Insights into the Mechanism of Biological Nitrogen Fixation through Characterization of the Nitrogenase Molybdenum-Iron Protein

Thesis by

Christine N. Morrison

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2017

(Defended 1 May 2017)

Christine N. Morrison
ORCID: 0000-0002-4180-8407

ACKNOWLEDGMENTS

"Science does not happen in a vacuum", as the saying goes. The work presented in this thesis would not have been possible without the support, time, and energy of many people. My thesis committee, including Professor Douglas C. Rees (my research advisor), Professor Jonas C. Peters (the chair of my committee), Professor William M. Clemons, and Professor Harry B. Gray, provided unwavering support and constructive feedback throughout my Ph.D. Many skills I learned from Dr. Limei Zhang, who mentored me during the first few years of my Ph.D. I also worked with Dr. Julie Hoy and Dr. Jens Kaiser to learn X-ray crystallography – a core technique of my thesis. "Team Nitro" (Dr. Limei Zhang, Dr. Thomas Spatzal, Dr. Kathryn Perez, Dr. Helen Segal, Belinda Wenke, Renee Arias, and Camille McAvoy) assisted with experimental design and execution, and they made lab a fun place to be. The Rees lab was a great group to be a part of as everyone was positive and quick to help when possible. There were many other people at Caltech who supported me in various ways during my Ph.D., including Agnes Tong, Felicia Hunt, Sue Chiarchiaro, Joe Drew, Anne Penney, Jeff Groseth, Dr. Angelo Di Bilio, and Dr. Paul Oyala. My work was financially supported by fellowships from the National Science Foundation Graduate Research Fellowship Program, the American Association of University Women, and the Philanthropic Educational Organization.

My family, including my partner, parents, brother, grandparents, cousins, aunts and uncles, Deborah Castleman, and friends, provided tremendous support over the years. There were also several communities that became like family to me, including my speed skating team, my scuba diving buddies, my student affairs team, the Chemistry Graduate Studies Committee, and my housemates.

I would like to thank everyone for their support, time, and energy.

You have my deepest gratitude.

ABSTRACT

Nitrogen fixation, the process of converting dinitrogen to ammonia, is performed industrially and biologically by the Haber-Bosch process and nitrogenase, respectively. The resulting ammonia is largely used as fertilizer. Since there is a finite amount of ammonia produced by nitrogenase, we are heavily dependent on the Haber-Bosch process – only two-fifths of the world's population could be fed without it. Although the importance of the Haber-Bosch process cannot be overstated, our dependence on it has several drawbacks, including significant energy costs (~5% of the annual natural gas consumption), greenhouse gas emissions, and nitrate runoffs. By understanding the biological mechanism of nitrogen fixation, we may be able to (1) develop more efficient nitrogen fixing catalysts to replace those in the Haber-Bosch process or (2) express *de novo* nitrogen fixing proteins in plants so crops can essentially fertilize themselves. The projects described in this thesis aim to contribute to our understanding of the mechanism of biological nitrogen fixation through structural studies of nitrogenase. Nitrogenase consists of the iron and molybdenum-iron (MoFe) proteins, the latter of which contains the active site, the FeMo-cofactor. Throughout my work, I compare the MoFe proteins from A*zotobacter vinelandii* (Av1) and *Clostridium pasteurianum* (Cp1), the two most structurally divergent molybdenum nitrogenases known. Determining the similarities and differences between these proteins may aid our understanding of biological nitrogen fixation. My first project (Chapter III) compares a 1.08 Å Cp1 X-ray structure to a previously published 1.0 Å Av1 structure. I determined that the center atom of the Cp1 FeMo-cofactor is carbon, showing conservation of cofactor structure among molybdenum nitrogenases. Next, I compared substrate pathways in Av1 and Cp1 via Xe pressurization and identification of small molecule binding sites (Chapter IV). My most significant results include the structural and electronic characterization of a reversible protonated resting state of Av1 and Cp1 (Chapter VII).

Published Content and Contributions

Zhang, L.-M., Morrison, C. N., Kaiser, J. T. & Rees, D. C. Nitrogenase MoFe Protein from *Clostridium pasteurianum* at 1.08 Å Resolution: Comparison with the *Azotobacter vinelandii* MoFe Protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71,** 274–282 (2015). doi: 10.1107/S1399004714025243.

C. N. M. participated in the conception of the project, solved and analyzed the crystal structures, prepared some of the data, and participated in the writing of the manuscript.

Morrison, C. N., Hoy, J. A., Zhang, L., Einsle, O. & Rees, D. C. Substrate Pathways in the Nitrogenase MoFe Protein by Experimental Identification of Small Molecule Binding Sites. *Biochemistry* **54,** 2052–2060 (2015). doi: 10.1021/bi501313k.

C. N. M. led the project, collected some of the X-ray data, solved and analyzed the crystal structures, prepared the data, and wrote the manuscript.

Morrison, C. N., Spatzal, T., & Rees, D. C. Reversible Protonated Resting State of the Nitrogenase Active Site. Reproduced with permission from *Journal of the American Chemical Society*. Copyright 2017 American Chemical Society.

C. N. M. led the project, collected the X-ray data, solved and analyzed the crystal structures, prepared EPR samples, collected and analyzed EPR data, prepared the data, and wrote the manuscript.

TABLE OF CONTENTS

KEY NOMENCLATURE

ADP: Adenosine diphosphate, a nucleotide, is the product of ATP hydrolysis. ADP temporarily exists in two binding sites in the Fe protein until it can be replaced with ATP.

ATP: Adenosine triphosphate is a nucleotide. There are two ATP binding sites in the Fe protein. During nitrogen reduction, 16 ATP molecules are hydrolyzed to ADP. The resulting energy drives electron transfer between the Fe and MoFe proteins.

Av1: The MoFe protein from *Azotobacter vinelandii.*

Av2: The Fe protein from *Azotobacter vinelandii.*

Azotobacter vinelandii: An aerobic soil bacterium that expresses nitrogenase; the most commonly studied species of nitrogenase in structural, spectroscopic, and biosynthetic investigations.

***Clostridium pasteurianum*:** An anaerobic soil bacterium that expresses nitrogenase; this species of C*lostridium* is nonpathogenic.

Cp1: The MoFe protein from *Clostridium pasteurianum.*

Cp2: The Fe protein from *Clostridium pasteurianum.*

EPR: In this work, electron paramagnetic resonance (EPR) spectroscopy is used to probe the electronic structure of the FeMo-cofactor.

Fe protein: The iron protein of nitrogenase; a homodimeric protein containing one [4Fe:4S] cluster and two nucleotide binding sites (one in each subunit).

FeMo-cofactor: A [7Fe:Mo:9S:C:*R*-homocitrate] cluster that resides in the alpha subunit of the MoFe protein and is the active site of nitrogenase.

Klebsiella pneumoniae: An aerobic soil bacterium that expresses nitrogenase.

Kp1: The MoFe protein from *Klebsiella pneumoniae.*

Kp2: The Fe protein from *Klebsiella pneumoniae.*

MoFe protein: The molybdenum-iron protein of nitrogenase; a heterodimeric protein in which each dimer contains one [8Fe:7S] P-cluster and one [7Fe:Mo:9S:C:*R*-homocitrate] FeMo-cofactor.

Neutron crystallography: A technique used to determine the atomic structure of crystallized molecules based on the diffraction neutrons from the nuclei of atoms in the molecules.

P-cluster: A [8Fe:7S] cluster that resides at the interface of the alpha and beta subunits in the MoFe protein. This cluster receives electrons from the [4Fe:4S] cluster in the Fe protein and transfers them to the FeMo-cofactor.

X-ray crystallography: A technique used to determine the atomic structure of crystallized molecules based on the diffraction of X-rays off of electrons in the molecule.

Chapter I

Conceptual Framework

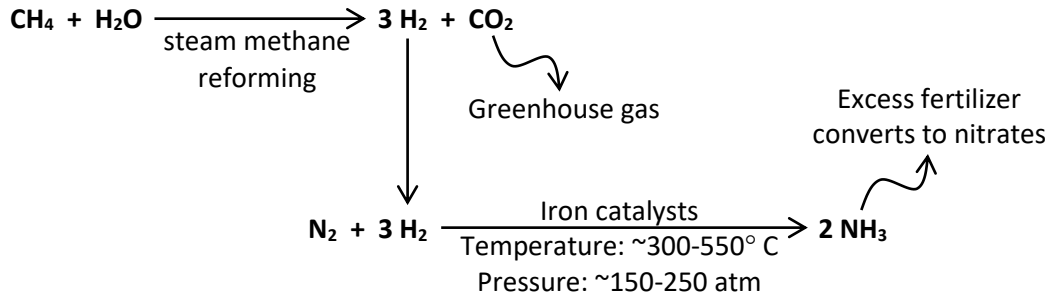
I.A. Motivation

The first mention of nitrogenase in published scientific work is in 1934.[1,2] Over 80 years later, its mysteries are being probed with ever increasing vigor. This enzyme catalyzes the reduction of atmospheric dinitrogen to its bioavailable form, ammonia, in a process called nitrogen fixation. Nitrogenase has intrigued so many scientists over the years because of its role in the nitrogen cycle, its complex structure and function, and its potential impact on sustainable agriculture. Its complexity is illustrated by the fact that after decades of work, the mechanism of biological nitrogen fixation remains enigmatic. The work described in this thesis aims to contribute to our understanding of the mechanism of biological nitrogen fixation through studies of the nitrogenase component proteins that (1) compare and contrast the structure of two different species of nitrogenase to understand what structural features are conserved and may therefore be mechanistically important, (2) identify small molecule binding sites within the proteins as a means of elucidating potential pathways and other functionally relevant aspects of the enzyme, and (3) investigate the atomic and electronic structure of a protonated resting state of the active site to learn about the protonation events on and near the active site during catalytic turnover.

A complete understanding of the biological nitrogen fixation mechanism has the potential to greatly impact sustainable agriculture. Quoting Vaclav Smil, environmental scientist, policy analyst, and author of Enriching the Earth: Fritz Haber, Carl Bosch, and the

Transformation of World Food Production, "Agriculture's principal objective is the production of digestible nitrogen".[3] Digestible nitrogen – ammonia – has been manufactured by the Haber-Bosch process since its invention in the early 1900's. Today, the Haber-Bosch process (Scheme I-1) yields about 275 billion pounds of ammonia per year,[4] the vast majority of which is used in synthetic fertilizers. Less than two-fifths of the current world population could be fed without synthetic nitrogen fertilizers, making the Haber-Bosch process a chief component of agriculture.[3] It is estimated that roughly half of the nitrogen atoms in a human body originate from biological nitrogen fixation and the other half from synthetic fertilizers.[3] While the importance of synthetic fertilizers can be overstated, we face the problem of the unsustainable energetic and environmental costs of this process. For example, the Haber-Bosch process is responsible for ~5% of the world's annual consumption of natural gas, whose steam reforming releases greenhouse gases.[3,4] Also, excess fertilizer that is not incorporated into crops leads to nitrate runoffs resulting in algal blooms that can be toxic.[5,6]

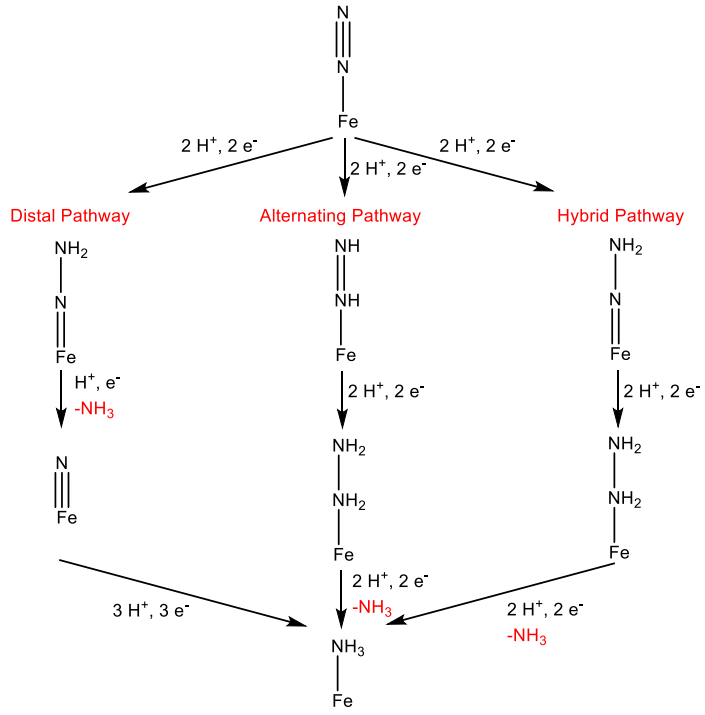
Scheme I-1. Outline of synthetic ammonia production via the Haber-Bosch process



We may be able to lessen our dependence on the Haber-Bosch process by understanding how biological nitrogen fixation works and incorporating this knowledge into new industrial catalysts or genetic engineering of crops to express *de novo* nitrogen-fixing

proteins that allow plants to make their own ammonia. Regarding the former, some synthetic inorganic groups are exploring non-biological nitrogen-fixing catalysts. For example, the formation and isolation of intermediates in small molecule nitrogen fixing catalysts suggests that dinitrogen reduction occurs by a mixed alternating and distal mechanism (Scheme I-2).[7,8] Related studies have explained the importance of a labile bond to the iron, which ecludiates the purpose of the interstitial carbide in the FeMo-cofactor.[9,10]

Scheme I-2. Distal and alternating pathways for dinitrogen reduction as well as the hybrid crossover pathway.



What we currently understand about nitrogenase has already inspired creative approaches to sustainable ammonia production.[11] For example, scientists are using our knowledge of the nitrogenase biosynthetic pathway[12] to engineer the enzyme in cereal crops.[13–15] A hurdle in expressing functional nitrogenase in non-native species is ensuring that all of the

20+ proteins required to yield active nitrogenase are artificially present in the genome and properly expressed in an anaerobic cellular environment.[12,13] As an alternative, imagine an engineered nitrogen-fixing protein with a simpler genome. Such a protein would likely consist of an iron-sulfur cluster similar to the active site of nitrogenase. Work is being done elsewhere on the self-assembly of proteins;[16–18] however, a thorough understanding of the nitrogenase mechanism could inform us on ideal geometry and coordination of metal centers and the nearby protein environment of *de novo* nitrogen-fixing proteins. Engineering such a protein for expression in crops would essentially allow plants to provide themselves with fertilizer; thus, reaching a state of sustainable agriculture.

I.B. Structure of nitrogenase

Nitrogenase (Figure I-1) is an enzyme consisting of two proteins.[19] The simpler of the two is the homodimeric iron (Fe) protein, which has a mass of about 60 kDa and contains a [4Fe:4S] cubane cluster at the interface of the two subunits as well as two nucleotide binding sites (one in each subunit) that bind ATP.[20] The molybdenum-iron (MoFe) protein, about 240 kDa, is a heterodimer containing two copies each of α- and β-subunits. In each αβ-dimer of the MoFe protein, there is one [8Fe:7S] "P-cluster" at the αβ-subunit interface and one [7Fe:Mo:9S:C:*R*-homocitrate] "FeMo-cofactor" buried in the α-subunit, the latter of which is the active site.[21,22] Additionally, there is a mono-nuclear iron site, named Fe16, at the two interfaces of the β- and β'-subunits. The purpose of Fe16 is not known.[23]

Nitrogenase proteins are named based on their bacterial species and the protein component number, where component 1 and 2 refer to the MoFe and Fe proteins, respectively. For example, the MoFe and Fe proteins expressed in *Clostridium pasteurianum* are called Cp1

and Cp2, respectively. Nitrogenases are divided into six groups based on their structure: (1) MoFe proteins homologous to Av1, (2) MoFe proteins homologous to Cp1, (3 and 4) MoFe proteins with small insertions or deletions in the α- and β-subunits, (Anf) FeFe proteins, and (Vnf) FeV proteins.[24] FeFe and FeV refer to component 1 proteins containing Fe or V at the molybdenum site in the FeMo-cofactor. The focus of this work is on MoFe proteins from groups 1 and 2. Most structural, spectroscopic, and biosynthetic investigations of nitrogenase are performed with Av nitrogenase. Cp1 differs by two insertion/deletion chains of about 50 residues each; it has a sequence identity of ~36% compared to Av1. The cofactors are conserved both in structure and location, as is most of the protein secondary structure elements. The similarities and differences between Cp1 and Av1 may shed light on the features important to the mechanism of biological nitrogen fixation and is fundamental to the work presented herein.

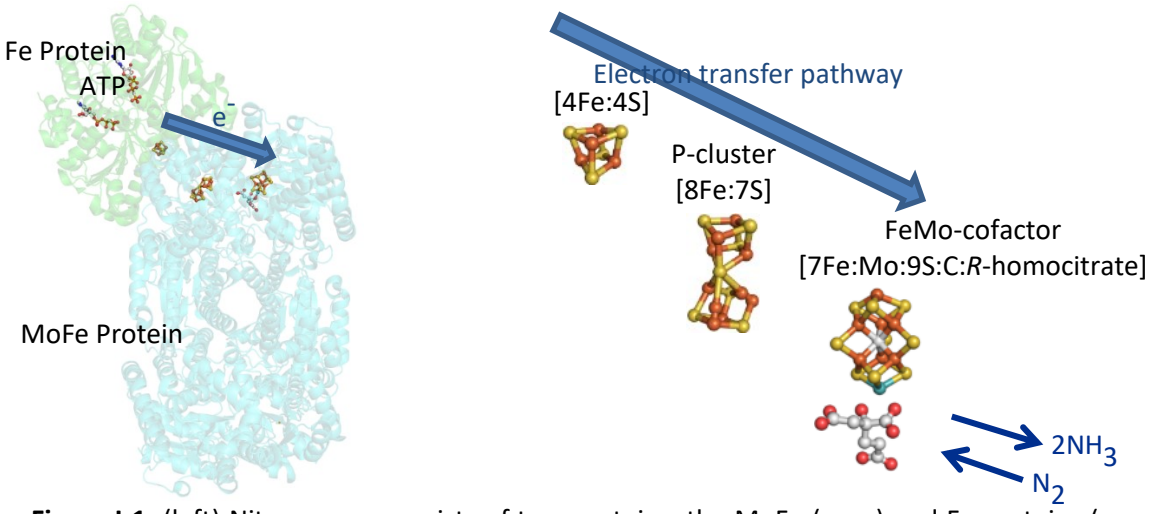


Figure I-1. (left) Nitrogenase consists of two proteins, the MoFe (cyan) and Fe proteins (green). During turnover, ATP hydrolyzes and electrons transfer from the [4Fe:4S] cluster in the Fe protein to the P-cluster and FeMo-cofactor in the MoFe protein (right). Substrates bind and are reduced at the FeMo-cofactor. For simplicity, only one heterodimer of the MoFe protein is illustrated with cofactors and Fe protein binding. The proteins are shown in cartoon representation, and the cofactors are shown in ball-and-stick models and colored by element (yellow: S; orange: Fe; gray: C; cyan: Mo; red: O).

I.C. Mechanism of biological nitrogen fixation

Scheme I-3. Standard model of biological nitrogen fixation

$$N_2 + 8\,H^+ + 8\,e^- + 16\,ATP \rightarrow 2\,NH_3 + H_2 + 16\,(ADP + P_i)$$

The standard model of biological nitrogen fixation is given in Scheme I-3. For substrate reduction to occur, sufficient buildup of electrons and protons (including bridging hydrides) at the FeMo-cofactor is required. This is achieved through the following steps:[25,26]

(1) The Fe protein binds with the MoFe protein by hinging at the [4Fe:4S] cluster and gripping the MoFe protein so that the shortest possible distance between the [4Fe:4S] cluster and the P-cluster is achieved.[27,28] Structural studies on Av1 and Av2 complexes with no nucleotide, MgADP, and MgAMPPCP suggest that the Fe protein may rock across the Av1-Av2 interface during the process of electron transfer.[28]

(2) One electron transfers from the [4Fe:4S] cluster to the P-cluster upon hydrolysis of two ATP molecules located in the Fe protein.[29–32]

(3) The Fe protein and MoFe protein dissociate from each other.

(4) The Fe protein exchanges two ADP for two ATP and is re-reduced by ferredoxin or flavodoxin.[29]

(5) In the MoFe protein, electrons transfer from the P-cluster to the FeMo-cofactor, and substrates are reduced at the FeMo-cofactor.

This process repeats eight times to reduce dinitrogen.[33] Six of these electrons reduce dinitrogen to ammonia, and two are used for hydrogen evolution that occurs upon substrate binding to the FeMo-cofactor.[34,35] Although it is unknown exactly where substrates are bound and reduced, the nitrogenase scientific community generally favors an iron-based reduction mechanism over a

molybdenum-based mechanism; however, both types are observed using small molecules.[9,36] Recent studies with CO and Se suggest that a substrate molecule may bind to Fe2 and/or Fe6 by displacing a belt sulfur.[37,38] Regarding the flow of electrons during substrate turnover, a "deficit-spending" model has been proposed in which electrons first transfer from the P-cluster to the FeMo-cofactor, which gives the P-cluster a sufficiently low reduction potential to attract electrons from the [4Fe:4S] cluster.[39]

The accepted scheme for the kinetics of biological fixation is given by the Lowe-Thorneley model and reparametrized by the Watt group (Figure I-2).[40–45] The E_n states, in which n varies from 0-7, refers to the number of electrons transferred to the MoFe protein; E_0 is the resting state. During the progression from the resting state to more reduced states of the MoFe protein, molecular hydrogen may be formed at E_n (where $n > 1$) and released from the FeMo-cofactor, causing relaxation to E_{n-2}. Dinitrogen binding occurs at E_3 and E_4. Acetylene, which is commonly used to study the substrate turnover activity of nitrogenase, can bind and be reduced to ethylene in the E_1 and E_2 state.

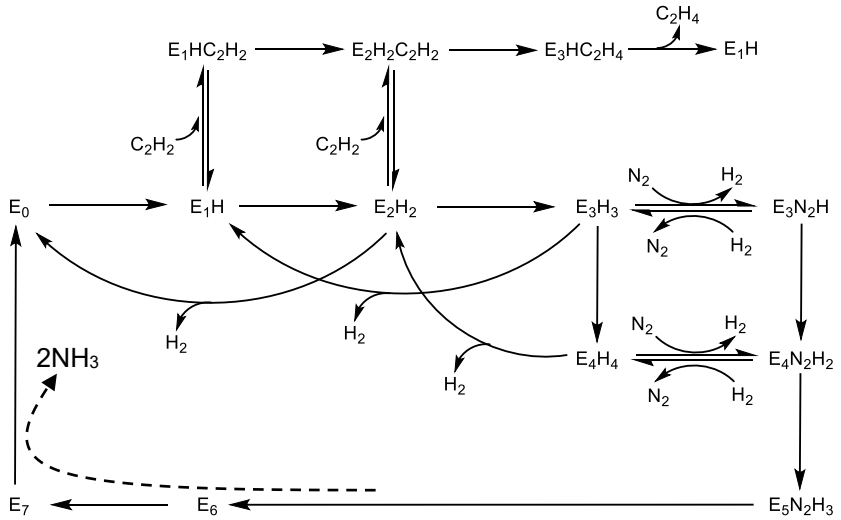


Figure I-2. A summary of the Lowe-Thorneley model.[45] Modeled from Scheme 3 of the following reference.[40]

I hope you enjoy the following chapters, which describe general methodology employed for working with nitrogenase proteins as well as detailed descriptions of several projects that aim to contribute to our understanding of the mechanism of biological nitrogen fixation.

Chapter II

Methodology

II.A. Introduction to and characterization of Cp

Clostridium pasteurianum (Cp) is a Gram-positive bacterium with a low-G+C content from the class *Firmibacteria.*[46] It is a strict anaerobe found widespread in soil and a diazotroph capable of producing endospores.[46] The Rees group originally obtained lyophilized Cp (strain W5, accession number 6013) from the American Type Culture Collection (ATCC) in the early 1990's.[20] To verify the identity of the bacterium, I performed several experiments. Microscopy images of a liquid culture show the general rod-like shape of Cp when it is healthy and at the spore-producing phase (Figure II-1). The gel from polymerase chain reaction (PCR) using bacterial and universal primers for the RNA of the small ribosomal subunit (16S rRNA) gene (Figure II-2) shows one band at ~1500 nucleotides with the bacterial primer and one band at ~1000 nucleotides with the universal primer. Lack of additional bands indicates that the Cp culture is free from non-bacterial contamination. Also, RNA sequencing of the small ribosomal subunit gave an identity of 99% (589/596), confirming that the culture is Cp.

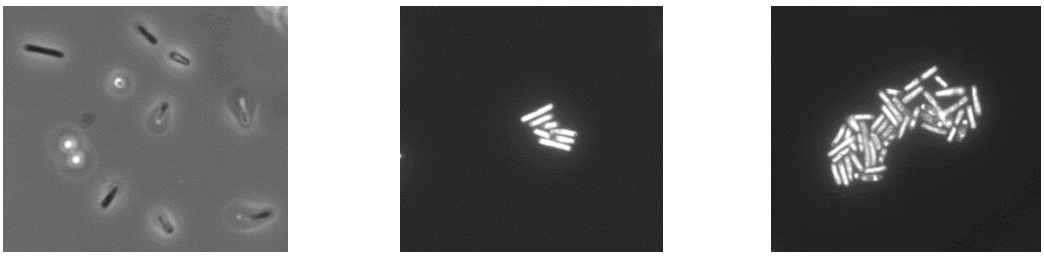


Figure II-1. All images are at 1000x magnification (left) Live view of Cp culture. Short rods and spherical objects are Cp cells orientated at an angle. (middle) DAPI stained Cp culture fixed on a glass slide. (right) DAPI stained Cp culture at the spore-producing phase.

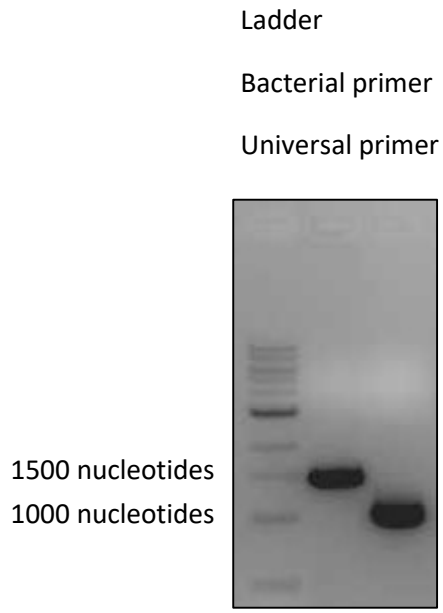


Figure II-2. Gel of PCR product using bacterial and universal primers for the RNA of the small ribosomal subunit (16S) gene. The gels show that there is no non-bacterial contamination.

II.B. Cell growths

II.B.i. Growth in mineral and potato media

Cp cultures were permanently stored in 1 mL aliquots (0.5 mL culture and 0.5 mL 50% sterilized glycerol) in liquid nitrogen. The frozen cell stock was replenished from 100 mL mineral medium cultures in the growth phase (<24 hours growth). Live cultures were maintained on mineral medium agar plates with no ammonium source at 4° C. New plates were streaked once per month. Every six months, new plates were streaked from frozen cell pellets; two consecutive generations of plate growths were required to obtain healthy cells. Agar plates were prepared with mineral medium according Table II-1. Contrary to liquid medium cultures, no ammonium chloride or calcium carbonate was added. 1.5% agar was added to the medium for plates prior to autoclaving. Calcium chloride was not added until after autoclaving, and a 100x calcium chloride stock was used rather than 1000x as for liquid medium cultures. This prevents calcium

phosphate precipitation, which appears as small white flakes in the cooled gel. Media for plates was generally made at volumes of 400 mL (20 plates).

To prepare cells for harvesting from a 15-L growth in a carboy, five new plates were streaked near an open flame with a sterilized loop. Plates were incubated upside-down at 30° C in a sealed jar kept anaerobic with an Oxoid AnaeroGen sachet for 36-48 hours. If plates were streaked from freshly grown plates at 30° C, only 36 hours was required for full growth. If plates were streaked from plates stored at 4° C, 48 hours was required for growth.

While the plates incubated, all media, long needles (~18 Gauge; used for purging the media of oxygen), rubber stoppers, and filters (made by inserting half a cotton ball into a Luer-Lok syringe) were prepared and autoclaved. Liquid mineral media were prepared in Erlenmeyer flasks with at least three times the headspace volume as liquid volume to allow room for gas production during cell growth. The openings of the media flasks were covered with aluminum foil and sealed with parafilm upon cooling to room temperature after autoclaving. Media stored at room temperature was not used if made more than three weeks in advance.

Once the plates were ready, two flasks of 100 mL mineral media were purged with filtered nitrogen gas. The flask opening remained completely covered with aluminum foil to avoid exposure to contaminants and air. The required amounts of phosphate, biotin, and ammonium chloride were added at the start of the purging. The phosphate was *slowly* added and while shaking the flask. After about 30 minutes, cells were inoculated into media from plates near an open flame. Upon opening the anaerobic jar, one plate was immediately sealed with parafilm and stored at 4° C as the new master plate. To the other four plates, working two plates at a time, 4 mL of the purged media was transferred to the plate. Cells were suspended by gently rubbing a cooled, flame-sterilized cell spreader around the plate. The suspended cells were inoculated into the purged media. The cultures continued to purge for another 20

minutes, after which the purging needle was slowly removed as a sterilized rubber stopper was simultaneously depressed into the flask on top of the foil. The stopper was pushed down as far as possible on top of the foil, and then the flask was completely sealed with parafilm and tape. Cultures were incubated at 30° C and 200 rpm for 18 - 24 hours. The optical density (OD) of the cultures was monitored visually or with a UV-vis spectrophotometer at 600 nm. Cultures grew to an OD_{600} of 6-10. Healthy cultures had a small layer of foam (0.5-1 cm). Generally, two liquid cultures were grown in duplicate. The pH range for optimum growth is 5.8 to 6.5.[47] The calcium carbonate and phosphate in the liquid mineral medium act as pH buffers.[47]

Cultures were scaled up to 250 mL by inoculating from the 100 mL starter cultures at a dilution of 1:10. 250 mL cultures were grown for ~12 hours. The culture was then scaled up to 1.5 L with a starting OD_{600} of 0.5-0.8. For the 1.5-L cultures, the media was purged using a tubed septum rather than needles, as illustrated in Figure II-3. 1.5 L cultures were grown for 8-9 hours.

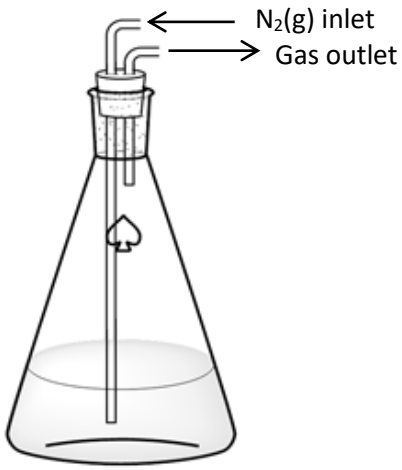


Figure II-3. 1.5 L cultures were purged through glass tubes fitted into the rubber stopper.

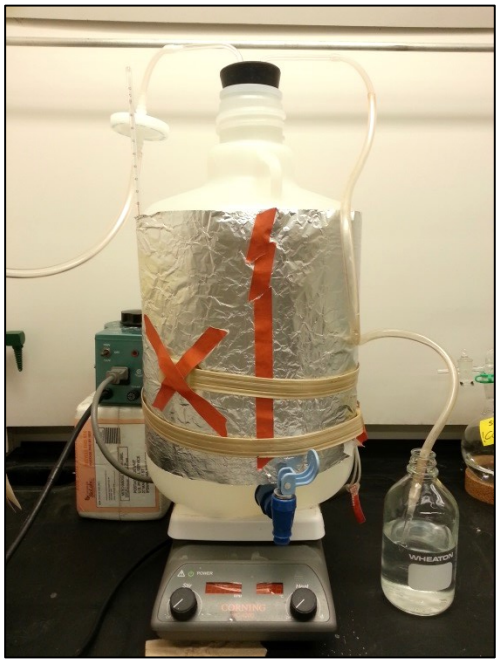


Figure II-4. The carboy setup for 15-L cell growths. Heating tape was wrapped around the carboy. Another layer of aluminum foil was wrapped around the outside of the carboy (not shown) to avoid heat loss to the atmosphere. A stir bar at 300 rpm was used to keep the culture agitated. A thermometer between the first aluminum foil layer and the carboy was used to monitor the approximate temperature of the culture. Continuous nitrogen was given to the carboy through rubber tubing. The nitrogen flow rate was monitored from bubbling the out-gas into a flask of water.

While the 1.5 L culture was growing, the 20 L carboy was prepared with 15 L of mineral medium. The setup is shown and described in Figure II-4. The carboy was purged with nitrogen gas for at least six hours prior to inoculation. During this time, the temperature stabilized at 30° C. One hour before inoculation, the appropriate amount of phosphate, biotin, and ammonium chloride was added. Only a 2.5 mM ammonium supplement was provided to activate nitrogenase expression. (The pre-cultures may also be grown on limited or no ammonium supplement; however, it was easier to keep cells alive with an ammonium supplement.) After adding all solutions to the carboy, the nitrogen flow rate was increased to a vigorous bubble: higher flow rates increased nitrogenase activity.

The OD_{600} of the carboy was measured at inoculation and every hour starting at 5 hours. Nitrogenase activity (section II.D.) was measured every hour starting at 5 hours. Just after the activity/OD_{600} peaked, cells were harvested by collecting the culture in 1-L centrifuge bottles on ice. (Stirring was turned off during collection to avoid collecting calcium carbonate.) The culture was centrifuged in several batches at 7000 rpm for 20 minutes and 4° C. The supernatant was discarded and the cell pellets were immediately frozen in liquid nitrogen. From a 15 L culture, it was common to obtain >50 g cell pellet, which was stored in liquid nitrogen or at -80° C. The cell pellets are spongy as a result of gas escaping from the cells and are light brown in color.

For all cell growths, the starting OD_{600} of each culture was controlled to avoid overgrowth. Overgrown cultures sometimes experienced an acid crash as a result of too much butyric and other acids. This causes the Cp metabolism to shift toward the production of alcohols and is characterized by a sweeter smell than the typical rancid smell of butyric acid.[49–52]

Table II-1. Recipe used to make mineral media

Components of mineral medium	Quantity needed for various culture sizes			
	100 mL	250 mL	1.5 L	15 L
Sucrose (g)	2	5	30	300
Water (mL)	90	225	1350	13.5 L
Mineral stock (1000x, mL)	0.1	0.25	1.5	15
FeMoCa (500x, mL)	0.2	0.5	3	30
$CaCl_2$ stock (1000x, mL)	0.1	0.25	1.5	15
$MgSO_4$ stock (1000x, mL)	0.1	0.25	1.5	15
$CaCO_3$ (g)	0.5	1.25	7.5	~17 **
Phosphate stock (10x, mL) *	10	25	150	1500
NH_4Cl (3M, mL) for 10 mM *	0.333	0.8325	4.995	12.5 (2.5 mM)
Biotin (100x, mL) *	1	2.5	15	150

* Chemicals added before inoculation when the media is purged with nitrogen.
** Limited $CaCO_3$ is added to the carboy media to prevent the stir bar from sticking in the solid.

Table II-2. Recipes for mineral media stock solutions

Chemical	Mol. Wt. (g/mol)	Conc. (mM)	Mass per volume (g)			Notes
			500 mL	1 L	1.5 L	
Mineral stock (1000x)						Do not autoclave
$CoCl_2 \cdot 6H_2O$	237.93	1		0.24		
$CuSO_4$	159.61	1		0.16		
$ZnCl_2$	136.28	1		0.14		
$MnSO_4 \cdot H_2O$	169.02	10		1.69		
$NiSO_4 \cdot 6H_2O$	262.86	1		0.26		
Phosphate stock pH6.4 (10x)						Autoclave separately
KH_2PO_4	136.09	800	54.44	108.87	163.31	
Na_2HPO_4	141.96	200	14.20	28.39	42.59	
3 M NH_4Cl stock						Autoclave separately; make batches 500 mL or less
NH_4Cl	53.49	3000	80.24			
0.9M $CaCl_2 \cdot 2H_2O$ (1000x)						Only make in 500-mL batches
$CaCl_2 \cdot 2H_2O$	131.02	900	66.2			
1.67M $MgSO_4 \cdot 7H_2O$ (1000x)						Only make in 500-mL batches
$MgSO_4 \cdot 7H_2O$	246.47	1670	205.8			
FeMoCa (500x)						Dissolve citric acid in water first, then add Fe and Mo
Citric Acid$\cdot H_2O$	210.14	90		18.91		
$FeSO_4 \cdot 7H_2O$	278.01	18		5.00		
$Na_2MoO_4 \cdot 2H_2O$	241.95	1		0.24		
Biotin (100x)						Dissolve in 100% ethanol. Store at 4° C.
Biotin	244.31	1		0.24431		
Citric Acid						Do not autoclave
Citric acid monohydrate	210.14	1 M	105	210	315	

Potato medium was sometimes employed to jumpstart Cp growth from agar plates to liquid media. Although Cp readily grow in potato medium, the subsequent growths in mineral medium were more difficult to control, likely due to potato contaminations. To prepare potato medium (Figure II-5), 3-4 mm layer of calcium carbonate was added to the bottom of each of ten test tubes (16 x 150 mm). A fresh (Russet) potato from the supermarket was sliced into small cubes about 0.5 x 0.5 x 0.5 cm^3. Each test tube was filled one-third (~5 cm) with potato cubes and then two-thirds (~10 cm) with 2% sucrose solution. Tubes were capped with a metal cap and autoclaved.

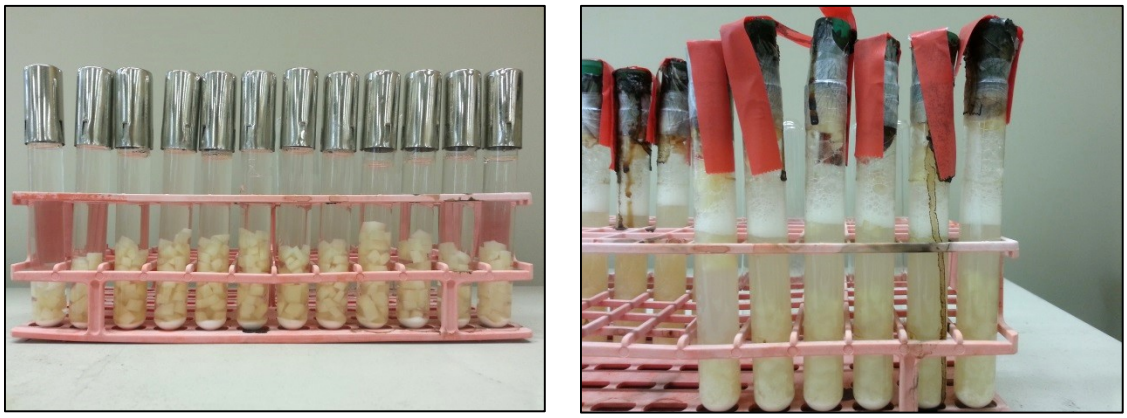


Figure II-5. (left) Potato tubes before autoclaving, showing the relative amounts of $CaCO_3$, potato cubes, and sucrose solution. (right) Potato tubes with fully grown Cp ready to be inoculated into mineral media.

Immediately after autoclaving, the metal caps were removed, and the top of each tube was plugged with half a sterile cotton ball. ~5 drops each of concentrated potassium carbonate and pyrogallol were added to each cotton ball to remove oxygen from the tube. Tubes were promptly and tightly capped with a rubber septum and then allowed to stand at room temperature to cool. While cooling, ~5 mL mineral medium was added to plates to suspend cells. To inoculate the potato tubes, the cotton plug was removed and 1 mL suspended cells was added to each tube. A new cotton plug was placed in each tube with the same treatment of alkaline pyrogallol. The rubber stoppers were placed back in each tube, and each tube was sealed with parafilm and tape. (The tape is necessary so the rubber stoppers do not eject during cell growth.) Cells were ready to inoculate into mineral media once vigorous bubbling was observed and the potato culture was opaque (Figure II-5). During inoculation into mineral media, care was taken to not transfer potato chunks.

II.B.ii. Growth curves

Cp growth patterns depend on the amount of ammonium supplied to the mineral media.[53] To determine the optimum level of ammonium, carboy growths were performed at different initial concentrations of ammonium chloride. The cultures were monitored until activity ceased. The results are shown in Charts II-1, 2, and 3. A description of activity calculation is provided in section II.D.ii. Although lack of ammonium supplement produces the highest activity, the peak activity occurs at a low OD (Chart II-1), which is unfavorable for harvesting large quantities of cells. The opposite is observed with 10 mM ammonium supplement: the peak activity is low and occurs at a high OD (Figure II-3). The best condition is with a limited ammonium supplement in which the peak activity per OD occurs at a reasonably high activity and OD (Chart II-2).[53]

All 15 L harvest cultures were provided with 2.5 mM ammonium supplement (Chart II-4). Growth increases at a steady rate while the initial ammonium supplement is consumed (~0-4 hours), after which the growth reaches a lag phase where the cells begin to express nitrogenase (~4-6 hours). Once a sufficient amount of active nitrogenase is available, cell growth resumes using of the ammonium produced by nitrogenase (~6-9 hours). Cells are harvested after the activity/OD peaks, ~9 hours.

Chart II-1. 15 L culture with no ammonium supplement

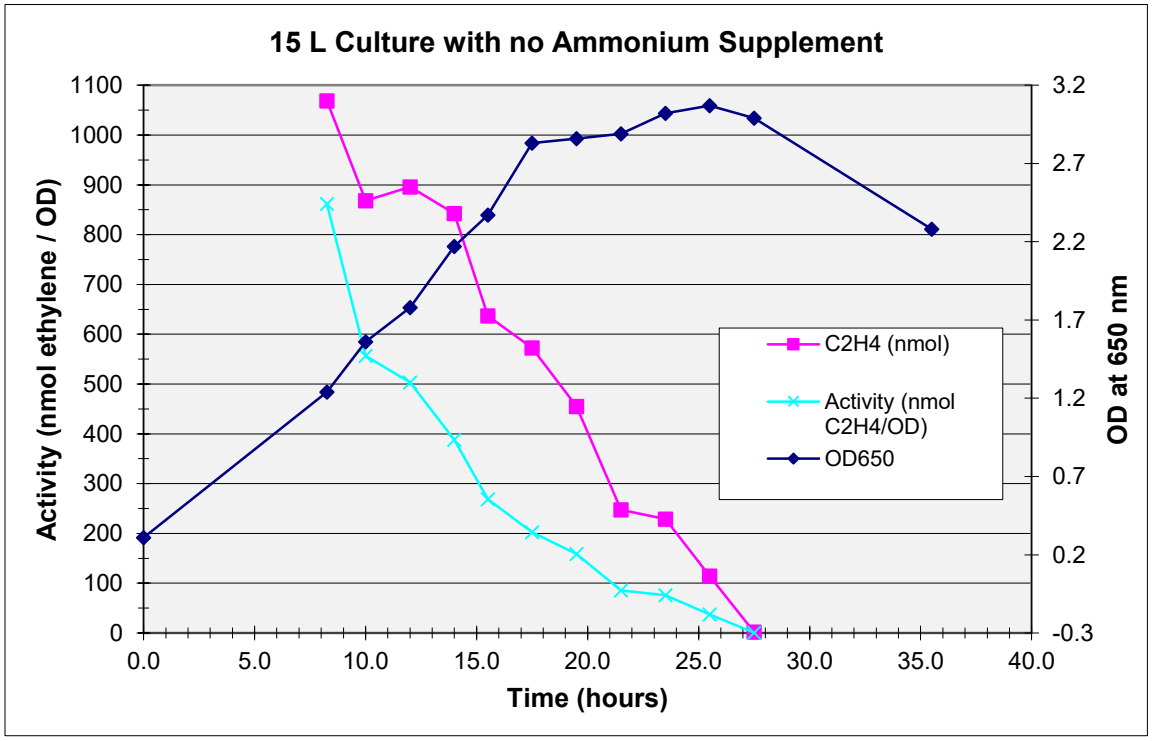


Chart II-2. 15 L culture with 2.5 mM ammonium supplement

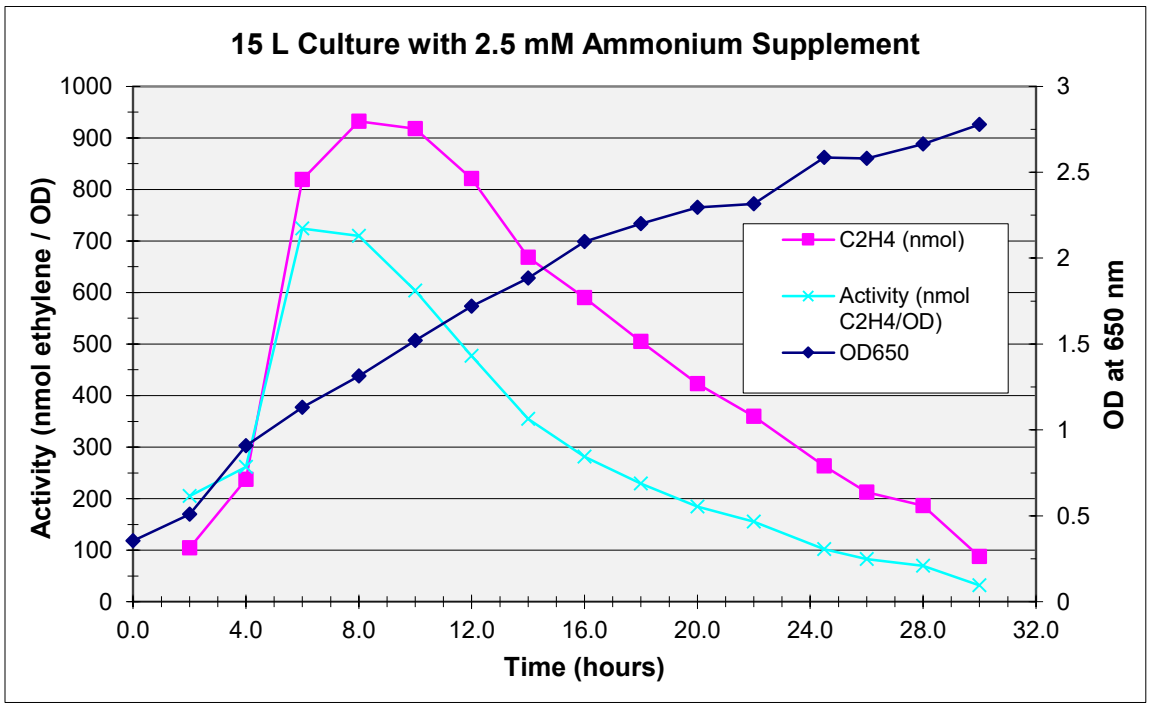


Chart II-3. 15 L culture with 10 mM ammonium supplement

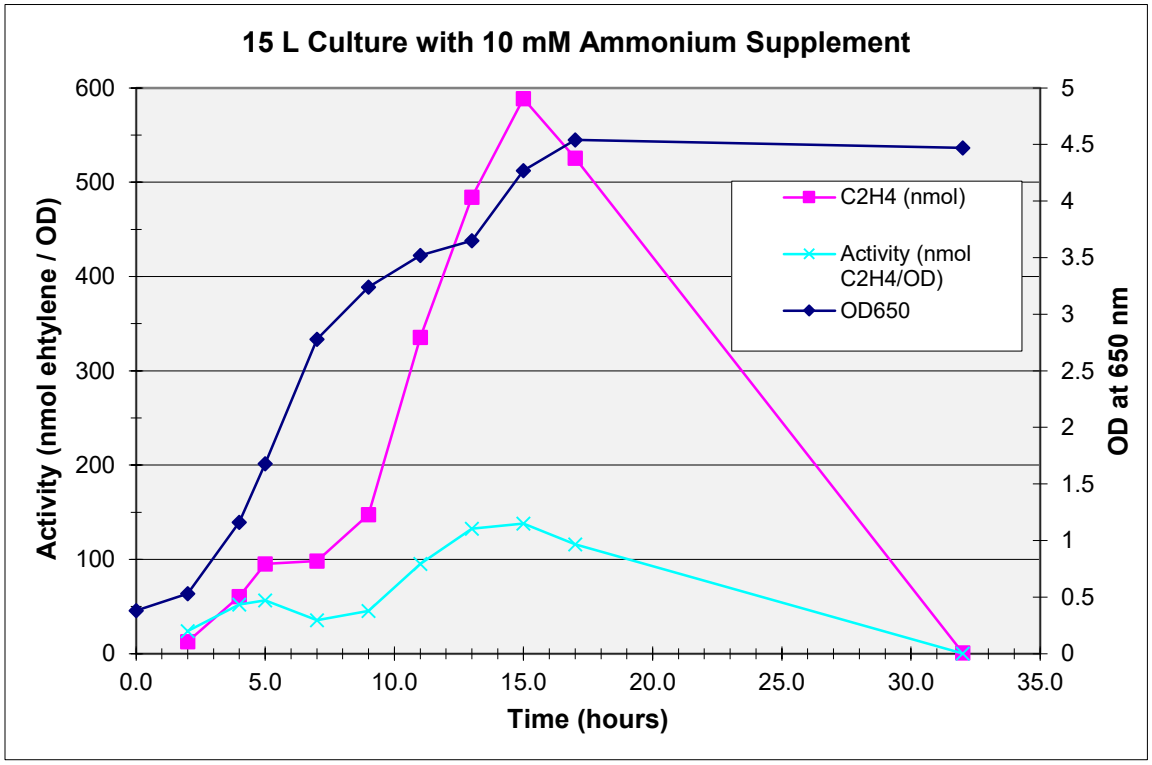
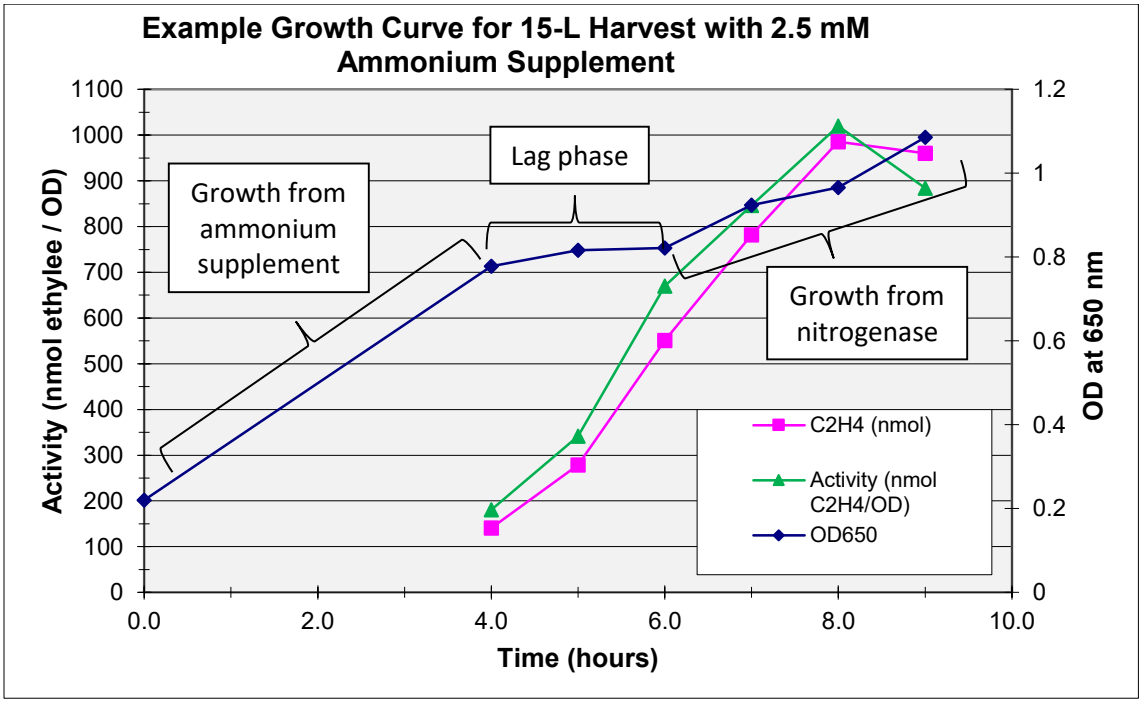


Chart II-4. Example growth curve for harvest cultures at 2.5 mM ammonium supplement



II.C. Protein purification

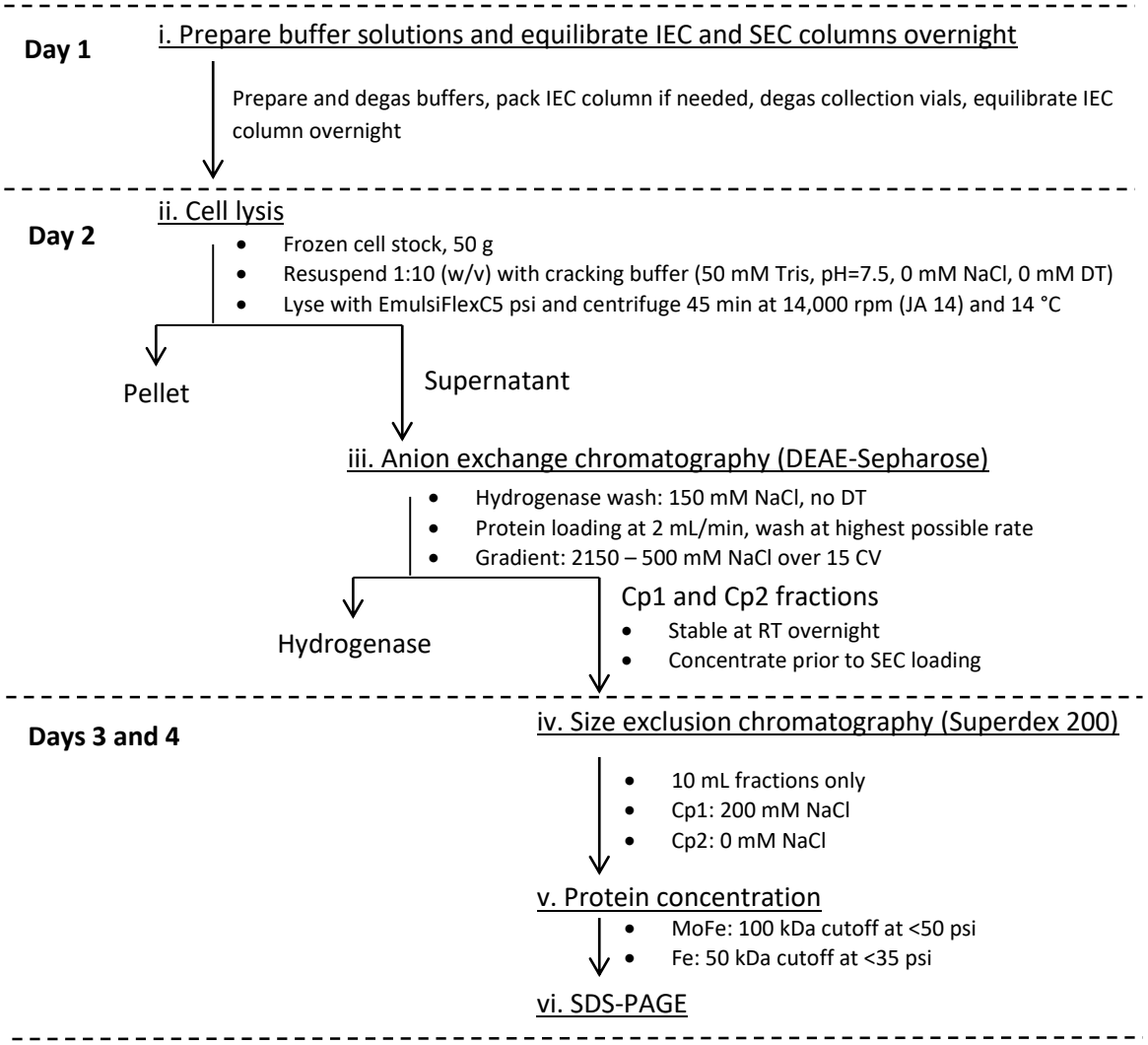


Figure II-6. Flow diagram of the Cp1 and Cp2 purification process.

II.C.i. Preparation of IEC and SEC columns and buffer solutions

II.C.i.a. IEC column

Diethylaminoethyl (DEAE) sepharose, a positively charged resin, was used for ion exchange column chromatography (IEC). If the column had no cracks and protein did not precipitate during the previous purification, the resin was recycled without unpacking the

column. In this case, 2-3 column volumes each of the following solutions (in order) were run through the column: 1 M NaCl, 100 mM NaCl, 1 M NaCl, 100 mM NaCl, water, 0.5 M NaOH, water, 0.5 M HCl. The column was stored in 20% ethanol if not in use for longer than a few days.

Unpacked resin was stored in a plastic bottle at 4° C in 30% ethanol with 50 mM Tris/HCl pH 8.0. To recycle the resin for use, the resin was soaked in 6 M urea for at least two hours. Next, the following washing procedure was executed: wash resin with 8 L water; suspend resin in 1 L of 0.5 M sodium hydroxide in a plastic beaker for 30 minutes; wash with 10 L water; suspend in 1 L of 0.5 M HCl for 30 minutes; wash with 12 L water. All wash cycles were performed using vacuum filtration and a large Büchner funnel with filter paper. The resin was never allowed to dry out while filtering. During the suspension stages, the resin was stirred occasionally with a glass rod. Prior to packing the column, the recycled resin was degassed on the vacuum manifold with eight cycles of 5-minute vacuum and 30-seconds argon. Instead of using a stir bar, which could sheer the resin particles, the flask was manually shaken. Generally, a column diameter of 2.5 cm was selected and packed no taller than 20 cm (~100 mL of packed resin). After packing, the IEC column was washed with water to remove any ethanol, which could cause NaCl in the buffer solutions to precipitate. After several column volumes of water were washed through the column, it was equilibrated overnight with a buffer of the same salt concentration as the cracking buffer (buffer A).

II.C.i.b. SEC column

A typical size exclusion column (SEC) included ~475 mL of Superdex S200 resin. After washing with water, the column was equilibrated overnight with the E_F (elution) buffer. After use, the column was flushed with more elution buffer, washed with several column volumes of water, and then stored under 20% ethanol. Flow rates never exceeded 2 mL/min.

II.C.i.c. Buffer solutions

Buffer solutions were prepared according to Table II-3. First, NaCl and Tris base were dissolved in water. After the solids completely dissolved, the pH was lowered with 6 M HCl. The buffers were diluted to the appropriate volume, after which all buffers (excluding the cracking buffer) were vacuum filtered. The cracking buffer and small volume of buffer A were transferred to 2 L and 1 L round bottom flasks, respectively, with a stir bar and capped with a septum. The other buffers were transferred to large round bottom flasks (3-5 L) and capped with an adapter that supports the inward flow of argon and the outward, anaerobic flow of buffer to the column. The buffers were degassed on the vacuum manifold with 10 cycles of 10-minutes vacuum and 2-minutes argon. Flasks were occasionally manually shaken during the vacuum cycles. Wheaton vials used to collect fractions eluted from the columns were also degassed on the Schlenk line.

After degassing was complete, the cracking buffer and small volume of buffer A were transferred to the anaerobic tent. The cracking buffer was divided into two portions: 500 mL and 250 mL. To the three 5-L flasks still on the Schlenk line, sodium dithionite ($Na_2S_2O_4$) solutions were added according to Table II-4, after which three additional cycles of vacuum and argon were performed.

The IEC column was equilibrated overnight using a combination of buffers A and B equal to the cracking buffer salt concentration. Also, the following materials were brought into the anaerobic tent in preparation for use: protein concentrating systems, four 250-mL centrifuge bottles, homogenizer, 100% glycerol (degassed), three 1000-mL round bottom flasks (for cell suspension, collecting lysed cells, and IEC loading protein), 1000 mL beaker and stir bar, 500 mL round bottom flask, assorted sizes of auto-pipettes, and a 500 mL graduated cylinder.

Table II-3. Buffer solutions for protein purification experiments

Buffer	Purpose	NaCl		Tris base		pH	Amount (L)	Filtered?
		mM	g	mM	g			
Cracking	Lyse cells	50	3.0	50	6.06	7.75	1	no
Buffer A	IEC	0	0	50	18.2	7.75	3	yes
Buffer A	Dilute Fe protein after IEC	0	0	50	3.03	7.75	0.5	yes
Buffer E_F	SEC, Fe protein	50	8.8	50	18.2	7.75	3	yes
Buffer E_M	SEC, MoFe protein	200	35.1	50	18.2	7.75	3	yes
Buffer B	IEC	500	58.5	50	12.1	7.75	2	yes

Table II-4. Sodium dithionite solutions added to buffers

Buffer	Amount	$Na_2S_2O_4$ (mM)	0.5 M Tris base (mL)	$Na_2S_2O_4$ (g)	Quantity
Cracking buffer & buffer A (for Fe protein dilution)	500 mL	5	2.5	0.435	2
Cracking buffer	250 mL	50	12.5	2.176	1
Buffers A, E_F, E_M	3 L	5	15	2.612	3
Buffer B	2 L	5	10	1.741	1

II.C.ii. Cell lysis

Cells were lysed under pressure with an EmulsiFlex C5 (Figure II-7). Frozen cell stock was thawed in cracking buffer (typically 50 g cells in 500 mL cracking buffer) with stirring until the suspension was homogenous (~20 minutes). 200 µL glycerol was added as an anti-foam reagent. In 50 mL portions, the cell suspension was homogenized and collected in a 1000-mL round bottom flask. The cell suspension, an empty 1000 mL flask, and 250 mL cracking buffer with 50 mM dithionite (flush solution for EmulsiFlex) were removed from the tent, put under positive argon pressure, and placed in an ice bath (to dissipate heat acquired from the EmulsiFlex during lysis). The EmulsiFlex was washed with water and the flush solution, after which the cell suspension was ruptured at ~2500 psi. The lysate was collected in the empty 1000 mL flask. After use, the EmulsiFlex was washed with flush solution and water and stored under 20%

ethanol. The cell suspension was centrifuged at 30,000 × g for 45 minutes at 15° C. A successful lysis yielded a dark brown supernatant and a pellet with three layers (lipids, cytosol, calcium carbonate). The supernatant was transferred to a round bottom flask for loading onto the IEC.

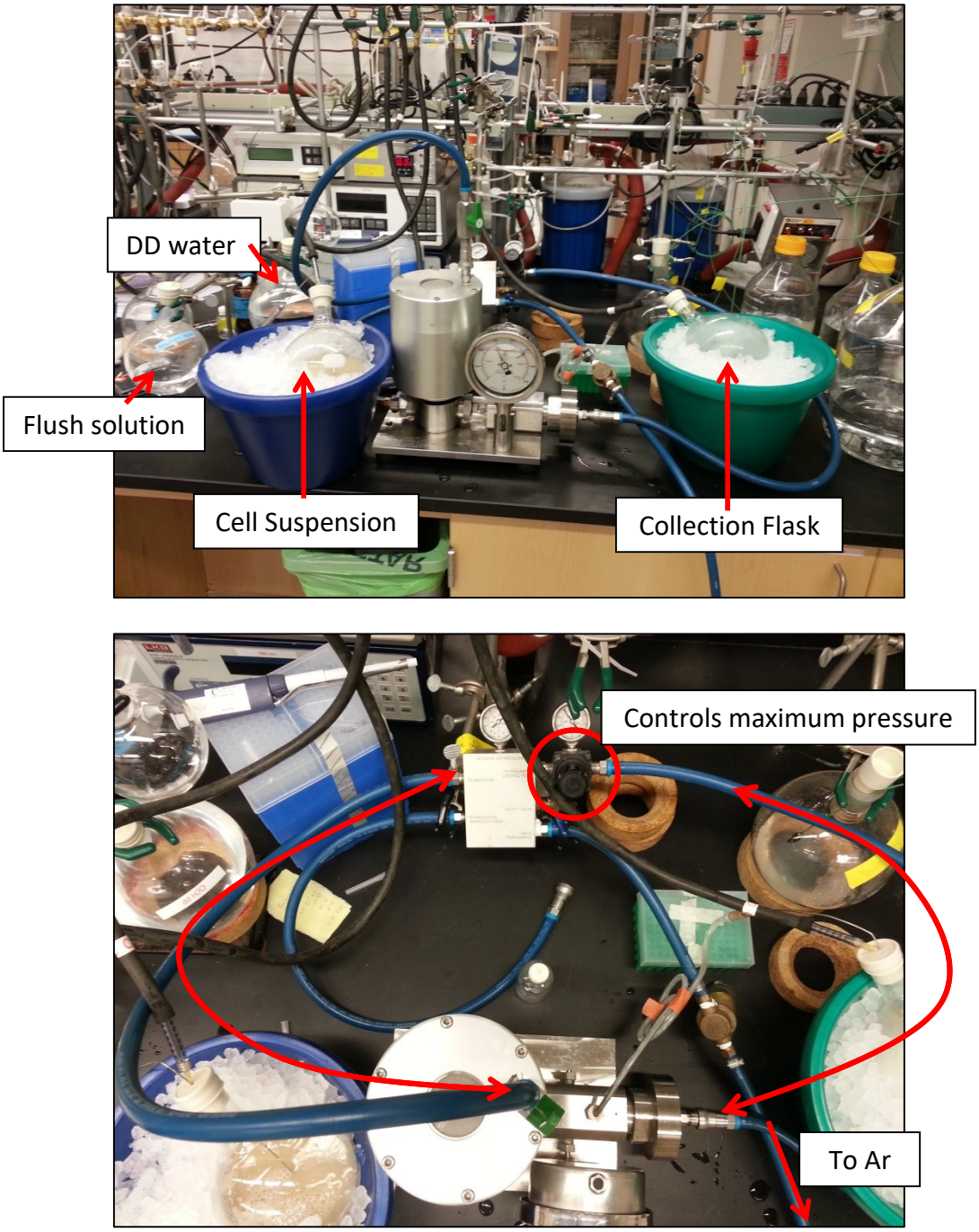
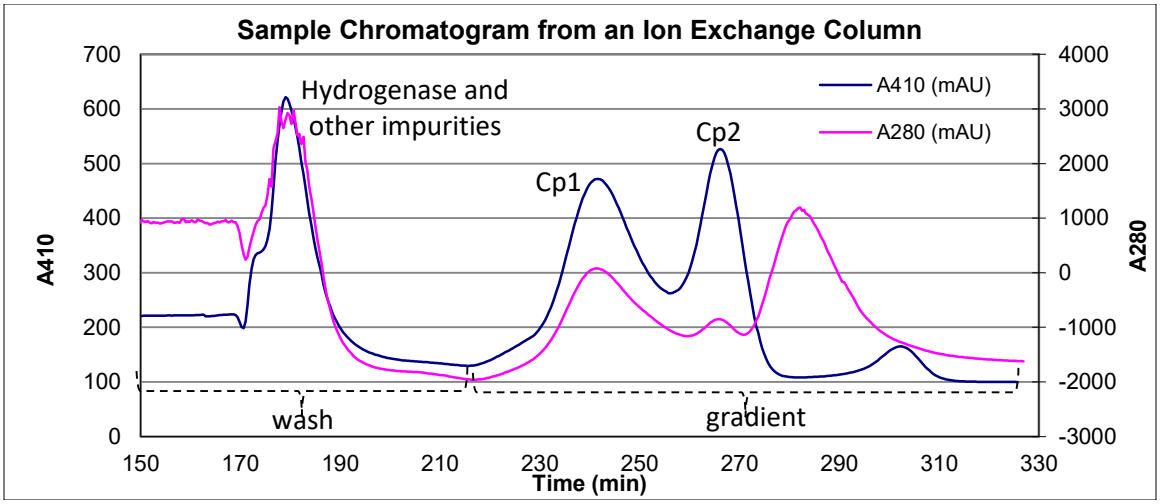


Figure II-7. Setup of cell lysis via the EmulsiFlex C5.

II.C.iii. Anion exchange chromatography

The flask containing protein was removed from the anaerobic tent and immediately put under positive argon pressure and placed on ice to avoid protein precipitation. Without introducing any air bubbles into the system, the protein was loaded onto the IEC at 2 mL/min through a long needle using an ÄKTA system (Pharmacia Biotech ÄKTA Explorer Air100). Since there is typically ~500 mL of protein solution, the loading process takes several hours. When experimenting with new buffers, the protein was loaded using an FPLC pump (which is less easily clogged compared to the ÄKTA) in case protein or lipids precipitated during loading.

Chart II-5. Sample chromatogram from the IEC



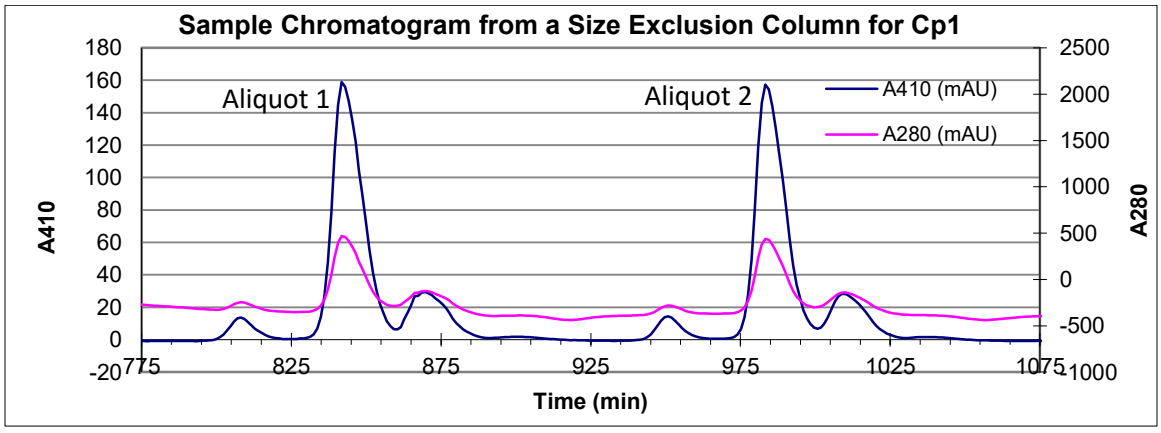
After protein loading was complete, elution was monitored at 280, 410, and 550 nm. For the column wash and elution, the flowrate was as fast as possible without exceeding a backpressure of 0.5 MPa (usually, 3.0-4.0 mL/min). The column was washed with 5-10 column volumes of 150 mM NaCl (mixture of buffers A and B) to elute hydrogenase and other unwanted material from the crude extract. After the column looked clean and no more peaks were visible in the chromatogram, a gradient was started using buffers A and B. A typical gradient was 150-

500 mM NaCl over 15 column volumes with flow rate of ~3-4 mL/min. Cp1 elutes first, typically ~250 mM NaCl; Cp2 follows closely behind at ~300 mM NaCl; ferredoxin elutes ~450 mM NaCl. A sample chromatogram is shown in Chart II-5. All bands were some shade of brown or black. Protein fractions were collected in anaerobic Wheaton vials. Cp1 and Cp2 were left overnight at room temperature if proceeding to SEC the following day; otherwise, the proteins were frozen and stored in liquid nitrogen. After running several column volumes of 500 mM NaCl over the IEC, it was washed according to the procedure previously described.

II.C.iv. Size exclusion chromatography

The SEC was washed with several column volumes of water and equilibrated overnight with E_F buffer. In the morning, the Cp1 and Cp2 solutions were concentrated in an Amicon Stirred Cell under argon pressure with a 100 and 50 kDa membrane, respectively. Concentrated protein was loaded on the SEC in 10-mL fractions at 1-1.5 mL/min and eluted at 2 mL/min using the buffers stated in Table II-3. The backpressure was not allowed to exceed 0.25 MPa. After collecting purified protein, the SEC column was flushed with additional elution buffer, washed with water, and stored in 20% ethanol. A typical SEC chromatogram is shown in Chart II-6.

Chart II-6. Sample SEC chromatogram



II.C.v. Protein concentration

Cp1 and Cp2 fractions were concentrated in an Amicon Stirred Cell under argon pressure using 100 kDa and 30 kDa cutoff filters, respectively. The pressure during Cp1 and Cp2 concentration was <50 psi and <35 psi, respectively. A Cary 300 Scan UV-visible spectrophotometer was used to determine the concentration of the purified protein. The instrument was first calibrated to the SEC elution buffer with 100% and 0% transmission. To 980 µL of the buffer, 20 µL of concentrated, purified protein was added. The cuvette was capped with a septum so maintain an anaerobic environment. The entire absorbance spectrum was collected. To determine the specific concentration, Beer's Law was employed: (sample dilution) * [(absorbance at 410 nm) - (absorbance at 280 nm)] * (extinction coefficient) * (path length, 1 cm) = protein concentration. The extinction coefficients for the Fe and MoFe proteins are 6.702 (at 430 nm) and 3.066 $mM^{-1}cm^{-1}$, respectively.[54] The target protein concentration was 25-35 mg/mL. The purified protein was frozen and stored in liquid nitrogen in 0.1 mL aliquots.

II.C.vi. SDS-PAGE

To confirm the purity and identity of the purified protein, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed on small samples from all stages of the protein purification. Samples were prepared by adding 20-30 µL of loading dye to 50 µL of each sample. Highly concentrated samples, such as the purified protein, were diluted with water before adding loading dye. 1% SDS was added to any sample containing insoluble material, such as the cell pellet. 10 µL of each sample was loaded into 12% Tris/HCl gels as well as a protein ladder. Gels were run at 200 V for ~45 minutes, after which the gels stained for one hour and then de-stained for at least one hour. A typical gel is shown in Figure II-8.

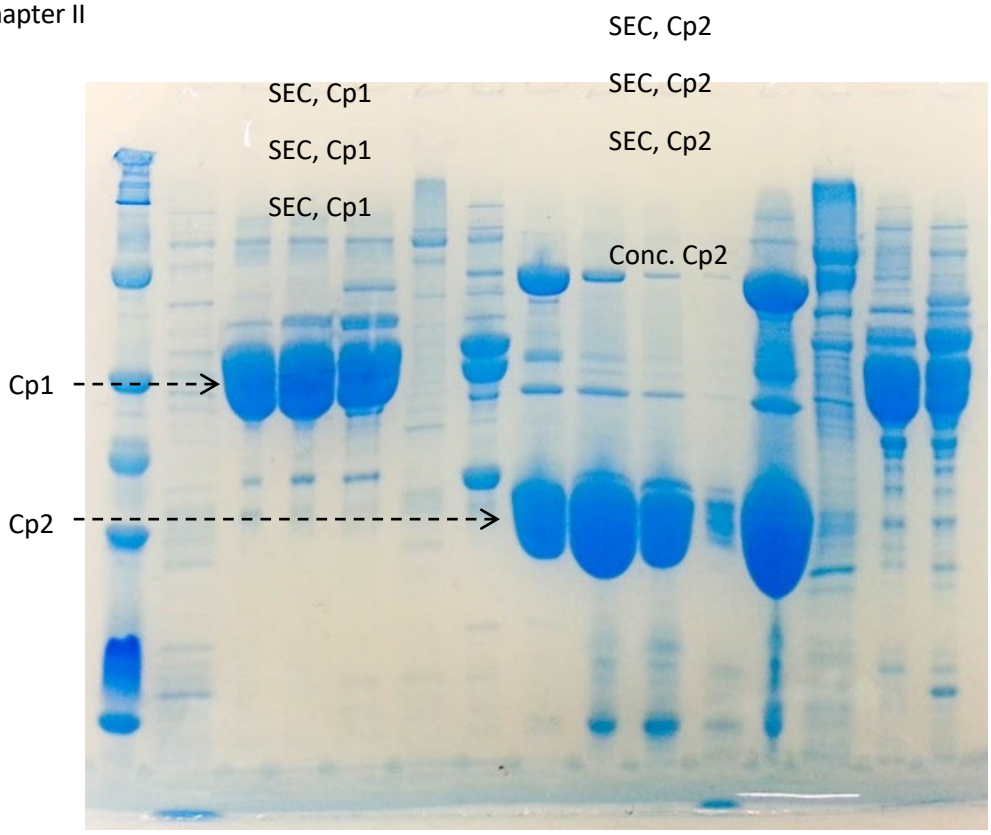


Figure II-8. Sample gel for protein purification.

II.D. Acetylene reduction activity assays

Although there are many substrates of nitrogenase,[55] acetylene reduction to ethylene is typically used to characterize the activity of nitrogenase because acetylene and ethylene are easily detected with high sensitivity using gas chromatography (GC).

II.D.i. Whole cell acetylene reduction assays

Whole cell assays were used to monitor nitrogenase activity during cell growth. Prior to starting the activity assays, a 500-mL sealed round bottom flask was filled with acetylene by evacuation of water with acetylene pressure. 2.5 mL of 5 mM sodium dithionite in 500 mM Tris base was injected into the flask. The dithionite solution was swirled around the flask, and the flask stood at room temperature for approximately one hour to remove excess oxygen.

Activity assay vials that have been in the anaerobic tent for at least 24 hours were capped with small rubber septa inside the anaerobic tent. Once removed from the tent, they were kept under positive argon pressure on the vacuum manifold. To perform the assays, 2 mL cell culture was injected into each of two assay vials. 1 mL acetylene was injected into the vials, and they were immediately placed in a water bath at 30° C and 120 rpm. After exactly 20 minutes, the assays were quenched with 0.25 mL concentrated citric acid. A negative control containing 2 mL water instead of cell culture was also prepared. A 50 µL sample of the headspace of each vial was injected into the GC. All GC runs were done at 110° C for 5 minutes. The area and percent area of the three peaks (methane/impurity at ~0.16 minutes, ethylene at ~0.8 minutes, and acetylene at ~2.3 minutes) were detected with a flame ionization detector (FID) and recorded. Calculations were done as described in section II.D.ii.

II.D.ii. Protein component acetylene reduction assays

Protein component acetylene reduction assays were performed to characterize the activity of purified protein and MoFe and Fe protein titrations. Each assay includes the MoFe protein, the Fe protein, an electron source (such as dithionite), ATP, an ATP regenerating system (creatine phosphokinase, phosphocreatine disodium, and magnesium chloride), and acetylene. The ATP regenerating system is required since ADP can inhibit nitrogenase.[56] Experimental details are provided in Table II-5. Solutions of 250 mM ATP, 2.5 mg/mL creatine phosphokinase (PCK), 100 mM phosphocreatine disodium (PC), and 500 mM $MgCl_2$ were made by dissolving the solid in assay buffer (0.5 M Tris/HCl at pH 7.5, 200 mM NaCl). ATP, PCK, and PC were kept on ice until added to the assay mixture in order to avoid degradation. Higher Tris concentration is used for the assays to ensure the pH is properly maintained, thereby avoiding decomposition of dithionite upon acidic conditions.

Experiment	Component ratio	Cp1 (mg)	Cp1 (nmole)	Cp2 (mg)	Cp2 (nmole)	Cp2/Cp1	Cp1 (µL)	Cp2 (µL)	Buffer (µL)
control	0	0.03	0.135	0	0	0	50	0	250
1	0.25	0.03	0.135	0.004	0.068	0.5	50	12.5	237.5
2	0.5	0.03	0.135	0.008	0.135	1	50	25	225
3	1	0.03	0.135	0.016	0.271	2	50	50	200
4	1.5	0.03	0.135	0.024	0.406	3	50	75	175
5	2	0.03	0.135	0.032	0.542	4	50	100	150
6	3	0.03	0.135	0.048	0.812	6	50	150	100
7	5	0.03	0.135	0.081	1.354	10	50	62.5 *	187.5
8	8	0.03	0.135	0.129	2.167	16	50	100 *	150
9	10	0.03	0.135	0.162	2.708	20	50	150 *	100
10	15	0.03	0.135	0.242	4.062	30	50	187.5 *	62.5
11	20	0.03	0.135	0.323	5.416	40	50	250 *	0

Table II-5. The following conditions lead to a total reaction volume of 1.300 mL (including 1.000 mL assay buffer and 0.300 mL protein and SEC buffer). Typically, experiments were performed in triplicate. The buffer used in the assays consists of 50 mM Tris/HCl and 200 mM NaCl at pH 7.5. Three protein stocks were prepared: Cp1 at 0.6 mg/mL, Cp2 at 1.29 mg/mL, and Cp2 at 0.32 mg/mL. Experiments requiring the more concentrated Cp2 stock solution are marked with an asterisk. A more concentrated Cp2 stock solution is necessary for the larger component ratios because otherwise an excessive amount of the more dilute Cp2 stock solution would be required in the assay mixture.

To 7.2 mL of a 50 mM Tris buffer (pH 7.5), 2 mL of the PC solution, 200 µL of the ATP solution, 100 µL of the $MgCl_2$ solution, and 500 µL of the PCK solution were added in the stated order (to avoid precipitation). 1 mL of this solution was added to each assay vial. Vials were purged on a Schlenk line with 12 cycles of 2.5-minutes vacuum and 0.5-minutes argon flush. During the purging, the acetylene flask was prepared as described in the previous section. The MoFe and Fe protein stock solutions were diluted in the anaerobic chamber as needed with a 50 mM Tris/HCl pH 7.5, 200 mM NaCl solution. The final protein dilutions contained 5 mM sodium dithionite. After purging, 50 µL of 500 mM sodium dithionite solution (dissolved in 0.5 M Tris/HCl pH 7.5, 200 mM NaCl) was added to each vial. The final composition of the assay mixture was 50 mM Tris/HCl at pH 7.5, 5 mM ATP, 5 mM $MgCl_2$, 20 mM PC, 2.5 mg/mL PCK, and ~19 mM sodium dithionite. The nitrogenase proteins are sensitive to salt concentration; therefore, the ionic strength should be kept constant (~at 48 mM NaCl) in each assay vial. Adjustments may be made using the salt concentration of the protein-free buffer as needed.

Prior to performing the titration experiments, a calibration curve was constructed by injecting 20, 30, and 40 µL of acetylene into the GC and determining the moles of acetylene in each measurement from the resulting peak area (Chart II-7). Three trials of each protein component ratio were typically run. For each assay vial, 1 mL headspace was replaced with 1 mL acetylene after which the vial was incubated at 30° C for 5 minutes. After incubating, MoFe protein and SEC buffer were added. Addition of Fe protein initiated the reaction. Assays were run for 10 minutes while incubating a water bath at 30° C and 120 rpm. After exactly 10 minutes, reactions were quenched with 0.25 mL concentrated citric acid. 50 µL of headspace was injected into the GC. All GC runs were done at 110° C for 5 minutes. The area and percent area of the methane/impurity, ethylene, and acetylene peaks were detected with a flame ionization detector (FID) and recorded.

To calculate the specific activity of the protein as a function of component ratio or dilution, ethylene formation is calculated for each vial by $v = (P/216.92)*((8.80-t)/0.05)/10$, where P is the original ethylene peak area reading and t is total volume (mL) of liquid in the vial (1.3 mL). 216.92 is the correction factor for converting the peak area to nanomoles ethylene; 8.80 refers to the total volume of the activity assay volume in mL; 0.05 refers to 50 µL headspace injected into the GC, and 10 is the reaction time in minutes. The ethylene peak area should be corrected by subtracting the ethylene peak area of a control assay to account for ethylene impurities in the acetylene. The maximum ethylene formation velocity is evaluated by Michaelis-Menton fitting of the resulting titration curve (titrant amount vs. v). The maximum velocity divided by the amount of titrated protein in the vial is the final specific activity presented in nanomole ethylene per minute per milligram protein. A sample titration curve is shown in Chart II-8.

Chart II-7. Sample calibration curve for acetylene reduction activity assays

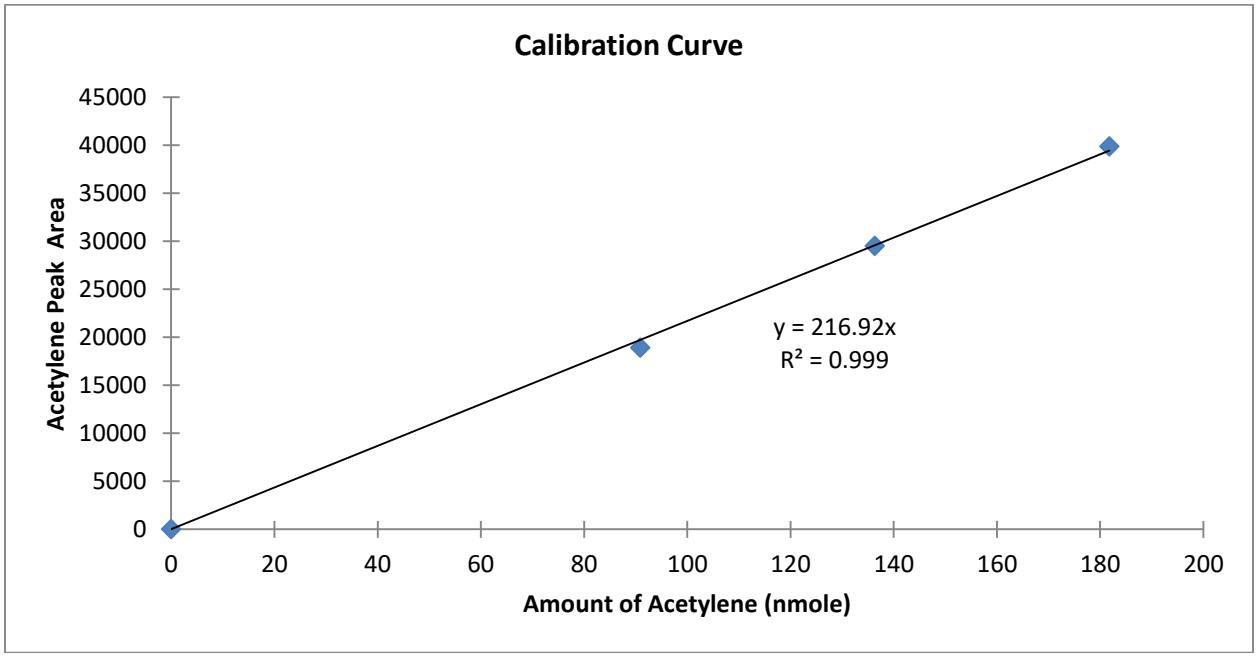
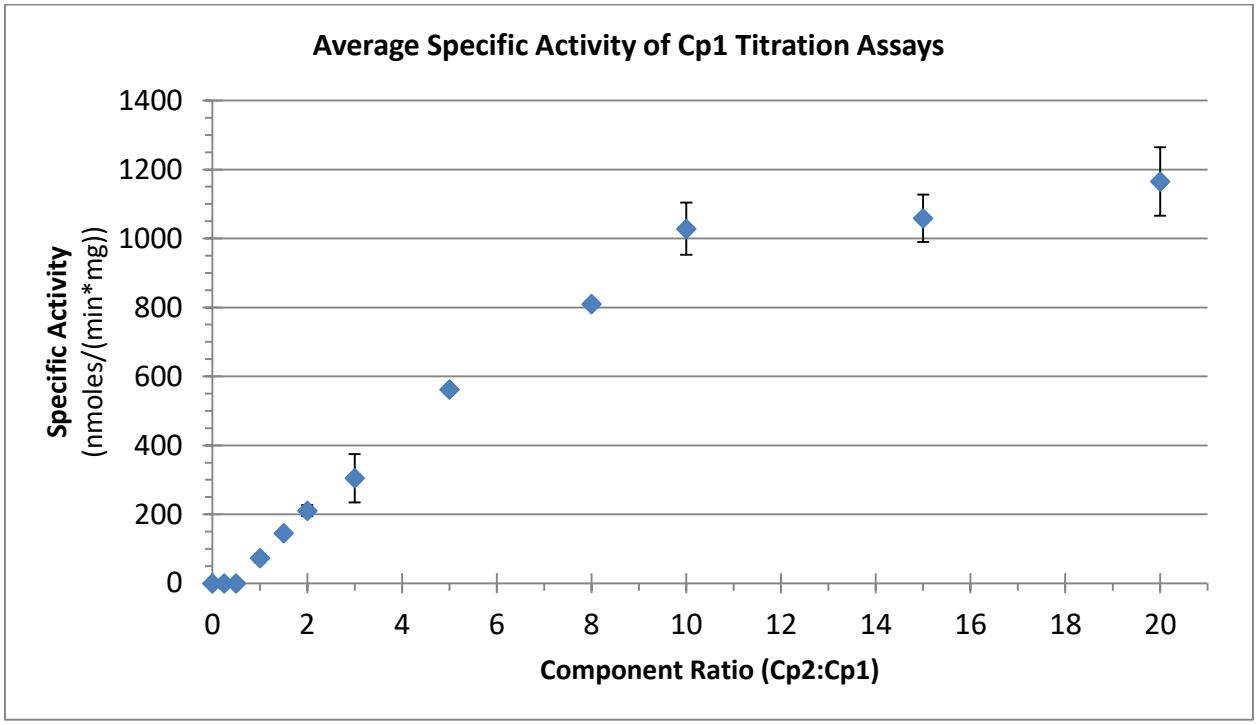


Chart II-8. Average specific activity of Cp1 titration assays



II.E. Crystallization

24-well Cryschem Plates were purchased from Hampton Research. Plates and 1.5-mL Eppendorf tubes were brought into the anaerobic tent (95% argon, 5% hydrogen) at least two days prior to use to allow all oxygen to diffuse out of the plastic. All solutions used in crystallization were filtered with a 0.22 µm, 50-mL Steriflip sterile centrifuge tube (EMD Millipore) and were then purged of oxygen using Schlenk lines (10 cycles of 3-minutes vacuum and 40-seconds argon) with stirring. Care was taken when subjecting poly(ethylene)glycol (PEG) solutions to vacuum so as not to suck up bubbles into the vacuum line.

In the anaerobic tent, crystallization solutions for each well were prepared in 1.5 mL Eppendorf tubes. The appropriate amount of double-distilled water and PEG was added to each tube. Tubes were then closed and inverted 20 times to create a homogenous solution. All remaining components, such as salt and Tris, were added. Additionally, freshly prepared 5 mM

sodium dithionite was added to each condition. Tubes were closed and inverted another 20 times. Tubes were allowed to sit for at least 45 minutes at room temperature before setting trays. During that time, protein was removed from the liquid nitrogen dewars and brought into the anaerobic tent to thaw.

Crystallization plates were setup using the sitting drop vapor diffusion method (Figure II-11). To each reservoir of the first 12 wells, 500 or 1000 µL crystallization solution was added. 0.5-2.0 µL protein was then added to the crystal well followed by an equal amount of reservoir solution. These 12 wells were then sealed with clear tape. The process was repeated for the remaining 12 wells. Plates stood undisturbed at room temperature. Most crystals started to form within 24 hours. Plates were checked for crystals using a low power setting on a microscope in the anaerobic tent to avoid heat or radiation damage to the crystals. Typical crystal morphologies observed included bars, triangular prisms, and hexagonal prisms (Figure II-9).

Figure II-9. (left) In sitting drop vapor diffusion crystallization, each well is sealed and water in the crystallization drop evaporates. At critical concentration, the protein precipitates or crystallizes. (right) Typical crystal morphologies observed include blocks, triangular prisms, and hexagonal prisms.

II.F. X-ray crystallography and structure solution

II.F.i. Crystal harvesting onto loops for cryogenic X-ray or neutron diffraction

The most reliable cryo-protection method was adding 2-methyl-2,4-pentanediol (MPD) and a few drops of Fomblin Y 16/6 oil directly to the crystallization drop. First, the sealing tape was sliced away. Under a microscope, the skin (a thin film of protein) of the crystal drop was removed by gently dragging a needle over the surface of the drop. Next, 10 µL reservoir solution and 1-2 µL degassed 100% MPD was added to the drop. Under a microscope, a needle was used to gently mix the solution around the crystal. A small layer of degassed Fomblin Y 16/6 oil (~3 drops) was added to the crystal drop, which remained on the top of the drop. Cryogenic loops were used to fish crystals out of the wells. Once looped, a crystal was slowly dragged through the oil layer. If the crystal was surrounded by an aqueous solution bubble, it was carefully dragged through the drop until the bubble disappeared. Once removed from the well, the looped crystal was immersed in liquid nitrogen as quickly as possible. After equilibrating to the liquid nitrogen temperature, it was carefully placed inside a cryo-vial. Crystals were stored in liquid nitrogen dewars.

II.F.ii. Crystal harvesting into capillaries for room temperature X-ray or neutron diffraction

For room temperature experiments, crystals must be in air-tight quartz capillaries. First, quartz capillaries were brought into the anaerobic tent at least 48 hours prior to harvesting to allow any oxygen in the capillary to diffuse out. At the time of harvesting, the tape on the crystallization well was removed, as was the skin on top of the crystal drop, as described in section II.F.i. 10 µL reservoir solution was added to the crystal drop. An additional 20 µL reservoir solution was transferred to a clean, random spot on the top of the plate. The crystal was fished with a loop and transferred to the drop on the top of the plate.

Next, the capillary, with a diameter about 1/3 larger than the crystal, was prepared. A tiny piece of the capillary was snipped off the small opening. If the opening was not even, it was gently tapped on a smooth surface or re-snipped. The larger opening of the capillary was inserted into piece of rubber tubing about 18 inches long, which was connected to a pipette tip with its small end snipped off by a few millimeters. Using a pipette, the crystal was *gently* sucked into the middle of the capillary.

Once the crystal was in the capillary, it was situated to roughly the middle of the capillary using the pipette. Next, most of the solution was removed from the capillary using wicks. The crystal remained in a drop of solution as small as possible (to minimize background scattering) with a nearby solution plug of at least a few millimeters in length to keep the capillary environment humid (Figure II-10). After the crystal and solution plug were positioned, the small opening of the capillary was sealed with wax. The large end was then snipped off and sealed. Care was taken to not heat the crystal while applying hot wax to the capillary. At this point, the capillaries could be removed from the tent. Capillaries were placed level in a small box and loosely held in place with a wadded Kim wipe.

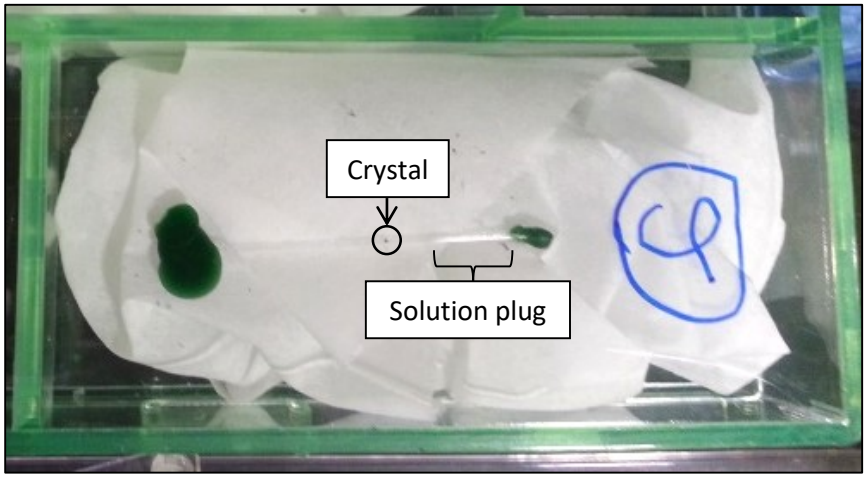


Figure II-10. The crystal is positioned in the middle with a small plug of solution nearby. The capillary is stored in a small box and is loosely held in place with a wadded Kim wipe.

II.F.iii. Data collection

Diffraction data was either collected in-house on a Rigaku MicroMax 007-HF X-ray generator with a Rigaku RAXIS-IV++ detector or remotely from the Stanford Synchrotron Radiation Lightsource (SSRL) on beamline 12-2 with a DECTRIS Pilatus 6M detector. For in-house experiments, 360 diffraction images were collected with an oscillation angle of 1° and 360° rotation. In a typical diffraction experiment at the SSRL, sets of 1440 diffraction images were collected at 12,999.97 eV with an oscillation angle of 0.25° (about half the mosaicity) and 360° rotation.

II.F.iv. Structure solution

Data sets were indexed and integrated with the XDS program package or MOSFLM.[57,58] Typically, MOSFLM was used to assess the quality of the reflections; indexing and integrating were performed with XDS. Scaling and merging were carried out with the CCP4 suite (POINTLESS and SCALA or AIMLESS), from which the resolution cutoff was determined by various metrics.[58] Phasing was calculated experimentally or by molecular replacement (PHASER or MOLREP). Initial structure refinement was carried out with CNS, and alternative conformations and isotropic *B*-factors were refined with REFMAC5.[59–61] For data sets near 1 Å resolution, anisotropic refinement was also performed. Simulated annealing was calculated using PHENIX.[62] Structures were further refined in COOT using the validation tools, including examining Ramachandran plot, searching for un-modeled blobs with electron density above 0.7 e/Å^3, and checking the difference density, waters, geometry, peptide omega angles, *B*-factors, Gln and Asn outliers, rotamer probability, and density fit. Figures were made in PyMOL.[63] For anomalous data, Friedel's Law was set to false during initial data processing. Following merging, structure factors were calculated using SFALL and Fourier transformation was carried out using FFT in CCP4.

Chapter III

Nitrogenase MoFe Protein from *Clostridium pasteurianum* at 1.08 Å Resolution: Comparison to the *Azotobacter vinelandii* MoFe Protein

This chapter summarizes my contributions to the publication of the same title written by Prof. Limei Zhang.[64] doi: 10.1107/S1399004714025243.

III.A. Abstract

The MoFe protein from *Azotobacter vinelandii* (Av1) is the most heavily studied nitrogenase protein for its structural and spectroscopic properties. Herein, we report the structural characterization of the most structurally divergent MoFe protein known, that from *Clostridium pasteurianum* (Cp1). The Cp1 structure was published in 1992 for the first (and, until this work, only) time with a resolution of 3.0 Å.[65] While this was useful for comparisons of the secondary protein structure, a detailed comparison of atomic positions, especially at the cofactors, could not be made. This work includes detailed comparisons of Av1 and Cp1, with an emphasis on the Fe protein docking site on the MoFe protein, the identity of the interstitial atom, and the P-cluster binding pocket. The results show that the structure of the cofactors is strictly conserved between Av1 and Cp1 homologs; however, differences in the secondary structure at the Fe protein docking site and in the P-cluster binding pocket may suggest some variation in the function of Av1 and Cp1.

III.B. Introduction

There are dozens of bacterial species that express nitrogenase; they are identified by the presence of nitrogen-fixing, "nif", genes in their genome.[24] Based on sequence alignment, these nitrogenases have been divided into six groups.[24] The two major groups are those with sequences similar to *Azotobacter vinelandii* (Av, group I) and *Clostridium pasteurianum* (Cp, group II). Groups III and IV are similar to Av but have small insertions or deletions; groups V and VI are nitrogenases that contain Fe or V in the FeMo-cofactor instead of Mo. The vast majority of nitrogenase work is performed using Av proteins.[19,22,28,37,38] By identifying similarities and differences between the MoFe proteins in different groups, we may further our understanding of the biological nitrogen fixation mechanism. Prior to this study, the only structure of a MoFe protein not from group I was the MoFe protein from Cp (Cp1) at 3.00 Å (PDB ID 1MIO).[21] This was a major accomplishment in understanding the greater structural differences between group I and II homologs; however, the resolution was too low to make definitive conclusions on atomic positions or the structure of the FeMo-cofactor. In this work, we report a high resolution structure of Cp1 and compare it to the high resolution structure of Av1 (PDB ID 3U7Q).[22]

III.B. Results

The published Cp1 structure, PDB ID 4WES, was solved to 1.08 Å resolution.[64] I used a dataset from a different crystal to create a structure solved to 1.12 Å resolution. The 1.12-Å Cp1 structure was determined by experimental phasing from multi-wavelength anomalous diffraction (MAD) data. Data processing and refinement statistics are provided in Table III-1. 10 of the 11 outliers in the Ramachandran plot include α-Ile48, α-Ser181, α-Gly345, α-Ser346, and α-Ala463, all of which are found in the interior of the protein and have well-defined electron density. These residues are not conserved in Av1, and may therefore be a characteristic of Cp1.

Four residues in the Cp1 structure have a peptide angle significantly distorted from 180°: α-Leu240 (~6°), α-Gly488 (~10°), β-Phe201 (~339°), and β-Phe411 (~351°). None of these residues are in areas of the protein identified as mechanistically important. The latter three exist in loop regions and are followed by a proline. α-Leu240 occurs between two threonine residues, the three of which make up the terminal strand of a beta sheet. α-Leu240 and α-Thr241 make a non-proline *cis*-peptide bond; this is the only non-proline *cis*-peptide bond observed in the MoFe protein.[22,66] In group I, the *cis*-peptide residues (α-Trp253 and α-Ser254 in Av1) are conserved 87%; those in group II (α-Leu240 and α-Thr241 in Cp1) are conserved 39%, indicating that the *cis*-peptide is more conserved among the Av1 homologs than the Cp1 homologs.

Table III-1. Data processing and refinement statistics for Cp1 at 1.12 Å

	Cp1 at 1.12 Å	
Data Collection Statistics		
Resolution range (Å)	38.91 – 1.12 (1.18-1.12)	
Wavelength (Å)	0.9537	
Space group	$P2_1$	
Unit cell constants	a = 72.8 Å	α = 90.00°
	b = 170.7 Å	β = 91.64°
	c = 87.5 Å	γ = 90.00°
Unique reflections	737,219 (86,732)*	
Completeness (%)	90.7 (73.1)*	
Redundancy	3.5 (3.4)*	
$I/\sigma(I)$	17.3 (3.2)	
R_{merge}	0.030 (0.317)*	
Refinement Statistics		
Protein residues	1951/1982	
Mean B value (Å^2)	11.6	
R_{work}	0.129	
R_{free}	0.151	
Ramachandran outliers	11 (0.59%)	
RMSD bond lengths (Å)	0.015	
RMSD bond angles (°)	1.80	

In the publication of the original Cp1 structure, two crystal forms, obtained from crystallization with magnesium chloride or cesium chloride (PDB ID 1MIO), were reported.[65] The unit cell dimensions of the structure presented here are similar to cesium chloride form; however, no cesium was used in the crystallization condition. Superposition of the secondary structure elements (1917/1951 residues) of 1MIO and the 1.12-Å structure gives a root mean square deviation of 0.152 Å.

III.C. Comparison to the Av1 structure

III.C.i. Insertion sequences

The shared sequence identity of Cp1 compared to Av1 is 37.7%. Superposition of the secondary structure elements (1738/1951 residues) of Av1 (PDB ID 3U7Q) and the 1.12-Å structure gives a root mean square deviation of 1.56 Å. The metalloclusters are identical, and the secondary structure elements are similar with two notable exceptions (Figure III-1): (1) an insertion sequence in the Cp1 α-subunit on the protein surface involving Cp1 residues α-Glu376–α-Gly428 and (2) a deletion in the Cp1 β-subunit on the protein surface involving Av1 residues β-Met1–β-Thr48.[21] Herein, these sequences are referred to as the Cp1 and Av1 insertion sequences, respectively.

The Cp1 insertion sequence consists of a flexible loop flanked with alpha helices; loop residues α-387–α-391 block the docking of the Fe protein when compared to Av1-Av2 complexes.[28,67] This is according to overlays with four different Av1-Av2 complexes (Figure III-2). These complexes were formed using additives to mimic nucleotide-free and nucleotide-bound conditions representing different conformations at various stages of electron transfer during substrate reduction.[28,67] Notably, Av1 and Cp2 form an inactive complex,[68,69] suggesting that there is something unique about the Cp1-Cp2 complex compared to the other nitrogenase

species, which likely involves the insertion sequence. Perhaps a different docking formation is adopted, or perhaps the loop flexes away from the protein in solution to open up the docking site. The latter is supported by the fact that there are only four polar contacts between residues in the loop portion of the insertion sequence and the rest of the protein.

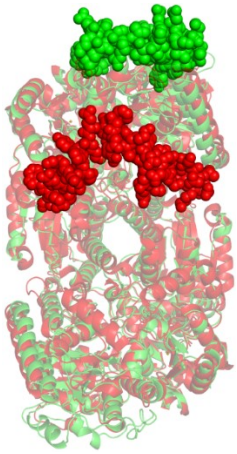


Figure III-1. Overlay of Cp1 (green) and Av1 (red, PDB ID 3U7Q) with insertion loops displayed as spheres. The insertion loops are shown in only one of the heterodimers.

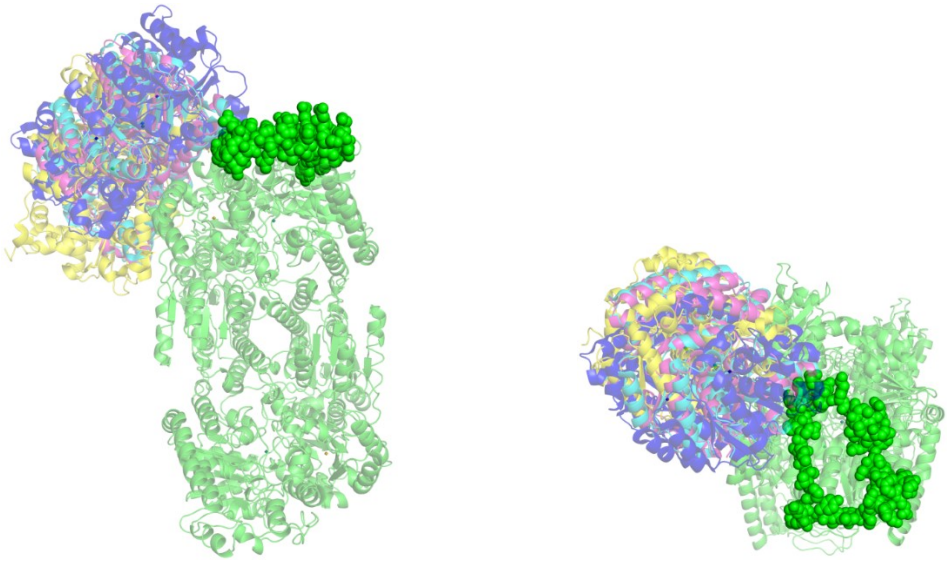


Figure III-2. Two orientations of overlays of Av2 (various colors) and Cp1 (green). The complexes include ADP·AlF_4^- stabilized Av1-Av2 (PDB ID 1M34, magenta), nucleotide-free Av1-Av2 (PDB ID 2AFH, yellow), MgAMPPCP-bound Av1-Av2 (PDB ID 2AFK, cyan), and MgADP-bound Av1-Av2 (PDB ID 2AFI, blue).[27,28] Av1 was used to create the superposition, but is hidden in the figure.

The Av1 insertion sequence initiates the β-subunit: the first quarter consists of a loop, followed by an alpha helix, another loop section, and another alpha helix. The Av1 insertion sequence partially covers opening of the interstitial water channel that connects the protein surface to the active site. This may impede access of small molecules to and from the active site. There are many polar contacts between the Av1 insertion sequence and the rest of the protein, which may indicate that the insertion sequence is not flexible.

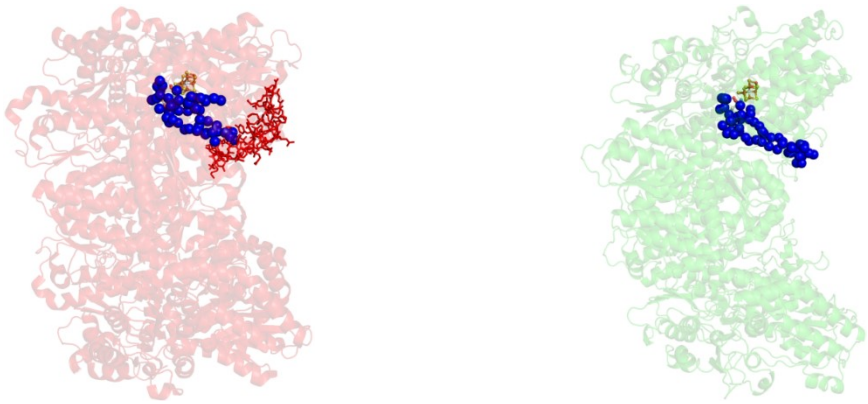


Figure III-3. (left) Av1 and (right) Cp1 with the water molecules of the interstitial channel shown in blue spheres. The residues of the Av1 insertion loops are displayed as sticks in the left figure. The Av1 insertion loop partially blocks the mouth of the interstitial channel. The FeMo-cofactor in both figures is shown with ball and stick and colored by element.

III.C.ii. Inter-subunit contacts

The contacts between the α, β, α', and β' subunits of Cp1 were determined using the program NCONT in the CCP4 suite (Table III-2).[58] A contact is defined as a distance of ≤3.5 Å between two atoms of different subunits. Contacts that are hydrogen bonds or electrostatic interactions are identified as salt bridges. There is a similar number of total contacts between all subunits in Av1 and Cp1, and most are salt bridges. The total number of salt bridges between all subunits is also similar, except in the case of the Cp1 β and β', which have about 50% more salt

bridges compared to other subunit interactions. The conservation of total contacts and salt bridges indicates that the interaction between subunits is not highly conserved.

Table III-2. Inter-subunit contacts in Av1 and Cp1

Subunits	Av1 β, β'	Cp1 β, β'	Av1 α, β	Cp1 α, β
Total number of contacts	58	59	58	59
Number of salt bridges	33	46	31	33
Conservation of total contacts in Av1 and Cp1	24 (41%)		29 (54%)	
Conservation of salt bridges in Av1 and Cp1	25 (63%)		17 (55%)	

III.C.iii. Conservation of tryptophan

The conservation of tryptophan residues was investigated since this amino acid is often involved in protein electron transfer reactions.[70] There are only six tryptophan residues in Cp1 (Figure III-4), compared to 18 in Av1 and 15 in Kp1. Of the six Cp1 tryptophan residues, four are conserved in Av1 and Kp1, including α-Trp223, α-Trp281, α-Trp512, and α-Trp519 (Table III-3). Only α-Trp84 resides near a cofactor; thus, this residue may have a role for in the substrate reduction mechanism. The five tryptophan residues in the α-subunit are all highly conserved.

Table III-3. Conservation of Cp1 tryptophan residues in groups I and II

	α-Trp84	α-Trp223	α-Trp281	α-Trp512	α-Trp519	β-Trp380
Conservation in group I (single variant mutant)	82% (Arg)	100%	100%	98% (Ile)	87%	0%
Conservation in group II (single variant mutant)	100%	72%	100%	94% (Phe)	89% (Phe)	67%
Overall conservation	87%	92%	100%	97%	87%	19%
Distance to FeMo-co (Å)	12	15	15	21	38	26
Distance to P-cluster (Å)	11	28	30	30	41	34

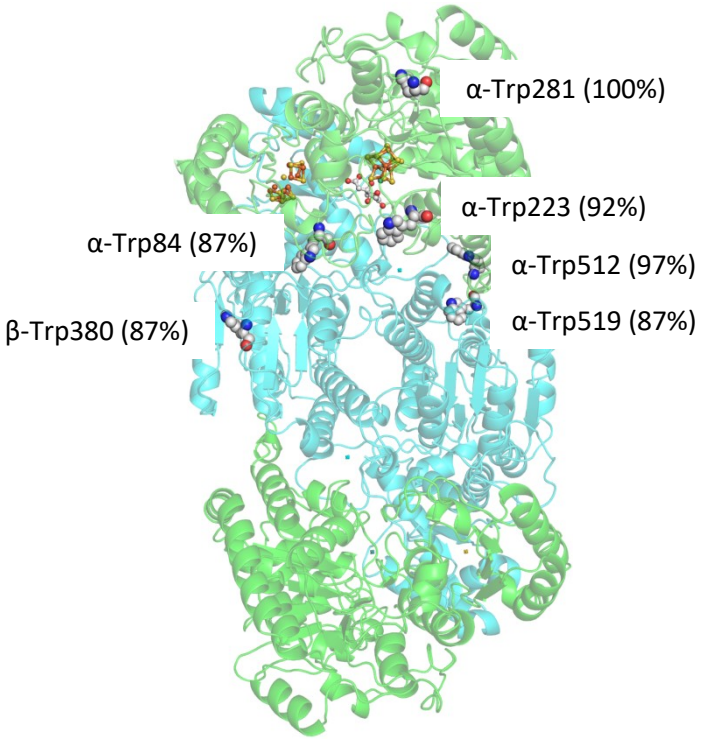


Figure III-4. Location of Cp1 Trp residues; their conservation in groups I and II are given in parentheses. The α subunits are shown in green and the β subunits are shown in cyan. The cofactors and the tryptophan residues are displayed with ball-and-stick and are colored by element.

III.C.iv. FeMo-cofactor insertion pathway

A pathway has been postulated for FeMo-cofactor insertion into the apo-MoFe protein.[71] This pathway was found in a structure of the Δ*nifB* protein, which lacks the FeMo-cofactor; the pathway is of sufficient width and is positively charged so as to accommodate the size and negative charge of the FeMo-cofactor. This pathway consists of three sections: (1) the lid-loop region (α-353–α-364), which may guide the FeMo-cofactor to the pathway, (2) the His-triad (α-His274, α-His442, and α-His451), which may provide binding site for the FeMo-cofactor during its journey, and (3) the switch/lock region (α-His442 and α-Trp444), which may secure the FeMo-cofactor in place by movement of residue side chains. Conservation of residues in the lid-loop, His-triad, and switch/lock regions are well conserved in group I (the Av1 homologs),

with a total of 88% conservation. The corresponding residues of the lid-loop, His-triad, and switch/lock regions in Cp1 are conserved in group II at 71%, suggesting that the FeMo-cofactor may have a different insertion pathway in Cp1 compared to Av1. Mutation studies as well as structural analysis of the Δ*nifB* Cp1 protein would be more conclusive.

III.C.v. Conservation of the contacts between the MoFe and Fe proteins

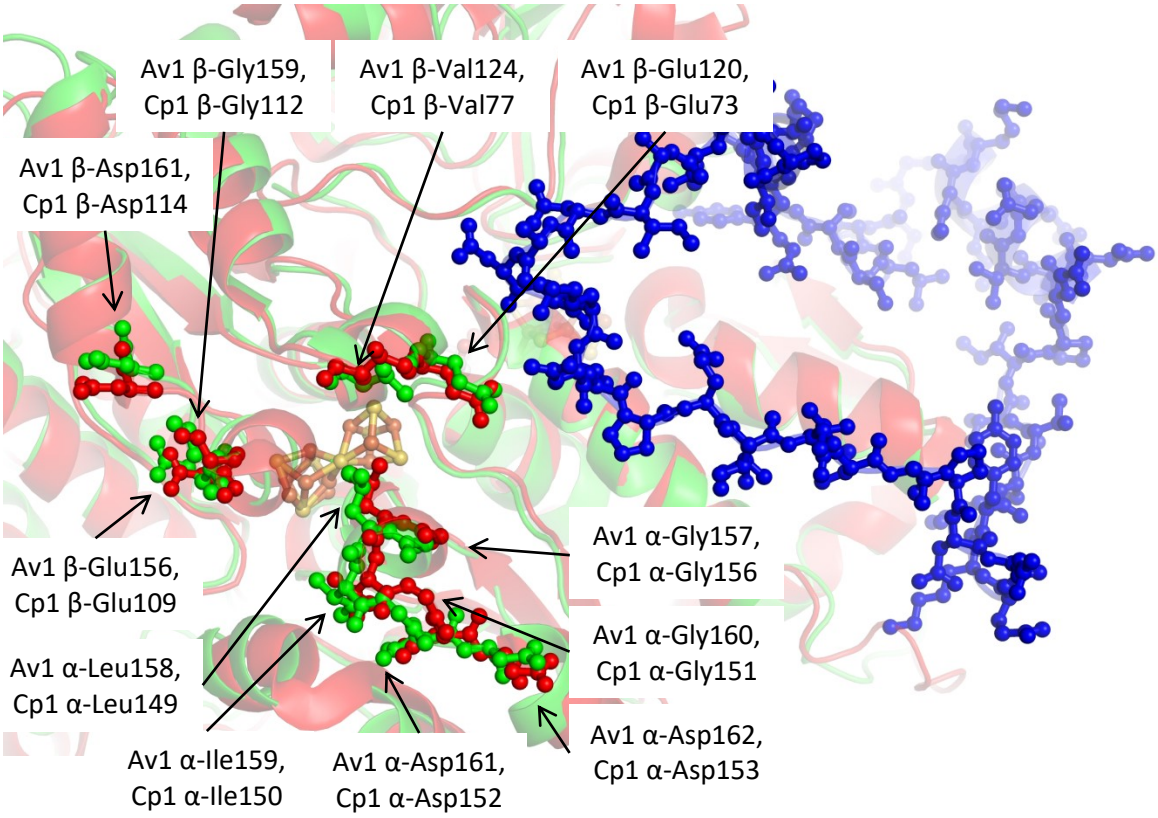


Figure III-5. Overlay of Av1 (red) and Cp1 (green). Of the 38 Av1 residues that have contact with Av2, 11 are conserved 100% among group I and II homologs. These residues are shown in ball-and-stick and are labeled. The P-cluster is shown in ball-and-stick and colored by element, and residues of the Cp1 insertion loop are shown in blue ball-and-sticks.

38 surface Av1 residues have contacts with Av2,[27,28] of which 15 are conserved 100% among group I homologs and may therefore serve in Fe protein docking. No Cp1-Cp2 structure has yet been reported. It is expected that the docking may have a different conformation due to

the Cp1 insertion sequence. Conservation of the Cp1 residues corresponding to the 18 Av1 residues with highly conserved contacts to Av2 was investigated. 11 of these 18 residues are conserved 100% in group II (Figure III-5). These residues are within close proximity of the P-cluster and are not blocked surface access by the Cp1 insertion loop. Because of this and the high conservation of these residues, they may be involved in the docking of Cp2 onto Cp1.

III.D. The interstitial atom of the FeMo-cofactor

Identification of the interstitial atom in the Av1 FeMo-cofactor as carbon was achieved in 2011 through a combination of X-ray crystallography, electron spin echo envelope modulation, and X-ray excitation spectroscopy.[22,72] Accurate electron density calculations from an X-ray crystal structure require atomic resolution due to Fourier series termination effects from the six iron atoms encasing the interstitial atom in a trigonal prism. This effect led to the masking of *all* electron density at the interstitial cavity for structures with resolutions worse than 1.5 Å and the incorrect identification of the interstitial atom as nitrogen in a 1.16 Å structure.[21,66,65,73–75]

Using the 1.08 Å Cp1 crystal structure, the electron density as a function of the radius from the atom center was calculated for all carbon, nitrogen, and oxygen atoms in the protein structure with a *B*-factor <30 Å^2 (Figure III-6). The electron density of the interstitial atom aligns best with the average carbon in the protein structure (Figure III-7). Although the interstitial atom of the FeMo-cofactor is likely electronically different from the average carbon in the protein scaffold, these methods were previously verified as reliable by spectroscopic techniques for use on structures at atomic resolution.[72]

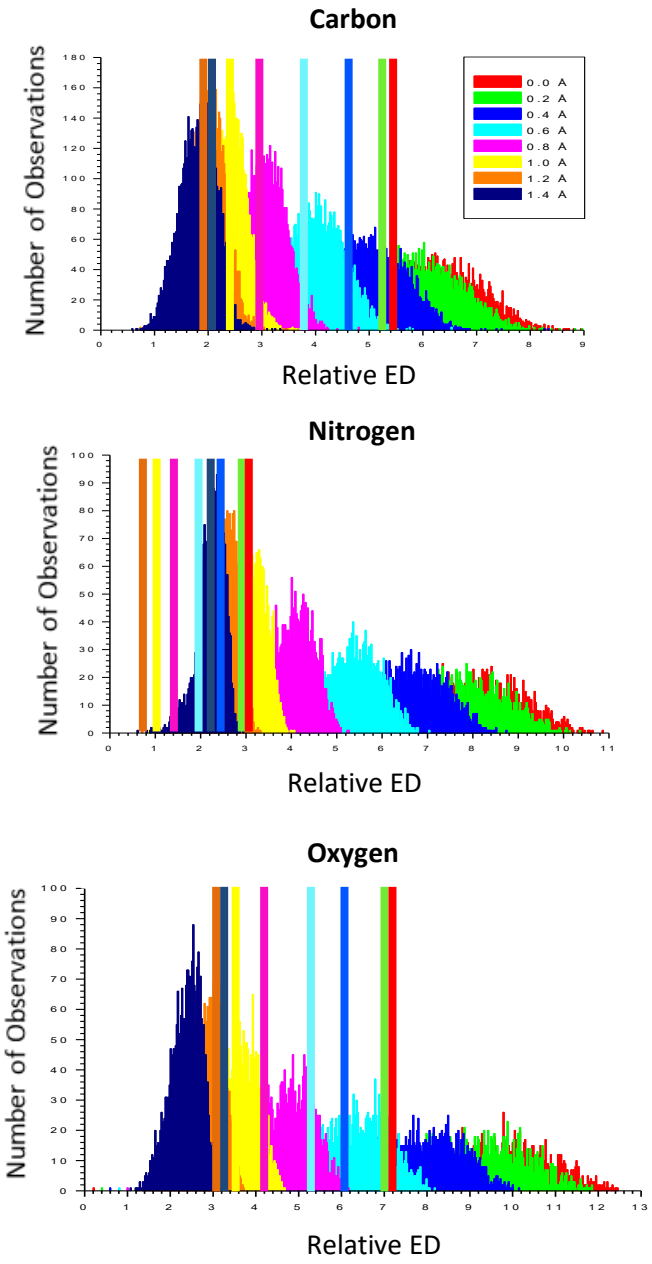


Figure III-6. Distribution plots of the electron density of all carbon, nitrogen, and oxygen atoms as a function of integration radius. The electron density of the interstitial atoms (solid lines) is overlaid on each plot. The overlay with the highest correlation is carbon. Two patterns are apparent from examination of the distribution plots. First, the distribution broadens as the integration radius decreases. This occurs because at larger integration radii, different distributions of the electrons in each electron cloud cause greater variations in the overall electron density. Second, the electron density decreases as the integration radius increases. Because the electron cloud is the most concentrated near the atom center, the average electron density at any given point in the electron cloud decreases at larger integration radii.

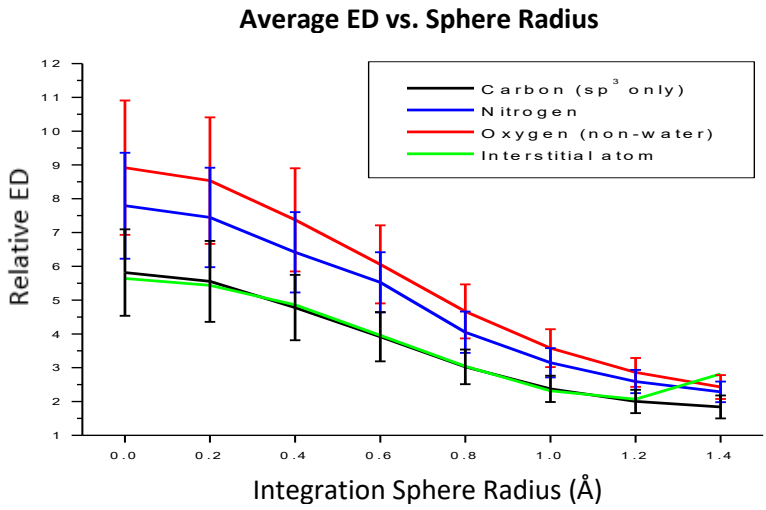


Figure III-7. Summary of results from Figure III-6 with Gaussian fitting. The error bars show one standard deviation from the mean electron density at each integration radius. The best correlation for the interstitial atom is carbon. The electron density of the interstitial atoms increases quite significantly at an integration radius 1.4 Å compared to other carbon atoms in the protein structure. This may result from incorporation of electron density of neighboring Fe atoms in the FeMo-cofactor, which are 2 Å from the interstitial carbon.

It was evident that there are different types of carbon and oxygen atoms in the protein structure from shoulders in the distribution plots of carbon and oxygen (Figure III-8). Further analysis showed that the right-hand shoulder in the carbon distributions is from sp^2 carbon atoms, which have greater electron density than sp^3 carbon atoms due to shorter bond lengths of double bonds compared to single bonds. The electron density of the interstitial atom correlates best with sp^3 carbon atoms compared to sp^2 carbon atoms. Although the interstitial carbon atom is coordinated to six iron atoms and exists as carbide (C^{4-}), the electron density distributions suggest that the iron atoms pull enough electron density away from the interstitial carbon such that it mimics the electron density of sp^3 carbons. Regarding the oxygen distributions, the left-hand shoulder is from oxygen atoms in water molecules (Figure III-8). The water oxygen atoms have less electron density than the protein backbone oxygen atoms

because the hydrogen atoms in a water molecule provide very little additional electron density compared to the carbon atoms in the protein scaffold.

From the X-ray crystallography data, the electron density at the atom center of each carbon, nitrogen, and oxygen atom was calculated as a function of *B*-factor and compared to that of the interstitial atoms (Figure III-9). Based on this comparison, the best fit for the interstitial atoms is carbon, as also concluded from electron density calculations and observed in Av1. This aligns with the observed minimal deviation in bond length (<0.09 Å) between all atoms in the cofactors (Table III-S1), which suggests that atomic identity is conserved in Av1 and Cp1.

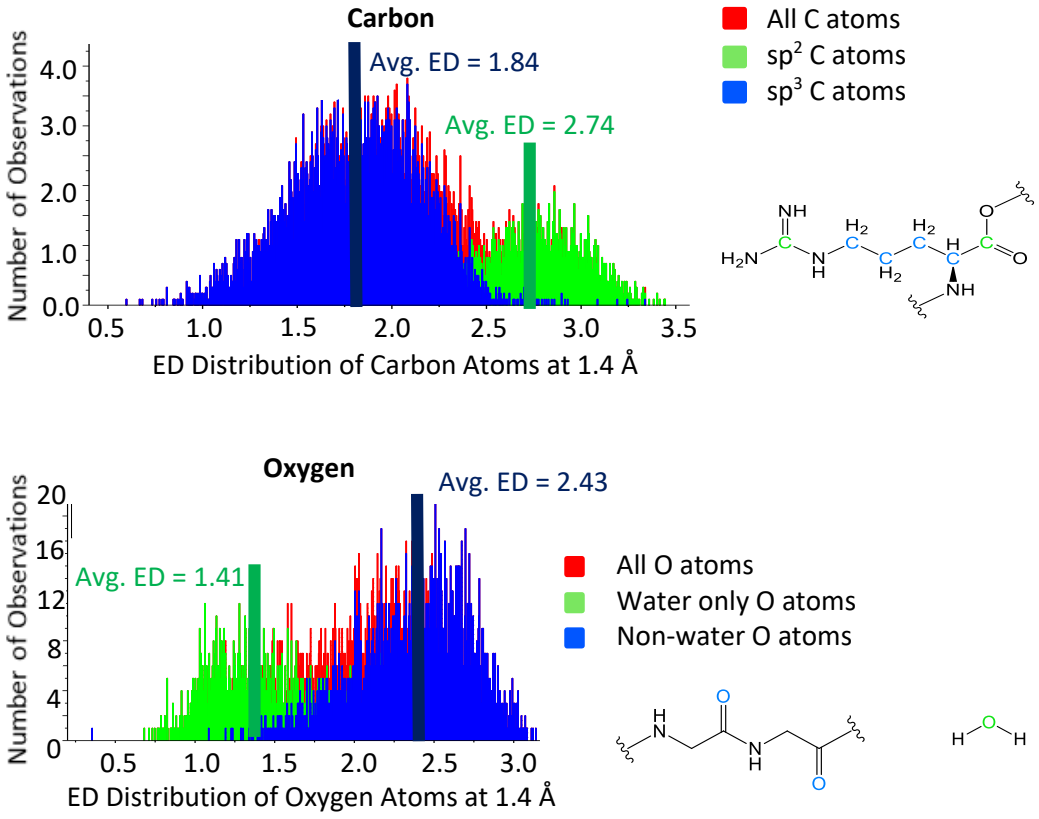


Figure III-8. Electron density distributions for carbon and oxygen at 1.4 Å show that different bonding environments lead to different electron densities.

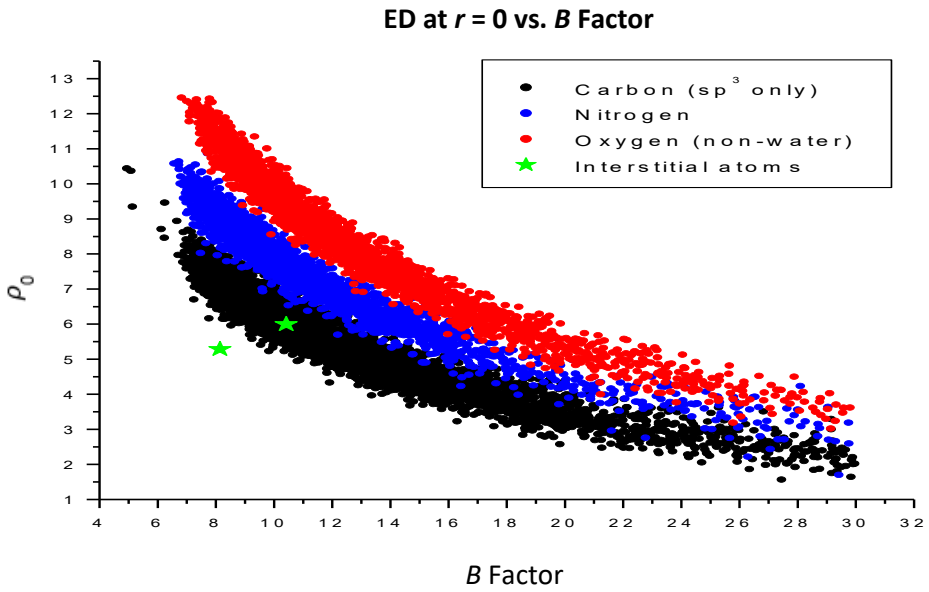


Figure III-9. Electron density of carbon, nitrogen, and oxygen atoms as a function of *B* factor.

III.F. The P-cluster

III.F.i. Comparing the reduced and oxidized states

As illustrated in Figure III-10, the P-cluster is stabilized by the following contacts in Av1: α-Cys154 to Fe1, α-Cys62 to Fe3, β-Cys95 to Fe2 and Fe8, α-Cys88 to Fe4, β-Cys70 to Fe7 (corresponding to Cp1 residues α-Cys145, α-Cys53, β-Cys48, α-Cys79, β-Cys23, respectively). In the open state of the P-cluster, Fe5 and Fe6 are stabilized by Av1 α-Cys88 (Cp1 α-Cys79) and Av1 β-Ser188 (Cp1 β-Ser141), respectively. In the closed state, Fe5 and Fe6 are stabilized by Av1 α-Cys88 (Cp1 α-Cys79) and Av1 β-Cys153 (Cp1 β-Cys106), respectively. The open and closed states of the P-cluster are the oxidized and dithionite-reduced forms, respectively.[75] X-ray crystal structures of Cp1 and Av1 typically show that the P-cluster is in the dithionite-reduced state or a mixed state. In a mixed state, Fe5 and Fe6 are modeled in both the closed and open conformations with a combined occupancy of 100%.

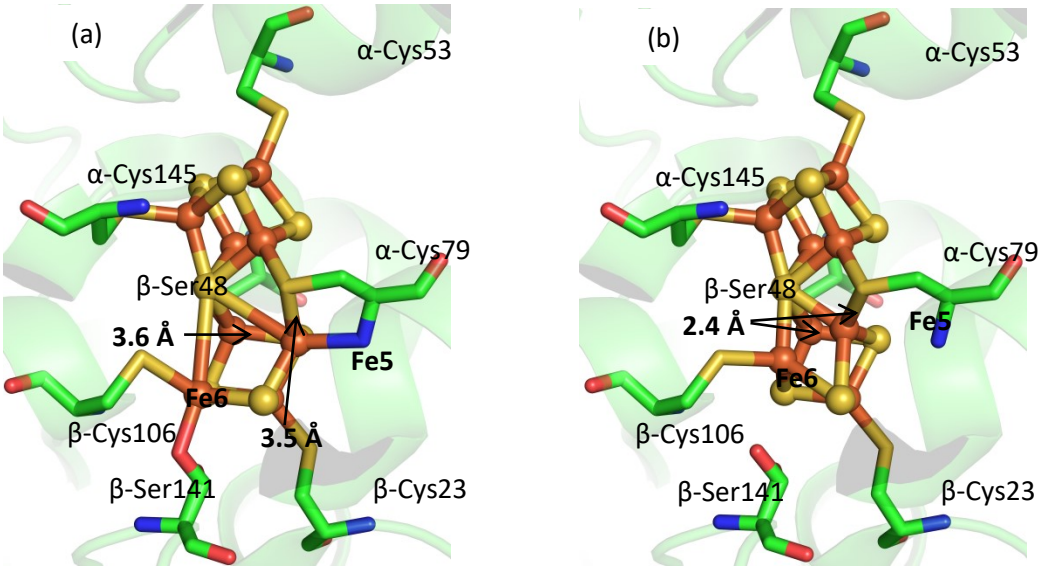


Figure III-10. The open (a) and closed (b) states of the P-cluster, corresponding to the putative oxidized and dithionite-reduced states, respectively. Fe5 and Fe6 are labeled as well as the bond length between these atoms and S1. The distance between S1 and Fe5,6 decreases from ~3.6 Å to 2.4 Å when going from the open to closed state. In both figures, the protein scaffold is shown in green cartoon and residues coordinated to the P-cluster are shown in sticks. The atoms of the P-cluster are represented with spheres and colored by element.

Residue β-Ser92 in Av1 may have a functional role in the MoFe protein, as deduced from its two alternate conformations in the reduced and oxidized Kp1 structures.[66] In the oxidized-state, the serine sidechain hydrogen bonds to a water molecule; in the reduced-state structure, the sidechain hydrogen bonds with S2A in the P-cluster. In the (dithionite-reduced) Cp1 structure presented here, the distance between Cp1 β-Ser45 and S2A is 2.9 Å. The dependence of this residue on the oxidation state of the cofactors and its strict conservation in groups I and II suggests that it could be involved in redox reactions and/or may provide structural support for the reduced-state P-cluster.

When the P-cluster is in the reduced state, Fe5 and Fe6 open up so as to form more hydrogen bonds between Fe6 and β-Ser188 in Av1 and β-Ser141 in Cp1. In group II, this serine is strictly conserved. In group I, however, β-Ser188 is conserved only 60%, and the only variant is

alanine. Although alanine is roughly the same size as serine, it is not capable of electrostatic interactions with the P-cluster. Interestingly, there are no nearby residues able to interact with Fe6 in the case of the Ser188Ala substitutions in Av1. This suggests that this serine may not be essential to the structure or function of nitrogenase.

III.G. Conclusions

This work showcases the first high resolution structure of a MoFe protein from group II, allowing detailed comparisons to be made between the most structurally divergent nitrogenase MoFe protein groups known. Notably, the interstitial carbon atom in the FeMo-cofactor in conserved, from which we can conclude that the structure of the cofactors (including the P-cluster) are identical in groups I and II. There are, however, differences in the structure of the protein scaffold that suggest alternative Fe protein docking, substrate and product pathways, FeMo-cofactor insertion pathways, and cofactor environment. These differences and their potential impact on the mechanism of substrate reduction will be better understood by characterizing Av1 and Cp1 under turnover conditions as well as Cp1-Cp2 complexes.

III.H. Experimental (see publication for further detail)[64]

III.H.i. Protein purification and crystallization

Purification of Cp1 was performed as previously described.[21] Cp1 was crystallized using the sitting-drop vapor diffusion method with a reservoir solution containing 0.2 M lithium citrate and 20% (w/v) PEG 3350 at room temperature in an anaerobic chamber. The resulting crystals were soaked in reservoir solution containing 10% 2-methyl-2,4-pentanediol (MPD) and 5 mM sodium dithionite for 15 min before flash-freezing in liquid nitrogen.

III.H.ii. Data collection and processing

X-ray diffraction data were collected on beam line 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL), with a Pilatus 6M pixel array detector. The 1.12 Å-resolution diffraction data set was collected at 14,000 eV with an oscillation angle of 0.15°. The data were integrated with the XDS program package.[57] The model was built manually in COOT using the electron density map from experimental phasing,[21,76] and the structure was refined with REFMAC5.[60] Structure validation analysis was done using MolProbity v4.02.[77] Intermolecular contacts were analyzed using the CCP4 program CONTACT.[58] Figures were prepared in PyMOL.[63]

Electron density analyses was carried out using proprietary software. First, all atoms of interest were selected from the PDB file, which include those with *B*-factor <30 $Å^2$ and a correlation factor of 1. The electron density map was then divided into a 400 x 600 x 400 grid, yielding about 9.6 x 10^7 cubes, each with dimensions of 0.182 x 0.284 x 0.219 Å. Due to the small size of each cube compared to the size of an atom, it is assumed that each cube has uniform electron density. The masked electron density distribution file was then overlaid onto the PDB file. The program calculated the electron density of each atom at a given radius from the atom center by averaging the electron density of each box contained in the given radius. The electron density was calculated for each atom from the atom center (0.0 Å) to 1.4 Å with 0.2 Å intervals.

III.H.iii. Protein Data Bank accession numbers

Atomic coordinates and structure factors for the 1.08 Å structure were deposited in the Protein Data Bank with ID 4WES.

III.I. Acknowledgments

The work was supported by NIH grant GM45162 to D.C.R., and the National Sciences and Engineering Research Council of Canada and Caltech Center for Environmental Microbial Interactions fellowships to L.M.Z. We thank the staff at beam line 12–2, Stanford Synchrotron Radiation Lightsource (SSRL) for their assistance. We acknowledge the Gordon and Betty Moore Foundation, the Beckman Institute, and the Sanofi–Aventis Bioengineering Research Program at Caltech for their generous support of the Caltech Molecular Observatory.

III.J. Supporting information

Table III-S1. Comparison of Bond Lengths in the FeMo-cofactor and P-cluster of the Atomic Resolution structures of Av1 (PDB ID 3U7Q) and Cp1[22]

Bond	3U7Q	1.12 Å Cp1	Difference
FeMo-cofactor			
FE1-S1A	2.31	2.29	0.02
FE1-S2A	2.27	2.31	0.04
FE1-S4A	2.29	2.30	0.005
S1A-FE2	2.26	2.28	0.02
S1A-FE4	2.29	2.28	0.015
S2A-FE2	2.25	2.30	0.045
S2A-FE3	2.28	2.28	0.005
S4A-FE4	2.29	2.29	0.005
S4A-FE3	2.25	2.29	0.035
FE2-S2B	2.21	2.24	0.03
FE2-CX	2.01	2.02	0.01
FE3-S5A	2.23	2.23	0
FE3-CX	1.99	2.00	0.01
FE4-S3A	2.24	2.24	0.005
FE4-CX	2.00	1.97	0.03
FE5-S3A	2.26	2.20	0.06
FE5-CX	2.01	1.98	0.03
FE5-S1B	2.27	2.22	0.05
FE5-S4B	2.26	2.25	0.01
FE6-S2B	2.18	2.23	0.05
FE6-CX	2.02	2.03	0.015

FE6-S3B	2.22	2.25	0.025
FE6-S1B	2.24	2.25	0.01
FE7-S5A	2.21	2.23	0.015
FE7-CX	1.99	2.01	0.02
FE7-S4B	2.22	2.24	0.02
FE7-S3B	2.25	2.23	0.02
S1B-MO1	2.36	2.35	0.005
S3B-MO1	2.37	2.35	0.02
S4B-MO1	2.35	2.38	0.03
P-Cluster			
FE3-S2A	2.29	2.33	0.04
FE3-S3A	2.32	2.35	0.03
FE3-S4A	2.26	2.30	0.04
S2A-FE1	2.31	2.35	0.04
S2A-FE2	2.30	2.31	0.01
S3A-FE4	2.30	2.33	0.03
S3A-FE1	2.31	2.34	0.025
S4A-FE2	2.28	2.32	0.035
S4A-FE4	2.29	2.29	0
FE1-S1	2.39	2.38	0.01
FE2-S1	2.48	2.48	0.005
FE4-S1	2.38	2.39	0.01
S1-FE5,B	2.60	2.54	0.065
S1-FE5,A	3.81	3.72	0.085
S1-FE6,B	2.56	2.53	0.035
S1-FE6,A	3.86	3.80	0.065
S1-FE8	2.36	2.39	0.03
FE5,A-S4B	2.31	2.34	0.025
FE5,B-S4B	2.21	2.25	0.04
FE5,A-S2B	2.27	2.22	0.05
FE5,B-S2B	2.45	2.47	0.02
FE6,A-S2B	2.33	2.35	0.015
FE6,B-S2B	2.57	2.52	0.045
FE6,A-S3B	2.36	2.34	0.015
FE6,B-S3B	2.32	2.33	0.01
FE8-S3B	2.28	2.28	0.005
FE8-S4B	2.32	2.33	0.005
S2B-FE7	2.32	2.32	0
S3B-FE7	2.32	2.33	0.005
S4B-FE7	2.29	2.32	0.03

Chapter IV

Substrate Pathways in the Nitrogenase MoFe Protein by Experimental Identification of Small Molecule Binding Sites

This chapter is a copy of the article published in *Biochemistry* under the ACS AuthorChoice license (http://pubs.acs.org/doi/full/10.1021/bi501313k).[78] Some additional work is also included herein, and it is noted accordingly.

IV.A. Abstract

In the nitrogenase molybdenum-iron (MoFe) protein, we have identified five potential substrate access pathways from the protein surface to the FeMo-cofactor (the active site) or the P-cluster using experimental structures of Xe pressurized into MoFe protein crystals from *Azotobacter vinelandii* and *Clostridium pasteurianum*. Additionally, all published structures of the MoFe protein, including those from *Klebsiella pneumoniae*, were analyzed for the presence of non-water, small molecules bound to the protein interior. Each pathway is based on identification of plausible routes from buried small molecule binding sites to both the protein surface and a metallocluster. Of these five pathways, two have been previously suggested as substrate access pathways. While the small molecule binding sites are not conserved among the three species of MoFe protein, residues lining the pathways are generally conserved, indicating that the proposed pathways may be accessible in all three species. These observations imply that there is unlikely a unique pathway utilized for substrate access from the protein surface to the active site; however, there may be preferred pathways such as those described here.

IV.B. Introduction

Nitrogen fixation is the process by which atmospheric dinitrogen (N_2) is reduced to a biologically active form of nitrogen, ammonia (NH_3).[79] This reaction is achieved on the industrial scale by the Haber-Bosch process, producing enough ammonia for nitrogen fertilizers to sustain 27−40% of the world's population.[80] Because of the dependence of the Haber-Bosch process on molecular hydrogen obtained from natural gas, this process accounts for more than 1.5% of the global energy consumption each year.[80] Industrial nitrogen fixation uses heterogeneous iron catalysts, pressures near 250 atm, and temperatures between 400 and 600 °C to reduce dinitrogen.[80,81] In contrast, the biological catalyst, nitrogenase, reduces N_2 to NH_3 at ambient temperature and atmospheric pressure. Understanding the process by which nitrogenase functions may facilitate the development of environmentally cleaner alternatives to the Haber-Bosch process, making nitrogenase an attractive enzyme to study for biotechnological NH_3 production. Although the enzyme has been studied for many decades, the detailed mechanism of N_2 reduction remains poorly understood. For example, certain stoichiometric aspects of the standard model of biological nitrogen fixation (Scheme IV-1) are still under discussion, including the ATP/e^- ratio and the obligatory nature of H_2 evolution.[82] Furthermore, N_2 reduction requires at least six protons, but the specific form of ammonia evolved (NH_3 versus NH_4^+) and the possibility of H_2 evolution may require up to 10 or more protons for N_2 reduction.

Scheme IV-1. The standard model for biological nitrogen fixation

$$N_2 + 8\,H^+ + 8\,e^- + 16\,ATP \longrightarrow 2\,NH_3 + H_2 + 16\,ADP + 16\,P_i$$

Nitrogenase consists of two proteins: the hetero-tetrameric molybdenum-iron (MoFe) protein and the homodimeric iron (Fe) protein. The Fe protein houses two ATP binding sites and

the [4Fe:4S] cubane cluster. The MoFe protein consists of two αβ dimers and contains three types of metal centers: (1) two [8Fe:7S] "P-clusters" at the α- and β-subunit interfaces, (2) two active sites, a [7Fe:9S:Mo:C:*R*-homocitrate] cluster called the FeMo-cofactor in the α-subunits, and (3) two mononuclear iron sites, named Fe16, between the β- and β'-subunits.[23] During substrate turnover, electrons flow from the [4Fe:4S] cluster to the P-cluster to the FeMo-cofactor, at which most, if not all, substrate reduction occurs upon sufficient buildup of protons and electrons.[30] Considering only the inorganic components, the FeMo-cofactor adopts near C_{3v} symmetry, with a central, trigonal prismatic core composed of three faces and three edges parallel to the C_3 axis that are made from six Fe atoms, numbered Fe2, 3, 4, 5, 6, and 7. A particular face may be identified by listing the four Fe atoms composing that face, such as Fe2,3,6,7. Similarly, an edge is identified by listing the two Fe atoms composing that edge. Examined in isolation, the edges and faces would be indistinguishable from each other, but variation in neighboring protein residues creates nonequivalent environments around the FeMo-cofactor inside the protein. The edges and faces may therefore each have different mechanistic roles. For example, CO has been shown to bind in a bridging fashion to Fe2,6.[37]

In this work, five possible pathways for substrate access from the protein surface to the FeMo-cofactor or P-cluster are experimentally identified. To accomplish this, *Azotobacter vinelandii* (Av) and *Clostridium pasteurianum* (Cp) MoFe protein crystals were pressurized with xenon (Xe) gas. Additionally, all published structures of the MoFe protein, including those from *Klebsiella pneumoniae* (Kp), were analyzed for the presence of nonwater, small molecules bound to the protein interior. The native Av, Cp, and Kp MoFe proteins are called Av1, Cp1, and Kp1, respectively, and the corresponding Xe-pressurized proteins are referred to as Av1-Xe and Cp1-Xe. Kp1 and Av1 are structurally similar (73% sequence identity), while their comparison to

Cp1 shows an insertion and a deletion, each ~50 residues, as well as primary structure differences (36% sequence identity between Cp1 and Av1).

Although Xe is monatomic, unlike nitrogenase substrates, several advantages exist for using Xe as a model for nitrogenase substrates such as N_2: (1) Xe and N_2 are neutral, polarizable, water-soluble, and unable to form hydrogen bonds; (2) the atomic radius of Xe (1.08 Å) is comparable to the NN bond distance (1.10 Å), so sterically, N_2 may travel similar pathways as those penetrable by Xe; and (3) Xe is easily detected by X-ray crystallography due to its high electron density and strong anomalous scattering. Furthermore, Xe pressurization is well established as a tool for probing gaseous substrate pathways in several biological molecules, including myoglobin,[83–86] copper amine oxidase,[87,88] laccase,[89] methane monooxygenase hydroxylase,[90,91] cytochrome c oxidase,[92] cytochrome ba3 oxidase,[93,94] acetyl-CoA synthase/carbon monoxide dehydrogenase,[95] and antibodies that oxidize water.[96] Finally, several molecular dynamic simulations support using Xe binding sites as markers for substrate pathways.[97–101]

X-ray crystallography, ^{129}Xe-NMR, and computational work on previous Xe studies indicate that Xe typically induces minimal distortion in the protein structure and occupies existing cavities in a protein either by displacing water molecules or filling otherwise empty pockets.[83–85,87–90,92,95,96] Reflecting the inertness of Xe, it tends to bind to the protein using mostly noncovalent, weak van der Waals forces with limited polarization interactions.[97] The Xe binding sites are usually hydrophobic, and the closest contacts (3.5-6.0 Å) are typically aliphatic and aromatic side-chains but can be polar groups.[79–81,83–86,88,91,92] These studies also show that pathways tend to travel parallel to secondary structure elements rather than through them.[97]

Previous studies on nitrogenase have identified four possible pathways to the active site. First, a water channel extending from the protein surface to the FeMo-cofactor, called the

interstitial channel, was identified from structural analysis of Av1, Cp1, and Kp1, and has been expected to facilitate access to the active site for protons and possibly larger substrates.[102–106] Use of the water-filled interstitial channel as an access pathway to the active site does not preclude the existence of other pathways since nonpolar substrates may prefer a less polar route. Second, Seefeldt and co-workers used the program CAVENV from the CCP4 suite with a probe radius of 2.5 Å to identify a hydrophobic substrate pathway, herein called the IS pathway (for authors Igarashi and Seefeldt).[29] The third previously proposed pathway, called the NH_3 egress pathway by its authors, was identified by a computational cavity analysis of Av1 and Kp1; it extends through the protein scaffold from the β-subunit surface to the FeMo-cofactor.[102] Fourth, molecular dynamic calculations yielded a possible substrate pathway that traces the shortest path from the protein surface to the FeMo-cofactor.[101] Like the NH_3 egress and IS pathways, it does not utilize any water channels but rather tunnels through the protein scaffold. Since nitrogenase has a relatively leisurely turnover rate of about 1 N_2/sec per active site, migration through the protein scaffold in the absence of permanent pathways should not be rate limiting, by analogy to O_2 binding to the buried heme of myoglobin and hemoglobin.[19,107] Compared to the four previously proposed pathways, only the interstitial water channel and the IS pathway coincide with pathways identified in the present work, suggesting that there are multiple potential pathways connecting the surface to the active site of nitrogenase.

IV.C. Results and discussion

Xe sites were determined from the X-ray crystal structures of one Av1 and two Cp1 Xe-pressurized crystals (Figure IV-1 and Table IV-1). Similar to previous Xe binding studies,[79–81,83–86,88,91,92] the Xe atoms in the MoFe proteins displace water or other small molecules or fill empty pockets, rather than displacing residue side chains (Table VI-2). The three Xe sites in Av1-Xe are

conserved in both crystallographically independent αβ dimers of the protein, as are the three Xe binding sites in Cp1-Xe. The root-mean-square deviations (RMSD) between the Xe-pressurized protein structures compared to their native structures is ∼0.20 Å. The RMSD of only the Xe binding pockets is between 0.15 and 0.23 Å, indicating little protein distortion from Xe incorporation (Table IV-S1).

Table IV-1. X-ray crystallographic data collection and refinement statistics for Av1 and Cp1

	Av1-Xe (4WNA)	**Cp1-Xe (4WN9)**
Data Collection Statistics		
Resolution range (Å)	39.62 – 2.00 (2.00-2.11)	39.75 – 1.90 (1.90-1.93)
Wavelength (Å)	1.5418	0.9537
Space group	$P2_1$	$P1$
Unit cell constants	a = 77.12 Å α = 90° b = 129.8 Å β = 108.9° c = 107.5 Å γ = 90°	a = 67.31 Å α = 73.47° b = 73.45 Å β = 87.56° c = 108.7 Å γ = 83.98°
Unique reflections	133,045 (6,500) *	137,866 (6,709) *
Completeness (%)	98.6 (98.6) *	96.7 (86.1) *
Redundancy	3.5 (3.3) *	3.9 (3.7) *
$I/\sigma(I)$	9.2 (3.2) *	16.0 (2.0) *
R_{merge}	0.079 (0.301) *	0.055 (0.660) *
Refinement Statistics		
Protein residues	1998/2054	1951/1984
Mean *B* value (Å^2)	23.0	31.0
R_{work}	0.177	0.192
R_{free}	0.225	0.246
Ramachandran outliers	10 (0.51%)	14 (0.73%)
RMSD bond lengths (Å)	0.009	0.008
RMSD bond angles (°)	1.29	1.21

*Numbers in parentheses represent data in the highest resolution shell.

Xe binding sites in Av1-Xe (PDB ID 4WNA) and Cp1-Xe (PDB ID 4WN9), as well as the PRL site in the Cp1-Xe structure, were analyzed for potential access routes to the FeMo-cofactor. Access to the P-cluster was also explored. In addition, imidazole (IMD), ethylene glycol (EDO), carbon monoxide (CO), and sulfur (S) binding sites from previously reported Av1, Cp1, and Kp1

structures were examined (Table IV-S2).[22,37,64,66] For all these small molecule binding sites (with the exception of the S site that may be derived from a cofactor sulfur[37]), we can conclude with certainty that a route from the protein surface to the binding site exists; however, routes from the binding sites to the cofactors are inherently less certain, and indeed, may not exist. Potential pathways were generated using the program CAVER. For reference, Figure IV-2 and the movie provided in the Supporting Information (section IV.I.) summarize all of the known water channels and proposed substrate pathways from this study and previously published studies. Residues involved in all pathways are provided in Table IV-S3, and close contacts for each small molecule binding site are provided in Tables IV-S6 – IV-S13.

Table IV-2. Properties of small molecule binding sites in Av1 and Cp1 Xe-pressurized crystals

Crystal	Site	Displaced species in native protein	*B*-factor (Å^2)	Occupancy (%)	Peak heights in anomalous Fourier map (σ)	Distance* to (Å)	
						FeMo-cofactor	surface
Av1-Xe (4WNA)	Xe1	Fills empty pocket	27.30	77	18.51	11.2	10.9
			28.16	81	20.23		
	Xe2	HOH	31.29	49	8.48	15.1	0
			31.40	46	7.12		
	Xe3	Imidazole	31.48	62	9.62	33.1	0
			29.51	54	10.84		
Cp1-Xe (4WN9)	Xe1	HOH	41.89	28	5.20	28.1	0
			39.52	34	5.78		
	Xe2	MPD	64.15	46	5.93	23.8	0
			44.89	50	7.64		
	Xe3	HOH	29.20	22	4.18	13.6	7.4
			33.79	19	5.53		
	PRL	HOH	20.65	100	N/A	24.5	7.7
			23.76	100	N/A		

* Distances were measured from Xe or PRL to the closest non-solvent atom in the FeMo-cofactor or at the protein surface.

IV.D.i. AI/IS and AII pathways: FeMo-cofactor access based on Xe binding sites

Xe binding sites in Cp1 and Av1 were examined to identify potential substrate pathways to the active site. We focused on buried Xe, as these sites have already penetrated into the protein interior. We further focused on Xe sites found in the α-subunit, since these are closer to the FeMo-cofactor compared to the two Xe residing in the β-subunit: Xe in the α-subunit (Av1-Xe1, Av1-Xe2, Cp1-Xe2) are 15, 13, and 23 Å away from the closest Fe atom in the FeMo-cofactor, respectively, while Xe in the β-subunit (Cp1-Xe1 and Av1-Xe3) are 28 and 33 Å away, respectively. These constraints narrow the relevant Xe sites to Av1-Xe1 and Cp1-Xe3.

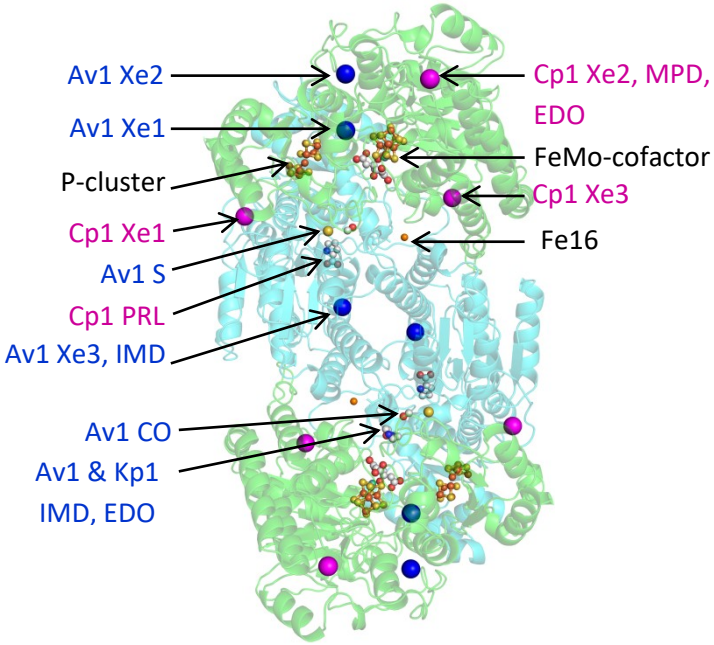


Figure IV-1. Ribbon representation of Cp1 showing locations for all Xe, PRL, and other small molecule binding sites. The α- and β-subunits are colored in green and cyan, respectively. The FeMo-cofactor, P-cluster, Fe16, and small molecules are displayed as small spheres colored by element. Av1 and Cp1 Xe sites are shown in large blue and magenta spheres, respectively. Binding sites observed in Av1 and Kp1 structures are superposed onto the Cp1 structure. (Cp1-Xe PDB ID: 4WN9; Av1-Xe PDB ID: 4WNA).

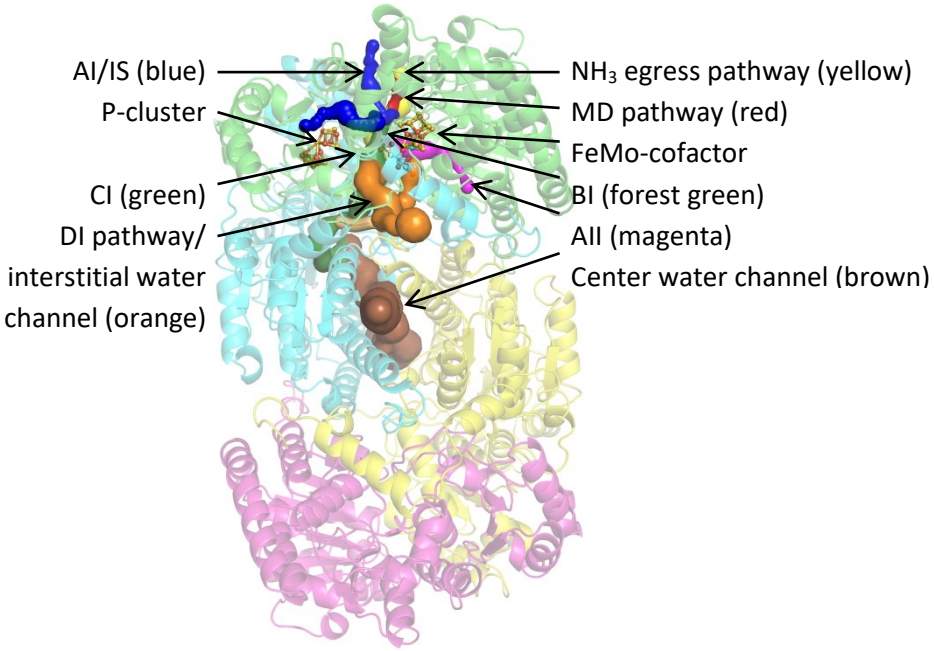
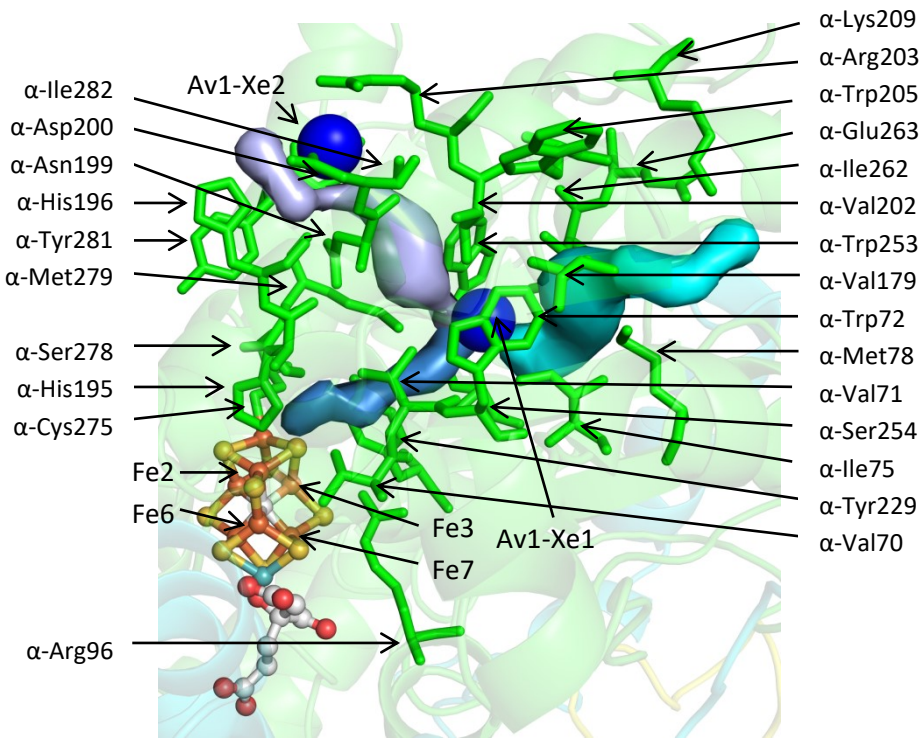


Figure IV-2. Ribbon representation of Av1 illustrating the channels and pathways discussed in this study. The α-subunits are colored green and magenta, and the β-subunits are colored cyan and yellow. The cofactors are shown in ball-and-stick representation colored by element. Pathways were calculated using CAVER and are displayed as surfaces within the protein structure.[108] This figure was created in PyMOL.[63]

We propose two substrate pathways based on Av1-Xe1 and Cp1-Xe3, called AI and AII, respectively (Figure IV-3). In both pathways, two routes were calculated: one from the protein surface to the binding site and another from the binding site to the FeMo-cofactor. The AI pathway, as calculated by CAVER, may include the surface site Av1-Xe2 as the point of substrate penetration through the protein surface; it is in close proximity (11 Å) to Av1-Xe1. A portion of the AI pathway is conserved in the previously published pathway by Seefeldt and co-workers (herein called the IS pathway) using the program CAVENV.[29] The IS pathway differs from the AI pathway at the point that it breaches the protein surface. The program utilized in this paper, CAVER, also identified the breaching point of the IS pathway as more favorable than the breaching point of the AI pathway: the average bottleneck radius of the IS and AI pathways from the Av1-Xe1 atom to the protein surface are 0.83 and 0.48 Å, respectively, and the lengths of

the pathways are 18.5 and 21.2 Å, respectively. In fact, the IS pathway is the most favored pathway as calculated by CAVER. However, binding of the Av1-Xe2 suggests that there may be multiple entry routes for the AI/IS pathway. Therefore, we present both entry/exit points as possible substrate pathways. Notably, the AI/IS pathway is predominantly hydrophobic, with the exception of residues at the protein surface and a couple around the water surrounding the FeMo-cofactor. These features support Seefeldt's postulation that this pathway is likely used by nonpolar substrates and/or reduction products. The AI/IS and AII pathways provide access to two of the three faces of the FeMo-cofactor, namely, the Fe2,3,6,7 and Fe3,4,5,7 faces. Since Xe is nonpolar and the interior Xe sites do not overlap with any polar species from other MoFe crystal structures, the AI and AII pathways may be primarily used by nonpolar substrates and/or reaction products.

(a)



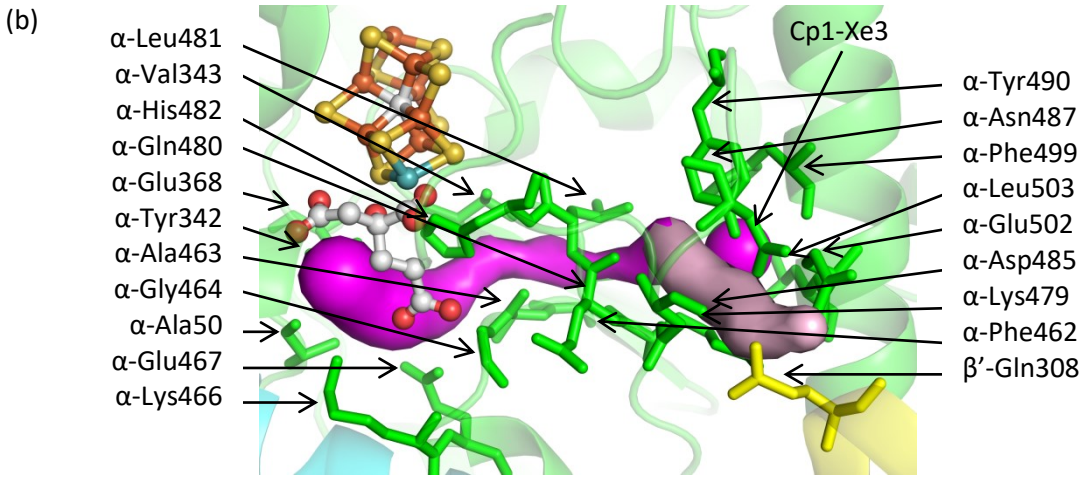
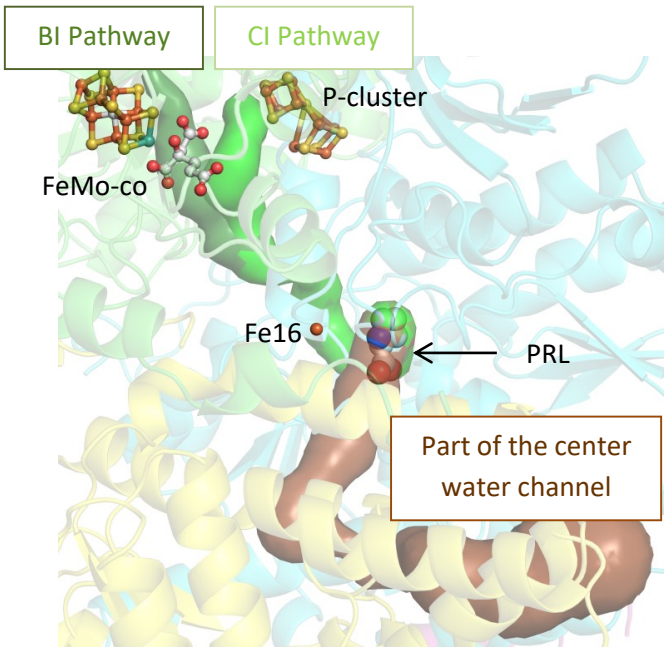


Figure IV-3. (a) The AI/IS pathway. From the protein surface to the Xe binding site, the AI and IS pathways follow the light purple and cyan pathways, respectively. From the Xe binding site to the FeMo-cofactor, the pathways (slate blue) are the same. Substrates may penetrate the protein surface at the Av1-Xe2 binding site following the light blue pathway or as illustrated by the cyan pathway. Upon reaching the Av1-Xe1 binding site, substrates may continue toward the FeMo-cofactor following the slate blue pathway. The Fe atoms accessed on the FeMo-cofactor by this pathway are labelled. Av1-Xe1 and Av1-Xe2 are displayed as large blue spheres. (b) The AII pathway (magenta and light pink surfaces). Substrates may reach the Cp1-Xe3 binding site following the light pink pathway and then continue toward the FeMo-cofactor following the magenta pathway. Cp1-Xe3 is displayed as a large magenta sphere. In both figures, residues lining the pathways are labelled. The α, β, and β' subunits are shown in green, cyan, and yellow, respectively.

IV.D.ii. BI pathway: FeMo-cofactor access based on PRL binding sites

The nearest neighbors of PRL in Cp1 are five aromatic residues, together with β-Glu323 and β-Lys424 (Figure IV-S2 and Table IV-S12). PRL resides in an arm of the center channel that curves toward the cofactors (Figure IV-4). The arm terminates before reaching the FeMo-cofactor; thus, substrates would need to continue through the protein scaffold to reach the active site. With CAVER, we deduced a possible substrate pathway from the PRL binding site to the FeMo-cofactor (pathway BI), which accesses the FeMo-cofactor at the Fe2,3,6,7 face.

(a)



(Figure continues on next page)

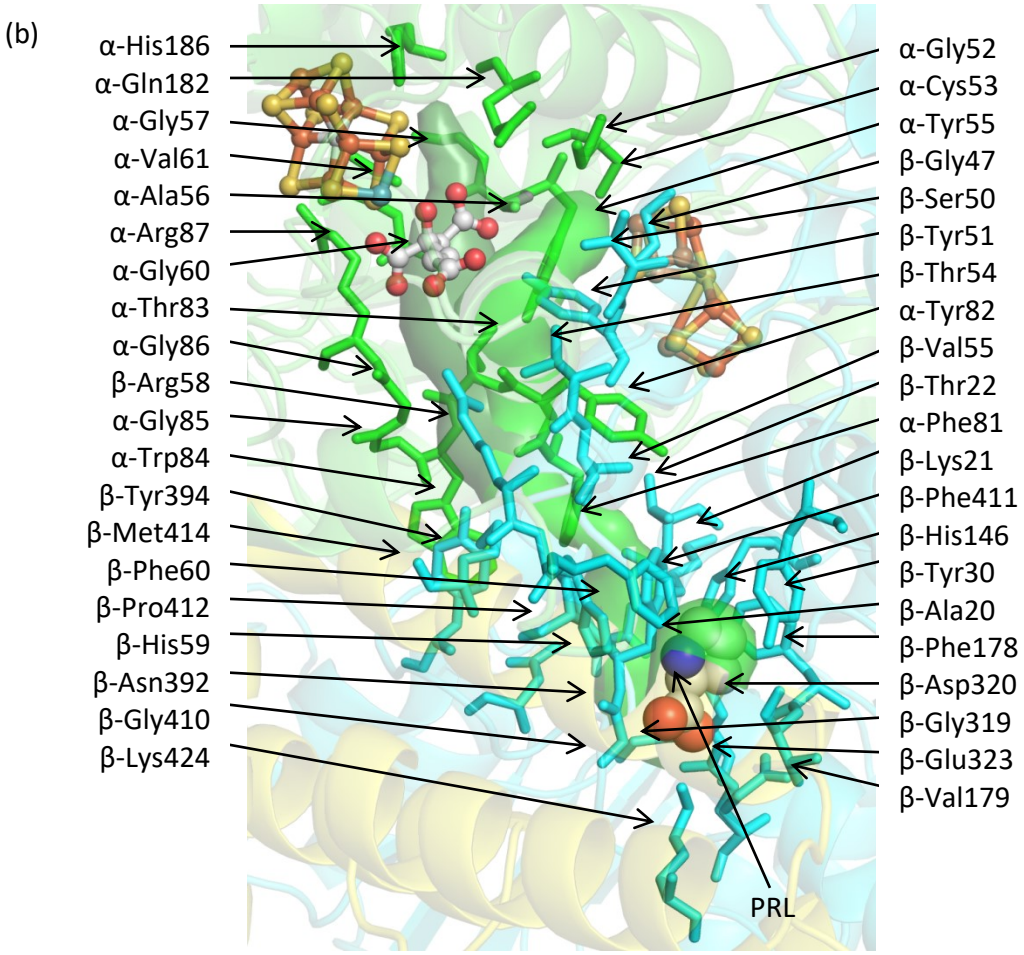


Figure IV-4. (a) PRL binds in an arm of the center channel (brown surface) that reaches toward the cofactors. The arm terminates before reaching the FeMo-cofactor, so substrates must continue to the cofactors within the protein scaffold. The most likely pathway (in terms of size) is between the α- and β-subunits. Branching from this pathway, substrates may either head toward the FeMo-cofactor (forest green surface, pathway BI) or toward the P-cluster (green surface, pathway CI). (b) A close-up view of the BI and CI pathways leading to the FeMo-cofactor and P-cluster, respectively. In both figures, the α, β, and β' subunits are shown in green, cyan, and yellow, respectively. Residues lining the pathway are shown in sticks and labelled. The cofactors and PRL are displayed as spheres colored by element.

IV.D.iii. CI pathway: P-cluster access based on PRL binding sites

Although generally considered as functioning in electron transfer between the Fe protein and FeMo-cofactor, ligand access to and from the P-cluster may be necessary since there is likely elimination of a sulfur atom during P-cluster biosynthesis.[109] Also, the redox

properties of the P-cluster indicate that it could potentially reduce protons and perhaps other substrates.[82,110] A possible substrate pathway, CI, from the PRL site to the P-cluster (Figure IV-4) was calculated by CAVER. The PRL is 16 Å away from the P-cluster and 24 Å away from the FeMo-cofactor; however, the closest metal center to the PRL binding site is Fe16, at a distance of 14 Å. Although the identity of this third metal site has been confirmed, its function is not currently known.[23]

Because the volume of the center channel is in excess of 1500 $Å^3$, the walls of the center channel are essentially an extension of the protein surface. As such, the MoFe protein resembles an oblong donut, in which the center channel is the donut hole.[21] Water molecules, nonpolar atoms (Xe), and polar molecules (IMD, EDO, MPD) are all found on the protein surface so it is not unexpected that these species have binding sites in the center channel as well. Therefore, the BI and CI pathways may facilitate access to the FeMo-cofactor for all species.

IV.D.iv. DI pathway/interstitial water channel: FeMo-cofactor access based on IMD, EDO, CO, and S binding sites

All published structures of native Av1, Cp1, and Kp1 were investigated for additional non-water, small molecule binding sites. Those containing small molecules are listed in Table IV-S2. These guest molecules come from crystallization solutions, cryo-protectants, or pressurized gas and include imidazole (IMD), 2-methyl-2,4-pentanediol (MPD), Mg^{2+}, 1,2-ethanediol (EDO), and carbon monoxide (CO); the sulfur (S) may be derived from the FeMo-cofactor.[22,37,64,66]

We focused on small molecules bound to the protein interior, of which there are five from previously reported MoFe protein structures (Table IV-SB). Several of these bind in or near the interstitial channel (Figure IV-5), which has been previously proposed to function as a substrate access pathway, as deduced from the conserved water network in Av1, Cp1, and Kp1.

The exogenous small molecules observed to bind within this channel include IMD in Av1 (3U7Q) and EDO in Kp1 (1QGU) (Table IV-S4).[22,66] The CO and S in Av1 (4TKV) bind in a protrusion from the interstitial channel. This protrusion (purple surface in Figure IV-S5) extends through the β-subunit to the protein surface; however, it is narrower and longer than the DI pathway. Therefore, only the short protrusion from the DI pathway containing the CO and S atoms is shown. The IMD and EDO sites directly overlap and are 4.8 and 7.7 Å from the CO and S, respectively. The polarity of IMD, EDO, and CO indicates that this channel may be utilized as an access pathway for polar substrates in addition to water or protons. The DI pathway accesses both the Fe2,3,6,7 and Fe3,4,5,7 faces, which are also accessed by the AI and AII pathways.

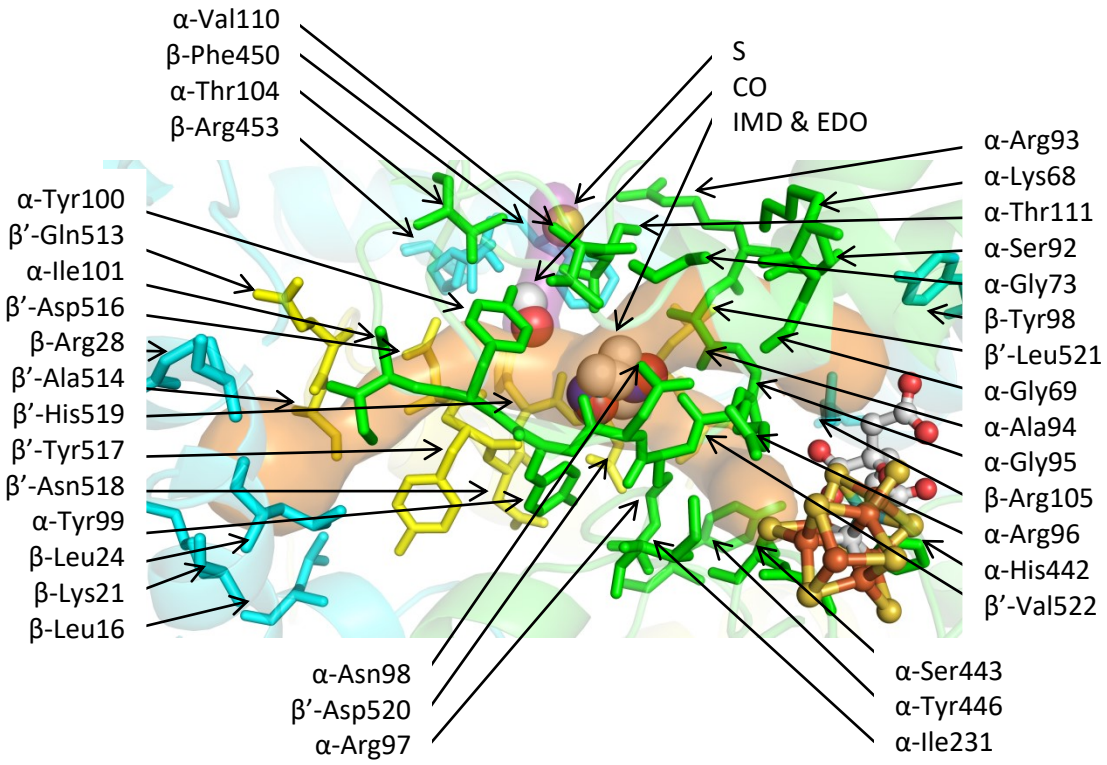


Figure IV-5. The DI pathway (orange surface). The EDO and IMD molecules bind in the DI pathway/interstitial channel that connects the protein surface to the Fe2,3,6,7 and Fe3,4,5,7 FeMo-cofactor faces (orange surface). The CO and S bind in a channel that extends from the DI pathway (purple surface). The α, β, and β' subunits are shown in green, cyan, and yellow, respectively. Residues lining the pathway are shown in sticks and labelled. The substrates and FeMo-cofactor are displayed as spheres colored by element.

IV.D.v. Pathway conservation

It is noteworthy that the Av1 and Cp1 Xe binding sites differ, given that Xe is used as an electron dense surrogate for crystallographic analysis of gas binding sites in enzymes. Furthermore, while a diverse set of small molecules have been found to bind to the protein interiors of Cp1, Av1, and Kp1, it is also the case that these binding sites are not identical between the structures. To assess whether the different pathways may be generally relevant to the functioning of nitrogenase or instead primarily reflect the behavior of specific MoFe proteins, the conservations of specific residue and residue type (hydrophobic or hydrophilic) in Av1 and Cp1 were evaluated for (1) all residues in the protein, (2) surface residues, and (3) nonsurface residues (Table IV-S5). This was compared to the conservation of specific residues and residue type for residues lining the substrate binding pockets and proposed pathways (Tables IV-SN – SU). The conservation of specific residue for all nonsurface residues compared to that of residues lining the substrate binding pockets and pathways is 38% and 63% respectively, and the conservation of residue type is 63% and 86%, respectively. This shows higher conservation of specific residue and residue type at the substrate pockets and in the proposed pathways, indicating that the AI, AII, BI, and CI pathways may be accessible in Av1 and Cp1. The differences in Xe and another small molecule binding sites in these structures may reflect details of the surrounding residues that alter the thermodynamics of ligand binding, but not necessarily the dynamic accessibility.

IV.D.vi. Comparison to other forms of the MoFe protein

The access pathways were compared to two other forms of the MoFe protein. First, examination of the proposed access pathways in complexes of the MoFe and Fe proteins indicates that the docking of the Fe protein onto the MoFe protein does not block any of the

proposed substrate pathways (Figure IV-S3).[28,67] This suggests that binding of the Fe protein may not sterically interfere with substrate access between the protein interior and exterior. This observation is of interest since the Thorneley-Lowe kinetic model assumes that substrates and products can only bind/leave the free MoFe protein.[44] Of course, differences in internal structure or protein dynamics could alter the behavior of the MoFe-protein between free and complexed states. Second, an overlay of the FeMo-cofactor-deficient Av1 protein structure shows that the AII pathway partially overlaps with the channel utilized by the FeMo-cofactor to access its binding pocket (Figure IV-S4).[111] This suggests that the funnel between the α-subunit domains mediating the transfer of the FeMo-cofactor into the active site region of the FeMo-cofactor-less protein may have multiple roles.

IV.D. Conclusion

On the basis of the Xe binding sites, we have identified in Av1 and Cp1, together with small molecule binding sites observed in Av1, Cp1, and Kp1, three new substrate and/or product pathways that can potentially connect the protein surface and the nitrogenase metalloclusters. The AI and AII pathways, deduced from Xe binding sites, are possible pathways for nonpolar substrates. Notably, the AI pathway is mostly conserved in the previously published pathway based on computational analysis by Seefeldt and co-workers. From the PRL binding site, there is a possible pathway to both the FeMocofactor (pathway BI) and the P-cluster (CI), the latter of which may provide a pathway for proton access. Both pathways contain part of the center water channel and then extend into the protein scaffold toward the metalloclusters. Given the polarity of the small molecules and the binding pocket environment, the BI and CI pathways may facilitate metallocluster access for both polar and nonpolar substrates/ products. IMD, EDO, CO, and S sites in Av1 and Kp1 suggest that the DI pathway/interstitial channel may be used as a

polar substrate pathway; it is conserved in all MoFe proteins. All pathways access the Fe2,3,6,7 and/or Fe3,4,5,7 faces of the FeMo-cofactor; however, this does not necessarily indicate that these faces are the primary targets for substrate binding since substrates may be able to move around the FeMo-cofactor. Overall, our studies establish that a variety of small molecules can access the interior of the MoFe-protein through multiple pathways (see the movie in the Supporting Information). This is based on experimental identification of nonwater, small molecule binding sites in the interior of Av1 and Cp1, which are two of the most structurally divergent bacterial MoFe proteins known. While there may be more favored pathways, given the variety of potential routes available, these observations indicate that there is unlikely to be a unique pathway utilized for substrate access from the protein surface to the active site; in effect, this is a molecular-level example of "all roads lead to Rome".

IV.E. Experimental

IV.C.i. Cell growth and protein purification

Av1 and Cp1 protein were obtained using cell growth and protein purification procedures previously described.[64,73]

IV.C.ii. Crystallization

Crystals were grown in 24-well plates using the sitting-drop method at room temperature in an anaerobic chamber with an atmosphere of ~95% argon and ~5% hydrogen. All crystallization solutions were purged with argon prior to use. Av1 crystals were obtained as described previously.[73] For Cp1, the reservoir and crystallization solutions consisted of double-distilled water, 23% polyethylene glycol (MW 3350 g/mol, Hampton Research), 0.2 M lithium

citrate (Aldrich), and 5 mM sodium dithionite (J.T. Baker). Several crystals of hexagonal and block morphology formed after 2 days, but only the block crystals diffracted well.

IV.C.iii. Xenon pressurization (expanded from published article)

Crystals were grown inside an anaerobic chamber as described above. Working with one well at a time, the tape covering the well was cut and removed. 100 µL of the reservoir solution was transferred to an Eppendorf tube. A thin film on top of the crystallization drop was removed with a needle. 2-3 drops of the Fomblin Y mineral oil (cryo-protectant, Sigma-Aldrich) was added to the top of the crystallization drop. The well was re-sealed with tape, and the tray was removed from the tent. A gas pressurization device was made by Michael Stowell.[112] On the pressurization device (Figure IV-6), a small wad of Kim wipe was placed at the bottom of the pressurization chamber and was soaked with a couple drops of the reservoir solution from the Eppendorf tube. This is necessary to maintain a relatively constant vapor pressure during the pressurization process.

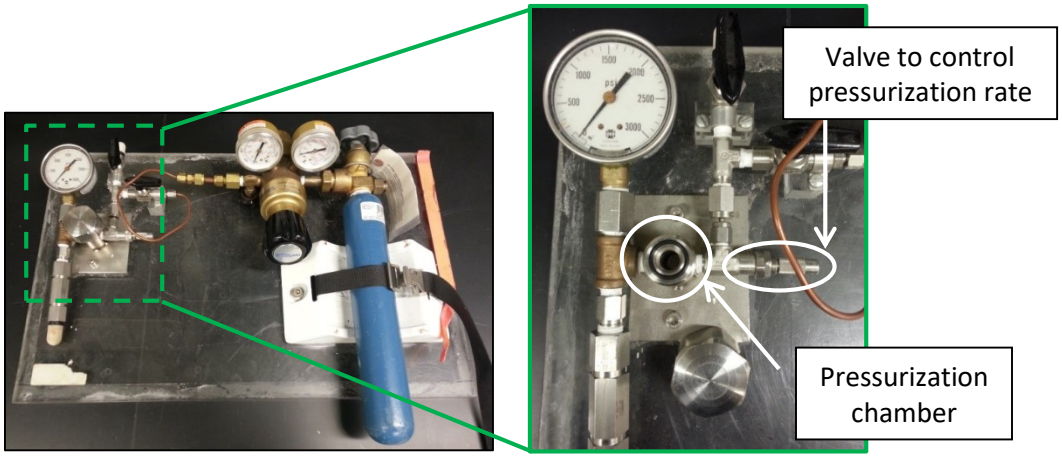


Figure IV-6. Homemade gas pressurization device.[112]

The sealing tape covering the well was cut and removed. Using a microscope, the crystal was collected on a nylon loop. The crystal was slowly lifted through the cryo-protectant layer of the drop and then very quickly moved to the pressurization chamber of the pressurization device. While lightly venting Xe (Matheson) through the device, the cap to the pressurization chamber was tightened. This process must be done carefully to ensure that all oxygen is forced out of the device without increasing the pressure too quickly, which causes the crystal to crack or become lost from the loop. After fully closing the device to the atmosphere, the pressure was increased to 14 atm over approximately six seconds. The device stood undisturbed for 10-15 minutes after which the pressure in the chamber was slowly released. Here also the pressure change must be done slowly to avoid losing or cracking the crystal. Once the pressure was equilibrated to atmospheric pressure, the crystal was very quickly transferred to liquid nitrogen for storage.

IV.C.iv. Data collection and refinement

Diffraction data for Cp1-Xe was collected remotely from the Stanford Synchrotron Radiation Lightsource (SSRL) on beamline 12-2 with a DECTRIS Pilatus 6 M detector. Reference sets of 1440 diffraction images were collected at 12999.97 eV with an oscillation angle of 0.25° over 360° rotation. To confirm the identity of the Xe sites, diffraction data were also collected at 6690.11 eV using the same strategy. Although well above the Ledge, Xe exhibits significant anomalous scattering at this energy with $\Delta f'' \approx 10$ electrons. Diffraction data for Av1-Xe was collected in-house on a Rigaku MicroMax 007-HF X-ray generator with a Rigaku RAXIS-IV++ detector. All data sets were integrated with the XDS program package.[57] Scaling was carried out with the CCP4 suite,[58] and phasing was determined by molecular replacement against Av1 (PDB ID 3U7Q) and Cp1 (4WES).[22,64] Initial refinement was carried out with CNS,[59] and alternative

conformations and isotropic B-factors were refined with REFMAC5.[60,61] Simulated annealing was performed using PHENIX.[62]

IV.C.v. Determination of small molecule binding sites

The presence and occupancy of each Xe site were evaluated by examination of electron density maps, anomalous difference Fourier peaks, and the B-factor of Xe and the surrounding residues (Table IV-2). Electron density and anomalous difference maps are shown for each Xe binding site in Figure IV-S1.

Nonprotein electron density was evident in the Cp1-Xe structure that was modeled as a proline ligand (PRL, Figure IV-S2). We propose PRL for use in the model because it nicely fits the observed electron density; however, the actual identity and origin of this species are not conclusively known. It is unlikely that the electron density represents ill-defined water molecules because neighboring water molecules are well-defined, and 6−7 water molecules would be needed to accurately model the observed electron and difference density, which is more than the space can accommodate. Furthermore, the ring of PRL could favorably interact with the five neighboring aromatic side chains in this binding pocket.[113,114] Finally, the acid group could interact with β-Lys424 and β-Glu323 via hydrogen bonding. While we will refer to the species throughout the manuscript as PRL (to distinguish it from proline in the peptide chain), we cannot unambiguously identify the species at this site. As it is clearly a nonwater ligand, however, we include it in the present analysis.

IV.C.vi. Pathway calculations and display

Pathways were calculated using the software CAVER.[108] Coordinates of the small molecules were provided as the starting point for pathway calculations. CAVER calculates

pathways from the grid point closest to the provided coordinates, so some starting points are slightly offset from the Xe atoms. The probe radius, shell radius, and shell depth were set to 0.5, 4, and 5 Å, respectively. For each small molecule, two pathways were selected: one from the protein surface to the small molecule binding site and the other from the binding site to a cofactor. For any given starting point, many pathways exist; however, the most probable pathways are those with the shortest length and largest width, and are prioritized by CAVER. Pathways are displayed throughout the manuscript as surfaces generated in PyMOL.[63]

IV.F. Accession codes

Accession Codes The structural model and structure factors have been deposited with the Protein Data Bank. The PDB ID for the Av1-Xe and Cp1-Xe structures are 4WNA and 4WN9, respectively.

IV.G. Acknowledgments

We acknowledge the Gordon and Betty Moore Foundation, the Beckman Institute, and the Sanofi-Aventis Bioengineering Research Program 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 thank Jens Kaiser and James Howard for helpful discussions as well as Thomas Spatzal for discussions and his assistance with the xenon pressurization.

IV.H. Abbreviations

Cp, Clostridium pasteurianum; Av, Azotobacter vinelandii; Kp, Klebsiella pneumonia; IMD, imidazole; EDO, 1,2-ethanediol; MPD, 2-methyl-2,4-pentanediol; PRL, proline (ligand)

IV.I. Supporting information

IV.I.i. Movie

A movie is accessible, free of charge, at http://pubs.acs.org/doi/suppl/10.1021/bi501313k. In this movie, the MoFe protein is rotated around the vertical axis to show the location of all pathways discussed in the manuscript, including newly identified and previously published pathways (blue: AI/IS pathway; magenta: AII pathway; forest green: BI pathway; green: CI pathway; orange: DI pathway/interstitial water channel; brown: center water channel; yellow: ammonia egress pathway; red: molecular dynamics pathway). The α subunits of the protein are shown as green and magenta ribbon, and the β subunits are shown in cyan and yellow ribbon. One FeMo-cofactor and P-cluster are shown as ball and sticks colored by element (yellow: sulfur; orange: iron; cyan: molybdenum; red: oxygen; grey: carbon). The two black Fe atoms in the FeMo-cofactor are Fe2 and Fe6, which bind the reversible inhibitor, carbon monoxide, in a bridging fashion. The three large black spheres are Xe binding sites in Av1 and Cp1. For clarity, the pathways, cofactors, and Xe binding sites are only displayed in one αβ-subunit of the protein. The protein scaffold is that of Av1 (PDB ID 4WNA) onto which the Cp1 pathways and Xe binding site have been superposed. The movie was made in PyMOL.

IV.I.ii. Tables

Table IV-S1. RMSD of MoFe protein Xe binding pockets compared to native structures

	Av1-Xe structure compared to Av1 (1.0 Å, 3U7Q)	Cp1-Xe structure compared to Cp1 (1.08 Å, 4WES)
RMSD of whole protein (Å)	0.192	0.215
RMSD of Xe1 binding site (Å)*	0.153	0.210
RMSD of Xe2 binding site (Å)*	0.151	0.232
RMSD of Xe3 binding site (Å)*	0.119	0.211

* Residues within 6 Å of the Xe site were used in the RMSD calculations.

Table IV-S2. Small molecules and ions in published MoFe protein structures

Protein (resolution)	PDB ID	Small molecule (total number observed)	Binding location
Av1 (1.0 Å)	3U7Q[22]	IMD (4)	Surface
		Mg^{2+} (2)	Surface
		IMD	Interior
Av1 (1.5 Å)	4TKV[37]	CO (3)	Surface
			Interior (2)
		S	Interior
Cp1 (1.08 Å)	4WES[64]	MPD	Surface
Kp1 (1.60 Å)	1QGU, 1QH8[66]	Mg^{2+} (5)	Surface
		EDO (14)	Surface
		EDO	Surface
		EDO	Interior

Table IV-S3. Residues involved in proposed pathways from previous and current studies

Pathway target	Pathway	Residues lining pathway (Av1 numbering)	
FeMo-cofactor	AI	α	Val70, Val71, Trp72, Ile75, His195, His196, Asn199, Asp200, Val202, Arg203, Tyr229, Trp253, Ser254, Cys275, Ser278, Met279, Tyr281, Ile282
	AII	α	Ile59, Tyr354, Ile355, Glu380, Gly422, Ser423, Gly424, Lys426, Glu427, Arg439, Glu440, Met441, His442, Asp445, Ser447, Tyr450, Phe459, Asp462, Met463
		β'	Thr360
	BI	α	Ala65, Gly66, Gly69, Val70, Gln90, Tyr91, Ser92, Arg93, Ala94, Gly95, Arg96, Gln191, His195
		β	Ala67, Lys68, Ala69, Leu67, Tyr102, Arg105, His106, Phe107, His193, Phe230, Thr232, Gly368, Asp369, Phe372, Asn445, Tyr447, Gly470, Phe471, Pro472, Phe474, Thr484
P-cluster	CI	α	Gly61, Cys62, Tyr64, Ala65, Gln90, Tyr91, Ser92
		β	Ala67, Lys68, Ala69, Leu77, Gly94, Ala97, Tyr98, Ser100, Tyr102, Arg105, His106, Phe107, His193, Phe230, Glu231, Gly368, Asp369, Phe372, Asn445, Tyr447, Gly470, Phe471, Pro472, Phe474, Thr484
FeMo-cofactor	Interstitial water channel[104]/DI	α	Lys68, Gly69, Gly73, Ser92, Arg93, Ala94, Gly95, Arg96, Arg97, Asn98, Tyr99, Tyr100, Ile101, Thr104, Val110, Thr111, Ile231, His442, Ser443, Tyr446
		β	Leu16, Lys21, Leu24, Arg28, Tyr98, Arg105, Phe450, Arg453
		β'	Gln513, Ala514, Asp516, Tyr517, Asn518, His519, Asp520, Leu521, Val522
	IS	α	Val70, Val71, Trp72, Ile75, Met78, Arg96, Val179, His195, Val202, Trp205, Lys209, Tyr229, Trp253, Ile262, Glu263, Met279
	NH_3 egress[102]	α	Gln53, Met57, Thr58, Ile59, Arg60, Ala65, Gln191, Glu380, Asp403, Lys426
		β	Gln93, Gly94, Ser115, Ser117
	MD simulation[101]	α	Asn49, Gly66, Val70, Ser190, Gln191, Ser192, Leu193, His195, His196, Asn199, Arg277, Ser278, Met279, Asn280, Tyr281, Gly357, Phe381, Ala382, His383

Table IV-S4. IMD, CO, S, and EDO binding sites in Av1 and Kp1

Crystal (PDB ID)	Site	Displaced species in native protein	Occupancy (%)	Distance* to (Å)	
				FeMo-cofactor	surface
Av1 (3U7Q)	IMD	HOH	100	11.5	11.6
Av1 (4TKV)	CO	HOH	60%	17.0	8.7
Av1 (4TKV)	S	Empty pocket	100%	19.0	10.9
Kp1 (1QGU)	EDO	HOH	100	11.6	10.1

* Distances were measured from the closest substrate atom to the closest metal in the FeMo-cofactor or to the closest non-solvent atom on the protein surface.

Table IV-S5. Conservation of specific residues and residue type in Av1 and Cp1

Residue conservation	Native protein			Xe/PRL binding pockets and AI, AII, BI, CI pathways
	All residues	Non-surface residues	Surface residues	
Conservation of specific residue	38.0%	38.4%	24.8%	62.5%
Conservation of residue type (hydrophobic or hydrophilic)	73.7%	63.4%	53.2%	85.6%

Table IV-S6. Close contacts and residue conservation for Av1-Xe1 (protein interior)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Cp1
1	α-Val71	CG1	3.88	α-Val60
2	α-Trp72	CE3	4.48	α-Met61
3	α-Ile75	CD1	3.29	α-Ile64
4	α-Ala198	O	5.95	α-Ala187
5	α-Val202	CG2	3.59	α-Val191
6	α-Trp253	CE3	4.34	α-Leu238
7	α-Ser254	CB	4.31	α-Thr239
8	α-Ile262	CD1	4.59	α-Val247
9	α-Met279	CE	4.49	α-Ile264
Hydrophobic residues: 89% Conservation of specific residue: 44% Conservation of residue type*: 100%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S7. Close contacts and residue conservation for Av1-Xe2 (protein surface)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Cp1
1	α-His196	CD2	5.20	α-His185
2	α-Asn199	O	3.98	α-Asn188
3	α-Asp200	OD1	3.08	α-Asn189
4	α-Arg203	NE	4.17	α-Met192
5	α-Tyr281	O	3.67	α-Tyr266
6	α-Ile282	CD1	3.87	α-Ile267
7	α-His285	CB	4.18	α-Met270
Hydrophobic residues: 29% Conservation of specific residue: 57% Conservation of residue type*: 71%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S8. Close contacts and residue conservation for Av1-Xe3 (protein surface)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Cp1
1	β-Tyr233	OH	5.24	β-Gly180
2	β-Arg468	O	5.61	β-Arg408
3	β-Ile469	CD1	4.36	β-Phe409
4	β-Ser482	CB	4.50	β-Asn422
5	β-Thr483	C	4.81	β-Pro423
6	β-Thr484	O	3.92	β-Lys424
7	β-Gly489	O	5.10	β-Gly429
8	β-Gln492	CB	4.19	β-Arg432
9	β-Ile493	CG1	3.83	β-Leu433
10	β-Thr496	OG1	4.44	β-Glu436
Hydrophobic residues: 40% Conservation of specific residue: 20% Conservation of residue type*: 90%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S9. Close contacts and residue conservation for Cp1-Xe1 (protein surface)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Av1
1	α-Asp107	OD2	5.84	α-Asp117
2	α-Lys120	CE	5.99	α-Lys130
3	α-Glu127	OE1	5.40	α-Glu137
4	β-Arg12	O	3.81	β-Arg59
5	β-Lys13	C	4.11	β-Glu60
6	β-Ala14	N	3.93	β-Ala61
7	β-Leu15	O	3.51	β-Leu62
8	β-Arg16	NH1	3.73	β-Thr63
9	β-Ile17	N	5.81	β-Val64
10	β-Glu371	O	5.84	β-Gly424
11	β-Asp373	OD2	5.09	β-Asp426
Hydrophobic residues: 27% Conservation of specific residue: 64% Conservation of residue type*: 91%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S10. Close contacts and residue conservation for Cp1-Xe2 (protein surface)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Av1
1	α-Asp254	O	3.57	α-Lys269
2	α-Leu255	CB	4.30	α-Leu270
3	α-Pro278	CG	5.41	α-Pro293
4	α-Cys300	O	3.38	α-Lys315
5	α-Phe301	CZ	3.84	α-Phe316
6	α-Val406	CG1	3.90	n/a
7	α-Ile407	CD1	4.18	n/a
Hydrophobic residues: 71% Conservation of specific residue: 43% Conservation of residue type*: 80%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S11. Close contacts and residue conservation for Cp1-Xe3 (protein interior)

	Residue	Atom	Distance from Xe (Å)	Corresponding residue in Av1
1	α-Phe460	CE1	4.63	α-Gly422
2	α-Lys477	CG	3.95	α-Arg439
3	α-Leu479	CD2	5.55	α-Met441
4	α-Asp483	OD2	5.94	α-Asp445
5	α-Tyr488	OH	3.61	α-Tyr450
6	α-Asn496	O	5.39	α-Ile458
7	α-Phe497	CD1	3.29	α-Phe459
8	α-Gly498	N	5.37	α-Ala460
9	α-Glu500	CB	3.83	α-Asp462
10	α-Leu501	N	3.71	α-Met463
Hydrophobic residues: 60% Conservation of specific residue: 30% Conservation of residue type*: 90%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S12. Close contacts for Cp1-PRL

	Residue	Atom	Distance from ligand (Å)	Corresponding residue in Av1
1	β-Tyr30	CZ	3.7	β-Leu77
2	β-His59	NE2	3.7	β-His106
3	β-Phe60	CZ	3.8	β-Phe107
4	β-Phe178	CE2	3.4	β-Phe230
5	β-Val179	CA	3.5	β-Glu231
6	β-Gly180	C3	3.4	β-Thr232
7	β-Asp183	OD2	4.1	β-Asn236
8	β-Glu323	OE2	3.0	β-Phe375
9	β-Gly410	O	3.8	β-Gly470
10	β-Phe411	C1	3.8	β-Phe471
11	β-Lys424	NZ	2.7	β-Thr424
Conservation of specific residue: 46% Conservation of residue type*: 73%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S13. Close contacts for Av1-IMD

	Residue	Atom	Distance from ligand (Å)	Corresponding residue in Cp1
1	α-Ala94	O	4.1	α-Gly85
2	α-Gly95	N	5.2	α-Gly86
3	α-Arg96	O	3.1	α-Arg87
4	α-Arg97	CA	3.9	α-Arg88
5	α-Asn98	ND2	3.9	α-Phe89
6	α-Tyr99	CE1	3.3	α-Lys90
7	α-Tyr100	CE2	4.2	α-Pro91
8	α-Val110	CG1	4.2	α-Asn102
9	α-Thr111	CG2	3.9	α-Glu103
10	β'-Asp516	O	5.7	β'-Glu451
11	β'-Tyr517	C5	4.1	β'-Glu452
12	β'-Asn518	N	6.0	β'-Asp453
13	β'-His519	C	5.4	β'-Phe454
14	β'-Asp520	OD1	3.7	β'-Glu455
15	β'-Leu521	N	4.9	β'-Val456
Conservation of specific residue: 20% Conservation of residue type*: 60%				

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S14. Residues lining the AI pathway in Av1 and Cp1

AI pathway in Av1	AI pathway in Cp1
α-Val70	α-Val59
α-Val71	α-Val60
α-Trp72	α-Met61
α-Ile75	α-Ile64
α-His195	α-His184
α-His196	α-His185
α-Asn199	α-Asn188
α-Asp200	α-Asn189
α-Val202	α-Val191
α-Arg203	α-Met192
α-Tyr229	α-Tyr214
α-Trp253	α-Leu238
α-Ser254	α-Thr239
α-Cys275	α-Cys260
α-Ser278	α-Ser263
α-Met279	α-Ile264
α-Tyr281	α-Tyr266
α-Ile282	α-Ile267
Conservation of specific residue: 67% Conservation of residue type*: 94%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S15. Residues lining the AII pathway in Cp1 and Av1

AII pathway in Av1	AII pathway in Cp1
α-Ile59	α-Ala48
α-Tyr354	α-Tyr340
α-Ile355	α-Val341
α-Glu380	α-Glu366
α-Gly422	α-Phe460
α-Ser423	α-Ala461
α-Gly424	α-Gly462
α-Lys426	α-Lys464
α-Glu427	α-Glu465
α-Arg439	α-Lys477
α-Glu440	α-Gln478
α-Met441	α-Leu479
α-His442	α-His480
α-Asp445	α-Asp483
α-Ser447	α-Asn485
α-Tyr450	α-Tyr488
α-Phe459	α-Phe497
α-Asp462	α-Glu500
α-Met463	α-Leu501
β'-Thr360	β'-Gln308
Conservation of specific residue: 45% Conservation of residue type*: 90%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S16. Residues lining the BI pathway in Cp1 and Av1

BI pathway in Av1	BI pathway in Cp1
α-Ala65	α-Ala54
α-Gly66	α-Gly55
α-Gly69	α-Gly58
α-Val70	α-Val59
α-Gln90	α-Phe79
α-Tyr91	α-Tyr80
α-Ser92	α-Thr81
α-Arg93	α-Trp82
α-Ala94	α-Gly83
α-Gly95	α-Gly84
α-Arg96	α-Arg85
α-Gln191	α-Gln180
α-His195	α-His184
β-Ala67	β-Ala20
β-Lys68	β-Lys21
β-Ala69	β-Thr22
β-Leu67	β-Tyr30
β-Tyr102	β-Val55
β-Arg105	β-Arg58
β-His106	β-His59
β-Phe107	β-Phe60
β-His193	β-His146
β-Phe230	β-Phe178
β-Thr232	β-Val179
β-Gly368	β-Gly319
β-Asp369	β-Asp320
β-Phe372	β-Glu323
β-Asn445	β-Asn392
β-Tyr447	β-Tyr394
β-Gly470	β-Gly410
β-Phe471	β-Phe411
β-Pro472	β-Pro412
β-Phe474	β-Met414
β-Thr484	β-Lys424
Conservation of specific residue: 68% Conservation of residue type*: 82%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S17. Residues lining the CI pathway in Cp1 and Av1

CI pathway in Av1	CI pathway in Cp1
α-Gly61	α-Gly50
α-Cys62	α-Cys51
α-Tyr64	α-Tyr53
α-Ala65	α-Ala54
α-Gln90	α-Phe79
α-Tyr91	α-Tyr80
α-Ser92	α-Thr82
β-Ala67	β-Ala20
β-Lys68	β-Lys21
β-Ala69	β-Thr22
β-Leu77	β-Tyr30
β-Gly94	β-Gly47
β-Ala97	β-Ser50
β-Tyr98	β-Tyr51
β-Ser100	β-Thr54
β-Tyr102	β-Val55
β-Arg105	β-Arg58
β-His106	β-His59
β-Phe107	β-Phe60
β-His193	β-His146
β-Phe230	β-Phe178
β-Glu231	β-Val179
β-Gly368	β-Gly319
β-Asp369	β-Asp320
β-Phe372	β-Glu323
β-Asn445	β-Asn392
β-Tyr447	β-Tyr394
β-Gly470	β-Gly410
β-Phe471	β-Phe411
β-Pro472	β-Pro412
β-Phe474	β-Met414
β-Thr484	β-Lys424
Conservation of specific residue: 66% Conservation of residue type*: 81%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S18. Residues lining the DI pathway/interstitial water channel in Cp1 and Av1

DI pathway in Av1	DI pathway in Cp1
α-Lys68	α-Lys57
α-Gly69	α-Gly58
α-Gly73	α-Gly62
α-Ser92	α-Thr81
α-Arg93	α-Trp82
α-Arg96	α-Arg85
α-Arg97	α-Arg86
α-Asn98	α-Ser89
α-Tyr99	α-Lys90
α-Tyr100	α-Pro91
α-Ile101	α-Glu92
α-Thr104	α-Thr95
α-Val110	α-Asn100
α-Thr111	α-Glu101
α-Ile231	α-Ile216
α-His442	α-His480
α-Ser443	α-Ser481
α-Tyr446	α-Tyr482
β-Leu16	n/a
β-Lys21	n/a
β-Leu24	n/a
β-Arg28	n/a
β-Tyr98	β-Tyr50
β-Arg105	β-Arg57
β-Phe450	β-Phe397
β-Arg453	β-Arg400
β'-Gln513	n/a
β'-Ala514	β'-Cys449
β'-Asp516	β'-Glu451
β'-Tyr517	β'-Glu452
β'-Asn518	β'-Asp453
β'-His519	β'-Phe454
β'-Asp520	β'-Glu455
β'-Leu521	β'-Val456
β'-Val522	β'-Val457
Conservation of specific residue: 50% Conservation of residue type*: 87%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S19. Residues lining the IS pathway in Av1 and Cp1

IS pathway in Av1	IS pathway in Cp1
α-Val70	α-Val59
α-Val71	α-Val60
α-Trp72	α-Met61
α-Ile75	α-Ile64
α-Met78	α-Met67
α-Arg96	α-Arg85
α-Val179	α-His168
α-His195	α-His184
α-Val202	α-Val191
α-Trp205	α-Glu194
α-Lys209	α-Lys198
α-Tyr229	α-Tyr214
α-Trp253	α-Leu238
α-Ile262	α-Val247
α-Glu263	α-Gln248
α-Met279	α-Ile264
Conservation of specific residue: 56% Conservation of residue type*: 94%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S20. Residues lining Dance's NH_3 egress pathway in Av1 and Cp1[102]

NH_3 egress pathway in Av1	NH_3 egress pathway in Cp1
α-Gln53	α-Val42
α-Met57	α-Ile46
α-Thr58	α-Thr47
α-Ile59	α-Ala48
α-Arg60	α-Arg49
α-Ala65	α-Ala54
α-Gln191	α-Gln180
α-Glu380	α-Glu366
α-Asp403	α-Asp441
α-Lys426	α-Lys464
β-Gln93	β-Gln46
β-Gly94	β-Gly47
β-Ser115	β-Thr68
β-Ser117	β-Ser70
Conservation of specific residue: 71% Conservation of residue type*: 93%	

***** Residues are categorized as hydrophilic or hydrophobic

Table IV-S21. Residues lining Smith's molecular dynamics pathway in Cp1 and Av1[101]

MD pathway in Av1	MD pathway in Cp1
α-Asn49	α-Asn38
α-Gly66	α-Gly55
α-Val70	α-Val59
α-Ser190	α-Ser179
α-Gln191	α-Gln180
α-Ser192	α-Ser181
α-Leu193	α-Ala182
α-His195	α-His184
α-His196	α-His185
α-Asn199	α-Asn188
α-Arg277	α-Arg262
α-Ser278	α-Ser263
α-Met279	α-Ile264
α-Asn280	α-Asn265
α-Tyr281	α-Tyr268
α-Gly357	α-Gly343
α-Phe381	α-Phe367
α-Ala382	α-Ala368
α-His383	α-His369
Conservation of specific residue: 90% Conservation of residue type*: 100%	

***** Residues are categorized as hydrophilic or hydrophobic

IV.I.iii. Figures

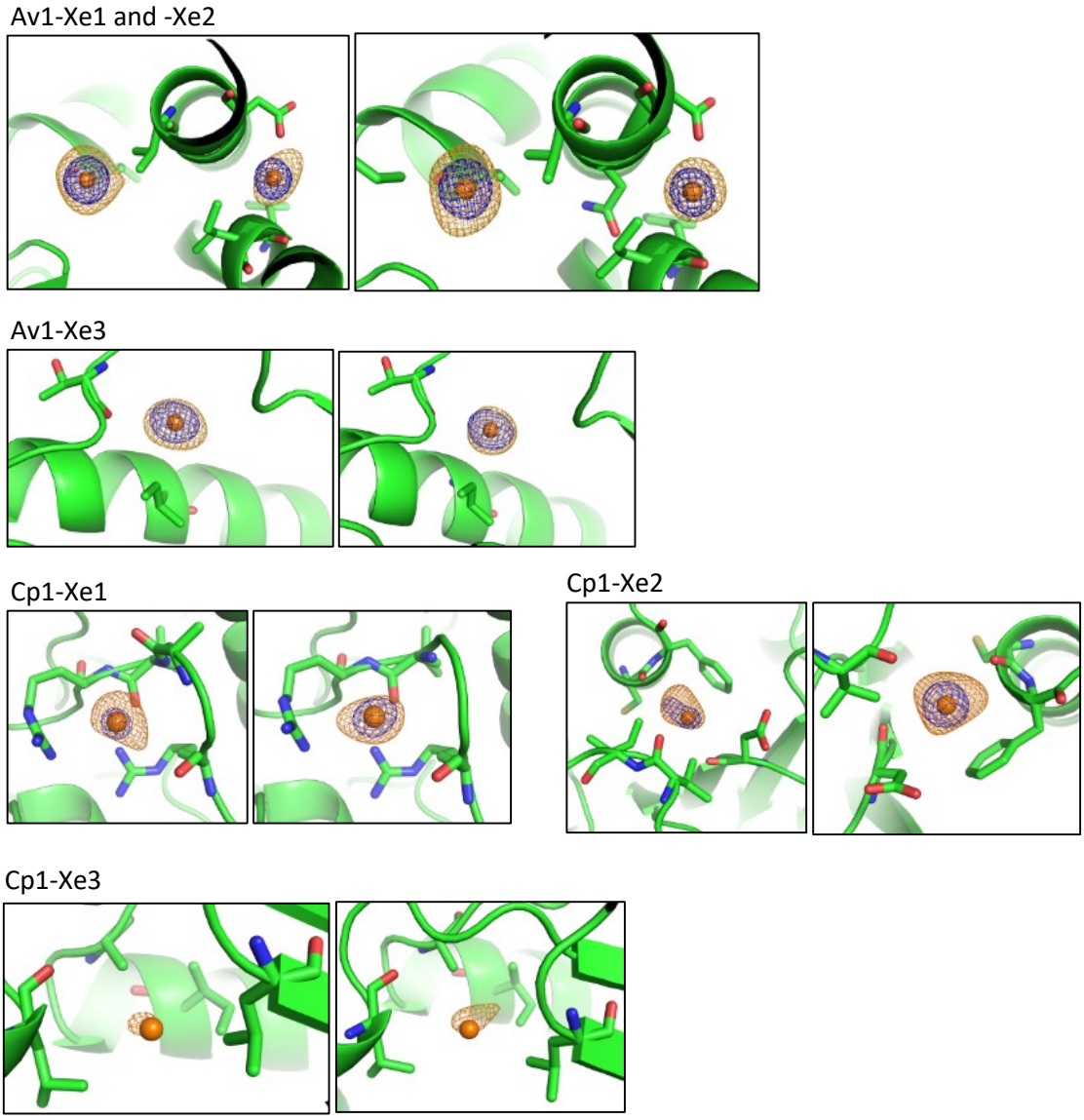


Figure IV-S1. Electron density maps (blue mesh) and anomalous difference maps (orange mesh) for each Xe binding site, including noncrystallographically-related pairs, and their surrounding environment. Both maps are contoured to 3 σ. Residues with at least one atom that is 4 Å or closer to the Xe atom are shown in sticks. Note that the electron density for Cp1-Xe3 is not visible above the orange Xe atom at 3 σ due to low site occupancy.

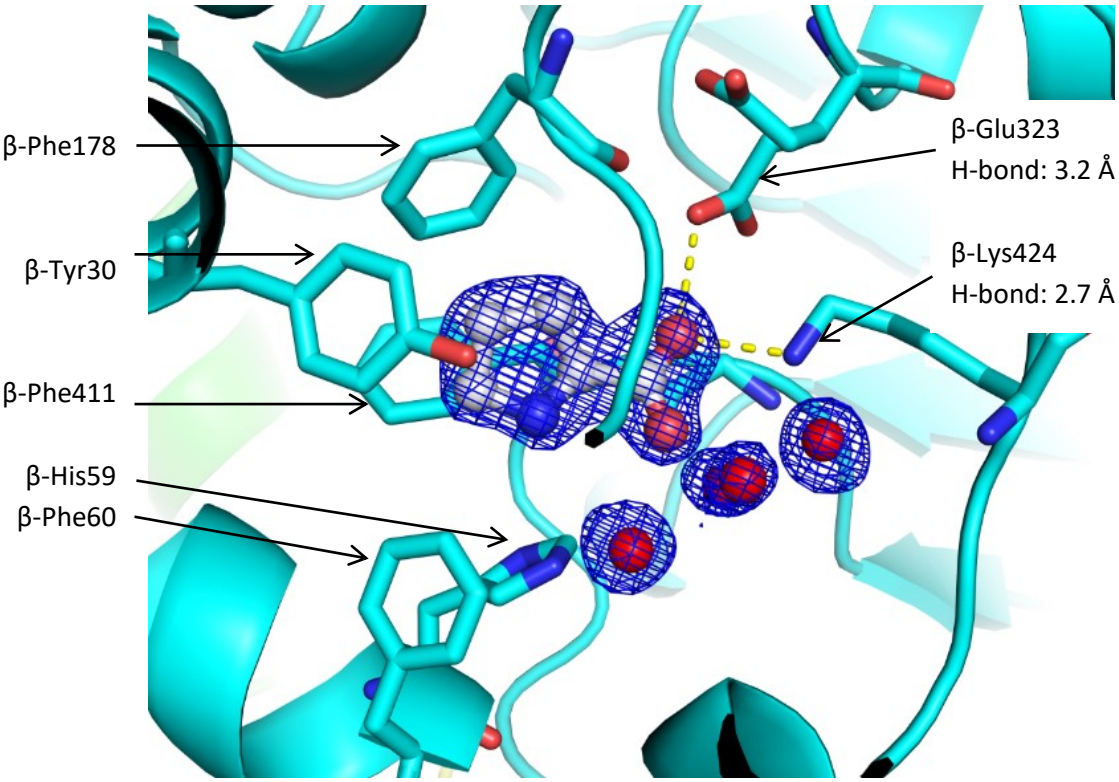


Figure IV-S2. PRL binding site with the neighboring residues in sticks. The PRL molecule is shown with spheres and sticks and is colored by element. The electron density of the PRL and nearby water molecules are shown in blue mesh contoured to 2.0 σ. Distances between the closest N atom of PRL and the N of β-Lys424 and OE2 of β-Glu323 are given. Note the well-defined electron density of the water molecules. See Experimental section of the main text for further discussion.

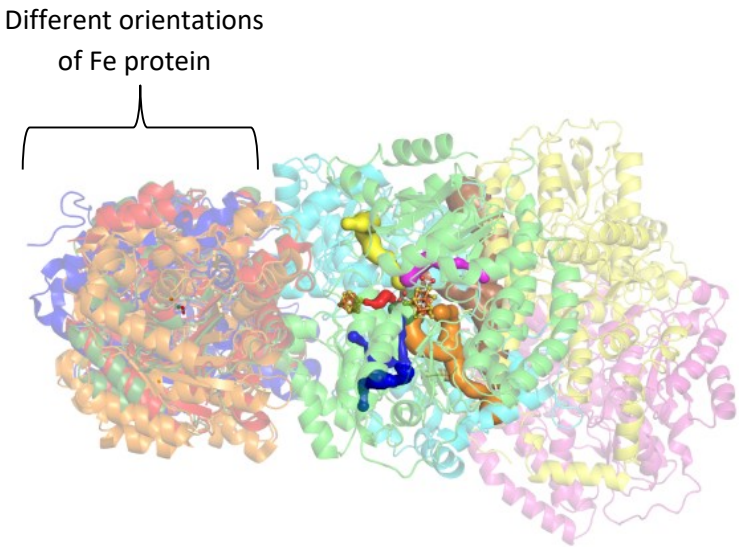


Figure IV-S3. Overlays of four different Av1-Av2 complexes: ADP·AlF_4^- stabilized Av1-Av2 (PDB ID 1M34, red cartoon), nucleotide-free Av1-Av2 (PDB ID 2AFH, blue cartoon), MgAMPPCP-bound Av1-Av2 (PDB ID 2AFK, forest green cartoon), and MgADP-bound Av1-Av2 (PDB ID 2AFI, orange cartoon). The docking of the Fe protein does not block any of the proposed substrate pathways.[28,67]

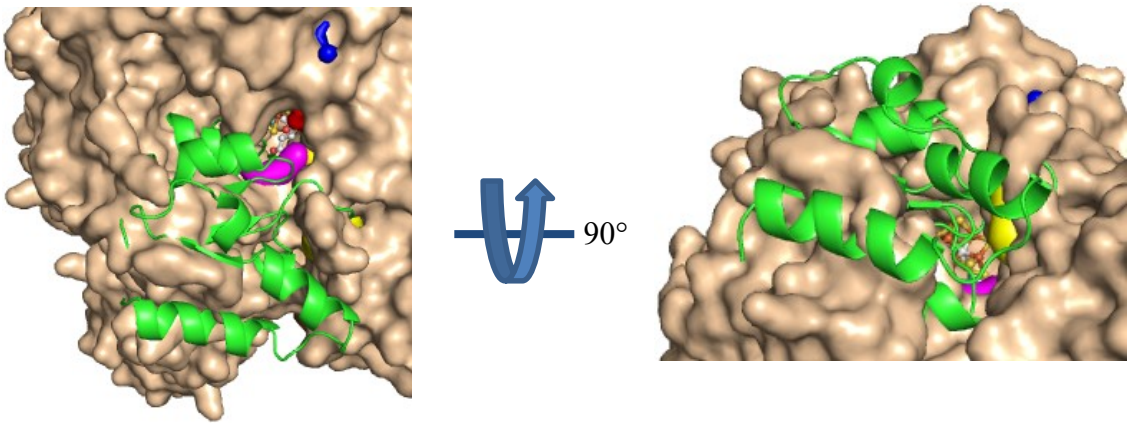


Figure IV-S4. Overlay of the molecular surface of the FeMo-cofactor-deficient Av1 protein (wheat surface, PDB ID 1L5H).[111] The channel used for transportation of the FeMo-cofactor is visible as the hole in the wheat surface that connects the protein surface to the location of the FeMo-cofactor in the mature protein. The green protein backbone is the α-subunit of the Av1-Xe structure, and the other colored surfaces belong to the substrate access pathways. Three are visible: the AI (blue surface), AII (magenta surface), and NH_3 egress (yellow surface) pathways. Part of the AII pathway overlaps with the FeMo-cofactor insertion pathway.

Chapter V

Thiocyanate and Methylamine Binding Sites in the MoFe Protein

V.A. Abstract

One of the most intriguing mysteries of nitrogenase is the mechanism by which it functions. Pieces of this puzzle may be revealed by determining the binding sites of small molecules that mimic substrates and products of nitrogenase. These binding sites yield information on how substrates and products may access and interact with the protein and its metal clusters. The binding sites of two small molecules, thiocyanate (SCN) and methylamine, are reported in this work. Understanding the sulfur binding sites is motivated by the observation that a sulfur atom is displaced from the FeMo-cofactor to accommodate CO-binding during inhibition; a sulfur returns to the belt position upon release of CO from the FeMo-cofactor.[37] Structural data from the MoFe protein of *Azotobacter vinelandii* (Av1) obtained at the resting, CO-inhibited, and re-activated states as well as the Av1-SCN co-crystal structure suggest that the displaced sulfur atom may reversibly bind in a positively charged pocket about 22 Å away from the FeMo-cofactor.[37] The work presented here aims to identify sulfur binding sites that may be present in the MoFe protein from *Clostridium pasteurianum* (Cp1). No sulfur binding sites in Cp1-SCN co-crystals were observed, suggesting that the displaced sulfur may roam about the protein or the water pocket at the active site during inhibition and turnover. Also presented in this work is a methylamine binding site. Methylamine is an analogue of hydrazine, a likely intermediate in the nitrogen fixation mechanism.[8] One methylamine binding site was observed

in Cp1, residing in the interstitial water channel. This provides further support for use of the interstitial water channel as a pathway for substrates and products.[78]

V.B. Introduction

The mechanism of biological nitrogen fixation remains inconclusive despite decades of work. A thorough understanding of the mechanism is highly sought for its potential impact on creating a sustainable alternative to the Haber-Bosch process of industrial ammonia production for fertilizers.[115] A recent breakthrough in the elucidation of the nitrogenase mechanism is the X-ray structure of CO bound to the Av1FeMo-cofactor – the first structure of a non-resting state MoFe protein.[37] CO is usually a non-competitive, reversible inhibitor in nitrogenase for all substrates except for protons.[116–118] In the CO-bound structure, one of the belt sulfur atoms of the Av1 FeMo-cofactor, S2B, is displaced by a CO molecule such that the carbon atom bridges Fe2 and Fe6. Inhibition of nitrogenase activity of this state was confirmed by enzyme activity assays. Activity was regained when the CO atmosphere was removed. The crystal structure of the resulting reactivated nitrogenase shows re-occupation of S2B in the FeMo-cofactor; thus, the displacement of S2B is reversible during inhibition. This study, along with a study on selenium substitution of belt sulfurs in the presence of CO,[38] suggests that substrate reduction may occur by substrate substitution of one or more belt sulfurs. The question remains: to where do the sulfur atoms relocate upon their displacement from the FeMo-cofactor?

By inspection of the anomalous difference Fourier map calculated from X-ray data of the CO-inhibited structure of the MoFe protein obtained at 7100 eV, a sulfur binding site was observed that is usually occupied by a water molecule in the Av1 resting state. This binding site is separated from the FeMo-cofactor by about 22 Å and resides in a small binding pocket between the α- and β-subunits and in a non-continuous water channel connected to the

interstitial water channel. The positive surface charge of this cavity suggests that the sulfur species may be HS^- or S^{2-}. The sulfur binding site is far away from the FeMo-cofactor, and the surrounding residues are poorly conserved among other nitrogenase species, which could indicate that the sulfur binding site is a general anion binding site. However, the density of the sulfur in this site decreased in the reactivated Av1 structure, suggesting that the sulfur binding site could be a sulfur reservoir. The location of the Av1 sulfur binding site was reproduced in co-crystallizations with SCN. To establish if Cp1 exhibits the same or similar sulfur binding sites, Cp1 was co-crystallized with SCN. The results are reported here. Also reported herein is the co-crystallization of Cp1 with methylamine, $H_3C\text{-}NH_2$. Methylamine is an analogue for hydrazine, $H_2N\text{-}NH_2$, which is an intermediate of the nitrogen fixation mechanism.[8] Methylamine binding sites provide clues on how polar substrates and products, such as NH_4^+, move between the protein surface and the active site.

V.C. Results and discussion

V.C.i. Cp1-SCN co-crystallization

X-ray crystallographic data for two Cp1 crystals made in the presence with 10 mM KSCN (called Cp1-SCN1 and Cp1-SCN2) were collected; structures were built by molecular replacement and refined (Table V-1). Inspection of the $2F_o\text{-}F_c$ and anomalous Fourier maps showed no blobs that could be modeled by a sulfur species anywhere within the protein other than in the cofactors or in methionine or cysteine residues. The structural overlay of Av1 and Cp1 (Figure V-1) shows that the Av1 sulfur binding site is blocked in Cp1 by residue sidechains. Cp1 was also crystallized in the presence of KSeCN and Na_2Se. No Se binding sites were observed at physiological pH, which confirms the lack of sulfur binding sites observed in the Cp1-SCN co-crystals.

Table V-1. X-ray crystallographic data collection and refinement statistics for Cp1-SCN1 and Cp1-SCN2

	Cp1-SCN1	Cp1-SCN2
Data Collection Statistics		
Resolution range (Å)	37.87-1.38 (1.38-1.40)*	39.63-2.05 (2.05-2.09)*
Wavelength (Å)	0.9537	0.9537
Space group	P1	P1
Unit cell constants	a = 67.01 Å α = 73.58° b = 72.78 Å β = 87.40° c = 108.33 Å γ = 84.23°	a = 66.99 Å α = 73.49° b = 73.21 Å β = 87.31° c = 108.48 Å γ = 83.92°
Unique reflections	384,397 (18,531)*	121,025 (6,016)*
Completeness (%)	98.9 (92.9)*	98.2 (97.0)*
Redundancy	7.0 (7.6)*	6.8 (6.9)*
$I/\sigma(I)$	15.2 (2.1)*	6.6 (2.0)
R_{merge}	0.076 (0.929)*	6.03 (70.349)*
Refinement Statistics		
Protein residues	1950/1984	1950/1984
Mean *B* value (Å^2)	11.5	14.3
R_{work}	0.146	0.232
R_{free}	0.164	0.274
Ramachandran outliers	9 (0.47%)	8 (0.41%)
RMSD bond lengths (Å)	0.012	0.009
RMSD bond angles (°)	1.54	1.18

*Numbers in parentheses represent data in the highest resolution shell.

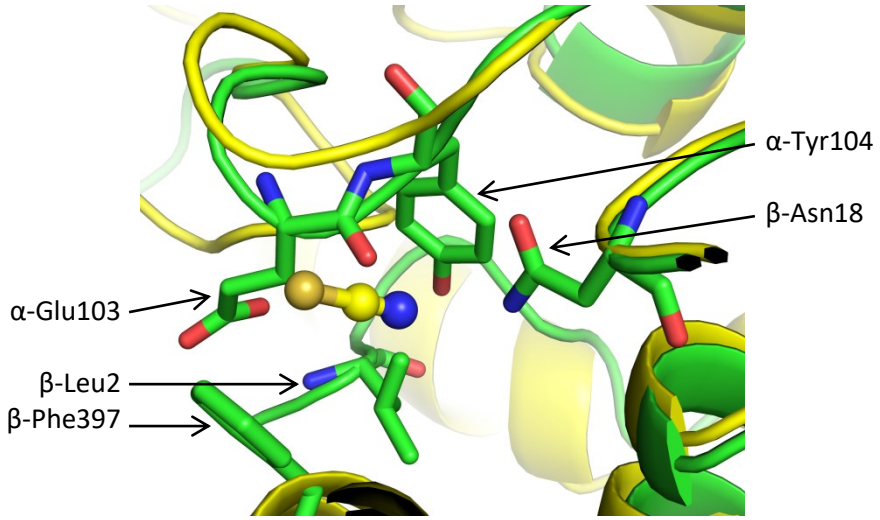


Figure V-1. Overlay of Av1-SCN and Cp1-SCN1. The yellow and green cartoons are that of Av1-SCN and Cp1-SCN1, respectively. The SCN molecule is displayed in ball and stick representation. Cp1-SCN residues blocking the Av1 SCN binding site are shown with sticks and labeled.

V.C.ii. Cp1-methylamine co-crystallization

Cp1-CH_3NH_2 co-crystals were grown in 5 mM methylamine. X-ray crystallographic data was collected, and the structure was built by molecular replacement and refined (Table V-2). Inspection of the $2F_o$-F_c map showed an elongated electron density blob within the α-subunit. A methylamine molecule modeled at this site has *B*-factor of 20 Å^2 – a little larger than the average *B*-factor of all protein molecules, 16.7 Å^2. This is reasonable considering that the methylamine exists in a channel. In this water channel, the methylamine nitrogen atom is 2.9 Å and 3.3 Å from the backbone carbonyl oxygen of α-Arg87 and α-Thr83, respectively (Figure V-2). The methylamine is 7 Å from the homocitrate, 8.5 Å from the FeMo-cofactor, and 11.5 Å from the P-cluster. All measurements were made using the two closest atoms of each species.

Table V-2. X-ray crystallographic data collection and refinement statistics for Cp1-CH_3NH_2

	Cp1-CH_3NH_2	
Data Collection Statistics		
Resolution range (Å)	39.10-1.73 (1.76-1.73)*	
Wavelength (Å)	0.9537	
Space group	P2	
Unit cell constants	a = 69.39 Å	α = 90°
	b = 149.7 Å	β = 103.6°
	c = 116.6 Å	γ = 90°
Unique reflections	235,226 (11,568)*	
Completeness (%)	98.0 (97.9)*	
Redundancy	6.9 (6.9)*	
$I/\sigma(I)$	10.8 (2.0)*	
R_{merge}	0.122 (1.000)*	
Refinement Statistics		
Protein residues	1950/1984	
Mean *B* value (Å^2)	16.7	
R_{work}	0.159	
R_{free}	0.184	
Ramachandran outliers	4 (0.21%)	
RMSD bond lengths (Å)	0.013	
RMSD bond angles (°)	1.46	

*Numbers in parentheses represent data in the highest resolution shell.

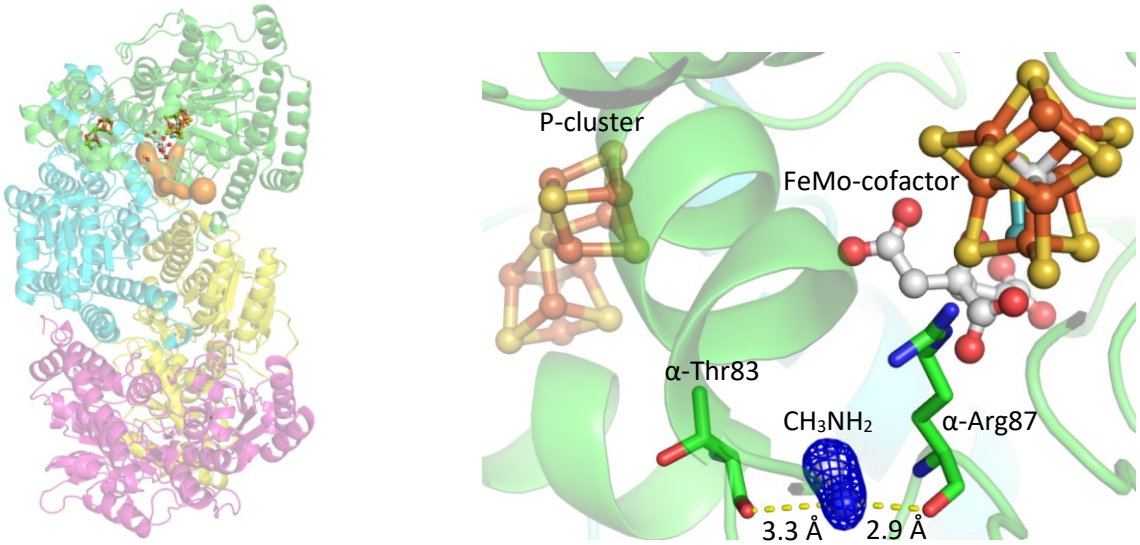


Figure V-2. Ribbon structure of Cp1-methylamine. The α-subunits are colored green and magenta, and the β-subunits are colored cyan and yellow. The cofactors and methylamine are shown with ball and stick and colored by element. The electron density of the methylamine is shown in blue mesh contoured to 1.5 σ. Close contacts to the methylamine are shown in sticks in the right image.

Cp1-CH_3NH_2 was superposed on structures of Av1-SCN, Av1-CO (PDB ID 4KTV), Cp1-Xe (PDB ID 4WN9), and Av1-Xe (PDB ID 4WNA).[37,78] No overlap was observed between the SCN, CO, or Xe atoms with methylamine. Additionally, no overlap was observed with the binding sites of other small molecules, such as sulfur, imidazole, ethane-di-ol, and the Cp1-proline ligand.[78]

V.D. Conclusions

The lack of S and Se binding sites in Cp1 co-crystals indicates that there may not be a defined sulfur binding site in Cp1. Perhaps a sulfur atom displaced from the FeMo-cofactor during turnover in Cp1 roams the water bath surrounding the FeMo-cofactor or elsewhere in the protein. When the MoFe protein returns to the resting state, the S2B binding site may be filled by any nearby sulfur atom, not necessarily by the same sulfur atom that was originally displaced from the FeMo-cofactor. The Cp1-methylamine structure shows a single methylamine

binding site in the interstitial water channel. This provides further support for use of this channel by polar substrates and products traveling between the protein surface and the active site.

V.E. Experimental

V.E.i. Protein purification and crystallization

Purification of Cp1 was performed as previously described.[21] Cp1 was crystallized using sitting-drop vapor diffusion in a Coy anaerobic chamber at room temperature (~295 K). The Cp1-SCN crystals were grown in a solution containing 0.1 M magnesium chloride, 24% (w/v) polyethylene glycol (PEG) with a molecular weight of 3350 g/mol, 10 mM thiocyanate (SCN), 80 mM Tris at pH 8.0, and 5 mM sodium dithionite. The Cp1-methylamine crystals were grown in a solution containing 0.3-0.5 M magnesium chloride, 14-18% (w/v) PEG 3350 g/mol, 5 mM methylamine, 80 mM Tris at pH 8.0, and 5 mM sodium dithionite. For both co-crystallizations, single crystals with block morphology appeared after one day. During crystal harvesting, crystals were first soaked in reservoir solution containing 10% 2-methyl-2,4-pentanediol (MPD) for 15 minutes, after which they were looped and flash-frozen in liquid nitrogen.

V.E.ii. Data collection and processing

Crystallographic diffraction data were collected on beam line 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL), with a Pilatus 6 M pixel array detector. The data were collected at 12,658 eV using an oscillation angle of 0.15° and 0.25° for Cp1-SCN and Cp1-methylamine, respectively. Because the Cp1-SCN crystals were in the P1 space group, two full data sets (360°) were collected, the second of which was horizontally and vertically offset from the first by 5 mm. In addition, two full sets of anomalous diffraction data (360°) were also

collected at 6.699 eV (the lowest possible energy of beamline 12-2) using an oscillation angle of 0.15° to identify the sulfur atoms. Like with the high energy data sets, the two data sets were offset by 5 mm horizontally and vertically. Although the X-ray wavelength of the low energy data sets is well above the K-edge of sulfur (2.4720 keV), the signal is still strong enough to observe anomalous scattering of the sulfur atoms, using the sulfur of known protein and inorganic atoms as a control. The crystallographic data were integrated with the XDS program package.[57] Scaling and integration were performed with AIMLESS from the CCP4 suite.[58] The model was built using PHASER, a molecular replacement program in CCP4, against the 1.08-Å Cp1 structure, 4WES.[58,64] The model was refined in COOT and with REFMAC5.[60,61,76] PyMOL was used to prepare the figures.[63]

Chapter VI

Progress toward Determining the Location of Hydrogen Atoms in the MoFe Protein using Neutron Diffraction

VI.A. Abstract

The heart of nitrogenase activity involves the transfer of electrons and hydrogen atoms between and around the metal clusters. Although X-ray crystallography is a powerful technique with high throughput, ultra-high resolution structures (<1.0 Å) are required to observe hydrogen atoms, which have thus far not been attained for nitrogenase. Alternatively, neutron diffraction may be used to observe hydrogen and deuterium atoms in a crystal structure at ~2 Å resolution. This work establishes conditions and crystal size required to obtain adequate neutron data sets of partially deuterated MoFe protein to observe the location of hydrogen and deuterium atoms. Using a spallation neutron source and a high flux isotope reactor at Oak Ridge National Laboratory, initial room temperature screens on partially deuterated Cp1 crystals show diffraction resolution >4 Å; cryogenic screens exhibit poorer resolution and diffraction quality. The next step is to increase the crystal size to ~1 mm^3, which should be sufficient to obtain ~2 Å resolution data and determine the location of hydrogen and deuterium atoms in the protein. This information would provide insight into where hydrogen is stored and how it is accessed during turnover in the MoFe protein. The long term goal of this project is to obtain neutron diffraction data of non-resting state structures of the MoFe protein and compare the location of hydrogen atoms to that of the resting state structure in order to observe how hydrogen moves around the cofactors during substrate turnover.

VI.B. Introduction

Nitrogenase catalyzes the reduction of atmospheric dinitrogen into its bioavailable form, ammonia, in a process called nitrogen fixation. The enzyme consists of the Fe and MoFe proteins. The MoFe protein contains two types of metal clusters, the [8Fe:7S] P-cluster and the [7Fe:Mo:9S:C:*R*-homocitrate] FeMo-cofactor, the latter of which is the active site. Two of the most commonly studied bacterial species that express nitrogenase are *Azotobacter vinelandii* (Av) and *Clostridium pasteurianum* (Cp), for which the MoFe proteins are named Av1 and Cp1, respectively. Due to the complexity of nitrogenase, the mechanism of biological nitrogen fixation remains elusive. High resolution X-ray structures of the MoFe and Fe proteins have revealed the identity and location of all atoms heavier than hydrogen with structures refined to 1.0 and 1.08 Å for Av1 and Cp1, respectively.[22,64] The high resolution X-ray data, combined with electron nuclear double resonance spectroscopy, was used to identify the central atom in the FeMo-cofactor as a carbon species.[22,64] The electron density of this atom was obscured in all previous structures due to perturbations from the neighboring electron-dense Fe atoms. Additionally, the binding mode of the inhibitor, carbon monoxide, to the FeMo-cofactor of Av1 was determined using X-ray diffraction.[37] This was the first observation of any species bound to the nitrogenase active site.

Studies on the binding, activation, and cleavage of N_2 have concluded that the participation of protons and/or hydrides at different steps could be highly important to elucidating the complete mechanism of biological nitrogen fixation.[22,34] Neither 1.0-Å resolution X-ray data nor data from other techniques have established the positions of hydrogens on the metalloclusters or solvent that are critical for proton transfer processes in nitrogenase. X-ray diffraction does not usually detect hydrogen because it involves the detection of X-rays elastically scattered from electrons in a crystal to yield an electron density map; the scattering is

roughly proportional to the atomic number. Additionally, X-ray radiation damages protein crystals, which limits the dose of radiation one can apply to a crystal. Liquid state NMR can also be used to determine protein structure, although this becomes increasingly difficult with larger molecules due to overlapping signals; thus, atomic resolution of structures cannot be reasonably solved for proteins as large as the 220-kDa MoFe protein.[119]

Given current technology, the most suitable method to determine the location of hydrogen atoms is neutron diffraction combined with X-ray crystallography. Neutron diffraction data are gathered by detecting neutrons that are elastically scattered from nuclei in crystallized proteins. Fourier transform of the reflections from diffraction patterns gives a nuclear density map to which a protein structure can be modeled. Deuterium and hydrogen in a protein structure can be observed at resolutions around 2-3 Å, depending on the protein size and deuteration level.[120] Hydrogen and deuterium are easy to distinguish from each other because hydrogen has a negative incoherent (inelastic) scattering length of $-3.74\ 10^{-15}$ m while deuterium has a positive coherent (elastic) scattering length of $6.67\ 10^{-15}$ m.[121] The incoherent scattering from hydrogen results in high background noise, but the signal-to-noise ratio can be reduced by soaking the crystals in D_2O-based solutions, crystallization in D_2O-based solutions, or growing the bacteria in D_2O-replaced media – techniques that offer increasingly better deuteration and therefore stronger coherent scattering signals.[122] Another advantage of neutron diffraction is that it does not cause radiation damage to the crystal.[123]

Despite the advantages of neutron diffraction, the technique is still in its infancy: the first neutron structure, that of myoglobin, was determined in 1969,[124] and the second structure, trypsin, was solved in 1980.[125] Currently, there are less than 100 neutron structures in the Protein Data Bank, compared to approximately 114,000 X-ray structures. The use of neutron diffraction is becoming increasingly popular due to ongoing technological progress allowing

smaller and smaller crystals with larger unit cell volumes to be imaged. Neutron data up to 1.1-Å resolution and with up to 95% visibility of hydrogen and deuterium have been reported.[126] Furthermore, newly engineered instruments with higher flux and increased detector sensitivity have decreased the data collection time period from months to days or weeks, depending on the size of the crystal, crystal symmetry, volume of the unit cell, and deuteration level.[127] Some instruments also provide neutron time-of-flight data so that the data can be binned for optimal signal-to-noise. When using neutron diffraction, large crystal sizes (>0.3 mm^3) and smaller unit cell volumes (<1 x 10^5 $Å^3$) are advantageous to compensate for the inherent weak flux of neutron sources, which is several orders of magnitude weaker than that of synchrotron X-ray sources.[123] The structure refinement of neutron datasets has also been an issue because there are roughly twice as many parameters compared to X-ray data, but new refinement programs specifically for solving neutron structures are available.[123]

The goal of this project is to observe the position of protons and/or hydrides near the metal centers in the resting state Cp1 protein using neutron diffraction. This knowledge will (1) provide information on the overall dynamics of the enzyme by showing the exchangeability of protons on amino acid residues possibly involved in proton transport, and (2) dictate the required crystal size and deuteration level for adequate neutron diffraction. Looking ahead, this will inform neutron diffraction experiments on non-resting state Cp1, which will aid in the understanding of hydride intermediates occurring on the metal clusters.

VI.C. Instrumentation

The neutron diffractometers used in this work are IMAGINE and MaNDi, which are instruments on the High Flux Isotope Reactor (HFIR) and Spallation Neutron Source (SNS), respectively, at Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee. IMAGINE

provides a constant flux of neutrons from a Uranium-235 core with a flux of 2.5 x 10^{15} neutrons/cm^2/s. The instrument operates under quasi-Laue conditions, detecting neutrons with wavelengths between 2-10 Å. It is advantageous to detect longer wavelength neutrons because they scatter better than shorter wavelength neutrons. The detector is a neutron image plate (Gd_2O_3 doped BaF(Br.I):Eu^{2+}). MaNDi operates with a spallation source with a frequency of 60 Hz (~10^7 neutrons/cm^2/s). MaNDi is also quasi-Laue, but has a smaller range of neutron detection: 2-4 Å. The detectors on MaNDi are anger lithium glass scintillators. When a neutron collides with the detector, it causes emission of ~50 photons, which produces a current that is measured by the detector. Cumulative reflections can be observed in real time. Additionally, because time-of-flight data is stored for each neutron, data is binned according to neutron wavelength. Neutrons of higher wavelength are better scatterers and allow better spacing of low resolution reflections, but also lead to more noise. Thus, by binning the data, the signal-to-noise ratio can be optimized.

VI.D. Progress to date

VI.D.i. Growing crystals

Typical crystallization experiments included 11.5-15% PEG with molecular weight 3350 g/mol, 0.2-0.5 M magnesium chloride, 0.08 M Tris/HCl at pH 8.0, and 5 mM sodium dithionite in D_2O using the sitting drop, vapor diffusion crystallization method in an anaerobic chamber at room temperature. Crystals shaped as rectangular prisms started growing within 24 hours. Most reached maximum size within a couple weeks. No crystal degradation was observed for crystals left in the crystallization drop up to six months. Crystal growth required that the crystallization drop consist of a protein stock to reservoir solution ratio of 1:1 (typically 2.0 or 2.5 µL of each). It was found that more highly concentrated protein stock produced larger crystals.

Crystals sent to ORNL for screening were typically around 0.1 mm^3. Several seeding techniques were employed to increase crystal size, including streak seeding from crushed crystals, feeding (adding additional protein stock to wells containing a single crystal), and macro-seeding (soaking a single crystal in fresh reservoir solution to dissolve the outer layer of protein molecules and then transporting the crystal to a fresh protein drop). Only streak seeding was successful. Feeding sometimes increased crystal size but mosaicity decreased. Macro-seeding tended to partially dissolve the crystal.

VI.D.ii. Deuteration

Increased deuteration of a protein crystal leads to decreased noise from inelastic scattering off of hydrogen atoms. Partial deuteration was achieved by crystallizing the protein stock in deuterated solvents. With a 2 µL protein drop exposed to 1 mL of deuterated reservoir solution, the ratio of H_2O to D_2O is approximately 2:1000 – enough to replace all exchangeable protons with deuterium in the growing protein crystal. To achieve complete deuteration (perdeuteration), the bacteria must be grown in deuterated media with a deuterated carbon source. This is an expensive endeavor as well as challenging, since the use of D_2O instead of H_2O significantly impacts the kinetics of cell growth. Due to these limitations, only a few species of bacteria have been optimized for growth in deuterated media, none of which have been shown to express nitrogenase.

VI.D.iii. Mounting, storing, and shipping crystals in capillaries

A detailed description for mounting and storing crystals in capillaries is provided in Chapter II. Capillaries arrived intact and undamaged to ORNL when shipped via FedEx Ground and FedEx Air.

VI.D.iv. Determining crystal stability and ideal screening temperature

At the time of screening, both IMAGINE and MaNDi operated on room temperature conditions but only MaNDi could operate at cryogenic temperatures. Cryogenic temperatures are favorable because they protect nitrogenase crystals from oxidation and degradation. Since neutrons cause no radiation damage, room temperature neutron data collections are feasible given sufficient crystal stability. Crystals were assessed for diffraction resolution at 298 K and 100 K on an in-house X-ray generator (Table VI-1), which has fluxes comparable to neutron sources. The results show no significant difference in diffraction resolution. For comparison, crystals were also screened at the Stanford Synchrotron Radiation Lightsource (SSRL) at 100 K.

Collecting a full neutron data set on a MoFe protein unit cell with $P2_1$ symmetry would require a whole cycle of beam time (~23 days). Crystals must be stable for at least two months, to allow time for sample preparation, shipping, and data collection. Crystals stability in capillaries at room temperature was investigated. In this experiment, diffraction quality was assessed at various time points after transfer to capillary. It was found that crystals could not be exposed to the X-ray beam more than once, even if for a single screen shot, due to radiation damage. Therefore, 30 crystals from the same seeding wells were mounted in capillaries on the same day. Every few days, a fresh crystal was screened on the in-house generator. Crystals started exhibiting poor diffraction quality after about six weeks; thus, a room temperature neutron dataset would likely have to be made from two or more merged datasets.

Table VI-1. X-ray crystallographic data collection and refinement statistics

	Room temp, in-house	**100 K, in-house**	**100 K, SSRL**
Data Collection Statistics			
Resolution range (Å)	29.25 – 2.06 (2.06 – 2.17)	19.77-2.00 (2.03-2.00)	36.84 – 1.62 (1.66 – 1.62)
Wavelength (Å)	1.5418	1.5418	0.9537
Space group	$P2_1$	$P2_1$	$P2_1$
Unit cell constants	a = 69.97 Å b = 151.3 Å c = 117.8 Å α = 90.00° β = 103.5° γ = 90.00°	a = 69.4 Å b = 149.6 Å c = 116.7 Å α = 90.00° β = 103.8° γ = 90.00°	a = 69.4 Å b = 148.8 Å c = 116.4 Å α = 90.00° β = 103.9° γ = 90.00°
Unique reflections	138,737	149394 (7181)*	3311 (21193)*
Completeness (%)	94.7 (92.5)*	96.2 (93.4)*	99.3 (99.0)*
Redundancy	2.9 (2.8)*	7.7 (7.8)*	7.2 (5.1)*
$I/\sigma(I)$	6.2 (2.0)*	13.6 (3.0)*	39.1 (2.3)*
R_{merge}	0.128 (0.636)*	0.150 (0.696)*	0.030 (0.593)*
Refinement Statistics			
Protein residues	1950/1984	1950/1984	1950/1984
Mean B value (Å^2)	20.0	13.8	18.8
R_{work}	0.176	0.184	0.167
R_{free}	0.221	0.223	0.196
RMSD bond lengths (Å)	0.011	0.012	0.012
RMSD bond angles (°)	1.47	1.85	1.78
Ramachandran outliers	10 (0.52%)	5 (0.26%)	5 (0.26%)
Water molecules	494	1181	559

*Numbers in parentheses represent data in the highest resolution shell.

VI.D.v. Screening crystals

Screening was performed on IMAGINE using a beam size of 2.5 x 3.0 mm^2. A crystal with dimensions of 0.5 x 0.5 x 0.3 mm (0.075 mm^3) was subjected to neutron wavelengths of 2-10 Å, and the best diffraction was 7-8 Å resolution (Figure VI-1). Larger crystals (~0.1 mm^3) were screened using a restricted band pass, 2.8-4.0 Å, which yielded better diffraction quality and resolution (~4 Å); however, higher resolution and more complete diffraction patterns are still needed. Three months later, the crystals diffracting to ~4 Å were screened again on IMAGINE.

No diffraction was observed, and the crystals were noticeably lighter in color, indicating that oxidation had occurred. No obvious damage to the capillary was apparent. Any future data collections on IMAGINE would require multiple crystals to collect a complete dataset.

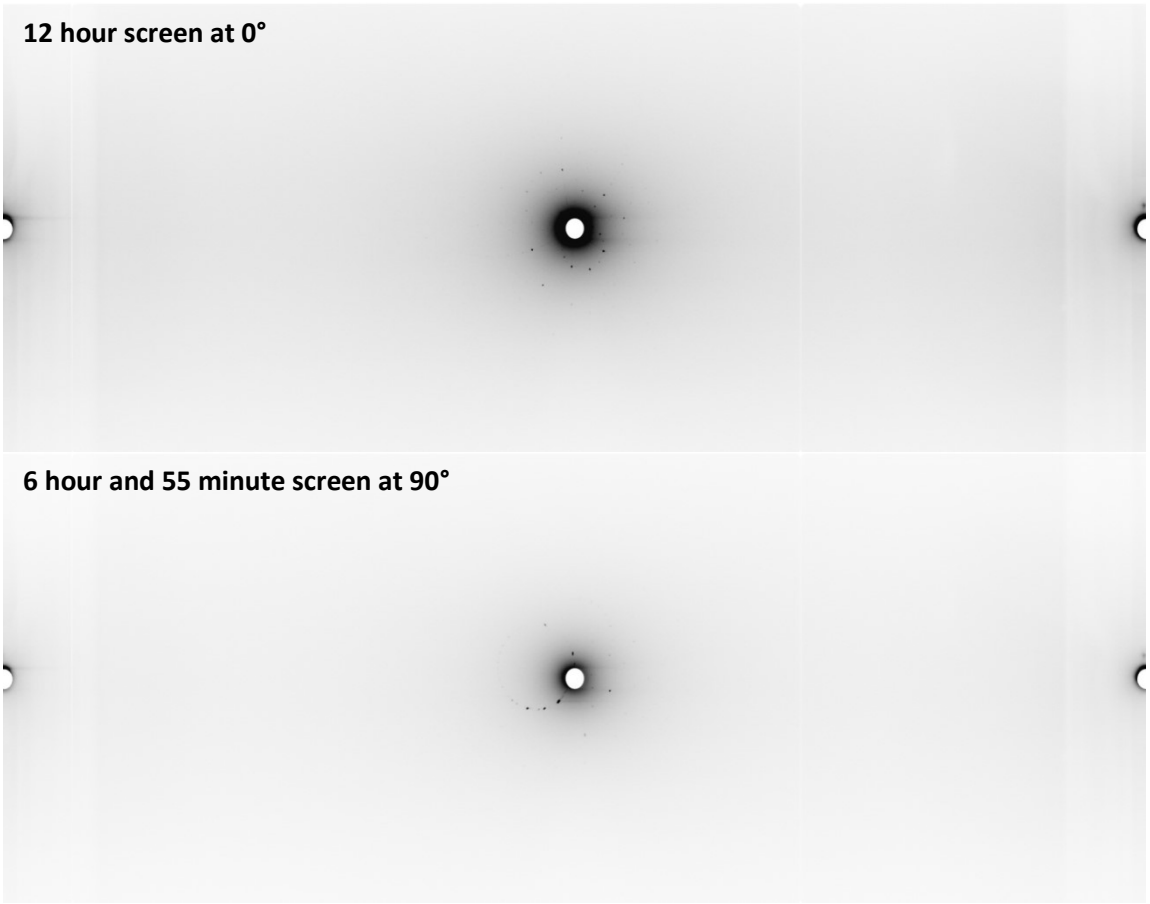


Figure VI-1. The best diffraction observed with Cp1 crystals on IMAGINE was 7-8 Å. Longer screen times yield better diffraction but increased noise. The screening images are cylindrical to reflect the shape of the detector.

Screening on MaNDi was performed to assess the potential of the cryogenic system. Despite our best efforts, we were unable to prevent ice buildup. After a 22-hour screen with a 1 mm beam aperture, 5-7 Å diffraction resolution was observed, but it was very noisy. Thus, room temperature neutron experiments are more favorable compared to cryogenic temperatures.

VI.E. Conclusions and next steps

Although Cp1 crystals are more stable at cryogenic temperatures, the cryogenic systems on MaNDi were not reliable at the time of screening. For room temperature data collection, stability tests indicate that multiple crystals would need to be imaged and their datasets merged to create a full neutron dataset. Regarding instrument selection, IMAGINE is preferred over MaNDi due to its ability to detect longer wavelength neutrons, which are better scatterers and therefore advantageous for proteins with large unit cells.

The immediate next step in this work is to increase crystal size to ~1 mm^3, which is a ten-fold increase from current crystal sizes. This scale increase necessitates a new approach to growing crystals, such as batch and large-protein drop (1 mL or more) vapor diffusion methods. These large scale crystallization methods significantly impact crystallization kinetics, so factors such as feeding, temperature, and protein-to-buffer ratios will have to be meticulously controlled during the scale-up process to determine conditions that facilitate large, single crystal growth. If the appropriate crystal size can be achieved and a whole neutron dataset resolved to at least 2 Å can be obtained, information on the location of hydrogen atoms on or near the cofactors may be observed. This will inform us on how hydrogen may be used and accessed during the course of substrate turnover, which would be invaluable information in the goal of elucidating the mechanism of biological nitrogen fixation.

VI.F. Acknowledgments

Many thanks to Dr. Flora Meilluer and Dr. Matthew Cuneo for their assistance in screening experiments on IMAGINE and MaNDi, respectively, and their input on this project.

Chapter VII

Reversible Protonated Resting State of the Nitrogenase Active Site

VII.A. Abstract

Protonated states of the nitrogenase active site are mechanistically significant since substrate reduction is invariably accompanied by proton uptake. We report the low pH characterization by X-ray crystallography and EPR spectroscopy of the nitrogenase molybdenum iron (MoFe) proteins from two phylogenetically distinct nitrogenases (*Azotobacter vinelandii*, Av, and *Clostridium pasteurianum*, Cp) at pHs between 4.5 to 8. X-ray data at pHs of 4.5 – 6 reveal the repositioning of sidechains along one side of the FeMo-cofactor, and the corresponding EPR data shows a new S = 3/2 spin state with spectral features similar to a state previously observed during catalytic turnover. The structural changes suggest that FeMo-cofactor belt sulfurs S3A or S5A are potential protonation sites. Notably, the observed structural and electronic low pH changes are correlated and reversible. The detailed structural rearrangements differ between the two proteins, which may reflect differences in potential protonation sites at the active site among nitrogenase species. These observations emphasize the benefits of investigating multiple nitrogenase species. Our experimental data suggests that reversible protonation of the resting state is likely occurring, and we term this state "E_0H^+", following the Lowe-Thorneley naming scheme.

VII.B. Introduction

Nitrogen fixation is the process of breaking the kinetically inert N-N triple bond via either reduction or oxidation of molecular dinitrogen. Biologically, nitrogen fixation is accomplished by the enzyme nitrogenase to yield ammonia, with an overall reaction stoichiometry conventionally described by Equation VII-1.

Equation VII-1. Biological nitrogen fixation

$$\mathrm{N_2 + 10H^+ + 8e^- + 16ATP \rightarrow 2NH_4^+ + H_2 + 16ADP + 16P_i}$$

Nitrogenase is a highly oxygen-sensitive enzyme present in specialized microorganisms; it consists of two proteins called the molybdenum-iron (MoFe) and iron (Fe) proteins.[26,33,128] The Fe protein contains two nucleotide binding sites and a 4Fe:4S cluster. The MoFe protein incorporates two 8Fe:7S "P-clusters" and two 7Fe:9S:C:Mo:*R*-homocitrate "FeMo-cofactors", the latter of which represents the active site where substrates bind and are reduced. ATP-dependent electron transfer occurs from the 4Fe:4S cluster to the P-cluster during docking interactions between the Fe and MoFe proteins, after which the proteins separate.[31,40,45] Substrates can only bind to forms of the FeMo-cofactor more reduced than the resting state. These states are conventionally designated as E_n, where *n* represents the number of electrons transferred to the MoFe protein (per active site); E_0 is the resting state.[40] Following the Lowe-Thorneley model, dinitrogen binds to the FeMo-cofactor in the E_3 and E_4 states; however, other substrates, such as acetylene, may bind to the FeMo-cofactor in less highly reduced states.[45]

Electron paramagnetic resonance (EPR) is a powerful tool for studying the electronic states of the FeMo-cofactor since the E_0 state exhibits a strong, unique rhombic spectrum, resulting from transitions within the ±1/2 ground-state Kramers' doublet of a $S = 3/2$ system.[129]

In contrast, the P-cluster is diamagnetic in the dithionite-reduced form (P^N) and exhibits a weak resonance at g = 12 in the oxidized form (P^{ox}).[130,131] The reported EPR spectra of the FeMo-cofactor under turnover conditions include three spin states called 1a, 1b, and 1c.[132–134] 1a is the resting state (E_0); 1b and 1c, which are in equilibrium with 1a,[134] are attributed to E_2 and are thought to represent different states of the FeMo-cofactor during turnover. More specifically, 1c is assigned to protonation of the FeMo-cofactor.[133] The E_1 state is EPR-silent.

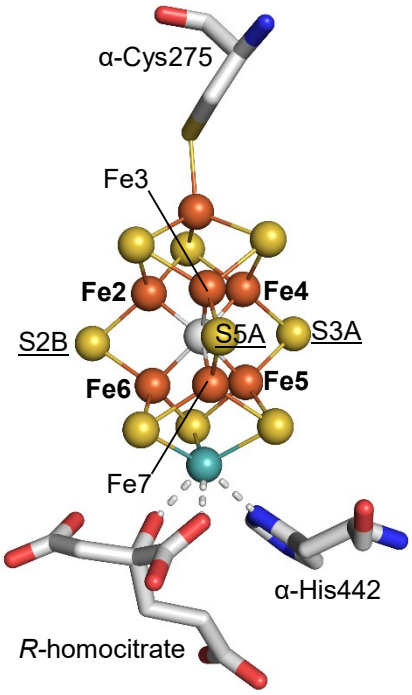


Figure 1. Structure of the FeMo-cofactor. The atoms of the cluster are shown in spheres and colored by element (Fe, orange; S, yellow; C, gray; Mo, cyan). Fe sites in the trigonal prism around the interstitial carbide are labeled. Belt S are also labeled and underlined. Coordinating residues and the R-homocitrate are shown in sticks and colored by element (C, gray; O, red; N, blue).

The FeMo-cofactor (Figure VII-1) contains a 7Fe:9S:C:Mo:homocitrate cluster exhibiting approximate C_{3v} symmetry. The core of the FeMo-cofactor is provided by a trigonal prism of six

Fe atoms (Fe2-7) surrounding an interstitial carbide.[22,72,135] Each face of the trigonal prism is bridged by one of three "belt" S labeled S2B, S3A, and S5A. Crystallographic evidence for turnover-dependent rearrangements of belt sulfurs is demonstrated by the reversible displacement of S2B upon CO inhibition.[37] Se from selenocyanate may also substitute S2B.[38] In the presence of substrate and under turnover conditions, interchange of the belt sulfurs was established such that Se originally at S2B migrates to S5A and S3A before ultimately exiting the FeMo-cofactor.[38] Intriguingly, the S2B site displaced by CO spans Fe2 and Fe6, which have been shown to be more oxidized in the resting state,[136] suggesting that their reduction is critical for ligand binding at this site.

There is still a high level of uncertainty in the mechanistic description of biological nitrogen fixation, including possible structural rearrangements in the FeMo-cofactor. The challenge has been to generate significant populations of higher E_n states competent for substrate binding. As formation of these states is associated with proton uptake, we reasoned that by studying the MoFe protein at low pH (high proton concentration), features of the active site that are characteristic of more highly reduced forms might be stabilized through Le Chatelier's principle. The effects of low pH (pH ≤ 5) on the X-ray structure and EPR spectra of the MoFe protein have not to our knowledge been described, likely because it has been reported that the MoFe protein is inactivated below pH 6.2.[137] However, our study shows that impacts to the atomic and electronic structure are reversible between pH 4.5 and pH 8.

In this study, we examine the two phylogenetically distinct nitrogenase MoFe proteins from *Azotobacter vinelandii* (Av1) and *Clostridium pasteurianum* (Cp1), which have a sequence identity of ~36%.[24] Working with Cp1 and Av1, we combine a structural approach with EPR spectroscopy to examine the atomic and electronic structure of MoFe proteins at pH 5 where the proton concentration is 2-3 orders of magnitude greater than that of typical enzyme activity

measurements. Changes occurring in the MoFe protein at low pH might therefore provide crucial information about the atomic and electronic structure of the protein at an early stage of substrate reduction.

VII.C. Results and Discussion

Over the pH range between 4.5-5.75, X-ray crystal structures of Cp1 and Av1 (Table VII-1) reveal structural rearrangements near the Fe3,4,5,7 face of the FeMo-cofactor (Figure VII-2) that are fully reversible upon returning to pH 8. For these studies, the purified protein was resuspended in a low pH tri-buffer system,[138] allowing the pH of the protein solution to be varied from pH 2 to pH 7 with minimal variation in the ionic strength and buffer components. Av1 and Cp1 exhibit a partially and fully occupied low pH conformer, respectively, when pH $\leq$ 5. We determined the pH 5 structures of Cp1 and Av1 at resolutions of 1.85 Å and 2.30 Å, PDB IDs 5VPW and 5VQ4, respectively. At pH ~6.5, Cp1 exhibits both conformations; the PDB ID for this structure is 5VQ3.

The conversion of the pH 8 to the low pH conformer under different pH and ionic strength was explored in Cp1 over a large number of conditions. It was found that higher ionic strength contributes to increased occupancy of the low pH conformer, which occurred at pH 5.8 or lower, depending on ionic strength. In view of the dependence of conformer occupancy on pH and ionic strength, as well as the challenges of measuring pH in small volumes around crystals and the uncertainties in extrapolating pH values measured to room temperature to the cryogenic temperatures used for crystallography and EPR, for simplicity, the acid-induced Av1 and Cp1 structural rearrangements are herein referred to as the low pH or pH 5 conformers.

At low pH in Cp1, a peptide flip[139] occurs between α-Ser346 and α-Arg347 (corresponding to Av1 residues, α-Leu358 and α-Arg359, respectively), causing the arginine

sidechain to reposition away from the face of the FeMo-cofactor (Figures VII-2b,3). Notably, this low pH rearrangement causes changes in hydrogen bonding interactions between sidechain atoms of α-Arg347 and S3A and S5A in Cp1 (Figure VII-3): (1) S5A loses its only hydrogen bond to NH1, (2) S3A loses its contact with the amide NH, and (3-4) S3A gains contacts with NH1 and NE. In Av1 at low pH, the sidechain of Av1 α-His274 (adjacent to the FeMo-cofactor ligand α-Cys275 and corresponding to Cp1 α-Gln261) moves closer to the FeMo-cofactor and displaces a water molecule. At this new position, a water molecule bridges the Av1 α-His274 sidechain and S5A of the FeMo-cofactor (Figure VII-2c). Of the two residues most affected by low pH in Cp1 and Av1, Cp1 α-Arg347 is invariant in all nitrogenases, whereas Av1 α-His274 is variant.[24] Mutagenesis of these residues in Av1 significantly reduces substrate reduction,[140,141] and α-His274 has been implicated in FeMo-cofactor insertion during Av1 assembly.[109]

Table VII-1. X-ray crystallographic data collection and refinement statistics

	Av1 at pH 5 (5VQ4)	Cp1 at pH 5 (5VPW)	Cp1 at pH 6.5 (5VQ3)
Data Collection Statistics			
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions a, b, c (Å); α, β, γ (°)	81.32, 128.9, 108.4 90, 110.9, 90	69.62, 146.3, 116.7 90, 103.6, 90	69.48, 148.0, 116.7 90, 103.5, 90
Resolution (Å)	39.54-2.30 (2.30-2.34)*	39.19-1.85 (1.88-1.85)*	39.83-1.75 (1.75-1.72)*
R_{merge}	0.174 (0.720)*	0.105 (0.684)*	0.079 (0.682)*
I/σ(I)	9.2 (3.1)*	11.6 (2.5)*	13.5 (2.9)*
Completeness (%)	98.8 (99.4)*	98.4 (95.4)*	98.4 (98.4)*
No. unique reflections	91,309 (4,321)*	189,858 (1,197)*	238,230 (11,876)*
Redundancy	6.7 (7.1)*	6.5 (6.2)*	6.8 (7.0)*
Refinement Statistics			
R_{work} / R_{free}	0.176 / 0.226	0.167 / 0.201	0.159 / 0.185
Average B-factor	24.0	30.0	29.0
R.m.s. bond lengths (Å)	0.011	0.012	0.013
R.m.s bond angles (°)	1.39	1.41	1.52

*Highest resolution shell is shown in parentheses.

The low pH structural rearrangements only occur on the face of the FeMo-cofactor that is exposed to water molecules (Fe3,4,5,7), potentially implicating this water pool (and likely the water channel that connects this pool to the protein surface) in proton transport between the active site and the exterior.[78,102,104] Additionally, there is slight movement (<1 Å) of the C1 carboxyl of the *R*-homocitrate *away* from α-Gln191 in Av1. A previously reported structure of Av1 at high pH shows slight movement of the C1 carboxyl *toward* α-Gln191,[142] which, in combination with results reported herein, indicates conformational flexibility in the C1 arm of the *R*-homocitrate in response to pH, possibly due to change in protonation state of the carboxylate group.

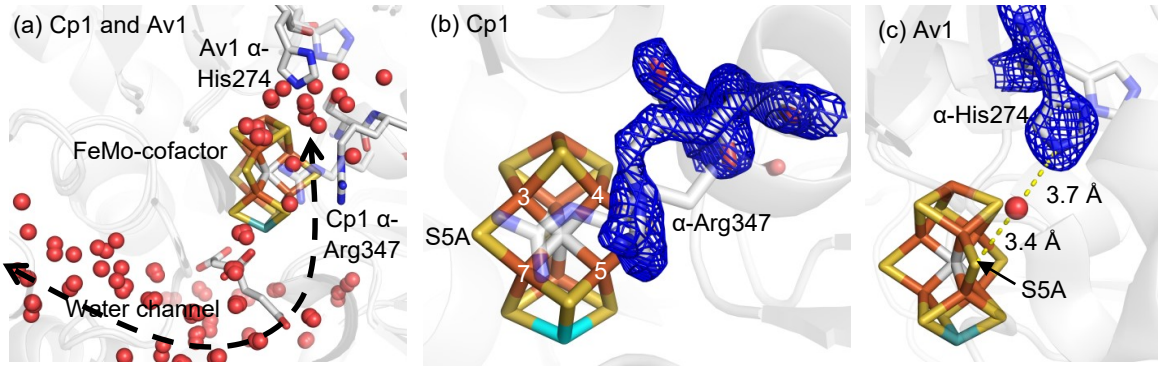


Figure VII-2. (a) Overview of the structural rearrangements observed at low pH at the active sites of Cp1 and Av1. Both changes occur on the Fe3,4,5,7 face of the FeMo-cofactor, which is the same face that is exposed to water molecules and connects to the interstitial water channel (dashed black line). (b) In Cp1, a peptide flip occurs between α-Arg347 and α-Ser346. Also, the Arg sidechain relinquishes its hydrogen bond with S5A and withdraws from the Fe3,4,5,7 face of the FeMo-cofactor. (c) In Av1, the α-His274 sidechain swings closer to the FeMo-cofactor and displaces a water molecule; two water molecules fill the former α-His274 sidechain position. The α-His274 coordinates to S5A of the FeMo-cofactor through a hydrogen bond bridge with a water molecule. In all images, transparent gray represents physiological pH structures. Non-transparent gray sticks show the low pH structural changes. The FeMo-cofactor and pH-affected residues are displayed as sticks and colored by element (yellow, S; orange, Fe;, cyan, Mo; gray, C). Water molecules are represented as red spheres. The blue meshes in (b) and (c) show the electron density maps of the pH affected residues contoured to 2.0 and 1.5 σ, respectively.

The low pH conformation changes could be triggered by proton binding to either the protein (possibly the sidechains of His, Glu, and Asp), as well as water and/or sites on the FeMo-cofactor such as the sulfurs and/or homocitrate. Without direct visualization of hydrogens, it is not possible to establish unambiguously which atoms are protonated to trigger the observed structural rearrangements. After close examination of the FeMo-cofactor and active site residues in the low pH and physiological pH structures, we see no obvious indicators for protonation of sidechains. It is also conceivable that protonation could be coupled to anion binding, such as buffer or counter-ion components in the buffer, but we see no evidence for this possibility, based on no appearance of new or shifted peaks in the solvent region. This leaves the possibility that the low pH rearrangements may reflect protonation of water and/or the FeMo-cofactor. Other than the homocitrate, the sulfurs represent the most likely site of protonation on the cofactor based on the pH titration properties of synthetic and protein-based clusters.[143–145]

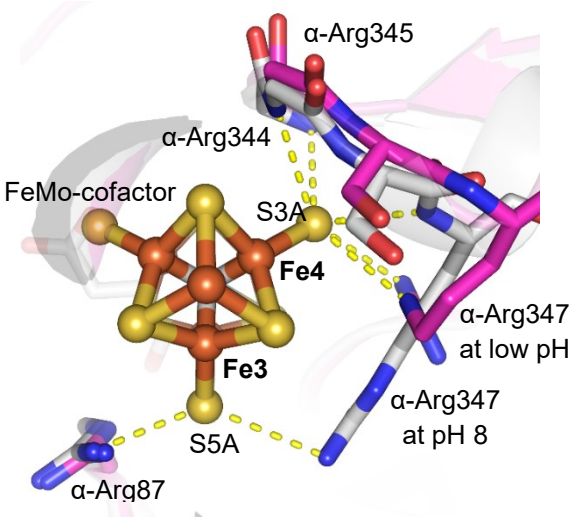


Figure VII-3. Structure of the Cp1 FeMo-cofactor as viewed down the C_3 axis. α-Arg347 at pH 8 (gray) and low pH (magenta) is shown in sticks. Contacts with the FeMo-cofactor at pH 8 and low pH are indicated with dashed yellow lines. All contact distances are ≤3.5 Å. The atoms of the cluster are shown in spheres and colored by element (Fe, orange; S, yellow; C, gray; Mo, cyan). Relevant Fe and S atoms are labeled. Coordinating residues and the *R*-homocitrate are shown in sticks and colored by element (C, gray; O, red; N, blue).

Following structural characterization by X-ray crystallography, EPR spectroscopy was performed on Cp1 and Av1 in solution at pH 8, 6.5, pH 5 (Figure VII-4). From simulations, the effective *g* values and *E/D* ratios were determined (Table VII-2 and Figure VII-S1). In both Cp1 and Av1, low pH conditions induce a second rhombic spin system with higher rhombicity compared to the resting state spin system at physiological pH. Line broadening is also observed in the low pH spectra, reminiscent of the EPR spectra of Cp1 FeMo-cofactor isolated in N-methylformamide. The two spin systems are in equilibrium with each other (Figure VII-4). All low pH EPR changes are reversible in both Cp1 and Av1 (Figure VII-S2). Power sweeps at low and physiological pH on Av1 and Cp1 show similar changes in peak area with change in power, indicating similar relaxation behavior of the resting state and low pH spin systems. Similar relaxation behavior is only expected for spin systems of the same cluster. Thus, the power sweeps demonstrate that both spin systems result from electronic changes at the FeMo-cofactor (Figure VII-S3).

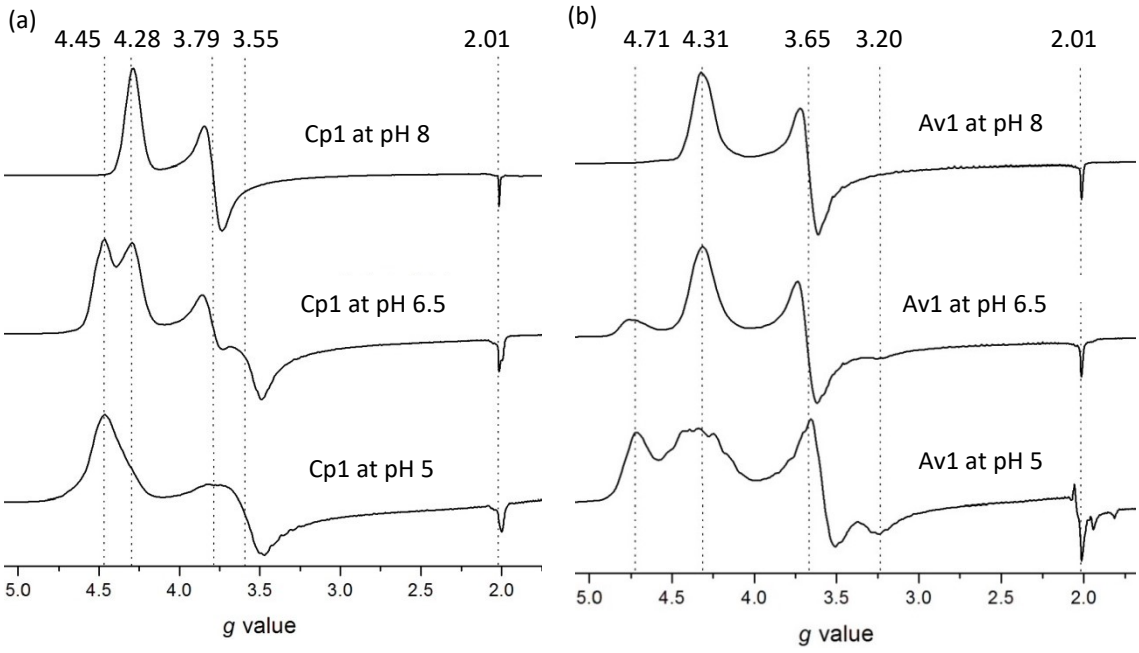


Figure VII-4. (a) Comparison of Cp1 at pH 8, pH 6.5, and pH 5. The same is shown in (b) for Av1.

The Av1 low pH spin state is an S = 3/2 system with zero-field splitting parameters similar to those reported for the 1c spin state. The 1c spin state emerges after 1b, putatively during the accumulation of electrons from the E0 to E2 state.[132] The 1c and 1b spin states form under turnover conditions (ATP regenerating system, Fe protein, and reductant) but without added substrate beyond H^+. 1c was never observed without 1b present,[132] and these signals relax with the same decay constant.[134] Follow-up studies on the EPR of states more reduced than E_0 suggest that 1c is a result of protonation of the FeMo-cofactor.[133]

Table VII-2. Summary of EPR data

Protein		g_{eff} values*	*E/D*	References
Resting state Cp1 at pH 8		4.28, 3.79, 2.01	0.041	This work
		4.29, 3.76, 2.01	Not reported	[146]
Cp1 at pH 6.5		4.28, 3.79, 2.01 (physiological pH spin state)	0.041	This work
		4.45, 3.55, 2.00 (low pH spin state)	0.077	
Cp1 at pH 5		4.45, 3.60, 2.00	0.070	This work
Av1 resting state at pH 8		4.30, 3.65, 2.01	0.053	This work
		4.31, 3.65, 2.01	0.053	[147]
Av1 under turnover conditions at pH 8	1a	4.32, 3.66, 2.01	Not reported	[34,132,134]
	1b	4.21, 3.76, 1.97	Not reported	[34,133,134]
	1b	4.27, 3.73, 2.02	Not reported	[132]
	1c	4.7 or 4.69, ~3.2-3.4, ~2.0	Not reported	[132–134,148]
Av1 at pH 6.5		4.31, 3.67, 2.01 (physiological pH spin state)	0.053	This work
		4.72, 3.30, 2.01 (low pH spin state)	0.124	
Av1 at pH 5		4.32, 3.57, 2.01 (physiological pH spin state)	0.064	This work
		4.71, 3.30, 2.01 (low pH spin state)	0.120	

* Effective *g* values are reported for simulated EPR spectra.

The equivalence between the crystal structure and the protein in solution was accomplished by measuring the EPR spectrum of a solution and polycrystalline protein sample under the same conditions used for the low pH X-ray crystallographic experiments (Figure VII-

S4). The resulting spectra exhibit the same features, thereby confirming that the low pH structural changes observed by X-ray crystallography correlate to the low pH electronic changes observed by EPR spectroscopy.

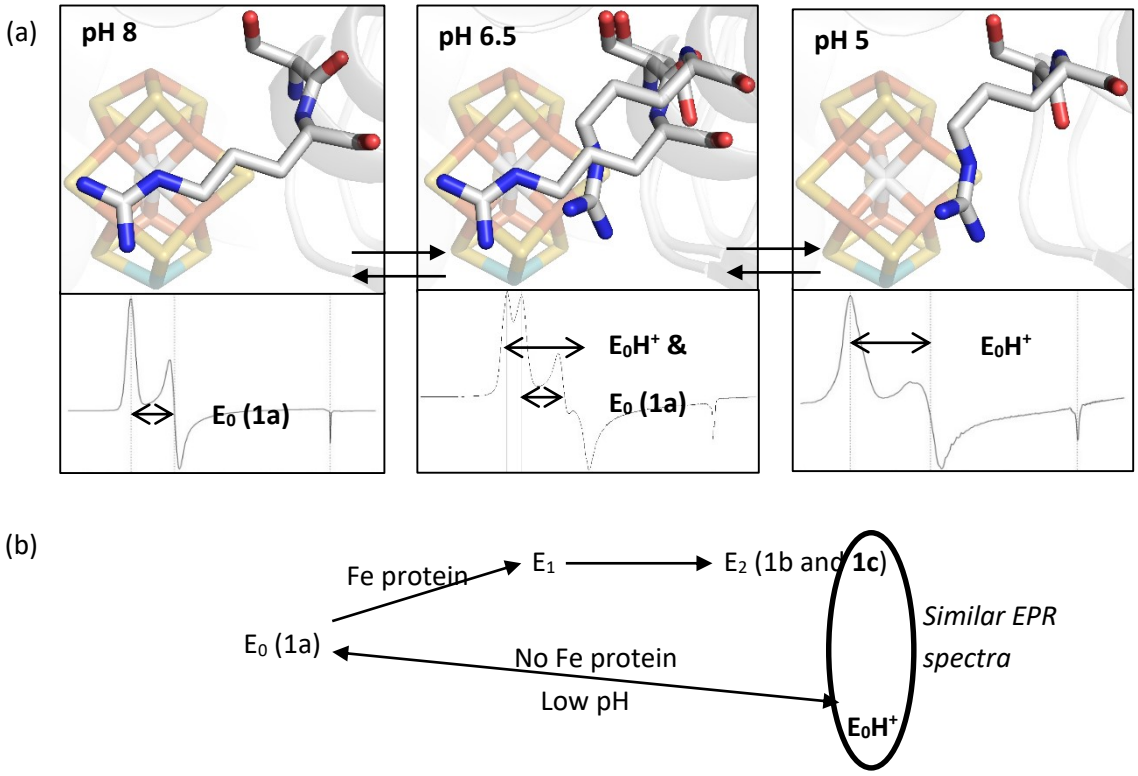


Figure 5. (a) A summary of the Cp1 data presented in this manuscript. At pH 8, the typical resting state X-ray diffraction structure and EPR signal are observed. At pH 5, a peptide flip and repositioning of the α-Arg357 sidechain away from the Fe3,4,5,7 face of the FeMo-cofactor is observed as well as an EPR signal similar to that reported for the E_2 state. At intermediate pH, both structural conformations and EPR states are observed. The EPR signals and X-ray structures are reversible and correlated. (b) The 1c peak has been attributed to the E_2 state and is hypothesized to result from protonation of the FeMo-cofactor. Our experimental conditions include only a proton source and not an electron source, so it is unlikely that these conditions achieve a reduced state, such as E_2. Given the similarity to the 1c peak and our experimental conditions, it is more likely that the resting state is protonated, which we have termed "E_0H^+".

Because our experimental conditions do not include the Fe protein and ATP regenerating system, and because all data obtained with and without dithionite are comparable, a net flow of

electrons to the FeMo-cofactor is likely not occurring in our low pH experimental conditions. Therefore, we conclude that the low pH state is a protonated resting state. We call it "E_0H^+", following the Lowe-Thorneley naming scheme. We would like to emphasize that this name is a generic designation for a protonated form of the resting state; we cannot determine the number of protons added to the FeMo-cofactor at low pH. Figure VII-5 depicts a summary of the relationships between different forms of the resting state, together with their major X-ray and EPR features.

VII.D. Conclusion

The active site of the MoFe protein exhibits pH-dependent structural and electronic rearrangements localized at the active site. The low pH structural rearrangements involve residues α-Arg347 from Cp1 and α-His274 from Av1, both of which participate in hydrogen bond networks with FeMo-cofactor belt sulfurs. The structural and electronic changes are reversible with pH and directly correlated, the latter of which was demonstrated by performing EPR spectroscopy on polycrystalline samples. The low pH Av1 spin system has EPR spectral features similar to that observed for 1c, which is one of the two spin states assigned to E_2. Given the observed structural rearrangements and the absence of net flow of electrons through nitrogenase at low pH without the Fe-protein, we conclude that reversible protonation of the resting state of the FeMo-cofactor occurs at low pH to generate "E_0H^+."

This study demonstrates the advantage of comparing more than one species of nitrogenase MoFe protein, despite having the same cofactor structures, when addressing the mechanism of substrate reduction. This is supported by the fact that the low pH structural and electronic changes of Cp1 and Av1 are similar but not identical: the low pH structural changes occur on the same face of the FeMo-factor but not the same residue, and the low pH spin states

show similar but not identical *g* values and *E/D* ratios. In both Cp1 and Av1, however, the structural data suggest protonation of the resting state may occur at one of the two belt sulfurs that are not replaced by CO or Se, which may facilitate rearrangements of the cofactor during turnover.

VII.E. Experimental

VII.E.i. Cell growth and protein purification

Av1 and Cp1 protein were obtained using cell growth and protein purification procedures previously described.[64,73]

VII.E.ii. Crystallization

Protein crystals were grown in 24-well plates using the sitting-drop method with a 1:1 ratio of protein stock to reservoir solution, at room temperature, and in an anaerobic chamber with an atmosphere of ~95% argon and ~5% hydrogen. All solutions were made anaerobic through a series of vacuum and argon cycles. The reservoir solution for Av1 crystals consisted of double-distilled water, 15% polyethylene glycol (MW 4000 g/mol, Hampton Research), 0.5-0.8 M sodium chloride (VWR), 0.2 M imidazole/malate at pH 8 (Sigma-Aldrich), and 5 mM sodium dithionite (J.T. Baker). The reservoir solution for Cp1 crystals consisted of double-distilled water, 13.5-14% polyethylene glycol (MW 3350 g/mol, Hampton Research), 0.3-0.5 M magnesium chloride (Mallinckrodt), 0.08 M Tris at pH 8 (Fisher Scientific), and 5 mM sodium dithionite. Av1 and Cp1 crystals of block morphology formed overnight.

VII.E.iii. Tri-buffer preparation

A tri-buffer consists of three different buffers, such that the buffering capacity extends over a large pH range while maintaining a nearly constant ionic strength. We created a tri-buffer from 0.05 M glycylglycine (pK_a = 3.14, Acros Organics), 0.05 M acetic acid (pK_a = 4.76, Sigma-Aldrich), 0.10 M Bis-Tris (pK_a = 6.46, Sigma) based off work on by Ellis and Morrison. The tri-buffer was adjusted to pH 6, 5, 4, 3, and 2 using HCl and maintained an ionic strength of ~0.1 M.

VII.E.iv. pH measurements

The pH of solutions surrounding crystals was measured using litmus paper at room temperature. Since the experimental conditions were 100 K (X-ray crystallography) and 4-8 K (EPR), the pH of the samples under cryogenic conditions will likely be greater than measured at room temperature.[149]

VII.E.v. X-ray sample preparation

A low pH solution was made according to the recipe for each well's reservoir solution except that a tri-buffer at low pH was substituted for Tris/HCl at pH 8. 10 µL of low pH solution was added to each well containing crystals as well as 1 µL of 2-methyl-2,4-pentanediol (cryo-protectant, Acros Organics). Also, three drops of Fomblin Y 16/6 mineral oil (Sigma-Aldrich) were added to the top of each crystal drop for additional cryo-protection. Crystals soaked for at least five minutes in the low pH solution before flash freezing in liquid nitrogen on nylon loops. The percentage of protein molecules exhibiting the low pH structural rearrangements was not impacted by soaking duration, provided that the crystals soaked for at least five minutes before freezing. Because the crystal wells contain Tris at pH 8 as part of the crystallization recipe, the actual pH that the crystals soaked in upon addition of low pH buffer was higher than the pH of

the added tri-buffers. To illustrate, in order to soak a crystal at pH 5, tri-buffer at pH 2 must be added to the crystallization well, since Tris at pH 8 is also present. Attempts to transfer crystals from the crystal well to a low pH buffer resulted in crystal cracking.

To check for reversibility in the crystallized state, crystals were soaked at low pH as described for 10 minutes, transferred to a well containing fresh reservoir solution at pH 8, and then flash frozen in liquid nitrogen after soaking for 5 minutes.

VII.E.vi. X-ray data collection and refinement

Diffraction data for Cp1 was collected remotely from the Stanford Synchrotron Radiation Light source (SSRL) on beamline 12-2 with a DECTRIS Pilatus 6 M detector. Reference sets of 1440 diffraction images were collected at 12999.97 eV with an oscillation angle of 0.25° over 360° rotation. Diffraction data for Av1 was collected in-house on a Rigaku MicroMax 007-HF X-ray generator with a Rigaku RAXIS-IV++ detector. All data sets were integrated with the XDS program package.[57] Scaling was carried out with the CCP4 suite,[58] and phasing was determined by molecular replacement against high resolution Av1 (PDB ID 3U7Q) and Cp1 (PDB ID 4WES) structures.[22,64] Initial refinement was carried out with CNS,[59] and alternative conformations and isotropic B-factors were refined with REFMAC5.[60,61] All figures were made in PyMOL.[63]

VII.E.vii. EPR sample preparation

After solubility tests, the following solution was chosen for low pH EPR studies: 100 mM tri-buffer at pH 2, 500 mM $MgCl_2$, and 5 mM sodium dithionite. To prepare the EPR samples, protein stock was concentrated 50% and then diluted with the low pH EPR solution. Samples were allowed to equilibrate for at least 30 minutes prior to freezing in liquid nitrogen. 200 µl of each sample (~30 mg/mL) was transferred to an EPR tube in an anaerobic tent. The samples

were carefully frozen in liquid nitrogen inside the anaerobic tent and then stored in a liquid nitrogen dewar until use.

After obtaining an EPR spectrum of the low pH Av1 and Cp1 samples, the samples were thawed and transferred to physiological pH by repeatedly concentrating the protein solution and then diluting it with the protein storage solution (200 mM NaCl, 50 mM Tris at pH 8, and 5 mM sodium dithionite). EPR spectroscopy was performed on the protein resuspended at physiological pH to check for reversibility.

To test if the structural changes observed by X-ray crystallography are related to changes observed in the solution state by EPR, polycrystalline samples of Av1 and Cp1 were made by collecting crystals from six plates of seeded crystals, crushing the crystals, and transferring them to low pH solutions used for the X-ray studies: (Av1) 15% PEG 4000 g/mol, 0.5 M $MgCl_2$, 0.1 M tri-buffer at pH 2, 5 mM sodium dithionite; (Cp1 partial peptide flip) 14% PEG 3350 g/mol, 0.3 M $MgCl_2$, 0.02 M tri-buffer at pH 2, 5 mM sodium dithionite; (Cp1 complete peptide flip) 13.5% PEG 3350 g/mol, 0.5 M $MgCl_2$, 0.08 M tri-buffer at pH 2, 5 mM sodium dithionite.

VII.E.viii. EPR spectroscopy

EPR spectra were recorded with an X-band Bruker EMX spectrometer equipped with an ER 4119HS cavity. The Bruker Win-EPR software suite version 3.0 was used. Variable temperature experiments were performed with an Oxford (ESR900) helium cryostat (temperature range 4-8 K). All spectra were recorded at 9.37 GHz with a microwave power of 1 mW, a modulation amplitude of 2 G, and a modulation frequency of 100 kHz at 4 K. For the power sweep data, the power was varied from 0.02 mW to 20 mW and the temperature was set to 5 K and 8 K for Av1 and Cp1, respectively. Simulations were performed with the EasySpin software suite (Figure VII-S6).[150] For all simulations, the S = 3/2 real spin system (axial *g*-tensor) and S = 1/2 effective spin

system (rhombic *g*-tensor) were matched to the experimental spectra. From the S = 3/2 model, the *E/D* ratio was determined; from the S = 1/2 model, the effective *g* values were determined. For spectra exhibiting two spin states, simulations were calculated by combining two spin systems with their own *E/D* ratios and *g* values. The relative weight of the spin systems and line widths were varied by inspection. All parameters for the simulations are provided in Table VII-S2.

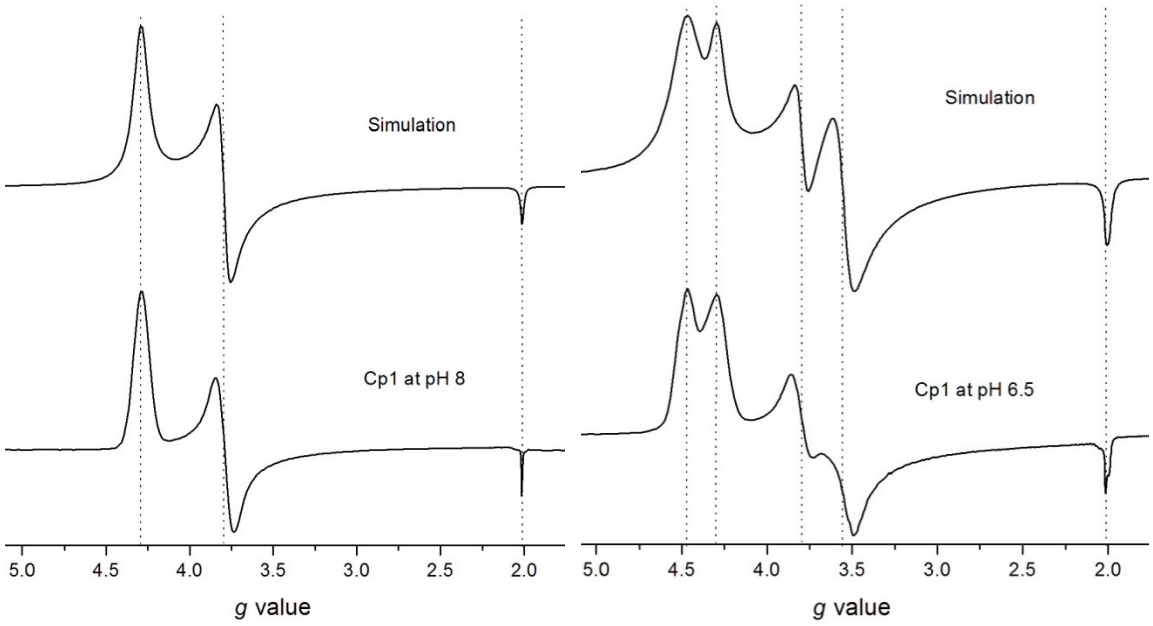
VII.F. Acknowledgement

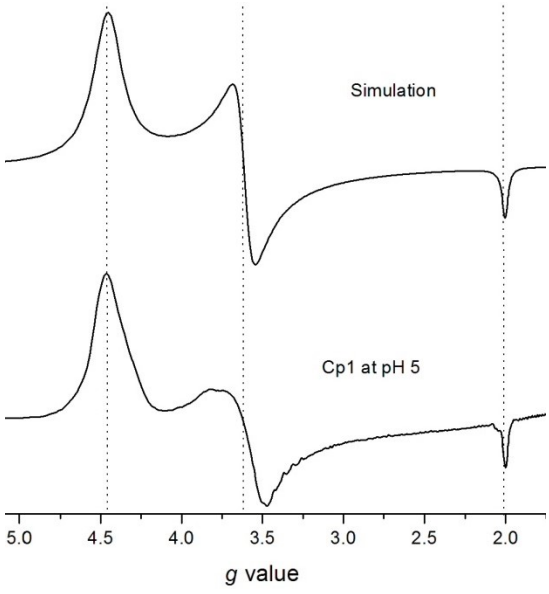
We acknowledge the Gordon and Betty Moore Foundation and the Beckman Institute at Caltech for their generous support of the Molecular Observatory at Caltech. We thank the staff at Beamline 12-2, Stanford Synchrotron Radiation Lightsource (SSRL), operated for the DOE and supported by its OBER and by the NIH, NIGMS (P41GM103393), and the NCRR (P41RR001209). We thank Dr. Angelo Di Bilio for assisting with the EPR experiments; Dr. Lorenz Heidinger for providing a script for EPR simulations; Dr. Paul Oyala for assistance with the EPR simulations and analysis; Dr. Jens Kaiser, Dr. Kathryn Perez, Dr. Helen Segal, Belinda Wenke, and Renee Arias for helpful discussions.

VII.G. Supporting Information

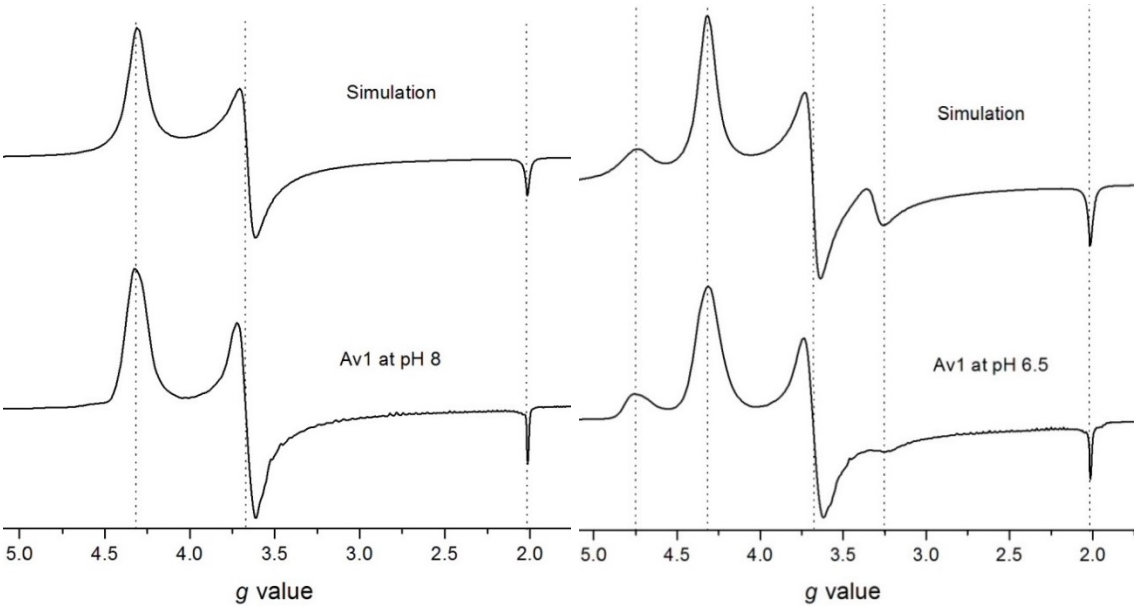
Figure VII-S1. EPR simulations (see Table VII-S1 for simulation parameters)

(a) Cp1 EPR simulations at pH 8, 6.5, and 5





(b) Av1 EPR simulations at pH 8, 6.5, and 5



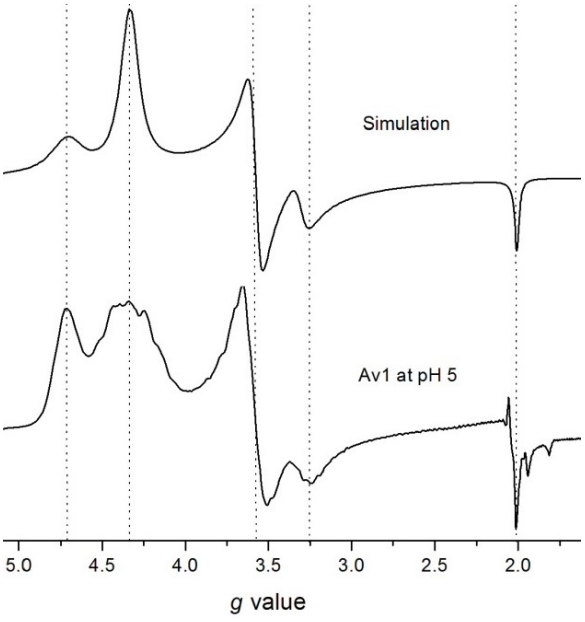
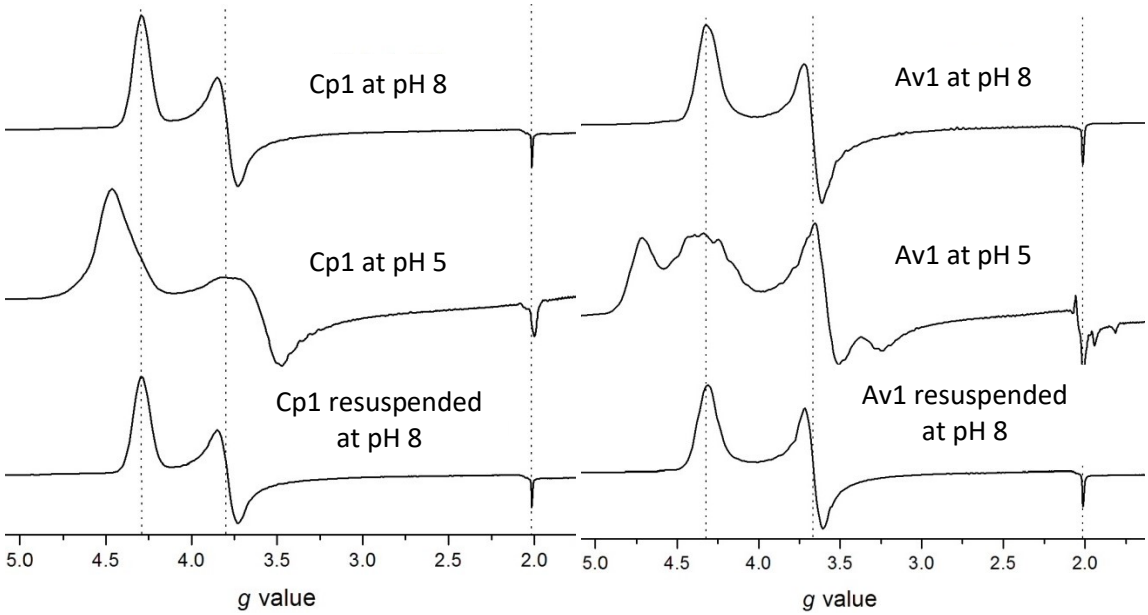


Table VII-S1. Parameters of EPR simulations

	Cp1				**Av1**				
	pH 8	pH 6.5		pH 5	pH 8	pH 6.5		pH 5	
		A	B			A	B	A	B
***S* = 3/2 model**									
E/D	0.041	0.041	0.077	0.07	0.0545	0.053	0.12	0.064	0.12
line width	1, 2	1, 2	1, 4	1, 4	1, 2.5	1, 2.5	1, 4	1, 2.5	1, 4
***S* = 1/2 model**									
g values	4.28 3.79 2.01	4.28 3.79 2.01	4.45 3.55 2.00	4.45 3.60 2.00	4.30 3.65 2.01	4.31 3.67 2.01	4.72 3.30 2.01	4.32 3.57 2.01	4.71 3.30 2.01
weight	n/a	1	4	n/a	n/a	3	1.5	3	1.5

For all simulations, the *S* = 3/2 real spin system (axial *g*-tensor) and *S* = 1/2 effective spin system (rhombic *g*-tensor) were matched to the experimental spectra. To determine the *E/D* ratio from the *S* = 3/2 model, the *E/D* ratio was varied while the *g* values were kept within 1% of [2.00 2.00 2.02]. From the S = 1/2 model, the effective *g* values were determined. For spectra exhibiting two spin states, simulations were calculated by combining two spin systems with their own *E/D* ratios and *g* values. The relative weight of the spin systems and line widths were varied by inspection. (Copied from Experimental Section in the main text.)

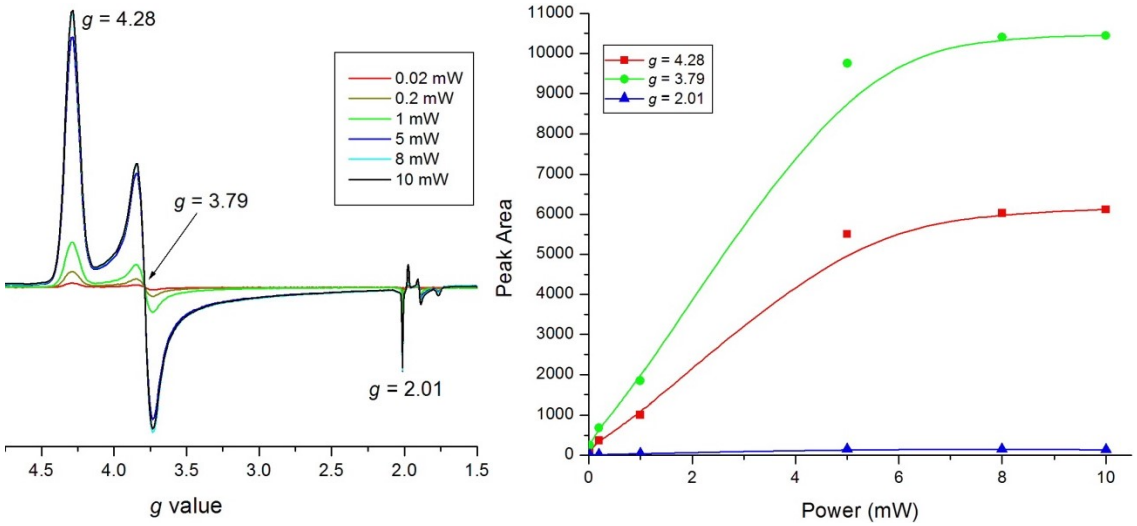
Figure VII-S2. The electronic changes at the FeMo-cofactor observed by EPR at low pH are reversible in Cp1 and Av1



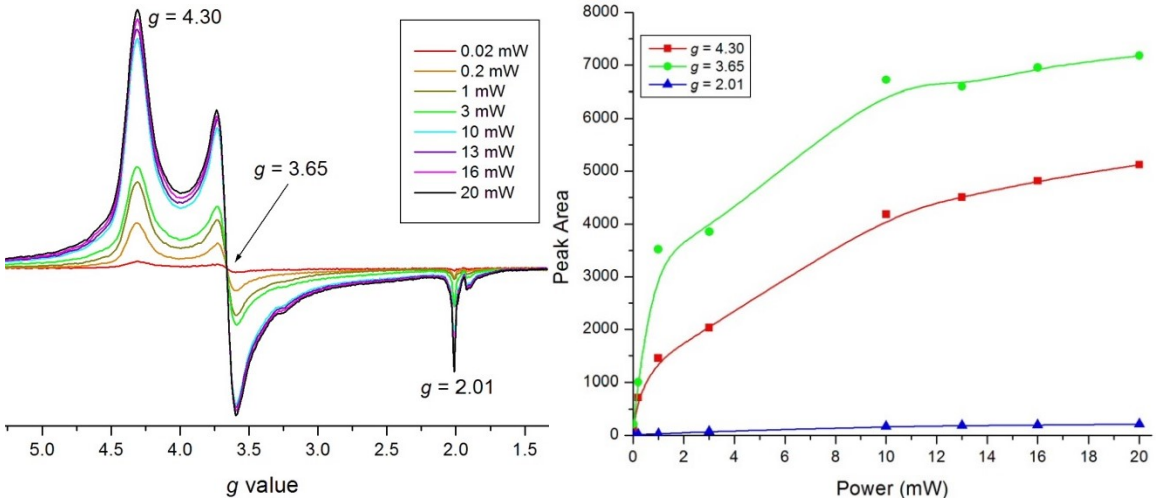
Low pH Av1 and Cp1 samples were resuspended at pH 8. This EPR of the reconstituted protein matches the EPR spectra of native protein, indicating that the electronic changes at the FeMo-cofactor are reversible with pH.

Figure VII-S3. Power Sweep of Cp1 and Av1 at physiological and low pH

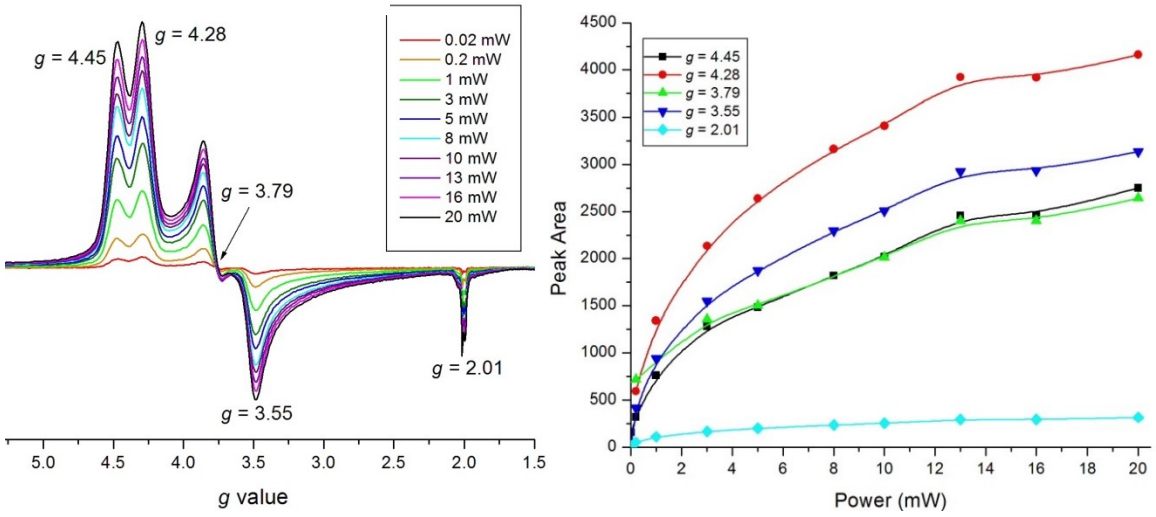
(a) Cp1 Power Sweep at pH 8



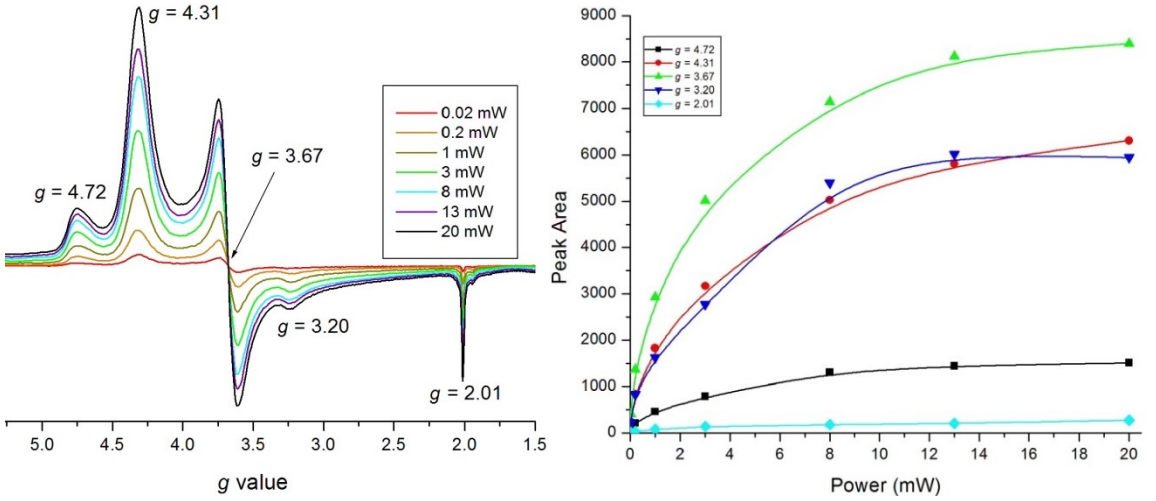
(b) Av1 Power Sweep at pH 8



(c) Cp1 Power Sweep at pH 6.5

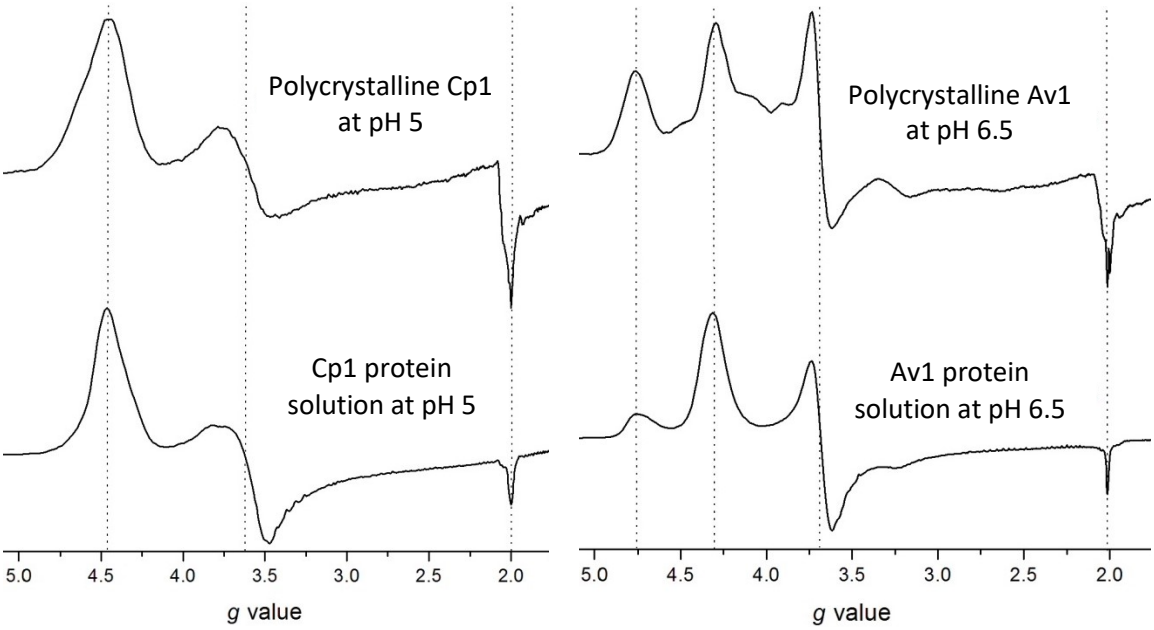


(d) Av1 Power Sweep at pH 6.5



All of the peak area curves show the similar relaxation behavior, indicating that all observed peaks are a result of unpaired spins in the FeMo-cofactor and its environment.

Figure VII-S4. Comparison of polycrystalline and solution protein for Av1 and Cp1 at low pH



The EPR spectra of polycrystalline and solution protein at low pH match well for both Av1 and Cp1, indicating that the low pH structural changes observed by X-ray diffraction are the cause of changes observed in the EPR spectra at low pH.

CONCLUDING PERSPECTIVES

Nitrogenase is a fascinating enzyme. It houses not one but two unique iron-sulfur clusters, the FeMo-cofactor and the P-cluster. Neither cluster has been successfully synthesized outside a cellular environment. Due to the complexity of the clusters' structure and their interactions with substrate, electrons, and protons, the mechanism of biological nitrogen fixation has attracted the attention of numerous scientists for the better part of a century but yet remains elusive. Solving the mechanism of biological nitrogen fixation will not only advance the fields of chemistry and biochemistry at a fundamental level and through the invention of new characterization techniques, it also has the potential to significantly impact agriculture by leading to new and cleaner methods of ammonia production for fertilizer.

The work described in this thesis contributes to our understanding of nitrogenase using structural and spectroscopic techniques. Major results include the structural characterization of a 1.08 Å Cp1 structure and its comparison to Av1, analysis of pathways within the MoFe protein, and the characterization of a reversible protonated resting state of the MoFe protein. An underlying principle in all the work described herein is the importance of investigating multiple nitrogenase species that are phylogenetically distinct. While the structure of the cofactors is conserved among all known MoFe proteins, their enzyme activity and structure are not wholly conserved. Understanding the similarities and differences between nitrogenases elucidates the relationship of the structure and function of nitrogenase.

Looking ahead, much work is still needed to achieve a conclusive and complete understanding of the mechanism of biological nitrogen fixation. New technology and creative approaches will undoubtedly get the job done, and I look forward to seeing the results!

References

1. Lineweaver, H., Burk, D. & Deming, W. E. The Dissociation Constant of Nitrogen-Nitrogenase in Azotobacter. *J. Am. Chem. Soc.* **56,** 225–230 (1934).

2. Burk, D., Lineweaver, H. & Horner, C. K. The Specific Influence of Acidity on the Mechanism of Nitrogen Fixation by Azotobacter. *J. Bacteriol.* **27,** 325–340 (1934).

3. Smil, V. *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production*. (The MIT Press, 2001).

4. Ritter, S. K. The Haber-Bosch Reaction: An Early Chemical Impact On Sustainability. *C&EN* **86,** (2008).

5. Zhang, X. Biogeochemistry: A Plan for Efficient use of Nitrogen Fertilizers. *Nature* **543,** 322–323 (2017).

6. Loxton, S. The Haber Process Nitrogen Fertilizer from the Air. *The Compost Gardener* (2008). Available at: http://www.the-compost-gardener.com/haber-process.html. (Accessed: 3rd August 2017)

7. Anderson, J. S., Rittle, J. & Peters, J. C. Catalytic Conversion of Nitrogen to Ammonia by an Iron Model Complex. *Nature* **501,** 84–87 (2013).

8. Rittle, J. & Peters, J. C. An Fe-N_2 Complex that Generates Hydrazine and Ammonia via Fe═NNH_2: Demonstrating a Hybrid Distal-to-Alternating Pathway for N_2 Reduction. *J. Am. Chem. Soc.* **138,** 4243–4248 (2016).

9. Creutz, S. E. & Peters, J. C. Catalytic Reduction of N_2 to NH_3 by an Fe–N_2 Complex Featuring a C-Atom Anchor. *J. Am. Chem. Soc.* **136,** 1105–1115 (2014).

10. Rittle, J. & Peters, J. C. Fe–N_2/CO Complexes that Model a Possible Role for the Interstitial C Atom of FeMo-cofactor (FeMoco). *Proc. Natl. Acad. Sci. U. S. A.* **110,** 15898–15903 (2013).

11. Withers, N. Nitrogenase: Carbon in the Middle. *Nat. Chem.* **4,** 68 (2012).

12. Ribbe, M. W., Hu, Y., Hodgson, K. O. & Hedman, B. Biosynthesis of Nitrogenase Metalloclusters. *Chem. Rev.* **114,** 4063–4080 (2014).

13. Oldroyd, G. E. D. & Dixon, R. Biotechnological Solutions to the Nitrogen Problem. *Curr. Opin. Biotechnol.* **26,** 19–24 (2014).

14. Beatty, P. H. & Good, A. G. Future Prospects for Cereals that Fix Nitrogen. *Science* **333,** 416–417 (2011).

15. Temme, K., Zhao, D. & Voigt, C. A. Refactoring the Nitrogen Fixation Gene Cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci.* **109,** 7085–7090 (2012).

16. Song, W. J. & Tezcan, F. A. A Designed Supramolecular Protein Assembly with *in vivo*

Enzymatic Activity. *Science* **346,** 1525–1528 (2014).

17. Sontz, P. A., Song, W. J. & Tezcan, F. A. Interfacial Metal Coordination in Engineered Protein and Peptide Assemblies. *Curr. Opin. Chem. Biol.* **19,** 42–49 (2014).

18. Song, W. J., Sontz, P. A., Ambroggio, X. I. & Tezcan, F. A. Metals in Protein–Protein Interfaces. *Annu. Rev. Biophys.* **43,** 409–431 (2014).

19. Rees, D. C. *et al.* in *Molybdenum Enzymes, Cofactors, and Model Systems* **535,** 11–170 (American Chemical Society, 1993).

20. Woo, D. Crystal Structure of the Nitrogenase Iron Protein from *Clostridium pasteurianum*. (University of California, Los Angeles, 1995).

21. Kim, J., Woo, D. & Rees, D. C. X-Ray Crystal Structure of the Nitrogenase Molybdenum-Iron Protein from *Clostridium pasteurianum* at 3.0-Å Resolution. *Biochemistry* **32,** 7104–7115 (1993).

22. Spatzal, T. *et al.* Evidence for Interstitial Carbon in Nitrogenase FeMo Cofactor. *Science* **334,** 940 (2011).

23. Zhang, L. *et al.* The Sixteenth Iron in the Nitrogenase MoFe Protein. *Angew. Chemie Int. Ed.* **52,** 10529–10532 (2013).

24. Howard, J. B., Kechris, K. J., Rees, D. C. & Glazer, A. N. Multiple Amino Acid Sequence Alignment Nitrogenase Component 1: Insights into Phylogenetics and Structure-Function Relationships. *PLoS One* **8,** e72751 (2013).

25. Seefeldt, L. C., Hoffman, B. M. & Dean, D. R. Mechanism of Mo-Dependent Nitrogenase. *Annu. Rev. Biochem.* **78,** 701 (2009).

26. Burgess, B. K. & Lowe, D. J. Mechanism of Molybdenum Nitrogenase. *Chem. Rev.* **96,** 2983–3012 (1996).

27. Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B. & Rees, D. C. Structure of ADP-AlF_4^--Stabilized Nitrogenase Complex and its Implications for Signal Transduction. *Nature* **387,** 370–376 (1997).

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

29. Igarashi, R. Y. & Seefeldt, L. C. Nitrogen Fixation: The Mechanism of the Mo-Dependent Nitrogenase. *Crit. Rev. Biochem. Mol. Biol.* **38,** 351–384 (2003).

30. Duval, S. *et al.* Electron Transfer Precedes ATP Hydrolysis during Nitrogenase Catalysis. *Proc. Natl. Acad. Sci.* **110,** 16414–16419 (2013).

31. Hageman, R. V, Orme-Johnson, W. H. & Burris, R. H. Role of Magnesium Adenosine 5'-Triphosphate in the Hydrogen Evolution Reaction Catalyzed by Nitrogenase from *Azotobacter vinelandii*. *Biochemistry* **19,** 2333–2342 (1980).

32. Thorneley, R. N., Lowe, D. J., Eday, R. R. & Miller, R. W. The Coupling of Electron Transfer in Nitrogenase to the Hydrolysis of Magnesium Adenosine Triphosphate. *Biochem. Soc. Trans.* **7,** 633–636 (1979).

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

34. Lukoyanov, D. *et al.* Identification of a Key Catalytic Intermediate Demonstrates That Nitrogenase is Activated by the Reversible Exchange of N_2 for H_2. *J. Am. Chem. Soc.* **137,** 3610–3615 (2015).

35. Simpson, F. B. & Burris, R. H. A Nitrogen Pressure of 50 Atmospheres does not Prevent Evolution of Hydrogen by Nitrogenase. *Science* **224,** 1095–1097 (1984).

36. Schrock, R. R. Nitrogen Reduction: Molybdenum does it Again. *Nat. Chem.* **3,** 95–96 (2011).

37. Spatzal, T., Perez, K. A., Einsle, O., Howard, J. B. & Rees, D. C. Ligand Binding to the FeMo-Cofactor: Structures of CO-Bound and Reactivated Nitrogenase. *Science* **345,** 1620–1623 (2014).

38. Spatzal, T., Perez, K. A., Howard, J. B. & Rees, D. C. Catalysis-Dependent Selenium Incorporation and Migration in the Nitrogenase Active Site. *Elife* **4,** e11620 (2015).

39. Danyal, K., Dean, D. R., Hoffman, B. M. & Seefeldt, L. C. Electron Transfer within Nitrogenase: Evidence for a Deficit-Spending Mechanism. *Biochemistry* **50,** 9255–9263 (2011).

40. Wilson, P. E., Nyborg, A. C. & Watt, G. D. Duplication and Extension of the Thorneley and Lowe Kinetic Model for *Klebsiella pneumoniae* Nitrogenase Catalysis using a MATHEMATICA Software Platform. *Biophys. Chem.* **91,** 281–304 (2001).

41. Lowe, D. J. & Thorneley, R. N. The Mechanism of *Klebsiella pneumoniae* Nitrogenase Action: Pre-Steady-State Kinetics of H_2 Formation. *Biochem. J.* **224,** 877–886 (1984).

42. Thorneley, R. N. & Lowe, D. J. The Mechanism of *Klebsiella pneumoniae* Nitrogenase Action: Pre-Steady-State Kinetics of an Enzyme-Bound Intermediate in N_2 Reduction and of NH_3 Formation. *Biochem. J.* **224,** 887–894 (1984).

43. Lowe, D. J. & Thorneley, R. N. F. The Mechanism of *Klebsiella pneumoniae* Nitrogenase Action: The Determination of Rate Constants Required for the Simulation of the Kinetics of N_2 Reduction and H_2 Evolution. *Biochem. J.* **224,** 895–901 (1984).

44. Thorneley, R. N. & Lowe, D. J. Nitrogenase of *Klebsiella pneumoniae*. Kinetics of the Dissociation of Oxidized Iron Protein from Molybdenum-Iron Protein: Identification of the Rate-Limiting Step for Substrate Reduction. *Biochem. J.* **215,** 393–403 (1983).

45. Thorneley, R. N. & Lowe, D. J. *Molybdenum Enzymes*. (John Wiley & Sons, 1985).

46. Young, J. P. W. in *Biological Nitrogen Fixation* (eds. Stacey, G., Burris, R. H. & Evans, H. J.) 43–86 (Routledge, Chapman and Hall, 1992).

47. Mallette, M. F., Reece, P. & Dawes, E. A. Culture of *Clostridium pasteurianum* in Defined Medium and Growth as a Function of Sulfate Concentration. *Appl. Microbiol.* **28,** 999–1003 (1974).

48. Rogers, P. & Palosaari, N. *Clostridium acetobutylicum* Mutants that Produce Butyraldehyde and Altered Quantities of Solvents. *Appl. Environ. Microbiol.* **53,** 2761–2766 (1987).

49. Maddox, I. S. *et al.* The cause of 'Acid-Crash' and 'Acidogenic Fermentations' during the Batch Acetone-Butanol-Ethanol (ABE-) Fermentation Process. *J. Mol. Microbiol. Biotechnol.* **2,** 95–100 (2000).

50. Sabra, W., Groeger, C., Sharma, P. N. & Zeng, A.-P. Improved *n*-Butanol Production by a Non-Acetone Producing *Clostridium pasteurianum* DSMZ 525 in Mixed Substrate Fermentation. *Appl. Microbiol. Biotechnol.* **98,** 4267–4276 (2014).

51. Biebl, H. Fermentation of Glycerol by *Clostridium pasteurianum*: Batch and Continuous Culture Studies. *J. Ind. Microbiol. Biotechnol.* **27,** 18–26 (2001).

52. Bahadur, K. & Saroj, K. K. A Study of the Influence of Hydrogen Ion Concentration of the Culture Media o the Formation of Acetone and Butanol by *Clostridium pasteurianum*, *Clostridium butylicus*, and *Clostridium acetobutylicum* using Sucrose as Substrate. *Jpn. J. Microbiol.* **4,** 341–349 (1960).

53. Daesch, G. & Mortenson, L. E. Effect of Ammonia on the Synthesis and Function of the N_2-Fixing Enzyme System in *Clostridium pasteurianum*. *J. Bacteriol.* **110,** 103–109 (1972).

54. Ljones, T. & Burris, R. H. Nitrogenase: The Reaction between Iron Protein and Bathophenanthrolinedisulfonate as a Probe for Interactions with MgATP. *Biochemistry* **17,** 1866–1872 (1978).

55. Seefeldt, L. C., Yang, Z.-Y., Duval, S. & Dean, D. R. Nitrogenase Reduction of Carbon-Containing Compounds. *BBA-Bioenergetics* **1827,** 1102–1111 (2013).

56. Weston, M. F., Kotake, S. & Davis, L. C. Interaction of Nitrogenase with Nucleotide Analogs of ATP and ADP and the Effect of Metal Ions on ADP Inhibition. *Arch. Biochem. Biophys.* **225,** 809–817 (1983).

57. Kabsch, W. XDS . *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

58. CCP4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763 (1994).

59. Brunger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta. Cryst.* **54,** 905–921 (1998).

60. Murshudov, G. N. *et al.* REFMAC 5 for the Refinement of Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

61. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–

255 (1997).

62. Afonine, P. V *et al.* Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

63. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

64. Zhang, L. L.-M., Morrison, C. N., Kaiser, J. T. & Rees, D. C. Nitrogenase MoFe Protein from *Clostridium pasteurianum* at 1.08 Å Resolution: Comparison with the *Azotobacter vinelandii* MoFe Protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71,** 274–282 (2015).

65. Kim, J. & Rees, D. C. Structural Models for the Metal Centers in the Nitrogenase Molybdenum-Iron Protein. *Science* **257,** 1677–1682 (1992).

66. Mayer, S. M., Lawson, D. M., Gormal, C. A., Roe, S. M. & Smith, B. E. New Insights into Structure-Function Relationships in Nitrogenase: A 1.6 Å Resolution X-Ray Crystallographic Study of *Klebsiella pneumoniae* MoFe-Protein. *J. Mol. Biol.* **292,** 871–891 (1999).

67. Schmid, B. *et al.* Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF_4^--Stabilized Structure. *Biochemistry* **41,** 15557–15565 (2002).

68. Emerich, D. W. & Burris, R. H. Interactions of Heterologous Nitrogenase Components that Generate Catalytically Inactive Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **73,** 4369–4373 (1976).

69. Eady, R. R., Smith, B. E., Cook, K. A. & Postgate, J. R. Nitrogenase of *Klebsiella pneumoniae*: Purification and Properties of the Component Proteins. *Biochem. J.* **128,** 655–675 (1972).

70. Winkler, J. R. & Gray, H. B. Could Tyrosine and Tryptophan serve Multiple Roles in Biological Redox Processes? *Philos. Trans. A. Math. Phys. Eng. Sci.* **373,** 20140178 (2015).

71. Hu, Y., Fay, A. W., Lee, C. C., Yoshizawa, J. & Ribbe, M. W. Assembly of Nitrogenase MoFe Protein. *Biochemistry* **47,** 3973–3981 (2008).

72. Lancaster, K. M. *et al.* X-ray Emission Spectroscopy Evidences a Central Carbon in the Nitrogenase Iron-Molybdenum Cofactor. *Science* **334,** 974–977 (2011).

73. Einsle, O. *et al.* Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science* **297,** 1696–1700 (2002).

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

75. Peters, J. W. *et al.* Redox-Dependent Structural Changes in the Nitrogenase P-Clustes. *Biochemistry* **36,** 1181–1187 (1997).

76. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

77. Chen, V. B. *et al.* MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

78. Morrison, C. N., Hoy, J. A., Zhang, L., Einsle, O. & Rees, D. C. Substrate Pathways in the Nitrogenase MoFe Protein by Experimental Identification of Small Molecule Binding Sites. *Biochemistry* **54,** 2052–2060 (2015).

79. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a Century of Ammonia Synthesis Changed the World. *Nat. Geosci.* **1,** 636–639 (2008).

80. Kitano, M. *et al.* Ammonia Synthesis using a Stable Electride as an Electron Donor and Reversible Hydrogen Store. *Nat. Chem.* **4,** 934–940 (2012).

81. Hartwig, J. F. *Organotransition Metal Chemistry: From Bonding to Catalysis*. (Science Books, 2010).

82. Rees, D. C. & Howard, J. B. Nitrogenase: Standing at the Crossroads. *Curr. Opin. Chem. Biol.* **4,** 559–566 (2000).

83. Scott, E. E., Gibson, Q. H. & Olson, J. S. Mapping the Pathways for O_2 Entry Into and Exit from Myoglobin. *J. Biol. Chem.* **276,** 5177–5188 (2001).

84. Radding, W. & Phillips, G. N. Kinetic Proofreading by the Cavity System of Myoglobin: Protection from Poisoning. *BioEssays* **26,** 422–433 (2004).

85. Teeter, M. M. Myoglobin Cavities Provide Interior Ligand Pathway. *Protein Sci.* **13,** 313–318 (2004).

86. Schoenborn, B. P., Watson, H. C. & Kendrew, J. C. Binding of Xenon to Sperm Whale Myoglobin. *Nature* **207,** 28–30 (1965).

87. Duff, A. P. *et al.* Using Xenon as a Probe for Dioxygen-binding Sites in Copper Amine Oxidases. *J. Mol. Biol.* **344,** 599–607 (2004).

88. Johnson, B. J. *et al.* Exploring Molecular Oxygen Pathways in *Hansenula polymorpha* Copper-Containing Amine Oxidase. *J. Biol. Chem.* **282,** 17767–17776 (2007).

89. Kallio, J. P., Rouvinen, J., Kruus, K. & Hakulinen, N. Probing the Dioxygen Route in *Melanocarpus albomyces* Laccase with Pressurized Xenon Gas. *Biochemistry* **50,** 4396–4398 (2011).

90. Whittington, D. A., Rosenzweig, A. C., Frederick, C. A. & Lippard, S. J. Xenon and Halogenated Alkanes Track Putative Substrate Binding Cavities in the Soluble Methane Monooxygenase Hydroxylase. *Biochemistry* **40,** 3476–3482 (2001).

91. McCormick, M. S. & Lippard, S. J. Analysis of Substrate Access to Active Sites in Bacterial Multicomponent Monooxygenase Hydroxylases: X-ray Crystal Structure of Xenon-Pressurized Phenol Hydroxylase from Pseudomonas sp. OX1. *Biochemistry* **50,** 11058–11069 (2011).

92. Svensson-Ek, M. *et al.* The X-ray Crystal Structures of Wild-type and EQ(I-286) Mutant

Cytochrome *c* Oxidases from *Rhodobacter sphaeroides*. *J. Mol. Biol.* **321,** 329–339 (2002).

93. Luna, V. M., Chen, Y., Fee, J. A. & Stout, C. D. Crystallographic Studies of Xe and Kr Binding within the Large Internal Cavity of Cytochrome *ba3* from *Thermus thermophilus*: Structural Analysis and Role of Oxygen Transport Channels in the Heme−Cu Oxidases. *Biochemistry* **47,** 4657–4665 (2008).

94. Luna, V. M., Fee, J. A., Deniz, A. A. & Stout, C. D. Mobility of Xe Atoms within the Oxygen Diffusion Channel of Cytochrome *ba3* Oxidase. *Biochemistry* **51,** 4669–4676 (2012).

95. Darnault, C. *et al.* Ni-Zn-[Fe_4-S_4] and Ni-Ni-[Fe_4-S_4] Clusters in Closed and Open α-Subunits of Acetyl-CoA Synthase/Carbon Monoxide Dehydrogenase. *Nat. Struct. Mol. Biol.* **10,** 271–279 (2003).

96. Wentworth, P. *et al.* Antibody Catalysis of the Oxidation of Water. *Science* **293,** 1806–1811 (2001).

97. Tilton, R. F. *et al.* Computational Studies of the Interaction of Myoglobin and Xenon. *J. Mol. Biol.* **192,** 443–456 (1986).

98. Tilton, R. F., Singh, U. C., Kuntz, I. D. & Kollman, P. A. Protein-Ligand Dynamics. *J. Mol. Biol.* **199,** 195–211 (1988).

99. Montet, Y. *et al.* Gas Access to the Active Site of Ni-Fe Hydrogenases Probed by X-Ray Crystallography and Molecular Dynamics. *Nat. Struct. Mol. Biol.* **4,** 523–526 (1997).

100. Czerminski, R. & Elber, R. Computational Studies of Ligand Diffusion in Globins: *I. Leghemoglobin*. *Proteins Struct. Funct. Bioinforma.* **10,** 70–80 (1991).

101. Smith, D., Danyal, K., Raugei, S. & Seefeldt, L. C. Substrate Channel in Nitrogenase Revealed by a Molecular Dynamics Approach. *Biochemistry* **53,** 2278–2285 (2014).

102. Dance, I. A Molecular Pathway for the Egress of Ammonia Produced by Nitrogenase. *Sci. Rep.* **3,** 3237 (2013).

103. Durrant, M. C. Controlled Protonation of Iron-Molybdenum Cofactor by Nitrogenase: A Structural and Theoretical Analysis. *Biochem. J.* **355,** 569–576 (2001).

104. Barney, B. M., Yurth, M. G., Dos Santos, P. C., Dean, D. R. & Seefeldt, L. C. A Substrate Channel in the Nitrogenase MoFe Protein. *J. Biol. Inorg. Chem.* **14,** 1015–1022 (2009).

105. Dance, I. The Controlled Relay of Multiple Protons Required at the Active Site of Nitrogenase. *Dalt. Trans.* **41,** 7647–7659 (2012).

106. Dance, I. Nitrogenase: A General Hydrogenator of Small Molecules. *Chem. Commun.* **49,** 10893–10907 (2013).

107. Ringe, D., Petsko, G. A., Kerr, D. E. & Ortiz de Montellano, P. R. Reaction of Myoglobin with Phenylhydrazine: A Molecular Doorstop. *Biochemistry* **23,** 2–4 (1984).

108. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in

Dynamic Protein Structures. *PLOS Comput. Biol.* **8,** e1002708 (2012).

109. Hu, Y. & Ribbe, M. W. Nitrogenase Assembly. *Biochim. Biophys. Acta* **1827,** 1112–1122 (2013).

110. Rupnik, K. *et al.* Nonenzymatic Synthesis of the P-Cluster in the Nitrogenase MoFe Protein: Evidence of the Involvement of All-Ferrous $[Fe_4S_4]^0$ Intermediates. *Biochemistry* **53,** 1108–1116 (2014).

111. Schmid, B. *et al.* Structure of a Cofactor-Deficient Nitrogenase MoFe Protein. *Science* **296,** 352–356 (2002).

112. Stowell, M. H. B. *et al.* A Simple Device for Studying Macromolecular Crystals under Moderate Gas Pressures (0.1–10 MPa). *J. Appl. Crystallogr.* **29,** 608–613 (1996).

113. Zondlo, N. J. Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.* **46,** 1039–1049 (2013).

114. Leitgeb, B. & Tóth, G. Aromatic–Aromatic and Proline–Aromatic Interactions in Endomorphin-1 and Endomorphin-2. *Eur. J. Med. Chem.* **40,** 674–686 (2005).

115. Udvardi, M., Brodie, E. L., Riley, W., Kaeppler, S. & Lynch, J. Impacts of Agricultural Nitrogen on the Environment and Strategies to Reduce these Impacts. *Procedia Environ. Sci.* **29,** 303 (2015).

116. George, S. J., Ashby, G. A., Wharton, C. W. & Thorneley, R. N. F. Time-Resolved Binding of Carbon Monoxide to Nitrogenase Monitored by Stopped-Flow Infrared Spectroscopy. *J. Am. Chem. Soc.* **119,** 6450–6451 (1997).

117. Meyer, J. Comparison of Carbon Monoxide, Nitric Oxide, and Nitrite as Inhibitors of the Nitrogenase from *Clostridium pasteurianum*. *Arch. Biochem. Biophys.* **210,** 246–256 (1981).

118. Rivera-Ortiz, J. M. & Burris, R. H. Interactions among Substrates and Inhibitors of Nitrogenase. *J. Bacteriol.* **123,** 537–545 (1975).

119. Sattler, M. *Introduction to Biomolecular NMR Spectroscopy*. (2004).

120. Thiyagarajan, P. *et al.* Design and Expected Performance of a High Resolution Macromolecular Neutron Diffractometer (MaNDi) at the Spallation Neutron Source. *Trans. Am. Crystallogr. Assoc.* **38,** 67–76 (2003).

121. Hainbuchner, M. & Jericha, E. *Bound Coherent Neutron Scattering Lengths*. (2001).

122. Wilson, C. C. Towards Large Molecule Crystallography at ISIS. *Trans. Am. Crystallogr. Assoc.* **38,** 14–25 (2003).

123. Teixeira, S. C. M. *Neutron Crystallography*. (2010).

124. Schoenborn, B. P. Neutron Diffraction Analysis of Myoglobin. *Nature* **224,** 143–146 (1969).

125. Kossiakoff, A. A. & Spencer, S. A. Neutron Diffraction Identifies His57 as the Catalytic Base in Trypsin. *Nature* **288,** 414–416 (1980).

126. Chen, J. C.-H., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. Direct Observation of Hydrogen Atom Dynamics and Interactions by Ultrahigh Resolution Neutron Protein Crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15301–15306 (2012).

127. Kovalevsky, A. Y. *et al.* Direct Determination of Protonation States of Histidine Residues in a 2 Å Neutron Structure of Deoxy-Human Normal Adult Hemoglobin and Implications for the Bohr Effect. *J. Mol. Biol.* **398,** 276–291 (2010).

128. Howard, J. B. & Rees, D. C. How many Metals does it take to Fix N_2? A Mechanistic Overview of Biological Nitrogen Fixation. *Proc. Natl. Acad. Sci.* **103,** 17088–17093 (2006).

129. Münck, E. *et al.* Nitrogenase: VIII. Mössbauer and EPR spectroscopy. The MoFe Protein Component from *Azotobacter vinelandii*. *BBA - Protein Struct.* **400,** 32–53 (1975).

130. Surerus, K. K. *et al.* Mössbauer and Integer-Spin EPR of the Oxidized P-Clusters of Nitrogenase: P^{OX} is a Non-Kramers System with a nearly Degenerate Ground Doublet. *J. Am. Chem. Soc.* **114,** 8579–8590 (1992).

131. Pierik, A. J., Wassink, H., Haaker, H. & Hagen, W. R. Redox Properties and EPR Spectroscopy of the P-Clusters of *Azotobacter vinelandii* MoFe Protein. *Eur. J. Biochem.* **212,** 51–61 (1993).

132. Fisher, K., Newton, W. E. & Lowe, D. J. Electron Paramagnetic Resonance Analysis of Different Azotobacter vinelandii Nitrogenase MoFe-Protein Conformations Generated during Enzyme Turnover: Evidence for *S* = 3/2 Spin States from Reduced MoFe-Protein Intermediates. *Biochemistry* **40,** 3333–3339 (2001).

133. Fisher, K. *et al.* Conformations Generated during Turnover of the *Azotobacter vinelandii* Nitrogenase MoFe Protein and their Relationship to Physiological Function. *J. Inorg. Biochem.* **101,** 1649–56 (2007).

134. Lukoyanov, D. *et al.* A Confirmation of the Quench-Cryoannealing Relaxation Protocol for Identifying Reduction States of Freeze-Trapped Nitrogenase Intermediates. *Inorg. Chem.* **53,** 3688–3693 (2014).

135. Wiig, J. A., Hu, Y. & Ribbe, M. W. Refining the Pathway of Carbide Insertion into the Nitrogenase M-Cluster. *Nat. Commun.* **6,** 8034 (2015).

136. Spatzal, T. *et al.* Nitrogenase FeMoco Investigated by Spatially Resolved Anomalous Dispersion Refinement. *Nat. Commun.* **7,** 10902 (2016).

137. Pham, D. N. & Burgess, B. K. Nitrogenase Reactivity: Effects of pH on Substrate Reduction and Carbon Monoxide Inhibition. *Biochemistry* **32,** 13725–13731 (1993).

138. Ellis, K. J. & Morrison, J. F. Buffers of cConstant Ionic Strength for Studying pH-Dependent Processes. *Methods Enzymol.* **87,** 405–426 (1982).

139. Hayward, S. Peptide-Plane Flipping in Proteins. *Protein Sci.* **10,** 2219–2227 (2001).

140. Fay, A. W., Hu, Y., Schmid, B. & Ribbe, M. W. Molecular Insights into Nitrogenase FeMoco Insertion: The Role of His274 and His451 of MoFe Protein α-Subunit. *J. Inorg. Biochem.* **101,** 1630–1641 (2007).

141. Smith, B. E. *et al.* Exploring the Reactivity of the Isolated Iron-Molybdenum Cofactor of Nitrogenase. *Coord. Chem. Rev.* **185–186,** 669–687 (1999).

142. Yang, K. Y., Haynes, C. A., Spatzal, T., Rees, D. C. & Howard, J. B. Turnover-Dependent Inactivation of the Nitrogenase MoFe-Protein at High pH. *Biochemistry* **53,** 333–343 (2014).

143. Bruice, T. C., Maskiewicz, R. & Job, R. The Acid-Base Properties, Hydrolytic Mechanism, and Susceptibility to O_2 Oxidation of $Fe_4S_4(SR)_4^{-2}$ Clusters. *Proc. Natl. Acad. Sci. U. S. A.* **72,** 231–234 (1975).

144. Bates, K., Garrett, B. & Henderson, R. A. Rates of Proton Transfer to Fe−S-Based Clusters: Comparison of Clusters Containing $[MFe(\mu_2\text{-}S)_2]_n^+$ and $[MFe_3(\mu_3\text{-}S)_4]_n^+$ (M = Fe, Mo, or W) Cores. *Inorg. Chem.* **46,** 11145–11155 (2007).

145. Chen, K. *et al.* Atomically Defined Mechanism for Proton Transfer to a Buried Redox Centre in a Protein. *Nature* **405,** 814–817 (2000).

146. Morgan, T. V, Mortenson, L. E., McDonald, J. W. & Watt, G. D. Comparison of Redox and EPR Properties of the Molybdenum Iron Proteins of *Clostridium pasteurianum* and *Azotobacter vinelandii* Nitrogenases. *J. Inorg. Biochem.* **33,** 111–120 (1988).

147. Spatzal, T., Einsle, O. & Andrade, S. L. A. Analysis of the Magnetic Properties of Nitrogenase FeMo Cofactor by Single-Crystal EPR Spectroscopy. *Angew. Chemie Int. Ed.* **52,** 10116–10119 (2013).

148. Smith, B. E., Lowe, D. J. & Bray, R. C. Studies by Electron Paramagnetic Resonance on the Catalytic Mechanism of Nitrogenase of *Klebsiella pneumoniae*. *Biochem. J* **135,** 331–341 (1973).

149. Good, N. E. *et al.* Hydrogen Ion Buffers for Biological Research. *Biochemistry* **5,** 467–477 (1966).

150. Stoll, S. & Schweiger, A. EasySpin: A Comprehensive Software Package for Spectral Simulation and Analysis in EPR. *J. Magn. Reson.* **178,** 42–55 (2006).