Examination of Selenium Incorporation and Product Formation in the Nitrogenase FeMo-cofactor

Thesis by

Renee Arias

In Partial Fulfillment of the Requirements for
the degree of
Biochemistry and Molecular Biophysics

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 14, 2018

© 2018

Renee J Arias

ORCID: 0000-0002-6505-7513

ACKNOWLEDGEMENTS

A main proponent of my work is of course, Professor Douglas Rees. Doug, I want to sincerely thank you for taking me into your lab, being a caring and supportive advisor, and for ultimately guiding me to achieve the coveted Caltech PhD. I am so lucky to have landed in your lab, and I will never forget my time here. And on the same vein, I would like to thank my committee members Harry Gray, Pamela Bjorkman, and Shu-ou Shan for their support and encouragement over the years.

To Belinda---I never want to embarrass you in front of others---but you are the smartest person I have ever met. I am so thankful to have had you sitting next to me these past few years. I love your sense of humor so much, I really do look forward to seeing you every day just for a laugh! We definitely have given each other sympathy stress, but we have also had so much fun exploring LA, indulging our love of cats, and of course the occasional fancy lady day. I will miss you so much, but I have set up a flight tracker specifically for good Madison, WI flights.

To Andy, you are the second smartest person I have ever met (behind Belinda, of course). I admire your curiosity---in both science and life in general---your inspiration to be a teacher, and your general thoughtful manner. I count myself very lucky to have you as a friend, and I am so glad I got to celebrate yours and Belinda's wedding. Looking forward to the next few years!

To Janet, Selene, Max, and Sam, you are all such beacons of happiness and joy for me. Although we live far apart now, I know we have many more years of watching the kids grow, and growing old friends ourselves.

To Kana, my very first true mentor at Caltech. You are such a professional and personal inspiration for me. You never stopped encouraging me to learn and grow, and I hope to instill that energy into my future mentees and friends.

To Phong, Amy, and Vian, I love seeing your smiling faces every chance I get. You two are some of the most caring and thoughtful people I have ever met. I secretly wish that you will end up in Oregon with us, but will accept the best outcome for you, of course! Congratulations on your beautiful son. I am looking forward to the next three!

To Mom and Dad, you always knew I could do whatever I wanted. You encouraged me when I wanted to become an art historian, and you encouraged me when I ultimately became a scientist. You were always full of reassurance and support, in everything from swim team to graduate school. You have always emphasized learning and working hard, reading books and not watching TV, and being active as much as possible. Bad grades were never really an option for us, not because you were strict, but because you knew we could always do better. Thank you to you both. Love too.

To Caity and Jake, the best brother and sister a person could ask for. I always reflect on our childhood as some of our very best times, with amazing summers at the pool, the ocean, the rivers, the lakes, and of course our own backyard, complete with super-soakers and bare feet. Love you both. See you again in Oregon soon.

To Kathryn and Thomas, where do I start? You two are explorers, which made our time together in LA even more fun. You always wanted to try something new, whether it was out in nature, some new and interesting food, or some exotic place. I hope to visit you soon in Germany and to have many more adventures together.

To Ailiena and Trixia, our overlap time in Team Nitro seemed so short! But I'm so glad we got to work together as much as we did. You both are so ready to work hard and have fun, which I always appreciate in a work environment. We need to squeeze in a few more trips to Din Tai Fung, taco trucks (with our favorite cat, Taco), and Magnolia House.

To Helen, Eric, and Matt, we are scattered across the country now, but we will always convene for those milestones like graduation, weddings, and trivia! Helen, we always watched the best CW shows and drank the best coffee. Never change, Miss Helen. And soon-to-be Mr. Segal, one of my best memories is walking around Pasadena with you and the group playing Pokémon Go. Should we re-download it for nostalgia's sake? Matt, thanks for

the donuts, the pizza, the awesome nights of trivia, and the much-needed gripe sessions. See you in San Francisco, and someday, see you in New York.

To Yuqing and Jasper, from college to grad school to…adulthood? We are so happy to be moving back to Portland for a lot of reasons, but you are definitely at the top of the list. Can't wait to go house shopping with you, can't wait to eat new and amazing foods with you, can't wait to go to more festivals with you, can't wait to relive Oregon with you!

To Camille and Christine, thank you for our fun times in Team Nitro. It wouldn't have been the same without you!

To Dr. Justin Henthorn and Professor Serena DeBeer, collaborating with you has been great and I hope we have many more projects together!

To Dr. Jens Kaiser, who is the most knowledgeable crystallographer I have ever met. You always approach problems with humor and determination, and I thank you for showing me the possibilities of crystallography.

To Phoebe Ray, who most likely does not get thanked enough for all of the work that you do. The Rees lab---and I'm sure most of the labs on the third floor---would not function without you.

To Camilo and the entire custodial staff, thank you so much for putting up with us! We could not get nearly any of our work done without you, and we are truly thankful.

And to save the greatest acknowledgement for last:

To Keith, I don't have enough space in the world to fully express how much I love and respect you, so I will keep it simple. We have gone so far together with mutual love, support, and ambition. You share my determination for a career, to see the world, and to start a family. You are my best friend and my favorite person. I love *you* the most.

ABSTRACT

Nitrogenase is the only known enzyme to convert the triply bonded atmospheric dinitrogen (N_2) to bioavailable ammonia (NH_3) in an ambient environment, breaking one of the strongest chemical bond in nature in the process. Industrially, the Haber-Bosch process is also capable of reducing dinitrogen to ammonia, and is essential for worldwide food production [1,2]. Due to the high temperatures and pressures required for the Haber-Bosch process (between 300-550ºC and 15-25 MPa) and its requirement for molecular hydrogen, it has become paramount to scientifically investigate the biological processes of nitrogen fixation to ultimately develop more efficient methods to produce bioavailable ammonia. Nitrogenase utilizes two component proteins, the Fe-protein and the MoFe-protein, to reduce ammonia in an ATP-hydrolysis dependent and electron-intensive reaction. Besides the canonical dinitrogen reduction reaction, nitrogenase can reduce a variety of other substrates including: acetylene (C_2H_2), carbon dioxide (CO_2), carbon monoxide (CO), carbonyl sulfide (COS), nitrous oxide (N_2O), diazene (N_2H_2), and more [3–11]. CO has long been of interest to the study of the mechanism of nitrogenase, owing to its isoelectronic identity to N_2, and its potent inhibitor properties at well as its ability to serve as a weak substrate [12,13]. Like CO, cyanide compounds (X-CN) are also of interest to the study of nitrogenase due to the isoelectronic nature of CN^- to N_2. However, cyanide compounds serve as particularly interesting spectroscopic and crystallographic tools, because X in X-CN can be substituted for more significant sulfur or selenium (Se). In this study, we investigate the substrate properties of $SeCN^-$, with Se-incorporation into the active site FeMo-cofactor and concurrent reduction of $SeCN^-$ to methane (CH_4). This study serves as yet another link between substrate reduction in nitrogenase. Part of this work describes the incorporation of Se into the cofactor as a vehicle for high-resolution study of nitrogenase under turnover using spectroscopy and crystallography, while another part describes a proposal for future work on the trapping of enzyme intermediates by fast-growing crystallography.

PUBLISHED CONTENT AND CONTRIBUTIONS

Protocols for the Crystallization of Nitrogenase. Wenke BB*, Arias RJ*, Spatzal T. *Complex Metalloproteins, Methods in Molecular Biology*. Springer Nature. In press as of April 2018. J.K.B.C.

*These authors contributed equally to this work.

R.J.A. participated in the research, organization, and writing of this manuscript.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Table Number *Page*

CHAPTER 1: INTRODUCTION

Brief

Nitrogenase is the only known enzyme to convert the triply bonded atmospheric dinitrogen to bioavailable ammonia in an ambient environment, breaking one of strongest chemical bonds in nature in the process. Approximately 1×10^{14} g of nitrogen are fixed annually via biological processes from the available atmospheric reservoir of 4×10^{21} g, a process which before the development of the Haber-Bosch process in 1913 was the only large scale pathway to fix nitrogen [1,2]. The Haber-Bosch process has since been attributed as the agricultural stimulus driving the incredible increase in world population by more than 5 billion people over the past century [14]. Due to the high temperatures and pressures required for the Haber-Bosch process (between 300-550º C and 15-25 MPa) and its requirement for molecular hydrogen, it has become paramount to scientifically investigate the biological processes of nitrogen fixation to ultimately develop more effective methods to produce bioavailable ammonia.

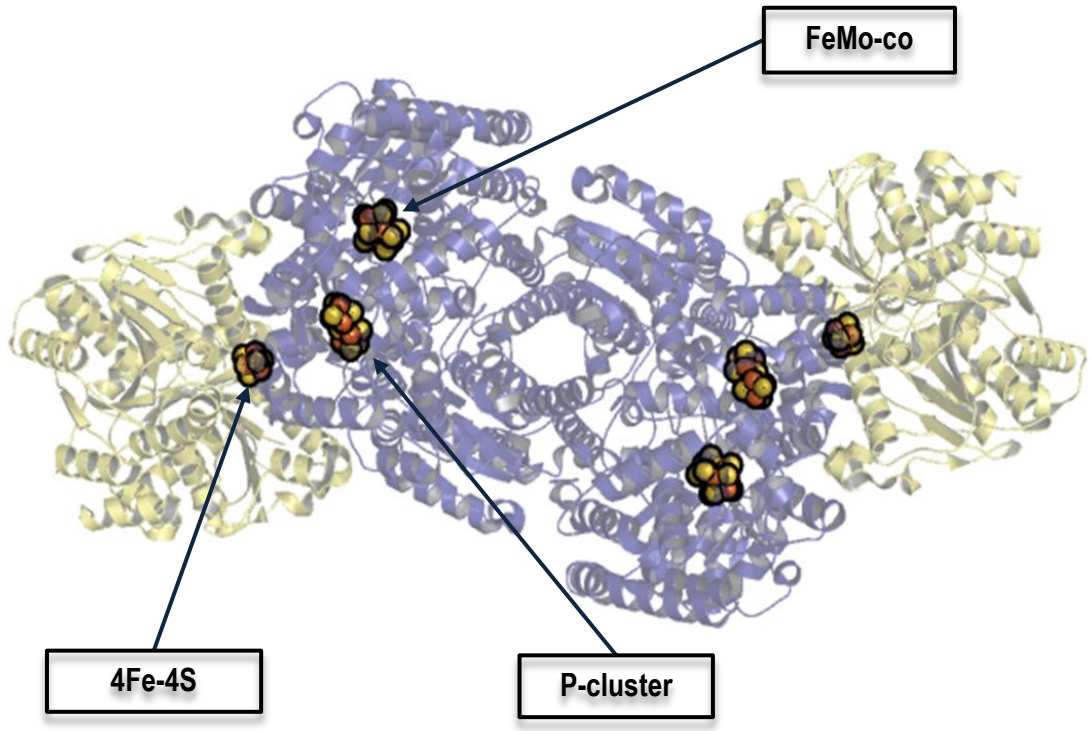


Figure 1: Structure of nitrogenase enzyme from *Azotobacter vinelandii*. Image adapted from PDB: 1N2C, Schindelin, *et al.*, *1997* [15]. The MoFe-protein is shown in blue. The Fe-protein is shown in yellow. As indicated by the arrows, the FeMo-cofactor, P-cluster, and 4Fe-4S clusters are shown in orange and yellow spheres. (Part of an alpha subunit has been removed to show the 4Fe:4S cluster.) Resolution is 3 Å.

Nitrogenase catalyzes the reaction:

$$N_2 + 8H^+ + 8e^- + 16Mg{\bullet}ATP \rightarrow 2NH_3 + H_2 + 16Mg{\bullet}ADP + 16P_i \ ,$$

while the Gibbs free energy of formation for NH_3 is favorable:

$$\Delta G_f^\circ (NH_3)_{gas} = -16 \frac{kJ}{mol} \text{ at 298 K}$$

and even more favorable for the aqueous form:

$$\Delta G_f^{\circ}(NH_3)_{aq} = -26\frac{kJ}{mol}.$$

The activation energy for the above reaction is a staggering $230-420\frac{kJ}{mol}$ [16]. Kinetically speaking, nitrogen fixation is very difficult, especially considering the extremely stable bond energy of dinitrogen, at $941\frac{kJ}{mol}$ for the N≡N triple bond at 298K [17]. Capturing intermediates has also proven to be difficult, since they are often transient and unstable, and can only be achieved with the appropriate amount of electrons transferred to the FeMo-cofactor, the presence of the Fe-protein, and ATP.

The first structure to show a ligand-bound nitrogenase active site was one with a carbon monoxide (CO) inhibited FeMo-cofactor [18]. CO serves as a potent inhibitor for nitrogenase studies because it readily binds to the cofactor under turnover conditions, and is isoelectronic to the physiological N_2 substrate. As observed in the study, the S2B belt sulfur atom becomes reversibly displaced by the CO ligand. The holding place for the displaced sulfur anion is still under investigation.

Another interesting report from George *et al.*, *1997*, showed spectroscopically via stopped-flow Fourier transform infrared spectroscopy (SF-FTIR) that several transient FeMo-cofactor-bound CO species exist in working nitrogenase [20]. These spectroscopic and other kinetic studies indicate that known ligands of nitrogenase may not all bind to a common site on the cofactor [10,21]. Understanding the manipulation of substrates by the cofactor under turnover will provide essential information on nitrogenase function and mechanism. Observation of substrate coordination under turnover is also predicted to provide insight into potential access routes to the FeMo-cofactor.

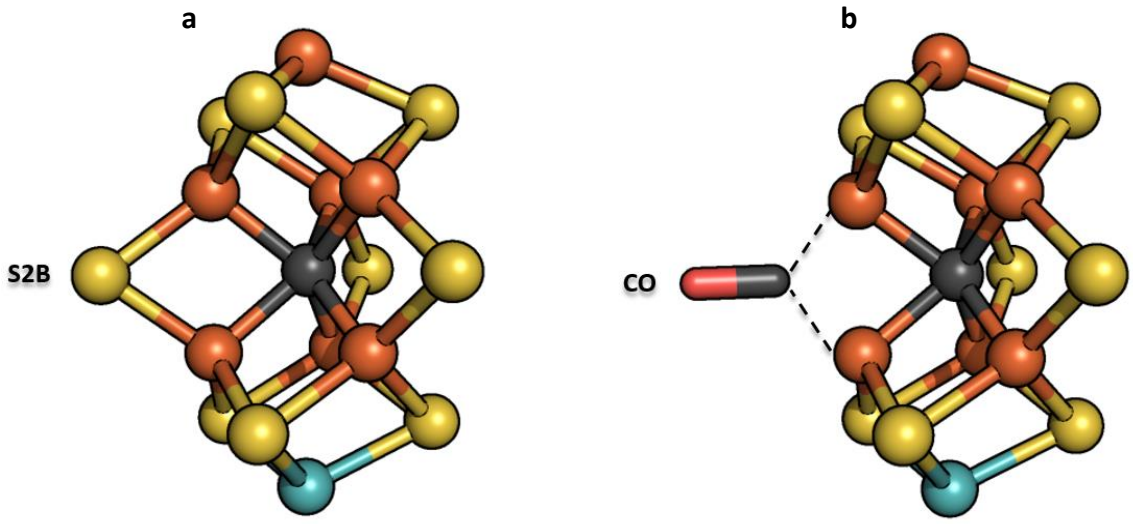


Figure 2: The CO-inhibited FeMo-cofactor. Image adapted from PDB: 4TKV and 3U7Q, Spatzal, *et al., 2014, 2011,* [18,19]. a) PDB: 3U7Q. The resting, "as-isolated" state of the FeMo-cofactor. Resolution: 1.00 Å. b) PDB: 4TKV. Structure of the CO-inhibited FeMo-cofactor. Resolution 1.5 Å. Crystals of *Azotobacter vinelandii* nitrogenase were grown from active assay mixtures of acetylene, MoFe-protein (Av1), and Fe-protein (Av2). After establishment of acetylene reduction activity, the assay mixture was saturated with CO, and complete inhibition was confirmed by absence of ethylene production. Crystals were then grown from the CO-inhibited protein mixture.

An appealing method for observing ligand binding on the FeMo-cofactor is to incorporate a nonmetal probe system directly in the active site. Spatzal, Perez *et al., 2015*, showed that it was possible to incorporate selenium into the cofactor at the S2B site, effectively serving as a crystallographic and spectroscopic probe [22]. Under turnover conditions, the selenium was observed to migrate through the cofactor to the other belt sulfur positions (3A and 5A), before completely migrating out of the protein. We will investigate the structural, spectroscopic, and energetic rearrangements required to incorporate Se into the cofactor using this probe system.

Nitrogen Availability: Anthropogenic Contributions and the Nitrogen Cycle

Nitrogen---like water---is an essential element for life. It is intercalated into almost every aspect of a living organism, from proteins to nucleic acids to carbohydrates. Comprising 78% of the atmosphere, nitrogen is also one of the most abundant elements on earth. Intuitively, one may assume that one of the most abundant elements on earth would also be the most widely available for biological organisms. This is however not the case, since only a few bacteria and archaea over evolutionary history have evolved the capability to convert atmospheric dinitrogen to reactive, bioavailable nitrogen. Many ecosystems and human agriculture are undeniably reliant on and are limited by nitrogen.

Reactive nitrogen availability has been heavily influenced by the anthropogenic activities of food and energy production [23]. Crop cultivation thrives on the symbioses of legumes with nitrogen-fixing organisms, and this type of biological nitrogen fixation was stimulated a great deal by anaerobic environments promoted by rice agronomy. Energy production---specifically the combustion of fossil fuels---releases nitrogen into the atmosphere as NO, a waste product from the oxidation of atmospheric dinitrogen or organic nitrogen as fuel [24,25]. The necessity of nitrogen fixation for food production became evident in the early 20th century, eventually leading to the development of the Haber-Bosch process in 1913 [14]. Industrial nitrogen fixation has now surpassed natural fixation processes, with industrial fertilizer production yielding about 140 Tg N per year, and natural processes producing about 110 Tg N per year [26]. About half of global nitrogen fixation is now composed of nitrogen from anthropogenic activities [26,27].

The Haber-Bosch Process

Fixation by the Haber-Bosch process occurs via an iron catalyst at high temperatures and pressures (300-500° C and 15-25 MPa, respectively). In stark contrast is biological nitrogen fixation, which occurs in ambient conditions (25° C and 0.1 MPa). Further understanding of the biological process is necessary to curtail the detrimental consequences of the Haber-Bosch process such as leaching of byproducts into soils, release of greenhouse gases, and ocean accumulation.

There are numerous variations to the Haber-Bosch process for production of ammonia, but all follow a common circuit. The majority involve an Fe-O precursor catalyst (most use Fe_3O_4), 15 MPa pressure, and 400-475° C temperatures. A summary of the development of ammonia synthesis catalysts is shown in Table 1 [28].

Although the precursors, promoters, and often the feedstock for ammonia synthesis via the Haber-Bosch differ, the general procedures are essentially the same as those developed by Haber and Bosch more than a century ago. First, synthesis gas containing molecular hydrogen and carbon monoxide is produced by the reaction of steam with solid fuels. Oxygen rich air is combusted with the fuels, and the residual nitrogen remains in the gas for subsequent ammonia synthesis. The synthesis gas is then purified of sulfur- and carbon-containing compounds with the use of various solvents. Following purification, the synthesis gas---now composed of nitrogen and hydrogen---is compressed to 10-30 MPa. Efficiency of conversion is very low during compression, so the gas must be compressed several times. Lastly, the synthesis gas is introduced into a hot reactor to form ammonia over the relevant catalyst and is subsequently cooled.

The exact mechanism of ammonia formation by the Haber-Bosch is not completely understood. Nitrogen is thought to adsorb to the iron catalyst surface and undergo early bond breakage, where then the solid adsorbed nitrogen atoms are reduced to ammonia by hydrogen. This type of mechanism can conceivably be applied to understanding biological nitrogen fixation.

Table 1: History and development of ammonia synthesis catalysts

			Chemical component/%wt	Activity[1]/%NH_3			
Country	**Date**	**Precursor**	**Promoters**	**400°C**	**425°C**	**450°C**	**475°C**
Germany	1913	Fe_3O_4	$Al_2O_3 + K_2O$	-	-	-	-
China	1958	Fe_3O_4	$Al_2O_3 + K_2O + CaO$	11.46	13.46	15.59	15.76
United States	1960s	Fe_3O_4	$Al_2O_3 + K_2O + CaO$	11.86	14.39	15.73	15.76
Denmark	1964	Fe_3O_4	$Al_2O_3 + K_2O + CaO$	12.21	14.84	16.12	15.89
UK	1966	Fe_3O_4	$Al_2O_3 + K_2O + CaO + MgO$	-	-	-	-
Denmark	1966	Fe_3O_4	$Al_2O_3 + K_2O + CaO + MgO$	-	-	-	-
China	1969	Fe_3O_4	$Al_2O_3 + K_2O + CaO + MgO$	11.75	14.39	16.39	16.00
China	1979	Fe_3O_4	$Al_2O_3 + K_2O + CaO$	12.80	15.34	16.45	16.09
UK	1979	Fe_3O_4	$Al_2O_3 + K_2O + CaO + CoO$	13.73	16.06	16.99	16.31
China	1984	Fe_3O_4	$Al_2O_3 + K_2O + CaO + CoO$	12.77	14.67	15.69	-
China	1994	Fe_3O_4	$Al_2O_3 + K_2O + CaO + CoO + Ce_2O_3$	13.60	15.14	16.40	-
China	1992	$Fe_{1-x}O$	$Al_2O_3 + K_2O + CaO + \ldots$	15.20	17.59	18.12	16.97
China	1998	$Fe_{1-x}O$	$Al_2O_3 + K_2O + CaO + \ldots$	16.68	18.50	19.15	17.36
UK/Japan	1992	Ru/AC	Ba-K				

[1]Activity testing under 15MPa and space velocity 3 x 10^4 h^{-1} with 75% H_2 and 25% N_2. Fe_3O_4-based catalysts are shown in purple, $Fe_{1-x}O$-based catalysts in orange, and Ru-based catalysts in black. Adapted from [28].

The Nitrogen Cycle

The nitrogen cycle consists of many well-studied metabolic transformations (Figure 3), the main processes being fixation, nitrification, anammox, and denitrification [29].

Nitrogen Fixation

As mentioned previously, the triple bond present in N_2 is extremely stable and requires large amounts of energy to break. Only a select group of prokaryotes are capable of carrying out the energetically taxing process; a process requiring eight electrons and at least 16 ATP molecules when coupled to obligatory H_2 evolution:

$$N_2 + 8H^+ + 8e^- \rightarrow 2NH_3 + H_2 \ .$$

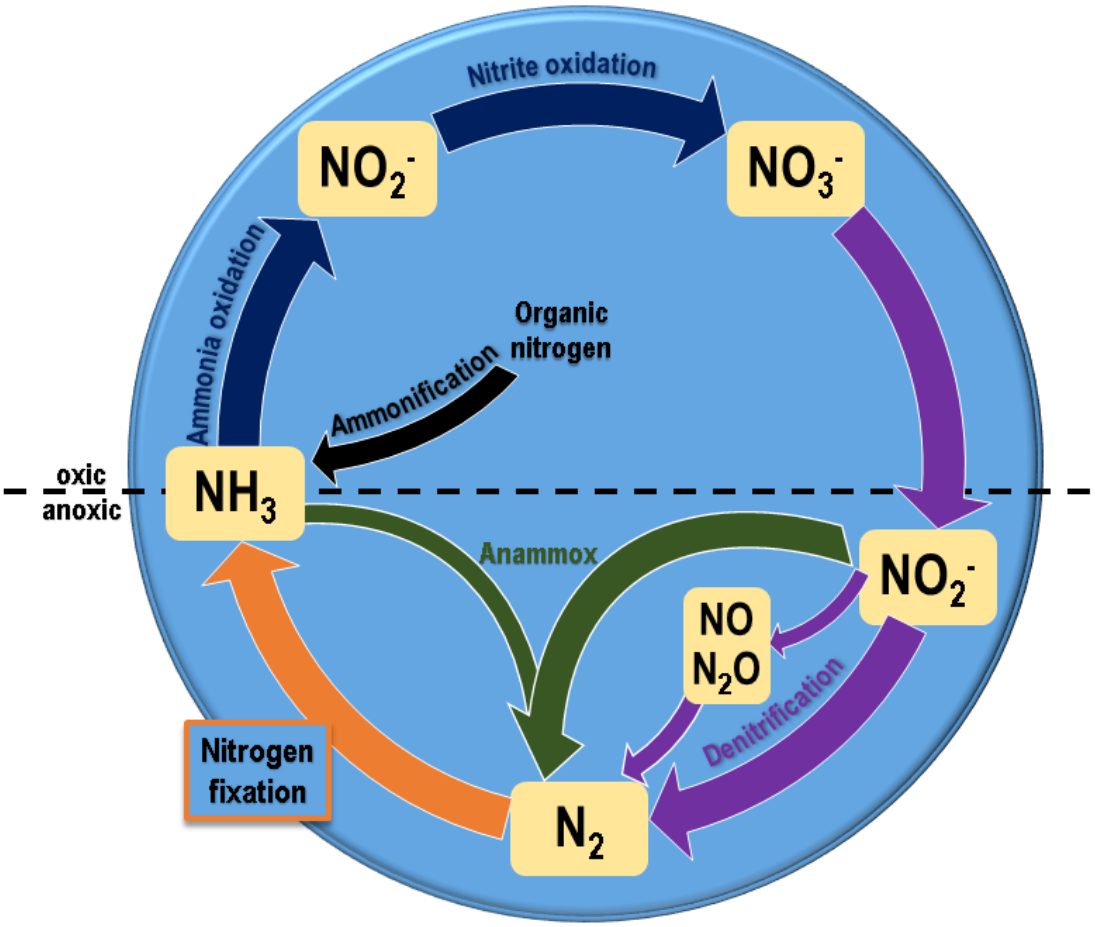


Figure 3: Nitrogen cycle schematic. Adapted from Bernhard, A. (2010) [29].

Most organisms obtain their nitrogen directly as NH_4^+ from the process of nitrogen fixation, or through assimilatory nitrate reduction of NO_3^- to NH_4^+. Organisms capable of fixing nitrogen are diazotrophs, which can exist in free-living or symbiotic forms. The most widespread nitrogen fixers on earth are cyanobacteria (a subset of free-living diazotrophs), which exert most of their influence on freshwater and marine systems, but also grow and fix nitrogen in terrestrial environments [30]. Nitrogen fixation by aquatic cyanobacteria depend on the ratio of nitrogen to phosphorous [31]. When the ratio is low, certain species of cyanobacteria dominate the planktonic community and fix nitrogen. Some ecosystems do not depend on phosphorous, but are still limited by nitrogen. Benthic cyanobacteria (cyanobacteria inhabiting the lowest level of a body of water) exhibit one of the highest rates of nitrogen

fixation, thought to more readily fix nitrogen due to the reduction of turbulence and grazing as compared to that observed with estuarial cyanobacteria. Terrestrial cyanobacteria on the other hand are thought to inhabit extreme environments unsuitable to vascular plants, such as dry or cold regions like deserts, grasslands, and tundra [30]. They are also found to inhabit other areas, but more extreme environments tend to increase their spatial range and magnitude of nitrogen fixation.

Nitrogen fixation in terrestrial ecosystems is most heavily dependent on symbioses involving legumes and rhizobia. One model put forward by Vitousek and Field, *1999*, suggests that nitrogen input may occur via precipitation, and output may occur via leaching and denitrification of excess available nitrogen [32]. Also thought to be important to terrestrial nitrogen fixation are legumes in tropical forests. The abundance of legumes in these regions correlates heavily with the abundance of nitrogen, which may also serve as an indicator of the high nitrogen usage of these organisms.

Besides the free-living diazotrophs that are cyanobacteria, another subset of bacteria also fixes nitrogen. The bacterium *Clostridium pastuerianum* fixes nitrogen anaerobically, while soil bacteria *Klebsiella pneumoniae* and *Azotobacter vinelandii* fix nitrogen aerobically.

Nitrification

In the presence of oxygen, NH_4^+ is oxidized to hydroxylamine (H_2NOH) by specific groups of archaea and bacteria containing the ammonium monooxygenase enzyme. H_2NOH is then oxidized to NO_2^- by hydroxylamine oxidoreductase, and then NO_2^- is finally oxidized to NO_3^- by nitrite oxidoreductase. The reactions for ammonium are as follows:

$$2NH_4^+ + 3O_2 \rightarrow 2NO_2^- + 2H_2O + 4H^+$$

$$2NO_2^- + O_2 \rightarrow 2NO_3^- + 2H_2O + 4H^+ .$$

The reactions for ammonia are:

$$NH_3 + O_2 \rightarrow NO_2^- + 3H^+ + 2e^-$$

$$NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ + 2e^- .$$

Some organisms capable of nitrification are Nitrosomonas, Nitrobacter, and Nitrospira. Complete conversion of ammonia into nitrate by a single microorganism is referred to as Comammox, for which some organisms within the Nitrospira are capable. An important by-product of nitrification is N_2O, an atmospheric greenhouse gas [33].

Denitrification

In the absence of oxygen, NO_3^- can be used as a respiratory electron acceptor by many microbes to eventually yield N_2 gas. Bacteria, archaea, and some eukaryotes are capable of denitrification, and the key enzymes are dissimilatory nitrate reductase, nitrite reductase, nitric oxide reductase, and nitrous oxide reductase. The complete denitrification process is shown as follows:

$$2NO_3^- + 10e^- + 12H^+ \rightarrow N_2 + 6H_2O .$$

Typical environments for denitrification include groundwater, soil, wetlands, and seafloor sediments, where oxygen consumption exceeds oxygen supply and sufficient quantities of nitrate are present. N_2O also serves as an intermediate in this process, with some escaping to the atmosphere as another source of the greenhouse gas [34].

Anammox

Anammox is an alternative route from fixed nitrogen to N_2, and occurs in a group of bacteria called Planctomycetes. Here, NH_4^+ oxidation is coupled to NO_2^- reduction in an exergonic reaction used for chemoautotrophic growth [35]. The anammox reaction is as follows:

$$NH_4^+ + NO_2^- \rightarrow N_2 + 2H_2O .$$

Denitrification and anammox close the nitrogen cycle and bring N_2 gas back to the atmosphere.

Nitrogenase: An Enzyme for Accessible, Bioavailable Nitrogen

Here, we focus primarily on the nitrogenase enzyme utilized by the free-living diazotroph *Azotobacter vinelandii*. Nitrogenase is composed of two component metalloproteins that catalyze the ATP-dependent reduction of nitrogen to ammonia: the iron protein (Fe) and the molybdenum iron protein (MoFe), which in *Azotobacter vinelandii* are also referred to as *Av2* and *Av1*, respectively.

The Fe-protein is a homodimer containing two ATP binding sites, and couples ATP hydrolysis to electron transfer. In a solvent-exposed cleft of the Fe-protein resides the [4Fe:4S] cluster (Figure 4), thought to transfer electrons from the Fe-protein to a second metallocluster, the P-cluster. The P-cluster---an [8Fe:7S] entity---resides in an interface between a pair of α- and β-subunits of the MoFe-protein. The P-cluster accepts electrons donated from the Fe-protein during complex formation, and transfers them to the FeMo-cofactor during turnover (Figure 5) [15,36,37].

Multiple associations and dissociations of the Fe-protein and concomitant ATP hydrolysis are required to generate the reduced form---by two to four electrons---of the MoFe-protein relative to the resting state. Only this reduced form of the MoFe-protein can then bind dinitrogen and proceed through the rest of the proposed 8-electron cycle to generate bioavailable ammonia [38,39]. Other substrates of nitrogenase are also known, and are proposed to bind the active site in a similar fashion. The MoFe-protein is an $\alpha_2\beta_2$ heterotetramer with one active site per αβ heterodimer (Figure 1) [15,40].

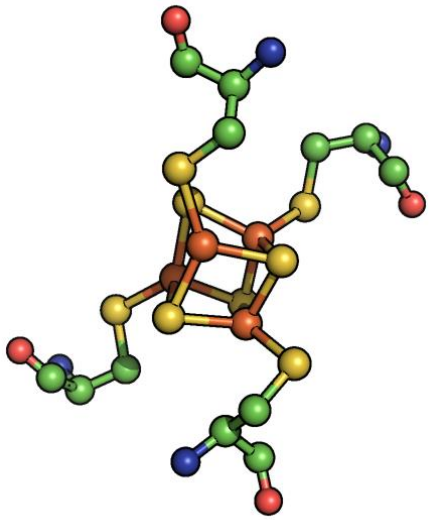


Figure 4: *Azotobacter vinelandii* 4Fe-4S cluster of nitrogenase. Coordinating thiols from Cys-97 and Cys-132 of each subunit are shown in green. PDB: 1NIP.

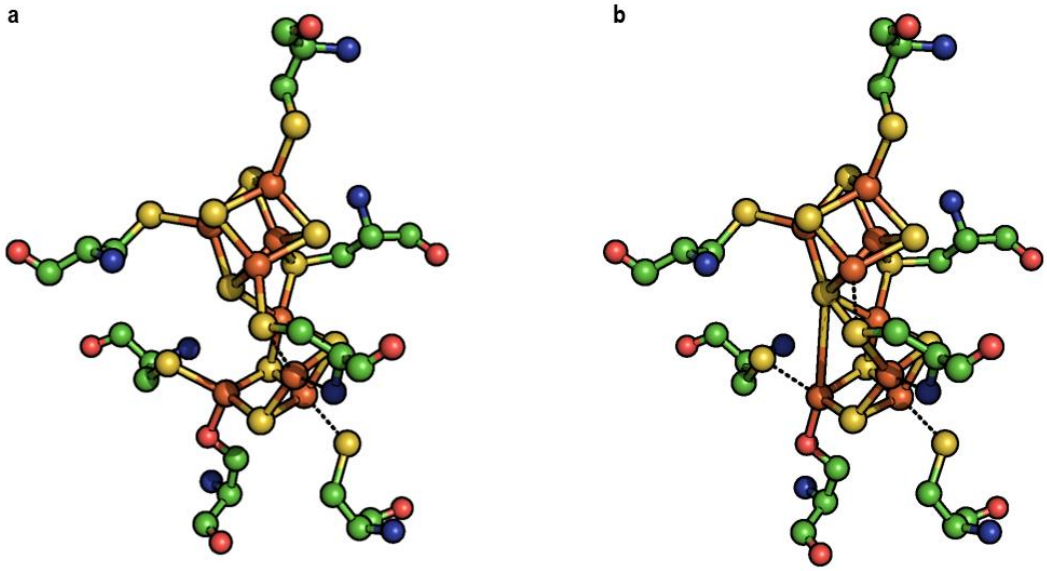


Figure 5: *Azotobacter vinelandii* P-cluster of nitrogenase. a) P^{OX} 2+ oxidized structure from PDB: 2MIN. b) P^{N} $2e^{-}$ reduced structure from PDB: 3U7Q. Shown in green are coordinating residues from the α and β subunits of the MoFe protein: α-Cys^{154}, α-Cys^{62}, α-Cys^{88}, β-Ser^{188}, β-Cys^{153}, β-Cys^{95}, and β-Cys^{70}.

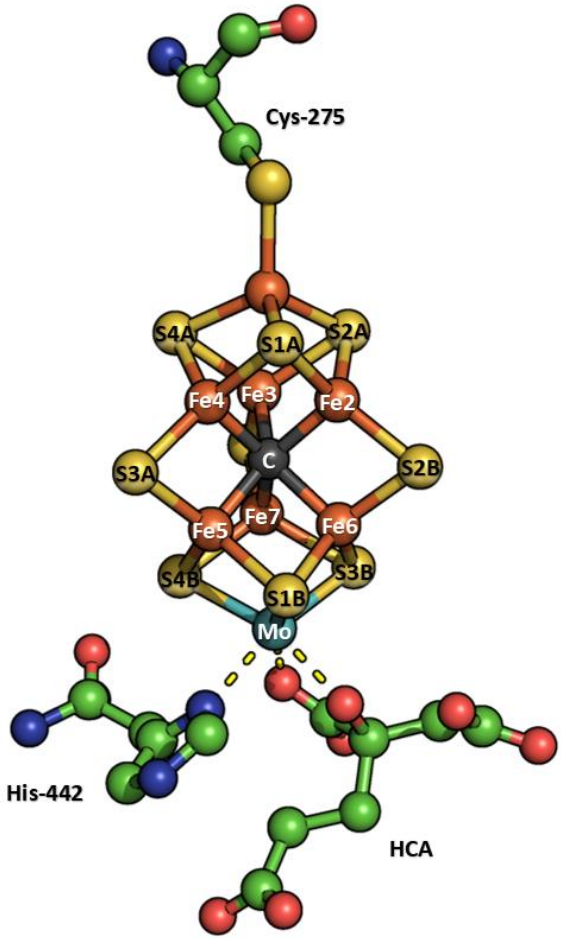


Figure 6: *Azotobacter vinelandii* FeMo-cofactor of nitrogenase. α-Cys[275], α-His[442], and *R*-homocitrate (HCA) are shown in green. PDB: 3U7Q. The interstitial carbon is shown in dark grey, molybdenum is shown in cyan [19].

The active site for substrate reduction is a metallocluster denoted as the FeMo-cofactor, a [7Fe:9S:C:Mo]-R-homocitrate cluster (Figure 2a and Figure 6) [19,40–43].

The focus of this study is to examine the manipulations of substrate by the FeMo-cofactor during turnover. To this end, we utilize previous research in kinetics, spectroscopy, structural biology, and molecular dynamics.

Mechanisms of Biological Nitrogen Fixation: Haber-Bosch vs Chatt Chemistry

The mechanism of biological nitrogen fixation is under much contention, but two main avenues are more commonly deliberated. One could consider a mechanism in which the FeMo-cofactor of nitrogenase resembles the iron-oxide catalyst of the Haber-Bosch process (Figure 7). In this case, dinitrogen could effectively adsorb to the cofactor where bond breakage would be expected to occur early. The dinitrogen bond being broken, subsequent electron/proton addition to reduce each adsorbed nitrogen would then ensue to eventually release ammonia.

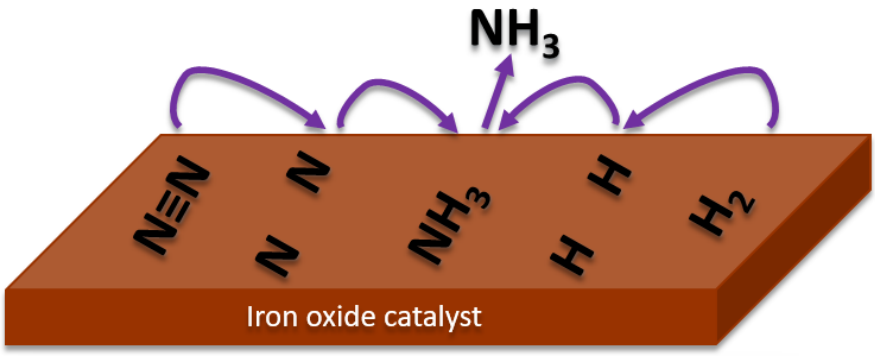


Figure 7: The proposed mechanism for the Haber-Bosch process.

Another plausible method for nitrogen fixation by nitrogenase would resemble Chatt-type chemistry (Figure 8) [44,45].

In this scheme, dinitrogen coordinates to an iron center on the FeMo-cofactor. The dinitrogen is then gradually reduced by alternating proton/electron addition for eventual release of two ammonia molecules.

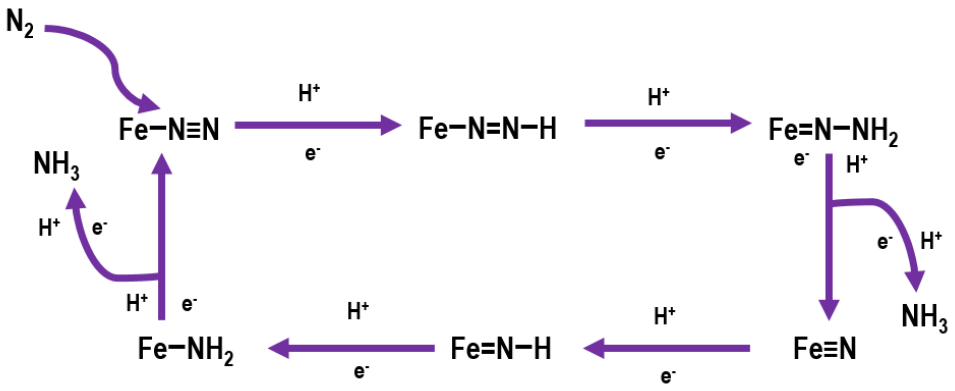


Figure 8: Chatt-type chemistry model of dinitrogen reduction. Adapted from [46].

As evidenced by the information reviewed here, a central mechanistic difference between a Haber-Bosch type scheme and a Chatt-chemistry method is the presence of dinitrogen intermediates. Since bond breakage occurs early in the Haber-Bosch mechanism, and reduction occurs at single nitrogen centers, one would not expect dinitrogen intermediates on the cofactor in such a scheme. Conversely, a Chatt-type mechanism would exhibit such intermediates. To truly delve into the possibilities of these two proposed mechanisms, it is essential to develop kinetic, spectroscopic, and structural methods for visualizing intermediate states of nitrogenase at very high resolutions.

Nitrogen Fixation: Mechanistic Conclusions in Kinetics, Spectroscopy, Computation, and Structure

The Lowe-Thorneley Model and the Substrates of Nitrogenase

After the 1893 discovery of free-living nitrogen-fixing bacteria, it was nearly 100 years later that Thorneley and Lowe published their culmination of kinetic experiments and computation of nitrogenase turnover in a scheme now known as the Lowe-Thorneley model (Figure 9) [38,39].

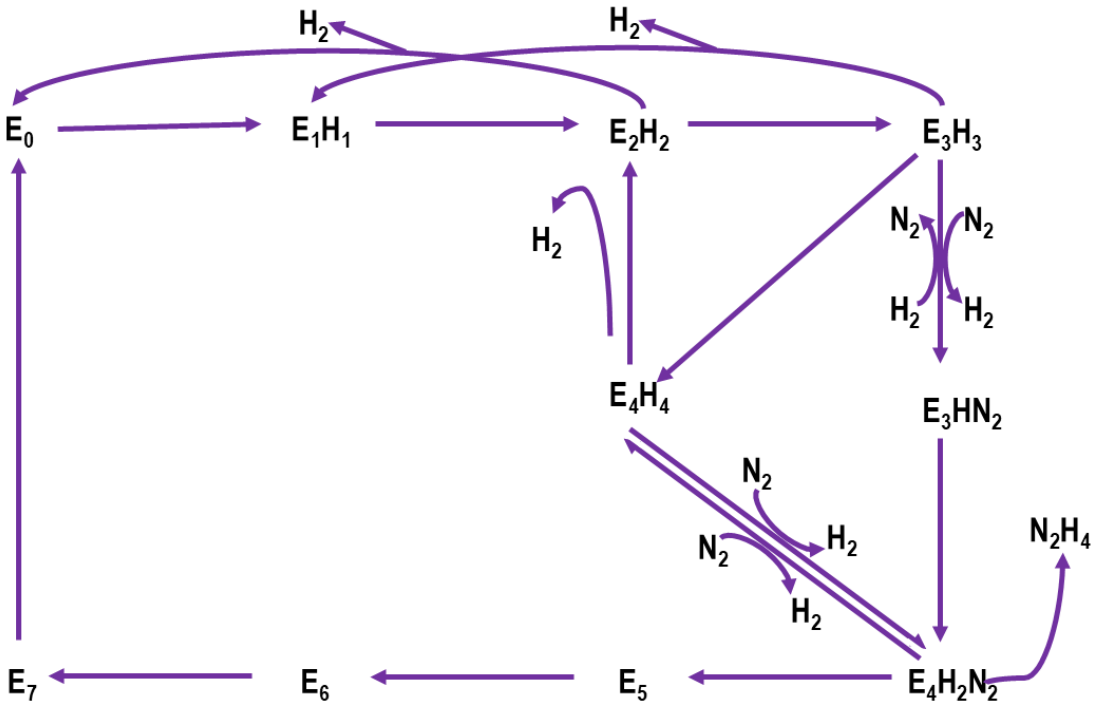


Figure 9: The Lowe-Thorneley model [39]. Adapted from [38].

The model includes eight Fe-protein cycles that are required to transfer eight electrons and eight protons to one active site of the MoFe-protein, leading to the reduction of one N_2 to $2NH_3$, and the concomitant evolution of one H_2. Other substrates are known to bind at various oxidation states of the MoFe protein, but N_2 binds after a three- or four-electron reduction (E_3, E_4). Compared to other substrates which bind more oxidized states of MoFe, N_2 is only efficiently reduced at high electron flux.

Nitrogenase is capable of reducing a wide variety of substrates from $2H^+$ to the isonitrile (CH_2=CHNC). This compound has a K_m ~0.8 mM, which indicates that it binds more tightly than N_2O or N_3^- [47]. Some other known substrates and inhibitors of nitrogenase are carbon monoxide (CO), cyanide (HCN, CN^-), and thiocyanate (SCN^-) [10,48]. There exists some kinetic evidence that some substrates such as CH_3NC may bind in more than one orientation to the same site on the cofactor [47]. Even more evidence suggests that two molecules---whether they be substrate or inhibitor---can be bound simultaneously [47]. In 1975, Rivera-Ortiz *et al*. demonstrated kinetically that N_2O is competitive with N_2 in partially

purified *Azotobacter vinelandii* nitrogenase; CO and acetylene are noncompetitive with N_2; CO, CN, and N_2O are noncompetitive with acetylene; and N_2 is competitive with acetylene [10]. In 2003, Han *et al*. proposed two binding sites on the active site for acetylene based on stereoselectivity and proton addition [21]. Nitrogenase can reduce a variety of other substrates including: carbon dioxide (CO_2), carbon monoxide (CO), carbonyl sulfide (COS), nitrous oxide (N_2O), diazene (N_2H_2), and more [3–11]. The common thread between observed substrates of nitrogenase (except for H^+) has been the presence of C-C, C-N, N-N, and N-O double or triple bonds [38,47,49]. CO has long been of interest to the study of the mechanism of nitrogenase, owing to its isoelectronic identity to N_2, and its potent inhibitor properties as well as its ability to serve as a weak substrate [12,13]. Like CO, cyanide compounds (X-CN) are also of interest to the study of nitrogenase due to the isoelectronic nature of CN to N_2. However, cyanide compounds serve as particularly interesting spectroscopic and crystallographic tools, because X in X-CN can be substituted for more significant sulfur or selenium (Se) [22]. These kinetic observations serve as motivation for further investigation of substrate manipulation by the FeMo-cofactor.

The current most commonly used method to investigate nitrogenase turnover is the acetylene reduction assay. Using gas chromatography, activity is monitored in the headspace of reaction mixtures containing an ATP regeneration system (creatine phosphate, ATP, phosphocreatine kinase, $MgCl_2$), a component for protein reduction (dithionite), and the protein mixture (Fe-protein and MoFe-protein) [50,51]. We utilize this method, as well as a similar N_2 reduction, proton reduction, and CH_4 production assays to investigate substrate manipulation at the cofactor [6,22,52].

Spectroscopy

Besides kinetic measurements of substrates and inhibitors on the cofactor, a great deal of research has been done on nitrogenase with spectroscopy. The complex electronic nature of the FeMo-cofactor complicates spectroscopic analysis. However, probing the electronic state of the cluster is achievable with the incorporation of a ligand. Therefore spectroscopic measurements are indispensable to the examination of ligand-bound states of the cofactor.

Ligand-bound spectroscopic studies arose after the initial discovery of CO as a specific and potent inhibitor of nitrogenase [53]. George *et al., 1997,* showed spectroscopically via stopped-flow Fourier transform infrared spectroscopy (SF-FTIR) that several transient FeMo-cofactor-bound CO species exist [20].

The as-isolated, dithionite-reduced FeMo-cofactor with S=3/2 has been extensively studied with Mössbauer, XANES, EXAFS, ENDOR, EPR, and more recently high energy resolution fluorescence detection (HERFD) [54–59]. Likely oxidation states for the entire cluster were recently reassessed by a report that identified Mo as Mo^{III} rather than the previously assumed Mo^{IV} [58]. The most recent overall oxidation state for the as-isolated FeMo-cofactor is $3Fe^{II}4Fe^{III}Mo^{III}$.

Also more recently were spectroscopic studies that unambiguously identified carbon as the interstitial ligand at the center of the FeMo-cofactor [19]. A 1.0 Å resolution structure of Av1 was supplemented by electron spin echo envelope modification (ESEEM) data of Av1, isotopically labeled with ^{13}C and ^{15}N. First, the cofactor center of the high resolution structure was integrated for electron density use a probe radius of 1.4 Å, the approximate van der Waals radius for a candidate atom. Varying the probe radius on a fine electron density grid gave statistical certainty that the central atom was indeed a carbon. These data were complemented by ESEEM, an electron paramagnetic resonance tool for the detection of weak hyperfine couplings in paramagnetic moieties like the clusters of nitrogenase. Two uniformly isotope labeled nitrogenases (^{13}C and ^{15}N) were compared to the wild-type using this method. The U-^{13}C-labeled nitrogenase revealed a spectral pattern at the free ^{13}C Larmor frequency (3.7 MHz) with a splitting of 2.5 MHz that was not detected in other samples. The ESEEM data showed strong ^{13}C hyperfine coupling with significant unpaired electron spin density, which correspond to a carbon within or near the FeMo-cofactor. Since there are no other carbons indicated in the crystal structure within the first coordination sphere, this coupling was determined to be from the central atom.

Another study published the same year identified the interstitial ligand as carbon using x-ray emission spectroscopy (XES) [42]. Here, the authors performed Fe Kβ valence-to-core (V2C) XES of intact MoFe protein, isolated FeMo-co, and the FeMo-co deficient Δ*nifB* protein. The Kβ" or "satellite" transitions had been previously shown to provide a signature

for directly coordinating ligands in Cr and Mn complexes [60,61]. Lancaster and colleagues developed V2C experimental and theoretical protocols with applications toward mono- and multinuclear iron complexes, including the study of a six iron cluster with a central μ_6-C^{4-} [62]. Data from these studies revealed a feature at 7099 eV, attributed to a transition from the μ_6-C^{4-} 2s orbital. This feature was calculated to shift to 7094 eV for a μ_6-N^{3-}, and to 7088 for a μ_6-O^{2-}. Comparisons of spectra from MoFe protein, isolated FeMo-co, and FeMo-co deficient Δ*nifB* protein to calculated and experimental spectra for μ_6-C^{4-}, μ_6-N^{3-}, and μ_6-O^{2-} coordinated species confirmed that the central ligand in the FeMo-co was a carbon.

Structural Biology: Studies of Nitrogenase under Turnover Conditions

At present, the best way to visualize changes to the FeMo-cofactor and its substrates, as well as to observe global changes to the nitrogenase protein, is with X-ray crystallography. The initial structural determination of the Fe-protein was published in 1992 at 2.9 Å resolution, along with the structures of the MoFe-protein at 2.7 Å and models of its metal centers [63–65]. Since then, structures from different organisms and different states have also been made available for both the Fe- and MoFe-proteins, as well as complexes between the two proteins [15,19,41,42,63,64,66–75]. All of these aforementioned studies provide important structures of the nitrogenase resting state, as well as identified the interstitial FeMo-cofactor ligand as carbon.

In 2014, the first ligand-bound structure of the FeMo-cofactor was published (Figure 2) [18]. Carbon monoxide (a potent inhibitor and weak substrate of nitrogenase) was observed to bind to the FeMo-cofactor on the Fe2-Fe6 face, displacing the S2B belt sulfur. Most interestingly, the S2B was demonstrated to be reversibly displaced. After removing the CO atmosphere from the reaction conditions, the bridging sulfur reappeared and the enzyme was reactivated. The 2014 study established an excellent model for future structural investigations of nitrogenase under turnover conditions, which we exploit in this report.

In 2017, another ligand-bound FeMo-cofactor was published (Figure 10) [76]. The vanadium-variant of the FeMo-co, the FeV-co, was shown to bind a μ-1,3 carbonate ligand in the 3A position. Interestingly, this position is obstructed in molybdenum nitrogenase,

emphasizing the role of the protein surrounding environment in the binding and turnover of various substrates.

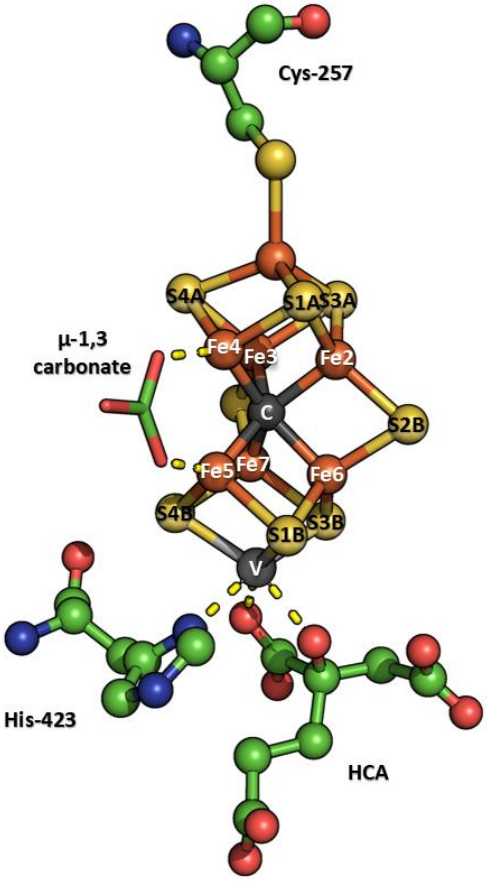


Figure 10: Binding of μ-1,3 carbonate at the 3A site of vanadium nitrogenase [76]. PDB: 5N6Y.

Incorporation of a Se-Probe

As stated previously, the complex nature of the FeMo-cofactor constrains efforts to study it spectroscopically. The particularly interesting finding of reversible belt-sulfur displacement under turnover has also prompted the need for a better probe to track substrate and the moving sulfur. Efforts have been made to track the sulfur with anomalous scattering, but the results have been inconclusive.

In 2015, Spatzal, Perez *et al.* successfully incorporated a selenium into the S2B site of the FeMo-cofactor in *Azotobacter vinelandii* nitrogenase (Figure 11) [22]. The Se2B

incorporated cofactor was active (85% of wild-type activity) in acetylene reduction assays, as well as in N_2 reduction, H_2 production, and CH_4 production.

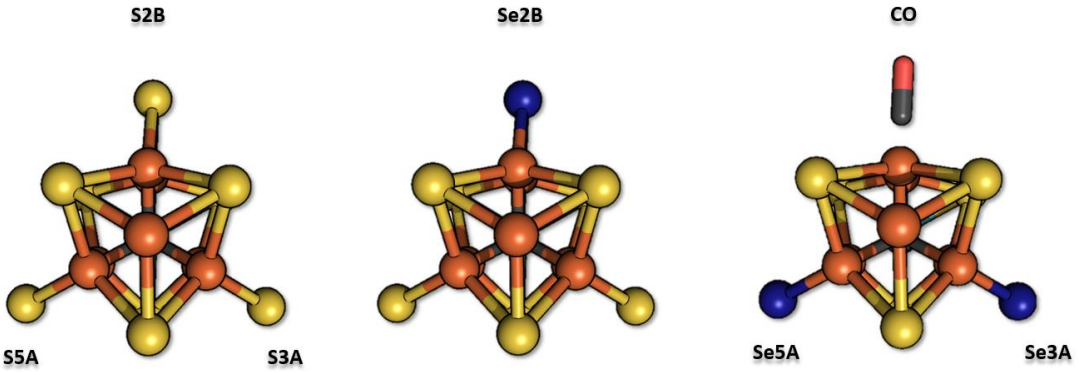


Figure 11: Selenium-labeled FeMo-cofactor [22]. Looking down the Fe1-C-Mo axis. Left) FeMo-co at the resting state. Middle) FeMo-co labeled with selenium at the 2B site under turnover conditions with 10 mM SeCN (Av1-Se2B). PDB: 5BVG. Right) Av1-Se2B under turnover conditions inhibited with CO. Selenium is shown in blue. PDB: 5BVH.

Incredibly, Spatzal, Perez *et al.*, *2015,* were also able to observe the selenium probe transfer around the cofactor to the other two belt sulfur sites under turnover conditions, before finally transferring out of the cofactor completely [22]. They were also able to show displacement of selenium by carbon monoxide to the other two belt sulfur sites (Figure 11b). Regardless of being active in terms of acetylene reduction, kinetic studies did show a longer initial lag phase than wild-type protein.

The importance of these findings for the tracking of substrates with the aid of the selenium probe cannot be understated. The selenium probe can serve as a significant beacon in crystallography as a strong anomalous scatterer. Kinetically, the observation of the longer lag phase with the selenium probe may have important implications for examining the earliest stages of substrate reduction.

In this study we utilize the selenium-probed system, where we will investigate the kinetics of nitrogenase turnover in the earliest stages, particularly Se movement into and out of the cofactor during turnover. Although many synthetic groups have attempted to make

compounds mimicking the FeMo-cofactor and its function, the actual cluster still defies synthesis [77–81]. To this end, molecular dynamics and simulation have been used to monitor ligands and the cofactor under turnover. A recent molecular dynamics study has put forward several predictions and possible mechanisms of the Se-probed system under turnover [82].

Lastly, we will use the selenium-incorporated system to perform high resolution spectroscopic measurements for a more thorough understanding of the electronic nature of the FeMo-cofactor. High energy resolution fluorescence detection (HERFD) has shown a great deal of promise for resolving features in the pre-edge region of XAS spectra, which give information regarding oxidation state and coordination geometry [58].

X-ray Crystallography

To obtain a visual representation of a molecule at atomic resolution, a technique capable of atomic resolution is necessary. Various wavelengths of light are available to distinguish objects, from the radio wave to gamma rays. In order for an object to diffract light and be subsequently visualized under magnification, the wavelength of this light must be no larger than the object, thus necessitating very short wavelengths to resolve very small objects (like atoms). As the wavelength of light decreases, the option for high resolution increases. In this report, we use X-rays to obtain high resolution molecular datasets. Figure 12 highlights the wavelength and frequencies of various categories of light, as well as the corresponding objects the types of light can resolve.

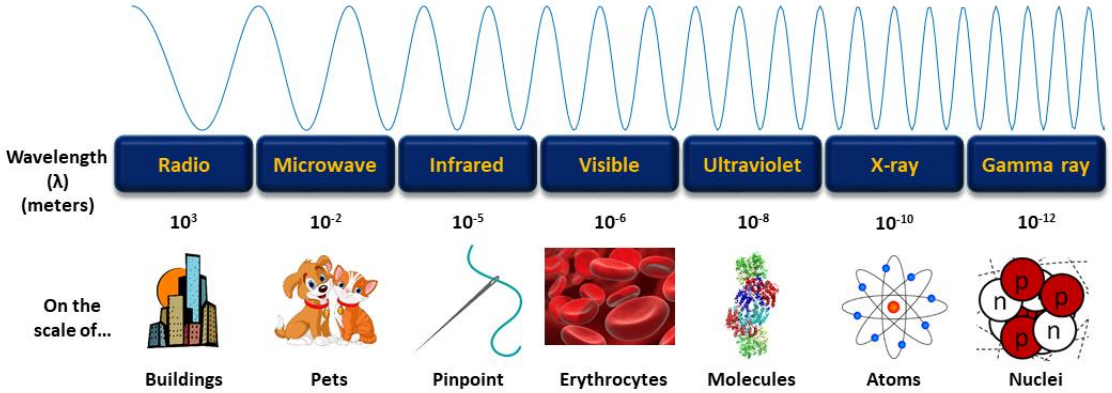


Figure 12: Frequencies and wavelengths of the various colors of light. Adapted from [83].

There are several additional caveats to working with X-rays in order to image proteins at the atomic level. One such caveat is that since X-rays cannot be focused by lenses, data must be collected on the diffraction plane and various computations must be performed to arrive at a final image of the molecule. A second caveat is that a lone protein molecule is a very weak scatterer of X-rays. Crystallographers must strengthen the diffraction potential of the protein molecule by growing large, well-ordered crystals. These crystals ideally will contain enough identically oriented protein molecules so that collectively, the molecules will diffract the incoming beam to produce a strong, detectable signal.

X-rays are generated at large synchrotron facilities, where electrons are accelerated to extremely high energy. The electrons are kept coherent and focused by passing through different types of magnets. The electrons at extremely high energies are also passed through undulators, which are a complex array of magnets, which force the electrons to accelerate, thereby producing X-rays in a focused, brilliant beam of radiation.

Protein crystals are mounted between the X-ray source and the detector, so that the crystal is directly in the X-ray path (Figure 13). A well-diffracting crystal will diffract the source beam into many discrete beams, each of which will produce a distinct spot (reflection) on the detector. The greater the intensity of the X-ray beam that reaches a position on the detector, the darker the reflection will appear. The position and intensity are then measured computationally, giving three-dimensional coordinates *h*, *k*, and *l* for each reflection, and relative intensity to one another.

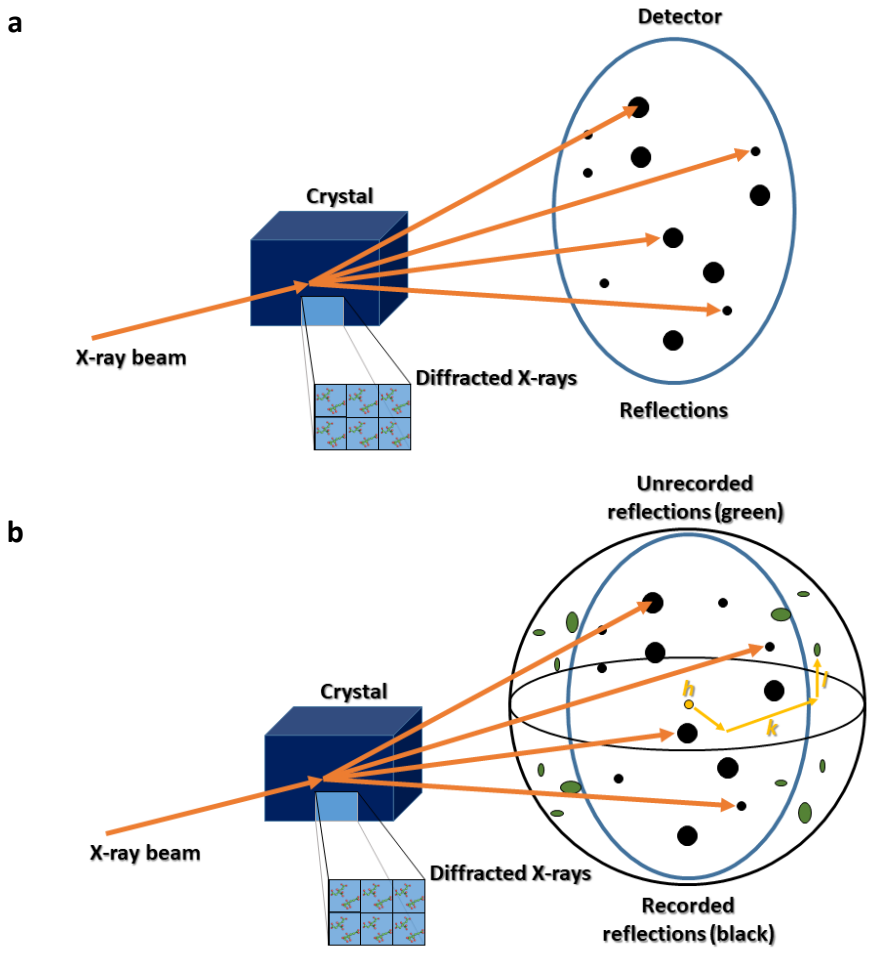


Figure 13: Crystallographic data collection. a) Recorded reflections from a protein crystal from a two-dimensional perspective. b) Recorded and unrecorded reflections from a protein crystal from a three-dimensional perspective.

The spacing of reflections (the *reciprocal lattice*) is inversely proportional to the spacing of the unit cell (pictured in Figure 13) in the crystalline lattice (the *real lattice*). From the spacing of the reciprocal lattice, we can calculate the unit cell of the crystal in angstroms.

In practice the protein crystal is rotated while within the beam, so multiple sets of reflections can be recorded from different orientations of the real lattice. A depiction of the three-dimensional diffraction pattern is shown in Figure 13b, with each reflection designated

at a position *h, k, l* in reciprocal space. The parameters used to analyze the diffraction pattern are the position *hkl* and the intensity I_{hkl} of each reflection.

Bragg's Law: A Necessity for Diffraction

To produce strongly diffracted rays, the incident light upon the crystal lattice must constructively interfere. This constructive interference is dependent upon the angle of incidence of the incoming light, and can be represented mathematically with Bragg's law (Figure 14).

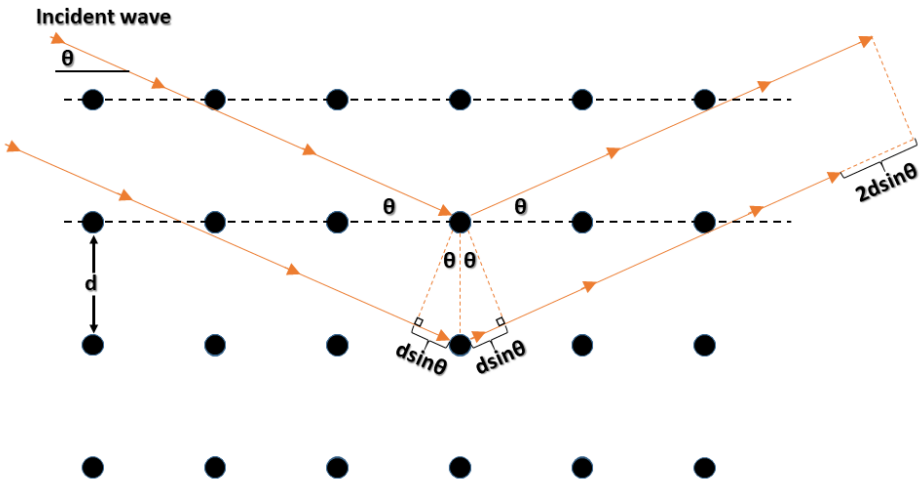


Figure 14: Bragg's law and the constructive interference of light from a crystal lattice.

As can be observed in Figure 14, an incident wave is diffracted from a point in the lattice. A second light wave is diffracted from another point in the lattice, and in order for these two light waves to constructively interfere, the extra path length traveled by the second light wave ($n\lambda = 2d\sin\Theta$) must be an integer multiple of the wavelength of the incident light ($n\lambda$). So Bragg's law can be written as:

$$n\lambda = 2d\sin\Theta,$$

where n is a positive integer, λ is wavelength, d is the lattice spacing distance, and θ is the angle of the incident light.

Bragg's Law in Reciprocal Space

Reciprocal lattice points allow us to compute the direction of diffracted beams from all sets of parallel planes in the crystal lattice. Reciprocal lattice points are depicted in Figure 15.

Reciprocal space is the space occupied by reflections, and within the reciprocal lattice are the locations of all Bragg reflections. In Figure 15a, an X-ray beam (depicted by the arrow AO) impacts the crystal along a plane. Point O is chosen as the origin of the both the reciprocal and the real lattice. A circle representing the wavelength of X-rays in reciprocal space can then be drawn, with a radius of 1/λ, and with its center C on the AO line. Rotation of the crystal about O brings reciprocal lattice points like R and R' into contact with the circle (Figure 15b). When point R (with indices *hkl*) is in contact with the circle, lines OR and BR can be drawn. The angle RBO is θ, and:

$$\sin\Theta = \frac{OR}{OB} = \frac{OR}{\frac{2}{\lambda}} \ ,$$

which with rearrangement becomes:

$$2\frac{1}{OR}\sin\Theta = \lambda \, .$$

The length of the line OR is $1/d_{hkl}$, so $1/OR = d_{hkl}$. Substituting this into the previous equation gives Bragg's law with n = 1:

$$2d_{hkl}\sin\Theta = \lambda \ .$$

When the crystal is rotated about O, and a reciprocal lattice point touches the circle, Bragg's law is satisfied and a reflection occurs. The sphere that emerges by rotating the circle of radius 1/λ about the X-ray beam is called the *Ewald sphere*, depicted in Figure 15.

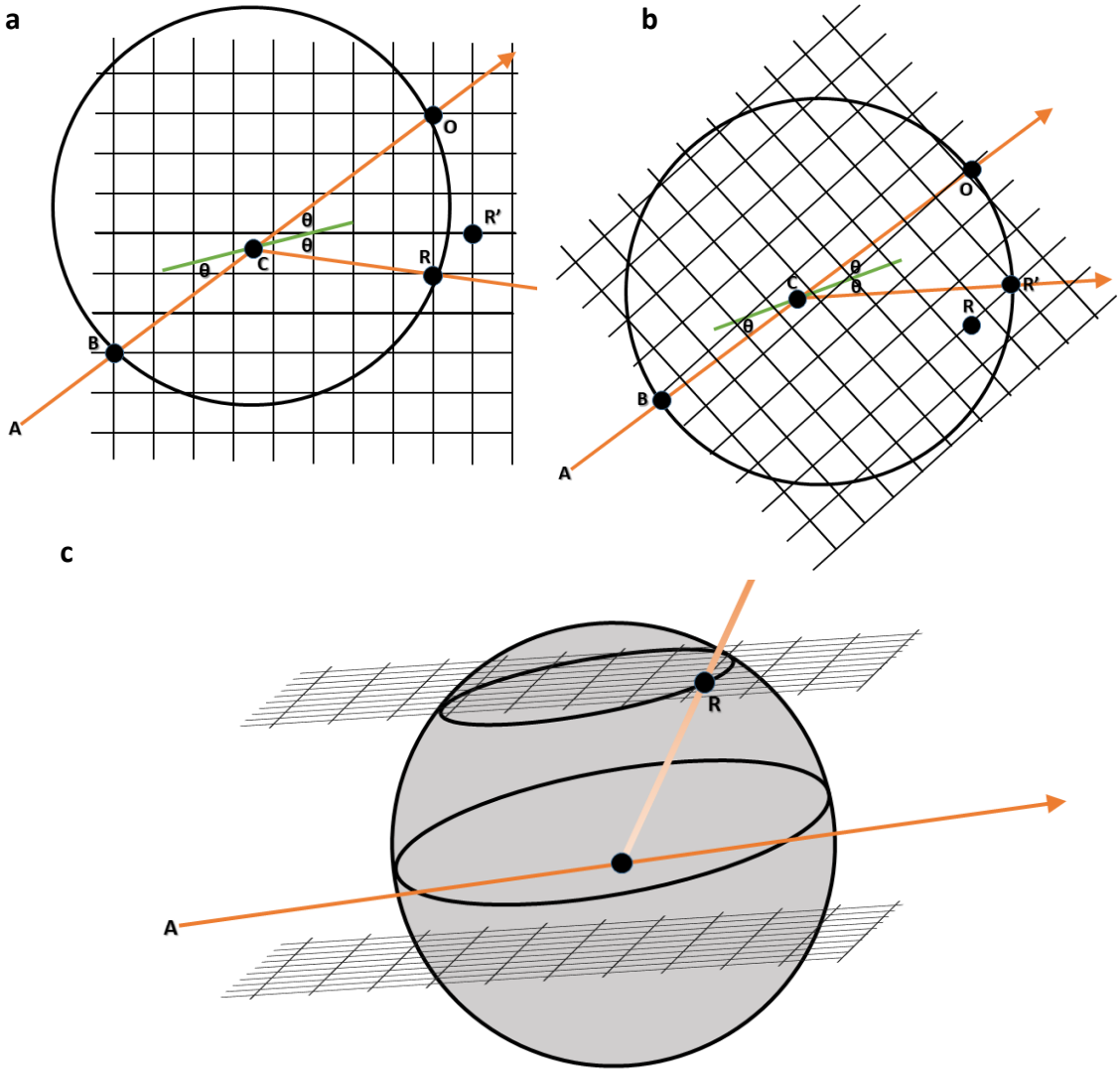


Figure 15: Bragg's law in reciprocal space. a) A ray emerges from the crystal when reciprocal lattice point R intersects the circle. b) Rotating the crystal around the origin (O), point R' comes to intersect the circle, producing a ray. c) Sphere of reflection with reciprocal lattice point R. *Adapted from* Rhodes, 2006 [84].

Diffraction Data to Electron Density: Fourier Sums and the Fourier Transform

Each recorded reflection can be described by a *structure-factor* equation, which is an equation describing the sum of the contributions from all the scatterers in the unit cell. For the reflection *hkl*, the computed sum is F_{hkl}. This structure factor F_{hkl} is a wave created by the superposition of many individual waves, each generated by diffraction from the electron cloud in the molecules of a crystal. Therefore, diffraction is related to and a direct result of the electron density of the molecules in the crystal. Given the ordered array of molecules in a protein crystal, the electron density in a crystal can be described with a three-dimensional periodic function, $\rho(x,y,z)$, with a ρ value for the electron density at every position in the unit cell.

The structure factor F_{hkl} can be written as a Fourier sum, composed of contributions from single atoms, f_{hkl}:

$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)},$$

where f_j is the scattering factor of the atom j. Since each diffracted X-ray is the sum of the diffracted contributions for all atoms in the unit cell, the structure factor for a single reflection F_{hkl} is written:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}.$$

The structure factor is therefore composed of a Fourier sum, in which each term represents the contribution of one atom to the reflection *hkl*. The contribution of each atom is treated as a simple sphere of electron density, and depends on the identity of the element and its position in the unit cell. The identity of the element determines the amplitude of the contribution (f_j), and the position within the unit cell (x_j, y_j, z_j) ascertains the phase.

We can then write the structure factor equation in terms of electron density over the volume of the unit cell:

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx+ky+lz)} dV \ .$$

We see that F_{hkl} is the Fourier transform of $\rho(x, y, z)$ on the set of real-lattice planes (*hkl*). And since the Fourier transform is a reversible function, we can show that the electron density is the transform of the structure factors, each entry term a reflection of the diffraction pattern:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \ .$$

The Phase Problem

As deduced from the preceding equations, we know that the function F_{hkl} is a periodic function, and must therefore contain terms for amplitude, frequency, and phase. Practically speaking, however, data collected from the reflections of the diffraction pattern only directly contain information about frequency and amplitude.

Frequency of the wave function are *h*, *k*, and *l*, indexing the reflection F_{hkl}. Therefore the frequency of the structure factor is $1/d_{hkl}$. The intensity of the reflection is related to the amplitude of the wave function, as shown in Figure 16.

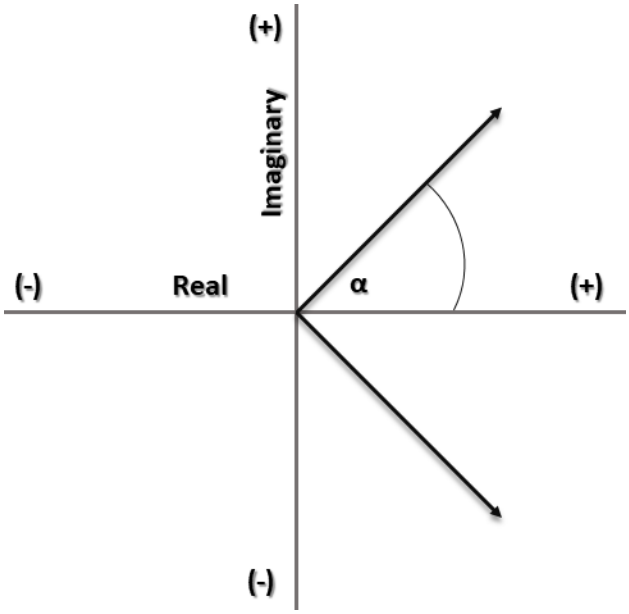


Figure 16. Argand diagram depicting vectors as complex numbers.

The length of the vector represents the amplitude of the structure factor and is proportional to the intensity I as: $|F_{hkl}| \propto (I_{hkl})^{\frac{1}{2}}$. Phase in the Argand diagram in Figure 16 is represented by α, the phase angle. To calculate electron density as a function of intensities and phases, we can write:

$$F_{hkl} = |F_{hkl}|\, e^{i\alpha},$$

and then substituting this into the electron density equation, we get:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} |F_{hkl}|\, e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}.$$

This final equation gives an expression for electron density in terms of the known amplitudes $|F_{hkl}|$ and unknown phases α_{hkl} of each reflection.

A fundamental problem of crystallography is the so-called phase problem, in which crystallographers use various tools to determine the phase angle α_{hkl} for each reflection.

Isomorphous Replacement

One such method to determine the phase angle is isomorphous replacement, or the heavy-atom method. Isomorphous replacement involves adding in a heavy atom, using the perturbation to the signal as a method to determine phase.

Often the crystallographer will soak protein crystals in solutions of heavy ions such as ionic complexes of Hg, Pt, or Au. Ideally these heavy ions will bind to sites within the crystal without perturbing the original structure. Figure 17 shows a representation of an Argand diagram with structure factors for the protein derivative reflection $\uparrow F_{PH}$ and the protein without derivative reflection $\uparrow F_P$.

Logically, the structure factor for the protein with the heavy atom derivative ($\uparrow F_{PH}$) would be equal to the sum of the structure factor for the heavy atom derivative ($\uparrow F_H$) and the structure factor for the protein alone ($\uparrow F_P$):

$$\uparrow F_{PH} = \uparrow F_P + \uparrow F_H .$$

To obtain the phases from isomorphous replacement, we rearrange the previous equation to solve for the protein structure factor:

$$\uparrow F_P = \uparrow F_{PH} - \uparrow F_H .$$

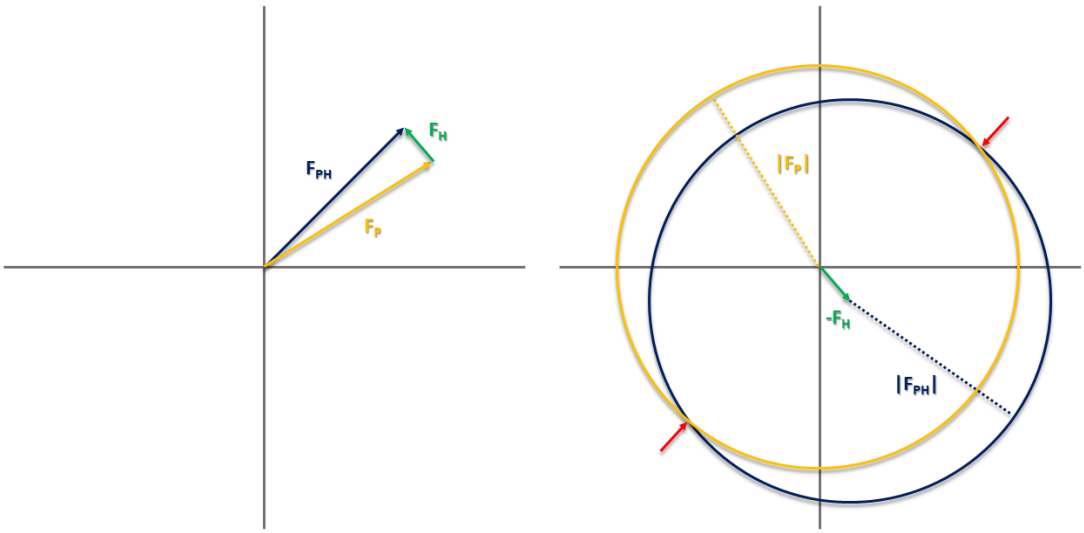


Figure 17. Harker diagram to obtain phases from heavy atom derivatives. Two possible phase angles would be associated with the vectors drawn to the red arrows. Adapted from Rhodes, 2006 [84].

From this equation and the Harker diagram in Figure 17, we can obtain the phase angle of $\uparrow F_P$ by first drawing a circle of radius $\uparrow F_{PH}$ with the point of vector $-\uparrow F_H$ at the center. A second circle is then drawn around the origin with radius $\uparrow F_P$. $\uparrow F_P = \uparrow F_{PH} - \uparrow F_H$ is only true at the points where the two circles intersect, and the corresponding phase angles are associated with these two possible vectors. If the two intersections lie near each other, sometimes it is sufficient to average the phase angles from these two vector points, thereby solving the phase angle by *single isomorphous replacement*. However, sometimes it is necessary to use multiple heavy atom derivatives to obtain accurate phase angles. The aforementioned process is repeated with the second heavy atom hopefully binding at a different place than the first heavy atom, and one phase becomes more likely than the other. This process is called *multiple isomorphous replacement* (MIR).

Anomalous Scattering

A useful physical property of heavy-atoms is that they have the capacity to absorb X-rays of a specified wavelength. Anomalous scattering is therefore the scattering of light at an absorption edge, or the point where the intensity of absorption rises sharply as a function of X-ray wavelength.

With coherent scattering, a set of reflections are produced from the lattice planes. The reflection intensities produced by the *hkl* lattice plane are identical to the reflection intensities produced from the *-h, -k, -l* lattice plane. This describes Friedel's law, where these reflections have equal amplitudes and opposite phase angles:

$$| F_{hkl} | = | F_{-h-k-l} | .$$

Friedel's law with coherent scattering is depicted in Figure 18a. Also under coherent scattering, the scattering factor is given by:

$$f = f_0 ,$$

where f_0 is the normal scattering factor and is a real number.

When an incident X-ray beam is instead absorbed near the absorption edge, a phase shift occurs which breaks Friedel's law and $| F_{hkl} | \neq | F_{-h-k-l} |$, Figure 18bc. Two correction factors must then be employed to calculate f. The scattering factor under anomalous dispersion becomes:

$$f = f_0 + \Delta f' + i\Delta f'' .$$

Again, f_0 is the normal scattering factor and is a real number. $\Delta f'$ is the real part of the scattering factor, while $\Delta f''$ is the imaginary part. Both $\Delta f'$ and $\Delta f''$ are wavelength dependent.

The full structure factor can then be written as:

$$F_{hkl} = [f_0 + \Delta f'(\lambda) + if''(\lambda)].$$

The real component $\Delta f'$ is 180° out of phase with f_0, and the imaginary component $\Delta f''$ is 90° out of phase with f_0.

F_{hkl} can be solved once again using Harker diagrams (Figure 18c), where the point of intersection of the nonanomalous structure-factor circle and the anomalous structure factor circle establishes two phase possibilities. The two circles were drawn according to the fact that:

$$F_{hkl}^{\lambda 1+} = F_{hkl}^{\lambda 2+} - \Delta f'^{+} - \Delta f''^{+}.$$

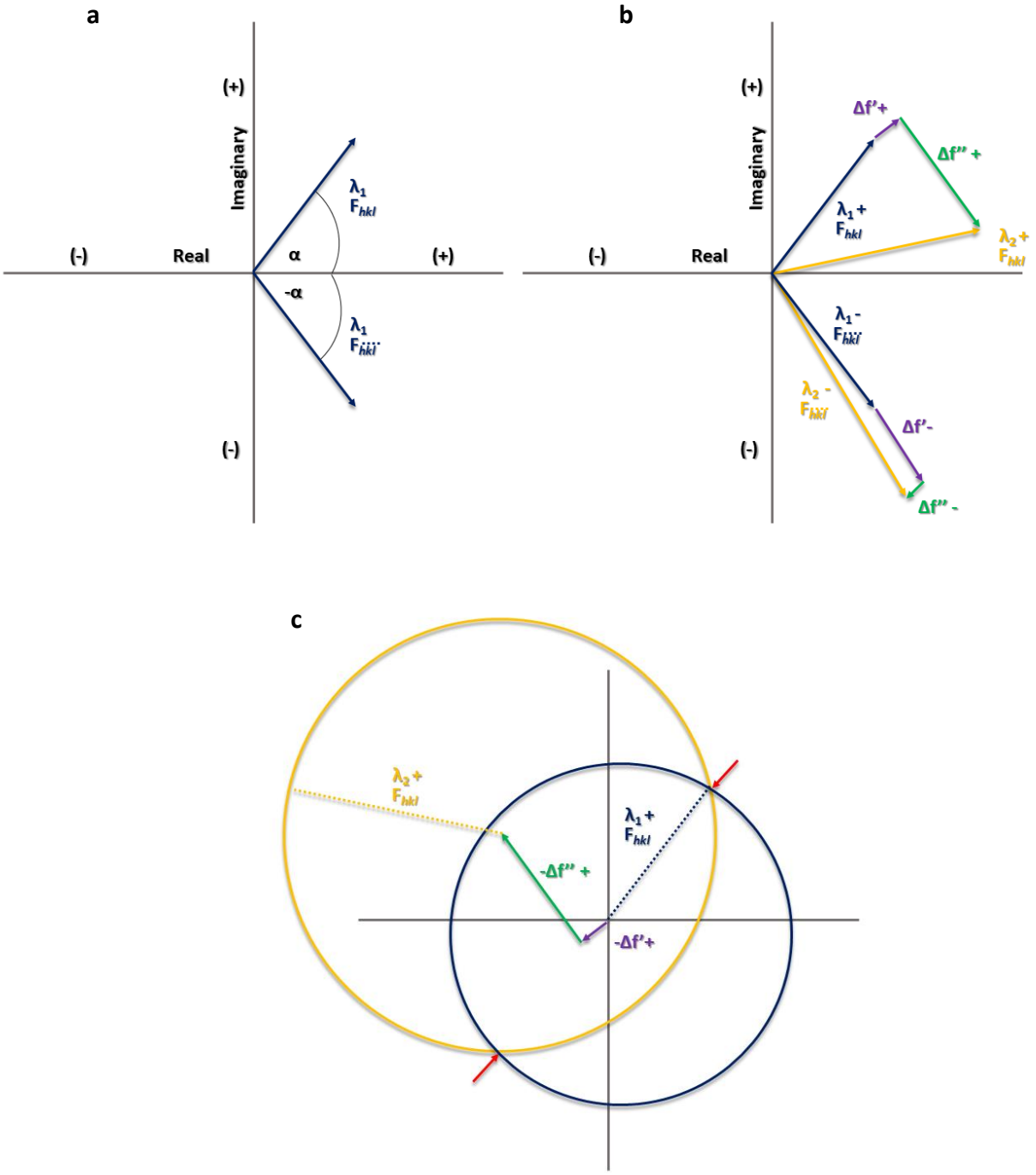


Figure 18. Friedel's law and anomalous diffraction. a) Friedel's law with coherent scattering. b) Friedel's law is broken with anomalous scattering. c) Harker diagram to solve the phase for F_{hkl}.

To distinguish which phase is correct, we can reflect $F_{hkl}^{\lambda 2-}$ (the Friedel pair in this situation) across the real axis, along with all of its vector components (Figure 19). Multiple iterations may be necessary to rigorously determine the phase, but the general process is summarized in Figure 18 and 19.

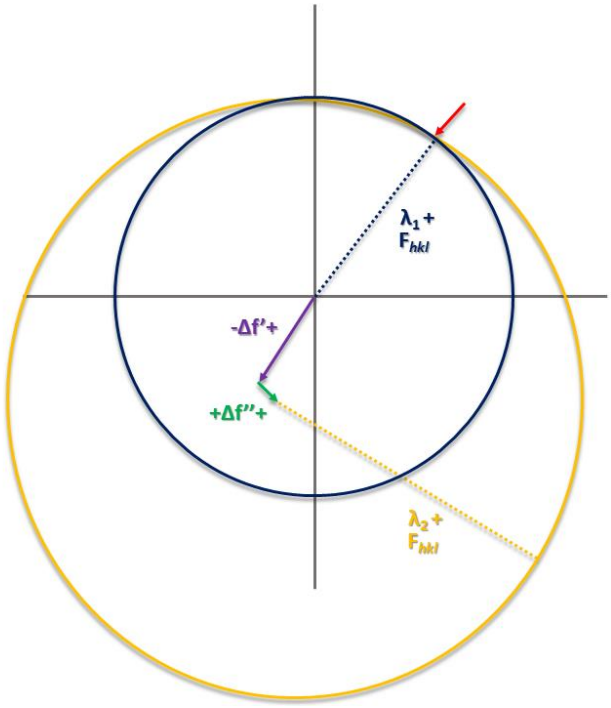


Figure 19. Second iteration of anomalous scattering Harker diagram to determine phases.

Molecular Replacement

Molecular replacement is a convenient method for determining phases in a new structure. The method utilizes existing structure factors of a known protein, and calculates initial phases of the new model by placing these existing structure factors into the unit cell of the new protein. If the new protein target and the phasing model are isomorphous, the native intensities $|F_{hkl}^{new}|$ of the new protein are used in a Fourier synthesis with phases from the model protein, α^{model}:

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}^{new}| e^{-2\pi i(hx+ky+lz-\alpha_{hkl}^{model})} .$$

Similarly, for non-isomorphous instances, the Fourier synthesis above may be applied, but first the orientation and position of the new protein within the unit cell must be determined using rotation and translation functions. This requires a computational search in which, for each position and orientation attempted in the calculation, structure factors F_{calc} of the model are calculated with the experimental amplitudes $|F_{obs}|$ obtained from the new protein. This process is repeated until the best match is obtained, and the resulting phases α_{calc} from the calculation are applied to the new protein. Computationally, it is better to perform orientation and position calculations separately.

Comparison of different orientations in the search to find the optimal combination is done by computing a rotation function. The rotation function makes use of the Patterson function, which is a Fourier sum determining the heavy-atom coordinates in the unit cell of a heavy-atom derivative crystal. The Patterson function $P(u,v,w)$ is a Fourier sum without phases, where the amplitude of each term is the square of one structure factor:

$$P(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)} .$$

A point-by-point comparison of two Patterson functions is calculated within a given radius of the origin. One of the Pattersons is then rotated, and a new point-by-point comparison is performed. This is repeated, until finally integration of the products of the two maps yields large numbers where the maps have coincident peaks, and small numbers where there is little overlap. This rotation search is repeated with a rotation matrix, finding the best orientation between the model and the new protein until the most optimal orientation is determined.

To determine the location of the new protein within the unit cell, the expected structure-factor amplitudes from the model at a given location and the amplitudes from the new protein are compared. The so-called R-factor compares the overall agreement between the two datasets:

$$R = \frac{\sum || F_{obs} | - | F_{calc} ||}{\sum | F_{obs} |}.$$

CHAPTER 2: NITROGENASE METHODOLOGIES

Various methodologies are used in this project including X-ray crystallography, liquid chromatography, and gas chromatography.

Cell Growth: *Azotobacter vinelandii*

Pre-cultures of *Azotobacter vinelandii* were cultivated in 50 mL of modified Burke's medium containing 10 mM NH_4Cl as a nitrogen source, 20 g/L sucrose, 0.2 mM $FeSO_4 \cdot 7H_2O$, 3 µM $Na_2MoO_4 \cdot 2H_2O$, 1.67 mM $MgSO_4$, 0.9 mM $CaCl_2$, and 10 mM KH_2PO_4/K_2HPO_4 (pH 7.5) at 30º C with shaking at 180 rpm [117]. Pre-cultures were grown from frozen glycerol stocks containing 1:1 (v/v) cell stock at an optical density (OD_{600}) between 1.5-2 and 50% glycerol. Main cultures (15 L) were also grown in Burke's medium, except with a depleted source of nitrogen (1.3 mM NH_4Cl) to stimulate nitrogenase expression. The cultures were grown in an Eppendorf BioFlo Bioreactor with a dissolved oxygen (DO) cascade designed to increase agitation to 800 rpm when the dissolved oxygen fell below 80%. Air flow into the media was set to 20 L/hour, and minimum agitation was set to 300 rpm. The temperature was maintained at 30º C via a heat jacket, and excess foam was avoided by injecting 200 uL of sterile anti-foaming agent (polyethylene glycol 1000). Cells were harvested by centrifugation at 8000 rpm and 4º C when the cells reached an OD of 1.5. The workflow for cell growth as well as the BioFlo Bioreactor are pictured in Figure 20.

Nitrogenase protein activity was assessed via gas chromatography in which 2 mL of whole cells were placed into a sealed 9 mL Wheaton vial. 1 mL of headspace was removed from the vial and replaced with 1 mL of pure acetylene at 1 atm. The vials were allowed to react for 30 min at 30º C with shaking at 180 rpm. Optimal nitrogenase activity was observed at raw peak areas of ethylene of 15- to 25-fold above a water control containing no cells.

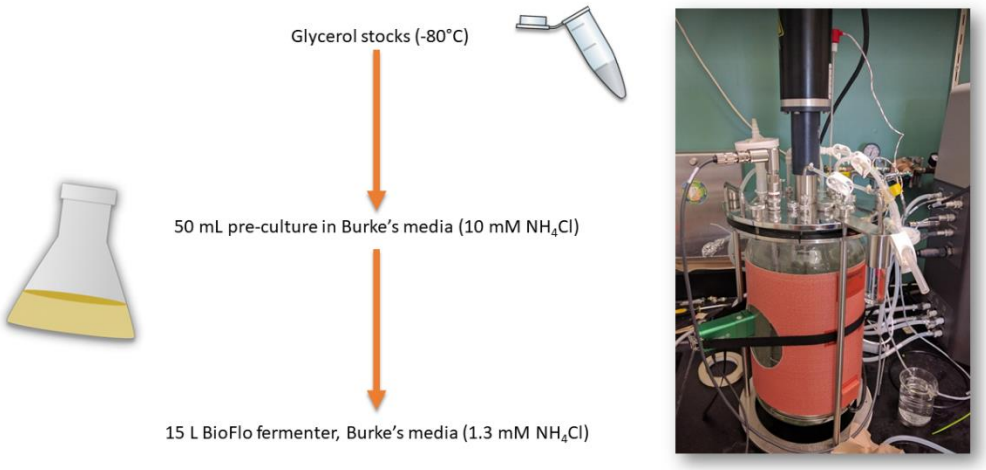


Figure 20. Cultivation of *Azotobacter vinelandii* using a BioFlo 115 Bioreactor.

Purification of the MoFe- and Fe-proteins

Isolation of the MoFe- and Fe-proteins were performed anaerobically, either on a manifold with argon, or in an anaerobic chamber with 95% argon / 5% hydrogen atmosphere. The purification was modified from Spatzal and Einsle, 2012 [85] (Figure 21).

All buffers were degassed using automated argon-vacuum cycles on a custom-made high-vacuum Schlenk line. All liquid lines to the purification system (AKTA) were pre-equilibrated with 5 mM $Na_2S_2O_4$ in degassed purification buffers, and all purification buffers were kept under argon. Cells were homogenized in an anaerobic chamber with 50 mM Tris/HCl (pH 7.5), 150 mM NaCL, and 5 mM $Na_2S_2O_4$. Cells were then lysed using a high pressure homogenizer (Emulsiflex C5, Avestin), pressurized with argon. The cell lysate was then transferred to centrifuge bottles in the anaerobic chamber, and centrifuged at 14,000 rpm (JA-14 rotor). Supernatant was loaded onto a HiTrap-Q anion exchange column (70 mL, GE-Healthcare), which had been pre-equilibrated with loading buffer (50 mM Tris/HCl, pH 7.5; 150 mM NaCl; 5 mM $Na_2S_2O_4$). A linear NaCl gradient over 300 min was then applied to the column using elution buffer (50 mM Tris/HCl, pH 7.5; 1 M NaCl; 5 mM $Na_2S_2O_4$). MoFe-protein (Av1) eluted at approximately 350 mM NaCl, while Fe-protein (Av2) eluted at approximately 475 mM.

After collection from the anion exchange column, the proteins were each concentrated using an Amicon concentrator pressurized with 5 bar argon in the anaerobic

chamber. The MoFe-protein was concentrated to approximately 10 mL with 100,000 kDa MWCO filters (Millipore Ultracell), while the Fe-protein was concentrated to 10 mL with 30,000 MWCO filters. After concentration, each protein was loaded onto a size exclusion column (S200, 26/60, GE Healthcare, 450 mL) which had been pre-equilibrated with size exclusion buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$). The pure proteins were then collected and concentrated to less than 2 mL of total volume, for MoFe-protein final concentrations of ~30 mg/mL, and Fe-protein concentrations of ~40 mg/mL. Protein concentrations were determined with UV-vis absorbance at 410 nm, with extinction coefficients of 76 $mM^{-1}cm^{-1}$ for the MoFe-protein and of 9.4 $mM^{-1}cm^{-1}$ for the Fe-protein. Nitrogenase *in vitro* activity was monitored via the acetylene reduction assay (ARA), described in the next section.

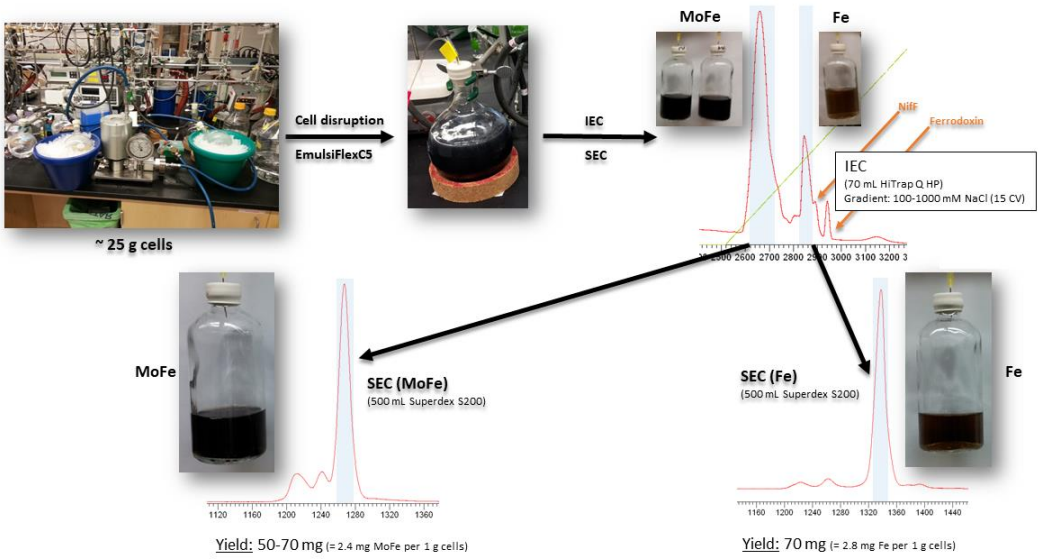


Figure 21. Overview of the purification procedure for the MoFe- and Fe-proteins. From the top left: Cell disruption with high pressure Emulsiflex C5, cell lysate collected under argon, ion exchange (IEC) of proteins, and size exclusion (SEC) of each protein with approximate yields.

Acetylene Reduction Assay

Nitrogenase *in vitro* activity was monitored via the acetylene reduction assay (ARA), in which the reduction of acetylene ($HC{\equiv}CH$) is reduced to ethylene ($H_2C{=}CH_2$) [50,51] (Figure 22). Activity mixtures consist of 20 mM creatine phosphate, 5 mM ATP, 5 mM $MgCl_2$, 25 units/mL phosphocreatine kinase, and 25 mM $Na_2S_2O_4$. The stock solutions for creatine phosphate, ATP, and phosphocreatine kinase were all prepared in 50 mM Tris/HCl (pH 7.5), while the $Na_2S_2O_4$ stock solution was prepared in 500 mM Tris/HCl (pH 7.5). Activity mixtures were dispensed in 1 mL volumes into sealed 9 mL Wheaton vials and degassed via Schlenk line and kept under argon. 1 mL of argon headspace was replaced by 1 mL of acetylene and incubated at 30° C for 4 minutes with shaking at 180 rpm. The reaction was then initiated by addition of the component proteins (MoFe=Av1, Fe=Av2). The standard component ratio for checking nitrogenase activity in this study was Av2:Av1=2:1, active site ratio 1:1, 0.25 mg Av1 : 0.27 mg Av2. The volumes of the reaction mixtures were kept constant by adding varying amounts of degassed size exclusion buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$). Reactions were terminated by addition of 3 M citric acid at varying timepoints, and ethylene and acetylene were assessed via gas chromatography (activated alumina 60/80 mesh column, flame ionization detector, column temperature at 110° C, detector temperature 150° C, and He carrier) from the headspace of the reaction vials. Peak integration was performed using the Peak Simple software (SRI Instruments), with defined amounts of acetylene in protein-free assay mixtures serving as calibration curves.

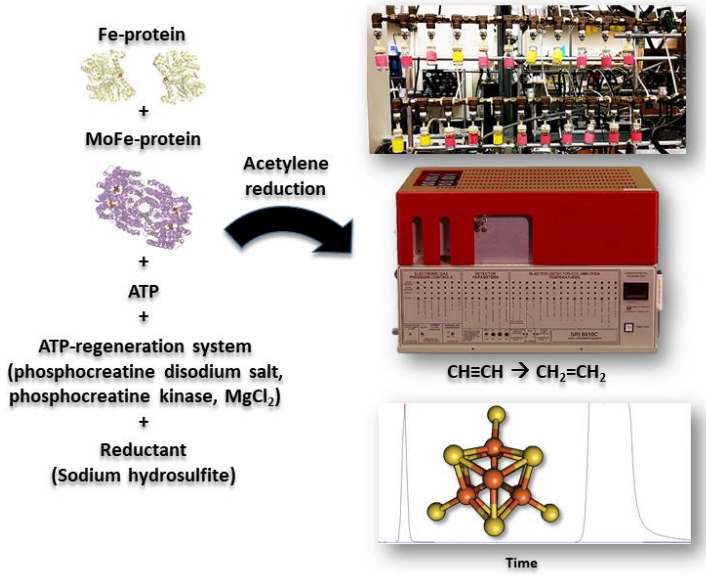


Figure 22. Overview of the acetylene reduction assay. From the left: activity assay mixtures, image of reaction vials on custom Schlenk line, gas chromatograph, and sample gas chromatogram for active protein.

Dinitrogen Reduction Assay

Nitrogenase *in vitro* activity of dinitrogen turnover was monitored by ammonia formation using a fluorescence method with an *o*-phthalaldehyde mercaptoethanol reagent [6,52]. The assay conditions were similar to the acetylene reduction assay, with a few modifications. 1 mL assay mixtures consisted of 20 mM creatine phosphate, 5 mM ATP, 5 mM MgCl, 25 units/mL phosphocreatine kinase, and 25 mM $Na_2S_2O_4$. The stock solutions for the creatine phosphate, ATP, and phosphocreatine kinase were all prepared in 150 mM MOPS buffer (pH 7.5). The stock solution for $Na_2S_2O_4$ was prepared in 500 mM MOPS buffer (pH 7.5). The headspace of the reaction vials were flushed with N_2 and incubated at 30° C for 4 min with shaking at 180 rpm. Reactions were initiated by addition of the component proteins (Av2:Av1=2:1, active site ratio 1:1) and terminated at desired timepoints with 300 uL of 0.5M EDTA (pH 8.0). After quenching the reaction, all following manipulations were performed aerobically. The assay mixture was filtered using 3 kDa MWCO centrifugal filters (Amicon Ultra) for 20-25 min in a tabletop centrifuge at maximum speed (13,000 rpm). A fluorescence reaction mixture was prepared containing

20 mM phthalic dicarboxyaldehyde, 3.5 mM 2-mercaptoethanol, 5% v/v ethanol, and 200 mM potassium phosphate (pH 7.3). In black Eppendorf tubes to protect from light, 25 uL of the filtrate from the activity mixtures was added to 975 uL of the fluorescence mixture, and were thoroughly mixed by inverting the tubes several times. The mixture was allowed to react in the dark for 30 minutes. Lastly, 100 uL of each reaction mixture were placed into 96-well black Corning plates (clear bottom), and fluorescence measurements were performed using a Flexstation 3 plate reader with SoftMax Pro Microplate Data Acquisition and Analysis software (Molecular Devices). Emission was set to λ_{Em}=472 nm, and excitation to λ_{Ex}=410 nm. A calibration curve was generated using NH_4Cl standards subjected to the same fluorescence mixture method.

Proton Reduction Assay

The proton reduction assay is identical to the acetylene reduction assay except for the detection method, the omission of acetylene, and use of argon as the carrier gas. Dihydrogen production from proton reduction was monitored via gas chromatography using a molecular sieve 5A-80/100 column and a thermal conductivity detector. Calibration curves were generated using 5% H_2 / 95% Ar standards in standard assay mixtures.

Se-Labeling of the FeMo-cofactor of MoFe

Se-incorporated Av1-FeMo-co (Av1-Se) was prepared with slight modifications to methods previously described [22,86]. The activity mixtures are identical to the proton reduction assay mixtures, with added KSeCN in varying concentrations. The KSeCN stock solution was prepared in size exclusion buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$). Inhibition was assessed by running parallel acetylene reduction assays with varying KSeCN, and observing reduction of the acetylene reduction to ethylene as evidence of inhibition (Figure 23). Depending on the desired timepoint, Av1-Se mixtures were either flash-frozen in liquid nitrogen, or were chilled to 4° C before concentration.

Av1-Se was then isolated from the activity components in the anaerobic chamber by ultrafiltration using a 100 kDa MWCO filter. The mixture was kept on ice to slow any further reactions in the assay mixture, and was washed 2-5 times in cold size exclusion

buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$). The final protein concentrations were adjusted to 30 mg/mL for crystallography.

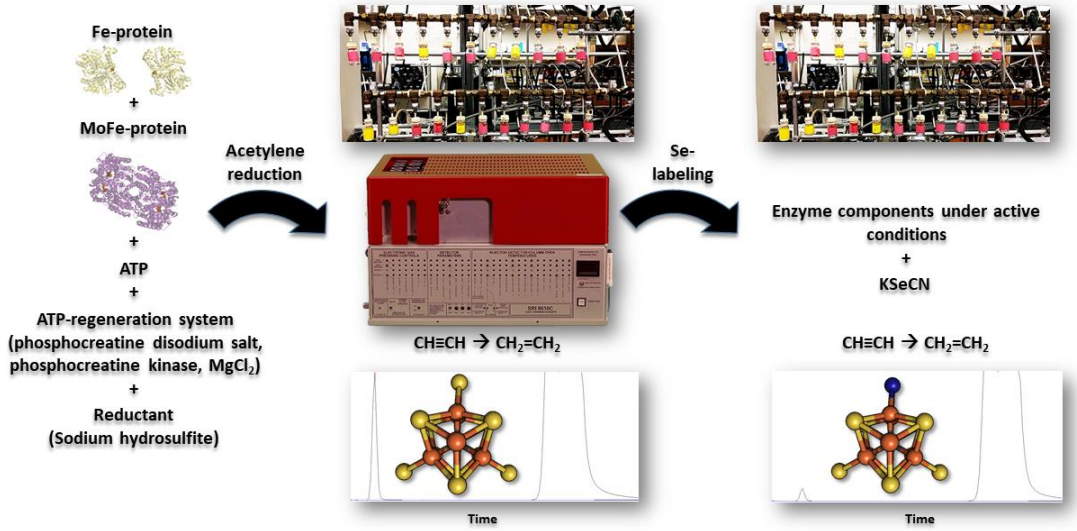


Figure 23. Se-labeling of the FeMo-co. From the left: activity mixtures with all assay components, acetylene reduction activity without KSeCN, and acetylene reduction activity with added KSeCN.

Crystallization of MoFe

All Av1-Se samples were crystallized using the sitting drop vapor diffusion method at 21° C in an anaerobic chamber (5% H_2 / 95% Ar), with protein concentration at 30 mg/mL. The reservoir solution consisted of: 24-28% PEG 6000 (v/v), 750-850 mM NaCl, 100 mM imidazole/malic acid (pH 8.0), and 5 mM $Na_2S_2O_4$ [18,18,19,22]. For CO-inhibited samples, the reservoir solutions were bubbled through with CO before setting the drops. After the drops were set, clear duct tape was used to seal the reservoirs. The tape was then punctured with a syringe and the wells were quickly filled with an additional 2 mL of CO, and resealed with a second layer of clear Duct tape. A seeding strategy using crushed Av1 crystals was employed to accelerate crystallization and to generate higher quality crystals. Cryo-protection of the crystals before freezing in liquid nitrogen consisted of dipping the crystals into a 5 uL drop of reservoir solution containing 8-12% MPD (v/v).

Quantification of Se-occupancy from Anomalous Data

Se-occupancies within all Av1-Se crystals were calculated using a modified script from [22]. Diffraction data were collected at 7110 eV (1.743 Å) and 12662 eV (0.979 Å) at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2 equipped with a Dectris Pilatus 6M detector. At 12662 eV, an energy experimentally determined by MAD scans to be above the Se K-edge. The Fe atoms in the MoFe-protein were assumed to be at 100% occupancy, and were then used as internal standards to estimate Se-occupancy using the relationship:

$$f = f_0 + \Delta f' + i\Delta f''.$$

At 12662 eV, the atoms of interest were determined to have experimental scattering factors $f_{Fe}'' = 1.5$, $f_{Se}'' = 3.84$, and $f_S'' = 0.24$. The data were indexed, integrated, and scaled using XDS and Aimless [87–89]. Molecular replacement was used to obtain phase information, with the 1.0 Å Av1 structure (PDB: 3U7Q) used as a model [19]. Structure refinement and modeling was performed using the CCP4-embedded programs REFMAC5 and Coot [88,90,91].

Se-anomalous density maps were calculated based on the data collected at 12662 eV, while the S-anomalous maps were calculated based on the data collected at 7110 eV. Refmac-refined mtz files were merged with anomalous mtz files using the CCP4-embedded program CAD. Maps were generated using fast-Fourier transform (FFT) of the merged mtz files and the Refmac-refined pdb coordinates. The CCP4 program Mapmask was used to orient the model based on the data resolution limits. Finally, the program Mapman performs map normalization and extracts the density values on an arbitrary scale at all atomic positions within a set sphere [92] (Figure 24). The structures were then rendered for publication in PYMOL.

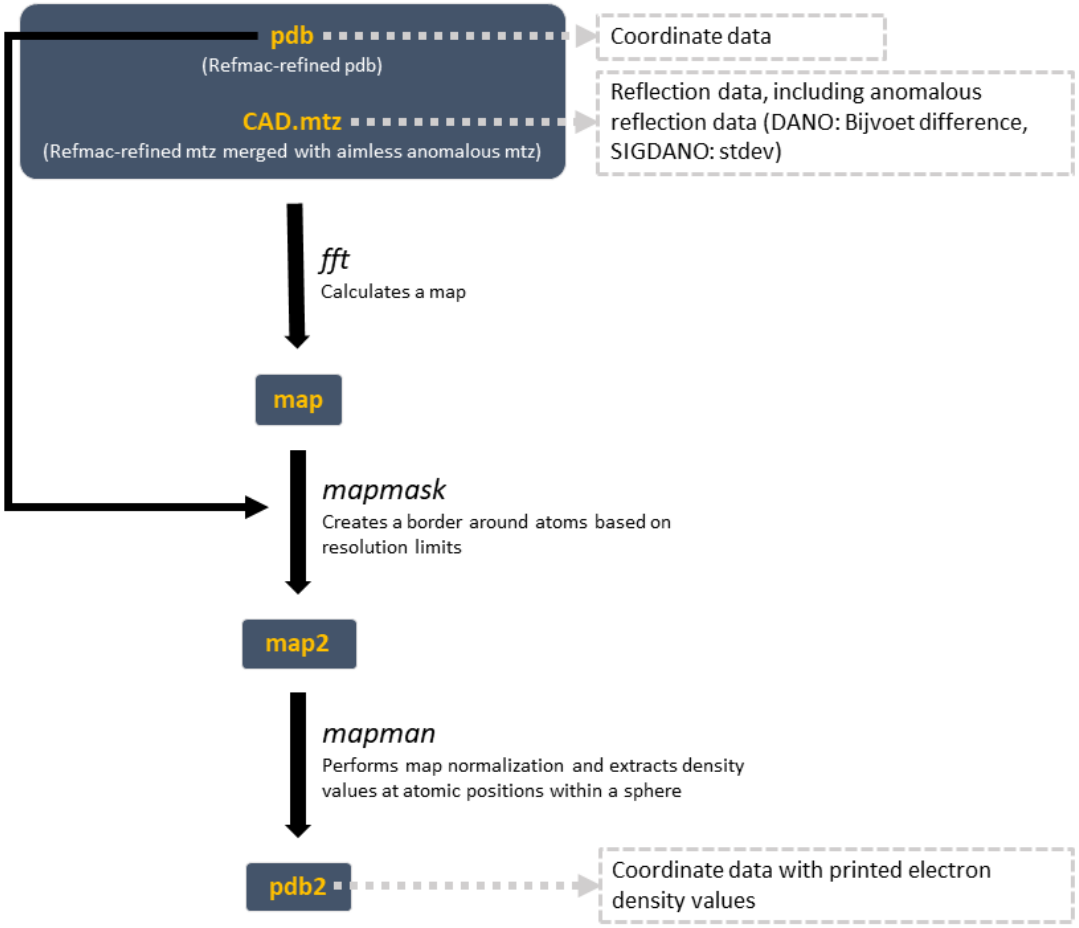


Figure 24. Quantification of Se-occupancies using a Mapman-dependent script. The program allows for input of peak-sphere radius and B-factor cut-off. This script was conceived by and written by Dr. Thomas Spatzal [22,85].

CHAPTER 3: HIGH ENERGY FLUORESCENCE DETECTION

Introduction

The two-fold achievements of high energy resolution fluoresence detection (HERFD) spectroscopy in identifying Mo as Mo^{III} in the FeMo-co, and the high-resolution crystallography and ESEEM determination of carbon as the interstitial FeMo-co atom, inspired the utilization of HERFD on the Se-incorporated FeMo-co. In general, the information gained from the Se K-edge is limited by significant lifetime broadening, which diminishes the resolution possible for this spectroscopic method. However, by using a high-resolution crystal analyzer, it is possible to tune to the Kα1 emission line, and therefore gain high-resolution pre-edge data for the Se K-edge.

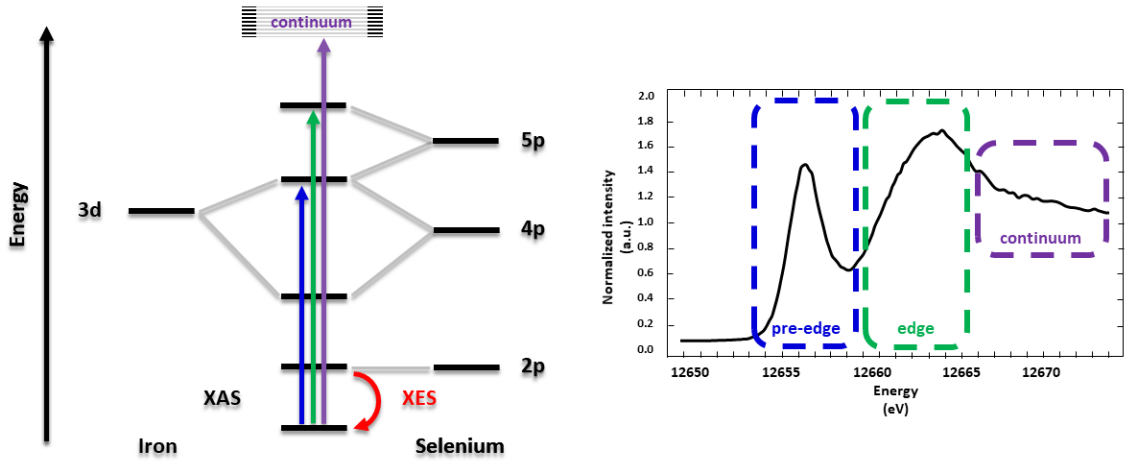


Figure 25: Theory of high energy resolution fluorescence detection. Left) Relative molecular orbitals of iron and selenium, demonstrating the predicted energy transitions for the pre-edge (blue), the edge (green), and the continuum (purple). Right) A sample selenium XAS spectrum with highlighted pre-edge, edge, and continuum. Adapted from DeBeer [93] and Hämäläinen *et al.*, 1991 [94].

The Se K-edge (labeled "edge" in Figure 25) corresponds to a 1s to 5p transition for absorption, and when Se is bound to an open shell transition metal like iron, the pre-edge is expected to be visible in the spectrum. Pre-edges for first row transition metals like iron are caused by 1s to 3d transitions, thereby giving information about the lowest unoccupied (LUMO) state of the metal. The LUMO states then reflect the coordination geometry and electronic configuration of the transition metal. In this system, the selenium 4p and 5p levels interact with the 3d level of iron, and should therefore be visible in the pre-edge region of the spectrum. The usual complication with observing pre-edges is that the features in this region of the spectrum are usually smeared away due to the core-hole lifetime broadening of the 1s electron hole. However with HERFD, the broadening is reduced by detecting the X-ray absorption spectrum using the intensity of the emitted fluorescence in a narrow energy bandwidth [95]. With this method, the resolution then becomes dominated by the 2p rather than the 1s electron core hole.

The pre-edge is expected to shift in both energy and intensity upon changes in the oxidation state of the coordinating irons. More reduced irons will yield higher energy and less intense pre-edges than more oxidized irons. Investigation of the Se K-edge using HERFD spectroscopy serves as an exclusive probe system to study the electronic structure of the Se-substituted FeMo-co, both in a resting state with no external ligands, or in a more reduced form of the ligand-bound enzyme. The CO-bound FeMo-co is an appealing selection for the Se K-edge study, since the binding is stable and is presumed to be coordinated to a more reduced form of the FeMo-co [18].

Methods

Se-labeling for HERFD

Se-labeling for HERFD was as mentioned above, but was subjected to an additional purification step and higher final protein concentrations. For Av1Se2B and Av1Se2BCO: after washing the samples via ultrafiltration with cold size exclusion buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$), the protein was loaded onto a size exclusion column (S200, 26/60, GE Healthcare, 450 mL) which had been pre-equilibrated with degassed size exclusion buffer. The protein was then collected and concentration was

adjusted to either 20 mg/mL or 200 mg/mL for various experiments. Av1-Se2B-CO was constantly bubbled through with CO during the washing steps.

CO-Inhibition of the Se-Labeled FeMo-cofactor

CO-inhibition of Av1-Se was achieved with methods previously described with some adjustments [18,22,86] (Figure 26). The assay mixtures were as described in the previous section, with the headspace of the reaction vials replaced with CO. For these reactions, timepoints were not a factor, so the reaction volumes were scaled up in three-neck round bottom flasks. For crystallography, the reactions were scaled up to be 12 mL total, and for HERFD, the reactions were scaled up to 380 mL. The three-neck flasks were degassed via Schlenk line technique while constantly stirring.

First, Av1Se was tested for activity using the procedure for the acetylene reduction assay. Protein concentrations were either Av2:Av1=2:1, or to conserve Av2 and slow the reaction, Av2:Av1=1:1. Then the headspaces of the three-neck flasks were replaced by CO, and the assay mixtures were allowed to react at room temperature with stirring for 20 minutes. The mixtures were then were concentrated by ultrafiltration with 100,000 kDa MWCO filters and washed 3-5 times with cold size exclusion buffer (50 mM Tris/HCl, pH 7.5; 200 mM NaCl; 5 mM $Na_2S_2O_4$) in the anaerobic chamber. CO was regularly bubbled through the solution until flash-frozen for future use. For crystallography, Av1 concentration was adjusted to 30 mg/mL. For HERFD, Av1 concentration was adjusted to 200 mg/mL.

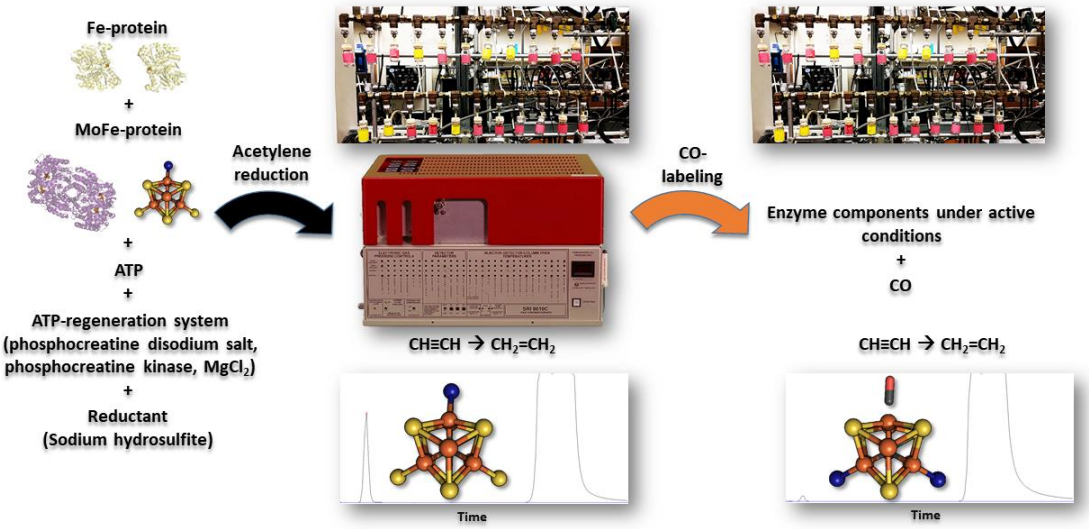


Figure 26. CO-inhibition of the Se-labeled FeMo-cofactor. From the left: activity mixtures with all assay components and protein from the Se-labeling assay, acetylene reduction activity from reactivated Se-labeled protein, and CO-inhibition of Se-labeled protein.

Results

High Energy Fluorescence Detection of Se-incorporated MoFe

To establish whether Se-incorporation could be used as a probe in the nitrogenase FeMo-cofactor, we prepared two versions of the Se-incorporated FeMo-cofactor. First, we prepared a sample with high concentrations of KSeCN in an effort to get full selenium incorporation in the S2B bridging sulfur site (referred to here as Av1Se2B). In order to assess the FeMo-cofactor in a ligand-bound state, we also prepared selenium-incorporated FeMo-cofactor inhibited with CO, referred to here as Av1SeCO. Pre-edge spectra of these samples were collected at SSRL beamline 6-2 using a high-resolution Si(844) crystal analyzer in collaboration with Dr. Justin Henthorn and Professor Serena DeBeer.

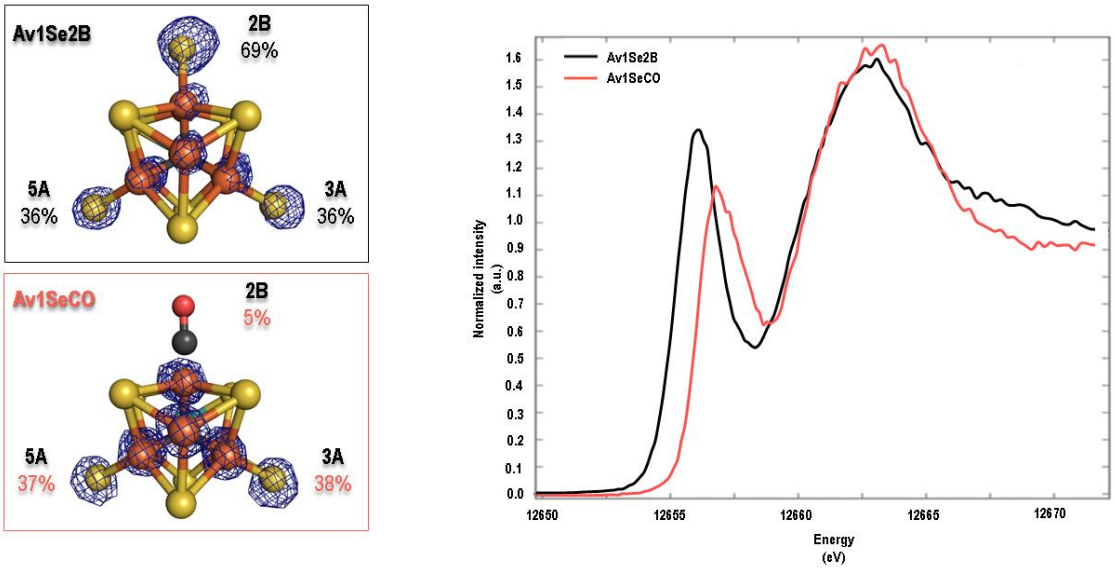


Figure 27. High energy resolution fluorescence detection around the Se-edge. Top left) Av1Se2B structure of the FeMo-cofactor, viewed down the Fe1-C-Mo axis. Bottom left) Av1SeCO structure viewed down the Fe1-C-Mo axis. Right) HERFD spectrum of Av1Se2B and Av1SeCO around the Se-edge.

The spectrum for Av1SeCO has shifted significantly from the Av1Se2B spectrum in both intensity and energy. The energy for Av1SeCO has increased in energy by ~1 eV, while the intensity has decreased by ~0.2 a.u.

Complementary structural data was also obtained from crystals grown from the same preparations of HERFD samples, but diluted to 30 mg/mL to be amenable to crystal growth. The data collection and processing statistics are shown in Table 2 and Table 3 for Av1Se2B and Av1SeCO, respectively.

Table 2: X-ray data collection statistics for the Av1Se2B crystal structure.

Av1Se2B	
Data collection statistics	**High resolution**
Wavelength (Å)	0.97915
Resolution range (Å)	38.84-1.36 (1.38-1.36)
Unique reflections	417029 (10556)
Completeness (%)	95.4 (48.6)
Anomalous completeness (%)	94.9 (42.9)
Multiplicity	6.7 (4.1)
Anomalous multiplicity	3.4 (2.2)
Space group	$P2_1$
Unit cell parameters	
a, b, c	77.74, 131.29, 107.73
α, β, γ	90.00, 108.83, 90.00
R_{merge}	0.268 (5.061)
$R_{p.i.m.}$	0.169 (4.082)
I/σ(I)	4.4 (0.2)
Data processing statistics	
R_{work} / R_{free}	0.226 – 0.257
r.m.s.d. bond lengths (Å)	0.02
r.m.s.d. bond angles (°)	1.962
Average B-factor ($Å^2$)	18.043

Table 3: X-ray data collection statistics for the Av1SeCO crystal structure.

Av1SeCO	
Data collection statistics	**High resolution**
Wavelength (Å)	0.97915
Resolution range (Å)	38.58-1.38 (1.38-1.36)
Unique reflections	408729 (11518)
Completeness (%)	96.2 (54.8)
Anomalous completeness (%)	95.5 (54.8)
Multiplicity	6.8 (4.6)
Anomalous multiplicity	3.4 (2.5)
Space group	$P2_1$
Unit cell parameters	
a, b, c	77.19, 128.00, 107.63
α, β, γ	90.00, 109.15, 90.00
R_{merge}	0.167 (2.711)
$R_{p.i.m.}$	0.105 (2.109)
I/σ(I)	6.4 (0.5)
Data processing statistics	
R_{work} / R_{free}(%)	0.175 / 0.199
r.m.s.d. bond lengths (Å)	0.022
r.m.s.d. bond angles (°)	2.092
Average B-factor ($Å^2$)	15.368

Discussion

High Energy Resolution Fluorescence Detection on a Se-Probe

We sought to establish whether Se-incorporation into the FeMo-cofactor could serve as a probe to the FeMo-cofactor ligand-binding system. The symmetrical nature of the FeMo-cofactor makes it a complicated target for spectroscopy, and therefore we set out

to break that symmetry by incorporating a Se-probe. We also assessed ligand-binding with high energy resolution fluorescence detection (HERFD) and the Se-incorporated, carbon-monoxide inhibited FeMo-cofactor.

The selenium K-edge corresponds to a 1s to 5p transition, and a pre-edge was predicted to be detectable when Se is bound to an open shell transition metal like iron. The pre-edge of Se should also be sensitive to the oxidation state of the iron. Therefore, if the oxidation state of the iron changes, the pre-edge would be expected to shift in both energy and intensity. More oxidized iron is predicted to lower the energy of the Se pre-edge, while more reduced irons are predicted to increase the energy.

We also sought to use the Se-incorporated FeMo-co as a spectator during ligand binding. Since CO can only bind to a more reduced form of the FeMo-co, the Se pre-edge shifts should give us unique insights into the electronic structure of the FeMo-co during turnover.

Indeed, as seen in Figure 27, the CO-bound Se pre-edge has shifted to higher energy and lower intensity. Higher energy corresponds to what was expected of CO binding to a more reduced form of the FeMo-co. The lower intensity agrees with what is expected of mixing and covalency in the pre-edge. That is, intensity of the pre-edge region is expected to increase with greater mixing of electrons. As more electrons are added to the system--- or as it becomes more reduced--- more mixing occurs, and therefore, leads to decreased intensity. On the converse side, if the system becomes more oxidized, covalency would increase, but mixing of electrons would decrease, and intensity would also increase.

Comparison of these results to DFT calculations from the literature poses some interesting questions. Bjornsson *et al., 2017*, performed a rigorous set of DFT calculations on the resting state form of FeMo-co, and arrived at a lowest energy broken-symmetry solution (Figure 28) [96]. It should be noted that Bjornsson and coworkers found three spin isomers for their lowest energy solution, with the one shown in Figure 28 being the most stable solution. The calculations show delocalized Fe(2.5) pairs on Fe2-Fe6, and well as on Fe3-Fe7, and show Fe(3) on Fe4-Fe5.

We presumed that the Se-incorporated FeMo-co (Av1Se2B) would give us insight into the resting state of the FeMo-co, while the Se-incorporated FeMo-co inhibited with

CO (Av1SeCO) would give us insight into the more reduced form. In Av1Se2B, the largest Se signal should arise from the 2B site, where the most Se has incorporated (~70% occupied according to the anomalous calculations). In Av1SeCO, the largest signals arise from two places: from the 5A and 3A sites, where Se has incorporated to ~40%. From the experimental spectrum of Av1Se2B, we would expect the Se to be between two more oxidized irons, since it has shifted to the lowest energy and a greater intensity. Conversely, we would expect the Se's of Av1SeCO to be between more reduced irons.

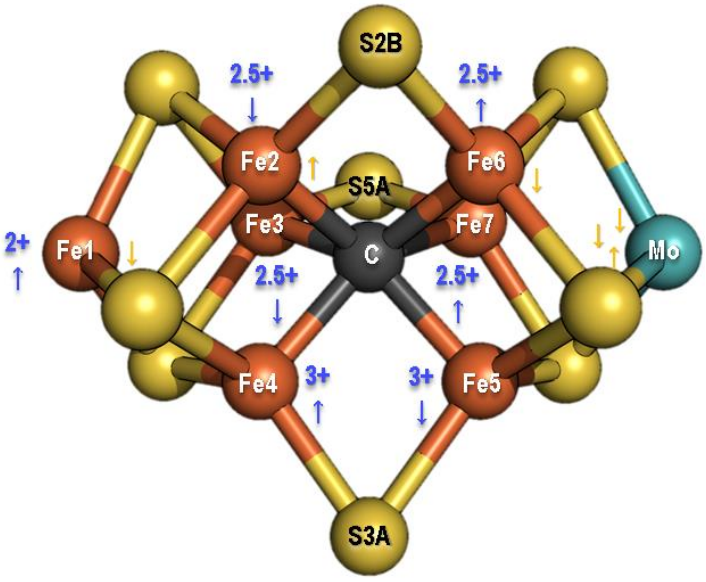


Figure 28. DFT calculations of the FeMo-cofactor oxidation and spin states. *Adapted from Bjornsson et al., 2017* [96]. The calculated oxidation states for the irons is shown in blue text. The majority spin vectors (spin 5/2) for the irons are shown as blue arrows, and the delocalized electron spin vectors (spin 1/2) are shown as yellow arrows.

Interestingly, this prediction does not seem to agree with the DFT-calculated structure in Figure 28. Rather, the calculations suggest that Se in Av1Se2B may encounter more reduced irons (Fe(2.5)), while the Se's in Av1SeCO may encounter more oxidized irons (Fe(2.5) and Fe(3)). These observations are all with the caveat that the structure in Figure 28 is the calculated resting state structure of the FeMo-co. Nevertheless, these

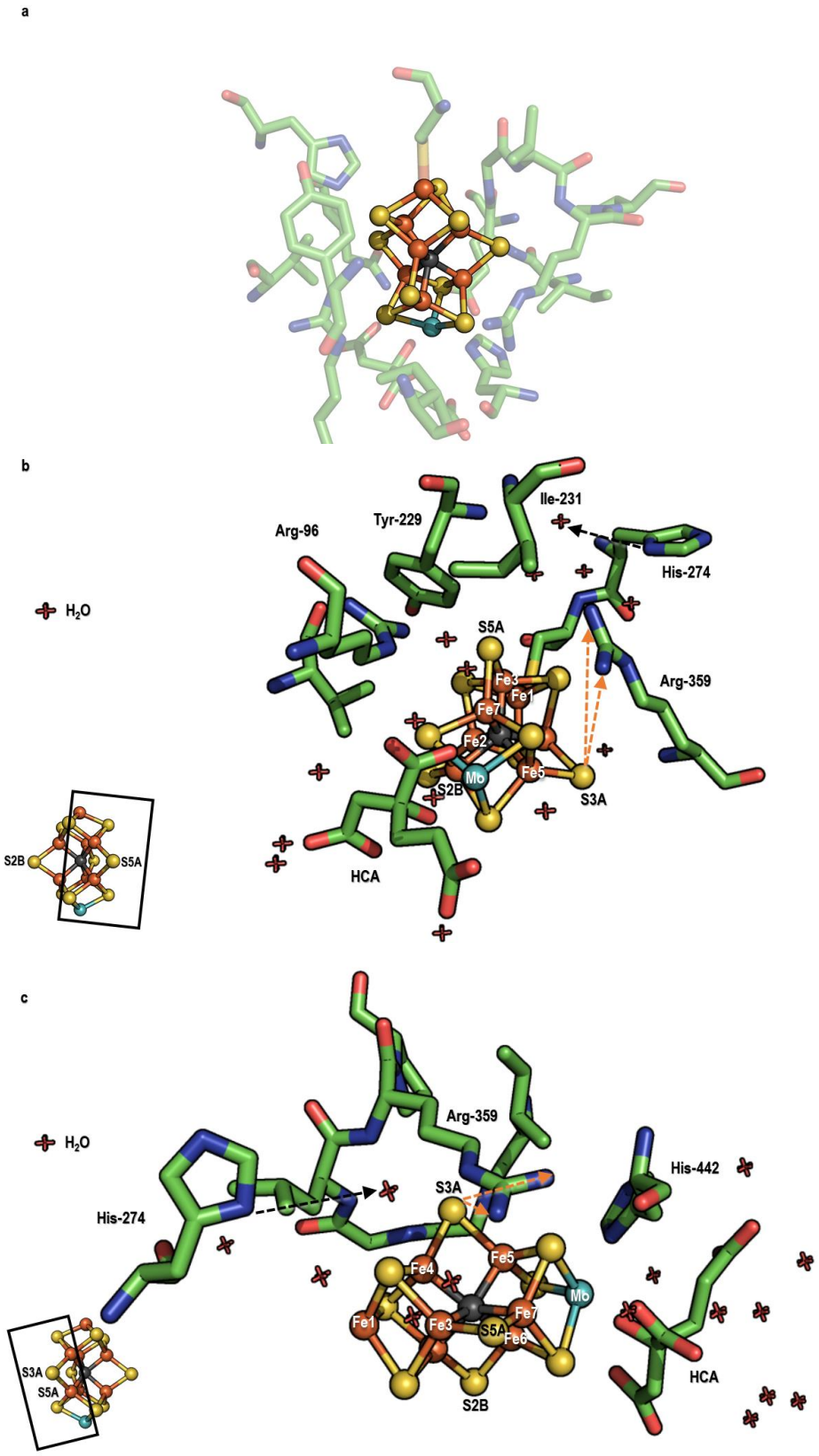
observations introduce some appealing concepts regarding the MoFe-protein in general. That is, the protein environment around the FeMo-co most likely tunes the expected oxidation states of the metal cluster during catalysis.

The S5A and S3A environments have been shown to engage in some interesting changes during turnover at varying pH's, which could help to explain the ligand-binding behavior at the 2B site. In the *Clostridium pasteurianum* component protein 1 (Cp1) at low pH, Arg-359 gets positioned away from the 5A and 3A face of the cofactor (Figure 29bc). S5A then loses its only hydrogen bond, and S3A loses one contact and gains another. Also at low pH, His-274 moves closer to the FeMo-co and displaces a water molecule [97]. These pH changes occur only on the Fe4-Fe5 and Fe3-Fe7 faces of the cofactor, which have potential routes for water access. The proximity to water at the 5A site could provide potential routes for protonation at the 5A bridging sulfide or to nearby waters. Protonation at the 3A bridging sulfide is also possible, but perhaps less likely due to the peptide cage surrounding the S3A site (Figure 29c).

The S2B site also has some possible proton transfer pathways through His-195 and the homocitrate (Figure 30). These well-positioned residues for S2B protonation pose as prospective explanations as to why the S2B site behaves like it's more oxidized than calculations suggest, and may also help to explain why CO and Se preferentially bind at this position, and why the 3A and 5A sites may be more prone to protonation.

Ligands have been observed to bind to other faces than the S2B face on the FeMo-cofactor, notably in vanadium nitrogenase. *Sippel et al.* 2017, observed the binding of a carbonate ligand in the 3A position of FeV-co [76], which in molybdenum nitrogenase is obstructed by the peptide cage and by Pro-360 in a corresponding loop region.

Figure 29 (next page). The protein environment around the FeMo-cofactor. a) Overview of the protein environment around the FeMo-cofactor. b) View from the S5A face of the FeMo-cofactor. c) View from the S3A face of the FeMo-cofactor. Inset orients the viewer to FeMo-cofactor face. Orange dotted lines show direction of Arg-359 movement away from FeMo-co at low pH. Black dotted lines show the movement of His-274 towards the FeMo-co at low pH. PDB: 3U7Q.


a
b
Arg-96
Tyr-229
Ile-231
His-274
H2O
S5A
Fe3
Fe1
Fe7
Fe2
Mo
Fe5
S2B
S3A
Arg-359
HCA
S2B
S5A
c
H2O
Arg-359
His-442
S3A
His-274
Fe4
Fe5
Mo
Fe1
Fe3
S5A
Fe7
Fe6
S2B
HCA
S3A
S5A

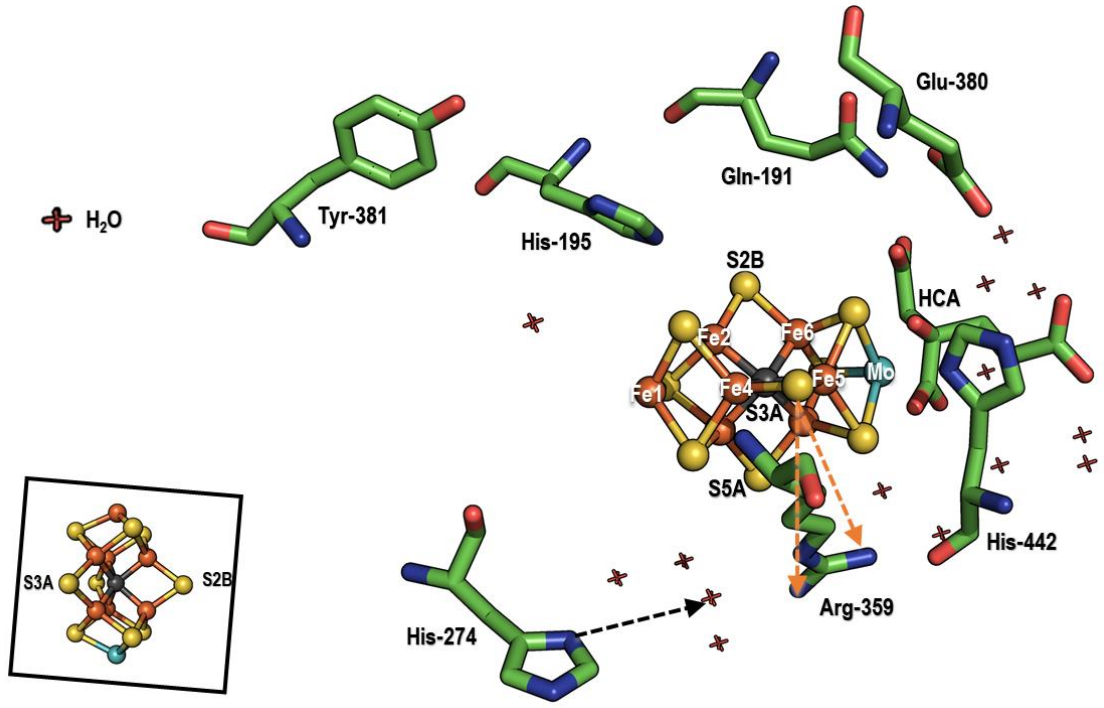


Figure 30. The protein environment around the S2B site of the FeMo-cofactor. Inset orients the viewer to the FeMo-co face. Orange dotted lines show direction of Arg-359 movement away from FeMo-co at low pH. Black dotted lines show the movement of His-274 towards the FeMo-co at low pH. PDB: 3U7Q.

With this study, we have established the Se-incorporated FeMo-co as a useful tool in probing the FeMo-co oxidation state and electronic configuration. We have also established the Se-incorporated FeMo-co as a tool to observe ligand binding during nitrogenase turnover. Future studies, including more rigorous DFT calculations, and comparisons to model compounds, will further the uses of this new and exciting technique.

CHAPTER 4: SELENIUM INCORPORATION AND METHANE PRODUCTION BY NITROGENASE

Introduction

Kinetic Studies of X-CN Compounds

As mentioned in the main introduction, several kinetic studies using assorted cyanide compounds are relevant to this study. An early study was conducted by Li *et al.* in 1982, examining sodium cyanide (NaCN) as a substrate for nitrogenase [98]. NaCN as a substrate releases methane as a product, with the proposed intermediates of: methanimine (CH_2NH), formaldehyde (H_2CO), methylamine (CH_3NH_2), and ammonia (NH_3) (Figure 31).

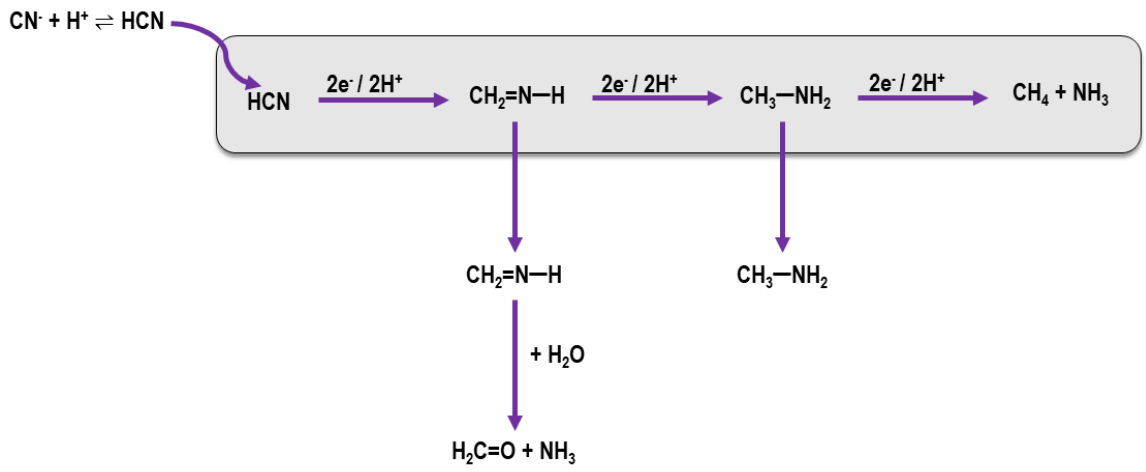


Figure 31: Methane production and proposed intermediates by nitrogenase using sodium cyanide (NaCN) as a substrate [98]. The grey box denotes the substrate bound to the FeMo-cofactor. Outside of the grey box, the intermediate has been released.

Another applicable study utilizing compounds related to cyanide was conducted by Rasche *et al.* in 1997, examining thiocyanate (KSCN) as a substrate for nitrogenase [5]. The proposed mechanism is similar to that of NaCN (Figure 31), except with an additional $2e^- / 2H^+$ step and release of hydrogen sulfide (H_2S) (Figure 32).

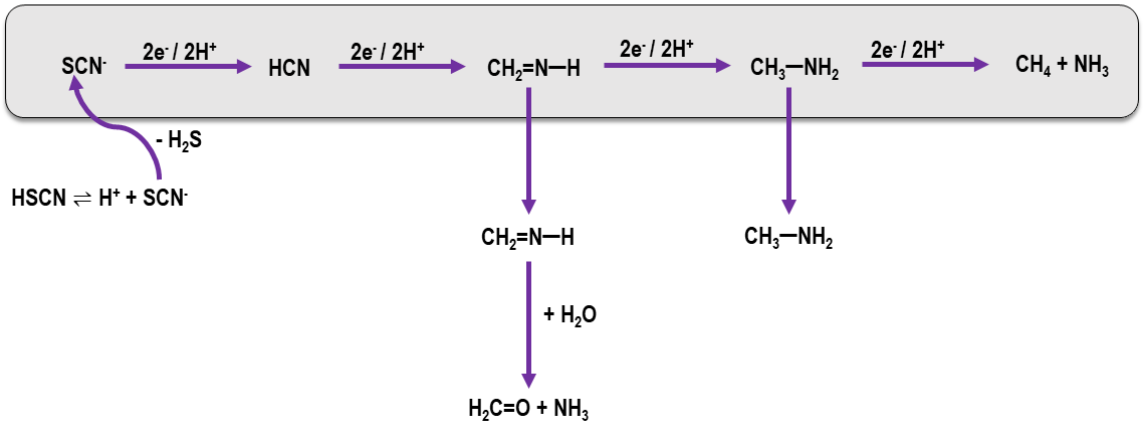


Figure 32: Production formation by the nitrogenase catalyzed reduction of potassium thiocyanate (KSCN) as a substrate[5].

Most recently, a kinetic analysis of KSeCN was conducted by Spatzal and Perez *et al., 2015* [22]. A summary of the observed specific activities for all three substrates is summarized in Figure 33. It is interesting to compare the specific activities of the three substrates despite the differences in component protein ratio (Av2:Av1). It is feasible that the high component ratio for the NaCN studies could account for the increase in specific activity, owing to the increased electron flux through the enzyme system.

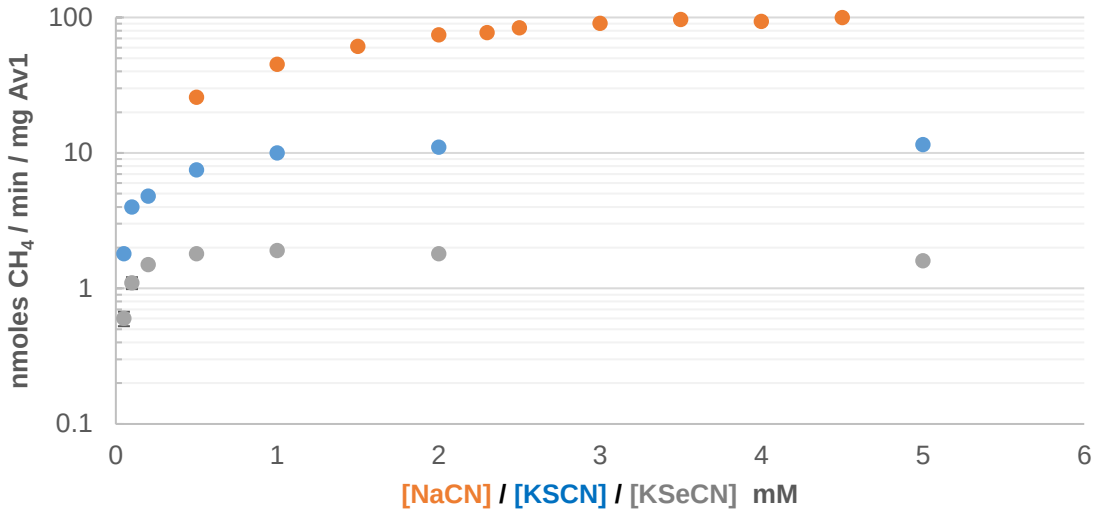


Figure 33: NaCN, KSCN, and KSeCN as substrates for nitrogenase [5,22,98]. The specific activities were all adjusted from the original data to be in nmoles CH_4 / min / mg Av1. NaCN studies were performed with at CR = 8, KSCN at CR = 0.8, and KSeCN at CR = 2.

The component ratios (Av2:Av1) used for the original studies were 8 for NaCN, 0.8 for KSCN, and 2 for KSeCN. The high component ratio for NaCN could presumably explain the differences in specificity activity owing to the increased electron flux through the protein system. However, it doesn't explain the differences in specific activity between the KSCN and KSeCN studies. The three substrates also have substantial differences in pKa's (Table 2), which could also explain the differences in reactivity. However, more exact measurements of these pKa's would be necessary to make direct comparisons. Pertinent to this study is the substitution of X in X-CN, which will be useful spectroscopically and crystallographically in the study of nitrogenase turnover.

Table 4: pKa's of cyanic acid (HCN), thiocyanic acid (HSCN), and selenocyanic acid (HSeCN)

Substrate (acid conjugate)	pKa
NaCN (HCN)	9.2 [99]
KSCN (HSCN)	-2 [100–102]
KSeCN (HSeCN)	<<1 [101,102]

Table 4: Cyanide-related substrates of nitrogenase with associated pKa's.

Methods

Methane Production using KSeCN as a Substrate

Methane production was monitored via gas chromatography from activity assay mixtures using KSeCN as a substrate. The assay mixtures were identical to that of the proton reduction assay, except with varying concentrations of KSeCN. Methane was measured from the headspace of the assay mixtures, and calibration curves were generated using pure methane gas.

Analytical ICP-MS of Se-Incorporated Av1

To further quantify methane production with SeCN as a substrate, we directly quantified selenium incorporation from protein crystals using inductively coupled plasma mass spectrometry (ICP-MS). 21 conditions for Se-incorporation were utilized to yield crystals with various levels of Se-incorporation. Protein crystals grown from assays for Se-incorporation were then re-suspended and dissolved in approximately 50 uL of ddH_2O. The estimated protein concentration for ICP-MS was about 0.15 mg/mL, harvesting crystals from 10 of 12 wells in a given condition. The resuspended crystal and water mixture was then added to 3 mL of 2% nitric acid (HNO_3), and analyzed using a quadrupole-based inductively coupled plasma-mass spectrometer (Agilent 8800 ICP-QQQ) at the Environmental Analysis Center at Caltech. Samples were analyzed in helium mode to avoid any argon-argon dimers that may form in the plasma. Standards were generated by comparing ion counts from prepared selenium, molybdenum, and iron standards, all prepared in 2% HNO_3.

Results

Methane Production with KSeCN as a Substrate

Once we had established a robust method for selenium incorporation into the FeMo-cofactor, we decided to focus on the product profile of the nitrogenase enzyme system using KSeCN as a substrate. We decided to focus on methane, the observed final product of nitrogenase turnover with a cyanide compound (Figure 31, Figure 32). Methane is readily quantifiable in our established assay conditions, and samples can be taken directly from the headspace of assay vials.

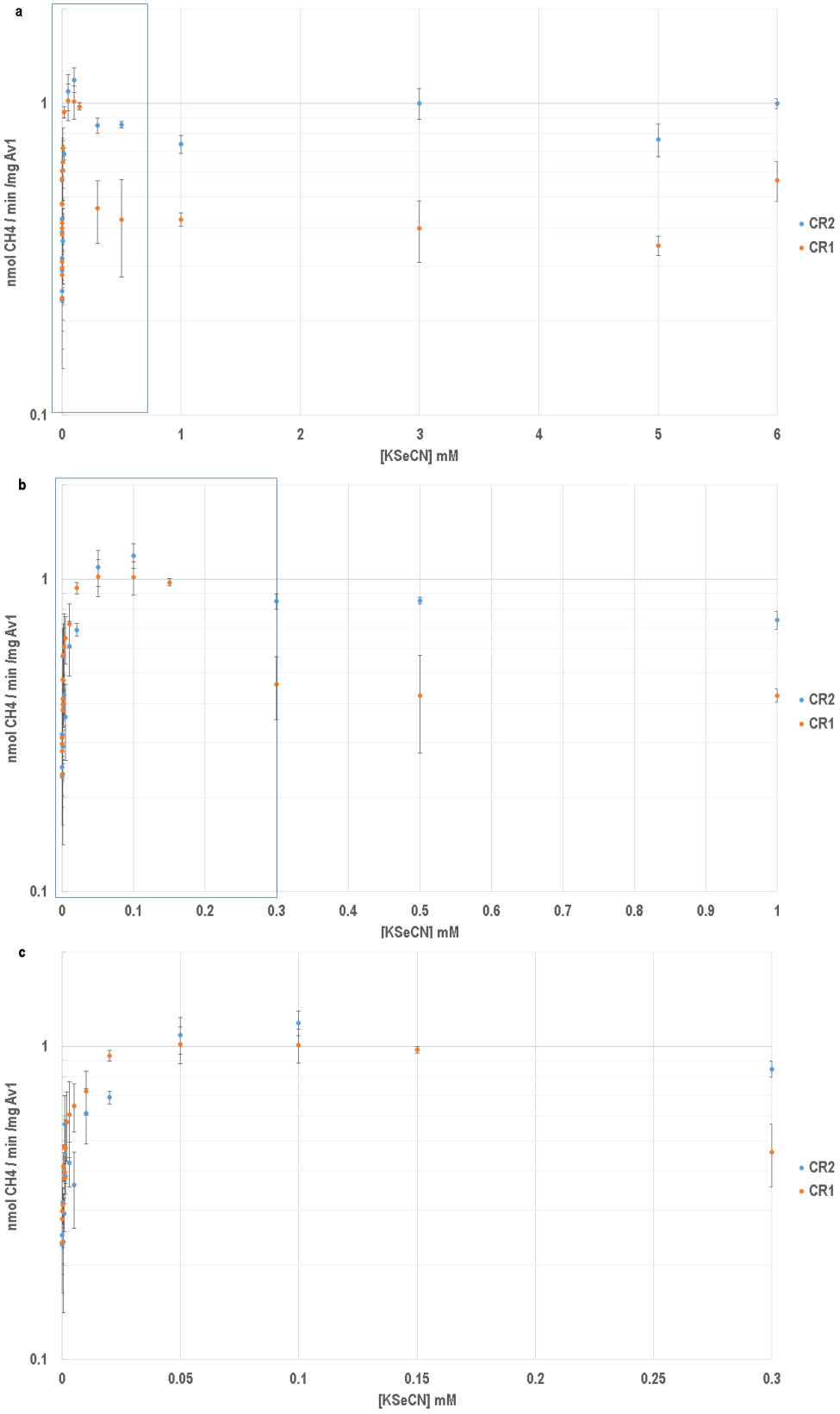
First, we performed titrations of various concentrations of KSeCN (Figure 34). To complement the methane studies, we also looked at proton reduction and ethylene production in the presence of KSeCN (Figure 35).

As can be observed from Figure 34, methane production increases until approximately 0.1 mM KSeCN in both component ratios. Methane production then decreases consistently between the component ratios at 0.3 mM KSeCN, and remains at a lower basal level out to 6 mM KSeCN. A maximum production of CH_4 was observed at ~1.2 (± 0.10) nmol CH_4 / min / mg Av1 for CR2, and ~1.0 (± 0.10) nmol / min / mg Av1 for CR1. The basal level of methane production for CR1 was lower than the basal level of methane production for CR2, averaging at ~0.85 (± 0.05) nmol CH_4 / min / mg Av1 for CR2, and 0.42 (± 0.10) nmol CH_4 / min / mg Av1 for CR1. The standard deviations reported are from triplicate experiments.

Owing to the similarity between the two sets of component ratios for methane production, we decided to carry out proton reduction assays only using CR2 (Figure 35a). Maximum H_2 production was observed before any addition of KSeCN, at ~1881 (± 228) nmol H_2 / min / mg Av1 at 0 mM KSeCN. H_2 production continued to decrease until a basal background level of ~ 194 (± 83) nmol H_2 / min / mg Av1 at 0.01 mM KSeCN. The standard deviations reported are from duplicate experiments.

We also performed parallel acetylene reduction experiments to monitor the decrease in ethylene upon KSeCN titration (Figure 35b). Maximum ethylene production was observed at ~1694 (± 117) nmol C_2H_4 / min / mg Av1 at 0 mM KSeCN. C_2H_4 production continued to decrease out to 5 mM KSeCN, where the specific activity was ~82 (± 0.3) nmol C_2H_4 / min / mg Av1.

Figure 34 (next page). Methane production with KSeCN titration with the nitrogenase active enzyme system. KSeCN was titrated into assay mixtures containing the ATP-regeneration system, and the nitrogenase component proteins in either a 2:1 component ratio (CR) (Av2:Av1) or a 1:1 component ratio. a) KSeCN titration (mM) up to 6 mM with nitrogenase proteins in CR2 and CR1. b) Inset of a, showing the titration of KSeCN up to 1 mM. c) Inset of b, showing the titration of KSeCN up to 0.3 mM. Methane production is shown in nmol CH_4 / min / mg Av1.


a
nmol CH4 / min /mg Av1
1
0.1
0
1
2
3
4
5
6
[KSeCN] mM
CR2
CR1
b
nmol CH4 / min /mg Av1
1
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
[KSeCN] mM
CR2
CR1
c
nmol CH4 / min /mg Av1
1
0.1
0
0.05
0.1
0.15
0.2
0.25
0.3
[KSeCN] mM
CR2
CR1

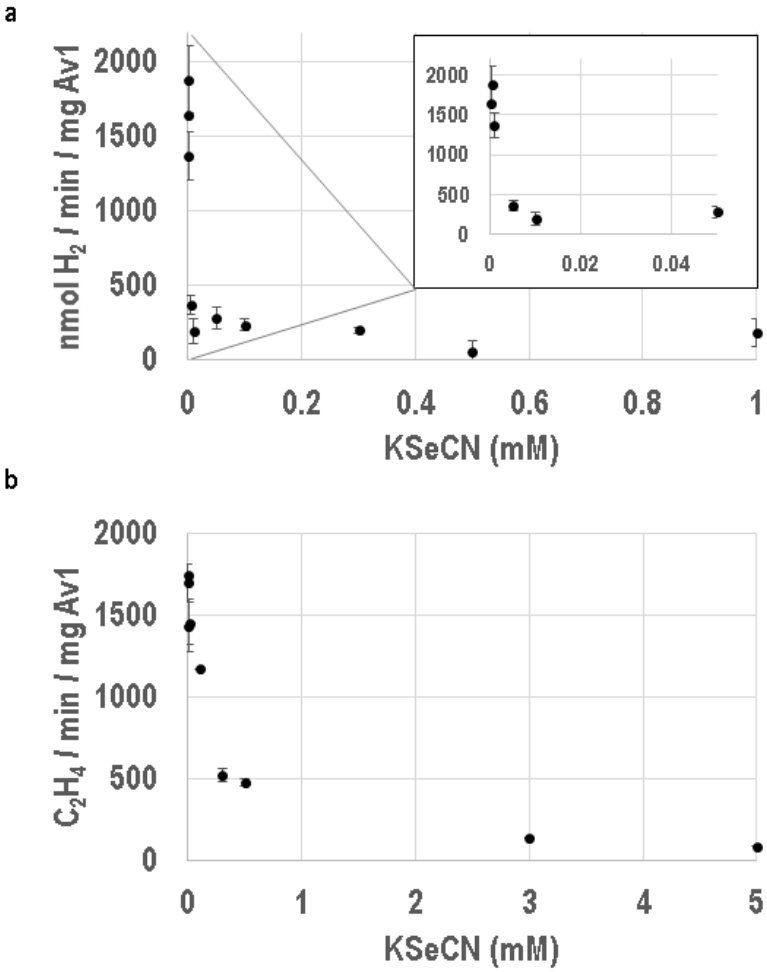


Figure 35. Hydrogen and ethylene production with KSeCN titration with the nitrogenase active enzyme system. KSeCN was titrated into assay mixtures containing the ATP-regeneration system, and the nitrogenase component proteins in a 2:1 component ratio (CR) (Av2:Av1). a) KSeCN titration (mM) up to 1 mM with nitrogenase proteins in CR2, quantifying H_2 production with no acetylene (C_2H_2) present in nmol / min / mg Av1. Inset showing KSeCN up to 0.05 mM. b) KSeCN titration (mM) up to 5 mM with nitrogenase proteins in CR2, quantifying C_2H_4 production from acetylene (C_2H_2) in nmol / min / mg Av1.

We also attempted to do various timepoints with excess KSeCN (1 mM), but the results were less consistent (data summarized in Figure 39 and Figure 40).

Se-Occupancy as a Function of Methane Production

Next we decided to assess whether selenium incorporation into the FeMo-cofactor could be correlated to methane production. To achieve this, we performed both KSeCN titrations, and timepoints with excess KSeCN. Activity assays at various concentrations of KSeCN and selected timepoints were pooled together to obtain enough material for crystallization. Rather than quenching with citric acid, we instead devised a freeze-quench technique. First, the headspace was used to measure methane production at each concentration of KSeCN and each timepoint, after which the assay mixtures were immediately frozen in liquid nitrogen. To avoid rapid expansion of nitrogen gas upon thawing of the samples, we pierced the septa of each vial under liquid nitrogen to purge any nitrogen that may have seeped through. The samples were then thawed in the anaerobic chamber, pooled, and concentrated over ice.

Eight conditions at KSeCN concentrations of 0.0003, 0.0008, 0.003, 0.005, 0.01, 0.05, 0.1, and 0.3 mM were crystallized and assessed for Se anomalous occupancies. Six conditions at excess KSeCN (1 mM) at timepoints of 5 sec, 10 sec, 30 sec, 1 min, 3 min, and 10 min were also crystallized and assessed for anomalous occupancies. Two strategies were employed to assess selenium occupancies.

The first strategy to assess anomalous occupancies was to calculate Se-occupancies using a modified script from [22]. At 12662 eV, an energy experimentally determined by MAD scans to be above the Se K-edge, Fe atoms in the MoFe-protein were assumed to be present at 100% occupancy. These atoms were then used as internal standards to estimate Se-occupancy using the relationship:

$$f = f_0 + \Delta f' + i\Delta f''.$$

At 12662 eV, the atoms of interest have theoretical scattering factors $f_{Fe}'' = 1.5$, $f_{Se}'' = 3.84$, and $f_S'' = 0.24$ [103–107]. However, 12662 eV is very close to the Se edge, so that the absorption curve may be influenced by chemical effects [108,109]. Indeed, we observed significant shifts in the pre-edge of the Se-absorption spectrum in our previously mentioned

HERFD experiments. To experimentally determine f_{Se}" at 12662 eV, we performed two absorption edge scans on crystals from KSeCN titration assays at 0.3 mM KSeCN (Figure 36). As can be observed from the scans, the f_{Se}" at 12662 eV varies quite substantially, from 4.9 for one measurement, and 3.9 for a second measurement. The average f_{Se}" at 12662 was determined to be ~4.42, which is significantly higher than the theoretical predictions for an isolated Se atom.

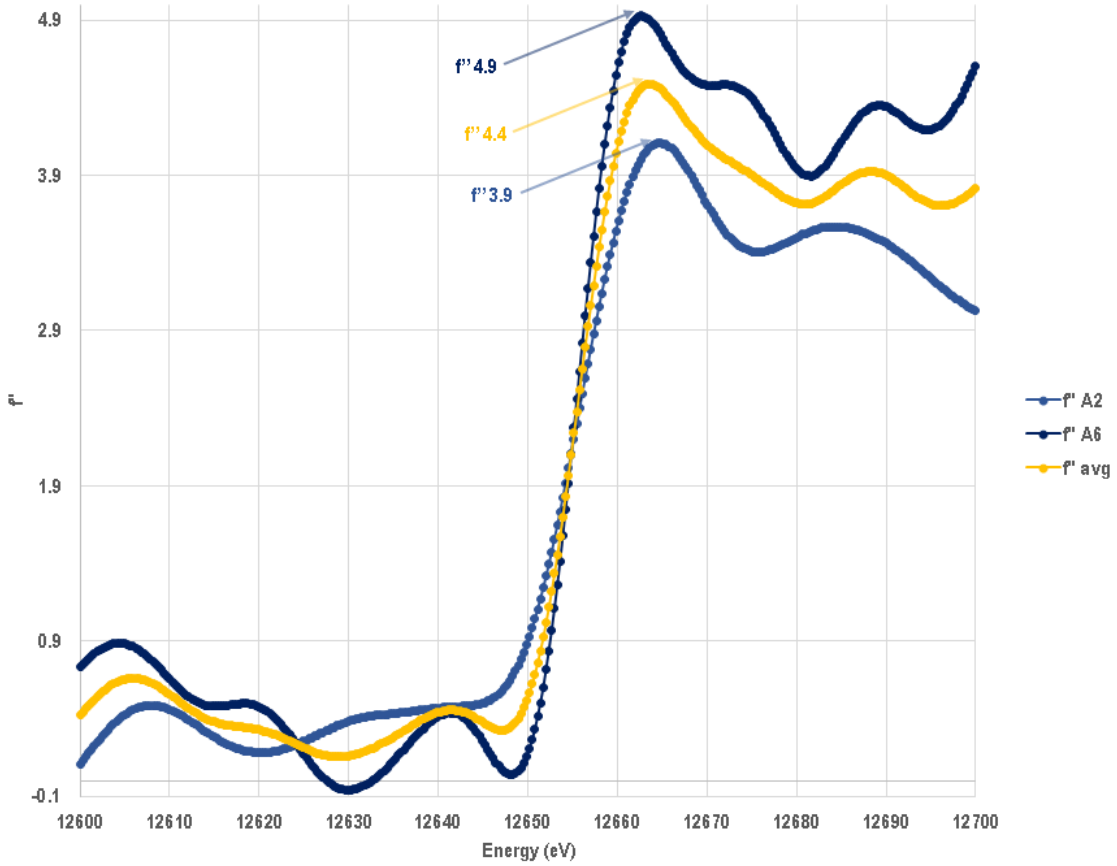


Figure 36. Se absorption edge scans. Two scans were performed on two separate crystals (A2 and A6) at SSRL beamline 12-2 at and around the Se-edge. The f'' at each wavelength was determined using the SSRL-provided script. The two scans were then averaged at every point for a third calculated curve.

We used $f_{Se}" = 4.42$ for our first strategy to calculate Se-anomalous occupancies at the belt-sulfur positions within the FeMo-cofactor (Figure 37). Se-anomalous density maps were calculated based on the data collected at 12662 eV. Refmac-refined mtz files were merged with anomalous mtz files using the CCP4-embedded program CAD. Maps were generated using fast-Fourier transform (FFT) of the merged mtz files and the Refmac-refined pdb coordinates. The CCP4 program Mapmask was used to process the model based on the

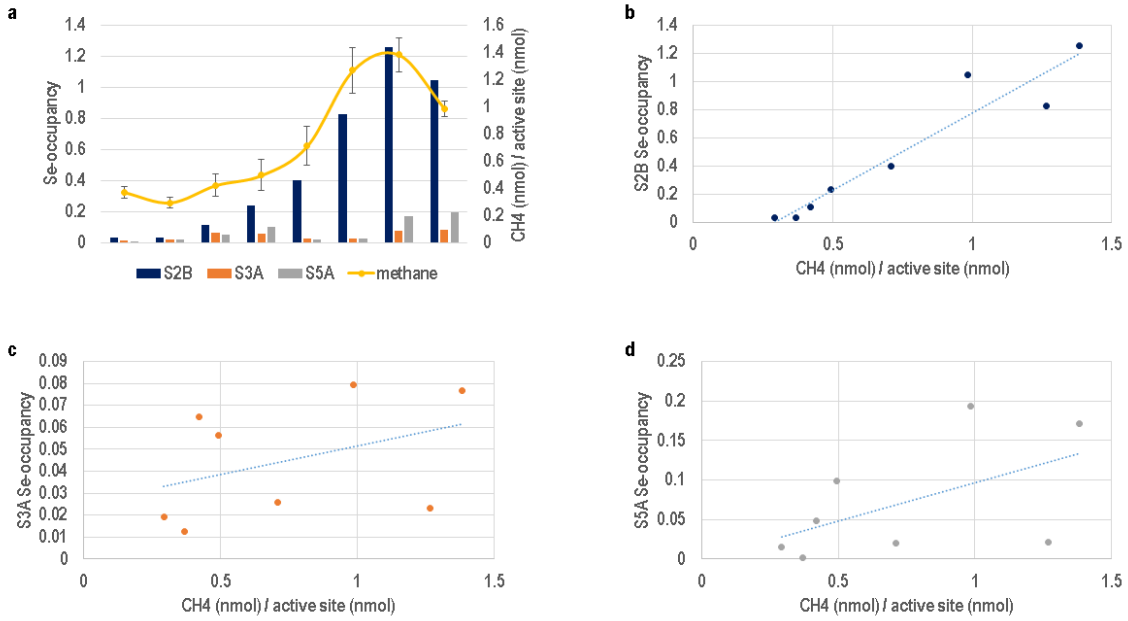


Figure 37. Se-occupancies correlated with methane production at various concentrations. a) Total selenium occupancy for each belt sulfur atom is displayed in clustered bars with the left primary axis, with S2B Se-occupancy in dark blue, S3A Se-occupancy in orange, and S5A Se-occupancy in gray. Each clustered bar set corresponds to 0.0003 mM, 0.0008 mM, 0.003 mM, 0.005 mM, 0.01 mM, 0.05 mM, 0.1 mM, and 0.3 mM KSeCN. Total methane (nmol) production per nmol of active site for each dataset is shown on the right primary axis in yellow. b) S2B Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.907$. c) S3A Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.160$. d) S5A Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.300$. All anomalous occupancies were calculated with $f_{Se}" = 4.42$.

data resolution limits. Finally, the program Mapman performs map normalization and extracts the density values on an arbitrary scale at all atomic positions within a set sphere [92]. The structures were then rendered for publication in PYMOL.

As can be observed in Figure 37, methane production correlates very well with Se-occupancy into the S2B site of the FeMo-cofactor, and Se-occupancy in S3A and S5A is poorly correlated with methane production, which can also be attributed to low incorporation at these sites. Visual inspection of the anomalous maps also shows increases in Se-incorporation with higher methane production (Figure 38).

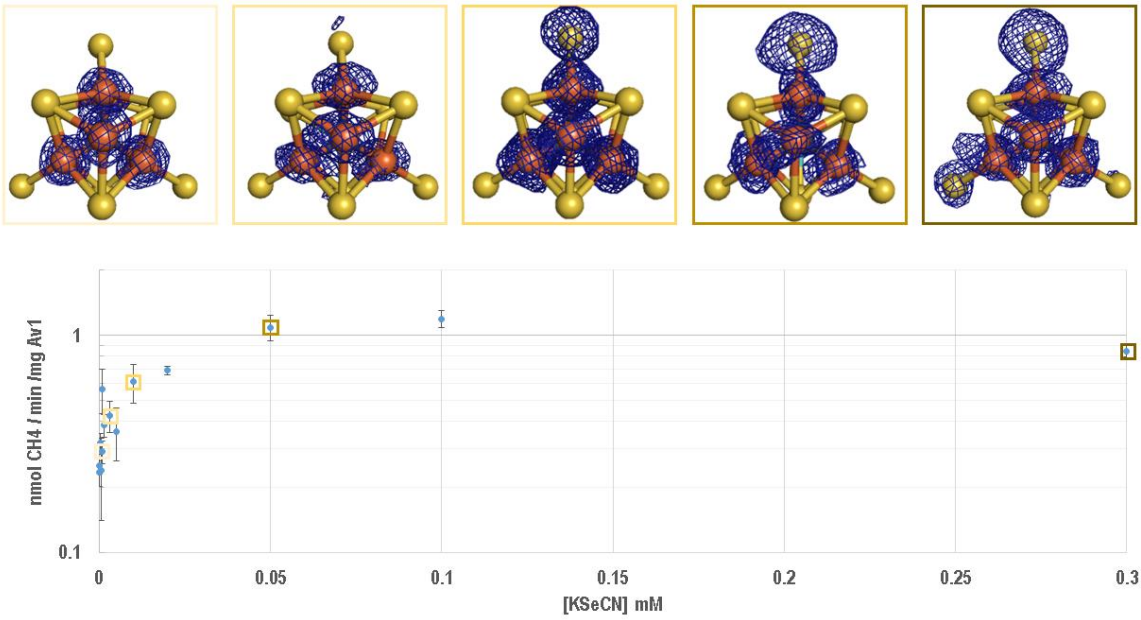


Figure 38. Se-occupancies at various concentrations of KSeCN. From top left to top right: 0.0008 mM (resolution estimate 1.88 Å), 0.003 mM (2.40 Å), 0.01 mM (1.81 Å), 0.05 mM (2.19 Å), and 0.3 mM (1.94 Å) KSeCN. The cofactor is viewed down the Fe1-C-Mo axis, with S2B as the top bridging sulfur, and then clockwise S3A, and S5A. Bottom: titration of KSeCN from which protein was directly sourced by freeze-quench for crystallography. CH_4 production is shown in nmol / min / mg Av1. The corresponding crystal structures are highlighted by shades of yellow, with least KSeCN in light yellow, and most KSeCN in dark gold. Maps are scaled to 4.0 σ.

Interestingly, the anomalous Se-signal at the S2B bridging sulfur site follows methane production very closely. Very low supply of KSeCN (0.0008 mM) showed minimal methane production (0.29 ± 0.04 nmol / nmol active site), and also showed the least Se-incorporation (3.3 ± 2%) at the S2B site. Maximal methane production was observed at 0.1 mM KSeCN (1.38 ± 0.1 nmol / nmol active site), with a decrease observed at 0.3 mM KSeCN (0.99 ± 0.06 nmol / nmol active site). Near maximal methane production, Se-incorporation at the S2B site is also at its highest (125%). When methane production begins to decrease at 0.3 mM KSeCN, Se-incorporation also decreases (107 ± 3%) in the S2B site, with occupancy also observed at the S5A site (29 ± 14%) and in S3A (16 ± 11%). The occupancies in S3A and S5A showed some interesting trends, but were also less consistent than the occupancies observed in S2B undoubtedly due to the low occupancies. The standard deviations reported here are from two different experiments with two different final crystal datasets. We attribute the percentages above 100% to the uncertainty in the anomalous scattering factor as described with Figure 36.

Various timepoints with excess KSeCN (1 mM) were also attempted, but the results were less consistent (Figure 39 and Figure 40). Timed trials were performed at 5 seconds, 10 seconds, 30 seconds, 1 minute, 3 minutes, and 10 minutes. Interestingly, Se-occupancy in the S5A and S3A sites during timed trials seem to be more correlated to methane production than Se-occupancy in the S2B site, presumably because the S2B site has long been occupied. Also of note is that the S2B site under excess KSeCN gets incorporated with selenium to nearly full occupancy (~82%) within 30 seconds. Anomalous data at 5 seconds showed about 10% incorporation at S2B, and ~72% at 10 seconds. These data are not included due to poor crystal data.

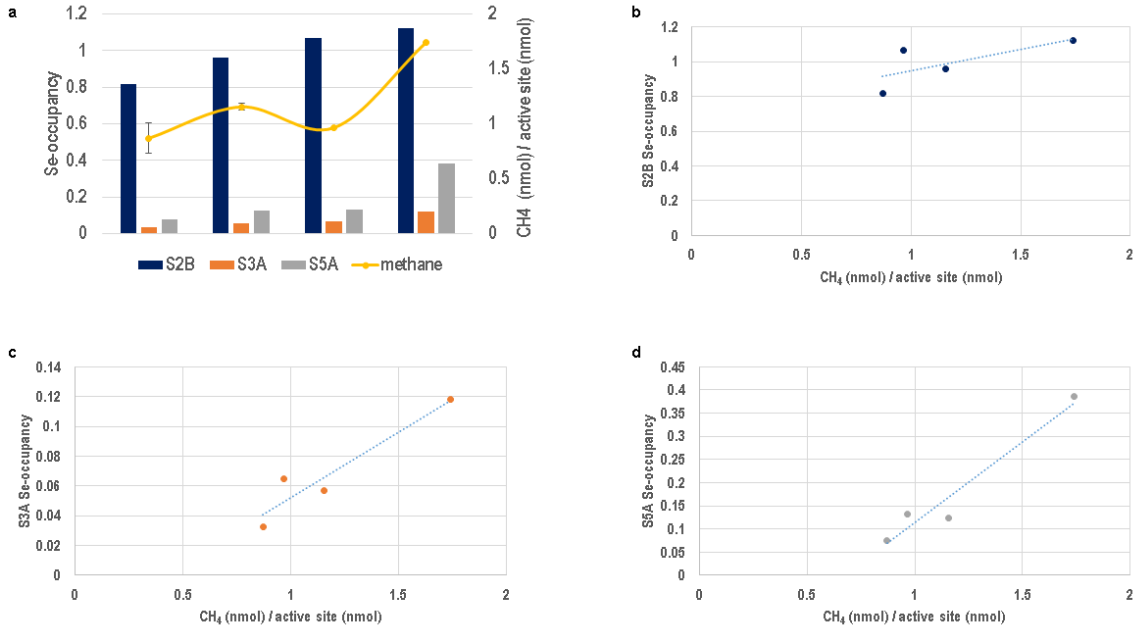


Figure 39. Se-occupancies correlated with methane production at various timepoints and excess KSeCN (1 mM). a) Total selenium occupancy for each belt sulfur atom is displayed in clustered bars with the left primary axis, with S2B Se-occupancy in dark blue, S3A Se-occupancy in orange, and S5A Se-occupancy in gray. Each clustered bar from left to right corresponds to 30 seconds, 1 minute, 3 minute, and 10 minute timepoints. Total methane production (nmol) per nmol of active site for each dataset is shown on the right primary axis in yellow. b) S2B Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.509$. c) S3A Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.891$. d) S5A Se-occupancy correlated with nmol of CH_4 produced per nmol of active site with $R^2 = 0.947$.

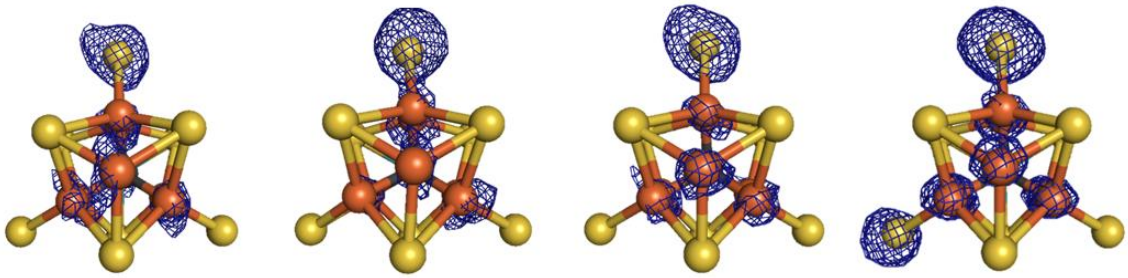


Figure 40. Se-occupancies with excess (1 mM) KSeCN at various timepoints. From left to right, 30 seconds (resolution estimate 1.94 Å), 1 minute (1.97 Å), 3 minute (2.05 Å), and 10 minute (2.08 Å) assays with 1 mM KSeCN. The FeMo-cofactor is being viewed down the Fe1-C-Mo axis, with S2B as the top bridging sulfur, and then clockwise is S3A and S5A. Maps are scaled to 4.0 – 6.0 σ.

As can be seen in Figure 39 and Figure 40, Se-incorporation stays relatively constant in S2B, fluctuating from ~82% to ~112% over the four timepoints. Se-occupancy in S5A fluctuations between ~7% at 1 minute and 13% at 3 minutes. By 10 minutes, Se-occupancy at S5A is at ~ 38%. During all timepoints, Se-occupancy at S3A fluctuated between ~3% and ~12%.

Analytical ICP-MS of Se-Incorporation into Av1

The second method for assessing selenium occupancy is analytically assessing Se-incorporation via ICP-MS. First, samples were assessed for quality by Mo:Fe ratio. The Mo:Fe ratio in intact Av1 protein is normally about 0.06, accounting for two Mo and 30 Fe per Av1 protein. Any sample that significantly deviated from this number was not used for Se-quantification (Figure 41c).

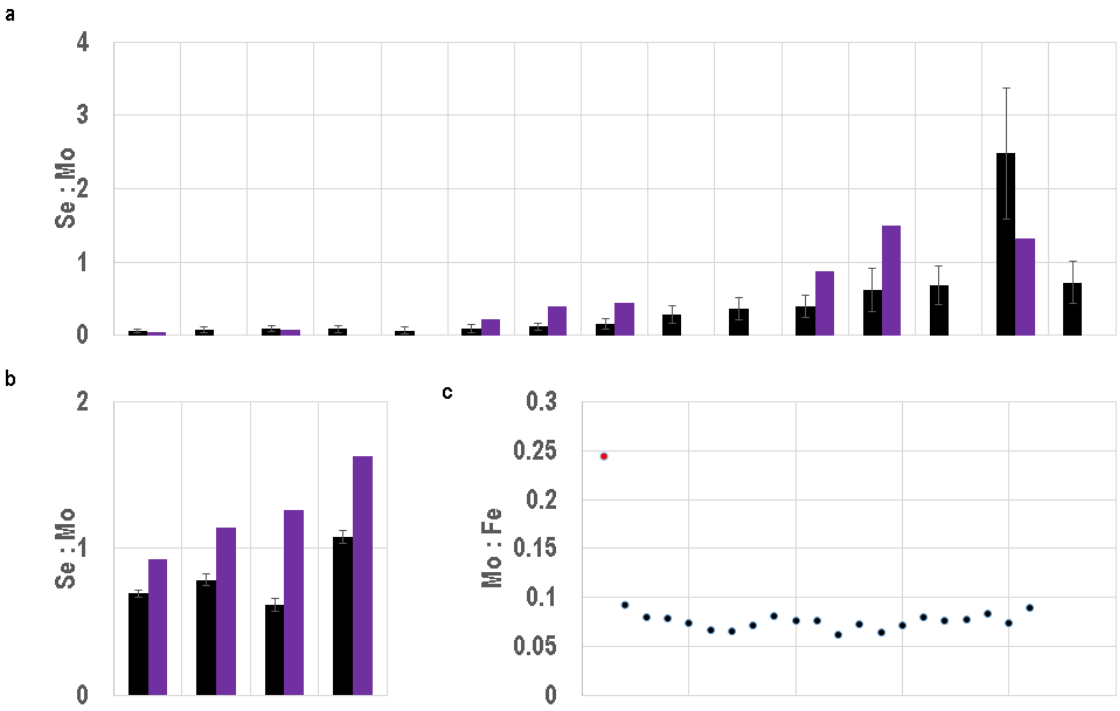


Figure 41. Comparison of Se:Mo ratios in the anomalous quantification experiment and ICP-MS experiment. a) Se:Mo ratios from KSeCN titration (each clustered group corresponds to 0.0003, 0.0005, 0.0008, 0.001, 0.0015, 0.003, 0.005, 0.01, 0.02, 0.03, 0.05, 0.1, 0.15, 0.3, and 0.5 mM KSeCN) is shown for the anomalous quantification in the crystals (purple) and for the crystals subjected to ICP-MS (black). b) Se:Mo ratios from excess KSeCN (1 mM) (each clustered group correspond to 30 seconds, 1 min, 3 min, and 10 min) is shown for the anomalous quantification in the crystals (purple) and for the crystals subjected to ICP-MS (black). c) Mo to Fe ratios to assess viable protein for Se-incorporation. 21 crystal samples of various Se-incorporation were analyzed via ICP-MS. The outlier that was not used for Se-quantification is highlighted in red. All other samples were used to assess Se-incorporation.

The analytical results were then compared to the anomalous results, with the analytical results giving the representative error bars (Figure 41). A trend for increasing Se:Mo ratios can be seen in both the anomalous quantification and the ICP-MS data, with a peak in Se:Mo at 0.1 mM for the anomalous data, and at 0.3 mM for the ICP-MS data. We

would expect the overall Se:Mo ratio to increase until a point and stabilize, and overall the results were within error.

The Se:Mo ratios in the timed trials with excess (1 mM) KSeCN generally increase with anomalous quantification, and fluctuate slightly within the ICP-MS data. The timed trial data are only slightly consistent with each other between the anomalous quantification and the ICP-MS data. The anomalous data appear to be either overestimated in the anomalous data, or underestimated in the ICP-MS data.

Discussion

Methane Turnover and Se-Incorporation

After investigating the spectroscopic implications of Se-incorporation into the FeMo-cofactor, we then turned to crystallography and kinetics to explore how selenium is integrated and how incorporation is related to methane production. Figures 31 and 32 illustrate the possible intermediates of X-CN^- reduction by the FeMo-cofactor. Of note in the reaction scheme are methanimine (CH_2NH), formaldehyde (H_2CO), methylamine (CH_3NH_2), ammonia (NH_3), and methane (CH_4). The reaction schemes are proposed for NaCN and KSCN substrates, and the reaction intermediates are thought to be the same for KSeCN.

One proposal of Se-incorporation comes from Ian Dance (2016), in which he proposed a bridging $SeCN^-$ exchange at Fe2 and Fe6 with loss of SCN^- [82] (Figure 42). Dance used density functional (DF) calculations that were all electron, spin-unrestricted, with no imposed symmetry, and arrived at an interested possibility for incorporation of Se into the cofactor.

Of note for this study is the dual presence of sulfur and selenium at two points in the reaction scheme. First, when $SeCN^-$ coordinates to the Fe2 and Fe6 sites, and second, where the cyanide group has migrated to the 2B bridging sulfur. Loss of SCN^- is predicted, which would allow for a reaction scheme like those proposed in Figures 31 and 32. This reaction scheme is also an appealing way to explain why some crystal occupancies measure above 100% for selenium incorporated FeMo-cofactors. It may be possible to have some residual sulfur in the 2B site, which would increase the occupancy measurement. At the selenium

absorption edge, the anomalous scattering factor for sulfur is ~ 0.24, which in the occupancy calculation would contribute about 5% to the final occupancy.

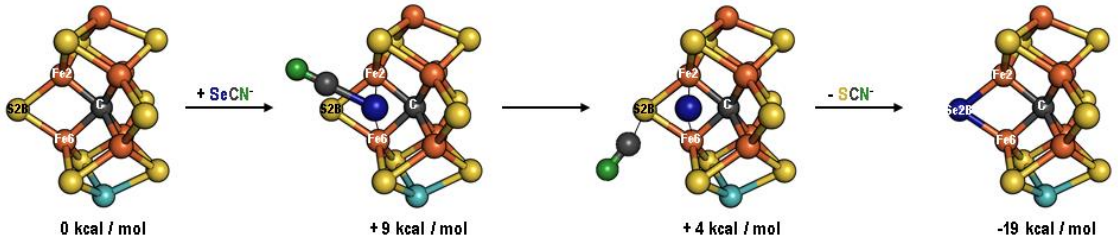


Figure 42. Calculated mechanism for the formation of the Se2B bridge from $SeCN^-$. A proposed mechanism for how Se-incorporation occurs at the S2B bridge site of the FeMo-co. Selenium is shown in dark blue, carbon in dark grey, and nitrogen in forest green. The energy barriers over the transition states are in kcal / mol. Adapted from *Dance, 2016* [82].

Following the mechanism for Dance, *2016*, it is possible that a similar mechanism may allow for reincorporation of sulfur into the 2B belt sulfur site (Figure 43). This model explains how Se gets incorporated, and provides a possibility for sulfur reincorporation. This model may also explain why extra sulfur species are not observed in nearby binding pockets in our study, since the process may be repeated between $SeCN^-$ and SCN^- *ad infinitum*. If that were the case, an appreciable amount of sulfide or selenide species may not be observable at any one time.

The proposed mechanism in Figure 43 may provide a means for selenium incorporation and sulfur reincorporation, but it does not account for methane production. Our methane production studies give some appealing insight into the mechanism of Se-incorporation and substrate reduction. Our data show that Se-incorporation into the S2B bridging site to methane production occurs on roughly a 1:1 scale (Figure 37). Once selenium incorporates to the active site, methane production must happen very early, before SCN^- can leave the cofactor for other chemistry. We propose Figure 44 as a possibility for Se incorporation and methane production.

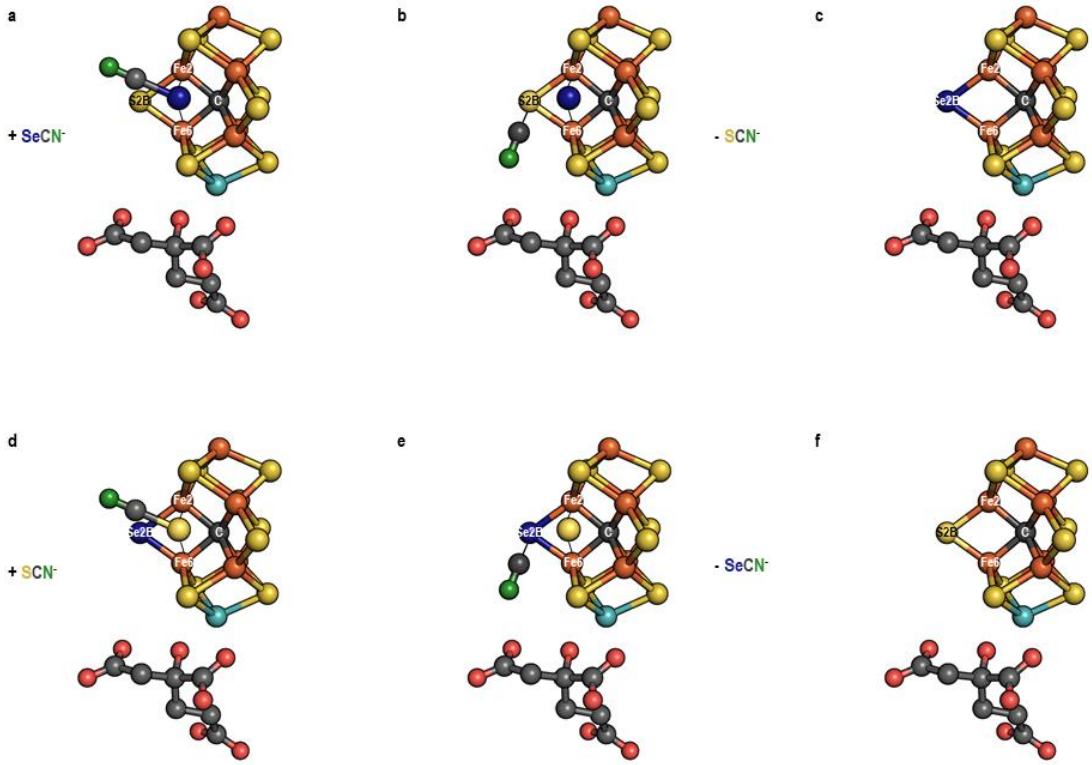


Figure 43. Possible selenium incorporation and sulfur reincorporation into the FeMo-cofactor. a) $SeCN^-$ is added to the reaction mixture. The Se atom coordinates between Fe2 and Fe6. b) The cyanide species migrates to S2B while Se remains bridged between Fe2 and Fe6. c) SCN^- is released from the cofactor, leaving a bridging Se between Fe2 and Fe6. d) SCN^- undergoes a similar mechanism as $SeCN^-$, with the sulfur atom coordinating between Fe2 and Fe6. e) The cyanide species migrates to Se2B while S remains bridged between Fe2 and Fe6. f) $SeCN^-$ is released from the cofactor, leaving a bridging S between Fe2 and Fe6. Se is shown in dark blue, S in yellow-orange, nitrogen in forest green, and carbon in dark grey. The homocitrate is included in the model for orientation in dark grey and red.

Figure 44 proposes that $SeCN^-$ binds to the cofactor preferentially at the 2B site, with selenium coordinating in a bridge between Fe2 and Fe6. Protonation of the 2B sulfur site can then occur (Figure 44bc) via His-195 (from Tyr-281) or by the homocitrate (HCA), with the ultimate release of CN^- (Figure 44c). Since Se-incorporation can only occur under turnover

conditions, we assume that the initial state of the FeMo-co is in a more reduced state (like E_3 or E_4 as proposed in the Lowe-Thornely model). The conditions of the experiment at pH 7.5 predict that the species HCN will be present at about 40 times the concentration of CN^-, according to: $pKa^{30^\circ C} = 9.11 = pH + \log(\frac{[HCN]}{[CN^-]})$. Therefore it is possible that once the HCN species has been produced, it can follow the mechanism to produce methane summarized in Figure 31 [98].

We have shown that Se preferentially incorporates at the 2B site, and from our studies with Se-HERFD, we deem the 2B site as the most favored for protonation and substrate binding. The second most likely site for protonation is the 5A site, where these series of reactions are also likely to occur. The 3A site is the least likely site for protonation and Se-incorporation, which could explain why it shows the least Se-incorporation in our studies.

Reincorporation of sulfur into the FeMo-co is still unclear in our studies, but we propose it is likely to occur via some combination of Figure 43 and Figure 44, with SCN^- being one of the possible sources of sulfur. Other sources of sulfur include dithionite and decomposed clusters. However, SCN^- is likely to be depleted upon formation of methane, so sulfur reincorporation would probably have to occur via a different sulfur anion. Previous groups have proposed potential binding pockets for sulfur anions for eventual reincorporation into the cofactor [18,22,110]. In our study we did not observe any sulfur or selenium densities elsewhere in the protein, but we reason that Se-incorporation and methane production can occur at any of the belt sulfur sites (with preference for the 2B site), which would deplete any sulfur or selenium anions in potential binding pockets.

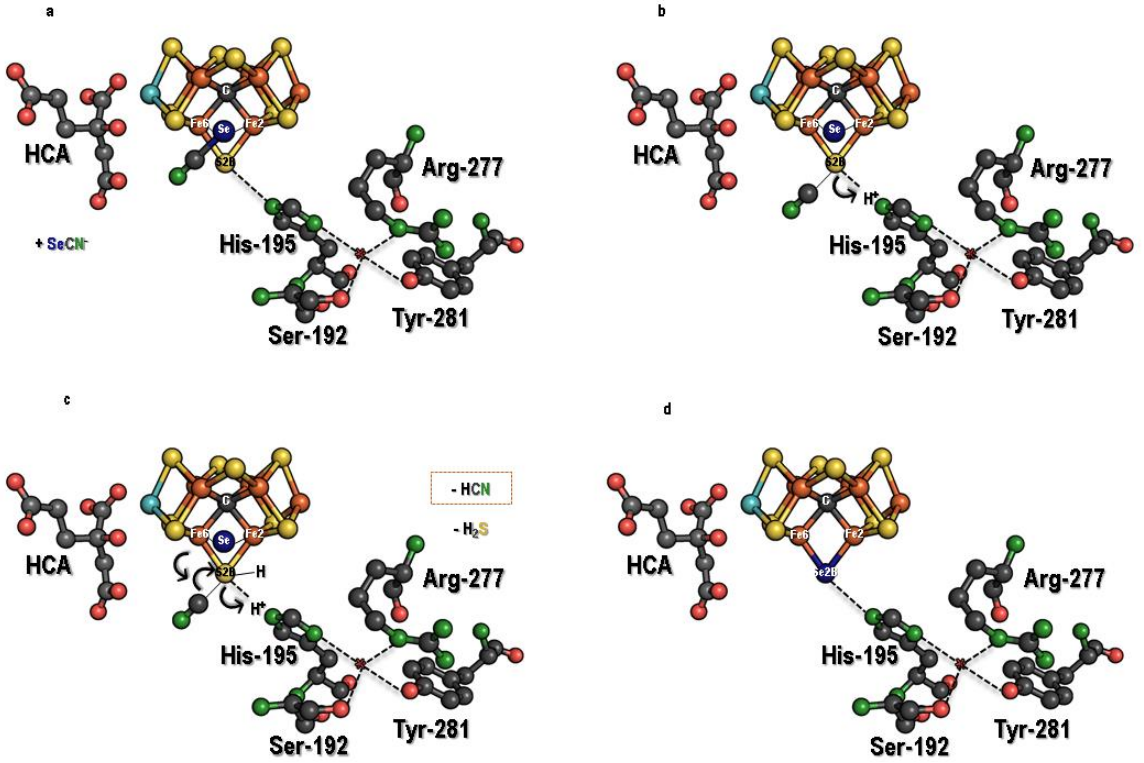


Figure 44. Proposed mechanism for Se-incorporation and HCN production from the FeMo-cofactor. a) $SeCN^-$ is added to the reaction mixture. The Se atom coordinates between Fe2 and Fe6. b) The cyanide species migrates to S2B while Se remains bridged between Fe2 and Fe6. S2B becomes protonated via His-195 (via Tyr-281). c) S2B becomes protonated a second time via His-195 (via Tyr-281). The bond between S2B and the cyanide group is broken. HCN and H_2S are released (HCN is formed from CN^- and excess protons). d) Se2B remains bridged between Fe2 and Fe6. Se is shown in dark blue, S in yellow-orange, nitrogen in forest green, and carbon in dark grey. The homocitrate is included in the model for orientation in dark grey and red. HCN is highlighted with orange dashed lines. His-195, Ser-192, Try-281, and Arg-277 are also shown. Dashed black lines shown hydrogen bonding contacts between His-195, H_2O, the S2B site, Ser-192, Tyr-281, and Arg-277.

We know that bridging sulfur is also reintegrated into the FeMo-co after going through several rounds of turnover and from being reactivated from CO [18,22]. According to an interesting report from Sippel *et al, 2017*, in the vanadium nitrogenase the sulfur may

reside in a small binding pocket 7.0 Å away from the belt sulfur site on the VFe-protein (Figure 45) [76]. Sippel and colleagues observed a light atom binding at the 2B site under turnover-like conditions.

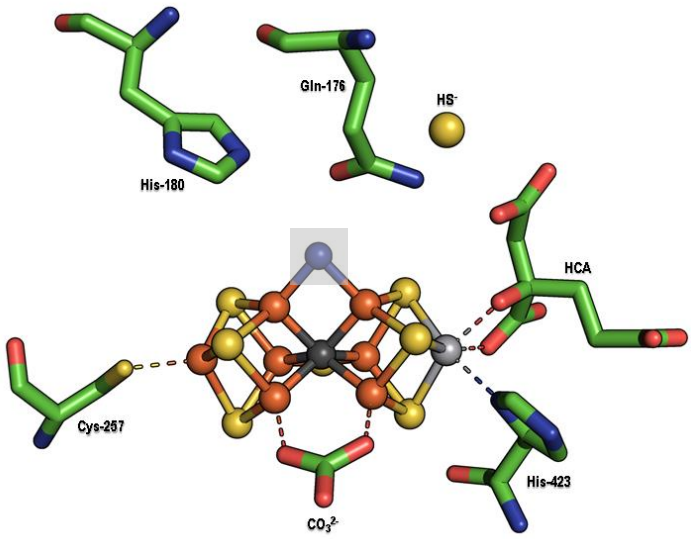


Figure 45. Potential sulfur binding site in vanadium nitrogenase. Adapted from Sippel *et al, 2017* [76]. PDB: 6FEA. The blue bridging atom is highlighted with a grey box, and was reasoned by the authors to be a light atom such as oxygen or nitrogen.

Like the MoFe nitrogenase, the light atom binds between Fe2 and Fe6 of the FeV-co. Sippel and colleagues classified this enzyme system as under turnover, and observed that Gln-176 rotated toward the FeV-co, thereby creating a binding pocket for a hydrosulfide anion (HS^-). The presence of the bound carbonate ligand at the 3A site of vanadium nitrogenase may have allowed for accumulation of HS^- at the observed binding site, since turnover events are not occurring at the carbonate-bound site. No such extra density corresponding to HS^- was observed in our structures, and neither was the movement of glutamine (Gln-191) in our *Azotobacter vinelandii* MoFe protein.

CHAPTER 5: FAST CRYSTALLIZATION OF LYSOZYME

Introduction

Crystal Growth Kinetics

As a future prospect of the study of enzymes under turnover conditions, we also investigated the possibility of trapping enzymes via fast crystallography. In principle, since nitrogenase is such a slow-turnover enzyme, crystals formed in a diffusion-limited environment could presumably be formed within one catalytic cycle.

New innovations in structural biology are expanding the capabilities for structure determination and for understanding the function of various proteins through diffraction studies on sub-micron sized crystals utilizing X-ray Free Electron Lasers (XFEL) and micro-electron diffraction (MicroED) [41,111–114]. With these new capabilities, so has evolved the opportunity for the crystallographic characterization of enzymatic reaction intermediates. With traditional diffraction sources, the study of reaction intermediates through time-resolved crystallography has been limited to those exceptional cases where either suitable photoactivatable substrates are available or where the enzyme kinetics are sufficiently slow that the reaction may be triggered by diffusion of substrates into the crystal [115–120]; the lattice must be able to accommodate conformational changes accompanying turnover. The timescale for the diffusion of substrates into crystals depends on the accessibility of a substrate to the active site in the crystal, and diffusion times for various enzymes and substrates have been reported that range from 20 seconds to several hours [12,118,122–126]. As the typical kcat for enzyme catalysis with physiological substrates is ~10 s^{-1}, it is therefore not possible to characterize intermediate states by substrate diffusion into crystals, which has motivated a number of approaches for preparing stable complexes using combinations of mutations, inhibitors and solution conditions[127]. Furthermore, for some reactions, such as nitrogenase, it is not possible to simply diffuse in substrates to trigger the reaction, as even though the physiological substrate is the diatomic N_2, the enzyme mechanism involves multiple proteins and large conformational changes that are incompatible with the crystal

lattice. As an alternative approach to ligand diffusion, we consider the possibility of trapping intermediates by rapid crystallization from an enzyme solution undergoing turnover.

Crystallization is a complex process and extensive experimental and theoretical models have been developed, taking into account key parameters in crystallization conditions: protein concentration, pH, temperature, salt, etc[128–131]. The kinetics of crystal growth reflects the rates of the underlying processes including nucleation, diffusion of sample from the surrounding solution to the crystal surface, and the surface kinetics by which molecules add to the growing surface[128,130,132–135]. The "speed limit" for crystal growth is ultimately set by the diffusion rate, which will occur when all other processes (nucleation and surface kinetics) are faster. To estimate this limit, we propose use of the "diffusion-to-capture" model, and first estimate the timescales for diffusion-controlled crystal growth.

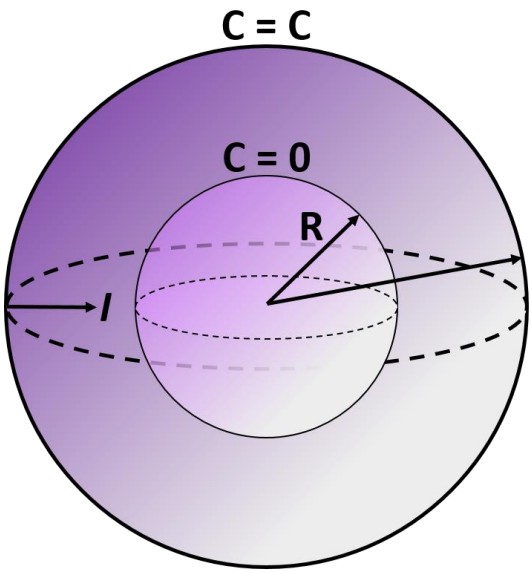


Figure 46: Diffusion-to-capture [136]. A spherical shell source with concentration C surrounds a spherical adsorber with $C = 0$ and radius R. *I* describes the diffusion current from source to adsorber.

The time, t, required for a spherical crystal to reach a radius r under diffusion-limited conditions can be estimated using the "diffusion-to-capture" model based on the steady state solution to Fick's second law [136] (Figure 46). The diffusion current describes the rate at which

mass, M, is transferred to the crystal from solution at a bulk concentration of C_0, and is given by:

$$\frac{dM}{dt} = 4\pi DC_0 \ ,$$

where D is the diffusion coefficient and r is radius; in SI units, M, D, C_0 and r have units of kg, $m^2\ s^{-1}$, kg m^{-3} (= mg ml^{-1}) and m, respectively. The mass of a spherical crystal will depend on r and the crystal density ρ (kg m^{-3}):

$$M = \frac{4}{3}\pi r^2 \rho$$

and the time derivatives of the crystal mass and radius may be related through:

$$\frac{dM}{dt} = 4\pi r^2 \rho \frac{dr}{dt} \ .$$

The increase in radius with time may be derived:

$$\frac{dr}{dt} = \frac{1}{4\pi r^2}\frac{dM}{dt} = \frac{4\pi r DC_0}{4\pi r^2 \rho} = \frac{DC_0}{r\rho} \ .$$

Consequently, the growth rate of a spherical crystal under diffusion-limited conditions is not constant, but varies inversely with the radius of the crystal[132]. The time required for a spherical crystal to grow to from radius 0 to R may be derived by integrating the previous equation:

$$t(R) = \frac{\rho}{2DC_0} R^2 .$$

This analysis predicts that the cross sectional area of a spherical crystal under diffusion controlled conditions will increase linearly with time. For a "typical" protein in water, D ~ 10^{-10} m^2 s^{-1}, ρ ~ 1300 kg m^{-3} and C ~ 10 kg m^{-3} (= 10 mg ml^{-1}), so that:

$$\begin{aligned} t(R) &\equiv \alpha A \\ &= 6x10^{-11} R^2 = 2x10^{-11} A(in\ \mathrm{m}^2) \\ &= 0.6R^2 = 0.2A(in\ \mu^2) \ . \end{aligned}$$

This analysis indicates that under conditions when diffusion is rate limiting, crystals of radius 1, 10 and 100 μ could grow in ~ 1s, 1 min and 2 hrs, respectively. A proof-of-concept study with a protein amenable to crystallography could therefore be devised to test the diffusion-to-capture theory and ultimately apply to more complex enzymes systems like nitrogenase.

Methods

Growing Fast Crystals of Lyozyme

Microcrystals of lysozyme were grown in 50 mM sodium acetate (pH 4.5) at ~90 mg/mL. A crystallization solution of 15% PEG 6000, 3.4 M NaCl, and 1 M sodium acetate (pH 4.5) was used to generate crystals on silicon hanging drop caps. The droplets were pipetted into Eppendorf tubes and stored until imaged. Microcrystals were then pipetted onto scanning electron microscope (SEM) grids, and gold-sputtered prior to imaging. The crystals were imaged using a FEI Quanta 200F Scanning Electron Microscope (Kavli Nanoscience Institute, Caltech). Crystal sizes were quantified in ImageJ by manually measuring all of the visible crystals in a given image. Averages were plotted in Excel.

To fluorescently label the lysozyme protein, lysozyme from chicken egg white (Sigma lyophilized powder, protein ≥ 90 %, ≥40,000 units/mg protein) was prepared in 50 mM sodium acetate (pH 4.5) at 12.5 mg/mL concentration. The protein was labeled with

carboxyrhodamine red using the Trace Fluorescence Labeling kit from Molecular Dimensions (MD1-73) [137]. The labeling procedure was repeated approximately 5 times in order to concentrate the final labeled product to a final concentration of ~ 60 mg/mL, which was done using Amicon centrifugal filters with MWCO 3,000. The protein was quantified at 60 mg/mL with UV-VIS and the extinction coefficient of lysozyme at 280 nm, E^{mM}=36 [16].

Lysozyme microcrystals were formed in a 16-channel Microlytic Crystal Former (Anatrace, CF-O-20) with ~60 mg/mL labeled lysozyme in 50 mM sodium acetate, pH 4.5, with less than 0.025% labeling efficiency according to the kit [137]. The crystallization solution was 15% PEG 6000, 3.4 M NaCl, and 1 M sodium acetate (pH 4.5). Crystals were formed by pipetting 0.5 uL of labeled lysozyme into one end of the crystal former channel, and pipetting 0.5 uL of crystallization solution into the other end. By capillary action the two solutions mix to form crystals.

Crystal growth rates were observed using a Leica light microscope and Hamamatsu camera (C8484-05G02) in the Bjorkman lab at Caltech, set to excite carboxyrhodamine red at ~530 nm. Timelapse videos were collecting using the software Slidebook to capture growing crystals as soon as the crystallization solution was pipetted into the Microlytic Crystal Former. Videos was collected at 10X magnification.

298 crystal growth videos were collected. Of the 298 videos, 168 were readily analyzable by ImageJ software (no background interference, no frame gaps, no interfering crystals, etc). The image scale was calibrated using the dimensions of the crystal former (100 μm X 150 μm X 10 mm) in ImageJ. The average frame time was recorded in the Slidebook software, and applied to every frame in a selected timelapse. Each crystal in the 10X video was cropped into individual timelapses and analyzed separately. First, the color threshold was adjusted to highlight the crystal. Then the image was converted to a black and white image using the Binary function in ImageJ. From the binary image, each particle was analyzed for cross-sectional area over time using the Analyze Particles function in ImageJ. Each crystal growth rate was plotted in Excel using the average frame rate recorded from SlideBook, and the cross-sectional area obtained from ImageJ. Growth rates were then reported collectively from 168 crystals.

Results

Fast Crystallization of Lysozyme

To experimentally validate the theoretical growth rates of protein crystals based on the diffusion-to-capture model, I devised a method to monitor growth rates of protein crystals in real time. We chose lysozyme as a model protein for this study, as it is readily crystallized and is amenable to high protein concentrations.

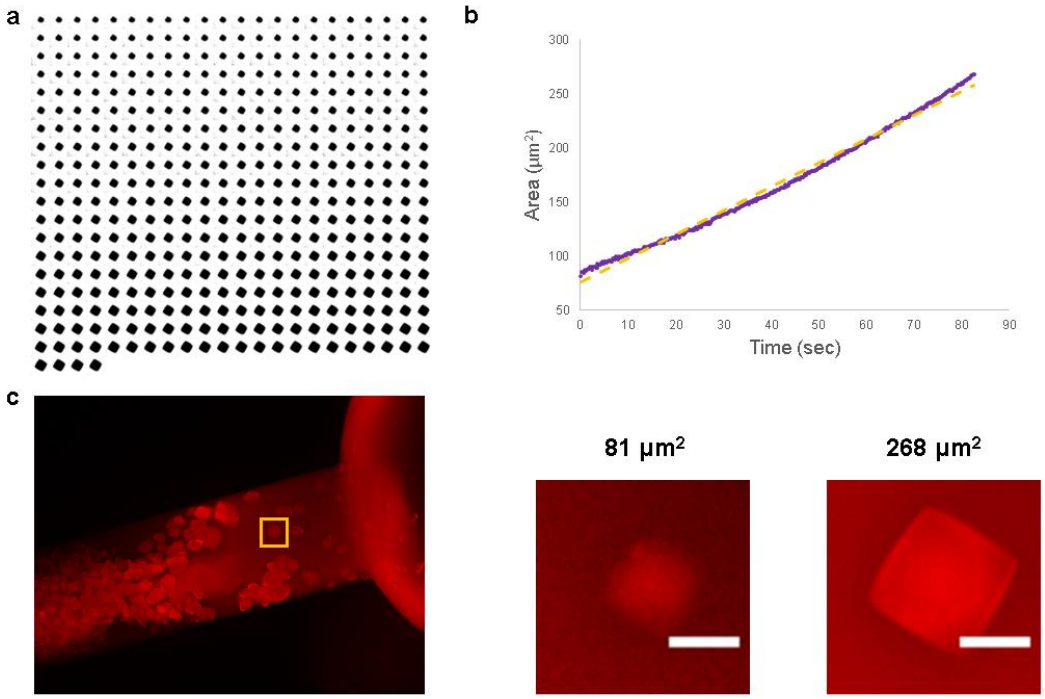


Figure 47. Single crystal growth of carboxyrhodamine-red labeled lysozyme. a) Still image, binary montage of a single labeled-lysozyme crystal growing over 82 seconds. b) Plot showing the change in area (μm^2) over time (sec). The orange trend line shows an approximate fit of $2.2\frac{\mu m^2}{\text{sec}}$ with $R^2 = 0.99$. c) Left: 10X magnification still image from time-lapse of labeled lysozyme crystals. Middle: Still image cropped to desired crystal at beginning of time-lapse video. Right: Still image of cropped crystal at end of time-lapse

video. Crystals were grown with ~60 mg/mL lysozyme in 50 mM sodium acetate, pH 4.5, with less than 0.025% labeling efficiency according to the kit [137]. The crystallization solution was 15% PEG 6000, 3.4 M NaCl, and 1 M sodium acetate, pH 4.5. Scale bar is $10\ \mu m$.

Several attempts were made to visualize crystal growth under regular light microscope conditions, but the contrast was not acceptable to perform rigorous quantification of growth rates on many crystals. To visualize the growing protein crystals with high accuracy, carboxyrhodamine-red labeled lysozyme was used to experimentally measure the kinetics of crystal growth (Figure 47). We visualized the growing crystals in the capillaries of Microlytic crystal formers, which alleviated any focusing issues we had with the microscope in larger crystal wells.

As evidenced from the data in Figure 47, the single crystal portrayed grows by $1\mu m^2$ in $0.3\ sec$. We continued to investigate the growth potentials of a great number of labeled crystals, and collected another 298 crystal videos. Of these 298 videos, 168 were analyzed in ImageJ for crystal growth rates (Figure 48).

We found that on average, we could grow lysozyme protein crystals at a growth rate of $2.1\mu m^2/\text{sec}$ with a maximum observed growth rate at $4.6\mu m^2/\text{sec}$ (Figure 48ab). Overall, we can grow crystals with an increased cross-sectional area of $1\mu m^2$ in $\sim 0.5\ sec$, with the potential to achieve this $\sim 0.2\ sec$.

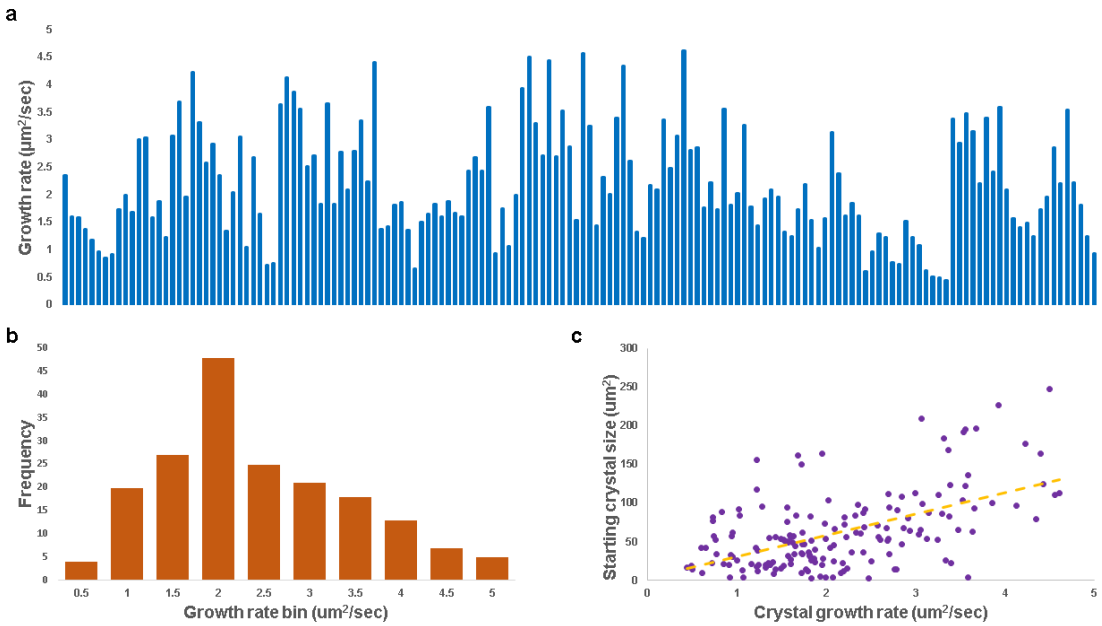


Figure 48. Growth rates for 168 carboxyrhodamine-red labeled lysozyme protein crystals. a) All growth rates for 168 crystals, with growth rate shown in $\mu m^2 / \sec$. The crystals were analyzed for cross-sectional area over time in ImageJ. The average growth rate was $2.1 \mu m^2 / \sec$ with a standard deviation of ± 0.98. b) Growth rates separated into bins of $0.5 \mu m^2 / \sec$). Growth rates correlated to starting crystal size in μm^2. The linear trendline was fit to: $27.7x + 3.02$. Pearson coefficient: 0.552 .

To further verify that the crystals we observed in the growth videos were indeed lysozyme protein, we retrieved 5 crystals directly from the crystal former wells. We sent these crystals to the Stanford Synchrotron Radiation Lightsource (SSRL), beamline 12-2 to obtain their crystal structure. We successfully solved the structure of lysozyme to 2.0 Å resolution and observed no significant changes from previously published structures (Figure 49). The carboxyrhodamine-red label was not resolved in the structure, which was unsurprising considering that the labeling efficiency was estimated at 0.025%[137].

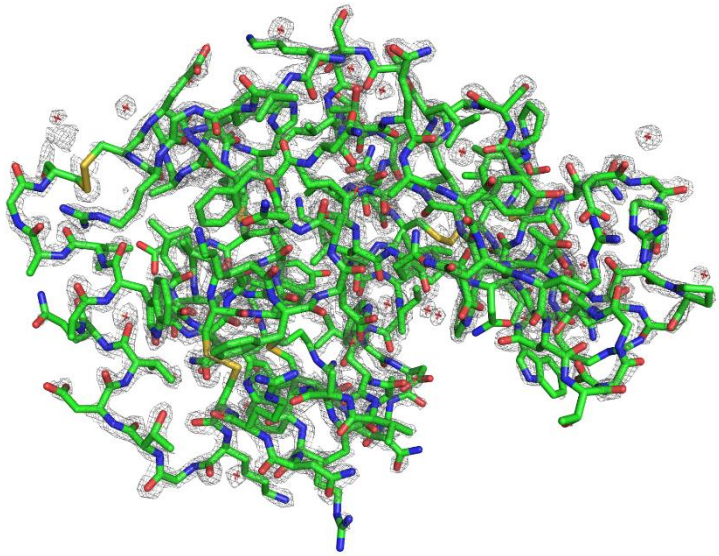


Figure 49. Lysozyme structure solved to 2.0 Å directly from fast-growing solution. Space group: P 4_3 2_1 2. R-factor: 0.212, R-free: 0.236. Unit cell: a=b=79.118 Å, c=36.760 Å.

Discussion

Future Studies using Crystallography as an Enzyme-Turnover Tool

We have shown that crystals can increase their cross-sectional area at a rate of 1 μm^2/sec, which under diffusion-limited conditions could possibly yield microcrystals within the turnover times of many enzymes. A histogram shows that a large number of crystals observed in this study show rates of cross-sectional area increase in the $1.5\mu m^2/\sec$ to $2.5\mu m^2/\sec$ range (Figure 48b). Growth rates for different crystal faces of lysozyme have been typically measured as ~ $0.01-0.08\mu m/\sec$, and are typically only measured from a single face of the growing crystal[128,138,139]. Our theoretical analysis indicates that under conditions when diffusion is rate limiting, crystals of radius 1, 10 and 100 μ could grow in ~ 1s, 1 min and 2 hrs, respectively. The experimentally observed growth times may be slowed for a number of reasons including drag attributed to the lysozyme molecule itself and viscosity of the buffer solutions. We attempted to vary the concentration of labeled lysozyme in solution to further probe our theoretical rates, but were unable to produce viable crystals to measure.

We also investigated the possibility of a correlation between the starting crystal size and the growth rate observed. One could predict that if the crystal began small, the growth rate would also be diminished due to less surface area available for further growth. According to the diffusion-to-capture model, we would not predict any correlation between crystal size and growth rate. Despite this, we observe a slight positive correlation between crystal size and growth rate (Figure 48c). More experiments over a very broad range of crystal sizes would have to be performed fully understand this finding.

With these results, we propose that microfluidic systems may be utilized to trap enzymes via crystallization under turnover conditions. For example, a system similar to the one devised by Zheng *et al., 2003* or Gerdts *et al., 2006,* could be a feasible way to initiate crystallization. In these works, serpentine channels and nanoliter sized droplets are used to control both nucleation and crystallization [23,24]. Injection ports on a microfluidic device can be customized for protein solution injection and precipitant injection, and outlet ports can be optimized for droplet collection at the desired size and/or timepoint. We propose that a crystal seeding port may also be devised to initiate the nucleation process and further accelerate the crystal growth process.

The crystals may be analyzed in several ways, but we believe the most feasible methods will be to utilize MicroED or XFEL. For use with MicroED, the crystals may be collected from the microfluidic device, frozen on electron microscopy grids, and evaluated in most 200 keV electron microscopes. Diffraction data can be collected and assessed using standard X-ray crystallography software[140]. For use with the XFEL, an in-house sample injector may be used to collect diffraction data from crystals generated under turnover conditions in real time [141].

Fast crystallization could be another technique in the structural biology repertoire that would give insight into structural changes within enzymes during turnover. XFEL and MicroED are both techniques that necessitate small crystals, and can each give resolutions that contend with synchrotron sources [112,113,142–146]. While the XFEL will also be able to resolve higher resolution questions pertaining to enzyme turnover, future improvements to the technique of MicroED will also eventually achieve higher resolutions. MicroED has shown great potential in the field of structural biology, currently resolving several structures,

including lysozyme at 1.8 Å, bovine liver catalase at 3.2 Å, proteinase K at 1.6 Å, and trypsin at 1.7 Å [112,113,147].

We propose that fast crystallization be applied to slow enzymes in general, or be used on enzymes of interest in which turnovers can be manipulated. In our case we propose these methods be used on nitrogenase in the future, as nitrogenase is a relatively slow enzyme, crystallizes well, and has turnover events that are very appealing for this method.

S U M M A R Y

The results shown here make various contributions to the kinetics of crystallography for the study of enzymes, the usage of new HERFD spectroscopies on the Se-incorporated FeMo-co, and the insertion of Se into the cofactor in terms of methane production. Further investigation into the Se-system is required, but the data here provide valuable insights into the electronic and chemical mechanism of nitrogenase.

REFERENCES

1. Bernhardt, E. S. & Schlesinger, W. H. *Biogeochemistry: an analysis of global change*. (Waltham, MA, USA and Oxford, UK: Elsevier Inc, 2013).
2. Tezcan, F. A. *et al.* Nitrogenase complexes: multiple docking sites for a nucleotide switch protein. *Science* **309,** 1377–1380 (2005).
3. Fisher, K., Dilworth, M. J. & Newton, W. E. Differential effects on N2 binding and reduction, HD formation, and azide reduction with α-195His-and α-191Gln-substituted MoFe proteins of Azotobacter vinelandii nitrogenase. *Biochemistry (Mosc.)* **39,** 15570–15577 (2000).
4. Fisher, K., Dilworth, M. J. & Newton, W. E. Azotobacter vinelandii vanadium nitrogenase: formaldehyde is a product of catalyzed HCN reduction, and excess ammonia arises directly from catalyzed azide reduction. *Biochemistry (Mosc.)* **45,** 4190–4198 (2006).
5. Rasche, M. E. & Seefeldt, L. C. Reduction of thiocyanate, cyanate, and carbon disulfide by nitrogenase: kinetic characterization and EPR spectroscopic analysis. *Biochemistry (Mosc.)* **36,** 8574–8585 (1997).
6. Barney, B. M. *et al.* Trapping a hydrazine reduction intermediate on the nitrogenase active site. *Biochemistry (Mosc.)* **44,** 8030–8037 (2005).
7. Barney, B. M. *et al.* Diazene (HN NH) Is a Substrate for Nitrogenase: Insights into the Pathway of N2 Reduction. *Biochemistry (Mosc.)* **46,** 6784–6794 (2007).
8. Dos Santos, P. C., Mayer, S. M., Barney, B. M., Seefeldt, L. C. & Dean, D. R. Alkyne substrate interaction within the nitrogenase MoFe protein. *J. Inorg. Biochem.* **101,** 1642–1648 (2007).

9. Seefeldt, L. C., Rasche, M. E. & Ensign, S. A. Carbonyl sulfide and carbon dioxide as new substrates, and carbon disulfide as a new inhibitor, of nitrogenase. *Biochemistry (Mosc.)* **34,** 5382–5389 (1995).

10. Rivera-Ortiz, J. M. & Burris, R. H. Interactions among substrates and inhibitors of nitrogenase. *J. Bacteriol.* **123,** 537–545 (1975).

11. Jensen, B. B. & Burris, R. H. Nitrous oxide as a substrate and as a competitive inhibitor of nitrogenase. *Biochemistry (Mosc.)* **25,** 1083–1088 (1986).

12. Hwang, J. C., Chen, C. H. & Burris, R. H. Inhibition of nitrogenase-catalyzed reductions. *Biochim. Biophys. Acta BBA-Bioenerg.* **292,** 256–270 (1973).

13. Cameron, L. M. & Hales, B. J. Investigation of CO binding and release from Mo-nitrogenase during catalytic turnover. *Biochemistry (Mosc.)* **37,** 9449–9456 (1998).

14. Smil, V. *Enriching the earth: Fritz Haber, Carl Bosch, and the transformation of world food production*. (MIT press, 2004).

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

16. Davies, R. C., Neuberger, A. & Wilson, B. M. The dependence of lysozyme activity on pH and ionic strength. *Biochim. Biophys. Acta BBA-Enzymol.* **178,** 294–305 (1969).

17. Voet, D. & Voet, J. *Biochemistry*. (Wiley, 2004).

18. 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).

19. Spatzal, T. *et al.* Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science* **334,** 940–940 (2011).

20. George, S. J., Ashby, G. A., Wharton, C. W. & Thorneley, R. N. Time-resolved binding of carbon monoxide to nitrogenase monitored by stopped-flow infrared spectroscopy. *J. Am. Chem. Soc.* **119,** 6450–6451 (1997).

21. Han, J. & Newton, W. E. Differentiation of acetylene-reduction sites by stereoselective proton addition during Azotobacter vinelandii nitrogenase-catalyzed C2D2 reduction. *Biochemistry (Mosc.)* **43,** 2947–2956 (2004).

22. Spatzal, T., Perez, K. A., Howard, J. B. & Rees, D. C. Catalysis-dependent selenium incorporation and migration in the nitrogenase active site iron-molybdenum cofactor. *Elife* **4,** (2015).

23. Galloway, J. N. *et al.* Nitrogen cycles: past, present, and future. *Biogeochemistry* **70,** 153–226 (2004).

24. Socolow, R. H. Nitrogen management and the future of food: lessons from the management of energy and carbon. *Proc. Natl. Acad. Sci.* **96,** 6001–6008 (1999).

25. Galloway, J. N. & Cowling, E. B. Reactive nitrogen and the world: 200 years of change. *AMBIO J. Hum. Environ.* **31,** 64–71 (2002).

26. Canfield, D. E., Glazer, A. N. & Falkowski, P. G. The evolution and future of Earth's nitrogen cycle. *science* **330,** 192–196 (2010).

27. Gruber, N. & Galloway, J. N. An Earth-system perspective of the global nitrogen cycle. *Nature* **451,** 293 (2008).

28. Liu, H. *Ammonia synthesis catalysts: innovation and practice.* (World Scientific, 2013).

29. Bernhard, A. The nitrogen cycle: processes, players, and human impact [WWW Document]. *Nat Educ Knowl URL Httpwww Nat. Comscitableknowledgelibrary-Nitrogen-Cycle-Process.--Hum.-15644632 Accessed 73 14* (2010).

30. Vitousek, P. M. *et al.* Towards an ecological understanding of biological nitrogen fixation. in *The Nitrogen Cycle at Regional to Global Scales* 1–45 (Springer, 2002).

31. Smith, V. H. & Bennett, S. J. Nitrogen: phosphorus supply ratios and phytoplankton community structure in lakes. *Arch. Für Hydrobiol.* 37–53 (1999).

32. Vitousek, P. M. & Field, C. B. Ecosystem constraints to symbiotic nitrogen fixers: a simple model and its implications. in *New Perspectives on Nitrogen Cycling in the Temperate and Tropical Americas* 179–202 (Springer, 1999).

33. Bouwman, A. F. Soils and the greenhouse effect. (1990).

34. Andreae, M. O. & Schimel, D. S. Exchange of trace gases between terrestrial ecosystems and the atmosphere. (1990).

35. Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440,** 790 (2006).

36. Peters, J. W., Fisher, K., Newton, W. E. & Dean, D. R. Involvement of the P cluster in intramolecular electron transfer within the nitrogenase MoFe protein. *J. Biol. Chem.* **270,** 27007–27013 (1995).

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

38. Burgess, B. K. & Lowe, D. J. Mechanism of molybdenum nitrogenase. *Chem. Rev.* **96,** 2983–3012 (1996).

39. Thorneley, R. N. & Lowe, D. J. Kinetics and mechanism of the nitrogenase enzyme system. *Molybdenum Enzym.* **7,** 89–116 (1985).

40. Kirn, J. & Rees, D. C. Crystallographic structure and functional implications of the nitrogenase molybdenum–iron protein from Azotobacter vinelandii. *Nature* **360,** 553 (1992).

41. Einsle, O. *et al.* Nitrogenase MoFe-protein at 1.16 \AA resolution: a central ligand in the FeMo-cofactor. *Science* **297,** 1696–1700 (2002).

42. Lancaster, K. M. *et al.* X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* **334,** 974–977 (2011).

43. Hu, Y. & Ribbe, M. W. Biosynthesis of nitrogenase FeMoco. *Coord. Chem. Rev.* **255,** 1218–1224 (2011).

44. Chatt, J. & Leigh, G. J. Nitrogen fixation. *Chem. Soc. Rev.* **1,** 121–144 (1972).

45. Chatt, J., Dilworth, J. R. & Richards, R. L. Recent advances in the chemistry of nitrogen fixation. *Chem. Rev.* **78,** 589–625 (1978).

46. Barney, B. M. *et al.* Breaking the N 2 triple bond: insights into the nitrogenase mechanism. *Dalton Trans.* 2277–2284 (2006).

47. Burgess, B. K. *Molybdenum Enzymes.* (Wiley, 1985).

48. Howard, J. B. & Rees, D. C. How many metals does it take to fix N2? A mechanistic overview of biological nitrogen fixation. *Proc. Natl. Acad. Sci.* **103,** 17088–17093 (2006).

49. McKenna, C. E., Simeonov, A. M., Eran, H. & Bravo-Leerabhandh, M. Reduction of cyclic and acyclic diazene derivatives by Azotobacter vinelandii nitrogenase: diazirine and trans-dimethyldiazene. *Biochemistry (Mosc.)* **35,** 4502–4514 (1996).

50. 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 (Mosc.)* **53,** 333–343 (2014).

51. Wolle, D., Kim, C., Dean, D. & Howard, J. B. Ionic interactions in the nitrogenase complex. Properties of Fe-protein containing substitutions for Arg-100. *J. Biol. Chem.* **267,** 3667–3673 (1992).

52. Corbin, J. L. Liquid chromatographic-fluorescence determination of ammonia from nitrogenase reactions: A 2-min assay. *Appl. Environ. Microbiol.* **47,** 1027–1030 (1984).

53. Lind, C. J. & Wilson, P. W. Mechanism of Biological Nitrogen Fixation. VIII. Carbon Monoxide as an Inhibitor for Nitrogen Fixation by Red Clover1. *J. Am. Chem. Soc.* **63,** 3511–3514 (1941).

54. Lee, H.-I., Hales, B. J. & Hoffman, B. M. Metal-ion valencies of the FeMo cofactor in CO-inhibited and resting state nitrogenase by 57Fe Q-band ENDOR. *J. Am. Chem. Soc.* **119,** 11395–11400 (1997).

55. Cramer, S. P. *et al.* The molybdenum site of nitrogenase. 2. A comparative study of molybdenum-iron proteins and the iron-molybdenum cofactor by x-ray absorption spectroscopy. *J. Am. Chem. Soc.* **100,** 3814–3819 (1978).

56. Yoo, S. J., Angove, H. C., Papaefthymiou, V., Burgess, B. K. & Münck, E. Mössbauer study of the MoFe protein of nitrogenase from Azotobacter vinelandii using selective 57Fe enrichment of the M-centers. *J. Am. Chem. Soc.* **122,** 4926–4936 (2000).

57. Harris, T. V. & Szilagyi, R. K. Comparative assessment of the composition and charge state of nitrogenase FeMo-cofactor. *Inorg. Chem.* **50,** 4811–4824 (2011).

58. Bjornsson, R. *et al.* Identification of a spin-coupled Mo (III) in the nitrogenase iron–molybdenum cofactor. *Chem. Sci.* **5,** 3096–3103 (2014).

59. Cramer, S. P., Hodgson, K. O., Gillum, W. O. & Mortenson, L. E. The molybdenum site of nitrogenase. Preliminary structural evidence from X-ray absorption spectroscopy. *J. Am. Chem. Soc.* **100,** 3398–3407 (1978).

60. Smolentsev, G. *et al.* X-ray emission spectroscopy to study ligand valence orbitals in Mn coordination complexes. *J. Am. Chem. Soc.* **131,** 13161–13167 (2009).

61. Eeckhout, S. G. *et al.* Cr local environment by valence-to-core X-ray emission spectroscopy. *J. Anal. At. Spectrom.* **24,** 215–223 (2009).

62. Delgado-Jaime, M. U. *et al.* Identification of a single light atom within a multinuclear metal cluster using valence-to-core X-ray emission spectroscopy. *Inorg. Chem.* **50,** 10709–10717 (2011).

63. Georgiadis, M. M. *et al.* Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science* **257,** 1653–1659 (1992).

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

65. Kim, J. & Rees, D. C. Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science* **257,** 1677–1682 (1992).

66. Bolin, J. T., Campobasso, N., Muchmore, S. W., Morgan, T. V. & Mortenson, L. E. Structure and Environment of Metal Clusters in the Nitrogenase Molybdenum—Iron Protein from Clostridium pasteurianum. in (ACS Publications, 1993).

67. Peters, J. W. *et al.* Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry (Mosc.)* **36,** 1181–1187 (1997).

68. 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 \AA resolution X-ray crystallographic study of Klebsiella pneumoniae MoFe-protein1. *J. Mol. Biol.* **292,** 871–891 (1999).

69. Sørlie, M. *et al.* Mechanistic features and structure of the nitrogenase α-Gln195 MoFe protein. *Biochemistry (Mosc.)* **40,** 1540–1549 (2001).

70. Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B. & Rees, D. C. Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum1. *J. Mol. Biol.* **280,** 669–685 (1998).

71. Jang, S. B., Seefeldt, L. C. & Peters, J. W. Modulating the midpoint potential of the [4Fe-4S] cluster of the nitrogenase Fe protein. *Biochemistry (Mosc.)* **39,** 641–648 (2000).

72. Strop, P. *et al.* Crystal structure of the all-ferrous [4Fe-4S] 0 form of the nitrogenase iron protein from Azotobacter vinelandii. *Biochemistry (Mosc.)* **40,** 651–656 (2001).

73. Sen, S. *et al.* A conformational mimic of the MgATP-bound "on state" of the nitrogenase iron protein. *Biochemistry (Mosc.)* **43,** 1787–1797 (2004).

74. Chiu, H.-J. *et al.* MgATP-bound and nucleotide-free structures of a nitrogenase protein complex between the Leu 127Δ-Fe-protein and the MoFe-protein. *Biochemistry (Mosc.)* **40,** 641–650 (2001).

75. Schmid, B. *et al.* Biochemical and structural characterization of the cross-linked complex of nitrogenase: comparison to the ADP-AlF4–stabilized structure. *Biochemistry (Mosc.)* **41,** 15557–15565 (2002).

76. Sippel, D. & Einsle, O. The structure of vanadium nitrogenase reveals an unusual bridging ligand. *Nat. Chem. Biol.* **13,** 956 (2017).

77. Rittle, J. & Peters, J. C. Proton-Coupled Reduction of an Iron Cyanide Complex to Methane and Ammonia. *Angew. Chem.* **128,** 12450–12453 (2016).

78. Del Castillo, T. J., Thompson, N. B. & Peters, J. C. A synthetic single-site Fe nitrogenase: high turnover, freeze-quench 57Fe Mossbauer data, and a hydride resting state. *J. Am. Chem. Soc.* **138,** 5341–5350 (2016).

79. Bazhenova, T. A. & Shilov, A. E. Nitrogen fixation in solution. *Coord. Chem. Rev.* **144,** 69–145 (1995).

80. Ung, G. & Peters, J. C. Low-Temperature N2 Binding to Two-Coordinate L2Fe0 Enables Reductive Trapping of L2FeN2- and NH3 Generation. *Angew. Chem. Int. Ed.* **54,** 532–535 (2015).

81. Arashiba, K. *et al.* Catalytic reduction of dinitrogen to ammonia by use of molybdenum–nitride complexes bearing a tridentate triphosphine as catalysts. *J. Am. Chem. Soc.* **137,** 5666–5669 (2015).

82. Dance, I. Mechanisms of the S/CO/Se interchange reactions at FeMo-co, the active site cluster of nitrogenase. *Dalton Trans.* **45,** 14285–14300 (2016).

83. NASA. The Electromagnetic Spectrum. (2017). Available at: https://mynasadata.larc.nasa.gov/.

84. Rhodes, G. *Crystallography Made Crystal Clear.* (Elsevier, 2006).

85. Spatzal, T. & Einsle, O. Structural, electronic and magnetic characterization of the nitrogenase active site FeMo-cofactor and crystallization of the ketol-acid reductoisomerase from Azotobacter vinelandii. (Universitat Freiburg, 2012).

86. Perez, K. Structural and Biochemical Characterization of Ligand Bound States of the FeMo-cofactor of Nitrogenase. (California Institute of Technology, 2016).

87. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 48–57 (2006).

88. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).

89. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 133–144 (2010).

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

91. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).

92. Kleywegt, G. J. & Jones, T. A. xdlMAPMAN and xdlDATAMAN–programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D Biol. Crystallogr.* **52,** 826–828 (1996).

93. DeBeer, S. Advanced X-Ray Spectroscopy. *Penn State Inorganic Workshop* (2016). Available at: http://sites.psu.edu/bioinorganic/wp-content/uploads/sites/29389/2015/08/2016-04-DeBeer-advanced-x-ray-spectroscopy.pdf.

94. Hämäläinen, K., Siddons, D. P., Hastings, J. B. & Berman, L. E. Elimination of the inner-shell lifetime broadening in x-ray-absorption spectroscopy. *Phys. Rev. Lett.* **67,** 2850 (1991).

95. Eisenberger, P., Platzman, P. M. & Winick, H. X-ray resonant Raman scattering: Observation of characteristic radiation narrower than the lifetime width. *Phys. Rev. Lett.* **36,** 623 (1976).

96. Bjornsson, R., Neese, F. & DeBeer, S. Revisiting the Mössbauer Isomer Shifts of the FeMoco Cluster of Nitrogenase and the Cofactor Charge. *Inorg. Chem.* **56,** 1470–1477 (2017).

97. Morrison, C. N., Spatzal, T. & Rees, D. C. Reversible Protonated Resting State of the Nitrogenase Active Site. *J. Am. Chem. Soc.* **139,** 10856–10862 (2017).

98. Li, J., Burgess, B. K. & Corbin, J. L. Nitrogenase reactivity: cyanide as substrate and inhibitor. *Biochemistry (Mosc.)* **21,** 4393–4402 (1982).

99. Smith, R. M. & Martell, A. E. Inorganic Ligands. in *Critical Stability Constants* 426–461 (Springer, Boston, MA, 1989). doi:10.1007/978-1-4615-6764-6_24

100. Davis, F. A. & Chen, B. C. Asymmetric hydroxylation of enolates with N-sulfonyloxaziridines. *Chem. Rev.* **92,** 919–934 (1992).

101. *HALOGENOID COMPLEXES*. (Defense Technical Information Center, 1965).

102. Acides, D. *et al. Dissociation Constants of Inorganic Acids and Bases in Aqueous Solution*.

103. Cromer, D. T. & Liberman, D. A. Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. *Acta Crystallogr. A* **37,** 267–268 (1981).

104. Cromer, D. T. & Mann, J. B. X-ray scattering factors computed from numerical Hartree–Fock wave functions. *Acta Crystallogr. A* **24,** 321–324 (1968).

105. Balyuzi, H. H. M. Analytic approximation to incoherently scattered X-ray intensities. *Acta Crystallogr. A* **31,** 600–602 (1975).

106. Hoyt, J. J., De Fontaine, D. & Warburton, W. K. Determination of the anomalous scattering factors for Cu, Ni and Ti using the dispersion relation. *J. Appl. Crystallogr.* **17,** 344–351 (1984).

107. Kissel, L. & Pratt, R. H. Corrections to tabulated anomalous-scattering factors. *Acta Crystallogr. A* **46,** 170–175 (1990).

108.Salem, S. I., Chang, C. N., Lee, P. L. & Severson, V. Energy shift of the K-absorption edge of Mn and Fe compounds. *J. Phys. C Solid State Phys.* **11,** 4085–4093 (1978).

109.Agarwal, B. K. & Verma, L. P. A rule for chemical shifts of X-ray absorption edges. *J. Phys. C Solid State Phys.* **3,** 535–537 (1970).

110.Sippel, D. *et al.* A bound reaction intermediate sheds light on the mechanism of nitrogenase. *Science* **359,** 1484–1489 (2018).

111.Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *Elife* **2,** (2013).

112.Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. Structure of catalase determined by MicroED. *Elife* **3,** (2014).

113.Nannenga, B. L., Shi, D., Leslie, A. G. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11,** 927 (2014).

114.Kupitz, C. *et al.* Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513,** 261 (2014).

115.Barends, T. R. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350,** 445–450 (2015).

116.Kern, J. *et al.* Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340,** 491–495 (2013).

117.Kern, J. *et al.* Taking snapshots of photosynthetic water oxidation using femtosecond X-ray diffraction and spectroscopy. *Nat. Commun.* **5,** 4371 (2014).

118.Hajdu, J. & Johnson, L. N. Progress with Laue diffraction studies on protein and virus crystals. *Biochemistry (Mosc.)* **29,** 1669–1678 (1990).

119. Alber, T., Petsko, G. A. & Tsernoglou, D. Crystal structure of elastase–substrate complex at- 55° C. *Nature* **263,** 297 (1976).

120. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C.-A. Structures of Product and inhibitor complexes of Streptomyces griseus protease A at 1.8 \AA Resolution: A Model for Serine Protease Catalysis. *J. Mol. Biol.* **144,** 43–88 (1980).

121. Geremia, S., Campagnolo, M., Demitri, N. & Johnson, L. N. Simulation of diffusion time of small molecules in protein crystals. *Structure* **14,** 393–400 (2006).

122. O'Hara, P., Goodwin, P. & Stoddard, B. L. Direct measurement of diffusion rates in enzyme crystals by video absorbance spectroscopy. *J. Appl. Crystallogr.* **28,** 829–834 (1995).

123. Wyckoff, H. W. *et al.* Design of a diffractometer and flow cell system for X-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S. *J. Mol. Biol.* **27,** 563–578 (1967).

124. Westbrook, E. M. & Sigler, P. B. Enzymatic function in crystals of delta 5-3-ketosteroid isomerase. Catalytic activity and binding of competitive inhibitors. *J. Biol. Chem.* **259,** 9090–9095 (1984).

125. Cvetkovic, A. *et al.* Quantifying anisotropic solute transport in protein crystals using 3-D laser scanning confocal microscopy visualization. *Biotechnol. Bioeng.* **86,** 389–398 (2004).

126. Velev, O. D., Kaler, E. W. & Lenhoff, A. M. Surfactant diffusion into lysozyme crystal matrices investigated by quantitative fluorescence microscopy. *J. Phys. Chem. B* **104,** 9267–9275 (2000).

127. Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry (Mosc.)* **50,** 4402–4410 (2011).

128.Schmit, J. D. & Dill, K. Growth rates of protein crystals. *J. Am. Chem. Soc.* **134,** 3934–3937 (2012).

129.Durbin, S. D. & Feher, G. Protein Crystallization. *Annu. Rev. Phys. Chem.* **47,** 171–204 (1996).

130.Gorti, S., Konnert, J., Forsythe, E. L. & Pusey, M. L. Effects of Kinetic Roughening and Liquid−Liquid Phase Transition on Lysozyme Crystal Growth Velocities. *Cryst. Growth Des.* **5,** 535–545 (2005).

131.Schmit, J. D., Whitelam, S. & Dill, K. Electrostatics and aggregation: How charge can turn a crystal into a gel. *J. Chem. Phys.* **135,** 08B626 (2011).

132.Saito, Y. *Statistical physics of crystal growth*. (World Scientific, 1996).

133.Schmit, J. D. & Dill, K. A. The stabilities of protein crystals. *J. Phys. Chem. B* **114,** 4020–4027 (2010).

134.George, A. & Wilson, W. W. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 361–365 (1994).

135.Sear, R. P. On the interpretation of quantitative experimental data on nucleation rates using classical nucleation theory. *J. Phys. Chem. B* **110,** 21944–21949 (2006).

136.Berg, H. *Random Walks in Biology*. (Princeton University Press, 1993).

137.The Trace Fluorescence Labeling Kit. (2018).

138.Durbin, S. D. & Feher, G. Crystal growth studies of lysozyme as a model for protein crystallization. *J. Cryst. Growth* **76,** 583–592 (1986).

139.Koszelak, S., Martin, D., Ng, J. & McPherson, A. Protein crystal growth rates determined by time lapse microphotography. *J. Cryst. Growth* **110,** 177–181 (1991).

140.Nannenga, B. L. & Gonen, T. MicroED: a versatile cryoEM method for structure determination. *Emerg. Top. Life Sci.* ETLS20170082 (2018). doi:10.1042/ETLS20170082

141. Yano, J. Water Oxidation Reaction in Photosystem II Studies with XFELs. *Biophys. J.* **114,** 206a (2018).

142. Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* 1217737 (2012).

143. Schlichting, I. Serial femtosecond crystallography: the first five years. *IUCrJ* **2,** 246–255 (2015).

144. Johansson, L. C. *et al.* Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography. *Nat. Commun.* **4,** 2911 (2013).

145. Kern, J. *et al.* Room temperature femtosecond X-ray diffraction of photosystem II microcrystals. *Proc. Natl. Acad. Sci.* **109,** 9721–9726 (2012).

146. Shi, D. *et al.* The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11,** 895 (2016).

147. de la Cruz, M. J. *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14,** 399 (2017).