SUPPLEMENTARY INFORMATION

Chiral recognition of 2,3-dihydroxypropanesulfonate by bacterial transport proteins adapted to distinct ecological niches

Table of Contents

SUPPLEMENTARY INFORMATION

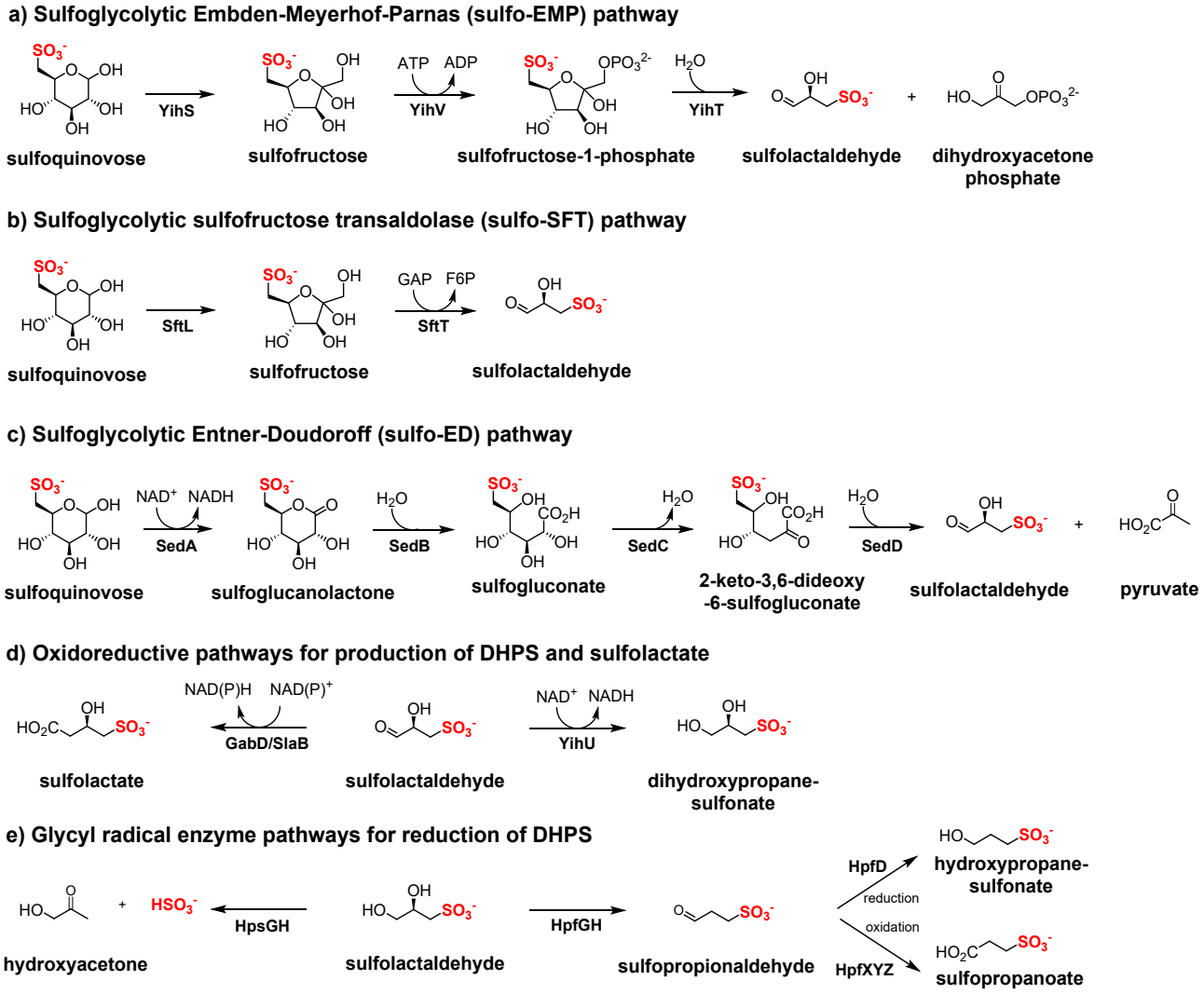


Figure S1. Summary of sulfoglycolytic pathways to sulfoacetaldehyde, formation of DHPS, and glycyl radical enzyme reduction of DHPS. a) Sulfoglycolytic Embden-Meyerhof-Parnas (sulfo-EMP) pathway, b) sulfoglycolytic sulfofructose transaldolase (sulfo-SFT) pathway, c) sulfoglycolytic Entner-Doudoroff (sulfo-ED) pathway, d) pathways for oxidation or reduction of sulfolactaldehyde, e) glycyl radical enzyme pathways for reduction of DHPS.

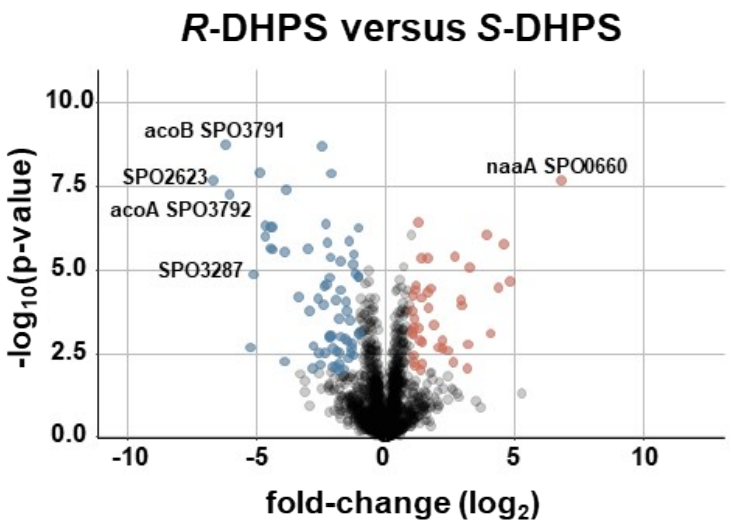


Figure S2. Comparative proteomics of *Ruegeria pomeroyi* DSS-3 reveals differences in growth on *R*- and *S*-DHPS. Volcano plot showing proteins increased in abundance when grown on *R*-DHPS versus *S*-DHPS.

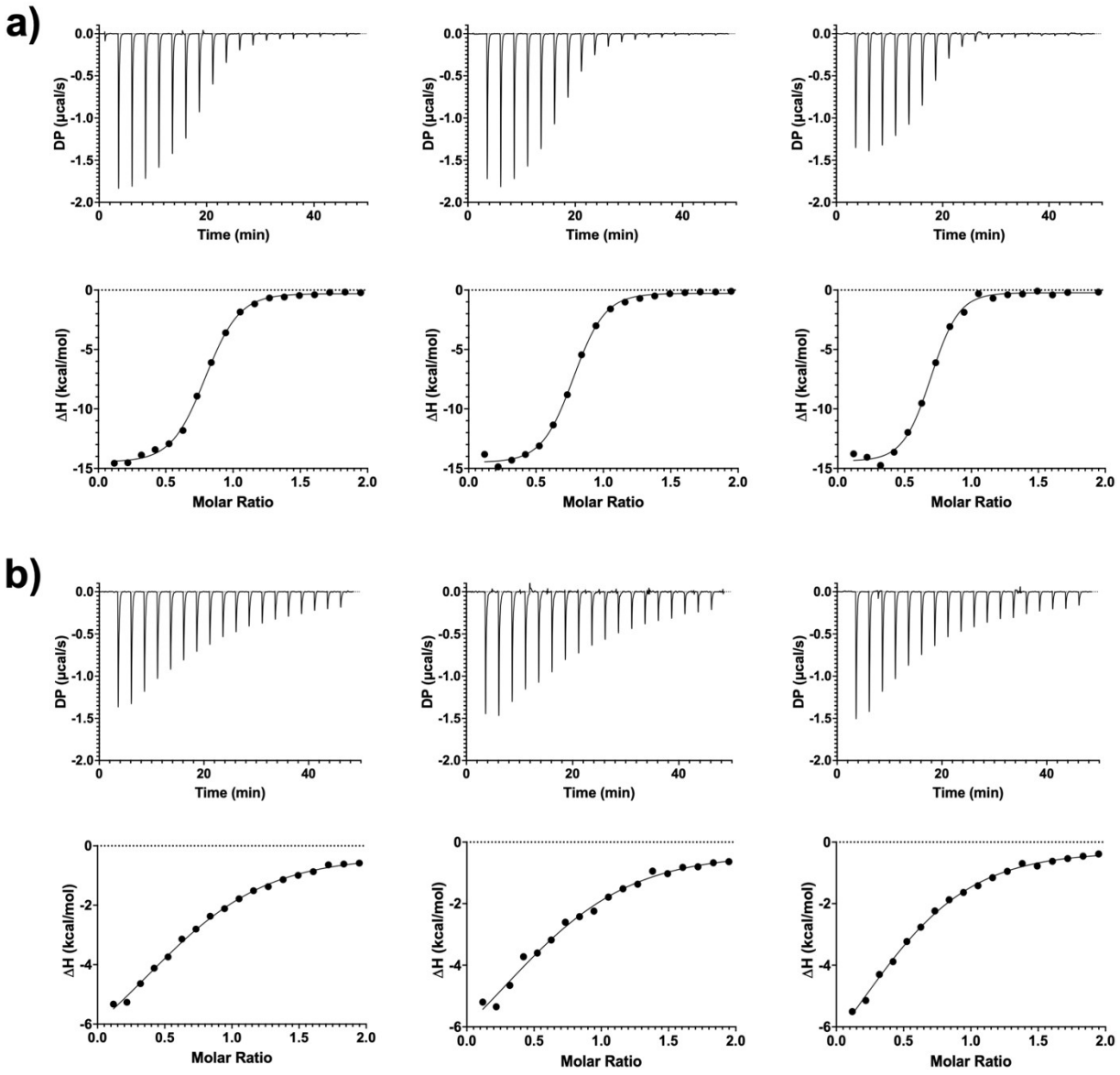


Figure S3. Isothermal calorimetry data of *Rp*HpsK titrated with hydroxypropanesulfonate and isethionate. Three independent replicates showing calorimetric traces and fitted binding isotherms for **a**) hydroxypropanesulfonate and **b**) isethionate.

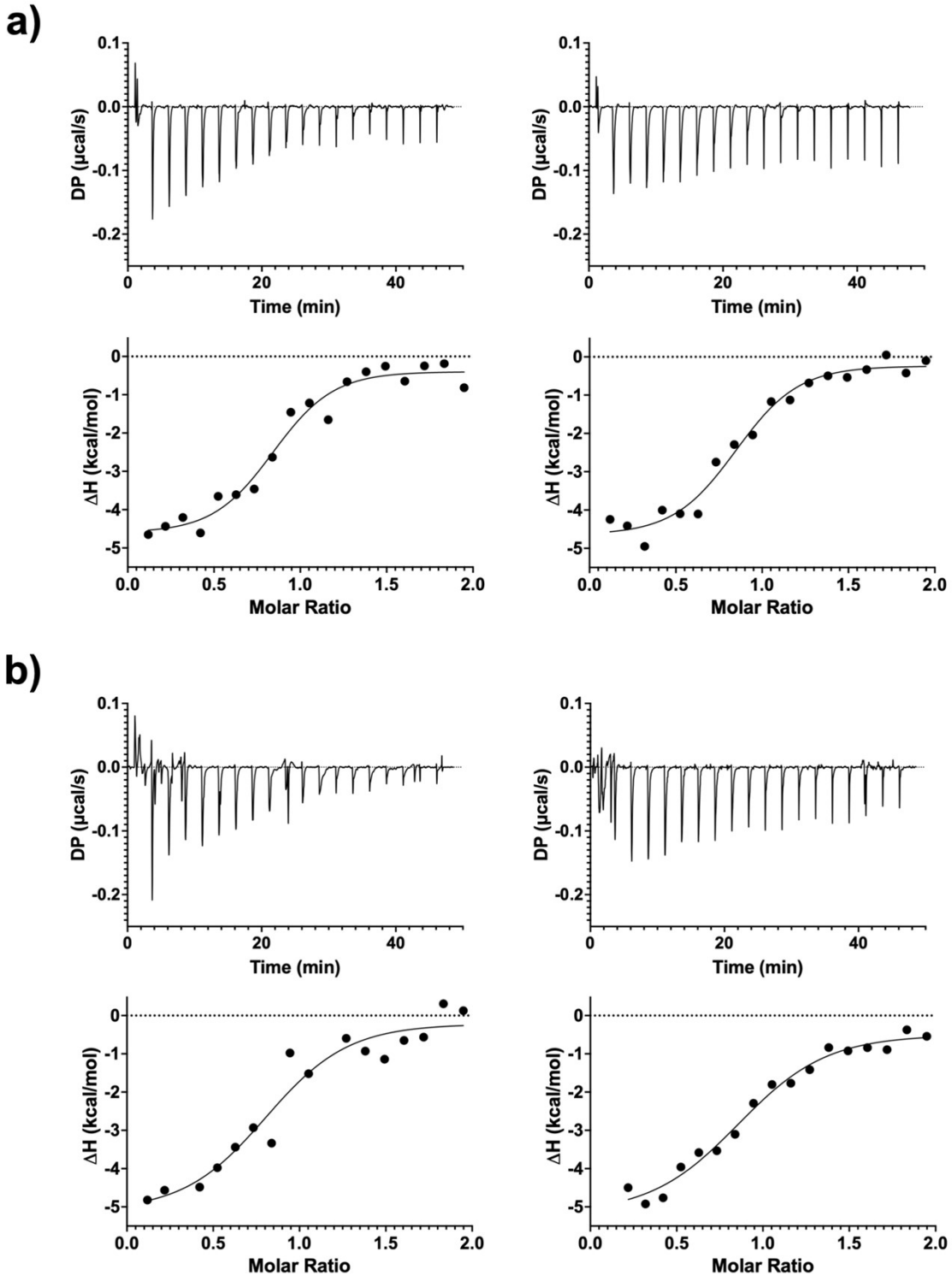


Figure S4. Isothermal calorimetry data for displacement isothermal calorimetry of *Rp*HpsK titrated with DHPS enantiomers. Two independent replicates showing calorimetric traces and fitted binding isotherms for **a**) *R*-DHPS and **b**) *S*-DHPS, each in the presence of 200 μM HPS.

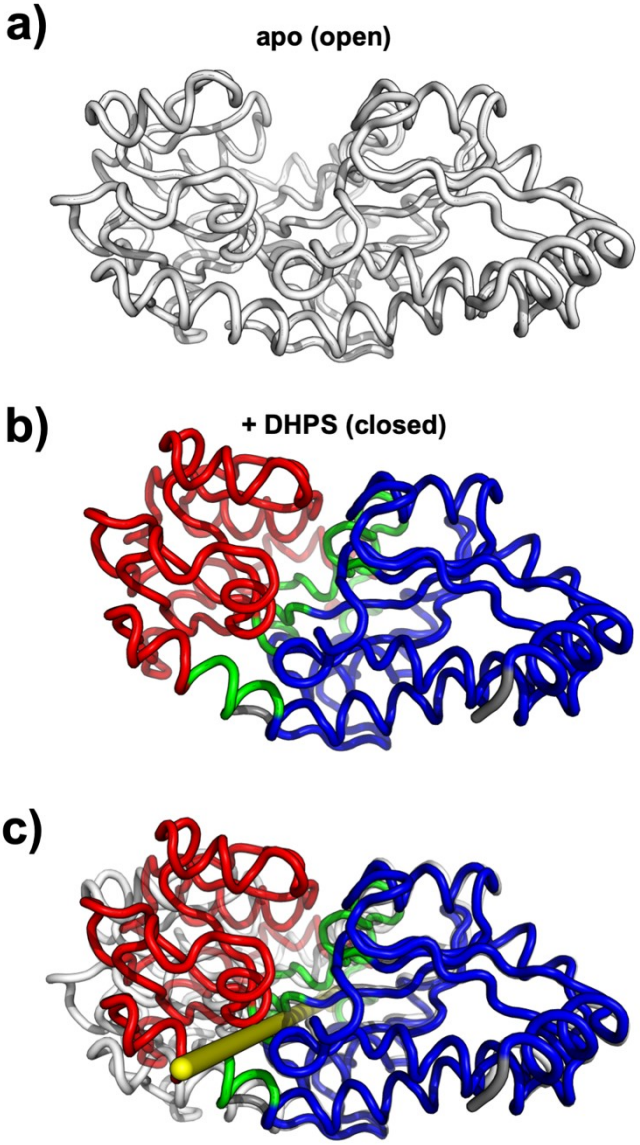


Figure S5. Domain movement analysis of *Rp*HpsK. a) Apo-*Rp*HpsK structure in the open conformation (chain A from crystal form 1), **b**) *Rp*HpsK•*S*-DHPS complex structure with the closed conformation with domains coloured in blue (fixed domain), red (moving domain), and green (bending regions), **c**) Superposition of panels (**a**) and (**b**) with the domain rotation axis indicated by a yellow cylinder. DynDom server[1, 2] was used for domain movement analysis.

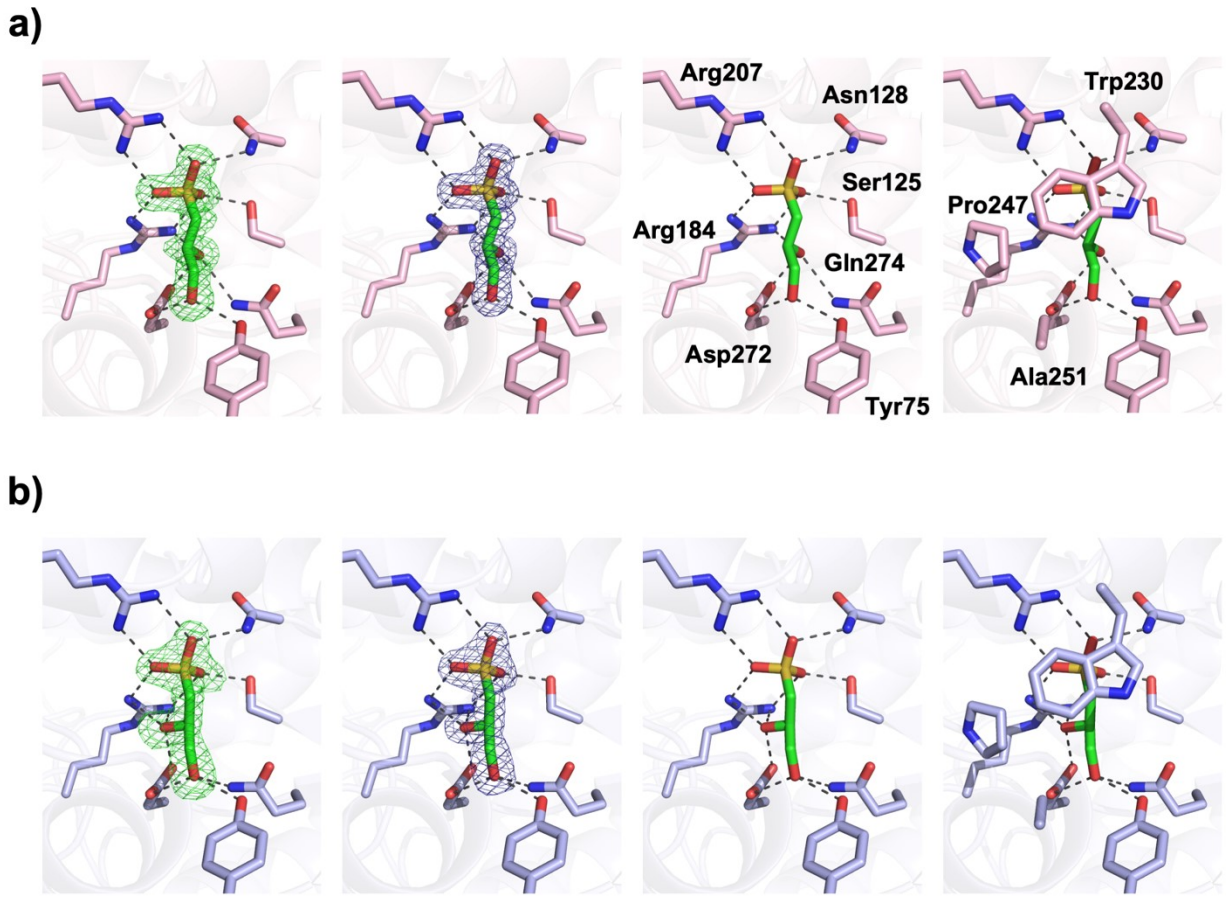


Figure S6. DHPS-binding mode in *Rp*HpsK–*S*-DHPS (**a**) and *R*-DHPS (**b**). From left to right, *F*o–*F*c map (green mesh, contoured at 3 σ) before modelling in, 2*F*o–*F*c map (blue mesh at 1 σ) in the final refinement, hydrogen-bond network, and hydrophobic interaction in the binding site, respectively.

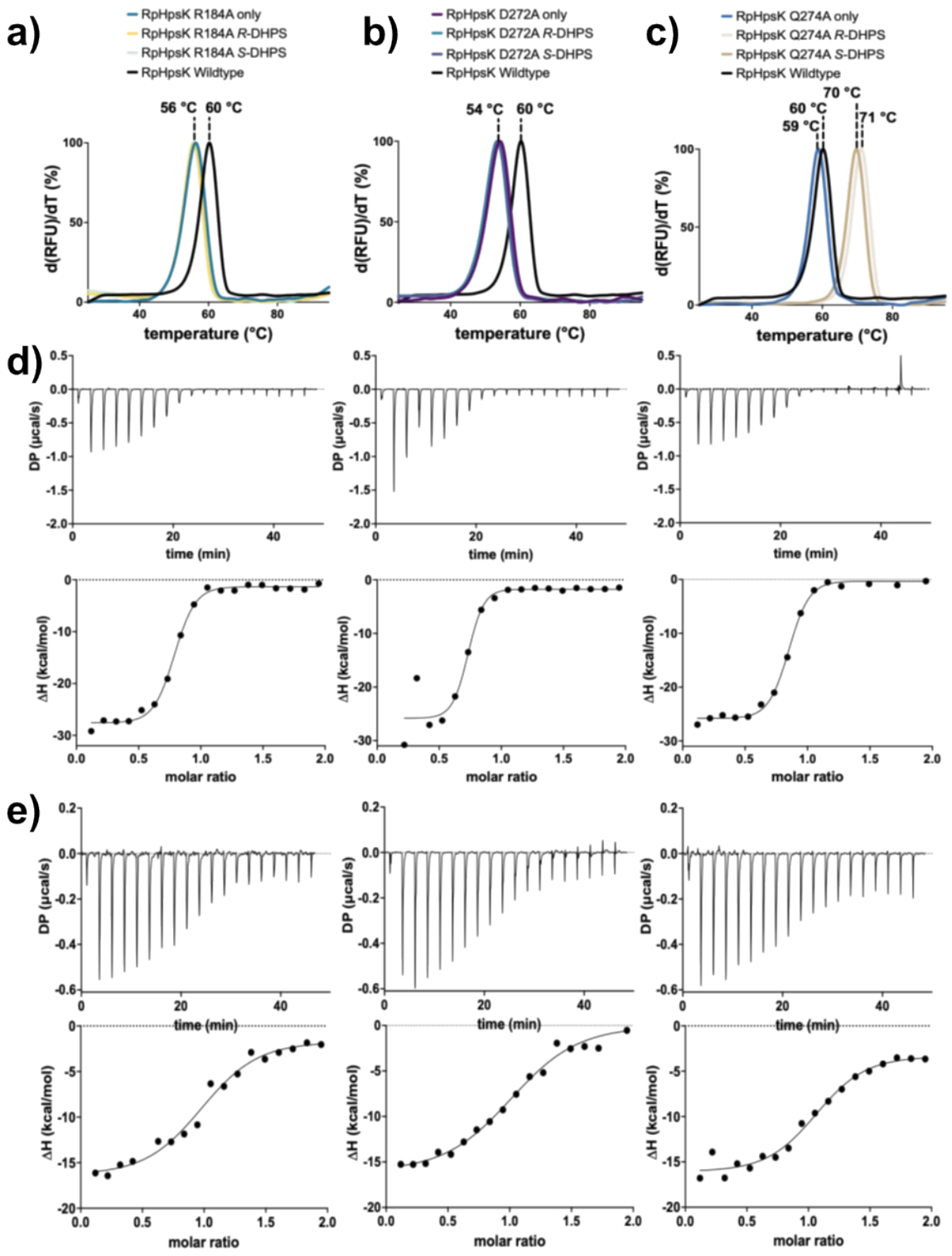


Figure S7. Biophysical characterisation of *Rp*HpsK ligand binding-site mutants. a–c) NanoDSF analysis of melting temperature of *Rp*HpsK mutants in the absence and presence of *R*- and *S*-DHPS: **a)** *Rp*HpsK R184A, **b)** *Rp*HpsK D272A, and c) *Rp*HpsK Q274A. **d,e)** Triplicate ITC experiments for *Rp*HpsK Q274A with d) *R*-DHPS and e) *S*-DHPS. Traces are shown as independent replicates.

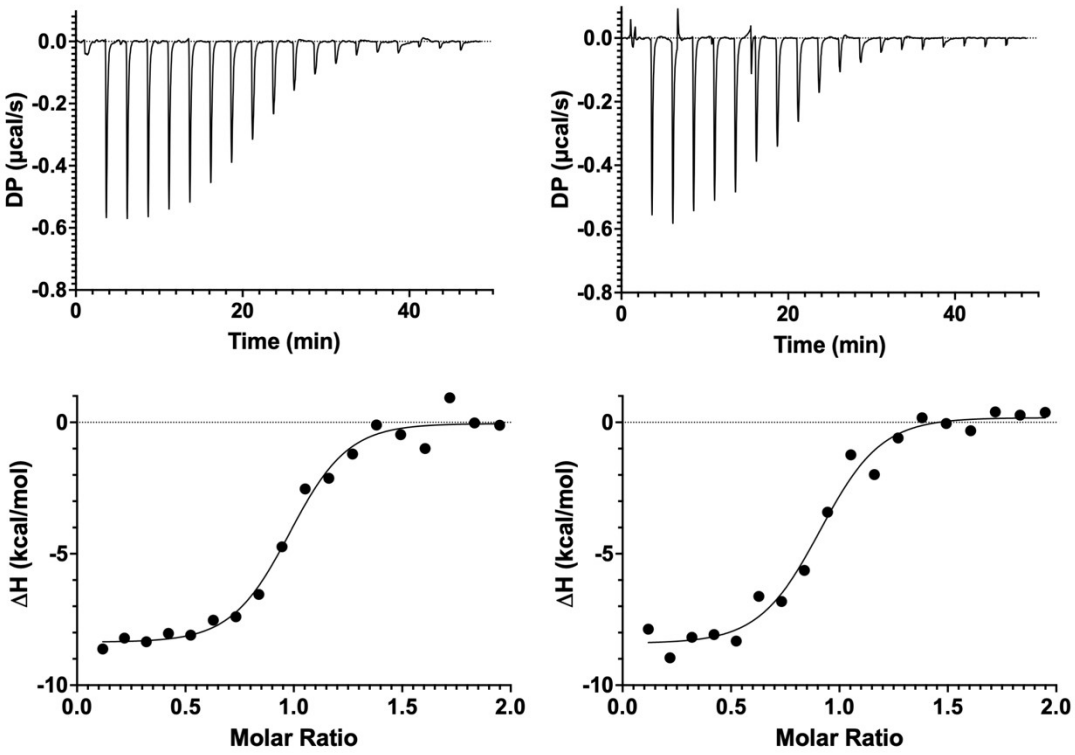


Figure S8. Isothermal calorimetry data for displacement isothermal calorimetry of *Bw*HpsK titrated with *S*-DHPS. Two independent replicates showing calorimetric traces and fitted binding isotherms.

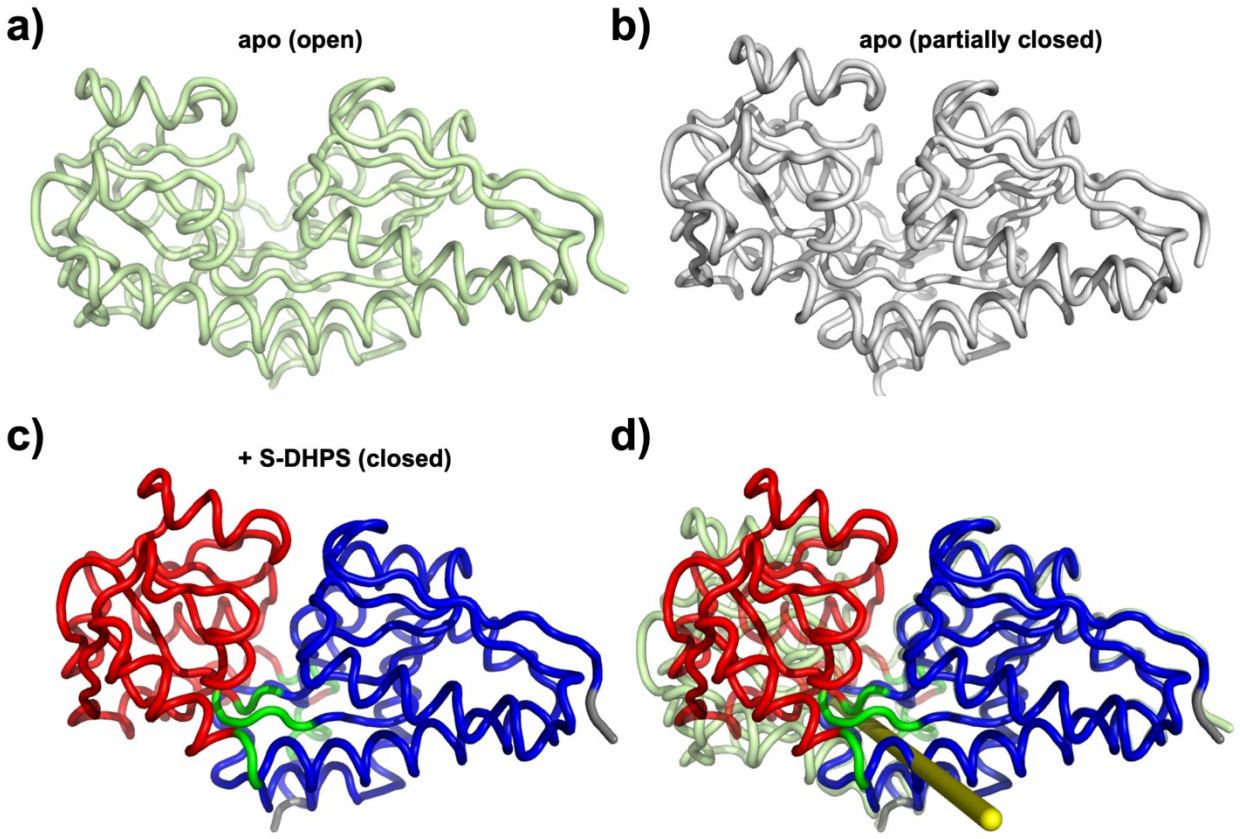


Figure S9. Domain movement analysis of *Bw*HpsK. a, b) Apo-*Bw*HpsK structures in the open conformation (**a**) and partially closed conformation (**b**), **c**) *Bw*HpsK•*S*-DHPS complex structure with the closed conformation with domains coloured in blue (fixed domain), red (moving domain), and green (bending regions), **d**) Superposition of panels (**a**) and (**b**) with the domain rotation axis indicated by a yellow cylinder.

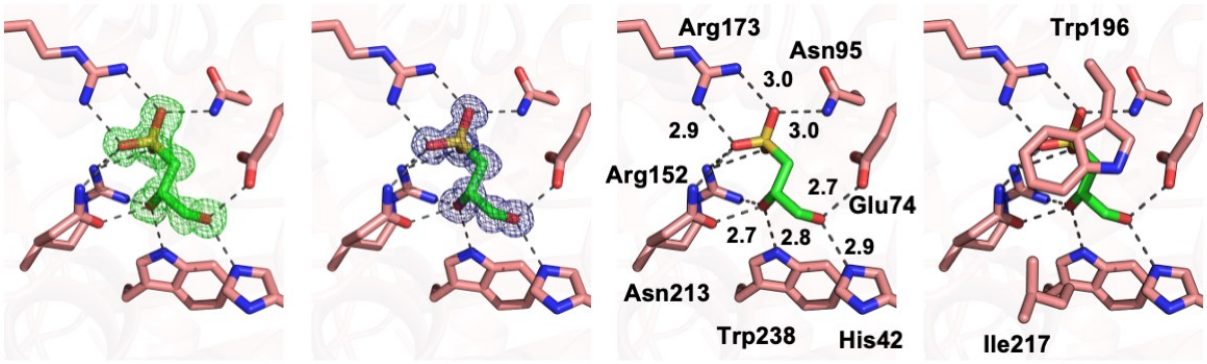


Figure S10. *S*-DHPS-binding mode in *Bw*HpsK. From left to right, *F*o–*F*c map (green mesh, contoured at 3 σ) before modelling in, 2*F*o–*F*c map (blue mesh at 1 σ) in the final refinement, hydrogen-bond network (numbers next to the dashed lines are hydrogen-bond distances in Å), and hydrophobic interaction in the binding site, respectively.

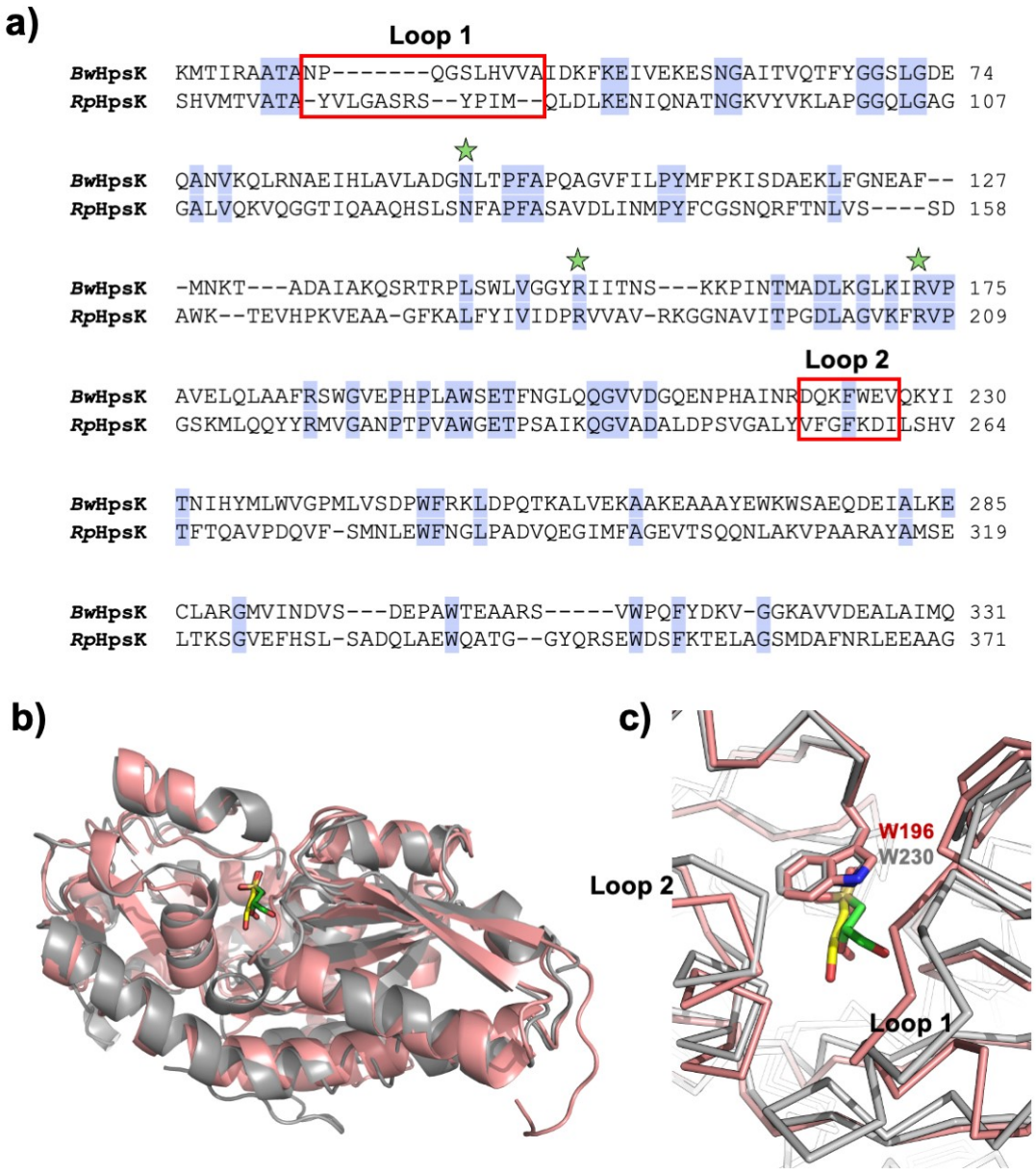


Figure S11. Comparison of *Rp*HpsK and *Bw*HpsK. a) Structure-based sequence alignment of *Rp*HpsK and *Bw*HpsK showing low sequence conservation (20.3% sequence identity). Conserved sulfonate-coordinating residues are marked with green stars, and loops that differ between the two proteins (shown in panel **c**) are highlighted with red boxes. **b**) Overall superposition of the two proteins in complex with *S*-DHPS (*Rp*HpsK shown in grey with *S*-DHPS carbons in green; *Bw*HpsK in salmon with *S*-DHPS carbons in yellow). **c**) Close-up view of the binding site, highlighting differences in the surrounding loops. The side chain of the conserved tryptophan residue (Trp230 in *Rp*HpsK; Trp196 in *Bw*HpsK) that closes the binding pocket is shown. Protein backbones are displayed in ribbon representation.

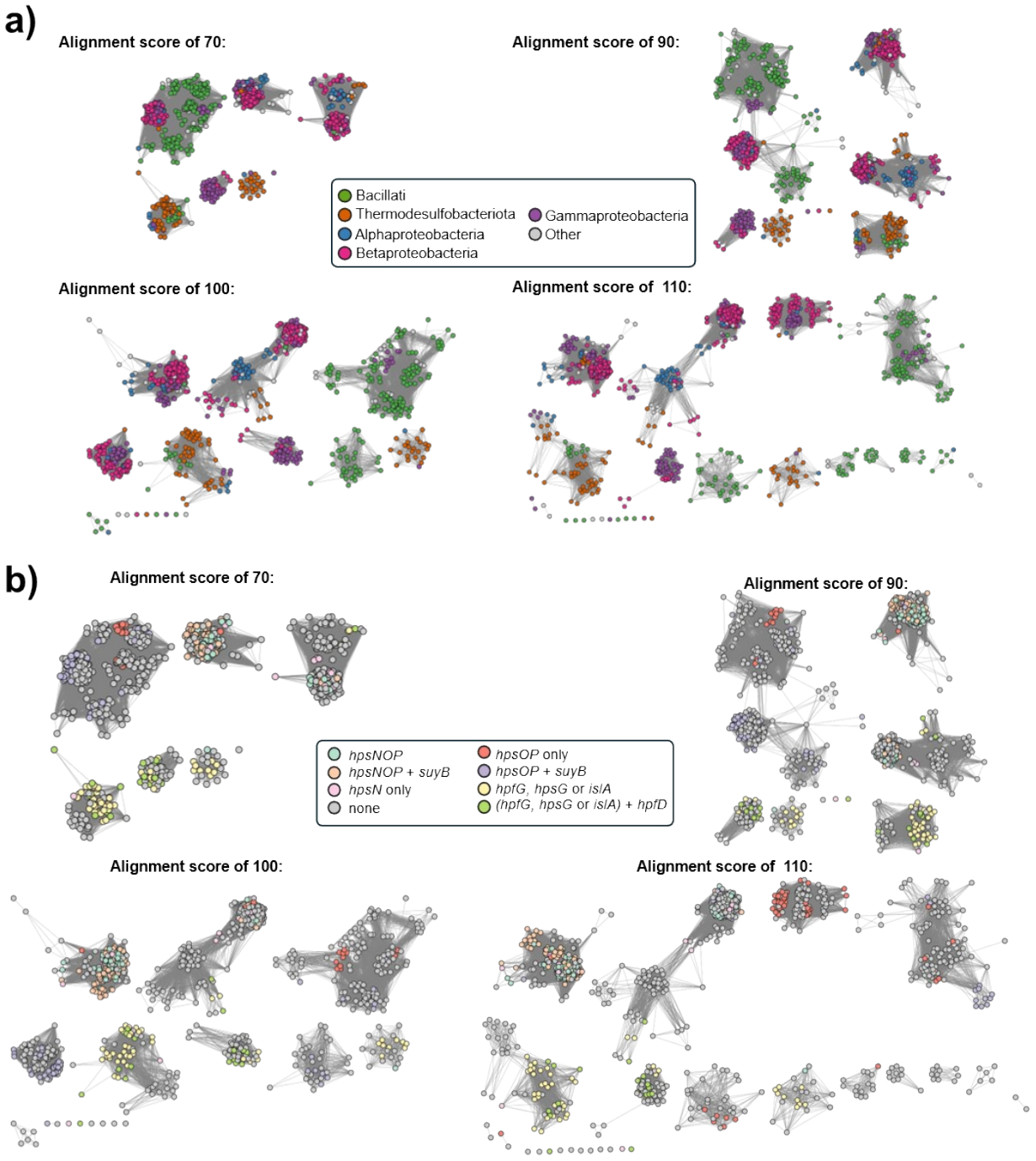


Figure S12. Effect of alignment score on taxonomic and Pfam clustering in sequence similarity network of HpsK homologs. Sequence similarity networks (SSNs) for HpsK homologues were generated using alignment scores of 70, 80 (see main manuscript), 90, 100, and 110. Networks are coloured by **a)** taxonomy and **b)** neighbouring Pfams as shown in the legend. An alignment score of 80 was selected for further analysis based on optimal cluster resolution and retention of biologically meaningful groupings.

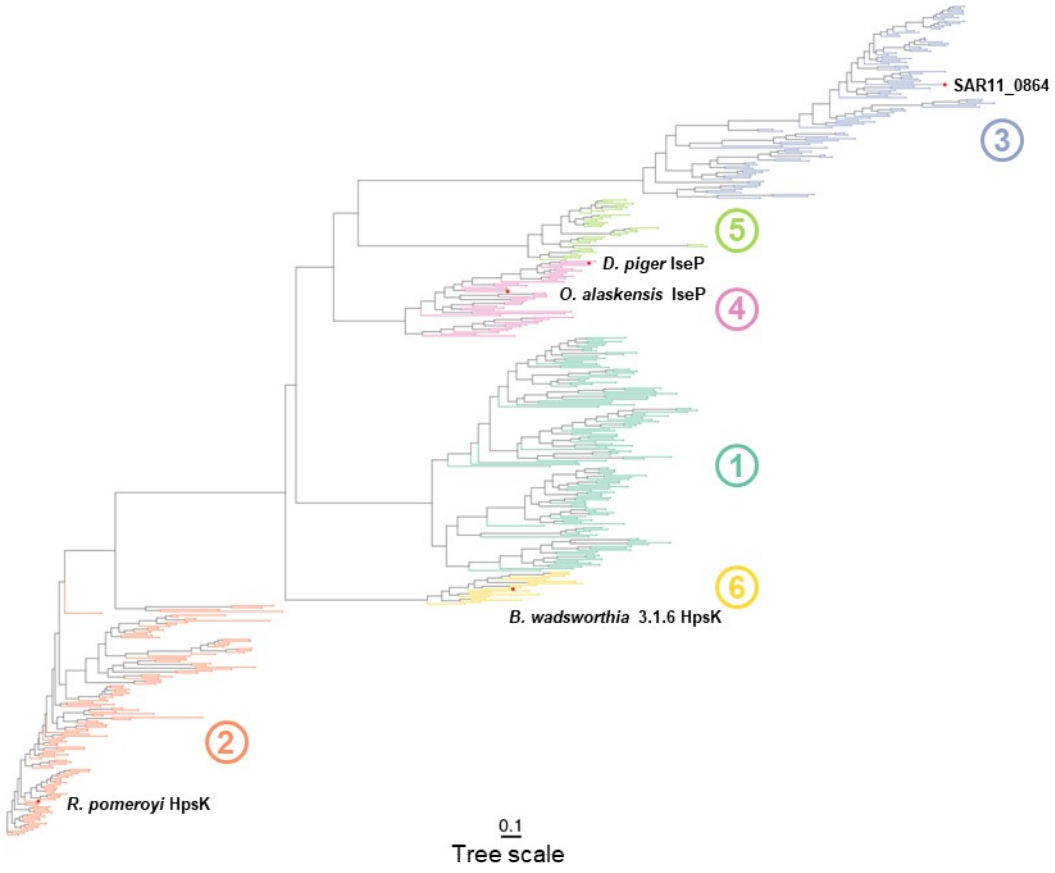


Figure S13. Phylogenetic analysis of HpsK homologues. Phylogenetic tree constructed from HpsK homologues clustered at the UniProt90 similarity level, showing the relationships among the SSN-defined clusters (colours and cluster numbers from Figure X). Experimentally characterised DHPS- and isethionate-binding proteins are marked with red tip points. The full tree, including bootstrap support values, is provided as a Nexus file in the Supporting Information.

SUPPLEMENTARY INFORMATION

Table S1. Ligand induced change in melting temperature for *Rp*HpsK measured using NanoDSF. Ligand concentration was 2 mM.

Ligand	**ΔT_m (K)**	
	***Rp*HpsK**	***Bw*HpsK**
S-DHPS	15.7	13.7
R-DHPS	15.7	1.5
3-hydroxy-propanesulfonate (HPS)	9.9	1.5
sulfolactaldehyde	6.0	6.4
isethionate	3.4	0.5
(*S*)-2-amino-3-hydroxypropane-1-sulfonic acid (*S*-AHPS)	3.3	0.9
R-AHPS	2.3	1.0
L-cysteinolic acid	0.3	0.9
D-cysteinolic acid	0.1	2.0
L-cysteate	-0.1	1.4
taurine	0.1	0.5
homotaurine	0.2	0.4
γ-aminobutyrate acid (GABA)	0.1	0.4
R-sulfolactate	-0.2	1.5
S-sulfolactate	-0.7	2.1
sulfopropanoate	-0.3	2.8
sulfoquinovose	0.2	0.3

Table S2. Isothermal titration calorimetry binding data for *Rp*HpsK. Error estimates for individual experiments are for standard deviation in curve fit. Error estimates for mean are the standard deviation of the individual replicates.

	N	K_D (nM)	ΔH (kcal/mol)	ΔG (kcal/mol)	ΔS (kcal/mol/K)
***R*-DHPS**					
1	0.90 ± 0.02[a]	4.1 ± 1.0	-19.2 ± 0.2	-11.4	-0.026
2	0.84 ± 0.03	5.0 ± 1.4	-19.7 ± 0.2	-11.3	-0.028
3	0.86 ± 0.04	6.2 ± 2.1	-19.8 ± 0.3	-11.2	-0.029
mean	**0.87 ± 0.03[b]**	**5.1 ± 1.0**	**-19.6 ± 0.3**	**-11.30 ± 0.1**	**-0.028 ± 0.001**
***S*-DHPS**					
1	0.79 ± 0.03	6.8 ± 1.8	-20.4 ± 0.3	-11.2	-0.031
2	0.83 ± 0.06	7.8 ± 3.7	-20.1 ± 0.5	-11.1	-0.030
3	0.92 ± 0.04	13.2 ± 3.2	-20.6 ± 0.4	-10.8	-0.033
mean	**0.85 ± 0.07**	**9.2 ± 3.5**	**-20.4 ± 0.3**	**-11.0 ± 0.2**	**-0.031 ± 0.001**
HPS					
1	0.76 ± (3.7×10^{-3})	(1.05 ± 0.06) × 10^3	-14.8 ± 0.1	-8.2	-0.022
2	0.74 ± (5.2×10^{-3})	(0.92 ± 0.08) × 10^3	-14.8 ± 0.15	-8.2	-0.022
3	0.66 ± (7.4×10^{-3})	(0.81 ± 0.18) × 10^3	-14.7 ± 0.24	-8.3	-0.022
mean	**0.72 ± 0.03**	**(0.92 ± 0.07) × 10^3**	**-14.8 ± 0.03**	**-8.2 ± 0.08**	**-0.022 ± 0.000**
Isethionate					
1	0.72 ± 0.03	(58.5 ± 5.5) × 10^3	-9.11 ± 0.5	-5.8	-0.011
2	0.69 ± 0.07	(66.0 ± 12.7) × 10^3	-9.61 ± 1.3	-5.7	-0.013
3	0.53 ± 0.03	(62.6 ± 6.1) × 10^3	-11.2 ± 0.9	-5.7	-0.018
mean	**0.64 ± 0.10**	**(62.4 ± 3.8) × 10^3**	**-10.0 ± 1.1**	**-5.7 ± 0.04**	**-0.014 ± 0.004**

Table S3. Diffraction data and refinement statistics for *R. pomeroyi* HpsK.[a]

	Apo *Rp*HpsK – crystal form 1	Apo *Rp*HpsK – crystal form 2	*Rp*HpsK•*S*-DHPS	*Rp*HpsK•*R*-DHPS
Data collection				
Space group	*P*1	$P2_1$	*C*2	*C*2
Unit cell parameters (Å; °)	54.02, 58.87, 67.70; 72.29, 70.89, 67.75	54.85, 126.1, 58.83; β = 117.29	84.79, 68.76, 71.38; β = 120.32	84.79, 69.46, 71.47; β = 120.42
Resolution (Å)	46.83–1.47 (1.50–1.47)	48.74–1.90 (1.94–1.90)	42.17–1.65 (1.68–1.65)	48.07–1.96 (2.01–1.96)
No. of observations	438,279 (9,673)	194,528 (11,974)	221,494 (10,973)	89,125 (6,061)
No. of unique reflections	112,178 (5,188)	55,825 (3,452)	42,559 (2,071)	25,729 (1,747)
Completeness (%)	96.2 (90.7)	99.6 (96.6)	99.9 (100.0)	99.7 (98.9)
Redundancy	3.9 (1.9)	3.5 (3.5)	5.2 (5.3)	3.5 (3.5)
R_{merge} (%)	2.1 (15.0)	4.5 (20.5)	5.8 (38.3)	4.8 (24.0)
R_{pim} (%)	1.5 (15.0)	3.3 (14.6)	3.1 (21.3)	3.5 (17.1)
$CC_{1/2}$	1.00 (0.93)	1.00 (0.94)	1.00 (0.90)	1.00 (0.93)
Average $I/\sigma(I)$	18.5 (2.2)	11.5 (3.5)	16.1 (3.4)	12.1 (3.3)
Refinement				
R (%)	12.0 (16.2)	15.9 (19.9)	14.2 (17.3)	14.4 (18.9)
R_{free} (%)	15.5 (20.0)	19.8 (24.3)	16.4 (22.6)	17.8 (20.7)
No. (%) of reflections in test set	5,558 (5.0)	2,874 (5.2)	2,110 (5.0)	1,218 (4.7)
No. of protein molecules per asu	2	2	1	1
R.m.s.d. bond length (Å)	0.006	0.006	0.006	0.006
R.m.s.d. bond angle (°)	1.494	1.461	1.506	1.427
Average *B*-factors ($Å^2$)[b]	22.1	19.4	15.2	16.9
Protein molecules	20.1	18.4	13.4	15.6
Ligand (DHPS)	-	-	8.3	10.4
Water molecules	35.0	26.7	26.4	26.5
continued	Apo *Rp*HpsK – crystal form 1	Apo *Rp*HpsK – crystal form 2	*Rp*HpsK•*S*-DHPS	*Rp*HpsK•*R*-DHPS

Ramachandran plot[c]				
Residues other than Gly and Pro in:				
Most favoured regions (%)	97.7	97.1	98.0	98.2
Additionally allowed regions (%)	2.0	2.6	1.7	1.5
Outliers (%)[d]	0.3	0.3	0.3	0.3
PDB entry code	24GC	24GD	24GE	24GF

[a] Values in parentheses are for the highest-resolution shell.

[b] Calculated by BAVERAGE in CCP4 Suite.[3]

[c] Calculated using MolProbity.[4]

[d] Asp246 in each chain
- Chain A Asp246: -122.0 (phi), -0.2 (psi); Chain B Asp246: -127.2 (phi), -0.9 (psi) in apo-*Rp*HpsK, crystal form 1
- Chain A Asp246: -117.7 (phi), -4.1 (psi); Chain B Asp246: -122.8 (phi), -3.6(psi) in apo-*Rp*HpsK crystal form 2
- Chain A Asp246: -125.9 (phi), -2.0 (psi) in *Rp*HpsK•*S*-DHPS
- Chain A Asp246: -124.3 (phi), -4.9 (psi) in *Rp*HpsK•*R*-DHPS

Table S4. Ligand induced change in melting temperature for *Rp*HpsK site-directed mutants measured using NanoDSF. ΔT_m versus wild type indicates the change in melting temperature of the apo mutant relative to wild-type *Rp*HpsK. Ligand-induced ΔT_m values indicate the change in melting temperature of each mutant in the presence of 2 mM concentration of the indicated DHPS enantiomer relative to the corresponding apo mutant.

***Rp*HpsK mutant**	**Mutant protein T_m (°C)**	**ΔT_m versus wild type (K)**	**ΔT_m with *R*-DHPS (K)**	**ΔT_m with *S*-DHPS (K)**
R184A	56.2	-4.1	-0.6	-0.5
D272A	54.2	-6.1	-0.8	-0.5
Q274A	58.8	-1.5	+12.3	+10.9

Table S5. Isothermal titration calorimetry binding data for *Rp*HpsK Q274A site-directed mutant. Error estimates for individual experiments are for standard deviation in curve fit. Error estimates for mean are the standard deviation of the individual replicates.

	N	**K_D (μM)**	**ΔG (kcal/mol)**	**ΔH (kcal/mol)**	**ΔS (kcal/mol/K)**
***R*-DHPS**					
1	0.75 ± 0.01	0.20 ± 0.04	-9.2	-28.5 ± 0.5	0.065
2	0.69 ± 0.03	0.13 ± 0.10	-9.4	-26.7 ± 1.8	0.058
3	0.81 ± 0.01	0.12 ± 0.02	-9.5	-26.2 ± 0.3	0.056
mean	**0.75 ± 0.06**	**0.15 ± 0.04**	**-9.3 ± 0.2**	**-27.1 ± 1.2**	**0.060 ± 0.004**
***S*-DHPS**					
1	1.04 ± 0.04	1.5 ± 0.4	-7.9	-18.0 ± 0.9	0.034
2	1.02 ± 0.02	1.1 ± 0.2	-8.1	-16.3 ± 0.5	0.027
3	1.22 ± 0.04	2.1 ± 0.6	-7.8	-18.2 ± 0.9	0.035
mean	1.09 ± 0.01	**1.6 ± 0.5**	**-7.9 ± 0.2**	**-17.5 ± 1.0**	**0.032 ± 0.004**

Table S6. ITC binding data for *Bw*HpsK titrated with *S*-DHPS. Error estimates for individual experiments are for standard deviation in curve fit. Error estimates for mean are the standard deviation of the individual replicates.

	N	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (kcal/mol/K)
***S*-DHPS**					
1	0.83 ± 0.01[a]	0.55 ± 0.12	-8.6	-8.9 ± 0.2	-0.0010
2	0.94 ± 0.02	0.54 ± 0.15	-8.6	-8.5 ± 0.2	0.0001
3	0.86 ± 0.02	0.50 ± 0.18	-8.6	-8.5 ± 0.3	0.0005
	0.88 ± 0.03[b]	**0.53 ± 0.02**	**-8.6 ± 0.03**	**-8.6 ± 0.1**	**0.00015 ± 0.001**

Table S7. Diffraction data and refinement statistics for *B. wadsworthia* HpsK.[a]

	Apo *Bw*HpsK	*Bw*HpsK•*S*-DHPS
Data collection		
Space group	$P2_1$	$P1$
Unit cell parameters (Å; °)	55.02, 87.81, 65.95; β = 108.84	51.75, 52.05, 66.64; 70.31, 88.75, 80.14
Resolution (Å)	48.41–2.01 (2.06–2.01)	46.62–1.05 (1.07–1.05)
No. of observations	275,967 (19,556)	1,023,252 (49,040)
No. of unique reflections	39,014 (2,772)	285,217 (13,723)
Completeness (%)	98.2 (94.0)	94.7 (91.9)
Redundancy	7.1 (7.1)	3.6 (3.6)
R_{merge} (%)	17.6 (98.9)	2.8 (34.1)
R_{pim} (%)	7.7 (42.8)	2.1 (25.6)
$CC_{1/2}$	0.99 (0.64)	1.00 (0.86)
Average $I/\sigma(I)$	6.6 (1.5)	11.3 (1.9)
Refinement		
R (%)	17.5 (26.4)	11.8 (18.6)
R_{free} (%)	22.7 (31.1)	13.6 (19.3)
No. (%) of reflections in test set	1,989 (5.1)	14,616 (5.1)
No. of protein molecules per asu	2	2
R.m.s.d. bond length (Å)	0.006	0.012
R.m.s.d. bond angle (°)	1.451	1.907
Average *B*-factors (Å^2)[b]	23.0	13.6
Protein molecules	22.8	11.9
Ligand (*S*-DHPS)	-	6.2 (chain A); 6.6 (chain B)
Water molecules	26.0	19.3
Ramachandran plot[c]		
Residues other than Gly and Pro in:		
Most favoured regions (%)	98.1	98.6
Additionally allowed regions (%)	1.9	1.4
Outliers (%)	0	0
PDB entry code	24GG	24GH

[a] Values in parentheses are for the highest-resolution shell.

[b] Calculated by BAVERAGE in CCP4 Suite.[3]

[c] Calculated using MolProbity.[4]

SUPPLEMENTARY INFORMATION

Experimental Methods

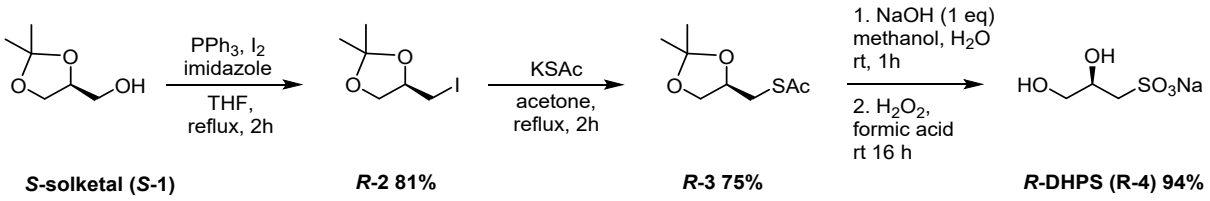
Materials

The following were synthesized as described previously: rac-DHPS,[5] AHPS and cysteinolic acids,[6] sulfolactaldehyde and sulfoquinovose.[7] Isethionate, taurine, homotaurine, hydroxypropanesulfonic acid and sodium propanesulfonate were purchased from Sigma Aldrich.

Synthesis of *R*- and *S*-DHPS

Dry solvents were obtained from a solvent purification system [8]. Reactions were monitored by thin-layer chromatography (TLC) using aluminium-backed silica gel 60 F_{254} plates. Compounds were visualized under UV light (254 nm) and/or by staining with 10% phosphomolybdic acid in ethanol followed by heating. Flash column chromatography was performed using silica gel (60 μm particle size). Optical rotations were measured as $[\alpha]_D$ ($deg \cdot mL \cdot g^{-1} \cdot dm^{-1}$). NMR spectra were recorded on a 400 MHz (^1H, 400 MHz; ^{13}C, 100 MHz) spectrometer. Chemical shifts (δ) are reported in ppm and referenced to residual solvent signals: $CDCl_3$ (^1H δ 7.26 ppm; ^{13}C δ 77.16 ppm) or D_2O (^1H δ 4.79 ppm). High-resolution mass spectra (HRMS) were acquired using an Orbitrap mass spectrometer, and data were processed using Xcalibur 4.0.27.10 (ThermoFisher Scientific).

Synthesis of sodium (*R*)-dihydroxypropanesulfonate (*R*-4):



Scheme S1. Synthesis of sodium (*R*)-dihydroxypropanesulfonate (*R*-**4**).

Synthesis of (*R*)-3-iodo-1,2-isopropylidene-glycerol (*R*-2):

A solution of I_2 (3.7 g, 14.5 mmol) in dry THF (5 mL) was added to a solution of *S*-1,2-isopropylidene-glycerol (***S*-1)** (1.5 g, 11.2 mmol), PPh_3 (3.5 g, 13.4 mmol) and imidazole (2.3 g, 33.6 mmol) in dry THF (100 mL) and the resultant mixture was heated to reflux for 2 h. After cooling to room temperature, the reaction mixture was quenched with aqueous $Na_2S_2O_3$ until decolourized, and the solvent was removed *in vacuo*. The residue was partitioned between Et_2O (50 mL) and H_2O (50 mL). The organic layer was separated, washed with water (3 × 100 mL) and brine (1 × 50 mL), dried over $MgSO_4$, filtered and concentrated The crude product was purified by flash column chromatography (CH_2Cl_2) to afford (*R*)-3-iodo-1,2-isopropylidene-glycerol (***R*-2**) (2.20 g, 81 %) as a colourless oil. $[\alpha]_D^{21}$ = +43.5 (*c* 2.46, $CHCl_3$, lit.[9]

+36.1, *c* 3.75, $CHCl_3$); ^1H NMR (400 MHz, $CDCl_3$) δ 4.30-4.23 (1 H, m), 4.15 (1 H, dd, *J* = 8.6, 6.1 Hz), 3.79 (1 H, dd, *J* = 8.6, 5.5 Hz), 3.25 (1 H, dd, *J* = 9.8, 4.7 Hz), 3.14 (1 H, dd, *J* = 9.8, 8.4 Hz), 1.46 (3 H, s), 1.35 (3 H, s) ppm; $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$) δ 110.7, 75.8, 69.8, 27.3, 25.7, 6.8 ppm. HRMS $(ESI)^+$ *m/z* $[M + H]^+$ calcd for $C_6H_{12}O_2I$ 242.9879; found 242.9882.

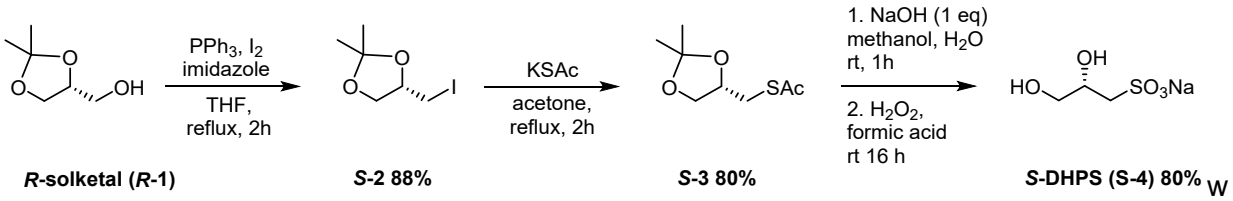
Synthesis of (*R*)-3-thioacetate-1,2-isopropylidene-glycerol (*R*-3):

A solution of ***R*-2** (2.00 g, 8.26 mmol) and potassium thioacetate (1.9 g, 16.5 mmol) in acetone (100 mL) was heated to reflux for 2 h. After cooling to room temperature, the solvent removed *in vacuo*. The resultant solid was triturated with Et_2O (3 × 50 mL) to afford (*R*)-3-thioacetate-1,2-isopropylidene-glycerol (***R*-3**) (1.3 g, 75%) as an orange oil. $[\alpha]_D^{20}$ = –8.33 (*c* 0.91, $CHCl_3$, lit.[10] –8.33 °, *c* 14, $CHCl_3$); ^1H NMR (400 MHz, $CDCl_3$) δ 4.22 (1 H, q, *J* = 6.0 Hz), 4.06 (1 H, dd, *J* = 8.5, 6.2 Hz), 3.63 (1 H, q, *J* = 8.5, 6.2 Hz), 3.15–3.03 (2 H, m), 2.35 (3 H, s), 1.43 (3 H, s), 1.34 (3 H, s) ppm; $^{13}C\{^1H\}$ NMR (100 MHz, $CDCl_3$) δ 195.3, 110.1, 74.7, 68.6, 32.3, 30.7, 27.0, 25.6 ppm. HRMS $(ESI)^+$ *m/z* $[M + H]^+$ calcd for $C_8H_{15}O_{3S}S$ 223.0463; found 223.0470.

Synthesis of sodium (*R*)-dihydroxypropanesulfonate (*R*-4):

A solution of NaOH (299 mg, 7.46 mmol) in water (16 mL) was added to a stirred solution of ***R*-3** (1.4 g, 7.46 mmol) in methanol (50 mL). The resultant solution was stirred at room temperature for 1 h, then hydrogen peroxide (32%, 20 mL) was added, resulting in formation of a colourless precipitate. The mixture was stirred for a further 1 h, during which time the precipitate dissolved. Formic acid (20 mL) was then added, and the resulting colourless solution was stirred for 16 h. The solvent removed *in vacuo* to afford a colourless residue. The residue was dissolved in water (5 mL) and precipitated by the addition of methanol (100 mL). After standing at 4 °C for 24 h, the precipitate was collected, redissolved in water (10 mL) and lyophilised to give sodium (*R*)-dihydroxypropanesulfonate (***R*-4**) (1.3 g, 94%) as a hygroscopic colourless crystalline $[\alpha]_D^{21}$ = –1.32 (*c* 1.00, H_2O); ^1H NMR (400 MHz, D_2O) δ 4.18 (1 H, ddt, *J* = 7.2, 6.2, 4.1 Hz), 3.75-3.70 (1H, m), 3.62 (1H, dd, *J* = 11.7, 6.3 Hz), 3.15 (1 H, dd, *J* = 14.5, 4.5 Hz), 3.06 (1 H, dd, *J* = 14.4, 7.6 Hz) ppm; $^{13}C\{^1H\}$ NMR (100 MHz, D_2O) δ 68.2, 64.8, 53.7 ppm. HRMS $(ESI)^+$ *m/z* $[M – Na]^-$ calcd for $C_3H_7O_5S$ 155.0014; found 155.0009.

Synthesis of sodium (*S*)-dihydroxypropanesulfonate (*S*-4):



Scheme S2. Synthesis of sodium (*S*)-dihydroxypropanesulfonate (*R*-4).

Synthesis of (*S*)-3-iodo-1,2-isopropylidene-glycerol (*S*-2):

A solution of *R*-1,2-isopropylidene-glycerol (***R*-1)** (1.5 g, 11.2 mmol), PPh_3 (3.5 g, 13.4 mmol) and imidazole (2.3 g, 33.6 mmol) in dry THF (100 mL) was heated to reflux. A solution of I_2 (3.7 g, 14.5 mmol) in dry THF (5 mL) was added dropwise and the resultant suspension refluxed for 2 h. After cooling to room temperature, the reaction was quenched with $Na_2S_2O_3$ and the solvent was removed *in vacuo*. The residue was partitioned between Et_2O (50 mL) and H_2O (50 mL), and the organic layer was washed with water (3 × 100 mL) and brine (50 mL), then dried over $MgSO_4$. Filtration and evaporation of the solvent afforded a crude residue that was purified by flash chromatography (CH_2Cl_2) to give (*S*)-3-iodo-1,2-isopropylidene-glycerol (***S*-2**) (2.4 g, 88%) as a colourless oil. $[\alpha]_D^{23}$ = –35.9 (*c* 1.21, $CHCl_3$, lit.[9] + 36.1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.30-4.23 (1 H, m), 4.13 (1 H, dd, *J* = 8.7, 6.0 Hz), 3.77 (1 H, dd, *J* = 8.6, 5.4 Hz), 3.24 (1 H, dd, *J* = 9.8, 4.7 Hz), 3.12 (1 H, dd, *J* = 9.8, 8.5 Hz), 1.43 (3H, s), 1.33 (3H, s) ppm; ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ 110.6, 75.8, 69.7, 27.3, 25.7, 6.8 ppm. HRMS (ESI)$^+$ *m/z* [M + H]$^+$ calcd for $C_6H_{12}O_2I$ 242.9882; found 242.9886.

Synthesis of (*S*)-3-thioacetate-1,2-isopropylidene-glycerol (*S*-3):

A solution of ***S*-2** (2.0 g, 8.26 mmol) and potassium thioacetate (1.9 g, 16.5 mmol) in acetone (100 mL) was heated at reflux for 2 h. After cooling to room temperature, the solvent was removed in vacuo. The resulting solid was triturated with Et_2O (3 × 50 mL) to afford (*S*)-3-thioacetate-1,2-isopropylidene-glycerol (***S*-3**) (1.4 g, 80%) as an orange oil. $[\alpha]_D^{20}$ = +7.24 (*c* 0.61, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.26–4.19 (1 H, m), 4.06 (1 H, dd, *J* = 8.4, 6.1 Hz), 3.63 (1 H, q, *J* = 8.4, 6.1 Hz), 3.15–3.03 (2 H, m), 2.35 (3H, s), 1.43 (3H, s), 1.34 (3H, s) ppm; ^{13}C{^{1}H} NMR (100 MHz, $CDCl_3$) δ 195.3, 110.1, 74.7, 68.6, 32.3, 30.7, 27.0, 25.6 ppm. HRMS (ESI)$^+$ *m/z* [M + H]$^+$ calcd for $C_8H_{15}O_{3S}S$ 223.0463; found 223.0461.

Synthesis of sodium (*S*)-dihydroxypropanesulfonate (*S*-4):

A solution of NaOH (299 mg, 7.46 mmol) in water (16 mL) was added to a stirred solution of ***S*-3** (1.4 g, 7.46 mmol) in methanol (50 mL). The mixture was stirred for 1 h, after hich hydrogen peroxide (32%, 20

mL) was added and a colourless precipitate was observed. The reaction was further stirred for an hour, during which the precipitate dissolved. Formic acid (20 mL) was then added and the resultant colourless solution stirred for 16 h. The solvent removed *in vacuo* to afford a colourless residue. The residue was dissolved in water (5 mL) and precipitated by the addition of methanol (100 mL). After standing at 4 °C for 24 h, the precipitate was collected, redissolved in water (10 mL) and lyophilised to afford the sodium salt of (*S*)-dihydroxypropanesulfonate (***R*-4**) (1.1 g, 80%) as a hygroscopic colourless crystalline solid $[\alpha]_D^{24}$ = +0.91 (*c* 1.00, H_2O); ^{1}H NMR (400 MHz, D_2O) δ 4.26–4.16 (1 H, m), 3.76–3.71 (1 H, m), 3.63 (1 H, ddd, *J* = 11.7, 6.4, 1.1 Hz), 3.16 (1 H, ddd, *J* = 14.5, 4.4, 1.1 Hz), 3.07 (1 H, ddd, *J* = 14.6, 7.5, 1.0 Hz) ppm; ^{13}C{^{1}H} NMR (100 MHz, D_2O) δ 68.2, 64.8, 53.7 ppm. HRMS (ESI)$^+$ *m/z* [M – Na]$^-$ calcd for $C_3H_7O_5S$ 155.0014; found 155.0015.

Growth of *R. pomeroyi* DSS-3 and proteomic analysis

R. pomeroyi DMZ 15171 (DSS-3) was grown in marine basal media (MBM), which was prepared containing (per L): 20 g sea salt, 250 mL basal media, 50 mL Na[FeEDTA] stock solution (500 mg/L, pre-autoclaved) and 700 mL deionized H_2O. The final MBM was supplemented with 0.1% vitamin mixture.
Basal media stock before dilution contains: 150 mL Tris.HCl (1.0 M, pH 7.5), 375 mL deionized H_2O, 87 mg KH_2PO_4 (final concentration 0.26 mM) and 1.5 g NH_4Cl (final concentration 75 mM).
Vitamin mixture stock before dilution contains: biotin (82 μM), folic acid (45 μM), pyridoxine hydrochloride (490 μM), riboflavin (130 μM), thiamine (190 μM), nicotinic acid (4.1 mM), pantothenic acid (230 μM), cyanocobalamin (0.74 μM) and p-aminobenzoic acid (360 μM).

Five independent cultures of *R. pomeroyi* were grown on glucose, *R*-DHPS, or *S*-DHPS in MBM supplemented with 5 mM sterile-filtered substrate, at 25 °C with shaking at 200 rpm. Cultures were inoculated at 0.1% (v/v) using stationary-phase *R. pomeroyi* DMZ 15171 (DSS-3) pre-grown under identical conditions with the corresponding substrate. Cells were harvested at mid-logarithmic phase (A_{600} ≈ 0.3x stationary phase A_{600}) by centrifugation at 4000 *g* for 10 min. Cell pellets were washed twice by resuspension in ice-cold 1 × phosphate-buffered saline (PBS) and frozen.

Frozen whole bacterial pellets were solubilized in 4% (w/v) SDS, 100 mM HEPES (pH 8.0) by heating at 95 °C for 10 min. Protein concentrations were determined using a bicinchoninic acid (BCA) assay (Thermo Fisher Scientific). For each independent replicate, 100 μg total protein was processed for proteomic analysis using S-Trap Mini columns (Protifi, USA) according to the manufacturer's protocol. Samples were reduced with 10 mM dithiothreitol (DTT) at 95 °C for 10 min and alkylated with 40 mM iodoacetamide (IAA) in the dark for 1 h at room temperature. Samples were acidified to a final concentration of 1.2% phosphoric acid and diluted with seven volumes of S-Trap binding buffer (90%

methanol, 100 mM tetraethylammonium bromide [TEAB], pH 7.1). The mixtures were loaded onto S-Trap columns and washed three times with 400 µL S-Trap wash buffer. Proteins were digested on-column with 2 µg Trypsin/Lys-C (Promega; 1:50 protease:protein ratio) in 100 mM TEAB overnight at 37 °C. Peptides were sequentially eluted by centrifugation with: 100 mM TEAB; 0.2% formic acid; and 0.2% formic acid in 50% acetonitrile. Eluates were dried under vacuum and further purified using C18 StageTips [11, 12] to remove salts and particulate contaminants.

Purified peptides were resuspended in Buffer A* (2% acetonitrile, 0.01% trifluoroacetic acid) and separated using a two-column nanoLC setup comprised of a PepMap100 C18 trap column (20 mm × 75 µm) and PepMap C18 analytical column (500 mm × 75 µm). Chromatography was performed on a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific). Samples were loaded onto the trap column at 5 µL/min for 6 min using Buffer A (0.1% formic acid, 2% DMSO), followed by analytical separation at 300 nL/min using an 89 min gradient on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) with the following profile: 2% to 22% Buffer B (0.1% formic acid, 77.9% acetonitrile, 2% DMSO) over 59 min; 22% to 40% B over 10 min; 40% to 80% B over 5 min; hold at 80% B for 5 min; decrease to 2% B over 2 min; re-equilibrate at 2% B for 8 min. The mass spectrometer was operated in data-dependent acquisition mode with a 3 s cycle time. A single Orbitrap MS scan (*m/z* 300–1600, resolution 60,000) was followed by Orbitrap HCD MS/MS scans (resolution 15,000, fixed NCE 35%, maximum injection time 54 ms).

Raw data were processed using MaxQuant (v2.2.0.0)(9) against the *Ruegeria pomeroyi* DSS-3 proteome (UniProt: UP000001023). Search parameters included: fixed modification: carbamidomethylation of cysteine; variable modification: oxidation of methionine; enzyme specificity: trypsin; LFQ enabled; “Match between runs” enabled. The resulting data were analyzed using Perseus (v1.4.0.6).[13] Missing values were imputed from a normal distribution based on total peptide intensities (width 0.3σ, downshift 2.0σ). Statistical comparisons between groups were performed using Student’s t-tests. Data visualization was carried out in R using the ggplot package.

Growth of *B. wadsworthia* 3.1.6 and proteomic analysis

Bilophila wadsworthia 3.1.6 was grown in a carbonate-buffered mineral salt medium[14] containing (per L): 1.0 g NaCl, 0.4 g $MgCl_2.6\ H_2O$, 0.2 g KH_2PO_4, 0.25 g NH_4Cl, 0.5 g KCl, 0.15 g $CaCl_2.2\ H_2O$ in double distilled water (ddH_2O). As redox indicator, 1 ml of a 0.4 mg/ml Na-resazurine solution was added. The medium was autoclaved at 121 °C for 30 min. Then, 50 mM bicarbonate buffer (final concentration) was added, and the medium was reduced with 1 mM titan-(III)-nitrilotriacetate (final concentration).[15] As additional sulfur source, 200 µM Na_2S was added (final concentration). 1 ml trace element solution SL10 (DSMZ medium 141), 1 ml selenite-tungstate solution (DSMZ medium 385), and 0.5 ml 7-vitamin

solution[14] were additionally added. The medium was supplemented with 200 µg L^{-1} of 1,4-naphthoquinone [16]. Finally, the pH was adjusted to 7 with 1 M HCl.

For comparative proteomics *B. wadsworthia* 3.1.6 was grown in 5-ml cultures in Hungate tubes. The headspace contained N_2/CO_2 gas (79/21). Cultures (n = 3) grew with 10 mM *rac*-DHPS or 10 mM Na_2SO_3 as electron acceptor and 20 mM L-lactate as the respective electron donor (final concentrations). After two days the cultures were harvested by centrifugation (13,000 x *g*, 10 min, 4 °C) and the resulting pellet was washed with 1 mL 10 mM Tris, pH 8.0 before resuspending the pellet in the same buffer. Resuspended pellets were stored at -20 °C until further processed. 200 µl of sterilized zirconia/silica beads (0.1 mm diameter, ROTH) were added to the suspended cells in screw cap tubes and lysed by bead beating with a FastPrep F120 (Savant Bio 101). Cells underwent four rounds of 45 sec of beating at force 6.5 followed by 2 min on ice, to cool the sample. Sufficient disruption of the cells was checked by microscopy. Intact cells and cell debris were removed by centrifugation (10,000 x *g,* 10 min, 4 °C). The supernatant was transferred into new sterile tubes and protein concentrations were measured with a Qubit 4 Fluorometer (Invitrogen) using the Qubit Protein BR assay kit (Invitrogen). Protein concentrations were adjusted by dilution with 10 mM Tris, pH8.0 to achieve the same protein concentration in each sample (210 µg mL^{-1}).

Proteomic analyses were performed by the Proteomics Core Facility at the University of Konstanz (www.biologie.uni-konstanz.de/proteomics-centre/) to assess the total proteome. For sample preparation, proteins were reduced with 10 mM dithiothreitol (DTT) for 30 min at 56 °C and alkylated with 50 mM chloroacetamide for 60 min at room temperature. Proteolytic digestion was carried out using trypsin at 0.01 µg $µl^{-1}$ trypsin for 18 h at 33 °C. Peptide digests were analysed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to an Easy-nLC 1200 nanoflow liquid chromatography system (Thermo Fisher Scientific, Bremen, Germany). Samples were reconstituted in 0.1% formic acid and loaded directly onto a heated analytical column (75 µm × 50 cm, 50 °C). Peptides were separated at a flow rate of 150 nL min^{-1} using a linear gradient of 6−48% solvent B (0.1% formic acid in 80% acetonitrile) over 108 min. Data-independent acquisition (DIA) was performed with full MS scans acquired over an *m/z* range of 300−1650 at a resolution of 120,000. The mass range was segmented into 24 variable isolation windows for fragmentation (window widths between 27 and 470 *m/z*), and MS/MS spectra were acquired at a resolution of 30,000. Fragmentation was achieved by higher-energy collision dissociation (HCD) with a normalized collision energy of 28%. Raw data were searched against the *Bilophila wadsworthia* 3.1.6 proteome (UniProt: UP000006034) using Spectronaut (Biognosys). Protein abundances from the two growth conditions (*rac*-DHPS + L-lactate and Na_2SO_3 + L-lactate) were statistically compared using the DEP (Differential Enrichment analysis of Proteomics data) package (v1.32.0) in R (v4.4.2) within RStudio. Data were variance-stabilizing normalized (vsn), and missing values

were imputed using the MinProb method (q = 0.01). Differentially abundant proteins were defined using a threshold of $-\log_{10}$(p-value) ≥ 2 (corresponding to p ≤ 0.01) and an absolute $\log_2$ fold change ≥ 1.

Cloning, protein expression and purification of HpsK from *Ruegeria pomeroyi* DSS-3 and *Bilophila wadsworthia* 3.1.6

The codon-optimized DNA sequences encoding *Rp*HpsK (WP_011046347.1), *Rp*HpsK mutants (R184A, D272A and Q274A) and *Bw*HpsK (WP_005027949.1), excluding signal peptides and incorporating an N-terminal hexa-histidine (His_6) tag followed by a TEV protease recognition site, were synthesized and cloned into pET28a between the NcoI and XhoI restriction sites. The plasmids were transformed into *E. coli* BL21(DE3). Cultures were grown in LB medium with 50 μg/ml kanamycin at 37 °C and induced with 0.3 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at A_{600} ≈ 0.8, followed by incubation at 16 °C for 16 h. Cells were harvested by centrifugation at 6000 × *g*. The pellet was resuspended in buffer A (20 mM Tris · HCl, pH 8.5, 500 mM NaCl, 25 mM imidazole, 10% glycerol), lysed using a TS Series benchtop cell disruptor, and clarified by centrifugation at 35000 × *g*. The soluble fraction was applied to a Ni^{2+}-charged chelating column (Hi-Trap HP, Cytiva) and eluted using a 40–500 mM imidazole gradient in buffer A. Pooled fractions from the affinity chromatography step were concentrated and applied to size-exclusion chromatography. For crystallization trials, selected protein batches were treated with tobacco etch virus (TEV) protease at a 1:20 (w:w) ratio in dialysis tubing (10 kDa MWCO) and dialyzed against buffer B (20 mM Tris · HCl, pH 8.5, 200 mM NaCl) for 2 h at room temperature. Uncleaved protein and His_6-tagged TEV protease were removed by reapplication to a Ni^{2+}-charged Hi-Trap column, and unbound fractions were collected. Further purification was performed using size-exclusion chromatography with a HiLoad 16/600 Superdex 75 pg column or Superdex 200 Increase 10/300 GL column equilibrated with buffer B. Final protein samples were concentrated to 10–33 mg/ml for *Rp*HpsK and 44–69 mg/ml for *Bw*HpsK, snap-frozen using liquid nitrogen and stored at − 80 °C.

Nano-differential scanning fluorescence (nanoDSF) analysis

Thermal stability analysis for HpsK in the presence and absence of each organosulfonate at 2 mM was performed on a Prometheus Panta (NanoTemper) at 30% excitation, scanning from 25 °C to 95 °C at 1°C per min. Protein samples at 4 mg/ml were mixed 1:1 with 4 mM ligand solution, which prepared in 20 mM Tris-HCl, pH 8.5, 0.2 M NaCl, to give a final protein concentration of 2 mg/mL and final ligand concentration of 2 mM. Data acquisition and analysis was performed with PR.Panta Analysis (NanoTemper) software.

Isothermal titration calorimetry (ITC)

ITC experiments were performed using a MicroCal PEAQ-ITC (Malvern Panalytical). HpsK samples were diluted in the range of 20–150 μM with the size-exclusion chromatography buffer (20 mM Tris · HCl, pH 8.5, 200 mM NaCl) and loaded into the sample cell. Ligand solutions were prepared with the same buffer at 10-fold higher concentrations than that of the protein to achieve a final ITC molar ratio of 2.

Experiments for *Bw*HpsK titrated with *S*-DHPS, *Rp*HpsK Q274A titrated with *R*- and *S*-DHPS, and wildtype *Rp*HpsK titrated with HPS or isethionate were conducted as standard ITC experiments: 40 μM *Bw*HpsK was titrated with 400 μM *S*-DHPS, 20 μM *Rp*HpsK Q274A with 200 μM *R*- or *S*-DHPS, whilst 60 μM wildtype *Rp*HpsK was titrated with 600 μM HPS and 150 μM wildtype *Rp*HpsK with 1.5 mM isethionate. Due to extremely low binding affinity (K_D) of wildtype *Rp*HpsK with *R*- and *S*-DHPS, competitive displacement ITC experiments were conducted by pre-mixing 200 μM HPS with 20 μM *Rp*HpsK, in order to increase the apparent K_D to a level more appropriate for ITC.[17] Aliquots of 200 μM *R*- or *S*-DHPS was then titrated in following standard procedure.

The system was allowed to equilibrate with stirring at 750 rpm at 25 °C, with reference power of 10 μ cal/s. For the experiment, there was an initial delay of 60s followed by an initial 0.4 μl injection, then by 18 injections of 2 μl at 150s intervals. Each injection lasted 4s. Control experiments were conducted by injecting buffer into protein solution. All data analysis was performed in MicroCal PEAQ-ITC Analysis software (Malvern Panalytical). Control data were subtracted from the experimental data, with baseline-corrected binding isotherms then fitted to a one-site binding model to derive the binding affinity (K_D), stoichiometry (n), enthalpy change (ΔH), and entropy change (ΔS).

Protein crystallization and X-ray diffraction data collection

Apo-crystals of *Rp*HpsK were grown by hanging-drop vapor diffusion at 20 °C. A 2 μl aliquot of protein (33 mg/ml) was mixed with an equal volume of reservoir solution (24–28% (w/v) PEG3350, 0.2 M sodium formate for crystal form 1 in space group $P1$; 20–25% (w/v) PEG3350, 0.2 M ammonium sulfate, 0.1 M Bis-Tris, pH 5.5 for crystal form 2 in space group $P2_1$ and equilibrated against 0.5 ml reservoir solution. Crystals of *Rp*HpsK in complex with *R*- or *S*-DHPS were grown by sitting-drop vapor diffusion at 20 °C. The protein sample was preincubated on ice for 30 min with 10 mM *R*- or *S*-DHPS (final concentration of protein at 30 mg/mL). Crystallization drops consisted of 0.75 μl protein-ligand complex and an equal volume of reservoir solution (20–25% (w/v) PEG3350, 0.1 M Bis-Tris, pH 6.5) were equilibrated against 0.25 ml reservoir solution.

Both apo- and *S*-DHPS-bound crystals of *Bw*HpsK were obtained during initial screening using the Index screen (Hampton Research). Apo-*Bw*HpsK was used at 44 mg/ml, while the preincubated *Bw*HpsK•*S*-DHPS complex was prepared at a final concentration of protein at 62 mg/ml with 10 mM *S*-

DHPS. Sitting-drop plates (96-well) were set up at 20 °C with crystallization drops composed of 0.15 µl protein solution and 0.15 µl reservoir solution, equilibrated against 50 µl reservoir solution. Conditions G1(25% (w/v) PEG3350, 0.2 M NaCl, 0.1 M Tris · HCl, pH 8.5) and F10 (25% (w/v) PEG3350, 0.2 M NaCl, 0.1 M Bis-Tris, pH 5.5) produced crystals for apo-*Bw*HpsK and *Bw*HpsK•*S*-DHPS, respectively.

Apo-crystals of *Rp*HpsK were cryoprotected by stepwise transfer to artificial reservoir solutions containing 20% ethylene glycol in 10% increments, before cryocooled with liquid nitrogen. Ligand-bound *Rp*HpsK crystals and all *Bw*HpsK crystals were harvested without additional cryoprotection. Diffraction data were collected at 100 K on beamline MX2[18] at the Australian Synchrotron using a wavelength of 0.954 Å for all except *Bw*HpsK•*S*-DHPS data set (0.827 Å used). Data were processed with XDS,[19] and merged and scaled using AIMLESS.[20] Data collection and merging statistics are summarized in **Tables S3** and **S7**.

Crystal structure determination and refinement

The crystal structures of *Rp*HpsK were solved by molecular replacement using PHASER[21] within the CCP4 suite.[3] A modified AlphaFold2-predicted model of *Rp*HpsK (AF-Q5LVV4-F1) was used as the initial search model. For the *S*-DHPS-bound structure, a solution with a log-likelihood gain (LLG) of 1,956 and a Z-score of 37.5, was obtained, locating one molecule in the asymmetric unit. The *R*-DHPS-bound structure was refined using the protein chain from the *S*-DHPS-bound structure after removing all non-protein atoms and setting the temperature factor to 30 Å^2. Iterative manual model building and refinement using COOT[22] and REFMAC[23], respectively, was carried out.

For the apo-structure from crystal form 1, the same initial search model yielded a solution with a LLG of 170 and a Z-score of 13.6, locating two molecules in the asymmetric unit. Due to the 'open' conformation, extensive model rebuilding was required for the apo-structure. Auto-building with density modification was performed using Phenix[24] followed by iterative manual model building and refinement using COOT and REFMAC. The apo-structure from crystal form 2 was solved using chain A of the apo-structure from crystal form 1 as the search model (LLG of 9,749 and a Z-score of 70.2, two molecules per asymmetric unit).

The crystal structures of *Bw*HpsK were also solved and refined similarly. A modified AlphaFold2 predicted model of *Bw*HpsK (AF-E5Y7I1-F1) located two molecules per asymmetric unit in both structures (apo-*Bw*HpsK: LLG of 935 and a Z-score of 28.8; *Bw*HpsK•*S*-DHPS complex: LLG of 520 and a Z score of 20.4). Auto-building with density modification using Phenix followed by iterative manual model building and refinement using COOT and REFMAC was carried out. Model quality was validated using MOLPROBITY.[4] Refinement statistics are included in **Tables S3** and **S7**.

Sequence similarity network analysis of HpsK

Following an approach similar to that of Vetting and co-workers,[25] we analysed a sequence similarity network (SSN) and genome neighbourhood network (GNN) comprising 92,628 TRAP-associated solute-binding protein (SBP) sequences from the InterPro family IPR018389 using the Enzyme Function Initiative (EFI) web server[26], with UniRef50 applied to reduce network complexity to 8672 nodes. SSN and GNN (±20 genes) were visualised using Cytoscape at an alignment score of 90, with this level of stringency chosen to match that used by Vetting and co-workers, enabling direct comparison with experimentally characterised SBPs that cluster into isofunctional groups (SI Table Excel). Clusters with a high proportion of HpsNOP (Pfams PF00815, PF13561, PF00107-PF08240) or HpsGH/HpfGH (Pfams PF01228-PF02901, PF04055) in their genome neighbourhoods were taken as proposed HpsK clusters and separated into a proposed HpsK SSN and GNN of 1,036 TRAP-associated SBP sequences. To improve resolution within these functionally relevant clusters, UniRef50 IDs were expended to their full UniProt sequences and used to construct a refined SSN at an alignment score of 80 (approximately 40% sequence identity), with this stringency chosen to reproduce the cluster distribution observed in the GNN analysis.

HpsK and IseP phylogenetic tree

HpsK and IseP sequences from the refined SSN were download at UniRef90 identity and underwent multi-sequence alignment using the ClustalW method. A phylogenetic tree was constructed and optimised by the maximum likelihood method, performed using the phangorn package in RStudio. Branch support was assessed using 100 bootstrap replicates.

SUPPLEMENTARY INFORMATION

> Signal peptide-removed *Rp*HpsK wild-type_codon optimized DNA sequence (DNA sequence coding for the N-terminal His6-tag and TEV protease cleavage site in capital letters)

ATGGGCAGCAGCCATCACCATCATCACCACGAAAACCTGTATTTTCAGGGCcaaactgcaaaagaggaga
aagaacgcgaggcggcagcttcacacgttatgaccgttgcaactgcttacgttcttggtgctagccgaag
ttacccaataatgcaattggatctgaaagaaaatattcaaaatgcgactaacgggaaagtttacgttaag
ttagccccaggtggacaacttggtgcgggcggcgccctggtccagaaagtacaaggtggtactatacaag
ccgcccaacatagtttatccaactttgcgccgtttgcatcagcggtggatttaattaatatgccatactt
ttgtggaagcaatcaacgttttacgaatttagtaagcagcgacgcttggaaaactgaagtccacccgaaa
gttgaagcagcagggtttaaagccttgttttacattgtgatagatccccgggttgtggcggttcggaaag
gcggaaatgccgttattactcccggggacttagcaggcgtgaaatttcgtgttccaggttctaaaatgct
ccagcaatactatcgtatggttggggcgaatcccacaccagtagcatggggcgaaactccgtcagccatt
aaacaaggtgtttgcagacgccctggatccctctgtaggggccctgtacgttttcgggtttaaagacattt
taagtcacgttacattcacgcaagctgttccggattcacaagtatttagtatgaatttagagtggtttaa
tggtctgccagcggacgtccaagaggggattatgtttgcaggggaagtgacgtctcaacaaaatcttgca
aaagtcccggcagctagagcatacgccatgagtgagttaacaaagagtggtgttgaatttcactccctgt
ctgctgatcaattggctgagtggcaagccacgggtggttaccaaagaagtgagtgggattcctttaaaac
ggagttagcgggcagtatggatgcgtttaatagattggaagaagcagcgggtacaatgggtcgttactat
gttcatgatgcgtaa

* Three side-directed mutants (R184A, D272A, and Q274A) have the same nucleotide sequence except the mutation site marked with single-underline, double-underline, and wave-underline, respectively. These sites are substituted as following: R184A (cgg → gcg), D272A (gat → gct), and Q274A (caa → gca), respectively.

> Signal peptide-removed *Rp*HpsK wild-type_amino acid sequence (N-terminal His6-tag and TEV protease cleavage site underlined)

MGSSHHHHHHENLYFQGQTAKEEKEREAAASHVMTVATAYVLGASRSYPIMQLDLKENIQNATNGKVYVK
LAPGGQLGAGGALVQKVQGGTIQAAQHSLSNFAPFASAVDLINMPYFCGSNQRFTNLVSSDAWKTEVHPK
VEAAGFKALFYIVIDPRVVAVRKGGNAVITPGDLAGVKFRVPGSKMLQQYYRMVGANPTPVAWGETPSAI
KQGVADALDPSVGALYVFGFKDILSHVTFTQAVPDSQVFSMNLEWFNGLPADVQEGIMFAGEVTSQQNLA
KVPAARAYAMSELTKSGVEFHSLSADQLAEWQATGGYQRSEWDSFKTELAGSMDAFNRLEEAAGTMGRYY
VHDA*

SUPPLEMENTARY INFORMATION

> Signal peptide-removed *Bw*HpsK_codon optimized DNA sequence (DNA sequence coding for the N-terminal His6-tag and TEV protease cleavage site in capital letters)

ATGGGCAGCAGCCATCACCATCATCACCACGAAAACCTGTATTTTCAGGGCgctgagtataagaaaatga
cgatacgggcagcaacggctaatccccaaggttcattacacgtcgtggcaattgataaatttaaagaaat
tgtggagaaagagagtaatggagccattactgttcaaacgttttacggtgggtcattaggggatgagcaa
gcaaatgtaaagcaattacgaaatgctgaaattcacctggctgtcttagctgatggtaatttaaccccat
ttgccccacaagctggagtctttattttaccctatatgtttcccaaaataagtgatgcagagaaattgtt
tggtaacgaggcctttatgaataaaacggcagatgcgatcgcaaaacaatcccgcactcgacctttgtca
tggctcgttggtgggtatcggataattacgaatagcaagaaacctattaatacaatggctgatctgaaag
gtttgaaaatacgagtcccagcagtcgaactgcaactggcggcgtttcgatcttggggtgtagagccgca
ccctttggcctggtctgagacttttaatggtttacaacaaggagttgtggatggacaagagaatccgcat
gcaattaatcgggatcagaaattttgggaagtgcaaaagtatattactaatatacactacatgttatggg
tgggacctatgttggtatccgatccgtggtttcgaaaactggatcctcagacgaaagctttggtcgagaa
agcagctaaagaggcggctgcttatgagtggaaatggagtgccgagcaagacgagattgctcttaaagag
tgtttagcccgtggtatggtaattaacgatgtttctgatgagccggcatggactgaagcagcaagaagcg
tatggccacaattttatgataaagtaggtggaaaagctgtagtcgacgaggcgctggcgattatgcaaca
atga

> Signal peptide-removed *Bw*HpsK _amino acid sequence (N-terminal His6-tag and TEV protease cleavage site underlined)

MGSSHHHHHHENLYFQGAEYKKMTIRAATANPQGSLHVVAIDKFKEIVEKESNGAITVQTFYGGSLGDEQ
ANVKQLRNAEIHLAVLADGNLTPFAPQAGVFILPYMFPKISDAEKLFGNEAFMNKTADAIAKQSRTRPLS
WLVGGYRIITNSKKPINTMADLKGLKIRVPAVELQLAAFRSWGVEPHPLAWSETFNGLQQGVVDGQENPH
AINRDQKFWEVQKYITNIHYMLWVGPMLVSDPWFRKLDPQTKALVEKAAKEAAAYEWKWSAEQDEIALKE
CLARGMVINDVSDEPAWTEAARSVWPQFYDKVGGKAVVDEALAIMQQ*

Supplementary References

1. S. Hayward and H. J. C. Berendsen, *Proteins*, 1998, **30**, 144-154.
2. C. Girdlestone and S. Hayward, *J. Comput. Biol.*, 2016, **23**, 21-26.
3. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D*, 2011, **67**, 235-242.
4. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall Iii, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson and D. C. Richardson, *Protein Sci.*, 2018, **27**, 293-315.
5. J. Mayer, T. Huhn, M. Habeck, K. Denger, K. Hollemeyer and A. M. Cook, *Microbiology*, 2010, **156**, 1556-1564.
6. L. Burchill, L. Zudich, P. L. van der Peet, J. M. White and S. J. Williams, *J. Org. Chem.*, 2022, **87**, 4333-4342.
7. 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.
8. A. B. Pangborn, M. A. Giardello, R. H. Grubbs, R. K. Rosen and F. J. Timmers, *Organometallics*, 1996, **15**, 1518-1520.
9. R. d. S. Martins, M. P. Pereira, P. P. de Castro and F. I. Bombonato, *Tetrahedron*, 2020, **76**, 130855.
10. A. K. M. Anisuzzaman and L. N. Owen, *J. Chem. Soc., Perkin Trans. 1*, 1967, DOI: 10.1039/J39670001021, 1021-1026.
11. J. Rappsilber, Y. Ishihama and M. Mann, *Anal. Chem.*, 2003, **75**, 663-670.
12. J. Rappsilber, M. Mann and Y. Ishihama, *Nat. Protoc.*, 2007, **2**, 1896-1906.
13. S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox, *Nat. Methods*, 2016, **13**, 731.
14. F. Widdel and N. Pfennig, *Arch. Microbiol.*, 1981, **129**, 395-400.
15. T. T. Moench and J. G. Zeikus, *J. Microbiol. Methods*, 1983, **1**, 199-202.
16. H. Laue, K. Denger and A. M. Cook, *Appl. Environ. Microbiol.*, 1997, **63**, 2016-2021.
17. A. Velazquez-Campoy and E. Freire, *Nat. Protoc.*, 2006, **1**, 186-191.
18. D. Aragao, J. Aishima, H. Cherukuvada, R. Clarken, M. Clift, N. P. Cowieson, D. J. Ericsson, C. L. Gee, S. Macedo, N. Mudie, S. Panjikar, J. R. Price, A. Riboldi-Tunnicliffe, R. Rostan, R. Williamson and T. T. Caradoc-Davies, *J. Synchrotron Radiation*, 2018, **25**, 885-891.
19. W. Kabsch, *Acta Crystallogr., Section D: Biol. Crystallogr.*, 2010, **66**, 125-132.
20. P. R. Evans and G. N. Murshudov, *Acta Crystallogr. Sect. D*, 2013, **69**, 1204-1214.
21. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658-674.
22. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. Sect. D*, 2010, **66**, 486-501.
23. M. D. Winn, G. N. Murshudov and M. Z. Papiz, *Methods Enzymol.*, 2003, **374**, 300-321.
24. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams and P. D. Adams, *Acta Crystallogr. D Struct. Biol.*, 2019, **75**, 861-877.
25. M. W. Vetting, N. Al-Obaidi, S. Zhao, B. San Francisco, J. Kim, D. J. Wichelecki, J. T. Bouvier, J. O. Solbiati, H. Vu and X. Zhang, *Biochemistry*, 2015, **54**, 909-931.

26. J. A. Gerlt, J. T. Bouvier, D. B. Davidson, H. J. Imker, B. Sadkhin, D. R. Slater and K. L. Whalen, *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics*, 2015, **1854**, 1019-1037.