The Many Roles of the Nitrogenase Iron Protein

Thesis by

Belinda B. Wenke

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2019

Defended January 18, 2019

Belinda B. Wenke
ORCID: 0000-0003-3214-6197

ACKNOWLEDGMENTS

The following work would not have been possible without the tutelage, support and collaboration of many people. I am greatly appreciative of my PhD advisor, Prof. Douglas Rees, for facilitating a productive and positive environment for a collection of outstanding scientists and students. He maintains a relentlessly curious and thoughtful approach to science that has enabled (in research) bold experiments in new fields, and (in teaching) careful examination and appreciation for the decades of work that has contributed to our collective scientific knowledge. I could not be more grateful. I am also appreciative of my committee: Profs. Pamela Bjorkman, Harry Gray and Jared Leadbetter, for providing seemingly endless expertise in highly disparate fields and support in and out of the lab. Thank you to Phoebe Ray for keeping the group in good order and the lights on (not a trivial task), and to Dr. Jens Kaiser and Welison Floriano for computing expertise and patience.

I could not be more thankful for the mentors I have had in Team Nitrogenase, whose exacting methods and approach to science and life are a constant source of happiness. Thank you to Dr. Thomas Spatzal and Dr. Kathryn Perez for building the scientific foundation and standards in the lab as well as past, present and future friendship, advice and adventure at home and abroad. Thank you to Dr. Renee Arias and Dr. Helen Segal for fostering an environment of thoughtful research, compassion, humor and spice. Thank you to the cast and crew of Team Nitrogenase and the greater Rees lab, all who have contributed in a myriad of ways, large, small and ongoing: Dr. Limei Zhang, Dr. Christine Morrison, Dr. Camille McAvoy, Dr. Trixia Buscagan and proto-Dr. Ailiena Maggiolo. The cryo-EM experiments also owe a debt of gratitude to many members of Prof. Grant Jensen's lab, foremost Sara Weaver and Dr. Songye Chen, among many others, for guidance, technical help and many hours in sundry basements. Thank you to my BMB classmates, especially Emily Blythe, Ferdinand Huber and Chengcheng Fan for perspective, experimental and moral support. Your talent is incomparable.

My endless thanks and appreciation go to my parents: Julia Chuang and Dr. Gottfried Wenke, who were the first scientists I knew, and truly embody the unattainable immigrant spirit of perseverance, preparation, passion, stubbornness, generational wisdom, love, and family. Thank you to my little brother Michael; your kindness and intelligence are more than enough. Lastly, thank you to my husband (Prof.!) Andrew Buller for, among many contributions, patience, scientific curiosity, fluency in first principles, care, cats, and compassion. We are all lucky to know you, and you are my home.

ABSTRACT

Nitrogenase is the only known enzyme capable of reducing atmospheric nitrogen (N_2) into ammonia (NH_3) for incorporation into cellular material. N_2 reduction by nitrogenase is accomplished by sequential electron transfer between two component proteins: the substrate reductase (the MoFe-protein), and a specialized low-potential electron donor (the Fe-protein). The MoFe-protein contains the active site for nitrogen reduction, the FeMo-cofactor (FeMo-co). During nitrogen reduction, each Fe-protein dimer docks onto the MoFe-protein, transferring electrons to an intermediate cluster (P-cluster), and ultimately to the FeMo-co.

Strikingly, the Fe-protein has another critical role in nitrogen fixation. The Fe-protein is required for the biosynthesis of the two unique metalloclusters of the MoFe-protein: the P-cluster [8Fe:7S] and the active site FeMo-co ([Mo:7Fe:9S:C]-*R*-homocitrate) cluster. During FeMo-co-cluster maturation, the Fe-protein forms a complex with NifEN, a scaffolding protein homologous to the MoFe-protein, catalyzing the final step in the FeMo-co biosynthesis. Studies indicate that the Fe-protein catalyzes insertion of molybdenum and *R*-homocitrate into an all-iron FeMo-co precursor in a reductant and nucleotide-dependent manner. The remaining questions about the cellular functions of the Fe-protein include how the Fe-protein interacts with other maturation proteins in distinct (or similar) ways compared to the MoFe-protein, and how the Fe-protein contributes to the activation and insertion of molybdenum into the FeMo-co.

PUBLISHED CONTENT AND CONTRIBUTIONS

Wenke B.B.*, Spatzal T*, Rees D.C. (2019) Site-specific oxidation state assignments of the irons in the [4Fe:4S]$^{2+/1+/0}$ states of the nitrogenase Fe-protein. *Angewandte Chemie.* doi: https://doi.org/10.1002/ange.201813966

B.B.W. participated in sample preparation, data collection and analysis, and the preparation and writing of the manuscript.

* These authors contributed equally.

Wenke B.B.*, Arias R.J.*, and Spatzal T. (2019) Crystallization of Nitrogenase Proteins. In *Metalloproteins*, pp. 155-165. Humana Press, New York, NY. doi: https://doi.org/10.1007/978-1-4939-8864-8_10

B.B.W. participated in the research and writing of the manuscript.

* These authors contributed equally.

Wenke B.B., Spatzal T. (2018) Looking at Nitrogenase: Insights from Modern Structural Approaches. In: Structure and Bonding. Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/430_2018_28

B.B.W. participated in the writing of the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 The Nitrogen Cycle

Life on earth is limited by the availability of nutrients. Nitrogen is an essential element, used in vital biological material such as proteins and nucleic acids[1]. Nitrogen compounds move through the biosphere in a cyclic pattern, known as the nitrogen cycle. The largest reserve of nitrogen is sequestered in the atmosphere. The atmosphere contains ~78% nitrogen gas, but the bioavailability is limited due to the highly stable N-N triple bond. Once the cleavage of the N-N triple bond is achieved, multiple pathways exist to distribute nitrogen compounds (Figure 1).

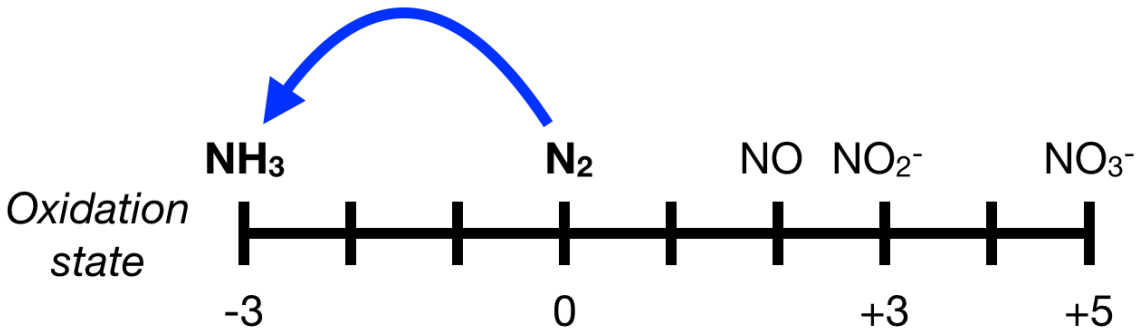


Figure 1. Select compounds found in the global nitrogen cycle, plotted against oxidation state. The process of nitrogen fixation from nitrogenase enzymes is indicated by an arrow.[2]

Nitrogen fixation (indicated in blue, Figure 1) is the sole biological process that converts nitrogen from the atmosphere into ammonia. Certain species of prokaryotes (e.g. soil bacteria, such as *Azotobacter vinelandii* and *Clostridium pasteurianum*) are able to perform nitrogen fixation (collectively called diazotrophs), although nitrogen is also fixed abiotically by lighting[3]. While the diazotrophs are a physiologically diverse group of

prokaryotes, a common feature is the enzyme nitrogenase, which is used as a genetic marker to predict nitrogen fixation competency. Ammonia is a key intermediate in the cycle, as it is the metabolite used for the production of biomass (e.g. amino and nucleic acids). Ammonia is also an important industrial commodity and is used in agricultural fertilizers. The remaining nitrogen-containing intermediates in the cycle are similarly crucial metabolites used for energy transduction by bacteria and archaea.

Nitrification describes the oxidation of ammonia or hydroxylamine to nitrite and nitrate. The conversion is catalyzed by two groups of chemolithoautotrophic microorganisms: the ammonia oxidizing microorganisms (bacteria and archaea) that generate nitrite from ammonia, and nitrite-oxidizing bacteria that produce nitrate from nitrite[3]. The ammonia-metabolizing bacteria prevent the toxic buildup of soluble ammonia in aquatic environments. Aerobic ammonia oxidization proceeds through a hydroxylamine intermediate, catalyzed by the enzymes ammonia monooxygenase and hydroxylamine oxidoreductase. Nitrification is an aerobic process that yields nitrate for the anaerobic processes of dissimilatory nitrate reduction, denitrification, and anaerobic ammonia oxidation.

Denitrification describes the conversion of nitrite (NO_2^-), from nitrate (NO_3^-), to nitric oxide (NO), which gets combined with nitrous oxide (N_2O) to form nitrogen gas (N_2). Overall, the process catabolizes nitrogen nutrients and returns nitrogen to its stable reservoir in the atmosphere. Denitrification is performed anaerobically by a variety of prokaryotes (e.g. *Pseudomonas aeruginosa*). The process is vital in removing nitrates from aquatic environments, where its buildup can disrupt environmental homeostasis.

Dissimilatory nitrate ammonification converts inorganic nitrate or nitrite to ammonia, and occurs anaerobically. Nitrate assimilation is the process by which nitrate is converted to ammonia directly to produce cellular material. A parallel process of ammonification involves the mineralization of organic nitrogen (from biological material) into ammonia through the decay of living tissue[4].

Anaerobic ammonia oxidation (ANAMMOX) is an alternative to the aerobic ammonia oxidation-nitrification pathway. ANAMMOX catalyzes the oxidation of ammonia using nitrite as the electron acceptor, to produce nitrogen gas (N_2) and water. The process occurs primarily in aquatic anoxic zones, and contributes to returning dinitrogen gas to the atmosphere. ANAMMOX is thought to generate 30-50% of the nitrogen gas originating from the oceans, with denitrification generating 70-50%[1].

1.2 Energetics of Nitrogen Fixation

Nitrogen is an absolute requirement for life on earth. Nitrogen compounds are found in at least seven oxidation states, but occur commonly in four states ranging from -3 (NH_3) to +5 (NO_3^-). The -3 oxidation state is most common in biological compounds, such as nucleic and amino acids[1]. The most abundant form of nitrogen is atmospheric dinitrogen, which makes up ~78% of earth's atmosphere[2]. The process of incorporating nitrogen into biological compounds is nitrification, which generates nitrate and nitrite from ammonia at great energetic cost. The (H-N) bond dissociation energy of ammonia is 385 kJ/mol at 298K. Nitrification is dependent on initial ammonia generation from dinitrogen (N-N), which has a bond dissociation energy of 941 kJ/mol at 298K. Nitrogen reduction to ammonia is thermodynamically favorable (Eq. 1), but kinetically unfavorable due to unstable intermediates and the activation energy is insurmountable without a catalyst[3].

$$N_2 + 3H_2 \rightarrow 2NH_3 \qquad \Delta G^{0\prime}_{aq}(298K) = -34 \text{ kJ/mol} \qquad \text{Eq. 1}$$

The only known enzyme to carry out biological nitrogen fixation is nitrogenase. Nitrogenases convert atmospheric dinitrogen to ammonia, producing 1 equivalent of hydrogen (H_2) with every 2 equivalents of ammonia produced from 1 equivalent of dinitrogen. Nitrogenase is a two-protein component enzyme, composed of the MoFe- and Fe-proteins. Nitrogenase catalysis requires cycles of binding and dissociation between the two proteins, using ATP hydrolysis to drive the binding-dependent electron transfer.

Current knowledge of the nitrogenase cycle assumes a stoichiometry of 2 molecules of ATP hydrolyzed per electron transferred to the substrate. Discrepancies in the stoichiometry of electron donation originate from the unusual range of redox couples used by the Fe-protein. The [4Fe:4S] cluster can adopt three oxidation states: 0, +1 and +2. The stoichiometry of 2:1 ATP:electron arises from the assumption that the Fe-protein is cycling between the +1/+2 overall states. However, a 2:2 ATP:electron can be achieved using the 0/+2 couple, and this possibility remains an active area of research. *In vivo*, the Fe-protein can be reduced by ferredoxin or flavodoxin. Ferredoxins only use the +1/+2 couple, but the flavin cofactor in flavodoxin has three reduction states (quinone, semiquinone, and hydroquinone), opening up the possibility for two-electron transfers[5–8]. For the purposes of this work, the 2:1 stoichiometry assumed by a one-electron transfer from ferredoxin is used and summarized in Eq. 2.[9]

$$N_2 + 8e^- + 8H^+ + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i \qquad \text{Eq. 2}$$

1.3 Nitrogenase

Nitrogenase is a two-component enzyme (Figure 1), consisting of the active-site containing protein and the reductase protein. Nitrogenases can use different transition metals in the active site cofactor: molybdenum, iron, or vanadium.[10] The best-characterized nitrogenase is the molybdenum-iron nitrogenase, whose components are named after the metal clusters they contain: the active-site protein is the molybdenum-iron (MoFe) protein, and the reductase is the iron (Fe) protein.

Nitrogen reduction by nitrogenase involves sequential electron transfer from a specialized electron donor, the Fe-protein, to the MoFe-protein, which contains the active site for nitrogen reduction: the FeMo-cofactor (FeMo-co). Structural studies of nitrogenase show that the MoFe-protein has two active sites per heterotetramer due to its intrinsic two-fold (C2) symmetry, allowing two Fe-protein dimers to bind (Figure 2).

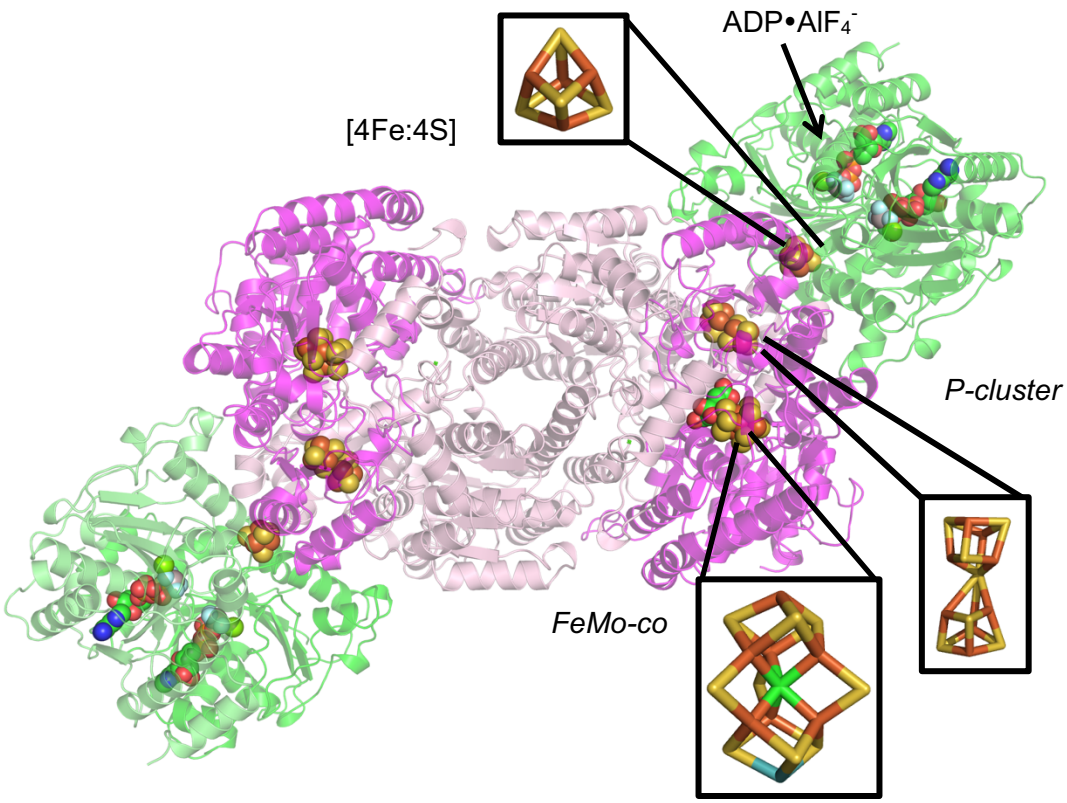


Figure 2. Crystal structure of the molybdenum nitrogenase from *Azotobacter vinelandii*, with an ATP transition state analog bound. The MoFe-protein is colored in magenta and light pink (α and β subunits, respectively). The Fe-protein is colored in green and lime. Cofactors are represented as sticks. PDBID 1M34.[11]

The MoFe-protein is an $\alpha_2\beta_2$ heterotetramer and has two biologically unique and highly elaborate metalloclusters: the FeMo-co ([7Fe:9S:C:Mo]-*R*-homocitrate) and the P-cluster ([8Fe:7S]). The MoFe-protein is a dimer of αβ hetero-dimers, and has a molecular weight of ~230 kDa (α-subunit is 55.3 kDa, and β-subunit is 59.5 kDa). Each αβ subunit coordinates one FeMo-co, located between the three domains of the α subunit. The P-cluster is located at the interface between α and β subunits, ~10 Å from the surface of the protein. The distance between the P-cluster and FeMo-co is ~14 Å, supporting a model of efficient electron transfer. The P-cluster can be considered as a used 2[4Fe:4S] cluster, with one sulfur removed. In the as-isolated state, the cluster is coordinated by seven protein residues: αCys62, αCys88, αCys154, βCys70, βCys95, βCys153 and βSer188. Ligation of the serine residue varies with oxidation state. As-isolated, the P-cluster irons are all-ferrous,

but the cluster can be oxidized using chemical oxidants. P-cluster oxidation has been associated with conformation changes, determined crystallographically. The active-site FeMo-co is a similarly fused cubane system, with a central carbon atom and three belt sulfur atoms connected two three- and four-iron cubanes. As a feature of the molybdenum nitrogenases, the FeMo-co contains a molybdenum atom at the base of the structure, which is coordinated to an organic *R*-homocitrate moiety. The FeMo-co is coordinated to the protein by only two ligands: αCys275 at the apical iron, and αHis442 at the molybdenum atom, completing its ligand sphere with homocitrate. The structure and dynamics of the FeMo-co, through interactions with inhibitors and substrates has been an active area of interest. Recent efforts have yielded remarkable structures identifying the central carbon atom and demonstrating the reversible lability of the belt sulfurs[12–14] in a turnover-dependent fashion.

The Fe-protein is a γ_2 homodimer and contains one [4Fe:4S] cluster and two ATP binding sites. The Fe-protein is 64 kDa in molecular weight and the [4Fe:4S] cluster is coordinated by Cys97 and Cys132, sitting at the dimer interface. The Fe-protein nucleotide binding site is characterized as a deviant Walker A motif, which contains a conserved KGG protein sequence. The nucleotide state is hypothesized to be coupled to the [4Fe:4S] cluster through the Switch I and II regions, corresponding to residues 38-43 and 125-135, respectively. Complex formation with the MoFe-protein results in a displacement of ~5 Å of the [4Fe:4S] cluster towards the interaction face with the MoFe-protein, as well as a ~13 degree tilting of the protein subunits towards the dimer two-fold axis (in total a 26 degree movement of one subunit relative to the other).

During catalysis, ATP binding and hydrolysis results in the Fe-protein binding to the MoFe-protein, and transferring electrons from the [4Fe:4S] cluster to the intermediate P-cluster. The P-cluster transfers electrons to the active-site FeMo-co. The FeMo-co, once in a sufficiently reduced state, is competent to bind substrates. Multiple associations and dissociations of the Fe-protein to the MoFe-protein in the presence of ATP and external reductant (physiologically flavodoxin or ferredoxin, *in vitro* with dithionite) yields

catalysis of multi-electron reduction of substrates. The wild-type reaction of nitrogenase is the reduction of dinitrogen to ammonia, but a variety of small molecules are also accommodated (e.g. acetylene to ethylene).

Despite the longevity of nitrogenase research, there remain fundamental questions regarding the nitrogenase catalytic cycle, mechanism of substrate reduction, and cofactor assembly.

1.3.1 Models of the Nitrogenase Mechanism

Current models for nitrogen reduction suggest that each Fe-protein dimer interacts independently with the MoFe-protein, resulting in a non-cooperative mechanism. The Fe-protein accepts low-potential electrons from a flavodoxin or ferredoxin and passes them, with 2 ATP hydrolysis reactions, to the MoFe-protein. Recent evidence suggests that dissociation of the inorganic phosphate after ATP hydrolysis by the Fe-protein is the rate-limiting step in the reaction[15]. The Fe-protein is the only physiological source of electrons capable of producing a reduced MoFe-protein that can efficiently reduce dinitrogen to ammonia.

The model of this catalytic cycle is summarized by the Thorneley-Lowe model[2,16], in which each electron transfer event is accompanied by ATP hydrolysis and dissociation of the Fe-protein (Figure 3). The enzyme begins in the E_0 "as-isolated" state, referring to the dithionite-purified enzyme. The E_0 state is not capable of substrate binding.

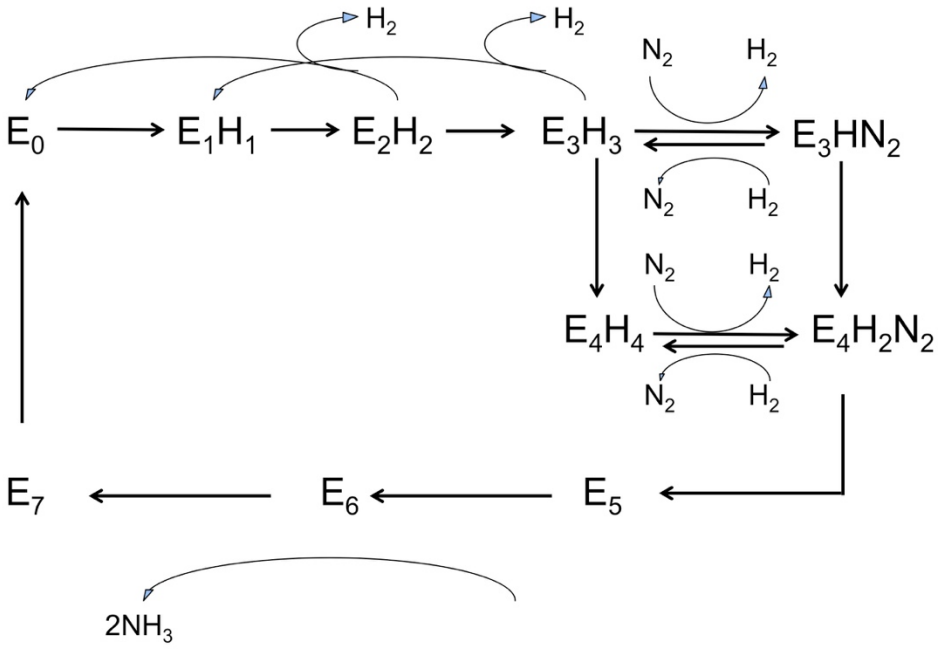


Figure 3. Thorneley-Lowe kinetic model of the MoFe cycle[16,17].

The Thorneley-Lowe model describes an 8-electron cycle, with obligatory hydrogen formation upon nitrogen reduction. The model assumes that substrates and products only bind to the free MoFe-protein, and that the Fe-protein transfers 1 electron per 2 MgATP hydrolyzed before dissociation. Both active sites of MoFe-protein are treated independently, and the Fe-protein undergoes its own cycle independent of the MoFe-protein. The endogenous source of electrons for nitrogenase originate from a flavodoxin or ferredoxin and are passed through a series of redox-active cofactors before being delivered to the active site[2,17]. The overall electron transfer pathway of electrons per E-state transition is shown as follows (Figure 4).

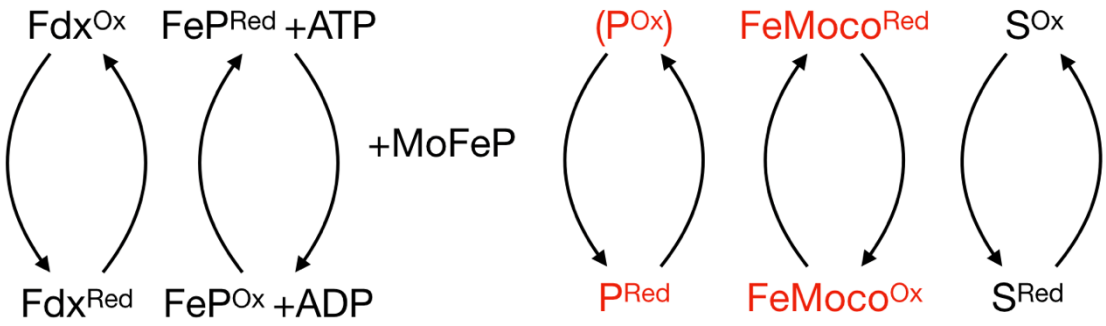


Figure 4. General redox cycling in the nitrogenase reaction. One electron from a ferredoxin is passed to the Fe-protein, which binds ATP in the presence of the MoFe-protein. In this model, upon hydrolysis, an electron is transferred to the P-cluster, which then transfers the electron to the FeMo-co for substrate reduction. In the scheme, (P^{ox}) indicates an oxidized form of the P-cluster, and not specifically the two-electron oxidized state, P^{ox}. Superscripts indicate generic reduction states, and cofactors are in red.

1.3.2 Mechanism and Electronics of the Iron Protein

The nitrogenase Fe-protein is a specialized reductase that initiates the transfer of electrons to the active site, coupled with ATP hydrolysis. The Fe-protein contains a unique [4Fe4S] cluster that can transfer 1 electron through the 2+/1+ couple, as well as a 2-electron transfer with a 2+/0+ couple. The biological relevance of the 0+ state is still debated, but the possibility of using the 2-electron transfer reduces the ATP demand by half[18]. Thorough biochemical studies of *in vitro* use of the 2+/1+ versus 2+/0+ couple have shown that the rate of electron transfer from the Fe-protein to the substrate-binding MoFe-protein is equivalent, as the MoFe-protein dictates electron transfer internally for substrate reduction. Additionally, the affinity of the Fe-protein for the MoFe-protein was demonstrated to be equivalent in the 1+ and 0+ states[19]. A major remaining question is how the redox state of the iron protein can influence ATP usage.

Examining the distribution of electrons in the cluster of the Fe-protein is an important first step in understanding the relationship to activity. The Fe-protein [4Fe4S] cluster has cysteine sulfur ligands, and is situated in a cleft between the two monomeric protein

subunits, such that one face of the cluster is exposed to solvent, and the other face is buried in the protein matrix. Bulk spectroscopic techniques such as electron paramagnetic resonance (EPR) have been able to detect changes in the [4Fe4S] cluster based on the nucleotide state. However, only the 1+ state is EPR accessible and paramagnetic; the 2+ state is diamagnetic (S = 0), and the 0+ state is an integer spin (S = 4)[5,20,21]. The 1+, nucleotide-free state has *g* values of 2.04, 1.94, 1.87, and 4.8. The *g* = 1.94 signal has been assigned to an S = 1/2 ground state, and the *g* = 4.8 signal has been attributed to an S = 3/2 population, although the low-field signal is often broadened. Signals at *g* = 4.3 have been reported for the ATP-bound state of the Fe-protein, and observed under freeze-quench turnover conditions with the MoFe-protein and substrate. The ratio of the S = 3/2 to 1/2 signal intensity changes depending on nucleotide state and perturbing buffer conditions (e.g. glycerol or urea), reflecting the solvent accessibility of the cluster or conformational state.[20,21] Notably, the *g* = 4.3 signal can also arise from contaminating ferric iron (in the case of Lindahl et al., the signal was removed by stimulation of ATP hydrolysis, suggesting the signal did not arise from ferric iron, but a state dependent on the nucleotide bound). The mixture of spin states and sensitivity to solvent content have presented significant challenges in unambiguous characterization of Fe-protein electronics.

Studies using model compounds have characterized many properties of synthetic FeS compounds: broken-symmetry density functional theory (DFT) calculations and spectroscopic techniques (Mössbauer and EPR) indicate that the [4Fe4S] cluster can subdivide into two-[2Fe-2S] pairs.[20,22] The 2+ state is modeled as a pair of high-spin Fe coupled ferromagnetically, and a second pair of two low-spin Fe coupled antiferromagnetically, with the sixth d-electron delocalized between the irons in each sublayer. Conclusions are difficult to translate into the protein environment due to solvent and ligand effects, and current protein studies are limited by techniques that cannot distinguish localized versus delocalized electrons in the cluster.

1.4 The *nif* Operon

A common model system for studying nitrogenase is *Azotobacter vinelandii*, a Gram-negative soil bacterium. *A. vinelandii* is an obligate aerobe that is capable of overexpressing the nitrogenase enzyme under nitrogen-fixing/ammonia starvation conditions such that 10% of its cellular protein is nitrogenase[19] . Roughly 1% of the genome of *A. vinelandii* is devoted to the regulation, assembly, and activity of nitrogen fixation.

The nitrogen fixation (*nif*) operon is divided into a major and minor gene cluster, located close to and equidistant from the origin of replication. The major cluster contains the genes for the molybdenum-iron nitrogenase, as well as most of the associated biosynthetic machinery. The minor cluster contains some biosynthetic genes and regulatory factors, including those involved in molybdenum trafficking. *A. vinelandii* also contains distinct structural genes for the vanadium- and all iron-nitrogenase on a separate part of the genome.

Fewer than 50% of *nif* gene products are functionally characterized, but eighteen *nif* genes have been identified as being essential for the nitrogenase maturation pathway[23]. For clarity, the *nif* gene products are tabulated below (Table 1).

Gene	**Product**
nifH	Fe-protein
nifDK	MoFe-protein
nifEN	Assembly protein
nifX, nifS, nifU, nifV, nifW, nifZ, nifB	Assembly/4Fe:4S transport proteins
nifF, nifJ, nifH	Assembly proteins
nifL, nifA	Regulation protein

Table 1. Protein products of major *nif* genes[23–25].

1.5 P-cluster maturation

The functional role of the Fe-protein is complex, and in addition to substrate reduction, it is essential in nitrogenase metallocluster biosynthesis. The Fe-protein is essential for the biosynthesis of the two unique metalloclusters in the MoFe-protein, the P-cluster [8Fe:7S] and the active site FeMo-co [7Fe:9S:Mo:C]-*R*-homocitrate cluster (Figure 5, and described below)[24]. The P-cluster is structurally described as two partial [4Fe:3S] cubanes, bridged by a μ_6-sulfide (in the P^N state), and can exist in three oxidation states (P^N, P^{1+}, P^{ox}). As-isolated in dithionite, the P-cluster is in the all-ferrous, diamagnetic state (P^N). The FeMo-co (or M-cluster) is structurally a [4Fe:3S] and [Mo:3Fe:3S] partial cubane, bridged by three μ_2-sulfurs and a μ_6-interstitial carbon. An *R*-homocitrate molecule coordinates the molybdenum through its 2-hydroxyl and 2-carboxyl groups.

The maturase activity of the Fe-protein was determined by isolating MoFe-protein in a *ΔnifH* deletion strain. The *ΔnifH* MoFe-protein contained immature P- (and FeMo-co) clusters, later identified as two $[4Fe:4S]^{1+}$ clusters by electron paramagnetic resonance (EPR) and Fe K-edge X-ray absorption spectroscopy (XAS) [26]. XAS revealed that the clusters were composed of a normal [4Fe:4S] and a distorted [4Fe:4S], potentially with a bridging cysteine residue. Reconstitution assays consisting of Fe-protein, reductant, and ATP yielded functional P-clusters. These studies suggest that the Fe-protein can reductively couple the two [4Fe:4S] subclusters. Many structural and mechanistic details of this process are missing, including the docking orientation of the Fe-protein on the *ΔnifH* MoFe-protein, and elimination of the "8^{th} sulfur" during coupling of 2[4Fe:4S] to the mature [8Fe:7S] (summarized in Figure 5). Furthermore, EPR studies suggest that the P-cluster fusion occurs in a step-wise fashion, breaking the symmetry of the MoFe-protein observed in its holo form. In a double knockout of two other maturation factors (*ΔnifBΔnifZ)*, a MoFe-protein was isolated with one immature two-subcluster "P-cluster," and one mature P-cluster[27]. The discovery of the asymmetric maturation of the P-clusters suggests a complicated, conformationally distinct binding of the Fe-protein in contrast to substrate reduction.

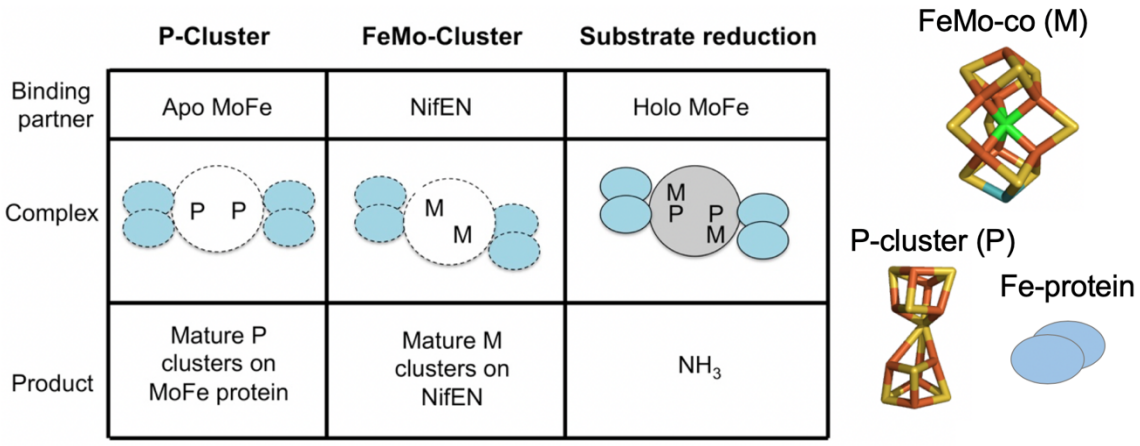


Figure 5. The proposed roles of the Fe-protein in substrate reduction and cofactor maturation. Fe-protein is shown in blue. The crystal structure of the ADP•AlF_4^- -stabilized complex is shown in gray, and the proposed cluster maturation complexes for the P-cluster and FeMo-co are shown in white. P and M indicate locations of the P- and FeMo-co clusters, respectively[24].

1.6 FeMo-co maturation

The biosynthesis of the FeMo-co involves multiple *nif* gene products and the exchange of immature cofactors previously identified by deletion strains of *Azotobacter vinelandii*. First, an 8Fe core is generated from two 4Fe clusters on NifB, which mediates the radical SAM insertion of the interstitial carbon. NifB passes the immature FeS core to NifEN, a scaffolding protein homologous to the MoFe-protein. At this stage, the 8Fe core is "matured" by addition of molybdenum and homocitrate from the Fe-protein. The fully assembled FeMo-co is then passed from NifEN to the apo-MoFe-protein through a direct protein-protein interaction, as modeled above (Figure 5).

Evidence for Fe-protein involvement is derived from cluster reconstitution assays using cofactor-deficient proteins. In an assay where incomplete cofactors on NifEN were incubated with apo MoFe-protein and Fe-protein, only Fe-protein incubated with molybdate/homocitrate, ATP, and reductant were competent to generate holo MoFe-protein (as assayed by acetylene reduction).[24] Furthermore, EPR and molybdenum K-edge X-ray absorption spectroscopy (XAS) spectroscopic studies of Fe-protein incubated with

molybdate, homocitrate, ATP, and reductant yield distinct spectra from the spectra of free molybdate and Fe-protein alone with ATP and reductant. These experiments indicate that the molybdate species is altered upon binding to the Fe-protein[28]. Homocitrate is pre-formed by condensation of acetyl-CoA and α-ketoglutarate by NifV. It is unclear how homocitrate is inserted, but from reconstitution assays, it is inserted during Fe-protein-NifEN complex formation in parallel with molybdenum. The mobilization of molybdenum and insertion is not fully understood.

In addition, the first crystal structure of the Fe-protein from *Azotobacter vinelandii* was crystallized in the presence of 0.2 M sodium molybdate as a precipitant.[29] Density (including anomalous density observed at the copper K-edge) assigned to a partially occupied molybdate was identified in the phosphate-binding loop (P-loop), in the position usually occupied by the γ-phosphate of ATP (Figure 6). The P-loop has a signature deviant Walker A motif common to many nucleotide-binding proteins, and contains the sequence GXXXXGK(T/S), where T/S may be threonine or serine. In the *A. vinelandii* Fe-protein, the sequence is GKGGIGKS, starting with glycine 9. The structure provides an intriguing possibility of phosphate-molybdate promiscuity at that binding site, and may be relevant in cofactor maturation.

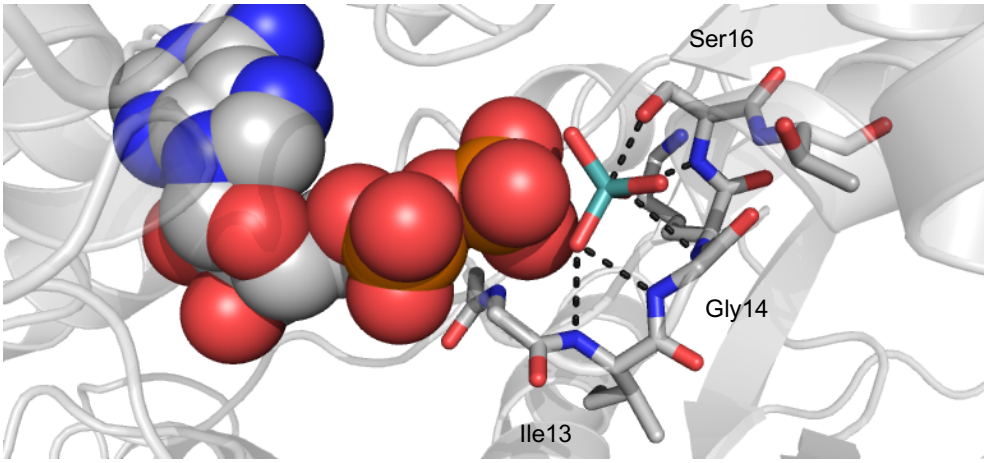


Figure 6. Nucleotide binding region in the Fe-protein (P-loop). The molybdate modeled from the first Fe-protein structure is shown in sticks[29]. Coordinating residues are shown in a stick representation, ADP is shown as spheres.

Multiple questions arise concerning the varied roles of the Fe-protein. How the Fe-protein achieves the highly concerted reaction of molybdenum and homocitrate insertion is not known, but it is of key mechanistic importance for the detailed understanding of complex metal center maturation and nitrogenase assembly.

1.7 X-ray Crystallography

Macromolecular X-ray crystallography is a technique that uses the diffraction of X-rays from a protein crystal to reconstruct the three-dimensional atomic structure of the crystalline material (protein in a crystal lattice). The technique is commonly used to solve biological structures of proteins and nucleic acids, as well as small molecules. Hard X-ray radiation is used because the wavelength of light is on the order of the length of atomic bonds (10^{-10} m, or Ångstrom range), and is thus appropriate for resolving the bonded structure of the material, which, in the case of proteins, include C-C, C-O, and C-N bonds (average 1.5 Å in length).[30,31] Single molecule diffraction of X-rays is extremely weak in intensity, and crystals are used to amplify the diffraction signal. Diffraction intensity is strongest when there is constructive interference in a three-dimensional periodic array of repeating units, given a set of geometric conditions. Although determining structures using crystals renders a solution that is an average structure of an ensemble, the gain in intensity allows for the detection of reflections at very high resolution, but only if the average structure is sufficiently well defined.

1.7.1 Protein Crystallization

Protein crystals are self-assembled periodic arrays of macromolecules that are held together through predominantly electrostatic interactions. Self-assembly occurs through a nucleation process in a super-saturated protein solution. In order to generate a solution that favors macromolecular nucleation, precipitants are added to a protein solution and dehydrated (Figure 7) through the vapor diffusion technique[30]. Other techniques include batch methods (direct addition of precipitant), dialysis solvent exchange, free-interface

diffusion, and pH changes. In vapor diffusion, water diffuses out of the protein solution, in the thermodynamically favorable direction and into an adjacent solution of concentrated precipitant.

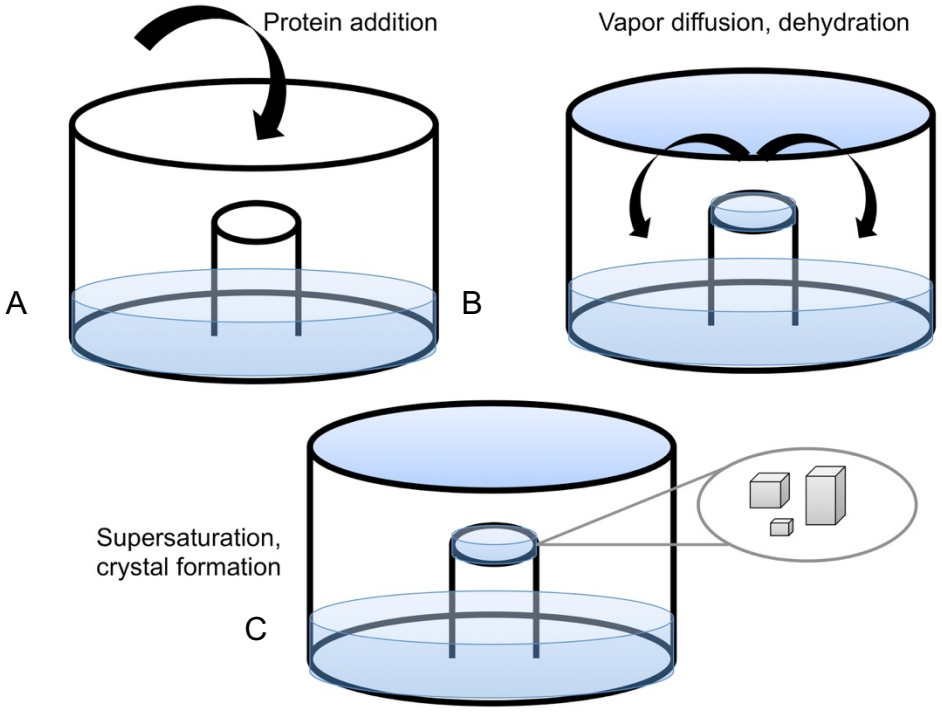


Figure 7. A schematic of a sitting-drop vapor diffusion experiment. (A) A well inside a sealed chamber contains a precipitant solution surrounding a pedestal. The protein solution of interest is mixed with a precipitant solution and pipetted into the pedestal, before the chamber is sealed. (B) Water evaporates from the protein/precipitant solution, which contains a lower concentration of precipitant than the well solution. The diffusion of water out of the protein drop concentrates the protein solution slowly over time, thus allowing for protein crowding in solution. (C) Under the proper conditions, nucleation occurs, after which crystal growth is thermodynamically favorable[31].

The protein crystallization process can be illustrated with a phase diagram (Figure 8). As protein and precipitant concentrations are increased, the solution equilibrium transitions from one phase (in the undersaturated region, bottom left) to at least two phases (top right) where protein in the solution and protein in solid form exist in equilibrium. As proteins nucleate into solids, they grow in three dimensions from still-soluble protein.

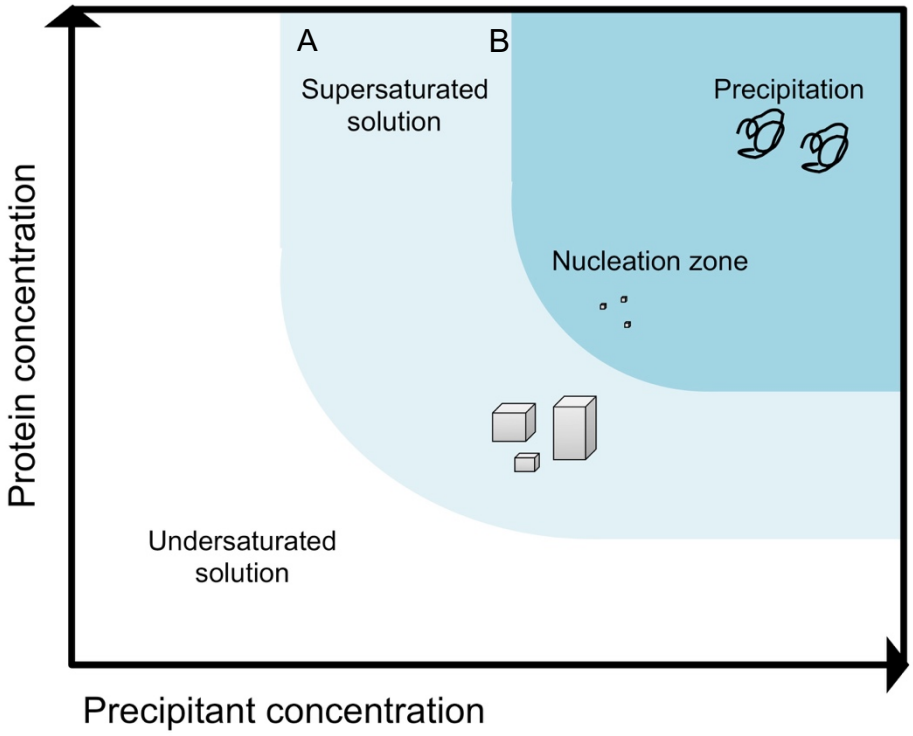


Figure 8. Schematic phase diagram of a crystallization experiment at constant temperature (although the principle applies to other variables, including temperature, pH, etc.). The white region (A) indicates the solubility zone, and the blue-shaded areas (B) indicate the instability zone that denotes where the solution becomes bi-phasic (solid crystals or solid aggregates in a solution that is in equilibrium with soluble protein) [30]. The light blue shaded area represents the labile and meta stable zones, where spontaneous nucleation and crystal growth occur. Crystals will continue to grow until they reach the saturation point, which is the boundary between (A) and (B).

In the formation of a crystal, there is an entropic penalty due to the confinement of a protein in a specific conformation, and a modest gain in enthalpy due to the weak protein-protein interactions in the crystal.[30] During crystallization, there is a gain in entropy of the bulk solvent as ordered waters are shed from the protein when adding to the crystal lattice. Ultimately, the free energy change is dependent on many weak interactions, so small changes to the protein (e.g. mutagenesis of surface residues, shielding effects from precipitants and solvent) can have dramatic effects on crystallization.[31]

Crystallization is also dependent on the kinetics of nucleation and growth. Crystal nucleation occurs when protein molecules in a supersaturated solution collide and form a metastable structure that can decompose, or be sufficiently stable to incorporate new proteins. After nucleation, crystal growth can occur at lower concentrations of protein (compared to the initial nucleation concentration)[31], thus seeding with microcrystals to circumvent nucleation facilitates growth and is a commonly used technique.

The resulting crystal is made up of a three-dimensional lattice, which in itself is a repeating array of protein (and solvent) molecules. The unit cell is the simplest non-repeating unit, defined here by the basis vectors **a**, **b**, and **c**, that can fully describe the entire crystal lattice by translation. The minimal arrangement of molecules that are related to one another within the unit cell by any symmetry operation is referred to as the asymmetric unit[30].

1.7.2 Diffraction

Scattered X-rays from a crystal lattice that constructively interfere manifest as reflections in a diffraction pattern. Points of constructive interference occur when Bragg's law is satisfied. Bragg's law is given by:

$$n\lambda = 2dsin\theta \qquad \text{Eq. 3}$$

Where n denotes an integer, λ is the wavelength of incident radiation, d denotes the interplanar spacing between atoms in the lattice, and θ denotes the angle between the incident radiation and the lattice plane (Figure 9). Coherent scattering is maximally produced when the $2dsin\theta$ term is an integer of the wavelength, λ. The $dsin\theta$ term reflects the path length difference (for a total of $2dsin\theta$) due to the spacing between the scattering planes[30].

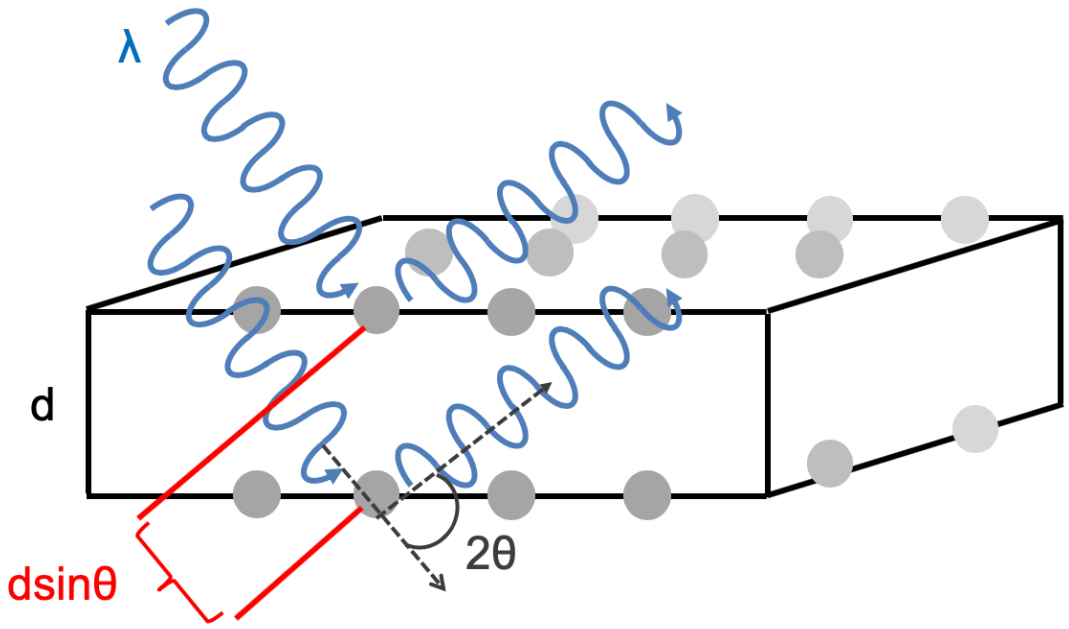


Figure 9. A diagram illustrating the Bragg condition for diffraction. Radiation of wavelength λ interacts with two lattice planes (*h,k,l*) made up of atoms (gray) and a spacing *d*. Maximal coherent scattering is achieved when the conditions $n\lambda = 2sin\theta$ are satisfied.

This construction demonstrates that every atom in the crystal lattice contributes to the intensity of the reflection. For large, imperfect crystals built of mosaic blocks, the intensity is directly proportional to the square of the scattering factor, (F_s). The total scattering of a crystal is described by the scattering factor:

$$F_s = \sum_{j=1}^{atoms} f_{S,j}^{0} e^{2\pi i S r_j} \qquad \text{Eq. 4}$$

Where *S* is the scattering vector and r_j are the coordinate vectors for the lattice over each atom, *j*. In reciprocal space, the set of lattice planes separated by *d* are converted into a single reciprocal lattice point. This phenomenon can be described mathematically by Fourier transformation where the reciprocal lattice points form reciprocal lattice planes. The Ewald sphere is a convenient construction to understand the Fourier relationship between the diffraction pattern and the real-space organization of a crystal.[30] From Bragg's law, the relationship can be defined:

$$d^*_{hkl} = \frac{1}{d_{hkl}} = \frac{2sin\theta}{n\lambda}$$ Eq. 5

Where d^*_{hkl} is the reciprocal lattice vector, normal to the lattice planes in real space (x,y,z). The scattering vector, *S,* can be included to satisfy Bragg's condition:

$$d^*_{hkl} = \frac{2sin\theta}{n\lambda} = |S|$$ Eq. 6

The Ewald sphere construct describes where points of constructive interference form a crystal result in a diffraction pattern. In an experiment, the crystal is centered at the origin of the Ewald sphere (Figure 10), and the vectors *S* and d^*_{hkl} are parallel (note that the scattering vector *S* is the vector difference between the scattered and incident vector $S = S_1 - S_0$). At every point where Bragg's law is satisfied, a reflection will be generated for a set of lattice planes when the reciprocal lattice point intersects the so-called Ewald sphere, which has a radius (1/λ). Coincidentally, diffraction patterns can appear as rounded lunes, which are reminiscent of slices in a sphere.

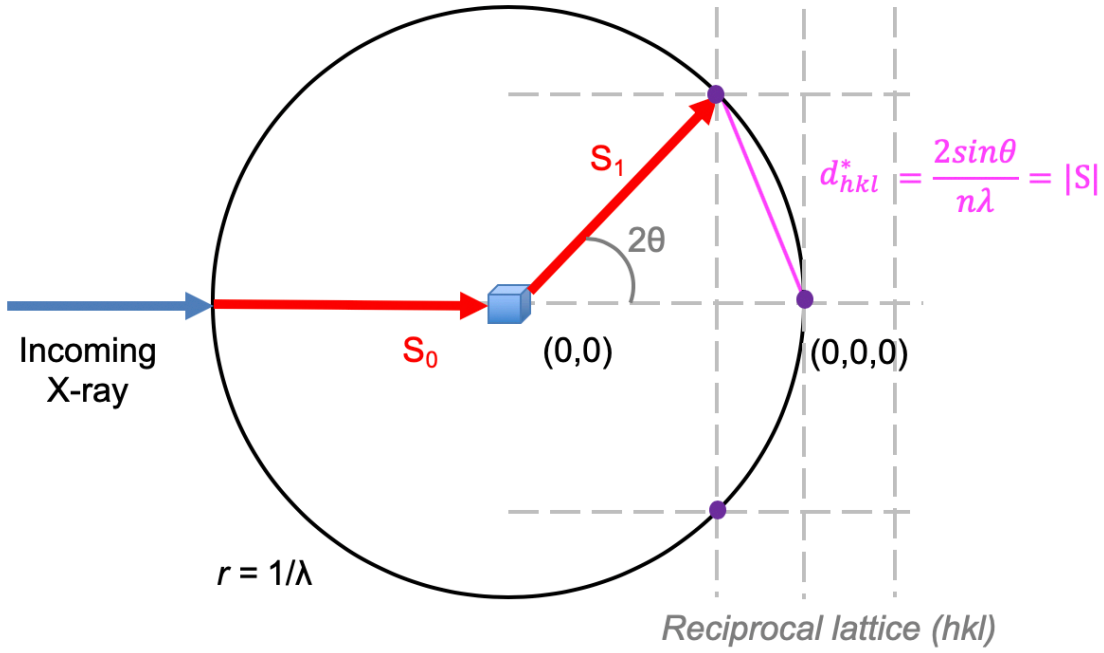


Figure 10. Diagram of the Ewald sphere construction. A crystal is centered in real space at the origin of the Ewald sphere, which has a radius of 1/λ. The total scattering vector, *S*, is made up of S_0 and S_1, which are each equal to 1/λ. When $d^*_{hkl} = \frac{2sin\theta}{n\lambda} = |S|$, the Ewald sphere and the reciprocal lattice point intersect, generating a reflection[30].

In order to obtain a complete data set, all of the reciprocal lattice points must be sampled. This is achieved by rotating the crystal in real space. As the crystal rotates, the reciprocal lattice rotates. The diffraction pattern of the crystal at each orientation is collected on a detector, which records the position and intensity of each reflection. The maximum angle where diffraction is observed is *2θ* (180 degrees), or backward scattering.

1.7.3 The Scattering Factor

The contents of the unit cell can be calculated by measuring the position and intensity of each reflection on the detector. The intensity of the reflection is dependent on the scattering material in the crystal, and specifically, the contents of the unit cell.[30] The total scattering of every atom in the unit cell is represented by the structure factor, and is contained in each reflection:

$$F_h = \sum_{j=1}^{atoms} f_{s,j}^0 e^{2\pi i h x_j} \quad \text{Eq. 7}$$

where h is the direction of scattering and x_j is the fractional coordinate vector that describes the positions of the atoms relative to the origin. The structure factor contains information of the phase of each contributing scattering wave[30]. Another representation of the structure factor is as a summation of the scattering from the electron density inside the unit cell:

$$F_h = V \iiint \rho_{xyz} e^{2\pi i(hx+ky+lz)} dxdydz \quad \text{Eq. 8}$$

As follows, the Fourier transform of the structure factor would then yield the electron density:

$$\rho_{xyz} = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \text{Eq. 9}$$

In returning to interpreting the diffraction pattern, the intensity of the reflection is dependent on the contents of the unit cell. The intensity can be more rigorously related to the structure factor as proportional to the square of the amplitude:

$$I \propto |F_{hkl}|^2 \quad \text{Eq. 10}$$

The scattered wave has a phase component, which cannot be extracted from the intensity recorded in a diffraction pattern.[30] Although direct methods may be applied to small molecules (<200 atoms), macromolecular diffraction has too many reflections and requires an additional set of experiments. For macromolecules, several experiments may be used to determine phase information, including molecular replacement (MR), single-wavelength anomalous dispersion (SAD), and multi-wavelength anomalous dispersion (MAD).

1.7.4 Molecular Replacement

Molecular replacement (MR) is a method to determine the phases of an unknown structure by using phases calculated from a model, typically derived from a previously solved structure. If the MR model is isomorphous (it contains the same contents with identical unit cell parameters) to the target structure, the calculated phases from the model can be used directly to predict the phases for the new structure. However, if the model is non-isomorphous, the Patterson map is used to relate the orientation of the MR model to the new data so that the calculated phases of the model can be used. The Patterson map contains interatomic distance vectors, and is generated solely from the experimental intensities.[30] The Patterson function is:

$$P(x) = \frac{1}{V}\sum_h |F(h)|^2 e^{-2\pi i h x} \qquad \text{Eq. 11}$$

The calculated Patterson map of the model is rotated and translated to generate the Patterson map that closely matches the unphased structure. The correlation between the known and unknown Patterson maps is used to score the solutions. Once the orientation is determined, the MR phases are applied to the new data set. Generally, a MR model should be structurally homologous and be ~30% identical in sequence to obtain a close fit to the unphased structure.

1.7.5 Anomalous Scattering

In the case of anomalous experiments, phase information is ultimately elucidated from the differential scattering behavior of heavy atoms within a structure. Returning to the atomic scattering factor, for a heavy atom the total scattering factor is:

$$f(\lambda) = f^0 + f'(\lambda) + if''(\lambda) \qquad \text{Eq. 12}$$

where the scattering factor $f(\lambda)$ contains contributions from f^0, which is independent of the wavelength and dependent on the scattering angle, as well as $f'(\lambda)$ and $f''(\lambda)$, which are wavelength-dependent terms that are independent of the scattering angle[30]. The $f'(\lambda)$ and $f''(\lambda)$ terms are respectively the real and imaginary terms of the scattering factor, where $f'(\lambda)$ is the dispersive component and $f''(\lambda)$ is the anomalous component of the atomic scattering factor. The extra wavelength-dependent terms are needed to account for the X-ray absorption (inelastic scattering) by the heavy atom. The absorption behavior of heavy atoms is related to their atomic number. At specific wavelengths of X-rays, the resonance energy matches the energy gap required to promote transitions of electrons in the heavy atom from low to high energy shells. The absorption promotes transitions from the n = 1, 2, 3 shells, resulting in the so-called K, L, and M X-ray absorption edges, respectively (Figure 11). The energy difference is apparent in the valence versus core electron transitions; optical absorbance promotes transitions of valence electrons, whereas X-ray photons are much higher in energy and affect core electrons[30]. The sharp increase in absorption at the K-edge of Se, shown below, indicates where the energy of the incident X-ray is resonant with the Se 1s transition. At this point, the energy of the incident photon is dominated by an inelastic scattering event.

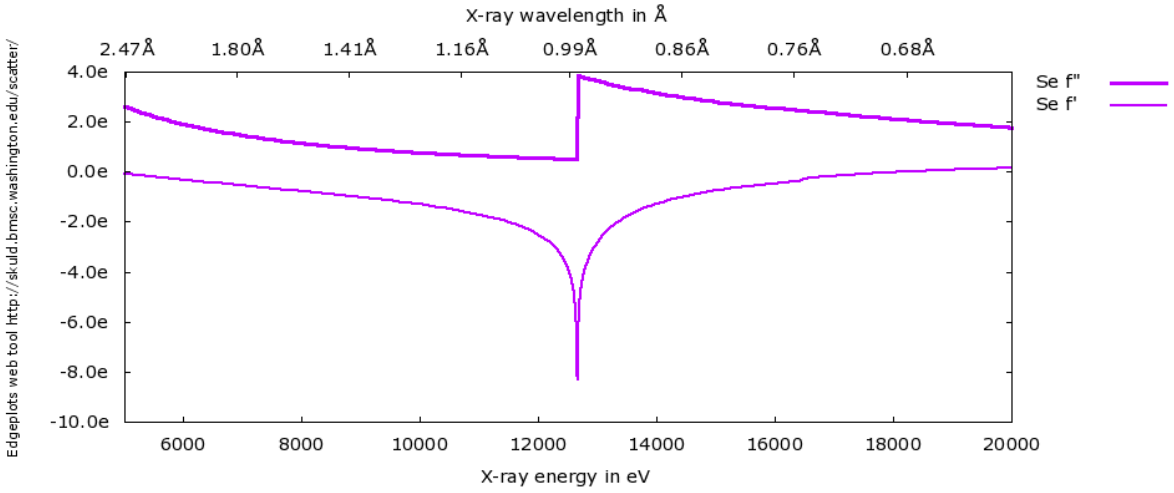


Figure 11. The calculated Se K-edge X-ray absorption spectrum, from the University of Washington plotting tool (http://skuld.bmsc.washington.edu).

In the absence of significant X-ray absorption, the reciprocal lattice is centrosymmetric. That is, the reflection *(h,k,l)* is equivalent to *(-h,-k,-l)* and the intensity ($I_h = I_{-h}$). The phase angle is equivalent in magnitude but of opposite sign (φ = -φ). The pair of equivalent reflections is referred to as a Friedel pair. When a heavy atom is present and anomalous absorption occurs, there is a breakdown of the Friedel pair (Figure 12), where they are no longer equivalent and are referred to as a Bijvoet pair. The heavy atom does not affect the phase of the normal contribution, but the anomalous contribution changes the phase by 90°[30]. The overall structure factor with the heavy atom has an intensity and phase change, breaking the symmetry of the Friedel pair.

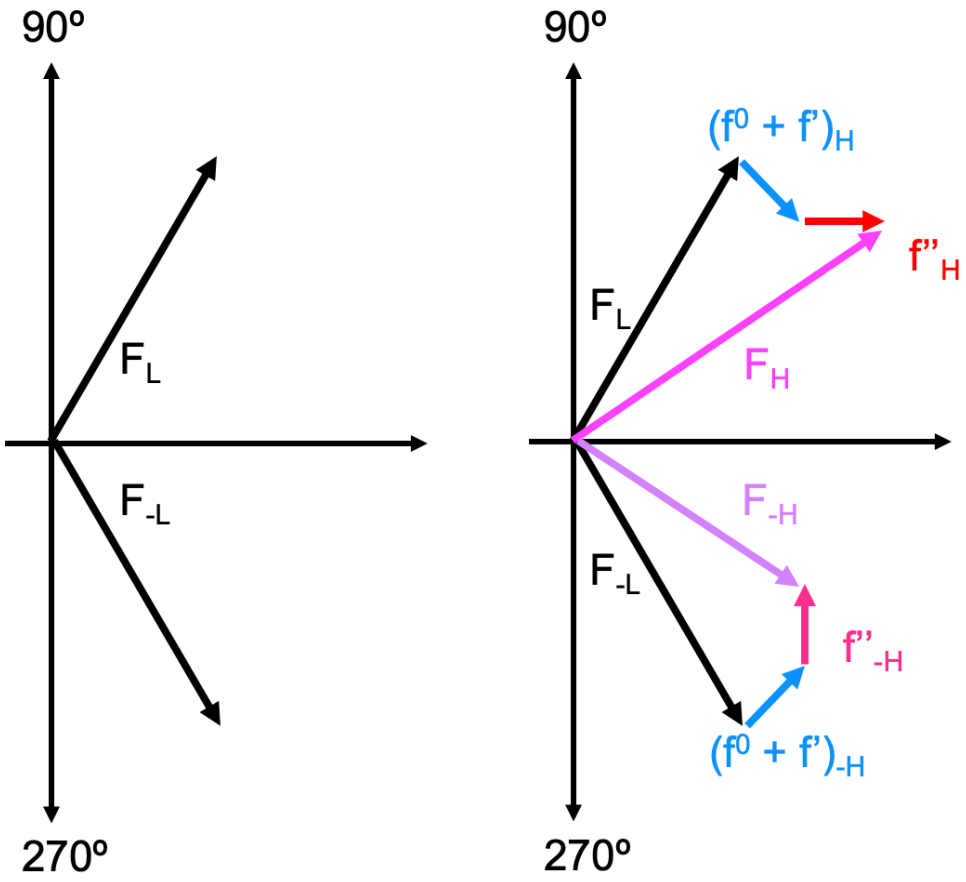


Figure 12. An Argand diagram illustrating the effect of the anomalous contribution of a heavy atom to the scattering factor. The arrows represent the total scattering vector, and the subscripts L and -L denote the conjugate pair (*h,k,l*) and (*-h,-k,-l*). The f'' contribution is at a right angle to $f^0 + f'$ due to the phase shift of +90°. (Left) Friedel pair of the scattering factors of a light atom. (Right) Heavy atom replacement contains an additional anomalous contribution contributes to an overall break in symmetry such that $|F_H| \neq |F_{-H}|$[30].

Anomalous differences can be used to determine the substructure of heavy atoms (that were soaked in or co-crystallized if not naturally present) in a structure (e.g. SAD and MAD), which can be applied to solve the phase of an isomorphous light-atom structure. For example, seleno-methionine can be substituted for methionine in a protein, and the phases calculated from the selenium-containing dataset can be used to phase the native protein data.

1.7.6 Spatially Resolved Anomalous Dispersion Refinement (SpReAD)

The anomalous contribution to the scattering factor has also been used by Einsle and authors to determine relative oxidation states of heavy atoms within a protein scaffold based on work by Coppens[32–34]. The scattering factor can be separated into the light atom (protein) and heavy atom contributions:

$$F_{h,\lambda} = \left\{\sum_m f_m e^{2\pi i r_m h}\right\} + \left\{\sum_n f_\lambda e^{2\pi i r_n h}\right\} \quad \text{Eq. 13}$$

where the first bracketed term describes the light atom (protein) scattering, and the second bracketed term describes the heavy atom scattering for *n* heavy atoms in a structure of *m* atoms in the unit cell. The dispersive and anomalous differences can also be represented as:

$$\Delta_{disp}(h,\lambda) = |F(h,\lambda)| - |F_{calc}(h)|$$
$$\Delta_{ano}(h,\lambda) = |F(h,\lambda)| - |F(-h,\lambda)| \quad \text{Eq. 14}$$

where F_{calc} is calculated omitting f". $\Delta_{disp}(h,\lambda)$ and $\Delta_{ano}(h,\lambda)$ can be combined with the expression for the phase of the wavelength-independent scattering to give:

$$\Delta_{disp}(h,\lambda) = \sum_n \frac{\Delta f_n'(\lambda)}{f_n^0}(a_n(h)cos\alpha(h) - b_n(h)sin\alpha(h))$$
$$\Delta_{ano}(h,\lambda) = -2\sum_n \frac{\Delta f_n''(\lambda)}{f_n^0}(b_n(h)cos\alpha(h) - a_n(h)sin\alpha(h)) \quad \text{Eq. 15}$$

The $\Delta f'$and $\Delta f''$can be refined for individual scatters from full datasets collected at different wavelengths along an absorption edge of a heavy atom, yielding individual $\Delta f'(\lambda)$ and $\Delta f''(\lambda)$. The anomalous contributions are robust and are proportional to the absorption of the heavy atom. Plotting the $\Delta f''(\lambda)$ values for individual heavy atoms as a function of X-ray wavelength yields site-specific X-ray absorption curves. The method was tested in the case of the [2Fe:2S] cluster of a ferredoxin from *Aquifex aeolicus*[34]. Full

crystallographic datasets were collected on a single crystal at seven energies that span the iron absorption edge: 7114, 7116, 7118, 7120, 7122, 7124, 7126 eV. A high-resolution dataset was collected at 7140 eV to calculate the structure factors for the light atoms. The analysis recapitulated the expected absorption edge of the iron heavy atoms, and the $\Delta f''(\lambda)$ values were consistent with a low-energy shift of one iron relative to the other. In X-ray absorption spectroscopy, edges at high energy are consistent with more oxidized irons, and low energy-shifted curves are consistent with more reduced irons (assuming all other variables, such as local environment, are equivalent). The SpReAD profile of *A. aeolicus* ferredoxin was assigned to an Fe(3+)Fe(2+) system, which is consistent with spectroscopic studies. The [2Fe:2S] yields an S=1/2 assignment in the reduced form, consistent with Fe(3+)Fe(2+) localized model (as opposed to a delocalized Fe(2.5+)Fe(2.5+) model)[34]. The utility of the method comes from the ability to measure X-ray absorption (classically an ensemble method in solution) in a spatially resolved manner.

1.8 Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) spectroscopy is a technique that probes molecules with unpaired electrons. EPR can be useful for characterizing metalloproteins by providing evidence to identify metal identity, establishing the spin states and probing ligand state. EPR probes the interactions of unpaired electrons by promoting transitions in a magnetic field.[35,36] The transitions have associated energies as described by Planck's law:

$$\Delta E = h\nu \qquad \text{Eq. 16}$$

where the energy change *(ΔE)* is related to Planck's constant h and the frequency of radiation, ν. The radiation frequency used by EPR is in the gigahertz (GHz) range. Transitions are promoted in the magnetic field due to the magnetic moment of the unpaired electron, which has an associated charge and angular momentum. The electron thus has a magnetic moment, which can align with (parallel) or against (antiparallel) an applied

external magnetic field (Figure 13). The electron spin has a quantum number, *s,* with magnitude ½ and a direction (m_s = + ½ or m_s = - ½).[35] The orientation in an applied external field is associated with an energy difference, or the Zeeman effect, described by:

$$\Delta E = h\nu = g_e \mu_\beta B_0 \qquad \text{Eq. 17}$$

where g_e is the electron g-factor, 2.0023 for a free electron, μ_β is the Bohr magneton, and B_0 is the applied magnetic field.

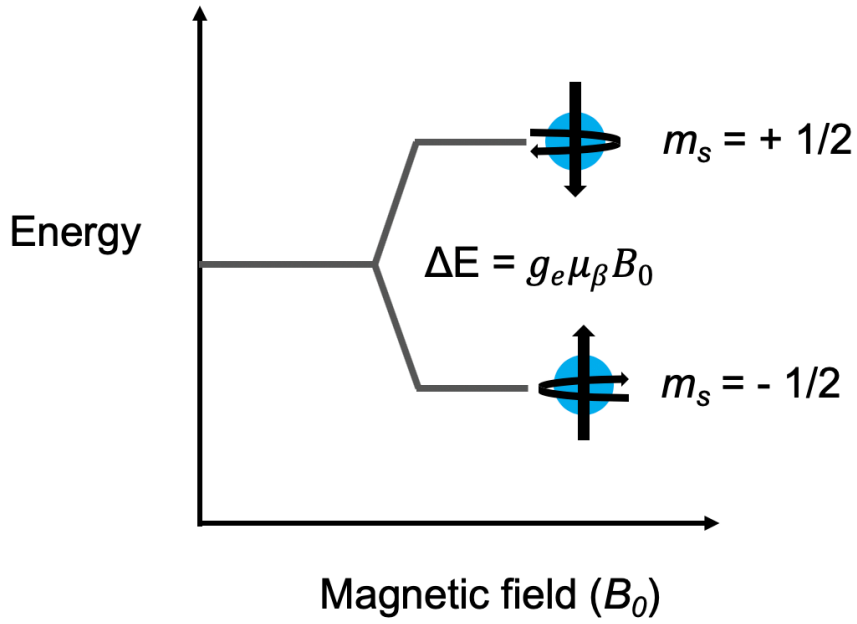


Figure 13. An illustration of the Zeeman effect. The magnetic moments can align with or against the applied external magnetic field, resulting in an energy difference. The spin aligned parallel to the external field has a lower energy than the spin aligned antiparallel to the applied field[35].

In order to promote the transition of lower-energy electrons to higher energy, microwave frequencies are used. In continuous wave EPR (cw-EPR), a fixed microwave frequency is applied and the external magnetic field is swept, or varied, resulting in energy absorption when the resonance condition or energy gap between the transitions is satisfied[36]. The spectra are given as first derivatives of the absorption peaks. The electron g-value can be calculated from:

$$g = \frac{h\nu}{\beta B_0} \quad \text{Eq. 18}$$

where h is Planck's constant (6.626 x 10^{-34} J·s), β is the Bohr magneton (9.274 x 10^{-28} J·G^{-1}), and B_0 is the applied field. The g-value can vary due to the electronic environment and is thus an important parameter to determine for a given molecule[36].

The g-value of an electron can change, resulting in environmental anisotropy. The g-value is split into its principal components g_x, g_y, and g_z, which give the spectra characteristic features. For isotropic samples, the principal components are equal ($g_x = g_y = g_z$). In the case $g_x = g_y < g_z$, or $g_x = g_y > g_z$, the spectra are axial. In the case where none of the principal components are equal ($g_x \neq g_y \neq g_z$), the spectra are rhombic.[37] The g-value and its principal components contain information about the electron environment, and can be diagnostic for spin systems within metalloproteins[36].

1.9 Electron Microscopy

Electron microscopy is a technique that has been used extensively in biology and material science to image samples at high resolution. Recently, cryo-transmission electron microscopy (cryo-TEM) has emerged as a robust method for solving protein structures to near-atomic resolutions rivaling X-ray crystallography[38].

Transmission electron microscopes operate analogously to light microscopes, but use electrons as an illumination source. The wavelength of an electron is dependent on the momentum, from the de Broglie relationship[39]:

$$\lambda = \frac{h}{p} \quad \text{Eq. 19}$$

Where h is Planck's constant (6.626 x 10^{-34} J·s), p is momentum, and λ is the wavelength. The momentum of the electron can be calculated based on its acceleration voltage, or velocity:

$$q = \frac{1}{2}mv^2 \qquad \text{Eq. 20}$$

Substituting the expressions yields an expression for the wavelength in terms of acceleration voltage (E, in volts):

$$\lambda = \frac{h}{\sqrt{2mE}} \qquad \text{Eq. 21}$$

The velocities of electrons in the microscope reach a significant fraction of the speed of light, so relativistic effects are also considered. The acceleration voltage E can be corrected to E*, including the velocity of light (c).

$$E^* = E\left(1 + \frac{e}{\sqrt{2mc^2}}E\right) \qquad \text{Eq. 21.1}$$

For high-end transmission electron microscopes, the acceleration voltage is 300 keV. Using a mass of 9.1 x 10^{-31} kg and 1.6 x 10^{-19} C for the elementary charge, e, gives a wavelength of 1.96 pm.

The scale of cellular structures is in the nanometer (10^{-9} m) regime, and protein substructure is in the Ångstrom (10^{-10} m) range. The sample resolutions attainable by electron illumination are not limited by the wavelength of electrons, but by sample stability and detectors[40].

Electrons interact strongly with matter[41] and can be scattered by air. Electron microscopes operate under high vacuum to reduce the non-productive scattering unrelated to the sample. In elastic X-ray scattering, X-rays do not interact with the nucleus, and instead scatter from

the electron cloud. Electron scattering does interact with the nucleus, as well as with the electron cloud, and can scatter to high angles (e.g. the technique high-angle annular-detector dark field imaging (HAADF) observes the electron-nuclear scattering)[41]. Additional concerns arising from electron sources include non-productive excitations that can add noise to the signal through inelastic scattering, or damage the sample. Valence electrons can be ejected as so-called secondary electrons, leading to ionization of the surface and damage of the sample (e.g. random bond breakage, radiolysis, etc.), affecting beam interactions during data collection (positive charge buildup from secondary electron emission can cause significant movement of the sample in the beam)[41]. To mitigate radiation damage, TEM is performed with vitrified biological samples (flash-frozen to liquid nitrogen temperatures, < -170 C) and data collection is done with a low dose rate (total dose ~40 $e^-/Å^2$)[38] on highly sensitive cameras (direct electron detectors).

One method of protein structure solution by cryo-electron microscopy is single particle analysis (SPA). In an SPA experiment, proteins are purified to homogeneity by chromatographic methods and flash-frozen with liquid ethane(/propane) in a frozen-hydrated state. Ice vitrification is critical, since crystalline ice can disrupt biological material and electrons are strongly scattered by cubic and hexagonal ice. Grids (typically containing a periodic array of holes) are used as sample support and are made hydrophilic through plasma cleaning to increase wettability. The sample is intended to be as thin as possible to increase the rate of cooling, so the sample volume is in the microliter (μL) range and blotted with filter paper. The specimen is blotted in a high-humidity chamber to prevent significant evaporation, and the sample becomes suspended stably in the grid holes by surface tension.

Liquid ethane and propane were discovered to vitrify water layers up to 1 μm thick[42], due to the high thermal conductivity of the cryogen. The rate of cooling to produce vitreous ice was determined to be ~10^5 K/s [43]. Liquid nitrogen (77 K) is unable to vitrify samples of appropriate thickness, as its thermal conductivity is ~400 K/s[42], whereas liquid ethane (90 K) is ~14 kK/s. Liquid nitrogen is a convenient storage coolant, as well as a primary coolant

for liquid ethane, but since the freezing point of ethane is higher than nitrogen, ethane will freeze over the course of the experiment. Jensen and authors[42] developed a formulation of 37% ethane and 63% propane that remains liquid in the presence of liquid nitrogen for TEM sample vitrification. Plunge-freezing is used as a rapid-freezing technique to capture a distribution of single particles in random orientations in the vitrified ice. Thin, vitrified ice is also critical for obtaining sufficient contrast of biomolecules, which solely originates in a phase shift, and additional stain is not commonly used.

An important breakthrough was the implementation of direct electron detectors, which afforded the use of low dose imaging modes to record data of low-contrast biological samples. The charge-coupled device (CCD) detectors used previously are relay devices that convert the electrons hitting the surface ("primary electrons") to photons with a scintillator that can be read by a sensor. In the relay, the signal becomes amplified to a pixel size physically larger than the original signal due to lateral scattering. The result is a loss of imaging resolution, as the signal is read by the CCD as multiple pixels instead of one[44]. Direct detectors have the substantial improvement of using thin layers that transmit the signal with minimal lateral scattering. For example, the Gatan K2 Summit direct detector has a 5 μm physical pixel size, producing 3838 x 3710 pixels for every frame and a full sensor readout of 400 frames/second. The detector is also capable of imaging in super-resolution mode, effectively splitting one pixel into four sub-pixels (generating a 7676 x 7420 pixel full frame), and assigning the centroid of an electron to one of the sub-pixels[45].

1.9.1 Single Particle Analysis (SPA)

Structures of biomolecules can be determined by single particle analysis (SPA), using samples in the vitrified, frozen-hydrated state randomly oriented in a thin layer of ice. Image formation is obtained from the interactions of the electron beam with the sample. Electrons can elastically and inelastically scatter. Elastic scattering is an event that has no transfer of energy, has a wide angular distribution, and contributes to high-resolution

structural information. Inelastic scattering transfers energy, has a narrow angular distribution, and does not contribute to useful signal. The model of electron beam interactions with the sample follows the phase object approximation, in which the electrons encountering the sample undergo a phase shift due to the Coulomb potential of the sample[39]. The elastic scattering contribution is described by:

$$\Phi(r) = \int_{-\infty}^{\infty} \Phi(r,z)dz \qquad \text{Eq. 22}$$

where the phase shift is $\Phi(r)$, the wave is travelling in the z direction, r is a two-dimensional vector, and $\Phi(r,z)$ is the three-dimensional Coulomb potential distribution in the object. The weak phase approximation is that the phase shift $\Phi(r)$ is very small (<< 1).[39] The incident wave, ψ, is related to the emitted wave:

$$\psi e^{i\Phi(x,y)} \qquad \text{Eq. 23}$$

Applying the weak phase approximation allows for the expansion:

$$\psi e^{\Phi(x,y)} = \psi(1 + i\Phi(x,y) + \cdots) \qquad \text{Eq. 24}$$

where the transmitted wave is composed of the unscattered first term, and a 90° phase-shifted scattered second term[39].

In the microscope, the imaging system can be schematically represented in a ray diagram as below:

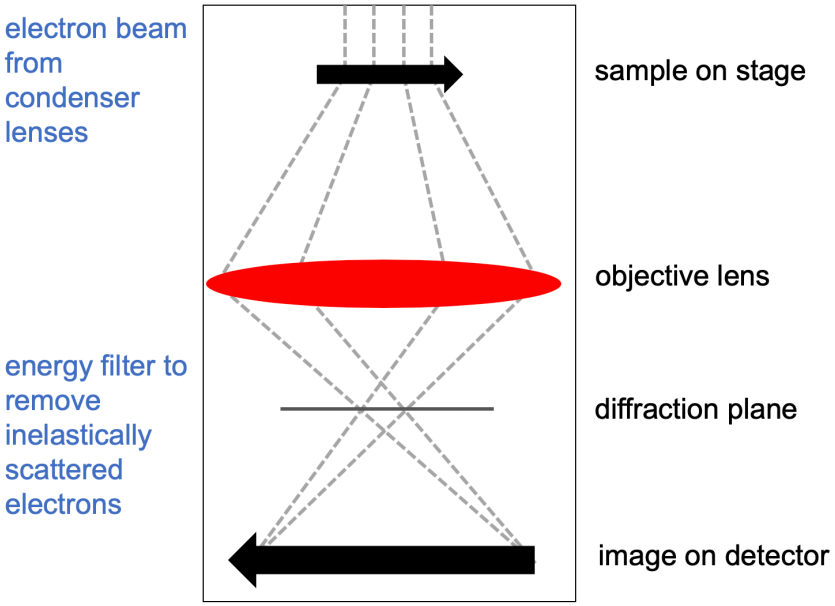


Figure 14. A general ray diagram of a microscope imaging system. Not pictured to the left: the electron source and the condenser lens system above the sample, and the energy filter before the detector.

As a property of lens systems, the image formed is also a convolution of the real space image and the point spread function:

$$I(r) = \Phi \otimes h(r) \qquad \text{Eq. 25}$$

where the symbol $\otimes$ denotes the convolution of the image with the point spread function, $h(r)$. The sample diffraction is physically limited by the circular aperture, and the effect from the Fourier transform of the scattered electrons results in a so-called Airy disc pattern on the image plane.[39] In an ideal lens system, a sample will scatter radiation that converges to single points on the diffraction plane, which, after inverse Fourier transform, recreate the image on the image plane. However, the point spread function convolution leads to an intensity distribution on the image plane instead of a single point.

$$F(k) = F\{I(r)\} \qquad \text{Eq. 26}$$

$$F(k) = F\{\Phi(r)A(k)2sin\gamma(k)\} \qquad \text{Eq. 27}$$

where $F\{\ \}$ denotes Fourier transform of the image. The $2\sin\gamma(k)$ term is the contrast transfer function. In Fourier space, the convolution between the image and the point spread function becomes multiplication. The Fourier transform of the point spread function is the contrast transfer function. The contrast transfer function (CTF) is a periodic function that oscillates with increasing frequency towards higher spatial frequencies, with the periodicity of the oscillations increasing with larger defocus. The effect of the CTF is an attenuation of the high-resolution features, and requires fitting and correction.[39]

The data collected on the image plane are 2D projections of randomly-oriented molecules on the sample plane. In order to reconstruct a 3D object from 2D projections, the projection theorem is applied. The projection theorem states that the projection of a 2D function ($f(x,y)$) can be obtained from the inverse two-dimensional Fourier transform of a central section through its 2D Fourier transform ($F(k_x, k_y) = F\{f(x,y)\}$). In SPA, 2D projections can be aligned in Fourier space along a common line[39]. Given the accurate assignment of the angles contributing to the 2D projections, the 3D Fourier space can be sampled by the aligned 2D projections and inverse-Fourier transformed to give the 3D structure in real space (Figure 15). The reconstruction is also dependent on an even distribution of unique views of the sample.

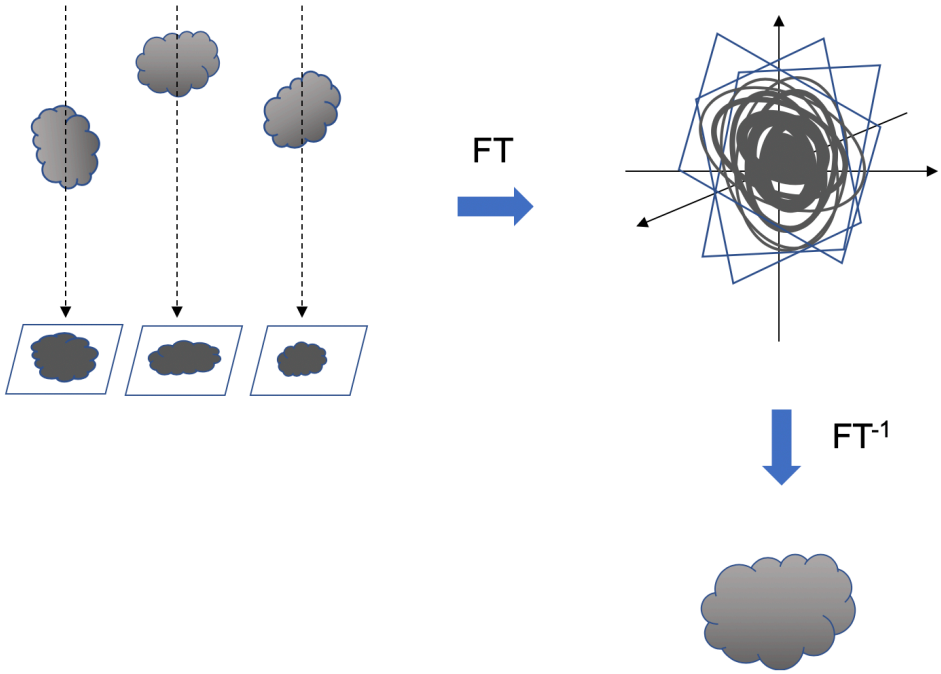


Figure 15. Schematic of the projection theorem applied to SPA. Particles in the electron beam generate 2D projections. The 2D projections are aligned to their common line in Fourier space, and a 3D-inverse Fourier transform yields the 3D structure in real space.

The 2D projections have a low signal-to-noise ratio (SNR), but the signal can be enhanced by averaging projections of the same view. The averaged views are binned into classes, and the angles of each projection are refined during reconstruction.

1.10 Significance of Study

The maturation role of the Fe-protein is mechanistically and structurally not understood. By exploring the functions of the Fe-protein in its maturation and nitrogenase roles, we are interested in the mechanism of discrimination between the Fe-protein functions. A central question is how one enzyme can catalyze disparate reactions by interacting with binding partner proteins that are topologically similar. Additionally, the mechanistic details of how the redox state at the [4Fe:4S] cluster is influenced by a bound nucleotide is not clear. The project goal is to address the following key questions regarding the mechanism of

metallocluster biosynthesis: (1) what are the oxidation state changes in the Fe-protein [4Fe:4S] cluster as a function of nucleotide state, and (2) is the binding interface between NifEN and Fe-protein similar to that of MoFe and Fe-protein during substrate turnover? These studies will help to understand the roles of the Fe-protein in substrate reduction and in cofactor biosynthesis, and ultimately provide insight into the overall organization of the nitrogenase-associated proteins in the cell.

In a broader context, the proposed research will address fundamental questions about a multifunctional catalyst in substrate reduction and the assembly of the two unique metalloclusters responsible for nitrogen fixation. Inorganic catalysts capable of converting nitrogen to ammonia at physiological levels have not been synthesized thus far. Studying the assembly of the FeMo-co will lead to insights on how to synthetically recapitulate N_2 and CO reduction to form ammonia and hydrocarbons, respectively.

C h a p t e r 2

METHODS

2.1 Growth and Isolation of *Azotobacter vinelandii*

Nitrogenase is purified from wild-type *Azotobacter vinelandii*, cultured in Burke medium (Table 2).

Sucrose		20	g/L
$FeSO_4 \cdot 7H_2O$		0.2	mM
$Na_2MoO_4 \cdot 2H_2O$		3	μM
$MgSO_4$		1.67	mM
$CaCl_2$		0.9	mM
KH_2PO_4/K_2HPO_4 pH 7.5		10	mM
NH_4Cl	Starter culture	10.5	mM
	Fermenter culture	1.3	mM

Table 2. Burke's minimal medium components and final concentrations.

A 50 mL starter culture is prepared using Burke's minimal media and inoculated with 1 mL of frozen culture (50% v/v glycerol). The culture is incubated at 30°C, shaking at 180 rpm. At an OD_{600} between 1 and 2, 1% inoculum is used to scale up a 500 mL culture of the same medium composition. The scaled culture is used to inoculate 60 L of medium (1% inoculum) in a fermenter at an OD_{600} of 1-2. The 60 L of growth medium is harvested at an OD_{600} of 0.8-1 to maximize Fe-protein yield. A typical 60 L growth yields ~80 g of *A. vinelandii* cells, which have a rich brown color indicating the presence of iron-sulfur proteins. The cell pellets are frozen and stored at liquid nitrogen temperatures. Recently, we have begun using a 10 L fermenter with proportionally similar yields.

The purification buffers after lysis require strict anaerobic preparation. Buffers are made anaerobic or degassed using standard Schlenk line technique. With nitrogenase, argon (Ar) gas is used, since nitrogen is a substrate. For purification, cells are transferred to an anaerobic atmosphere and resuspended in anaerobic lysis buffer (50 mM Tris/Cl pH 7.5, 150 mM NaCl, 5 mM dithionite). Cells are lysed using an Emulsiflex C5 (pressurized with Ar gas) and centrifuged to removed cell debris. To protect the lysate from the high temperatures generated by lysis, the lysate is kept on ice. The *A. vinelandii* lysate is dark brown to black in color, indicating the presence of iron-sulfur proteins.

Chromatography is performed using an Äkta FPLC system. The supernatant is loaded onto a HiTrap Q HP anion exchange column and eluted using a linear gradient from 0-1 M NaCl over 10 column volumes. The MoFe-protein elutes around ~300 mM NaCl, and the Fe-protein shortly thereafter at ~450 mM NaCl. The proteins are collected by hand into flasks vacuum/Ar-purged (12 cycles, 4 mins vacuum, 30 sec Ar), and sealed with rubber septa. Separated proteins are further purified using size exclusion chromatography on a Superdex 200 equilibrated in anaerobic 50 mM Tris/Cl pH 7.5, 200 mM NaCl, 5 mM dithionite. Representative elution profiles from a purification on 22-Dec-14 are shown below (Figure 16).

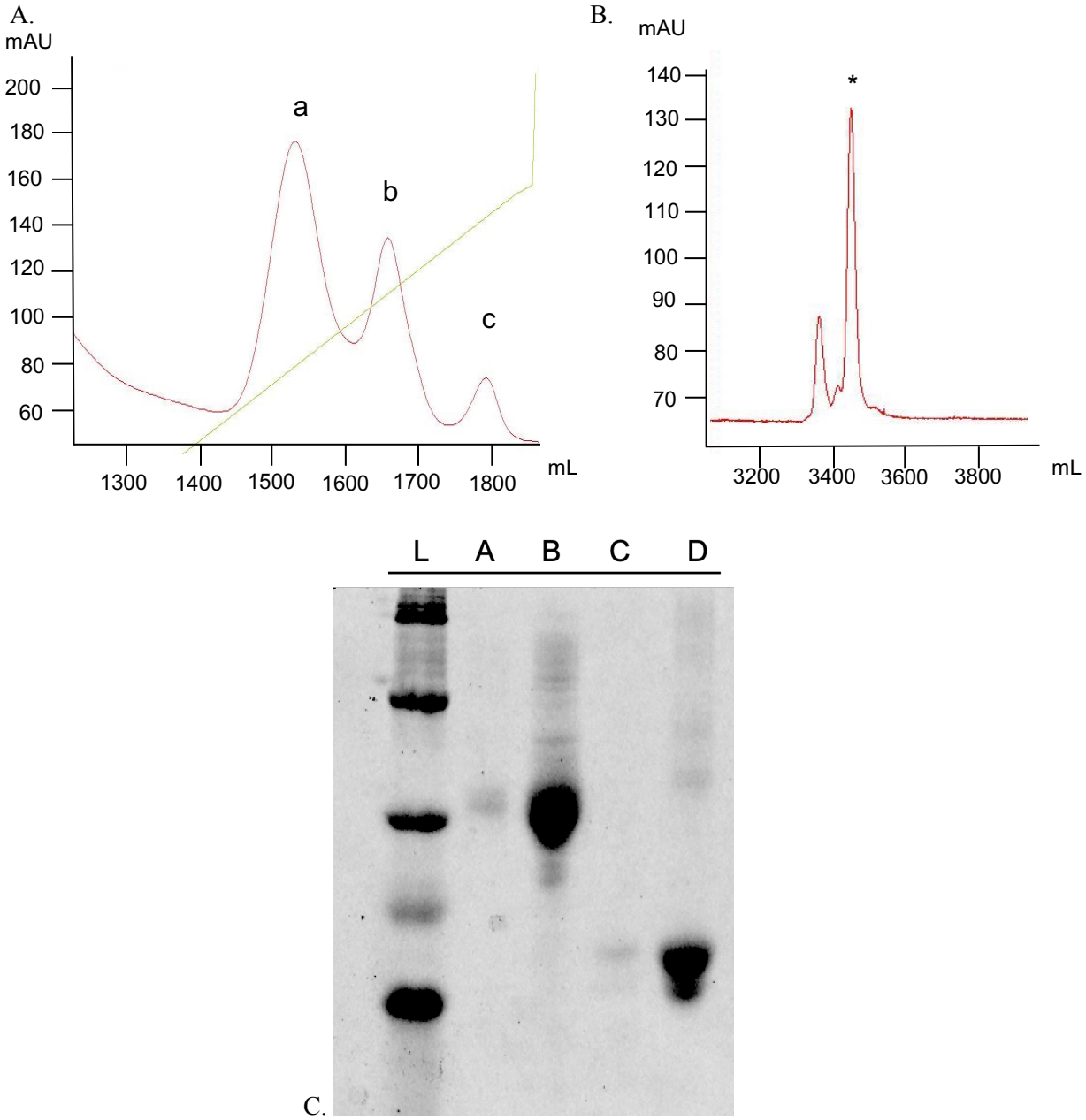


Figure 16. (A) Ion exchange elution profile: a) MoFe-protein, b) Fe-protein, c) flavodoxin (NifF). (B) Size-exclusion profile for MoFe-protein purification. The main peak is indicated with an asterisk, and pure protein is collected by hand for only the symmetric peak. (C) Representative gel is shown, with ladder (**L**), unconcentrated MoFe-protein (**A**), concentrated MoFe-protein (**B**), unconcentrated Fe-protein (**C**) and concentrated Fe-protein (**D**).

After elution, the purified proteins are concentrated using an over-pressure concentration unit in the anaerobic chamber (Amicon), with filter sizes to match the molecular weight of

the protein of interest. MoFe-protein (~230 kDa) is concentrated using a 100 kDa molecular weight cutoff (Millipore) using 5 bar of Ar. Fe-protein (~60 kDa) is concentrated using a 30 kDa molecular weight cutoff (Millipore). The purify of the final product is confirmed using gel electrophoresis (Figure 16). The protein concentrations are determined by UV-visible spectroscopy based on the absorbance at 410 by the iron-sulfur clusters. The extinction coefficient of MoFe-protein is 76 mM^{-1} cm^{-1} and 9.4 mM^{-1} cm^{-1} for Fe-protein. The product activity can be determined using enzymatic assays.

2.2 Acetylene Reduction Assay

Nitrogenase can catalyze the reduction of acetylene to ethylene as shown below (Eq. 30). The production of ethylene has been used to test and standardize the activity of nitrogenase preparations and measure certain kinetic parameters.

$$C_2H_2 + 2e^- + 2H^+ + 4ATP \rightarrow C_2H_4 + 4ADP + 4P_i \qquad \text{Eq. 28}$$

ADP is a potent inhibitor of nitrogenase, so to prevent the accumulation of ADP, the assays include an ATP regeneration system, made up of phosphocreatine kinase (PCK) and phosphocreatine salt (PC), which carries out the reaction:

$$PCK + PC + ADP \rightarrow PCK + C + ATP \qquad \text{Eq. 29}$$

In the assay, there are 20 units of PCK enzyme, 20 μmoles of PC, and 5 μmoles ATP. The four-fold excess of PC ensures each mole of ATP is regenerated four times. Dithionite is treated as a two electron donor, such that 1 μmole of dithionite reduces 1 μmole of acetylene. Using a generous assumption of 5 ATP/2 e^-, the overall reaction is: (Eq. 32)

$$25\ \mu mol\ ATP + 25\ \mu mol\ Na_2S_2O_4 + \textit{excess}\ C_2H_2 \rightarrow$$
$$5\ \mu mol\ C_2H_4 + 25\ \mu mol\ ADPP_i + 15\ \mu mol\ Na_2S_2O_4 + \textit{excess}\ C_2H_2 \qquad \text{Eq. 30}$$

In this treatment, the theoretical maximum is 5 μmol ethylene (C_2H_4) produced, with ATP as the limiting component, and not dithionite or acetylene. It is useful to note that in our reaction conditions, 5 mM is equivalent to 5 μmole per 1 mL reaction volume.

Reduction assay mixtures are prepared using protein in 50 mM Tris/Cl pH 7.5, 200 mM NaCl. Components of the standard assay mixture include: 50 mM Tris/Cl, pH 7.5, 5 mM ATP, 5 mM $MgCl_2$, 20 mM phosphocreatine, 0.1 mg/mL phosphocreatine kinase, and 20 mM sodium dithionite. 0.11 mg of MoFe-protein and 0.12 mg Fe-protein are used per reaction vial, at a component ration of 2, according to the equation below (Eq. 33).

$$CR = 1.82\ (C_{Fe}/C_{MoFe}) \qquad \text{Eq. 31}$$

where C is concentration in mg/mL, and component ratio (CR) describes the moles of Fe-protein active sites per mole MoFe-protein active site. Commonly, a component ratio of 2 is used for inhibition studies of nitrogenase, but the ratio can be varied to measure certain kinetic parameters or affect activity.

After vacuum/Ar purging of 1 mL assay mixtures (1.5 min vacuum, 30 s Ar), dithionite is added. Each vial is overpressured with Ar to preserve an anaerobic environment. 1 mL of headspace is removed, and 1 mL of acetylene gas is injected. Vials with acetylene gas are incubated in a heated water bath at 30°C, shaking at 180 rpm. In a standard assay, protein is added and allowed to react for 10 mins and quenched with 1 mL of 3 M citric acid. For data collection, 50 μL of headspace is injected onto a gas chromatograph using an activated alumina column (60/80 mesh) and a flame ionization detector. The column oven is operated at 110° C. A sample GC chromatogram is shown below from assays run on 9-Jan-15 (Figure 17).

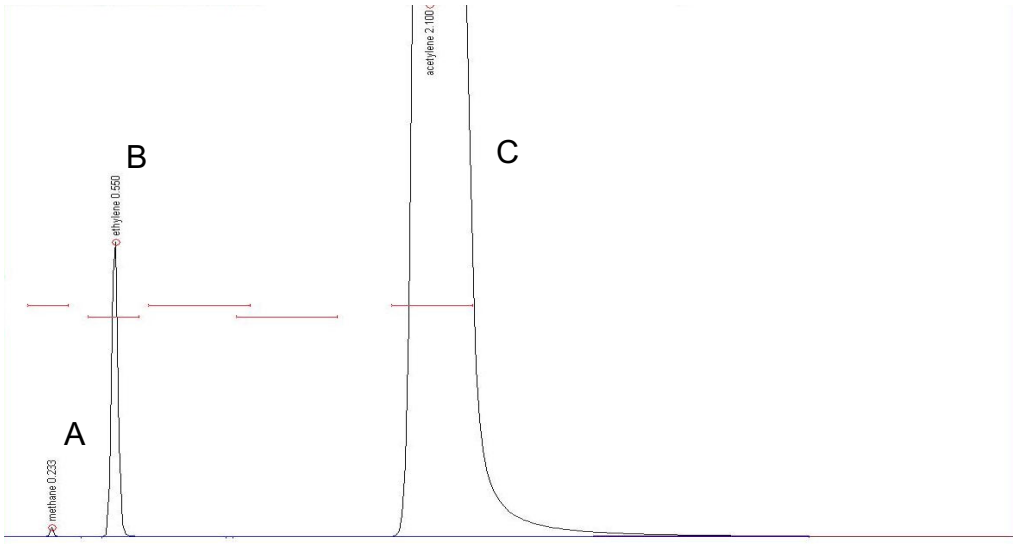


Figure 17. Sample GC chromatogram using the flame-ionization detector and helium carrier gas. The peaks are: (A) methane, (B) ethylene, and (C) acetylene.

The gas chromatograms are integrated using the Peak Simple software (SRI Instruments) for conversion to moles of product using the standard curve. The gases produced in the headspace can be quantified using standard curves of pure acetylene (Figure 18). To measure a calibration curve, a standard volume of acetylene gas is prepared by flushing a 1L round-bottom flask to 1 atm. Blank assay vials containing buffer but no protein components are prepared in parallel to experimental samples. Syringes are used to inject 20, 30, 40, and 50 µL of acetylene into the GC, and the integrated peak areas are recorded. Assuming 1 atm of pressure in the flask, the moles of acetylene in each injection are calculated using the ideal gas law. The curve is fit to an equation to relate the integrated peak area to nmoles of gas. In an acetylene reduction experiment using purified protein, the ethylene produced can be quantitated using the acetylene calibration curve. The nmoles of ethylene are calculated per injection volume (typically 50 µL), and total nmoles are calculated by a ratio to the total headspace. The total nmoles of product are related to the specific activity based on the milligrams of protein used in the experiment, yielding activities in units of nmoles of ethylene min^{-1} mg^{-1} protein.

The activity of the MoFe- and Fe-proteins can be examined through titrations of one of the components against the other. The activity observed of the nitrogenase protein complex varies as a function of a variety of variables, including the component ratio, method of reduction and substrate reduced. For a typical protein preparation, the acetylene reduction activity of MoFe-protein is roughly 2000 nmol ethylene min^{-1} mg^{-1} and the activity of Fe-protein is around 1900 nmol ethylene min^{-1} mg^{-1}. The highest activities observed for both proteins were around 2700 nmol ethylene min^{-1} mg^{-1}. The activity of both proteins is, unsurprisingly, oxygen sensitive, and the Fe-protein is observed to a have a much lower tolerance of trace oxygen compared to the MoFe-protein. When performing full titrations as shown in Figures 19 and 20, it was necessary to continuously maintain argon gas over-pressure in the sample vials and carefully inspect Hamilton syringes for damage (which results in exposure to atmosphere). The nitrogenase proteins are not particularly heat labile, and purifications are performed at room temperature, but lowering the temperature does improve the stability of the Fe-protein, especially during the course of long titrations.

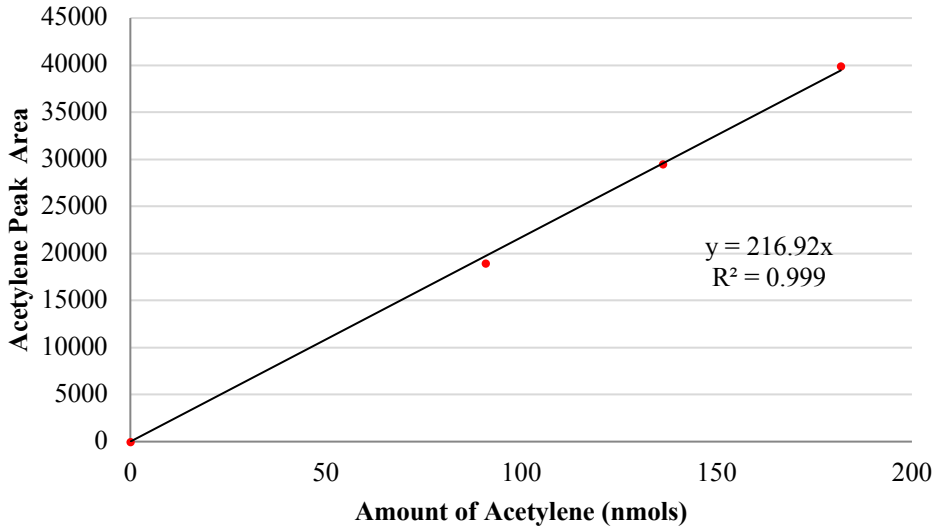


Figure 18. Representative calibration curve to relate integrated peak area to nmoles of acetylene.

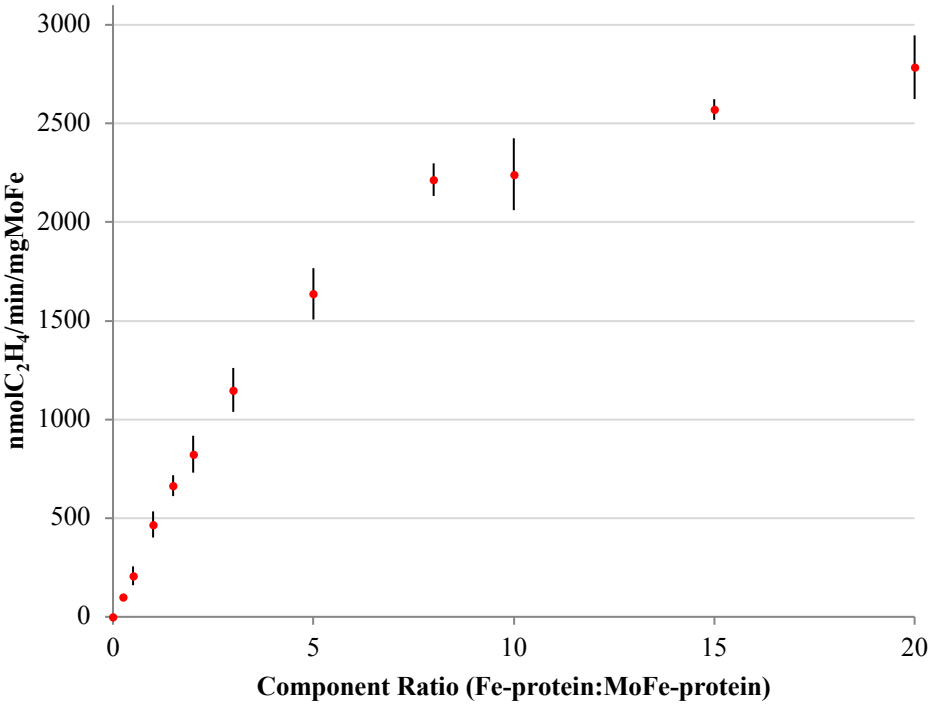


Figure 19. MoFe-protein by Fe-protein titration with highly active protein.

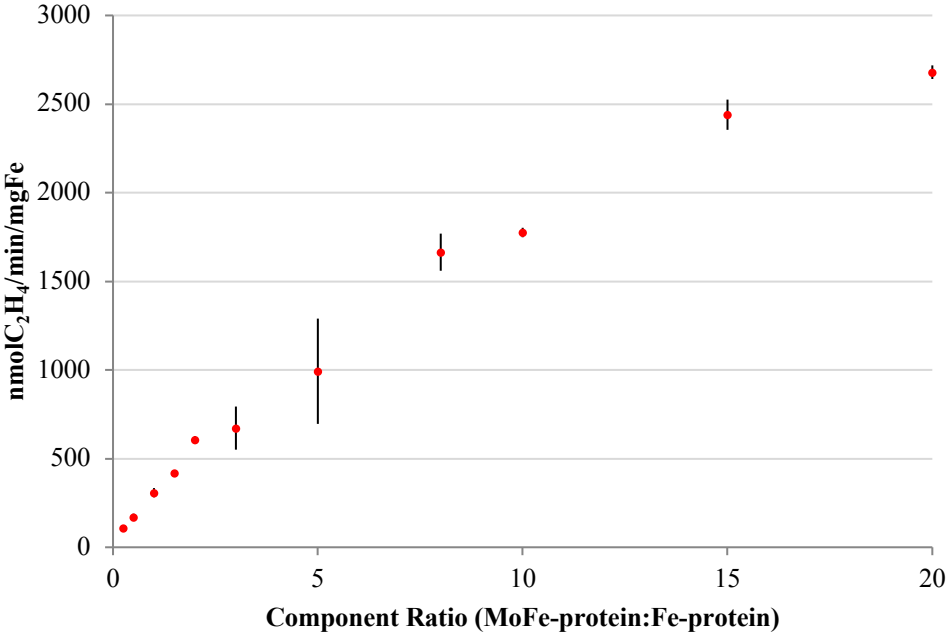


Figure 20. Fe-protein by MoFe-protein titration with highly active protein.

The activity of nitrogenase against multiple substrates is of continuing interest to the Rees group. While acetylene is the most prevalent substrate to assess and benchmark protein activity, we can detect the products of hydrogen gas (from protons) and ammonia (from nitrogen). Additionally, the activity can be assayed in a combinatorial fashion by measuring two products formed from the reaction. The most common secondary product is hydrogen formation, which is detected in the wild-type reduction of nitrogen. However, hydrogen reduction is not detected in a background of acetylene reduction, leading to intriguing questions about substrate discrimination.

Nitrogenase activity is also dependent on the type of reductant used. Most commonly, the reductant dithionite is used in the purifications to maintain anaerobicity. In activity assays, dithionite molecules ($S_2O_4^{2-}$) reduce the Fe-protein to the $[4Fe:4S]^{1+}$ state, which support the reduction of acetylene, protons and nitrogen. The reductant Ti(III)-citrate is a more potent reductant (the midpoint potential of dithionite is around -0.66 V vs NHE at pH 7, and the midpoint potential of Ti(III)-citrate is less than -0.8 V at pH 7[1]) and supports reduction of the Fe-protein to the $[4Fe:4S]^0$ state. The use of Ti(III)-citrate reductant also supports a model of twice-as-efficient ATP usage in the nitrogenase catalytic cycle: Ti(III)-citrate has been measured to use a ratio of $ATP/2e^- = 2$, compared to dithionite, which uses a ratio $ATP/2e^- = 4$[46]. In endpoint measurements, Ti(III)-citrate and dithionite produce similar reduction activities against all substrates tested (Figure 21).

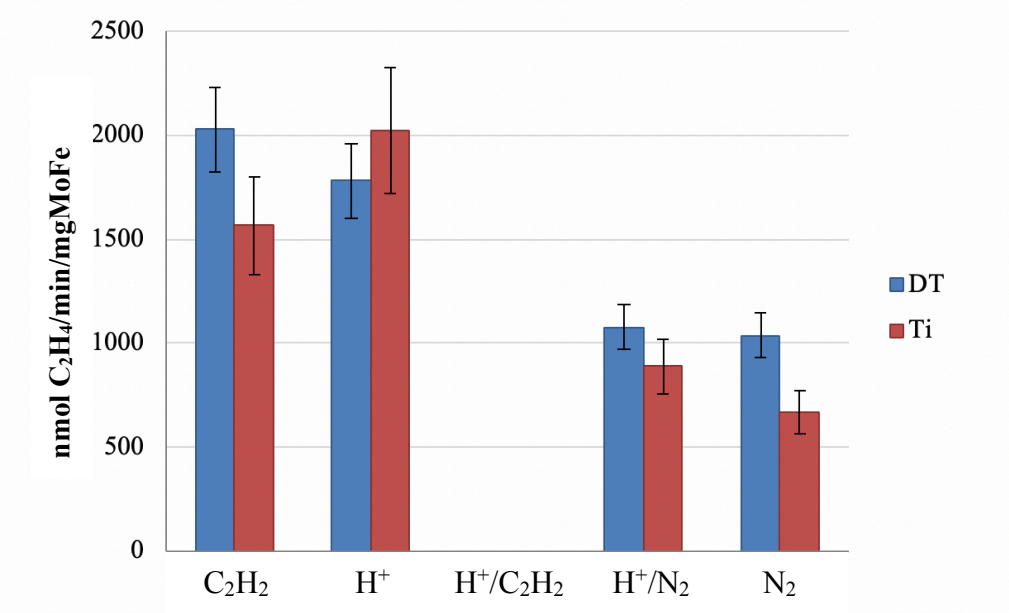


Figure 21. Activities of MoFe-protein at a component ratio of 2 with various substrates using dithionite reductant (blue) or Ti(III) citrate (red).

2.3 Inhibition of Nitrogenase

Examining the inhibition properties of nitrogenase is an important tool in discriminating the mechanism of substrate reduction. For example, nitrogenase is inhibited by the presence of salts, which disrupt the interaction between the MoFe- and Fe-protein (Figure 22).

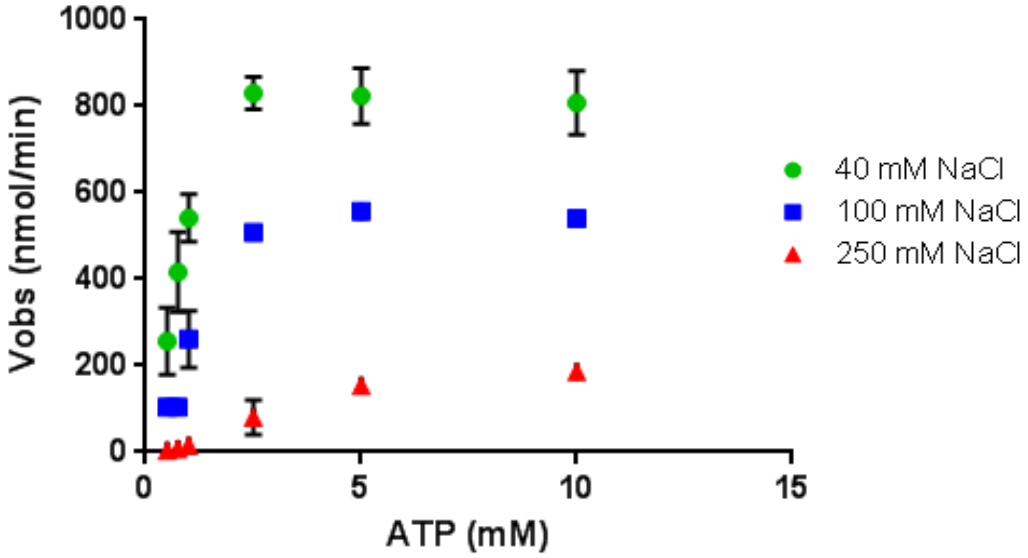


Figure 22. Inhibition assay of nitrogenase in the presence of fixed concentrations of sodium chloride, titrating ATP.

The NaCl inhibition assays present a mixed inhibition mode; the V_{max} of the reaction drops with increasing NaCl concentration, and the K_M shifts. The salt inhibition behavior has an intuitive effect based on our knowledge of nitrogenase complex information. A similar analysis can be performed with Fe-protein in the context of maturation activity. Nitrogenase activity is also inhibited by small molecules that bind at the active site, and ADP (from binding to the Fe-protein). In order to probe the interactions between Fe-protein and molybdate, a hypothesized interaction based on maturation assays, we assayed the inhibition properties. We observed inhibition by molybdate, but the effect is likely from the formation of polyanions due to reduction by dithionite in the reaction medium (Figure 23). We assayed the inhibition properties of molybdate at fixed ATP concentrations, with and with an ATP-regeneration system (ARS).

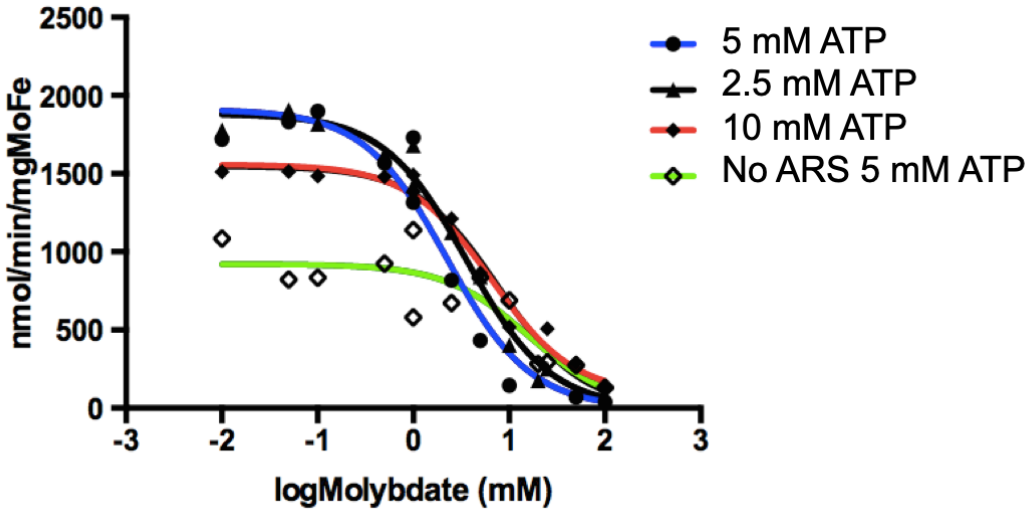


Figure 23. Inhibition assay of nitrogenase in the presence of increasing concentrations of molybdate, with fixed ATP concentration. The fit IC50s are 2.2 (5 mM ATP, blue), 3.6 (2.5 mM ATP, black), 6.6 (10 mM ATP, red), 16 (in absence of ATP regeneration system, 5 mM ATP, green) ±1 mM molybdate.

Evidence of polyanions formation in the presence of dithionite was detected due to a color change of the solution (red/pink) and we subsequently crystallized poly-oxo-metallate clusters (Figure 23B) containing primarily molybdenum and phosphate.

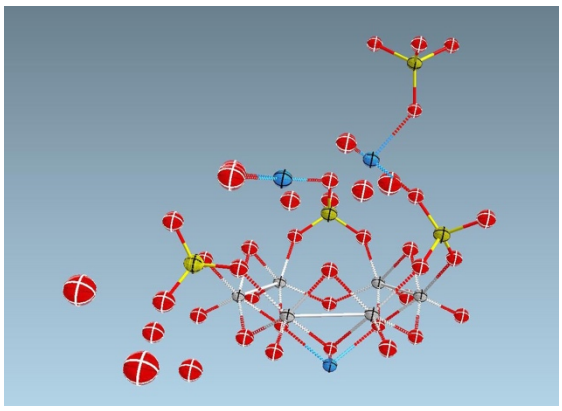
A.  B. 

Figure 23B. (A) Crystals obtained using molybdate-inhibited assay mixture. (B) Polyoxometalate structure. Molybdenum is colored in gray, oxygen in red, sulfur in yellow and magnesium in blue.

Further support of dithionite interference with molybdate was confirmed through cyclic voltammetry, performed in collaboration with Dr. Helen Segal. Molybdate reduction was tested using a glassy carbon electrode with an Ag/AgCl reference (Figure 24). Reduction

was tested in activity assay buffer (without MoFe/Fe-protein present) and in buffer with dithionite. The presence of dithionite led to catalysis at cathodic potentials, indicating a cascade of dithionite reducing molybdate. These experiments provide evidence that using dithionite with molybdate to study molybdenum insertion is interfering with the assays. A better reductant will need to be used to study the process, ideally from the natural reductant (reduced flavodoxin or ferredoxin), or Ti(III)-citrate.

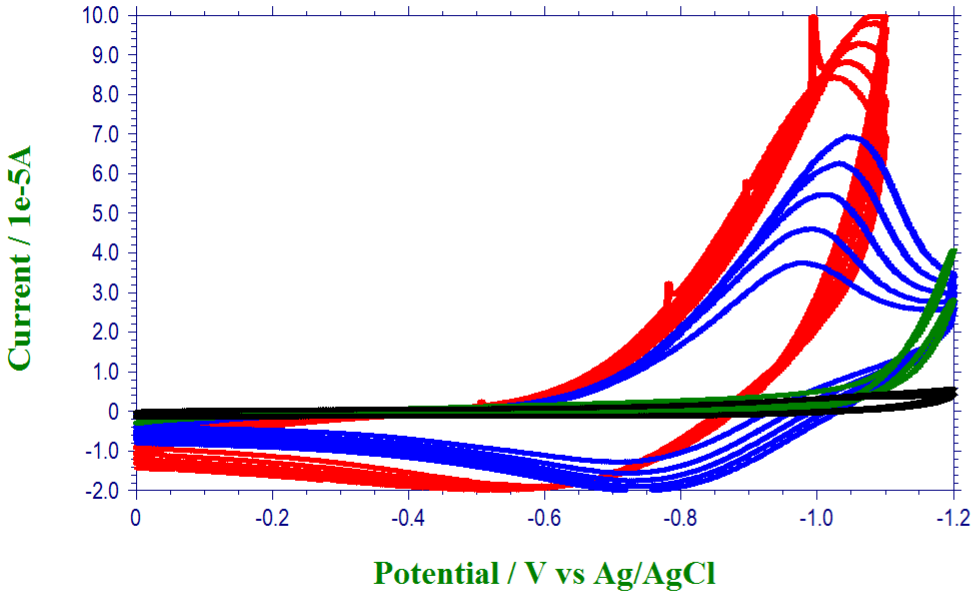


Figure 24. Electrochemistry experiments testing dithionite reduction of molybdate. Black is assay buffer, blue is 10 mM molybdate in assay buffer, green is assay buffer with 25 mM dithionite, and red is 10 mM molybdate in assay buffer with 25 mM dithionite.

2.4 Anaerobic Grid Preparation

For negative stain grid preparation, Ultrathin C Film on holey carbon support film, 400 Cu mesh grids are used. Grids are glow-discharged/plasma cleaned using a Pelco EasiGlow glow discharge cleaning system at 0.19 mBar, 15 mA for 1 min (N.B. glow-discharged grids should be used within an hour of plasma cleaning). The grids are cycled into the anaerobic chamber (97% argon, 5% hydrogen gas, Coy Labs) using 20 manual cycles of vacuum and argon for 30 seconds each.

If using frozen protein, the sample is centrifuged at 14,000 rpm for 1.5 minutes (Eppendorf) to remove precipitate accumulated from the freeze/thawing cycle. 3 μL sample is applied to the glow-discharged face of the grid, and incubated for 45 seconds. The sample is then blotted from the edge of the grid using No. 1 Whatman filter paper. 3 μL of 1% uranyl acetate (prepared in water, centrifuged before use, and stored in a light-blocking container) is applied to the grid face, and incubated for 45 seconds. (N.B. uranyl acetate degrades over time and should be replaced every couple months.) Stain is blotted in the same manner described above, and the stain is allowed to completely dry in the anaerobic chamber before imaging. Additional wash steps using water before applying uranyl acetate stain can improve grid quality.

For preparing anaerobic grids for cryo-imaging, we modified an anaerobic chamber to accommodate use of flammable cryogens. In our set-up, ports were installed in the chamber that contained heavy rubber stoppers with copper tubing threaded through the center. One-way valves (Swagelok) were installed on either end for a gas-tight seal. The copper tubing is attached to a snorkel exhaust line, which is connected to a chemical-resistant vacuum pump (Welch). The vacuum pump is vented into the hood and run throughout grid preparation to remove the buildup of flammable ethane/propane gas. To accommodate the flammable gas storage in our laboratory, it is necessary to condense liquid ethane/propane (37% ethane/63% propane, Airgas) outside of the anaerobic chamber in a tube (Falcon) submerged in a dewar of liquid nitrogen, and brought into the anaerobic chamber using the

short cycle vacuum/argon purging. The condensed ethane/propane mix can be decanted into the Vitrobot cup (FEI ThermoFisher) that has been pre-cooled using liquid nitrogen and the heat bridge connected to the ethane pot. The anaerobic chamber is large enough to accommodate the Mark IV Vitrobot (FEI ThermoFisher). Degassed/argon purged water is used for the humidifier, and No. 1 Whatman filter paper is stored inside the chamber to maintain an oxygen-free blotting surface. Proteins are diluted with buffers containing 2-5 mM dithionite inside the anaerobic chamber. To minimize transfer ice contamination, the anaerobic chamber is kept at <20% humidity using calcium chloride desiccant.

The results of extensive negative stain testing and ice optimization from cryo-imaging studies are shown below (Figure 25). Through optimization of sample preparation and freezing, along with modifying grid conditions, uniform distributions of proteins could be obtained reliably.

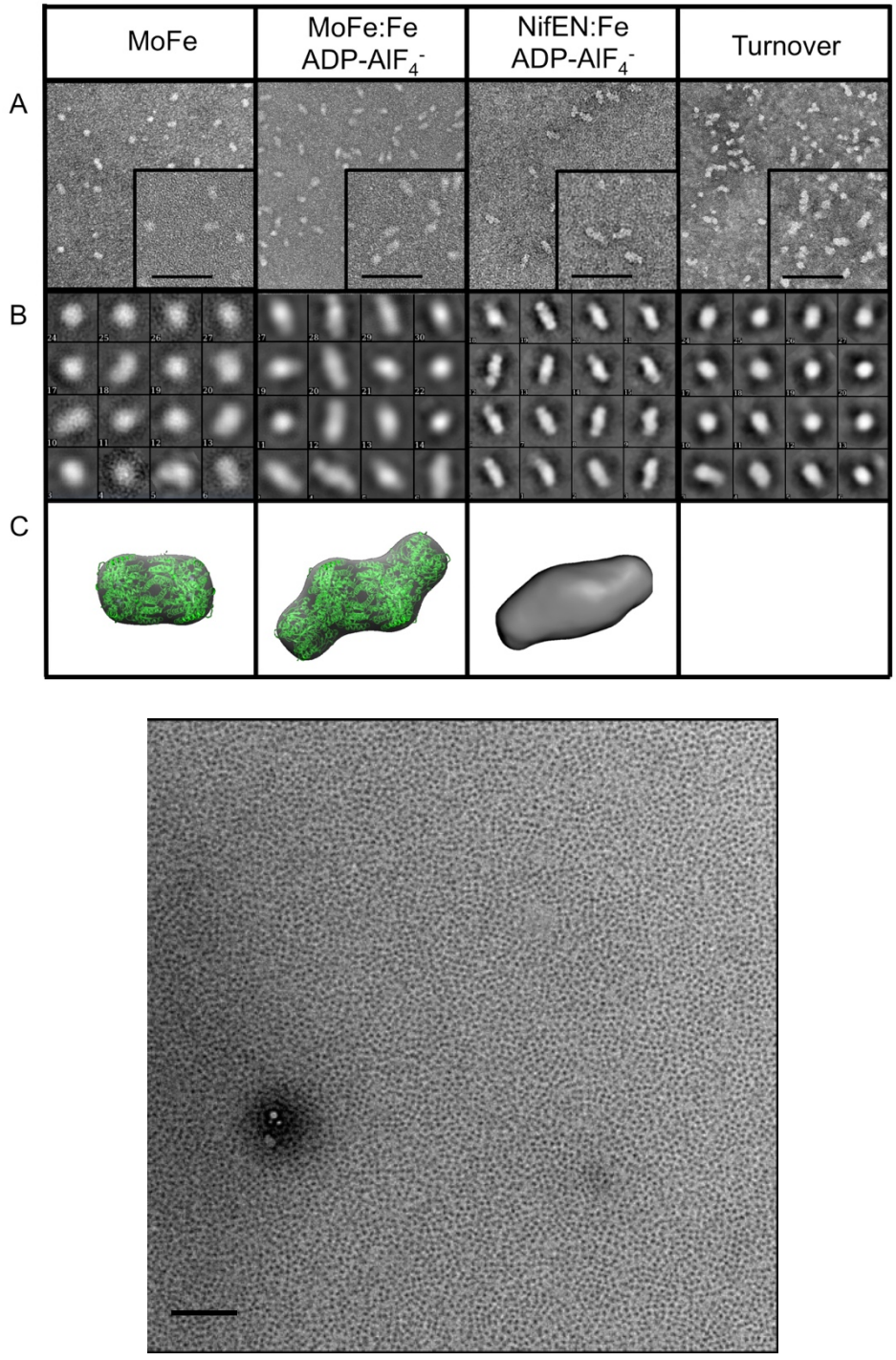


Figure 25. (Top) Overview figure of negative stain data. In panel A, the negative stain grid is shown, with a scale bar at 50 nm. Panel B shows a subset of the class averages (14 nm x 14 nm), and C shows initial reconstructions of MoFe-protein (NifDK), as well as the ADP-AlF_4 inhibited complexes of MoFe-Fe protein (NifDKH) and NifENH. The last column contains preliminary

negative stain data of an assay mixture under active turnover conditions, in which we would expect a mixture of free and Fe-protein bound MoFe-protein. (Bottom) Test cryo-EM sample of MoFe-protein at 0.1 mg/mL at 30,000x, scale bar represents 100 nm. Particles are overcrowded at this concentration; a series of samples from 0.1 to 0.02 mg/mL have a more suitable distribution.

The negative stain screening of samples was important to identify heterogeneous particles for cryo-imaging. However, given an intentionally heterogeneous samples, it is possible to make observations about the population distribution of protein complexes. We collected preliminary negative-stain micrographs (Figure 25) of an assay mixture of MoFe-protein Fe-protein in the presence of ATP and reductant (the ATP regeneration system was omitted to reduce background proteins on the grid). Ideally, classification of the particles can be related to the population of bound versus free MoFe-proteins during turnover conditions. Such an approach requires further optimization and biochemical evidence to justify the ensembles captured in the negative-stain sample.

C h a p t e r 3

SITE-SPECIFIC OXIDATION STATE ASSIGNMENTS OF THE IRONS IN THE [4FE:4S]$^{2+/1+/0}$ STATES OF THE NITROGENASE FE-PROTEIN

Wenke B.B.*, Spatzal T.S.*, Rees D.C. (2019) *Angewandte Chemie*.

3.1 Abstract

The nitrogenase iron protein (Fe-protein) serves as the electron donor for biological nitrogen fixation. Understanding how the Fe-protein controls electron transfer to the active site is critical in addressing the mechanism of substrate reduction. The Fe-protein contains an unusual [4Fe:4S] iron-sulfur cluster that is stable in three oxidation states: 2+, 1+, and 0. Here, we combine structural and spectroscopic techniques, including spatially-resolved anomalous dispersion refinement (SpReAD), to report oxidation assignments for individual irons in the cluster for each overall state. Additionally, we report the 1.13-Å resolution structure for the Fe-protein with bound ADP, the highest resolution Fe-protein structure presently determined. In the dithionite-reduced [4Fe:4S]$^{1+}$ state, the SpReAD analysis supports the oxidation state assignment of a delocalized $Fe^{2.5+}$ pair and a reduced Fe^{2+} pair. Our work identifies the $Fe^{2.5+}$ pair as coordinated by the solvent exposed Cys97, while the Fe^{2+} pair faces the protein interior and is coordinated by Cys132. It is proposed that binding of ATP to the Fe-protein promotes an internal redox rearrangement such that the solvent-exposed Fe becomes reduced, thereby facilitating electron transfer to the nitrogenase molybdenum iron-protein. The SpReAD analysis supports a uniform oxidation state assignment of Fe^{2+} for all irons in the titanium citrate-reduced [4Fe:4S]^{0} state, while all irons in the IDS oxidized [4Fe:4S]$^{2+}$ state are assigned to the valence delocalized $Fe^{2.5+}$ state.

3.2 Introduction

Biological nitrogen fixation is a multi-electron redox process carried out by nitrogenase[9,47,48], a two-protein enzyme system comprised of the iron- and molybdenum iron-proteins (Fe-protein and MoFe-protein, respectively). In the catalytic cycle, electrons originating from an electron donor (ferredoxin or flavodoxin *in vivo* or dithionite *in vitro*) are passed to the Fe-protein, which subsequently mediates the ATP-dependent reduction of the MoFe-protein. The MoFe-protein contains two unique metallocofactors: an intermediate [8Fe:7S] P-cluster and the active-site [Mo:7Fe:9S:C]-*R*-homocitrate FeMo-cofactor. Substrate reduction is accomplished through multiple binding and dissociation events between the MoFe-protein and Fe-protein, with each cycle involving electron transfer from the Fe-protein to the MoFe-protein, ultimately reducing substrate bound to the FeMo-cofactor. The timing of the electron transfers between the constituent clusters in the overall substrate reduction mechanism is an active area of investigation.

The Fe-protein is the only electron donor currently known to efficiently support N_2 reduction by nitrogenase. As a homodimer, the Fe-protein contains a single [4Fe:4S] cluster coordinated to the side chains of Cys97 and Cys132 of each subunit[29,49]. The cluster has been characterized in three different oxidation states: the oxidized $[4Fe:4S]^{2+}$ and dithionite-reduced $[4Fe:4S]^{1+}$ forms identified in early work on nitrogenase[47], and the subsequently characterized all-ferrous $[4Fe:4S]^{0}$ form[5,50]. It is generally considered that electron transfer from the Fe-protein to the MoFe-protein cycles through the 2+ and 1+ states, although a role for the all-ferrous form has been proposed.[5,18,46,51] Changes in the Fe-protein [4Fe:4S] cluster environment can be monitored spectroscopically[52–54] and through the accessibility of the cluster to chelation[55–57].

Mössbauer and EPR spectroscopies have been particularly valuable in deciphering the oxidation state description of [4Fe:4S] clusters, including the Fe-protein[20,58–60]. From the isomer shifts observed in Mössbauer spectra, the Fe in [4Fe:4S] clusters may be assigned as either valence localized sites (Fe^{2+} or Fe^{3+}) or as valence delocalized $Fe^{2.5+}Fe^{2.5+}$ pairs[59,60]. In the case of the Fe-protein, the S = 1/2 spin state of the $[4Fe:4S]^{1+}$ form

consists of two types of Fe sites: a delocalized $Fe^{2.5+}Fe^{2.5+}$ pair and a pair of Fe^{2+} irons (the overall cluster charge of +1 reflects the sum of these Fe oxidation states together with the four sulfides (S^{2-})). In contrast, only one type of Fe site occurs in each of the other two cluster forms: the S = 0 state of $[4Fe:4S]^{2+}$ clusters contains delocalized $Fe^{2.5+}$ pairs[20], while in the S = 4 state of the $[4Fe:4S]^{0}$ cluster, the iron sites are all Fe^{2+}[22]. Additional complexities in the electronic/magnetic structure arise from the variable spin coupling between the Fe. For example, the $[4Fe:4S]^{1+}$ cluster exists as a mixture of S = 1/2 and S = 3/2 spin states with the precise distribution dependent on the solvent and nucleotide conditions[20].

A full understanding of the nitrogenase mechanism must include the detailed oxidation state assignments of the component metalloclusters during substrate reduction. Assigning oxidation states to individual metals in a metallocluster is difficult using conventional spectroscopic techniques, due to the ambiguity in assigning spectral features to specific metal sites. To circumvent these limitations, we have developed SpReAD (Spatially Resolved Anomalous Dispersion Refinement), a combination of X-ray crystallography and X-ray absorption spectroscopy to determine site-specific X-ray absorption spectra[34,61], building on the pioneering work by Coppens[33]. By refining structures against diffraction data collected at a series of energies across an absorption edge, the X-ray absorption spectra of each individual metal site can be obtained from the corresponding Δf" values. All other factors being equal, oxidation should lead to a shift in the absorption spectrum to higher energies. For the Fe-protein, oxidation leads to a shift of the main edge position to higher energies by ~1 eV per 0.5 change in oxidation state[62]. It should be noted, however, that edge positions are sensitive to many factors including the ligand sphere, and studies on Fe-S clusters have demonstrated convoluted relationships between oxidation state and edge position[63]. The SpReAD method has previously been used to identify the reduced iron site in the $[2Fe:2S]^{1+}$ cluster of a ferredoxin[34] from *Aquifex aeolicus*, as well as Fe oxidation states in the resting form of the FeMo-cofactor[61] and a mononuclear Fe site in the *Azotobacter vinelandii* MoFe-protein[64]. Here we present a comprehensive picture of the

spatially resolved electronic properties of the *A. vinelandii* Fe-protein in its three possible oxidation states.

3.3 Results and Discussion

High resolution crystal structures are a prerequisite for the SpReAD analysis to ensure that the electron density for the absorbing iron-site is sufficiently well defined. While attempting to crystallize an ADP-AlF_4 stabilized complex between the MoFe-protein and Fe-protein with an excess of Fe-protein, conditions were identified that permitted the first atomic resolution (1.13 Å) structure of the ADP-bound *Azotobacter vinelandii* Fe-protein (see Experimental Section), the highest resolution reported for the Fe-protein. At this resolution, the individual Fe and S electron densities of the [4Fe:4S] cluster are resolved well (Figure 26).

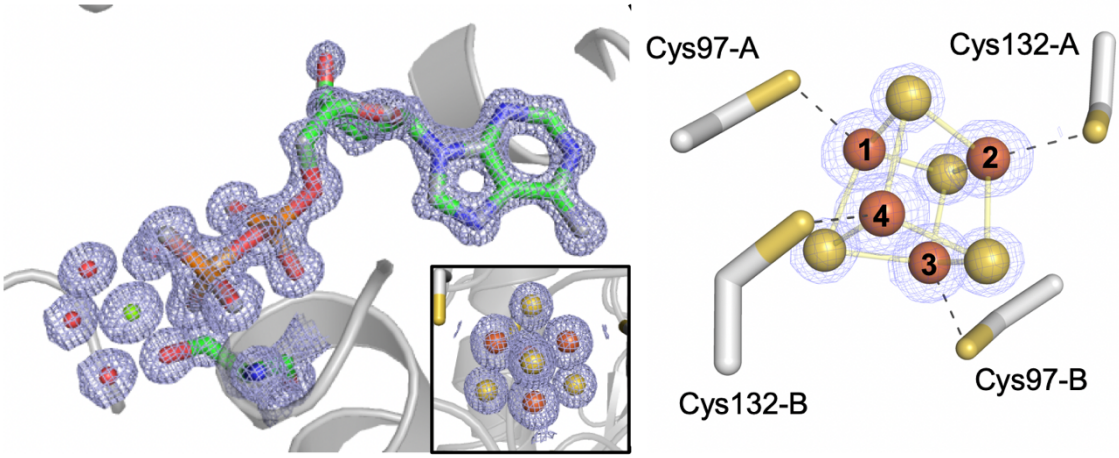


Figure 26. (Left) Electron density of MgADP molecule and [4Fe:4S] cluster from the dithionite-reduced, ADP-bound structure at 1.13 Å-resolution. The cluster bridges the dimer interface and $2F_o$-F_c map is contoured at 1σ. (Right) The Fe labelling scheme used in this work is indicated.

The overall subunit fold not surprisingly resembles previously determined Fe-protein structures[29,65,66], with Cα root mean square deviations (RMSDs) of ≤ 1 Å to existing nucleotide free and ADP-bound forms (Table 3).

	0+	1+	1+ (MgADP)	2+ (MgADP)
1G5P	0.627	0.666	0.745	0.687
1G1M	0.322	0.570	0.727	0.648
1FP6	0.693	0.731	0.576	0.449

Table 3. Cα RMSD (Å) between Fe-protein in different oxidation and nucleotide states reported in here and previously, including the nucleotide free forms of the $[4Fe:4S]^{1+}$ (1G5P) and $[4Fe:4S]^{0}$ (1G1M) oxidation states, and the ADP-bound form of the $[4Fe:4S]^{1+}$ state (1FP6) using Chimera[67].

Aside from the significant increase in resolution, there are two novel features of this structure: (i) the crystals were obtained in the space group $P22_12_1$, with one subunit per asymmetric unit (previous Fe-protein crystals have exclusively contained the full dimer in the asymmetric unit) and (ii) the C-terminus is fully ordered and exclusively interacts with the subunit to which is it covalently attached, in contrast to previous structures where the C-termini are domain swapped to interact with the opposite monomer (Figure 27).

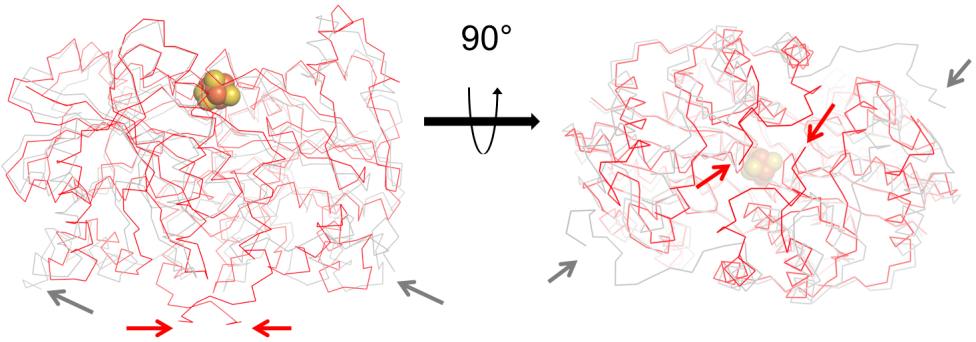
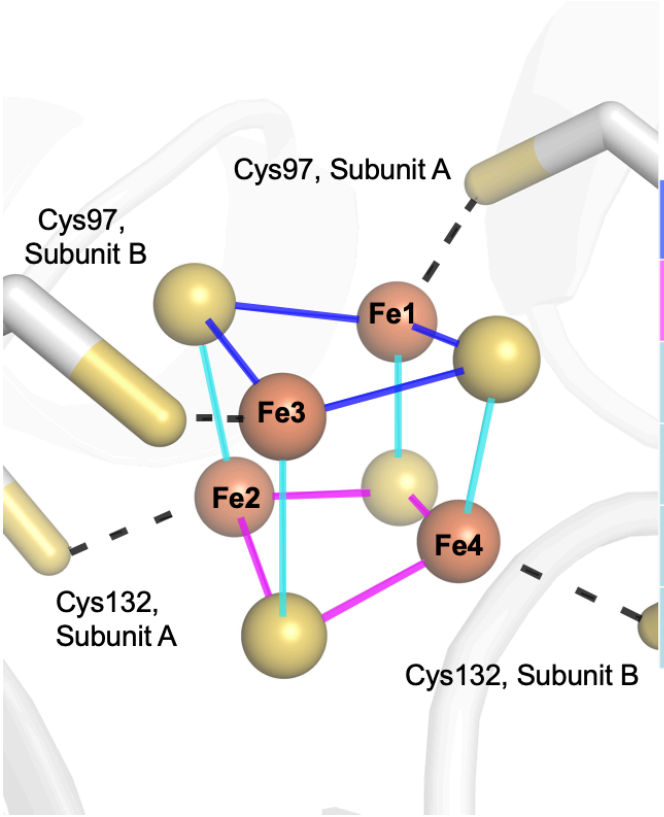


Figure 27. Overlay of the previously reported Fe-protein structure (1G5P, gray) and the 1.13 Å-resolution, Mg-ADP-bound structure determined in this work (red). C-termini are indicated with gray and red arrows, respectively.

The average iron-iron distances in the atomic-resolution cluster (Figure 28) for the dithionite-reduced 1+ state with bound ADP are measured to be 2.70 Å, which is close to the 2.73 Å distances observed by extended X-ray absorption fine structure (EXAFS) for this oxidation state in the absence of MgADP[21,62].



	Cluster charge			
D (Å)	0	1+	1+ (ADP)	2+ (ADP)
Fe1-Fe3	2.86	2.68	2.76	2.73
Fe2-Fe4	2.82	2.64	2.78	2.77
Fe1-Fe2	2.81	2.64	2.70	2.69
Fe1-Fe4	2.84	2.67	2.63	2.61
Fe2-Fe3	2.85	2.70	2.63	2.61
Fe3-Fe4	2.84	2.64	2.70	2.69
AVG	2.84	2.66	2.70	2.68
STDEV	0.02	0.03	0.06	0.06

Figure 28. Fe-S distances in the cluster across oxidation states. Fe1-Fe4 are coordinated by γS of Cys A97, Cys A132, Cys B97 and Cys B132, respectively.

To assess the site-specific oxidation state assignments for Fe in the [4Fe:4S] cluster of the Fe-protein as a function of overall cluster oxidation state (2+, 1+, 0), and in the presence or absence of ADP, SpReAD analyses were conducted. Appropriate diffraction data sets were collected for four forms of the Fe- protein: i) the dithionite (DT)-reduced $[4Fe:4S]^{1+}$ state with ADP, ii) the dithionite-reduced $[4Fe:4S]^{1+}$ state without ADP, iii) indigodisulfonic acid (IDS)-oxidized $[4Fe:4S]^{2+}$ with ADP, and iv) the all-ferrous Ti(III)-citrate reduced $[4Fe:4S]^{0}$ state without ADP. Suitable quality data sets could not be obtained for either the $[4Fe:4S]^{0}$ form in the presence of ADP or for the $[4Fe:4S]^{2+}$ state in the absence of ADP. Intriguingly, these forms are less likely to be mechanistically relevant as ADP and oxidized Fe-protein are the physiological products of electron transfer to the

MoFe-protein. Diffraction data were collected along the rising iron edge centered at 7120 keV using procedures described previously[34,61] (Table 4, 5).

	0+ (-ADP)	**0+ SpReAD**	**1+ (-ADP)**	**1+ SpReAD**	**1+ (+ADP)**	**1+ SpReAD**	**2+ (+ADP)**	**2+ SpReAD**
Wavelength (eV)	12000	7100-7130	12000	7100-7130	14000	7080-8010	12000	7080-8010
Resolution range (Å)	44.68-1.95 (2.06-1.95)	[(2.27-2.05)]	39.29-1.76 (1.82-1.76)	[(2.06-2.05)]	39.01-1.13 (1.15-1.13)	[(2.0-1.80)]	39.03-1.58 (1.67-1.58)	[(2.11-1.92)]
Space group	$P2_1$		C2		$P22_12_1$		$P22_12_1$	
Unit cell parameters	57.25 Å		103.81 Å		45.59 Å		45.81 Å	
	93.05 Å		45.30 Å		74.63 Å		74.58 Å	
	60.77 Å		109.05 Å		75.40 Å		75.02 Å	
	90° 98.5° 90°		90° 96.6° 90°		90° 90° 90°		90° 90° 90°	
Mean I/σ	11.2 (2.2)		8.1 (1.7)		17.6 (3.0)		31.2 (3.7)	
CC(1/2)	0.999 (0.930)		0.998 (0.847)		0.999 (0.915)		1.000 (0.897)	
Completeness (%)	90.04 (92.4)	[96.6-93.5 (89.8-79.5)]	96.1 (55.7)	[92.8-92.2 (76.1-75.9)]	100 (99.9)	[99.1-85.4 (90.0-90.3)]	98.9 (97.8)	[95.0-94.4 (71.3-68.9)]
R_{pim} (all I+ & I-)	0.026 (0.211)	[0.023-0.021 (0.384-0.217)]	0.019 (0.278)	[0.040-0.039 (0.187-0.143])	0.020 (0.239)	[0.020-0.018 (0.357-0.107)]	0.016 (0.250)	[0.017-0.016 (0.343-0.235)]
Total unique reflections	41566 (6160)	[31666-31542 (4527-2373)]	51971 (1562)	[60171-59309 (7230-7046)]	96897 (13986)	[21154-17961 (1165-1172]	35382 (5009)	[23855-16472 (8519-8092)]
Bond length RMSD (Å)	0.021		0.017		0.029		0.023	
Bond angles RMSD (°)	2.430		2.232		2.658		2.158	
R_{factor}	0.2108		0.1684		0.1364		0.1763	
R_{free}	0.2471		0.2197		0.1539		0.1908	

Table 4. Crystallographic data table for Fe-protein in 0+, 1+, and 2+ oxidation states. Statistics in the high-resolution shell are indicated with round brackets. The data used in the SpReAD analysis is given as a range of values, denoted with square brackets.

Energy (eV)	P-cluster (2+)	Fe Mo-co (2.57+)	S	Fe 1 (TiCi)	Fe 2 (TiCi)	Fe 3 (TiCi)	Fe 4 (TiCi)	Fe 1 (DT)	Fe 2 (DT)	Fe 3 (DT)	Fe 4 (DT)
7108	0.58	0.52	0.45	0.36	0.45	0.40	0.48	0.37	0.42	0.36	0.41
7110	0.44	0.45	0.49	0.45	0.45	0.33	0.36	0.39	0.46	0.39	0.35
7112	0.57	0.52	0.50	0.26	0.41	0.31	0.43	0.46	0.54	0.58	0.51
7114	0.62	0.82	0.51	0.80	0.79	0.73	0.73	0.76	0.91	0.83	0.85
7116	1.10	1.10	0.54	0.77	0.74	0.63	0.75	0.71	0.68	0.65	0.66
7118	1.95	1.59	0.53	1.95	1.10	1.80	1.79	1.41	1.80	1.32	1.73
7120	2.72	2.14	0.59	2.68	2.73	2.59	2.64	2.47	2.76	2.32	2.65
7122	3.17	2.73	0.57	3.32	3.32	3.24	3.45	2.98	3.27	2.97	3.22
7124	3.61	3.17	0.57	3.63	3.64	3.55	3.56	3.51	3.70	3.48	3.64
7126	3.78	3.46	0.53	3.96	3.91	3.77	3.88	3.78	3.93	3.62	3.68
7128	3.80	3.64	0.53	3.94	3.97	3.92	3.96	3.65	3.92	3.88	3.77
7133	3.77	3.72	0.53	3.96	3.89	3.99	4.07	3.77	3.80	3.65	3.74

Energy (eV)	P-cluster (2+)	FeMo-cofactor (2.57+)	S	Fe 1,3 (DT +ADP)	Fe 2,4 (DT+ADP)	Fe 1,3 (IDS + ADP)	Fe 2,4 (IDS + ADP)
7108	0.58	0.52	0.45	--	--	0.37	0.35
7110	0.44	0.45	0.49	0.38	0.37	0.39	0.47
7112	0.56	0.52	0.50	0.58	0.66	1.00	0.88
7114	0.62	0.82	0.51	0.60	0.52	0.57	0.43
7116	1.10	1.10	0.52	0.70	0.86	0.69	0.54
7118	1.95	1.59	0.53	1.38	1.70	0.50	0.61
7120	2.72	2.14	0.58	2.27	2.53	2.02	1.88
7122	3.17	2.73	0.57	2.89	3.16	3.02	3.21
7124	3.61	3.17	0.57	3.56	3.67	3.68	3.53
7126	3.78	3.46	0.53	3.93	3.76	4.12	4.16
7128	3.80	3.64	0.53	--	--	4.20	4.20
7133	3.77	3.72	0.53	4.00	4.02	4.47	4.47

Table 5. f'' values as a function of energy for observed oxidation states, nucleotide free (top) and ADP-bound (bottom).

To benchmark the SpReAD analysis and assign discrete oxidation states, reference curves were derived from the MoFe-protein metalloclusters, with the P-cluster and FeMo-cofactor serving as the reference for reduced Fe^{2+} and oxidized $Fe^{\sim 2.57+}$, respectively[61] (Figure 29, 30).

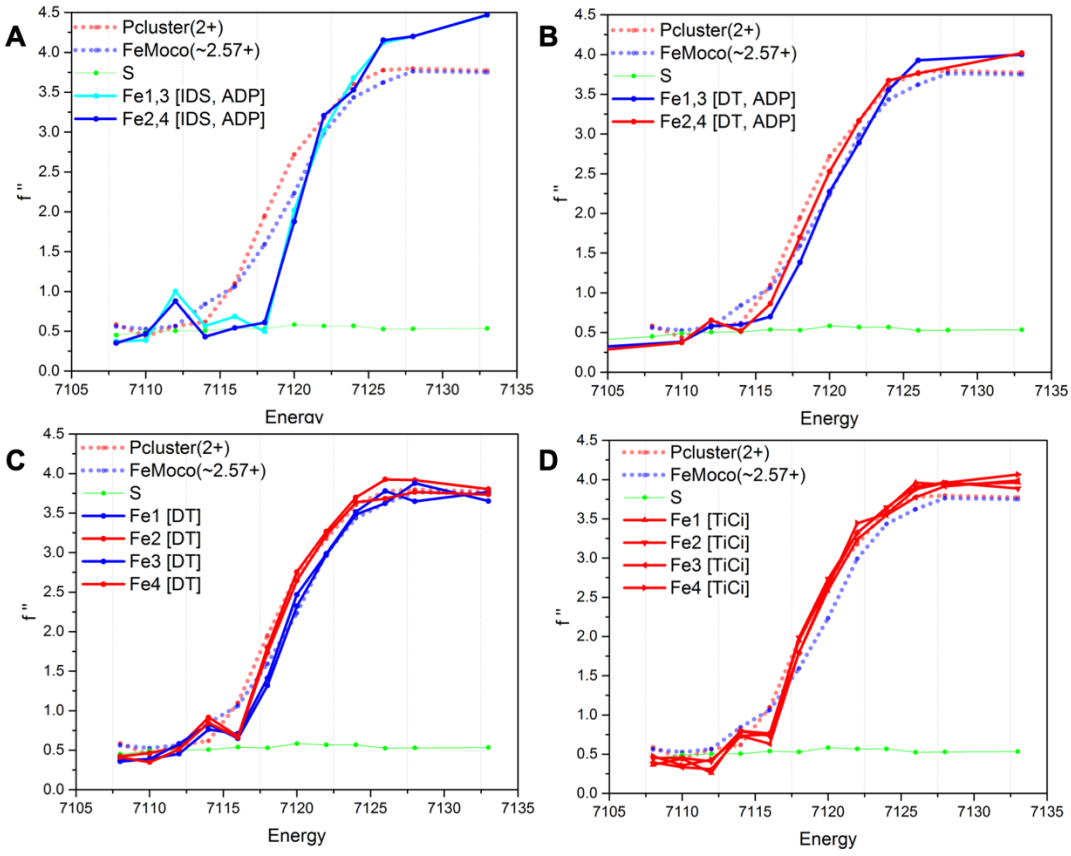


Figure 29. SpReAD curves for the three possible oxidation states of the Fe-protein: (A) the IDS-oxidized $[4Fe:4S]^{2+}$ state with ADP, (B) the dithionite (DT)-reduced $[4Fe:4S]^{1+}$ state with ADP, (C) the dithionite-reduced $[4Fe:4S]^{1+}$ state (nucleotide free) and (D) the Ti(III)-citrate-reduced $[4Fe:4S]^{0+}$ state (nucleotide free). The structures used for the data depicted in panels (A) and (B) contain a crystallographic two-fold axis of symmetry through the cluster and so Fe1 and Fe3 are equivalent, as are Fe2 and Fe4. The structures used for the data depicted in panels (C) and (D) contain an entire [4Fe:4S] cluster in the asymmetric unit, and so the absorption curves for each Fe are presented. More-reduced profiles are coded in red, more-oxidized profiles are coded in blue. For reference, the SpReAD profiles of the dithionite-reduced state of the MoFe-protein metalloclusters are plotted: the average Fe curve for the all-ferrous P-cluster (red dots) and the more oxidized average of all Fe in the FeMo-cofactor ($\sim Fe^{2.57+}$) (blue dots). The sulfur anomalous (green)

is also shown; at 7100 eV the expected value of f'' is 0.70 e (skuld.bmsc.washington.edu/scatter/data/S.dat).

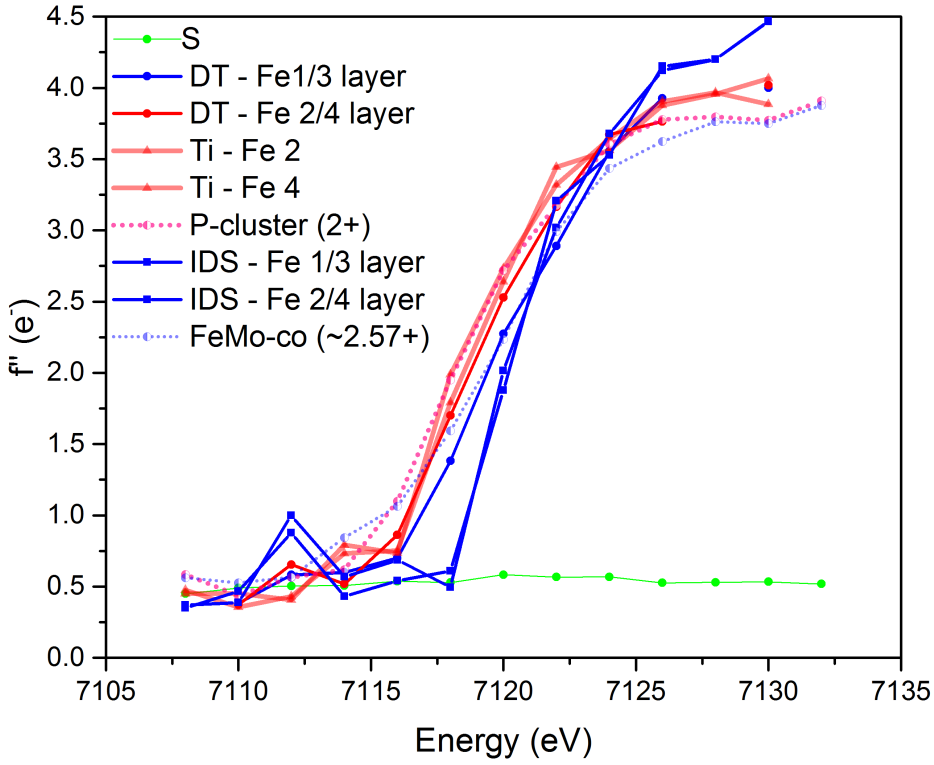


Figure 30. SpReAD profiles for $[4Fe:4S]^{2+}$, $^{1+}$ and ^{0} states with average profiles for the P-cluster and FeMo-cofactor[61].

The average oxidation state of the iron sites in the FeMo-cofactor (~2.57+) was taken as the average of 4 Fe^{3+} and 3 Fe^{2+}[61]. Absorption curves were compiled for each iron site in different forms of the Fe-protein [4Fe:4S] cluster and overlaid with the reference spectra (Figure 29, 30). The comparison to the P-cluster absorption spectra yielded an unambiguous population of Fe^{2+}-like curves, and a second set of curves were shifted to higher energy by roughly 1 eV, indicating a more oxidized state comparable to that of the average Fe-sites in the FeMo-cofactor. The curves that aligned with the P-cluster were

assigned to the Fe^{2+} state, while the data shifted to higher energies are assigned to the expected $Fe^{2.5+}$ state.

The SpReAD analysis of the dithionite-reduced state $[4Fe:4S]^{1+}$ was completed on the MgADP-bound and nucleotide free structures of Fe-protein. In both cases, the site-specific X-ray absorption spectra support the presence of two distinct [2Fe:2S] sublayers perpendicular to the C2 axis of the protein dimer. The sublayer facing the protein surface (coordinated by the Cys 97 Sγ)) is assigned to the more oxidized valence delocalized $Fe^{2.5+}Fe^{2.5+}$ iron sites, while the sublayer facing the interior of the protein (coordinated by Cys 132 Sγ) is assigned as two Fe^{2+}. In contrast, the absorption curves for the Fe in the fully-reduced $[4Fe4S]^{0}$ agree well with the P-cluster reference state, supporting a consistent oxidation state description of Fe^{2+} for all sites. The absorption curves for the fully oxidized $[4Fe4S]^{2+}$ state are all shifted to higher energy as anticipated for the $Fe^{2.5+}$ state, but the curves do differ in shape from those observed from the $Fe^{2.5+}Fe^{2.5+}$ layer of the $[4Fe4S]^{1+}$ state and from the reference FeMo-cofactor (Figure 29, 30).

To confirm that the electronic state of the $[4Fe:4S]^{1+}$ cluster in the crystal reflected the solution properties, perpendicular-mode electron paramagnetic resonance (EPR) was used. A polycrystalline sample of the $[4Fe:4S]^{1+}$ nucleotide-free protein crystals (used in SpReAD experiments) was prepared and compared to the respective protein-solution sample by EPR (Figure 31, see Methods). Solution-state samples produced a rhombic spectrum with a strong S = 1/2 feature at $g = 1.94$ with no other low-field features. Spectra of Fe-protein crystals similarly showed a predominant S = 1/2 feature, with an additional signal at $g = 4.3$ (4% of the $g = 1.9$ integrated peak area) composed of multiple unresolved peaks. Previous studies have attributed the $g = 4.3$ signal to adventitious Fe^{3+}, or an S = 5/2 system by Mössbauer spectroscopy[21] and associate its appearance with the ATP-bound state or conformational heterogeneity. However, the crystal structure of our sample is well ordered and shows no bound nucleotide bound, suggesting the $g = 4.3$ signal arises from a minor spin population. The appearance of additional low-field features is also associated with changes in solvent composition. Spectra were also collected in the presence of the

crystallization solution including PEG, but no additional low-field signals were observed. The resting state of the Fe-protein is generally considered to be a mixture of S = 1/2 and S = 3/2 spin states[20], and the absence of the S = 3/2 state signal may reflect the solvent dependence of the equilibrium, with PEG favouring the S = 1/2 state as observed with ethylene glycol[20].

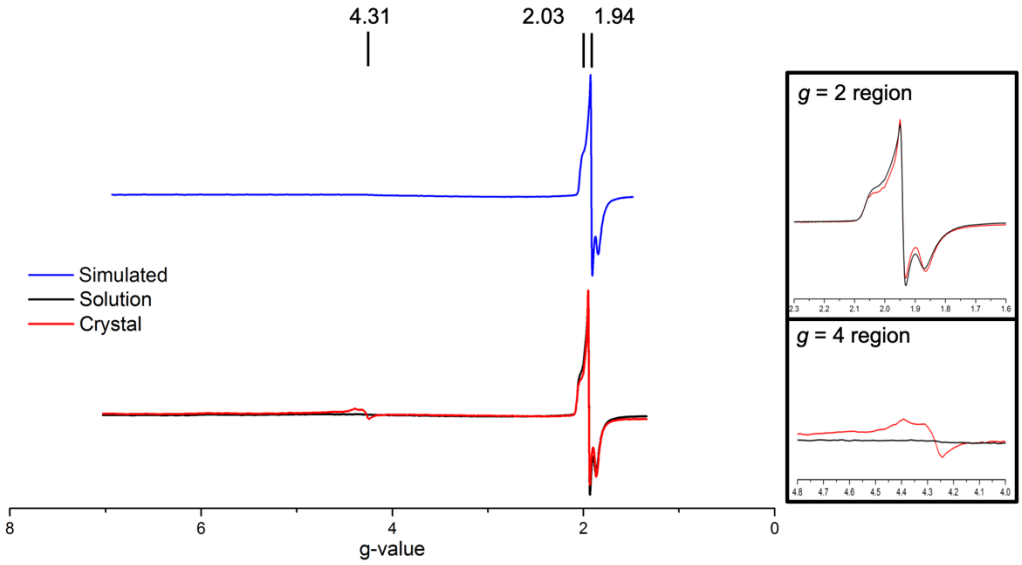


Figure 31. EPR spectra of Fe-protein in the $[4Fe4S]^{1+}$ state in solution (black) and in a crystal slurry (red). Simulated spectrum in blue. Inserts show the expanded $g = 2$ and $g = 4$ regions.

During the process of substrate reduction by nitrogenase, the Fe-protein cycles through a series of conformations coupled to distinct nucleotide and the cluster oxidation states. This coupling is mediated through the interactions between the protein and the cluster. The crystal structures reveal distinct patterns of hydrogen bonds surrounding the cluster that are associated with the presence or absence of MgADP (Figure 32). Although the underlying basis for the preferential stability of one hydrogen bond pattern relative to the other is not evident, the MgADP conformation is observed in the +1 and +2 oxidation states, while the nucleotide free conformation is observed in +0 and +1 states, suggesting

that the MgADP conformation interacts preferentially with the more oxidized forms of the [4Fe:4S] cluster, serving as a redox sensitive conformational switch.

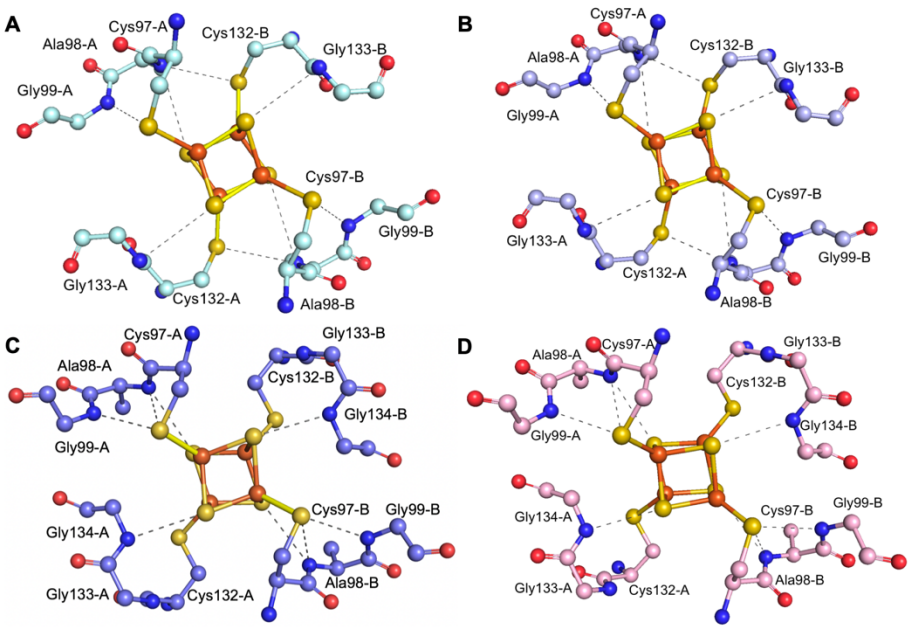


Figure 32. Hydrogen bonding patterns observed in the Fe-protein structures: (A) the IDS-oxidized $[4Fe:4S]^{2+}$ state with ADP, (B) the dithionite-reduced $[4Fe:4S]^{1+}$ state with ADP, (C) the dithionite-reduced $[4Fe:4S]^{1+}$ state (nucleotide free), and (D) the Ti(III)-citrate-reduced $[4Fe:4S]^{0+}$ state (nucleotide free). Distinct hydrogen bonding patterns are observed for the ADP-bound (A,B) and nucleotide free forms (C,D), with little difference associated with the oxidation state changes from +2 to +1 in the presence of ADP (A,B), or from +1 to +0 in the nucleotide free form (C,D).

An important feature of the nitrogenase mechanism is that the $[4Fe:4S]^{+1}$ state of the cluster serves as the electron donor to the MoFe-protein only in the MgATP bound form of the Fe-protein, but not in the nucleotide free and MgADP-bound states. Interestingly, the SpReAD analysis on the latter two structures establishes that the two Fe in the [4Fe:4S] cluster closest to the surface are more oxidized relative to the more buried Fe. The redox competent form of the Fe-protein might be expected to have the reduced Fe nearer the surface to facilitate electron transfer to the MoFe-protein. If so, the role of MgATP could be to promote an internal redox rearrangement such that the positions of the Fe^{2+} and $Fe^{2.5+}$

are switched. The binding of ATP has been associated with a decrease (~100 mV) in the reduction potential of the cluster[68], as well as changes in the EPR spectra that may reflect these changes. A further intriguing observation is that in the $[4Fe:4S]^{1+}$ form, despite the surface exposed nature of the cluster [21], only the MgATP bound form is readily chelated by α,α-dipyridyl and related compounds. These chelators have a higher affinity for Fe^{2+} relative to Fe^{3+} (estimated to be ~ 10^6, based on the reduction potential of 1.12 V[69]). One explanation for this behavior is that the more surface-exposed Fe are reduced to Fe^{2+} in the MgATP bound form and that therefore, the enhanced reactivity with the chelator reflects their reduced state.

Complete understanding of the Fe-protein catalytic cycle will require the structural analysis of the MgATP-bound form that has evaded crystallization for over 25 years. The SpReAD analysis of this form, together with the results described in this paper and the previous work on the dithionite-reduced form of the MoFe-protein[61], will provide a conclusive picture of the site specific redox states of the resting form of the nitrogenase proteins; the next step will then be to pursue non-resting states with substrates, intermediates, and inhibitors bound to the active-site [13,14,70,71].

3.4 Experimental Section

Following previously described protocols[72,12,13], *Azotobacter vinelandii* cells were grown aerobically and nitrogenase expression was induced by ammonia depletion. Nitrogenase proteins from *A. vinelandii* were purified anaerobically using Schlenk line techniques and assessed for activity using acetylene reduction assays. The specific activity for acetylene reduction was on average 2010 +/- 100 nmol of ethylene $min^{-1}mg^{-1}$ for Mo-Fe-protein and 1970 +/- 100 nmol of ethylene $min^{-1}mg^{-1}$ for Fe-protein. Nitrogenase protein solutions were buffered in 50 mM Tris/HCl pH 7.5, 200 mM NaCl, 5 mM sodium dithionite. Ti(III) citrate was prepared from a 1.5% solution of Ti(III) chloride in 0.2 M sodium citrate, neutralized to pH 7.0 by titrating in sodium carbonate.

3.4.1 Crystallization

All crystals were obtained by vapor-diffusion sitting-drop techniques in 95% Ar, 5% H_2 atmosphere, 25°C. *A. vinelandii* Fe-protein was used at concentrations between 20-40 mg/mL.

3.4.1.1 ADP-bound Crystal Formation

The dithionite-reduced, ADP-bound crystals were identified from a condition containing Fe-protein and MoFe-protein at a 4:1 molar ratio, with 1 mM $AlCl_3$, 10 mM NaF, 5 mM ATP, and 10 mM $MgCl_2$ to form a stabilized protein complex with AlF_3^-. A 100 kDa molecular weight cut-off concentration unit (EMD Millipore) was used to separate free Fe-protein, but Fe-protein was retained in the mixture used for crystal trials. A population of Fe-protein hydrolyzed MgATP, resulting in crystalized Fe-protein-MgADP. The crystals were obtained in 40% PEG 400, 0.17 mM Cymal 7, 0.1 M HEPES pH 7.5, 5 mM dithionite. Oxidized structures were obtained by first growing crystals in the dithionite-reduced, MgADP condition described, and soaking the crystals in 5 mM indigocarmine (IDS) for 1 – 5 min. The progress of the crystal soaks was apparent by the change in color of the crystals from ~brown to ~green over time. Crystals were cryo-protected with 2-methyl-2, 4-pentanediol (MPD). Ti(III) citrate-reduced protein was screened for an ADP-bound form, but the trials did not yield crystals.

3.4.1.2 Nucleotide-free Crystal Formation

Crystals of the dithionite-reduced, nucleotide free state were grown in a solution of 40% PEG 1000, 0.2 M NaCl, 0.1 M MES/OH pH 6.5, 10% 2,2,2-trifluoroethanol, 5 mM dithionite, and cryo-protected with 2-methyl-2, 4-pentanediol (MPD). All-ferrous crystal structures were grown in conditions of 37% PEG 3350, 0.05 M NaCl, 0.1 M Bis/Tris pH 5.5, 5 mM Ti(III) citrate. Crystals of the IDS-oxidized nucleotide-free form were prepared from the dithionite-reduced nucleotide free crystals, but yielded poor diffraction.

3.4.2 Structure Determination

Crystallographic data were collected at the Stanford Synchrotron Radiation Lightsource at beamline 12-2, using 0.5° steps over 360°. The high-resolution Fe-protein structure was obtained in the orthorhombic space group $P22_12_1$ with cell constants of a = 45.71 Å, b= 74.6 Å, c= 75.2 Å. Datasets were collected at multiple wavelengths (see Table 5) across the iron K-edge (7121 eV). Spatially resolved anomalous dispersion refinement (SpReAD) analysis was performed as previously reported by refining Δf'' values against the anomalous differences for individual structure factors[34,61], with phases calculated from the structure refined against the high-resolution dataset collected at 12,000 eV. The data were integrated with XDS[73] and merged with SCALA[74]. The high-resolution structures collected at 12,000 eV were phased with PHASER by molecular replacement using one monomer of the previously determined Fe-protein structure (1G5P)[65], with the [4Fe:4S] cluster omitted. For the refinement, the SF4 and FES were used as standard ligands from the CCP4 library (FES Fe-Fe bond distance: 3.06 Å, Fe-S bond lengths: 2.20 Å. SF4 Fe-Fe bond lengths: 3.10 Å, Fe-S bond distance: 2.19 Å). We note that the Fe-Fe distances in these standard groups are significantly longer than observed in actual FeS clusters (~2.70 Å). The holo structures were refined with Refmac5 in the CCP4 suite, and provided the phases for the high-energy datasets[75]. Model building was performed in Coot[76]. For an overlay of all SpReAD profiles, see Figure S3. The nucleotide observed was assigned as MgADP based on the coordination sphere of the Mg^{2+} ion; we observed a hexacoordinate ion adjacent to the ADP molecule, liganded to four waters, Ser16, and a phosphate from ADP. The Mg^{2+}-ligand distances are ~2.1 Å, consistent with $Mg^{2+}(H_2O)$ complexes[77]. An AlF_4^- molecule would have predicted ~1.7 Å bond lengths for Al-F[78], with a square planar geometry and a binding site distinct from the Mg^{2+} ion, as observed in the 1M34 complex structure[11].

The coordinates and structure factors have been deposited in the RCSB Protein Data Bank entries: 6N4J (Ti(III)-citrate reduced, nucleotide free form), 6N4K (DT-reduced,

nucleotide free form), 6N4L (DT-reduced, ADP-bound form), 6N4M (IDS-oxidized, ADP-bound form).

3.4.3 Electron Paramagnetic Resonance

EPR experiments were performed on an X-band Bruker EMX CW-EPR spectrometer with an ESR 900 liquid helium/nitrogen flow-through cryostat (Oxford Instruments) at powers ranging from 2 mW to 10 mW at 10 K. The modulation amplitude was 3 G and the modulation frequency 100 kHz, with a 40 ms time constant and 167.7 s sweep time. In the solution-state sample, the protein concentration was 20 mg/mL (0.31 mM) in 18.25% PEG 1000, 0.1 M sodium chloride, 0.05 M MES/OH pH 6.5 and 10 mM sodium dithionite. The crystal slurry sample was prepared by growing *A. vinelandii* Fe-protein crystals in 2 μL sitting drops. 115 clear drops with crystals were pooled in reservoir buffer (36.5% PEG 1000, 0.2 M sodium chloride, 0.1 M MES/OH pH 6.5, 10 mM sodium dithionite), and crystals were broken using a cryo-loop. Samples were frozen in liquid nitrogen in 4 mm thin-wall quartz tubes (Wilmad-Lab Glass). Simulations were performed using GeeStrain5[37]. Simulation parameters: g = (1.85, 1.927, 2.00), σ_g = (0.023, 0.010, 0.045).

3.5 Acknowledgments

We thank Prof. Dr. O. Einsle, Prof. Dr. L. Zhang, Dr. K. Perez, Dr. R. Arias, A. Maggiolo and Prof. Dr. J. B. Howard for useful discussions, and K. H. Lee for the initial identification of the ADP-bound Fe-protein crystals. We acknowledge the Gordon and Betty Moore Foundation and the Beckman Institute at Caltech for their generous support of the Molecular Observatory at Caltech. We thank the staff at Beamline 12-2, Stanford Synchrotron Radiation Lightsource (SSRL), operated for the DOE and supported by its OBER and by the NIH, NIGMS (P41GM103393), and the NCRR (P41RR001209). We also thank Dr. A. Di Bilio for assisting with EPR experiments. The Caltech EPR Facility was supported by NSF-1531940. This work was supported by NIH grant GM045162 and the Howard Hughes Medical Institute.

Chapter 4

SOLUTION OF AN ANAEROBIC CRYO-EM STRUCTURE OF A NITROGENASE MATURATION COMPLEX

4.1 Introduction

The recent improvements in cryo-electron microscopy (cryo-EM) has led to an explosion of protein structures[79] in a frozen-hydrated state, capturing difficult-to-crystallize complexes due to size and heterogeneity. Recently, a cryo-EM structure of photosystem complex I yielded insights into electron transfer by ferredoxin[80]. Many metalloproteins participate in dynamic relays for electron transfer or metallocofactor assembly. The proteins involved in the process frequently require domain rearrangements and must be flexible. In electron transfer and cofactor assembly, binding and catalysis involve reactive metal species that necessitate protection from the bulk solvent as well as side-reactions.

A classic system that involves complex transformation of metal species is the assembly of the unique metallocofactors in the enzyme nitrogenase. Nitrogenase is the two-component enzyme that catalyzes the biological reduction of nitrogen to ammonia, and carbon monoxide to hydrocarbons, among other substrates[47,81–83]. Nitrogenase is composed of the MoFe (NifDK) and Fe (NifH) proteins, the substrate-binding and reductase proteins, respectively. NifDK contains two elaborate metallocofactors: the FeMo-cofactor (FeMo-co, [7Fe:9S:C:Mo-*R* homocitrate]) and the P-cluster ([8Fe:7S]), both of which require specialized protein machinery to assemble and insert a matured cluster into apo-NifDK[24]. In the final steps of FeMo-co assembly, apo-NifDK receives the matured FeMo-co from NifEN[84,85], which requires the binding and dissociation of the radical *S*-adenosyl methionine (SAM) protein NifB[86], and the reductase NifH[28]. One of the most intriguing steps of FeMo-co assembly is the insertion of the central carbon atom by the assembly protein NifB, which generates the so-called L-cluster ([8Fe:9S:C]) from the K-cluster (2[4Fe:4S]) and *S*-adenosyl methionine[87] . Directly following the carbon insertion is an

interaction between NifB, NifEN, and NifH. The interactions between NifB, NifEN, and NifH ultimately convert the L-cluster into FeMo-co (M-cluster, [7Fe:9S:C:Mo-*R* homocitrate])[23,24]. Catalysis by NifB can be followed spectroscopically by electron paramagnetic resonance (EPR) and X-ray emission spectroscopy (XES). Spectroscopic experiments support a model in which the K-cluster (2[4Fe:4S]) is converted to the L-cluster ([8Fe:9S:C]) for transfer to NifEN, which interacts with NifH to yield mature M-cluster ([7Fe:9S:C:Mo-*R* homocitrate])[88]. Overall, the process involves dynamic protein complexes containing oxygen-labile metal clusters. Here, we report the cryo-EM structure of the intermediate complex NifENH-ADP•AlF_4^-, formed in the FeMo-co maturation pathway, using an anaerobic grid preparation technique.

4.2 Anaerobic Single Particle Cryo-Electron Microscopy

Initially, samples were prepared in an excess of dithionite (5 mM) and vitrified in open atmosphere. The samples were observed to dissociate into constituent monomers that could be explained by cluster oxidation from oxygen exposure and loss of clusters that bridge subunits. To eliminate the introduction of oxygen to oxygen-labile samples, we developed a method to prepare samples for single particle analysis anaerobically.

In order to validate our strategies for preparing intact assembly proteins, we solved structures of the nitrogenase proteins (see Experimental Section). We prepared grids of MoFe-protein (NifDK) and observed strong orientation preference and, surprisingly, dissociation of the tetramer (Figure 33). We hypothesized that the protein was partially denaturing at the air-water interface and introducing artifacts. Multiple strategies for optimizing samples for single particle analysis have been described, including modifying the grid support with graphene oxide[89] or antibody conjugates[90], adding detergents to improve orientation preference and reduce aggregation[91], and applying multiple rounds of sample to improve particle density (to compensate for sample being lost to adsorption on the carbon support)[92].

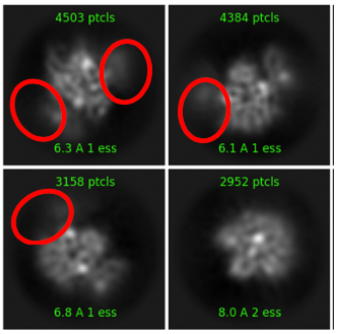


Figure 33. In red, subunits of the tetramer are missing from the class averages.

In order to alter the stability of NifDK on the grid, we added a non-specific, sacrificial ~20 kDa Fab (see Experimental Section) in a 1:1 molar ratio. We hypothesized that the Fab would compete with the NifDK protein for denaturation at the air-water interface and adsorption to the carbon support between holes. Ultimately, we solved a 4.5-Å resolution structure of NifDK in isolation (Figure 34).

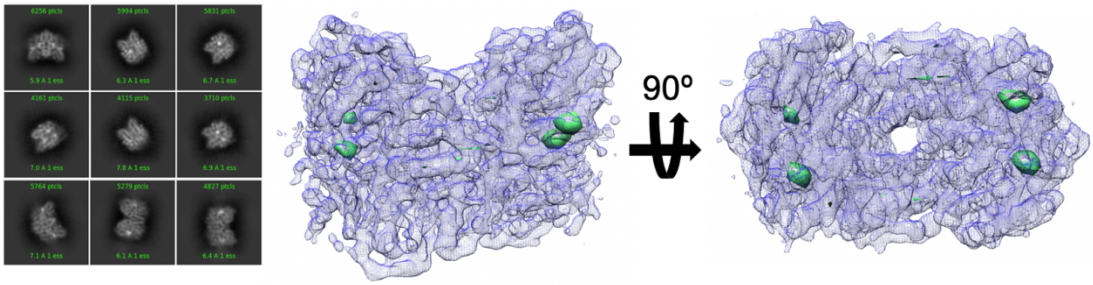


Figure 34. Anaerobic cryo-EM structure of NifDK. (Left) Representative 2D classes from cryoSPARC. (Right) Map in blue, with metal clusters in green. The blue map is contoured to a threshold of 0.58 in Chimera, corresponding to a calculated 4.8 σ (standard deviation = 0.12). The green map is contoured to a threshold of 2.33, or 19 σ.

The structure is superimposable with the crystal structure, 3U7Q[12], with the notable exception of the N-terminal 50 residues, which were disordered in the EM structure. At lower contour levels, the N-terminal element of the α subunit is present, but the density is significantly weaker compared to the core of the protein.

We collected a dataset of 630 micrographs on a 300 keV F30 Polara and solved a 5.7-Å resolution structure of the MoFe-Fe-protein (NifDKH) complex with an ATP transition state analog, ADP•AlF_4^- (Figure 35).

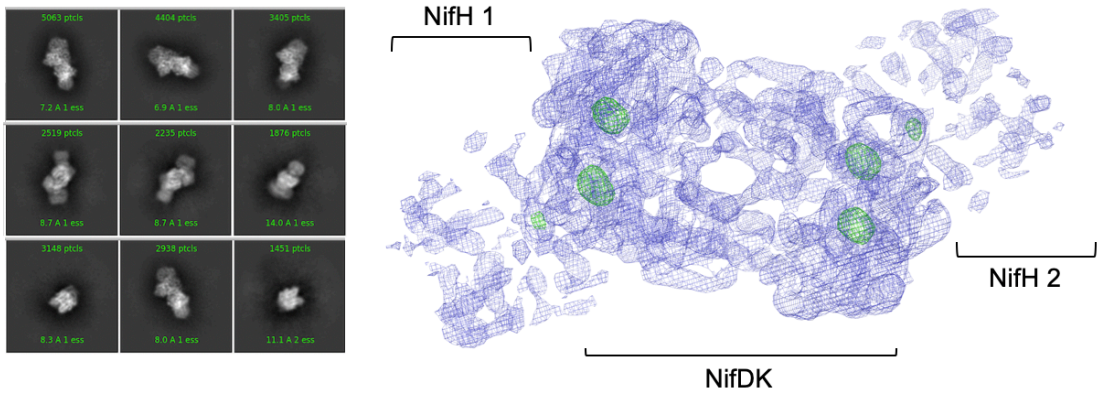


Figure 35. Anaerobic cryo-EM structure of NifDKH-ADP•AlF_4^-. (Left) Representative 2D classes from cryoSPARC. (Right) Final map in blue, with metal clusters highlighted in green. The blue map is contoured to a threshold of 0.43 in Chimera, corresponding to a calculated 5.6 σ (standard deviation = 0.08). The green map is contoured to a threshold of 1.94, or 24 σ.

The protein density agrees well with the 2.3-Å resolution crystal structure of the complex, 1M34[11], and importantly, all six metal clusters were identified. The structure was solved in the absence of a sacrificial protein or other additives, due to the full occupancy of all subunits.

4.3 Cryo-EM Structure of Resolved NifENH-ADP•AlF_4^- Complex

We sought to solve a structure of a fusion protein of NifBENH bound with the ATP-transition state analog ADP•AlF_4^- (see Experimental Section). Through various grid optimization steps, we were ultimately able to reliably resolve the NifENH complex (Figure 36) and solved a 4.7-Å resolution structure. Although the NifB subunit was not resolved, this is the first structure of a complete NifENH complex.

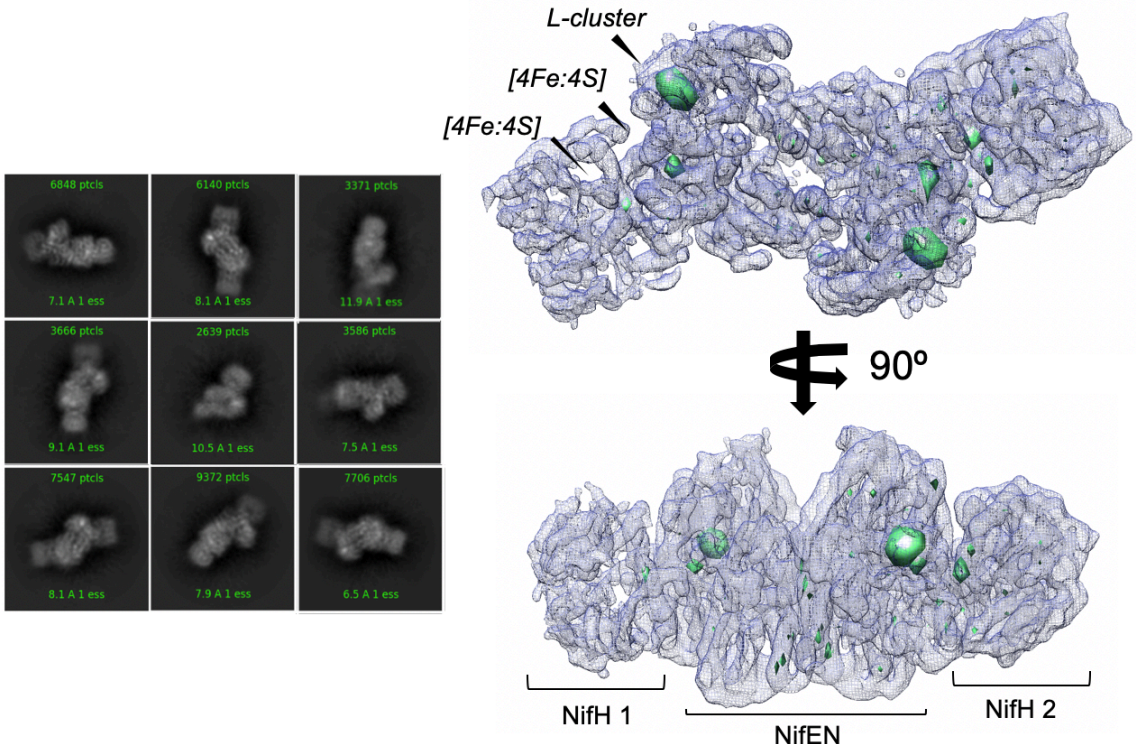


Figure 36. Anaerobic cryo-EM structure of NiENH- ADP•AlF_4^-. (Left) Representative 2D classes from cryoSPARC. (Right) Map in blue, with metal clusters in green. Putative cluster assignments in italics. (N.B. Some strong density from the core helices becomes visible at a similar contour to the [4Fe:4S] cluster of NifH). The blue map is contoured to a threshold of 0.41 in Chimera, corresponding to a calculated 3.4 σ (standard deviation = 0.12). The green map is contoured to a threshold of 2.06, or 17 σ.

The density was well modeled by each of the constituent subunit structures (NifEN, 3PDI[84] and NifH, 1M34[11]). The resolution of the EM structure is sufficient to identify secondary structural elements, and the map could not accommodate the predicted TIM-barrel motif of the canonical radical SAM proteins (e.g. MoaA[93]) of a similar molecular weight to NifB.

Rigid-body docking and refinement of the NifEN and NifH structures into the map builds a complex with a stoichiometry of 2:1 NifH:NifEN (Figure 37, Table 6), with C2 symmetry (no symmetry was applied during the map generation, and imposing symmetry did not significantly improve the structure resolution).

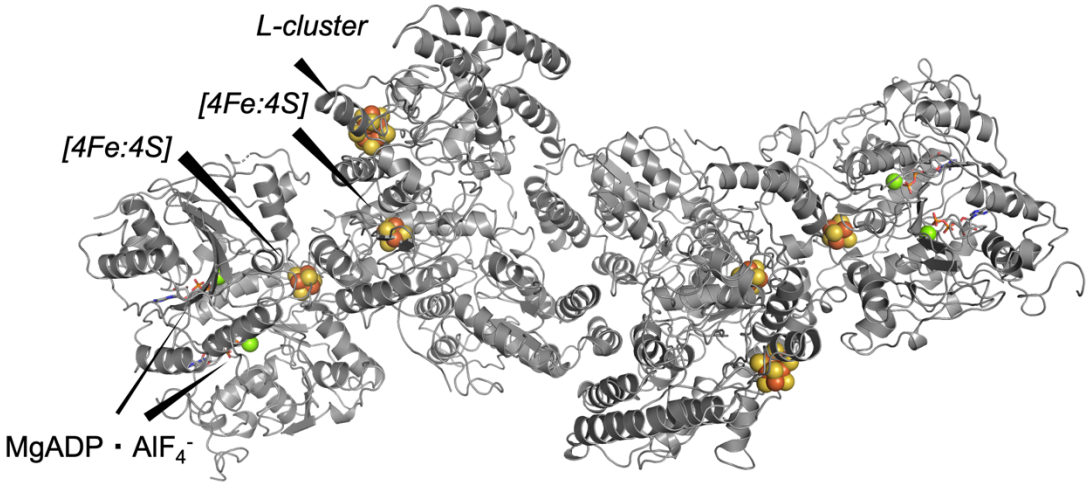


Figure 37. Ribbon model of NifENH based on the EM map. The cofactors are indicated.

Atoms	21362
Bonds (RMSD)	
Length (Å)	0.014
Angles (°)	1.391
MolProbity Score	2.98
Clash Score	11.9
Ramachandran plot (%)	
Outliers	0.29
Favored	89.6
Rotamer outliers (%)	12.52

Table 6. Validation table for the NifENH model.

The complex structure shows that NifH binds to NifEN in a similar fashion to NifDK. The NifH proteins have been observed to bind in distinct, but overlapping conformations in the nitrogenase complex based on nucleotide state[94]. The dihedral angle calculated between the NifH cofactors ($[4Fe:4S]^{NifH}$-$[4Fe:4S]^{NifEN}$-$[4Fe:4S]^{NifEN}$-$[4Fe:4S]^{NifH}$) across the complex is 148°, compared to 150° calculated for the NifDKH complex ($[4Fe4S]^{NifH}$-

$[8Fe:7S]^{NifDK}$-$[8Fe:7S]^{NifDK}$-$[4Fe:4S]^{NifH}$, using coordinates from 1M34[11]). This dihedral angle across the complex indicates a highly similar binding orientation with respect to the primary ($[4Fe:4S]^{NifH}$) and intermediate ($[4Fe:4S]^{NifEN}$, $[8Fe:7S]^{NifDK}$/P-cluster) clusters.

The interaction face between NifH and NifEN involves two short helices connecting the intermediate cluster ($[4Fe:4S]^{NifEN}$) to the protein surface. These helices are conserved in NifDK, and they interact in an analogous fashion in the NifENH complex structure. In NifDK, one helix from the α subunit and one from the β subunit connect the P-cluster to the protein surface (Cysα154-Glyα160, Thrβ152-Glyβ159), terminating in loop regions that contain a Gly-Asp-Asp motif. The two GDD motifs present an electronegative patch on the protein surface that aligns with the docking site of NifH as observed by the complex crystal structure (Figure 38).

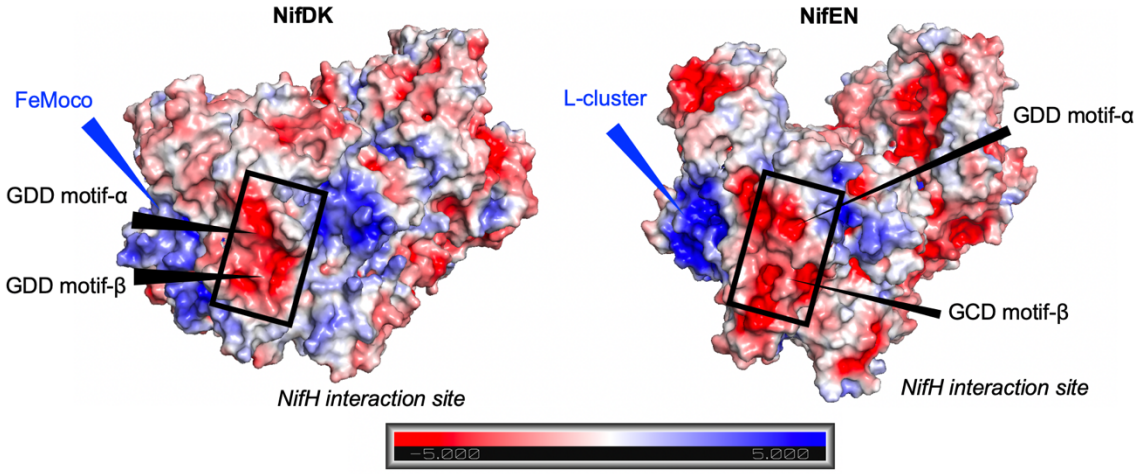


Figure 38. (Left) NifDK electrostatic potential map, with the NifH binding site indicated. (Right) NifEN electrostatic potential map, with NifH binding site indicated. The electrostatic scale bar is (red) -5 to (blue) +5 k_BT. Map was generated using PDB2PQR[95].

Similarly, NifEN contains two helices connecting the intermediate cluster ([4Fe:4S]^{NifEN}) to the protein surface: Cysα124-Glyα-130 and Thrβ101-Aspβ110. Both helices also terminate in a loop region containing Gly-Asp-Asp in chain α and Gly-Cys-Asp in chain β. The region forms the interaction face with NifH.

The global conformation of NifH is similar to published crystal structures (Table 7), with RMSD values around 1 Å. Similarly, the NifEN portion of the model has an RMSD around 1 Å compared to the published crystal structure. However, fixing one NifH monomer during alignment reveals deviations previously observed for nucleotide-dependent conformations of NifH[94] (Figure 39). The conformation of NifH modeled into the EM structure most closely matches the structure of NifH bound to an ATP transition state analog, ADP-AlF_4^-, and in complex with NifDK (PDB 1M34). Compared to the nucleotide-free state, the ADP-AlF_4^- conformation is characterized predominantly by a hinge motion along the dimer axis. A secondary twisting mode of the monomers relative to each other is also observed, but the twisting mode is a larger component of the ADP- and AMPPCP-bound forms. A pairwise alignment of the monomers indicates that the NifH dimers in the EM structure are most similar to the ADP-AlF_4^- conformation, but the EM model has a larger displacement of the hinge motion, resulting in the ~1 Å deviation with the NifH from 1M34 (Figure 39).

	Global NifH RMSD (Å)	one NifH subunit fixed RMSD (Å)
1G5P[65]	1.00	3.84
1M34[11]	1.21	1.79
4WZA[94]	1.08	2.20

	RMSD with NifEN model (Å)
3PDI[84]	1.18

Table 7. Cα-RMSD comparisons with published (top) NifH crystal structures, and (bottom) NifEN, calculated with Chimera[67] and PyMOL (Schrödinger LLC).

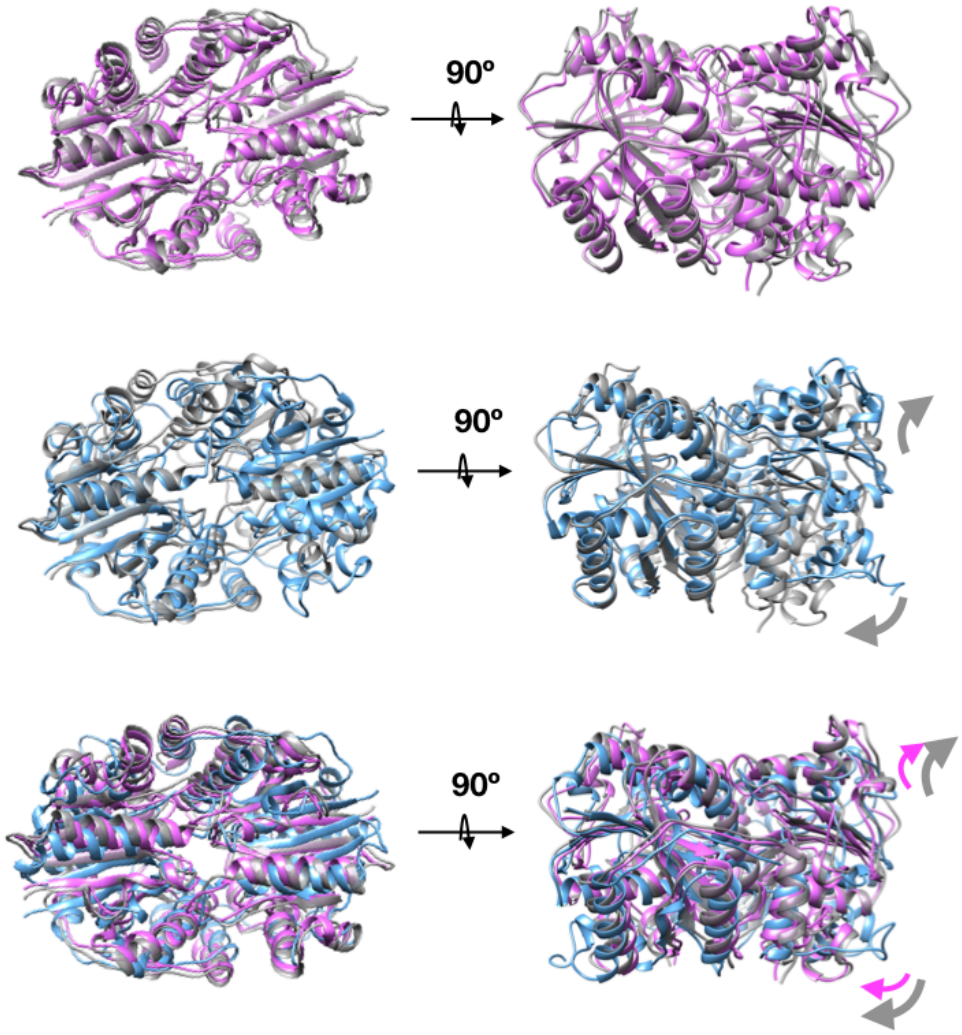


Figure 39. Conformational comparisons of NifH bound to NifEN (gray), with one subunit aligned to NifH-ADPAlF$_4^-$ (1M34, magenta) and nucleotide-free NifH (1G5P, blue). Arrows indicate the hinge domain motion of NifH in complex with NifEN. Size of the arrows indicate the magnitude of the displacement.

The regions of highest RMSD deviation (Figure 40) between NifH bound to NifEN and NifH bound to NifDK in the ADP-AlF$_4^-$ conformation include: a short helical element from Asp43-Leu49, a loop region from Ala114-Asp117 and the C-terminus. Importantly, RMSD deviations in the P-loop (Lys10-Lys15) and the Switch II region (Asp129-Val130), which reflect the nucleotide state of the protein, are very low and the backbones are superimposable. Overall, the maximum displacement was 4.34 Å, the minimum displacement was 0.19 Å.

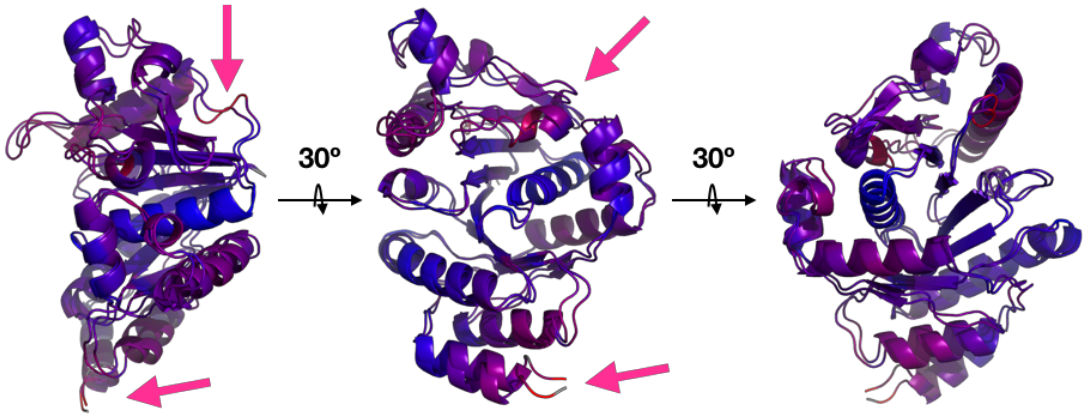


Figure 40. Single monomers aligned from NifH in bound to NifDK in the ADP-AlF_4^- conformation (1M34) and bound to NifEN. The structures are colored by the distance between the Cα pairs, with blue corresponding to a minimum value and red corresponding to a maximum value. Regions of high RMSD deviations are further highlighted with pink arrows. Alignment and RMSD calculations were performed with Pymol (Schrödinger LLC).

NifDK with and without NifH bound are nearly identical (~0.4 Å RMSD between NifDK alone and NifDK in the complex NifDKH), whereas NifEN compared to the NifENH complex has displacements (~2-5 Å) of many helices on the exterior of the protein. The overall effect is an expansion of NifEN within the complex, and a spreading of exterior helices away from the core of the protein. The calculated volume[104] of NifEN alone is ~280,000 $Å^3$ (using the 3PDI crystal structure), which increases to ~300, $Å^3$ when bound to NifH in the EM map. A similar comparison of NifDK bound (1M34[11]) and unbound to NifH (3U7Q[12]) reveals fewer analogous changes, with an overall Cα RMSD of 0.39 Å. Changes to NifDK are localized to the immediate binding interface with NifH and minor deviations in the position of the N-terminus, while the analogous comparison with NifEN contains changes distributed throughout the surface helices of the protein. Further detailed interpretation of the structural differences is limited by the resolution of the structures.

The NifENH structure contains two areas of unmodeled density. One region is at the N-terminus of the NifE domain, near the position of the putative L-cluster (Figure 41). The

cluster is surface-exposed in the crystal structure, with no density apparent for the first 25 N-terminal residues. In the EM map, there is disordered but significant density between the cluster and solvent, possibly corresponding to the unmodeled N-terminus. The loop region Glu-368-Leu377 is similarly disordered in the EM map.

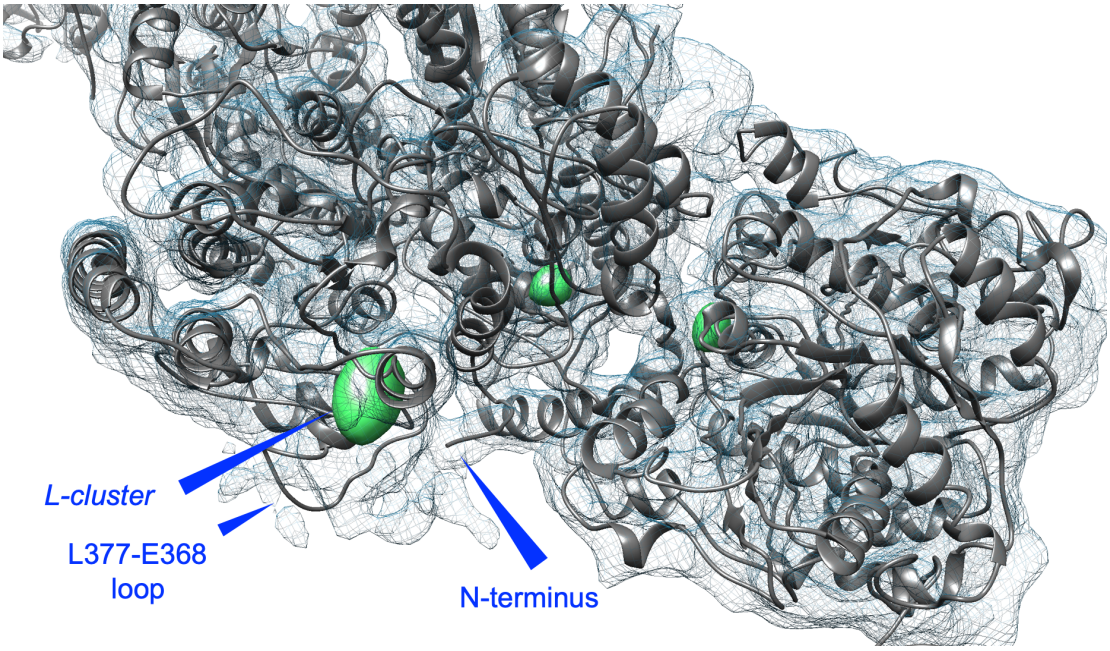


Figure 41. NifENH model within the EM map. Metal clusters are shown in green.

A second region of unmodeled density is in the core of the protein, extending from the C-terminus of NifN (Figure 42). The location corresponds to the site of NifB covalent attachment via genetic fusion. Interestingly, the density extends for roughly 25 amino acids (assuming helical secondary structure), and no additional density corresponding to the remaining ~30 kDa NifB protein is observed. Speculative modeling of NifB would place the protein in the large groove formed by the NifEN tetramer, or extending above the NifE or NifN domains. If the NifB domain were to project away from the core of the NifEN tetramer, and instead bind to the solvent-exposed loop regions (e.g. Arg179-His196) of NifE, it is possible that NifB could interact with the putative L-cluster binding site to transfer the matured L-cluster to NifEN. Transfer of the L-cluster to NifEN could trigger

dissociation, leading to a loss in stability of NifB. The potentially unstable apo-NifB could then be susceptible to denaturation or multiple disordered conformations that lead to ambiguity in the EM maps.

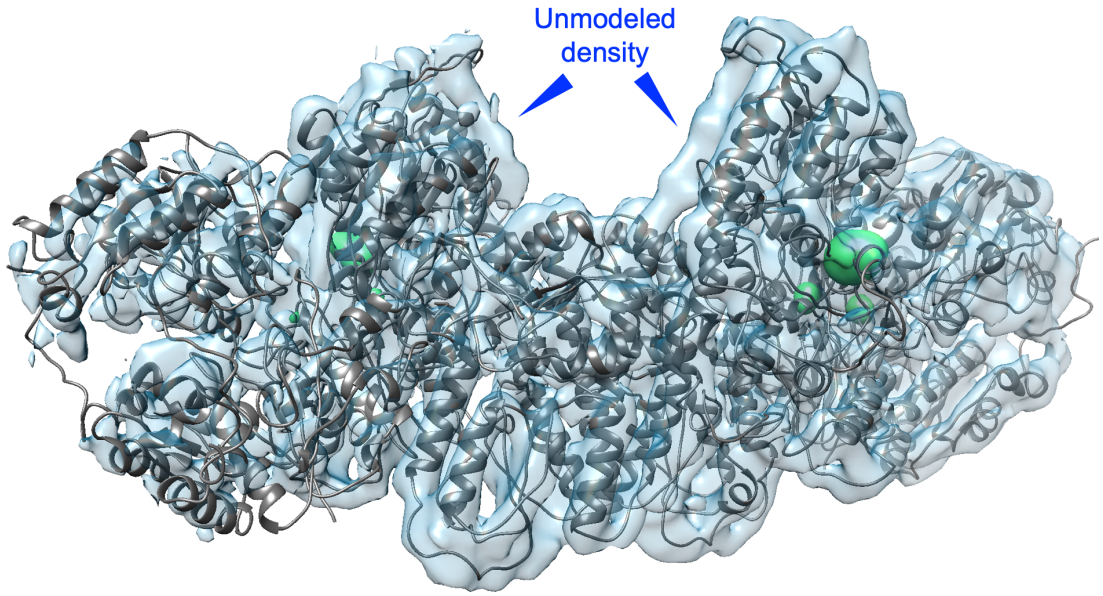


Figure 42. Unassigned density in the core of the protein, corresponding to the N-terminal region of NifB.

4.4 Discussion

Our results provide, to our knowledge, the first cryo-EM structure of an anaerobic protein complex. While the full target complex was not resolved, the proteins NifEN-NifH-ADP•AlF_4^- were identified. The NifENH- ADP•AlF_4^- complex contains a 2:1 stoichiometry of NifH:NifEN (overall $\alpha_2\beta_2\gamma_4$), and 6 metal clusters were identified based on the component crystal structures: two putative L-clusters ([8Fe:9S:C]) and two [4Fe:4S] clusters per NifEN, and one [4Fe:4S] cluster per NifH. The function of NifEN is to serve as the maturase of the FeMo-co ([7Fe:9S:C:Mo-*R*-homocitrate)[84], the cofactor used for nitrogen reduction by NifDK. NifEN interacts with NifH to catalyze the final maturation steps: the installation of the molybdenum atom in place of the 8th iron, and *R*-homocitrate. Biochemically, maturation assays have demonstrated the NifH-dependence of the

molybdenum and homocitrate insertion[28], but the spatial arrangement of the complex components was missing.

In our structure of NifENH- ADP•AlF_4^- , the arrangement of the proteins is analogous to the nitrogenase complex NifDKH ADP•AlF_4^- [11], in which the position of the auxiliary (non-L) clusters support an electron transfer role through NifEN to the putative L-cluster. In our model, NifH binds and transfers electrons in an ATP-dependent manner, passing electrons to the intermediate [4Fe:4S] cluster in NifEN, which then passes electrons to the L-cluster. An electrostatic potential surface analysis of NifEN templates the binding of NifH through electronegative patches similar to NifDK. Overall, the surface potential of NifEN presents multiple acidic regions that may allow for more than one binding mode of NifH, but no heterogeneity in the map supports alternative binding models. While the role of electron transfer is not fully understood in the NifENH complex, the L-cluster may require reduction for removal of the 8th iron. Additionally, maturation assays have utilized molybdate (MoO_4^{2-}) as the molybdenum donor, which contains a Mo(VI). Based on high-energy resolution fluorescence detected K-edge Mo X-ray absorption spectroscopic methods, the molybdenum in FeMo-co has been identified as a Mo(III)[96]. The conversion of molybdate to Mo(III) within the cofactor requires a three-electron reduction and the difficult removal of oxo ligands and replacement with sulfur ligands. It is possible that reduction of the L-cluster is required for molybdenum mobilization, facilitated by electron transfer through NifEN by NifH.

Our structure also provides evidence that the L-cluster is less surface-exposed than observed in the crystal structure, and is likely shielded in a poorly-ordered manner by the N-terminus. The state captured in the EM structure may be related to positional changes in the L-cluster in response to NifH binding, and requires further experiments to confirm. Unfortunately, our structure does not provide structural information of NifB, and only contains the first ~25 residues arranged in a helical structure. It is possible that the state captured reflects a deliberately-dissociated NifB, having already transferred the immature FeMo-co to NifEN. The transfer of the immature FeMo-co may be highly destabilizing to

NifB, and contributes to the disorder in the structure. The electron-transfer role of NifH may be related to a conformational change that disrupts the NifEN-NifB interaction to signal a complete transfer of immature cofactor.

Thusfar, NifH has been biochemically implicated in multiple roles related to supporting nitrogen reduction in the nitrogenase complex, and maturing FeMo-co precursors in the maturation pathway. Structurally, we have evidence of NifH in the ADP-AlF_4^- state bound to NifEN in an orientation evocative of its electron transfer role in nitrogenase. Our structure suggests that NifH is catalyzing a nucleotide-dependent electron transfer role in the L-cluster maturation. The ADP-AlF_4^- state captured by EM does not support a direct molybdate or homocitrate insertion, but the structure was solved in absence of the molybdate or homocitrate substrates, and further structures may illustrate the nature of these specific interactions.

The NifH binding mode we have identified with NifEN, in combination with the existing *Av* nitrogenase structures provides a starting point for identifying the determinants for the electron transfer binding mode of NifH. While there are no complex structures of the vanadium nitrogenase (VFe protein) and the MoFe protein from *Clostridium pasteurianum*, electrostatic potential map analysis of the structures reveals similar motifs observed in *Av* NifDK (MoFe protein) and NifEN. VFe protein[70] contains two Gly-Asp-Asp motifs within proximity to its intermediate cluster: Glyα144, Aspα145, Aspα146 and Glyβ121, Aspβ122, Aspβ123. Similarly, MoFe protein from *Clostridium pasteurianum*[97] presents two Gly-Asp-Asp motifs connecting its P-cluster to the solvent: Glyα151, Aspα152, Aspα153, and Glyβ112, Aspβ113, Aspβ114. Although the positions of the residues are not strictly conserved, the motifs are superimposable across *Av* NifDK, NifEN, VFe protein and *Cp* MoFe protein (Figure 43).

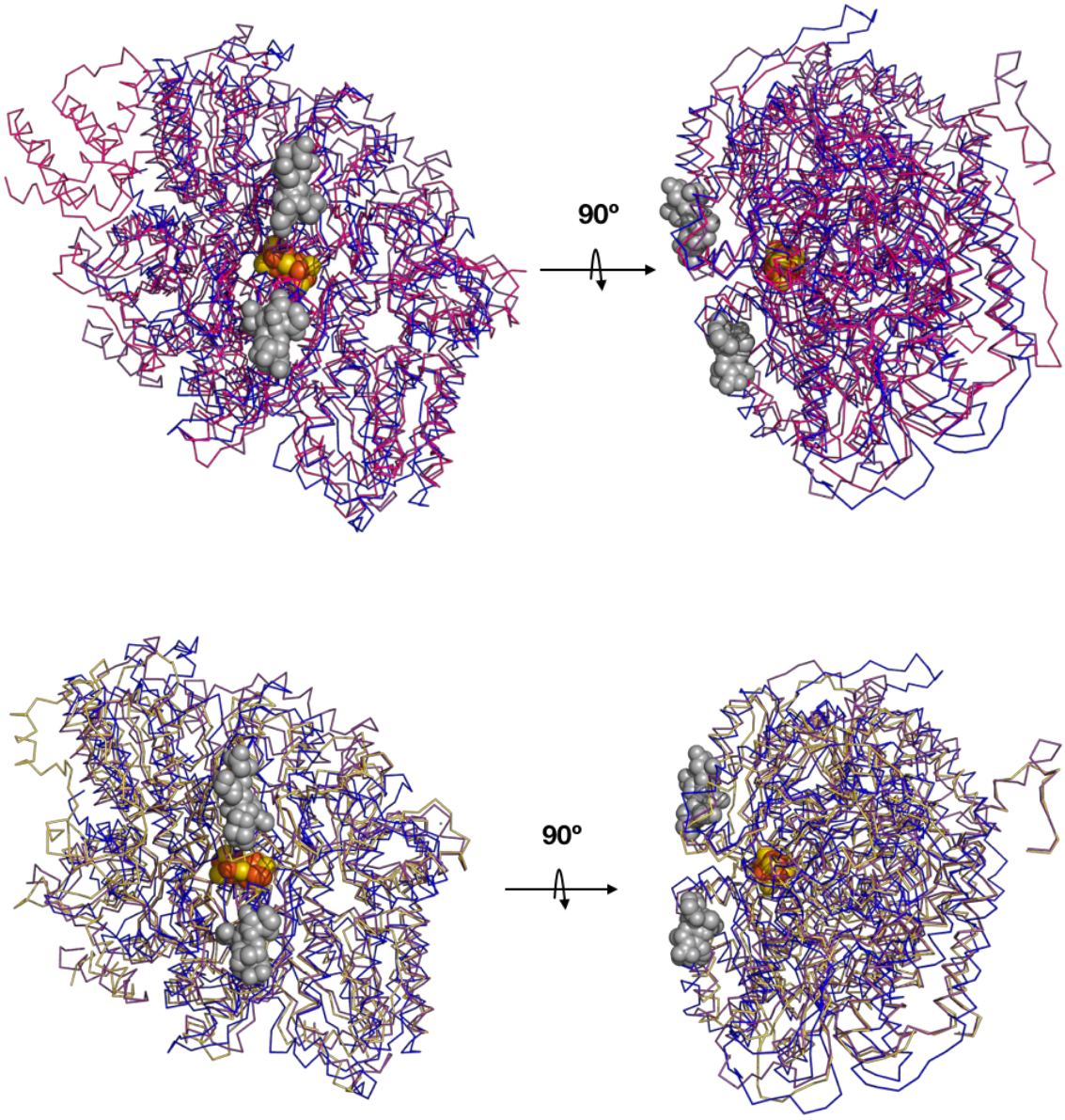


Figure 43. (Top) Alignment of NifEN from EM (blue), *Av* NifDK (3U7Q, purple) and VFe protein (5N6Y, pink). (Bottom) Alignment of NifEN from EM (blue), *Av* NifDK (3U7Q, purple) and *Cp* NifDK (4WES, yellow). GDD motifs in each structure are displayed as gray spheres. P-cluster is shown in yellow and orange spheres.

Future studies will focus on resolving the missing NifB component and clarifying the interactions of NifB, NifEN and NifH during FeMo-co maturation. The anaerobic cryo-

EM technique is a versatile tool for resolving difficult-to-crystalize complexes, and provides a promising avenue for determining labile maturation protein structures.

4.5 Experimental Section

The NifBEN fusion protein was isolated from *Azotobacter vinelandii* strains as described previously[87], and provided by Lee Rettberg at UC Irvine. The NifB protein construct is a truncation of the wild-type that lacks the X-domain at its C-terminus, a preparation that has been validated experimentally and supports cofactor maturation[98].

4.5.1 Cryo-EM Data Collection

In order to eliminate exposure to atmosphere while preparing samples, we modified an anaerobic glove bag (Coy Labs) to accommodate a Mark IV Vitrobot (FEI Thermo Fisher) and a gas-tight snorkel exhaust for removing flammable ethane gas build-up. A snorkel hood (vacuum line with a conical opening, designed to function as a local fume hood) was connected via copper tubing to a chemical-resistant diaphragm vacuum pump (VWR) to remove ethane vapor. Pre-condensed ethane/propane mix (37% ethane, 63% propane, Airgas) was introduced to the chamber for vitrification. Proteins were purified anaerobically through Schlenk line technique described previously[12] and then applied to grids for negative stain or cryo screening. In the modified anaerobic chamber, samples were diluted to 1 mg/mL with 50 mM Tris/Cl pH 7.5, 200 mM NaCl, and 5 mM dithionite directly before freezing.

For the NifDK structure, improving the orientation preference of the particles was a challenge. As part of a panel of additives, which included detergents, we identified an improvement in particle distribution (data not shown) in the presence of a non-specific 47.8 kDa Fab at a 1:1 molar ratio. The improvement in orientation preference could be due to competition at the air-water interface, but a thorough investigation of the effect was not performed. The sequence of 17b Fab (EC 65110) is as follows:

Light chain:
DIVMTQSPATLSVSPGERATLSCRASESVSSDLAWYQQKPGQAPRLLIYGASTRA
TGVPARFSGSGSGAEFTLTISSLQSEDFAVYYCQQYNNWPPRYTFGQGTRLEIKR
TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQES
VTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRG

Heavy chain:
EVQLVESGAEVKKPGSSVKVSCKASGDTFIRYSFTWVRQAPGQGLEWMGRIITIL
DVAHYAPHLQGRVTITADKSTSTVYLELRNLRSDDTAVYFCAGVYEGEADEGEY
DNNGFLKHWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPE
PVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPS
NTKVDKKVEPK

Quantifoil 200 mesh 2/2 copper grids (Electron Microscopy Services) were glow discharged using the PELCO easiGlow Glow Discharge System (Ted Pella). 3 μL of sample was applied to the grids before plunge-freezing using a Vitrobot Mark IV (FEI Thermo Fisher).

For the NifDK structure, 1405 micrographs were collected on a 200 keV Talos Arctica with a Falcon III direct electron detector at a pixel size of 0.863 Å, at a dose per frame of 1 e^-/A^2 (for a total dose of 40 e^-/A^2) with defocus values ranging from -0.8 μm – (-3 μm). From 808,446 particles, 47,340 particles were used in the final refinement.

For the NifDKH structure, 630 micrographs were collected on the 300 keV F30 Polara with a K2 Summit direct electron detector at a pixel size of 1.4 Å, at a dose around 1.5 e^-/A^2 with defocus values ranging from -1 μm – (-4 μm). From 328,026 particles, 43,745 particles were used in the final refinement.

For the NifENH structure, 2,246 micrographs were collected on a 200 keV Talos Arctica with a Falcon III direct electron detector at a pixel size of 1.15 Å, at a dose per frame of 1 e^-/A^2 with defocus values ranging from -0.8 μm – (-3 μm). From 571,048 particles, 14,947 particles were used in the final refinement.

4.5.2 Cryo-EM Data Processing

All movies were motion-corrected using MotionCor2[99], and the first and last frames were removed. CTF estimation was completed using CTFFIND4[100] on non-dose weighted micrographs. Particles were picked for autopicking templates in RELION[101] using dose weighted micrographs. Autopicked particles were submitted for 2D classification in cryoSPARC[102]. Well-resolved classes were used for *ab initio* model generation in cryoSPARC to avoid model bias, and refined. Resolution was determined using the Fourier shell correlation between two independent half-maps in cryoSPARC, which uses a 0.143 cutoff (Figure 44, 45, 46). Final maps were exported to PHENIX[103] for model building. Coordinates for NifEN (3PDI[84]) and NifH (1G5P[65]) were rigid-body docked into the map to generate a starting model using Chimera[67]. The resolution was insufficient for *de novo* model building, specifically in the N-terminal region of NifE, and the density remains unmodeled. Maps were converted using phenix.map_to_structure_factors and refined with phenix.real_space_refine. The validation report was generated using phenix.validation_cryoem. Protein volumes were calculated using the 3V server[104], and electrostatic potential maps were calculated using APBS in PyMol (Schrödinger LLC).

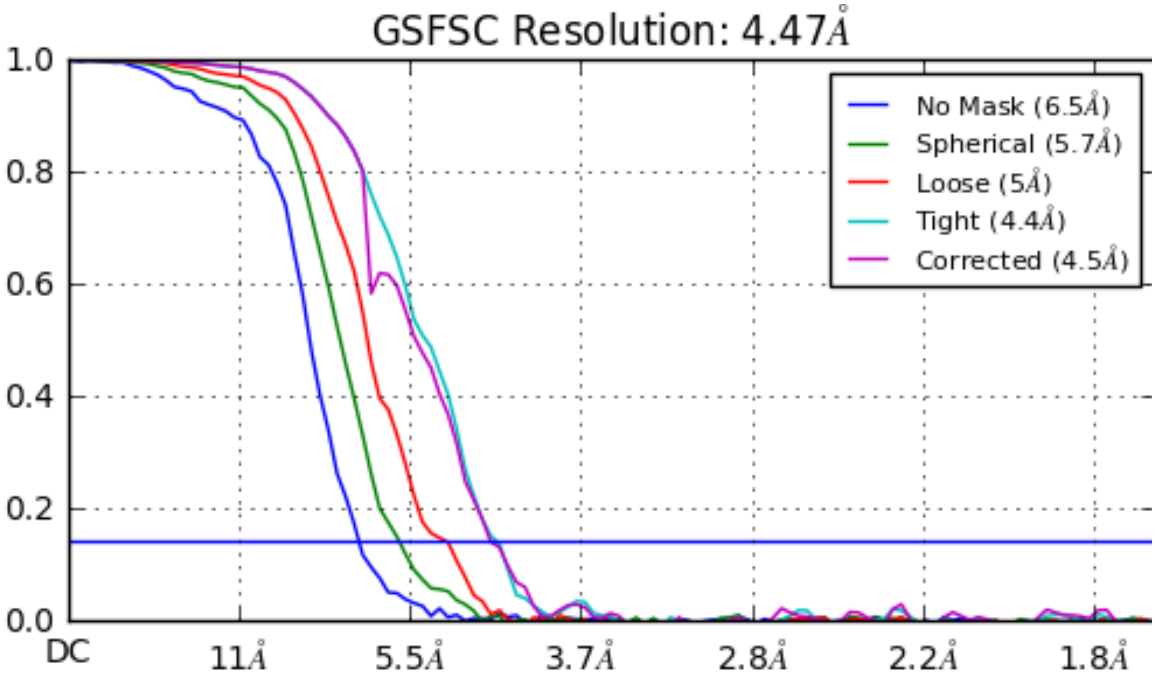


Figure 44. FSC curve for the cryo-EM NifDK structure.

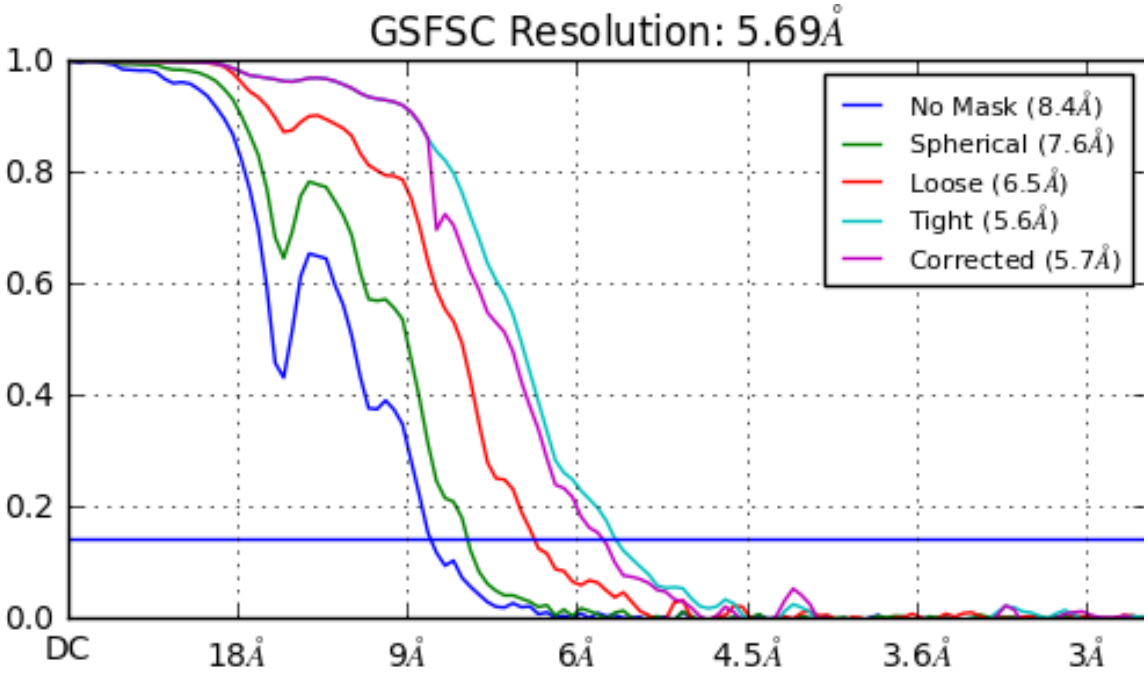


Figure 45. FSC curve for the cryo-EM NifDKH-ADPAlF_4^- structure.

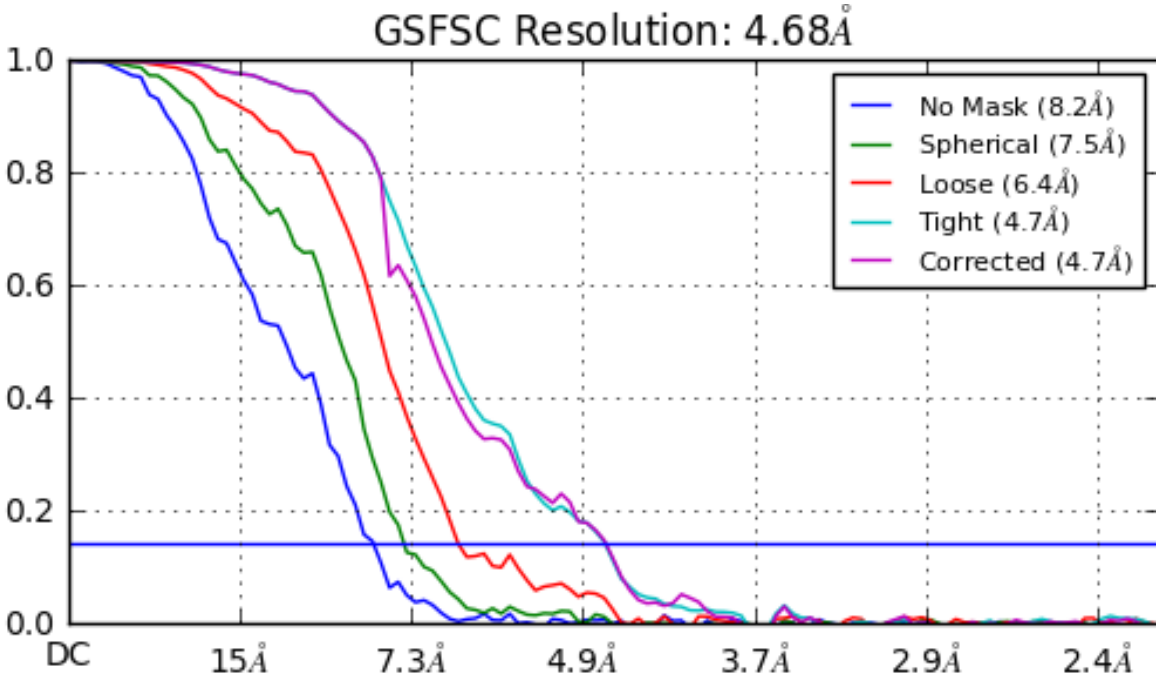


Figure 46. FSC curve for the cryo-EM NifENH- ADPAlF_4^- structure.

4.6 Acknowledgments

This work was performed in collaboration with Lee Rettberg, Dr. Markus Ribbe and Dr. Yilin Hu at the University of California at Irvine. The authors thank Dr. Haoqing Wang and Dr. Pamela Bjorkman for supplying the sacrificial Fab. We also thank Sara Weaver, Dr. Alasdair McDowall, Dr. Songye Chen, and Dr. Andrey Malyutin for technical expertise and helpful discussion. Data were collected at the Caltech Beckman Institute Resource Center for Transmission Electron Microscopy.

Chapter 5

STRUCTURAL STUDIES OF A P-CLUSTER MATURATION INTERMEDIATE: ΔBΔZ-MOFE-PROTEIN

5.1 Introduction

Nitrogenase catalyzes a crucial step in the global nitrogen cycle: the conversion of nitrogen to ammonia[47]. Nitrogenase is a two-component enzyme composed of the MoFe- and Fe-proteins, which are so-named for the metallocofactors they contain. The MoFe-protein (NifDK, using gene product nomenclature) contains two unique metal clusters: the FeMo-cofactor, the active site [7Fe:9S:C:Mo:*R*-homocitrate] cluster, and the P-cluster, an intermediate [8Fe:7S] cluster. The Fe-protein (NifH) contains a [4Fe:4S] cluster, and during nitrogen reduction, binds and transfers electrons to the P-cluster, which reduces the FeMo-cofactor (FeMo-co). The MoFe-protein is an $\alpha_2\beta_2$ tetramer with a C2 axis of symmetry, containing two P-clusters and two FeMo-cos. The P-cluster and FeMo-co require specialized cellular machinery for assembly, and the mechanism of cluster formation is an active area of research[24].

The stepwise generation of holo MoFe-protein has been studied extensively through deletion strains in *Azotobacter vinelandii*[23]. First, the P-clusters are assembled on the apo MoFe-protein in a stepwise fashion from [4Fe:4S] units. The initial step was identified from an Fe-protein, Δ*nifH*, deletion strain of *A. vinelandii*. In the absence of Fe-protein, Δ*nifH* MoFe-protein (ΔHNifDK) has spectroscopic features consistent with spin-coupled, individual [4Fe:4S][105]. A second strain with deletions of the radical SAM protein NifB (the progenitor of the central carbon in FeMo-co) and the maturation protein NifZ produced Δ*nifB* Δ*nifZ* MoFe-protein (ΔBΔZNifDK) and was identified as a model for the second step in the maturation process. By variable-temperature, variable-field magnetic circular dichroism (VTVH MCD), and electron paramagnetic resonance (EPR), ΔBΔZNifDK

contains features consistent with a mature [8Fe:7S] P-cluster and two [4Fe:4S][106,107], leading to a model of asymmetric P-cluster fusion. A crystal structure of the MoFe-protein isolated from the single Δ*nifB* deletion strain contained two mature P-clusters and was FeMo-co-deficient[108], representing the final stage of P-cluster maturation. Interestingly, the ΔBNifDK structure adopted an open conformation in the α-subunit to accommodate the mature FeMo-co. Unlike some copper-containing proteins, which utilize a subunit-subunit exchange mechanism to generate mature holoprotein[109], ΔBNifDK is pre-arranged to receive and presumably bury the cofactor. To contextualize the second maturation step represented by the isolated ΔBΔZNifDK, structural studies are necessary. In order to visualize the conformational state of ΔBΔZNifDK, we used anaerobic cryo-electron microscopy (cryo-EM) to solve a 4.4-Å resolution structure.

5.2 Cryo-EM Structure of ΔBΔZNifDK

Using our anaerobic grid preparation technique described previously, ΔBΔZNifDK samples were screened and optimized for single particle analysis. We collected a dataset of 1408 micrographs 200 keV Talos Arctica, resulting in a 4.4-Å resolution map (see Experimental Section, Figure 47).

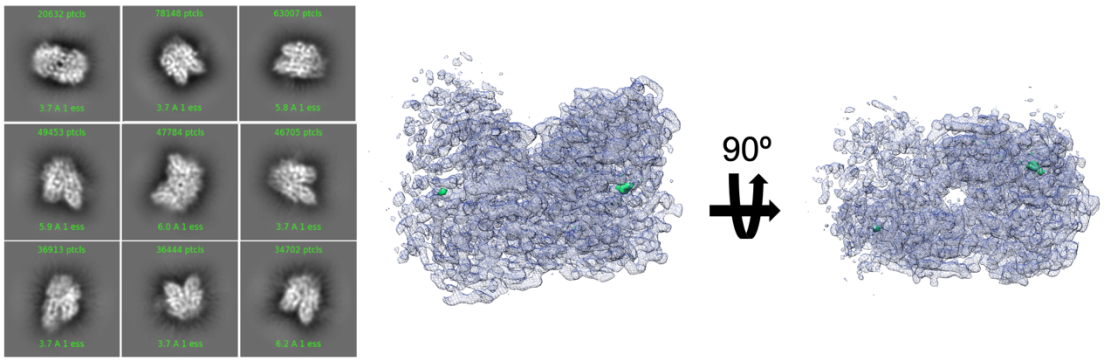


Figure 47. Anaerobic cryo-EM structure of ΔBΔZNifDK. (Left) Representative 2D classes from cryoSPARC. (Right) Map in blue, with metal clusters in green. The blue map is contoured to a threshold of 1.39 in Chimera, corresponding to a calculated 4.3 σ (standard deviation = 0.32). The green map is contoured to a threshold of 4.91, or 15 σ.

From the map, two metal clusters were identified: an elongated density that can accommodate a [8Fe:7S] P-cluster, and a smaller peak that can accommodate a single [4Fe:4S] (Figure 48, 49).

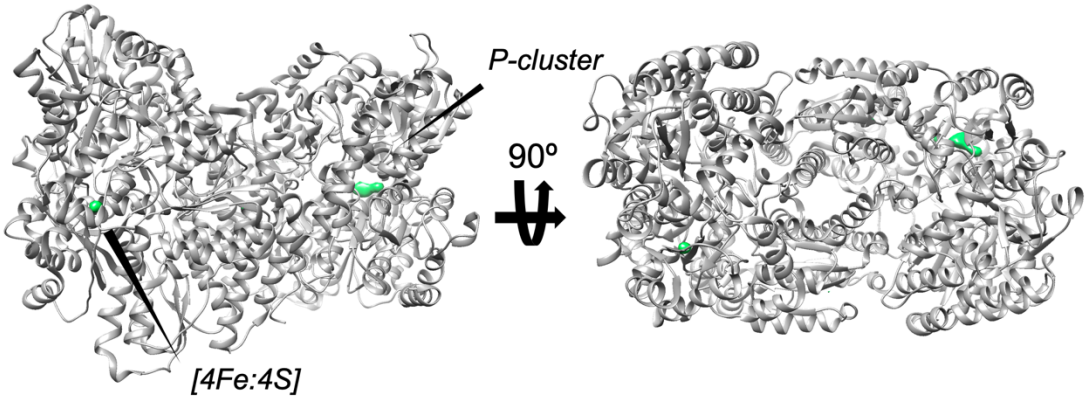


Figure 48. Protein model with the clusters indicated in green.

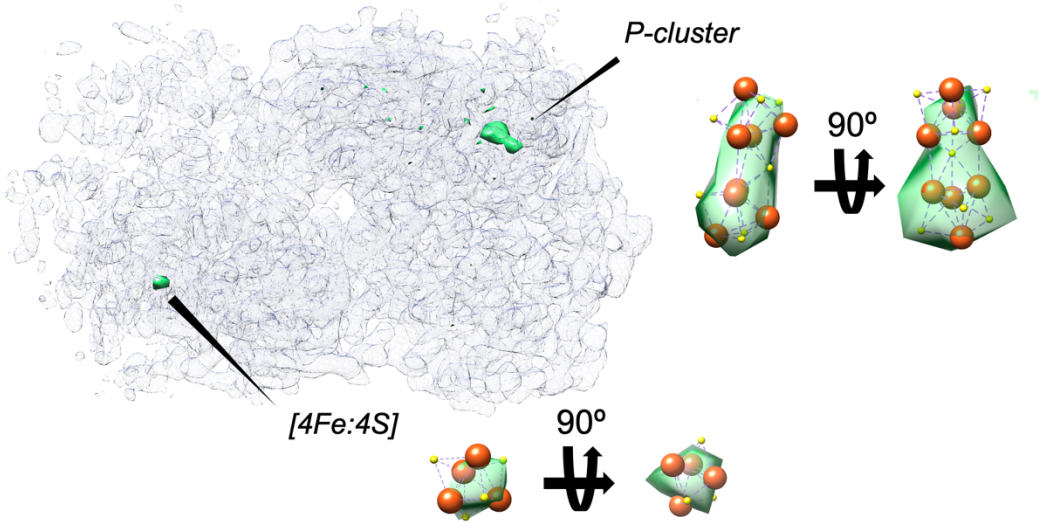


Figure 49. View of two cofactors identified at high map contour (the map is contoured equivalently for both cofactors). (Top) P-cluster is modeled into the elongated density at the αβ-subunit bridge. (Bottom) [4Fe:4S] is modeled into a smaller peak, which does not elongate upon lower contour.

The two peaks of density occur at the αβ-subunit interface, where the P-cluster is positioned in the matured MoFe-protein[12]. The subunits coordinating the elongated density are fully occupied, but the junction between the smaller peak assigned to [4Fe:4S] is disordered (Figure 49). The entire α-subunit is partially occupied (Figure 50), leading to a break in symmetry across the $\alpha_2\beta_2$ tetramer. A minor population (~10% of total particles) yielded a ~4.6 Å resolution map in an $\alpha\beta_2$ conformation that is completely missing the α-subunit (not shown) but is otherwise identical to the 4.4-Å resolution map. A map that contained a fully occupied α-subunit could not be obtained, suggesting that an α-subunit is missing or disordered in a fraction of the ΔBΔZNifDK molecules and could not be resolved.

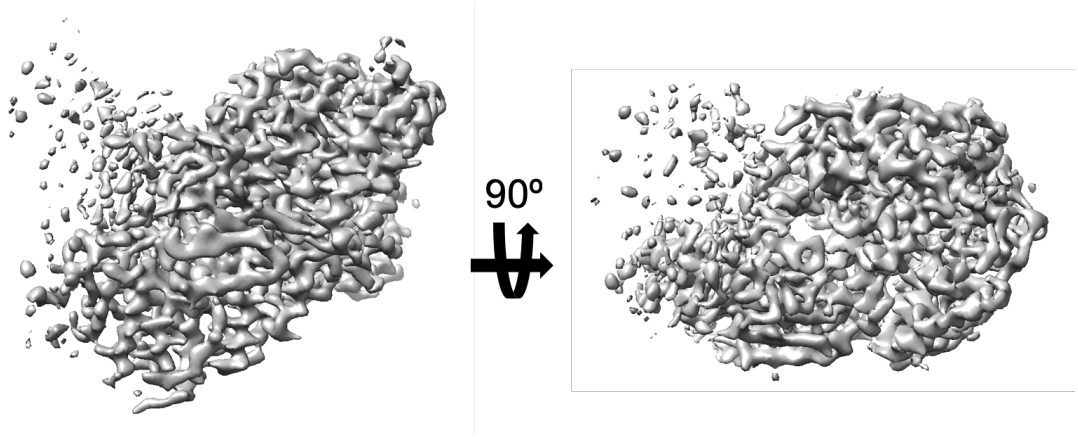


Figure 50. The incompletely occupied α-subunit becomes apparent at a lower map contour.

We real-space refined a model of the protein (Table 8), using the crystal structure of the FeMo-co-deficient ΔBNifDK as starting coordinates (Figure 48).

Bonds (RMSD)	
Length (Å)	0.007
Angles (°)	1.302
MolProbity Score	2.35
Clash Score	15.46
Ramachandran plot (%)	
Outliers	0.22

Favored	84.91
Rotamer outliers (%)	0.63

Table 8. Validation report for the ΔBΔZNifDK model.

The global Cα-RMSD between the refined model and the starting coordinates is 1.37 Å. Major deviations between the cofactor-deficient structures are localized to the α-subunit, although the major helix in the β-subunit (βTyr487-βArg510) that comprises the core of the tetramer has displacements on the order of 2-3 Å. The largest motions in the α-subunit are observed in the helices that surround the region poised to receive the FeMo-co: αArg277-αTyr290, αThr303-αPhe316, and αSer319-αPro344. The solvent-facing helices (including the N-terminus) that block the FeMo-co channel in the holo MoFe protein structure are disordered, as observed in the ΔBNifDK structure (αMet4-αAsn49, and αAsp385-αLys392). These regions were similarly unmodeled in the ΔBNifDK crystal structure, and reveal a positively-charged funnel hypothesized to receive the FeMo-co[108]. Between the crystal structures, the RMSD between the α-subunit of the high-resolution MoFe protein and ΔBNifDK is 0.39 Å. Overall, the RMSD between the refined model and the high-resolution MoFe protein crystal structure is 1.38 Å. Omitting the α-subunit, the Cα positions of the holo structure and ΔBΔZNifDK are highly similar, but deviations do occur in the surface helix βLys400-βAla411 and helix-loop βGlu202-βLys222. The displacement of the helices ~ 2 Å at a site distal to and opposite from the FeMo-co bindings site suggests a region of flexibility could be related to P-cluster maturation (Figure 51).

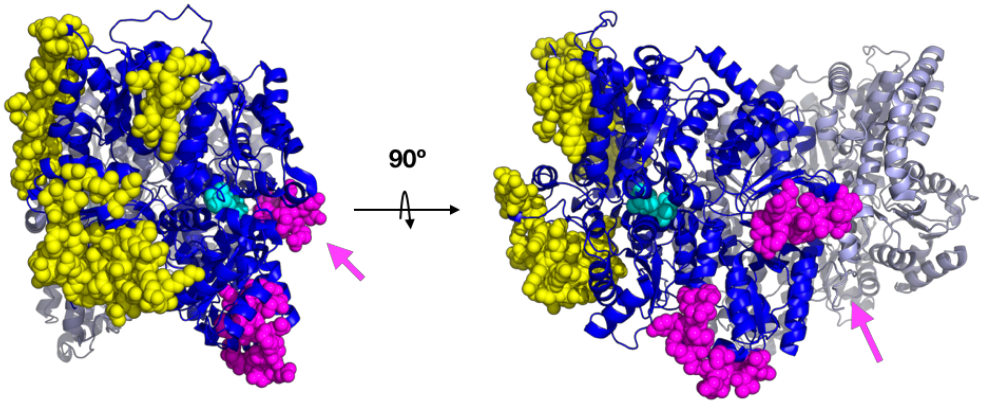


Figure 51. Regions of high Cα RMSD in the ΔBΔZNifDK compared to holo NifDK (blue). One αβ dimer is shown in dark blue, the second αβ dimer is shown in gray. The regions of high CαRMSD (or entirely disordered) are shown in yellow, and are in close proximity to the (unoccupied) FeMo-co binding site. The regions of higher RMSD are shown in magenta, and are found on a distinct face of NifDK, in the β-subunit. Mangenta arrows indicate the helix region βLys400-βIle409. The mature P-cluster is shown in cyan.

The three lysines at positions β400, β403 and β404 generate a small basic, electropositive patch that could be related to P-cluster maturation. However, rearrangements observed in the ΔBΔZNifDK structure, with and without the α-subunit modeled, do not present any dramatically positively-charged regions, as is found in the area surrounding the FeMo-co binding site.

5.3 Discussion

The maturation of the elaborate nitrogenase cofactors, the FeMo-co ([7Fe:9S:C:Mo:*R*-homocitrate]) and the P-cluster ([8Fe:7S]), is a critical cellular process that is an intriguing synthetic challenge, as well as a delicate logistical operation involving 15 gene products[108] and labile metals. The initial steps in the process to mature apo-MoFe-protein into a functional component of the nitrogenase complex are dominated by the step-wise generation of the P-clusters. Unlike the FeMo-co, which is matured on a separate scaffold, NifEN[84], the P-clusters are formed directly in the apo-MoFe-protein from [4Fe:4S] units. Spectroscopy has provided evidence of a mixture of [4Fe:4S] units and matured [8Fe:7S] clusters based on proteins isolated from *A. vinelandii* strains containing maturation protein

deletions. The earliest step in the process captured crystallographically contains fully matured P-clusters and is FeMo-co deficient.

Here, we present structural evidence of a MoFe-protein maturation intermediate immediately preceding the FeMo-co deficient ΔBNifDK. We solved a 4.4-Å resolution cryo-EM structure of the MoFe-protein isolated from a Δ*nifB* Δ*nifZ* strain. Our structure is consistent with one matured P-cluster, and one [4Fe:4S] unit bridging the αβ-subunits. The α-subunit is disordered and does not exclude the possibility of a second [4Fe:4S]. The deletion of the NifZ gene product yields an asymmetric maturation of the P-cluster in the NifB deletion background, raising important questions concerning the mechanism of fusion. The lability of the α-subunit may indicate a structural role of the coupled 2[4Fe:4S] precursor or the mature [P-cluster] in structurally stabilizing the αβ interface.

Our results provide structural evidence of the spectroscopic prediction of asymmetric P-cluster maturation, and identify disorder in the α-subunit that does not contained the matured P-cluster. Further experiments will be needed to explore the origin of the asymmetric 2[4Fe:4S] fusion, including how the partially-matured MoFe-protein is identified for further processing.

5.4 Experimental Section

ΔBΔZNifDK was isolated from genetically modified *Azotobacter vinelandii* strains as described previously[27] and prepared by Lee Rettberg at UC Irvine.

5.4.1 Cryo-EM Data Collection

Grids were prepared anaerobically as described previously. Samples were diluted to 1 mg/mL with 50 mM Tris/Cl pH 7.5, 200 mM NaCl and 5 mM dithionite directly before freezing. Quantifoil 200 mesh 2/2 copper grids (Electron Microscopy Services) were glow discharged using the PELCO easiGlow Glow Discharge System (Ted Pella). 3 μL of

sample was applied to the grids before plunge-freezing using a Vitrobot Mark IV (FEI Thermo Fisher). 1408 micrographs were collected on a 200 keV Talos Arctica with a Falcon III direct detector at a pixel size of 1.15 Å. Movies were collected at a dose per frame of 1 e^-/A^2 (for a total dose of 40 e^-/A^2) with defocus values ranging from -0.8 μm – (-3.5 μm). 2.8 M particles were selected for processing. 602,236 particles contributed to the 4.4-Å resolution structure of the tetramer ΔBΔZNifDK with the α-subunit incompletely occupied. 93,292 particles were used in the final refinement of the 4.6- Å resolution trimer ($\alpha\beta_2$) of ΔBΔZNifDK.

5.4.2 Cryo-EM Data Processing

All movies were motion-corrected using MotionCor2[99], removing the first and last frame due to beam-induced motion and radiation damage, respectively. CTF estimation was performed using non-dose corrected micrographs with CTFFIND4[100]. Autopicking was performed using RELION[101], and particle stacks were exported to cryoSPARC[102] for classification, *ab initio* model generation and refinement. Final resolution was determined in cryoSPARC by Fourier shell correlation between two independent half-datasets, using the 0.143 cutoff (Figure 52, 53). Model refinement was performed with PHENIX[103], using the holo MoFe-protein structure[12] and the FeMo-co-deficient structure[108]. Maps were converted using phenix.map_to_structure_factors, refined with phenix.real_space_refine and validated using phenix.validation_cryoem.

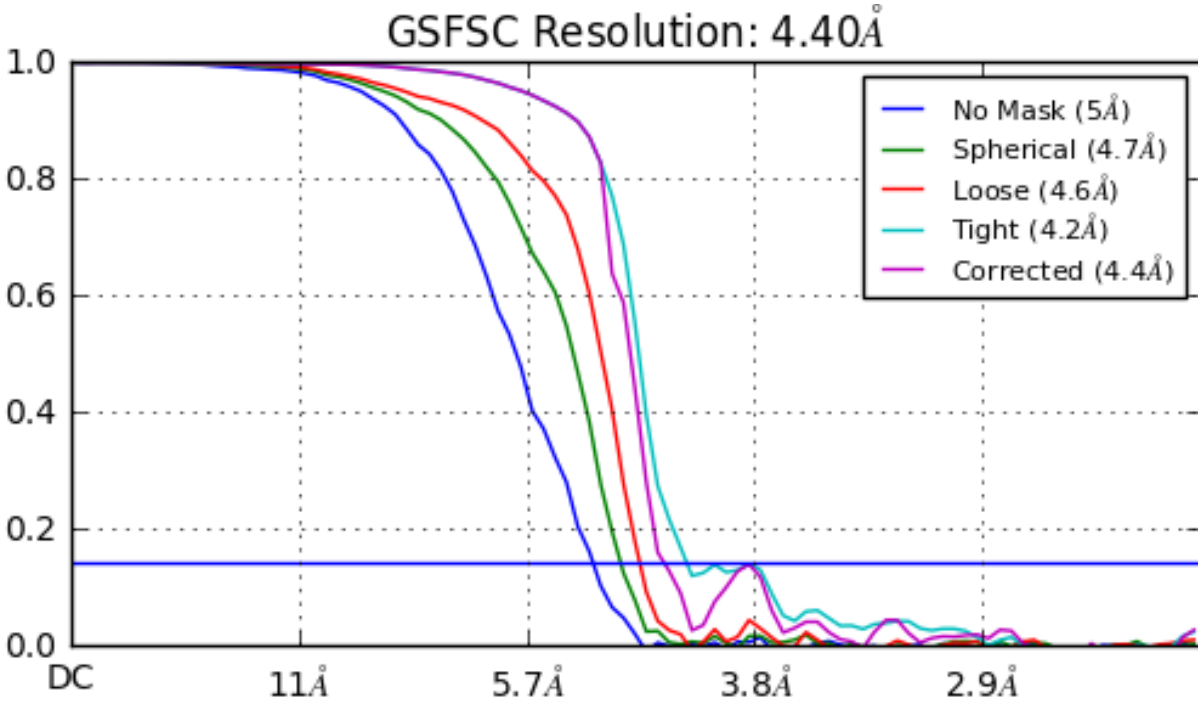


Figure 52. FSC curve for the tetramer ΔBΔZNifDK (with incompletely occupied α-subunit).

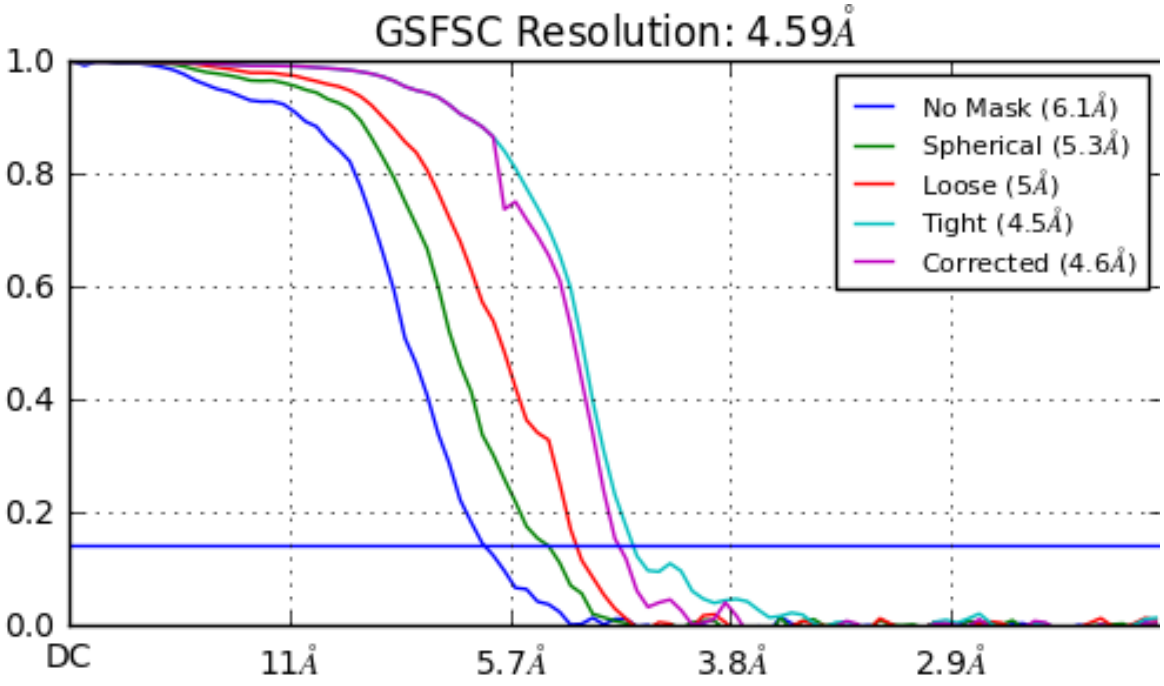


Figure 53. FSC curve for the trimer ΔBΔZNifDK (with missing α-subunit).

5.5 Acknowledgments

This work was performed in collaboration with Lee Rettberg, Dr. Markus Ribbe and Dr. Yilin Hu at the University of California at Irvine. We thank Dr. Songye Chen and Dr. Andrey Malyutin for technical expertise and helpful discussion, and the University of Michigan CryoEM Workshop for data processing expertise. Data were collected at the Caltech Beckman Institute Resource Center for Transmission Electron Microscopy.

REFERENCES

[1] M. S. Booth, J. M. Stark, E. Rastetter, *Ecological monographs* **2005**, *75*, 139–157.
[2] L. E. Mortenson, R. N. Thorneley, *Annual review of biochemistry* **1979**, *48*, 387–418.
[3] B. S. Chauhan, *Principles of Biochemistry and Biophysics*, Firewall Media, **2008**.
[4] Y.-R. Luo, *Comprehensive Handbook of Chemical Bond Energies*, CRC Press, **2007**.
[5] G. D. Watt, K. R. N. Reddy, *Journal of inorganic biochemistry* **1994**, *53*, 281–294.
[6] A. Braaksma, H. Haaker, H. J. Grande, C. Veeger, *European journal of biochemistry* **1982**, *121*, 483–491.
[7] M. Guo, F. Sulc, M. W. Ribbe, P. J. Farmer, B. K. Burgess, *Journal of the American Chemical Society* **2002**, *124*, 12100–12101.
[8] J. Deistung, R. N. Thorneley, *Biochemical Journal* **1986**, *239*, 69–75.
[9] J. B. Howard, D. C. Rees, *Proceedings of the National Academy of Sciences* **2006**, *103*, 17088–17093.
[10] B. J. Hales, E. E. Case, J. E. Morningstar, M. F. Dzeda, L. A. Mauterer, *Biochemistry* **1986**, *25*, 7251–7255.
[11] B. Schmid, O. Einsle, H.-J. Chiu, A. Willing, M. Yoshida, J. B. Howard, D. C. Rees, *Biochemistry* **2002**, *41*, 15557–15565.
[12] T. Spatzal, M. Aksoyoglu, L. Zhang, S. L. Andrade, E. Schleicher, S. Weber, D. C. Rees, O. Einsle, *Science* **2011**, *334*, 940–940.
[13] T. Spatzal, K. A. Perez, O. Einsle, J. B. Howard, D. C. Rees, *Science* **2014**, *345*, 1620–1623.
[14] T. Spatzal, K. A. Perez, J. B. Howard, D. C. Rees, *Elife* **2015**, *4*, e11620.
[15] Z.-Y. Yang, R. Ledbetter, S. Shaw, N. Pence, M. Tokmina-Lukaszewska, B. Eilers, Q. Guo, N. Pokhrel, V. L. Cash, D. R. Dean, et al., *Biochemistry* **2016**, *55*, 3625–3635.
[16] R. N. Thorneley, D. J. Lowe, John Wiley, New York, NY USA, **1985**.
[17] S. Hill, E. P. Kavanagh, *Journal of bacteriology* **1980**, *141*, 470–475.
[18] J. A. Erickson, A. C. Nyborg, J. L. Johnson, S. M. Truscott, A. Gunn, F. R. Nordmeyer, G. D. Watt, *Biochemistry* **1999**, *38*, 14279–14285.
[19] J. C. Setubal, P. dos Santos, B. S. Goldman, H. Ertesv\a ag, G. Espin, L. M. Rubio, S. Valla, N. F. Almeida, D. Balasubramanian, L. Cromes, *Journal of bacteriology* **2009**, *191*, 4534–4545.
[20] P. A. Lindahl, E. P. Day, T. A. Kent, W. H. Orme-Johnson, E. Münck, *Journal of Biological Chemistry* **1985**, *260*, 11160–11173.
[21] P. A. Lindahl, N. J. Gorelick, E. Münck, W. H. Orme-Johnson, *Journal of Biological Chemistry* **1987**, *262*, 14945–14953.
[22] S. J. Yoo, H. C. Angove, B. K. Burgess, M. P. Hendrich, E. Münck, *Journal of the American Chemical Society* **1999**, *121*, 2534–2545.

[23] Y. Hu, M. W. Ribbe, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2013**, *1827*, 1112–1122.
[24] Y. Hu, M. W. Ribbe, *Annual review of biochemistry* **2016**, *85*, 455–483.
[25] N. S. Sickerman, M. W. Ribbe, Y. Hu, *Accounts of chemical research* **2017**, *50*, 2834–2841.
[26] K. Rupnik, C. C. Lee, J. A. Wiig, Y. Hu, M. W. Ribbe, B. J. Hales, *Biochemistry* **2014**, *53*, 1108–1116.
[27] Y. Hu, A. W. Fay, P. C. Dos Santos, F. Naderi, M. W. Ribbe, *Journal of Biological Chemistry* **2004**.
[28] Y. Hu, M. C. Corbett, A. W. Fay, J. A. Webber, K. O. Hodgson, B. Hedman, M. W. Ribbe, *Proceedings of the National Academy of Sciences* **2006**, *103*, 17125–17130.
[29] M. M. Georgiadis, H. Komiya, P. Chakrabarti, D. Woo, J. J. Kornuc, D. C. Rees, *Science* **1992**, *257*, 1653–1659.
[30] B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, **2009**.
[31] G. Rhodes, *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*, Elsevier, **2010**.
[32] P. Coppens, T. N. Guru Row, P. Leung, E. D. Stevens, P. t Becker, Y. W. Yang, *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **1979**, *35*, 63–72.
[33] G. Wu, Y. Zhang, L. Ribaud, P. Coppens, C. Wilson, B. B. Iversen, F. K. Larsen, *Inorganic chemistry* **1998**, *37*, 6078–6083.
[34] O. Einsle, S. L. Andrade, H. Dobbek, J. Meyer, D. C. Rees, *Journal of the American Chemical Society* **2007**, *129*, 2210–2211.
[35] J. R. Pilbrow, G. R. Hanson, in *Methods in Enzymology*, Elsevier, **1993**, pp. 330–353.
[36] G. Palmer, *The Electron Paramagnetic Resonance of Metalloproteins*, Portland Press Limited, **1985**.
[37] W. R. Hagen, *Biomolecular EPR Spectroscopy*, CRC Press, **2008**.
[38] Y. Cheng, N. Grigorieff, P. A. Penczek, T. Walz, *Cell* **2015**, *161*, 438–449.
[39] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, Oxford University Press, **2006**.
[40] R. Fernandez-Leiro, S. H. W. Scheres, *Nature* **2016**, *537*, 339–346.
[41] J. C. H. Spence, *Structural Dynamics* **2017**, *4*, 044027.
[42] M. J. Dobro, L. A. Melanson, G. J. Jensen, A. W. McDowall, in *Methods in Enzymology* (Ed.: G.J. Jensen), Academic Press, **2010**, pp. 63–82.
[43] J. Dubochet, A. W. McDowall, *Journal of Microscopy* **1981**, *124*, 3–4.
[44] R. N. Clough, G. Moldovan, A. I. Kirkland, *J. Phys.: Conf. Ser.* **2014**, *522*, 012046.
[45] M. Kuijper, G. van Hoften, B. Janssen, R. Geurink, S. De Carlo, M. Vos, G. van Duinen, B. van Haeringen, M. Storms, *Journal of Structural Biology* **2015**, *192*, 179–187.

[46] T. J. Lowery, P. E. Wilson, B. Zhang, J. Bunker, R. G. Harrison, A. C. Nyborg, D. Thiriot, G. D. Watt, *Proceedings of the National Academy of Sciences* **2006**, *103*, 17131–17136.
[47] B. K. Burgess, D. J. Lowe, *Chemical reviews* **1996**, *96*, 2983–3012.
[48] L. C. Seefeldt, B. M. Hoffman, J. W. Peters, S. Raugei, D. N. Beratan, E. Antony, D. R. Dean, *Accounts of chemical research* **2018**, *51*, 2179–2186.
[49] R. P. Hausinger, J. B. Howard, *Journal of Biological Chemistry* **1983**, *258*, 13486–13492.
[50] H. C. Angove, S. J. Yoo, E. Münck, B. K. Burgess, *Journal of Biological Chemistry* **1998**, *273*, 26330–26337.
[51] A. C. Nyborg, J. L. Johnson, A. Gunn, G. D. Watt, *Journal of Biological Chemistry* **2000**, *275*, 39307–39312.
[52] W. G. Zumft, G. Palmer, L. E. Mortenson, *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1973**, *292*, 413–421.
[53] M. J. Ryle, W. N. Lanzilotta, L. C. Seefeldt, R. C. Scarrow, G. M. Jensen, *Journal of Biological Chemistry* **1996**, *271*, 1551–1557.
[54] P. J. Stephens, C. E. McKenna, B. E. Smith, H. T. Nguyen, M. C. McKenna, A. J. Thomson, F. Devlin, J. B. Jones, *Proceedings of the National Academy of Sciences* **1979**, *76*, 2585–2589.
[55] T. L. Deits, J. B. Howard, *Journal of Biological Chemistry* **1989**, *264*, 6619–6628.
[56] G. A. Walker, L. E. Mortenson, *Biochemistry* **1974**, *13*, 2382–2388.
[57] T. Ljones, R. H. Burris, *Biochemistry* **1978**, *17*, 1866–1872.
[58] H. C. Angove, S. J. Yoo, B. K. Burgess, E. Münck, *Journal of the American Chemical Society* **1997**, *119*, 8730–8731.
[59] H. Beinert, R. H. Holm, E. Münck, *Science* **1997**, *277*, 653–659.
[60] L. Noodleman, C. Y. Peng, D. A. Case, J.-M. Mouesca, *Coordination Chemistry Reviews* **1995**, *144*, 199–244.
[61] T. Spatzal, J. Schlesier, E.-M. Burger, D. Sippel, L. Zhang, S. L. Andrade, D. C. Rees, O. Einsle, *Nature communications* **2016**, *7*, 10902.
[62] K. B. Musgrave, H. C. Angove, B. K. Burgess, B. Hedman, K. O. Hodgson, *Journal of the American Chemical Society* **1998**, *120*, 5325–5326.
[63] J. K. Kowalska, A. W. Hahn, A. Albers, C. E. Schiewer, R. Bjornsson, F. A. Lima, F. Meyer, S. DeBeer, *Inorganic chemistry* **2016**, *55*, 4485–4497.
[64] L. Zhang, J. T. Kaiser, G. Meloni, K.-Y. Yang, T. Spatzal, S. L. Andrade, O. Einsle, J. B. Howard, D. C. Rees, *Angewandte Chemie International Edition* **2013**, *52*, 10529–10532.
[65] P. Strop, P. M. Takahara, H.-J. Chiu, H. C. Angove, B. K. Burgess, D. C. Rees, *Biochemistry* **2001**, *40*, 651–656.
[66] S. B. Jang, L. C. Seefeldt, J. W. Peters, *Biochemistry* **2000**, *39*, 14745–14752.
[67] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *Journal of computational chemistry* **2004**, *25*, 1605–1612.
[68] W. G. Zumft, L. E. Mortenson, G. Palmer, *European journal of biochemistry* **1974**, *46*, 525–535.
[69] D. C. Harris, *Quantitative Chemical Analysis*, Macmillan, **2010**.

[70] D. Sippel, O. Einsle, *Nature chemical biology* **2017**, *13*, 956.
[71] D. Sippel, M. Rohde, J. Netzer, C. Trncik, J. Gies, K. Grunau, I. Djurdjevic, L. Decamps, S. L. Andrade, O. Einsle, *Science* **2018**, *359*, 1484–1489.
[72] D. Wolle, C. Kim, D. Dean, J. B. Howard, *Journal of Biological Chemistry* **1992**, *267*, 3667–3673.
[73] W. Kabsch, *Acta Crystallographica Section D* **2010**, *66*, 133–144.
[74] P. Evans, *Acta Crystallographica Section D: Biological Crystallography* **2006**, *62*, 72–82.
[75] M. D. Winn, G. N. Murshudov, M. Z. Papiz, in *Methods in Enzymology*, Elsevier, **2003**, pp. 300–321.
[76] P. Emsley, K. Cowtan, *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60*, 2126–2132.
[77] D. Rutkowska-Zbik, M. Witko, L. Fiedor, *Journal of molecular modeling* **2013**, *19*, 4661–4667.
[78] Y.-W. Xu, S. Moréra, J. Janin, J. Cherfils, *Proceedings of the National Academy of Sciences* **1997**, *94*, 3579–3583.
[79] J. L. Vilas, N. Tabassum, J. Mota, D. Maluenda, A. Jiménez-Moreno, T. Majtner, J. M. Carazo, S. T. Acton, C. O. S. Sorzano, *Current Opinion in Structural Biology* **2018**, *52*, 127–145.
[80] J. M. Schuller, J. A. Birrell, H. Tanaka, T. Konuma, H. Wulfhorst, N. Cox, S. K. Schuller, J. Thiemann, W. Lubitz, P. Sétif, et al., *Science* **2018**, eaau3613.
[81] C. K. Rofer-DePoorter, *Chemical Reviews* **1981**, *81*, 447–474.
[82] Y. Hu, C. C. Lee, M. W. Ribbe, *Science* **2011**, *333*, 753–755.
[83] C. C. Lee, Y. Hu, M. W. Ribbe, *Angewandte Chemie* **2015**, *127*, 1235–1238.
[84] J. T. Kaiser, Y. Hu, J. A. Wiig, D. C. Rees, M. W. Ribbe, *Science* **2011**, *331*, 91–94.
[85] A. W. Fay, M. A. Blank, C. C. Lee, Y. Hu, K. O. Hodgson, B. Hedman, M. W. Ribbe, *Angewandte Chemie International Edition* **2011**, *50*, 7787–7790.
[86] J. A. Wiig, Y. Hu, C. C. Lee, M. W. Ribbe, *Science* **2012**, *337*, 1672–1675.
[87] J. A. Wiig, Y. Hu, M. W. Ribbe, *Proceedings of the National Academy of Sciences* **2011**, *108*, 8623–8627.
[88] K. M. Lancaster, Y. Hu, U. Bergmann, M. W. Ribbe, S. DeBeer, *Journal of the American Chemical Society* **2013**, *135*, 610–612.
[89] R. S. Pantelic, J. C. Meyer, U. Kaiser, W. Baumeister, J. M. Plitzko, *Journal of Structural Biology* **2010**, *170*, 152–156.
[90] G. Yu, K. Li, P. Huang, X. Jiang, W. Jiang, *Structure* **2016**, *24*, 1984–1990.
[91] S. Chowdhury, S. A. Ketcham, T. A. Schroer, G. C. Lander, *Nature Structural & Molecular Biology* **2015**, *22*, 345–347.
[92] J. Snijder, A. J. Borst, A. Dosey, A. C. Walls, A. Burrell, V. S. Reddy, J. M. Kollman, D. Veesler, *Journal of Structural Biology* **2017**, *198*, 38–42.
[93] P. Hänzelmann, H. Schindelin, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12870–12875.
[94] F. A. Tezcan, J. T. Kaiser, J. B. Howard, D. C. Rees, *J. Am. Chem. Soc.* **2015**, *137*, 146–149.

[95] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res* **2004**, *32*, W665–W667.
[96] R. Bjornsson, F. A. Lima, T. Spatzal, T. Weyhermüller, P. Glatzel, E. Bill, O. Einsle, F. Neese, S. DeBeer, *Chemical Science* **2014**, *5*, 3096–3103.
[97] L.-M. Zhang, C. N. Morrison, J. T. Kaiser, D. C. Rees, *Acta Cryst D* **2015**, *71*, 274–282.
[98] A. W. Fay, J. A. Wiig, C. C. Lee, Y. Hu, *PNAS* **2015**, *112*, 14829–14833.
[99] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, *Nature Methods* **2017**, *14*, 331–332.
[100] A. Rohou, N. Grigorieff, *Journal of Structural Biology* **2015**, *192*, 216–221.
[101] S. H. Scheres, *Journal of structural biology* **2012**, *180*, 519–530.
[102] A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nature methods* **2017**, *14*, 290.
[103] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, et al., *Acta Cryst D* **2010**, *66*, 213–221.
[104] N. R. Voss, M. Gerstein, *Nucleic Acids Res* **2010**, *38*, W555–W562.
[105] M. C. Corbett, Y. Hu, F. Naderi, M. W. Ribbe, B. Hedman, K. O. Hodgson, *J. Biol. Chem.* **2004**, *279*, 28276–28282.
[106] M. S. Cotton, K. Rupnik, R. B. Broach, Y. Hu, A. W. Fay, M. W. Ribbe, B. J. Hales, *J. Am. Chem. Soc.* **2009**, *131*, 4558–4559.
[107] K. Rupnik, C. C. Lee, Y. Hu, M. W. Ribbe, B. J. Hales, *Inorg. Chem.* **2018**, *57*, 4719–4725.
[108] B. Schmid, M. W. Ribbe, O. Einsle, M. Yoshida, L. M. Thomas, D. R. Dean, D. C. Rees, B. K. Burgess, *Science* **2002**, *296*, 352–356.
[109] Z. Zhou, Y. Hashimoto, K. Shiraki, M. Kobayashi, *PNAS* **2008**, *105*, 14849–14854.