

**The Electron Bifurcating FixABCX Protein Complex from *Azotobacter vinelandii*:
Generation of Low-Potential Reducing Equivalents for Nitrogenase Catalysis**

Rhesa N. Ledbetter¹, Amaya M. Garcia Costas², Carolyn E. Lubner³, David W. Mulder³,
Monika Tokmina-Lukaszewska², Jacob H. Artz⁴, Angela Patterson², Timothy S.
Magnuson⁵, Zackary J. Jay⁶, H. Diessel Duan⁷, Jacquelyn Miller², Mary H. Plunkett⁸,
John P. Hoben⁷, Brett M. Barney⁸, Ross P. Carlson⁶, Anne-Frances Miller⁷, Brian
Bothner², Paul W. King³, John W. Peters^{2,4}, and Lance C. Seefeldt^{1*}

¹Department of Chemistry and Biochemistry, Logan, UT 84322

²Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT
59717

³National Renewable Energy Laboratory, Golden, CO 80401

⁴Institute of Biological Chemistry, Washington State University, Pullman, WA 99163

⁵Department of Biological Sciences, Idaho State University, Pocatello, ID 83201

⁶Department of Chemical and Biological Engineering, Center for Biofilm Engineering
and Thermal Biology Institute, Montana State University, Bozeman, MT 59717

⁷Department of Chemistry, University of Kentucky, Lexington, KY 40506

⁸Department of Bioproducts and Biosystems Engineering and Biotechnology Institute,
University of Minnesota, St. Paul, MN 55108

*Corresponding author: Lance C. Seefeldt, 0300 Old Main Hill, Logan, Utah, 84322

Phone: +1.435.797.3964, Email: lance.seefeldt@usu.edu

Table S1. Primers for Δfix mutant generation.

| Primer name | Sequence | Purpose |
|-------------|---|--------------|
| fix101 | 5'-GGTGGGAATTCCGATCTGCATGGCGCCCG-3' | Construct |
| fix102 | 5'-CGGC <u>GGATCC</u> GGCATAGGAAGTGGCCAGGGT-3' | Construct |
| fix103 | 5'-TCA <u>AGGATCC</u> CTTCATCGCCTCACGTTCG-3' | Construct |
| fix104 | 5'-CGT <u>GAAGCTT</u> CTGC <u>GGCG</u> CTGTACTTC-3' | Construct |
| fix107 | 5'-ACATT <u>CGGT</u> GCAATGGTCGG-3' | Verification |
| fix108 | 5'-CCAT <u>CATGCGCC</u> AGGGTG-3' | Verification |
| fix105 | 5'-GTCACGT <u>CTGGGT</u> CTGCAT -3' | Verification |
| fix106 | 5'-CGGTCAGT <u>GCCGG</u> CAGG-3' | Verification |
| fix109 | 5'-GCCACCGACAGCGCCAG-3' | Verification |

Table S2. Primers for overexpression of FixABCX.

| Primer name | Sequence | Purpose |
|-------------|---|--------------|
| BB2290 | 5'- NNNTCTAGACATATGCCCTACAAGATCAACGGTTCC-3' | Construct |
| BB2291 | 5'-NNNGATCCTCAGTTCTGGCCGAGCGG-3' | Construct |
| BBP2286 | 5'-CTGGAAGGCACCAATGAAATGCG-3' | Verification |
| BBP2287 | 5'-GCTGTGCACCACCATCCTCACCCAG-3' | Verification |
| BBP2288 | 5'-CGCCATCGAGAAGATCATTCCCGAC-3' | Verification |
| BBP2289 | 5'-GTCAACGGCATCCATCGCGAAG-3' | Verification |
| BB2283 | 5'-TCAGAGGGCTTGGTAGCGCGGTAC-3' | Verification |

Table S3 (inserted at the end of this document due to size). Protein-protein interactions captured, (A) the inter-cross-links and (B) intra-cross-links, within FixABCX complex during cross-linking reaction with BS3. Mapping of intra-subunit cross-links onto the homology models was consistent with the expected spanning distance of 20-25 Å between C-alpha chains. It should be noted that a subset of cross-links mapped to regions spanning 26-30 Å (for Lys-Lys) and higher. This is to be expected and has been explained by protein flexibility using known systems.¹

Table S4. Overview of the EPR signals observed for FixABCX from *A. vinelandii*.

| Signal Name | g-values | T _{opt} (K) | Samples present | Percent contribution | Possible source |
|----------------------|------------------------|----------------------|--------------------------------------|----------------------|--|
| Fast relaxing, broad | * | 5 K | As-purified NADH | * | Unknown |
| Rhombic 2.07 | 2.072, 1.940, 1.895 | 15 K | NADH Na-dithionite | 86 87 | [4Fe-4S] ¹⁺ |
| Axial 2.04 | 2.041, 1.944, 1.944 | 5 – 10 K | NADH Na-dithionite | 9 12 | [4Fe-4S] ¹⁺ |
| Isotropic 2.005 | 2.005 | > 50 K | As-purified NADH Na-dithionite | * | Flavin radical |
| Isotropic 2.005 | 2.005 | > 50 K | As-purified NADH Na-dithionite | 1 1 | Flavin radical |
| Axial 2.03 | 2.030, 2.00, 2.00 | 10 K | NADH | 4 | Interacting flavin radical and [4Fe-4S] ¹⁺ |

*Not determined

Table S5. Calculated Gibbs free energies of three possible reactions catalyzed by the Fix complex and three possible non-enzymatic reactions.

| Reaction | ΔG_{rxn}° (kJ/mol) |
|--|---------------------------------|
| $2 \text{NADH} + 2 \text{Fld}^{\text{Sq}} + \text{CoQ} = 2 \text{Fld}^{\text{Hq}} + \text{CoQH}_2 + 2 \text{NAD}^+$ | -36.7 |
| $\text{NADH} + \text{CoQH}_2 + 4 \text{Fld}^{\text{Sq}} = 4 \text{Fld}^{\text{Hq}} + \text{CoQ} + \text{NAD}^+ + 3 \text{H}^+$ | 118 |
| $\text{NADH} + 2 \text{Fld}^{\text{Hq}} + 2 \text{CoQ} + 3 \text{H}^+ = 2 \text{CoQH}_2 + 2 \text{Fld}^{\text{Sq}} + \text{NAD}^+$ | -154 |
| $2 \text{Fld}^{\text{Sq}} + \text{CoQ} + 2 \text{H}^+ = 2 \text{Fld}^{\text{Ox}} + \text{CoQH}_2$ | -36.7 |
| $\text{NADH} + \text{CoQ} + \text{H}^+ = \text{CoQH}_2 + \text{NAD}^+$ | -63.7 |
| $\text{NADH} + 2 \text{Fld}^{\text{Ox}} = 2 \text{Fld}^{\text{Sq}} + \text{NAD}^+ + \text{H}^+$ | -27.0 |

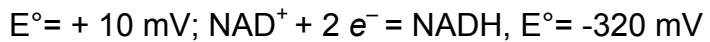
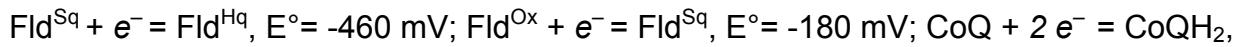


Table S6. Fitting of the ASQ kinetic curves for *R. rubrum* FixAB and *A. vinelandii* FixABCX reveal a biphasic decay, with one component having a lifetime of tens of picoseconds and a slightly longer lifetime of one thousand picoseconds. Fitting was performed in Igor Pro using a double exponential fit function.

| Sample | ASQ τ_1 | ASQ τ_2 |
|------------------------------|--------------------------------|--------------------------------|
| <i>R. rubrum</i> FixAB | 14 ps | 970 ps |
| <i>A. vinelandii</i> FixABCX | 14 ps | 1140 ps |

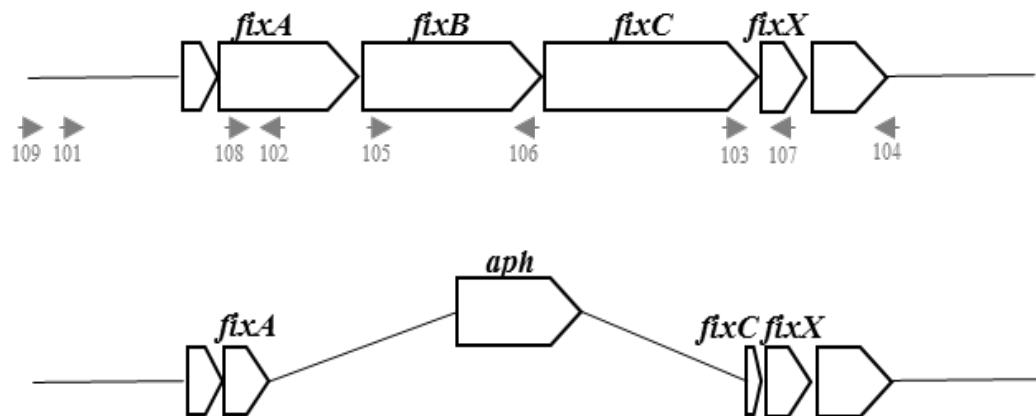


Figure S1. Scheme for Δfix mutant generation.

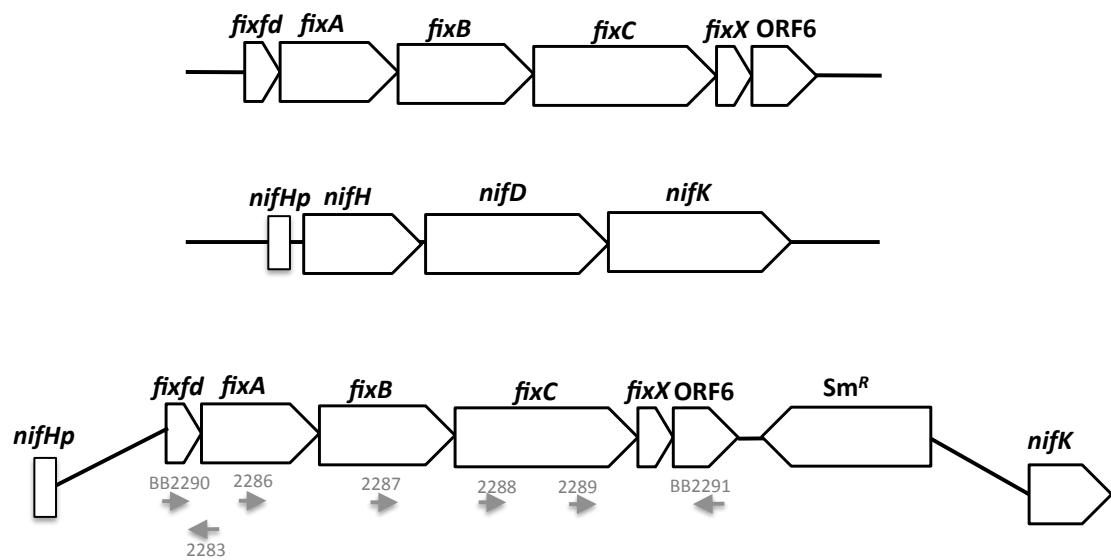


Figure S2. Scheme for overexpression of *fix* genes under the *nifH* promoter in *A. vinelandii*.

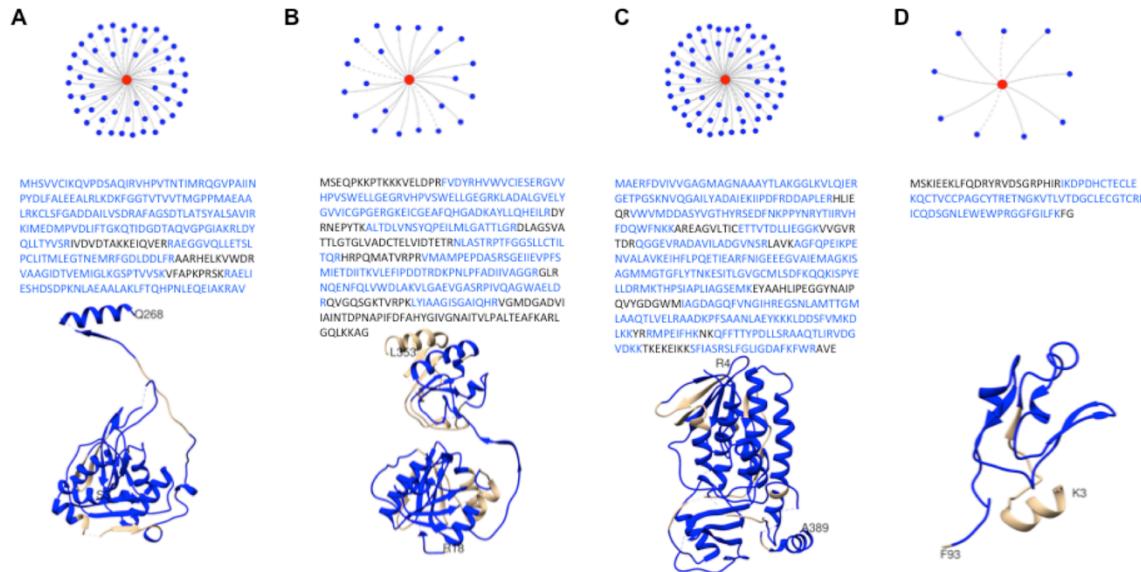


Figure S3. Protein identification within FixABCX complex purified from *A. vinelandii*.

The identity of FixA (panel A), FixB (panel B), FixC (panel C), and FixX (panel D) subunits was confirmed based on 54, 22, 59, and 9 unique tryptic peptides that correspond to 79%, 63%, 85%, and 73% sequence coverage (top row). Peptides and proteins are represented by blue and red spheres, respectively. Identified peptides were rendered in blue on primary sequence (middle row) and homology models of corresponding subunits (bottom row). Homology models were generated based on following protein templates: FixA, 4L2I; FixB, 3CLR; FixC, 4K2X; FixX, 2GMH.

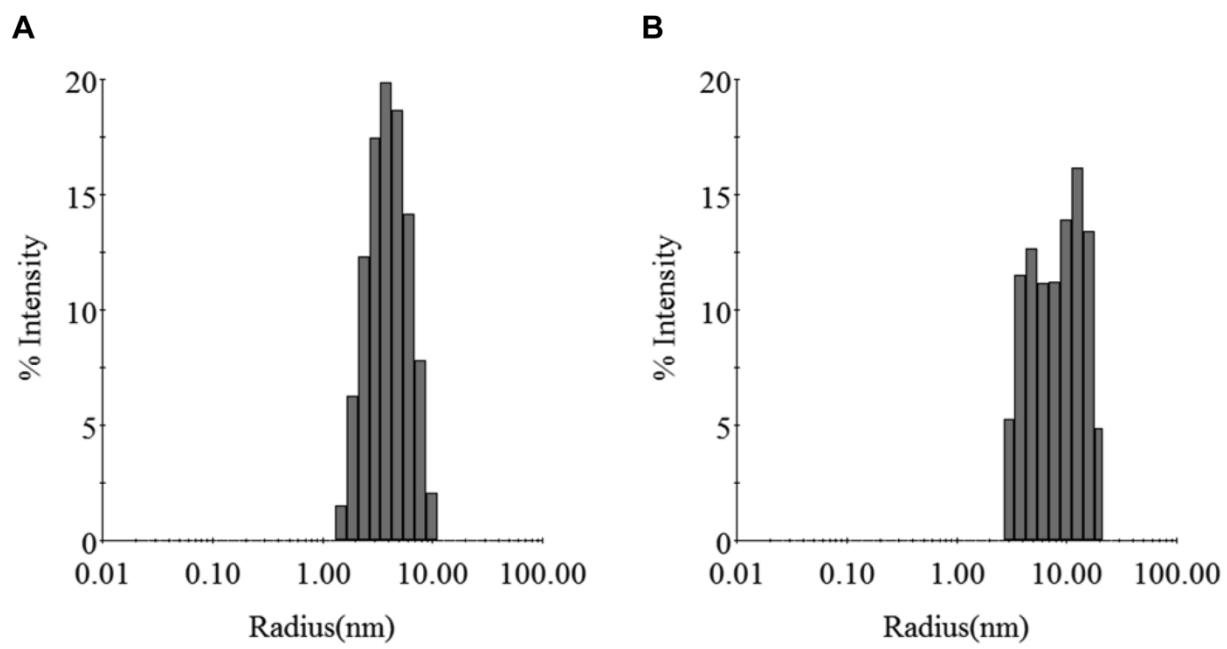


Figure S4. Size distribution of protein complexes as revealed by dynamic light scattering, comparing buffer containing DDM (negative control) and the purified FixABCX complex. (A) Buffer containing 0.02% DDM (w/v) with a determined hydrodynamic radius of 4.32 nm. (B) 1.0 mg/mL of FixABCX complex with a determined hydrodynamic radius of 4.69 nm accounted for 99.4% of the total population. Triplicate measurements yielded an average hydrodynamic radius of 5.1 ± 0.3 nm.

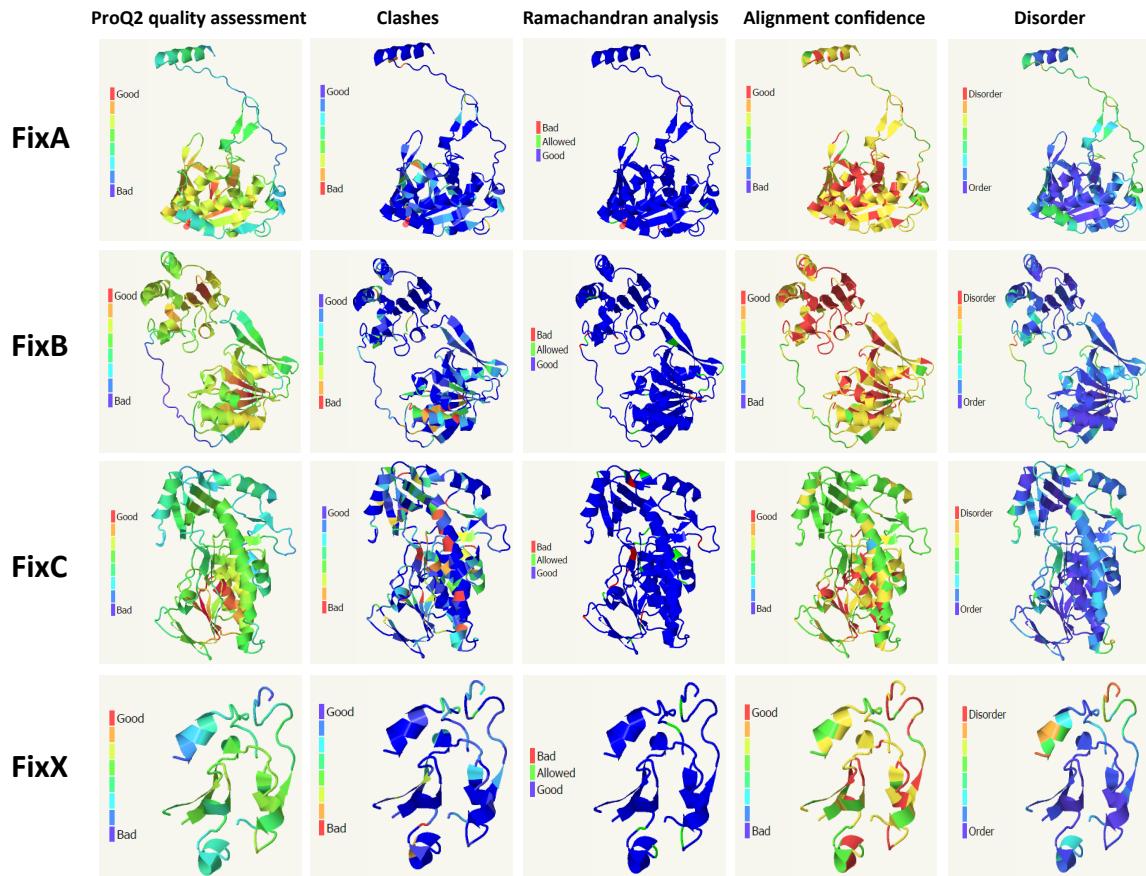


Figure S5. Evaluation of subunit homology models. *ProQ2 quality assessment*: a support vector machine is used to predict local and global quality of the model;² *Clashes*: indicates which residues lie too close to one another due to incorrect side chain and/or backbone placement; *Ramachandran analysis*: provides information on residues position (favorable/allowed/disallowed) in the model and potential problems with the backbone phi/psi angles; *Alignment confidence*: reliability of the pair-wise query-template alignment as reported by HHsearch. The confidence values are obtained from the posterior probabilities calculated in the forward-backward algorithm; *Disorder*: identifies dynamically flexible regions. This prediction has been made by the knowledge-based Disopred method (PSIPRED server).³ Overall, models for each subunit were obtained with the highest template confidence of 100% (with the exception

of 99.8% for FixX). However, it was impossible to attain complete sequence coverage, primarily due to insufficient structural information in the database. For example, sequence coverage for FixA, FixB, FixC, and FixX were 92%, 88%, 82%, and 96%, respectively.

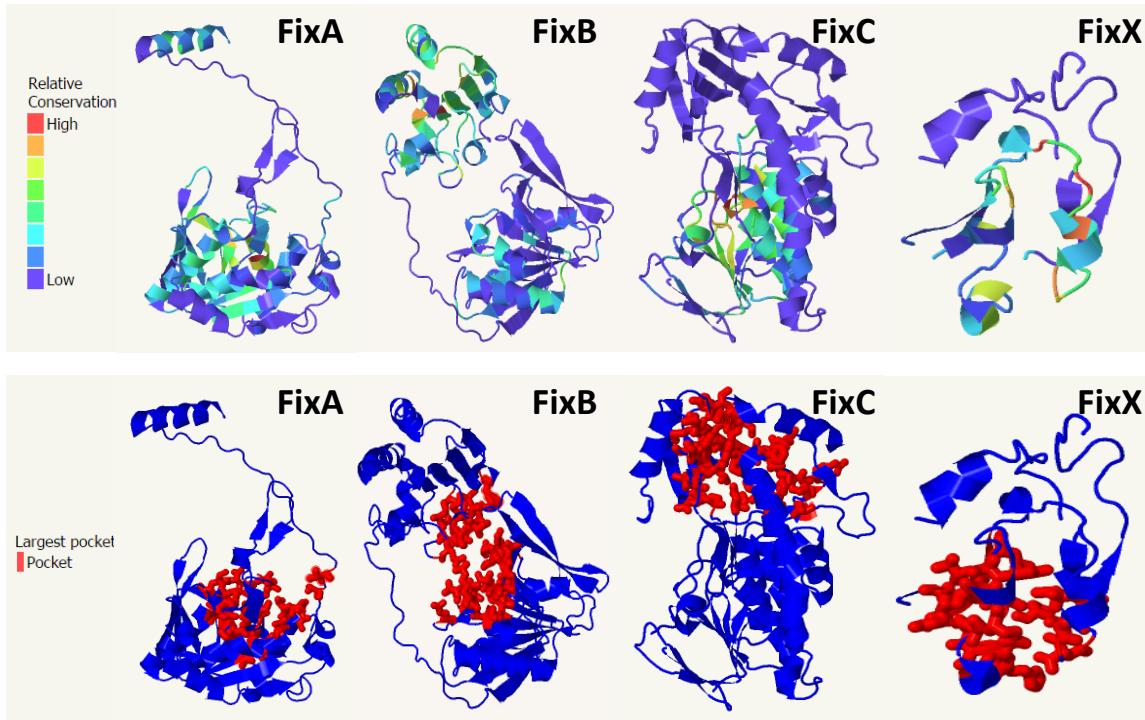


Figure S6. Predicted structural features in Fix subunits. *Conservation:* sequence analysis based on Jensen-Shannon divergence;⁴ *Largest pocket:* the pockets are detected by the fpocket2 program and highlighted in red.⁵ This provides a tool for tracking small molecule binding sites, pockets for molecular docking or detection of subpockets of conformational ensembles. The highest degree of conservation (from green to red color) is associated with cofactor binding sites. A helix (Ala338-Leu353) on the C-terminal domain of FixB appears to be critical for A and B subunit interactions. Further, large pockets are detected around FixA, FixB and FixX cofactor binding sites. In addition to flavin binding sites, two large cavities are present between the N-terminal and C-terminal domain of FixB (important for interactions with FixA) and near the predicted transmembrane region (Met299-Leu325) on FixC.

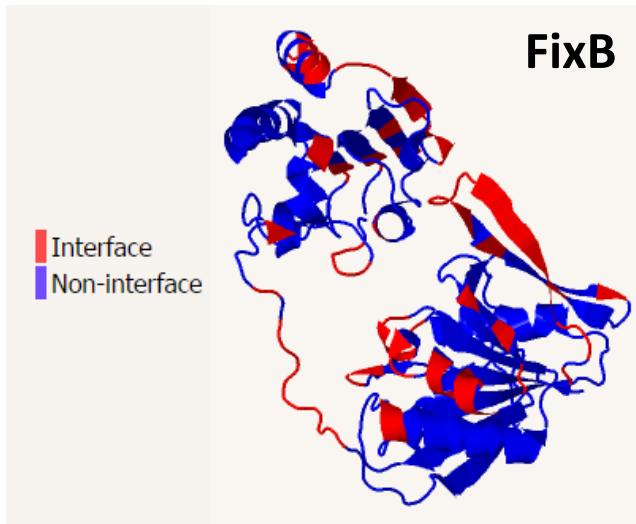


Figure S7. Protein interface. PI-site interface residues⁶ and ProtinDB interface residues provide information on prediction of protein-protein interaction sites based on information extracted from protein complexes stored in the Protein Data Bank. Interface residues are colored in red.

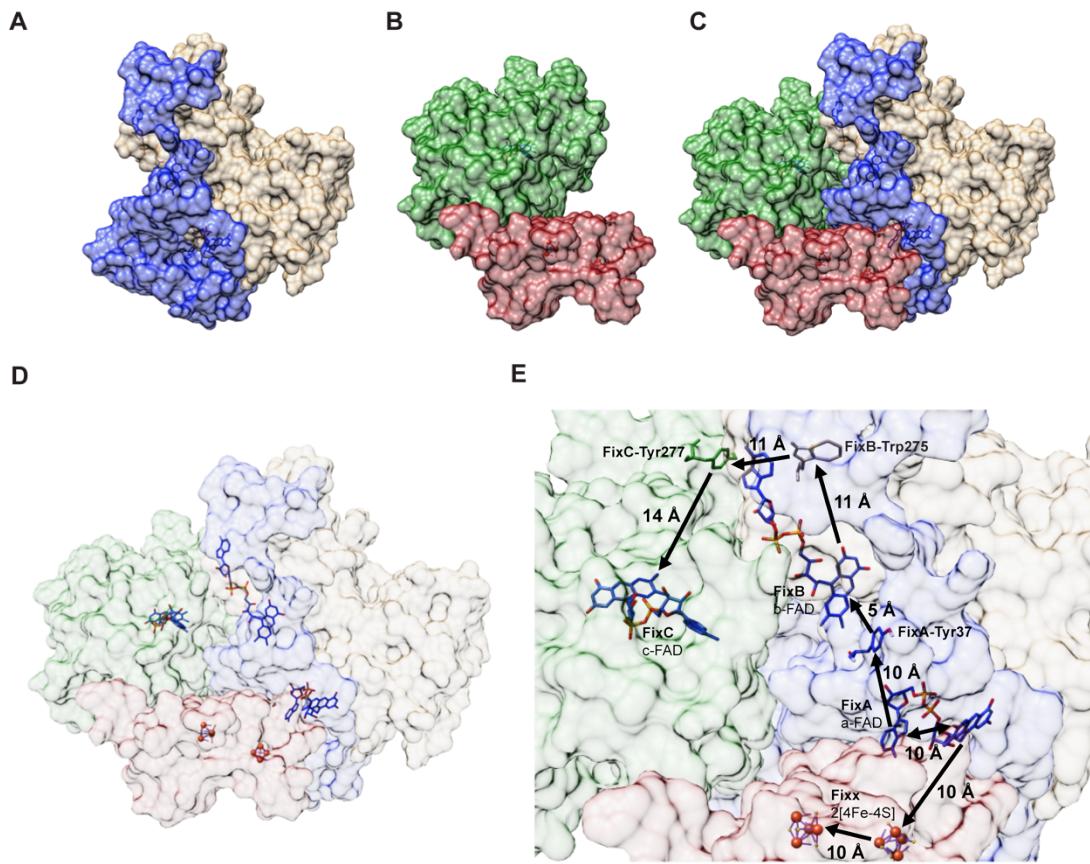


Figure S8. Structural model of FixABCX from *A. vinelandii*. Fix subunit homology models were built using Phyre2 and had the closest match to following templates from the Protein Data Bank (PDB): FixA, 4L2I; FixB, 3CLR; FixC, 4K2X; FixX, 2GMH. Chemical cross-linking data was used to produce distance constraints for generating structural models. (A) ClusPro2 docking model of FixAB subunits (visualized in Chimera). FixA in blue, FixB in tan. (B) ClusPro2 docking model of FixCX subunits (visualized in Chimera). FixC in green, FixX in red. (C) FixABCX complex model generated by docking (ClusPro2) four homology models (Phyre2). Restraints derived from the chemical cross-linking experiment. The final model was evaluated and satisfied two requirements: 1) Most of the cross-links could be explained by the model and 2) cofactor distances of <14 Å that enable electron transfer. (D) Cofactors and protein

residues that could function in electron transfer. (E) Distances between cofactors taking into account the presence of conserved Trp and Tyr side chains that may convey electrons between cofactors. FixA-Tyr37 is conserved as Tyr in 44 of the 53 FixAs of which we are aware, FixB-Trp275 is conserved as Trp in 53 out of 53 FixBs of which we are aware, and FixC-Tyr277 is Tyr in 8/8 sequences examined.

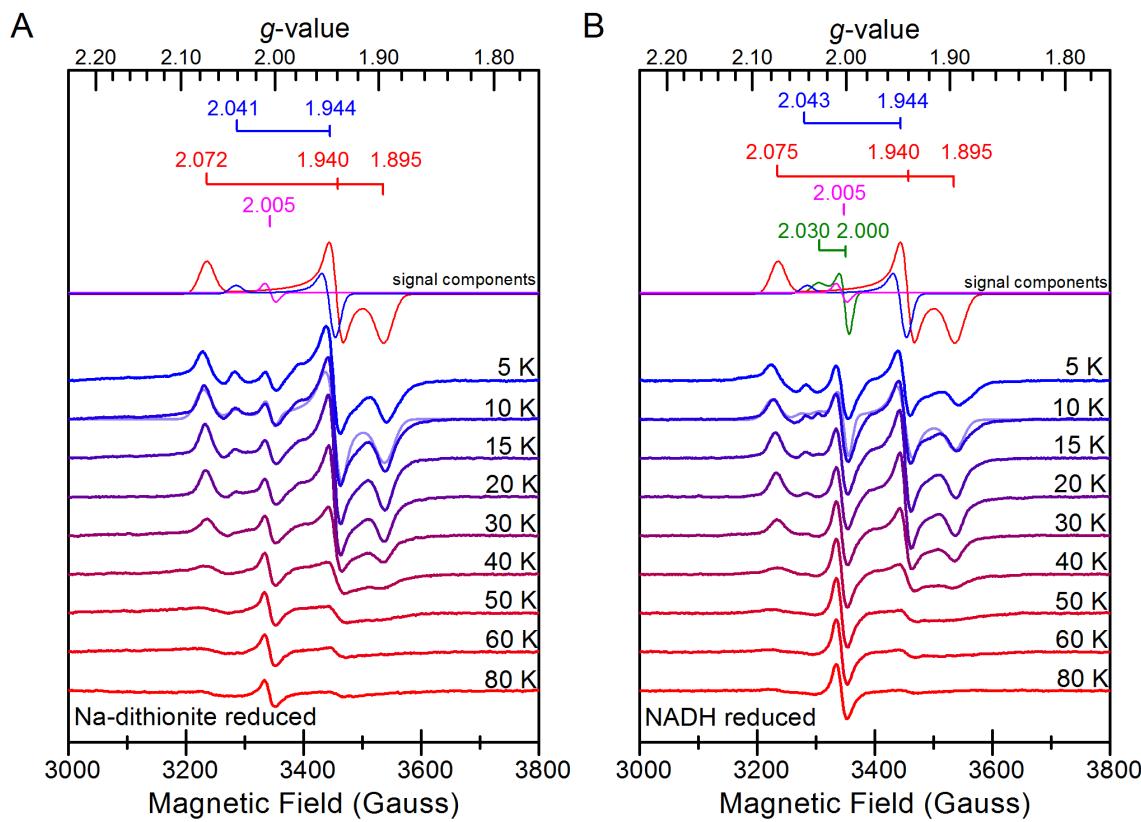


Figure S9. EPR temperature profiles of FixABCX from *A. vinelandii* reduced with either Na-dithionite or NADH. (A) FixABCX (100 μ M) reduced with Na-dithionite (10 mM). (B) FixABCX (100 μ M) reduced with NADH (1 mM). The spectra were recorded at 5, 10, 15, 20, 30, 40, 50, 60 and 80 K. Simulations of the overall spectra at 10 K are shown in the lighter shade of blue. The individual signal components and *g*-values comprising the overall simulated spectra at 10 K are shown at the top (blue, axial 2.04; red, rhombic 2.07 magenta, isotropic 2.0; green, axial 2.03) and varied slightly for each treatment (Table S4). The overall signal intensities for NADH (0.40 spins/mol protein) and Na-dithionite treatments (0.45 spins/mol protein) were similar at 10 K, indicating partial reduction of the protein complex. Microwave frequency, 9.38 GHz; microwave power, 1 mW; modulation frequency, 100 kHz; modulation amplitude, 10.0 G.

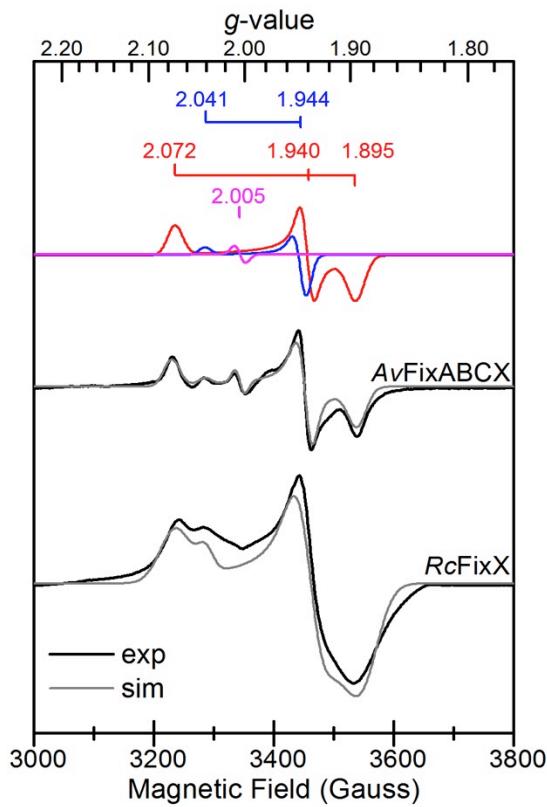


Figure S10. Comparison of EPR spectra of reduced FixABCX from *A. vinelandii* (Av) and the reduced, individual subunit FixX from *Ro. castenholzii* (Rc). FixABCX from *A. vinelandii* (100 μ M) was reduced 10 mM Na-dithionite. FixX from *Ro. castenholzii* (200 μ M) was reduced with 5 mM Na-dithionite. Signal intensities were normalized for sample concentration. Experimental (exp) spectra are colored black and simulated (sim) spectra are colored gray. The individual signal components and *g*-values comprising the overall simulated spectra are shown at the top (blue, axial 2.04; red, rhombic 2.07 magenta, isotropic 2.0; green, axial 2.03). Microwave frequency, 9.38 GHz; microwave power, 1 mW; modulation frequency, 100 kHz; modulation amplitude, 10.0 G; sample temperature, 10 K.

Supplemental information references

- (1) Merkley, E. D., Rysavy, S., Kahraman, A., Hafen, R. P., Daggett, V., and Adkins, J. N. (2014) Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* 23, 747–759.
- (2) Ray, A., Lindahl, E., and Wallner, B. (2012) Improved model quality assessment using ProQ2. *BMC Bioinformatics* 13, 224.
- (3) Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucl. Acids Res.* 41, W349–W357.
- (4) Capra, J. A., and Singh, M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics* 23, 1875–1882.
- (5) Schmidtke, P., Bidon-Chanal, A., Luque, F. J., and Barril, X. (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 27, 3276–3285.
- (6) Higurashi, M., Ishida, T., and Kinoshita, K. (2009) PiSite: a database of protein interaction sites using multiple binding states in the PDB. *Nucl. Acids Res.* 37, D360–D364.

| | | | | | | | | | | |
|----------|-------|------|-----|----|--|------|------|-----|-----|------|
| 7024.405 | 9.94 | 17.4 | 43 | 55 | IKDPDHCTECKQCTVC(57.02146)CPAGCYTRETNKG - ESI(LGVGCM(LSDFK(156.0786)QQKIS(156.0786)PYELLDR | FIXX | FIXC | 36 | 249 | 45.8 |
| 7063.375 | 7.51 | 8.9 | 40 | 34 | VLQIERGETPGSKNVQGAILYADAIEK(156.0786)IIPDFR - IK(156.0786)DPDHCTECKQCTVC(57.02146)PAGCYTR | FIXX | FIXX | 42 | 36 | 47.3 |
| 7024.405 | 9.94 | 17.4 | 44 | 54 | IKDPDHCTECK(156.0786)QCTVC(57.02146)CPAGCYTRETNKG - ESI(LGVGCM(LSDFK(156.0786)PYELLDR | FIXX | FIXC | 25 | 249 | 47.8 |
| 7063.375 | 7.22 | 3.2 | 47 | 24 | ETNGKVTLTDGCLECGTCRIC(57.02146)QDSGNLEWEWPR - AADKPFSAANLAEYKKLDDS(156.0786)FVMK | FIXX | FIXC | 54 | 338 | 49.1 |
| 4185.174 | 7.20 | 14.5 | 52 | 19 | KAREAGVLTICETTVTDLLLEGK - IKDPDHCTECK | FIXC | FIXX | 116 | 25 | 52.4 |
| 7024.405 | 9.94 | 1.9 | 59 | 39 | ETNGKVTLTDGC(57.02146)LECGTCRIC(CQDSGNLEWEWPR - GETPGSKNVQGAILYADAIEKIPDFR | FIXX | FIXC | 54 | 41 | 54.8 |
| 7024.405 | 10.04 | 1.9 | 59 | 40 | ETNGKVTLTDGC(57.02146)LECGTCRIC(CQDSGNLEWEWPR - GETPGSKNVQGAILYADAIEKIPDFR | FIXX | FIXC | 54 | 42 | 57.5 |
| 7024.405 | 9.63 | 1.9 | 59 | 36 | ETNGKVTLTDGC(57.02146)LECGTCRIC(CQDSGNLEWEWPR - GETPGSKNVQGAILYADAIEKIPDFR | FIXX | FIXC | 54 | 56 | 61.4 |
| 6949.570 | 14.44 | 15.6 | 127 | 16 | THPS(156.0786)IAPLIAGSEMKEYAAHLIPEGGYNAPIQVYGDGWMIAGDAGQFVNIGHR - VDSGRPHIR | FIXC | FIXX | 275 | 17 | 66.4 |

| | | | | | | | | |
|----------|-------|------|----|----|---|-----|-----|------|
| 6080.245 | 7.93 | 9.6 | 48 | 30 | VAAGIDTVEMIGLKGSPTVSKVFAPK(156.0786)PR - VFAPK(156.0786)PRSKRAELIESHDS(156.0786)DPK | 232 | 242 | 28.7 |
| 5500.014 | 14.69 | 17.7 | 95 | 50 | HELKWDVRVAAGIDTVEMIGLKGSPTVVS(156.0786)K - MHSVVC(57.02146)IKQVPDSAQR | 225 | 3 | 30.9 |
| 6080.255 | 9.21 | 12.0 | 48 | 43 | HELKWDVRVAAGIDTVEMIGLKGSPTVSK - PMAEAALRKCLS(156.0786)F6GADDAILVSDR | 232 | 72 | 34.1 |
| 6949.570 | 14.33 | 17.6 | 76 | 66 | VWDRVAAGIDTVEMIGLKGSPTVSKVFAPKPR - KIMEDMPV/DLIFTGKQTIDGDTAQVGPGIAK | 233 | 107 | 36.4 |
| 5500.014 | 14.07 | 10.9 | 92 | 47 | VAAGIDTVEMIGLKGSPTVSKVFAPK(156.0786)PRSK - K(156.0786)CLSGADDAILVS(156.0786)DR | 225 | 75 | 36.5 |
| 6080.255 | 9.14 | 10.0 | 59 | 31 | C(57.02146)LSFGADDAILVSDRAFGS(156.0786)DLATSVASAVIR - GSPTVSK(156.0786)VFAPKPRSK(156.0786)R | 102 | 238 | 37.3 |
| 6080.245 | 7.82 | 9.6 | 46 | 31 | VAAGIDTVEMIGLKGSPTVSKVFAPK(156.0786)PR - VFAPKPRSKRAELIES(156.0786)HDS(156.0786)DPK | 227 | 242 | 39.0 |
| 6080.245 | 8.03 | 9.6 | 49 | 30 | VAAGIDTVEMIGLKGSPTVSKVFAPK(156.0786)PR - VFAPKPRSKRAELIESHDS(156.0786)DPK | 233 | 249 | 46.0 |
| 4190.183 | 8.44 | 5.6 | 49 | 34 | LDYQLLTVVSRIVDVTAKK - SKRAELIESHDSDPK | 157 | 241 | 52.0 |
| 5500.014 | 14.88 | 2.1 | 73 | 74 | NLAEAALAKLFTQHPNLEQEIAK(156.0786)RAV - VAAGIDTVEMIGLKGSPTVVS(156.0786)K | 264 | 227 | 53.4 |
| 5500.014 | 14.88 | 2.1 | 73 | 74 | NLAEAALAKLFTQHPNLEQEIAK(156.0786)RAV - VAAGIDTVEMIGLKGSPTVVS(156.0786)K | 264 | 225 | 58.2 |
| 6080.245 | 7.72 | 9.6 | 43 | 33 | VAAGIDTVEMIGLKGSPTVSK(156.0786)VFAPK(156.0786)PR - VFAPKPRSKRAELIESHDS(156.0786)DPK | 225 | 249 | 60.1 |
| FixB | | | | | | | | |
| 6080.255 | 9.13 | 13.3 | 58 | 32 | NLAS(156.0786)TRPTFGSLLCTILTQRHRPQMATORPR - TVRPPLYIAIGIS(156.0786)GAIQHR | 166 | 292 | 21.6 |
| 4034.089 | 9.04 | 14.3 | 62 | 27 | NLASTRPTFGSLLC(57.02146)TILTOR - GKEICGEAFQHGADK | 166 | 71 | 38.8 |