

SUPPLEMENTARY INFORMATION

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

Jinling Li,^a Mahima Sharma,^b Richard Meek,^b Amani Alhifthi,^{a,c} Zachary Armstrong,^a
Niccolay Madiedo Soler,^d Mihwa Lee,^a Ethan D. Goddard-Borger,^{d,e} Jamesw 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

SUPPLEMENTARY INFORMATION

Experimental section

Heterologous expression and protein purification

The sequence encoding RIGabD (WP_017967313.1) was amplified from *Rhizobium leguminosarum* SRDI-565 genomic DNA by PCR using the primers 5'-TTACTCATATGCCATCCAACCTATGACAGC-3' and 5'-TCAGACTCGAGGTTCTGCGTCAACCCGGC-3'. The amplicon was cloned into the MCS of pET-29b(+) (Novagen) using the *NdeI* and *XhoI* sites, and sequence-verified by Sanger sequencing. This plasmid was transformed into chemically competent *E. coli* BL21(DE3) cells and starter cultures grown in LB-medium (5 mL) containing 100 $\mu\text{g mL}^{-1}$ 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 A_{600} of 0.6-0.8 was reached. *RIGabD* expression was induced by addition of IPTG (0.5 mM) and shaking continued overnight at 18 °C at 200 r.p.m. The cells were harvested by centrifugation at 5000 g for 20 min and the pellet 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 \times 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 His₆-tagged protein was purified by immobilized-metal affinity chromatography (IMAC) using a Ni-NTA column, followed by size exclusion chromatography (SEC). 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. *RIGabD* 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 Vivaspın® 6 with a 30 kDa MW cut-off membrane, to a final concentration of 8 mg mL⁻¹ for structural studies.

Generation of active-site variant constructs

The active-site variant constructs, *RIGabD* Glu261Ala and *RIGabD* Cys295Ala, were generated with the Q5 site-directed mutagenesis kit (New England Biolabs) using the primers 5'-cgTTGGGCGGCAATGCGC-3' and 5'-cCAGCGAGAGGCGCTTG-3' for Glu261Ala and 5'-CGGCCAGACCgcgGTTTGCGCCA-3' and 5'-GCATTGCGGAATTTGGAG-3' for Cys295Ala, respectively (lower-case letters indicate the mutated sequence). Mutagenesis was verified by DNA sequencing and the proteins were purified by the same method used for the wild-type protein.

SUPPLEMENTARY INFORMATION

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 *R/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 ¹H 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 D₂O) of methyl β -glucoside at 100, 50, and 10 mM. The absolute integration of H1 was 15738, 7753.06, 1249.15, respectively (**Fig. S1A**). The SLA stock solution was diluted 1:1 with D₂O. The absolute integration of H2 (δ 3.79 ppm in D₂O) of SLA in the ¹H NMR

SUPPLEMENTARY INFORMATION

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 *R/GabD*

The production of NADH from oxidation of SLA catalyzed by *R/GabD* was monitored using UV-Vis spectrophotometer at 340 nm, in triplicate. 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 [*R/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 (**Fig. S1B**). More *R/GabD* was added to a final [*R/GabD*] = 205 nM, and the reaction was monitored for 100 min with no further increase in absorbance. The extinction coefficient used for NADH was 6363 M⁻¹ cm⁻¹. Error is standard error mean.

Michaelis-Menten kinetic analysis of *R/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 *R/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 *R/GabD*, constant 0.25 mM D-SLA and varying concentrations of NAD⁺/NADP⁺ (0.05-1 mM). For the Michaelis-Menten kinetics of D-SLA, constant 0.25 mM NAD⁺/NADP⁺ and varying concentrations of D-SLA (0.025-0.25 mM) were used. 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. 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 >0.5 mM [D-SLA] and >0.25 mM [NADP⁺]. This appears to be a result of inhibition.

The activity of Cys295Ala and Glu261Ala variants were measured at 0.5 mM NAD⁺ and 0.25 mM D-SLA, or 0.5 mM NADP⁺ and 0.25 mM D-SLA) in 0.1% BSA 30 mM Tris buffer pH 7.12 at 30 °C. The concentration of each variant enzyme was 3000 nM.

Cryo-EM 3D-structure of *R/GabD*•NADH

A frozen aliquot of *R/GabD* was prepared at 4 mg/mL and complexed with 2 mM NADH for

SUPPLEMENTARY INFORMATION

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 μ L 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 *R/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 *R/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 μ m. The autofocus function was run every 10 μ m, and the dataset was collected with a total dose of 50 electrons per Å².

Image processing and 3D reconstruction

Movie frames of the *R/GabD*•NADH were motion corrected without binning, using a pixel size of 0.574 Å, and dose-weighted using the Motioncorr2 program.¹ Contrast transfer function (CTF) corrections were performed using CTFFIND 4.1.² Most of the subsequent processing steps were carried out using RELION 3.³ 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 refinement⁴ was performed for magnification anisotropy; optical aberrations (up to the fourth order); and per-particle defocus and per-micrograph astigmatism. Bayesian Polishing⁵ 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 Fig. S6.

Model building, refinement, and validation

An initial model was built using the map_to_model function in Phenix.⁶ The initial model was

SUPPLEMENTARY INFORMATION

modified using *Coot*.⁷ Cycles of real space refinement in *Phenix* and interactive model building in *Coot* were used to fit the structure to the map.⁶ Model geometries were assessed with MolProbity.⁸ Structures and maps in the Fig. were rendered with ccp4mg,⁹ PyMOL (<http://www.pymol.org/>) or ChimeraX.¹⁰

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/>),¹¹ and Genome neighborhood analysis was carried out using the web-based EFI Genome Neighborhood Tool (EFI-GNT) (<https://efi.igb.illinois.edu/efi-gnt/>).¹² All networks were visualized using Cytoscape v3.9.¹³ The SSN created using the alignment score 150 was used to generate genome neighbourhood diagrams (GND) with open reading frame (ORF) \pm 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; **Fig. S18**). 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, PF00294), 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.

SUPPLEMENTARY INFORMATION

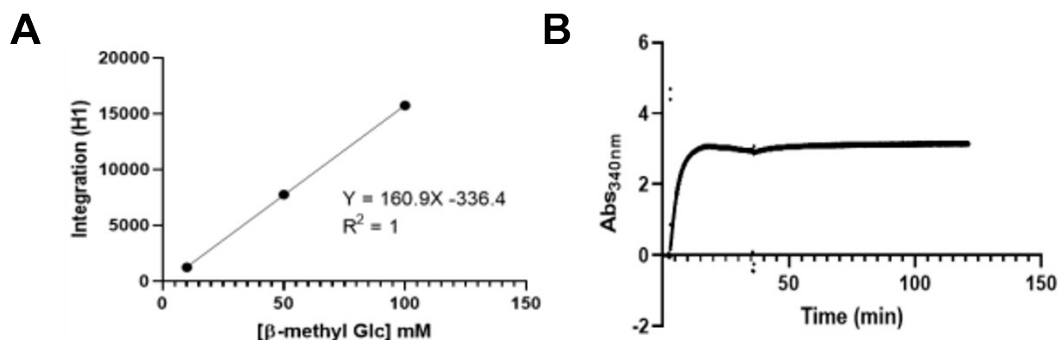


Figure S1. **A)** Calibration curve of the integration of the anomeric proton in the ^1H NMR spectrum of methyl β -glucoside as a function of concentration, used for determination of [SLA]. **B)** Representative plot (performed in triplicate) for increase in absorption (corresponding to production of NADH) from oxidation of SLA catalyzed by *R/GabD* SLA dehydrogenase monitored using UV-Vis spectrophotometer at 340 nm. Reactions contained 30 mM Tris buffer pH 7 at 30 °C $[\text{NAD}^+] = 1.5$ mM and $[\text{SLA}] = 1$ mM, and were initiated by addition of *R/GabD* to final concentration of 15.8 nM. The reaction appeared complete by 40 min. At this time, more *R/GabD* was added to a final concentration of 205 nM, and the reaction was monitored for 100 min with no further increase in absorbance. Maximum absorbance (A_{340}) = 3.081, corresponding to consumption of 48.7% of the SLA.

SUPPLEMENTARY INFORMATION

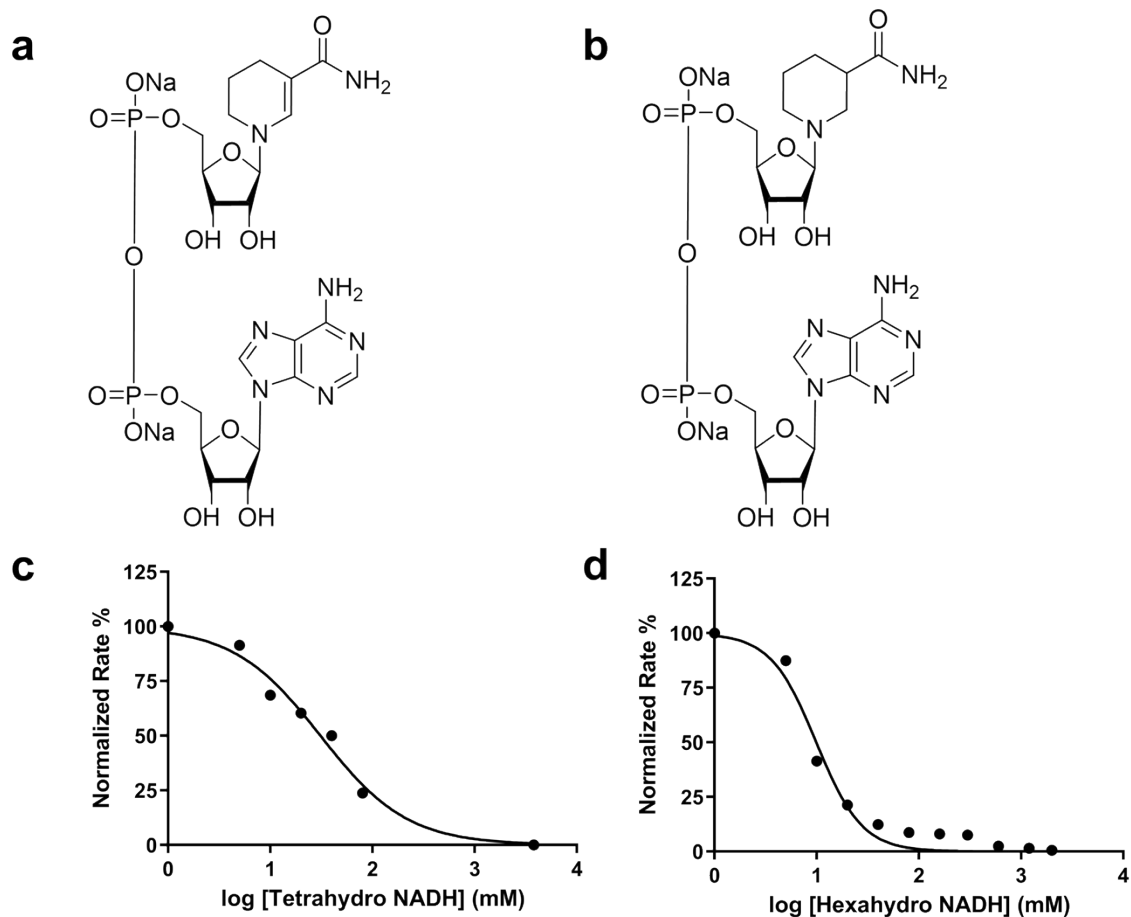


Figure S2. Structures of **a**) tetrahydro-NADH and **b**) hexahydro-NADH. Tetrahydro-NADH and hexahydro-NADH were assessed as inhibitors of *R/GabD*. IC_{50} values were determined in 30 mM Tris buffer and 30 mM KCl (pH 7.12) with 0.1% BSA at 30 °C using constant [SLA] (at $K_M^{SLA}/10$) and constant $[NAD^+]$ (at $K_M^{NAD^+}$) using a UV/Vis spectrophotometer at 340 nm. **c**) Inhibition of *R/GabD* at constant NAD^+ and SLA by tetrahydro NADH, IC_{50} (tetrahydro-NADH) = 28.2 μ M; and **d**) hexahydro NADH, IC_{50} (hexahydro-NADH) = 9.1 μ M.

SUPPLEMENTARY INFORMATION

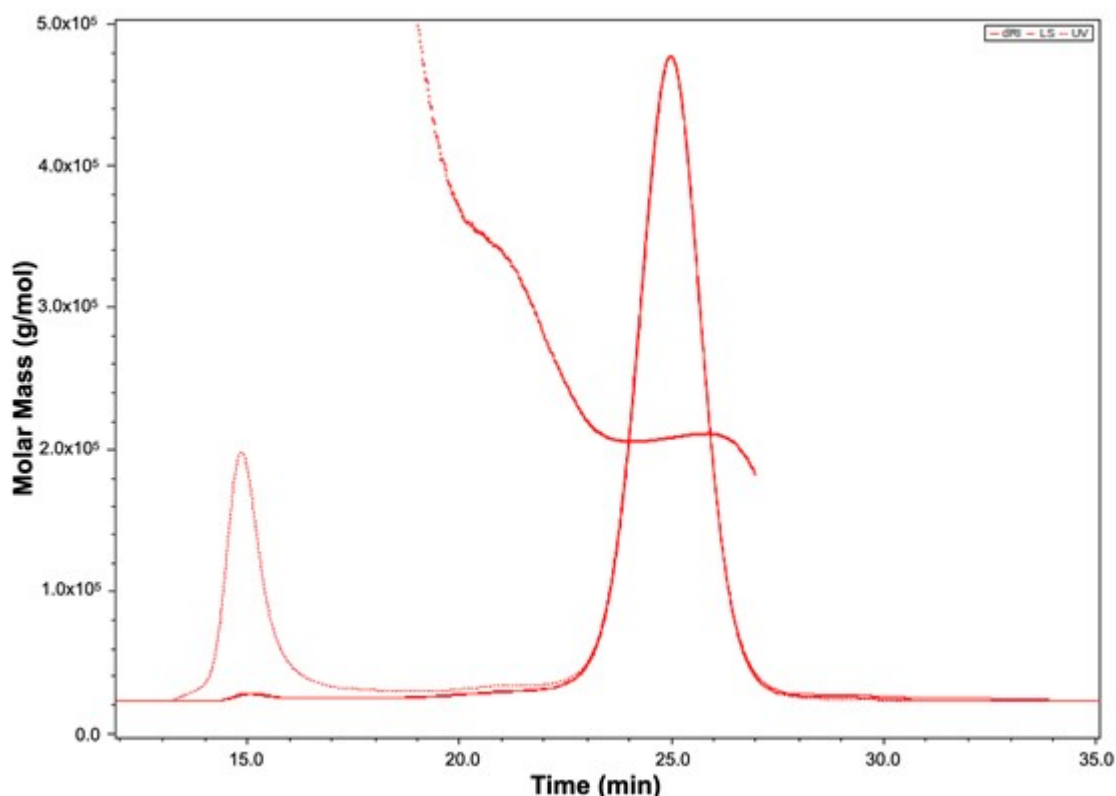
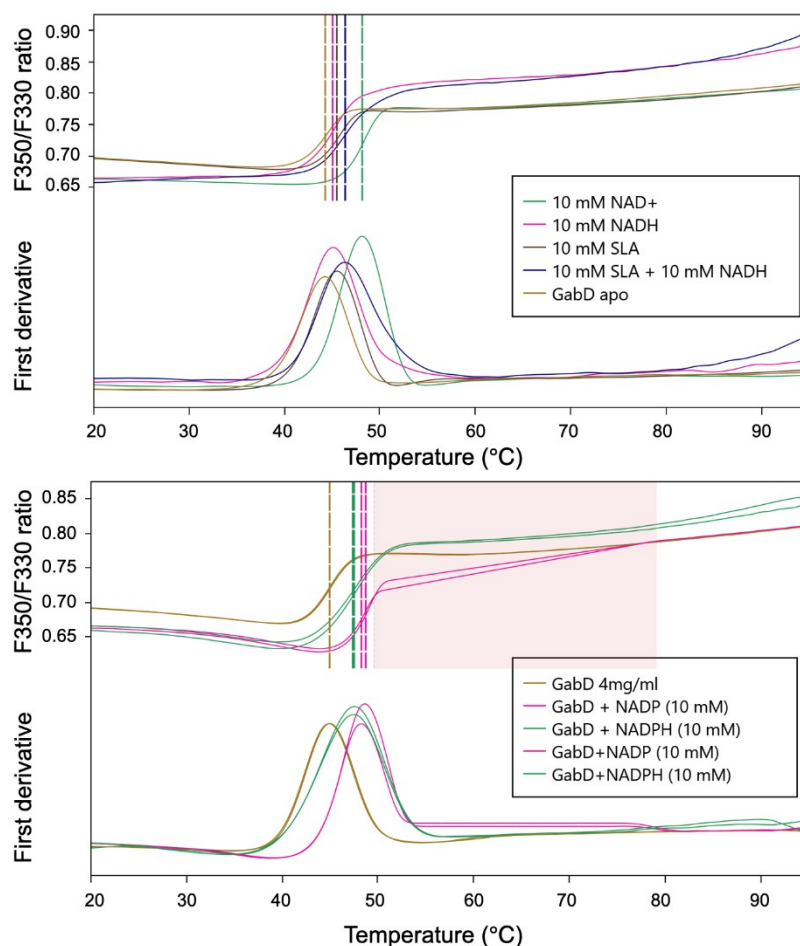


Figure S3. SEC-MALLS plot reveals the oligomeric state of *RIGabD* in solution. UV-trace and an average molecular weight trace (red), calculated from the refractive index and light scattering signal gave mass estimation of 207 kDa, which corresponds to a homotetramer (with monomer molecular weight of 51568 Da). The area eluted under the major peak corresponds to 337 μ g, which comprises ~96% of the eluted material confirming near homogeneity of the sample. A small amount of material at the void volume has MW >8 MDa, corresponding to some larger aggregates.

SUPPLEMENTARY INFORMATION



Sample	Inflection point for F350/F330 (Replicate 1)	Inflection point for F350/F330 (Replicate 2)	Trend	ΔT_m (Avg)
<i>Interaction of RIGabD with substrates SLA, NAD(H) and analogs</i>				
1 RIGabD apo 4mg/mL	44.2°C	44.0°C		
2 RIGabD + NAD (10 mM)	48.1°C	48.8°C	↑	4.3°C
3 RIGabD + NADH (10 mM)	45.1°C	45.6°C	↑	1.5°C
4 RIGabD + SLA (10 mM)	45.5°C	45.4°C	↑	1.3°C
5 RIGabD + NADH + SLA (10 mM each)	46.4°C	47.1°C	↑	2.6°C
6 RIGabD + NADH ₃ (10 mM)	43.7°C	43.9°C	↓	0.3°C
7 RIGabD + NADH ₄ (10 mM)	43.6°C	43.9°C	↓	0.4°C
<i>Interaction of RIGabD with NADP(H)</i>				
1 RIGabD apo 4mg/ml	45.0°C	44.9°C		
2 RIGabD + NADP (10 mM)	48.7°C	48.3°C	↑	3.5°C
3 RIGabD + NADPH (10 mM)	47.5°C	47.4°C	↑	2.5°C

Figure S4. NanoDSF studies with RIGabD. T_m shifts occurring upon binding cofactors NAD(P)H (and analogs) and substrate SLA.

SUPPLEMENTARY INFORMATION

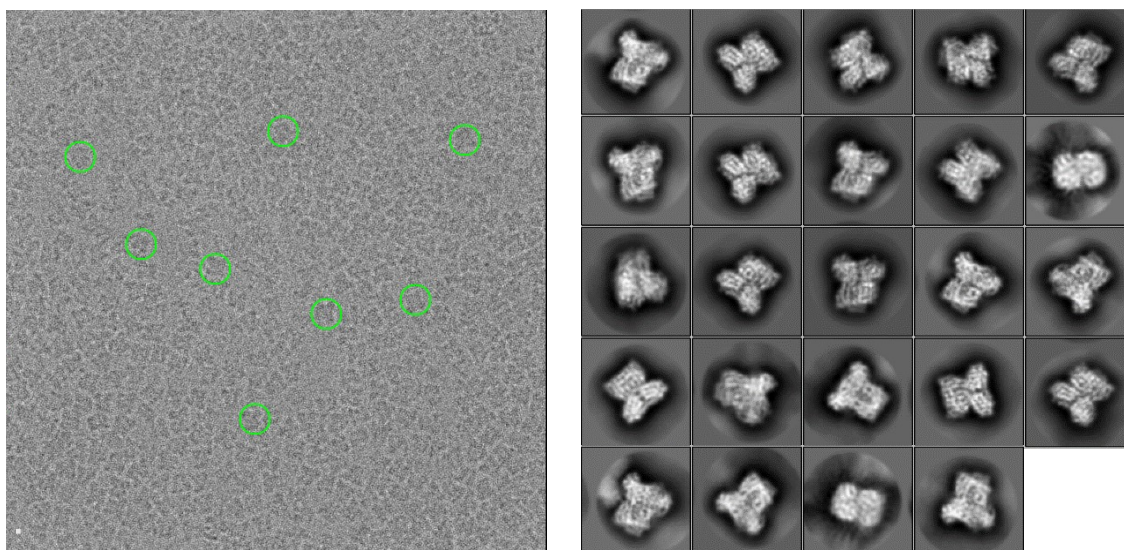
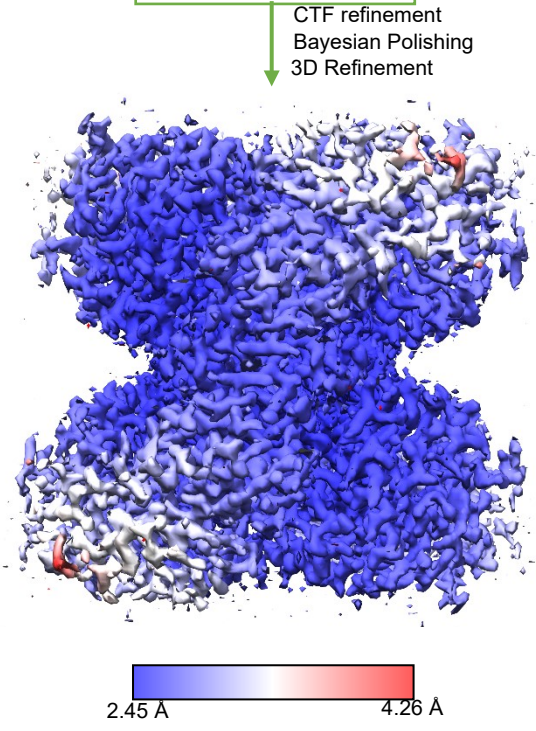
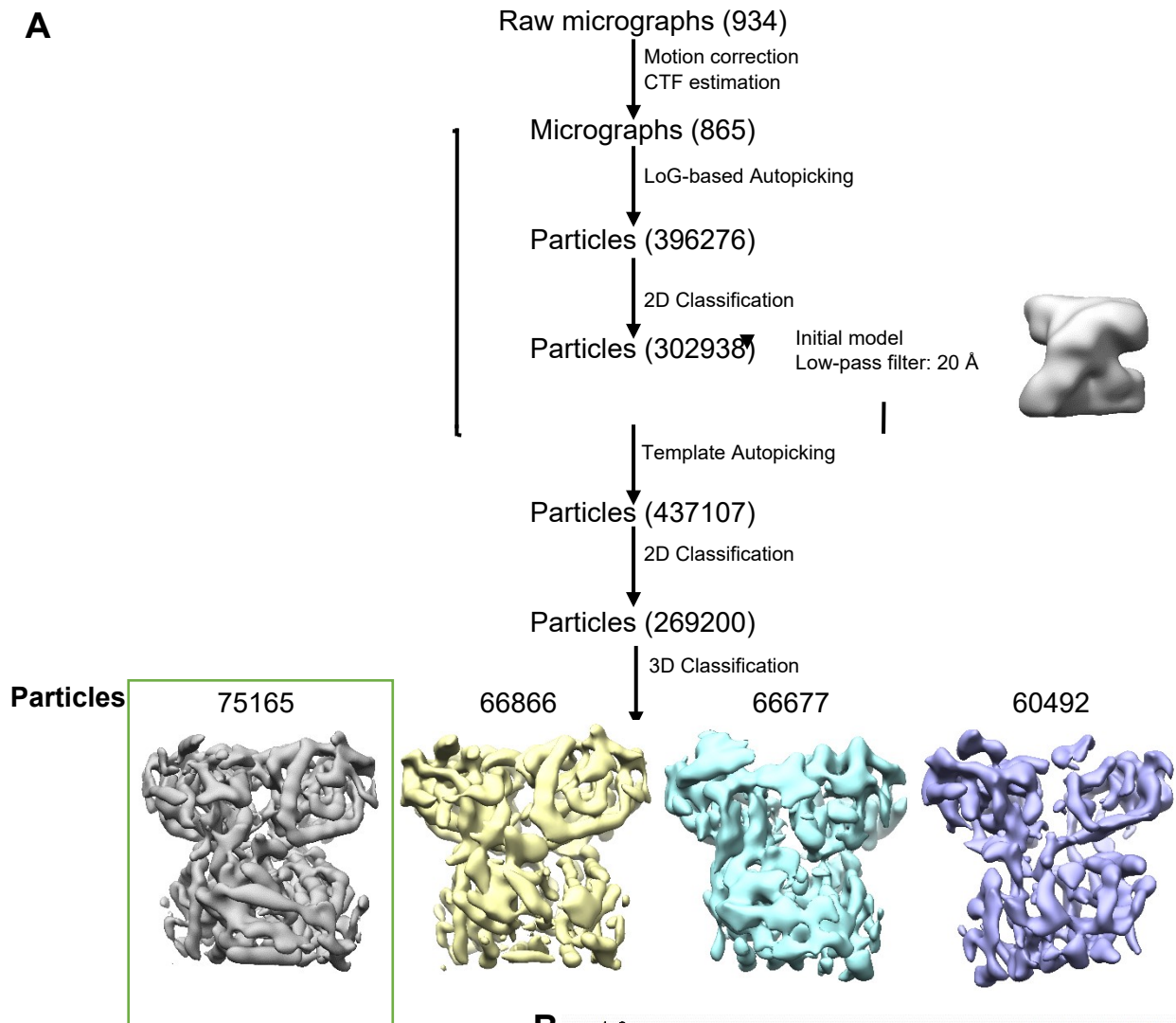


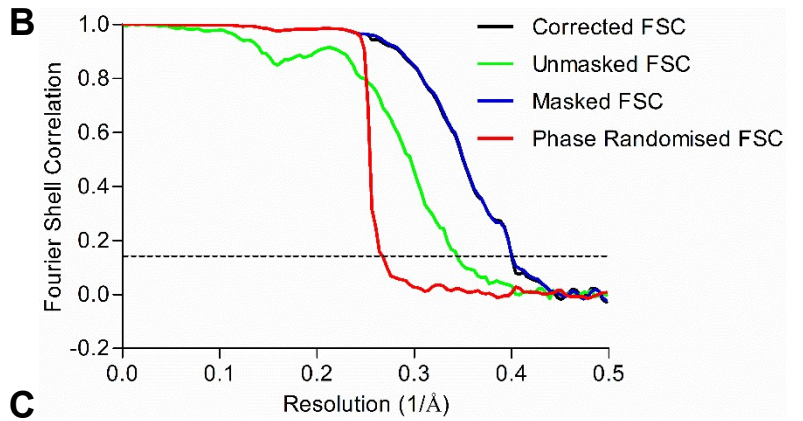
Figure S5. Cryo-EM data processing. (A) Representative micrograph of GabD collected from an UltrAuFoil R1.2/1.3, 300 mesh gold grid. Green circles highlight a few picked particles from the micrograph. White scale bar is 150 Å. Total micrographs collected = 934. (B) 2D class averages of GabD used for 3D classification.

SUPPLEMENTARY INFORMATION

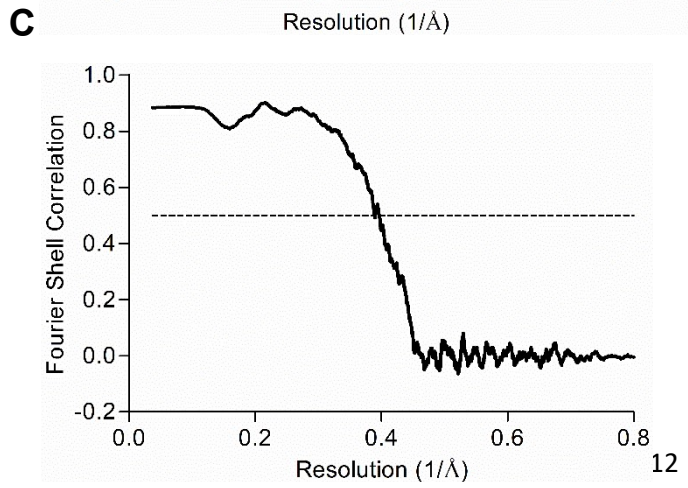
A



B



C



SUPPLEMENTARY INFORMATION

Figure S6. Data processing and refinement workflow of *R/GabD* Cryo-EM structure. (A) Data processing steps from initial data collection to final reconstruction. Numbers in Parentheses refer to total number of micrographs or particles. Final reconstruction coloured by local resolution as estimated through Relion. (B) Fourier Shell Correlation (FSC) between the two independently refined maps. Horizontal dotted black line, FSC=0.143 (C) FSC between the final map reconstruction and *R/GabD* model. Horizontal dotted black line, FSC=0.5.

SUPPLEMENTARY INFORMATION

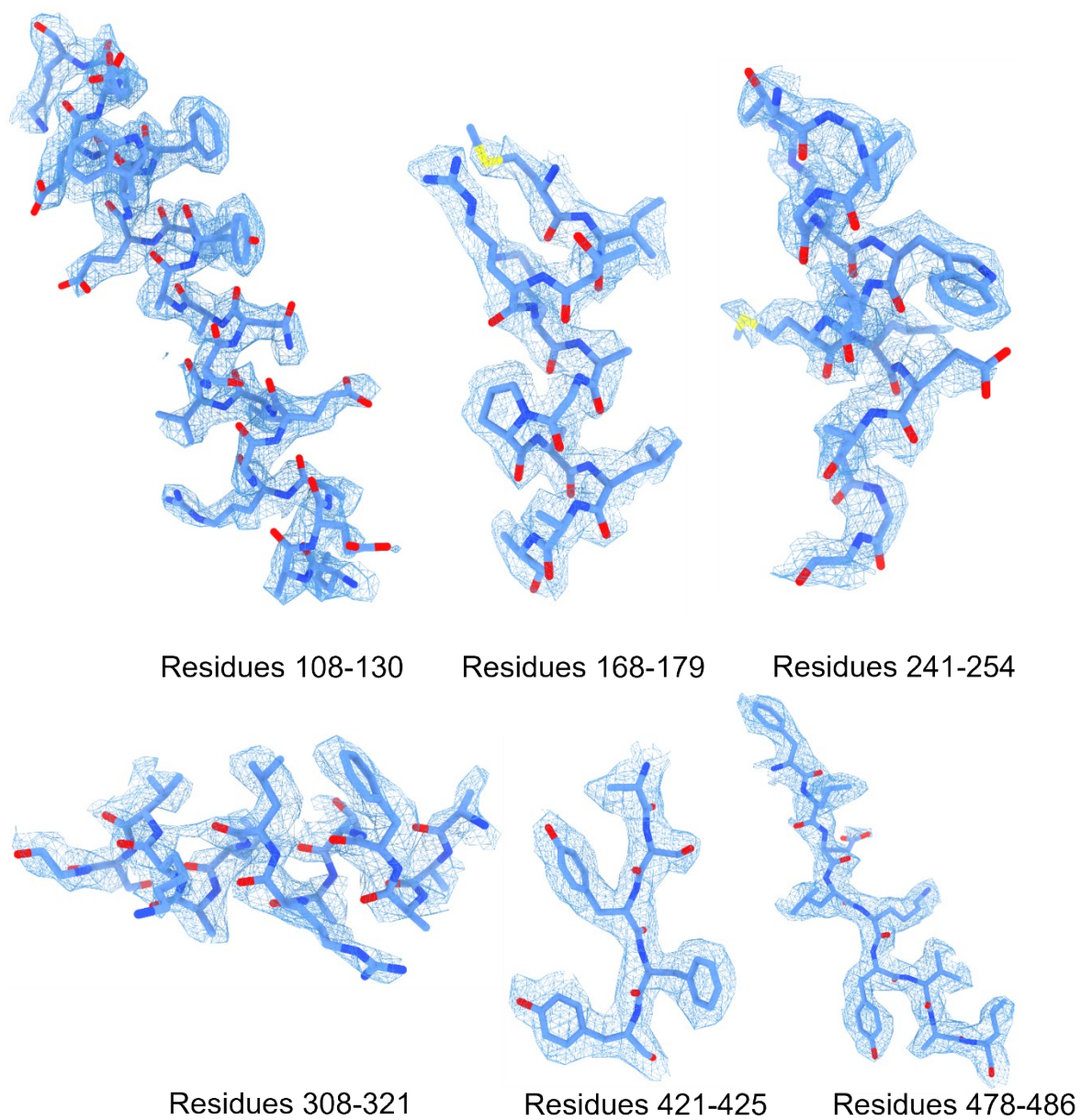


Figure S7. Representative electron densities of different regions of *R/GabD*. Densities carved at a 1.4-1.6 Å radius around residues at a threshold of 0.04-0.05.

SUPPLEMENTARY INFORMATION

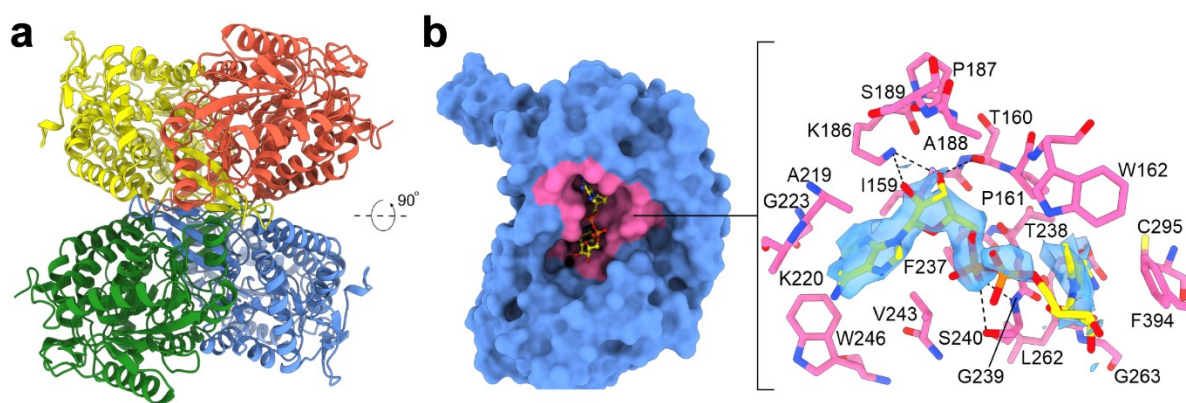


Figure S8. Cryo-EM structure of *R*/GabD•NADH complex. a) Quaternary structure of *R*/GabD•NADH complex shown in cartoon representation. b) Surface representation of *R*/GabD•NADH complex showing the NADH binding site (left); residues lining the NADH binding site are shown in stick format. Blue mesh is the density map at a threshold of 0.037.

SUPPLEMENTARY INFORMATION

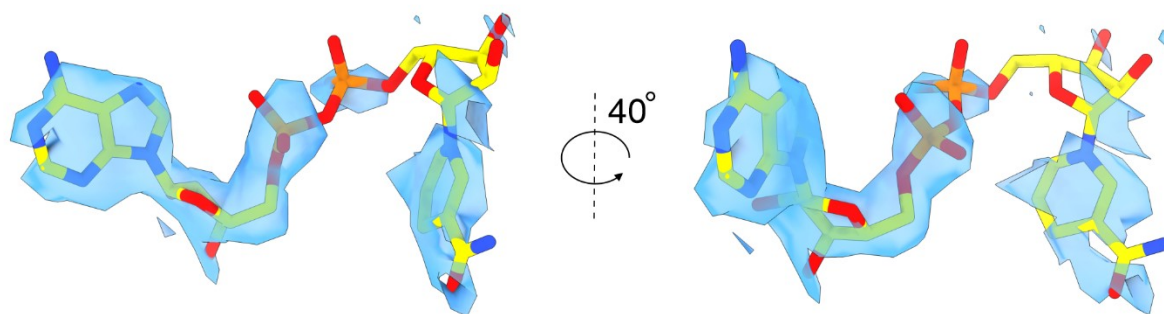


Figure S9. Ligand density for NADH bound to *R/GabD*. While the density is well defined for the adenosine group and attached phosphate, the density is poor for the second phosphate and nicotinamide group. The density is almost totally absent for the phosphate-to-nicotinamide linking ribose.

SUPPLEMENTARY INFORMATION

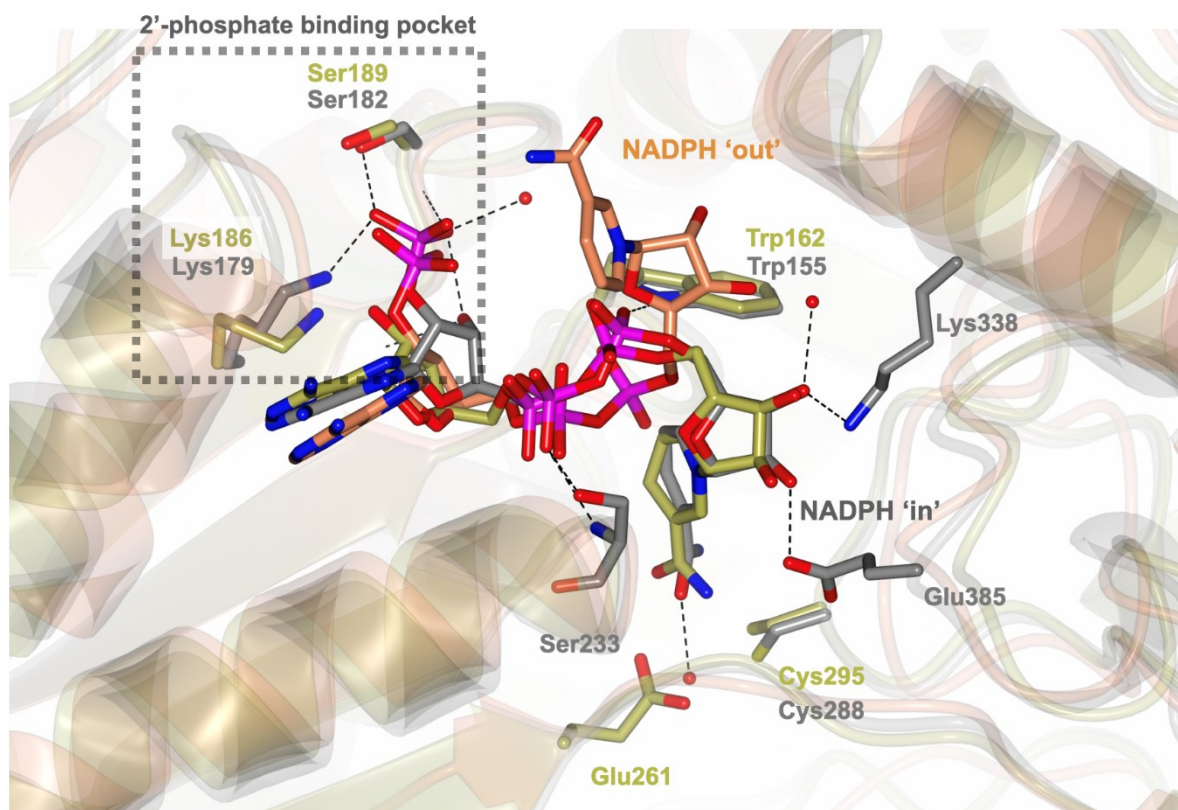


Figure S10. Superposition of crystal structures of binary complexes of *R/GabD*•NADH (in gold), *E. coli* SSADH•NADPH (PDB: 3JZ4) in grey and *E. coli* lactaldehyde dehydrogenase•NADPH (PDB: 2ILU) in coral showing the 2'-phosphate binding pocket and different conformations of the nicotinamide ring of NAD(P)H cofactor in aldehyde dehydrogenases.

SUPPLEMENTARY INFORMATION

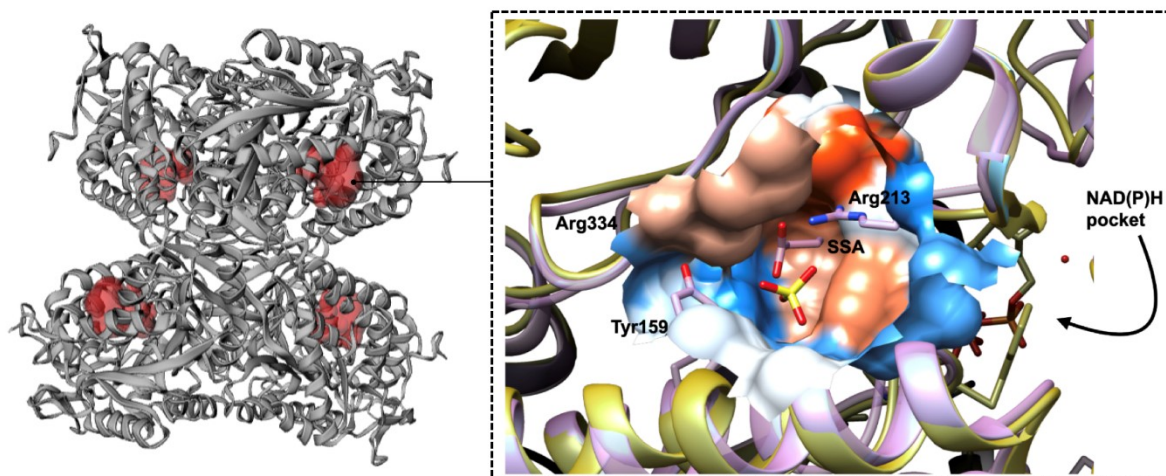


Figure S11. Imprint of sulfonate substrate binding pockets depicted in *R/GabD* tetramer (in grey, left), as computed by CastP 3.0 server shown in red. Overlay of *R/GabD* substrate (SLA) binding pocket with the crystal structure of human SSADH with bound succinic semialdehyde (SSA) (pdb: 2w8q). *R/GabD* and human SSADH are shown in yellow and purple, respectively, and residues in the vicinity of bound SSA are shown in purple.

SUPPLEMENTARY INFORMATION

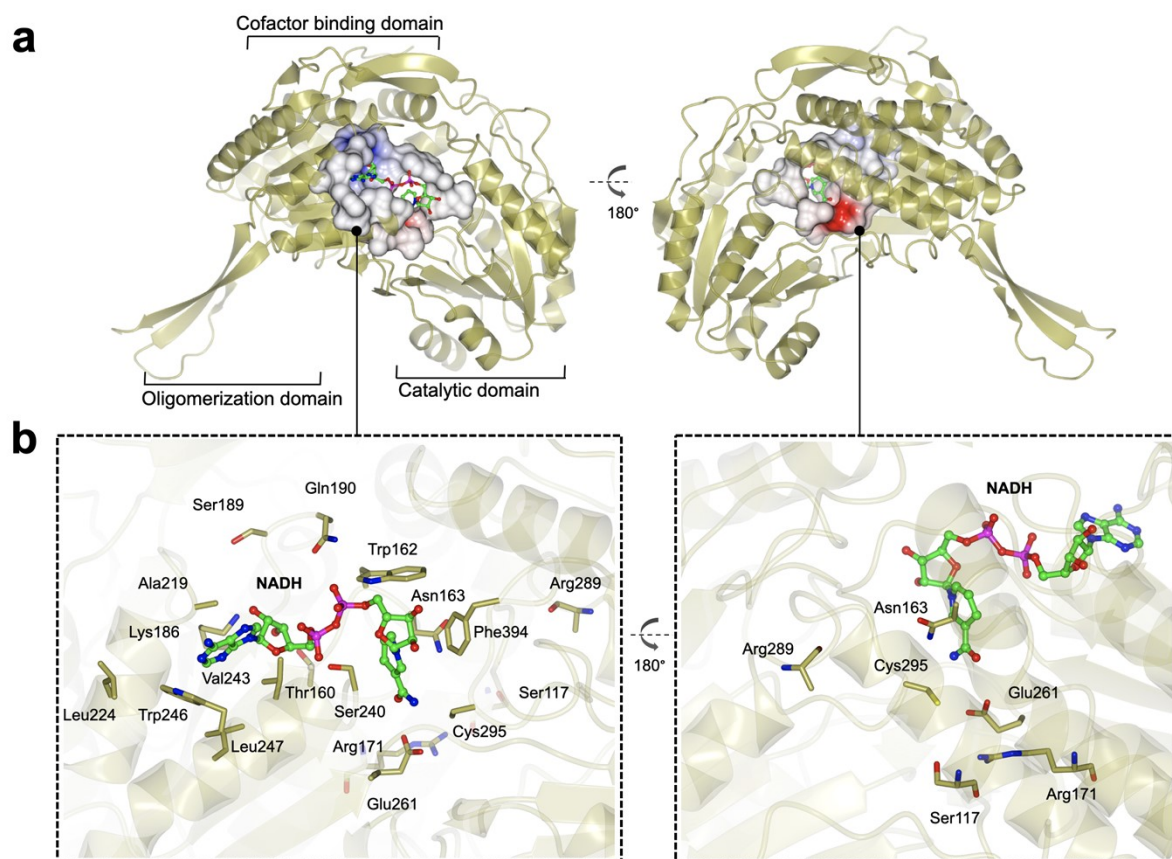


Figure S12. Additional views of NADH binding pocket from front and back. a) *RiGabD* subunit showing the three-domain architecture and the two views of the NADH binding pocket represented as the electrostatic potential of the residues lining the pocket. The red region corresponds to the catalytic Glu-Cys dyad. **b)** Close up view of the NAD(H) binding pocket depicting the residues lining the NAD(H) and SLA binding pockets.

SUPPLEMENTARY INFORMATION

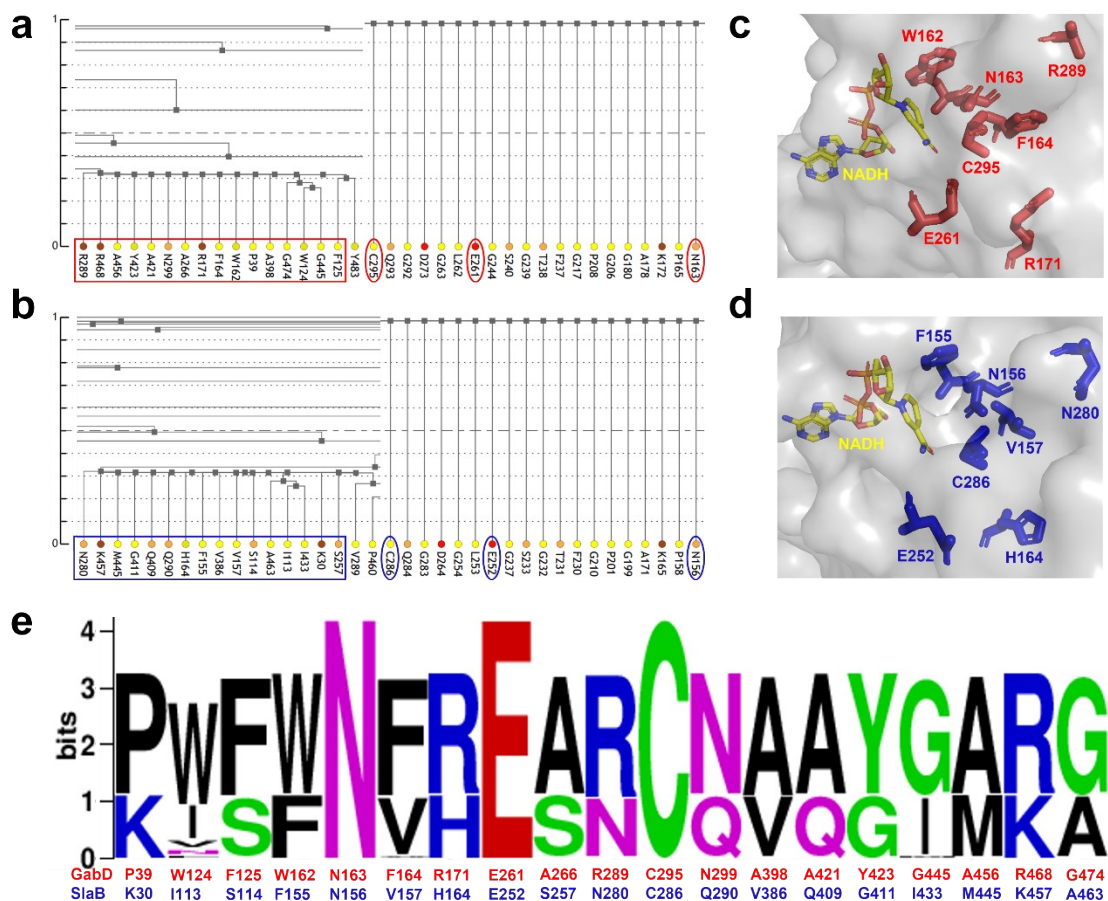


Figure S13. Co-evolution analysis of SLADH. Snapshot of cluster tree diagrams of **a)** *Rhizobium leguminosarum* GabD and **b)** *Bacillus megaterium* SlaB, generated by CoeViz.¹⁴ The cluster trees were generated using the complete linkage hierarchical clustering algorithm based on Mutual Information (MI), with scores are converted to distances by 1-score transformation.¹⁵ Circled residues are conserved residues involved in the reaction mechanism. Boxed residues are identified as co-evolving with R171/R289 (*Ri*GabD) and H164/N280 (*Bm*SlaB). Location of conserved mechanism residues, and co-evolved residues proximal to the proposed SLA binding site shown on the **c)** cryo-EM structures of *Rhizobium leguminosarum* GabD and **d)** the AlphaFold2 predicted structure of *Bacillus megaterium* SlaB. **e)** Sequence logo of conserved mechanism residues and residues co-evolving with R171/R289 (*Ri*GabD) and H164/N280 (*Bm*SlaB), generated by WebLogo (<https://weblogo.berkeley.edu/logo.cgi>)

SUPPLEMENTARY INFORMATION

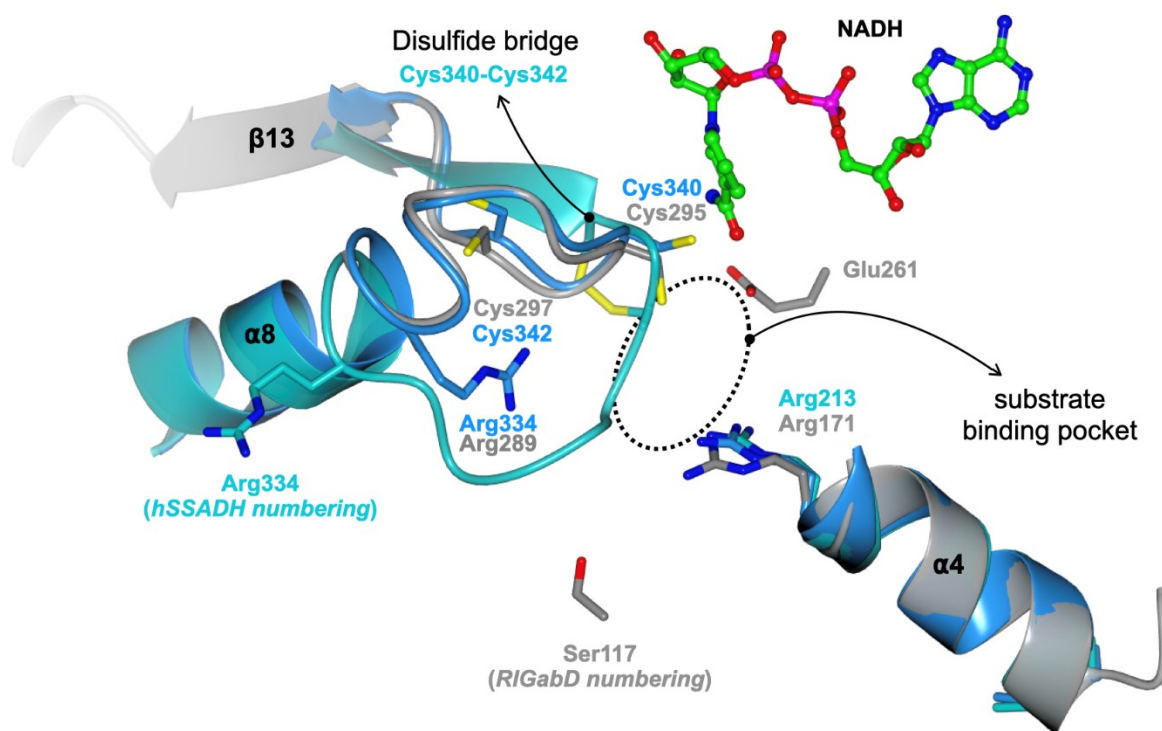


Figure S14. Overlay and close-up view of active site of SLA dehydrogenase *R/GabD* compared to human SSADH in oxidized 'closed' (PDB: 2W8N, shown in turquoise) and reduced 'open' (PDB: 2W8O, shown in blue) forms. In hSSADH the loop containing the catalytic cysteine nucleophile can undergo redox modulation accompanied with large loop position changes due to reversible disulfide bridge formation with adjacent cysteine. In GabD, the equivalent loop is present in the 'open' form with active site residues poised for substrate binding and catalysis. NB: in our structure of *R/GabD* the side-chain of Arg289 was disordered and was not modelled.

SUPPLEMENTARY INFORMATION

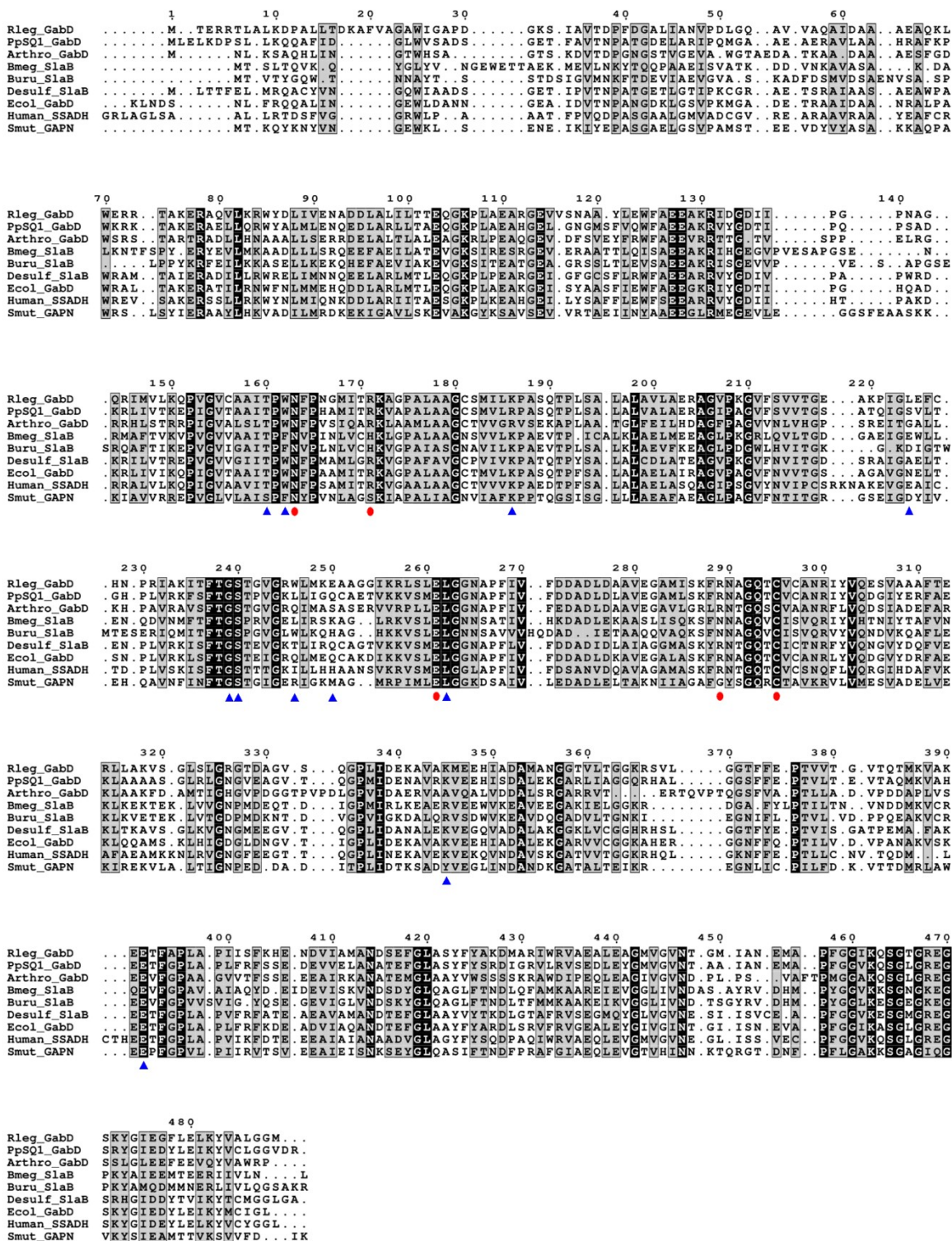


Figure S15. Multiple sequences alignment of SlaB/GabD proteins including SSADH (*E. coli*) and SLA dehydrogenases (*Arthrobacter* sp., *Rhizobium leguminosarum*, *Pseudomonas putida*, *Herbaspirillum seropedicae*, *Bacillus megaterium*, *Bacillus urumqiensis*). Secondary structural elements are annotated from the structure of *E. coli* SSADH (PDB 3JZ4). Annotated by red dots are residues shown to be important in the catalytic mechanism. The residues of NADP⁺ binding to *E. coli* SSADH are annotated by blue arrows.

SUPPLEMENTARY INFORMATION

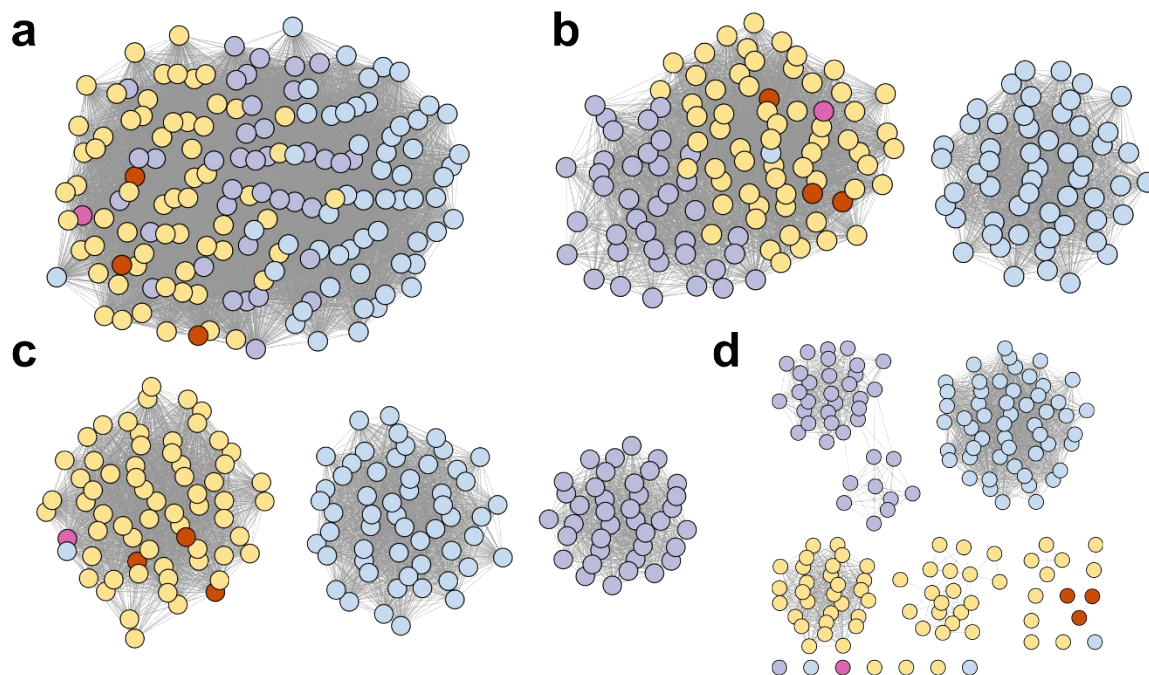


Figure S16. SSNs created using different alignment scores. **a)** At alignment score 75, the sequences form a single cluster. **b)** At alignment score 100, the sequences form two clusters. **c)** At alignment score 125-150, the SSN breaks into three clusters that separate into three main Phyla, with Chloroflexi and Candidatus Dormibacteraeota segregating with Actinobacteria. **d)** At higher alignment threshold (175), the SSN produces a large number of singletons. Colouring scheme: Actinobacteria (yellow), Chloroflexi (red), Candidatus Dormibacteraeota (pink), Firmicutes (light blue), Proteobacteria (light purple).

SUPPLEMENTARY INFORMATION

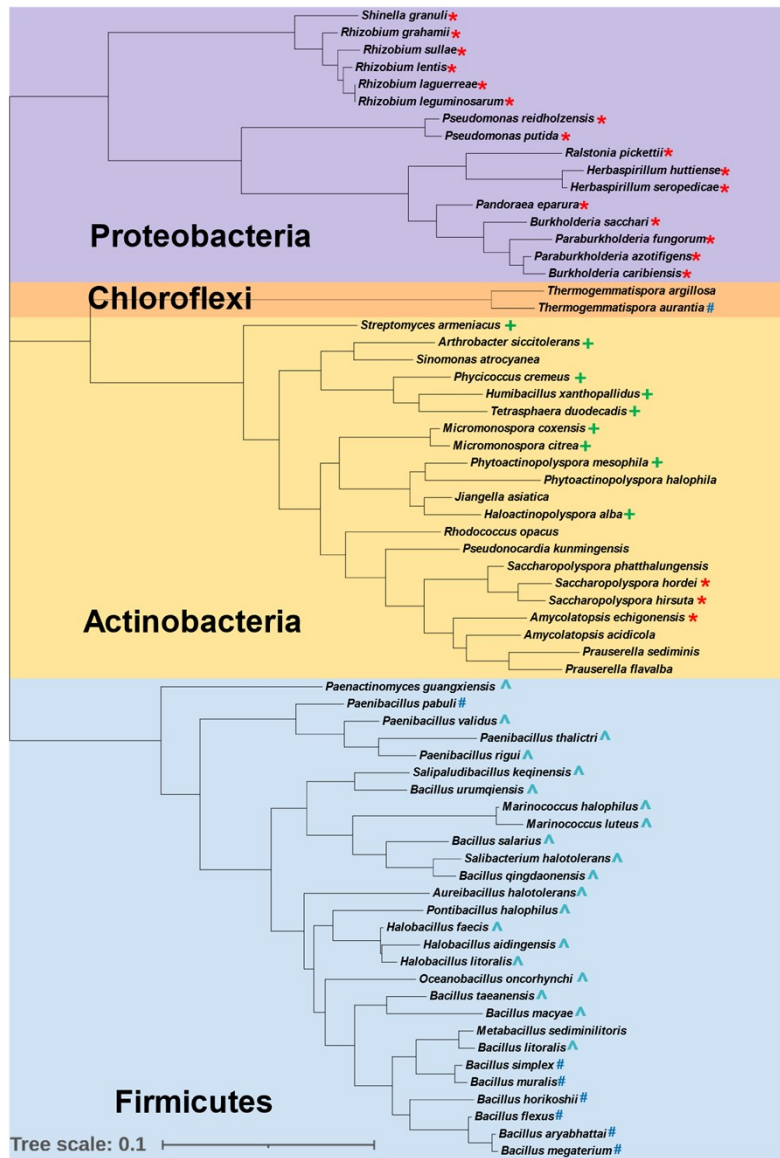


Figure S17. A phylogenetic tree based on 16S ribosomal RNA of organisms from four different phyla containing putative SLADH proteins in 4 different sulfoglycolytic pathways: sulfo-ED (*), sulfo-SFT (#), sulfo-EMP2 (^) and sulfo-EMP3 (+). Note: Candidatus could not be included as its 16S ribosomal RNA data was unavailable.

SUPPLEMENTARY INFORMATION

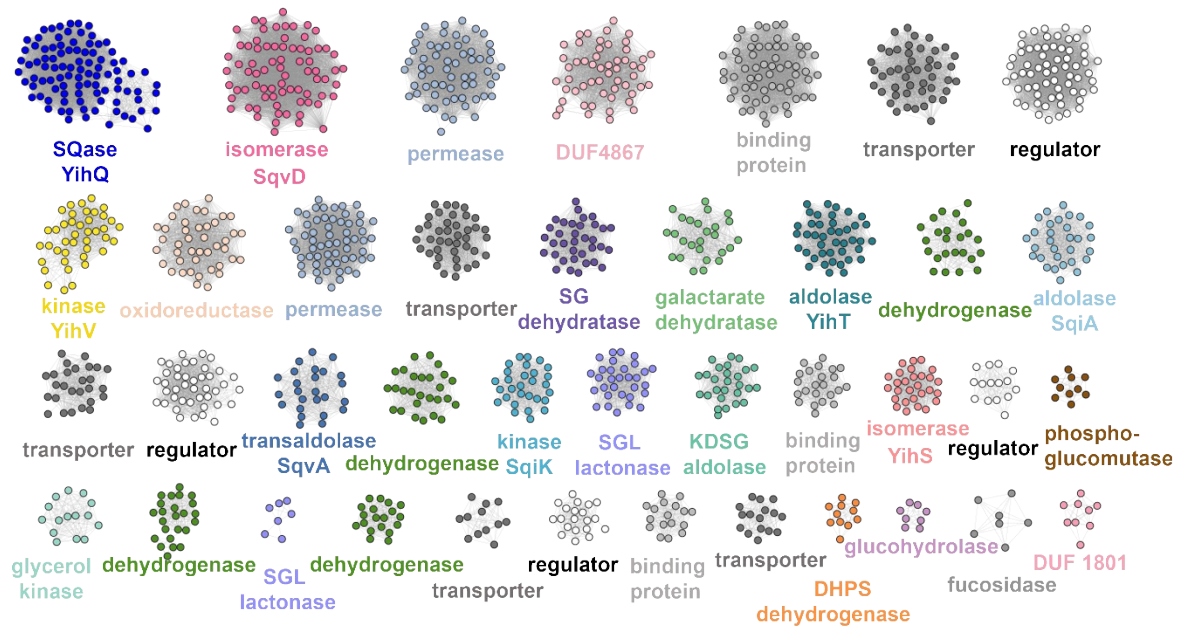


Figure S18. SSN of SLADH neighbors, displayed at alignment score of 50 while the percent identity is 40 and the length of the sequences are between 45 to 1151. Clusters are colored according to their functions. Key proteins from the four sulfoglycolytic pathways are: SQ isomerase (YihS, pink; SqvD, red pink), SQase (blue), aldolase (SqiA, light blue; KDSG aldolase, light green), SF kinase (YihV, yellow PF00294; SqiK, aquamarine PF00365), dehydratase (SG dehydratase, purple PF00920), lactonase (SGL lactonase, light purple), transaldolase (SqvA, dark blue PF00923).

SUPPLEMENTARY INFORMATION

Table S1. Data collection, processing and refinement statistics for *R/GabD*•NADH complex structure

<i>R/GabD</i> complexed with NADH (PDB:8C54)	
Data collection and processing	
Nominal magnification	240,000
Voltage (kV)	200
Electron exposure (e-/Å ²)	50
Defocus range (µm)	-2, -1.75, -1.5, -1.25, -1.0, -0.9, -0.7
Pixel Size (Å)	0.574
Symmetry imposed	D2
Initial number of particles	437,107
Final number of particles	75,165
Map resolution (Å) (FSC threshold)	2.52 (0.143)
Refinement	
Initial model used (PDB)	N/A
Model resolution (Å) (FSC threshold)	2.57 (0.5)
Map sharpening <i>B</i> factor (Å ²)	-44.85
<i>Model composition</i>	
Non-hydrogen atoms	14266
Protein residues	1928
Ligands	4
<i>RMSDs</i>	
Bond lengths (Å)	0.003
Bond angles (°)	0.488
<i>Ramachandran plot</i>	
Outliers (%)	0
Allowed (%)	2.2
Favoured (%)	97.8
<i>Validation</i>	
Rotamer outliers (%)	3.43
MolProbity	1.82
Clash score	6.47
Model vs. Data (CC mask)	0.84

SUPPLEMENTARY INFORMATION

References

1. 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*, 331-332.
2. Rohou, A.; Grigorieff, N., CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **2015**, *192*, 216-221.
3. Scheres, S. H. W., RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180*, 519-530.
4. 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*.
5. Zivanov, J.; Nakane, T.; Scheres, S. H. W., A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr* **2019**, *6*, 5-17.
6. 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.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **2019**, *75*, 861-877.
7. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. Sect. D* **2010**, *66*, 486-501.
8. 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.; Jain, S.; Lewis, S. M.; Arendall Iii, W. B.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293-315.
9. McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E. M., Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **2011**, *67*, 386-394.
10. 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*, 70-82.
11. 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*, 1019-37.
12. Zhao, S.; Sakai, A.; Zhang, X.; Vetting, M. W.; Kumar, R.; Hillerich, B.; San Francisco, B.; Solbiati, J.; Steves, A.; Brown, S.; Akiva, E.; Barber, A.; Seidel, R. D.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A.; Jacobson, M. P., Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *eLife* **2014**, *3*, e03275.
13. 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*, 2498-2504.
14. Baker, F. N.; Porollo, A., CoeViz: a web-based tool for coevolution analysis of protein residues. *BMC Bioinformatics* **2016**, *17*, 119.
15. Buslje, C. M.; Santos, J.; Delfino, J. M.; Nielsen, M., Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. *Bioinformatics* **2009**, *25*, 1125-1131.