Supporting Information for Newton *et al.*, "Structural and functional innovations in the real-time evolution of new $(\beta\alpha)_8$ barrel enzymes"

The Supporting Information comprises SI Materials and Methods, Figures S1-S4, Tables S1-S2 and SI References.

SI Materials and Methods

Molecular biology. The primers for introducing the D7N mutation into the genes encoding HisA(dup13-15/D10G) and HisA(dup13-15/D10G/G102A/Q24L/V106L) were 5′-CCG GCA TTA AAC TTA ATT GGC GGC ACC GTG GTG CG-3′ and 5′-CGC ACC ACG GTG CCG CCA ATT AAG TTT AAT GCC GG-3′ (mutated codon underlined).

Enzyme kinetics. For HisA and TrpF activity assays, two independently prepared batches of each enzyme were assayed with a range of substrate concentrations; technical triplicates were carried out at each concentration. All assays were at 25°C and the HisA assays were conducted as described previously (1).

The TrpF activity of each enzyme was quantified using a coupled spectrophotometric assay (2). Addition of the downstream enzyme TrpC (indoleglycerol phosphate synthase) allowed TrpF activity to be quantified *via* the formation of the TrpC product indoleglycerol phosphate (InGP), which absorbs strongly at 278 nm (ε = 5,590 $M^{-1}.cm^{-1}$). The TrpC enzyme was from *Pseudomonas aeruginosa* and its purification has been described previously (3). Reaction mixtures contained Tris·HCl pH 7.5 (50 mM), $MgCl_2$ (4 mM), β-mercaptoethanol (5 mM) and TrpC at 10 μM. TrpD (anthranilate phosphoribosyltransferase; 5 μM) from *Acinetobacter baylyi* (2) was used to synthesize the labile substrate, PRA, in the cuvette at concentrations from 0 to 2.0 mM, from the corresponding amount of anthranilate (Sigma Chemical Co.) and a 5-fold molar excess of phosphoribosyl pyrophosphate (Sigma Chemical Co.). Where it was not possible to saturate the enzyme with substrate (*i.e.* $K_M^{PRA} > 2$ mM), k_{cat}/K_M was estimated from the slope of the regression line when the initial velocity was plotted against the substrate concentration. Controls confirmed that the enzyme with TrpF activity was catalyzing the rate determining step in InGP formation.

Relative quantification of enzyme expression levels. The relative expression level of each HisA variant was determined using multiplexed tandem mass tagging. The 11 *S. enterica* strains used in this analysis were constructed previously (4): DA25516; DA25518; DA25520; DA25522; DA25524; DA25530; DA25540; DA25544; DA25568; DA25570; and DA25578. Overnight cultures grown in LB supplemented with 0.2% glucose were diluted 125-fold into the same medium. The cultures were grown to OD_{600} = 0.4 before being chilled in an ice-water bath. The cells were pelleted, washed three times in PBS, and frozen at -80°C. Sample preparation and relative quantification of peptides by TMT10plex (Thermo Fisher Scientific) was performed by the Proteomics Core Facility at the Sahlgrenska Academy, University of Gothenburg, as described previously (5). Only peptides unique for a given protein were used for relative quantification and ratios were normalized using the protein median. Only peptides that were common between all HisA variants and were detected in all samples in the same 10-plex were used in the analysis.

Crystallization screening, data collection and refinement. The HisA mutants were subjected to commercial crystallization screens (JCSG-*plus* HT-96, Structure Screen 1 + 2 HT-96, and Morpheus HT-96), using the vapor diffusion method at 8°C and at room temperature. The protein concentrations used for screening were 10-30 $mg.mL^{-1}$. Crystallization conditions for each mutant are listed in Table S2. All crystals were cryoprotected in reservoir solution supplemented with 15% glycerol before being vitrified in liquid nitrogen, unless mentioned otherwise in Table S2.

Data were collected at ESRF (Grenoble, France), Diamond Light Source (Didcot, UK) and PETRA (Hamburg, Germany), and processed using XDS (6) and Aimless (7). The structures were solved by molecular replacement using Phaser (8), as described for wild type HisA (1), then rebuilt using Coot (9) and refined with phenix.refine (10). Data and refinement statistics are listed in Table S2.

NMR experiments. ^{15}N Carr-Purcell-Meiboom-Gill (CPMG) (11, 12) relaxation dispersion experiments were recorded for ^{15}N labeled HisA(dup13-15) and HisA(dup13-15/D10G) at a static magnetic field strength of 14.1 T and a sample temperature of 25°C. The sample conditions for both variants were 0.7 mM protein in 50 mM Tris·HCl pH 7.5, 150 mM Na_2SO_4, 5 mM β-mercaptoethanol, 2 mM NaN_3, 10% D_2O. Data sets were obtained with a constant time relaxation delay of T = 20 ms and effective fields, ν_{CPMG}, between 50 and

1000 Hz. In both cases a total of 18 two-dimensional planes, including four duplicate data points and a reference experiment with the relaxation delay omitted, were recorded. The data was processed with NMRPipe (13) and peaks integrated using PINT (14). Effective transverse relaxation rates were calculated as $R_{2,eff}(\nu_{CPMG}) = \ln(I_0/I(\nu_{CPMG}))\,/T$ where $I(\nu_{CPMG})$ is the peak intensity for different effective fields and I_0 is the peak intensity in the reference experiment. Here, $\nu_{CPMG} = 1/(2\tau_{CP})$, where τ_{CP} is the interval between successive 180° refocusing pulses during the CPMG pulse train. Uncertainties in $R_{2,eff}(\nu_{CPMG})$ values were estimated from the duplicate measurements. All dispersion data was fitted on a per-residue basis to the Bloch-McConnell equations (15) including or excluding two-site exchange. Conformational dynamics were established by F-tests at a significance level of $p < 0.01$, and dispersion sizes were calculated as the difference between the effective transverse relaxation rates at the lowest effective field and in the limit of infinitely fast pulsing.

Figure S1 (following page). Overview of structural results arranged according to the evolutionary trajectory (Fig. 1B). Additional structures are at the bottom left. Numbered footnotes refer to the five different crystal forms, as described below. All structures were superposed to the apo crystal structure of wild type HisA (1), which has 231 ordered residues. Structure superpositioning and calculation of the root mean square deviation (RMSD) for the matching C_α atoms was done using the LSQ commands in O (16, 17).

[#1] Space group $P6_122$; a = b ≈ 87 Å, c ≈ 122 Å; 1 molecule per asymmetric unit.
Loops 1 and 6 are disordered in this crystal form and loop 5 is involved in crystal packing, preventing formation of the HisA-active conformation even in presence of ProFAR.

[#2] Space group $P3_121$; a = b ≈ 47 Å, c ≈ 198 Å; 1 molecule per asymmetric unit.

[#3] Space group $P6_122$; a = b ≈ 62 Å, c ≈ 607 Å; 2 molecules per asymmetric unit.

[#4] Space group $P3_221$; a = b ≈ 61 Å, c ≈ 393 Å; 3 molecules per asymmetric unit.

[#5] Space group $P2_12_12_1$; a ≈ 49 Å, b ≈ 90 Å, c ≈ 118 Å; 2 molecules per asymmetric unit.

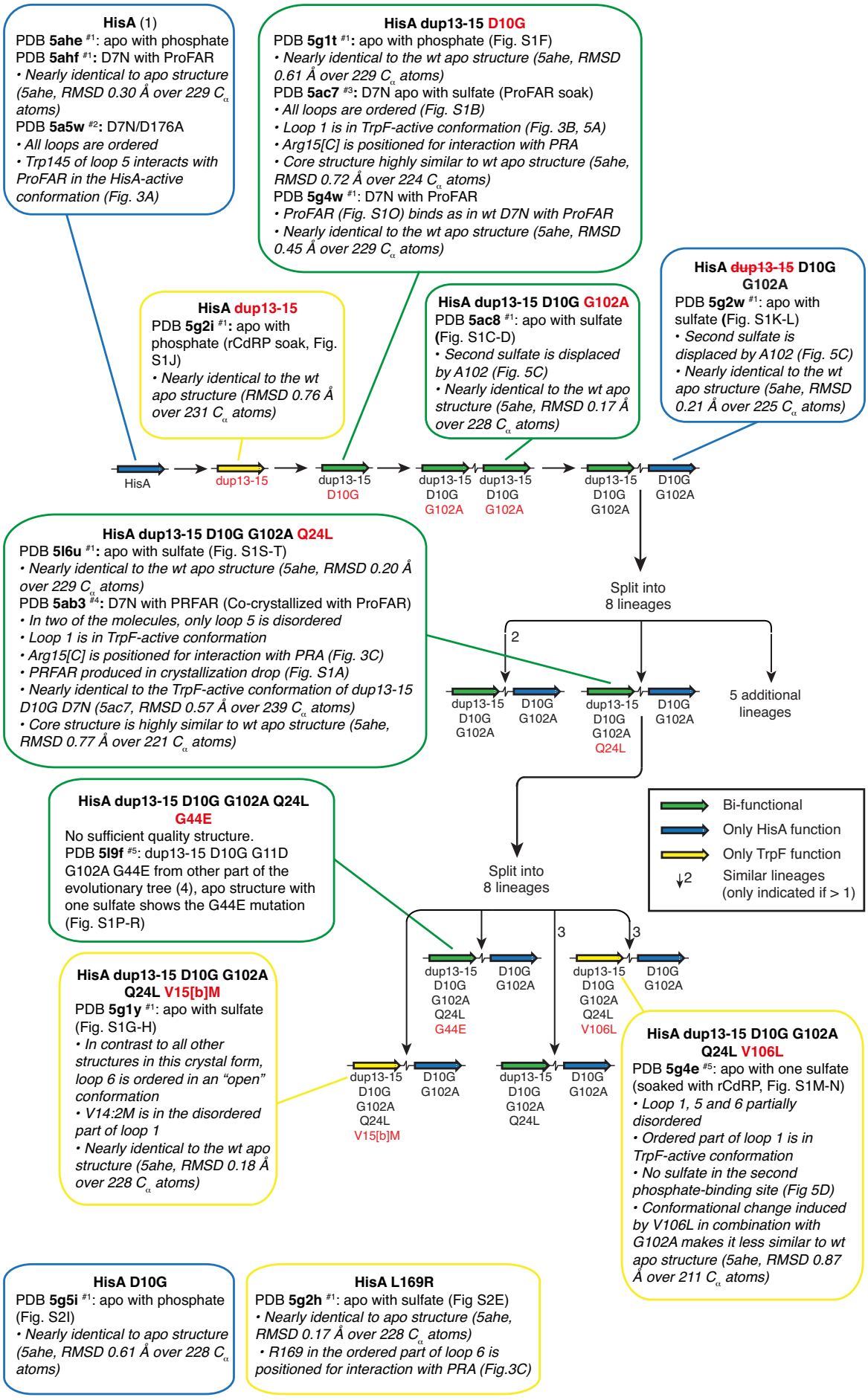
HisA (1)
PDB 5ahe #1: apo with phosphate
PDB 5ahf #1: D7N with ProFAR
• Nearly identical to apo structure (5ahe, RMSD 0.30 Å over 229 Cα atoms)
PDB 5a5w #2: D7N/D176A
• All loops are ordered
• Trp145 of loop 5 interacts with ProFAR in the HisA-active conformation (Fig. 3A)
HisA dup13-15 D10G
PDB 5g1t #1: apo with phosphate (Fig. S1F)
• Nearly identical to the wt apo structure (5ahe, RMSD 0.61 Å over 229 Cα atoms)
PDB 5ac7 #3: D7N apo with sulfate (ProFAR soak)
• All loops are ordered (Fig. S1B)
• Loop 1 is in TrpF-active conformation (Fig. 3B, 5A)
• Arg15[C] is positioned for interaction with PRA
• Core structure highly similar to wt apo structure (5ahe, RMSD 0.72 Å over 224 Cα atoms)
PDB 5g4w #1: D7N with ProFAR
• ProFAR (Fig. S1O) binds as in wt D7N with ProFAR
• Nearly identical to the wt apo structure (5ahe, RMSD 0.45 Å over 229 Cα atoms)
HisA dup13-15
PDB 5g2i #1: apo with phosphate (rCdRP soak, Fig. S1J)
• Nearly identical to the wt apo structure (RMSD 0.76 Å over 231 Cα atoms)
HisA dup13-15 D10G G102A
PDB 5ac8 #1: apo with sulfate (Fig. S1C-D)
• Second sulfate is displaced by A102 (Fig. 5C)
• Nearly identical to the wt apo structure (5ahe, RMSD 0.17 Å over 228 Cα atoms)
HisA dup13-15 D10G G102A
PDB 5g2w #1: apo with sulfate (Fig. S1K-L)
• Second sulfate is displaced by A102 (Fig. 5C)
• Nearly identical to the wt apo structure (5ahe, RMSD 0.21 Å over 225 Cα atoms)
HisA
dup13-15
dup13-15 D10G
dup13-15 D10G G102A
dup13-15 D10G G102A
dup13-15 D10G G102A
D10G G102A
HisA dup13-15 D10G G102A Q24L
PDB 5l6u #1: apo with sulfate (Fig. S1S-T)
• Nearly identical to the wt apo structure (5ahe, RMSD 0.20 Å over 229 Cα atoms)
PDB 5ab3 #4: D7N with PRFAR (Co-crystallized with ProFAR)
• In two of the molecules, only loop 5 is disordered
• Loop 1 is in TrpF-active conformation
• Arg15[C] is positioned for interaction with PRA (Fig. 3C)
• PRFAR produced in crystallization drop (Fig. S1A)
• Nearly identical to the TrpF-active conformation of dup13-15 D10G D7N (5ac7, RMSD 0.57 Å over 239 Cα atoms)
• Core structure is highly similar to wt apo structure (5ahe, RMSD 0.77 Å over 221 Cα atoms)
Split into 8 lineages
2
dup13-15 D10G G102A
D10G G102A
dup13-15 D10G G102A Q24L
D10G G102A
5 additional lineages
HisA dup13-15 D10G G102A Q24L G44E
No sufficient quality structure.
PDB 5l9f #5: dup13-15 D10G G11D G102A G44E from other part of the evolutionary tree (4), apo structure with one sulfate shows the G44E mutation (Fig. S1P-R)
Split into 8 lineages
3
3
dup13-15 D10G G102A Q24L G44E
D10G G102A
dup13-15 D10G G102A Q24L V106L
D10G G102A
HisA dup13-15 D10G G102A Q24L V15[b]M
PDB 5g1y #1: apo with sulfate (Fig. S1G-H)
• In contrast to all other structures in this crystal form, loop 6 is ordered in an "open" conformation
• V14:2M is in the disordered part of loop 1
• Nearly identical to the wt apo structure (5ahe, RMSD 0.18 Å over 228 Cα atoms)
dup13-15 D10G G102A Q24L V15[b]M
D10G G102A
dup13-15 D10G G102A Q24L
D10G G102A
HisA dup13-15 D10G G102A Q24L V106L
PDB 5g4e #5: apo with one sulfate (soaked with rCdRP, Fig. S1M-N)
• Loop 1, 5 and 6 partially disordered
• Ordered part of loop 1 is in TrpF-active conformation
• No sulfate in the second phosphate-binding site (Fig 5D)
• Conformational change induced by V106L in combination with G102A makes it less similar to wt apo structure (5ahe, RMSD 0.87 Å over 211 Cα atoms)
Bi-functional
Only HisA function
Only TrpF function
↓2 Similar lineages (only indicated if > 1)
HisA D10G
PDB 5g5i #1: apo with phosphate (Fig. S2I)
• Nearly identical to apo structure (5ahe, RMSD 0.61 Å over 228 Cα atoms)
HisA L169R
PDB 5g2h #1: apo with sulfate (Fig S2E)
• Nearly identical to apo structure (5ahe, RMSD 0.17 Å over 228 Cα atoms)
• R169 in the ordered part of loop 6 is positioned for interaction with PRA (Fig.3C)

Figure S2 (following two pages). Electron density maps for each structure, showing ligands and mutation sites. Green mesh: *Fo-Fc* omit map contoured at 3.0 σ; gray mesh: *2Fo-Fc* map contoured at 1.0 σ. (**A**) PRFAR from HisA(D7N/D10G/dup13-15/Q24L/G102A) that was co-crystallized with its substrate, ProFAR (converted to the product PRFAR during crystallization). (**B**) Mutation sites from the HisA(D7N/D10G/dup13-15) structure co-crystallized with ProFAR. (**C-D**) Mutation sites from the HisA(D10G/dup13-15/G102A) apo structure. (**E**) Mutation site from the HisA(L169R) structure soaked with rCdRP. (**F**) Mutation sites from the HisA(D10G/dup13-15) apo structure. (**G-H**) Mutation sites from the HisA(D10G/dup13-15/V15[b]M/Q24L/G102A) apo structure. (**I**) Mutation site from the HisA(D10G) apo structure. (**J**) Mutation sites from HisA (dup13-15) structure soaked with rCdRP. (**K-L**) Mutation sites from the HisA(D10G/G102A) apo structure. (**M-N**) Mutation sites from the HisA(D10G/dup13-15/Q24L/G102A/V106L) structure soaked with rCdRP. (**O**) ProFAR from the structure of HisA(D7N/D10G/dup13-15) soaked with ProFAR. (**P-R**) Mutation sites from the HisA(D10G/G11D/dup13-15/G44E/G102A) apo structure. (**S-T**) Mutation sites from the HisA(D10G/dup13-15/Q24L/G102A) apo structure.

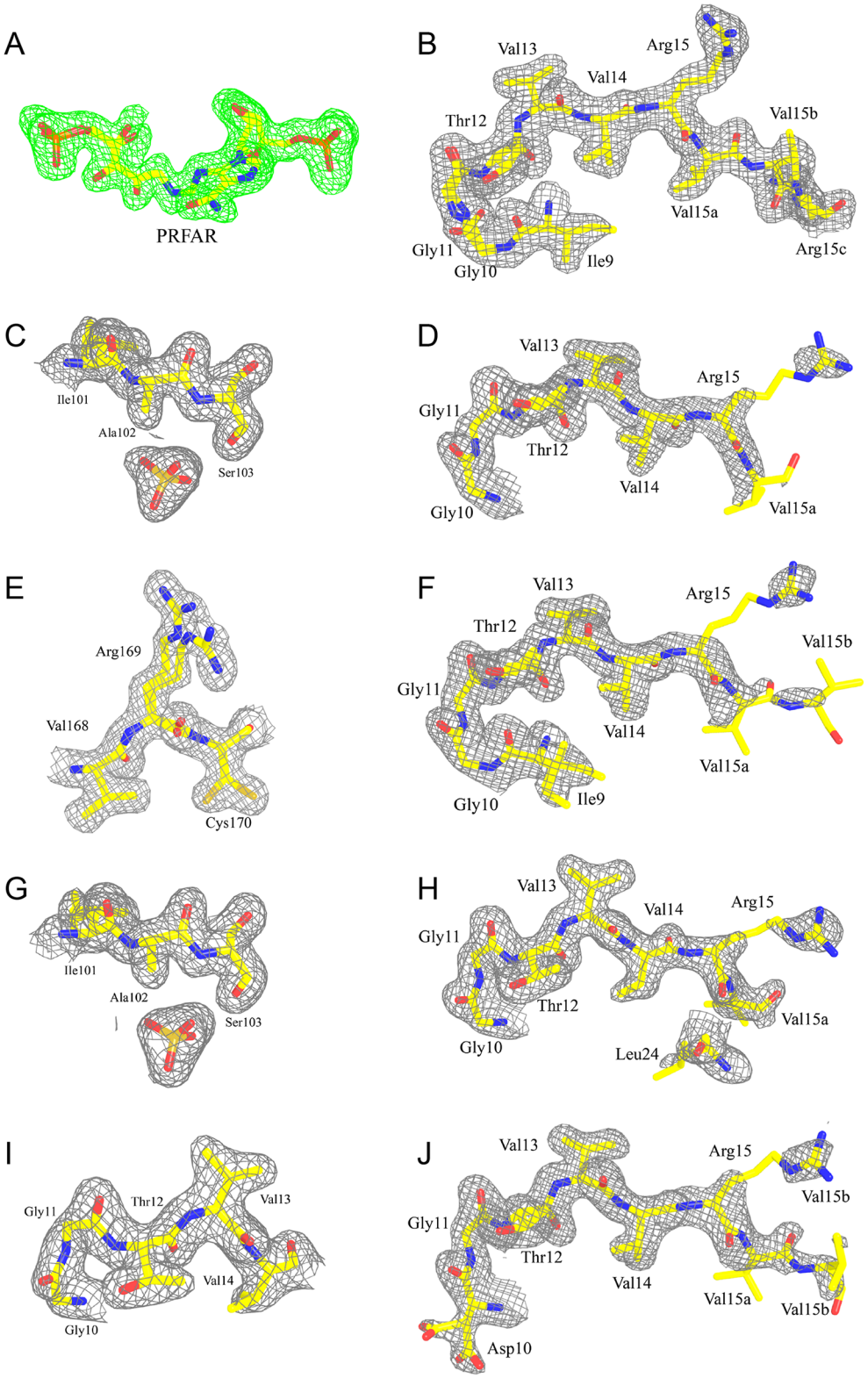

Figure S2. (legend on previous page.)

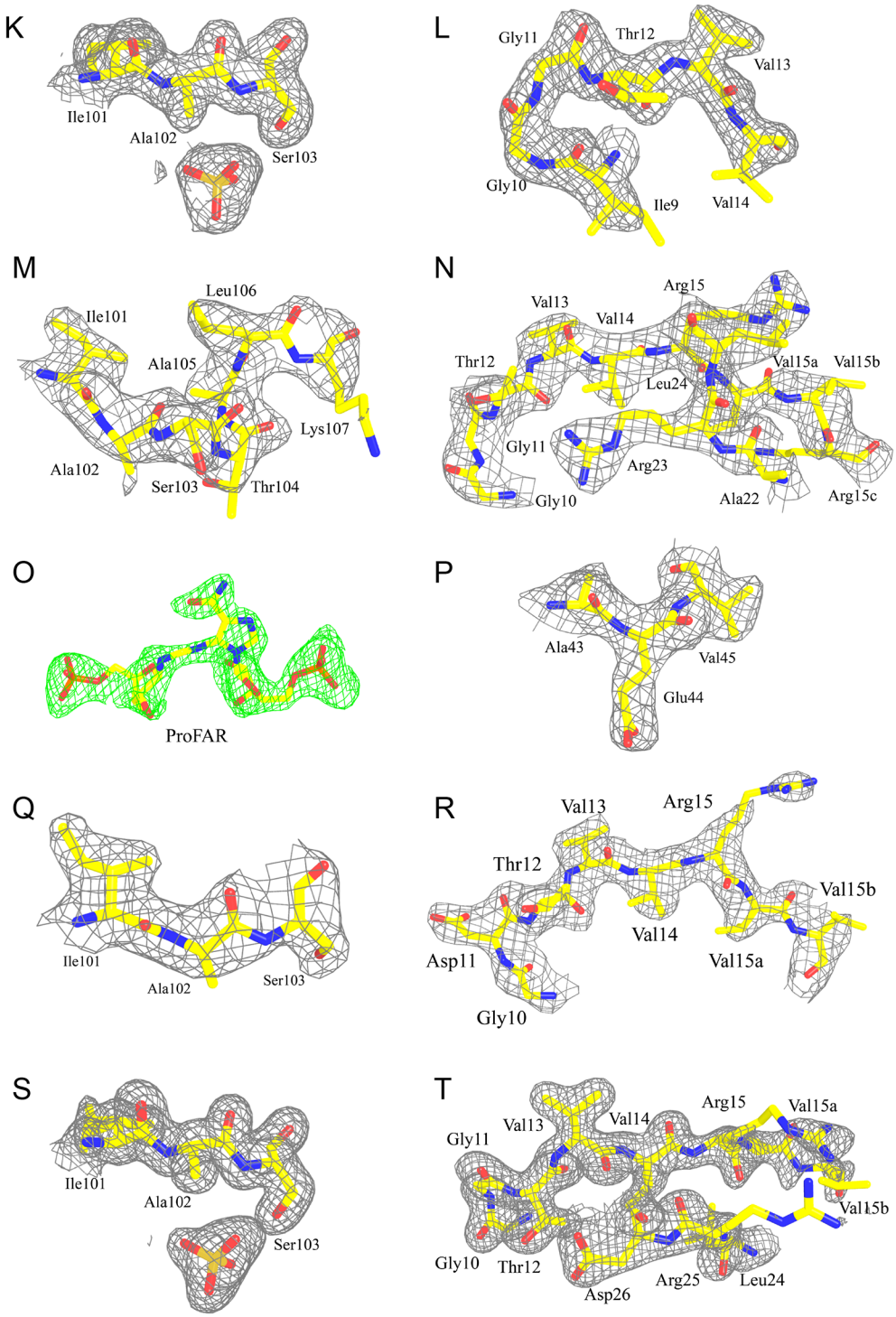

Figure S2 (continued).

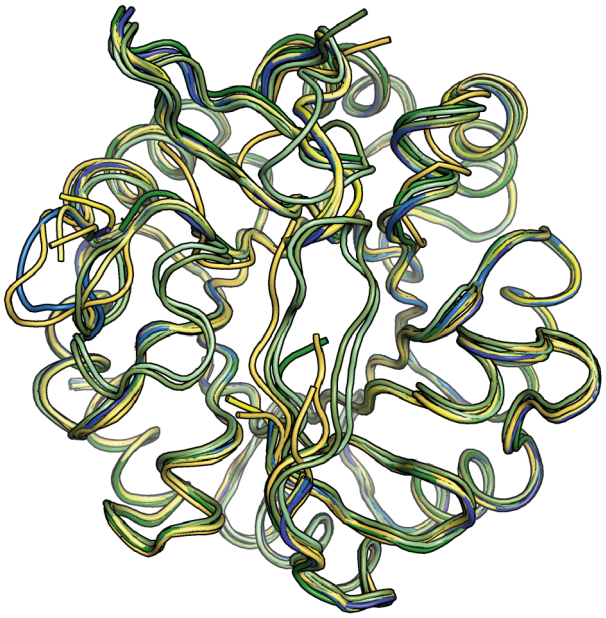

Figure S3. Overlay of all HisA mutant structures (listed in Table S2). HisA specialists are colored in blue, TrpF specialists in yellow and bi-functional enzymes in green. The catalytic face of the enzyme is oriented towards the viewer.

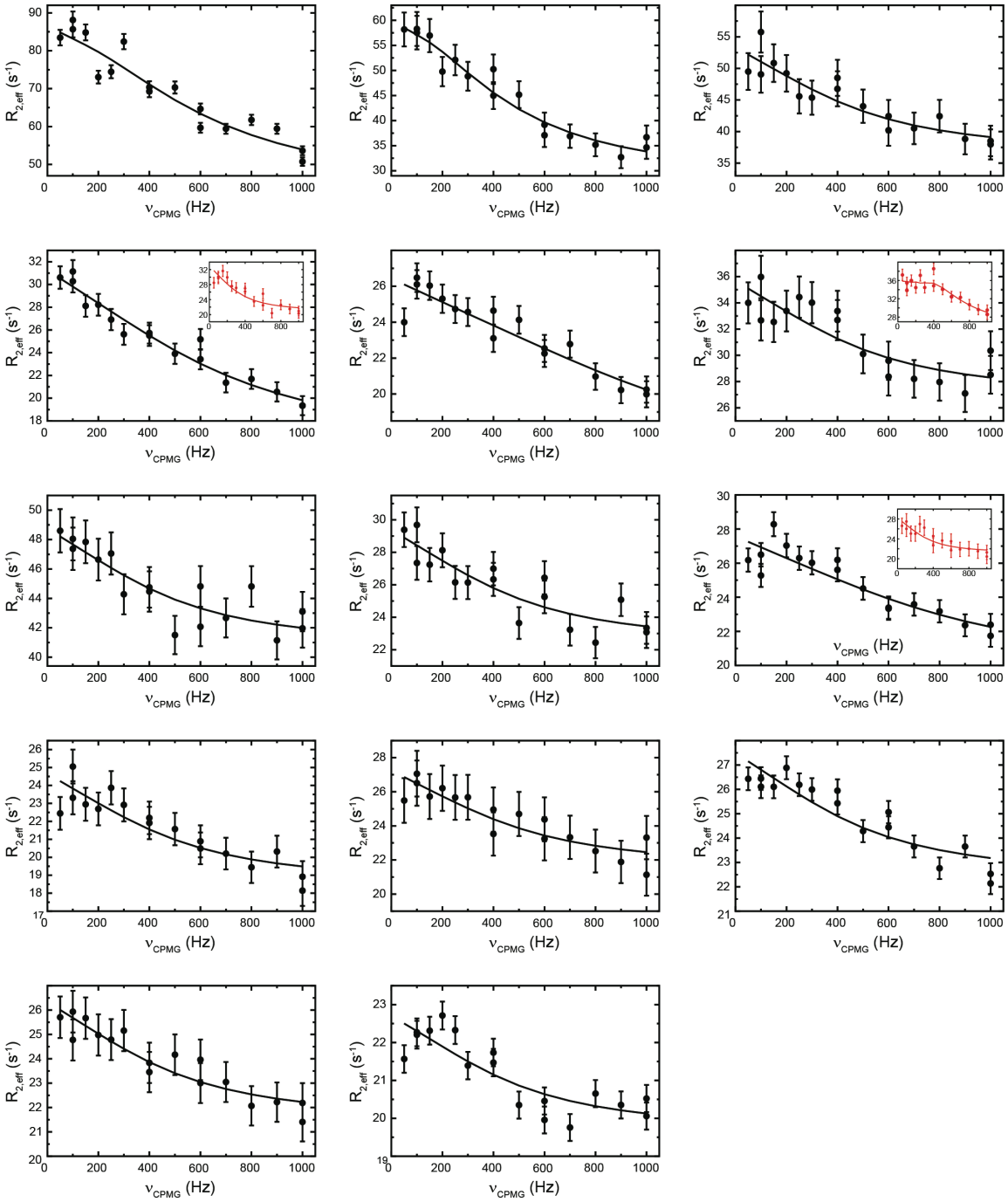

Figure S4. NMR ^{15}N CPMG relaxation dispersions for resonances with significant conformational dynamics for HisA(dup13-15/D10G). Effective transverse relaxation rates, $R_{2,eff}$, at different effective fields, ν_{CPMG}, are shown as filled circles and the lines represent the best fit to the Bloch-McConnell equations for two-site exchange. The insets show corresponding significant relaxation dispersions for HisA(dup13-15). The ordering of the panels is the same as the order of the bars in Fig. 3D.

Table S1. Characteristics of bi-functional HisA mutants.

Enzyme	$(k_{cat}/K_M)^{HisA}$ $(s^{-1}.M^{-1})$	$(k_{cat}/K_M)^{TrpF}$ $(s^{-1}.M^{-1})$	$(k_{cat}/K_M)^{HisA} \div (k_{cat}/K_M)^{TrpF}$	**Growth rate as HisA (+Trp), h^{-1}**	**Growth rate as TrpF (+His), h^{-1}**	**Growth rate when bi-functional, h^{-1}**
HisA(dup13-15/D10G)	2.8×10^4	51	550	0.189	0.295	0.143
HisA(dup13-15/D10G/G102A)	9.2×10^3	220	42	0.270	0.409	0.212
HisA(dup13-15/D10G/G102A/Q24L)	5.1×10^3	260	20	0.162	0.509	0.155
HisA(dup13-15/D10G/G102A/Q24L/G44E)	5.2×10^3	140	37	0.238	0.534	0.256
HisA(dup13-15/D10G/G102A/G11D/G44E)	6.7×10^3	130	52	0.526	0.531	0.423

Specificity constants are reproduced from Table 1. Growth rate data were reported previously (4).

Table S2. Summary of crystallographic data, refinement statistics and crystallization.

***S. enterica* HisA variant**	**D10G, dup13-15**	**D7N, D10G, dup13-15**	**D10G, dup13-15, G102A**	**D10G, G11D, dup13-15, G44E, G102A**
Data collection				
Beam line	ID23-2	I02	ID23-1	ID23-2
Detector	CCD	Pilatus 6M-F	Pilatus 6M	CCD
Space group	$P6_122$	$P6_122$	$P6_122$	$P2_12_12_1$
Unit cell parameters:				
a, b, c (Å)	87.2, 87.2, 121.9	62.0, 62.0, 607.1	86.4, 86.4, 122.0	49.2, 91.1, 117.9
α, β, γ (°)	90, 90, 120	90, 90,120	90, 90, 120	90, 90, 90
Molecules per asymmetric unit	1	2	1	2
Matthew's coefficient ($Å^3$/Da)	2.3	3.3	2.3	2.4
Resolution range (Å)*	50.00-1.70 (1.80-1.70)	50.00-1.90 (1.95-1.90)	50.00-1.70 (1.73-1.70)	50.00-2.59 (2.71-2.59)
Wavelength	0.8726	0.9795	0.9763	0.8726
Total reflections*	330372 (52037)	2137134 (138797)	578357 (21225)	82382 (6916)
Unique reflections*	30762 (4752)	56836 (3669)	30311 (1516)	17013 (2007)
Completeness (%)*	99.9 (100)	99.8 (97.8)	99.8 (97.1)	99.7 (98.1)
Redundancy*	10.7 (11.0)	37.6 (37.8)	19.1 (14.0)	4.8 (4.8)
R_{meas}(%)*	4.9 (45.6)	9.1 (115.9)	15.7 (216.5)	20.4 (79.5)
I/σ(I)*	30.7 (5.4)	30.0 (4.1)	13.9 (1.3)	9.0 (2.7)
Wilson B-factor ($Å^2$)	29.2	27.8	19.6	26.5
CC ½ (%)	100 (95.2)	100 (93.6)	99.9 (40.7)	98.8 (80.0)
Data scaling software	XDS	Aimless	Aimless	Aimless
Refinement				
Resolution range	28.55-1.70	49.10-1.90	47.30-1.70	49.50-2.59
Reflection / test set	30762 / 1539	56576 / 2868	30248 / 1529	16926 / 856
Number of non-hydrogen atoms	2007	3981	1905	3420
macromolecules	1829	3720	1779	3361
ligands	30	28	15	12
water	147	233	109	47
R_{work}/R_{free} (%)	17.2 / 20.0	18.6/21.4	19.6/20.9	22.2/28.9
Average B-factors:				
overall	28.2	41.9	27.9	32.0
macromolecules	27.4	41.9	27.5	32.1
ligands	41.4	59.6	44.7	42.5
water	35.3	40.4	30.9	25.6
RMSD from ideal bond lengths	0.007	0.007	0.008	0.009
RMSD from ideal bond angles	1.06	1.10	1.15	1.27
Ramachandran plot:				
Preferred (%)	99	98	98	97
Allowed (%)	1	2	2	3
Outliers (%)	0	0	0	0
Model				
Protein residues#	1-15[b], 24-174, 183-244	ChainA: 1-243; ChainB 1-244	1-15[a], 24-179,184-244	ChainA:1-15[b], 23-142, 151-174, 183-244; ChainB: 1-15[b], 23-141, 150-173, 183-244
Ligands	HEPES, phosphate	Sulfate	Sulfate	Sulfate
Crystallization	0.1 M HEPES pH 7.5, 0.8 M sodium phosphate, 0.8 M potassium phosphate	2.0 M ammonium sulfate, 0.1 M Bis-Tris pH 5.5	2.0 M ammonium sulfate, 0.1 M sodium acetate pH 4.6	0.2 M lithium sulfate, 0.1 M sodium acetate pH 4.5, 30% w/v PEG8000
Ligand addition		Co-crystallized with ProFAR		
Comments	—	—	—	—
PDB accession number	**5G1T**	**5AC7**	**5AC8**	**5L9F**

* Values in parentheses refer to the highest resolution shell.

Residues 10-30 represent loop 1, 129-152 loop 5 and 172-183 loop 6.

***S. enterica* HisA variant**	**D7N, D10G, dup13-15, Q24L, G102A**	**D10G, dup13-15, V15[b]M, Q24L, G102A**	**D10G, dup13-15, Q24L, G102A**	**D10G, dup13-15, Q24L, G102A, V106L**
Data collection				
Beam line	I02	ID14-4	P14	ID23-2
Detector	Pilatus 6M-F	CCD	Pilatus	CCD
Space group	P3221	P6122	P6122	P212121
Unit cell parameters:				
a, b, c (Å)	61.2, 61.2, 391.9	86.0, 86.0, 122.2	85.9, 85.9, 121.0	48.8, 88.8, 117.8
α, β, γ (°)	90,90,120	90, 90, 120	90, 90, 120	90, 90, 90
Molecules per asymmetric unit	3	1	1	2
Matthew's coefficient ($Å^3$/Da)	2.8	2.3	2.4	2.3
Resolution range (Å)*	50.00-1.80 (1.84-1.80)	50.00-1.80 (1.84-1.80)	74.43-1.60 (1.70-1.60)	50.00-2.40 (2.50-2.40)
Wavelength	0.9795	0.9763	0.9763	0.8726
Total reflections*	755447 (41474)	337881 (11073)	1359688 (223319)	81952 (7078)
Unique reflections*	80382 (4268)	25362 (1367)	35399 (5723)	19789 (791)
Completeness (%)*	99.6 (94.8)	99.5 (93.2)	99.8 (99.5)	95.5 (76.9)
Redundancy*	9.4 (9.7)	13.3 (8.1)	38.4 (38.8)	4.1 (8.9)
R_{meas}(%)*	7.2 (95.9)	9.9 (109.9)	8.2 (162.4)	16.8 (101.1)
I/σ(I)*	15.7 (2.4)	19.0 (2.0)	36.1 (2.77)	9.9 (1.6)
Wilson B-factor ($Å^2$)	27.9	17.0	30.5	44.3
CC ½ (%)	99.9 (88.0)	99.9 (64.5)	100 (80.7)	99.5 (65.6)
Data scaling software	Aimless	Aimless	XDS	Aimless
Refinement				
Resolution range	49.10-1.80	47.25 - 1.80	74.43-1.60	45.10-2.65
Reflection / test set	80205 / 4024	25335 / 1291	35398 / 1769	15276 / 790
Number of non-hydrogen atoms	6206	1962	2015	3230
macromolecules	5521	1823	1886	3176
ligands	152	33	15	28
water	532	105	114	26
R_{work}/R_{free} (%)	17.2/20.6	16.6/21.0	15.5/18.9	24.0/29.9
Average B-factors:				
overall	40.3	28.8	31.2	45.6
macromolecules	40.2	28.1	30.6	45.7
ligands	34.3	47.5	52.8	49.0
water	42.4	34.7	38.6	37.8
RMSD from ideal bond lengths	0.007	0.037	0.027	0.004
RMSD from ideal bond angles	1.08	1.81	2.55	0.760
Ramachandran plot:				
Preferred (%)	97	98	98	96
Allowed (%)	3	2	2	3.75
Outliers (%)	0	0	0	0.25
Model				
Protein residues#	ChainA: 1-142, 149-244; ChainB: 1-16, 19-141, 150-244; ChainC: 1-142, 147-244	1-15[a], 24-174, 178-180, 182-244	1-15B, 22-174, 180-181, 183-244	Chain A: 1-17, 21-129, 138-141, 149-172, 182-244; Chain B: 1-15[a], 25-129, 151-172, 182-244
Ligands	PRFAR	Sulfate, glycerol	Sulfate	
Crystallization	0.1 M MOPS/HEPES-Na pH7.5, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 10% w/v PEG20000, 20% v/v PEG MME 550	0.2 M ammonium sulfate, 0.1 M sodium acetate pH 4.6, 25% w/v PEG4000	0.2 M ammonium sulfate, 0.1 M sodium acetate pH 4.8, 25% w/v PEG4000	0.18 M ammonium acetate, 0.09 M sodium acetate, pH 5.15, 27% w/v PEG4000
Ligand addition	Co-crystallized with ProFAR			Soaked with rCdRP
Comments	—	—	—	138-141 in chain A built as polyalanine
PDB accession number	**5AB3**	**5G1Y**	**5L6U**	**5G4E**

* Values in parentheses refer to the highest resolution shell.

Residues 10-30 represent loop 1, 129-152 loop 5 and 172-183 loop 6.

***S. enterica* HisA variant**	**D7N, D10G, dup13-15**	**dup13-15**	**D10G, G102A**	**L169R**
Data collection				
Beam line	ID23-2	BM30	ID14-4	IO2
Detector	Pilatus	ADSC	ADSC	Pilatus M6
Space group	P6122	P6122	P6122	P6122
Unit cell parameters:				
a, b, c (Å)	87.0, 87.0, 120.5	86.6, 86.6, 121.4	86.0, 86.0, 121.6	86.0, 86.0, 122.7
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Molecules per asymmetric unit	1	1	1	1
Matthew's coefficient ($Å^3$/Da)	2.4	2.4	2.4	2.4
Resolution range (Å)*	50.00-2.30 (2.44-2.30)	50.00-1.60 (1.70-1.60)	50.00-2.10 (2.23-2.10)	50.00-1.90 (2.02-1.90)
Wavelength	0.8726	0.9797	0.9763	0.9795
Total reflections*	170686 (27112)	754966 (123820)	277592 (44157)	314134 (49510)
Unique reflections*	22042 (3547)	35938 (5789)	16100 (2515)	40024 (6473)
Completeness (%)*	97.3 (97.2)	99.4 (98.4)	99.8 (99.4)	100 (99.8)
Redundancy*	7.7 (7.6)	21.0 (21.4)	17.2 (17.6)	7.8 (7.6)
R_{meas}(%)*	26.6 (117.8)	5.5 (62.4)	9.0 (62.2)	17.8 (75.8)
I/σ(I)*	8.5 (1.8)	42.1 (6.1)	29.2 (5.3)	9.2 (2.3)
Wilson B-factor ($Å^2$)	34.7	26.2	35.8	25.7
CC ½ (%)	99.1 (57.8)	100 (95.7)	99.9 (93.7)	99.5 (79.9)
Data scaling software	XDS	XDS	XDS	XDS
Refinement				
Resolution range	47.00-2.30	47.18-1.60	47.10-2.10	47.40-1.90
Reflection / test set	22041 / 1100	35934 / 1797	16099 / 805	40017 / 1982
Number of non-hydrogen atoms	1865	1952	1828	1987
macromolecules	1747	1809	1766	1827
ligands	49	16	10	39
water	69	125	52	121
R_{work}/R_{free} (%)	19.9/26.4	18.3/20.6	16.8/22.0	16.9/21.2
Average B-factors:				
overall	30.1	25.4	36.6	25.8
macromolecules	29.8	24.8	36.4	25.1
ligands	38.5	28.3	57.7	37.8
water	31.6	32.8	38.5	32.6
RMSD from ideal bond lengths	0.004	0.006	0.009	0.010
RMSD from ideal bond angles	0.792	1.06	1.07	1.19
Ramachandran plot:				
Preferred (%)	97	99	98	98
Allowed (%)	3	1	2	2
Outliers (%)	0	0	0	0
Model				
Protein residues#	1-15[a], 24-174, 180-244	1-15[b], 24-174, 182-244	1-14, 26-174, 183-244	1-16, 24-175, 178-244
Ligands	ProFAR	Phosphate	Sulfate	Sulfate, glycerol
Crystallization	2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5	0.8 M sodium phosphate monobasic, 0.8 M potassium phosphate monobasic, 0.1 M HEPES pH 7.5	0.2 M ammonium acetate, 0.1 M sodium acetate, pH 5.15, 20% PEG4000	2.3 M ammonium sulfate, 0.1 M sodium acetate pH 4.6
Ligand addition	Soaked with ProFAR in cryo buffer: 50 mM HEPES sodium pH 7, 150 mM sodium chloride, 30% PEG4000, 15% glycerol	Soaked with rCdRP		Soaked with rCdRP in cryo buffer: 50 mM HEPES sodium pH 7, 150 mM sodium chloride, 30% PEG4000, 15% glycerol
Comments	—	—	—	—
PDB accession number	**5G4W**	**5G2I**	**5G2W**	**5G2H**

* Values in parentheses refer to the highest resolution shell.

Residues 10-30 represent loop 1, 129-152 loop 5 and 172-183 loop 6.

***S. enterica* HisA variant**	**D10G**
Data collection	
Beam line	ID14-4
Detector	CCD
Space group	P6122
Unit cell parameters:	
a, b, c (Å)	85.9, 85.9, 121,4
α, β, γ (°)	90, 90, 120
Molecules per asymmetric unit	1
Matthew's coefficient ($Å^3$/Da)	2.4
Resolution range (Å)*	50.00-2.00 (2.10-2.00)
Wavelength	0.9164
Total reflections*	216796 (28674)
Unique reflections*	18459 (2409)
Completeness (%)*	99.7 (98.2)
Redundancy*	11.7 (11.7)
R_{meas}(%)*	12.2 (85.6)
I/σ(I)*	17.5 (3.4)
Wilson B-factor ($Å^2$)	32.5
CC ½ (%)	99.9 (87.8)
Data scaling software	XDS
Refinement	
Resolution range	47.00-2.00
Reflection / test set	18458 / 923
Number of non-hydrogen atoms	1866
macromolecules	1755
ligands	30
water	81
R_{work}/R_{free} (%)	17.1/21.0
Average B-factors:	
overall	32.1
macromolecules	31.6
ligands	47.3
water	35.9
RMSD from ideal bond lengths	0.008
RMSD from ideal bond angles	1.13
Ramachandran plot:	
Preferred (%)	98
Allowed (%)	2
Outliers (%)	0
Model	
Protein residues#	1-15, 24-175, 184-244
Ligands	Phosphate
Crystallization	0.8 M sodium phosphate monobasic, 0.8 M potassium phosphate monobasic, 0.1 M HEPES pH 7.5
Ligand addition	
Comments	—
PDB accession number	**5G5I**

* Values in parentheses refer to the highest resolution shell.

Residues 10-30 represent loop 1, 129-152 loop 5 and 172-183 loop 6.

SI References

1. Söderholm A, et al. (2015) Two-step ligand binding in a $(\beta\alpha)_8$ barrel enzyme: substrate-bound structures shed new light on the catalytic cycle of HisA. *J Biol Chem* 290:24657-24668.
2. Patrick WM, Matsumura I (2008) A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J Mol Biol* 377:323-336.
3. Gerth ML, Nigon LV, Patrick WM (2012) Characterization of the indole-3-glycerol phosphate synthase from *Pseudomonas aeruginosa* PAO1. *Protein J* 31:359-365.
4. Näsvall J, Sun L, Roth JR, Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338:384-387.
5. Knöppel A, Näsvall J, Andersson DI (2016) Compensating the fitness costs of synonymous mutations. *Mol Biol Evol* 33:1461-1477.
6. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66:125-132.
7. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235-242.
8. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
9. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.
10. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68:352-367.
11. Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* 94:630-638.
12. Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29:688-691.
13. Delaglio F, et al. (1995) NMRPipe - a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-293.
14. Ahlner A, Carlsson M, Jonsson BH, Lundström P (2013) PINT: a software for integration of peak volumes and extraction of relaxation rates. *J Biomol NMR* 56:191-202.
15. McConnell HM (1958) Reaction rates by nuclear magnetic resonance. *J Chem Phys* 28:430-431.

16. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica* A47:110-119.
17. Kleywegt GJ, Jones TA (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol* 277:525-545.