UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA
ALIMENTARIA Y DE BIOSISTEMAS
DEPARTAMENTO DE BIOTECNOLOGÍA-BIOLOGÍA VEGETAL

POLITÉCNICA

TESIS DOCTORAL

**Caracterización de una nueva proteína hipertermófila NifB y su
papel en el mecanismo de síntesis de NifB-co, precursor del grupo
metálico de FeMo-co**

Author: Alessandro Scandurra

Director: Prof. Luis Manuel Rubio Herrero

Co-director: Dr. Simon Jean Marius Arragain

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA
ALIMENTARIA Y DE BIOSISTEMAS
DEPARTAMENTO DE BIOTECNOLOGÍA-BIOLOGÍA VEGETAL

Memoria presentada por **D. Alessandro Scandurra**
para optar al grado de Doctor

Director
Dr. Luis Manuel Rubio Herrero
Profesor Titular UPM

Madrid, 2017

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rest with the author and that no quotation from the thesis, or any information derived therefrom may be published without the author's prior, written consent.

Acknowledgments

As each journey, also this one has an end (Thanks goodness).

It was a long trip, with several changes of direction. Like for many other people, the PhD adventure was not easy, but I have to admit that I was lucky enough to meet really unique people, that became special friendships.

First of all, I want to thank my family, especially my mother and my father, who always push me to go forward in my life and to follow my dreams, ready to pay the sacrifice of the distance. This experience has helped me to understand how easy is to be a son and how challenging is to be a parent.

But I would never reach the conclusion of my Ph.D without the support of my lovely Lola, the princess of my favorite fable, my life. Her support, encouragement, tenacious patience and unwavering love were undeniably the bedrock upon which the past six years of my life have been built. During the Ph.D time, thanks to her temperamental Andalusian character, I always admired her capacity to shake me from my torpor's moments and bring me down to the earth from my Pindaric flights after a long day of work.

With a background in Marine biology, I started a PhD in biochemistry. At first sight, this decision did not seem to be very promising, and nowadays, this choice looks for me to be pure madness, but I think this was the typical impulsive decision of a young guy without too much experience in the professional world.

I have to admit that my analytical side has always been a bit limited. Poorly trained, my rational intelligence was always compensated by my emotional intelligence. Indeed my curiosity, fomented by an insatiable fantasy, pushed me to take decisions during my life and still today encourages me to go on.

Moreover I find fun and witty to compare my PhD experience as one of "*Goldoni's novel*", divided in two principal acts, where unlikely characters develop a plot that at first time seems overwhelmed, but at the end is concluded with a brilliant ending.

In the *first act*, I started as "*neophyte*" my work experience in the laboratory with Emilio, another PhD student. Regarding our work relationship, I always thought that we were the antithesis of a "*noble principle*" (acid and base, electron and proton, communism and capitalism, ying and yang). Despite we argued constantly and we had many personal clashes, I have always admired his relentless tenacity, which until now remains incomparable. In the same time another two key actors have played a fundamental role at this time. The first was Carlos, a veteran post doc, with an unbridled passion for sailing and for the sea life in general. Perhaps for this passion, when we were working together, I always compared our lab during the endless experiments imposed by the ruthless protein purification

protocols of NifB as a “*Spanish galleon*” in the middle of a storm and Carlos was the impetuous “*boatswain*” that shouts the orders to the crew. At the beginning I misunderstood this attitude, but nowadays I have realized that without such determination, we never have come out from the stormy tempest of such complex experiments. The second actor was Cesar, a post doc with many experiences on his shoulders and a strong capacity to be empathetic with the people around him. He always was close to me and he helped me out on my first steps in the molecular biology world.

But it was in the *second act*, when the principal comedians of this story, group of people that maybe influenced more my scientific and personal formation, entered in the *scenario*. This was the “*Landsknecht post docs team*”, so called because they were “mercenary researchers” coming from the most exotic countries (Morocco, India, US and France). In this group, there was Nasser, a perspicacious guy, from whom I learned to be attentive to the details and cautious in my decisions; Hemant, sunny and cheerful, always ready to comfort me with his maxims of Indian wisdom; Ryan, my mythical friend, pensive and precise scientist, used to communicate an idea with few good words. His ability to remain shiny and cold in the most complicated experiments has always reminded me *Clint Eastwood* in the movie “*Inspector Callaghan*”. Last but perhaps most important, Simon. This “*big boy*”, coming from the French mountains, despite his young age, from the first time has impressed me for his deep scientific maturity, combined with the passion and curiosity of a children. Not only he has helped me to grow up as a researcher, but he represented for me the “*fortress of the north*”, always ready to support me in my moments of discouragement. Curious but true, his surname *Arragain*, though French, presents some reminiscences with the Basque language and properly translated means “*stone*”.

A special thanks to all the other friends of the CBGP (Giammi, Nils, Manuel, Marcela and Felipe) that represented for me the best group of friends that I could imagine find, with whom I spent unforgettable nights, walking in the lost alleys and locales of Madrid as truly “*gatos madrileños*”.

Obviously, all of that will not happen without the help of my thesis director, Luis, that decided to take me aboard on this adventure. I have to admit that he had a lot of patience with me and maybe, in some occasion, I have taken advantage of that. Maybe after this experience, the only my regret is the fact that I was not able to create a constructive relationship with him. Indeed, I missed the opportunity to learn from him not only at the scientific level but also, and maybe more important, at personal.

And finally, last but by no means least, also to everyone in the Luis Lab, Institut de Biologie Structurale (Grenoble – France), David Britt lab at UC Davis (California - US) that had a direct or indirect impact in my work. It was great sharing laboratory with all of you during these four years.

A special thank to the Spanish Government that placed the special grant FPI (Formación de Personal Investigador), fundamental for the financing of my PhD research.

Thanks for all your encouragement!

TABLE OF CONTENTS

Abstract.....	5
Abbreviations and symbols.....	7
1. Introduction.....	11
1.1. The importance of nitrogen.....	12
1.1.1. The chemical nature of dinitrogen.....	12
1.1.2. Biogeochemical cycle of nitrogen.....	14
1.2. Diazotrophic microorganisms.....	15
1.2.1. Symbiotic diazotrophs.....	15
1.2.2. Free-living diazotrophs.....	16
1.2.3. Nitrogen fixing Archaea.....	17
1.2.4. Nitrogenase evolutionary history.....	17
1.3. The Mo-nitrogenase enzyme.....	18
1.3.1. Different structures and functions of [Fe-S] clusters.....	22
1.4. Roles of nitrogen fixation proteins in FeMo-co synthesis.....	23
1.4.1. Proteins that provides substrates: NifV, NifQ and NifS.....	24
1.4.2. Scaffold proteins: NifU, NifB, and NifEN.....	25
1.4.3. Metallocluster carrier proteins: NifX and NafY.....	26
1.4.4. Additional proteins required for the maturation of FeMo-co: NifH.....	27
1.5. NifB, a member of the Radical SAM superfamily.....	27
1.5.1. General overview of Radical SAM enzymes.....	28
1.5.2. A functional example of the auxiliary [Fe-S] clusters in radical SAM enzymes: LipA.....	30
1.5.3. Timeline of NifB and NifB-co discoveries.....	32
2. Objectives.....	35
3. Materials and Methods.....	38
3.1. Reagents.....	39
3.2. Bacterial strains and growth conditions.....	39
3.3. Plasmids.....	41
3.4. Data mining and phylogenetic analysis.....	43
3.5. NifB_{Mi} purifications.....	50
3.5.1. IMAC chromatography.....	52
3.5.2. HiPrep 26/10 desalting chromatography.....	53
3.5.3. Superdex 200 size exclusion chromatography.....	53
3.6. Preparation of NifB_{Mi} apo-protein.....	54
3.7. Reconstitution of Fe-S clusters.....	54
3.8. Reduction of NifB_{Mi} [Fe-S] clusters.....	55
3.9. Growth and preparation of UW140 cell-free extracts for in vitro FeMo-co insertion and nitrogenase reconstitution assays.....	55
3.10. NifB_{Mi} <i>in vitro</i> activity assays.....	56

3.11. Gas Chromatography.....	58
3.12. SDS Poly Acrylamide Gel Electrophoresis.....	59
3.13. Protein determination.....	59
3.13.1. Fe quantification.....	59
3.13.2. Sulfur quantification.....	59
3.13.3. Mass Spectroscopy.....	60
3.13.4. UPLC/ESI-qTOF-MS analysis of reaction sub-products.....	60
3.13.5. Deuterium and ¹⁵ N-SAM substitution experiments.....	61
3.13.6. UV-Visible spectroscopy.....	61
3.13.7. Electron paramagnetic resonance (EPR)	62
3.13.8. One-dimensional (1D) electron-spin echo envelope modulation (ESEEM)	64
3.13.9. Two-dimensional (2D) electron-spin echo envelope modulation (HYSORE)	64
3.13.11. Crystallography screenings.....	65
 4. Results.....	 70
 Chapter I: Phylogenetic studies and purification of the NifB _{Mi} protein.....	 71
4.1. Phylogenetic Distribution of Three Distinct NifB Domain Architectures	71
4.1.2. Organization of <i>nif</i> genes in the genome of <i>M. infernus</i>	72
4.2. NifB _{Mi} is functional <i>in vivo</i> when expressed in the aerobic mesophilic host <i>A. vinelandii</i>	75
4.2.1. Heterologous expression and purification of wild type NifB from <i>Methanocaldococcus infernus</i> (NifB _{Mi}) and the site-directed variants SM, AM and DM	77
4.3. Elution profile of NifB _{Mi} wild type on Superdex 200 column used for Crystallography and Spectroscopic studies.....	80
 Chapter II: Biochemical characterization of NifB _{Mi} WT and site- directed variants.....	 83
4.4. UV-visible absorption and metal content of NifB _{Mi}	83
4.4.1. Site-directed NifB _{Mi} variants.....	85
4.5. NifB _{Mi} enzymatic activity.....	88
4.5.1. Preliminary study of wild-type NifB _{Mi} with UW140 crude extract at different temperatures during FeMo-co synthesis.....	88
4.5.2. <i>In vitro</i> NifB _{Mi} assays with different levels of substrate and UW140 proteins.....	90
4.5.3. NifB _{Mi} dependent FeMo-co synthesis and NifDK activation assays in a system with purified Nif proteins.....	92
4.5.4. Characterization of the essential elements for the <i>in vitro</i> NifB _{Mi} assay with purified Nif proteins.....	94
4.6. Detection of NifB _{Mi} radical SAM activity by-products.....	97

4.6.1. SAH and AdoH production correlated with NifB _{Mi} - dependent activation of nitrogenase over time.....	98
Chapter III: Spectroscopic studies of NifB_{Mi} WT and site-directed variants.....	101
4.7. Electron paramagnetic resonance (EPR) of NifB _{Mi} wild type and mutant variants.....	101
4.7.1. Temperature dependence and spectral subtractions of WT and site-directed variants of NifB _{Mi}	103
4.8. ESEEM of NifB _{Mi} wild type and site-directed mutant variants.....	105
4.9. Preliminary HYSCORE studies of NifB _{Mi} WT and DM proteins.....	107
Chapter IV: NifB_{Mi} crystalization trials.....	109
4.10. Crystallization trials.....	109
5. Discussion.....	114
5.1. NifB phylogeny and architecture.....	115
5.2. Ancestral NifB protein from a strict anaerobic and thermophilic Archaea functions <i>in vivo</i> in an aerobic mesophilic host.....	116
5.3. Purified NifB _{Mi} is active in the in vitro FeMo-co synthesis assay.....	117
5.4. Spectroscopic characterization of NifB _{Mi} and mechanistic implications.....	118
5.5. Crystallogenesis trials.....	120
6. Conclusions.....	122
7. References.....	125

Abstract

Nitrogen is a limiting nutrient for plant growth, both in natural and agricultural ecosystems. Biological nitrogen fixation (BNF) is a process catalyzed by metallo enzymes called nitrogenases. There are three types of nitrogenases depending on the metal group present at the active site, the most abundant being the Mo-nitrogenase, which carries the iron-molybdenum cofactor (FeMo-co). The genes encoding the nitrogenase related proteins are found exclusively in prokaryotes and are organized into operons, termed *nif*. Mo-nitrogenase is a two-component system composed of a NifD₂K₂ heterodimer, where each NifDK half carries two metal cofactors (one P-cluster and one FeMo-co), and a NifH homodimer that carries a [Fe₄S₄] at the subunit interface. FeMo-co is arguably the most complex modified [Fe-S] cluster in Biology. Its biosynthesis requires highly specific reactions that are tightly regulated. NifB catalyzes the first essential reaction in FeMo-co biosynthesis. Interestingly, NifB activity is required for all nitrogenases. NifB is a radical *S*-adenosylmethionine (SAM) enzyme that synthesizes the [Fe₈S₉C] precursor of FeMo-co, named NifB-co. NifB contains the signature [Fe₄S₄]-SAM binding motif at its N terminus and a number of additional cysteinyl and histidiny residues that could coordinate additional [Fe-S] clusters for NifB-co synthesis. Due to its extreme O₂sensitivity and protein instability, NifB must be purified under strict anaerobic conditions by rapid chromatography methods. Our studies and overlapping studies of other group show that, in the presence of SAM, NifB couples two [Fe₄S₄] clusters and generates and inserts a carbide atom to form NifB-co. In this thesis, we have contributed and developed methods for NifB purification from cells of the facultative anaerobic non-nitrogen fixing bacterium *Escherichia coli* expressing *nifB* from the thermophilic archaea *Methanocaldococcus infernus* (NifB_{Mi}). The stability of NifB_{Mi} allowed a deep biochemical and spectroscopic characterization of this enzyme. In our study, we demonstrated that NifB_{Mi} is a radical SAM enzyme able to reductively cleave SAM to 5'-deoxyadenosine radical and is competent in FeMo-co maturation. Furthermore we characterize the metal cluster organization of this protein through electron paramagnetic resonance spectroscopy and electron spin echo envelope modulation, offering mechanistic insights into NifB-co synthesis.

Abbreviations and symbols

Å	Angstrom
ADP	Adenosine diphosphate
Ado•	5'-deoxyadenosyl radical
AdoH	5'- deoxyadenosine
AM	Auxiliary cluster NifB _{Mt} mutant
ATP	Adenosine triphosphate
BioB	Biotin synthase
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
Da	Dalton
DNA	Deoxyribonucleic acid
DM	SAM and Auxiliary cluster NifB _{Mt} double mutant
DMPD	Dimethyl-4-phenylenediamine
DTH	Sodium dithionite
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
ESE	Electro Spin Echo
ESEEM	Electron spin echo envelope modulation
FeFe-co	Iron-iron cofactor
FeMo-co	Iron-molybdenum cofactor
FeV-co	Iron-vanadium cofactor
GHz	Gigahertz
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
HPLC	High performance liquid chromatography
HemN	Heme biosynthesis enzyme
HYSCORE	2D hyperfine sub-level correlation

LAM	Lysine 2,3-aminomutase
Ma	<i>Methanosarcina acetivorans</i>
Mi	<i>Methanocaldococcus infernus</i>
MoaA	Molybdopterin biosynthesis protein
Mt	<i>Methanobacterium thermoautotrophicum</i>
NifB-co	NifB cofactor
Nif	Gene designation for molybdenum-dependent nitrogen fixation
ORF	Open reading frame
PCR	Polymerase chain reaction
PLP	Pyridoxal phosphate
ppm	Part per million
OD	Optical Density
Radical-SAM	Enzyme with a [Fe ₄ S ₄] center able to form Ado• from SAM
SAM	S-adenosylmethionine
SDS	Sodium dodecyl sulfate
SM	SAM cluster NifB _{Mi} mutant
TAE	Tris-acetate-EDTA buffer
Tris	tris(hydroxymethyl)aminomethane
TFA	Trifluoroacetic acid
UV	Ultraviolet

Introduction

1.1. The importance of nitrogen

Nitrogen is found in a wide variety of organic compounds such as proteins, nucleic acids, porphyrins, coenzymes, and many other cellular components. However, nitrogen is often a limiting nutrient, despite its ubiquitous presence as nitrogen gas (N_2) in air and water. Biologically available nitrogen is essential for all living organisms. On early Earth, fixed nitrogen was generated exclusively by abiotic processes such as lightning (1). The appearance of biological nitrogen fixation suggests that the demand for fixed nitrogen eventually exceeded the supply from abiotic sources. Nitrogen fixation was and still is fundamental to the biosphere because it converts the inert N_2 into a chemical form accessible to metabolic enzymes for biosynthesis of the diverse set of organo-nitrogen compounds necessary for life.

1.1.1. The chemical nature of dinitrogen

Nitrogen is a Group 15 element. The N_2 molecule exhibits a triple bond consisting of one σ and two π orbitals that are all occupied (Figure 1). The corresponding antibonding orbitals are empty.

Figure 1. Density of states and shape of N_2 orbitals. The $p-\pi^*$ and the $p-\sigma^*$ states are unoccupied (2).

This element, locked in the kinetically stable form N_2 , seemed so chemically inert that Lavoisier named it azote, meaning "without life". The energetic requirement of

nitrogen fixation is limited by its thermodynamics with the overall reaction being thermodynamically favorable (3). The standard free-energy change (ΔG°) equals -7.7 kcal/mol at 298 K and 1 atm pressure.

The biological reduction of N_2 to ammonia is also thermodynamically favorable with ΔG° equal to -15.2 kcal/mol at 298 K and pH 7.

Although thermodynamically favorable, the stability of N_2 triple bond severely affects reaction kinetics. The difference between the bond energy levels of the triple, double, and single bonds of N_2 (225, 100, and 39 kcal/mol, respectively) is also crucial in making N_2 practically inert. Thus, the primary energetic barrier to N_2 reduction is the triple bond while the subsequent reduction of N_2 double and single bonds is energetically favorable (4). The significant stability of the N_2 triple bond compared to the double and single bonds is reflected in the enthalpies of formation of the following compounds:

1.1.2. Biogeochemical cycle of nitrogen

There are two main reservoirs of inorganic nitrogen in nature: the atmosphere, which is 78% N_2 , and nitrate ion (NO_3^-) found in aerobic and anaerobic waters and soils. The biogeochemical cycle of nitrogen can be outlined in four major processes (Figure 2).

Figure 2. Simplified scheme of the nitrogen cycle (5).

Nitrification is the biological process that converts, by oxidation, ammonia to nitrite and nitrate. Two main bacterial groups are responsible for nitrification: *Nitrosomonas*, which oxidize ammonia to nitrite, and *Nitrobacter* that oxidize nitrite to nitrate (6). *Denitrification* is the production of gaseous N_2 from the reduction of nitrate and nitrite. This is an anaerobic respiratory process carried out by many bacteria, fungi and yeast whereby N-oxides act as terminal electron acceptors of respiration (7). *Nitrogen fixation* can occur through biotic (biological) or abiotic processes. Abiotic nitrogen fixation occurs mainly by lightning, and fixes more than 10^{12} g N yr^{-1} (8). The biological nitrogen fixation (BNF) accounts for the vast majority of fixed N_2 and is carried out by a limited, but diverse, group of Bacteria and Archaea (9). Thus, N_2 enters the biological nitrogen cycle mainly through catalysis by the prokaryotic enzyme nitrogenase, which we will detail in the next pages. Nitrogenase catalysis is extremely energy demanding and requires low potential electrons and energy released by ATP hydrolysis (10). The minimal stoichiometry for the reduction of N_2 to NH_3 with the concomitant formation of one molecule of H_2 is usually depicted as follows:

1.2. Diazotrophic microorganisms

Although there is an enormous reservoir of N_2 in the atmosphere, it is non reactive and most organisms are unable to metabolize it. Organisms able to fix N_2 are defined as *diazotrophs* (nitrogen eaters). Thus, input of metabolically usable forms of nitrogen by the activity of diazotrophs is essential to all ecosystems.

BNF is exclusively prokaryotic. For many years, it was thought that only a few bacterial species were nitrogen fixers. However, more recently, it has been shown that most Bacteria phyla and also methanogenic Archaea have representatives capable of performing BNF (11). In fact, diazotrophs are present in practically all ecosystems and environments as diverse as aerobic soils, anaerobic muds, all sorts of water ecosystems and plant rhizosphere. Importantly, nitrogenase is irreversibly inactivated in the presence of O_2 (10) and diazotrophs living in aerobic environments have developed different ways of dealing with O_2 damage. Due to this O_2 sensitivity and to the high energetic cost of BNF, there are a high variety of nitrogenase-regulating schemes in the different nitrogen-fixing species.

1.2.1. Symbiotic diazotrophs

In the mutualistic symbioses between plants and bacteria, the N_2 -fixing bacteria symbionts exist in specialized structures (nodules) inside the roots of the host plants. In these associations, the transfer of fixed N from the bacterium to the plant is relatively low, and the relationship could be viewed as opportunistic for the bacterium rather than mutualistic (12). The ability to fix nitrogen symbiotically in nodules of vascular plants is found in three major groups of Eubacteria.

The Rhizobia group includes species from the genera *Allorhizobium*, *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, and *Sinorhizobium*. They associate essentially with leguminous plants of the Fabaceae family (13).

The Frankia group associates with a broader spectrum of plants. Whereas host-

specificity, i.e. the selectivity of the plant host for specific bacteria species, is useful to designate rhizobia species, this is not as effective in *Frankia* (14). This group of bacteria infects actinorhizal plants forming nodules that consist of several lobes each one having a structure similar to a lateral root.

The third group is formed by symbiotic cyanobacteria forming well-developed root-symbioses. The dominant microsymbiont within this group is *Nostoc/Anabaena*. Some cyanobacteria are also associated with fungi as lichens, liverworts, fern, cycads (15), corals (16), sponges (17), and insects (18). An example of symbiotic interaction important for the agriculture is the association of *Anabaena* with the water fern *Azolla* harboring. Cyanobacteria do not trigger nodule formation but instead fixed nitrogen in specialized anaerobic cells named heterocysts.

1.2.2. Free-living diazotrophs

Free-living diazotrophs are ubiquitous in soil, lakes and seas. Their great ecological diversity extends to their physiological diversity, as N_2 fixation is carried out by chemotrophs and phototrophs, as well as by autotrophs and heterotrophs. Usually, free-living diazotrophs are classified into three groups according to their relationships with O_2 :

Anaerobes that live exclusively in habitats with low O_2 concentration, such as ocean sediments and soils, and carry out fermentation or anaerobic respiratory or photosynthetic processes. Examples of this group are species from the genus *Clostridium*, *Desulfovibrio* and *Desulfobacter* (13). Facultative anaerobes, which grow either using or not using O_2 but that only fix N_2 anaerobically. Examples include *Klebsiella oxytoca*, *Bacillus polymyxa* and *Bacillus macerans* (13). Finally, strict aerobes that require O_2 to grow although their nitrogenase is still damaged if exposed to O_2 . *Azotobacter vinelandii* is the most studied of this group. It uses very high respiration rates and protective proteins to prevent damage by O_2 . Many other species also reduce the O_2 levels in similar ways but they exhibit lower respiratory rates and lower O_2 tolerance (13).

1.2.3. Nitrogen fixing Archaea

Archaea are distributed in two phylogenetic branches based on 16S rRNA sequence comparisons: the Euryarchaeota, which contains the anaerobic methanogens and extreme halophiles, and the Crenarchaeota that contains the extreme thermophiles. Methanogenic archaea presents distinct characteristics, including a strictly anaerobic lifestyle and a tendency toward thermophily (19). Methanogenic archaea contains N₂-fixing species, a fact that broadens the perspective of nitrogen fixation. Nitrogen fixation in Archaea and Bacteria are evolutionarily related and their molybdenum nitrogenases share the same fundamental mechanisms. Among this class of organisms, *Methanocaldococcus infernus* is an example of extreme thermophile isolated from a 3,000-meter deep-sea hydrothermal chimney in the Mid-Atlantic Ridge. *M. infernus* grows between 55°C and 91°C, with optimum growth at 85°C. According to the 16S gene sequence analysis, *M. infernus* is closely related to the non-nitrogen fixing *Methanococcus jannaschii*.

1.2.4. Nitrogenase evolutionary history

The origin of nitrogenase enzyme is a controversial issue and its complex genetic history is replete of events of gene duplication, gene recruitment, gene fusion, and horizontal gene transfer. To understand nitrogenase evolution it is important to consider selective pressures that might have affected the stability and distribution of diazotrophy. The principal source on organic nitrogen in early prebiotic environments was generated by lightning discharges (between 10⁷ and 4 x 10⁷ tons of N per year) (1). It is thought that this amount of abiotically generated organic nitrogen was sufficient for the development of the first forms of life. However the sudden reduction in the supply of fixed nitrogen that might have occurred soon after the origin of life, as the prebiotic source of organic material was depleted by the emerging life forms, may have forced the innovation of BNF (20, 21).

An interesting hypothesis suggested that Anf or Vnf nitrogenases, rather than the Nif nitrogenase, were responsible for BNF before the rise of O₂ (2.5 Gy). In these conditions oceans were depleted in Mo (22, 23) due to the insolubility of Mo-sulfides under anoxic conditions (24). More recent studies strongly suggest that the Mo-nitrogenase originated in methanogens and predated the V and the Fe-only

nitrogenases. These studies also indicate that the duplication of *nifDK* to *nifEN*, and thus the ability to mature NifB-co into FeMo-co in an external scaffold, occurred in an ancestor of hydrogenotrophic methanogens. Further movement into the Bacteria domain occurred through a lateral gene transfer event between Methanogen and Firmicute ancestors (25). It should be noted that, although it is convenient to classify nitrogenases according to their metal content, it is possible that some Mo-containing nitrogenases may share more similarity to either V nitrogenases or Fe-only nitrogenases than to conventional Mo-nitrogenases (26).

1.3. The Mo-nitrogenase enzyme

Mo-nitrogenase consists of two component proteins (Figure 3) named dinitrogenase reductase (also called component II or Fe protein) and dinitrogenase (also called component I or MoFe protein). Dinitrogenase reductase is a 60-kDa homodimer of the *nifH* gene product with a [Fe₄S₄] cluster located between the two subunits and one site for MgATP binding and hydrolysis in each subunit (27). The Fe protein is the obligate electron donor to the MoFe-protein. In addition, it is required for P-cluster and FeMo-co assembly. Dinitrogenase is a 220-kDa $\alpha_2\beta_2$ heterotetramer of the *nifD* and *nifK* gene products. The MoFe protein contains two pairs of two complex metal centers: the [Fe₈S₇] P-clusters working in inter-protein electron transfer and the iron-molybdenum cofactors (FeMo-co) that are the sites for substrate binding and reduction (28, 29, 30). The closest edge-to-edge distance from the P-cluster to the inorganic portion of FeMo-co is 14 Å, with this space largely occupied by FeMo-co homocitrate and a network of hydrogen-bonded water molecules. Both clusters are shielded from direct exposure to water by the surrounding NifDK polypeptides. Inter and intramolecular electron transfer, as well as substrate entry and product release, involves a carefully regulated sequence of interactions between the MoFe protein and the Fe protein.

Nitrogenase metallocenters have served as models for biophysicists and synthetic inorganic chemists due to their unique chemical and physical properties among [Fe-S] clusters. Numerous spectroscopic and X-ray absorption studies have been conducted to describe their structure, cluster oxidation states and reactivity. The P-cluster, which participates as reducing agent inside the MoFe-protein, presents cysteinyl thiol

ligands and a potential interaction with the hydroxyl group of a serine residue (β -Ser¹⁸⁸) (Figure 4) suggesting that neighboring residues could also participate in cluster integrity (32, 33).

Figure 3. Mo-nitrogenase component proteins. (A) $\alpha_2\beta_2$ tetramer of the *A. vinelandii* MoFe protein. α subunits are blue and green, and β subunits are red and orange. (B) Detail of one of the α subunits of the MoFe protein showing the positions of domains I (blue), II (light blue), and III (dark blue). (C) *A. vinelandii* Fe protein (subunits in blue and green). Atom color is: iron, magenta; molybdenum, grey; sulfur, yellow; carbon, green; and oxygen, red (31).

Figure 4. Stick and sphere representation of the P-cluster. Yellow: sulfur; Purple: iron. P-cluster coordinating MoFe protein amino acid residues are shown to the right.

FeMo-co consists of a $[\text{MoFe}_7\text{S}_9\text{C}]$ core that can be viewed as $[\text{MoFe}_3\text{S}_3]$ and $[\text{Fe}_4\text{S}_3]$ sub clusters bridged by three μ_2 -sulfides and one μ_6 -carbide (C^{4-}) ion with its Mo atom further coordinated by *R*-homocitrate (34) (Figure 5). As for the P-cluster, the neighboring amino acid residues profoundly affect FeMo-co spectroscopic features. The resting state of this cofactor inside the MoFe protein has a characteristic $S = 3/2$ EPR signal with g values close to 4.3, 3.7, and 2.0. Although FeMo-co was first isolated by Shah and Brill in 1977 (35), and then crystallized within the MoFe protein in 1992 (36), its complete atomic structure remained ambiguous until 2002 (37), due to the incapacity to identify the C^{4-} atom in the center of the cofactor. FeMo-co isolated from the protein by various solvent extractions present single electron redox chemistry (38, 39, 40, 41).

Figure 5. Stick and sphere representation of FeMo-co. Color atoms legend: Purple: iron, Yellow: sulfur, Green: carbon, Red: oxygen, Black: molybdenum. FeMo-co coordinating MoFe protein amino acid residues are shown to the right.

NifH is required both to assemble the MoFe protein P-clusters and for FeMo-co biosynthesis (42, 43). In addition, the synergistic interaction between NifH and NifDK involves the association and dissociation of these two component proteins during the reduction of N_2 to NH_3 . The reduction of a single molecule of N_2 requires multiple electron transfer events, and dissociation of the complex likely represents the rate limiting step during turnover, at least in the experiments carried out *in vitro* (44). The Fe protein integrates nucleotide binding and MgATP hydrolysis to electron transfer between its $[Fe_4S_4]$ cluster and the P-cluster of the MoFe protein. It has been demonstrated that, during catalysis of N_2 , electrons are delivered one at a time from the Fe protein to the MoFe protein in a gated process that involves association and dissociation of the component proteins and the hydrolysis of at least two MgATP molecules for each electron transfer event.

1.3.1. Different structures and functions of [Fe-S] clusters

The [Fe-S] clusters are probably one of the most ancient and versatile prosthetic groups present in all of biology. Different types of [Fe-S] clusters have been found in proteins (45) and can be classified according to their metal center complexity. Helmut Beinert pioneered the detection of [Fe-S] clusters in proteins by using EPR spectroscopy (46). In the recent years, knowledge of these biomolecules has grown exponentially due to a combination of biochemical, spectroscopic and crystallographic studies. These studies have revealed a wide variety of protein-bound centers [Fe-S] in terms of their structures and their biological functions. The most common forms found in Fe-S proteins are the rhombic [Fe₂S₂] and the cubane [Fe₄S₄], which are normally coordinated to the polypeptides by the thiolate side chains of cysteine residues (Figure 7). In some cases [Fe-S] clusters can be coordinated by a carboxylate, like in ferredoxin from the *Pyrococcus furiosus* (3 Cys, 1 Asp) (47, 48) or nitrogen from a histidine residue, like in the Rieske protein (2 Cys, 2 His) (49). It is also interesting that a number of [Fe-S] proteins carry several [Fe-S] clusters. A extreme example is the respiratory Complex I (NADH-ubiquinone oxidoreductase), which contains 8 or 9 [Fe-S] clusters (50).

The [Fe-S] proteins are wide spread in Archaea, Bacteria and Eukaryotes (51) where they carry out a variety of physiological functions. [Fe-S] proteins are involved in electron transfer, redox and non-redox catalysis, gene regulation, interaction and activation of substrate, heterometal trafficking or even maintaining protein structures (52).

The [Fe-S] clusters result from the assembly of iron atoms in different oxidation states (ferrous iron or ferric iron Fe²⁺/Fe³⁺) and sulfide ions (S²⁻) in variable stoichiometric. In the cell, [Fe-S] clusters are assembled on scaffold proteins and later on incorporated into the apo-forms of related protein metalloproteins. However, they can often be assembled *in vitro* directly of the apo-protein by the simple addition of Fe²⁺ and S²⁻ under reducing conditions (53) (Figure 6).

More complex [Fe-S] clusters, with nuclearity higher than 4, present distinct structures and are found in a few specific enzymes. They are frequently associated to other metals, like Ni and Mo. Examples are the “C-cluster” of the carbon monoxide dehydrogenase, characterized by a [Fe₃S₄] cluster associated with a Ni-Fe diatomic

center (54), the “A-cluster” of acetyl-CoA synthase formed by [Fe₄S₄] cluster associated with Ni-Ni center, and the FeMo-co of nitrogenase, which contains a [Fe₇S₉C] cluster associated to a Mo atom and a molecule of *R*-homocitrate.

[1Fe-1S]	[2Fe-2S]	[3Fe-3S]	[4Fe-4S]

Redox state			
Fe ²⁺ Fe ³⁺	Fe ^{3+/2+} → [2Fe-2S] ²⁺ Fe ^{3+/2+} → [2Fe-2S] ¹⁺	3Fe ³⁺ → [3Fe-4S] ¹⁺ 2Fe ^{3+/2+} → [3Fe-4S] ⁰	3Fe ^{3+/2+} → [4Fe-4S] ³⁺ 2Fe ^{3+/2+} → [4Fe-4S] ²⁺ Fe ^{3+/2+} → [4Fe-4S] ¹⁺
Potential redox			
+20 à -100 mV	-100 à +400 mV	-60 à -460 mV	-100 à -700 mV p.H 7

Figure 6. Most common [Fe-S] cluster structures. Geometry, composition, redox states and redox potentials for biologically relevant [Fe-S] clusters are shown (55).

1.4.Roles of nitrogen fixation proteins in FeMo-co synthesis

The cofactor of the nitrogenase, FeMo-co, and the products of nitrogenase structural genes (*nifHDK*) are not active without a complex network of *nif*-specific gene products that cooperate in processing nitrogenase structural components to their active forms. Different proteins participate in FeMo-co biosynthesis. Among them, NifB plays a crucial role in the synthesis of the FeMo-co precursor, NifB-co, and its catalytic activity will be explained in the following sections.

Two different major strategies for metal cofactor biosynthesis are found in nature. In some cases, the cofactor is assembled directly into the final target protein. The nitrogenase [Fe₈S₇] P-cluster is an example of proteins where in situ cofactor assembly takes place (56). In the case of FeMo-co, the cofactor is independently of the production of the structural polypeptides in a biosynthetic pathway comprising many *nif* gene products (57). At the end, fully assembled FeMo-co is inserted into apo-NifDK to render mature, active MoFe protein (58, 59).

We can classify the Nif proteins involved in FeMo-co synthesis into four groups: *i*) enzymes that provide and regulate substrate availability (NifS, NifV and NifQ), *ii*) scaffolds proteins for the synthesis and maturation of the FeMo-co (NifU, NifB and NifEN), *iii*) metallocluster carrier proteins involved in the transport of the FeMo-co intermediates between scaffolds and into apo-NifDK (NifX and NafY); and *iv*) proteins involved in electron transfer and maturation of the FeMo-co intermediates (NifH). Among them, NifH, NifB, NifE, NifN, and NifV are fundamental *in vitro* and *in vivo* and, without their participation, only apo-NifDK is synthesized (Figure 7).

1.4.1. Proteins that provides substrates: NifV, NifQ and NifS

The organic part of FeMo-co is constituted by the tricarboxylic acid *R*-homocitrate, which is synthesized by the homocitrate synthase NifV from the condensation of acetyl coenzyme and 2-OG (60). Mutants defective in *nifV* exhibit have altered substrate reduction properties (61) and very decreased activities of Mo-, V- and Fe-only nitrogenases (62), suggesting that homocitrate is also part of FeV-co and FeFe-co.

FeMo-co synthesis assays with purified components showed that NifQ could serve as unique Mo source for FeMo-co synthesis *in vitro* (59). In fact, this protein is present in many diazotrophic organisms. Primary sequence analysis shows that NifQ is unrelated to others proteins involved in molybdenum trafficking. NifQ coordinates a [Fe₃S₄Mo] cluster (63) probably through a conserved Cx₄Cx₂Cx₅C motif.

NifS, is a pyridoxal phosphate (PLP) containing homodimer able to desulfurate L-cysteine into L-alanine creating as final product a protein-bound persulfide at a conserved Cys residue (64). NifS provides sulfur atoms for the assembly of [Fe-S] clusters, mainly on the NifU scaffold (65). The persulfide is presumed to be the activated form of sulfur required by the scaffold protein NifU for [Fe-S] cluster assembly (65).

Figure 7. Nitrogenase metallocluster biosynthesis. The scheme illustrates the steps in nitrogenase metallocluster assembly. Early steps involve NifS and NifU activities for the assembly of NifH [Fe₄S₄] cluster and the precursors of P-cluster and FeMo-co of NifDK. P-cluster biosynthesis occurs in situ by the NifH-NifZ dependent condensation of the [Fe₄S₄] pairs. FeMo-co biosynthesis occurs outside NifDK. NifB catalyzes the first committed step of FeMo-co biosynthesis by assembling NifB-co, the central [Fe₈S₉C] core of FeMo-co. NifB-co is then transferred to the NifEN scaffold via NifX. Maturation to FeMo-co occurs within a putative NifEN/NifH complex by sequential addition of Fe, Mo and homocitrate. Finally, NafY transfers FeMo-co to apo-NifDK to generate active NifDK protein.

1.4.2. Scaffold proteins: NifU, NifB, and NifEN

NifU is a homodimer consisting of three defined and conserved domains: a central domain with a permanent [Fe₂S₂] cluster coordinated by four conserved cysteine residues and N-terminal and C-terminal domains with two and three conserved cysteine residues, respectively, that participate in the assembly of transient [Fe-S] clusters (66). The function of this scaffold protein is the transient assembly of [Fe-S] clusters destined to the apo-forms of nitrogenase structural and biosynthetic proteins. In vitro, in the presence of L-cysteine, Fe²⁺ and NifS, NifU is able to assemble and transfer transient [Fe₄S₄] to the apo-form of NifH (67).

Another scaffold protein fundamental to FeMo-co synthesis *in vivo* (68) and *in vitro* (59) is NifEN. This protein is $\alpha_2\beta_2$ heterotetramer of 200 kDa, carrying two [Fe₄S₄] clusters located at the interface of both subunits (69). Several similarities exist between NifEN and NifDK, including their amino acid sequences (70), the location of metal clusters within the proteins (71) and the ability to catalyze C₂H₂ reduction albeit at very low rate (72). All these elements suggest that their encoding genes derived by an in-tandem paralogous duplication of NifDK (9).

NifEN has a central role in the synthesis of FeMo-co participating in the maturation of NifB-co into FeMo-co, passing through an intermediate state called the VK-cluster. Similar to NifB-co, the VK cluster is composed of [Fe₈S₉C] their electronic properties are different, since NifB-co is diamagnetic (73) and the VK-cluster paramagnetic (74). The fine structure underlying this difference is not known.

The NifB scaffold belongs to the family of SAM-radical proteins and its function is explained in detailed in the following sections.

1.4.3 Metallocluster carrier proteins: NifX and NafY

FeMo-co biosynthetic intermediates are mobilized between scaffold proteins. Specific carrier proteins participate in the mobilization and transfer of these intermediates for several reason: *i*) the prosthetic groups are O₂ sensitive and are expected to be always protein bound to proteins within the cell (75), *ii*) the high energy demand of nitrogen fixation requires perfect synchronism and efficient coordination among the nitrogenase proteins that participate in FeMo-co biosynthesis.

The 17-kDa protein NifX is able to bind complex metal clusters as FeMo-co, is part of a conserved small group of nitrogenase related proteins that includes VnfX, NifY, NafY, and VnfY (76). Several lines of evidence indicate this protein has two principal functions carried out through transient interactions with NifB and NifEN: first, the transfer of NifB-co from NifB to NifEN, and second the storage of FeMo-co precursors under stress conditions thus regulating their availability.

The 26-kDa protein NafY presents two distinct domains, each one of them has specific and different roles. Its 12 kDa N-terminal domain is a four-helix bundle that participates in protein–protein interaction and stabilization of apo-NifDK (77). Its 14

kDa C-terminal domain binds FeMo-co without the participation of any intermediates and promotes its insertion into apo-NifDK. In particular, mutational analysis indicated that a conserved His¹²¹ of NafY is directly involved in FeMo-co binding (78).

In vitro studies have demonstrated that NifX and NafY present high affinity for only one species of metal cluster intermediates, specifically binding to NifB-co and FeMo-co (76, 77). This incompatible substrate affinity among the nitrogenase metal cluster transporters in combination with the apo-NifDK capacity to bind NifB-co or/and FeMo-co, could be explained as the natural mechanism to prevent the insertion of prosthetic intermediates in to the nitrogenase active site.

1.4.4. Additional proteins required for the maturation of FeMo-co: NifH

NifH holds at least three different key functions in the nitrogenase system: *i*) electron transfer to the MoFe protein during catalysis with concomitant hydrolysis of MgATP, *ii*) assembly of P-clusters from pairs of [Fe₄S₄] clusters precursors, and *iii*) it is fundamental in FeMo-co biosynthesis through its interaction with NifEN. The capacity of NifH to transfer electrons is only required for the catalytic reduction of N₂ and not in P-cluster or FeMo-co biosyntheses (79, 80). Importantly, the incorporation of Mo into FeMo-co precursor is dependent of NifH, although the exact mechanism of how this happens is not known (59).

1.5. NifB, a member of the Radical SAM superfamily

The synthesis of complex precursors of FeMo-co from small [Fe-S] cluster precursors is attributed to the NifB enzyme/scaffold. As was early demonstrated by Shah and colleagues (35) this protein plays a key function in the synthesis of NifB-co (81), an early precursor to FeMo-co (63, 82). The inability for diazotrophic growth of strains carrying *nifB* deletion (regardless of the availability of Mo and V) and the absence of a *nifB* analogue in the V- and Fe-only nitrogenase gene clusters of *A. vinelandii*, suggested that *nifB* was essential for the maturation of all nitrogenases (58, 83). The idea of a shared NifB enzyme for the three nitrogenases is also supported by the *nifB* location being in a separate cistron from that of the Mo-nitrogenase genes, *nifHDKENXUSV* (Figure 8).

Figure 8. Simplified depiction of *nif* gene clusters in *A. vinelandii*. The *nif* gene cluster is under the control of NifA and NifL proteins sensitive to the levels of oxygen and fixed nitrogen (C/N ratio).

NifB has eluded detailed biochemical characterization due to the extreme lability of its [Fe-S] clusters to O₂ and the instability of the enzyme, which limit purification procedures. Surprisingly, NifB catalytic product (NifB-co) was purified earlier than the NifB protein, and its role as intermediate in FeMo-co biosynthesis clearly established (84). It was Curatti and collaborators who first succeeded in purifying NifB from *A. vinelandii* (NifB_{Av}) demonstrating its essential role in the in vitro FeMo-co synthesis assay and its S-adenosylmethionine dependence, which confirmed the involvement of radical chemistry in NifB-co synthesis (85).

1.5.1. General overview of Radical SAM enzymes

Multiple sequence alignment studies of NifB homologs reveal several regions of high identity (86). The N-terminal region shows very high conservation with three cysteine residues within the motif CX₃CX₂C being completely conserved. This “Cys box” motif defines the Radical SAM superfamily (Figure 9). This super family currently comprises more than 2,800 proteins that participate in more than 40 distinct biochemical transformations in metabolism regulation, cofactor biosynthesis, nucleic-acid repair and modification, although most members have not been biochemically characterized (87). Phylogenetic studies indicate that the Radical SAM enzymes are of ancient origin, probably among the earliest biological catalysts to function by radical mechanisms (87).

Figure 9. Conserved amino acids of selected Radical SAM superfamily members and the mechanism of cleavage of SAM. The conserved cysteines in the CX₃CX₂C motif are displayed in red. On the panel to the top of the sequence alignment is a [Fe₄S₄] cluster bound to the conserved cysteines. SAM is ligated through the amino and carboxylate groups of [Fe₄S₄]. Atoms are depicted in brown for Fe, yellow for S, dark blue for methyl group CH₃, black for O and N. From (86, 88, 89).

S-adenosyl methionine (SAM) plays a key role in this family of enzymes. The initiating step in the SAM Radical super family is the reductive cleavage of SAM into 5'-deoxyadenosyl radical (Ado•) and methionine catalyzed by a SAM-bound [Fe₄S₄] cluster. The highly reactive Ado• intermediate readily reacts by performing stereoselective H-abstraction on a substrate (90, 91). Among these proteins, the most extensively studied are represented in Figure 10. HemN is the coproporphyrinogen oxidase of anaerobic bacteria (91, 92) that catalyzes the oxidative decarboxylation of coproporphyrinogen. BioB is involved in biotin biosynthesis catalyzing the enzymatic insertion of sulfur atoms into inactivated C–H bonds to form the thiophane ring of biotin. MoaA is involved in the first step of molybdopterin biosynthesis, the precursor of the molybdenum cofactor (Mo-co). This protein contains two [Fe₄S₄] clusters, the SAM-bound cluster and a second [Fe₄S₄] cluster that binds the guanine ring of GTP to produce precursor Z (93, 94, 95). The lysine 2,3-aminomutase (LAM) has a role in the metabolism of L-lysine in Clostridia. LAM contains pyridoxal-phosphate (PLP) and [Fe-S] clusters and works in the interconversion of L-lysine and L-β-lysine. In this case, LAM employs SAM as a true coenzyme to facilitate catalysis (96).

Figure 10. Reactions catalyzed by studied radical SAM enzymes (96). In the reaction of HemN (a), SAM is a substrate consumed as an oxidant in the oxidative decarboxylation of propionate side chains of coproporphyrinogen to vinyl groups. In reactions of BioB (b), SAM is a true substrate that is consumed as an oxidant in the insertion of sulfur into unreactive C–H bonds. The exact roles of SAM in the reactions of MoaA (c) are not clarified. In reactions of LAM (D), SAM functions catalytically as a temporary oxidant and is regenerated at the end of each catalytic cycle. Note the diversity of reaction types: isomerizations (D), insertion of sulfur atoms into inactivated C–H bonds (B), oxidative decarboxylation (A), rearrangement and fusion of heterocyclic rings (C).

1.5.2. A functional example of the auxiliary [Fe-S] clusters in radical SAM enzymes: LipA

A subset of radical SAM enzymes contains one or more [Fe-S] clusters in addition to the SAM-[Fe₄S₄] cluster. In many cases, these auxiliary clusters have proven essential for enzyme activity, as they are required as a source of sulfur for the reaction, to bind or to activate the substrate for catalysis, or for electron transfer.

One of the best-known systems for functionality of the auxiliary [Fe-S] clusters in the Radical SAM family is the sulfur donor protein LipA. As BioB, LipA is involved in the insertion of one or two S atoms into an organic substrate. In the reaction of BioB, the insertion of a sulfur atom between the C6 and C9 of dethiobiotin requires the

presence of a $[\text{Fe}_2\text{S}_2]$ cluster in addition to the SAM- $[\text{Fe}_4\text{S}_4]$ cluster. The $[\text{Fe}_2\text{S}_2]$ cluster has a sacrificial role exhibiting a unique ligation environment comprised of three cysteines and one arginine (97, 98). Instead of the $[\text{Fe}_2\text{S}_2]$ cluster found in BioB, LipA contains an auxiliary $[\text{Fe}_4\text{S}_4]$ cluster (99). Using isotopically labeled substrate it was demonstrated that LipA is able to cleave two molecules of SAM for each molecule of product produced, catalyzing the insertion of two sulfur atoms into an octanoyl chain attached to a lipoyl carrier protein (Figure 11). The first hydrogen atom ($\text{H}^\bullet$) abstraction from the C6 of the octanoyl chain, generates a transient carbon-centered substrate radical (100, 101) that afterwards induces a covalently cross-linked intermediate between the octanoyl chain of the LCP and the $[\text{Fe}_4\text{S}_4]$ cluster of LipA, thanks to the bridging μ -sulfido ion of the auxiliary $[\text{Fe}_4\text{S}_4]$ cluster. The second $\text{H}^\bullet$ abstraction occurs from C8 of the octanoyl substrate, and the resulting C8 radical attacks a second bridging μ -sulfido ion of the auxiliary cluster. At the end the synthesis of the lipoyl cofactor is completed upon the addition of two protons.

Figure 11. LipA mechanism. The first 5'-dA• generated abstracts a hydrogen atom ($\text{H}^\bullet$) from C6 of the octanoyl chain, resulting in a transient carbon-centered substrate radical. The second 5'-dA• abstracts a $\text{H}^\bullet$ from C8 of the octanoyl substrate, and the resulting C8 radical attacks a second bridging μ -sulfido ion of the auxiliary cluster (89).

1.5.3. Timeline of NifB and NifB-co discoveries

In early 70s, Nagatani and colleagues (102) demonstrated that *nifB* was a fundamental gene for the biosynthesis of FeMo-co (Figure 12). Shah and colleagues showed in 1994 that the biosynthesis of the iron-molybdenum cofactor could be satisfied by the addition of the low molecular weight product of NifB, named NifB-co (84). Allen and Ludden determined that NifB-co provided most if not all Fe content of FeMo-co (103). NifB-co, was also hypothesized to be the precursor for the synthesis of FeV-co and FeFe-co, suggesting the role of core portion, common among all three nitrogenase active site cofactors (104). Subsequent bioinformatics methods classified *Azotobacter vinelandii* NifB part of the Radical SAM superfamily (86). Soon after, Curatti purified NifB from *A. vinelandii* and confirmed empirically the SAM dependency by *in vitro* studies demonstrating the importance of the radical chemistry activity of this protein (85).

The first crystal structure of nitrogenase (34) revealed that FeMo-co resembled two [Fe₄S₄] cubes linked by an additional sulfur atom and with one Fe atom replaced by Mo. In 2002, near-atomic resolution crystallographic data demonstrated that the central atom that coordinates the trigonal core prismatic arrangement of the six central iron atoms, was a light element atom, potentially C, N or O (37, 104). Spectroscopic studies then showed that NifB-co comprised at least the [Fe₆S₉X] core of FeMo-co (63).

Among the candidates for the FeMo-co central atom, it was later demonstrated that C fits best the XES data (105). It was then found that the light atom contained inside the Fe-S cage of NifB-co was carbide, product of the methyl group of SAM activated by NifB radical activity (106). These discoveries added a new layer of complexity to the FeMo-co biosynthesis, highlighting that in the NifB reaction, two different processes are generated: *i*) the 5-deoxyadenosine radical (5-dA•) and *ii*) the carbon donation. Functional analyses with deuterium substitution and ¹⁴C, remarked that the methyl group (-CH₃) of one equivalent of SAM was used as a source of the C atom and a second equivalent of SAM generated a transient -CH₂ • radical from the 5-dA• activity of the protein.

Figure 12. Schematics showing of the timeline of discoveries about NifB

The complete biosynthesis process was not at all demonstrated but two principal hypotheses were formulated about carbon insertion in NifB-co. The first mechanism considers the transfer of a methyl group from SAM to a sulfide of an [Fe-S] cluster present in NifB, generating S-adenosyl homocysteine (SAH) and a cluster ligated to methanethiol. The reductive cleavage of a second SAM molecule generates the 5-dA• radical that abstracts a hydrogen atom from the methyl group generating 5'-dAH and a methylenethiol. The second mechanism would be similar to the radical activity involved in Vitamin B12 synthesis (107, 108). It involves direct formation of Fe-C, starting from the reductive cleavage of the methyl group from SAM on an Fe atom of one of the [Fe-S] cluster precursors of NifB. Despite these advances towards elucidating structure-function properties of NifB, the proposed mechanisms for carbide insertion lack fundamental information related to the potential reductant requirements for the initial methyl transfer and the processing of the methyl group into carbide.

Discoveries afterwards overlap with the work of this thesis and will be presented in the Results and Discussion sections.

Objectives

This work aims to deepen our understanding of NifB. This enzyme catalyzes the first committed step in the biosynthesis of the active-site cofactors of all known nitrogenases and is required for all biological nitrogen fixation activity on Earth. As such, NifB holds as much agronomical importance as nitrogenase itself. We aim to understand the [Fe-S] cluster rearrangements that lead to the formation of NifB-co, a unique [Fe₈S₉C] cluster of importance in bioinorganic chemistry. We also aim to find simpler and more robust NifB proteins that may help engineering nitrogenase biosynthesis in non-nitrogen fixing host such as yeasts and plants.

The specific objectives of this thesis are:

- To establish a better model for the study of NifB and the reaction that catalyzes.
- To determine *in vivo* and *in vitro* functionality of *M. infernus* NifB.
- To investigate the essentiality of conserved amino acid residues, motifs and protein domains in the synthesis of NifB-co.
- To determine the nature and number of [Fe-S] clusters in *M. infernus* NifB as base to understand the process of NifB-co formation.
- To crystallize NifB and determine its three-dimensional molecular structure.

Materials and Methods

3.1. Reagents

All chemicals were purchased from Sigma-Aldrich, VWR, New England Biolabs, Waters, Merck or Invitrogen unless specifically stated. All reagents purchased were of appropriate quality for their use.

3.2. Bacterial strains and growth conditions

A. vinelandii DJ (wild-type) (109), UW140 ($\Delta nifB$) (85), and UW418 (*PnifH::his-nifB_{Mi}*) were used in this study. Strain UW140 was used in complementation experiments to demonstrate the functionality of NifB_{Mi} *in vivo*. Strain UW418 is a UW140 derivative in which *nifB_{Mi}* expression is controlled by an extra copy of the *nifH* promoter. The genotypes of UW140 and UW418 were confirmed by PCR analysis.

For diazotrophic growth rate determinations, *A. vinelandii* strains were grown at 30°C on N-free Burk's medium. When a fixed nitrogen source was required, ammonium acetate was added to a final concentration of 29 mM. Antibiotics were added at standard concentrations: Km, 0.5 µg/ml; Sp, 10 µg/ml; Ap 50 µg/ml; and Rif 5 µg/ml. Growth was estimated as OD_{600nm} using an Ultrospec 3300 Pro spectrophotometer (Amersham). The exponential growth rate constant corresponds to $\ln 2/td$, where *td* represents the doubling time.

For *A. vinelandii* competent cell preparation and strain transformation, cells were grown on solid Burk's modified medium lacking molybdenum for 2-3 days at 30°C. Afterwards, 3-4 loops of cells were resuspended in 1 ml of Burk's-modified liquid medium lacking molybdenum and iron, vortexed vigorously and resuspended in 10 ml of Burk's-modified liquid medium lacking molybdenum and iron. Cells were incubated overnight in a 50 ml Falcon tube at 30°C and 180 rpm shaking until an OD₅₈₀ of 1.0 was reached. In general, *A. vinelandii* cells cultured under these conditions, and at this optical density, exhibit a bright greenish color due to the excretion of the iron siderophore azotobactin. Two hundred µl of this cell suspension were incubated with 200 µl of transformation buffer (20 mM MOPS pH 7.4, 16 mM MgSO₄) and 5-10 µl of plasmid DNA in a sterile 1.5 ml centrifuge tube in a shaker at 180 rpm and 30°C for 45 min. The transformation mixtures were transferred to 12.5 ml of Burk's fresh media in 100-ml flasks and incubated overnight at 30°C and 200

rpm. In order to select the recombinant strains, 1000 μ l, 100 μ l, and 10 μ l aliquots of the 12.5 ml culture were withdrawn and plated on solid Burk's medium containing the appropriate antibiotics. Positive colonies were streak again on fresh plates and then on liquid Burk's medium.

E. coli DH5 α (F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nup* G Φ 80*dlacZ* Δ M15 Δ (*lacZYA-argF*)U169, *hsdR*17($r_K^- m_K^+$), λ^-) was used for plasmid cloning. *E. coli* BL21 (DE3) (F⁻ *ompT gal dcm lon hsdSB*($r_B^- m_B^-$) λ (DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])) was used to overexpress proteins encoded by genes cloned under the control of a T7 promoter. The LacI repressor inhibits transcription of the gene of interest and the addition of IPTG (isopropyl β -D-1-thiogalactopyranoside) relieves this inhibition when added in the culture medium.

For cloning experiments, *E. coli* strains were routinely cultivated in Luria-Bertani medium (10 g tryptone, 5 g yeast extract, 10 g NaCl in one liter of distilled water and pH adjusted to 7.5 with sodium hydroxide) at 37°C with shaking (200 rpm). Antibiotics were added at standard concentrations (110). Solid media was formed by adding 15 g agar per liter of LB and pouring approximately 25 ml into sterile Petri dishes. Petri dishes were stored at 4°C until used.

For overexpression of *MiNifB* WT and the site-directed variants, *E. coli* BL21 (DE3) carrying pRSFIscMetKDuet-1 was transformed with plasmids pRHB509, pRHB705, pRHB706 or pRHB707. To determine the best expression time, a freshly transformed single colony from agar plate was used to inoculate 5 ml of TB broth with appropriate selective antibiotic in a 25 ml tube. The culture was grown at OD₆₀₀ of 10 and induced for respectively 1, 2 and 3 hours with the corresponding amount of IPTG inducer. Aliquots of 250 μ l of culture were collected by centrifugation (13,000 rpm, 2 min) at different points during time course point experiments. Samples were resuspended in 250 μ l SDS-sample buffer, boiled for 5 min, and 15 to 25 μ l loaded onto SDS-PAGE.

In large-scale expression experiments, recombinant *E. coli* cells were routinely grown in 50-l batches of Terrific Broth (12 g/l tryptone, 24 g/l yeast extract, 4 g/l glycerol adjusted to pH 7.1 with 17 mM KH₂PO₄ and 72 mM K₂HPO₄). The fermentation processes were initially performed under aerobic conditions (sparging 50 l of air per min and shaking at 200 rpm) for around 17 h until an OD₆₀₀ value of 10 was reached. Cultures were then sparged with N₂ and induced for NifB expression for an additional

3 h by adding 3 g/l lactose, 0.2 mM ammonium iron citrate, 2 mM cysteine, and 25 mM di-sodium fumarate. After the fermentation processes, cells (ca. 500 g) were collected under anaerobic conditions in a GE Hollow Fiber system and sedimented by centrifugation. Solid cell paste was immediately frozen and stored under liquid N₂.

3.3. Plasmids

For NifB_{Mi} purifications, *E. coli* codon-optimized sequences of wild-type *M. infernus* *nifB* (*metin_0554* gene; <http://www.ncbi.nlm.nih.gov/nuccore/CP002009>), and site-directed variants substituting Ala for Cys in the N-terminal radical-SAM motif (AX₃AX₂A; called SM), a C-terminal Cys-rich motif (AX₂A, called AM), and both motifs (AX₃AX₂A/AX₂A, called DM) were ordered from GenScript (Table 1) and cloned into the *Nde*I and *Xho*I sites of pET16b expression vector to generate plasmids pRHB509, pRHB707, pRHB705, and pRHB706, respectively (Table 2). Plasmid pRSFiscMetKDuet-1 (kindly donated by Juan Fontecilla) is a pRSFDuet-1 (Novagen) derivative generated by inserting *E. coli* K12-MG1655 *isc* gene cluster (lacking the *iscR* gene) into its *Nco*I and *Not*I sites, and the SAM synthetase gene *metK* gene into its *Nde*I and *Xho*I sites.

For *A. vinelandii* Δ *nifB* genetic complementation experiments, the synthetic wild-type *nifB*_{Mi} gene was cloned into the *Nde*I and *Eco*RI sites of pRHB258 to generate plasmid pRHB558, which was used for the expression of *his₉-nifB*_{Mi} under the control of the *nifH* promoter in *A. vinelandii* (69). Plasmid pRHB558 was inserted into the chromosome of *A. vinelandii* UW140 (Δ *nifB*) by homologous recombination at the D-sequence, a 1.1-kb DNA fragment from the chromosomal region downstream of Avin02530 (59), to generate strain UW418. Transformants were selected in agar plates of NH₄⁺-free Burk's modified medium (111) containing 50 μ g/ml ampicillin.

NifB _{Mi} version	DNA sequence ^(a)	Amino primary sequence ^(b)
Wild type	CATATGGAAAAAATGTCAAAATTCTCGCATCTGCTGAAAGCGCACCCGTGCTTCA ACGAAAAAGTCCACGATAAAATACGGTCGCGTCCACCTGCCGGTTGCACCGCGTTG CAACATCGCTTGCAAAATTTGTAAACGCGAGCGTCTCTAAAGAATGCTGTGAACAT CGTCCGGGCGTCTCACTGGGTGTGCTGAAACCGGAAGATGTGGAAGACTACCTGA AGAAAAATCTGAAAGAAATGCCGAACATCAAAGTGGTTGGCATTGCGGGCCCGG GTGATTCTCGTGTTTAATAAAGAAACGTTGCAAAACCTGAAAAATTATCGACGAAAA ATTCCCGAACCTGATTAATGTATCAGTACGAATGGTCTGCTGCTGTCCAAATATT ACAAAGATCTGGCAAAACCTGAATGTTGCGACCATCACGGTTACCGTCAACGCTAT TAAACCGGAAATCTGGAAGAAATTTGGATTGGGTTTATTACGACAAAAAACTG TACCGTGGCCTGGAAGGTGCGAAACTGCTGATCGAAAAACAGATCGAAGGCATC AAAAAAGCCTCTGAAGAAGACTTCATCATCAAAATCAACACGGTCTGATTCCCG AAATCAATATGGATCACGTCGTGGAATTTGCGAAATTTTCAAAGACTATGCCTA CGTTCAGAAATATTATCCCGCTGATTCCGCAATATAAAATGAAAGAACTGCGCGCA CCGACCTGCGAAGAAATCAAAAAAGTGCCTAAAGAATGTGAAAAATACATTCCG CAGTTTCGCGCCTGTGGTCAATGTGCTGCGGATGCCGTTGGTCTGATTAAAGAAA AAGAACTGCTGAAAGAATTTTCAAAGAAAAAGAATAAAGAAAAAGAACATTAAC TGGAAGTCTTTGACCTGAAACACTTCTCCCACTGACTCGAG	MEKMSKFSHLLKAHPCFN EKVHDKYGRVHLPVAPRC NIACKFCRKSVSKECEHR PGVSLGVLPEDVEDYLYK KILKEMPNIKVVGIAGPGD SLFNKETFETLKIIDEKFPN LIKICISTNGLLLSKYYKDLE NLNVRTITVTVNAIKPEILE KIVDWVYYDKKLYRGLEG AKLLIEKQIEGIKKASEEDFI IKINTVLIPEINMDHVVEIA KFFKDYAYVQNIPLIPQYK MKELRAPTCIEIKKVRKEC EKYIPQFRACGQCRADAV GLIKEKELLKEFFKEKNKE KNIKLEVFDLKHFSH
SAM mutant (SM)	CATATGGAAAAAATGTCAAAATTCTCGCATCTGCTGAAAGCGCACCCGTGCTTCA ACGAAAAAGTCCACGATAAAATACGGTCGCGTCCACCTGCCGGTTGCACCGCGTGC CAACATCGCTGCCAAATTTGTCTAAACGCGAGCGTCTCTAAAGAATGCTGTGAACAT CGTCCGGGCGTCTCACTGGGTGTGCTGAAACCGGAAGATGTGGAAGACTACCTGA AGAAAAATCTGAAAGAAATGCCGAACATCAAAGTGGTTGGCATTGCGGGCCCGG GTGATTCTCGTGTTTAATAAAGAAACGTTGCAAAACCTGAAAAATTATCGACGAAAA ATTCCCGAACCTGATTAATGTATCAGTACGAATGGTCTGCTGCTGTCCAAATATT ACAAAGATCTGGCAAAACCTGAATGTTGCGACCATCACGGTTACCGTCAACGCTAT TAAACCGGAAATCTGGAAGAAATTTGGATTGGGTTTATTACGACAAAAAACTG TACCGTGGCCTGGAAGGTGCGAAACTGCTGATCGAAAAACAGATCGAAGGCATC AAAAAAGCCTCTGAAGAAGACTTCATCATCAAAATCAACACGGTCTGATTCCCG AAATCAATATGGATCACGTCGTGGAATTTGCGAAATTTTCAAAGACTATGCCTA CGTTCAGAAATATTATCCCGCTGATTCCGCAATATAAAATGAAAGAACTGCGCGCA CCGACCTGCGAAGAAATCAAAAAAGTGCCTAAAGAATGTGAAAAATACATTCCG CAGTTTCGCGCCTGTGGTCAATGTGCTGCGGATGCCGTTGGTCTGATTAAAGAAA AAGAACTGCTGAAAGAATTTTCAAAGAAAAAGAATAAAGAAAAAGAACATTAAC TGGAAGTCTTTGACCTGAAACACTTCTCCCACTGACTCGAG	MEKMSKFSHLLKAHPCFN EKVHDKYGRVHLPVAPRA NIAAKFAKRSVSKECEHR PGVSLGVLPEDVEDYLYK KILKEMPNIKVVGIAGPGD SLFNKETFETLKIIDEKFPN LIKICISTNGLLLSKYYKDLE NLNVRTITVTVNAIKPEILE KIVDWVYYDKKLYRGLEG AKLLIEKQIEGIKKASEEDFI IKINTVLIPEINMDHVVEIA KFFKDYAYVQNIPLIPQYK MKELRAPTCIEIKKVRKEC EKYIPQFRACGQCRADAV GLIKEKELLKEFFKEKNKE KNIKLEVFDLKHFSH
Auxiliar y motif mutant (AM)	CATATGGAAAAAATGTCAAAATTCTCGCATCTGCTGAAAGCGCACCCGTGCTTCA ACGAAAAAGTCCACGATAAAATACGGTCGCGTCCACCTGCCGGTTGCACCGCGTTG CAACATCGCTTGCAAAATTTGTAAACGCGAGCGTCTCTAAAGAATGCTGTGAACAT CGTCCGGGCGTCTCACTGGGTGTGCTGAAACCGGAAGATGTGGAAGACTACCTGA AGAAAAATCTGAAAGAAATGCCGAACATCAAAGTGGTTGGCATTGCGGGCCCGG GTGATTCTCGTGTTTAATAAAGAAACGTTGCAAAACCTGAAAAATTATCGACGAAAA ATTCCCGAACCTGATTAATGTATCAGTACGAATGGTCTGCTGCTGTCCAAATATT ACAAAGATCTGGCAAAACCTGAATGTTGCGACCATCACGGTTACCGTCAACGCTAT TAAACCGGAAATCTGGAAGAAATTTGGATTGGGTTTATTACGACAAAAAACTG TACCGTGGCCTGGAAGGTGCGAAACTGCTGATCGAAAAACAGATCGAAGGCATC AAAAAAGCCTCTGAAGAAGACTTCATCATCAAAATCAACACGGTCTGATTCCCG AAATCAATATGGATCACGTCGTGGAATTTGCGAAATTTTCAAAGACTATGCCTA CGTTCAGAAATATTATCCCGCTGATTCCGCAATATAAAATGAAAGAACTGCGCGCA CCGACCTGCGAAGAAATCAAAAAAGTGCCTAAAGAATGTGAAAAATACATTCCG CAGTTTCGCGCCTGTGGTCAATGTGCTGCGGATGCCGTTGGTCTGATTAAAGAAA AAGAACTGCTGAAAGAATTTTCAAAGAAAAAGAATAAAGAAAAAGAACATTAAC TGGAAGTCTTTGACCTGAAACACTTCTCCCACTGACTCGAG	MEKMSKFSHLLKAHPCFN EKVHDKYGRVHLPVAPRC NIACKFCRKSVSKECEHR PGVSLGVLPEDVEDYLYK KILKEMPNIKVVGIAGPGD SLFNKETFETLKIIDEKFPN LIKICISTNGLLLSKYYKDLE NLNVRTITVTVNAIKPEILE KIVDWVYYDKKLYRGLEG AKLLIEKQIEGIKKASEEDFI IKINTVLIPEINMDHVVEIA KFFKDYAYVQNIPLIPQYK MKELRAPTCIEIKKVRKEC EKYIPQFRACGQCRADAV GLIKEKELLKEFFKEKNKE KNIKLEVFDLKHFSH
Double mutant (DM)	CATATGGAAAAAATGTCAAAATTCTCGCATCTGCTGAAAGCGCACCCGTGCTTCA ACGAAAAAGTCCACGATAAAATACGGTCGCGTCCACCTGCCGGTTGCACCGCGTGC GAACATCGCTGCGAAATTTGCGAAACGCGAGCGTCTCTAAAGAATGCTGTGAACAT CGTCCGGGCGTCTCACTGGGTGTGCTGAAACCGGAAGATGTGGAAGACTACCTGA AGAAAAATCTGAAAGAAATGCCGAACATCAAAGTGGTTGGCATTGCGGGCCCGG GTGATTCTCGTGTTTAATAAAGAAACGTTGCAAAACCTGAAAAATTATCGACGAAAA ATTCCCGAACCTGATTAATGTATCAGTACGAATGGTCTGCTGCTGTCCAAATATT ACAAAGATCTGGCAAAACCTGAATGTTGCGACCATCACGGTTACCGTCAACGCTAT TAAACCGGAAATCTGGAAGAAATTTGGATTGGGTTTATTACGACAAAAAACTG TACCGTGGCCTGGAAGGTGCGAAACTGCTGATCGAAAAACAGATCGAAGGCATC AAAAAAGCCTCTGAAGAAGACTTCATCATCAAAATCAACACGGTCTGATTCCCG AAATCAATATGGATCACGTCGTGGAATTTGCGAAATTTTCAAAGACTATGCCTA CGTTCAGAAATATTATCCCGCTGATTCCGCAATATAAAATGAAAGAACTGCGCGCA CCGACCTGCGAAGAAATCAAAAAAGTGCCTAAAGAATGTGAAAAATACATTCCG CAGTTTCGCGCCTGTGGTCAATGTGCTGCGGATGCCGTTGGTCTGATTAAAGAAA AAGAACTGCTGAAAGAATTTTCAAAGAAAAAGAATAAAGAAAAAGAACATTAAC TGGAAGTCTTTGACCTGAAACACTTCTCCCACTGACTCGAG	MEKMSKFSHLLKAHPCFN EKVHDKYGRVHLPVAPRA NIAAKFAKRSVSKECEHR PGVSLGVLPEDVEDYLYK KILKEMPNIKVVGIAGPGD SLFNKETFETLKIIDEKFPN LIKICISTNGLLLSKYYKDLE NLNVRTITVTVNAIKPEILE KIVDWVYYDKKLYRGLEG AKLLIEKQIEGIKKASEEDFI IKINTVLIPEINMDHVVEIA KFFKDYAYVQNIPLIPQYK MKELRAPTCIEIKKVRKEC EKYIPQFRACGQCRADAV GLIKEKELLKEFFKEKNKE KNIKLEVFDLKHFSH

Table 1. Gene (a) and amino acid sequences (b) of synthetic NifB_{Mi} proteins.

Plasmid	Relevant genes cloned	Vector	Reference or source
pRHB509	NifB _{Mt} WT	pET16	This study
pRHB707	NifB _{Mt} SM	pET16	This study
pRHB705	NifB _{Mt} AM	pET16	This study
pRHB706	NifB _{Mt} DM	pET16	This study
pRSFIscMet KDuet-1	IscS, IscU, IscA, HscB, HscA, fdx, IscX, and MetK	pRSFDuet-1	Juan Fontecilla

Table 2. List of plasmids used for the overexpression of the different NifB_{Mt} variants in *E. coli*.

3.4. Data mining and phylogenetic analysis

The 390 annotated NifB sequences retrieved from the Structure and Function Linkage Database (SFLD) (111) and UniProt (<http://uniprot.org>) are shown in Table 3. To exclude potentially faulty annotated sequences, the following filtering procedure was applied to the dataset. First, amino acid sequences of experimentally proven NifB proteins, including *A. vinelandii* (NifB_{Av}) (85), *K. oxytoca* (NifB_{Ko}) (112), *Clostridium pasteurianum* (NifB_{Cp}) (113, 114), *M. infernus* (NifB_{Mi}) (115), *Methanosarcina acetivorans* (NifB_{Ma}) (116), *Methanobacterium thermoautotrophicum* (NifB_{Mt}) (116), and *C. tepidum* (117) were aligned to determine conserved motifs. These NifB fingerprint motifs localized in the SAM-radical domain and included an HPC motif, the AdoMet Cx₃Cx₂C motif, an ExRP motif, an AGPG motif, a TxTxN motif, and a Cx₂CRxDaxG motif (Figure 19). The full NifB dataset was then analyzed for the presence of these fingerprints, reducing the initial 390 sequences to 289 (Table 3). Protein domain architecture was analyzed using the PFAM database (<http://pfam.xfam.org>) (118). The frequency of appearance of each one of the different NifB domains in diazotrophic phyla shown in Figure 19 was represented by overlapping data from Table 3 with a 3-domain taxonomic tree of life.

SFLD id	Species	Phylum	Class	Uniprot acc num	Length (AA)	Discarded
280102	<i>Methanocaldococcus jannaschii</i> DSM 2661	Euryarchaeota	Methanococci	Q58493	300	
381033	<i>Bradyrhizobium japonicum</i> USDA 110	Proteobacteria	Alphaproteobacteria	P06770	499	
381034					521	
381035	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	Proteobacteria	Alphaproteobacteria	P24427	490	
381036	<i>Cloning vector</i> SBA_000534, <i>Klebsiella pneumoniae</i>	Proteobacteria	Gammaproteobacteria	P10390	468	
381037	<i>Rhizobium leguminosarum</i>	Proteobacteria	Alphaproteobacteria	P07748	359	
381038	<i>Rhizobium</i> sp. <i>NGR234</i>	Proteobacteria	Alphaproteobacteria	Q53205	493	
381039	<i>Rhodobacter capsulatus</i>	Proteobacteria	Alphaproteobacteria	P17434	452	
381040	<i>Azotobacter vinelandii</i>	Proteobacteria	Gammaproteobacteria	P11067	502	
381041	<i>Anabaena variabilis</i>	Cyanobacteria		Q43883	475	
381043	<i>Herbaspirillum seropedicae</i>	Proteobacteria	Betaproteobacteria	P27714	525	
381044	<i>Sinorhizobium meliloti</i> 1021	Proteobacteria	Alphaproteobacteria	P09824	490	
381045	<i>Nostoc</i> sp. <i>PCC 7120</i>	Cyanobacteria		P20627	475	
381046	<i>Frankia alni</i>	Actinobacteria	Actinobacteria	P46044	510	
385641	<i>Rhodomicrobium vannielii</i> ATCC 17100	Proteobacteria	Alphaproteobacteria	E313H6	519	
385653	<i>Gloeotheca</i> sp. <i>KO68DGA</i>	Cyanobacteria		A2V895	490	
385663	<i>Novosphingobium nitrogenifigens</i> DSM 19370	Proteobacteria	Alphaproteobacteria	F1Z4T3	522	
385707	<i>Bradyrhizobium</i> sp. <i>ORS278</i>	Proteobacteria	Alphaproteobacteria	A4YYX1	520	
385780	<i>Klebsiella oxytoca</i> KCTC 1686	Proteobacteria	Gammaproteobacteria	G8W9F3	468	
385807	<i>Hydrogenobaculum</i> sp. <i>Y04AAS1</i>	Aquificae	Aquificae	B4U920	469	
385893	<i>Azospirillum amazonense</i> Y2	Proteobacteria	Alphaproteobacteria	G1XWD1	519	
386002	<i>Nostoc punctiforme</i> PCC 73102	Cyanobacteria		B2J6Z1	483	
386061	<i>Trichodesmium erythraeum</i> IMS101	Cyanobacteria		Q10X83	489	
386075	<i>Thiorhodospira sibirica</i> ATCC 700588	Proteobacteria	Gammaproteobacteria	G4E8L1	518	
412273	<i>Desulfurobacterium thermolithotrophum</i> DSM 11699	Aquificae	Aquificae	F0S0T8	322	
412274	<i>Klebsiella variicola</i> At-22	Proteobacteria	Gammaproteobacteria	D1KS27	468	
412275	<i>Clostridium papyrosolvens</i> DSM 2782	Firmicutes	Clostridia	F1TH64	884	
412276	<i>Mesorhizobium loti</i> R7A	Proteobacteria	Alphaproteobacteria	Q8KJ58	493	
412277	<i>Rhodopseudomonas palustris</i> DX-1	Proteobacteria	Alphaproteobacteria	E6VEV0	518	
412278	<i>Azoarcus</i> sp. <i>BH72</i>	Proteobacteria	Betaproteobacteria	A1K2T5	498	
412279	<i>Methanoterris igneus</i> Kol 5	Euryarchaeota	Methanococci	F6BAP3	294	
412280	<i>Klebsiella</i> sp. <i>L_1_55</i>	Proteobacteria	Gammaproteobacteria	D6GAU2	468	
412281	<i>Methanococcus maripaludis</i> C7	Euryarchaeota	Methanococci	A6VJW4	297	
412282	<i>Candidatus Azobacteroides pseudotrichonymphae</i> genomovar. <i>CFP2</i>	Bacteroidetes	Bacteroidia	B6YRJ3	419	
412283	<i>Cyanothece</i> sp. <i>PCC 8801</i>	Cyanobacteria		B7JWX7	491	
412284	<i>Chlorobium phaeobacteroides</i> BSI	Chlorobi	Chlorobia	B3EL88	424	
412285	<i>Pelodictyon phaeoclathratiforme</i> BU-1	Chlorobi	Chlorobia	B4SC66	423	
412286	<i>Thermoanaerobacterium thermosaccharolyticum</i> DSM 571	Firmicutes	Clostridia	D9TS57	349	
412287	<i>Chlorobaculum parvum</i> NCIB 8327	Chlorobi	Chlorobia	B3QQ19	424	
412288	<i>Anaeromyxobacter</i> sp. <i>Fw109-5</i>	Proteobacteria	Deltaproteobacteria	A7HES5	495	
412289	<i>Desulfovibrio vulgaris</i> str. 'Miyazaki F'	Proteobacteria	Deltaproteobacteria	B8DR83	427	x
412290	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304	Proteobacteria	Alphaproteobacteria	B6A0Y0	363	
412291	<i>Bradyrhizobium</i> sp. <i>BTAi1</i>	Proteobacteria	Alphaproteobacteria	A5ENV2	520	
412292	<i>Treponema azotonutricium</i> ZAS-9	Spirochaetes	Spirochaetia	F5YEM2	306	
412293	<i>Clostridium thermocellum</i> DSM 1313	Firmicutes	Clostridia	C7HF06	291	x
412294	<i>Phascolarctobacterium</i> sp. <i>YIT 12068</i>	Firmicutes	Negativicutes	E8LBQ5	222	x
412295	<i>Campylobacter showae</i> RM3277	Proteobacteria	Epsilonproteobacteria	C6RGS7	275	
412296	<i>Thiorhodococcus drewsii</i> AZ1	Proteobacteria	Gammaproteobacteria	G2E6G2	497	
412297	<i>Methanobrevibacter ruminantium</i> M1	Euryarchaeota	Methanobacteria	D3DZ69	296	
412298	<i>Megasphaera micronuciformis</i> F0359	Firmicutes	Negativicutes	E2ZAI0	232	x
412299	<i>Candidatus Accumulibacter phosphatis</i> clade <i>IIA</i> str. <i>UW-1</i>	Proteobacteria	Betaproteobacteria	C7RRE1	505	
412300	<i>Clostridium ljungdahlii</i> DSM 13528	Firmicutes	Clostridia	D8GL53	423	
412301	<i>Acidithiobacillus ferrooxidans</i> ATCC 53993	Proteobacteria	Acidithiobacillia	B5ER81	522	
412302	<i>Methanocaldococcus vulcanius</i> M7	Euryarchaeota	Methanococci	C9RI91	300	
412303	<i>Desulfovibrio magneticus</i> RS-1	Proteobacteria	Deltaproteobacteria	C4XQ76	422	x
412304	<i>Caldicellulosiruptor kristjanssonii</i> 177R1B	Firmicutes	Clostridia	E4SAV8	266	x
412305	<i>Caldicellulosiruptor hydrothermalis</i> 108	Firmicutes	Clostridia	E4Q7K0	266	x
412306	<i>Caldicellulosiruptor lactoaceticus</i> 6A	Firmicutes	Clostridia	G2PWQ2	266	x
412307	<i>Ethanoligenens harbinense</i> YUAN-3	Firmicutes	Clostridia	E6U4U5	419	x
412308	<i>Clostridium cellulolyticum</i> H10	Firmicutes	Clostridia	B8I2G9	269	x
412309	<i>Marichromatium purpuratum</i> 984	Proteobacteria	Gammaproteobacteria	F9TXW1	497	
412310	<i>Gluconacetobacter diazotrophicus</i> PAI 5	Proteobacteria	Alphaproteobacteria	A9H5T3	490	
412311	<i>Rhizobium etli</i> CIAT 652	Proteobacteria	Alphaproteobacteria	B3Q2S5	492	
412312	<i>Campylobacter concisus</i> 13826	Proteobacteria	Epsilonproteobacteria	A7ZEI0	275	

412313	<i>Arcobacter nitrofigilis</i> DSM 7299	Proteobacteria	Epsilonproteobacteria	D5V3N9	482	
412314	<i>Pantoea</i> sp. At-9b	Proteobacteria	Gammaproteobacteria	D1KS47	467	
412315	<i>Thiocapsa marina</i> 5811	Proteobacteria	Gammaproteobacteria	F9UCV2	503	
412316	<i>Methanobacterium</i> sp. AL-21	Euryarchaeota	Methanobacteria	F0TAR5	283	
412317	<i>Pectobacterium atrosepticum</i> SCRI1043	Proteobacteria	Gammaproteobacteria	Q6D307	468	
412318	<i>Methanococcus maripaludis</i> C6	Euryarchaeota	Methanococci	A9A7C6	297	
412319	<i>Sulfuricurvum kujense</i> DSM 16994	Proteobacteria	Epsilonproteobacteria	E4TZR5	471	
412320	<i>Rhodobacter sphaeroides</i> ATCC 17029	Proteobacteria	Alphaproteobacteria	A3PLT4	491	
412321	<i>Geobacter bemidjensis</i> Bem	Proteobacteria	Deltaproteobacteria	B5ECR1	285	
412322	<i>Spirochaeta smaragdinae</i> DSM 11293	Spirochaetes	Spirochaetia	E1R3Y2	421	x
412323	<i>Roseiflexus</i> sp. RS-1	Chloroflexi	Chloroflexia	A5USK4	313	
412324	<i>Rhodopseudomonas palustris</i> TIE-1	Proteobacteria	Alphaproteobacteria	B3QIR1	296	x
412325	<i>Dickeya dadantii</i> Ech703	Proteobacteria	Gammaproteobacteria	C6C8V3	468	
412326	<i>Syntrophobotulus glycolicus</i> DSM 8271	Firmicutes	Clostridia	F0SYG8	296	
412327	<i>Acidithiobacillus ferrivorans</i> SS3	Proteobacteria	Acidithiobacillia	G0JSZ3	522	
412328	<i>Desulfobacca acetoxidans</i> DSM 11109	Proteobacteria	Deltaproteobacteria	F2NEQ9	422	x
412329	<i>Desulfovibrio vulgaris</i> str. 'Miyazaki F'	Proteobacteria	Deltaproteobacteria	B8DR79	419	x
412330	<i>Desulfovibrio</i> sp. ND132	Proteobacteria	Deltaproteobacteria	F0JD05	422	x
412331	<i>Candidatus Desulforudis audaxviator</i> MP104C	Firmicutes	Clostridia	B110Z4	285	x
412332	<i>Desulfovibrio africanus</i> str. Walvis Bay	Proteobacteria	Deltaproteobacteria	F3YUU4	377	x
412333	<i>Cyanothece</i> sp. PCC 7424	Cyanobacteria		B7KG80	483	
412334	<i>Geobacter</i> sp. FRC-32	Proteobacteria	Deltaproteobacteria	B9M029	291	
412335	<i>Pantoea agglomerans</i>	Proteobacteria	Gammaproteobacteria	Q57037	465	
412336	<i>Desulfovibrio</i> sp. A2	Proteobacteria	Deltaproteobacteria	G2HBA1	427	x
412337	<i>Clostridium acetobutylicum</i> ATCC 824	Firmicutes	Clostridia	Q97MD9	882	
412338	<i>Rhodospirillum rubrum</i> ATCC 11170	Proteobacteria	Alphaproteobacteria	Q2RVQ0	503	
412339	<i>Burkholderia vietnamiensis</i> G4	Proteobacteria	Betaproteobacteria	A4JRL6	530	
412340	<i>Rhodobacter</i> sp. SW2	Proteobacteria	Alphaproteobacteria	C8S0U0	486	
412341	<i>Sideroxydans lithotrophicus</i> ES-1	Proteobacteria	Betaproteobacteria	D5CPD5	501	
412342	<i>Desulfosporosinus</i> sp. OT	Firmicutes	Clostridia	G2FSQ1	278	
412343	<i>Methanosarcina barkeri</i> str. Fusaro	Euryarchaeota	Methanomicrobia	Q46G74	321	
412344	<i>Nodularia spumigena</i> CCY9414	Cyanobacteria		A0ZAX0	445	
412345	<i>Mesorhizobium australicum</i> WSM2073	Proteobacteria	Alphaproteobacteria	F7Y557	493	
412346	<i>Clostridium ljungdahlii</i> DSM 13528	Firmicutes	Clostridia	D8GMG8	895	
412347	<i>Geobacter</i> sp. M18	Proteobacteria	Deltaproteobacteria	E8WHN4	285	
412348	<i>Clostridium beijerinckii</i> NCIMB 8052	Firmicutes	Clostridia	A6LUZ5	896	x
412349	<i>Methylomonas methanica</i> MC09	Proteobacteria	Gammaproteobacteria	G0A0V5	497	
412350	<i>Gluconacetobacter diazotrophicus</i> PAI 5	Proteobacteria	Alphaproteobacteria	Q9F1E1	490	
412351	<i>Methanosphaera stadtmanae</i> DSM 3091	Euryarchaeota	Methanobacteria	Q2NFI5	291	
412352	<i>delta proteobacterium</i> NaphS2	Proteobacteria	Deltaproteobacteria	D8EZD5	420	x
412353	<i>Methanothermococcus okinawensis</i> IH1	Euryarchaeota	Methanococci	F8ANM0	299	
412354	<i>Rhizobium leguminosarum</i>	Proteobacteria	Alphaproteobacteria	Q52836	189	x
412355	<i>Frankia alni</i> ACN14a	Actinobacteria	Actinobacteria	Q0RAW4	470	
412356	<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ZM4	Proteobacteria	Alphaproteobacteria	Q5NLG9	489	
412357	<i>Azotobacter vinelandii</i> DJ	Proteobacteria	Gammaproteobacteria	C1DMB1	503	
412358	<i>Desulfovibrio</i> sp. A2	Proteobacteria	Deltaproteobacteria	G2HCH1	306	
412359	<i>Verrucomicrobiae bacterium</i> DG1235	Verrucomicrobia	Verrucomicrobiae	B5JMT4	438	x
412360	<i>Methanocaldococcus infernus</i> ME	Euryarchaeota	Methanococci	D5VRM1	302	
412361	<i>Rhodopseudomonas palustris</i> CGA009	Proteobacteria	Alphaproteobacteria	Q6N6I1	298	x
412362	<i>Rhodospirillum rubrum</i> ATCC 11170	Proteobacteria	Alphaproteobacteria	G2TBP3	331	x
412363	<i>Desulfotomaculum nigrificans</i> DSM 574	Firmicutes	Clostridia	F0DNT9	280	x
412364	<i>Fibrobacter succinogenes</i>	Fibrobacteres	Fibrobacteria	C9RQ04	290	x
412365	<i>Rhodobacter sphaeroides</i> 2.4.1	Proteobacteria	Alphaproteobacteria	Q3J0G6	491	
412366	<i>Sinorhizobium meliloti</i> AK83	Proteobacteria	Alphaproteobacteria	F6EB47	490	
412367	<i>Desulfarculus baarsii</i> DSM 2075	Proteobacteria	Deltaproteobacteria	E1QE22	428	x
412368	<i>Synechococcus</i> sp. PCC 7335	Cyanobacteria		B4WRV7	397	x
412369	<i>Clostridium lentocellum</i> DSM 5427	Firmicutes	Clostridia	F2JLB1	897	x
412370	<i>Tolomonas auensis</i> DSM 9187	Proteobacteria	Gammaproteobacteria	C4LAM9	476	
412371	<i>Methanosarcina mazei</i>	Euryarchaeota	Methanomicrobia	Q8PYU2	335	
412372	<i>Methanocorpusculum labreanum</i> Z	Euryarchaeota	Methanomicrobia	A2SR95	310	
412373	<i>Frankia</i> sp. CcI6	Actinobacteria	Actinobacteria	Q2J4G8	527	
412374	<i>Methanosarcina acetivorans</i> C2A	Euryarchaeota	Methanomicrobia	Q8TIF7	323	
412375	<i>Magnetospirillum gryphiswaldense</i> MSR-1	Proteobacteria	Alphaproteobacteria	A4TZU0	490	
412376	<i>Cyanothece</i> sp. ATCC 51142	Cyanobacteria		A1KYD1	490	
412377	<i>Chlorobium tepidum</i> TLS	Chlorobi	Chlorobia	Q8KC85	424	x
412378	<i>Clostridium kluyveri</i> NBRC 12016	Firmicutes	Clostridia	B9E2E8	451	
412379	<i>Halorhodospira halophila</i> SL1	Proteobacteria	Gammaproteobacteria	A1WTR3	351	
412380	<i>Syntrophobotulus glycolicus</i> DSM 8271	Firmicutes	Clostridia	F0SYJ2	299	
412381	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i>	Proteobacteria	Alphaproteobacteria	D3GDK0	497	

412382	<i>Desulfobacterium autotrophicum HRM2</i>	Proteobacteria	Deltaproteobacteria	C0QGQ4	411	x
412383	<i>Clostridium thermocellum ATCC 27405</i>	Firmicutes	Clostridia	A3DFR5	291	x
412384	<i>Mesorhizobium loti</i>	Proteobacteria	Alphaproteobacteria	Q98AT9	493	
412385	<i>Dechloromonas aromatica RCB</i>	Proteobacteria	Betaproteobacteria	Q47G28	500	
412386	<i>Candidatus Methanoregula boonei</i>	Euryarchaeota	Methanomicrobia	A7I622	297	
412387	<i>Desulfurivibrio alkaphilus AHT2</i>	Proteobacteria	Deltaproteobacteria	D6Z0A8	424	x
412388	<i>Desulfobacterium autotrophicum HRM2</i>	Proteobacteria	Deltaproteobacteria	C0QKL5	424	x
412389	<i>Desulfurispirillum indicum</i>	Chrysiogenetes	Chrysiogenetes	E6W2X1	308	
412390	<i>Chlorobium phaeobacteroides DSM 266</i>	Chlorobi	Chlorobia	A1BEG4	423	x
412391	<i>Rhodobacter sphaeroides WS8N</i>	Proteobacteria	Alphaproteobacteria	F5M4C1	491	
412392	<i>Geobacter sulfurreducens PCA</i>	Proteobacteria	Deltaproteobacteria	Q749E4	290	
412393	<i>Clostridium cellulovorans</i>	Firmicutes	Clostridia	D9SS79	885	
412394	<i>Rhodospirillum centenum</i>	Proteobacteria	Alphaproteobacteria	B6IXL5	492	
412395	<i>Desulfovibrio desulfuricans</i>	Proteobacteria	Deltaproteobacteria	B8IY89	296	x
412396	<i>Clostridium kluyveri DSM 555</i>	Firmicutes	Clostridia	ASN1T7	890	
412397	<i>Geobacter sp. M21</i>	Proteobacteria	Deltaproteobacteria	C6E992	285	
412398	<i>Sinorhizobium meliloti</i>	Proteobacteria	Alphaproteobacteria	F6BZ98	490	
412399	<i>Desulfovibrio vulgaris</i>	Proteobacteria	Deltaproteobacteria	E3IT28	514	x
412400	<i>Chloroherpeton thalassium ATCC 35110</i>	Chlorobi	Chlorobia	B3QXW4	423	x
412401	<i>Cyanobacterium UCYN-A</i>	Cyanobacteria		B7TB39	491	
412402	<i>Zymomonas mobilis subsp. pomaceae ATCC 29192</i>	Proteobacteria	Alphaproteobacteria	F8EUR1	490	
412403	<i>Chlorobium limicola DSM 245</i>	Chlorobi	Chlorobia	B3EH81	424	x
412404	<i>Methanohalobium evestigatum</i>	Euryarchaeota	Methanomicrobia	D7E943	315	
412405	<i>Pseudomonas stutzeri</i>	Proteobacteria	Gammaproteobacteria	Q93JV6	504	
412406	<i>Methylobacterium nodulans</i>	Proteobacteria	Alphaproteobacteria	B8ITH4	498	
412407	<i>Thermicola potens</i>	Firmicutes	Clostridia	D5XC16	428	
412408	<i>Geobacter daltonii</i>	Proteobacteria	Deltaproteobacteria	B9M0T8	286	
412409	<i>Magnetospirillum magneticum</i>	Proteobacteria	Alphaproteobacteria	Q2W6Y8	498	
412410	<i>Geobacter lovleyi</i>	Proteobacteria	Deltaproteobacteria	B3E297	273	
412411	<i>Anabaena variabilis ATCC 29413</i>	Cyanobacteria		Q44481	484	
412412	<i>Syntrophobacter fumaroxidans</i>	Proteobacteria	Deltaproteobacteria	A0LH03	423	x
412413	<i>Thermodesulfovibrio yellowstonii DSM 11347</i>	Nitrospirae	Nitrospira	B5YH11	293	
412414	<i>Pseudomonas stutzeri</i>	Proteobacteria	Gammaproteobacteria	A4VJ50	447	x
412415	<i>Methanocella paludicola</i>	Euryarchaeota	Methanomicrobia	D1YZZ3	298	
412416	<i>Syntrophothermus lipocalidus DSM 12680</i>	Firmicutes	Clostridia	D7CIY9	271	
412417	<i>Caldicellulosiruptor saccharolyticus DSM 8903</i>	Firmicutes	Clostridia	A4XMA2	267	x
412418	<i>Acidithiobacillus ferrooxidans ATCC 23270</i>	Proteobacteria	Acidithiobacillia	B7JAA6	499	
412419	<i>Frankia sp. EUN1f</i>	Actinobacteria	Actinobacteria	D3CZR2	672	
412420	<i>Azospirillum lipoferum</i>	Proteobacteria	Alphaproteobacteria	D3NWW6	521	
412421	<i>Rhodopseudomonas palustris</i>	Proteobacteria	Alphaproteobacteria	B3QG74	307	x
412422	<i>Desulfovibrio sp. A2</i>	Proteobacteria	Deltaproteobacteria	G2HB97	445	x
412423	<i>Methanococcoides burtonii DSM 6242</i>	Euryarchaeota	Methanomicrobia	Q12WI7	312	
412424	<i>Ruminococcus albus</i>	Firmicutes	Clostridia	E6UE53	293	
412425	<i>Lyngbya sp. PCC 8106</i>	Cyanobacteria		A0YP01	486	
412426	<i>Campylobacter gracilis</i>	Proteobacteria	Epsilonproteobacteria	C8PHT9	275	
412427	<i>delta proteobacterium MLMS-1</i>	Proteobacteria	Deltaproteobacteria	Q1NMU0	424	x
412428	<i>Teredinibacter turnerae</i>	Proteobacteria	Gammaproteobacteria	C5BT94	498	
412429	<i>Geobacter uramireducens</i>	Proteobacteria	Deltaproteobacteria	A5GA14	288	
412430	<i>Rubrivivax benzoatilyticus</i>	Proteobacteria	Betaproteobacteria	F3LRJ1	524	
412431	<i>Methylacidiphilum infernorum</i>	Verrucomicrobia	Methylacidiphilae	B3DYY9	523	
412432	<i>Methylococcus capsulatus</i>	Proteobacteria	Gammaproteobacteria	Q60CA5	522	
412433	<i>Microcoleus chthonoplastes</i>	Cyanobacteria		B4VJN9	427	x
412434	<i>Pelobacter propionicus DSM 2379</i>	Proteobacteria	Deltaproteobacteria	A1AUN5	289	
412435	<i>Rhizobium leguminosarum bv. trifolii</i>	Proteobacteria	Alphaproteobacteria	A1Z1Y5	495	
412436	<i>Methylocella silvestris</i>	Proteobacteria	Alphaproteobacteria	B8EJ39	519	
412437	<i>Arcobacter sp. L</i>	Proteobacteria	Epsilonproteobacteria	G2HZ63	494	
412438	<i>Rhizobium etli</i>	Proteobacteria	Alphaproteobacteria	B3Q2N6	277	x
412439	<i>Desulfobacca acetoxidans DSM 11109</i>	Proteobacteria	Deltaproteobacteria	F2NG51	291	
412440	<i>Anaeromyxobacter sp. K</i>	Proteobacteria	Deltaproteobacteria	B4ULY8	482	
412441	<i>Methanobrevibacter smithii ATCC 35061</i>	Euryarchaeota	Methanobacteria	B9AEE1	288	
412442	<i>Leptothrix cholodnii</i>	Proteobacteria	Betaproteobacteria	B1Y745	529	
412443	<i>Veillonella parvula ATCC 17745</i>	Firmicutes	Negativicutes	D1YN93	276	
412444	<i>Rhizobium etli</i>	Proteobacteria	Alphaproteobacteria	Q8KLB8	492	
412445	<i>Rhodobacter sphaeroides</i>	Proteobacteria	Alphaproteobacteria	B9KKU5	491	
412446	<i>Azorhizobium caulinodans</i>	Proteobacteria	Alphaproteobacteria	A8IIJ5	519	
412447	<i>Paenibacillus sabiniae</i>	Firmicutes	Bacilli	E1ABV1	458	
412448	<i>Clostridium butyricum</i>	Firmicutes	Clostridia	C4IC68	888	x
412449	<i>Syntrophobacter fumaroxidans</i>	Proteobacteria	Deltaproteobacteria	A0LJP4	293	

412450	<i>Denitrovibrio acetiphilus</i> DSM 12809	Deferribacteres	Deferribacteres	D4H6W5	446	
412451	<i>Prevotella bryantii</i>	Bacteroidetes	Bacteroidia	D8DUM7	247	x
412452	<i>Methanocella arvoryzae</i>	Euryarchaeota	Methanomicrobia	Q0W760	292	
412453	<i>Methanococcus vannielii</i>	Euryarchaeota	Methanococci	A6USF4	297	
412454	<i>Desulfonatronospora thiodismutans</i>	Proteobacteria	Deltaproteobacteria	D6SLD9	430	x
412455	<i>Rhizobium leguminosarum</i>	Proteobacteria	Alphaproteobacteria	C6B4Y4	490	
412456	<i>Cyanothece</i> sp. PCC 7822	Cyanobacteria		E0UH09	483	
412457	<i>Desulfotomaculum kuznetsovii</i> DSM 6115	Firmicutes	Clostridia	F6CNC8	271	
412458	<i>Frankia</i> sp. EAN1pec	Actinobacteria	Actinobacteria	A8L2B4	655	
412459	<i>Methanohalophilus mahii</i> DSM 5219	Euryarchaeota	Methanomicrobia	D5E7C7	322	
412460	<i>Clostridium beijerinckii</i>	Firmicutes	Clostridia	A6LR35	290	
412461	<i>Methylobacter tundripaludum</i>	Proteobacteria	Gammaproteobacteria	G3IZB1	496	
412462	<i>Desulfotomaculum carboxydivorans</i>	Firmicutes	Clostridia	F6B3H9	308	
412463	<i>Methanoculleus marisnigri</i>	Euryarchaeota	Methanomicrobia	A3CT87	303	
412464	<i>Desulfitobacterium hafniense</i>	Firmicutes	Clostridia	Q24P18	364	x
412465	<i>Desulfotomaculum reducens</i>	Firmicutes	Clostridia	A4J8B5	299	
412466	<i>Rhizobium etli</i> CNPAF512	Proteobacteria	Alphaproteobacteria	F2AGG4	492	
412467					519	x
412468	<i>Burkholderia phymatum</i>	Proteobacteria	Betaproteobacteria	B2JYC0	535	
412469	<i>Chlorobium luteolum</i> DSM 273	Chlorobi	Chlorobia	Q3B2N9	420	x
412470	<i>Methanosalsum chilinae</i> DSM 4017	Euryarchaeota	Methanomicrobia	F7XPI1	314	
412471	<i>Rhodopseudomonas palustris</i>	Proteobacteria	Alphaproteobacteria	B3QB57	518	
412472	<i>Treponema azotonutricium</i>	Spirochaetes	Spirochaetia	F5Y9P9	422	x
412473	<i>Klebsiella pneumoniae</i> 342	Proteobacteria	Gammaproteobacteria	B5XPF5	468	
412474	<i>Coralimargarita akajimensis</i> DSM 45221	Verrucomicrobia	Opitutae	D5EQN3	428	x
412475	<i>Xanthobacter autotrophicus</i>	Proteobacteria	Alphaproteobacteria	A7IBH7	520	
412476	<i>Dysgonomonas gadei</i> ATCC BAA-286	Bacteroidetes	Bacteroidia	F5IU62	420	x
412477	<i>Prosthecochloris aestuarii</i> DSM 271	Chlorobi	Chlorobia	B4S9I2	424	x
412478	<i>Chlorobium ferrooxidans</i> DSM 13031	Chlorobi	Chlorobia	Q0YUY3	424	x
412479	<i>Clostridium kluyveri</i> DSM 555	Firmicutes	Clostridia	A5N910	448	
412480	<i>Dethiobacter alkaliphilus</i> AHT 1	Firmicutes	Clostridia	C0GKD1	418	
412481	<i>Methanosarcina mazei</i>	Euryarchaeota	Methanomicrobia	Q8PYU4	328	
412482	<i>Frankia symbiont of Datisca glomerata</i>	Actinobacteria	Actinobacteria	F8B5P3	558	
412483	<i>Methanobrevibacter smithii</i> DSM 2374	Euryarchaeota	Methanobacteria	D2ZQ66	288	
412484	<i>Desulfovibrio vulgaris</i>	Proteobacteria	Deltaproteobacteria	B8DJB4	325	
412485	<i>Methanopyrus kandleri</i> AV19	Euryarchaeota	Methanopyri	Q8TXS8	282	x
412486	<i>Desulfitobacterium hafniense</i>	Firmicutes	Clostridia	B8FZP2	280	
412487	<i>Thermocrinis albus</i> DSM 14484	Aquificae	Aquificae	D3SPG1	472	
412488	<i>Cylindrospermopsis raciborskii</i> CS-505	Cyanobacteria		D4TKJ5	477	
412489	<i>Desulfitobacterium hafniense</i> Y51	Firmicutes	Clostridia	Q24Q99	306	
412490	<i>Magnetococcus</i> sp. MC-1	Proteobacteria	Alphaproteobacteria	A0L6X4	464	
412491	<i>Rhodopseudomonas palustris</i> BisB5	Proteobacteria	Alphaproteobacteria	Q13C88	519	
412492	<i>Rhodobacter capsulatus</i> SB 1003	Proteobacteria	Alphaproteobacteria	D5ANH7	497	
412493	<i>Rhodopseudomonas palustris</i> BisA53	Proteobacteria	Alphaproteobacteria	Q07HX3	518	
412494	<i>Veillonella</i> sp. oral taxon 158	Firmicutes	Negativicutes	E4LE39	276	
412495	<i>Clostridium kluyveri</i> DSM 555	Firmicutes	Clostridia	A5N6Z7	423	
412496	<i>Rhodobacter sphaeroides</i> ATCC 17025	Proteobacteria	Alphaproteobacteria	A4WRX8	491	
412497	<i>Clostridium beijerinckii</i>	Firmicutes	Clostridia	Q7BKZ0	896	x
412498	<i>Roseiflexus castenholzii</i> DSM 13941	Chloroflexi	Chloroflexia	A7NR79	313	
412499	<i>Methanosphaerula palustris</i> EI-9c	Euryarchaeota	Methanomicrobia	B8GJM8	303	
412500	<i>Paenibacillus fujiensis</i>	Firmicutes	Bacilli	B9X299	458	
412501	<i>Zymomonas mobilis</i> subsp. mobilis NCIMB 11163	Proteobacteria	Alphaproteobacteria	C8WE20	489	
412502	<i>Pelobacter carbinolicus</i> DSM 2380	Proteobacteria	Deltaproteobacteria	Q3A2R0	297	x
412503	<i>Methanothermobacter thermotrophicus</i> str. Delta H	Euryarchaeota	Methanobacteria	O27899	288	
412504	<i>Desulfosporosinus</i> sp. OT	Firmicutes	Clostridia	G2FXR3	312	
412505	<i>Methanospirillum hungatei</i> JF-1	Euryarchaeota	Methanomicrobia	Q2FLI5	298	
412506	<i>Dickeya dadantii</i> 3937	Proteobacteria	Gammaproteobacteria	E0SHZ3	465	
412507	<i>Treponema primitia</i> ZAS-2	Spirochaetes	Spirochaetia	F5YPB0	428	x
412508	<i>Cyanothece</i> sp. PCC 7425	Cyanobacteria		B8HWE0	495	
412509	<i>Rhodopseudomonas palustris</i> HaA2	Proteobacteria	Alphaproteobacteria	Q2JIJ1	518	
412510	<i>Ethanoligenens harbinense</i> YUAN-3	Firmicutes	Clostridia	E6U878	907	x
412511	<i>delta proteobacterium</i> MLMS-1	Proteobacteria	Deltaproteobacteria	Q1NW55	424	x
412512	<i>Veillonella parvula</i> DSM 2008	Firmicutes	Negativicutes	D1BR62	276	
412513	<i>Polaronomonas naphthalenivorans</i> CJ2	Proteobacteria	Betaproteobacteria	A1VML0	531	
412514	uncultured archaeon			D1JAT1	304	
412515	<i>Desulfotomaculum acetoxidans</i> DSM 771	Firmicutes	Clostridia	C8W448	304	
412516	<i>Chlorobium chlorochromatii</i> CaD3	Chlorobi	Chlorobia	Q3AR63	423	x
412517	<i>Methanothermobacter fervidus</i> DSM 2088	Euryarchaeota	Methanobacteria	E3GWP9	284	x
412518	<i>Methanococcus aeolicus</i> Nankai-3	Euryarchaeota	Methanococci	A6UUJ9	318	

412519	<i>Rhodopseudomonas palustris</i> CGA009	Proteobacteria	Alphaproteobacteria	Q6N0X9	518	
412520	<i>Methanococcus voltae</i> A3	Euryarchaeota	Methanococci	D7DSI7	297	
412521	<i>Ilyobacter polytropus</i> DSM 2926	Fusobacteria	Fusobacteriia	E3HAJ5	426	
412522	<i>Rhodimicrobium vannielii</i> ATCC 17100	Proteobacteria	Alphaproteobacteria	E3I4T9	297	
412523	<i>Beijerinckia indica</i> subsp. <i>indica</i> ATCC 9039	Proteobacteria	Alphaproteobacteria	B2IE25	527	
412524	<i>Cyanothece</i> sp. CCY0110	Cyanobacteria		A3IL23	491	
412525	<i>Paludibacter propionigenes</i> WB4	Bacteroidetes	Bacteroidia	E4T4D4	424	x
412526	<i>Methylobacterium</i> sp. 4-46	Proteobacteria	Alphaproteobacteria	B0UBK8	498	
412527	<i>Caldicellulosiruptor kronotskyensis</i> 2002	Firmicutes	Clostridia	E4SGT3	266	x
412528	<i>Anabaena azollae</i> , 'Nostoc azollae' 0708	Cyanobacteria		D7E3U6	479	
412529	<i>Oscillatoria</i> sp. PCC 6506	Cyanobacteria			484	
412530	<i>Synechococcus</i> sp. JA-3-3Ab	Cyanobacteria		Q2JTL3	478	
412531	<i>Methanocaldococcus fervens</i>	Euryarchaeota	Methanococci	C7P9B4	299	x
412532	<i>Anaerotruncus colihominis</i> DSM 17241	Firmicutes	Clostridia	B0PE91	207	x
412533	<i>Methanothermobacter marburgensis</i>	Euryarchaeota	Methanobacteria	D9PUZ6	288	x
412534	<i>Desulfovibrio africanus</i>	Proteobacteria	Deltaproteobacteria	F3YV05	445	x
412535	<i>Desulfovibrio africanus</i> str. Walvis Bay	Proteobacteria	Deltaproteobacteria	F3YWU7	291	
412536	<i>Alkaliphilus metalliredigens</i>	Firmicutes	Clostridia	A6TTX6	415	
412537	<i>Thiocystis violascens</i> DSM 198	Proteobacteria	Gammaproteobacteria	I3YAR0	503	
412538	<i>Rhizobium leguminosarum</i>	Proteobacteria	Alphaproteobacteria	Q8GBU8	497	
412539	<i>Desulfovibrio vulgaris</i>	Proteobacteria	Deltaproteobacteria	A1V143	473	x
412540	<i>Crocospaera watsonii</i>	Cyanobacteria		Q4C2Z1	490	
412541	<i>Halorhodospira halophila</i>	Proteobacteria	Gammaproteobacteria	Q588Z7	263	x
412542	<i>Desulfarculus baarsii</i> DSM 2075	Proteobacteria	Deltaproteobacteria	E1QE25	384	x
412543	uncultured <i>Desulfobacterium</i> sp.	Proteobacteria	Deltaproteobacteria	E1Y8C8	424	x
412544	<i>Ruminococcus albus</i> 8	Firmicutes	Clostridia	E9SF62	291	
412545	<i>Methanococcus maripaludis</i> C5	Euryarchaeota	Methanococci	A4FYD9	297	
412546	<i>Burkholderia</i> sp. CCGE1002	Proteobacteria	Betaproteobacteria	D5WN35	532	
412547	<i>Chlorobium phaeovibrioides</i> DSM 265	Chlorobi	Chlorobia	A4SFV3	420	x
412548	<i>Hyphomicrobium</i> sp. MC1	Proteobacteria	Alphaproteobacteria	F8JB82	519	
412549	<i>Calditerrivibrio nitroreducens</i> DSM 19672	Deferribacteres	Deferribacteres	E4TG34	295	
412550	<i>Desulfobulbus propionicus</i> DSM 2032	Proteobacteria	Deltaproteobacteria	E8RBV7	425	x
412551	<i>Methylosinus trichosporium</i>	Proteobacteria	Alphaproteobacteria	D5QKH0	519	
412552	<i>Oscillochloris trichoides</i> DG6	Chloroflexi	Chloroflexia	E1I9R3	307	
412553	<i>Cupriavidus taiwanensis</i> LMG 19424	Proteobacteria	Betaproteobacteria	B2AIY9	523	
412554	<i>Desulfotomaculum ruminis</i> DSM 2154	Firmicutes	Clostridia	F6DLK8	313	
412555	<i>cyanobacterium endosymbiont of Rhopalodia gibba</i>	Cyanobacteria		A1KYJ3	483	
412556					440	x
412557	<i>Clostridium pasteurianum</i> DSM 525	Firmicutes	Clostridia	Q6J2L8	929	
412558	<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ATCC 10988	Proteobacteria	Alphaproteobacteria	F8DSC8	489	
412559	<i>Methanosaela thermophila</i> PT	Euryarchaeota	Methanomicrobia	A0B690	299	
412560	<i>Sinorhizobium medicae</i>	Proteobacteria	Alphaproteobacteria	A6UMF6	490	
412561	<i>Geobacter metallireducens</i>	Proteobacteria	Deltaproteobacteria	Q39XV1	290	
412562	<i>Herbaspirillum seropedicae</i> SmR1	Proteobacteria	Betaproteobacteria	R0EHS3	525	
412563	<i>Wolinella succinogenes</i> DSM 1740	Proteobacteria	Epsilonproteobacteria	Q7MRF5	449	
412564	<i>Rhodopseudomonas palustris</i>	Proteobacteria	Alphaproteobacteria	Q20XZ1	518	
412565	<i>Rhodopseudomonas palustris</i> CGA009	Proteobacteria	Alphaproteobacteria	Q6N7A4	307	x
412566	<i>Methanococcus maripaludis</i> S2	Euryarchaeota	Methanococci	Q6LZH0	297	
412567	<i>Burkholderia xenovorans</i> LB400	Proteobacteria	Betaproteobacteria	Q13N21	531	
412568	<i>Rhizobium leguminosarum</i>	Proteobacteria	Alphaproteobacteria	P09825	162	x
412569	<i>Paenibacillus abekawaensis</i>	Firmicutes	Bacilli	B9X2B9	458	
412570	<i>Acetonea longum</i> DSM 6540	Firmicutes	Negativicutes	F7NPZ3	299	
412571	<i>Desulfuromonas acetoxidans</i> DSM 684	Proteobacteria	Deltaproteobacteria	Q1K1F3	287	
412572	<i>Allochroamatium vinosum</i> DSM 180	Proteobacteria	Gammaproteobacteria	D3RS02	500	
412573	<i>Spirochaeta thermophila</i> DSM 6192	Spirochaetes	Spirochaetia	E0RPC5	423	x
412574	<i>Methanosaela concilii</i> GP6	Euryarchaeota	Methanomicrobia	F4BYQ2	302	
412575	<i>Pseudomonas stutzeri</i> DSM 4166	Proteobacteria	Gammaproteobacteria	F2MZW2	504	
412576	<i>Helioibacterium modesticaldum</i> Ice1	Firmicutes	Clostridia	B0TAQ0	277	
412577	<i>Methanoplanus petrolearius</i> DSM 11571	Euryarchaeota	Methanomicrobia	E1RJ30	302	
412578	<i>Desulfatibacillum alkenivorans</i> AK-01	Proteobacteria	Deltaproteobacteria	B8FAB7	424	x
412579	<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	Cyanobacteria		Q2JP82	478	
412580	<i>Desulfovibrio salaxigens</i> DSM 2638	Proteobacteria	Deltaproteobacteria	C6BX24	423	x
412581	<i>Methanocaldococcus</i> sp. FS406-22	Euryarchaeota	Methanococci	D3S752	295	
412582	<i>Thermodesulfator indicus</i> DSM 15286	Thermodesulfo bacteria	Thermodesulfobacteria	F8ADV8	276	x
412583	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841	Proteobacteria	Alphaproteobacteria	Q1M7V9	497	
412584	<i>Desulfovibrio</i> sp. ND132	Proteobacteria	Deltaproteobacteria	F0JD10	391	x
412585	<i>Desulfomicrobium baculatum</i> DSM 4028	Proteobacteria	Deltaproteobacteria	C7LPB8	418	x
412586	<i>Acetonea longum</i> DSM 6540	Firmicutes	Negativicutes	F7NEX4	293	
412587	<i>Leptospirillum ferrodiazotrophum</i>	Nitrospirae	Nitrospira	C6HVB8	409	x

412588	<i>Desulfotobacterium hafniense DCB-2</i>	Firmicutes	Clostridia	B8G298	318	
412589	<i>Dehalococcoides ethenogenes 195</i>	Chloroflexi	Dehalococcoidia	Q3Z7D7	276	
412590	<i>Desulfovibrio aespoensis Asp-2</i>	Proteobacteria	Deltaproteobacteria	E6VSZ6	423	x
412591	<i>Rhizobium leguminosarum bv. trifolii</i>	Proteobacteria	Alphaproteobacteria	A1Z1W7	490	
412592	<i>Paenibacillus durus</i>	Firmicutes	Bacilli	Q9AKT9	458	
412593	<i>Methanobacterium sp. SWAN-1</i>	Euryarchaeota	Methanobacteria	F6D1V2	283	
412594	<i>Helibacterium chlorum</i>	Firmicutes	Clostridia	Q53UA6	277	
412595	<i>Methanococcus maripaludis X1</i>	Euryarchaeota	Methanococci	G0H409	297	
412596	<i>Desulfovibrio fructosovorans JJ</i>	Proteobacteria	Deltaproteobacteria	E1JXA2	425	x
412597	<i>Frankia sp. EulK1</i>	Actinobacteria	Actinobacteria	Q9Z5X8	658	
412598	<i>Cyanosphaera sp. PCC 8801</i>	Cyanobacteria		O09254	558	
412599	<i>Paenibacillus massiliensis</i>	Firmicutes	Bacilli	Q56JU5	473	
435526	<i>Crocospira watsonii</i>	Cyanobacteria			421	x
435527	<i>Phaeosporillum molischianum DSM 120</i>	Proteobacteria	Alphaproteobacteria		490	
435528	<i>Sinorhizobium meliloti CCNWSX0020</i>	Proteobacteria	Alphaproteobacteria		494	
435529	<i>Clostridium sp. DL-VIII</i>	Firmicutes	Clostridia		899	x
435530	<i>Mesorhizobium amorphae</i>	Proteobacteria	Alphaproteobacteria		493	
435531	<i>Sinorhizobium fredii HH103</i>	Proteobacteria	Alphaproteobacteria		493	
435532	<i>Clostridium clariflavum DSM 19732</i>	Firmicutes	Clostridia		900	
435533	<i>Desulfotomaculum gibsoniae DSM 7213</i>	Firmicutes	Clostridia		318	
435534	<i>Brenneria sp. EniD312</i>	Proteobacteria	Gammaproteobacteria		466	
435535	<i>Methanoseta harundinacea 6Ac</i>	Euryarchaeota	Methanomicrobia		305	
435536	<i>Bradyrhizobium sp. ORS285</i>	Proteobacteria	Alphaproteobacteria		520	
435537	<i>Fischerella sp. JSC-11</i>	Cyanobacteria			482	
435538	<i>Campylobacter sp. 10_I_50</i>	Proteobacteria	Epsilonproteobacteria		275	
435539	<i>Rhodospirillum photometricum DSM 122</i>	Proteobacteria	Alphaproteobacteria		500	
435540	<i>Bradyrhizobium sp. STM 3809</i>	Proteobacteria	Alphaproteobacteria		520	
435541	<i>Azospirillum lipoferum 4B</i>	Proteobacteria	Alphaproteobacteria		520	
435542	<i>Klebsiella oxytoca 10-5250</i>	Proteobacteria	Gammaproteobacteria		468	
435543	<i>Bradyrhizobium sp. WSM471</i>	Proteobacteria	Alphaproteobacteria		517	
435544	<i>Paenibacillus sp. HPL-003</i>	Firmicutes	Bacilli		499	
435545	<i>Bradyrhizobium sp. STM 3843</i>	Proteobacteria	Alphaproteobacteria		520	
435546	<i>Paenibacillus sp. Aloe-11</i>	Firmicutes	Bacilli		489	
435547	<i>Clostridium clariflavum DSM 19732</i>	Firmicutes	Clostridia		570	
435548	<i>Azoarcus sp. KH32C</i>	Proteobacteria	Betaproteobacteria		504	
435549	<i>Acetobacterium woodii DSM 1030</i>	Firmicutes	Clostridia		906	x
435550	<i>Rahnella aquatilis ATCC 33071</i>	Proteobacteria	Gammaproteobacteria		464	
435551	<i>Thiorhodovibrio sp. 970</i>	Proteobacteria	Gammaproteobacteria		501	
435552	<i>Ectothiorhodospira sp. PHS-1</i>	Proteobacteria	Gammaproteobacteria		505	
435553	<i>Rahnella aquatilis HX2</i>	Proteobacteria	Gammaproteobacteria		464	
435554	<i>Burkholderiales bacterium JOSHI_001</i>	Proteobacteria	Betaproteobacteria		528	
435555	<i>Klebsiella oxytoca 10-5242</i>	Proteobacteria	Gammaproteobacteria		468	
435556	<i>Methanoplanus limicola DSM 2279</i>	Euryarchaeota	Methanomicrobia		298	
435557	<i>Azospirillum brasilense</i>	Proteobacteria	Alphaproteobacteria		518	
435558	<i>Azoarcus sp. KH32C</i>	Proteobacteria	Betaproteobacteria		521	
435559	<i>Pseudomonas putida</i>	Proteobacteria	Gammaproteobacteria		485	
435560	<i>Azospira oryzae</i>	Proteobacteria	Betaproteobacteria		507	
435561	<i>Holophaga foetida DSM 6591</i>	Acidobacteria	Holophagae		428	x
435562	<i>Methanolinea tarda NOBI-1</i>	Euryarchaeota	Methanomicrobia		297	
435563	<i>Bradyrhizobium sp. ORS375</i>	Proteobacteria	Alphaproteobacteria		520	
435564	<i>Bradyrhizobium japonicum USDA 6</i>	Proteobacteria	Alphaproteobacteria		527	

Table 3. NifB amino acid sequences used to determine strictly conserved residue and protein architecture.

Twenty-eight NifB proteins representing all phylogenetic groups known to contain diazotrophs (9) were selected from the reduced list and used to investigate taxonomy versus architecture correlation. The taxonomy of diazotrophic groups was resolved using PhyloT (<http://phylot.biobyte.de/>), an online tool that uses the full NCBI taxonomy to generate phylogenetic trees.

Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to generate protein alignments and neighbor joining (NJ) phylogenetic trees (119). Maximum likelihood (ML) trees were produced using the IQ-Tree web server (<http://www.iqtree.org>) (120). Gblocks (121) was used to remove non-conserved aligned segments leaving a 315 contiguous amino acid sequence that was used to generate the SAM-radical domain tree and a 64 contiguous amino acid sequence used to generate the NifX-like domain trees. Phylogenetic trees were resolved using the Interactive Tree of Life online tool (<http://itol.embl.de>) (122) and FigTree.

3.5. NifB_{Mi} purifications

A scheme of NifB_{Mi} purification steps along with the equipment used is shown in Figure 13. Buffers used were:

Buffer A – Tris–HCl 50 mM pH 8, 200 mM KCl, 30 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol, 0.05% n-dodecyl-β-D-maltopyranoside, 0.1% Triton X-100, and 0.1% Tween 20.

Buffer B – 50 mM Tris–HCl, pH 8, 200 mM KCl, 30 mM imidazole, 10% glycerol, and 5 mM β-mercaptoethanol.

Buffer C – 50 mM Tris–HCl, pH 8, 200 mM KCl, 100 mM imidazole, 10% glycerol, and 5 mM β-mercaptoethanol.

Buffer D – 50 mM Tris–HCl, pH 8, 200 mM KCl, 250 mM imidazole, 10% glycerol, and 5 mM β-mercaptoethanol.

Buffer E – 50 mM Tris–HCl, pH 8, 200 mM KCl, 10% glycerol, and 5 mM β-mercaptoethanol.

Except when specified as aerobic NifB_{Mi} purification, all purification steps were performed at room temperature inside a glovebox. A standard NifB_{Mi} anaerobic purification procedure started with 300 g of cells from *E. coli* BL21 (DE3) pRSFiscMetKDuet-1 pRHB509 resuspended for 30 minutes in 900 ml buffer A supplemented with lysozyme (0.5 mg/ml), 100 μM phenylmethylsulfonyl fluoride (PMSF), leupeptine (0.5 μg/ml), and DNase I (5 μg/ml)) inside a glovebox at <2 ppm O₂.

The resuspended pellet was lysed by sonication (10 s pulse/15 s pause cycles during 10 min at 90% power output) using a UP100H sonicator equipped with a 12 mm tip. Lysate was incubated for 10 minutes on ice to facilitate DNA precipitation and then centrifuged at 70,000 x *g* for 30 minutes (Beckman JA 25.50 rotor). The supernatant was recovered and filtrated through 0.44 μ M filters to obtain the cell-free extract (CFE). CFE was loaded into a previously equilibrated IMAC HI trap Ni²⁺- charged column following manufacturer's instructions. The column was washed with 5 column volumes (CV) of buffer B and 2 CV of buffer C. Elution was performed with 1-2 CV of buffer D. Elution fractions were analyzed by SDS-PAGE. Valid eluted fractions, containing pure NifB_{Mi}, were pooled and desalted in a GE HiPrep 26/10 desalting column using buffer E. The final sample was concentrated in 3 kDa centricons and stored in liquid N₂ drops.

In some cases, NifB_{Mi} preparations were heated for 30 min at 65°C in buffer E and repurified by gel filtration on HiPrep 26/10 columns. Heat-treated preparations were analyzed by Superdex 200 chromatography to verify prevalence of the monomeric fraction. NifB_{Mi} preparations were then concentrated in a YM30Amicon cell, aliquoted and frozen in liquid nitrogen.

3.5.1. IMAC chromatography

Histidine-tagged proteins have a selective affinity for metal ions, such as nickel, that can be immobilized on chromatographic media using chelating ligands. To optimize the tag placement on the protein and reduce the potential problem of inaccessibility of the protein tag to the immobilized metal due to tag occlusion in the folded protein, the polyhistidine tag was located to the N-terminus. A variety of immobilized metal matrices are available for use in IMAC. In this work we used nickel-nitrilotriacetic acid (Ni²⁺-NTA). This matrix coordinates metal ions through four coordination sites while leaving two of the transition metal coordination sites exposed to interact with histidine residues in the affinity tag (Figure 14). Filtrated cell-free extracts from *E. coli* cells overexpressing NifB_{Mi} were loaded onto a Ni²⁺-charged 5-ml HiTrap column previously equilibrated with 10 column volumes (CV) of buffer B.

Figure 14. Ni^{2+} -NTA coordinated to agarose beads. The Ni^{2+} -NTA ligand is covalently attached to a cross-linked agarose matrix for the selective purification of polyhistidine-tagged proteins. Left: histidine residues can coordinate to the Ni^{2+} ion by replacing the bound water molecules (indicated by red arrows). Right: scheme showing coordination by two histidine residues (123).

The column was washed with 5 CV of buffer B and 2 CV of buffer C. Elution was performed with 1-2 CV of buffer D. In general, the various NifB_{Mi} mutant forms presented the same elution profile as the wild type and the protein elution was observed after 15- 20 ml of buffer D application. Elution fractions were analyzed by SDS-PAGE. Valid eluted fractions that contained the higher amount of NifB_{Mi} and less contaminant proteins were pooled and desalted as described in the next section.

3.5.2. HiPrep 26/10 desalting chromatography

To separate and remove small molecules that can interfere with downstream biochemical assays or protein stability (such as imidazole), 10 ml of NifB_{Mi} protein obtained from IMAC chromatography was loaded on a HiPrep 26/10 column. This column is packed with the gel filtration medium Sephadex G-25 Superfine, a matrix based on cross-linked dextran beads. Before loading the protein, the desalting column was equilibrated with 10 CV of buffer E. The chromatography was followed using the UV absorbance detector of an AKTA FPLC apparatus. Desalted fractions were collected, pooled, and then concentrated in an Amicon cell equipped with YM30 membranes (Diaflo). Desalted NifB_{Mi} preparations were stored in liquid nitrogen inside anaerobic hermetic 1 ml vials.

3.5.3. Superdex 200 size exclusion chromatography

The HiLoad 16/600 Superdex 200 column is a pre-packed size exclusion chromatography column for high-resolution small-scale preparative and

analytical separations of biomolecules. Superdex 200 has a separation range for molecules with molecular weights between 10,000 and 600,000 Da. Before loading the protein sample (1.25 mg of NifB_{Mi} in 500 μ l), the desalting column was equilibrated with 10 CV of 50 mM Tris-HCl pH 8, 200 mM KCl, 5 mM DTT. The chromatography was run at a flow rate of 1 ml/min and followed using the UV absorbance detector of an AKTA FPLC apparatus.

3.6. Preparation of NifB_{Mi} apo-protein

To remove [Fe-S] clusters and generate apo-NifB_{Mi}, purified enzyme preparations were treated in the presence of 10 mM EDTA inside the MBRAUN glove box (2 hours, 17°C) under reducing conditions with 5 mM DTH. After the incubation step, in order to remove the excess of Fe and Sulfur precipitated after the exposition at strong reduced conditions, the proteins were centrifuged at 11,000 rpm for 10 minutes using a mini centrifuge inside the glove box. The clarified protein solution was desalted on 5 ml PD10 columns equilibrated with buffer E (50 mM Tris-HCl, 200 mM KCl, 10% glycerol, and 5 mM β -mercaptoethanol). This chromatography step removes EDTA-Fe and DTH. Once desalted, NifB_{Mi} was concentrated by membrane ultrafiltration (Amicon cell, YM30, Diaflo) inside the glove box and frozen in liquid N₂.

3.7. Reconstitution of Fe-S clusters

The [Fe-S] centers of the various NifB_{Mi} variants were reconstituted in atmosphere controlled glove box (O₂ < 2 ppm) at 20°C in Buffer E. NifB_{Mi} samples (10 μ M) were incubated during 2 h at 50°C with a 12-fold molar excess of Fe²⁺ ((NH₄)₂Fe(SO₄)₂) and S²⁻ (Na₂S) in the presence of 10 mM DTT. The reconstitution reactions inside the glove box were followed by UV-visible spectroscopy by using optical fibers that connected the reading cell (optical path of 1 mm) to a Shimadzu spectrophotometer. After 2 hours of incubation, the reactions were stopped and the solutions were loaded onto a HiPrep 26/10 Desalting column (GE Healthcare life Science) equilibrated in buffer E. This filtration eliminated the low molecular weight molecules (cysteine, alanine, DTT, Fe²⁺) that were in excess. The resulting dark brown protein solutions were then concentrated by centrifugation (30-kDa Vivaspinn, Sartorius Stedim).

3.8. Reduction of NifB_{Mi} [Fe-S] clusters

NifB_{Mi} [Fe-S] clusters could be reduced *in vitro* by adding 10 molar excess DTH. A DTH/EDTA 100x stock solution was used (50 mM Tris- HCl, pH 8.5% glycerol and 480 mM EDTA; degassed with N₂ for 2 hours and DTH added to a final concentration of 480 mM in anaerobic glovebox. Solution must be used within 30 min after the addition of DTH). Reduction reactions were carried out on protein solutions in Buffer E. Reactions were monitored by UV-visible spectroscopy. The decay of the charge transfer band at 400 nm reflects the reduction of [Fe-S] clusters.

3.9. Growth and preparation of UW140 cell-free extracts for *in vitro* FeMo-co insertion and nitrogenase reconstitution assays

A. vinelandii is an aerobic microorganism with a slower duplication phase comparing with *E. coli*, estimated around 3 hours per duplication. In order to prepare the UW140 free extract that was used in the *in vitro* nitrogenase assays UW140 inoculated in a Burk's medium plate and incubated at 30°C for two days, until small colonies appear in the plate. An individual colony was used to inoculate a 250 ml flask with 50 ml of Burk's medium supplemented with ammonia, which was incubated at 30°C overnight with 200 rpm shaking. When the culture reached an OD of 1.2, it was used to inoculate 1-liter flask with 250 ml of Burk's medium supplemented with ammonia at a initial OD₆₀₀ of 0,002. This culture was incubated overnight until reaching an OD of 1.2, at which point the cells were derepressed for nitrogenase. Nitrogenase derepression involved harvesting by centrifugation at 4000 rpm for 8 minutes and resuspension in 250 ml of Burk's media without ammonia in 1-liter flasks. The culture was incubated under diazotrophic conditions for 3.5 hours at 30°C and 200 rpm, after which cells were harvested by centrifugation at 4000 rpm for 8 minutes and frozen at -80°C until use.

Cell-free extract preparations were performed under anaerobic conditions. Cell disruption was performed in 40-ml centrifuge tubes containing 5 g of cells, 20 ml of glycerol buffer (4M glycerol, 1 mM DTH), and around 30 glass beads. The mixture was incubated on ice for 30 minutes and then spun down for 15 min at 15,000 rpm at 4°C. The supernatant was carefully removed and the pellet detached from tube's wall using a spatula. Three ml of lysis buffer (Table 4) was added per g of cells and

immediately vortexed vigorously for 1 min. Cell debris was removed by centrifugation at 18,500 rpm and 4°C for 30 min. The supernatant (cell-free extract) was removed and combined with 1 ml of glycerol buffer per gram of cells. If not used immediately, the cell-free extract was stored at -80°C in sealed vials under anaerobic conditions.

Compound	Initial concentration	Volume used
Tris-HCl pH 7.5	25 mM	40 ml
PMSF	100 mM	160 µl
Leupeptine	1 mg/ml	20 µl
DTH	60 mM	800 µl

Table 4. Lysis buffer used for preparation of UW140 cell-free extracts

3.10. NifB_{Mi} *in vitro* activity assays

In order to reproduce the nitrogenase biosynthetic system, it is widely accepted that the *in vitro* synthesis of FeMo-co requires a minimum compendium of molecules as homocitrate, molybdate, MgATP, DTH as well as Nif proteins (at a minimum NifB, NifH, NifEN, and FeMo-co deficient apo-NifDK). These proteins can be added as a multicomponent mixture if using *A. vinelandii* cell-free extracts or as single pure proteins (59). All assays were performed in 9-ml serum vials under argon, or argon/acetylene (96%/4%) atmospheres, were crimp sealed with a butyl rubber stopper. In order to move all potential O₂ contaminants from the vials, they were left overnight inside the glovebox.

NifB_{Mi} activity was determined under strict anaerobic conditions inside an MBraun glovebox (O₂ < 0.1 ppm). The *in vitro* assay designed to determine NifB_{Mi} activity (i.e. synthesis of NifB-co followed by synthesis of FeMo-co, insertion of FeMo-co into apo-NifDK, and determination of reconstituted NifDK activity) has evolved over the development of this thesis. Indeed, the conditions used in the assays below were adjusted from the original protocol (69), by balancing the physiological requirements of proteins of mesophilic origin (*A. vinelandii*) and hyperthermophilic origin (*M. infernus*). As depicted in Figure 16, assays included three different steps: *i*) Pre-

activation of NifB_{Mi} (AB), *ii*) NifB-co dependent FeMo-co synthesis and apo-NifDK activation (CSDKF), and *iii*) determination of reconstituted holo-NifDK activity (ADK). The first phase was designed to deal with the temperature incompatibility between NifB_{Mi} and the rest of Nif proteins.

For temperature dependence assays with *A. vinelandii* UW140 cell-free extracts we performed the last two steps (CSDKF + ADK) using four different temperatures (30°C, 40°C, 50°C, and 60°C) during the SS.

For temperature dependence assays with pure Nif proteins we performed all three steps (AB + CSDKF + ADK) using three different temperatures (30°C, 40°C, and 60°C) in the AB (Figure 15). If another temperature is not specified the AB step was carried out at 60°C. First, 80 µM NifB_{Mi} was incubated for 30 min at 60°C in 0.2 ml of a reaction mixture containing 5 mM DTH, 1 mM SAM, 500 µM FeSO₄, and 500 µM Na₂S. Control reactions substituted NifB-co (84) for NifB_{Mi}. Fifty µl aliquots of NifB_{Mi} were then transferred to reaction mixtures for *in vitro* FeMo-co synthesis and Holo-NifDK formation (CSDKF) and incubated for 30 min at 30°C. Complete reaction mixtures (0.4 ml) contained: 22 mM Tris-HCl pH 7.4, 3.5% glycerol, 17.5 µM Na₂MoO₄, 175 µM *R*-homocitrate, 10 µM NifB_{Mi}, 3 µM NifX, 1.2 µM NifY, 3.5 µM apo-NifEN, 3.0 µM NifH, and 0.6 µM apo-NifDK. Activity of reconstituted NifDK (ADK) was analyzed by the acetylene reduction assay after addition of excess NifH and 0.42 ml of ATP-regenerating mixture (1.32 mM ATP, 18 mM phosphocreatine, 2.2 mM MgCl₂, 3 mM DTH and 40 µg creatine phosphokinase). Ethylene formation was measured in a Shimadzu GC-2014 gas chromatograph equipped with a Porapak N 80/100 column (124). Activity controls with pure NifDK were carried out with 0.04 mg of NifDK incubated for 30 minutes with excess NifH and 0.42 ml of ATP-regenerating mixture at the indicated temperatures.

For time-course experiments with simultaneous detection of ethylene, SAH and AdoH, the 60°C NifB_{Mi} pre-activation step was omitted. NifB_{Mi}-dependent FeMo-co synthesis and apo-NifDK activation reactions were performed at 40°C in a series of 0.8 ml reaction mixtures as described above. Reactions were split into 0.4 ml aliquots at different times (0, 15, 30, 45 and 60 min), stopped, and processed differently. In aliquots used to determine NifDK reconstitution, reactions were stopped by addition of 0.3 mM (NH₄)₂MoS₄ and the acetylene reduction assay performed after addition of

excess NifH and ATP-regenerating mixture (125). In aliquots used to quantify SAH and AdoH, reactions were stopped by addition of 50 μ l of 1 M TFA and exposure to air, and then analyzed by High Pressure Liquid Chromatography (HPLC). Reaction samples (0.25 ml) were injected onto a Zorbax SB-C18 column (protected with a pre-guard SB-C18 column) equilibrated with 0.1% TFA and resolved with a 0-100% linear gradient of acetonitrile in 0.1% TFA at 1 ml/min for 25 min. AdoH and SAH were identified by UPLC/ESI-qTOF-MS by comparing their profiles to those of commercial standards (Sigma-Aldrich D1771 and A9384, respectively). The standards have the following retention times: SAM (7 min), SAH (12 min), and AdoH (13 min).

Figure 15. Summary of experimental conditions for *in vitro* FeMo-co biosynthesis assays.

3.11. Gas Chromatography

Nitrogenase activity was measured by C₂H₂ reduction to C₂H₄ and quantification of the C₂H₄ produced using a Shimadzu GC-2014 gas chromatograph fitted with Porapak N 80/100 column.

3.12. SDS Poly Acrylamide Gel Electrophoresis

The discontinuous buffer system of Laemmli (126) was used to analyze proteins on a molecular weight basis via SDS polyacrylamide gel electrophoresis. SDS sample buffer stock contained in a total volume of 7 ml: 1 ml Tris-HCl 0.5 M (pH 6.8), 2 ml glycerol, 1.6 ml of SDS 10% (w/v), 0.4 ml β -mercaptoethanol, and 2 ml 0.05% bromophenol blue (w/v).

3.13. Protein determination

Protein concentration was determined by bicinchoninic acid method (Pierce), with bovine serum albumin as the standard (127). In the case of purified NifB_{Mi} preparations, protein concentration was additionally determined by using the molar extinction coefficients of the apo- and holo- forms.

3.13.1. Fe quantification

The amount of iron in protein preparations was assayed by the method of Fish (128). To establish the standard curve, a stock solution of 500 mM Fe²⁺ was used to prepare standards containing 0, 2, 4, 6, 8 and 10 nmol Fe in a final sample volume of 65 μ l. Likewise, different amounts of protein, from 0.5 to 2 nmol, were dissolved in a final volume of 65 μ l. To improve the precision of the assay, samples and standards were performed in duplicate. All samples and standards were denatured by incubation with 45 μ l of 1M perchloric acid for 1 h. After centrifugation (13,000 rpm, 10 min) 90 μ l of solution were separated into another tube and combined with 72 μ l of batho-phenanthroline disulfonate (1.7 mg / ml), 36 μ l of sodium ascorbate (38 mg / ml) and 27 μ l of an ammonium acetate saturated solution diluted to 1/3. After 30 minutes incubation at room temperature the DO₅₃₅- DO₆₈₀ absorbance was measured on a Shimadzu spectrometer. The amount of Fe per enzyme monomer was obtained by comparing absorbance obtained to the Fe standard.

3.13.2 Sulfur quantification

Protein labile sulfur content was determined by the method of Beinert (129). A translucent sodium sulfide crystal (Na₂S.9H₂O, M = 240 g/mol) was weighed and

dissolved in 1 M sodium hydroxide to a concentration of 100 mM Na₂S. A working solution of 200 μ M Na₂S was obtained from the stock solution. Standards of 0, 2, 4, 6, 8 and 10 nmol S were prepared in a final volume of 100 μ l. The same amounts of protein as used for concomitant Fe determinations were prepared in a final volume of 100 μ l. Protein samples and standards were made in duplicate in order to improve dosing accuracy. Protein samples were denatured for 3 h with simultaneous addition of 300 μ l of 1% zinc acetate and 15 μ l of 3 M sodium hydroxide, causing the immediate precipitation of zinc hydroxide. After the denaturation step, samples were taken over by the simultaneous addition of 75 μ l of DMPD (from a stock solution containing 1 mg DMPD in one ml of 5 N HCl) and 16 μ l of FeCl₃ (from a stock solution containing 3.75 mg FeCl₃ in 1 ml of 1 N HCl) leading to the slow formation of methylene blue. After an incubation of 30 minutes at room temperature, the absorbance DO₆₇₀ - DO₈₅₀ was measured on a Shimadzu spectrometer. The amount of labile S per enzyme monomer was obtained by comparing absorbance obtained to the sulfur standard.

3.13.3. Mass Spectroscopy

As isolated NifB_{Mi} was analyzed in a 4800 Proteomics Analyzer with TOF/TOF mass spectrometer at the Unidad de Proteómica of Parque Científico de Madrid.

3.13.4. UPLC/ESI-qTOF-MS analysis of reaction sub-products

These experiments were performed by Dr. Stephan Pollmann (CBGP, Madrid).

Dried samples were re-suspended in 50 μ l acidified methanol (methanol, 0.1% formic acid [v/v]). Five microliters of the solution were separated using a Dionex UltiMate3000 RSLC system (Dionex, Sunnyvale, USA). The column used was a 50 mm x 2.1 mm i.d., 1.7 μ m, Acquity UPLC BEH C18 column with a 5 mm x 2.1 mm i.d., 1.7 μ m, VanGuard pre-column of the same material, and a 0.2 μ M in-line filter set (Waters, Milford, USA). The following binary gradient was applied: 0 to 1 min isocratic 100% A (deionized water, 0.1% formic acid [v/v]); 1 to 10 min linear gradient to 28% B (acetonitrile, 0.1% formic acid [v/v]); 10 to 15 min linear gradient to 100% B; then 100% B isocratic for 2 min; re-conditioning of the column, 2.5 min

isocratic 100% A. The flow rate was 400 μ l/min. The auto injector temperature was set to 4°C, the column compartment was maintained at 30°C.

Eluted compounds were detected by a micrOTOF-Q II mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an electrospray ionization source in positive ion mode. Typical instrument settings were as follows: capillary voltage, 4500 V; capillary exit, 130 V; dry gas temperature, 200 °C; dry gas flow, 10 l/min. Ions were detected from m/z 50 to 1000 at a repetition rate of 2 Hz. Mass calibration was performed using sodium formate clusters (10 mM solution of NaOH in 50/50% v/v isopropanol/water containing 0.2% formic acid).

3.13.5. Deuterium and ^{15}N -SAM substitution experiments

Each reaction contained, in a total volume of 150 μ l, 25 mM Tris-HCl (pH 7.4), 10% glycerol, 500 μ M β -Mercaptoethanol, 5 mM DTH, and 10 μ M *MiNifB*. The reaction was initiated by the addition of 5 mM unlabeled SAM (Sigma-Aldrich) or 500 μ M $^{13}\text{CD}_3$ -SAM incubated for 15 min at 65°C with intermittent mixing, and subsequently terminated by removing *MiNifB* from the solutions using Amicon Ultra 30,000 MW centrifugal filters. Filtered samples were then analyzed by HPLC and LC-MS at CBGP Metabolomics Facility. SAM was synthesized starting with unlabeled methionine, or with labeled $^{13}\text{CD}_3$ -methionine or ^{15}N -methionine and purified as described (130). Synthesized $^{13}\text{CD}_3$ -SAM and ^{15}N -SAM were kindly donated by Dr. Etienne Mulliez.

3.13.6. UV-Visible spectroscopy

UV-visible spectroscopy was used to determine the concentration of apo-proteins and holo-proteins. Proteins containing a $[\text{Fe}_4\text{S}_4]^{2+}$ have a characteristic absorption band at 400 nm. The reconstruction of the centers $[\text{Fe-S}]$ of *NifB_{Mi}* is followed in the M'Braun glove box with a cell connected to a spectrophotometer Shimadzu UV-2600 equipped with optical fibers spectrometer UV4 fitted with diode array using 0.5 ml quartz cuvettes and scans were performed between 250 nm and 700 nm.

3.13.7. Electron paramagnetic resonance (EPR)

EPR spectroscopy makes it possible to study the so-called paramagnetic species, which contain one or more unpaired electrons. The paramagnetic entity is placed in a variable magnetic field H and receives an electromagnetic wave of energy $h\nu$. When the energy $h\nu$ of the wave is equal to the energy necessary to allow the transition between two states of spin of the electron, the phenomenon of paramagnetic resonance occurs. The transition energy is equal to $g\beta H$, where g corresponds to the Landé factor ($g_e = 2.0023$ for the free electron) and β to the Bohr magneton. The unpaired electron in a molecular orbital has a factor g whose value differs from that of g_e as a function of the paramagnetic center and its molecular environment. The factor g is anisotropic, which means that it can be assimilated to a tensorial variable defined by three components g_x , g_y and g_z . These three components correspond to the points of inflexion and to the extremes of the absorption spectrum recorded during the scanning of the values of the magnetic field (Figure 16). In fact, the spectrum obtained experimentally is the derivative of the absorption signal for which the determination of the values of g is more precise. The X-band EPR spectroscopy allows the study of paramagnetic species that present one or more unpaired single electrons. Applied to the study of the centers [Fe-S], EPR informs us about the type of center, its redox state and its quantity since the signal can be integrated and the value of the integration compared to that of a standard solution of known concentration. A singular electron presents a electronic spin: $S=1/2$. In the case an atom is surrounded by several electrons, this can be coupled by two spin states ($m_s + 1/2$ and $-1/2$). In this case the total spin (S) of the molecule is zero. But when is not the case, the molecule spin is multiple of $1/2$ and in this case the spin is $\neq 0$. This last situation is common in the free radicals (an unpaired electron) or with molecules incorporating metallic atoms (one or more unpaired electrons depending on the nature of the metallic atom (s) involved). This technique was used to characterize the [Fe-S] clusters of NifB_{Mt} wild type and all the mutant forms in collaboration with Dr. Jarett Wilcoxon in the laboratory of David Britt (UC-Davis, California, USA). Experimentally, a volume between 100 μ l - 200 μ l of sample was reduced and then frozen inside the glovebox. Samples were prepared in a Vacuum Atmospheres glove box (VAC, model LM02) under nitrogen atmosphere by mixing the reaction components and filling a 4.0 mm O.D. x 10 cm quartz tube, sealed with rubber septa,

and freezing in liquid nitrogen or an acetone/dry ice bath immediately upon removal from the glove box. EPR spectra were measured at the CalEPR Center at the University of California, Davis. Continuous-wave (CW) X-band spectra were acquired with a Bruker Elexsys-II E500 spectrometer (Bruker, Billerica, MA) under non-saturating conditions using a Super-High Q resonator (ER 4122SHQE). Cryogenic temperatures were achieved and maintained using an Oxford Instruments ESR900 liquid helium cryostat in conjunction with an Oxford Instruments ITC503 temperature and gas flow controller.

Figure 16. Absorption signal χ'' and its derivative $d\chi''/dH$ obtained by varying the field H . The derivative of the absorption signal makes it possible to determine the tensor quantities g_x , g_y and g_z , which are characteristic of the paramagnetic center studied (55).

Unless indicated, CW spectra were measured at 9.38 GHz microwave frequency, 1 mW power, 5 G modulation amplitude, at 15 K.

Concentrations of EPR samples of NifB were calculated using a 300 μ M Cu(II)-EDTA standard collected at 100K, 1 mW microwave power, and 3G modulation amplitude. Spectra intensities were quantified by taking the double integral of the collected data and adjusting for temperature, power and modulation amplitude by the equation:

$$I = I_{\text{rel}} * T / (\text{sqrt}(P) * A)$$

Where I is the parameter adjusted intensity, I_{rel} is the unadjusted maximum of the double integral of the spectrum, T is temperature in kelvin, P is the microwave power and A is the modulation amplitude.

3.13.8. One-dimensional (1D) electron-spin echo envelope modulation (ESEEM)

ESEEM is a pulsed EPR technique that monitors nuclear magnetic resonance (NMR) transitions indirectly through EPR transitions (131, 132). In the experiment, the electron spin-echo amplitude is monitored as it is modulated by the nuclear spin of nearby atoms at characteristic electron-nuclear double resonance (ENDOR) frequencies. Typically, ESEEM spectroscopy has been utilized to study metal binding, coordination, radical formation, and metalloenzyme mechanisms.

3P-ESEEM data were acquired with a Bruker EleXsys E580 pulse EPR spectrometer equipped with an Oxford-CF935 liquid helium cryostat and an ITC-503 temperature controller using the pulse sequence $\pi/2$ - τ - $\pi/2$ - T - $\pi/2$ -echo where the delay time, T , is varied. 3P-ESEEM were collected at 10 K, $\tau = 128$ ns, $\pi/2 = 12$ ns, and a magnetic field of 340 mT. Data processing and spectra simulations were performed with Matlab using the EasySpin 4.0 toolbox (133).

3.13.9. Two-dimensional (2D) electron-spin echo envelope modulation (HYSCORE)

The HYSCORE spectroscopy (Hyperfine Sublevel Correlation) detects hyperfine coupling between a paramagnetic metal cluster $[\text{Fe}_4\text{S}_4]^+$ and related species having a non-zero spin which can be an atom (^1H , ^2N , ^{13}C , ^{14}N , ^{15}N) or a radical species (134). HYSCORE highlights the presence of a ligand to the $[\text{Fe}_4\text{S}_4]^+$ cluster differentiating between weak or strong interaction.

NifB_{Mi} wild type (WT) or NifB_{Mi} double mutant (DM) (311 μM protein per sample) were incubated with 3 mM ^{15}N -SAM in the presence of 10 mM DTH. HYSCORE parameters for NifB_{Mi} WT were: temp 128t, 10 K, field 350 mT, depth 3.00 and 3.30. For the Holo NifB_{Mi} DM were: temp 128t, 10 K, field 350 mT, depth 3.00.

3.13.11. Crystallography screenings

The X-ray crystallography is used to access the three-dimensional structure of the ordered material. The interatomic distance is to the order of 0.1 nm and the same atoms perturb the rays used by this technique, which have a wavelength between 0.01 and 10 nm. Since the intensity of the rays diffused by a molecule is too low, this technique requires crystals in which the protein, stacked periodically, is found repeatedly in the three dimensions of the space. Waves diffused in phase (constructive interference) will be added to reach a measurable signal on a detector.

Protein crystals are obtained when the protein passes from a liquid state to an organized solid state (135). The most widely used crystallization technique is the vapor diffusion, either as a seated drop or as a suspended drop (Figure 17). For the latter procedure, a drop of a few μl of a protein / precipitating mixture is deposited on a siliconized slide (to prevent the droplet spreading) on the top of a well containing the same precipitating agent in volume 50 to 1000 times larger. The silicon slat is sealed to the well with grease to make the system watertight.

Since the precipitating agent concentration is less in the drop than in the well, exchanges of vapor (water or organic solvent) take place between the reservoir solution in the well and the drop in order to obtain an osmotic equilibrium. When the ratio of protein/precipitating agent was 2:1, the drop volume at equilibrium is approximately halved, which implies that the concentration of the protein doubles as well.

Figure 17. Diffusion vapor techniques used during crystallogenesis (135).

In normal conditions, nucleation phenomenon can occur, leading the crystals formation. In the case of the seated drop, the drop is deposited on a support containing a silicone cavity, which contains the precipitating agent. The well is closed with scotch tape.

The parameters that affect protein crystallization and/or solubility are: super saturation (protein and precipitating agent concentration), temperature, pH, time (equilibration and crystal growth), ionic strength and purity of chemicals (nature of precipitant, buffer, or additive), and purity and homogeneity of the protein.

The crystallization solution may be composed of a stable pH buffer, precipitating agent, additives (alcohol, salts, etc.), cofactors, inhibitors or detergents. The precipitating agents may be of different natures: organic solvents (isopropanol, MPD, ethanol, etc.) that decrease the dielectric constant of the medium; salts (ammonium sulfate, NaCl, etc.) that modify the ionic strength and increase the hydrophobic interactions between the protein molecules while excluding the solvent; and organic polymers (PEG, PEGMME).

Figure 18. describes in detail the phase diagram representing the solubility of the protein as a function of the precipitating agent concentration. The solubility curve separates two main zones, sub-saturation and super saturation. In the sub-saturation zone, the protein concentration is too low to allow the crystals formation. In the super saturation zone, constituted by the i) metastable, ii) nucleation and iii) precipitation

zones, the protein excess will have either the possibility to precipitate or crystallize. In the precipitation zone, as its name indicates, we have the precipitation of the protein. In the nucleation zone, the super saturation level of the protein is sufficient for the appearance of crystals “neo” forms and their subsequent growth. The driving force in this process is the super saturation level, which provides the energy necessary for the molecules association. In the metastable zone, the kinetic frequency of molecular interaction is lower, with the consequence of a slower crystal formation. In this area the crystals' nuclei can growth in order of months or even years. Figure 18 shows the area where crystal formation occurs (path i', j', k') or not (path i, j). The concentration of protein in solution decreases until reaching the solubility, where there is a balance between the protein in solution and the crystal.

Figure 18. Vapor diffusion diagram and experimental paths. The drop presents an equivalent volume protein/precipitating agent. C_P , C'_P Protein concentration; C_A , C'_A : precipitating agent concentration. The points i and i' are the two initial conditions (135).

Crystallization tests were performed with the use of crystallization robot inside a Jacomex SAS to quickly cover a greater number of different conditions. Each different buffer for the optimization conditions was previously degassed inside Jacomex with modified gas system. Various conventional matrices of crystallization conditions (96 conditions in total) were tested under anaerobic conditions in order to cover different types of precipitants (polyethylene glycols, ammonium sulfate, methyl

pentanediol, etc.) and pH. These first tests allowed us to determine the most favorable conditions for the appearance of crystals. The first experiments with NifB_{Mi} protein were carried out at different range concentration (5, 10, 15, 25 mg/ml protein), previously incubated for 2 hours with 1 mM SAM and 1 mM FAD for a total of 13 plates for each NifB_{Mi} concentration tested.

Once the crystals of the protein of interest were obtained, they were rapidly frozen in liquid nitrogen or liquid propane using a cryoprotectant. The NifB_{Mi} proteins were sensitive to oxygen. For this reason we used a freezing protocol specific for the glove box (136). The crystals obtained were cryoprotected with 0.1 M solution of sodium cacodylate buffer pH 5.6 and 40% (v/v) of PEG 400. The addition of a cryoprotectant during freezing prevents the formation of crystalline ice on the outside and inside the crystals, which contain 30 to 80% of water.

During the X-ray diffraction experiment, the crystals, previously frozen, were kept under nitrogen gas at about 100 K. The very low temperature slows down the propagation of the radioactive damage in the crystals, caused by exposure to high-intensity X-rays. A dataset was collected at the ESRF (European Synchrotron Radiation Facility), Grenoble – France, on line ID14-eh4. The data were processed with the XDS program.

The entire crystallographic study (crystal growth, crystallogensis optimization and data collection) of NifB_{Mi} was conducted in collaboration in the laboratory of Dr. Juan Fontecilla and Dr. Yvain Nicolet at the IBS (Institut de Biologie Structurale, Grenoble, France).

Results

Chapter I: Phylogenetic studies and purification of the NifB_{Mi} protein

In this chapter we study the diversity and phylogeny of naturally occurring NifB proteins, their protein architecture and the functions of the distinct NifB domains in order to understand what defines a catalytically active NifB. Focus is on NifB from the hyperthermophile *M. infernus* (single-domain architecture), showing *in silico* characterization of their *nif* gene clusters, conserved NifB motifs, and *in vivo* functionality. Moreover, we present methods for the expression and purification of NifB_{Mi} in recombinant *E. coli* cells showing that NifB_{Mi} is more robust than NifB proteins from *A. vinelandii* and *K. oxytoca* purified to date. Parts of this chapter have been published in Methods in Molecular Biology 1122: 19-31. Other parts have been submitted for publication at Frontiers in Plant Science.

4.1. Phylogenetic Distribution of Three Distinct NifB Domain Architectures

The primary sequence of a protein can be used to predict its structural and functional properties. Furthermore, the observation that biological sequences evolve at measurable and relatively constant rates permits the analysis of molecular evolution at the sequence level (137, 138). This information is derived by comparing the amino acid sequences of proteins and can be used to infer functional and evolutionary links between them. To understand what defines the catalytically active NifB, we studied the diversity and phylogeny of naturally occurring NifB proteins, their protein architecture and the role of its distinct domains. The 390 putative NifB sequences found in the Structure and Function Linkage Database (SFLD) (111) are shown in Table 3. By aligning experimentally proven NifB proteins from *A. vinelandii* (NifB_{Av}) (85), *K. oxytoca* (NifB_{Ko}) (112), *Clostridium acetobutylicum* (NifB_{Ca}) (113, 114), *M. infernus* (NifB_{Mi}) (115), *Methanosarcina acetivorans* (NifB_{Ma}) (116), *Methanobacterium thermoautotrophicum* (NifB_{Mt}) (116), and *C. tepidum* (NifB_{Ct}) (117), a number of conserved motifs were identified in the SAM-radical domain including an HPC motif, the AdoMet motif (Cx₃Cx₂C) common to all SAM-radical proteins, an ExRP motif, an AGPG motif, a TxTxN motif and a Cx₂CRxDaxG motif. Putative NifB proteins that did not present all these motifs

were eliminated from the dataset, which was then reduced by 28% down to 289 sequences (Table 3 and Figure 19).

As shown in Figure 19, the most widely occurring NifB domain architecture consists of an N-terminal SAM-radical domain (green dot) linked to a C-terminal NifX-like domain (red dot). This protein configuration accounted for 73% of NifB sequences in the Bacteria domain of the curated database. This configuration has been proposed to emerge after an ancestral gene fusion event (85, 139). The functionality of this NifB subfamily has been demonstrated *in vivo* in many bacteria and *in vitro* in the cases of NifB_{Av} and NifB_{Ko} (85, 112). A second NifB subfamily that included an additional NifN-like domain (blue dot) was found in 6 NifB sequences in the Bacteria domain (2.4% of the curated database). This NifB subfamily was first described in Clostridia (113) and then proven functional in *A. vinelandii* carrying an engineered NifN-B fusion that mimicked the Clostridium protein (114). Finally, a stand-alone SAM-radical domain was found in 104 NifB sequences, accounting for a 100% NifB in the Euryarchaeota and 24% in the Bacteria domain. The functionality of this NifB subfamily was not demonstrated at the time this thesis was being carried out. Importantly, the *Clostridium* genus of the Firmicutes phylum is unique in that it contains all three NifB architectures. The curated NifB database contains 45 Firmicutes species likely to be diazotrophic organisms. Among these, 55% carry the stand-alone SAM-radical domain, 33% carry the two-domain architecture, and 13% carry the three-domain architecture.

4.1.2. Organization of *nif* genes in the genome of *M. infernus*

The Archaea *M. infernus* *nif* gene cluster reveals differences and similarities to bacterial *nif* clusters (Figure 20). The order of the *nifH*, *nifD*, and *nifK* genes is similar to bacteria but the presence of two *nifI* genes between *nifH* and *nifD* is unique to methanogens. The *nifI* gene products are *glnB* homologs that have regulatory functions. Nitrogenase post-transcriptional regulation in Archaea and Bacteria are very similar (27). However the regulation of *nif* transcription in methanogens occurs by repression rather than by NifA activation. Importantly, genes with similarity to *nifS*, *nifU*, *nifH* and *nifB* have been found in non-diazotrophic Archaea suggesting that the presence of these genes alone does not imply a function in N₂ fixation. The

absence of *nifN*, but not *nifE* in *M. infernus* is striking as they are supposed to evolve from duplication of a *nifDK* cluster and their products usually form heterotetramers.

Figure 19. Occurrence of NifB architectures in diazotrophs. Top: Overall distribution and relative frequency of NifB architectures in putative diazotrophs. Green dot represents the SAM-radical domain; red dot represents the NifX-like domain; blue dot represents the NifN-like domain. **Bottom:** Scheme of the three different NifB architectures and representative species known to carry them. The SAM-radical is shown in green, the NifX-like domain in red, and the NifN-like domain in blue. Color dots represent motifs strictly conserved in the SAM-radical domain. The scale bar represents 100 amino acids.

4.2. NifB_{Mi} is functional *in vivo* when expressed in the aerobic mesophilic host *A. vinelandii*

Genetic complementation analysis was performed by expressing a synthetic codon-optimized *nifB_{Mi}* gene in the *A. vinelandii* UW140 ($\Delta nifB$) strain under the control of the *nifH* promoter (Figure 21). *A. vinelandii* is a strict aerobe with optimum growth temperature of 30°C and was used here to provide a test of NifB functionality. This work was carried out by Emilio Jiménez but is included in this thesis because it is unpublished material of relevance to understand the context of the *in vitro* work described below.

Figure 21. Schematic diagram of the incorporation of *nifB_{Mi}* into the *A. vinelandii* UW140 ($\Delta nifB$) genome through homologous recombination. The gene expression cassette had a *nifB* gene under the control of a *nifH* promoter and was placed in the so-called D sequence (downstream of the *vnfY* gene).

Strain UW418 ($\Delta nifB$, *P_{nifH}::nifB_{Mi}*) exhibited diazotrophic growth both in solid and liquid culture media (Figure 22), in contrast to the Nif⁻ phenotype of the parental strain UW140 ($\Delta nifB$). Diazotrophic growth rates ($\ln 2/t_d$) of UW418 are shown in Table 5 together with the rates of the wild type and UW140. This data shows that, although NifB_{Mi} originates from a strict anaerobic and thermophilic microbe, it was functional and could complement the *A. vinelandii* $\Delta nifB$ mutant phenotype. The simplest explanation is that gene product of *nifB_{Mi}* is able to synthesize *in vivo* the FeMo-co precursor and support diazotrophic growth of the aerobic, mesophilic microorganism *A. vinelandii*. However, it is important to note that recombinant

NifB_{Mi} did not support a growth rate at 30°C similar to that of the *A. vinelandii* wild type strain, possibly due to the almost 55°C difference in optimal growth temperature between *M. infernus* (140) and *A. vinelandii*. No differences in growth rate among tested strains were observed when using NH₄⁺ as nitrogen source (Table 5).

Figure 22. Genetic complementation of *A. vinelandii* $\Delta nifB$ strain with $nifB_{Mi}$. (A) Petri dishes with solid Burk media showing growth of *A. vinelandii* DJ (wild-type strain, WT), UW140 ($\Delta nifB$), UW418 ($\Delta nifB$, $P_{nifH}::nifB_{Mi}$) and UW422 ($\Delta nifB$, $P_{nifH}::nifB_{Ct}$) in the presence (left plate) or absence (right plate) of ammonium as nitrogen source. $nifB_{Ct}$ stands for $nifB$ from the thermophilic bacterium *Chlorobium tepidum*. Strain UW422 is out of the scope of this study and will not be discussed here. (B-C) Growth curves of DJ, UW140, and UW418 strains using ammonium (B) or N₂ (C) as source of nitrogen. Data in (C and D) represent means $\pm$ SD ($n \geq 2$).

Azotobacter Strain	N ₂	NH ₄ ⁺
DJ (Wild type)	0.23	0.31
UW140 ($\Delta nifB$)	<0.001	0.29
UW418($\Delta nifB$, $P_{nifH}::nifB_{Mi}$)	0.015	0.30

Table 5. Diazotrophic growth rates using NH₄⁺ or N₂ (diazotrophic conditions) as nitrogen source. Growth rates are calculated as ln2/td.

4.2.1. Heterologous expression and purification of wild type NifB from *Methanocaldococcus infernus* (NifB_{Mi}) and the site-directed variants SM, AM and DM

Overexpression of Fe-S proteins in *E. coli* usually yields partially loaded enzymes due to incomplete incorporation of [Fe-S] clusters. To increase [Fe-S] cluster incorporation, [Fe-S] cluster biosynthesis machinery can also be overexpressed (141). We have used plasmid pRSFIcsDuet-1 (kindly provided by Juan Fontecilla) for the overexpression of *E. coli* ISC machinery and SAM synthase encoding gene *metK*. This plasmid has been successfully used before to overexpress radical-SAM proteins in *E. coli*. Overexpression of proteins from thermophilic organisms in *E. coli* can also be difficult due to codon bias differences. Although the genetic code is strictly conserved in all living organisms, the proportion of each tRNA is different in different organisms, which may prove to be a limiting factor in the overproduction of a protein whose gene contains "rare" codons in *E. coli*. For this reason, synthetic codon optimization was performed to achieve the highest possible level of for NifB_{Mi} production.

E. coli cells co-transformed with plasmids pRHB509 and pRSFIscMetK Duet-1 were used to inoculate 10-liter cultures. Time-course expression of NifB_{Mi} in combination with MetK and Isc proteins is shown in Figure 23. Although higher amount of protein was obtained through induction with 1 mM IPTG (lanes 12-14), the major part of it was insoluble (data not shown). Indeed the excess of IPTG triggers rapid protein accumulation and increases the chance of yielding unfolded enzymes. Therefore in the optimized NifB_{Mi} overexpression protocol, the culture was induced with lactose to decrease expression rates. The optimized induction protocol consisted of induction by the addition of lactose and incubation for 8-9 hours at 26°C under low bubbling conditions that resulted in O₂ depletion from the medium. In these conditions NifB_{Mi} was found mostly in soluble form.

Figure 23. Co-expression of NifB_{Mi} with MetK and Isc proteins. Lane 1, molecular weight markers; lane 2, whole cell protein profile of a non induced culture; lanes 3-5, time course of lactose induction in 1 hour increments; lanes 6-8, time course of IPTG induction (50 μ M) in 1 hour increments; lanes 9-11, time course of IPTG induction (100 μ M) in 1 hour increments; lanes 12-14, time course of IPTG induction (1 mM) in 1 hour increments. Expected molecular weights for each overexpressed protein are shown to the right.

A DNA sequence encoding a poly-histidine tag was added to the 5' terminal end of *nifB_{Mi}* when constructing the expression plasmid. Although a potential risk exists for the formation of artifacts, such as unspecific binding of [Fe-S] clusters to the polyhistidine tag, we proceeded with this strategy to achieve rapid purification by IMAC, thus facilitating the purification process and increasing the amount and quality of pure protein. We also took advantage of NifB_{Mi} thermostability and subjected the *E. coli* crude extract to a step of heating at 75°C for 20 min to eliminate as many contaminants as possible previous to the IMAC. At this temperature the contaminating proteins of *E. coli* were denatured and could be removed by centrifugation. The use of a long polyhistidine tag (10 His) resulted in increased purity due to the ability to use more stringent washing (250 mM imidazole) in IMAC without apparently altering protein activity. The NifB_{Mi} protein eluted at a fairly wide peak in the imidazole gradient. Elution fractions showing purest and more concentrated NifB_{Mi} were pooled, concentrated by centrifugation in Centricon devices and subject to desalting chromatography on HiPrep 16/60 column.

SDS-PAGE analysis showed a protein band with estimated molecular weight of 37.5 kDa and purity > 95% (Figure 24). A similar protocol was used to purify three NifB_{Mi} variants generated by site-directed mutagenesis: a variant in which the Cys residues of the signature SAM-binding motif were replaced by Ala residues (SM variant), a variant in which the Cys residues of a conserved Cys-rich motif present at the C-terminal end of the protein were replaced by Ala residues (Auxiliary or AM variant), and a variant carrying both changes (Double or DM variant). Typically, 150 mg of pure protein were obtained from 10 liters of culture expressing either wild type or any of the variant proteins (Table 6), which was satisfactory for carrying out the biochemical characterization, crystallization and spectroscopic assays.

Figure 24. Purification of NifB_{Mi} wild type and site-directed variants. (A) SDS-PAGE analysis of IMAC purification steps. CFE, cell-free extract; SN, supernatant; FT, flow through the IMAC column; W1-W3, fractions containing low imidazole; E, eluted fraction (B) SDS-PAGE analysis of purified preparations of NifB_{Mi} wild type (WT) and the site-directed variants SM, AM and DM. About 7-10 μ g of protein were loaded per lane.

NifB protein variant			
NifB _{Mi} WT	NifB _{Mi} SM	NifB _{Mi} AM	NifB _{Mi} DM
150 mg	130 mg	140 mg	100 mg

Table 6. Typical amount of pure proteins obtained from 10 liter of *E. coli* BL21 DE3 cultures expressing the different NifB_{Mi} variants.

4.3. Elution profile of NifB_{Mi} wild type on Superdex 200 column used for crystallography and spectroscopic studies.

To obtain homogenous protein preparations suitable for crystallographic trials, purified NifB_{Mi} samples were subjected to an *in vitro* [Fe-S] cluster reconstitution procedure followed by desalting to remove excess Fe and S. Immediately after desalting, reconstituted samples were repurified on a Superdex-200, equilibrated with 50 mM Tris-HCl, pH 8, 200 mM KCl, 10 % glycerol, and 5 mM β -mercaptoethanol. Under these conditions, all NifB_{Mi} proteins, WT and site-directed variants, eluted mostly as a single peak at 82-87 ml which was collected and stored in liquid nitrogen (Figure 25). Calibration of the Superdex-200 column with proteins of known molecular weight confirmed that this peak corresponded to the monomeric form of NifB_{Mi} (Table 7). A small amount of dimeric form (18%) was also present, consistent with the result TOF/TOF analysis (Figure 26).

Figure 25. Elution profile of NifB_{Mi} on Superdex 200 column. The color lines corresponds to UV-visible absorbance wavelengths tracked during chromatography: 280 nm (blue), 400 nm (red), and 600 nm (green). Purified NifB_{Mi} proteins (1.25 mg in 500 μ l sample) were injected onto a preparative Superdex 200 column previously equilibrated in 50 mM Tris-HCl buffer pH 8, 200 mM KCl, and 5 mM DTT. The chromatography was carried out at a flow rate of 1 ml/min using AKTA Purifier.

Sample	Elution volume (ml)	K _{av}	Measured MW (Da) ^a	Theoretical MW (Da)	Probable oligomer state
Aprotinin	103.75	0.7930	6,843	6,500	
Ribonuclease A	95.21	0.6843	13,559	13,700	
Carbonic anhydrase	87.27	0.5832	25,613	29,000	
Ovalbumin	78.71	0.4742	50,849	43,000	
Conalbumin	74.74	0.4237	69,865	75,000	
NifB _{Mi} WT	85.29	0.5581	29,995	37,543	monomer
NifB _{Mi} SM variant	83.59	0.5364	34,383	37,446	monomer
NifB _{Mi} AM variant	85.25	0.5575	30,108	37,478	monomer
NifB _{Mi} DM variant	85.30	0.5582	29,976	37,382	monomer

Table 7. Estimated molecular weights of WT NifB_{Mi} and the site-directed variants AM, SM and DM. Table includes the molecular weight standards used to calibrate the Superdex 200.^a K_{av} = 2.1967 - 0.366 * log MW; R² = 0.989. Column geometric volume = 120 ml. Column void volume = 41.46 ml.

Figure 26. TOF/TOF analysis of NifB_{Mi}. Analysis was carried out in a 4800 Proteomics Analyzer. NifB_{Mi} yielded a main peak corresponding to a molecular weight of 37,483 Da and a second peak corresponding to a molecular weight of 75,180 Da that could account for a NifB dimer.

In the next chapter we will present studies and results related to the biochemical characterization of the wild type and mutant forms of NifB_{Mi}.

Chapter II: Biochemical characterization of NifB_{Mi} WT and the site-directed variants

4.4. UV-visible absorption and metal content of NifB_{Mi}

Purified NifB_{Mi} variants could be highly concentrated without evidence of precipitation. All of them presented intense dark color suggesting that the metal centers of NifB_{Mi} were quite stable (also in the mutant versions). In low concentration samples, all purified NifB_{Mi} proteins (wild type and variants) exhibited similar UV-visible spectra. The UV-visible absorption spectrum, recorded for [Fe-S] cluster reconstituted NifB_{Mi} WT is presented, as an example, in Figure 27. In this spectrum, a single peak at 280 nm is observed which corresponds to the absorption of the aromatic residues of the polypeptide chain. The spectrum also shows a shoulder at 320 nm and a broad peak at 420 nm. This absorption corresponds to the region of the charge transfer bands $S \rightarrow Fe(III)$ that are characteristic of the presence of [Fe-S] centers. Addition of DTH reduced the [Fe-S] centers present in NifB_{Mi} rapidly decreasing the UV absorbance between 300 nm and 500 nm. This response indicated that NifB_{Mi} [Fe-S] clusters were redox responsive.

Figure 27. UV-visible absorption spectra of reconstituted NifB_{Mi}. The [Fe-S] clusters of protein samples (10 μ M monomer) were chemically reconstituted as described in Materials and Methods. UV spectra were recorded at 1, 5 and 25 minutes after the addition of 5 mM DTH.

In order to determine the average metal content of the NifB_{Mi} wild type, we performed purifications under anaerobic and aerobic conditions. Aerobically purified NifB_{Mi} presented a Fe and Sulfur content of 0.8 ± 0.1 and 1.2 ± 0.3 whereas apo-NifB_{Mi} had negligible levels of Fe and S. These results demonstrate that also NifB_{Mi} is O₂ sensitive and when exposed to air, is able to preserve a limited Fe and S content. In contrast, the anaerobically “as isolated” form of NifB_{Mi} contained an average of 3.3 ± 0.1 Fe atoms and 4.0 ± 0.2 sulfur atoms per monomer (Figure 28B), much less than expected from comparison to other NifB proteins purified from different sources, for example *K. oxytoca*.

A

B

	Fe	S
As isolated	3.3 ± 0.1	4.0 ± 0.2
Reconstituted with Fe and S	9.2 ± 0.3	8.0 ± 0.2
apo-NifB_{Mi}	< 0.1	< 0.1
apo-NifB _{Mi} reconstituted with Fe and S	10 ± 0.2	11.2 ± 0.4
Aerobically purified NifB_{Mi}	0.8 ± 0.1	1.2 ± 0.3

Figure 28. A) UV-visible absorption spectrum of NifB_{Mi} wild type. Black line: Apo form obtained by treatment with 10 mM EDTA to remove all Fe bound to the protein; Green line: protein purified under aerobic conditions; Blue line: as isolated protein purified under anaerobic conditions; Red line: protein after reconstitution of [Fe-S] clusters with an excess of DTT, Fe and S. Spectra were recorded for protein samples (45 μ M NifB_{Mi}) in 50 mM Tris-HCl, pH 8, 200 mM KCl, 10 % glycerol, and 5 mM β -mercaptoethanol with a 1 cm optical path. B) Fe and S content of the different NifB_{Mi} wild type forms. Data represent the mean $\pm$ SD.

Thus, despite co-overexpression of *E. coli* ISC operon and purification of the NifB_{Mi} protein under anaerobic conditions, the incorporation of the [Fe-S] centers was not complete. One possible explanation could be that some NifB_{Mi} molecules had lost their [Fe-S] centers during purification, probably during the heating step outside the glove compartment, despite the use of airtight tubes. In order to have homogeneous preparations, we had to resort to chemical *in vitro* reconstitution of [Fe-S] centers. NifB_{Mi} wild type contained 9.2 ± 0.3 Fe atoms and 8.0 ± 0.2 S atoms per NifB_{Mi} after reconstitution and its UV-visible spectrum (Figure 28A) was consistent with

successful reconstitution of [Fe-S] centers since the 315/280 nm ratio increased from 0.38 in the as isolated sample to 0.6 in the reconstituted sample. This result was consistent with the presence in wild type NifB_{Mi} of, at minimum, two [Fe₄S₄] clusters in addition to the SAM-binding [Fe₄S₄] cluster. In addition, to remove partially degraded clusters and increase sample homogeneity, apo-NifB_{Mi} (enzyme lacking its metal cofactors) was prepared from anaerobically purified protein and used in [Fe-S] cluster reconstitution experiments. These samples exhibited the highest Fe and S content and were used for further spectroscopic analysis (10 ± 0.2 Fe atoms and 11.2 ± 0.4 S atoms per NifB_{Mi}, respectively). In the course of our work, a competing team published in 2015 (116) a preliminary characterization of NifB proteins from two hyperthermophilic organisms (*M. acetivorans* and *M. thermoautotrophicum*) obtaining similar results for Fe and S content of wild type NifB after reconstitution of its [Fe-S] clusters *in vitro*.

4.4.1 Site-directed NifB_{Mi} variants

As one of the objectives of this research was to identify the auxiliary [Fe-S] clusters within NifB_{Mi} and describe their function, it was fundamental to quantify the Fe and S content of the generated NifB_{Mi} mutant variants (Table 8) and to establish their UV-visible absorption properties (Figure 29). Comparison of NifB amino acid sequences revealed the presence of a number of conserved Cys residues that could potentially coordinate multiple [Fe-S] clusters. Three of these residues belong to the Cx₃Cx₂C motif responsible for the coordination of the catalytic [Fe₄S₄]-radical SAM center. Substitution of these Cys residues (C36A/ C40A/C43A) in variant SM resulted in reduction of Fe and S content, changes in the UV-visible spectrum (Figure 29), and inability to activate apo-NifDK (see Figure 34 below). In addition, we hypothesized that two Cys residues conserved at the C-terminal end of the protein could be ligands for a second center of the [Fe₄S₄] or [Fe₂S₂] type, and these were mutated as well to generate the AM variant.

Name	Variant	Substitutions
WT	Wild type	none
SM	SAM motif variant	C36A/C40A/C43A
AM	Auxiliary motif variant	C260A/C263A
DM	Double variant	C36A/C40A/C43A C260A/C263A

Table 8. NifB_{Mi} variant nomenclature

A third variant combining both mutations and expected to be heavily impaired in [Fe-S] cluster coordination was also generated and called DM. UV-visible spectra overlay of as isolated and reconstituted forms of all NifB_{Mi} variants, normalized to 280 nm, is shown in Figure 29 to enable visual comparison of their differences in the 320-500 nm region. The spectra of all NifB_{Mi} variants show absorbance maxima at 280 nm, 315 nm, 413 nm a distinct plateau leading from 380 nm to 450 nm. Comparison of the absorbance at 315 nm and 280 nm showed a marked decrease in the 315 nm maxima for the Double mutant (DM) (0.24) that was relatively increased after the metal cluster reconstitution (0.43) and a smaller decrease for SAM mutant (0.45) that remained practically constant after the reconstitution (0.54) (Figure 29, bottom chart). The 413nm/280nm ratio showed little variation between the reconstituted forms wild type (0.40) and the SAM mutant (0.42). In contrast, if we compare ratio variations among as isolated and reconstituted forms for SAM mutant and wild type, the wild type demonstrates a significant increase in the ratio (0.25 to 0.40) compared with the SM mutant (0.36 to 0.42).

Comparing the data obtained by the UV-visible absorption spectra and Fe/S quantification, the auxiliary domain mutant (AM) seems to maintain at least one [Fe₄S₄] center. Indeed the characteristic UV absorption (Figure 29), that remain constant after the reconstitution and the Fe/S amount quantifications, demonstrate that there is not any significant metal content difference between the reconstituted and as isolated forms of AM (As isolated Fe/S content of 5.3/6.3 versus reconstituted Fe/S content of 5.5/6.4). Among the three mutants, the double mutant (DM) retains the lesser amount of Fe/S atoms, 2.6 ± 0.2 and 3.5 ± 0.3 respectively, lower than expected if the protein were to coordinate a single [Fe₄S₄] cluster. Consistently, the DM variant exhibits minimum shoulder at 320-400nm. After *in vitro* [Fe-S] cluster reconstitution, DM preparations exhibited UV-visible features as well as Fe and S content consistent with the presence of one [Fe₄S₄] cluster, (Figure 29B), with amount of Fe/S atoms of 4.6 ± 0.4 and 5.5 ± 0.3 respectively. These results reinforce the hypothesis of the presence in NifB_{Mi} of at least one additional [Fe-S] cluster-binding motif (that we designate as AC2), in addition to the [Fe₄S₄]-SAM binding motif (SM) and the C-terminal auxiliary motif (AC1). Thus, the increase of [Fe-S] content and 413nm/280nm ratio in the reconstituted form of DM might correspond to the formation of a full complement of the remaining auxiliary secondary cluster (AC2).

As isolated	Fe	S
WT	3.3 ± 0.1	4.0 ± 0.2
SM	3.3 ± 0.3	4.5 ± 0.4
AM	5.3 ± 0.4	6.3 ± 0.1
DM	2.6 ± 0.2	3.5 ± 0.3
Reconstituted with Fe and S		
WT	9.2 ± 0.3	8.0 ± 0.2
SM	6.1 ± 0.3	7.4 ± 0.4
AM	5.5 ± 0.2	6.4 ± 0.2
DM	4.6 ± 0.4	5.5 ± 0.3

Figure 29. UV-visible absorption spectra and Fe and S content of NifB_{Mi}. Wild type (Red line), Double mutant (Blue line), Auxiliary mutant (Green line) and SAM mutant (Black line) A) “As” isolated and B) “Reconstituted” . Protein (45 μM) recorded with a 1 cm optical path; To the right: Fe and S content of as-isolated and reconstituted NifB_{Mi} purified under anaerobic conditions. Data represent mean ± SD. Bottom: Bar chart shows ratios of 315 nm / 280 nm, 413 nm / 280 nm and 413 nm / 315 nm for relative comparison between maxima of the proteins. All peaks were normalized to 280 nm values.

4.5. NifB_{Mi} enzymatic activity

In order to evaluate the capacity of the hyperthermophilic NifB_{Mi} to synthesize the precursor of FeMo-co and the impact of its standalone SAM Radical architecture on functionality, we performed the NifB-dependent *in vitro* FeMo-co synthesis assay, using either biochemical complementation of cell-free extracts from an *A. vinelandii* $\Delta nifB$ mutant strain or the minimal and composition-defined reaction mixtures previously described by Curatti (69). The composition-defined mixtures contained a fixed amount of NifB_{Mi} protein plus apo-NifEN (NifB-co lacking NifEN), apo-NifDK (FeMo-co lacking NifDK), NifH, NafY, and NifX, all of them in purified form, and aims to reconstitute active Mo-nitrogenase that is further tested by the acetylene reduction assay.

Importantly, when the experiments of this work were performed, there was no a body of results from previous studies carried out with the *A. vinelandii* NifB (69) demonstrating that the addition of specific molecules as SAM, Fe, S, Mo, and the reducing agent DTH, were fundamental for the NifB-dependent *in vitro* synthesis of FeMo-co.

4.5.1. Preliminary study of wild-type NifB_{Mi} activity at different temperatures using the UW140 ($\Delta nifB$) crude extract complementation assay

The first experiments were conducted to test at least a minimum capability of the heterologous NifB_{Mi} to carry out NifB-co synthesis in *in vitro*, and complement the missing NifB activity of the *A. vinelandii* UW140 ($\Delta nifB$) cell-free extract. This “classical” *in vitro* assay is constituted of two different phases: *i*) FeMo-co synthesis and insertion into apo-NifDK to reconstitute holo-NifDK (hereafter simply called NifDK), and *ii*) NifDK activity as determined by the reduction of acetylene to ethylene in the presence of NifH and an ATP-regenerating mixture. The function of NifB is needed in the first phase, where the strain UW140 lacks NifB-co activity but provides, molybdenum, homocitrate, and the rest of the protein components required for FeMo-co synthesis and apo-NifDK activation. The assay more commonly utilized as positive control for the NifB-dependent FeMo-co synthesis and insertion assay, typically involves the addition of purified NifB-co to the UW140 extract. NifB-co is isolated from *K. oxytoca* cells lacking *nifE* or *nifN* genes, which makes them unable to

further process NifB-co into FeMo-co. When *in vitro* FeMo-co synthesis occurs, de novo-synthesized FeMo-co is incorporated into apo-NifDK to reconstitute active NifDK. In the second step, NifDK activity is estimated as described above.

It is important consider that the assay developed by Curatti (69), designed to provide direct biochemical evidence of the minimal reaction requirements for the synthesis *in vitro* of FeMo-co, was performed at the optimal temperature for protein from the mesophilic *A. vinelandii* (30°C). Determining the optimal condition where NifB_{Mi} can express its activity in the synthesis of NifB-co and guarantee the structural and functional stability of the other Nif proteins was a critical point of this work. As shown in Figure 30 we tested the capacity of the “as isolated” and “reconstituted” forms of NifB_{Mi} to enable FeMo-co synthesis in the UW140 crude extract. The reactions mixtures were supplemented with 800 μ M SAM, 500 μ M FeSO₄, 500 μ M Na₂S, 17 μ M molybdenum, 17 μ M homocitrate and ATP mix. Four different temperatures (30°C, 40°C, 50°C, 60°C) were tested for the FeMo-co synthesis step in order to accommodate the differences in activity and stability between NifB_{Mi} and the rest of the proteins required for FeMo-co synthesis. Temperature in the NifDK activity determination phase was maintained at 30°C. The results obtained from the *in vitro* assays, shown the highest activity at 50°C for the sample complemented with reconstituted NifB_{Mi} (1.2 ± 0.1). This activity was similar to that of the positive control obtained by adding purified NifB-co to the reaction mix. Conversely the as isolated NifB_{Mi} form seemed unable to maintain its activity at this temperature. As discussed in the previous chapter, despite the presence of *E. coli* Isc cluster in the overexpressing strain, this result suggests that the Fe content of as isolated NifB_{Mi} corresponds to a group of partially assembled and not totally functional [Fe-S] clusters. Surprisingly NifDK exhibited the activity at 50°C than 30°C (2319 ± 29 nmol of ethylene·min⁻¹·mg⁻¹ NifDK). Taking these data into consideration, we chose 40°C as optimal temperature for the FeMo-co synthesis and holo-NifDK formation phase in the *in vitro* assay with *A. vinelandii* cell-free extracts.

Figure 30. Temperature dependence of the *in vitro* complementation assay of UW140 (*AnifB*) with NifB_{Mi}. Two hundred µl of UW140 cell-free extract (1.4 mg/ml) were used in each assay.: FeMo-co synthesis and NifDK formation (step 1) conditions: 45 min incubation at 30°C, 40°C, 50°C, or 60°C. NifDK acetylene reduction activity (step 2) conditions: 30 min incubation at 30°C. UW140 (blue line) UW140 + as isolated WT NifB_{Mi} (10 µM) (red line), UW140 + reconstituted WT NifB_{Mi} (10 µM) (green line), UW140 + NifB-co (4.47µM Fe) (purple line). Specific activity of NifDK (0.043 mg) under the same reaction conditions was calculated as control of the effect of buffer and temperature on NifDK activity (gray columns and scale bar to the right).

4.5.2. *In vitro* NifB_{Mi} assays with different levels of substrate and UW140 proteins

Although the presence of SAM stimulates FeMo-co synthesis, the excess of this substrate can also inhibit the enzymatic activity of NifB (Table 9). It is interesting to consider that, in previous work, the addition of the nonreactive analog S-adenosyl homocysteine (SAH) was shown to inhibit the reaction and that addition of excess SAM relieved SAH inhibition (69). The specific requirement for SAM was previously demonstrated by using desalted cell-free extracts lacking low molecular weight compounds. However, as the *in vitro* experiments with UW140 extracts performed here were carried out to merely to demonstrate the capacity of NifB_{Mi} to provide NifB-co compatible with the *A. vinelandii* Nif machinery, we decided not to use desalted extracts.

The initial *in vitro* complementation studies were performed with limited amounts of UW140 cell-free extracts (Figure 30). These assays showed that very low protein concentrations in cell-free extracts, which provided limited amounts of Nif proteins could result in very low activities, probably by decreasing the chance of protein-protein interactions or by decreasing protein stability.

FeMo-co synthesis and holo-NifDK formation (40°C)	NifDK activity (nmol C ₂ H ₄ x min ⁻¹ assay ⁻¹)
Assay 1	Crude extract (1.4 mg/ml)
-SAM, -Fe ²⁺ , -S ²⁻	0.20 ± 0.1
-SAM	1.02 ± 0.07
-Fe ²⁺ , -S ²⁻	1.5 ± 0.1
Complete (SAM 0.5 mM)	1.03 ± 0.2
Complete (SAM 1 mM)	2.26 ± 0.1
Complete (SAM 2 mM)	1.61 ± 0.3
Complete (SAM 3 mM)	1.53 ± 0.1
Complete (SAM 6 mM)	1.33 ± 0.2
-NifB, -SAM, -Fe ²⁺ , -S ²⁻	ND
+NifB-co (4.4 µM Fe)	1.56 ± 0.09
Assay 2	Crude extract (4.4 mg/ml)
Complete (SAM 1 mM)	7.4 ± 0.4
+NifB-co (8.55 µM Fe)	15.22 ± 0.4

Table 9. Effect of SAM, Fe, and S on the *in vitro* different substrates concentrations. The acetylene reduction activity of reconstituted NifDK was used as proxy for FeMo-co synthesis (125). Temperature was maintained at 40°C for FeMo-co synthesis and holo-NifDK formation. Data are the average of two independent determinations. ND, not determined.

Using 3-fold concentrated UW140 cell-free extracts (4.4 mg/ml) significantly increased NifDK reconstitution and acetylene reduction activity without increasing

the amount of NifB_{Mi} (Table 9). This result, in combination with the positive control result (UW140 + NifB-co), where the NifB-co is at a final concentration of 8.55 μ M, support the hypothesis that UW140 crude extract could represent a limiting factor in the *in vitro* assay and stimulated us to continue the analysis using only purified Nif proteins.

4.5.3. NifB_{Mi} dependent FeMo-co synthesis and NifDK activation assays in a system with purified Nif proteins

An important aspect of performing complete *in vitro* FeMo-co biosynthesis and nitrogenase reconstitution using only purified proteins, and not crude cell extract, is the possibility of addressing whether or not a particular protein has direct participation in the assembly of FeMo-co. Previous work regarding the *in vitro* assay complementation with pure Nif proteins (69) highlighted how the protein components involved in this process, could be divided in three different groups. The first group is constituted by the NifB, NifEN, and NifH proteins, which are essential for the complete *in vitro* synthesis of FeMo-co. The second group included the NifX and NafY protein that participate actively in FeMo-co biosynthesis or insertion but that were not essential.

Figure 31. SDS-PAGE of purified FeMo-co biosynthesis component enzymes used in the *in vitro* FeMo-co synthesis assay. In the upper part is represented the major Nif cluster and the corresponding oligomer state for each protein. About 7-10 μ g of protein were loaded for NifH and NafY, 3 μ g for NifDK, and 20 μ g for NifX.

A third different group included proteins that provided the chemical substrates for FeMo-co synthesis. This group included NifU and NifS providing [Fe-S] cluster precursors, NifQ providing Mo and NifV providing homocitrate. These proteins can be easily substituted in the *in vitro* assay by the addition of substrate's excess and were not tested here. Purified preparations of proteins in the first two groups (with the absence of NifB) are shown in Figure 31 along with their proposed oligomer states and the corresponding location of their encoding genes in the *nif* genes cluster.

In the previous experiments with cell-free extracts, failure to obtain high acetylene reduction activities could be due to two main reasons: 1) the low concentration of Nif proteins in the crude extracts could limit substrate availability and protein-protein interactions needed for the synthesis of the NifB-co intermediates and the NifB-co trafficking between NifB and apo-NifEN, and later between apo-NifEN and apo-NifDK; 2) The constant temperature (40°C) used during the FeMo-co synthesis and NifDK reconstitution phase could be, on one side, affecting the stability of *A. vinelandii* Nif proteins and, on the other side, impairing the kinetics of the hyperthermophilic NifB_{Mi} protein during NifB-co synthesis. For these reason, we decided to modify the original *in vitro* assay, and included an additional pre-step, defined as "NifB_{Mi} Activation". During this phase, NifB_{Mi} was incubated at different temperatures (30°C, 40°C, or 60°C) with 5 mM DTH, 1 mM SAM, 500 µM FeSO₄, and 500 µM Na₂S prior addition of the rest of purified Nif proteins. The others two steps were maintained as in the original protocol.

The results presented in Figure 32 indicate that the pre-incubation step at of NifB_{Mi} with excess of SAM, Fe/S and the reducing agent DTH improved NifB-dependent FeMo-co synthesis. At 60°C, the acetylene reduction achieves the highest level after an incubation of 30 minutes (10.15 nmol C₂H₄ formed per min⁻¹ x ml⁻¹) 5-fold higher than the activity obtained by pre-incubating NifB_{Mi} at 30°C (1.8 nmol C₂H₄ formed per min⁻¹ x ml⁻¹). A pre-incubation of 15 minutes seem enough to exponentially increase NifB_{Mi} activity. Although it is likely that NifB_{Mi} reaches maximum activity near 90°C, we did not test temperatures above 60°C in our assays.

4.5.4 Characterization of the essential elements for the *in vitro* NifB_{Mi} dependent FeMo-co synthesis assay with purified Nif proteins

A reaction containing NifB-co in place of NifB_{Mi} (+NifB-co) was carried out as control of pathway functionality downstream NifB. Protein and substrate requirements were analyzed by removing one or more substrates/proteins at a time from the reaction mixtures, as indicated in the Figure 33. In this assay the activity refers to nmol ethylene formed per min and mg of NifDK. According with the results previously obtained from the temperature dependence analysis of NifB_{Mi} kinetics, the *in vitro* complementation assays with pure Nif proteins consisted of a 30 minutes and 60°C NifB_{Mi} pre-incubation phase (that in the complete version contained Fe, S, SAM, DTH, and NifB_{Mi}) followed by a 30°C FeMo-co synthesis and apo-NifDK reconstitution phase (that in the complete version contained Mo, homocitrate, DTH, NifX, apo-NifEN, NifH, NafY, and apo-NifDK).

Figure 32. Temperature dependence of NifB_{Mi} activity in the pre-activation phase. WT reconstituted NifB_{Mi} was incubated for 30 min at different temperatures prior the addition of pure Nif proteins. Reaction mixtures (0.2 ml) contained: 5 mM DTH, 1 mM SAM, 500 μ M FeSO₄, and 500 μ M Na₂S. Yellow line, incubation at 30°C; Blue line, incubation at 40°C; Green line, incubation at 60°C. Concentrations of Apo-NifEN, NifH, apo-NifDK, NifX and NafY are described in Material and Methods. The final concentration of NifB_{Mi} in the FeMo-co synthesis and NifDK reconstitution phase was 10 μ M. Data are the average of two independent determinations.

Consistent with all experiments reported above, apo-NifDK activation was strictly dependent on addition of NifB_{Mi} or NifB-co. Addition of either apo-NifB_{Mi} or aerobically purified NifB_{Mi} did not activate apo-NifDK (data not shown). Both as isolated and [Fe-S] cluster reconstituted NifB_{Mi} forms had capacity to synthesize a limited amount of NifB-co in the absence of Fe/S supplement (6.5 nmol ethylene formed⁻¹ mg⁻¹ of NifDK), although this activity was much lower in the as isolated form. These results underline how reconstituted [Fe-S] clusters in the protein were able to operate as precursor clusters for the formation of NifB-co. Conversely, the absence of SAM represented a clear substrate deficit, and its addition to the *in vitro* system with purified components was essential for NifB_{Mi} activity.

Figure 33. Protein and substrate requirements for NifB_{Mi}-dependent *in vitro* reconstitution of apo-NifDK. Detailed assay conditions are described in the Material and Methods. Activities are shown in nmol of ethylene·min⁻¹·mg⁻¹ NifDK. Data are the mean ± SD of at least two independent experiments.

As it was the case with the *A. vinelandii* NifB (69), the NifEN and NifH proteins were essential elements for NifB_{Mi} dependent *in vitro* FeMo-co synthesis (data not shown). Other proteins involved in NifB-co or FeMo-co trafficking, such as NifX and NafY, although not essential for the *in vitro* activation of apo-NifDK, influence positively the formation of a holo-NifDK (Figure 33). Indeed the FeMo-co mobilization from

NifEN to apo-NifDK mediated by NafY seems to play a relevant role and in the absence of this protein, the formation of NifDK was significantly reduced (50.2 nmol ethylene formed⁻¹ mg⁻¹ of NifDK). This result confirms how FeMo-co and its precursors are extremely unstable and sensitive to oxidation (142, 69). Overall, these results support a model in which NifB_{Mi} assembles a metal cluster that serves as a FeMo-co precursor upon incubation with Fe, S, and SAM *in vitro* but it is not capable of retaining this metal precursor after synthesis.

To understand how conserved Cysteine residues (supposedly coordinated the protein [Fe-S] clusters) could influence NifB_{Mi} activity, we tested the *in vitro* activities of the three generated NifB_{Mi} variants. The results obtained and presented in Figure 34 show that substitution of Cys residues by Ala (C36A/ C40A/C43A) coordinating the [Fe₄-S₄]-SAM cluster in variant SM render a protein unable to activate apo-NifDK. NifDK activation ability was also lost in the AM variant carrying amino acid substitutions in the C-terminal CRXCAXG motif, which is similar to those shown to bind [Fe-S] clusters for storage and transfer (89, 96) and in the DM variant with amino acid substitutions in both domains (Table 8).

Figure 34. NifB_{Mi}-dependent apo-NifDK reconstitution assays with NifB_{Mi} WT and site-directed variants. Activities are nmol of ethylene · min⁻¹ · mg⁻¹ NifDK (mean ± SD; n ≥ 2)

4.6. Detection of NifB_{Mi} radical SAM activity by-products

Unfortunately, in the case of the *in vitro* FeMo-co synthesis assay it is not possible to correlate directly the consumption of substrate (SAM, Fe, S) by NifB and the production of NifB-co. As previously described, NifB activity must be measured in a coupled assay, in which NifB-co is matured into FeMo-co and then inserted into apo-NifDK to restore nitrogenase activity in passing through a complex pathway where numerous other participants are acting (NifX, NifEN, NafY, NifH and NifDK). However, it is possible to detect and quantify by-products of a SAM-radical protein such as NifB. Consistently, the by-products of the NifB_{Mi} reaction were analyzed on a C18 column by HPLC. Scheme n.1 represents the capacity of NifB_{Mi} to generate SAH and 5'-deoxyadenosine (5'-dAH) from SAM in the presence of reductant during the formation of NifB-co.

The chromatogram presented in Figure 35 shows the separation of SAM, AdoH, and SAH produced by NifB catalysis during the pre-activation phase. The observation of identical SAM cleavage products as observed in other studied NifB proteins suggests that NifB_{Mi} follows the same mechanism that was proposed for NifB_{Av}, NifB_{Mt} and NifB_{Ma}, i.e. mobilization of the methyl group of one SAM substrate and subsequent abstraction of a hydrogen atom from this methyl group by the 5-dA[•] radical generated by the homolytic cleavage of a second equivalent of SAM (116).

Scheme 1. Minimal mechanism of NifB-co synthesis in NifB

Figure 35. HPLC analysis of the by-products of NifB_{Mi} reaction catalyzed in the pre-activation phase. After 15 minutes of incubation at 60°C, the samples were centrifuged (13,000 rpm for 5 min) and then injected on a C18 column. The retention times of each compound present in the reaction medium were assigned by injection of commercial standards. The production of AdoH, depends on the presence of all constituents of the enzyme activity test.

4.6.1. SAH and AdoH production correlated with NifB_{Mi}- dependent activation of NifDK over time.

The capacity of NifB_{Mi} to transfer a methyl group from SAM, to generate 5' dAdo•, and to initiate FeMo-co synthesis were determined by measuring concomitant production of *S*-deoxyadenosyl-L-homocysteine (SAH), 5-deoxyadenosine (dAdoH), and reconstituted NifDK in a single assay. For time-course experiments with simultaneous detection of ethylene, SAH and AdoH, the 60°C NifB_{Mi} pre-incubation phase described in the previous chapter, was omitted and the reactions were performed at 40°C in a series of 0.8 ml reaction mixtures as described in Materials and Methods. From the injection of SAH, AdoH standards into HPLC, standard curves were obtained making it possible to quantify these compounds (Figure 36).

The kinetics of AdoH, SAH formation was sigmoidal. For an amount of 4 nmol of enzyme, the maximum level of AdoH production was 0.2 nmol, a ratio of 0.05

turnovers per enzyme molecule. This result underlines again how this enzyme was not able to perform efficiently at a temperature of 40°C. Interestingly, demethylation of SAM proceeded at higher rates than dAdoH formation (3:2 SAH to dAdoH ratio), and this difference was maintained over time.

Figure 36. Time course of SAH and AdoH formation during NifB_{Mi}-dependent FeMo-co synthesis and apo-NifDK reconstitution. Activity refers to nmol ethylene formed per min and mg of NifDK. Data are the mean $\pm$ SD of at least two independent experiments.

This “asymmetric” formation of SAM cleavage products suggests that, NifB_{Mi}, as other NifB proteins (116) catalyzes the removal of methyl group from SAM at faster rates than it forms the 5'-dAH. Another possibility to explain this result is that decomposition of SAM into different products varies in proportion to the increase in temperature and incubation time.

Incubation of NifB_{Mi} with [methyl-d3]-SAM, where the three hydrogen atoms of the methyl group were labeled with deuterium, resulted in a mixture of deuterium-enriched 5'-dAD and unlabeled 5'-dAH. Incubation with ¹³CD³-SAM resulted in production of deuterated 5'-deoxyadenosine (dAdoD) demonstrating abstraction of a H atom from the methyl group of SAM (Figure 37). The observation of the SAM cleavage products demonstrates that NifB_{Mi} was capable to catalyze the SAM-dependent reaction, mobilizing the methyl group of one equivalent of SAM and subsequently abstracting a hydrogen atom from this methyl group by a 5'-dA· radical that is derived from a second equivalent of SAM. The additional detection of 5'-dAH suggests either *i*) the occurrence of abortive cleavage of SAM, in which the 5'-dA·

abstracts a solvent-derived hydrogen atom, *ii*) the abstraction by 5'-dA $\cdot$ from an unlabeled methyl group generated on NifB through the action of unlabeled SAM during cell growth that remains associated with NifB after purification; or *iii*) a combination of both events.

Figure 37. Electrospray Ionization Mass Spectrometry (ESI-MS) analysis showing the shift in m/z 253 abundance between the 5'-dAdoH control and 5'-dAdoD generated by incubating NifB_{Mi} with CD₃-SAM. 5'-dAdoH elution profile (A) and MS analysis (B). Panel (C) shows the products formed by incubation of NifB_{Mi} with CD₃-SAM. (D) MS analysis of the peak eluted at 3.8 min corresponding to 5'-dAdoD.

Chapter III: Spectroscopic studies of NifB_{Mi} WT and site-directed variants

4.7. Electron paramagnetic resonance (EPR) of NifB_{Mi} wild type and site-directed variants

Continuous wave X-band (9.4 GHz) EPR analysis was performed on as isolated and reconstituted enzyme preparations to investigate the composition of the [Fe-S] clusters present in WT NifB_{Mi} and the site-directed SM, AM and DM variants. Protein samples reduced by dithionite in the presence or absence of SAM exhibited spectra with components representative of multiple $S = 1/2$ [Fe-S] clusters (Figure 38)

Figure 38. To the left: X-band CW EPR spectroscopy of DTH-reduced NifB_{Mi} WT and site-directed variants. (A) WT NifB_{Mi} (274 μ M), (B) WT NifB_{Mi} in the presence of 3 mM SAM, (C) SM variant (159 μ M), (D) AM variant (159 μ M), and (E) DM variant (252 μ M). Data are shown in black with composite spectral simulations in red. To the right: EPR spectra of [Fe-S] clusters reconstituted NifB_{Mi} variants. X-band CW EPR measurements of NifB_{Mi} protein samples prepared anaerobically with 10 mM DTH in the absence (Blue) or Presence (Red) of 3 mM SAM. **Top Left:** Wild type NifB_{Mi} **Top Right:** Auxiliary cluster substituted variant (AM); **Bottom Left:** SAM Binding cluster substituted variant (SM); **Bottom Right:** Double cluster substituted variant (DM). Typical conditions for recording spectra were 9.40 GHz microwave frequency, 1 mW power, 10 G modulation amplitude, and temperature of 15K.

The addition of SAM altered the EPR spectrum of the wild type enzyme. None of the variant exhibited altered spectra in the presence of SAM. This result was expected for the SM and DM mutants but was surprising for the AM mutant. Confronted with the result obtained by the *in vitro* activity assay with pure Nif proteins and NifB_{Mi} variants, this result also supports a fundamental role for AC1 in NifB-co synthesis.

The EPR spectrum of the DM reconstituted variant was consistent with a single [Fe₄S₄] cluster, in agreement with its Fe and S content (4.6 ± 0.4 Fe; 5.5 ± 0.3 S atoms). The experimental spectra of WT and all the variants were simulated using the DM variant as starting point. Individual component spectra, simulation parameters, and spin quantitation are shown in Figure 39, Table 10, and Table 11, respectively.

Figure 39. Components of simulations for [Fe-S] cluster reconstituted NifB_{Mi}. X-band CW EPR measurements of NifB_{Mi} protein samples prepared anaerobically with 10 mM DTH and frozen in an acetone/dry ice bath. (A) WT (274 μM), (B) AM (159 μM), (C) SM (159 μM), and (D) DM (252 μM). Data are presented in Black with composite spectral simulations in Blue for the WT NifB_{Mi} and each of the variants. Contributions of each of the individual component clusters are included with the SAM binding cluster RS (Red), the auxiliary cluster AC1 (Fuchsia), the auxiliary cluster AC2 (Light Blue), and the degradation cluster, ACx, (Dark Green). Typical conditions for recording spectra were 9.40 GHz microwave frequency, 1 mW power, 5G modulation amplitude, and temperature of 15K.

	g_1	g_2	g_3	g_{av}
RS	2.038	1.931	1.916	1.962
AC1	2.062	1.917	1.875	1.951
AC2	2.039	1.923	1.886	1.949
ACx	2.058	1.985	1.909	1.984

WT			
	σ_1	σ_2	σ_3
RS	0.028	0.015	0.04
AC1	0.015	0.015	0.018
AC2	0.03	0.03	0.05
ACx	-	-	-

	AM			SM			DM		
	σ_1	σ_2	σ_3	σ_1	σ_2	σ_3	σ_1	σ_2	σ_3
RS	0.028	0.018	0.04	-	-	-	-	-	-
AC1	-	-	-	0.025	0.018	0.018	-	-	-
AC2	0.03	0.04	0.08	0.03	0.03	0.08	0.03	0.04	0.08
ACx	-	-	-	0.017	0.018	0.03	-	-	-

Table 10. EPR simulation parameters. RS, SAM binding cluster; AC1, Auxiliary cluster 1; AC2, Auxiliary cluster 2; ACx, degradation cluster.

	-SAM (μM)	+SAM (μM)
WT	52	111
SM	65	82
AM	66	92
DM	53	59

Table 11. Spin quantitation of NifB_{Mi} variants in the absence or presence of SAM

4.7.1. Temperature dependence and spectral subtractions of WT and site-directed variants of NifB_{Mi}

Temperature-dependent EPR, which has proven useful to establish the nuclearity of [Fe–S] clusters (143, 144) was utilized to identify individual species present. Data shown in Figure 40 suggest that the three clusters present in NifB_{Mi} are [Fe₄S₄] clusters. Additionally, the g_1 feature of each cluster becomes clear when examining the temperature dependence of the different spectra, providing spectroscopic handles for these clusters.

Deconvolution through spectral subtractions was performed to support individual cluster assignments. Unique features for each of the species present, and the positions of g1 and g3 for each of the clusters were verified (Figure 41). The cluster present in the DM variant, termed AC2, showed features typical of a $[\text{Fe}_4\text{S}_4]$ cluster, such as a fairly axial g- tensor ($g = [2.039 \ 1.923 \ 1.886]$) and signal disappearance above 40 K. Three clusters were identified in the SM variant. The $[\text{Fe}_4\text{S}_4]$ cluster described above for the DM variant (AC2), an additional $[\text{Fe}_4\text{S}_4]$ cluster, termed AC1, with g-values of $[2.062 \ 1.917 \ 1.875]$, and a third cluster that was only observed above 50 K and had a much more rhombic g-tensor of $[2.058 \ 1.985 \ 1.909]$ consistent with a $[\text{Fe}_2\text{S}_2]$. This $[\text{Fe}_2\text{S}_2]$ cluster was only apparent in the SM variant and was tentatively assigned to a degraded cluster, termed ACx. The EPR spectrum of the AM variant could be simulated with a combination of AC2 and a new $[\text{Fe-S}]$ cluster with g-values of $[2.038 \ 1.931 \ 1.916]$, which has been assigned to the SAM-binding $[\text{Fe}_4\text{-S}_4]$ cluster (RS). The g-values and temperature dependence of the EPR spectrum for RS are in agreement with reported SAM binding clusters (144).

Figure 40. Temperature dependence and spectral subtractions of WT and site-directed variants of NifB_{Mi}. All NifB_{Mi} samples were reduced with 10 mM DTH. **(Top Left)** Temperature dependence of 274 μM WT NifB_{Mi}. **(Top Right)** Temperature dependence of 159 μM AM NifB_{Mi}. **(Bottom Left)** Temperature dependence of 159 μM SM NifB_{Mi}. **(Bottom Right)** Temperature dependence of 252 μM DM NifB_{Mi}. Typical conditions for recording spectra are 9.40 GHz microwave frequency, 1 mW power,

and 5G modulation amplitude. Spectra were recorded at various temperatures indicated to the right of each panel.

Figure 41. Spectral subtractions of WT NifB_{Mi} and variants utilized in determining the clusters present by differences in the spectral composition of the variants. A) WT NifB_{Mi}; (B) AM variant; (C) SM variant; (D) DM variant; (E) WT minus AM; (F) WT minus SM; (G) Difference spectrum F minus DM to remove unsubtracted component and highlight the EPR signal from AC1 cluster. Asterisk marks g₂ of ACx degradation cluster. Dashed lines are included to indicate defining g₁ and g₃ spectral features used in deconvolution. Typical conditions for recording spectra are 9.40 GHz microwave frequency, 1 mW power, 5G modulation amplitude, and a temperature of 15 K.

4.8. ESEEM of NifB_{Mi} wild type and site-directed variants

Examination of WT NifB_{Mi} by three-pulse electron spin echo envelope modulation (3P-ESEEM) revealed nitrogen modulations to the echo from one or more of the [Fe–S] clusters (Figure 42). Incubation with isotopically enriched ¹⁵N-SAM, carried out to investigate changes in spectral features attributed to SAM coordination, resulted in no change in modulations (Figure 42, blue trace) as would be expected if this modulation derived from coupling to SAM nitrogen. This lack of ¹⁴N/¹⁵N SAM isotope sensitivity of the ESEEM spectra suggests proteinaceous nitrogen coordination to one or more of the [Fe–S] clusters. Due to its simplified EPR spectrum and cluster content, the DM variant was selected to investigate nitrogen

coordination from the protein to the AC2 cluster. The nitrogen modulations observed in the DM variant (Figure 42, red trace) were very similar to those of the WT, suggesting that AC2 is coordinated by at least one nitrogen ligand. While the nitrogen modulations arising from AC2 are present in the spectra of WT NifB_{Mi} with peaks at 2 and 0.7 MHz (Figure 42), there was also evidence for additional nitrogen modulations from a second cluster (peaks at 2.3 and 1 MHz), which was tentatively assigned to AC1.

ESEEM data indicated a weakly bound ¹⁵N from SAM and ¹⁴N from protein. The source of ¹⁴N needs to be verified by removing the histidine tag from the protein and exchanging the Tris by another buffer. Nitrogen coupling from the protein has implications in redox potential of the cluster and might have mechanistic implications about how the NifB-co is transferred to NifEN.

Figure 42. X-band (9.4 GHz) 3P-ESEEM time domain (top) and Fourier transform of the time domain (bottom) of NifB_{Mi}. Black trace, WT NifB_{Mi} reduced with 10 mM DTH in the presence of 3 mM ¹⁴N-SAM. Blue trace, WT NifB_{Mi} reduced with 10 mM DTH in the presence of 3 mM ¹⁵N-SAM. Red trace, DM variant reduced with 10 mM DTH in the absence of SAM.

4.9. Preliminary HYSCORE studies of NifB_{Mi} WT and DM proteins

After reconstitution of its [Fe₄S₄] centers, NifB_{Mi} wild type was reduced by DTH, incubated with 3 mM of ¹⁵N-SAM and analyzed by HYSCORE spectroscopy. HYSCORE is a two-dimensional EPR technique that monitors nuclei interacting with an $S = \frac{1}{2}$ [Fe₄S₄]⁺ system. We also changed the original buffer conditions with HEPES as the Tris can interfere with the analysis. The analysis of the spectrum obtained (Figure 43) revealed a potential proteinaceous nitrogen interaction with one or more of the [Fe–S] clusters attributed to a nitrogen atom weakly coupled to a paramagnetic metal center. This result suggested the presence of a nitrogen atom close to the paramagnetic center but probably not coordinating this center.

The ¹⁵N atom presented numerous absorption peaks limiting the interpretation of HYSCORE results. On order to diminish interferences arising from the radical-SAM cluster (RS) and the auxiliary cluster 1 (AC1), we also analyzed the NifB_{Mi} DM variant by HYSCORE. Weak nitrogen coordination was also evident in the DM variant (Figure 44), suggesting that the second auxiliary cluster (AC2) is the metal center coordinated by at least one nitrogen ligand from the protein.

Figure 43. HYSCORE spectrum of NifB_{Mi} incubated with ¹⁵N-SAM in the quadrant (+, +) and (+, -). Top: image resolution at temp 128 t 10 K, field 350 mT, depth 3.00; 128 t 10 K, field 350 mT, depth 3.30. The NifB_{Mi} in the presence of 3 mM ¹⁵N-SAM (311 μM), is reduced with 10 mM DTH.

Figure 44. HYSCORE spectrum of NifB_{Mi} DM incubated with ¹⁵N-SAM in the quadrant (+, +) and (+, -). Top: image resolution at temp 128t, 10 K, field 350 mT, depth 3.00; 140t, 10 K, field 342 mT, depth 3.50. NifB_{Mi} DM in the presence of 3 mM ¹⁵N-SAM (286 μM), is reduced with 10 mM DTH.

Chapter IV: NifB_{Mi} crystallization trials

4.10. Crystallization trials

During my internship at the Metalloproteins unit of the Institut de Biologie Structurale in Grenoble, I focused my research on designing and performing preliminary crystallography tests on NifB_{Mi} wild type. These experiments were conducted at the end of 2013, one year before the internship at the David Britt lab and, at this time, we had not yet generated the site-directed NifB variants used for spectroscopic and biochemical studies, an event that limited considerably our crystallography study. As these studies were performed before obtaining the spectroscopic and enzymatic characterization of NifB_{Mi}, we hypothesized that this NifB_{Mi} could potentially coordinate other substrates, in addition to SAM and the [Fe₄-S₄] cluster precursors during the assembly of the [Fe₈-S₉-C] NifB-co. In particular we were interested in a possible role of FAD as it had been detected in some of our NifB preparations and it had been previously demonstrated that another metallocluster protein complex, the SufBC₂D complex of *E. coli*, coordinates a molecule of FADH₂ during the [Fe-S] cluster assembly (145). FADH₂ is fundamental not only for the enzymatic activity of this enzyme but also for its structural stability (145).

Figure 45 shows some examples of the three potential scenarios extracted from the first NifB_{Mi} crystallographic tests. In the first case (A) the sample at 10 mg/ml was unable to initiate the nucleation process, consequence of unequal protein/ precipitant buffer ratio, maintaining the structural state in the metastable zone. Conversely, in the scenario (C), the kinetic frequency of molecular interactions is too high during the nucleation process and as a consequence the protein precipitates. In the third scenario (B), the NifB_{Mi} protein showed the initiation of a nucleation process after 4 days of incubation, potentially related with the outset for crystals formation.

Among the 1248 tests performed, we selected two conditions potentially suitable for optimization (Table 12 and Figure 46). In the first condition, at a NifB_{Mi} concentration of 10 mg/ml, a small yellow brown hexagonal crystal was present after one day of incubation.

Figure 45. Examples of crystallization during the first attempts performed on NifB_{Mi}: A & C) 10 mg/ml NifB_{Mi} and B) 5 mg/ml NifB_{Mi}. All samples were incubated with 1 mM SAM and 1 mM FAD.

Figure 46. Time-course of the two conditions selected as suitable for the crystallogensis optimization. Both samples were incubated previously for 2 hours with 1 mM SAM and 1 mM FAD.

Condition	Protein condition	Precipitant	Buffer	Salt
1	10 mg/ml NifB _{Mi} + (1 mM SAM + 1 mM FAD)	45% 2-methyl 2,4- pentanediol (MPD)	0.1 M bis Tris pH 6,5	0.2 M CaCl ₂
2	15 mg/ml NifB _{Mi} + (1 mM SAM)	(2.5 M) NaCl	0.1 M Tris pH 7	0.2 M MgCl ₂

Table 12. Precipitant and protein concentration of the two conditions selected for the crystallogenesi's optimization among the initial 1248 combinations tested.

As condition 1 of Table 12 rendered the best amount and quality of crystals formed, we decided to optimize it by manually modifying the pH, MPD and protein concentration in a restricted range around the initial conditions inside a glove box. A total of 24 optimization conditions varying the precipitant and buffer parameters were tested and are presented in Table 13. This approach proved fruitful for the first group of conditions (NifB_{Mi} 10 mg/ml +1 mM SAM + 1 mM FAD) since we were able to obtain crystals similar to the initial screening at two different MPD concentrations (35% and 40% (w/v)). The crystals thus obtained were light yellow (Figure 47) suggesting the presence of FAD inside the crystal. This was expected as a result of the incubation of the protein for 2 hours with this substrate. Unfortunately, the second (NifB_{Mi} 5 mg/ml +1 mM SAM + 1 mM FAD) and fourth (NifB_{Mi} 15 mg/ml +1 mM SAM) groups of conditions did not show any crystal nuclearization (data not shown).

The groups of conditions 1 and 3 showed that crystal formation after five days of incubation was clearly influenced by buffer composition and protein concentration. Indeed at precipitant buffer concentration of 35% MPD and pH 8.5, we obtained evident yellow crystals of considerable dimensions. These crystals were analyzed at the European Synchrotron Radiation Facility. Unfortunately after the diffraction analysis demonstrated that the crystals obtained were of FAD coordinated to Cl₂ molecules. Other members of the group continuing this work are currently carrying out further crystallization tests.

Condition group	Protein condition	Precipitant concentration	pH	Salt
1	10 mg/ml NifB _{Mi} + (1 mM SAM + 1 mM FAD)	MPD: 35, 40, 45, 50, 55, 60%	4.5, 5.5, 6.5, 7.5	0.2 M CaCl ₂
2	5 mg/ml NifB _{Mi} + (1 mM SAM + 1 mM FAD)	MPD: 25, 30, 35, 40, 45, 50%	6.5, 7.5, 8.5, 9.5	0.2 M CaCl ₂
3	15 mg/ml NifB _{Mi} + (1 mM SAM + 1 mM FAD)	MPD: 25, 30, 35, 40, 45, 50%	6.5, 7.5, 8.5, 9.5	0.2 M CaCl ₂
4	15 mg/ml NifB _{Mi} + (1 mM SAM)	NaCl: 1, 1.5, 2, 2.5, 3, 3.5 M	6.5, 7.5, 8.5, 9.5	0.2 M MgCl ₂

Table 13. Precipitant and protein concentrations for the manual optimization conditions.

Figure 47. Crystals obtained in the crystallogensis optimization of NifB_{Mi} samples (10 and 15 mg/ml) previously incubated with 1 mM SAM and 1 mM FAD. Precipitant /buffer conditions are: (A) 0.01 M Tris pH 7.5; 35% (w/v) MPD, and 0.2 M CaCl₂. (B) 0.01 M Tris pH 7.5; 40% (w/v) MPD, and 0.2 M CaCl₂. (C) 0.01 M Tris pH 8.5; 35% (w/v) MPD, and 0.2 M CaCl₂. (D) 0.01 M Tris pH 8.5; 40% (w/v) MPD, and 0.2 M CaCl₂.

Discussion

5.1. NifB phylogeny and architecture

To our knowledge, this work presents the largest compilation of NifB proteins described to date. The NifB database was stringently filtered to exclude faulty annotated proteins and the curated dataset provides insights about NifB origin, taxonomy and architecture that complement previous work (9, 139, 146). In this study we demonstrate that the SAM-radical domain of NifB is sufficient to support FeMo-co biosynthesis *in vivo* in the model organisms *A. vinelandii* and *K. oxytoca*.

A strict filter, based on motifs exhibited by experimentally confirmed NifB proteins, was applied to the initial database. As a result, 28% NifB sequences were excluded from further analysis. Although these criteria might be too strict, we reasoned that it was better to miss some true-positives than to risk including false-positives. Most excluded NifB proteins lacked the conserved Cx₃Cx₂C motif required for SAM-radical catalysis. On the contrary, the NifX domain was identified in each one of them and we think that these faulty annotated NifB proteins instead are NifX proteins. This confusion originates from the fact that the NifX domain is present in NifB, NafY, NifY as well as NifX proteins.

Three distinct NifB protein architectures exist. The most widespread in Bacteria consists of an N-terminal SAM-radical domain followed by a C-terminal NifX-like domain. However, this configuration is absent in Archaea, which present smaller NifB proteins consisting of a stand-alone SAM-radical domain. Boyd and collaborators investigated the lineage of the stand-alone SAM-radical domain in Archaea NifB proteins and compared it to the two-domain architecture favored in Bacteria (139). The authors traced an event that suggested that a methanogen donated its *nif* cluster via LGT to a Firmicutes ancestor that co-existed in the same ecological niche. Then, a fusion event happened that resulted in the *nifB-nifX* protein occurring in Firmicutes. It was later suggested that the widespread of the *nifB-nifX* fusion protein in Bacteria was independent of the selective pressure associated with aerobic diazotrophy (147). An additional fusion event between *nifN* and *nifB-nifX* also occurred in Firmicutes leading to the three-domain NifB architecture. This last event was confined to Firmicutes, which is the only phylum presenting all three types of NifB architecture. It is surprising that the three-domain NifB was not widespread in Bacteria. From knowledge gained through *in vitro* FeMo-co synthesis studies (114), it

could be assumed that a NifENB fusion protein would be beneficial by protecting labile NifB-co and streamlining FeMo-co synthesis. However, it is possible that a NifENB fusion might not allow fine-tuning of precursor biosynthesis.

Based on the phylogeny of independent NifX proteins, another early *nifB* LGT was detected between Methanosarcinales and Chloroflexi. This event was also apparent in the SAM-radical domain phylogenetic tree, with Chloroflexi rooting deeper than any other group. The short distance between Methanosarcinales and Chloroflexi NifB lineages was also observed by Boyd and colleagues (139).

5.2. Ancestral NifB protein from a strict anaerobic and thermophilic Archaea functions *in vivo* in an aerobic mesophilic host

Although our studies and those of the Ribbe group have demonstrated that stand-alone SAM-radical domain NifB proteins catalyze NifB-co synthesis *in vitro* (115, 116), *M. thermautotrophicus*, *M. acetivorans*, and *M. infernus*, which are not experimentally confirmed diazotrophs, have not been proven capable of sustaining diazotrophic growth. It was also unclear whether or not this NifB family would function in mesophilic and aerobic environment. The Nif⁺ phenotype exhibited by the *A. vinelandii* Δ *nifB* strain complemented with *nifB*_{Mi} shown in this thesis is definitive proof of *in vivo* functionality in mesophilic aerobic bacteria.

As expected, stronger Nif⁺ phenotype was achieved by complementation with NifB_{Ct}. (compare growth of UW418 and UW422). *C. tepidum* is a mild thermophile with optimum growth temperature of 48°C and therefore much closer to the 30°C optimum of *A. vinelandii*. In addition, NifB_{Ct} has a two-domain NifB architecture similar to NifB_{Av}. Interestingly, NifB_{Ct} was a monomer, similar to the archaea single-domain NifB proteins and different from the NifB_{Av} and NifB_{Ko} homodimers. Although constrained by the limited set of available experimental data, it appears that NifB monomers might be more stable and therefore favored in thermophilic organisms regardless of protein architecture. Importantly, both configurations are functional *in vivo* in a mesophilic host.

The capacity of the gene products of *nifB*_{Mi} to synthesize *in vivo* the FeMo-co precursor and support the diazotrophic growth of the aerobic, mesophilic microorganism *A. vinelandii* was an remarkable result as this protein is naturally

expressed by strict anaerobic and thermophilic microbes. However, in contrast to UW422, the UW418 strain did not exhibit measurable *in vivo* acetylene reduction assay (data not shown).

5.3. Purified NifB_{Mi} is active in the *in vitro* FeMo-co synthesis assay

The wild type and three site-directed variants of NifB_{Mi} were purified with high yield and good purity of the product. This monomeric protein, as it is the case of other recombinant proteins from hyperthermophilic organisms, was overexpressed in soluble form in the *E. coli*. Although NifB_{Mi} belongs to the Radical SAM family, which is characterized for its extreme O₂ sensitivity, the His-tagged NifB_{Mi} was successfully co-expressed with the *E. coli* (Fe-S) machinery assembly in BL21 (DE3) yielding a dark brown cell-free extract and purified protein partially full of its metal content.

After purification, all NifB_{Mi} variants remained stable in highly concentrated preparations and maintaining quite stable [Fe-S] centers. Moreover these NifB_{Mi} forms, including the double mutant variant, showed UV-visible spectra characteristic of the presence of [Fe₄S₄]²⁺ centers. The comparison of OD 413 nm / 280 nm ratios and Fe/S contents of as isolated and reconstituted NifB_{Mi} variants is consistent with the presence in wild-type NifB_{Mi} of a minimum of two [Fe₄S₄] clusters in addition to the SAM-binding [Fe₄S₄] cluster.

The *in vitro* activity assays with NifB_{Mi} demonstrated how the temperature played a key role in the activity of this hyperthermophilic enzyme, as expected. Moreover, protein concentration in the *A. vinelandii* cell-free extract and substrate availability seemed to limit NifB_{Mi} activity in the reconstitution of NifDK in complementation of UW140 (Δ nifB) extracts. Similarly, substrate availability (SAM, Fe, and S) limited NifB_{Mi} activity in the *in vitro* FeMo-co synthesis and nitrogenase reconstitution assays containing only purified components. Pre-activation of NifB_{Mi} by reconstituting its [Fe-S] clusters at 60°C prior to the activity assay increased enzyme activity but did not relieve the need for SAM, Fe and S in the FeMo-co synthesis phase, indicating that the NifB-co formed during the pre-activation phase did not remain bound to the protein and was not useful for further FeMo-co synthesis. This fact limited the maximum activity observed in the assays since productive NifB-co

synthesis by NifB_{Mi} could not be carried out at temperatures allowing rapid enzyme turnover (e.g. 60°C-80°C) due to deleterious effects of the other enzymes of the FeMo-co pathway.

Wild type NifB_{Mi}, both in the as isolated and reconstituted forms, was the only variant able to produce NifB-co and sustain FeMo-co synthesis *in vitro*. This results underline not only the importance of the SAM radical domain, extensively demonstrated to be fundamental for the NifB catalysis (69, 115, 116) but also how the AC1 [Fe-S] cluster, which is absent in the AM and DM variants, plays a crucial role in activity. We favor a hypothesis in which AC1 is in facts a NifB-co precursor.

The quantification of S-deoxyadenosyl-L-homocysteine (SAH), 5-deoxyadenosine (dAdoH) by-products demonstrated asymmetric formation of SAM cleavage products. NifB_{Mi} catalyzed the removal of methyl group from SAM at much faster rates than it formed the 5'-dAH. This observation could be explained by the fact the NifB catalyzes not only the condensation of two [Fe₄-S₄] clusters into NifB-co but also the insertion of a carbide atom at the center of it. The origin of the carbide atom is the methyl group of SAM. Therefore NifB_{Mi} demethylation activity may be independent of radical formation produced by the homolytic cleavage of a second SAM molecule.

Overall, the data presented here provide direct proof of NifB_{Mi} serving as heterologous NifB-co donor for FeMo-co assembly both *in vitro* and *in vivo* in *A. vinelandii*. NifB from *M. infernus* represents a simplified homologue (lacking the NifX domain) of NifB proteins from model diazotrophs *A. vinelandii* and *K. oxytoca*. These results are in agreement with those recently reported for *M. acetivorans* and *M. thermoautotrophicum*.

5.4. Spectroscopic characterization of and mechanistic implications

EPR and biochemical analyses unambiguously demonstrate the presence of three [Fe₄S₄] clusters in NifB_{Mi}. One of these clusters is the [Fe₄S₄] SAM binding cluster (RS) coordinated by the canonical CX₃CX₂C motif (g1: 2.038, g2: 1.931, g3: 1.916). While the auxiliary clusters AC1 (g1: 2.06, g2: 1.917, g3: 1.875) and AC2 (g1: 2.039, g2: 1.923, g3: 1.886) may be present as conduits for electron transfer, our preferred model is that they are in-fact precursors to NifB-co. Based on the temperature behavior of EPR spectra, reporting information on the relaxation dynamics of the

cluster, we propose that both auxiliary clusters are present as [Fe₄S₄] clusters. While the temperature dependence of an [Fe–S] cluster can help discriminate between a [Fe₄S₄] and [Fe₂S₂] cluster (a [Fe₄S₄] cluster having faster relaxation rates than a [Fe₂S₂] cluster), this analysis can become clouded when multiple clusters are present and near each other causing increased relaxation rates from dipolar coupling between them. Therefore, the presence of one [Fe₂S₂] cluster cannot be completely ruled out. The assignment of NifB-co as a fusion of two [Fe₄S₄] clusters is consistent with the current proposal of NifB-co comprising the [Fe₈S₉C] core of FeMo-co (73).

Indeed NifB_{Mi} contains all cluster species that are required to facilitate the conversion of intermediate [Fe₄S₄] clusters into NifB-co in the presence of SAM. Similar conclusions were obtained in a parallel study with NifB from *M. acetivorans* (NifB_{Ma}) and *M. thermoautotrophicum* (NifB_{Mt}) (116) although in that case the assignment was mostly based on Fe/S contents rather than in a detailed spectroscopic analysis.

Based on the number of strictly conserved cysteines in NifB_{Mi}, there are not enough to fully coordinate the SAM- binding cluster plus two additional [Fe₄S₄] clusters. This would indicate that the two auxiliary clusters are either not saturated in Fe coordination or that some non-cysteinyll residues may coordinate the auxiliary clusters as suggested by 3P-ESEEM data. Preliminary HYSCORE results also underline the potential interaction of one nitrogen atom from an amino acid residue of NifB_{Mi} with AC2. This result led us to think about the potential involvement of one of the histidine residues that are conserved in the NifB family in the coordination of the metal centers during NifB-co synthesis. NifB_{Mi} has two conserved His residues as well as several conserved Arg residues. Coordination by His or Arg residues is well-known for [Fe₂S₂] clusters and has precedence for an auxiliary [Fe₄S₄] cluster of Cp1 Fe–Fe hydrogenase, as well as in the hydrogenase cofactor maturase HydF (148, 149). Additionally, in BioB, IscU and MitoNEET nitrogen coordination is utilized to bind transient [Fe–S] clusters (150) Nitrogen coordination to the cluster would provide a convenient path for cluster release through a simple protonation of the nitrogen ligand and release of NifB-co to NifEN or NifX. Like NifB from *M. acetivorans* and *M. thermoautotrophicum*, (116) NifB_{Mi} produces more SAH than AdoH in turnover assays (a 3:2 ratio). We note that SAM cleavage to AdoH has never been observed to be in excess of SAH in NifB suggesting a single H abstraction followed by oxidation/deprotonation events to remove the remaining H from the transferred

methyl to form carbide. Interestingly, we observed a shift in the g3 feature at 352 mT of the WT EPR spectrum in the presence of SAM (Figure 38 line B), but no indication of binding of SAM to the RS cluster in the 3P-ESEEM, suggesting the existence of additional steps in the reaction mechanism to form NifB-co. In the mechanism that has been proposed for NifB (106, 151) SAM transfers a methyl group to one of the clusters that will be fused to form NifB-co (Scheme S1). We suggest the observed perturbation to WT NifB EPR spectrum in the presence of SAM is due to methyl transfer to one of the auxiliary [Fe-S] clusters, poised for NifB-co formation, thus ensuring that, when 5'dAdo• is formed, the site of hydrogen abstraction is present preventing unwanted side reactions from the highly reactive primary carbon radical.

5.5. Crystallogenesis trials

We have carried out crystallogenesis optimization tests in order to obtain NifB_{Mi} crystals usable for structure determination through X-ray diffraction. In collaboration with the Fontecilla/Nicolet group at the IBS of Grenoble, a robot installed inside a Jacomex glovebox tried more than 1,200 different conditions for crystallogenesis. For each condition, parameters as precipitant, pH and protein concentration were altered. Two crystallogenesis conditions resulting in a light yellow hexagon and an intense brown elongated crystal were obtained were identified as promising for the formation of NifB_{Mi} crystals and manually optimized. Unfortunately, after X-ray diffraction analysis at the ESRF, these structures demonstrated to be the association of Cl₂ with a molecule of FAD. This result was not completely surprising. Indeed the presence of high buffer salt concentration during the crystallogenesis represents a problematic aspect. Frequently the salts can interact with other molecules present in the sample, creating artifacts during the nucleation phase. Moreover, previous studies have shown that Radical SAM enzymes are normally crystallized in their apo-forms, indicating that they initiate crystallogenesis when deprived of their metal clusters (152). Therefore it is possible that the NifB_{Mi} protein carrying a complete set of reconstituted [Fe₄S₄] centers cannot be with crystallized. As this form seems difficult to crystallize, further work with NifB_{Mi} site-directed variants (SM, AM and DM) in their apo and reconstituted forms, as well as new crystallization tests with NifB proteins from other organisms would be required.

Conclusions

1. Here we study the diversity and phylogeny of naturally occurring NifB proteins, their protein architecture and the functions of the distinct NifB domains in order to understand what defines a catalytically active NifB.
2. NifB from the thermophilic Archaea *Methanocaldococcus infernus* is functional *in vivo* in nitrogenase cofactor maturation and represents a simplified homolog (lacking the NifX domain) of NifB proteins from model diazotrophs *A. vinelandii* and *K. oxytoca*.
3. Highly purified preparations of NifB_{Mi} wild type and site-directed mutants were obtained.
4. NifB_{Mi} was established as a more robust model to study early steps in nitrogenase cofactor biosynthesis.
5. NifB_{Mi} is a radical SAM enzyme able to reductively cleave SAM to 5'-deoxyadenosine radical and is competent in the *in vitro* FeMo-co maturation assay.
6. NifB_{Mi} carries three [Fe₄S₄] clusters: one SAM binding cluster and two auxiliary clusters probably acting as substrates for NifB-co formation. Nitrogen coordination to one or more of the auxiliary clusters was observed.
7. Preliminary evidence is shown for the transfer of a methyl group to one of the auxiliary [Fe-S] clusters, preparing it for NifB-co formation before the highly reactive 5'dAdo• is formed.

References

1. Yung, Y. L. & Mcelroy, M. B. (1979) Fixation of nitrogen in the prebiotic atmosphere, *Science*. **203**, 1002 – 1004.
2. Schimpl, J. K. (2004) Dissertation: “Biological Nitrogen Fixation – Simulation of the Reaction Mechanism of Nitrogenase from First Principles, Mathematisch-Naturwissenschaftlichen Fakultät“ at der Technischen Universität at Clausthal.
3. Smith, B. E., Lowe, D. J. & Bray, R. C. (1973) Studies by electron paramagnetic resonance on the catalytic mechanism of nitrogenase of *Klebsiella pneumonia*, *Biochemical Journal*. **135**, 331 – 341.
4. Howard, J. B. & Rees, D. C. (1996) Structural Basis of Biological Nitrogen Fixation, *Chem Rev*. **96**, 2965 – 2982.
5. Bonete, M. J., Martínez-Espinosa, R. M., Pire C., Zafrilla, B. & Richardson, D. J. (2008) Nitrogen metabolism in haloarchaea, *Saline Systems*. **1**, 4 - 9.
6. Poret – Peterson, A. T., Graham, J. E., Gullledge, J. & Klotz, M. G. (2008) Transcription of nitrification genes by the methane-oxidizing bacterium, *Methylococcus capsulatus* strain Bath, *ISME J*. **2**, 1213 - 1220.
7. Zumft, W. G. (1997) Cell biology and molecular basis of denitrification, *Microbiol Mol Biol Rev*. **61**, 533 - 616.
8. Navarro-González, R., McKay, C. P. & Mvondo, D. N. (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning, *Nature*. **412**, 61- 4.
9. Boyd, E. S. & Peters, J. W. (2013) New insights into the evolutionary history of biological nitrogen fixation, *Front. Microbiol*. **4**: 201.
10. Burgess, B. K. & Lowe, D. J. (1996) Mechanism of Molybdenum Nitrogenase, *Chem Rev*. **96**, 2983 - 3011.
11. Montagna E. & Torres B. B. (2008) Expanding ecological possibilities. *Biochemistry and Molecular Biology Education*. **36**, 99 – 105.
12. Stacey, G., Burris, R. H. & Evans, H. J. (1992) Biological Nitrogen Fixation, Springer Science & Business Media.
13. Sprent, J. I. & James, E. K. (2007) Legume Evolution: Where Do Nodules and Mycorrhizas Fit In ?, *Plant Physiology*. **144**, 575 – 581.
14. Benson, D. R. & Silvester, W. B. (1993) Biology of Frankia strains, actinomycete symbionts of actinorhizal plants, *Microbio. Rev*. **57**, 293 - 319.
15. Rasmussen, U. & Nilsson, M. (2002) Cyanobacterial diversity and specificity in plant symbioses, *Cyanobacteria in Symbiosis*. **313** - 328.
16. Williams, W. M., Viner, A. B. & Broughton, W. J. (1987) Nitrogen fixation (acetylene reduction) associated with the living coral *Acropora variabilis*, *Mar Biol.*, **94**, 531 - 535.
17. Steindler, L., Huchon, D., Avni, A. & Ilan, M. (2005) 16S rRNA phylogeny of sponge-associated cyanobacteria, *Appl Environ Microbiol*. **71**, 4127 - 4131.
18. Stingl, U., Radek, R., Yang, H. & Brune, A. (2005) Endomicrobia: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes, *Appl Environ Microbiol*. **71**, 1473 - 1479.
19. Leigh, J. A. (2000) Nitrogen fixation in methanogens: the archaeal perspective, *Curr Issues Mol Biol*. **2**, 125 - 131.
20. Hart, M. H. (1978) The Evolution of the Atmosphere of the Earth, *Icarus*. **33**, 23 - 39.
21. Souillard, N. & Sibold, L. (1989) Primary structure, functional organization and expression of nitrogenase structural genes of the thermophilic archaebacterium *Methanococcus thermolithotrophicus*, *Mol Microbiol*. **3**, 441 - 552.
22. Raven, J. A. & Yin, Z. H. (1998) The past, present and future of nitrogenous compounds in the atmosphere and their interactions with plants, *New Phytol*. **139**, 205 - 219.
23. Anbar, A. D., Duan, Y., Lyons, T. W., Arnold, G. L., Kendall, B., Creaser, R. A., Kaufman, A. J., Gordon, G. W., Scott, C., Garvin, J. & Buick, R. (2007) A whiff of oxygen before the great oxidation event, *Science*. **317**, 1903 - 1906.
24. Scott, C., Lyons, T. W., Bekker, A., Shen, Y., Poulton, S. W., Chu, X. & Anbar, A. D. (2008) Tracing stepwise oxygenation of the Proterozoic ocean, *Nature*. **452**, 456 - 459.
25. Smith, B. E., Eady, R. R., Lowe, D. J. & Gormal, C., (1988). The vanadium - iron protein of vanadium nitrogenase from *Azotobacter chroococcum* contains an iron-vanadium cofactor, *Biochem J*. **250**, 299 - 302.
26. Brigle, K. E., Setterquist, R. A., Dean, D. R., Cantwell, S., Weiss, M. C. & Newton, W. E. (1987) Site-directed mutagenesis of the nitrogenase MoFe protein of *Azotobacter vinelandii*.

- Proceeding of the National Academy of Sciences of the United States of America. **84**, 7066 - 7069.
27. Benton, P. M. C. & Peters J. W. (2004) The Structures of The Nitrogenase Proteins and Stabilized Complexes, *Catalysts for Nitrogen Fixation*, part of *Nitrogen Fixation: Origins, Applications, and Research Progress*. **1**, 77 - 96.
 28. Gillum, W. O., Mortenson, L. E., Chen, J. S. & Holm, R. H. (1977) Quantitative extrusion of the Fe₄S₄ cores of active sites of ferredoxins and the hydrogenase of *Clostridium pasteurianum*, *J Am Chem Soc.* **99**, 584 - 95.
 29. James, E. K. (2000) Nitrogen fixation in endophytic and associative symbiosis. *Field Crops Research*. **65**, 197 - 209.
 30. Lindahl, P. A., Papaefthymiou, V., Orme-Johnson, W. H. & Münck E. (1988) Mössbauer studies of solid thionin-oxidized MoFe protein of nitrogenase. *J Biol Chem.* **263**, 19412 - 19418.
 31. Rubio, L. M. & Ludden, P. W. (2005) Maturation of Nitrogenase: a Biochemical Puzzle, *J Bacteriol.* **187**, 405 - 414.
 32. Stiefel, E. I. (1993) Molybdenum Enzymes, Cofactors and Chemistry, An Introductory Survey, *American Chemical Society*, **535**, 1 - 19.
 33. Peters, J. W., Stowell, M. H. B., Soltis, S. M., Finnegan, M. G., Johnson, M. K. & Rees, D. C. (1997) Redox-Dependent Structural Changes in the Nitrogenase P-Cluster, *Biochemistry*. **36**, 1181- 1187.
 34. Kim, J. & Rees, D. C. (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum - iron protein from *Azotobacter vinelandii*, *Nature*. **360**, 553 - 560.
 35. Shah, V. K. & Brill, W. J. (1977) Isolation of an iron - molybdenum cofactor from nitrogenase. *Proceeding of the National Academy of Sciences of the United States of America*. **74**, 3249 - 3253.
 36. Hinnemann, B. & Nørskov, J. K. (2003) Modeling a Central Ligand in the Nitrogenase FeMo Cofactor, *J Am Chem Soc.* **125**, 1466 - 1467.
 37. Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B. & Rees, D. C. (2002) Nitrogenase MoFe - Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor, *Science*. **297**, 1696 - 1700.
 38. Kim, C. H., Newton, W. E. & Dean, D. R. (1995) Role of the MoFe Protein. Alpha - Subunit Histidine-195 Residue in FeMo-cofactor Binding and Nitrogenase Catalysis, *Biochemistry*. **34**, 2798 - 2808.
 39. Morgan, T. V., Mortenson, L. E., McDonald, J. W. & Watt, G. D. J. (1988) Comparison of redox and EPR properties of the molybdenum iron proteins of *Clostridium pasteurianum* and *Azotobacter vinelandii* nitrogenases, *Inorg Biochem.* **33**, 111 - 120.
 40. Wink, D. A., McLean, P. A., Hickman, A. B. & Orme-Johnson, W. H., (1989) A new method for extraction of iron-molybdenum cofactor (FeMoco) from nitrogenase adsorbed to DEAE-cellulose. Solubilization of FeMoco in a wide range of organic solvents, *Biochemistry*. **28**, 9407 - 9412.
 41. Huang, H. Q., Kofford, M., Simpson, F. B. & Watt, G. D. J. (1993) Purification, Composition, Charge, and Molecular weight of the FeMo cofactor from *Azotobacter vinelandii* nitrogenase, *Inorg Biochem.* **52**, 59 - 75.
 42. Rangaraj, P., Shah, V. K., & Ludden, P. W., (1997) ApoNifH functions in iron-molybdenum cofactor synthesis and apodinitrogenase maturation, *Proceeding of the National Academy of Sciences of the United States of America*. **94**, 11250 -11255.
 43. Hu Y., Fay A. W. & Ribbe M. W. (2005) Identification of a nitrogenase FeMo cofactor precursor on NifEN complex, *Proceeding of the National Academy of Sciences of the United States of America*. **102**, 3236 - 3241.
 44. Hoffman, B. M., Lukoyanov, D., Yang, Z., Y., Dean, D. R. & Seefeldt, L. C. (2014) Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage, *Chem Rev.* **114**, 4041 - 4062.
 45. Messerschmidt, A., Huber, R., Poulos, T. & Wieghardt, K., (2001) *Handbook of metalloproteins*, John Wiley and Sons, Ltd.
 46. Beinert, H. & Page, E. (1957) On the mechanism of dehydrogenation of fatty acyl derivatives of coenzyme A. V. Oxidations-reductions of the flavoproteins, *J Biol Chem.* **225**, 479-497.
 47. Calzolari, L., Gorst, C. M., Zhao, Z. H., Tong, Q., Adams, M. W. W. & La Mar, G., N. (1995) 1H NMR investigation of the electronic and molecular structure of the four-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus*. Identification of Asp 14 as a cluster ligand in each of the four redox states, *Biochemistry*. **34**, 11373 - 11384.

48. Hu, Y., Faham, S., Roy, R., Adams, M. W. W. & Rees, D. C. (1999) Formaldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus*: the 1.85 Å resolution crystal structure and its mechanistic implications, *J Mol Biol.* **286**, 899-914.
49. Iwata, S., Saynovits, M., Link, T. A. & Michel, H. (1996) Structure of a water soluble fragment of the 'Rieske' iron-sulfur protein of the bovine heart mitochondrial cytochrome bc₁ complex determined by MAD phasing at 1.5 Å resolution, *Structure.* **4**, 567 - 579.
50. Rubach, J. K., Brazzolotto, X., Gaillard, J. & Fontecave, M. (2005) Biochemical characterization of the HydE and HydG iron-only hydrogenase maturation enzymes from *Thermatoga maritima*, *FEBS Letters.* **579**, 5055 - 5060.
51. Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters, *Ann Rev Biochem.* **74**, 247 - 281.
52. Wachtershauser, G. (1990) Evolution of the first metabolic cycles, *Proceeding of the National Academy of Sciences of the United States of America.* **87**, 200 - 204.
53. Malkin, R. & Rabinowitz, J. C. (1966) The reconstitution of clostridial ferredoxin, *Biochem Biophys Res Commun.* **23**, 822 - 827.
54. Dobbek, H., Svetlitchnyi, V., Liss, J. & Meyer, O. (2004) Carbon monoxide induced decomposition of the active site [Ni-4Fe-5S] cluster of CO dehydrogenase, *J Am Chem Soc.* **126**, 5382 - 5387.
55. Sendra, M. (2007) Dissertation: Etude mécanistique de la biosynthèse des centres [Fe-S] chez *Escherichia coli*: Quel rôle pour la protéine SufA ? Université Joseph Fourier, Grenoble – France, Sciences et Géographie école doctorale chimie et sciences du vivant.
56. Chan, M. K., Kim, J., & Rees, D. C. (1993) The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 Å resolution structures, *Science.* **260**, 792 - 794.
57. Rubio, L. M. & Ludden, W.L. (2015) Biosynthesis of the Iron - Molybdenum Cofactor of Nitrogenase, *Annu Rev Microbiol.* **62**, 75 – 84.
58. Hu, Y. & Ribbe, M. W. (2013) Biosynthesis of the Iron - Molybdenum Cofactor of Nitrogenase, *J Biol Chem.* **288**, 13173 – 13177.
59. Hernandez, J. A., Curatti, L., Aznar, C. P., Perova, Z., Britt, R. D. & Rubio, L. M. (2008) Metal trafficking for nitrogen fixation: NifQ donates molybdenum to NifEN/NifH for the biosynthesis of the nitrogenase FeMo-cofactor, *Proceeding of the National Academy of Sciences of the United States of America.* **105**, 11679 - 11684.
60. Zheng, L. & Dean, D. R. (1994) Catalytic formation of a nitrogenase iron-sulfur cluster, *J Biol Chem.* **269**, 18723 – 18726.
61. McLean, P. A., Smith, B. E. & Dixon, R. A., (1983) Nitrogenase of *Klebsiella pneumoniae* nifV mutants, *Biochem J.*, **211**, 589 – 597.
62. Zheng, L., White, R. H., & Dean, D. R., (1997) Purification of the *Azotobacter vinelandii* nifV-encoded homocitrate synthase, *J Bacteriol.* **179**, 5963 - 5966.
63. George, S. J., Igarashi, R. Y., Xiao, Y., Hernandez, J. A., Demuez, M., Zhao D., Yoda, Y., Ludden, P. W., Rubio, L. M. & Cramer, S. P. (2008) Extended X-ray absorption fine structure and nuclear resonance vibrational spectroscopy reveal that NifB-co, a FeMo-co precursor, comprises a 6Fe core with an interstitial light atom, *J Am Chem Soc.* **130**, 5673 - 5680.
64. Yuvaniyama, P., Agar, J. N., Cash, V. L., Johnson, M. K., & Dean, D. R. (2000) NifS-directed assembly of a transient (2Fe-2S) cluster within the NifU protein, *Proceeding of the National Academy of Sciences of the United States of America.* **97**, 599 - 604.
65. Zheng, L., White, R. H., Cash, V. L & Dean, D. R. (1994) Mechanism for the Desulfurization of L-Cysteine Catalyzed by the nifS Gene Product, *Biochemistry.* **33**, 4714 - 4720.
66. Fontecave, M. & Ollagnier-de-Choudens, S. (2008) Iron-sulfur cluster biosynthesis in bacteria: Mechanisms of cluster assembly and transfer, *Arch of Biochem Biophys.* **474**, 226 – 237.
67. Smith, A. D., Jameson, G. N., Dos Santos, P. C., Agar, J. N., Naik, S., Krebs, C., Frazzon, J., Dean, D. R., Huynh, B. H. & Johnson, M. K., (2005) NifS-mediated assembly of (4Fe-4S) clusters in the N- and C- terminal domains of the NifU scaffold protein, *Biochemistry.* **44**, 12955 - 12969.
68. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., & Dean, D. R., (1989) Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*, *J Bacteriol.* **171**, 1017 - 1027.
69. Curatti, L., Hernandez, J. A., Igarashi, R. Y., Soboh, B., Zhao, D. & Rubio, L. M., (2007) In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins, *Proceeding of the National Academy of Sciences of the United States of America.* **104**, 17626 - 17631.

70. Goodwin, P. J., Agar, J. N., Roll, J. T., Roberts, G. P., Johnson, M. K. & Dean, D. R. (1998) The *Azotobacter vinelandii* NifEN complex contains two identical (4Fe-4S) clusters, *Biochemistry*. **37**, 10420 - 10428.
71. Brigle, K. E., Weiss, M. C., Newton, W. E. & Dean, D. R. (1987) Products of the iron-molybdenum cofactor - specific biosynthetic genes, *nifE* and *nifN*, are structurally homologous to the products of the nitrogenase molybdenum - iron protein genes, *nifD* and *nifK*, *J Bacteriol*. **169**, 1547 - 1553.
72. Kaiser, J. T., Hu, Y., Wiig, J. A., Rees, D. C. & Ribbe, M. W. (2011) Structure of precursor-bound NifEN: a nitrogenase FeMo cofactor maturase/insertase, *Science*. **331**, 91 - 94.
73. Guo, Y., Echavarri-Erasun, C., Demuez, M., Jiménez-Vicente, E., Bominaar, E. L., Rubio L. M. (2016) The Nitrogenase FeMo-Cofactor Precursor Formed by NifB Protein: A Diamagnetic Cluster Containing Eight Iron Atoms, *Angew Chem Int Ed Engl*. **55**, 12764 - 12767.
74. Hernandez, J., Igarashi, R. Y., Soboh, B., Curatti, L., Dean, D. R., Ludden, P.W. & Rubio, L. M. (2007) NifX and NifEN exchange NifB cofactor and the VK-cluster, a newly isolated intermediate of the iron-molybdenum cofactor biosynthetic pathway, *Mol Microbiol*. **63**, 77 - 92.
75. Hu, Y., Corbett, M. C., Fay, A. W., Webber, J. A., Hodgson, K. O., Hedman, B. & Ribbe, M. W. (2006) FeMo cofactor maturation on NifEN, *Proceeding of the National Academy of Sciences of the United States of America*. **103**, 17119 - 17124.
76. Rubio, L. M., Ragaraj, P., Homer, M. J., Roberts, G. P. & Ludden, P. W. (2002) Cloning and mutational analysis of the gamma gene from *Azotobacter vinelandii* defines a new family of proteins capable of metallocluster binding and protein stabilization, *J Biol Chem*. **277**, 14299 - 305.
77. Hernandez, J. A., Phillips, A. H., Erbil, W. K., Zhao, D., Demuez, M., Zeymer, C., Pelton, J. G., Wemmer, D. E. & Rubio, L. M. (2011) A sterile alpha-motif domain in NafY targets apo-NifDK for iron-molybdenum cofactor delivery via tethered domain, *J Biol Chem*. **286**, 6321-6328.
78. Rubio, L. M., Singer, S. W. & Ludden, P. W. (2004) Purification and characterization of NafY (apodinitrogenase gamma subunit) from *Azotobacter vinelandii*, *J Biol Chem*. **279**, 19739 - 46.
79. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., & Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*, *Science*. **257**, 1653 - 1659.
80. Roberts, G. P., MacNeil, T., MacNeil, D. & Brill, W. J., (1978) Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*, *J Bacteriol*. **136**, 267 - 279.
81. Gevorkyan, J. (2013) Dissertation: Biochemical characterization of the NifB enzyme and NifB-cofactor, University of Central Florida , U.S - Orlando, Department of Chemistry.
82. Hedderich, R., Albracht, S. P., Linder, D., Koch, J. & Thauer, R. K. (1992) Isolation and characterization of polyferredoxin from *Methanobacterium thermoautotrophicum*. The *mvhB* gene product of the methylviologen-reducing hydrogenase operon, *FEBS Letters*. **298**, 65 - 68.
83. Joerger, R. D. & Bishop, P. E. (1988) Bacterial alternative nitrogen fixation systems, *Crit Rev Microbiol*. **16**, 1-14.
84. Shah, V. K., Allen, J. R., Spangler, N. J & Ludden, P.W. (1994) In Vitro Synthesis of the Iron-Molybdenum Cofactor of Nitrogenase Purification and Characterization of NifB cofactor, the product of NifB protein. *J Biol Chem*. **269**, 1554 - 1558.
85. Curatti, L., Ludden, P. W. & Rubio, L. M. (2006) NifB-dependent in vitro synthesis of the iron-molybdenum cofactor of nitrogenase. *Proceeding of the National Academy of Sciences of the United States of America*. **103**, 5297 - 5301.
86. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods, *Nucleic Acids Res*. **29**, 1097 - 1106.
87. Frey, P. A., Hegeman, A. D & Ruzicka, F. J. (2008) The Radical SAM Superfamily, *Crit Rev Biochem Mol Biol*. **43**, 63 - 88.
88. Beinert, H. & Kennedy, M. C. (1989) Engineering of protein bound iron-sulfur clusters, a tool for the study of protein and cluster chemistry and mechanism of iron-sulfur enzymes, *Eur J Biochem*. **186**, 163-173.

89. Lanz, N. D. & Booker, S. J. (2015) Auxiliary iron–sulfur cofactors in radical SAM enzymes. *Biochimica et Biophysica Acta journal*, **1853**, 1316 - 1334.
90. Fontecave, M., Mulliez, E. & Ollagnier-de-Choudens, S. (2001) Adenosylmethionine as a source of 5'-deoxyadenosyl radicals, *Curr Opin Chem Biol.* **5**, 506 - 511.
91. Layer, G. Moser, J., Heinz, D. W., Jahn, D. & Schubert, W. D. (2003) Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes, *EMBO J.* **22**, 6214 – 6224.
92. Layer, G., Kervio, E., Morlock, G., Heinz, D. W., Jahn D., Retey, J. & Schubert, W. D., (2005) Structural and functional comparison of HemN to other radical SAM enzymes, *Biol Chem.* **386**, 971 – 980.
93. Hanzelmann, P. & Schindelin, H. (2004) Crystal structure of the S - adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans, *Proceeding of the National Academy of Sciences of the United States of America.* **101**, 12870 – 12875.
94. Hanzelmann, P. & Schindelin, H. (2006) Binding of 5-GTP to the C- terminal FeS cluster of the radical S-adenosylmethionine enzyme MoaA provides insight into its mechanism. *Proceeding of the National Academy of Sciences of the United States of America.* **103**, 6829 – 6834.
95. Santamaria-Araujo, J. A., Fischer, B., Otte, T., Nimtz, M., Mendel, R. R., Wray, V. & Schwarz, G. (2004) The tetrahydropyranopterin structure of the sulfur-free and metal - free molybdenum cofactor precursor, *J Biol Chem.* **279**, 15994 – 15999.
96. Wang, S. C. & Frey, P. A. (2007) S-adenosylmethionine as an oxidant: the radical SAM superfamily, *Trends in Biochemical Sciences*, **32**, 102 – 110.
97. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. (2004) Crystal structure of biotin synthase, an S-adenosylmethionine - dependent radical enzyme, *Science.* **303**, 76 - 79.
98. Cosper, M. M., Jameson, G. N. L., Eidsness, M. K., Huynh, B. H. & Johnson, M. K. (2002) Recombinant *Escherichia coli* biotin synthase is a $[2\text{Fe}-2\text{S}]^{2+}$ protein in whole cells, *FEBS Letters.* **529**, 332 - 336.
99. Cicchillo, R. M., Iwig, D. F., Jones, A. D., Nesbitt, N. M., Baleanu-Gogonea, C., Souder, M. G., Tu, L. & Booker, S. J. (2004) Lipoyl synthase requires two equivalents of S-adenosyl-L-methionine to synthesize one equivalent of lipoic acid, *Biochemistry.* **43**, 6378 - 6386.
100. Miller, J. R., Busby, R. W., Jordan, S. W., Cheek, J., Henshaw, T. F., Ashley, G. W., Broderick, J. B., Cronan, J. E. & Marletta, M. A. (2000) *Escherichia coli* LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein, *Biochemistry.* **39**, 15166 - 15178.
101. Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A. & Cronan, J. E., (2003) Assembly of the covalent linkage between lipoic acid and its cognate enzymes, *Chem Biol.* **10**, 1293 - 1302.
102. Nagatani, H. H. & Brill, W. J. (1974) *Biochimica et Biophysica Acta journal.* **360**, 160 – 166.
103. Allen, R. M., Chatterjee, R., Ludden, P.W. & Shah, V. K. (1995) Incorporation of Iron and Sulfur from NifB Cofactor into the Iron-Molybdenum Cofactor of Dinitrogenase, *J Biol Chem.* **270**, 26890 - 26896.
104. Hu, Y., Yoshizawa, J. M., Fay, A. W., Lee, C. C., Wiig, J. A. & Ribbe, M. W. (2009) Catalytic activities of NifEN: implications for the nitrogenase evolution and mechanism, *Proceeding of the National Academy of Sciences of the United States of America.* **106**, 16962 - 16966.
105. Lancaster, K. M., Roemel, M., Ettenhuber, P., Huc, Y., Ribbe, M. W., Neese, F., Bergmann, U. & DeBeer, S. (2012) X-ray Emission Spectroscopy Evidences a Central Carbon in the Nitrogenase Iron-Molybdenum Cofactor, *Science.* **334**, 974 - 977.
106. Wiig, J. A., Hu, Y., Lee, C. C. & Ribbe, M. W. (2012) Radical SAM-Dependent Carbon Insertion into the Nitrogenase M-Cluster, *Science.* **337**, 1672 - 1675.
107. Reed, G. H. (2004) Radical mechanisms in adenosylcobalamin-dependent enzymes, *Curr Opin Chem Biol.* **8**, 477 - 83.
108. Banerjee, R. (2003) Radical Carbon Skeleton Rearrangements: Catalysis by Coenzyme B12-Dependent Mutases, *Chem Rev.* **103**, 2083 - 2094.
109. Setubal, J. C., Dos Santos, P., Goldman, B. S., Ertesvag, H., Espin, G., Rubio, L. M., Valla, S., Almeida, N. F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner – Burris, K., Hernandez, J. A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T. J., Latreille, P., Ligon, L. S., Lu, J., Maerk, M., Miller, N. M., Norton, S., Carroll, I. P., Paulsen, I., Raulfs, E. C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S. L.,

- Studholme, D. J., Sun, J., Viana, C. J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D. R., Dixon, R. & Wood, D. (2009). Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic, *J Bacteriol.* **191**, 4534 – 4545.
110. Sambrook, J. & David, W. (2001) Molecular Cloning: a laboratory manual, Volume 1.
 111. Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., Custer, A. F., Hicks, M. A., Huang, C. C., Lauck, F., Mashiyama, S. T., Meng, E. C., Mischel, D., Morris, J. H., Ojha, S., Schnoes, A. M., Stryke, D., Yunes, J. M., Ferrin, T. E., Holliday, G. L. & Babbitt, P. C., (2014) The Structure–Function Linkage Database, *Nucleic Acids Res.* **42**, 521 – 530.
 112. Zhao D., Curatti L. & Rubio L. M., (2007) Evidence for nifU and nifS Participation in the Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase, *J Biol Chem.* **282**, 37016 – 37025.
 113. Chen, J. S., Toth, J. & Kasap, M. (2001) Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *Clostridium beijerinckii*, *J Ind Microbiol Biotechnol.* **27**, 281 – 286.
 114. Wiig, J. A., Hu, Y. & Ribbe, M. W. (2011) NifEN-B complex of *Azotobacter vinelandii* is fully functional in nitrogenase FeMo cofactor assembly, *Proceeding of the National Academy of Sciences of the United States of America.* **108**, 8623 - 8627.
 115. Wilcoxon, J., Arragain, S., Scandurra A. A., Jimenez-Vicente, E., Echavarri-Erasun, C., Pollmann, S., Britt, D. & Rubio, L. M. (2016) Electron Paramagnetic Resonance Characterization of Three Iron - Sulfur Clusters Present in the Nitrogenase Cofactor Maturase NifB from *Methanocaldococcus infernus*, *J Am Chem Soc.* **138**, 7468 – 7471.
 116. Fay, A. W., Wiig, J. A., Chi, Chung, L. & Hu, Y. (2015) Identification and characterization of functional homologs of nitrogenase cofactor biosynthesis protein NifB from methanogens, *Proceeding of the National Academy of Sciences of the United States of America.* **112**, 14829 - 14833.
 117. Arragain, S., Jiménez-Vicente, E., Scandurra, A. A., Burén, S., Rubio, L. M. & Echavarri-Erasun, C. (2017) Diversity and functional analyses of the FeMo-cofactor maturase NifB for its implementation in eukaryotic cells. *Frontiers Plant Science* (Submitted).
 118. Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.* **44**, 279 – 285.
 119. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol.* **539**, 1 – 6.
 120. Trifinopoulos, J., Nguyen, L. T., Von Haeseler, A. & Minh, B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis, *Nucleic Acids Res.* **44**, 232 – 235.
 121. Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, *Systematic Biology.* **56**, 564 - 577.
 122. Letunic, I. & Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation, *Bioinformatics.* **23**, 127 – 128.
 123. Bornhorst, J. A. & Falke, J. J., (2000) Purification of Proteins Using Polyhistidine Affinity Tags, *Methods Enzymol.* **326**, 245 – 254.
 124. Shah V. K. & Brill W. J. (1973) Nitrogenase IV. Simple method of purification to homogeneity of nitrogenase components from *Azotobacter vinelandii*, *Biochimica et Biophysica Acta journal.* **305**, 445 - 454.
 125. Stewart, W. D., Fitzgerald, G. P. & Burris, R. H. (1967) In situ studies on N₂ fixation using the acetylene reduction technique, *Proceeding of the National Academy of Sciences of the United States of America.* **58**, 2071 - 2078.
 126. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.*
 127. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid, *Anal Biochem.* **150**, 76 - 85.
 128. Fish, W. W. (1988) Methods in Enzymology. **158**, 357- 364.
 129. Beinert, H. (1983) Semi-micro methods for analysis of labile sulfide and of labile sulfide plus sulfane sulfur in unusually stable iron-sulfur proteins, *Anal Biochem.* **131**, 373 – 378.
 130. Iwig, D. F. & Booker, S. J. (2004) Insight into the Polar Reactivity of the Onium Chalcogen Analogues of S-Adenosyl-L-methionine, *Biochemistry.* **43**, 13496 – 13509.

131. Cieslak, J., Focia, P. J. & Gross, A. (2010) Electron spin echo envelope modulation (ESEEM) reveals water and phosphate interactions with the KcsA potassium channel, *Biochemistry*. **49**, 1486 – 1494.
132. Harmer, J., Mitrikas, G. & Schweiger, A. (2009) Advanced Pulse EPR Methods for the Characterization of Metalloproteins, *High Resolution EPR*, Springer, **28**, 13 – 61.
133. Stoll, S. & Schweiger, A. (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR, *J Magn Reson.* **178**, 42 - 55.
134. Gambarelli S., Mulliez, E. & Fontecave, M. (2010) Metals in biology-applications of High-resolution EPR to Metalloenzymes, Hanson, Graeme, Berliner, Lawrence (éd.) *Biol Magn Reson.* **29**, 53-75.
135. Ducruix, A. & Giegé, R. (1999) Crystallization of Nucleic Acids and Proteins, A Practical Approach, Second Edition.
136. Vernede, X. & Fontecilla – Camps, J. C. (1999) A method to stabilize reduced and or gas-treated protein crystals by flash - cooling under a controlled atmosphere, *J Appl Crystallogr.* **32**, 505 - 509.
137. Zuckerkandl, E. & Pauling, L. (1962) Molecular disease, evolution and genetic heterogeneity, *Horizons in Biochemistry*, 189 - 225.
138. Zuckerkandl E., (1965) Remarks on the Evolution of Polynucleotides Compared to That of Polypeptides, *Bull Soc Chim Biol.* **47**, 1729 - 1730.
139. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M. & Peters, J. W. (2011) A late methanogen origin for molybdenum-dependent nitrogenase, *Geobiology*. **9**, 221, 232.
140. Jeanthon, C., L'Haridon, S., Reysenbach, A. L., Vernet, M., Messner, P., Sleytr, U. B. & Prieur, D., (1998) *Methanococcus infernus* sp. nov., a novel hyperthermophilic lithotrophic methanogen isolated from a deep-sea hydrothermal vent, *Int J Syst Bacteriol.* **48**, 913 - 919.
141. Nakamura, M., Saeki, K. & Takahashi, Y. (1999) Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster, *J Biochem.* **126**, 10 - 18.
142. Averill, B. A. (2007) Fe-S and Mo-Fe-S clusters as models for the active site of nitrogenase, Copper, Molybdenum, *Copper, Molybdenum, and Vanadium in Biological Systems.* **53**, 59 - 103.
143. Hanson, G. & Berliner, L. (2010) Metals in Biology: Applications of High-Resolution EPR to Metalloenzymes, *Biol Magn Reson.* Springer, **29**, 11 – 44.
144. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. (2014) Radical S-adenosylmethionine enzymes, *Chem Rev.* **114**, 4229 - 4317.
145. Wollers, S., Layer G., Garcia-Serres R., Signor L., Clemancey M., Latour J. M., Fontecave M. & Ollagnier de Choudens S. (2010) Iron-Sulfur (Fe-S) Cluster Assembly, The SufBCD complex is a new type of Fe – S Scaffold with a flavin redox cofactor, *J Biol Chem.* **285**, 23331 – 23341.
146. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W. & Rubio, L. M. (2010) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Letters.* **584**, 1487 - 1492.
147. Boyd, E. S., Garcia Costas, A. M., Hamilton, T. L., Mus, F. & Peters, J. W. (2015) Evolution of Molybdenum Nitrogenase during the Transition from Anaerobic to Aerobic Metabolism, *J Bacteriol.* **197**, 1690 - 1699.
148. Cammack, R. & MacMillan, F. (2010) Metals in Biology: Electron magnetic resonance of iron-sulfur proteins in electron transfer chains - resolving complexity, *BIMR.* **29**, 11 – 44.
149. Peters, J. W., Lanzilotta, W. N., Lemon, B. J. & Seefeldt, L. C. (1998) X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 angstrom resolution, *Science.* **282**, 1853 – 1858.
150. Bak, D. W. & Elliott, S. J. (2014) Alternative FeS cluster ligands: tuning redox potentials and chemistry, *Curr Opin Chem Biol.* **19**, 50 – 58.
151. Wiig, J. A., Hu, Y. & Ribbe, M. W. (2015) Refining the pathway of carbide insertion into the nitrogenase M-cluster, *Nat Commun.* **6**, 8034.
152. Marinoni, E. (2011) Dissertation: Maturation de sites métalliques de protéines par les protéines à radical S-Adenosyl-L-méthionine et la machinerie de fabrication des centres fer-soufre. Université de Grenoble – France, Sciences agricoles.