Sulfolactaldehyde dehydrogenase of *Rhizobium leguminosarum:* Kinetic and structural analysis

Jinling Li,[a] Mahima Sharma,[b] Richard Meek,[b] Amani Alhifthi,[a,c] Zachary Armstrong,[a] Niccolay Madiedo Soler,[d] Ethan D. Goddard-Borger,[d,e] Jamie N. Blaza,[b] Gideon J. Davies,[b,*] Spencer J. Williams[a,*]

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

[b] York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, United Kingdom

[c] Chemistry Department, Faculty of Science (Female section), Jazan University, Jazan 82621, Saudi Arabia

[d] ACRF Chemical Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3010, Australia

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

* Corresponding authors
Spencer J. Williams: sjwill@unimelb.edu.au
Gideon J. Davies: gideon.davies@york.ac.uk

Keywords: biogeochemical sulfur cycle; enzyme mechanism; structural biology; organosulfur metabolism; bioinformatics

Abstract

Sulfolactate (SL) is a short-chain organosulfonate that is an important reservoir of sulfur in the biosphere. SL is produced by oxidation of sulfolactaldehyde (SLA), which in turns derives from sulfoglycolysis of the sulfosugar sulfoquinovose, or through oxidation of 2,3-dihydroxypropanesulfonate, which are themselves important organosulfur compounds in the biogeochemical cycle. Oxidation of SLA is catalyzed by SLA dehydrogenases belonging to the aldehyde dehydrogenase superfamily. We report that SLA dehydrogenase *Rl*GabD from the sulfoglycolytic bacterium *Rhizobium leguminsarum* SRDI565 can use both NAD^+ and $NADP^+$ as cofactor to oxidize SLA, and operates through a rapid equilibrium ordered mechanism. We report the cryo-EM structure of *Rl*GabD bound to NADH, revealing a tetrameric quaternary structure and supporting proposal of organosulfonate binding residues in the active site, and a catalytic mechanism. Sequence based homology searches identified putative SLADH homologs in a range of sulfoglycolytic gene clusters in bacteria predominantly from the phyla Actinobacteria, Firmicutes, and Proteobacteria.

Introduction

Sulfur is the tenth most common element by mass in the universe, and the sixteenth most common in the Earth's crust, and the sixth most abundant in seawater.[1] It joins nitrogen, phosphorus, and potassium as the fourth macronutrient required for plants. The breakdown of C3-organosulfonates, primarily 2,3-dihydroxypropanesulfonate (DHPS) and sulfolactate (SL), allows the recycling of the element sulfur.[2] DHPS and SL are produced from the reduction or oxidation, respectively, of sulfolactaldehyde (SLA).[3] SLA in turn is produced in the pathways of sulfoglycolysis, in which the C6-organosulfonate sulfoquinovose (SQ) is catabolized (**Figure 1a**).[4] Alternatively, SLA may be produced by anaerobic DHPS degrading bacteria through the oxidation of DHPS (**Figure 1b**).[5] Reduction of SLA to DHPS is catalyzed by SLA reductase, an NADH-dependent enzyme, which has been biochemically and structurally characterized.[6, 7] On the other hand oxidation of SLA to SL is poorly studied, with only basic evidence for the formation of product in coupled assays (*vide infra*).

Three sulfoglycolytic pathways produce SLA by cleaving the 6-carbon chain of SQ into two C3 chains, namely the sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP/EMP2),[6, 8] Entner-Doudoroff (sulfo-ED)[9] and sulfofructose transaldolase (sulfo-SFT) pathways (**Figure 1a**).[10, 11] These pathways generate dihydroxyacetone phosphate (DHAP), pyruvate or fructose-6-phosphate (by transfer of a C3-glycerone moiety to glyceraldehyde-3-phosphate (GAP)), which are utilized by the host, and SLA, which is either reduced (to DHPS) or oxidized (to SL), and excreted. Examples of SL producing sulfoglycolytic organisms include: the sulfo-ED pathway (*Pseudomonas putida* SQ1,[9] and *Rhizobium leguminosarum* bv. trifolii SRDI565[8]); the sulfo-EMP/EMP2 pathway (*Bacillus urumquiensis*,[7] *Arthrobacter* spp.[9]); and the sulfo-SFT pathway (*Bacillus aryabhattai* SOS1,[11] *Bacillus megaterium* DSM1804,[12] and *Enterococcus gilvus*[10]); gene clusters encoding these pathways are shown in **Figure 2**). Excreted DHPS and SL are substrates for biomineralization bacteria. In the DHPS degradation pathway used by *Desulfovibrio* sp. strain DF1, DHPS is oxidized to SLA, and then SLA dehydrogenase SlaB oxidizes SLA to SL (**Figure 1b**).[5] SL is a substrate for SL lyase, which cleaves the C–S bond, producing pyruvate and sulfite. Other bacteria, such as *Roseovarius nubinhibens* and *Paracoccus pantotrophus*, utilize SL as a substrate for growth through the direct action of SL lyase.[5, 12, 13]

SLA dehydrogenases belong to the sequence-based protein family within the Pfam database PF00171 (sometimes annotated as GabD or SlaB), which are members of the aldehyde dehydrogenase superfamily.[13] Proteins of this superfamily oxidize the oxo group of aldehyde substrates to carboxylic acids, and use either NAD^+ or $NADP^+$ as hydride acceptors. Other activities within family PF00171 include succinate-semialdehyde dehydrogenase (SSADH),[14] non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPDH),[15] and glutarate semialdehyde reductase.[16] The potential cross-reactivity of SLA dehydrogenase with the structurally similar glycolytic intermediate

glyceraldehyde-3-phosphate has not been reported.

Recombinant SLA reductase from *P. putida* SQ1 reduced SLA formed in situ in a coupled assay with both NAD^+ and $NADP^+$ cofactors,[7] with a preference for the former, while SLA dehydrogenases from *B. aryabhattai* SOS1 (SftD),[11] *B. megaterium* (SlaB)[10] and *Desulfovibrio* sp. strain DF1 (SlaB) were described as NAD^+ dependent, although it is unclear whether their ability to utilize $NADP^+$ was assessed. In all cases accurate kinetic parameters have not been reported for any SLA dehydrogenase as SLA was not available in pure form. Recently, our group synthesized SLA from glycidol diethyl acetal using a chemical method,[17] meaning that a comprehensive kinetic characterization of SLA dehydrogenase is now possible.

Here, we report the structure and reactivity of SLA dehydrogenase from *R. leguminasarum* SRDI565 (*Rl*GabD), which oxidizes SLA produced in a sulfo-ED pathway in this organism.[8] We measure Michaelis-Menten kinetics and show its ability to use both NAD^+/$NADP^+$ as cofactors, its cross-reactivity to the structurally-related glycolytic metabolite GAP, and its sensitivity to inhibition by reduced NADH analogues. We determine its kinetic reaction order and show that it follows an equilibrium ordered mechanism in the forward direction. We report the 3D structure of SLA dehydrogenase using CryoEM and define its quaternary structure and substrate binding pocket, allowing proposal of a chemical mechanism of catalysis. Finally, we explore the sequence-based taxonomic distribution of SLA dehydrogenases across sulfoglycolytic and DHPS-degrading pathways using sequence similarity network analysis.

Results and Discussion

*Rl*GabD is an NAD(P)$^+$-dependent SLA oxidase

The gene encoding GabD from *Rhizobium leguminosarum* (*Rl*GabD) was cloned, recombinantly expressed in *E. coli*, and purified to homogeneity. Reaction rates for the oxidation of SLA catalyzed by *Rl*GabD were measured using chemically-synthesized racemic D/L-SLA[17] and monitoring reduction of NAD(P)$^+$ to NAD(P)H using a UV/Vis spectrophotometer. Incubation of a solution of racemic SLA (1 mM) in Tris buffer with *Rl*GabD and excess NAD$^+$ gave a progress curve that indicated complete reaction after 40 min; addition of more *Rl*GabD did not result in further conversion (**Figure S1**). Based on the change in absorbance and the extinction coefficient for NAD$^+$ we calculate that 48.7% of the SLA was consumed and conclude that *Rl*GabD is stereospecific for D-SLA. All subsequent analysis used the calculated D-SLA concentration (ie [SLA]/2).

Apparent Michaelis-Menten parameters were measured for D-SLA, NAD$^+$ and NADP$^+$ under pseudo first order conditions, in which one substrate was held at a constant concentration while that of the other was varied (**Figure 3a-d, Table 1**). At 0.25 mM D-SLA, the pseudo first order parameters for NAD$^+$ are: k_{cat}^{app} = 17.7 s^{-1}, K_M^{app} = 0.081 mM and $(k_{cat}/K_M)^{app}$ = 210 mM^{-1} s^{-1} and for NADP$^+$: k_{cat}^{app} = 4.1 s^{-1}, K_M^{app} = 0.017 mM and $(k_{cat}/K_M)^{app}$ = 240 mM^{-1} s^{-1}. Thus, while NADP$^+$ has a lower K_M^{app} value, the $(k_{cat}/K_M)^{app}$ values of the two nucleotides are essentially identical. For variable D-SLA the Michaelis-Menten parameters at constant concentration (25 mM) of nucleotide were, NAD$^+$: k_{cat}^{app} = 17.8 s^{-1}, K_M^{app} = 0.13 mM and $(k_{cat}/K_M)^{app}$ = 137 mM^{-1} s^{-1}; and NADP$^+$: k_{cat}^{app} = 4.8 s^{-1}, K_M^{app} = 0.16 mM and $(k_{cat}/K_M)^{app}$ = 30 mM^{-1} s^{-1}. Comparison of $(k_{cat}/K_M)^{app}$ reveal a modest preference for NAD$^+$. Above 0.25 mM SLA, we observed substrate inhibition and so data were fit to rates measured at concentrations below this limit.

*Rl*GabD oxidizes GAP and binds reduced NADH analogues

GAP is produced in glycolysis/gluconeogenesis during sulfoglycolytic growth and has a similar structure to SLA. Therefore, we investigated if *Rl*GabD can catalyze the oxidation of GAP. GAP was synthesized from racemic glyceraldehyde-3-phosphate diethyl acetal barium salt.[18,19] Apparent Michaelis-Menten kinetics for oxidation of racemic GAP measured at constant concentration (25 mM) of nucleotide (**Figure 3e,f**). These data reveal that the apparent second order rate constants were similar for NAD$^+$ $(k_{cat}/K_M)^{app}$ = 4.19 s^{-1} mM^{-1} and NADP$^+$ $(k_{cat}/K_M)^{app}$ = 10.5 s^{-1} mM^{-1} with a modest preference for NADP$^+$, opposite to that seen for SLA (**Table 1**). As for the kinetics with SLA, inhibition was also observed when the concentration of GAP was higher than 0.25 mM, and so rate data used for Michaelis-Menten analysis was used below this limit. The ratio of apparent second order rate constants ($(k_{cat}/K_M)^{app}$) for SLA and GAP at constant nucleotide concentration reveals that the activity on GAP is approx. 30-fold lower than SLA.

To explore the ability of *Rl*GabD to bind analogues of NADH we synthesized tetrahydro- and hexahydro-NADH by reduction of NADH following the procedure of Dave.[20] IC_{50} values were measured at constant [SLA] (fixed around K_M^{SLA} /10) and constant [NAD^+] (fixed around K_M^{NAD+}) (**Figure S2a,b**). For tetrahydro-NADH, IC_{50} = 28.2 μM, and for hexahydro-NADH, IC_{50} = 9.1 μM, indicating the latter binds more tightly (**Figure S2c,d**).

***Rl*GabD follows a rapid equilibrium ordered kinetic mechanism**

*Rl*GabD is a bisubstrate enzyme that acts on two substrates (NAD(P)H and SLA) and produces two products (SL and reduced NAD(P)H) and so its kinetic mechanism is described as Bi-Bi. Such Bi-Bi reactions can occur through non-sequential (Ping Pong) or sequential (ordered, steady-state random, and Theorell-Chance (a special case of ordered reactions where the steady-state level of central complexes is low)) mechanisms.[21] Analysis of initial rates is a powerful way to distinguish reaction mechanisms. For bisubstrate enzymes with substrates A and B, a plot of $1/\upsilon_0$ versus 1/[A] at various concentrations of substrate B, or $1/\upsilon_0$ versus 1/[B] at various constant concentrations of substrate A can help determine the kinetic mechanism. For a Ping Pong reaction, the plot of $1/\upsilon_0$ versus 1/[A] will afford a series of straight lines with constant slope = $K_M(A)/V_{max}$. On the other hand, for a sequential mechanism (ordered or random) the same plot will produce a family of straight lines with slope dependent on the concentration of B and that intersect to the left of the *y* axis, or in the case of a rapid equilibrium sequential mechanism, on the *y* axis itself.

To gain insight into the kinetic mechanism we measured rate data at various constant concentrations of NAD^+, and *vice versa* **(Figure 4a, d)**. The data were replotted as double reciprocal plots ($1/\upsilon_0$ versus 1/[SLA]) **(Figure 4b, e)**. These primary double reciprocal plots gave a series of intersecting straight lines, consistent with a sequential mechanism. The position of the intersection provides insight into the nature of the sequential or ordered mechanism. The plot of 1/[NAD^+] versus 1/*V* intersected close to the *y*-axis **(Figure 4b)**, while the plot of 1/[SLA] versus 1/*V* intersected to the left of the *y*-axis (**Figure 4e**). This is consistent with a rapid equilibrium ordered mechanism.[22]

Secondary plot analysis involves replotting the slope data from the primary double reciprocal plots. Thus, the slopes of each line in the double reciprocal plot were plotted versus the reciprocal concentrations of the other substrate **(Figure 4c, f)**. For the plot of slopes of the 1/[SLA] versus 1/[NAD^+] plot, the line intercepted the *y*-axis above the origin, while the plot of slopes of the 1/[NAD^+] plot versus 1/[SLA] passed through the origin (see Figure legend for a more detailed analysis of the lines of fit).[23] This data supports the conclusion of a rapid equilibrium ordered reaction, with NAD^+ binding first to enzyme.

***Rl*GabD adopts a tetrameric assembly of a classic aldehyde dehydrogenase fold**

*Rl*GabD belongs to the family of aldehyde dehydrogenase (ALDHs), which usually exist and function as homodimers and homotetramers.[14] A molar plot of *Rl*GabD (MW 52,000 Da) using size exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) analysis showed a solution-state species with molecular weight of approximately 210 kDa (**Figure S3**). *Rl*GabD thus exists as tetramer in solution, presenting it as a suitable candidate for structural characterisation using cryogenic electron microscopy (cryo-EM).

To define conditions for imaging of a possible complex, we studied the interaction of *Rl*GabD with NAD(H) using nanoscale differential scanning fluorimetry (nano-DSF). This revealed a thermal shift of 4.3 °C for NAD^+ (and for SLA+NADH, ΔTm of 2.6 °C), while $NADP^+$ produced a ΔTm of 3.5 °C (and for NADPH ΔTm of 2.5 °C) (**Figure S4**). These results guided experiments to image a binary complex. Thus, cryo-EM grids were optimised and prepared with 1 mg/mL *Rl*GabD pre-incubated with 2 mM NADH. Data collection and refinement statistics of single particle cryo-EM analysis for *Rl*GabD-NADH complex are provided in **Table S1**. A total of 865 micrographs were used for auto-picking. Particles picked from these micrographs were used to generate 2D-class averages, which displayed distinct orientations (**Figure S5**). Downstream processing and refinement with D2 symmetry gave a final 3D reconstruction of *Rl*GabD at an overall resolution of 2.52 Å at Fourier shell correlation threshold of 0.143 (**Figure 5a**, **Figures S6, S7**).

The *Rl*GabD tetramer assembles as a pair of dimers (**Figure 5a,b**). Each protomer of *Rl*GabD adopts the canonical ALDH class I/II fold with three domains.[15, 16] Each L-shaped protomer comprises an α/β N-terminal cofactor binding domain [residues 8-133, 153-263], an α/β catalytic domain containing the conserved Cys-Glu dyad [residues 264-476], and a smaller, anti-parallel β-sheet oligomerization domain [residues 134-152, 477-489], which interacts with two other subunits. The *Rl*GabD dimer is formed through domain swapping interactions of the three-stranded oligomerization domain of subunit A with the catalytic domain of partner subunit (B) forming a ten-stranded β-sheet. Oligomerization domains of subunits A+C (and B+D) form extended β-sheets stabilising the final pair-of-dimers assembly.

A binary *Rl*GabD-NADH complex reveals the active site architecture

The *Rl*GabD-NADH complex shows the binding mode and positioning of the cofactor. NADH is bound in an extended mode, as seen in *E. coli* SSADH GabD,[17] with the nicotinamide group protruding into the active site located in the cleft between the two major domains (**Figure 5c**). The nicotinamide ring was modelled in close vicinity to the catalytic dyad; however, poor density was evident for the second phosphate and of the ribosyl-nicotinamide group (**Figure S9**). The adenine ring occupies a hydrophobic pocket lined by Ala 219, Leu224, Val243, Trp246, and Leu247. 2'-OH of the adenosine of NADH forms hydrogen bonds with Lys186 (2.5 Å) and a water molecule, which

in turn engages in hydrogen bonding interactions with Ser189 (2.6 Å). The 3'-OH of adenosine is hydrogen-bonded to Lys186 (3 Å) and the backbone carbonyl of Thr160 (2.4 Å). The pyrophosphate group of NADH forms hydrogen-bonding interactions with backbone carbonyl and hydroxyl group of Ser240. Some additional density is seen at active site Cys295 possibly indicating oxidation and weaker side-chain density of Glu261 owing to radiation damage or cryo-EM density weakness of negatively charges of carboxylate groups. [18, 19]

The poor density in the core of the bound NADH may reflect multiple binding modes of the cofactor. At least two discrete conformations have been reported for the nicotinamide ring in members of ALDH class I/II families. *Rl*GabD shares high sequence and structural and functional similarities with representative ALDH members such as *E. coli* SSADH (PDB: 3JZ4, core RMSD of 0.65 Å and 61% sequence ID),[17] *E. coli* lactaldehyde DH (PDB: 2ILU, RMSD 1.3 Å and 35% sequence ID),[20] and the reduced form of human SSADH (PDB: 2W8R, RMSD 0.76 Å and 55% sequence ID).[21] Structural comparison of the *Rl*GabD-NADH complex with *E. coli* SSADH and lactaldehyde DH clearly demonstrates the two discrete 'in' and 'out' cofactor conformations (**Figure S10**). The *Rl*GabD-NADH complex displays the catalytically-relevant 'in' conformation with the nicotinamide ring pointing into the active site, and C4 of nicotinamide approx. 6.7 Å from catalytic Cys295. Further, the 2'-phosphate binding residues Ser179 and Lys182 (*E. coli* SSADH numbering) are conserved in *Rl*GabD, contributing the dual cofactor specificity [NAD(P)H] of *Rl*GabD.

To propose active site residues involved in catalysis we compared the sequence alignments of *Rl*GabD with the NADP-dependent non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase (GAPN) from *Streptococcus mutans* (**Figure S11**). Like *Rl*GabD, GAPN operates through an ordered sequential mechanism in which the cofactor binds first.[22] The catalytic mechanism for oxidation of GAP by GAPN has been described in detail and involves two main steps.[23] In the first step, nucleophilic addition of Cys302 to the aldehyde of GAP forms a hemithioacetal oxyanion, which is stabilized by an 'oxyanion hole' formed from the terminal NH_2 of Asn169 and the backbone N-H of Cys302. The hemithioacetal oxyanion is activated to transfer hydride to NADP, forming an acyl enzyme and NADPH. In the second step, Glu268 acts as general base to assist the nucleophilic addition of water to the acyl enzyme, forming a tetrahedral intermediate oxyanion, which eliminates Cys302 to give the product, 3-phosphoglycerate. All of the catalytic residues involved in GAPN catalysis are conserved with *Rl*GabD. Moreover, the 3D structure reveals that they are in an appropriate position adjacent to the nicotinamide headgroup of NADH to adopt similar roles (**Figure 5**). Namely, Cys295 is the catalytic nucleophile, Glu261 is the general base, and the oxyanion hole is formed from Asn163 and the backbone NH of Cys295.

Prediction of the SLA binding pocket in SLADH enzymes

To predict the location of the SLA binding pocket, we computed the interior cavities and channels in the *Rl*GabD-NADH tetramer using the CASTp server 3.0[24] (**Figure S11, S12**). This predicted a hydrophilic, active-site pocket with solvent-accessible area of approx. 283 $Å^2$ and volume of 184 $Å^3$ (reported as Richard's solvent accessible surface area/volume[25]), with helix α3 [residues 106-121] and surface loop [residues 449-458] forming the mouth of the opening. The base of this cavity is buried in the cleft between the two domains and immediately precedes the conserved nucleophile Cys295. Superposition of the predicted SLA binding pocket computed by CASTp with the 3D structure of human SSADH (in which the catalytic nucleophile Cys340 was converted to Ala) with bound succinic semialdehyde (SSA) (PDB 2W8Q) shows SSA occupies this cavity in an orientation where aldehyde group directly points towards the nucleophile Cys295 of *Rl*GabD, suggesting a similar orientation for sulfolactaldehyde (SLA). The carboxylate group of SSA is H-bonded to Arg213, Arg334 and Ser498 (hSSADH numbering). These residues are incompletely conserved with SLADH enzymes. *Rl*GabD and several other SLADH candidates including *P. putida*, *Arthrobacter* sp., and *Desulfovibrio* sp. contain Arg residues at equivalent positions (*Rl*GabD: Arg171 and Arg289), but *B. megaterium* and *B. urumqiensis* have Arg213 is replaced by His, and Arg334 by Asn. The multiple sequence alignment at hSSADH position Ser498 is poorly aligned with insertions/deletions, and no clear consensus. We propose Arg171/Arg289 as sulfonate binding residues in *Rl*GabD and some SLADH enzymes, with the equivalent positions as His/Asn in other SLADH enzymes fulfilling a similar role. To explore whether other residues are associated with the Arg171/Arg289 pair in *Rl*GabD, and the His164/Asn280 pair in *B. megaterium* SlaB we conducted coevolution analysis using CoeViz.[26] Using the multiple sequence alignment of 158 putative SLADH enzymes (*vide infra*) we identified a clique of 16 residues (including the above pairs) that independently co-evolve with the above pairs (**Figure S13**). Mapping of these coevolving cliques onto the cryo-EM structure of *Rhizobium leguminosarum* GabD and the AlphaFold2[27, 28] predicted structure of *Bacillus megaterium* SlaB identified additional partially conserved residues in proximity to the proposed SLA binding pocket that contain the oxyanion hole stabilizing residue, namely Trp162-Asn163-Phe164 in *Rl*GabD and Phe155-Asn156-Val157 in *B. megaterium* SlaB.

Human SSADH (hSSADH) shares a similar fold and the conserved catalytic residues with *E. coli* GabD SSADH and *Rl*GabD SLADH, including the catalytic cysteine (Cys340 in hSSADH). hSSADH contains a second cysteine (C342) two residues downstream in a redox active mobile loop that can engage in a disulfide bond with the nucleophilic cysteine. Oxidation to the disulfide results in hSSADH adopting an 'open' conformation, while reduction to the cysteine causes loop movement and a 'closed conformation' (**Figure S14**). The second cysteine residue is not conserved in *E. coli* SSADH nor some SLADH enzymes (e.g. *Arthrobacter* spp, *B. urumqiensis*, and *B. megaterium*), but is present in *Rl*GabD and SLADHs from *Desulfovibrio* sp., and *P. putida* (**Figure S15**). In the

*Rl*GabD-NADH structure, the catalytic loop of *Rl*GabD adopts the 'open' conformation, with the two cysteine residues 8.4 Å apart and the catalytic dyad Cys295/Glu262) poised for catalysis (**Figure S14**). It is unknown whether Cys295/297 in bacterial SLADH proteins undergo comparable oxidation and associated loop movement as seen for hSSADH.

Sequence similarity network analysis reveals the taxonomic range and functional distribution of SLA dehydrogenases in the pathways of sulfoglycolysis

SLA dehydrogenases occur in several sulfoglycolytic pathways: sulfo-ED, sulfo-SFT, and sulfo-EMP pathways, as well as within DHPS degradation pathways. To explore the distribution and evolution of SLA reductases we performed sequence similarity network (SSN) analysis[29] using the EFI enzyme similarity tool (EFI-EST).[30, 31] Using the individual SLA dehydrogenase sequences from seven experimentally-verified sulfoglycolytic organisms (*B. urumquiensis*, *P. putida*, *R. leguminosarum*, *H. seropedicae*, *B. megaterium*, *Arthrobacter* sp. AK01, *B. aryabhattai*) and one DHPS degrading organism (*Desulfovibrio* sp. DF1) we separately conducted BLASTp searches and combined the results to obtain a total of 158 sequences with >37-64% sequence identity to the search queries. To visualize and study the distribution of these sequences, we used SSNs. Initially, we explored the construction of SSNs using different alignment scores (**Figure S16**). At alignment score 75, the sequences form a single cluster, while in the range 125-150, the SSN breaks into three clusters that almost perfectly separate the three main Phyla: Actinobacteria, Firmicutes and Proteobacteria, with the three Chlorflexi members, the sole Candidatus Dormibacteraeota member and one spurious Firmicutes member clustering with Actinobacteria. At even higher alignment threshold (175), the Chloroflexi members segregate, but the Actinobacteria fragment and the SSN spawns many singletons that limit its utility. The SSN generated at alignment score 150 (corresponding to minimum identity >53%; **Figure 6a**) was coloured based on the pathway encoded by the proposed function of the gene cluster in which the SLA dehydrogenase gene was located (**Figure 6b**). The sulfo-EMP pathway organisms are limited to Actinobacteria, sulfo-EMP2 pathway organisms are mainly limited to Firmicutes but with several members within the Actinobacteria, while hybrid sulfo-EMP pathways comprised of various combinations of genes from the sulfo-EMP and sulfo-EMP2 pathways are more broadly distributed across Actinobacteria, Firmicutes, and Proteobacteria. Sulfo-ED organisms occur mainly within proteobacteria but with several members within the Actinobacteria. Sulfo-SFT organisms are mainly Firmicutes but with two Chloroflexi and one Candidatus Dormibacteraeota members. Finally, DHPS degradation pathways are mainly limited to Actinobacteria with a sole Proteobacteria representative. We highlight the Proteobacteria member *Ensifer* sp. HO-A22 contains both sulfo-ED and DHPS degrading gene clusters, suggesting that this organism achieves the complete biomineralization of SQ to sulfite. A tree showing the phylogenetic relationships based on

16S RNA sequences between bacteria that contain SLADH genes within sulfoglycolytic gene clusters is shown in **Figure S17**.

We used the sequences from the SSN to identify the genes that flank the 158 SLADH genes in the genomes of the host organisms. Using the EFI-EST tools, we identified 1287 gene neighbours located ± 10 ORF from the query SLADH sequences. These were analysed by creation of a sequence similarity network of neighbors (SSNN) into isofunctional proteins that were assigned a function based on manual inspection (**Figure S18**). To organize and visualize the sulfoquinovose and DHPS degrading gene clusters we constructed a genome neighborhood similarity network (GNSN) using the EFI-GNT tool (**Figure 6c**). In this network each node corresponds to a single SLA dehydrogenase protein that is connected by an edge to another SLA dehydrogenase if they share >4 isofunctional genes in their genome neighborhood. The GNSN shows that SQase proteins (coloured blue) are encoded in the gene clusters for most sulfoglycolytic organisms, with the exception of some sulfo-EMP organisms, consistent with the role of SQases as a gateway to sulfoglycolysis through cleavage of SQ-glycosides (**Figure 6c(i)**).[32, 33] Characteristic enzymes encoded by sulfo-EMP (SF kinase YihV, yellow), sulfo-EMP2 (SF kinase SqiK, aquamarine), sulfo-ED (SG dehydratase, purple), sulfo-SFT (SF transaldolase, dark blue) and DHPS degradation (DHPS dehydrogenase, orange) pathways distribute across the GNSN into clusters (**Figures 6c(ii-v)**). The sulfoglycolytic clusters are mutually exclusive to the DHPS degrading clusters, except for *Ensifer* sp. HO-A22, which occurs within the main sulfo-ED cluster. When the GNSN was coloured for the five phyla identified in the SSN (Actinobacteria, Firmicutes, Proteobacteria, Chloroflexi, and Candidatus Dormibacteraeota) (**Figure 6c(vi)**), we observed coloured clusters that recapitulated the taxonomic clustering of the SSN of SLADH sequences in **Figure 6a**.

Conclusions

SLADH enzymes catalyze the oxidation of SLA to SL, the final step of sulfoglycolytic pathways that lead to excretion of SL,[4] and the second step in the oxidation of DHPS to SL in *Desulfovibrio* sp.,[5] which activates this substrate for sulfur-carbon bond scission in the DHPS degradation pathway to produce sulfite and pyruvate. Our data demonstrates that *Rl*GabD has dual cofactor activity, and a 30-fold preference for oxidation of SLA versus the structurally-related phosphate analogue glyceraldehyde phosphate, a key intermediate in glycolysis/gluconeogenesis. The weak activity on GAP will lead to production of 3-phospho-D-glycerate, an intermediate in glycolysis, and thus this low-level activity likely has little consequence for cellular metabolism. Nonetheless, the activity of SLA dehydrogenase on GAP stands in contrast to SLA reductase, which had no detectable activity on GAP.[7]

Similar to the well-characterized GAP dehydrogenase from *S. mutans*,[22] *Rl*GabD uses a rapid

equilibrium ordered mechanism, in which $NAD(P)^+$ is the first substrate to bind. Knowledge of reaction order, a large change in protein melting temperature upon binding NADH, and the identification of a tetramer in the solution state, guided our approach to determining the 3D structure of the *Rl*GabD-NADH complex using cryo-EM. This complex revealed sequence and spatial conservation of amino acid residues involved in catalysis, and allows a proposal of a mechanism for catalysis. Binding of NAD+, and then SLA gives the Michaelis complex. In the first step, nucleophilic addition of Cys295 to the aldehyde of SLA forms a hemithioacetal oxyanion, stabilized by an 'oxyanion hole' formed from the terminal NH_2 of Asn163 and the backbone N-H of Cys295. The hemithioacetal oxyanion is activated to transfer hydride to NADP, forming an acyl enzyme and NADPH. In the second step, Glu261 provides general base catalysis, assisting the nucleophilic addition of water to the acyl enzyme, forming a tetrahedral intermediate oxyanion, which eliminates Cys295 to give SL. Based on their proximity to the active site, we propose that Ser117-Arg171-Arg289 comprise the SLA sulfonate binding residues in *Rl*GabD. Arginine residues are common in a wide range of other sulfonate binding proteins and enzymes from various sulfoglycolytic pathways.

Because SL as an endproduct of sulfoglycolysis, a nutrient for SL degrading bacteria, and an intermediate in DHPS degradation, the oxidation of SLA to SL catalyzed by SLADH is an important step in the breakdown of the C6-organosulfonate sulfoquinovose and the C3-organosulfonate DHPS. Sulfoglycolytic gene clusters containing genes encoding SLADH enzymes are distributed across Actinobacteria, Firmicutes, Proteobacteria, Chloroflexi, and Candidatus Dormibacteraeota, while DHPS degradation gene clusters containing SLADH homologues are limited to Proteobacteria and Actinobacteria. The present work provides a structural and biochemical view of SLADH enzymes that complements our knowledge of SLA reductases, and enriches our understanding of a critical step in the organosulfur cycle.

Experimental section

Heterologous expression and protein purification

The gene sequence coding for *Rhizobium leguminosarum* SRDI565 GabD (NCBI locus tag WP_017967313.1) was synthesised commercially with codon harmonisation for expression in *E. coli* within a pET-29a(+) vector with C-terminal His-tag (GenScript). The plasmid containing the gene for target enzyme was used to transform *E. coli* BL21(DE3) competent cells for gene expression. Starter cultures were grown in LB-medium (5 mL) containing 100 µg mL^{-1} kanamycin for 18 h at 37°C with shaking at 200 r.p.m. 1 L volume cultures were inoculated with the starter culture (5 mL) and incubated at 37 °C, with shaking at 200 r.p.m. until an OD_{600} of 0.6-0.8 was reached. Gene expression was induced by addition of IPTG (0.5-1 mM) and shaking was continued overnight at 18 °C with shaking at 200 r.p.m. The cells were harvested by centrifugation at 5000 *g* for 20 min and

resuspended in 50 mM sodium phosphate buffer pH 7.4, containing 500 mM NaCl and 30 mM imidazole. Cells were disrupted by ultrasonication for 3 × 5 min, 30 s on; 30 s off cycles, and the suspension was centrifuged at 50,000 *g* for 30 min to yield a clear lysate. The C-terminal His_6-tagged protein was purified using immobilized-metal affinity chromatography (IMAC) using a Ni-NTA column, followed by size exclusion chromatography (SEC) (Supplementary Figure 1). For IMAC, the lysate was loaded onto a pre-equilibrated Ni-NTA column, followed by washing with a load buffer (50 mM Tris, 500 mM NaCl, 30 mM imidazole pH 7.5). The bound protein was eluted using a linear gradient with buffer containing 300 mM imidazole. *Rl*GabD fractions were pooled, concentrated and loaded onto a HiLoad 16/600 Superdex 75 gel filtration column pre-equilibrated with 50 mM Tris, 300 mM NaCl pH 7.5 buffer. The protein was concentrated using a Vivaspin® 6 with a 30 kDa MW cut-off membrane, to a final concentration of 8 mg mL^{-1} for structural studies.

SEC MALLS analysis

Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu 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±2 °C). Solvent was 0.2 µm filtered before use and a further 0.1 µm filter was present in the flow path. The column was equilibrated with at least 2 column volumes of buffer (50 mM NaPi, 300 mM NaCl pH 7.4) before use and flow was continued at the working flow rate until baselines for UV, light scattering and refractive index detectors were stable. Sample injection volume was 100 µL *Rl*GabD at 6 mg mL^{-1} in 50 mM Tris buffer, 300 mM NaCl pH 7.4 containing 2 mM NADH; Shimadzu LC Solutions software was used to control the HPLC 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 synchronisation. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data were analysed using the Astra V software. MWs were estimated using the Zimm fit method with degree 1. A value of 0.182 was used for protein refractive index increment (dn/dc).

Nanoscale Differential Scanning Fluorimetry (nanoDSF)

NanoDSF studies were performed on a Prometheus NT.48 (NanoTemper). Data recording and initial analysis was performed with PR.ThermControl software. All protein samples were at 2 mg.ml-1 in 50 mM Tris, 150 mM NaCl at pH 7.4, with a 15 µl capillary load per sample. Temperature was ramped from 15 °C to 95 °C, at 1.0 °C/min with 10% excitation power. Experiments were performed in duplicate.

Synthesis of SLA and determination of concentration

SLA was synthesized as a solution in water as reported, at a nominal concentration of 109 mM.[17] The concentration of the SLA solution was measured by quantitative ^{1}H NMR spectroscopy using a calibration curve of methyl β-glucoside to determine the instrument sensitivity. The calibration curve was generated by measuring the absolute integration of H1 (δ 4.25 ppm in D_2O) of methyl β-glucoside at 100, 50, and 10 mM. The absolute integration of H1 was 15738, 7753.06, 1249.15, respectively **(Figure S1A)**. The SLA stock solution was diluted 1:1 with D_2O. The absolute integration of H2 (δ 3.79 ppm in D_2O) of SLA in the ^{1}H NMR spectrum was 8647.6. The calculated concentration of SLA was 110 mM, which was used for all further calculations.

Measurement of consumption of SLA by *Rl*GabD

The production of NADH from oxidation of SLA catalyzed by *Rl*GabD was monitored using UV-Vis spectrophotometer at 340 nm. The reaction was carried out in 30 mM Tris buffer, 30 mM KCl pH 7 at 30 °C [NAD^+] = 1.5 mM, [DL-SLA] = 1 mM, and [*Rl*GabD] = 15.8 nM. Phosphate buffer was not used as it has been shown to have a negative impact on NADH stability.[24] The reaction appeared complete after 40 min. More *Rl*GabD was added to a final [*Rl*GabD] = 205 nM, and the reaction was monitored for 100 min with no further increase in absorbance. Using an extinction coefficient for NADH of 6363 M^{-1} cm^{-1}, the max absorbance is 3.081, allowing calculation of 48.4% SLA consumed **(Figure S1B)**.

Michaelis-Menten kinetic analysis of *Rl*GabD

Michaelis-Menten kinetic analysis were performed for SLA, GAP, NAD^+ and $NADP^+$ under pseudo first-order conditions in which the concentration of one substrate was varied while the other was held constant, and *vice versa*. The production of NADH/NADPH from oxidation of SLA catalyzed by *Rl*GabD SLA dehydrogenase was monitored using a UV-Vis spectrophotometer at 340 nm. For the Michaelis-Menten kinetics of NAD^+/$NADP^+$, reactions were conducted in 0.1% BSA 30 mM Tris buffer pH 7.12 at 30 °C with 3 nM *Rl*GabD, constant 0.25 mM D-SLA and varying concentrations of NAD^+/$NADP^+$ (0.05-1 mM). For the Michaelis-Menten kinetics of SLA, constant 0.25 mM NAD^+/$NADP^+$ and varying concentrations of SLA (0.05-0.5 mM) were used **(Figure 2b, c)**. For the Michaelis-Menten kinetics of racemic GAP, constant 0.25 mM NAD^+/$NADP^+$ and varying concentrations of GAP (0.05-0.25 mM) were used **(Figure 2e, f)**. The apparent kinetic parameters, k_{cat}, K_M, and k_{cat}/K_M were calculated using the Prism 9 software package (GraphPad Scientific Software) **(Table 1)**. No reaction was observed for >1 mM [SLA] and >0.25 mM [$NADP^+$]. This appears to be a result of inhibition.

Cryo-EM 3D-structure of *Rl*GabD-NADH

A frozen aliquot of *Rl*GabD was prepared at 4 mg/mL and complexed with 2 mM NADH for screening. R1.2/1.3 300 mesh UltrAuFoil gold grids (Quantifoil) were glow-discharged for 3 min at 20 mAmp/0.38 mBar. A total of 2.5 µL of this sample was applied to glow-discharged UltrAuFoil gold grids, which were subsequently blotted for 2 s with a blot force of 10, then were plunged into liquid ethane cooled by liquid nitrogen. Plunge-freezing was performed using a Vitrobot Mark IV (Thermo Fisher Scientific) at 100% humidity and 22 °C. EER formatted movies of *Rl*GabD-NADH complex were acquired on the Glacios microscope (Thermo Fisher Scientific), housed in the York Structural Biology Laboratory. The microscope was operated at an accelerating voltage of 200 kV with a Falcon 4 direct electron detector. The *Rl*GabD-NADH dataset was acquired at a dose rate of 2.98 electrons per pixel per second, and a pixel size of 0.574 Å; target defocus values were -2 to -0.7 µm. The autofocus function was run every 10 µm, and the dataset was collected with a total dose of 50 electrons per $Å^2$.

Image processing and 3D reconstruction

Movie frames of the *Rl*GabD-NADH were motion corrected without binning, using a pixel size of 0.574 Å, and dose-weighted using the Motioncorr2 program.[34] Contrast transfer function (CTF) corrections were performed using CTFFIND 4.1.[35] Most of the subsequent processing steps were carried out using RELION 3.[36] Laplacian-of Gaussian (LoG) based automated particle picking was performed on the data. Particles were extracted and subjected to 2D classification. 2D classes showing sharp structural features were chosen to build an initial 3D model. This initial model was used for template-based picking of particles (low pass-filtered at 20 Å). Picked particles were 2D classified to remove poor particles. Particles were 3D classified, without the use of symmetry constraints. The class showing well-defined structural features was then selected for 3D refinement with D2 symmetry imposed. This gave a reconstruction with a resolution of 3.37Å (FSC threshold of 0.143). Subsequently, CTF refinement[37] was performed for magnification anisotropy; optical aberrations (up to the fourth order); and per-particle defocus and per-micrograph astigmatism. Bayesian Polishing[38] was also used to optimise per-particle beam-induced motion tracks, followed by another round of auto-refinement. CTF refinement, Bayesian Polishing and 3D refinement steps were repeated to yield a 2.52 Å map (FSC threshold of 0.143). Further details of the image processing and 3D reconstruction can be found in Table S1 and Figure S6.

Model building, refinement, and validation

An initial model was built using the map_to_model function in Phenix.[39] The initial model was modified using *Coot*.[40] Cycles of real space refinement in *Phenix* and interactive model building in

Coot were used to fit the structure to the map.[39] Model geometries were assessed with MolProbity.[41] Structures and maps in the figures were rendered with ccp4mg,[42] PyMOL (http://www.pymol.org/) or ChimeraX.[43]

Sequence similarity network analysis

Sequence Similarity Network (SSN) analysis was carried out using the web-based Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) (https://efi.igb.illinois.edu/efi-est/),[44] and Genome neighborhood analysis was carried out using the web-based EFI Genome Neighborhood Tool (EFI-GNT; (https://efi.igb.illinois.edu/efi-gnt/).[45] All networks were visualized using Cytoscape v3.9.[46] The SSN created using the alignment score 150 was used to generate genome neighbourhood diagrams (GND) with open reading frame (ORF) ± 10 neighbours using the EFI-GNT. A script was used to extract the accession codes of the retrieved neighbours. The SLA dehydrogenase neighbours were used to generate an SSN of neighbours (SSNN; **Figure S16**). The resulting SSNN was plotted using Cytoscape v3.9, with minimum alignment score of 50. We used the SSNN to build a genome neighborhood similarity network (GNSN) in which each node corresponded to a single SLA dehydrogenase that was connected by an edge to another SLA dehydrogenase if they shared >4 isofunctional genes in their genome neighborhood. This condition results in clustering of nodes involved in similar SQ degradation pathways. To annotate the GNSN we coloured the nodes according to whether the SLA dehydrogenase gene possessed specific neighboring genes encoding SQ degradation pathway enzymes in the genome neighborhood, namely the gateway enzyme SQase (PF01055) for the sulfo-EMP (SF kinase YihV), sulfo-EMP2 (SF kinase SqiK, PF00365), sulfo-ED (SG dehydratase, PF00920), sulfo-SFT (SF transaldolase, PF00923), and DHPS dehydrogenase (DHPS degradation pathway, PF03446), as well as manual inspection. A separate GNSN was coloured according to phylum of the host bacterium.

Associated content

Supporting information

The supporting information is available free of charge at https://pubs.acs.org/doi/

Additional details on enzyme kinetics (Figures S1-2), protein biochemistry (Figures S3-S4), Cryo-EM (Figures S5-S9), 3D structural data (Figures S10-S12), bioinformatics (Figures S13-S18), and structural statistics (Table S1)

Author contributions

Commented [SW1]: Where does ZA belong in this list?

G.J.D. and S.J.W. conceived the project. N.M.S. and M.S. performed molecular biology and protein

purification. M.S., Z.A., and R.M. prepared the cryo-EM grids and performed the image acquisition and processing. J.L. and A.A. synthesized SLA and inhibitors, and conducted enzyme kinetics. J.N.B. helped with the image acquisition, processing, and model building. J.L., M.S., A.A., E.D.G.-B., J.N.B., G.J.D. and S.J.W. analysed the data and wrote the manuscript.

Notes

The authors declare no competing financial interest.

Acknowledgements

This work was supported by the Australian Research Council (DP210100233, DP210100235)), the Biotechnology and Biological Sciences Research Council (BB/W003805/1), the UKRI Future Leader Fellowship Program (MR/T040742/1), and the Royal Society for the Ken Murray Research Professorship to G.J. D and the associated PDRA funding (RP\EA\180016) for RWM. We thank the Wellcome Trust for funding the Glacios electron microscope (grant number 206161/Z/17/Z) and Dr Johan Turkenburg and Sam Hart for assistance with cryo-EM data collection. We also acknowledge Dr. Andrew Leech at the University of York Bioscience Technology Facility for assistance with SEC-MALLS analysis.

Abbreviations

DHPS, 2,3-dihydroxypropanesulfonate

SL, sulfolactate

SQ, sulfoquinovose

References

(1) *CRC Handbook of Chemistry and Physics*; CRC Press, 2017.

(2) Cook, A. M.; Denger, K.; Smits, T. H. Dissimilation of C3-sulfonates. *Arch. Microbiol.* **2006**, *185* (2), 83-90. DOI: 10.1007/s00203-005-0069-1 From NLM.

(3) Goddard-Borger, E. D.; Williams, S. J. Sulfoquinovose in the biosphere: occurrence, metabolism and functions. *Biochem. J.* **2017**, *474*, 827–849. DOI: 10.1042/BCJ20160508.

(4) Snow, A. J. D.; Burchill, L.; Sharma, M.; Davies, G. J.; Williams, S. J. Sulfoglycolysis: catabolic pathways for metabolism of sulfoquinovose. *Chem. Soc. Rev.* **2021**, *50*, 13628–13645 DOI: 10.1039/D1CS00846C.

(5) Burrichter, A.; Denger, K.; Franchini, P.; Huhn, T.; Müller, N.; Spiteller, D.; Schleheck, D. 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.* **2018**, *9* (2792), Original Research. DOI: 10.3389/fmicb.2018.02792.

(6) Denger, K.; Weiss, M.; Felux, A. K.; Schneider, A.; Mayer, C.; Spiteller, D.; Huhn, T.; Cook, A. M.; Schleheck, D. Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **2014**, *507* (7490), 114-117. DOI: 10.1038/nature12947 From NLM.

(7) Sharma, M.; Abayakoon, P.; Lingford, J. P.; Epa, R.; John, A.; Jin, Y.; Goddard-Borger, E. D.; Davies, G. J.; Williams, S. J. Dynamic Structural Changes Accompany the Production of Dihydroxypropanesulfonate by Sulfolactaldehyde Reductase. *ACS Catalysis* **2020**, *10* (4), 2826-2836. DOI: 10.1021/acscatal.9b04427.

(8) Liu, J.; Wei, Y.; Ma, K.; An, J.; Liu, X.; Liu, Y.; Ang, E. L.; Zhao, H.; Zhang, Y. Mechanistically Diverse Pathways for Sulfoquinovose Degradation in Bacteria. *ACS Catalysis* **2021**, 14740-14750. DOI: 10.1021/acscatal.1c04321.

(9) Felux, A. K.; Spiteller, D.; Klebensberger, J.; Schleheck, D. Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proc. Natl. Acad. Sci. USA* **2015**, *112* (31), E4298-4305. DOI: 10.1073/pnas.1507049112 From NLM.
(10) Frommeyer, B.; Fiedler, A. W.; Oehler, S. R.; Hanson, B. T.; Loy, A.; Franchini, P.; Spiteller, D.; Schleheck, D. Environmental and Intestinal Phylum Firmicutes Bacteria Metabolize the Plant Sugar Sulfoquinovose via a 6-Deoxy-6-sulfofructose Transaldolase Pathway. *iScience* **2020**, *23* (9), 101510. DOI: 10.1016/j.isci.2020.101510 From NLM.
(11) Liu, Y.; Wei, Y.; Zhou, Y.; Ang, E. L.; Zhao, H.; Zhang, Y. A transaldolase-dependent sulfoglycolysis pathway in *Bacillus megaterium* DSM 1804. *Biochem. Biophys. Res. Commun.* **2020**, *533*, 1109-1114. DOI: https://doi.org/10.1016/j.bbrc.2020.09.124.
(12) Denger, K.; Mayer, J.; Buhmann, M.; Weinitschke, S.; Smits, T. H.; Cook, A. M. Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (*S*)-cysteate sulfolyase. *J. Bacteriol.* **2009**, *191* (18), 5648-5656. DOI: 10.1128/jb.00569-09 From NLM.
(13) Rein, U.; Gueta, R.; Denger, K.; Ruff, J.; Hollemeyer, K.; Cook, A. M. Dissimilation of cysteate via 3-sulfolactate sulfo-lyase and a sulfate exporter in *Paracoccus pantotrophus* NKNCYSA. *Microbiology (Reading, England)* **2005**, *151* (Pt 3), 737-747. DOI: 10.1099/mic.0.27548-0 From NLM.
(14) Shortall, K.; Djeghader, A.; Magner, E.; Soulimane, T. Insights into aldehyde dehydrogenase enzymes: a structural perspective. *Front. Mol. Biosci.* **2021**, *8*, 659550. DOI: 10.3389/fmolb.2021.659550 From NLM.
(15) Moore, S. A.; Baker, H. M.; Blythe, T. J.; Kitson, K. E.; Kitson, T. M.; Baker, E. N. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* **1998**, *6* (12), 1541-1551. DOI: 10.1016/S0969-2126(98)00152-X (acccessed 2023/02/17).
(16) Steinmetz, C. G.; Xie, P.; Weiner, H.; Hurley, T. D. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* **1997**, *5* (5), 701-711. DOI: 10.1016/S0969-2126(97)00224-4 (acccessed 2023/02/17).
(17) Langendorf, C. G.; Key, T. L. G.; Fenalti, G.; Kan, W.-T.; Buckle, A. M.; Caradoc-Davies, T.; Tuck, K. L.; Law, R. H. P.; Whisstock, J. C. The X-ray crystal structure of *Escherichia coli* succinic semialdehyde dehydrogenase; Structural insights into $NADP^+$/enzyme interactions. *PLOS ONE* **2010**, *5* (2), e9280. DOI: 10.1371/journal.pone.0009280.
(18) Wang, J. On the appearance of carboxylates in electrostatic potential maps. *Protein Sci.* **2017**, *26* (3), 396-402, https://doi.org/10.1002/pro.3093. DOI: https://doi.org/10.1002/pro.3093 (acccessed 2023/03/09).
(19) Beckers, M.; Mann, D.; Sachse, C. Structural interpretation of cryo-EM image reconstructions. *Prog. Biophys. Mol. Biol.* **2021**, *160*, 26-36. DOI: https://doi.org/10.1016/j.pbiomolbio.2020.07.004.
(20) Di Costanzo, L.; Gomez, G. A.; Christianson, D. W. Crystal structure of lactaldehyde dehydrogenase from *Escherichia coli* and inferences regarding substrate and cofactor specificity. *J. Mol. Biol.* **2007**, *366* (2), 481-493. DOI: 10.1016/j.jmb.2006.11.023 From NLM.
(21) Kim, Y. G.; Lee, S.; Kwon, O. S.; Park, S. Y.; Lee, S. J.; Park, B. J.; Kim, K. J. Redox-switch modulation of human SSADH by dynamic catalytic loop. *EMBO J.* **2009**, *28* (7), 959-968. DOI: 10.1038/emboj.2009.40 From NLM.
(22) Marchal, S.; Rahuel-Clermont, S.; Branlant, G. Role of glutamate-268 in the catalytic mechanism of nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*. *Biochemistry* **2000**, *39* (12), 3327-3335. DOI: 10.1021/bi9914208 From NLM.
(23) D'Ambrosio, K.; Pailot, A.; Talfournier, F.; Didierjean, C.; Benedetti, E.; Aubry, A.; Branlant, G.; Corbier, C. The first crystal structure of a thioacylenzyme intermediate in the ALDH family: new coenzyme conformation and relevance to catalysis. *Biochemistry* **2006**, *45* (9), 2978-2986. DOI: 10.1021/bi0515117 From NLM.
(24) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W363-W367. DOI: 10.1093/nar/gky473 (acccessed 2/17/2023).
(25) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55* (3), 379-400. DOI: 10.1016/0022-2836(71)90324-x From NLM.
(26) Baker, F. N.; Porollo, A. CoeViz: a web-based tool for coevolution analysis of protein residues. *BMC Bioinformatics* **2016**, *17* (1), 119. DOI: 10.1186/s12859-016-0975-z.
(27) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2 From NLM.
(28) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-

sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50* (D1), D439-D444. DOI: 10.1093/nar/gkab1061 (acccessed 1/28/2023).
(29) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* **2009**, *4* (2), e4345. DOI: 10.1371/journal.pone.0004345 From NLM.
(30) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **2015**, *1854* (8), 1019-1037. DOI: 10.1016/j.bbapap.2015.04.015.
(31) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **2019**, *58* (41), 4169-4182. DOI: 10.1021/acs.biochem.9b00735.
(32) 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.; et al. Structural and Biochemical Insights into the Function and Evolution of Sulfoquinovosidases. *ACS Cent. Sci.* **2018**, *4* (9), 1266-1273. DOI: 10.1021/acscentsci.8b00453 From NLM.
(33) 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.* **2016**, *12*, 215-217. DOI: 10.1038/nchembio.2023.
(34) Zheng, S. Q.; Palovcak, E.; Armache, J.-P.; Verba, K. A.; Cheng, Y.; Agard, D. A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **2017**, *14* (4), 331-332. DOI: 10.1038/nmeth.4193.
(35) Rohou, A.; Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **2015**, *192* (2), 216-221. DOI: https://doi.org/10.1016/j.jsb.2015.08.008.
(36) Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180* (3), 519-530. DOI: https://doi.org/10.1016/j.jsb.2012.09.006.
(37) Zivanov, J.; Nakane, T.; Forsberg, B. O.; Kimanius, D.; Hagen, W. J.; Lindahl, E.; Scheres, S. H. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **2018**, *7*. DOI: 10.7554/eLife.42166 From NLM.
(38) Zivanov, J.; Nakane, T.; Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **2019**, *6* (Pt 1), 5-17. DOI: 10.1107/s205225251801463x From NLM.
(39) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **2019**, *75* (Pt 10), 861-877. DOI: 10.1107/s2059798319011471 From NLM.
(40) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **2010**, *66* (Pt 4), 486-501. DOI: 10.1107/s0907444910007493.
(41) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27* (1), 293-315, https://doi.org/10.1002/pro.3330. DOI: https://doi.org/10.1002/pro.3330 (acccessed 2023/02/09).
(42) McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **2011**, *67* (4), 386-394. DOI: doi:10.1107/S0907444911007281.
(43) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30* (1), 70-82. DOI: 10.1002/pro.3943 From NLM.
(44) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **2015**, *1854* (8), 1019-1037. DOI: 10.1016/j.bbapap.2015.04.015 From NLM.
(45) Zhao, S.; Sakai, A.; Zhang, X.; Vetting, M. W.; Kumar, R.; Hillerich, B.; San Francisco, B.; Solbiati, J.; Steves, A.; Brown, S.; et al. Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *eLife* **2014**, *3*, e03275. DOI: 10.7554/eLife.03275.
(46) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498-2504, 10.1101/gr.1239303.

Table 1. Apparent first order kinetic parameters determined for D-SLA, NAD^+, $NADP^+$ and GAP.

Variable substrate	K_M **(mM)**	k_{cat} **(s^{-1})**	k_{cat}/K_M **(mM^{-1} s^{-1})**
[a] NAD^+	0.081±0.019	17.7±3.2	210±63
[a] $NADP^+$	0.017±0.0041	4.06±0.78	238±72.9
[b] D-SLA	0.13±0.045	17.8±7.0	137±71.8
[c] D-SLA	0.16±0.035	4.76±1.1	30±9.41
[b] GAP	0.29±0.031	0.73±0.20	4.19±1.22
[c] GAP	0.17±0.028	3.02±0.43	10.48±2.28

[a] [NAD^+]/ [$NADP^+$] was varied, while [D-SLA] was held constant at 0.25 mM.

[b] [D-SLA]/ [GAP] was varied, while [NAD^+] was held constant at 0.25 mM.

[c] [D-SLA]/ [GAP] was varied, while [$NADP^+$] was held constant at 0.25 mM.

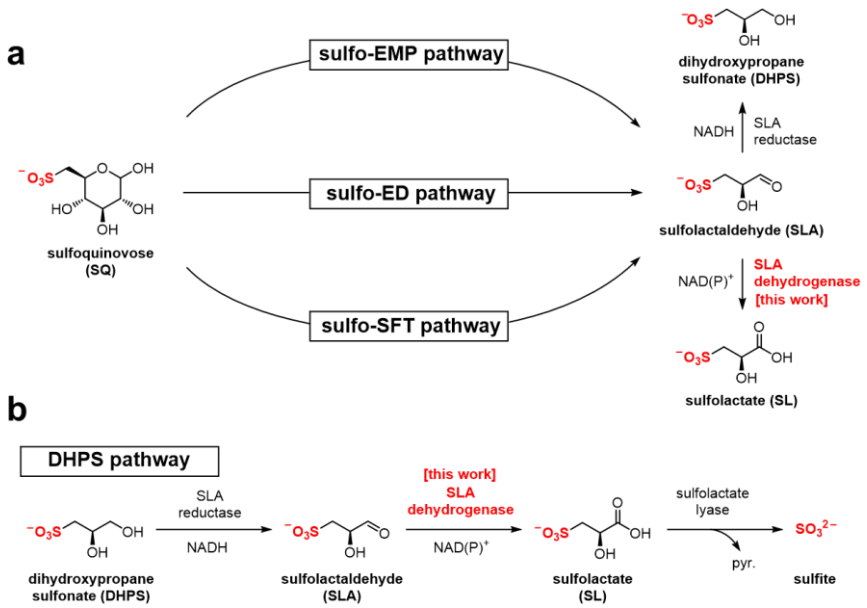

Figure 1. (a) Formation of SL and DHPS through the pathways of sulfoglycolysis from sulfoquinovose (SQ). (b) Formation and degradation of sulfolactate by catabolism of DHPS. (c) (a) Gene clusters containing SLA dehydrogenase (SLADH) from bacteria that degrade SQ (through sulfo-EMP, sulfo-EMP2, sulfo-ED, and sulfo-SFT pathways) and DHPS (through DHPS degradation pathway).

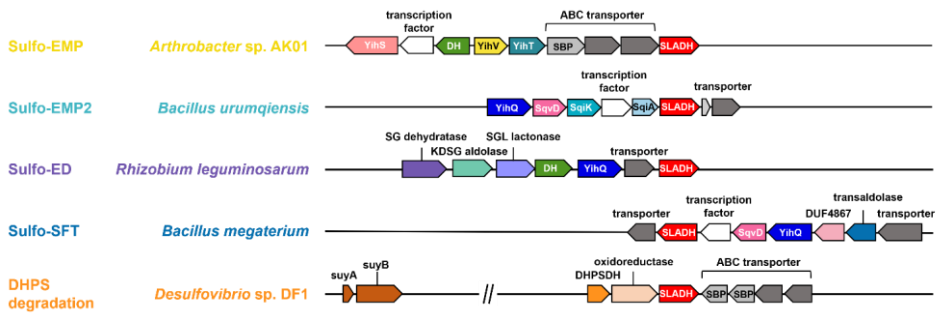

Figure 2. (a) Gene clusters of bacteria containing SLA dehydrogenase (SLADH) genes that degrade SQ (through sulfo-EMP, sulfo-EMP2, sulfo-ED, and sulfo-SFT pathways) and DHPS (through DHPS degradation pathway).

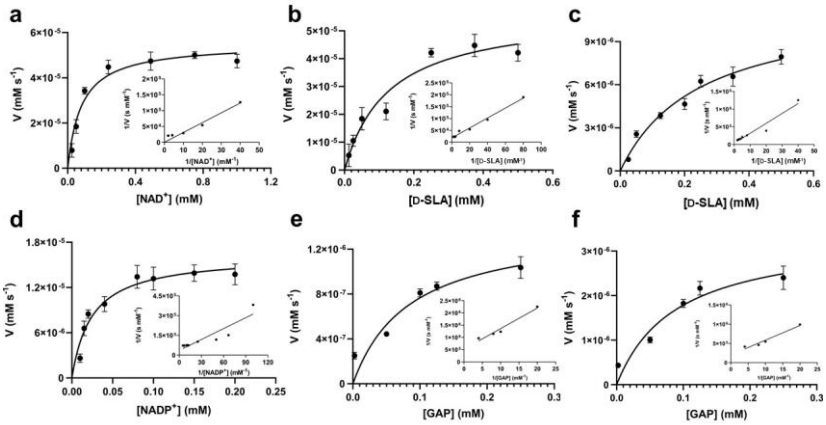

Figure 3. Michaelis-Menten kinetic analysis for SLA dehydrogenase using NAD^+, D-SLA, $NADP^+$ and GAP as substrates. (a, b) Michaelis-Menten and Lineweaver-Burk (inset) plots for oxidation of SLA by *Rl*GabD under pseudo first-order conditions of [D-SLA] = 0.25 mM (for panel a) and [NAD^+] = 0.25 mM (for panel b). (c, d) Michaelis-Menten and Lineweaver-Burk (inset) plots for oxidation of SLA by *Rl*GabD under pseudo first-order conditions of [D-SLA] = 0.25 mM (for panel c) and [$NADP^+$] = 0.25 mM (for panel d). For panel (d), dotted line is calculated by least squares linear regression through all data, solid line is fit through data for [SLA] = 0.05-0.25 mM. (e, f) Michaelis-Menten and Lineweaver-Burk (inset) plots for oxidation of GAP by *Rl*GabD under pseudo first-order conditions of [NAD^+] = 0.25 mM (for panel e) and [$NADP^+$] = 0.25 mM (for panel f).

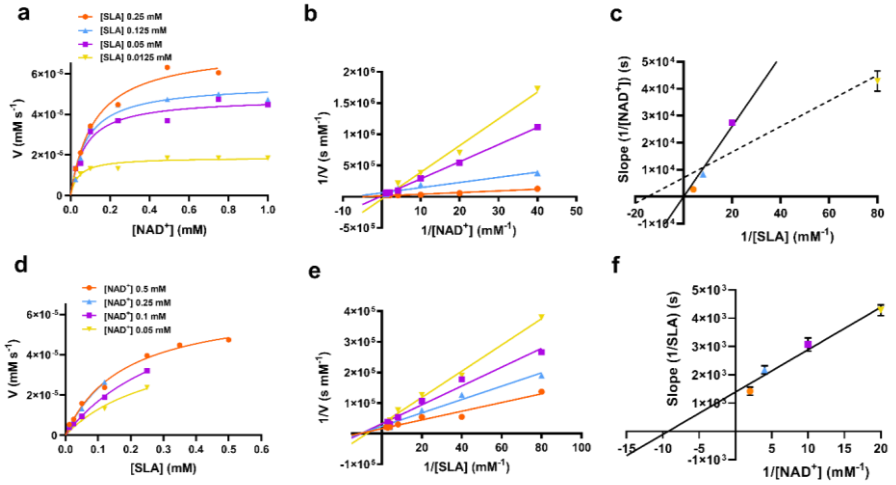

Figure 4. a) Rate data for reactions catalyzed by *Rl*GabD when [NAD^+] was varied under several different fixed concentrations of D-SLA (0.0125−0.5 mM). **b)** Double reciprocal plots for the data from a. **c)** Secondary plot of slopes from the double reciprocal plot (b). Errors are standard error mean. **d)** Rate data for reactions catalyzed by *Rl*GabD when [D-SLA] was varied under several different fixed concentrations of NAD^+ (0.05−1 mM). **e)** Double reciprocal plots for the data from d. **f)** Secondary plot of slopes from the double reciprocal plot (e). Errors are standard error mean.

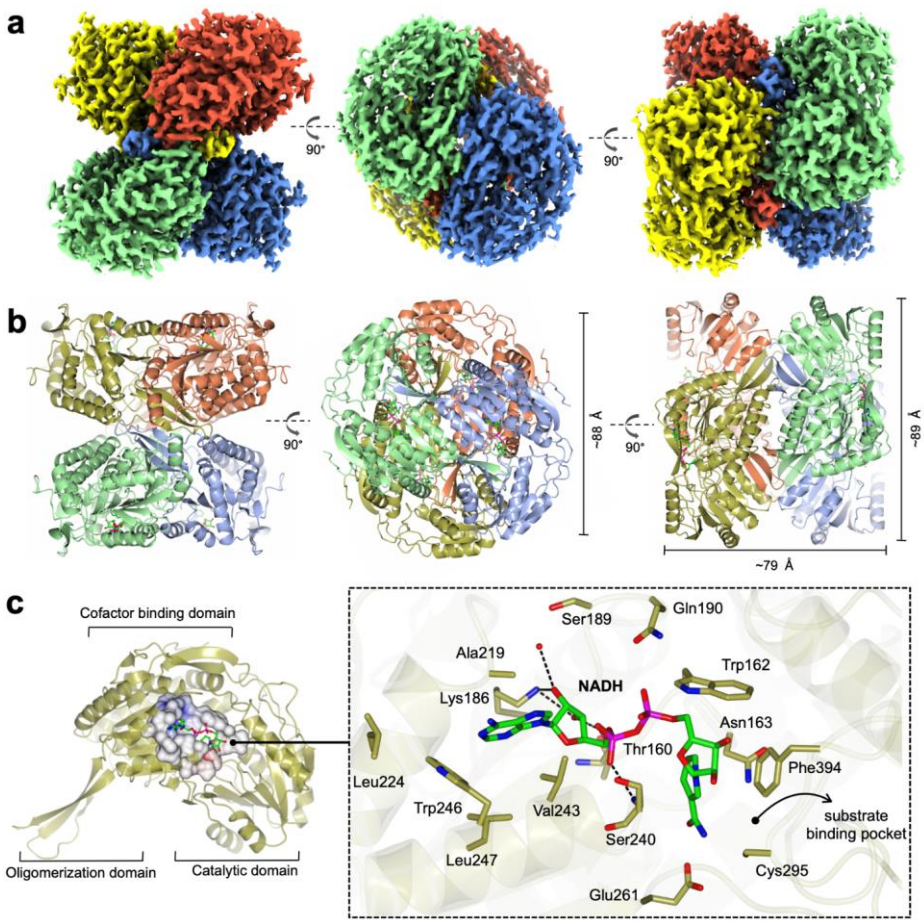

Figure 5. Cryo-EM structure of *Rl*GabD-NADH complex. a) Single-particle cryo-EM reconstruction of the *Rl*GabD-NADH tetramer depicting front and side views. Map is contoured at a threshold of 0.04, and four protomers are coloured in blue, red, green and yellow. **b**) Quaternary structure of *Rl*GabD tetramer depicting front, top and side views. c) Ribbon representation of an *Rl*GabD monomer showing β-strand oligomerization domain, the NAD-cofactor binding domain, and the catalytic domain. Inset: zoom of active site showing bound NADH molecule and residues lining the active site. The location of the SLA binding pocket is indicated. Cys295 is predicted to be the nucleophile, and Glu261 is predicted to be the catalytic general acid/base.

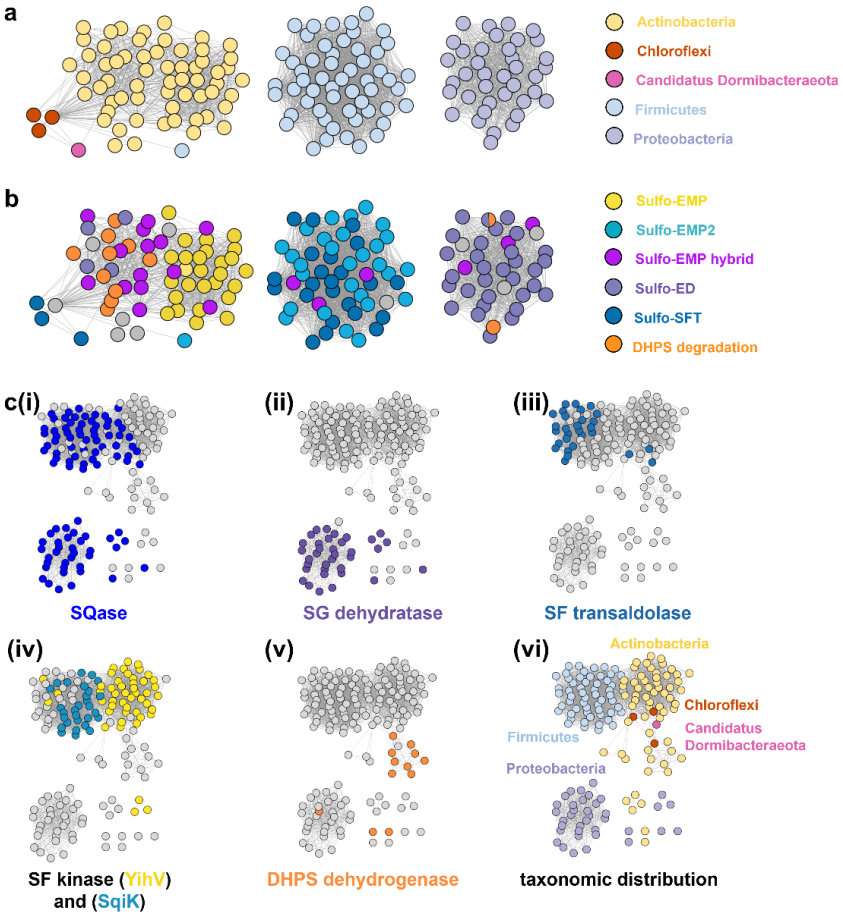

Figure 6. SSN of SLA dehydrogenase proteins (family PF00171) showing distribution in Actinobacteria, Firmicutes, Proteobacteria, Chloroflexi and Candidatus Dormibacteraeota. Nodes are individual SLADH proteins. **a)** Coloured according to occurrence within indicated SQ or DHPS degradation pathways. **b)** Coloured according to distribution across five phyla. **c)** Genome neighborhood similarity network (GNSN) of SLA dehydrogenase proteins. Each node corresponds to an SLA dehydrogenase ortholog in the family PF00171. Nodes are colored according to the presence of genes encoding SQ or DHPS degradation enzymes within a ±10-ORF window of the gene encoding SLA dehydrogenase. Each SQ or DHPS degradation enzyme corresponds to a specific cluster in the SSNN in **Figure S16**. Edges connect nodes that share >4 isofunctional genes in their genome neighborhood. Nodes are coloured in each panel if a specific enzyme belonging to a PFAM is found in the genome neighborhood of the SLA dehydrogenase otholog: (i) SQase (PF01055, royal blue); (ii) SG dehydratase (PF00920, purple); (iii) SF transaldolase (PF00923, dark blue); (iv) YihV-type SF kinase (PF00294, yellow), SqiK-type SF kinase (PF00365, light blue); (v) DHPS dehydrogenase (PF03446, orange); (vi) nodes are coloured according to the Phyla of the host organism.

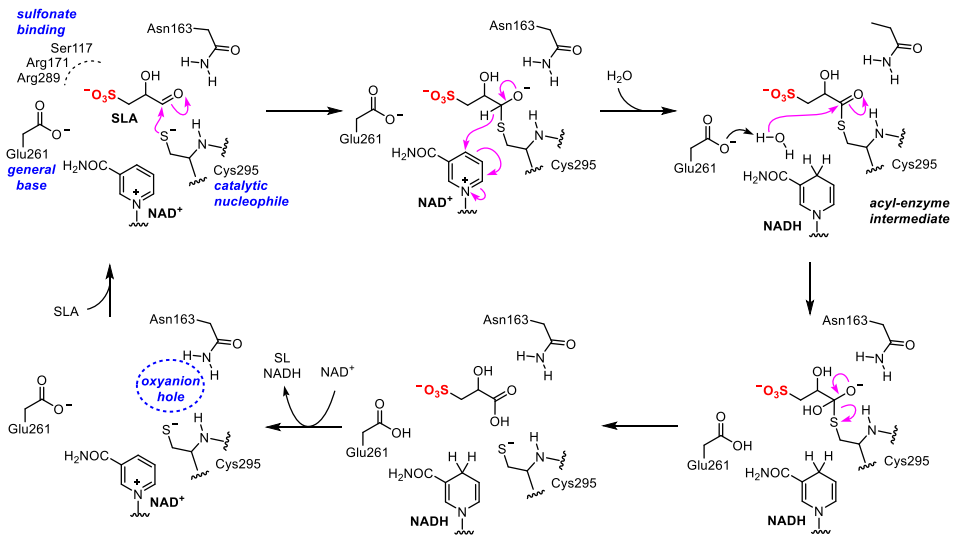

Figure 7. Proposed mechanism and active site residue roles for *Rl*GabD SLA dehydrogenase.