Crystallographic Studies of Nitrogenase

Thesis by
Ailiena O. Maggiolo

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 24, 2023

Ailiena O. Maggiolo
ORCID: 0000-0003-1707-5060

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Doug Rees, for his invaluable mentorship and support throughout my graduate studies. I am deeply thankful for his encouragement to explore my own interests, and it has been a privilege to have been trained by such a skilled mentor. I would like to thank Bil Clemons, who has generously provided a kind and welcoming environment and consistently embodies the principles he advocates through his exemplary leadership. I thank Jim Howard for his mentorship and unwavering enthusiasm for nitrogenase, which has been an inspiration to me and many others. I also thank Harry Gray and Pamela Bjorkman for serving on my thesis committee and for their valuable insights.

Words cannot adequately express how thankful I am for the guidance, instruction, and relentless patience I received from Jens Kaiser. Jens is a wealth of knowledge and a remarkable individual. I thank Welison Floriano for his training, expertise, and inspiration to travel the world. The entire nitrogenase team is a collection of extraordinary people. It has been an absolute pleasure to work side-by-side with Rebeccah Warmack, who is clever, creative, and an exceptional scientist. Many productive years were made possible because of the diligent contributions and support by Jeff Lai. I thank him and his family and consider them important role models for life outside the lab. I thank Allen Lee, Chengcheng Fan, Renee Arias, Belinda Wenke, Kathryn Perez, Limei Zhang, Thomas Spatzal, Phong Nguyen, Naima Sharaf, Siobhán MacArdle, Stephanie Threatt, and many other Rees lab members and alumni. This thesis has benefited from the foundation laid by generations of former students and researchers who solved a multitude of nitrogenase structures. This is evident from the fact that all the structures presented herein were solved using the method of molecular replacement, which requires the utilization of a known starting model. Therefore, I am indebted to those who came before me and grateful for the opportunity to have made a modest contribution to field of structural biology related to nitrogenase.

My research would not have been possible without the dedicated support staff at Caltech. I would like to especially thank Joe Drew, Nate Siladke, and Alison Ross.

I thank the many collaborators I have had the fortune of working with over my years as a graduate student. I thank Shivansh Mahajan for many hours of stimulating

conversation and structural analysis comparing nitrogenase iron protein, Get3, and ArsA, which culminated in a review that we wrote together. This project was especially enjoyable as it brought together the Rees lab and the Clemons lab in a scientific collaboration that coincided with the physical merger of the two labs. I thank my colleague and good friend, Juner Zhang, for an exciting and fruitful collaboration with the Arnold lab. Thank you to Michelle Fry and Shyam Saladi for their encouragement, support, collaboration, and friendship. I am continually impressed by their rigor and intellect. I thank Jan Kern, Stephen Keable, and the entire team of researchers that made the XFEL experiments possible. I give special thanks to Christopher Barnes for his mentorship and for introducing me to wonderful team of structural biologists at SSRL.

I am profoundly grateful for the mentorship of Amie Boal. Amie inspired me to become a structural biologist and, among many things, taught me X-ray crystallography. There is no doubt that she changed my life. I thank Carsten Krebs and Marty Bollinger who also mentored me during my time as an undergraduate and technician and greatly influenced my trajectory. I am also grateful for the guidance and friendship of Hannah Rose, Andrew Mitchell, and Noah Dunham, Christopher Pollock, Bo Zhang, and the other members of the Boal, Bollinger, and Krebs labs. I thank Brad Rogers for years of friendship, mentorship, and support.

Some of the most cherished moments during my time in graduate school have been the experiences shared with friends and fellow members of my cohort at Caltech. I especially thank my roommates, Alex Shimozono, Sarah Bevilacqua, Travis DeLano, Veeral Shah, Morgan Abernathy (our roommate for one day!), and Felix Horns. Our time in the blue house on Cordova St. was full of unforgettable memories. Like one of those jokes. . . a structural biologist, an electrochemist, and two synthetic organic chemists live in a house. . . anything could follow. I thank all our friends and friends-of-friends who came together to enjoy food, drink, and conversation in that wonderful backyard.

I have been fortunate to have several wonderful friends from home who have continued to support me from over 2,600 miles away. I especially thank Hayley Dodds. I thank Axl LeVan for our hikes and adventures. Thank you to Adrian Borbon for his friendship, encouragement to begin computer programming, and providing fantastic resources.

I give my endless thanks and appreciation for my family, particularly my parents and my aunts. I thank the gifted artist and gardener, Marie Fox, whose mark I strive

to infuse in my scientific figures.

I reserve my deepest appreciation for my partner Felix Horns who has been a constant source of love. The hardest worker I know, he is an endlessly impressive, yet humble person. I am so thankful for him and that we have so much to look forward to.

Finally, I thank my father, Mark Maggiolo, who was the first scientist I knew. Although not formerly trained, he is a prolific reader and a true polygot, with fluency in several languages. As a musician and a farmer, he has developed a keen understanding of the natural world, in a way that many accomplished scientists will never know. Thank you, Dad.

ABSTRACT

Nitrogenase is the only enzyme known to reduce atmospheric dinitrogen to ammonia, producing a biologically available form of nitrogen. The primary component of nitrogenase, the molybdenum-iron (MoFe) protein, binds and turns over substrate after reduction by multiple electron equivalents, which are supplied by the obligate reductase, the iron (Fe) protein. Previous high-resolution X-ray crystal structures have provided pictures of the nitrogenase proteins and revealed the geometry of their metalloclusters. In this thesis, we study MoFe protein crystal isoforms and their crystal pathologies to classify isomorphous candidates suitable for crystallographic merging. We then leverage this classification to determine a high-resolution structure of the MoFe protein with improved geometric accuracy and lower coordinate error than currently available models. The reduced states of the MoFe protein are challenging to capture and therefore have been minimally characterized. We explore the structural consequences of introducing reductants and oxidants into MoFe protein crystals and study the effects of X-ray induced photoreduction on the metalloclusters. Further, we determine the radiation damage-free X-ray crystal structure of MoFe protein. The Fe protein adopts various conformational states as it functions in ATP-coupled electron transfer to the MoFe protein. We examine a set of proteins that are evolutionarily related to the Fe protein and have diverse functionalities, but retain similarity in their ATP-dependent function and allostery as a result of their conserved structural features. Finally, we characterize the structural and functional aspects of the Fe protein lacking an iron-sulfur cluster. These studies expand our understanding of the structural properties of nitrogenase and shed light on previously uncharacterized states of these proteins.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Maggiolo, A. O.*; Mahajan, S.*; Rees, D. C.; Clemons, W. M. Intradimeric Walker A ATPases: Conserved features of a functionally diverse family. *Journal of Molecular Biology* **2023**, Accepted.
*Indicates equal contributions.
A.O.M. performed structural analysis, wrote, and edited the review.

(2) Warmack, R. A.; Maggiolo, A. O.; Orta, A.; Wenke, B. B.; Howard, J. B.; Rees, D. C. Structural consequences of turnover-induced homocitrate loss in nitrogenase. *Nature Communications* **2023**, 14, 1091.
A.O.M. participated in model building, structural interpretation, and editing the manuscript.

(3) Zhang, J.; Maggiolo, A. O.; Alfonzo, E.; Mao, R.; Porter, N. J.; Abney, N. M.; Arnold, F. H. Chemodivergent C(sp^3)–H and C(sp^2)–H cyanomethylation using engineered carbene transferases. *Nature Catalysis* **2023**, 6, 152–160.
A.O.M. performed crystallography experiments, interpreted structural data, and participated in writing the manuscript.

(4) Fry, M. Y.; Najdrová, V.; Maggiolo, A. O.; Saladi, S. M.; Doležal, P.; Clemons, W. M. Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. *Nature Structural & Molecular Biology* **2022**, 29, 820–830.
A.O.M. participated in crystallographic data processing, model building, structural interpretation, and editing the manuscript.

(5) Buscagan, T. M.; Perez, K. A.; Maggiolo, A. O.; Rees, D. C.; Spatzal, T. Structural characterization of two CO Molecules bound to the nitrogenase active site. *Angewandte Chemie* **2021**, 133, 5768–5771.
A.O.M. performed experiments and participated in editing the manuscript.

(6) Saladi, S. M.; Maggiolo, A. O.; Radford, K.; Clemons, W. M. Structural biologists, let's mind our colors; preprint, *bioRxiv*, **2020**.
A.O.M. participated in writing the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Nitrogen is an essential component of living organisms as it is woven into the most fundamental biomolecules of life, such as amino acids and nucleotides. Although there is an abundance of nitrogen in Earth's atmosphere in the form of N_2 gas, it must first be "fixed" or converted to a reduced form before it can be used by most organisms. For billions of years this process of nitrogen fixation was performed in small part by the geophysical process of lightning and in large part by a special set of microorganisms, called diazotrophs.

Diazotrophs are widespread in Earth's soils and oceans. They can even be found in the planet's most specialized environments, such as within a fern's leaflet [1, 2] or inside the skeletons of coral [3] where they facilitate symbiotic growth.[4] These diverse diazotrophs comprise both archaea and bacteria and have in common a specialized set of proteins that make up the enzyme nitrogenase. Nitrogenase catalyzes nitrogen fixation by the conversion of N_2 to bio-available ammonia. Remarkably, nitrogenase stands as the only known enzyme able to activate the otherwise inert N_2 molecule, making nitrogenase integral to the biogeochemical nitrogen cycle. In this thesis, we study nitrogenase isolated from the Gram-negative soil bacterium, *Azotobacter vinelandii* (*Av*),[5] which has been established as a model system for nitrogenase.

Nitrogenase consists of two primary enzymatic components **Fig. 1**.[6, 7] These are the larger (230 kDa in *Av*) molybdenum-iron (MoFe) protein and the smaller (64 kDa in *Av*) iron (Fe) protein, which serves as the obligate reductase for the MoFe protein. Research on nitrogenases unveiled the presence of two unique metalloclusters in the MoFe protein. The first is the active site iron-molybdenum cofactor (FeMo-cofactor) where substrates bind [8, 9] and the second is the P-cluster, which relays electrons from the Fe protein to the MoFe protein for catalysis.[10–13] More recently, alternative forms of the MoFe protein have been discovered, whereby molybdenum is replaced by either vanadium or iron.[14] Our understanding of these alternative nitrogenases has greatly expanded during the time of my graduate research,[15–20] which has been a pleasure to witness; however, my work has only focused on the molybdenum-containing form.

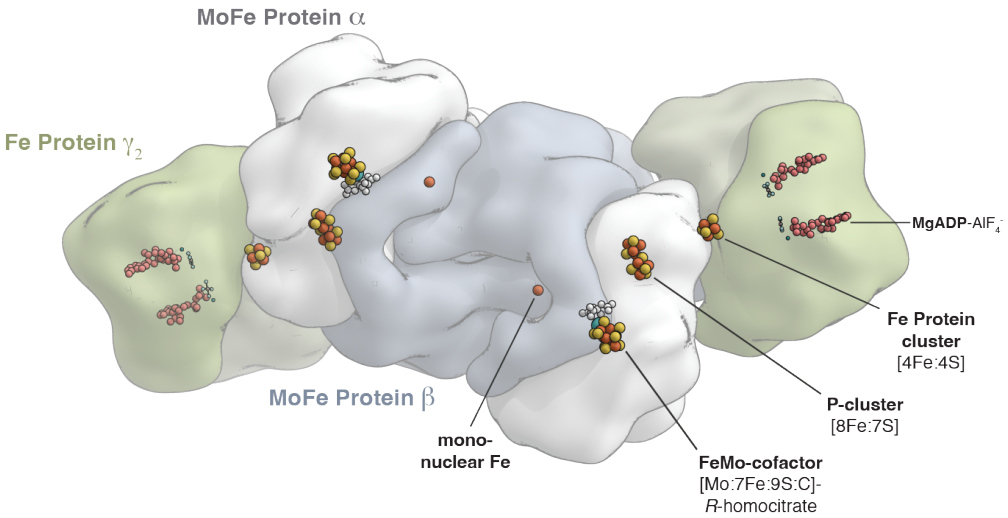


Figure 1.1: Surface representation of the $\alpha_2\beta_2$ MoFe protein heterotetramer (white and blue) complexed to two copies of the γ_2 Fe protein homodimer (green). The metalloclusters are shown as spheres and colored as follows: iron (orange), sulfur (yellow), molybdenum (teal). Homocitrate molecules are shown as white, AlF_4^- as white and gray, magnesium as green, and ADP molecules as pink spheres. This figure was generated using the coordinates from Schmid, *et al.* (PDB 1M34).[21]

The last several decades of intense spectroscopic and structural studies have revealed the nature and geometry of these metalloclusters. Some of those structural studies will be briefly discussed in this introduction. For in-depth reviews on structural studies, see Einsle & Rees (2020) [22] and Lycus *et al.* (2023).[23] For reviews on spectroscopic studies, see Hoffman *et al.* (2014) [24] and Van Stappen *et al.* (2020).[25]

There were multiple early attempts to discern the structure of the nitrogenase proteins using X-ray crystallography.[22] These studies were facilitated by the ability to obtain crystals and preliminary diffraction data.[26–30] But it was not until 1992 when Kim *et al.* published the first phased structure of the MoFe protein that a picture of nitrogenase came into view.[10, 31] The MoFe protein is composed of an $\alpha_2\beta_2$-heterotetramer of two $\alpha\beta$-heterodimers related by twofold symmetry. The genes that encode the α-subunit (*nifD*) and the β-subunit (*nifK*) have 24% amino acid sequence identity (in *Av*) and likely arose from an early gene duplication event.[32] Both the α- and β-subunits have three globular domains that each adopt a Rossmann-fold ($\beta\alpha\beta$-fold), which consist of a four or five-stranded parallel β-sheet flanked on both sides by α-helices. A key discovery from early MoFe protein structures was the

relative location of the FeMo-cofactors and P-clusters. At the time, it was known from spectroscopic and elemental studies that the MoFe protein contained two copies of the FeMo-cofactor and two copies of the P-cluster and that the FeMo-cofactors were approximately 70 Å apart.[33] However, the crystal structures revealed that the FeMo-cofactors are buried in the α-subunits and the P-clusters straddle the α/β-subunit interface.[10, 11, 31, 34] Thus, the FeMo-cofactor and P-cluster are approximately 15 Å apart. Furthermore, the structures suggested that each α/β-dimer likely functions separately from the other heterodimer. However, these initial structures were not yet high enough resolution to unambiguously determine the constituents and the geometry of the clusters.

The crystal structure of the Fe protein was determined at the same time as the MoFe protein.[35] The Fe protein is encoded by the *nifH* gene and forms a γ_2-homodimer. The two monomers are connected by a bridging [4Fe:4S] cluster, which is covalently linked by Cys97 and Cys132 (numbering from *Av*) from each monomer. The overall architecture of each subunit comprises an eight-stranded β-sheet surrounded by α-helices on both sides and harbors an N-terminal nucleotide binding site. Three loops at the ends of β-strands sit at the dimer interface between the nucleotide binding site and the [4Fe:4S] cluster. These loops adopt discrete nucleotide-dependent conformations that enable the [4Fe:4S] cluster to be directly coupled to the nucleotide binding site.[20, 35, 36] Although most iron-sulfur proteins have buried clusters, the [4Fe:4S] cluster of the Fe protein is surface exposed, making it highly sensitive to degradation by oxygen and accessible to chemical chelators. The surface exposure of the cluster enables efficient electron transfer across a protein-protein interface, which became apparent from subsequent structures of the Fe protein complexed to MoFe protein.

The Fe protein transiently binds the MoFe protein to deliver electrons sequestered from endogenous sources within the cell.[37] As this electron delivery step is coupled to ATP-hydrolysis, the Fe protein could be stably complexed to MoFe protein by introducing the ATP transition-state analogue, ADP-AlF_4^-.[38] Structures using ATP-analogues for this purpose revealed two Fe protein dimers bound to the MoFe protein, closest to the P-cluster.[21, 36] Remarkably, the [4Fe:4S] cluster of the Fe protein was shifted 4 Å towards the MoFe protein, putting the [4Fe:4S] cluster and the P-cluster only 16 Å apart. Subsequent complexed structures revealed Fe protein binding modes that are choreographed with nucleotide states.[39–43]

High resolution X-ray crystal structures of the nitrogenase MoFe protein have revealed a wealth of information. The original structures at 2.7 Å resolution [10, 11] were subsequently improved to 2.2 Å resolution [34] and then 2.0 Å resolution,[44] which enabled the proper assignment of the P-cluster as an [8Fe:7S] fused double-cubane cluster that undergoes redox-dependent conformational changes. However, a watershed moment for the field came with the 1.16 Å resolution structure of the *Av* MoFe protein.[45] This structure revealed a central light atom encased within the trigonal prismatic core the FeMo-cofactor. It took another 10 years of study and a seemingly modest increase in resolution to 1.0 Å before the identity of the central atom was determined.[46] The most likely candidates for the central atom were nitrogen, oxygen, or carbon. Ultimately, a combination of Fe Kβ X-ray emission spectroscopy,[47] biochemical studies of the precursor clusters and biosynthetic machinery,[48] ^{13}C-labeled MoFe protein probed by ESEEM electron paramagnetic resonance,[46] and the 1.0 Å resolution crystal structure definitively established the mystery atom as a central carbide. Thus, the constituents of the FeMo-cofactor are [Mo:7Fe:9S:C]-*R*-homocitrate.

Drawing on the wealth of knowledge derived from the 1.16 and 1.0 Å resolution MoFe protein structures, one can envision that profound insights await discovery through further improvement in resolution. Some outstanding questions involve the number and locations of hydrogens throughout the structure, particularly around the cofactor. In **Chapter 2**, we describe the analysis of *Av* MoFe protein crystals and their crystal pathologies to classify common crystal isoforms. We leveraged this classification to identify the most isomorphous datasets as candidates for crystallographic merging. Although each MoFe protein crystallographic dataset was near 1.0 Å resolution individually, this computational analysis and merging procedure improved the electron density maps. The merged dataset of 50 million observations had higher redundancy and yielded a model with lower coordinate error than that of the previously published 1.0 Å resolution structure. Further, the resultant electron density maps contained information about the locations of some hydrogens throughout the structure, including in important regions such as the water pool around homocitrate. This improved structure could be used in future studies to explore mechanisms of nitrogenase through computation modeling.

In **Chapter 3**, we explore the effects of chemical reductants, oxidants, and X-ray induced photoreduction on MoFe protein crystals. We found that the potent reductants, Eu(II)-DTPA and Ti(III)-citrate, convert the MoFe protein P-cluster

from the closed (P^N) conformation to the open (P^{2+}) conformation. We identified radiation damage effects on MoFe protein crystals from extended X-ray irradiation using a combination of diffraction and X-ray absorption to monitor damage levels. Further, we obtained X-ray radiation damage-free structures of the MoFe protein using an X-ray free electron laser. This damage-free structure of resting state MoFe protein provides a foundation for future work to capture structures of activated MoFe protein.

An analysis of the Fe protein dimer dynamics and its connection to ancestrally related Walker A proteins is discussed in **Chapter 4**. In **Chapter 5** we present the structure of the Fe protein lacking the [4Fe:4S] cluster and characterize the Fe protein in this apo form. Taken together, these studies expand our understanding of nitrogenase proteins and the structural states that they adopt.

Finally, in **Chapter 6**, we structurally characterize an engineered cytochrome P450 that is capable of new-to-nature C(sp^3)-H cyanomethylation. Through our investigation of its structure, we identify modifications in the active site that affect the shape and electrostatics of the substrate binding pocket. These observed alterations likely account for the distinctive reactivity exhibited by this engineered P450 enzyme in comparison to the native enzyme and its closely related counterparts.

References

(1) Temmink, R. J. M.; Harpenslager, S. F.; Smolders, A. J. P.; van Dijk, G.; Peters, R. C. J. H.; Lamers, L. P. M.; van Kempen, M. M. L. Azolla along a phosphorus gradient: biphasic growth response linked to diazotroph traits and phosphorus-induced iron chlorosis. *Scientific Reports* **2018**, *8*, 4451.

(2) Baker, J. A.; Entsch, B.; McKay, D. B. The cyanobiont in an *Azolla* fern is neither *Anabaena* nor *Nostoc*. *FEMS Microbiology Letters* **2003**, *229*, 43–47.

(3) Moynihan, M. A.; Goodkin, N. F.; Morgan, K. M.; Kho, P. Y. Y.; Lopes Dos Santos, A.; Lauro, F. M.; Baker, D. M.; Martin, P. Coral-associated nitrogen fixation rates and diazotrophic diversity on a nutrient-replete equatorial reef. *The ISME Journal* **2022**, *16*, 233–246.

(4) Smil, V., *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production*; The MIT Press: 2000.

(5) Kennedy, C.; Rudnick, P.; MacDonald, M.; Melton, T., *Bergey's manual of systematic bacteriology*, Garrity G. M.; Genus Azotobacter part B, Vol. 2; Spring Verlag, New York, NY: 2005; 384-401.

(6) Bulen, W. A.; LeComte, J. R. The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proceedings of the National Academy of Sciences* **1966**, *56*, 979–986.

(7) Mortenson, L. Components of cell-free extracts of clostridium pasteurianum required for ATP-dependent H2 evolution from dithionite and for N2 fixation. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1966**, *127*, 18–25.

(8) Mortenson, L. E. FERREDOXIN AND ATP, REQUIREMENTS FOR NITROGEN FIXATION IN CELL-FREE EXTRACTS OF CLOSTRIDIUM PASTEURIANUM. *Proceedings of the National Academy of Sciences* **1964**, *52*, 272–279.

(9) Burgess, B. K. The iron-molybdenum cofactor of nitrogenase. *Chemical Reviews* **1990**, *90*, 1377–1406.

(10) Kim, J.; Rees, D. C. Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from azotobacter vinelandii. *Nature* **1992**, *360*, 553–560.

(11) Kim, J.; Rees, D. C. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science (New York, N.Y.)* **1992**, *257*, 1677–1682.

(12) Peters, J. W.; Fisher, K.; Newton, W. E.; Dean, D. R. Involvement of the P Cluster in Intramolecular Electron Transfer within the Nitrogenase MoFe Protein. *Journal of Biological Chemistry* **1995**, *270*, 27007–27013.

(13) Lowe, D. J.; Fisher, K.; Thorneley, R. N. F. *Klebsiella pneumoniae* nitrogenase: pre-steady-state absorbance changes show that redox changes occur in the MoFe protein that depend on substrate and component protein ratio; a role for P-centres in reducing dinitrogen? *Biochemical Journal* **1993**, *292*, 93–98.

(14) Eady, R. R. Structure-Function Relationships of Alternative Nitrogenases. *Chemical Reviews* **1996**, *96*, 3013–3030.

(15) Sippel, D.; Einsle, O. The structure of vanadium nitrogenase reveals an unusual bridging ligand. *Nature Chemical Biology* **2017**, *13*, 956–960.

(16) Sippel, D.; Rohde, M.; Netzer, J.; Trncik, C.; Gies, J.; Grunau, K.; Djurdjevic, I.; Decamps, L.; Andrade, S. L. A.; Einsle, O. A bound reaction intermediate sheds light on the mechanism of nitrogenase. *Science* **2018**, *359*, 1484–1489.

(17) Rohde, M.; Trncik, C.; Sippel, D.; Gerhardt, S.; Einsle, O. Crystal structure of VnfH, the iron protein component of vanadium nitrogenase. *JBIC Journal of Biological Inorganic Chemistry* **2018**, *23*, 1049–1056.

(18) Trncik, C.; Müller, T.; Franke, P.; Einsle, O. Structural analysis of the reductase component AnfH of iron-only nitrogenase from Azotobacter vinelandii. *Journal of Inorganic Biochemistry* **2022**, *227*, 111690.

(19) Trncik, C.; Detemple, F.; Einsle, O. Iron-only Fe-nitrogenase underscores common catalytic principles in biological nitrogen fixation. *Nature Catalysis* **2023**, DOI: 10.1038/s41929-023-00952-1.

(20) Schmidt, F. V.; Schulz, L.; Zarzycki, J.; Oehlmann, N. N.; Prinz, S.; Erb, T. J.; Rebelein, J. G. *Structural Insights into the Iron Nitrogenase Complex*; preprint; Biochemistry, 2023.

(21) Schmid, B.; Einsle, O.; Chiu, H.-J.; Willing, A.; Yoshida, M.; Howard, J. B.; Rees, D. C. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF_4^--Stabilized Structure † ‡. *Biochemistry* **2002**, *41*, 15557–15565.

(22) Einsle, O.; Rees, D. C. Structural Enzymology of Nitrogenase Enzymes. *Chemical Reviews* **2020**, *120*, 4969–5004.

(23) Lycus, P.; Einsle, O.; Zhang, L. Structural biology of proteins involved in nitrogen cycling. *Current Opinion in Chemical Biology* **2023**, *74*, 102278.

(24) Hoffman, B. M.; Lukoyanov, D.; Yang, Z.-Y.; Dean, D. R.; Seefeldt, L. C. Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical Reviews* **2014**, *114*, 4041–4062.

(25) Van Stappen, C.; Decamps, L.; Cutsail, G. E.; Bjornsson, R.; Henthorn, J. T.; Birrell, J. A.; DeBeer, S. The Spectroscopy of Nitrogenases. *Chemical Reviews* **2020**, *120*, 5005–5081.

(26) Burns, R.; Holsten, R.; Hardy, R. Isolation by crystallization of the MoFe protein of Azotobacter nitrogenase. *Biochemical and Biophysical Research Communications* **1970**, *39*, 90–99.

(27) Burgess, B. K.; Jacobs, D. B.; Stiefel, E. I. Large-scale purification of high activity Azotobacter vinelandii nitrogenase. *Biochimica et Biophysica Acta (BBA) - Enzymology* **1980**, *614*, 196–209.

(28) Weininger, M. S.; Mortenson, L. E. Crystallographic properties of the MoFe proteins of nitrogenase from Clostridium pasteurianum and Azotobacter vinelandii. *Proceedings of the National Academy of Sciences* **1982**, *79*, 378–380.

(29) Yamane, T.; Weininger, M. S.; Mortenson, L. E.; Rossmann, M. G. Molecular symmetry of the MoFe protein of nitrogenase. Structural homology/nitrogen fixation/x-ray crystallography. *The Journal of Biological Chemistry* **1982**, *257*, 1221–1223.

(30) Rees, D. C.; Howard, J. B. Crystallization of the Azotobacter vinelandii nitrogenase iron protein. *The Journal of Biological Chemistry* **1983**, *258*, 12733–12734.

(31) Kim, J.; Woo, D.; Rees, D. C. X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. *Biochemistry* **1993**, *32*, 7104–7115.

(32) Fani, R.; Gallo, R.; Liò, P. Molecular Evolution of Nitrogen Fixation: The Evolutionary History of the nifD, nifK, nifE, and nifN Genes. *Journal of Molecular Evolution* **2000**, *51*, 1–11.

(33) Bolin, J. T.; Ronco, A. E.; Morgan, T. V.; Mortenson, L. E.; Xuong, N. H. The unusual metal clusters of nitrogenase: structural features revealed by x-ray anomalous diffraction studies of the MoFe protein from Clostridium pasteurianum. *Proceedings of the National Academy of Sciences* **1993**, *90*, 1078–1082.

(34) Chan, M. K.; Kim, J.; Rees, D. C. The Nitrogenase FeMo-Cofactor and P-Cluster pair: 2.2 Å Resolution Structures. *Science* **1993**, *260*, 792–794.

(35) Georgiadis, M. M.; Komiya, H.; Chakrabarti, P.; Kornuc, J. J.; Rees, D. C. Crystallographic Structure of the Nitrogenase Iron Protein from Azo tobacter vinelandli. **1992**, *257*, 8.

(36) Schindelin, H.; Kisker, C.; Schlessman, J. L.; Howard, J. B.; Rees, D. C. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. *Nature* **1997**, *387*, 370–376.

(37) Hageman, R. V.; Burris, R. H. Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. *Proceedings of the National Academy of Sciences* **1978**, *75*, 2699–2702.

(38) Renner, K. A.; Howard, J. B. Aluminum Fluoride Inhibition of Nitrogenase: Stabilization of a Nucleotide·Fe-Protein·MoFe-Protein Complex. *Biochemistry* **1996**, *35*, 5353–5358.

(39) Rutledge, H. L.; Tezcan, F. A. Electron Transfer in Nitrogenase. *Chemical Reviews* **2020**, *120*, 5158–5193.

(40) Chiu, H.-J.; Peters, J. W.; Lanzilotta, W. N.; Ryle, M. J.; Seefeldt, L. C.; Howard, J. B.; Rees, D. C. MgATP-Bound and Nucleotide-Free Structures of a Nitrogenase Protein Complex between the Leu 127Δ-Fe-Protein and the MoFe-Protein . *Biochemistry* **2001**, *40*, 641–650.

(41) Tezcan, F. A. Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein. *Science* **2005**, *309*, 1377–1380.

(42) Tezcan, F. A.; Kaiser, J. T.; Howard, J. B.; Rees, D. C. Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase. *Journal of the American Chemical Society* **2015**, *137*, 146–149.

(43) Rutledge, H. L.; Cook, B. D.; Nguyen, H. P. M.; Herzik, M. A.; Tezcan, F. A. Structures of the nitrogenase complex prepared under catalytic turnover conditions. *Science* **2022**, *377*, 865–869.

(44) Peters, J. W.; Stowell, M. H. B.; Soltis, S. M.; Finnegan, M. G.; Johnson, M. K.; Rees, D. C. Redox-Dependent Structural Changes in the Nitrogenase P-Cluster † ‡. *Biochemistry* **1997**, *36*, 1181–1187.

(45) Einsle, O.; Tezcan, F. A.; Andrade, S. L. A.; Schmid, B.; Yoshida, M.; Howard, J. B.; Rees, D. C. Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science* **2002**, *297*, 1696–1700.

(46) Spatzal, T.; Aksoyoglu, M.; Zhang, L.; Andrade, S. L. A.; Schleicher, E.; Weber, S.; Rees, D. C.; Einsle, O. Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science* **2011**, *334*, 940–940.

(47) Lancaster, K. M.; Roemelt, M.; Ettenhuber, P.; Hu, Y.; Ribbe, M. W.; Neese, F.; Bergmann, U.; DeBeer, S. X-ray Emission Spectroscopy Evidences a Central Carbon in the Nitrogenase Iron-Molybdenum Cofactor. *Science* **2011**, *334*, 974–977.

(48) Wiig, J. A.; Hu, Y.; Lee, C. C.; Ribbe, M. W. Radical SAM-Dependent Carbon Insertion into the Nitrogenase M-Cluster. *Science* **2012**, *337*, 1672–1675.

Chapter 2

VARIABLE UNIT CELL ANALYSIS AND HIGH-RESOLUTION STRUCTURE DETERMINATION OF *AV* MOFE PROTEIN CRYSTALS

2.1 Introduction

The proposed mechanisms of N_2 reduction by nitrogenase are founded on models that infer various protonation states of N_2, its reduced intermediates, and the FeMoco during catalysis. However, measurements of the protonation states of the FeMoco are challenging due to the large combination of possible electronic and geometric states that the FeMoco can access. Although the FeMoco is likely unprotonated in the resting state,[1] FeMoco may become protonated in its more reduced states during catalysis.[1–3] During catalysis, hydrogens can be extracted from neighboring residues, homocitrate, or a nearby reservoir of water molecules. Thus, there is considerable interest in mapping their locations as this may influence the catalytic mechanism. In theory, this can be achieved by solving very high-resolution structures of the MoFe protein. However, in practice, even a 1 Å resolution structure [4] of the resting state protein has not provided definitive evidence for all of the hydrogen locations. The visualization of hydrogens is further complicated by Fourier series termination ripples that occur around electron rich metal centers and obscure the density for light atoms.[5]

As we have obtained several crystallographic datasets near, but not significantly below 1 Å resolution, we focused on computational efforts to combine these datasets to increase the data quality. This chapter discusses an analysis of *Azotobacter vinelandii* MoFe protein crystals and their crystal defects, which led us to systematically characterize the various crystal isoforms. We leveraged the classification of these variations to identify the most isomorphous datasets as candidates for crystallographic merging. A single merged dataset originating from 6 crystals and 50 million observations gave rise to a high data-to-parameter ratio X-ray crystal structure of the MoFe protein. Although this dataset is ostensibly the same resolution as the best MoFe protein crystal structure published to date (the 1 Å resolution PDB 3U7Q)[4], our merged dataset that is derived from more diffraction data has higher redundancy (34.4 versus 6.4 for 3U7Q) and the resultant model has lower coordinate

error (0.018 versus 0.027 for 3U7Q) than that of PDB 3U7Q. Our merged dataset has electron difference density for hydrogens around the most ordered regions of the protein and some difference density for hydrogens around the active site, homocitrate, and active site water pool. The noise level and reliability of the difference maps around these important regions in the active site are discussed herein. Altogether, this variable unit cell classification algorithm is applicable where many nearly isomorphous crystallographic datasets are collected on a single protein target, which is the case for the *Av* MoFe protein. Further, the resultant electron density maps that contain information of the location of some hydrogens may serve as the basis for improved models for computational analysis on the mechanism of nitrogen reduction.

2.2 Results

Variable unit cell parameters of *Av* MoFe protein crystals

Many *Av* MoFe protein crystals are plate-shaped and adopt a monoclinic unit cell. Monoclinic unit cells have three unequal cell lengths (a, b, and c) and one interaxial angle (β) that is not equal to 90°. Although it is not unusual for crystals to have slight variations in their unit cell parameters, I found that my MoFe protein crystals had relatively large variations in their unit cell parameters, while remaining in the $P2_1$ space group. For example, two crystals listed as set 1 and set 2 in **Table 1**, have a-axis lengths that differ by ~6%, amounting to a roughly 5 Å difference in the total axis lengths (76.3 versus 80.9 Å). Furthermore, some crystals such as set 3 in **Table 1** had unit cell parameters with double the typical a-axis length, being roughly 160 Å in one direction rather than 80 Å. These crystals with large variations in their cell parameters would sometimes come from the same crystal drop, treated with the same cryoprotectant, and had no differences in their visual appearance.

To learn whether this was a property of just my crystals or whether this variation in unit cells extends to MoFe protein crystals made by other members of our lab and other labs, I looked at the unit cell parameters of previously published *Av* MoFe protein crystals, several of which are listed in **Table 1**. From these selected examples, it appears that other researchers also crystallize MoFe protein that adopt a variety of unit cell parameters (sets 4-12). These parameters are not resolution dependent, as exhibited by sets 4 and 9, which only differ in their a-axis lengths by 0.5%. Other individual researchers also obtain crystals with multiple unit cell parameters. For example, the crystals from sets 5 and 6 that were generated by the same crystallographer under the same crystallization condition have ~5% differences

Table 2.1: Selected MoFe protein crystals and their unit cell parameters.

Set	Name	Unit Cell	Res	PDB	Reference
1	*Av* MoFe "short"	76.3, 127.5, 107.0, β =109.1˚	1.5	-	This work
2	*Av* MoFe "long"	80.9, 130.9, 107.1, β =110.6 ˚	1.3	-	This work
3	*Av* MoFe "doubled"	107.1, 130.4, 159.6, β =108˚	1.1	-	This work
4	*Av* MoFe high-res	81.2, 130.7, 107.2, β = 110.7˚	1.00	3U7Q	Spatzal, *et al.* Science **2011**
5	*Av* MoFe Se, CO	77.2, 128.0, 107.6, β = 109.2˚	1.7	6OP1	Henthorn, *et al.* J.A.C.S. **2019**
6	*Av* MoFe Lo [Se]	80.9, 131.0, 107.0, β = 110.˚	1.6	6OP3	Henthorn, *et al.* J.A.C.S. **2019**
7	*Av* MoFe high pH	76.5, 127.9, 107.1, β = 108.9˚	2.0	4ND8	Yang, *et al.* Biochem. **2014**
8	*Av* MoFe F99Y/S188A	76.7, 128.6, 107.5, β = 108.9˚	1.9	6O7O	Rutledge, *et al.* J.A.C.S. **2019**
9	*Av* MoFe variant	80.8 130.8, 108.1, β = 111.1˚	2.0	4XPI	Danyal, *et al.* Biochem. **2015**
10	*Av* MoFe Q96X	76.8, 128.3, 107.2, β = 109.1˚	1.7	6BBL	Zadvornyy, *et al.* J. Inorg. Biochem. **2017**
11	*Av* MoFe	108.3, 131.6, 159.2, β = 108.4˚	1.16	1M1N	Einsle, *et al.* Science **2002**
12	*Av* MoFe	108.4, 130.5, 81.5, β = 110.8˚	2.20	1MIN	Kim, Thesis Dissertation (Caltech) **1994**

***all $P2_1$*

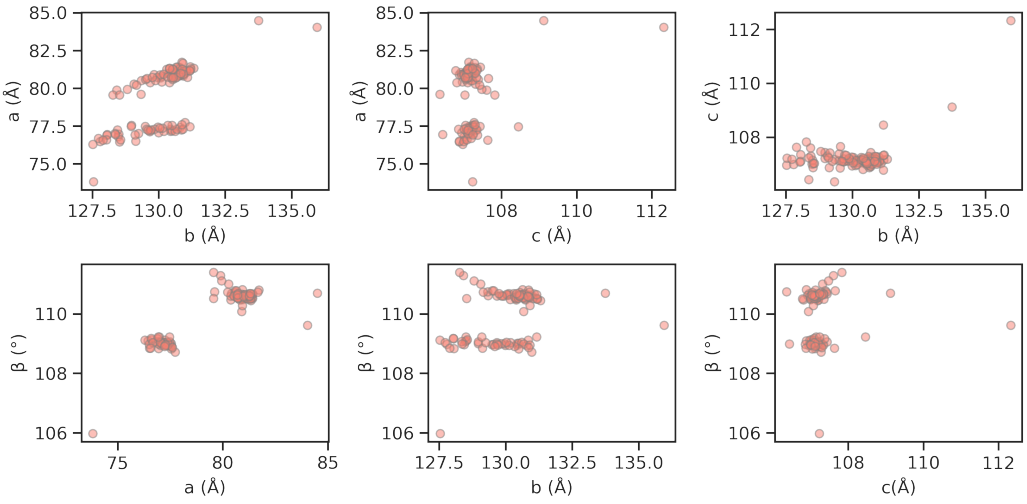


Figure 2.1: Correlation plots of the unit cell parameters (a, b, c, and β) from MoFe protein crystals. For simplicity, only ~100 datasets are shown. The unit cell parameters cluster into two distinct populations.

in their a-axis lengths (77.2 versus 80.9 Å). Furthermore, the variable unit cell parameters are not dependent on the MoFe protein preparation, as demonstrated by MoFe protein inhibited states (sets 5 and 7) and sequence variants (sets 8-9). Thus, it is possible that the variable $P2_1$ unit cell parameters are general phenomenon of

Av MoFe protein crystals.

The variations in the unit cell parameters result in seemingly insignificant differences in the final structure solution of the MoFe protein. More specifically, the resultant coordinates from sets 1 and 2 overlay with a root mean standard deviation of 0.16 Å. However, the unit cell parameters were not isomorphous enough for successful crystallographic merging. Thus, we systematically characterized several hundred MoFe protein crystal diffraction datasets to classify and cluster them into the most isomorphous unit cells.

From the correlations between each MoFe protein crystal unit cell parameter, we found that the unit cells clustered into two distinct populations **(Fig. 1)**. These two populations were generally distributed into a cluster of unit cells with either a "short" a-axis (colored blue in **Fig. 2**) with an average length of 77.1 +/- 0.05 Å or a "long" a-axis (colored orange in **Fig. 2**) with an average length of 80.9 +/- 0.05 Å.

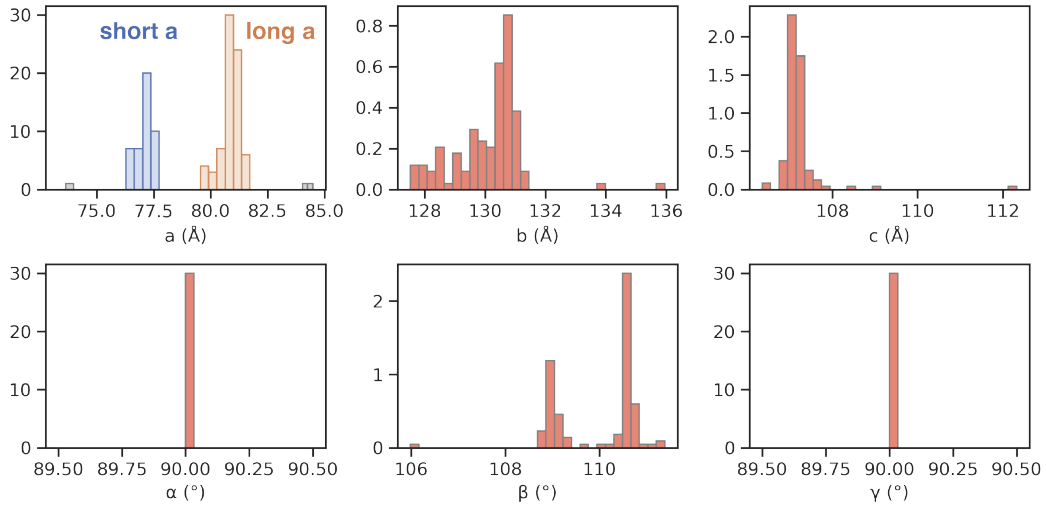


Figure 2.2: Distributions of each unit cell parameter (a, b, c, α, β, and γ) from the same MoFe protein crystals as shown in **Figure 1**. The a-axis is distributed into two populations based on their a-axis lengths, designated "short a" (blue) and "long a" (orange). The α and γ angles are uniformly distributed at 90°, as all the datasets are from monoclinic $P2_1$ crystals.

The pairwise correlations and distributions of these two populations between each unit cell parameter are shown in **Fig. 3A**, where the unit cells with a short a-axis are colored in blue and the unit cells with a long a-axis are colored in orange. To discern the structural differences between members of these clusters, we selected three datasets that were on the extreme ends of each cluster for further analysis.

Since all the c-axis lengths were uniformly distributed as a single species with an average length of 107.2 +/- 0.05 Å, we selected candidates based only on the differences in their a- and b-axes lengths. The first dataset we chose has long a- and long b-axes (80.9, 130.9, 107.1 Å and β = 110.6°) (teal hexagon in **Fig. 3A**).

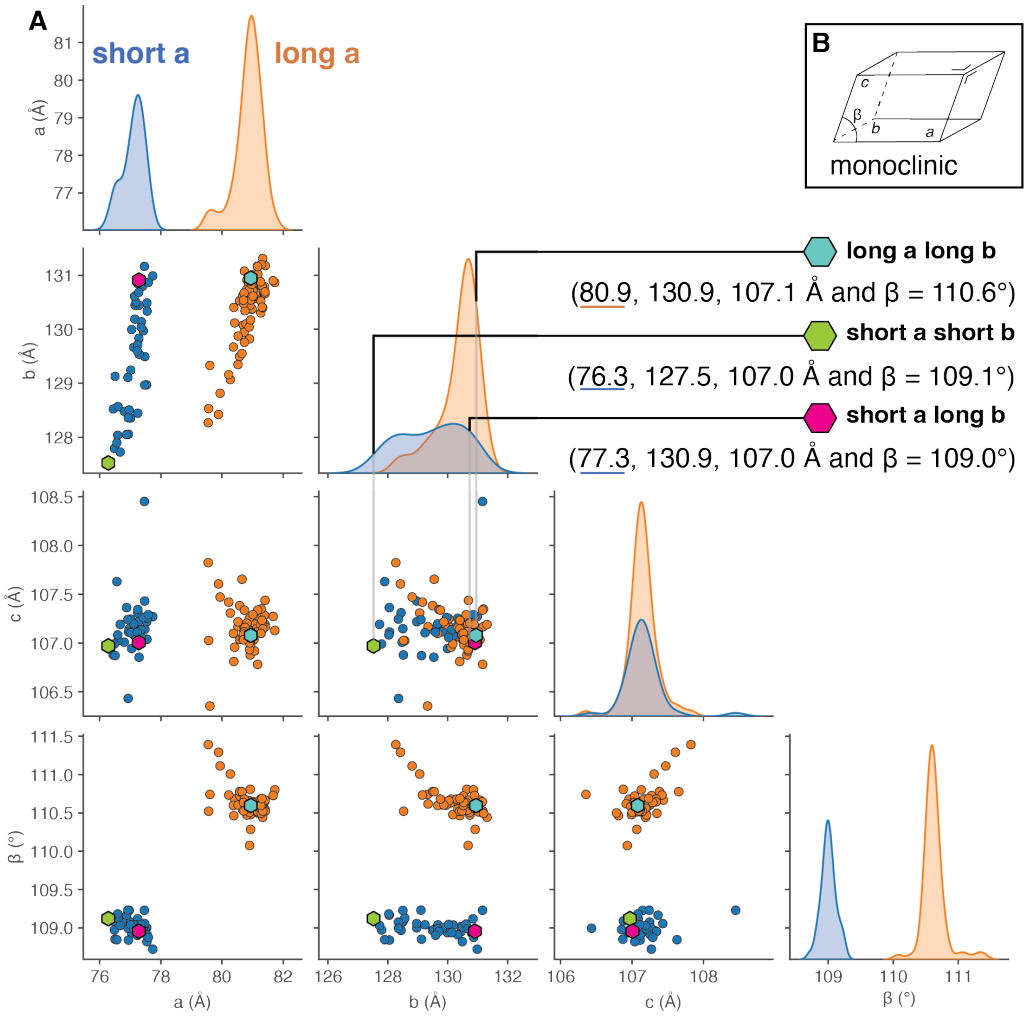


Figure 2.3: **A)** Pairplot of the pairwise relationship between and distributions of the unit cell parameters of the same MoFe protein crystals as shown in Figure 1, colored by whether they contain a short (blue) or long (orange) a-axis. All the unit cells have a normally distributed c-axis. The unit cells with long b-axes generally also have long a-axes. The unit cells with short a-axes are bimodally distributed in their b-axes lengths. Thus, we selected three datasets for further analysis to compare their crystallographic packing arrangements with: 1) long a and long b (teal hexagon), 2) short a and short b (green hexagon), and 3) short a and long b (pink hexagon) axes lengths. **B)** Schematic of a monoclinic unit cell with sides a, b, and c and angle β. Angles α and γ (not shown) are 90°.

Because the unit cells with short a-axis lengths were bimodally distributed in their b-axis lengths, we chose two datasets with short a-axis lengths: one that had both a short a- and short b-axis (76.3, 127.5, 107.0 Å and β = 109.1°) (green hexagon in **Fig. 3A**) and one that had a short a- and long b-axis (77.3, 130.9, 107.0 Å and β = 109.0°) (pink hexagon in **Fig. 3A**).

We used these three representative datasets to identify differences in the solved structures from each unit cell parameter cluster and to analyze their crystal packings. Each dataset was processed using the following methods: the data were indexed, integrated, and scaled using HKL3000.[6] We obtained phases by molecular replacement with CCP4 PHASER [7] using only the protein atoms of PDB 3U7Q. Model building was performed using Coot [8] and further refined using REFMAC5 [9] and Phenix [10]. The three resultant structures had an all-atom super position root mean standard deviation of ~0.15 Å, which was consistent with there being minimal differences in their structures. However, there were differences in the lattice packing, which accounted for changes in the unit cell lengths. **Fig. 4A** shows the superposition of the MoFe proteins in the "short a short b" and "short a long b" datasets. By looking at the symmetry mates, we found that the neighboring molecules were farther apart in the b-c plane and the b-a plane in the dataset with the longer b-axis parameter. The same was true in the comparison between the "short a short b" and "long a long b" datasets **(Fig. 4B)**. Because of the differing amounts of space between the neighboring molecules, there were slight differences in these three datasets in the helices and loops (α-subunit residues 35-41, 121-124, 344-348 and β-subunit residues 400-414) that reside at the lattice packing interface. In the case of the datasets that contained short a-axes, the interfacial α-subunit residues 35-41 were disordered **(Fig. 4C)**.

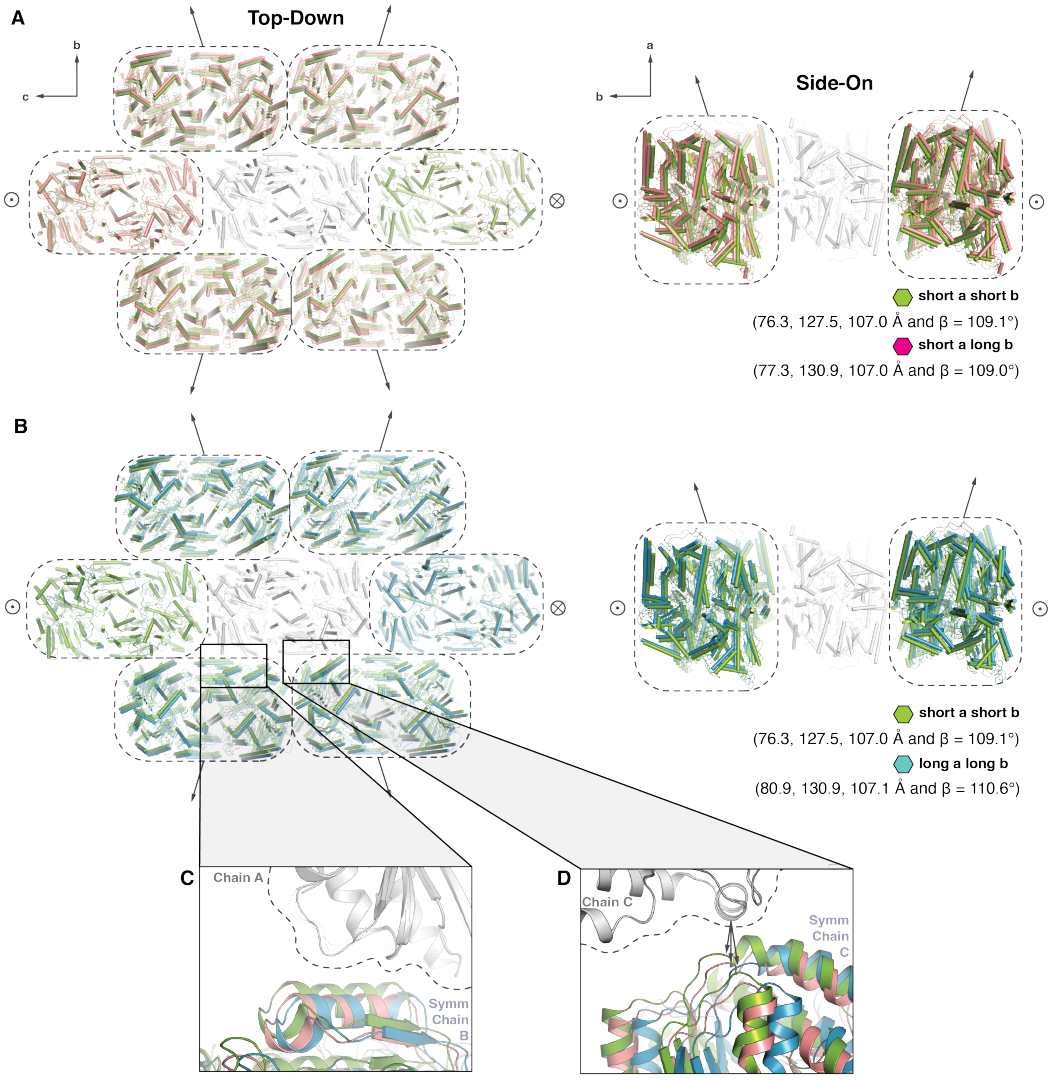


Figure 2.4: **A)** Top-down and side-on views of symmetry mates of two datasets: "short a and short b" (green) and "short a long b" (pink) that were aligned on a single MoFe protein tetramer (white). The "short a long b" (pink) symmetry mates were farther from the central molecule (white) in the direction of the arrows shown, which accounted for the longer unit cell lengths. **B)** Top-down and side-on views of symmetry mates of "short a and short b" (green) and "long a long b" (teal) that were aligned on a single MoFe protein tetramer (white). The "long a long b" (teal) symmetry mates were farther from the central molecule (white) in the direction of the arrows shown. Bullseye indicates direction of the aligned symmetry mate coming out of the page, and cross indicates the direction going into the page. Alignment of all three datasets on a single MoFe protein tetramer (white) shows the difference in the interface **C)** between chain A and symmetry mate chain B and **D)** chain C and symmetry mate chain C among the three datasets. Note, views of the insets C and D have been rotated for clarity.

Av MoFe protein crystal pathologies

In addition to small (~5 Å) variations in the unit cell parameters, some *Av* MoFe protein crystals contain lattice defects that lead to extra spots in their diffraction patterns. An example of a "standard" diffraction image of a MoFe protein crystal without this crystal defect is shown in **Fig. 5**.

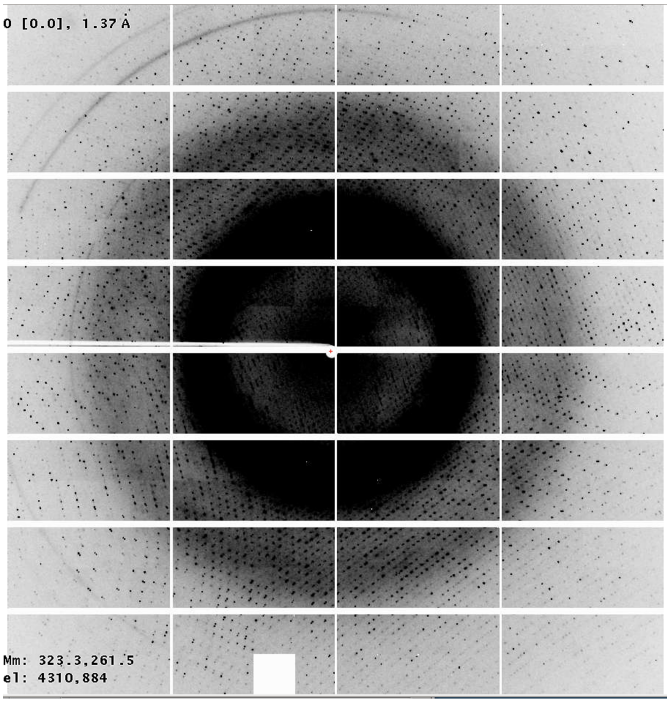


Figure 2.5: X-ray diffraction image collected at the Advanced Photon Source (Lemont, IL) beamline 23-ID-B on March 31, 2021 of a nitrogenase MoFe protein crystal at 12 keV, with detector distance of 200 mm.

A crystal with this defect resulted in a diffraction pattern that could be indexed in either the "standard" unit cell (for example with parameters a = 81.1, b = 130.6, c = 107.2 Å, α = 90, β = 110.6, γ = 90°) or with a "doubled" unit cell, where additional spots appeared between the lunes of the "standard" diffraction pattern **(Fig. 6)**. In the doubled unit cell, the a-axis was doubled in length (now ~160.0 rather than 80.0 Å) and due to crystallographic convention, was no longer the smallest axis so it became the c-axis. This resulted in a "doubled" unit cell with parameters such as a = 107.1, b = 130.4, c = 159.6 Å, α = 90, β = 108.3, γ = 90°. Both the "standard" and "doubled" unit cells had $P2_1$ symmetry. Although structures could be solved using either of these indexing solutions, one (typically the doubled unit cell) had higher R-factors in their structure solutions. Furthermore, the indexing solution for the "standard" unit cell typically extended to higher resolution. Thus, initially indexing

only the high-resolution reflections (< 1.8 Å), could assist in finding the "standard" unit cell parameters for any given MoFe protein dataset.

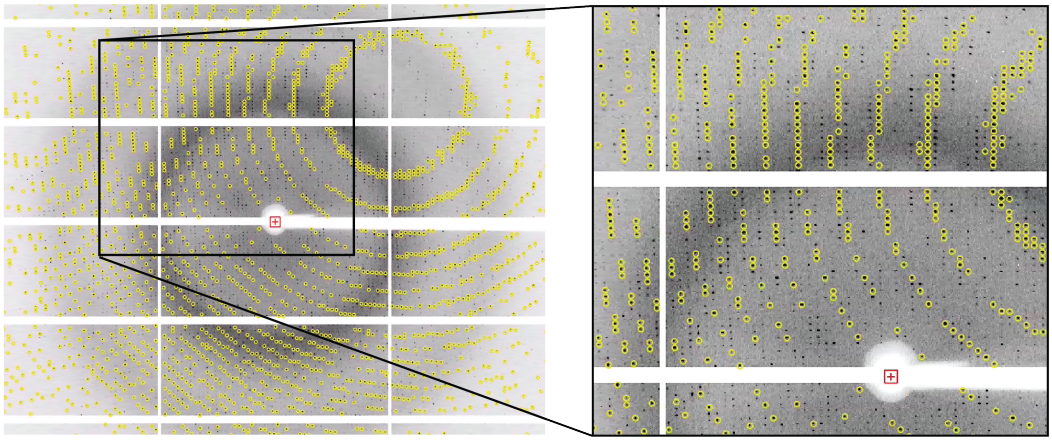


Figure 2.6: **Left)** Diffraction pattern of a MoFe protein crystal exhibiting lattice translocation disorder. Yellow circles indicate the spots indexed that result in the "standard" unit cell parameters (81.1, 130.6, 107.2 Å, 90, 110.6, and 90°). **Right)** Zoom-in shows the second lattice of spots between the spots that make the "standard" unit cell. If these spots are indexed in addition to the "standard" unit cell spots, the unit cell a-axis is doubled to yield parameters of 107.1, 130.4, 159.6 Å, 90, 108.3, 90°.

Density for translated metalloclusters in MoFe protein crystals

In some MoFe protein crystal data datasets, we found extensive positive difference electron density (F_o-F_c) around the FeMoco clusters **(Fig. 7A)**. Initially, we explored whether the difference density resulted from partial occupancy of the FeMoco in altered conformations; however, upon further examination this was ruled out. When visualizing only the F_o-F_c density around the FeMoco **(Fig. 7B)** we found that the difference density could be adequately modeled by two additional copies of the FeMoco **(Fig. 7C)**. This resulted in a model with three copies of the FeMoco per α/β-subunit that were related by translations of about +/- 4 Å from the "original" FeMoco position **(Fig. 7D)**. The same phenomenon was observed in the other α/β-subunit, but the clusters were translated in a side-to-side direction **(Fig. 7 bottom row)**. Similarly, two additional copies of translated P-clusters were modeled in the positive difference density around the P-cluster in each α/β-subunits. Modeling these additional clusters fully accounted for the unexplained positive difference density around the clusters. This translated cluster phenomenon does not appear to be dependent on the unit cell parameters. In other words, datasets from each classified

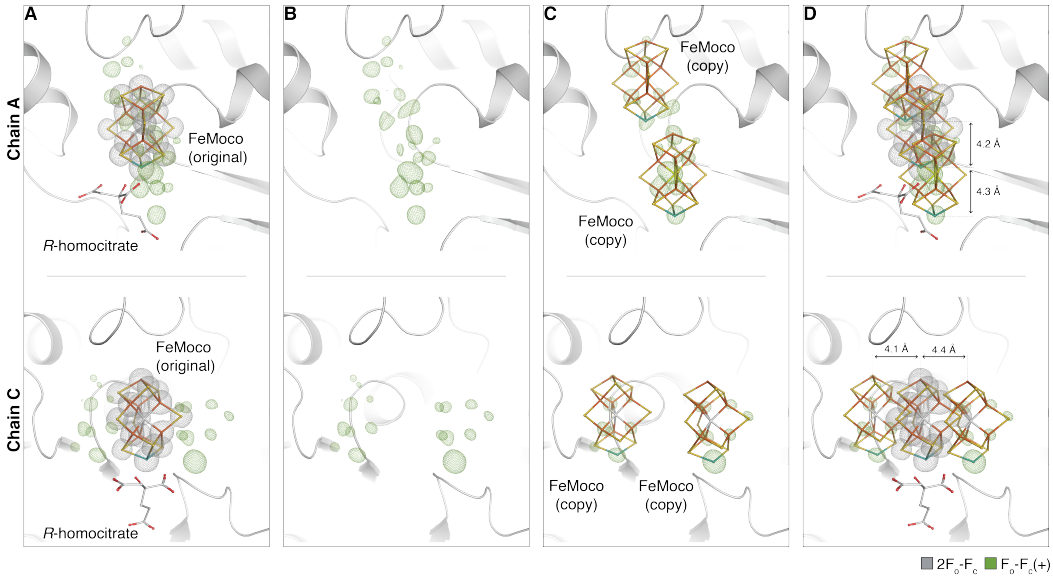


Figure 2.7: Positive difference electron density around FeMoco resulting from additional copies of the FeMoco. **A)** Example of a MoFe protein crystallographic dataset with significant positive difference electron density (green mesh) around the FeMoco in each α/β-subunit after modeling and refining one copy of the FeMoco (designated "original"). The top panel shows the FeMoco in chain A and the bottom panel in chain B. $2F_o$-F_c density is shown as gray mesh. **B)** Removal of the original FeMoco and its $2F_o$-F_c map enables visualization of the positive difference electron density (green mesh) that resembles two additional copies of the FeMoco. **C)** These difference density peaks were unequivocally modeled as two copies of the FeMoco. **D)** Modeling three copies of the FeMoco each translated ~4 Å fully accounts for electron density around the FeMoco active site.

unit cell population (from the unit cell cluster analysis described previously) had examples where difference density for translated clusters were found.

Although the clusters were translated in different directions in each α/β-subunit, they translated in the same direction relative the unit cell. It is likely that the rest of the MoFe protein is translated in accordance with the translated FeMoco, but due to the high scattering factor of the heavy atoms, only the clusters could be adequately modeled. Nonetheless, there was no clear evidence in the electron density maps for the translated protein atoms. Surprisingly, electron density for sulfur in the axial cysteine (α-subunit Cys275) that directly coordinates the FeMoco was not discernible, even though density for the translated cluster sulfurs was apparent. As an additional layer of complication, the translation of the clusters was not always localized directly around the cluster. In some datasets, the "translated clusters" were located

elsewhere in the protein, with the translated FeMoco molecules predominantly near α-subunit Arg93.

Av MoFe protein crystallographic merging for high-resolution and high-quality structure determination

X-ray crystal structures have been pivotal to the study of nitrogenase.[5, 11–18] High-resolution crystal structures were instrumental in determining the constituents and geometries of the MoFe protein clusters.[4, 5] The determination of the MoFe protein structure to 1.0 Å resolution enabled the visualization and assignment of the interstitial carbon atom within the FeMoco.[4] MoFe protein crystals rarely diffract to < 1.1 Å resolution, thus limiting our ability to observe sub-atomic features such as hydrogen atoms. As such, structures of the MoFe protein with resolution better than 1 Å are lacking. Therefore, we developed an algorithm to leverage many datasets of the same protein to improve crystallographic resolution and quality. Applying this algorithm to MoFe protein crystal datasets yielded a structure that has similar resolution, but higher redundancy and lower coordinate error in comparison to the highest resolution structure published to date.[4]

Based on the results of our variable unit cell clustering analysis, we chose to focus on a unit cell cluster (the "long a long b" cluster) that had the most complete datasets with better than 1.2 Å resolution diffraction. From the long a long b cluster, we picked the six highest resolution datasets from different MoFe protein crystals (listed in **Table 2**) as candidates for merging.

Table 2.2: Unit cell parameters from the crystal datasets chosen for merging from the "long a long b" cluster.

No.	**a** (Å)	**b** (Å)	**c** (Å)	**α** (°)	**β** (°)	**γ** (°)
1	80.88	130.68	107.09	90.00	110.60	90.00
2	80.99	130.67	107.15	90.00	110.59	90.00
3	80.96	130.66	107.18	90.00	110.61	90.00
4	80.92	130.64	106.88	90.00	110.81	90.00
5	80.99	130.38	107.06	90.00	110.51	90.00
6	80.98	130.63	107.11	90.00	110.58	90.00

Each crystallographic dataset was processed using the general workflow depicted in **Fig. 8**. The datasets were individually indexed, integrated, and scaled using XDS. [19] The Laue group and point-group symmetries were determined using POINTLESS [20] in the CCP4 program suite.[21] During this step, the six individual datasets were combined into a single mtz file containing over 50 million observations. Next, we performed merging and scaling with CCP4 AIMLESS, [20] which reduced the number of observations to 1.4 million unique reflections. For the initial analysis, we chose not to impose a maximum resolution cut off, thus the resolution ranged from 50.16 to 0.91 Å. Within this resolution range, the overall merging R-factor was 0.283 and the Rpim was 0.069. The overall completeness and multiplicity were 98.1% and 33.9, respectively. The average unit cell parameters were 81.00 Å, 130.77 Å, 107.20 Å, 90.00, 110.63, 90.00 ° in space group $P2_1$ and had an average mosaicity of 0.11.

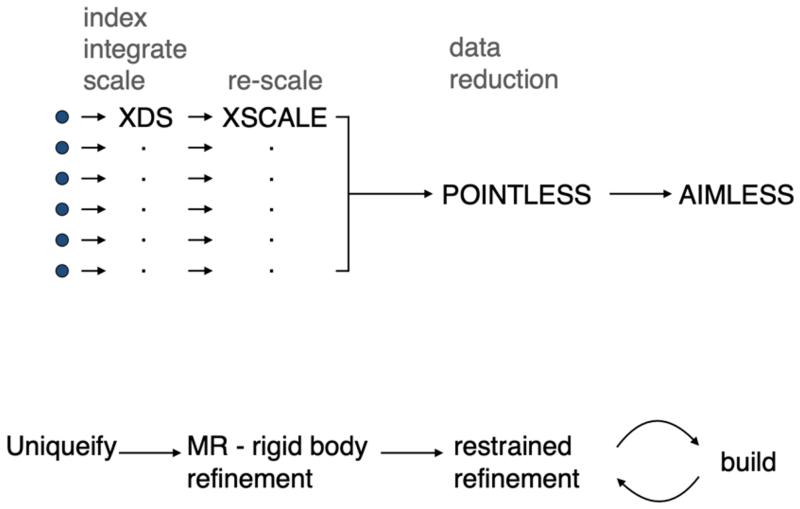


Figure 2.8: General workflow for processing the crystallographic datasets chosen for merging.

As the final step in processing our reflections file, we generated FreeR flags from 5% of the total data using CCP4 UNIQUE. We obtained phases by performing rigid body refinement against the protein atoms of a previously published MoFe protein structure PDB 3U7Q using CCP4 REFMAC5.[9] All waters, ions, and metallocofactors were removed from the molecular replacement model. Model building was performed in Coot [8] and both REFMAC5 and Phenix [10] were used for further refinement. After approximately 150 cycles of refinement and model building, positive difference electron density peaks appeared along some parts of the peptide backbone and aliphatic side chains **(Fig. 9)**.

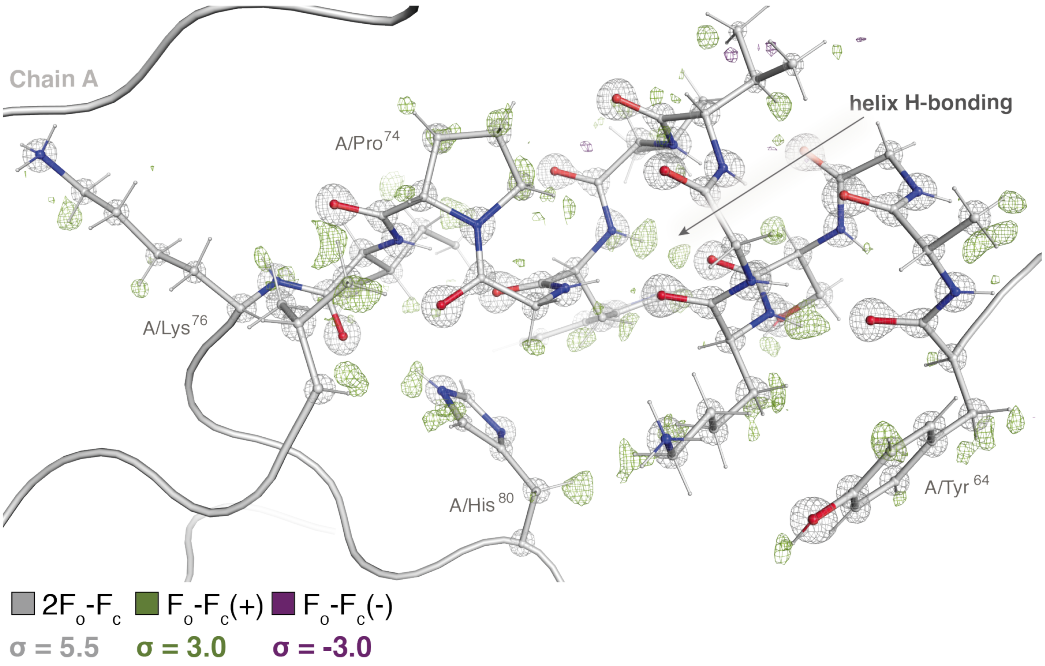


Figure 2.9: Positive difference electron density (green mesh) peaks at the positions of explicit hydrogens in a well-ordered α-helix of the MoFe protein in our merged crystallographic dataset. The model with explicit hydrogens (thin white sticks) are shown, but the map was refined against a model without hydrogens, thus resulting in difference density peaks where hydrogens are present. The negative F_o-F_c density is shown as purple mesh and the $2F_o$-F_c density is shown as gray mesh.

The FeMoco **(Fig. 10 A)** and P-cluster **(Fig. 10 B)** can be unequivocally modeled into the $2F_o$-F_c density near their respective active sites. However, the difference electron density around these clusters was noisy relative to the rest of the protein **(Fig. 10 C** and **D)**. Although we were most interested in learning the location of the hydrogens around the active site, the level of noise in this region, likely exacerbated by the Fourier series termination ripples from the metal centers,[5] prevented us from unequivocally assigning the positions of every hydrogen atom.

Nonetheless, there were positive difference density peaks at several explicit hydrogen positions around the active site, including the R-homocitrate that coordinates the FeMoco. Around the homocitrate, there was a pool of water molecules on which we could model several hydrogens that inform the orientation of waters and intricate hydrogen bonding network in this reservoir **(Fig. 11)**. This water pool has previously garnered interest among theorists who use molecular simulations to calculate proton ingress and product egress in nitrogenase active sites.[22, 23] However, their models are based on crystal structures lacking hydrogens. These

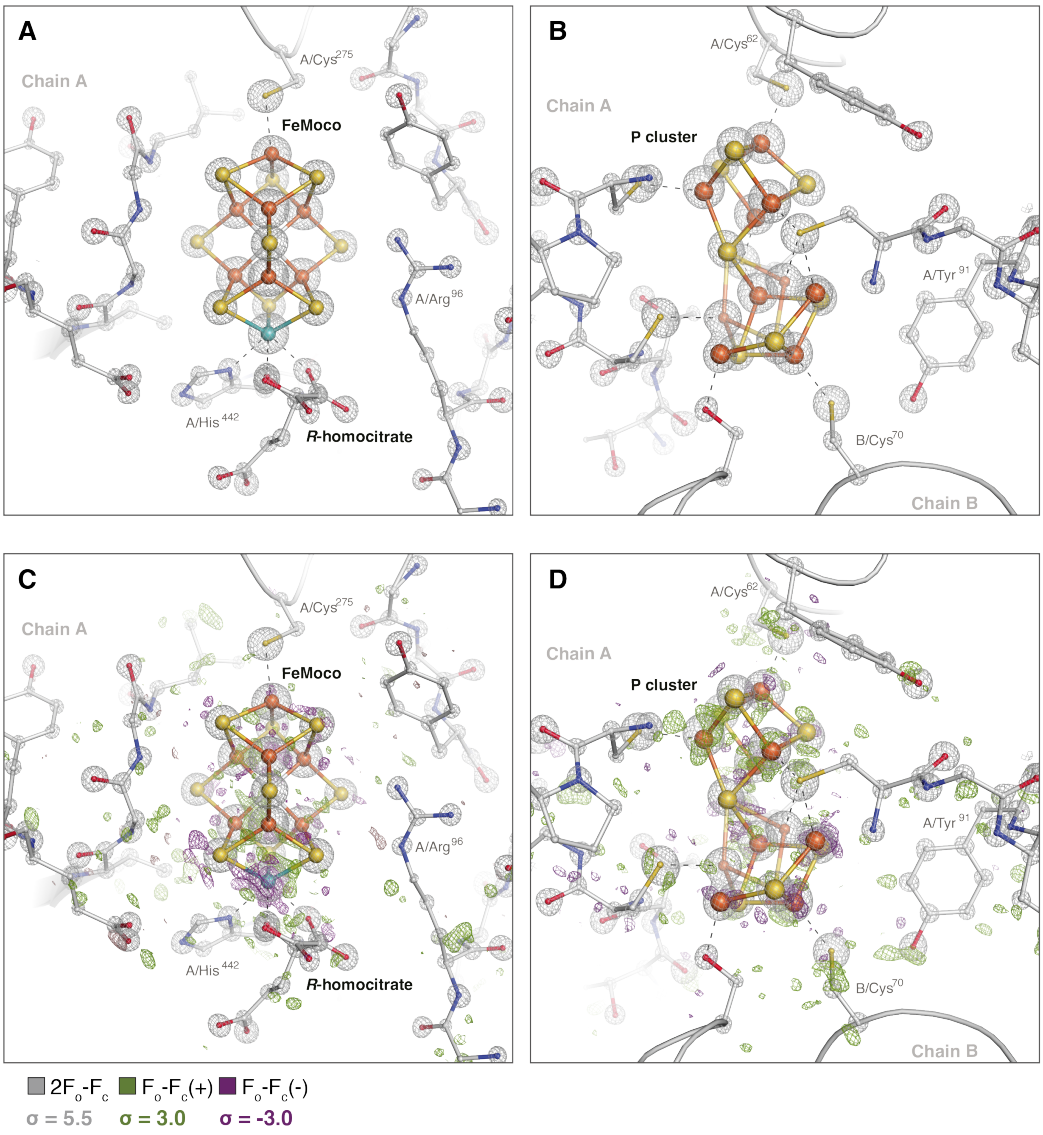


Figure 2.10: $2F_o$-F_c density (gray mesh) for the **A)** FeMoco and **B)** P-cluster sites in our merged datasets. **C)** Same as panel A but with positive (green mesh) and negative (purple mesh) F_o-F_c density shown. **D)** Same as panel B but with F_o-F_c density shown.

calculations should benefit from our new dataset where we have modeled some hydrogens, albeit with low confidence, as many of these difference density peaks were shrouded in noise. However, even these low confidence estimates could serve as improved initial conditions for further calculations.

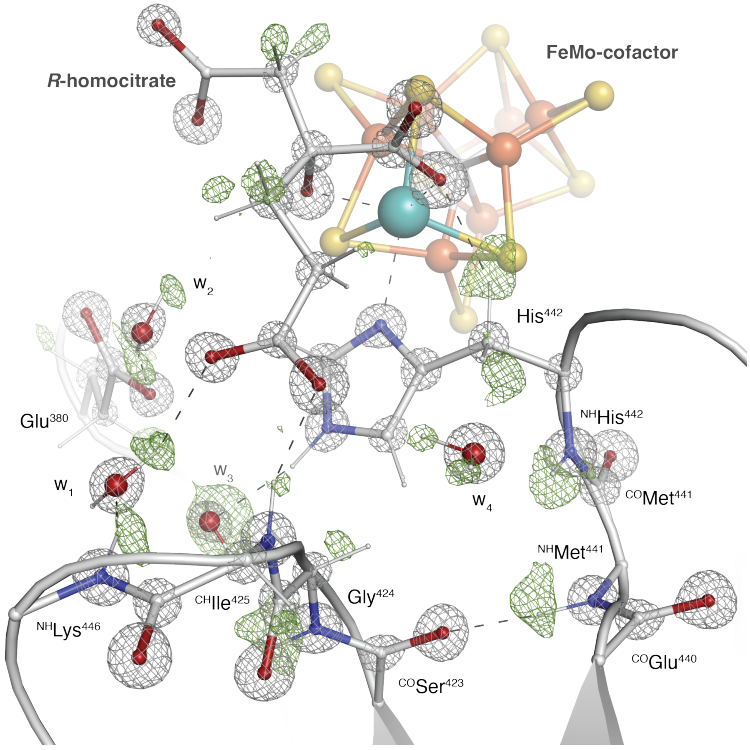


Figure 2.11: Hydrogen bonding network around the *R*-homocitrate, water molecules, and surrounding residues can be estimated using information from positive difference density (shown as green mesh) in our merged dataset. Hydrogens are shown as thin sticks. The maps were refined against a model lacking hydrogen atoms to visualize positive difference density peaks in the places of hydrogen atoms.

Resolution of the merged MoFe protein crystallographic dataset

The question of the resolution of our merged MoFe protein dataset is challenging to address. The standard metrics for determining the maximum resolution resulted in conflicting resolution cut offs. For example, we can considered the cut off using the mean I/σI of our unmerged reflections determined from POINTLESS, which had I/σI = 2.2 at 1.41 Å resolution and I/σI = 1.0 at 1.10 Å resolution. However, this did not coincide with the maximum resolution estimated by AIMLESS being either 1.21 Å based on an CC1/2 score of 0.65 or 1.61 Å based on mean I/σI of 4.28. Indeed, all of these resolution cut offs appeared to be conservative when considering

the features that were visible in the final electron density maps. We re-processed the data imposing resolution cut offs at 0.91, 0.93, and 0.95 Å to determine how the density for the hydrogen atoms were affected **(Fig. 12)**. Based on this analysis, it appeared that the most information was retained in the 0.91 Å resolution dataset. Although there was significantly more noise in the difference density maps in the 0.91 Å rather than in the 0.95 Å resolution dataset, there was significant gain of difference electron density for some hydrogen atoms when cut at 0.91 Å resolution. This was demonstrated in **Fig. 12**, which depicts a well-ordered tyrosine residue (α-subunit Tyr 99) with positive difference density peaks (green mesh) for all the aromatic hydrogens in the 0.91 Å resolution dataset. Although the 0.95 Å resolution dataset was much cleaner in terms of noise level, the positive difference density for the aromatic hydrogen indicated by the arrow was not visible at the same map σ-level (3.0).

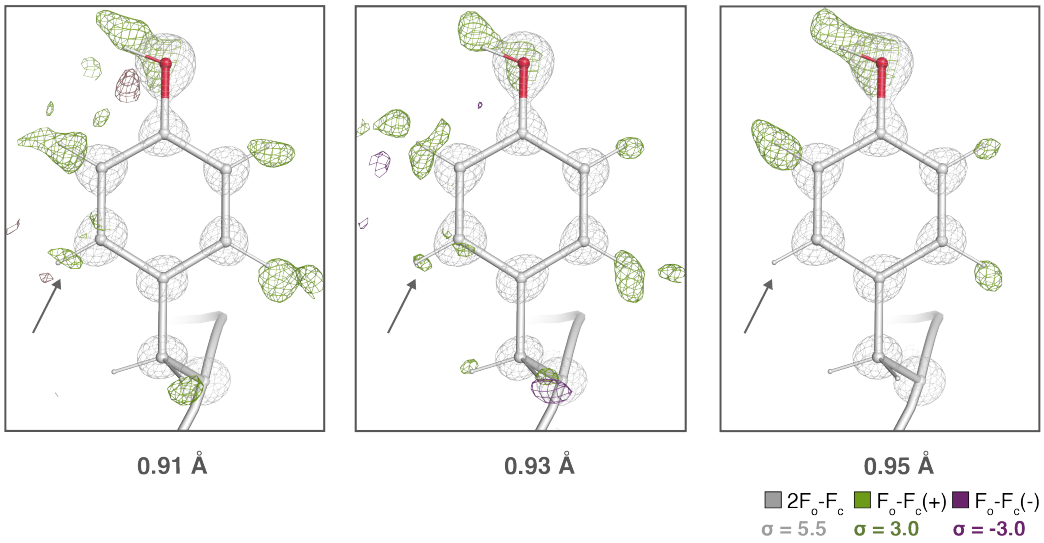


Figure 2.12: Comparison of the $2F_o$-F_c (gray mesh) and difference density (green mesh for positive and purple for negative) of a tyrosine residue in the merged dataset re-processed with 0.91, 0.93, and 0.95 Å resolution cut offs. Hydrogens are shown as thin sticks. The maps were refined against models lacking hydrogens to visualize positive difference density peaks in place of hydrogen atoms. The 0.91 Å resolution map is the noisiest, but also has the most signal for hydrogen atoms. The 0.95 Å resolution dataset has less noise but loses the signal for the aromatic hydrogen indicated by the arrow.

However, the apparent optical resolution did not coincide with acceptable statistics in the highest resolution bins. For the highest resolution bin (0.92 to 0.91 Å), the Rmerge and Rpim were negative (-14.1 and -8.0, respectively), which would typi-

cally be considered nonsensical. Notably, the total number of observations in this resolution bin is 393,518 (49,605 unique), yielding a multiplicity of 7.9. Further AIMLESS statistics are reported in **Table 3**. Perhaps merging six large crystallographic datasets with ~8 million reflections each and potentially variable diffraction quality at very high resolution, inadequately weights the contribution of the reflections across all the resolution bins, consequently resulting in an underestimation of the proper resolution cutoff.

Table 2.3: AIMLESS merging statistics.

	Overall	**InnerShell**	**OuterShell**
Low resolution limit	50.16	50.16	0.92
High resolution limit	0.91	4.98	0.91
Rmerge (within I+/I-)	0.283	0.094	-14.168
Rmerge (all I+ and I-)	0.285	0.095	-14.253
Rmeas (within I+/I-)	0.292	0.096	-16.481
Rmeas (all I+ & I-)	0.290	0.096	-15.564
Rpim (within I+/I-)	0.069	0.020	-8.017
Rpim (all I+ & I-)	0.049	0.014	-5.757
Total number of observation	49641142	409595	393518
Total number unique	1462977	9278	49605
Mean((I)/sd(I))]	7.8	37.7	-0.1
Mn(I) half-set correlation CC(1/2)	0.921	0.999	0.152
Anomalous completeness	97.6	99.9	59.7
Anomalous multiplicity	17.0	22.6	4.3

The average R-factor versus resolution **(Fig. 13)** determined by refinement in REFMAC5 shows that the average R-factor was 0.46 at 0.92 Å resolution, where an average R-factor below 0.5 is considered to contain meaningful data. Additionally, the completeness in the highest resolution bin (0.93 – 0.92 Å) was 98.1% as shown in **Table 4**.

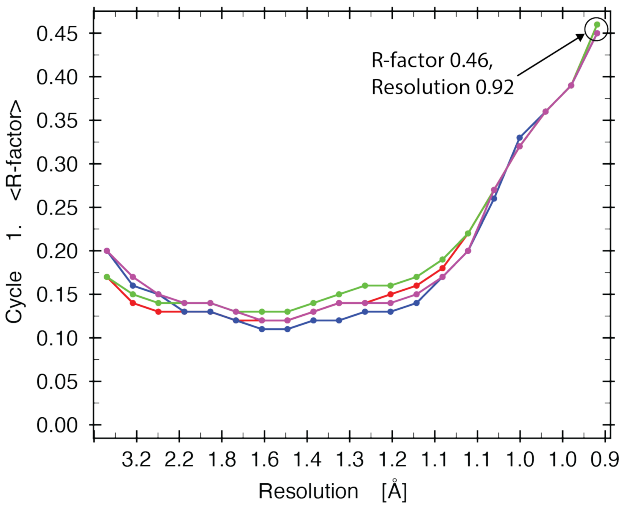


Figure 2.13: The average R-factor versus resolution plot from structure refinement using REFMAC5 indicates the average R-factor is 0.46 at the highest resolution bin 0.92 for the merged dataset.

Comparison to the 1 Å resolution MoFe protein structure (PDB 3U7Q)

Our merged dataset had improved accuracy and precision in comparison to the highest resolution MoFe protein structure published to date (PDB 3U7Q),[4] as evidenced by the following metrics. Owing to the large number of recorded observations (over 50 million) that make up this dataset, there was a substantial gain in the redundancy in our merged data (34.4 versus 6.4 for PDB 3U7Q). Our merged dataset had roughly 40% more unique reflections than PDB 3U7Q (1.4 versus 1.0 million for PDB 3U7Q) and higher completeness in the high resolution bins (94.7% in the 1.05-0.91 Å bin and 98.8% in the 1.17-1.00 bin for PDB 3U7Q). Additionally, the model derived from our merged dataset had lower coordinate error than that of PDB 3U7Q (Cruickshank DPI of 0.018 versus 0.027 for PDB 3U7Q).

Our merged dataset also had sharpened density for a structurally variable region of the MoFe protein, enabling us to model two conformations of an exterior helix (β-subunit residues 400-415) for the first time **(Fig. 14 A)**. This helix typically harbors elongated density in both the peptide backbone and sidechains, such as that shown in **Fig. 14 B** from PBD 3U7Q. However, the two conformations were clearly resolved in our merged dataset.

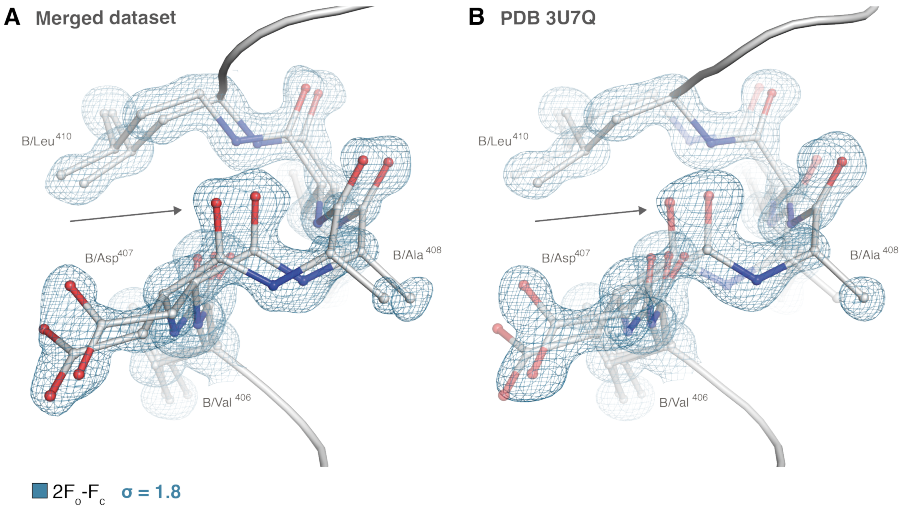


Figure 2.14: **A)** 2F$_o$-F$_c$ electron density (blue mesh) of a portion of an external helix in the MoFe protein β-subunit that can be modeled in two conformations. The arrow indicates a peptide carbonyl for which there is clear density for two distinct conformations. **B)** This second conformation is not resolvable in the 2F$_o$-F$_c$ electron density for the same region in PDB 3U7Q. The carbonyl indicated by the arrow has elongated density, but cannot be unequivocally modeled as a second conformation (shown as transparent sticks).

The second conformation of the helix is shifted in the direction towards the center of the protein, further interacting with a neighboring loop (β-subunit 389-391) **(Fig. 15)**. The neighboring loop contains a proline (β-subunit Pro 390) and valine (β-subunit Val 391) in two conformations, related by an intervening peptide flip. This helix resides near the tetramer interface between the two α/β-subunits, thus it is tempting to consider that changes in the structure in this region could contribute to coupling between one α/β-subunit to the other.

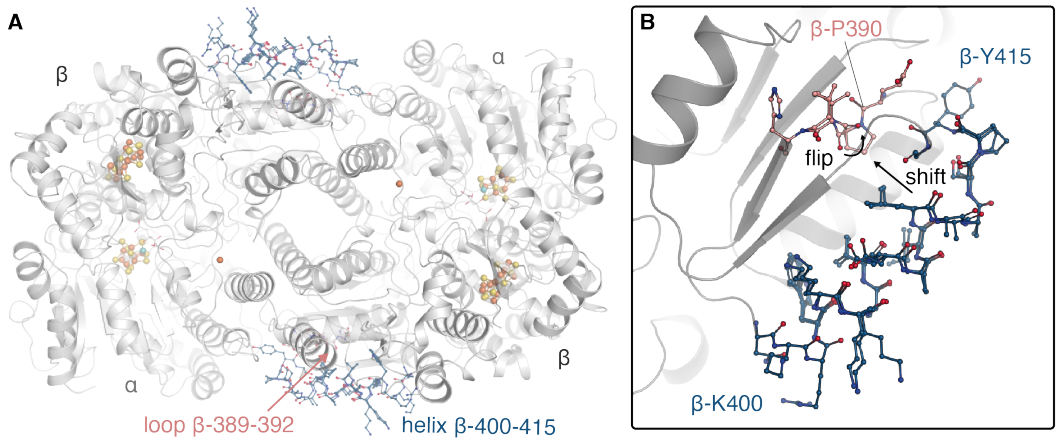


Figure 2.15: Cartoon representation of the MoFe protein from our merged dataset. Metalloclusters are shown as spheres. The external helices (β-subunit 400-415) and neighboring loop (β-subunit 389-392) in multiple conformations are shown as blue and pink sticks, respectively. Inset shows a zoomed in view of the multiple conformations of the shifted helix.

Altogether, analysis of *Av* MoFe protein crystal isoforms and their crystal pathologies led us to develop an algorithm to classify variable unit cell parameters from hundreds of MoFe protein crystallographic datasets. This variable unit cell clustering analysis enabled us to identify 6 different isomorphous datasets that were ideal candidates for crystallographic merging. The resultant merged dataset has improved data quality relative to any other MoFe protein structure published to date. These data revealed positive difference electron density for hydrogens in important regions of the protein and multiple conformations of an α-helix at the tetramer interface. The high resolution and low coordinate error of this merged dataset enable precise and accurate estimation of the geometry of the metallocofactor and neighboring residues. These improved measurements may inform the theoretical calculations that explore the potential mechanisms of nitrogen fixation.

Table 2.4: Completeness and multiplicity versus resolution of the merged dataset.

N	1/d^2	Dmid	Nmeas	Nref	Ncent	%poss	C%poss	Mlplct	AnoCmp
1	0.0202	7.04	409595	9278	522	99.9	99.9	44.1	99.9
2	0.0606	4.06	625699	16782	521	100	100	37.3	100
3	0.101	3.15	873037	21668	522	100	100	40.3	100
4	0.1413	2.66	1070106	25571	515	100	100	41.8	100
5	0.1817	2.35	1135769	28983	522	100	100	39.2	100
6	0.2221	2.12	1209985	32012	513	100	100	37.8	100
7	0.2625	1.95	1428405	34769	519	100	100	41.1	100
8	0.3029	1.82	1573278	37356	518	100	100	42.1	100
9	0.3433	1.71	1670877	39771	515	100	100	42	100
10	0.3836	1.61	1673298	42041	526	100	100	39.8	100
11	0.424	1.54	1728682	44128	509	100	100	39.2	100
12	0.4644	1.47	1898573	46199	529	100	100	41.1	100
13	0.5048	1.41	2001859	48189	512	100	100	41.5	100
14	0.5452	1.35	2042244	49969	515	100	100	40.9	100
15	0.5856	1.31	2125880	51813	518	100	100	41	100
16	0.6259	1.26	1990685	53558	516	100	100	37.2	100
17	0.6663	1.23	2193536	55337	520	100	100	39.6	100
18	0.7067	1.19	2267854	56889	516	100	100	39.9	100
19	0.7471	1.16	2346805	58485	503	100	100	40.1	100
20	0.7875	1.13	2431943	60160	532	100	100	40.4	100
21	0.8279	1.1	2277668	61627	518	100	100	37	100
22	0.8682	1.07	2413166	63034	515	100	100	38.3	100
23	0.9086	1.05	2395238	64445	518	100	100	37.2	100
24	0.949	1.03	2382022	66002	515	100	100	36.1	100
25	0.9894	1.01	2293611	67340	522	100	100	34.1	100
26	1.0298	0.99	1757066	68600	501	100	100	25.6	100
27	1.0702	0.97	1241468	69687	516	99.6	100	17.8	99.4
28	1.1105	0.95	964596	70311	476	98.6	99.9	13.7	98.3
29	1.1509	0.93	824679	69368	453	95.5	99.7	11.9	94.8
30	1.1913	0.92	393518	49605	303	67.2	98.1	7.9	59.7
		Overall:	49641142	1462977	15200	98.1	98.1	33.9	97.6

References

(1) Cao, L.; Caldararu, O.; Ryde, U. Protonation and Reduction of the FeMo Cluster in Nitrogenase Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations. *Journal of Chemical Theory and Computation* **2018**, *14*, 6653–6678.

(2) Hoffman, B. M.; Lukoyanov, D.; Yang, Z.-Y.; Dean, D. R.; Seefeldt, L. C. Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical Reviews* **2014**, *114*, 4041–4062.

(3) Thorneley, R.; Lowe, D. Molybdenum enzymes. *Metal Ions in Biology* **1985**, *7*, 221–284.

(4) Spatzal, T.; Aksoyoglu, M.; Zhang, L.; Andrade, S. L. A.; Schleicher, E.; Weber, S.; Rees, D. C.; Einsle, O. Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science* **2011**, *334*, 940–940.

(5) Einsle, O.; Tezcan, F. A.; Andrade, S. L. A.; Schmid, B.; Yoshida, M.; Howard, J. B.; Rees, D. C. Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science* **2002**, *297*, 1696–1700.

(6) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. *HKL* -3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography* **2006**, *62*, 859–866.

(7) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Phaser* crystallographic software. *Journal of Applied Crystallography* **2007**, *40*, 658–674.

(8) Emsley, P.; Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60*, 2126–2132.

(9) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **1997**, *53*, 240–255.

(10) Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology* **2019**, *75*, 861–877.

(11) Georgiadis, M. M.; Komiya, H.; Chakrabarti, P.; Kornuc, J. J.; Rees, D. C. Crystallographic Structure of the Nitrogenase Iron Protein from Azo tobacter vinelandli. **1992**, *257*, 8.

(12) Kim, J.; Woo, D.; Rees, D. C. X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. *Biochemistry* **1993**, *32*, 7104–7115.

(13) Schindelin, H.; Kisker, C.; Schlessman, J. L.; Howard, J. B.; Rees, D. C. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. *Nature* **1997**, *387*, 370–376.

(14) Tezcan, F. A. Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein. *Science* **2005**, *309*, 1377–1380.

(15) Spatzal, T.; Perez, K. A.; Einsle, O.; Howard, J. B.; Rees, D. C. Ligand binding to the FeMo-cofactor: Structures of CO-bound and reactivated nitrogenase. *Science* **2014**, *345*, 1620–1623.

(16) Spatzal, T.; Perez, K. A.; Howard, J. B.; Rees, D. C. Catalysis-dependent selenium incorporation and migration in the nitrogenase active site iron-molybdenum cofactor. *eLife* **2015**, *4*, DOI: 10.7554/eLife.11620.

(17) Sippel, D.; Rohde, M.; Netzer, J.; Trncik, C.; Gies, J.; Grunau, K.; Djurdjevic, I.; Decamps, L.; Andrade, S. L. A.; Einsle, O. A bound reaction intermediate sheds light on the mechanism of nitrogenase. *Science* **2018**, *359*, 1484–1489.

(18) Trncik, C.; Detemple, F.; Einsle, O. Iron-only Fe-nitrogenase underscores common catalytic principles in biological nitrogen fixation. *Nature Catalysis* **2023**, DOI: 10.1038/s41929-023-00952-1.

(19) Kabsch, W. *XDS*. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66*, 125–132.

(20) Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography* **2006**, *62*, 72–82.

(21) Agirre, J. et al. The *CCP* 4 suite: integrative software for macromolecular crystallography. *Acta Crystallographica Section D Structural Biology* **2023**, *79*, 449–461.

(22) Dance, I. A molecular pathway for the egress of ammonia produced by nitrogenase. *Scientific Reports* **2013**, *3*, 3237.

(23) Dance, I. What is the role of the isolated small water pool near FeMo-co, the active site of nitrogenase? *The FEBS Journal* **2018**, *285*, 2972–2986.

Chapter 3

REDOX-DEPENDENT STRUCTURAL STATES OF NITROGENASE MOFE PROTEIN

3.1 Introduction

Nitrogenase catalysis proceeds through a cycle of sequential steps in which electrons are transferred to the MoFe protein P-cluster and ultimately accumulate on the FeMo-cofactor.[1] Based on extensive biochemical and kinetic data, Lowe and Thorneley developed a comprehensive kinetic model for the nitrogenase reaction.[2, 3] This model describes the reduction of a single molecule of N_2 to two molecules of ammonia, while simultaneously generating H_2, via multiple MoFe protein electronic and protonation states E_n (where n ranges from 0 to 7). As the reduced (E_1-E_7) states of MoFe protein are difficult to capture, the most well characterized state is the resting-state (E_0); however, the resting state does not directly bind N_2. It must first be reduced by 3-4 cycles of the Fe protein, which delivers electrons to the MoFe protein in an ATP-hydrolysis dependent manner.[4–8] In turn, the oxidized Fe protein is reduced by endogenous ferredoxins and flavodoxins in the cell.[7–10]

The Fe protein functions in the $[4Fe{:}4S]^{2+/1+}$ redox couple, which has a reduction potential ($E^{\circ\prime}$) of -0.30 V vs. the standard hydrogen electrode (SHE) in the nucleotide-free state.[11–13] Upon binding of MgATP and subsequently the MoFe protein, the $[4Fe{:}4S]^{2+/1+}$ $E^{\circ\prime}$ decreases to -0.43 V and -0.62 V vs. SHE, respectively,[12, 14] making the Fe protein a more effective electron donor after forming the MgATP-MoFe protein complex. The receptor cluster, the P-cluster, functions in the $P^{N/1+}$ and $P^{1+/2+}$ redox couples, which have similar $E^{\circ\prime}$s of -0.31 V vs. SHE.[15–17] In vitro, the ferredoxins and flavodoxins that reduce Fe protein can be replaced by chemical reductants. Sodium dithionite, which has $E^{\circ\prime}$ of ~-0.66 V vs. SHE is most commonly used.[18] However, strong reductants such as Eu(II)-diethylenetriaminepentaacetate (Eu(II)-DTPA) ($E^{\circ\prime}$= -1.1 V vs. SHE)[19, 20] and Ti(III)-citrate ($E^{\circ\prime}{}_{Ti(III)/(II)}$ = -0.8 V vs. SHE),[21] have been used to reduce Fe protein, namely to probe the all ferrous state of the Fe protein.[19, 22–24] A study by Danyal *et al.* reported the ability of a MoFe protein variant (βY98H) to reduce hydrazine to ammonia without the need for Fe protein when Eu(II)-DTPA was used as the reductant.[20] Although Eu(II)-DTPA was unable to facilitate N_2 reduction

to ammonia in that particular system, these studies indicate a growing interest in employing potent reductants to capture further reduced MoFe protein states for characterization. However, little work has probed the potential structural effects of these strong reductants on the MoFe protein.

In this chapter we investigate the structural consequences of introducing chemical reductants to MoFe protein crystals. In addition, we explore the effects of photoreduction from X-rays on MoFe protein crystals, using X-ray crystallography and X-ray absorption spectroscopy. Finally, we determined a room-temperature radiation damage free structure of MoFe protein using an X-ray free electron laser.

3.2 Results

Redox-dependent conformational changes of the MoFe protein P-cluster

Substantial prior works have established the structural changes that occur in the MoFe protein P-cluster in response to changes in the oxidation state. The P-cluster is an [8Fe:7S] cluster liganded by six cysteines (αCys62, αCys88, αCys154, βCys70, βCys95, βCy153) across the MoFe protein α/β dimer interface.[25] In the dithionite-reduced resting state, the P-cluster adopts a closed double-cubane structure (P^N) where all the irons are in the 2+ oxidation state.[26–28] Upon single electron oxidation, the P^{1+} cluster undergoes a structural change whereby the Fe6 coordinated to βCy153 opens the cluster and coordinates βSer188 **(Fig. 1)**.[29] Further oxidation to the P^{2+} state results in an additional structural change, where both Fe6 and Fe5 open the cluster and Fe5 coordinates to the backbone amide of αCys88.[30] The structure of the P^{1+} state was captured by application of an electrical charge over MoFe protein crystals in an electrochemical cell.[29] In contrast, the P^{2+} state was obtained by either applying the oxidant indigo disulfonate [31] or by depletion of dithionite over a prolonged amount of time.[30]

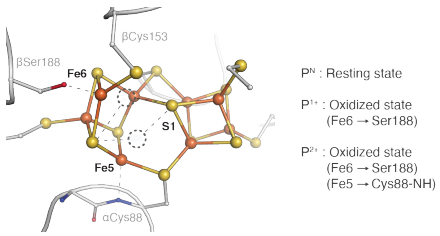


Figure 3.1: The P-cluster adopts three known structural states: P^N, P^{1+}, P^{2+}. In the P^{1+} state, Fe6 moves away from the closed position (indicated by a dashed circle) and coordinates βSer188. In the P^{2+} state, both Fe6 and Fe5 move away from their closed positions and Fe5 coordinates the backbone amide of αCys88.

We were able to reproduce structures with oxidized P-clusters by exposing MoFe protein crystals to air for various periods of time before flash freezing them in liquid nitrogen for structural characterization **(Fig. 2)**. The crystals that were incubated in air for up to ~40 minutes reliably diffracted. Incubation in air for greater than 40 minutes led to loss of brown color in the crystals and precipitous loss of diffraction. Most structures of the MoFe protein have a combination of fully closed (P^N) and fully open (P^{2+}) P-clusters, such as those seen in PDB 3U7Q.[32] However, using this short air-exposure method, we could reproducibly obtain fully opened (P^{2+}) P-cluster that diffracted to high resolution (<1.5 Å) without the addition of the oxidant indigo disulfonate.

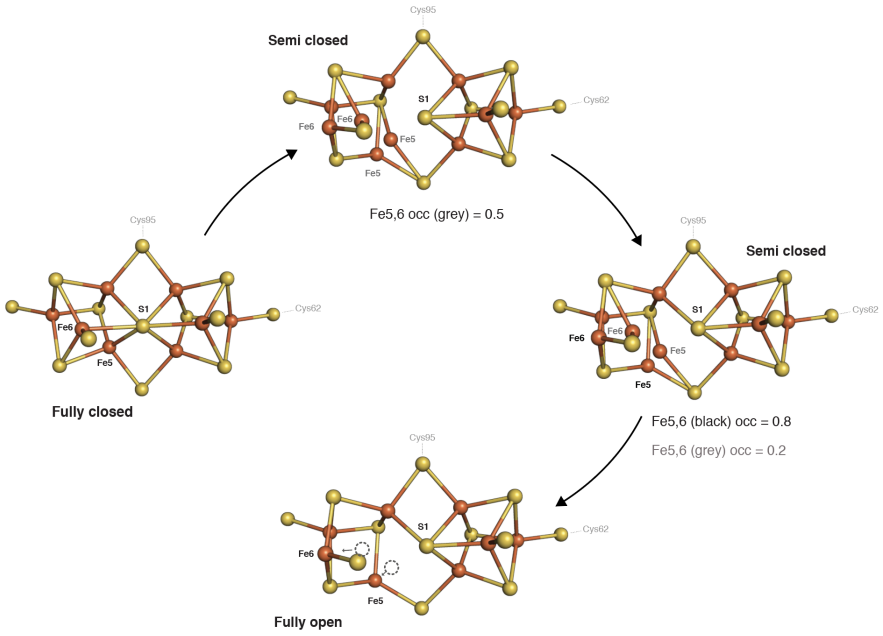


Figure 3.2: Structures of the fully closed (P^N) and fully open (P^{2+}) P-clusters in various occupancies. Typically, MoFe protein crystals have either fully closed or a combination of fully closed and fully open P-clusters. Our air-exposed MoFe protein crystals enabled us to obtain fully open P-cluster without the use of the oxidant indigo disulfonate.

The diffraction quality of MoFe protein crystals was not diminished by air-exposure, provided that they were frozen within the first 40 minutes of air-incubation. **Fig. 3** shows the $2F_o$-F_c electron density map of an air-exposed MoFe protein crystal determined to 1.18 Å resolution. The FeMo-cofactor remained unchanged and there were virtually no differences in the first or second coordination sphere around the cofactor. The primary structural differences were in the P-cluster, which adopted the fully opened (P^{2+}) conformational state.

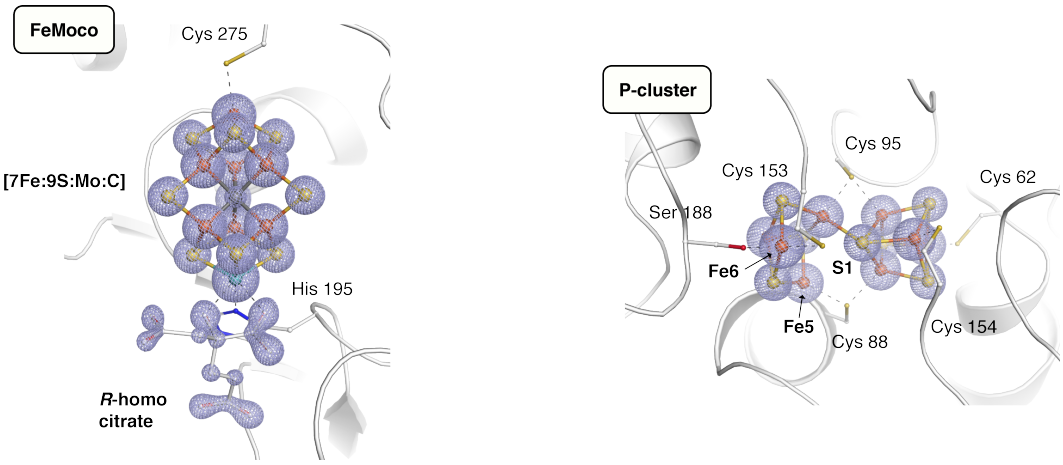


Figure 3.3: $2F_o$-F_c electron density (blue mesh) for the FeMo-cofactor (left) and the P-cluster (right) in the air-exposed MoFe protein crystal structure determined to 1.18 Å resolution. The P-cluster adopted a fully opened (P^{2+}) conformation.

After establishing a method to generate the fully opened (P^{2+}) P-cluster, we designed an experiment to learn what structural effects the addition of Eu(II)-DTPA and Ti(III)-citrate had on the MoFe protein and whether addition of these strong reductants would close the oxidized P-cluster. Thus, we started with MoFe protein crystals that were grown under strict anoxic conditions in an argon atmosphere (condition 1 in **Fig. 4**). The anaerobic addition of Eu(II)-DTPA to MoFe protein crystals (condition 2 in **Fig. 4**) had the surprising effect of converting the P-cluster from the closed to the open conformation, leading to a mixture of open and closed P-clusters. Although the Eu(II)-DTPA is a reductant, the crystals (n = 8) that were treated with Eu(II)-DTPA invariably had opened P-cluster in roughly 70% occupancy. Notably, Eu(II)-DTPA treatment had no effect on the conformation of the FeMo-cofactor. To determine if Eu(II)-DTPA could close an opened P-cluster, we generated MoFe protein crystals with fully opened P-cluster by brief air-exposure (condition 3 in **Fig. 4**) and re-equilibrated them into the glove box (condition 4 in **Fig 4.**). We found that anaerobically incubating Eu(II)-DTPA into MoFe protein

crystals with opened (P^{2+}) P-cluster did not close the P-cluster (condition 5 in **Fig. 4**). In addition, we found the same was true when oxidized MoFe protein was incubated with Ti(III)-citrate (condition 6 in **Fig. 4**). During the course of these experiments, we did not see accumulation of the (P^{1+}) state. Altogether, the addition of reductants Eu(II)-DTPA and Ti(III) citrate induced the open conformation of the P-cluster. Furthermore, when strong reductants were added to MoFe protein crystals that already had opened P-cluster, the cluster did not revert to the closed state.

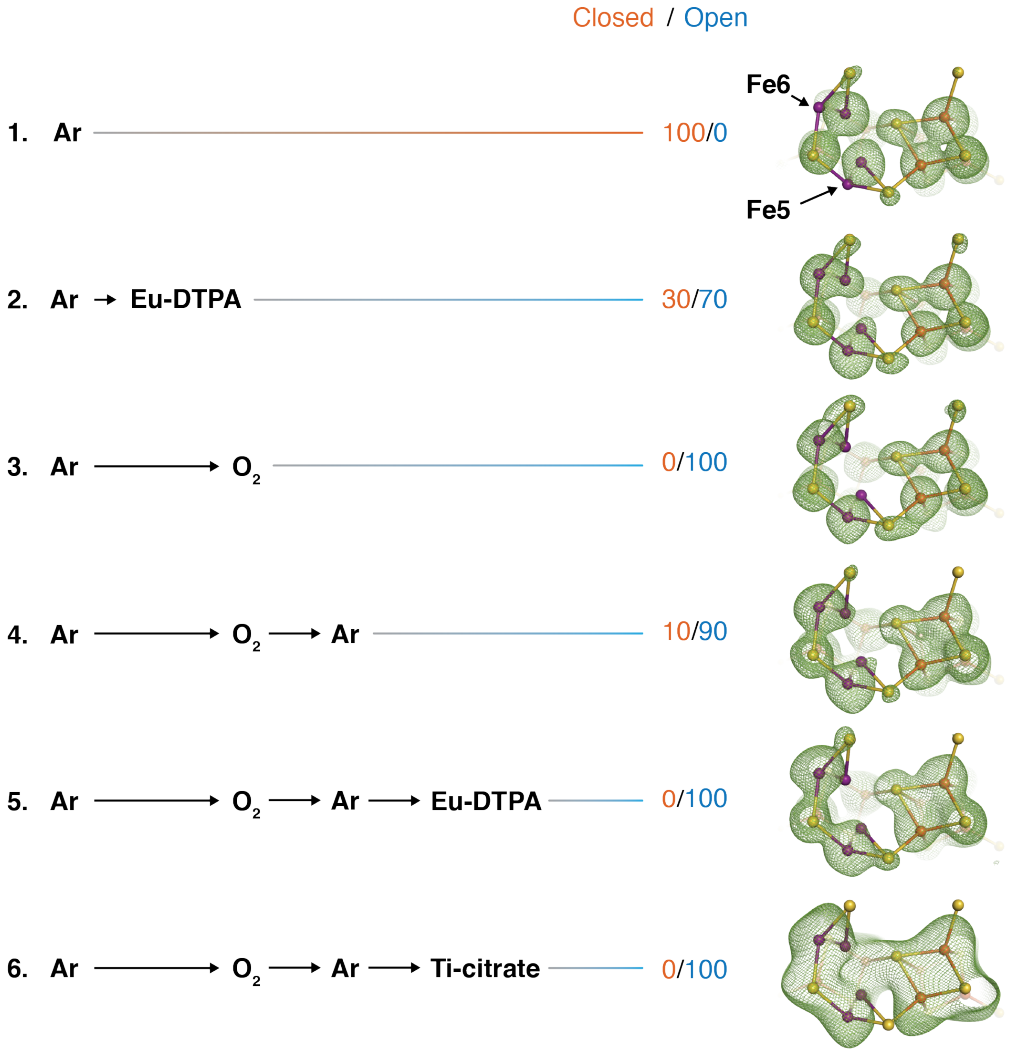


Figure 3.4: Representative F_o-F_c difference electron density (green mesh) for P-clusters with oxidant (O_2) and reductant (Eu(II)-DTPA and Ti(III)-citrate) treatments. The opened and closed conformations of the clusters are indicated by two purple spheres for Fe5 and Fe6. The occupancies of each P-cluster conformation are show as orange values for closed and blue values for opened.

MoFe protein crystals with photosensitizers

Inspired by the works of Syrtsova and Likhtenstein who reported the reduction of nitrogenase using an eosin dye combined with NADH and activated with light,[33, 34] we performed soaking experiments with eosin dyes on MoFe protein crystals. We collected several crystallographic datasets of MoFe protein crystals anaerobically soaked with either eosin-Y or eosin-B. X-ray data collection caused the crystals and surrounding mother liquor, which were dyed red as a result of the eosin soaks, to turn clear **(Fig. 5A** and **5B)**. The crystals that were soaked with the eosin dye had no differences in their MoFe protein structures compared to unsoaked crystals.

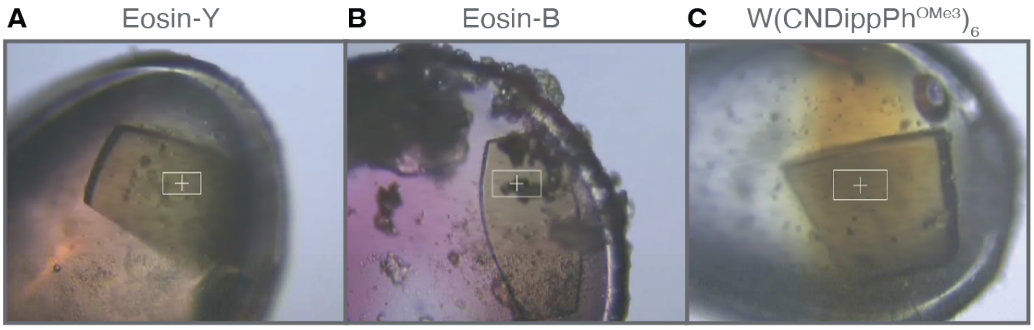


Figure 3.5: MoFe protein crystals soaked with **A)** eosin-Y, **B)** eosin-B, and **C)** W(CNDippPh^{OMe3})$_6$. Images by synchrotron hutch in-line camera taken post-X-ray data collection. The path of the X-ray beam can be viewed by the dark orange streak in panel **C**.

Further, we preformed crystal soaking experiments with a tungsten (0) aryliso-cyanide photoreductant that was developed in lab of Dr. Harry B. Gray. The compound W(CNDippPh^{OMe3})$_6$ (CNDippPh^{OMe3} = 4-(3,4,5-trimethoxyphenyl)-2,6-diisopropylphenylisocyanide)[35] was synthesized by Dr. Javier Fajardo and generously shared with us for soaking experiments. We established that MoFe protein crystals could be soaked with the W-photosensitizer by solubilization of the compound in a mixture of crystal mother liquor (containing 15% (w/v) PEG4000, 0.2 M NaCl, and 0.1 M imidazole/malate pH 8.0) and dimethylsulfoxide (DMSO) or ethanol. MoFe protein crystals withstood soaking conditions with up to 50% DMSO or 20% ethanol without visible crystal damage. MoFe protein crystals did not tolerate the addition of toluene to the crystal drop. We collected full crystallographic datasets on 15 out of 50 W-photosensitizer soaked crystals that diffracted to 1.9 Å resolution or better. By single-crystal fluorescence excitation, we confirmed the presence of W in our crystal loop. However, from both visual inspection of the resultant electron density maps and from W-anomalous data collection (at 10.000

and 10.208 keV) we found that the W-photosensitizer was not site-specifically binding to the MoFe protein in appreciable amounts from crystal soaking experiments. The structures revealed that there were no major changes to the overall MoFe protein architecture and no changes in the conformation of the FeMo-cofactors or the P-clusters upon addition of the W-photosensitizer. These are encouraging results for further experiments whereby the W(CNDippPh^{OMe3})$_6$ could be photoactivated directly in the crystal in an attempt to capture further reduced MoFe protein structures.

Anaerobic illumination of the eosin dye-soaked crystals with 1 to 10 minutes of blue LED light caused a change in the conformation of the P-cluster from the closed (P^N) state to the open (P^{1+}) state. Only a small fraction (~5%) of the P-cluster was in the P^{1+} state, as indicated by the green difference electron density peak in the position of Fe6 in the open conformation in **Fig. 6**. Although it is not unusual for there to be multiple conformations of the P-cluster in MoFe protein crystals, it is unusual for there to be a clear population of the P^{1+} and not the P^{2+} state. Thus, we found the combination of eosin with light enabled the visualization of the P^{1+} without the need for applying a potential across a crystal using an electrochemical cell. This was an unexpected result, considering that the eosin dyes were expected to affect Fe protein only, not MoFe protein in photoactivation experiments.[33, 36]

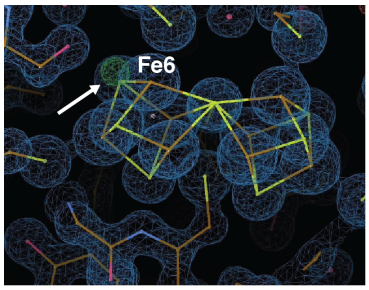


Figure 3.6: The P-cluster of a MoFe protein crystal anaerobically soaked with eosin-Y and illuminated with blue LED light for 1 minute. The F$_o$-F$_c$ difference map (green mesh) shows positive difference density (indicated by an arrow) for the Fe6 in the P^{1+} conformation. 2F$_o$-F$_c$ electron density is shown as blue mesh.

Extended X-ray irradiation of MoFe protein crystals

Researchers strive to minimize photoreduction by X-ray irradiation of proteins during crystallographic data collection, which leads to a decrease in diffraction quality.[37, 38] This is especially true for proteins with metallocenters, which are often sensitive to radiation effects.[39–41] However, in the case of initially oxidized met-

alloproteins that require the capture of the reduced structure, this very effect has proven to be advantageous.[42–46] Thus, we explored whether MoFe protein crystals could be reduced by X-ray irradiation and whether this effect would manifest as changes in the conformations of the FeMo-cofactor or P-cluster.

We collected X-ray crystallographic datasets for extended periods of time and over many degrees of rotation on MoFe protein crystals. The data were then fractionated into individual crystallographic data sets and processed separately to generate "snapshots" of the crystal structures over time **(Fig. 7)**. In addition, we measured Fe X-ray absorption spectra of the crystals before and after each crystallographic dataset to monitor changes in the overall oxidation state of the clusters.

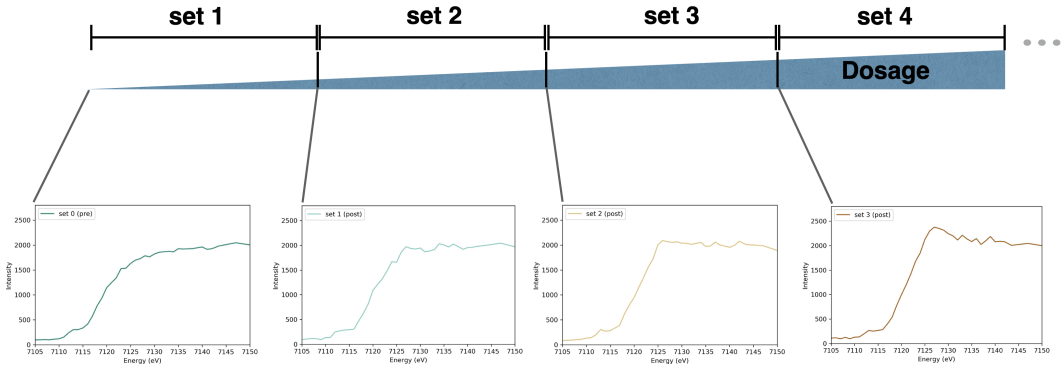


Figure 3.7: Experimental scheme for extended X-ray irradiation of MoFe protein crystals. Several full crystallographic datasets were collected back-to-back, equating to at least 4 full spheres of diffraction data. In between each diffraction dataset, a MAD scan was collected to monitor the Fe-edge position through the irradiation experiment. This experiment is described in detail following.

X-ray absorption spectroscopy (XAS) on protein crystals is possible at Stanford Synchrotron Radiation Lightsource beamline 12-2 for the purpose of determining optimal energies for multiwavelength anomalous diffraction (MAD) experiments. As the XAS measurements do not need to be highly accurate to estimate MAD energies, the number of points measured at the edge and the counting time were much lower than that of standard experiment done at a dedicated XAS beamline. However, we were able to improve the quality of our XAS measurements by increasing the counting time for each point (from 1 to at least 3 seconds) and by taking multiple scans that extend the spectrum to higher energies to enable better background subtraction **(Fig. 8)**.

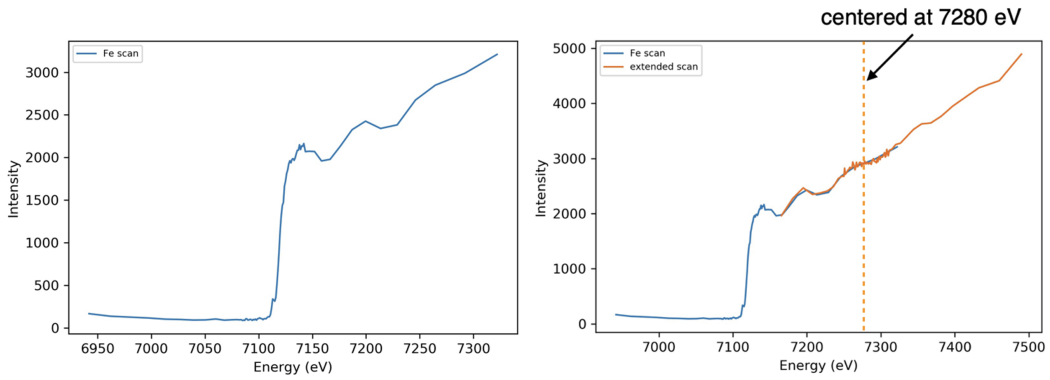


Figure 3.8: **Left)** XAS measured around the Fe X-ray absorption edge of a MoFe protein crystal at SSRL BL 12-2. **Right)** A second scan (orange) centered at the end of the previous scan (blue) extends the spectrum for better background subtraction.

To determine the upper limit of radiation MoFe protein crystals could withstand, we irradiated crystals continuously until there was almost complete loss of diffraction **(Fig. 9)**. Using a 50x50 µm beam at 7140 eV with detector distance of 200 mm, we collected full crystallographic data sets (360° rotation) at an oscillation angle of either 0.2° or 0.4°.

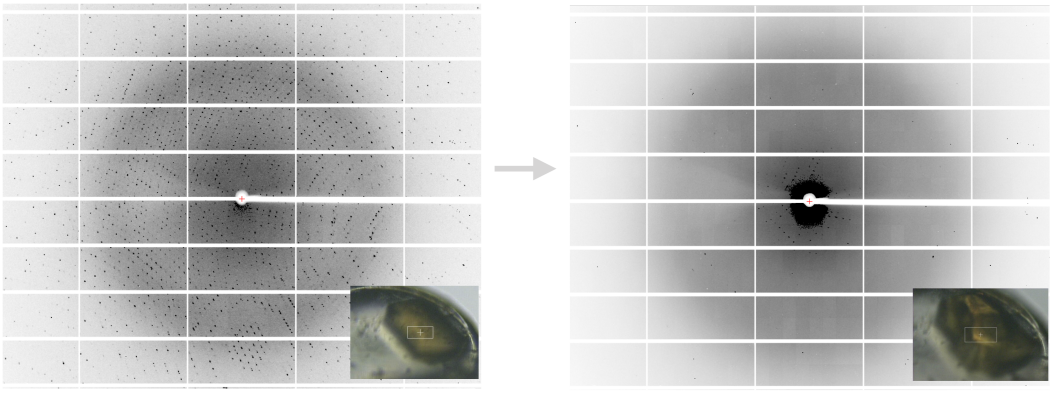


Figure 3.9: **Left)** Diffraction of a MoFe protein crystal before extended irradiation and **Right)** after extended irradiation. Inset shows the in-line camera image of the crystal. The crystal was enlarged after extended irradiation.

We collected XAS spectra before and after each dataset **(Fig. 10)**, taking care to collect XAS scans at the same goniometer position, since the crystal orientation could impact the details of the spectrum.[47] In this experiment XAS scan "set 0" was measured before any diffraction data collection and the other scans were collected at intervals between each diffraction dataset. In total, the crystal was irradiated for 110 min and experienced an estimated 5.03 x 10^{15} photons before almost entirely losing

diffraction. Using RADDOSE,[48] we estimated this amounted to 1.3 x 10^3 MGy, where the Garman limit before a protein typically experiences significant radiation damage is considered to be 30 MGy.[37]

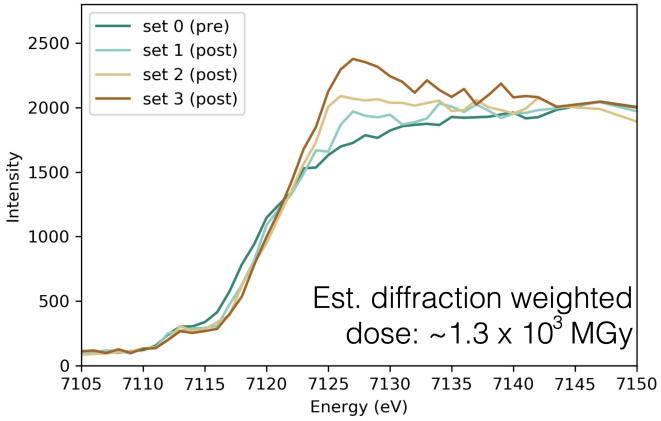
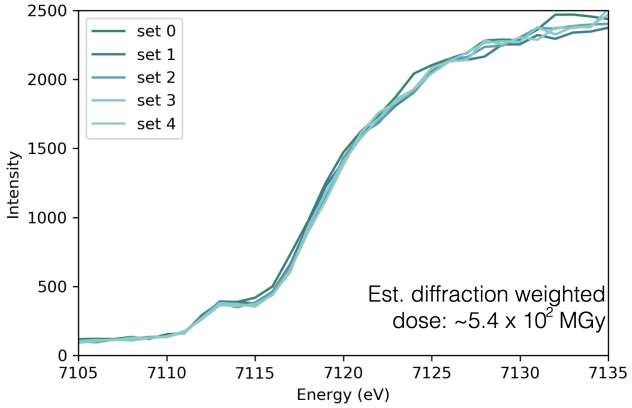


Figure 3.10: XAS of MoFe protein crystals around the Fe-edge before (set 0, dark blue) and at various points after (sets 1-3) extended X-ray irradiation.

Having established the upper limit of radiation a MoFe protein could withstand before almost entirely losing diffraction, we repeated the same experiment on a new crystal with smaller X-ray doses per diffraction dataset. Using the XAS scans between each diffraction dataset as a guide, we were able to collect 5 full crystallographic datasets before significantly changing the shape and position of Fe-edge **(Fig. 11**, left**)**. This crystal was irradiated for 18 mins in total and experienced an estimated 1.58 x 10^{15} photons and a diffraction weighted dose of roughly 5.4 x 10^2 MGy.



	Resolution (Å)	mean I/σI (highest res bin)
Set 1	2.1	16.1(8.5)
Set 2	2.1	11.5(6.6)
Set 3	2.1	12.7(3.3)
Set 4	2.1	11.8(1.6)
Set 5	2.2	11.5(1.1)

Figure 3.11: **Left)** XAS of a MoFe protein crystal before and after collection of multiple complete diffraction data sets. **Right)** Each resolution of each dataset and the mean I/σI of each dataset. The I/σI of the highest resolution bin of each dataset is shown in parenthesis.

We processed each individual crystallographic dataset to look for X-ray irradiation induced structural changes. Because the irradiation/diffraction data collection was performed at 7140 eV, the best resolution obtainable was 2.1 Å due to the SSRL 12-2 beamline specific parameters. However, each dataset (even set 5) had diffraction to the detector edge and had I/σI > 1 in the highest resolution bin **(Fig. 12**, right**)**. After having determined that all 5 datasets were high enough quality for further interpretation, we solved the MoFe protein structure from each data set. We compared the first X-ray diffraction data set (set 1) with the last (set 5) and found no major structural differences in the FeMo-cofactor itself nor in its surrounding residues.

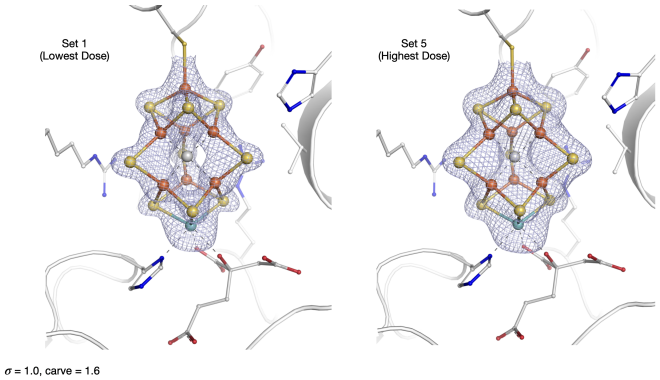


Figure 3.12: $2F_o$-F_c electron density (blue mesh) for the FeMo-cofactor from the **Left)** first dataset (set 1) and from the **Right)** last dataset (set 5). No major structural changes were observed in set 5, even after significant (5.4 x 10^2 MGy) of X-ray irradiation.

However, there were subtle changes in the P-clusters that could be observed by inspection of the Fe-anomalous difference maps. The comparison of the Fe-anomalous difference maps for each P-cluster between the first dataset (set 1, gray mesh in **Fig. 13**) and the last dataset (set 5, dark blue mesh in **Fig. 13**) revealed diminished electron density around Fe2, which coordinates to βCys95. When the Fe-anomalous maps for each dataset are displayed at the same σ-level (**Fig. 13** bottom), we saw that the anomalous signal for Fe2 diminished more than the anomalous signal for the neighboring atom Fe8.

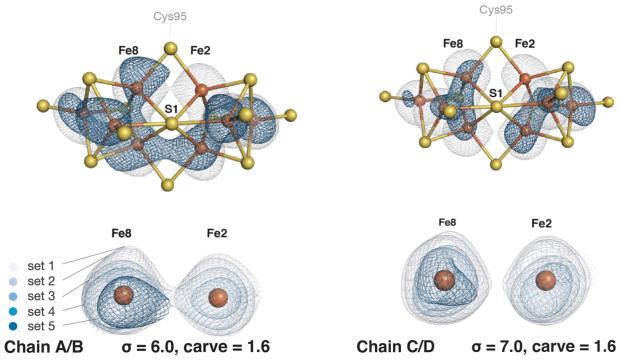


Figure 3.13: **Top)** Fe-anomalous difference density for set 1 (gray mesh) overlayed with Fe-anomalous difference density for set 5 (dark blue mesh) in each P-cluster. Anomalous signal is diminished for Fe2. **Bottom)** Fe-anomalous difference density for Fe8 and Fe2 overlaid from each dataset. The carve threshold indicates the radius of the sphere used to display electron density.

When the same extended irradiation experiment described above was performed on MoFe protein crystals that had opened (P^{2+}) P-clusters, we again saw diminished Fe-anomalous signal for Fe2. In contrast to the experiment starting with the closed P-cluster, when we started with opened P-cluster, we also saw diminished Fe-anomalous signal for Fe6 **(Fig. 14**, top).

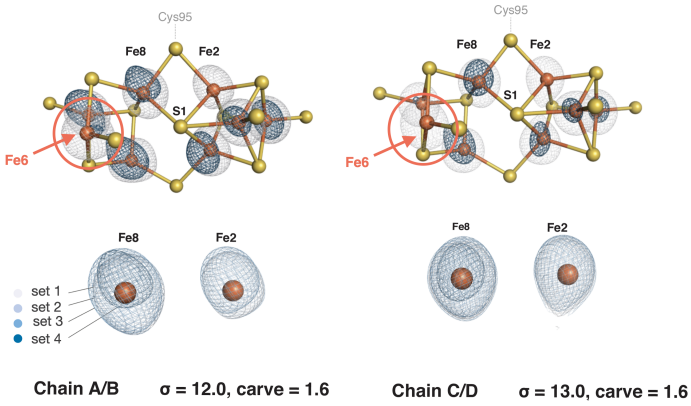


Figure 3.14: **Top)** Fe-anomalous difference density for set 1 (gray mesh) overlayed with Fe-anomalous difference density for set 5 (dark blue mesh) in each P-cluster from a MoFe protein crystal starting with fully opened (P^{2+}) P-cluster. Anomalous signal is diminished for both Fe2 and Fe6.

As these experiments consume extensive beamtime, we have limited datasets for further comparison. Thus, we do not make any inferences on what the site-specific decrease in Fe-anomalous signal could mean in the context of nitrogenase without further experiments. As a demonstration for future studies, we collected XAS spectra at various crystal orientations and observed differences in absorption intensity, emphasizing the importance of controlling for crystal orientation effects when comparing multiple spectra in these experiments **(Fig. 15)**.

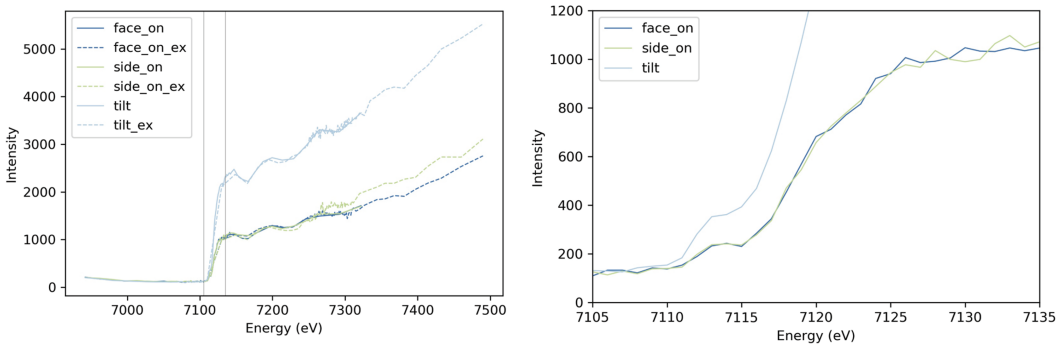


Figure 3.15: **Left)** Single crystal XAS of a MoFe protein crystal in three different orientations. **Right)** Zoom-in on the rising Fe-edge centered at 7120 eV.

The MoFe protein structures in the high-dose X-ray diffraction datasets had no large-scale structural changes compared to the low-lose datasets. And aside from differences in the Fe-anomalous maps, we saw no changes even in the residues surrounding the electron-dense metalloclusters. This raises the question of whether the MoFe protein is perhaps more resilient against radiation damage than other proteins. To learn what types of radiation damage a MoFe protein crystal might experience under standard diffraction data collection conditions, we calculated radiation damage difference electron density maps, using a function available in HKL3000 that highlights radiation damage-induced changes.[49] Using a standard data collection strategy with parameters such as 10% beam transmission at 12.685 keV at SSRL BL12-2 for 1800 frames and an exposure time of 0.2 s per frame, we calculated the radiation damage difference electron density map for MoFe protein. **Fig. 16** shows an example of a MoFe protein crystal structure overlaid with its radiation damage difference electron density map where negative peaks (blue mesh) represent atoms that were displaced, and positive peaks (orange mesh) represent atoms that become more ordered or electron dense as a result of X-ray irradiation. We saw negative (blue) difference density peaks around solvent exposed residues, such as αGlu427, which likely became decarboxylated. αGlu427 sits in the water channel that leads to the pool of waters around *R*-homocitrate. Moreover, we saw positive (orange) difference density peaks around both the FeMo-cofactors and P-clusters. Altogether, these data reveal, as expected, that changes do occur in MoFe protein crystals in response to X-ray irradiation. However, what is remarkable is the ability for MoFe protein crystals to maintain diffraction well beyond the estimated Garman limit, a point at which most crystals would typically not survive.

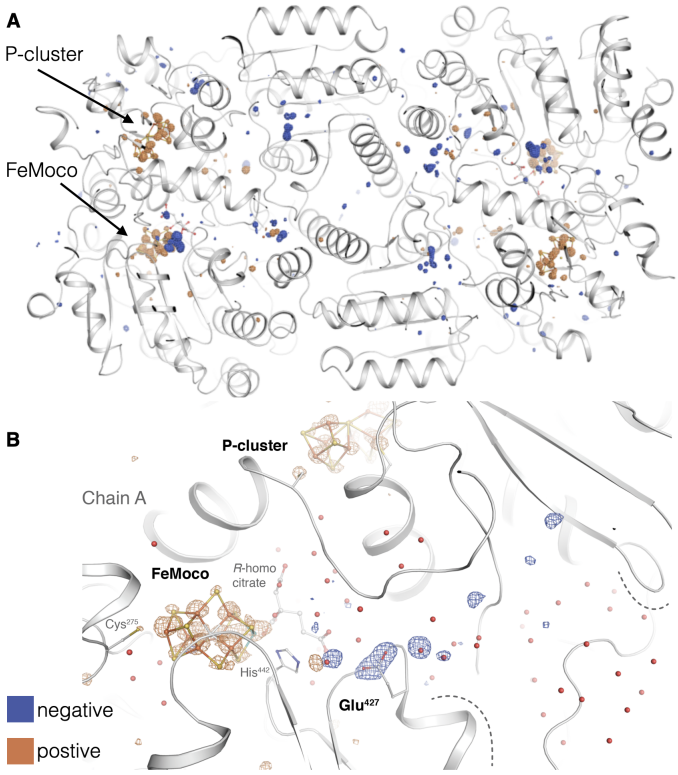


Figure 3.16: **Top)** Overall cartoon structure of the MoFe protein overlaid with the radiation damage difference electron density map for a crystal that was collected under standard conditions (not under highly extended X-ray irradiation). **Bottom)** Zoomed-in view of the FeMo-cofactor and water channel on one side of the MoFe protein tetramer. Negative difference density peaks (blue mesh) indicate regions with atoms that become disordered through the data collection. Positive difference density peaks (orange mesh) indicate regions that become more ordered or more electron rich through the data collection.

Radiation damage-free X-ray crystal structure of the MoFe protein

The advent of X-ray free electron lasers (XFELs) and serial diffraction methods allowed protein crystallographers to determine damage-free structures. Although XFELs are 10 orders of magnitude brighter than synchrotrons,[50] their ultra-short (fs) X-ray pulses enable the collection of a single diffraction image before significant sample damage has occurred.[51] The trade-off to "outrunning" radiation damage is the destruction of the sample, which undergoes Coulomb explosion after a single shot. Therefore, the use of serial crystallography and advancements in sample delivery have been pivotal to the success of XFEL experiments.

Sample delivery technology has been rapidly evolving. In recent years, structural biologists have favored three techniques to get their samples into the XFEL beam.

The first technique is widely used and involves a liquid jet stream of crystal slurries, which either continually flows or flows in pulses that form a series of droplets.[52, 53] The stream of droplets can be timed to coincide with the pulses of the XFEL beam. A major downside of this method is the requirement for large amounts of sample to sustain a constant stream, and the low hit-rate of crystals in the sample leads to significant sample loss. Thus, other methods have emerged that require less sample. A second method involves a fixed-target goniometer mounted with prefabricated mesh loops, which resemble a more standard synchrotron-based crystallography experiment.[54, 55] A third method, pioneered by the Kern lab, utilizes a tape-drive system that uses an acoustic droplet handler to deposit nanoliter-sized droplets of crystal slurries onto a continuous roll of Kapton tape.[56, 57] The tape then moves the droplets into the X-ray beam in time with the XFEL pulses.

To obtain radiation-damage free structures of the nitrogenase MoFe protein, we collected crystallographic data using the latter two sample delivery methods: using a fixed-target goniometer and a tape-drive system. The fixed-target experiments were performed in collaboration with Dr. Christopher O. Barnes, Dr. Pamela J. Bjorkman, Dr. Aina E. Cohen, and the Linac Coherent Light Source (LCLS) Macromolecular Femtosecond Crystallography team at SLAC National Accelerator Laboratory. The tape-drive experiments were performed in collaboration with Dr. Jan Kern, Dr. Stephen Keable, the Kern lab tape-drive sample delivery team, and the Coherent X-ray Imaging team at SLAC National Accelerator Laboratory.

Fixed-target serial diffraction studies of nitrogenase MoFe protein

Our initial XFEL experiments implemented a goniometer-based set up that enabled us to bring crystal samples that were pre-loaded in mesh loops.[54, 55] These loops were sent to Caltech prior to the allocated beamtime and we were able to load them with our MoFe protein crystals under strictly anoxic conditions and freeze them for shipping. Each loop accommodated several crystals whose positions were annotated in an atlas using the Blu-Ice graphical interface. Because LCLS beamtimes are allocated in 12 hr shifts that are alternated with neighboring experimental hutches, the atlas annotation, which did not require X-ray beam, was completed during the off-shift to maximize the efficiency of the active beamtime. During beamtime, the atlas for each pin was then loaded and still diffraction images were collected in rastering mode, yielding an almost perfect hit-rate. This method has the least sample loss compared to those previously mentioned; however, the mesh format of the loop led to preferential orientation of the crystals. Since our MoFe protein

crystals had plate-shaped morphology and because we could not rotate the mesh loop more than +/- 30 degrees, our crystals, which diffracted past the detector edge (at 1.6 Å resolution), had a missing wedge of data that led to difficulties in structure refinement. Therefore, we turned to the tape-drive sample delivery system that implemented MoFe protein crystal slurries.

Tape-drive serial diffraction studies of nitrogenase MoFe protein

Slurries of MoFe protein crystals were prepared by setting up hanging drop vapor diffusion experiments in 24-well tissue culture trays **(Fig. 17)** in a Coy Labs anoxic chamber containing a 97% Ar and 3% H_2 atmosphere. As the crystal drops were eventually pooled together, a single optimized crystallization condition was used for all experiments, containing 14% (w/v) PEG 4000, 0.6 M NaCl, 0.2 M imidazole/malate pH 8.0. The protein solution was prepared in buffer containing 50 mM Tris-HCl pH 7.5 and 200 mM NaCl, supplemented with 5 mM dithionite. To minimize reagents and space, 5-8 crystal drops of 2 µL volume were applied per hanging drop well containing up to 30 mg/mL MoFe protein (mixed 1:1 with reservoir solution). After crystallization, each crystal drop was resuspended with mother liquor and carefully collected and pooled into Eppendorf tubes. After the crystals settled, roughly 30% of each sample contained visible solid crystal pellet, which was dark brown in color.

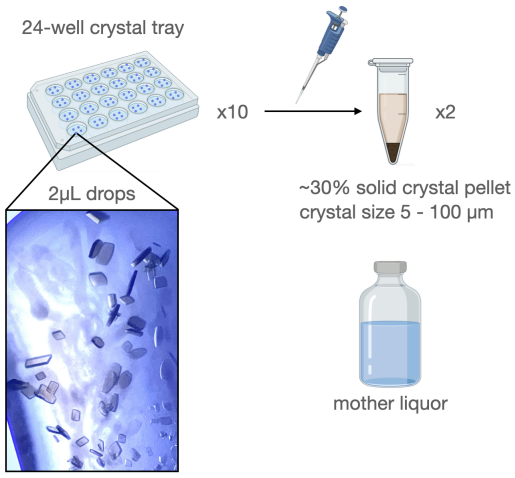


Figure 3.17: Workflow for preparation of MoFe protein crystal slurries for XFEL beamtime. Crystals were prepared by hanging drop vapor diffusion and collected by pipetting into Eppendorf tubes. Excess mother liquor was allocated for down-stream sample preparation and buffer matching at the beamline.

In total, three samples were prepared for the beamtime: 1) 500 µL of sample containing 150 µL of solid crystal pellet, 2) 600 µL of sample containing 250 µL of pellet, and 3) 300 µL of sample containing 50 µL of pellet. Sample 3) had lower quality, mosaic crystals and was kept as a separate sample to test further manipulations. Excess mother liquor was stored for additional experiments and to sparge the sample transfer line at the XFEL beamline.

Transmission electron microscopy of MoFe protein crystal slurries

To determine whether the crystals were in the target size range that was needed for the tape drive delivery system (20-100 µm, optimally 50-75 µm) and to test various crystal slurry preparations, we used transmission electron microscopy (TEM) to visualize our slurries.[58] Four samples were prepared: 1) crystal slurry without crushing, 2) slurry crushed with a glass crystal pestle, 3) slurry vortexed with a glass bead at max speed for 1 min, and 4) a mother liquor control. Each sample was treated and then diluted with mother liquor before transferring to negative stain grids. The TEM grid preparation and imaging were graciously performed by Dr. Rebeccah Warmack. The samples were applied to freshly glow discharged Ultrathin C film on holey carbon support film 400 mesh, copper grids (Ted Pella, Inc.) and stained with 3% uranyl acetate **(Fig. 18)**.

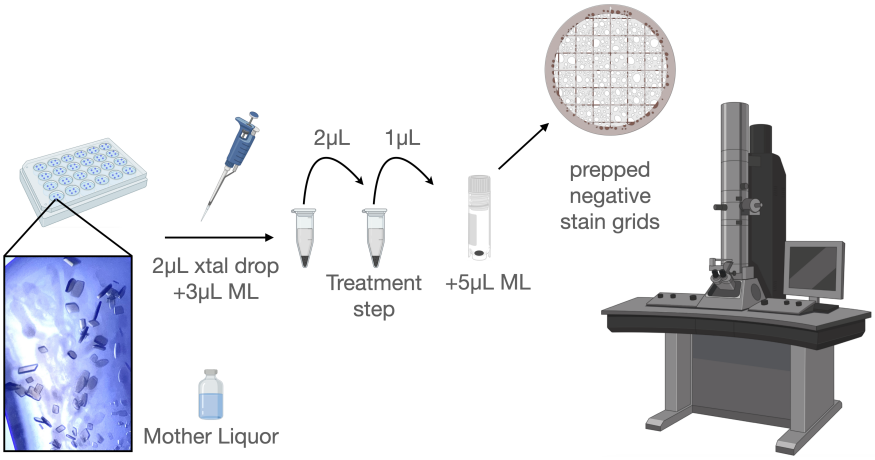


Figure 3.18: Workflow for crystal slurry sample treatment and grid preparation for TEM.

The crystals that were not further crushed were the best quality and within the target size range (roughly 20 μm in diameter). It is likely that the shear forces from pipetting the crystallization drops to generate the crystal slurries broke the crystals into smaller sizes, as they were initially about 200-300 μm in diameter. The TEM images revealed lattice planes and defined edges of the MoFe protein crystals even after slurry preparation **(Fig. 19)**.

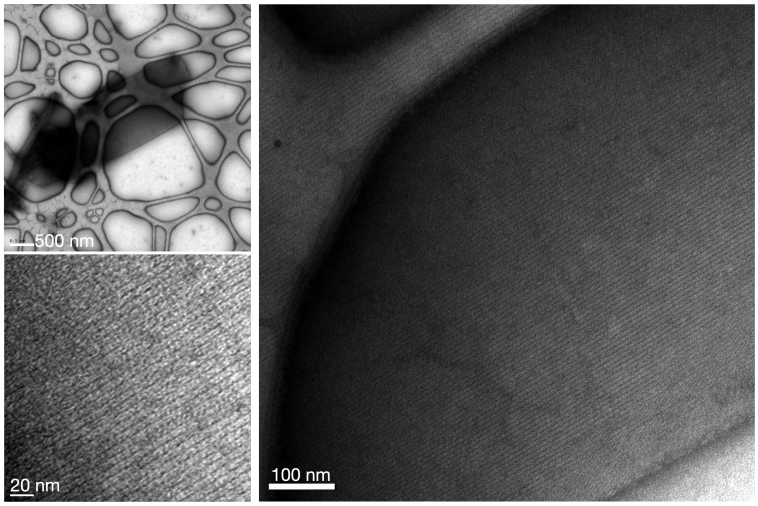


Figure 3.19: Negative stain TEM images of MoFe protein crystals after crystal slurry preparation.

As crystal slurries could not be frozen, they were transported to SLAC National Laboratory via air-tight 2.5 L rectangular jars (Thermo Scientific) and kept anoxic using disposable AnaeroPack gas generators (Thermo Scientific). Each box was monitored for anaerobicity with resazurin indicator strips, which turn from white to magenta in the presence of oxygen. The crystal slurries were transferred to an anaerobic chamber immediately upon arrival at the facility. Roughly 5 days after the initial preparation of the crystal slurries, cloudy white precipitate formed in the mother liquor layer above the crystal pellet. Crystal slurries were washed by decanting the cloudy white precipitate layer and replenished with fresh mother liquor supplemented with 5 mM dithionite.

To prepare the crystal slurries for the XFEL experiment, excess mother liquor was decanted such that roughly 50% of the volume of the crystal slurry contained solid crystal pellet. The slurry was then filtered by passage through a 250 μm diameter capillary attached to a 1 mL Hamilton lock syringe. Filtration was necessary to ensure the sample would not clog the microfluidics system used to transfer the

sample to the acoustic delivery system coupled to the tape drive. The filtrate was then loaded into a fresh Hamilton syringe for transfer to the beamline. A small aliquot of this sample was viewed under a microscope and revealed that the crystals in the slurry were starting to deteriorate **(Fig. 20B)**. The final sample contained 400 µL of crystal slurry **(Fig. 20C)**. Slurry 3) with the mosaic crystals was excluded from the final sample. A second syringe of fresh mother liquor supplemented with 5 mM dithionite was prepared to purge the sample transfer lines at the beamline **(Fig. 20E)**. After the final sample was prepared, the sample turned white, presumably due to additional precipitation during the subsequent 2 hour-wait for the beam to become available for our sample run.

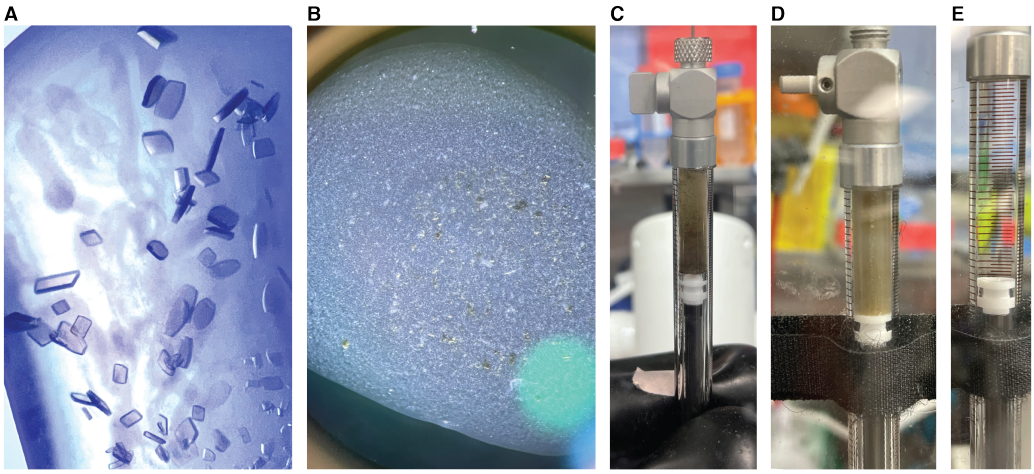


Figure 3.20: **A)** Initial MoFe protein crystals. **B)** Crystal slurry 5 days after initial slurry preparation and after filtration. **C)** Final 400 µL sample after filtration. **D)** Sample turned from dark to light brown during the wait for beamtime. **E)** Sample matching buffer used to sparge the fluidics for crystal slurry delivery.

The MoFe protein crystal slurry was deposited onto the tape drive at a frequency of 30 Hz and belt speed of 200 mm/s. The drop delivery device was encased in a tank with helium atmosphere with roughly 240 ppm O_2. In total, 63,668 images were collected with 31,779 hits, of which 18,342 images were merged. The nitrogenase crystals adopted a uniform unit cell consisting of unit cell parameters a = 81.8 Å, b = 130.4 Å, c = 108.5 Å and $\alpha = 90°$, $\beta = 110.9°$, $\gamma = 90°$ **(Fig. 21)**.

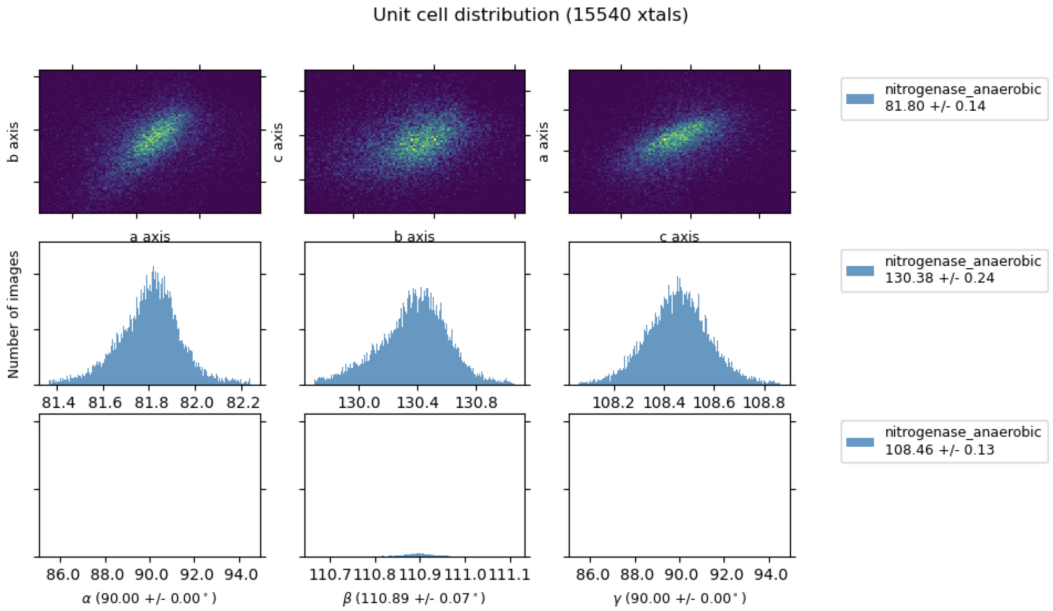


Figure 3.21: Unit cell distributions for MoFe protein crystal slurry.

Fe Kα X-ray emission spectroscopy (XES) of the MoFe protein crystals was collected simultaneously to the diffraction data collection. The MoFe protein, which has 32 Fe atoms per heterotetramer,[59] had strong signal for Fe **(Fig. 22)**.

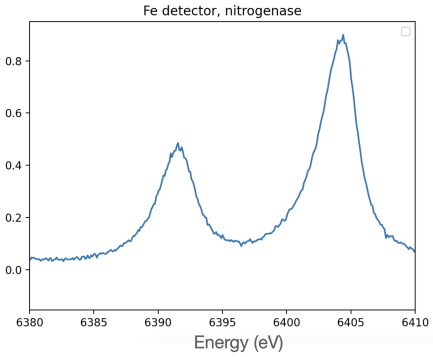


Figure 3.22: Fe Kα_1 (right peak) and Kα_2 (left peak) XES spectra of the MoFe protein microcrystals.

The data were cut at the highest resolution bin with 10 times multiplicity and determined to be 1.85 Å resolution. Phases were obtained using the protein atoms of PDB 3U7Q. The damage-free structure of MoFe protein had no major conformational changes relative to the synchrotron-based structures, such as PDB 3U7Q **(Fig. 23)**. Unambiguous electron density was observed for all the cofactor and P-cluster atoms. The P-clusters in both subunits were in the open conformation, indicating

the sample had been oxidized before data collection.

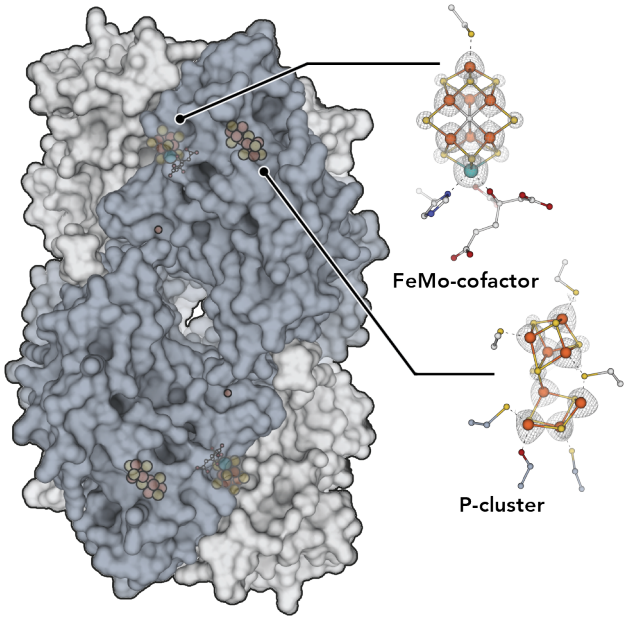


Figure 3.23: Damage-free X-ray crystal structure of *Av* MoFe protein by XFEL radiation. The P-cluster adopts the fully open conformation. $2F_o$-F_c density for the metalloclusters shown as gray mesh.

The damage-free structure of MoFe protein confirms that our synchrotron crystallographic datasets have minimal radiation damage, as the XFEL structure had no significant differences compared to the synchrotron-acquired structure, PDB 3U7Q. The two structures which align with a root mean square deviation of 0.28 Å and are nearly identical. The majority of the displacements in the structure appear at the periphery of the heterotetramer. Vectors drawn between the aligned Cα of all atoms in the two structures exhibit the largest differences in the helices and sheets in the α-subunit, with the damage-free structure being slightly expanded compared to PDB 3U7Q (**Fig. 24**). As the vector lengths (which correspond to the distance between the Cαs in the two structures) are not homogeneous throughout the structure, it is unlikely that the expansion of the unit cell could be explained by the beamline parameters alone.

In summary, we found that introducing strong chemical reductants to MoFe protein crystals resulted in conversion of the P-cluster from the closed to open conformation. Soaking Eu(II)-DTPA and Ti(III)-citrate to oxidized MoFe protein crystals did not convert the opened P-cluster back to the closed conformation. We found that soaking MoFe protein crystals with eosin dyes, followed by illumination with blue

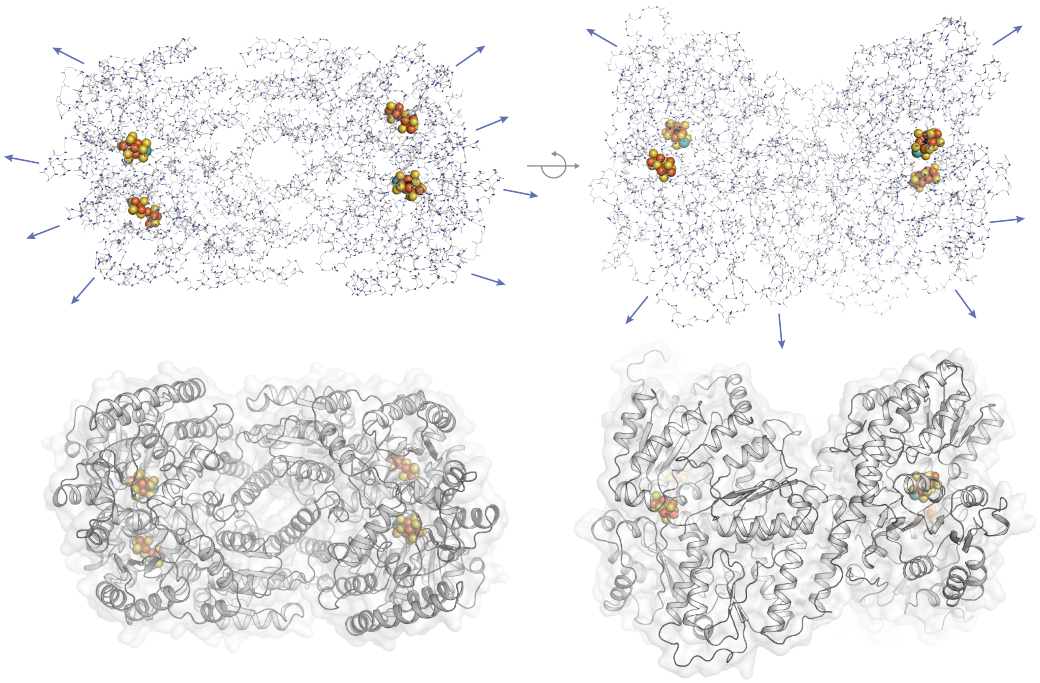


Figure 3.24: **Top)** The vector lines corresponding to equivalent alignment atoms in PyMOL represent the difference in the atom position between the damage-free structure relative to PDB 3U7Q (not shown). Arrows indicate direction of greatest alignment difference. **Bottom)** depicts the ribbon diagram of the damage-free structure for reference.

LED light, resulted in a small population of P^{1+} P-cluster formation. Extended X-ray irradiation of MoFe protein crystals resulted in diminished electron density for Fe2 of the P-cluster in the reduced MoFe protein and diminished electron density for Fe2 and Fe6 of the P-cluster in the oxidized MoFe protein. Finally, we determined the damage-free, room temperature X-ray crystal structure of the resting state MoFe protein using an X-ray free electron laser, which serves as a valuable basis set for future experiments investigating activated states of nitrogenase.

3.3 Methods

Av MoFe protein purification for crystallography

Nitrogenase MoFe protein was expressed and purified from *Azotobacter vinelandii* (*Av*) using previously described procedures.[32, 60] Briefly, cultures of *Av* were grown in Burk's minimal medium containing nitrogen limiting conditions (1.3 mM NH_4Cl). All steps after cell lysis were performed under strict anaerobic conditions using either anoxic chambers (COY Laboratory products) containing a 97/3% Ar/H_2 gas mixture or a 100% Ar modified Schlenk line. Cells were lysed by Emulsiflex

(Avestin) and subsequently centrifuged to separate cell debris. MoFe protein was isolated from the soluble fraction by ion exchange chromatography and further purified by size exclusion chromatography. Purified protein was stored in buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and 5mM dithionite and flash frozen in liquid nitrogen until further use.

Av MoFe protein crystallization

Purified MoFe protein was concentrated to 30-40 mg/mL in buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and freshly supplemented 5 mM sodium dithionite. MoFe protein concentrations were estimated using the 410 nm absorbance (extinction coefficient 76,000 $M^{-1}cm^{-1}$). Crystallization trays were prepared by hanging drop vapor diffusion in pre-greased 24-well tissue culture plates in a Coy Labs anaerobic chamber containing 97% Ar and 3% H_2 atmosphere. Crystal drops were prepared by addition of 1 µL of protein stock solution mixed 1:1 with reservoir solution. Each drop was streak seeded with MoFe protein seed crystals using a natural fiber seeding tool. Optimized MoFe protein crystallization solution contained 14% (w/v) PEG 4000, 0.6 M NaCl, and 0.2 M Imidazole/malate pH 8.0. Seed stocks were prepared by resuspention of one 2 µL crystal drop containing small (<0.1 mm diameter), well nucleated crystals into 50 µL of reservoir solution. Crystals were crushed by vortexing with glass seed beads for ~10 s at max speed. MoFe protein crystals were dark brown and grew to ~0.1-0.2 mm in diameter within 24 hours under these conditions. MoFe protein crystals grow up to 0.4 mm in diameter if not harvested within 1 week. Crystals were cryo-protected by sequentially soaking into solutions of mother liquor supplemented with increasing MPD concentration (8-12% MPD).

Air incubation of MoFe protein crystals

Air-exposure of the MoFe protein crystals was performed by removing MoFe protein crystals from the glove box and flash-freezing them in liquid nitrogen at 30 second intervals. The crystals were cryo-protected using aerobically prepared cryo-solutions before freezing. After ~40 minutes the crystals turned from brown to clear in color. Clear crystals exhibited a significant loss of diffraction quality.

Eu(II)-DTPA, Ti(III)-citrate, and O_2 reactions MoFe protein crystals

Eu(II)-diethylenetriaminepenta-acetic acid (Eu(II)-DPTA) soak experiments into MoFe protein crystals were performed by preparing a 10 mM stock of Eu(II) chloride

and a 10 mM stock of DTPA in buffer containing 50 mM Tris-HCl pH 7.5 and 100 mM NaCl. 1-2 µL of each of these stocks were directly added to the 2 µL crystal drops and allowed to soak for up to 2 hours before the crystals were cryo-protected and flash frozen in liquid nitrogen. An 88 mM stock of Ti(III)-citrate that had been previously prepared and frozen in liquid nitrogen was anaerobically thawed and diluted to 10 mM in buffer containing 50 mM Tris-HCl pH 7.5 and 100 mM NaCl. 1-2 µL of the 10 mM Ti(III)-citrate stock was directly added to the 2 µL crystal drops. Crystals that were first oxidized by air-exposure were taken out of the glove box and incubated in air for less than 10 minutes before being cycled back into the glove box and immediately soaked with reductant before freezing.

W-photosensitizer soaks in MoFe protein crystals

Tungsten (0) arylisocyanide photoreductant, W(CNDippPh^{OMe3})$_6$ (CNDippPh^{OMe3} = 4-(3,4,5-trimethoxyphenyl)-2,6-diisopropylphenylisocyanide)[35] that was developed in lab of Dr. Harry B. Gray and synthesized by Dr. Javier Fajardo was anaerobically dissolved by addition of a grain of the synthesized compound (< 1 mg) in solutions containing a mixture of crystal mother liquor (15% (w/v) PEG4000, 0.2 M NaCl, and 0.1 M imidazole/malate pH 8.0) and 20-50% dimethylsulfoxide (DMSO) or 20% ethanol. Crystals were soaked for up to 1 hour before being cryo-protected and flash-frozen in liquid nitrogen.

Eosin dye soaks in MoFe protein crystals

Eosin-Y (691.85 g/mol) and eosin-B (624.06 g/mol) stocks were prepared by anaerobically making 50 mM solutions in buffer containing 50 mM Tris-HCl pH 7.5 and 100 mM NaCl. The crystal soaking solutions of eosin were prepared by addition of 2 µL of the 50 mM eosin stock to 16 µL of well solution and 2 µl of the cryo-protectant MDP (for a final concentration of 5 mM eosin) or addition of 4 µL of the 50 mM eosin stock to 14 µL of well solution and 2 µL of MDP (for a final concentration of 10 mM eosin). As a control, crystal soaking solutions with MDP, lacking eosin dyes were prepared. The crystals were soaked at various time-points for up to 1 hour in eosin soaking solutions before being flash frozen in liquid nitrogen. For the photoactivation experiments, eosin soaked crystals were anaerobically illuminated with blue LED light for 0 mins (control), 3 mins, or 10 mins before being flash frozen in liquid nitrogen.

MoFe protein X-ray data collection and structure solution

X-ray diffraction datasets were collected at 100 K at either the Stanford Synchrotron Radiation Light Source (SSRL) at beamline 12-2 or at the Argonne National Laboratory Advanced Photon Source (APS) GM/CA beamline 23-ID-B. Data collected at SSRL had a general collection strategy of 90% attenuation, with a 50x50 micron beam set at an energy of 12.658 keV. Each dataset had a minimum of 1800 frames (1 full sphere) at 0.2° oscillation. Very high resolution sets (~1.2 Å) were collected at an energy of 15.0 or 17.0 keV. Data collected at APS had 1x attenuation, with a 25x25 micron beam set at an energy of 12.66keV. Either one or three full spheres of data were collected at 0.4° oscillation. Data were indexed, integrated, and scaled using HKL3000 [61] or XDS.[62] Phases were obtained by either molecular replacement (MR) with PHASER [63] or by rigid body refinement. Initial MR was performed using the protein atoms from a previously solved MoFe protein structure, PDB 3U7Q,[32] and subsequent MR was performed using the high resolution MoFe protein structure described in **Chapter 2**. Model building was performed in Coot [64] and both REFMAC5 [65] and Phenix [66] were used for refinement. Software was installed and configured using the SBGrid package manager.[67] All crystallographic figures were prepared using PyMOL molecular graphics software package (Schödinger, LLC).

References

(1) Burgess, B. K.; Lowe, D. J. Mechanism of Molybdenum Nitrogenase. *Chemical Reviews* **1996**, *96*, 2983–3012.

(2) Lowe, D. J.; Thorneley, R. N. The mechanism of *Klebsiella pneumoniae* nitrogenase action. Pre-steady-state kinetics of H_2 formation. *Biochemical Journal* **1984**, *224*, 877–886.

(3) Thorneley, R.; Lowe, D. Molybdenum enzymes. *Metal Ions in Biology* **1985**, *7*, 221–284.

(4) Ljones, T.; Burris, R. H. Atp hydrolysis and electron transfer in the nitrogenase reaction with different combinations of the iron protein and the molybdenum-iron protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1972**, *275*, 93–101.

(5) Bulen, W. A.; Burns, R. C.; LeComte, J. R. NITROGEN FIXATION: HYDROSULFITE AS ELECTRON DONOR WITH CELL-FREE PREPARATIONS OF AZOTOBACTER VINELANDII AND RHODOSPIRILLUM RUBRUM. *Proceedings of the National Academy of Sciences* **1965**, *53*, 532–539.

(6) Burns, R.; Bulen, W. ATP-Dependent hydrogen evolution by cell-free preparations of Azotobacter Vinelandii. *Biochimica et Biophysica Acta (BBA) - Enzymology and Biological Oxidation* **1965**, *105*, 437–445.

(7) Mortenson, L. E. FERREDOXIN AND ATP, REQUIREMENTS FOR NITROGEN FIXATION IN CELL-FREE EXTRACTS OF CLOSTRIDIUM PASTEURIANUM. *Proceedings of the National Academy of Sciences* **1964**, *52*, 272–279.

(8) Mortenson, L. E. FERREDOXIN AND ATP, REQUIREMENTS FOR NITROGEN FIXATION IN CELL-FREE EXTRACTS OF CLOSTRIDIUM PASTEURIANUM. *Proceedings of the National Academy of Sciences* **1964**, *52*, 272–279.

(9) Martin, A. E.; Burgess, B. K.; Iismaa, S. E.; Smartt, C. T.; Jacobson, M. R.; Dean, D. R. Construction and characterization of an Azotobacter vinelandii strain with mutations in the genes encoding flavodoxin and ferredoxin I. *Journal of Bacteriology* **1989**, *171*, 3162–3167.

(10) Yates, M. Electron transport to nitrogenase in *Azotobacter chroococcum* : Azotobacter flavodoxin hydroquinone as an electron donor. *FEBS Letters* **1972**, *27*, 63–67.

(11) Lowe, D. J.; Ashby, G. A.; Brune, M.; Knights, H.; Webb, M. R.; Thorneley, R. N. F. In *Nitrogen Fixation: Fundamentals and Applications*, Tikhonovich, I. A., Provorov, N. A., Romanov, V. I., Newton, W. E., Eds., Series Title: Current Plant Science and Biotechnology in Agriculture; Springer Netherlands: Dordrecht, 1995; Vol. 27, pp 103–108.

(12) Lough, S.; Burns, A.; Watt, G. D. Redox reactions of the nitrogenase complex from Azotobacter vinelandii. *Biochemistry* **1983**, *22*, 4062–4066.

(13) Rutledge, H. L.; Tezcan, F. A. Electron Transfer in Nitrogenase. *Chemical Reviews* **2020**, *120*, 5158–5193.

(14) Lanzilotta, W. N.; Seefeldt, L. C. Changes in the Midpoint Potentials of the Nitrogenase Metal Centers as a Result of Iron Protein-Molybdenum-Iron Protein Complex Formation. *Biochemistry* **1997**, *36*, 12976–12983.

(15) Pierik, A. J.; Wassink, H.; Haaker, H.; Hagen, W. R. Redox properties and EPR spectroscopy of the P clusters of Azotobacter vinelandii MoFe protein. *European Journal of Biochemistry* **1993**, *212*, 51–61.

(16) Watt, G. D.; Burns, A.; Lough, S.; Tennent, D. L. Redox and spectroscopic properties of oxidized MoFe protein from Azotobacter vinelandii. *Biochemistry* **1980**, *19*, 4926–4932.

(17) Lanzilotta, W. N.; Christiansen, J.; Dean, D. R.; Seefeldt, L. C. Evidence for coupled electron and proton transfer in the [8Fe-7S] cluster of nitrogenase. *Biochemistry* **1998**, *37*, 11376–11384.

(18) Mayhew, S. G. The Redox Potential of Dithionite and SO-2 from Equilibrium Reactions with Flavodoxins, Methyl Viologen and Hydrogen plus Hydrogenase. *European Journal of Biochemistry* **1978**, *85*, 535–547.

(19) Vincent, K. A.; Tilley, G. J.; Quammie, N. C.; Streeter, I.; Burgess, B. K.; Cheesman, M. R.; Armstrong, F. A. Instantaneous, stoichiometric generation of powerfully reducing states of protein active sites using Eu(ii) and polyaminocarboxylate ligands. *Chemical Communications* **2003**, 2590.

(20) Danyal, K.; Inglet, B. S.; Vincent, K. A.; Barney, B. M.; Hoffman, B. M.; Armstrong, F. A.; Dean, D. R.; Seefeldt, L. C. Uncoupling Nitrogenase: Catalytic Reduction of Hydrazine to Ammonia by a MoFe Protein in the Absence of Fe Protein-ATP. *Journal of the American Chemical Society* **2010**, *132*, 13197–13199.

(21) Uppal, R.; Incarvito, C. D.; Lakshmi, K. V.; Valentine, A. M. Aqueous Spectroscopy and Redox Properties of Carboxylate-Bound Titanium. *Inorganic Chemistry* **2006**, *45*, 1795–1804.

(22) Solomon, J. B.; Rasekh, M. F.; Hiller, C. J.; Lee, C. C.; Tanifuji, K.; Ribbe, M. W.; Hu, Y. Probing the All-Ferrous States of Methanogen Nitrogenase Iron Proteins. *JACS Au* **2021**, *1*, 119–123.

(23) Angove, H. C.; Yoo, S. J.; Münck, E.; Burgess, B. K. An All-ferrous State of the Fe Protein of Nitrogenase: INTERACTION WITH NUCLEOTIDES AND ELECTRON TRANSFER TO THE MoFe PROTEIN. *Journal of Biological Chemistry* **1998**, *273*, 26330–26337.

(24) Yang, Z.-Y.; Badalyan, A.; Hoffman, B. M.; Dean, D. R.; Seefeldt, L. C. The Fe Protein Cycle Associated with Nitrogenase Catalysis Requires the Hydrolysis of Two ATP for Each Single Electron Transfer Event. *Journal of the American Chemical Society* **2023**, *145*, 5637–5644.

(25) Kim, J.; Woo, D.; Rees, D. C. X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. *Biochemistry* **1993**, *32*, 7104–7115.

(26) Smith, B. E.; Lang, G. Mössbauer spectroscopy of the nitrogenase proteins from *Klebsiella pneumoniae* . Structural assignments and mechanistic conclusions. *Biochemical Journal* **1974**, *137*, 169–180.

(27) Zimmermann, R.; Münck, E.; Brill, W. J.; Shah, V. K.; Henzl, M. T.; Rawlings, J.; Orme-Johnson, W. H. Nitrogenase X: Mössbauer and EPR studies on reversibly oxidized MoFe protein from Azotobacter vinelandii OP Nature of the iron centers. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1978**, *537*, 185–207.

(28) McLean, P. A.; Papaefthymiou, V.; Orme-Johnson, W. H.; Münck, E. Isotopic hybrids of nitrogenase. Mössbauer study of MoFe protein with selective 57Fe enrichment of the P-cluster. *Journal of Biological Chemistry* **1987**, *262*, 12900–12903.

(29) Keable, S. M.; Zadvornyy, O. A.; Johnson, L. E.; Ginovska, B.; Rasmussen, A. J.; Danyal, K.; Eilers, B. J.; Prussia, G. A.; LeVan, A. X.; Raugei, S.; Seefeldt, L. C.; Peters, J. W. Structural characterization of the P1+ intermediate state of the P-cluster of nitrogenase. *Journal of Biological Chemistry* **2018**, *293*, 9629–9635.

(30) Peters, J. W.; Stowell, M. H. B.; Soltis, S. M.; Finnegan, M. G.; Johnson, M. K.; Rees, D. C. Redox-Dependent Structural Changes in the Nitrogenase P-Cluster † ‡. *Biochemistry* **1997**, *36*, 1181–1187.

(31) Owens, C. P.; Katz, F. E. H.; Carter, C. H.; Oswald, V. F.; Tezcan, F. A. Tyrosine-Coordinated P-Cluster in *G. diazotrophicus* Nitrogenase: Evidence for the Importance of O-Based Ligands in Conformationally Gated Electron Transfer. *Journal of the American Chemical Society* **2016**, *138*, 10124–10127.

(32) Spatzal, T.; Aksoyoglu, M.; Zhang, L.; Andrade, S. L. A.; Schleicher, E.; Weber, S.; Rees, D. C.; Einsle, O. Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science* **2011**, *334*, 940–940.

(33) SYu, D.; Syrtsova, L. A.; Uzenskaja, A. M.; Likhtenstein, G. I. The photoreduction of nitrogenase. *Biochemical Journal* **1993**, *290*, 627–631.

(34) Syrtsova, L. A.; Timofeeva, E. A. Electron transfer coupled with ATP hydrolysis in nitrogenase. *Russian Chemical Bulletin* **2001**, *50*, 1789–1794.

(35) Fajardo, J.; Barth, A. T.; Morales, M.; Takase, M. K.; Winkler, J. R.; Gray, H. B. Photoredox Catalysis Mediated by Tungsten(0) Arylisocyanides. *Journal of the American Chemical Society* **2021**, *143*, 19389–19398.

(36) Syrtsova, L. A.; Nadtochenko, V. A.; Denisov, N. N.; Timofeeva, E. A.; Shkondina, N. I.; Gak, V. Y. Kinetics of elementary steps of electron transfer in nitrogenase in the presence of a photodonor. *Biochemistry. Biokhimiia* **2000**, *65*, 1145–1152.

(37) Owen, R. L.; Rudiño-Piñera, E.; Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences* **2006**, *103*, 4912–4917.

(38) De la Mora, E.; Coquelle, N.; Bury, C. S.; Rosenthal, M.; Holton, J. M.; Carmichael, I.; Garman, E. F.; Burghammer, M.; Colletier, J.-P.; Weik, M. Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proceedings of the National Academy of Sciences* **2020**, *117*, 4142–4151.

(39) Corbett, M. C.; Latimer, M. J.; Poulos, T. L.; Sevrioukova, I. F.; Hodgson, K. O.; Hedman, B. Photoreduction of the active site of the metalloprotein putidaredoxin by synchrotron radiation. *Acta Crystallographica Section D Biological Crystallography* **2007**, *63*, 951–960.

(40) Yano, J. K.; Hsu, M.-H.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nature Structural & Molecular Biology* **2005**, *12*, 822–823.

(41) Champloy, F.; Gruber, K.; Jogl, G.; Kratky, C. XAS spectroscopy reveals X-ray-induced photoreduction of free and protein-bound B $_{12}$ cofactors. *Journal of Synchrotron Radiation* **2000**, *7*, 267–273.

(42) Schlichting, I. The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science* **2000**, *287*, 1615–1622.

(43) Berglund, G. I.; Carlsson, G. H.; Smith, A. T.; Szöke, H.; Henriksen, A.; Hajdu, J. The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **2002**, *417*, 463–468.

(44) Logan, D. T.; Su, X.-D.; Åberg, A.; Regnström, K.; Hajdu, J.; Eklund, H.; Nordlund, P. Crystal structure of reduced protein R2 of ribonucleotide reductase: the structural basis for oxygen activation at a dinuclear iron site. *Structure* **1996**, *4*, 1053–1064.

(45) Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J. Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *The EMBO journal* **1996**, *15*, 4081–4092.

(46) Eriksson, M.; Jordan, A.; Eklund, H. Structure of *Salmonella typhimurium nrdF* Ribonucleotide Reductase in Its Oxidized and Reduced Forms . *Biochemistry* **1998**, *37*, 13359–13369.

(47) Yano, J.; Robblee, J.; Pushkar, Y.; Marcus, M. A.; Bendix, J.; Workman, J. M.; Collins, T. J.; Solomon, E. I.; DeBeer George, S.; Yachandra, V. K. Polarized X-ray Absorption Spectroscopy of Single-Crystal Mn(V) Complexes Relevant to the Oxygen-Evolving Complex of Photosystem II. *Journal of the American Chemical Society* **2007**, *129*, 12989–13000.

(48) Zeldin, O. B.; Gerstel, M.; Garman, E. F. *RADDOSE-3D* : time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of Applied Crystallography* **2013**, *46*, 1225–1230.

(49) Borek, D.; Cymborowski, M.; Machius, M.; Minor, W.; Otwinowski, Z. Diffraction data analysis in the presence of radiation damage. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66*, 426–436.

(50) Boutet, S.; Yabashi, M. In *X-ray Free Electron Lasers*, Boutet, S., Fromme, P., Hunter, M. S., Eds.; Springer International Publishing: Cham, 2018, pp 1–21.

(51) Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **2000**, *406*, 752–757.

(52) Martin-Garcia, J. M.; Conrad, C. E.; Coe, J.; Roy-Chowdhury, S.; Fromme, P. Serial femtosecond crystallography: A revolution in structural biology. *Archives of Biochemistry and Biophysics* **2016**, *602*, 32–47.

(53) DePonte, D. P.; Weierstall, U.; Schmidt, K.; Warner, J.; Starodub, D.; Spence, J. C. H.; Doak, R. B. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *Journal of Physics D: Applied Physics* **2008**, *41*, 195505.

(54) Cohen, A. E. et al. Goniometer-based femtosecond crystallography with X-ray free electron lasers. *Proceedings of the National Academy of Sciences* **2014**, *111*, 17122–17127.

(55) Barnes, C. O.; Wu, Y.; Song, J.; Lin, G.; Baxter, E. L.; Brewster, A. S.; Nagarajan, V.; Holmes, A.; Soltis, S. M.; Sauter, N. K.; Ahn, J.; Cohen, A. E.; Calero, G. The crystal structure of dGTPase reveals the molecular basis of dGTP selectivity. *Proceedings of the National Academy of Sciences* **2019**, *116*, 9333–9339.

(56) Fuller, F. D. et al. Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nature Methods* **2017**, *14*, 443–449.

(57) Butryn, A. et al. An on-demand, drop-on-drop method for studying enzyme catalysis by serial crystallography. *Nature Communications* **2021**, *12*, 4461.

(58) Stevenson, H. P. et al. Use of transmission electron microscopy to identify nanocrystals of challenging protein targets. *Proceedings of the National Academy of Sciences* **2014**, *111*, 8470–8475.

(59) Zhang, L.; Kaiser, J. T.; Meloni, G.; Yang, K.-Y.; Spatzal, T.; Andrade, S. L. A.; Einsle, O.; Howard, J. B.; Rees, D. C. The Sixteenth Iron in the Nitrogenase MoFe Protein. *Angewandte Chemie International Edition* **2013**, *52*, 10529–10532.

(60) Burgess, B. K.; Jacobs, D. B.; Stiefel, E. I. Large-scale purification of high activity Azotobacter vinelandii nitrogenase. *Biochimica et Biophysica Acta (BBA) - Enzymology* **1980**, *614*, 196–209.

(61) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. *HKL* -3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography* **2006**, *62*, 859–866.

(62) Kabsch, W. *XDS*. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66*, 125–132.

(63) McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D Biological Crystallography* **2005**, *61*, 458–464.

(64) Emsley, P.; Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60*, 2126–2132.

(65) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **1997**, *53*, 240–255.

(66) Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology* **2019**, *75*, 861–877.

(67) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. Collaboration gets the most out of software. *eLife* **2013**, *2*, e01456.

Chapter 4

INTRADIMERIC WALKER A ATPASES: CONSERVED FEATURES OF A FUNCTIONALLY DIVERSE FAMILY

Material for this chapter is published in: Maggiolo, A. O.*; Mahajan, S.*; Rees, D. C.; Clemons, W. M. Intradimeric Walker A ATPases: Conserved features of a functionally diverse family. *Journal of Molecular Biology* **2023**.
*Indicates equal contributions.
A.O.M. performed structural analysis, wrote, and edited the review.

4.1 Introduction

The hydrolysis of nucleoside triphosphates (NTPs) to nucleoside diphosphates (NDPs) supports the progression of a multitude of cellular processes. P-loop NTPases, named after the characteristic functional loop within the fold that coordinates the phosphate moiety of the bound NTP, are a large superfamily that contain a distinctive three-layered αβα-sandwich fold structure.[1–3] Members of the P-loop NTPases perform a wide range of cellular functions, including transcription, translation, cellular regulation, DNA repair, cell division, and membrane transport.[1, 4]

The P-loop, also known as the Walker A motif, was first identified by Walker and coworkers in several ATP-hydrolyzing proteins (ATPases), including the α and β subunits of ATP synthase, myosin, and adenylate kinase.[5] The Walker A motif contains the consensus sequence G-X-X-X-X-G-K-[T/S] (where X is any amino acid) that directly follows an N-terminal β-strand **(Fig. 1A)** . The invariant Lys and Thr/Ser residues of this motif fix the position of the nucleotide by stabilizing the negative charge on the NTP γ-phosphate and coordinating the associated Mg^{2+} ion, respectively.[1] The structured glycine-rich P-loop plays an essential role in orienting phosphates for hydrolysis.[1, 6–8] Walker A proteins also contain a conserved Walker B motif that harbors an aspartate residue that coordinates the associated Mg^{2+} ion through a bridging water molecule. After the original Walker A characterization in ATPases, several small GTP-binding proteins were identified to contain both Walker A and B motifs, including the elongation factor Tu and the H-Ras GTPase.[9, 10] In these systems, the Walker A motif interacts with the phosphate groups of GTP, extending the P-loop terminology to both ATP and GTP hydrolyzing proteins.

In recent years, the ‘Walker A motif’ and ‘P-loop’ terminology have been used interchangeably in the literature.

P-loop NTPases form a distinctive three-layer nucleotide binding fold that consists of a parallel β-sheet that is flanked on both sides by a layer of α-helices. This architecture is proposed to have emerged before the last universal common ancestor.[3, 16]

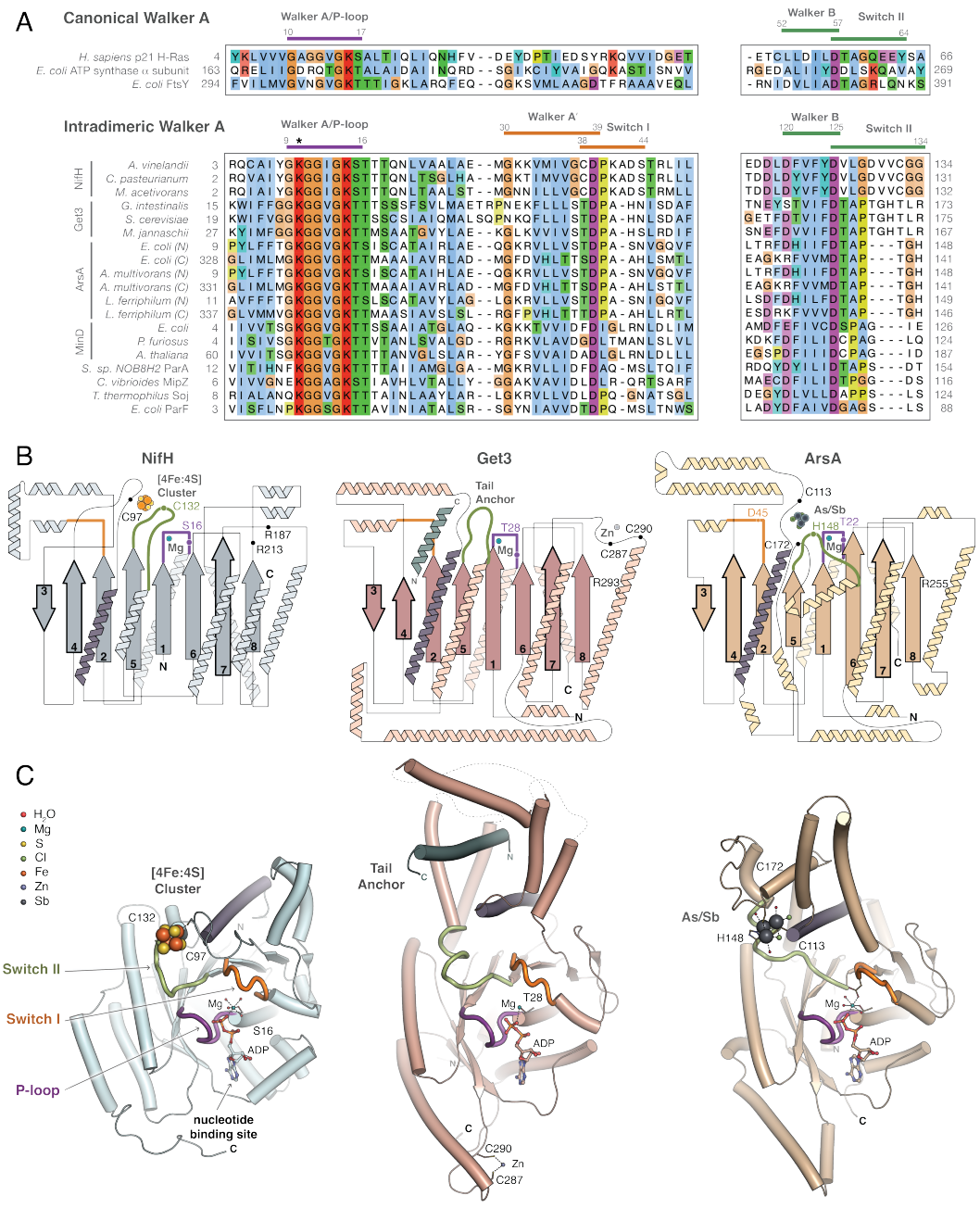


Figure 4.1: **A)** (Top) Sequence alignment of representative members of the canonical Walker A NTPase superfamily and (bottom) of the intradimeric Walker A (IWA) family, highlighting the conserved motifs. Positions of the motifs are numbered based on p21 H-Ras and *Av*NifH for canonical Walker A and IWA sequences, respectively. Asterisk indicates the location of the IWA signature lysine. Full names of the abbreviated organisms can be found in SI Table 1. Alignments were constructed using a structure-based alignment method with PROMALS3D,[11] and residues are colored based on ClustalX.[12] **B)** Topology diagrams of NifH (based on PDB 6N4L), Get3 (PDB 7SPY and PDB 7SQ0 for tail-anchor positioning), and ArsA (PDB 1F48) show the IWA proteins adopt a three-layer αβα-sandwich fold, commonly found in P-loop NTPases.[13–15] The relative β-sheet and α-helix lengths are shown to scale. The connecting loops are not to scale. Select residues are indicated using single letter codes. Bold outlined β-strands indicate insertions to the core β-sheet in the canonical P-loop NTPase fold. Purple-colored core α-helix indicates the 'substrate-helix.' **C)** Cartoon representation of one monomer of *Av*NifH (left, PDB 6N4L), *Gi*Get3 (middle, PDB 7SQ0), and the N-terminal subunit of the *Ec*ArsA (right, PDB 1F48) pseudodimer. The IWA proteins contain a P-loop (purple), Switch I loop (orange), and Switch II loop (green) which couple the nucleotide binding site to the substrate/cargo binding site.

The P-loop fold is structurally similar to another ancient fold, the Rossmann fold, differing primarily in the strand order of the β-sheet, with P-loop folds most commonly having a β2-β3-β1-β4-β5 topology.[3, 17] Both folds serve as a scaffold for functional loops that succeed the β-strands. Residues in these loops either directly interact with nucleotide (as in the case of the P-loop), or interact with a water or proteinaceous network. In the latter case, these loops are commonly termed 'Switch I' and 'Switch II,' as they adopt discrete conformations that can be correlated to nucleotide state.[18] Several comprehensive works provide a framework for the classification and mechanism of P-loop NTPases (for further reading, see [8, 19]).

While P-loop NTPases are ubiquitous among the tree of life, several evolutionary clades have been identified with variations in the Walker A sequence.[20] Based on sequence and structural analysis, Leipe and coworkers broadly distributed the GTPase superclass of the P-loop NTPase superfamily into two classes: TRAFAC, which includes several well-characterized translation factor-related GTPases, and SIMIBI, which includes the signal recognition GTPases, MinD- and BioD-related NTPases.[2] It is known that SIMIBI proteins, that include both GTPases and ATPases, predominantly function as homo- or heterodimers.

A unique subset of proteins, later classified within the SIMIBI proteins, was first identified in the bacterial plasmid partitioning systems, having an uncommon alter-

ation in the first and last glycine of the Walker A motif.[21] This family of proteins with a distinct Walker A sequence, X-K-G-G-X-G-K-S/T , was discussed in detail by E. Koonin in 1993 and expanded to include a handful of evolutionarily related ATPases, which have disparate functions and are found in all three domains of life.[22] Importantly, Koonin identified a second invariant lysine positioned near the beginning of the Walker A motif, which has since been regarded as the 'signature lysine' for members of this subset **(Fig. 1A**, bottom**)**. This subset of ATPases, due to their divergence from other Walker A sequences, was termed the 'deviant Walker A' proteins; however, the question of ancestral origin is unresolved.

In this report, we discuss the commonalities among this distinct subset of P-loop NTPases. We coin the term 'intradimeric Walker A' (IWA) proteins as they are functional homodimers with important catalytic residues that span across the dimer interface. Some of the well characterized members of this group include the nitrogenase iron protein (NifH), the enzyme that supplies electrons during biological nitrogen reduction; the guided entry of tail-anchored proteins factor 3 (Get3), which shepherds tail-anchored integral membrane proteins to their destined membrane for insertion; the cytosolic subunit of the bacterial arsenite efflux pump (ArsA), which transfers arsenite to its partner membrane transporter for cellular detoxification; and the protein involved in the regulation of bacterial cell division (MinD). Other members of this family include DNA partitioning proteins such as ParA and Soj, cell division regulator, MipZ, and flagellar regulation protein, FleN. Due to the plethora of structural information, particularly for NifH and Get3, we use NifH, Get3, and ArsA as representative examples to examine the conserved features of the IWA proteins at the sequence, structural, and functional levels. We find that the proteins in this special subset, with a common evolutionary origin but diverse functionalities, retain remarkable similarity in their ATP-dependent function and allostery as a product of their conserved structural features.

4.2 Sequence Conservation

In comparison to the classical Walker A motif **(Fig. 1A**, top**)**, the IWA proteins differ in the positions of the conserved glycine residues in the P-loop region and have a second invariant lysine after the first β-strand, making the consensus sequence for this subset of Walker A proteins β-strand 1- X-K-G-G-X-G-K-S/T **(Fig. 1A**, bottom**)**.[22] Although there are several hallmarks that differentiate the IWA proteins from the larger SIMIBI family, the initial lysine has generally been considered the defining feature of this subset. Similar to other SIMIBI NTPases, the IWA proteins contain

a conserved motif located ~20 residues downstream of the Walker A motif on the second β-strand of the core β-sheet, designated as the Walker A' motif. This motif overlaps with Switch I and is characterized by a conserved aspartate, implicated in general base catalysis to activate the nucleophilic water for ATP hydrolysis in these proteins.[19]

The general topology of P-loop NTPases in the IWA and greater SIMIBI family consists of a 5-stranded parallel β-sheet, where the first three β-strands make up the Walker A, Walker A', and the Walker B motifs. Each of these β-strands is followed by functional loops, which are designated the P-loop, Switch I, and Switch II, respectively **(Fig. 1B)**. The IWA topology differs from the most common P-loop NTPase topology in that the IWA proteins have three additional β-strands in the core 5-stranded β-sheet from the insertion of two β-strands between β2 and β3 and a third β-strand between β4 and β5 **(Fig. 1B**, bold outlined β-strands**)**. Nonetheless, the P-loop, Switch I, and Switch II come together on the same side of the β-sheet, such that the loops directly connect the nucleotide binding site to a second functional site that is specific to each protein **(Fig. 1C)**.

4.3 Structural Conservation

The IWA proteins have high structural conservation, which is dominated by the common P-loop NTPase fold. Each IWA protein has auxiliary helices and loops, which decorate the core αβα-sandwich and are specific to the function of each protein. For example, Get3 furnishes a series of flexible loops and helices, which form a hydrophobic groove used to capture tail-anchored protein cargo, while NifH utilizes helices to form the interaction surface with its partner protein, NifDK.

Aside from their common P-loop NTPase fold, a key conserved feature of IWA proteins is the requisite homodimeric state **(Fig. 2A)**. In the case of NifH, the homodimer is covalently linked by coordination of the active site cluster, containing four iron and four sulfur atoms ([4Fe:4S]). Similarly, Get3 is a homodimer that is stabilized by a zinc ion. In both NifH and Get3, a number of X-ray crystal structures are available that span multiple functional states and delineate changes at the dimer interface, that are critical for function. ATP-dependent dimerization is an essential feature for the function of MinD, which exists as a monomer prior to ATP binding.[23] In these systems, large dimer conformational changes occur in response to the bound nucleotide state. Thus, there appears to be communication from one monomer to the other throughout the nucleotide hydrolysis cycle. ArsA differs in

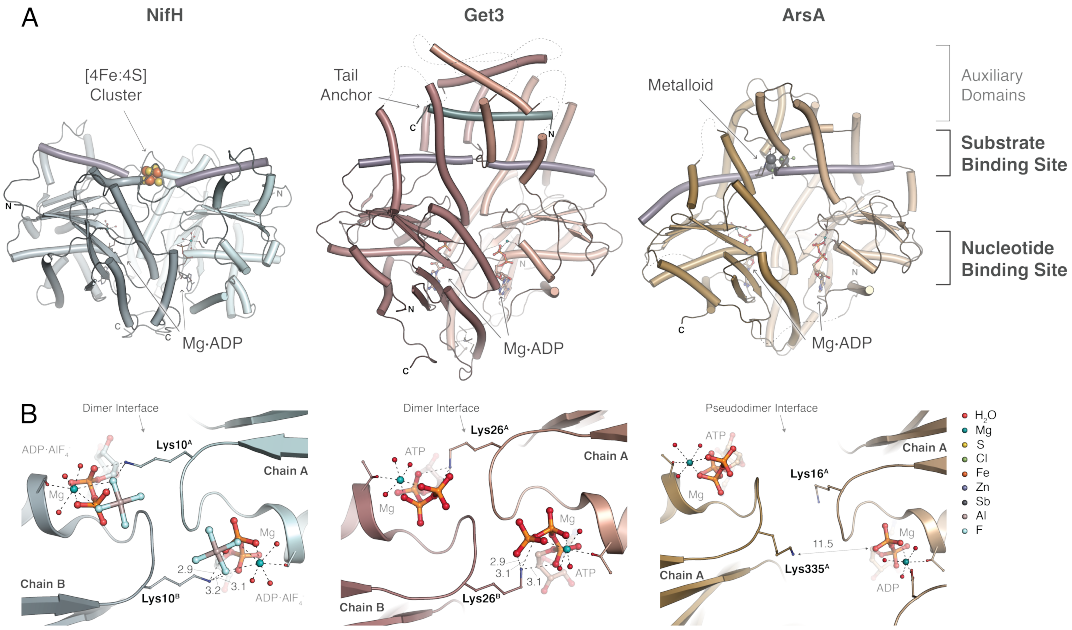


Figure 4.2: **A)** IWA proteins have disparate functions, but commonly act as 'shuttle proteins' that are driven by ATP hydrolysis. In this regard, they have two functional sites: one for nucleotide binding and one for substrate binding. The substrate binding site is tailored to accommodate the cargo specific to each protein, in some cases by the addition of an auxiliary domain. Cartoon representation of three IWA proteins with their nucleotide and substrate binding sites indicated. NifH (PDB 6N4L) carries electrons to its partner protein complex, NifDK, for the reduction of dinitrogen; Get3 (PDB 7SQ0) delivers tail-anchored proteins to the Get1 and Get2 complex for targeted insertion into the membrane; and ArsA (PDB 1F48) scavenges and transports toxic metalloids to the ArsB efflux pump for extrusion out of the cell. The substrate-helix is indicated in purple. **B)** The signature N-terminal lysine in IWA proteins reaches across the dimer interface and interacts with the phosphate moiety of the nucleotide in its partner monomer. Representative examples shown are NifH (left, PDB 1M34), Get3 (middle, PDB 7SPY), and ArsA (right, PDB 1II0).[14, 25, 26] The lysine provides the positive charge necessary to initiate ATP hydrolysis. In addition, it plays a role in neutralizing the build-up of negative charge after ATP hydrolysis and communicates the nucleotide phosphorylation state across the dimer interface. Distances are indicated in angstroms.

this respect as it is expressed as a single chain that effectively forms a covalent dimer, often referred to as a pseudodimer, with two nucleotide binding domains connected by a flexible linker. This has been suggested to be a result of gene duplication and fusion of an ancestor.[24] The pseudosubunits have ~30% sequence identity and together form a dimer interface that mimics the conserved features found in NifH and Get3.

A critical interaction across the dimer interface is provided by the IWA signature lysine. This was first recognized in the X-ray crystal structure of NifH complexed to an ATP-transition state analogue, ADP·AlF_4^-.[27] In this structure, the amino group of Lys10 in each NifH monomer interacts with the negatively charged phosphate group of the nucleotide in the adjacent monomer **(Fig. 2B**, left**)**. A similar interaction is seen in the ATP-bound Get3 dimer **(Fig. 2B**, center**)**. Thus, the lysine was proposed to act as both a charge neutralizer for the nucleotide phosphate moieties and to lower the activation barrier of ATP hydrolysis by stabilizing the negative charge build-up on the phosphate leaving group in the transition state. This interaction is structurally equivalent to the 'arginine finger' of GTPase activating proteins that stimulate GTP hydrolysis in Ras proteins.[28] The location of the catalytic lysines on the opposing subunits serves to facilitate communication about the nucleotide phosphorylation state across a dimer interface and regulate hydrolysis, which can only proceed once the dimer is in the appropriate conformational state. A number of non-nucleotide bound structures of NifH and Get3 reveal that the dimer splays apart in preparation for nucleotide uptake. In these structures, the dimer adopts an 'open' conformation and the residues that make up the dimer interface are oriented such that the IWA lysines are too far apart to interact with the opposing subunit. After nucleotide binding, the dimer interface restructures as the monomers twist together to enclose the nucleotides and form a 'closed' dimer conformation **(Fig. 3A-B)**. This mechanism ensures that nucleotide hydrolysis is directly coupled to the conformational state of the dimer.

In contrast to the structural commonalities at the nucleotide binding domains among the IWA proteins, the auxiliary domain outside the nucleotide binding domain have evolved to adopt functions unique to each protein. Thus, the IWA proteins can be thought of as having two functional sites: (i) the site of ATP hydrolysis and (ii) the site of substrate/cargo binding that is specific to each protein (simplified as substrate for the rest of the text) **(Fig. 2A)**. The two sites communicate to one another by the intervening loops (Switch I and Switch II) and the overall dimer conformation, which is how substrate binding and nucleotide hydrolysis are coupled. For example, NifH has a [4Fe:4S] cluster that enables inter-protein electron transfer between NifH and its partner protein, NifDK. To prevent futile redox cycles, electron transfer only occurs when NifH is properly coordinated to its partner protein and poised to deliver electrons. Structures of NifH complexed to NifDK and the ATP analog, ADP·AlF_4^-, reveal that ATP binding pushes the cluster 4 Å towards the surface of the protein.[27] This shift of the cluster enables electron transfer by minimizing the

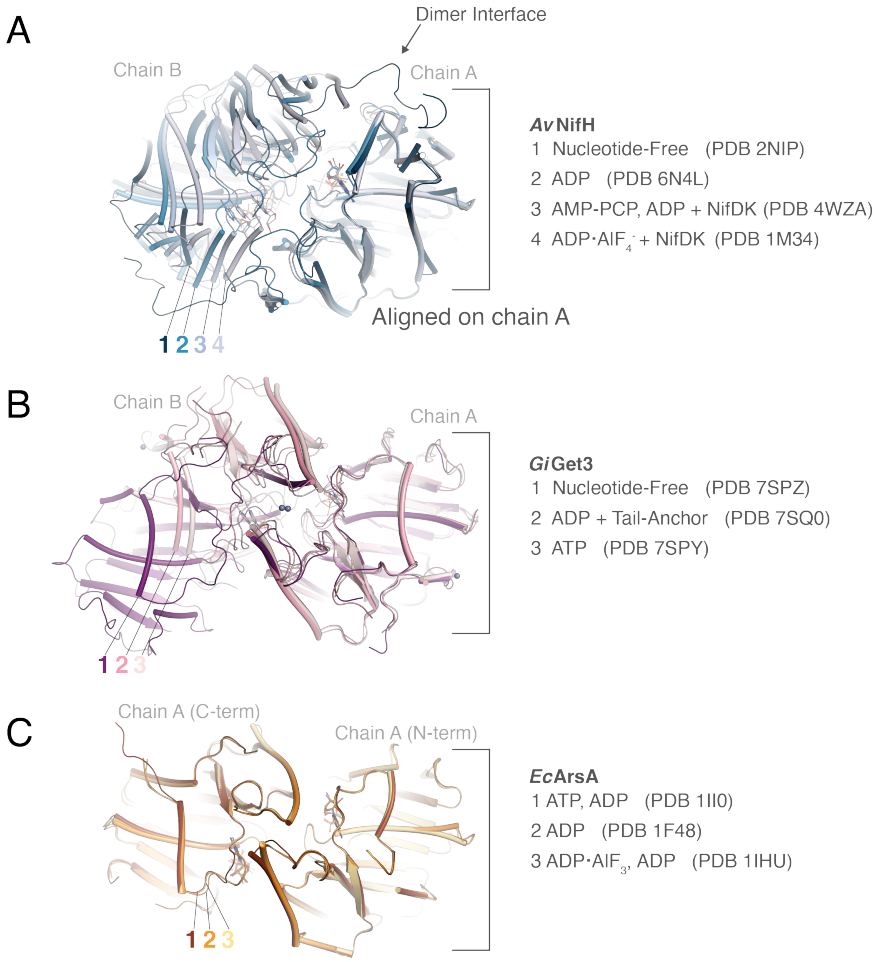


Figure 4.3: The dimer dynamics of IWA proteins can be visualized by alignment on chain A of several nucleotide-dependent dimer conformations of **A)** *Av*NifH, **B)** *Gi*Get3, and **C)** *Ec*ArsA. The line of sight for the dimer is from the bottom relative to the dimer conformation as viewed in Fig. 2B.

distance between the [4Fe:4S] cluster of NifH and the receptor iron sulfur cluster in NifDK. The movement of the [4Fe:4S] cluster in NifH is facilitated by a concerted shift of the Switch I and II loops in response to ATP binding **(Fig. 4A-D)**. Thus, the two active sites are coupled by the components of the IWA motif that relay the nucleotide binding signal to the [4Fe:4S] cluster site and enable the overall activity to be modulated by ATP hydrolysis.

Likewise in Get3, signal is propagated from the ATP binding site to the tail-anchored protein binding site, predominantly through changes that happen in the Switch II loop. Upon binding ATP, the Switch II loop shifts, affecting changes in residues that will subsequently form the hydrophobic cleft at the dimer interface **(Fig. 4E-H)**.[14]

Accessory proteins, such as Get4, and the tail-anchored protein likely facilitate structural rearrangements that form the groove for tail-anchored protein binding, which then stimulates ATP hydrolysis.[29, 30] This is evident in the post-hydrolysis tail-anchored protein-bound state of *Giardia intestinalis (Gi)*Get3, wherein the C-terminal end of the Switch II loop has restructured into an α-helix **(Fig. 4G-H)**.[14] Furthermore, in both NifH and Get3, the helix that precedes the β-strand harboring the Walker B motif (residues 96-112 in *Av*NifH and 139-157 in *Gi*Get3), herein referred to as the 'substrate-helix,' moves as part of the conformational changes that occur at the substrate binding site **(Fig. 1B-C**, purple helices**)**. This is most apparent in structures of Get3, where this helix shifts down towards the nucleotide binding site upon tail-anchored protein binding, forming the base of the hydrophobic cleft.[14, 30] In NifH, the change in conformation is more subtle, as rearrangements in this helix are coupled to repositioning of the [4Fe:4S] cluster to facilitate intermolecular electron transfer when complexed to NifDK. Switch II likely coordinates with the substrate-helix to mediate communication between the nucleotide binding site and the substrate binding site.

ArsA is structurally more similar to Get3 than the other well-characterized IWA proteins. Like Get3, ArsA contains the auxiliary helical domain that extends above the nucleotide binding sites **(Fig. 1C**, center and right**)**. In Get3 this domain makes up the hydrophobic cleft for tail-anchored protein binding. In ArsA, it encloses a cysteine-rich site located at the dimer interface that forms the binding site for its physiological metalloid substrates, arsenite and antimonite.[15, 31] The presence of three conserved cysteines at this site and the absence of the features necessary for tail-anchored protein binding are the hallmarks that distinguish ArsA from Get3, all of which were indiscriminately annotated as ArsA in databases. Mutagenesis and tryptophan fluorescence kinetic studies revealed that the metalloid binding site of ArsA is allosterically coupled to ATP binding and hydrolysis.[31, 32] Analogous to NifH and Get3, the substrate binding site of ArsA is structurally linked to the ATP binding site via the Switch loops **(Fig. 1B-C**, right**)**. X-ray crystal structures of *Escherichia coli (Ec)*ArsA bound to various nucleotides have been reported.[15, 26] Surprisingly, there are no significant conformational changes between these structures, unlike the dramatic changes seen in the nucleotide-free versus nucleotide-bound states of Get3 and NifH **(Fig. 3C)**. Likewise, the signature lysines in the structures of ArsA bound to ADP·AlF_3 or ATP are >10 Å apart and are oriented away from the nucleotide, resembling the open dimer conformation **(Fig. 2B)**, right**)**. It is plausible that the crystal packing was not conducive to explore the

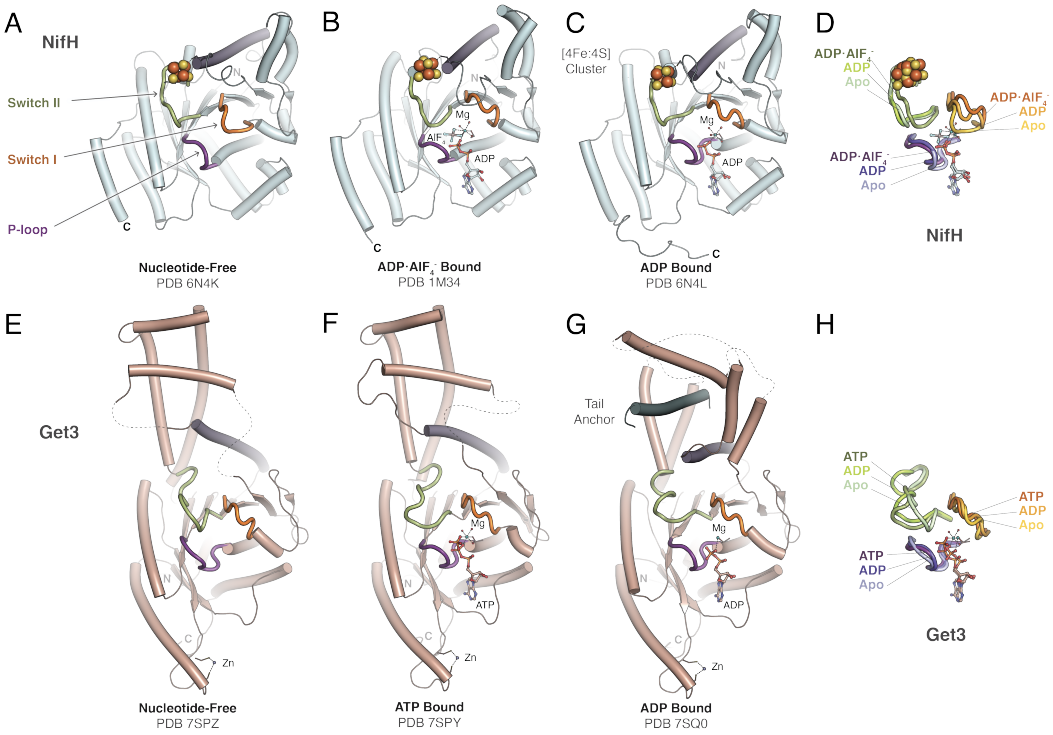


Figure 4.4: (Top) Conformations of the nucleotide-dependent functional loops (P-loop (purple), Switch I (orange), and Switch II (green)) in one monomer of NifH in the **A)** nucleotide-free, **B)** ADP·AlF_4^-, and **C)** ADP bound states. **D)** Direct comparison of the three functional loops to demonstrate their nucleotide-dependent conformational changes. (Bottom) Conformations of the nucleotide-dependent functional loops in one monomer of Get3 in the **E)** nucleotide-free, **F)** ATP, and **G)** ADP and tail-anchor substrate bound states. **H)** Comparison of the Get3 functional loops.

full conformational landscape of ArsA. Nevertheless, the presence of all the IWA structural motifs in ArsA strongly indicate that the ATP-dependent conformation rearrangements are likely preserved and required to catalyze the transfer of the metalloid to the membrane efflux pump, ArsB.

4.4 Functional Conservation

In IWA proteins, ATP binding and hydrolysis impose changes at the dimer interface that span a range of nucleotide-dependent dimer conformational states **(Fig. 3A-B)**. An abundance of structural data, particularly among the NifH and Get3 proteins, supports that the members of the IWA family exhibit common nucleotide-dependent dimer dynamics. In the nucleotide-free structures of NifH and Get3, the dimers adopt the open conformation with minimal protein interactions across the dimer interface.[13, 14] In the structures of NifH and Get3 in the pre-hydrolytic ATP-

bound state, the dimers are stabilized in the closed conformation, wherein the dimer interface constricts, bringing the IWA signature lysine in proximity to the nucleotides in preparation for hydrolysis.[14, 33] In these systems, docking of a partner protein or substrate stabilizes the closed dimer conformation and serves as a trigger for ATP hydrolysis. In NifH, this trigger is provided by complexation to the partner protein, NifDK.[25] In Get3, tail-anchored protein binding and subsequent release from a binding partner initiates hydrolysis.[14, 34] In MinD, binding of the partner protein, MinE, activates hydrolysis.[35]

At the hydrolysis step, a water molecule positioned adjacent to the γ-phosphate and trans to the β-phosphate, serves as the nucleophile for ATP hydrolysis. In NifH, several aspartate residues within the nucleotide binding site are candidates for activation of the catalytic water. The structure of the ADP·AlF_4^--bound state of NifH, contains a hydrogen bonding network involving the conserved aspartate residues on the Switch I and II loops, the water molecules coordinating the Mg^{2+} ion, and the conserved Ser/Thr residue on the P-loop **(Fig. 5A-B)**. It is difficult to tell from structures alone whether a single aspartate is responsible for water activation. Based on both structural interpretation and mutagenesis studies, it is more likely that several apartates are essential and an intricate hydrogen bonding network surrounding the catalytic water facilitates its activation via a proton relay mechanism.[19, 36–38] Similar to the IWA signature lysine, Asp129 reaches across the dimer interface to stabilize the catalytic water in the ADP·AlF_4^--bound state, but not in the ADP-bound state **(Fig. 5C)**, further affirming the importance of nucleotide-dependent intradimeric interactions. A similar arrangement of asparates forming a hydrogen bonding network is also present in the ATP-bound Get3 structure.[14] Following hydrolysis and phosphate release, the dimer expands, returning to its open conformation, as represented in the ADP-bound structures of NifH and Get3.[13, 14]

The ATPase cycle mediates discrete conformational rearrangements of the Switch loops. This allosterically induces structural changes at the substrate binding site involved in the primary biological function of the respective proteins. This typically involves transfer of the substrate to its downstream binding partner as characterized for NifH, Get3, and ArsA. On the other hand, MinD facilitates localization of an entire protein, MinC, to the site on the plasma membrane where non-productive cell divison must be prevented.[23, 39] The combined oscillation of MinD and MinC between the poles of a dividing cell is required for cell division regulation. Based

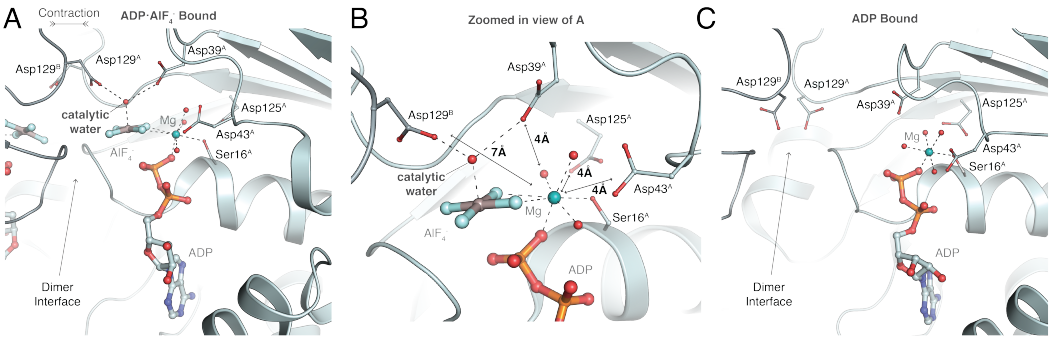


Figure 4.5: **A)** Asp39 and 43 (Switch I), and Asp125 and 129 (Switch II) participate in an intricate hydrogen bonding network around coordinated ADP·AlF_4^-. Asp43, 125, and 129 are essential for the Mg·ATP hydrolysis activity of *Av*NifH (PDB 1M34).[8, 36–38] Mutation of Asp39 has been shown to affect the FeMo-cofactor synthesis and maturation activities of NifH.[38] **B)** Zoom-in of panel A indicating the distances from the aspartates to the Mg^{2+} ion (PDB 1M34). **C)** Aspartate residues around the nucleotide binding site of *Av*NifH when ADP is bound (PDB 6N4L).

on several structures, the substrate binding site of MinD differs from other IWA members in that MinC binds at the dimer interface near the nucleotide binding sites of MinD,[40–42] which is a distinct region from the substrate binding site of NifH, Get3, and ArsA. Nonetheless, the presence of the IWA features indicates that functionally relevant interactions at the analogous IWA substrate binding site on MinD must be conserved and the corresponding details await future characterization. As the ATPase cycle regulates transfer of a substrate from an IWA proteins to either a partner protein or a different location in the cell, these proteins, in principle, can be thought of as 'shuttle proteins' that have evolved to catalyze the transport (shuttling) of their respective substrates. Although a similar description has previously been made with reference to DNA segregation proteins (MinD and ParA),[43] we identify that this functional role is conserved across the IWA family.

The catalytic steps of the ATPase cycle regulating the shuttle function of the IWA proteins are mediated by partner proteins that are functionally analogous to effectors. NifH is an example where the catalytically competent conformation required for both ATP hydrolysis and electron transfer is stabilized only when it is bound to NifDK; this property effectively prevents futile ATPase cycles.[27] After ATP hydrolysis and electron transfer, the NifH dimer dissociates from NifDK and transitions to the open dimer conformation. In this state, ATP exchanges for ADP and the [4Fe:4S] cluster can be reduced by the appropriate reductant.[44] Although both Get3 and

ArsA exhibit basal ATPase activity in the absence of substrate, effector binding in both cases has been shown to regulate the ATPase cycle.[34, 45] Binding of the effector protein, Get4, to Get3 stabilizes its ATP-bound closed dimer conformation, exposing the substrate binding site for tail-anchored protein binding. Hydrolysis is inhibited until the tail-anchor is successfully bound.[29, 30] Subsequently, Get3 binds to the Get1 and Get2 endoplasmic reticulum membrane complex, which stabilizes the open dimer conformation, facilitating the transfer of the tail-anchored protein into the membrane.[46] In the case of ArsA, a metallochaperone, ArsD, putatively binds ArsA to facilitate transfer of arsenite to the substrate binding site of ArsA.[47] Additionally, by analogy to NifH-NifDK, interactions between ArsA and its membrane partner ArsB may also be critical to regulate the nucleotide state of ArsA suitable for arsenite transfer. Yet another mechanism of regulation of IWA proteins, especially notable in Get3, MinD, and chromosomal segregation proteins such as ParA and Soj, has been suggested to be conferred by membrane association of these proteins. NifH has been found to associate with membranes in nitrogen-fixing bacteria under low oxygen tension.[48] Indeed, it has been proposed that lipids can trigger conformational changes in these systems.[49]

The presence of such regulatory features in IWA proteins that control their switching between the open and closed dimer conformations might serve as a critical mechanistic feature that ensures precise timing of the substrate shuttling function. That these proteins act as 'molecular clocks' is well established for MinD, which oscillates between the poles of the cell and prevents non-productive division septa formation necessary for cell division. The ATP-dependent dimerization linked to binding of MinC and MinE proteins helps regulate the periodicity of MinD function.[23] A similar mechanistic role of ATP hydrolysis has been discussed previously in light of the nitrogenase mechanism, suggesting that ATP hydrolysis is thermodynamically non-essential for electron transfer and instead acts as a conformational gate to maintain the unidirectionality of electron transfer from NifH to NifDK.[50] We propose that the well regulated ATPase cycle is critical for controlling the flux of the substrate being shuttled, therefore, ensuring precise timing in other IWA proteins, as well.

4.5 Structurally Conserved Arginine Residues

Beyond the residues in the conserved sequence motifs defining this family, IWA proteins have additional conserved interactions across the dimer interface that correlate with nucleotide-dependent conformational states. Of particular interest are

interactions involving arginine residues that are frequently found at the dimer interface of IWA proteins. One striking example, described in *Av*NifH by Tezcan, Rees, and coworkers, involves a pair of arginine residues (Arg187 and Arg213) from each NifH subunit in a complex of NifH with NifDK, stabilized by the asymmetric binding of AMP-PCP and ADP.[51] Arg187 and Arg213 are positioned on loops following β-strand 7 and β-strand 8, respectively **(Fig. 1B)**. The intriguing aspect of these arginines was the prominent chain of π-stacking interactions involving the adenine bases and the guanidinium groups of the arginines to form the π-stacked chain adenine - Arg213 - Arg187 - Arg187 - Arg213 - adenine linking the two nucleotide binding sites across the homodimer interface **(Fig. 6A)**. This observation prompted examination of these residues in a series of NifH structures to evaluate the relationship between the locations of these arginines and the nucleotide dependent NifH dimer conformation. Arg187 and Arg213 were found to interact via discrete hydrogen bonding and salt bridge interactions with each other and with a neighboring glutamate (Glu154 in *Av*NifH) in structures of NifH complexed to NifDK with ADP·AlF_4^-,[25, 27] ADP,[33] and AMP-PCP.[51] Structures of nucleotide-free NifH do not exhibit direct arginine-arginine interactions across the dimer interface as the dimer is splayed too far apart; additionally, the adenine appears to provide an important handle for arginine stacking **(Fig. 6B)**.[13, 52] We note that mutagenesis of Arg213 to cysteine renders the protein inactive as measured by *in vivo* acetylene reduction assays.[53] The interactions between these four arginine residues and the nucleotide could serve to control nucleotide positioning across an interface coupling the dimer conformation to the nucleotide state. Although the interaction between two positively charged side chains has an unfavorable electrostatic component, arginine-arginine interactions across protein interfaces have been found to be thermodynamically favorable,[54–56] thereby facilitating the formation of the closed dimer conformation.

In addition to NifH, a survey of IWA motif-containing proteins in the Protein Data Bank (PDB) revealed the presence of an arginine residue (sometimes a pair of arginine residues) that is structurally analogous to NifH Arg187 and 213. Similar to the IWA signature lysine, these arginines are found to participate in extensive inter-subunit interactions with both the protein and the bound nucleotides. Arginine-arginine interactions are maintained in MinD and more broadly in the Par family of proteins **(Fig. 6C-D)**. Upon examination of the available MinD structures from different organisms, we find inter-subunit hydrogen bonds made by the two dimer-related arginines that also interact with nucleotides. In *Pyrococcus fu-*

*riosus(Pf)*MinD, a pair of arginines (residues Arg172 and Arg175) stack over the nucleotide in the monomeric ADP-bound structure **(Fig. 6C)**. As there are no dimeric ATP-bound structures of *Pf*MinD available, we examined the ATP-bound, dimeric *Ec*MinD that has arginines homologous to those of *Pf*MinD (Arg182 and Arg187, respectively). In this structure, Arg187 is dissociated from the stack upon

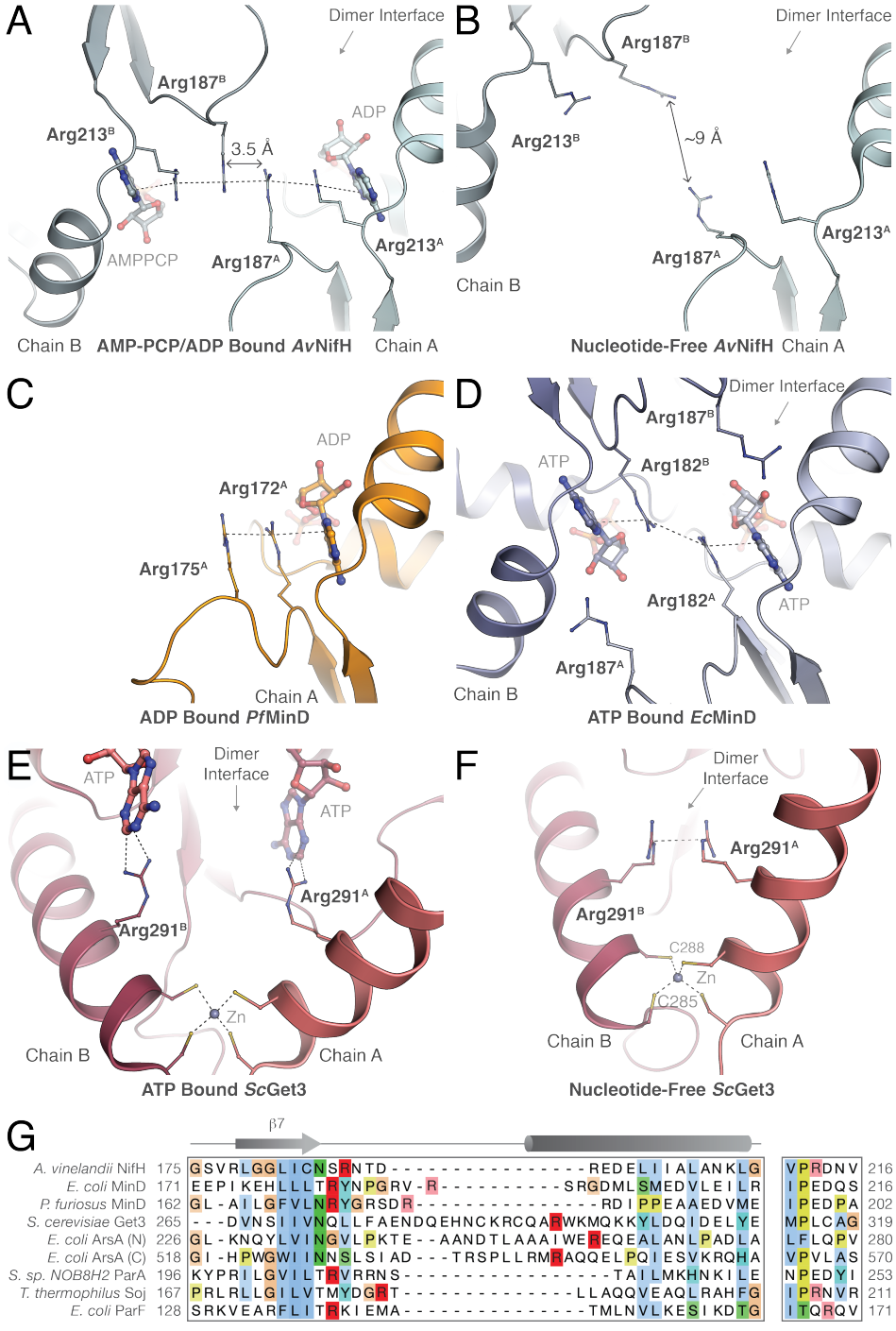


Figure 4.6: **A)** Four arginine residues (two from each monomer) reside at the dimer interface of *Av*NifH. In the absence of nucleotide (PDB 6N4K), the dimer adopts an 'open' conformation with the arginines too widely separated to interact. Dimer interface distances are indicated as solid arrows. **B)** Upon nucleotide binding, the NifH dimer constricts and the four arginines π-π stack over the nucleotides and across the dimer interface (PDB 4WZA). **C)** Two arginine residues from the same monomer that structurally align with the IWA arginine residues in *Av*NifH, adopt a similar stacking motif in the monomeric *Pf*MinD. **D)** In the dimeric ATP-bound *Ec*MinD structure, one pair of arginines stack over the bound nucleotides and extend interactions across the dimer interface. **E)** Arg291 in *Sc*Get3 interacts with the nucleotide in the neighboring monomer when ATP is bound (PDB 4XVU) and **F)** interact at the dimer interface when either no nucleotide or ADP is bound (PDB 3SJA). **G)** The IWA arginine residues (highlighted in red) are structurally conserved but are not discernible from the sequence alignment. Full names of the abbreviated organisms can be found in SI Table 1.

dimerization and Arg182 is hydrogen bonded to Arg187 from the adjacent monomer **(Fig. 6D)**.

In Get3, arginine interactions at the dimer interface are most clear in *Saccharomyces cerevisiae (Sc)*Get3. There is a single arginine residue (Arg291 in *Sc*Get3) that participates in nucleotide-dependent intradimeric contacts. In structures of *Sc*Get3 with ATP bound, Arg291 from each monomer interacts with the adenine base of the ATP in the opposing subunit **(Fig. 6E)**. In the structures with no nucleotide bound, Arg291 from each monomer interacts across the dimer interface via direct hydrogen bonding or π-π stacking **(Fig. 6F)**. An investigation of 69 mutations in *Sc*Get3 by Suloway, Clemons, and coworkers found that mutation of Arg291 resulted in a strong loss of function phenotype.[57] Incubation of R291A with wildtype protein could not recover function, indicating mutation of Arg291 had a dominant negative effect (unpublished results). The corresponding arginines in *Ec*ArsA interact with the adenine in the opposing monomer in the single dimer conformation seen for ArsA; however, a π-π stacking interaction between the arginines is not observed.[15, 26] As discussed previously, the conformational states of ArsA do not appear to have been fully explored and it is likely that these types of interactions will be observed in conformational states of ArsA yet to be characterized.

The absence of a conserved sequence motif surrounding these arginines has precluded recognition of the significance of these interactions **(Fig. 6G**, dark red). Nevertheless, the arginines are structurally conserved with respect to the nucleotide binding site and are present on the loop or the α-helix following β-strand 7. In

some IWA proteins, a second arginine is present at the dimer interface and typically resides on a different structural element **(Fig. 6G**, light red**)**. Given the striking conformation of these arginine residues, which interact across a dimer interface at the nucleotide binding site in several members of the IWA subclass, and that the intra-subunit interactions correlate with multiple structural intermediate states in NifH, the arginine residues further support the importance of cross-dimeric interactions in the IWA family.

4.6 Conclusions

The IWA proteins had previously been ascribed to a subset of Walker A proteins that have an additional, fully conserved N-terminal lysine residue that is essential in ATP hydrolysis in these enzymes that function as obligate homodimers. Due to their common P-loop NTPase fold, they share a conserved nucleotide binding site. The IWA proteins have a second functional site, also at the dimer interface and typically at the opposing end of the dimer from the nucleotide binding site. This second functional site serves as the location for substrate binding and is tailored with auxiliary structural components to accommodate substrates that are specific to each protein system. Within a monomer, a series of loops that straddle the nucleotide and substrate binding sites directly couple the substrate transport activity to the ATPase cycle. As illustrated by studies of MinD, regulation of the ATPase cycle of IWA proteins by interactions with partner proteins and/or the cell membrane ensures precise timing and unidirectionality of substrate transport. This timing mechanism might be a critical feature that enables IWA proteins to carry out their respective biological functions efficiently.

Structural analysis reveals that the nucleotide-dependent conformations of the homodimer are an important feature in the progression of the ATPase cycle in IWA proteins. Several residues that make intradimeric interactions, including the IWA signature lysine, are essential in facilitating communication about the phosphorylation state of the coordinated nucleotides across the dimer interface to induce concerted dimer motions. Herein, we identify arginine residues that are present at the dimer interface in the IWA proteins, but not conserved in the IWA sequence motif. It is tempting to speculate on their potential role in the IWA proteins, which may not necessarily be the same for all IWA members. Although these arginines appear to be unique to IWA proteins, a highly conserved lysine residue in the related NTPase, p21 Ras, (Lys117 in *Homo sapiens*) is structurally homologous to Arg213 in NifH.[50, 58] Lys117 in Ras positions directly over the base of the bound

nucleotide and plays an important role in guanine base recognition.[58, 59] The connection between Arg213 in NifH and Lys117 in Ras has been made before as one of the residues in the P-loop NTPase fold responsible for adenine versus guanine base specificity.[60] Whether these arginines partake in either base specificity or dimer dynamics, or both, in NifH and other IWA proteins is worth further investigation.

The commonalities described here between the members in this special subset of Walker A proteins is seemingly at odds with their divergent functions. On some level, it is challenging to see the similarity between electron, metalloid, and hydrophobic tail-anchored protein transfer. Moreover, the sequence of steps of the ATPase cycle-mediated substrate transfer may not necessarily be conserved and may be specific to the mechanism of each IWA protein. Nonetheless, the ability of these proteins to transport various substrates seems less surprising considering the core of the protein remains virtually unchanged and only the regions of the protein that harbor various cargo appeared to undergo considerable evolutionary alterations. Thus, the IWA proteins retain their most integral feature, which is substrate transport driven by ATP-dependent structural changes. We predict that further structural characterization of the substrate and nucleotide-bound states of other IWA proteins will reveal conservation of these mechanistic features. An interesting note is that Get3 is the only IWA protein known to date found in humans and yeast, unlike the many different IWA proteins found in bacteria and archeae. An outstanding question concerns the evolution of the IWA family in the P-loop NTPase evolutionary timeline and its functionally critical homodimeric architecture that has evolved to adapt remarkably diverse biological functions.

Table 4.1: Full names of organisms mentioned throughout the text.

H. sapiens	*Homo sapiens*
E. coli	*Escherichia coli*
A. vinelandii	*Azotobacter vinelandii*
C. pasteurianum	*Clostridium pasteurianum*
M. acetivorans	*Methanosarcina acetivorans*
G. intestinalis	*Giardia intestinalis*
S. cerevisiae	*Saccharomyces cerevisiae*
M. jannaschii	*Methanococcus jannaschii*
A. multivorans	*Acidiphilium multivorans*
L. ferriphilum	*Leptospirillum ferriphilum*
P. furiosus	*Pyrococcus furiosus*
A. thaliana	*Arabidopsis thaliana*
S. sp. NOB8H2	*Sulfolobus species NOB8H2*
C. vibrioides	*Caulobacter vibrioides*
T. thermophilus	*Thermus thermophilus*

References

(1) Saraste, M.; Sibbald, P. R.; Wittinghofer, A. The P-loop — a common motif in ATP- and GTP-binding proteins. *Trends in Biochemical Sciences* **1990**, *15*, 430–434.

(2) Leipe, D. D.; Wolf, Y. I.; Koonin, E. V.; Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *Journal of Molecular Biology* **2002**, *317*, 41–72.

(3) Longo, L. M.; Jabłońska, J.; Vyas, P.; Kanade, M.; Kolodny, R.; Ben-Tal, N.; Tawfik, D. S. On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. *eLife* **2020**, *9*, e64415.

(4) Vetter, I. R.; Wittinghofer, A. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. *Quarterly Reviews of Biophysics* **1999**, *32*, 1–56.

(5) Walker, J.; Saraste, M.; Runswick, M.; Gay, N. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal* **1982**, *1*, 945–951.

(6) Mimura, C. S.; Holbrook, S. R.; Ames, G. F. Structural model of the nucleotide-binding conserved component of periplasmic permeases. *Proceedings of the National Academy of Sciences* **1991**, *88*, 84–88.

(7) Georgiadis, M. M.; Komiya, H.; Chakrabarti, P.; Kornuc, J. J.; Rees, D. C. Crystallographic Structure of the Nitrogenase Iron Protein from Azo tobacter vinelandli. **1992**, *257*, 8.

(8) Kozlova, M. I.; Shalaeva, D. N.; Dibrova, D. V.; Mulkidjanian, A. Y. Common Patterns of Hydrolysis Initiation in P-loop Fold Nucleoside Triphosphatases. *Biomolecules* **2022**, *12*, 1345.

(9) Jurnak, F. Structure of the GDP Domain of EF-Tu and Location of the Amino Acids Homologous to *ras* Oncogene Proteins. *Science* **1985**, *230*, 32–36.

(10) Kim, S.-H.; de Vos, A.; Tong, L.; Milburn, M.; Matias, P.; Jancarik, J.; Ohtsuka, E.; Nishimura, S. ras Oncogene Proteins: Three-dimensional Structures, Functional Implications, and a Model for Signal Transducer. *Cold Spring Harbor Symposia on Quantitative Biology* **1988**, *53*, 273–281.

(11) Pei, J.; Kim, B.-H.; Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* **2008**, *36*, 2295–2300.

(12) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **1997**, *25*, 4876–4882.

(13) Wenke, B. B.; Spatzal, T.; Rees, D. C. Site Specific Oxidation State Assignments of the Iron Atoms in the [4Fe:4S] $^{2+/1+/0}$ States of the Nitrogenase Fe Protein. *Angewandte Chemie International Edition* **2019**, *58*, 3894–3897.

(14) Fry, M. Y.; Najdrová, V.; Maggiolo, A. O.; Saladi, S. M.; Doležal, P.; Clemons, W. M. Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. *Nature Structural & Molecular Biology* **2022**, *29*, 820–830.

(15) Zhou, T. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *The EMBO Journal* **2000**, *19*, 4838–4845.

(16) Aravind, L.; Iyer, L. M.; Leipe, D. D.; Koonin, E. V. A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. *Genome Biology* **2004**, *5*, R30.

(17) Rossmann, M. G.; Moras, D.; Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **1974**, *250*, 194–199.

(18) Milburn, M. V.; Tong, L.; deVos, A. M.; Brünger, A.; Yamaizumi, Z.; Nishimura, S.; Kim, S.-H. Molecular Switch for Signal Transduction: Structural Differences Between Active and Inactive Forms of Protooncogenic *ras* Proteins. *Science* **1990**, *247*, 939–945.

(19) Kozlova, M. I.; Shalaeva, D. N.; Dibrova, D. V.; Mulkidjanian, A. Y. Common Mechanism of Activated Catalysis in P-loop Fold Nucleoside Triphosphatases—United in Diversity. *Biomolecules* **2022**, *12*, 1346.

(20) Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites: Rossmann Fold in FAD, NAD and NADP Binding Domains. *Biochemistry and Molecular Biology Education* **2015**, *43*, 206–209.

(21) Motallebi-Veshareh, M.; Rouch, D. A.; Thomas, C. M. A family of ATPases involved in active partitioning of diverse bacterial plasmids. *Molecular Microbiology* **1990**, *4*, 1455–1463.

(22) Koonin, E. V. A Superfamily of ATPases with Diverse Functions Containing Either Classical or Deviant ATP-binding Motif. *Journal of Molecular Biology* **1993**, *229*, 1165–1174.

(23) Lutkenhaus, J.; Sundaramoorthy, M. MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. *Molecular Microbiology* **2003**, *48*.

(24) Chen, C. M.; Misra, T. K.; Silver, S.; Rosen, B. P. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *Journal of Biological Chemistry* **1986**, *261*, 15030–15038.

(25) Schmid, B.; Einsle, O.; Chiu, H.-J.; Willing, A.; Yoshida, M.; Howard, J. B.; Rees, D. C. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF_4^--Stabilized Structure † ‡. *Biochemistry* **2002**, *41*, 15557–15565.

(26) Zhou, T.; Radaev, S.; Rosen, B. P.; Gatti, D. L. Conformational Changes in Four Regions of the Escherichia coli ArsA ATPase Link ATP Hydrolysis to Ion Translocation. *Journal of Biological Chemistry* **2001**, *276*, 30414–30422.

(27) Schindelin, H.; Kisker, C.; Schlessman, J. L.; Howard, J. B.; Rees, D. C. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. *Nature* **1997**, *387*, 370–376.

(28) Ahmadian, M. R.; Stege, P.; Scheffzek, K.; Wittinghofer, A. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature Structural Biology* **1997**, *4*, 686–689.

(29) Gristick, H. B.; Rao, M.; Chartron, J. W.; Rome, M. E.; Shan, S.-o.; Clemons, W. M. Crystal structure of ATP-bound Get3–Get4–Get5 complex reveals regulation of Get3 by Get4. *Nature Structural & Molecular Biology* **2014**, *21*, 437–442.

(30) Keszei, A. F. A.; Yip, M. C. J.; Hsieh, T.-C.; Shao, S. Structural insights into metazoan pretargeting GET complexes. *Nature Structural & Molecular Biology* **2021**, *28*, 1029–1037.

(31) Ruan, X.; Bhattacharjee, H.; Rosen, B. P. Cys-113 and Cys-422 Form a High Affinity Metalloid Binding Site in the ArsA ATPase. *Journal of Biological Chemistry* **2006**, *281*, 9925–9934.

(32) Walmsley, A. R.; Zhou, T.; Borges-Walmsley, M. I.; Rosen, B. P. Antimonite regulation of the ATPase activity of ArsA, the catalytic subunit of the arsenical pump. *Biochemical Journal* **2001**, *360*, 589–597.

(33) Tezcan, F. A. Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein. *Science* **2005**, *309*, 1377–1380.

(34) Chio, U. S.; Chung, S.; Weiss, S.; Shan, S.-o. A protean clamp guides membrane targeting of tail-anchored proteins. *Proceedings of the National Academy of Sciences* **2017**, *114*.

(35) Hu, Z.; Lutkenhaus, J. Topological Regulation of Cell Division in E. coli: Spatiotemporal Oscillation of MinD Requires Stimulation of Its ATPase by MinE and Phospholipid. *Molecular Cell* **2001**, *7*, 1337–1343.

(36) Lanzilotta, W. N.; Ryle, M. J.; Seefeldt, L. C. Nucleotide Hydrolysis and Protein Conformational Changes in Azotobacter vinelandii Nitrogenase Iron Protein: Defining the Function of Aspartate 129. *Biochemistry* **1995**, *34*, 10713–10723.

(37) Wolle, D.; Dean, D. R.; Howard, J. B. Nucleotide-Iron-Sulfur Cluster Signal Transduction in the Nitrogenase Iron-Protein: the Role of Asp ^{125}. *Science* **1992**, *258*, 992–995.

(38) Rangaraj, P.; Ryle, M. J.; Lanzilotta, W. N.; Ludden, P. W.; Shah, V. K. In Vitro Biosynthesis of Iron-Molybdenum Cofactor and Maturation of the nif-encoded Apodinitrogenase. *Journal of Biological Chemistry* **1999**, *274*, 19778–19784.

(39) Hu, Z.; Saez, C.; Lutkenhaus, J. Recruitment of MinC, an Inhibitor of Z-Ring Formation, to the Membrane in *Escherichia coli* : Role of MinD and MinE. *Journal of Bacteriology* **2003**, *185*, 196–203.

(40) Wu, W.; Park, K.-T.; Holyoak, T.; Lutkenhaus, J. Determination of the structure of the MinD-ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC: Structure of MinD-ATP complex. *Molecular Microbiology* **2011**, *79*, 1515–1528.

(41) Ghosal, D.; Trambaiolo, D.; Amos, L. A.; Löwe, J. MinCD cell division proteins form alternating copolymeric cytomotive filaments. *Nature Communications* **2014**, *5*, 5341.

(42) Szewczak–Harris, A.; Wagstaff, J.; Löwe, J. Cryo-EM structure of the MinCD copolymeric filament from *Pseudomonas aeruginosa* at 3.1 Å resolution. *FEBS Letters* **2019**, *593*, 1915–1926.

(43) Lutkenhaus, J. The ParA/MinD family puts things in their place. *Trends in Microbiology* **2012**, *20*, 411–418.

(44) Segal, H. M.; Spatzal, T.; Hill, M. G.; Udit, A. K.; Rees, D. C. Electrochemical and structural characterization of Azotobacter vinelandii flavodoxin II. *Protein Science: A Publication of the Protein Society* **2017**, *26*, 1984–1993.

(45) Rosen, B. P.; Bhattacharjee, H.; Zhou, T.; Walmsley, A. R. Mechanism of the ArsA ATPase. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1999**, *1461*, 207–215.

(46) Mariappan, M.; Mateja, A.; Dobosz, M.; Bove, E.; Hegde, R. S.; Keenan, R. J. The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **2011**, *477*, 61.

(47) Lin, Y.-F.; Walmsley, A. R.; Rosen, B. P. An arsenic metallochaperone for an arsenic detoxification pump. *Proceedings of the National Academy of Sciences* **2006**, *103*, 15617–15622.

(48) Hurek, T.; Montagu, M.; Kellenberger, E.; Reinhold-Hurek, B. Induction of complex intracytoplasmic membranes related to nitrogen fixation in *Azoarcus* sp. BH72. *Molecular Microbiology* **1995**, *18*, 225–236.

(49) Bange, G.; Sinning, I. SIMIBI twins in protein targeting and localization. *Nature Structural & Molecular Biology* **2013**, *20*, 776–780.

(50) Howard, J. B.; Rees, D. C. NITROGENASE: A Nucleotide-Dependent Molecular Switch. *Annual Review of Biochemistry* **1994**, *63*, 235–264.

(51) Tezcan, F. A.; Kaiser, J. T.; Howard, J. B.; Rees, D. C. Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase. *Journal of the American Chemical Society* **2015**, *137*, 146–149.

(52) Schlessman, J. L.; Woo, D.; Joshua-Tor, L.; Howard, J. B.; Rees, D. C. Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum. *Journal of Molecular Biology* **1998**, *280*, 669–685.

(53) Chang, C. L.; Davis, L. C.; Rider, M.; Takemoto, D. J. Characterization of nifH mutations of Klebsiella pneumoniae. *Journal of Bacteriology* **1988**, *170*, 4015–4022.

(54) Pednekar, D.; Tendulkar, A.; Durani, S. Electrostatics-defying interaction between arginine termini as a thermodynamic driving force in protein-protein interaction. *Proteins: Structure, Function, and Bioinformatics* **2009**, *74*, 155–163.

(55) Lee, D.; Lee, J.; Seok, C. What stabilizes close arginine pairing in proteins? *Physical Chemistry Chemical Physics* **2013**, *15*, 5844.

(56) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **2018**, *7*, e31486.

(57) Suloway, C. J. M.; Chartron, J. W.; Zaslaver, M.; Clemons, W. M. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences* **2009**, *106*, 14849–14854.

(58) Osaka, N.; Hirota, Y.; Ito, D.; Ikeda, Y.; Kamata, R.; Fujii, Y.; Chirasani, V. R.; Campbell, S. L.; Takeuchi, K.; Senda, T.; Sasaki, A. T. Divergent Mechanisms Activating RAS and Small GTPases Through Post-translational Modification. *Frontiers in Molecular Biosciences* **2021**, *8*, 707439.

(59) Baker, R.; Wilkerson, E. M.; Sumita, K.; Isom, D. G.; Sasaki, A. T.; Dohlman, H. G.; Campbell, S. L. Differences in the Regulation of K-Ras and H-Ras Isoforms by Monoubiquitination. *Journal of Biological Chemistry* **2013**, *288*, 36856–36862.

(60) Denessiouk, K. A.; Johnson, M. S. When fold is not important: A common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins: Structure, Function, and Genetics* **2000**, *38*, 310–326.

Chapter 5

THE APO FORM OF THE NITROGENASE IRON PROTEIN

This project was done in collaboration with Siobhán G. MacArdle who developed and performed ATPase activity assays using ion chromatography and performed inductively coupled plasma mass spectrometry to determine protein metal content.

5.1 Introduction

Proteins that contain small inorganic clusters composed of iron and sulfur atoms (FeS clusters) are among the oldest on Earth and have evolved to perform a wide range of biological functions.[1] The FeS clusters typically serve a structural role to support its protein scaffold and also enable complex chemical reactivity that is otherwise inaccessible to non-metalloproteins. Because FeS clusters can adopt multiple redox states and move electrons within the cluster, they can participate in electron transfer and perform acid-base catalysis.[2]

These extraordinary chemical properties enable the FeS cluster containing nitrogenase proteins to catalyze the conversion of dinitrogen to ammonia. The nitrogenase iron protein (Fe protein), which contains a four iron and four sulfur cluster ([4Fe:4S]) harvests low potential electrons from endogenous sources in the cell and delivers them to the nitrogenase molybdenum-iron protein (MoFe protein) for catalysis. [3–5] The Fe protein is a 64 kDa nucleotide-binding homodimer, that symmetrically coordinates its [4Fe:4S] cluster at the dimer interface through two conserved cysteines (Cys97 and Cys132 in *Azotobacter vinelandii* (*Av*)) from each subunit **(Fig. 1A)**.[6]

Although most FeS clusters are buried within a protein to be protected from deleterious oxidation, the cluster of the Fe protein is an exception. In the Fe protein, the [4Fe:4S] cluster bridges a dimer interface at the protein surface. Although this may seem disadvantageous to nitrogenases which are expressed in obligate aerobes, the surface exposure of the cluster enables efficient electron transfer across a protein-protein interface when the Fe protein transiently binds to the MoFe protein. However, there is a careful balance permitting facile electron transfer and preventing futile redox cycles or cluster degradation. Thus, the position and degree of surface exposure of the [4Fe:4S] cluster is modulated by changes driven by the Fe protein

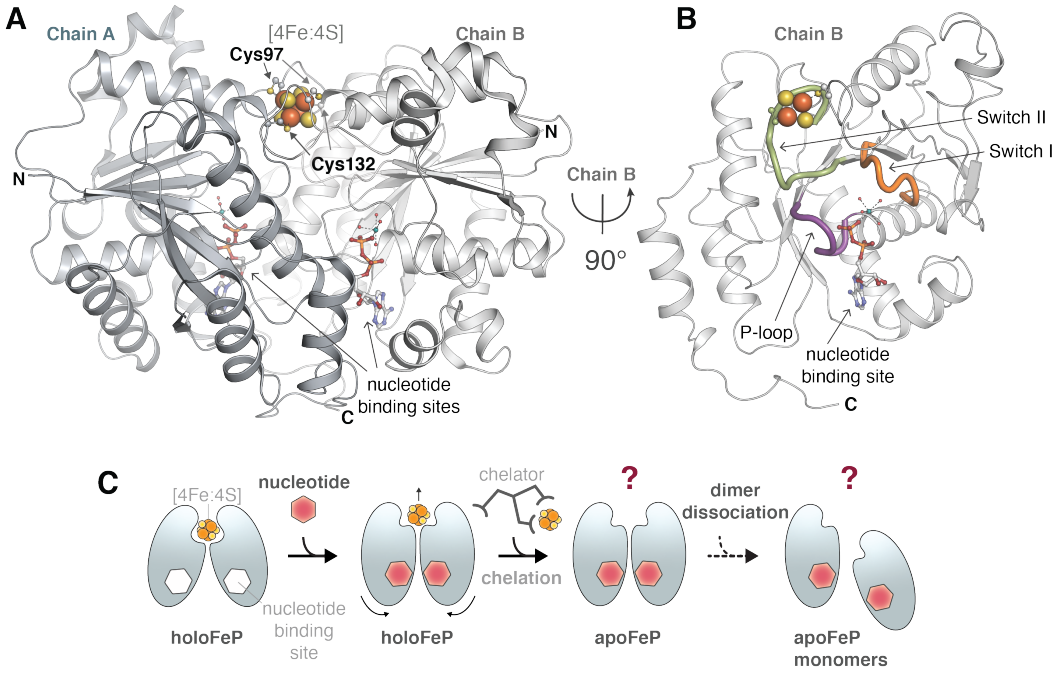


Figure 5.1: **A)** X-ray crystal structure of the homodimeric holo Fe protein with MgADP bound (PDB 6N4L).[7] **B)** One monomer of the holo Fe protein rotated 90 degrees to highlight the three loops intervening the nucleotide binding site and the FeS cluster, the P-loop (purple), switch I (orange), switch II (green). **C)** Schematic of the holo Fe protein cluster chelation. ATP binding to the holo Fe protein causes the [4Fe:4S] cluster to become surface exposed. The surface exposure allows the cluster to be removed by a chemical chelator, forming apo Fe protein. The small surface area that makes up the Fe protein dimer interface could theoretically cause the apo Fe protein to disassociate into monomers, however, apo Fe protein remains dimeric. Question marks indicate how the apo Fe protein remains dimeric is unknown.

nucleotide binding site.[8–10] In this way, electron transfer is tightly coupled to ATP-hydrolysis and only proceeds when the Fe protein is in the correct redox state and docked to the MoFe protein.[11] ATP binding at the Fe protein dimer interface is coupled to the FeS cluster through a series of intervening loops, the P-loop (purple), switch I (orange), and switch II (green) **(Fig. 1B)**. These loops together shift the cluster toward the surface of the protein when complexed to the MoFe protein.[10] This structural rearrangement in conjunction with the lowered redox potential of the cluster,[12, 13] facilitates electron transfer to the MoFe protein.

The surface exposure of the FeS cluster makes the cluster susceptible to removal by chemical chelators, which yields apo Fe protein. Leveraging this feature revealed that the addition of MgATP enhanced cluster chelation, while addition of MgADP

inhibited chelation.[14–16] Thus, it was evident a conformational change associated with ATP binding occurs. However, the structures of both the ATP-bound free Fe protein and the apo Fe protein have not been solved, leaving the full conformational landscape associated with nucleotide-dependent dynamics unresolved.

Further, the identical protein subunits of the Fe protein dimer have minimal protein-protein contacts in its nucleotide-free form but are held together by covalent linkage through its FeS cluster.[6] This suggests that cluster loss would cause dissociation of the dimer into its monomeric subunits **(Fig. 1C)**. However, the apo Fe protein remains dimeric.[17] This raises the question of how the dimer is held together without its cluster, which could be addressed by structural characterization of the apo Fe protein.

The functional roles of the apo Fe protein also remain unclear. During nitrogen fixation, the Fe protein performs both electron transfer and ATP-hydrolysis. Because the apo Fe protein is missing its FeS cluster, it is inactive in electron transfer. However, whether the apo Fe protein retains ATP-hydrolysis activity is not known.

Here, we report the X-ray crystal structure and catalytic ATP-hydrolysis activity of the apo Fe protein. We found crystallization conditions for the apo Fe protein and determined its structure to 1.7 Å resolution. We developed a protocol to isolate pure apo Fe protein from dissolved crystals for further characterization. The apo Fe protein adopts a significantly altered dimer conformation relative to known structures of the holo Fe protein. We found that this new dimer conformation is stabilized by a restructured interface, mediated by nucleotide binding. We further demonstrated that the apo Fe protein is capable of hydrolyzing ATP in the presence of the MoFe protein, using ion chromatography to quantify free phosphate derived from ATPase activity in nitrogenase. Finally, we solved the structure of apo Fe protein with ATP bound. These results reveal an unexpected activity of the Fe protein and broaden our understanding of its conformational landscape. Additionally, our findings widen our perspective of FeS cluster proteins that function without their clusters and highlight the importance of characterizing FeS cluster proteins in their apo forms.

5.2 Results

Generation of apo Fe protein crystals from holo Fe protein crystals

Crystallization of the apo Fe protein began with an effort to determine novel structures of Fe protein by crystallizing the holo Fe protein under varying conditions. As is typical, the holo Fe protein formed plate-shaped brown crystals, reflecting

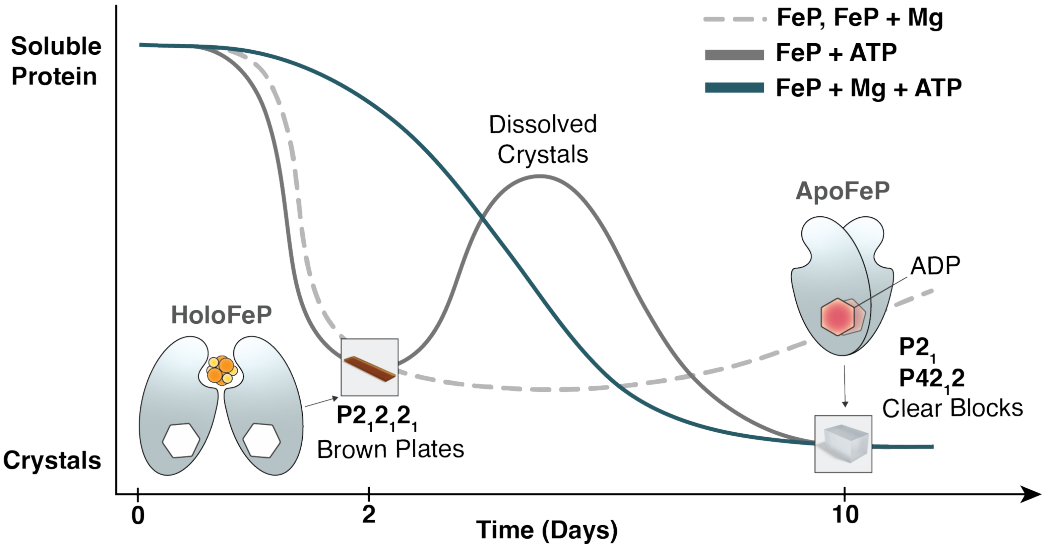


Figure 5.2: Schematic of crystallization kinetics. Both holo Fe protein alone and holo Fe protein co-crystallized with $MgCl_2$ (dashed gray line) form brown, plate-like crystals within 2 days. These crystals that have $P2_12_12_1$ space group symmetry contain Fe protein with intact [4Fe:4S] and no nucleotide bound. However, addition of ATP to holo Fe protein (solid gray line) before crystallization generated plate-like crystals after 2 days that then spontaneously dissolved back into the crystal drop solution. After 1 week, new crystals appeared that were clear and have either $P2_1$ or $P42_12$ space group symmetry. Both new crystal forms had Fe protein that were lacking the [4Fe:4S] cluster (apo Fe protein) and had ADP bound. Finally, addition of MgATP to holo Fe protein (solid green line) before crystallization generated clear apo Fe protein crystals without ever forming brown, plate-like crystals.

the presence of an intact [4Fe:4S] cluster. To our surprise, we discovered crystallization conditions in which holo Fe protein formed crystals that were only stable for the first 48 hours, after which the crystals dissolved and re-crystallized into a new morphology (**Fig. 2**, solid gray line). These new crystals formed from the same mother liquor, but in contrast to the original crystals, were clear in color. We suspected the loss of color might indicate the loss of the FeS cluster. Although this seemed unlikely considering the crystal drops were not perturbed with additional reactants and were kept under a strict anoxic environment throughout the experiment, we harvested these crystals for structural determination. The subsequent analysis confirmed that the clear crystals contained apo Fe protein that lacked the [4Fe:4S] cluster.

To better understand the kinetics of apo Fe protein crystal formation, we monitored crystallization under various conditions over several days. As a general trend,

depicted in **Fig. 2**, the formation of clear crystals was dependent on the presence of nucleotide. Holo Fe protein alone or holo Fe protein preincubated with $MgCl_2$, but without nucleotide, yielded brown crystals, that remained stable for several weeks **(Fig. 2**, gray dashed line**)**. When holo Fe protein was preincubated with ATP or ADP, similar brown crystals formed within two days. However, these crystals dissolved, and clear crystals appeared several days later from the same crystallization drop **(Fig. 2**, solid gray line**)**. We harvested both the initial brown crystals and the re-formed clear crystals and determined their structures using X-ray crystallography. We confirmed that the brown crystals contained Fe protein with cluster and the clear crystals contained apo Fe protein. We also found that the first brown crystal formation step could be by-passed to directly form clear crystals if the holo Fe protein was pre-incubated with both $MgCl_2$ and ATP **(Fig. 2**, solid green line**)**. This MgATP dependence is consistent with previous findings whereby cluster chelation of the Fe protein is facilitated by the presence of MgATP.[14, 15, 18]

X-ray crystal structure of the holo Fe protein

We solved the holo Fe protein structure to 2.06 Å resolution from the brown crystals that grew within the first two days of the crystallization experiment. These crystals had $P2_12_12_1$ space group symmetry and were phased by molecular replacement using only the protein atoms of PDB 6N4L.[7] The resultant structure had two copies of the holo Fe protein dimer **(Fig. 3A)** and contained unambiguous electron difference density for an intact [4Fe:4S] cluster. Although the holo Fe protein was co-crystallized with ATP, there was no density for bound nucleotide and these structures closely resembled other nucleotide-free holo Fe protein structures.[6]

X-ray crystal structure of the apo Fe protein reveals a new dimer orientation

We determined the structure of apo Fe protein to 1.62 Å resolution from the clear crystals. These crystals exhibited either $P2_1$ or $P42_12$ space group symmetry and were phased using only the protein atoms of a previously published holo Fe protein structure (PDB 6N4L).[7] The apo Fe protein had no difference electron density for FeS cluster and the loops that harbor the cluster coordinating cysteines, Cys97 and Cys132, were disordered and splayed 40 Å apart **(Fig. 3B)**. Although Cys97 and its neighboring residues were disordered, the density for Cys132 was still visible, confirming that no cluster or partial cluster was coordinated. The preceding loop (residues 127-132) that normally extends toward the cluster in the holo Fe protein, was folded back into the monomer and packed against α-helix 5 (residue

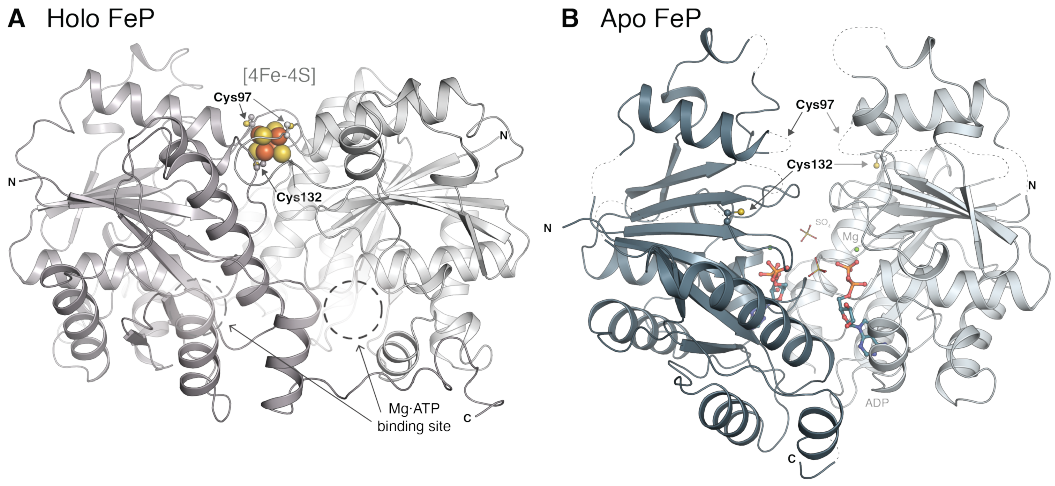


Figure 5.3: **A)** 2.1 Å resolution holo Fe protein crystal structure from the brown crystals which form after 2 days. The holo Fe protein crystals were co-crystallized with ATP, but have no nucleotide bound. Their nucleotide bindings sites are indicated. The [4Fe:4S] cluster is shown as spheres and the cluster coordinating cysteines (Cys97 and Cys132) are shown as sticks. **B)** 1.62 Å resolution apo Fe protein structure from clear crystals lacks the [4Fe:4S] cluster and has ADP bound (shown as sticks). The cluster coordinating loops are disordered and are indicated as dotted lines.

154-173). The apo Fe protein crystals that were co-crystallized with ATP had unequivocal difference electron density for ADP bound in the nucleotide binding site. Additionally, the apo Fe protein exhibited a major conformational change relative to the holo Fe protein, where the monomers were rotated roughly 70 degrees about the nucleotide binding domain of each monomer.

This novel dimer conformation was stabilized by a buried surface area of 830 $Å^2$ around the nucleotide binding domains. The cross-subunit interactions at this interface involve hydrogen bonds between Glu154A and Arg213B and between Lys10A with both the bound nucleotide and Glu221B from the partner monomer. Additionally, two arginine residues (Arg187 and Arg213) from each monomer that reside at the Fe protein dimer interface associate upon nucleotide binding. The guanidinium groups of the arginines from each monomer aligned to π-stack across the dimer interface in line with the base of each bound nucleotide **(SI Fig. 2A)**. This striking conformation of four stacked arginine residues resembles the nucleotide binding conformation in structures where Fe protein is trapped in complex with MoFe protein using nucleotide analogues **(SI Fig. 2B)**.[8–10, 19] Together, the restructured dimer interface of the apo Fe protein and its drastic change in the

overall dimer conformation render the cluster coordinating loops too far apart to coordinate [4Fe:4S] cluster.

Characterization of the apo Fe protein from dissolved crystals

To determine whether the apo Fe protein remains a dimer in solution, crystals of the apo Fe protein were anaerobically dissolved and subjected to anaerobic size exclusion chromatography (SEC). The apo Fe protein eluted as a dimer **(Fig. 4A**, blue**)** similarly to the 64 kDa dimeric holo Fe protein **(Fig. 4A**, dark gray**)**. The major apo Fe protein peak eluted slightly later than holo Fe protein, likely due to its change in dimer conformation, which may have increased its hydrodynamic radius. A second peak in the apo Fe protein chromatogram at the column total liquid volume resulted from the remaining crystallization reagents and free nucleotide that absorb at 280 nm. As a control, we measured the elution profile of free ADP that matched the second peak in the apo Fe protein sample **(Fig. 4A**, light gray**)**. Together with the elution volume calibration curve of standardized protein samples **(Fig. 4B)**, the SEC indicated apo Fe protein is dimeric in solution.

The Fe content of our apo Fe protein samples were measured by inductively coupled plasma mass spectrometry (ICP-MS). As expected, the holo Fe protein controls contained roughly four stoichiometric equivalents of Fe per Fe protein dimer, consistent with the presence of a [4Fe:4S] cluster **(SI Table 1)**. We found the apo Fe protein samples contained roughly two equivalents of Fe per dimer. This observation could be explained by three possible scenarios. The first scenario would be that 50% of our apo Fe protein was actually holo Fe protein; the second, would be that the apo Fe protein sample contained predominantly [2Fe:2S] cluster; and the third would be that Fe was adventitiously binding to the apo Fe protein.

To investigate the first two possibilities, we measured our samples by electron paramagnetic resonance (EPR). The holo Fe protein [4Fe:4S] cluster had a distinctive S = ½, g = 2.049, 1.939, 1.864 EPR signal **(SI Fig. 3**, dark red**)**, which was diminished by >90% in the apo Fe protein sample **(SI Fig. 3**, light red**)**. Thus, we determined that the two equivalents of Fe in apo Fe protein was not coming from partially occupied holo Fe protein. We also saw no evidence for [2Fe:2S] cluster by EPR in both the 10K perpendicular and 5K parallel mode spectra **(SI Fig. 3B)**. Oxidized [2Fe:2S] cluster containing Fe protein is EPR silent; however, the samples were prepared with addition of dithionite reductant, which would result in EPR active reduced [2Fe:2S] cluster. Reduced [2Fe:2S] Fe protein has a distinctive

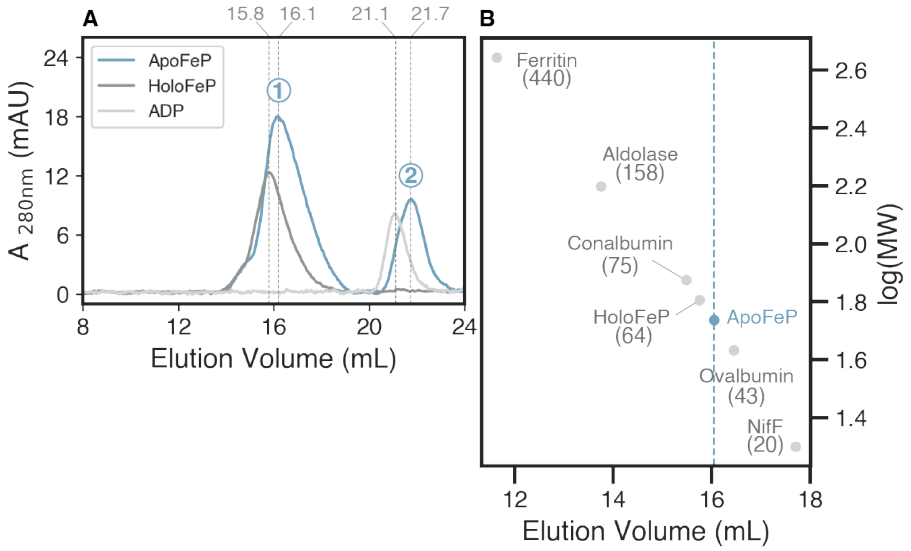


Figure 5.4: **A)** Apo Fe protein (blue) prepared from dissolved clear crystals elutes near the retention volume of holo Fe protein (dark gray), indicating the apo Fe protein is a dimer in solution. The second peak in the apo Fe protein sample at ~21.5 mL likely resulted from free nucleotide or additional crystallization buffers and eluted similarly to ADP alone (light gray). **B)** Elution volume calibration curve of standardized protein samples and their molecular weights (in kDa) for the same size exclusion column indicated apo Fe protein elutes near holo Fe protein (64 kDa). Because the loss of [4Fe:4S] cluster led to the lack of an absorption handle at 410 nm we turned to amino acid analysis for accurate determination of protein concentration. Using only the most reliable amino acids for Fe protein quantitation (alanine, lysine, proline, tyrosine, and valine),[20] we were able to accurately quantitate Fe protein concentrations. This step was critical for then determining the Fe content and ATPase activity of the apo Fe protein samples.

EPR signal with resonances at g = 2.00, 1.94, 1.92, which we did not observe.[21] In addition, we ruled out the possibility that our apo Fe protein sample contained predominantly [2Fe:2S] cluster since we do not observe [2Fe:2S] cluster in the apo Fe protein crystal structures. Although the loops that coordinate the cluster were disordered in the apo Fe protein structures, there was sufficient electron density for Cys132 to confirm that it was not coordinating FeS cluster.

The third possibility, that Fe is adventitiously binding to the apo Fe protein was the most likely explanation and we found that Fe binds strongly in place of Mg in the Mg-ADP binding site. Using X-ray crystallographic Fe-anomalous diffraction, we found signal for Fe both Mg binding sites per apo Fe protein dimer **(SI Fig. 4)**. Double difference anomalous maps were calculated at 7110 eV and 7130 eV, resulting in an electron density peak selective for Fe. Given these pieces of data, we

concluded that the apo Fe protein lacked the [4Fe:4S] cluster and the two equivalents of Fe retained in the sample were binding to the apo Fe protein predominantly in the Mg-binding site.

Apo Fe protein ATP hydrolysis activity

Having established that our apo Fe protein samples were devoid of FeS cluster and having a firm sense of the protein concentration, we sought to determine if the apo Fe protein was functional in ATP hydrolysis. As the holo Fe protein performs ATP-dependent electron transfer, the holo Fe protein only hydrolyzes ATP in the presence of the MoFe protein.[3] Thus, we wondered if apo Fe protein would perform ATP hydrolysis with and without MoFe protein even though it lacks a cluster. Typically, nitrogenase activity assays are performed with an ATP-regeneration system to minimize the accumulation of ADP, which is inhibitory to nitrogenase. This is achieved by the addition of phosphocreatine and phosphocreatine kinase, which catalytically converts ADP to ATP by transfer of the terminal phosphate from phosphocreatine. However, we found the ATP regeneration system interfered with colorimetric-based ATPase assays that involved a color change upon reaction with free phosphate. Thus, we turned to other methods of free phosphate quantification.[22]

We developed a new protocol for quantification of P_i release by the Fe protein ATPase activity using ion chromatography. We favored this method over previous methods for quantifying ATPase activity because no additional reagents were required after completion of the assay and could be measured by automated sample injection into the ion chromatography instrument. We found apo Fe protein was capable of catalytically hydrolyzing ATP and its activity was dependent on the presence of MoFe protein, which was consistent with the behavior of the holo Fe protein (**SI Fig. 5**, open circles). At a Fe protein to MoFe protein component ratio (CR) of 0.05, which corresponds to 10 molar equivalents of MoFe protein active sites per apo Fe protein, we assumed apo Fe protein binding sites were saturated and undergo pseudo first order kinetics. Under these conditions, the apo Fe protein hydrolyzed ATP at a rate of 0.89 ± 0.01 μmol ATP min^{-1} mg Fe $protein^{-1}$ **(Fig. 5A)**. However, some ATPase activity was observed in the MoFe protein alone control, most likely due to a contamination from Fe protein or other ATPases that were not sufficiently removed during the nitrogenase purification. The MoFe protein alone controls were measured at the same concentration of MoFe protein used in the Fe protein to MoFe protein CR 0.05 samples and had an ATP hydrolysis rate of 0.32 ± 0.02 μmol ATP min^{-1} mg Fe $protein^{-1}$. This rate was subtracted from the apo Fe protein hydrolysis

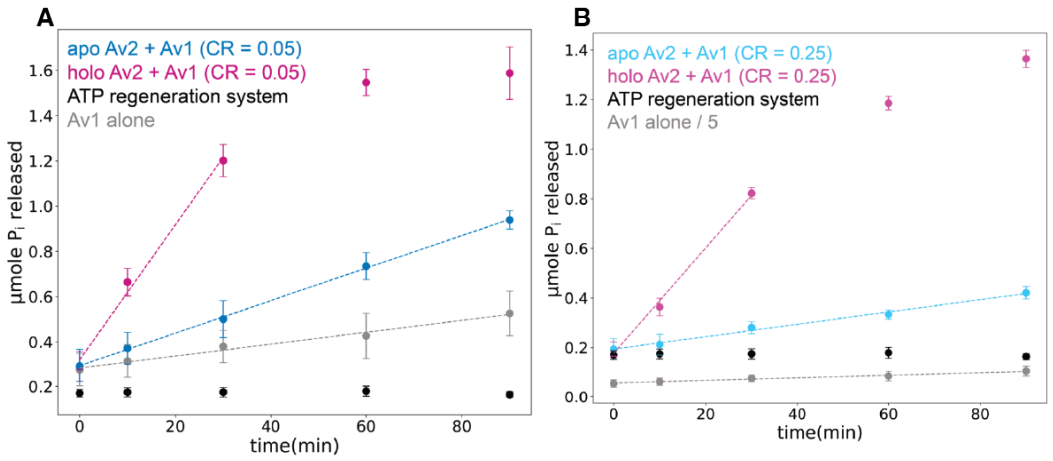


Figure 5.5: ATPase activity assays of apo Fe protein (blue) and holo Fe protein (red) at Left) a component ratio (CR) of 0.05 that corresponds to 10 molar equivalents of MoFe protein active sites per apo Fe protein and **Right)** a CR of 0.25 corresponds to 2 molar equivalents of MoFe protein per apo Fe protein. ATPase activity was determined by measuring P_i concentration in assay samples by ion chromatography. Each point and error bars are the average and standard deviation, respectively of three trials performed on the same day. The ATP regeneration system control (black) had only basal levels of P_i. MoFe protein alone controls contained the same concentration of MoFe protein as the CR of 0.05 samples and 5 times as much MoFe protein as the CR of 0.25 samples. Dashed lines are linear least squares fits of the data (only the linear region of holo Fe protein was used), the slope of which is divided by the amount of Fe protein in the assays (0.008 mg) to obtain the rate of ATPase activity by Fe protein: $\upsilon_{holoFeprotein}$ = 3.7 ± 0.3 µmol P_i min^{-1} mg Fe $protein^{-1}$, $\upsilon_{apoFeprotein}$ = 0.90 ± 0.01 µmol P_i min^{-1} mg Fe $protein^{-1}$, $\upsilon_{MoFeprotein}$ = 0.33 ± 0.02 µmol P_i min^{-1} mg Fe $protein^{-1}$ (for the purpose of comparison with assays with Fe protein, the rate of P_i release was divided by 0.008 mg Fe protein in the MoFe protein alone sample even though no Fe protein was added to this assay).

rate, yielding an adjusted rate of 0.57 ± 0.02 µmol ATP min^{-1} mg Fe $protein^{-1}$.

We found the apo Fe protein ATP hydrolysis rate at a CR of 0.05 is roughly 6 times lower than the hydrolysis rate of holo Fe protein (3.4 ± 0.3 µmol ATP min^{-1} mg Fe $protein^{-1}$) under identical conditions. This discrepancy is further exaggerated at a component ratio of 0.25, which corresponds to two molar equivalents of MoFe protein per Fe protein. The adjusted rates for holo and apo Fe protein were 2.6 ± 0.2 µmol ATP min^{-1} mg Fe $protein^{-1}$ and 0.25 ± 0.02 µmol ATP min^{-1} mg Fe $protein^{-1}$, respectively. This corresponds to an apo Fe protein ATP hydrolysis rate being roughly 10 times lower than that of holo Fe protein. Controls of ATP only, ATP regeneration system only, and all buffer components only, including

phosphocreatine, had only baseline quantities of P_i, which did not increase over time. All together these results indicated apo Fe protein catalytically hydrolyzes ATP, albeit at a lower rate than the holo Fe protein.

X-ray crystal structure of ATP-bound apo Fe protein

While structures of holo Fe protein complexed with ATP analogues when bound to the MoFe protein have provided valuable insights into the ATP-dependent electron transfer process from Fe protein to MoFe protein, obtaining the structure of free Fe protein with ATP bound has proven difficult. After discovering that apo Fe protein is capable of binding ATP in solution, we attempted to capture the ATP-bound apo Fe protein crystal structure by anaerobically soaking crystals with ATP immediately before flash freezing in liquid nitrogen. This method allowed us to determine the structure of free apo Fe protein with ATP bound. The resultant structure contained two molecules of bound nucleotide per dimer with clear density for the ATP γ-phosphate **(Fig. 6B)**, which could not be modeled as ADP or ADP and a water molecule **(SI Fig. 6)**. Further inspection of the ATP binding mode revealed the γ-phosphate of ATP extended out of the nucleotide binding pocket towards the dimer interface. This caused the two terminal phosphates from each monomer to be only 3.0 Å away from each other **(Fig. 6C)**. This seemingly unstable build-up of negative charge was not neutralized by any direct coordination by positive residues. Our ATP-bound apo Fe protein structure revealed a previously unrecognized feature, the formation of a tunnel, which could facilitate release of cleaved phosphate. Several positive residues including Lys41, Lys10, and Arg46 arranged at the surface of the dimer, made a positively charged trap at the end of the tunnel, which extended from the nucleotide binding site to the surface of the dimer **(Fig. 6C, D)**. In fact, a sulfate ion, likely present from the high concentration of ammonium sulfate in the crystallization condition, was found snared in this trap (**Fig. 6C, D**). The surface electrostatics representation shows this tunnel is highly positively charged **(Fig. 6D)**.

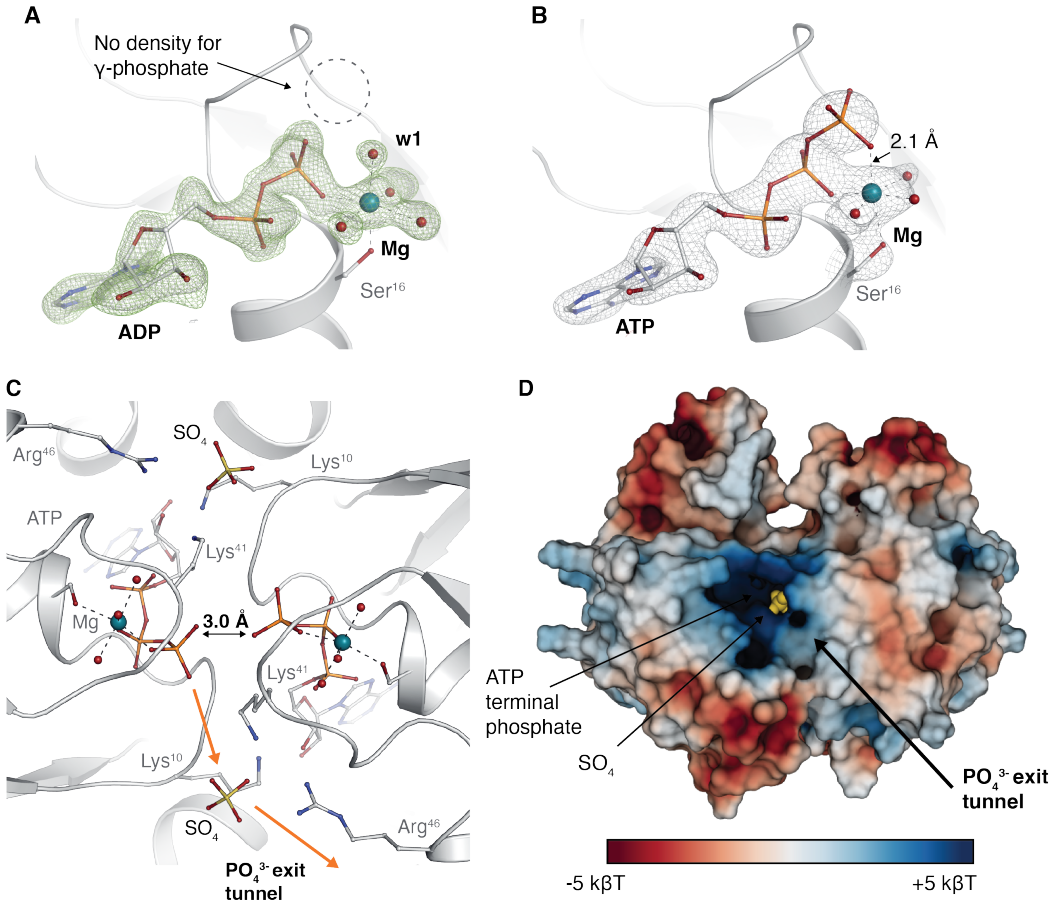


Figure 5.6: ATP-bound apo Fe protein structure revealed the formation of a cleaved phosphate exit tunnel. **A)** Co-crystallization with MgATP yields apo Fe protein crystals with MgADP bound. There is clear density for octahedral Mg coordinated by Ser16, ADP γ-PO_4, and four water molecules. **B)** ATP-bound apo Fe protein has unambiguous electron density for the ATP γ-phosphate coordinated to Mg. **C)** The terminal phosphates of the ATP molecules from each monomer are 3 Å apart from each other across dimer interface causing a buildup of localized negative charge. Cleaved phosphate could shuttle to the protein surface via a positively charged exit tunnel. A sulfate ion from the crystallization solution is trapped in the tunnel. **D)** Surface electrostatics shows the tunnel is positively charged.

5.3 Discussion

The holo Fe protein uses its [4Fe:4S] cluster to transfer electrons to the MoFe protein for nitrogenase substrate reduction. However, the position and degree of surface exposure of the cluster is modulated by changes driven by the Fe protein nucleotide binding site, making the holo Fe protein susceptible to cluster loss in the presence of MgATP. Under newly identified crystallization conditions, the holo Fe protein co-crystallized with ATP resulted in loss of its cluster and crystals of apo Fe protein.

This fortuitous discovery enabled us to structurally characterize the apo Fe protein for the first time.

In the X-ray crystal structure of the apo Fe protein, we observed that the loops containing the conserved the cysteines (Cys97 and Cys132) that coordinate the [4Fe:4S] cluster were disordered and had no intact cluster. The apo Fe protein adopted a dimer conformation not seen in any holo Fe protein structure, where each monomer was rotated such that the cluster binding ligands were separated by 40 Å. The dimer interface was composed of a small surface area of interacting residues near the nucleotide binding site. Apo Fe protein co-crystallized with either ADP or ATP had ADP bound to each monomer. Whether ATP was auto-hydrolyzed on the time-scale of the crystallization (anywhere from 1 to 10 days), or whether Fe protein is catalyzing the hydrolysis is unknown. In solution studies of both the holo Fe protein[23–25] and apo Fe protein (this work), the Fe protein does not hydrolyze ATP at meaningful rates without the presence of the MoFe protein. In the apo Fe protein structures, the bound nucleotide likely contributes to the dimer stabilization through π-stacking interactions with Arg187 and Arg213 from each monomer **(SI Fig. 2A)**.

Although the structure of the apo Fe protein generated from α,α'-dipyridyl chemical chelation has not been solved, studies that mapped the location of surface exposed cysteines by iodo[2-14C] acetate labeling rightly predicted a conformational change occurs upon cluster loss.[16] Our results are consistent with the observation made 40 years ago by Hausinger *et al.* that the apo Fe protein has several surface exposed cysteines, where the holo Fe protein has all buried cysteines (not including the two that coordinate the cluster) **(SI Fig. 7)**.

As the holo Fe protein performs ATP-dependent electron transfer, it is reasonable to assume that the apo Fe protein, which cannot do electron transfer, would not hydrolyze ATP. However, the resemblance of the apo Fe protein nucleotide binding site to that of the holo Fe protein with ATP analogs bound led us to wonder if apo Fe protein retained ATP hydrolysis activity. Thus, we measured the apo Fe protein ATPase activity both with and without MoFe protein and found that the apo Fe protein hydrolyzes ATP in the presence of MoFe protein at a rate of 0.57 ± 0.02 μmol ATP min^{-1} mg Fe $protein^{-1}$. When the MoFe protein is in 10 times molar excess of Fe protein, the ATP hydrolysis rate of apo Fe protein is 6 times lower than that of holo Fe protein.

By soaking apo Fe protein crystals with excess ATP before freezing, we were able

to trap an ATP-bound structure. In this structure, two ATP molecules per apo Fe protein dimer were coordinated with the γ-phosphate of each ATP extended towards the dimer interface. This arrangement resulted in the terminal phosphates being only 3 Å apart, leading to a substantial accumulation of negative charge that was not neutralized by any nearby positively charged residues. This highly unfavorable conformation may lower the activation energy needed to cleave the γ-phosphate. The ATP-bound apo Fe protein structure revealed the formation of a positively charged tunnel that could facilitate the rapid dissociation of cleaved phosphate following ATP hydrolysis **(Fig. 6C, D)**.

This structure inspired us to look for the formation of this phosphate exit tunnel in other structures of the Fe protein. By analyzing six previously published Fe protein structures in different states, we reconstructed the complete ATP hydrolytic cycle. These states include: 1) nucleotide-free,[7] 2) pre-ATP hydrolysis captured by two equivalents of the nonhydrolyzable ATP analog (AMP-PCP),[10] 3) mid-ATP hydrolysis with the ATP transition state analog ADP-AlF_4^-,[8] 4) post-hydrolysis of one ATP molecule captured by an asymmetric complex with one molecule of AMP-PCP and one molecule of ADP bound,[10] 5) post hydrolysis of both ATP molecules with two equivalents of ADP bound while Fe protein is bound to MoFe protein,[9] 6) free Fe protein with ADP bound, no longer complexed with MoFe protein[7] **(Fig. 7)**. The phosphate exit tunnel can be clearly visualized by surface electrostatics calculated using the Adaptive Poisson-Boltzmann Solver.[26] The tunnel and the positive patch at its surface forms just prior to the first ATP hydrolysis step and subsequently closes, reverting to a neutral cavity after the last ATP hydrolysis step is complete.

The hydrolysis of ATP by apo Fe protein raises questions about how ATP is utilized and whether apo Fe protein couples ATP-hydrolysis to some other process, aside from electron transfer. It is known that the Fe protein has alternate functions, namely in the biosynthesis and insertion of the MoFe protein metalloclusters.[27–30] Notably, Fe protein that is unable to perform electron transfer (due to mutagenesis) is still capable of performing its biosynthetic role.[27–29, 31–33] Further, the Fe protein missing its cluster entirely (apo Fe protein) has been shown to be active in the biosynthesis of the MoFe protein metallocofactors.[34–36]

As the apo Fe protein exists in the cell after translation, but before cluster insertion, it is tempting to consider whether this apo form is the relevant form for MoFe protein metallocluster processing. Perhaps the dimer conformation seen in the apo

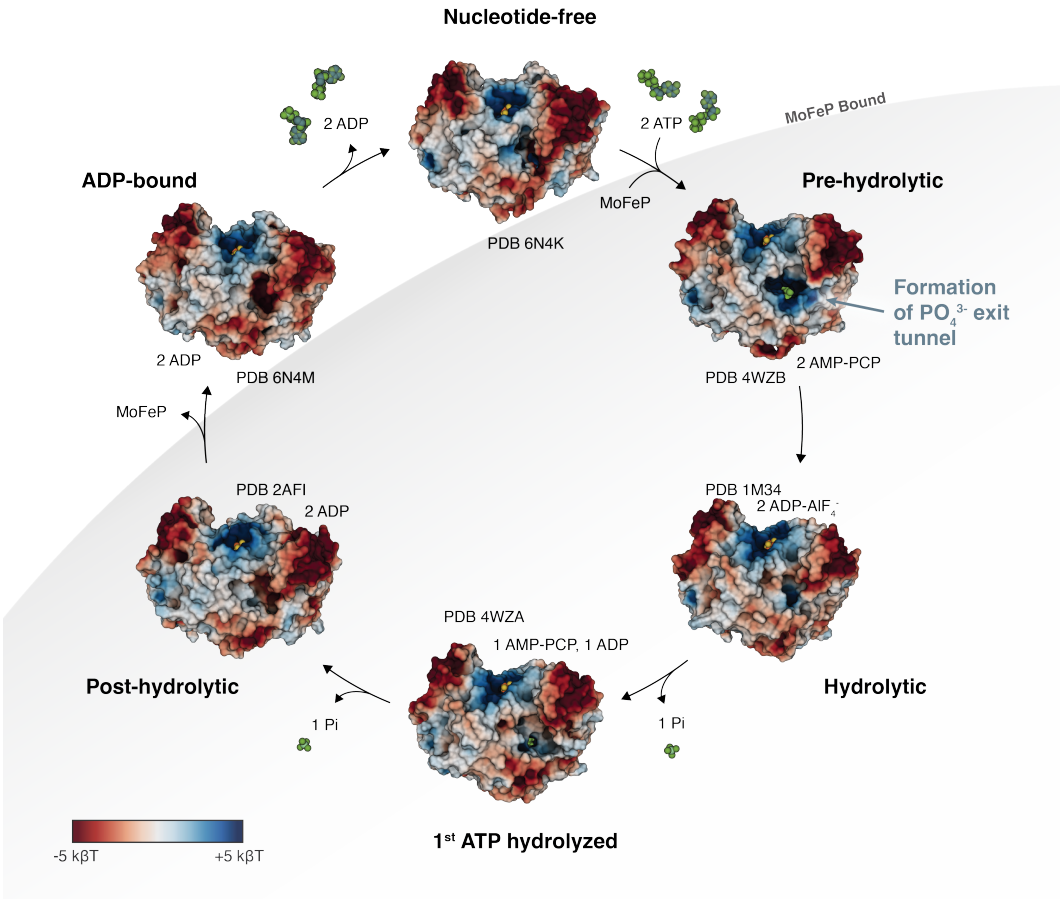


Figure 5.7: Surface electrostatics of the holo Fe protein at six different stages in the ATP hydrolysis cycle. ATP analogs are used to capture ATPases in pre-hydrolytic, hydrolytic, and post-hydrolytic states. The formation of a positively charged phosphate exit tunnel begins at the pre-hydrolytic state and progressively becomes neutral after the first and second ATP hydrolysis steps.

Fe protein structure is by design whereby the cysteines that typically coordinate cluster are forced apart to prevent [4Fe:4S] formation. This design would have the useful benefit of freeing the cluster chelating cysteines for other functionalities, such as binding, storing, or transporting free metals or immature clusters. Furthermore, the apo Fe protein is the protein substrate for the FeS cluster insertion machinery, NifU-NifS. Thus, these structural studies prompt consideration of how the apo Fe protein facilitates coordination with the insertion machinery for [4Fe:4S] cluster loading. Characterizing FeS cluster proteins in their apo forms, such as the apo Fe protein, can help us better understand the capabilities and functional versatility of these proteins, shedding light on their potential biological roles beyond their well-established functions.

5.4 Methods

Growth of *A. vinelandii* and nitrogenase purification

Nitrogenase Fe protein and MoFe protein were natively expressed and purified from *Azotobacter vinelandii* using previously described procedures.[37, 38] Briefly, cultures *A. vinelandii* were grown in Burk's minimal medium containing nitrogen limiting conditions (1.3 mM NH_4Cl). Cells were harvested, pelleted, and flash frozen in liquid nitrogen for storage. All steps after pelleted cells were thawed were performed under strict anaerobic conditions either in anoxic chambers (COY Laboratory products) containing a 97/3% Ar/H2 gas mixture or 100% Ar modified Schlenk line. Cells were lysed by Emulsiflex (Avestin), and subsequently centrifuged to separate cell debris. Fe protein and MoFe protein were individually isolated from the soluble fraction by ion exchange chromatography and further purified by size exclusion chromatography. Both proteins were stored in buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and 5mM dithionite. Samples were flash frozen in liquid nitrogen until further use.

Protein crystallization and structure determination

Purified holo Fe protein concentrated to 15 mg/mL in buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl, and freshly supplemented 5 mM sodium dithionite was crystallized by hanging drop vapor diffusion method. Both holo and apo Fe protein crystals were generated by addition of 1 µL of protein sample to 1 µL of precipitant solution containing 0.2 M ammonium sulfate, 0.1 M Bis-Tris pH 5.5, and 22% (w/v) PEG3350. Holo Fe protein crystallized as brown, plate-shaped crystals after two days. Holo Fe protein that was co-crystallized with 5 mM ATP or 5 mM MgATP crystallized as clear block-shaped crystals after about 10 days. These clear crystals contain apo Fe protein. The ATP-bound apo Fe protein crystals were prepared by soaking crystals by direct addition of MgATP to the crystal drop to a final concentration of 5 mM immediately before freezing. Harvested crystals were cryoprotected in crystallization solution supplemented with 30% ethylene glycol prior to freezing in liquid nitrogen. All crystal growth and manipulations were performed under strictly anoxic conditions.

X-ray diffraction datasets were collected at 100 K at the Stanford Synchrotron Radiation Light Source at beamline 12-2. Data were indexed, integrated, and scaled using HKL3000.[39] Phases were obtained by molecular replacement (MR) with PHASER.[40] The initial MR model was chosen using BALBES.[41] Initial MR was performed using the protein atoms from a previously solved Fe protein structure

PDB 2C8V [42] and subsequent MR was performed using the protein atoms of PDB 6N4L.[7] Model building was performed in Coot [43] and both REFMAC5[44] and Phenix [45] were used for refinement. Software was installed and configured using the SBGrid package manager.[46] All crystallographic figures were prepared using PyMOL molecular graphics software package (Schödinger, LLC). Electrostatics calculations were performed using the APBS PyMOL plugin.[47]

Size exclusion chromatography of dissolved apo Fe protein crystals

Crystals of apo Fe protein were anaerobically dissolved by addition of size exclusion chromatography (SEC) buffer (50 mM Tris-HCl pH 7.5 and 100 mM NaCl) and collected into a 1.5 mL Eppendorf tube. The crystallization solution was removed by buffer exchange into fresh SEC buffer by spin column filtration using a 30 kDa cut off 0.5 mL Amicon Ultra centrifugal filters (MilliporeSigma). All column manipulations were performed under an overpressure of Ar using Schlenk technique. A 24 mL GE S200 increase 10/300 column was prepared by washing with 5 column volumes (CV) of anaerobic SEC buffer, then 5 CV anaerobic SEC buffer supplemented with 5 mM dithionite to remove any residual oxygen. The dithionite was removed by washing again with 5 CV of anaerobic SEC buffer. The apo Fe protein sample was loaded into the injection loop using an airtight syringe immediately before column injection. The samples were eluted at 0.5 mL/min and monitored at 280 and 410 nm.

Inductively coupled plasma mass spectrometry

The iron content of the Fe protein samples were quantified using a Hewlett-Packard 45000 ICP mass spectrometer (Agilent Technologies) with a CETAC ASX-5000 autosampler (CETAC) at the Caltech Resnick Water and Environmental Lab. Both apo (10 µL of 4.74 mg/mL) and holo (2.5 µL or 18.36 mg/mL) Fe protein samples were prepared by addition of 285 µl of 70% nitric acid and subsequent incubation at 50 degrees C for 2 hr. Samples were diluted to a final volume of 10 mL prior to analysis. ICP-MS was performed in both helium mode (kinetic energy discrimination) and oxygen mode (mass shift). Protein concentrations were determined by amino acid analysis at the UC Davis Molecular Structure Facility.

Electron paramagnetic resonance

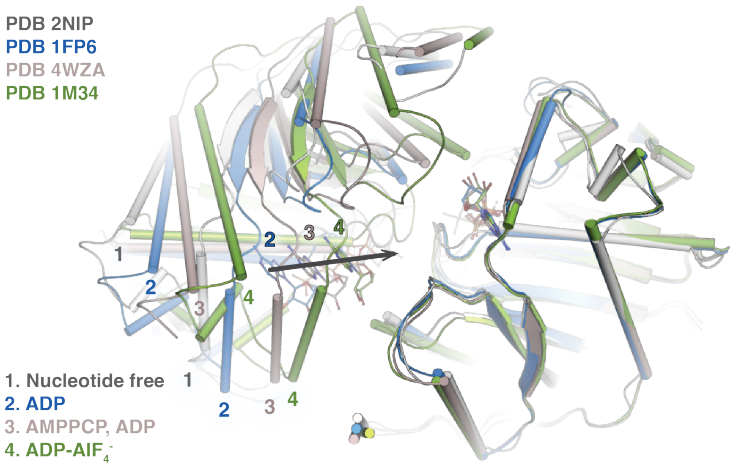
Apo (260 µL of 51.9 mM) and holo (260 µL of 52.1 mM) Fe protein were anaerobically prepared in 4 mm quartz EPR tubes. Both continuous-wave and parallel-mode

EPR were performed at the Caltech EPR Facility on a Bruker EMX spectrometer using Bruker Win-EPR software for data acquisition.

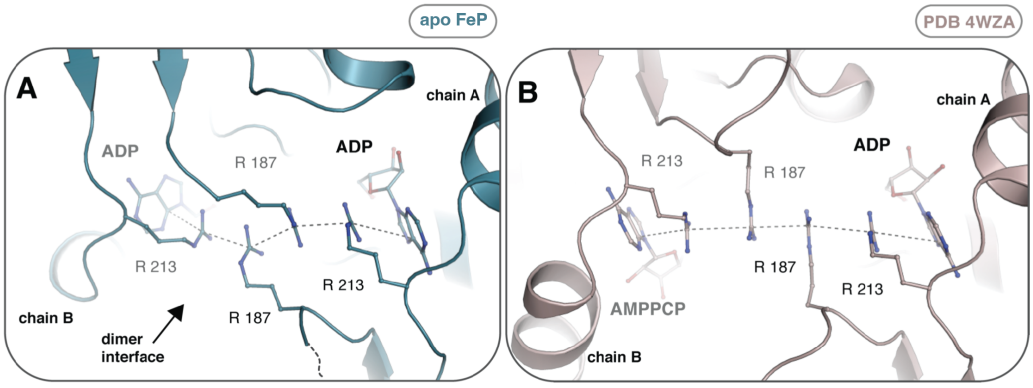
Ion chromatography and ATPase assay

ATP-hydrolysis assays were performed anaerobically (97% Ar and 3% H2 atmosphere). Apo and holo Fe protein were prepared at component ratios (CR) of 0.05 and 0.25 was pre-incubated with the ATP regeneration system (12 mM creatine phosphate, 5 mM $MgCl_2$, 19 units of creatine kinase) and 20 mM dithionite. The reaction was initiated by the addition of ATP (5 mM final concentration). Each 100 µL reaction was quenched by the addition of 25 µL of 400 mM EDTA (pH 7.8). The quenched samples were diluted 20x in milliQ water and analyzed by ion chromatography using a Thermo Scientific Dionex Integrion HPIC with a Dionex VWD. The samples were analyzed by isocratic method and eluents were detected by 214 nm absorbance. Phosphate ion was quantified using a standard curve of phosphate standards.

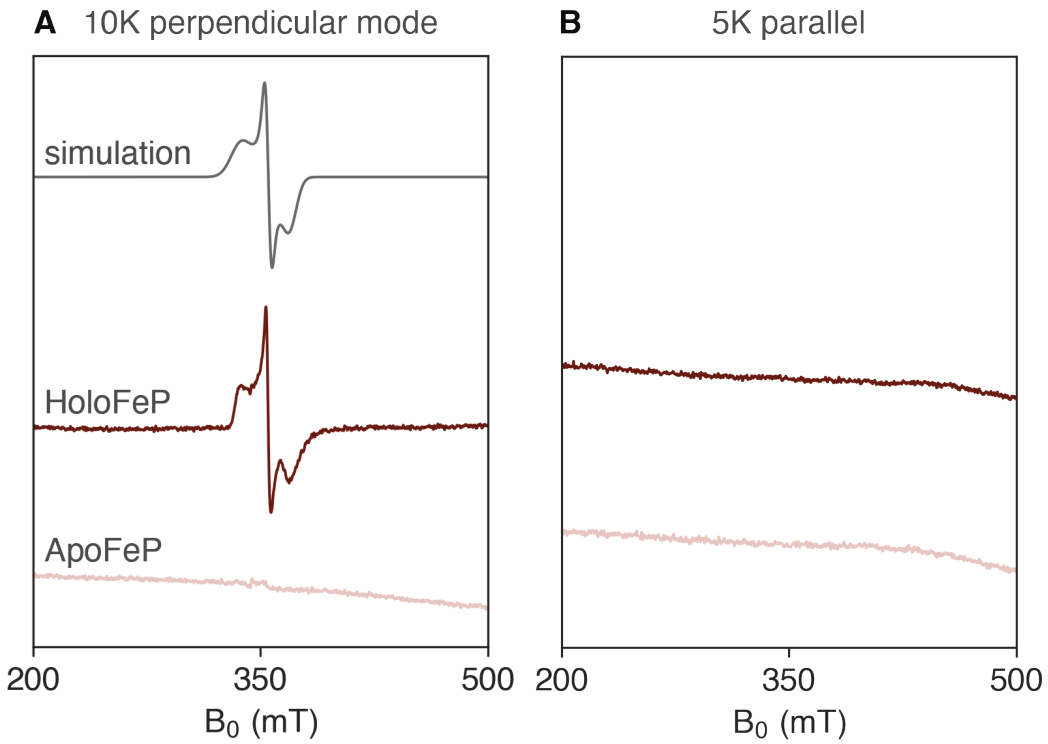
5.5 Supplemental Information



Supplementary Figure 5.1: Overlay of holo Fe protein X-ray crystal structures aligned on one monomer. Structures include Fe protein with 1) no nucleotide bound (PDB 2NIP, gray)[19], 2) ADP-bound (PDB 1FP6, blue)[48], 3) AMP-PCP/ADP-bound (PDB 4WZA, tan)[10], 4) ADP-AlF4- -bound (1M34, green)[49]. Alignment of the structures show the dimer contracts as nucleotides bind and throughout the ATP hydrolysis cycle.



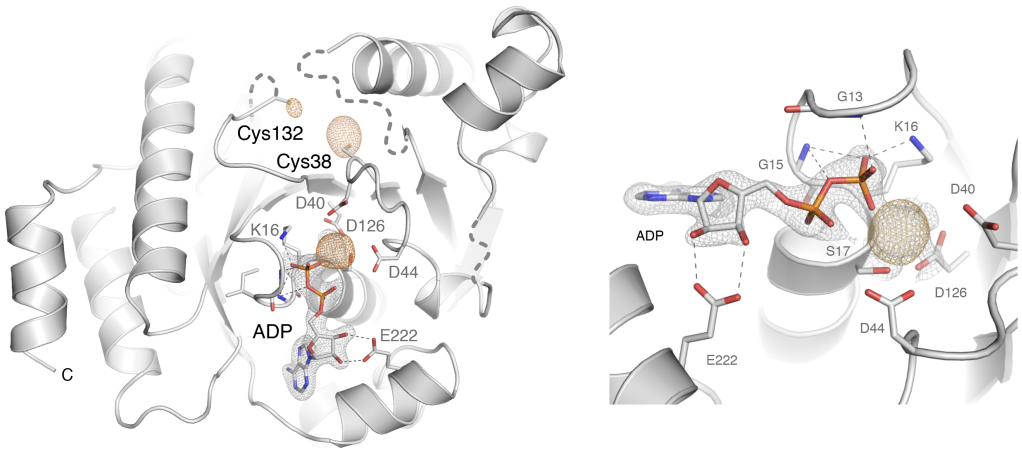
Supplementary Figure 5.2: **A)** A 1.74 Å X-ray crystal structure of the apo Fe protein with ADP bound exhibits arginine stacking that is reminiscent of **B)** the asymmetric nucleotide binding structure of holo Fe protein (PDB 4WZA) with AMP-PCP and ADP bound.[10] The asymmetric structure represents an intermediate in step-wise ATP hydrolysis and P_i release.



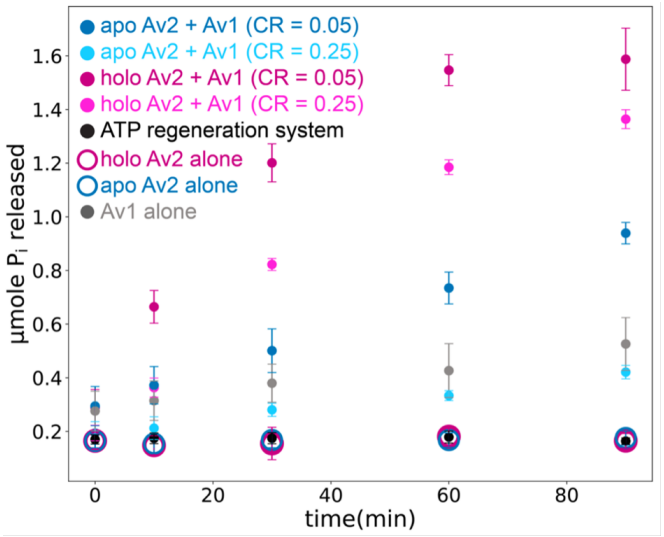
Supplementary Figure 5.3: **A)** Electron paramagnetic resonance (EPR) spectrum of Fe protein samples. Experimental data of holo Fe protein (dark red) and apo Fe protein (light red) at 10K in perpendicular mode and **B)** 5K parallel mode. Frozen solution spectra were collected at 9.38 GHz with a microwave power of 8.7 mW, a modulation amplitude of 8.00 G, a modulation frequency of 100 KHz and conversion time of 42 ms. Simulation of holo Fe protein in gray. Holo Fe protein simulation parameters are S=1/2, g=[2.049, 1.939, 1.864], gaussian linewidth 2.5 mT, Hstrain=[439, 100, 280] MHz. Simulated EPR spectra agree with previously reported values.[50]

Table 5.1: Ratio of Fe atoms to Fe protein dimers in apo and holo Fe protein samples measured by ICP-MS.

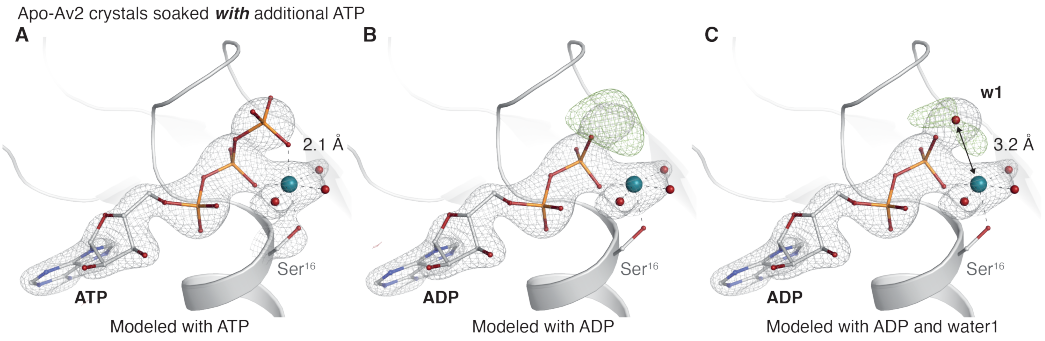
Sample	Fe:Av2 dimer (He)	Fe:Av2 dimer (O_2)
apo Av2 batch 1	2.5 ± 0.3	2.1 ± 0.3
holo Av2	5.3 ± 0.6	4.4 ± 0.5



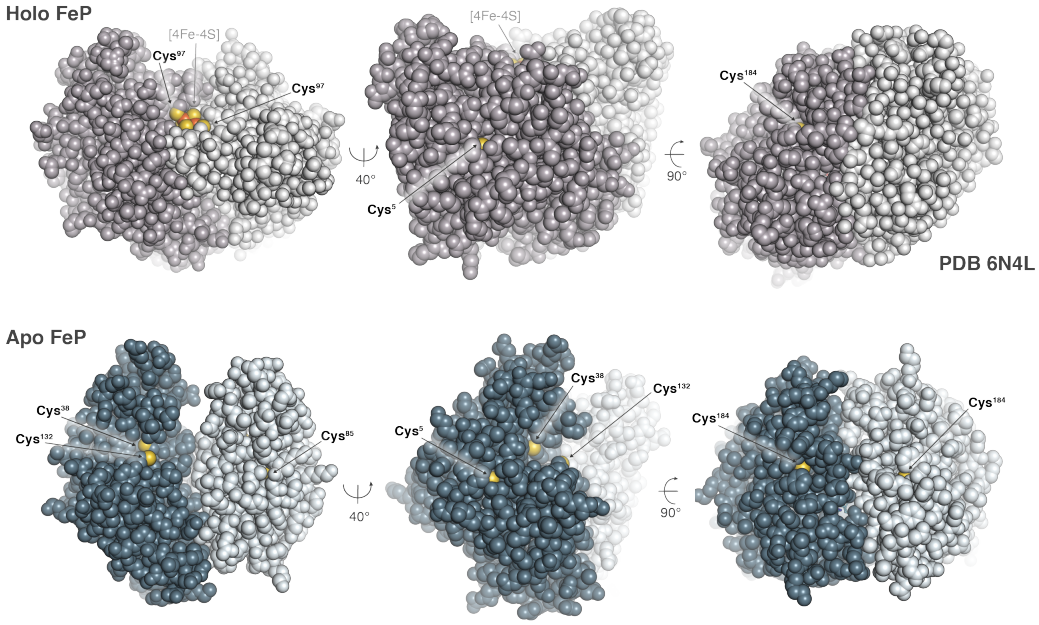
Supplementary Figure 5.4: **A)** One monomer of the apo Fe protein protein dimer is shown in gray cartoon. $2F_o$-F_c density map for bound ADP is shown as gray mesh. Anomalous difference density at the Mg^{2+} binding site is show as orange mesh. Two additional anomalous peaks appear at surface cysteines, Cys38 and Cys132. **B)** Zoom-in of the Mg^{2+} binding site.



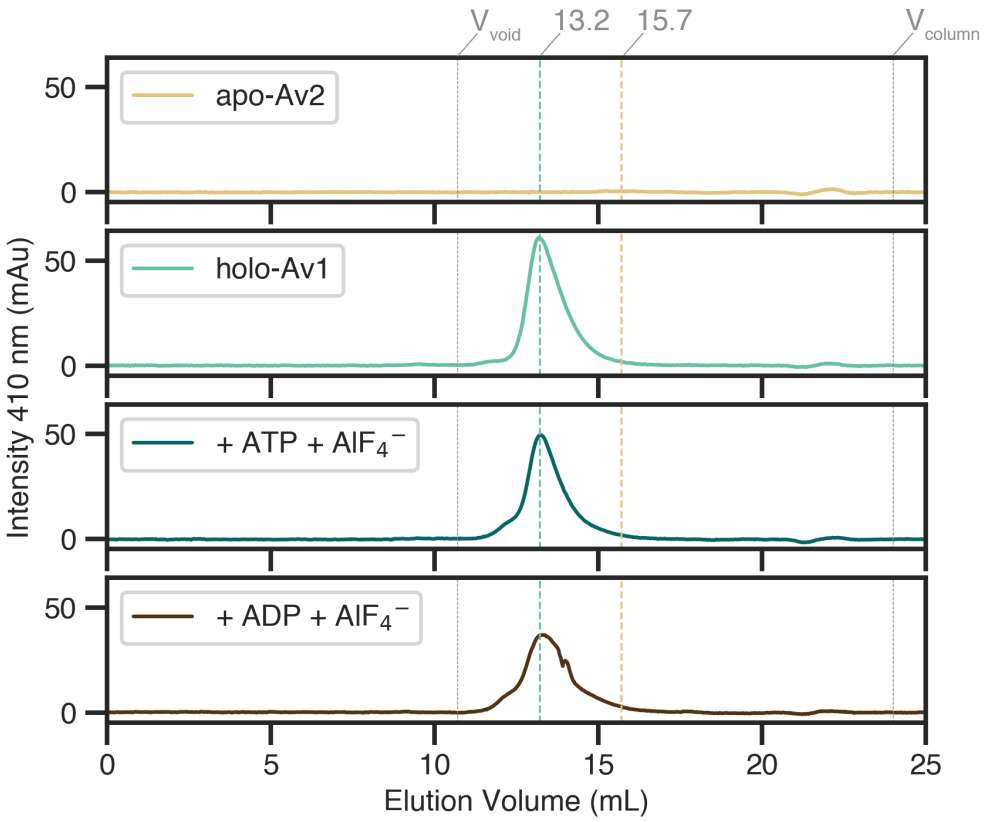
Supplementary Figure 5.5: ATPase activity assays of apo Fe protein (solid blue) and holo Fe protein (solid red) at two different component ratios (CR). A CR of 0.05 corresponds to 10 molar equivalents of MoFe protein per apo Fe protein and a CR of 0.25 corresponds to 2 molar equivalents of MoFe protein per apo Fe protein. ATPase activity was determined by measuring P_i concentration in assay samples by ion chromatography. Each point and error bars are the average and standard deviation, respectively of three trials performed on the same day. Controls of apo Fe protein alone (open blue), holo Fe protein alone (open red), and the ATP regeneration system (solid black) have only basal levels of Pi. MoFe protein alone controls contains the same concentration of MoFe protein as the CR of 0.05 samples and 5 times as much MoFe protein as the CR of 0.25 samples.



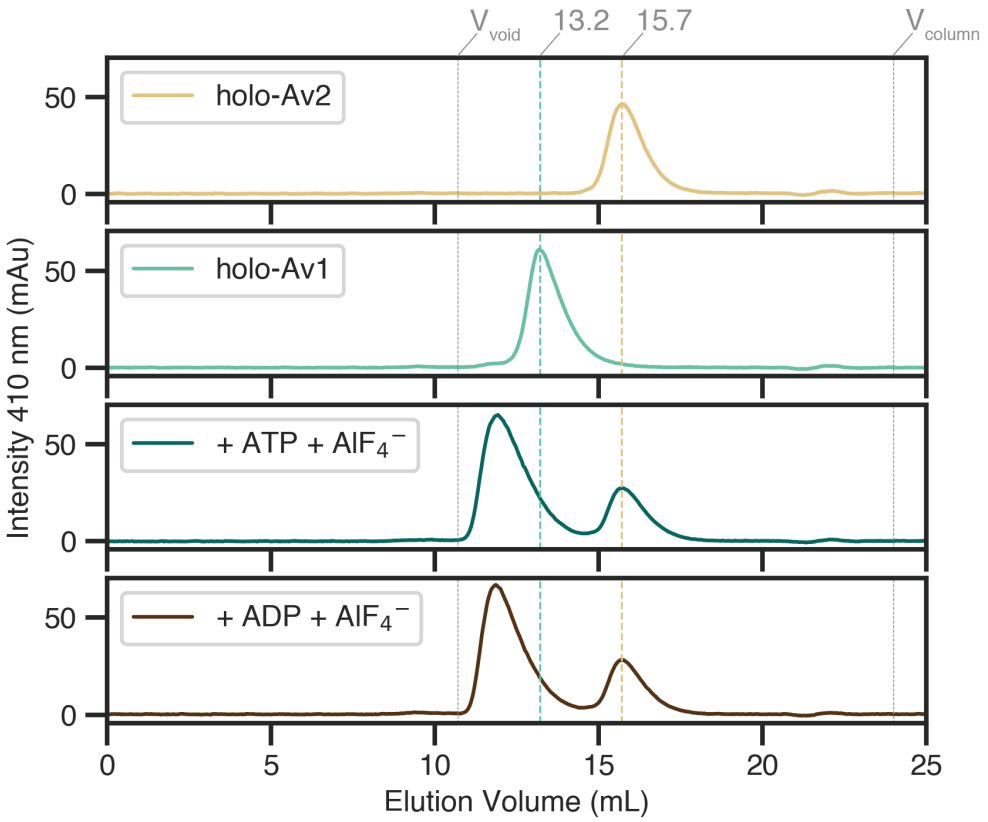
Supplementary Figure 5.6: Refinement of ATP soaked apo Fe protein crystals with **A)** ATP bound, **B)** ADP with no water trans to Ser16, **C)** ADP and an additional water molecule. $2F_o$-F_c shown as gray mesh, 2.0 σ, carve 1.6, F_o-F_c shown as green (positive) and purple (negative) mesh, 3.0 σ, carve 1.6. C is white, O is red, N is blue, P is orange, Mg is teal. The nucleotide density is best modeled as ATP.



Supplementary Figure 5.7: Space filing model of holo Fe protein (top) and apo Fe protein (side) from three angles. Cysteines, aside from the [4Fe:4S] cluster coordinating Cys97 and Cys132, are not surface exposed in the holo Fe protein. Cys5, Cys38, and Cys85 becomes exposed when the Fe protein is apo. Cys132 is also solvent accessible in the apo Fe protein and Cys97 is disordered.



Supplementary Figure 5.8: Anaerobic size exclusion chromatograms apo Fe protein control (yellow), MoFe protein control (light green), reaction mixture after ATP, NaF, $AlCl_3$ incubation with apo Fe protein and MoFe protein (dark green), and reaction mixture after ADP, NaF, $AlCl_3$ incubation with apo Fe protein and MoFe protein (brown). A small shoulder to the left of the MoFe protein peak (at c.a. 12.5 mL) may indicate some complex formation.



Supplementary Figure 5.9: Anaerobic size exclusion chromatograms holo Fe protein control (yellow), MoFe protein control (light green), reaction mixture after ATP, NaF, $AlCl_3$ incubation with holo Fe protein and MoFe protein (dark green), and reaction mixture after ADP, NaF, $AlCl_3$ incubation with holo Fe protein and MoFe protein (brown). The shift of the MoFe protein peak (13.2 mL) to higher elution volume (c.a. 12 mL) indicates the full complexation of holo Fe protein to MoFe protein.

Table 5.2: X-ray crystallographic data collection and refinement statistics.

	Holo FeP	Apo FeP
Data collection		
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1$
Wavelength (Å)	0.97946	0.97946
Cell dimensions		
a, *b*, *c* (Å)	55.85, 137.38, 161.05	55.27, 103.50, 103.71
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.20, 90.0
Resolution (Å)*	50.0-2.06 (2.10-2.06)	50.0-1.74 (1.77-1.74)
R_{merge}	0.143 (0.822)	0.084 (0.925)
R_{pim}	0.039 (0.237)	0.033 (0382)
$I / \sigma I$	20.0 (3.0)	21.1 (2.2)
$CC_{1/2}$	0.999 (0.835)	0.996 (0.784)
Completeness (%)	98.2 (97.7)	90.6 (56.7)
Redundancy	12.9 (11.6)	6.4 (5.8)
Refinement		
Resolution (Å)	43.57-2.06	48.80-1.74
No. reflections	73398	103925
R_{work} / R_{free}**	0.198/0.234	0.237/0.270
No. atoms	9186	
Protein	8772	7019
Ligand/ion	68	128
Water/solvent	346	143
B-factors		
Protein	29.4	21.7
Water/solvent	30.1	22.0
R.m.s deviations		
Bond lengths (Å)	0.003	0.010
Bond angles (°)	1.209	1.656

*Values in parentheses are for highest-resolution shell.

References

(1) Rees, D. C.; Howard, J. B. The Interface Between the Biological and Inorganic Worlds: Iron-Sulfur Metalloclusters. *Science* **2003**, *300*, 929–931.

(2) Beinert, H. Iron-sulfur proteins: ancient structures, still full of surprises. *JBIC Journal of Biological Inorganic Chemistry* **2000**, *5*, 2–15.

(3) Burgess, B. K.; Lowe, D. J. Mechanism of Molybdenum Nitrogenase. *Chemical Reviews* **1996**, *96*, 2983–3012.

(4) Hageman, R. V.; Burris, R. H. Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. *Proceedings of the National Academy of Sciences* **1978**, *75*, 2699–2702.

(5) Hageman, R.; Burris, R. Changes in the EPR signal of dinitrogenase from Azotobacter vinelandii during the lag period before hydrogen evolution begins. *Journal of Biological Chemistry* **1979**, *254*, 11189–11192.

(6) Georgiadis, M. M.; Komiya, H.; Chakrabarti, P.; Kornuc, J. J.; Rees, D. C. Crystallographic Structure of the Nitrogenase Iron Protein from Azo tobacter vinelandli. **1992**, *257*, 8.

(7) Wenke, B. B.; Spatzal, T.; Rees, D. C. Site Specific Oxidation State Assignments of the Iron Atoms in the [4Fe:4S] $^{2+/1+/0}$ States of the Nitrogenase Fe Protein. *Angewandte Chemie International Edition* **2019**, *58*, 3894–3897.

(8) Schindelin, H.; Kisker, C.; Schlessman, J. L.; Howard, J. B.; Rees, D. C. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. *Nature* **1997**, *387*, 370–376.

(9) Tezcan, F. A. Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein. *Science* **2005**, *309*, 1377–1380.

(10) Tezcan, F. A.; Kaiser, J. T.; Howard, J. B.; Rees, D. C. Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase. *Journal of the American Chemical Society* **2015**, *137*, 146–149.

(11) Eady, R. R.; Lowe, D. J.; Thorneley, R. N. Nitrogenase of *klebsiella pneumoniae* : A pre-steady state burst of ATP hydrolysis is coupled to electron transfer between the component proteins. *FEBS Letters* **1978**, *95*, 211–213.

(12) Lough, S.; Burns, A.; Watt, G. D. Redox reactions of the nitrogenase complex from Azotobacter vinelandii. *Biochemistry* **1983**, *22*, 4062–4066.

(13) Lanzilotta, W. N.; Seefeldt, L. C. Changes in the Midpoint Potentials of the Nitrogenase Metal Centers as a Result of Iron Protein-Molybdenum-Iron Protein Complex Formation. *Biochemistry* **1997**, *36*, 12976–12983.

(14) Walker, G. A.; Mortenson, L. E. Effect of magnesium adenosine 5'-triphosphate on the accessibility of the iron of clostridial azoferredoxin, a component of nitrogenase. *Biochemistry* **1974**, *13*, 2382–2388.

(15) Ljones, T.; Burris, R. H. Nitrogenase: the reaction between iron protein and bathophenanthrolinedisulfonate as a probe for interactions with MgATP. *Biochemistry* **1978**, *17*, 1866–1872.

(16) Hausinger, R. P.; Howard, J. B. Thiol reactivity of the nitrogenase Fe-protein from Azotobacter vinelandii. *Journal of Biological Chemistry* **1983**, *258*, 13486–13492.

(17) Howard, J. B.; Davis, R.; Moldenhauer, B.; Cash, V. L.; Dean, D. Fe:S Cluster Ligands Are the Only Cysteines Required for Nitrogenase Fe-Protein Activities. *Journal of Biological Chemistry* **1989**, *264*, 11270–11274.

(18) Deits, T. L.; Howard, J. B. Kinetics of MgATP-dependent Iron Chelation from the Fe-protein of the Azotobacter vinelandii Nitrogenase Complex. *Journal of Biological Chemistry* **1989**, *264*, 6619–6628.

(19) Schlessman, J. L.; Woo, D.; Joshua-Tor, L.; Howard, J. B.; Rees, D. C. Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum. *Journal of Molecular Biology* **1998**, *280*, 669–685.

(20) Hausinger, R. P.; Howard, J. B. The amino acid sequence of the nitrogenase iron protein from Azotobacter vinelandii. *Journal of Biological Chemistry* **1982**, *257*, 2483–2490.

(21) Anderson, G. L.; Howard, J. B. Reactions with the oxidized iron protein of Azotobacter vinelandii nitrogenase: formation of a 2Fe center. *Biochemistry* **1984**, *23*, 2118–2122.

(22) Katz, F. E.; Shi, X.; Owens, C. P.; Joseph, S.; Tezcan, F. A. Determination of nucleoside triphosphatase activities from measurement of true inorganic phosphate in the presence of labile phosphate compounds. *Analytical Biochemistry* **2017**, *520*, 62–67.

(23) Bulen, W. A.; LeComte, J. R. The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proceedings of the National Academy of Sciences* **1966**, *56*, 979–986.

(24) Cordewener, J.; Haaker, H.; Ewijk, P. V.; Veeger, C. Properties of the MgATP and MgADP binding sites on the Fe protein of nitrogenase from Azotobacter vinelandii. *European Journal of Biochemistry* **1985**, *148*, 499–508.

(25) Cordewener, J.; Kruse-Wolters, M.; Wassink, H.; Haaker, H.; Veeger, C. The role of MgATP hydrolysis in nitrogenase catalysis. *European Journal of Biochemistry* **1988**, *172*, 739–745.

(26) Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Science* **2018**, *27*, 112–128.

(27) Filler, W. A.; Kemp, R. M.; Ng, J. C.; Hawkes, T. R.; Dixon, R. A.; Smith, B. E. The nifH gene product is required for the synthesis or stability of the iron-molybdenum cofactor of nitrogenase from Klebsiella pneumoniae. *European Journal of Biochemistry* **1986**, *160*, 371–377.

(28) Robinson, A. C.; Burgess, B. K.; Dean, D. R. Activity, reconstitution, and accumulation of nitrogenase components in Azotobacter vinelandii mutant strains containing defined deletions within the nitrogenase structural gene cluster. *Journal of Bacteriology* **1986**, *166*, 180–186.

(29) Robinson, A. C.; Dean, D. R.; Burgess, B. K. Iron-molybdenum cofactor biosynthesis in Azotobacter vinelandii requires the iron protein of nitrogenase. *The Journal of Biological Chemistry* **1987**, *262*, 14327–14332.

(30) Robinson, A. C.; Chun, T. W.; Li, J. G.; Burgess, B. K. Iron-molybdenum cofactor insertion into the Apo-MoFe protein of nitrogenase involves the iron protein-MgATP complex. *The Journal of Biological Chemistry* **1989**, *264*, 10088–10095.

(31) Shah, V. K.; Davis, L.; Gordon, J. K.; Orme-Johnson, W.; Brill, W. J. Nitrogenase. III. Nitrogenaseless mutants of Azotobacter vinelandii: Activities, cross-reactions and epr spectra. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1973**, *292*, 246–255.

(32) Roberts, G. P.; Brill, W. J. Gene-product relationships of the nif regulon of Klebsiella pneumoniae. *Journal of Bacteriology* **1980**, *144*, 210–216.

(33) Howard, K. S.; McLean, P. A.; Hansen, F. B.; Lemley, P. V.; Koblan, K. S.; Orme-Johnson, W. H. Klebsiella pneumoniae nifM gene product is required for stabilization and activation of nitrogenase iron protein in Escherichia coli. *Journal of Biological Chemistry* **1986**, *261*, 772–778.

(34) Rangaraj, P.; Shah, V. K.; Ludden, P. W. ApoNifH functions in iron–molybdenum cofactor synthesis and apodinitrogenase maturation. *Proceedings of the National Academy of Sciences* **1997**, *94*, 11250–11255.

(35) Curatti, L.; Brown, C. S.; Ludden, P. W.; Rubio, L. M. Genes required for rapid expression of nitrogenase activity in *Azotobacter vinelandii*. *Proceedings of the National Academy of Sciences* **2005**, *102*, 6291–6296.

(36) Rubio, L. M.; Ludden, P. W. Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase. *Annual Review of Microbiology* **2008**, *62*, 93–111.

(37) Burgess, B. K.; Jacobs, D. B.; Stiefel, E. I. Large-scale purification of high activity Azotobacter vinelandii nitrogenase. *Biochimica et Biophysica Acta (BBA) - Enzymology* **1980**, *614*, 196–209.

(38) Spatzal, T.; Aksoyoglu, M.; Zhang, L.; Andrade, S. L. A.; Schleicher, E.; Weber, S.; Rees, D. C.; Einsle, O. Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science* **2011**, *334*, 940–940.

(39) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. *HKL* -3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography* **2006**, *62*, 859–866.

(40) McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D Biological Crystallography* **2005**, *61*, 458–464.

(41) Long, F.; Vagin, A. A.; Young, P.; Murshudov, G. N. *BALBES* : a molecular-replacement pipeline. *Acta Crystallographica Section D Biological Crystallography* **2008**, *64*, 125–132.

(42) Sen, S.; Krishnakumar, A.; McClead, J.; Johnson, M. K.; Seefeldt, L. C.; Szilagyi, R. K.; Peters, J. W. Insights into the role of nucleotide-dependent conformational change in nitrogenase catalysis: Structural characterization of the nitrogenase Fe protein Leu127 deletion variant with bound MgATP. *Journal of Inorganic Biochemistry* **2006**, *100*, 1041–1052.

(43) Emsley, P.; Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60*, 2126–2132.

(44) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **1997**, *53*, 240–255.

(45) Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology* **2019**, *75*, 861–877.

(46) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. Collaboration gets the most out of software. *eLife* **2013**, *2*, e01456.

(47) Holst, M.; Saied, F. Multigrid solution of the Poisson?Boltzmann equation. *Journal of Computational Chemistry* **1993**, *14*, 105–113.

(48) Jang, S. B.; Seefeldt, L. C.; Peters, J. W. Insights into Nucleotide Signal Transduction in Nitrogenase: Structure of an Iron Protein with MgADP Bound . *Biochemistry* **2000**, *39*, 14745–14752.

(49) Schmid, B.; Einsle, O.; Chiu, H.-J.; Willing, A.; Yoshida, M.; Howard, J. B.; Rees, D. C. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF $_4$ $^-$ -Stabilized Structure † ‡. *Biochemistry* **2002**, *41*, 15557–15565.

(50) Lindahl, P. A.; Gorelick, N. J.; Münck, E.; Orme-Johnson, W. H. EPR and Mössbauer studies of nucleotide-bound nitrogenase iron protein from Azotobacter vinelandii. *Journal of Biological Chemistry* **1987**, *262*, 14945–14953.

Chapter 6

CHEMODIVERGENT C(SP^3)-H AND C(SP^2)-H CYANOMETHYLATION USING ENGINEERED CARBENE TRANSFERASES

Material for this chapter is published in: Zhang, J.; Maggiolo, A. O.; Alfonzo, E.; Mao, R.; Porter, N. J.; Abney, N.; Arnold, F. H. Chemodivergent C(sp^3)–H and C(sp^2)–H cyanomethylation using engineered carbene transferases, *Nature Catalysis* **2023**, 6, 152-160 and the full supplemental information can be found therein. J.Z. initiated this project and conducted directed evolution experiments. A.O.M. obtained the X-ray crystal structure of the engineered protein and performed structural analysis.

6.1 Introduction

Given the ubiquity of C–H bonds in organic molecules, advancing selective C–H functionalization methodologies can fundamentally simplify chemical synthesis.[1–5] Ideally, precise and divergent alterations to a molecular scaffold would be achieved by using a panel of highly selective catalysts that can functionalize each unique C–H bond in a molecule, including both C(sp^3)–H and C(sp^2)–H bonds.[6] Methods of this kind are promising in many settings, such as late-stage derivatizations of pharmaceuticals, agrochemicals and materials.[4, 7–9] Notwithstanding the appeal of this approach, design principles and methodologies featuring small-molecule catalysts with complementary selectivity are scarce. Successful examples have exploited directing groups to guide the functionalization of the desired C–H bonds.[10–13] Divergent, highly selective methods that act on desired substrates absent of guiding functional groups are desired.[3]

Enzymes are ideal candidates to address unmet selectivity challenges in C–H functionalization reactions: catalyst-controlled selectivity can be achieved and reprogrammed in enzyme active sites through fine-tuned substrate alignments that enable complementary reaction outcomes. C–H hydroxylases and halogenases exemplify this impressive capability, with divergent chemo-, regio- and stereoselectivity either found in nature or engineered using directed evolution **(Fig. 1a)**.[14–21] Over the past decade, our group and others have broadened the scope of enzymatic C–H functionalization reactions by introducing new transformations originally developed by

synthetic chemists.[22] Examples include repurposing haem proteins and non-haem Fe enzymes to selectively alkylate or aminate C–H bonds via carbene- or nitrene-transfer reactions[23–26] and exploiting non-haem Fe enzymes for radical-mediated selective C–H azidation and nitration reactions.[20, 27, 28] However, most of these efforts have focused on targeting a specific sp^3-hybridized C–H bond **(Fig. 1b)**; chemodivergent approaches to functionalizing C(sp^3)–H bonds and arene C(sp^2)–H bonds present in the same molecule are lacking.

Here, we describe two complementary P450-based carbene transferases that can selectively introduce a cyanomethyl group into a C(sp^3)–H bond or a nearby arene C(sp^2)–H bond **(Fig. 1c)**. Whereas enzymatic and transition metal-catalysed C(sp^3)–H carbene insertion are well documented,[1, 29–35] examples of highly selective intermolecular carbene transfer to an arene C(sp^2)–H bond remain rare.[36, 37] In small-molecule catalysis, high regioselectivity is challenging to achieve via the postulated Friedel–Crafts-like electrophilic aromatic substitution mechanism.[38–41] Site-selective transformations often occur at the least sterically demanding para-position.[42–47] Meanwhile, state-of-the-art biocatalytic systems have been limited to electron-rich heteroaromatics48,49. Inspired by these precedents, we set out to engineer carbene transferases that favour arene C–H functionalization, to complement reported 'C(sp^3)–H alkylases'. This complementary, chemodivergent enzymatic platform can enable straightforward generation of constitutional isomers, which are otherwise laborious to prepare and would require wholly different strategies and starting materials to access. Additionally, nitriles and their derivatives (for example, amides) are well-established functional groups in medicinal chemistry that can be diversified readily in complexity-building transformations.[48, 49] We see this enzymatic platform as a proof of principle and starting point from which to generate new 'C–H cyanomethylases' for functionalization of targeted C–H bonds in complex bioactive molecules.

6.2 Results

Initial activity screening and reaction discovery

We commenced by evaluating the biocatalytic C–H carbene-transfer reaction with N-phenyl-morpholine **1a** and diazoacetonitrile **2**, which has multiple C(sp^3)–H and C(sp^2)–H bonds where cyanomethylation could occur **(Fig. 1d)**. This transformation relies on the generation and transfer of an α-cyanocarbene intermediate,[50] which has been reported in both chemocatalytic and biocatalytic systems for cyclopropanation, N–H and S–H insertion, and indole alkylation reactions.[50–52] C–H

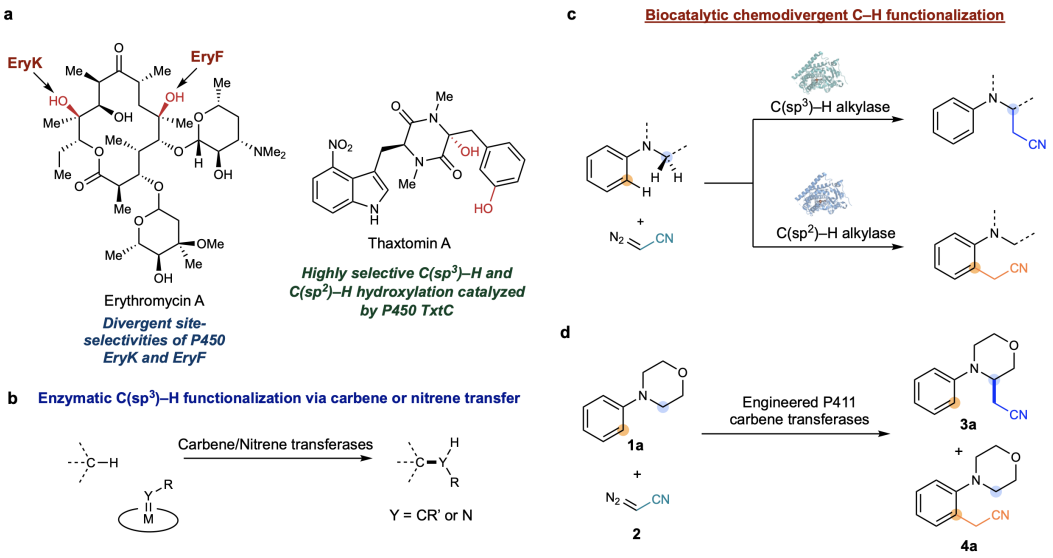


Figure 6.1: Reaction design. **A)** Enzymatic C–H functionalization reactions can be highly selective and divergent. **B)** Previous work on abiological enzymatic C–H functionalization has mainly focused on modifying C(sp^3)–H bonds. **C)** The goal of this study is to demonstrate that closely related C–H alkylases can enable divergent functionalization of arene C(sp^2)–H bonds and nearby C(sp^3)–H bonds. **D)** Model reaction for the initial activity discovery.

cyanomethylation via catalytic carbene transfer has not been reported. In initial studies, we examined the enzymatic C(sp^3)–H cyanomethylation reaction with a panel of 82 different variants of axial serine-ligated cytochromes P450 (P411s, previously engineered for abiological carbene-transfer reactions). The haem cofactor alone is not an active catalyst for this reaction. Many of the P411 variants, however, catalysed the carbene insertion into the α-amino C(sp^3)–H bond of **1a**, affording **3a** with moderate yields. Notably, P411-PFA, a carbene transferase previously engineered to catalyse α-amino C(sp^3)–H fluoroalkylation,[31] afforded the α-amino C(sp^3)–H cyanomethylation product with 11% yield.

We were fascinated to observe that a small number of P411 variants also exhibited basal-level *ortho*-arene C–H functionalization activity affording **4a (Fig. 1d)**. This activity was not seen in previous studies with diazoesters (for example, ethyl diazoacetate) or perfluorinated diazo compounds (for example, 2,2,2-triflurodiazoethane).[30–32] The formation of 4a is not catalysed by the haem cofactor, nor is it produced by the cellular background. Notably, whereas P411-PFA is a selective C(sp^3)–H alkylase, variant P411-FA-E3, which differs by only a single mutation (P401L) to P411-PFA, exhibits the highest selectivity to yield **4a** among

the enzymes tested (**3a**:**4a** = 2:1). Even though P411-FA-E3's carbene-transfer selectivity towards the arene C–H bond was still poor, this result encouraged us to explore the extent to which arene C–H functionalization selectivity could be improved by directed evolution, and whether we could develop a set of chemodivergent C(sp^3)–H and C(sp^2)–H carbene transferases.

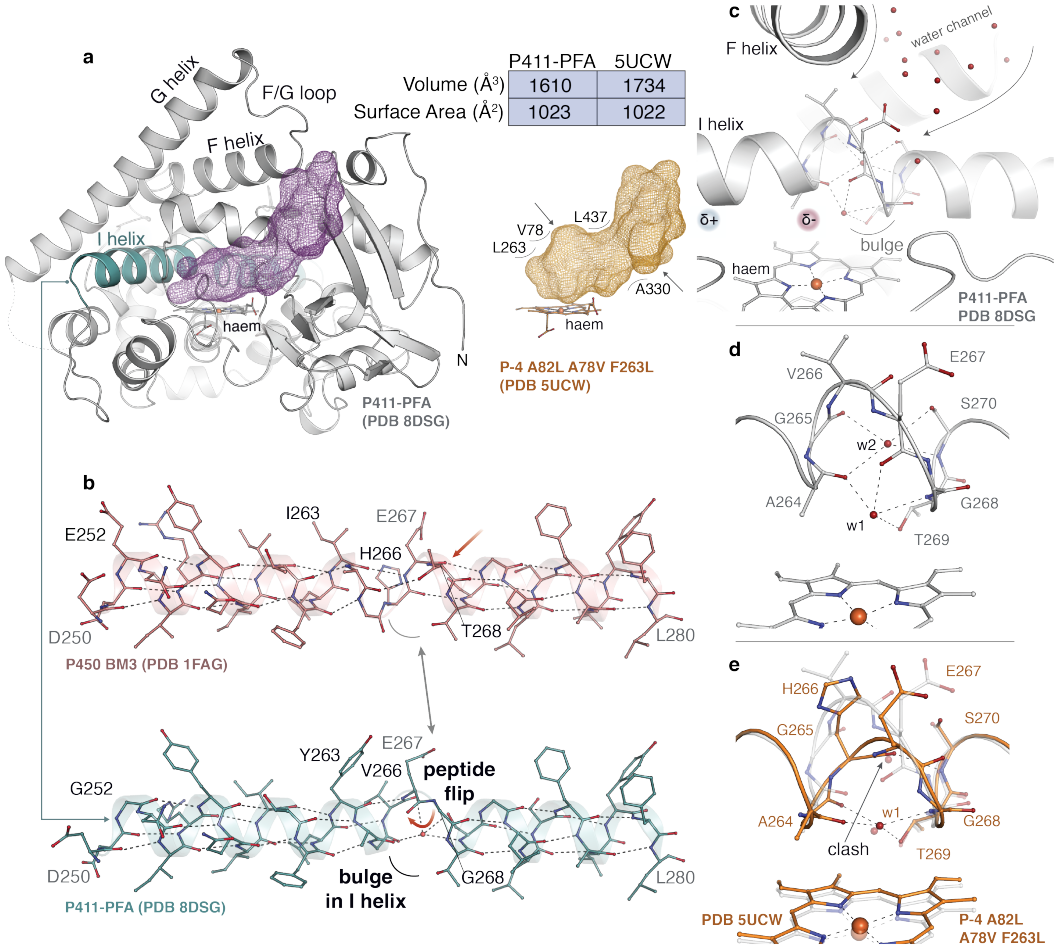


Figure 6.2: Crystallographic studies of P411-PFA (PDB 8DSG). **A)** Overall fold (white cartoon) and active-site cavity of P411-PFA (left, purple mesh) compared to a closely related C–H aminase, P-4 A82L A78V F263L (PDB 5UCW) (right, orange mesh). **B)** An unusual backbone conformation, a peptide flip, is observed in the I helix (teal) of P411-PFA at residue position E267. The E267 backbone carbonyl is indicated by a red arrow. Mutations on the I helix from wild-type P450 BM3 (pink) to P411-PFA are indicated as black residue labels. **C)** Close-up view of the helical bulge that resides at the bottom of a water channel. The disruption of the I helix generates a partial dipole near the active site, which may make the local electronic environment more favourable for diazo binding and activation. **D,E)** Waters (red spheres) stabilizing the bulged helix are shown in **D)** P411-PFA (white) and **E)** P411-PFA overlaid on P411 P-4 A82L A78V F263L (orange).

Structural studies of P411-PFA

To gain structural insight into the α-amino C(sp^3)–H cyanomethylase and help guide the directed evolution of a highly selective arene C–H cyanomethylase, we determined the X-ray crystal structure of P411-PFA's haem domain to 1.87-Å resolution (Protein Data Bank (PDB) 8DSG). P411-PFA adopts an overall architecture similar to previously solved structures of cytochromes P450 and P411 (refs.[53–56]). P411-PFA contains 13 mutations relative to the closest related variant with a determined structure, P-4 A82L A78V F263L (PDB 5UCW), an enzyme capable of nitrene insertion into C–H bonds.[55] Most of the 13 mutations in P411-PFA are near the active site **(Supplementary Fig. 1)** and reshape the substrate cavity into a shallow pocket directly above the haem **(Fig. 2a** and **Supplementary Fig. 2)**. This probably facilitates binding of the smaller carbene precursors and organizes steric interactions to orient the substrate for carbene insertion at the selected position.

Among the structural changes introduced through directed evolution, the most striking feature is an unusual helical backbone conformation observed in the structure of P411-PFA. **(Fig. 2b** and **Supplementary Figs. 3** and **4)**. Specifically, a peptide flip at residue position 267 fully disrupts the helical hydrogen bonding pattern of the I helix. Such drastic structural rearrangements in this region led us to analyse the progression of the I helix hydrogen bonding pattern over generations of directed evolution. In the structure of wild-type P450 BM3 (PDB 1FAG),[56] this region of the I helix exhibits a conformation reminiscent of the i+5→i hydrogen bonding pattern present in a π helix,[57] resulting in a bulge around residue 267. In the evolved P411 P-4 A82L A78V F263L aminase structure, this helical bulge is stabilized by a water molecule (w1) probably present but not modelled in wild-type P450 BM3, indicating a feature that has persisted over many generations of directed evolution of this enzyme **(Fig. 2c–e** and **Supplementary Fig. 3)**.[53–56] Notably, however, this distortion is further expanded in the structure of P411-PFA; the peptide flip present at position 267, which fully disrupts the helical hydrogen bonding network, traps a second water molecule (w2) within an expanded coil and breaks the standard I helix into two distinct helices. This breakage results in the accumulation of complementary dipoles on either side of this expanded coil, altering the electrostatics in the active site **(Fig. 2c** and **Supplementary Fig. 5)**. In the wild-type monooxygenase, residues within this region have been suggested to influence C–H hydroxylation by facilitating oxygen binding and activation.[58] The helical disruption and resultant complementary macrodipoles in this pocket of the carbene transferase may promote binding or activation of the zwitterionic diazo compounds (carbene precursors).

Whereas P411-PFA contains 26 haem-domain mutations compared to wild-type P450 BM3, E267 had not been changed **(Supplementary Fig. 3)**. Instead, mutations flanking the helical bulge around this site, L263Y, H266V and T268G, which have been introduced to enhance carbene and nitrene C–H insertion activities,[30, 54, 55] probably stabilize the otherwise unfavourable flipped conformation of the E267 peptide.

Directed evolution of a selective arene C–H cyanomethylase

To develop highly active and selective arene C(sp^2)–H carbene transferases, we used P411-FA-E3 as the starting enzyme for directed evolution by sequential rounds of site-saturation mutagenesis, error-prone PCR mutagenesis and screening (by gas chromatography–flame ionization detection (GC–FID)) **(Fig. 3a)**. To preserve the helical disruption found in P411-PFA, presumed to be beneficial for carbene-transfer activity, we targeted amino acid residues for site-saturation mutagenesis on the opposite side of the active site from the disrupted I helix and those previously found to affect abiological carbene- and nitrene-transfer activities **(Fig. 3b,c)**. We found beneficial mutations within the proximal active-site pocket (A87V, M177Y, W325C, V330C and M354V) as well as distal ones (S118Q, G252L) that affect the chemoselectivity of the protein catalysts.

Meanwhile, in comparison to P411-FA-E3, α-amino C(sp^3)–H cyanomethylase P411-PFA contains an additional L401P mutation in the axial 'Cys pocket', which disrupts the hydrogen bond between the amide hydrogen of residue 402 and axial serine ligand **(Supplementary Fig. 6)**; mutations in this pocket are known to tune the electronic properties of haem.[53, 59, 60] The different selectivities of these two variants encouraged us to investigate whether mutations at other residues in the Cys pocket improve arene C(sp^2)–H functionalization selectivity. F393W was identified as a beneficial mutation from site-saturation mutagenesis of this region. In summary, a total of seven rounds of mutagenesis and screening yielded the final variant P411-ACHF (arene C–H functionalization enzyme), which contains eight additional mutations relative to P411-PFA **(Fig. 3b)**. Under yield-optimized reaction conditions, arene C(sp^2)–H cyanomethylase P411-ACHF delivers **4a** in 51% yield and excellent chemoselectivity (**3a**:**4a** = 1:17).

Like many small-molecule arene C–H carbene-transfer catalysts,[38–42, 47, 61–63] P411-ACHF was found to exhibit (low) Buchner ring expansion activity. This activity was not seen in the parental C(sp^3)–H alkylase, P411-PFA. We observed

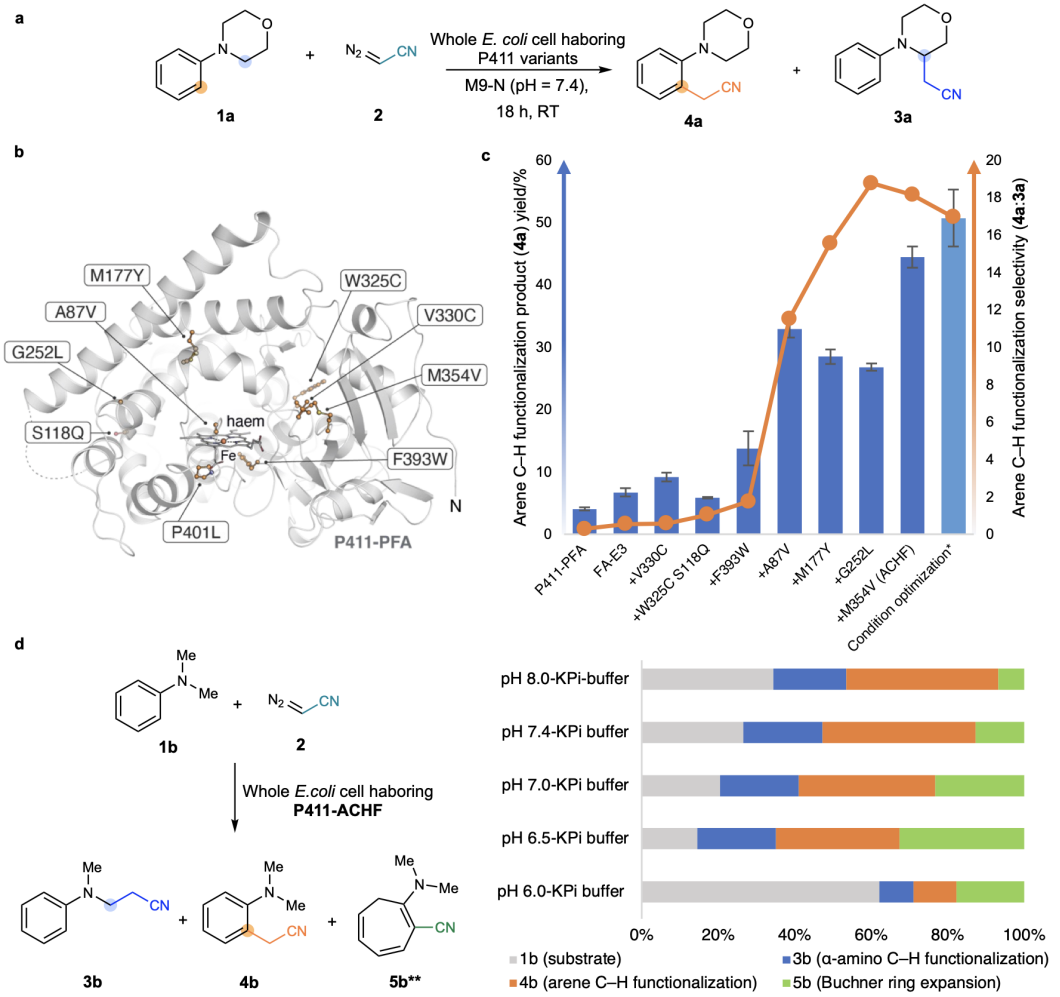


Figure 6.3: Directed evolution of a selective arene C–H cyanomethylase. Reaction conditions: anaerobic; 5 mM **1a** or **1b**; 48 mM **2** (9.63 equiv.); *E. coli* whole cell harbouring P411 variants (OD_{600} = 30) suspended in M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (cosolvent); room temperature; 18–20 h. **A)** Initial activity screening revealed the formation of C(sp^3)–H insertion product **3a** and *ortho*-arene C–H alkylation product **4a**. **B)** Locations of beneficial mutations are highlighted in the P411-PFA structure. **C)** Directed evolution of a highly selective arene C(sp^2)–H cyanomethylase; *E. coli* whole cell harbouring P411- ACHF (OD_{600} = 30) suspended in M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (cosolvent); room temperature; 18–20 h). Reactions were performed in triplicate (n = 3). Yields reported are the means of three independent experiments. Error bars indicate the standard deviation of the yields. **D)** Observed product distribution with 1b includes ring-expanded Buchner products (**5b**, **inseparable mixture of tautomers).

the highest level of Buchner product formation when the enzyme was challenged with **1b** and **2** **(Fig. 3d)**. In that case, P411-ACHF affords cycloheptatriene **5b**,

presumably through cyclopropanation, electrocyclic ring opening and tautomerization. The enzymatic Buchner ring expansion reaction was found to be sensitive to the pH of the reaction buffer **(Fig. 3d)**. It was observed that basic conditions (pH 7.4–8.0) suppressed Bucher product formation. Thus, the scope of arene C–H functionalization with P411-ACHF was examined under basic conditions.

Substrate scope studies of P411-PFA and P411-ACHF

We investigated the substrate scope of the reaction under yield-optimized conditions with the two highly selective but distinct cyanomethylases, P411-PFA and P411-ACHF. As shown in **Fig. 4a**, these enzymes are capable of alkylating N-substituted arenes with complementary chemoselectivity. Without additional protein engineering, the previously engineered C(sp^3)–H fluoroalkylase P411-PFA efficiently installs cyanomethyl groups to secondary, α-amino C(sp^3)–H bonds with high chemo-, regio- and enantioselectivity. Whereas P411-PFA functionalizes the C(sp^3)–H bonds of these N-substituted arenes (**3a–e**; **Fig. 4a**), P411-ACHF alkylates their arene *ortho*-C–H bonds (**4a–e**). Notably, even though introduction of the cyanomethyl group to an arene C(sp^2)–H bond does not establish a new stereocentre, when a racemic mixture of **1j** is used as the substrate, P411-ACHF preferentially converts the (*R*)-enantiomer into the *ortho*-cyanomethylation product (**4j**) with 81:19 e.r. via a kinetic resolution **(Fig. 4a)**.

Substrate recognition in the enzyme active site is key to carbene-transfer selectivity. P411-PFA preferentially functionalizes the secondary α-amino C(sp^3)–H bonds in the presence of other C–H bonds with similar bond-dissociation energies (for example, benzylic C–H bonds in **1d** and **1e**). Meanwhile, P411-ACHF exhibits exclusive selectivity towards *ortho*-C–H bonds (to nitrogen) over other activated positions (positions *ortho* to –OMe in **4m**). Transition metal-catalysed carbene transfer to C(sp^2)–H bonds of N-substituted arenes, in contrast, favours the kinetically more accessible *para*-positions.[42–47] The *para*-positions are probably sterically occluded by the enzyme, leading to the observed *ortho*-selectivity. Overall, the substrate scope studies reveal that the multiple potential non-covalent interactions between the enzymes and the substrate are key to determining carbene-transfer selectivity, and a small number of mutations can cause these new-to-nature enzymes to discriminate nearby C–H bonds.

Although chemoselectivity of P411-ACHF is not yet excellent with some substrates, further directed evolution should be able to improve activity and selectivity on

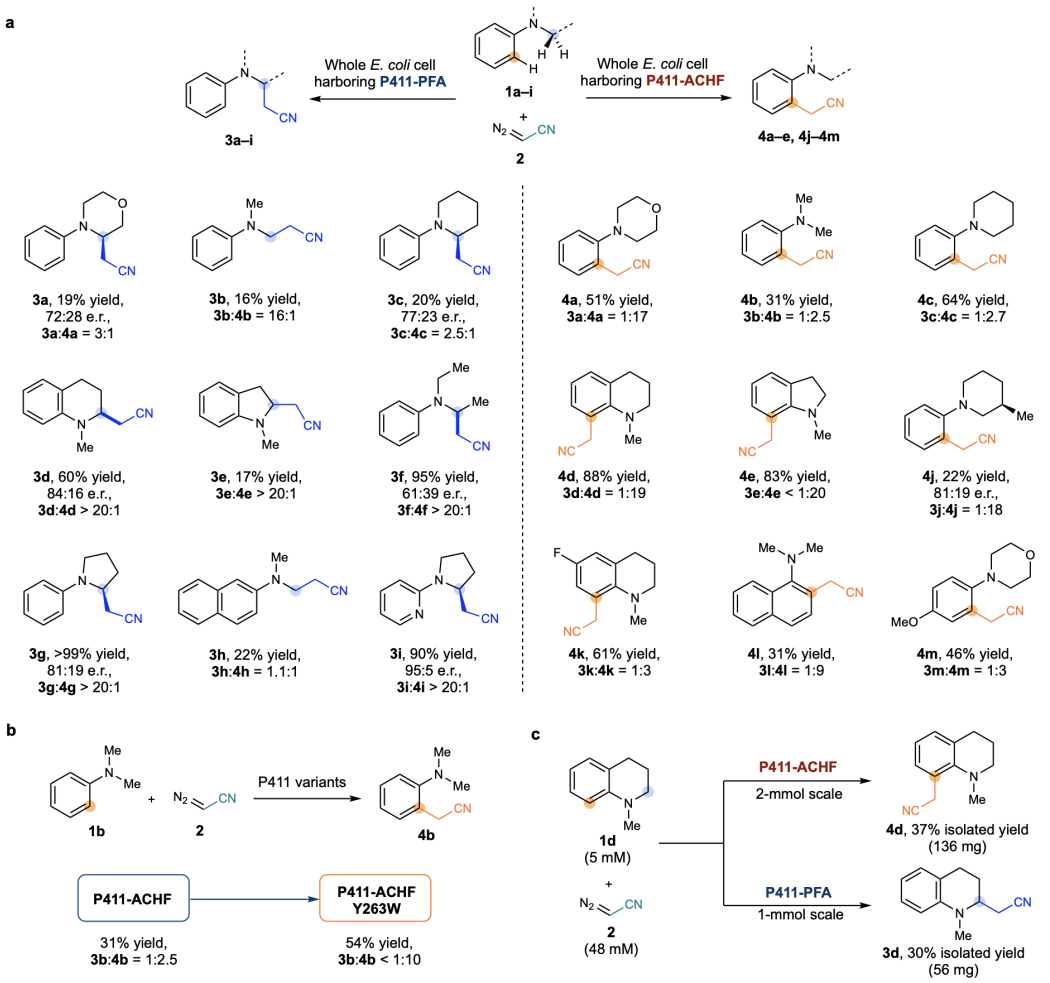


Figure 6.4: Substrate scope study. **A)** Substrate scope of P411-PFA and P411-ACHF. Analytical yields of desired products were determined by using their GC calibration curves. Reaction conditions: anaerobic; 2.5 mM **1**; 51.6 mM **2** (20.63 equiv.); *E. coli* whole cells harbouring P411 variants (OD_{600} = 30) M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (cosolvent); room temperature; 18–20 h. Yields are reported as the means of three independent experiments (n = 3). **B)** The arene C(sp^2)–H functionalization activity and selectivity can be further optimized on individual target molecules. **C)** Chemodivergent, preparative-scale syntheses using P411-PFA and P411-ACHF. Reaction conditions: anaerobic; 5 mM 1d; 48 mM 2 (9.63 equiv.); *E. coli* whole cells harbouring P411 variants (OD_{600} = 30) M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (cosolvent); room temperature; 18–20 h. e.r., enantiomeric ratio.

any specific substrate. To demonstrate this, we performed an additional round of directed evolution using **1b** as the new model substrate and identified Y263W as a beneficial mutation. The optimized variant, P411-ACHF Y263W, with decreased active-site volume, delivers **4b** with 54% yield and <1:10 **3b**:**4b** ratio **(Fig. 4b)**.

To demonstrate the synthetic use of these enzymes, we performed chemodivergent derivatization of **1d** using P411-PFA and P411-ACHF on a preparative scale, where the enzymes yielded constitutional isomers **3d** and **4d** with 30% (1-mmol scale, 56 mg) and 37% (2-mmol scale, 136 mg) isolated yield, respectively **(Fig. 4c)**.

6.3 Discussion

We have demonstrated that complementary, closely related P450-based carbene transferases can selectively functionalize either a C(sp^3)–H bond or an arene C(sp^2)–H bond present in the same molecule. These results support our proposal that the divergent, native C–H hydroxylating selectivity widely exhibited by P450s can be re-established for a non-native enzymatic reaction. Given these results, it is reasonable to expect that these enzymes can serve as starting points from which to evolve C–H alkylases to divergently functionalize C(sp^3)–H and C(sp^2)–H bonds with high selectivity, provided some promiscuous activity for the targeted position can be found. In fact, as we have shown here, less selective generalist enzymes are excellent starting points from which to evolve catalysts that are highly selective for a desired transformation.

Furthermore, X-ray crystallographic studies of the C(sp^3)–H carbene transferase, P411-PFA, revealed an unusual helical disruption in the active site of the cyanomethylase that is neither observed in reported P450 structures nor captured by computational structure predictions (AlphaFold2; **Supplementary Fig. 3**). This underscores the advantages of using directed evolution to establish new-to-nature functions in engineered enzymes, which often calls for new structural features that differ from the native proteins and may be challenging to predict or design. In summary, we hope that these complementary C–H functionalization carbene transferases promote the use of bio-catalysts to perform challenging, divergent C–H alkylation reactions with selectivity unmatched by chemical catalysts.

6.4 Methods

Cloning and site-saturation mutagenesis

The genes encoding all enzymes described in this study were cloned using Gibson assembly[64] into vector pET22b(+) (Novagen) between restriction sites *Nde*I and *Xho*I in frame with a C-terminal 6× His-tag. Site-saturation mutagenesis was performed using the '22c trick'67 or 'NNK' as degenerative codons. The PCR products were digested with *Dpn*I, gel purified and ligated using Gibson Mix. The ligation mixture was used to directly transform electrocompetent *Escherichia coli* BL21 (DE3) cells (Lucigen).

Protein expression of P450 and P411 variants in 96-well plates

Single colonies from Luria-Bertani (LB) agar plates with ampicillin were picked using sterile toothpicks and cultured in deep-well 96-well plates containing LB medium with ampicillin (400 µl per well) at 37 °C, 80% humidity and 220 r.p.m. shaking overnight. Subsequently, Hyper Broth medium (AthenaES) with ampicillin (1,080 µl per well) in a deep-well plate was inoculated with an aliquot (120 µl per well) of these over-night cultures and allowed to shake for 3 h at 37 °C, 80% humidity and 220 r.p.m. The plates were then cooled on ice for 30 min and the cultures were induced with 0.5 mM isopropyl-β-d-1-thiogalactopyranoside (IPTG) and 1.0 mM 5-aminolevulinic acid (final concentrations). Expression was then conducted at 20 °C, 220 r.p.m. for 18–20 h.

Reaction screening in whole-cell format

E. coli cells harbouring P411 variants in deep-well 96-well plates were pelleted (3,500 g, 5 min, 4 °C) and resuspended in M9-N buffer (375 µl) by gentle vortexing. The 96-well plates were then transferred to an anaerobic chamber. To deep-well plates of cell suspensions were added the *N*-phenyl-morpholine substrate (5 µl per well, 400 mM in EtOH) and diazoacetonitrile (35 µl per well, 550 mM in EtOH). During the addition, the stock solution of diazoacetonitrile was kept on an ice bath to minimize evaporation. The plates were sealed with aluminium sealing foil immediately after the addition and shaken in the anaerobic chamber at room temperature and 600 r.p.m. After 18–24 h, the seals were removed and 610 µl of 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal standard was added. The plates were tightly sealed with silicone mats, vigorously vortexed and centrifuged (5,000 g, 5 min) to separate the organic and aqueous layers. The organic layer of each well was then transferred to individual

vials equipped with autosampler vial inserts and analysed by GC–FID or GC–mass spectrometry (GC–MS).

Protein expression of P411 variants

E. coli cells carrying plasmid encoding the appropriate P411 variant were grown overnight in 5 ml of LB culture with ampicillin. The preculture (1.5 ml) was used to inoculate 28.5 ml of Hyper Broth media with ampicillin (5% v/v inoculation) in a 125-ml Erlenmeyer flask; this culture was incubated at 37 °C, 230 r.p.m. for 2.5 h. The culture was then cooled on ice for 30 min and induced with 0.5 mM IPTG and 1.0 mM 5-aminolevulinic acid (final concentrations). Protein expression was conducted at 20 °C, 150 r.p.m. for 16–18 h. Subsequently, the *E. coli* cells were pelleted by centrifugation (3,000 g, 5 min, 4 °C). Media was removed, and the resulting cell pellet was resuspended in M9-N buffer to an optical density (OD_{600}) of 30, unless otherwise specified.

Biotransformations using whole *E. coli* cells

Suspension of *E. coli* cells expressing the appropriate haem protein variant in M9-N or potassium phosphate (KPi) buffer (typically OD_{600} = 30) were transferred to a reaction vial. The headspace of the reaction vial was purged with a stream of argon for at least 15 min. Enzymatic reactions were then set up in an anaerobic chamber. To a 2-ml vial were added degassed suspension of *E. coli* expressing P411 variant (typically OD_{600} = 30, 360 µl), N-substituted arene substrate (typically 2.5 or 5 µl of 400 mM stock solution in EtOH) and the diazoacetonitrile solution (typically 37.5 or 35 µl of 550 mM stock solution in EtOH). The concentration of the solution was measured by ^{1}H-NMR and adjusted to 550 mM. The diazoacetonitrile solution was kept in an ice bath and added last. The final volume of the biotransformation was 400 µl, with 10% vol EtOH. The reaction vials were then capped and shaken in the anaerobic chamber at room temperature and 600 r.p.m. for 18–20 h. After the completion of the reaction, 610 µl of 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal standard was added to the vial. The resulting mixture was transferred to a 1.5-ml microcentrifuge tube, vigorously vortexed and centrifuged (20,000 g, 10 min) to separate the organic and aqueous layers. The organic layer was transferred to a vial equipped with an autosampler vial insert and analysed by GC–FID or GC–MS.

Enzymatic preparative syntheses

The *E. coli* cell suspension harbouring either P411-PFA or P411-ACHF was prepared as described above. The cell suspension was placed in a sealed flask on ice, and its headspace was purged with a stream of argon for at least 30 min. The reaction flask was then transferred to an anaerobic chamber. To every 100-ml cell suspension (OD_{600} = 30), N-arylamine substrate (1.25 ml from 400 mM EtOH stock solution, final concentration of 5 mM, 1 mmol, 1.0 equiv.) and diazoacetonitrile (7.5 ml of 550 mM EtOH stock solution, 8.25 mmol, 8.25 equiv.) were added. The reaction vial was immediately capped and sealed with parafilm, removed from the anaerobic chamber and shaken at room temperature at 200 r.p.m. for 18 h. The reaction solution was then extracted with 100 ml of 1:1 hexane/ethylacetate three times. The combined organic layer was then washed with brine and dried over anhydrous $MgSO_4$. The solvent was removed in vacuo, and the product was purified by flash chromatography.

Author contributions

J.Z. designed the overall research with F.H.A. providing guidance. J.Z. designed and conducted the initial screening of haem proteins. J.Z. and N.M.A. performed the directed evolution experiments. J.Z., E.A. and R.M. designed and performed the substrate scope studies and analysis. A.O.M. obtained and analysed the X-ray crystal structure of the engineered proteins with N.J.P. providing help. J.Z. and F.H.A. wrote the manuscript with input from all authors.

6.5 Supplemental Information

Crystallography of P411-PFA haem domain

The construct employed for crystallization is the haem domain of P411-PFA (residue 1– 463 of the holoprotein, followed by an *Xho*I site and a C-terminal 6×His-tag). The amino acid sequence is:

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRY
LSSQRLIKEACDESRFDKTLSQGLKFLRDFLGDGLATSWTHEKNWKKAHNI
LLPSFSQQAMKGYHASMVDIAVQLVQKWERLNADEHIEVSEDMTRLTLDTI
GLCGFNYRFNSFYRDQPHPFIISMVRALDEVMNKLQRANPDDPAYDENKRQ
FQEDIKVMNDLVDKIIADRKATGEQSDDLLTQMLNGKDPETGEPLDDGNIR
YQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEEAARVLVDPVPSYKQV
KQLKYVGMVLNEALRLWPVVPVFSLYAKEDTVLGGEYPLEKGDEVMVLIPQ
LHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASPGQQFALHEAT
LVLGMMLKHFDFEDHTNYELDIKELQTLKPKGFVVKAKSKKIPLGGIPSPST
LEHHHHHH

Protein expression of the P411-PFA haem domain follows the protocol described in the Methods section in the main text (3 L Hyper Broth expression culture). The cells were pelleted, frozen, and stored at -20 degrees before use. For protein purification, the frozen pellets were resuspended in His-trap buffer A (25 mM Tris-HCl, 100 mM NaCl, 25 mM imidazole, pH = 7.5 at room temperature, 4 mL/g of cell wet weight), added with DNaseI, and lysed by sonication. Cell debris was removed by centrifugation (20,000 x g, 20 min, at 4 °C). The lysate was further clarified by filtration.

For protein purification, the clarified lysate harboring P411-PFA haem domain with C- terminal 6×His-tag was first subjected to a Ni-NTA column (5 mL HisTrap HP, GE Healthcare) on an AKTA purifier FPLC system (GE Healthcare). The enzyme was eluted with a linear gradient from 100% HisTrap buffer A to 100% HisTrap buffer B (25 mM Tris-HCl, 100 mM NaCl, 300 mM imidazole, pH = 7.5 at room temperature) over 10 column volumes.

The enzyme was further purified by anion exchange chromatography. Fractions contain the haem protein (absorption at 280 nm and 417 nm was monitored) were pooled and buffer exchanged (3 times) to anion exchange buffer A (25 mM Tris-HCl, 25 mM NaCl, pH = 7.5) using centrifugal spin filters (10 kDa molecular weight cut-off, Amicon Ultra, Merck Millipore). Subsequently, the protein was purified with an anion exchange Q Sepharose column (5 mL HiTrap Q HP, GE Healthcare).

Enzymes are eluted with a linear gradient by increasing NaCl concentration (25 mM to 300 mM NaCl) over 10 column volumes. Fractions containing the desired protein were pooled, buffer exchanged into storage buffer (25 mM Tris-HCl, 25 mM NaCl, pH 7.5 at room temperature), and concentrated using centrifugal spin filters. Protein concentration was determined by BCA assay using a bovine serum albumin standard curve. Purified P411-PFA haem domain protein was aliquoted, flash-frozen on powered dry ice, and stored at -80 degrees C before use.

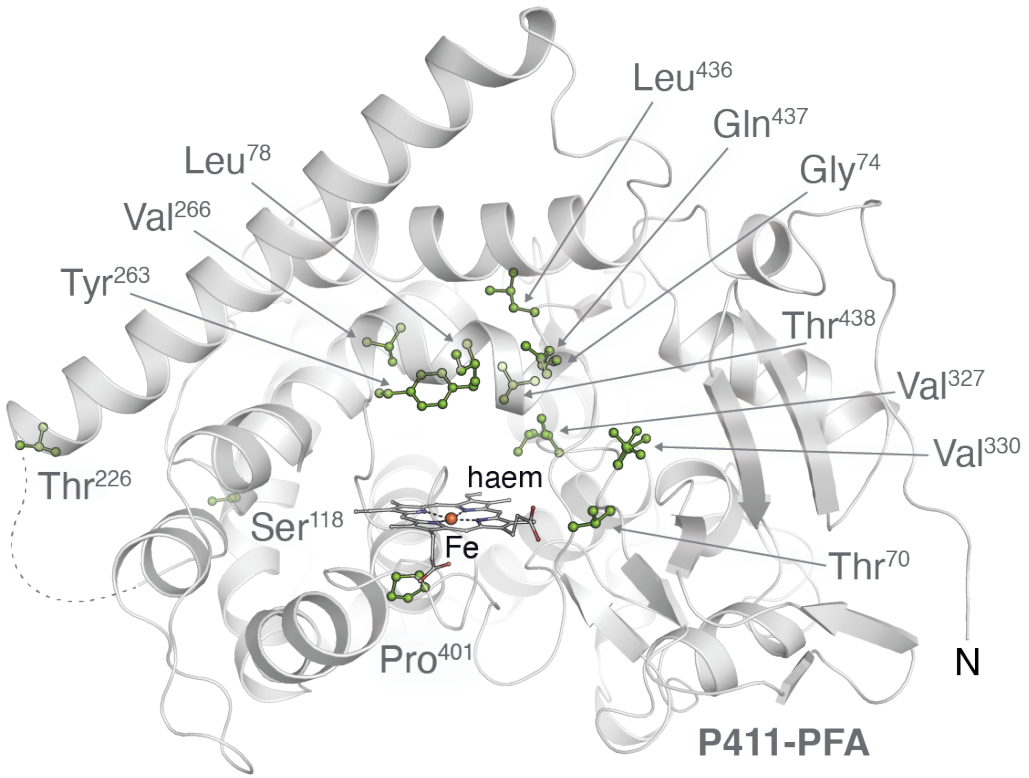
All crystallization experiments were performed in an anoxic chamber (COY Laboratory Products) with a 97% Ar and 3% H2 atmosphere. Purified P411-PFA concentrated to 10 mg/mL in stabilization buffer containing 250 mM Tris-HCl pH 7.5 and 25 mM NaCl was crystallized by sitting drop vapor diffusion method. Initial crystallization conditions were identified from the commercial screening kit, NeXtal JCSG+ (Qiagen). Crystals were generated by addition of 1 µL of protein sample to 1 µL of a precipitant solution containing 0.2 mM ammonium sulfate, 0.1 M Bis-Tris pH 5.5, and 20% (w/v) PEG 3350 and subsequent streak seeding. Brown, rod-shaped crystals appeared after 4 days. Harvested crystals were soaked in cryoprotectant containing the well solution supplemented with 20% (v/v) ethylene glycol prior to flash freezing by direct plunge into liquid N_2.

X-ray diffraction data sets were collected at 100K at the Stanford Synchrotron Radiation Lightsource (SSRL) at beamline 12-2 equipped with a Dectris Pilatus 6M detector. Data were indexed, integrated, and scaled using HKL3000.[65] Two complete crystallographic data sets from two parts of the same crystal were merged and scaled together for the final analysis. Phases were obtained by molecular replacement (MR) with PHASER [66] implemented within the CCP4 crystallographic suite.[67] Proteinic atoms from a previously solved cytochrome P411 structure (PDB accession code 5UCW) was used as the MR input model. Model building was performed in Coot,[68] and both REFMAC5 [69] and Phenix [70] were used for refinement. The MolProbity [71] server was used to identify Ramachandran outliers and analyze geometric parameters. Software was installed and configured using the SBGrid package manager.[72] All crystallographic figures were prepared using the PyMOL molecular graphics software package (Schrödinger, LLC). Cavity analysis was performed with Hollow and electrostatics calculations were performed using the APBS PyMOL plugin.[73, 74] The P411-PFA coordinates and electron density maps were deposited in the Protein Data Bank with the accession code 8DSG.

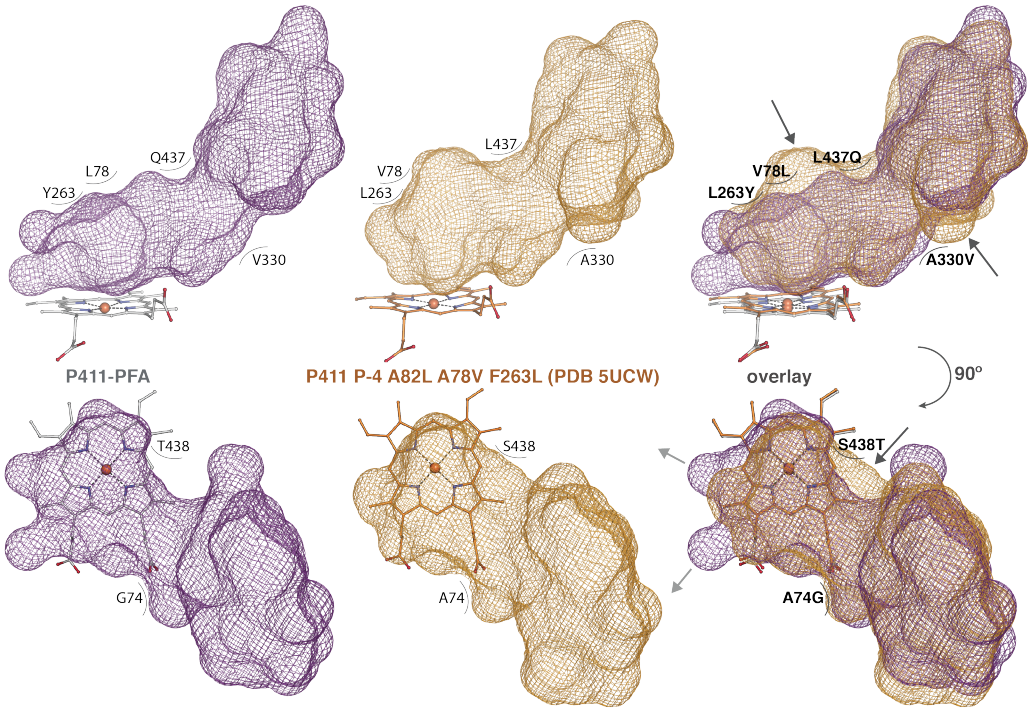
The haem domain of P411-ACHF was purified using the same procedure. However, structural determination attempts of the haem domain of P411-ACHF were unsuccessful under similar purification and crystallization conditions.

Protein Structure prediction using AlphaFold2

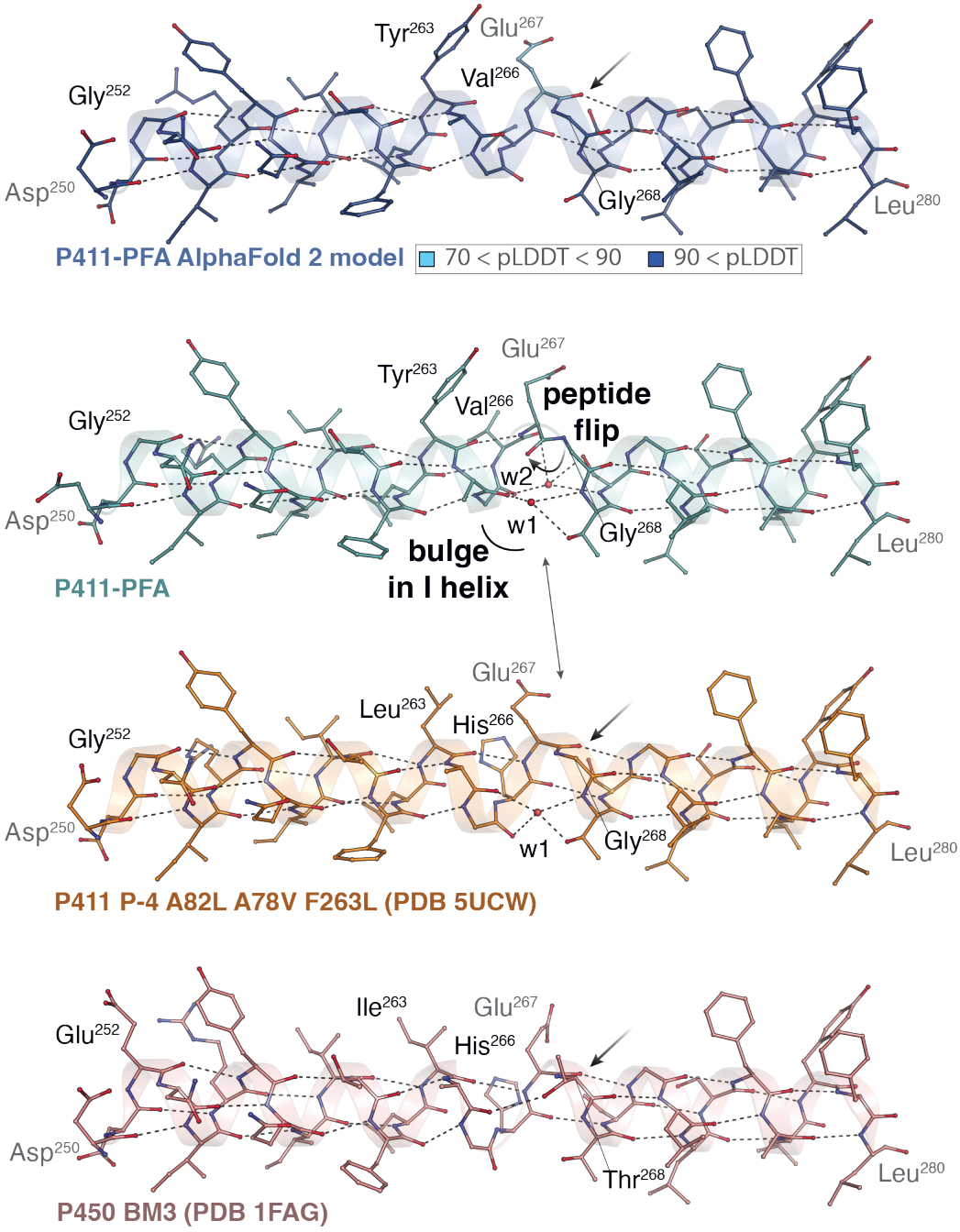
The AlphaFold2 prediction of the haem domain of P411-PFA was performed using the ColabFold web tool.[75] The I helix of the predicted model was colored by atom pLDDT scores using the alphafold_coloring.py PyMOL extension, available on GitHub.



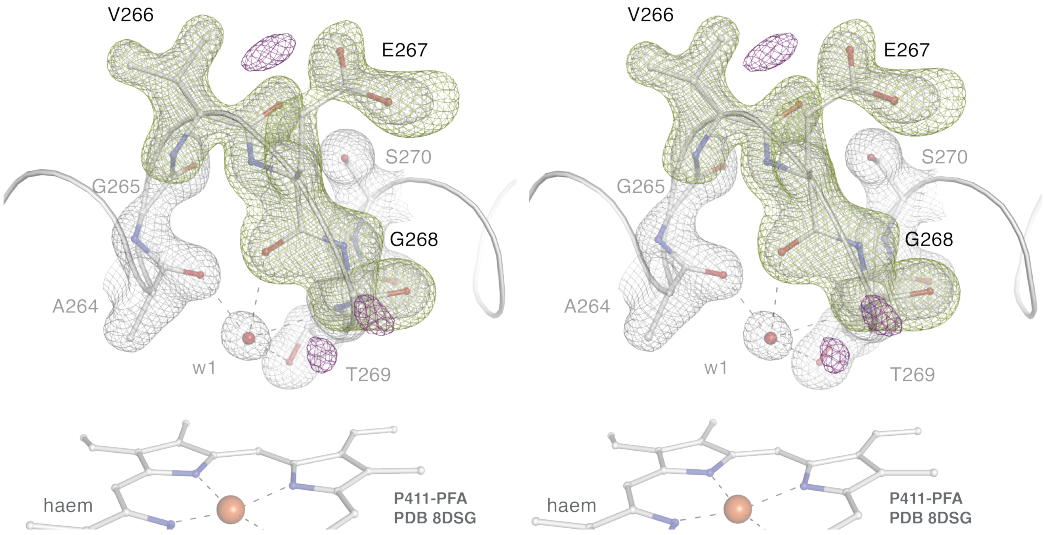
Supplementary Figure 6.1: The 13 mutations in P411-PFA (PDB 8DSG) relative to the closely related C–H aminase, P-4 A82L A78V F263L (PDB 5UCW).[55] P411-PFA is shown in cartoon representation and mutated residues are highlighted as green sticks. V330 is modeled in two conformations. The haem cofactor is represented in sticks with carbon (white), oxygen (red), nitrogen (blue), and Fe (orange) coloring.



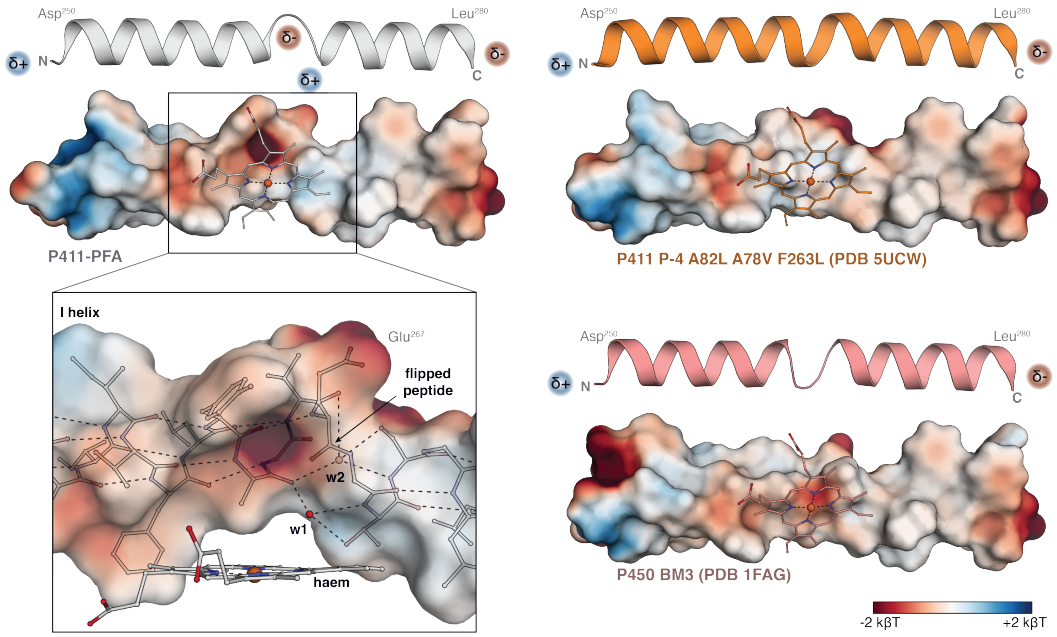
Supplementary Figure 6.2: Side-on (above) and top-down (below) view of the active site cavity comparison between P411-PFA (PDB 8DSG) and P-4 A82L A78V F263L (PDB 5UCW).[55] Residues which significantly shape the active site are noted near their cavity contact site. Cavity surface representations were calculated using Hollow[73] and depicted in PyMOL.



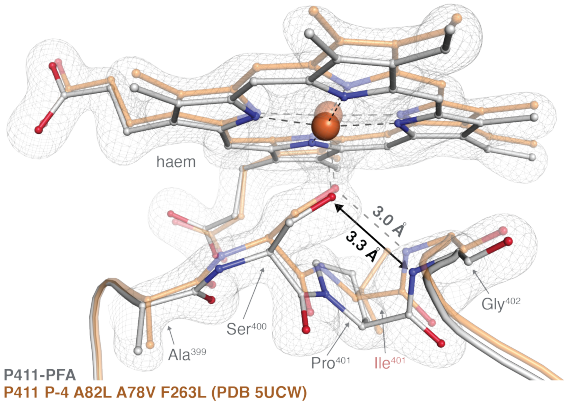
Supplementary Figure 6.3: Comparison of I helices of (starting from the bottom): wild-type P450 BM3 (PDB 1FAG, red);[56] the P411 C-H aminase, P-4 A82L A78V F263L (PDB 5UCW, orange),[55] which is closely related to P411-PFA; P411-PFA (PDB 8DSG, teal); and AlphaFold 2 predicted model of P411-PFA (blue, colored by pLDDT score). PyMOL script to color AlphaFold2 predicted models by pLDDT scores is available at: https://github.com/ailienamaggiolo/alphafold_coloring. The hydrogen bonding of the I helix is disrupted between residue positions 263 and 268. This disruption is stabilized by an intercalated water molecule (w1), modeled in PDB 5UCW and likely present in PDB 1FAG, although not modeled. The peptide flip of E267 in P411-PFA is stabilized by a second water molecule (w2). Although the template-based model prediction of P411-PFA using AlphaFold 2 captured the disrupted hydrogen bonding of the I helix, AlphaFold 2 failed to predict the E267 peptide flip.



Supplementary Figure 6.4: Stereo image of the simulated annealing omit map of a portion of the I helix calculated after removal of residues 266-268. The F_o-F_c (green mesh (positive, 3.5 σ, carve 2.0), purple mesh (negative, 3.5 σ, carve 2.0)) and $2F_o$-F_c (gray mesh, 1.5 σ, carve 1.5) densities show the peptide of E267 has flipped to the opposite orientation relative to the neighboring carbonyl in the helix, disrupting the normal hydrogen bonding pattern of the α-helix.



Supplementary Figure 6.5: Local electrostatics of the I helix shown as cartoon and surface representations. Top left: P411-PFA (PDB 8DSG, white); top right: the P411 C–H aminase, P-4 A82L A78V F263L (PDB 5UCW, orange),[55] which is closely related to P411-PFA; and bottom right: wild-type P450 BM3 (PDB 1FAG, red).[56] Bottom left: zoomed-in view of the electrostatics directly above the haem reveals a negatively charged patch resulting from the peptide flip of residue E267. Surface electrostatics were calculated using the APBS PyMOL plugin.[74]



Supplementary Figure 6.6: Structural difference between P411-PFA (PDB 8DSG, white) and P-4 A82L A78V F263L (PDB 5UCW, orange)[55] at the "Cys pocket" region. Residues in this region are highly conserved and affect the electronic properties of haem. Mutation L401P in P411-PFA has disrupted the hydrogen bond between the backbone amide of G402 to the axial S400 (an effective hydrogen bond should be < 3.3 Å). This structural change has led to the enzyme activity and selectivity differences between P411-PFA (P401) and FA-E3 (L401). We postulated that other residues in this region can be targeted to further improve arene C–H functionalization selectivity and identified F393W as a beneficial mutation.

Table 6.1: X-ray crystallographic data collection and refinement statistics.

	P411-PFA PDB **8DSG**
Data collection	
Space group	$C\,2\,2\,2_1$
Wavelength (Å)	0.97946
Cell dimensions	
a, *b*, *c* (Å)	105.25, 172.46, 226.62
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)*	50.0-1.87 (1.90-1.87)
R_{merge}	0.090 (0.399)
R_{pim}	0.056 (0.282)
$I / \sigma I$	21.6 (2.3)
$CC_{1/2}$	0.989 (0.805)
Completeness (%)	91.9 (46.2)
Redundancy	2.4 (1.6)
Refinement	
Resolution (Å)	38.60-1.87
No. reflections	147073
R_{work} / R_{free}**	0.187/0.218
No. atoms	15955
Protein	14704
Ligand/ion	266
Water/solvent	985
B-factors	
Protein	22.17
Ligand/ion	24.32
Water/solvent	29.72
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	1.216
Molprobity clash score	0.91 (100^{th} percentile)
Rotamer outliers (%)	0.00
Ramachandran favored (%)	97.52

*Values in parentheses are for highest-resolution shell.

References

(1) Davies, H. M. L.; Manning, J. R. Catalytic C–H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **2008**, *451*, DOI: 10.1038/nature06485.

(2) Davies, H. M. L.; Bois, J. D.; Yu, J. .-. Q. C–H functionalization in organic synthesis. *Chem. Soc. Rev.* **2011**, *40*, DOI: 10.1039/c1cs90010b.

(3) Hartwig, J. F.; Larsen, M. A. Undirected, homogeneous C–H bond functionalization: challenges and opportunities. *ACS Cent. Sci.* **2016**, *2*, DOI: 10.1021/acscentsci.6b00032.

(4) Yamaguchi, J.; Yamaguchi, A. D.; Itami, K. C–H bond functionalization: emerging synthetic tools for natural products and pharmaceuticals. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, DOI: 10.1002/anie.201201666.

(5) Dalton, T.; Faber, T.; Glorius, F. C–H activation: toward sustainability and applications. *ACS Cent. Sci.* **2021**, *7*, DOI: 10.1021/acscentsci.0c01413.

(6) Börgel, J.; Ritter, T. Late-stage functionalization. *Chem* **2020**, *6*, DOI: 10.1016/j.chempr.2020.07.007.

(7) Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **2016**, *45*, DOI: 10.1039/C5CS00628G.

(8) Guillemard, L.; Kaplaneris, N.; Ackermann, L.; Johansson, M. J. Late-stage C–H functionalization offers new opportunities in drug discovery. *Nat. Rev. Chem.* **2021**, *5*, DOI: 10.1038/s41570-021-00300-6.

(9) Fazekas, T. J. Diversification of aliphatic C–H bonds in small molecules and polyolefins through radical chain transfer. *Science* **2022**, *375*, DOI: 10.1126/science.abh4308.

(10) Neufeldt, S. R.; Sanford, M. S. Controlling site selectivity in palladium-catalyzed C–H bond functionalization. *Acc. Chem. Res.* **2012**, *45*, DOI: 10.1021/ar300014f.

(11) Dai, H. .-. X.; Stepan, A. F.; Plummer, M. S.; Zhang, Y. .-. H.; Yu, J. .-. Q. Divergent C–H functionalizations directed by sulfonamide pharmacophores: late-stage diversification as a tool for drug discovery. *J. Am. Chem. Soc.* **2011**, *133*, DOI: 10.1021/ja201708f.

(12) Sambiagio, C. A comprehensive overview of directing groups applied in metal-catalysed C–H functionalisation chemistry. *Chem. Soc. Rev.* **2018**, *47*, DOI: 10.1039/C8CS00201K.

(13) Yu, Z.; Li, Y.; Zhang, P.; Liu, L.; Zhang, J. Ligand and counteranion enabled regiodivergent C–H bond functionalization of naphthols with α-aryl-α-diazoesters. *Chem. Sci.* **2019**, *10*, DOI: 10.1039/C9SC01657K.

(14) Greule, A.; Stok, J. E.; Voss, J. J. D.; Cryle, M. J. Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat. Prod. Rep.* **2018**, *35*, DOI: 10.1039/C7NP00063D.

(15) Kille, S.; Zilly, F. E.; Acevedo, J. P.; Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **2011**, *3*, DOI: 10.1038/nchem.1113.

(16) Zhang, K.; Shafer, B. M.; Demars, M. D.; Stern, H. A.; Fasan, R. Controlled oxidation of remote sp3 C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **2012**, *134*, DOI: 10.1021/ja3073462.

(17) Lukowski, A. L. C–H hydroxylation in paralytic shellfish toxin biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, DOI: 10.1021/jacs.8b08901.

(18) Zhang, X. Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science* **2020**, *369*, DOI: 10.1126/science.abb8271.

(19) Andorfer, M. C.; Park, H. J.; Vergara-Coll, J.; Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci.* **2016**, *7*, DOI: 10.1039/C5SC04680G.

(20) Neugebauer, M. E. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, *15*, DOI: 10.1038/s41589-019-0355-x.

(21) Hayashi, T. Evolved aliphatic halogenases enable regiocomplementary C-H functionalization of a pharmaceutically relevant compound. *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, DOI: 10.1002/anie.201907245.

(22) Chen, K.; Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **2020**, *3*, DOI: 10.1038/s41929-019-0385-5.

(23) Yang, Y.; Arnold, F. H. Navigating the unnatural reaction space: directed evolution of heme proteins for selective carbene and nitrene transfer. *Acc. Chem. Res.* **2021**, *54*, DOI: 10.1021/acs.accounts.0c00591.

(24) Zhang, R. K.; Huang, X.; Arnold, F. H. Selective C–H bond functionalization with engineered heme proteins: new tools to generate complexity. *Curr. Opin. Chem. Biol.* **2019**, *49*, DOI: 10.1016/j.cbpa.2018.10.004.

(25) Brandenberg, O. F.; Fasan, R.; Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr. Opin. Biotechnol.* **2017**, *47*, DOI: 10.1016/j.copbio.2017.06.005.

(26) Natoli, S. N.; Hartwig, J. F. Noble-metal substitution in hemoproteins: an emerging strategy for abiological catalysis. *Acc. Chem. Res.* **2019**, *52*, DOI: 10.1021/acs.accounts.8b00586.

(27) Matthews, M. L. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **2014**, *10*, DOI: 10.1038/nchembio.1438.

(28) Rui, J. Directed evolution of nonheme iron enzymes to access abiological radical-relay C(sp3)-H azidation. *Science* **2022**, *376*, DOI: 10.1126/science.abj2830.

(29) Doyle, M. P.; Duffy, R.; Ratnikov, M.; Zhou, L. Catalytic carbene insertion into C-H bonds. *Chem. Rev.* **2010**, *110*, DOI: 10.1021/cr900239n.

(30) Zhang, R. K. Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp3 C–H functionalization. *Nature* **2019**, *565*, DOI: 10.1038/s41586-018-0808-5.

(31) Zhang, J.; Huang, X.; Zhang, R. K.; Arnold, F. H. Enantiodivergent α-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.* **2019**, *141*, DOI: 10.1021/jacs.9b04344.

(32) Zhou, A. Z.; Chen, K.; Arnold, F. H. Enzymatic lactone-carbene C–H insertion to build contiguous chiral centers. *ACS Catal.* **2020**, *10*, DOI: 10.1021/acscatal.0c01349.

(33) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **2016**, *534*, DOI: 10.1038/nature17968.

(34) Dydio, P. An artificial metalloenzyme with the kinetics of native enzymes. *Science* **2016**, *354*, DOI: 10.1126/science.aah4427.

(35) Gu, Y.; Natoli, S. N.; Liu, Z.; Clark, D. S.; Hartwig, J. F. Site-Selective Functionalization of (sp^3)C-H Bonds Catalyzed by Artificial Metalloenzymes Containing an Iridium-Porphyrin Cofactor. *Angewandte Chemie International Edition* **2019**, *58*, 13954–13960.

(36) Caballero, A. Catalytic functionalization of low reactive C(sp3)–H and C(sp2)–H bonds of alkanes and arenes by carbene transfer from diazo compounds. *Dalton Trans.* **2015**, *44*, DOI: 10.1039/C5DT03450G.

(37) Liu, Z.; Wang, J. Cross-coupling reactions involving metal carbene: from C-C/C–C bond formation to C–H bond functionalization. *J. Org. Chem.* **2013**, *78*, DOI: 10.1021/jo401850q.

(38) Fructos, M. R.; Belderrain, T. R.; De Frémont, P.; Scott, N. M.; Nolan, S. P.; Díaz-Requejo, M. M.; Pérez, P. J. A Gold Catalyst for Carbene-Transfer Reactions from Ethyl Diazoacetate. *Angewandte Chemie International Edition* **2005**, *44*, 5284–5288.

(39) Fructos, M. R. Mechanistic studies on gold-catalyzed direct arene C–H bond functionalization by carbene insertion: the coinage-metal effect. *Organometallics* **2017**, *36*, DOI: 10.1021/acs.organomet.6b00604.

(40) Conde, A.; Sabenya, G.; Rodríguez, M.; Postils, V.; Luis, J. M.; Díaz-Requejo, M. M.; Costas, M.; Pérez, P. J. Iron and Manganese Catalysts for the Selective Functionalization of Arene C(sp ^{2})-H Bonds by Carbene Insertion. *Angewandte Chemie International Edition* **2016**, *55*, 6530–6534.

(41) Postils, V. Mechanism of the selective Fe-catalyzed arene carbon–hydrogen bond functionalization. *ACS Catal.* **2018**, *8*, DOI: 10.1021/acscatal.7b03935.

(42) Xi, Y.; Su, Y.; Yu, Z.; Dong, B.; McClain, E. J.; Lan, Y.; Shi, X. Chemoselective Carbophilic Addition of α-Diazoesters through Ligand-Controlled Gold Catalysis. *Angewandte Chemie International Edition* **2014**, *53*, 9817–9821.

(43) Yu, Z. Highly site-selective direct C–H bond functionalization of phenols with α-aryl-α-diazoacetates and diazooxindoles via gold catalysis. *J. Am. Chem. Soc.* **2014**, *136*, DOI: 10.1021/ja503163k.

(44) Xu, B.; Li, M. .-. L.; Zuo, X. .-. D.; Zhu, S. .-. F.; Zhou, Q. .-. L. Catalytic asymmetric arylation of α-aryl-α-diazoacetates with aniline derivatives. *J. Am. Chem. Soc.* **2015**, *137*, DOI: 10.1021/jacs.5b05086.

(45) Ma, B.; Chu, Z.; Huang, B.; Liu, Z.; Liu, L.; Zhang, J. Highly *para* -Selective C-H Alkylation of Benzene Derivatives with 2,2,2-Trifluoroethyl α-Aryl-α-Diazoesters. *Angewandte Chemie International Edition* **2017**, *56*, 2749–2753.

(46) Holmberg-Douglas, N.; Onuska, N. P. R.; Nicewicz, D. A. Regioselective Arene C-H Alkylation Enabled by Organic Photoredox Catalysis. *Angewandte Chemie International Edition* **2020**, *59*, 7425–7429.

(47) Pizarro, J. D.; Schmidtke, I. L.; Nova, A.; Fructos, M. R.; Pérez, P. J. Selective functionalization of arene C(sp2)–H bonds by gold catalysis: the role of carbene substituents. *ACS Catal.* **2022**, *12*, DOI: 10.1021/acscatal.2c01713.

(48) Zhang, W. Enantioselective cyanation of benzylic C–H bonds via copper-catalyzed radical relay. *Science* **2016**, *353*, DOI: 10.1126/science.aaf7783.

(49) Lennox, A. J. J. Electrochemical aminoxyl-mediated α-cyanation of secondary piperidines for pharmaceutical building block diversification. *J. Am. Chem. Soc.* **2018**, *140*, DOI: 10.1021/jacs.8b08145.

(50) Mykhailiuk, P. K.; Koenigs, R. M. Diazoacetonitrile (N $_2$ CHCN): A Long Forgotten but Valuable Reagent for Organic Synthesis. *Chemistry – A European Journal* **2020**, *26*, 89–101.

(51) Chandgude, A. L.; Fasan, R. Highly Diastereo- and Enantioselective Synthesis of Nitrile-Substituted Cyclopropanes by Myoglobin-Mediated Carbene Transfer Catalysis. *Angewandte Chemie International Edition* **2018**, *57*, 15852–15856.

(52) Hock, K. J.; Knorrscheidt, A.; Hommelsheim, R.; Ho, J.; Weissenborn, M. J.; Koenigs, R. M. Tryptamine Synthesis by Iron Porphyrin Catalyzed C-H Functionalization of Indoles with Diazoacetonitrile. *Angewandte Chemie International Edition* **2019**, *58*, 3630–3634.

(53) Coelho, P. S. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **2013**, *9*, DOI: 10.1038/nchembio.1278.

(54) Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. Enzyme-controlled nitrogen-atom transfer enables regiodivergent C–H amination. *J. Am. Chem. Soc.* **2014**, *136*, DOI: 10.1021/ja509308v.

(55) Prier, C. K.; Zhang, R. K.; Buller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. *Nature Chemistry* **2017**, *9*, 629–634.

(56) Li, H.; Poulos, T. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Mol. Biol.* **1997**, *4*, DOI: 10.1038/nsb0297-140.

(57) Cooley, R. B.; Arp, D. J.; Karplus, P. A. Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. *J. Mol. Biol.* **2010**, *404*, DOI: 10.1016/j.jmb.2010.09.034.

(58) Yeom, H.; Sligar, S. G.; Li, H.; Poulos, T. L.; Fulco, A. J. The role of Thr268 in oxygen activation of cytochrome P450BM-3. *Biochemistry* **1995**, *34*, DOI: 10.1021/bi00045a014.

(59) Ost, T. W. B. Structural and spectroscopic analysis of the F393H mutant of flavocytochrome P450 BM3. *Biochemistry* **2001**, *40*, DOI: 10.1021/bi010717e.

(60) Krest, C. M. Significantly shorter Fe–S bond in cytochrome P450-I is consistent with greater reactivity relative to chloroperoxidase. *Nat. Chem.* **2015**, *7*, DOI: 10.1038/nchem.2306.

(61) Ford, A. Modern organic synthesis with α-diazocarbonyl compounds. *Chem. Rev.* **2015**, *115*, DOI: 10.1021/acs.chemrev.5b00121.

(62) Reisman, S. E.; Nani, R. R.; Levin, S. Buchner and beyond: arene cyclopropanation as applied to natural product total synthesis. *Synlett* **2011**, *2011*, DOI: 10.1055/s-0031-1289520.

(63) Fleming, G. S.; Beeler, A. B. Regioselective and enantioselective intermolecular Buchner ring expansion in flow. *Org. Lett.* **2017**, *2017*, DOI: 10.1021/acs.orglett.7b02537.

(64) Gibson, D. G. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, DOI: 10.1038/nmeth.1318.

(65) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. *HKL* -3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography* **2006**, *62*, 859–866.

(66) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Phaser* crystallographic software. *Journal of Applied Crystallography* **2007**, *40*, 658–674.

(67) Winn, M. D. et al. Overview of the *CCP* 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **2011**, *67*, 235–242.

(68) Emsley, P.; Cowtan, K. *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **2004**, *60*, 2126–2132.

(69) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **1997**, *53*, 240–255.

(70) Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology* **2019**, *75*, 861–877.

(71) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **2010**, *66*, 12–21.

(72) Morin, A.; Eisenbraun, B.; Key, J.; Sanschagrin, P. C.; Timony, M. A.; Ottaviano, M.; Sliz, P. Collaboration gets the most out of software. *eLife* **2013**, *2*, e01456.

(73) Ho, B. K.; Gruswitz, F. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Structural Biology* **2008**, *8*, 49.

(74) Holst, M.; Saied, F. Multigrid solution of the Poisson?Boltzmann equation. *Journal of Computational Chemistry* **1993**, *14*, 105–113.

(75) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nature Methods* **2022**, *19*, 679–682.