Supporting information

X-ray Crystallographic and EPR Spectroscopic Analysis of HydG, a Maturase in [FeFe]-Hydrogenase H-Cluster Assembly

Pedro Dinis[a], Daniel L. M. Suess[b], Stephen J. Fox[a], Jenny E. Harmer[a], Rebecca C. Driesener[a], Liliana De La Paz[c], James R. Swartz[c,d], Jonathan W. Essex[a], R. David Britt[b], and Peter L. Roach[a,*].

[a]Chemistry and the Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK, [b]Department of Chemistry, University of California, Davis, Davis, CA 95616, USA, [c]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA, [d]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA.

METHODS

Overexpression, purification, and crystallization of *Thermoanaerobacter italicus* HydG (*Ti*HydG) with reconstituted iron sulfur clusters. A commercially supplied codon-optimized synthetic *Thermoanaerobacter italicus hydG* encoded a hexahistidine tag at the N-terminus. The *hydG* was ligated into the plasmid pFM024, a derivative of pMK024 (1), to yield pRD003. This plasmid permits the co-expression of *Ti*HydG and the *E. coli isc* operon under control of the pBAD promoter. Chemically competent *E. coli* BL21 (DE3) were transformed with pRD003 and used for expression of *Ti*HydG. Overnight cultures (100 mL) were used as a 1% inoculum for larger scale cell culture (5 L) in 2TY media using a fermenter (Brunswick Bioflo 110) system at a constant temperature (27 ºC) and O_2 levels (40 %). When the OD_{600} of the culture reached 0.6, expression was induced by 1 % L-arabinose. Cell pellets were harvested by centrifugation.

All purification steps were performed in an anaerobic environment, using a glove box (Belle Technology MR1). The cell paste was re-suspended in Buffer A (3 mL/g, 25 mM Hepes, pH 7.4, 500 mM NaCl, 20 mM imidazole) and lysozyme added (10 mg/100 mL) and EDTA-free protease inhibitor tablets (1 tablet/50 mL) and stirred for 1 hour. The cells were lysed by sonication (10 minutes, 1 second pulse, 20W) and the cell debris removed by centrifugation. The cleared supernatant was applied to a Ni-sepharose column pre-equilibrated with Buffer A. After washing, the proteins were eluted with a gradient of Buffer B (25 mM Hepes, pH 7.4, 500 mM NaCl, 250 mM imidazole). Red-brown fractions, containing *Ti*HydG, were immediately desalted into Buffer C (25 mM Hepes, pH 7.4, 500 mM NaCl, 5 mM DTT) using a Superdex 75 column (XK26/10, 50 mL). The clusters were reconstituted by the dropwise addition of DTT (5 mM) and then $FeCl_3$ and Na_2S solutions (10 equivalents). Precipitated iron sulfide was removed by centrifugation and the supernatant concentrated to ~3 mL (>30 mg/mL). The protein was then applied to a Superdex 75 column (XK26/60, 300 mL) and fractions collected. The purity of HydG was determined by SDS-PAGE. The purest fractions were pooled, reconstituted (as above, but with 5 equivalents) and concentrated (to 1 mM). The degree of reconstitution was assessed from the UV-visible spectrum of protein ($OD_{410/280} \approx 0.5$). Crystallization conditions were identified by a sitting drop screen (1 μL protein plus 1 μL precipitant solution) in plates stored at 20 ºC in the presence of SAM (10 mM). After 5 days, crystals of *Ti*HydG grew in the presence of 0.1M Bis-Tris Propane, pH 6.5, 0.2 M sodium fluoride and 20% (w/v) PEG 3350.

Overexpression and purification of Wild Type (WT) *Shewanella oneidensis* HydG (*So*HydG) and the mutant *So*HydG^{XN}. The *So* HydG mutant lacking the N-terminal cluster, "*So*HydG^{XN}", was prepared by mutating the three conserved cysteines of the N-terminal Fe-S cluster, CX_3CX_2C, to serines, similar to Driesener *et al*. (2). Using WT *So*HydG as a template, the mutants C103/107/110S were prepared using QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies, La

Jolla, CA) on a pET-21b derived vector (Novagen) and transformed into *E. coli* BL21 (DE3) ΔiscR chemically competent cells, and the sequence confirmed (ElimBio, Hayward, CA). Anaerobic expression and purification of *So*HydG^{XN} in *E. coli* strain BL21(DE3) *ΔiscR::kan* was performed using methods previously described for high-yield production of metalloproteins with Fe−S clusters (3).

The wild type, codon optimized *So*HydG was also heterologously expressed in *E. coli* BL21(DE3) *ΔiscR::kan*, as previously described (3). Briefly, media was supplemented with 2 mM ferric ammonium citrate, 2 mM cysteine, and induced with 200 µM IPTG (Life Technologies, Carlsbad, CA) for 10 hours in the anaerobic glove box. Cells were harvested by centrifugation at 6000 x *g* for 30 min, lysed with Bug Buster Master Mix (2 mL lysis buffer/ gram cell pellet). The lysate was clarified by centrifugation at 15000 x *g,* purified on a Streptactin column (IBA Life Sciences, Goettingen, Germany), and eluted with desthiobiotin (3 mM). The purified product was concentrated in an Amicon stirred cell with a pressure based sample concentration (30 kDa MWCO, EMD Millipore, Darmstadt, Germany), formulated with 5% (w/v) trehalose, flash frozen, and stored at –80°C.

Structure Determination and Refinement of *Ti*HydG. The crystals were flash frozen directly in liquid nitrogen. Native and Fe-SAD data sets were collected at 100 K on beamline I03 at the Diamond Light Source. The iron absorption maximum (7162 eV) was identified using an X-ray fluorescence scan and used to collect the anomalous data set. All data was automatically processed with Xia2 (4). The structure was solved, built and refined using the Phenix package (5). The Fe-SAD data set was used to determine initial experimental phases using Phenix.AutoSol, followed by automatic building with Phenix.AutoBuild, manual (re-)building using the program WinCoot (6) and refinement with Phenix.Refine. Two Ramachandran outliers were observed, Gly194 and Gly326 in each monomer.

Computational Methods. Hydrogen atoms were added to the crystal structure and possible tyrosine binding sites around the [4Fe-4S] and [5Fe-5S] clusters were located using moe2013 (7). A pocket capable of encapsulating tyrosine was located near to SAM, pointing towards the auxiliary cluster. Valid models of tyrosine bound with either the phenolic or α-amino hydrogen atoms adjacent to the 5’-carbon of SAM could be achieved. The model presented in the main text (Fig. 2E) placed an α-amino hydrogen atom close to the 5’-carbon of SAM by analogy to a recent mechanistic proposal for NosL (8). Tyrosine was manually built into the pocket, and along with neighboring residue side chains, minimized using the MMFF94 (9) to a gradient of 0.1 kcal/mol/Å^2. Crystallographic waters that overlapped with the tyrosine were removed and the remaining waters relaxed to the same gradient to achieve a hydrogen bonding network which further stabilized the tyrosine.

EPR Spectroscopy Methods. All EPR spectra were collected at the CalEPR Center in the Department of Chemistry at the University of California, Davis. X-band measurements were performed with a Bruker Biospin EleXsys E500 spectrometer equipped with a cylindrical TE011-

mode resonator (SHQE-W), an ESR-900 liquid helium cryostat, and an ITC-5 temperature controller (Oxford Instruments). Hyperfine sublevel correlation (HYSCORE) spectra were measured using an Elexsys E580 spectrometer (Bruker) with a split-ring (MS5) resonator using the pulse sequence π/2-τ-π/2-t1-π-t2-π/2-τ-echo wherein both the inversion pulse length (tπ) and the π/2 pulse (tπ/2) are identical (16 ns). Four-step phase cycling was employed. Time-domain spectra were baseline-corrected (third-order polynomial), apodized with a Hamming window, zero-filled to eight-fold points, and fast Fourier-transformed to yield the frequency-domain spectra. Spectral simulations were performed with Matlab using the EasySpin toolbox (10). Custom simulation scripts were written in MatLab^{TM} to make use of matrix diagonalization, perturbation theory, and optimization functions of the EasySpin routines.

References for the Supporting Information: Methods.

1. Bryant P, Kriek M, Wood RJ, & Roach PL (2006) The activity of a thermostable lipoyl synthase from Sulfolobus solfataricus with a synthetic octanoyl substrate. *Anal Biochem* 351(1):44-49.
2. Driesener RC*, et al.* (2013) Biochemical and Kinetic Characterization of Radical S-Adenosyl-L-methionine Enzyme HydG. *Biochemistry* 52(48):8696-8707.
3. Kuchenreuther JM*, et al.* (2010) High-Yield Expression of Heterologous [FeFe] Hydrogenases in Escherichia coli. *PLoS ONE* 5(11):e15491.
4. Winter G, Lobley CM, & Prince SM (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1260-1273.
5. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
6. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
7. Anonymous (2013) Molecular Operating Environment, 2013.08 (Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7).
8. Nicolet Y, Zeppieri L, Amara P, & Fontecilla-Camps JC (2014) Crystal Structure of Tryptophan Lyase (NosL): Evidence for Radical Formation at the Amino Group of Tryptophan. *Angew Chem Int Ed Engl* 53(44):11840-11844.
9. Halgren TA (1996) Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17(5-6):490-519.
10. Stoll S & Schweiger A (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Mag Res* 178(1):42-55.

Fig. S1. Alignment of Aromatic Amino Acid Lyase Enzymes. Residues are shaded by increasing conservation. The cysteine ligands for the radical SAM and auxiliary clusters are shaded red and orange respectively. The histidine ligand to the labile iron is shaded purple. Residues proposed to interact with the aromatic amino acid substrate are shaded green (8). Secondary structural features taken from *Ti*HydG are shown below the alignment and colored as follows: N-terminal extension (helices HA to HD), pink; radical SAM ¾ TIM barrel core (strand S1 to helix H6), green; C-terminal region (strand S7 to helix HJ), blue. Residues that interact with SAM include the 'CX3CXϕC' (cysteines shaded red) (1), the 'GGE' motif (red circles below the alignment), the 'GxIGxxE' motif (red triangles) and the conserved structural motif (red square) (2, 3). Sequences for each enzyme are taken from the following database entries (from top to bottom): *Thermoanaerobacter italicus* HydG, UniProt accession code D3T7F1; *Shewanella oneidensis* HydG, Q8EAH9; *Clostridium acetobutylicum* HydG, F7ZVC7; *Escherichia coli* ThiH, P30140; *Salmonella typhimurium* ThiH, Q9S498; *Streptomyces actuosus* NosL, C6FX5; *Nocardia sp.* ATCC 202099 NocL, E5DUI3; *Methanocaldococcus jannaschii* CofH, Q58826 and *Candidatus Nitrosoarchaeum limnia* BG20 CofH, S2E5F8. Sequences were aligned with Multalin (5) and the figure was prepared with Jalview (5).

References

1. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, & Miller NE (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res 29(5):1097-1106.

2. Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. Chem Rev 111(4):2487-2506.

3. Dowling DP, Vey JL, Croft AK, & Drennan CL (2012) Structural diversity in the AdoMet radical enzyme superfamily. Biochim Biophys Acta 1824(11):1178-1195.

4. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16(22):10881-10890.

5. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189-1191.

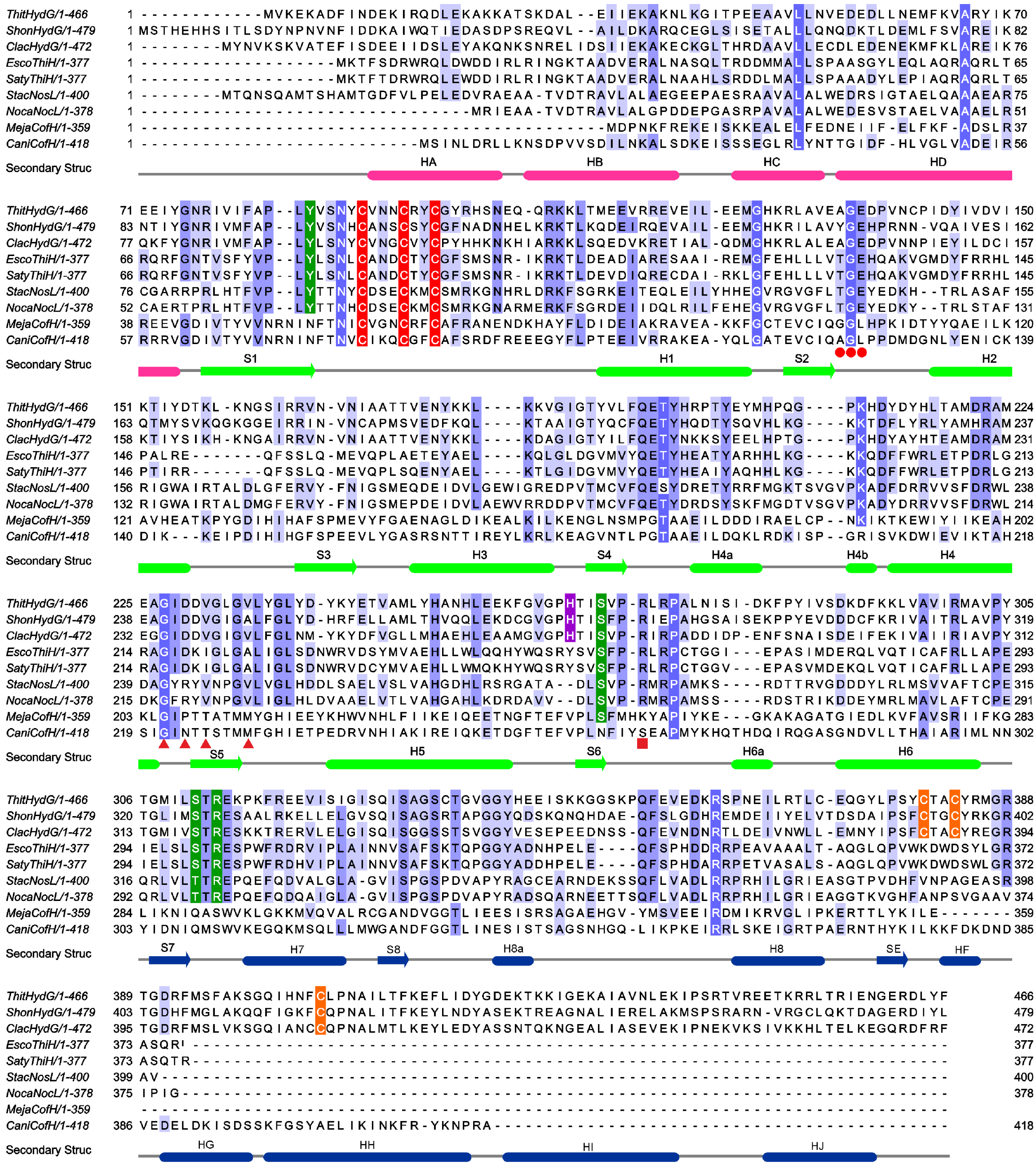

ThitHydG/1-466
ShonHydG/1-479
ClacHydG/1-472
EscoThiH/1-377
SatyThiH/1-377
StacNosL/1-400
NocaNocL/1-378
MejaCofH/1-359
CaniCofH/1-418
Secondary Struc
HA
HB
HC
HD
S1
H1
S2
H2
S3
H3
S4
H4a
H4b
H4
S5
H5
S6
H6a
H6
S7
H7
S8
H8a
H8
SE
HF
HG
HH
HI
HJ

Fig. S2. *Ti*HydG Fold and Clusters. The secondary structure is colored as follows: N terminal extension, pink; the radical SAM 3/4 TIM barrel core, green; C-terminal extension, blue. The cluster binding residues are highlighted in yellow (Cys) and green (His). The disordered loop is numbered in red.

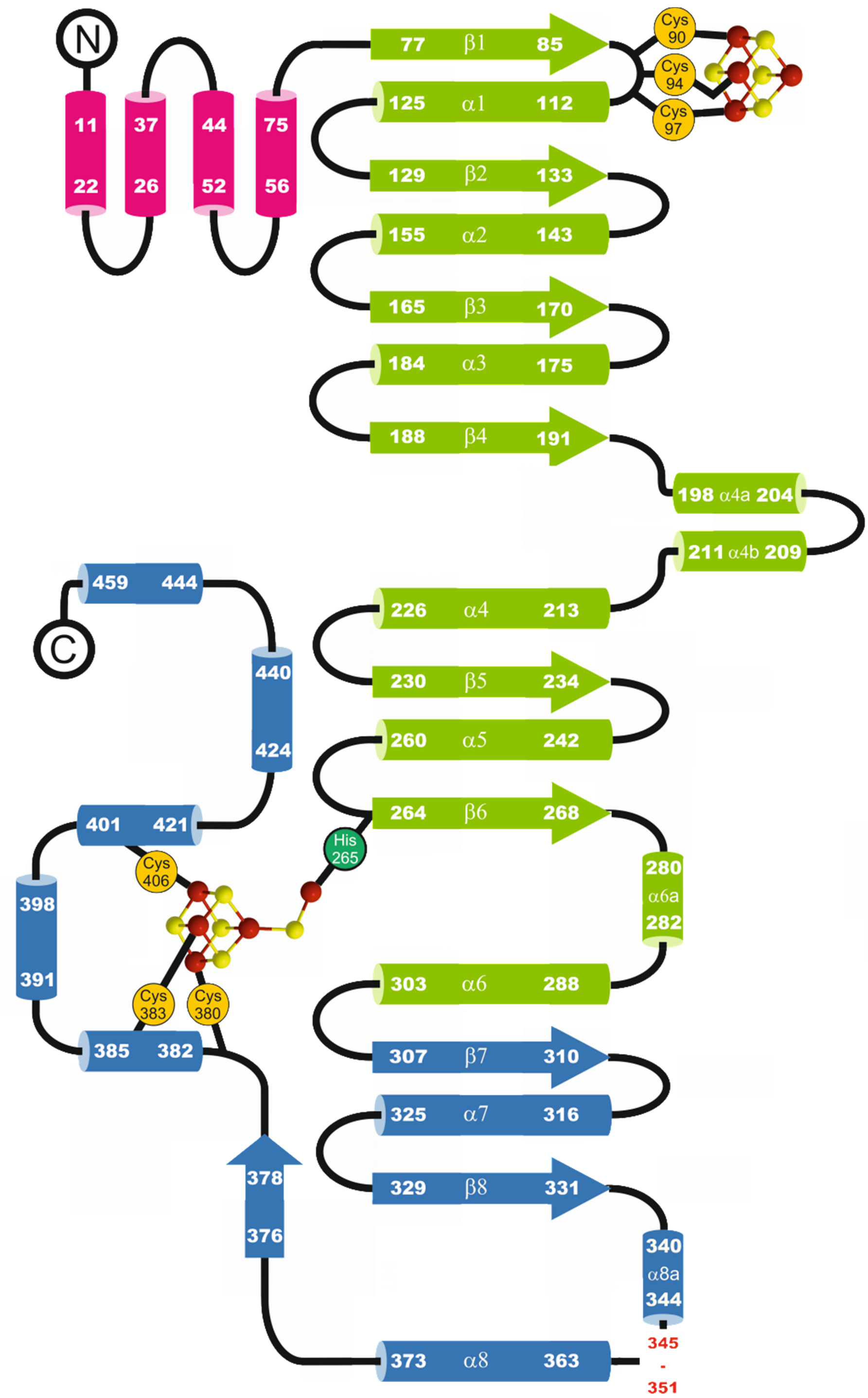


Fig. S3. Metal binding sites and ligands. The 2Fo-Fc map is contoured at 1.0 σ around the cluster and ligands. (*A*). The [4Fe-4S]$_{RS}$ site of monomer A with methionine bound. (*B*). The [4Fe-4S]$_{RS}$ site of monomer B with SAM bound (same orientation as monomer A). (*C*). The [5Fe-5S]$_{Aux}$ cluster from monomer A, showing the coordinating water molecules, amino acid ligand and His265. (*D*). The [4Fe-4S]$_{Aux}$ cluster from monomer B.

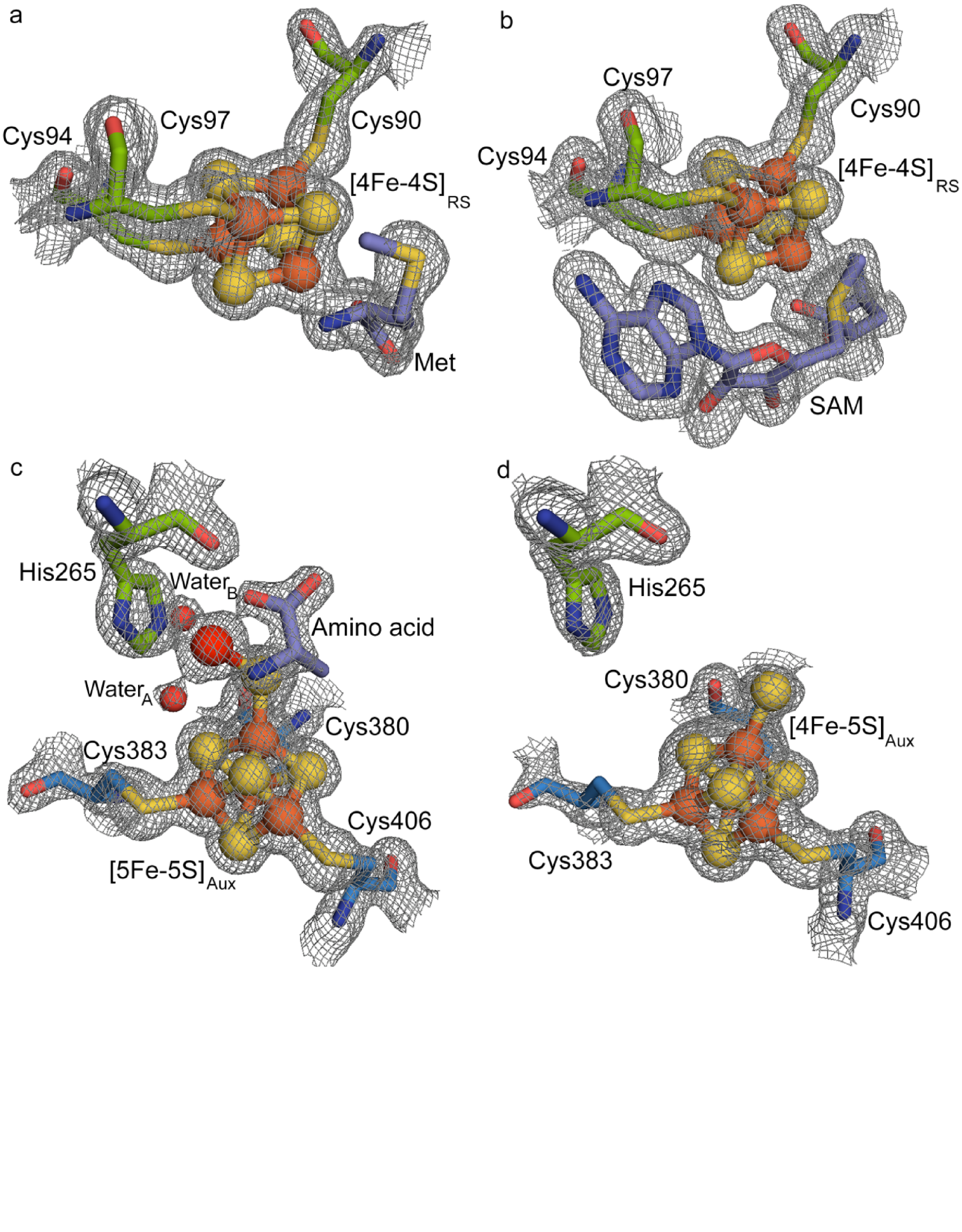
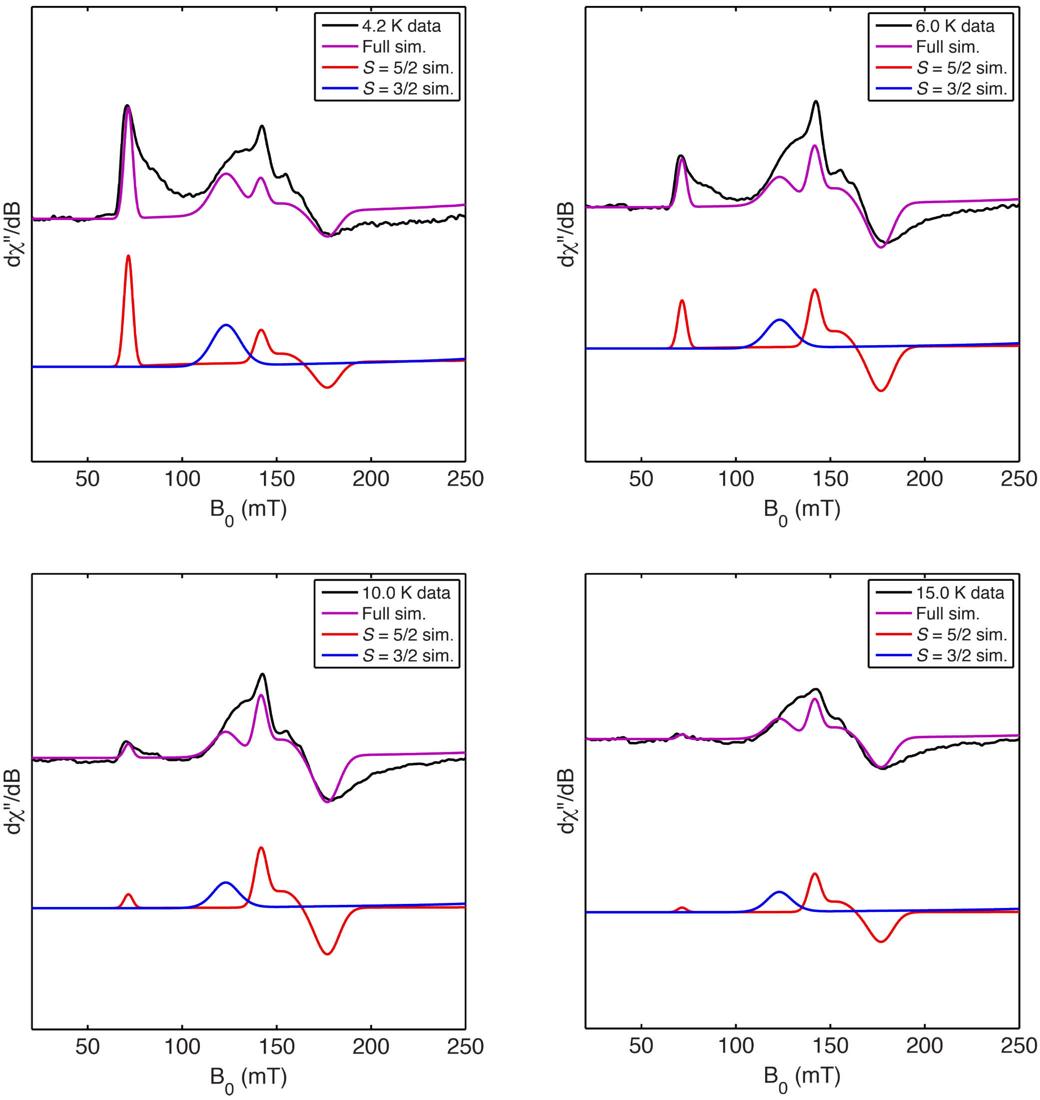


Fig. S4. Estimation of the zero-field splitting parameters of the $S = {}^5/_2$ form of the [5Fe–5S] cluster by simulation of variable-temperature EPR data. X-band CW EPR spectra of dithionite-reduced *So*HydG^{XN} recorded at 9.397 GHz with 5.00 mW power. The primary simulated component is shown in red with $S = {}^5/_2$, $D = +4.5$ cm^{-1}, $E/D = 0.255$, and $g = 2$ (σ = [0.25, 0.2, 0.1]). The positions of the features from this component allow for E/D to be determined. D was estimated by the relative intensity of the $m_S = \pm {}^1/_2$ feature at ~70 mT compared with the $m_S = \pm {}^3/_2$ features between 140 and 180 mT. The broadness of the features (seen very clearly between 70 and 100 mT and > 180 mT) is ascribed to a distribution in D values owing to heterogeneity in the local environment of the [5Fe–5S] cluster. Note that EPR spectroscopy tends to overemphasize such heterogeneity because the transition probabilities are greater for more axial systems (lower E/D values) (1). An $S = {}^3/_2$ component (blue traces) was also included with $D >> h\nu$, $E/D = 0.33$, and $g = 2$ (σ = [0.2, 0.2, 0.2]) to show how some of the extra intensity may also be ascribed to an $S = {}^3/_2$ component. The relative proportions of the $S = {}^5/_2$ and $S = {}^3/_2$ components in the simulations are not indicative of their actual relative proportions owing to their different relaxation properties and the broadness of the signals.



Reference

1. Kennedy MC, et al. (1984) Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. J Biol Chem 259(23):14463-14471.

Fig. S5. X-band CW EPR spectra of *So*HydG^{XN}. Top spectra: with 10 mM dithionite. Bottom spectra: with 3 mM KCN and 10 mM dithionite. Experimental parameters: 9.397 GHz, 10 K, and 5 mW (left spectra) or 0.162 mW (right spectra). A decrease in the intensity of the $S = {}^{5}/_{2}$ [5Fe–5S] signal and an increase in the intensity of $S = {}^{1}/_{2}$ [4Fe–4S] clusters signals is observed upon addition of cyanide.

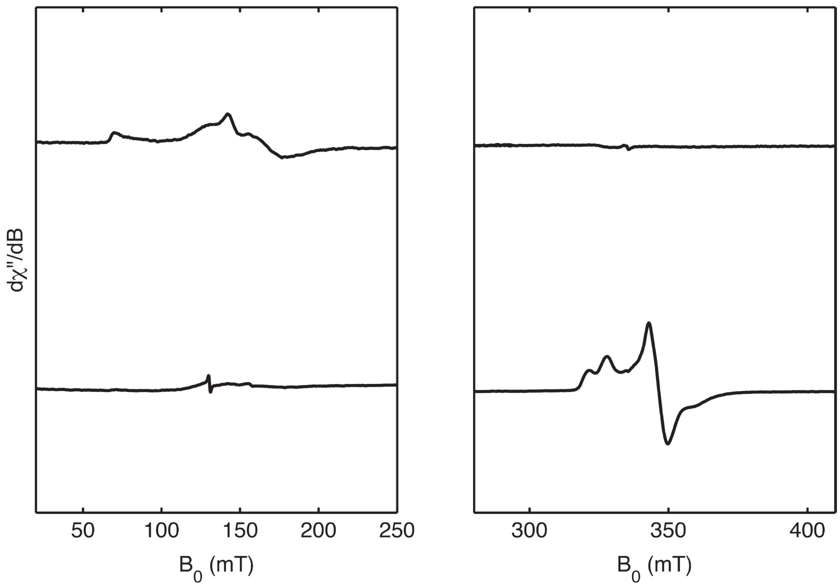


Fig. S6. Simulation of EPR spectrum of *So*HydG desalted three times in buffer containing 15 mM $K^{13}CN$, 10 mM dithionite, and 3 mM SAM as described in the main text. Top traces: experimental data (black) and simulation (red). Middle traces: simulation of auxiliary [4Fe–4S](^{13}CN) contribution (blue, g = [2.09, 1.94, 1.93] with σ = [0.03, 0.02, 0.04], 43% of total), simulation of $[4Fe–4S]_{RS}$(SAM) (purple, g = [2.01, 1.88, 1.85] with σ = [0.02, 0.03, 0.04], 28% of total), and simulation of the unidentified species (teal, g = [2.06, 1.95, 1.93] with σ = [0.03, 0.02, 0.04], 29% of total) that may correspond to cyanide bound to the $[4Fe–4S]_{RS}$ cluster. Bottom trace: residual of experimental data minus simulation (orange). Spectra were recorded at 9.397 GHz, 20 K, and 0.126 mW.

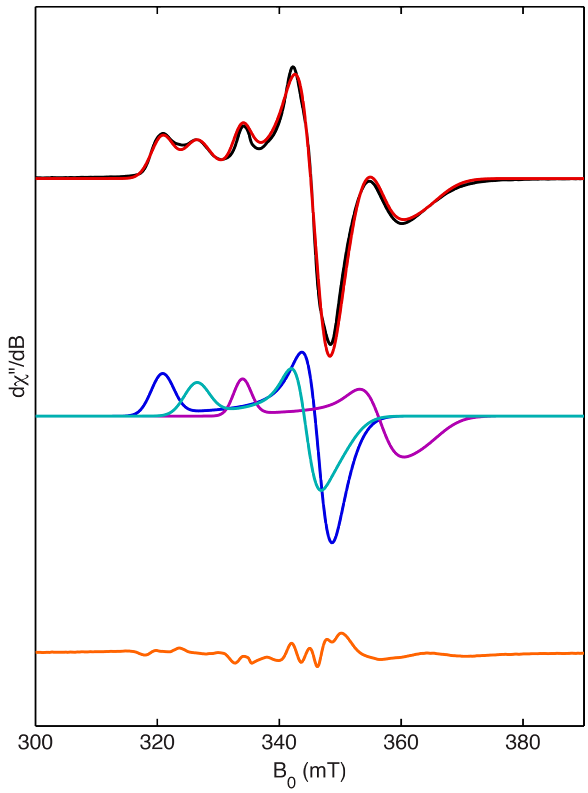


Figure S7. Orientation-selective X-band HYSCORE of *So*HydG desalted three times in buffer containing 15 mM $K^{13}CN$, 10 mM dithionite, and 3 mM SAM as described in the main text. Top row: experimental data. Bottom row: simulation of the auxiliary [4Fe–4S][^{13}CN] cluster overlaid onto experimental data (gray). Experimental conditions: 9.730 GHz, π/2 = π = 16 ns, 10 K. Simulation parameters: A = [–5.0, –4.0, 0.9] MHz, Euler angles of [–90˚, –40˚, 0˚], g = [2.09, 1.94, 1.93] (blue). The ^{13}C correlation ridges in HYSCORE spectra acquired at higher fields are obscured by ^{14}N correlation ridges from SAM bound to the [4Fe–4S]$_{RS}$ cluster. This effect does not occur in samples of *So*HydG^{XN} (see Fig. S8).

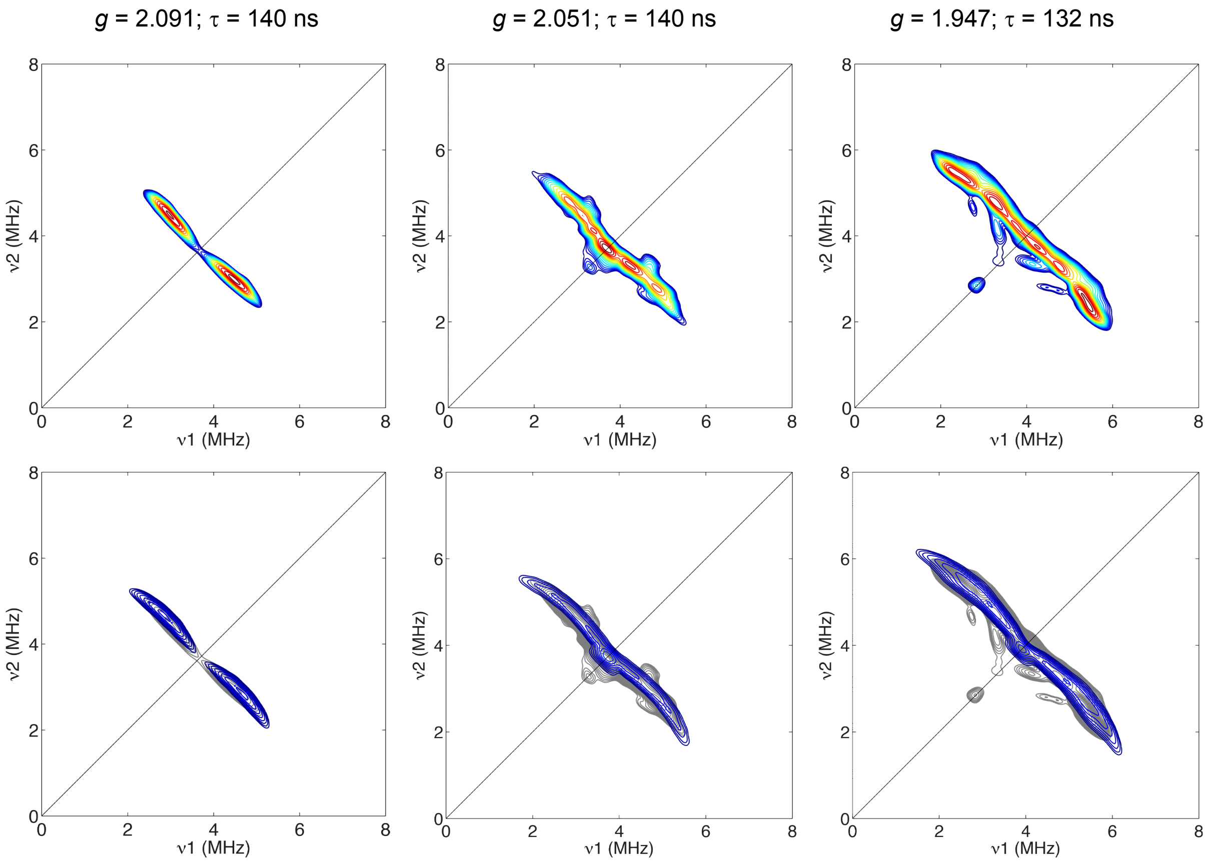


Fig. S8. Orientation-selective X-band HYSCORE of *So*HydG^{XN} in the presence of 20 mM K^{13}CN and 10 mM dithionite. Top row: experimental data. Bottom row: simulation of the auxiliary [4Fe–4S][^{13}CN] cluster overlaid onto experimental data (gray). Experimental conditions: 9.730 GHz, $\pi/2 = \pi = 16$ ns, 10 K. Simulation parameters are the same as in Fig. S7: A = [–5.0, –4.0, 0.9] MHz, Euler angles of [-90˚, -40˚, 0˚], g = [2.09, 1.94, 1.93] (blue).

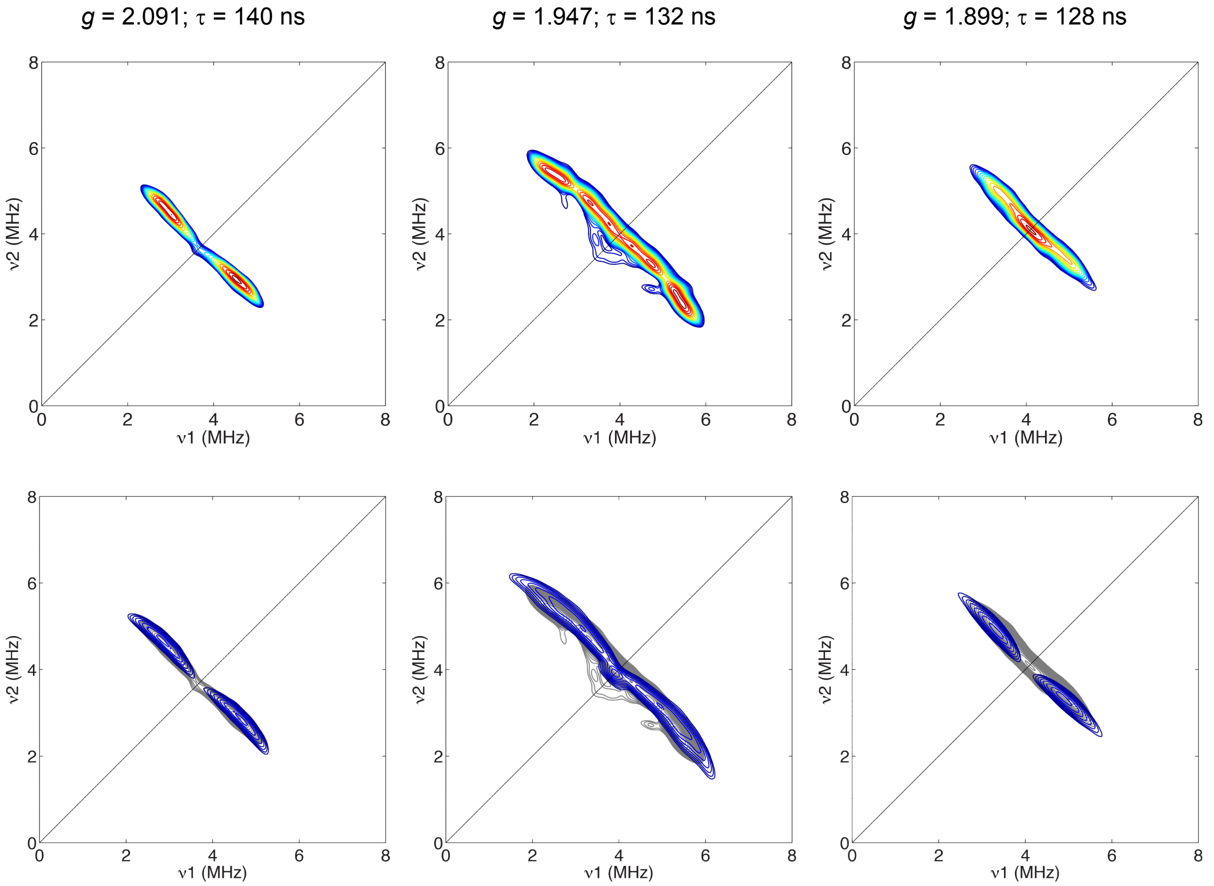


Table S1. Data collection and refinement statistics.

	HydG	HydG Fe (SAD)
PDB accession code	4WCX	
Data collection		
Space group	P1	
Cell dimensions		
a, *b*, *c* (Å)	54.13, 56.19, 84.92	
α, β, γ (°)	89.59, 83.62, 66.84	
		Peak
Wavelength	0.97625	1.73891
Resolution (Å)	84.33-1.59 (7.11-1.59)	84.41-1.84 (8.23-1.84)
R_{sym} or R_{merge}	0.039 (0.033-0.349)	0.049 (0.043-0.484)
$I / \sigma I$	12.3 (30.6-2.2)	14.6 (26.9-2.1)
Completeness (%)	96.2 (96.4-94.5)	69.4 (94.0-5.6)
Redundancy	2.7 (2.7-2.7)	3.3 (3.3-2.3)
Refinement		
Resolution (Å)	84.44-1.59	
No. reflections (free set)	116166 (1995)	
R_{work} / R_{free}	0.18/0.22	
No. atoms:	15,504	
Chain A / Chain B	7,331 + 7,247	
Ligand/ion	48 + 63	
Water	815	
B-factors:		
Chain A / Chain B	32.0 / 28.4	
Ligand/ion	28.5 / 22.2	
Water	37.8	
R.m.s deviations:		
Bond lengths (Å)	0.027	
Bond angles (°)	2.083	

Table S2. Ligand interaction distances.

Monomer	Cluster	Ligand	Bond	Observed distance (Å)	Expected distance (Å)	Reference
A	RS	Methionine	Fe_U-N	2.7	-	
			Fe_U-O	2.5	-	
	Aux	-	Fe_L-μ_2-S	2.5	-	
			μ_2-S-Fe_U	2.3	-	
		Amino acid	Fe_L-N	2.1	-	
			Fe_L-O	2.1	2.01	(1)
		Water A	Fe_L-$Water_A$	2.2	2.09	(1)
		Water B	Fe_L-$Water_B$	2.1	2.09	(1)
		His265	Fe_L-N	2.2	2.08	(1)
B	RS	SAM	Fe_U-N	2.3	2.0-2.6	(2)
			Fe_U-O	2.2	2.0-2.5	(2)
			Fe_U-S	3.4	3.4-4.2	(2)
	Aux	-	S-Fe_U	2.2	-	

Abbreviations: RS, Radical SAM; Aux, auxiliary; $Fe_{U,,}$ iron atom from a [4Fe-4S] core without a cysteine ligand; Fe_L, labile iron of the auxillary cluster and μ_2-S, bridging sulfide ion of the auxiliary cluster.

References

(1) Harding MM (2000) The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. Acta Crystallogr D Biol Crystallogr 56(Pt 7):857-867

(2) Vey JL & Drennan CL (2011) Structural insights into radical generation by the radical SAM superfamily. Chem Rev 111(4):2487-2506.