Erschienen in: Nature ; 507 (2014), 7490. - S. 114-117

Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle

Karin Denger[1], Michael Weiss[2], Ann-Katrin Felux[2], Alexander Schneider[3], Christoph Mayer[3], Dieter Spiteller[1], Thomas Huhn[4], Alasdair M. Cook[1] & David Schleheck[1]

Sulphoquinovose (SQ, 6 deoxy 6 sulphoglucose) has been known for 50 years as the polar headgroup of the plant sulpholipid[1,2] in the photosynthetic membranes of all higher plants, mosses, ferns, algae and most photosynthetic bacteria[3]. It is also found in some non photosynthetic bacteria[4], and SQ is part of the surface layer of some Archaea[5]. The estimated annual production of SQ[4] is 10,000,000,000 tonnes (10 petagrams), thus it comprises a major portion of the organo sulphur in nature, where SQ is degraded by bacteria[6,7]. However, despite evidence for at least three different degradative pathways in bacteria[6 8], no enzymic reaction or gene in any pathway has been defined, although a sulphoglycolytic pathway has been proposed[7]. Here we show that *Escherichia coli* K 12, the most widely studied prokaryotic model organism, performs sulphoglycolysis, in addi tion to standard glycolysis. SQ is catabolised through four newly discovered reactions that we established using purified, heterolo gously expressed enzymes: SQ isomerase, 6 deoxy 6 sulphofructose (SF) kinase, 6 deoxy 6 sulphofructose 1 phosphate (SFP) aldolase, and 3 sulpholactaldehyde (SLA) reductase. The enzymes are encoded in a ten gene cluster, which probably also encodes regulation, trans port and degradation of the whole sulpholipid; the gene cluster is present in almost all (>91%) available *E. coli* genomes, and is wide spread in Enterobacteriaceae. The pathway yields dihydroxyacetone phosphate (DHAP), which powers energy conservation and growth of *E. coli*, and the sulphonate product 2,3 dihydroxypropane 1 sulphonate (DHPS), which is excreted. DHPS is mineralized by other bacteria, thus closing the sulphur cycle within a bacterial community.

Recent work showed that environmental isolates of *Klebsiella* spp. (Enterobacteriaceae) convert SQ quantitatively to DHPS[7,8], and we pro posed that utilization of SQ might be a property of Enterobacteriaceae. We found that four genome sequenced *E. coli* K 12 substrains (BW25113, DH1, MG1655 and W3100), after subculturing, grew with SQ within 1 to 3 days. We chose to work (largely) with the fastest growing sub strain, MG1655. The organism used SQ as a sole source of carbon and energy with a molar growth yield of 3 g of protein per mol of SQ carbon, whereas glucose gave approximately 6 g of protein per mol of carbon; the latter value represented mass balance of glucose carbon as biomass and CO_2 (ref. 9). However, approximately 1 mol of DHPS per mol of SQ was released into the growth medium (Fig. 1a), as observed with *Klebsiella oxytoca*[8]. Thus, there was complete mass balance for carbon and for sulphur from SQ. The growth rate with SQ was 0.13 h^{-1} (0.5 h^{-1} with glucose), and the specific degradation rate for SQ *in vivo* was 120 mU per mg of protein (1 mU = 1 nmol min^{-1}). We concluded that SQ is metabolized to a C_3 sulphonate, which is excreted as DHPS, and that the remainder of the molecule is used for growth (Fig. 2a).

The out grown culture was filter sterilized and inoculated with *Cupriavidus pinatubonensis* JMP134, which can utilize DHPS for growth[10], but cannot utilize SQ[8]. *C. pinatubonensis* grew with the DHPS formed from SQ by *E. coli*, and released its sulphonate sulphur quantitatively as sulphate (Fig. 1b) using a pathway described elsewhere[10]. We thus demonstrated mineralization of SQ in a laboratory model system.

Proteins from whole cells of *E. coli* K 12 grown with glucose or SQ were subjected to two dimensional polyacrylamide gel electropho resis (2D PAGE) (Extended Data Fig. 1) and examined by peptide fingerprinting mass spectrometry (PF MS) (Extended Data Table 1). The immediately relevant, apparently SQ inducible proteins (see Extended Data Fig. 1 and Extended Data Table 1) were attributed to b3878 (also known as *yihQ* (b numbers are locus tags); predicted to be an α glucosidase), b3879 (also known as *yihR*; predicted to be an epimerase), b3880 (*yihS*; predicted to be an isomerase), b3881 (*yihT*; predicted to be an aldolase) and b3882 (*yihU*; predicted to be an NAD^+/NADH linked dehydrogenase/reductase). Transcriptional analyses for the gene cluster b3879 b3882, as well as for b3883 (also known as *yihV*; predicted to be a sugar kinase), confirmed a strong inducible transcription during

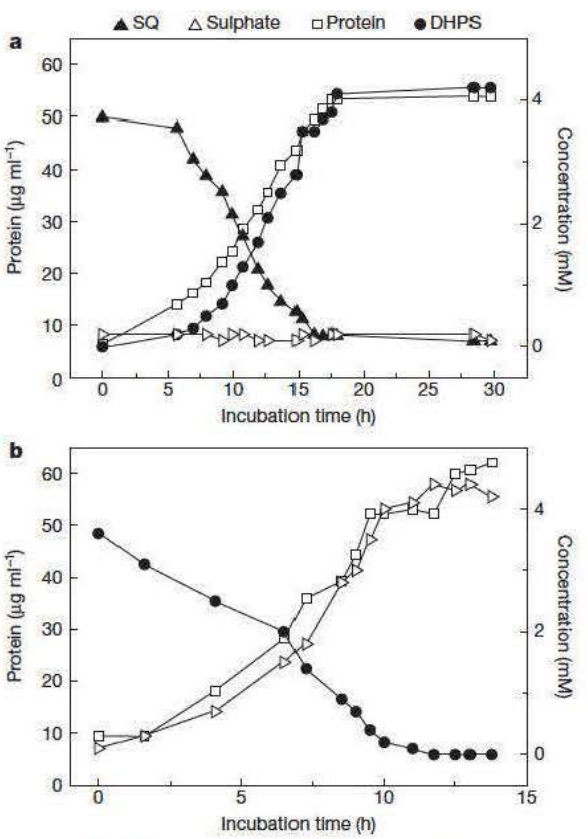


Figure 1 | Complete degradation of sulphoquinovose during growth. a, Growth of *E. coli* K 12 substrain MG1655 with SQ and excretion of 2,3 dihydroxypropane 1 sulphonate (DHPS). b, Growth of *C. pinatubonensis* JMP134 with the DHPS formed from SQ by *E. coli*. Data from representative growth experiments ($n = 3$) are shown. To allow a compact graph, sulphate release and not total sulphate is shown.

[1]Department of Biology, University of Konstanz, D-78457 Konstanz, Germany. [2]Konstanz Research School Chemical Biology, University of Konstanz, D-78457 Konstanz, Germany. [3]Interfaculty Institute of Microbiology and Infection Medicine, University of Tubingen, D-72076 Tubingen, Germany. [4]Department of Chemistry, University of Konstanz, D-78457 Konstanz, Germany.

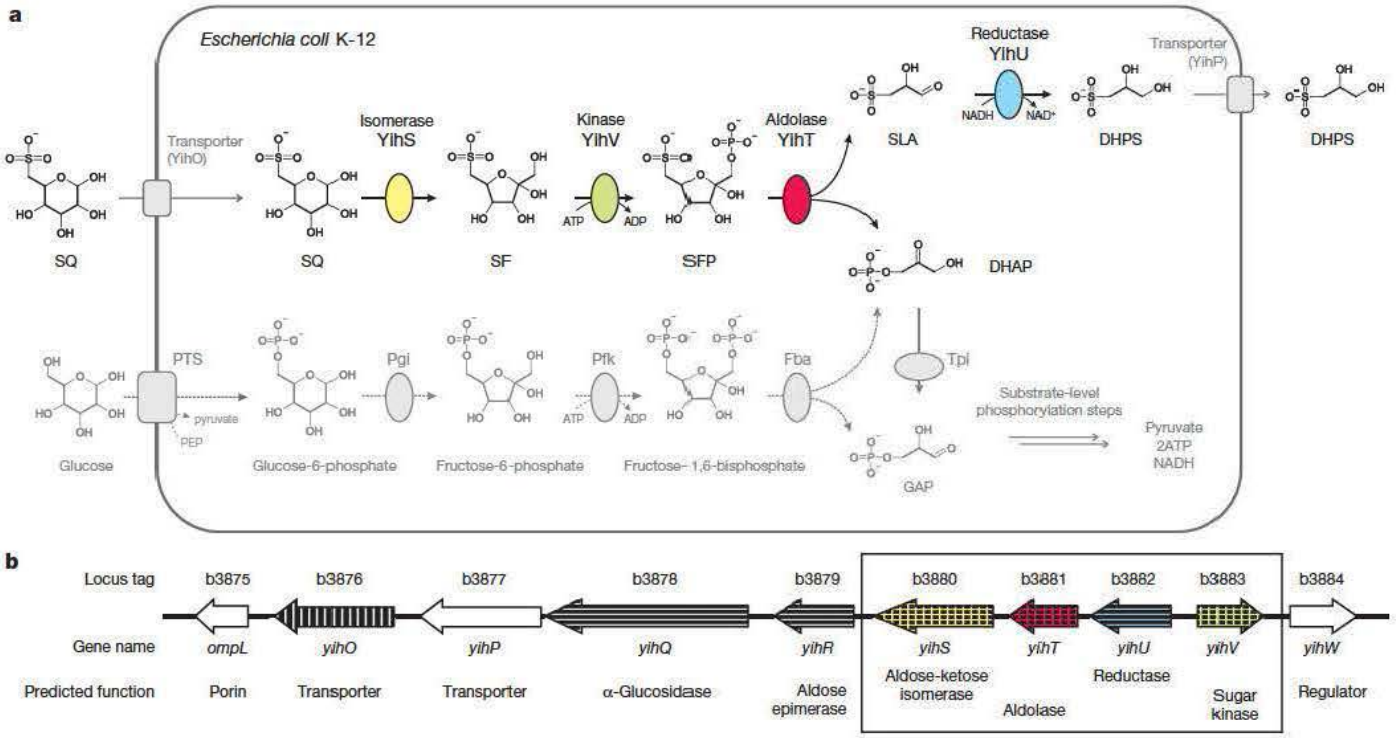


Figure 2 | The four core enzyme reactions of sulphoglycolysis, with transport, and the corresponding genes in a ten gene cluster in *E. coli* K 12. a, SQ is metabolized by four enzymes (shown in colour) to a C_3 sulphonate, DHPS, which is excreted, and the remainder of the molecule is used for growth. For comparison, the analogous enzyme reactions for the catabolism of (unsubstituted) glucose through the glycolytic pathway in *E. coli* are also shown (dashed arrows). Fba, fructose bisphosphate aldolase; GAP, glyceraldehyde 3 phosphate; Pfk, phosphofructokinase; Pgi, phosphoglucose isomerase; PTS, phosphotransferase system permease; Tpi, triose phosphate isomerase. **b**, The EcoGene *E. coli* website (http://www.EcoGene.org) uses the abbreviation *yih* for most of these genes; we have retained this nomenclature. Vertical stripes, genes confirmed as being essential for growth with SQ by mutational analysis; horizontal stripes, genes confirmed as being inducible for growth with SQ by proteomic and/or transcriptional analyses; box framed genes, genes encoding the four core enzymes of the pathway (shown in **a**) that were subject of heterologous expression, purification and functional characterization.

growth with SQ, but not during growth with glucose (Extended Data Fig. 2). Furthermore, single gene knockouts (in substrain BW25113 (ref. 11)) in genes b3876 (also known as *yihO*; predicted to be a major facilitator superfamily (MFS) type transporter), b3880, b3881 and b3883 did not grow with SQ, which confirmed and expanded on the proteomic and transcriptional data (Fig. 2b).

We thus identified a gene cluster in *E. coli* K 12 that contained SQ inducible, essential genes for catabolism of SQ, but we still did not know which pathway was involved. A sulphoglycolytic pathway would involve a hypothetical 3 sulpholactaldehyde (SLA) reductase to yield DHPS in the final reaction (apart from export) (Figs 1a and 2a), whereas a hypothetical SQ dehydrogenase as the first reaction would lead into hypothetical Entner Doudoroff type or pentose phosphate type pathways, or another novel pathway. An SLA reductase was detected (assayed as DHPS oxidation) in cell free extracts of SQ grown substrain MG1655 at a specific activity of 420 mU per mg of protein, which exceeds the specific degradation rate for SQ *in vivo* and, thus, was sufficient to explain growth. This enzyme activity was not detected in extracts of glucose grown cells. Thus the enzyme was confirmed to be inducible, and it was specific for NAD^+; $NADP^+$ was not a substrate. Furthermore, SQ did not lead to reduction of NAD^+ or of $NADP^+$ in the extracts of SQ or glucose grown cells, hence, hypothetical SQ dehydrogenase was not detectable. These data led us to predict the sulphoglycolytic pathway depicted in Fig. 2a, including the requirement for sulphonate import and export across the cell membrane[12 14].

The four predicted core enzymes of the pathway (Fig. 2a) were heterologously expressed and purified as His tagged proteins, b3880 (putative isomerase), b3883 (putative sugar kinase), b3881 (putative aldolase) and b3882 (putative reductase) (Extended Data Fig. 3). Protein b3882 was shown to encode SLA reductase. First, we partially purified and identified (PF MS) the wild type enzyme in cell free extracts of substrain MG1655 (see above), and second, we examined the recombinant protein (see below). In both cases we identified that b3882 represents an SLA reductase; the enzyme showed no activity with 4 hydroxybutyrate[15].

The heterologously expressed and purified putative isomerase (b3880) caused about one fourth of the SQ in the reaction mixture to disappear, as observed by high pressure liquid chromatography mass spectrometry (HPLC MS), and a new peak was formed that eluted with shorter retention time, but exhibited the same relative mass (M_r 244 Dalton (Da); observed as a quasi molecular ion in the negative ion mode ([M H]) at a mass to charge ratio (m/z) of 243) (Fig. 3a, b). The new peak was confirmed to represent 6 deoxy 6 sulphofructose (SF), as proposed elsewhere[7], by the HPLC separation pattern (Extended Data Fig. 4), by the matching exact mass of the [M H] ion (Extended Data Fig. 4), and by its MS MS fragmentation pattern (Extended Data Fig. 5). Thus we confirmed that b3880 catalysed the SQ isomerase reaction.

The reaction mixture was augmented with ATP and the putative sugar kinase (b3883). The peaks of SQ and SF partially disappeared and a new peak was formed (Fig. 3c). This new peak was confirmed to represent 6 deoxy 6 sulphofructose 1 phosphate (SFP), as proposed elsewhere[7], by the matching exact mass of the [M H] ion (observed mass, 322.9877 Da; theoretical mass of $C_6H_{12}O_{11}PS$, 322.9843 Da) and by its fragmentation pattern (Extended Data Fig. 6). HPLC confirmed that ATP disappeared and ADP was formed during the reaction and, furthermore, that SFP was converted back to SF when alkaline phosphatase was added to a preparation of SFP (not shown). Thus, with b3883, we expressed an ATP dependent kinase that phosphorylated SF to SFP.

The reaction mixture was augmented with the putative aldolase (b3881). The peak for SFP partially disappeared, and two new peaks were formed

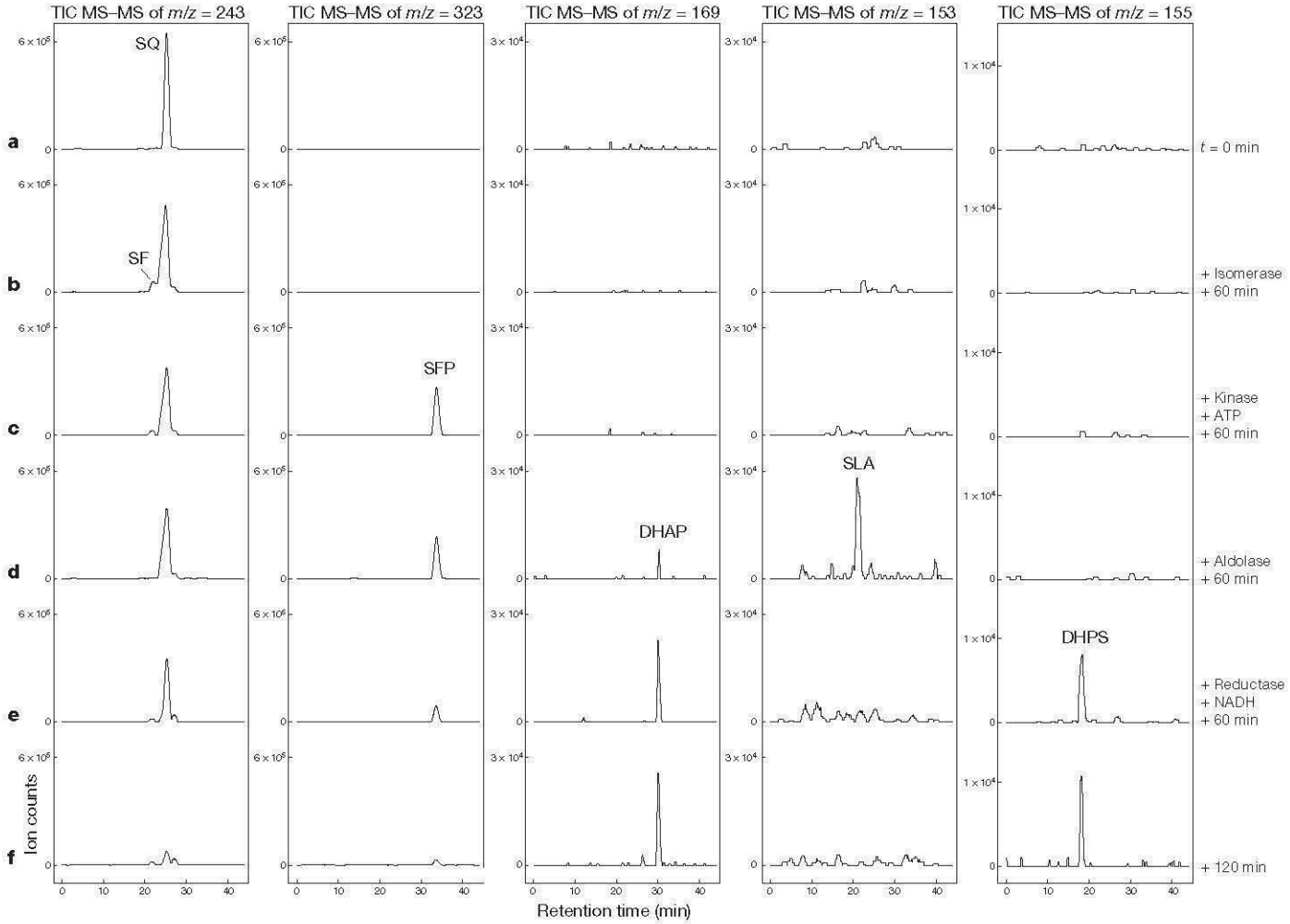


Figure 3 | Illustration of the reactions of the four core enzymes of sulphoglycolysis *in vitro*. The transformation of SQ to SF, SFP, DHAP and SLA, and DHPS, by successive addition of recombinantly expressed pathway enzymes was followed by HPLC ESI MS. **a**, Sample of SQ in reaction buffer (t = 0 min). **b**, Sample after addition of isomerase (b3880) (t = 60 min). **c**, Sample after addition of ATP and kinase (b3883) (t = 120 min). **d**, Sample after addition of aldolase (b3881) (t = 180 min). **e**, Sample after addition of NADH and reductase (b3882) (t = 240 min). **f**, Sample after extended incubation of the four enzyme reaction (t = 360 min). The total ion chromatograms (TICs) recorded in the negative ion mode from the MS MS fragmentation of the quasi molecular ions [M H] of SQ and SF, SFP, DHAP, SLA and DHPS, from a representative experiment (n = 5) are shown. For representative MS MS fragmentation patterns of the [M H] ions of SQ and SF, SFP, and SLA and DHPS, see Extended Data Figs 5, 6 and 7, respectively.

(Fig. 3d). The first new peak was identified to represent DHAP, as proposed elsewhere[7], with an authentic DHAP standard. The second new peak was confirmed to represent SLA, as proposed elsewhere[7], by the matching mass of the [M H] ion (M_r 154 Da; observed as [M H] ion at m/z 153) and by its fragmentation pattern (Extended Data Fig. 7); the same peak was observed when we used recombinant SLA reductase in reverse to oxidize DHPS to SLA (see above). Thus, with b3881, we expressed an aldolase that cleaved SFP into DHAP and SLA. The SFP turnover was incomplete (Fig. 3d); the equilibrium of the corresponding enzyme reaction in glycolysis (fructose 1,6 bisphosphate aldolase) lies far to the left[16], that is, hardly any products are formed.

However, when NADH and the recombinant SLA reductase (b3882) were added, the peak for SFP was further diminished, as was the peak for SQ, and that for the aldolase reaction product DHAP was further increased (Fig. 3e). In addition, the peak for SLA had disappeared, and the peak for the anticipated sulphonate product, DHPS, was formed (Fig. 3e and Extended Data Fig. 7). After an extended incubation of the four enzyme reaction (see Fig. 3e, f), the peaks for SQ and SFP had almost completely disappeared, and the peaks for DHAP and DHPS had further increased.

Together, the results show that the SQ pathway in *E. coli* K 12 (Fig. 2a) does not involve a desulphonation reaction and that no substrate level phosphorylation of the sulphonated C_3 intermediate occurs, which has been used previously[7] as a default hypothesis. Furthermore, we deduce that there are ten genes in the gene cluster (Fig. 2b). The core pathway comprises a SQ transporter (for example, b3876, YihO), SQ isomerase (b3880, YihS), SF kinase (b3883, YihV), SFP aldolase (b3881, YihT), SLA reductase (b3882, YihU) and a DHPS exporter (for example, b3877, YihP), which could be under the putative control of repressor b3884 (YihW). We propose a sulpholipid porin (b3875, OmpL), a sulpholipid α glucosidase (b3878, YihQ), and an epimerase (b3879, YihR) to funnel other SQ derivatives into the pathway; for example, the whole sulpholipid (see Extended Data Fig. 8).

The gene cluster is found in at least 1,009 (>91%) of the 1,110 genome sequences of commensal *E. coli*, as well as pathogenic *E. coli* (for example, EHEC) strains, that were available in November 2013 (finished and draft genome sequences) in the Integrated Microbial Genomes (IMG) and Human Microbiome Project (HMP) databases (that is, gene clusters with syntenic *yihTUVW* and collinear homologues of *yihSRQPO* and *ompL* in variable order). Hence, the gene cluster is a feature of the core genome of *E. coli* species. It can also be found in a wide range of other Enterobacteriaceae (for example, *Chronobacter sakazakii* ATCC BAA 894, *Klebsiella oxytoca* 10 5242, *Pantoea ananatis* LMG 20103 and *Salmonella enterica* LT2). We therefore suspect that the pathway has a significant role in bacteria in the alimentary tract of all omnivores and herbivores, that the pathway occurs in bacteria in excrement from

these animals, and that it occurs in plant pathogens, which would explain part of the widespread occurrence of microbial degradation observed[6 8].

SQ is produced in large amounts in nature and thus represents a significant proportion of the organic sulphur cycle[4], and it is degraded in similar amounts by both bacteria[6 8] and algae[17], or it would accu mulate in the environment. We see here that the Enterobacteriaceae use one pathway (Fig. 2a) to initiate degradation of SQ, and that a com munity is required for complete degradation (Fig. 1b)[8]. This covers a variety of habitats, but we know that other pathways exist. A previous paper[7] presented evidence for SQ dehydrogenase, which we also observe in our SQ using strain of *Pseudomonas putida* (A. K.F. unpublished observations). Notably, another group[6] reported complete SQ degra dation, including desulphonation, in a single organism; however, this organism has been lost[18].

In summary, we have established that sulphoglycolysis, which was named but not defined in a previous report[19], converts SQ to DHPS in the most widely studied prokaryotic model organism, *E. coli* K 12, rep resenting many Enterobacteriaceae (Fig. 2a). We have identified a gene cluster in *E. coli* K 12 (Fig. 2b) that encodes the pathway. The core pathway, for SQ, involves four newly discovered enzymes, two newly identified transporters and three newly characterized intermediates (Fig. 2a, b). We know that the pathway is regulated (Extended Data Figs 1 and 2) and we suspect that it includes the degradation of the whole sulpholipid (Extended Data Fig. 8). The pathway represents a substantial part of the biogeochemical sulphur cycle, and the pathway is likely to have a significant role in bacteria in the alimentary tract of all omnivores and herbivores, and in plant pathogens. We and others[6 8] anticipate other degradative pathways for SQ in nature; for example, in bacteria of all marine, freshwater and terrestrial habitats where SQ is produced and degraded. We now provide the tools to elucidate these degradative pathways.

METHODS SUMMARY

SQ and DHPS were synthesized chemically and identified by NMR and mass spectrometry[8,10]. Cultivation, preparation of cell free extracts, enzyme purification, 2D PAGE and PF MS, RNA preparation and polymerase chain reaction with reverse transcription (RT PCR), and expression and purification of His tagged proteins, are described in the online Methods. SQ, SF, SFP, SLA, DHAP and DHPS were separated using hydrophilic interaction liquid chromatography (HILIC)[8] and detected by an evaporative light scattering detector (ELSD)[8] or electrospray ion ization (ESI) time of flight (TOF) MS or ESI iontrap MS (see Methods). The enzyme reaction mixture (Fig. 3) was 3 mM SQ in 50 mM ammonium acetate buffer (pH 8.7), and 8 mM ATP, 0.5 mM $MgCl_2$ and 8 mM NADH supplemented with the corresponding enzymes (each 50 $\mu g\ ml^{-1}$).

1. Benson, A. A. The plant sulfolipid. *Adv. Lipid Res.* **1,** 387 394 (1963).
2. Benning, C. Questions remaining in sulfolipid biosynthesis: a historical perspective. *Photosynth. Res.* **92,** 199 203 (2007).
3. Benning, C. Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 53 75 (1998).
4. Harwood, J. L. & Nicholls, R. G. The plant sulpholipid a major component of the sulphur cycle. *Biochem. Soc. Trans.* **7,** 440 447 (1979).
5. Meyer, B. H. *et al.* Sulfoquinovose synthase an important enzyme in the *N* glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol. Microbiol.* **82,** 1150 1163 (2011).
6. Martelli, H. L. Oxidation of sulphonic compounds by aquatic bacteria isolated from rivers of the Amazon region. *Nature* **216,** 1238 1239 (1967).
7. 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).
8. Denger, K., Huhn, T., Hollemeyer, K., Schleheck, D. & Cook, A. M. Sulfoquinovose degraded by pure cultures of bacteria with release of C_3 organosulfonates: complete degradation in two member communities. *FEMS Microbiol. Lett.* **328,** 39 45 (2012).
9. Cook, A. M. Biodegradation of s triazine xenobiotics. *FEMS Microbiol. Rev.* **46,** 93 116 (1987).
10. Mayer, J. *et al.* 2,3 Dihydroxypropane 1 sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: purification of dihydroxypropanesulfonate 3 dehydrogenase. *Microbiology* **156,** 1556 1564 (2010).
11. Baba, T. *et al.* Construction of *Escherichia coli* K 12 in frame, single gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
12. Graham, D. E., Xu, H. & White, R. H. Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate biosynthesizing enzymes. *J. Biol. Chem.* **277,** 13421 13429 (2002).
13. Mampel, J. *et al.* A novel outer membrane anion channel (porin) as part of the putatively two component transport system for *p* toluenesulfonate in *Comamonas testosteroni* T 2. *Biochem. J.* **383,** 91 99 (2004).
14. Mayer, J. & Cook, A. M. Homotaurine metabolized to 3 sulfopropanoate in *Cupriavidus necator* H16: enzymes and genes in a patchwork pathway. *J. Bacteriol.* **191,** 6052 6058 (2009).
15. Saito, N. *et al.* Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in *Escherichia coli*. *J. Biol. Chem.* **284,** 16442 16451 (2009).
16. Cornish Bowden, A. Thermodynamic aspects of glycolysis. *Biochem. Educ.* **9,** 133 137 (1981).
17. Sugimoto, K., Sato, N. & Tsuzuki, M. Utilization of a chloroplast membrane sulfolipid as a major internal sulfur source for protein synthesis in the early phase of sulfur starvation in *Chlamydomonas reinhardti*. *FEBS Lett.* **581,** 4519 4522 (2007).
18. Cook, A. M. & Denger, K. Dissimilation of the C_2 sulfonates. *Arch. Microbiol.* **179,** 1 6 (2002).
19. Benson, A. A. & Shibuya, I. Sulfocarbohydrate metabolism. *Fed. Proc.* **20,** 79 (1961).

Acknowledgements We thank E. Deuerling for substrain MG1655, J. Klebensberger for substrain BW25113 and its knockout mutants, and K. Leitner for help with growth experiments. The work of M.W. and A. K.F. was supported by the Konstanz Research School Chemical Biology (KoRS CB), the work of C.M. by German Research Foundation (DFG) grants (MA2436/4 and SFB766/A15) and by the Baden Württemberg Stiftung (P BWS Glyko11), and the work of D.Sc. by a DFG grant (SCHL 1936/1 1) and by the University of Konstanz and the Konstanz Young Scholar Fund.

Author Contributions K.D. carried out most of the enzymic experiments, together with M.W. and A. K.F., who carried out the heterologous expressions and RT PCR. A. K.F., A.S., C.M. and D.Sp. carried out the LC MS analyses, and T.H. the chemical syntheses and NMR. D.Sc. set up the HILIC separation and carried out the 2D PAGE, growth physiology and mutant analyses. A.M.C. and D.Sc. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.C. (alasdair.cook@uni konstanz.de) or D.Sc. (david.schleheck@uni konstanz.de).

METHODS

Chemicals. SQ and DHPS were synthesized chemically and identified by NMR and mass spectrometry as described previously[8,10]. Dihydroxyacetone phosphate dilithium salt, D fructose 6 phosphate disodium salt hydrate, D glucose 6 phosphate disodium salt hydrate, and D fructose 1,6 bisphosphate trisodium salt octahydrate were from Sigma. Other biochemicals (NADH, NADPH, NAD^+, $NADP^+$, ATP, ADP) were purchased from Sigma, Fluka, Merck or Biomol.

Bacteria and growth conditions. *Escherichia coli* K 12 substrains W3100 (DSM 5911, ATCC 27325) and DH1 (DSM 4235, ATCC 33849), and *Cupriavidus pinatubonensis* JMP134 (DSM 4058)[20] were purchased from the Leibniz Institute DSMZ Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH. *E. coli* K 12 substrain MG1655 (DSM 18039) was a gift from E. Deuerling, and *E. coli* K 12 substrain BW25113 and its single gene knockouts from the *E. coli* Keio Knockout Collection[11] were a gift from J. Klebensberger. The growth medium was a phosphate buffered mineral salts medium[21] (pH 7.2) with SQ or glucose as the sole carbon sources. Cultures were inoculated (1%) with pre culture grown with the same substrate, and grown aerobically at 30 °C. Cultures in 3 ml volume were grown in screw cap tubes (30 ml) in a roller, and cultures in 50 ml or 200 ml volume in capped Erlenmeyer flasks (0.3 or 1.0 litre volume, respectively) on a horizontal shaker; for the Erlenmeyer flask cultures, 0.8 ml samples were taken at intervals to determine optical density (attenuance D at 580 nm; $D_{580\ nm}$) and substrate and product concentrations (HILIC HPLC, see below). For the growth experiments to demonstrate mineralization of SQ (see Fig. 1), *E. coli* K 12 substrain MG1655 was grown with SQ (4 mM; 50 ml scale), and after growth had been completed, the cellular biomass was removed from the culture fluid by centrifugation (20,000*g*, 30 min, 4 °C) followed by filter sterilization (pore size, 0.2 μm). The culture fluid was then inoculated with *C. pinatubonensis* JMP134. During the growth experiments, samples were taken at intervals to monitor the growth ($D_{580\ nm}$) and to determine total protein, substrate, and product concentrations (see below). All growth experiments were replicated twice (n 3).

Preparation of crude extract, soluble fraction and enzyme enrichment. *E. coli* cells from growth experiments with SQ or glucose were collected at an $D_{580\ nm}$ of approximately 0.4 by centrifugation (20,000*g*, 15 min, 4 °C) and disrupted by three passages through a chilled French pressure cell (140 megapascals (MPa); Aminco) in the presence of DNase (25 μg ml $^{-1}$). Cell debris was removed by centrifugation (11,000*g*, 5 min, 4 °C) and the membrane fragments collected by ultracentrifugation (70,000*g*, 1 h, 4 °C); the supernatant was called the soluble fraction. For enzyme enrichment, soluble fraction was loaded onto an anion exchange chromatography column (MonoQ HR 10/10 column, Pharmacia) equilibrated with 20 mM Tris/H_2SO_4 buffer, pH 9.0, at a flow rate of 1 ml min $^{-1}$, and bound proteins eluted by a linear Na_2SO_4 gradient (from 0 M to 0.2 M in 45 min, and to 0.5 M in 10 min) and fractions (2 ml) collected; the SLA reductase activity eluted at about 0.12 M Na_2SO_4.

Two dimensional gel electrophoresis and peptide fingerprinting mass spectrometry. 2D PAGE and PF MS were carried out according to our previously published protocols[22]. In brief, soluble protein fractions from *E. coli* cells grown with SQ or glucose (see above) were desalted (PD 10 Desalting Columns, GE Healthcare Life Sciences) and precipitated by addition of acetone (four volumes 100% acetone, −20 °C, overnight); each 1 mg of precipitated protein was solubilised in rehydration buffer (300 μl) and loaded onto isoelectric focusing (IEF) strips (BioRad ReadyStrip IPG system) overnight; IEF involved a voltage ramp to 10,000 V during 3 h, and a total focusing of 40,000 Volt hours (Vh); the strips were equilibrated in SDS equilibration buffers I and II (with DTT and iodoacetamide, respectively) and placed onto SDS PAGE gels using an overlay of SDS gel buffer solidified with agarose (0.5%); SDS PAGE gels contained 12% polyacrylamide (no stacking gel), and were stained with Coomassie brilliant blue R 250 (ref. 23). Stained protein spots of interest were excised from gels and submitted to PF MS at the Proteomics Facility of the University of Konstanz to identify the corresponding genes; the MASCOT engine (Matrix Science) was used to match each peptide fingerprint against a local database of all predicted protein sequences of the annotated *E. coli* K 12 substrain MG1655 genome (IMG version 2011 08 16).

Total RNA preparation and PCR with reverse transcription. RNA preparation and RT PCR were carried out according to our previously published protocols[24]. In brief, cells were grown in the appropriate selective medium (3 ml) and collected in the mid exponential growth phase ($D_{580\ nm} \approx 0.3$); the cell pellets were stored at −20 °C in RNA*later* RNA stabilization solution (Applied Biosystems); total RNA was prepared using the E.Z.N.A. Bacterial RNA Kit (Omega Bio Tek) following the manufacturer's instructions; the RNA preparation (40 μl) was treated with RNase free DNase (2 units, 30 min, 37 °C) (Fermentas). For complementary DNA (cDNA) synthesis, the Maxima reverse transcriptase (Fermentas) was used following the manufacturer's instructions; the reactions contained 0.2 μg total RNA and 20 pmol sequence specific primer (see below). PCR reactions (20 μl volume) were carried out using Taq DNA polymerase (Fermentas) and the manufacturer's standard reaction mixture (including 2.5 mM $MgCl_2$); cDNA from reverse transcription reactions was used as template (2 μl of reverse transcription reaction mixture), or genomic DNA (4 ng DNA) for PCR positive controls, or non reverse transcribed total RNA (2 μl) for the confirmation of an absence of DNA impurities in the RNA preparations.

Heterologous expression and purification of His tagged proteins. Heterologous expression of candidate genes and purification of the recombinant proteins were carried out according to our previously published protocol[25]. In brief, chromosomal DNA was isolated using the Illustra bacteria genomicPrep Mini Spin Kit (GE Healthcare) and the target genes amplified by PCR using Phusion HF DNA Polymerase (Finnzymes) and the primer pairs given below. The PCR conditions were 30 cycles of 18 s denaturation at 98 °C, 20 s annealing at 58 °C, and 60 s elongation at 72 °C for gene b3880, or 45 s elongation at 72 °C for genes b3881, b3882, and b3883. The PCR products were then separated by agarose gel electrophoresis, excised, and purified using the QIAquick gel extraction kit (Qiagen), and ligated into the amino terminal His_6 tag expression vector pET100 (Invitrogen); correct integration of the inserts was confirmed by sequencing (GATC Biotech). For expression, BL21 Star (DE3) OneShot *E. coli* cells (Invitrogen) were transformed with the constructs and grown at 37 °C in lysogeny broth medium containing 100 mg l $^{-1}$ ampicillin; at an $D_{580\ nm} \approx 0.6$, the cultures were induced by addition of 0.5 mM IPTG (isopropyl β D thiogalactoside), and the cells grown for additional 4 to 5 h at 20 °C, collected by centrifugation (15,000*g*, 20 min, 4 °C), and stored frozen (−20 °C). Cells were resuspended in buffer A (20 mM Tris HCl, pH 8.0, 100 mM KCl) that contained 50 μg ml $^{-1}$ DNase I, and disrupted by four passages through a pre cooled French pressure cell (140 MPa). The cell extracts were centrifuged (15,000*g*, 10 min, 4 °C) and ultracentrifuged (70,000*g*, 1 h, 4 °C), and the soluble protein fractions loaded onto 1 ml Ni^{2+} chelating Agarose affinity columns (Macherey Nagel) pre equilibrated with buffer A (see above). After a washing step (30 mM imidazole in buffer A), the His tagged proteins were eluted (200 mM imidazole in buffer A), concentrated in a Vivaspin concentrator (Sartorius), and, after addition of 30% glycerol (v/v), stored in aliquots at −20 °C.

Enzyme assays. SLA reductase activity was assayed photometrically at 365 nm in 1 ml cuvettes in 50 mM Tris HCl buffer, pH 9.0, with 1 mM NAD^+ and 5 mM DHPS as substrates. The reaction was started with the addition of protein (0.01 0.1 mg ml $^{-1}$) and the reduction of NAD^+ was recorded. Enzyme assays with recombinant proteins (each 50 μg protein ml $^{-1}$) for analysis by HILIC HPLC were carried out in 1 ml volume in 50 mM ammonium acetate buffer, pH 8.7, stirred at room temperature (approximately 20 23 °C); samples were taken at intervals, for which the reactions were stopped by addition of 30% acetonitrile. SQ (3 mM) and recombinant isomerase were incubated for 60 min, after which ATP (8 mM), $MgCl_2$ (0.5 mM) and recombinant kinase were added. After additional 60 min, recombinant aldolase was added, and after another 60 min NADH (8 mM) and the recombinant reductase.

Analytical methods. Total protein was determined according to a protocol based on a method reported previously[26], and soluble protein was determined by protein dye binding[27], each using bovine serum albumin (BSA) as the standard. Sulphate release during growth was quantified turbidimetrically[28] as a suspension of $BaSO_4$. For HPLC ESI MS MS, an Agilent 1100 HPLC system fitted with a ZIC HILIC column (5 μm, 200 Å, 150 × 4.6 mm; Merck) was connected to an LCQ ion trap mass spectrometer (Thermofisher). The HPLC conditions for the LCQ ion trap were: from 90% B to 65% B in 25 min, 65% B for 10 min, in 0.5 min back to 90% B, 90% B column equilibration for 9.5 min; solvent A, 90% 0.1 M NH_4Ac, 10% acetonitrile; solvent B, acetonitrile; flow rate, 0.75 ml min $^{-1}$. The mass spectrometer was run in ESI negative mode. The retention times and ESI MS MS fragmentation patterns of the analytes were observed as follows: SQ retention time, 25.4 min; SQ ESI MS m/z (per cent base peak) 243 (100); SQ ESI MS MS of [M H] 243: 243 (4), 225 (11), 207 (34), 183 (100), 153 (54), 143 (1), 123 (16), 101 (8), 81 (6). SF retention time, 21.9 min; SF ESI MS, 243 (100); SF ESI MS MS of [M H] 243: 243 (1), 225 (37), 207 (38), 183 (21), 153 (100), 143 (3), 123 (24), 101 (13), 81 (5). SFP retention time, 33.4 min; SFP ESI MS, 323 (100); SFP ESI MS MS of [M H] 323: 305 (34), 287 (3), 233 (2), 225 (100), 207 (32), 153 (4). SLA retention time, 21 min; SLA ESI MS, 153 (100); SLA ESI MS MS of [M H] 153: 153 (9), 81 (100), 71 (18). DHPS retention time, 18 min; DHPS ESI MS, 155 (100), 95 (4); DHPS ESI MS MS of [M H] 155: 155 (100), 137 (18), 95 (40). DHAP retention time, 30.2 min; DHAP ESI MS: 169 (100)); DHAP ESI MS MS of [M H] 169: 169 (2), 125 (2), 97 (100). HPLC for ESI TOF MS (MicrO TOF II, Bruker) involved the same column and gradient system, but a different gradient program (from 90% B to 65% B in 20 min, and further to 55% B in 20 min), which resulted in retention times (see Extended Data Fig. 4) of 22.4 min for SQ, 21.8 min for SF, 16.3 min for fructose, 18.5 min for glucose, 35.2 min for fructose 6 phosphate, and 39.8 min for glucose 6 phosphate.

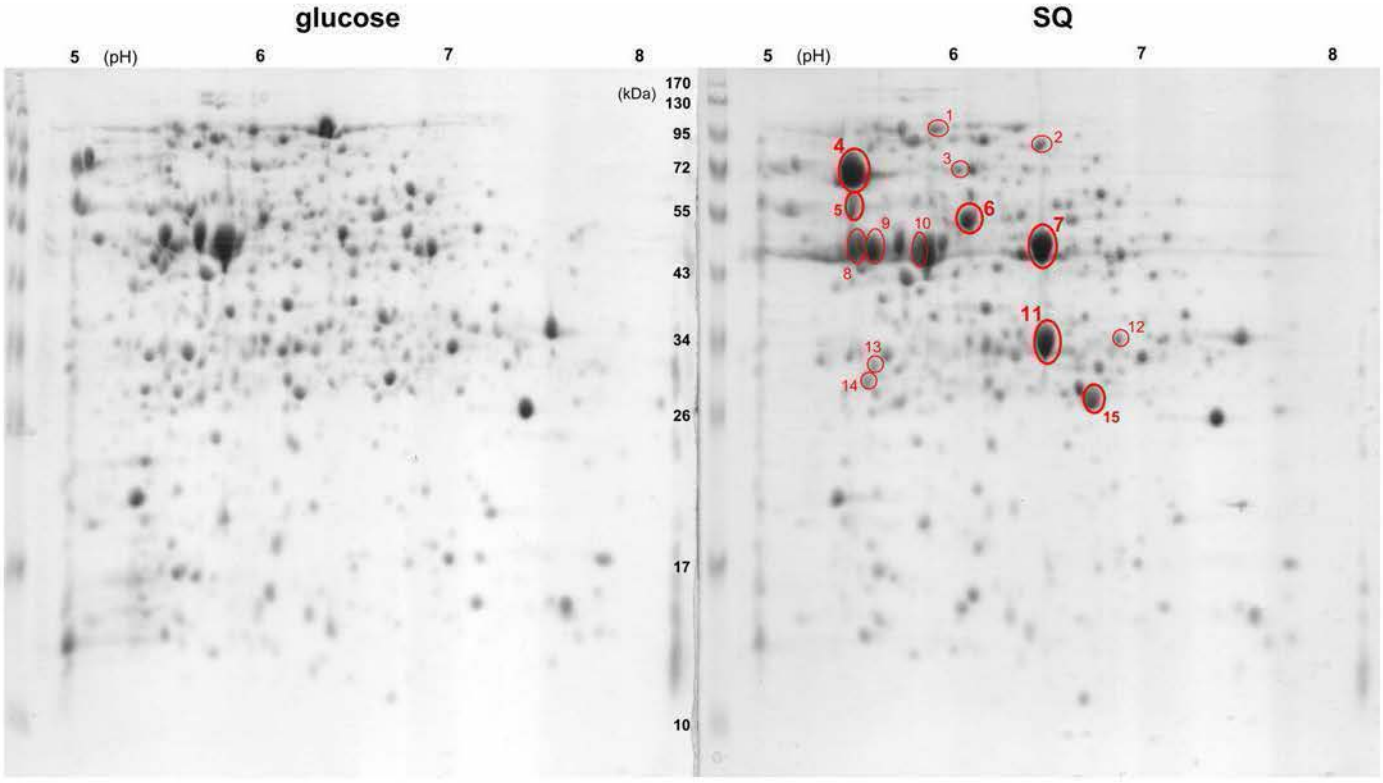
PCR primers. Primers were purchased from Microsynth (Balgach). The sequences of the primers pairs for RT PCR (see above) were (product length in bp): b3879 forward, 5′ CCTTATGGCGTGGGTATTCATCC 3′, b3879 reverse, 5′ TTAGG

CGGGCAACTCATAGGTTC 3′ (353); b3880 forward, 5′ ACGCGGTGGAAG CTTTCTTGAT 3′, b3880 reverse, 5′ CACGGTGGCGTTAAACAGACCTT 3′ (332); b3881 forward, 5′ TGTCGCCGCCGATGAGTTC 3′, b3881 reverse, 5′ C TTTGTAGAGGTCAGCGCCAC TGT 3′ (320); b3882 forward, 5′ GGCGCAG GCCGCTAAAGA 3′, b3882 reverse, 5′ AAGATTCAGGGCTTCGCACAAAA 3′ (439); b3883 forward, 5′ GGCACGACGGCGCTAAAAA 3′, b3883 reverse, 5′ TGACTCCGCTAAATCCCCACTTG 3′ (374); b0720 forward, 5′ CGCTGGCG GCGTTCTATCA 3′, b0720 reverse, 5′ ATTTTCAGCGCCGCTTCGTTAG 3′ (403). The sequences of the primers pairs for TOPO cloning and heterologous expression (see above) were (the directional overhang is underlined): b3880 for ward, 5′ CACCGGAATGAAATGGTTTAACACCCTAAG 3′, b3880 reverse, 5′ AACCCGCACCCTATTTTCAG 3′; b3881 forward, 5′ CACCATGAATAA GTACACCATCAACGACATTACG 3′, b3881 reverse 5′ ACCATTTCATTCC TTTTATCCTCATCTT 3′; b3882 forward, 5′ CACCATGGCAGCAATCGCGT TTATCG 3′, b3882 reverse, 5′ CGCGTAATGTCGTTGATGGTGTA 3′; b3883 forward, 5′ CACCATGATTCGTGTTGCTTGTGTAGGT 3′, b3883 reverse, 5′ TGAAAATTCCTCGAAAAACCATCA 3′.

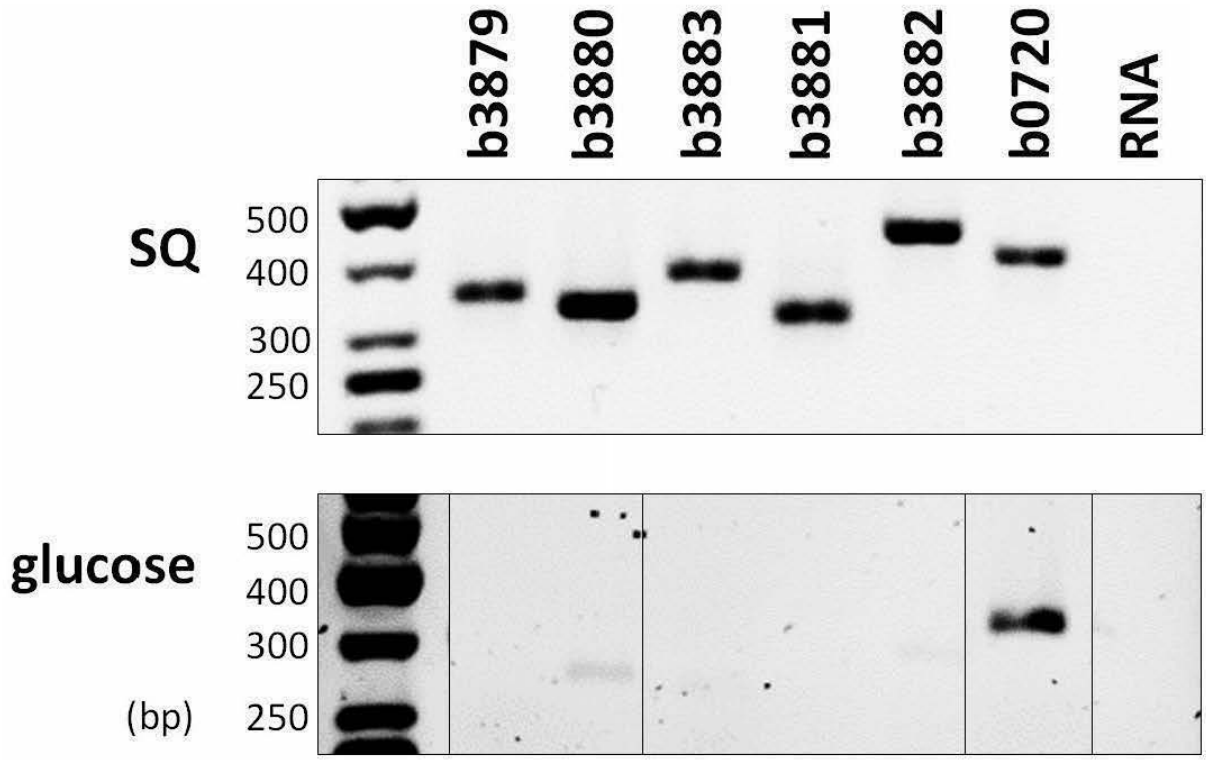
Genome analyses. Analysis of genomes for orthologous gene clusters was carried out through the gene cassette search and neighbourhood regions search options of the Integrated Microbial Genomes (IMG) and IMG Human Microbiome Project (IMG HMP) platforms (http://img.jgi.doe.gov/). Basic sequence analyses were done using NCBI's BLAST tools (http://blast.ncbi.nlm.nih.gov) and the Lasergene DNAstar software package (http://www.dnastar.com).

Enzyme nomenclature. We suggest that sulpholactaldehyde reductase belongs to the NC IUBMB subgroup EC 1.1.1., with the name sulpholactaldehyde 3 reductase (systematic name 2,3 dihydroxypropane 1 sulphonate:NAD^+ 3 oxidoreductase). The sulphofructose kinase would then belong to EC 2.7.1., with the name sulphofruc tose 1 kinase (systematic name ATP:6 deoxy 6 sulphofructose 1 phosphotransferase). The sulphofructosephosphate aldolase would belong to EC 4.1.2., with the name sulphofructosephosphate aldolase (systematic name 6 deoxy 6 sulphofructose 1 phosphate 2 hydroxy 3 oxopropane 1 sulphonate lyase (glycerone phosphate forming)). Sulphoquinovose isomerase would belong to EC 5.3.1., with the name sulphoquinovose isomerase (systematic name 6 deoxy 6 sulphoglucose aldose ketose isomerase).

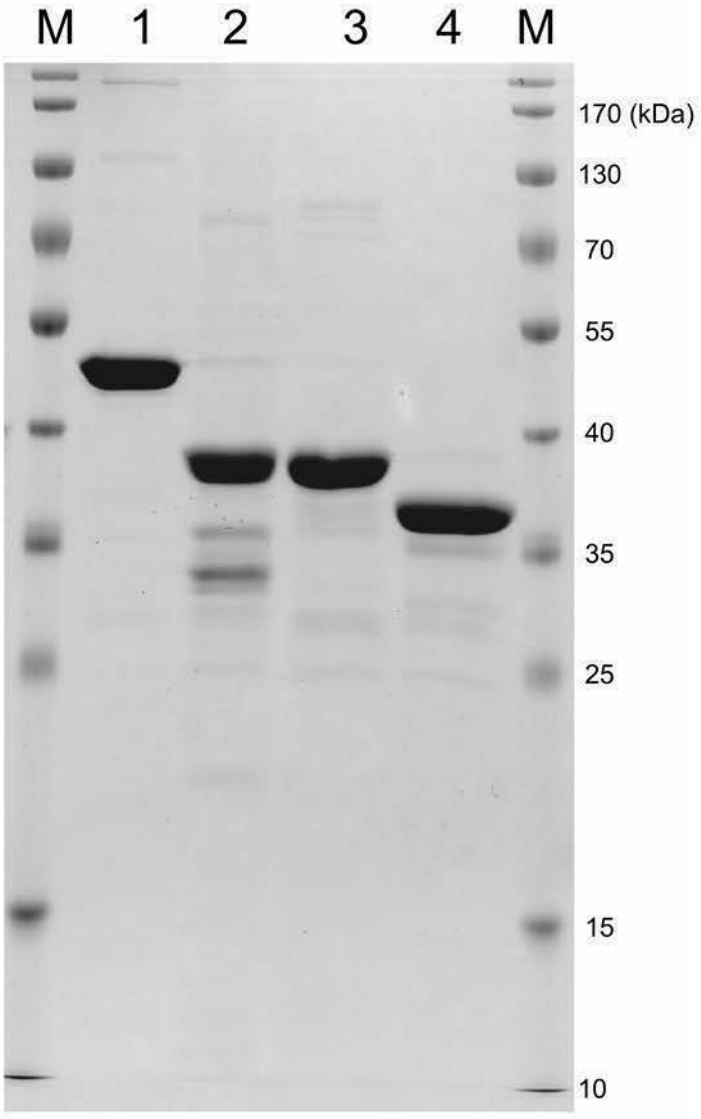
20. Sato, Y. *et al. Cupriavidus pinatubonensis* sp. nov. and *Cupriavidus laharis* sp. nov., novel hydrogen oxidizing, facultatively chemolithotrophic bacteria isolated from volcanic mudflow deposits from Mt. Pinatubo in the Philippines. *Int. J. Syst. Evol. Microbiol.* **56,** 973 978 (2006).
21. Thurnheer, T., Köhler, T., Cook, A. M. & Leisinger, T. Orthanilic acid and analogues as carbon sources for bacteria: growth physiology and enzymic desulphonation. *J. Gen. Microbiol.* **132,** 1215 1220 (1986).
22. Schmidt, A., Müller, N., Schink, B. & Schleheck, D. A proteomic view at the biochemistry of syntrophic butyrate oxidation in *Syntrophomonas wolfei*. *PLoS ONE* **8,** e56905 (2013).
23. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680 685 (1970).
24. Weiss, M., Denger, K., Huhn, T. & Schleheck, D. Two enzymes of a complete degradation pathway for linear alkylbenzenesulfonate (LAS) surfactants: 4 sulfoacetophenone Baeyer Villiger monooxygenase and 4 sulfophenylacetate esterase in *Comamonas testosteroni* KF 1. *Appl. Environ. Microbiol.* **78,** 8254 8263 (2012).
25. Felux, A. K., Denger, K., Weiss, M. & Cook, A. M & Schleheck, D. *Paracoccus denitrificans* PD1222 utilizes hypotaurine via transamination followed by spontaneous desulfination to yield acetaldehyde, and finally acetate for growth. *J. Bacteriol.* **195,** 2921 2930 (2013).
26. Kennedy, S. I. T. & Fewson, C. A. Enzymes of the mandelate pathway in bacterium N.C.I.B. 8250. *Biochem. J.* **107,** 497 506 (1968).
27. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72,** 248 254 (1976).
28. Sörbo, B. Sulfate: turbidimetric and nephelometric methods. *Methods Enzymol.* **143,** 3 6 (1987).



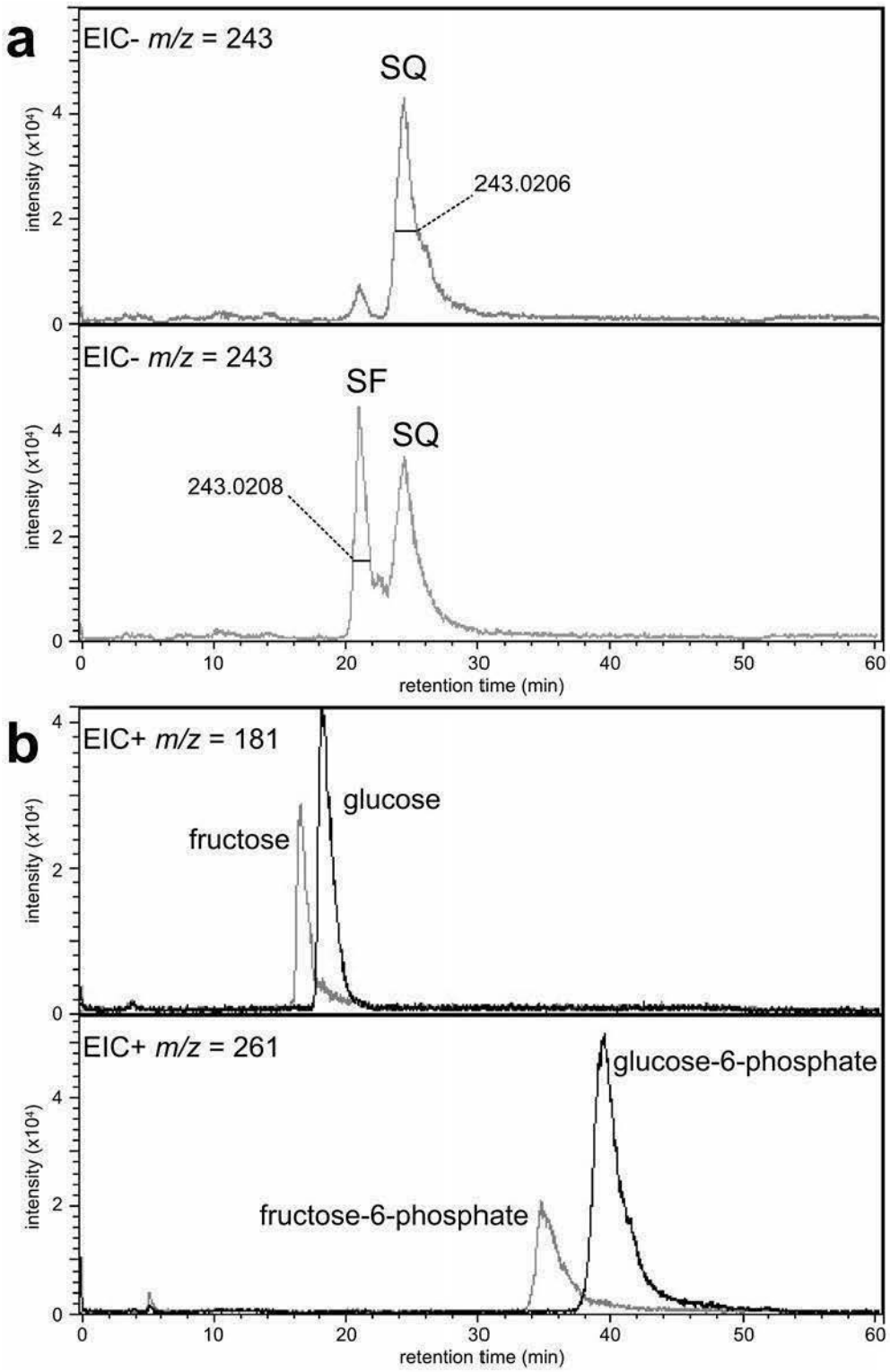
Extended Data Figure 1 | Soluble proteins in glucose or SQ grown cells of *E. coli* K 12 MG1655 separated by 2D PAGE. All prominent protein spots on the gel from SQ grown cells that suggested inducibly expressed proteins were excised and submitted to PF MS (see Extended Data Table 1). The PF MS identifications were replicated in an independent growth experiment and gel electrophoresis run.



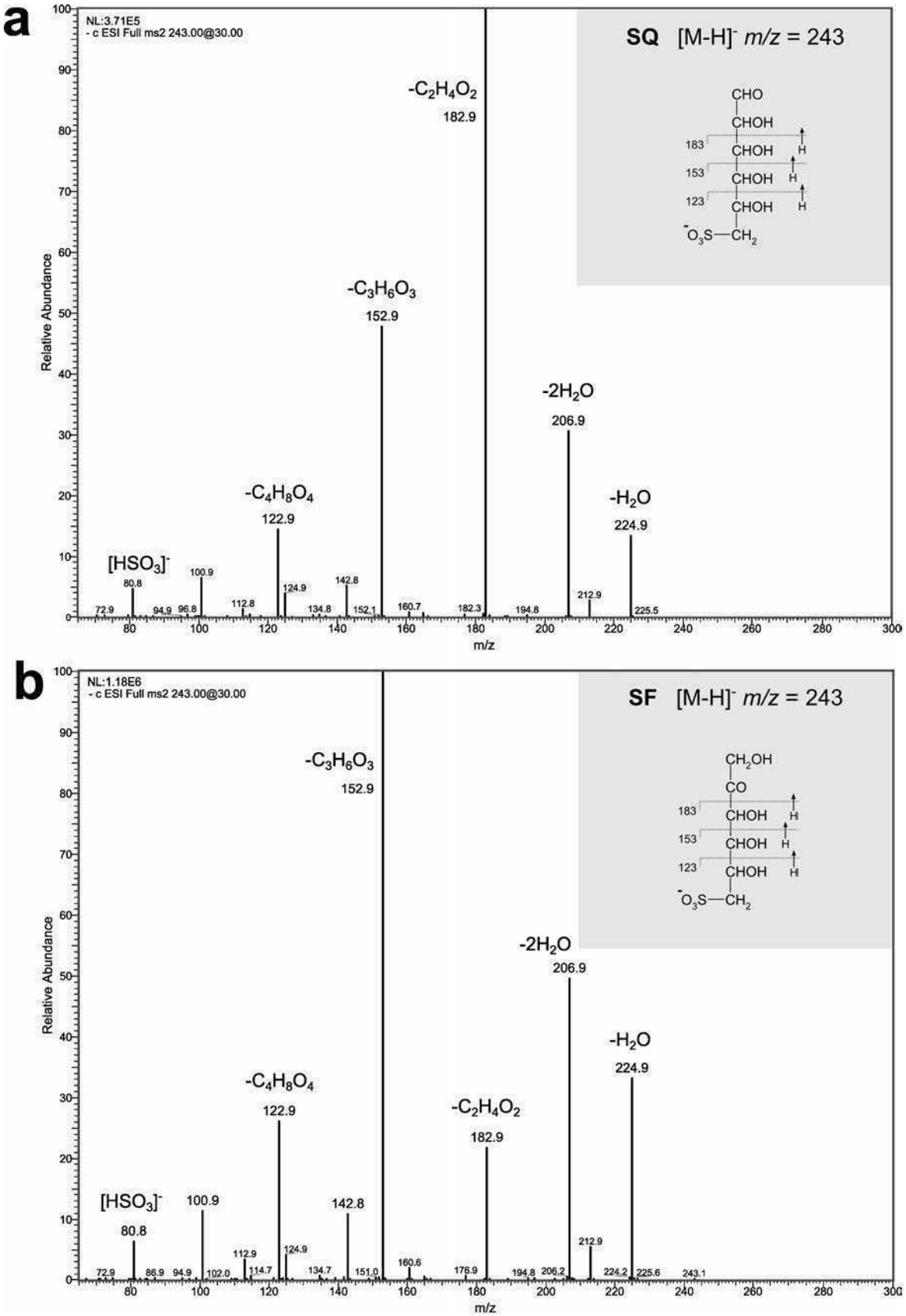
Extended Data Figure 2 | Differential transcriptional analysis of the genes encoding the central pathway of sulphoglycolysis in *E. coli* K 12. RT PCR of the inducible transcription of genes b3879 (epimerase), b3880 (isomerase), b3883 (kinase), b3881 (aldolase) and b3882 (reductase) in cells grown with SQ in comparison to glucose grown cells. A positive control was a constitutively expressed gene (b0720; *gltA*, citrate synthase), and a negative control was a PCR without reverse transcription (RNA) to confirm the absence of DNA contamination in the RNA preparations used. The results were replicated starting from an independent growth experiment.



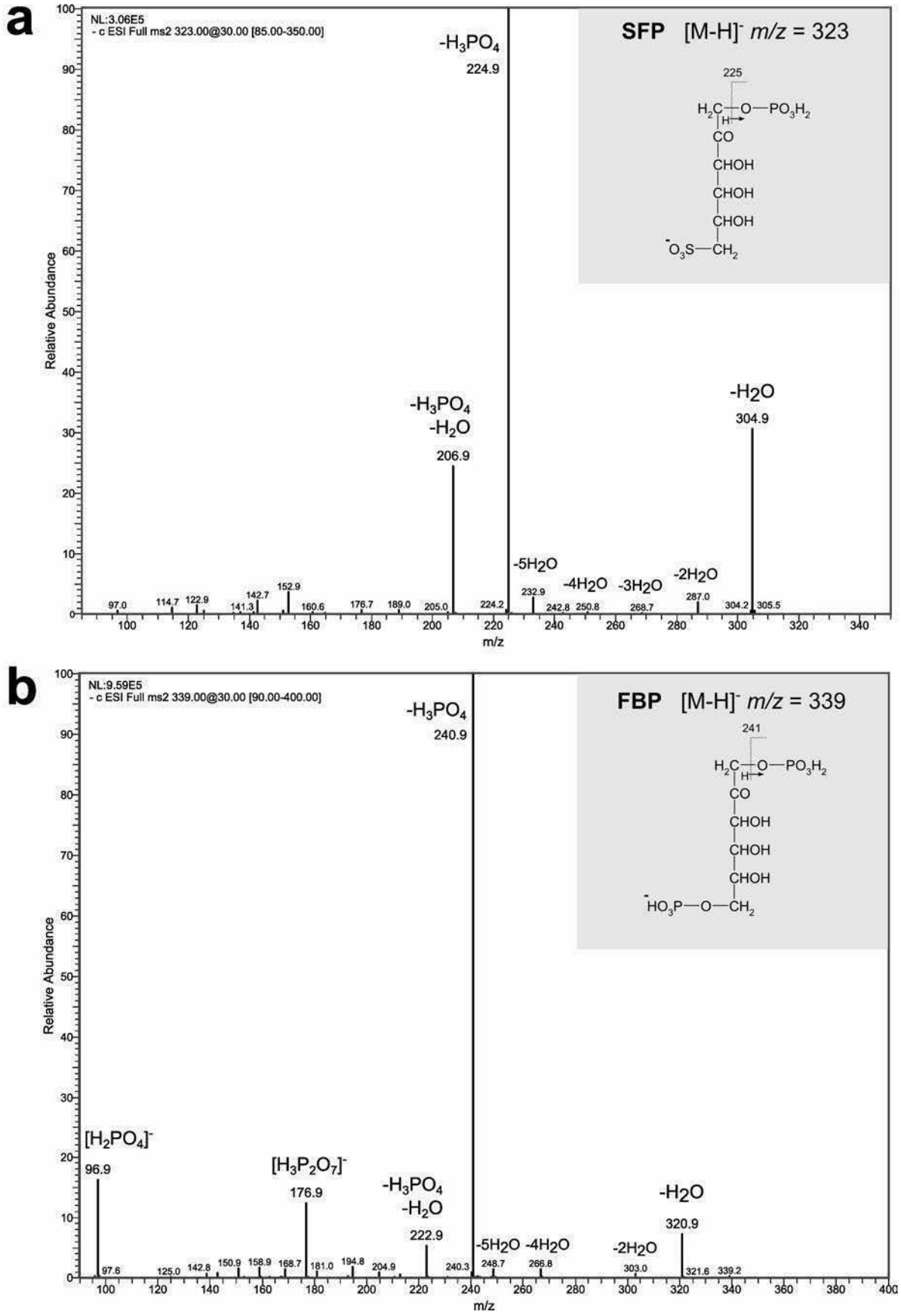
Extended Data Figure 3 | Purity of heterologously expressed enzymes using SDS PAGE. M, marker proteins; 1, b3880 (isomerase); 2, b3883 (sugar kinase); 3, b3881 (aldolase); 4, b3882 (reductase). A representative SDS gel is shown (n 2). Each enzyme, 20 μg.



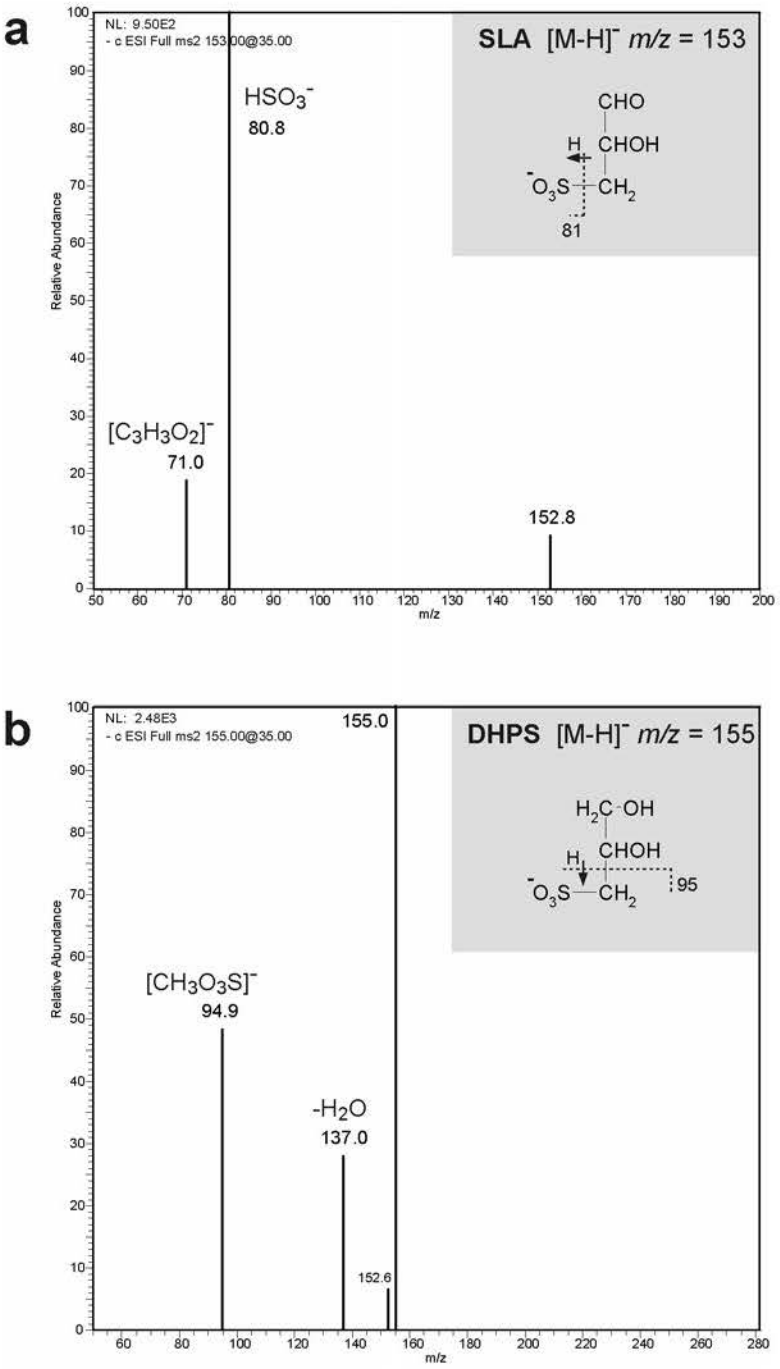
Extended Data Figure 4 | Separation of SQ, SF and analogues by HILIC HPLC and detection by ESI TOF MS. **a**, Extracted ion chromatogram of SQ in reaction buffer before (top) and after (bottom) addition of recombinant isomerase b3880 to generate SF. The exact masses determined for the [M H] ions of SQ and SF are indicated (in Da); the theoretical exact masses (monoisotopic masses) for the [M H] ions of both SQ and SF (each $C_6H_{11}O_8S$) is 243.0180 Da. The data from a representative experiment are shown; the results were replicated (n 3) with samples from independent enzyme reactions. **b**, Samples of reference substances fructose and glucose (top), and fructose 6 phosphate and glucose 6 phosphate (bottom). The reference substances illustrate the chromatographic separation by HILIC; that is, the ketoses (for example, SF) elute before the aldoses (for example, SQ).



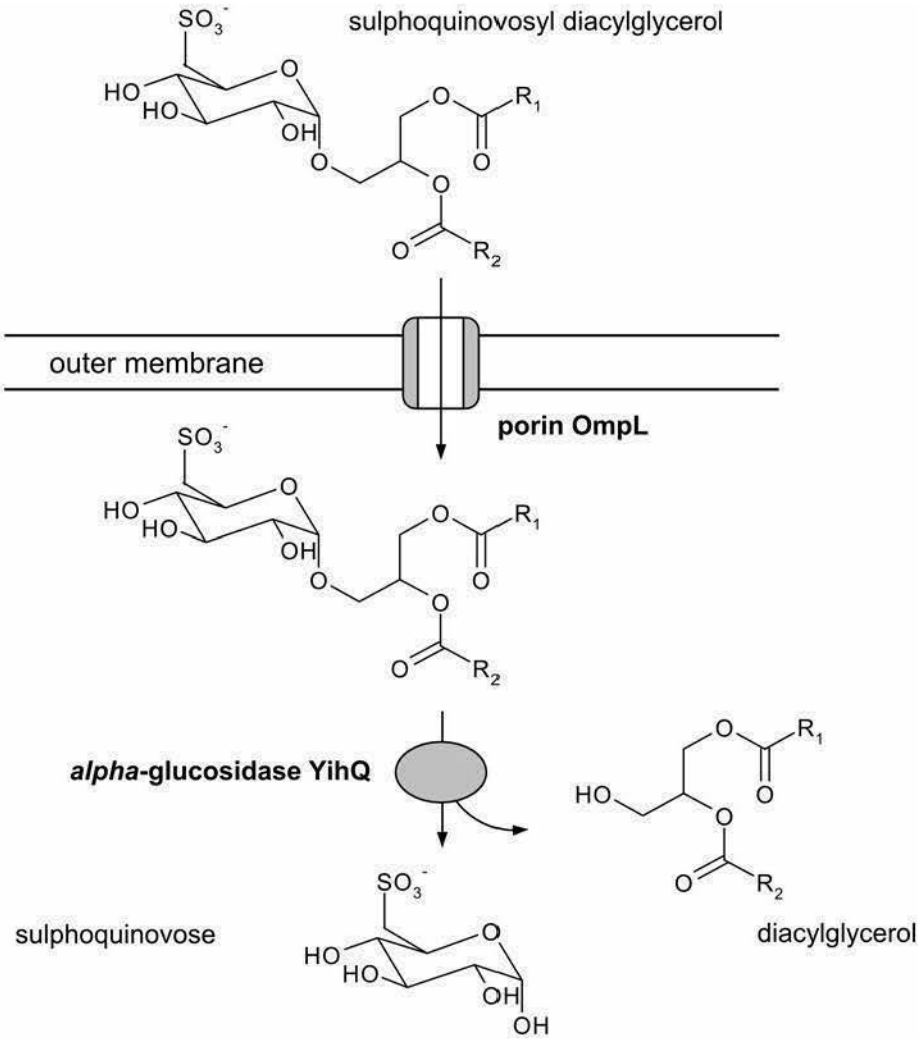
Extended Data Figure 5 | MS MS fragmentation of SQ and SF. a, Fragment ions of the [M H] ions of SQ. b, Fragment ions of the [M H] ions of SF. The fragmentation pattern of both compounds, SQ and SF, is dominated by the loss of water (−18), $C_2H_4O_2$ (−60), $C_3H_6O_3$ (−90) and $C_4H_8O_4$ (−120), and by the formation of HSO_3 (81) ions. The MS MS spectra of SQ and SF differ in their peak heights, in particular in their base peak for SQ (183) and of SF (153) corresponding to a preferred formation of $C_4H_7O_6S$ and $C_3H_5O_5S$, respectively; both fragments might be formed by McLafferty like rearrangement reactions corresponding to the different positions of the keto groups in SQ and SF. Representative data are shown (n 5; see Fig. 3).



Extended Data Figure 6 | MS MS fragmentation of SFP and FBP. **a**, Fragment ions of the [M H] ions of SFP. **b**, Fragment ions of the [M H] ions of the analogue fructose 1,6 bisphosphate (FBP) for comparison. Fragmentations led to loss of water (−18) and/or loss of phosphoric acid (−98), and to the formation of dihydrogen phosphate (97) and pyrophosphate (177). Representative data are shown (n 5; see Fig. 3).



Extended Data Figure 7 | MS MS fragmentation of SLA and DHPS. **a**, Fragment ions of the [M H] ions of SLA. **b**, Fragment ions of the [M H] ions of DHPS. Fragmentation of SLA led to a cleavage of the carbon sulphur bond and formation of HSO_3 (81) and $[C_3H_3O_2]$ (71) ions. The fragmentation of DHPS is characterized by an initial loss of water (−18) and concomitant fragmentation of the enol (or the methylketone) to ketene (−42) and formation of $[CH_3O_3S]$ (95) ions; the same fragmentation pattern was observed for authentic DHPS standard (not shown). Representative data are shown (*n* 5; see Fig. 3).



Extended Data Figure 8 | Hypothetical degradation of the sulpholipid sulphoquinovosyl diacylglycerol in *E. coli* K 12. Presumed functions of OmpL (b3875) and YihQ (b3878); apart from OmpL, the subcellular location of this pathway is unknown. R_1 and R_2 indicate acyl chains of different length and degree of unsaturation.

Extended Data Table 1 | Identifications by peptide fingerprinting–mass spectrometry of protein spots excised from 2D-PAGE gels of SQ-grown *E. coli* cells

Protein spot (no.)	Apparent mass on gel (kDa)	Apparent pI on gel (pH)	PF-MS identification					
			Gene (locus tag)	Annotation	Predicted mass (kDa)	Predicted pI (pH)	Score	Sequence coverage (%)
1	95	6	b0114	pyruvate dehydrogenase, decarboxylase component (*aceE*)	99,948	5.46	999	54
			b3829	homocysteine transmethylase (*metE*)	85,020	5.61	246	35
2	80	6.5	b0903	pyruvate formate lyase (*pflB*)	85,588	5.69	757	57
3	70	6	b2935	transketolase, thiamine-binding (*tktA*)	72,451	5.43	683	59
4	70	5.5	**b3878**	***alpha*-glucosidase (*yihQ*) / glycosyl hydrolase family 31**	77,853	5.06	1198	76
5	55	5.5	b1415	NAD-linked aldehyde dehydrogenase (*aldA*)	52,411	5.07	1475	70
6	50	6.0	b2095	tagatose-1,6-bisphosphate aldolase, non-catalytic subunit (*gatZ*)	47,535	5.50	1135	84
7	45	6.5	**b3880**	**aldose-ketose isomerase (*yihS*)**	47,687	5.71	839	84
8	45	5.3	b3880	aldose-ketose isomerase (*yihS*) / D-mannose isomerase	47,687	5.71	554	73
			b1136	$NADP^+$-linked isocitrate dehydrogenase	46,070	5.15	410	55
			b4015	isocitrate lyase (*aceA*)	47,777	5.16	324	36
			b2942	*S*-adenosylmethionine synthetase (*metK*)	42,153	5.10	222	47
			b2029	6-phosphogluconate dehydrogenase, decarboxylating	51,563	5.06	207	46
9	45	5.5	b4015	isocitrate lyase (*aceA*)	47,777	5.15	1256	91
			b1136	isocitrate dehydrogenase, specific for $NADP^+$	46,070	5.16	108	13
10	45	5.7	b3339	protein chain elongation factor EF-Tu	43,427	5.30	803	74
			b4177	adenylosuccinate synthetase (*purA*)	47,543	5.31	511	52
			b1493	glutamate decarboxylase (*gadB*)	53,204	5.29	277	48
11	35	6.5	**b3881**	**predicted aldolase (*yihT*)**	32,248	5.74	1487	92
12	35	7	**b3879**	**predicted aldose-1-epimerase (*yihR*)**	34,387	6.07	155	35
			b0729	succinyl-CoA synthetase, alpha subunit	30,044	6.32	152	29
13	30	5.5	b2150	D-galactose-binding periplasmic protein (*mglB*)	35,690	5.68	733	87
14	30	5.5	b2310	lysine/arginine/ornithine-binding periplasmic protein (*argT*)	28,088	5.62	914	83
			b1130	transcriptional regulatory (*phoP*)	25,519	5.10	199	37
15	25	7	**b3882**	***gamma*-hydroxybutyrate dehydrogenase (*yihU*)**	31,525	5.86	1293	89

Protein spots were sorted according to their apparent molecular mass on the gel (see Extended Data Fig. 1).