achieve a similar effect¹⁰¹. The closed state of MBP (fig 2-2, a) pre-MalFGK₂ binding appears to be crucial to MalFGK₂ conformation

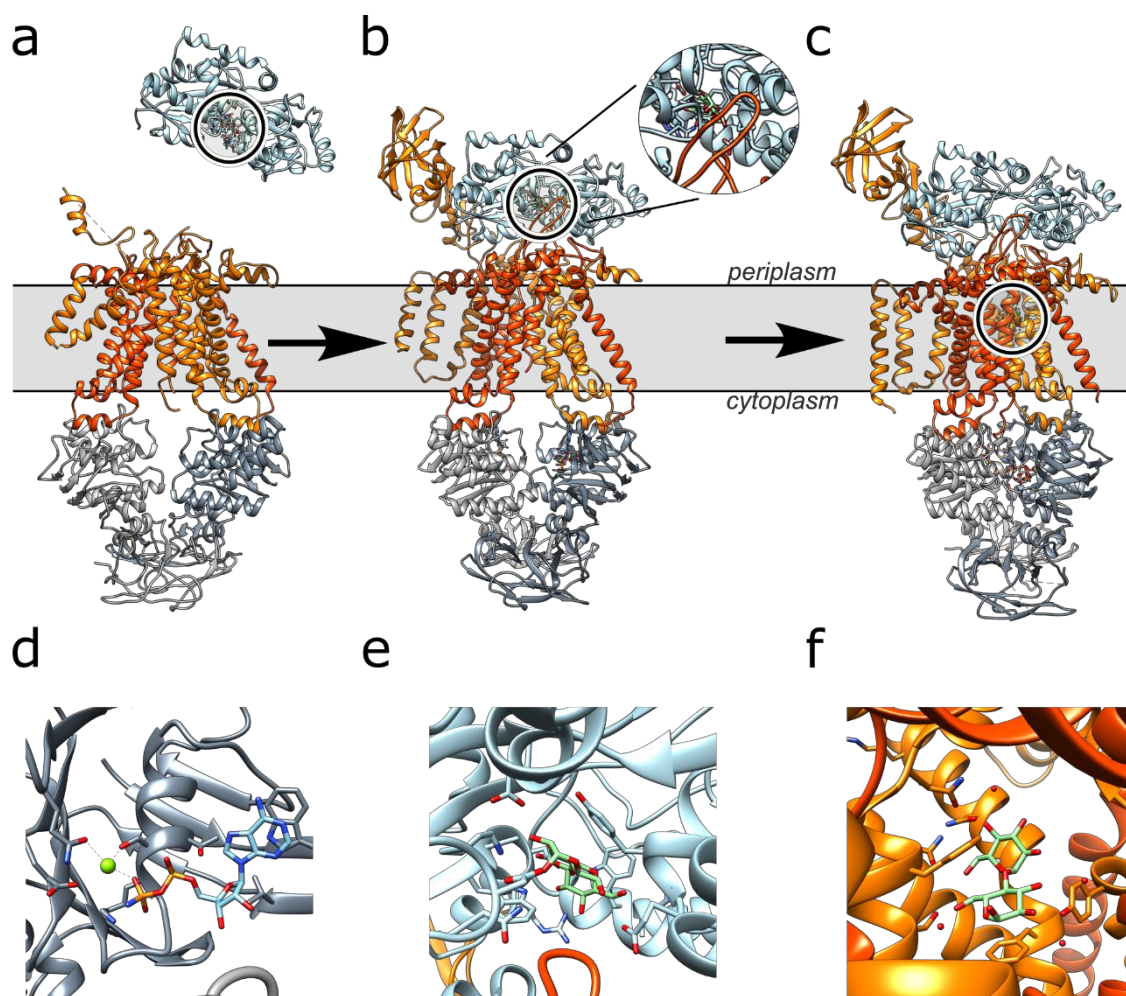


Fig 2-2: Transport cycle of the MalFGK₂ ABC transporter. MalF is orange, MalG red, MalK grey and MalE blue. Location of maltose during transport is circled, and maltose is green. **A,D.** MalE binds maltose and adopts the closed conformation. The transporter is P-open. ATP binds MalK (ADP shown in crystal structure). **B,E.** MalE binds MalF and adopts an open conformation. This forms a channel for maltose to dissociate, which is encouraged by the P-loop (orange, inset). The transporter is P-closed. **C,F.** Maltose moves through the MalFG pore. After transport ATP is hydrolysed causing reversion to P-open, and MalE dissociates.

change to a periplasmic open (P-open) state, as binding by open, ligand-free MBP causes the formation of a complex with an open MBP but MalFGK₂ in a P-closed conformation. Crystal structures of a P-open state MalGFK₂-MBP-Maltodexrose complex reveal a solvent-filled cavity of volume approximately 6500 Å³. This is large enough to enable transport of maltoheptaose, the largest substrate known to be capable of transport (fig 2-2, f). This transmembrane cavity is entirely shielded from the membrane by amphipathic α-helices and shielded from the bulk of the periplasm by the open-state MBP, though solvent channels with a maximum radius of about 5 Å allow for water molecule diffusion (fig 2-2, b,c). Maltose is present at the bottom of the transmembrane cavity bound to MalF (fig 2-2, e). The binding mode of this MalF-maltodextrose complex contains similar motifs to that used by the SBP MalE including a π-π interaction with a local Tyr side chain (fig 2-2, e,f) ¹⁰². Overall this presents a very different transport mode to transporters including maltophorin, which uses a "greasy slide" to enable maltose diffusion across the outer membrane ¹⁰³. MalFGK₂ appears to have a mechanism reliant on lowering MBP-substrate affinity to encourage movement of a single substrate to an intermediate binding site. The lowering of MBP-maltodexrose affinity is achieved by both the opening of the transporter, which removes many of the contacts present when closed, and protrusion of the P3 loop of MalG into the MBP binding site (fig 2-2, b,e). Attempts to model any experimentally confirmed MBP substrate into the open site with the P3 loop present result in a steric clash, suggesting a physical displacement of the substrate using the loop is required in addition to the lowered affinity (fig 2-2, e) ¹⁰¹.

2.2.3.2. *Coupling ATP hydrolysis to substrate transport*

Nucleotide binding domains (NBDs) are a major component of ABC transporters, with a sequence identity of 30-50% between prokaryotic and eukaryotic variants ¹⁰⁴. NBDs contain a larger and smaller domain. The larger is structurally similar to a RecA-like motor ATPase while the smaller is helical and unique to ABC transporters. The RecA-like domain typically contains two beta-sheets and 6 α-helices. It also contains the Walker-A motif (GxxGxGKS/T) and the Walker B motif (ϕϕϕϕD where ϕ represents any hydrophobic residue). The NBDs interact with the rest of the transporter largely through the cytosol exposed, conserved EAAxxxGxxxxxxxxlxLP region ¹⁰⁵, which contains a coupling helix which is responsible for many of the contacts between each NBD and the transmembrane domain above it ⁹⁴. This is true in the MalFGK₂E transporter + binding protein complex as well as others including the

ModB₂C₂A complex¹⁰². Crystal structures including these consistently show dimerised NBDs at the same time as a P-open conformation. ATP hydrolysis requires the dimerisation of each NBD¹⁰⁶. Crystal structures of inactive mutants show contacts between ATP and the Walker-A motif of one subunit and the LSGGQ motif of the other one. This pattern is present in the NBDs alone and in a structure of a full-length, ABC transporter-like protein¹⁰⁴. The activity has also been confirmed in solution using vanadate-mediated photocleavage of the NBD, when vanadate has been introduced to the ATP binding site. Cleavage was found on both the Walker-A and LSGGQ loop¹⁰⁷. The structures also contain a coordinated Mg²⁺ near the ATP binding site¹⁰⁸. ATP hydrolysis by an NBD is achieved through conformational changes within the domain, as noted by crystal structures of NBDs associated with many ABC transporters including MalFGK₂^{101,109,110}. Co-crystallisation of The *E.coli* MalK with ATP, ADP or neither revealed a conformational change on ATP binding, and subsequent hydrolysis, but that dimerization was not ATP dependent. ATP binds along the dimer interface, with a movement of the LSGGQ loop to contact ATP and partially bury it within the interface¹¹⁰. A comparison of ATP and ADP -bound crystal structures in the MJ1267 NBD revealed a movement of an α -helical sub-domain away from the position occupied by the γ -phosphate in the ATP-bound structure, suggesting a movement to release Pi upon ATP hydrolysis¹⁰¹. The domains also undergo a larger conformational change on ATP binding, described as a 'tweezers' motion of the recA-like regions around the regulatory regions. Distance measurements between H89 residues on each MalK between an ATP-bound and nucleotide-free structure reveal a closure of 16 Å with ATP binding. This increases the number of contacts between each monomer along the NBD regions, though the number of contacts on the regulatory domain is unchanged. Overall, this explains the increase in affinity observed for ATP-bound NBDs over nucleotide-free¹¹⁰. The positive charge of the coordinated Mg²⁺ may also contribute to the electrostatic interactions stabilising dimerization, as ATP-bound NBDs do not dimerize in the absence of it. Post hydrolysis, the negative charge on the Pi product may then cause electrostatic repulsion and therefore dissociate. Alternatively, the strong interactions between the gamma phosphate and LSGGQ motif may prevent transient dissociation post-hydrolysis, so this negative charge may instead help force the closed NBD conformation back to an open state¹¹¹. The ADP product and Mg²⁺ have been confirmed through crystal structures to remain bound¹⁰⁸. MsbA is another ABC transporter, used as a "flippase" to move lipopolysaccharides (LPSs) from the inner leaflet post-synthesis to the outer leaflet of Gram-

negative bacterial plasma membranes. The overall mechanism for transport is similar to that of MalK, with a P-open and closed transmembrane conformation that flips the substrate between periplasm and cytoplasm exposure based on the binding and hydrolysis of ATP by a homodimeric NBD pair ¹¹². Through use of time-resolved Fourier transform infrared spectroscopy (FTIR) a V_{max} of 45 nmol.mg⁻¹.min⁻¹ was found for ATP hydrolysis for the NBD of MsbA alone. A similar value is found for the full transporter. FTIR found two major rate constants. The faster of the two (0.01 s) representing formation of the MalK dimer and the slower (~1 s) representing ATP hydrolysis. A Hill coefficient of 1.49 also presented the idea of positive cooperativity between NBDs. Overall this shows a common method between ABC transporter NBD activity, with dimerisation preceding, and being necessary for, ATP hydrolysis ¹¹³.

2.2.4. Substrate binding proteins

Substrate-binding proteins (SBPs) are the third component of the ABC transporter. Unlike the others, SBPs are often separate proteins, in the periplasm (though they are sometimes linked to the transporter, in which case they are named substrate binding domains, SBDs). These proteins follow a three-domain architecture, with each domain featuring a β -sheet. These act as a “clamshell” that closes around the ligand upon binding. The proteins are often in the 40-50 kDa size range, with this size being independent of the size of intended ligand. While the general fold remains similar, the sequence of SBPs appears dependent more on the intended ligand type than the source organism. A comparison between more than 50 bacterial SBPs identified 8 clusters of similar sequence, which roughly sort into the class of molecule the SBP binds ¹¹⁴.

2.2.4.1. Substrate recognition and binding

The variation in preferred substrate for SBPs gives rise to a large variety of different binding modes. OppA is an example of how SBP binding modes vary. The protein binds peptides 2-5 residues in length in a mostly side-chain independent manner. The binding appears to mostly come from the backbone of the ligand forming an antiparallel β -sheet complementary to the ligand backbone between residues 413-419 (fig 2-3, a-b)). The ligand is held in an extended

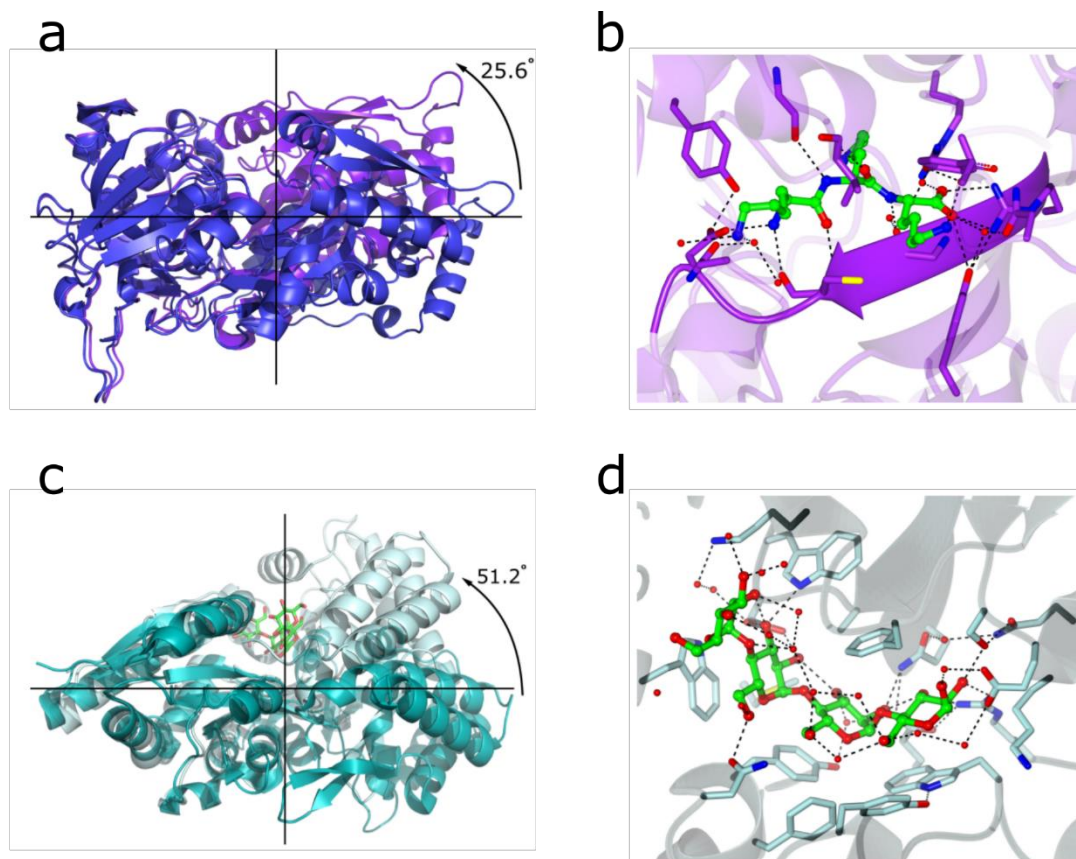


Fig 2-3: Conformation changes and ligand binding pattern of OppA with tri-lysine and MalE with maltotetraose. **A.** Superposition of ligand-free MalE (cyan) and MalE with maltotetraose (light blue). A hinge motion was found of 51.2° between the two conformations achieving enclosure of the binding site. **B.** Maltotetraose binding site of MalE₂. Primary binding occurs from contacts between Asp₆₆, Glu₁₁₁ and Arg₃₀₃ with the C2 and 3 hydroxyls of maltose monomers 2, 3 and 4. **C.** Superposition of ligand-free OppA (blue) and OppA with tri-lysine (purple) crystal structures. A hinge motion of 25.6° was detected between the open and closed conformations, resulting in full ligand enclosure. **D.** Detail of the tri-lysine binding site. The ligand adopts an extended conformation, acting as an antiparallel β-sheet against residues 413-419. The lysine side chains are further stabilised by hydrogen bonding to the clamshell domains. PDBs: 1RKM (ligand-free OppA) 2OLB (trilys), 6DTT (ligand-free MalE₂) 6DTS (maltotetraose).

conformation by ionic bonds at the N and C termini. Once bound, the ligand is enclosed by the clamshell domains, which interact with the side chains. This conformational change entirely buries the ligand. Further cocrystals with a Lys-Lys-Lys-Ala peptide and uranyl anions determined the clamshells contain large, hydrated cavities to contain the side chains in a non-specific manner. Binding affinities vary by peptide length; a dialanine will bind with about 100x the strength of trialanine⁹³. This binding model fits the cell wall peptide recycling function of the protein. Dipeptide binding-protein precursor (DppA), a close structural homologue of OppA, has a similar binding model, though with a preference for dipeptides.

The preference is achieved as DppA lacks binding residues R413, H371 and K307, with equivalent positions being filled with neutral residues ¹¹⁵. This multi-specificity is common to DppA across many Gram-negative bacteria, as the residues involved in backbone binding are conserved. However, the individual dipeptide affinities are modulated as side-chain binding and accommodating residues are often not conserved. The DppA variant found in deep-sea *Pseudoaltermonas* binds 8 different dipeptides out of 25 tested, with 4 crystal structures conforming a similar backbone-based binding model ¹¹⁶. MalE is a maltose binding protein from *T. maritima* binding maltotetraose, and one of three isoforms in the bacteria, which bind di-, tri- and tetrasaccharides ¹¹⁷. All three use a similar binding model, with the ability to bind larger oligosaccharides arising from a longer binding pocket. Maltotetraose, when binding MalE adopts an extended conformation, with recognition arising from coordination of the C2 and 3 hydroxyls of maltose monomers 2, 3 and 4 with Asp66, Glu111 and Arg303 (fig 2-3, c-d). This is a common method of binding for carbohydrate binding proteins. Carbohydrate-binding SBPs can also exhibit the multi-specificity shown in OppA. MalE1 in *T. thermophilus* binds trehalose in order to enable ABC transporter-mediated uptake with a K_M of 109 nM, as well as maltose at 114 nM and sucrose at 424 nM ¹¹⁸. MalE1 is highly thermally stable, with a T_m of 90°C.

2.2.4.2. Conformation changes upon SBP-ligand binding

Ligand binding for any SBP results in a large conformational change, in which the ligand is enclosed by the clamshell domains and stabilised, often totally excluding it from the bulk solvent. A comparison of crystal structures of apo and holo-SBPs show a motion typically between 25° and 55°. This is not associated with large-scale changes in the domains, which retain their individual conformations ¹¹⁷. The dynamics of this change are more challenging to visualise. Using Paramagnetic Relaxation Enhancement (PRE) NMR, the apo MBP from *E. coli* was observed in an equilibrium between an open and partial closed conformation. This change is reported to take place on the ns to μ s scale on a spin-labelled, D41C and S211C MBP mutant. MBP has a 95% favourably towards the open conformation in the ligand-free state ¹¹⁹.

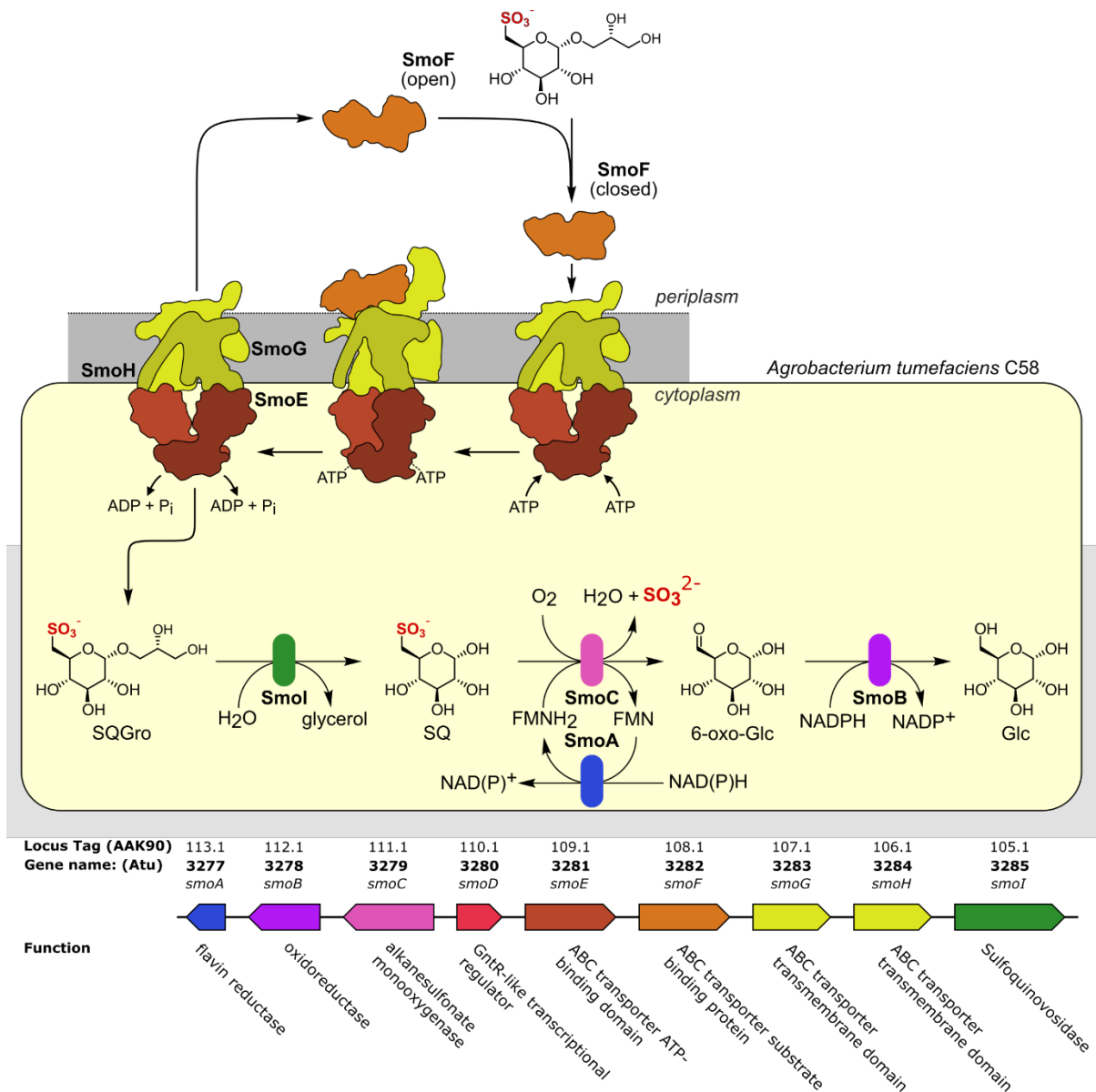


Fig 2-4: Role of SmoF in the sulfo-SMO pathway. SQGro (top) is sequestered by SmoF, and the SmoGHE₂ ABC transporter enacts the active transport of it into the cell. There, glycerol is hydrolysed by the sulfoquinovosidase SmoI followed by oxidative desulfurisation by SmoCA. The product 6-oxo-glucose is then reduced by SmoB.

2.2.5. SBPs and sulfoglycolysis: Aims of this chapter

The first SQBP, SmoF was discovered in *A.tumefaciens* C58 as part of the sulfo-SMO pathway present there. The same operon contains genes for an ABC transporter, SmoGHE₂ which together would work to transport SQGro into the cell. After this, glycosidase activity would produce SQ, which following oxidative desulfurisation to form 6-oxo-glucose would be reduced producing the final product of the pathway, glucose (fig 2-4). While SmoF was

expected to behave in a similar capacity to other SBPs, the mode of ligand recognition, affinity, specificity and structural changes were unknown. Here, the SQBP SmoF is characterized for the first time, with structures produced in the ligand-free state and with a range of sulfoquinovose derivatives and glycosides. These data are combined with solution binding assays and bioinformatic analyses to produce a detailed picture of the structure and function.

2.3. Materials and Methods

2.3.1. *Expression and purification of SmoF*

The gene for SmoF in a pET29a vector was transformed into *E.coli* BL21(DE3) competent cells for expression, using the standard NEB protocol, and plated onto LB-agar with 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin. These were grown at 37 °C over 18 hours. Expression began with 18-hour pre-cultures in 5 ml LB containing 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, at 37 °C, shaken at 180 rpm. This was scaled up to 2L expressions through inoculation of two autoclaved 1L LB expression jugs with half of the 5 ml pre-culture each. Kanamycin was added for a final 30 $\mu\text{g}\cdot\text{ml}^{-1}$ concentration. These were grown at 37 °C, 225 rpm until an OD600 of 0.6-0.8 was reached. Induction then took place with the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final 1 mM concentration, and the temperature and shaking were lowered to 18 °C, 180 rpm for 18 hours. Selenomethionine labelled expression began with 18-hour pre-cultures in 5 ml LB containing 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, at 37 °C, shaken at 180 rpm. The cells were pelleted at 4000 rpm for 10 minutes at 10 °C and resuspended in 10 ml M9 minimal medium. 2 ml of this was added to each of two flasks containing 500 ml M9 medium, and the cells were grown at 37 °C, 225 rpm until an OD600 of 0.5 was reached. 5 ml of "magic 6" amino acids (100 mg Lysine, Phenylalanine and Threonine, 50 mg Isoleucine, Leucine and Valine in 10 ml H₂O) plus 60 mg selenomethionine was added to each flask, and growth conditions were resumed for a further 15 minutes. 250 μl 1 M IPTG was added, and induction was carried out at 18 °C, 180 rpm for 18 hours. The cells were harvested by centrifugation at 5000 g for 20 minutes, and resuspended in binding buffer (50 mM Tris HCl, 300 mM NaCl, 30 mM Imidazole, pH 7.4). Pellets could be snap frozen and stored in -80 °C or lysed at this point. Cell lysis used a disruptor at 25 MPa, with the lysate collected on ice. Debris was pelleted at 18,000 g for 30 minutes at 6 °C. The pellet was discarded. The soluble fraction loaded to a pre-equilibrated 5

ml Ni-NTA HisTrap column, using a sample pump at 5 ml.min⁻¹. The column was then washed with binding buffer at 5 ml.min⁻¹ until the A280 read less than 20.0 mAU. His-tagged protein was eluted at 5 ml.min⁻¹ in a linear gradient of increasing elution buffer (50 mM TRIS, 300 mM NaCl, 250 mM Imidazole, pH 5.5). Fractions were collected in a 96-well deep block. Fractions identified as containing protein of the correct size by SDS-PAGE were pooled and concentrated in 20k MWCO Vivaspin® concentrators to a 2 ml final volume. This was further purified by size-exclusion chromatography on a HiLoad Superdex 16/600 S200 column. This was performed in 50 mM citrate, 150 mM NaCl, pH 5.5. The resultant fractions with a high absorbance at 280 nm were run on SDS-PAGE for purity and size assessment. All purified protein aliquots were flash frozen and stored at -20 °C, at 30 mg.ml⁻¹. Defrosted protein was stored at 4 °C when not in immediate use, for a maximum time of one week. All SDS-PAGEs for all proteins were performed using 12% acrylamide gels and run for 45 minutes at 200 V with 0.1% SDS TRIS/Glycine running buffer. Gels were subsequently stained with Coomassie magic blue dye. Unless otherwise indicated, 7 µl of sample was loaded, with 5 µl loading dye and 3 µl water for a 15 µl total load.

2.3.2. *Intact ESI mass spectroscopy*

Purity and mass estimates were obtained using electrospray-ionisation mass-spectroscopy (ESI-MS) with a Bruker Maxis-HD qTOF mass spectrometer. All samples were prepared to 1 µM with 50% by volume acetonitrile, 1% by volume formic acid. Infusion was at 3 µl.min⁻¹. Spectra were averaged over 1 minute and smoothed in 1 cycle using a Gaussian method, then the baseline was subtracted with a flatness of 0.8. Compass 1.7 software was used for this. Protein concentration was determined through absorbance at 280 nm, followed by calculation via extinction coefficient. 100 µl aliquots at 25 and 50 mg.ml⁻¹ were flash frozen in liquid nitrogen before storage at -20 °C.

2.3.3. *NanoDSF*

NanoDSF was performed in 10 µl sample capillaries on a Prometheus NT.48 (NanoTemper). Excitation was dependent on the sample, and presence of NADPH. 330/350 nm ratio of fluorescence was recorded between 15 °C and 95 °C, at 1 °C.min⁻¹. Data collection and preliminary analysis performed using ThermalControl (NanoTemper). All SmoF samples were at 1 mg.ml⁻¹ in 50 mM citrate, 150 mM NaCl pH 5.5. SQDG-(C4:0/C16:0) and SQDG-

(C16:1/C18:0) were contained in 100% DMSO at 250 mM, and diluted into working concentrations with 50 mM citrate, 150 mM NaCl pH 5.5. All samples were centrifuged at 13,000 rpm for 5 minutes prior to loading.

2.3.4. X-ray crystallography and structure determination

Ligand-free SmoF crystals were grown by sitting drop at 50 mg.ml⁻¹ protein in 0.3 M ammonium acetate, 0.1 M bis-tris (pH 5.5) and 25% PEG 3350 at 6 °C. A 1:1 ratio of protein and mother liquor was used. Selenomethionine-labelled SQGro cocrystals were grown by sitting drop at 50 mg.ml⁻¹ protein in 0.3 M MgCl₂, 0.1 M bis-tris (pH 5.5) and 25% PEG 3350 at 20 °C, with a final 1:1 mother liquor: protein ratio. SQ cocrystals were grown by sitting drop using 20 mg.ml⁻¹, in 0.1 M NaBr, NaI, 0.1M imidazole, MES pH 6.9, 13.5% MPD, PEG 1000, PEG 3350 at 10 °C, with a 5:6 mother liquor: protein ratio. Crystallisation of SQ-Me cocrystals used 50 mg.ml⁻¹ protein in 0.3 M sodium acetate, 0.1 M bis-tris (pH 5.5) and 35% PEG 2000 MME in a sitting drop, with a 1:1 protein:mother liquor ratio in-drop. SQ-oct cocrystals were grown using 50 mg.ml⁻¹ protein in 0.2 M calcium acetate, 0.1 M bis-tris (pH 6.5), 25% PEG 2000 MME, at 20 °C, with a 1:1 protein:mother liquor ratio. For SQGro, SQ, SQ-me and SQ-oct, 2 mM ligand was added to protein 10 minutes prior to drop formation. SQDG-(C4:0/C16:0) cocrystals of SmoF were grown by sitting drop at 6 °C, with 25 mg.ml⁻¹ protein and 2.5 mM SQDG-(C4:0/C16:0) dissolved in 100% DMSO, in 25 mM NaPi, 150 mM NaCl pH 7.0, incubated at room temperature with 2.5 mM SQDG-(C4:0/C16:0) for 10 minutes prior to crystallisation. Diffraction quality crystals were collected from a direct scale up of the Morpheus screen (Molecular Dimensions), condition H12. This contains 0.1 M amino acids (0.2M L-Na-Glutamate, 0.2M Alanine, 0.2M Glycine, 0.2M Lysine hydrochloride, 0.2M Serine), 0.1 M buffer system 3 pH 8.5 (1 M TRIS, 1 M bicine) and 50% v/v precipitant mix 4 (25% MPD, 25% PEG 1000, 25% PEG 3350). Crystals only formed in premade mother liquor. No cryoprotectant was used on the resultant crystals due to the presence of cryoprotecting PEG solutions in the mother liquor. Crystals were fished then flash cooled in liquid nitrogen, using nylon CryoLoops™ (Hampton). All crystals were tested using the in-house Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV. X-ray data collection took place at Diamond Light Source, using beamline I-03 during collection mx18598-51. Data collection statistics are

available in Table 2-1,2-2.. Data indexing and initial processing was performed at Diamond Light Source, using either dials or 3dii pipelines from the Xia2 package ¹²⁰. This was used to index and scale data in all cases except for SmoF with SQDG-(C4:0/C16:0), where indexing was performed manually using DUI ¹²¹. Data reduction was performed with AIMLESS. Assigning the resolution cutoff was challenging because of considerable anisotropy in the data, so a higher CC1/2 than the normal 0.5 was chosen ¹²². In all cases AIMLESS was used for data reduction and quality assessment. Molecular replacement used either MOLREP or PHASER ^{123,124}, and early model building was automated using BUCCANEER ¹²⁵. Model refinement was performed using REFMAC5, and all interactive modelling and validation performed in COOT ^{126,127}. All steps excluding manual data integration were performed from within the CCP4i2 system ¹²⁸. Structural analysis and diagram preparation for molecular models was performed using Chimera, CCP4MG or Pymol, depending on the desired outcome ^{129–131}.

2.3.5. Isothermal Titration Calorimetry

All samples of SmoF were dialysed into ITC buffer (25mM NaPi, pH 7.5, 3 rounds of 400ml) at 4 °C over 6 hours using 10 kDa MWCO Visking tubing. It was then diluted to a final [X] μ M stock concentration using the final dialysate buffer. Protein concentration was assessed by A280, and the accuracy of this measurement was assessed by BCA protein concentration assay (Thermo Fisher). Stocks of SQ, SQMe and SQOct were dissolved from dry to a 50 mM stock concentration using this same dialysate buffer, ultrafiltered through a 3 kDa MWCO concentrator to remove all protein. Buffer stocks for ITC syringe and cell prewashing were also filtered in this fashion. All samples were degassed and checked for aggregation by centrifugation at 13,000 g for 10 minutes prior to loading in the MicroCal autoITC200 (Malvern Panalytical). This was used to automate 20 injections, the first containing 0.4 μ l of ligand, and a wait 4/16 of 180 seconds, and all subsequent injecting 3 μ l ligand over 6 seconds, with a 150 second wait. A water-water run and a buffer wash were performed before every new experiment. A ligand-buffer control was also run for each ligand to provide a baseline heat plot for subtraction. All data analysis and experimental simulation was performed in MicroCal and all plotting in Origin.

2.3.6. Bioinformatics

Initial pathway discovery was performed by collaborators in the group of Spencer Williams and are presented in Snow *et al*, 2022¹³². All MSAs were produced using Clustal Omega, and processed using Jalview 2.11.2.0¹³³. All protein structure prediction was performed using AlphaFold 2.0¹³⁴.

2.4. Results

2.4.1. Expression, purification and early characterization

Due to the lack of existing crystal structures for this protein or homologues for viable molecular replacement models, experimental phasing was an anticipated requirement to solve the initial SmoF structure. To this end, the protein was expressed natively and with selenomethionine (SeMet) labelling. Due to the use of the “magic 6” pathway inhibition method, a methionine-auxotrophic *E.coli* cell line was not required for labelling, and BL21(DE3) was used in both cases, as was the pET28a expression vector. In both cases an estimated 95% sample purity was achieved using immobilised metal affinity chromatography (IMAC) to isolate the 6xHis-tagged protein, followed by polishing by size-exclusion chromatography (fig 2-5). This produced an average of 40 mg.ml⁻¹ protein per litre culture in native expression. Initial purity and yield estimates were taken with SDS-PAGE. The fractions were loaded without dilution leading to a high protein concentration on-gel, resulting both in the large bands and the obvious impurities. The largest bands were in the expected location given the sequence-derived Mw of 44 kDa, and A280 measurements during size-exclusion chromatography confirmed protein elution at around 70 minutes, again expected for a protein of that size. Electrospray ionisation mass-spectroscopy (ESI-MS) further confirmed the protein was of the correct size to be SmoF with the highest intensity m/z peak at 44064.3 Da. Comparing the mass of the highest intensity m/z peaks in both MS traces highlighted a difference of 377.3 Da between them. This extra mass in the SeMet expression corresponds to approximately 8 SeMet residues replacing methionine residues in the protein. This

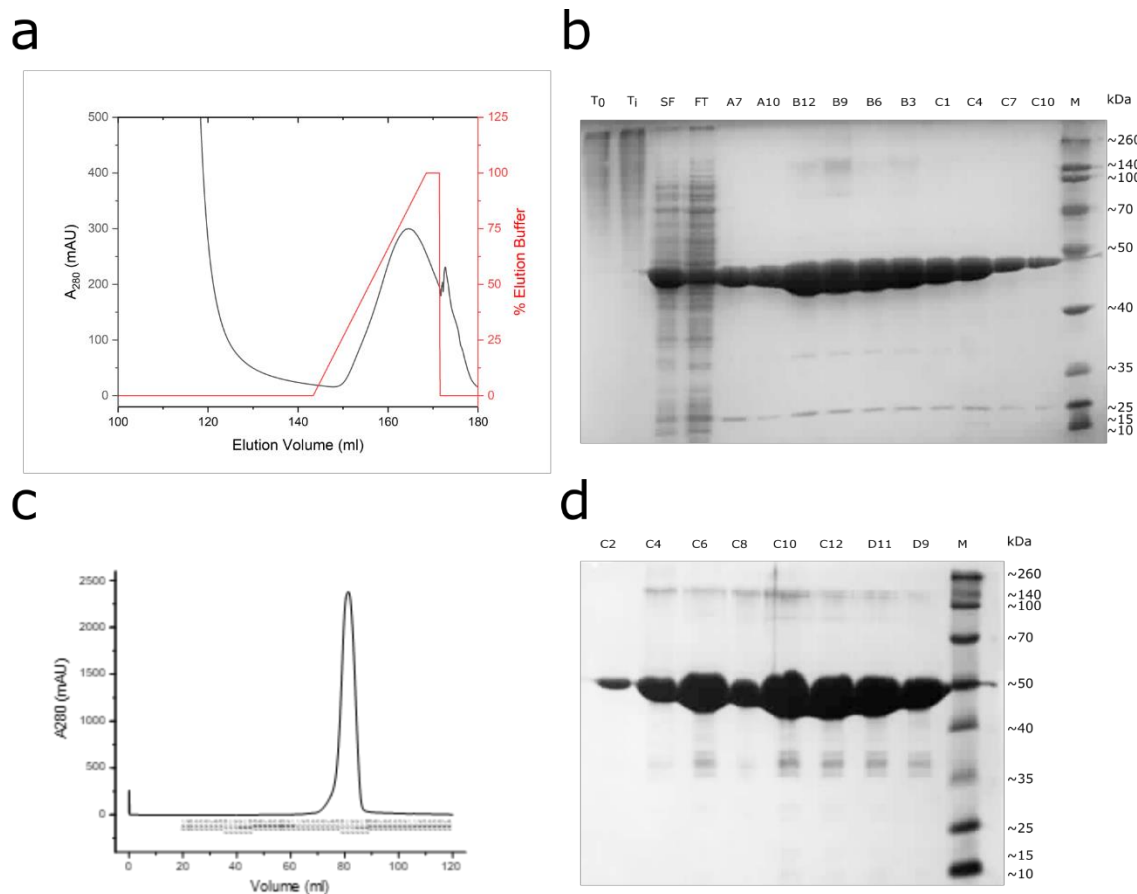


Fig 2-5: Expression and purification of SmoF. **A.** A280 elution profile of SmoF during IMAC with increasing % of elution buffer (red). **B.** SDS-PAGE of SmoF growth and IMAC. From left to right: Lysed cells at time of inoculation (T_0), time of induction (T_i), soluble fraction (SF), flow-through (FT), elution fractions A7-C10 and marker. A band can be clearly seen during IMAC elution at ~ 44 kDa, which corresponds to the Mw of SmoF. **C.** A280 elution profile of SmoF during SEC. **D.** SDS-PAGE of SmoF polishing. From left to right: Elution fractions C2-D9, marker. A band can be clearly seen at ~ 44 kDa, which corresponds to the Mw of SmoF.

indicates a successful incorporation with high specificity. The lack of other meaningful peaks in the SeMet expression further indicated a high purity for that protein. Similar purity was not achieved in the native expression trace.

2.4.2. *SmoF binds SQGro, SQDG, SQ and C1 SQ derivatives*

NanoDSF was used to determine a T_m of 48.2°C for ligand-free SmoF, which increased by 10.3°C to 58.5°C with 2 mM SQ (fig 2-6, a; fig 2-7, a-b). This increase in thermal stability suggests a high affinity binding. A similar result was seen with 2.5 mM SQGro, which provided

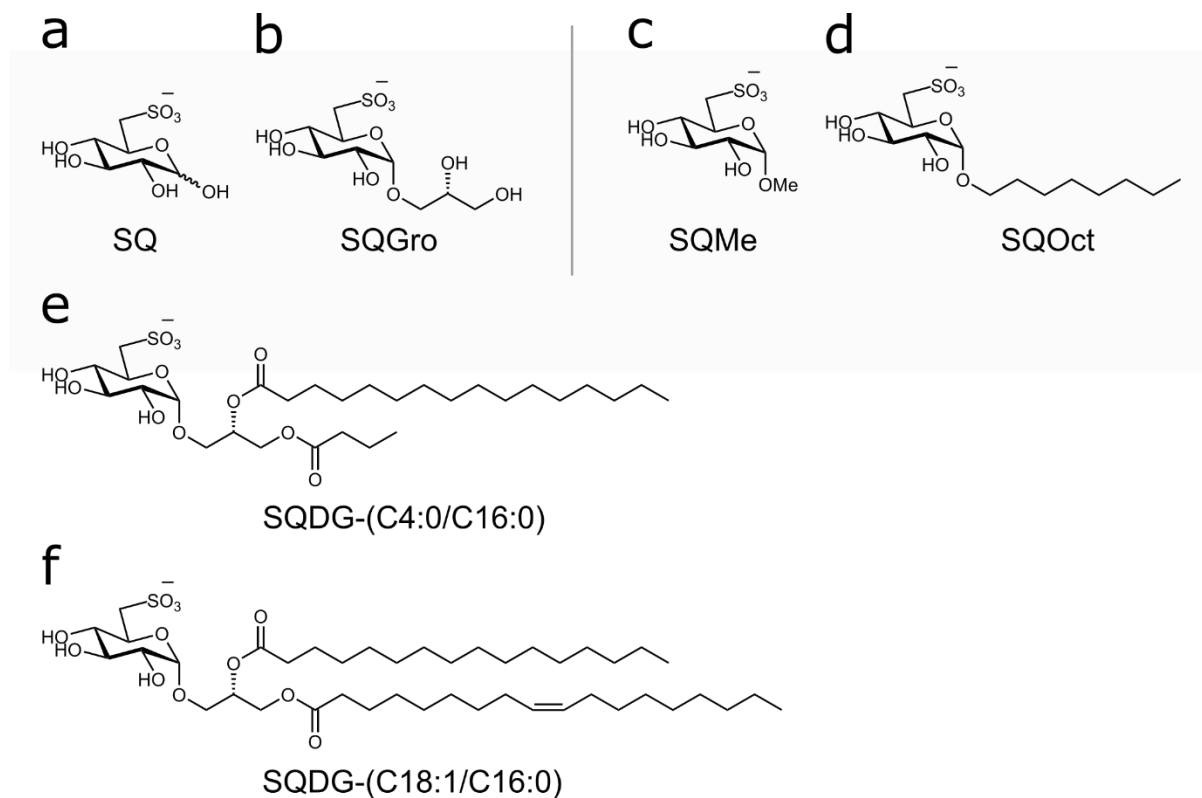


Fig 2-6: Sugars and lipids used in this study. **SQ** sulfoquinovose. **SQGro** sulfoquinovosyl glycerol. **SQMe** C1-methylated sulfoquinovose. **SQOct** C1-octylated sulfoquinovose. **SQDG-(C4:0/C16:0)** sulfoquinovosyl diacylglycerol (butanoyl/palmitoyl chains). **SQDG-(C18:1/C16:0)** sulfoquinovosyl diacylglycerol (oleoyl/palmitoyl chains). All ligands except SQ were kindly provided by the lab of Spencer Williams.

a T_m shift of 11.2 °C over ligand-free SmoF (fig 2-6, b; fig 2-7, b). This again suggests a high binding affinity, and a K_D of 8.7 μM for SmoF with SQ was subsequently confirmed by isothermal titration calorimetry (ITC). This is around 20-fold lower than the K_D for SmoF with SQGro, at 0.29 μM (fig 2-8)⁸⁰. To better determine the adaptability of the binding cleft to larger moieties, a C1 methylated SQ (SQ-Me) and a C1 octylated SQ (SQ-Oct) were synthesised. These produced a ΔT_m of 14.6 °C and 15.3 °C respectively (fig 2-6, c-d; fig 2-7 a-b). The capacity for SmoF to bind sulfoquinovosyl diacylglycerol (SQDG) compounds, and therefore whether delipidation must occur before transport, was an important question both for testing the observed capacity for C1 moiety accommodation in SmoF and for confirming the mechanism of the pathway itself. SQDGs with a butyl fatty acid/palmitic acid and palmitic/oleic acid were produced (fig 2-6, e-f)¹³⁵. These are labelled SQDG-(C4:0/C16:0) and SQDG-(16:1/C18:0) respectively. The T_m of SmoF with SQDG-(C4:0/C16:0) increased by 7.7 °C over the ligand-free protein in nanoDSF (fig 2-7, b). 2 mM SQDG-(C16:1/C18:0) failed to increase the T_m , and measurements with up to 10 mM also showed no change suggesting a

lack of even weak binding. Though the capacity to bind a variety of C1 SQ glycosides was at this point confirmed, the ability of SmoF to bind other sulfonates and sugars was unknown. The T_m of the protein with 2 mM sulfofructose (SF), glucose, HEPES, MES, PIPES and MOPS was measured. No meaningful change in T_m was detected in any case. It is highly likely that SmoF requires both the 6C sulfonate and glucose moieties to be present in a ligand to bind (fig 2-7, c).

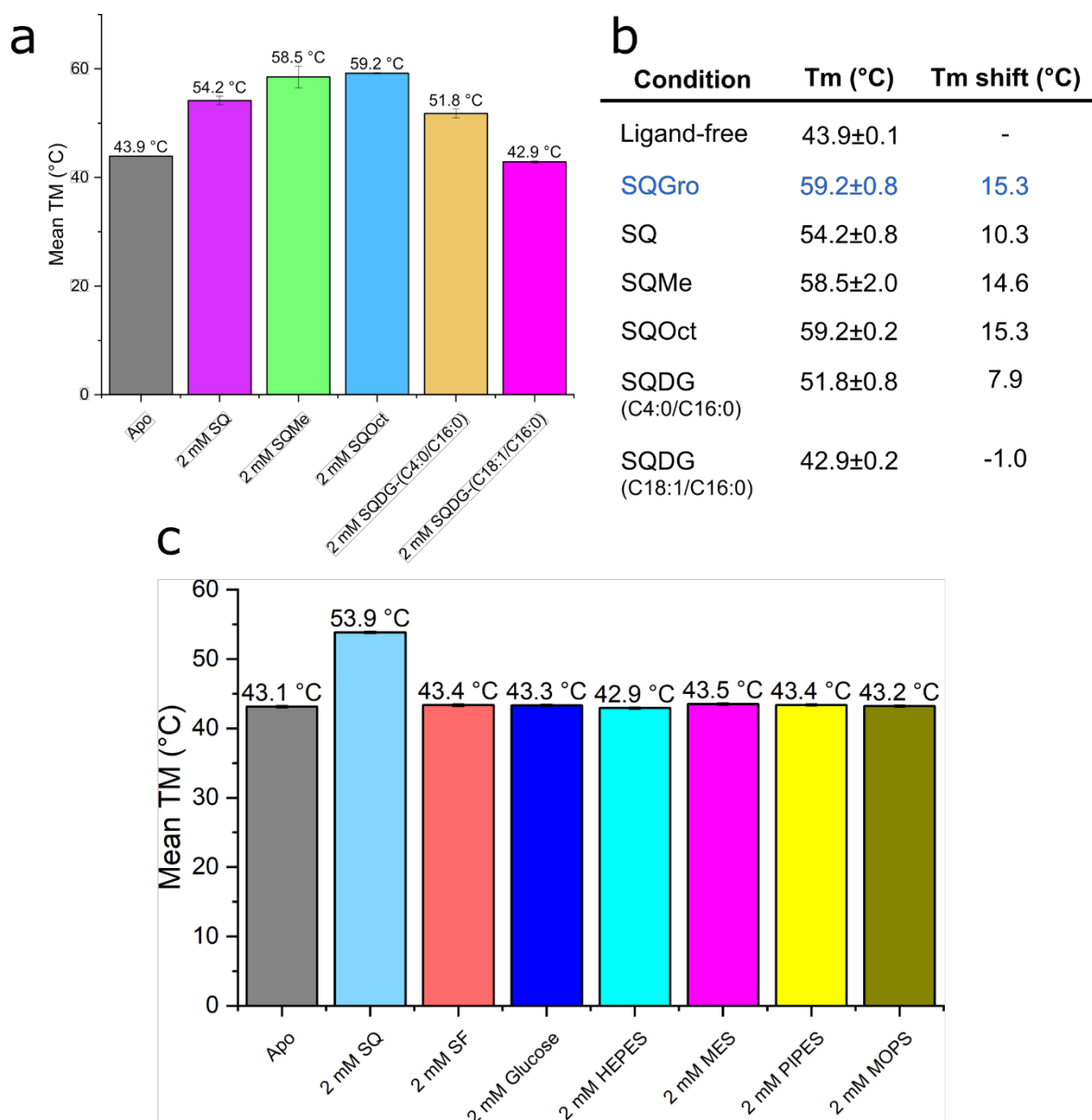


Fig 2-7: Thermal stability assays of SmoF with ligands. A. T_m of SmoF with temperature ligand free, with SQ, SQMe, SQOct, SQDG-(C4:0/C16:0) and SQDG-(C4:0/C16:0). **B.** Table of T_m changes with ligands. SQGro was obtained at 2.5 mM and is highlighted in blue. **C.** Bar chart of T_m with SQ (positive control), SF, glucose, HEPES, MES, PIPES and MOPS.

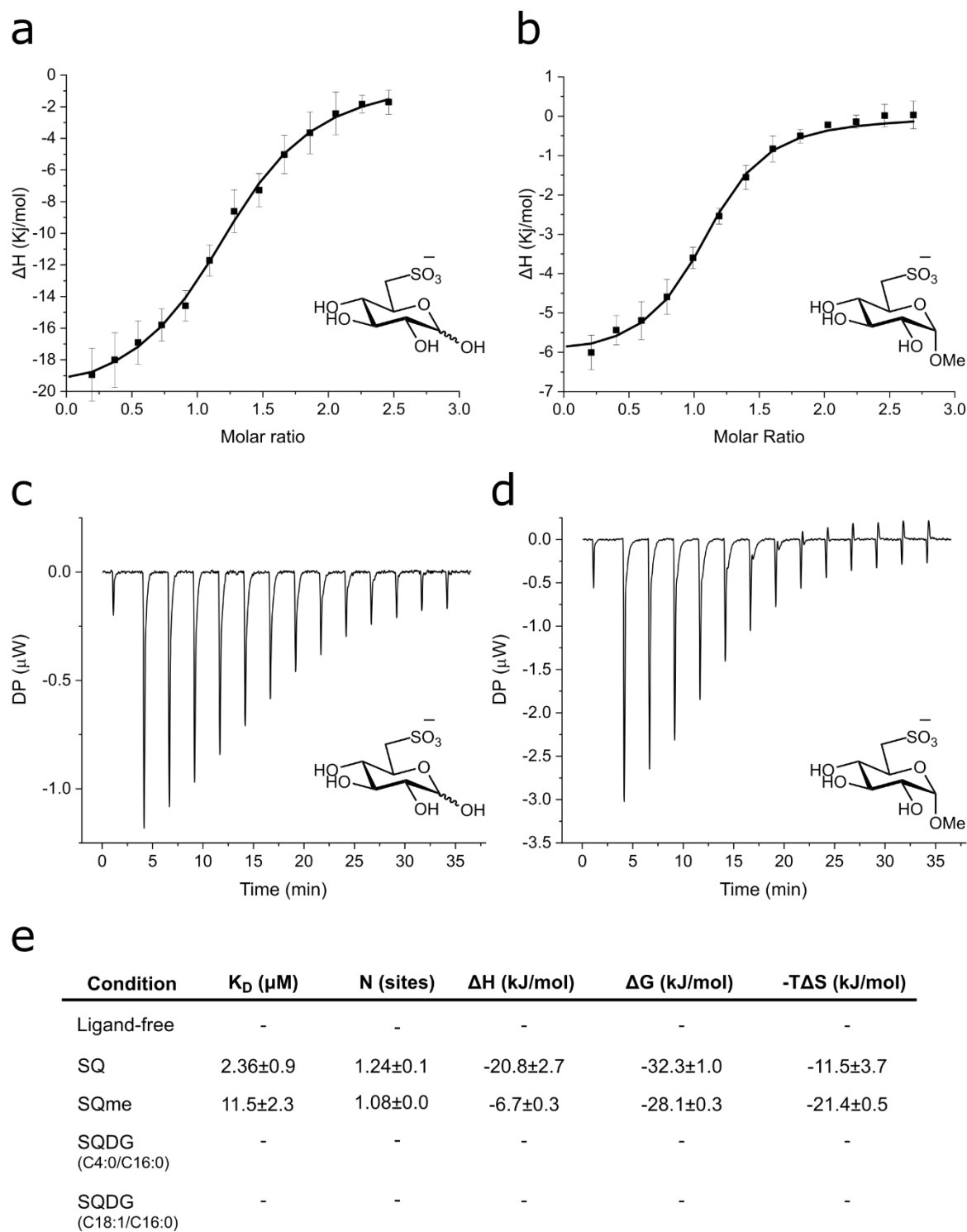


Fig 2-8: Thermodynamic analysis of SmoF with SQ, SQMe. A-B. Integrated heat plots of SmoF with **A.** SQ and **B.** SQMe. **C-D.** Baseline-subtracted thermograms of SmoF titrated with **C.** SQ and **D.** SQMe. **E.** Thermodynamic properties of SmoF with regards to ligands.

2.4.3. Structural characterization of SmoF

Crystallography of ligand-free SmoF did not initially result in harvestable crystals in any screening condition. However, crystals did form at room temperature in a variety of conditions when co-crystallised with SQ suggesting a possible conformational stabilisation effect on ligand binding, allowing for the sample homogeneity required for crystal growth. There was no difference in conditions leading to crystal formation between native and SeMet-labelled SmoF, leading to the harvesting of diffracting crystals of labelled protein with SQGro for experimental phasing. This however proved unnecessary as a collaborator produced a native structure of SmoF with SQGro⁸⁰. NMR data concerning the interdomain motion of the apo MBP, which has a similar (global RMSD = 2.6 Å) suggested SmoF was undergoing a rapid change from an open to partially closed conformation¹¹⁹. This in turn would introduce heterogeneity in the crystal condition inhibiting crystal formation. Crystallisation at 6 °C produced diffracting crystals. These crystals proved unstable at any higher temperature but

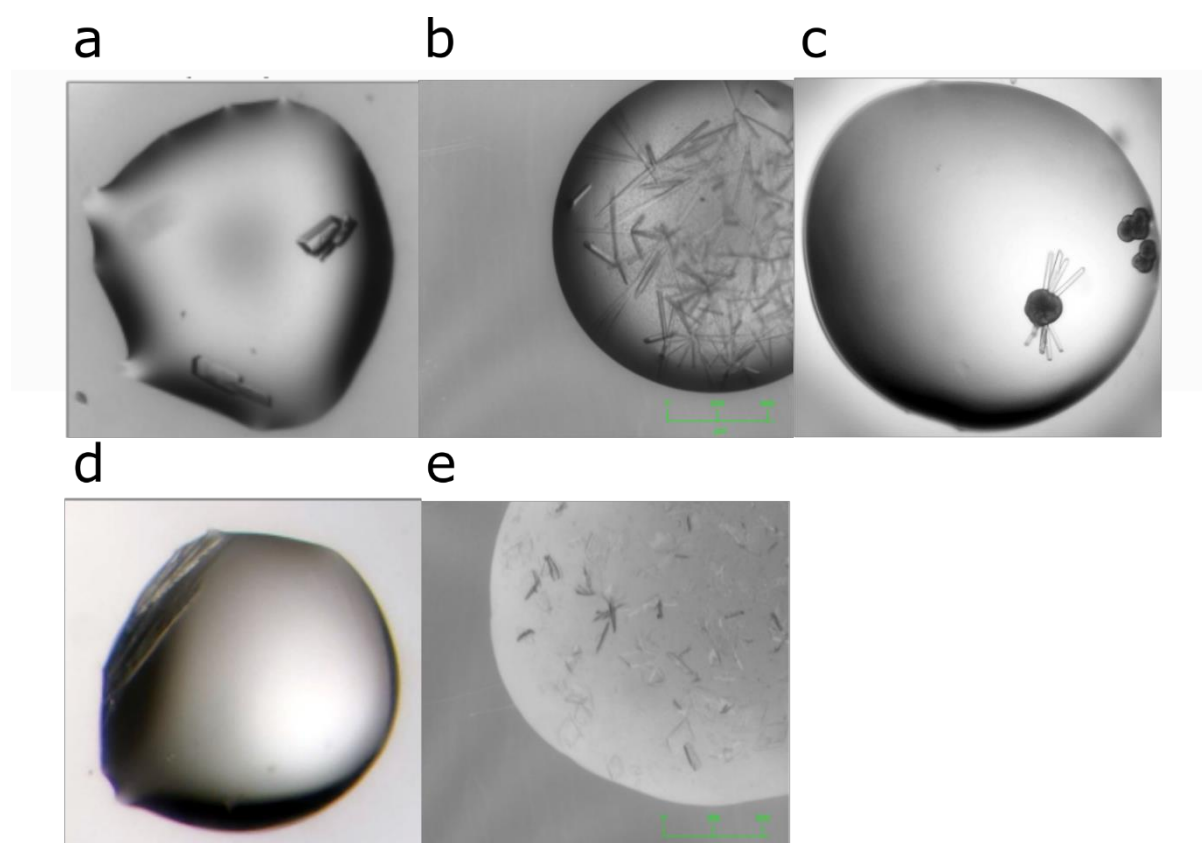


Fig 2-9: Crystals of SmoF with ligands. Micrographs of **A.** Ligand-free SmoF. **B.** SmoF•SQ **C.** SmoF•SQMe **D.** SmoF•SQOct **E.** SmoF•SQDG-(C4:0/C16:0).

	Ligand free SmoF	SmoF•SQ	SmoF•SQMe
Data collection			
Space group	P 1	P 3 ₁ 2 1	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions a, b, c (Å)	54.21, 78.02, 83.32	102.2, 102.2, 67.96	53.76, 66.27, 99.38
α, β, γ (°)	109.0, 106.9, 104.7	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution (Å)	72.0-1.30 (1.37-1.35)	88.5-1.80 (1.84-1.80)	49.6-1.59(1.62-1.59)
R _{merge}	0.029 (0.832)	0.179 (2.77)	0.280 (1.38)
R _{pim}	0.015 (0.637)	0.06 (0.93)	0.148 (0.909)
I / σI	13.1(1.1)	10.1 (1.3)	7.7 (1.3)
CC1/2	1.0 (0.91)	1.0 (0.65)	0.98(0.68)
Completeness (%)	99.1(67.5)	100 (100)	99.4 (96.1)
Redundancy	3.3(2.5)	19.2 (19.1)	6.9 (5.6)
Refinement			
Resolution (Å)	1.30	1.8	1.59
No. unique reflections	270846	38847	48233
R _{work} / R _{free}	0.18/0.20	0.20/0.23	0.24/0.27
Protein atoms	8977	5740	5796
Ligand/ion atoms	8	27	30
Water atoms	1470	138	299
B-factors (Å²)			
Protein	16.6	32	20
Ligand/ion	30.7	23	18
Water	27.6	34	27
R.M.S. deviations			
Bond lengths (Å)	0.0135	0.0146	0.0143
Bond angles (°)	1.911	1.85	1.79
Ramachandran Plot Residues			
Most favorable regions (%)	98.61	98.4	98.1
In allowed regions (%)	1.29	1.6	1.6
Outliers	0	0.0	0.3
PDB code	7NBZ	7YZS	7YZU

Table 2-1: Data collection and refinement statistics for SmoF ligand free, SQ and SQMe. Values in parentheses are for the highest resolution shell. All values are within I/σI where applicable.

	SmoF•SQOct	SmoF•SQDG-(C4:0/C16:0)
Data collection		
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁
Cell dimensions a, b, c (Å)	53.78, 65.85, 107.0	53.22 69.59 104.57
α, β, γ (°)	90.0,90.0,90.0	90.0, 91.54, 90.0
Resolution (Å)	56.1-1.80 (1.84-1.80)	69.6-2.14 (2.20-2.14)
R _{merge}	0.116 (0.470)	0.093 (0.305)
R _{pim}	0.065 (0.262)	0.081 (0.265)
I / σI	10.1 (3.3)	8.9 (3.6)
CC1/2	0.965 (0.952)	0.99(0.94)
Completeness (%)	100 (100)	99.9 (100)
Redundancy	8.0 (7.9)	4.1 (4.1)
Refinement		
Resolution (Å)	1.80	2.14
No. unique reflections	36036	24477
R _{work} / R _{free}	0.23/0.27	0.20/0.26
Protein atoms	2850	11648
Ligand/ion atoms	23	142
Water atoms	243	239
B-factors (Å²)		
Protein	21	27
Ligand/ion	19	26
Water	28	29
R.M.S. deviations		
Bond lengths (Å)	0.0145	0.0075
Bond angles (°)	1.85	1.50
Ramachandran Plot Residues		
Most favorable regions (%)	97.04	97.8
In allowed regions (%)	2.7	2.2
Outliers	0.27	0.0
PDB code	-	7QHV

Table 2-2: Data collection and refinement statistics for SmoF•SQOct and SQDG-(C16:0/C4:0). Values in parentheses are for the highest resolution shell. All values are within I/σI where applicable.

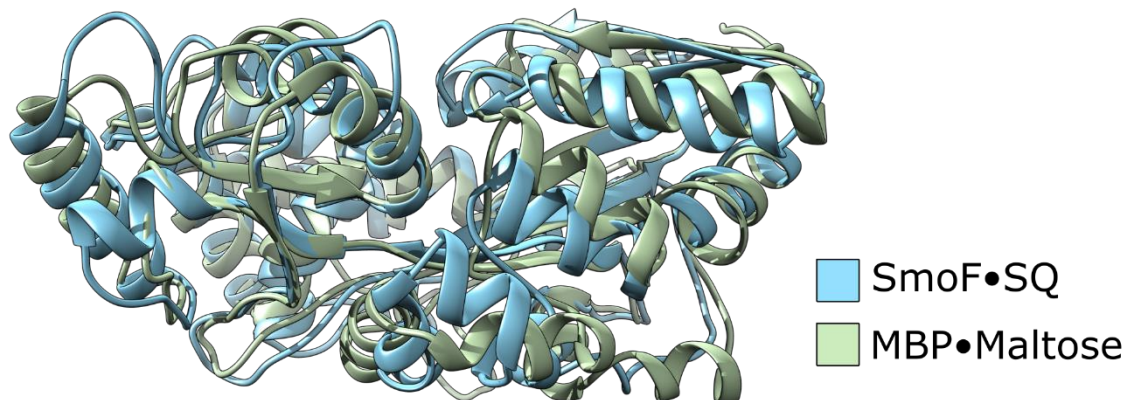


Fig 2-10: Crystal structure of SmoF•SQ, superposed with MBP•Maltodextrose. SmoF in blue, MBP in green.

were large enough to successfully harvest and well-ordered enough for data collection (fig 2-9) The resultant dataset was solved by molecular replacement using a structure of SmoB-SQGro (collected by the lab of Spencer Williams). To account for the conformation change the structure was dissected into the constituent domains and the dataset solved one domain at a time. The resultant structure is in P1 and contains 3 monomers in the ASU. SmoF shows a similar domain arrangement to other SBPs such as MBP, with two primarily α -helical “clamshell” domains linked together. These domains are not self-contained, rather forming in a non-sequence contiguous manner (fig 2-10). In SmoF, SQ adopts a 4C_1 chair conformation and exists as the α -anomer. Binding is achieved through both the C6 sulfonate and C1-4 hydroxyl groups. The sulfonate occupies a pocket comprised of Gln₁₂-Ser₄₃-Gly₁₆₆-Thr₂₂₀. The side-chain nitrogen of Gln₁₂ and side-chain hydroxyl of Thr₂₂₀ form a hydrogen bond with the first sulfonate oxygen (2.6 Å, 2.9 Å) and the second sulfonate oxygen forms a hydrogen bond with the backbone secondary amine of Gly₁₆₆ (2.8 Å). The third sulfonate oxyanion forms a hydrogen bond to the backbone secondary amide of Ser₄₃ (2.9 Å) as well as an ordered water (3.0 Å). The C1 hydroxyl forms a hydrogen bond with a side-chain secondary amine on His₁₃ (2.7 Å). The C2 hydroxyl group forms hydrogen bonds to the side chain carboxyl of Asp₁₁₃ (2.7 Å) and the indole nitrogen of Trp₂₇₆ (2.8 Å). The C3 and C4 hydroxyls each bind one nitrogen of Arg₃₄₅ and the carboxyl oxygen of Asp₆₇ (2.9, 2.7, 3.0, 2.5 Å). These interactions are identical to those observed with SQGro, with the only notable differences being in the interactions of Gln₁₂ and His₁₃. These form hydrogen bonds with the proximal hydroxyl of the

glycerol moiety and an ordered water respectively. The water in turn forms a hydrogen bond to the same sulfonate oxygen as Gln₄₆. These extra interactions may provide a structural basis for the higher binding affinity seen with SQGro as opposed to SQ. Much like SQGro, SQ is totally enclosed by interacting residues.

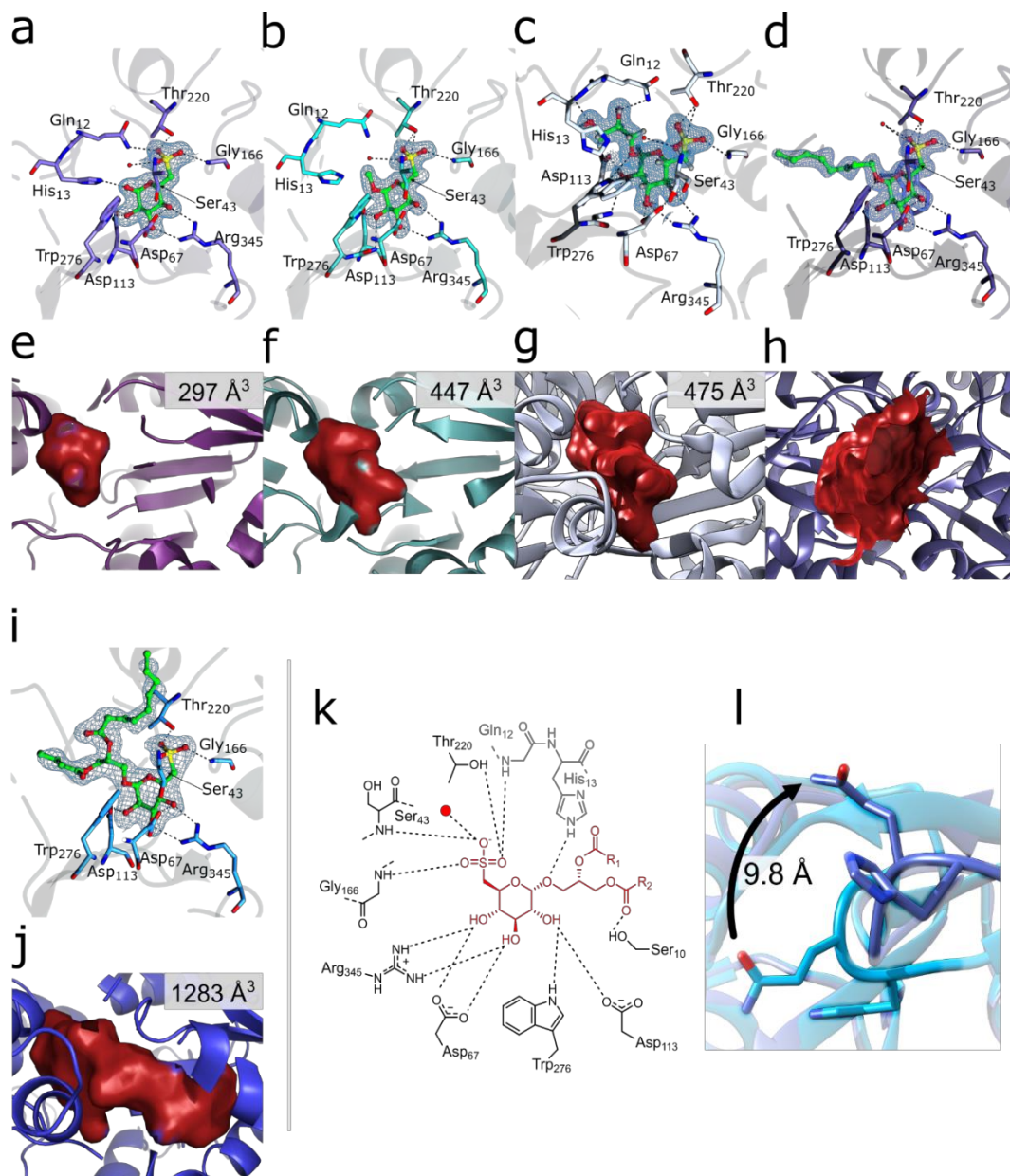


Fig 2-11: Binding mode and accommodation of SmoF. Ligand binding site of SmoF with A. SQ, B. SQMe, C. SQGro (data included as a comparative point from Sharma *et al*, 2022), D. SQOct. Internal cavity produced with E. SQ, F. SQMe, G. SQGro, H. SQOct. I, J. Ligand binding site and internal cavity respectively respectively of SmoF•SQDG-(C4:0/C16:0). K. generalised ligand binding schematic of SQ and derivatives with SmoF. Residues in grey bind SQ but no other ligand. L. Movement of QH loop between SmoF•SQ (cyan) and SmoF•SQDG-(C4:0/C16:0) (blue). All maps are 2Fo-Fc and contoured to 1.0 σ (SmoF-SQ, 0.44 e/Å²; SmoF-SQMe, 0.61 0.44 e/Å²; SmoF-SQGro 0.44 0.44 e/Å²; SmoF-SQOct 0.61 0.44 e/Å²; SmoF-SQDG-(C16:0/C4:0), 0.44 0.44 e/Å²).

2.4.4. SmoF binds C1-methylated and octylated SQ derivatives

Crystals of SmoF with SQMe diffracted to 1.55 Å, and a 1.50 Å crystal structure was also obtained with SQOct. In both cases SmoF adopts the closed conformation. In both cases the ligand occupies the same general location and has the same conformation and anomeric state as with SQ and SQGro. The acyl chain of SQOct is well resolved for a chain of its length, adopting an unambiguously extended conformation. SQMe and SQOct form almost identical interactions with SmoF as SQGro. The presence of the C1 methyl/octyl group has displaced the loop containing Gln₁₂ and His₁₃ such that Gln₁₂ no longer interacts directly with the 6C sulfonate. In both structures it may interact instead with an ordered water, which in turn forms a hydrogen bond to a sulfonate oxygen. His₁₃ no longer forms any hydrogen bonds to either ligand.

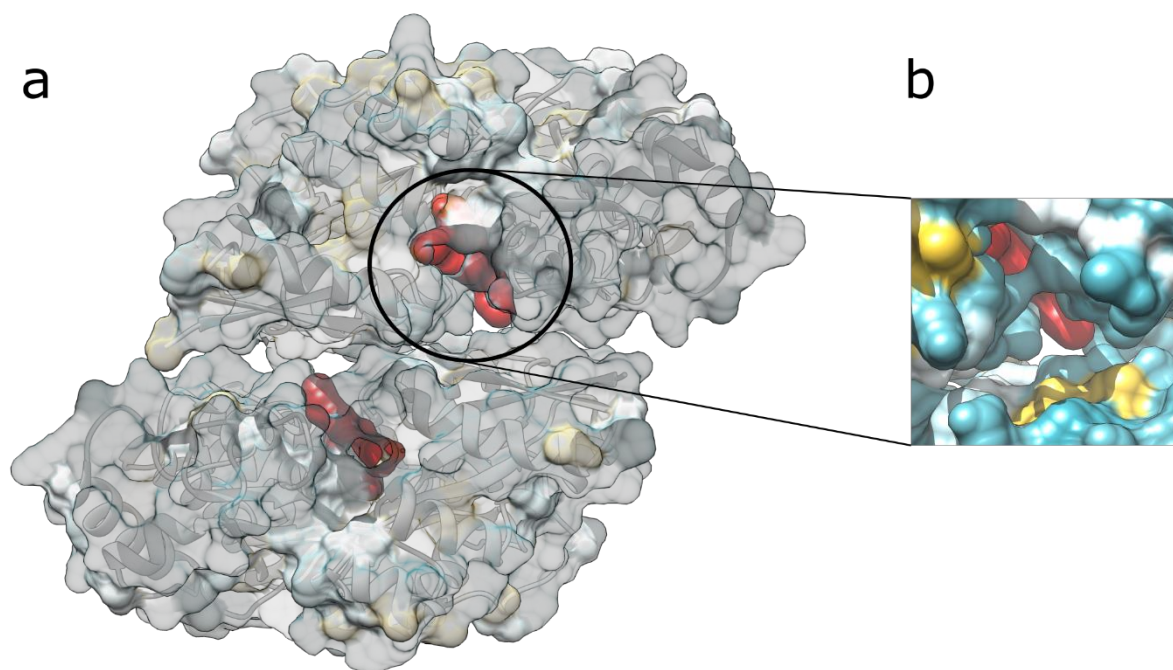


Fig 2-12: Crystal packing observed in SmoF•SQDG-(C18:1/C16:0). **A.** Surface view of the asymmetric unit seen in the SmoF•SQDG-(C18:1/C16:0) crystal structure. SQDG-(C18:1/C16:0) highlighted in red. **B.** Inset view of ligand. All electron density is from 2Fo-Fc maps, and contoured to 1.0 σ .

2.4.5. *SmoF and SQDG*

Initially, co-crystallisation was attempted with both ligands. For both, crystals were grown at 6 °C as it was reasoned the large acyl chains would be mobile, and a lower temperature may slow the kinetics of this movement sufficiently for crystal growth. Datasets obtained through co-crystallisation with 2 mM SQDG-(C16:1/C18:0) and solved by MR resulted in ligand-free structures in the open conformation. No crystal grown including SQDG-(16:1/C18:0) resulted in acquisition of a ligand-bound structure.

SmoF adopts the closed conformation with SQDG-(C4:0/C16:0), though the ligand is not entirely enclosed by the protein as observed with previously tested ligands (fig 2-11, a-d). SQDG-(C4:0/C16:0) binds SmoF with the same pattern observed with SQMe, though the interactions with Gln₁₂ and His₁₃ are notably absent.

2.4.6. *SmoF undergoes conformational changes to accommodate C1 SQ glycosides*

SQ, SQGro and SQMe all reside within a sealed cavity in SmoF. The internal volume displaced for SQ is 297 Å³. The pocket for SQGro is 476 Å³, 37% larger than SQ. For SQMe the pocket is 447 Å³, 33% larger (fig 2-11, e-h). In all these cases the extra internal volume arises from conformational changes in the internal structure, with a particularly large deflection observed in the loop containing Gln₁₂ and His₁₃ (the QH loop). SmoF does not fully enclose SQOct, with the C1 octyl moiety residing in a deep cleft leading towards the binding site. This cleft is large enough that use of the CASTp server did not result in a meaningful cavity being detected. Despite this, SmoF does adopt the closed conformation, with few differences to the overall interdomain closure. In the complex with SQDG-(C4:0/C16:0), the QH loop is displaced as far as 9.8 Å away from its location in the SQ complex. This movement is facilitated by an upwards movement of helix α 1, and results in a total internal volume of 1283 Å³. The acyl chains then occupy the space left by this movement, though as with SQOct this is not sufficient to allow full ligand enclosure. Instead, the crystal structure shows the palmitoyl chain protruding from one monomer and being partially enclosed by another along a crystal contact (fig 2-12). This state is unlikely to be reflective of a solution structure as SBPs are monomeric, so the conformation of this chain in solution remains unknown. A superposition of all ligand-bound structures of SmoF shows that α -helix 1 and especially the QH loop adopt the most varied

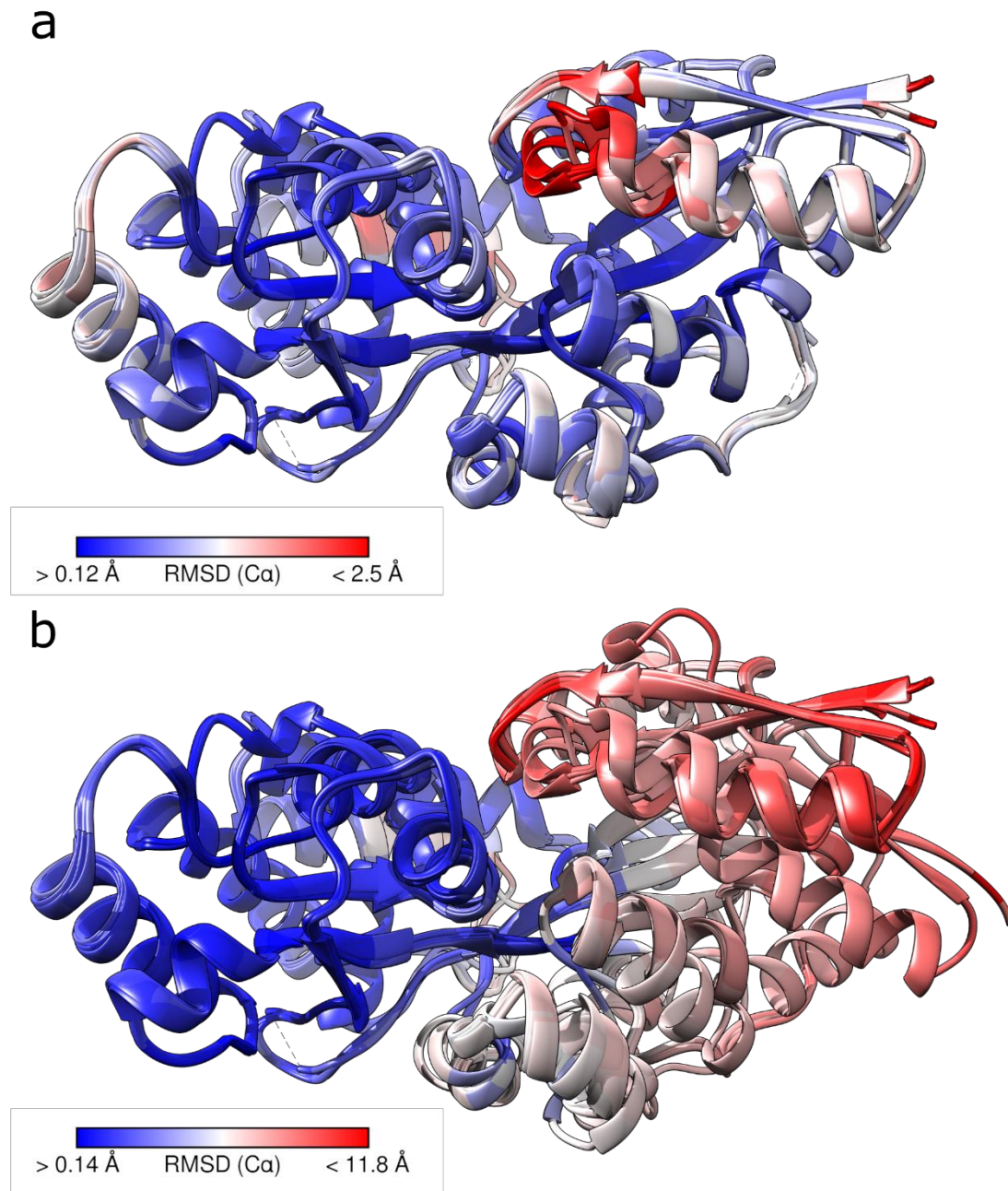


Fig 2-13: Ligand-induced conformation changes in SmoF. Heatmaps of RMSD between **A.** All ligand-bound states **B.** Ligand-free and SQ-bound SmoF. In both cases blue indicates low RMSD between conformers, red indicates high.

conformations. In contrast, most of the clamshell domains do not adopt any ligand-specific conformations (fig 2-13).

A common trait in SBPs is the large-scale conformational changes they undergo upon ligand binding. When complexed with any of the ligands texted except SQDG-(C4:0/C16:0), SmoF undergoes a hinge motion centred on the flexible domains Thr₂₇₁-Leu₂₇₃ and Ser₁₁₄-Met₁₁₅-

Phe₁₁₆, pivoting each clamshell domain to fully enclose the ligand (fig 2-14). During this motion the individual clamshell domains undergo few small-scale changes, with by far the largest non-hinge change in the QH loop. Using the DynDom server the motion from the open (ligand-free) and closed conformations is estimated at 31-34°. The ligand binding site is comprised of residues from both clamshell domains, allowing for the observed ligand enclosure. Despite failing to fully enclose SQOct, SmoF does undergo a comparable interdomain motion to that seen with smaller ligands at 34°.

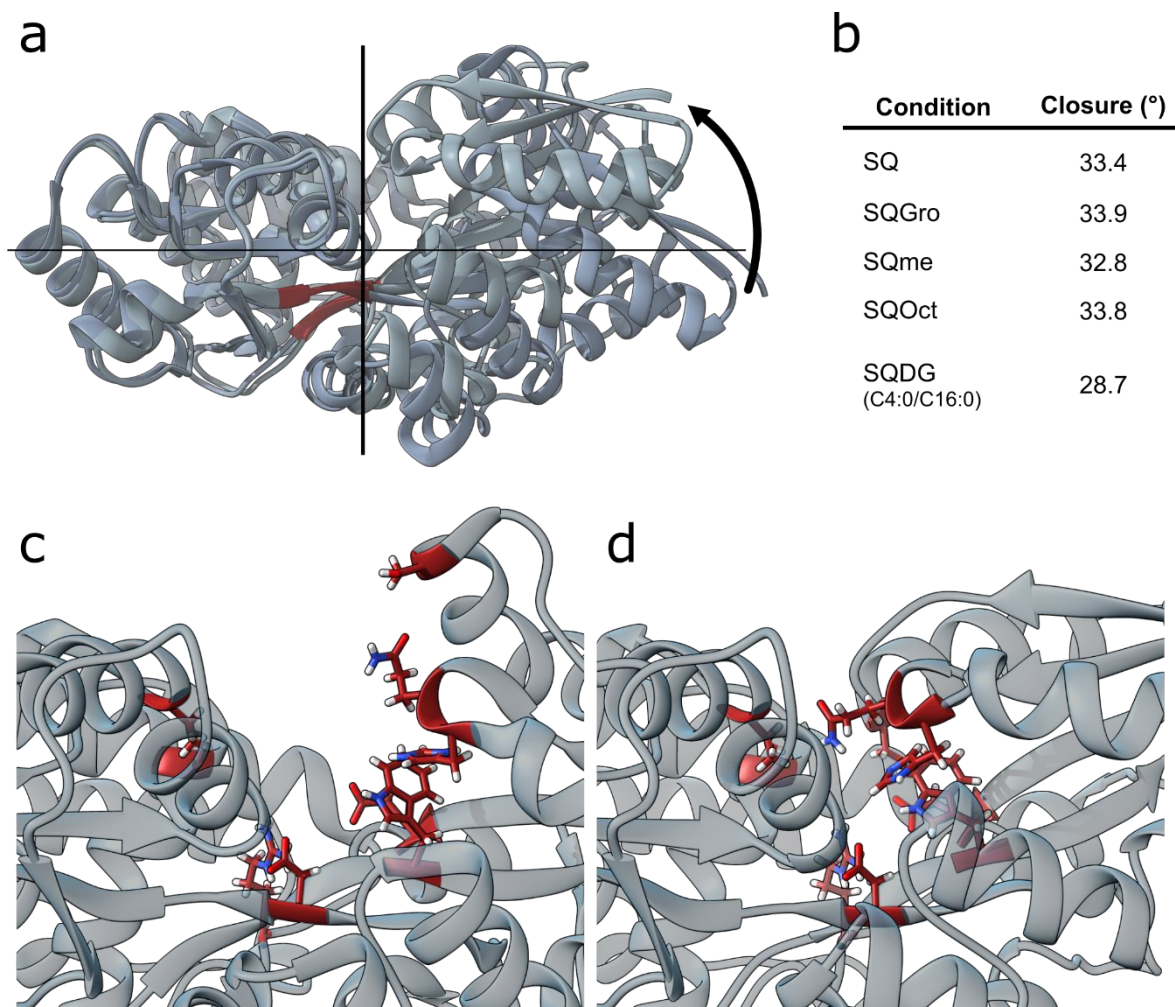


Fig 2-14: Large-scale “hinge” motion of SmoF. As analysed by the DynDom server. **A.** Comparison of ligand-free (darker) and SQ-bound (lighter) SmoF. Residues forming the hinge regions highlighted in red. **B.** Table of closure angles for SmoF with ligands, as compared to the ligand-free conformation. **C-D.** Comparison of the ligand binding residues between the **C.** ligand-free and **D.** SQ-bound protein. All residues forming hydrogen bonds with SQ are highlighted in red.

2.4.7. SmoF bioinformatics

The structural analyses thus far conducted have identified a sulfonate binding pocket. To test whether this was a conserved feature (and by extension, a possible tool for identifying more SQBPs) known examples of sulfoglycolytic operons featuring specific transport systems were compared (fig 2-15, a). Sequences were compared for *Agrobacterium tumefaciens*, *Rhizobium oryzae*, *Rhizobium leguminosarum*, *Vibrio barjaei*, *Microlunatus phosphovorus*, *Neorhizobium galagae*, *Tetrasphaera*, *Clostridium pastueranium*, *Thermotoga neopolitana*, *Tenericutes* and *Dictyoglomus tugidium*. These encode genes for sulfo-ED, sulfo-EMP, sulfo-TAL, sulfo-TK and sulfo-SMO pathways. Thus, all the genes for SBPs in these operons are likely to be SQBPs. A cladogram shows SmoF has the greatest sequence similarity to other SBPs in *Rhizobiales*, with a second branch containing all the other genes with the exception of *Clostridium pastueranium*, which is not featured in either group (fig 2-15, b). Of particular note are the sulfo-ED operon in *R.leguminosarum*, which contains an ABC SBP with 79% sequence identity to SmoB and the *V.barjaei* sulfo-EMP operon, featuring a TRAP system with no structural characterization. This binding protein with accession code OIN26723 is predicted by AlphaFold to form a classical TRAP binding protein architecture (fig 2-15, c) ¹³. Docking SQ into this structure yields an approximation of a complex, which would feature a sulfonate binding pocket of Asn₃₈-Gln₈₇-Glu₉₂. C2-4 hydroxyl stabilisation would involve either Arg₁₅₁ or Arg₁₇₂ with a similar interaction not that seen in SmoF (fig 2-15, d). Tyr₃₇ and Asp₂₁₆ may form hydrogen bonds to the C1-2 hydroxyls. Finally, a π - π stacking interaction could occur to Trp₁₉₅. These predictions hint at an alternate form of sulfosugar recognition. To aid in confirming this, a multiple-sequence alignment was also performed with all the pathways. Also included here was TauE, a taurine binding protein from *E.coli* K-12. TauE is a part of the 4-toluene sulfonate uptake permease (TSUP) family which is also found in conjunction with sulfo-EMP and sulfo-ED pathways ^{40,41,137}. It was included in order to determine whether a common mode of sulfonate recognition exists between the two importer families (fig 2-16). Residues in SmoF involved in sulfonate recognition are conserved consistently only in *Rhizobium oryzae* and *R.leguminosarum*. The other member of *Rhizobiales* featured, *N.galagae*, did not contain any of these residues. Partial conservation was also observed in *C.pasteranium*, *T.neopolitana* and *Dictyglomas*. From this, the mode of sulfonate recognition appears to be poorly conserved among putative SQBPs except for those in *Rhizobiales*, but not consistently even then. It is

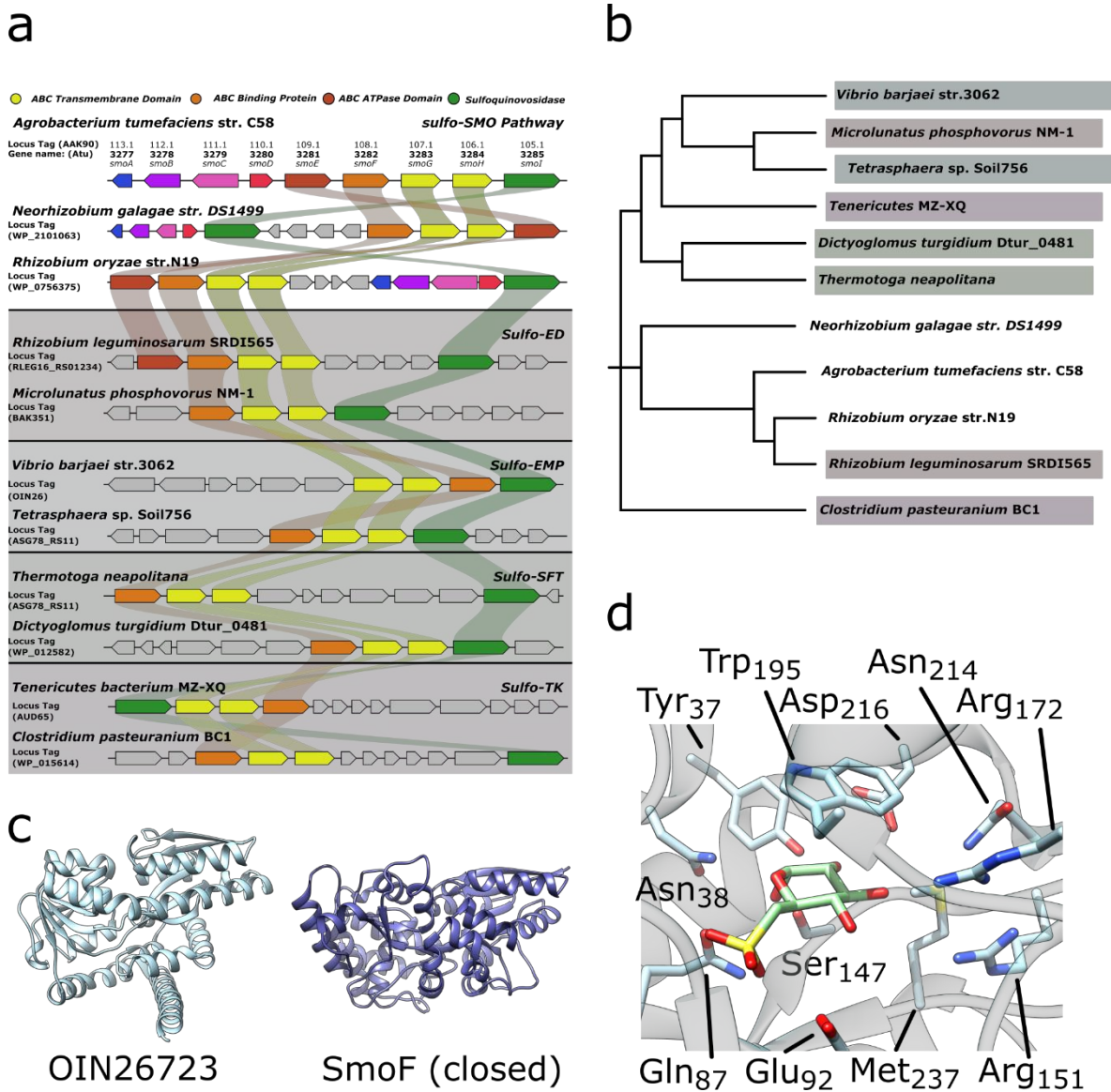


Fig 2-15: Sulfoglycolytic pathways containing putative SQBPs. A. Operons containing sulfoglycolytic operons featuring import systems. Pathways have been sorted by the pathway type they are thought to represent. Colouration of Importer components and sulfo-SMO pathway enzymes is consistent with fig 2-4. **B.** Cladogram of binding proteins found in a. **C** Comparison of AlphaFold prediction of a novel SQ-specific TRAP transporter in *V. barjaei* (accession code OIN26723, cyan) and SQ-bound SmoF (blue) **D.** Result of docking SQ into the TRAP structure in c.

likely sulfonate recognition in SQBPs is an example of independent evolution, and therefore searches for sulfosugar binding proteins by recognition motif is unlikely to be a useful method of sulfoglycolytic pathway discovery, unlike SQases such as YihQ⁵⁵.

2.5. Conclusion and future directions

Based on two independent solution assays, SmoF clearly binds sulfoquinovosyl glycerol, and as such can be confirmed as the prototypical sulfoquinovosyl binding protein (SQBP) and the

first step in the sulfo-SMO pathway. Additionally, SmoF can bind SQ as well as C1 methylated and octylated SQ derivatives, as well as SQDG-(C4:0/C16:0). Crystal structures of SmoF with all these ligands provide a clear method for recognition of the C6-sulfonate as well as a common recognition method for the C2-4 hydroxyl groups. SmoF undergoes small-scale conformation changes within loops proximal to the binding site to accommodate the varied C1 moieties tested. Ligand recognition is accompanied by a large interdomain rotation, between 31° and 33° to a closed conformation, in comparison to the open conformation observed with the ligand-free protein, for which a crystal structure was also obtained. The residues involved in ligand binding are spread across both domains, coming together to entirely enclose SQ, SQGro and SQMe. SQOct and SQDG-(C4:0/C16:0) are too large to allow full enclosure, though in both cases a closed conformation is achieved. SmoF is unable to bind non-sulfonated α -D-hexoses and other sulfonated solutes, limiting the binding promiscuity to

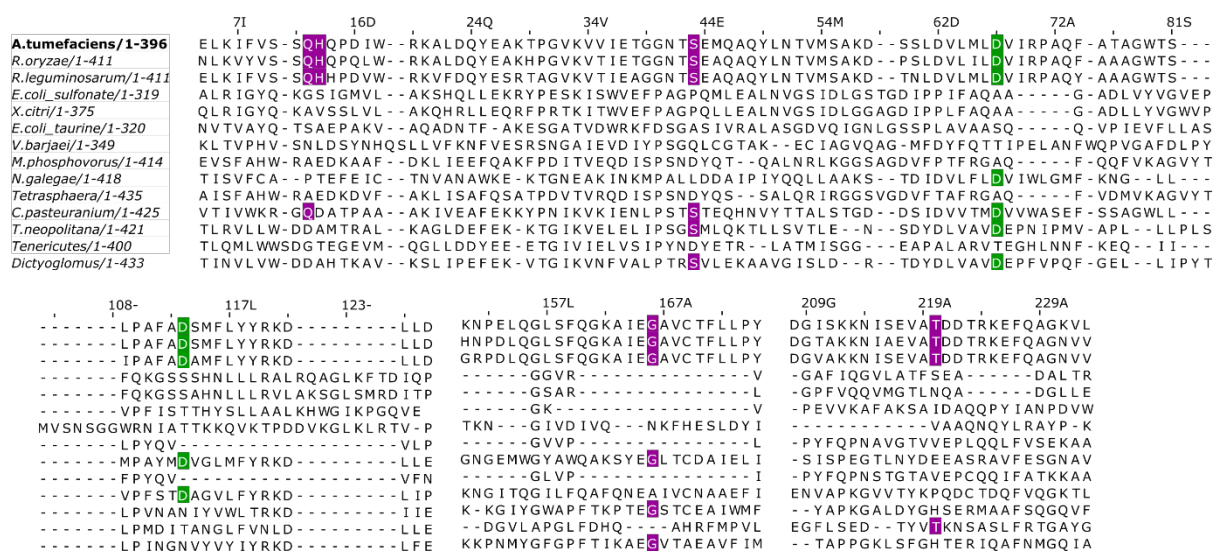


Fig 2-16: Multiple sequence alignment of SmoF and putative SQBPs. Residues involved in sulfonate binding are highlighted in purple, and residues binding the C2-4 hydroxyl groups are highlighted in green.

SQ and derivatives. This capacity to bind SQ derivatives with long acyl chains may allow for an SQ-based protein purification method, analogous to the MBP column system. The capacity to bind diverse SQ derivatives may also indicate a capacity for their transport by the ABC transporter encoded in the sulfo-SMO operon in *A.tumefaciens* C58. Putative SQBPs exist in a range of bacteria, and across all currently known sulfoglycolysis pathways, though a lack of consistent sequence conservation in the motifs responsible for ligand recognition in SmoF means this may not provide a useful bioinformatic tool for sulfoglycolytic pathway discovery. Rather, SQ glycoside recognition and import may be achieved by a range of importer types including TsuP-family importers and TRAP transporters. This has yet to be confirmed experimentally, however.

3. Chapter 3: 6-oxo-glucose reductase SmoB

3.1. Abstract

Aldo-keto reductases are a common superfamily of oxidoreductases that enact the reduction of a substrate by oxidizing a cofactor, or vice versa. In order to achieve this process in a controllable and specific manner, AKRs show a variety of substrate specificities, though a conserved catalytic tetrad of His-Tyr-Lys-Asp. They can utilize either NADH or NADPH exclusively or be able to use both. The sulfo-SMO pathway in *A. tumefaciens* C58 contains an AKR, SmoB which represents the last enzymatic step of the pathway. Here, SmoB is confirmed to be a 6-oxo-glucose reductase, which in turn confirms glucose as the final product of the sulfo-SMO pathway. The enzyme is observed as a solution trimer with the TIM-barrel fold common to the AKR superfamily, and structural characterization by X-ray crystallography was used to produce ligand-free, NADPH-bound and NADPH, glucose-bound structures. These show an extended, *anti*- conformation for NADPH across the central TIM barrel pore. Glucose is then bound near the nicotinamide moiety of NADPH, and in position for a standard “push-pull” mechanism using the catalytic tetrad. NADPH specificity was also confirmed in solution through isothermal titration calorimetry. Overall, these data provide a solid baseline for SmoB fold and function, and the exact nature of cofactor specificity could be explored further.

3.2. Introduction

3.2.1. Oxidoreductases

The chemical processes of oxidation and reduction are ubiquitous in biochemistry, and occur on a vast variety of substrates. As such, a variety of enzymes exist with varying mechanisms and specificities to handle it. These come under the umbrella of the oxidoreductase – an

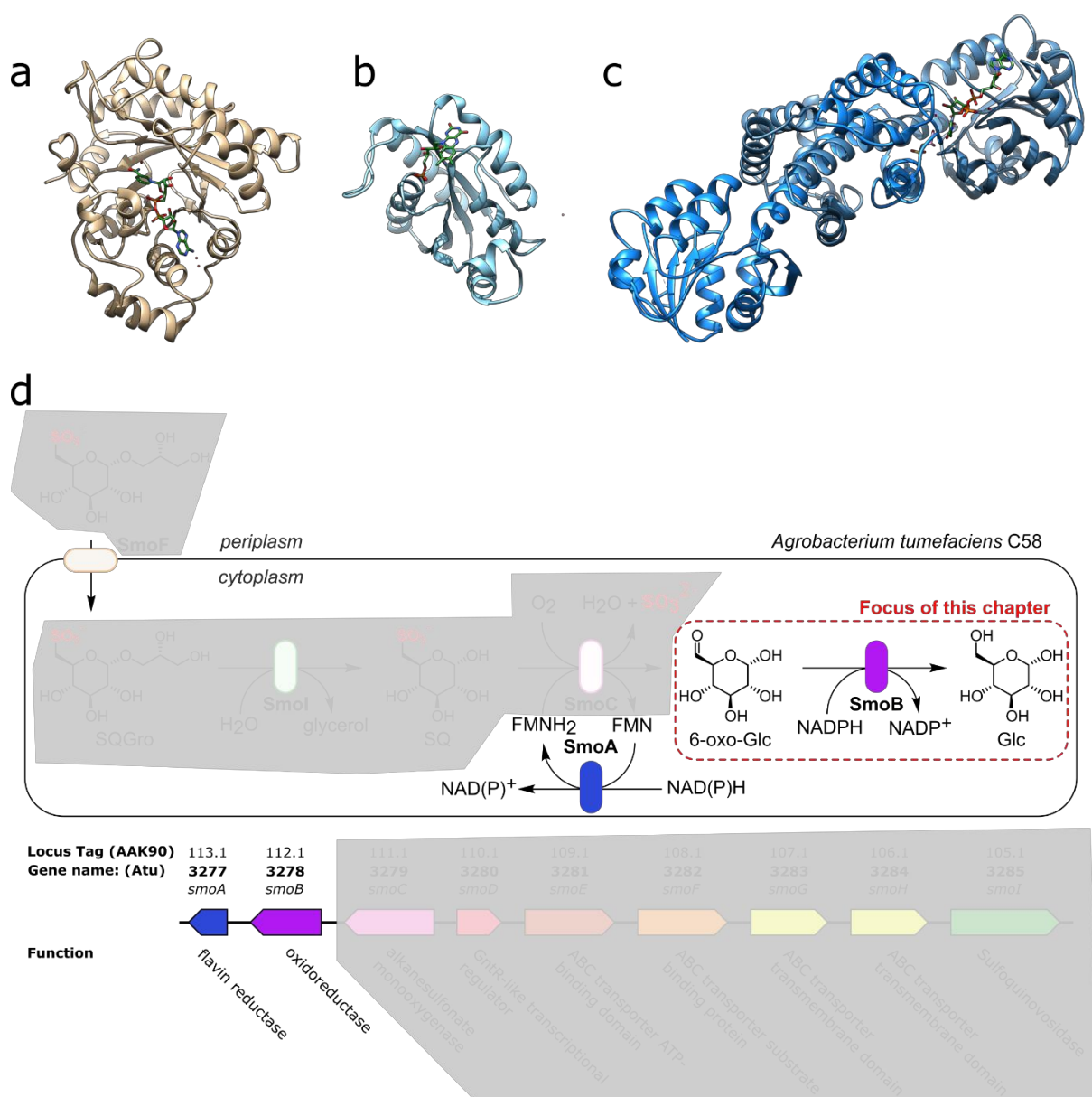


Fig 3-1: Oxidoreductases across nature and sulfoglycolysis. Structures of **A.** Xylose reductase **B.** YihU **C.** SsuE. PDBs: 1MI3, 6SMY, 4UP0. All structures are depicted with cofactors (NADH, FMN, NADPH) bound. **D.** The sulfo-SMO pathway as seen in *A. tumefaciens* C58. All enzymatic steps are greyed out with the exception of SmoA and SmoB, which is highlighted as the focus of this work.

enzyme that couple a redox transfer between a substrate and cofactor, which will be a source or sink for a hydrogen depending on the enzyme direction. The majority of oxidoreductases use an adenine dinucleotide cofactor – NADH, NADPH or FADH₂^{138,13}. This variety in cofactor specificity is highly advantageous to the cell as it allows for concurrent redox actions within the same compartment while maintaining independent control of each. The subsequent lack of “expensive” isolation systems such as membrane bound organelles or juxta-membrane localisation gives this system an energetic advantage, particularly in bacteria where such systems are often out of reach anyway. These factors synergise where membrane compartments and localisation to them are possible, such as in mitochondrial electron transport. Such cofactor specificity can be completely rigid, though many oxidoreductases exist capable of using either NADH or NADPH, with a difference in affinity often giving the enzyme a preference for one cofactor over the other¹⁴⁰.

3.2.2. Redox reactions in sulfoglycolysis

Redox steps are commonly found in both EMP and ED glycolysis, so the appearance of oxidoreductases in the analogous sulfo-EMP and ED pathways was no surprise. Generally, these are short-chain dehydrogenases found after the (trans)aldolase/ketolase steps in the sulfo-EMP, sulfo-ED, sulfo-TAL and sulfo-TK pathways^{40,41,75,78}. A prototypical example is YihU, a sulfolactaldehyde reductase found in the sulfo-EMP operon of *E.coli* k12 (Figure 3-1 b). YihU is a member of the β -hydroxyacid dehydrogenase (β -HAD) family, which employ a Rossmann fold⁸⁴. The sulfo-SMO pathway employs a monooxygenase (SmoC) and FAD reductase (SmoA) as well as SmoB. These enzymes work much like the SsuD-SsuE system in *E.coli* (Figure 3-1 c), with SmoA using NADH to supply FMNH₂ to SmoC, using a hydride ion transfer mechanism to do so. Then SmoC effects the oxidative cleavage of the C6 sulfonate group from SQ, though the exact mechanism for this remains unclear⁸⁰. SmoB can be found at the very end of the sulfo-SMO pathway, as observed in *A.tumefaciens* C58. The gene is upregulated based on the presence of environmental SQ, much as the other enzymes of the pathway are, and based on the actions of these enzymes together (Figure 3-1 d), SmoB is most likely a 6-oxo-glucose (6-OG) reductase or oxidase, producing glucose or glucuronic acid (GA) for further use by the cell. A distinct difference observed between SmoB and any other redox

enzyme in these pathways is that of the ones that directly work on the substrate it is the only one that acts on a C6 species with no sulfonate group. Therefore, SmoB is not expected to have any motifs that allow for sulfonate recognition, unlike any of the other enzymes featured in this thesis.

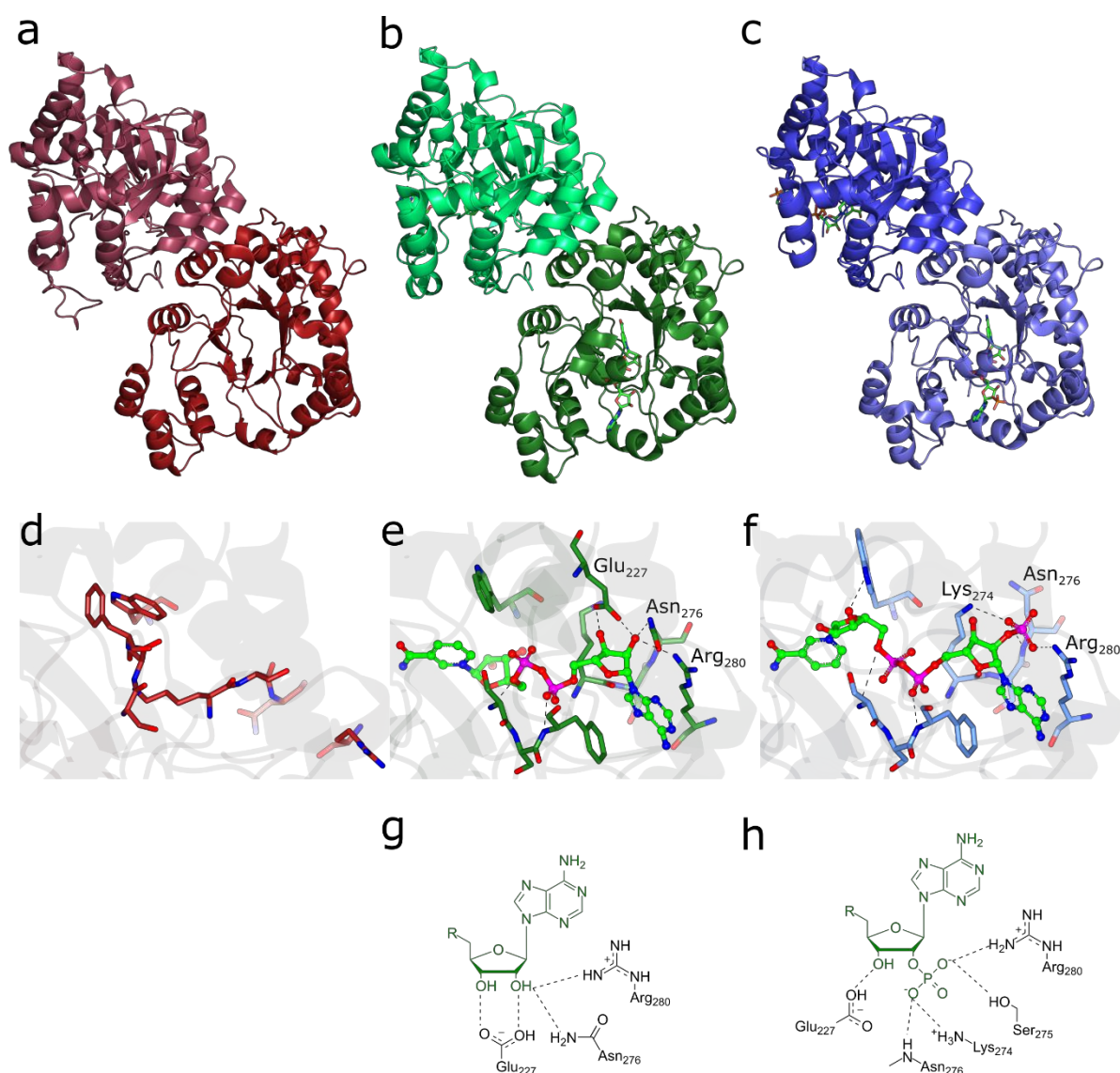


Fig 3-2: Fold and structural basis of cofactor dual-specificity of xylose reductase. Crystal structures of dimeric XR in **A.** ligand-free **B.** NADH **C.** NADPH-bound forms. **D-F.** Cofactor binding site of XR ligand-free, with NADH and with NADPH. **G-H.** 2D representation of the 2' phosphate binding pocket with NADH, NADPH respectively.

3.2.3. Aldo-keto reductases

Based on its sequence, SmoB belongs to the aldo-keto reductase (AKR) superfamily. Unlike many nucleotide-binding proteins, which typically contain a Rossmann fold¹⁴, AKRs feature an $\alpha_8\beta_8$ Triose phosphate isomerase (TIM) barrel fold¹⁴². While this fold remains a common feature, the superfamily shows a high variation in sequence. This has led to the development of a dedicated AKR naming convention^{143,144}, with families and subfamilies defined by 40% and 60% sequence identity respectively, and noted with a number, then a letter. Finally, a number is added for a unique identifier between similar AKRs. For example, xylose reductase from *Candida tenuis* (XR, Figure 3-1 a) has the designation AKR2B5: AKR superfamily, family 2, subfamily B, unique protein 5¹⁴⁵. Many AKRs have been submitted to a curated database for easy comparison and reference¹⁴⁶ (Figure 3-3). As of 2022, there are 14 known AKR families. Proteins within these families often have similar functions including hydroxysteroid dehydrogenases (AKR1), mannose reductases (AKR2) and aflatoxin reductases (AKR6)¹⁴⁷.

They are also often monomeric, and 34-37 kDa in size^{148,14}, though some exist as dimers¹⁴⁰, trimers¹⁵⁰, tetramers¹⁵¹ and even octamers¹⁵². Multimer formation is typically achieved through hydrophobic protein-protein interfaces, as is common in proteins of all folds throughout nature. However, it is not a universal property in AKRs. AKR2B5 is a solution dimer with a network of hydrogen bonds between monomers (Figure 3-2 a-c). The interface is polar enough that a crystal structure showed it including 11 ordered waters. Analytical SEC showed no monomeric populations, meaning this interface to be strong enough to exist in a polar, buffered environment rather than being an exclusively crystallographic phenomenon^{140,145}.

Of the (as of 2015) over 190 known AKRs, the majority exclusively use NADPH as a cofactor. Despite large variation being present among AKR sequences, the active site is positionally conserved near the C-terminus. This is thought to be an indication of common ancestry in the superfamily, as there is little reason for this to arise through convergent evolution¹⁵. Some proteins of the AKR1,2 and 6 families have dual-specificity for NAD(P)H. One such protein is AKR2B5. It has a 5-fold preference for NADPH, with a K_M of 4.8 μ M versus 25.4 μ M for NADH¹⁴⁵. A closer look at the binding site reveals that the general pattern for NADH and NADPH binding is similar, however the 2'-phosphate group in NADPH requires three further residues to provide adequate hydrogen bonding, which may provide an explanation for the preference (Figure 3-2 d-h). Unlike the other common binding residues, Glu₂₂₇ is notable as it adopts a

different conformation for NADPH, as the phosphate would cause electrostatic repulsion were it to remain in the NADH conformation. Therefore, Glu₂₂₇ must be part of a flexible loop. Such an interaction, whether from one residue or a motif, is common amongst AKRs with dual specificity, and a lack of flexibility would cause the protein to become fully NADH specific¹⁴⁸. Upon cofactor binding, AKRs undergo a conformation change that stabilises the binding and buries the cofactor, to the point where only 20 Å² is solvent-exposed in a human aldose reductase¹⁴⁹. This has been described as a “safety belt” motif¹⁵⁴. AKRs follow a sequential bi-bi reaction model, with the cofactor binding first and leaving last¹⁵⁵. This was further subdivided through fluorescence stopped-flow of rat liver AKR 3α-HSD. This identified a transient NADPH-enzyme complex, which then stabilises through a conformational change to form the binary “fuelled” complex, which can accept substrates. The same transient was not observed with NADH, despite 3α-HSD having dual-specificity. Arg₂₇₆ appears to be crucial to formation of this transient, as mutation to alanine prohibited its formation. An explanation is that the residue forms a part of the 2'-phosphate binding pocket¹⁴². This would make it unlikely to interact with NADH which provides an explanation for why the NADH interaction lacks the transient binding and conformation change seen with NADPH¹⁵⁶. The shift from transient to bound occurs faster than the reduction reaction, and so is not rate-limiting. For both NADH and NADPH, the affinity of the binary complex is high, with a K_D of 80 nM and 50 nM respectively. This high affinity may be advantageous as a driving factor in conferring substrate promiscuity in the class. This is because the energy required to reduce a carbonyl is

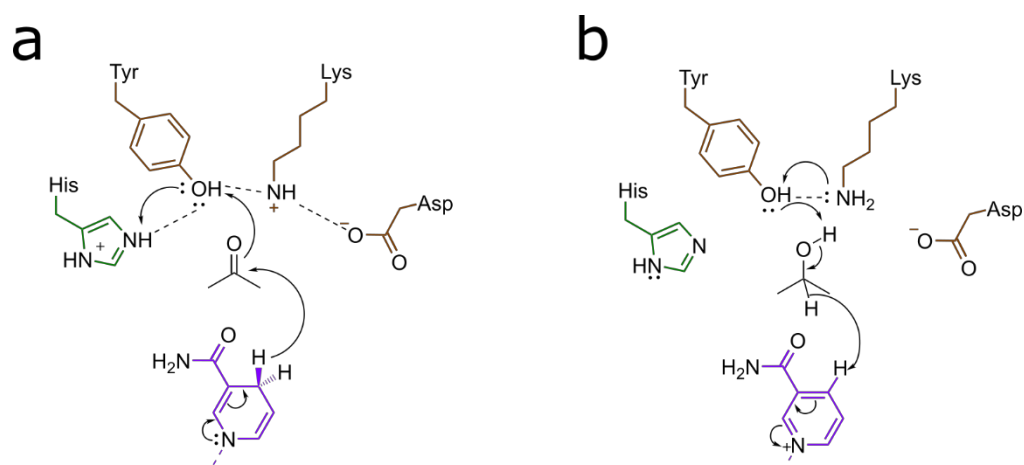


Fig 3-3: Generalised reaction mechanism based on 3α-HSD for AKRs. Mechanism for **A.** reduction **B.** oxidation. Here the active quartet are in brown (His in green), NAD(P)H in purple and the substrate in black.

largely provided by cofactor binding in this case rather than the substrate. Aldehydes have the capacity to be more widely reactive with nucleophiles than alcohols, so AKRs employing this system are useful as they allow rapid and efficient metabolic detoxification to a cell employing them ¹⁵⁷. This is a very common use of human AKRs ¹⁵⁸.

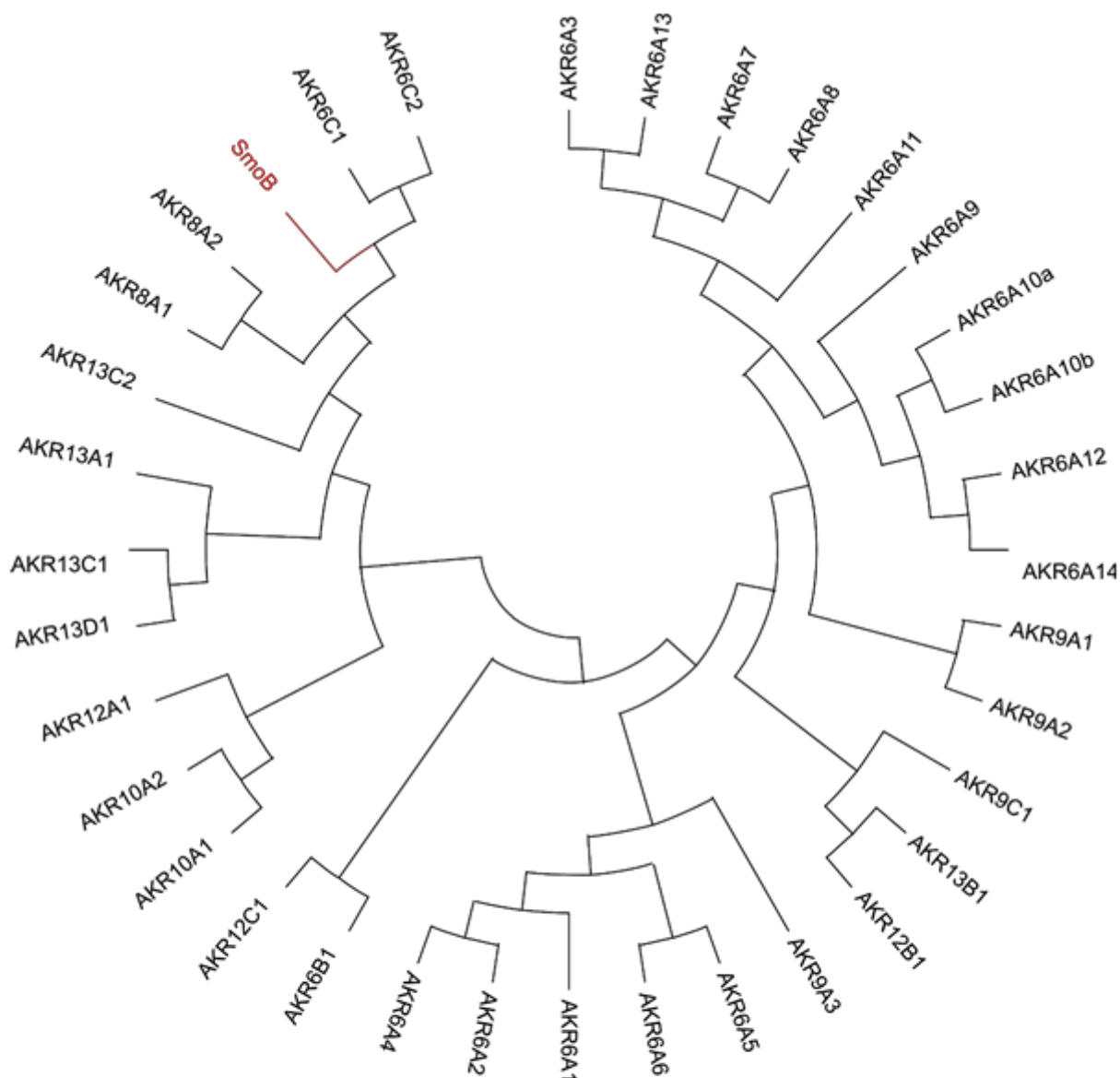


Fig 3-4: Dendrogram of bacterial AKRs, with SmoB included. Alignment was performed based on a maximum-likelihood method, using a JTT substitution model. SmoF is highlighted in red, and all other entries are represented by their AKR nomenclature entries.

3.2.4. AKR mechanisms

AKRs feature a positionally conserved tetrad of active site residues, with all four close by to the nicotinamide moiety of the cofactor. These four residues (Tyr, Lys, His and Asp) are all able to be ionised, with most mechanisms relying on the Tyr acting as a general acid/ base. The pKa of Tyrosine is 10.5, which in isolation would be too high to act as such (a pKa of 7.5-8.5 is required for carbonyl reduction) however it is lowered through the formation of a hydrogen bond to Lys¹⁵. In turn, this forms an ionic bond to Asp. Mutation of the Tyr or Lys to Ala in ARK1C9 resulted in total activity loss, and mutation of any of the active quartet except Tyr resulted in an inability for the substrate to bind (testosterone in this case) though Tyr did not preclude it when mutated. A push-pull mechanism was established for these enzymes, where His₁₁₇ allows proton donation and Tyr₄₈ proton removal¹⁶⁰ (Figure 3-4). Mutation of the Tyr to His results in a reduction in activity by 5 orders of magnitude. A structure of the mutant reveals the new His occupying the approximate space of Tyr in the native active site, in addition to an ordered water which directly occupies the space the hydroxyl in the Tyr would have and hydrogen bonds both the Lys and His. A similar mutation to Phe resulted in a total loss of activity¹⁶¹.

The direction these enzymes work is pH dependent, with a typical reductive pH of 6.0 to 7.4 and an oxidative pH of 9.0^{160,16}. This is a large reason why these oxidoreductases tend to be thought of as "aldo – keto reductases". The reductive reaction involves hydride ion transport from N/FAD(P)H to the substrate, followed by protonation of the catalytic Tyrosine's side-chain hydroxyl, and finally deprotonation of the Tyrosine through protonation of the catalytic Lysine by the same Tyrosine. The oxidative mechanism is essentially the same in reverse: A lone pair on the Lysine causes protonation of the tyrosine, resulting in the nucleophilic attack of the substrate hydroxyl by a lone pair on the tyrosine, causing oxidation of the substrate to a ketone. This in turn causes reduction of the cofactor through hydride ion transport to it¹⁶³.

3.2.5. Chapter Aims

In this chapter, the AKR and putative 6-OG reductase SmoB is structurally and biophysically characterized. This will be performed to determine the multimeric state, cofactor specificity and affinity. Then, steady-state kinetics will be performed to determine whether SmoB can

oxidise glucose to 6-OG, and therefore whether glucose is a likely natural product. Additionally, the structural basis of cofactor and ligand recognition will be explored through X-ray crystallography. This will be useful for determining the status of the active quartet in SmoB, and thus whether the enzyme is likely to employ the same “push-pull” mechanism seen in similar AKRs. Together these data will provide a detailed picture of the final step in the sulfo-SMO pathway.

3.3. Materials and Methods

3.3.1. *Expression and purification of Atu3278*

The gene for SmoB in a pET29a vector was transformed into *E.coli* BL21(DE3) competent cells for expression using the standard NEB protocol and plated onto LB-agar with 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin. These were grown at 37 °C over 18 hours. Expression began with 18-hour pre-cultures in 5 ml LB containing 30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, at 37 °C, shaken at 180 rpm. This was scaled up to 2L expressions through inoculation of two autoclaved 1L LB expression jugs with half of the 5 ml pre-culture each. Kanamycin was added for a final 30 $\mu\text{g}\cdot\text{ml}^{-1}$ concentration. These were grown at 37 °C, 225 rpm until an OD600 of 0.6-0.8 was reached. Induction then took place with the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final 1 mM concentration, and the temperature and shaking were lowered to 18 °C, 180 rpm for 18 hours. The cells were harvested by centrifugation at 5000 g for 20 minutes, and resuspended in binding buffer (50 mM NaPi, 300 mM NaCl, 30 mM Imidazole, pH 7.4). Pellets could be snap frozen and stored in -80 °C or lysed at this point. Cell lysis used a disruptor at 25 MPa, with the lysate collected on ice. Debris was pelleted at 18,000 g for 30 minutes at 6 °C. The pellet was discarded. The soluble fraction loaded to a pre-equilibrated 5 ml Ni-NTA HisTrap column, using a sample pump at 5 $\text{ml}\cdot\text{min}^{-1}$. The column was then washed with binding buffer at 5 $\text{ml}\cdot\text{min}^{-1}$ until the A280 read less than 20.0 mAU. His-tagged protein was eluted at 5 $\text{ml}\cdot\text{min}^{-1}$ in a linear gradient of increasing elution buffer (50 mM NaPi, 300 mM NaCl, 250 mM Imidazole, pH 7.5). Fractions were collected in a 96-well deep block. Fractions identified as containing protein of the correct size by SDS-PAGE were pooled and concentrated in 20k MWCO Vivaspin® concentrators to a 2 ml final volume. This was further purified by size-exclusion chromatography on a HiLoad Superdex 16/600 S200 column. This was performed in 50 mM NaPi, 150 mM NaCl, pH 7.4. The resultant fractions with a high absorbance at 280 nm

were run on SDS-PAGE for purity and size assessment. All purified protein aliquots were flash frozen and stored at $-20\text{ }^{\circ}\text{C}$, at $30\text{ mg}\cdot\text{ml}^{-1}$. Defrosted protein was stored at $4\text{ }^{\circ}\text{C}$ when not in immediate use, for a maximum time of one week. All SDS-PAGEs for all proteins were performed using 12% acrylamide gels and run for 45 minutes at 200 V with 0.1% SDS TRIS/Glycine running buffer. Gels were subsequently stained with Coomassie magic blue dye. Unless otherwise indicated, $7\text{ }\mu\text{l}$ of sample was loaded, with $5\text{ }\mu\text{l}$ loading dye and $3\text{ }\mu\text{l}$ water for a $15\text{ }\mu\text{l}$ total load.

3.3.2. SEC-MALLS

Size-Exclusion Chromatography with Multiple Angle Laser Light Scattering (SEC-MALLS) was performed using $120\text{ }\mu\text{l}$ samples of SmoB at $5\text{ mg}\cdot\text{ml}^{-1}$, in 50 mM NaPi , 300 mM NaCl , pH 7.4. The 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 HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Work was conducted at room temperature ($20 \pm 2\text{ }^{\circ}\text{C}$). Solvent was $0.2\text{ }\mu\text{m}$ filtered before use and a further $0.1\text{ }\mu\text{m}$ filter was present in the flow path. The column was equilibrated with 2 column volumes 50 mM NaPi , 300 mM NaCl , pH 7.4 before use and flow was continued at the working flow rate until baselines for UV, light scattering and refractive index detectors were all stable. Sample injection volume was $100\text{ }\mu\text{l}$. LC Solutions (Shimadzu) was used to control the HPLC and Astra V was used for the HELEOS-II and rEX detectors. The Astra data collection was 1 minute shorter than the LC solutions run to maintain synchronisation. TM 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).

3.3.3. NanoDSF

All SmoB samples were at $1\text{ mg}\cdot\text{ml}^{-1}$ in 50 mM TRIS , 150 mM NaCl , pH 7.5. Glucose, NADH and NADPH were weighed dry and hydrated with 50 mM TRIS , 150 mM NaCl , pH 7.4 to a stock concentration of 50 mM . All ligands and cofactors were added 10 minutes prior to loading to

a final concentration of 2 mM, and three technical repeats were run of each sample, with each replicate loaded to a standard glass capillary. The temperature was increased from 20 °C to 95 °C at 1 °C.min⁻¹ in a linear gradient and Trp fluorescence was measured by a NanoTemper Prometheus NT.48. All data processing and graphing was performed in Origin.

3.3.4. X-ray crystallography and structure determination

The NADPH cocrystal for SmoB (pET29a construct) was grown at 13 mg.ml⁻¹ in 0.056 M Sodium phosphate monobasic monohydrate, 1.344 M Potassium phosphate dibasic, pH 8.2. Crystals of SmoB (LiC3C construct) were grown in the ligand-free state at 20 mg.ml⁻¹ in 0.1 M magnesium formate dihydrate. 15% w/v PEG 3350, at room temperature. NADH cocrystals were grown using 20 mg.ml⁻¹ protein, 2 mM NADH, 0.1 M HEPES, 25% PEG 3350 at pH 7.0. NADPH cocrystals were grown using 20 mg.ml⁻¹ protein, 2 mM NADPH, in 0.1 M SPG (Qiagen), 25% W/V PEG 1500. Crystals at pH 4 were harvested for the pure NADPH-bound structure, and crystals at pH 6 were harvested and used in glucose soak experiments. All SmoB crystals were grown at 20 °C. Crystals were harvested into liquid nitrogen, using nylon CryoLoopsTM (Hampton). All soaking took place over 5 seconds, in mother liquor with glucose added dry. Cryoprotection took place over 5 seconds and used 10% glycerol, or 10% ethylene glycol in mother liquor. All crystals were tested using the in-house Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV. X-ray data collection took place at Diamond Light Source on beamlines I-03 and I-04. Data collection statistics are available in Table 3-1. Data indexing and initial processing was performed using either DIALS or 3dii pipelines from the Xia2 package^{120,121}. Data reduction was performed with AIMLESS, and resolution was cut until CC1/2 = 0.5¹²². Molecular replacement was performed using MOLREP, with monomeric SeYdhF (4R90, 43% sequence identity to SmoF) used to solve an initial ligand-free structure¹²³. This was then used to solve subsequent structures. Diagram preparation for molecular models was performed using CCP4MG, Chimera or Pymol, depending on the desired outcome

129–131.

3.3.5. Molecular Cloning

All cloning of SmoB was performed using an In-fusion cloning kit (Takara). Extraction of the plasmid and gene fragments were achieved using a KOD hot-start polymerase PCR kit (NEB), then purified using a gel extraction kit (NEB). Ligation was performed using the standard Takara protocol, and reactions were subsequently transformed into STELLAR cloning cells (Takara) using the standard protocol. Colonies were assayed for the gene using colony PCR: Half of the pellet was picked and suspended in 10 μ l nuclease free water, then heated to 95 °C for 10 minutes to lyse the cells. Debris was pelleted by a 5-minute, 13,000 rpm centrifugation step and the supernatant was used in a standard PCR protocol (NEB). All gene products were viewed on a 2% agarose/TBE gel with 0.01% cybrSAFE and run at 120v for 40 minutes. All primers used in this study, and peptide sequences of the pET29a and LiC3C constructs of SmoB can be found in Tables 3-2 and 3-3.

3.3.6. Isothermal Titration Calorimetry

All SmoB samples for ITC were dialysed using 10 kDa MWCO visking tubing into ITC buffer (25 mM NaPi, pH 7.5, 3 rounds of 400 ml), and diluted to 1.3 mg.ml⁻¹, approximated by taking A₂₈₀ measurements and the accuracy of this measurement was assessed by BCA protein concentration assay (Thermo Fisher). NADPH was bought dry and dissolved in ITC buffer to 1 mM concentration. All samples were degassed and checked for aggregation by centrifugation at 13,000 g for 10 minutes prior to loading in the MicroCal autoITC200 (Malvern Panalytical). This was used to automate 20 injections, the first containing 0.4 μ l of ligand, and a wait 4/16 of 180 seconds, and all subsequent injecting 3 μ l ligand over 6 seconds, with a 150 second wait. A water-water run and a buffer wash was performed before every new experiment. All data analysis and experimental design was performed using Malvern PEAQ-ITC, and all graphing was performed in Origin.

3.3.7. Steady-state kinetics

All NADP⁺ -based kinetics used NADPH absorbance at 340 nm as is commonly used elsewhere¹⁶. All kinetic analysis of SmoB took place in kinetics buffer (25 mM Glycine/Cl) and began with

pH optimisation between pH 8.4 and 9.6 in steps of 0.2. The pH with the highest activity was 9.6, which was used going forward. Optimal values for SmoB concentration and reaction temperature were found through the same method and maintained for all subsequent experiments, at 37 °C and 0.45 mg.ml⁻¹ respectively. For determination of NADP⁺ kinetic parameters glucose concentration was held at 20 mM. Determination of values with respect to glucose an NADP⁺ concentration of 3 mM was used. All data collection took place on a Clariostar Plus microplate reader (BMG Labtech), on a Corning 96-well plate using a 200 µl final reaction volume. Reactions were allowed to reach operating temperature for 5 minutes prior to reaction start. All reactions were started by the addition of 10 µl of SmoB stock. Concentration was confirmed using absorbance at 280 nm, and stocks were prepared by dialysis into kinetics buffer at pH 9.8 at 4 °C. Data were collected 75 times over 375 seconds, with a 5 second, 50 rpm shake at the start in all instances. Initial data processing used MARS software (BMG Labtech) and final kinetic analysis and curve fitting used OriginPro 2019b (OriginLab). All final specific activity measurements are the mean of four repeats.

3.4. Results

3.4.1. Gene expression, protein purification and early characterization

Initially SmoB was expressed using the pET29a expression vector, with successful purification by IMAC achieved using the included C-terminal 6xHis tag and polishing to ~95% purity by SEC. This was clearly visible as a band at around 35 kDa, which corresponds to the sequence-

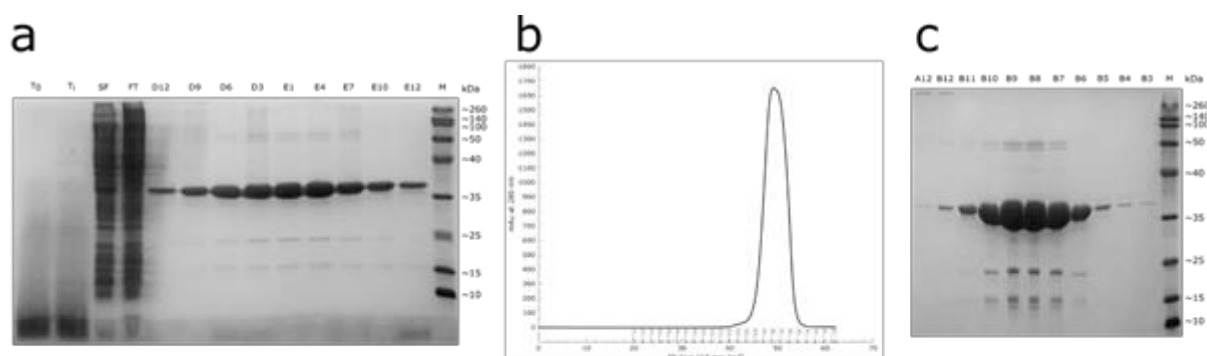


Fig 3-5: Expression and purification of SmoB in the pET29a construct. **A.** SDS-PAGE of SmoB expression. Left to right: Culture at time of inoculation (T₀), induction (T_i), soluble fraction (SF), flow-through from IMAC column (FT), elution fractions (D12-E12) and marker (M). **B.** A₂₈₀ elution profile of SmoB during SEC. **C.** SDS-PAGE of SEC elution fractions.

derived Mw of SmoF. Impurities in the 50, 25 and 15 kDa ranges were also detected, but were not deemed likely to cause issues in crystallisation due to their low relative concentrations. Overall, a yield of 7 mg per 1g of wet cell pellet was achieved (Figure 3-5).

The protein proved readily able to form crystals across a range of commercial screening conditions, both ligand-free and co-crystallised with NADH, NADPH. From these a dataset of SmoB with NADH was collected and solved at 1.83 Å, with one copy in the ASU. The protein forms a 3-way symmetrical arrangement by crystallographic symmetry and exhibits the TIM-barrel fold and α helix pair seen in nearly all AKRs. In particular, the fold is reminiscent of YdhF, a trimeric putative AKR from *S. enterica* (PDB 1OG6, C α RMSD = 1.7 Å). No condition yielded structures in anything except a ligand-free state. These structures contain clearly resolved electron density showing the C-terminal 6xHis tag occupying the cofactor site of an adjacent monomer (Figure 3-6). This appears to be stabilised by a π - π interaction between the side-chain imidazole of His₃₀₀ of one subunit and side-chain indole of Trp₂₁₀ of the other. Much of the nicotinamide-binding region of the cofactor site is also hydrophobic. Together, these may explain the stability of the interaction (as seen through the clear electron density for the tag) and apparent inability to displace it with NAD(P)H by co-crystallisation. The presence of Asn₂₈₅₋₂₈₆ and Arg₂₈₉ close to the likely cofactor binding site suggest a capacity to bind NADPH, based on similar motifs in AKRs with NADPH or dual specificity.

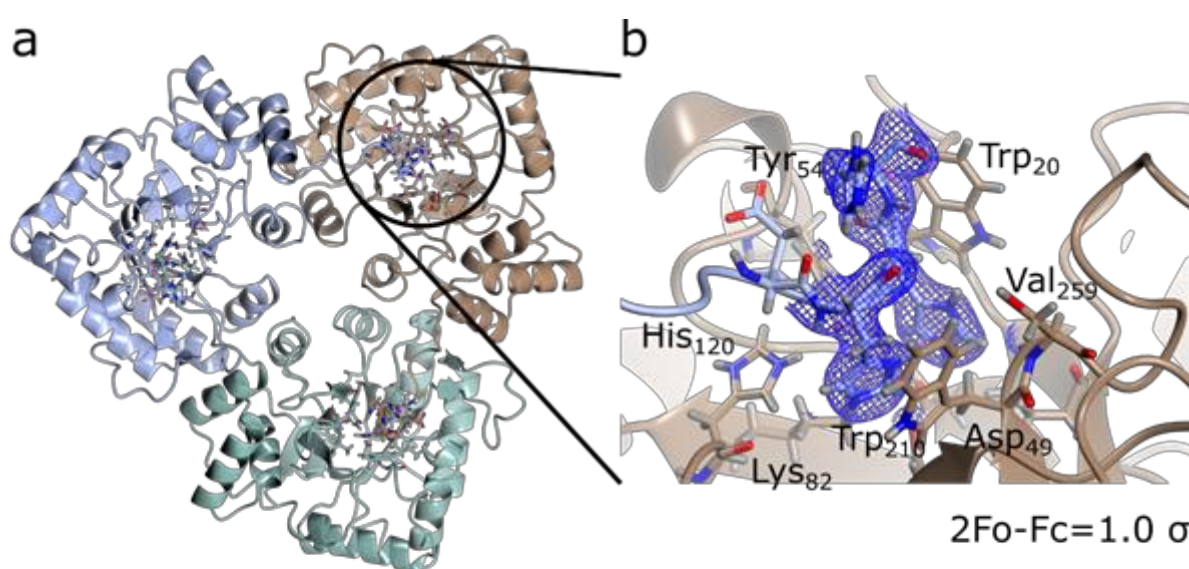


Fig 3-6: Crystal structure of trimeric SmoB in pET29a construct. A. View of SmoB trimer. **B.** View of the cofactor binding site occupied by the 6xHis tag of the adjacent subunit. Density maps (2Fo-Fc) are contoured to 1.0 σ (0.35 e/Å³)

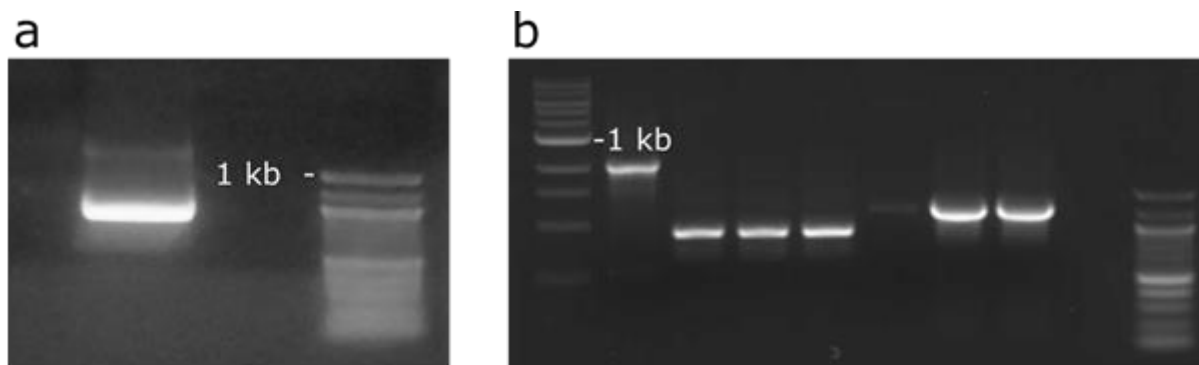


Fig 3-7: Molecular cloning of *SmoB* into pET-YSBLic3C. A. Agarose/TBE gel for inverse PCR amplification of the *SmoB* gene. **B.** Agarose/TBE gel for colony PCR of successful ligation/transformation.

The pET29a construct does not include a cleavage site for the tag, precluding the possibility of obtaining ligand-bound structures through a protease digest and re-purification. These structures did still prove useful however in confirming the general fold. *SmoB* was subsequently subcloned into the pET-YSBLic-3C expression vector to take advantage of the N-terminal, TEV protease cleavable 6xHis tag included in the vector (Figure 3-7). After successful subcloning, *SmoB* was re-expressed. As pET-YSBLic-3C is T7-based, the same expression and purification method was used as with the original pET29a construct. This again resulted in a yield of 7 mg per 1 g wet cell pellet, and a clear band of 35 kDa in all post-induction samples run on SDS-PAGE (Fig. 3-8). The impurities observed before are largely missing, though bands at around 20 kDa suggest at least one protein was co-purified. Again, this was not immediately purified out as the relatively low concentration (the sample was again about 95% pure) suggested it was of little consequence to future experiments.

To determine the solution multimeric state, SEC-MALLS was performed on both constructs of *SmoB*. This identified a M_w of ~100 kDa during elution for both, which corresponds to a solution trimer. Therefore, the presence of the C-terminal 6xHis tag in the adjacent cofactor binding site is unlikely to be a major influence on trimer formation (Figure 3-9).

3.4.2. Binding assays and cofactor specificity

Crystal structures in the pET29a construct contain motifs suggestive of NADPH or dual cofactor specificity, however as yet this ability had not been shown. To this end NanoDSF was used as a high-throughput T_m screen with NADH and NADPH. 2 mM NADPH produced a 2.0 °C increase in T_m over the ligand free protein ($T_m = 51.5$ °C) while 2 mM NADH only produced a 0.9 °C increase. The ΔT_m with NADPH over ligand-free increased to 3.0 °C with the addition of 2 mM glucose suggesting it was a likely product. The same experiment was also performed with SmoB in the pET29a construct, which resulted in no significant T_m shifts with any ligand (Figure 3-10). It is likely that the 6xHis tag could not have been displaced readily enough to allow for more sensitive experiments such as ITC and steady-state kinetics. Some samples containing NADH or NADPH showed a larger standard deviation in T_m than other conditions

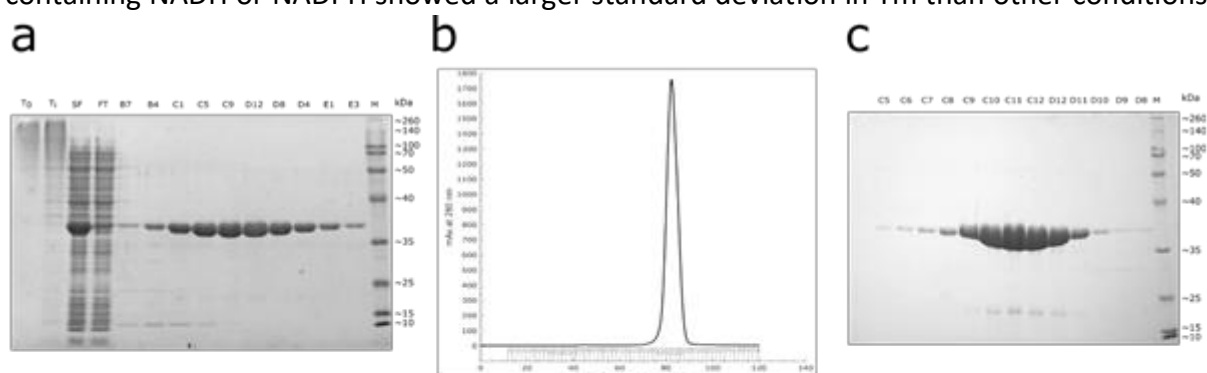


Fig 3-8: Expression and purification of SmoB in the pET-YSBLiC3C construct. A. SDS-PAGE of SmoB expression. Left to right: Culture at time of inoculation (T_0), induction (T_i), soluble fraction (SF), flow-through from IMAC column (FT), elution fractions (B7-E3) and marker (M). **B.** A_{280} elution profile of SmoB during SEC. **C.** SDS-PAGE of SEC elution fractions.

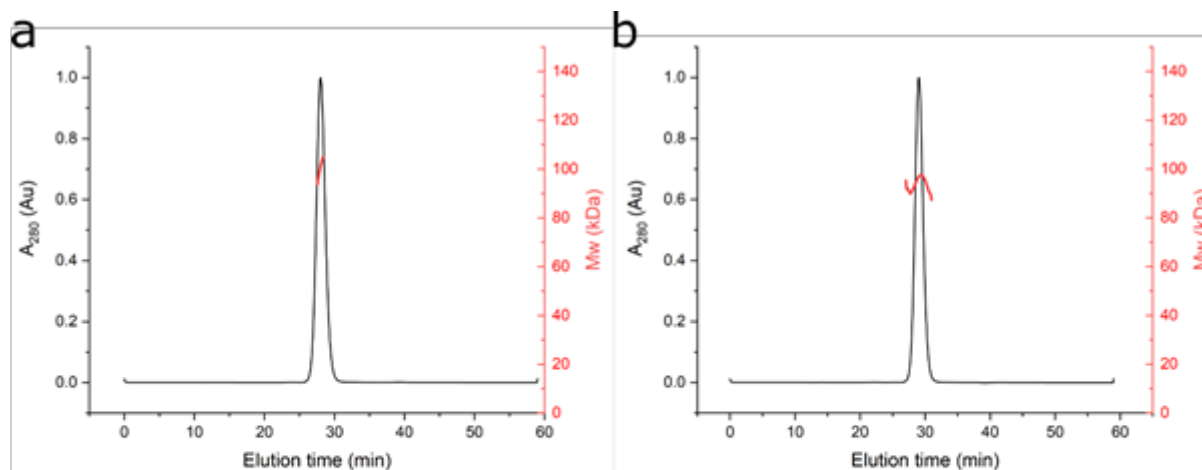


Fig 3-9: SEC-MALLS of SmoB. SEC-MALLS traces in the **A.** pET29a and **B.** pET-YSBlic3C constructs. A_{280} trace for elution is shown in black, estimated Mw based on scattering in red. Scattering trace has been clipped to region at elution only.

despite using the same number of technical repeats and running in the same experimental run as the others. A possible reason may be the absorbance of NAD(P)H at 340 nm, which could interfere with the Trp fluorescence measured at 330 and 350 nm in nanoDSF. However, the reason for the inconsistency in this observation is unknown, and it was reasoned to be of little consequence due to the more detailed calorimetric experiments subsequently performed.

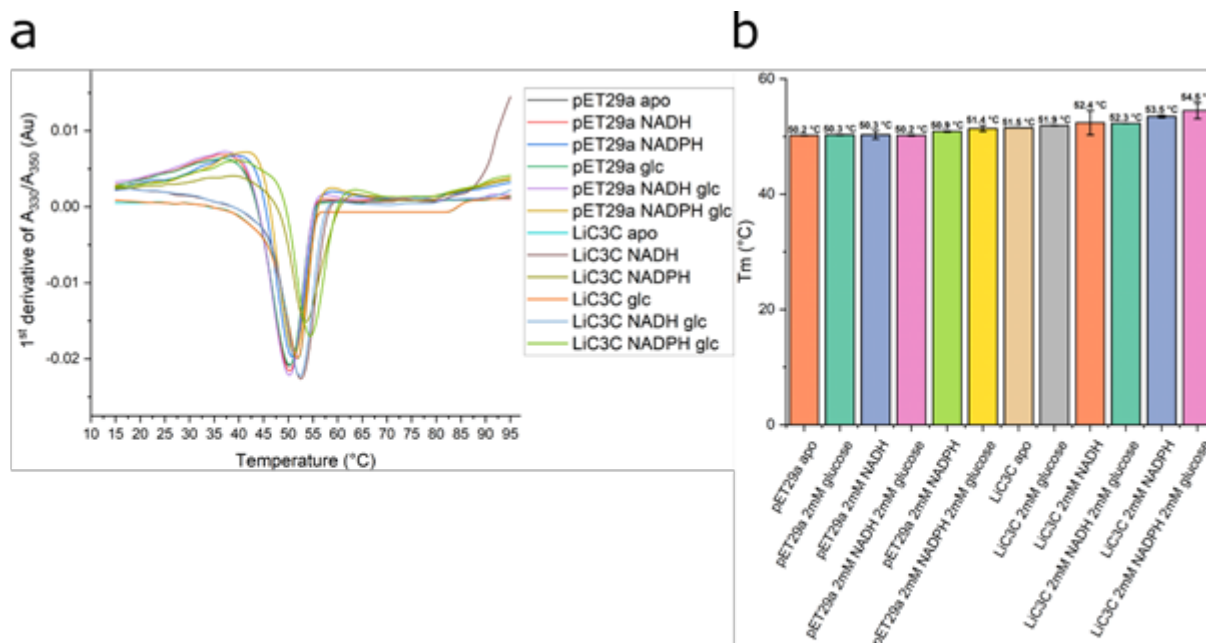


Fig 3-10: Cofactor and glucose binding assays for SmoB. A. Raw plots of 1st derivative of A₃₃₀/A₃₅₀ with increasing temperature. In all cases the mean of all repeats was plotted. **B.** Bar chart of T_m of SmoB in pET29a/pET-YSBLiC3C constructs with 2 mM NADH, 2 mM NADPH, 2 mM glucose. In all cases the inflection point of the raw data was averaged across all repeats, with error bars plotting the standard deviation.

Isothermal titration calorimetry (ITC) was used to conclusively determine the cofactor specificity of SmoB. A titration of NADPH into SmoB shows a clear single site interaction with a K_D of $1.2 \pm 0.04 \mu\text{M}$ (Figure 3-11). This is about 24x lower than the affinity for NADPH observed in rat liver AKR 3 α -HSD at 50 nM¹⁶, suggesting it may not employ the same high binding affinity/ substrate promiscuity strategy employed by 3 α -HSD. Based on the observed concentrations of SmoB, NADPH, the stoichiometry N-value for the interaction was 0.76 ± 0.01 . Based on the crystal structure in the pET29a construct as well as structures of other AKRs a 1:1 binding stoichiometry is very likely, and the variation observed is most likely due to a discrepancy between the calculated concentration of NADPH and the real value. The residual peaks after the trace plateaus are caused by dilution of the sample with increasing

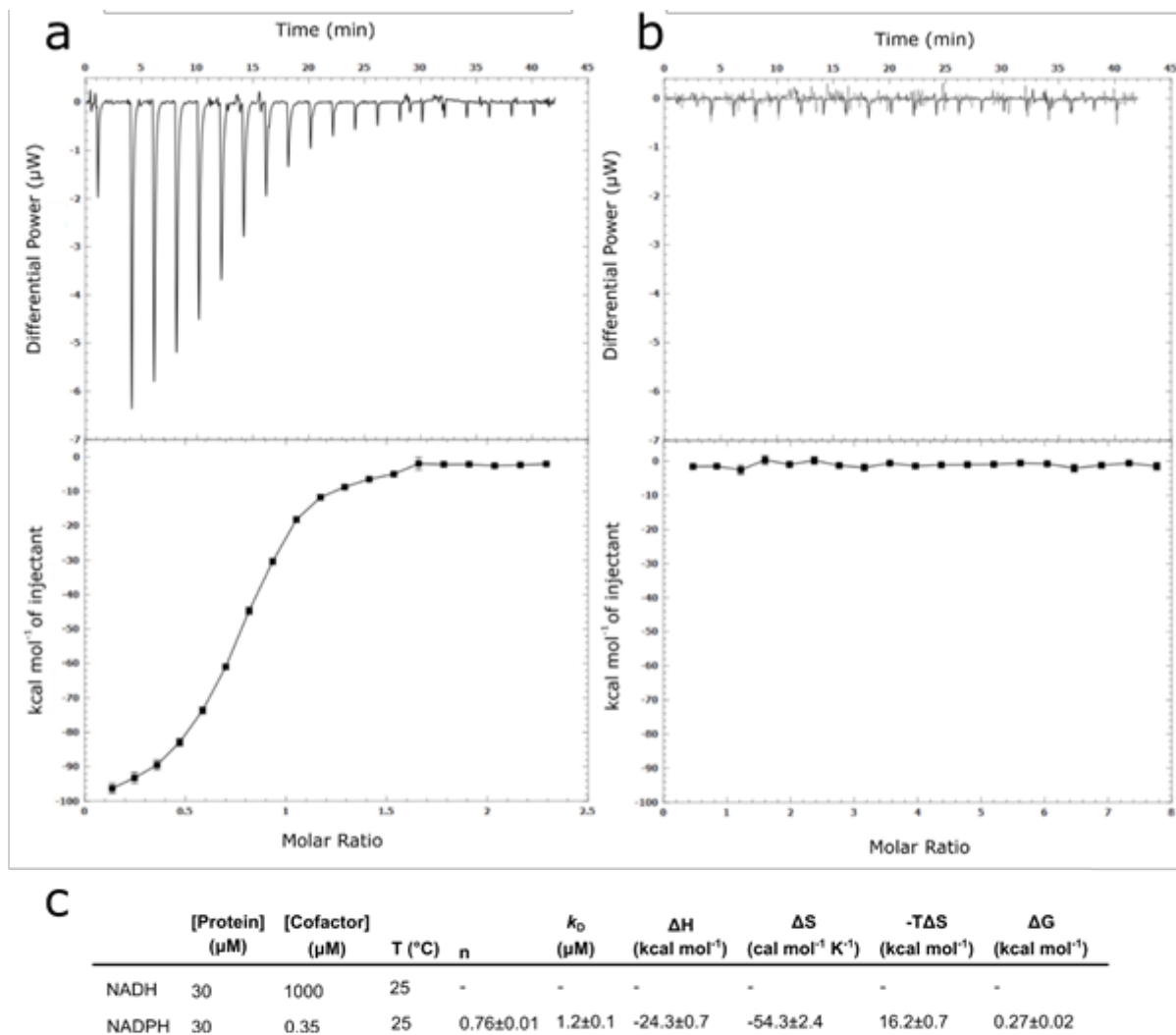


Fig 3-11: Determination of cofactor specificity in SmoB. Thermograms for raw heats (top) and enthalpy (bottom) with time for SmoB with **A.** NADPH **B.** NADH. **C.** Table of experimental and thermodynamic values obtained through experiments **A,B.**

titrant rather than any residual binding events. A titration of NADH into SmoB produced a thermogram that could not be fitted to any binding model, and indeed, no changes in differential power were detected beyond a consistent dilution of titrant into the cell. Based on these data NADH is unable to bind SmoB in solution, making the protein specific to NADPH.

3.4.3. SmoB and cofactor recognition

Like in the pET29a construct, SmoF in LiC3C readily crystallised both ligand-free and co-crystallised with NADH and NADPH in commercial screens. From these conditions optimised

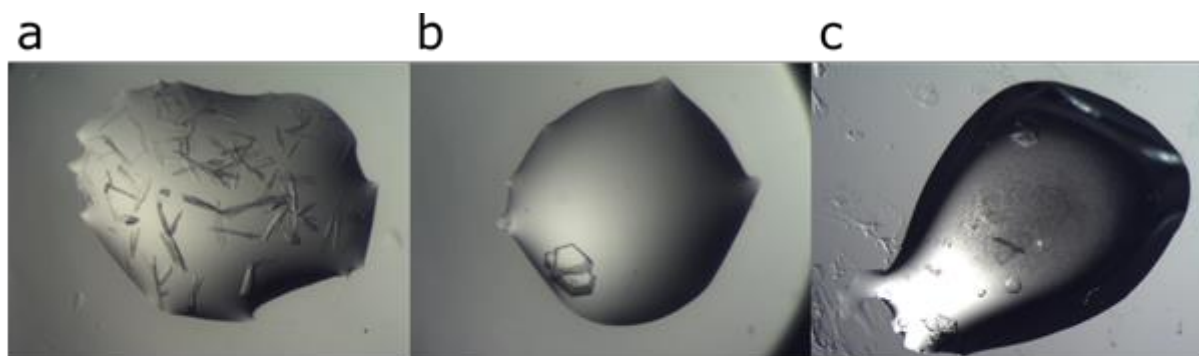


Fig 3-12: Micrographs of SmoB crystals in the pET-YSBLIC3C construct. Light micrographs of **A.** Ligand-free SmoB. **B.** SmoB co-crystallised with NADPH. **C.** SmoB-NADPH crystal grown from a scaled-up formulation of the same mother liquor used in B. Glucose powder was soaked into crystals prior to harvesting from this drop.

crystal growth was achieved resulting in larger crystals and successful harvesting in all three conditions (Figure 3-12). Of these, ligand-free and NADPH yielded unique structures, with NADH resulting in a second ligand-free structure at 1.77 Å. NADPH produced a structure at 1.27 Å resolution. In both cases, SmoB adopts the same trimeric state observed in the pET29a construct and predicted in solution by SEC-MALLS. As with that structure, this state is produced by crystallographic symmetry rather than being present in the asymmetric unit, though the structures contain 1 and 2 copies in the ASU respectively. The same condition used to produce the NADPH crystals was also used to produce a ternary complex by soaking powdered glucose in the drop prior to harvesting. This structure is at 1.5 Å, with one copy in the ASU. As expected, this structure is in the same spacegroup ($P 6_3$) as the NADPH binary structure. The ligand-free structure is in $C 2 2 2_1$, distinct from the pET29a structure, $P 2_1 3$. A full list of collection and refinement statistics are available as Table 3-1. As with other AKRs, the cofactor binding site surrounds the extended NADPH, with residues mostly burying it in the bound conformation (Figure 3-13 a-c). A 2' phosphate binding pocket is present, with all terminal phosphate oxygens hydrogen bound the guanidium group of Arg₂₈₉, the side-chain hydroxyl of Asn₂₈₅ via an ordered water and the peptide bond amine linking Thr₂₈₄ and Asn₂₈₅. Each of these form a hydrogen bond to a phosphate oxygen (Figure 3-13 d,e; Figure 3-14, a). This pocket undergoes no significant conformation change between the cofactor-free and bound states, with the exception of Arg₂₈₉, which rotates almost 90° to form a hydrogen bond to the phosphate and avoid a steric clash with the adenosine moiety (Figure 3-14, c, d). No further significant conformation changes occur with the addition of glucose. Unlike the pockets seen in dual-specific AKRs, this loop does not show any evidence of inherent large-

	SmoB-apo (YSBLIC3C)	SmoB• NADPH	SmoB•NADPH•Glc	SmoB-apo (pET29a)
Data collection				
Space group	C 2 2 2 ₁	P 6 ₃	P 6 ₃	P 2 ₁ 3
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	89.34, 137.6, 153.8	82.93, 82.93, 77.52	82.85, 82.85, 77.26	108.93, 108.93, 108.93
α , β , γ (°)	90,90,90	90,90,120	90,90,120	90,90,90
Resolution (Å)	1.77(1.80-1.77)	1.27(1.29-1.27)	1.50(1.53-1.50)	1.83(1.87-1.83)
R _{merge}	0.120(1.877)	0.108(0.932)	0.079(1.548)	0.189(4.196)
R _{pim}	0.064(0.99)	0.046(0.799)	0.035(0.801)	0.039(0.865)
<i>I</i> / σ <i>I</i>	11.2(1.2)	9.6(0.6)	14.7(1.4)	14.1(1.0)
CC1/2	0.998(0.547)	0.999(0.439)	0.999(0.582)	0.999(0.439)
Completeness (%)	100(100)	98.6(85.5)	100(99.9)	100(100)
Redundancy	8.4(8.3)	10.1(3.2)	11.9(9.0)	24.5(24.9)
Refinement				
Resolution (Å)	1.77	1.27	1.50	1.83
No. unique reflections	92215	78502	48123	38208
R _{work} / R _{free}	0.18/0.21	0.18/0.20	0.17/0.19	0.24/0.27
No. atoms				
Protein	6698	2179	2178	2237
Ligand/ion	0	48	60	10
Water	536	307	230	183
<i>B</i> -factors (Å ²)				
Protein	29	18	22	34
Ligand/ion	-	16	24	54
Water	32	29	32	37
R.m.s. deviations				
Bond lengths (Å)	0.0151	0.0178	0.0120	0.0145
Bond angles (°)	1.90	2.11	1.73	1.810
Ramachandran Plot Residues				
Favourable (%)	97.97	98.96	99.31	97.97
allowed (%)	2.03	1.04	0.69	1.36
Outliers	0	0	0	2
PDB code	7BBZ	7BC0	7BC1	7BBY

Table 3- 1 Data collection and refinement statistics for all SmoB crystal structures. All values are within *I*/ σ *I* where applicable. Values in parentheses represent the outer resolution shell.

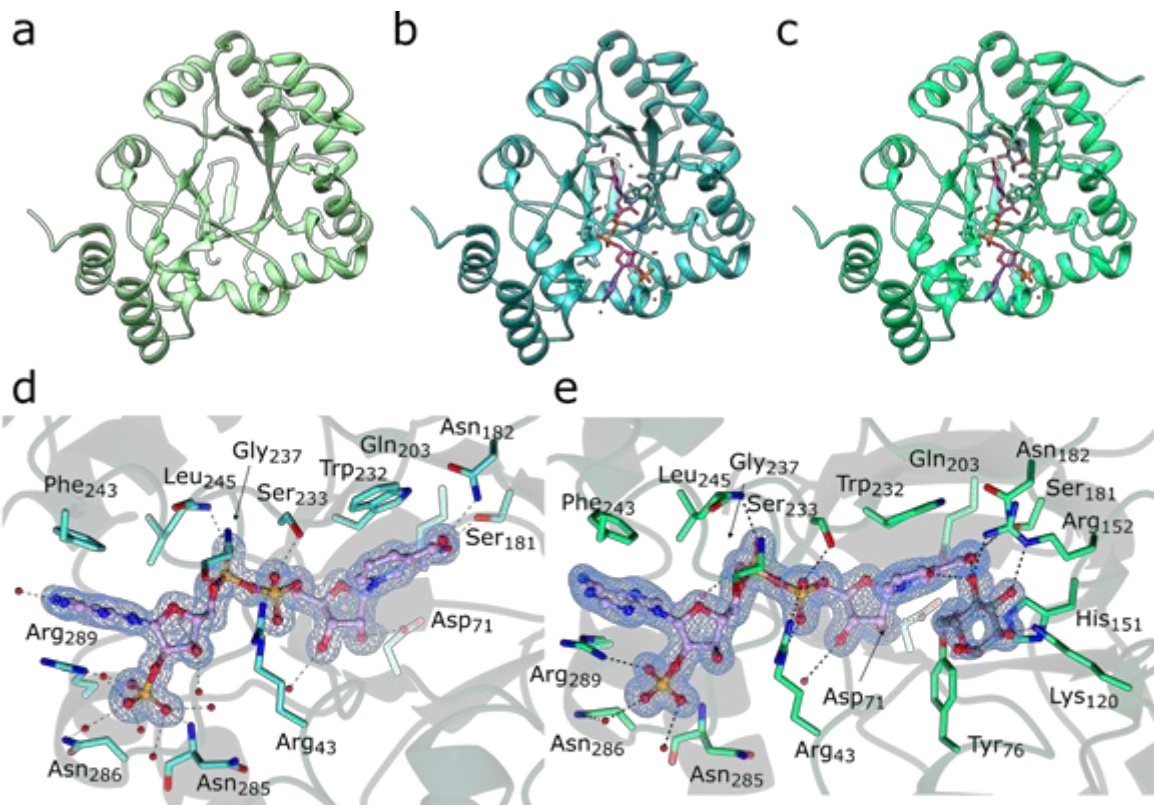


Fig 3-13: Structural overview of SmoB – ligand complexes. Crystal structures of SmoB **A.** ligand-free, **B.** NADPH, **C.** NADPH, glucose. **D.** Binding site detail of NADPH (purple). **E.** Binding site detail of NADPH (purple), glucose (grey). All density is represented by 2Fo-Fc maps, contoured to 1.0 σ (0.44 e/Å³ and 0.26 e/Å³ for the binary and ternary complexes respectively).

scale flexibility, providing a structural basis for the inability to bind NADH observed. The pyrophosphate linker of NADPH is bound by hydrogen bonds to the terminal oxygens much as the 2'-phosphate is. The adenosine-adjacent phosphate forms hydrogen bonds to the backbone amide of Leu₂₄₅ and Gly₂₃₇, and the nicotinamide-adjacent phosphate from the guanidium group of Arg₄₃ and the side-chain hydroxyl of Ser₂₃₃. π - π stacking interactions appear to be largely responsible for the extended NADPH conformation. Trp₂₃₂ and Phe₂₄₁ lay 3.3 Å and 4.9 Å from the adenine and nicotinamide moieties of NADPH respectively. In both cases, these interactions appear to hold their respective moieties in the *anti*-conformation, relative to the ribose sugars (Figure 3-13, d,e).

3.4.4. Structural basis of trimer formation

SmoB forms a trimer both as part of all known crystal forms and in solution. Probing a structure of the biological assembly for all residues within 2 Å of an adjacent residue reveals a clear interface between each monomer. This interface is comprised of a hydrophobic core surrounded by polar and neutral residues, with Leu₁₅₆ of one forming the core of a

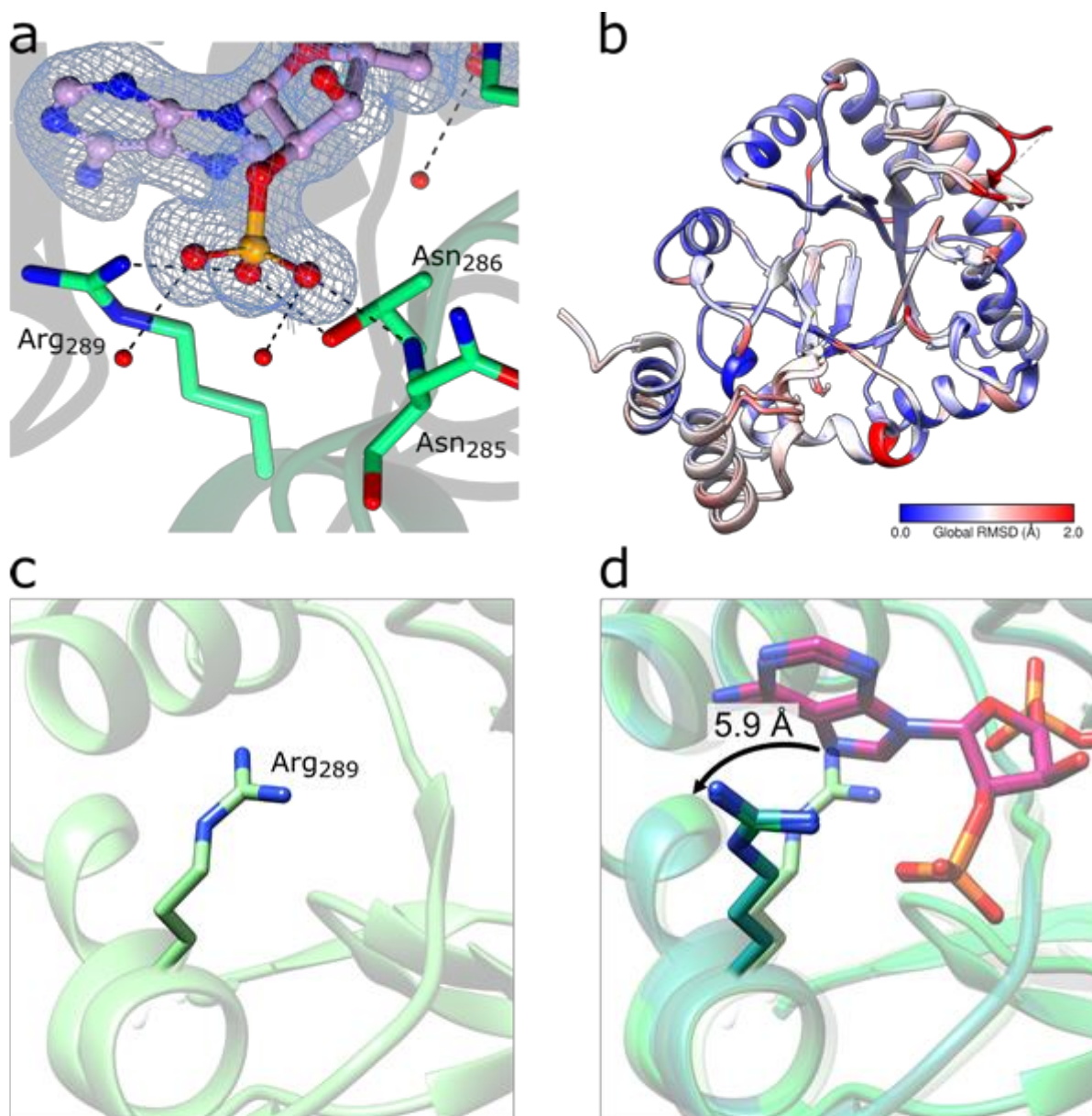


Fig 3-14: Structural basis of NADPH selectivity. **A.** 2' Phosphate binding pocket of SmoB (green) recognising the 2' phosphate of NADPH (purple). Density is represented by a 2Fo-Fc map contoured to 1.0 σ (0.26 e/Å³). **B.** Superposition of ligand-free SmoB, binary and ternary complexes, coloured by global RMSD relative to the ligand-free structure. Blue represents low RMSD, red 2 Å or higher. **C-D.** Movement of Arg₂₈₉ between the ligand-free (C) and binary, ternary structures (D) by 5.9 Å.

hydrophobic protrusion, accepted by a pocket in the adjacent monomer (Figure 3-15 c). The protrusion is located on the C-terminus of the protein, along the surface edges of helices α_9

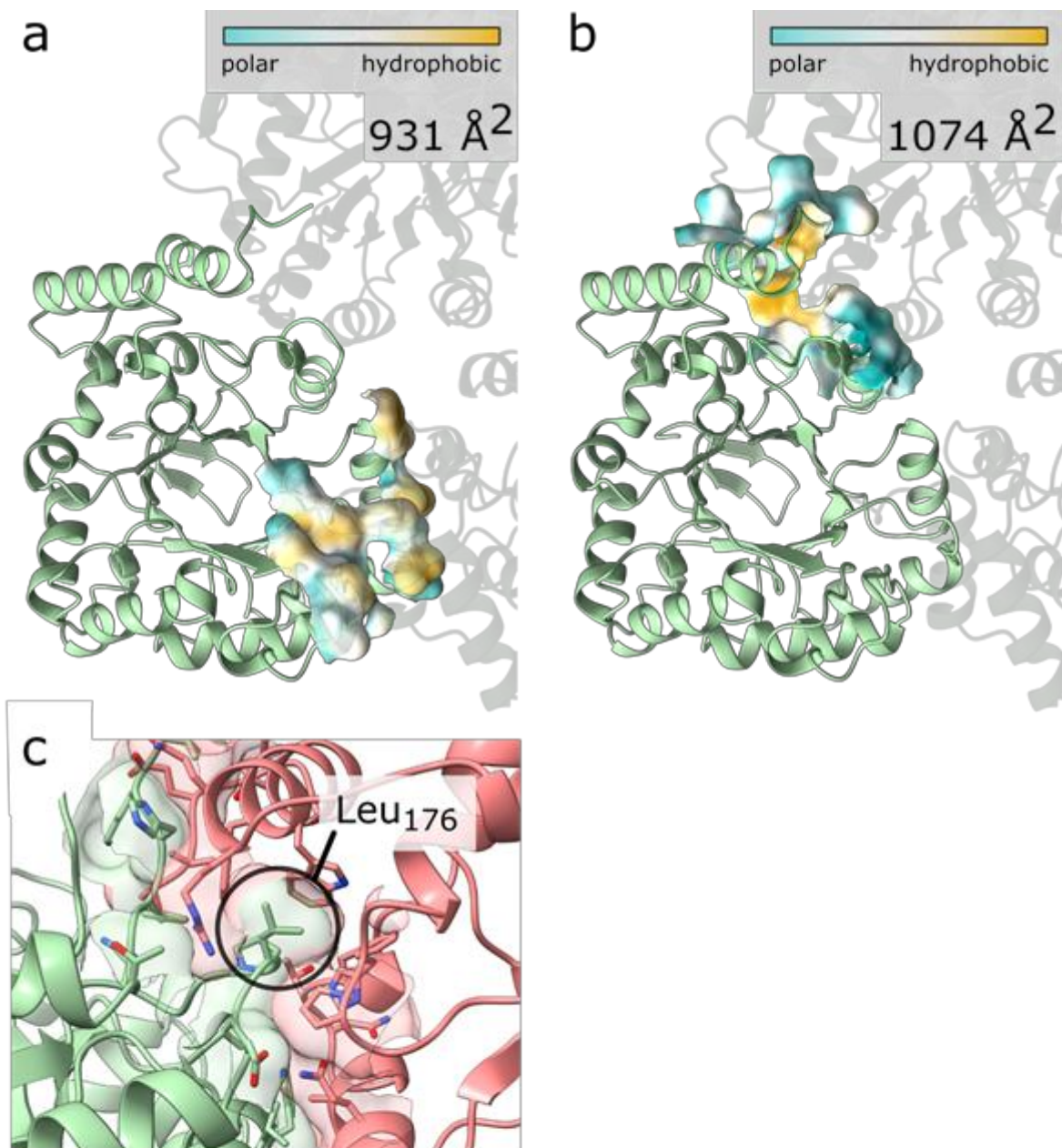


Fig 3-15: Structural basis of SmoB trimer formation. SmoB trimer (green) with surfaces mapped for **A.** the protruding interface **B.** the pocket interface. In both cases the surfaces are coloured by hydrophobicity, blue for polar and red for non-polar. A single SmoB monomer is highlighted to more clearly show the interface. **C.** Detail of protrusion/pocket interface. SmoB monomer highlighting the protrusion is in green, monomer highlighting the pocket in red. Leu₁₇₆ is highlighted as the basis of the protrusion.

and α_{12} . The acceptor is on the $\alpha_8\beta_8$ barrel, primarily formed from surface residues along helices α_4 , α_5 , and the C-terminus of α_3 . Each monomer has one protrusion and one pocket

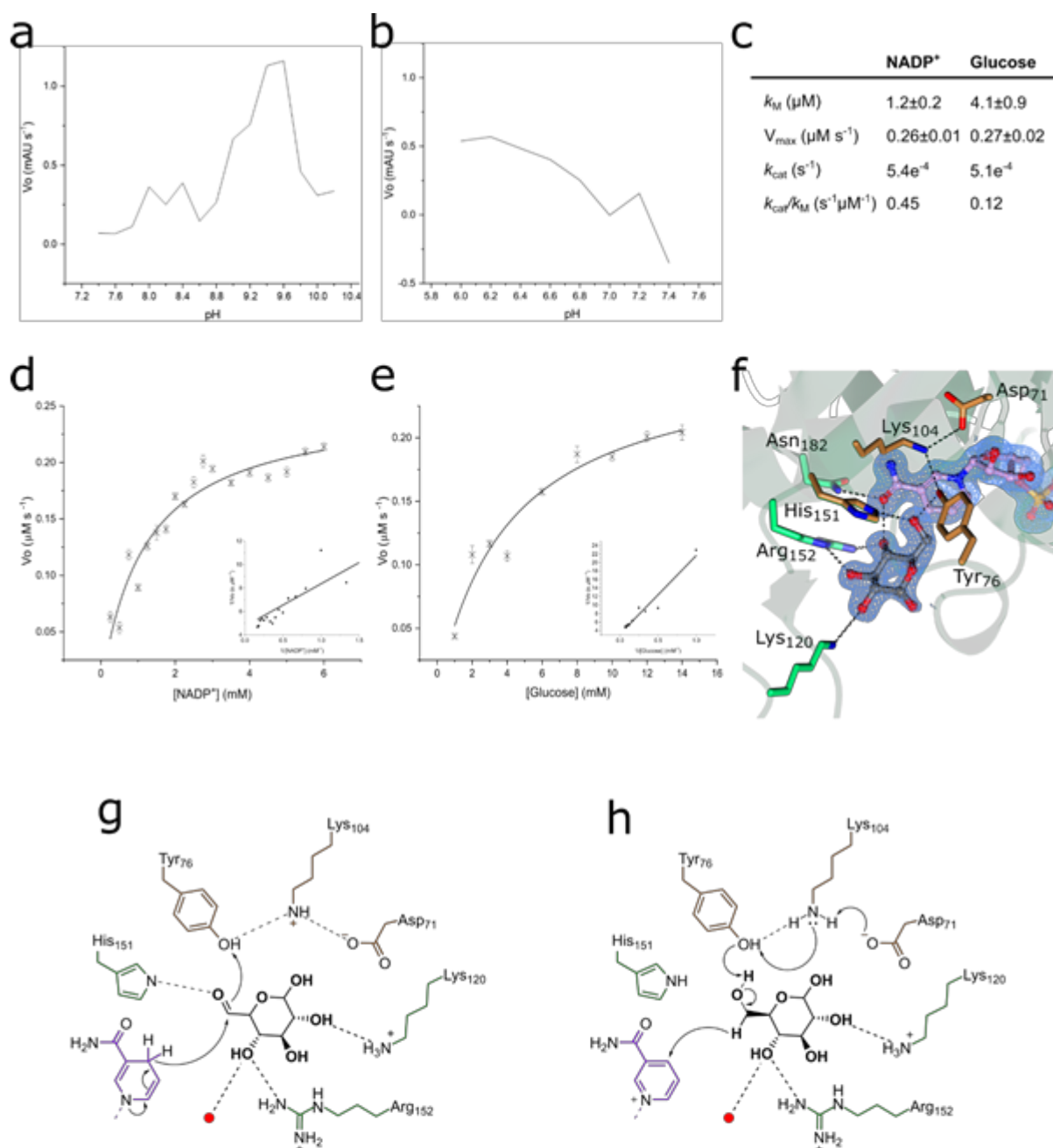


Fig 3-16: Structural and kinetic basis for SmoB activity. pH-activity profiles with **A.** NADP⁺/glucose **B.** NADPH/GA. These were used to determine an optimal pH for NADP⁺/glucose, and the basis for concluding SmoB is not active against GA. **C.** Kinetic parameters for SmoB with respect to NADP⁺, glucose. **D-E.** Michaelis-Menten plots for SmoB with respect to NADP⁺, glucose respectively. Inset double-reciprocal plots are also included (inset). **F.** Structural view of sugar binding site, occupied by glucose. The catalytic tetrad are depicted in brown, and other residues interacting with glucose in green. All electron density used 2Fo-Fc maps, contoured to 1.0 σ . NADPH is in purple, glucose in grey. **G-H.** Proposed mechanism for SmoB in the likely *in vivo* reductive capacity (G) and observed oxidative capacity (H). All residue colours are represented identically to F.

with an approximate included angle of 120°. This allows three monomers to form the observed trimer. The overall surface area of the protrusion is 931 Å², and the acceptor is slightly larger at 1074 Å² (Figure 3-15 a,b). The total area per monomer is on average 12640 Å². Therefore, approximately 16% of the surface area of each monomer forms an interface: 7.4% protrusion and 8.6% acceptor. The solvent exposed area of the complex is 31700 Å². As no monomeric populations of SmoB have been observed under any tested conditions it can be surmised that the affinity of this interaction is more than sufficient to maintain the solution trimer. The inclusion of the C-terminus in the protrusion interface provides further evidence that the pET29a construct, with the non-cleavable C-terminal 6xHis tag was unlikely to ever produce a biologically relevant complex as a tag of any kind would always be in a prime location to interact with the adjacent monomer.

3.4.5. Activity of SmoB and sugar binding

Based on its location in the sulfo-SMO pathway, the likely substrate of SmoB is 6-OG. However, this species is not readily available. Therefore, reverse kinetics were used to determine whether SmoB acts as an oxidase or reductase *in vivo* and produces glucuronic acid or glucose as a product respectively. Early pH profiles determined a peak in activity with glucose and NADP⁺ at pH 9.6 but no clearly active pH with GA and NADPH (Figure 3-17 a,b). Therefore, Michalis-Menten kinetics were optimised for the glucose / NADP⁺ reaction only, resulting in the collection of separate kinetic parameters with respect to [NADP⁺] and [glucose] (Figure 3-17 c,d,e). These provide a K_M of 1.2 and 4.1 mM respectively. The V_{max} was almost identical with respect to both glucose (0.26 μM s⁻¹) and NADP⁺ (0.27 μM s⁻¹). This is entirely expected as both cofactor and substrate have a binding stoichiometry of 1, and therefore should process at an equimolar maximal rate. This was considered a good indicator of the validity of the other parameters obtained. A first-order rate constant, k_{cat} of 5.4x10⁻⁴ and 5.1x10⁻⁴ s⁻¹ for NADP⁺ and glucose respectively were also obtained, as well as a second order rate constant, k_{cat}/K_M of 0.45 s⁻¹ μM⁻¹ and 0.12 s⁻¹ μM⁻¹ with regards to NADP⁺ and glucose respectively. Similarly to V_{max} the similarity in first-order rate constant is expected due to the identical stoichiometry of glucose and NADP⁺, so SmoB would be expected to

perform at the maximum rate for both when the cofactor and substrate sites are both saturated.

Based on the lack of activity over any pH, SmoB cannot utilise GA as a substrate and is therefore unlikely to produce it as a product. This confirms SmoB as a 6-OG reductase and glucose as the final product of the sulfo-SMO pathway. The basis and likely mechanism for the observed activity are clearly visible in a crystal structure of a ternary complex of SmoB with NADPH and glucose (Figure 3-17 f). This was obtained by producing crystals identical to those used for the NADPH binary structure, then soaking solid glucose for 5 minutes prior to harvesting. Here, glucose is clearly visible in the active site next to the nicotinamide moiety of NADPH and forms a hydrogen bond (2.6 Å) between the hydroxyl of that moiety and its own C5 hydroxyl. The binding site is completed by hydrogen bonds from Lys₁₂₀ to the C2 hydroxyl and Arg₁₅₂ to the C3-4 hydroxyls. The active triad seen in other AKRs is present as Tyr₇₆-Lys₁₀₄-Asp₇₁, as well as His₁₅₁, which forms a hydrogen bond to the C6 hydroxyl and Asn₁₈₂ which forms a hydrogen bond to the nicotinamide hydroxyl of NADPH. Based on this as well as the mechanisms of comparable AKRs it is likely that Tyr₇₆ acts as a general base. *In*

<i>SmoB</i>	TTGGACATATGCAACGTATCGCTCTTTCTGAC	sense
	AGGATCTCGAGCGCCACCTCTTTTCCAATCG	antisense
pET-YSBLiC3C	ATGCAACGTATCGCTCTTTCTG	sense
	TTACGCCACCTCTTTTCCAATC	antisense

Table 3-3: Oligonucleotide sequences used for amplification of *SmoB* gene and pET-YSBLiC3C plasmid.

SmoB (pET29a)	MQRIALSDKLELSRIVYGMWRIGDDADTSPAHVQAKIEACLAQGITTMDQADIYGGY TAEAILGGGLKAAPGLRDKIEIVTKCGIVAPAGRHSARVKHYDTTAGHINVSVEAS LRDMGTDHVDLLLIHRPDPLIDAEETGKALDALVASGKVKAVGVSNFRPWDFSLQ AMSRLVTNQIEMSLLATDTFTNGDLAYLQEKRVSPMAWSPLGGGSLFSGAYGGTMA ALQRIGKEQGVDATAVAIAWLLRHPAKIVPVLGTNNLERIRTAADALRVMTDRQTF ELYTLAIGKEVALEHHHHHH
SmoB (pET-YSBLiC3C)	MGSSHHHHHSSGLEVLFGQPAMQRIALSDKLELSRIVYGMWRIGDDADTSPAHVQA KIEACLAQGITTMDQADIYGGYTAEAILGGGLKAAPGLRDKIEIVTKCGIVAPAGR HSARVKHYDTTAGHINVSVEASLRDMGTDHVDLLLIHRPDPLIDAEETGKALDALVA SGKVKAVGVSNFRPWDFSLQSAMSLVTNQIEMSLLATDTFTNGDLAYLQEKRV SPMAWSPLGGGSLFSGAYGGTMAALQRIGKEQGVDATAVAIAWLLRHPAKIVPVLGT NNLERIRTAADALRVMTDRQTFELYTLAIGKEVA

Table 3-3: Amino acid sequences of *SmoB* gene in pET29a and pET-YSBLiC3C

in vivo, the reductive mechanism of SmoB is likely to proceed starting with a nucleophilic attack from the C=O double bond of 6-OG and to the Tyr₇₆ -OH hydrogen, then hydride ion transfer by NADPH, forming NADP⁺ and glucose (Figure 3-17g). The oxidative mechanism observed

through Michaelis-Menten kinetics is likely to begin by nucleophilic attack from Asp₇₁ to the side-chain amine of Lys₁₀₄ in turn, a lone pair on Lys₁₀₄ attacks Tyr₇₆ which attacks the 6C -OH of glucose causing formation of a C=O double bond. Finally, hydride transfer from the glucose to NADP⁺ forming NADPH and 6-OG (Figure 3-17, h).

3.5. Conclusion and future directions

The investigations described above have clearly shown SmoB is a trimeric AKR with the characteristic $\alpha_8\beta_8$ core and C-terminal α domain the superfamily is known for. The protein forms a solution homotrimer, which is also observable through symmetry in all obtained crystal structures. This arises from a hydrophobic protrusion and acceptor pair on each monomer with an included 120° angle, leading to the 3-way symmetrical arrangement observed. SmoB binds NADPH as the exclusive cofactor, as is common for the class. The affinity of this interaction is high at 1.2 μ M, though not as high as in AKRs known to have high substrate promiscuity such as α 3-HSD. The cofactor is arranged similar to most NADPH-binding AKRs; in an extended, *anti*-conformation along the top of the barrel, and NADPH is selected for by the inclusion of a 5' phosphate binding pocket. This contains an Arg₂₈₉, which undergoes a 5.9 Å rotation away from the protein to bind the phosphate upon NADPH binding. Though similar flexibility is a common mechanism for NAD(P)H dual-specificity, this capacity for movement is seemingly insufficient to grant it in this case. Soaking glucose into SmoF•NADPH crystals produced a ternary complex structure. Here, glucose is located next to the nicotinamide moiety of NADPH and is closely associated with the active tetrad. SmoB is active against glucose, using an NADP⁺ cofactor to act as a glucose oxidase. A lack of observed activity against GA with NADPH marks glucose as the end product of the *in vivo* reductive mechanism. While SmoB has been extensively characterized structurally, the nature of the underlying catalytic mechanism could be further explored through mutagenesis of the active tetrad, particularly Tyr₇₆. Additionally, engineering of the Asn₂₈₅ – Asn₂₈₆ – Arg₂₈₉ loop to improve flexibility could be investigated to determine if SmoB can have NAD(P)H dual-specificity conferred, in the same way as that performed on 2,5-diketo-*D*-gluconic acid reductase¹⁶.

Unlike YihQ and SmoF, SmoB does not process any sulfonated species, and therefore lacks a distinct sulfonate binding pocket. Therefore, use of SmoB as a template enzyme for a bioinformatic approach to sulfoglycolytic pathway identification would not be effective. This stands the enzyme out as a rare case in all known sulfoglycolysis pathways as being a core protein induced by environmental [SQ] but in no way able to recognise the sulfonate group.

4. Chapter 4: SF transaldolase SqvA

4.1. Abstract

Transaldolases are a ubiquitous family of enzymes used for the transfer of dihydroxyacetone between a donor and acceptor substrate in a reversible manner. This is performed using a near-identical mechanism to that employed by class I aldolases; the formation of a Schiff base using a conserved lysine. TALs display a variety of multimeric states and a subfamily-dependent level of sequence conservation but always the same $\alpha_8\beta_8$ barrel fold. Recently TALs in *Bacillus megaterium* and *B. aryahabbatai* were discovered in sulfoglycolytic operons, forming the basis of a novel sulfo-TAL pathway. After isomerization of sulfoquinovose to sulfofructose they transfer dihydroxyacetone from sulfofructose to glyceraldehyde-3-phosphate, producing sulfolactaldehyde and fructose-6-phosphate. Sulfolactaldehyde is subsequently reduced and excreted. Fructose-6-phosphate is used in downstream glycolysis, in turn regenerating glyceraldehyde-3-phosphate. While this activity has been confirmed in these sulfo-TALs, no detailed structural or biophysical characterization has been performed. Here, the structure of the decameric sulfo-TAL SqvA is solved by single particle cryo-EM and X-ray crystallography in ligand-free (2.6 Å) and SF-bound (2.0 Å) states. SqvA activity against SF is structurally confirmed by the formation of a Schiff base. The active triad is conserved between SqvA and other TALs, however the motif binding the sulfonate group shows some differences, notably the addition of an arginine, which may undergo conformation changes to accommodate either the sulfonate of sulfofructose or phosphate of fructose-6-phosphate. These data conclusively present SqvA as a classical TAL with a previously unseen sulfonate binding pocket and provide a strong foundation for in-depth kinetic and mechanistic analysis.

4.2. Introduction

4.2.1. Aldolases, Transaldolases and Transketolases

Aldolases are enzymes that can be used to combine two carbonyls to form a single, longer β -hydroxyl carbonyl product in an aldol condensation reaction¹⁶. They are found ubiquitously in glycolysis and the TCA cycle. However, in these capacities they are primarily used to perform the reverse reaction: The aldol cleavage of a β -hydroxyl carbonyl into two smaller products (fig 4-1, a)¹⁶⁷. The mechanisms employed to do this can be grouped into two classes. Class I aldolases use a catalytic Lys to form an imine with the general structure of $R_1R_2C=NR_3$, where R_1 is a hydrogen and R_2 the rest of the Lys side chain. R_3 is the substrate. This arrangement is also known as a Schiff base. While this is a common natural occurrence of these bases, the only proviso on whether an imine with $R_1R_2C=R_3$ is a Schiff base is that R_3 cannot be a hydrogen¹⁶⁸. Class II aldolases do not form Schiff bases. They are metallozymes which natively coordinate Zn^{2+} . Chelation of the zinc results in inactivation of the enzyme, though substitution with Co^{2+} , Fe^{2+} , Ni^{2+} or Mn^{2+} retain enzyme activity¹⁶⁹. The substrate carbonyl acts as a bidentate ligand through the enol and hydroxamate oxygens, which in addition to a histidine triad fully coordinates the Zn^{2+} ¹⁷⁰.

Transaldolases (TALs) differ from aldolases with their use of a co-substrate, which they use to enact the reversible transfer of a dihydroxyacetone moiety from a 6 or 7 carbon donor (6/7C) species to a 3/4C acceptor, producing a 3/4C product and a new 6/7C product (fig 4-1, b). As with aldolases, this is done without the use of ATP or cofactors. They are further differentiated from transketolases (TKs) which enact the transfer of a 2C species rather than the 3C dihydroxyacetone.

TALs and TKs see almost universal use throughout nature in the non-oxidative pentose-5-phosphshate (P-5-P) pathway (Fig 4-1). They can be found in all three domains of life, and have even been discovered in viruses¹⁷¹. In the P-5-P pathway, TALs and TKs are used to enact dihydroxyacetone between ketose phosphates including fructose-6-phosphate (F6P) and sedoheptulose-7-phosphate (S7P) and aldose phosphates including glyceraldehyde-3-

phosphate (G3P) and erythrose-4-phosphate (E4P)^{172,173}. TAL mechanisms are reversible allowing for efficient interconversion between these sugars as needed. The P-5-P pathway

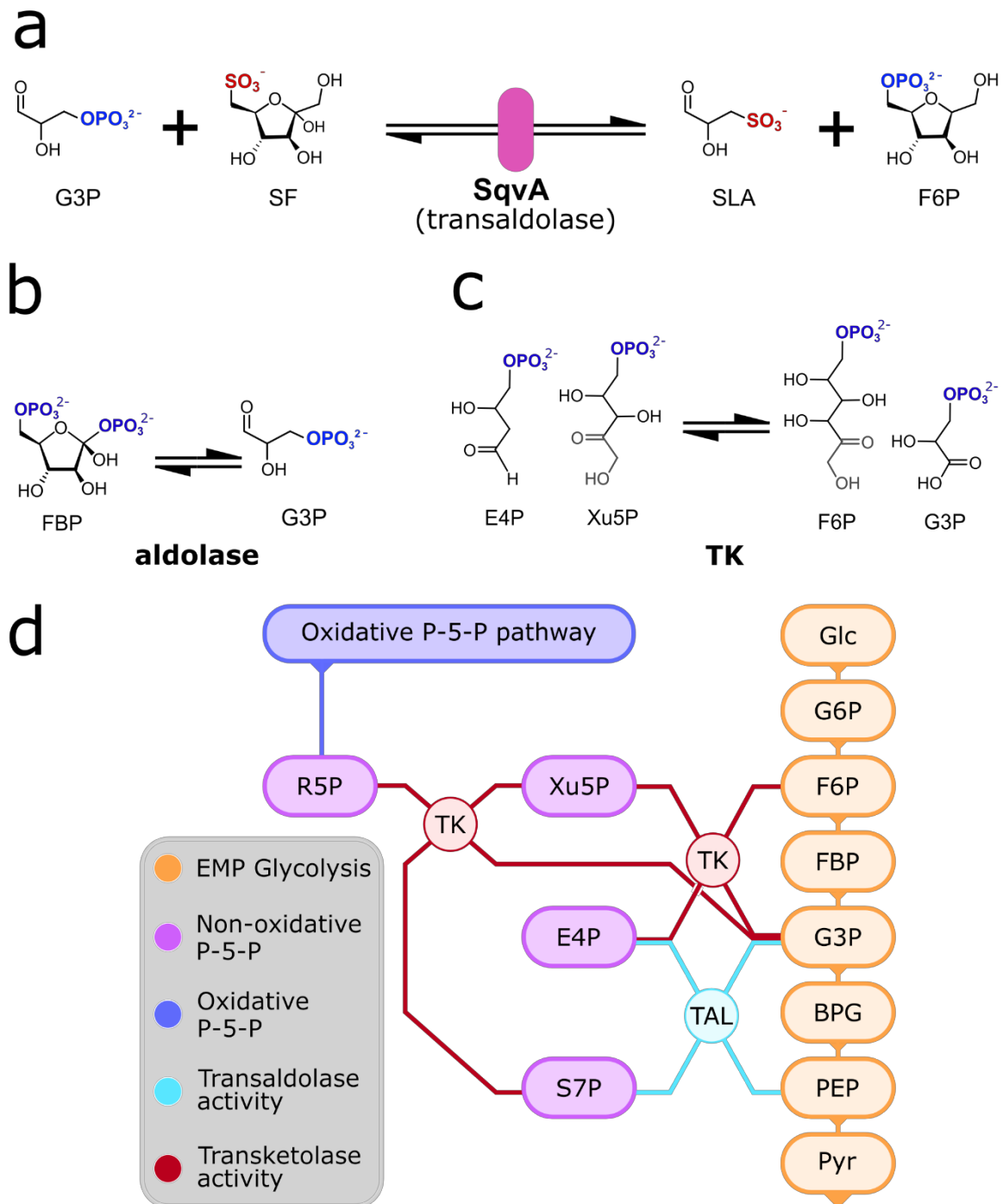


Figure 4-1: Aldolases, TALs and TKs, and the non-oxidative pentose-5-phosphate pathway. Reaction overviews of a TAL (**a**), aldolase (**b**) and TK (**c**). **D.** The Embden-Meyerhoff-Parnas glycolytic pathway is in orange, non-oxidative P-5-P in purple, oxidative in blue. The non-oxidative P-5-P interfaces with EMP glycolysis through fructose-6-phosphate (F6P), glyceraldehyde-3-phosphate (G3P) and phosphoenolpyruvate (PEP).

also uses a similar enzyme, the transketolase (TK). Here, TKs are used to interconvert between F6P and xylose-5-phosphate (Xu5P), and G3P and E4P. An additional TK interconversion exists between Xu5P and S7P, and ribose-5-phosphate (R5P) and G3P. TKs work in a very similar fashion to TALs, allowing the transfer of a 2C species between the donor and acceptor substrates in contrast to the 3C dihydroxyacetone used by TALs (fig 4-1, b,c). The P-5-P pathway is subdivided into the oxidative and non-oxidative P-5-P. Of these, the non-oxidative is by far the more prevalent to the point of being practically universal. The non-oxidative P-5-P pathway is one of the oldest known, with evidence suggesting the processes involved may predate enzyme-catalysed metabolism itself ¹⁷⁴. It overlaps with both Embden-Meyerhoff-Parnas (EMP) glycolysis and the Calvin cycle and produces the precursors to polynucleotide backbones and aromatic amino acids and vitamin B6, R5P and E4P respectively ¹⁷⁵. The oxidative P-5-P pathway is much less common, though still common in most eukaryotes ¹⁷². The pathway produces Ribulose-5-phosphate (Ru5P) through a pair of oxidation steps using NADPH as a cofactor: G6P is oxidised using NADP⁺ by G6P dehydrogenase to form 6-phosphogluconolactone which is subsequently converted to 6-phosphogluconate by 6-phosphogluconolactone lactonase. Finally, 6-phosphogluconate dehydrogenase oxidises 6-phosphogluconate to produce Ru5P. The discrepancy in prevalence between oxidative and non-oxidative P-5-P is evidence that the non-oxidative pathway evolved first ¹⁷⁴. Given that TALs and TKs in the P-5-P pathway always interconvert between P-5-P and ED glycolytic metabolites, the two enzyme classes are considered important to maintaining the balance of the two systems. S7P is also a product of other processes including synthesis by S7P isomerase ¹⁷⁶.

TALs are generally subdivided into 5 subfamilies, based on sequence similarity ¹⁷. Subfamily 1 contains TALs used in the P-5-P pathway across all of life and labelled “classical TALs”. These TALs generally feature a well-conserved sequence and similar activities and mechanisms ¹⁷⁷. Subfamily 2 still bears close sequence similarity to subfamily 1, however in the few cases observed, the active site is not shared, making the subfamily inactive as TALs. Enzymes of this subfamily are typically isoenzymes in plants. The activity of these TALs is currently unknown ¹⁷⁸. Subfamily 3 are generally larger than TALs of other subfamilies (~380 aa vs 310-350 aa for subfamily 1). They are found in plants and bacteria, and exhibit activity similar to classical

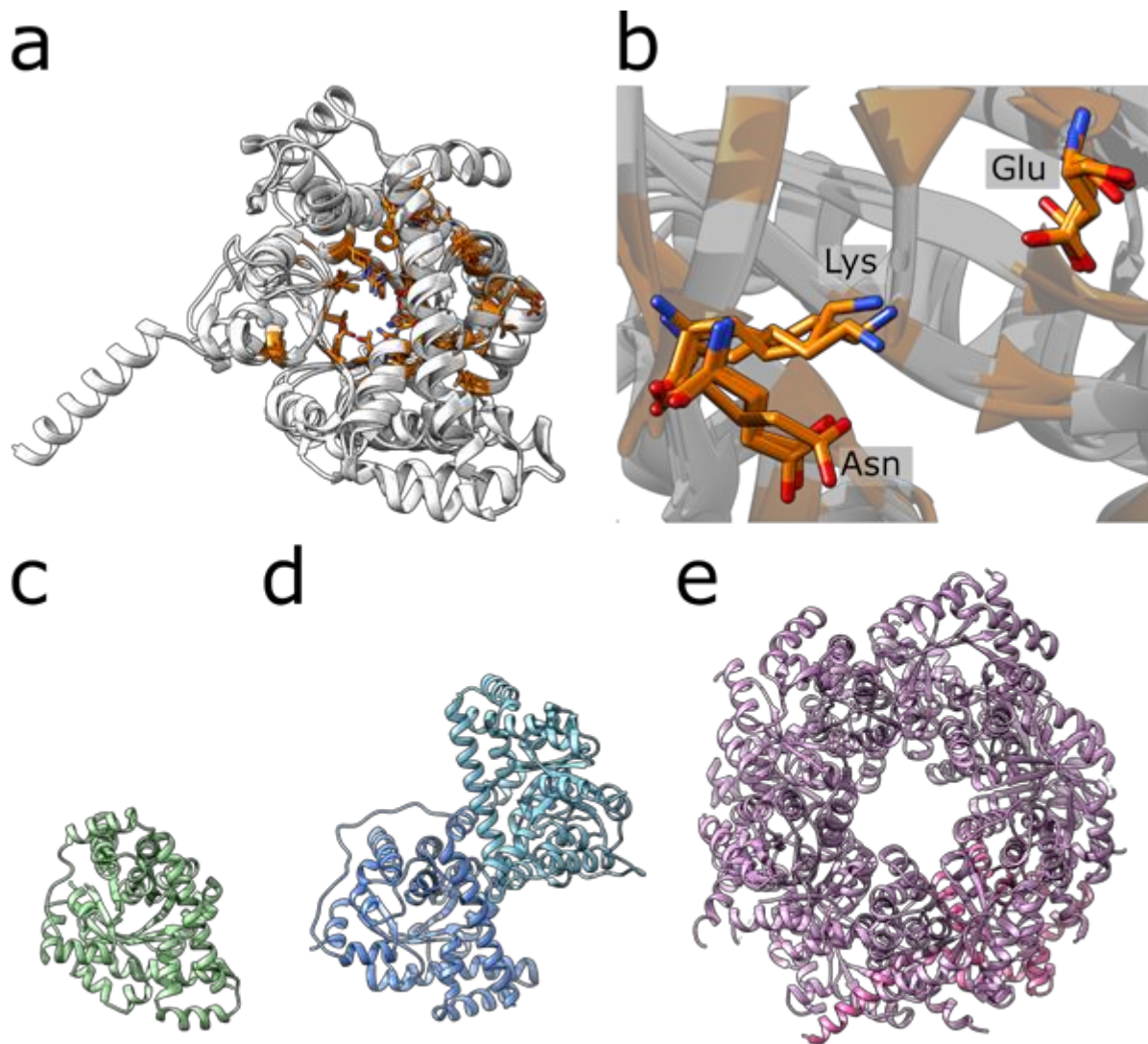


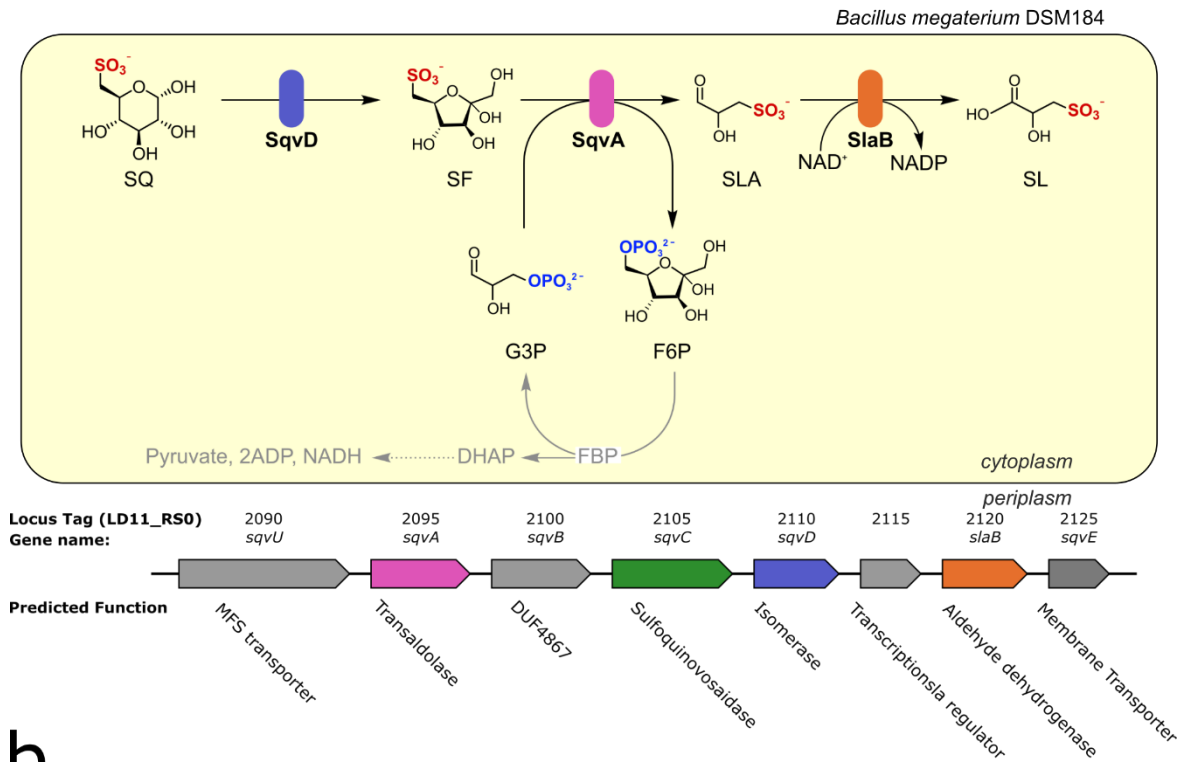
Figure 4-2: Active site conservation and multimer formation in classical TALs. A. Superposition of a representative monomeric, dimeric and decameric TAL. Residues conserved between all three are highlighted in orange. B. View of monomeric, dimeric and decameric TAL active triad. The same colour scheme was used as in A. C-E. Structures of the representative complexes formed by monomeric, dimeric and decameric TALs. In all cases NgTAL (PDB 3CLM), EcTALB (PDB 1ONR) and TmTAL (PDB 1VPX) were used for the monomeric, dimeric and decameric TAL examples respectively.

TALs. Generally, an organism that encodes a gene for a subfamily 3 TAL will also have a gene for one from subfamily 1¹⁷⁹. Subfamily 4 encapsulates TALs of ~220 aa from Gram positive bacteria and show classical TAL-like activity¹⁸⁰. Subfamily 5 is unusual as it does not contain any TALs. Rather, it is made up of enzymes exhibiting F6P aldolase activity. These enzymes share a similar fold and size to subfamily 4 TALs, but differ greatly in the active site, with a consensus sequence of V/A/I/L/T-A- P-Y-V/L-N/G-R-I/V/L/M-D, versus that of subfamily 4 TALs, S-P-F-I/V/L-G-R-I/V/L/M-D-D. Enzymes of subfamily 5 are typically found in γ -proteobacteria¹⁸¹. TALs are thought to be regulated through phosphorylation, as several sites

are predicted to exist on conserved sites in classical TALs including *E.coli* Tal A and B. Only one site, Ser₃₇, has actually found to be phosphorylated *in vivo* in OrfU, a TAL in *B.subtilis*¹⁸². While the sequence of TALs can vary heavily, the overall fold of an $\alpha_8\beta_8$ barrel is well conserved. Additionally, many residues within the centre of the barrel are positionally conserved. Typically these are involved in substrate recognition, though three of the residues, Lys-Glu-Asp form an active triad responsible for the TAL mechanism¹⁸².

Most TALs are homodimers, though they can also be monomeric (Fig 4-2, green), or form tetramers or decamers (Fig 4-2)¹⁸. These quaternary structures do not appear to confer any cooperativity between subunits, and mutation of key residues to dissolve the complex, such as an R300A mutant of *E.coli* TalB still result in active enzyme¹⁸³. Multimer formation relies heavily on the conformation of the C-terminal α -helix (CTH). In dimeric TALs (Fig 4-2, blue), this folds back along the top of the TIM barrel on the C-terminal side, partially occluding the centre and active site. Two monomers then associate with the CTHs antiparallel relative to each other using a hydrophobic surface. Decameric TALs (Fig 4-2, purple) have an amphipathic CTH in an extended conformation. This is used to bind a hydrophobic cleft on an adjacent monomer at an included angle of 72°, which when repeated 5 times forms a closed ring. Two rings then associate using hydrophobic surfaces on one side to form a 2-member ring, with each ring facing away from the other. Decameric TALs and class I F-1,6-P aldolases form very similar arrangements and even feature highly similar mechanisms¹⁸⁴. A major difference between the two is the location of the catalytic Lys. In both cases, this is located on the inner surface of the barrel however in TALs it is found two β -sheets removed (β_4) compared to F6P aldolase (β_6). This circular permutation is not common across the entire binding region, indeed the phosphate binding sites of TalB and F6P aldolase occupy the same positions – the cleft between each loop following sheets β_8 and β_1 ¹⁸⁵. The active site residues of both enzymes are also positionally conserved, with the exception of a second active site Lys, which is present in F6P aldolase but not TALs. This second Lys may bind the C1 phosphate, which is present on aldolase substrates but not TAL substrates¹⁸⁶. The residue in the equivalent position in TALs is a conserved Ser which may form part of the substrate binding region. This difference has led to the proposal that the second

a



b

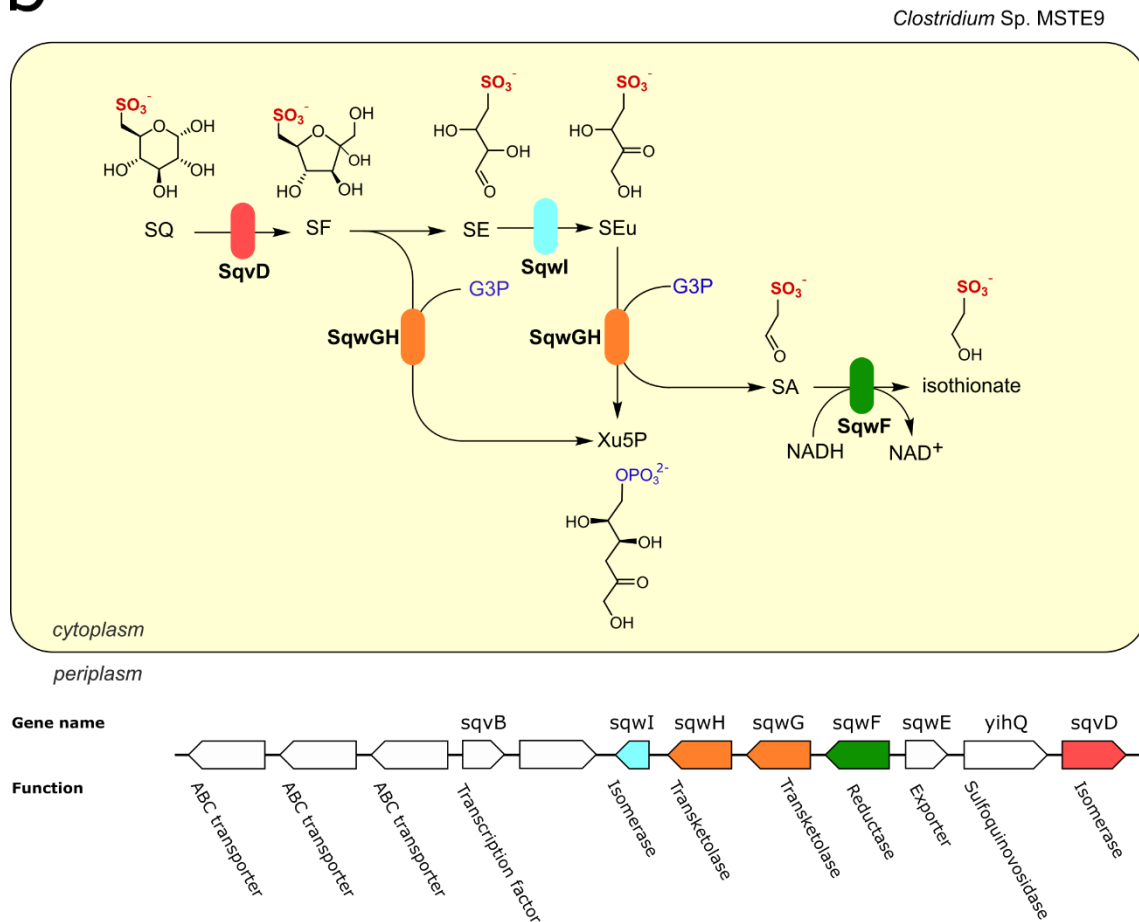


Figure 4-3: Sulfoglycolytic pathways featuring TALs (A.) and TKs (B.). Operons are presented below each pathway.

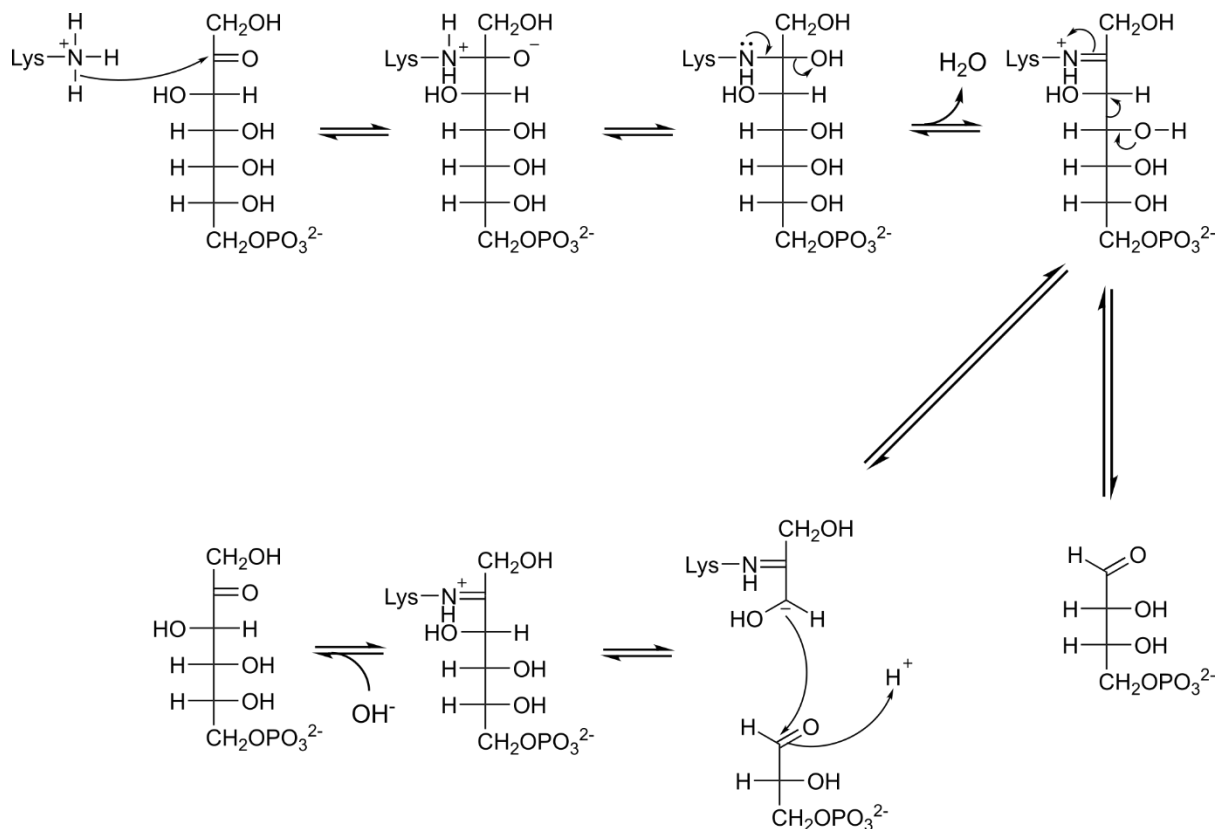


Figure 4-4: Proposed activity for SqvA. Generalized mechanism of classical TALs, with F6P as a donor substrate and G3P as a donor.

d Lys is acting as an acid/base catalyst to effect Schiff base cleavage ^{185,186}.

TALs are active against a large variety of acceptor and donor substrates, and these are processed with a variety of K_M values. The most common acceptor substrate is G3P, though it only achieves 8% of the highest recorded V_{max} with E4P.

4.2.2. TALs in sulfoglycolysis

In 2020, two pathways were discovered in *Bacillus megaterium* DSM1804 and intestinal *Firmicutes* such as *Clostridium symbiosum* and *Eubacterium rectale*. These work around a novel 6-deoxy-6-sulfofructose (SF) TAL. LC-MS pathway characterization assays, similar to those used to first characterize the sulfo-ED, sulfo-EMP and sulfo-SMO pathways, ^{40,41,80} were used to show the production of Fructose-6-phosphate (F6P) and Sulfolactaldehyde (SLA) with the addition of SF and G3P ^{78,79}. These sulfo-TALs therefore effectively allow the net transfer of the sulfonate group from a 6C donor to a 3C acceptor (Fig 4-3,a, 4-4,a).

Recently, a sulfo-TK was discovered in *Clostridium* sp. MSTE9. This appears to be a heterodimer, SqwGH, and it produces xylose-5-phosphate (Xu5P), G3P and 4-deoxy-4-sulfoerythrose (SE). SE then spontaneously isomerises to 4-deoxy-4-sulfoerythulose (SEu). SEu is then used by SqwGH again to transfer a second 2C ketol to G3P forming sulfoacetaldehyde (SA). Thus, the final products are SA and Xu5P, and the former is reduced to isethionate and excreted while the latter is used by the cell (Fig 4-3,b) ⁷⁵.

4.2.3. *Substrate recognition and TAL activity*

The active site of classical TALs is highly conserved across all branches of life. In particular, the catalytic Lysine used to form the Schiff base is positionally conserved.

A generalised mechanism has been proposed as follows: The ϵ -N of the catalytic Lys performs a nucleophilic attack on the donor compound C2 forming a carbinolamine. The catalytic Lys then transfers a proton to the active site Glu via a water. This causes formation of hydrogen bonds between the C2 oxygen and the water as well as the side chain of a nearby Thr residue. Removal of water causes the formation of the Schiff base. Aldol cleavage is then performed between C3 and C4 on the donor causing liberation of the C3 product, G3P. The remains of the donor remain stable on the enzyme until addition of the acceptor substrate ¹⁸, at which point the mechanism reverses, ending with the hydrolysis of the Schiff base and release of the acceptor product (Fig 4-4,b) ^{182,187}.

4.2.4. *Goals of this chapter*

While SqvA has been confirmed to be active on SF, details of the exact mechanism and structure remain unclear. Here, the SF transaldolase SqvA will be structurally and biophysically characterized, with the specific goals of conforming the solution multimeric state and the capacity to bind SF and other likely substrates. Activity against SF should be plainly visible in any structure as the ligand will form a Schiff base with the catalytic Lys. Notably, due to the reversible nature of the typical TAL mechanism and use of a phosphorylated cosubstrate, SqvA should be able to recognise phosphate groups as well as sulfonates. Therefore, structural characterization will be carried out not only to confirm the fold and activity but to assess the exact method for this phosphate/sulfonate co-specificity.

4.3. Materials and Methods

4.3.1. Gene Expression and protein purification

The gene for SqvA was codon optimised for *E.coli* expression and ordered from Eurofins Genomics in a standard pET28a vector. This was transformed into BL21(DE3) competent *E.coli* (NEB) using the standard NEB protocol, onto LB-agar with 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, and grown at 37 °C overnight. Isolated colonies were picked and used to form liquid precultures in 20ml LB with 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, then grown at 37 °C, 180rpm overnight. 5ml of preculture was used to inoculate every 1L of expression media (2x YT + 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin). Expression cultures were grown at 37 °C, 225 rpm until an OD₆₀₀ of 0.6-0.8 was achieved. Then, isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final 1mM concentration and induction took place over 18 hours, 18°C, shaken at 180rpm. Cells were harvested by centrifugation at 4500 rpm, 20 minutes, 4 °C. The supernatant was discarded, and the pellet resuspended in 40 ml binding buffer (50 mM tris, 300 mM NaCl, 30 mM imidazole, pH 7.4) per 1L culture. 1 tablet of EDTA-free protease inhibitor was also added per 1L. Cells were lysed either by sonication (5 pulses of 100% amplitude, 60s on, 60s off while on ice) or disruptor (25 kPSI, 4 °C, eluted into an ice-cooled flask). Cell debris was then pelleted by centrifugation at 18,000 rpm, 40 minutes, 4 °C. The pellet was discarded, and the soluble fraction was loaded to a 5ml Ni-NTA Crude FF column at 2 ml.min⁻¹. The column was pre-equilibrated with binding buffer. A linear gradient into elution buffer (50 mM tris, 300 mM NaCl, 500 mM imidazole, pH 7.4) was used to elute his-tagged protein into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. Protein-containing fractions were then pooled and buffer-exchanged into binding buffer due to high imidazole sensitivity of SqvA. The pooled fractions were then concentrated to a 2ml final volume and loaded to a HiLoad superdex 16/600 S200 column for size-exclusion chromatography. The column was pre-equilibrated into SEC buffer (50 mM tris, 300 mM NaCl, pH 7.4). Elution took place using SEC buffer at 1 ml.min⁻¹ into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. SqvA-containing fractions were pooled and concentrated to a working concentration of 38 mg.ml⁻¹, confirmed by A₂₈₀, then snap-frozen with liquid nitrogen. These were subsequently stored at -70 °C. For selenomethionine-labelled protein, all steps were identical with the following differences: M9 minimal media was used for expression, to which 10ml of preculture was added. Upon reaching an OD₆₀₀ of 0.6-0.8, 5 ml of "magic 6" amino acids (100 mg Lysine, Phenylalanine and

Threonine, 50 mg Isoleucine, Leucine and Valine in 10 ml H₂O) plus 60 mg selenomethionine was added per 1L media, and growth temperature was maintained for a further 15 minutes, followed by induction with IPTG as standard.

All SDS-PAGEs for all proteins were performed using 12% acrylamide gels and run for 45 minutes at 200 V with 0.1% SDS TRIS/Glycine running buffer. Gels were subsequently stained with Coomassie magic blue dye. Unless otherwise indicated, 7 μ l of sample was loaded, with 5 μ l loading dye and 3 μ l water for a 15 μ l total load.

4.3.2. *Intact ESI mass spectroscopy*

Purity, mass, and labelling extent estimates were obtained using electrospray-ionisation mass-spectroscopy (ESI-MS) with a Bruker Maxis-HD qTOF mass spectrometer. All samples were prepared to 1 μ M with 50% by volume acetonitrile, 1% by volume formic acid. Infusion was at 3 μ l.min⁻¹. Spectra were averaged over 1 minute and smoothed in 1 cycle using a Gaussian method, then the baseline was subtracted with a flatness of 0.8. Compass 1.7 software was used for this.

4.3.3. *SEC-MALLS*

Size-Exclusion Chromatography with Multiple Angle Laser Light Scattering (SEC-MALLS) was performed using 120 μ l samples of SqvA at 2 mg.ml⁻¹, in 50 mM TRIS, 300 mM NaCl, pH 7.4. The 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 HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Work was conducted at room temperature (20 \pm 2 $^{\circ}$ C). Solvent was 0.2 μ m filtered before use and a further 0.1 μ m filter was present in the flow path. The column was equilibrated with 2 column volumes 50 mM TRIS, 300 mM NaCl, pH 7.4 before use and flow was continued at the working flow rate until baselines for UV, light scattering and refractive index detectors were all stable. Sample injection volume was 100 μ l. LC Solutions (Shimadzu) was used to control the HPLC and Astra V was used for the HELEOS-II and rEX detectors. The Astra data collection was 1 minute shorter than the LC solutions run to maintain synchronisation. TM 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).

4.3.4. NanoDSF

All SqvA samples were at 1 mg.ml⁻¹ in 50 mM TRIS, 150 mM NaCl, pH 7.5. All ligands were weighed dry and hydrated with 50 mM TRIS, 150 mM NaCl, pH 7.4 to a stock concentration of 50 mM. All ligands and cofactors were added 10 minutes prior to loading to a final concentration of 2 mM, and three technical repeats were run of each sample, with each replicate loaded to a standard glass capillary. The temperature was increased from 20 °C to 95 °C at 1 °C.min⁻¹ in a linear gradient and Trp fluorescence was measured by a NanoTemper Prometheus NT.48. All data processing and graphing was performed in Origin.

4.3.5. X-ray crystallography and structure determination

All SqvA crystals were grown at 20 °C. Crystals were harvested into liquid nitrogen, using nylon CryoLoopsTM (Hampton). All soaking took place over 5 seconds, in mother liquor with glucose added dry. Cryoprotection took place over 5 seconds and used 10% glycerol, or 10% ethylene glycol in mother liquor. All crystals were tested using the in-house Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV. X-ray data collection took place at Diamond Light Source on beamlines I-03 and I-04. Data collection statistics are available in the results section. Data indexing and initial processing was performed using either DIALS or 3dii pipelines from the Xia2 package ^{120,121}. Data reduction was performed with AIMLESS, and resolution was cut until CC1/2 = 0.5 ¹²². Molecular replacement was performed using MOLREP, using a monomer of the ligand-free EM structure of SqvA as a search model ¹²³. Automated model building was performed using BUCCANEER ¹²⁵. Diagram preparation for molecular models was performed using CCP4MG, Chimera or Pymol, depending on the desired outcome ¹²⁹⁻¹³¹.

4.3.6. Cryo-Electron Microscopy and structure determination

Prior to sample loading, all grids were glow-discharged at 20 mA, 10 sec hold, 3 min discharge and 0.38 mBar pressure. All grids were either UltraAuFoil R1.2/1.3 or UltraAuFoil R2.0/2.0,

depending on the experiment. Sample vitrification was performed using a Vitrobot mk IV (ThermoFisher Scientific), with a blot force of -10 arbitrary units, and blot time of 4 or 6 seconds. Humidity within the sample chamber was 100% and temperature was 4 °C. 2 µl of sample was added in all cases prior to blotting. All grids were vitrified into liquid ethane, then stored and handled in liquid nitrogen. The highest-resolution ligand-free SqvA dataset was produced with a grid prepared at 2.5 mg.ml⁻¹ protein in 50 mM TRIS, 300 mM NaCl pH 7.4. The highest resolution SF dataset used 2.5 mg.ml⁻¹ protein, 2 mM SF in the same buffer.

Data collection took place on a Glacios microscope (ThermoFisher Scientific) (ThermoFisher Scientific) at 200 kV. Automated data collection was performed using EPU (ThermoFisher Scientific). Experiment-specific data collection parameters are available in the results section. Image processing and 3D reconstruction was performed using Relion v.3.0¹⁸. Motion correction was performed using MotionCorr v.2.0, CTF estimation by GCTF v.1.06¹⁸⁹. All post-reconstruction processing and model was performed using PHENIX¹⁹⁰. Map sharpening was achieved by Phenix.autosharpen, and model validation by molprobity^{191,192}. Fourier-shell correlation plots can be found as fig 4-13.

All model building and validation was performed using Coot¹²⁷, and diagram preparation was performed identically to XRD structures.

4.4. Results

4.4.1. Expression, purification, and early characterization

The gene for SqvA was ordered into pET28a, allowing for initial purification using Ni-NTA resin due to the inclusion of a C-terminal 6xHis tag. The protein produced in *E.coli* and proved straightforward to separate from the soluble fraction and polish using IMAC and SEC respectively (Fig 4-5,a-d). SqvA is stable in solution post-purification at concentrations as high as 38 mg.ml⁻¹ however it may be sensitive to high imidazole concentrations as it quickly aggregated in the high-imidazole IMAC elution buffer. Low-temperature handling and immediate buffer-exchange to an imidazole-free buffer were sufficient to overcome the effect. The protein presents as a single band at ~25 kDa on an SDS-PAGE, with few visible

impurities. Overall, the purity of SqvA is estimated at <95%. Based on the observed monomeric Mw SqvA falls under subfamily 4, suggesting a classical TAL mechanism.

TALs typically form either solution dimers or tetramers, or large multimeric complexes resembling rings. Based on SEC-MALLS data, SqvA falls firmly into the latter category, with a clear elution peak corresponding to approximately 280 kDa. The sequence-derived Mw of SqvA is 25 kDa, which would correspond to an 11-mer in solution. However, given the typical dimer or pentamers seen in multimeric TALs, a decamer was considered more likely. The discrepancy observed here may be due to discrepancies between the analysis model and light scattering observed (Fig 4-5,e).

In order to inform structural studies, the specificity of SqvA was explored using nanoDSF. Based on early work the most likely 6C substrate for SqvA is sulfofructose^{75,78,79}. SLA is a likely product. SQ was not considered likely to bind due to the presence of an upstream SQ isomerase in the sulfo-TAL pathway and was used as a negative control to compare to the ligand-free protein. Of these, only SF caused a meaningful Tm increase, from 78.7 °C ligand-free to 84.3°C ($\Delta T_m=5.6$ °C) (Fig 4-6,a-b). No meaningful Tm shifts were observed with any

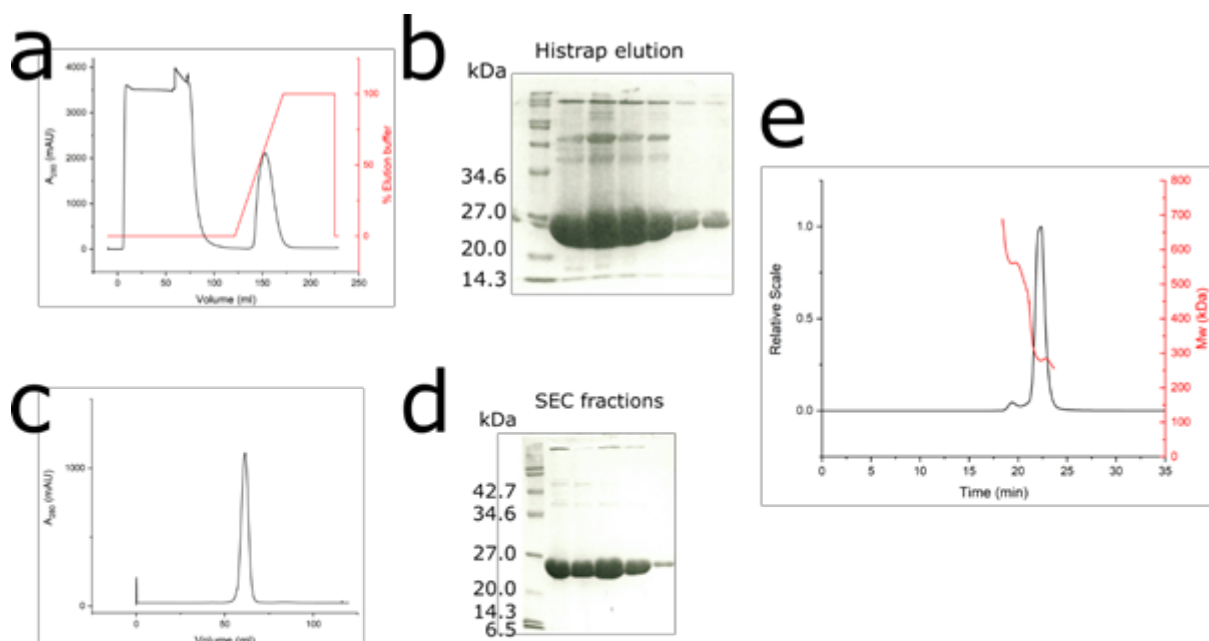


Figure 4-5: Expression, purification and solution state of SqvA. **A.** A280 elution profile of SqvA during IMAC purification. Elution buffer % with elution is shown in red. **B.** 12% SDS-PAGE of elution fractions, with broad range Mw ladder on the left. **C.** A280 elution profile of size-exclusion chromatography for SqvA. **D.** 12% SDS-PAGE of fractions under SEC peak. Broad range Mw ladder on the left. **E.** SEC-MALLS elution and scattering profiles of SqvA. Scattering profile is in red, and clipped to elution A280 peaks.

other ligand. Therefore, SqvA was considered likely to form a strong enough interaction with SF to allow structural characterization.

Many classical TALs show wide substrate promiscuity, and to assess if this was the case for SqvA, nanoDSF was also used to assay binding with 2 mM fructose-6-phosphate (F6P), sedoheptulose-7-phosphate (S7P) and glyceraldehyde-3-phosphate (G3P). The former two are substrates of many classical TALs, and G3P is a common donor substrate, and likely used for SqvA. None of these showed any meaningful ΔT_m compared to a ligand-free control. Arabinose-5-phosphate (A5P) forms a covalent complex with classical TALs but cannot be

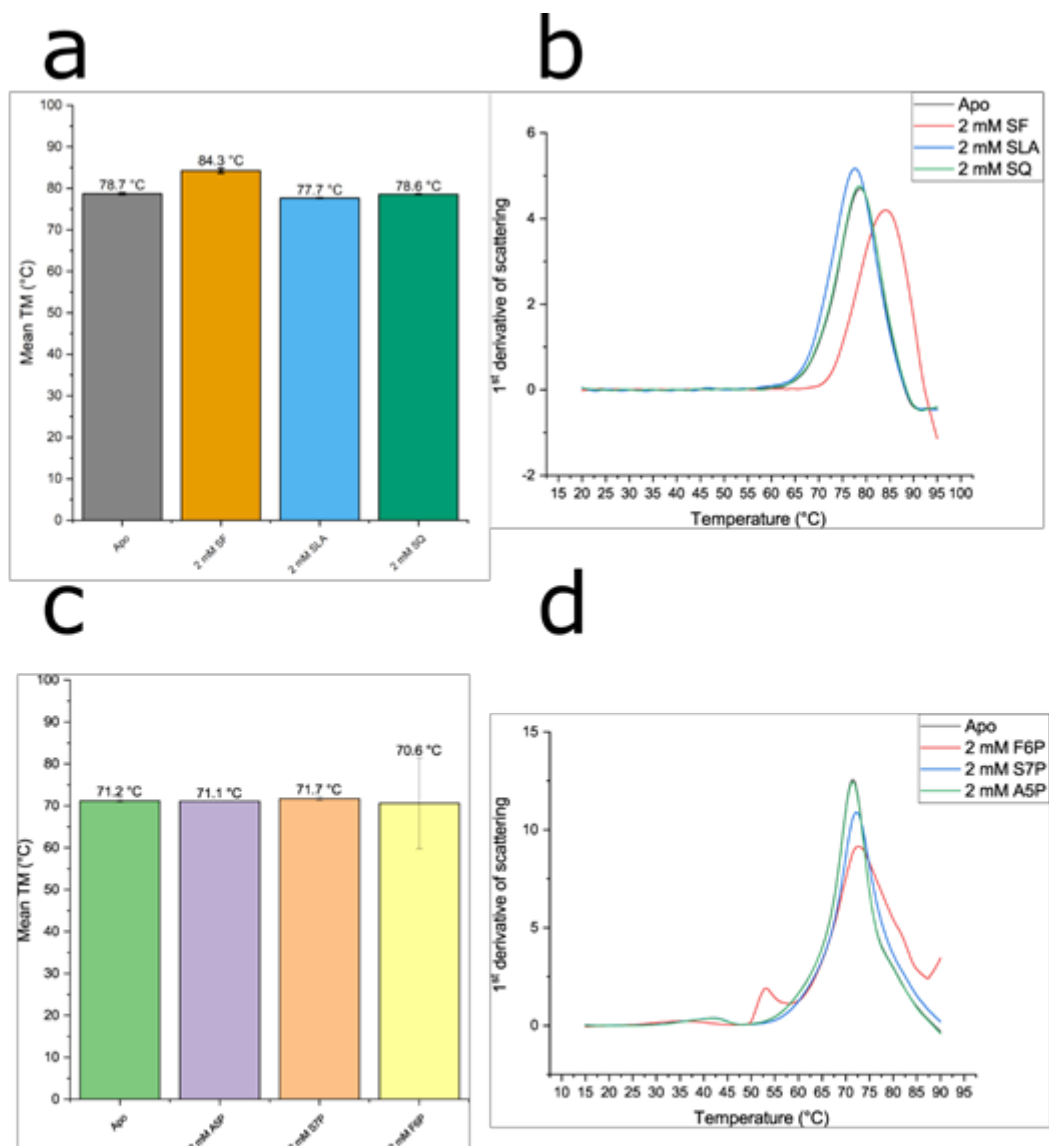


Figure 4-6: T_m shift assays of SqvA. **A.** Mean T_m of SqvA in ligand-free state, and with 2 mM SF, SLA and SQ. **B.** Differentiated mean scattering profiles for A. **C.** Mean T_m of SqvA in ligand-free state, and 2 mM A5P, S7P and F6P. **D.** Differentiated mean scattering profiles for C. Datasets are split as the ligands used in C, D required a lower instrument sensitivity (9%) than A,B (30%).

used as a dihydroxyacetone donor, making it an effective inhibitor of the classical TAL mechanism^{193,19}. However, it also failed to produce a T_m shift indicative of binding (Fig 4-6,c-d). For all substrates, a lack of T_m shift was not considered definitive evidence of no binding, merely a likelihood that any binding present would be of low affinity, a conclusion based on the low number of additional bonds between ligand and protein compared to the size and strength of the full complex seen in other decameric TALs⁷⁷. Therefore, structural characterization was still attempted with all listed ligands.

Typically, the ratio of 350/330nm fluorescence is used in nanoDSF to determine T_m , however with SqvA the scattering at 330nm was more consistent between repeats. This may be due to the decameric solution state of SqvA having multiple Trp shifts during unfolding, reflecting the dissolution of the complex prior to or during denaturation. Scattering is reflective of the onset of aggregation, and so may be higher than the true T_m . This is irrelevant however as ligand binding could still increase thermal stability as evidenced by the addition of 2 mM SF.

4.4.2. A two-pronged approach to structural characterization

Early hits in the JCSG and PACT commercial screens showed SqvA readily crystallises, with optimisations providing crystals large enough (~0.5 mm) to reliably harvest. However, these never resulted in solvable data collections. To rule out a bad MR model, experimental phasing was attempted by producing selenomethionine-labelled protein, SeSqvA. SeSqvA purified identically to the native protein and had similar tolerances to protein and imidazole concentration. Intact ESI mass spectroscopy confirmed replacement of all 3 methionine residues with selenomethionine (Fig 4-7,b). SeSqvA produced crystals in similar conditions to those producing large native protein, with the farthest-diffracting dataset produced in 0.1 M MMT pH 5.4, 22% PEG1500 in a 1:1 protein:mother liquor ratio. This dataset failed to produce a sufficient anomalous signal for experimental phasing (fig 4-7, a).

Due to its status as a solution decamer SqvA was considered a large enough target for cryo-EM. Early micrographs of ligand-free protein, as well as incubated with SF, G3P, A5P, F6P and E7P showed SqvA is stable on gold grids, forming complexes with a clear two-layer ring structure (fig 4-8, a). The symmetry inherent to the shape did however mean the

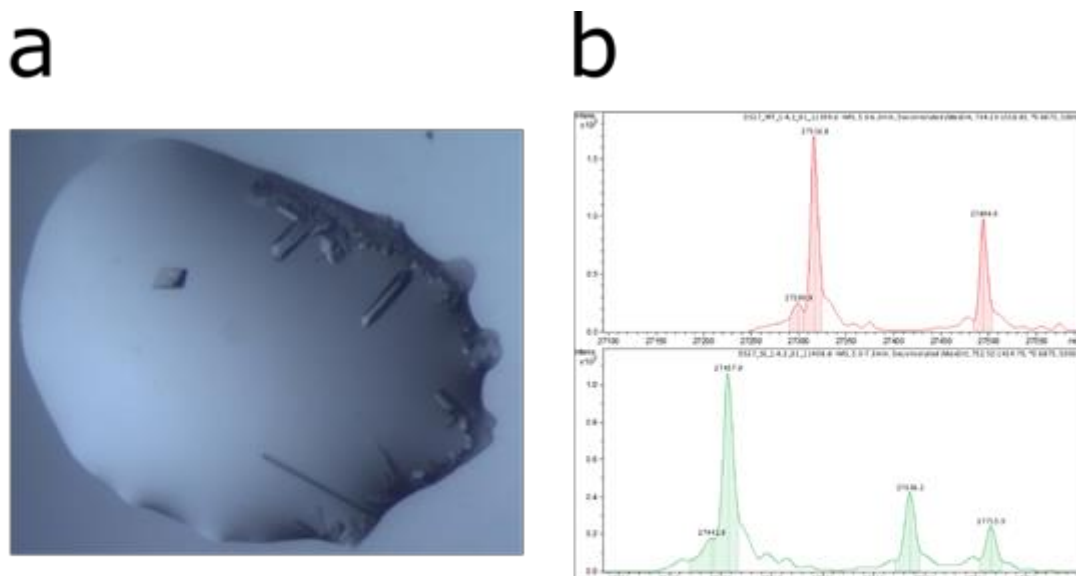


Figure 4-7: Crystallography of SqvA. A. Micrograph of farthest diffracting SeSqvA crystal drop. **B.** Intact MS-MS traces of SeSqvA (red) and native SqvA (green).

overwhelming majority of particles were either side-on or facing, though enough angled particles were present to form classes (fig 4-8, b). Additionally, particles with the side-on orientation often stack into a tubular arrangement (fig 4-8, b-c) This appears to be dependent on the packing of particles on the grid and is not ordered. No larger complexes were observed in SEC-MALLS, so this behaviour is thought to be an artefact of the unnatural conditions on the grid rather than being of any physiological relevance. This produced “fountains” stemming from each pole of all side-on 2D class averages, as the circular soft mask added to 2D class averages cannot fully remove the stacked particles without clipping the target particle.

Early 3D reconstructions of the ligand-free structure showed a clear pair of pentameric rings, facing away from one another, and offset by about 36° from one another on the Z-axis (fig 4-8, c). This allowed for an assignment to symmetry group D₅. This fivefold symmetry proved valuable as further refinements could be done under the assumption that the structure repeated exactly 5 times across the ring, effectively pentupling the amount of available data in a collection. This, combined with the high contrast and particle count in the micrographs allowed for reconstruction of the ligand-free complex to 2.6 Å (fig 4-8, d). The same process was repeated with SF producing a structure at 2.1 Å. The “fountains” observed in 2D class averages were also present in all unmasked maps, though they did not refine into any useful data. Masking after 3D refinement effectively removed them.

	SqvA apo	SqvA·SF (D5)	SqvA·SF (C1)
Data collection and processing			
Magnification	240000	310000	310000
Voltage (kV)	200	200	200
Electron exposure (e ⁻ /Å ²)	50	50	50
Pixel size (Å ²)	0.54	0.45	0.45
Symmetry	D5	D5	C1
Initial particles	153063	77489	77489
Final particles	59063	53450	53450
Map resolution (Å)	2.6	2.0	2.7
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	34.1-2.02	44.69-1.90	30.69-2.50
Refinement			
Initial model	-	SqvA apo	SqvA·SF (D5)
Map sharpening <i>B</i> factor (Å ²)	-66.85	-39.15	-11.39
Model composition:			
Non-H atoms	15550	16326	16124
Residues	2180	2160	2160
Ligands	0	10	10
Water	6	216	14
Mean <i>B</i> factors (Å ²):			
Protein	56.5	55.5	64.6
Ligand	-	30.0	64.7
Water	56.5	55.1	69.0
RMS deviations:			
Bond length (Å)	0.012	0.039	0.005
Bond angles (°)	1.85	2.21	0.945
Validation			
MolProbity score	2.8	2.4	2.3
Clashscore	4.98	13.0	20.3
Poor rotamers %	18.8	6.12	0
Ramachandran plot:			
Favoured %	89.25	97.14	97.91
Allowed %	10.28	2.86	2.29
Outliers %	0.47	0	0
PDB code	-	-	-

Table 4-1: Micrograph collection parameters, reconstruction, and model validation statistics for EM structures of SqvA.

	SqvA apo
Data collection and indexing	
Space group	P 2 ₁
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.89, 120.8, 152.6
α , β , γ (°)	90,90.4,90
Resolution (Å)	3.2
R _{merge}	0.165(0.593)
R _{pim}	0.099(0.337)
<i>I</i> / σI	7.5(2.5)
CC 1/2	0.98(0.68)
Completeness (%)	100(100)
Redundancy	6.8(7.2)
Refinement	
Resolution (Å)	3.2
No. unique reflections	48733
<i>R</i> _{work} / <i>R</i> _{free}	0.24/0.29
Atom Count	
Protein	15550
Ligand/ion	0
Water	61
<i>B</i> factors (Å²)	
Protein	57
Ligand/ion	-
Water	32
Bond lengths (Å)	0.0158
Bond angles (°)	2.08
Ramachandran residues	
In most favourable regions (%)	95.1
In allowed regions (%)	4.5
Outliers	0.4
PDB code	-

Table 4-2: XRD data collection and refinement statistics for ligand-free SqvA. Values in parentheses are for the highest resolution shells. All values are within *I*/ σI where applicable.

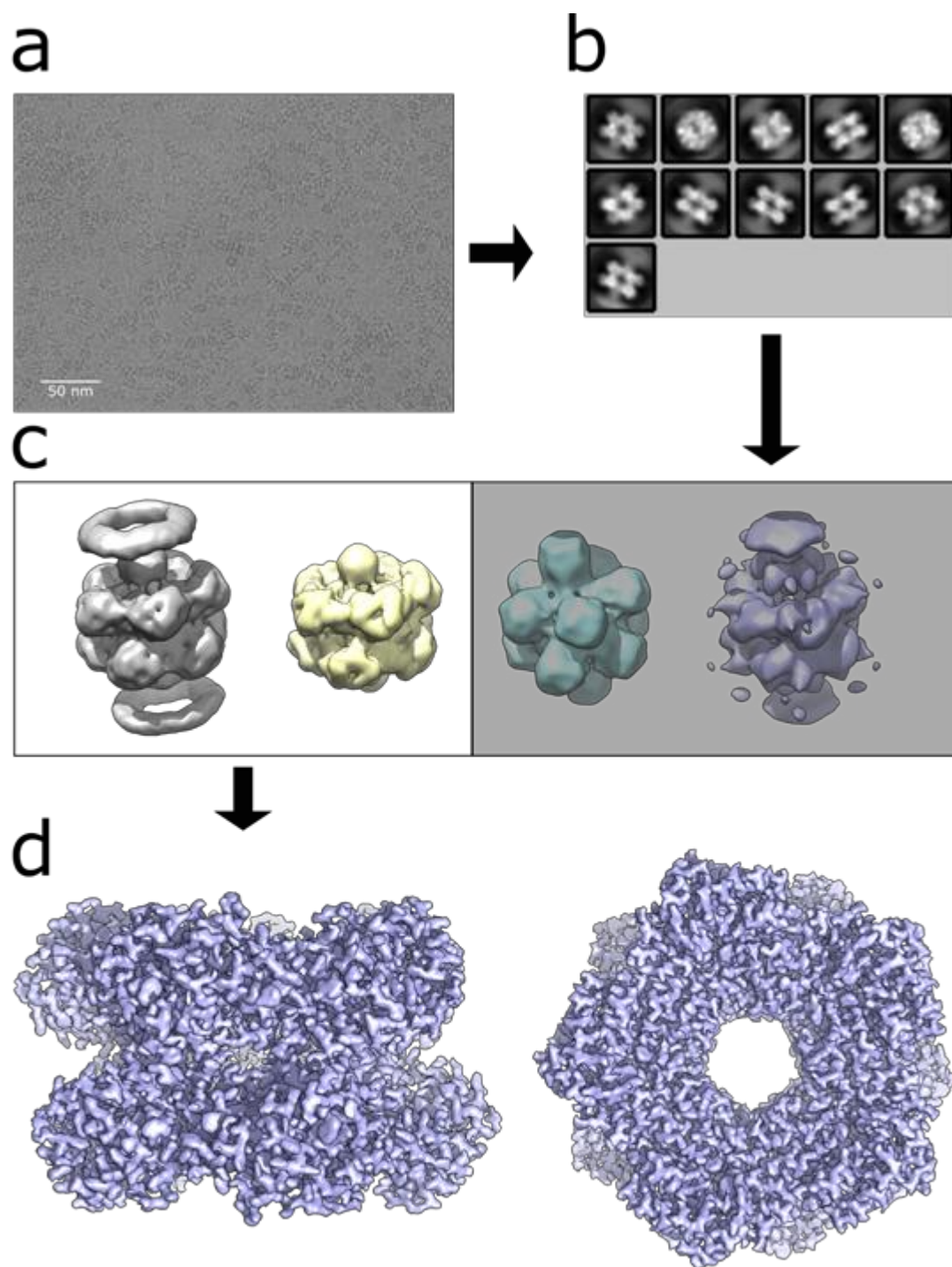


Figure 4-8: 3D reconstruction of ligand-free SqvA through *Relion*. **A.** Representative raw micrograph of ligand-free SqvA on-grid. Taken at 0.45 Å/px resolution, 2 μm defocus. **B.** Final 2D classes selected for 3D reconstruction at 1.8 Å/px resolution. **C.** 3D classification of SqvA reconstruction at 1.8 Å/px particle resolution. Greyed-out classes were discarded due to excessive deviation from the expected structure and low number of constituent particles. **D.** Final D5 reconstruction of ligand-free SqvA at 0.9 Å/px particle resolution. A side-on and top-down view is shown.

Automated model building and symmetry-based density clipping using PHENIX were at first unsuccessful, producing models and density unrecognisable as a viable protein. This standard model building pathway may have failed due to the symmetry of the overall structure being in D5 despite that symmetry operator not encompassing any complete chains. Boxing the density based on D5 instead produces density for one monomer without the CTH, and one bisected down the centre of the TIM barrel, preventing model fitting either by alignment or chain tracing. Automated model building of the entire decamer was deemed too computationally expensive when another option was available. A successful monomeric base model was instead produced using AlphaFold 2.0. The CTH was then removed from this to aid in alignment, and the prediction was aligned to an early α -helical fragment produced in PHENIX Map-to-model. The AlphaFold model was then modified to fit the density and copied to form the decameric complex by alignment to a copy of the density lowpass-filtered to a 5 Å resolution. Any further refinements were applied to chain A, then copied to the rest using the NCS copy utility in Coot (fig 4-9, a-b). The refined ligand-free SqvA monomer and unmodified AlphaFold model share an RMSD of 0.54 Å

4.4.3. SqvA exhibits the same general fold and extended CTH seen in other decameric TALs

SqvA monomers share the $\alpha_8\beta_8$ TIM barrel fold observed in all TALs. However, they can be distinguished by the extended C-terminal α -helix (CTH), which lies away from the main fold. Both the ligand-free and SF-bound maps show well-resolved and clear side chain density and a continuous and traceable series of backbones for each monomer. The density is notably weaker in quality than the surrounding areas, possibly indicating SF is not bound in all sites or that the ligand shows conformational flexibility. All 10 monomers in each structure have an identical fold due to the symmetry used in map refinement and NCS copying used in model building. These methods both work on the assumption that all 10 monomers work identically, and no inter-subunit cooperativity exists such as that famously exhibited by haemoglobin¹⁹. As no cooperativity of that nature has been described in TALs, this assumption was considered valid. However, refinement of the SF structure was attempted without symmetry, resulting in a structure of 2.7 Å resolution. This map is notably less complete than the higher resolution one refined in D5, with a greater number of missing side chains and more rarely, missing backbone regions. While SF is present in every binding

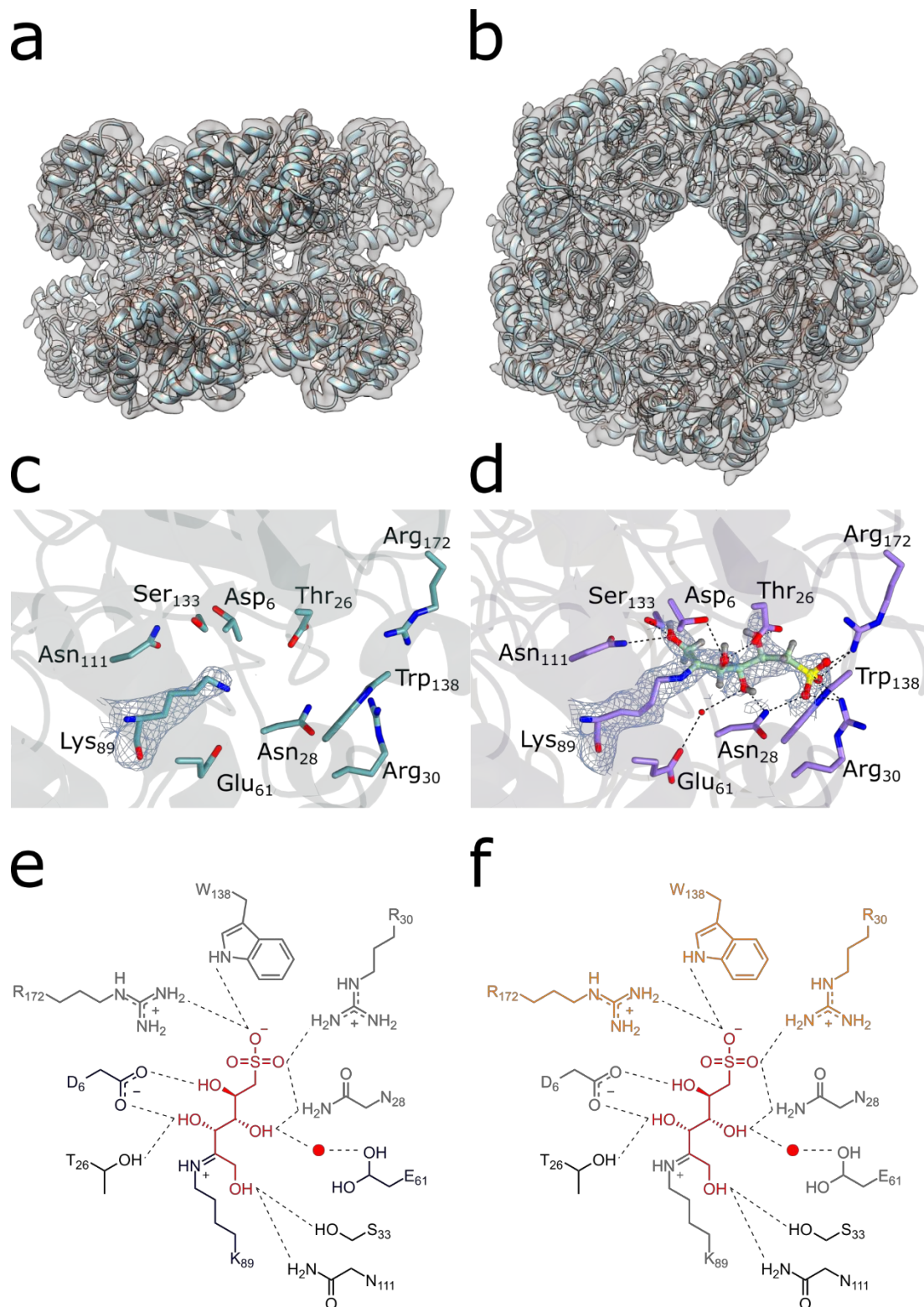


Figure 4-9: Fold and binding site of SqvA by cryo-EM. A-B. Side-on and top-down view of ligand-free SqvA final model, fitted to sharpened map. **C-D.** Ligand binding site of SqvA ligand-free (C) and SF-bound (D). In both cases density is clipped to Lys₈₉ and SF and contoured at 1.0 σ (0.27 e/ \AA^3). **E-F.** 2D representations of the SqvA-SF complex. The active triad is in blue (E.) and sulfonate binding pocket in orange (F).

site, the density is of poor quality in most monomers, with only 60% having density along the entire sugar chain. When resolved in D5 this area improves greatly, showing the full sugar chain and a clear sulfonate, however this region is still poorly resolved in comparison to the rest of each SqvA core, for which the density is unambiguous and detailed. This effect is clearly visible in the average B -factors for the final models with and without D5 symmetry, which increases from 55.5 \AA^2 in all chains and 30.0 \AA^2 for ligands in D5 to an average 64.6 \AA^2 and 69.0 \AA^2 respectively ($\Delta B_{\text{chain}} = 9.1 \text{ \AA}^2$, $\Delta B_{\text{ligand}} = 39.0 \text{ \AA}^2$). Given the small overall change in fold (global RMSD = 0.372 \AA) this large increase must be primarily caused by the decrease in density quality. Based on this it seems likely that not every SqvA complex used in the reconstruction was ligand-bound. An improvement in ligand density could therefore be expected through a 3D classification using particles with lower scaling, allowing for subtraction of any particles not producing ligand density. A more effective solution however may be to increase the concentration of SF used when preparing grids. To this end further attempts at producing ligand-bound SqvA structures were performed using 50 mM of the relevant ligand rather than the 2 mM used with SF.

4.4.4. Structural basis of SF recognition and insights into SqvA activity

The binding site of SqvA is located in the centre of the $\alpha_8\beta_8$ barrel, as observed with every other TAL studied. All TALs also feature a conserved active triad of Asp-Lys-Glu, and SqvA is no exception. A comparison of the density between ligand-free and SF-bound SqvA shows SF in an open-chain form and with an extended conformation, with clear evidence of a covalent bond formed to Lys₈₉. SqvA must, therefore, be forming a Schiff base, confirming the activity previously seen against SF⁷⁹. The Schiff base forms in place of the previous C2 hydroxyl of SF. The active site is completed by hydrogen bonding between the C3 and 5 hydroxyls and side chain hydroxyls of Asp₆, and C4 hydroxyl and a side chain hydroxyl of Glu₆₁, via an ordered water. This binding mode is identical to that seen in Tal-B against E4P, including the coordinated water (fig 4-9, c,e)⁷⁷. This supports the idea that SqvA follows the same catalytic mechanism (fig 4-4), wherein the ordered water is the product of the condensation involved in Schiff-base formation. The mechanism would continue with the nucleophilic attack of G3P by a cleaved dihydroxyacetone, producing F6P, which then dissociates following dissolution of the Schiff base. Additional hydrogen bonds are formed between the terminal hydroxyl of

SF and side-chain hydroxyl of Ser₁₃₃, and side chain primary amine of Asn₁₁₁. The side chain hydroxyl of Thr₂₆ also forms a hydrogen bond to the SF C3 hydroxyl.

The sulfonate binding pocket appears to coordinate two of the three sulfonate oxygens. The first of these forms hydrogen bonds to the indole secondary amine of Trp₁₃₆ and side-chain primary amine of Arg₁₇₂ (fig 4-9, c,d,f). The second is bound by the side-chain primary amines of Arg₃₀ and Asn₂₈. As classical TAL mechanisms, which SqvA is likely to also use are always reversible, F6P should be capable of forming an active complex with SqvA, and the sulfonate binding pocket must also bind phosphate. However, no binding was observed with nanoDSF, and no condition tested in grid preparation produced an SqvA•F6P complex. Therefore, the differences in sulfonate and phosphate binding modes could not be directly characterized. Additionally, no complexes with G3P or SLA could be produced. This was less surprising however as TALs generally must bind the donor substrate first. As addition of the acceptor substrate to the donor binary complex would result in active enzyme, a ternary complex would not be feasible using traditional grid vitrification methods. A time-resolved system can reduce the delay between substrate addition and vitrification to the ms timescale and may provide the only viable method of obtaining these. No structures were produced with S7P or A5P, which may indicate a higher degree of substrate specificity than typically observed with TALs and differences to the mechanism respectively. However, these may simply have not bound due to a lack of optimised grid conditions, or bound in a minority of particles, preventing resolution of the ligand density.

4.4.5. Structural solutions by cryo-EM and XRD show variations in SqvA fold and conformation

Initially no XRD dataset could be solved due to the lack of an MR model of sufficient quality and lack of crystallised SeMet-labelled protein, precluding the straightforward use of experimental phasing by MAD. However, the advent of AlphaFold and near-simultaneous solution of a ligand-free EM structure allowed a solution on the highest resolution XRD dataset available. This dataset diffracted to 3.1 Å resolution in spacegroup P 2₁. Solution by MR used chain A of the refined ligand-free EM structure. The resulting crystal structure contains one full biological assembly in the ASU (10 monomers) forming the same structure as observed with EM and was refined to an R/R_{free} of 0.24/0.29. The crystal is densely

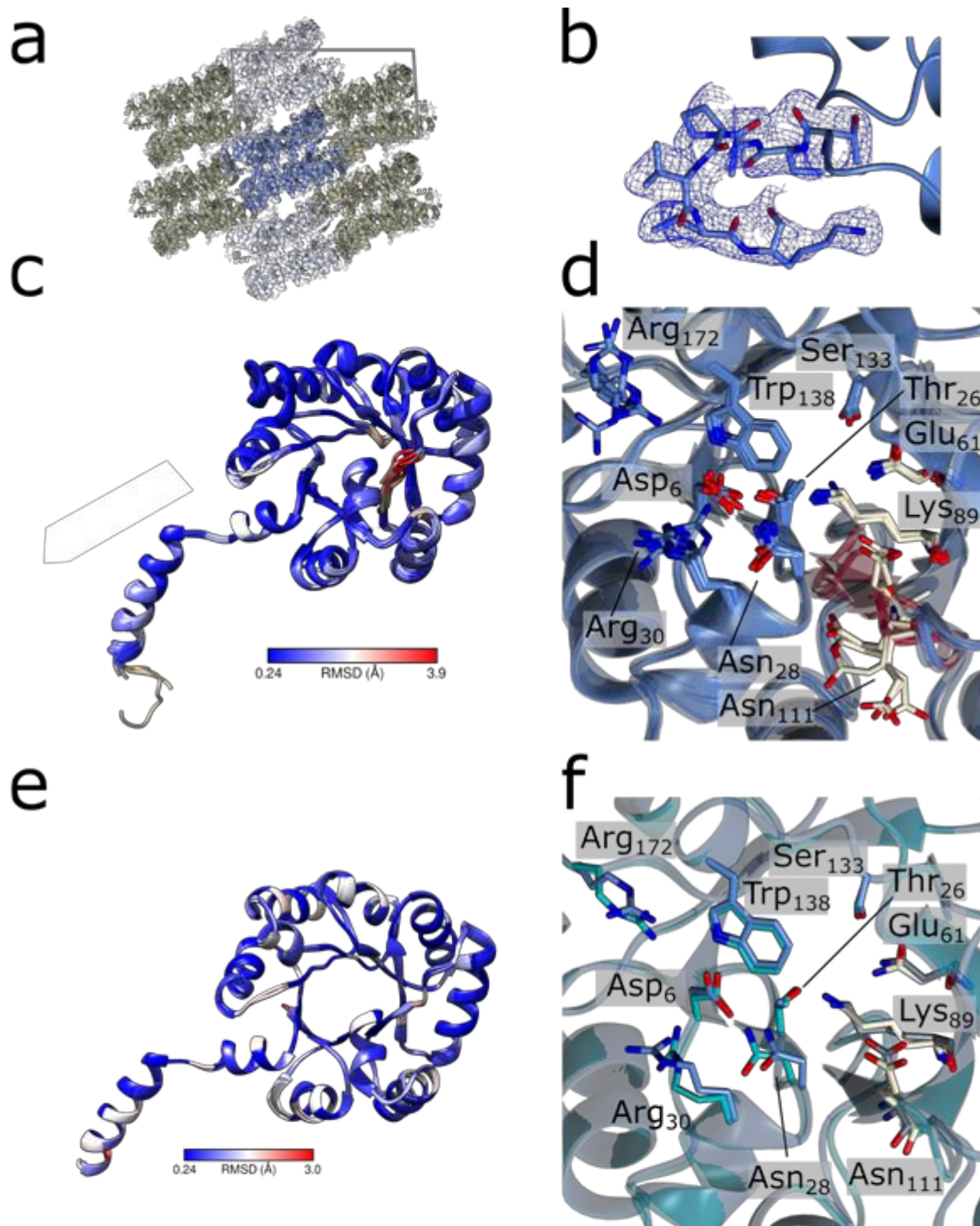


Figure 4-10: Crystal structure of ligand-free SqvA, and differences relative to EM reconstructions. A. Crystal packing of SqvA. Modelled complex is in blue, and the unit cell is shown. **B.** C-terminal region of SqvA. Electron density is contoured at 1.0σ ($0.21 \text{ e}/\text{\AA}^2$). **C.** All-atom RMSD heatmap of all SqvA monomers in the ASU, where blue denotes a low (minimum of 0.24 \AA) RMSD and red denotes high (maximum of 3.9 \AA). **D.** Superposition of all active sites in the ASU. Active triad are in white, and mainchain colouring is retained from C. **E.** Global RMSD heatmap between ligand-free SqvA solved by EM and XRD. A random chain from the XRD structure was used. Colouring is retained from C. (maximum RMSD 3.0 \AA). **F.** Superposition of EM (cyan) and XRD (blue) SqvA active sites. Active triads are in white (grey for EM).

packed, each assembly forming crystal contacts with 6 neighbours (fig 4-10, a). This structure contains details not found on any other SqvA structure, notably the last 6 residues of the CTH are present in 2 of the monomers (fig 4-10, b). These are typically averaged out during map refinement during EM reconstruction. The electron density shows slight differences between monomers, indicating slight differences in fold across all monomers in the ASU. These are especially prevalent in sheet β_3 , which contains the catalytically important Asn₁₁₁ (fig 4-10, c). This residue as well as binding site residues Arg₃₀ and Arg₁₇₂ display a variety of side chain conformations (fig 4-10, d). The remainder of the active site residues all retain approximately the same conformation between chains. All three of the moving residues are naturally flexible and may not be sufficiently constrained when SqvA is ligand-free to adopt a consistent conformation. It is worth noting that the limited resolution of the dataset prevents any biologically relevant conclusions beyond this from being assigned. Additionally, the backbone of β_3 shows more variation than anywhere else in the protein. Again, this may arise from the limited resolution of the dataset though the mainchain in the core of a protein is usually better resolved. A superposition of the ligand free XRD and EM structures shows only the N-terminus having a mean global Δ RMSD of $>0.5 \text{ \AA}$. There is no significant change in conformation for residues in the active site and substrate binding region, with the exception of chain-specific changes in the XRD structure (fig 4-10, e-f).

4.4.6. *Structural basis of decamer formation*

SqvA forms a dimer of pentamers, giving rise to the overall decameric, two-member ring complex. Pentamer formation relies primarily on the C-terminal α -Helix (CTH), which lies in an extended conformation away from the main fold of each monomer (fig 4-11, a-b). This helix is amphipathic, with the hydrophobic face placed inwards. The adjacent SqvA monomer contains a hydrophobic groove along the top of the $\alpha_8\beta_8$ barrel, which accommodates the CTH. This interaction places the CTH near the active site, however no residue from it directly interacts with the substrate. Therefore, SqvA does not use the “half site” model previously proposed for some decameric TALs¹⁹. This hydrophobic interaction is further stabilised by hydrophobic interfaces along helix α_2 and sheet β_6 . The overall included angle of this interface is 72° , leading SqvA to repeat exactly 5 times until each monomer both binds a cleft

and has its own cleft bound, which produces the pentameric ring (fig 4-11, c). A pair of rings forms the full decamer by binding facing away from one another, with a Z-axis rotation of 36°. Unlike with pentamer formation this is a mostly polar interaction involving helix $\alpha 7$ on each monomer. These face inwards and contain a range of polar amino acid side chains that form a network of hydrogen bonds to an acceptor region between $\alpha 7$ and $\beta 6$, making for a total of 10 $\alpha 7/\alpha 7\beta 6$ interactions. The crystal structure of ligand-free SqvA includes an average of three ordered waters within this interface per monomer pair (fig 4-11, d-e). The lack of any monomeric or pentameric peaks on SEC-MALLS indicates this interaction is strong enough to hold the decamer together even in a fully solvated environment.

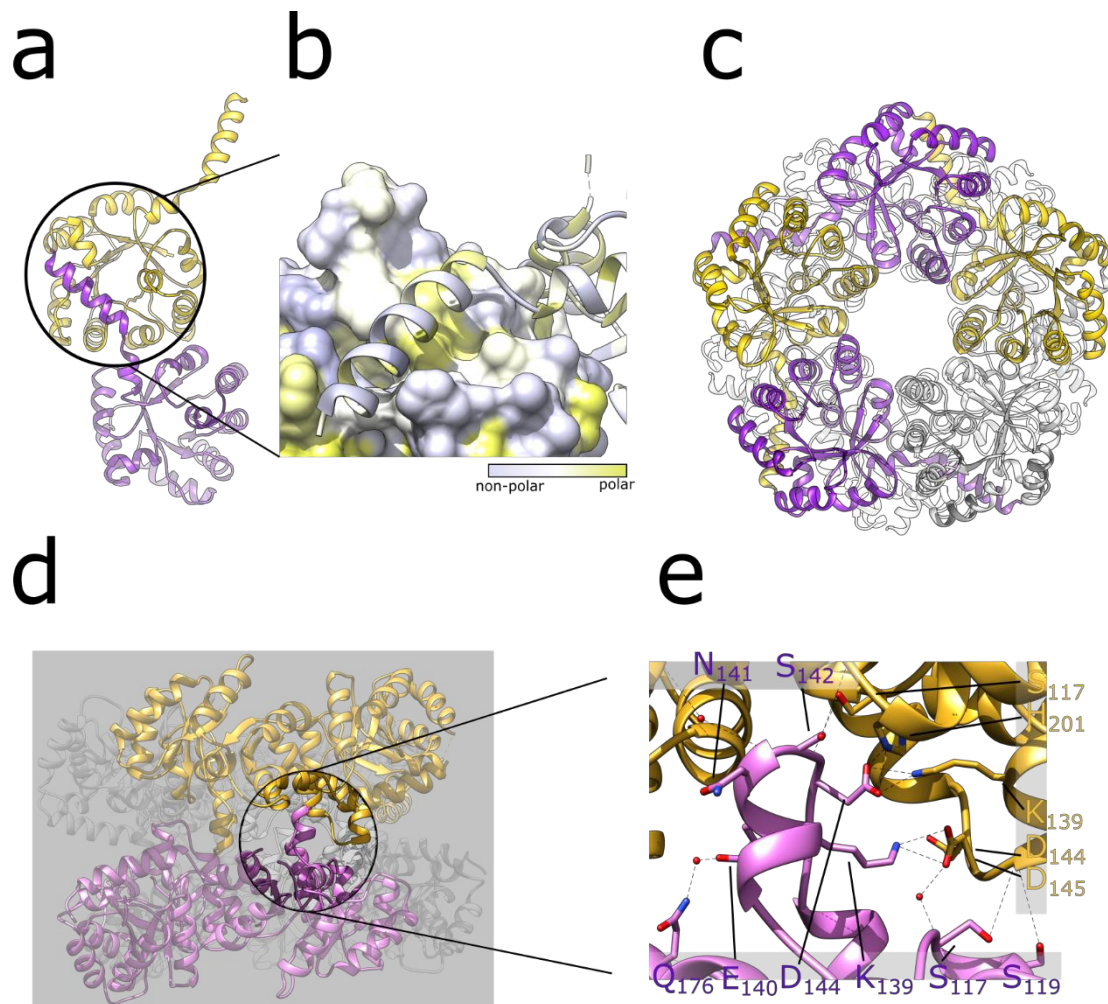


Figure 4-11: Structural basis of SqvA decamer formation. **A.** View of two SqvA monomers in the same ring. **B** Inset heatmap of CTH binding to the cleft. **C.** Assembly of 5 monomers to form one pentameric ring. Second ring is outlined behind. **D.** Assembly of two pentamers. **E.** Detail view of the hydrogen bond network and ordered waters as observed in the XRD structure.

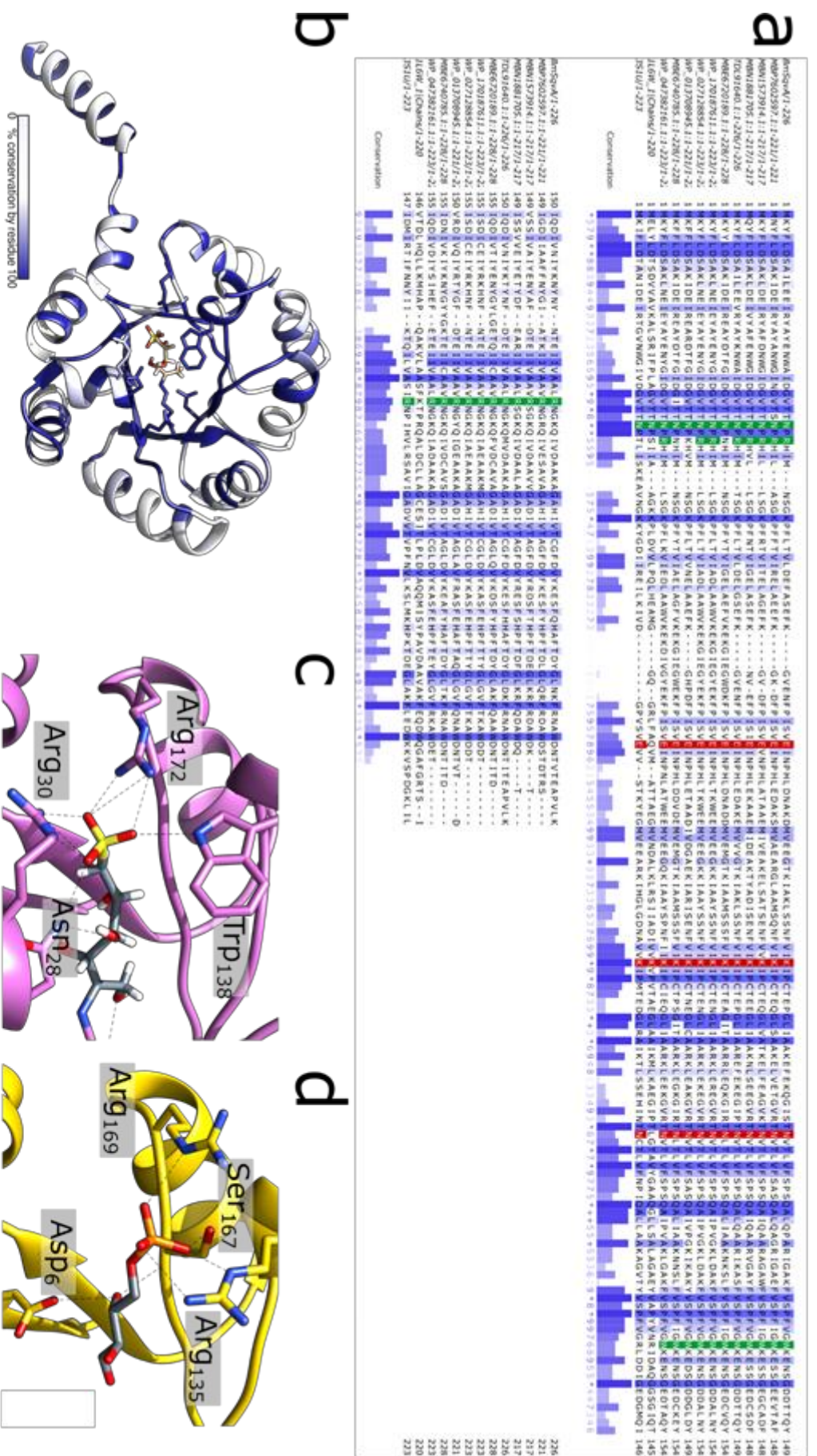


Figure 4-12: Bioinformatic analysis of SqvA active triad and sulfonate binding pocket. A. Multiple sequence alignment of SqvA and the 10 closest matching, non-*Bacillus* TALs, chosen to represent a variety of bacterial genomes. F6p aldolase (PDB 1L6W) and a structurally characterized decameric TAL ECTALB (3HSU) were also included. Conserved active triad residues are highlighted in red, and sulfonate binding residues in green. All other residues are shaded by conservation, with darker blues denoting higher levels. This is also presented in a graph below the alignment. B. Conservation heatmap of SqvA and ECTALB monomers with identical colouring to A. C-D. Ligand complexes formed by SqvA (C, purple) and ECTALB (yellow, D), which bind SF and E4P in their respective structures.

4.4.7. *SqvA informatics*

SqvA has a well-conserved sequence when compared to other TALs of the same subfamily, with most of this conservation focussed on the internal surface of the $\alpha_8\beta_8$ barrel. The active triad are conserved in all TALs assessed. However, the residues involved in sulfonate binding are less well conserved, with only Asn₂₈ and Arg₁₇₂ present in all TALs (fig 4-12, a). Arg₃₀ is only present in 66% of cases, and Trp₁₃₈ is absent in a structurally characterized TAL from *T.acidophilum* (TaTAL). TaTAL can bind F6P and S7P, and a structure exists in a Michaelis complex with E4P, the acceptor substrate (fig 4-12, b,d) ⁷⁷. The protein has never been observed binding a sulfonated species. A comparison of the binding sites of *SqvA* and TaTAL shows the only common residue in sulfonate/phosphate binding is Arg₁₇₂. Asn₂₈ cannot bind the phosphate in E4P due to the extra distance imposed by the presence of the phospho-ester. Trp₁₃₈ and Arg₃₀ both bind the sulfonate and given the flexibility of the Arg side chain, it may be possible for Arg₃₀ to conform and bind a phosphate group, conferring the dual-specificity to the site that must exist to allow observed *SqvA* activity. An alignment between *SqvA* and a decameric F6P aldolase show a general fold alignment as expected. The active site additionally features the circular permutation previously noted as a major difference between TalB and aldolases ¹⁸, as seen in the lack of alignment across the entire active triad (fig 4-12, c-d).

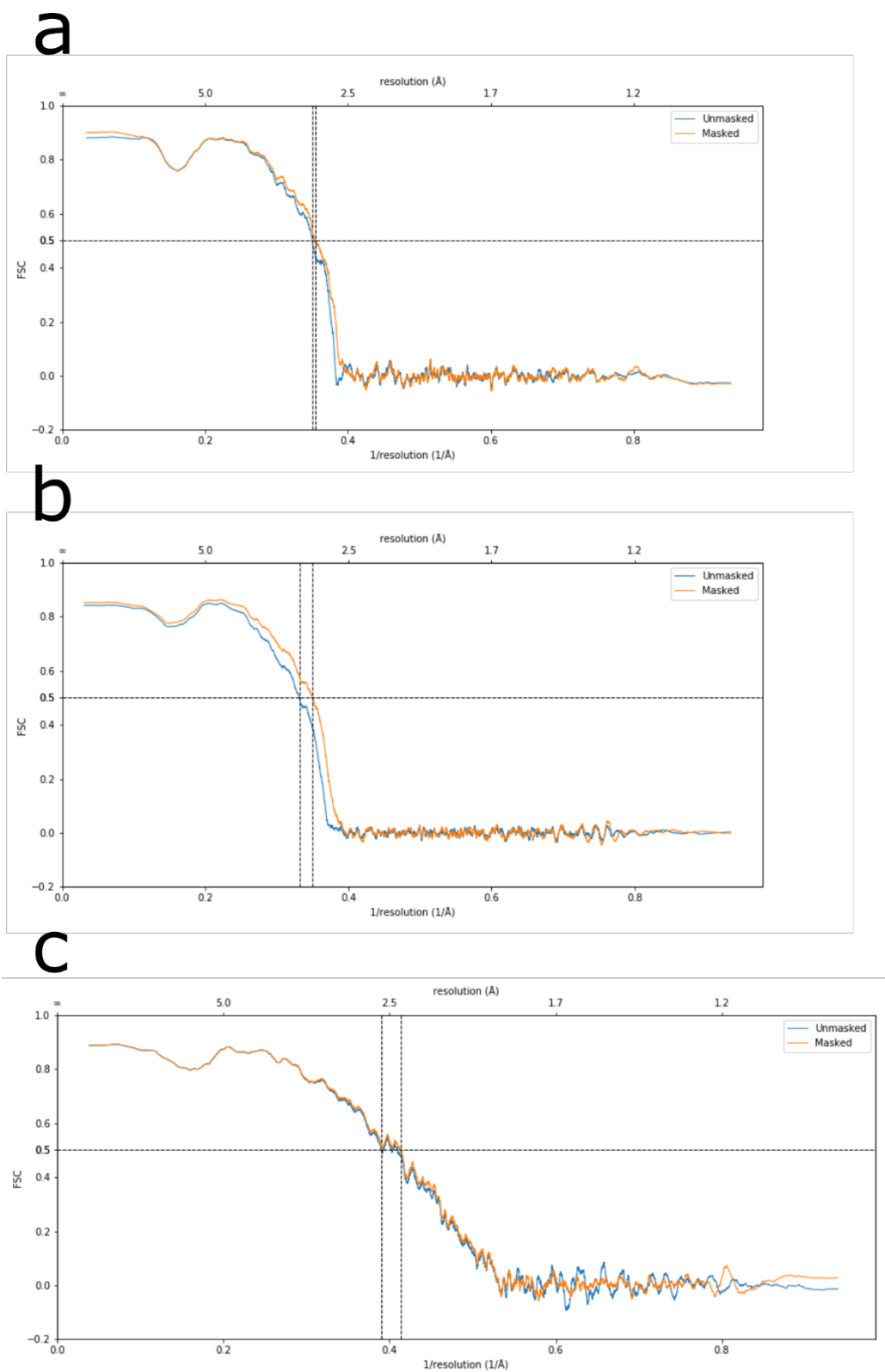


Figure 4-13: FSC curves for EM datasets. A. FSC curve for SqvA ligand-free. **B.** FSC curve for SqvA-SF (D5 symmetry). **C.** FSC curve for SqvA-SF (C1 symmetry). In all cases blue denotes the unmasked map, and orange masked. Dashed lines denote an FCS of 0.5, where resolution was cut.

4.5. Conclusion and future directions

SqvA is a classical TAL with a novel capacity and apparent selectivity for SF, with independently determined activity using SF as a donor and G3P as an acceptor. The protein expresses and purifies in *E.coli* as a solution decamer. X-ray crystallography and cryo-EM produced ligand free structures at 3.1 Å and 2.6 Å resolution respectively, which both show the SqvA complex as a stack of pentamers, with two facing outwards to form the full structure. Pentamer formation is primarily achieved by a hydrophobic interaction between a hydrophobic cleft of one monomer, and the nonpolar face of an amphipathic C-terminal α -helix on the adjacent, which forms a pentamer when repeated 5 times. Ring joining is achieved through a hydrogen bond network between $\alpha 7$ of one monomer and $\alpha 7\beta 6$ of the next.

A cryo-EM structure of SqvA with SF revealed SqvA shares the conserved active triad Asn-Lys-Glu as seen in all classical TALs. Lys₈₉ forms a Schiff base with SF in an extended, open-chain conformation clearly suggesting a classical TAL mechanism. Sulfonate binding is achieved through an Arg-Trp-Arg-Thr pocket. These motifs are well conserved in other classical TALs with the exception of Arg₃₀ which may provide the means for sulfonate recognition.

The work presented in this chapter provides valuable insights into the fold, multimeric state and SF binding action of SqvA, but a great deal more work can be done to fully characterize this protein. Enzyme activity can be measured by assaying for produced F6P using a commercial F6P quantitative assay kit. This can be done with respect to both SF and G3P. These rates when compared to those of other TALs with the same likely catalytic mechanism could shed light into any role SqvA may play in overall substrate concentration throughout the sulfo-TAL pathway. Additionally, arabinose-5-phosphate (A5P) is a known inhibitor of classical TAL mechanisms and while binding was not observed by nanoDSF, an inhibition assay and EM structure could serve to further confirm the mechanism.

A structure with a phosphorylated substrate such as G3P or F6P would be a helpful next step in confirming the sulfonate/phosphate dual-specificity that must exist to support the observed activity of SqvA. Attempts at EM were attempted with both substrates without success, however sufficient binding may be possible through concentration, pH and buffer

contents and incubation time optimisation. The same may allow for an A5P complex structure. Additionally, the sulfo-transketolase SqwGH as seen in *Clostridium* SP. MSTE9 has not yet been structurally characterized. The enzyme fulfils a similar niche in the sulfo-TK pathway as SqvA in sulfo-TAL, and would make for an interesting comparison.

5. Chapter 5: Conclusion s and future perspectives

5.1. Impetus and development of this research

Of the five discrete sulfoglycolysis pathways known currently, only two were published in any capacity at the beginning of this project. The sulfo-EMP and sulfo-ED pathways, which most closely mirror traditional glycolysis were first characterized in detail in 2014 and 2015 respectively ^{40,41}. The sulfo-SMO pathway had been theorized by Spencer Williams and was backed up by a comparative proteomic study as well as some early recombinant expression and purification of the core enzymes identified. From here, structural study was split between me and Dr. Mahima Sharma. These studies, and the collaboration with the labs of Spencer Williams and Ethan Goddard-Borger led to the publication of a comprehensive paper ⁸⁰. However, by this time another group had published a reconstituted pathway assay identifying the sulfo-SMO pathway, though not the structural or in-depth biochemical studies ⁷⁵. The same paper described a variant of sulfo-EMP, labelled sulfo-EMP2, and a sulfo-transketolase-dependent pathway (sulfo-TK) for the first time. Additionally, the sulfo-TAL pathway was described, though this had been described previously in two organisms ^{78,79}.

The discovery of so many novel sulfoglycolysis pathways in such a short time serves to highlight the fast-developing nature of this branch of biochemistry. This fast pace is partly caused by the advent of powerful bioinformatic techniques centred around the conserved sulfonate binding motif in SQases, first described in YihQ ⁴². The discovery of so many new pathways changed the scope of the project greatly across its progression. The best example of this is SqvA, which despite being one of the best characterized proteins described in this thesis had not even been discovered until more than half of the funded time had elapsed.

5.2. SmoF

Based on the data presented in chapter 1, SmoF is a sulfoquinovosyl binding protein (SQBP) which works overall much like the prototypical SBP, MBP. Binding of SQ is achieved through hydrogen bonding to the C2-5 hydroxyl groups, in addition to a sulfonate binding pocket and

ordered water molecule. Based on crystal structures of SQ-bound and ligand-free SmoF, the protein undergoes a 31° interdomain rotation on ligand binding to adopt a closed conformation. This is a similar change to those seen in other SBPs. Based on the high degree of thermal instability seen in ligand-free SmoF crystals SmoF, may undergo the same shifts between an open and semi-closed state when ligand-free as observed by NMR in MBP. SmoF is capable of binding SQ derivatives and C1 glycosides including a C1 methylated and octylated derivative, sulfoquinovosyl glycerol and a form of sulfoquinovosyl diacylglycerol (SQDG-(C16:0/C4:0)). In all cases SmoF undergoes closure on ligand binding with dynamic motions of helix α 1 and the loop containing residues Gln₁₂ and His₁₃. This closure is complete in all cases except with the C1 octylation and SQDG, wherein the ligand displaces enough of helix α 1 to prevent full ligand enclosure. Overall, these data provide a valuable insight into the first stage of the sulfo-SMO pathway. The capacity of SmoF to bind and accommodate larger SQ derivatives presents an opportunity to design bespoke functional components including SQ-decorated resin, which could be used to form an affinity purification technique after the fashion of the MBP column.

5.3. SmoB

SmoB is a trimeric aldo-keto reductase, which catalyses the final reaction of the sulfo-SMO pathway. That is, the reduction of 6-oxo-glucose to form glucose in an NADPH-dependent manner. This status was confirmed through a series of crystal structures of SmoB in the ligand-free state, in a binary complex with NADPH and a ternary complex with NADPH and glucose. In all cases NADPH is bound in an extended, *anti* conformation along the top of the TIM barrel making up each monomer. A C5 phosphate binding pocket can be found providing the structural basis of NADPH specificity. Isothermal titration calorimetry confirmed SmoB cannot bind NADH, and unlike dual-specific AKRs such as xylose reductase the phosphate pocket is not flexible enough to accommodate NADH. Glucose occupies a site adjacent to the nicotinamide moiety of NADPH, with recognition of the C2-4 hydroxyls by hydrogen bonding forming the basis of substrate binding. NADPH and glucose are positioned to form hydrogen bonds with the active tetrad of Tyr₇₆-Lys₁₀₄-Asp₇₁-His₁₅₁ allowing hydride transfer in a mechanism shared between many AKRs. Steady-state kinetics of the reverse reaction (glucose + NADP⁺ → NADPH + 6-oxo-glucose) provide further evidence that SmoB is active exclusively

with NADP(H). These studies show SmoB as an archetypal AKR. Engineering of the phosphate binding pocket could in future allow for glucose-acting enzymes for use in industrial applications, easing the complexity inherent in sugar organic chemistry.

5.4. SqvA

Transaldolases are a common enzyme family responsible for the transfer of dihydroxyacetone between a donor and an acceptor substrate. Here, the first known sulfo-transaldolase SqvA has been recombinantly expressed and purified, then structurally characterized both in a ligand-free and sulfofructose-bound state by cryo-electron microscopy. These studies confirm a decameric, two-layer ring complex, which is present in solution. The basis for this complex formation lies in a hydrophobic interaction between a C-terminal amphipathic α -helix and an acceptor groove across the $\alpha_8\beta_8$ barrel comprising each monomer. Repeated 5 times, this produces a ring. Two rings associate with a 36° rotation, each ring facing outwards using a network of polar interactions, including a series of ordered water molecules. SqvA binds sulfofructose in an open-chain, extended conformation and is covalently bound to a conserved catalytic Lys, forming a Schiff base as observed in other TALS. SqvA contains the structurally conserved active triad of Asp-Lys-Glu featured in all TALS. The mechanism these enzymes employ is reversible, which combined with the proposed products, fructose-6-phosphate and sulfolactaldehyde implies that SqvA must accommodate sulfonate and phosphate groups. Though no structure with a phosphorylated product was obtained, a bioinformatic analysis suggests a flexible Arg in the sulfonate binding pocket may undergo a conformation change allowing for this dual specificity. These data provide a foundation for the function of SqvA, which would be strengthened by assessing the kinetics of fructose-6-phosphate formation. Additionally, a structure with a phosphorylated substrate such as fructose-6-phosphate and subsequent characterization of the phosphate binding pocket may elucidate the true nature of sulfonate-phosphate dual-specificity.

5.5. The contribution of AlphaFold to structural biology

The ability to rapidly and (often) accurately predict the fold of any given protein with AlphaFold (AF) has undoubtedly changed the face of structural biology. It has also led some to claim the original algorithm and its derivatives essentially make traditional structural characterization obsolete. However, I believe this is a flawed assessment for several reasons, the most important being that AF does not simulate folding in a choice solvent, as would be done in atomistic molecular dynamics. The algorithm rather uses the PDB database to correlate between the sequence of a motif and a likely fold, then merges these together to form a prediction. On its own this would not produce a chemically competent structure. To remedy this the backbone is assembled in order to prevent unacceptable bond lengths. Additionally, the entire structure is relaxed using a series of AMBER force fields which typically results in reasonable bond geometry. The approach taken by AF also means the algorithm may not always correctly account for the function of a secondary structural element, for example a transmembrane α -helix, which would form a very different fold in an *in vivo* membrane environment to a polar solvent. At the time of writing this conclusion, AF does not yet have the capacity to fully predict ligand binding sites of all types, though prediction models do exist for common cofactors such as NADPH. The same applies to metals as the electrochemistry determining the exact metal coordination in a protein, and often by extension its activity is quite complex and not easy to predict using the existing models.

It is my belief that the best uses of AF are twofold. Firstly, as a preliminary step to inform future experiments. Examples include prediction of a cofactor binding site location relative to a peptide chain terminus to inform whether an affinity tag added there will become occluded and non-functional. Had AF existed when SmoB was first expressed, this may have led to the use of a C-terminal 6xHis tag before initial characterization in the pET29a construct. Another example could be to determine flexible side chains on a protein surface to mutate when designing a surface entropy reduction mutant. Another use is where there are real data to assess the validity of the produced model. An obvious example is one I employed to produce a model for SqvA after producing the first refined EM map. The AF model and the final model from that map share a global RMSD of 0.54 Å, indicating few large-scale differences between the two and highlighting the potential of AF. It is worth noting that TALs share a very common core fold in the TIM barrel, for which a large variety of structures exist in the PDB, so this case may not be applicable for proteins with more unique folds. A similar use would be the use of

an AF model for molecular replacement, which I have used to solve structures of both At3280ct and SlaB. While these structures are of limited use due to their poor resolution, the ability to produce an advanced MR search model with a 100% sequence identity is highly desirable. There is a question of bias in the resulting density map from the model, but in my opinion, this can be a worse issue in the traditional use of a single PDB. Given AF models are always based on the consensus of a database of PDB structures, the use of an AF model can be considered similar to using an ensemble of PDBs as a single model, as is possible with Phaser. As has always been the case with MR, in my opinion so long as the quality of the model compared to the original data (which can be quantified using statistics such as an R_{free}) is used this is an acceptable use. This use has the side-effect of making many forms of experimental phasing mostly obsolete in many cases, including the selenomethionine labelling I used with SqvA and At3280ct. I predict these methods will still see use where an AF prediction cannot produce a viable structure, such as rare natural folds or synthetic designs without equivalents in the PDB.

5.6. Future perspectives

The five archetypal sulfoglycolytic pathways so far characterized contain a large variety of proteins, of which only a minority have been described in detail. Of notable interest are the transcriptional regulators in the sulfo-ED, EMP and SMO pathways. Of these only the GntR-like transcription factor Atu3280 has a structure (appendix 1), though no functional characterization to determine whether it is responsible for the SQ-inducibility observed in the pathway. Additionally, there is no data to suggest the effector specificity. CsqR is a DeoR-like transcription factor in the sulfo-EMP pathway that lacks structural characterization but has been confirmed through gel-shift assay to regulate the pathway and use SQ and sulforhamnose (SR) as an effector. A comparative study of these transcription factors could provide valuable insight into the varied mechanisms for sulfoglycolytic pathway regulation. Additionally, if the effector binding motifs are consistent between these two classes of enzymes they could provide a new means of pathway identification, similar to how the RWY motif seen in SQases is used.

Another target is the putative sulfo-transketolase SqwGH. Aside from a reconstituted pathway assay confirming activity this enzyme remains entirely uncharacterized. TKs have a very different fold to TALS despite the overall similar action. TKs also use a cofactor, thiamine diphosphate (TPP) unlike TALS. They are typically homodimeric and have a “half-site” completed through dimerization to form an active complex. In initial studies, SqwGH were co-expressed and purified. Assuming an active SqwGH complex maintains the typical monomeric Mw of a TK, ~70 kDa, SqwG and H are an N and C-terminal domain respectively of the SqwGH complex since the two combined have an Mw of ~75 kDa. Obtaining a structure of SqwGH could provide valuable insight into whether the enzyme is naturally heterodimeric, or whether it was split in this fashion for ease of study. The apparent addition of a ribosome binding site between SqwG and H in the study suggests the latter ⁷⁵.

5.7. Impact on the wider field

The scope of these studies has universally been based in molecular biology, biochemistry and structural biology, and while the larger-scale implications of these data have not been ignored, they have never been the focus. With the foundation of knowledge now collected however, this can begin to change.

The same principles used to bioinformatically search for new sulfoglycolysis pathways can be applied to a more general search for organisms capable of sulfoglycolysis. Recognition of a highly specific motif and enzymatic mechanism through activity-based probes (ABPs) has been an increasingly valuable tool in identifying organisms capable of specific glycosidase activities ¹⁹. The same technique can be applied to SQases, and these probes are in development. The last decade has proven that sulfoglycolysis is a much more common and diverse process than first thought, so it is my prediction that assaying for it in wild organisms will uncover many more examples of the five archetypal pathways, and perhaps find novel ones.

A thematically similar system can be employed with SmoF. MBP has been a frequently used tool in protein purification due to its high affinity for maltose, and tolerance to substrate modifications, allowing for technologies such as MBP affinity columns. SmoF shows many of

these qualities so an SQ-decorated resin could be used as a system for affinity chromatography. This may be hampered by the lack of SmoF affinity for sulfonated buffers though elution with SQ and resin-labelling with a lower-affinity derivative could still allow for efficient elution. This principle could also be used to form a pull-down assay wherein soluble extract from bacteria grown on SQ media (to encourage pathway upregulation as well as testing the effectiveness of any sulfoglycolytic pathway present) would be exposed to the resin. A wash and elution step could then purify any SQBPs present for further analysis. While this system could provide a useful tool for binding novel SQBPs, I do not see it as the primary use of an SQ column, which would be in my opinion as a specific form of bulk affinity chromatography for recombinantly expressed, SQBP-labelled fusion proteins.

5.8. Closing thoughts

It is the nature of science that no study is funded or worked by the same group of people for long enough to answer every question raised. The data presented in this thesis provide a foundation for three important enzymes across the sulfo-SMO and sulfo-TAL pathways, with each providing novel insights either to sulfoglycolysis and sulfosugar recognition, the class of enzyme it belongs to or both. However, there is much work to be done before the questions raised over these studies are answered to a satisfactory level. In my view, the priority for this facet of the sulfoglycolysis project is to publish data on SqvA, as the combination of a novel substrate specificity for a classic enzyme, high-resolution cryo-EM solutions and central role in the sulfo-TAL pathway makes it a prime target for many competing groups. Further study in the substrate specificity and activity would be highly desirable but shouldn't be taken as barriers to publication. The development of an "SQBP column" for SQ-based affinity chromatography is another high-priority target due to the potential for real-world applications, which though niche, are in my opinion a potentially lasting mid-term innovation resulting from this project.

It has been my privilege and pleasure to work on this project, and I look forward to seeing the developments made in the future.

6. Chapter 6: Appendices

6.1. Appendix abstract

While SmoF, SmoB and SqvA represent the three largest and most successful bodies of work in this document, the characterization of other proteins was attempted under the same project. These appendices present the limited success obtained through work on two proteins: Atu3280 and SlaB. The GntR-like transcription factor Atu3280 is thought to regulate the sulfo-SMO pathway in the SQ-mediated manner observed, using an N-terminal winged helix-turn-helix for DNA binding, and a C-terminal effector binding domain featuring a jelly fold. Here, a C-terminal truncation mutant is expressed and purified, with an ongoing effort into crystallisation yielding low quality crystals in the unlikely spacegroup F 2 3. The aldehyde dehydrogenase SlaB is also presented. This protein represents the final stage of the sulfo-TAL pathway in *Bacillus megaterium* – oxidation of sulfolactaldehyde to sulfolactone in an NADH-dependent manner. This protein is also expressed and purified, with early attempts at crystallography yielding low resolution crystals in P 1, for which limited success is obtained in phasing and model building, confirming the fold described for other proteins in the family if little else. Overall, these data provide a simple foundation on which more meaningful data may be obtained with time and persistence.

6.2. Materials and Methods

6.2.1. Gene expression and protein purification

6.2.1.1. *Atu3280ct*

The gene for *Atu3280-ct* was codon optimised and ordered from Eurofins Genomics in a standard pET28a vector, which was transformed into BL21(DE3) competent *E.coli* (NEB) using the standard NEB protocol. This was plated onto LB-agar with 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin and grown at 37 °C overnight. Isolated colonies were picked and used to form liquid precultures in 20ml LB with 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin, then grown at 37 °C, 180rpm overnight. 5ml of preculture was used to inoculate every 1L of expression media (LB + 35 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin). Expression cultures were grown at 37 °C, 225 rpm until an OD_{600} of 0.6-0.8 was achieved. Then, isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final 1mM concentration and induction took place over 18 hours, 18°C, shaken at 180rpm. Cells were harvested by centrifugation at 4500 rpm, 20 minutes, 4 °C. The supernatant was discarded, and the pellet resuspended in 40 ml binding buffer (50 mM tris. 300 mM NaCl, 30 mM imidazole, pH 7.5) per 1L culture. 1 tablet of EDTA-free protease inhibitor was also added per 1L. Cells were lysed either by sonication (5 pulses of 100% amplitude, 60s on, 60s off while on ice) or disruptor (25 kPSI, 4 °C, eluted into an ice-cooled flask). Cell debris was then pelleted by centrifugation at 18,000 rpm, 40 minutes, 4 °C. The pellet was discarded, and the soluble fraction was loaded to a 5ml Ni-NTA Crude FF column at 5 $\text{ml}\cdot\text{min}^{-1}$. The column was pre-equilibrated with binding buffer. A linear gradient into elution buffer (50 mM tris, 300 mM NaCl, 500 mM imidazole, pH 7.5) was used to elute his-tagged protein into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. Protein-containing fractions were then pooled and buffer-exchanged into binding buffer due to high imidazole sensitivity of SqvA. The pooled fractions were then concentrated to a 2ml final volume and loaded to a HiLoad superdex 16/600 S75 column for size-exclusion chromatography. The column was pre-equilibrated into SEC buffer (50 mM tris, 300 mM NaCl, pH 7.5). Elution took place using SEC buffer at 1 $\text{ml}\cdot\text{min}^{-1}$ into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. SqvA-containing fractions were pooled and concentrated to a working concentration of 38 $\text{mg}\cdot\text{ml}^{-1}$, confirmed by A_{280} , then snap-frozen with liquid nitrogen. These were subsequently stored at -70 °C.

The optimised tag-cleavage purification was identical to the original with key differences: Pooled IMAC purification fractions were concentrated to an approximate 2 mg.ml⁻¹ concentration as confirmed by A₂₈₀. Thrombin cleavage was performed over 6 hours at 30 °C, 100rpm using a 1/2000 dilution of commercial Thrombin. Uncleaved protein was then removed by IMAC using a 5 ml Ni-NTA purification, with the flow-through pooled and prepared for SEC. At 36 kDa, thrombin can be fully separated from Atu3280-ct on a 16/600 S75 column.

6.2.1.2. *SlaB*

The gene for SlaB was codon optimised for *E.coli* expression and ordered from Eurofins Genomics in a standard pET28a vector, and expression of the pET-YSBLiC3C construct was handled in an identical manner. In both cases expression began with transformation into BL21(DE3) competent *E.coli* (NEB) using the standard NEB protocol, onto LB-agar with 35 µg.ml⁻¹ kanamycin, and grown at 37 °C overnight. Isolated colonies were picked and used to form liquid precultures in 20ml LB with 35 µg.ml⁻¹ kanamycin, then grown at 37 °C, 180rpm overnight. 5ml of preculture was used to inoculate every 1L of expression media (LB + 35 µg.ml⁻¹ kanamycin). Expression cultures were grown at 37 °C, 225 rpm until an OD₆₀₀ of 0.6-0.8 was achieved. Then, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final 1mM concentration and induction took place over 18 hours, 18°C, shaken at 180rpm. Cells were harvested by centrifugation at 4500 rpm, 20 minutes, 4 °C. The supernatant was discarded, and the pellet resuspended in 40 ml binding buffer (50 mM NaPi, 300 mM NaCl, 30 mM imidazole, pH 7.4) per 1L culture. 1 tablet of EDTA-free protease inhibitor was also added per 1L. Cells were lysed either by sonication (5 pulses of 100% amplitude, 60s on, 60s off while on ice) or disruptor (25 kPSI, 4 °C, eluted into an ice-cooled flask). Cell debris was then pelleted by centrifugation at 18,000 rpm, 40 minutes, 4 °C. The pellet was discarded, and the soluble fraction was loaded to a 5ml Ni-NTA Crude FF column at 2 ml.min⁻¹. The column was pre-equilibrated with binding buffer. A linear gradient into elution buffer (50 mM NaPi, 300 mM NaCl, 500 mM imidazole, pH 7.4) was used to elute his-tagged protein into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. Protein-containing fractions were then pooled and concentrated to a 2ml final volume and loaded to a HiLoad superdex 16/600 S200 column for size-exclusion chromatography. The column was pre-equilibrated into SEC buffer (50 mM NaPi, 300 mM NaCl, pH 7.4). Elution took place using SEC buffer at 1

ml.min⁻¹ into a 96-well deep block, and protein purity and size were assessed by SDS-PAGE. SlaB-containing fractions were pooled and concentrated to a working concentration of 30 mg.ml⁻¹, confirmed by A₂₈₀, then snap-frozen with liquid nitrogen. These were subsequently stored at -70 °C.

All SDS-PAGEs for all proteins were performed using 12% acrylamide gels and run for 45 minutes at 200 V with 0.1% SDS TRIS/Glycine running buffer. Gels were subsequently stained with Coomassie magic blue dye. Unless otherwise indicated, 7 µl of sample was loaded, with 5 µl loading dye and 3 µl water for a 15 µl total load.

6.2.2. Molecular cloning of BmSlaB

All subcloning of SlaB was performed using In-fusion cloning (Takara). SlaB fragments were produced by PCR using KOD hot-start DNA polymerase with a standard protocol (NEB). 1 µl DMSO and 5 µl CG enhancer (NEB) were also added to improve primer specificity and amplification yield. Plasmid fragments were amplified from a sequenced stock using the same method. All gene product purification took place using a monarch gel digest kit using the standard protocol. All ligation and subsequent transformation to STELLAR cells took place using the standard protocol. Transformed colonies were assayed for ligated plasmid through colony PCR: A half colony was suspended in 5 µl H₂O and lysed at 95 °C, 5 minutes. Cell debris was pelleted at 13,000 rpm for 5 minutes, and 1 µl supernatant was loaded to an identical PCR protocol as used for gene amplification. Colonies containing the gene of interest were sequenced by Eurofins Genomics. All DNA products were visualised using a 2% agarose/TBE gel with 0.1% CYBRsafe. Samples were run at 120V, 50 minutes and visualised by UV lightbox.

6.2.3. ESI peptide-ID mass spectroscopy

Purity, mass, and labelling extent estimates were obtained using electrospray-ionisation mass-spectroscopy (ESI-MS) with a Bruker Maxis-HD qTOF mass spectrometer. All samples were prepared to 1 µM with 50% by volume acetonitrile, 1% by volume formic acid. Infusion was at 3 µl.min⁻¹. Spectra were averaged over 1 minute and smoothed in 1 cycle using a

Gaussian method, then the baseline was subtracted with a flatness of 0.8. Compass 1.7 software was used for this.

6.2.4. X-ray crystallography and structure solution

Ligand-free crystals of At3280ct were grown by hanging drop at 20 °C, with protein at 20mg.ml⁻¹ in 50 mM TRIS, 300 mM NaCl pH 7.5. Mother liquor contained 0.1 M HEPES pH 7.0, 15% Tacsimate (Hampton Research) pH 7.0, 0.01 M Urea, 1% PEG 3350. Diffraction quality crystals were developed after lowering protein concentration and streak seeding from a 1/1000 dilution seed stock, made from crystals at 30 mg.ml⁻¹ of the same mother liquor condition. Final drops were formed from 1 µl protein, 2 µl mother liquor. The resultant crystals were cryoprotected in a separate drop of 10% ethylene glycol in mother liquor for 30 seconds before flash cooling. Seed stocks were prepared on ice using the standard SeedBead protocol SQ cocrystals of At3280ct were grown by sitting drop containing 150 nl At3280ct at 30 mg.ml⁻¹, incubated for 10 minutes at room temperature with 5 mM SQ prior to crystallisation, and 150 nl mother liquor containing 0.8 M succinic acid at pH 7.0. Crystallisation occurred over 8 weeks at 20 °C. Resultant crystals were cryoprotected in a separate drop of mother liquor containing 10% glycerol over 30 seconds before flash cooling.

All crystals were tested using the in-house Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV. X-ray data collection took place at Diamond Light Source on beamlines I-03 and I-04. Data collection statistics are available in the results section. Data indexing and initial processing was performed using either DIALS or 3dii pipelines from the Xia2 package^{120,121}. Data reduction was performed with AIMLESS, and resolution was cut until $CC1/2 = 0.5$ ¹²². Molecular replacement was performed using MOLREP, using a monomer of the ligand-free EM structure of SqvA as a search model¹²³. Automated model building was performed using BUCCANEER¹²⁵. Diagram preparation for molecular models was performed using CCP4MG, Chimera or Pymol, depending on the desired outcome^{129–131}.

6.3. Transcription Factor Atu3280

6.3.1. Introduction

Transcription factors are a ubiquitous class of DNA-binding protein, typically responsible for the regulation of transcription of anything between single gene and entire operon. In this capacity they become major players in almost all biochemical process regulation, including carbon catabolite repression¹⁹. It is therefore unsurprising that an estimated 10% of all genes in free living bacteria such as *E.coli* K-12 encode TFs¹⁹⁹. The GntR family of transcription factors is named for the *B. subtilis* gluconate operon repressor²⁰⁰ and made up of about 270 members as of 2002²⁰¹.

GntRs are divided into an N-terminal helix-turn-helix (HTH) DNA binding domain, and a C-terminal effector binding domain (EBD) (fig 6-1, a-c). Of these, the HTH is far better conserved structurally, and is typically of the winged-HTH (wHTH) variety, which features 3 α -helices and two β -sheets, the former forming a DNA-binding tri-helical core (fig 6-1, c). These are labelled α D1,2 and 3, and β D1 and 2 respectively. This differs from a traditional HTH with the addition

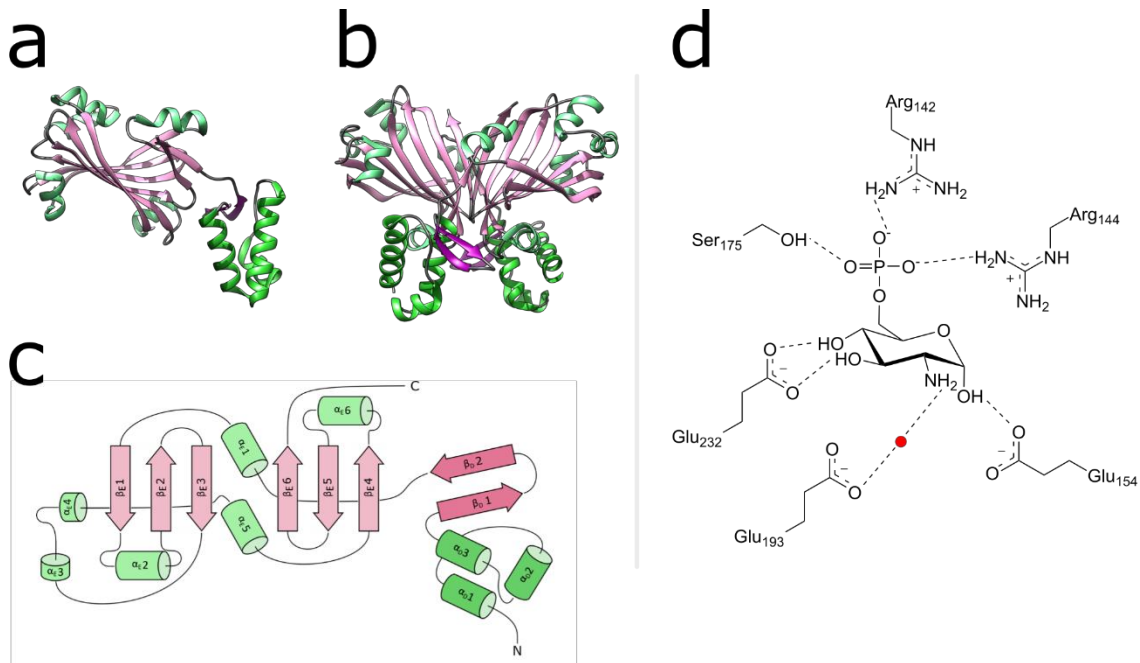


Fig 6-1: Structure of a representative GntR-type TF and effector recognition mechanism. A. Monomeric DasR. **B.** Active DasR dimer. **C.** Secondary structural schematic of DasR. The wHTH is coloured darker. A-C are coloured by secondary structure; α -helices in green, β sheets in purple. **D.** Effector recognition mode of DasR for GlcNAc.

of a C-terminal β -hairpin wing near the core. The core and wing then bind the major and minor groove of the target DNA respectively. wHTH domains can further vary with the addition of a

C-terminal α -helix, as found in drug-efflux pump TF MarR²⁰, or an additional β -strand between α D1 and α D2 as in ArsR, used in pH response²⁰³. In cases such as methionine aminopeptidase, β D1 forms the N-terminus instead of appearing after α D3²⁰⁴.

The DNA binding mode of these wHTH domains is shared among most GntRs, and similar to that of traditional HTH domains. A major helix, typically α D3 or the equivalent provides the largest contribution of DNA binding and recognition by binding the major groove. The majority of these interactions are from polar residues, and can be mediated by water²⁰. The wing of β D1 and 2 in wHTH domains allows a secondary binding to the minor groove, typically using charged residues along the strands. Many wHTH-containing proteins also bind DNA as a dimer, with each wHTH binding one half of an operator site consisting of a pair of inverse repeats²⁰⁶. The EBD is far more variable in structure, binding specificity and mode, and function. This variety is prevalent to the point that GntR naming convention notes the type of EBD as a subfamily. The most common is the GntR/FadR, making up about 40% of all proteins in the family. This is followed by GntR/HutC at 30%²⁰¹.

FadR is the archetypal GntR of the GntR/FasR subfamily, and is used as a regulator of fatty acid metabolism in *E.coli*. FadR does not bind DNA in the typical fashion for a wHTH despite the overall similarities. The wHTH domain undergoes no significant conformational change on DNA binding, and the protein is capable of dimerizing in the absence of the promoter sequence. As with other GntRs, the operator sequence is palindromic, though with a GC (1/1') base pair between the sequences. The binding pattern is symmetrical around this base pair

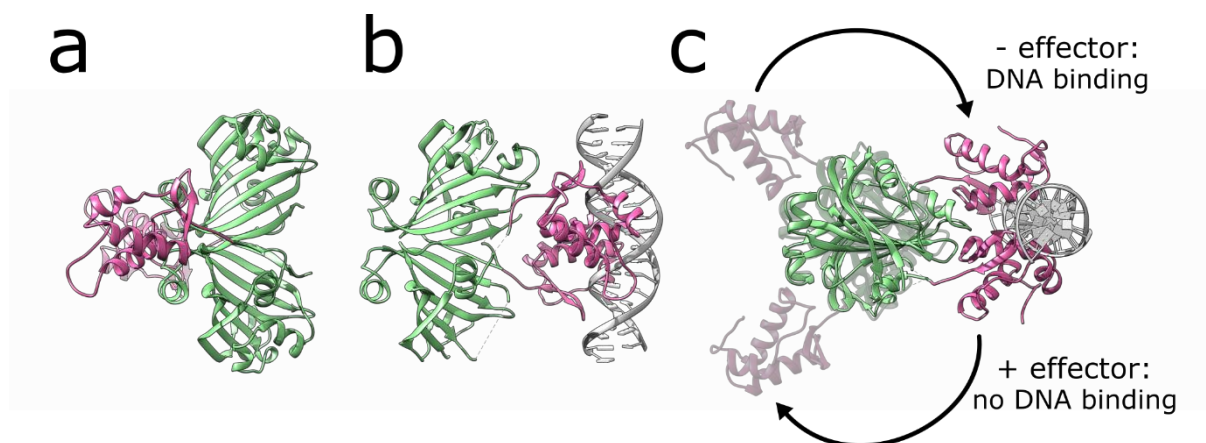


Fig 6-2: Conformation changes allowing DNA binding in DasR. A. DasR in effector-bound, DNA-unbound state. **B.** DasR in effector-free, DNA-bound state. **C.** Top view superposition of A,B (a greyed out) showing rotation of the wHTH domains to facilitate DNA binding.

as each domain forms two major interaction sites at identical base positions relative to the central 1/1' base-pair. These sites correspond to α D3 interactions with the major groove and interactions between the wing region and minor groove. The DNA maintains a B-conformation though with a 20° bend around the 1/1' position (fig 6-2, a-c). Operator recognition arises from the 6/16' (T/A), 7/15' (G/C) and 8/14' (G/C) base pairs, which form hydrogen bonds to the conserved wing residue His₆₅, and α D3 residues Arg₃₅ and Arg₄₅ of each wHTH. Other interactions appear to be with the sugar-phosphate backbone, and are nonspecific²⁰. As with other GntR/FadR proteins, the EBD is made up of 7 α -helices. Helices 1-4,5-7 form an antiparallel barrel, and 6 forms a lid near the N-terminus, in an arrangement similar to the tetracycline resistance operon regulator family TetR, which contains an HTH rather than wHTH DNA-binding domain. These proteins also dimerise forming a bipartite complex with DNA²⁰⁸.

DasR is a GntR from *S. coelicolor* which has the official subfamily notation GntR/HutC, after the histidine utilisation operon regulator HutC, a major player in bacterial histidine catalysis and archetypal member of the GntR family²⁰⁹⁻²¹. The EBD of DasR has a fold reminiscent of HutC with a clear antiparallel β -sheet, flanked by α -helices. The domain binds 6' phosphorylated glucosamines including glucosamine-6-phosphate (GlcN-6-P) and N-acetylglucosamine (GlcNAc) between α E1 and α E5, which upon effector binding move up 4-5 Å and down 1 Å respectively. This decreases the solvent accessibility of the binding site to the solvent enough to almost totally enclose the effector. DasR in a effector-bound conformation holds the wHTH domains in an "upwards" orientation unable to bind DNA, though structures in effector-free conformation appear to do the same. A comparison between DasR and NagR, another GntR/HutC with a 38.8% sequence identity and a crystal structure in complex with DNA, reveals a "downwards" orientation with the wHTH sitting below the EBD, well away from α E2 and α E6 (Figure 4 B.). Interestingly the structure lacks a effector, indicating that the "downwards" conformation may arise in the presence of the operator sequence. NagR also shares the method of 6'-phosphorylated glucosamine recognition seen in DasR. Overall DasR (and other GntR/HutC TFs) may have a native "upwards" conformation, which moves to a DNA-binding "downwards" conformation in the presence of the operator sequence, but is

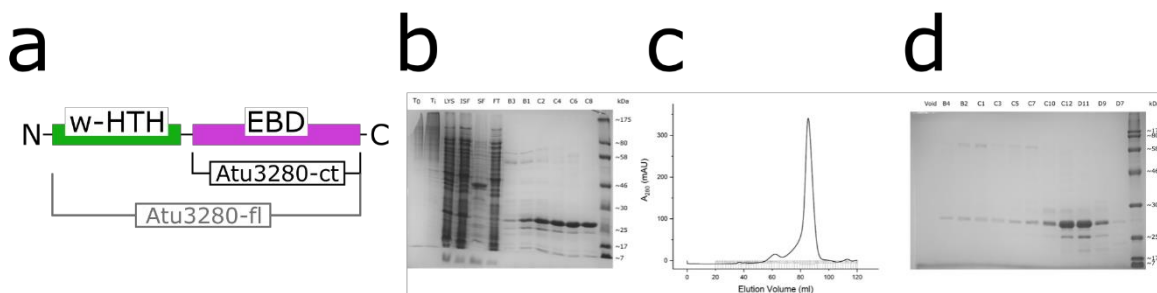


Fig 6-3: Construct design, expression, and purification of Atu3280-ct. **A.** Construct design schematic highlighting the difference between Atu3280-ct and Atu3280-fl. **B.** SDS-PAGE of expression, protein extraction and IMAC purification for Atu3280-ct. From left to right; colony at inoculation (T_0), at time of induction (T_1), cell lysate (LYS), insoluble fraction (ISF), soluble fraction (SF), IMAC flow-through (FT) and elution fractions (B3-C8), ladder. **C.** A_{280} chromatogram of Atu3280-ct polishing by SEC. **D.** SDS-PAGE of Atu3280-ct polishing. Left to right; void volume (Void), elution fractions (B4-D7), ladder.

locked "upwards" by effector binding. This would prevent DNA binding in a effector-dependent manner fulfilling the proteins purpose as a repressor²⁰⁹.

Here, the uncharacterized GntR transcription factor Atu3280 is expressed as a C-terminal EBD truncation, with ongoing efforts into obtaining a crystal structure outlined.

6.3.2. Results and discussion

Atu3280 is composed of an N-terminal wHTH (residues 1-60) and a C-terminal EBD (61- 234). Full-length GntRs have a propensity to form multimers in the absence of their operator sequences, which could complicate purification and crystallography. To simplify expression and determination of EBD - ligand selectivity the EBD was expressed as a truncated mutant called Atu3280-ct (fig 6-3, a). Expression of Atu3280 in *E.coli* BL21(DE3) and purification by Ni-NTA affinity through use of an N-terminal 6xHis tag, followed by size-exclusion chromatography yielded on average 7 mg per 1 L LB protein, with an estimated purity of 90-95% based on SDS-PAGE band intensity (fig 6-3, b-d.). All attempts to produce the full-length Atu3280 (Atu3280-fl) failed to yield any protein.

Initial screening resulted in cubic crystals too small to harvest, however crystals large enough were obtained after an additive screen revealed additional growth with 0.01 M urea (Figure 10). Increasing drop size to 1 μ l allowed for fishing but not diffraction. Streak seeding using previously collected non-diffracting crystals was then attempted to improve both size and

order of crystals produced in the optimised condition. The resultant crystals were large enough to harvest and ordered enough to diffract in-house, however the resolution was poor (fig 3-4, a-d). The low precipitant concentrations found in all optimised mother liquors was indicative of a high crystal water content, and a possible source of crystallographic disorder. Therefore, attempts were made to dehydrate the crystals prior to harvesting through soaking in drops of increasing PEG Mw and content. These typically caused dissolution of the crystals and the incidence of phase separation, indicating the crystals are highly sensitive to osmotic shock. Crystals soaked in 10% ethylene glycol did retain cohesion. While this typically did not increase resolution, one dataset returned from Diamond was high enough (3.4 Å) to process (fig 3-4, e). This was indexed into spacegroup F 2 3, with unit cell dimensions of 375 x 375 x 375 Å. Based on Mw of Atu3280-ct this suggested a high copy count of between 20 and 50, with 24 being statistically most likely. Most structures solved in F 2 3 are viral capsid components with icosahedral symmetry, and crystal structures deposited to the PDB in the spacegroup had similar folds or Mw. Initially this was thought to be the result of an impurity being crystallised. However, a representative crystal was harvested, dissolved, and confirmed to be Atu3280-ct by peptide-ID MS. The same sample run on an SDS-PAGE showed a band of the correct Mw to be intact protein. No MR solution was possible with this dataset due to the high search volume involved, even with an AF model. To remedy this crystallisation in different spacegroups was encouraged. Cleavage of the 6xHis tag did not meaningfully change the crystallisation behaviour. Modification of Atu3280-ct crystal contact formation was attempted through the addition of Anderson-Evans polyoxotungstenate ($\text{TeW}_6\text{O}_{24}$, TEW). TEW promotes crystal contacts by binding surface Glu side chains, allowing multimer formation and leading to nucleation²¹. The additive contains a tellurium atom at the core providing an anomalous signal, which can be used for experimental phasing. 2 mM TEW addition did produce new Atu3280-ct crystals though none diffracted.

SDS-PAGEs run after on sample stored after purification often contained low Mw bands. A sample kept at 4 °C for 20 days showed an almost total degradation (>10% purity of intact protein) compared to protein stored at -80 °C for the same length of time (<90%). These were all analysed by peptide ID MS and considered very likely to be degraded Atu3280ct. A single high Mw band was also identified as *E.coli* chaperone DnaK.

6.3.3. Conclusions and future perspectives

The appearance of Atu3280-ct crystals in F 2 3 in both co-crystallised with SQ and ligand-free. Timely use of Atu3280-ct in crystallography appears to be of particular importance to producing useful crystals. Furthermore, crystallisation at a lower temperature may improve protein stability and lead to higher-quality crystals. The current conditions may lead to a solution by using experimental phasing, either by heavy atom soaking or selenomethionine incorporation. The osmotic sensitivity of the crystals obtained so far to soaking suggests a selenomethionine expression may be a more reliable method. Atu3280-ct contains only two Met residues, though the high copy count in the ASU should provide an adequate anomalous signal. If no screening conditions can produce diffraction quality crystals in a different spacegroup, and experimental phasing fails, the construct may require surface-entropy reduction. The resultant engineered construct may form different crystal contacts allowing for a more useful spacegroup. The degradation of Atu3280-ct over time observed by MS between the 1 and 20-day old aliquots confirms the importance of timely use of the protein for any experiments. This is particularly important in crystallography where sample heterogeneity can be a major influence on crystal quality.

As expression of the full-length Atu3280 was unsuccessful, it was deemed a more efficient use of limited time to work on the EBD mutant. Though no changes in OD₆₀₀ or wet cell mass

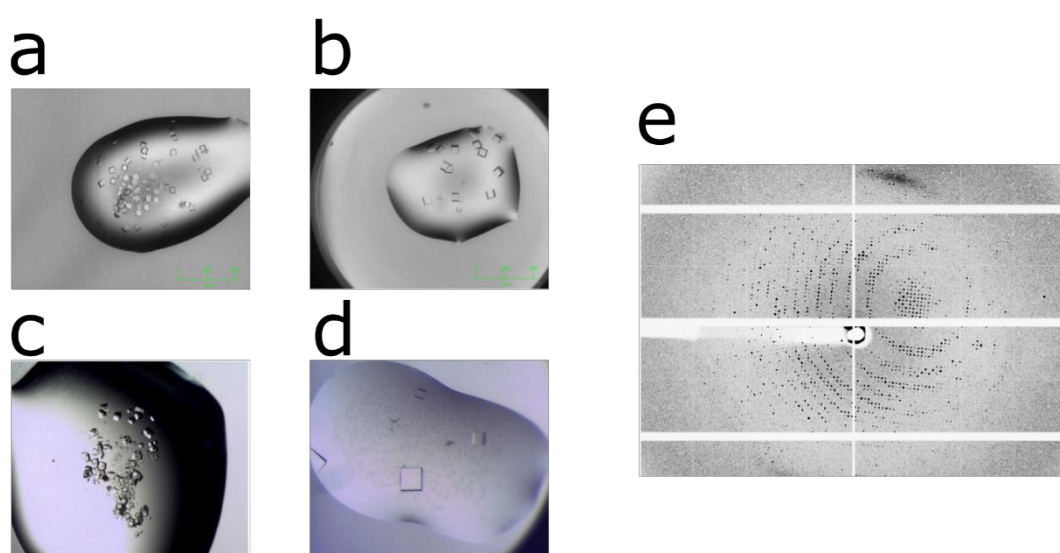


Fig 6-4: X-ray crystallography of Atu3280-ct. Crystals grown in initial screen (a) scaled-up optimisation (b), hanging drop post additive screen, streak seeding (c) optimisation of c (d). E. Representative diffraction image from best dataset collected.

suggestive of protein toxicity were observed, it is possible that the full-length protein caused cell death. A speculative mechanism would be that the palindromic binding region wHTH domains recognise is present in *E.coli* BL21(DE3). As the binding region is not known, this cannot be tested. A solution could be the use of a different bacterial strain, though with no way to test this hypothesis currently any choice of strain would be an educated guess. It is also possible that the codon optimisation used in gene synthesis caused an inability to express it. Given that this was specific to *E.coli* it is unlikely, but synthesis of a non-optimised gene may provide a solution. Should soluble Atu3280-fl be produced, SELEX followed by a gel-shift assay could be used to elucidate the DNA binding motif and assess effector specificity respectively, similar to the methods used with CsqR⁶². Synthesis of the DNA in question and its use in X-ray crystallography could then produce a complex structure, allowing for meaningful comparison between Atu3280 and other GntRs.

Overall, the investigations shown here represent a foundation and an exploration of some of the challenges this protein presents that stand in the way of successful characterization.

6.4. Aldehyde dehydrogenase SlaB

6.4.1. Introduction

Aldehyde dehydrogenases (ALDHs) are a superfamily of enzymes that enact the irreversible oxidation of an aldehyde to form a carboxylic acid. ALDHs are typically classified based on the tissues and sub-cellular locations the earliest characterized examples were found in, since ALDH research principally began with human variants. Classes 1 and 3 represent cytosolic ALDHs, while class 2 are constitutively expressed in the mitochondria, and found specifically

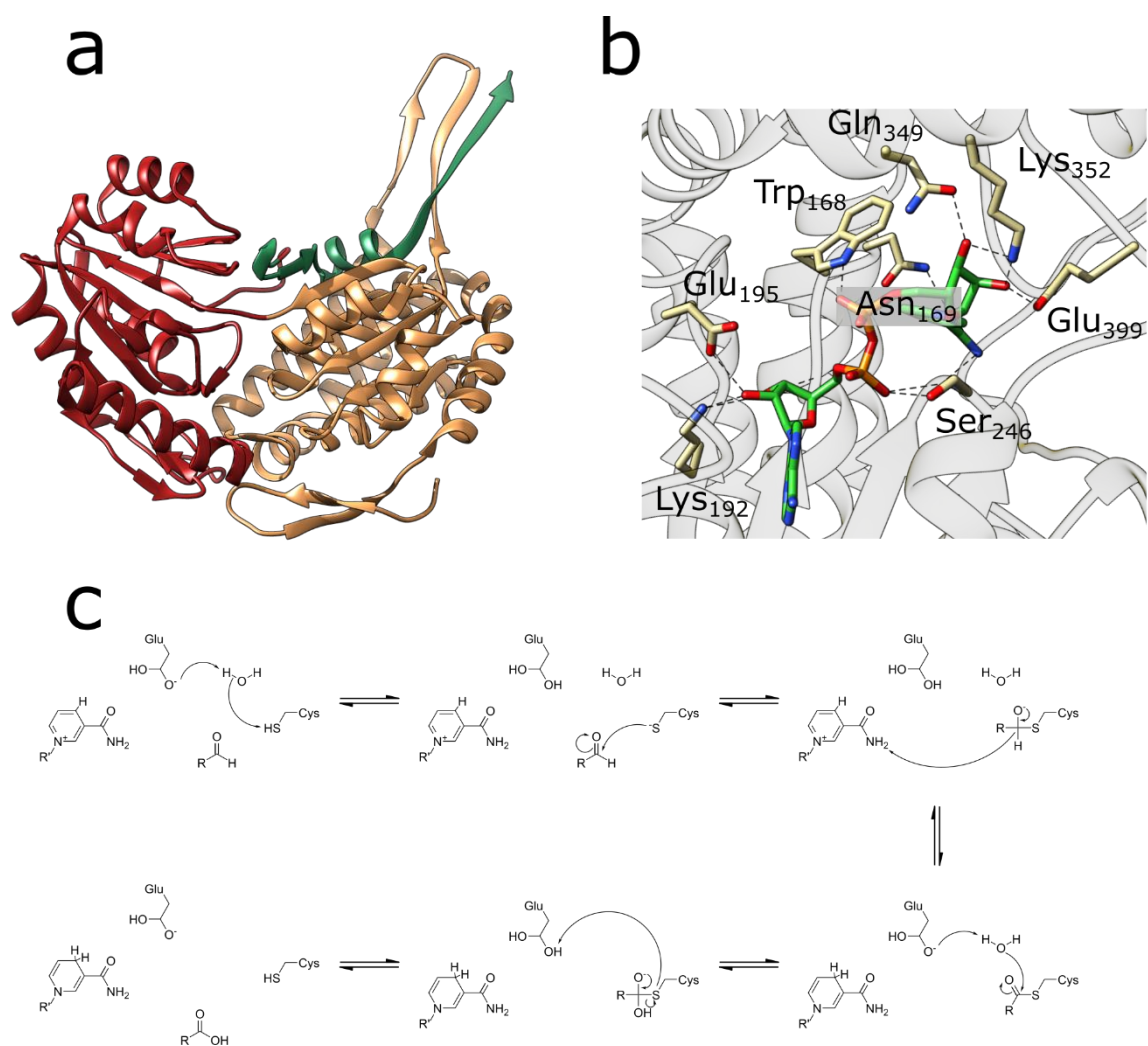


Fig 6-5: Structure, cofactor recognition and mechanism of GapN. **A.** GapN monomer coloured by domain: Nucleotide binding in yellow, catalytic in red, oligomerisation in green. **B.** Cofactor recognition site of GapN. **C.** Generalised mechanism employed by ALDHs including GapN.

in stomach and cornea cells, and in some tumours. Class 1 and 2 are generally homodimers,

and class 3 form tetramers. Classes 1 and 3 also generally show greater substrate promiscuity than class 2. The larger number of ALDHs discovered since have expanded the classification system such that any ALDH with <40% sequence identity to a class is added to it, and any within that have <60% sequence ID are further grouped into sub-classes ²¹.

ALDHs typically form dimers or tetramers, with the latter forming as a dimer of dimers. Formation of an ALDH homodimer occurs through the domain swap of a C-terminal “tail” domain (fig 6-5,a). The length and extent of this tail appears to determine whether ALDH dimers can form higher order structures, as dimeric ALDHs typically feature a 17-residue extension to the tail compared to tetrameric examples. This extension occupies the same location in dimeric ALDHs as part of the dimer-dimer interface in tetrameric complexes. Truncation of the tail in dimeric ALDHs does allow tetramer formation, but only in low-salt solvent. Addition of these residues to a tetrameric ALDH does preclude dimer-dimer association. Hexameric ALDHs have also been observed as a ring-shaped dimer triplet, complete with a distinct method for higher-order structure formation. An α -helical N-terminal extension is used in addition to the tail domain, acting as a surface-based binding domain. From these studies, it appears that the quaternary structure of an ALDH can be reliably predicted based on a sufficiently intact structure.

The cofactor binding domain uses a Rossmann fold, through which the cofactor binds along the loop between β 1 and α A, the “dinucleotide binding helix” which interacts with the pyrophosphate linker of the cofactor ²¹. ALDHs make few of the same interactions other NAD(P)H-utilising superfamilies such as AKRs use, and as such do not constrain the cofactor in an extended conformation (fig 6-5, b). Rather, a negatively charged patch allows for cofactor binding but not full conformational stabilisation. NAD(P)H is generally more ordered at the adenine moiety than the nicotinamide, with some structures even missing the density for the latter due to disorder. The mechanism of an ALDH means the cofactor must undergo movement to allow water entry and prevent clashes that would prevent proper action of the crowded active site ^{214,215}. NAD(P)H specificity arises from the charge of residues at the end of β 2; positive charge stabilises the 2' phosphate while negative charge prevents its binding allowing specificity to NADPH or NADH respectively ²¹⁶.

ALDHs share a common mechanism based on the use of a conserved catalytic Cys side chain, which forms a tetrahedral thiohemiacetal intermediate with the substrate ^{217,21}. Before this

can happen, the Cys must be activated by abstraction of a proton by the active site Glu side chain, which acts through an ordered water. Once the intermediate is formed, hydride transfer can occur to NAD(P)⁺ forming NAD(P)H, and a nucleophilic attack from the ordered water subsequently hydrolyses the thioester bond, allowing the removal of the product and regeneration of the catalytic Cys²¹⁸. The superfamily also share a common fold consisting of three domains (Figure 7-X, a). These are a nucleotide binding domain, responsible for interacting with and positioning the cofactor (yellow), a catalytic domain (red) and an oligomerisation domain (green)²¹⁵. The active site is situated at the bottom of a hydrophobic funnel structure situated close to the interface between the cofactor binding and catalytic domains²¹⁹ (fig 6-5, c).

The sulfo-TAL pathway contains an ALDH, which catalyses the oxidation of SLA to SL, which is then excreted⁷⁹. While employing a redox reaction as the final stage is a common feature in sulfoglycolysis, it is typically done using a semialdehyde reductase or aldo-keto reductase^{62,80}. Here, the gene for SLA ALDH SlaB is successfully expressed in *E.coli* and purified. X-ray crystallography is then used to produce the first known structure of SlaB, in a ligand-free state. These data represent an initial investigation that can be built upon, leading to an in-depth structural and biophysical characterization of the enzyme.

6.4.2. Results and discussion

Initial attempts at SlaB expression and purification did not yield purified protein. A band at ~50 kDa was present in the lysate and IMAC flow-through. SlaB has a sequence-derived Mw of 52 kDa making this band likely to be SlaB. This band was absent from all elution fractions indicating the C-terminal 6xHis-tagged protein could not bind the Ni-NTA resin (fig 6-6, a). The most probable reason was that the tag was being occluded by the protein and was therefore unavailable to bind the resin. The possibility that the tag was cleaved by a protease was also considered, though due to the use of a protease inhibitor added before cell lysis and lack of obvious sites at the C-terminus this was not considered likely. SlaB was instead sub-cloned into the pET-YSBLiC3C expression vector, which contains an N-terminal 6xHis tag.

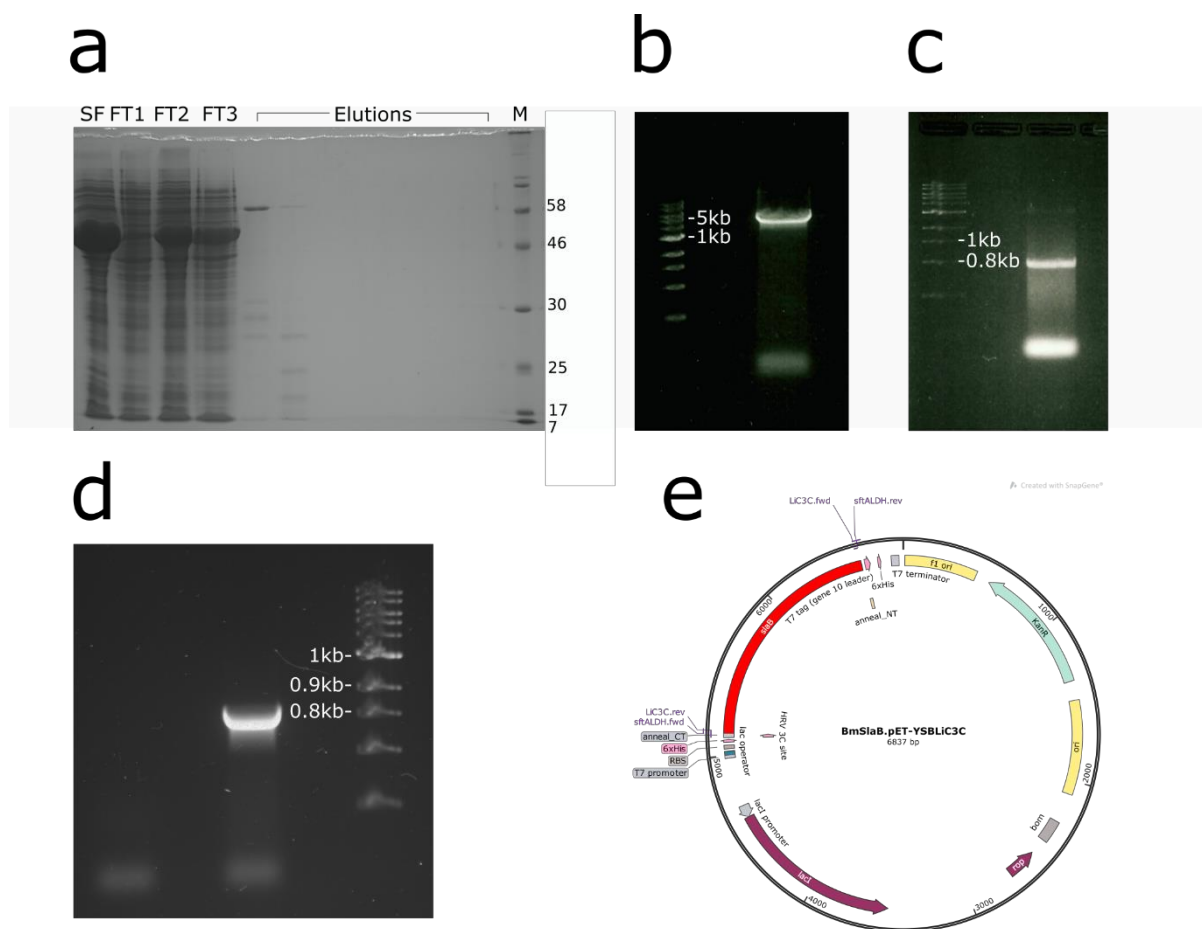


Fig 6-6: Initial expression and sub-cloning of SlaB in pET29a. **A.** SDS-PAGE of initial purification attempt. Left to right; Soluble fraction (SF), flow throughs FT1-3, elution fractions, ladder. **B.** PCR result of gene amplification. **C.** PCR result of vector amplification. **D.** Colony PCR result. **E.** Predicted plasmid for SlaB in pET-YSBLiC3C. This result was confirmed through sequencing.

Amplification of both the insert from the original pET28a vector and vector from a LiC3C stock were successfully performed using PCR, and ligation produced one colony confirmed through colony PCR and gene sequencing to contain SlaB.LiC3C with an identical sequence to the simulated result (fig 6-6, b-e).

SlaB was then re-expressed and purified in LiC3C with a final purity of < 95% as observed with SDS-PAGE (fig 6-7, a-c). While SlaB elutes in a clean peak from IMAC, the SEC elution profile contains several “shoulders” where close peaks blend. This indicates multiple populations are present. These could be impurities, but the relative intensity of each suggests more protein than seen for anything except SlaB in SDS-PAGE. The close packing of these peaks further

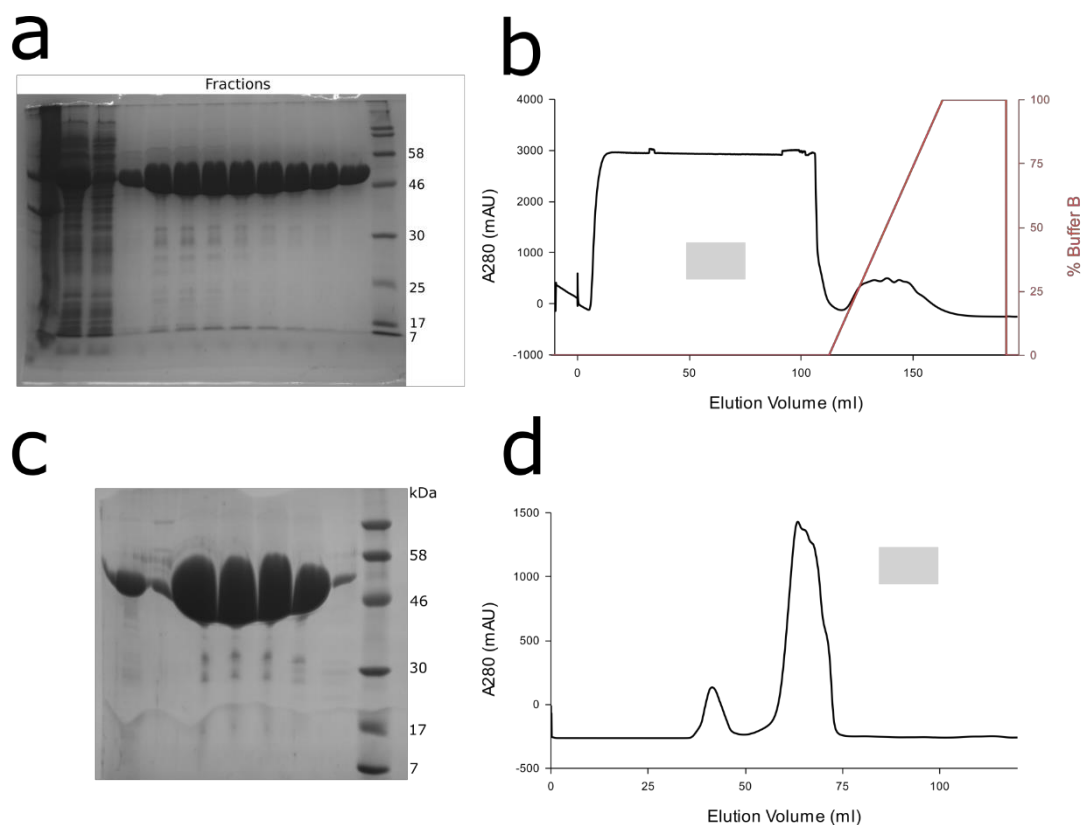


Fig 6-7: Expression and purification for SlaB in pET-YSBLiC3C. **A.** SDS-PAGE of extraction and purification of SlaB. Left to right; Insoluble fraction (ISF), soluble fraction (SF), flow through (FT), elution fractions, ladder. **B.** A₂₈₀ chromatogram of IMAC purification. % of high-imidazole elution buffer concurrent with elution in red. **C.** SDS-PAGE of SEC. Left to right; elution fractions, ladder. **D.** A₂₈₀ chromatogram of SEC elution.

implies these may be multimeric states for SlaB. As ALDHs typically form dimers or tetramers this was not unexpected. A higher Mw elution may be an impurity (fig 6-7, d).

Crystallography of SlaB began with the identification of a viable condition in JCSG when ligand-free, and one in the PEG/ION screen when co-crystallised with 2 mM SLA. However, no crystal harvested from these conditions diffracted. Co-crystallisation with 2 mM NADH produced crystals in PACT, and subsequent optimisation of the mother liquor allowed growth of a single crystal able to produce a diffraction pattern. The resultant dataset indexed into P1 with a comparatively large ASU (87 x 91 x 131 Å), indicating a high copy count was likely. Due to a poor MR model (sequence ID = 45%) and the poor resolution of the dataset (cut at 3.3 Å) no solution could be found at first. A solution became possible later using an AF prediction, though initial model quality statistics were poor ($R/R_f = 0.47/0.48$, table 6-1) suggesting this solution was far from perfect. The data are highly anisotropic ($CC1/2 = 0.5$ at 3.3 Å (d₁), 4.8 Å

	SlaB apo
Data collection and indexing	
Space group	P 1
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	87061, 91.45, 131.0
α , β , γ (°)	80.34, 74.53, 78.93
Resolution (Å)	3.3
R_{merge}	0.394(1.014)
R_{pim}	0.332(1.011)
$I / \sigma I$	4.3(1.1)
$CC1/2$	0.84(0.75)
Completeness (%)	98.1(87.4)
Redundancy	3.6(3.7)
Refinement	
Resolution (Å)	3.3
No. unique reflections	54616
$R_{\text{work}} / R_{\text{free}}$	0.37/0.45
Atom Count	
Protein	21347
Ligand/ion	0
Water	0
<i>B</i> factors (Å²)	
Protein	17.7
Ligand/ion	-
Water	-
Bond lengths (Å)	0.046
Bond angles (°)	2.29
Ramachandran residues	
In most favourable regions (%)	81.1
In allowed regions (%)	14.2
Outliers	4.7
PDB code	-

Table 7-1: Data collection and refinement statistics for SlaB. Values in parentheses are for the highest resolution shells. All values are within $I/\sigma I$ where applicable.

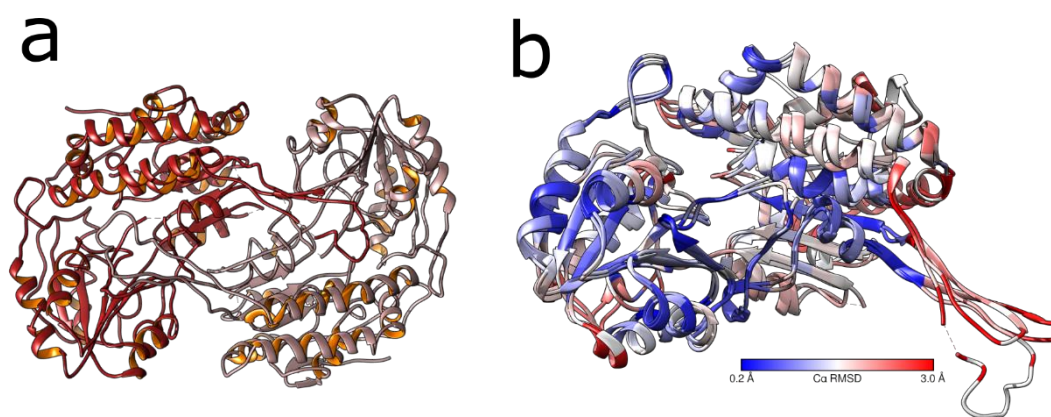


Fig 6-8: Crystal structure of SlaB. A. Observed dimer interaction of SlaB. **B.** RMSD ($C\alpha$) heatmap of SlaB relative to GapN. Lower RMSD residues in blue, higher in red.

(d2) and 3.5 Å (d3)), through with anisotropy correction this was considered a lesser problem than the overall poor resolution. Fitting a copy of the AF structure with all side chains removed to the solved electron density then using it as the basis for automated model building using BUCCANEER produced a monomeric model able to be copied to all copies present in the ASU, which were then refined using jelly-body and NCS restraints in an attempt to produce a useful model in a time-efficient manner. The resultant structure is still in an early stage of validation and contains 6 copies of SlaB in the ASU, 4 forming dimer interactions (fig 6-8, a). The remaining two monomers do not associate in the same way either within the ASU or through symmetry, so the arrangement overall is not likely to be biologically relevant. No density suggesting NADH can be seen in this model. The overall fold is similar to other ALDHs including GapN (global RMSD = 1.3 Å), though the sequence identity is low (34%) (fig 6-8, b). This is in line with other structurally characterized ALDHs. Based on this structure, SlaB is likely to be tetrameric since the tail domain is missing no residues relative to the tetrameric ALDH GapN²².

6.4.3. Conclusions and future perspectives

The data presented here represent an early stage in SlaB characterization. Expression was successful and was followed by efficient purification through changing the affinity tag location. However, attempts at X-ray crystallography have seen limited success with a single, low resolution and ligand-free dataset obtained. This was enough to confirm SlaB has a similar fold to other ALDHs, and is likely to follow a comparable binding scheme for the cofactor. However, the collection of a higher-resolution structure for SlaB should be a priority in the

future as this would allow a more in-depth structural analysis and comparison with other ALDHs than is possible at present. SlaB is also likely to be a viable cryo-EM target due to the large size of the complex (~100 kDa for a dimer, ~200 kDa for a tetramer). It is likely that SlaB forms a solution tetramer based on the length of the C-terminal tail domain, this requires further confirmation by a solution technique such as SEC-MALLS. Additionally, the collection of a ternary structure with SLA or SL would greatly enhance this study. As ALDH mechanisms are not reversible any complex involving NAD(P)H should be inactive. SL could alternatively be used in place of SLA should it bind as any structure is unlikely to reach sufficient resolution to differentiate between SLA and SL, given the results obtained this far for SlaB crystallography and extreme unlikelihood of obtaining a sub-angstrom resolution cryo-EM structure.

ALDHs have a well-understood mechanism that should be reasonably straightforward to assess in SlaB. Cofactor and substrate binding can be assayed using Tm analysis by nanoDSF, as seen in other targets in this thesis. Thermodynamic analyses can then be performed with ITC and compared to values such as those obtained for GapN. Furthermore, cofactor inhibition such as that observed for NADP⁺ ($K_D = 1 \mu\text{M}$) in GapN can be assessed, as can cofactor and ligand binding and catalytic residue conservation. The quality of the SlaB structure obtained here is not sufficient to confidently assess this. Finally, steady-state kinetic parameters may be obtained by measuring the absorbance change at 340nm, corresponding to reduction of NADP⁺ to NADPH.

7. Amino acid sequences used in this thesis

Protein	Sequence
SmoF.pET29b	MDAELKIFVSSQHQPDIWRKALDQYEAKTPGVKVIETGGNTSEMQAQYLNTVM SAKDSSLVDLMLDVIRPAQFATAGWTSDFSGKDL SAYLPTYAEANTVNGKIVAL PAFADSMFLYYRKDLLDKYGIKPPTTWDELKEASKKVMGEKKNPELQGLSFQ GK AIEGAVCTFLLPYWSEGKSLVENGKLNFDNKA AVDSLKLWKS FVDDGISKKNIS EVATDDTRKEFQAGKVLFAVNWSYAWTHFQ GKESQVNDKVGVARLPAVKGGEQT TCLGGWEFGVSAYSQQDEAKKLV EYLSQDVSKFMAINAALLPTYAALYK DAD VTKTIPWFADALPVVETAKARPVTPRYNEVSETIRTTVNGVLAGVMT PEDGAKQ MESRLRRVLRLEHHHHHH*
SmoB.pET29a	MQRIALSDKLELSRIVYGMWRIGDDADTSPA HVQAKIEACLAQGITTMDQADIY GGYTAEAILGGGLKAAPGLRDKIEIVTKCGIVAPAGR HSSARVKHYD TTAGHIN VSVEASLRDMGTDHVDLLL IHRPDPLIDAEETGKALDALVASGKVKAVGVS NFR PWDFSL LQSAMSNRLVTNQIEMSL LATDTFTNGDLAYLQEKRVSPMAWSPLGGG SLFSGAYGGTMAALQRIGKEQGV DATAVAIAWLLRHPAKIVPVLGTNNLERIRT AADALRV TMDRQTFEL YTLAIGKEVALEHHHHHH*
SmoB.pET- YSBLic3C	MGSSHHHHHHSSGLEVL FQGPAMQRIAL SDKLELSRIVYGMWRIGDDADTSPA H VQAKIEACLAQGITTMDQADIYGGYTAEAILGGGLKAAPGLRDKIEIVTKCGIV APAGR HSSARVKHYD TTAGHINVSVEASLRDMGTDHVDLLL IHRPDPLIDAEET GKALDALVASGKVKAVGVS NFRPWDFSL LQSAMSNRLVTNQIEMSL LATDTFTN GDLAYLQEKRVSPMAWSPLGGGSLFSGAYGGTMAALQRIGKEQGV DATAVAIAW LLRHPAKIVPVLGTNNLERIRTAADALRV TMDRQTFEL YTLAIGKEVA*
Atu3280- ct.pET29a	MGLLERVQSGSNYIRAVSDPQSVYAFFRLELI EGGGLPTAEVLDVARLAKPADL PAFGTSTEGHRIRRLRRIAGKPA AIEE IWL DGSYVDTIT IENMSESLYLYRTR LNLWISKAEDRIDLGDVPEWAPEVFGQKAGSSVPRVLR L SQAQDGAVAEVSWTW FDHTVARYVSRIRLEHHHHHH*
Atu3280- fl.pET29a	MKHTGGSLPMYMQIAEMLVREVAAGRLIDGEKLAPERDMAADLGI AVGTLRKSL AELQERGLLERVQSGSNYIRAVSDPQSVYAFFRLELI EGGGLPTAEVLDVARLA KPADLPAFGTSTEGHRIRRLRRIAGKPA AIEE IWL DGSYVDTIT IENMSESLYLY YRTRLNLWISKAEDRIDLGDVPEWAPEVFGQKAGSSVPRVLR L SQAQDGAVAE VSWTWFDHTVARYVSRIRLEHHHHHH*
SqvA.pEt29a	MKYFLDSAIL E EIRYAYENWAIDGVTTNPRHIMNSGKPFLTVLDEFASEFKGVE NFPISVEINPHLDNAKDMVEEGTKIAKLS SNFVIKIPCTEPGLIAAKEFEKQGI STNVTLVFSPSQALQPARIGAKFVSPFVGWKENSGDDTTQYIQDIVNIYKNYNY NTEIIVAALRNGKQIVDAKAGAHIVTCGFDVYKESFQHAFTDYGLNKFRNAWD NTVTEAPVLKLEHHHHHH*
SlaB.pET29a	MTSLTQVKQYGLYVNGEWETTAEKMEVLN KYTQQPAAEISVATKDDVNKAVASA KDALKNTFSPYERYEVL MKAADLLL SRQEEFAEILATEVGKSIRESRGEVERAA TTLQISAE EAKRIHGEGVPVESAPGSENMAFTVKVPVGVVAAITPFNVPINLV CHKLGPALAAGNSVVLKPAEVT PICALKLAELMEEAGL PKGRLQVLTGDGAEIG EWLLENQDVMFTFTGSPRVGELIRSKAGLRKVSLELGNSATIVHKDADLEKA ASLISQKSFNNAGQVCISVQRIYVHTNIYTA FVNKLKEKTEKLVVGNPMDEQTD IGPMIRLKEAERVEEWVKEAVE EGAKIELGGKRDGAFYLP TILT NVNDDMKVCR QEVFGPAVAIAQYDEIDEVISKVNDSDYGLQAGLFTNDLQFAMKAAREIEVGG IVNDASAYRVDHMPYGGVKKSGNGKEGPKYAIEEMTEER IIVLNLEHHHHHH*
SlaB.pET- YSBLic3C	MGSSHHHHHHSSGLEVL FQGPAHMTSLTQVKQYGLYVNGEWETTAEKMEVLN KY TQQPAAEISVATKDDVNKAVASAKDALKNTFSPYERYEVL MKAADLLL SRQEEF AEILATEVGKSIRESRGEVERAATTLQISAE EAKRIHGEGVPVESAPGSENMA FTVKVPVGVVAAITPFNVPINLVCHKLGPALAAGNSVVLKPAEVT PICALKLAEL MEEAGL PKGRLQVLTGDGAEIGEWLLENQDVMFTFTGSPRVGELIRSKAGLR

	KVSLELGNN SATIVHK DADLEKAAS LISQK SFNNAGQVCISVQRIYVHTNIYTA FVNKLKEKTEKLVVGNPMDEQTDIGPMIRLKEAERVEEWKEAVEEGAKIELGG KRDGAFYLP TILTNVNDDMKVCRQEVFGPAVAIAQYDEIDEVISKVNDSDYGLQ AGLFTNDLQFAMKAAREIEVGGLIVNDASAYRVDHMPYGGVKKSGNGKEGPKYA IEEMTEERIIVLNL*
--	---

Table 7-1: Amino acid sequences for the proteins used in this thesis

8. References

1. Hutzinger, O. & Zehnder, A. J. B. *The natural environment and the biogeochemical cycles*. Springer-Verlag **1**, (1980).
2. Kertesz, M. A. *Riding the sulfur cycle - Metabolism of sulfonates and sulfate esters in Gram-negative bacteria*. *FEMS Microbiology Reviews* **24**, 135–175 (Oxford Academic, 2000).
3. Harwood, J. L. & Nicholls, R. G. The plant sulfolipid-- a major component of the sulphur cycle. *Biochem. Soc. Trans.* **7**, 440–447 (1979).
4. Benson, A. A., Daniel, H. & Wiser, R. A sulfolipid in plants. *Proc. Natl. Acad. Sci.* **45**, 1582–1587 (1959).
5. Goddard-Borger, E. D. & Williams, S. J. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem. J.* **474**, 827–849 (2017).
6. Douce, R., Holtz, R. B. & Benson, A. A. Isolation and properties of the envelope of spinach chloroplasts. *J. Biol. Chem.* **248**, 7215–7222 (1973).
7. Radunz, A. über das Sulfochinovosyl-diacylglycerin aus höheren Pflanzen, Algen und Purpurbakterien. *Hoppe. Seylers. Z. Physiol. Chem.* **350**, 411–417 (1969).
8. Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T. & Devol, A. H. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8607–8612 (2006).
9. Naoki, S. & Norio, M. Lipid biosynthesis in the blue-green alga, *Anabaena variabilis*. II. Fatty acids and lipid molecular species. *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.* **710**, 279–289 (1982).
10. Keusgen, M. *et al.* Sulfoquinovosyl diacylglycerols from the alga *Heterosigma carterae*. *Lipids* **32**, 1101–1112 (1997).
11. Roughan, G. & Slack, R. Glycerolipid synthesis in leaves. *Trends Biochem. Sci.* **9**, 383–386 (1984).
12. Heinz, E. *et al.* Synthesis of different nucleoside 5'-diphospho-sulfoquinovoses and their use for studies on sulfolipid biosynthesis in chloroplasts. *Eur. J. Biochem.* **184**, 445–453 (1989).
13. Lewis, G. F. Lipid metabolism. *Current opinion in lipidology* **15**, 85–87 (2004).
14. Benning, C. & Somerville, C. R. Identification of an operon involved in sulfolipid biosynthesis in *Rhodobacter sphaeroides*. *J. Bacteriol.* **174**, 6479–6487 (1992).
15. Benning, C. & Somerville, C. R. Isolation and genetic complementation of a sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. *J. Bacteriol.* **174**, 2352–2360 (1992).
16. Benning, C., Beatty, J. T., Prince, R. C. & Somerville, C. R. The sulfolipid sulfoquinovosyldiacylglycerol is not required for photosynthetic electron transport in *Rhodobacter sphaeroides* but enhances growth under phosphate limitation. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1561–1565 (1993).
17. Kleppinger-Sparace, K. F. & Mudd, J. B. Biosynthesis of sulfoquinovosyldiacylglycerol in higher plants: Use of adenosine-5'-phosphosulfate and adenosine-3'-phosphate 5'-phosphosulfate as precursors. *Plant Physiol.* **93**, 256–263 (1990).

18. Rossak, M., Tietje, C., Heinz, E. & Benning, C. Accumulation of UDP-sulfoquinovose in a sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. *J. Biol. Chem.* **270**, 25792–25797 (1995).
19. Rossak, M., Schäfer, A., Xu, N., Gage, D. A. & Benning, C. Accumulation of sulfoquinovosyl-1-O-dihydroxyacetone in a sulfolipid-deficient mutant of *Rhodobacter sphaeroides* inactivated in *sqdC*. *Arch. Biochem. Biophys.* **340**, 219–230 (1997).
20. Güler, S., Seeliger, A., Härtel, H., Renger, G. & Benning, C. A null mutant of *Synechococcus* sp. PCC7942 deficient in the sulfolipid sulfoquinovosyl diacylglycerol. *J. Biol. Chem.* **271**, 7501–7507 (1996).
21. Güler, S., Essigmann, B. & Benning, C. A cyanobacterial gene, *sqdX*, required for biosynthesis of the sulfolipid sulfoquinovosyldiacylglycerol. *J. Bacteriol.* **182**, 543–545 (2000).
22. Yu, B., Xu, C. & Benning, C. Arabidopsis disrupted in *SQD2* encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5732–5737 (2002).
23. Sato, N. *et al.* Identification of genes for sulfolipid synthesis in primitive red alga *Cyanidioschyzon merolae*. *Biochem. Biophys. Res. Commun.* **470**, 123–129 (2016).
24. Murata, N. & Siegenthaler, P.-A. Lipids in Photosynthesis: An Overview. in *Lipids in Photosynthesis: Structure, Function and Genetics* 1–20 (Kluwer Academic Publishers, 2006). doi:10.1007/0-306-48087-5_1
25. Nußberger, S., Dörr, K., Wang, D. N. & Kühlbrandt, W. Lipid-protein interactions in crystals of plant light-harvesting complex. *J. Mol. Biol.* **234**, 347–356 (1993).
26. Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* **438**, 1040–1044 (2005).
27. Dubertret, G., Gerard-Hirne, C. & Trémolières, A. Importance of trans- Δ^3 -hexadecenoic acid containing phosphatidylglycerol in the formation of the trimeric light-harvesting complex in *Chlamydomonas*. *Plant Physiol. Biochem.* **40**, 829–836 (2002).
28. McCourt, P., Browse, J., Watson, J., Arntzen, C. J. & Somerville, C. R. Analysis of Photosynthetic Antenna Function in a Mutant of *Arabidopsis thaliana* (L.) Lacking trans - Hexadecenoic Acid. *Plant Physiol.* **78**, 853–858 (1985).
29. Minoda, A. *et al.* Role of sulfoquinovosyl diacylglycerol for the maintenance of photosystem II in *Chlamydomonas reinhardtii*. *Eur. J. Biochem.* **269**, 2353–2358 (2002).
30. Minoda, A., Sonoike, K., Okada, K., Sato, N. & Tsuzuki, M. Decrease in the efficiency of the electron donation to tyrosine Z of photosystem II in an SQDG-deficient mutant of *Chlamydomonas*. *FEBS Lett.* **553**, 109–112 (2003).
31. Voß, R., Radunz, A. & Schmid, G. H. Binding of Lipids onto Polypeptides of the Thylakoid Membrane I. Galactolipids and Sulfolipid as Prosthetic Groups of Core Peptides of the Photosystem II Complex. *Zeitschrift für Naturforsch. - Sect. C J. Biosci.* **47**, 406–415 (1992).
32. Sato, N., Sonoike, K., Tsuzuki, M. & Kawaguchi, A. Impaired Photosystem II in a Mutant of *Chlamydomonas Reinhardtii* Defective in Sulfoquinovosyl Diacylglycerol. *Eur. J. Biochem.* **234**, 16–23 (1995).
33. Yu, H. *et al.* Cryo-EM structure of monomeric photosystem II at 2.78 Å resolution reveals factors important for the formation of dimer. *Biochim. Biophys. Acta - Bioenerg.* **1862**, 148471 (2021).

34. Benning, C. Biosynthesis and Function of the Sulfolipid Sulfoquinovosyl Diacylglycerol. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 53–75 (1998).
35. Aoki, M., Sato, N., Meguro, A. & Tsuzuki, M. Differing involvement of sulfoquinovosyl diacylglycerol in photosystem II in two species of unicellular cyanobacteria. *Eur. J. Biochem.* **271**, 685–693 (2004).
36. Shibuya, I. & Benson, A. A. Hydrolysis of α -Sulphoquinovosides by β -Galactosidase. *Nature* **192**, 1186–1187 (1961).
37. Yagi, T. & Benson, A. A. Plant sulfolipid. V. Lysosulfolipid formation. *Biochim. Biophys. Acta* **57**, 601–603 (1962).
38. Dougal Burns, D., Galliard, T. & Harwood, J. L. Properties of acyl hydrolase enzymes from *Phaseolus multiflorus* leaves. *Phytochemistry* **19**, 2281–2285 (1980).
39. Andersson, L. *et al.* Hydrolysis of galactolipids by human pancreatic lipolytic enzymes and duodenal contents. *J. Lipid Res.* **36**, 1392–1400 (1995).
40. Denger, K. *et al.* Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **507**, 114–117 (2014).
41. Felux, A. K., Spitteller, D., Klebensberger, J. & Schleheck, D. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4298–E4305 (2015).
42. 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).
43. Laine, R. A. Invited commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10 structures for a reducing hexasaccharide: The Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**, 759–767 (1994).
44. Sinnott, M. L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem. Rev.* **90**, 1171–1202 (1990).
45. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
46. McCarter, J. D. & Stephen Withers, G. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885–892 (1994).
47. Koshland, D. E. STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416–436 (1953).
48. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316 (1991).
49. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788 (1993).
50. Henrissat, B. & Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal* **316**, 695–696 (1996).
51. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317 (2003).

52. CAZy. CAZy - Home. (2022). Available at: <http://www.cazy.org/>. (Accessed: 18th March 2022)
53. Okuyama, M., Mori, H., Chiba, S. & Kimura, A. Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from *Escherichia coli*. *Protein Expr. Purif.* **37**, 170–179 (2004).
54. Lovering, A. L., Seung, S. L., Kim, Y. W., Withers, S. G. & Strynadka, N. C. J. Mechanistic and structural analysis of a family 31 α -glycosidase and its glycosyl-enzyme intermediate. *J. Biol. Chem.* **280**, 2105–2115 (2005).
55. Abayakoon, P. *et al.* Structural and Biochemical Insights into the Function and Evolution of Sulfoquinovosidases. *ACS Cent. Sci.* **4**, 1266–1273 (2018).
56. Martelli, H. L. *Oxidation of sulphonic compounds by aquatic bacteria isolated from rivers of the Amazon region.* *Nature* **216**, 1238–1239 (1967).
57. Kok, L. J. De, Rennenberg, H., Brunold, C. & Rauser, W. E. *Sulfur Nutrition and Assimilation in Higher Plants; Regulatory, Agricultural and Environmental Aspects.* (SPB, 1993).
58. Roy, A. B., Ellis, A. J., White, G. F. & Harwood, J. L. Microbial degradation of the plant sulpholipid. in *Biochemical Society Transactions* **28**, 781–783 (2000).
59. Roy, A. B., Hewlins, M. J. E., 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).
60. Meyerhof, O. & Junowicz-Kocholaty, R. THE EQUILIBRIA OF ISOMERASE AND ALDOLASE, AND THE PROBLEM OF THE PHOSPHORYLATION OF GLYCERALDEHYDE PHOSPHATE. *J. Biol. Chem.* **149**, 71–92 (1943).
61. Cornish-Bowden, A. Thermodynamic aspects of glycolysis. *Biochem. Educ.* **9**, 133–137 (1981).
62. Sharma, M. *et al.* Molecular Basis of Sulfosugar Selectivity in Sulfoglycolysis. *ACS Cent. Sci.* **7**, 476–487 (2021).
63. Itoh, T., Mikami, B., Hashimoto, W. & Murata, K. Crystal Structure of YihS in Complex with d-Mannose: Structural Annotation of *Escherichia coli* and *Salmonella enterica* yihS-encoded Proteins to an Aldose-Ketose Isomerase. *J. Mol. Biol.* **377**, 1443–1459 (2008).
64. Sigrell, J. A., Cameron, A. D., Jones, T. A. & Mowbray, S. L. Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 Å resolution: Insights into a new family of kinase structures. *Structure* **6**, 183–193 (1998).
65. Cabrera, R., Babul, J. & Guixé Victoria, V. Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Arch. Biochem. Biophys.* **502**, 23–30 (2010).
66. Shimada, T. *et al.* Regulatory role of CsqR (YihW) in transcription of the genes for catabolism of the anionic sugar sulfoquinovose (SQ) in *Escherichia coli* K-12. *Microbiol. (United Kingdom)* **165**, 78–89 (2019).
67. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A. M. Sulfoquinovose degraded by pure cultures of bacteria with release of C3-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39–45 (2012).
68. Felux, A. K., Franchini, P. & Schleheck, D. Permanent draft genome sequence of sulfoquinovose-degrading *Pseudomonas putida* strain SQ1. *Stand. Genomic Sci.* **10**, 1–6 (2015).

69. Abayakoon, P. *et al.* Discovery and characterization of a sulfoquinovose mutarotase using kinetic analysis at equilibrium by exchange spectroscopy. *Biochem. J.* **475**, 1371–1383 (2018).
70. Li, J. *et al.* A sulfoglycolytic entner-doudoroff pathway in *Rhizobium leguminosarum* bv. trifolii SRDI565. *Appl. Environ. Microbiol.* **86**, (2020).
71. Thoden, J. B., Kim, J., Raushel, F. M. & Holden, H. M. The catalytic mechanism of galactose mutarotase. *Protein Sci.* **12**, 1051–1059 (2003).
72. Salas, M., Vinuela, E. & Sols, A. Spontaneous and Enzymatically Catalyzed Anomerization of Glucose 6-Phosphate and Anomeric Specificity of Related Enzymes. *J. Biol. Chem.* **240**, 561–568 (1965).
73. Osipiuk, J., Zhou, M., Moy, S., Collart, F. & Joachimiak, A. X-Ray crystal structure of GarR-tartronate semialdehyde reductase from *Salmonella typhimurium*. *J. Struct. Funct. Genomics* **10**, 249–253 (2009).
74. Sharma, M. *et al.* Dynamic structural changes accompany production of Dihydroxypropanesulfonate by Sulfolactaldehyde reductase. *ACS Catal.* **10**, 2826–2836 (2020).
75. Liu, J. *et al.* Mechanistically Diverse Pathways for Sulfoquinovose Degradation in Bacteria. *ACS Catal.* **11**, 14740–14750 (2021).
76. Burrichter, A. *et al.* Anaerobic degradation of the plant sugar sulfoquinovose concomitant with H₂S production: *Escherichia coli* k-12 and *desulfovibrio* sp. Strain DF1 as co-culture model. *Front. Microbiol.* **9**, 1–18 (2018).
77. Lehwess-Litzmann, A. *et al.* Twisted Schiff base intermediates and substrate locale revise transaldolase mechanism. *Nat. Chem. Biol.* **7**, 678–684 (2011).
78. 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).
79. Liu, Y. *et al.* A transaldolase-dependent sulfoglycolysis pathway in *Bacillus megaterium* DSM 1804. *Biochem. Biophys. Res. Commun.* **533**, 1109–1114 (2020).
80. Sharma, M. *et al.* Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proc. Natl. Acad. Sci.* **119**, e2116022119 (2022).
81. Van Der Ploeg, J. R., Iwanicka-Nowicka, R., Bykowski, T., Hryniewicz, M. M. & Leisinger, T. The *Escherichia coli* ssuEADCDB 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).
82. Driggers, C. M., Dayal, P. V., Ellis, H. R. & Andrew Karplus, P. A. Crystal structure of *Escherichia coli* SsuE: Defining a general catalytic cycle for FMN reductases of the flavodoxin-like superfamily. *Biochemistry* **53**, 3509–3519 (2014).
83. Abdurachim, K. & Ellis, H. R. Detection of protein-protein interactions in the alkanesulfonate monooxygenase system from *Escherichia coli*. *J. Bacteriol.* **188**, 8153–8159 (2006).
84. Snow, A. J. D., Burchill, L., Sharma, M., Davies, G. J. & Williams, S. J. Sulfoglycolysis: catabolic pathways for metabolism of sulfoquinovose. *Chem. Soc. Rev.* **50**, 13628–13645 (2021).
85. 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).
86. 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).
 87. Toplak, M., Matthews, A. & Teufel, R. The devil is in the details: The chemical basis and mechanistic versatility of flavoprotein monooxygenases. *Archives of Biochemistry and Biophysics* **698**, 108732 (2021).
 88. Vigonsky, E., Ovcharenko, E. & Lewinson, O. Two molybdate/tungstate ABC transporters that interact very differently with their substrate binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5440–5445 (2013).
 89. Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
 90. Scheepers, G. H., Lycklama a Nijeholt, J. A. & Poolman, B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* **590**, 4393–4401 (2016).
 91. Fischer, M., Zhang, Q. Y., Hubbard, R. E. & Thomas, G. H. Caught in a TRAP: Substrate-binding proteins in secondary transport. *Trends in Microbiology* **18**, 471–478 (2010).
 92. Oldham, M. L. Crystal Structure of the Maltose Intermediate State. *Science (80-)*. **1202**, 1202–1206 (2011).
 93. Tame, J. R. H. *et al.* The structural basis of sequence-independent peptide binding by OppA protein. *Science (80-)*. **264**, 1578–1581 (1994).
 94. Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216 (2007).
 95. Erkens, G. B. *et al.* The structural basis of modularity in ECF-type ABC transporters. *Nat. Struct. Mol. Biol.* **18**, 755–760 (2011).
 96. Szmelcman, S., Sassoon, N. & Hofnung, M. Residues in the α helix 7 of the bacterial maltose binding protein which are important in interactions with the Mal FGK2 complex. *Protein Sci.* **6**, 628–636 (1997).
 97. Martinez-Pomares, L. *et al.* Carbohydrate-independent recognition of collagens by the macrophage mannose receptor. *Eur. J. Immunol.* **36**, 1074–1082 (2006).
 98. Austermuhle, M. I., Hall, J. A., Klug, C. S. & Davidson, A. L. Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport. *J. Biol. Chem.* **279**, 28243–28250 (2004).
 99. Prossnitz, E., Nikaido, K., Ulbrich, S. J. & Ames, G. F. L. Formaldehyde and photoactivatable cross-linking of the periplasmic binding protein to a membrane component of the histidine transport system of *Salmonella typhimurium*. *J. Biol. Chem.* **263**, 17917–17920 (1988).
 100. Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of maltose transport in *Escherichia coli*: Transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2360–2364 (1992).
 101. Lu, G., Westbrooks, J. M., Davidson, A. L. & Chen, J. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17969–17974 (2005).
 102. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515–521 (2007).

103. Weiss, M. S. *et al.* Molecular architecture and electrostatic properties of a bacterial porin. *Science* (80-.). **254**, 1627–1630 (1991).
104. Hopfner, K. P. *et al.* Structural Biology of Rad50 ATPase: ATP-Driven Conformational Control in DNA Double-Strand Break Repair and the ABC-ATPase Superfamily Karl-Peter. *Cell* **101**, 789–800 (2000).
105. Mourez, M., Hofnung, M. & Dassa, E. Subunit interactions in ABC transporters: A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* **16**, 3066–3077 (1997).
106. Moody, J. E., Millen, L. & Binns, D. Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *J. Biol. Chem.* **277**, 21111–21114 (2002).
107. Fetsch, E. E. & Davidson, A. L. Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9685–9690 (2002).
108. Davidson, A. L. & Chen, J. ATP-binding cassette transporters in bacteria. *Annual Review of Biochemistry* **73**, 241–268 (2004).
109. Böhm, A., Diez, J., Diederichs, K., Welte, W. & Boos, W. Structural model of MalK, the ABC subunit of the maltose transporter of *Escherichia coli*: Implications for mal gene regulation, inducer exclusion, and subunit assembly. *J. Biol. Chem.* **277**, 3708–3717 (2002).
110. Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quijcho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12**, 651–661 (2003).
111. Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10**, 139–149 (2002).
112. Mi, W. *et al.* Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233–237 (2017).
113. Syberg, F., Suveyzdis, Y., Kötting, C., Gerwert, K. & Hofmann, E. Time-resolved fourier transform infrared spectroscopy of the nucleotide-binding domain from the ATP-binding cassette transporter MsbA: ATP hydrolysis is the rate-limiting step in the catalytic cycle. *J. Biol. Chem.* **287**, 23923–23931 (2012).
114. Tam, R. & Saier, M. H. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews* **57**, 320–346 (1993).
115. Tame, J. R., Dodson, E. J., Murshudov, G., Higgins, C. F. & Wilkinson, A. J. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **3**, 1395–1406 (1995).
116. Li, C. Y. *et al.* Structural insights into the multispecific recognition of dipeptides of deep-sea gram-negative bacterium *Pseudoalteromonas* sp. strain SM9913. *J. Bacteriol.* **197**, 1125–1134 (2015).
117. Shukla, S. *et al.* Differential Substrate Recognition by Maltose Binding Proteins Influenced by Structure and Dynamics. *Biochemistry* **57**, 5864–5876 (2018).
118. Silva, Z. *et al.* The high-affinity maltose/trehalose ABC transporter in the extremely thermophilic bacterium *Thermus thermophilus* HB27 also recognizes sucrose and palatinose. *J. Bacteriol.* **187**, 1210–1218 (2005).

119. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
120. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
121. Beilsten-Edmands, J. *et al.* Scaling diffraction data in the DIALS software package: Algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 385–399 (2020).
122. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
123. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 22–25 (2010).
124. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
125. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
126. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
127. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
128. Potterton, L. *et al.* CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 68–84 (2018).
129. Schrödinger, L. The PyMol Molecular Graphics System. (2015).
130. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
131. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394 (2011).
132. Snow, A. J. D. *et al.* The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids. *Curr. Res. Struct. Biol.* **4**, 51–58 (2022).
133. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
134. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
135. Zhang, Y. *et al.* Concise synthesis of sulfoquinovose and sulfoquinovosyl diacylglycerides, and development of a fluorogenic substrate for sulfoquinovosidases. *Org. Biomol. Chem.* **18**, 675–686 (2020).
136. Rosa, L. T., Bianconi, M. E., Thomas, G. H. & Kelly, D. J. Tripartite ATP-independent periplasmic (TRAP) transporters and Tripartite Tricarboxylate Transporters (TTT): From uptake to pathogenicity. *Frontiers in Cellular and Infection Microbiology* **8**, 33 (2018).
137. Shlykov, M. A., Zheng, W. H., Chen, J. S. & Saier, M. H. Bioinformatic characterization of the 4-Toluene Sulfonate Uptake Permease (TSUP) family of transmembrane proteins. *Biochim.*

- Biophys. Acta - Biomembr.* **1818**, 703–717 (2012).
138. Sellés Vidal, L., Kelly, C. L., Mordaka, P. M. & Heap, J. T. Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application. *Biochim. Biophys. Acta - Proteins Proteomics* **1866**, 327–347 (2018).
 139. Stehle, T., Claiborne, A. & Schulz, G. E. NADH binding site and catalysis of NADH peroxidase. *Eur. J. Biochem.* **211**, 221–226 (1993).
 140. Kavanagh, K. L., Klimacek, M., Nidetzky, B. & Wilson, D. K. The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry* **41**, 8785–8795 (2002).
 141. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* **43**, 206–209 (2015).
 142. Hoog, S. S., Pawlowski, J. E., Alzari, P. M., Penning, T. M. & Lewis, M. Three-dimensional structure of rat liver 3 α -hydroxysteroid/dihydrodiol dehydrogenase: A member of the aldo-keto reductase superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2517–2521 (1994).
 143. Jez, J. M., Flynn, T. G. & Penning, T. M. A new nomenclature for the aldo-keto reductase superfamily. *Biochemical Pharmacology* **54**, 639–647 (1997).
 144. Mindnich, R. D. & Penning, T. M. Aldo-keto reductase (AKR) superfamily: genomics and annotation. *Human genomics* **3**, 362–370 (2009).
 145. Wilson, D. K., Kavanagh, K. L., Klimacek, M. & Nidetzky, B. The xylose reductase (AKR2B5) structure: Homology and divergence from other aldo-keto reductases and opportunities for protein engineering. in *Chemico-Biological Interactions* **143–144**, 515–521 (Elsevier, 2003).
 146. Hyndman, D., Bauman, D. R., Heredia, V. V. & Penning, T. M. The aldo-keto reductase superfamily homepage. in *Chemico-Biological Interactions* **143–144**, 621–631 (Elsevier, 2003).
 147. Jez, J. M., Bennett, M. J., Schlegel, B. P., Lewis, M. & Penning, T. M. Comparative anatomy of the aldo-keto reductase superfamily. *Biochemical Journal* **326**, 625–636 (1997).
 148. Di Luccio, E., Elling, R. A. & Wilson, D. K. Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochem. J.* **400**, 105–114 (2006).
 149. Wilson, D. K., Rudolph, F. B. & Quijcho, F. A. Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutations. *Science (80-.).* **252**, 1278–1284 (1991).
 150. Kim, Y. *et al.* Crystal Structure of Putative Aldo/Keto Reductase from *Salmonella enterica*. *TO BE Publ.* (2019). doi:10.2210/PDB4R90/PDB
 151. Long, S. B., Campbell, E. B. & Mackinnon, R. Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K β Channel. *Science (80-.).* **309**, 897–903 (2005).
 152. Totir, M. *et al.* Macro-to-micro structural proteomics: Native source proteins for high-throughput crystallization. *PLoS One* **7**, e32498 (2012).
 153. Jürgens, C. *et al.* Directed evolution of a ($\beta\alpha$) $_8$ -barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9925–9930 (2000).
 154. Krump, C., Vogl, M., Brecker, L., Nidetzky, B. & Kratzer, R. Acceleration of an aldo-keto reductase by minimal loop engineering. *Protein Eng. Des. Sel.* **27**, 245–248 (2014).
 155. Penning, T. M. The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234**, 236–246

- (2015).
156. Kapila Ratnam, Haiching Ma, A. & Penning, T. M. The Arginine 276 Anchor for NADP(H) Dictates Fluorescence Kinetic Transients in 3 α -Hydroxysteroid Dehydrogenase, a Representative Aldo-Keto Reductase. *Biochemistry* **38**, 7856–7864 (1999).
 157. Bachur, N. R. Cytoplasmic aldo-keto reductases: A class of drug metabolizing enzymes. *Science (80-.)*. **193**, 595–597 (1976).
 158. Jin, Y. & Penning, T. M. Aldo-keto reductases and bioactivation/detoxication. *Annual Review of Pharmacology and Toxicology* **47**, 263–292 (2007).
 159. Kratzer, R., Wilson, D. K. & Nidetzky, B. Catalytic mechanism and substrate selectivity of aldo-keto reductases: Insights from structure-function studies of *Candida tenuis* xylose reductase. *IUBMB Life* **58**, 499–507 (2006).
 160. Schlegel, B. P., Jez, J. M. & Penning, T. M. Mutagenesis of 3 α -hydroxysteroid dehydrogenase reveals a ‘push -pull’ mechanism for proton transfer in aldo-keto reductases. *Biochemistry* **37**, 3538–3548 (1998).
 161. Bohren, K. M. *et al.* Tyrosine-48 Is the Proton Donor and Histidine-110 Directs Substrate Stereochemical Selectivity in the Reduction Reaction of Human Aldose Reductase: Enzyme Kinetics and Crystal Structure of the Y48H Mutant Enzyme. *Biochemistry* **33**, 2021–2032 (1994).
 162. Barski, O. A., Tipparaju, S. M. & Bhatnagar, A. The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. *Drug Metabolism Reviews* **40**, 553–624 (2008).
 163. Weng, J., Cao, Y., Moss, N. & Zhou, M. Modulation of voltage-dependent shaker family potassium channels by an aldo-keto reductase. *J. Biol. Chem.* **281**, 15194–15200 (2006).
 164. Machielsen, R., Uria, A. R., Serve, W., Kengen, M. & van der Oost, J. Production and Characterization of a Thermostable Alcohol Dehydrogenase That Belongs to the Aldo-Keto Reductase Superfamily. *Appl. Environ. Microbiol.* **72**, 233–238 (2006).
 165. Sanli, G. Structural alteration of cofactor specificity in *Corynebacterium* 2,5-diketo-D-gluconic acid reductase. *Protein Sci.* **13**, 504–512 (2004).
 166. Gefflaut, T., Blonski, C., Perie, J. & Willson, M. Class I aldolases: Substrate specificity, mechanism, inhibitors and structural aspects. *Progress in Biophysics and Molecular Biology* **63**, 301–340 (1995).
 167. Windle, C. L., Müller, M., Nelson, A. & Berry, A. Engineering aldolases as biocatalysts. *Current Opinion in Chemical Biology* **19**, 25–33 (2014).
 168. Schiff, H. Communications from the university laboratory in Pisa: 2. A new series of organic bases. *Ann. der Chemie und Pharm. (in Ger.)* **131**, 118–119 (1864).
 169. Kobes, R. D., Simpson, R. T., Vallee, B. L. & Rutter, W. J. A Functional Role of Metal Ions in A Class II Aldolase. *Biochemistry* **8**, 585–588 (1969).
 170. Fessner, W. D. *et al.* The mechanism of class II, metal-dependent aldolases. *Angew. Chemie (International Ed. English)* **35**, 2219–2221 (1996).
 171. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).
 172. Stincone, A. *et al.* The return of metabolism: Biochemistry and physiology of the pentose

- phosphate pathway. *Biol. Rev.* **90**, 927–963 (2015).
173. Horecker, B. L. The pentose phosphate pathway. *J. Biol. Chem.* **277**, 47965–47971 (2002).
 174. Keller, M. A., Turchyn, A. V. & Ralser, M. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.* **10**, (2014).
 175. Clasquin, M. F. *et al.* Riboneogenesis in yeast. *Cell* **145**, 969–980 (2011).
 176. Kneidinger, B., Graninger, M., Puchberger, M., Kosma, P. & Messner, P. Biosynthesis of Nucleotide-activated D-glycero-D-manno-Heptose. *J. Biol. Chem.* **276**, 20935–20944 (2001).
 177. Samland, A. K. & Sprenger, G. A. Transaldolase: From biochemistry to human disease. *International Journal of Biochemistry and Cell Biology* **41**, 1482–1494 (2009).
 178. Caillau, M. & Quick, W. P. New insights into plant transaldolase. *Plant J.* **43**, 1–16 (2005).
 179. Moehs, C. P., Allen, P. V., Friedman, M. & Belknap, W. R. Cloning and expression of transaldolase from potato. *Plant Mol. Biol.* **32**, 447–452 (1996).
 180. Soderberg, T. & Alver, R. C. Transaldolase of *Methanocaldococcus jannaschii*. *Archaea* **1**, 255–262 (2004).
 181. Schürmann, M. Biochemische Charakterisierung und Struktur-Funktionsbeziehungen bakterieller Transaldolasen und Fruktose-6-Phosphat Aldolasen. *Dissertation* Universität Düsseldorf (2001).
 182. Schörken, U. *et al.* Identification of catalytically important residues in the active site of *Escherichia coli* transaldolase. *Eur. J. Biochem.* **268**, 2408–2415 (2001).
 183. Schörken, U., Jia, J., Sahm, H., Sprenger, G. A. & Schneider, G. Disruption of *Escherichia coli* transaldolase into catalytically active monomers: Evidence against half-of-the-sites mechanism. *FEBS Lett.* **441**, 247–250 (1998).
 184. Thorell, S., Schürmann, M., Sprenger, G. A. & Schneider, G. Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *J. Mol. Biol.* **319**, 161–171 (2002).
 185. Jia, J. *et al.* Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family. *Structure* **4**, 715–724 (1996).
 186. Littlechild, J. A. & Watson, H. C. A data-based reaction mechanism for type I fructose bisphosphate aldolase. *Trends Biochem. Sci.* **18**, 36–39 (1993).
 187. Jia, J., Schörken, U., Lindqvist, Y., Sprenger, G. A. & Schneider, G. Crystal structure of the reduced Schiff-base intermediate complex of transaldolase B from *Escherichia coli*: Mechanistic implications for class I aldolases. *Protein Sci.* **6**, 119–124 (1997).
 188. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
 189. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
 190. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).
 191. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by

- maximization of detail and connectivity. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 545–559 (2018).
192. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
 193. Light, S. H. & Anderson, W. F. Arabinose 5-phosphate covalently inhibits transaldolase. *J. Struct. Funct. Genomics* **15**, 41–44 (2014).
 194. Ogawa, T., Murakami, K. & Yoshino, M. Inhibition by fructose 1,6-bisphosphate of transaldolase from *Escherichia coli*. *FEMS Microbiol. Lett.* **363**, 183 (2016).
 195. Ciaccio, C., Coletta, A., De Sanctis, G., Marini, S. & Coletta, M. Cooperativity and allostery in haemoglobin function. *IUBMB Life* **60**, 112–123 (2008).
 196. Tsolas, O. & Horecker, B. L. TRANSALDOLASE: A MODEL FOR STUDIES OF ISOENZYMES AND HALF-SITE ENZYMES*. *Mol. Cell. Biochem.* **1**, 3–13 (1973).
 197. Wu, L. *et al.* An overview of activity-based probes for glycosidases. *Current Opinion in Chemical Biology* **53**, 25–36 (2019).
 198. Moreno, M. S., Schneider, B. L., Maile, R. R., Weyler, W. & Saier, M. H. Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: Novel modes of regulation revealed by whole-genome analyses. *Mol. Microbiol.* **39**, 1366–1381 (2001).
 199. Pérez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**, 1838–1847 (2000).
 200. Haydon, D. J. & Guest, J. R. A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.* **79**, 291–296 (1991).
 201. Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J. Biol. Chem.* **277**, 12507–12515 (2002).
 202. Gao, Y. R. *et al.* Structural analysis of the regulatory mechanism of MarR protein Rv2887 in *M. tuberculosis*. *Sci. Rep.* **7**, (2017).
 203. Gupta, S. S., Borin, B. N., Cover, T. L. & Krezel, A. M. Structural analysis of the DNA-binding domain of the *Helicobacter pylori* response regulator ArsR. *J. Biol. Chem.* **284**, 6536–6545 (2009).
 204. Liu, S., Widom, J., Kemp, C. W., Crews, C. M. & Clardy, J. Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science (80-.)*. **282**, 1324–1327 (1998).
 205. Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Current Opinion in Structural Biology* **10**, 110–116 (2000).
 206. Cornille, F. *et al.* DNA binding properties of a chemically synthesized DNA binding domain of hRFX1. *Nucleic Acids Res.* **26**, 2143–2149 (1998).
 207. Xu, Y., Heath, R. J., Li, Z., Rock, C. O. & White, S. W. The Fad-RDNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J. Biol. Chem.* **276**, 17373–17379 (2001).
 208. Ramos, J. L. *et al.* The TetR Family of Transcriptional Repressors. *Microbiol. Mol. Biol. Rev.* **69**, 326–356 (2005).
 209. Fillenberg, S. B., Friess, M. D., Korner, S., Böckmann, R. A. & Müller, Y. A. Crystal structures of the global regulator dasR from *Streptomyces coelicolor*: Implications for the allosteric





- regulation of GntR/HutC Repressors. *PLoS One* **11**, e0157691 (2016).
210. Aravind, L. & Anantharaman, V. HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. *FEMS Microbiol. Lett.* **222**, 17–23 (2003).
 211. Bender, R. A. Regulation of the Histidine Utilization (Hut) System in Bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 565–584 (2012).
 212. Molitor, C., Bijelic, A. & Rompel, A. The potential of hexatungstotellurate(VI) to induce a significant entropic gain during protein crystallization. *IUCr* **4**, 734–740 (2017).
 213. Vasiliou, V., Reuter, S. F., Williams, S., Puga, A. & Nebert, D. W. Mouse cytosolic class 3 aldehyde dehydrogenase (Aldh3a1): Gene structure and regulation of constitutive and dioxin-inducible expression. *Pharmacogenetics* **9**, 569–580 (1999).
 214. Liu, Z. J. *et al.* The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat. Struct. Biol.* **4**, 317–326 (1997).
 215. Shortall, K., Djeghader, A., Magner, E. & Soulimane, T. Insights into Aldehyde Dehydrogenase Enzymes: A Structural Perspective. *Frontiers in Molecular Biosciences* **8**, 410 (2021).
 216. Perozich, J. *et al.* Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. *Eur. J. Biochem.* **267**, 6197–6203 (2000).
 217. Liu, L. K. & Tanner, J. J. Crystal Structure of Aldehyde Dehydrogenase 16 Reveals Trans-Hierarchical Structural Similarity and a New Dimer. *J. Mol. Biol.* **431**, 524–541 (2019).
 218. Farres, J., Wang, T. T. Y., Cunningham, S. J. & Weiner, H. Investigation of the Active Site Cysteine Residue of Rat Liver Mitochondrial Aldehyde Dehydrogenase by Site-Directed Mutagenesis. *Biochemistry* **34**, 2592–2598 (1995).
 219. Sobreira, T. J. P. *et al.* Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 226–231 (2011).
 220. Pohl, E., Brunner, N., Wilmanns, M. & Hensel, R. The crystal structure of the allosteric non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeum *Thermoproteus tenax*. *J. Biol. Chem.* **277**, 19938–19945 (2002).

9. Publications Released as part of this work

The following pages list all publications released as part of this project, in the order in which they were published.



A Sulfoglycolytic Entner-Doudoroff Pathway in *Rhizobium leguminosarum* bv. trifolii SRDI565

Jinling Li,^{a,b} Ruwan Epa,^{a,b}  Nichollas E. Scott,^e Dominik Skoneczny,^{b,f} Mahima Sharma,^g Alexander J. D. Snow,^g James P. Lingford,^{c,d}  Ethan D. Goddard-Borger,^{c,d}  Gideon J. Davies,^g Malcolm J. McConville,^{b,f}  Spencer J. Williams^{a,b}

^aSchool of Chemistry, University of Melbourne, Parkville, Victoria, Australia

^bBio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia

^cACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

^dDepartment of Medical Biology, University of Melbourne, Parkville, Victoria, Australia

^eDepartment of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Parkville, Victoria, Australia

^fDepartment of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia

^gYork Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, United Kingdom

ABSTRACT Rhizobia are nitrogen-fixing bacteria that engage in symbiotic relationships with plant hosts but can also persist as free-living bacteria in the soil and rhizosphere. Here, we show that free-living *Rhizobium leguminosarum* SRDI565 can grow on the sulfosugar sulfoquinovose (SQ) or the related glycoside SQ-glycerol using a sulfoglycolytic Entner-Doudoroff (sulfo-ED) pathway, resulting in production of sulfolactate (SL) as the major metabolic end product. Comparative proteomics supports the involvement of a sulfo-ED operon encoding an ABC transporter, sulfo-ED enzymes, and an SL exporter. Consistent with an oligotrophic lifestyle, proteomics data revealed little change in expression of the sulfo-ED proteins during growth on SQ versus mannitol, a result confirmed through biochemical assay of sulfoquinovosidase activity in cell lysates. Metabolomics analysis showed that growth on SQ involves gluconeogenesis to satisfy metabolic requirements for glucose-6-phosphate and fructose-6-phosphate. Metabolomics analysis also revealed the unexpected production of small amounts of sulfofructose and 2,3-dihydroxypropanesulfonate, which are proposed to arise from promiscuous activities of the glycolytic enzyme phosphoglucose isomerase and a nonspecific aldehyde reductase, respectively. The discovery of a rhizobium isolate with the ability to degrade SQ builds our knowledge of how these important symbiotic bacteria persist within soil.

IMPORTANCE Sulfonate sulfur is a major form of organic sulfur in soils but requires biomineralization before it can be utilized by plants. Very little is known about the biochemical processes used to mobilize sulfonate sulfur. We show that a rhizobial isolate from soil, *Rhizobium leguminosarum* SRDI565, possesses the ability to degrade the abundant phototroph-derived carbohydrate sulfonate SQ through a sulfoglycolytic Entner-Doudoroff pathway. Proteomics and metabolomics demonstrated the utilization of this pathway during growth on SQ and provided evidence for gluconeogenesis. Unexpectedly, off-cycle sulfoglycolytic species were also detected, pointing to the complexity of metabolic processes within cells under conditions of sulfoglycolysis. Thus, rhizobial metabolism of the abundant sulfosugar SQ may contribute to persistence of the bacteria in the soil and to mobilization of sulfur in the pedosphere.

KEYWORDS sulfoglycolysis, metabolomics, sulfur cycle, rhizobia, carbohydrates, metabolism

Sulfur is essential for plant growth and is the fourth most important macronutrient after nitrogen, phosphorus, and potassium. Up to 10 kg/ha/year of sulfur is deposited in rain, especially near industrialized areas (1). However, sulfur dioxide emissions

Citation Li J, Epa R, Scott NE, Skoneczny D, Sharma M, Snow AJD, Lingford JP, Goddard-Borger ED, Davies GJ, McConville MJ, Williams SJ. 2020. A sulfoglycolytic Entner-Doudoroff pathway in *Rhizobium leguminosarum* bv. trifolii SRDI565. *Appl Environ Microbiol* 86:e00750-20. <https://doi.org/10.1128/AEM.00750-20>.

Editor Emma R. Master, University of Toronto

Copyright © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Spencer J. Williams, sjwill@unimelb.edu.au.

Received 30 March 2020

Accepted 14 May 2020

Accepted manuscript posted online 22 May 2020

Published 20 July 2020

from industrial sources have decreased in recent decades as a result of pollution mitigation and the move to low-sulfur fuels and renewable energy sources, and quantities received from atmospheric sources are now at levels below that required by most crops (2). Sulfur deficiency in soils is primarily combated by application of sulfur-containing fertilizers, such as superphosphate, ammonium sulfate, and gypsum (3), which are applied across all major crop-growing and pasture areas worldwide (4). Soils contain significant amount of sulfur, yet plants can use sulfur only in the form of sulfate, and it has been shown that 95 to 98% of sulfur in soils is in the form of unavailable biological sulfur (4). Thus, effective microbial cycling of sulfur from biological to inorganic forms within the soil is important (5) and has the potential to enhance crop yields and reduce reliance on fertilizers.

X-ray absorption near-edge spectroscopy measurements have led to estimates that approximately 40% of sulfur within various sediments and humic substances exists as sulfonate (6). Chemical methods of analysis applied to a range of forest soils revealed that sulfonate sulfur accounted for 40% of the total organic sulfur pool in the majority of cases (7). Little detail is known about the speciation of organic sulfonates in soils, but one important input is phototroph-derived litter. It is estimated that around 10 billion tonnes of the sulfosugar sulfoquinovose (SQ) is produced annually by photosynthetic organisms, including plants, cyanobacteria, and algae (8). SQ is primarily found as the glycerolipid sulfoquinovosyl diacylglycerol (SQDG), and land plants can contain as much as 10% SQDG in their thylakoid membrane glycerolipids (9). Very little is known about how SQ is metabolized within soils, although it has been shown to undergo very rapid mineralization to inorganic sulfate (10).

Bacteria are likely to be primarily responsible for the biomineralization of SQ, possibly by using SQ as a carbon source and catabolizing it via a modified version of glycolysis, termed sulfoglycolysis (11). Two sulfoglycolytic processes have been described: the sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP) pathway (12), and the sulfoglycolytic Entner-Doudoroff (sulfo-ED) pathway (Fig. 1) (13). The sulfo-ED pathway was first reported in *Pseudomonas putida* strain SQ1, a bacterium isolated from freshwater sediment that catabolizes SQ with excretion of equimolar amounts of sulfolactate (SL) (13). The sulfo-ED operon of *P. putida* SQ1 contains 10 genes, including those corresponding to a transcriptional regulator, an SQ importer and SL exporter, a sulfoquinovosidase, SQ mutarotase, SQ dehydrogenase, SL lactonase, SG dehydratase, KDSG (2-keto-3,6-dideoxy-6-sulfogluconate) aldolase, and SLA (3-sulfolactaldehyde) dehydrogenase. Based on genome-wide annotation studies, the sulfo-ED pathway is predicted to occur in a range of alpha-, beta-, and gammaproteobacteria (13). However, no direct evidence for this pathway has been reported for any organism other than *P. putida* SQ1. Other members of the microbial community can catabolize SL and 2,3-dihydroxypropanesulfonate (DHPS; the product of the sulfo-EMP pathway) to inorganic sulfur (14), completing the biomineralization of SQ.

Rhizobium leguminosarum bv. *trifolii* SRDI565 (syn. N8-J) was isolated from a soil sample collected in western New South Wales, Australia, but has the capacity to colonize *Trifolium subterraneum* subsp. *subterraneum* (subterranean clover) and other *Trifolium* spp. (15). Colonization of *Trifolium* species with SRDI565 results in suboptimal nodulation and nitrogen fixation in some species and ineffective nitrogen fixation in others, leading to reduced shoot nitrogen content relative to other commercial strains (16). Interestingly, the genome of SRDI565 contains all the genes needed for a functional sulfo-ED pathway (17), although there is no evidence to show that this is operative and/or that SRDI565 can use SQ as a major carbon source.

Rhizobia participate in sophisticated symbiotic relationships with leguminous host plants that allow them to fix atmospheric dinitrogen to provide a growth advantage to the host (18). Symbiosis is triggered by molecular communication between the bacterium and the host, resulting in nodule formation on the root and colonization by the bacterium. Within nodule bacteroids, the energy-intensive fixation of nitrogen is supported by C₄-dicarboxylates (primarily malate, fumarate, and succinate) obtained from glycolysis of sucrose photosynthate within the plant host (18). Owing to the importance

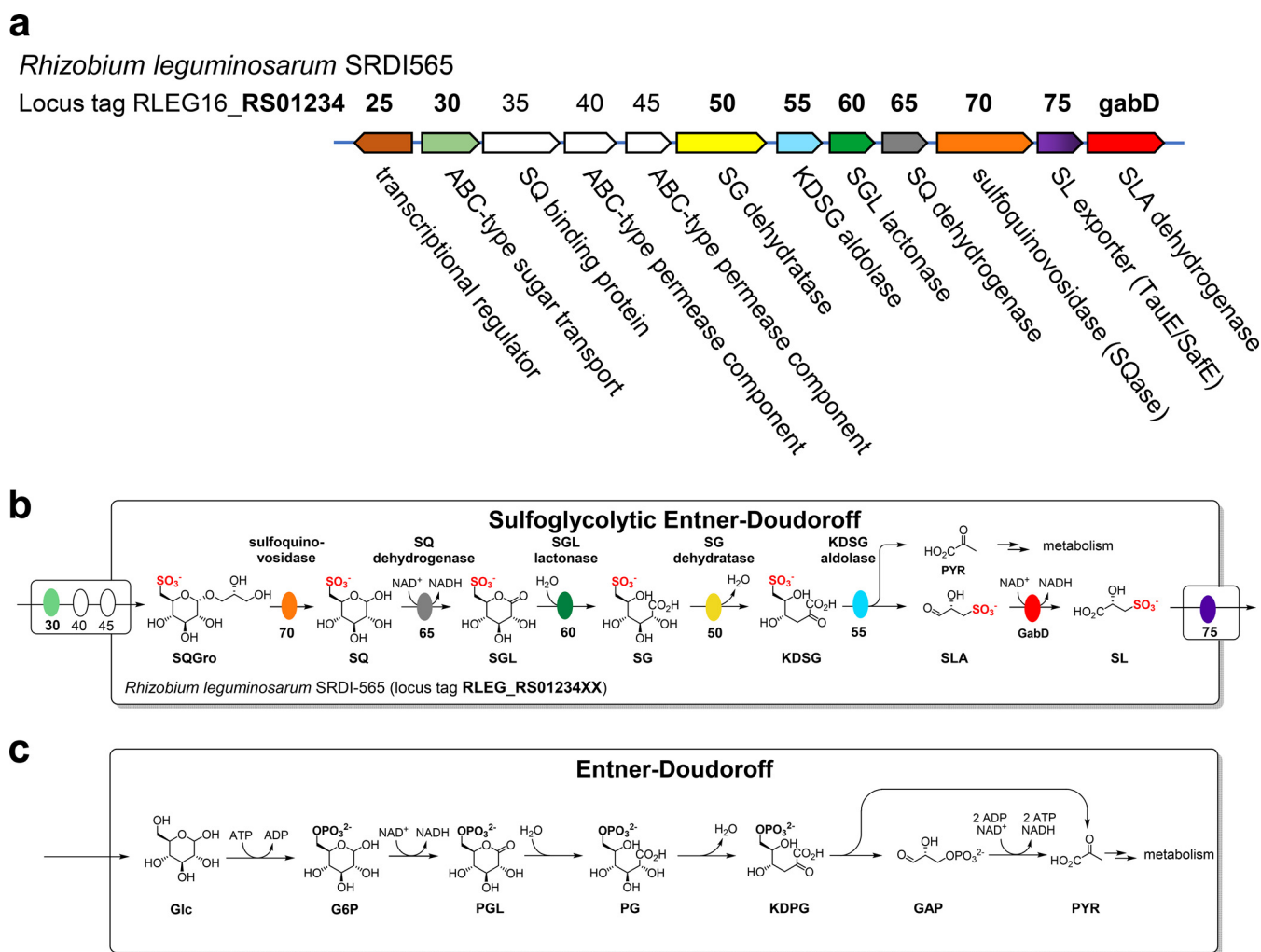


FIG 1 Proposed sulfoglycolytic Entner-Doudoroff (sulfo-ED) pathway in *Rhizobium leguminosarum* bv. trifolii SRDI565. (a) Operon encoding the sulfo-ED pathway. (b) Proposed sulfo-ED pathway. (c) Comparison with the Entner-Doudoroff pathway.

of biological nitrogen fixation for input of nitrogen into the biosphere, the symbiosis of rhizobia and leguminous hosts has been well studied. However, rhizobia can also exist as free-living bacteria within the soil and rhizosphere (19). Here, like other soil bacteria, they adopt a saprophytic and oligotrophic lifestyle, where they utilize a variety of alternative carbon sources, including a wide range of carbohydrates (20). Most likely, the ability of various rhizobia to persist in the pedosphere depends upon their ability to utilize diverse carbohydrate and noncarbohydrate substrates and establish an appropriate niche. SQ or its glycosides are likely to be a common soil constituent and nutrient, given their ubiquitous production by plants. Possibly, the sulfo-ED pathway in SRDI565 might provide it with the capacity to survive on plant-derived SQ or SQDG in the rhizosphere and in the soil.

Here, we investigated whether the sulfo-ED pathway is active in SRDI565 and its potential role in utilizing plant-derived SQ or SQDG in the rhizosphere and in the soil. We show that SRDI565 can grow on SQ and sulfoquinovosyl glycerol (SQGro) as the sole carbon source. Growth on SQ leads to excretion of SL into the growth medium, indicating active sulfoglycolysis. This was supported by proteomic analyses, which showed that several proteins encoded by the sulfo-ED operon exhibit increased expression when bacteria are grown on SQ, while metabolomic analyses confirmed the presence of characteristic intermediates of the sulfo-ED pathway, as well as the unexpected production of intracellular DHPS. Overall, we show that SRDI565 has an

active pathway for SQ utilization which may support growth of this bacterium in the environment; thus, this strain constitutes a new model organism for the study of the sulfo-ED pathway.

RESULTS

Analysis of the genome of SRDI565 revealed a sulfo-ED operon that had the same genes as but no synteny with the *P. putida* SQ1 operon (Fig. 1). Genes with high sequence identity to the *P. putida* genes included those encoding a putative SQase, SQ dehydrogenase, SL lactonase, SG dehydratase, KDSG aldolase, and SLA dehydrogenase and an SL exporter (see Fig. S1 to S6 in the supplemental material). The SRDI565 operon contains some important differences compared to that of *P. putida* SQ1. In particular, it lacks a putative SQ mutarotase gene (21) and appears to use an ABC transporter to import SQ/SQGro in place of an SQ/SQGro importer/permease. The putative sulfo-ED pathway in SRDI565 is consistent with the proposed protein functions outlined in Fig. 1b, with a comparison to the classical ED pathway in Fig. 1c.

Initial attempts were made to grow SRDI565 in completely defined medium, such as M9 minimal medium containing $125 \mu\text{g ml}^{-1}$ biotin (22), to allow assessment of the effects of different carbon sources on bacterial growth. However, optimal growth could be achieved only by using a yeast extract-based medium (16). In particular, robust growth was achieved using a 5% dilution of 1 g liter^{-1} yeast extract ($Y_{5\%}$ medium) containing 5 mmol mannitol ($Y_{5\%}\text{M}$), while no detectable bacterial growth was observed on $Y_{5\%}$ medium alone. Significantly, SRDI565 also grew robustly on $Y_{5\%}$ medium containing 5 mM SQ ($Y_{5\%}\text{SQ}$) and reached the same final optical density at 600 nm (OD_{600}) as in $Y_{5\%}\text{M}$ (Fig. 2a). SRDI565 also grew on $Y_{5\%}$ medium containing glucose, although to a lower final OD_{600} than in $Y_{5\%}\text{M}$ or $Y_{5\%}\text{SQ}$. ^{13}C nuclear magnetic resonance (NMR) spectroscopic analysis of the culture medium of stationary-phase SRDI565 grown in $Y_{5\%}\text{SQ}$ revealed the presence of three major signals corresponding to SL (Fig. 2b). A fourth signal was also observed but not assigned and was also present in stationary-phase medium of cells grown on $Y_{5\%}\text{M}$, suggesting that it is derived from other carbon sources in the yeast extract. SRDI565 also grew on $Y_{5\%}$ containing SQGro, but less robustly than on SQ.

We next examined changes in the proteome of SRDI565 cultivated on mannitol versus SQ. Label-free quantitative proteomic analysis of five experimental replicates of SRDI565 cultivated on each carbon source identified 2,954 proteins, with 1,943 proteins quantified in at least 3 experimental replicates under each growth condition (Table S1). Expression levels of 17 proteins potentially associated with SQ metabolism were significantly elevated [$-\log_{10}(P)$ of >2 and a fold change greater than $2 \log_2$ units] in bacteria cultivated in $Y_{5\%}\text{SQ}$ (Fig. 2e and f). In particular, a suspected KDSG aldolase (annotated as alpha-dehydro-beta-deoxy-D-glucarate aldolase; WP_017967308.1), a member of the proposed sulfo-ED pathway, was significantly increased [$-\log_{10}(P)$ of 4.74429 and a fold change of $2.38 \log_2$]. Consistent with the involvement of this pathway, we also observed a significant yet less dramatic increase in the proposed SQase (annotated alpha-glucosidase; WP_017967311.1) [$-\log_{10}(P)$ of 1.43643 and a fold change of $1.02 \log_2$]. Additional members of the predicted pathway expressed at higher levels in SQ-fed bacteria included the suspected SQ dehydrogenase (annotated as SDR family oxidoreductase; WP_017967310.1), identified by tandem mass spectrometry (MS/MS) events in 4 of 6 SQ experiments compared to 1 mannitol experiment, and the suspected SG dehydratase (annotated as dihydroxy-acid dehydratase, WP_017967307.1), identified by MS/MS events in 3 of 6 SQ experiments compared to 0 mannitol experiments; however, owing to their low abundance, they could not be accurately quantified (Fig. S7).

Other proteins that were significantly increased in SQ-fed bacteria included a NAD(P)-dependent oxidoreductase (WP_017965793.1), the NADH-quinone oxidoreductase subunit NuoH (WP_017963854.1), a NAD-dependent succinate-semialdehyde dehydrogenase (WP_017967313.1), and a citrate synthase/methylcitrate synthase

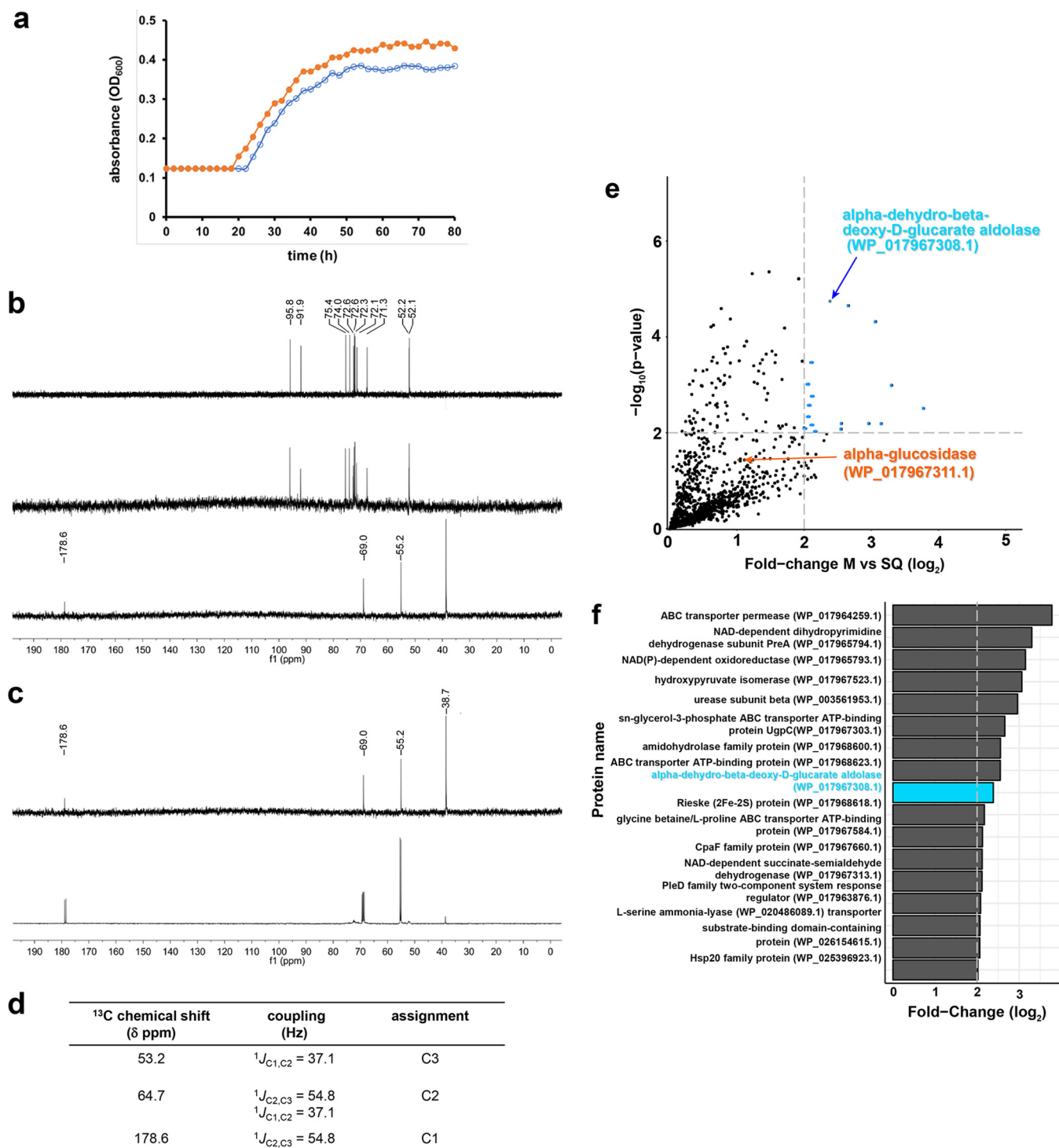


FIG 2 Growth of *Rhizobium leguminosarum* bv. trifolii SRDI565 on SQ produces SL as the major terminal metabolite. (a) Growth of SRDI565 on 5% yeast extract medium containing 5 mM SQ (solid circles) or 5 mM mannitol (open circles). The data are representative of 2 independent experiments. (b) ¹³C NMR (126 MHz) spectra of SQ (top), 5 mM SQ in 5% yeast extract medium (middle), and spent culture medium from growth of SRDI565 on 5 mM SQ (bottom). (c) ¹³C NMR (126 MHz) spectrum of spent culture medium from growth of SRDI565 on 5 mM [¹³C₆]SQ. The signal at δ 38.7 ppm is present in control experiments with SRDI565 grown on mannitol and is believed to derive from yeast extract. (d) Tabulated ¹³C NMR (126 MHz) data for [¹³C₃]SL from panel c. All samples contain 10% D₂O, added to allow frequency lock. (e) Quantitative proteomics was undertaken to identify proteins associated with sulfoquinovose catabolism versus mannitol. Examination of proteins observed to increase in abundance more than 4-fold revealed 17 proteins, including alpha-dehydro-beta-deoxy-D-glucarate aldolase (WP_017967308.1), highlighted in blue. (f) Growth in sulfoquinovose leads to the increase of multiple proteins associated with the TCA cycle, including NAD(P)-dependent oxidoreductase (WP_017965793.1), NADH-quinone oxidoreductase subunit NuoH (WP_017963854.1), NAD-dependent succinate-semialdehyde dehydrogenase (WP_017967313.1), and citrate synthase/methylcitrate synthase (WP_017964386.1).

(WP_017964386.1), supporting an alteration of the tricarboxylic acid (TCA) cycle and oxidative phosphorylation under conditions of growth on SQ (Fig. 2f).

To demonstrate activity for a representative sulfo-ED enzyme from SRDI565, we cloned and expressed the gene encoding the putative SQase. To support future structural studies, we expressed the N-terminal hexahistidine-tagged K375A/K376A variant, termed *RISQase**, a mutant enzyme whose design was guided by the Surface Entropy Reduction prediction (SERP) server (Fig. S8) (23). Size exclusion chromatography–multiple-angle light scattering (SEC-MALS) analysis of *RISQase** revealed that the protein exists as a dimer in solution (Fig. S8). Enzyme kinetics were analyzed using the chromogenic SQase substrate 4-nitrophenyl α -sulfoquinovoside (PNPSQ). PNPSQ was designed as an analogue of the natural substrate SQGro, and its hydrolysis results in release of the chromophore 4-nitrophenolate, which can be detected using UV-visible spectrophotometry with high sensitivity at 400 nm or at the isosbestic point, 348 nm (24, 25). *RISQase** exhibited a bell-shaped pH profile with an optimum at pH 7 to 8 and consistent with titration of catalytically important residues with a pK_{a1} value of 6.5 ± 0.4 and a pK_{a2} value of 8.6 ± 0.3 . The enzyme displayed saturation kinetics with Michaelis-Menten parameters: $k_{cat} = 1.08 \pm 0.17 \text{ s}^{-1}$, $K_m = 0.68 \pm 0.25 \text{ mM}$, and $k_{cat}/K_m = (1.59 \pm 0.83) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 3a and b). For comparison, the kinetic parameters for *Agrobacterium tumefaciens* SQase are as follows: $k_{cat} = 22.3 \pm 0.6 \text{ s}^{-1}$, $K_m = 0.21 \pm 0.03 \text{ mM}$, and $k_{cat}/K_m = (1.1 \pm 0.1) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$. Those for *Escherichia coli* SQase YihQ are as follows: $k_{cat} = 32.7 \pm 0.6 \text{ s}^{-1}$, $K_m = 0.15 \pm 0.01 \text{ mM}$, and $k_{cat}/K_m = (2.2 \pm 0.2) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ (24).

Direct evidence for enzymatic activity associated with the sulfo-ED operon in SRDI565 was obtained by measuring SQase enzyme activity in cell lysates. SRDI565 was grown to mid-logarithmic phase in $Y_{5\%}M$ and $Y_{5\%}SQ$ media, and the harvested cells were used to prepare a cell-free lysate containing soluble proteins. Incubation of both $Y_{5\%}M$ - and $Y_{5\%}SQ$ -derived lysates with PNPSQ resulted in production of 4-nitrophenolate at similar rates. The activity in the $Y_{5\%}SQ$ -derived lysate was inhibited by the addition of isofagomine-SQ (IFGSQ), an azasugar inhibitor of SQases that carries out key interactions in the active site that mimic those required for substrate recognition (Fig. 3c) (24). The similar levels of activity of SQase in both mannitol- and SQ-grown SRDI565 are consistent with the abundance of the putative SQase WP_017967311.1 detected by proteomic analysis.

To further confirm that a sulfo-ED pathway was operative in cells, a targeted metabolomics approach was used to detect expected intermediates in bacteria grown on $Y_{5\%}SQ$ medium. Detected intermediates were identified based on their liquid chromatography-MS/MS (LC-MS/MS) retention time and mass spectra with authentic reference standards of the sulfo-EMP and sulfo-ED pathway that were synthesized in-house. Sulfoluconate (SG) was synthesized by oxidation of SQ with iodine (26) (Fig. S9), while SQ, SF, sulfofructose-1-phosphate (SFP), DHPS, SLA, and SL were prepared as previously reported (27). SRDI565 was grown to mid-log phase in $Y_{5\%}M$ or $Y_{5\%}SQ$ and metabolically quenched, and extracted polar metabolites were analyzed by LC-MS/MS. SQ-grown bacteria contained SQ, SF, SG, SL, and DHPS, while SFP and SLA could not be detected (Fig. 4a to e). The detection of SG is characteristic of a sulfo-ED pathway and presumably arises from the action of the putative SQ dehydrogenase and SGL lactonase. The identification of DHPS and SF was unexpected, as these are intermediates or products of the sulfo-EMP pathway (12). BLAST analysis of the genome of SRDI565 did not identify putative genes for the sulfo-EMP pathway. SF may therefore be formed by the action of phosphoglucose isomerase (PGI), while DHPS could be the product of a promiscuous aldehyde reductase. SRDI565 was unable to utilize DHPS or SL as the sole carbon source in $Y_{5\%}$ medium, supporting the absence of an alternative pathway of sulfoglycolysis that utilizes these intermediates. Unexpectedly, cytosolic levels of DHPS were 20-fold higher than SL, suggesting that cells may lack a membrane transporter to export accumulated DHPS, in contrast to the SL transporter.

NMR and LC-MS/MS analysis of the culture supernatant of both unlabeled and $^{13}C_6$ -labeled SQ-cultivated SRDI565 confirmed that the substrate is almost completely consumed by the time bacteria reach stationary growth (final concentration

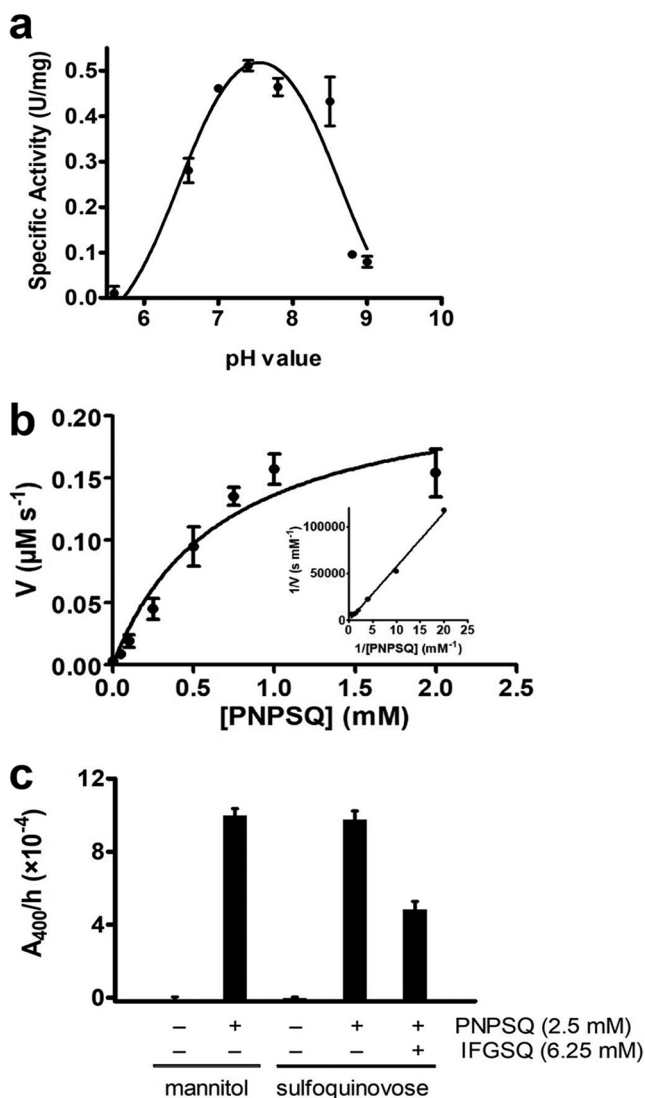
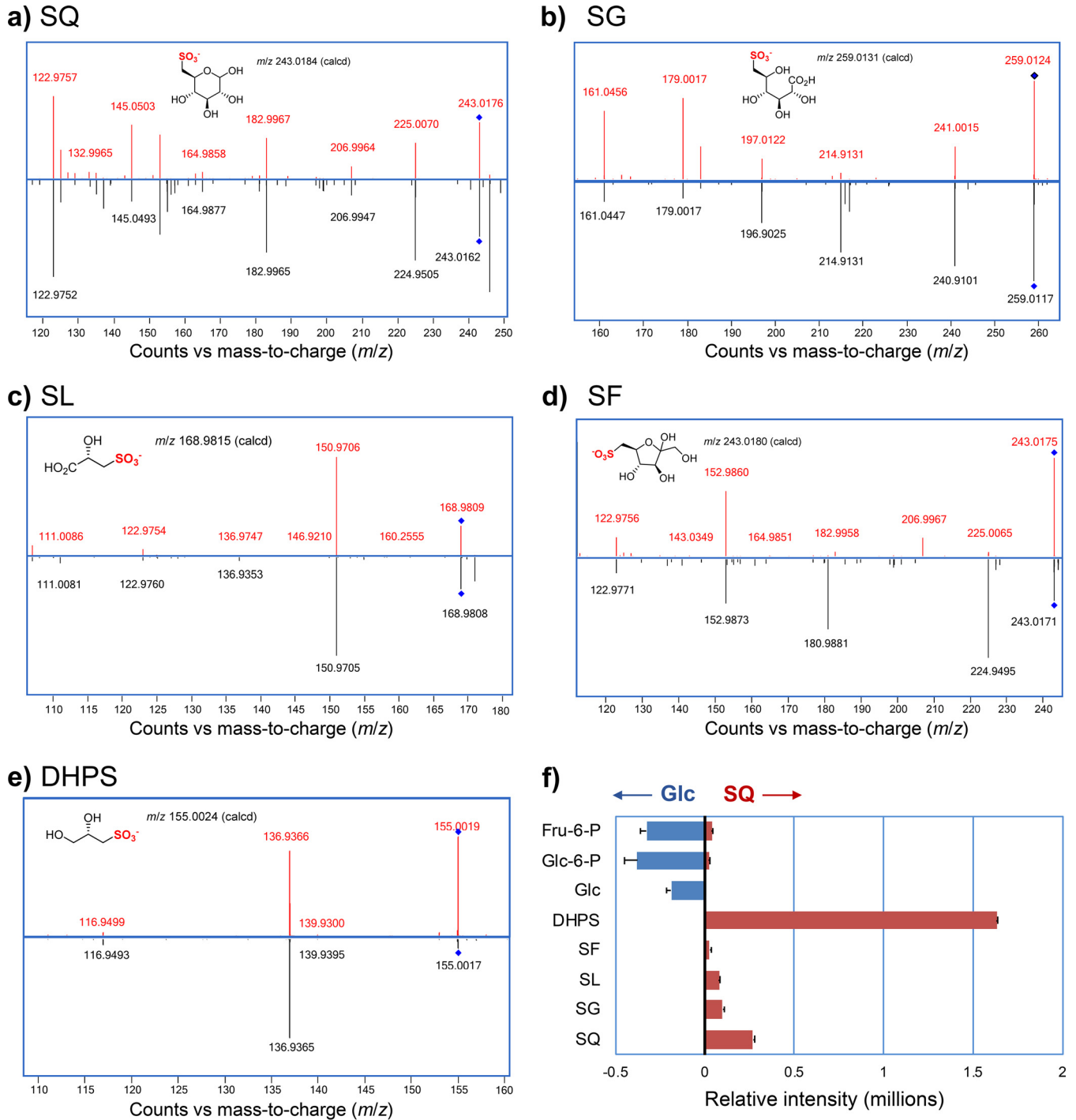


FIG 3 *Rhizobium leguminosarum* SRDI565 produces a functional sulfoquinovosidase that can be detected in cell lysates. (a) pH profile of *RISOase**. Specific activities were determined for hydrolysis of PNPSQ at the isobestic point, 348 nm. (b) Michaelis-Menten plot of kinetic parameters for *RISOase** for hydrolysis of PNPSQ at 400 nm. (c) Analysis of sulfoquinovosidase activity of SRDI565 lysate grown on sulfoquinovose and mannitol. Cell lysates of soluble proteins derived from growth on SQ or mannitol was standardized for equal protein and SQase activity, measured using the chromogenic substrate PNPSQ at 400 nm. SQase activity was confirmed by inhibition by the azasugar inhibitor IFGSQ. Error bars denote standard errors of the means.

of 0.006 ± 0.001 mM, compared to 5.0 ± 0.5 mM SQ in the starting medium) (Fig. S10). When a highly sensitive cryoprobe was used, ^{13}C NMR spectroscopic analysis revealed that both DHPS and SG were present in culture supernatant of $[^{13}\text{C}_6]$ SQ-cultivated SRDI565. Quantitative LC-MS/MS analysis showed that consumption of SQ was associated with production of SL (5.70 ± 0.12 mM) and low levels of DHPS (0.081 ± 0.010 mM), SG (0.172 ± 0.006 mM), and SF (0.002 ± 0.0001 mM) (Table 1). This experiment was repeated to assess the effect of growth of SRDI565 but with SQGro as the carbon source. As noted previously, SRDI565 grows inconsistently on SQGro, and complete consumption of SQGro could not be achieved. However, the results of partial consumption broadly agreed with the results for growth on SQ, namely, that SL is the major terminal metabolite detected in the culture medium, with much smaller amounts of SF, SG, and DHPS (Table 1).



Downloaded from <http://aem.asm.org/> on December 4, 2020 at UNIV OF YORK

FIG 4 Detection of sulfolglycolytic intermediates and end products in cytosolic extracts of SRDI565. SRDI565 was grown on $Y_{5\%}$ SQ medium and metabolically quenched by rapid cooling to 4°C, followed by extraction of cellular metabolites and LC-MS analysis. Sulfolglycolytic and glycolytic/neoglucogenic intermediates SQ (a), SG (b), SL (c), SF (d), and DHPS (e) were detected. In each panel, the upper portion corresponds to the collision-induced dissociation mass spectrum of chemically synthesized standard, while the lower portion is the equivalent mass spectrum for the metabolite identified in the cytosolic extract. (f) Relative mass spectrometric intensities of metabolites from cells grown on Glc or SQ.

DISCUSSION

We demonstrate here that SRDI565 has a functional sulfo-ED pathway that allows this bacterium to utilize SQ as its major carbon source. Catabolism of SQ is primarily or exclusively mediated by a sulfo-ED pathway, with production of SL as the major end product, similar to the situation in *P. putida* SQ1, the only other experimentally

TABLE 1 Analysis of sulfonate metabolites detected in spent culture medium of SRDI565 grown on 5.0 ± 0.5 mM SQ or SQGro

Metabolite	Metabolite concn (mM) ^a after growth on:	
	SQ	SQGro ^b
SL	5.70 \pm 0.12	3.14 \pm 0.03
DHPS	0.081 \pm 0.010	0.116 \pm 0.002
SQ	0.006 \pm 0.001	0.215 \pm 0.001
SF	0.002 \pm 0.0001	0.003 \pm 0.0001
SG	0.172 \pm 0.006	0.200 \pm 0.008
SQGro		2.16 \pm 0.06

^aMeasurements were performed in triplicate using LC-MS/MS. Values are means \pm standard errors of the means (standard error estimate).

^bGrowth on SQGro was incomplete.

described exemplar of this pathway (13). In contrast to *P. putida* SQ1, SRDI565 also produces trace amounts of DHPS, which could reflect the presence of enzymes which exhibit promiscuous activities similar to those in the conventional sulfo-EMP pathway. This observation is reminiscent of *Klebsiella* sp. strain ABR11, isolated from soil (28), which is also able to grow on SQ with production of both SL and DHPS. *Klebsiella* sp. strain ABR11 possesses an NAD⁺-specific sulfoquinovose-dehydrogenase activity (29), suggesting that it has an operative sulfo-ED pathway.

Various bacteria that can metabolize SQ have been isolated from soil, including *Agrobacterium* sp. (29), *Klebsiella* sp. (29), and *Flavobacterium* sp. (30), as well as *P. putida* SQ1 (13), which was isolated from a freshwater littoral sediment. These bacteria may work cooperatively with organisms, such as *Paracoccus pantotrophus* NKNCYSA, that can convert SL to mineral sulfur, leading to stoichiometric recovery of sulfite/sulfate (14). Together these bacterial communities achieve the complete mineralization of SQ to sulfate, which is available for use by plants.

Proteomic and biochemical evidence suggests that the sulfo-ED pathway is constitutively expressed in SRDI565, with only relatively small increases in protein expression, as shown by statistically significant increases in only KDSG aldolase and SQase in the presence of SQ. As SRDI565 in the soil is likely to be oligotrophic, constitutive expression of the sulfo-ED pathway may allow simultaneous usage of multiple nonglycolytic substrates without requirement for significant transcriptional changes. Consistent with this view, the proteomic abundance of the putative substrate-binding domain-containing protein LacI-type regulator WP_157386381.1 was unchanged between mannitol- and SQ-grown SRDI565. The sulfo-ED operon in SRDI565 differs from that described for *P. putida* SQ1 through the absence of a putative SQ mutarotase. SQ undergoes mutarotation with a half-life of approximately 6 h, which is much longer than that of the glycolytic intermediate Glc-6-P, which has a half-life of just seconds (21). Aldose mutarotases are often relatively nonspecific, and it is possible that a constitutive mutarotase not encoded in the sulfo-ED operon expressed by the cell provides this catalytic capacity. Alternatively, the SQ dehydrogenase may not be stereospecific, with the ability to act on both anomers of SQ, or it may even act on α -SQ (the product released from SQGro by an SQase) at a high rate, such that mutarotation to β -SQ is insignificant. A second difference in the sulfo-ED operon is the presence of an ABC transporter gene. ABC transporters are the most common solute transporters and can translocate their substrates in either a forward or reverse direction (31). While we propose that the ABC transporter operates in the forward direction, based on the presence of a signal sequence in the putative solute binding domain targeting it to the periplasm and consistent with a wide range of sugar import systems, the directionality of transport and thus the choice of substrate (SQ/SQGro versus SL) may depend on the relative abundance of these metabolites intra- and extracellularly.

Sulfoglycolysis in SRDI565 leads to production of pyruvate and the excretion of the C₃-organosulfonate SL (Fig. 5). In order to satisfy the demands of the pentose phosphate pathway and cell wall biogenesis, sulfoglycolytic cells must synthesize glucose-

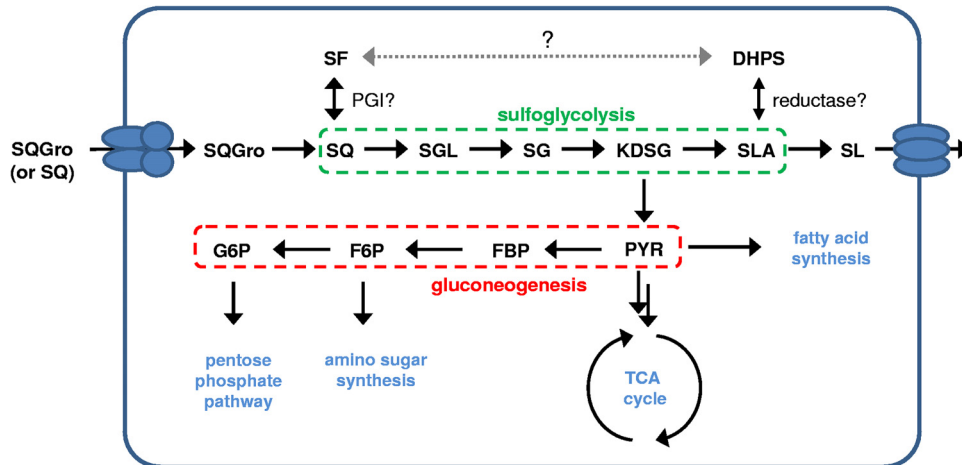


FIG 5 Proposed pathway for SQ metabolism in *Rhizobium leguminosarum* SLD1565.

based metabolites, such as glucose-6-phosphate and glucose-1-phosphate. Gluconeogenesis has been studied in *Rhizobium leguminosarum* strain MNF3841 and operates through a classical pathway involving fructose biphosphate aldolase (32). Action of phosphoglucose isomerase on SQ might lead to production of SF, thereby explaining the observation of this metabolite in SLD1565. This is not likely to be consequential, as the reversibility of this reaction should allow complete consumption of any SF through isomerization back to SQ. The formation of DHPS may result from a promiscuous aldehyde reductase. Analysis of spent culture medium reveals that the production of DHPS is minor in terms of total carbon balance. However, within the cytosol, DHPS accumulates to much higher levels than SL, presumably because of the absence of a dedicated exporter for the former. Possibly, reduction of SLA to DHPS is reversible and enables conversion of this metabolite to SL and subsequent excretion from the cell. The observation of SG, SF, and DHPS in the spent culture medium at low levels is suggestive of low levels of leakage of these metabolites from the cell, through either cell lysis or leaky export systems.

Given that SQ contains a significant portion of organic sulfur within plants, the pathways of SQ catabolism leading to release of its sulfur may be important to enable recycling of this important macronutrient. Plants can use only sulfate, which is poorly retained by most soils. Biomineralization of organic sulfur to sulfate is important to allow plants to access this element. As one of just two known pathways for the catabolism of SQ, the sulfo-ED pathway is likely to be an important part of environmental breakdown of SQ and may contribute to the persistence of symbiotic rhizobia within the pedosphere. The present work lays the groundwork for a more detailed investigation of sulfoglycolysis in a well-characterized bacterium with an established capability for symbiosis with a leguminous plant host.

MATERIALS AND METHODS

Reagents. SQ, [$^{13}\text{C}_6$]SQ, SF, SFP, SLA, SL, and DHPS were chemically and chemoenzymatically synthesized as described previously (27). IFGSQ was chemically synthesized as described previously (24).

Bacteria and culture conditions. *Rhizobium leguminosarum* bv. trifolii SLD1565 was a gift from Ross Ballard (South Australian Research and Development Institute, Adelaide, South Australia, Australia). Minimal salts medium consists of 0.5 g liter $^{-1}$ K $_2$ HPO $_4$, 0.2 g liter $^{-1}$ MgSO $_4$, 0.1 g liter $^{-1}$ NaCl, and 0.33 g liter $^{-1}$ CaCl $_2$, adjusted to pH 7.0. Y $_{59\%}$ M consists of minimal salts medium plus 50 mg liter $^{-1}$ yeast extract and 5 mM mannitol. Y $_{59\%}$ SQ consists of minimal salts medium plus 50 mg liter $^{-1}$ yeast extract and 5 mM SQ.

Growth curves were determined in a MicrobeMeter built in-house according to published plans (33) and blueprints available at <https://humanetechnologies.co.uk/download-microbemeter/>. The MicrobeMeter was calibrated by performing serial 2-fold dilutions across the detection range of the MicrobeMeter (0 to 1,023 U), starting with a culture of SLD1565 with an OD $_{600}$ of approximately 1. OD $_{600}$ measurements were made with a UV-visible spectrophotometer and plotted against the reading of the MicrobeMeter. The data were fitted to a polynomial to obtain a calibration curve.

(i) Proteomic sample preparation. Cells were washed 3 times in phosphate-buffered saline (PBS), collected by centrifugation at $10,000 \times g$ at 4°C , and then snap-frozen. Frozen whole-cell samples were resuspended in 4% sodium dodecyl sulfate (SDS)–100 mM Tris (pH 8.0)–20 mM dithiothreitol (DTT) and boiled at 95°C with shaking at 2,000 rpm for 10 min. Samples were then clarified by centrifugation at $17,000 \times g$ for 10 min, the supernatant was collected, and protein concentration was determined using a bicinchoninic acid assay (Thermo Scientific Pierce). One hundred micrograms of protein from each sample was cleaned using SP3-based purification according to previously published protocols (34). Briefly, reduced samples were cooled and then alkylated with 40 mM 2-chloroacetamide (CAA) for 1 h at room temperature in the dark. The alkylation reactions were quenched with 40 mM DTT for 10 min, and then samples were precipitated onto SeraMag Speed Beads (GE Healthcare, USA) with ethanol (final concentration, 50% [vol/vol]). Samples were shaken for 10 min to allow complete precipitation onto beads and then washed three times with 80% ethanol. The precipitated-protein-covered beads were resuspended in 100 mM ammonium bicarbonate containing $2 \mu\text{g}$ trypsin (1/50 [wt/wt]) and allowed to digest overnight at 37°C . Upon completion of the digests, samples were centrifuged at $14,000 \times g$ for 5 min to pellet the beads, and the supernatant was collected and desalted using homemade C_{18} stage tips (35). The eluted material was dried and stored until LC-MS analysis.

(ii) Proteomics analysis using reversed-phase LC-MS. Purified peptides prepared were resuspended in buffer A* (2% acetonitrile [ACN], 0.1% $\text{CF}_3\text{CO}_2\text{H}$) and separated using a two-column chromatography setup composed of a PepMap100 C_{18} 20-mm by $75\text{-}\mu\text{m}$ trap and a PepMap C_{18} 500-mm by $75\text{-}\mu\text{m}$ analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at $5 \mu\text{l}/\text{min}$ for 5 min and infused into an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). Gradients of 120 min were run in which the buffer composition was altered from 1% buffer B (80% ACN, 0.1% formic acid) to 28% B over 90 min, from 28% B to 40% B over 10 min, and from 40% B to 100% B over 2 min, held at 100% B for 3 min, dropped to 3% B over 5 min, and held at 3% B for another 10 min. The Elite Orbitrap 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 (collision-induced dissociation [CID] normalized collision energy, 35; maximum fill time, 100 ms; automatic gain control [AGC], 1×10^4).

(iii) Mass spectrometry data analysis. Proteomic comparison of growth with and without sulfoquinovose was accomplished using MaxQuant (v1.5.5.1) (36). Searches were performed against *Rhizobium leguminosarum* bv. trifolii SRDI565 (NCBI taxonomy ID 935549; downloaded 8 January 2019; 6,404 entries) with carbamidomethylation of cysteine set as a fixed modification. Searches were performed with trypsin cleavage, allowing 2 miscleavage events and the variable modifications of oxidation of methionine and acetylation of protein N termini. The precursor mass tolerance was set to 20 ppm for the first search and 10 ppm for the main search, 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 (37) was enabled in addition to the requantification module. The resulting peptide outputs were processed within the Perseus (v1.4.0.6) (38) analysis environment to remove reverse matches and common protein contaminants with missing values imputed.

Enzyme kinetics of *RISQase.** **(i) Michaelis-Menten plot.** Kinetic analysis of *RISQase** was performed using PNPSQ as the substrate, using a UV-visible spectrophotometer to measure the release of the 4-nitrophenolate ($\lambda = 348 \text{ nm}$). Assays were carried out in 50 mM sodium phosphate–150 mM NaCl (pH 7.2) at 30°C using 212 nM *RISQase** at substrate concentrations ranging from $0.05 \mu\text{M}$ to 4 mM. Using an extinction coefficient for 4-nitrophenolate of $5.125 \text{ mM}^{-1} \text{ cm}^{-1}$, kinetic parameters were calculated using Prism.

(ii) pH profile. For the determination of pH profile, specific activities of *RISQase** were monitored by measuring absorbance changes at a wavelength of 348 nm in the presence of sodium acetate buffer (pH 5.6), sodium phosphate buffer (pH 6.0 to 8.5), and glycine NaOH buffer (pH 8.8 to 9.2). The assays were performed at 30°C in duplicate, and specific activities were determined using an extinction coefficient for *p*-nitrophenol (PNP) of $5.125 \text{ mM}^{-1} \text{ cm}^{-1}$ at the isosbestic point (348 nm). One unit of *SQase* activity is defined as the amount of protein that releases $1 \mu\text{mol}$ PNP per min.

Cloning, expression, and kinetic analysis of *RISQase.** The gene sequence coding for the *RISQase** SERp mutant was synthesized with codon optimization for expression in *E. coli* and was cloned within a pET-28a(+) vector with C-terminal His tag through GenScript. The plasmid His₆-*RISQase**-pET-28a(+) containing the gene for target *RISQase** was transformed into *E. coli* BL21(DE3) cells for protein expression. Precultures were grown in LB medium (5 ml) containing $30 \mu\text{g}/\text{ml}$ for 18 h at 37°C and 200 rpm. Cultures (1 liter of LB medium supplemented with $30 \mu\text{g}/\text{ml}$ kanamycin) were inoculated with the preculture (5 ml) and incubated at 37°C and 200 rpm until an OD_{600} of 0.6 to 0.8 was reached. Protein expression was induced by addition of IPTG (isopropyl- β -D-thiogalactopyranoside; 1 mM), and shaking was continued overnight (20 to 22 h) at 18°C and 200 rpm. The cells were harvested by centrifugation (5,000 rpm, 4°C , 20 min), resuspended in 50 mM Tris–300 mM NaCl (pH 7.5) buffer, and subjected to further cell lysis. Cells were disrupted using a French press under a pressure of 20,000 lb/in², and the lysate was centrifuged at $50,000 \times g$ for 30 min.

The N-terminal His₆-tagged protein was purified by immobilized metal ion affinity chromatography, followed by size exclusion chromatography (SEC) (Fig. S8). The lysate was loaded onto a pre-equilibrated nickel-nitrilotriacetic acid (Ni-NTA) column, followed by washing with loading buffer (50 mM Tris-HCl, 300 mM NaCl, 30 mM imidazole [pH 7.5]). The bound protein was eluted using a linear gradient with buffer containing 500 mM imidazole. Protein-containing fractions were pooled, concentrated, and loaded

onto a HiLoad 16/600 Superdex 200 gel filtration column pre-equilibrated with 50 mM Tris-HCl–300 mM NaCl (pH 7.5) buffer. The protein was concentrated to a final concentration of 60 mg ml⁻¹ using a Vivaspin 6 with a 300-kDa molecular weight cutoff membrane for characterization and enzyme assays.

SEC-MALS analysis. SEC-MALS experiments were conducted on a system comprising a Wyatt HELEOS-II multiangle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu high-performance liquid chromatography (HPLC) 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- μ m filter prior to use, and a 0.1- μ m filter was present in the flow path. The column was equilibrated with at least 2 column volumes of buffer (50 mM Tris, 300 mM NaCl [pH 7.5]) 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 μ l R/SQase* at 6 mg/ml in 50 mM Tris buffer–300 mM NaCl (pH 7.5). Shimadzu LC Solutions software was used to control the HPLC, and Astra V software was used for the HELEOS-II and rEX detectors (Fig. S8). 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 carryover between sample runs. Data were analyzed using the Astra V software. Molar masses were estimated using the Zimm fit method with degree 1. A value of 0.158 was used for the protein refractive index increment (dn/dc).

Detection of SQase activity in cell lysates. SRDI565 was grown in 50 ml Y_{5%}M and Y_{5%}SQ media at 30°C to mid-log phase, equivalent to an OD₆₀₀ of approximately 0.2, measured using a Varian Cary50 UV-visible spectrophotometer. Cells were harvested by adding a 3 \times volume of ice-cold PBS to metabolically quench the samples and then centrifuged at 2,000 \times g and 4°C for 10 min. The supernatant was discarded, and the cells were washed 3 times with ice-cold PBS, with each wash involving resuspension and centrifugation at 2,000 \times g and 4°C for 10 min. The cells were collected once more by centrifugation at 10,000 \times g and 4°C for 1 min, snap-frozen in liquid nitrogen, and stored at -80°C.

Cells were lysed by addition of 1,000 μ l prechilled PBS, 1 μ l RNase A, 1 μ l DNase, 1 μ l hen egg white lysozyme (100 mg ml⁻¹; Sigma), and a 1 \times final concentration of Complete EDTA-free protease inhibitor cocktail (Roche) to the cell pellet. The cells were gently resuspended and mixed at 4°C for 10 min. The suspension was placed on ice, irradiated with a Sonoplus HD3200 MS 73 sonicator probe (Bandelin) at a frequency of 20 kHz, 20% amplitude, and a pulse of 2 s on and 8 s off, repeated for a total time of sonication to 150 s, and then incubated on ice for 5 min. The suspension was clarified by centrifuging at 14,000 \times g and 4°C for 1 min, and the supernatant was filtered through a Nanosep mini-centrifugal spin column with a 0.2- μ m filter (Pall) into a 1.5-ml Eppendorf tubes and stored at 4°C. Protein concentration was determined using a bicinchoninic acid assay.

SQase activity was measured in triplicate using PNPSQ and an Agilent Cary UV workstation (G5191-64000) at 30°C. Reaction mixtures contained buffer consisting of 50 mM NaP_i and 150 mM NaCl (pH 7.4) and 2.5 mM PNPSQ. Reactions were initiated by addition of SQ- or mannitol-derived lysate to a final concentration of 43.7 μ g ml⁻¹ protein, and absorbance was monitored at 400 nm for 3 h. After 3 h, IFGSQ was added to a final concentration of 6.25 mM to the SQ-lysate sample, and absorption was monitored for 3 h.

Metabolite analysis of *R. leguminosarum* cell extracts. (i) Metabolic quenching and extraction. SRDI565 was grown on Y_{5%}SQ or Y_{5%} medium containing 35 mM glucose to mid-logarithmic phase (OD₆₀₀ of approximately 0.15), calculated based on the OD₆₀₀ measured by a Cary 50 UV-visible spectrophotometer, and was rapidly quenched in a prechilled 15-ml Falcon tube containing phosphate-buffered saline (PBS) at 4°C. Ice-cold PBS (11 ml) was infused into cell culture medium (4 ml). The Falcon tubes were mixed by inversion, incubated in an ice-water slurry for 5 min, and then centrifuged at 2,000 \times g at 1°C for 10 min. The supernatant was removed by aspiration, and cell pellets were washed twice with 1 ml of ice-cold PBS (with resuspension each time) and transferred into 1.5-ml Eppendorf tubes. Cells were pelleted by centrifugation at 14,000 rpm, and residual solvent was carefully removed. Cell pellets were stored at -80°C until extraction. Cells were extracted in 200 μ l of extraction solution (methanol-water, 3:1 [vol/vol]) containing an internal standard, 5 μ M [¹³C₄]aspartate (Cambridge Isotopes), and subjected to 10 freeze-thaw cycles to facilitate cell lysis (30 s in liquid nitrogen followed by 30 s in a dry ice-ethanol bath). Debris was pelleted by centrifugation at 14,000 rpm (1°C for 5 min), and cell lysate was transferred into an HPLC vial insert for LC-MS analysis.

(ii) LC-MS analysis and identification of sulfonate metabolites. Separation and detection of polar metabolites were performed using an Agilent Technologies 1200 series high-performance liquid chromatograph (HPLC) coupled to a quadrupole time-of-flight mass spectrometer (6545 QTOF; Agilent Technologies) using a method modified from that of Masukagami et al. (39). Metabolite extracts were transferred into 2-ml autosampler vials with glass inserts and placed in the autosampler and held at 4°C prior to analysis. Metabolite separation was performed by injecting 7 μ l of the extract into a SeQuant ZIC-pHILIC PEEK (polyetheretherketone)-coated column (150 mm by 4.6 mm, 5- μ m polymer; Merck Millipore) maintained at 25°C, with a gradient of solvent A (20 mM ammonium carbonate [pH 9.0]; Sigma-Aldrich) and solvent B (100% acetonitrile, hypergrade for LC-MS [LiChrosolv; Merck]) at a flow rate of 0.3 ml/min. A 33.0-min gradient was set up as follows: 0 min, 80% B; 0.5 min, 80% B; 15.5 min, 50% B; 17.5 min, 30% B; 18.5 min, 5% B; 21.0 min, 5% B; 23.0 min, 80% B.

The LC flow was directed into an electrospray ionization (ESI) source with a capillary voltage of 2,500 V operating in negative ionization mode. Drying nitrogen gas flow was set to 10 liters/min, and sheath gas temperature and nebulizer pressure were set to 300°C and 20 lb/in², respectively. The voltages of fragmentor and skimmer were set at 125 V and 45 V, respectively. Data were acquired in MS and MS/MS mode, with scan ranges of 60 to 1,700 *m/z* and 100 to 1,700 *m/z*, respectively, at a rate of 1.5 spectra/s. MS/MS acquisition was performed with four collision energies (0, 10, 20, and 40 V). The mass spectrom-

eter was calibrated in negative mode prior to data acquisition, and mass accuracy during runs was ensured by a continuous infusion of reference mass solution at a flow rate of 0.06 ml/min (API-TOF reference mass solution kit; Agilent Technologies). Data quality was ensured by multiple injections of standards (each at a 1.5 μ M concentration) and a pooled biological sample (a composite of cell extracts) used to monitor the instrument performance. Samples were randomized prior to metabolite extraction and LC-MS analysis.

(iii) Standard preparation. Standards of selected metabolites (Table S1) were prepared at 10 μ M in 80% acetonitrile (hypergrade for LC-MS [LiChrosolv; Merck]) and injected separately into a column connected to the mass spectrometer interface. Retention time and detected molecular ion were used to create a targeted MS/MS acquisition method. The spectra, mass-to-charge ratio (m/z), and retention time were imported into a personal compound database and library (PCDL Manager, version B.07.00; Agilent Technologies) used in the data processing workflow.

(iv) Data analysis. Data were analyzed using MassHunter qualitative and quantitative analysis software (version B.07.00; Agilent Technologies). Identification of metabolites was performed in accordance with metabolite identification (Metabolomics Standard Initiative) level 1 based on the retention time and molecular masses matching authentic standards included in the personal database and library. Peak integration was performed with MassHunter quantitative software (version B.07.00; Agilent Technologies) on the spectra from identified metabolites.

Chemical synthesis of 6-deoxy-6-sulfo-D-gluconate. NaOH in methanol (4% [wt/vol]; 4 ml) was added dropwise to a stirred solution of sulfoquinovose (100 mg; 0.410 mmol) and iodine (209 mg; 1.65 mmol) in water (1 ml) and methanol (4 ml) held at 40°C. As the sodium hydroxide was added, the color of iodine dissipated. The solvent was evaporated under reduced pressure, and the crude residue was subjected to flash chromatography (ethyl acetate [EtOAc]-methanol [MeOH]-H₂O, 4:2:1 to 2:2:1, then water) to give the 6-deoxy-6-sulfolgluconate sodium salt (89.2 mg). An aqueous solution of the sodium salt was eluted through a column of Amberlite IR120 (H⁺ form) resin. The acidic eluate was collected and concentrated under reduced pressure to give SG (71.3 mg, 67%). ¹H NMR (400 MHz, D₂O): δ 4.23 to 4.15 (1H, m, H2), 4.13 (1 H, d, J = 3.3 Hz, H3), 4.05 (1 H, t, J = 3.2 Hz, H5), 3.74 (1 H, dd, J = 6.5, 3.4 Hz, H4), 3.35 (1 H, d, J = 14.6 Hz, H6a), 3.05 (1 H, dd, J = 14.6, 9.7 Hz, H6b); ¹³C (¹H) NMR (100 MHz, D₂O) δ 178.7 (C1), 74.2 (C4), 73.8 (C2), 70.8 (C3), 67.8 (C5), 53.4 (C6); HRMS (ESI⁻) for C₆H₁₁O₉S [M⁻]: calculated, 259.0129; found, 259.0131.

Quantitation of metabolite levels in spent culture medium. The metabolites (DHPS, SF, SQ, SL, and SG) present in spent culture medium were quantified against standard solutions of pure metabolites by LC-ESI-MS/MS. Quantification was done with the aid of calibration curves generated by dissolving the pure standards in spent medium from SRDI565 grown on Y_{5%}M. Spiked spent medium was diluted 100-fold with water and then analyzed by LC-MS/MS with α -MeSQ as the internal standard. For quantitation of metabolites, spent culture medium from SRDI565 grown in Y_{5%}SQ or Y_{5%}SQGro was diluted 100-fold with water and analyzed by LC-MS/MS with α -MeSQ as the internal standard.

LC-ESI-MS/MS analysis was performed using a TSQ Altis triple-quadrupole mass spectrometer (Thermo Fisher Scientific) coupled with a Vanquish Horizon UHPLC system (Thermo Fisher Scientific). The column was a ZIC hydrophilic interaction liquid chromatography (HILIC) column (5 μ m, 50 by 2.1 mm; Merck). The HPLC conditions were as follows: from 90% B to 40% B over 15 min, 40% B for 5 min, and back to 90% B over 1 min (solvent A, 20 mM NH₄OAc in 1% acetonitrile; solvent B, acetonitrile); flow rate, 0.30 ml min⁻¹; injection volume, 1 μ l. The mass spectrometer was operated in negative ionization mode. Quantitation was done using the MS/MS selected reaction monitoring (SRM) mode using Thermo Scientific Xcalibur software and normalized with respect to the internal standard, α -MeSQ. Prior to analysis, for each analyte, the sensitivity for each SRM-MS/MS transition was optimized. Results were as follows: for DHPS, ESI-MS/MS m/z of [M-H]⁻, 155; product ions, 137, 95; retention time, 4.91 min; for α -MeSQ (internal standard), ESI-MS/MS m/z of [M-H]⁻, 257; product ions, 166, 81; retention time, 6.31 min; for SF, ESI-MS/MS m/z of [M-H]⁻, 243; product ions, 207, 153; retention time, 6.81 min; for SQ, ESI-MS/MS m/z of [M-H]⁻, 243; product ions, 183, 123; retention time, 7.58 and 7.89 min for α/β ; for SL, ESI-MS/MS m/z of [M-H]⁻, 169; product ions, 107, 71; retention time, 9.26 min; for SG, ESI-MS/MS m/z of [M-H]⁻, 259; product ions, 241, 161; retention time, 9.66 min; and for SQGro, ESI-MS/MS m/z of [M-H]⁻, 317; product ions, 225, 165; retention time, 7.15 min.

Data availability. The mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier [PXD015822](https://www.ebi.ac.uk/PRIDE/archive/PXD015822).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.5 MB.

ACKNOWLEDGMENTS

This work was supported by grants from the Australian Research Council (DP180101957), the National Health and Medical Research Council of Australia (APP1100164 and GNT1139549), and the Leverhulme Trust; 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. M.J.M. is an

NHMRC Principal Research Fellow; G.J.D. is a Royal Society Ken Murray Research Fellow. J.L. is supported by a Ph.D. scholarship from the China Scholarship Council.

We thank Ross Ballard (SRDI) for supplying SRDI565, Humane Technologies for support with the MicrobeMeter, the Melbourne Mass Spectrometry and Proteomics and the Metabolomics Australia facilities of the Bio21 Institute at the University of Melbourne, Palika Abayakoon and Janice Mui for reagents, and Shuai Nie, Yunyang Zhang, and Alex Chen (Thermo Fisher) for technical support. Thermo Fisher Scientific Australia are acknowledged for access to the TSQ Altis triple quadrupole mass spectrometer.

REFERENCES

- Hu ZY, Zhao FJ, McGrath SP. 2005. Sulphur fractionation in calcareous soils and bioavailability to plants. *Plant Soil* 268:103–109. <https://doi.org/10.1007/s11104-004-0229-0>.
- Wilhelm Scherer H. 2009. Sulfur in soils. *J Plant Nutr Soil Sci* 172:326–335. <https://doi.org/10.1002/jpln.200900037>.
- Scherer HW. 2001. Sulphur in crop production—invited paper. *Eur J Agronomy* 14:81–111. [https://doi.org/10.1016/S1161-0301\(00\)00082-4](https://doi.org/10.1016/S1161-0301(00)00082-4).
- Tabatabai MA. 1984. Importance of sulphur in crop production. *Biogeochemistry* 1:45–62. <https://doi.org/10.1007/BF02181120>.
- Kertesz MA, Mirleau P. 2004. The role of soil microbes in plant sulphur nutrition. *J Exp Bot* 55:1939–1945. <https://doi.org/10.1093/jxb/erh176>.
- Kertesz MA. 2000. Riding the sulfur cycle—metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiol Rev* 24:135–175. [https://doi.org/10.1016/S0168-6445\(99\)00033-9](https://doi.org/10.1016/S0168-6445(99)00033-9).
- Autry AR, Fitzgerald JW. 1990. Sulfonate S: a major form of forest soil organic sulfur. *Biol Fertil Soils* 10:50–56.
- Harwood JL, Nicholls RG. 1979. The plant sulpholipid—a major component of the sulphur cycle. *Biochem Soc Trans* 7:440–447. <https://doi.org/10.1042/bst0070440>.
- Goss R, Nerlich J, Lepetit B, Schaller S, Vieler A, Wilhelm C. 2009. The lipid dependence of diadinoxanthin de-epoxidation presents new evidence for a macrodomain organization of the diatom thylakoid membrane. *J Plant Physiol* 166:1839–1854. <https://doi.org/10.1016/j.jplph.2009.05.017>.
- Strickland TC, Fitzgerald JW. 1983. Mineralization of sulphur in sulphoquinovose by forest soils. *Soil Biol Biochem* 15:347–349. [https://doi.org/10.1016/0038-0717\(83\)90082-2](https://doi.org/10.1016/0038-0717(83)90082-2).
- Goddard-Borger ED, Williams SJ. 2017. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem J* 474:827–849. <https://doi.org/10.1042/BCJ20160508>.
- Denger K, Weiss M, Felux AK, Schneider A, Mayer C, Spiteller D, Huhn T, Cook AM, Schleheck D. 2014. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* 507:114–117. <https://doi.org/10.1038/nature12947>.
- Felux AK, Spiteller D, Klebensberger J, Schleheck D. 2015. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc Natl Acad Sci U S A* 112:E4298–E4305. <https://doi.org/10.1073/pnas.1507049112>.
- Denger K, Huhn T, Hollemeyer K, Schleheck D, Cook AM. 2012. Sulfoquinovose degraded by pure cultures of bacteria with release of C₃-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol Lett* 328:39–45. <https://doi.org/10.1111/j.1574-6968.2011.02477.x>.
- Drew EA, Ballard RA. 2010. Improving N₂ fixation from the plant down: compatibility of *Trifolium subterraneum* L. cultivars with soil rhizobia can influence symbiotic performance. *Plant Soil* 327:261–277. <https://doi.org/10.1007/s11104-009-0052-8>.
- Melino VJ, Drew EA, Ballard RA, Reeve WG, Thomson G, White RG, O'Hara GW. 2012. Identifying abnormalities in symbiotic development between *Trifolium* spp. and *Rhizobium leguminosarum* bv. trifolii leading to sub-optimal and ineffective nodule phenotypes. *Ann Bot* 110:1559–1572. <https://doi.org/10.1093/aob/mcs206>.
- Reeve W, Drew E, Ballard R, Melino V, Tian R, De Meyer S, Brau L, Ninawi M, Teshima H, Goodwin L, Chain P, Liolios K, Pati A, Mavromatis K, Ivanova N, Markowitz V, Woyke T, Kyripides N. 2013. Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. trifolii strain SRDI565. *Stand Genomic Sci* 9:220–231. <https://doi.org/10.4056/signs.4468250>.
- Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* 64:781–805. <https://doi.org/10.1146/annurev-arplant-050312-120235>.
- Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* 16:291–303. <https://doi.org/10.1038/nrmicro.2017.171>.
- Stowers MD. 1985. Carbon metabolism in *Rhizobium* species. *Annu Rev Microbiol* 39:89–108. <https://doi.org/10.1146/annurev.mi.39.100185.000513>.
- Abayakoon P, Lingford JP, Jin Y, Bengt C, Davies GJ, Yao S, Goddard-Borger ED, Williams SJ. 2018. Discovery and characterization of a sulfoquinovose mutarotase using kinetic analysis at equilibrium by exchange spectroscopy. *Biochem J* 475:1371–1383. <https://doi.org/10.1042/BCJ20170947>.
- Bergersen FJ. 1961. The growth of rhizobium in synthetic media. *Aust J Biol Sci* 14:349–360. <https://doi.org/10.1071/B19610349>.
- Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. 2007. Toward rational protein crystallization: a Web server for the design of crystallizable protein variants. *Protein Sci* 16:1569–1576. <https://doi.org/10.1110/ps.072914007>.
- Abayakoon P, Jin Y, Lingford JP, Petricevic M, John A, Ryan E, Wai-Ying Mui J, Pires DEV, Ascher DB, Davies GJ, Goddard-Borger ED, Williams SJ. 2018. Structural and biochemical insights into the function and evolution of sulfoquinovosidases. *ACS Cent Sci* 4:1266–1273. <https://doi.org/10.1021/acscentsci.8b00453>.
- Speciale G, Jin Y, Davies GJ, Williams SJ, Goddard-Borger ED. 2016. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat Chem Biol* 12:215–217. <https://doi.org/10.1038/nchembio.2023>.
- Roy AB, Hewlins M. 1997. Sulfoquinovose and its aldonic acid: their preparation and oxidation to 2-sulfoacetaldehyde by periodate. *Carbohydr Res* 302:113–117. [https://doi.org/10.1016/S0008-6215\(97\)00112-2](https://doi.org/10.1016/S0008-6215(97)00112-2).
- Abayakoon P, Epa R, Petricevic M, Bengt C, Mui JWY, van der Peet PL, Zhang Y, Lingford JP, White JM, Goddard-Borger ED, Williams SJ. 2019. Comprehensive synthesis of substrates, intermediates and products of the sulfoglycolytic Embden-Meyerhoff-Parnas pathway. *J Org Chem* 84:2910–2910.
- Roy AB, Ellis AJ, White GF, Harwood JL. 2000. Microbial degradation of the plant sulpholipid. *Biochem Soc Trans* 28:781–783. <https://doi.org/10.1042/bst0280781>.
- Roy AB, Hewlins MJ, Ellis AJ, Harwood JL, White GF. 2003. Glycolytic breakdown of sulfoquinovose in bacteria: a missing link in the sulfur cycle. *Appl Environ Microbiol* 69:6434–6441. <https://doi.org/10.1128/aem.69.11.6434-6441.2003>.
- Martelli HL, Benson AA. 1964. Sulfocarbohydrate metabolism. I. Bacterial production and utilization of sulfoacetate. *Biochim Biophys Acta* 93:169–171. [https://doi.org/10.1016/0304-4165\(64\)90272-7](https://doi.org/10.1016/0304-4165(64)90272-7).
- Davidson AL, Dassa E, Orelle C, Chen J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* 72:317–364. <https://doi.org/10.1128/MMBR.00031-07>.
- McKay IA, Glenn AR, Dilworth MJ. 1985. Gluconeogenesis in *Rhizobium leguminosarum* MNF3841. *Microbiology* 131:2067–2073. <https://doi.org/10.1099/00221287-131-8-2067>.
- Sasidharan K, Martinez-Vernon AS, Chen J, Fu T, Soyer OS. 2018. A low-cost DIY device for high resolution, continuous measurement of microbial growth dynamics. *bioRxiv* <https://www.biorxiv.org/content/10.1101/407742v1>.
- Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J. 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* 14:68–85. <https://doi.org/10.1038/s41596-018-0082-x>.
- Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics

- using StageTips. *Nat Protoc* 2:1896–1906. <https://doi.org/10.1038/nprot.2007.261>.
36. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372. <https://doi.org/10.1038/nbt.1511>.
 37. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13:2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.
 38. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13:731–740. <https://doi.org/10.1038/nmeth.3901>.
 39. Masukagami Y, Nijagal B, Mahdizadeh S, Tseng CW, Dayalan S, Tivendale KA, Markham PF, Browning GF, Sansom FM. 2019. A combined metabolomic and bioinformatic approach to investigate the function of transport proteins of the important pathogen *Mycoplasma bovis*. *Vet Microbiol* 234:8–16. <https://doi.org/10.1016/j.vetmic.2019.05.008>.



Cite this: *Chem. Soc. Rev.*, 2021, 50, 13628

Sulfoglycolysis: catabolic pathways for metabolism of sulfoquinovose

Alexander J. D. Snow,^a Laura Burchill,^{bc} Mahima Sharma,^{id}*^a Gideon J. Davies^{id}*^a and Spencer J. Williams^{id}*^{bc}

Sulfoquinovose (SQ), a derivative of glucose with a C6-sulfonate, is produced by photosynthetic organisms and is the headgroup of the sulfolipid sulfoquinovosyl diacylglycerol. The degradation of SQ allows recycling of its elemental constituents and is important in the global sulfur and carbon biogeochemical cycles. Degradation of SQ by bacteria is achieved through a range of pathways that fall into two main groups. One group involves scission of the 6-carbon skeleton of SQ into two fragments with metabolic utilization of carbons 1–3 and excretion of carbons 4–6 as dihydroxypropanesulfonate or sulfolactate that is biomineralized to sulfite/sulfate by other members of the microbial community. The other involves the complete metabolism of SQ by desulfonation involving cleavage of the C–S bond to release sulfite and glucose, the latter of which can enter glycolysis. The discovery of sulfoglycolytic pathways has revealed a wide range of novel enzymes and SQ binding proteins. Biochemical and structural characterization of the proteins and enzymes in these pathways have illuminated how the sulfonate group is recognized by Nature's catalysts, supporting bioinformatic annotation of sulfoglycolytic enzymes, and has identified functional and structural relationships with the pathways of glycolysis.

Received 5th October 2021

DOI: 10.1039/d1cs00846c

rsc.li/chem-soc-rev

^a York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, YO10 5DD, UK. E-mail: mahima.sharma@york.ac.uk, gideon.davies@york.ac.uk

^b School of Chemistry, University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: sjwill@unimelb.edu.au

^c Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

1. Introduction

Sulfoquinovose (SQ) is a sulfonated hexose analogous to D-glucose, but which contains a sulfur–carbon bond (Fig. 1).^{1–3} SQ is produced by photosynthetic organisms and is the anionic headgroup of the sulfolipid sulfoquinovosyl diacylglycerol (SQDG) found in plants, algae, and cyanobacteria. SQ occurs



Alexander J. D. Snow

Alex Snow read BSc Biochemistry at the University of Southampton, then moved into the field of structural biology, reading MRes Structural Molecular Biology at Imperial College London. There he worked with Professor Kurt Drickamer, before moving to York to study for a PhD with Professor Gideon Davies at YSBL, University of York. During these degrees Alex has kindled a strong interest in structural biology as well as biophysical enzyme characterization. He is currently searching for postdoctoral positions and is excited to continue expanding his skills and knowledge.



Laura Burchill

Laura Burchill is a postdoctoral fellow at the University of Melbourne. She graduated Monash University with a BSc (Hons), and in 2021 completed her PhD on the biomimetic synthesis of natural products at the University of Adelaide, under the supervision of Assoc. Prof. Jonathan George. Currently her research in the laboratory of Professor Spencer Williams focuses on the synthesis of small molecule metabolites for the elucidation of biochemical pathways.

within the related metabolites sulfoquinovosyl monoglyceride (SQMG, lyso-sulfolipid),⁴ sulfoquinovosyl glycerol (SQGro),⁵ sulfoquinovosyl glyceryl ether,⁶ and 2'-*O*-acyl-SQDG.⁷ SQ is also present within the N-linked glycan of the archaeon *Sulfolobus acidocaldarius*.⁸ SQDG is localized to the thylakoid membrane, is found in intimate association with photosynthetic protein complexes,^{9,10} and assists in the function of photosystem II.¹¹ The biosynthesis of SQ occurs at the level of the sugar nucleotide diphosphate, by reaction of sulfite with UDP-glucose to give UDP-SQ (catalyzed by UDP-sulfoquinovose synthase), which is then transferred to diacylglycerol to give SQDG (catalyzed by the glycosyltransferase SQDG synthase).^{12,13} Owing to its production within photosynthetic tissues,^{1,12,14} SQ constitutes a major reservoir of organosulfur, with one estimate of its production standing at 10 billion tonnes annually.¹⁵ Regardless of the precise amount, this

places SQ among other important organosulfur compounds in the biosphere including the amino acids cysteine and methionine, the simple sulfonate 2,3-dihydroxypropane-sulfonate (DHPS)¹⁶ and the osmolytes dimethylsulfoniopropionate¹⁷ and dimethylsulfoxonium propionate.¹⁸

SQDG is a sulfur reservoir in photosynthetic organisms, and a carbon and sulfur source for diverse microbial communities. Shortly after the discovery of SQ, and amid growing appreciation for its ubiquity in photosynthetic organisms, it was realised that breakdown pathways must exist. Early hypotheses suggested that SQ deconstruction could be analogous to glycolysis, and thus was described as sulfoglycolysis.¹⁹ Subsequently, various organisms were isolated that could grow on SQ as sole carbon source and it was found that they produced a range of end-products including DHPS, sulfolactate (SL) and sulfate.^{20–22} While early studies identified possible breakdown products of SQ, until relatively recently it proved difficult to elucidate the individual steps.

In recent years, a series of breakthroughs have led to the identification of four pathways of sulfoglycolysis and has ushered in a new era of genetic, bioinformatic, structural, biochemical and microbiological studies of SQ (Fig. 1). These four pathways fall into two main groups depending on whether they allow microorganisms to use 3 or 6 of the carbons within the SQ skeleton. The first group is comprised of bacteria that cleave the 6-carbon chain of SQ into two C3 chains, leading to production of DHPS, pyruvate or F6P (by transfer of a glycerone fragment to GAP) that are utilized by the organism, and sulfolactaldehyde (SLA), which is oxidized to SL or reduced to DHPS then excreted. These organisms use one of three pathways: the sulfoglycolytic Embden–Meyerof–Parnas (sulfo-EMP) (Fig. 2),²³ sulfoglycolytic Entner–Doudoroff (sulfo-ED) (Fig. 3)²⁴ and sulfoglycolytic sulfofructose transaldolase (sulfo-SFT; also termed the sulfoglycolytic transaldolase, sulfo-TAL) (Fig. 4)^{25,26} pathways. The excreted SL or DHPS becomes available to organisms that utilize its carbon and effect the biomineralization of the sulfonate group to sulfite or sulfate.²⁷ The second



Mahima Sharma

Mahima Sharma is a post-doctoral researcher working with Prof. Gideon Davies at the University of York. She received her BSc Honours degree in Chemistry from the University of Delhi and her DPhil in Chemical Biology from the University of Oxford. She worked on designing artificial metalloenzymes for C–C cross-coupling reactions with Prof. Benjamin G. Davis, and on structural studies of reductive aminases enabling chiral amine synthesis in Prof. Gideon Grogan's group and was awarded the 2021 RSC Horizon Prize (Rita and John Cornforth Award for collaborative research). Her current work is focussed upon investigating the sulfoquinovose degradation pathways.

Mahima Sharma is a post-doctoral researcher working with Prof. Gideon Davies at the University of York. She received her BSc Honours degree in Chemistry from the University of Delhi and her DPhil in Chemical Biology from the University of Oxford. She worked on designing artificial metalloenzymes for C–C cross-coupling reactions with Prof. Benjamin G. Davis, and on structural studies of reductive aminases enabling chiral amine synthesis in Prof. Gideon Grogan's group and was awarded the 2021 RSC Horizon Prize (Rita and John Cornforth Award for collaborative research). Her current work is focussed upon investigating the sulfoquinovose degradation pathways.



Gideon J. Davies

Gideon Davies did his PhD at the University of Bristol with Herman Watson and Len Hall. Following a period at the EMBL Hamburg synchrotron outstation, he moved to York (with periods in Uppsala, Grenoble and Vancouver) where he was made a full professor in 2001. In 2016 he was made the Royal Society "Ken Murray" Research Professor. Gideon's work focusses on glycochemistry: carbohydrate-active enzymes, their mechanisms, and 3D structures – designing novel chemical biology approaches to their study – and applying enzymes and tools for societal benefits in biotechnology and human health.

Gideon Davies did his PhD at the University of Bristol with Herman Watson and Len Hall. Following a period at the EMBL Hamburg synchrotron outstation, he moved to York (with periods in Uppsala, Grenoble and Vancouver) where he was made a full professor in 2001. In 2016 he was made the Royal Society "Ken Murray" Research Professor. Gideon's work focusses on glycochemistry: carbohydrate-active enzymes, their mechanisms, and 3D structures – designing novel chemical biology approaches to their study – and applying enzymes and tools for societal benefits in biotechnology and human health.



Spencer J. Williams

Spencer Williams earned BSc (Hons) and PhD degrees with Prof Bob Stick at the University of Western Australia, and undertook postdoctoral training at the University of British Columbia (with Prof Stephen Withers) and University of California at Berkeley (with Prof Carolyn Bertozzi). His independent career has been spent at the University of Melbourne, where he focusses on the synthesis of bioactive molecules, biological chemistry and chemical biology. A collaborative spirit pervades his research and he is committed to training the next generation of scientists with fluency in chemistry and biology.

Spencer Williams earned BSc (Hons) and PhD degrees with Prof Bob Stick at the University of Western Australia, and undertook postdoctoral training at the University of British Columbia (with Prof Stephen Withers) and University of California at Berkeley (with Prof Carolyn Bertozzi). His independent career has been spent at the University of Melbourne, where he focusses on the synthesis of bioactive molecules, biological chemistry and chemical biology. A collaborative spirit pervades his research and he is committed to training the next generation of scientists with fluency in chemistry and biology.

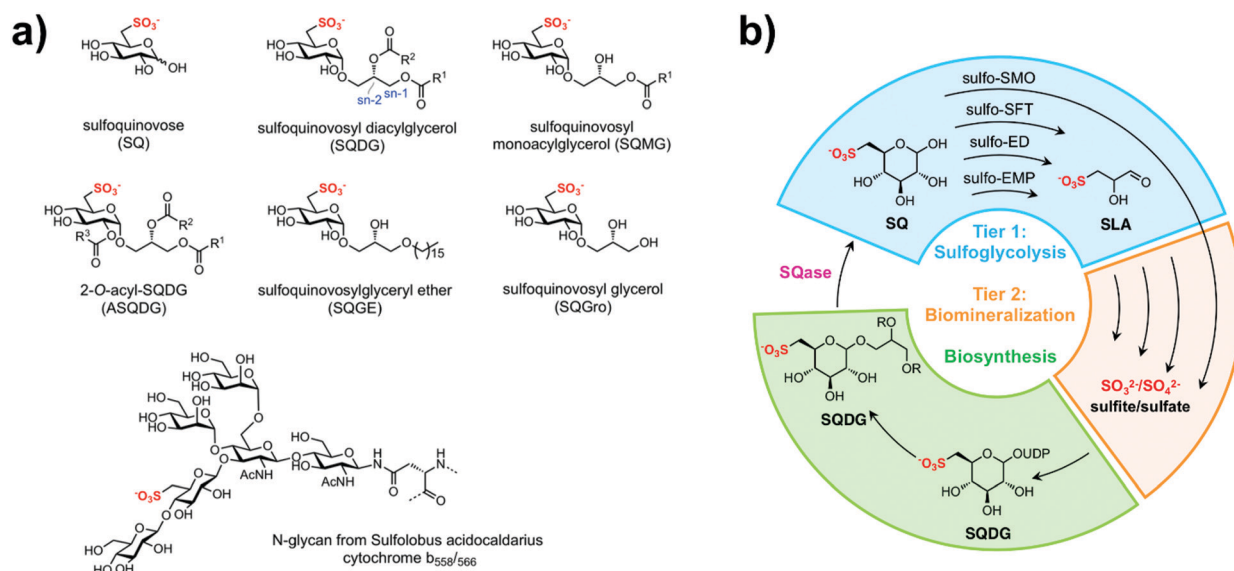


Fig. 1 Speciation and metabolism of sulfoquinovose (SQ). (a) Structures of sulfoquinovose and naturally occurring glycoconjugates. (b) Overview of the biosynthesis and catabolism of sulfoquinovosyl diglyceride (SQDG) and SQ.

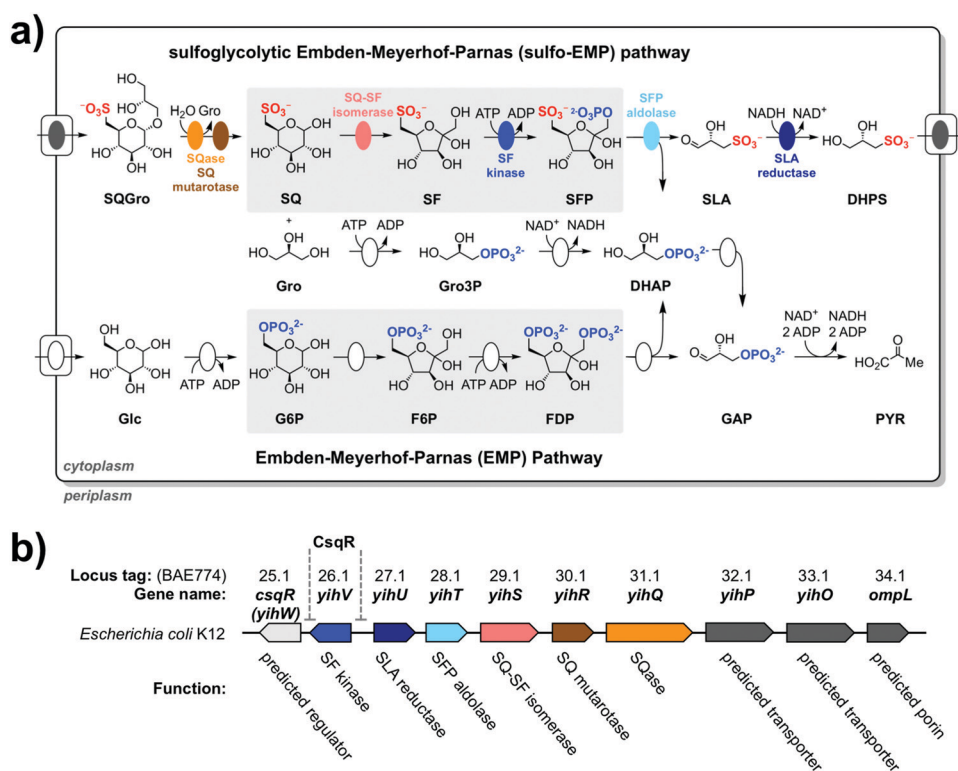


Fig. 2 The sulfoglycolytic Embden–Meyerhof–Parnas (sulfo-EMP) pathway. (a) Comparison of the sulfo-EMP pathway for sulfoglycolysis of SQ by *E. coli* with the EMP pathway for glycolysis of glucose. (b) Sulfo-EMP gene cluster from *E. coli* K-12, showing intergenic binding sites for the transcription factor *CsqR*.

group contains bacteria that use a fourth pathway, termed the sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway (Fig. 5), which produces glucose and sulfite and is the

only pathway in a single organism that results in cleavage of the sulfur–carbon bond of SQ to allow complete utilization of all six of its carbons.²⁸

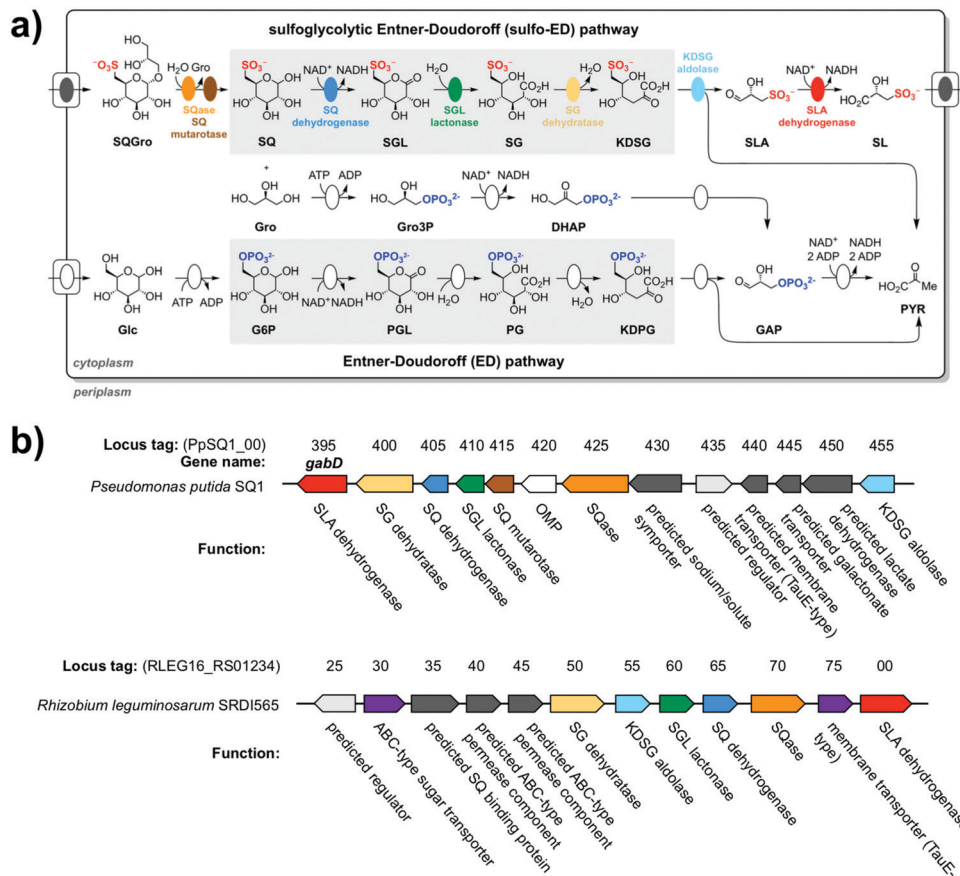


Fig. 3 The sulfoglycolytic Entner–Doudoroff (sulfo-ED) pathway. (a) Comparison of the sulfo-ED pathway for sulfoglycolysis of SQ by *P. putida* SQ1 with the ED pathway for glycolysis of glucose. (b) sulfo-ED gene clusters from *P. putida* SQ1 and *R. leguminosarum* SRDI565.

2. Early observations on the catabolism of sulfoquinovose and sulfoquinovosyl glycerol

Early studies of the metabolism of SQ and SQGro focused on algae and plants. The formation and breakdown of ^{35}S -SQDG was studied in the alga *Chlorella ellipsoidea* after labelling with ^{35}S -sulfate. Using 2D-thin layer chromatography a suite of radiolabelled products were observed that included SQGro, SQ, DHPS, SL and SLA.²⁹ SLA condensed with dihydroxyacetone phosphate in the presence of rabbit muscle aldolase to give a sulfoketose phosphate.²⁹ This data supported the existence of a sulfoglycolytic pathway in *Chlorella*.¹⁹ In alfalfa leaves, SQGro is broken down to SQ, SLA and SL, showing the existence of a sulfoglycolytic pathway in plants.³⁰ In coral tree (*Erythrina crista-galli*), the only product of SQGro breakdown observed was sulfoacetic acid, which was proposed to result from decarboxylation of SL and then oxidation.³⁰ A recent survey of algae identified DHPS in many but not all species investigated.³¹ Bioinformatic analysis of macro- and microalgae revealed only limited occurrence of SQase and SLA reductase homologues, but with no evidence of an intact sulfo-EMP pathway.³¹ Likewise, while homologues of SLA dehydrogenase, sulfogluconate

(SG) dehydratase and SQ dehydrogenase were observed in the majority of algal species, homologues of SQase and 2-keto-3-deoxysulfoquaconate (KDSG) aldolase were found only in a smaller subset and no complete sulfo-ED pathway could be identified. These data suggest that either algae use other pathways for SQ catabolism, or do not accomplish complete sulfoglycolysis.

Other early studies examined the breakdown of SQ by saprophytic soil microorganisms (*i.e.* those that digest decaying organic matter). Field studies involving the incubation of SQ in assorted forest soils revealed its rapid mineralization to sulfate, suggesting that sulfoglycolytic microorganisms are widespread.³² Martelli and Benson isolated a soil *Flavobacterium* sp. that converted methyl α -sulfoquinovoside to sulfoacetate in a phosphate-dependent manner.³³ Roy and co-workers isolated several SQ-metabolizing bacteria from soil and sewage sludge, which produced a range of metabolic end-products: SL, DHPS or sulfate.^{20,21} *Klebsiella* sp. ABR11 initially produced DHPS and SL, and eventually sulfate.²¹ Crude cell-free extracts from this bacterium exhibited ‘phosphofructokinase’ activity but using SQ as substrate, and NAD^+ -dependent SQ dehydrogenase activity. Based on these results it was proposed that ABR11 performed sulfoglycolysis using pathways that were variants

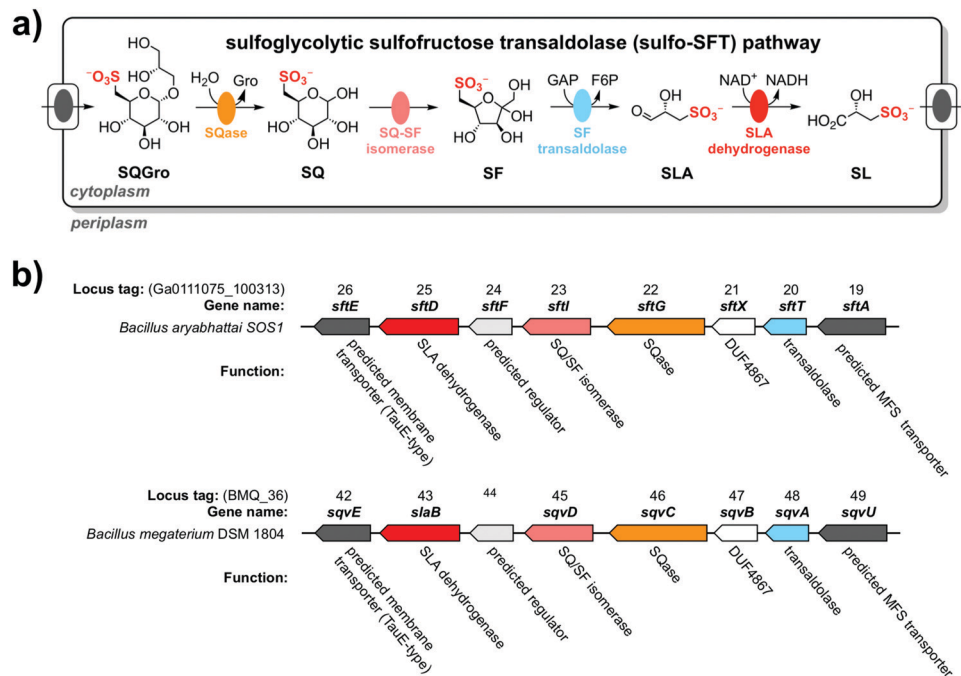


Fig. 4 The sulfoglycolytic sulfofructose transaldolase (Sulfo-SFT) pathway. (a) Sulfo-SFT pathway for sulfoglycolysis of SQ. (b) sulfo-SFT gene clusters from *B. aryabhatai* SOS1 and *B. megaterium* DSM 1804.

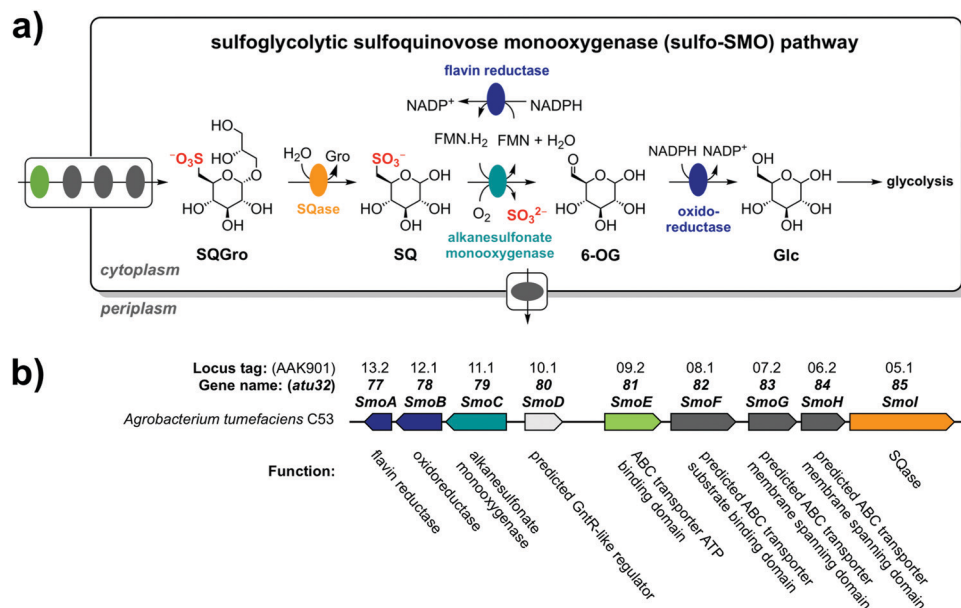


Fig. 5 The sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway. (a) Sulfo-SMO pathway for sulfoglycolysis of SQ in *A. tumefaciens* C58. (b) Sulfo-SMO gene cluster.

of the classical Embden–Meyerhof–Parnas (EMP) or Entner–Doudoroff (ED) glycolysis pathways.²¹ By contrast, *Agrobacterium* sp. ABR2 produced only sulfate and bicarbonate as metabolic products, with a significant delay between SQ consumption and sulfate production.²¹

Collectively, this work provided strong evidence for the occurrence of sulfoglycolytic pathways in plants, algae and bacteria, and a framework for understanding the molecular details of sulfoglycolysis. Recent breakthroughs have placed these early observations on firmer foundations and identified

dedicated sulfoglycolytic pathways in bacteria. The pathways used in plants and algae remain unclear.

3. Pre-sulfoglycolytic processing of SQDG: delipidation, SQ glycoside hydrolysis, and mutarotation

3.1 Delipidation of SQDG: formation of SQMG and SQGro

While specific sulfolipid lipases (sulfolipases) have not yet been discovered, there is evidence for enzymes that can effect the partial and complete delipidation of SQDG (Fig. 6a). Yagi and Benson reported that incubation of ^{35}S -SQDG with a crude extract of the algae *Scenedesmus obliquus* resulted in the sequential

formation of SQMG then SQGro.⁴ Similar lipase activity on SQDG was observed in *Chlorella ellipsoidea*,⁴ alfalfa (*Medicago sativa*),⁴ corn-roots extracts⁴ and *Chlorella pyrenoidosa*.³⁴ On the other hand a lipid hydrolase isolated from leaves of the scarlet runner bean, *Phaseolus multiflorus*, converted SQDG to SQGro with no SQMG observed.³⁵ Hazelwood and Dawson reported the isolation of a lipolytic fatty-acid requiring *Butyrivibrio* sp. S2 from sheep rumen that could catabolize SQDG but its further metabolism was not studied.³⁶ When *Chlorella protothecoides* was grown in light, levels of SQDG increased while levels of SQGro was stable.³⁷ By contrast, levels of SQDG declined in 'glucose bleached' *C. protothecoides* cells, while SQGro increased, suggesting that lipase action is connected with the bleaching action and disintegration of the photosynthetic architecture.

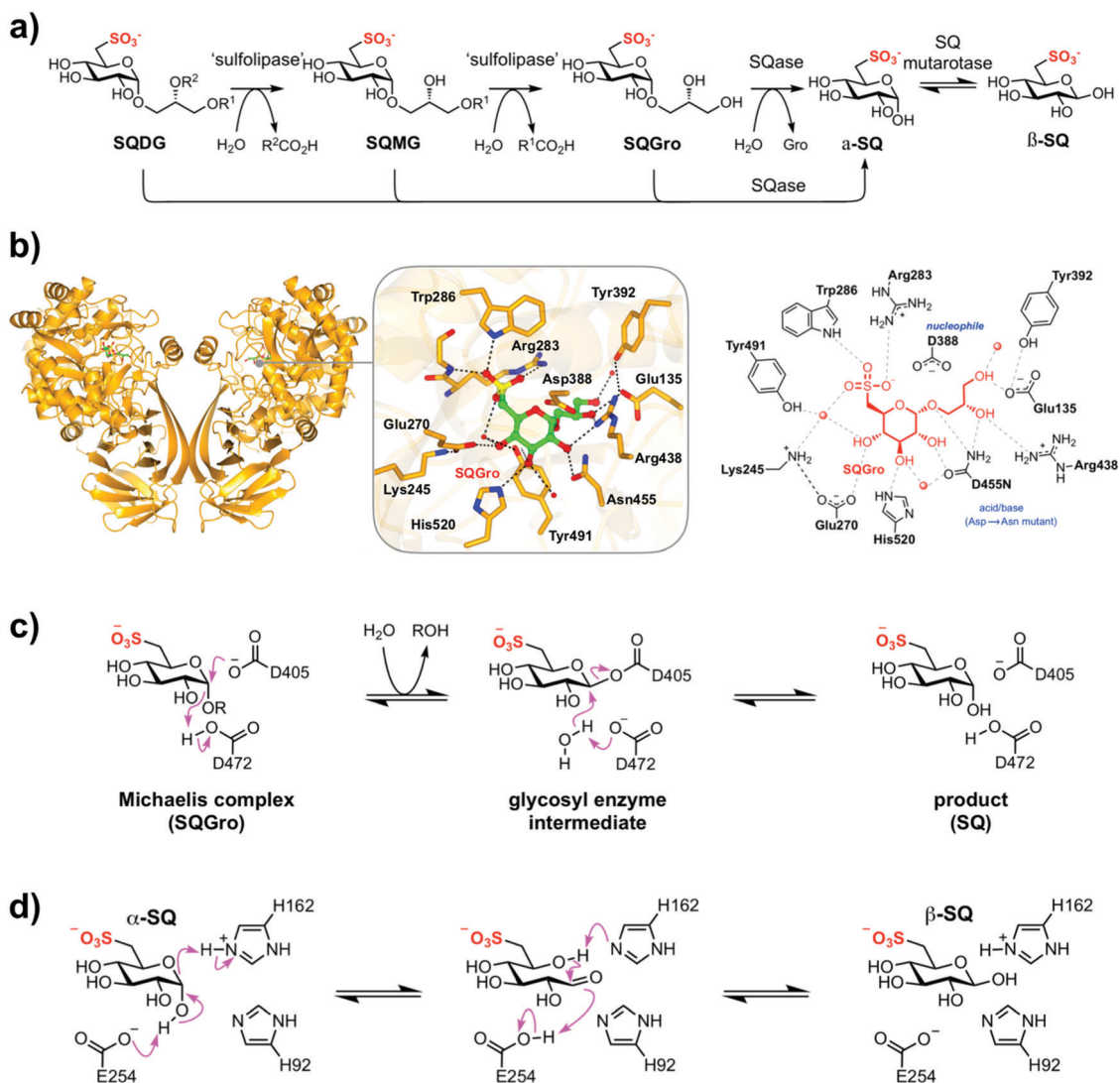


Fig. 6 SQDG catabolism preparatory phase. (a) The initial steps before entering sulfoglycolysis for breakdown of SQDG involve delipidation, glycoside cleavage, and SQ mutarotation. (b) Sulfoquinovosidases are a 'gateway' enzyme that liberates SQ for sulfoglycolysis. 3D structure of pseudo Michaelis complex of SQGro-bound to inactive acid/base mutant of Smol sulfoquinovosidase (SQase) (7OFX) from *A. tumefaciens*. Close-up view of the active site (centre) and cartoon (right) showing conserved RYW sulfonate binding motif and catalytic residues of SQases. (c) Catalytic retaining mechanism of SQases produce α -SQ; residue numbering for *A. tumefaciens* Smol. (d) Catalytic mechanism proposed for SQ mutarotase; residue numbering for *H. seropedicaea* SQM.

The hydrolysis of SQDG has been studied in mammalian systems. Pancreatin (an enzymatic extract from pig pancreas) can effect the deacylation of SQDG.⁴ Gupta and Sastry studied saline extracts of acetone-precipitated pancreas and intestinal mucosa from guinea pig, sheep and rat and observed stepwise deacylation of SQDG to SQMG (ascribed to 'sulfolipase A') and to SQGro (ascribed to 'sulfolipase B').³⁸ Administration of ³⁵S-SQDG to guinea pigs resulted in rapid conversion to water soluble forms: SQGro and SO₄²⁻ supporting the conclusion that C-S bond cleavage is brought about by intestinal microflora.³⁸

3.2 Hydrolysis of sulfolipid and SQ glycosides

Liberation of SQ from its glycosides is achieved by specialized glycosidases termed sulfoquinovosidases (SQases) (Fig. 6a). While early reports suggested that *E. coli* β -galactosidase can act as an SQase,³⁹ this observation was subsequently reconized to be erroneous,⁴⁰ and in retrospect the most reasonable explanation is contamination, perhaps by the *E. coli* YihQ protein. All characterized sulfoglycolysis gene clusters characterized to date contain a gene encoding an SQase belonging to glycoside hydrolase family 31 of the CAZy sequence-based classification (www.cazy.org;⁴¹ www.cazypedia.org⁴²): YihQ from *E. coli*,⁴³ SftG from *Bacillus aryabhatai*,²⁵ SqvC from *Bacillus megaterium*,²⁶ PpSQ1_00425 from *Pseudomonas putida* SQ1,²⁴ RISQase from *Rhizobium leguminosarum* SRDI565,⁴⁴ and SmoI (AtSQase) from *A. tumefaciens*.

Phylogenies of family GH31 proteins using hidden Markov models (a statistical method used for pattern recognition)⁴⁵ revealed that SQases occupy a small sub-group in this large family that is dominated by α -glucosidases and includes α -xylosidase, α -galactosidases, α -N-acetylgalactosaminidases and α -glucan lyases.^{46,47} SQases operate through a retaining mechanism involving a two-step double displacement *via* a glycosyl-enzyme intermediate.⁴³ In the first step a general acid residue (Asp472 in *E. coli* YihQ) assists the departure of the leaving group while a second carboxylate (Asp405) acts as a nucleophile to form the glycosyl-enzyme intermediate with inversion of anomeric stereochemistry. In the second step Asp472 acts as a general base to deprotonate a water molecule and assist the hydrolysis of the glycosyl-enzyme with a second inversion of anomeric stereochemistry, leading to release of α -SQ (Fig. 6c). YihQ SQase acts on both SQDG and SQGro and displays a 6-fold preference for the natural 2'R-SQGro stereoisomer, which is also preferred by SmoI.⁴⁸ 4-Nitrophenyl α -D-sulfoquinovoside (PNPSQ) is an effective chromogenic substrate for *E. coli* YihQ,⁴³ *A. tumefaciens* SQase⁴⁸ and *Rhizobium leguminosarum* SQase,⁴⁴ enabling their characterization using real-time assays with a UV/vis spectrophotometer. On the other hand the fluorogenic substrate 4-methylumbelliferyl α -D-sulfoquinovoside (MUSQ) is a poor substrate for YihQ and SmoI with $k_{\text{cat}}/K_{\text{M}}$ values some 10⁴–10⁵-fold lower, indicating that the SQase active site has difficulty accommodating the bulky methylumbelliferone group.⁴⁹

X-Ray structures of two SQases have been reported that reveal structurally homologous (α/β)₈ barrel folds.^{43,48} Complexes of wildtype *E. coli* YihQ and *A. tumefaciens* SmoI enzymes with the aza-sugar inhibitor sulfoquinovose-isofagomine (SQ-IFG), and of

the catalytically disabled acid–base mutant with the artificial substrate PNPSQ identified a conserved sulfonate binding pocket comprised of Arg301, Trp304 and Tyr508 *via* a bridging water molecule (*E. coli* YihQ numbering). The RWY sulfonate binding triad is present within predicted GH31 enzymes from plants, bacteria, fungi, animals and protists. Second sphere residues around the active site are largely conserved, but with two subgroups containing either QQ (*E. coli* Q262–Q288) or KE (*A. tumefaciens* K245–E270) residues.⁴⁸ Mutant enzymes with partially switched second-shell residues were around 1000-fold less active than wildtype. Mutants with fully swapped neutral pairs were less active than wildtype but were 10-fold more active than the partially swapped mutants, showing that charge neutrality is required for optimum SQase activity (Fig. 6b).

SQase activity serves to explain the ability of *E. coli* K-12, *R. leguminosarum* SRDI585, *P. putida* SQ1 and *B. aryabhatai* SOS1 to grow on SQGro as sole carbon source.^{25,44,48} In the case of *E. coli*, growth on SQGro led to cell densities similar to growth on glucose, while growth on SQ led to similar density to Gro, consistent with the utilization of 6C and 3C in the respective substrate pairs.⁴⁸

3.3 Sulfoquinovose mutarotase

Sulfoglycolysis gene clusters from sulfo-EMP and sulfo-ED pathways typically contain genes annotated as aldose-1-epimerase (also termed mutarotases) that are upregulated upon growth on SQ (Fig. 2 and 3).^{23,24} Mutarotases catalyse the interconversion of anomers of sugars, and in the case of SQ, mutarotation is a relatively slow process, with a half-life of 300 min under phosphate-free conditions at 26 °C.⁵⁰ An SQ mutarotase (*HsSQM*) from the putative sulfo-ED gene cluster from *Herbaspirillum seropedicae* AU14040 has been experimentally studied using the nuclear magnetic resonance spectroscopy technique of chemical exchange spectroscopy.⁵⁰ *HsSQM* catalyzes the mutarotation of SQ and glucose-6-phosphate (G6P), but not glucuronic acid, mannose, glucose or galactose (Fig. 6d). Unidirectional rate constants measured at equilibrium revealed that the $k_{\text{cat}}/K_{\text{M}}$ value for SQ mutarotation is 5-fold higher than for G6P. Combining these results with the reported rate for spontaneous mutarotation of G6P ($t_{1/2} = 6$ s),^{51,52} allowed calculation of the proficiency ratio ($k_{\text{cat}}/K_{\text{M}}/k_{\text{uncat}}$), which revealed that *HsSQM* is 17 000-fold more proficient as a catalyst for enhancing the rate of SQ mutarotation compared to G6P.⁵⁰ *HsSQM* shares highly conserved residues with other hexose mutarotases: His92, His162 and Glu254.⁵⁰ A histidine and Glu254 are proposed to act in roles of general acid and general base,^{53,54} respectively, in the first half of the reaction leading to the acyclic aldehyde.⁵⁰

4. The sulfoglycolytic Embden–Meyerhof–Parnas (sulfo-EMP) pathway

Denger *et al.* and co-workers reported that *E. coli* K-12 (strains BW25113, DH1, MG1655 and W3100) grow on SQ under aerobic conditions.²³ Working with strain MG1655 they observed that

consumption of SQ coincided with release of DHPS into the culture media. Addition of a DHPS-degrading bacterium, *Cupriavidus pinatubonensis* JMP134, to the spent culture medium resulted in consumption of DHPS and production of sulfate. By comparative two-dimensional polyacrylamide gel electrophoresis (2D SDS-PAGE) of glucose and SQ-grown *E. coli*, and peptide fingerprinting-mass spectrometry sequencing of differentially upregulated proteins excised from the gel, the responsible gene cluster was identified as *yihO-yihW* (Fig. 2).²³ This operon had previously been assigned as encoding O-antigen biosynthesis in *Salmonella enterica*.⁵⁵ Single-gene knockouts of *yihO*, *yihS*, *yihT* and *yihV* in strain BW25113 did not grow on SQ, supporting the contribution of these genes to SQ catabolism. This same gene cluster contributes to growth of *E. coli* K-12 under anaerobic conditions, where a mixed type of fermentation produces DHPS, succinate, acetate and formate.⁵⁶

The sulfo-EMP gene cluster of *E. coli* encodes: a predicted transporter (YihO) for importing SQ and SQ glycosides; an SQase (YihQ) to hydrolyze SQ glycosides (*vide supra*); SQ mutarotase (YihR) to catalyze conversion of α -SQ to β -SQ (*vide supra*); aldose-ketose isomerase (YihS) to isomerize SQ to sulfofructose (SF); SF kinase (YihV) to phosphorylate SF to SF-1-phosphate (SFP); SFP aldolase (YihT) to convert SFP into dihydroxyacetone phosphate and SLA; SLA reductase (YihU) to convert SLA into DHPS; and a second predicted transporter (YihP) that may export DHPS from the cell. *yihW* was renamed *csqR* and encodes a DeoR-type transcription factor for the operon. Overall, the sulfo-EMP pathway shares remarkable similarity with the EMP pathway of glycolysis.⁵⁷ However, consistent with specialization to act on SQ, the sulfo-EMP pathway is induced upon growth on SQ and enzymes within the pathway have undetectable reactivity on the corresponding intermediates in glycolysis.^{58,59} This is likely to be of significance as growth under sulfoglycolytic conditions requires gluconeogenesis to produce intermediates to supply the pentose phosphate pathway (PPP) and cell wall biogenesis. If the sulfoglycolytic enzyme SF kinase acted on G6P, this would result in a futile cycle that would consume ATP. Bioinformatics analysis showed that the sulfo-EMP pathway is present in the majority of commensal and pathogenic *E. coli* strains, and a wide range of other Enterobacteriaceae including *Salmonella enterica*, *Chronobacter sakazakii*, *Klebsiella oxytoca* and *Pantoea anantistis*, suggesting a role in sulfoglycolysis in the gastrointestinal tract of omnivores and herbivores.²³

4.1 Sulfoquinovose-sulfofructose isomerase (YihS)

SQ-SF isomerase activity was demonstrated for *E. coli* YihS using recombinantly expressed protein and demonstrating conversion by liquid chromatography-mass spectrometry.²³ Sharma *et al.* re-examined the activity of YihS (and the homolog from *S. enterica*, *SeYihS*) and showed that these proteins catalyze the equilibration of SQ to SF and sulforhamnose (SR, the C2-epimer of SQ) in an equilibrium ratio of 30:21:49, respectively.⁵⁸ YihS also acts as a glucose:fructose:mannose isomerase, with the activity for mannose as substrate 178-fold lower than for SQ.^{58,60} ¹H NMR spectroscopy was used to monitor H/D exchange at C2 catalyzed by YihS, which revealed

that YihS acts preferentially on β -SQ.⁵⁸ YihS is inactive on the glycolytic intermediate G6P.⁵⁸

YihS and *SeYihS* form hexamers in solution, with each monomer exhibiting an α_6/α_6 barrel fold.^{58,60} A 3D structure of a complex of *SeYihS*-H248A with β -SF helped define the active site architecture (Fig. 7a).⁵⁸ Overlay of the structure of the mutant complex with the wildtype structure showed that His248 and His383 are positioned on the 'α-face' of SF and two loops move to enclose the ligand in the active site. The sulfonate of SF is bound by Arg55-Gln379-Gln362; this active site architecture is shared with YihS. Collectively, the structural and biochemical data supports a mechanism involving deprotonation of C-2 in acyclic SQ to give a 1,2-enediol (Fig. 8a). Protonation at C-1 forms SF, protonation at C-2 from the bottom face forms SR while protonation at C-2 from the top face regenerates SQ.

4.2 Sulfofructose kinase (YihV)

Denger *et al.* demonstrated *E. coli* YihV is an SF kinase using recombinantly expressed protein and demonstrating ATP-dependent conversion of *in situ* generated SF to SFP by liquid chromatography-mass spectrometry,²³ which has been confirmed using chemically synthesized SF.⁶¹ Analysis of the thermal stability of YihV with SF, ADP, or both showed stabilisation upon sugar binding.⁵⁸ YihV activity towards SF is highly sensitive to ATP concentration: at [ATP] = 1 mM, K_M = 8 mM, while at [ATP] = 0.1 mM, K_M = 0.3 mM.⁵⁸ Modulation of activity by small molecules extends to other cellular metabolites: substrate inhibition by SF, product inhibition by ADP, and activation by SQ, SLA, fructose-6-phosphate (F6P), fructose biphosphate (FBP), phosphoenolpyruvate (PEP), dihydroxyacetone phosphate (DHAP), and citrate. This data suggest that YihV is an important control point for managing flux through sulfoglycolysis in the event of large fluctuations in substrate concentration and concentrations of cellular metabolites. This modulation of activity is reminiscent of the regulatory role played by phosphofructokinase (PFK), which catalyzes the first committed step in the EMP glycolysis pathway.^{62,63} YihV is inactive on the glycolytic intermediate F6P, showing no potential for cross-talk between sulfoglycolysis and glycolysis/gluconeogenesis.⁵⁸

A series of 3D structures of apo YihV, and complexes of YihV with SF and ATP analogue adenylyl-imidodiphosphate (AMPPNP) or ADP, and with SFP were determined by X-ray crystallography (Fig. 7b).⁵⁸ YihV is a pfkB-family ribokinase⁶⁴ with a characteristic nucleotide-binding domain and a β -domain forming a lid to enclose the substrate binding site. YihV is dimeric with a β -clasp forming the interface, with residues from the dimer interface on each subunit protruding into the active site of their counterpart, which indicates enzyme activity is dependent on dimerization. The lid and nucleotide binding domains have open and closed conformations, related by an interdomain rotation and closed upon binding of SF. SF binds in the cleft between the nucleotide binding domain and the lid, with a sulfonate-binding pocket formed by Arg138 and Asn109 from one subunit and Lys27 from the other dimer subunit. Lys27 also forms hydrogen bonds to the C1 hydroxyl and ring oxygen

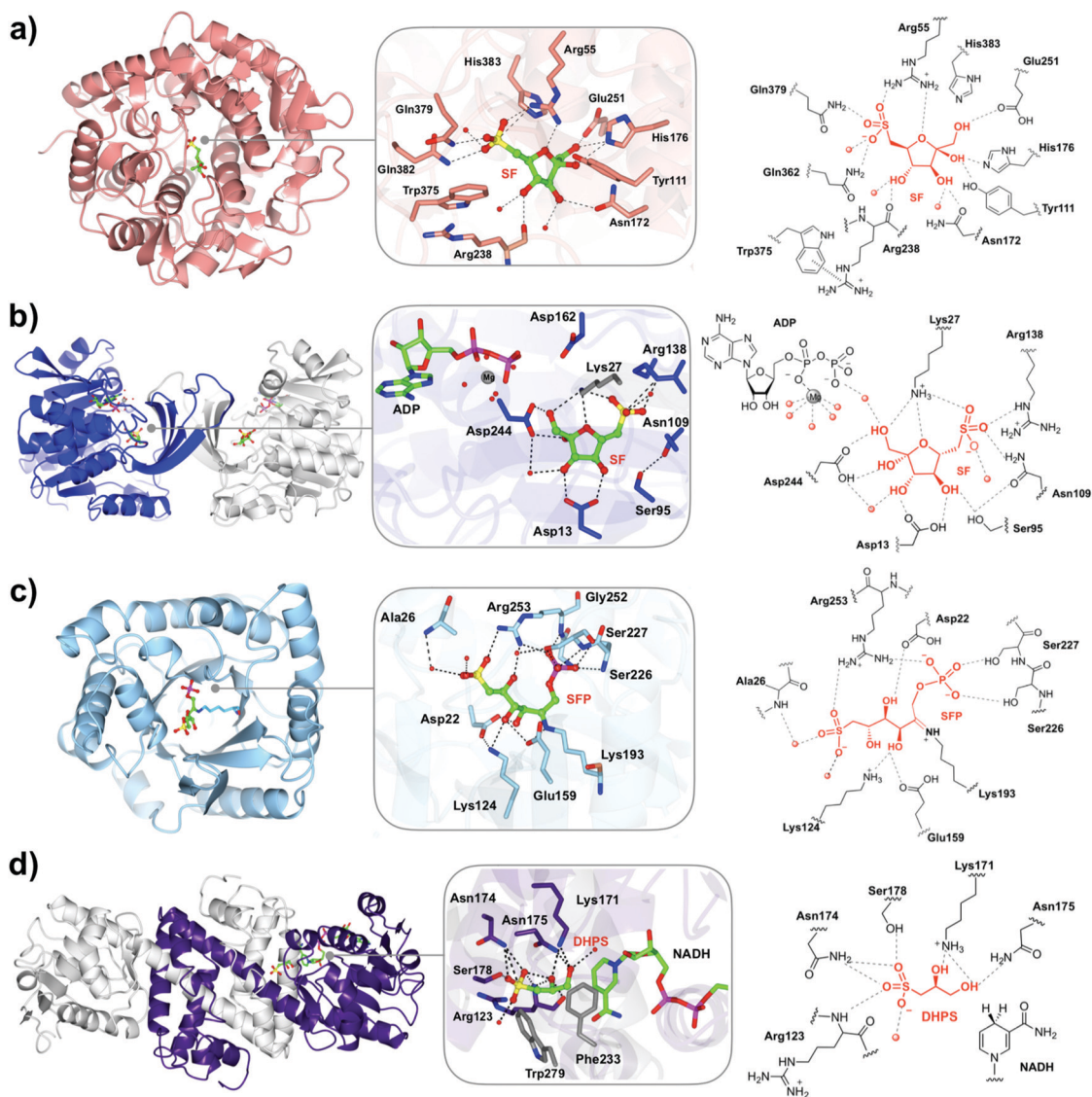


Fig. 7 3D structures of four enzymes mediating SQ breakdown through the sulfo-EMP pathway. (a) 3D structure of *S. enterica* SQ/SF isomerase (YihS) showing SF-bound active site view (7AG4) (left) and cartoon (right). (b) Crystal structure of *E. coli* SF kinase (YihV) bound to ADP, Mg^{2+} and SF (7AG6) depicting the β -barrel dimer motif in blue and grey, active site (centre) and cartoon (right). (c) 3D structure of *S. enterica* SFP aldolase (YihT) bound to SFP (7NE2) as a Schiff base with active site (centre) and cartoon (right). (d) 3D structure of *E. coli* SLA reductase (YihU) in complex with NADH and DHPS (6SMY) showing active site view (centre) and cartoon (right). The intimate dimer pair with reciprocal domain-sharing is shown in purple and grey; the major solution state is a tetramer.

of SF. The YihV-SF-ADP- Mg^{2+} quaternary complex adopts the closed conformation, with the putative catalytic base Asp244 forming a hydrogen bond to the C1 hydroxyl to facilitate the nucleophilic attack on the γ -phosphate of ATP and is consistent with a mechanism involving an in-line transfer of the phosphoryl group (Fig. 8b).

4.3 Sulfofructose-1-phosphate aldolase (YihT)

Denger *et al.* assigned SFP aldolase function to YihT using recombinantly-expressed protein and demonstrating conversion of *in situ* generated SFP to SLA and DHAP by liquid chromatography-mass spectrometry.²³ Activity has been demonstrated for the homologue *SeYihT* from *S. enterica* using

chemoenzymatically synthesized SFP,⁶¹ which was also shown to catalyze the reverse reaction, condensation of SLA and DHAP to give SFP,⁵⁸ in concordance with early observations by Benson using rabbit muscle aldolase.^{19,29} *SeYihT* is inactive on the glycolytic intermediate fructose-1,6-bisphosphate.⁵⁸

Sequence analysis defines YihT as a class I aldolase, a group of enzymes that catalyze a retro-aldol reaction using an active-site lysine to form a Schiff base intermediate.⁶⁵ *EcYihT* and *SeYihT* have a TIM-barrel fold and share close structural homology with other class I aldolases (Fig. 7c).⁵⁸ *SeYihT* is a tetramer in solution and crystallized with 12 independent molecules in the unit cell (three tetrameric assemblies). Upon soaking with SFP, a covalent Schiff base complex of SFP with Lys193 was obtained in two of the

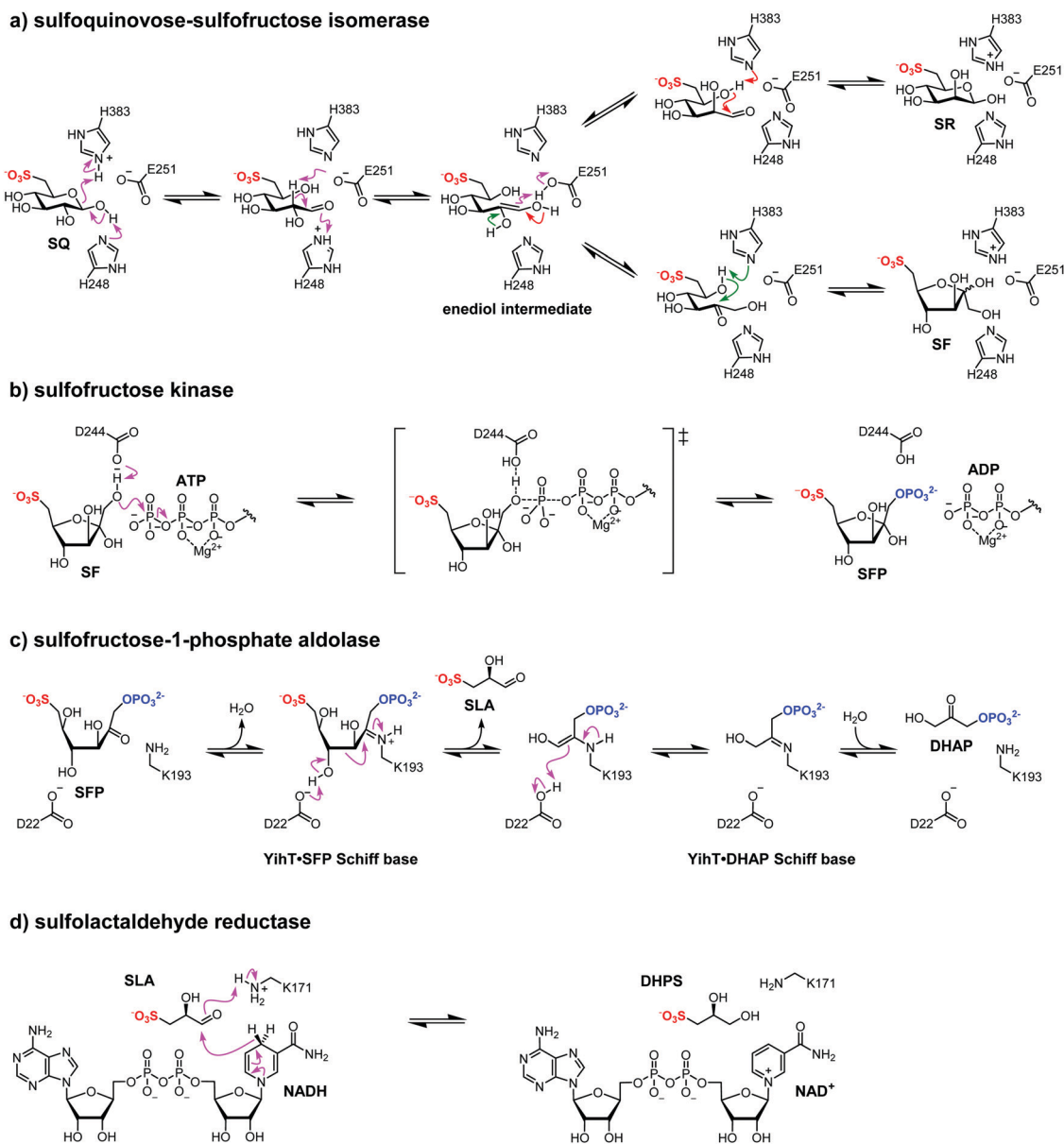


Fig. 8 Mechanism of enzymes in the sulfo-EMP pathway. (a) Mechanism for interconversion of SQ, SR and SF catalyzed by SQ-SF isomerase. (b) Mechanism for ATP-dependent SF kinase. (c) Mechanism for SFP aldolase, a class I aldolase. (d) Mechanism for NADH-dependent SLA reductase.

four subunits in the tetramers, and a covalent Schiff base complex with DHAP was obtained in the other two subunits.⁵⁸ Collectively, these data are consistent with the initial formation of a Schiff base with SFP, cleavage of the C3–C4 bonds to generate a Schiff base with DHAP and release of SLA, and finally hydrolysis of the DHAP Schiff base to release DHAP (Fig. 8c). The sulfonate group is recognized by a diad of conserved Arg253–Ala26, the latter *via* a bridging water molecule.

4.4 Sulfolactaldehyde reductase (YihU)

The final chemical step in the *E. coli* sulfo-EMP pathway is the reduction of SLA to DHPS. Denger *et al.* assigned YihT as DHPS reductase using recombinantly-expressed protein and demonstrating NADH-dependent conversion of *in situ*

generated SLA to DHPS by liquid chromatography–mass spectrometry; NADPH was not a substrate.²³ Activity of recombinant YihU was confirmed using chemically-synthesized SLA.⁵⁹ YihU is inactive on the glycolytic intermediate glyceraldehyde-3-phosphate (GAP), ensuring that YihU would not consume NADH by reduction of this substrate.⁵⁹ A random sequential Bi Bi mechanism was demonstrated for YihU involving sequential binding of NADH then SLA followed by the reduction to give DHPS.⁵⁹ YihU catalyzes succinate semialdehyde reduction, but with a catalytic efficiency 42 000-fold worse than measured for reduction of SLA.^{59,66} YihU is inhibited by the reduced NADH analogues: tetrahydro- and hexahydro-NADH.

YihU is a member of the β -hydroxyacid dehydrogenase (β -HAD) family.⁶⁷ YihU forms a solution tetramer and crystallised

as a pair of intimately associated dimers (Fig. 7d).⁵⁹ The protein has a two-domain architecture containing an N-terminal nucleotide binding domain forming a Rossmann fold, and a C-terminal helical bundle, connected by an interdomain helix. The protein dimerises with a domain swap of the C-terminus: helix $\alpha 8$ from one monomer inserts into the bundle of the counterpart. 3D structures of apo YihU and a complexes with NADH, or NADH and DHPS have been obtained by X-ray crystallography.⁵⁹ The YihU-NADH structure adopts a closed conformation when compared to the ligand-free protein, with an 8° hinge motion along the interdomain helix, resulting in enclosure of NADH. NADH specificity is ascribed to a hydrogen bond between the 2-hydroxyl on the NADH ribose ring and Asp31, which appears unable to accommodate the 2'-phosphate of NADPH. In the YihU-NADH complex, a 10.4 Å long channel decorated with a pair of positively charged residues at the entrance leads to the active site. These cationic residues may allow recruitment of anionic SLA to the YihU-NADH complex. A ternary complex of YihU-NADH-DHPS was produced by soaking DHPS into YihU-NADH crystals. This showed 2S-DHPS bound through its hydroxyl groups to Lys171, and with the terminal hydroxyl group close to the dihydronicotinamide ring, consistent with a mechanism for reduction involving direct hydride transfer to SLA (Fig. 8d). A dedicated sulfonate binding pocket was identified comprised of Arg123-Asn174-Ser178 that is common to all putative SLA reductases.

In the YihU-NADH-DHPS complex Arg123 within the triplet Gly122-Arg123-Thr124 interacts with one sulfonate oxygen,⁵⁹ whereas in the complex of the closely-related tartronate semi-aldehyde reductase GarR from *Salmonella typhimurium* bound to the substrate analogue L-tartrate the equivalent (and conserved among classical carboxylate-processing β -HADs) residues Ser123-Gly124-Gly125 exhibit a 180° flip in the central glycine allowing the carboxylate oxygens to bind to the triplet of residues Ser123-Gly124-Gly125.⁶⁸ Site-directed mutagenesis of YihU was used to probe the effect of individually converting each residue from Gly122-Arg123-Thr124 to the corresponding residue in GarR.⁵⁹ While only minor changes were noted for k_{cat}/K_M for varying [NADH] at constant [SLA], k_{cat}/K_M values changed more dramatically when varying [SLA] at constant [NADH], consistent with these residues being important in the binding of SLA.

4.5 DeoR-like transcription factor (CsqR)

The transcription factor for the sulfo-EMP operon in *E. coli* was initially named YihW and was renamed Csqr.⁶⁹ Ishihama and co-workers identified binding of Csqr at two genomic sites: one inside the spacer between the yihUTS operon and the yihV gene, and the other upstream of the yihW gene itself (Fig. 2). Csqr is a repressor for all sulfo-EMP transcription units and binding of Csqr was de-repressed by SQ and SQGro,⁶⁹ and more weakly by SR.⁵⁸ Lactose, glucose and galactose did not affect DNA binding. Atomic force microscopy of Csqr in the presence of DNA containing the Csqr binding sequence revealed large aggregates consistent with a high level of cooperativity in binding to the target DNA that were alleviated upon de-repression by SQ or SR.^{58,69}

5. The sulfoglycolytic Entner-Doudoroff (sulfo-ED) pathway

Pseudomonas putida SQ1 was isolated by enrichment culture using SQ as sole carbon source from nearshore sediment in Lake Constance (Germany).²² Growth of SQ1 was coincident with production of equimolar SL into the growth media, and led to cell density that was approximately half that of growth on glucose.²² Genome sequencing⁷⁰ followed by comparative 2D SDS-PAGE of glucose and SQ-grown *E. coli*, and peptide fingerprinting-mass spectrometry sequencing of differentially upregulated proteins excised from the gel, as well as total-proteome analysis, led to identification of a sulfoglycolytic Entner-Doudoroff pathway encoded by *PpsQ1_00088-00100* (Fig. 3).²⁴ Five steps of the sulfo-ED pathway were reconstituted *in vitro* using recombinantly produced proteins, with LC-MS used to demonstrate conversion of SQ to SGL, SG, KDSG, SLA and pyruvate, and finally SL.

The sulfo-ED gene cluster of *P. putida* SQ1 encodes: a predicted importer for SQ and SQ glycosides and an exporter for SL (*PpsQ1_00435*, *PpsQ1_00440*); an SQase (*PpsQ1_00425*) to hydrolyze SQ glycosides; an SQ mutarotase (*PpsQ1_00415*) to catalyze the equilibration of α -SQ and β -SQ; an NAD⁺-dependent SQ dehydrogenase (*PpsQ1_00405*) to convert SQ to 6-deoxy-6-sulfogluconolactone (SGL); SGL lactonase (*PpsQ1_00410*) to hydrolyze SGL to 6-deoxy-6-sulfogluconate (SG); SG dehydratase (*PpsQ1_00400*) to convert SG to 2-keto-3,6-dideoxy-6-sulfogluconate (KDSG); KDSG aldolase (*PpsQ1_00455*) to cleave KDSG to SLA and pyruvate; and an NAD(P)⁺-dependent SLA dehydrogenase (*PpsQ1_00395*, *GabD*) that converts SLA to SL. The sulfo-ED pathway shares a striking similarity with the ED pathway of glycolysis, yet SQ dehydrogenase, SG dehydratase and KDSG aldolase were inactive on G6P, phosphogluconate and keto-deoxy-phosphogluconate, respectively, showing that these enzymes are highly specific for the sulfonated substrates. Similar gene clusters were identified in a limited range of *Pseudomonas putida* strains, other γ -proteobacteria, as well as α - and β -proteobacteria.

Li and co-workers have studied a sulfo-ED pathway in *Rhizobium leguminosarum* SRDI565 (isolated from soil in eastern Australia) that produces SL upon growth on SQ or SQGro.⁴⁴ This pathway shares genes encoding putative enzymes that are homologous to the *P. putida* SQ1 proteins, with a key difference being the absence of a putative SQ mutarotase. SRDI565 shared the putative TauE-type SL exporter with SQ1 but it instead possessed an ABC solute importer cassette involving a putative SQ binding protein, and ABC-type permease components, also identified in the sulfo-SMO pathway.²⁸ This suggests that the SQ importation machinery can be interchanged between different sulfoglycolytic gene clusters. Metabolomics analysis of crude cell extracts from SRDI565 grown on SQ revealed the presence of the canonical sulfo-ED pathway intermediate SG and the endproduct SL, as well as G6P and F6P, providing evidence that gluconeogenesis is used to satisfy metabolic requirements for the PPP and cell wall biogenesis.⁴⁴ In addition, low amounts of SF and DHPS were also detected, which were

proposed to arise from moonlighting activities of phosphoglucose isomerase and a non-specific reductase.

6. The sulfoglycolytic sulfofructose transaldolase (sulfo-SFT) pathway

Sulfoglycolytic pathways that lack the SF kinase and SFP aldolase of the sulfo-EMP pathway, but instead possess a transaldolase active on SF have been identified, termed the sulfo-SFT pathway (or the sulfoglycolytic transaldolase, sulfo-TAL pathway) (Fig. 4). *Bacillus aryabhatai* SOS1, was isolated from a maple leaf (Konstanz, Germany) and grows aerobically on SQ to produce SL.²⁵ The draft genome sequence, comparative proteomics and 2D SDS-PAGE were used to identify upregulated proteins from growth on SQ leading to identification of the *sftATXGIFDE* gene cluster. Similar results were reported in a contemporaneous study by Zhang and co-workers using *Bacillus megaterium* DSM1804 for equivalent genes named *sqvUABCD-slaB-SqvE*.²⁶

The core steps of the sulfo-SFT pathway were reconstituted *in vitro* with recombinant enzymes.^{25,26} This pathway contains: a putative SQase (SftG, SqvC), SQ-SF isomerase (SftI, SqvD), and SF transaldolase (SftT, SqvA), which uses SF and GAP as substrate to produce F6P, which can enter glycolysis, or SF and erythrose-4-phosphate to produce sedoheptulose-7-phosphate, which can enter the PPP, in both cases through the transfer of a glycerone (dihydroxyacetone) fragment. The pathways also contain an NAD⁺-dependent SLA dehydrogenase (SftD, SlaB) that oxidizes SLA to SL, and which is excreted into the growth media. Sulfo-SFT gene clusters are predominantly present within the classes Bacilli and Clostridia of the phylum Firmicutes, and representatives were also found in individual genomes of members of the phyla Fusobacteria, Chloroflexi, Actinobacteria, Spirochaetes, and Thermotogae.²⁵ Human gut Firmicutes *Enterococcus gilvus*, *Clostridium symbiosum* and *Eubacterium rectale* use the sulfo-SFT pathway for fermentative growth on SQ. However, while *E. gilvus* produced SL during fermentative growth on SQ, *C. symbiosum* and *E. rectale* produced DHPS, consistent with their genomes instead encoding an SLA reductase. One unresolved question is the function of a conserved gene encoding DUF4867 (domain of unknown function) in the sulfo-SFT clusters (SftX in *B. aryabhatai*, SqvB in *B. megaterium*).^{25,26}

Human fecal slurry-derived microcosms grown under anoxic conditions rapidly consume SQ to transiently produce DHPS and culminate in production of H₂S.⁴⁷ Levels of *E. rectale* and *Bilophila wadsworthia* increased strongly, and correlated with disappearance of SQ and DHPS, respectively. DHPS production by *E. rectale* was associated with a sulfo-SFT pathway and DHPS consumption by *B. wadsworthia* with a sulfite-lyase pathway through the HpsG–HpsH system involving HspG, a glycol radical sulfite-lyase, and HspH as its cognate activator.⁷¹ Sulfo-SFT pathway expression was two orders of magnitude higher than that of proteobacterial sulfo-EMP pathway expression, suggesting that *Enterobacteriaceae* are only minor contributors to SQ degradation in the human gut.

7. The sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway

All pathways described above involve the degradation of SQ in two tiers, with production of the C3-fragments DHPS or SL through sulfoglycolysis. A distinct pathway termed the sulfo-SMO pathway, reported in *Agrobacterium tumefaciens* C58, achieves the complete degradation of SQ in a single organism (Fig. 5).²⁸ *A. tumefaciens* grows on SQ but the only observable carbon metabolite produced in the growth media is bicarbonate, and growth is coincident with release of sulfite, suggestive of complete degradation of SQ. Comparative proteomics revealed the upregulated proteins from growth on SQ derive from the gene cluster *smoABCDEFGHI* (*Atu3277-3285*). This gene cluster encodes an ATP binding cassette (ABC) transporter system (SmoEGH) with an associated periplasmic SQ-binding protein (SmoF), a previously characterized SQase (SmoI) for hydrolysis of SQ glycosides,⁴⁸ an NAD(P)H-dependent flavin mononucleotide (FMN) reductase (SmoA), an FMN-dependent SQ monooxygenase (SmoC) that converts SQ to 6-oxo-glucose (6-OG), and a 6-OG reductase (SmoB) that converts 6-OG to glucose, which enters glycolysis. This sulfo-SMP pathway is distributed across α - and β -proteobacteria and is particularly prevalent among members of the *Agrobacterium* and *Rhizobium* genus within the *Rhizobiales* order. Some of these pathways lack the ABC cassette seen in *A. tumefaciens* C58.

7.1 Sulfoquinovose-binding protein (SmoF)

Solute binding proteins deliver solutes to ABC transporters to move small molecules across membranes.⁷² These systems use a homodimeric nucleotide binding domain to enforce conformational changes on a heterodimeric transmembrane domain pair depending on the state of ATP binding, hydrolysis and ADP dissociation. This conformational change opens a channel to the periplasm upon ATP binding and to the cytosol upon ATP hydrolysis. The binding protein recognizes the substrate and docks to the transporter when the channel is in the periplasm-open state. Recombinant SmoF binds SQGro with $K_d = 290$ nM and does not bind the stereochemically-related monosaccharides D-glucose and D-glucuronic acid.²⁸ 3D X-ray structures reveal a globular fold featuring a pair of α/β domains (Fig. 9a).²⁸ SQGro binds SmoF in a cleft between domains. Binding of SQGro causes a 31° hinge motion of its two domains resulting in complete enclosure of the ligand. Recognition is achieved through a sulfonate binding pocket, in which the sulfonate oxygens form hydrogen bonds to Thr220, Gyl166 and Ser43. An ordered water mediates an interaction to His13. The sugar C2-4 hydroxyls are also recognized through hydrogen-bonding interactions.

7.2 Flavin mononucleotide reductase (SmoA)

The genes encoding flavin reductase SmoA and monooxygenase SmoC are present within sulfo-SMO gene clusters in multiple organisms including *Agrobacterium* sp., *Rhizobium oryzae*, and *Aureimonas flava* suggestive that SmoA and SmoC comprise a two-component system that effects the desulfurization and oxidation of SQ.²⁸ Recombinantly-expressed *A. tumefaciens* SmoA

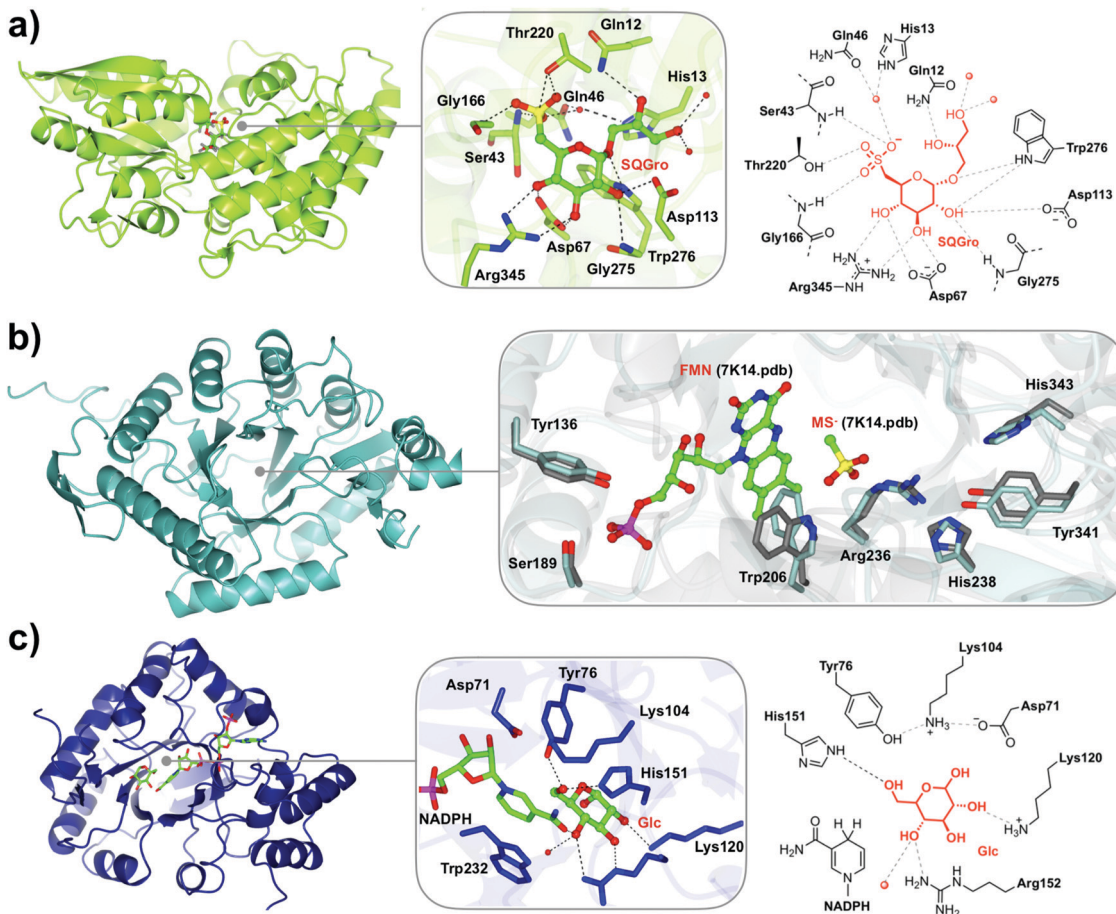


Fig. 9 3D structures of proteins mediating SQ transport and breakdown through the sulfo-SMO pathway. (a) X-ray structure of *A. tumefaciens* SmoF SQ binding protein (in green) showing SQGro-bound active site view (centre) and cartoon (right) (7OFY). (b) Overlay of X-ray structure of apo *R. oryzae* SmoC SQ monooxygenase (in cyan) vs a ternary complex of a model alkanesulfonate monooxygenase (7K14, in grey) showing putative sulfosugar and FMN binding residues in an active site view. (c) Crystal structure of *A. tumefaciens* SmoB 6-OG reductase (in dark blue) showing ternary complex with NADPH and glucose (7BC1).

is a pale yellow colour and heat-denaturing resulted in the release of FMN, identifying its preferred flavin. Kinetics with saturating FMN and NADH or NADPH revealed a preference for NADH. Thus, reduction of SmoA-bound FMN by hydride transfer from NAD(P)H supplies FMNH₂ to the monooxygenase partner SmoC. Similar results were obtained for the homologous enzyme from *R. oryzae*.

The SmoA and SmoC proteins share similarity with the *E. coli* SsuD and SsuE proteins expressed from the alkylsulfonate sulfur-utilization gene cluster (*ssuEADCB*).⁷³ SsuD–SsuE comprise a two-component alkane sulfonate monooxygenase active on a range of linear alkane sulfonates, which is upregulated under sulfur starvation.⁷⁴ Crystal structures and biophysical analysis of solution states (apo vs. FMN bound forms) of the smaller flavin reductase SsuE indicates FMN binding drives a tetramer–dimer equilibrium, possibly to enable association with the monooxygenase partner for transfer of reduced flavin.^{75,76} The mechanism and stoichiometry of association of SmoA–SmoC system is unknown.

7.3 Flavin mononucleotide-dependent sulfoquinovose monooxygenase (SmoC)

The oxidative cleavage of C–S bond of SQ is catalysed by the sulfoquinovose monooxygenase SmoC.²⁸ This reaction presumably

involves a C4a-(hydro)peroxyflavin intermediate as proposed for Lada⁷⁷ and SsuD⁷⁸ monooxygenases or an N5-peroxyflavin intermediate invoked for methanesulfonate monooxygenase MsuD and other monooxygenases.^{79,80} Combination of SmoC, FMN reductase SmoA, NADH or NADPH and FMN resulted in release of sulfite from SQ, with a preference for NADH. Activity appeared to be oxygen-dependent with conversion limited by the solubility of oxygen in aqueous solution. No activity was seen for SQGro or the sulfonate HEPES, thus revealing a clear preference for SQ. SmoC binds SQ with $K_d = 3 \mu\text{M}$, whereas no binding was detected for SQGro. The requirement for SmoA to reduce SmoC defines SmoC as a Category II two-component flavoprotein monooxygenase.⁸⁰

A 3D X-ray structure of a homologue of SmoC from the equivalent pathway from the syntenic gene cluster in *Rhizobium oryzae* revealed an $\alpha_8\beta_8$ barrel with three insertion regions (Fig. 9b). A low-resolution structure of SmoC from *A. tumefaciens* aligns well to this structure, showing high similarity. By overlay with the structurally-related methanesulfonate monooxygenase MsuD in complex with FMN a structural model of SmoC binding to FMN was generated. The isoalloxazine ring of FMNH₂ occupies a deep hydrophobic cleft present in both structures, and both

structures contain an identical sulfonate binding pocket comprised of Trp206, Arg236, His238, His343. For alkylsulfonate monooxygenase SsuD, substrate binding induces conformational changes through dynamic loop movements involving salt-bridge interactions of conserved arginines and glutamates distal to the active site.^{81,82}

The mechanism of sulfonate monooxygenases remains enigmatic, with disagreement even on the identity of the oxidized form of FMN and the site of oxygen attack. Possible mechanisms are proposed based on studies with MsuDs.⁷⁹ Initially, formation of a C4a-peroxy or N5-peroxyflavin species is proposed to occur by reaction with oxygen on-enzyme (Fig. 10a and c). One mechanism then proposes that the peroxide deprotonates C6, and the resulting carbanion is oxidized to an α -hydroxysulfonate that undergoes elimination to produce sulfite and 6-OG (Fig. 10a). An alternative mechanism suggests the terminal peroxide oxygen of either peroxyflavin

species attacks the sulfonate sulfur, which then undergoes rearrangement and effect C-S bond cleavage and release of the 6-OG and sulfite (Fig. 10b). Biophysical, kinetic and structural studies on the SmoA-SmoC two-component system are required to understand the kinetic mechanism, identifying protein-protein interactions and revealing residues that bind the sulfosugar and stabilize the C4a-(hydro)-peroxyflavin or N5-peroxyflavin intermediate involved in the catalytic mechanism of SmoC.

7.4 NADPH-dependent 6-oxo-glucose reductase (SmoB)

The final step of the sulfo-SMO pathway is reduction of 6-OG to form glucose, catalysed by SmoB.²⁸ SmoB is an NADPH-dependent reductase from the aldose-ketose reductase (AKR) superfamily.⁸³ SmoB bound NADPH with K_d 2 μ M and did not bind NADH. Incubation of SmoB with glucose, NADP⁺ and

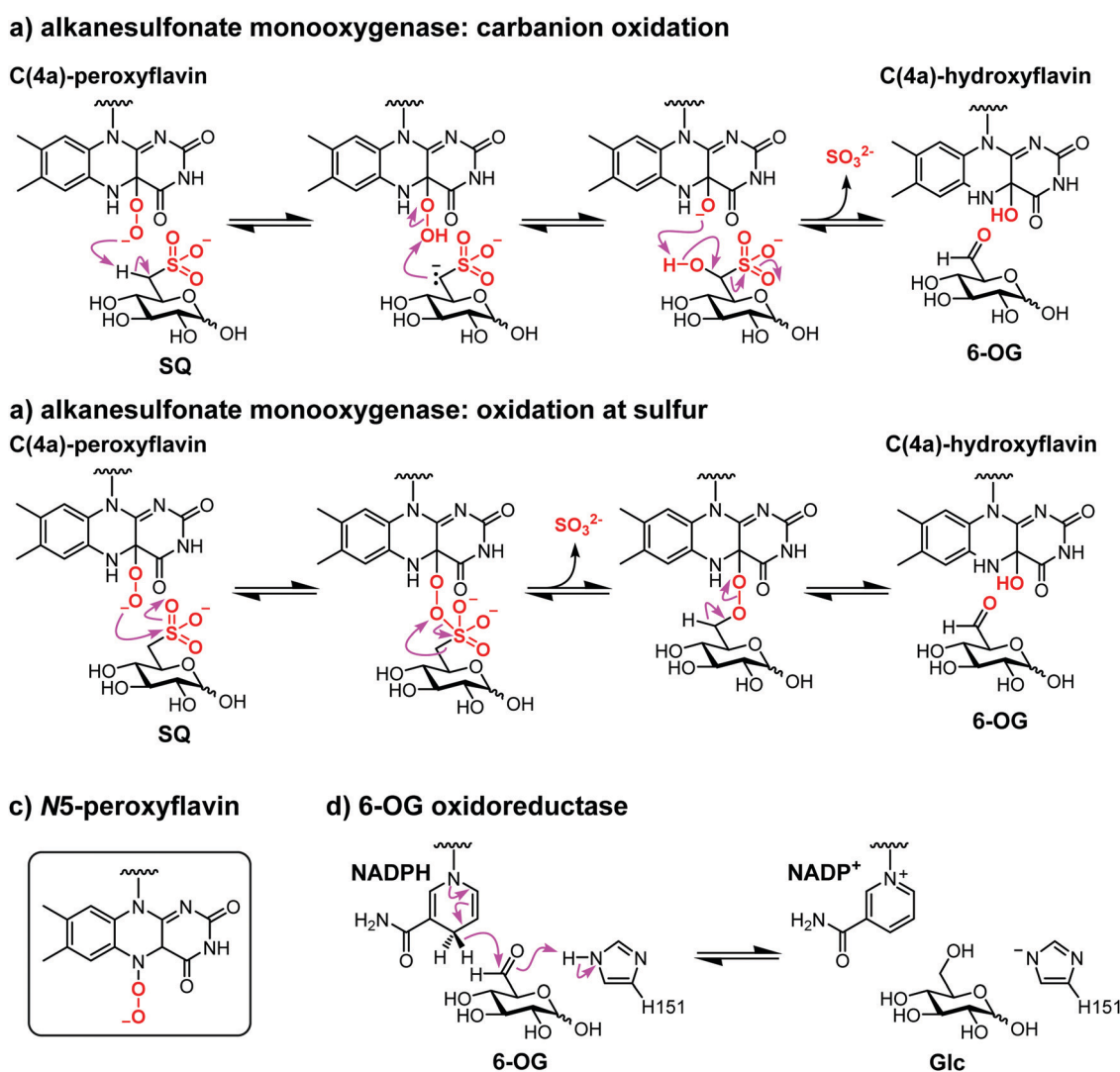


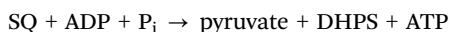
Fig. 10 Mechanism of enzymes in the sulfo-SMO pathway. Proposed mechanisms for conversion of SQ to 6-OG catalyzed by FMN-dependent SQ monooxygenase via a C4-peroxyflavin intermediate. (a) Carbanion mechanism. (b) Oxidation at sulfur mechanism. (c) Structure of an alternative N5-peroxyflavin. (d) Mechanism for NADPH-dependent 6-OG reductase.

H_2^{18}O resulted in incorporation of an ^{18}O -label at C6, as demonstrated by GC-MS (electron-ionisation).

Recombinant SmoB exists as a trimer in solution. The 3D structure of SmoB was obtained by X-ray crystallography and revealed a TIM-barrel fold (Fig. 9c). A ternary complex of SmoB, NADPH and glucose revealed the cofactor is held in an extended, *anti*-conformation over the center of the barrel by use of a C-terminal binding site, with the nicotinamide moiety positioned 3 Å from C6 of the substrate. In the SmoB-NADP⁺-glucose complex, glucose interacts with Lys120 (3 Å), 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. The complex is consistent with a proposed mechanism involving direct hydride transfer from NADH to the aldehyde of 6-OG (Fig. 10d).

8. Energy balance and carbon flux through sulfoglycolytic pathways.

Metabolism of sulfoquinovose may involve breakdown of SQDG, SQGro or SQ. The lack of lipases within sulfoglycolytic operons suggests that organisms have primarily evolved to breakdown SQGro or SQ after deacylation, which may be conducted by non-specific lipases providing an opportunistic fatty acid bounty for non-sulfoglycolytic and sulfoglycolytic organisms alike, with the metabolic yields as expected through the pathways of β -oxidation.⁸⁴ The four pathways of sulfoglycolysis involve SQ utilization with different outcomes in terms of carbon supply and production of reducing equivalents and ATP. The tier 1 pathways that utilize only half of the carbon of SQ and release C3 sulfonates are ascetic relative to the equivalent glycolysis pathways.²⁵ The sulfo-EMP pathway of *E. coli* under aerobic conditions (including substrate-level phosphorylation of released DHAP) can be represented by the equation:



For the sulfo-ED pathways of *P. putida* SQ1 and *R. leguminosarum* SRDI565 the reaction is:

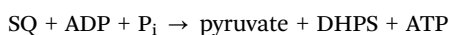


While the sulfo-ED pathway of *P. putida* SQ1 appears to use passive transporters, the pathway in *R. leguminosarum* SRDI565 contains an ABC transporter, suggesting that a single molecule of ATP is required to import SQ.

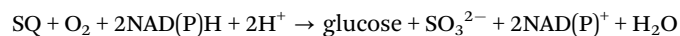
The sulfo-SFT pathways of *B. aryabhatai* SOS1 and *B. megaterium* DSM1804 including oxidation of SLA to SL and substrate level phosphorylation of DHAP to produce pyruvate the reaction is:



In the case of reduction of SLA to DHPS, such as in *E. rectale*, the sulfo-SFT process becomes:



The sulfo-SMO pathway of *A. tumefaciens* C58 is fundamentally different as the end-product of the reaction is glucose, which enters glycolysis, and can be represented as:



This analysis of the sulfo-SMO pathway excludes the investment of a single molecule of ATP during importation using the ABC transporter. For direct comparison with the tier 1 pathways, glycolytic breakdown of glucose to pyruvate must be included, as outlined in Table 1.

The metabolic logic of sulfoglycolysis changes for growth on SQGro, which is likely to be the more environmentally significant substrate. For the tier 1 sulfoglycolytic pathways, SQGro has 6 metabolizable carbons, and for the sulfo-SMO pathway 9 metabolizable carbons, assuming that the released glycerol is converted to GAP by the action of glycerol kinase and glycerol-3-phosphate (G3P) dehydrogenase, which may in turn act as an acceptor for SF transaldolase in the sulfo-SFT pathway or be transformed to pyruvate. Table 1 summarizes the metabolic yields using SQGro in each of these pathways. This analysis may lead to undue focus on the energetic and reducing equivalent yields of sulfoglycolysis and ignore the much greater yield from respiratory catabolism of pyruvate. Under aerobic conditions, pyruvate enters the citric acid cycle and undergoes oxidative phosphorylation leading to the net production of 32 ATP molecules. Sulfoglycolysis metabolic yields are likely to be more significant for growth under anaerobic conditions, where pyruvate may be subjected to fermentation (in the case of *E. coli* producing formate, acetate and succinate through mixed acid fermentation).⁵⁶

For bacteria grown exclusively under sulfoglycolytic conditions (*i.e.* growth on only SQ or SQGro), the carbon released from sulfoglycolysis needs to satisfy the energetic and metabolic requirements of growth. This requires partitioning of pyruvate into the citric acid cycle, fatty acid synthesis, cell wall biogenesis and the PPP. Unlike classical glycolysis pathways that produce G6P and F6P that can enter the PPP or support peptidoglycan synthesis, the sulfo-EMP and sulfo-ED pathways do not, and thus must generate these critical molecules by gluconeogenesis.

Table 1 Metabolic yields for sulfoglycolytic pathways utilizing SQ or SQGro, upon conversion to pyruvate

Pathway	Substrate	Pyruvate	ATP	NAD(P)H	Product
EMP	Glucose	2	2	2	
Sulfo-EMP	SQ	1	1	0	DHPS
	SQGro ^a	2	2	2	DHPS
ED	Glucose	2	1	2	
Sulfo-ED	SQ	1	0	2	SL
	SQGro ^a	2	1	4	SL
Sulfo-SFT	SQ	1	1	2	SL
	SQGro ^a	2	2	4	SL
Sulfo-SMO ^b	SQ + O ₂	2	2	-1	Sulfite
	SQGro ^a	3	3	1	Sulfite

^a Yields for SQGro assume conversion of released glycerol to pyruvate according to $\text{Gro} + \text{ADP} + 2\text{NAD}^+ + 2\text{H}^+ \rightarrow \text{pyruvate} + \text{ATP} + 2\text{NADH}$.

^b For the sulfo-SMO pathway, the metabolic yield is calculated for glycolysis of the released glucose to pyruvate using the EMP pathway.

Evidence for gluconeogenesis for *R. leguminosarum* SRDI565 (which uses a sulfo-ED pathway) has been obtained by metabolomic analysis of cells grown on SQ as sole carbon source, which revealed the presence of G6P and F6P, but at levels much lower than cells grown on glucose, consistent with a more ascetic lifestyle under sulfoglycolysis.⁴⁴ Both the sulfo-SFT and sulfo-SMO pathways are fundamentally different in that they generate hexose-6-P through either transaldolase/isomerase reactions or direct glycolysis, respectively, and cells therefore do not require gluconeogenesis to supply the PPP or cell wall biogenesis. The sulfo-SMO pathway represents an assimilation pathway, that supplies glucose to the cell, whereas the sulfo-SFT pathway rewires the metabolic connections, injecting F6P directly into the glycolytic pathway.

9. Conclusions

The discovery of new pathways for sulfoglycolysis has invigorated the study of SQ and its metabolites and helped close an important gap in the biogeochemical sulfur cycle. The striking similarity of sulfo-EMP and sulfo-ED pathways with their namesake glycolytic pathways hints at an evolutionary relationship, possibly arising from gene duplication and neofunctionalization. A connection exists with the key enzyme of the sulfo-SFT pathway, sulfofructose transaldolase, and transaldolase within the PPP, which converts sedoheptulose-7-phosphate and GAP to erythrose-4-phosphate and F6P. The SQase (SmoI), SQGro binding protein (SmoF) and SmoE/G/H (ABC transporter) of the sulfo-SMO pathway share similarity with MalP (maltodextrin phosphorylase), MalE (maltose binding protein) and MalF/G/K (ABC transporter) encoded by the *mal* operon of *E. coli* that degrades maltodextrin,⁸⁵ while SmoC (SQ monooxygenase) and SmoA (flavin reductase) of the SMO pathway are reminiscent of SsuD (FMN₂-dependent alkylsulfonate monooxygenase) and SsuE (NADPH-dependent FMN reductase) encoded by the *ssu* operon of *E. coli* that degrades alkanesulfonates.⁸⁶ Like the *smo* gene cluster, the *ssu* operon also encodes an ABC transporter encoded by *ssuABC* that constitutes an uptake system for alkane sulfonates.⁸⁶

The widespread distribution of sulfoglycolysis pathways in bacteria reflects the diverse niches in which photosynthetic tissues are found. These include within soils, waterways and the digestive tract of herbivores and omnivores, which appear to be replete with bacteria that catabolize only half of the SQ carbon through tier 1 pathways and require other bacteria to catabolise the released SL or DHPS through tier 2 biomineralization pathways.^{1,27,87} The tier 1 sulfoglycolytic pathways are 'generous' in the sense that they support diverse microbial communities by keystone SQ-degrading organisms that release C3-sulfonates. However, this support also operates in the reverse direction as biomineralized sulfur (as sulfate/sulfite) will supply assimilatory sulfur metabolism by tier 1 organisms. The sulfo-SMO pathway is distinct in that it allows the complete metabolism of SQ and release of sulfite that can enter sulfur assimilation pathways; thus this pathway may support a more

selfish microbial lifestyle within competitive environments. The occurrence of the sulfo-SMO pathway within specialized mutualistic (both symbiotic and pathobiotic) bacteria that grow on or within plants suggest that it may support nutrient acquisition even in the absence of other bacteria.

Sulfoglycolytic pathways are of fundamental interest as they inform our understanding of nutrient and energy acquisition from an unusual sugar with striking resemblance to glucose. However, the energy, reducing equivalents and carbon yields upon sulfoglycolysis *versus* glycolysis varies greatly. Understanding whether sulfoglycolysis leads to limitations in one or more of these outputs will be critical to understanding the metabolic consequences of utilizing sulfoglycolysis *versus* glycolysis. Our growing understanding of the structural basis for how nature recognizes the defining sulfonate group present in SQ and its metabolites is already enhancing the accuracy of bioinformatic methods and supporting new discoveries on the contribution of sulfoglycolysis to biomedically important processes.⁴⁷ Future work is needed to uncover the molecular basis of sulfoglycolysis pathways in plants and algae to help understand SQ cycling within phototrophs.

Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
DHAP	Dihydroxyacetone phosphate
DHPS	2,3-Dihydroxypropane sulfonate
FBP	Fructose biphosphate
FMN	Flavin mononucleotide
F6P	Fructose-6-phosphate
G6P	Glucose-6-phosphate
β-HAD	β-Hydroxyacid dehydrogenase
KDSG	2-Keto-3,6-deoxy-6-sulfoluconate
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate)
6-OG	6-Oxo-glucose
PPP	Pentose phosphate pathway
PEP	Phosphoenolpyruvate
PFK	Phosphofructokinase
SF	Sulfofructose
SFP	Sulfofructose-1-phosphate
SG	Sulfoluconate
SGL	6-Deoxy-6-sulfoluconolactone
SL	Sulfolactate
SLA	Sulfolactaldehyde
SQ	Sulfoquinovose
SQDG	Sulfoquinovosyl diacylglycerol
SQGro	Sulfoquinovosyl glycerol
SQMG	Sulfoquinovosyl monoglyceride
SR	Sulforhamnose
sulfo-ED	Sulfoglycolytic Entner-Doudoroff
sulfo-EMP	Sulfoglycolytic Embden-Meyerof-Parnas
sulfo-SFT	Sulfoglycolytic transaldolase
sulfo-SMO	Sulfoglycolytic sulfoquinovose monooxygenase
UDP	Uridine-5'-diphosphate.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

We thank the Australian Research Council (DP210100233, DP210100235; grants to SJW), The Leverhulme Trust (RPG-2017-190; support for A. J. D. S. and M. S.) and The Royal Society Chemistry for Ken Murray Research Professorship to G. J. D.

References

- 1 E. D. Goddard-Borger and S. J. Williams, *Biochem. J.*, 2017, **474**, 827–849.
- 2 C. Benning, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 1998, **49**, 53–75.
- 3 A. A. Benson, H. Daniel and R. Wisser, *Proc. Natl. Acad. Sci. U. S. A.*, 1959, **45**, 1582–1587.
- 4 T. Yagi and A. A. Benson, *Biochim. Biophys. Acta*, 1962, **57**, 601–603.
- 5 Y. H. N. Fusetani, *Agric. Biol. Chem.*, 1975, **39**, 2021–2025.
- 6 N. Oku, A. Hasada, K. Kimura, H. Honoki, R. Katsuta, A. Yajima, T. Nukada, K. Ishigami and Y. Igarashi, *Chem. – Asian J.*, 2021, **16**, 1493–1498.
- 7 W. R. Riekhof, M. E. Ruckle, T. A. Lydic, B. B. Sears and C. Benning, *Plant Physiol.*, 2003, **133**, 864–874.
- 8 U. Zähringer, H. Moll, T. Hettmann, Y. A. Knirel and G. Schäfer, *Eur. J. Biochem.*, 2000, **267**, 4144–4149.
- 9 K. Kobayashi, *J. Plant Res.*, 2016, **129**, 565–580.
- 10 N. Mizusawa and H. Wada, *Biochim. Biophys. Acta*, 2012, **1817**, 194–208.
- 11 Y. Nakajima, Y. Umena, R. Nagao, K. Endo, K. Kobayashi, F. Akita, M. Suga, H. Wada, T. Noguchi and J.-R. Shen, *J. Biol. Chem.*, 2018, **293**, 14786–14797.
- 12 G. Hözl and P. Dörmann, *Annu. Rev. Plant Biol.*, 2019, **70**, 51–81.
- 13 M. Frentzen, *Curr. Opin. Plant Biol.*, 2004, **7**, 270–276.
- 14 B. Kalisch, P. Dörmann and G. Hözl, *Subcell. Biochem.*, 2016, **86**, 51–83.
- 15 J. L. Harwood and R. G. Nicholls, *Biochem. Soc. Trans.*, 1979, **7**, 440–447.
- 16 E. Celik, M. Maczka, N. Bergen, T. Brinkhoff, S. Schulz and J. S. Dickschat, *Org. Biomol. Chem.*, 2017, **15**, 2919–2922.
- 17 J. S. Dickschat, P. Rabe and C. A. Citron, *Org. Biomol. Chem.*, 2015, **13**, 1954–1968.
- 18 K. Thume, B. Gebser, L. Chen, N. Meyer, D. J. Kieber and G. Pohnert, *Nature*, 2018, **563**, 412–415.
- 19 A. A. Benson and I. Shibuya, *Fed. Proc.*, 1961, **20**, 79.
- 20 A. B. Roy, A. J. Ellis, G. F. White and J. L. Harwood, *Biochem. Soc. Trans.*, 2000, **28**, 781–783.
- 21 A. B. Roy, M. J. Hewlins, A. J. Ellis, J. L. Harwood and G. F. White, *Appl. Environ. Microbiol.*, 2003, **69**, 6434–6441.
- 22 K. Denger, T. Huhn, K. Hollemeyer, D. Schleheck and A. M. Cook, *FEMS Microbiol. Lett.*, 2012, **328**, 39–45.
- 23 K. Denger, M. Weiss, A. K. Felux, A. Schneider, C. Mayer, D. Spiteller, T. Huhn, A. M. Cook and D. Schleheck, *Nature*, 2014, **507**, 114–117.
- 24 A. K. Felux, D. Spiteller, J. Klebensberger and D. Schleheck, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E4298–4305.
- 25 B. Frommeyer, A. W. Fiedler, S. R. Oehler, B. T. Hanson, A. Loy, P. Franchini, D. Spiteller and D. Schleheck, *iScience*, 2020, **23**, 101510.
- 26 Y. Liu, Y. Wei, Y. Zhou, E. L. Ang, H. Zhao and Y. Zhang, *Biochem. Biophys. Res. Commun.*, 2020, **533**, 1109–1114.
- 27 Y. Wei and Y. Zhang, *Annu. Rev. Biochem.*, 2021, **90**, 817–846.
- 28 M. Sharma, J. P. Lingford, M. Petricevic, A. J. P. Snow, Y. Zhang, M. Järvå, J. W.-Y. Mui, N. E. Scott, E. C. Saunders, R. Mao, R. Epa, B. M. da Silva, D. E. V. Pires, D. B. Ascher, M. J. McConville, G. J. Davies, S. J. Williams and E. D. Goddard-Borger, *ChemRxiv*, 2021, DOI: 10.33774/chemrxiv-32021-33708j33701.
- 29 I. Shibuya, T. Yagi and A. A. Benson, *Plant Cell Physiol.*, 1963, 627–636.
- 30 R. F. Lee and A. A. Benson, *Biochim. Biophys. Acta*, 1972, **261**, 35–37.
- 31 S. Scholz, M. Serif, D. Schleheck, M. D. J. Sayer, A. M. Cook and F. C. Küpper, *Bot. Mar.*, 2021, **64**, 301–312.
- 32 T. C. Strickland and J. W. Fitzgerald, *Soil Biol. Biochem.*, 1983, **15**, 347–349.
- 33 H. L. Martelli and A. A. Benson, *Biochim. Biophys. Acta*, 1964, **93**, 169–171.
- 34 M. G. Wolfersberger and R. A. Pieringer, *J. Lipid Res.*, 1974, **15**, 1–10.
- 35 D. Dougal Burns, T. Galliard and J. L. Harwood, *Phytochemistry*, 1980, **19**, 2281–2285.
- 36 G. Hazlewood and R. M. C. Dawson, *Microbiology*, 1979, **112**, 15–27.
- 37 I. Shibuya and E. Hase, *Plant Cell Physiol.*, 1965, **6**, 267–283.
- 38 S. D. Gupta and P. S. Sastry, *Arch. Biochem. Biophys.*, 1987, **259**, 510–519.
- 39 I. Shibuya and A. A. Benson, *Nature*, 1961, **192**, 1186–1187.
- 40 H. Daniel, M. Miyano, R. O. Mumma, T. Yagi, M. Lepage, I. Shibuya and A. A. Benson, *J. Am. Chem. Soc.*, 1961, **83**, 1765–1766.
- 41 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490–495.
- 42 The CAZypedia Consortium, *Glycobiology*, 2018, **28**, 3–8.
- 43 G. Speciale, Y. Jin, G. J. Davies, S. J. Williams and E. D. Goddard-Borger, *Nat. Chem. Biol.*, 2016, **12**, 215–217.
- 44 J. Li, R. Epa, N. E. Scott, D. Skoneczny, M. Sharma, A. J. D. Snow, J. P. Lingford, E. D. Goddard-Borger, G. J. Davies, M. J. McConville and S. J. Williams, *Appl. Environ. Microbiol.*, 2020, **86**, e00750.
- 45 S. R. Eddy, *Bioinformatics*, 1998, **14**, 755–763.
- 46 P. Rahfeld, J. F. Wardman, K. Mehr, D. Huff, C. Morgan-Lang, H. M. Chen, S. J. Hallam and S. G. Withers, *J. Biol. Chem.*, 2019, **294**, 16400–16415.
- 47 B. T. Hanson, K. Dimitri Kits, J. Löffler, A. G. Burrichter, A. Fiedler, K. Denger, B. Frommeyer, C. W. Herbold,

- T. Rattei, N. Karcher, N. Segata, D. Schleheck and A. Loy, *ISME J.*, 2021, **15**, 2779–2791.
- 48 P. Abayakoon, Y. Jin, J. P. Lingford, M. Petricevic, A. John, E. Ryan, J. Wai-Ying Mui, D. E. V. Pires, D. B. Ascher, G. J. Davies, E. D. Goddard-Borger and S. J. Williams, *ACS Cent. Sci.*, 2018, **4**, 1266–1273.
- 49 Y. Zhang, J. W. Mui, T. Arumaperuma, J. P. Lingford, E. D. Goddard-Borger, J. M. White and S. J. Williams, *Org. Biomol. Chem.*, 2020, **18**, 675–686.
- 50 P. Abayakoon, J. P. Lingford, Y. Jin, C. Bengt, G. J. Davies, S. Yao, E. D. Goddard-Borger and S. J. Williams, *Biochem. J.*, 2018, **475**, 1371–1383.
- 51 J. M. Bailey, P. H. Fishman and P. G. Pentchev, *J. Biol. Chem.*, 1967, **242**, 4263–4269.
- 52 J. M. Bailey, P. G. Pentchev and P. H. Fishman, *Fed. Proc.*, 1967, **26**, 854.
- 53 J. B. Thoden, J. Kim, F. M. Raushel and H. M. Holden, *J. Biol. Chem.*, 2002, **277**, 45458–45465.
- 54 J. A. Beebe and P. A. Frey, *Biochemistry*, 1998, **37**, 14989–14997.
- 55 D. L. Gibson, A. P. White, S. D. Snyder, S. Martin, C. Heiss, P. Azadi, M. Surette and W. W. Kay, *J. Bacteriol.*, 2006, **188**, 7722–7730.
- 56 A. Burrichter, K. Denger, P. Franchini, T. Huhn, N. Müller, D. Spiteller and D. Schleheck, *Front. Microbiol.*, 2018, **9**.
- 57 T. Conway, *FEMS Microbiol. Rev.*, 1992, **9**, 1–27.
- 58 M. Sharma, P. Abayakoon, R. Epa, Y. Jin, J. P. Lingford, T. Shimada, M. Nakano, J. W. Y. Mui, A. Ishihama, E. D. Goddard-Borger, G. J. Davies and S. J. Williams, *ACS Cent. Sci.*, 2021, **7**, 476–487.
- 59 M. Sharma, P. Abayakoon, J. P. Lingford, R. Epa, A. John, Y. Jin, E. D. Goddard-Borger, G. J. Davies and S. J. Williams, *ACS Catal.*, 2020, **10**, 2826–2836.
- 60 T. Itoh, B. Mikami, W. Hashimoto and K. Murata, *J. Mol. Biol.*, 2008, **377**, 1443–1459.
- 61 P. Abayakoon, R. Epa, M. Petricevic, C. Bengt, J. W. Y. Mui, P. L. van der Peet, Y. Zhang, J. P. Lingford, J. M. White, E. D. Goddard-Borger and S. J. Williams, *J. Org. Chem.*, 2019, **84**, 2901–2910.
- 62 V. Guixé and J. Babul, *J. Biol. Chem.*, 1985, **260**, 11001–11005.
- 63 R. L. Zheng and R. G. Kemp, *J. Biol. Chem.*, 1992, **267**, 23640–23645.
- 64 J. Park and R. S. Gupta, *Cell. Mol. Life Sci.*, 2008, **65**, 2875–2896.
- 65 T. Gefflaut, C. Blonski, J. Perie and M. Willson, *Prog. Biophys. Mol. Biol.*, 1995, **63**, 301–340.
- 66 N. Saito, M. Robert, H. Kochi, G. Matsuo, Y. Kakazu, T. Soga and M. Tomita, *J. Biol. Chem.*, 2009, **284**, 16442–16451.
- 67 R. K. Njau, C. A. Herndon and J. W. Hawes, *Chem.-Biol. Interact.*, 2001, **130-132**, 785–791.
- 68 J. Osipiuk, M. Zhou, S. Moy, F. Collart and A. Joachimiak, *J. Struct. Funct. Genomics*, 2009, **10**, 249–253.
- 69 T. Shimada, K. Yamamoto, M. Nakano, H. Watanabe, D. Schleheck and A. Ishihama, *Microbiology*, 2019, **165**, 78–89.
- 70 A.-K. Felux, P. Franchini and D. Schleheck, *Stand. Genomic Sci.*, 2015, **10**, 42.
- 71 J. Liu, Y. Wei, L. Lin, L. Teng, J. Yin, Q. Lu, J. Chen, Y. Zheng, Y. Li, R. Xu, W. Zhai, Y. Liu, Y. Liu, P. Cao, E. L. Ang, H. Zhao, Z. Yuchi and Y. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 15599–15608.
- 72 A. L. Davidson, E. Dassa, C. Orelle and J. Chen, *Microbiol. Mol. Biol. Rev.*, 2008, **72**, 317–364.
- 73 E. Eichhorn, J. R. van der Ploeg and T. Leisinger, *J. Biol. Chem.*, 1999, **274**, 26639–26646.
- 74 H. R. Ellis, *Bioorg. Chem.*, 2011, **39**, 178–184.
- 75 C. M. Driggers, P. V. Dayal, H. R. Ellis and P. A. Karplus, *Biochemistry*, 2014, **53**, 3509–3519.
- 76 K. Abdurachim and H. R. Ellis, *J. Bacteriol.*, 2006, **188**, 8153–8159.
- 77 L. Li, X. Liu, W. Yang, F. Xu, W. Wang, L. Feng, M. Bartlam, L. Wang and Z. Rao, *J. Mol. Biol.*, 2008, **376**, 453–465.
- 78 E. Eichhorn, C. A. Davey, D. F. Sargent, T. Leisinger and T. J. Richmond, *J. Mol. Biol.*, 2002, **324**, 457–468.
- 79 E. Romero, J. R. Gómez Castellanos, G. Gadda, M. W. Fraaije and A. Mattevi, *Chem. Rev.*, 2018, **118**, 1742–1769.
- 80 M. Toplak, A. Matthews and R. Teufel, *Arch. Biochem. Biophys.*, 2021, **698**, 108732.
- 81 A. Thakur, S. Somai, K. Yue, N. Ippolito, D. Pagan, J. Xiong, H. R. Ellis and O. Acevedo, *Biochemistry*, 2020, **59**, 3582–3593.
- 82 J. Xiong and H. R. Ellis, *Biochim. Biophys. Acta*, 2012, **1824**, 898–906.
- 83 J. M. Jez, M. J. Bennett, B. P. Schlegel, M. Lewis and T. M. Penning, *Biochem. J.*, 1997, **326**, 625–636.
- 84 J. K. Hiltunen and Y.-M. Qin, *Biochim. Biophys. Acta*, 2000, **1484**, 117–128.
- 85 R. Dippel and W. Boos, *J. Bacteriol.*, 2005, **187**, 8322.
- 86 J. R. van Der Ploeg, R. Iwanicka-Nowicka, T. Bykowski, M. M. Hryniewicz and T. Leisinger, *J. Biol. Chem.*, 1999, **274**, 29358–29365.
- 87 M. A. Moran and B. P. Durham, *Nat. Rev. Microbiol.*, 2019, **17**, 665–678.



Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria

Mahima Sharma^a, James P. Lingford^{b,c}, Marija Petricevic^{d,e}, Alexander J.D. Snow^a, Yunyang Zhang^{d,e}, Michael A. Järvå^{b,c}, Janice W.-Y. Mui^{d,e}, Nichollas E. Scott^f, Eleanor C. Saunders^g, Runyu Mao^{b,c}, Ruwan Epa^{d,e}, Bruna M. da Silva^{g,h}, Douglas E.V. Pires^{g,h}, David B. Ascher^{e,g}, Malcolm J. McConville^g, Gideon J. Davies^{a,1}, Spencer J. Williams^{d,e,1}, and Ethan D. Goddard-Borger^{b,c,1}

^aYork Structural Biology Laboratory, Department of Chemistry, University of York, Heslington YO10 5DD, United Kingdom; ^bThe Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia; ^cDepartment of Medical Biology, University of Melbourne, Parkville, VIC 3010, Australia; ^dSchool of Chemistry, University of Melbourne, Parkville, VIC 3010, Australia; ^eBio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia; ^fDepartment of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Parkville, VIC 3010, Australia; ^gDepartment of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, VIC 3010, Australia; and ^hSchool of Computing and Information Systems, University of Melbourne, Melbourne, VIC 3010, Australia

Edited by Carolyn Bertozzi, Department of Chemistry, ChEM-H, Stanford University, Stanford, CA; received September 2, 2021; accepted December 15, 2021

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 pathway for SQ degradation that involves oxidative desulfurization to release sulfite and enable utilization of the entire carbon skeleton of the sugar to support the growth of the plant pathogen *Agrobacterium tumefaciens*. SQ or its glycoside sulfoquinovosyl glycerol are imported into the cell by an ATP-binding cassette transporter system with an associated SQ binding protein. A sulfoquinovosidase hydrolyzes 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 NAD(P)H-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 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 because it enables the complete utilization of all carbons within SQ by a single organism with concomitant production of inorganic sulfite.

carbohydrate metabolism | sulfur cycle | oxidative desulfurization

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 tons produced each year in nature, and so its cycling is a significant component of the biogeochemical sulfur cycle (2). SQ is primarily found as the headgroup of the plant sulfolipid sulfoquinovosyl diacylglycerol, and its close association with photosynthetic membranes and proteins supports roles in their structure and function (3). Microbial communities play a dominant role in SQ cycling, and usually more than one tier of organism is required to completely assimilate this source of carbon and sulfur. Organisms with a tier 1 sulfoglycolysis pathway perform scission of the C3–C4 bond of SQ to give two three-carbon fragments; carbons 1 to 3 enter central metabolism, while carbons 4 to 6 bearing the sulfonate are excreted as dihydroxypropanesulfonate (DHPS) or sulfolactate (SL). Syntrophic organisms with a tier 2 pathway process DHPS and SL to access the remaining three-carbon fragment and release inorganic sulfur. To date, three tier 1 sulfoglycolysis pathways have been described: the sulfoglycolytic Embden-Meyerhof-Parnas (4), sulfoglycolytic Entner-Doudoroff (sulfo-ED) (5, 6), and sulfoglycolytic sulfofructose transaldolase pathways (7, 8). Tier 2

metabolism has been described for various specialized bacteria that "biomineralize" SL or DHPS to release inorganic sulfite, which under aerobic conditions is readily oxidized to sulfate (1). While many of the steps in the three 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 sulfoquinovosyl glycerol (SQGro), to release SQ (9, 10).

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 et al. 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 (11). We previously reported that *Agrobacterium tumefaciens* C58 contains a functional SQase, with the ability to hydrolyze SQGro (9). However, analysis of its genome did not reveal any genes homologous to those expected for known sulfoglycolysis pathways.

Significance

Sulfoquinovose, a sulfosugar derivative of glucose, is produced by most photosynthetic organisms and contains up to half of all sulfur in the biosphere. Several pathways for its breakdown are known, though they provide access to only half of the carbon in sulfoquinovose and none of its sulfur. Here, we describe a fundamentally different pathway within the plant pathogen *Agrobacterium tumefaciens* that features oxidative desulfurization of sulfoquinovose to access all carbon and sulfur within the molecule. Biochemical and structural analyses of the pathway's key proteins provided insights how the sulfosugar is recognized and degraded. Genes encoding this sulfoquinovose monooxygenase pathway are present in many plant pathogens and symbionts, alluding to a possible role for sulfoquinovose in plant host–bacteria interactions.

Author contributions: M.J.M., G.J.D., S.J.W., and E.D.G.-B. designed research; M.S., J.P.L., M.P., A.J.D.S., Y.Z., M.A.J., N.E.S., E.C.S., and R.M. performed research; Y.Z., R.E., B.M.d.S., D.E.V.P., and D.B.A. contributed new reagents/analytic tools; M.S., A.J.D.S., J.W.-Y.M., M.J.M., G.J.D., S.J.W., and E.D.G.-B. analyzed data; and M.S., G.J.D., S.J.W., and E.D.G.-B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: gideon.davies@york.ac.uk or sjwill@unimelb.edu.au or goddard-borger.e@wehi.edu.au.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116022119/-DCSupplemental>.

Published January 24, 2022.

Here, we investigate the “biomineralization” of SQ by *A. 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 an 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 (6-OG); and an NADPH-dependent oxidoreductase to reduce 6-OG 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 show through bioinformatics analyses that this pathway, which enables 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 SQ. 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 and *SI Appendix, Fig. S1*), as was previously reported by Roy et al. for *Agrobacterium* sp. strain ABR2 (11). Noting that sulfite is generally released from organosulfonate degradation pathways (1, 12), 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 $^{13}\text{C}_6$ -SQ (13) and analyzed the culture supernatant using ^{13}C nuclear magnetic resonance spectroscopy (*SI Appendix, Fig. S2*). The only significant ^{13}C -labeled product we could detect was ^{13}C -bicarbonate, which formed transiently during exponential phase growth, and the ^{13}C -labeled 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, 2-(*N*-morpholino)ethanesulfonic acid (MES), 3-(*N*-morpholino)propanesulfonic acid (MOPS), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES), cysteic acid, or methanesulfonic acid (*SI Appendix, Fig. S3*). Collectively, these data demonstrate 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 midlog 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 (14). Thus, the gene cluster *Atu3277-Atu3285*, which was subsequently renamed *smoA-smoF*, appeared to be important for growth on SQ (Fig. 1C). While the protein encoded by *Atu3285* was previously identified as an SQase (9), the proteins encoded by other genes in the cluster were not annotated with functions that were consistent with any sulfoglycolysis 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, short-chain dehydrogenase/reductase (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 (15). Accordingly, we produced recombinant SmoF (*SI Appendix, Fig. S4*) and demonstrated that it binds SQGro with $K_d = 0.29 \pm 0.17 \mu\text{M}$ ($\Delta H = -11 \pm 0.4 \text{ kcal mol}^{-1}$, $\Delta S = -7 \pm 2 \text{ cal mol}^{-1} \text{ deg}^{-1}$) (Fig. 2A and *SI Appendix, Fig. S5* and *Table S3*). 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 three-dimensional (3D) structure of SmoF in its ligand-free apo state and in complex with SQGro (Fig. 2B and *SI Appendix, Table S4*). 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 interdomain 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 (*SI Appendix, Fig. S6*). SQGro is buried deep within the interdomain cleft, and residues from both domains accommodate this ligand through a network of hydrogen-bonding interactions (Fig. 2C and D). The sulfonate of SQGro, which is the defining feature of this sulfosugar, is accommodated by hydrogen bonds to the sidechain 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 and 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 (*SI Appendix, Fig. S7*).

The Structural Basis of SQGro Recognition by the SQase Atu3285 (*smoI*). We previously reported that *Atu3285* (*smoI*) encodes an SQase that preferentially hydrolyzes 2'R-SQGro, the natural stereoisomer of this glycoside (9). 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 and 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 the Arg283/Trp286/Tyr491 triad⁸. 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 Reductase. SmoA, annotated as a flavin reductase, was recombinantly expressed in *Escherichia coli* and maintained a yellow color throughout

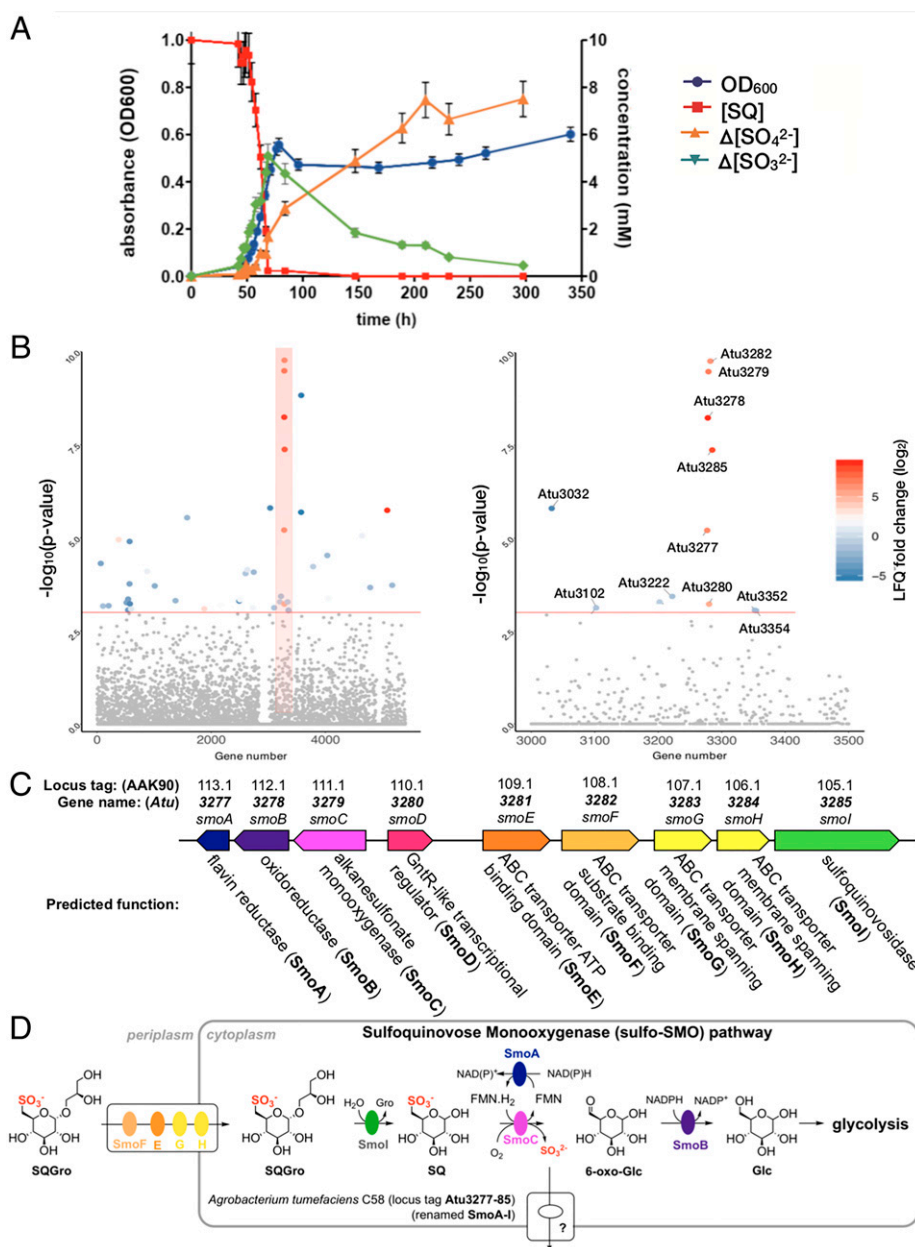


Fig. 1. *A. tumefaciens* utilizes SQ and its glycosides as a carbon source. (A) Optical density of *A. tumefaciens* C58 culture (blue) and concentration of SQ ([SQ]) (red), change in concentration of sulfite (Δ [sulfite]) (green) and change in concentration of sulfate (Δ [sulfate]) (yellow), with respect to time. This data are representative of two independent experiments (SI Appendix, Fig. S1); error bars denote observational error (derived by propagation of estimated random errors). (B) Manhattan plot of comparative proteomics data for *A. tumefaciens* C58 grown on SQ versus glucose, demonstrating that the most heavily up-regulated proteins belong to a single gene cluster. (C) An illustration of the up-regulated 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 biochemical pathway. (D) An illustration of the hypothetical roles played by the gene products of this pathway to complete the catabolism of SQGro.

purification, suggesting that it had copurified with a flavin cofactor. A sample of this protein was heat-denatured to release the cofactor and the supernatant analyzed by liquid chromatography–mass spectrometry (LC-MS) to reveal that flavin mononucleotide (FMN) was the sole detectable flavin (SI Appendix, Fig. S8). Michaelis-Menten kinetics were conducted for SmoA with saturating FMN and nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) to determine which of these reductants was preferred by the enzyme. With NADH, the kinetic parameters were $K_M = 35 \pm 5 \mu\text{M}$, $k_{\text{cat}} = 14.5 \pm 0.5 \text{ s}^{-1}$, and $k_{\text{cat}}/K_M = 4.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, while for NADPH, saturation was not observed and $k_{\text{cat}}/K_M = 6.8 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$, indicating that NADH is the

preferred cofactor for SmoA (Fig. 3A and SI Appendix, Figs. S9 and S10). Owing to difficulties in obtaining structural data for this enzyme, we also studied a close homolog from *Rhizobium oryzae* (*RoSmoA*, UniProt accession number: A0A1X7D6Q3), which possesses a syntenic gene cluster to *Atu3277-Atu3285*. Recombinant *RoSmoA* also copurified with FMN (SI Appendix, Fig. S8) and utilized the NADH cofactor with $K_M = 16 \pm 5 \mu\text{M}$, $k_{\text{cat}} = 33 \pm 2 \text{ s}^{-1}$, and $k_{\text{cat}}/K_M = 2.1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (SI Appendix, Fig. S9).

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

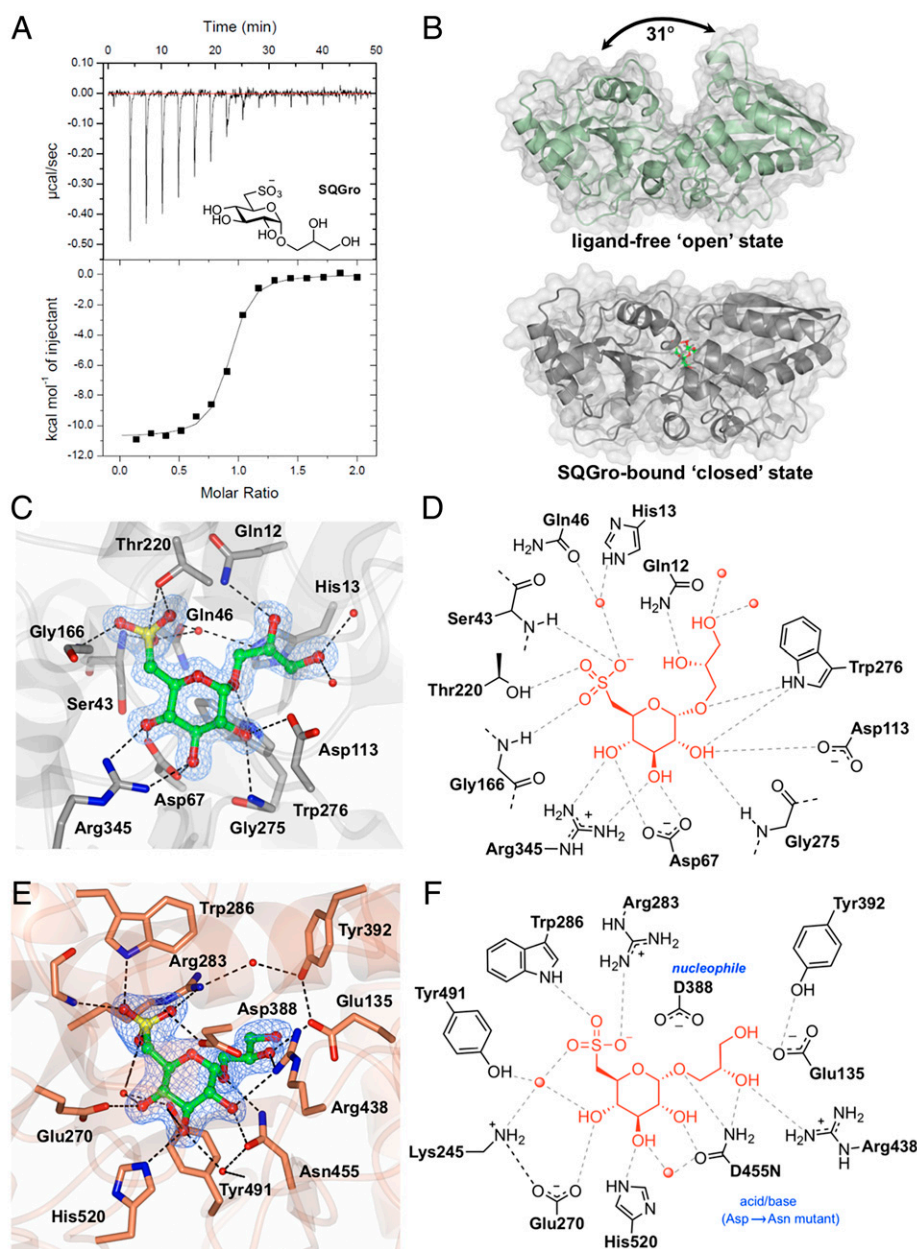


Fig. 2. Biochemical and structural analyses of the SQGro-binding protein SmoF (Atu3282) and SQase Smol (Atu3285). (A) Isothermal titration calorimogram for SmoF titrated against its cognate ligand 2'R-SQGro. The data are representative of two independent experiments (SI Appendix, Fig. S5). (B) Ribbon diagrams (with transparent surface) for the open and closed (liganded) conformations of SmoF. The 2'R-SQGro is bound tightly in the interdomain 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 gray, 2'R-SQGro is in green, and the 2Fo – Fc map at 1.5 σ is in blue. (D) An illustration highlighting key interactions from C. (E) Interactions between protein and ligand within the complex of Smol-D455N SQase and 2'R-SQGro: Smol is in gold, 2'R-SQGro is in green, and the 2Fo – Fc map at 1.5 σ is in blue. (F) An illustration highlighting key interactions from E: red spheres represent ordered water molecules; dotted lines represent proposed hydrogen bonds.

E. coli. SsuD catalyzes the reduced flavin mononucleotide (FMNH₂)- and O₂-dependent oxidation of alkanesulfonates to produce the corresponding aldehyde and sulfite, with a preference for pentanesulfonate (16). The mechanism of this and related enzymes has been intensively studied yet remains 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 fission and release of the aldehyde and sulfite products (SI Appendix, Fig. S124) (17). 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 (SI Appendix, Fig. S12B) (18). To demonstrate activity for recombinant SmoC (SI Appendix, Fig. S4), we adapted assays developed for SsuD that use Ellman's reagent to detect sulfite released by the enzyme (19). Direct detection of the putative sugar product, 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 FMNH₂ in situ, and the concentration of sulfite determined periodically using Ellman's reagent (Fig. 3B). Maximal substrate conversion was ~200 μ M (SI Appendix,

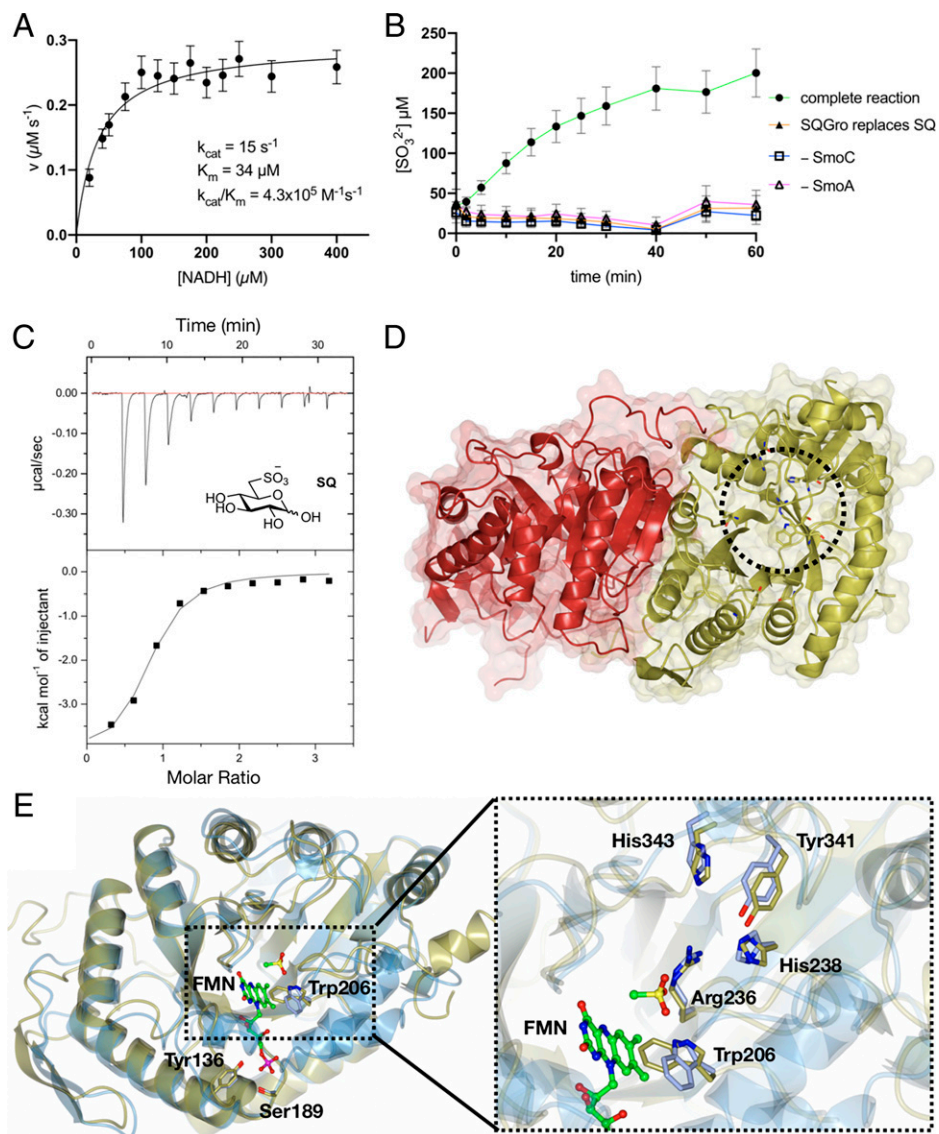


Fig. 3. Biochemical and structural analyses of the flavin reductase SmoA and SQ monoxygenase SmoC. (A) Michaelis-Menten kinetics for SmoA-catalyzed reduction of FMN by NADH. The data are representative of two independent replicates (SI Appendix, Fig. S10); error bars denote observational errors (derived by propagation of estimated random error). (B) SmoC activity assessed using sulfite release assay with Ellman's reagent in the presence of FMN, flavin reductase, NADH, and SQ. The data are representative of two independent experiments (SI Appendix, Fig. S11); error bars denote observational error (derived by propagation of estimated random error). (C) Isothermal titration calorimetry (ITC) titration of SmoC with SQ as determined by ITC. The data are representative of two independent experiments (SI Appendix, Fig. S13). (D) Transparent molecular surface and ribbon diagram of RoSmoC homodimer showing cofactor binding pocket and active site (dotted circle). (E) Alternative orientation of RoSmoC monomer (in gold) overlaid with the MsuD-FMN-CH₃SO₃⁻ complex (7K14.pdb in ice blue) showing FMN from the latter. Expansion shows view of proposed substrate-binding pocket and conserved residues lining the active site of RoSmoC.

Fig. S12C), which is commensurate with the solubility of molecular oxygen in water under standard conditions, with peak activity observed at pH 8.5 (SI Appendix, Fig. S12D). 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 (SI Appendix, Fig. S12C). As such, the hydrolysis of SQGro by SmoI necessarily precedes oxidative desulfurization by SmoC. This observation is further supported by isothermal titration calorimetry (ITC), where SQ was found to bind SmoC with $K_d = 3 \mu\text{M}$ in the absence of any flavin-based cofactors, whereas no binding was detected for SQGro (Fig. 3C and SI Appendix, Fig. S13 and Table S3). The unique SQ monoxygenase activity of SmoC defines this pathway: it is the

enzyme that effects fission of the C-S bond in SQ, and so it was chosen as the namesake for this gene cluster, and *Atu3277-Atu3285* was renamed the **SQ MonoOxygenase** 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 size exclusion chromatography-multiple angle light scattering (SEC-MALS) (SI Appendix, Fig. S14). To obtain structural information for an SQ monoxygenase, we turned to the homolog from *R. oryzae* (RoSmoC). Recombinant RoSmoC exhibited similar activity and substrate selectivity for SQ to SmoC (SI Appendix, Fig. S12E) and provided crystals that diffracted to 1.9 Å. Importantly, the low-resolution structure of *A. tumefaciens* SmoC superimposed with

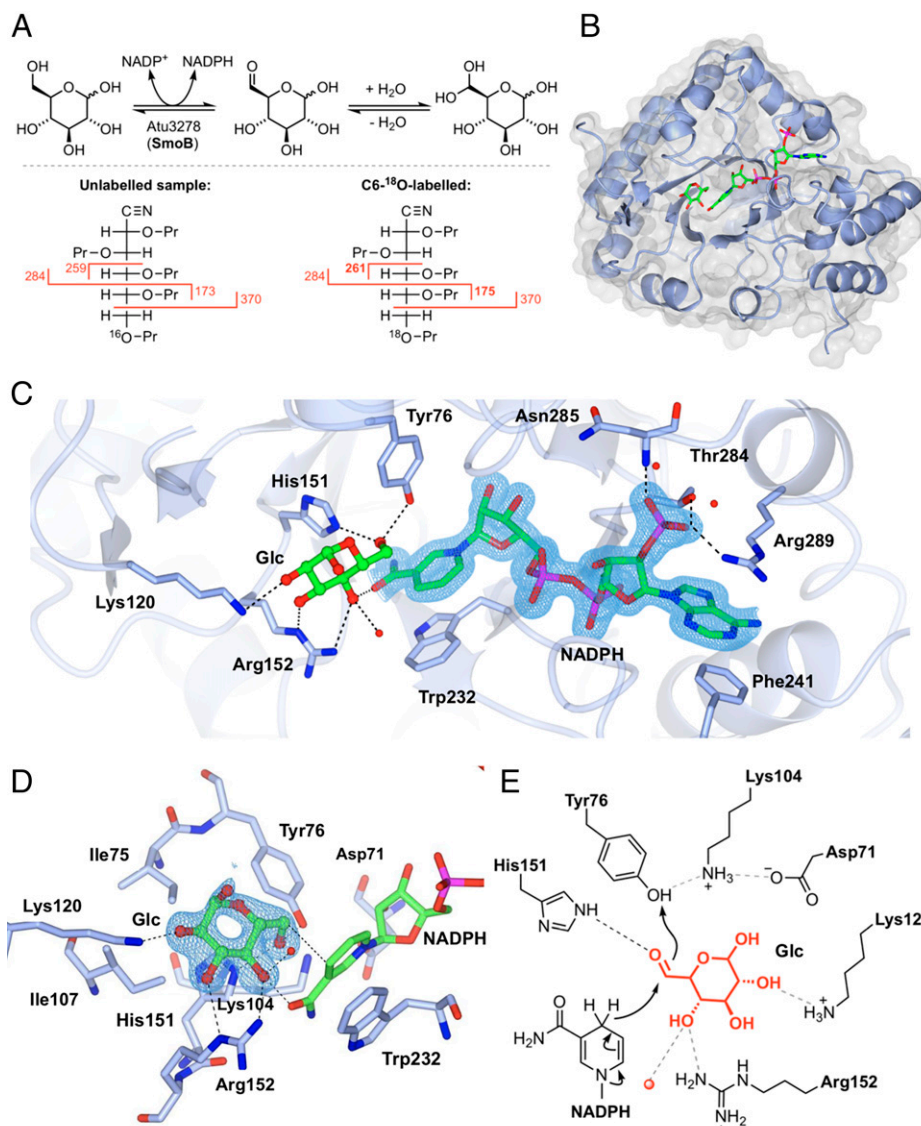


Fig. 4. Biochemical and structural analyses of 6-OG reductase SmoB. (A) Top: Equilibrium oxygen exchange at C-6 of Glc via 6-OG facilitated by SmoB when incubated with NADP⁺ in H₂¹⁸O. Bottom: Derivatization and MS fragmentation allows localization of ¹⁸O to C6 of Glc. (B) Transparent molecular surface 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.

the high-resolution *RoSmoC* 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 (*SI Appendix*, Fig. S15). Both SQ monooxygenases consist of a core (α/β)₈ triose-phosphate isomerase (TIM) barrel with three additional insertion regions, analogous to monooxygenases from the bacterial luciferase family. The protomers exist as a homodimer that buries 4,697 Å² of surface area, amounting to 18% of total accessible surface area for each protomer (Fig. 3D). Pairwise structural analysis using the distance matrix alignment (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, the FMN₂-dependent methane-sulfonate monooxygenase *MsuD* (7K14.pdb, rmsd 2.0/322

residues, Z score of 41.0), and the FMN₂-dependent alkane-sulfonate monooxygenase *SsuD* (1M41.pdb, rmsd 1.8/317 residues, Z score of 41.2).

Comparisons of the *RoSmoC* structure with *MsuD* (7K14.pdb) in complex with FMN enabled identification of the FMN binding site for *RoSmoC*: 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) (18). The close structural and functional relationship of *RoSmoC* to *MsuD* is evident from the conservation of a putative sulfonate binding site comprised of the side-chains Trp206, Arg236, His238, Tyr341, and His343 (18). 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 MsuD and SsuD (18, 19). Efforts to obtain crystals of a RoSmoC-SQ complex were unsuccessful, limiting further insights into the origin of enzyme specificity toward SQ over other sulfonates.

Atu3278 (smoB) Encodes an Oxidoreductase that Converts 6-OG to Glucose. SmoB is annotated as an 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 correct, and as a consequence of microscopic reversibility, incubation of SmoB with a nicotinamide cofactor and glucose in H₂¹⁸O should result in transient formation of 6-OG, rapid and reversible hydration/dehydration with H₂¹⁸O to compete-out ¹⁶O at C6 for ¹⁸O, and reduction to give 6-¹⁸O-glucose. In parallel to this process, ¹⁸O 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 nicotinamide cofactor was suitable for SmoB: NADPH bound to SmoB with *K*_d ~2 μM, while no binding was observed for NADH (SI Appendix, Fig. S16 and Table S3). Thus, glucose pre-equilibrated in H₂¹⁸O was incubated with SmoB and NADP⁺ and 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 ¹⁸O atoms into glucose. The crude reaction mixture was subjected to peracetylation (Ac₂O/pyridine) and then LC-MS analysis to confirm that the +4 Da product coeluted with authentic D-glucose-pentaacetate (SI Appendix, Fig. S17). To determine that the ¹⁸O label was incorporated at C6 of glucose, we used electron-impact gas chromatography-mass spectrometry (GC-MS), which required conversion of the reaction product to the acyclic pentapropionate aldonitrile (SI Appendix, Fig. S18) (20). This approach provided diagnostic C1-C5 and C5-C6 fragment ions. The ¹⁸O-labeled 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 unlabeled glucose reference (*m/z* 370), demonstrating that the ¹⁸O is incorporated at C6. Only enzymatic reactions conducted in the presence of NADP⁺ produced product labeled with ¹⁸O 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 His₆-tag in this construct occupied the putative active site of adjoining subunits, making cocrystallization with cofactors difficult (SI Appendix, Fig. S19). 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 (SI Appendix, Fig. S20). This SmoB construct was cocrystallized 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 (α/β)₈ TIM barrel fold with a C-terminal cofactor binding site. The overall fold has high structural conservation with members of the aldoketo 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 nonredundant protein set of the National Center for Biotechnology Information (NCBI) for gene clusters that contain homologous SQases and SQ monooxygenases. 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). Among these putative *smo* gene clusters, some were syntenic while others were substantially rearranged (nonsyntenic) 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 sulfo-ED pathway (5, 6). Indeed, sulfo-ED gene clusters have been identified in several Rhizobiales (5, 6), suggesting that there has been ample opportunity for genetic exchanges between these pathways during their evolution.

Discussion

While existing pathways for the breakdown of SQ require two different organisms and involve scission of the carbon chain into two three-carbon fragments, we describe here a fundamentally different approach that features complete utilization of the SQ carbon skeleton. The SMO pathway includes several proteins with hitherto undescribed activities, including an SQGro-binding protein; an FMNH₂- and O₂-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 (9, 10).

The SMO pathway is reminiscent of other sugar-metabolizing pathways in bacteria. For example, the SmoI (SQase), SmoF (SQGro-binding protein), and SmoE/G/H (ABC transporter) proteins encoded by the *smo* cluster are 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). Additionally, the SmoC (SQ monooxygenase) and SmoA (flavin reductase) proteins of the SMO pathway are reminiscent of the SsuD (FMNH₂-dependent alkylsulfonate monooxygenase) and SsuE (NADPH-dependent FMN reductase) pair encoded by the *ssu* operon of *E. coli* that degrades alkanesulfonates (16). Indeed, it is likely that the SMO pathway arose through the recombination and neofunctionalization of analogous sugar- and sulfonate-metabolizing pathways.

Through structural analysis, we 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 (9, 23, 24). This includes the Thr220-Gly166-Ser43-H₂O (His13-Gln46) cluster of SmoF for the recognition of SQGro; the Arg283-Trp286-H₂O (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 monooxygenases like SsuD.

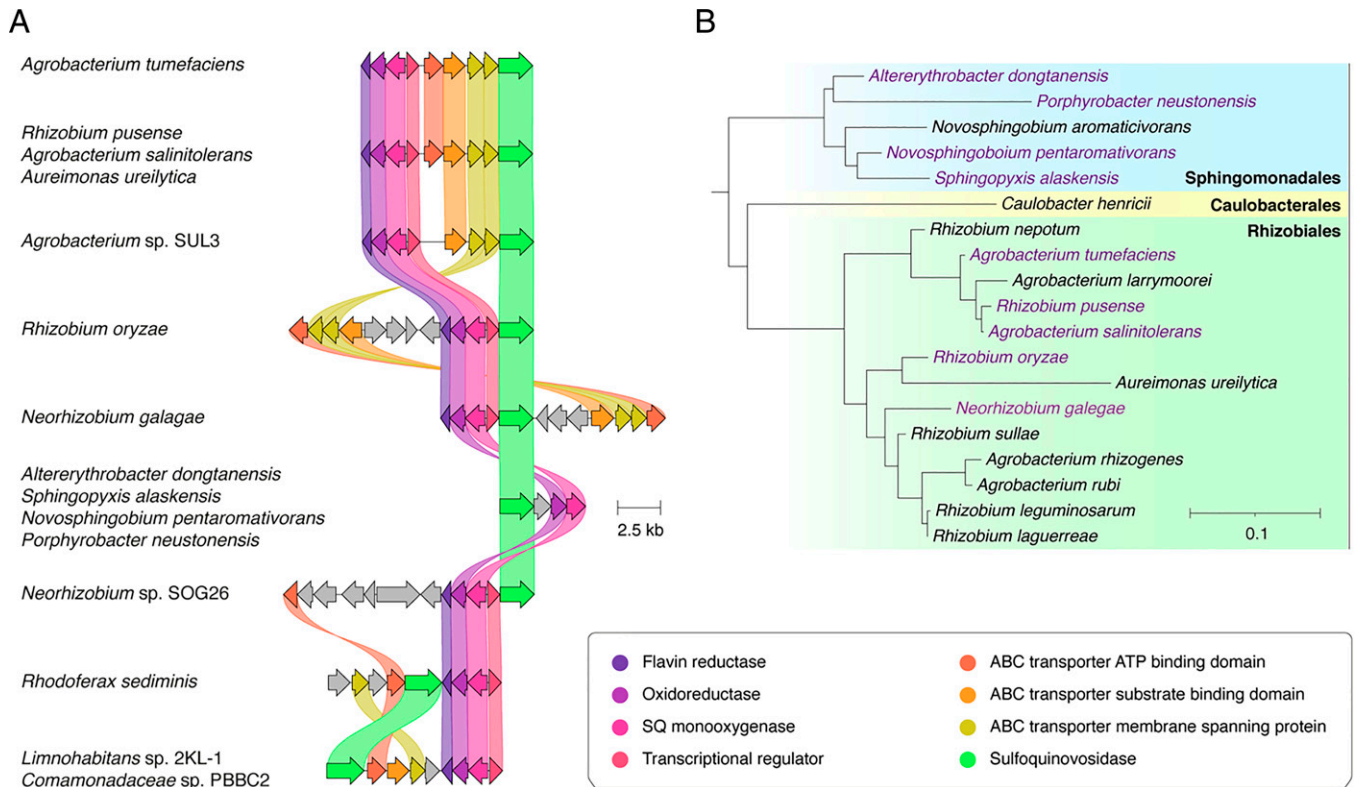


Fig. 5. Prevalence of the SMO pathway. (A) Architecture of the SMO gene cluster in *A. tumefaciens* and homologous gene clusters in other organisms. Colored links indicate $\geq 30\%$ protein sequence similarity. Only those clusters encoding putative SQ monooxygenases and SQases were annotated as putative 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/>).

The prevalence of the SMO pathway in Alphaproteobacteria of the Rhizobiales order is intriguing, since many bacteria of this order are plant symbionts or pathogens. Indeed, those bacteria that do not possess an SMO pathway often possess a complementary tier 1 sulfo-ED pathway (5). Accordingly, it appears that plant sulfolipid catabolism is important for Rhizobiales, 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). Speck et al. 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 sulfolipid 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. Combined with the pathway's requirement for molecular oxygen to effect C-S bond fission, this may 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 different environmental niches remains an important question, with answers that have significant implications for understanding plant diseases and symbioses as well as soil chemistry.

Materials and Methods

Specialist Reagents. SQ and methyl α -sulfoquinovoside were purchased from MCAT GmbH, and ($^{13}\text{C}_6$)SQ, glycer-1-yl α -sulfoquinovoside (SQGro), and dicyclohexylammonium sulfolactate and cyclohexylammonium dihydroxypropylsulfonate were synthesized as described (13, 28). All other sulfonates were purchased from Sigma-Aldrich.

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 to 3 d. These were subcultured (1% inoculum) into the same media (10 mL) and grown at 30 °C (250 rpm). Bacterial growth was quantitated using a Varian Cary50 ultraviolet (UV)/visible spectrophotometer to measure OD₆₀₀. 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 (29). This assay uses preprepared alkaline diluent and 4-hydroxybenzoic acid hydrazide (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, and 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 (30). This assay uses a barium sulfate–polyethylene glycol (Ba-PEG) reagent, which contains polyethylene glycol (PEG) to stabilize BaSO₄ crystals and a small amount of preformed BaSO₄ 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 BaCl₂ (42 mg, 0.20 mmol) and polyethylene glycol 6000 (0.75 g) in deionized water (5.0 mL). A small amount of Na₂SO₄ (10 μL, 50 mM) was added to this solution, with efficient magnetic stirring to generate preformed BaSO₄ seed crystals. Individual sulfate assays were conducted as follows. An aliquot of culture supernatant obtained after pelleting of cells for 5 min at 5,000 g (typically 100 μL, containing a maximum of 2.5 μmol of Na₂SO₄) 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 Na₂SO₄. This curve was linear up to 2.5 μmol of Na₂SO₄.

Colorimetric Fuchsin Sulfite Assay for Culture Supernatant The fuchsin sulfite assay was performed according to the procedures of Brychkova et al. (31) and Kurmanbayeva et al. (32). This procedure requires three preprepared 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% H₂SO₄ (1.25 mL). Reagent B was prepared by diluting formaldehyde (36% in H₂O, 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 Na₂SO₃ (12 μL), with the latter providing a stable background signal for reference. The sample was incubated at 20 to 22 °C 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 Na₂SO₃.

NMR Analysis of Metabolites Produced from (¹³C₆)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 (¹³C₆)SQ and the culture incubated at 30 °C (250 rpm). At OD₆₀₀ 0.27 and OD₆₀₀ 0.49, 950 μL samples of culture supernatant were diluted with 100 μL of D₂O and ¹³C-NMR spectra acquired using a 400 MHz spectrophotometer (100 MHz for ¹³C).

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), MeSQ, glycer-1-yl α-sulfoquinovoside (SQGro), dicyclohexylammonium sulfolactate, cyclohexylammonium dihydroxypropanesulfonate, sulfoacetic acid, taurine, sodium pentanesulfonate, cysteine acid, MOPS, HEPES, PIPES, MES, and methanesulfonic acid. Cultures were incubated for 30 d 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 resuspended in 500 μL lysis buffer (4% sodium dodecyl sulfate (SDS), 50 mM Tris pH 8.5, 10 mM dithiothreitol (DTT)) and boiled at 95 °C for 10 min with shaking at 2,000 rpm to shear DNA and inactivate protease activity. Lysates were cooled to room temperature and protein concentration determined using a bichinchoninic acid (BCA) assay. Each sample (200 μg of protein) was acetone precipitated by mixing four volumes of ice-cold acetone with one volume of sample. Samples were precipitated overnight at –20 °C and then centrifuged at 4,000 × 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 17,000 × g for 10 min at 4 °C to collect precipitated protein, and 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, and 40 mM NH₄HCO₃ and reduced/alkylated prior to digestion with Lys-C (1/200 wt/wt) and then trypsin (1/50 wt/wt) overnight as previously described (33). Digested samples were acidified to a final concentration of 0.5% formic acid and desalted using C18 stage tips (34) before analysis by LC-MS.

Quantitative Proteomics Using Reversed Phase LC-MS. Purified peptides were resuspended in Buffer A* (2% acetonitrile [MeCN], 0.1% trifluoroacetic acid [TFA]) and separated using a Proflow-equipped Dionex Ultimate 3000 Ultra-Performance Liquid Chromatography system (Thermo Fisher Scientific) with a two-column chromatography setup 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% formic acid [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, and 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 tandem mass spectrometry scans (higher energy collisional dissociation normalized collision energy 28, maximum fill time 40 ms, automatic gain control 2 × 10⁵ with a resolution of 15,000).

Mass Spectrometry Data Analysis. Proteomics datasets were searched using MaxQuant (v1.5.3.3) (35) against the *A. tumefaciens* C58 proteome (Uniprot proteome id UP000000813, downloaded 27 January 2018, 5,344 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 two miscleavage events with a maximum false discovery rate 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 (LFQ), the MaxLFQ option within MaxQuant (36) was enabled in addition to the requantification module. The resulting protein group output was processed within the Perseus (v1.4.0.6) (37) analysis environment to remove reverse matches and common protein contaminants 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 (38) 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 *SI Appendix, Table S1* and *A. tumefaciens* C58 gDNA as template. Oligonucleotides encoding RoSmoA and RoSmoC were synthesized (IDT) to provide the sequences listed in *SI Appendix, Table S1*. 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, RoSmoA, and RoSmoC. Due to interference from the SmoB C-terminal His₆-tag during structural studies, the *smoB* (*Atu3278*) gene was subcloned into the pET-Y5BLC3C vector (39) by PCR amplification with the relevant primers in *SI Appendix, Table S1* and In-Fusion cloning (Clontech Laboratories, Inc.) into linearized Y5BLC3C 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 LysoGeny Broth (LB)-agar (50 μg mL⁻¹ 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⁻¹ kanamycin, and the cultures were incubated at 37 °C for 16 h. These starter cultures were used to inoculate 1,000 mL of S-broth (35 g tryptone, 20 g yeast extract, 5 g NaCl, pH 7.4) containing 50 μg mL⁻¹ kanamycin, which was incubated with shaking (250 rpm) at 37 °C until it reached an OD₆₀₀ of 0.8. Each culture was cooled to room temperature, isopropyl thiogalactoside 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 and then resuspended in 40 mL binding buffer

(50 mM NaPi, 300 mM NaCl, 5 mM imidazole, pH 7.5) containing protease inhibitor (Roche cOmplete EDTA-free protease inhibitor mixture) and lysozyme (0.1 mg mL⁻¹) by nutating at 4°C for 30 min. Benzonase (1 µL, 250 U) 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, and then the protein was eluted using elution buffer (50 mM NaPi, 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 NaPi, 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 (SI Appendix, Fig. S2). Smol (Atu3285 or Smol) was prepared as previously described (9).

SEC-MALS Analyses. Experiments were conducted on a system comprising a Wyatt HELEOS-II multiangle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu liquid chromatography (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 carryover 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).

ITC. ITC experiments were performed using a MicroCal iTC200 (GE Healthcare) at 25°C, with a 750 rpm stirring speed and a reference power of 10 µCal·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 SmoB). Protein concentration was determined by BCA assay (Thermo Fisher) before initiating 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 preinjection 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 preinjection 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 preinjection 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).

Nanodifferential 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 Cofactor that Copurified with SmoA. An amount of 100 µL of recombinant flavin reductase (SmoA or RoSmoA) 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 LC-MS on an Agilent 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 C₁₈ 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⁻¹; DAD: 254 and 214 nm.

Michaelis-Menten Kinetic Analyses of SmoA and RoSmoA. Reactions were conducted at 25°C in 96-well plate format and involved the addition of SmoA or RoSmoA (final concentration of 20 nM for NADH and 500 nM for NADPH) to 20 to 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 (19) 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% (wt/vol) BSA, 100 nM SmoA or RoSmoA, and 300 nM SQ monooxygenase (SmoC or RoSmoC) 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 RoSmoA 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 to 1,000 µM Na₂SO₃. The activity of SQ monooxygenases at different pH was determined by modifying the buffer in the aforementioned 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 Labeling Using SmoB In order to prelabel the anomeric position, glucose was incubated in 98% H₂¹⁸O with heating at 80°C for 2 d, then evaporated to dryness to give C1-¹⁸O-labeled glucose. Labeling was determined to be 95% by mass spectrometry based on intensities of the M and M+2 peaks. Using H₂¹⁸O buffer (100 mM potassium phosphate, pH 7.0), NAD⁺ and NADP⁺ were each added at 0.05 molar equivalent to C1-¹⁸O-glucose and SmoB. Four control experiments were conducted: one without enzyme, one without NAD⁺ and NADP⁺, one in H₂¹⁶O, and one in H₂¹⁸O with unlabeled glucose. Reactions were monitored by mass spectrometry. Only in the experimental sample containing enzyme, H₂¹⁸O, 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 H₂¹⁸O, and only the reaction containing NADP⁺ generated the M+4 species. To confirm that the M+4 species was glucose with two ¹⁸O labels, we studied the product by HPLC. However, under aqueous HPLC conditions, the ¹⁸O-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 aforementioned reaction mixture 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 ethyl acetate and washed with saturated CuSO₄ to remove pyridine. The organic solution containing peracetylated glucose was analyzed by LC-MS on an Agilent 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 C₁₈ 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⁻¹. 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 Labeled Carbohydrates. A 0.1 µL aliquot of SmoB-glucose reaction mixture (containing ~2.5 nmol glucose) was transferred to a gas chromatography (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 20 to 22°C 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 ethyl acetate (40 µL). Control samples of U-¹²C-glucose, U-¹³C-glucose, 1,2-¹³C₂-glucose, and 6,6-²H₂-glucose were also prepared at a 2.5 nmol scale in the assay buffer mixture. Samples were blinded for analysis. The derivatized labeled glucose samples (SI Appendix, Fig. S13 and Table S7) 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). *SI Appendix, Fig. S12* shows the fraction of labeled fragments, corrected for isotope natural abundance by DEXSI analysis (40).

Protein Crystallization. Initial crystallization screening was performed using commercially available INDEX (Hampton Research), PACT premier, and CSS/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% wt/vol 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% [wt/vol] 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 optimization plate was set up with drops comprised of 0.1 µL of various mother liquors (28–36% [wt/vol] polyethylene glycol 4000, 0.2 M sodium acetate, 0.1 M Tris chloride, pH 7.1 to 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% (wt/vol) 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% (vol/vol) ethylene glycol as cryoprotectant.

Crystals of SmoL-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 wt/vol, 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 wt/vol, and 10 mM SQ-glucitol). Crystals of apo-RoSmoC 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 NaNO₃, 20% PEG 3350 wt/vol, 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% wt/vol 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% wt/vol 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% wt/vol 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 (41) and scaled using SCALA (42) included in the Xia2 processing system (43). Data reduction was performed with AIMLESS, and resolution was cut until CC1/2 = 0.5. The structure of the SmoL•SQGro complex was determined using molecular replacement using 5OHS (9) as the initial model. For SmoF, the structure was solved by molecular replacement using PHASER (44) with a search model created from Protein Data Bank (PDB) ID: 6DTQ (45). The structure of RoSmoC was solved by molecular replacement using the ensemble based on PDB ID: 1M41 (19) as an initial search model. The

structure of SmoB was determined using molecular replacement with the monomer of an AKR 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 (46) and REFMAC (47) or Phenix (48), the latter employing local noncrystallographic symmetry restraints. Following building and refinement of the protein and water molecules, clear residual density was observed in the omit maps for cocomplex structures, and respective ligands were modeled into these. The coordinate and refinement library files were prepared using ACEDRG (49). The final structures gave R_{cryst} and R_{free} values along with data and refinement statistics that are presented in *SI Appendix, Tables S4–S6*. Data were collected at Diamond Light Source, Didcot, Oxfordshire, UK, on beamlines I24 (SmoL-D455N•SQGro, to 2.15 Å; SmoF-apo, to 1.88 Å), I04 (RoSmoC 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 PDB with the coordinate accession numbers 7OFX (SmoL-D455N•SQGro), 7NBZ (SmoF-apo), 7OFY (SmoF•SQGro), 7OH2 (RoSmoC), 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 (50). Structural comparisons and structure-based sequence alignments were conducted using PDB25 search on DALI server against a representative subset of the PDB (51). All structure figures were generated using ccp4mg (52).

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 $\leq 1.19 \times 10^{-51}$. The corresponding nucleotide accession numbers for each protein from all nine searches were extracted and combined and duplicates removed to provide a list of candidate genome sequences. This was converted into a reference library for MultiGeneBLAST (53) and queried using the *A. tumefaciens* C58 SMO gene cluster. Clusters identified by this workflow with both an SQ monooxygenase and SQase homolog were regarded as putative SMO gene clusters. Clusters representative of the observed diversity were visualized using Clinker (54). 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 (55) using iTOL (56).

Data Availability. Structure coordinates have been deposited in the PDB (<https://www.rcsb.org/>) under accession codes 7OFX, 7OFY, 7NBZ, 7OH2, 7OLF, 7BBZ, 7BC0, 7BC1, and 7BBY. Proteomics data are available via ProteomeXchange (57, 58) (<http://www.proteomexchange.org/>) with the identifier PXD014115. Scripts used to screen for the related gene clusters listed in Fig. 5 are available on GitHub (<https://github.com/jmui-unimelb/Gene-Cluster-Search-Pipeline>) (59).

ACKNOWLEDGMENTS. Dr. Monica Doblin is thanked for the provision of *A. 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.), and GNT1139546 and GNT1139549 (E.D.G.-B.); the Leverhulme Trust grant RPG-2017-190 (G.J.D.); Australian Research Council grants 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 an Australian Research Council Future Fellowship (FT200100270), B.M.d.S. was supported by the Melbourne Research Scholarship, J.W.Y.M. was supported by a Sir John and Lady Higgins Scholarship, and M.P. was supported 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 Australia's Nuclear Science and Technology Organisation, for access to the MX-2 beamline, which made use of the Australian Cancer Research Foundation 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.

1. E. D. Goddard-Borger, S. J. Williams, Sulfoquinovose in the biosphere: Occurrence, metabolism and functions. *Biochem. J.* **474**, 827–849 (2017).
2. J. L. Harwood, R. G. Nicholls, The plant sulpholipid—A major component of the sulphur cycle. *Biochem. Soc. Trans.* **7**, 440–447 (1979).
3. N. Mizusawa, H. Wada, The role of lipids in photosystem II. *Biochim. Biophys. Acta* **1817**, 194–208 (2012).
4. K. Denger *et al.*, Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **507**, 114–117 (2014).
5. A. K. Felux, D. Spittler, J. Klebensberger, D. Schleheck, Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4298–E4305 (2015).
6. J. Li *et al.*, A sulfoglycolytic Entner-Doudoroff pathway in *Rhizobium leguminosarum* bv. *trifolii* SRDI565. *Appl. Environ. Microbiol.* **86**, e00750-20 (2020).
7. B. Frommeyer *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).
8. Y. Liu *et al.*, A transaldolase-dependent sulfoglycolysis pathway in *Bacillus megaterium* DSM 1804. *Biochem. Biophys. Res. Commun.* **533**, 1109–1114 (2020).
9. P. Abayakoon *et al.*, Structural and biochemical insights into the function and evolution of sulfoquinovosidases. *ACS Cent. Sci.* **4**, 1266–1273 (2018).
10. G. Speciale, Y. Jin, G. J. Davies, S. J. Williams, E. D. Goddard-Borger, YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **12**, 215–217 (2016).
11. A. B. Roy, M. J. Hewlins, A. J. Ellis, J. L. Harwood, G. F. White, Glycolytic breakdown of sulfoquinovose in bacteria: A missing link in the sulfur cycle. *Appl. Environ. Microbiol.* **69**, 6434–6441 (2003).
12. M. A. Kertesz, Riding the sulfur cycle—Metabolism of sulfonates and sulfate esters in gram-negative bacteria. *FEMS Microbiol. Rev.* **24**, 135–175 (2000).
13. P. Abayakoon *et al.*, Comprehensive synthesis of substrates, intermediates and products of the sulfoglycolytic Embden-Meyerhoff-Parnas pathway. *J. Org. Chem.* **84**, 2901–2910 (2019).
14. A. E. Speers, C. C. Wu, Proteomics of integral membrane proteins—Theory and application. *Chem. Rev.* **107**, 3687–3714 (2007).
15. A. L. Davidson, E. Dassa, C. Orelle, J. Chen, Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72**, 317–364 (2008).
16. J. R. van Der Ploeg, R. Iwanicka-Nowicka, T. Bykowski, M. M. Hryniewicz, T. Leisinger, 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).
17. A. Thakur *et al.*, Substrate-dependent mobile loop conformational changes in alkanesulfonate monooxygenase from accelerated molecular dynamics. *Biochemistry* **59**, 3582–3593 (2020).
18. J. J. M. Liew, I. M. El Saudi, S. V. Nguyen, D. K. Wicht, D. P. Dowling, 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**, 100823 (2021).
19. E. Eichhorn, C. A. Davey, D. F. Sargent, T. Leisinger, T. J. Richmond, Crystal structure of *Escherichia coli* alkanesulfonate monooxygenase SsuD. *J. Mol. Biol.* **324**, 457–468 (2002).
20. M. R. Antoniewicz, J. K. Kelleher, G. Stephanopoulos, Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal. Chem.* **83**, 3211–3216 (2011).
21. T. M. Penning, The aldo-keto reductases (AKRs): Overview. *Chem. Biol. Interact.* **234**, 236–246 (2015).
22. R. Dippel, W. Boos, The maltodextrin system of *Escherichia coli*: Metabolism and transport. *J. Bacteriol.* **187**, 8322–8331 (2005).
23. M. Sharma *et al.*, Dynamic structural changes accompany the production of dihydroxypropanesulfonate by sulfolactaldehyde reductase. *ACS Catal.* **10**, 2826–2836 (2020).
24. M. Sharma *et al.*, Molecular basis of sulfosugar selectivity in sulfoglycolysis. *ACS Cent. Sci.* **7**, 476–487 (2021).
25. M. Udvardi, P. S. Poole, Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).
26. J. J. Speck, E. K. James, M. Sugawara, M. J. Sadowsky, P. Gyaneshwar, An alkane sulfonate monooxygenase is required for symbiotic nitrogen fixation by *Bradyrhizobium diazoefficiens* (syn. *Bradyrhizobium japonicum*) USDA110^T. *Appl. Environ. Microbiol.* **85**, e01552-19 (2019).
27. K. Denger, T. Huhn, K. Hollemeyer, D. Schleheck, A. M. Cook, Sulfoquinovose degraded by pure cultures of bacteria with release of C₃-organosulfonates: Complete degradation in two-member communities. *FEMS Microbiol. Lett.* **328**, 39–45 (2012).
28. Y. Zhang *et al.*, Concise synthesis of sulfoquinovose and sulfoquinovosyl diacylglycerides, and development of a fluorogenic substrate for sulfoquinovosidases. *Org. Biomol. Chem.* **18**, 675–686 (2020).
29. A. B. Blakeney, L. L. Mutton, A simple colorimetric method for the determination of sugars in fruit and vegetables. *J. Sci. Food Agric.* **31**, 889–897 (1980).
30. B. Sörbo, "Sulfate: Turbidimetric and nephelometric methods" in *Methods in Enzymology* (Academic Press, 1987), Vol. **143**, pp. 3–6.
31. G. Brychkova, D. Yarmolinsky, R. Fluhr, M. Sagi, The determination of sulfite levels and its oxidation in plant leaves. *Plant Sci.* **190**, 123–130 (2012).
32. A. Kurmanbayeva *et al.*, "Determination of total sulfur, sulfate, sulfite, thiosulfate, and sulfolipids in plants" in *Plant Stress Tolerance: Methods and Protocols*, R. Sunkar Ed. (Springer New York, 2017), pp 253–271.
33. N. E. Scott *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).
34. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
35. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
36. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
37. S. Tyanova *et al.*, Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456 (2015).
38. Y. Perez-Riverol *et al.*, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47** (D1), D442–D450 (2019).
39. M. J. Fogg, A. J. Wilkinson, Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.* **36**, 771–775 (2008).
40. M. J. Dagley, M. J. McConville, DExSI: A new tool for the rapid quantitation of 13C-labelled metabolites detected by GC-MS. *Bioinformatics* **34**, 1957–1958 (2018).
41. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
42. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
43. G. Winter, xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2010).
44. L. C. Storoni, A. J. McCoy, R. J. Read, Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432–438 (2004).
45. S. Shukla *et al.*, Differential substrate recognition by maltose binding proteins influenced by structure and dynamics. *Biochemistry* **57**, 5864–5876 (2018).
46. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
48. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
49. F. Long *et al.*, AceDRG: A stereochemical description generator for ligands. *Acta Crystallogr. D Struct. Biol.* **73**, 112–122 (2017).
50. E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
51. L. Holm, P. Rosenström, Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–9 (2010).
52. S. McNicholas, E. Potterton, K. S. Wilson, M. E. Noble, Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 386–394 (2011).
53. M. H. Medema, E. Takano, R. Breitling, Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218–1223 (2013).
54. C. L. M. Gilchrist, Y. H. Chooi, Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics (Oxford, England)* **37**, 2473–2475 (2021).
55. P. Yilmaz *et al.*, The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).
56. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47** (W1), W256–W259 (2019).
57. E. W. Deutsch *et al.*, The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45** (D1), D1100–D1106 (2017).
58. E.D. Goddard-Borger, N. Scott, Agrobacterium sulfoquinovose metabolism. ProteomeXchange. <http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX0014115>. Deposited 4 June 2019.
59. J. W.-Y. Mui, Gene cluster search pipeline. GitHub. <https://github.com/jmui-unimelb/Gene-Cluster-Search-Pipeline>. Deposited 10 November 2021.



The sulfoquinovosyl glycerol binding protein SmoF binds and accommodates plant sulfolipids

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

^a York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5DD, UK

^b The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia

^c Department of Medical Biology, University of Melbourne, Parkville, Victoria, 3010, Australia

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

ARTICLE INFO

Handling editor: Natalie Strynadka

Keywords:

X-ray crystallography

Substrate-binding protein

Isothermal-titration calorimetry

Sulfoglycolysis

ABSTRACT

Sulfoquinovose (SQ) is the anionic headgroup of the ubiquitous plant sulfolipid, sulfoquinovosyl diacylglycerol (SQDG). SQDG can undergo delipidation to give sulfoquinovosyl glycerol (SQGro) and further glycoside cleavage to give SQ, which can be metabolized through microbial sulfoglycolytic pathways. Exogenous SQDG metabolites are imported into bacteria through membrane spanning transporter proteins. The recently discovered sulfoglycolytic sulfoquinovose monooxygenase (sulfo-SMO) pathway in *Agrobacterium tumefaciens* features a periplasmic sulfoquinovosyl glycerol binding protein, SmoF, and an ATP-binding cassette (ABC) transporter. Here, we use X-ray crystallography, differential scanning fluorimetry and isothermal titration calorimetry to study SQ glycoside recognition by SmoF. This work reveals that in addition to SQGro, SmoF can also bind SQ, a simple methyl glycoside and even a short-chain SQDG analogue. Molecular recognition of these substrates is achieved through conserved interactions with the SQ-headgroup together with more plastic interactions with the aglycones. This suggests that the solute binding protein of *A. tumefaciens*, and related SQ-binding proteins from other sulfoglycolytic pathways, can provide their host organisms direct access to most of the SQ metabolites known to be produced by phototrophs.

1. Introduction

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

The sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP) (Denger et al., 2012, 2014; Sharma et al., 2021), sulfoglycolytic Entner-Doudoroff (sulfo-ED) (Felux et al. 2015a,b; Li et al., 2020) and sulfoglycolytic

sulfofructose transaldolase (sulfo-SFT) (Frommeyer et al., 2020; Y. Liu et al., 2020) pathways involve scission of the six-carbon SQ backbone into two three-carbon fragments: carbons 1–3 are metabolized, while carbons 4–6 are excreted as a three-carbon sulfonate after subsequent reduction to 2,3-dihydroxypropanesulfonate DHPS (by NADH-dependent SLA reductase (YihU) (Sharma et al., 2020)) or oxidation to sulfolactate SL (by NAD⁺-dependent SLA dehydrogenase). The sulfoglycolytic transketolase (sulfo-TK pathway) involves stepwise scission of two carbons (carbons 1 and 2, and carbons 3 and 4) from sulfofructose and transfer to G3P, while carbons 5 and 6 form sulfoacetaldehyde, which is reduced and excreted as isethionate (J. Liu et al., 2021). In contrast to all other known pathways, the sulfoglycolytic SQ monooxygenase (sulfo-SMO) pathway involves the cleavage of the sulfur-carbon bond of SQ with excretion of inorganic sulfur (predominantly as sulfite) and enables the utilization of all six carbons of SQ (Sharma et al., 2022, J. Liu et al., 2021). For sulfoglycolytic pathways to catabolise SQ, it must first be

* Corresponding author.

E-mail address: gideon.davies@york.ac.uk (G.J. Davies).

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

Received 11 January 2022; Received in revised form 3 March 2022; Accepted 3 March 2022

2665-928X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

liberated from imported SQDG or its delipidated forms sulfoquinovosyl monacylglycerol (SQMG) and sulfoquinovosyl glycerol (SQGro). SQ is hydrolyzed from these molecules by ‘gateway’ sulfoquinovosidases, which belong to glycoside hydrolase family GH31 (www.cazy.org) (Abayakoon et al., 2018; Speciale et al., 2016), while import of SQ glycosides is mediated by specialized permeases or transport systems.

The sulfo-SMO pathway of *Agrobacterium tumefaciens* utilizes a two-component system, comprised of an FMN₂-dependent SQ monooxygenase and a flavin reductase, to cleave the carbon-sulfur bond of SQ to form 6-oxo-glucose (6-OG) and sulfite (Sharma et al., 2022), (Fig. 1). Reduction of 6-OG to glucose is catalyzed by an NADPH-dependent 6-OG reductase, enabling the product, glucose, to enter central metabolism. The *smo* gene cluster encodes an ATP-binding cassette (ABC) transport system consisting of a pair of identical ATPase domains (SmoE) and two distinct transmembrane domains (SmoG, SmoH). The ABC transporter engages with the periplasmic solute binding protein SmoF, which binds SQGro with sub-micromolar affinity and recruits it for import into the cell (Sharma et al., 2022). The sulfo-ED pathway gene cluster in *Rhizobium leguminosarum* also contains an ABC transporter and putative SQGro binding protein, suggesting that ABC transporters may be utilized in other sulfoglycolytic pathways in different organisms (Li et al., 2020).

Solute binding proteins, such as maltose binding protein (MalE) and SmoF, are associated with ABC transporters and are involved in the recruitment of the substrate ligand to the transmembrane domains to enable ATP-dependent transport across the membrane (Davidson et al., 2008) (Fig. 1). There are seven classes of ABC transporters (Thomas and

Tampé, 2020), with the specificity and mechanism of the type 1 ABC transporter maltose transporter MalEFGK₂ perhaps the best characterized. Maltose transporter operates in conjunction with a periplasmic substrate binding protein MalE, with maltooligosaccharide loaded-MalE docking with the membrane components MalFGK₂ (Quiocho et al., 1997; Spurlino et al., 1991). In free (apo) form MalE adopts an open conformation, and upon ligand binding MalE undergoes a hinge bending motion to a closed conformation. The adoption of the closed conformation is essential for productive interaction of MalE with the cytoplasmic-membrane components of the ABC transporter complex and importation of maltooligosaccharides across the membrane (Duan et al., 2001). The SQ binding protein SmoF, like other solute binding proteins, has a structural fold comprised of two globular lobes (interconnected by polypeptide chains), which undergo conformational changes upon ligand binding (Berntsson et al., 2010; Sharma et al., 2022). Studies of solute binding proteins show that the ligand-free form undergoes equilibration between open and semi-closed states (Tang et al., 2007). In the case of SmoF, once the open ligand-free form binds SQGro (Sharma et al., 2022), it undergoes a domain rotation to a closed conformation that encapsulates the ligand.

Here, we study the ligand specificity of SmoF, showing that in addition to SQGro, it can bind SQ, the simple glycoside SQMe and, unexpectedly, a short-chain derivative of SQDG. The thermodynamic and structural basis of binding for these ligands is explored. This work suggests that SmoF may facilitate the delivery of both SQ, SQGro and even plant sulfolipids to the ABC transporter and that this transporter system allows acquisition of a range of SQ glycosides by the host organism. We

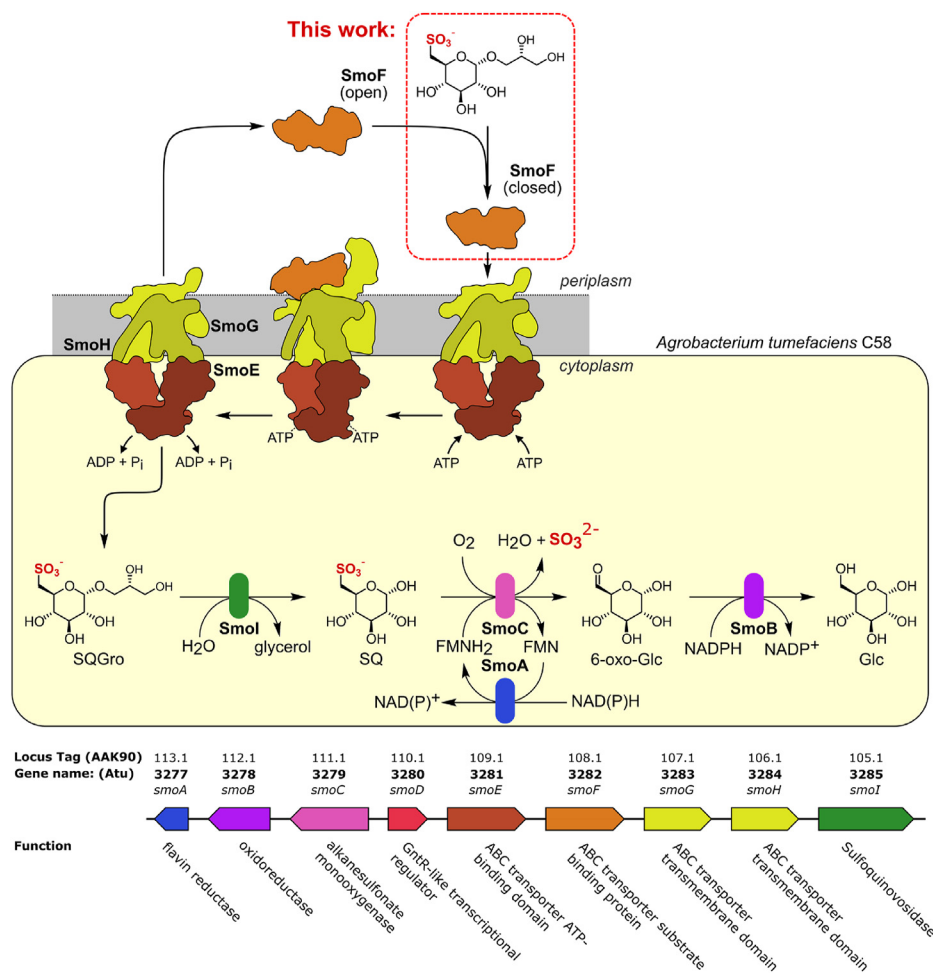


Fig. 1. The sulfoquinovose monooxygenase (sulfo-SMO) pathway of *Agrobacterium tumefaciens*. The SQ-Gro binding protein SmoF works in concert with the ABC transporter SmoE-SmoG-SmoH to transport SQ metabolites into the cytoplasm. SQ glycosides are cleaved by sulfoquinovosidase SmoI, and SQ is converted to glucose and sulfite by FMN₂-dependent monooxygenase SmoC and 6-oxoglucose reductase SmoB.

demonstrate that the residues involved in sulfonate binding are conserved within several other organisms in the taxon *Rhizobiales* but not in putative solute-binding proteins associated with ABC transporters in other sulfoglycolytic bacteria, suggesting that their sulfonate-binding proteins may have evolved other modes of SQ recognition.

2. Materials and methods

2.1. Gene expression and protein purification

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

2.2. Protein crystallization and optimization

SmoF•SQ crystals were grown in a sitting drop using 20 mg ml⁻¹, in 0.1 M NaBr, NaI, 0.1 M imidazole, MES pH 6.9, 13.5% MPD, PEG 1000, PEG 3350 at 10 °C, with a 5:6 mother liquor: protein ratio. SmoF•SQMe crystals were grown using 50 mg ml⁻¹ protein in 0.3 M sodium acetate, 0.1 M BIS-TRIS (pH 5.5) and 35% PEG 2000 MME in a sitting drop, with a 1:1 protein:mother liquor ratio in-drop. In both cases 2 mM ligand was added to protein 10 min prior to drop formation. SmoF•SQDG-(C4:0/C16:0) crystals were grown in sitting drop at 6 °C, with 25 mg ml⁻¹ protein and 2.5 mM SQDG-(C4:0/C16:0) dissolved in DMSO, in 25 mM NaPi, 150 mM NaCl pH 7.0, incubated at room temperature with 2.5 mM SQDG-(C4:0/C16:0) for 10 min prior to crystallization. Diffraction quality crystals were collected from a direct scale up of the Morpheus screen (Molecular Dimensions), condition H12. This contains 0.1 M amino acids (0.2M L-Na-glutamate, 0.2 M alanine, 0.2 M glycine, 0.2 M lysine hydrochloride, 0.2 M serine), 0.1 M buffer system 3 pH 8.5 (1 M TRIS, 1 M bicine) and 50% v/v precipitant mix 4 (25% MPD, 25% PEG 1000, 25% PEG 3350). Crystals only formed in premade mother liquor. No cryoprotectant was used on the resultant crystals due to the presence of cryoprotecting PEG solutions in the mother liquor. Crystals were harvested then flash cooled in liquid nitrogen, using nylon CryoLoops (Hampton).

2.3. Data collection and structure determination

All crystals were tested using a Rigaku MicroMax 007HF X-ray generator with an RAXIS IV++ imaging plate detector. Data was collected at 120 K using a 700 Series Cryostream (Oxford Cryosystems). Diffraction pattern quality assessment and resolution estimate performed using ADXV (Porebski et al., 2013). X-ray data collection occurred at the Diamond Light Source, using beamline I-03 during collection mx18598-51. Data collection statistics are available in Table 1. Data indexing and initial processing for SmoF•SQ and SmoF•SQMe was performed at Diamond, using either DIALS or 3dii automated pipelines from the Xia2 package (Beilsten-Edmands et al., 2020; Winter, 2010). For SmoF with SQDG-(C4:0/C16:0), indexing was performed manually using DUI (Beilsten-Edmands et al., 2020). In all cases AIMLESS was used for data reduction and quality assessment (Evans and Murshudov, 2013). Resolution was cut to CC1/2 = 0.5, or to the highest resolution possible while maintaining an outer shell completeness of 80% or higher. Molecular replacement used either MOLREP or PHASER (Vagin and

Table 1

Data collection and refinement statistics for SmoF structures complexed with SQ, SQMe, SQDG-(C4:0/C16:0).

	SmoF•SQ	SmoF•SQMe	SmoF•SQDG-(C4:0/C16:0)
Data collection			
Space group	P 3 ₁ 2 1	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁
Cell dimensions a, b, c (Å)	102.2, 102.2, 67.96	53.76, 66.27, 99.38	53.22, 69.59, 104.57
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 91.54, 90.0
Resolution (Å)	88.5–1.80 (1.84–1.80)	49.6–1.59 (1.62–1.59)	69.6–2.14 (2.20–2.14)
R _{merge}	0.179 (2.77)	0.280 (1.38)	0.093 (0.305)
R _{pim}	0.06 (0.93)	0.148 (0.909)	0.081 (0.265)
I/σI	10.1 (1.3)	7.7 (1.3)	8.9 (3.6)
CC1/2	1.0 (0.65)	0.98 (0.68)	0.99 (0.94)
Completeness (%)	100 (100)	99.4 (96.1)	99.9 (100)
Redundancy	19.2 (19.1)	6.9 (5.6)	4.1 (4.1)
Refinement			
Resolution (Å)	1.8	1.59	2.14
No. unique reflections	38847	48233	24477
R _{work} /R _{free}	0.20/0.23	0.24/0.27	0.20/0.26
No. atoms			
Protein	5740	5796	11648
Ligand/ion	27	30	142
Water	138	299	239
B-factors (Å ²)			
Protein	32	20	27
Ligand/ion	23	18	26
Water	34	27	29
R.M.S. deviations			
Bond lengths (Å)	0.0146	0.0143	0.0075
Bond angles (°)	1.85	1.79	1.50
Ramachandran Plot Residues			
In most favourable regions (%)	98.4	98.1	97.8
In allowed regions (%)	1.6	1.6	2.2
Outliers	0.0	0.3	0.0
PDB code	7YZS	7YZU	7QHV

Teplyakov, 2010; McCoy et al., 2007) The model used for the SmoF•SQMe structure was 70FY, and the SmoF•SQMe structure was then used for the other two datasets. Early model building was automated using BUCCANEER (Cowtan, 2006). Model refinement was performed using REFMAC5 employing local NCS restraints in the refinement cycles, and all interactive modelling and validation performed in COOT (Emsley & Cowtan, 2004; Murshudov et al., 2011). All steps excluding manual data integration were performed from within the CCP4i2 system (Potterton et al., 2018). Diagram preparation for molecular models was performed using CCP4MG, Pymol or UCSF Chimera, depending on the desired outcome (McNicholas et al., 2011; Pettersen et al., 2004; Schrödinger, 2015). Analyses of conformational changes and internal cavities were performed using the DynDom web server and the CASTp V.3.0 Pymol plugin, respectively (Girdlestone and Hayward, 2016; Tian et al., 2018). We detect anisotropy in SmoF•SQ and SmoF•SQMe datasets as evident from a much higher anisotropic B value for data along the c* axis, affecting the respective data processing statistics. The resolution cut-off for these datasets was chosen based on higher quality of maps and better refinement statistics.

2.4. NanoDSF

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

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

2.5. Isothermal titration calorimetry (ITC)

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

2.6. Bioinformatics

To find sulfo-SMO and sulfo-ED clusters containing SmoF homologues, the protein sequence of *A. tumefaciens* C58 SmoF was submitted to the NCBI psiBLAST algorithm, searching a non-redundant protein sequence (nr) database. Browsing the outputs allowed identification of homologues in sulfo-SMO and sulfo-ED clusters. To find sulfo-EMP clusters containing SmoF homologues, the *E. coli* sulfoquinovosidase (NP_418314.1, locus tag b3878), SQ mutarotase (NP_418315.3, locus tag b3879), SQ isomerase (NP_418316.4, locus tag b3880), SF kinase (NP_418319.2, locus tag b3883), SFP aldolase (NP_418317.1, locus tag b3881), SLA reductase (NP_418318.1, locus tag b3882) and sulfo-EMP regulator (NP_418320.2, locus tag b3884) were submitted separately as queries to the NCBI BLASTp tool. The database searched was the non-redundant protein sequence (nr) database, with *E. coli* (taxid: 562) sequences excluded. Standard algorithm parameters were used, except the maximum target sequences was set to 10,000. The results were filtered, with only protein sequences with E-value $\leq 5.41\text{e-}44$ retained. The corresponding nucleotide accession numbers for each protein from all seven searches were extracted, and the seven lists combined and duplicates removed to give a list of candidate genome sequences. This list was converted into a MultiGeneBLAST reference library and searched using the *E. coli* sulfo-EMP gene cluster as a query. Scripts for this pipeline are available on GitHub (<https://github.com/jmui-unimelb/Gene-Cluster-Search-Pipeline>). Gene clusters possessing a putative SQase, putative SQ isomerase, putative SF kinase and putative SFP aldolase were deemed putative sulfo-EMP operons. These putative sulfo-EMP operons were manually searched to identify their transporter types. Candidate SQBPs were submitted to Clustal 2.1 for multiple sequence alignment, and results were used to generate a cladogram.

3. Results and discussion

The SQ moiety exists in nature as the free sugar SQ, and as glycosides including SQDG and SQGro (Supplementary Fig. S1). To explore the ability of SmoF to bind to different glycosides, we synthesized methyl α -sulfoquinovoside (SQMe), and a naturally occurring SQDG, α -sulfoquinovosyl 1-oleoyl-2-palmitoylglycerol (SQDG-(C18:1/C16:0)) (Zhang et al., 2020). Because the full-length lipids endow this lipoform with poor aqueous solubility we also synthesized a more water-soluble analogue, α -sulfoquinovosyl 1-butanoyl-2-palmitoylglycerol (SQDG-(C4:0/C16:0)), which bears a shorter butanoyl lipid.

Initially, we assessed binding of the analogues to SmoF using nano

differential scanning fluorimetry (nanoDSF). NanoDSF uses tryptophan or tyrosine fluorescence to monitor protein unfolding as a function of temperature and allows calculation of a melting temperature (T_m) that describes the thermodynamic stability of the protein or protein-ligand complex. Ligand-free SmoF had T_m of 43.9 °C, which was raised to 54.2 °C in the presence of 2 mM SQ ($\Delta T_m = 10.3$ °C). Using 2 mM SQMe gave a T_m of 58.5 °C ($\Delta T_m = 14.6$ °C), and in the presence of 2 mM SQDG-(C4:0/C16:0) the T_m of SmoF increased to 51.8 °C ($\Delta T_m = 7.7$ °C). In contrast, 2 mM SQDG-(C18:1/C16:0) did not result in a significant change in T_m (Fig. 2c, Supplementary Fig. S2). This may indicate this long-chain SQDG does not bind, that it binds with no change, or may simply reflect the poor solubility of this glycolipid and the formation of micelles unable to bind SQBP (Supplementary Fig. S1). We next studied the direct binding of these ligands to SmoF by isothermal titration calorimetry (ITC). SQ bound with a K_d value of 2.4 μM , and SQMe bound with a K_d value of 11.5 μM , which are 10-fold and 40-fold weaker affinity compared to SQGro, respectively (Fig. 2a and b, Supplementary Fig. S3). ITC was attempted with SQDG-(C4:0/C16:0) but was unsuccessful, possibly due to the formation of lipid micelles leading to phase separation (Fig. 2c; Supplementary Table S1).

Crystals of complexes of SmoF with SQ, SQMe and SQDG-(C4:0/C16:0) were obtained by co-crystallization with SmoF and diffracted to 1.8, 1.59 and 2.14 Å resolution, respectively (Table 1). Crystal structure of SmoF with SQDG-(C4:0/C16:0) was obtained in P₂₁ space group and contains two copies in the asymmetric unit with no significant differences between the copies. The crystals of SmoF with SQ are in space-group P3₁21, and SmoF•SQMe was obtained in P2₁2₁2₁, each with one copy in the asymmetric unit. Of the 394 residues present in the protein, 386 were present in the SmoF•SQ structure, 387 in the SmoF•SQMe structure, and 392 in the SmoF•SQDG-(C4:0/C16:0) structure. In all three cases, 95% or more of the amino acids in the protein were observed in the electron density, with exceptions occurring primarily in flexible loops and the affinity tag. In all cases SmoF maintains the globular, primarily α -helical structure with the two-domain fold observed previously (Sharma et al., 2022). Within the SmoF•SQ complex, SQ is present as the α -anomer and adopts a ⁴C₁ (chair) conformation. Its binding interactions involve the C6 sulfonate and sugar C2, C3 and C4 hydroxyls (Fig. 3a). The sulfonate occupies a binding pocket comprised of Gln₁₂-Ser₄₃-Gly₁₆₆-Thr₂₂₀. Within this pocket, the side-chain nitrogen of Gln₁₂ and side-chain hydroxyl of Thr₂₂₀ engage in a hydrogen bond with one sulfonate oxygen (2.6 Å, 2.9 Å), the second sulfonate oxygen forms a hydrogen bond to the backbone secondary amine of Gly₁₆₆ (2.8 Å), and the third sulfonate oxyanion forms a hydrogen bond to the backbone secondary amide of Ser₄₃ (2.9 Å) and an ordered water (3.0 Å). The C1 hydroxyl forms a hydrogen bond with a side-chain secondary amine on His₁₃ (2.7 Å). The C2 hydroxyl group forms hydrogen bonds to the side chain carboxyl of Asp₁₁₃ (2.7 Å) and the indole nitrogen of Trp₂₇₆ (2.8 Å). The C3 and C4 hydroxyls each bind one nitrogen of Arg₃₄₅ and the carboxyl oxygen of Asp₆₇ (2.9, 2.7, 3.0, 2.5 Å) (Fig. 3a).

The SmoF•SQMe complex contains small differences in binding recognition compared to SQ. The sulfonate pocket is identical with the exception of Gln₁₂, which is too distant (3.6 Å) from the sulfonate oxygen to form a hydrogen bond. His₁₃, which is on the same loop, is unable to form a hydrogen bond with the C1 oxygen as it is now present as a glycoside in SQMe. All other interactions are identical to those observed with SQ (Fig. 3b). The SmoF•SQDG-(C4:0/C16:0) complex involves identical interactions as for SQMe with the C2-4 hydroxyl groups, and within the sulfonate binding pocket. The palmitic acid chain protrudes through the top of the binding pocket and forms a crystal contact with another SmoF molecule in the asymmetric unit (Supplementary Fig. S4).

SQ and SQMe reside within an internal cavity that entirely encloses the ligand (Fig. 3d and e). For SQ the volume is 297 Å³ and for SQMe is 447 Å³, 66% larger. The cavity for SQGro is 476 Å³, 6% larger than SQMe (Supplementary Fig. S5). As noted above, SQDG-(C4:0/C16:0) is not fully enclosed by the protein and the cavity features three openings. One of these is occupied by the protruding palmitoyl chain, while the other two

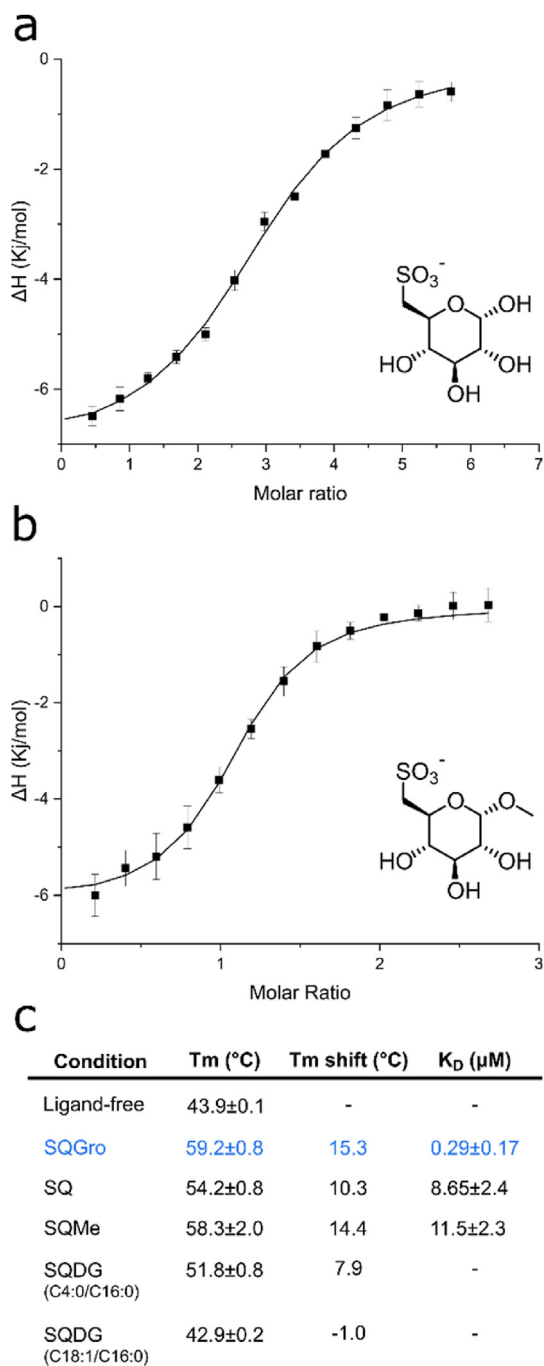


Fig. 2. Binding affinity of SQ and its glycosides for SmoF. a) Isothermal titration calorimetry showing titration of SQ into SmoF. b) Calorimogram of SQMe into SmoF. c) Melting temperature (T_m) of SmoF, as determined by differential scanning fluorimetry, the T_m shift relative to apo-SmoF, and K_d values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

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

The structures of the SmoF complexes with SQ, SQMe and SQDG-(C4:0/C16:0) show large conformational changes relative to the ligand-

free form, as observed previously with SQGro (Sharma et al., 2022), and undergo interdomain rotations of up to 33° compared to the unliganded state (Fig. 3g). This movement centers around a pair of hinges, which are found in other SBPs. MalE features a comparable closure angle upon ligand binding (37°) (Tang et al., 2007) (Supplementary Fig. S6). In the SmoF•SQDG-(C4:0/C16:0) complex, ligand binding is accompanied by an upwards deflection in α -helix 1 by 9.8 Å. This moves Gln₁₂ and His₁₃ away from SQDG-(C4:0/C16:0) allowing its large lipid groups to bind (Supplementary Figs. S7 and S8). Collectively, this data shows that SmoF retains its interactions around the sulfosugar yet has sufficient conformational plasticity to accommodate larger aglycones. For the diacyl glycerol substituent this enables binding even though the entire lipid chain cannot be contained within the binding pocket.

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

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

4. Conclusions and future work

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

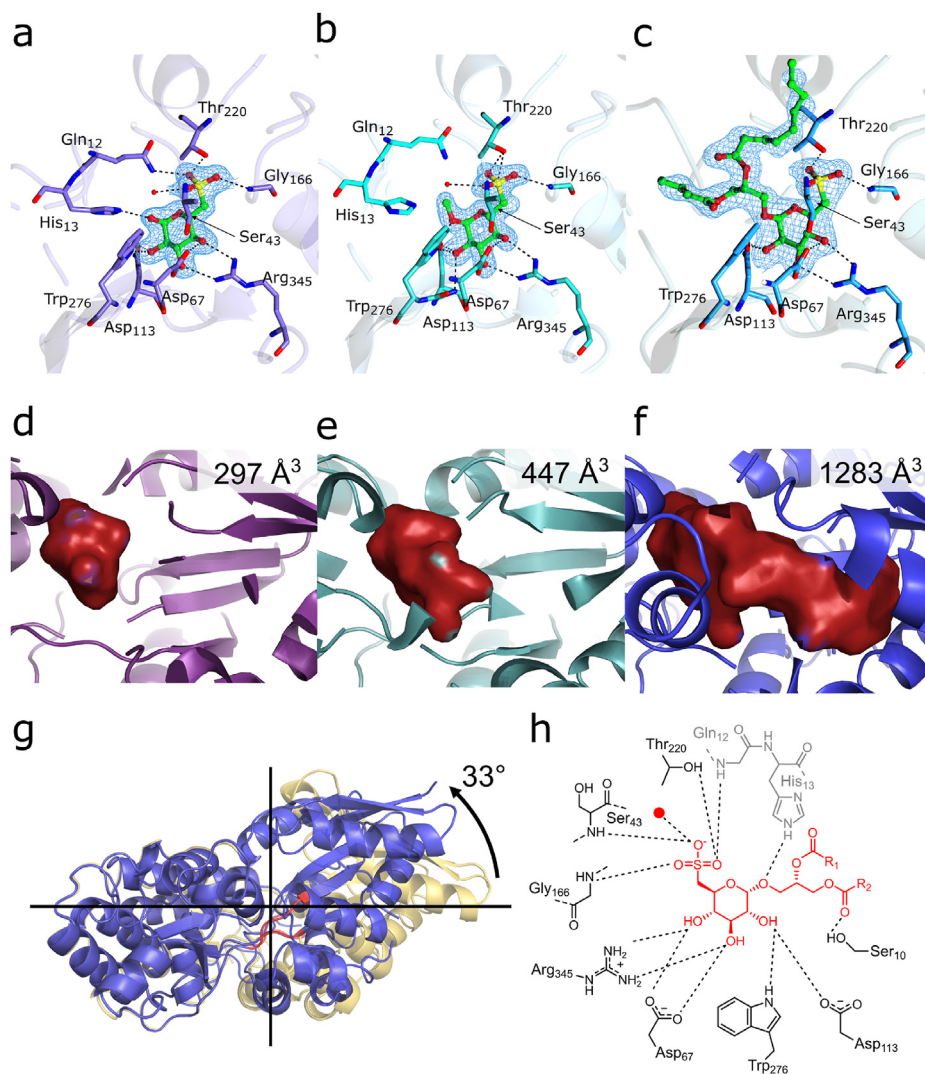


Fig. 3. Structural basis and induced conformational changes for binding of SQ and its derivatives to SmoF. SmoF-ligand complex formation with a) SQ, b) SQMe, and c) SQDG-(C4:0/C16:0). d-f) Internal cavities filled by SQ, SQMe and SQDG-(C4:0/C16:0), as detected using the CASTp server. g) Superposition of ligand-free SmoF (yellow) and complex with SQDG-(C4:0/C16:0) (blue). Hinge angle and domain selection performed using the DynDom web server, and hinge residues highlighted in red. h) Binding interactions of SQ and glycosides. Interactions present in SQ complex but not SQMe or SQDG-(C4:0/C16:0) in grey. In all cases electron density (2Fo-Fc) has been contoured to 1.0 σ or 0.44 $e/\text{\AA}^3$ for SmoF•SQ, 0.61 $e/\text{\AA}^3$ for SmoF•SQMe and 0.44 $e/\text{\AA}^3$ for SmoF•SQDG-(C16:0/C4:0). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

allow capture of free SQ, SQGro and even lipidated SQ glycosides such as SQMG and SQDG, allowing metabolism of the lipidic part in addition to the SQ and the glycerol. Previously, there has been no evidence that SQDG can be completely metabolized by sulfoglycolytic organisms. Instead, various non-specific lipases have been reported that can cleave the lipid chains (Snow et al., 2021), suggesting that the sulfo-SMO pathway is used in partnership with non-sulfoglycolytic organisms (possibly including plants) that excise and metabolize the energy-rich lipid chains of SQDG, releasing the sulfo-sugar SQGro. However, the ability of SmoF to bind SQDG suggests that *A. tumefaciens* can on its own achieve the import of SQDG and SQMG. Within this scenario, SmoF, working in concert with *A. tumefaciens* sulfoquinovosidase SmoI, which is expressed with a signal peptide that will direct expression to the periplasm, enables capture of the full carbon-content of SQDG/SQMG. Possibly, this could allow *A. tumefaciens* to utilize intact sulfolipids as a nutrient upon infecting a plant host.

A search for other ABC transporters and associated solute binding proteins in sulfoglycolytic gene clusters led to identification of ABC transporter systems similar to that of *A. tumefaciens* in organisms with gene clusters encoding sulfo-SMO, sulfo-ED, sulfo-EMP, sulfo-SFT and sulfo-TK pathways. This complements earlier reports showing that sulfo-EMP and sulfo-ED (Denger et al., 2014; Felux et al. 2015a,b) gene clusters also contained TauE transporters of the 4-toluene sulfonate uptake permease (TSUP) family (Shlykov et al., 2012) to import SQ and its glycosides. The occurrence of TSUP family or ABC transporter systems in

various sulfoglycolysis gene clusters suggests that the specific transporter used to import the sulfo-sugar substrate is not restricted to a particular pathway. Sequence alignment of putative SQ-binding proteins from this range of organisms revealed that SQ binding residues identified in *A. tumefaciens* SmoF are not well conserved, and thus that acquisition of SQ-binding function may have arisen through independent evolutionary events. Thus, sequence-based searches for new SQ-binding proteins may have poor predictive power, and will require consideration of genetic context and whether the solute binding protein and ABC transporter are associated with a sulfoglycolytic gene cluster. Finally, the ability of SmoF to bind SQ glycosides bearing extended lipid chains means it may be possible to exploit this SQ-binding protein to bind to SQ-linked structures for affinity-based protein capture and purification applications, in a way analogous to the use of maltose-binding protein that binds its cognate ligands ($K_d = 0.5\text{--}2\ \mu\text{m}$) with similar affinities.

Funding

GJD is supported by the Royal Society Ken Murray Research Professorship. MS and AS were funded by the Leverhulme Trust (grant RPG-2017-190) and MS subsequently by the Biotechnology and Biological Sciences Research Council (BB/W003805/1). SJW is supported by the Australian Research Council (DP180101957 and DP210100233). EDGB is supported by the Brian M. Davis Charitable Foundation Centenary Fellowship, National Health and Medical Research Council of Australia

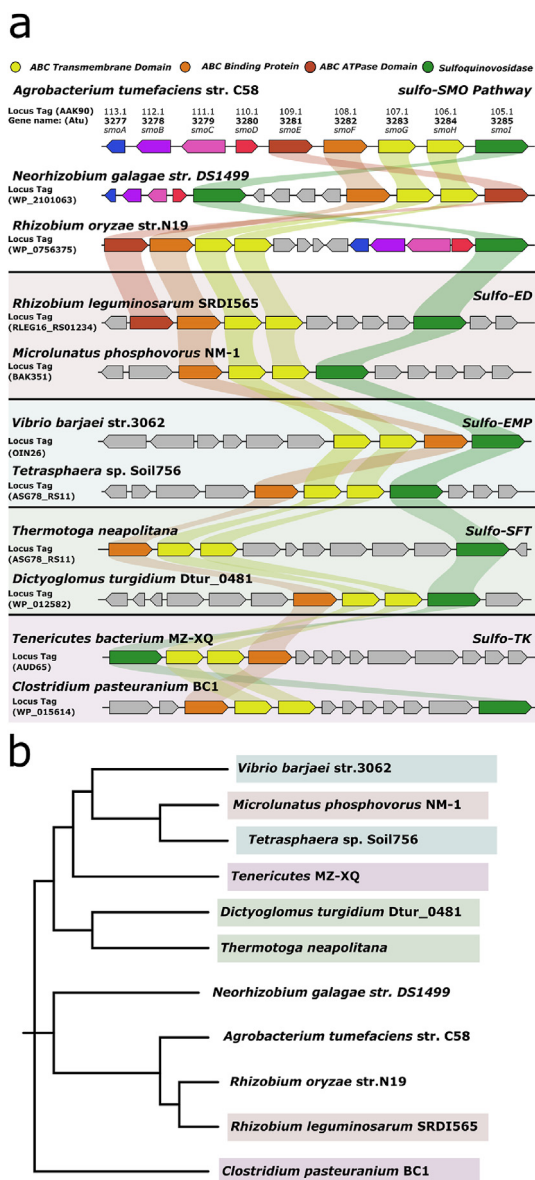


Fig. 4. Comparison of *A. tumefaciens* sulfo-SMO gene cluster with other proposed sulfoglycolytic gene clusters containing SmoF homologues. a) Gene clusters for sulfoglycolytic sulfo-SMO, sulfo-ED, sulfo-EMP, sulfo-TAL and sulfo-TK pathways containing ABC transporters featuring a sulfoquinovose binding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

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

CRedit authorship contribution statement

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

Declaration of competing interest

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

Acknowledgements

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

Appendix A. Supplementary data

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

References

- Abayakoon, P., Jin, Y., Lingford, J.P., Petricevic, M., John, A., Ryan, E., Wai-Ying Mui, J., Pires, D.E.V., Ascher, D.B., Davies, G.J., Goddard-Borger, E.D., Williams, S.J., 2018. Structural and biochemical insights into the function and evolution of sulfoquinovosidases. *ACS Cent. Sci.* 4, 1266–1273. <https://doi.org/10.1021/acscentsci.8b00453>.
- Beale, J., Lee, S.Y., Iwata, S., Beis, K., 2010. Structure of the aliphatic sulfonate-binding protein SsuA from *Escherichia coli*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 66, 391–396. <https://doi.org/10.1107/S1744309110006226>.
- Beilstein-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D., Evans, G., 2020. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr. Sect. D Struct. Biol.* 76, 385–399. <https://doi.org/10.1107/S2059798320003198>.
- Benson, A.A., Daniel, H., Wisner, R., 1959. A sulfolipid in plants. *Proc. Natl. Acad. Sci. Unit. States Am.* 45, 1582–1587. <https://doi.org/10.1073/pnas.45.11.1582>.
- Berntsson, R.P.A., Smits, S.H.J., Schmitt, L., Slotboom, D.J., Poolman, B., 2010. A structural classification of substrate-binding proteins. *FEBS Lett.* 584, 2606–2617. <https://doi.org/10.1016/j.febslet.2010.04.043>.
- Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 62, 1002–1011. <https://doi.org/10.1107/S0907444906022116>.
- Davidson, A.L., Dassa, E., Orelle, C., Chen, J., 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* 72, 317–364. <https://doi.org/10.1128/mmr.00031-07>.
- Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D., Cook, A.M., 2012. Sulfoquinovose degraded by pure cultures of bacteria with release of C3-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.* 328, 39–45. <https://doi.org/10.1111/j.1574-6968.2011.02477.x>.
- Denger, K., Weiss, M., Felux, A.K., Schneider, A., Mayer, C., Spitteller, D., Huhn, T., Cook, A.M., Schleheck, D., 2014. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* 507, 114–117. <https://doi.org/10.1038/nature12947>.
- Duan, X., Hall, J.A., Nikaido, H., Quiocho, F.A., 2001. Crystal structures of the maltodextrin/maltose-binding protein complexed with reduced oligosaccharides: flexibility of tertiary structure and ligand binding. *J. Mol. Biol.* 306, 1115–1126. <https://doi.org/10.1006/jmbi.2001.4456>.
- Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 1204–1214. <https://doi.org/10.1107/S0907444913000061>.
- Felux, A.K., Franchini, P., Schleheck, D., 2015a. Permanent draft genome sequence of sulfoquinovose-degrading *Pseudomonas putida* strain SQ1. *Stand. Genomic Sci.* 10, 1–6. <https://doi.org/10.1186/s40793-015-0033-x>.
- Felux, A.K., Spitteller, D., Klebensberger, J., Schleheck, D., 2015b. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4298–E4305. <https://doi.org/10.1073/pnas.1507049112>.
- Frommeyer, B., Fiedler, A.W., Oehler, S.R., Hanson, B.T., Loy, A., Franchini, P., Spitteller, D., Schleheck, D., Hanson, S.R., Loy, B.T., 2020. Environmental and intestinal phylum Firmicutes bacteria metabolize the plant sugar sulfoquinovose via a 6-deoxy-6-sulfofructose transaldolase pathway. *iScience* 23, 101510. <https://doi.org/10.1016/j.isci.2020.101510>.
- Girdlestone, C., Hayward, S., 2016. The DynDom3D webserver for the analysis of domain movements in multimeric proteins. *J. Comput. Biol.* 23, 21–26. <https://doi.org/10.1089/cmb.2015.0143>.
- Goddard-Borger, E.D., Williams, S.J., 2017. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem. J.* 474, 827–849. <https://doi.org/10.1042/BCJ20160508>.
- Harwood, J.L., Nicholls, R.G., 1979. The plant sulpholipid—a major component of the sulphur cycle. *Biochem. Soc. Trans.* 7, 440–447. <https://doi.org/10.1042/bst0070440>.

- Li, J., Epa, R., Scott, N.E., Skoneczny, D., Sharma, M., Snow, A.J.D., Lingford, J.P., Goddard-Borger, E.D., Davies, G.J., McConville, M.J., Williams, S.J., 2020. A sulfolipid biosynthetic pathway in *Rhizobium leguminosarum* bv. *trifolii* SRD1565. *Appl. Environ. Microbiol.* 86. <https://doi.org/10.1128/AEM.00750-20>.
- Liu, J., Wei, Y., Ma, K., An, J., Liu, X., Liu, Y., Ang, E.L., Zhao, H., Zhang, Y., 2021. Mechanistically diverse pathways for sulfoquinovose degradation in bacteria. *ACS Catal.* 11, 14740–14750. <https://doi.org/10.1021/acscatal.1c04321>.
- Liu, Y., Wei, Y., Zhou, Y., Ang, E.L., Zhao, H., Zhang, Y., 2020. A transaldolase-dependent sulfolipid biosynthesis pathway in *Bacillus megaterium* DSM 1804. *Biochem. Biophys. Res. Commun.* 533, 1109–1114. <https://doi.org/10.1016/j.bbrc.2020.09.124>.
- Loll, B., Kern, J., Saenger, W., Zouni, A., Biesiadka, J., 2005. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* 438, 1040–1044. <https://doi.org/10.1038/nature04224>.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40 (4), 658–674. <https://doi.org/10.1107/S0021889807021206>.
- McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 386–394. <https://doi.org/10.1107/S0907444911007281>.
- Mizusawa, N., Wada, H., 2012. The role of lipids in photosystem II. *Biochim. Biophys. Acta Bioenerg.* <https://doi.org/10.1016/j.bbabi.2011.04.008>.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 355–367. <https://doi.org/10.1107/S0907444911001314>.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- Porebski, B.T., Ho, B.K., Buckle, A.M., 2013. Interactive visualization tools for the structural biologist. *J. Appl. Crystallogr.* 46, 1518–1520. <https://doi.org/10.1107/S0021889813017858>.
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S.J., Nicholls, R.A., Noble, M., Pannu, N.S., Roth, C., Sheldrick, G., Skubak, P., Turkmenburg, J., Uski, V., Von Delft, F., Waterman, D., Wilson, K., Winn, M., Wojdyr, M., 2018. CCP 4 i 2: the new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D Struct. Biol.* 74, 68–84. <https://doi.org/10.1107/S2059798317016035>.
- Qu, F., ElOmari, K., Wagner, A., De Simone, A., Beis, K., 2019. Desolvation of the substrate-binding protein TauA dictates ligand specificity for the alkanesulfonate ABC importer TauABC. *Biochem. J.* 476, 3649–3660. <https://doi.org/10.1042/BCJ20190779>.
- Quioco, F.A., Spurlino, J.C., Rodseth, L.E., 1997. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* 5, 997–1015. [https://doi.org/10.1016/S0969-2126\(97\)00253-0](https://doi.org/10.1016/S0969-2126(97)00253-0).
- Sato, N., Kobayashi, S., Aoki, M., Umemura, T., Kobayashi, I., Tsuzuki, M., 2016. Identification of genes for sulfolipid synthesis in primitive red alga *Cyanidioschyzon merolae*. *Biochem. Biophys. Res. Commun.* 470, 123–129. <https://doi.org/10.1016/j.bbrc.2016.01.006>.
- Schrödinger, L., 2015. *The PyMol Molecular Graphics System*.
- Sharma, M., Abayakoon, P., Epa, R., Jin, Y., Lingford, J.P., Shimada, T., Nakano, M., Mui, J.W.Y.-Y., Ishihama, A., Goddard-Borger, E.D., Davies, G.J., Williams, S.J., 2021. Molecular basis of sulfosugar selectivity in sulfolipid biosynthesis. *ACS Cent. Sci.* 7, 476–487. <https://doi.org/10.1021/acscentsci.0c01285>.
- Sharma, M., Abayakoon, P., Lingford, J.P., Epa, R., John, A., Jin, Y., Goddard-Borger, E.D., Davies, G.J., Williams, S.J., 2020. Dynamic structural changes accompany production of dihydroxypropanesulfonate by Sulfolactaldehyde reductase. *ACS Catal.* 10, 2826–2836. <https://doi.org/10.26434/chemrxiv.9725120.v2>.
- Sharma, M., Lingford, J.P., Petricevic, M., Snow, A.J.D., Zhang, Y., Järvä, M.A., Mui, J.W.-Y., Scott, N.E., Saunders, E.C., Mao, R., Epa, R., Silva, B.M. da, Pires, D.E.V., Ascher, D.B., McConville, M.J., Davies, G.J., Williams, S.J., Goddard-Borger, E.D., 2022. Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proc. Natl. Acad. Sci. Unit. States Am.* 119, e2116022119. <https://doi.org/10.2185/jjrm.23.202>.
- Shlykov, M.A., Zheng, W.H., Chen, J.S., Saier, M.H., 2012. Bioinformatic characterization of the 4-Toluene Sulfonate Uptake Permease (TSUP) family of transmembrane proteins. *Biochim. Biophys. Acta Biomembr.* 1818, 703–717. <https://doi.org/10.1016/j.bbame.2011.12.005>.
- Snow, A.J.D., Burchill, L., Sharma, M., Davies, G.J., Williams, S.J., 2021. Sulfolipid biosynthesis: catabolic pathways for metabolism of sulfoquinovose. *Chem. Soc. Rev.* 50, 13628–13645.
- Speciale, G., Jin, Y., Davies, G.J., Williams, S.J., Goddard-Borger, E.D., 2016. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* 12, 215–217. <https://doi.org/10.1038/nchembio.2023>.
- Spurlino, J.C., Lu, G.Y., Quioco, F.A., 1991. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* 266, 5202–5219. [https://doi.org/10.1016/S0021-9258\(19\)67774-4](https://doi.org/10.1016/S0021-9258(19)67774-4).
- Tang, C., Schwieters, C.D., Clore, G.M., 2007. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449, 1078–1082. <https://doi.org/10.1038/nature06232>.
- Thomas, C., Tampé, R., 2020. Structural and mechanistic principles of ABC transporters. *Annu. Rev. Biochem.* 89, 605–636. <https://doi.org/10.1146/annurev-biochem-011520-105201>.
- Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., 2018. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46, W363–W367. <https://doi.org/10.1093/nar/gky473>.
- Tófoli De Araújo, F., Bolanos-Garcia, V.M., Pereira, C.T., Sanches, M., Oshiro, E.E., Ferreira, R.C.C., Chigardze, D.Y., Gonçalves Barbosa, J.A., Ferreira, L.C.D.S., Benedetti, C.E., Blundell, T.L., Balan, A., 2013. Structural and physiological analyses of the alkanesulfonate-binding protein (SsuA) of the citrus pathogen *Xanthomonas citri*. *PLoS One* 8, e80083. <https://doi.org/10.1371/journal.pone.0080083>.
- Vagin, A., Teplyakov, A., 2010. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (1), 22–25. <https://doi.org/10.1107/S0907444909042589>.
- Winter, G., 2010. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* 43 (1), 186–190. <https://doi.org/10.1107/S0021889809045701>.
- Zhang, Y., Mui, J.W.Y., Arumaperuma, T., Lingford, J.P., Goddard-Borger, E.D., White, J.M., Williams, S.J., 2020. Concise synthesis of sulfoquinovose and sulfoquinovosyl diacylglycerides, and development of a fluorogenic substrate for sulfoquinovosidases. *Org. Biomol. Chem.* 18, 675–686. <https://doi.org/10.1039/c9ob02540e>.