Supporting Information

Structural basis for specificity and promiscuity in a carrier protein/enzyme system from the sulfur cycle

Daniel B. Grabarczyk, Paul E. Chappell, Steven Johnson, Lukas S. Stelzl, Susan M. Lea, and Ben C. Berks

SI Materials and Methods

Genetic constructs

Expression plasmids based on pQE80L (Qiagen) and containing the *T. thermophilus* HB27 *soxB* (pVS048) or *soxYZ* (pVS056) genes have previously been described and produce proteins bearing a Strep II tag on SoxB and a hexa-histidine tag on SoxY (1). To increase expression and aid genetic manipulation analogous plasmids were constructed using codon-optimised *sox* genes. The codon-optimized genes were synthesised by Genscript and have the sequences shown in Fig. S14. The *soxB* coding region was synthesised with EcoRI and KpnI ends and used to replace the native *soxB* gene in pVS048 to generate pQE*soxB*$_{CO}$. Similarly *soxYZ* was synthesised with KpnI and HindIII sites to allow generation of pQE*soxYZ*co by replacement of the native *soxYZ* genes in pVS056.

All point mutants were constructed in the codon-optimised genes using the Quikchange method (Stratagene) and verified by sequencing. Replacement of the SoxZ Z-loop (residues 29-46) with a Gly-Ser-Gly linker was achieved by inverse PCR using primers 5′-ACGGCT*GGATCC*GGTTACATCAACCTGCTGGAAG-3′ and 5′-ACGGCT*GGATCC*ATGTTGTGCGACG ACTTG-3′ which contain BamH1 sites (underlined). The PCR product was digested with BamHI and circularised by ligation.

Production of protein samples

T. thermophilus SoxB and SoxYZ and all their variants were expressed and purified as previously described (1). All proteins were stored in 'ITC Buffer' (30 mM Tris-HCl, 200 mM NaCl, pH 8.0). Sox protein concentrations were routinely estimated by the Bradford assay (Biorad) performed in triplicate. The assay was calibrated with standards previously quantified by their A_{280nm} in 6 M guanadinium-HCl using extinction coefficients of 105 mM^{-1} cm^{-1} for SoxB and 7.45 mM^{-1} cm^{-1} for SoxYZ calculated with the program ProtParam (2).

Electrospray Ionization Mass Spectrometry

Samples were desalted prior to ESI-MS using a C_4 ZipTip (Millipore) according to the manufacturer's instruction. ESI-MS analysis was performed on a Micromass LCT (Waters) in positive ion mode in 50% v/v acetonitrile/water, 0.1% formic acid with an accuracy of ±0.1%. Spectra were acquired and analysed using MassLynx software (Waters Software).

Isothermal Titration Calorimetry

All ITC experiments were performed on a MicroCal iTC_{200} at 25°C with a reference power of 3 cal/s and a stirring speed of 1000 rpm. Experiments were carried out with SoxB in the cell and SoxYZ as the titrant. Both proteins were in ITC Buffer. For each protein variant control experiments were carried out to confirm that sample dilution did not cause systematic deviation from a flat baseline. Traces were integrated using OriginPro and then two replicate experiments were simultaneously fitted to a hetero-association model ($A + B \leftrightarrow AB$) with ΔH and K_D as fitting parameters using SEDPHAT (3). When the inflexion point was clear, inactive protein fractions were included as fitting parameters to account for small errors in protein concentration determination.

Surface plasmon resonance

All surface plasmon resonance (SPR) experiments were performed on a Biacore T200 (GE Healthcare) at 25°C. SoxB was immobilized to a CM5 chip using amine coupling chemistry. SoxB in sodium acetate pH 5.5 at 1 or 10 μg/mL was immobilised at a density of 100 and 200 response units, respectively. For each comparison a control reaction without SoxB was performed in an adjacent flow cell and the response subtracted from that of the SoxB-containing test cell. The sensogram from an injection of buffer onto the sensor chip was also subtracted

from each sample. Experiments were performed in 10 mM HEPES-NaOH, 150 mM NaCl, 0.005% Tween-20, pH 7.4. Analytes were injected in a randomized order for 30 seconds at a flow rate of 75 µL/min, and allowed to dissociate over 360 seconds. No regeneration step was required due to the fast dissociation rates. For each analyte loading density data were acquired for duplicate two-fold serial dilutions with 9 concentrations from 30 µM to 0.12 µM. The methods used to fit the data are described in the legend to Fig. S2.

Size exclusion chromatography - multi angle light scattering

Samples were loaded on a Superdex 200 10/300 size exclusion column (GE Healthcare) equilibrated in 30 mM Tris-HCl, 150 mM NaCl pH 8.0 and analysed by an in-line Dawn Heleos-II light scattering detector (Wyatt Technologies) and an Optilab-rex refractive index monitor (Wyatt Technologies). Molecular mass calculations were performed using ASTRA 5.3.4.14 (Wyatt Technologies) assuming a dn/dc value of 0.186 ml/g.

Building molecular models of the SoxB-SoxY(SSO_3^-)Z complex

Starting from the disulfide-linked structure, we restored native SoxB Trp175 in place of the introduced Cys residue and added an S-thiosulfonate group on to the SoxY carrier arm Cys. In one model we manually positioned the S-thiosulfonate group to co-ordinate the manganese ions using the position of thiosulfate in the previously-determined SoxB-thiosulfate structure as a guide (1). In a second model the SoxY C-terminal carboxylate ligation seen in the disulfide-linked SoxB-SoxYZ structure was retained. These models were solvated and then subjected to energy minimisation and position restrained molecular dynamics equilibration.

Molecular dynamics (MD) simulations used an AMBER 99SB force-field (4) with a TIP3P water model (5) run on GROMACS v4.5.6 (6). Previously described AMBER parameters were used for the manganese (II) ions (7). Cysteine-S-thiosulfonate partial charges were generated by quantum mechanical calculations using AmberTools13 (8) and Gaussian 03 (9), and bond

length, angle and dihedral parameters for the sulfonate group were taken from the AMBER GAFF force-field (10).

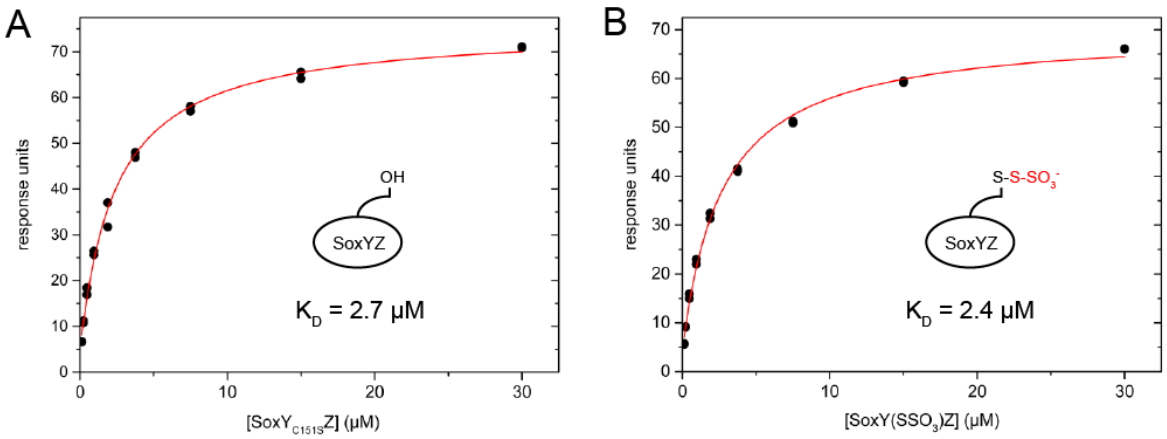


Figure S1. Affinities of SoxB-SoxYZ interactions determined by equilibrium SPR. (**A**) The $SoxY_{C151S}Z$ variant was used. (**B**) The $SoxY(SSO_3^-)Z$ adduct was used.

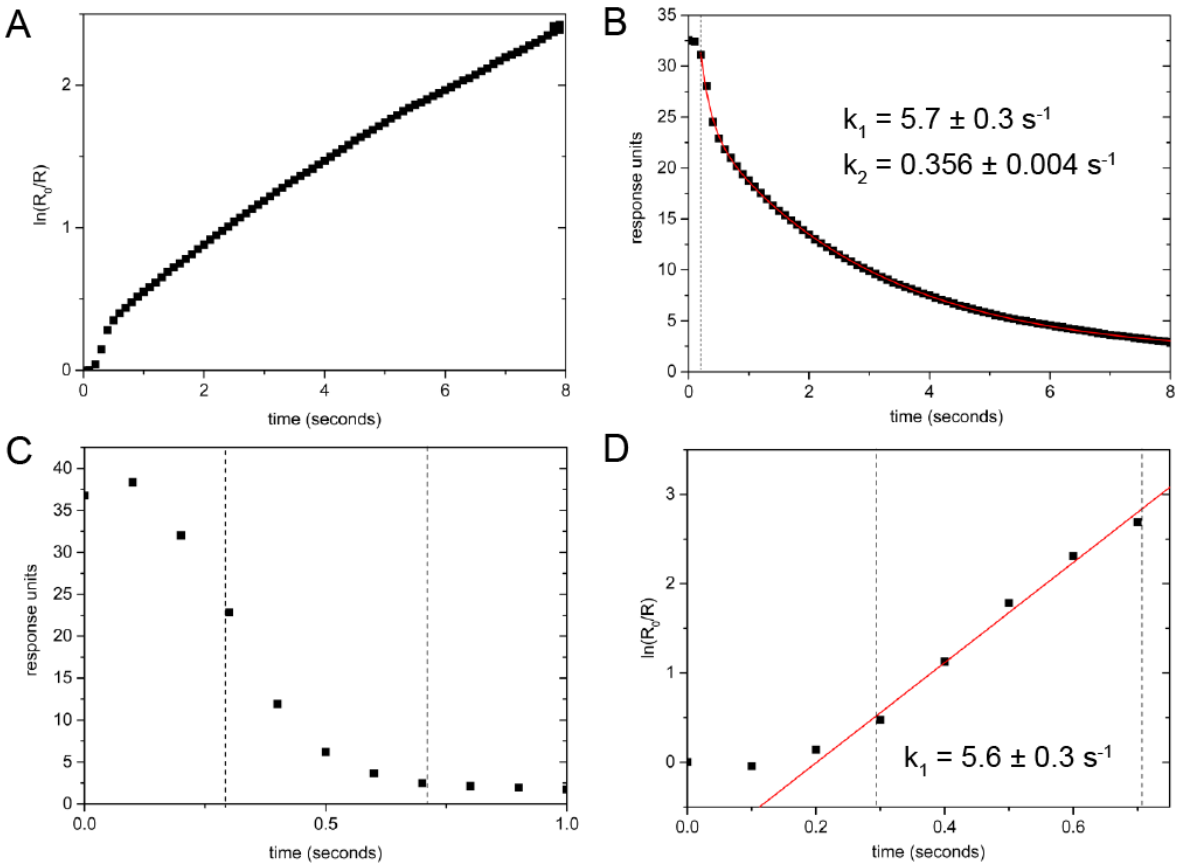


Figure S2. Representative SoxB-SoxYZ complex dissociation phase data showing the fitting methods used. Analysis of the SoxB-SoxY(SSO_3^-)Z complex is shown in (**A,B**) and analysis of the SoxB-SoxY$_{C151S}$Z complex in (**C,D**). The SoxYZ concentration in both cases was 1.88 µM. (**A**) A log plot of the dissociation phase for the SoxB-SoxY(SSO_3^-)Z complex shows deviation from linearity indicating biphasic behaviour with a burst phase before 0.5 seconds. The burst phase is associated with the dissociation of a minor proportion of the complexes and most likely arises from a small proportion of underivatized or incorrectly-derivatized SoxYZ in the sample. (**B**) The data in (A) were fit as a sum of two exponential decay processes using the equation:

$$Y = Y_0 + A_1.e^{k_1(x-x_0)} + A_2.e^{k_2(x-x_0)}$$

Here k_1 and k_2 represent the kinetic rate constants of each decay, and A_1 and A_2 the starting amplitude of each decay. The data is unreliable around the injection point. Therefore the fit does not include points on either side of the injection, and the time of injection is floated as the x_0 parameter. The A_1 and k_1 parameters are defined by only a few points and so are heavily influenced by injection artefacts. By contrast, the k_2 parameter can be robustly fitted across

different analyte concentrations and loading densities. The values of the kinetic rate constants given in the main text are the means and 95% confidence intervals of the fitted values at every concentration and their duplicates. The values shown in this figure derive from the fit to the single curve shown here. (**C**) The dissociation of the SoxB-SoxY$_{C151S}$Z complex is rapid and a significant proportion of the curve is affected by injection artefacts making it difficult to fit the data by non-linear regression fitting. (**D**) Consequently, a log plot was used to fit the k_1 value using only that part of the curve not influenced by the injection noise. The k_1 value in the main paper is the mean with 95% confidence intervals fitted at every concentration for two technical replicates.

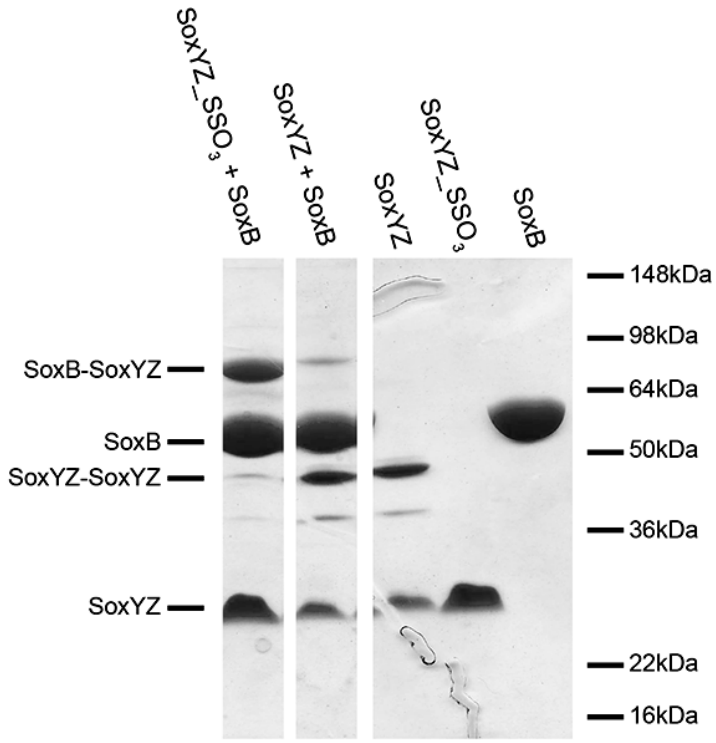
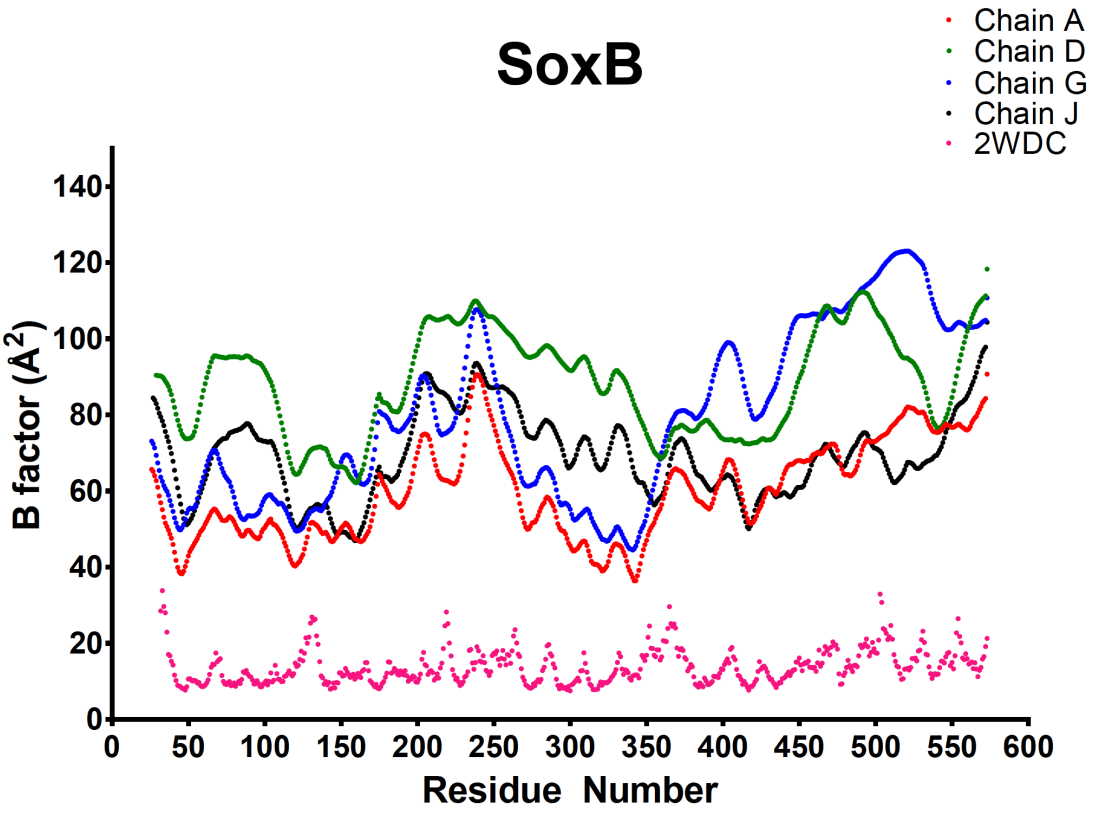
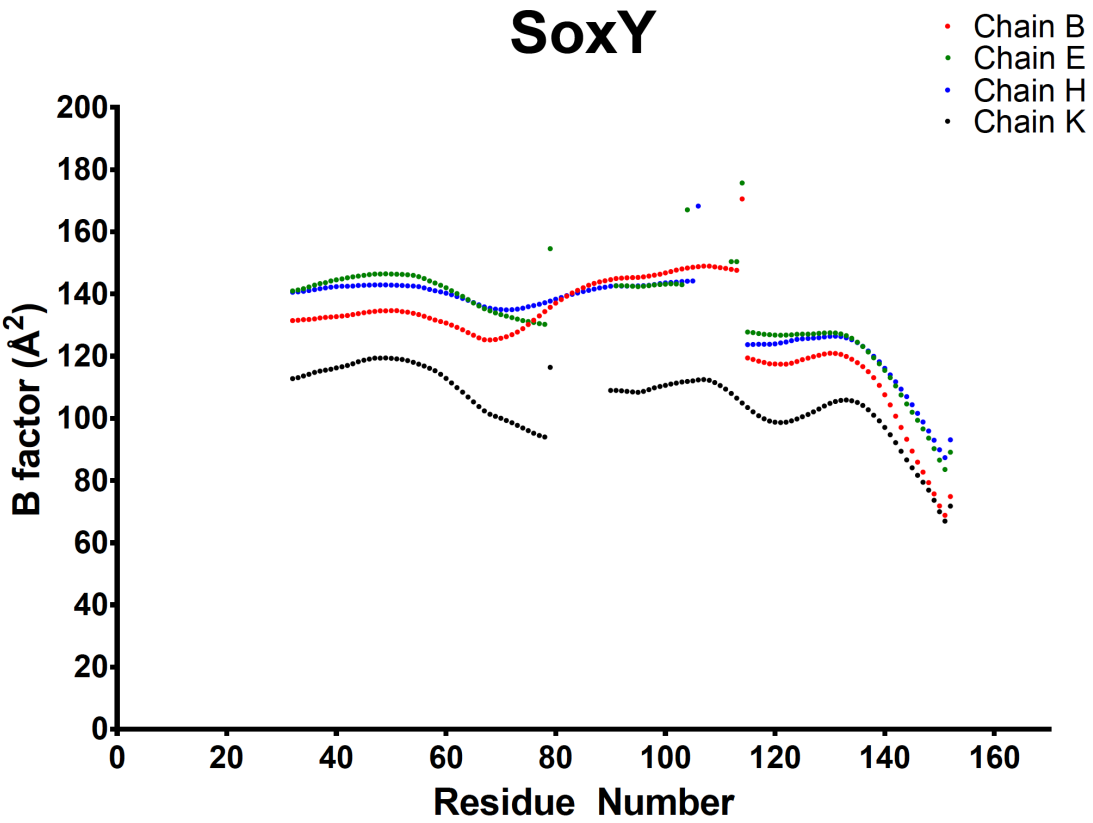


Figure S3. Disulfide cross-linking of SoxB and SoxYZ. Either 20 µM SoxYZ or SoxY(SSO_3^-)Z were incubated with 20 µM SoxB(Trp175Cys) at 70˚C for one hour. Cross-linking was visualised using a 12% polyacylamide concentration non-reducing SDS-PAGE gel and the samples were not heated prior to loading. Under these conditions the native SoxYZ complex does not dissociate.


SoxB
Chain A
Chain D
Chain G
Chain J
2WDC
B factor (Å2)
0
20
40
60
80
100
120
140
0
50
100
150
200
250
300
350
400
450
500
550
600
Residue Number


SoxY
Chain B
Chain E
Chain H
Chain K
B factor (Å2)
0
20
40
60
80
100
120
140
160
180
200
0
20
40
60
80
100
120
140
160
Residue Number

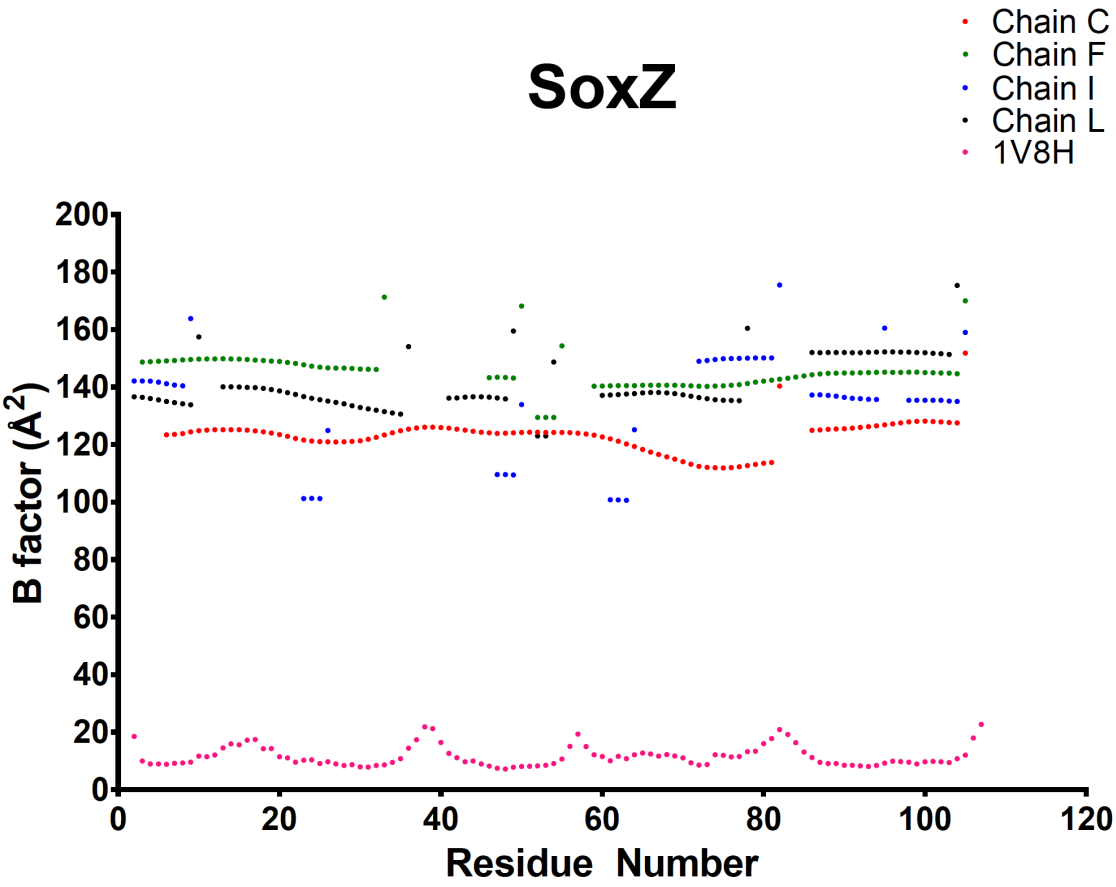


Figure S4. Residue-by-residue B-factors for each chain in the unit cell. Also shown are the B factors for the previously-determined high resolution structures of the isolated SoxB (PDB 2WDC; 1.5Å) and SoxZ (PDB 1V8H; 1.2Å) proteins.

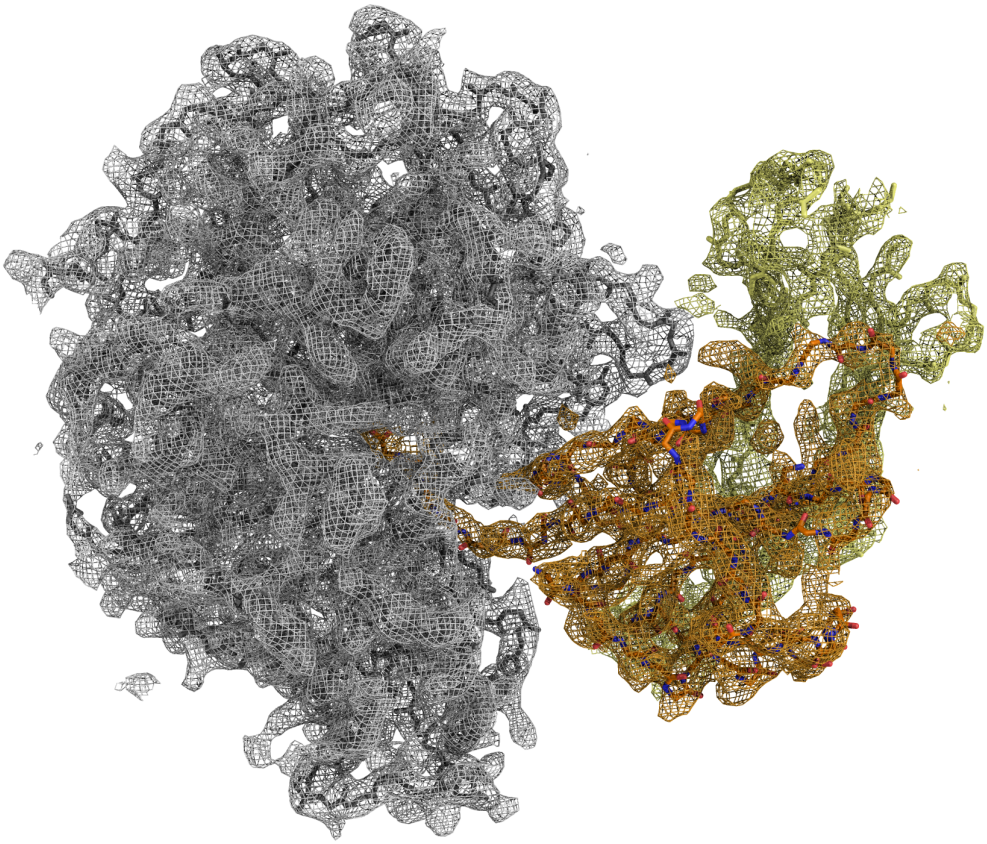


Figure S5. Quality of the electron density of the best-ordered SoxB-SoxYZ ternary complex in the unit cell. 2Fo-Fc αc electron density for chains SoxB (grey, chain A), SoxY (orange, chain B), SoxZ (yellow, chain C), contoured at 1.0 sigma.

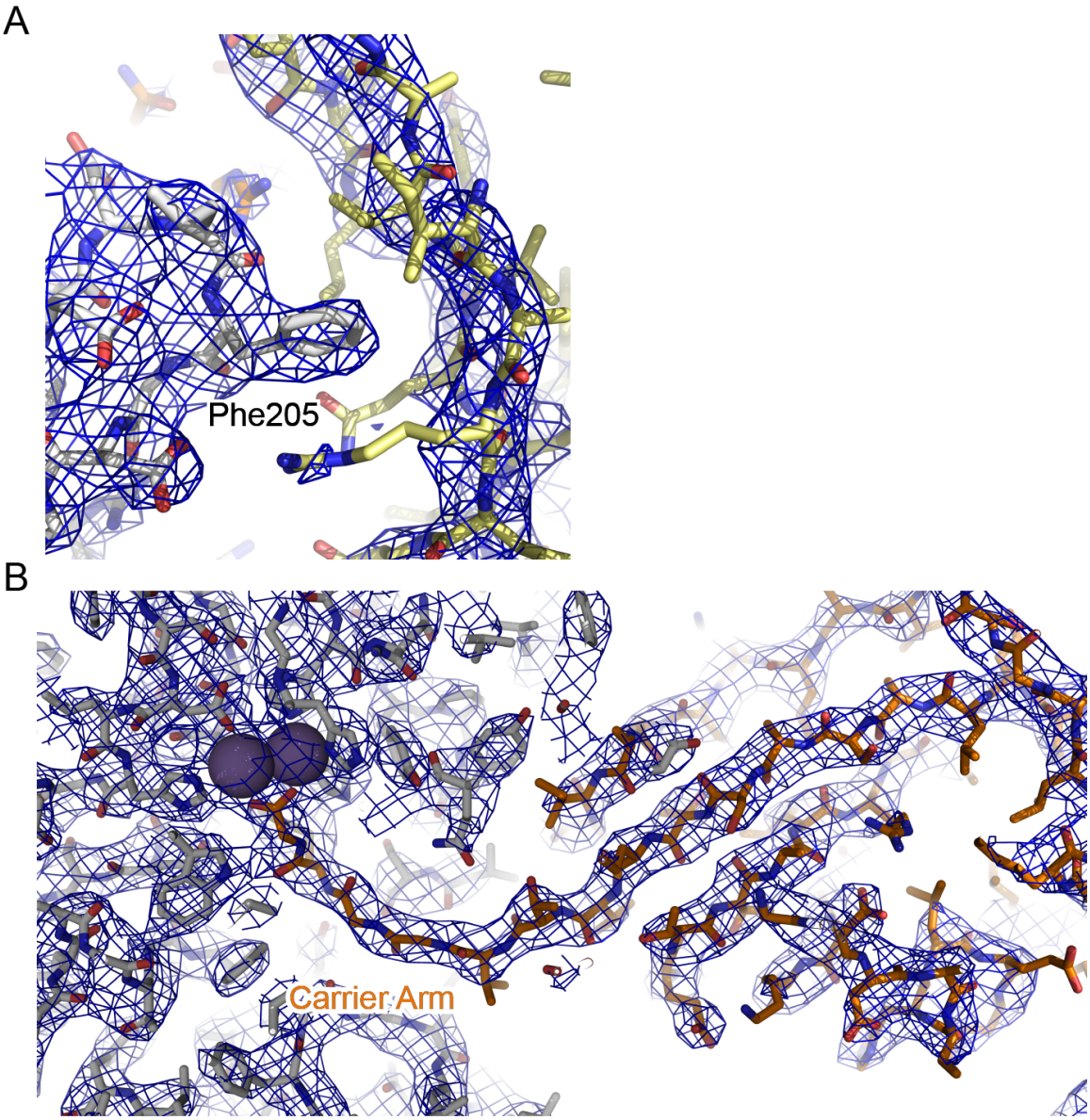


Figure S6. Quality of the electron density for the Z-patch interface and SoxY carrier arm shown in Figs. 6A and 7C. 2Fo-Fc αc electron density, contoured a 1σ, for the most ordered copy of the SoxBYZ complex (chains A, B, and C) around (A) the Z-patch interface and (B) the SoxY carrier arm.

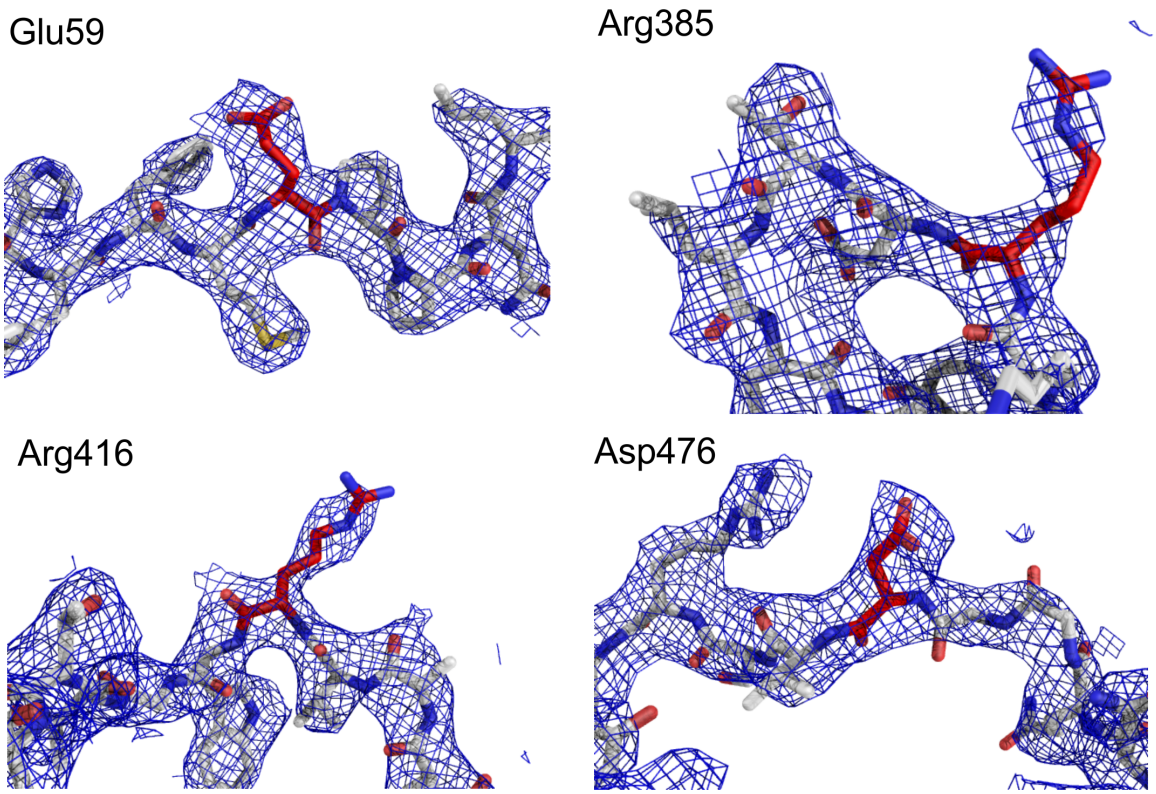


Figure S7. Quality of the electron density covering Arg416 and other important residues of the Sox B active site shown in Fig. 8B. 2Fo-Fc αc electron density is shown for (chain A) and contoured at 1.0σ. The carbon atoms in the residue of interest in each panel are colored red.

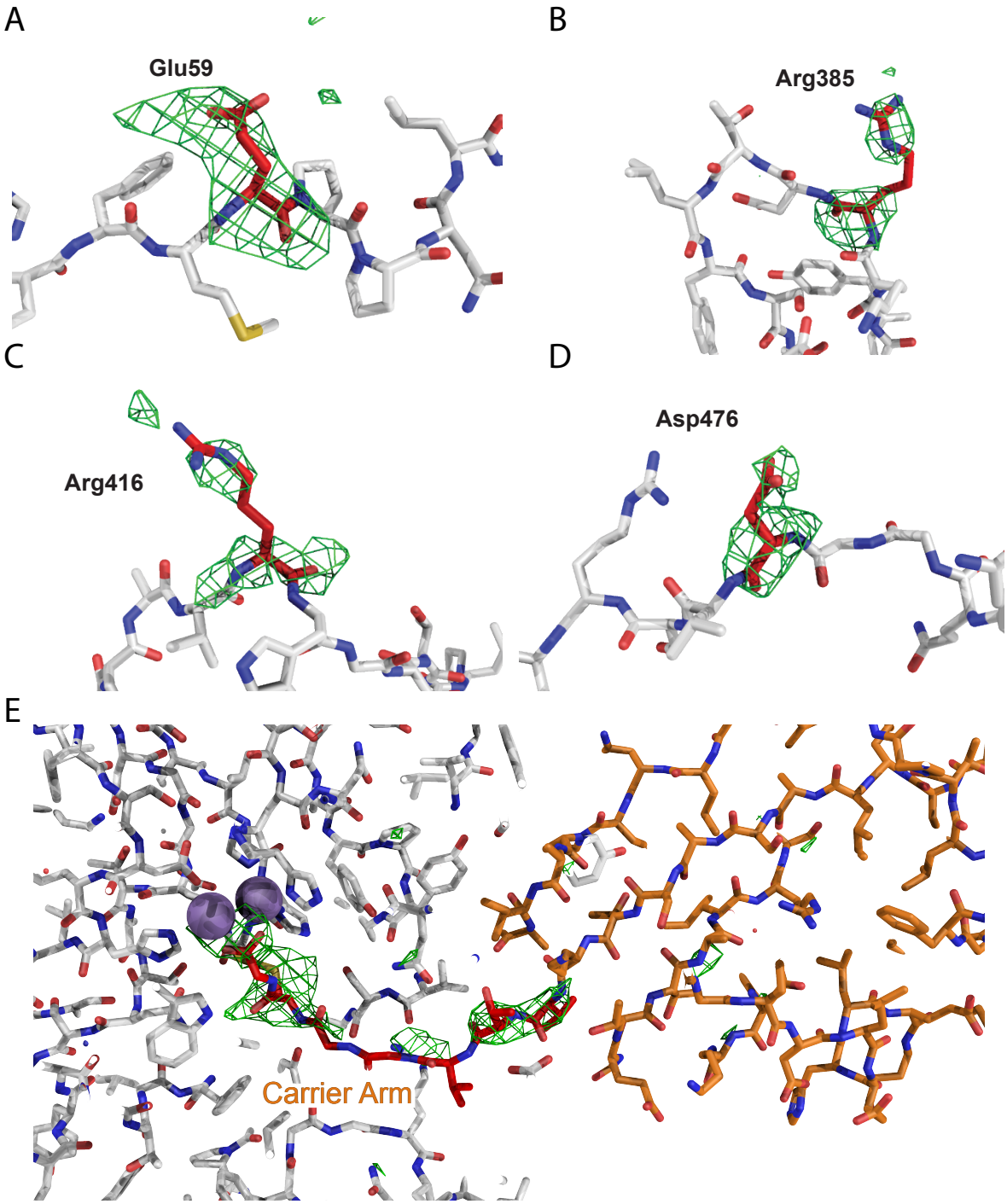


Figure S8. Simulated annealing omit maps for key structural features in the complex. Omit maps were calculated in Phenix (11) for the most ordered copy of the SoxBYZ complex. Residues colored in red were omitted in each panel and the density is shown contoured at 3σ (green). (A-D) SoxB active site residues. (E) SoxY carrier arm.

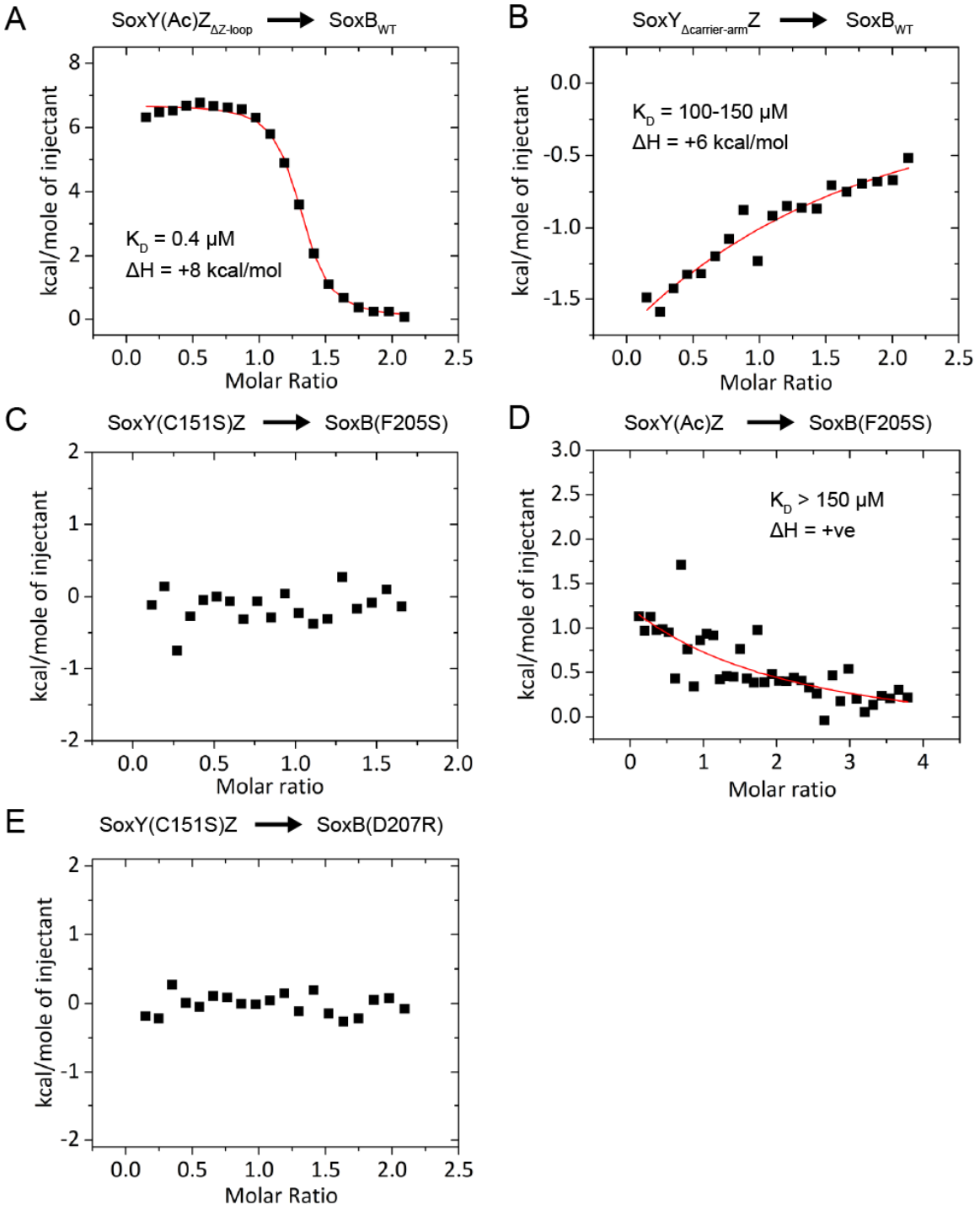


Figure S9. ITC analysis of the role of specific amino acids in SoxB-SoxYZ complex formation. In each case integrated heats from a representative experiment are shown. Where the data is fitted, the fit and corresponding K_D and enthalpy change values are calculated from duplicate experiments. The label above each panel shows which SoxYZ variant or derivative has been titrated into which SoxB variant. WT is the wild-type SoxB protein.

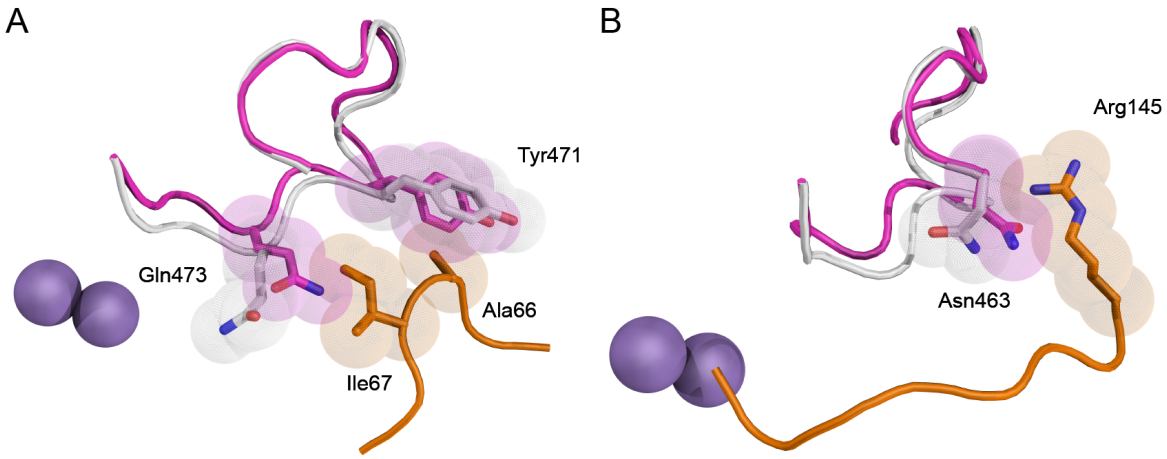


Figure S10. Alterations in SoxB side chain orientations that alleviate steric clashes with SoxY. Comparative positions of SoxB residues in the presence (gray; SoxB-SoxYZ complex structure) or absence (magenta; unliganded SoxB structure) of SoxY amino acids (orange). Calculated surfaces are shown in surface representation.

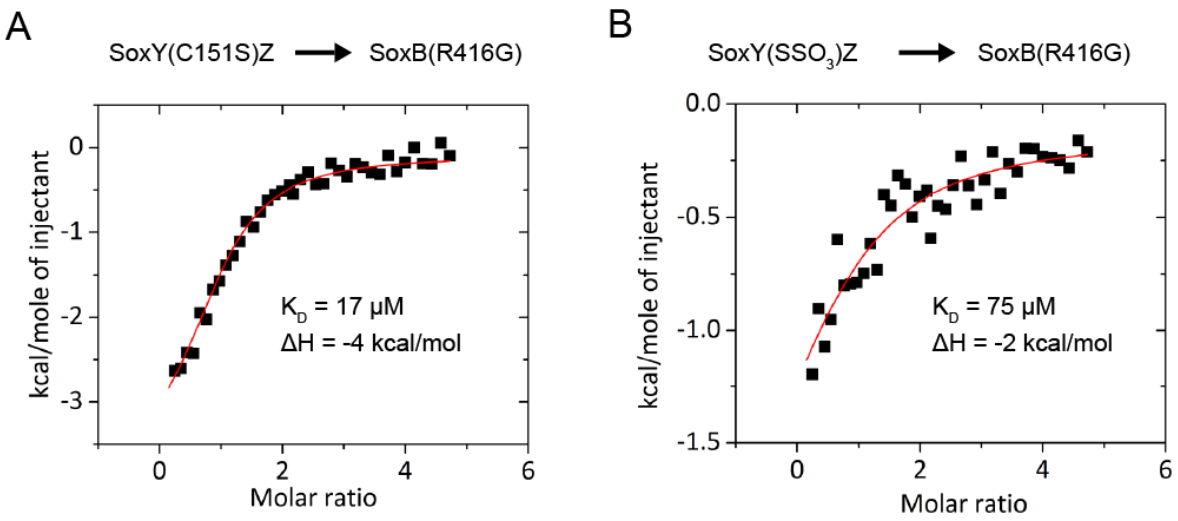


Figure S11. ITC analysis of the role of individual amino acids in substrate discrimination at the SoxB active site. In each case integrated heats from a representative experiment are shown. The fit in B, together with the corresponding K_D and enthalpy change values, are calculated from duplicate experiments. The label above the panels shows the SoxYZ variant or derivative that has been titrated into the SoxB R416G variant.

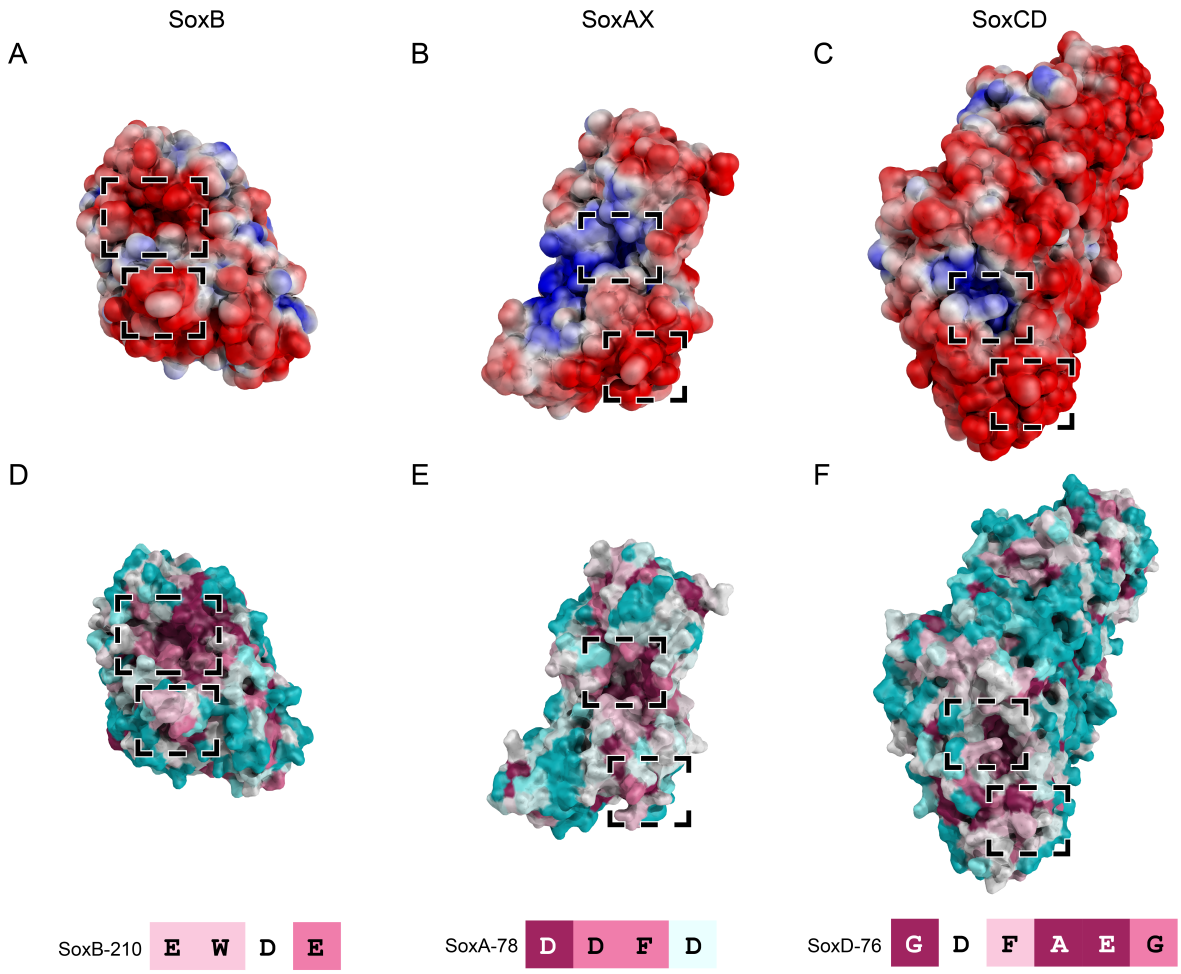
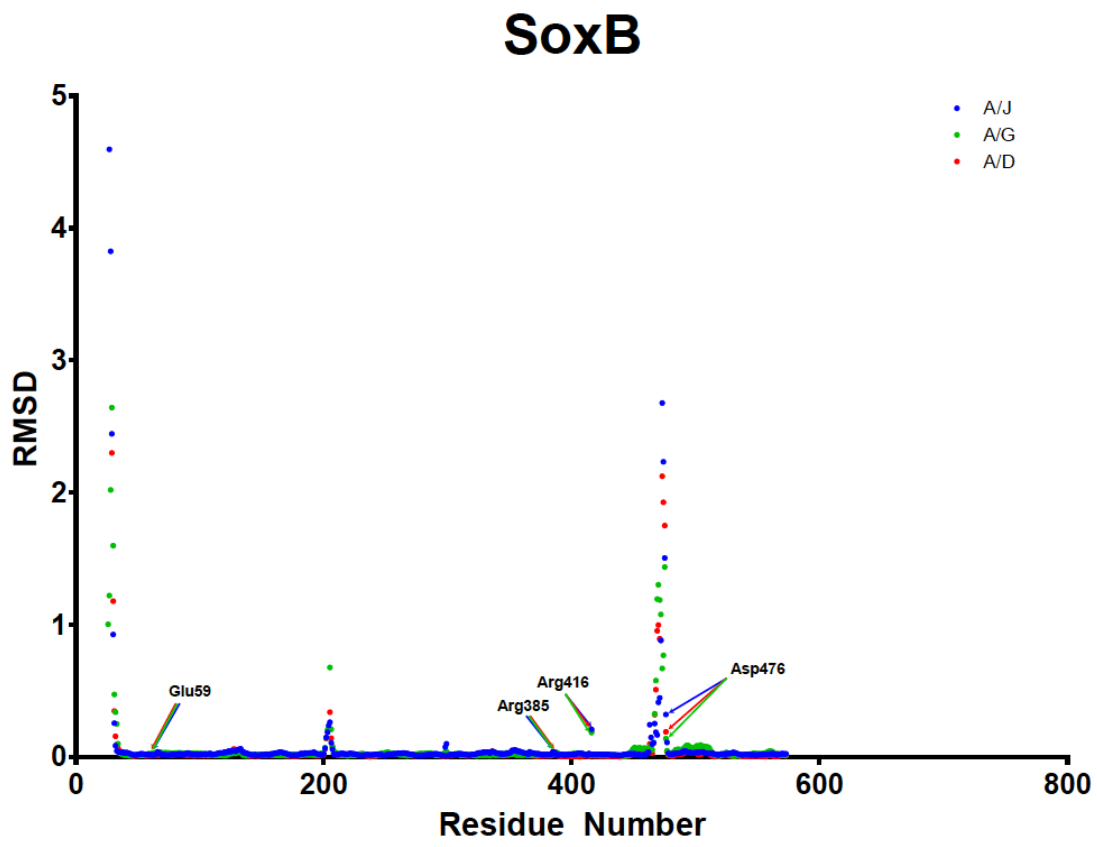
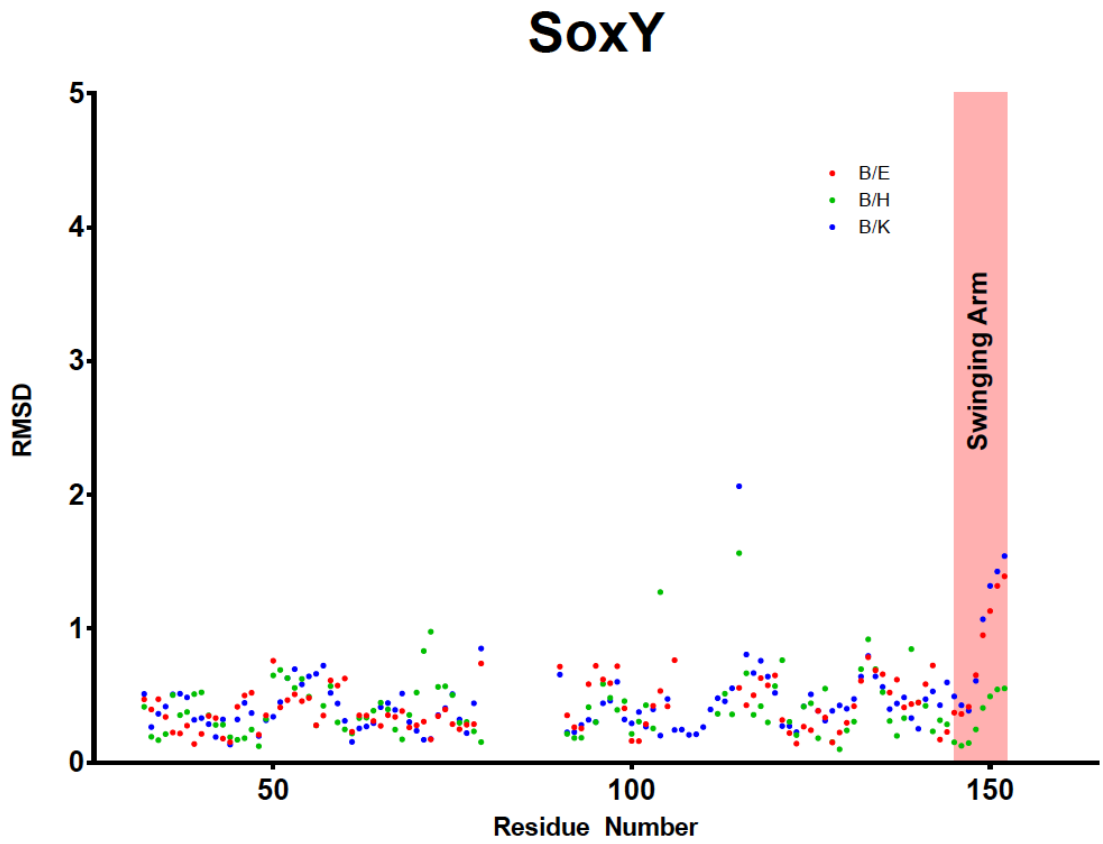


Figure S12. Comparative analysis of the SoxYZ-interacting surfaces of SoxYZ partner enzymes from the same species. SoxYZ-interacting enzymes of *Paracoccus pantotrophus* are shown in surface representation viewed from the face containing the active site entrance tunnel. The upper panels (***A-C***) show the electrostatic surface potential calculated using Adaptive Poisson-Boltzmann Solver (12) with positive potentials coloured blue and negative potentials red. The lower panels (***D-F***) show the same views of each protein but with the surface coloured according to sequence conservation using the program Consurf (13). Magenta indicates areas of highest sequence conservation and cyan the most variable sequences. In each panel the active site entrance is indicated by the upper, larger, box. The known (SoxB) or proposed (SoxAX and SoxCD) SoxZ contact site is indicated by the lower, smaller, box and the sequence of the loop from this contact is shown below at the foot of the figure with residues in Consurf colours. The structures shown are a homology model of SoxB produced using SWISS-MODEL (14) based on

the structure of *T. thermophilus* SoxB (PDB:2WDF)(1), *P. pantotrophus* SoxAX (PDB:2C1D)(15), and *P. pantotrophus* SoxCD (PDB:2XTS)(16).


SoxB
A/J
A/G
A/D
RMSD
Residue Number
Glu59
Arg385
Arg416
Asp476
0
1
2
3
4
5
0
200
400
600
800


SoxY
B/E
B/H
B/K
Swinging Arm
RMSD
Residue Number
0
1
2
3
4
5
50
100
150

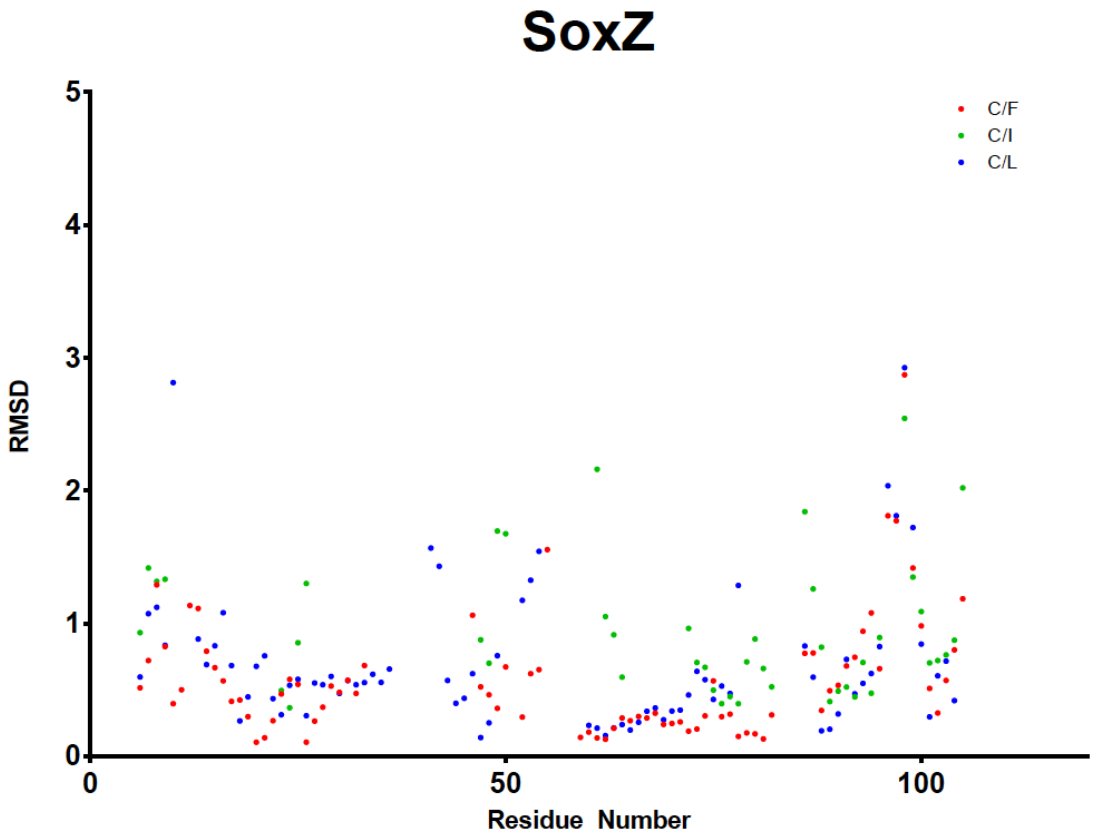


Figure S13. Residue-by-residue RMSD of the main chain atoms for each protein. Comparisons are between the most ordered chain and each of the three other chains in the unit cell. Structural features of interest are indicated.

A

gsSoxB

GAATTCATTAAAGAGGAGAAATTAACTATGGCCTCGTGGTCGCATCCGCAGTTTGAAAAAGGTGCCCTGG
AAGACCCGCGCTCGCTGTATGATCTGCCGCCGTATGGTGACGCAACCCTGCTGTACTTTTCCGATCTGCA
TGGCCAGGCTTTTCCGCACTATTTCATGGAACCGCCGAACCTGATCGCACCGAAACCGCTGATGGGTCGT
CCGGGTTATCTGACCGGTGAAGCGATTCTGCGCTATTACGGCGTGGAACGTGGTACGCCGCTGGCCTATC
TGCTGTCTTACGTGGATTTTGTTGAACTGGCACGTACCTTCGGTCCGATCGGCGGTATGGGTGCACTGAC
GGCACTGATTCGCGACCAGAAAGCTCGTGTCGAAGCGGAAGGCGGTAAAGCACTGGTGCTGGATGGCGGT
GACACCTGGACGAACAGCGGCCTGTCTCTGCTGACCCGCGGTGAAGCGGTGGTTCGTTGGCAAAATCTGG
TCGGCGTGGATCATATGGTGTCTCACTGGGAATGGACGCTGGGTCGTGAACGCGTTGAAGAACTGCTGGG
CCTGTTTCGCGGTGAATTTCTGAGTTATAATATCGTGGATGACCTGTTTGGCGATCCGCTGTTCCCGGCT
TACCGCATTCATCGTGTTGGTCCGTATGCTCTGGCAGTCGTGGGTGCAAGTTATCCGTACGTTAAAGTCA
GTCACCCGGAATCCTTTACCGAAGGTCTGTCCTTCGCCCTGGATGAACGTCGCCTGCAGGAAGCAGTGGA
CAAAGCCCGCGCAGAAGGCGCTAACGCGGTTGTCCTGCTGTCACATAATGGTATGCAACTGGATGCGGCC
CTGGCGGAACGTATTCGCGGCATCGATCTGATTCTGTCGGGTCATACCCACGACCTGACGCCGCGTCCGT
GGCGTGTGGGTAAAACCTGGATCGTTGCCGGCAGCGCAGCTGGTAAAGCACTGATGCGTGTGGATCTGAA
ACTGTGGAAAGGCGGTATTGCTAACCTGCGTGTGCGCGTTCTGCCGGTTCTGGCGGAACACCTGCCGAAA
GCTGAAGATGTCGAAGCGTTTCTGAAAGCTCAGCTGGCGCCGCATCAAGACCACCTGTTCACCCCGCTGG
CCGTTAGCGAAACGCTGCTGTATAAACGCGATACCCTGTACTCTACGTGGGACCAGCTGGTTGGTGAAGC
CGTCAAAGCAATCTACCCGGAAGTCGAAGTGGTTTTTAGCCCGGCAGTGCGTTGGGGCACCACGATCCTG
CCGGGTCAGGCTATTACCTGGGATCATCTGTATGCGTACACCGGCTTTACGTATCCGGAACTGTACCTGT
TTTATCTGCGCGGTGCCCAAATCAAAGCAGTTCTGGAAGACATTGCCTCAAACGTCTTTACCTCGGATCC
GTTCTACCAGCAAGGCGGTGACGTCAGTCGCGTGTTTGGCCTGCGTTATGTGCTGGATCCGGACGCACCG
ACGGGTGAACGTGTTCGCGAAGTCGAAGTGGGCGGTCGTCCGCTGGATCCGAATCGTCGCTATCTGGCAG
CAGCATACGGTGGTCGTCTGCAGCGTGTGGGTGAAGCCAAACCGGGTTATGAACCGCGCCCGATTTACGA
AGTGCTGGCAGAATATCTGCGTAGCGTTGGCCGTGTTCGCGTCCGTCCGGAACCGAATGTGAAAGTGATT
GGTCGCAACTATCGCCTGCCGGAAGTGACGGGCTAA**GGTACC**

B

gsSoxYZ

GGATCCCAAGGCCTGGAAGGCGAAGACCTGGAACATCTGGAACAAGCACTGAAAGAAGTTTTTGGTAAAG
GTTTTAAAGACCTGACCCCGTCGGATGCGGTGAAACTGAACATGCCGGCGATTGCCGAAAGCGGCGCGAA
TGTTCCGGCCGAAGTCGAAGTGGCCCTGCCGAAAGAACAGGTCAAAGCAATTCACCTGTTTGCTGACAAA
AACCCGACCCCGCACATCCTGGCATTCATGCCGATGAAAGCGGAACCGTATTACGCCACCCGTGTTCGCC
TGGCTGAAACCACGGCAATCCGTGCTGTGGTTGAAACGCAAGATGGCAAACTGCTGCTGGCGTCTGCTTC
AACCCGTGTGACCGTGGGCGGCTGCGGCTGAGGTACCATTAAAGAGGAGAAATTAACTATGCCGTTCCGC
ACCATTGCCCGTCTGAATCCGGCCAAACCGAAAGCAGGCGAAGAATTCCGTCTGCAAGTCGTCGCACAAC
ATCCGAACGAACCGGGCACCCGTCGCGATGCAGAAGGTAAACTGATTCCGGCCAAATACATCAACCTGGT
GGAAGTTTACTTCGAAGGCGAAAAAGTCGCAGAAGCACGTCCGGGTCCGAGCACGTCTGCAAACCCGCTG
TATGCCTTTAAATTCAAAGCAGAAAAAGCTGGCACCTTCACGATTAAACTGAAAGATACGGACGGTGATA
CGGGCGAAGCAAGTGTGAAACTGGAACTGGCGTGA**AAGCTT**

Figure S14. Sequences of codon-optimised genes used in this study. Restriction sites were added to beginning and end of each sequence as described in the text. These sites are marked in bold.

References

1. Sauvé V, *et al.* (2009) Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB. *Journal of Biological Chemistry* 284(32):21707-21718.
2. Gasteiger E, *et al.* (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook,* No Series), pp 571-607.
3. Houtman JCD, *et al.* (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. *Protein Science* 16(1):30-42.
4. Hornak V, *et al.* (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* 65(3):712-725.
5. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2):926-935.
6. Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4(3):435-447.
7. M. Bradbrook G, *et al.* (1998) X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *Journal of the Chemical Society, Faraday Transactions* 94(11):1603-1611.
8. Case DA, *et al.* (2012) AMBER 12.
9. Frisch MJ, *et al.* (2004) Guassian 03, Revision C.02, Guassian Inc., Wallingford CT).
10. Wang J, Wolf RM, Caldwell JW, Kollman PA, & Case DA (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry* 25(9):1157-1174.
11. Adams, PV, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica* D66:213-221.
12. Baker NA, Sept D, Joseph S, Holst MJ, & McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98(18):10037-10041.
13. Celniker G, *et al.* (2013) ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry* 53(3-4):199-206.
14. Biasini M, *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(Web Server issue):W252-258.
15. Dambe T, Quentmeier A, Rother D, Friedrich C, & Scheidig AJ (2005) Structure of the cytochrome complex SoxXA of *Paracoccus pantotrophus*, a heme enzyme initiating chemotrophic sulfur oxidation. *Journal of structural biology* 152(3):229-234.
16. Zander U, *et al.* (2011) Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. *The Journal of biological chemistry* 286(10):8349-8360.

Supplementary Table 1

X-ray data collection statistics	
X-ray source	Diamond - i04 (0.99990 Å)
Space group	P1
Cell dimensions	*a* = 70.2 Å, *b* = 116.0 Å, *c* = 120.9 Å
	α = 86.51°, β = 83.21°, γ = 89.77°
Resolution (Å)	59.81-3.28 (3.37-3.28)
Total reflections	112492
Number of unique reflections	57115
Completeness (%)	98.7 (98.4)
Multiplicity	2.0
R_{merge}	0.088 (0.302)
I / δ (I)	7.9 (1.9)
Refinement statistics	
R (%)	26.96 (27.76)
R_{free} (%)	27.47 (30.04)
Root mean square deviation from idealized covalent geometry	
Bond length (Å)	0.007
Bond angles (°)	0.84
Average whole structure B value (Å^2)	92.8
Average single chain B value (Å^2)	SoxB: A=64, D=93, G=84, J=74
	SoxY: B=110, E=138, H=138, K=134
	SoxZ: C=128, F=150, I=141, L=146
Ramachandran outliers (%)	0.24
Residues modelled	2952
Non-protein molecules	11 waters, 8 Mn^{2+} (100% occupancy)
Molprobity score	1.48 (100th centile for structures at this resolution)