PNAS
www.pnas.org

Supplementary Information for

Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria

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

Gideon J. Davies
Email: gideon.davies@york.ac.uk

Spencer J. Williams
Email: sjwill@unimelb.edu.au

Ethan D. Goddard-Borger
Email: goddard-borger.e@wehi.edu.au

This PDF file includes:

Figures S1 to S20
Tables S1 to S6

Supplementary Figures

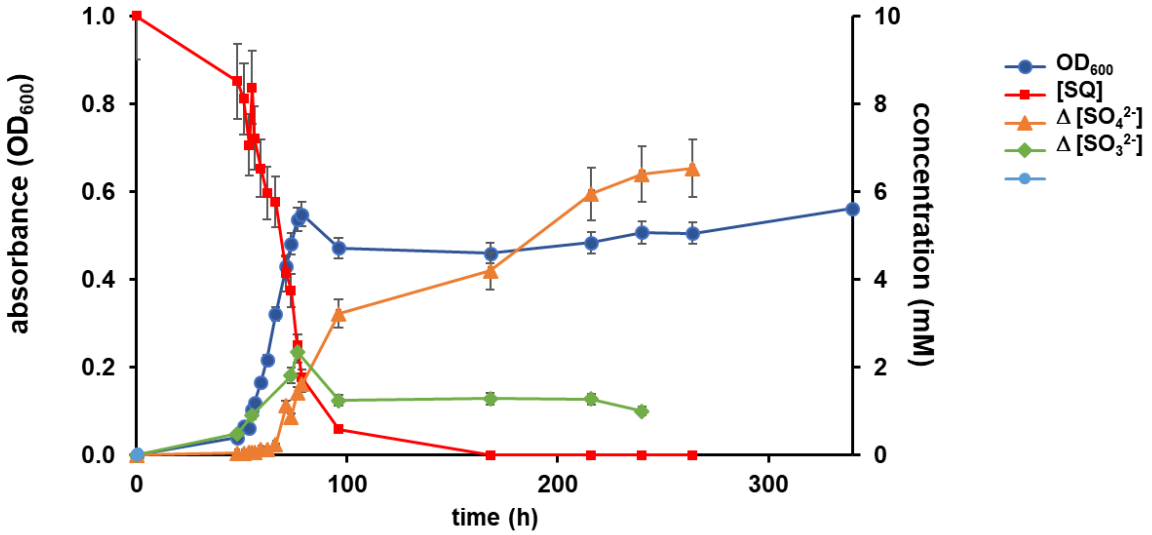


Fig. S1: Independent replicate of data in Figure 1a. Optical density of *A. tumefaciens* C58 culture (blue) and [SQ] (red), change in [sulfite] (green) and change in [sulfate] (yellow), with respect to time. Error bars denote observational error (derived by propagation of estimated random errors).

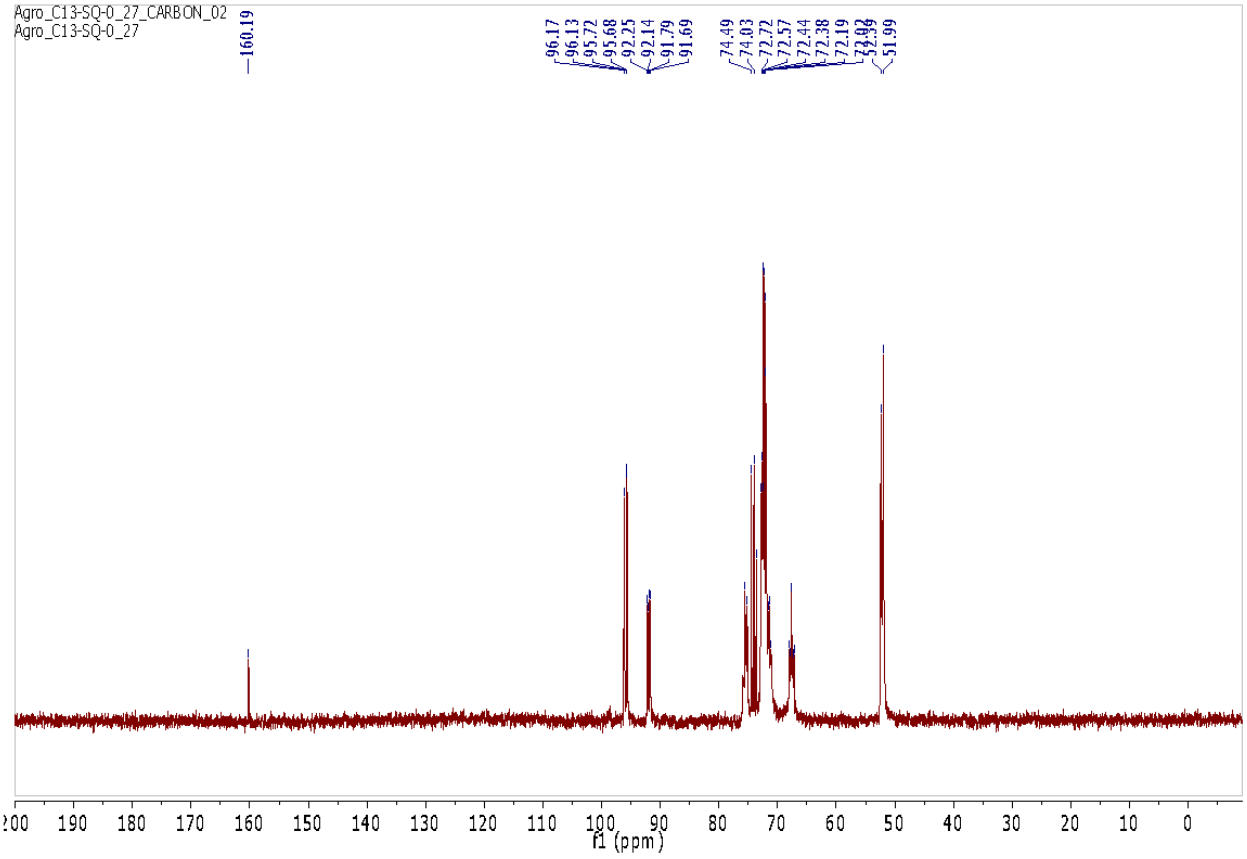


Fig. S2: $^{13}C\{^{1}H\}$ NMR (100 MHz) spectrum of *A. tumefaciens* C58 culture grown on M9 media supplemented with 10 mM ($^{13}C_6$)SQ at an OD_{600} of 0.49. ^{13}C-labelled bicarbonate is observed at δ = 160.2 ppm.

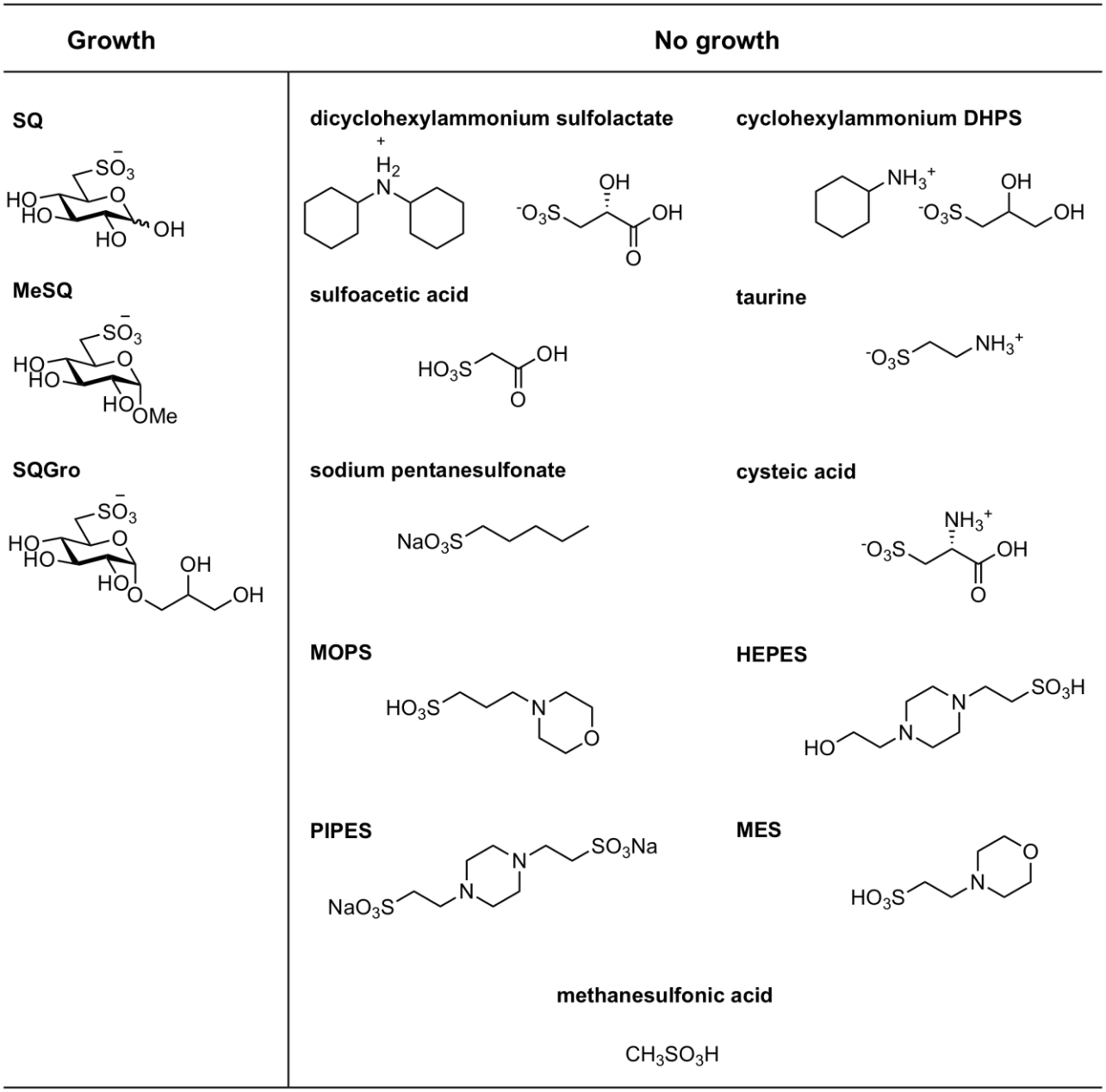


Fig. S3: Sulfonate substrates investigated as substrates for growth of *A. tumefaciens* C58 in M9 minimal media.

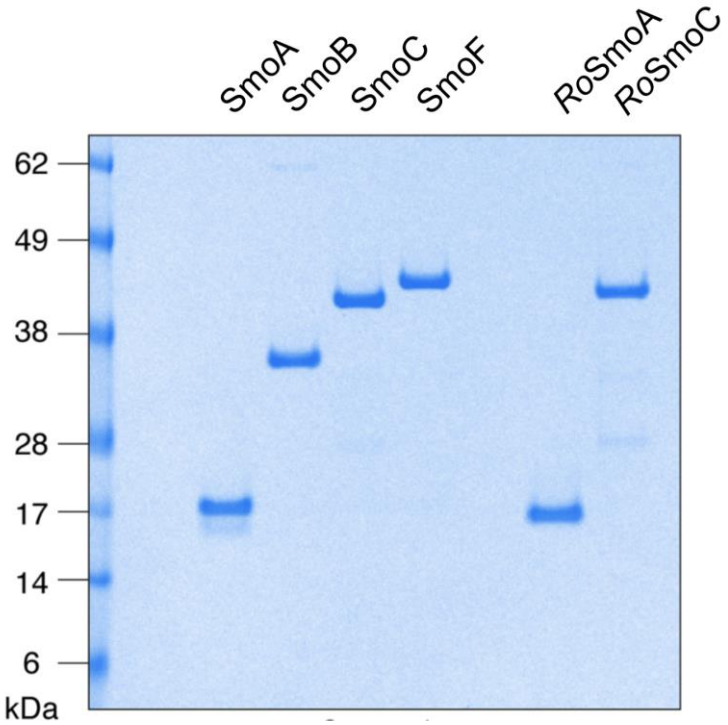


Fig. S4: Coomassie-stained SDS-PAGE gel of all recombinant proteins used in this study (5 µg loaded per well).

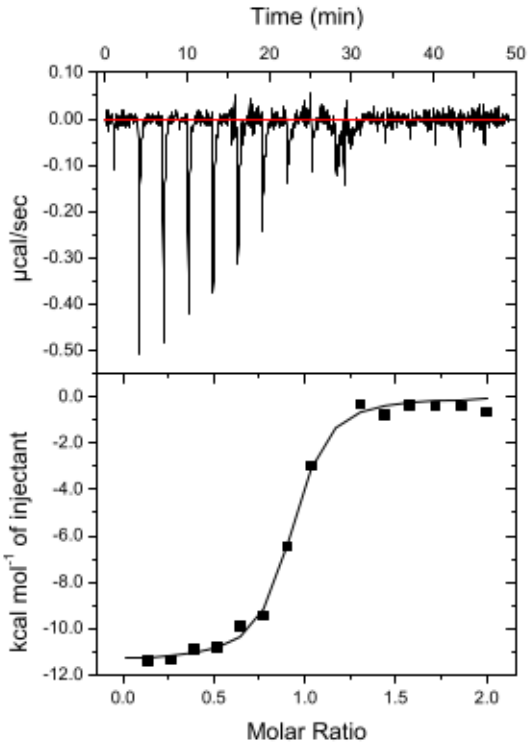
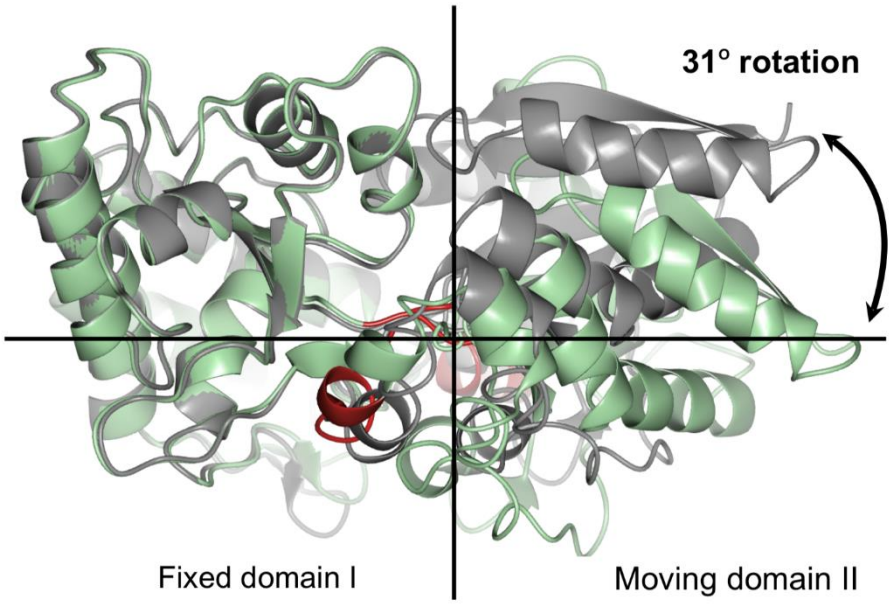


Fig. S5: Independent replicate of data in Figure 2a. Isothermal titration calorimogram for SmoF titrated against its cognate ligand 2'*R*-SQGro.



Property	Value
Fixed Domain (208 residues, r.m.s.d. 0.67 Å)	72 - 75, 113 - 275, 343 - 387
Moving Domain (168 residues, r.m.s.d. 1.87Å)	4 - 71, 76 - 112, 276 - 342
Rotation Angle (deg)	30.6
Translation (A)	-0.5
Closure (%)	99.7
Bending Residues (red)	71 - 76, 112 - 114, 273 - 276, 336 - 343

Fig. S6: Conformational changes occurring upon binding of 2'*R*-SQGro to SmoF. DynDom analysis of the dynamic domains and hinge-bending motion of the SmoF(Atu3282)-apo structure (green) *vs.* 2'*R*-SQGro•SmoF(Atu3282) complex structure in 'closed' conformation (grey). The *x* and *y* axes cross at the centre of rotation and the hinge (*z*) axis is perpendicular to the origin.

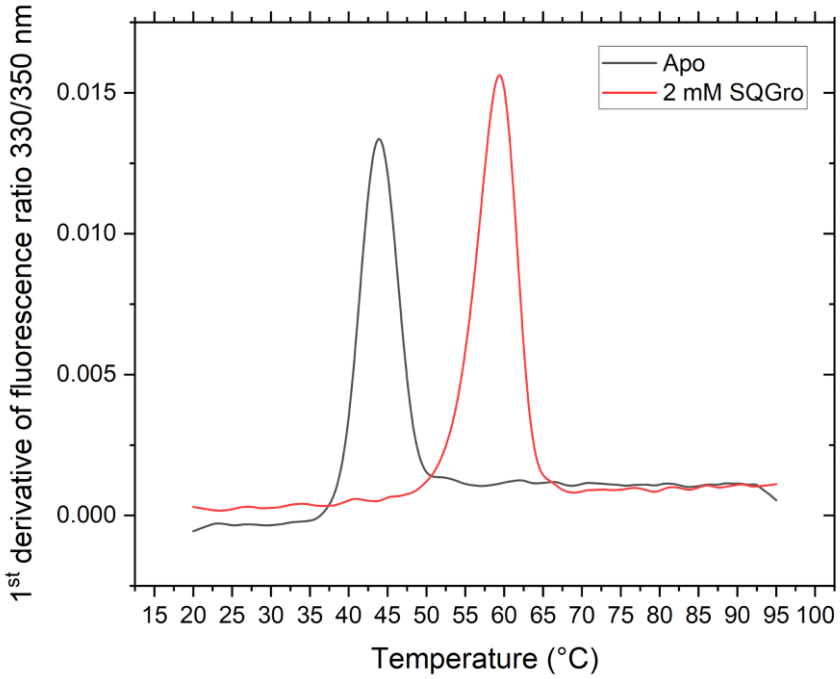


Fig. S7: Thermal stability of SmoF in ligand-free state and with bound SQGro. Differential scanning fluorimetry (DSF) was used to determine the melting temperature (T_m) of SmoF in the presence and absence of SQGro. SQGro binding to SmoF increased the protein's T_m by 15.3 °C (from 43.9 °C to 59.2 °C).

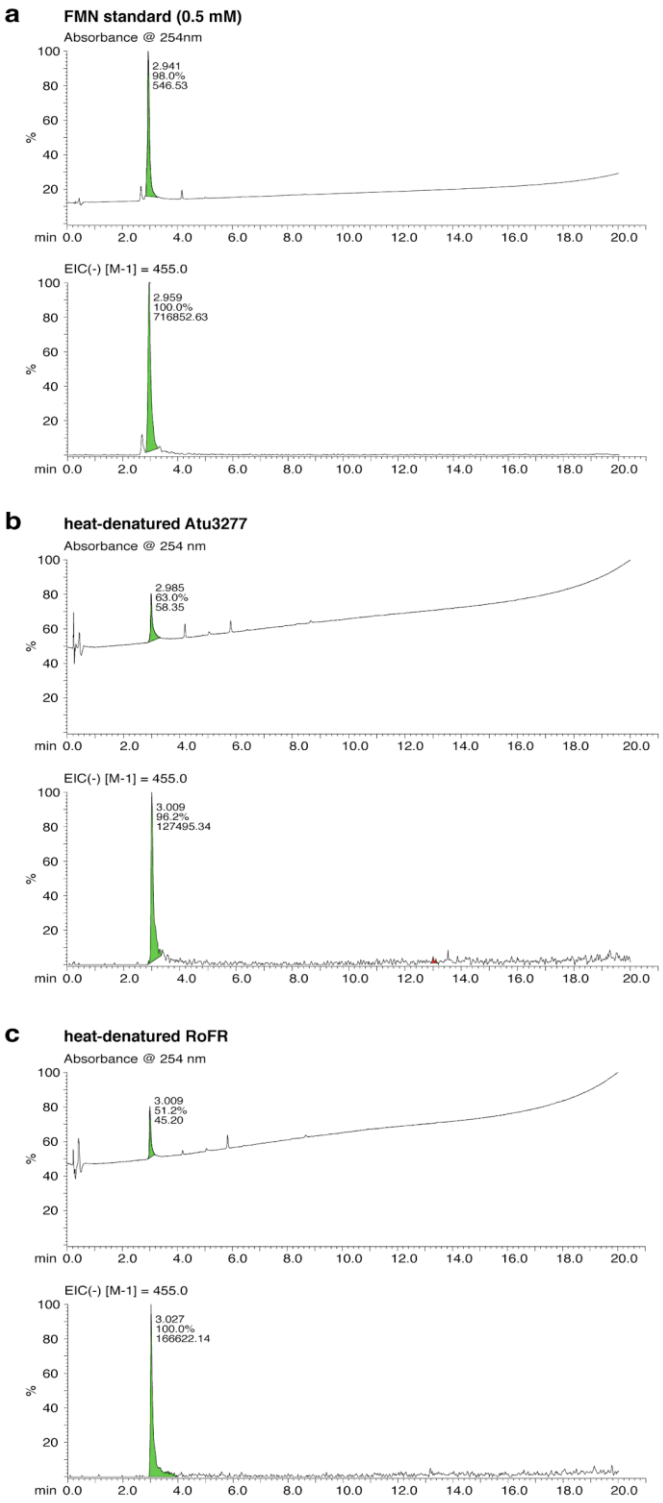


Fig. S8: LC-MS analysis of the supernatant from heat-denatured samples of recombinant SmoA (Atu3277) and *Ro*SmoA. (a) Chromatograms showing absorbance and *m*/*z* 455 signals for authentic FMN. (b) Chromatograms showing absorbance and *m*/*z* 455 signals for heat-denatured SmoA. (c) Chromatograms showing absorbance and *m*/*z* 455 signals for heat-denatured *Ro*SmoA.

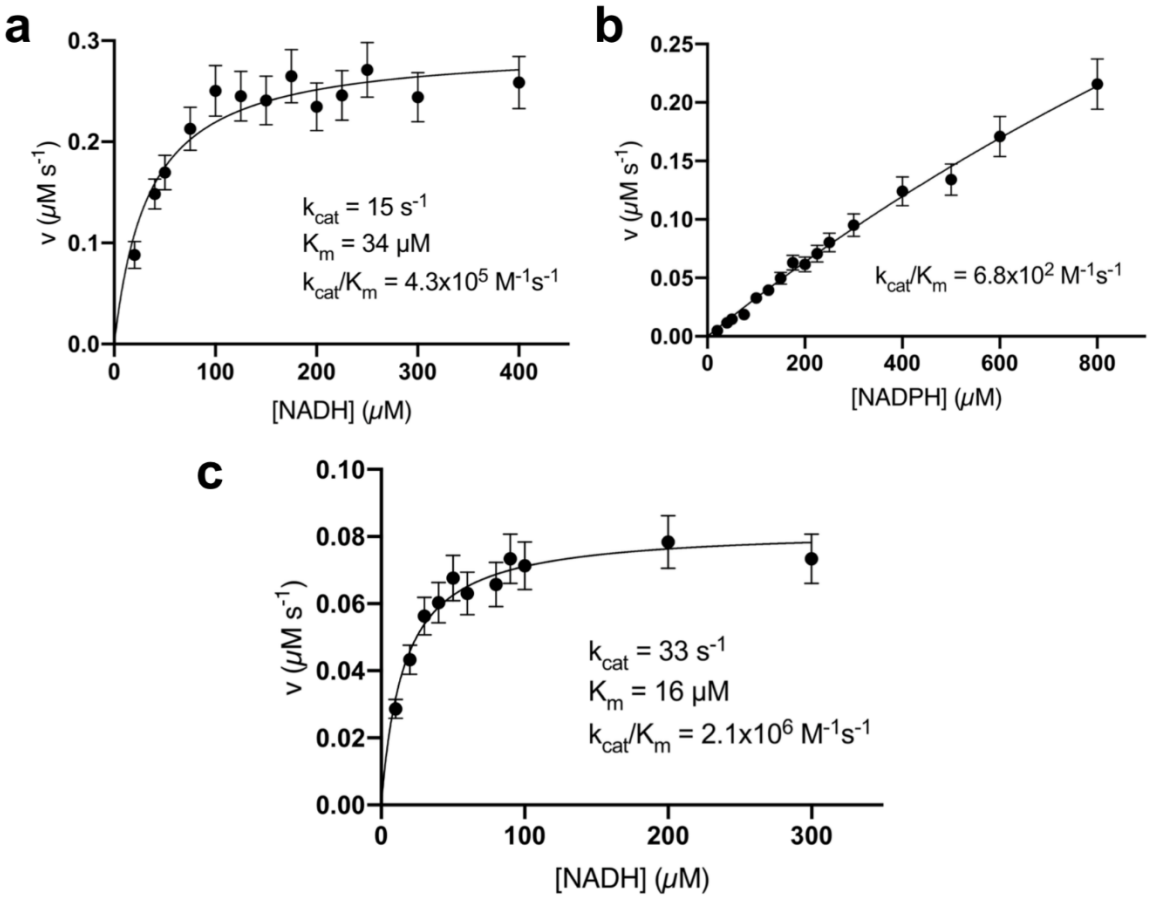


Fig. S9: Kinetic analysis of SmoA and *Ro*SmoA. Michaelis-Menten kinetics for SmoA-catalysed reduction of FMN (at 30 µM) by (a) NADH and (b) NADPH, illustrating that NADH is the preferred nicotinamide cofactor for SmoA. (c) Michaelis-Menten kinetics for *Ro*SmoA-catalysed reduction of FMN (at 30 µM) by NADH. Error bars denote observational errors (derived by propagation of estimated random error).

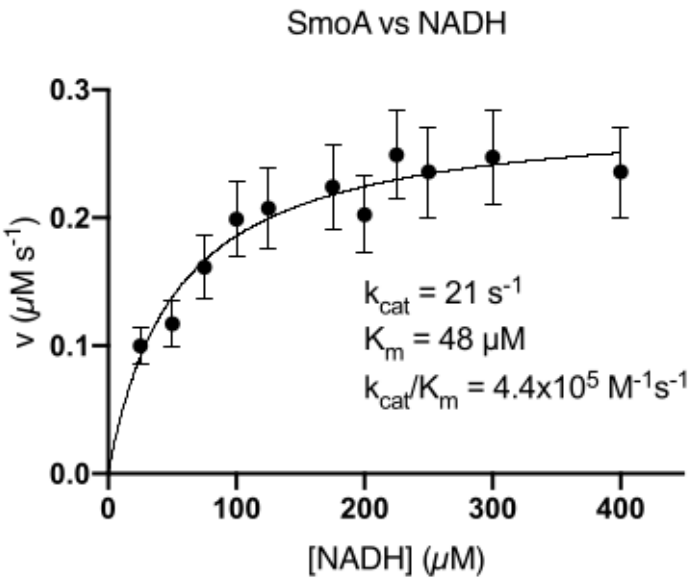


Fig. S10: Independent replicate of data in Figure 3a. Michaelis-Menten kinetics for SmoA-catalysed reduction of FMN by NADH. Error bars denote observational errors (derived by propagation of estimated random error).

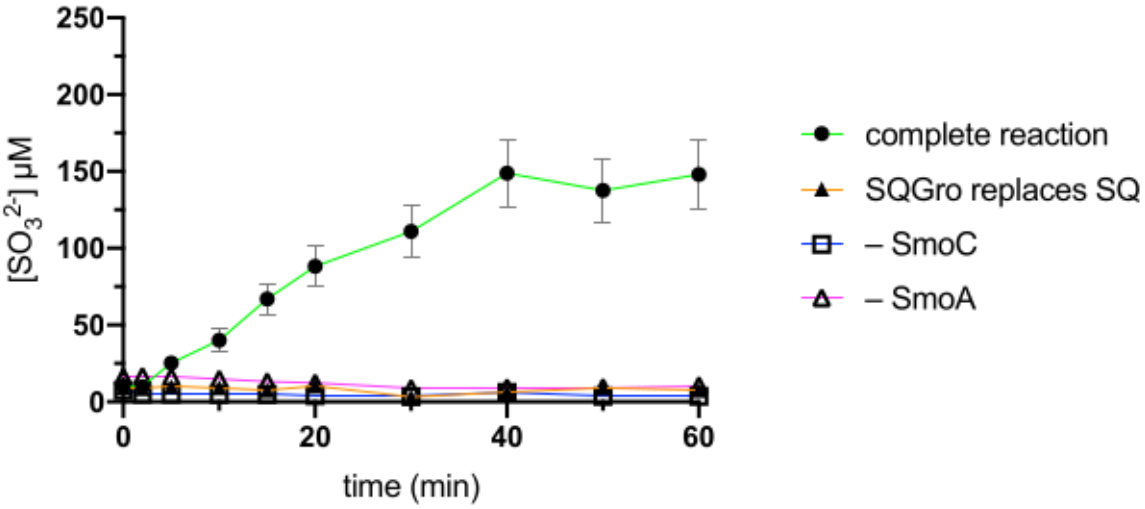


Fig. S11: Independent replicate of data in Figure 3b. SmoC activity assessed using sulfite release assay with Ellman's reagent in the presence of FMN, flavin reductase, NADH and SQ. The data is representative of three independent experiments, error bars denote observational error (derived by propagation of estimated random errors).

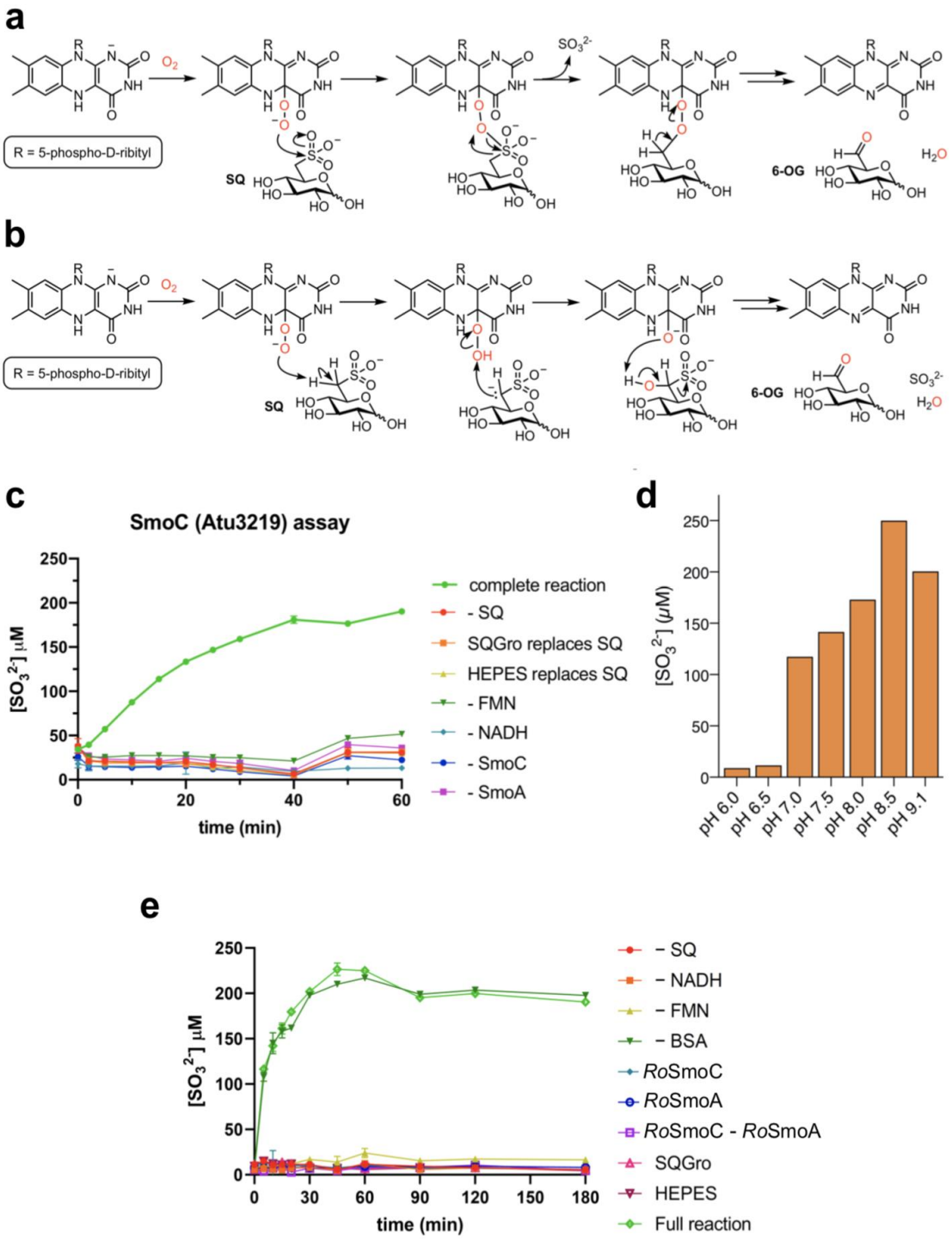


Fig. S12: Proposed mechanisms and biochemical characterization of SmoC and *Ro*SmoC. (a) Proposed SQ monooxygenase mechanism involving nucleophilic attack at sulfur by a C4a-peroxide (alternatively, an N5-peroxide may be invoked). (b) Proposed SQ monooxygenase mechanism involving oxidation at C6 of SQ (alternatively, an N5-peroxide may be invoked). (c) Sulfite release assay using Ellman's reagent to quantify SmoC activity in absence (–) or presence of FMN, flavin reductase (SmoA), NADH and SQ (extended controls). (d) pH profile for SmoC-catalyzed desulfurization of SQ. (e) Sulfite release assay using Ellman's reagent to quantify *Ro*SmoC activity in presence of FMN, flavin reductase (*Ro*SmoA), NADH and SQ, as well as other controls.

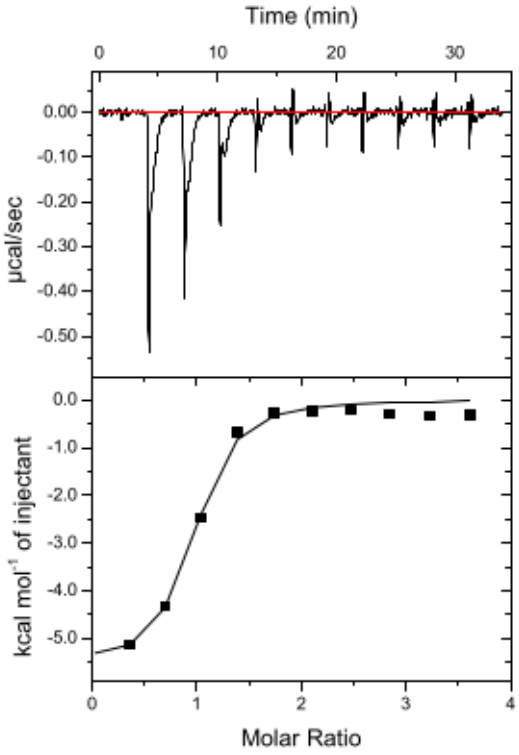
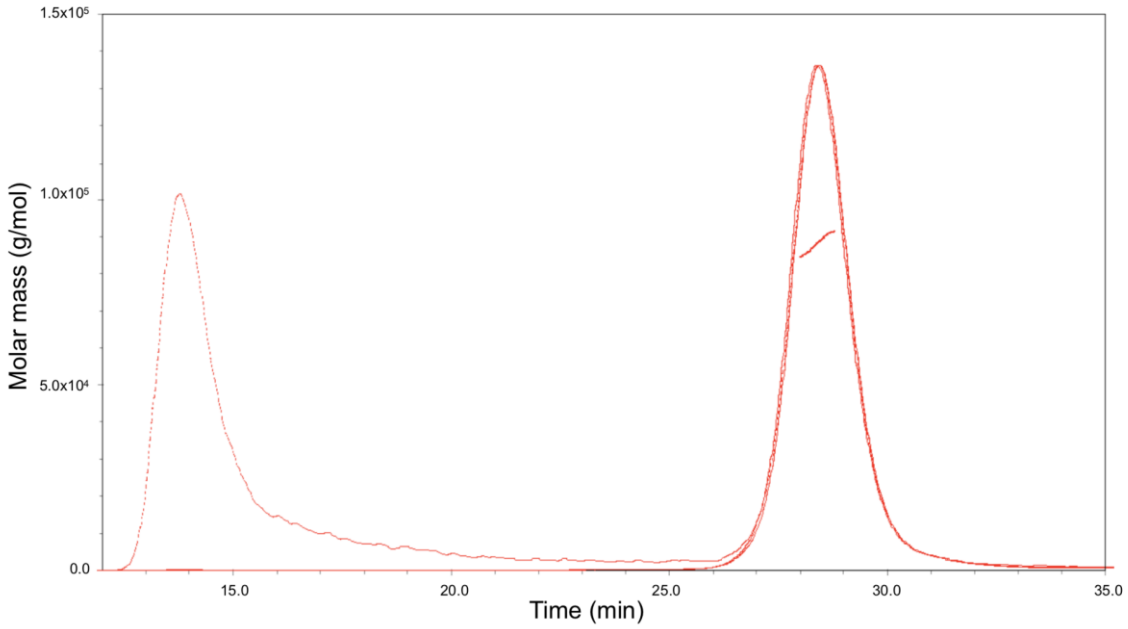


Fig. S13: Independent replicate of data in Figure 3c. Isothermal titration calorimogram of interaction of SmoC with SQ as determined by ITC.



Peak	from (min)	to (min)	Amount (µg)	MW (kDa)
Major Peak (99%)	26.00	28.80	373	88.1

Fig. S14: SEC-MALS for SmoC. Molar mass plots show the light scatter trace as a solid red line, the refractive index trace (concentration measurement) as a dashed red line, and the UV trace as a dotted red line, normalised to the largest peak.

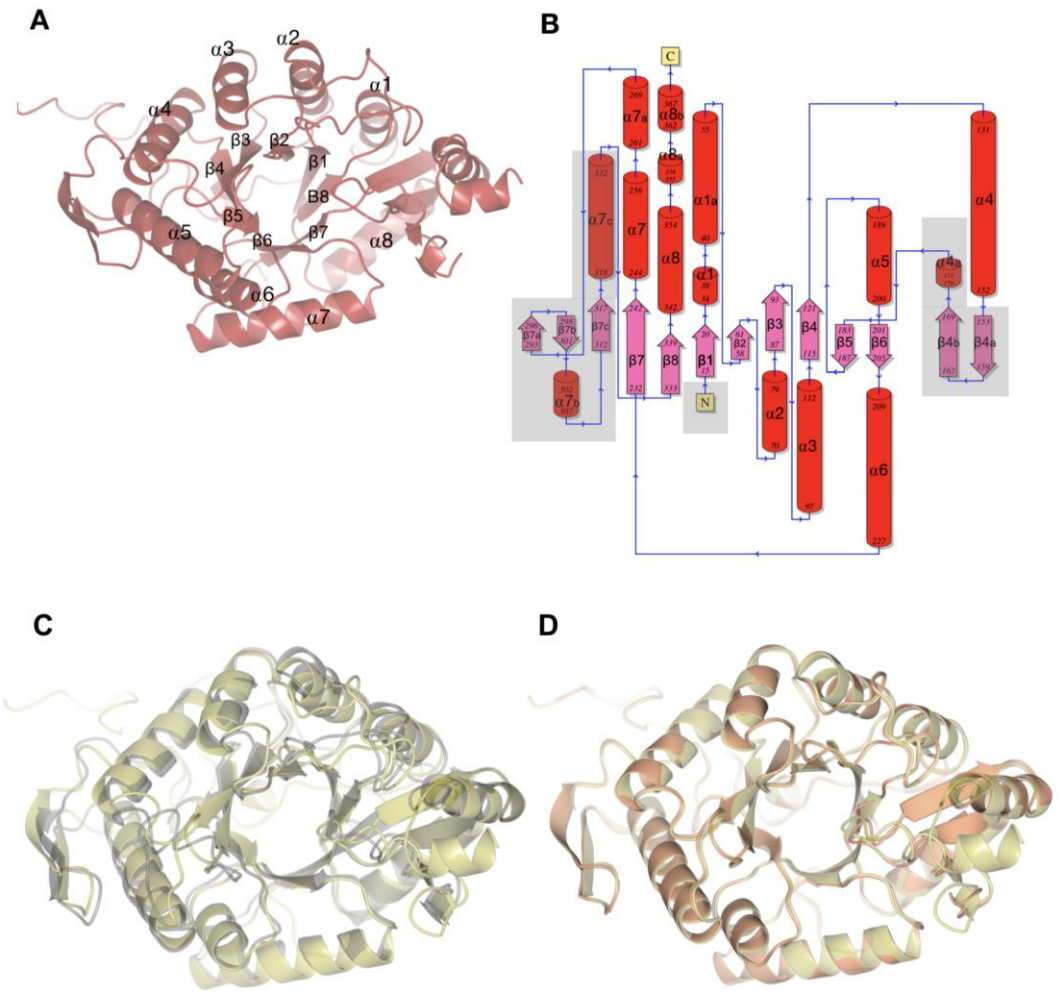


Fig. S15: Ribbon and topology diagrams of *Ro*SmoC. (a) Ribbon diagram of the *Ro*SmoC monomer showing $(\alpha/\beta)_8$ TIM barrel fold. (b) Secondary structure arrangement of *Ro*SmoC depicted in topology diagram showing $(\alpha/\beta)_8$ TIM barrel fold with additional inserts highlighted in grey (generated using the PDBsum server). (c) Overlay of *Ro*SmoC subunit A (in gold) with SsuD (PDB ID: 1M41, in grey): the peptide backbones have an r.m.s.d. of 1.3 Å. (d) Overlay of *Ro*SmoC subunit A (in gold) with SmoC (At3279) (in coral): the peptide backbones have an r.m.s.d. of 0.4 Å.

.

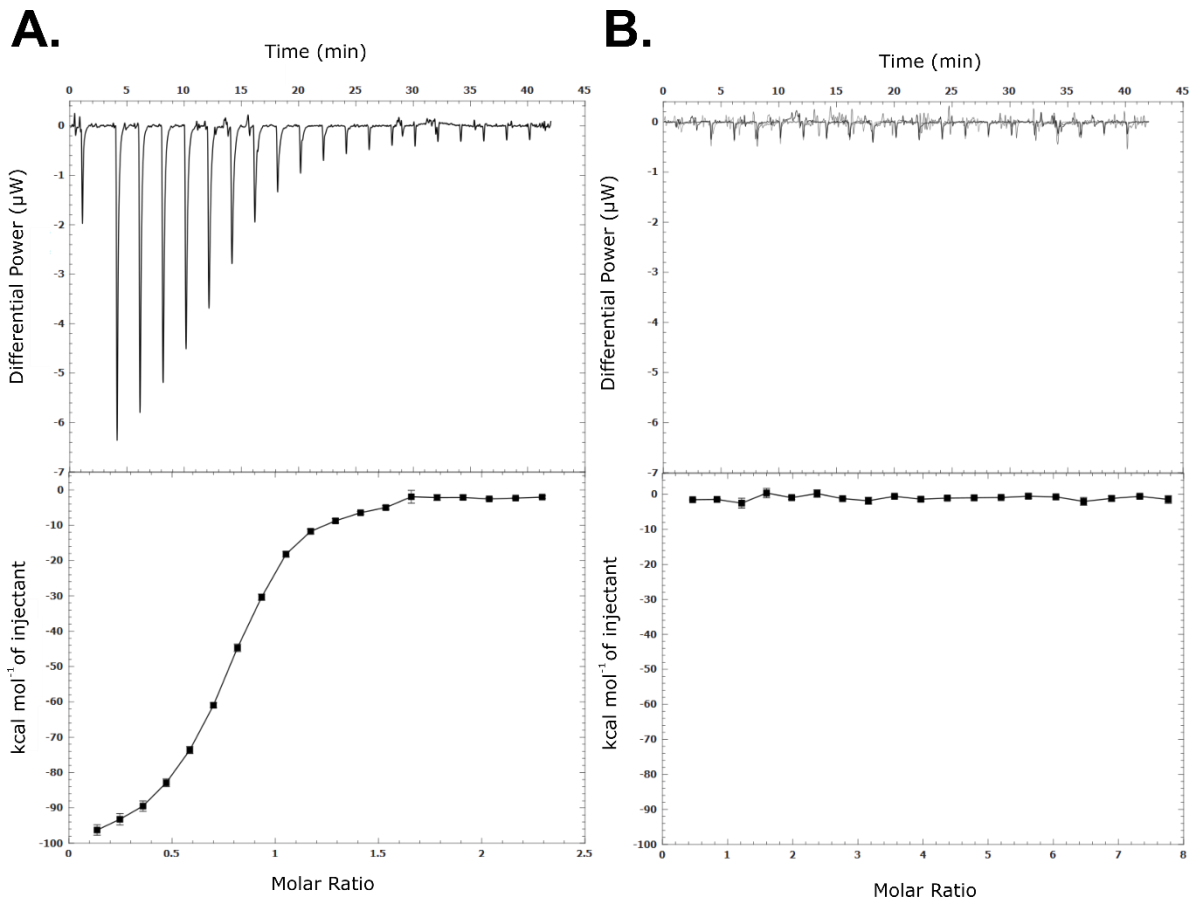


Fig. S16: Isothermal Titration Calorimetry for addition of NAD(P)H to recombinant SmoB. (a) Unprocessed heat and molar ratio plots of SmoB with NADPH. (b) Unprocessed heat and molar ratio plots of SmoB with NADH.

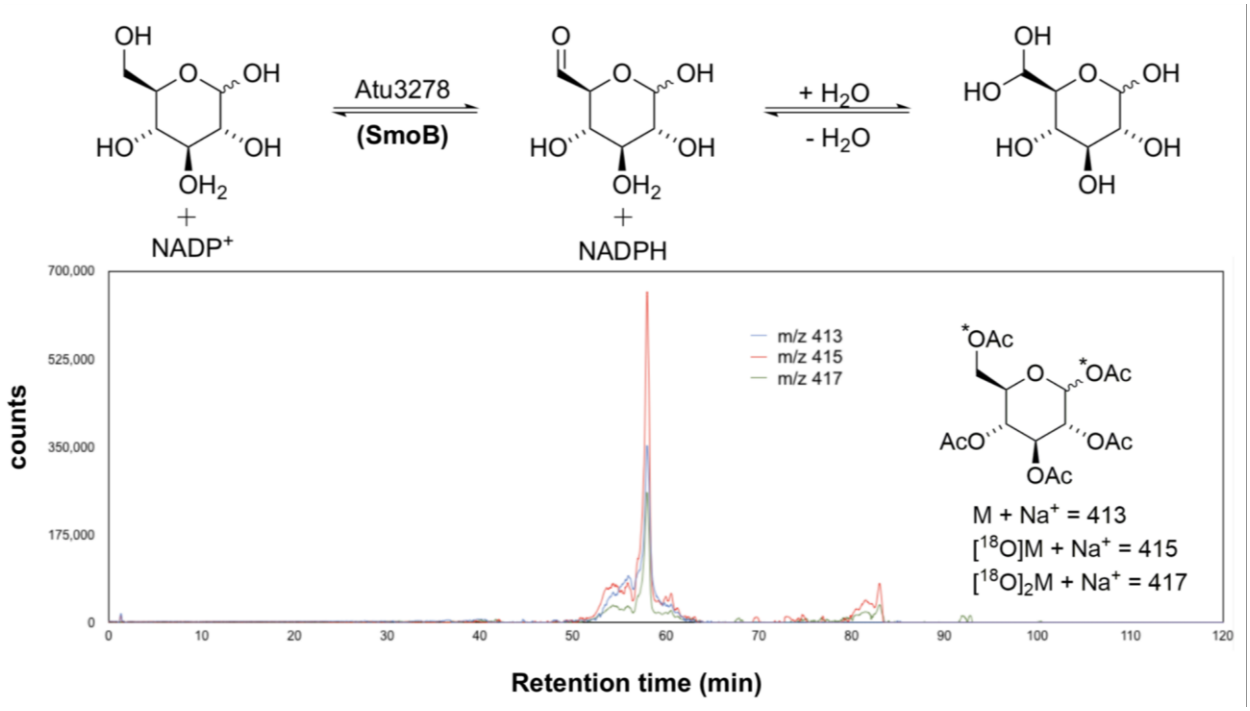


Fig. S17: SmoB (Atu3278) catalyzes formation of $^{18}O_2$-labelled glucose. Extracted ion chromatogram from LC-MS analysis of glucose incubated with SmoB and $NADP^+$ in $H_2{}^{18}O$ followed by peracetylation.

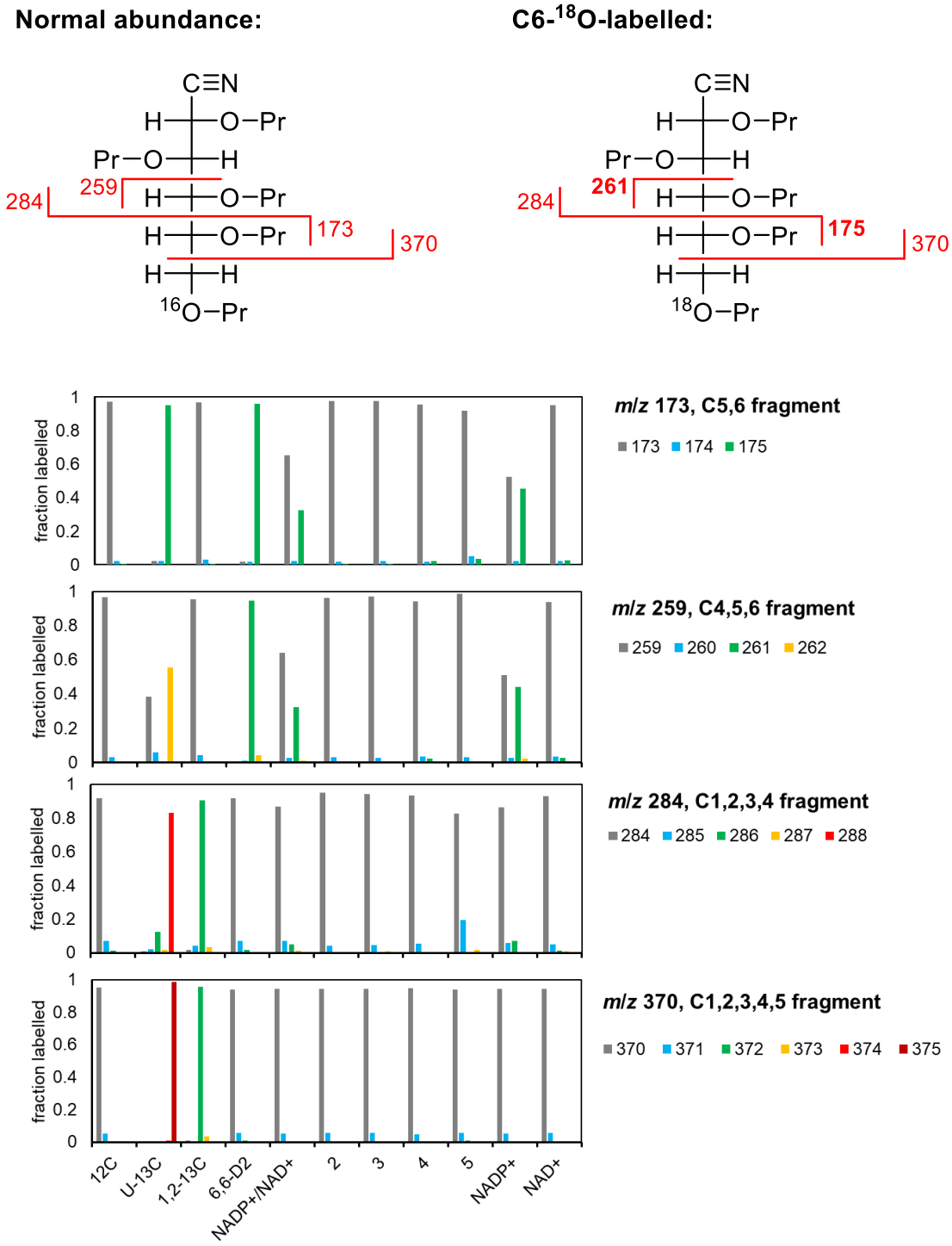


Fig. S18: Identification of ^{18}O-labelled glucose. (Top) Structure of aldonitrile pentapropionate glucose and predicted *m/z* values for key fragments involving ^{18}O incorporation at C6. (**Bottom**) Fraction of labelled fragments derived from incubation with SmoB and various control experiments, corrected for isotope natural abundance by DExSI analysis for 2.5 nmol of analysed glucose. Standards include: $^{12}C_6$-glucose, $^{13}C_6$-glucose, 1,2-$^{13}C_2$-glucose, 6,6-$^{2}H_2$-glucose. Only reactions containing enzyme, substrate and NADP$^+$ gave M+4 product and labelling consistent with introduction of ^{18}O at C6.

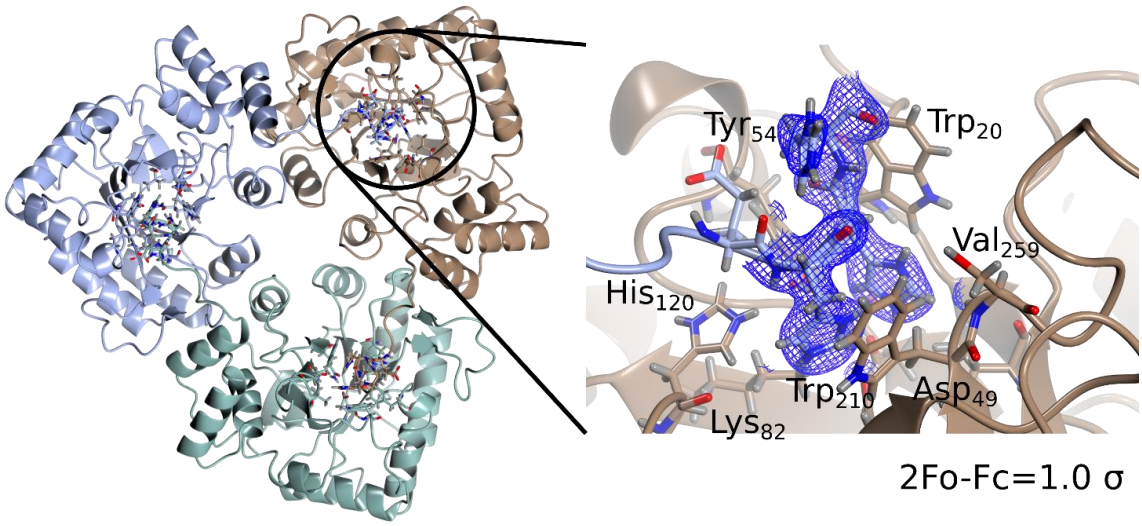
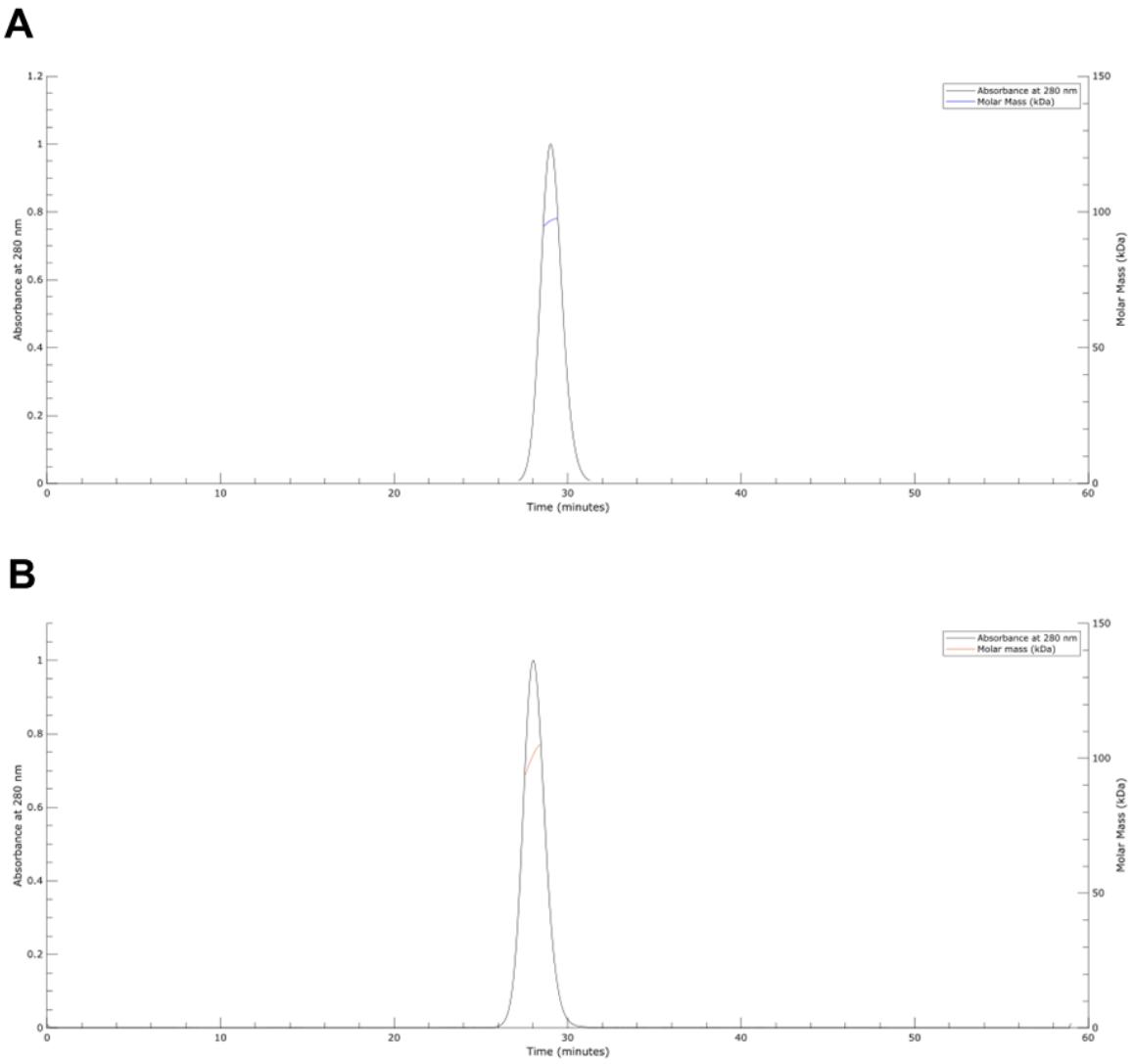


Fig. S19: 3D structures of SmoB-pET29a showing undesirable His_6-tag interactions. SmoB (pET29 construct) showing C-terminal His_6-tag (blue) from the adjoining subunit bound at the cofactor site within the physiologically-relevant trimer.



Peak	from (min)	to (min)	Amount (µg)	MW (kDa)
Atu3278 pET29a(+)	28.6	31.0	523.5	96.54
Atu3278 YSBLIC3C	26.2	30.0	301	100.3

Fig. S20: SEC-MALS plot of SmoB. *A,* pET29a(+) construct. *B,* YSBLIC3C construct (Molecular weight: 34 kDa). SEC-MALS plot reveals the oligomeric state of SmoB in solution. UV-trace and an average molecular weight trace (red), calculated from the refractive index and light scattering signal giving mass estimation of 96.5 and 100 kDa, respectively, which corresponds to a trimer.

Supplementary Tables

Table S1: Recombinant protein sequences.

Atu3277 (SmoA)	MTVVEAIKMPNEHVFVPGGENSRSFRNALGAFTTGVTVVTATTPEGPIGMTVNSFASVSLD PPLVLWSPAKSSSRHPAFSEATHFAIHVLSADQDVLSARFTRNGRAFDDLDWEINDEGVPV IPGTLARFECRRAAAHDAGDHTIIVGEVLRAAHRDGDPLCFSGGAFGRFSRQLEHHHHHH
Atu3282 (SmoF)	MDAELKIFVSSQHQPDIWRKALDQYEAKTPGVKVVIETGGNTSEMQAQYLNTVMSAKDSSL DVLMLDVIRPAQFATAGWTSDFSGKDLSAYLPTYAEANTVNGKIVALPAFADSMFLYYRKD LLDKYGIKPPTTWDELKEASKKVMEGEKNPELQGLSFQGKAIEGAVCTFLLPYWSEGKSLV ENGKLNFDNKAAVDSLKLWKSFVDDGISKKNISEVATDDTRKEFQAGKVLFAVNWSYAWTH FQGKESQVNDKVGVARLPAVKGGEQTTCLGGWEFGVSAYSKQQDEAKKLVEYLSSQDVSKF MAINAALLPTYAALYKDADVTKTIPWFADALPVVETAKARPVTPRYNEVSETIRTTVNGVL AGVMTPEDGAKQMESRLRRVLRLEHHHHHH
Atu3285 D455N E370A E371A (SmoI)	MHFETTKDGFTIAIGNRIILSHSPDKPAFFAGFGEERMDMYRGNFDIEDYVIERTALRHAE VSGDSVTLSSAPGQAPRLRLTLDGNAIRLTALDETINRLWLRVVAETDEHVWGGGEQMSYF DMRGRRFPLWTSEPGVGRDKTTEITFKSDVSGKAGGDYYNTNYPQPTWLSSRKYALHVETS AYSVFDFRNGDFHEIEIWAVPEKIEFFAGDSFADIVSALSLHFGRQPELPDWVYNGAIIGL KDGVNSFARLEKIRAAGTKVSGLWCEDWVGLRQTSFGARLFWDWQANDTRYPHLRQKIAEL ADQGIRFLGYVNPYLCVDGPLFPVAESAGYFATDVDGKTALVDFGEFDCGVVDFTNPAAAD WFAAAIIGKNMLDFGLSGWMADFGEYLPIDIKLSNGVDAKLMHNAWPTLWAEVNAKGVESR GKTGEALFFMRAGFTGVQAHCPLIWGGNQSVDFSRHDGLVTVICGALSSGLMGNAYHHSDI GGYTSLFGNVRTAELIMRWTEMAAFTPVMRTHEGNRPRDNLQIDQDETVLAHFARMTAIYV ALAPYLKSLSAEAAKTGLPVQRPLFLHYENEPQTYAVQDCYLYGADMLVAPVWKAGETQRS LYLPGHGEWVHLWSGKRHAGGRDITVETPLGEPAVFYRADSSHHRLFEQLRTIGLEHHHHH H N.B. active site D455N and double surface E370A/E371A mutations are highlighted in red
Atu3278 (SmoB) in pET29	MQRIALSDKLELSRIVYGMWRIGDDADTSPAHVQAKIEACLAQGITTMDQADIYGGYTAEA ILGGGLKAAPGLRDKIEIVTKCGIVAPAGRHSSARVKHYDTTAGHINVSVEASLRDMGTDH VDLLLIHRPDPLIDAEETGKALDALVASGKVKAVGVSNFRPWDFSLLQSAMSNRLVTNQIE MSLLATDTFTNGDLAYLQEKRVSPMAWSPLGGGSLFSGAYGGTMAALQRIGKEQGVDATAV AIAWLLRHPAKIVPVLGTNNLERIRTAADALRVTMDRQTWFELYTLAIGKEVALEHHHHHH
Atu3278 (SmoB) in YSBLIC3C	MGSSHHHHHHSSGLEVLFQGPAMQRIALSDKLELSRIVYGMWRIGDDADTSPAHVQAKIEA CLAQGITTMDQADIYGGYTAEAILGGGLKAAPGLRDKIEIVTKCGIVAPAGRHSSARVKHY DTTAGHINVSVEASLRDMGTDHVDLLLIHRPDPLIDAEETGKALDALVASGKVKAVGVSNF RPWDFSLLQSAMSNRLVTNQIEMSLLATDTFTNGDLAYLQEKRVSPMAWSPLGGGSLFSGA YGGTMAALQRIGKEQGVDATAVAIAWLLRHPAKIVPVLGTNNLERIRTAADALRVTMDRQT WFELYTLAIGKEVA
Atu3279 (SmoC)	MTVVPVTSADLDAAEVSWFSALCSDDYAYLGVPDGSLRSSFEHCSDIVKKAEELGFRNILC PSSYQVGQDTLSFVAGCAPISDRINFLAAIRCGEMQPIMLARTVATLDHMLKGRLTLNVIS SDFPGEVADSAFRYKRSHEVVEILRQAWTRDTIDHDGEIYQFKGVSTEPARPYQLNGGPLL YFGGYSPDALELCGAQCDVYLMWPETKDQLADRMRAAHERAAAHGRTLDYGLRVHMVVRDT EQEAREYADHLVSKLDDEYGQLIRNRAHDSGSLGVSHQARARELADKFGYVEPNLWTGIGR ARSGCGAALVGSTDQVLSALEEYQKMGIRAFILSGYPHLDEAEHFGTKVLPQMKTCSLPHA YGRVPSETPATPLGNGERHLEHHHHHH
***Ro*SmoA**	MTISADITHGLNEQVFIPDASTARHYRNALGTFTTGVAVVTARTPDGPIGMTVNSFTSVSL DPPLVLWSPAKSSSRHRAFTAASYFVIHVLSAEQDRLSARFTRNGAGFEGLDWIENMEGVP VIPGTLARFECERSDLHDAGDHTLILGRVLRAAHREGDPLCFSRGTFGRFQSH
***Ro*SmoC**	MTVVPITSPDLDAAEVSWFAALCSDDYAYLGVPDDALKSSFEHCSEIVTRAETLGFRNILC PSSYQVGQDTLSFVAACSQITERINLLAAIRCGEMQPIMLARTVATLDHMLKGRLTLNVIS SDFPGEVADSAFRYRRSHEVVQILRQAWTRDTIDHEGEVYNFKGVTTEPARPYQQNGGPLL YFGGYSPDALELCGAQCDVYLMWPEPKEQIAERMKAVHARAEAHGRTLDYGLRVHMIVRDT EKEARDYAEHLVSKLDDEYGRLIRSRAHDSTSLGVSHQARTRELADKFGYVERHLWTGIGR ARSGCGAALVGSTDQVLSEIEAYKKMGVRAFIFSGYPHLDEAEHFGKKVLPQLKTCSLPHI YGRVPADTPATPLGAGRRH

Table S2: Oligonucleotides used in this study.

ssDNA		
Atu3277 (*smoA*)	TTGGACATATGACAGTTGTAGAGGCAATCAAGATGC	sense
	AGGATCTCGAGTTGCCTTGAAAAACGACCGAAG	antisense
Atu3278 (*smoB*)	TTGGACATATGCAACGTATCGCTCTTTCTGAC	sense
	AGGATCTCGAGCGCCACCTCTTTTCCAATCG	antisense
At3278-LIC3C	TTCCAGGGACCAGCAATGCAACGTATCGCTCTTTCTG	sense
	CATGCTAGCCATATGTTACGCCACCTCTTTTCCAATC	antisense
Atu3279 (*smoC*)	TTGGACATATGACCGTCGTACCCGTTACATCTG	sense
	AGGATCTCGAGGTGGCGTTCCCCGTTGCCG	antisense
Atu3282 (*smoF*)	TTGGACATATGGACGCCGAACTGAAAATCTTCG	sense
	AGGATCTCGAGTCTCAGAACGCGCCGCAGACG	antisense
dsDNA		
RoSmoA	TGTAAAACGACGGCCAGTCATATGACCATCTCCGCAGATATCACCCACGGCCTGAACG AACAGGTGTTTATTCCGGATGCCAGCACCGCTCGTCATTATCGCAACGCGCTGGGCAC GTTTACCACCGGCGTCGCGGTGGTGACCGCCCGCACGCCGGACGGCCCGATTGGCATG ACGGTGAACAGCTTTACCAGCGTGAGCCTCGATCCGCCGCTGGTGCTGTGGAGCCCGG CGAAAAGCAGTAGCCGCCATCGCGCGTTTACGGCGGCGAGCTACTTTGTGATTCATGT GCTGAGCGCGGAACAGGATCGTCTGAGCGCGCGTTTCACCCGTAACGGCGCGGGCTTC GAAGGCCTGGACTGGATTGAAAATATGGAAGGCGTGCCGGTGATCCCGGGCACGCTGG CGCGTTTTGAATGCGAACGCAGCGATCTGCATGATGCCGGTGATCATACCCTGATTCT GGGCCGCGTGCTGCGCGCGGCGCATCGTGAAGGTGATCCGCTGTGCTTTAGCCGCGGC ACCTTTGGCCGCTTTCAGAGCCATCTCGAGCCGCTGAGCAATAACTAGC	
RoSmoC	TGTAAAACGACGGCCAGTCATATGACCGTGGTACCGATCACCTCCCCGGATCTGGATG CTGCTGAGGTCAGTTGGTTTGCGGCGCTCTGTTCAGATGATTATGCGTATCTGGGTGT CCCAGATGATGCCCTGAAATCATCGTTTGAACATTGCTCGGAGATTGTGACCCGTGCA GAAACGTTAGGTTTTCGGAATATTCTATGTCCCAGTTCGTATCAGGTTGGCCAGGATA CCCTGTCTTTCGTAGCGGCATGCAGCCAGATTACGGAACGTATTAACCTTTTGGCGGC GATCCGTTGCGGTGAAATGCAGCCTATTATGCTGGCCCGTACCGTAGCGACACTGGAT CACATGTTGAAAGGCCGCCTGACTTTGAACGTGATTAGCAGCGATTTTCCGGGTGAAG TGGCGGATAGCGCGTTCCGCTATCGCCGTAGCCACGAAGTGGTGCAGATTCTGCGCCA GGCGTGGACCCGTGATACGATTGATCATGAAGGCGAGGTATATAATTTCAAAGGTGTG ACCACCGAACCGGCACGTCCGTACCAGCAGAACGGCGGCCCGCTGCTGTATTTTGGCG GCTATTCACCAGACGCGCTGGAACTGTGCGGGGCACAGTGTGATGTGTACCTGATGTG GCCTGAACCGAAAGAACAGATTGCGGAACGCATGAAAGCCGTCCACGCGCGTGCGGAA GCCCACGGCCGCACGCTGGATTACGGCCTGCGTGTTCACATGATCGTTCGCGATACCG AAAAAGAAGCGCGTGATTATGCGGAACATCTGGTCAGCAAACTGGACGATGAATATGG TAGATTAATTCGCAGTCGTGCCCATGATTCCACGAGCCTGGGCGTGAGCCACCAGGCA CGCACCCGCGAACTGGCCGATAAATTCGGCTACGTTGAACGTCACCTGTGGACCGGCA TTGGCCGCGCCCGCAGCGGCTGTGGGGCGGCTCTGGTCGGCAGCACCGACCAGGTTCT GTCCGAAATCGAAGCCTATAAAAAAATGGGCGTTCGTGCGTTTATTTTTAGCGGTTAT CCGCACTTAGATGAAGCAGAACATTTTGGTAAAAAAGTTTTACCGCAGTTAAAAACCT GTAGCCTGCCGCATATTTATGGTCGTGTGCCGGCTGATACCCCGGCGACCCCGTTAGG TGCGGGTCGTCGTCATCTCGAGCCGCTGAGCAATAACTAGC	

Table S3: Parameters used for ITC and the thermodynamic terms determined for ligand binding to SmoB (Atu3277), SmoC (Atu3279) and SmoF (Atu3282). Errors are provided as standard deviations (n=3).

Protein (30 µM)	ligand	Conc. (µM)	T (°C)	n	K_d (µM)	ΔH (kcal mol^{-1})	ΔS (cal · mol^{-1} · K^{-1})	-TΔS (kcal · mol^{-1})	ΔG (kcal · mol^{-1})
SmoC	SQ	600	25	0.81 ± 0.06	3.27 ± 1.20	-4.17 ± 0.26	10.06 ± 1.53	-3.00 ± 0.46	-7.17 ± 0.52
SmoF	SQGro	200	25	0.85 ± 0.04	0.29 ± 0.17	-11.21 ± 0.40	-7.44 ± 2.26	2.22 ± 0.67	-9.00 ± 0.78
SmoB	NADH	1000	25	-	-	-	-	-	-
SmoB	NADPH	0.35	25	0.76 ± 0.01	1.22 ± 0.04	-24.25 ± 0.74	-54.34 ± 2.42	16.2 ± 0.72	-8.07 ± 0.04

Table S4: Data collection and refinement statistics for crystal structures of SmoF (Atu3282) and SmoI (Atu3285). Numbers in brackets refer to data for highest resolution shells.

	Atu3285-D455N•SQGro	**Atu3282-apo**	**Atu3282•SQGro**
Data collection			
Space group	P 1 21 1	*P 1*	P 1 21 1
Cell dimensions			
a, *b*, *c* (Å)	97.53, 168.40, 100.69	54.21, 78.02, 83.32	53.80 137.80 54.07
α, β, ϒ (°)	90.0, 116.53, 90.0	109.0,106.9,104.7	90.0, 118.70, 90.0
Resolution (Å)	61.51-2.15 (2.19-2.15)	1.30	47.19-1.70 (1.73-1.70)
R_{merge}	0.077 (0.538)	0.040(0.515)	0.067 (0.709)
R_{pim}	0.077 (0.534)	0.040(0.515)	0.064 (0.674)
I / σI	8.1 (1.6)	8.9(1.1)	9.8 (1.4)
Completeness (%)	97.4 (98.9)	99.1(67.5)	98.3 (96.8)
Redundancy	2.9 (2.9)	2.1(1.6)	3.4 (3.4)
Refinement			
Resolution (Å)	2.15	1.30	47.19-1.70
No. unique reflections	145522	270846	74,272
R_{work} / R_{free}	0.1946/0.2235	0.18/0.20	0.1494/0.1832
No. atoms			
Protein	20405	8977	6,096
Ligand/ion	80	8	40 (SQG), 40 (EDG)
Water	439	1470	573
B-factors (Å^2)			
Protein	31.38	16.6	24.4
Ligand/ion	22.24	30.7	14.8 (SQGro); 39.6 (EDG)
Water	23.37	27.6	37.6
R.m.s. deviations			
Bond lengths (Å)	0.0143	0.0135	0.006
Bond angles (°)	1.81	1.911	0.776
Ramachandran Plot Residues			
In most favourable regions (%)	96.56	98.61	99.35
In allowed regions (%)	3.1	1.29	0.65
Outliers	0.34	0	0
PDB code	**7OFX**	**7NBZ**	**7OFY**

Table S5: Data collection and refinement statistics for crystal structures of SQ monooxygenases SmoC (Atu3279) from *Agrobacterium tumefaciens* and *Ro*SmoC from *Rhizobium oryzae*. Numbers in brackets refer to data for highest resolution shells.

	***Ro*SmoC**	**Atu3279**
Data collection		
Space group	P $2_12_12_1$	P62 2 2
Cell dimensions		
a, *b*, *c* (Å)	71.16, 100.40, 104.65	203.76 203.76 110.73
α, β, ɣ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	72.45 - 1.9 (1.94-1.90)	48.94-3.40 (3.67-3.40)
R_{merge}	0.165 (1.739)	0.117 (0.344)
R_{pim}	0.067 (0.709)	0.042 (0.173)
I / σI	12.7 (3.6)	16 (5.1)
Completeness (%)	100.0 (100.0)	99.9 (99.9)
Redundancy	13.4 (13.8)	15.8 (8.9)
Refinement		
Resolution (Å)	72.01-1.90	48.00-3.40
No. unique reflections	59830 (3774)	19137 (3864)
R_{work} / R_{free}	0.1673/0.1961	0.2179/0.2615
No. atoms		
Protein	5558	5392
Ligand/ion	-	-
Water	491	-
B-factors (Å^2)		
Protein	25.65	85.03
Ligand/ion	-	-
Water	31.05	-
R.m.s. deviations		
Bond lengths (Å)	0.01473	0.0105
Bond angles (°)	1.80	2.01
Ramachandran Plot Residues		
In most favourable regions (%)	95.96	91.92
In allowed regions (%)	4.04	7.40
Outliers	0.0	0.68
PDB code	**7OH2**	**7OLF**

Table S6: Data collection and refinement statistics for 6-oxo-glucose reductase SmoB (Atu3278) structures. Numbers in brackets refer to data for highest resolution shells.

	Atu3278-apo (YSBLIC3C)	**Atu3278• NADPH**	**Atu3278•NADPH•Glc**	**Atu3278-apo (pET29a)**
Data collection				
Space group	C 2 2 2_1	P 6_3	P 6_3	P 2_1 3
Cell dimensions				
a, *b*, *c* (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, ɣ (°)	90,90,90	90,90,120	90,90,120	90,90,90
Resolution (Å)	1.77	1.27	1.50	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 / σI	11.2(1.2)	9.6(0.6)	14.7(1.4)	14.1(1.0)
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
B-factors (Å^2)				
Protein	29.06	17.71	22.22	34.32
Ligand/ion	-	15.67	24.01	54.13
Water	32.23	28.9	31.67	37.34
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				
In most favourable regions (%)	97.97	98.96	99.31	97.97
In allowed regions (%)	2.03	1.04	0.69	1.36
Outliers	0	0	0	2
PDB code	**7BBZ**	**7BC0**	**7BC1**	**7BBY**

Table S7: Glucose fragmentation data (for aldonitrile pentapropionate glucose).

sugar	retention time	ion m/z	relevant carbons
glucose	17.52	259	4,5,6
glucose	17.52	284	1,2,3,4
glucose	17.52	173	5,6
glucose	17.52	370	1,2,3,4,5