Supporting Information

Materials and Methods

Crystallization and Data Collection

The expression and purification of the MoFe-protein from *Azotobacter vinelandii* (Av) cells were described previously[1]. MoFe-protein was crystallized by sitting-drop vapor diffusion at room temperature (~ 295 K) in an anaerobic chamber. The MoFe-protein solution at 30 mg/ml was mixed in 1:1 volume ratio with the reservoir solution containing 13% polyethylene glycol (PEG) 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0, unless specified elsewhere. For flash-cooling, the crystals were transferred into harvesting solutions consisting of the reservoir solution and increasing amounts of 2-methyl-2,4-pentanediol (MPD). This was carried out in steps of 5 % MPD for 15 min each, until 15 % MPD were reached in the harvesting buffers.

Full sets of anomalous diffraction data were collected on Beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL), with a DECTRIS Pilatus 6M. A reference set of 1034 diffraction images was collected at 12,658 eV with an oscillation angle of 0.15°. Two sets of anomalous data were collected below and above the iron edge, at 7080 eV and at 7130 eV, respectively. Complete sets of diffraction data were collected at these energies with an oscillation angle of 0.15° yielding a total of 1034 images each for both standard and inverse beam modes of data collection.

Partially-complete multiple-wavelength anomalous diffraction (MAD) data sets were collected on Beamline 9-2 at SSRL, with an ADSC Q315 Detector. For the reference data set, 360 diffraction images were collected at 12,115 eV with an oscillation angle of 0.5°. MAD data were collected at 19 energies (7097, 7102, 7107, 7112, 7113, 7114, 7115, 7116, 7117, 7118, 7119, 7120, 7121, 7122, 7123, 7124, 7129, 7134 and 7139 eV) across the Fe K-edge. At every energy, a 10˚-wedge set of diffraction data (40 images with 0.5° rotation and inverse beam) was collected for refining only the $\Delta f'$ and $\Delta f''$ parameters of the heavy atoms; the atomic coordinates were fixed by the reference structure refined against the 12,115 eV data set.

Hard X-ray fluorescence emission spectra of the MoFe-protein crystals and solutions were collected on Beamline 12-2 at SSRL, with a silicon-drift detector.

Data Processing and Refinement

Data sets were integrated with the XDS program package,[2] fixing unit cell dimensions and crystal orientation for each energy to the values obtained from refining the reference data sets. All data sets were converted to structure factor amplitudes and scaled individually against the reference data set using programs of the CCP4 suite.[3] The diffraction data statistics are listed in Table S1 and S2 for full sets and 10˚ wedges of MAD data, respectively. The MoFe-protein model (PDB ID 3U7Q) was refined using REFMAC5[4] against the reference data set. The anomalous scattering factors around the Fe K-edge were refined for heavy atoms as described previously[5]. Molybdenum and sulfur were also included in the anomalous refinement procedure due to their significant anomalous signal in the Fe K-edge region; since the data sets were collected at energies remote from the sulfur edge, a single set of $\Delta f'$ and $\Delta f''$ parameters was used for all sulfur atoms.

X-ray Absorption Spectroscopy Data Collection and Analysis

Ferrous sulfate heptahydrate ($[FeSO_4]\bullet 7H_2O$, ≥ 99.0%) and ferric sulfate ($Fe_2(SO_4)_3]\bullet xH_2O$, 97%) used in this study were purchased from Sigma-Aldrich. The X-ray absorption spectra of ferrous sulfate heptahydrate and ferric sulfate hydrate were collected on the HXMA Beamline at the Canadian Light Source (CLS). The operational current in the storage ring of CLS is 200-250 mA at 2.9 GeV. The HXMA Beamline uses the synchrotron source with a wiggler field of 2 Tesla and employs a Si(220) double-crystal monochromator, a rhodium-coated vertical collimating mirror upstream of the monochromator and a toroidal focusing mirror (also rhodium-coated) downstream of the monochromator. Harmonic rejection at HXMA beamline was accomplished by detuning the second monochromator crystal to 50% of the maximum intensity. The incident and transmitted X-ray intensities were monitored using nitrogen-filled ionization chambers, and X-ray absorption was measured in transmittance mode. During data collection,

samples were maintained at a temperature of approximately 10 K using an Oxford instruments liquid helium flow cryostat. The energy was calibrated by reference to the absorption of a standard iron metal foil measured simultaneously with each scan, assuming the lowest energy inflection point of the iron foil to be 7111.2 eV.

The background reduction and normalization of X-ray absorption spectra were performed using EXAFSPAK (http://ssrl.slac.stanford.edu/~george/exafspak/exafs.htm). No smoothing, filtering, or related operations were performed on the data.

The X-ray absorption coefficient μ_α for both spectra are normalized at 7150 eV and converted to the imaginary part of the anomalous scattering factor f'' using the following equation:

$$f'' = mc\varepsilon_0 E\mu_\alpha / e^2\hbar \qquad (1)$$

where m is the electron mass, c is the speed of light in vacuum, ε_0 is the vacuum permittivity, e is the electron charge, E is the incident photon energy, $\hbar$ is Planck's constant.[6]

Metal Analyses by Inductively Coupled Plasma Mass Spectrometry

Inductively coupled plasma mass spectroscopy (ICP-MS) was conducted at the Caltech Environmental Analysis Center on a Hewlett-Packard 4500 ICP mass spectrometer (Agilent Technologies) with a CETAC ASX-500 autosampler (CETAC). Nitric acid, metal standards and water were from Sigma-Aldrich. For sample preparation, the desired amount of the MoFe-protein was mixed with 20% nitric acid and digested at 80 °C for >24 hours. Each digested solution was diluted to 10 ml solution to a 1% final concentration of nitric acid. The buffer for the MoFe-protein (200 mM NaCl and 50 mM Tris-HCl buffer at pH 7.8) was used as blank for background subtraction. The concentration of the MoFe-protein was determined by the absorbance at 410 nm with an absorption coefficient of 76000 $M^{-1}cm^{-1}$.

Sequence Alignment

BLAST sequence alignments were performed using the BLOSUM62 similarity matrix[7] against NCBI's referenced protein database. The aligned sequences were chosen with a total score cutoff of 100 and the subject sequence length ≥ 80% of the *Av* MoFe-protein *β*-subunit.

Supporting Tables

Table S1. Data processing statistics for full sets of MAD data.

	Higher-resolution Reference data	MAD data sets across Fe K-edge[1]
Resolution range (Å)	39.7 - 1.49	39.8 - 2.10
Unique reflections	320,774	135,027 (2381)
Completeness [%]	93.9	96.0 (0.4)
Anomalous completeness [%]	84.4	94.5 (0.5)
Rsym	0.077	0.070 (0.004)
Rsym (outer shell)[2]	0.507	0.186 (0.013)
I/σ(I)	8.8	15.5 (1.6)
I/σ(I) (outer shell)[2]	2	6.9 (0.3)
Multiplicity	2.9	5.5 (0.1)
Scale Factor	1	1.014 (0.001)

[1]Two MAD data sets were collected below and above the Fe-K edge, at 7080 eV and 7130 eV, respectively. The statistics shown in the table are the average of two data sets, with the standard deviations in parentheses.
[2]The outer shell ranges between 1.57 – 1.49 Å for the higher resolution reference data, and between 2.20 – 2.10 Å for the MAD data sets.

Table S2. Data processing statistics for 10° wedges of MAD data used for Figure 2.

Data Set	Higher-resolution Reference data	MAD data sets across Fe K-edge[2]
Resolution range (Å)	38.0 - 1.39	38.0 - 1.91
Unique reflections	830,984	122,405 (293)
Completeness [%]	98.9	18.4 (0.1)
Anomalous completeness [%]	-	16.8 (0.1)
Rsym	0.08	0.071 (0.004)
Rsym (outer shell)[1]	0.441	0.173 (0.010)
I/σ(I)	8.3	7.4 (0.4)
I/σ(I) (outer shell)[1]	3.6	3.6 (0.2)
Multiplicity	3.5	2.0 (0.04)
Scale Factor	1	1.175 (0.008)

[1]The MAD data sets were collected at 7097, 7102, 7107, 7112, 7113, 7114, 7115, 7116, 7117, 7118, 7119, 7120, 7121, 7122, 7123, 7124, 7129, 7134 and 7139 eV. The statistics shown in the table are the average of all data sets, with the standard deviations in parentheses.
[2]The outer shell ranges between 1.47 – 1.39 Å for the higher resolution reference data, and between 2.00 – 1.91 Å for the MAD data sets.

Table S3. The interatomic distances between the metals and ligands at the MMB–site.

	M····O Distance (Å)	
	Fe16	**Na1**
β–R108 O	2.09	2.34
β–E109 $O_{\varepsilon 2}$	2.28	–
β'–D353 $O_{\delta 1}$	–	2.44
β'–D353 $O_{\delta 2}$	2.34	2.23
β'–D357 $O_{\delta 2}$	2.28	2.47
HOH–1	2.23	–
HOH–2A	2.18	–
HOH–2B	–	2.38
Average	2.23	2.37

Table S4. Metal analyses on the MoFe-proteins by ICP-MS.

	1	**2**	**3**	**4**
Sample Description	Protein was from Prep-G, and in 50 mM Tris, 200 mM NaCl, pH8.0	Protein was from Prep-C, and in 50 mM Tris, 200 mM NaCl, pH8.0	Protein was from Prep-H, and in 50 mM Tris, 200 mM NaCl, pH8.0	Protein was from Prep-I, and in 50 mM Tris, 200 mM NaCl, pH8.0
[Fe]/[MoFe] [1]	31.77	34.89	32.44	34.20
[Mo]/[MoFe]	2.03	2.29	2.04	2.22
[Ca]/[MoFe]	1.47	0.76	0.89	1.54
[Mg]/[MoFe]	1.66	2.01	1.98	2.55
[Zn]/[MoFe]	0.14	0.13	0.23	0.27
[Fe]/[Mo]	15.67	15.27	15.91	15.40
[Ca]/[Mo]	0.74	0.33	0.44	0.70
[Mg]/[Mo]	0.82	0.88	0.98	1.15
[Zn]/[Mo]	0.07	0.06	0.11	0.12

[1][MoFe] is defined as the molar concentration of MoFe-protein tetramer.

Table S5. Estimation of the occupancy number of Fe16 in different MoFe-protein crystals by analysis of the MAD data

		1	2	3	4	5	6	7	8
Crystallization solution[1]		13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0	13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0	13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0	18% PEG 3325, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 9.5	18% PEG 3325, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 9.5	13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0	13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0	13% PEG 8000, 0.9 M NaCl and 0.1 M Tris-HCl buffer at pH 8.0
Protein Source[2]		Prep-A	Prep-B	Prep-C	Prep-D	Prep-C	Prep-E	Prep-F	Prep-D
Resolution Range[3] (Å)		50 - 1.39	40 - 1.49	40 - 1.5	40-1.52	40 - 1.57	50 - 1.99	40-2.1	40-2.1
B-Overall[4] (Å^2)		18.16	20.23	25.81	19.87	20.80	21.57	31.75	28.43
Edge Jump[5]	**AvgClustFe**	2.79 (0.22)[7]	4.24 (0.32)	3.14 (0.20)	4.16 (0.24)	3.68 (0.23)	2.57 (0.18)	2.72 (0.40)	3.15 (0.15)
	Fe16	1.69	2.85	2.4	3.14	1.18	1.24	4.18	0.25
Fe16 Occupancy[6]		0.4	0.5	0.6	0.5	0.2	0.3	1.1	0.1
Electron Density	**AvgClustFe[7]**	7.22 (0.11)[7]	5.50 (0.16)	3.15 (0.13)	4.92 (0.17)	4.77 (0.15)	6.66 (0.69)	2.97 (0.30)	2.94 (0.13)
	Fe16	4.9	3.9	2.2	2.43	1.88	5.8	3.2	1.62
	Fe16/ AvgClustFe	0.68	0.71	0.70	0.49	0.39	0.87	1.08	0.55

[1]The crystallization methods are those described in the Methods and Materials, except for the crystallization solutions given here.
[2]The protein batches used for data shown here were from purifications conducted in the years 2001 to 2011.
[3]The resolution range is for the higher resolution reference diffraction data. The highest resolution limits for the anomalous diffraction data collected across Fe K-edge are 1.9 – 2.1 Å.
[4]The overall B-factor is calculated for all the atoms in the crystallographic structure.
[5]The edge jump is defined by the difference of Δf" above and below Fe K-edge (at 7130 eV and 7080 eV respectively, except for Crystal #1 for which 7129 eV and 7097 eV are used and #6 for which 7129 eV and 7090 eV were used). All the Fe in the P-cluster and FeMo-cofactor of the MoFe-protein were used in calculating the average edge jump (AveClustFe).
[6]The occupancy number of Fe16 is estimated by comparing the ratio of the edge jump of Fe16 to the average of the Fe in the metallocluster to that of the normalized X-ray absorption spectra of ferrous sulfate heptahydrate and the MoFe-protein. When normalized at 7150 eV, the ratio of the defined edge jump between ferrous sulfate heptahydrate and the MoFe-protein is around 1.4.
[7]Fe5 and Fe6 in the P-cluster were excluded from the average due to the existence of alternative conformations with variable occupancy at these two Fe sites.
[8]The numbers in parentheses are the standard deviations.

Table S6. Conserved residue patterns in the Av MoFe-protein β-subunit (NifK) homologs.

Residues around the MMB-site		
α 432-EK-433	**β 107-RE-108**	**β 352-VDMMTD-357**
Nitrogenase		
[F|H|L|Q|T|V|Y]K (241)	[K|R]E (290)	XDXXXD (331)
[R|K]K (19)	[N|Q|H|A]E (25)	XDXXXN (2)
EK (14)	[K|R]D (19)	XNXXXN (1)
	QR (1)	XEXXXS (1)
Nitrogenase cofactor biosynthesis protein NifEN		
XK (194)	[K|R]E (99)	XDXXXD (121)
	[H|N|Q|W|S|C]E (16)	XEXXXS (1)
	[H|N]D (6)	XEXXXG (1)
	[K|R]Q (2)	XHXXXE (1)
	KH (1)	

1. The numbering of *Azotobacter vinelandii α*- and *β*-subunit sequences is used for the residues.
2. Residue **α**-Lys433 is the only **α**-subunit residue within ~4-Å range of the Fe16 site that is conserved in nitrogenase homologs.
3. The letters in the square brackets, separated by "|", list the variations of the given residue among the homologs. The numbers in the round brackets are the number of homologs from the sequence alignment.
4. X represents any amino acid residue.

Supporting Figures

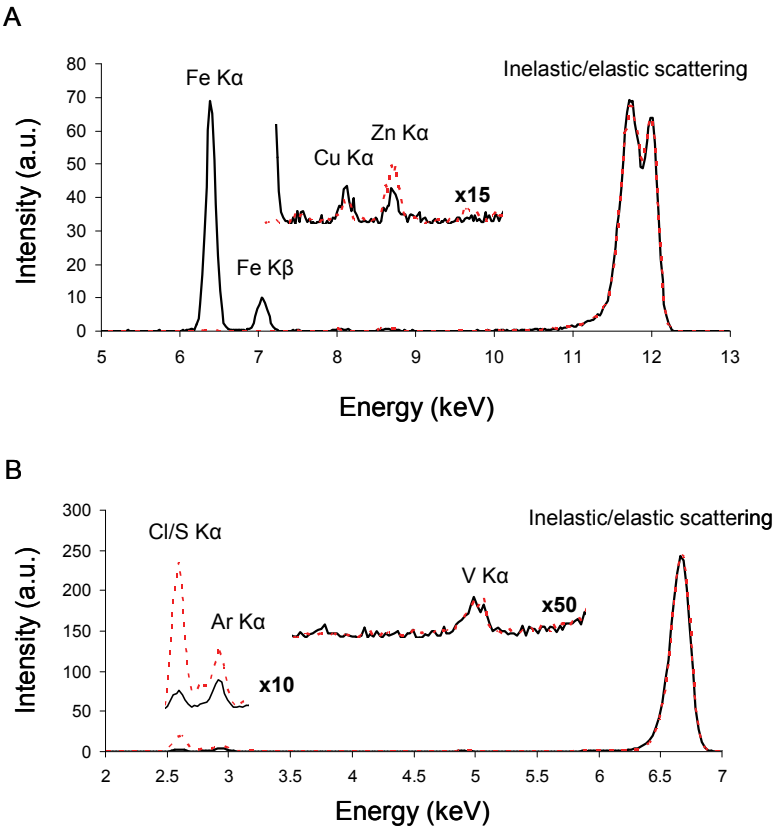

Figure S1. The X-ray emission spectra of the MoFe-protein crystal (solid curve in black) and the crystallization solution (dash curves in red) as control collected at 12 keV (A) and 6.67 keV (B), respectively. The X-ray emission spectrum of the crystallization solution is normalized to that of the MoFe-protein crystal at the scattering peak.

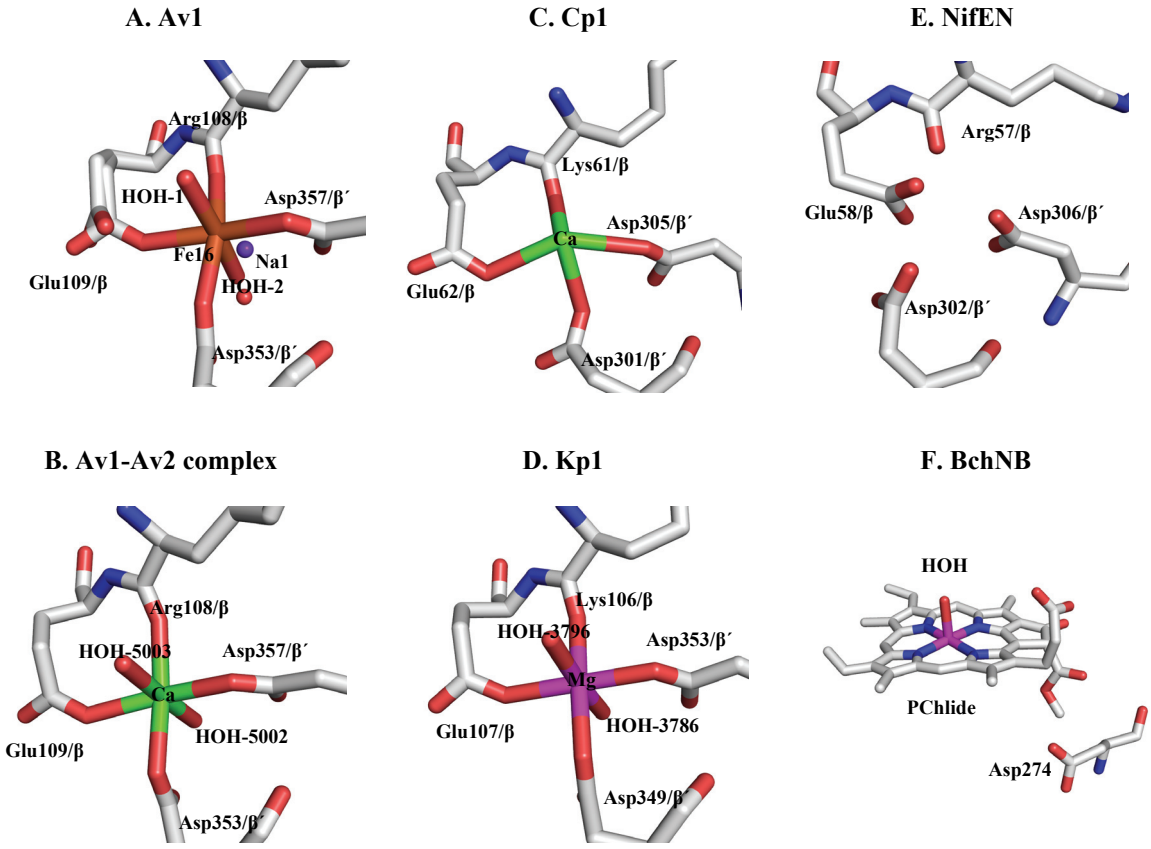

Figure S2. The MMB-site in the nitrogenase homologues A) *A. vinelandii* MoFe-protein (Av1) at 1.0 Å resolution (PDB code 3U7Q),[8] B) the nitrogenase complex Av1-Av2 at 2.3 Å resolution (PDB code 2AFK),[9], C) the MoFe-protein from *Clostridium pasteurianum* (Cp1) at 3.0 Å resolution (PDB code 1MIO),[10] D) the MoFe-protein from *Klebsiella pneumoniae* (Kp1) at 1.6 Å resolution (PDB code 1QGU),[11] E) the nitrogenase FeMo-cofactor biosynthesis protein NifEN at 2.4 Å resolution (PDB code 3PDI)[12] and F) the nitrogenase-like, light-independent protochlorophyllide reductase BchNB at 2.3 Å resolution (PDB code 3AEK).[13] Only one of the protein sidechains forming the MMB-site in the MoFe-protein (Asp357) is conserved in BchNB, and that residue (Asp274) is positioned adjacent to the substrate protochlorophyllide (PChlide). The remaining residues are > 10 Å from Asp274 and are omitted from the figure. No water molecules were included in the crystal structures of Cp1 (C) and NifEN (E). While no metal ion was modeled in the MMB site of NifEN due to the moderate quality of the diffraction data, the electron density in this region does not rule out a coordinated ion.

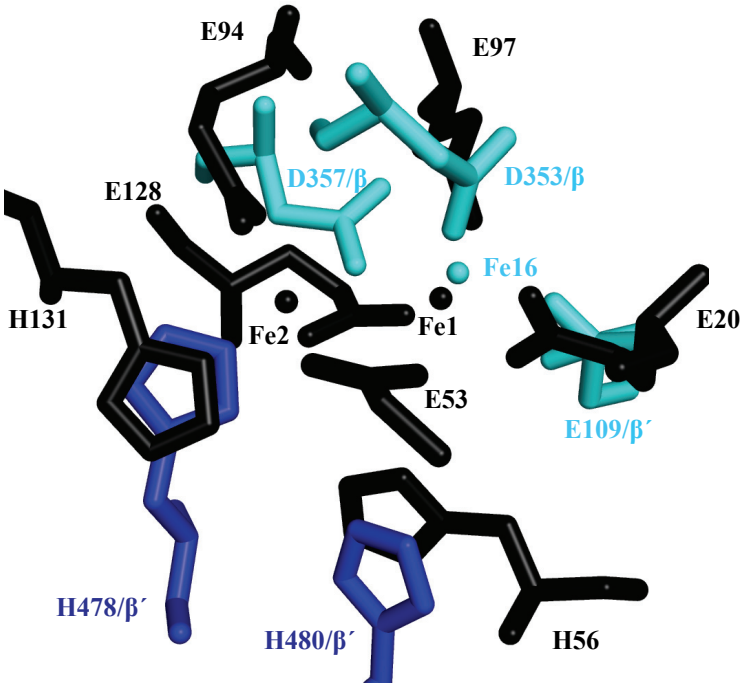

Figure S3. The MMB-site in *A. vinelandii* MoFe-protein at 1.0 Å resolution (PDB code 3U7Q)[8] (A) superposed with the di-iron center in the *Desulfovibrio vulgaris* rubrerythrin at 1.69 Å resolution (PDB code 1LKM)[14] (B). The residues in the MoFe-protein coordinated to Fe16 are highlighted in cyan, while those in the second shell to Fe16 are in dark blue color. All the residues near the di-iron site in the rubrerythrin structure are in black. The ligands that are not matching between the two structures (Arg108 and the water molecule at the MMB site, and the water molecule bridging the di-iron center) are not shown in the figure.

Supplementary References

[1] J. S. Kim, D. C. Rees, *Nature* **1992**, *360*, 553-560.
[2] W. Kabsch, *Acta Crystallogr. Sect. D* **2010**, *66*, 125-132.
[3] S. Bailey, *Acta Crystallogr. Sect. D* **1994**, *50*, 760-763.
[4] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240-255.
[5] O. Einsle, S. L. A. Andrade, H. Dobbek, J. Meyer, D. C. Rees, *J. Am. Chem. Soc.* **2007**, *129*, 2210-2211.
[6] R. W. James, in *The Optical Principles of the Diffraction of X-rays.*, G. Bell And Sons Limited, **1969**.
[7] S. Henikoff, J. G. Henikoff, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10915-10919.
[8] T. Spatzal, M. Aksoyoglu, L. M. Zhang, S. L. A. Andrade, E. Schleicher, S. Weber, D. C. Rees, O. Einsle, *Science* **2011**, *334*, 940-940.
[9] F. A. Tezcan, J. T. Kaiser, D. Mustafi, M. Y. Walton, J. B. Howard, D. C. Rees, *Science* **2005**, *309*, 1377-1380.
[10] J. Kim, D. Wood, D. C. Rees, *Biochemistry* **1993**, *32*, 7104-7115.
[11] D. M. Lawson, S. M. Mayer, C. A. Gormal, S. M. Roe, B. E. Smith, *J. Mol. Biol.* **1999**, *292*, 871-891.
[12] J. T. Kaiser, Y. L. Hu, J. A. Wiig, D. C. Rees, M. W. Ribbe, *Science* **2011**, *331*, 91-94.
[13] N. Muraki, J. Nomata, K. Ebata, T. Mizoguchi, T. Shiba, H. Tamiaki, G. Kurisu, Y. Fujita, *Nature* **2010**, *465*, 110-114.
[14] S. Jin, D. M. Kurtz, Jr., Z. J. Liu, J. Rose, B. C. Wang, *J. Am. Chem. Soc.* **2002**, *124*, 9845-9855.