Supporting Online Material for

Structure of precursor bound NifEN: a nitrogenase FeMo cofactor maturase/insertase*

Jens T. Kaiser,[1]** Yilin Hu,[2]** Jared A. Wiig[2], Douglas C. Rees,[1,3]† Markus W. Ribbe[2]†

[1] Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail Code 114-96, Pasadena, CA 91125, USA.

[2] Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3900

[3] Howard Hughes Medical Institute, California Institute of Technology, Mail Code 114-96, Pasadena, CA 91125, USA

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at http://www.sciencemag.org/.

** These authors contributed equally to this work.

†To whom correspondence should be addressed: E-mail: dcrees@caltech.edu (D.C.R.); mribbe@uci.edu (M.W.R.)

This PDF file includes:

MATERIALS AND METHODS

Data Collection. Crystals of NifEN were cryogenically protected by stepwise transfer from mother liquor to a 10 µL drop of mother liquor in which 1-2 µg of sucrose were dissolved. Subsequently, they were transferred to an artificial mother liquor containing 25% (w/v) sucrose and flash frozen in liquid nitrogen. NifEN remained catalytically competent in the crystallized state, showing maturation activities (*S1*) of ~800 nmol C_2H_4 formation per mg protein per min when dissolved in a 25 mM Tris-HCl buffer (pH 8.0). All protein and crystal manipulations up to this step were performed in an inert gas atmosphere of 90% Ar and 10% H_2. About 300 cryogenically protected crystals were screened for diffraction at beam lines 12-2 and 11-1 at SSRL. Generally, crystals diffracted anisotropically to resolutions between 2.0 Å and 3.5 Å, but patterns were streaky and split in one direction. About 10 crystals exhibited diffraction patterns of sufficient quality for the assignment of cell constants and the subsequent integration and scaling. The initial processing was performed in the primitive monoclinic system with cell constants in the ranges of a=145.5-147.3 Å, b=94.7-96.0 Å, c=146.7-154.4 Å, and β=91.1-100.1°. Some crystals could be integrated using a C-centered orthorhombic lattice (a≈196 Å, b≈223 Å, c≈96 Å), but merging R-factors were unsatisfactory below a resolution of ~6 Å. About 30 datasets at different wavelengths (0.98 Å, 1.033 Å, 1.742 Å, 1.737 Å and 1.738 Å) were collected, using varying beam sizes, scanning across crystals, different exposure times and frame widths. All data were finally processed in space group $P2_1$ with MOSFLM (*S2*) or XDS (*S3*), and scaled and merged with SCALA (*S4*). Final R-merges were between 8.5 and 17.7%. Datasets collected near the Fe-absorption edge typically showed good anomalous correlation coefficients to a resolution of 5.5 Å, depending on the dataset. Native Patterson maps showed no significant peaks longer than 10 Å. The best dataset (designated F8r) was used for refinement after anisotropy correction [anisotropy server (*S5*)]. Data processing statistics are shown in Table S1.

Structure Solution. Molecular replacement with MOLREP (*S6*) and PHASER (*S7*) were attempted using various modifications of PDB entry 1M1N [NifDK structure at 1.16 Å resolution (*S8*); NifDK shares approximately 28% sequence homology with NifEN]. A homology model was built with SCWRL (*S9*) using the FFAS server (*S10*). In one of the datasets, a molecular replacement solution with two tetramers per asymmetric unit could be found with either program, utilizing a mixed model (non-conserved residues substituted by serines) with the same tetrameric arrangement as NifDK. Anomalous difference Fourier maps with model phases were used to locate the cluster centers. Completion of the heavy atom substructure was achieved using the experimental phasing capabilities of PHASER (*S7*). Experimental maps were density modified with PARROT (*S11*) and cyclically averaged (2 x 4 fold) with DM (*S12*) or MAIN (*S13*). Initial NCS-operators were taken from the replacement model. The resulting electron density maps were of excellent quality and established that the conformation of NifEN is closer to that of apo NifDK (*S14*). The positions of the clusters were easily identified by peaks in the anomalous difference Fourier maps, and the corresponding $2F_o$-F_c maps allowed an assignment of the cluster geometries.

Model Building and Refinement. Despite the non-isomorphism between NifEN crystals, a free-R-set (5%) was defined in the early stages of processing up to a resolution of 2.0 Å and used for all datasets. Model building using MAIN and COOT (*S15*) and refinement using REFMAC (*S16*), MAIN, CNS (*S17*) and PHENIX (*S18*) were performed using various models and various datasets. Transfer of a model between datasets was done by either molecular replacement or rigid body refinement. Hydrogen bond restraints in MAIN and Ramachandran potentials in CNS were utilized to maintain satisfactory model geometry. The final model was extensively refined with PHENIX against the anisotropy corrected dataset F8r. Final statistics, as reported by the validation module of PHENIX, are shown in Table S2. It was difficult to unambiguously establish the structure of the L-cluster in this study or to determine whether or

not non-protein light atom ligands are present, including ones that might correspond to the homocitrate or the interstitial ligand of the M-cluster. While challenging, we would have expected a more definitive assignment at this resolution, so it is also possible that compositional heterogeneity and/or structural disorder of the cofactor and adjacent N-terminal residues of the α-subunit of NifEN may contribute to the diffuseness of the electron density in this region.

SUPPLEMENTARY TABLES

Table S1. Data Processing Statistics

	F8r	F8e
Space group	$P2_1$	
Wavelength (Å)	1.033	1.738
a (Å)	146.07	146.03
b (Å)	95.22	95.29
c (Å)	149.98	150.32
β (°)	95.50	95.41
d_{min} (Å)	2.6 (2.74)	3.4 (3.8)
R_{mrg} (%)	7.0 (38.9)	12.7 (30.5)
Completeness (%)	96.8 (84.6, 94.8)	98.0 (93.9)
Multiplicity	3.4 (2.4, 2.8)	6.0 (6.0)
I/σ(i)	7.7 (1.5, 2.4)	8.5 (4.7)
Anomalous Multiplicity	–	3.0 (2.4)
CC_{ano} < 0.3 (Å)	–	6.0

Table S2. Refinement Statistics

Resolution (Å)	39.2-2.6	2.64-2.6
R_{work} (%)	26.3	30.16
R_{free} (%) (10% of data)	30.3	35.3
Completeness (%)	98.5	83.5
RMS(bonds) (Å)	0.014	
RMS(Angles) (°)	1.77	
<B> (Å^2)	44.7	
Ramachandran Outliers (%)	4.9	
Clashscore	73.8	
Cβ outliers	33	

The refinement statistics are reported at a resolution of 2.6 Å by the validation module of Phenix (*S18*). The structure was refined against a data set corrected for anisotropy (*S5*) by ellipsoidal truncation to reciprocal resolutions of 1/2.4, 1/2.4, and 1/2.7 Å^{-1} along a*, b*, c*, respectively; anisotropic scaling with temperature factors of 12, 12 and -24 Å^2 along a*, b*, c* and application of an overall temperature factor of -16 Å^2 to restore the magnitude of the high resolution reflections diminished by anisotropic scaling.

SUPPLEMENTARY FIGURES

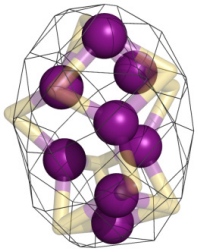

Fig. S1. Structure of L-cluster overlaid with anomalous difference electron density map. The L-cluster is illustrated as a ball-and-stick model, with Fe atoms shown in purple spheres and the remaining portion of the cluster in transparent sticks. PYMOL was used to prepare the figure (*S19*).

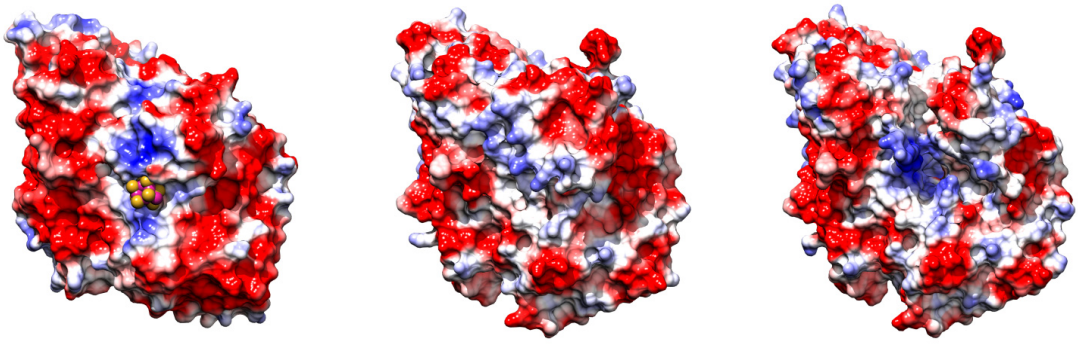

Fig. S2. Electrostatic surface potential representations of the αβ-pairs of NifEN (left), NifDK (middle) and apo NifDK (right). Negative and positive potentials are shown in red and blue, respectively. The structures are presented in the same orientation as those in Fig. 1B. The surface exposed L-cluster of NifEN is shown as a space-filling model, with atoms colored as follows: Fe, purple; S, yellow. CHIMERA (*S20*) was used to prepare the figure.

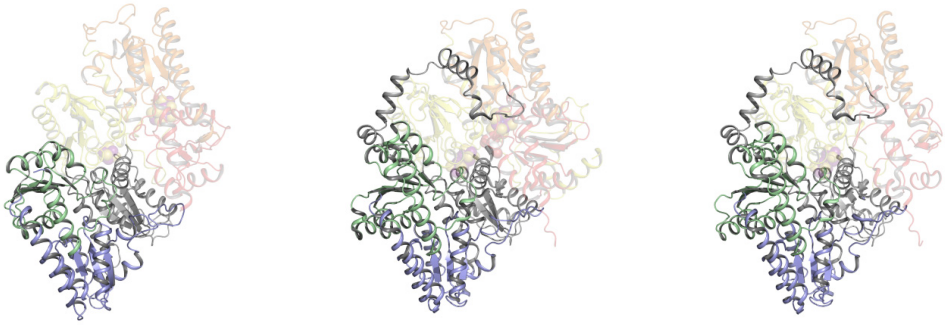

Fig. S3. Structures of the αβ-pairs of NifEN (left), NifDK (middle) and apo NifDK (right). The β-subunits are presented in the foreground, and the α-subunits are rendered transparent in the background. The domains of the subunits of all three proteins are colored as in Fig. 1A. PYMOL was used to prepare the figure (*S19*).

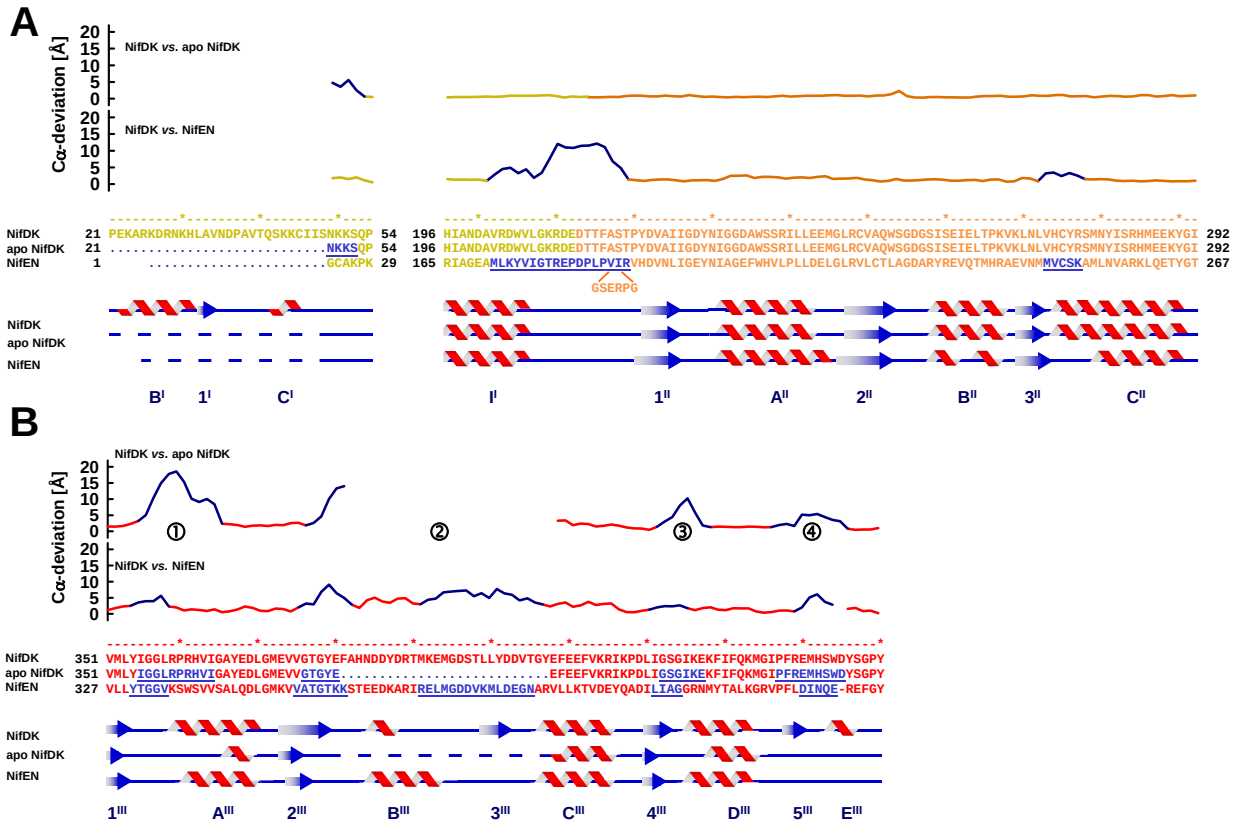

Fig. S4. **(A)** C_α-deviations of residues of the αI- and αII-domains of NifDK *vs.* apo NifDK (upper graph) and NifDK *vs.* NifEN (lower graph). The C_α-deviations were obtained after the least-square superposition of C_α-positions in subunits by CHIMERA (*S20*). The C_α-deviations of residues in the αI- and αII-domains are colored yellow and orange, respectively. Regions with significant deviations in C_α-positions are colored blue. The primary sequences and secondary structural elements of the αI- and αII-domains of NifDK, apo NifDK and NifEN that correspond to the regions in the C_α-deviation graphs are shown below. The α- and 3_{10}-helices are shown in red (labeled in consecutive letters with α-domains indicated in superscripts), and the β-strands are shown in blue (labeled in consecutive numbers with α-domains indicated in superscripts). Disordered regions are shown as broken lines. **(B)** C_α-deviations of residues of the αIII-domains of NifDK *vs.* apo NifDK (upper graph) and NifDK *vs.* NifEN (lower graph). The C_α-deviations were obtained after the least-square superposition of C_α-positions in subunits by CHIMERA (*S20*). Regions with significant deviations in C_α-positions are colored blue. The primary sequences and secondary structural elements of the αIII-domains of NifDK, apo NifDK and NifEN that correspond to the regions of the C_α-deviation graphs are shown below. The α- and 3_{10}-helices are shown in red (labeled in consecutive letters with superscripts indicating the αIII-domain) and β-strands in blue (labeled in consecutive numbers with superscripts indicating the αIII-domain). The disordered region is shown as a broken line.

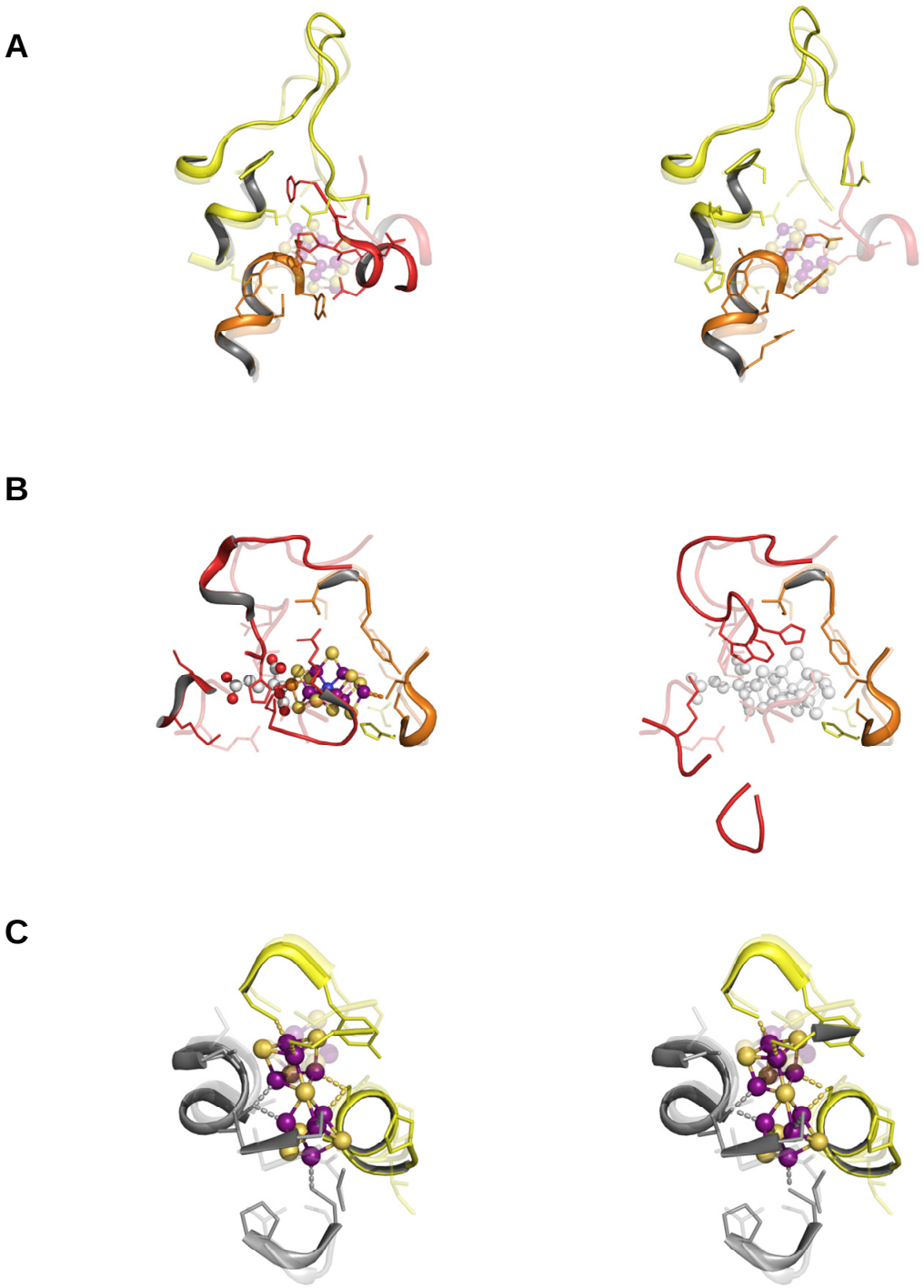

Fig. S5. Overlays of the L-cluster sites **(A)**, M-cluster sites **(B)** and O/P-cluster sites **(C)** between NifEN and NifDK (left) and between NifEN and apo NifDK (right). Shown are the same parts of the backbones and side chain residues as those in Fig. 4 in the close vicinities of the clusters. The presentation style and color coding are identical to those in Fig. 4, except that NifEN is rendered transparent. PYMOL was used to prepare the figure (*S19*).

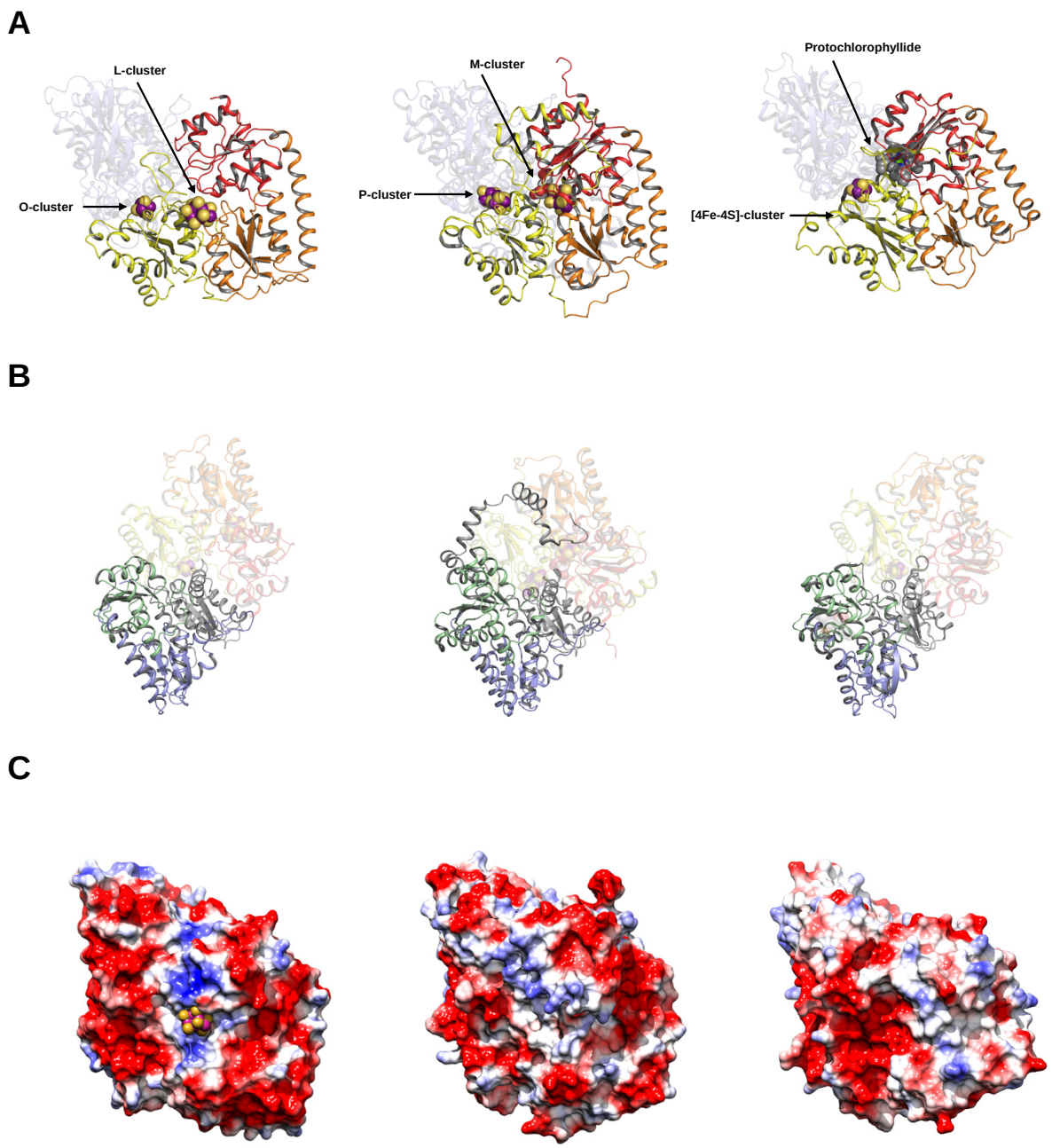

Fig. S6. **(A)** Structures of the αβ-pairs of NifEN (left), NifDK (middle, PDB entry 1M1N) and BchNB (right, PDB entry 3AEK). The α-subunits are presented in the foreground, and the β-subunits are rendered transparent in the background. The domains of the subunits of all three proteins are colored as in Fig. 1A. All clusters and protochlorophyllide are illustrated as space-filling models, with atoms colored as follows: Fe, purple; S, yellow; O, red; C, gray; N, blue; Mg, green. The Mo atom and the interstitial ligand of the M-cluster are not visible. **(B)** Structures of the αβ-pairs of NifEN (left), NifDK (middle) and apo BchNB (right). The β-subunits are presented in the foreground, and the α-subunits are rendered transparent in the background. The domains of the subunits of all three proteins are colored as in Fig. 1A. **(C)** Electrostatic surface potential representations of the αβ-pairs of NifEN (left), NifDK (middle) and BchNB (right). Negative and positive potentials are shown in red and blue, respectively. The structures are presented in the same orientation as those in A. The surface exposed L-cluster of NifEN is shown as a space-filling model, with atoms colored as follows: Fe, purple; S, yellow. PYMOL (*S19*) and CHIMERA (*S20*) were used to prepare the figure.

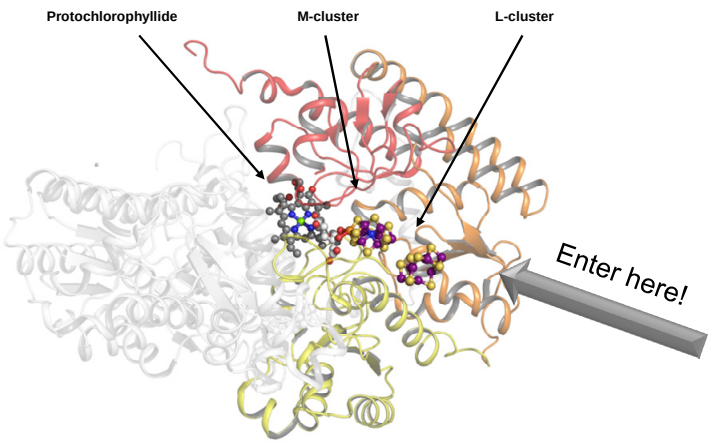

Fig. S7. Positions of L-cluster, M-cluster and protochlorophyllide along the insertion funnel in apo NifDK from the entrance all the way to the bottom. The positions of the clusters and protochlorophyllide were determined by a structural overlay of apo NifDK, NifEN, NifDK and BchNB. The domains of the subunits of apo NifDK are colored as in Fig. 1A. The β-subunit of apo NifDK is rendered transparent. All clusters and protochlorophyllide are illustrated as ball-and-stick models, with atoms colored as follows: Fe, purple; S, yellow; O, red; C, gray; N, blue; Mg, green; Mo, orange; center atom (X), dark blue. PYMOL was used to prepare the figure (*S19*).

SUPPLEMENTARY REFERENCES

S1. Y. Hu, A. W. Fay, M. W. Ribbe, *Proc. Natl. Acad. Sci. USA* **102**, 3236 (2005).

S2. A. G. W. Leslie, in *Crystallographic Computing V*, D. Moras, A. D. Podjarny, J. C. Thierry, Eds (Oxford University Press, UK, 1990), pp. 27-38.

S3. W. Kabsch, *J. Appl. Crystallogr.* **26**, 795 (1993).

S4. P. R. Evans, *Acta Crystallogr.* **D62**, 72 (2006).

S5. M. Strong et al., *Proc. Natl. Acad. Sci. USA* **103**, 8060 (2006).

S6. A. Vagin, A.Teplyakov, *J. Appl. Crystallogr.* **30**, 1022 (1997) .

S7. A. J. McCoy *et al.*, *J. Appl. Crystallogr.* **40**, 658 (2007).

S8. O. Einsle *et al.*, *Science* **297**, 1696 (2002).

S9. A. A. Canutescu, A. A. Shelenkov, R. L. Dunbrack, Jr. , *Protein Science* **12**, 2001 (2003).

S10. L. Jaroszewski, L. Rychlewski, Z. Li, W. Li, A. Godzik, *Nucl. Acids Res.* **33**, W284 (2005).

S11. K. Cowtan, *Acta Crystallogr.*, **D66**, 470 (2010).

S12. K. Cowtan, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34 (1993).

S13. D. Turk, Ph.D. thesis, Technical University, Munich, Germany (1992).

S14. B. Schmid *et al.*, *Science* **296**, 352 (2002).

S15. P. Emsley, B. Lohkamp, W. Scott, K. Cowtan, *Acta crystallogr.* **D66**, 486 (2010).

S16. G. N. Murshudov, A. A.Vagin, E. J. Dodson, *Acta Crystallogr.* **D53**, 240 (1997).

S17. A. T. Brunger, *Nat. Protoc.* **2**, 2728 (2007).

S18. P. D. Adams *et al.*, *Acta Crystallogr*. **D66**, 213 (2010).

S19. www.pymol.com

S20. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).