Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2023, **14**, 11429

All publication charges for this article have been paid for by the Royal Society of Chemistry

Molecular basis of sulfolactate synthesis by sulfolactaldehyde dehydrogenase from *Rhizobium leguminosarum*†

Jinling Li,[a] Mahima Sharma,[b] Richard Meek,[b] Amani Alhifthi,[ac] Zachary Armstrong,[a] Niccolay Madiedo Soler,[d] Mihwa Lee,[a] Ethan D. Goddard-Borger,[de] James N. Blaza,[b] Gideon J. Davies *[b] and Spencer J. Williams *[a]

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 turn derives from sulfoglycolysis of the sulfosugar sulfoquinovose, or through oxidation of 2,3-dihydroxypropanesulfonate. 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 indicatively 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 SLA dehydrogenase homologs in a range of putative sulfoglycolytic gene clusters in bacteria predominantly from the phyla Actinobacteria, Firmicutes, and Proteobacteria. This work provides a structural and biochemical view of SLA dehydrogenases to complement our knowledge of SLA reductases, and provide detailed insights into a critical step in the organosulfur cycle.

Received 27th March 2023
Accepted 25th August 2023

DOI: 10.1039/d3sc01594g

rsc.li/chemical-science

Introduction

Sulfur is the tenth most common element by mass in the universe, 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, through which the C6-organosulfonate sulfoquinovose (SQ) is catabolized (Fig. 1a).[4] Alternatively, SLA may be produced by anaerobic DHPS degrading bacteria through the oxidation of DHPS (Fig. 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),[6,8,9] Entner–Doudoroff (sulfo-ED)[10] and sulfofructose transaldolase (sulfo-SFT) pathways (Fig. 1a).[11,12] These pathways generate dihydroxyacetone phosphate, 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,[10] and *Rhizobium leguminosarum* bv. trifolii SRDI565 (ref. 13)); the sulfo-EMP/EMP2 pathways (*Escherichia coli*,[6] *Bacillus urumquiensis*,[7] *Arthrobacter* spp.[9]); and the sulfo-SFT pathway (*Bacillus aryabhattai* SOS1,[11] *Bacillus megaterium* DSM1804,[12] and *Enterococcus gilvus*[11]). Gene clusters encoding these pathways are shown in Fig. 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

[a]*School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: sjwill@unimelb.edu.au*

[b]*York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK. E-mail: gideon.davies@york.ac.uk*

[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*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3sc01594g

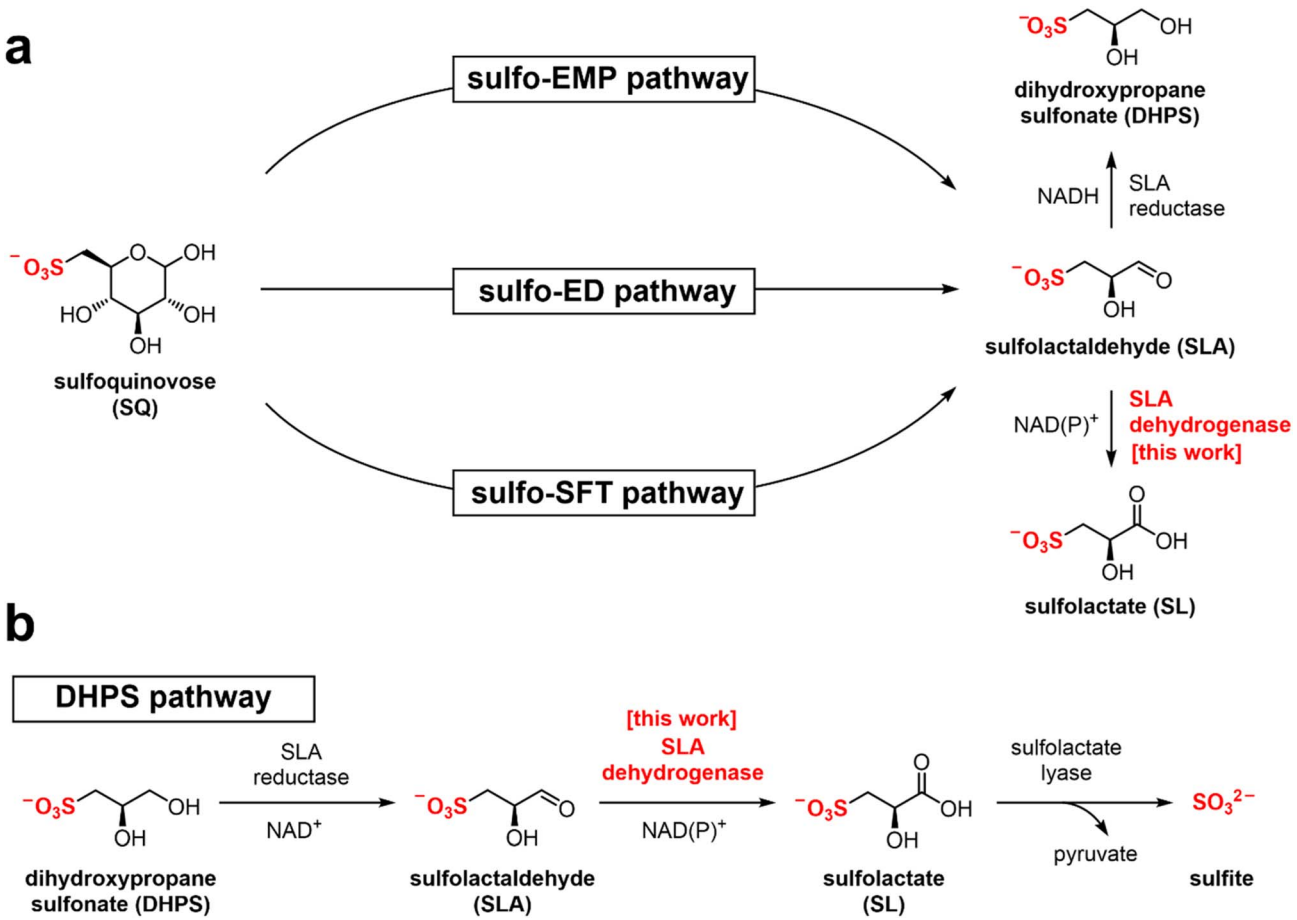

Fig. 1 (a) Formation of SL and DHPS through the pathways of sulfoglycolysis from sulfoquinovose (SQ). (b) Formation and degradation of sulfolactate by catabolism of DHPS.

SL (Fig. 1b).[5] SL is a substrate for SL lyase, which cleaves the C–S bond, producing pyruvate and sulfite.[14] Other bacteria, such as *Roseovarius nubinhibens* and *Paracoccus pantotrophus*, utilize SL as a substrate for growth through the direct action of SL lyase.[5,14,15]

SLA dehydrogenases (annotated as GabD or SlaB) belong to the sequence-based protein family PF00171 within the Pfam database, 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

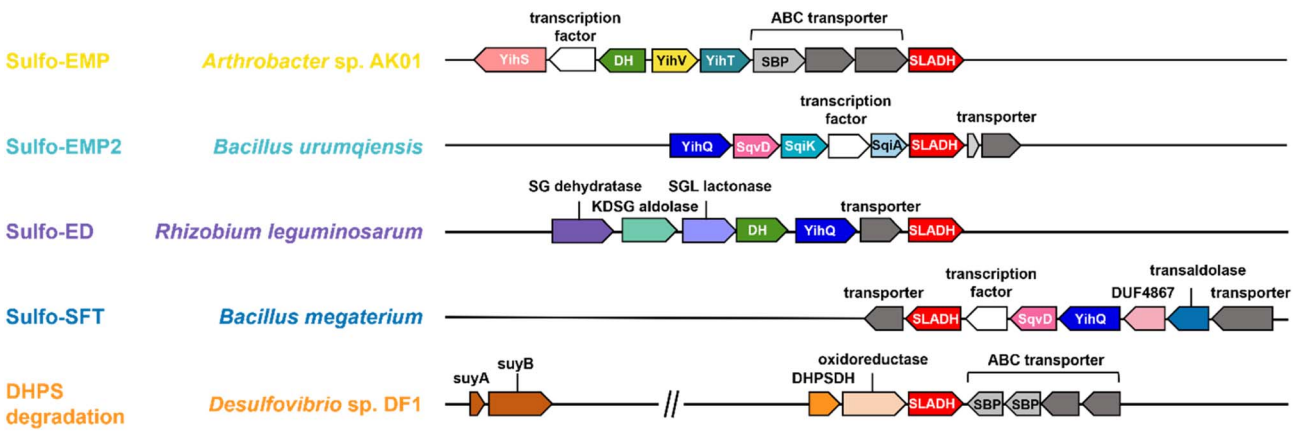

Fig. 2 Proposed gene clusters of bacteria containing SLA dehydrogenase (SLADH) genes that degrade SQ (through sulfo-EMP, sulfo-EMP2, sulfo-ED, and sulfo-SFT pathways) and DHPS (through DHPS degradation pathway).

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)[5] 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,[16] 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.[13] 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 provide evidence in support of an equilibrium ordered mechanism in the forward direction. We report the 3D structure of SLA dehydrogenase using cryogenic electron microscopy (Cryo-EM) and define its quaternary structure and infer the SLA 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, expressed in *E. coli*, and the recombinant protein was purified to homogeneity. Reaction rates for the oxidation of SLA catalyzed by *Rl*GabD were measured using chemically-synthesized racemic D/L-SLA[16] 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 (Fig. S1†). Based on the change in absorbance and the extinction coefficient for NAD^+ we calculate that 47 ± 2% 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 (Fig. 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 $(k_{cat}/K_M)^{app} = 240$

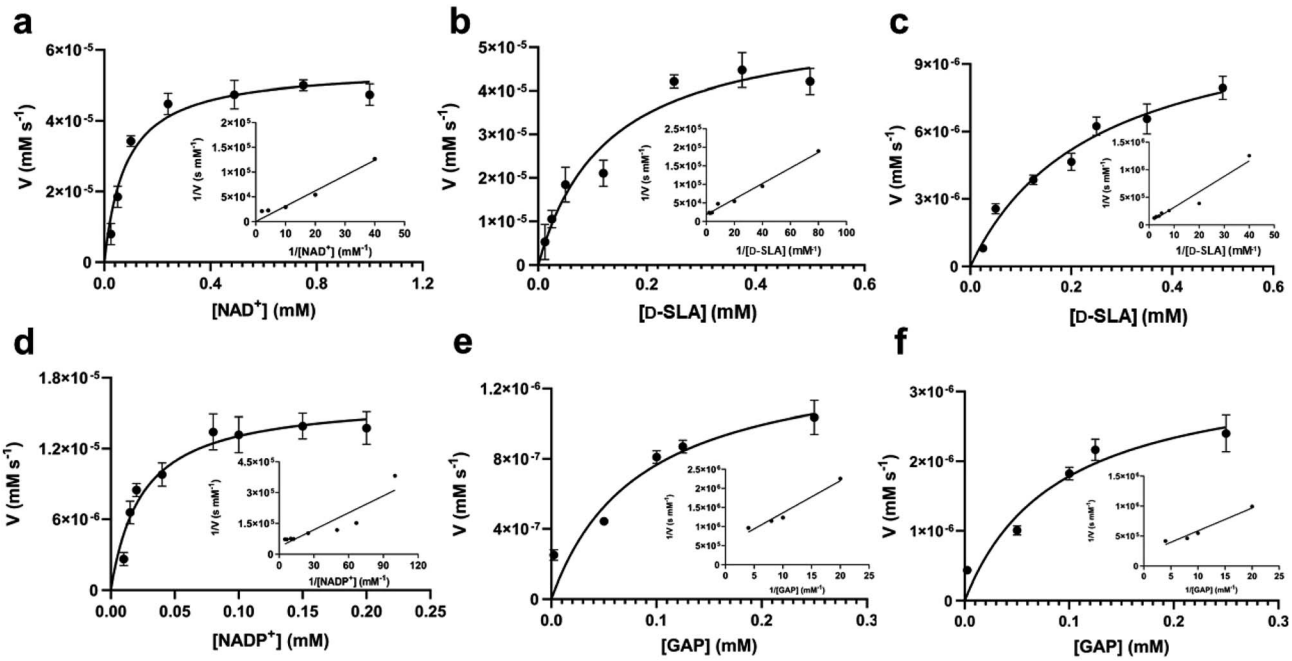

Fig. 3 Michaelis–Menten kinetic analysis for SLA dehydrogenase using NAD^+, D-SLA, $NADP^+$ and GAP as substrates. (a and b) Michaelis–Menten and Lineweaver–Burk (inset) plots for *Rl*GabD under pseudo first-order conditions of [D-SLA] = 0.25 mM (for panel a) and [NAD^+] = 0.25 mM (for panel b). (c and d) Michaelis–Menten and Lineweaver–Burk (inset) plots for *Rl*GabD under pseudo first-order conditions of [$NADP^+$] = 0.25 mM (for panel c) and [D-SLA] = 0.25 mM (for panel d). (e and 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).

Table 1 Apparent first order kinetic parameters for *Rl*GabD 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})
NAD^{+a}	0.081 ± 0.019	17.7 ± 3.2	210 ± 63
$NADP^{+a}$	0.017 ± 0.0041	4.1 ± 0.8	240 ± 73
D-SLA[b]	0.13 ± 0.04	17.8 ± 7.0	137 ± 72
D-SLA[c]	0.16 ± 0.035	4.7 ± 1.1	30 ± 9.4
GAP[b]	0.29 ± 0.031	0.73 ± 0.20	4.2 ± 1.2
GAP[c]	0.17 ± 0.028	3.0 ± 0.4	10.5 ± 2.3

[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.

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 (0.25 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 $NADP^+$: $k_{cat}^{app} = 4.7$ 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.[17] Apparent Michaelis–Menten kinetics for oxidation of racemic GAP were measured at constant concentration (0.25 mM) of nucleotide (Fig. 3e and f). These data reveal that the apparent second order rate constants were similar for NAD^+ $(k_{cat}/K_M)^{app} = 4.2$ 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 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.[18] IC_{50} values were measured at constant [SLA] (at $K_M^{SLA}/10$) and constant [NAD^+] (at $K_M^{NAD^+}$) (Fig. S2a and b†). For tetrahydro-NADH, $IC_{50} = 28$ μM, and for hexahydro-NADH, $IC_{50} = 9.1$ μM, indicating the latter binds more tightly (Fig. S2c and 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.[19] Analysis of initial rates is a powerful way to distinguish reaction mechanisms. For bisubstrate enzymes with substrates A and B, a plot of $1/v_0$ *versus* 1/[A] at various concentrations of substrate B, or $1/v_0$ *versus* 1/[B] at various constant concentrations of substrate A can help determine the kinetic mechanism. For a Ping Pong reaction, the plot of $1/v_0$ *versus* 1/[A] will afford a series of parallel 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 ordered mechanism, on the *y* axis itself.

To gain insight into the kinetic mechanism we measured rate data for varying [SLA] at several constant concentrations of NAD^+, and *vice versa* (Fig. 4a and d). The data were replotted as double reciprocal plots ($1/v_0$ *versus* 1/[SLA]) (Fig. 4b and 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 (Fig. 4b), while the plot of 1/[SLA] *versus* 1/*V* intersected to the left of the *y*-axis (Fig. 4e). While recognizing the difficulty of interpreting whether the intersection of the plot in Fig. 4b is on or close to the *y*-axis, we propose that this data is consistent with a rapid equilibrium ordered mechanism.[19]

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. For the plot of slopes from the 1/[NAD^+] *versus* 1/*V* plot (Fig. 4c, see Fig. legend for a more detailed analysis of the possible lines of fit), the line passed through the origin, while for the plot of slopes from the 1/[SLA] *versus* 1/*V* plot intercepted the *y*-axis above the origin (Fig. 4f). Again, recognizing the limits of this graphical approach to determining kinetic mechanism, this data is indicative of a rapid equilibrium ordered reaction, with NAD^+ binding first to enzyme.[17]

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

*Rl*GabD belongs to the family of aldehyde dehydrogenases (ALDHs), which usually exist and function as homodimers and homotetramers.[20] Size exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) analysis of *Rl*GabD (MW 52,000 Da) showed a solution-state species with molecular weight of approximately 210 kDa (Fig. S3†). *Rl*GabD thus exists as tetramer in solution, presenting it as a suitable candidate for structural characterisation using cryo-EM.

To define conditions for imaging the complex, we studied the interaction of *Rl*GabD with NAD(H) using nanoscale differential

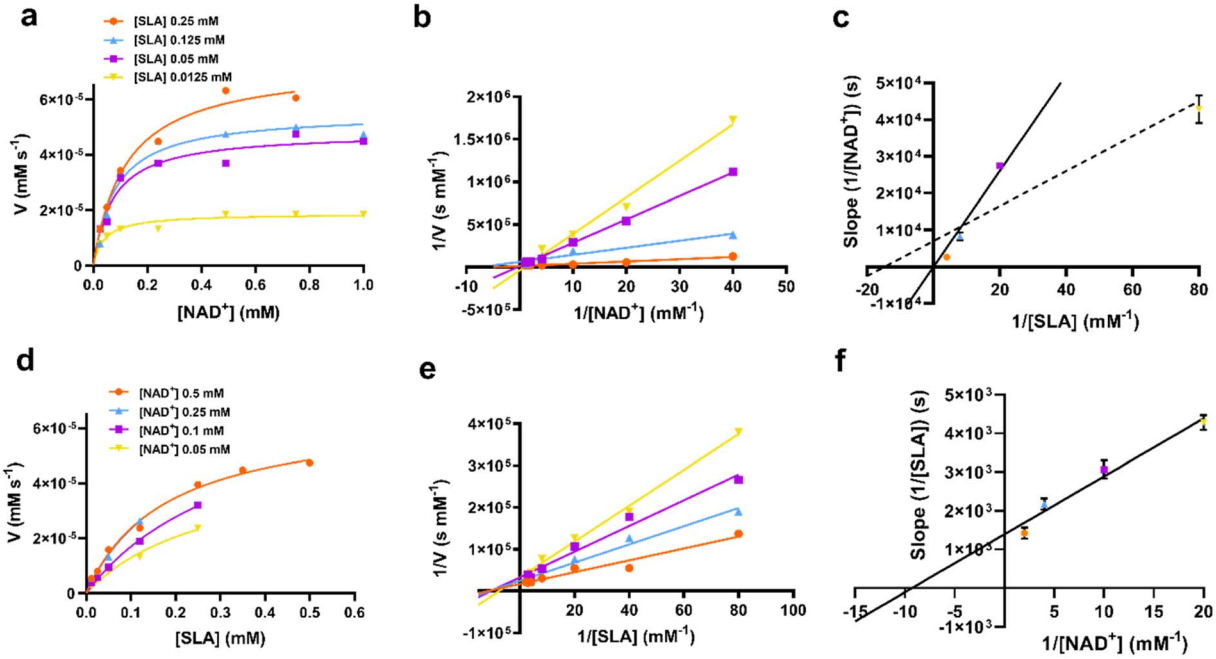

Fig. 4 (a) Rate data for reactions catalyzed by *Rl*GabD when [NAD^+] was varied under several different fixed concentrations of D-SLA (0.0125–0.25 mM). (b) Double reciprocal plots for the data from (a). (c) Secondary plot of slopes from the double reciprocal plot (b). The solid line is fit to the first three data points; the dotted line is fit to all four data points. (d) Rate data for reactions catalyzed by *Rl*GabD when [D-SLA] was varied under several different fixed concentrations of NAD^+ (0.05–0.25 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.

scanning fluorimetry (nano-DSF). nanoDSF uses intrinsic fluorescence to determine the melting temperature of proteins and can aide identification of the formation of protein complexes. nanoDSF revealed a thermal shift (ΔT_m) of 4.3 °C for NAD^+ (and for SLA + NADH, ΔT_m of 2.6 °C), while $NADP^+$ produced a ΔT_m of 3.5 °C (and for NADPH ΔT_m of 2.5 °C) (Fig. S4†). These results guided experiments to image a binary complex. Thus, cryo-EM grids were optimised and prepared with 1 mg mL^{-1} *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 (Fig. 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 (Fig. 5a S6 and S7†).

The *Rl*GabD tetramer assembles as a pair of dimers (Fig. 5a and b). Each protomer of *Rl*GabD adopts the canonical ALDH class I/II fold with three domains.[21,22] 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, antiparallel 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. Poor density was evident for the second phosphate and of the ribosyl-nicotinamide group (Fig. S9†). However, this still allowed NADH to be modelled with confidence, and showed that NADH is bound in an extended mode, as seen in *E. coli* SSADH GabD,[23] with the nicotinamide group protruding into the active site located in the cleft between the two major domains (Fig. 5c). The nicotinamide ring sits in close vicinity to the catalytic dyad, and the adenine ring occupies a hydrophobic pocket lined by Ala 219, Leu224, Val243, Trp246, and Leu247. Based on the distance between heteroatoms, the 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 the backbone carbonyl and hydroxyl group of Ser240. Some additional density is seen at the active site Cys295 residue, possibly indicating oxidation and weaker side-chain density of Glu261 owing to radiation damage or cryo-EM density weakness of negatively charged carboxylate groups.[24,25]

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

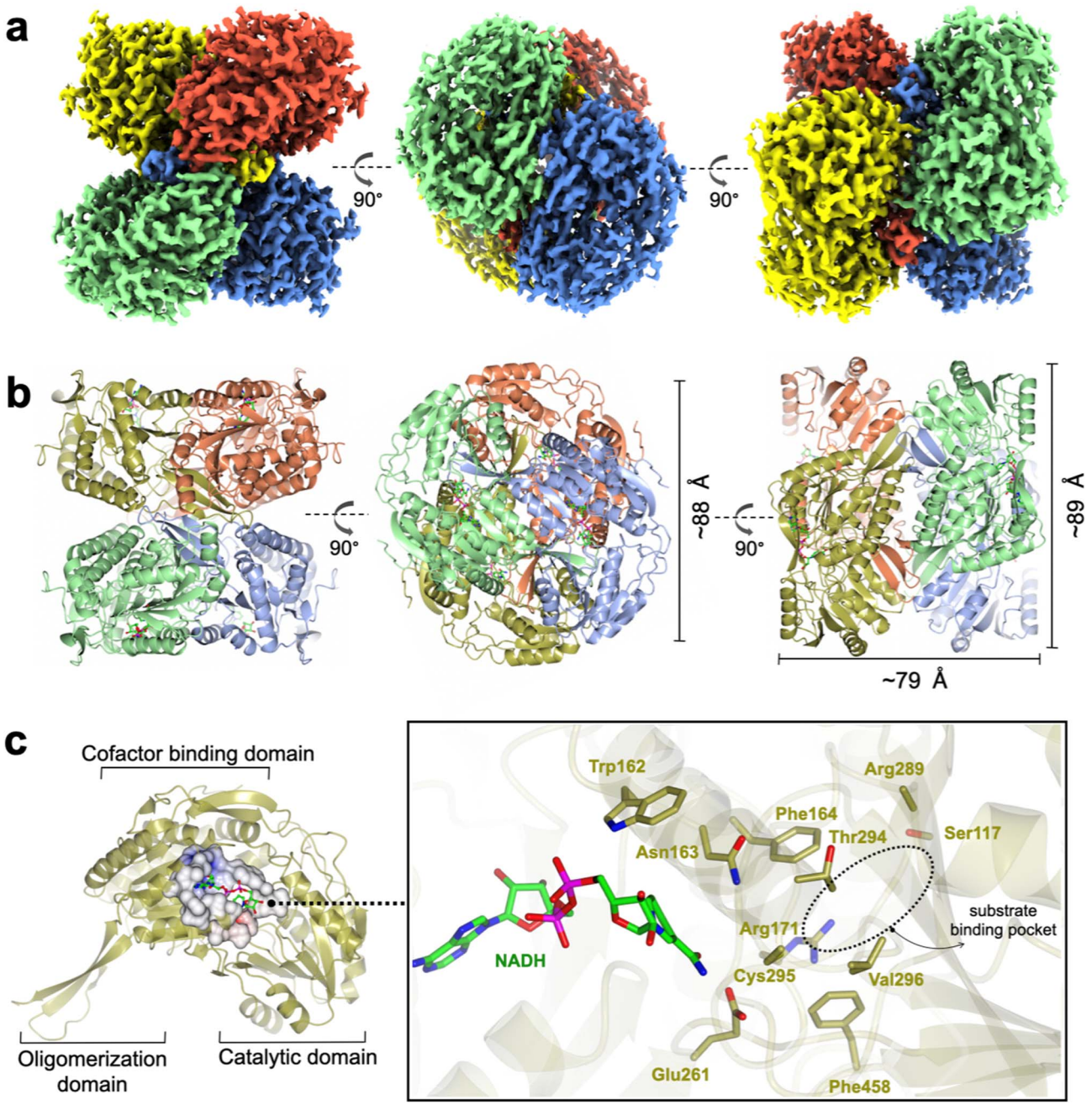

Fig. 5 Cryo-EM structure of *Rl*GabD·NADH complex. (a) Single-particle cryo-EM reconstruction of the *Rl*GabD·NADH tetramer depicting front and side views. Map is contoured at a threshold of 0.04, and four protomers are coloured in blue, red, green and yellow. (b) Quaternary structure of *Rl*GabD tetramer 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 the predicted catalytic nucleophile, and Glu261 is the predicted general acid/base.

representative ALDH members such as *E. coli* SSADH (PDB: 3JZ4, core RMSD of 0.65 Å and 61% sequence ID),[23] *E. coli* lactaldehyde DH (PDB: 2ILU, RMSD 1.3 Å and 35% sequence ID),[26] and the reduced form of human SSADH (PDB: 2W8R, RMSD 0.76 Å and 55% sequence ID).[27] Structural comparison of the *Rl*GabD·NADH complex with *E. coli* SSADH and lactaldehyde DH demonstrates the two discrete 'in' and 'out' cofactor conformations (Fig. 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 alignment of *Rl*GabD with the NADP-

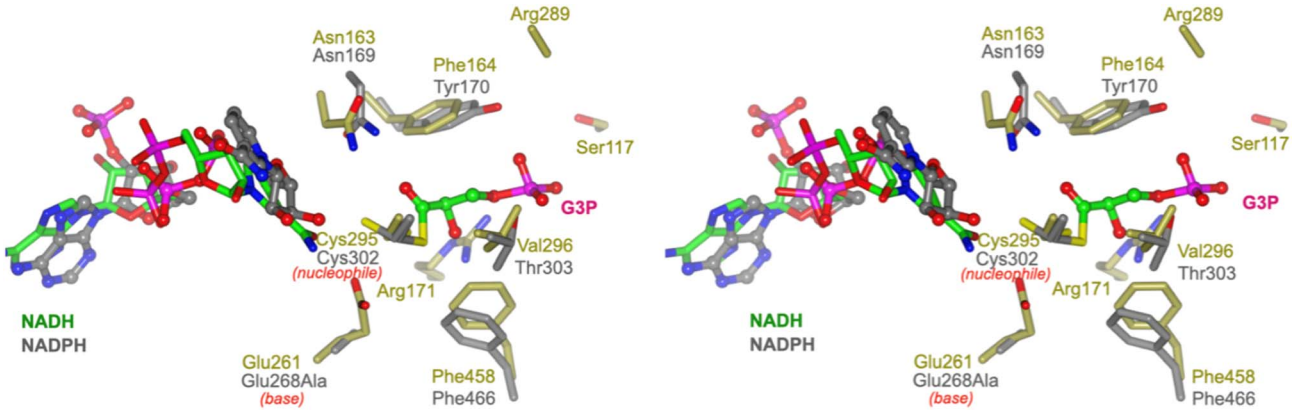

Fig. 6 Conservation of active site residues of *Rl*GabD with glyceraldehyde 3-phosphate dehydrogenase. Stereoview of *Rl*GabD·NADH (gold) and NADPH complex of the acyl enzyme intermediate formed on the Glu268Ala mutant of glyceraldehyde 3-phosphate dehydrogenase from *Streptococcus mutans* (PDB code: 2ESD, grey). The structures align with an RMSD of 1.26 Å over 453 residues. Cys295 is the predicted catalytic nucleophile, and Glu261 is the predicted general base.

dependent non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase (GAPN) from *Streptococcus mutans* (Fig. S15†).[28,29] Like *Rl*GabD, GAPN operates through an ordered sequential mechanism in which the cofactor binds first.[28] The catalytic mechanism for oxidation of GAP by GAPN has been described in detail and involves two main steps.[29] 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 residues involved in GAPN catalysis are conserved with *Rl*GabD and the 3D structure reveals that they are in an appropriate position adjacent to the nicotinamide headgroup of NADH to adopt similar roles (Fig. 5). Overlay of the 3D structures of *Rl*GabD and that of the covalent thioacyl adduct of the Glu268Ala mutant of GAPN[29] reveals spatial conservation of the bases (*Rl*GabD Glu261, GAPN Glu268Ala) and nucleophiles (*Rl*GabD Cys295, GAPN Cys284) (Fig. 6). Thus, we propose that Cys295 is the catalytic nucleophile, Glu261 is the general base, and the oxyanion hole is formed from Asn163 and the backbone NH of Cys295. We probed the importance of Cys295 and Glu261 for catalysis by site-directed mutagenesis. The relative activity *versus* wildtype for Cys295Ala was 1/120 000 and for Glu261Ala was 1/63 000 at 0.5 mM NAD^+ and 0.25 mM D-SLA (activity was undetectable with 0.5 mM $NADP^+$ and 0.25 mM D-SLA). These values approach the limits of site directed mutagenesis because of the complications of translational misincorporation by the heterologous host *E. coli*.[32] Thus, both the Cys295Ala and Glu261Ala variants are severely disabled catalysts, consistent with their critical roles in catalysis.

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 (ref. 30) (Fig. S11 and 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[31]), 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 adjoins 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 succinate semialdehyde (SSA) (PDB 2W8Q)[27] shows SSA occupies this cavity in an orientation where the 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 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.[33] 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 (Fig. S13†). Mapping of these coevolving cliques onto the cryo-EM structure of *Rl*GabD and the AlphaFold2 (ref. 34 and 35) model of *Bacillus megaterium* SlaB identified a tripeptide sequence of partially conserved residues in proximity to

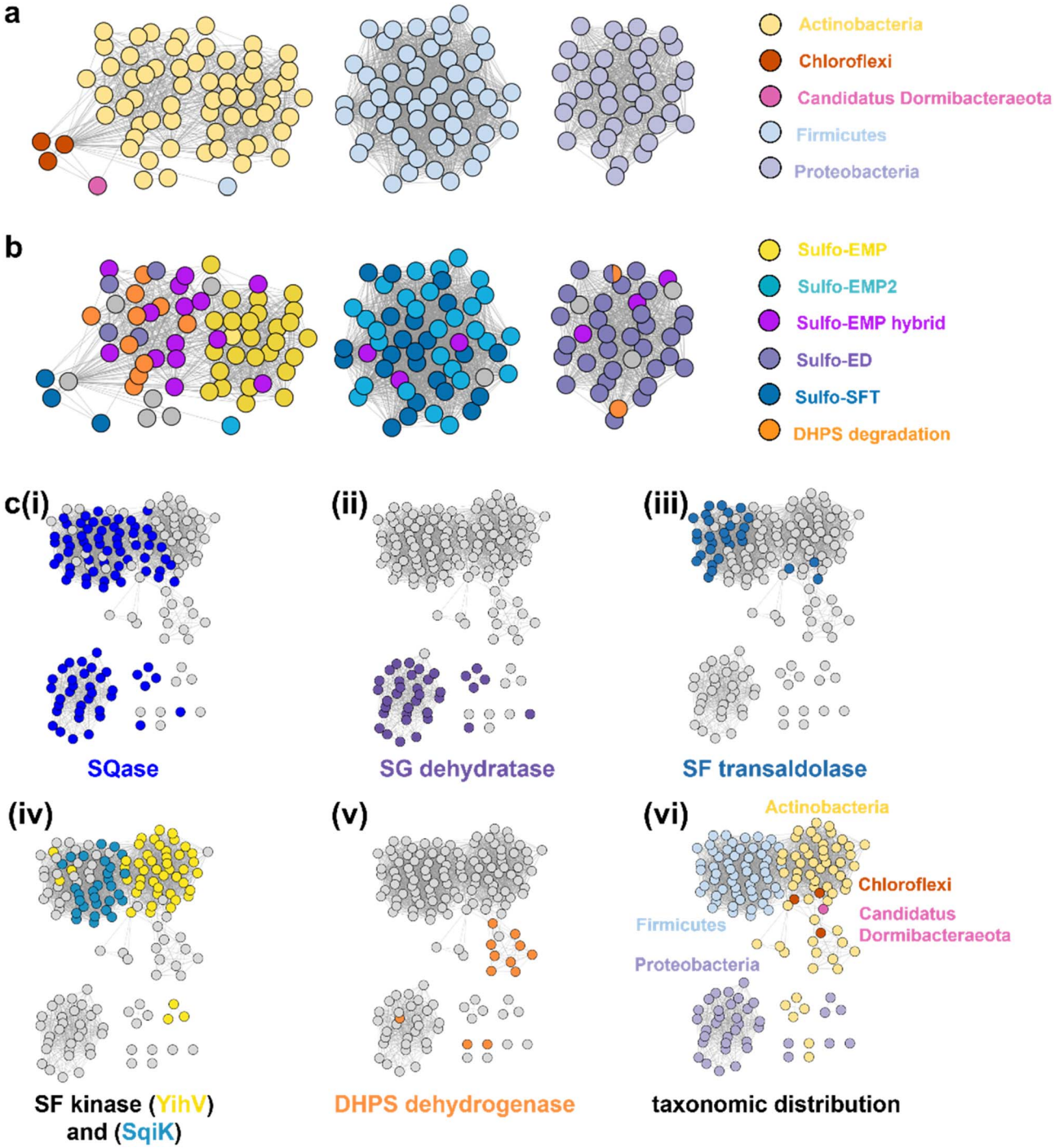

Fig. 7 SSN of SLA dehydrogenase proteins (family PF00171) showing distribution in Actinobacteria, Firmicutes, Proteobacteria, Chloroflexi and *Candidatus Dormibacteraeota*. Nodes are individual SLADH proteins that are coloured according to: (a) occurrence within indicated SQ or DHPS degradation pathways, or (b) 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 Fig. S18.† 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 ortholog: (i) Family GH31 SQase (PF01055); (ii) SG dehydratase (PF00920); (iii) SF transaldolase (PF00923); (iv) YihV-type SF kinase (PF00294), SqiK-type SF kinase (PF00365); (v) DHPS dehydrogenase (PF03446); (vi) nodes are coloured according to the Phyla of the host organism.

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 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 'closed' conformation, while reduction to cysteine causes loop movement and a 'open' conformation (Fig. 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* (Fig. 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/Glu261 poised for catalysis (Fig. 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[36] using the EFI enzyme similarity tool (EFI-EST).[37,38] 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 (Fig. S16†). At alignment score 75, the sequences form a single cluster; at alignment score 100 two clusters; 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 limits its utility. The SSN generated at alignment score 150 (corresponding to minimum identity >53%; Fig. 7a) was coloured based on the pathway encoded by the proposed function of the gene cluster in which the SLA dehydrogenase gene was located (Fig. 7b). 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 membership of two Chloroflexi and one *Candidatus Dormibacteraeota*. 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 ribosomal RNA sequences between bacteria that contain SLADH genes within sulfoglycolytic gene clusters is shown in Fig. 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 (Fig. 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 (Fig. 7c). 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 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 (Fig. 7c(i)).[39,40] Characteristic enzymes encoded by sulfo-EMP (SF kinase YihV), sulfo-EMP2 (SF kinase SqiK), sulfo-ED (SG dehydratase), sulfo-SFT (SF transaldolase) and DHPS degradation (DHPS dehydrogenase) pathways distribute across the GNSN into clusters (Fig. 7c(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*) (Fig. 6c(vi)), we observed coloured clusters that recapitulated the taxonomic clustering of the SSN of SLADH sequences in Fig. 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 $NAD(P)^+$ cofactor activity, and a 30-fold preference for oxidation of SLA *versus* the structurally-related phosphate analogue glyceraldehyde phosphate, a key intermediate in glycolysis/

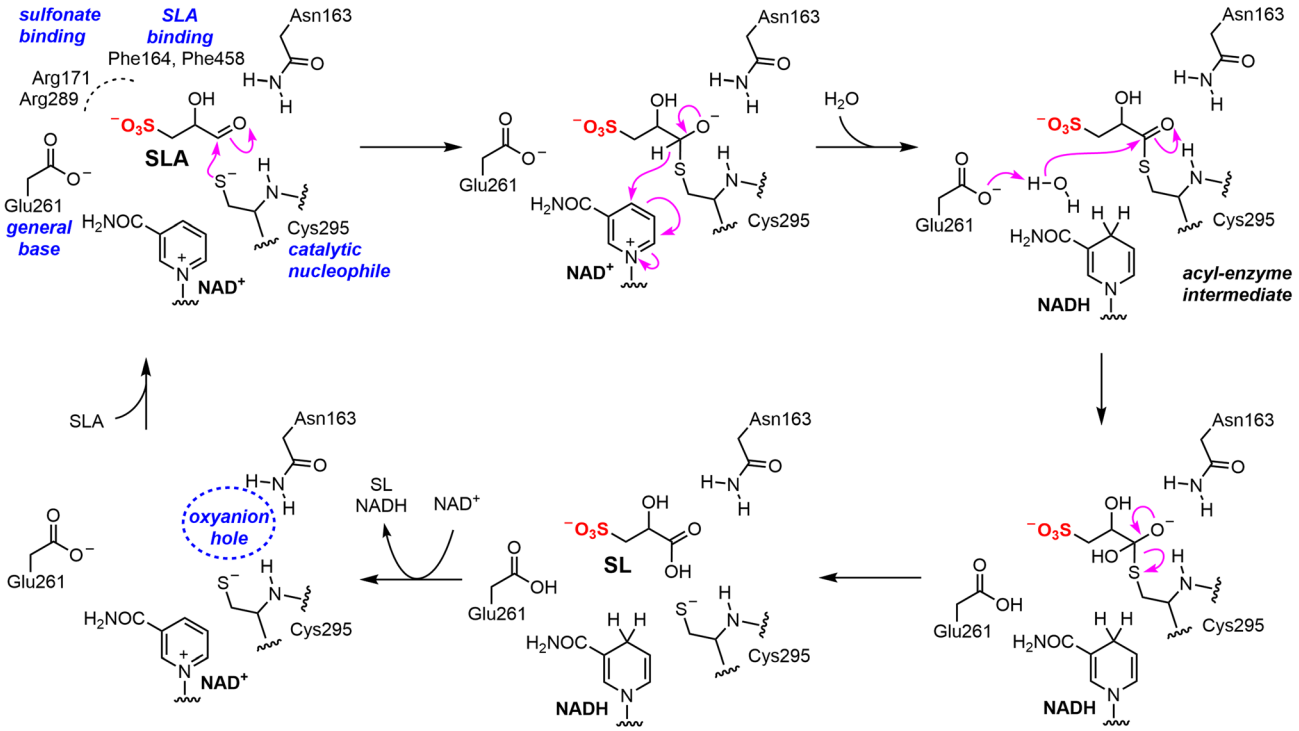

Fig. 8 Proposed mechanism and active site residue roles for *Rl*GabD SLA dehydrogenase.

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 *E. coli* SLA reductase, which had no detectable activity on GAP.[7]

Similar to the well-characterized GAP dehydrogenase from *S. mutans*,[28] our data suggests that *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 proposal of a mechanism for catalysis (Fig. 8). Binding of $NAD(P)^+$, and then SLA gives the Michaelis complex. In the first reaction 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 $NAD(P)^+$, forming an acyl enzyme and NAD(P)H. 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 Arg171-Arg289 comprise the sulfonate binding residues in *Rl*GabD, and the first and last residues within the tripeptide sequence Trp162-Asn163-Phe164 (containing the oxyanion stabilizing residue) are additional SLA binding residues. Arginine residues are common in a wide range of other sulfonate binding proteins and enzymes from various sulfoglycolytic pathways.

Because SL is 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.

Data availability

The ESI† includes experimental and additional details on enzyme kinetics (Fig. S1–S2†), protein biochemistry (Fig. S3–S4†), cryo-EM (Fig. S5–S9†), 3D structural data (Fig. S10–S12†), bioinformatics (Fig. S13–S18†), and structural statistics (Table S1†).

Author contributions

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. and M. S. conducted bioinformatics. 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.

Conflicts of interest

The authors declare no competing interests.

Abbreviations

DHPS	2,3-dihydroxypropanesulfonate
ED	Entner-Doudoroff
EMP	Embden-Meyerhof-Parnas
GAP	glyceraldehyde-3-phosphate
SFT	sulfofructose transaldolase
SL	sulfolactate
SLA	sulfolactaldehyde
SQ	sulfoquinovose
SSA	succinate semialdehyde
NAD(P)H	reduced nicotinamide adenine dinucleotide (phosphate)

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. J. L. is supported by the China Scholarship Council. E. D. G.-B. acknowledges support from The Walter and Eliza Hall Institute of Medical Research, National Health and Medical Research Council of Australia (NHMRC) project grant GNT2000517, the Australian Cancer Research Fund, and the Brian M. Davis Charitable Foundation Centenary Fellowship. 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 acknowledge Dr Andrew Leech at the University of York Bioscience Technology Facility for assistance with SEC-MALLS analysis, and Arashdeep Kaur for assistance with bioinformatics.

Notes and references

1 W. M. Haynes, D. R. Lide and T. J. Bruno, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, 97 edn, 2017.
2 A. M. Cook, K. Denger and T. H. Smits, *Arch. Microbiol.*, 2006, **185**, 83–90.
3 E. D. Goddard-Borger and S. J. Williams, *Biochem. J.*, 2017, **474**, 827–849.
4 A. J. D. Snow, L. Burchill, M. Sharma, G. J. Davies and S. J. Williams, *Chem. Soc. Rev.*, 2021, **50**, 13628–13645.
5 A. Burrichter, K. Denger, P. Franchini, T. Huhn, N. Müller, D. Spiteller and D. Schleheck, *Front. Microbiol.*, 2018, **9**, 2792.
6 K. Denger, M. Weiss, A. K. Felux, A. Schneider, C. Mayer, D. Spiteller, T. Huhn, A. M. Cook and D. Schleheck, *Nature*, 2014, **507**, 114–117.
7 M. Sharma, P. Abayakoon, J. P. Lingford, R. Epa, A. John, Y. Jin, E. D. Goddard-Borger, G. J. Davies and S. J. Williams, *ACS Catal.*, 2020, **10**, 2826–2836.
8 J. Liu, Y. Wei, K. Ma, J. An, X. Liu, Y. Liu, E. L. Ang, H. Zhao and Y. Zhang, *ACS Catal.*, 2021, **11**, 14740–14750.
9 M. Sharma, P. Abayakoon, R. Epa, Y. Jin, J. P. Lingford, T. Shimada, M. Nakano, J. W. Y. Mui, A. Ishihama, E. D. Goddard-Borger, G. J. Davies and S. J. Williams, *ACS Cent. Sci.*, 2021, **7**, 476–487.
10 A. K. Felux, D. Spiteller, J. Klebensberger and D. Schleheck, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E4298–E4305.
11 B. Frommeyer, A. W. Fiedler, S. R. Oehler, B. T. Hanson, A. Loy, P. Franchini, D. Spiteller and D. Schleheck, *iScience*, 2020, **23**, 101510.
12 Y. Liu, Y. Wei, Y. Zhou, E. L. Ang, H. Zhao and Y. Zhang, *Biochem. Biophys. Res. Commun.*, 2020, **533**, 1109–1114.
13 J. Li, R. Epa, N. E. Scott, D. Skoneczny, M. Sharma, A. J. D. Snow, J. P. Lingford, E. D. Goddard-Borger, G. J. Davies, M. J. McConville and S. J. Williams, *Appl. Environ. Microbiol.*, 2020, **86**, e00750.
14 U. Rein, R. Gueta, K. Denger, J. Ruff, K. Hollemeyer and A. M. Cook, *Microbiology*, 2005, **151**, 737–747.
15 K. Denger, J. Mayer, M. Buhmann, S. Weinitschke, T. H. Smits and A. M. Cook, *J. Bacteriol.*, 2009, **191**, 5648–5656.
16 P. Abayakoon, R. Epa, M. Petricevic, C. Bengt, J. W. Y. Mui, P. L. van der Peet, Y. Zhang, J. P. Lingford, J. M. White, E. D. Goddard-Borger and S. J. Williams, *J. Org. Chem.*, 2019, **84**, 2901–2910.
17 D. Gauss, B. Schoenenberger and R. Wohlgemuth, *Carbohydr. Res.*, 2014, **389**, 18–24.
18 K. G. Dave, R. B. Dunlap, M. K. Jain, E. H. Cordes and E. Wenkert, *J. Biol. Chem.*, 1968, **243**, 1073–1074.
19 V. Leskovac, *Comprehensive Enzyme Kinetics*, Kluwer Academic Publishers, New York, 2004.
20 K. Shortall, A. Djeghader, E. Magner and T. Soulimane, *Front. Mol. Biosci.*, 2021, **8**, 659550.
21 S. A. Moore, H. M. Baker, T. J. Blythe, K. E. Kitson, T. M. Kitson and E. N. Baker, *Structure*, 1998, **6**, 1541–1551.
22 C. G. Steinmetz, P. Xie, H. Weiner and T. D. Hurley, *Structure*, 1997, **5**, 701–711.
23 C. G. Langendorf, T. L. G. Key, G. Fenalti, W.-T. Kan, A. M. Buckle, T. Caradoc-Davies, K. L. Tuck, R. H. P. Law and J. C. Whisstock, *PLoS One*, 2010, **5**, e9280.
24 J. Wang, *Protein Sci.*, 2017, **26**, 396–402.
25 M. Beckers, D. Mann and C. Sachse, *Prog. Biophys. Mol. Biol.*, 2021, **160**, 26–36.
26 L. Di Costanzo, G. A. Gomez and D. W. Christianson, *J. Mol. Biol.*, 2007, **366**, 481–493.
27 Y. G. Kim, S. Lee, O. S. Kwon, S. Y. Park, S. J. Lee, B. J. Park and K. J. Kim, *EMBO J.*, 2009, **28**, 959–968.
28 S. Marchal, S. Rahuel-Clermont and G. Branlant, *Biochemistry*, 2000, **39**, 3327–3335.

29 K. D'Ambrosio, A. Pailot, F. Talfournier, C. Didierjean, E. Benedetti, A. Aubry, G. Branlant and C. Corbier, *Biochemistry*, 2006, **45**, 2978–2986.
30 W. Tian, C. Chen, X. Lei, J. Zhao and J. Liang, *Nucleic Acids Res.*, 2018, **46**, W363–W367.
31 B. Lee and F. M. Richards, *J. Mol. Biol.*, 1971, **55**, 379–400.
32 E. B. Kramer and P. J. Farabaugh, *RNA*, 2007, **13**, 87–96.
33 F. N. Baker and A. Porollo, *BMC Bioinf.*, 2016, **17**, 119.
34 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583–589.
35 M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis and S. Velankar, *Nucleic Acids Res.*, 2022, **50**, D439–D444.
36 H. J. Atkinson, J. H. Morris, T. E. Ferrin and P. C. Babbitt, *PLoS One*, 2009, **4**, e4345.
37 R. Zallot, N. Oberg and J. A. Gerlt, *Biochemistry*, 2019, **58**, 4169–4182.
38 J. A. Gerlt, J. T. Bouvier, D. B. Davidson, H. J. Imker, B. Sadkhin, D. R. Slater and K. L. Whalen, *Biochim. Biophys. Acta*, 2015, **1854**, 1019–1037.
39 P. Abayakoon, Y. Jin, J. P. Lingford, M. Petricevic, A. John, E. Ryan, J. Wai-Ying Mui, D. E. V. Pires, D. B. Ascher, G. J. Davies, E. D. Goddard-Borger and S. J. Williams, *ACS Cent. Sci.*, 2018, **4**, 1266–1273.
40 G. Speciale, Y. Jin, G. J. Davies, S. J. Williams and E. D. Goddard-Borger, *Nat. Chem. Biol.*, 2016, **12**, 215–217.