Addressing Grand Challenges in Rice Productivity and Sustainability through Biotechnology

Wenshu He

Universitat de Lleida

TESI DOCTORAL

Addressing Grand Challenges in Rice Productivity and Sustainability through Biotechnology

Wenshu He

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en Ciència i Tecnologia Agrària i Alimentària

Director/a
Paul Christou
Teresa Capell

Tutor/a
Paul Christou

2023

Front cover paintings by Teresa Capell

“Pink narcissus” (2015) oil on canvas 19x27 cm

“Narcissus bulb with pink gene” (2016) oil on canvas 19x27 cm

Back cover painting by Teresa Capell

“Colored grains” oil on canvas 19x27 cm

Supervisors: **Paul Christou**

Teresa Capell

Department of Plant Production and Forestry Science

School of Agricultural and Forestry Engineering

University of Lleida

2023

Signatures P. Christou T. Capell

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Dr Paul Christou and Dr Teresa Capell for their continuous support, valuable guidance, and enthusiastic encouragement in science and beyond. It has been an absolute pleasure to have your guidance in all the time of research and writing of this thesis. Without your guidance and encouragement this PhD would not have been possible.

I would like to thank Dr Luis Rubio [Universidad Politécnica de Madrid (UPM), Centro de Biotecnología y Genómica de Plantas (CBGP) – Instituto Nacional de Investigaciones Agrarias (INIA), Spain] for making available facilities for nitrogenase analysis and Dr Stefan Burén and Dr Xi Jiang (UPM, CBGP-INIA, Spain) for their kind help in carrying out experiments and data analysis on nitrogenase activity. I would also like to thank Dr Greta Nölke (Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Germany) for her kindness and contribution to my research.

I gratefully acknowledge the University of Lleida, especially the “Departament de Producció Vegetal i Ciència Forestal” for offering me the facilities for my PhD research. I would like to thank the “Agencia de Gestión de Ayudas Universitarias y de Investigación” (AGAUR, Spain) for a PhD fellowship. Also, thanks to the Bill and Melinda Gates Foundation and the European Union's Horizon 2020 research and innovation programme for their financial support.

I would like to thank Dr Changfu Zhu for helping me to get started in the lab, Dr Ludovic Bassie for helping with techniques, Núria Gabernet for organizing the official paperwork and Jaume Capell for taking care of all my plants. To my colleagues Can and Ashwin for helping me with all the experiments. To Xin Huang and Andrea for their kind support whenever I needed it and for being more than friends. To Vicky and Guillermo for being incomparable lab colleagues. To Hugo and Jessica for helping me with tissue culture. To all the lab colleagues from then and now: Xin Jin, Amaya, Maria, Derry, Pedro, Jaume, Jose, Alex, and Sheila, thank you for all the assistance and encouragement.

I would like to thank my entire family, mother and father, aunt, and uncle, thank you for all your love and support. Also, I would like to thank my boyfriend Herui, for understanding and supporting me when I needed it most. All my love and thanks for everything. Words are powerless to express my gratitude.

SUMMARY

This thesis focuses on addressing the grand challenges of contemporary agriculture in rice productivity and sustainability through biotechnology. The overall aims of my thesis were to (a) engineer a heterologous biological nitrogen fixation pathway to incorporate biological nitrogen fixation in rice and (b) engineer an ectopic oxygen scavenging pathway to reduce photorespiration and improve photosynthetic efficiency. Both objectives are congruent with improving rice productivity and sustainability in a durable and environmentally friendly manner and are consistent with key major Sustainability Development Goals as articulated by the UN charter on the environment.

Biological nitrogen fixation is catalyzed by nitrogenase. This enzyme is composed of molybdenum-iron (MoFe) protein (NifDK) and Fe protein (NifH). NifB synthesizes precursors of the nitrogenase active site cofactor (FeMo-co) that binds to nitrogen and reduces it. I generated transgenic rice plants expressing two different forms of NifB from the archaean *Methanocaldococcus infernus* and *Methanothermobacter thermautotrophicus*, together with *Azotobacter vinelandii* FdxN; NifH from *A. vinelandii* together with *A. vinelandii* NifM, NifS *and* NifU; NifH from *Hydrogenobacter thermophilus* with *A. vinelandii* NifM; and three *A. vinelandii* NifD variants ($NifD^{Av}$-Y99K, $NifD^{Av}$-Y99Q, $NifD^{Av}$-Y99QY100T) in separate sets of experiments as detailed in the subsequent chapter of the thesis. I determined that both NifB and NifH proteins accumulated as soluble proteins *in planta*. The purified NifB proteins were functional in the *in vitro* FeMo-co synthesis assay. Purified *H. thermophilus* NifH proteins were able to carry out the fundamental roles of the Fe protein required to engineer nitrogen fixation, including electron transfer to the nitrogenase component MoFe protein and the biosynthesis of FeMo-co. The three $NifD^{Av}$ variants were susceptible to cleavage by mitochondrial processing peptidase, resulting in the formation of only truncated proteins. Further studies need to focus on understanding the basis of this instability and producing intact NifD in rice. The expression of biologically active nitrogenase components in transgenic rice represents a critical step toward achieving biological nitrogen fixation *in planta* and represents the only report to date which resulted in the recovery of a stably transformed major crop expressing nitrogenase component enzymes.

Photorespiration reduces overall photosynthetic efficiency because it results in up to 50% loss of carbon fixed by photosynthesis. Reducing photorespiration, therefore, is a potential strategy to increase photosynthetic efficiency. I co-introduced *Lactobacillus*

buchneri lactate oxidase (*Lb*LOX), *Escherichia coli* lactate dehydrogenase (*Ec*LDH) and *Oryza sativa* catalase (*Os*CAT) into rice to establish a novel oxygen scavenging pathway to reduce photorespiration. Expression of the three transgenes was confirmed at mRNA levels. Only *Ec*LDH protein accumulation was detectable in callus and regenerated plants. Detection of *Lb*LOX and *Os*CAT protein accumulation was probably hampered by the detection tag used to identify the protein. Further experiments will be required to confirm the accumulation of both proteins by enzymatic assays and the replacement of the assay tags.

RESUMEN

Esta tesis doctoral está enfocada en dar respuesta a los grandes desafíos de la agricultura contemporánea, concretamente la productividad y la sostenibilidad del cultivo del arroz mediante la aplicación de técnicas biotecnológicas como la ingeniería genética. Los objetivos generales de esta tesis han sido (a) introducir una vía de fijación de nitrógeno heteróloga en arroz para que este sea capaz de fijar el nitrógeno biológico e (b) introducir una vía ectópica de captación de oxígeno para reducir la fotorespiración y mejorar la eficiencia fotosintética de las plantas de arroz. Ambos objetivos son congruentes con la mejora de la productividad y la sostenibilidad del arroz de una manera duradera y respetuosa con el medio ambiente y son coherentes con los principales Objetivos de Desarrollo Sostenible, tal como se articula en la carta de las Naciones Unidas sobre el medio ambiente.

La fijación de nitrógeno biología esta catalizada por la enzima nitrogenasa. Esta enzima está ligada al complejo hierro-molibdeno (MoFe), denominado proteína NifDK y a la ferritina, siendo denominado proteína NifH. La proteína NifB sintetiza precursores del cofactor del lugar activo de la nitrogenasa (FeMo-co) que se une al nitrógeno reduciéndolo. En esta tesis doctoral he regenerado plantas de arroz transgénicas que expresan dos formas diferentes de la proteína NifB, provenientes de las arqueobacterias *Methanocaldococcus infernus* y *Methanothermobacter thermautotrophicus*. Estas proteínas han sido combinadas con FdxN de *Azotobacter vinelandii*; NifH de *A. vinelandii* juntamente con NifM, NifS *y* NifU de *A. vinelandii*; y NifH de *Hydrogenobacter thermophilus* con NifM de *A. vinelandii*; además de tres variantes de NifD *A* ($NifD^{Av}$-Y99K, $NifD^{Av}$-Y99Q, $NifD^{Av}$-Y99QY100T) en experimentos detallados en cada capítulo. Los resultados que he obtenido han sido que ambas proteínas NifB y NifH se acumulan en forma de proteína soluble *in planta*. Las proteínas NifB purificadas son funcionales en el ensayo *in vitro* de síntesis de FeMo-co. La proteína NifH de *H. thermophilus* es capaz de desarrollar los papeles fundamentales que las proteínas complejadas a Fe requieren para la funcionalidad de la fijación de nitrógeno, incluida la transferencia de electrones a la proteína MoFe componente de la nitrogenasa y para la biosíntesis de FeMo-co. Las tres variantes $NifD^{Av}$ han demostrado ser susceptibles de cortes por la peptidasa de procesado mitocondrial, resultando solo en la formación de proteínas truncadas. Estudios posteriores se tienen que enfocar en elucidar el porqué de la inestabilidad de la proteína y de cómo producir NifD intacta en arroz. La expresión de componentes biológicamente activos de nitrogenasa en arroz transgénico representa un

paso criticó hacia conseguir la fijación de nitrógeno biológico *in planta* y representa el único informe hasta ahora que ha conseguido la regeneración de un cultivo mayoritario con la presencia de componentes enzimáticos de la nitrogenasa.

La fotorespiración reduce la eficiencia fotosintética porque resulta en hasta un 50% de pérdida del carbón fijado durante la fotosíntesis. En consecuencia, la reducción de la fotorespiración puede ser una estrategia potencial para incrementar la eficiencia fotosintética. En esta tesis doctoral he co-introducido la enzima lactato oxidasa de *Lactobacillus buchneri* (*Lb*LOX), la enzima lactato deshidrogenasa de *Escherichia coli* (*Ec*LDH) y la catalasa de *Oryza sativa* (*Os*CAT) en arroz, estableciendo una nueva ruta de captación de oxígeno para reducir la fotorespiración. La co-expresión de los tres genes fue confirmada a nivel de ARNm. Pero solamente la acumulación de la proteína *Ec*LDH fue detectable a nivel de tejido de callo y en las plantas regeneradas. La detección de la acumulación de las proteínas *Lb*LOX y *Os*CAT fue probablemente interferida por la tag-detectora de proteínas utilizada en el experimento. Se utilizarán análisis de actividad enzimática para confirmar la acumulación de proteínas en ambos casos y quizás se reemplacen las tags-detectoras de proteínas utilizadas.

RESUM

Aquesta tesis doctoral està enfocada en donar resposta als grans desafiaments de l'agricultura contemporània, concretament la productivitat i la sostenibilitat del cultivo de l'arròs mitjançant l'aplicació de tècniques biotecnològiques com l'enginyeria genètica. Els objectius generales d'aquesta tesis han estat (a) introduir una via de fixació de nitrogeno heteròloga en arròs per a que aquest sigui capaç de fixar el nitrogen biològic i (b) introduir una via ectòpica de captació d'oxigen per a reduir la fotorespiració i millorar la eficiència fotosintètica de les plantes de arròs. Tots dos objectius son congruents amb la millora de la productivitat i la sostenibilitat de l'arròs d'una forma duradora i respectuosa amb el medi ambient i son coherents amb els principals Objectius de Desenvolupament Sostenible, tal com s'articula en la carta de las Nacions Unides sobre el medi ambient.

La fixació de nitrogen biològica esta catalitzada por l'enzim nitrogenasa. Aquest enzim està lligat al complex ferro-molibdè (MoFe), denominada proteïna NifDK i a la ferritina, essent denominada proteïna NifH. La proteïna NifB sintetitza precursors del cofactor del lloc actiu de la nitrogenasa (FeMo-co) que s'uneix al nitrogen reduint-lo. En aquesta tesis doctoral he regenerat plantes d'arròs transgèniques que expressen dues formes diferents de la proteïna NifB, prevenients de les arqueobacteries *Methanocaldococcus infernus* i *Methanothermobacter thermautotrophicus*. Aquestes proteïnes han estat combinades amb FdxN de *Azotobacter vinelandii*; NifH de *A. vinelandii* juntament amb NifM, NifS i NifU de *A. vinelandii*; i NifH de *Hydrogenobacter thermophilus* amb NifM de *A. vinelandii*; a més a més de tres variants de NifD *A* ($NifD^{Av}$-Y99K, $NifD^{Av}$-Y99Q, $NifD^{Av}$-Y99QY100T) en experiments detallats en cada capítol. Els resultats que he obtingut han estat que ambdues proteïnes NifB i NifH s'acumulen en forma de proteïna soluble *in planta*. Les proteïnes NifB purificades son funcionals en l'assaig *in vitro* de síntesis de FeMo-co. La proteïna NifH de *H. thermophilus* és capaç de desenvolupar els papers fonamentals que las proteïnes complexades a Fe requereixen per a la funcionalitat de la fixació de nitrogen, inclosa la transferència de electrons a la proteïna MoFe component de la nitrogenasa i per a la biosíntesis de FeMo-co. Les tres variants $NifD^{Av}$ han demostrat ser susceptibles de talls per la peptidasa de processat mitocondrial, resultant únicament en la formació de proteïnes truncades. Estudis posteriors cal que s'enfoquin en elucidar el perquè de la inestabilitat de la proteïna i de cóm produir NifD intacta en arròs. La expressió de components biològicament actius de nitrogenasa en arròs transgènic representa un pas crític cap a aconseguir la fixació de nitrogen biològic *in planta* i quest representa l'únic

informe fins ara que ha aconseguit la regeneració d'un cultivo majoritari amb la presencia de components enzimàtics de la nitrogenasa.

La fotorespiració redueix l'eficiència fotosintètica perquè resulta a fins un 50% de pèrdua del carbó fixat durant la fotosíntesis. En conseqüència, la reducció de la fotorespiració pot ser una estratègia potencial per a incrementar l'eficiència fotosintètica. En aquesta tesis doctoral he co-introduit l'enzim lactat oxidasa de *Lactobacillus buchneri* (*Lb*LOX), l'enzim lactat deshidrogenasa de *Escherichia coli* (*Ec*LDH) i l'enzim catalasa de *Oryza sativa* (*Os*CAT) en arròs, establint una nova ruta de captació de oxigen per a reduir la fotorespiració. La co-expressió dels tres gens va ser confirmada a nivell de ARNm. Però solament l'acumulació de la proteïna *Ec*LDH va ser detectable a nivell de teixit de call i en les plantes regenerades. La detecció de l'acumulació de les proteïnes *Lb*LOX i *Os*CAT va ser probablement interferida por la tag-detectora de proteïnes utilitzada en l'experiment. S'utilitzarà anàlisis d'activitat enzimàtica per a confirmar l'acumulació de proteïnes en tots dos casos i potser és substitueixin les tags-detectores de proteïnes utilitzades.

TABLE OF CONTENTS

INDEX OF FIGURES

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

INDEX OF TABLES

ABBREVIATIONS

2-ME	2-mercaptoethanol
2PG	2-phosphoglycolate
3PGA	3-phosphoglyceric acid
ANOVA	One-way analysis of variance
ARA	Acetylene reduction assay
*At*GLO	*Arabidopsis thaliana* glycolate oxidase
BCT1	Bicarbonate transporter1
BicA	Bicarbonate uptake A
BNF	Biological nitrogen fixation
BST1-3	Bestrophin-like proteins 1-3
CA	Carbonic anhydrase
CAH3	Carbonic anhydrase 3
CAM	Crassulacean acid metabolism
CCM	CO_2-concentrating mechanism
CH2-THF	5,10-methylenetetrahydrofolate
C_i	Inorganic carbon
*Cm*MS	*Cucurbita maxima* malate synthase
Cox4	*Saccharomyces cerevisiae* cytochrome c oxidase subunit IV
CTP	Chloroplast stromal targeting peptide *S. tuberosum* Rubisco small subunit
DEPC	Phosphoenolpyruvate carboxylase
DTH	Sodium dithionite
*Ec*CAT	*Escherichia coli* catalase
*Ec*GCL	*Escherichia coli* glyoxylate carboligase
*Ec*GDH	*Escherichia coli* glycolate dehydrogenase
*Ec*LDH	*Escherichia coli* lactate dehydrogenase
*Ec*NifS^{Av}	*Saccharomyces cerevisiae*-derived *Azotobacter vinelandii* NifS
*Ec*NifU^{Av}	*Escherichia coli*-derived *Azotobacter vinelandii* NifU
*Ec*TSR	*Escherichia coli* tartronic semialdehyde reductase
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
FdxN	Ferredoxin
FdxN^{Av}	*Azotobacter vinelandii FdxN*
Fe	Iron
FeFe-co	Iron-Iron cofactor
FeMo-co	Iron-molybdenum cofactor
FeV-co	Iron-vanadium cofactor
GOGAT	Glutamate synthase
GS	Glutamine synthase

HA	Hemagglutinin
HCO_3^-	Bicarbonate
HLA3	High-light activated 3
Hpt	Hygromycin phosphotransferase gene
IFS	Isoflavone synthase
*Lb*LOX	*Lactobacillus buchneri* lactate oxidase
LCI1	low-CO2-inducible protein 1
LCIA	low-CO2-inducible protein A
LCIB	low-CO2-inducible protein B
LCIC	low-CO2-inducible protein C
LNP	Lectin nucleotide phosphohydrolase
LSD	Fisher's least significant difference
LYK3	LysM receptor kinase 3
Mo	Molybdenum
MoFe	Molybdenum iron
MPP	Mitochondrial processing peptidase
N	Nitrogen
NAD(P)-MDH	NAD(P)-malate dehydrogenase
NADH-MDH	NADH-malate dehydrogenase
NADP-ME	NADP-malic enzyme
NafY	Nitrogenase gamma subunit
NFP	Nodulation factor perception
NFR	Nodulation factor receptor
NFR1	Nodulation factor receptor 1
NFR5	Nodulation factor receptor 5
Nif	Nitrogen fixation
NifB	Nitrogenase cofactor maturase NifB
$NifB^{Mi}$	*Methanocaldococcus infernus NifB*
$NifB^{Mt}$	*Methanothermobacter thermautotrophicus NifB*
NifD	Nitrogenase molybdenum-iron protein subunit alpha
$NifD^{Av}$	*Azotobacter vinelandii NifD*
NifDK	Nitrogenase molybdenum iron protein
$NifDK^{Av}$	*Azotobacter vinelandii* NifDK
NifE	Nitrogenase iron-molybdenum cofactor biosynthesis protein NifE
$NifEN^{Av}$	*Azotobacter vinelandii* NifEN
NifF	Flavodoxin
NifH	Nitrogenase Iron protein
$NifH^{Av}$	*Azotobacter vinelandii NifH*
$NifH^{Ht}$	*Hydrogenobacter thermophilus NifH*
NifK	Nitrogenase molybdenum-iron protein subunit beta

NifM	Putative peptidyl-prolyl cis-trans isomerase NifM
NifM^{Av}	*Azotobacter vinelandii NifM*
NifN	Nitrogenase iron-molybdenum cofactor biosynthesis protein NifN
NifQ	Nitrogen fixation protein NifQ
NifS	Cysteine desulfurase
NifS^{Av}	*Azotobacter vinelandii NifS*
NifT	Nitrogen fixation protein NifT
NifU	Fe-S cluster assembly protein NifU
NifU^{Av}	*Azotobacter vinelandii NifU*
NifV	Homocitrate synthase
NifW	Nitrogenase-stabilizing/protective protein NifW
NifX	Nitrogen fixation protein NifX
NifY	Nitrogen fixation protein NifY
NifZ	Nitrogen fixation protein NifZ
Nod	Nodulation
NodD	Nodulation factor D
NUE	Nitrogen-use efficiency
OAA	Oxaloacetate
*Os*CAT	*Oryza sativa* catalase
*Os*CATC	*Oryza sativa* catalase C
*Os*FdxN^{Av}	*Oryza sativa*-derived *Azotobacter vinelandii* FdxN
*Os*GLO3	*Oryza sativa* glycolate oxidase 3
*Os*NifB	*Oryza sativa*-derived *Methanocaldococcus infernus* and *Methanothermobacter thermautotrophicus* NifB
*Os*NifB^{Mi}	*Oryza sativa Methanocaldococcus infernus* NifB
*Os*NifB^{Mt}	*Oryza sativa-derived Methanothermobacter thermautotrophicus* NifB
*Os*NifH	*Oryza sativa*-derived *Azotobacter vinelandii* and *Hydrogenobacter thermophilus* NifH
*Os*NifH^{Av}	*Oryza sativa-derived Azotobacter vinelandii NifH*
*Os*NifH^{Ht}	*Oryza sativa-derived Hydrogenobacter thermophilus* NifH
*Os*OXO3	*Oryza sativa* oxalate oxidase 3
PDH	Pyruvate dehydrogenase
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PMSF	Phenylmethanesulfonylfluoride
PPDK	Pyruvate phosphate dikinase
RC2	Chloroplast stromal targeting peptide *O. sativa* Rubisco small subunit
RT-qPCR	Real-time quantitative reverse transcription PCR
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
SAM	S-adenosyl-L-methionine

SBPase	Sedoheptulose 1,7-bisphosphate phosphatase
SbtA	Sodium-bicarbonate transporter A
*Sc*NifB^{Mi}	*Saccharomyces cerevisiae*-derived *Methanocaldococcus infernus* NifB
*Sc*NifB^{Mt}	*Saccharomyces cerevisiae*-derived *Methanothermobacter thermautotrophicus* NifB
*Sc*NifH^{Av}	*Saccharomyces cerevisiae*-derived *Azotobacter vinelandii* NifH
SD	Standard deviation
SDS	Sodium dodecylsulfate
STAC	Strep-tag affinity chromatography
Su9	*Neurospora crassa* mitochondrial ATPase subunit 9
TPI	Triose-phosphate isomerase
TS	Twinstrep
V	Vanadium
*Zj*PCK	*Zoysia japonica* phosphoenolpyruvate carboxykinase
*Zm*GLK	*Zea mays* GOLDEN2-LIKE transcription factor
Ω	CO_2/O_2 specificity of Rubisco

Chapter I.
General introduction

1.1 General introduction

Rice is the second most important cereal crop in the world after corn (Soto-Gómez and Pérez-Rodríguez 2022). Rice is the primary food staple, for over half of the world's population, particularly in Asia and South America. The 2022/23 global rice consumption forecast was 516.9 million tons. However, the global production forecast for the same period was 503.3 million tons (Childs and LeBeau 2022). Rice production has increased at an average annual growth rate of 0.45% over the period 2012-2021 (Childs 2012; Childs 2021). However, this rate of increase is not sufficient to keep pace with the growing global demand, which is expected to increase by 0.9% per year over the next decade (Mohidem et al. 2022). Climate change is also projected to increase global temperatures by up to 2°C in the next decades (Intergovernmental Panel on Climate Change, https://www.ipcc.ch). Similarly, the loss of agricultural land, water shortages, biotic and abiotic stresses, and other factors cause unprecedented challenges to global food security. To address these pressing challenges, improving crop productivity is critical. New approaches and strategies are urgently needed to further improve productivity. Nitrogen is an important component of chlorophyll, proteins, nucleic acids, and many metabolites and is an essential element necessary for the growth and development of all plants, including staple crops. In addition, plants assimilate carbon dioxide and water into organic matter through photosynthesis to complete carbon fixation. Thus, nitrogen fixation and photosynthesis are crucial to crop growth, development, and yield.

1.1.1. Engineering biological nitrogen fixation in crops

1.1.1.1 Importance of nitrogen in plants

Nitrogen (N) is the element required in the largest amount for plant growth and development. It is an integral constituent of proteins, DNA, RNA, chlorophyll, phytohormones and secondary metabolites, and accounts for about 1-5% of total plant dry matter (Hawkesford et al. 2012). N promotes leaf area growth, the key to obtaining high photosynthetic capacity. Depending on the supply of N, the difference among mature leaf size was shown to be up to 4-fold in potato and 3-fold in maize (Vos et al. 2005). A minimum nitrogen content is required in the development of new leaf area, and correspondingly, insufficient nitrogen supply to crops results directly in a decrease in final area of individual leaves (Sinclair 1990). Since photosynthetic enzymes account for a substantial portion of the leaf N and N is a constituent of chlorophyll, leaf CO_2

assimilation rates and leaf nitrogen content are also highly correlated (correlation coefficient over 0.75) (Sinclair and Horie 1989).

Furthermore, nitrogen plays an important role in plant disease defences. Firstly, N availability regulates cell structure and composition through its effects on primary and secondary metabolism in plants. A negative correlation was found between N input and the thickness of the physical barriers of lignin and waxy epidermis in plants (Sun et al. 2020). With the reinforcement of the cuticle, plants effectively prevent the penetration of pathogenic bacteria, thereby reducing the incidence of diseases (Kumar et al. 2016). In addition, N-mediated biochemical defences are associated with plant metabolites including phytoalexins, antimicrobial proteins, amino acids and organic acids, as well as defence-related (phenylalanine ammonolyses) and antioxidant enzymes (superoxide dismutase and catalase) (Sun et al. 2020).

1.1.1.2 Biological nitrogen fixation

Biological nitrogen fixation (BNF) involves the conversion of N_2 to NH_3 by the multi-component enzyme, nitrogenase (Halbleib and Ludden 2000). The overall enzymatic reaction is:

$$N_2 + 8H^+ + 8e^- + 16ATP \xrightarrow{\text{Nitrogenase}} 2NH_3 + H_2 + 16ADP + 16\ P_i$$

BNF is carried out by a group of prokaryotes belonging to the bacteria or archaea domains. No eukaryotes at present have been reported to convert N_2 into a biologically useful form directly (Aryal et al. 2022). BNF is the largest natural source of exogenous N to ecosystems (Sullivan et al. 2014). The BNF bacteria are classified into three main groups, namely free-living (e.g., *Azotobacter*), endophytic (e.g., *Azoarcus*), and symbiotic (e.g., *Rhizobium*) (Mus et al. 2016). Endophytic and symbiotic bacteria rely on their associated host plants to provide a carbon source, whereas free-living (rhizosphere-associative) bacteria typically gain their energy by oxidizing organic molecules released by roots, other organisms, or from decomposition of organic matter (Mus et al. 2016).

There are three main classes of nitrogenases with the same basic two component proteins, but different metal composition in the enzyme active site [iron-molybdenum, (named as Mo-nitrogenase), iron-vanadium (V-nitrogenase) and iron-only (Fe-nitrogenase)] (Rubio and Ludden 2008). The Mo-nitrogenase is the most abundant in nature and it has been the subject of most studies by far. The Mo-nitrogenase subunits are the iron (Fe) protein and

the molybdenum iron (MoFe) protein (**Figure 1.1**) (Ledbetter 2018). The nitrogen fixation (*Nif*) gene *NifD* and *NifK* encode for the MoFe protein and *NifH* encodes for the Fe protein (Rubio and Ludden 2008). MoFe protein is a $\alpha_2\beta_2$ heterotetramer that contains two metal clusters: the P-cluster, a [8Fe-7S] cluster, and the Iron-molybdenum cofactor (FeMo-co), a [7Fe-Mo-9S-C-R homocitrate] cluster (Burén et al. 2020). Fe protein is a dimeric ATPase which transfers electrons to the MoFe protein (Burén et al. 2020). In addition to N_2, nitrogenase is able to reduce a number of other small molecules such as acetylene and hydrazine (Pham and Burgess 1993). Therefore, nitrogenase activity can be determined directly by measuring the assimilation of $^{15}N_2$ or indirectly by measuring the reduction of acetylene. The acetylene reduction assay (ARA) is the most commonly used method and the BNF activity is quantified by gas chromatography (Senthilkumar et al. 2021).

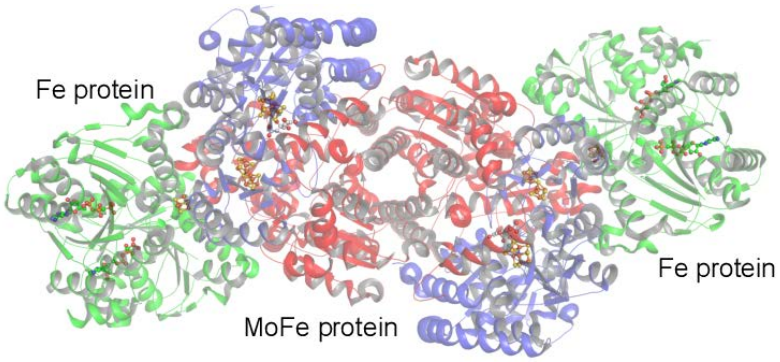

Figure 1.1 Schematic representation of Mo-nitrogenase complex from *Azotobacter vinelandii*. The Fe protein is shown in green and the MoFe protein in blue (α-subunit, NifD) and red (β-subunit, NifK). Modified from Ledbetter (2018).

1.1.1.3 Engineering biological nitrogen fixation strategies in plants

Crop production requires high levels of N in the form of chemical fertilizers. The global N fertilizer demand is approximately 110.8 million metric tons (Fernández 2022). However, nitrogen-use efficiency (NUE) is only 35% which implies that 65% of all applied N fertilizer is lost to the environment (Omara et al. 2019). This has a negative impact on the environment because of groundwater contamination and soil acidification, which in turn also impact human and animal health negatively. Engineering BNF in plants is an attractive strategy to reduce dependence on N fertilizers (Bloch et al. 2020). There are currently two different strategies being pursued to engineer crops to fix atmospheric nitrogen: 1) introduction of nitrogenase-encoding genes into the plant genome, without

the involvement of BNF bacteria; 2) induction of symbiosis between plants and BNF bacteria, by transferring the genes leguminous plants utilize to encourage symbiosis signalling pathways to form nodules on the roots (Bloch et al. 2020).

The complexity of nitrogenase biosynthesis and sensitivity of its component enzymes to oxygen are major challenges hindering the effective implementation of strategy 1. The biosynthesis of nitrogenase requires multiple genes. *The major Nif genes from Azotobacter vinelandii include NifH, NifD, NifK, NifT, NifY, NifE, NifN, NifX, NifU, NifS, NifV, NifW, NifZ, NifM, and NifF* (Jacobson et al. 1989). *NifH*, *NifD*, *NifK*, *Nif*E, *NifN*, and *NifB* are present in nearly all known diazotrophs, Dos Santos et al. (2012) proposed that these six highly conserved genes are the minimum set of genes for nitrogen fixation (Dos Santos et al. 2012). To meet the requirement of the high energy demand, in the form of ATP, for BNF, chloroplasts and mitochondria have been considered as potential subcellular locations for engineering nitrogenase components. Chloroplasts produce large amounts of ATP and reducing power in the form of NADPH through photosynthetic electron-transfer reactions. Although the sensitivity of nitrogenase to O_2, a major product of photosynthesis, would seem to preclude nitrogenase expression in chloroplasts. Photosynthetic cyanobacteria can also efficiently fix N_2 by separating nitrogen fixation from photosynthesis at two different periods of diurnal cycle or by compartmentalizing nitrogenase in heterocysts (Bogorad 2000). *NifH* and *NifM* were introduced into the tobacco chloroplast genome (Ivleva et al. 2016). The resulting transplastomic plants exhibited slight but reproducible Fe protein activity at a low O_2 concentration (10% O_2, which is half of the O_2 in atmosphere). However, the NifH protein isolated from plants growing in ambient air was inactive (Ivleva et al. 2016). In transiently expressed tobacco, NifH, NifM, NifU, and NifS proteins were imported into the chloroplast stroma by chloroplast targeting peptides (Eseverri et al. 2020). NifH protein was isolated from tobacco plants collected at the end of the dark period at ambient air condition. The isolated tobacco NifH protein showed 16% activity compared to *A. vinelandii* NifH after reconstitution *in vitro* (Eseverri et al. 2020). The respiration results in large amounts of energy and a low O_2 level inside mitochondria. This respiratory protection mechanism was found in *A. vinelandii* (Poole and Hill 1997). In addition, mitochondria contain an [Fe-S] cluster assembly machinery which is very similar to the NifUS system, which is involved in the early steps of nitrogenase metal-cluster biosynthesis (Lill and Mühlenhoff 2008). To date, Nif proteins (NifB, NifD, NifH, NifK, NifM, NifS, NifU etc.) were targeted to mitochondria using mitochondrial targeting peptides (Burén et al. 2017; Burén

et al. 2019; Jiang et al. 2021; Jiang et al. 2022). NifB, isolated from aerobically cultured yeast and ambient air cultured tobacco (transient expression) supported the FeMo-co synthesis *in vitro*. In these experiments, NifB was co-expressed with NifS, NifU and FdxN (Burén et al. 2017; Burén et al. 2019; Jiang et al. 2022). NifH isolated from yeast mitochondria (also expressing NifM, NifS and NifU) carried out the functions of electron transfer for substrate reduction, maturation of P-clusters, and biosynthesis of FeMo-co (Jiang et al. 2021). However, stable expression of a functional nitrogenase has not been achieved in cereals.

The process of nitrogen fixation through symbiosis is a highly complicated process involving multiple events and their regulation by both host and rhizobia including recognition of nodulation (Nod) factor, organogenesis of the root nodule and bacterial infection (Huisman and Geurts 2020). The recognition of Nod factors is an essential step to initiate symbiotic nitrogen fixation in plants. Nod factors are signalling molecules secreted by rhizobia and serve a key role in the induction of early responses (Long 1996). The rhizobia *Nod* gene expression is initiated when bacteria sense specific molecules, such as flavonoids, secreted by plant roots (Zuanazzi et al. 1998). Flavonoids activate the bacterial transcriptional regulator NodD which in turn induces the transcription of the other bacterial nodulation genes involved in the synthesis of Nod factors (Kobayashi et al. 2004). Cereal crops such as rice do not produce isoflavones naturally. To engineer rice plants with the ability to induce *Nod* expression in rhizobia, the soybean isoflavone synthase gene (IFS) was expressed into rice (Sreevidya et al. 2006). The IFS transgenic lines resulted in the production of the isoflavone genistein and also led to the induction of *Nod* gene expression (Sreevidya et al. 2006). Legumes perceive Nod factor signals by Nod factor receptors (NFRs), which are LysM domain receptor kinases [NFR5/nodulation factor perception (NFP) and NFR1/LysM receptor kinase 3 (LYK3)] anchored at the root cell plasma membrane. NFRs initiate the plant processes necessary for the development of root nodule symbiosis (Oldroyd 2013). Another unique receptor protein known as lectin nucleotide phosphohydrolase (LNP) has also been shown to participate in Nod factors perception (Roberts et al. 1999). The heterologous expression of legume-specific NFR genes, *Medicago truncatula NFP*, *M. truncatula LYK3*, and *Lotus japonicus LNP* in rice conferred to root hairs the ability to respond to Nod factors (Altúzar-Molina et al. 2020). Engineering *Rhizobium* infection and nodule organogenesis in cereals still remains elusive.

1.1.2. Photosynthetic efficiency enhancement in C_3 plants through biotechnological interventions

1.1.2.1 Importance of photosynthesis

Photosynthesis uses sunlight energy and atmospheric CO_2 and water to produce chemical energy and release O_2. This process is catalyzed by Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) which is present in most prokaryotic autotrophs (photosynthetic and chemoautotrophic bacteria, cyanobacteria and archaea) and eukaryotes such as algae and higher plants (Ellis 1979). Theoretically, 45% (maximum) of the energy of incident sunlight could be absorbed by photosynthesis, and 67% of the absorbed energy is stored as chemical energy (Melis 2009). Plant growth depends on photosynthesis and the photosynthetic assimilation of CO_2 impacts crop yield directly (Zelitch 1982). Increasing photosynthetic efficiency is arguably critical to improving crop yield, especially in C_3 plants (Long et al. 2015). Additionally, photosynthesis is the essential driver of the global carbon cycle, strongly coupled to the climate system through numerous feedback processes which are associated with extreme climate events such as heat waves, droughts and floods (Ryu et al. 2019).

1.1.2.2 Photosynthesis in C_3, C_4 and CAM plants

Higher plants are categorized as C_3, C_4, or crassulacean acid metabolism (CAM) photosynthetic organisms, depending on the initial product of CO_2 fixation as well as its timing and location (**Figure 1.2**) (Yamori et al. 2014). About 85% of all higher plants are C_3 species, including major crops such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), soybean (*Glycine max*), and potato (*Solanum tuberosum*); 5% are C_4 species, including maize (*Zea mays*) and sugarcane (*Saccharum officinarum*), and the remaining 10% are CAM species such as succulents (Yamori et al. 2014). Both the C_4 and the CAM photosynthesis evolved from the C_3 photosynthetic pathway (Sage 2004; Goolsby et al. 2018). To adapt to the increasing desertification and decreasing atmospheric CO_2 concentrations ca. 30 million years ago, C_4 and CAM plants evolved the biochemical CO_2-concentrating mechanism (CCM) (Edwards 2019). The CCM uses phosphoenolpyruvate carboxylase (PEPC) to fix atmospheric CO_2 into organic acids with subsequent release of high concentrations of CO_2 in the vicinity of Rubisco (**Figure 1.2**) (Yamori et al. 2014).

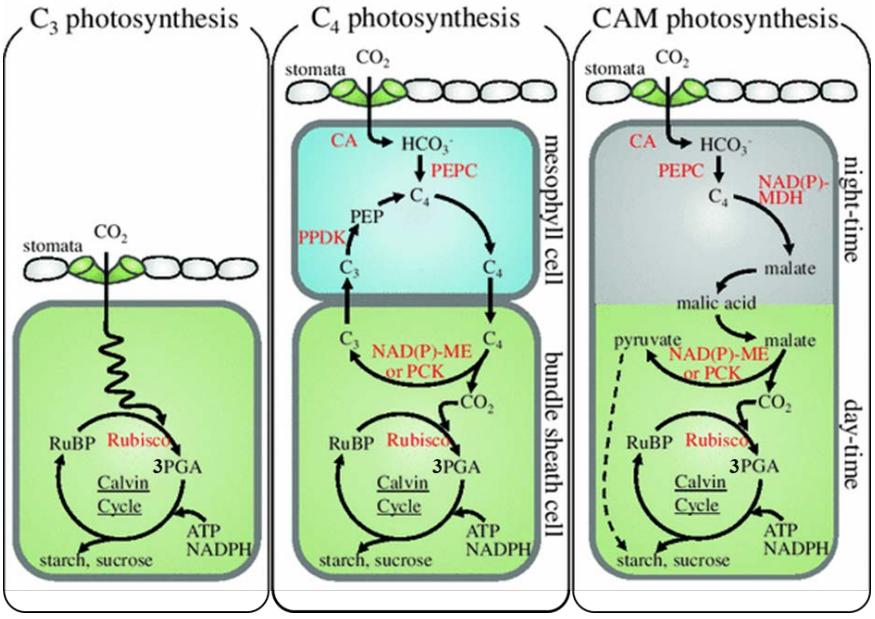

Figure 1.2 Schematic representation of C_3, C_4, and CAM photosynthesis. CA carbonic anhydrase, 3PGA phosphoglyceric acid, RuBP ribulose-1,5-bisphosphate, PEP phosphoenolpyruvate, Rubisco ribulose-1,5-bisphosphate carboxylase/oxygenase, PEPC phosphoenolpyruvate carboxylase, NAD(P)-ME NAD(P)-malic enzyme, PCK phosphoenolpyruvate carboxykinase, PPDK pyruvate phosphate dikinase, NAD(P)-MDH NAD(P)-malate dehydrogenase. The enzymes are highlighted in red. Modified from (Yamori et al. 2014)

In C_3 plants, atmospheric CO_2 diffuses to the stroma of leaf mesophyll cells through stomata and the air spaces between cells and is then fixed by Rubisco (**Figure 1.2**). Rubisco catalyzes the carboxylation of ribulose-1,5-bisphosphate (RuBP) and produces two molecules of 3-phosphoglyceric acid (3PGA) which contains a 3-Carbon unit. 3PGA is further phosphorylated and reduced to synthesize sucrose and starch and regenerates RuBP through the Calvin Cycle (Farquhar et al. 1980).

In C_4 plants, atmospheric CO_2 entering the leaves is hydrated to bicarbonate (HCO_3^-) in the cytosol of mesophyll cells by carbonic anhydrase (CA). HCO_3^- and phosphoenolpyruvate (PEP) are converted to oxaloacetate (OAA), a 4-carbon molecule by PEPC. OAA is further reduced to other C_4 acids (malate, aspartate, or alanine) which diffuse into the bundle sheath cells through plasmodesmata. Malate or aspartate are decarboxylated and release CO_2 to Rubisco and initiate the Calvin Cycle. The C_3 acids released from decarboxylation diffuse back to the mesophyll cells and regenerate to PEP by pyruvate phosphate dikinase (PPDK) (**Figure 1.2**).

In CAM plants, photosynthesis involves a temporal separation of the CO_2 fixation (night-time) and the Calvin cycle (daytime) within one single cell (**Figure 1.2**) (Kellogg 2013; Yamori et al. 2014). CAM takes place in four phases: Phase I is the atmospheric CO_2

assimilation by CA, PEPC and NAD(P)-malate dehydrogenase (NAD(P)-MDH) and accumulation of malic acid in the vacuole of the mesophyll tissue during the night. Phase II is the short period of light-stimulated CO_2 assimilation at dawn, which indicates the activity shift from PEPC to Rubisco. Phase III involves the decarboxylation of stored malic acid and releasing CO_2 to Rubisco for re-fixation while the stomata are tightly closed during the day. Phase IV is similar to Phase II, and is the short period of light-stimulated CO_2 assimilation but at dusk while the stomata open again (Osmond 1978).

1.1.2.3 Genetic engineering strategies to enhance photosynthetic efficiency in C_3 plants

In C_3 plants, the key reason for limiting carbon assimilation is the bifunctional catalytic property of Rubisco. In addition to the carboxylation reaction described earlier, Rubisco also reacts with O_2 leading to oxygenation and photorespiration. The ratio of the carboxylation to the oxygenation reaction depends on the ratio of CO_2 and O_2 concentration around Rubisco. The concentration of O_2 in the atmosphere is much higher than that of CO_2, resulting in a loss of 20% to 50% of the carbon fixed by photosynthesis due to photorespiration (Jordan and Ogren 1981; Mann 1999). In C_4 plants, Rubisco is isolated to bundle sheath cells, where C_4 acids are decarboxylated to release CO_2 in the vicinity of Rubisco. This serves as a light-energy-driven CCM, which largely eliminates photorespiration (Sage 2004). Conversion of a C_3 to C_4 crop would theoretically raise the maximum solar energy conversion efficiency from 4.6% to 6% (Zhu et al. 2008). The light conversion efficiency of C_3 crops is 1-2% versus 3-4% for C_4 crops under normal field conditions (Beadle and Long 1985). If Rubisco can be engineered to completely eliminate the oxygenation reaction, this would raise efficiency to 8.8% (Zhu et al. 2008). In addition, algae and cyanobacteria concentrate CO_2 near Rubisco to minimize oxygenation and photorespiration. respectively (Badger and Price 2003; Moroney James and Ynalvez Ruby 2007).

Therefore, in order to increase photosynthetic efficiency in C_3 plants, overcoming photorespiration is a necessary requirement. This may be achieved by introducing C_4 pathway to C_3 plants or adding algal or cyanobacterial CO_2 concentrating systems or synthesizing photorespiratory bypasses.

In order to introduce the C_4 pathway to C_3 plants, the genes encoding the key enzymes (*Zm*PEPC, *Zm*PPDK, *Sb*NADP-MDH and *Os*NADP-ME) in C_4 pathway have been expressed in rice leaf mesophyll cells (Taniguchi et al. 2008). However, the quadruple

transformants showed only slightly enhanced CO_2 assimilation and were stunted. Engineer a full Kranz two cell-type mechanism in rice is the current approach of engineer C_4 photosynthesis in rice (Ermakova et al. 2020). Due to the specific Kranz structure of C_4 plants, this approach requires not only the proper level of expression of each enzyme (under the control of suitable, tissue-specific promoters in the appropriate cell type) but also an increase in organelle volume in sheath cells surrounding leaf veins (Ermakova et al. 2020). For example, specifically expressing NADP-ME in bundle sheath cells is currently the greatest challenge for establishing a C_4 metabolic pathway in rice. The *Zoysia japonica* phosphoenolpyruvate carboxykinase (*ZjPCK*) promoter directed GUS expression only in bundle sheath cells, but did not retain light-dependent expression in rice (Nomura et al. 2005). A transitional Kranz rice was generated by overexpressing a maize GOLDEN2-LIKE gene (*ZmGLK*), a transcription factor regulating plant chloroplast development. This resulted in significant increases in the volume of bundle sheath cells and the plasmodesmatal connectivity at the bundle sheath/ mesophyll cells interface (Wang et al. 2017). Anatomical features of transgenic rice showed similar characteristics to those found in Kranz species (Wang et al. 2017). However, photosynthetic rate did not increase compared to the wild type when plants were growing in the greenhouse with an average light intensity of 824.75 μmol photons m^{-2} s^{-1} (Wang et al. 2017). Whereas a significantly photosynthetic rate increasing was observed at higher light intensity (over 1000 μmol photons m^{-2} s^{-1}) (Li et al. 2020). The results are probably because GLK upregulates the expression of nuclear photosynthetic genes involved in chlorophyll biosynthesis and light harvesting (Waters et al. 2009).

Cyanobacteria carry out CCM within unique micro-compartments known as carboxysomes, which contain Rubisco together with a carboxysomal CA (Badger and Price 2003). The inorganic carbon (C_i) transporters (bicarbonate transporter1, BCT1; bicarbonate uptake A, BicA; and sodium-bicarbonate transporter A, SbtA) accumulate a cytosolic pool of HCO_3^-. Subsequently, CA converts HCO_3^- into CO_2 coupled with a diffusive restriction system (NAD(P)H dehydrogenase complex NDH-1_3, and NDH-1_4) to increase CO_2 around the active site of Rubisco. Modelling experiments suggest that addition of the complete cyanobacteria CCM to C_3 plants is predicted to increase 60% leaf light-saturated CO_2 uptake and result in 60% increase in yield, whereas addition of a single HCO_3^- transporter is predicted to increase 9% leaf light-saturated CO_2 uptake (McGrath and Long 2014). However, *Synechococcus* BicA proteins were accumulated in tobacco chloroplasts via plastid transformation, but did not exhibit any discernible effect

on photosynthetic CO_2 assimilation rates and growth (Pengelly et al. 2014). Pyrenoids are found in eukaryotic algae and some non-vascular plants. They are surrounded by starch, and preform functions similar to carboxysome (Barrett et al. 2021). A functional pyrenoid CCM also requires C_i uptake systems (low-CO2-inducible protein 1, LCI1; high-light activated 3, HLA3; low-CO2-inducible protein A, LCIA; carbonic anhydrase 3, CAH3 and bestrophin-like proteins 1-3, BST1-3) and the CO_2 anti-leakage proteins (low-CO2-inducible protein B, LCIB and low-CO2-inducible protein C, LCIC) in *Chlamydomonas* (Barrett et al. 2021). Introduction of the passive CO_2 uptake system (LCIB, BST1-3 and CAH3) of pyrenoid CCM into C_3 plants is predicted to increase CO_2 assimilation rates threefold (Fei et al. 2022). Transgenic tobacco plants expressing either CAH3 or LCIA showed significantly increase in shoot biomass. The CO_2 assimilation rate, photosystem II efficiency, and chlorophyll content of the plants were increased by 15%, 18% and 19%, respectively (Nölke et al. 2019).

An alternative way to minimize the extra energy requirements and assimilated carbon loss during the physiological process of photorespiration is to introduce a photorespiratory bypass pathway in C_3 plants through genetic engineering. Two photorespiratory bypasses have been designed and tested in rice, namely a GOC bypass and a GCGT bypass (Shen et al. 2019; Wang et al. 2020). The GOC bypass converted glycolate to CO_2 via *O. sativa* glycolate oxidase 3 (*Os*GLO3), *O. sativa* catalase C (*Os*CATC) and *O. sativa* oxalate oxidase 3 (*Os*OXO3) within the chloroplast (Shen et al. 2019). The GCGT bypass redirected 75% of carbon from glycolate metabolism to the Calvin cycle utilizing *Os*GLO, *E. coli* catalase (*Ec*CAT), *E. coli* glyoxylate carboligase (*Ec*GCL) and *E. coli* tartronic semialdehyde reductase (*Ec*TSR) (Wang et al. 2020).Transgenic rice plants engineered with both bypass pathways, showed increased biomass by 14% - 35% and yield by 7% - 27% in the field (Shen et al. 2019; Wang et al. 2020). The light-saturated photosynthetic rate of GOC rice was increased by 23% (Shen et al. 2019). In addition, GCGT plants showed 19% decrease in the photorespiratory rate (Wang et al. 2020).

1.2 References

Altúzar-Molina A, Lozano L, Ortíz-Berrocal M, Ramírez M, Martínez L, de Lourdes Velázquez-Hernández M, Dhar-Ray S, Silvente S, Mariano N, Shishkova S. 2020. Expression of the legume-specific nod factor receptor proteins alters developmental and immune responses in rice. *Plant Mol. Biol. Report.* **38**: 262-281.

Aryal B, Gurung R, Camargo AF, Fongaro G, Treichel H, Mainali B, Angove MJ, Ngo HH, Guo W, Puadel SR. 2022. Nitrous oxide emission in altered nitrogen cycle and implications for climate change. *Environ. Pollut.* **314**: 120272.

Badger MR, Price GD. 2003. CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**: 609-622.

Barrett J, Girr P, Mackinder LCM. 2021. Pyrenoids: CO2-fixing phase separated liquid organelles. *Biochimica et Biophysica Acta (BBA) - Mol. Cell* **1868**: 118949.

Beadle C, Long S. 1985. Photosynthesis—is it limiting to biomass production? *Biomass* **8**: 119-168.

Bloch SE, Ryu M-H, Ozaydin B, Broglie R. 2020. Harnessing atmospheric nitrogen for cereal crop production. *Curr. Opin. Biotechnol* **62**: 181-188.

Bogorad L. 2000. Engineering chloroplasts: an alternative site for foreign genes, proteins, reactions and products. *Trends Biotechnol.* **18**: 257-263.

Burén S, Jiang X, López-Torrejón G, Echavarri-Erasun C, Rubio LM. 2017. Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NifB. *Front. Plant Sci* **8**: 1567.

Burén S, Jiménez-Vicente E, Echavarri-Erasun C, Rubio LM. 2020. Biosynthesis of nitrogenase cofactors. *Chem. Rev.* **120**: 4921-4968.

Burén S, Pratt K, Jiang X, Guo Y, Jimenez-Vicente E, Echavarri-Erasun C, Dean DR, Saaem I, Gordon DB, Voigt CA et al. 2019. Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **116**: 25078-25086.

Childs N. 2012. Rice Outlook. *Econ. Res. Rep.*

Childs N. 2021. Rice Outlook. *Econ. Res. Rep.*

Childs N, LeBeau B. 2022. *Econ. Res. Rep.*

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genom.* **13**: 162.

Edwards EJ. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytol.* **223**: 1742-1755.

Ellis RJ. 1979. The most abundant protein in the world. *Trends Biochem. Sci* **4**: 241-244.

Ermakova M, Danila FR, Furbank RT, von Caemmerer S. 2020. On the road to C4 rice: advances and perspectives. *Plant J.* **101**: 940-950.

Eseverri Á, López-Torrejón G, Jiang X, Burén S, Rubio LM, Caro E. 2020. Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnol. J.* **18**: 1882-1896.

Farquhar GD, von Caemmerer Sv, Berry JA. 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *planta* **149**: 78-90.

Fei C, Wilson AT, Mangan NM, Wingreen NS, Jonikas MC. 2022. Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat. Plants* **8**: 583-595.

Fernández L. 2022. Global fertilizer demand by nutrient 2011-2022. . Statista.

Goolsby EW, Moore AJ, Hancock LP, De Vos JM, Edwards EJ. 2018. Molecular evolution of key metabolic genes during transitions to C4 and CAM photosynthesis. *Am. J. Bot* **105**: 602-613.

Halbleib CM, Ludden PW. 2000. Regulation of Biological Nitrogen Fixation. *J. Nutr.* **130**: 1081-1084.

Hawkesford M, Horst W, Kichey T, Lambers H, Schjoerring J, Møller IS, White P. 2012. Functions of macronutrients. In *Marschner's mineral nutrition of higher plants*, pp. 135-189. Elsevier.

Huisman R, Geurts R. 2020. A Roadmap toward Engineered Nitrogen-Fixing Nodule Symbiosis. *Plant Commun.* **1**: 100019.

Ivleva NB, Groat J, Staub JM, Stephens M. 2016. Expression of active subunit of nitrogenase via integration into plant organelle genome. *PLoS One* **11**: e0160951.

Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR. 1989. Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. *J. Bacteriol.* **171**: 1017-1027.

Jiang X, Coroian D, Barahona E, Echavarri-Erasun C, Castellanos-Rueda R, Eseverri Á, Aznar-Moreno Jose A, Burén S, Rubio Luis M. 2022. Functional Nitrogenase Cofactor Maturase NifB in Mitochondria and Chloroplasts of Nicotiana benthamiana. *mBio* **13**: e00268-00222.

Jiang X, Payá-Tormo L, Coroian D, García-Rubio I, Castellanos-Rueda R, Eseverri Á, López-Torrejón G, Burén S, Rubio LM. 2021. Exploiting genetic diversity and gene synthesis to

identify superior nitrogenase NifH protein variants to engineer N2-fixation in plants. *Commun. Biol.* **4**: 4.

Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* **291**: 513-515.

Kellogg EA. 2013. C4 photosynthesis. *Curr. Biol.* **23**: R594-R599.

Kobayashi H, Graven YN, Broughton WJ, Perret X. 2004. Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in Rhizobium sp. NGR234. *Mol. Microbiol.* **51**: 335-347.

Kumar A, Yogendra KN, Karre S, Kushalappa AC, Dion Y, Choo TM. 2016. WAX INDUCER1 (HvWIN1) transcription factor regulates free fatty acid biosynthetic genes to reinforce cuticle to resist Fusarium head blight in barley spikelets. *J. Exp. Bot.* **67**: 4127-4139.

Ledbetter RN. 2018. Electron Flow and Management in Living Systems: Advancing Understanding of Electron Transfer to Nitrogenase (Doctoral dissertation, Utah State University).

Li X, Wang P, Li J, Wei S, Yan Y, Yang J, Zhao M, Langdale JA, Zhou W. 2020. Maize GOLDEN2-LIKE genes enhance biomass and grain yields in rice by improving photosynthesis and reducing photoinhibition. *Commun. Biol.* **3**: 151.

Lill R, Mühlenhoff U. 2008. Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu Rev Biochem* **77**: 669-700.

Long Stephen P, Marshall-Colon A, Zhu X-G. 2015. Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell* **161**: 56-66.

Long SR. 1996. Rhizobium symbiosis: nod factors in perspective. *Plant Cell* **8**: 1885.

Mann CC. 1999. Genetic Engineers Aim to Soup Up Crop Photosynthesis. *Science* **283**: 314-316.

McGrath JM, Long SP. 2014. Can the Cyanobacterial Carbon-Concentrating Mechanism Increase Photosynthesis in Crop Species? A Theoretical Analysis *Plant Physiol.* **164**: 2247-2261.

Melis A. 2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Sci.* **177**: 272-280.

Mohidem NA, Hashim N, Shamsudin R, Che Man H. 2022. Rice for Food Security: Revisiting Its Production, Diversity, Rice Milling Process and Nutrient Content. *Agriculture* **12**: 741.

Moroney James V, Ynalvez Ruby A. 2007. Proposed Carbon Dioxide Concentrating Mechanism in Chlamydomonas reinhardtii. *Eukaryotic Cell* **6**: 1251-1259.

Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GE, Poole PS. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**: 3698-3710.

Nölke G, Barsoum M, Houdelet M, Arcalís E, Kreuzaler F, Fischer R, Schillberg S. 2019. The Integration of Algal Carbon Concentration Mechanism Components into Tobacco Chloroplasts Increases Photosynthetic Efficiency and Biomass. *Biotechnol. J.* **14**: 1800170.

Nomura M, Higuchi T, Ishida Y, Ohta S, Komari T, Imaizumi N, Miyao-Tokutomi M, Matsuoka M, Tajima S. 2005. Differential Expression Pattern of C4 Bundle Sheath Expression Genes in Rice, a C3 Plant. *Plant Cell Physiol.* **46**: 754-761.

Oldroyd GE. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol* **11**: 252-263.

Omara P, Aula L, Oyebiyi F, Raun WR. 2019. World Cereal Nitrogen Use Efficiency Trends: Review and Current Knowledge. *Agrosyst. Geosci. Environ.* **2**: 180045.

Osmond C. 1978. Crassulacean acid metabolism: a curiosity in context. *Annu. Rev. Plant Physiol.* **29**: 379-414.

Pengelly JJL, Förster B, von Caemmerer S, Badger MR, Price GD, Whitney SM. 2014. Transplastomic integration of a cyanobacterial bicarbonate transporter into tobacco chloroplasts. *J. Exp. Bot.* **65**: 3071-3080.

Pham DN, Burgess BK. 1993. Nitrogenase reactivity: effects of pH on substrate reduction and carbon monoxide inhibition. *Biochem.* **32**: 13725-13731.

Poole RK, Hill S. 1997. Respiratory protection of nitrogenase activity in Azotobacter vinelandii—roles of the terminal oxidases. *Biosci. Rep.* **17**: 303-317.

Roberts N, Brigham J, Wu B, Murphy J, Volpin H, Phillips D, Etzler M. 1999. A Nod factor-binding lectin is a member of a distinct class of apyrases that may be unique to the legumes. *Mol. Genet. Genom.* **262**: 261-267.

Rubio LM, Ludden PW. 2008. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu Rev Microbiol* **62**: 93-111.

Ryu Y, Berry JA, Baldocchi DD. 2019. What is global photosynthesis? History, uncertainties and opportunities. *Remote Sens. Environ.* **223**: 95-114.

Sage RF. 2004. The evolution of C4 photosynthesis. *New Phytolo.* **161**: 341-370.

Senthilkumar M, Amaresan N, Sankaranarayanan A, Senthilkumar M, Amaresan N, Sankaranarayanan A. 2021. Quantitative Estimation of Nitrogenase Activity: Acetylene Reduction Assay. Springer.

Shen B-R, Wang L-M, Lin X-L, Yao Z, Xu H-W, Zhu C-H, Teng H-Y, Cui L-L, Liu E-E, Zhang J-J. 2019. Engineering a new chloroplastic photorespiratory bypass to increase photosynthetic efficiency and productivity in rice. *Mol. Plant* **12**: 199-214.

Sinclair T. 1990. Nitrogen influence on the physiology of crop yield. *Theoretical production ecology: Reflections and prospects* **181**.

Sinclair T, Horie T. 1989. Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review. *Crop Sci.* **29**: 90-98.

Soto-Gómez D, Pérez-Rodríguez P. 2022. Sustainable agriculture through perennial grains: Wheat, rice, maize, and other species. A review. *Agric Ecosyst Environ* **325**: 107747.

Sreevidya VS, Srinivasa Rao C, Sullia SB, Ladha JK, Reddy PM. 2006. Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. *J. Exp. Bot.* **57**: 1957-1969.

Sullivan BW, Smith WK, Townsend AR, Nasto MK, Reed SC, Chazdon RL, Cleveland CC. 2014. Spatially robust estimates of biological nitrogen (N) fixation imply substantial human alteration of the tropical N cycle. *Proc. Natl. Acad. Sci. U.S.A.* **111**: 8101-8106.

Sun Y, Wang M, Mur LAJ, Shen Q, Guo S. 2020. Unravelling the roles of nitrogen nutrition in plant disease defences. *Int. J. Mol. Sci.* **21**: 572.

Taniguchi Y, Ohkawa H, Masumoto C, Fukuda T, Tamai T, Lee K, Sudoh S, Tsuchida H, Sasaki H, Fukayama H et al. 2008. Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. *J. Exp. Bot.* **59**: 1799-1809.

Vos J, Van Der Putten P, Birch C. 2005. Effect of nitrogen supply on leaf appearance, leaf growth, leaf nitrogen economy and photosynthetic capacity in maize (Zea mays L.). *Field Crops Res.* **93**: 64-73.

Wang L-M, Shen B-R, Li B-D, Zhang C-L, Lin M, Tong P-P, Cui L-L, Zhang Z-S, Peng X-X. 2020. A Synthetic Photorespiratory Shortcut Enhances Photosynthesis to Boost Biomass and Grain Yield in Rice. *Mol. Plant* **13**: 1802-1815.

Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA. 2017. Re-creation of a key step in the evolutionary switch from C3 to C4 leaf anatomy. *Curr. Biol.* **27**: 3278-3287. e3276.

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009. GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in *Arabidopsis Plant Cell* **21**: 1109-1128.

Yamori W, Hikosaka K, Way DA. 2014. Temperature response of photosynthesis in C3, C4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynth. Res.* **119**: 101-117.

Zelitch I. 1982. The close relationship between net photosynthesis and crop yield. *Bioscience* **32**: 796-802.

Zhu X-G, Long SP, Ort DR. 2008. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol* **19**: 153-159.

Zuanazzi JAS, Clergeot PH, Quirion J-C, Husson H-P, Kondorosi A, Ratet P. 1998. Production of Sinorhizobium meliloti nod gene activator and repressor flavonoids from Medicago sativa roots. *Mol. Plant-Microbe Interact* **11**: 784-794.

AIMS AND OBJECTIVES

Aims

The overall aims of my thesis were to address two of the grand challenges in contemporary agriculture, to engineer a heterologous biological nitrogen fixation pathway to incorporate biological nitrogen fixation in rice and engineer an ectopic oxygen scavenging pathway to reduce photorespiration and enhance photosynthetic efficiency in rice through biotechnology.

Objectives

In order to reach these aims, the following specific objectives were pursued:

1) introduce the archaean *Methanocaldococcus infernus NifB* and archaean *Methanothermobacter thermautotrophicus NifB* separately and each case with bacterial *Azotobacter vinelandii FdxN* into rice and measure the enzymatic activity of the corresponding recombinant protein *in planta*;

2) introduce the bacterial *A. vinelandii NifH* with *A. vinelandii NifM, NifS and NifU*, and the bacterial *Hydrogenobacter thermophilus NifH* with *A. vinelandii NifM* to rice and measure the enzymatic activity of the corresponding recombinant protein *in planta*;

3) express bacterial *A. vinelandii NifD* variants to obtain an intact NifD protein in transgenic rice;

4) engineer a novel oxygen scavenging pathway consisting of *Lactobacillus buchneri* LOX, *Escherichia coli* LDH, and *Oryza sativa* CAT into rice and assess its impact on photosynthetic efficiency in rice plants.

Chapter II.
The nitrogenase cofactor maturase NifB isolated from transgenic rice is active in FeMo-co synthesis

2.0 Abstract

The engineering of nitrogen fixation in plants requires the assembly of an active prokaryotic nitrogenase complex, which is yet to be achieved. Nitrogenase biogenesis relies on NifB, which catalyzes the formation of the [8Fe-9S-C] metal cluster NifB-co. This is the first committed step in the biosynthesis of the iron-molybdenum cofactor (FeMo-co) found at the nitrogenase active site. The production of NifB in plants is challenging because this protein is often insoluble in eukaryotic cells, and its [Fe-S] clusters are extremely unstable and sensitive to O_2. As a first step to address this challenge, I generated transgenic rice plants expressing NifB from the archaean *Methanocaldococcus infernus* and *Methanothermobacter thermautotrophicus*. The recombinant proteins were targeted to the mitochondria to limit exposure to O_2 and to have access to essential [4Fe-4S] clusters required for NifB-co biosynthesis. *M. infernus* and *M. thermautotrophicus* NifB accumulated as soluble proteins *in planta* and the purified proteins were functional in the *in vitro* FeMo-co synthesis assay. I characterized NifB protein expression and purification from an engineered staple crop, representing a first step in the biosynthesis of a functional NifDK complex as required for independent biological nitrogen fixation in cereals

2.1. Introduction

2.1.1. Nitrogenase and its components

Biological nitrogen fixation (BNF) involves the conversion of inert atmospheric N_2 into biologically usable ammonia catalyzed by the enzyme nitrogenase (Halbleib and Ludden 2000). Nitrogenases are only found in some bacteria and archaea, so plants cannot fix their own nitrogen and must either form symbiotic relationships with nitrogen-fixing prokaryotes or obtain fixed nitrogen compounds from the soil (Dos Santos et al. 2012).

There are three types of nitrogenases: molybdenum (Mo), vanadium (V) and iron-only (Fe) nitrogenases (Harwood 2020). The Mo-nitrogenase is the most widespread and studied. It is a complex metalloenzyme composed of two metalloproteins: molybdenum-iron (MoFe) protein and Fe protein. The MoFe protein is a dinitrogenase encoded by *NifDK* which contains the catalytic site that binds and reduces N_2. The MoFe protein is an $\alpha_2\beta_2$ tetramer with a P-cluster [8Fe-7S] located at the α/β-subunit interface and an FeMo-co [7Fe-9S-C-Mo-*R* homocitrate] located within the α-subunit (Burén et al. 2020).

The Fe protein is a dinitrogenase reductase encoded by *NifH*, which provides electrons to the MoFe protein (Seefeldt et al. 2020). (The detail is described in chapter III)

Nitrogenase catalysis starts with ATP hydrolysis, and then the Fe protein donates electrons from its [4Fe-4S] cluster to the active site FeMo-co of MoFe protein for the final reduction of the N_2 (Burgess and Lowe 1996). In this process, the P-cluster acts as electron relay sites between Fe and MoFe proteins, coupling electron, and proton transfer (Peters et al. 1997). However, the later study showed that P-cluster was displaceable by other metal clusters in nitrogenase activity (Hu et al. 2005). Two P-cluster variants with similar [4Fe-4S]-like centers also supported substrate reduction at the catalytic site (Hu et al. 2005). In contrast, the FeMo-co cluster as an active site is irreplaceable for nitrogenase activity. The FeMo-co deficient FeMo protein showed no enzymatic activity, but it could be activated *in vitro* by adding FeMo-co (Hawkes and Smith 1983; Zhao et al. 2005).

2.1.2. The importance of NifB in MoFe protein maturation

The [4Fe-4S] clusters are synthesized by Fe-S cluster assembly protein NifU (NifU) and cysteine desulfurase (NifS), where NifU provides the scaffold for the assembly of [Fe-S] clusters and NifS mobilizes sulfur from cysteine for [Fe-S] cluster synthesis on NifU (Jacobson et al. 1989; Smith et al. 2005). The P-cluster is formed by the reductive coupling of two [4Fe-4S] cluster pairs at the MoFe protein in the presence of Fe protein, MgATP and a reductant (Hu et al. 2007; Jimenez-Vicente et al. 2019). FeMo-co is one of the most complex [Fe-S] clusters discovered in nature thus far, and is synthesized in a regulated and coordinated process depending on a multitude of proteins (Rubio and Ludden 2008).

The proteins involved in FeMo-co biosynthesis can be functionally divided into four classes (a) molecular scaffolds: NifU, NifB and the complex of nitrogenase iron-molybdenum cofactor biosynthesis proteins NifE and NifN (NifEN), (b) metallocluster carriers: nitrogen fixation protein NifX (NifX), nitrogenase gamma subunit (NafY) and nitrogen fixation protein NifY (NifY), (c) substrate providers: NifS, nitrogen fixation protein NifQ (NifQ) and homocitrate synthase (NifV) and (d) Fe protein (NifH) (Rubio and Ludden 2008). FeMo-co biosynthesis is initiated by NifU and NifS with the assembly of [4Fe-4S] cluster units that are transferred to NifB (Shah et al. 1994; Curatti et al. 2006; Zhao et al. 2007). NifB is an S-adenosyl-L-methionine (SAM) radical enzyme that carries a catalytic [4Fe-4S] cluster (called RS cluster) and two additional [4Fe-4S] clusters (called K1- and K2-clusters) used as substrates to generate the [8Fe-9S-C] product called

NifB-co (Wiig et al. 2012; Guo et al. 2016; Fajardo et al. 2020; Jenner et al. 2021). It has been shown that NifB-co production is 80% lower in *A. vinelandii FdxN* mutants, suggesting that ferredoxin (FdxN) is required for NifB-co biosynthesis (Jimenez-Vicente et al. 2014). Although the precise function of FdxN remains unclear, it is thought to provide NifB with electrons needed for NifB-co formation (Burén et al. 2020). NifB-co synthesized by NifB is matured into FeMo-co on NifEN/NifH complexes (Allen et al. 1995; Soboh et al. 2006), which is then inserted into the active site of the apo-MoFe protein to reconstitute active Mo-nitrogenase (Curatti et al. 2007). As the function of NifB is to catalyze the first committed step in the biosynthesis of FeMo-co (the production of NifB-co), *NifB* mutants lack FeMo-co (Shah et al. 1986; Christiansen et al. 1998). NifB-co is also the precursor cluster for the biosynthesis of FeV cofactor (FeV-co) and FeFe cofactor (FeFe-co) found at the active sites of the alternative V-nitrogenase and Fe-only nitrogenase (Joerger and Bishop 1988; Shah et al. 1994), and therefore required for all biological nitrogen fixation.

2.1.3. NifB characterization and its heterologous expression

The NifB proteins of *A. vinelandii* and *Klebsiella oxytoca* comprise an N-terminal SAM-radical domain containing the CxxxCxxC SAM-binding motif and a C-terminal NifX-like domain (Boyd ES 2011; Arragain et al. 2017). However, the simplest NifB architecture is a standalone SAM-radical domain because the NifX-like sequence is not essential for NifB activity. For example, the single-domain NifB proteins from *Methanosarcina acetivorans*, *M. thermoautotrophicum*, *M. infernus*, and *Methanothrix thermoacetophila* are functionally equivalent to *A. vinelandii* NifB (Fay et al. 2015; Arragain et al. 2017; Fajardo et al. 2020). This single-domain architecture facilitates the heterologous expression of soluble NifB in *Escherichia coli* (Fay et al. 2015; Wilcoxen et al. 2016). When a library of 28 NifB proteins was screened for expression in *Saccharomyces cerevisiae* mitochondria only six accumulated as predominantly soluble proteins, and four of these had single-domain architecture (Burén et al. 2019). When the same NifB variants were expressed transiently in *Nicotiana benthamiana* and targeted to the mitochondria or chloroplasts, only three accumulated as soluble proteins. All three were single-domain proteins, corroborating their superior expression and performance in eukaryotic hosts (Jiang et al. 2022).

Another obstacle to the heterologous expression of NifB is its vulnerability to oxygen. Thirty seconds of exposure to air, resulted in a loss of 90% of NifB activity (Curatti et al.

2006). However, mitochondria and chloroplast have been shown to be suitable sub-cellular organelles for the expression of NifB in *S. cerevisiae* and *N. benthamiana*. Active *M. infernus* NifB and *M. thermautotrophicus* NifB were produced in aerobically cultured *S. cerevisiae* by targeting NifB to mitochondria (Burén et al. 2017a; Burén et al. 2019). Furthermore, transiently expressed NifB isolated from *N. benthamiana* mitochondria or chloroplasts supported the FeMo-co synthesis *in vitro* (Jiang et al. 2022).

In this chapter, I describe the generation and characterization of engineered rice plants expressing *M. infernus* or *M. thermautotrophicus* NifB targeted to the mitochondria, in each case together with *A. vinelandii* FdxN. Both NifB proteins accumulated in a soluble form and their functionality was confirmed by FeMo-co synthesis *in vitro*. *M. thermautotrophicus* NifB exhibited higher NifB-co conversion activity *in vitro* compared with *M. infernus* NifB. The production of functional NifB in rice represents an important step toward the expression of active nitrogenase to achieve BNF in cereal crops.

2.2. Aims and objectives

The aims of this study were to express nitrogenase active cofactor maturase NifB in transgenic rice plants and to determine its activity. The specific objectives were to: 1) introduce the *M. infernus NifB* or *M. thermautotrophicus NifB* into rice, in each case co-transformed with *A. vinelandii FdxN*; 2) measure mRNA expression levels of the introduced transgenes in transgenic rice callus and plants; 3) confirm accumulation of *M. infernus* NifB, *M. thermautotrophicus* NifB and *A. vinelandii* FdxN protein in transgenic rice callus and plants; 4) purify *M. infernus* NifB and *M. thermautotrophicus* NifB from transgenic rice callus and plants; 5) determine the enzymatic activity of the purified *M. infernus* NifB and *M. thermautotrophicus* NifB in the FeMo-co synthesis *in vitro* assay.

2.3. Materials and methods

2.3.1. Expression constructs for rice transformation

The sequences of *M. infernus NifB* (*NifB*^{Mi}), *M. thermautotrophicus NifB* (*NifB*^{Mt}), *A. vinelandii FdxN* (*FdxN*^{Av}), mitochondrial targeting peptide *S. cerevisiae* cytochrome c oxidase subunit IV (Cox4), the Twinstrep (TS) and hemagglutinin (HA) tags were codon-optimized for *S. cerevisiae* using the GeneOptimizer tool (Thermo Fisher Scientific, Waltham, MA, USA) and synthesized by Thermo Fisher Scientific as described (Burén et al. 2017a; Burén et al. 2019). The empty vector pUC57 (GenScript Biotech, Piscataway, NJ, USA) was digested with Acc65I and SalI, allowing the insertion of the strong

constitutive *ZmUbi1*+1sti promoter. The Cox4-TS-*NifB*^{Mi}- t*Nos* cassette was generated by PCR using pN2XJ21 (Burén et al. 2017a) as the template, and introduced into the intermediate vector pUC57-*ZmUbi1*+1sti at *the* SalI *and* SphI sites to produce pMiNifB. The pMiNifB vector was then digested with BamHI and BstEII, allowing the insertion of the *NifB*^{Mt} sequence generated by PCR using pN2SB103 (Burén et al. 2019) as template, to produce pMtNifB. This was digested with Sal1 and BstEII, to allow insertion of the synthetic Cox4-*FdxN*^{Av}*-HA* cassette, to generate pAvFdxN. All restriction enzymes and T4 DNA ligase were obtained from Promega (Madison, WI, USA) or New England Biolabs (Ipswich, MA, USA). The plasmids were amplified in *E. coli* DH5α cells grown at 37°C in lysogeny broth medium supplemented with 100 μg/mL ampicillin. The fidelity of all DNA constructs was verified by Sanger sequencing (Stabvida, Caparica, Portugal). The sequences of cloning primers and DNA constructs for expression in rice are listed in **Tables 2.1 and 2.2**.

Table 2.1 Vectors and constructs used. (a) Plant expression vectors, (b) sequences of genetic constructs.

(a)

Plasmids	Expressed proteins	Promoter / terminator	Size (kDa, with tag)
PUC57	Cox4-TS-NifB^{Mi}	p*ZmUbi1*+1sti / t*Nos*	38
PUC57	Cox4-TS-NifB^{Mt}	p*ZmUbi1*+1sti / t*Nos*	35
PUC57	Cox4-FdxN^{Av}-HA	p*ZmUbi1*+1sti / t*Nos*	11

Cox4, *S. cerevisiae* cytochrome c oxidase subunit IV; TS, Twinstrep tag; HA, hemagglutinin tag; NifB^{Mi}, *M. infernus* NifB; NifB^{Mt}, *M. thermautotrophicus* NifB; FdxN^{Av}, *A. vinelandii* FdxN.

(b)

Genetic elements	Nucleotide sequence Amino acid sequence										
Cox4											
	ATG	CTT	TCA	CTT	AGA	CAA	TCT	ATT	AGA	TTT	TTC
	M	**L**	**S**	**L**	**R**	**Q**	**S**	**I**	**R**	**F**	**F**
	AAG	CCA	GCT	ACA	AGA	ACT	TTG	TGT	TCT	TCT	AGA
	K	**P**	**A**	**T**	**R**	**T**	**L**	**C**	**S**	**S**	**R**
	TAT	CTT	CTT	CAG	CAA	AAA	CCT				
	Y	**L**	**L**	**Q**	**Q**	**K**	**P**				
TS											
	TGG	AGT	CAT	CCT	CAG	TTT	GAG	AAA	GGT	GGA	GGT
	W	**S**	**H**	**P**	**Q**	**F**	**E**	**K**	**G**	**G**	**G**
	TCA	GGT	GGT	GGA	AGC	GGT	GGA	TCT	GCT	TGG	TCA
	S	**G**	**G**	**G**	**S**	**G**	**G**	**S**	**A**	**W**	**S**
	CAT	CCA	CAA	TTT	GAA	AAA					

	H	**P**	**Q**	**F**	**E**	**K**					
HA											
	TAT	CCA	TAT	GAT	GTT	CCA	GAT	TAT	GCT	TAA	
	Y	**P**	**Y**	**D**	**V**	**P**	**D**	**Y**	**A**	**-**	
NifB^{Mi}											
	ATG	GAA	AAG	ATG	TCC	AAG	TTC	TCC	CAT	TTG	TTG
	M	**E**	**K**	**M**	**S**	**K**	**F**	**S**	**H**	**L**	**L**
	AAA	GCT	CAT	CCA	TGC	TTC	AAC	GAA	AAG	GTT	CAT
	K	**A**	**H**	**P**	**C**	**F**	**N**	**E**	**K**	**V**	**H**
	GAT	AAG	TAC	GGT	AGA	GTT	CAT	TTG	CCA	GTT	GCT
	D	**K**	**Y**	**G**	**R**	**V**	**H**	**L**	**P**	**V**	**A**
	CCA	AGA	TGT	AAC	ATT	GCT	TGT	AAG	TTC	TGC	AAG
	P	**R**	**C**	**N**	**I**	**A**	**C**	**K**	**F**	**C**	**K**
	AGG	TCC	GTT	TCT	AAA	GAA	TGT	TGT	GAA	CAT	AGA
	R	**S**	**V**	**S**	**K**	**E**	**C**	**C**	**E**	**H**	**R**
	CCA	GGT	GTT	TCT	TTG	GGT	GTT	TTG	AAA	CCA	GAA
	P	**G**	**V**	**S**	**L**	**G**	**V**	**L**	**K**	**P**	**E**
	GAT	GTT	GAG	GAC	TAC	CTG	AAA	AAG	ATC	TTG	AAA
	D	**V**	**E**	**D**	**Y**	**L**	**K**	**K**	**I**	**L**	**K**
	GAG	ATG	CCA	AAC	ATC	AAG	GTT	GTT	GGT	ATT	GCT
	E	**M**	**P**	**N**	**I**	**K**	**V**	**V**	**G**	**I**	**A**
	GGT	CCT	GGT	GAT	TCT	CTG	TTT	AAC	AAA	GAA	ACT
	G	**P**	**G**	**D**	**S**	**L**	**F**	**N**	**K**	**E**	**T**
	TTC	GAA	ACC	CTG	AAG	ATC	ATC	GAC	GAA	AAG	TTT
	F	**E**	**T**	**L**	**K**	**I**	**I**	**D**	**E**	**K**	**F**
	CCC	AAC	TTG	ATT	AAG	TGC	ATT	TCC	ACC	AAC	GGT
	P	**N**	**L**	**I**	**K**	**C**	**I**	**S**	**T**	**N**	**G**
	CTG	TTG	TTG	TCT	AAG	TAC	TAC	AAG	GAT	TTG	GCC
	L	**L**	**L**	**S**	**K**	**Y**	**Y**	**K**	**D**	**L**	**A**
	AAC	TTG	AAC	GTT	AGA	ACT	ATT	ACC	GTT	ACT	GTC
	N	**L**	**N**	**V**	**R**	**T**	**I**	**T**	**V**	**T**	**V**
	AAC	GCC	ATT	AAG	CCA	GAA	ATC	TTG	GAA	AAA	ATC
	N	**A**	**I**	**K**	**P**	**E**	**I**	**L**	**E**	**K**	**I**
	GTT	GAC	TGG	GTT	TAC	TAC	GAC	AAG	AAG	TTG	TAT
	V	**D**	**W**	**V**	**Y**	**Y**	**D**	**K**	**K**	**L**	**Y**
	AGA	GGT	TTG	GAA	GGT	GCC	AAG	TTG	TTG	ATC	GAA
	R	**G**	**L**	**E**	**G**	**A**	**K**	**L**	**L**	**I**	**E**
	AAA	CAA	ATC	GAA	GGT	ATC	AAG	AAG	GCC	TCC	GAA
	K	**Q**	**I**	**E**	**G**	**I**	**K**	**K**	**A**	**S**	**E**
	GAA	GAT	TTC	ATT	ATC	AAG	ATC	AAC	ACC	GTC	TTG
	E	**D**	**F**	**I**	**I**	**K**	**I**	**N**	**T**	**V**	**L**
	ATC	CCA	GAA	ATC	AAC	ATG	GAT	CAC	GTT	GTT	GAA
	I	**P**	**E**	**I**	**N**	**M**	**D**	**H**	**V**	**V**	**E**
	ATT	GCC	AAG	TTC	TTC	AAG	GAT	TAC	GCC	TAC	GTT
	I	**A**	**K**	**F**	**F**	**K**	**D**	**Y**	**A**	**Y**	**V**
	CAA	AAC	ATC	ATT	CCA	TTG	ATT	CCA	CAG	TAC	AAG
	Q	**N**	**I**	**I**	**P**	**L**	**I**	**P**	**Q**	**Y**	**K**
	ATG	AAG	GAA	TTG	AGA	GCA	CCA	ACT	TGC	GAA	GAA
	M	**K**	**E**	**L**	**R**	**A**	**P**	**T**	**C**	**E**	**E**

	ATC	AAA	AAG	GTC	AGA	AAA	GAG	TGC	GAG	AAG	TAC
	I	**K**	**K**	**V**	**R**	**K**	**E**	**C**	**E**	**K**	**Y**
	ATC	CCA	CAA	TTC	AGA	GCT	TGT	GGT	CAA	TGT	AGA
	I	**P**	**Q**	**F**	**R**	**A**	**C**	**G**	**Q**	**C**	**R**
	GCT	GAT	GCT	GTT	GGT	CTG	ATC	AAA	GAA	AAA	GAG
	A	**D**	**A**	**V**	**G**	**L**	**I**	**K**	**E**	**K**	**E**
	CTG	TTG	AAA	GAG	TTT	TTC	AAA	GAG	AAG	AAC	AAA
	L	**L**	**K**	**E**	**F**	**F**	**K**	**E**	**K**	**N**	**K**
	GAA	AAG	AAC	ATC	AAG	CTG	GAA	GTG	TTC	GAC	TTG
	E	**K**	**N**	**I**	**K**	**L**	**E**	**V**	**F**	**D**	**L**
	AAG	CAC	TTC	TCT	CAT	TGA					
	K	**H**	**F**	**S**	**H**	**-**					
NifB^{Mt}											
	ATG	CCA	GAT	CAA	AGA	CAA	ACC	AGA	TTC	GCT	CAT
	M	**P**	**D**	**Q**	**R**	**Q**	**T**	**R**	**F**	**A**	**H**
	ATT	ACT	AAG	GCT	CAT	CCA	TGC	TTC	AAC	GAA	AAG
	I	**T**	**K**	**A**	**H**	**P**	**C**	**F**	**N**	**E**	**K**
	TTG	CAT	GAT	AGA	GTT	GGT	AGA	GTT	CAT	GTT	CCA
	L	**H**	**D**	**R**	**V**	**G**	**R**	**V**	**H**	**V**	**P**
	ATT	GCT	CCA	AGA	TGT	AAC	ATC	CAT	TGC	AAG	TTC
	I	**A**	**P**	**R**	**C**	**N**	**I**	**H**	**C**	**K**	**F**
	TGT	ACC	AGA	GAT	ATC	AAC	GAA	TGT	GAA	AGA	CGT
	C	**T**	**R**	**D**	**I**	**N**	**E**	**C**	**E**	**R**	**R**
	CCA	GGT	GTT	ACT	GGT	AGA	TTG	ATG	ACT	GCT	GAT
	P	**G**	**V**	**T**	**G**	**R**	**L**	**M**	**T**	**A**	**D**
	GAT	GCT	ATT	AAG	CAC	GTC	GAA	AAG	GTC	AAA	GAA
	D	**A**	**I**	**K**	**H**	**V**	**E**	**K**	**V**	**K**	**E**
	GAA	ATG	CCA	ATT	TCC	GTT	ATT	GGT	GTT	GCT	GGT
	E	**M**	**P**	**I**	**S**	**V**	**I**	**G**	**V**	**A**	**G**
	CCA	GGT	GAT	GCT	TTG	GCT	AAT	GAA	GAA	ACT	TTC
	P	**G**	**D**	**A**	**L**	**A**	**N**	**E**	**E**	**T**	**F**
	GAG	TTC	TTC	AAG	AAG	GCC	TCT	AAG	AAG	TTT	CCA
	E	**F**	**F**	**K**	**K**	**A**	**S**	**K**	**K**	**F**	**P**
	GAT	TTG	TTG	AAG	TGT	ATG	TCC	ACC	AAC	GGT	TTG
	D	**L**	**L**	**K**	**C**	**M**	**S**	**T**	**N**	**G**	**L**
	TTG	TTG	CCA	GAT	AGA	GCT	GAT	GAA	TTG	GCT	GAA
	L	**L**	**P**	**D**	**R**	**A**	**D**	**E**	**L**	**A**	**E**
	TTG	GGT	ATT	AAC	ACT	GTT	ACT	GTT	ACC	GTT	AAC
	L	**G**	**I**	**N**	**T**	**V**	**T**	**V**	**T**	**V**	**N**
	GCT	GTT	GAT	CCA	GAA	ATT	GGT	GAA	AAG	ATC	TAC
	A	**V**	**D**	**P**	**E**	**I**	**G**	**E**	**K**	**I**	**Y**
	TCC	TTC	GTT	GTC	TAC	AAG	GAT	AAG	GTT	TAT	CAT
	S	**F**	**V**	**V**	**Y**	**K**	**D**	**K**	**V**	**Y**	**H**
	GGT	AGA	GAA	GCC	TTC	GAA	GTG	TTG	TCT	AGA	AAT
	G	**R**	**E**	**A**	**F**	**E**	**V**	**L**	**S**	**R**	**N**
	CAA	TTG	GAA	GGC	ATT	GAA	AAG	TTG	GCC	GAA	AGA
	Q	**L**	**E**	**G**	**I**	**E**	**K**	**L**	**A**	**E**	**R**
	GGT	ATT	ATC	GTC	AAG	GTT	AAC	TCT	GTT	TTG	ATC
	G	**I**	**I**	**V**	**K**	**V**	**N**	**S**	**V**	**L**	**I**
	CCA	GGT	TTG	AAC	GAT	GAA	CAT	ATT	GTC	GAT	ATT

	P	**G**	**L**	**N**	**D**	**E**	**H**	**I**	**V**	**D**	**I**
	GCC	CGT	GAA	GTT	AAG	AAA	AGG	GGT	GCT	TCT	TTG
	A	**R**	**E**	**V**	**K**	**K**	**R**	**G**	**A**	**S**	**L**
	ATG	AAC	ATC	ATT	CCA	TTG	ATT	CCA	ATG	GGT	GAG
	M	**N**	**I**	**I**	**P**	**L**	**I**	**P**	**M**	**G**	**E**
	ATG	AAG	GAT	TAT	CCA	AGA	CCA	ACC	TGT	GAA	CAA
	M	**K**	**D**	**Y**	**P**	**R**	**P**	**T**	**C**	**E**	**Q**
	ATC	GAA	AGA	GTT	AGA	AAC	GAA	GTC	GAG	AAG	ATC
	I	**E**	**R**	**V**	**R**	**N**	**E**	**V**	**E**	**K**	**I**
	ATC	CCA	GTT	TTT	AGA	GCT	TGT	ACT	CAA	TGT	AGA
	I	**P**	**V**	**F**	**R**	**A**	**C**	**T**	**Q**	**C**	**R**
	GCA	GAT	GCT	TAT	GGT	ATC	CCA	GGT	AAA	AAA	GAA
	A	**D**	**A**	**Y**	**G**	**I**	**P**	**G**	**K**	**K**	**E**
	GCT	GAT	AAG	CAC	TTG	GAT	ATG	ACC	CCA	GCT	TCT
	A	**D**	**K**	**H**	**L**	**D**	**M**	**T**	**P**	**A**	**S**
	CAT	TAC	TAA								
	H	**Y**	**-**								
FdxN^{Av}											
	ATG	GCT	CTT	AAG	ATA	GTT	GAG	TCT	TGT	GTG	AAC
	M	**A**	**L**	**K**	**I**	**V**	**E**	**S**	**C**	**V**	**N**
	TGC	TGG	GCA	TGT	GTT	GAT	GTG	TGC	CCA	AGT	GAG
	C	**W**	**A**	**C**	**V**	**D**	**V**	**C**	**P**	**S**	**E**
	GCT	ATA	TCC	TTG	GCA	GGT	CCT	CAT	TTT	GAA	ATT
	A	**I**	**S**	**L**	**A**	**G**	**P**	**H**	**F**	**E**	**I**
	TCT	GCT	TCA	AAA	TGC	ACC	GAG	TGT	GAT	GGA	GAC
	S	**A**	**S**	**K**	**C**	**T**	**E**	**C**	**D**	**G**	**D**
	TAT	GCT	GAA	AAG	CAA	TGC	GCA	TCT	ATT	TGT	CCA
	Y	**A**	**E**	**K**	**Q**	**C**	**A**	**S**	**I**	**C**	**P**
	GTT	GAA	GGT	GCT	ATC	TTG	TTA	GCA	GAC	GGA	ACT
	V	**E**	**G**	**A**	**I**	**L**	**L**	**A**	**D**	**G**	**T**
	CCT	GCT	AAC	CCA	CCT	GGT	TCA	CTT	ACA	GGA	ATC
	P	**A**	**N**	**P**	**P**	**G**	**S**	**L**	**T**	**G**	**I**
	CCA	CCT	GAA	AGA	TTG	GCT	GAG	GCA	ATG	AGA	GAA
	P	**P**	**E**	**R**	**L**	**A**	**E**	**A**	**M**	**R**	**E**
	ATA	CAG	GCA	AGG							
	I	**Q**	**A**	**R**							

Table 2.2 Primers used for vector construction.

Genetic element	Primer sequence (5′→3′)	Restriction enzymes
p*ZmUbi1*+1sti	F: TAAGCAGGATCCGGAGTGCAGTGCAGCGTGA	*Acc*65I
	R: TGCTTACTGCAGAAGTAACACCAAACAACAG	*Sal*I
Cox4-TS-*NifB*^{Mi}	F: TAAGCACTCGACATGCTTTCACTTAGACAATC	*Sal*I
-t*Nos*	R:TGCTTAGCATGAGATCTAGTAACATAGATGAC	*Sph*I
NifB^{Mt}	F: TAAGCAGGATCCATGCCAGATCAAAGACAAAC	*BamH*I
	R:TGCTTAGGTNACCTTAGTAATGAGAAGCTGGGG	*BstE*II
Cox4-*FdxN*^{Av}	F: TAAGCACTCGACATGCTTTCACTTAGACAATC	*Sal*I
-HA	R: TGCTTAGGTNACCTTAAGCATAATCTGGAACAT	*BstE*II

p*ZmUbi1*+1st I, maize ubiquitin promoter and first intron; Cox4, *S. cerevisiae* cytochrome c oxidase subunit IV; TS, Twinstrep tag; HA, hemagglutinin tag; *NifB*^{Mi}, *M. infernus NifB*; *NifB*^{Mt}, *M. thermautotrophicus NifB*; *FdxN*^{Av}, *A. vinelandii FdxN*; t*Nos*, nopaline synthase terminator.

2.3.2. Transformation of rice explants, callus recovery and regeneration of transgenic plants

Seven-day-old mature rice embryos (*Oryza sativa* cv. Nipponbare) were isolated for particle bombardment. The embryos were transferred to Murashige & Skoog (MS) osmoticum (MSO) medium for 4 h in the dark before transformation with 10 mg gold particles coated with the transgene constructs (pMiNifB and pAvFdxN, or pMtNifB and pAvFdxN) and the hygromycin phosphotransferase (*Hpt*) selectable marker at a molar ratio of 3:3:1. The bombarded embryos were maintained on MSO medium for 16 h in the dark, then transferred to MS selection medium for 4 weeks in the dark, with one subculture after 2 weeks. Half of the resistant callus were kept under selection and the other half were transferred to MS regeneration medium with a 12 h photoperiod for 3-4 weeks to regenerate transgenic plantlets. The transgenic plantlets were transferred to rooting medium (HMS) with a 12 h photoperiod for 2 weeks and planted to soil in the greenhouse with a 12 h photoperiod and 80% relative humidity. Media composition is shown in **Table 2.3**.

Table 2.3 Composition of media for *in vitro* culture (for 1 L total volume).

Item	MSP	MSO	MSS	MSR	HMS
MS powder	4.4 g	4.4 g	4.4 g	4.4 g	2.2 g
Casein hydrolysate	300 mg	300 mg	300 mg	100 mg	
Proline	500 mg	500 mg	500 mg		
Sucrose	30 g	30 g	30 g		10 g
Maltose				30 g	
Mannitol		72.8 g			
2.4-D (5 mg/mL)	500 μL	500 μL	500 μL		
			Adjust pH to 5.8		
Phytagel	5 g	3 g	5 g	4 g	3 g + 2 g agar
			Autoclave at 121°C for 20 min		
BAP				3 mg	
NAA				0.5 mg	
Hygromycin			1 mg	1 mg	1 mg

MS, Murashige & Skoog medium with vitamins; MSP, medium for callus induction; MSO, osmoticum medium prior to bombardment; MSS, selection and callus proliferation medium; MSR, plant regeneration medium; HMS, rooting medium; 2,4-D, 2,4-dichlorophenoxyacetic acid; BAP, 6-benzylaminopurine (systematic name *N*-(phenylmethyl)-7*H*-purin-6-amine), NAA, 1-naphthaleneacetic acid.

2.3.3. Gene expression analysis by real-time quantitative reverse transcription PCR

Total RNA was isolated from rice callus and leaves of the corresponding regenerated plants using the LiCl method (Creissen and Mullineaux 1995). The cDNA was synthesized from 1 µg of total RNA using Maxima™ H minus cDNA Synthesis Master Mix with dsDNase kit (Thermo Fisher Scientific). The cDNA products were diluted 5-fold prior to use in real-time quantitative reverse transcription PCR (RT-qPCR). RT-qPCR was performed in 96-well plates in the CFX96 system (Bio-Rad Laboratories, Hercules, CA, USA) using a 15 µl mixture containing 1.0 µl 5-fold diluted cDNA template, 2× iTaq Universal SYBR Green Supermix (Bio-Rad), and 0.5 µM of each primer. The RT-qPCR cycling conditions and melt curve protocol were as previously described in Jin et al. (2019). All reactions were carried out in triplicate for each cDNA sample. Primer specificity was confirmed by melting curve analysis of the final PCR products and the identity of the PCR products was confirmed by sequencing. Primer Expression levels were normalized against *OsActin* mRNA. The RT-qPCR primers are listed in **Table 2.4.**

Table 2.4 Primers used for real-time quantitative reverse transcription PCR analysis.

Gene	Primer sequence (5′→3′)	Product size (bp)
$Actin^{Os}$	F: TCATGTCCCTCACAATTTCC	181
	R: GACTCTGGTGATGGTGTCAGC	
$NifB^{Mi}$	F: CTGTCAACGCCATTAAGCCA	178
	R: TCGGAGGCCTTCTTGATACCT	
$NifB^{Mt}$	F: ACGGTTTGTTGTTGCCAGAT	136
	R: GCCAACTTTTCAATGCCTTCC	
$FdxN^{Av}$	F: GCTGGGCATGTGTTGATGTG	199
	R: GCAGGAGTTCCGTCTGCTAA	

$Actin^{Os}$, O. sativa Actin; $NifB^{Mi}$, M. infernus NifB; *$NifB^{Mt}$, M. thermautotrophicus NifB*; *$FdxN^{Av}$, A. vinelandii FdxN*.

2.3.4. Protein extraction and immunoblot analysis

Total rice protein extracts for initial expression analysis from callus were prepared by grinding 0.1-0.2 g callus in liquid nitrogen and thawing the powder in 0.2–0.4 mL extraction buffer comprising 20 mM Tris-HCl (pH 7.5), 5 mM ethylenediaminetetraacetic acid (EDTA), 0.1% Tween-20, 0.1% sodium dodecylsulfate (SDS), and 2 mM phenylmethanesulfonylfluoride (PMSF). The mixture was vortexed for 1 h at 4 °C. Cell debris was removed by centrifugation at 13,000 rpm for 20 min at 4 °C, and the supernatant was collected and stored at -80 °C. The protein concentration in the

supernatants was determined using the Bradford method (AppliChem, Darmstadt, Germany).

Soluble rice leaf protein extracts were prepared by grinding ca. 50 mg rice tissue (snap-frozen in liquid N_2) in 2-mL Eppendorf tubes using 3-mm BeadBug steel balls and a microtube homogenizer (Benchmark Scientific, Edison, NJ, USA) operating at 400 rpm for 20 s. Leaf powder was resuspended in 7 volumes (v/w) of extraction buffer comprising 100 mM Tris-HCl (pH 8.6), 200 mM NaCl, 10% glycerol, 1 mM PMSF, 1 µg/mL leupeptin and 5 mM EDTA, and homogenized twice. Cell debris was removed by centrifugation (20,000*g*, 5 min, 4 °C) and the supernatant was collected and stored at -80 °C. Soluble rice callus extracts were prepared using a blender (see below 2.3.5 section).

Rice proteins were separated by SDS-PAGE and then immunoblotted to Protran Premium 0.45 µm nitrocellulose membranes (GE Healthcare, Chicago, IL, USA) using a semi-dry transfer apparatus (Bio-Rad) at 20 V for 45 min. Loading equivalence was confirmed by staining polyacrylamide gels with Coomassie brilliant blue or nitrocellulose membranes with Ponceau S. The membranes were blocked with 5% non-fat milk in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.02% Tween-20 (TBS-T) for 1 h at room temperature before incubation with primary antibodies overnight at 4 °C. Primary polyclonal antibodies against NifB^{Mi} (generated at the Centre for Plant Biotechnology and Genomics, CBGP) and monoclonal antibodies against strep-tag II (2-1507-001, IBA Lifesciences, Göttingen, Germany), the HA tag (H6908, Sigma-Aldrich, St Louis, MO, USA) or Rubisco (as loading control) were diluted at 1: 2,000-1: 5,000 in TBS-T supplemented with 5% bovine serum albumin (BSA). Secondary antibodies (Thermo Fisher Scientific) were diluted at 1: 20,000 in TBS-T supplemented with 2% non-fat milk and incubated for 2 h at room temperature. Membranes were developed on medical X-ray films (AGFA, Mortsel, Belgium) using enhanced chemiluminescence.

2.3.5. Purification of *Os*NifB proteins by strep-tag affinity chromatography

*Os*NifB^{Mi} (hereafter *Os*NifB^{Mi} to indicate expression in rice) and *Os*NifB^{Mt} (hereafter *Os*NifB^{Mt}) were prepared for strep-tag affinity chromatography (STAC) purification at O_2 levels below 1 ppm in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI, USA or MBraun, Garching, Germany). Callus was disrupted in lysis buffer comprising 100 mM Tris-HCl (pH 8.5), 300 mM NaCl, 10% glycerol, 3 mM sodium dithionite (DTH), 5 mM 2-mercaptoethanol, 1 mM PMSF, 1 µg/mL leupeptin, 10 µg/mL DNAse I, and 1:200 (v/v) BioLock solution (IBA Lifesciences Göttingen, Germany) at a ratio of 1:3

(w/v). Total extracts were prepared by lysing the cell suspensions under anaerobic conditions using the Oster 4655 blender (Newell Brands, Atlanta, GA, USA) modified with a water-cooling system operating at full speed in 4 cycles of 2 min at 4 ℃. Extracts were transferred to centrifuge tubes equipped with sealing closures (Beckman Coulter, Brea, CA, USA) and centrifuged (50,000*g*, 1.5 h, 4°C) using a Beckman Coulter Avanti J-26 XP device. The supernatant was passed through Nalgene 0.2 µm filter cups (Thermo Fisher Scientific) to yield a cell-free extract of soluble proteins. This was loaded at 2.5 mL/min onto a 5-mL Strep-Tactin XP column (IBA Lifesciences) attached to an ÄKTA FPLC system (GE Healthcare). The column was washed with 150 mL of 100 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10% glycerol, 2 mM DTH and 5 mM 2-mercaptoethanol at 16°C, and bound proteins were eluted with 15-20 mL of the same wash buffer supplemented with 50 mM biotin (IBA LifeSciences). The elution fraction was concentrated using the Amicon Ultra centrifugal filter (Millipore Sigma, Burlington, MA, USA) with a cut-off size of 10 kDa. Biotin was removed by passing the protein through PD-10 desalting columns (GE Healthcare) equilibrated with wash buffer. The desalted eluate was concentrated and snap-frozen in Nalgene cryovials and stored in liquid nitrogen.

2.3.6. Quantification of purified *Os*NifB proteins

The yields of the purified *Os*NifB^{Mi} and *Os*NifB^{Mt} were quantified by the densitometric analysis of Coomassie gel titration. Images of the Coomassie SDS-PAGE gels were analysed using ImageJ (Davarinejad 2015). The standards were the purified *S. cerevisiae* NifB (*Sc*NifB^{Mi} and *Sc*NifB^{Mt}) proteins. The standard loading amounts were 2 µg, 1 µg, 0.5 µg, 0. 25 µg, 125 ng and 62.5 ng. The concentration of purified *Os*NifB was determined by the comparison of the optical density of bands among *Os*NifB proteins and standard.

2.3.7. N-terminal sequencing of purified OsNifB proteins

N-terminal amino acid sequencing of purified *Os*NifB^{Mi} and *Os*NifB^{Mi} was determined by Edman degradation (Centro de Investigaciones Biológicas, Madrid, Spain). 25 pmol *Os*NifB protein was separated by SDS-PAGE, transferred to 0.2 µm Sequi-Blot PVDF membranes (Thermo Fisher Scientific) in 50 mM borate buffer (pH 9.0), stained with freshly prepared 0.1% Coomassie R-250 (Sigma-Aldrich) in 40% methanol and 10% acetic acid, and then destained using 50% methanol.

2.3.8. FeMo-co synthesis and apo-NifDK reconstitution in vitro

FeMo-co synthesis and apo-NifDK reconstitution assays were carried out *in vitro* in an anaerobic chamber, as previously described (Burén et al. 2019). For the *in vitro* synthesis of FeMo-co, each 100-µL reaction contained 3 µM NifH^{Av}, 1 µM *Os*NifB, 1.5 µM apo-NifEN^{Av}, 0.6 µM apo-NifDK^{Av}, 17.5 µM Na_2MoO_4, 175 µM *R*-homocitrate, 9 µM [Fe_4-S_4] cluster-loaded *Ec*NifU^{Av} (holo- *Ec*NifU^{Av}), 125 µM SAM, 1 mg/mL BSA, and the ATP-regenerating mixture (1.23 mM ATP, 18 mM phosphocreatine disodium salt, 2.2 mM $MgCl_2$, 3 mM dithionite, 46 µg/mL creatine phosphokinase). For the positive control FeMo-co synthesis assay, holo-NifU^{Ec} was omitted and *Os*NifB was replaced with 2.5 µM NifB-co. The reactants were incubated for 60 min at 30 ℃. For the acetylene reduction assays, 500 µL of the ATP-regenerating mixture and 2.0 µM NifH^{Av} were added to the reaction tube. The reaction mixture was then transferred to 9 mL serum vials under an argon/acetylene (94%/6%) atmosphere. The reaction was incubated for 20 min at 30 ℃. To measure ethylene formation, 50 µL of the gas phase was taken from the reaction vials and injected in the Shimadzu GC-2014 gas chromatographer equipped with the Porapak N 80/100 column (Shimadzu, Kyoto, Japan).

2.3.9. Statistical analysis

Standard deviation (SD) of relative mRNA expression data was calculated based on three technical replicates. SD of *in vitro* activity data was calculated based on two biological replicates (each one with two technical replicates).

2.4. Results

2.4.1. Expression constructs for rice transformation

S. cerevisiae codon-optimized *NifB* from *M. infernus* (*NifB*^{Mi}) and *M. thermautotrophicus* (*NifB*^{Mt}), and *FdxN* from *A. vinelandii* (*FdxN*^{Av}) were used for rice transformation because there are no rare codons in these three gene sequences in the context of rice codon usage (Nakamura et al. 2000). The *NifB*^{Mi}, *NifB*^{Mt} and *FdxN*^{Av} were introduced into separate vectors for rice transformation. Expression was driven by the strong constitutive *ZmUbi1*+1sti promoter. An N-terminal mitochondrial leader sequence Cox4 was added to direct the proteins to mitochondria because this sequence was previously shown to target recombinant enhanced green fluorescent protein (eGFP) to rice mitochondria effectively (Baysal et al. 2020). An N-terminal TS tag was added between the Cox4 signal and the *Os*NifB^{Mi} or *Os*NifB^{Mt} proteins for detection and purification of *Os*NifB. A C-terminal

HA tag was added to *Os*FdxN^{Av} (hereafter *Os*FdxN^{Av} to indicate expression in rice) to enable immunodetection. The *NifB*^{Mi} and *NifB*^{Mt} constructs were used separately, in each case combined with the *FdxN*^{Av} and a third construct carrying the selectable marker *Hpt* gene. Transgenic rice callus expressing *Nif* transgenes were produced by direct DNA transfer as described (Sudhakar et al. 1998; Valdez et al. 1998). Plantlets were regenerated from the corresponding callus lines under hygromycin selection and grown to maturity as described (Sudhakar et al. 1998; Valdez et al. 1998).

2.4.2. *NifB* and *FdxN* mRNA expression in transgenic rice callus and plants

NifB^{Mi}, *NifB*^{Mt} and *FdxN*^{Av} expression in callus and plants was confirmed by RT-qPCR analysis. Eight independent lines (MiB29, MiB32, MiB84, MiB115, MtB35, MtB194, MtB198, and MtB235, four from each NifB transgene) were used for RT-qPCR analysis (**Figure. 2.1**). The mRNA expression analysis in callus and the corresponding regenerated plants were performed in three independent lines (MiB32, MiB115 and MtB35) (**Figure. 2.2**). The relative expression levels of *NifB*^{Mi} and *NifB*^{Mt} were similar in callus and in the corresponding regenerated plants. However, at the plant level both *NifB*^{Mi} and *NifB*^{Mt} mRNA expression was two orders of magnitude higher compared to callus. *FdxN*^{Av} co-expression was confirmed in callus and plants expressing *NifB*^{Mi} or *NifB*^{Mt} (**Figures 2.1 and 2.2**).

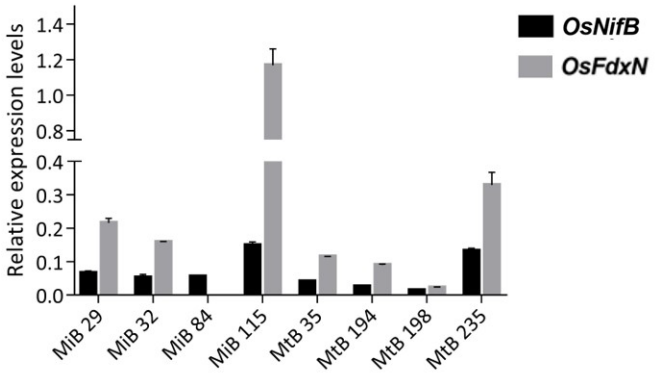

Figure 2.1 *NifB* and *FdxN* mRNA accumulation in representative Os*NifB*^{Mi} and *OsNifB*^{Mt} callus lines. Four independent lines from each transgene were used for RT-qPCR analysis**.** Relative mRNA expression levels were normalized to *OsActin*. Data are means ± SD (n = 3). *OsNifB*, *O. sativa*-derived *M. infernus NifB* and *O. sativa*-derived *M. thermautotrophicus NifB*; *OsFdxN*, *O. sativa*-derived *A. vinelandii FdxN*. MiB29, MiB32, MiB84, MiB115, MtB35, MtB194, MtB198, and MtB235 are eight independent lines.

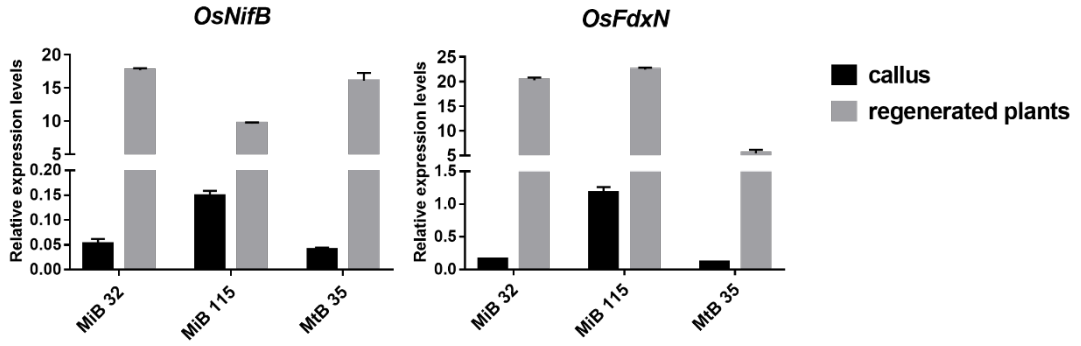

Figure 2.2 *NifB* and *FdxN* mRNA accumulation in callus and corresponding regenerated plants. Three independent lines were used for RT-qPCR analysis**.** Relative mRNA expression levels were normalized to *OsActin*. Data are means ± SD (n = 3). *OsNifB*, *O. sativa*-derived *M. infernus NifB* and *O. sativa*-derived *M. thermautotrophicus NifB*; *OsFdxN*, *O. sativa*-derived *A. vinelandii FdxN*. MiB32, MiB115 and MtB35 are three independent lines.

2.4.3. Accumulation of *Os*NifB and *Os*FdxN protein in transgenic rice callus and plants

Accumulation of *Os*NifB^{Mi}, *Os*NifB^{Mt} and *Os*FdxN^{Av} protein in callus and the corresponding regenerated plants was confirmed by immunoblot analysis using antibodies specific for the NifB^{Mi} and TS, and HA tags from lines MiB32, MiB35 and MtB35 which accumulated the recombinant proteins (*Os*NifB and *Os*FdxN) at the highest levels (**Figure 2.3**). Based on their SDS-gel migration patterns, the recombinant proteins were correctly processed, with the expected molecular weights of 38 kDa (*Os*NifB^{Mi}), 35 kDa (*Os*NifB^{Mt}) and 11 kDa (*Os*FdxN^{Av}). As reported previously when expressed in *S. cerevisiae* (Burén et al. 2017a; Burén et al. 2019), the migration of the HA-tagged FdxN^{Av} protein was less distinct, probably due to its smaller size. I thus confirmed that *Os*NifB^{Mi}, *Os*NifB^{Mt} and *Os*FdxN^{Av} proteins accumulated in soluble form in rice callus (**Figure 2.3a**). *Os*NifB^{Mi} and *Os*FdxN^{Av} were soluble in regenerated plants as well (**Figure 2.3b**). I was not able to detect accumulation of *Os*NifB^{Mt} in regenerated MtB35 plants.

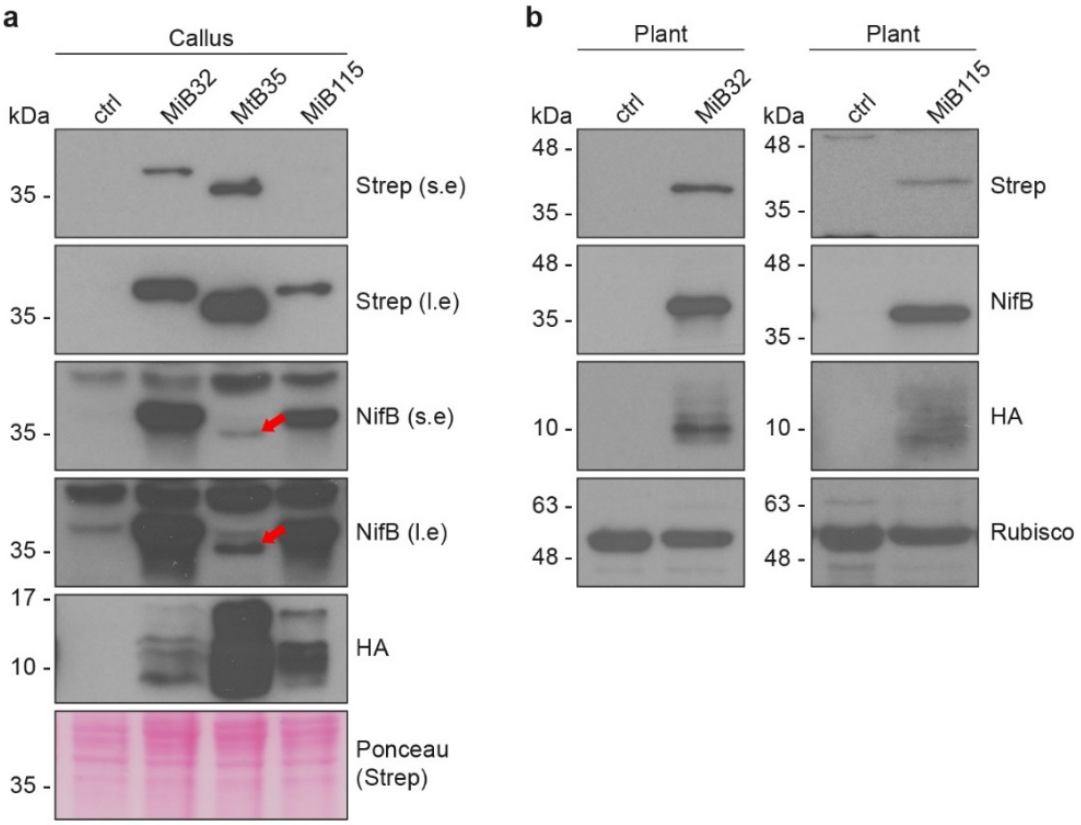

Figure 2.3 Accumulation of *Os*NifB^{Mi}, *Os*NifB^{Mt} and *Os*FdxN^{Av} protein in callus and plants. Immunoblot analysis of cell-free extracts prepared from callus (a) and plants (b). *Os*NifB^{Mi} and *Os*NifB^{Mt} were detected with antibodies against NifB^{Mi} and the N-terminal TS tag. *Os*FdxN^{Av} was detected with antibodies against the C-terminal HA tag. The red arrow indicates the signal from *Os*NifB^{Mt} detected with antibodies against NifB^{Mi}. *Os*NifB^{Mi}, *O. sativa*-derived *M. infernus* NifB; *Os*NifB^{Mt}, *O. sativa*-derived *M. thermautotrophicus* NifB; *Os*FdxN^{Av}, *O. sativa*-derived *A. vinelandii* FdxN; s.e., short exposure during immunoblot detection; l.e., long exposure during immunoblot detection; MiB32, MiB115, and MtB35 are three independent lines. N.B. *Os*NifB^{Mt} was not detectable in multiple regenerated siblings from line MtB35.

2.4.4. Purification of *Os*NifB^{Mi} and *Os*NifB^{Mt} produced in rice tissue

The *Os*NifB^{Mi} and *Os*NifB^{Mt} proteins were purified from rice callus (lines MiB32, MiB115 and MtB35) by strep-tag affinity chromatography (STAC) and the purification process was monitored by sampling the total extract, cell-free extract, flow-through, wash, and elution fractions for analysis by SDS-PAGE and immunoblotting (**Figure 2.4a-c**). No significant amount of protein was lost during centrifugation and filtration of the cell extract, confirming that both *Os*NifB^{Mi} and *Os*NifB^{Mt} proteins were soluble in the mitochondria. The elution fractions featured a band matching the anticipated size of correctly processed *Os*NifB^{Mi} *or Os*NifB^{Mt}. Isolated NifB yields were 44 and 87 µg per 100 g fresh weight callus for *Os*NifB^{Mi} and *Os*NifB^{Mt}, respectively (**Figure 2.5**). Side by side comparison of NifB^{Mi} and NifB^{Mt} proteins isolated from yeast and rice confirmed the correct targeting and the specific processing of Cox4 signals from both *Os*NifB proteins (**Figure 2.4d, e**) (Burén et al. 2019). The exact cleavage site of the Cox4 sequence was

further investigated by N-terminal sequencing. Cleavage was specific after amino acid 26 in the Cox4 peptide (**Figure 2.4f, g**), which is only one amino acid away from where the endogenous Cox4 protein is processed in *S. cerevisiae* (Geier et al. 1995), generating a single *Os*NifB protein moiety. Removal of the Cox4 signal following the import of *Os*NifB^{Mi} and *Os*NifB^{Mt} to the mitochondria, therefore confirmed successful targeting.

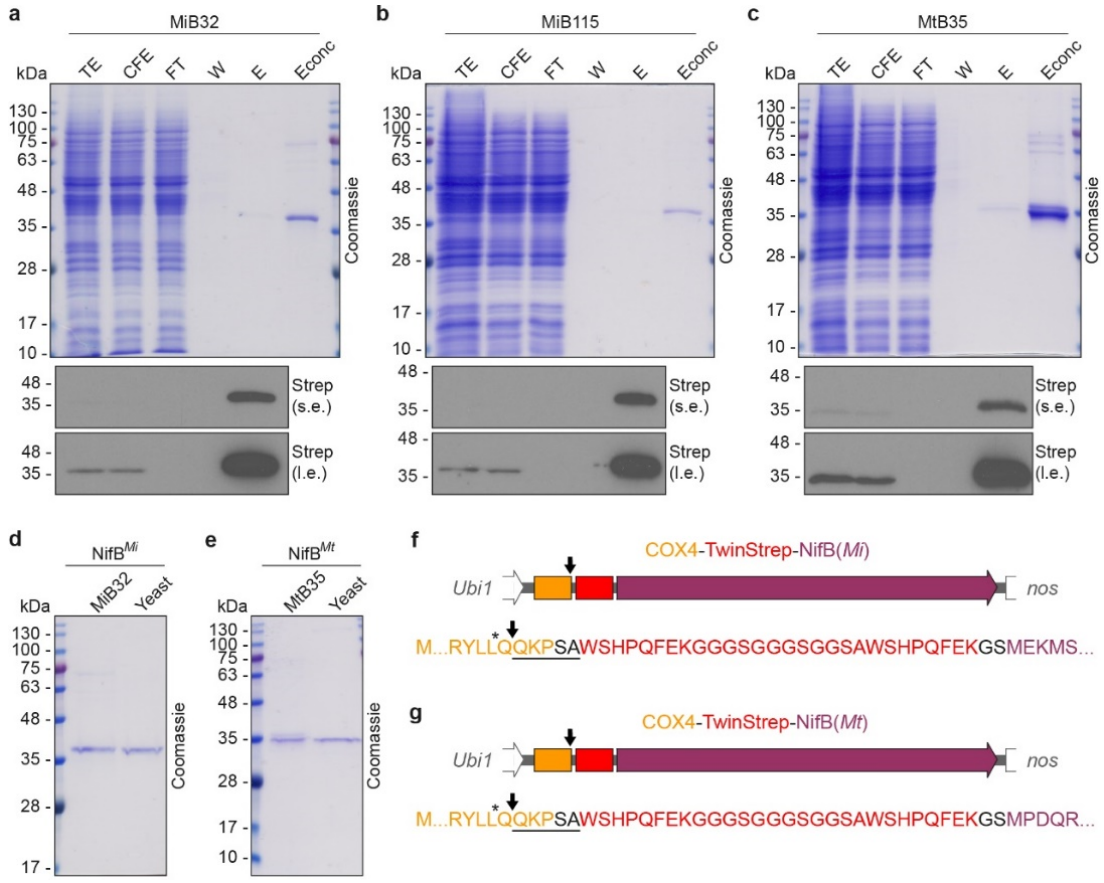

Figure 2.4 STAC purification and N-terminal sequence of *Os*NifB^{Mi} and *Os*NifB^{Mt}. Purification of *Os*NifB from MiB32 (a), MiB115 (b) and MtB35 (c) callus. TE: total extract, CFE: soluble cell-free extract, FT: flow-through fraction, W: wash fraction, and E: elution fraction. Fractions were analyzed by SDS-PAGE, followed by Coomassie gel staining or immunoblot analysis using antibodies detecting the TS tag. Migration of STAC-purified *Os*NifB^{Mi} (MiB32) and *Sc*NifB^{Mi} (d), and *Os*NifB^{Mt} (MtB35) and *Sc*NifB^{Mt} (e). Cleavage sites of Cox4 in *Os*NifB^{Mi} (f) and *Os*NifB^{Mt} (g). The black arrow indicates the N-terminal processing site as determined by N-terminal sequencing. The underlined amino acids represent those detected by the Edman degradation procedure. The black stars indicate the cleavage site for endogenous Cox4 in *S. cerevisiae*.

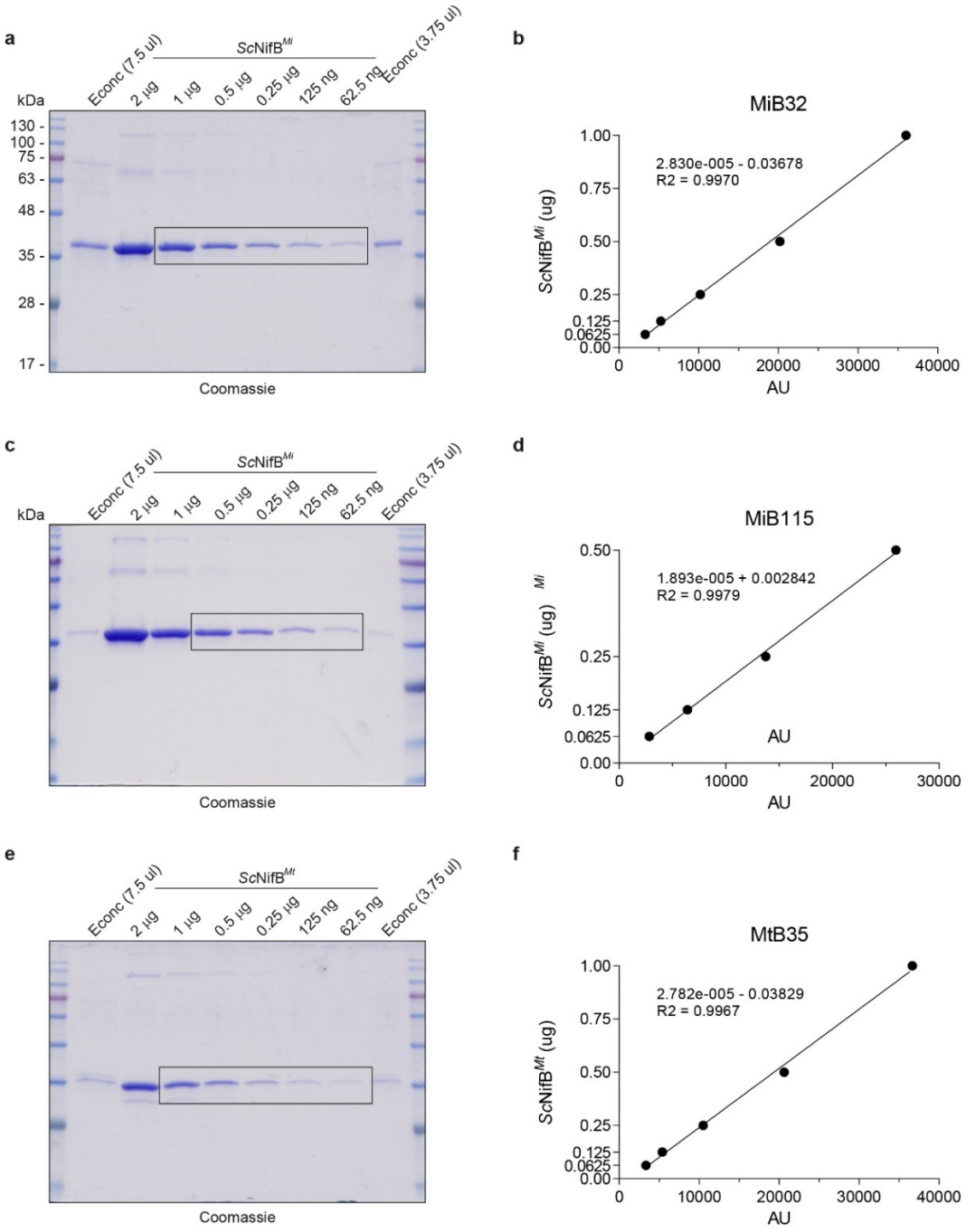

Figure 2.5 Coomassie staining and quantification of purified *Os*NifB. *Os*NifB^{Mi} and *Os*NifB^{Mt} were purified from line MiB32 (a and b), MiB115 (c and d) and MtB35 (e and f). Coomassie stained gels were loaded with known amounts of *Sc*NifB^{Mi} (a and c) and *Sc*NifB^{Mt} (e) to quantify the concentration of the eluted and concentrated *Os*NifB^{Mi} (b and d) and *Os*NifB^{Mt} (f) using linear regression. The slope, intercept and the R2 for the linear regression analyses are indicated. Black boxes indicate the Coomassie stained *Sc*NifB proteins used for the analyses.

*Os*NifB^{Mi} was also purified from regenerated plants following a similar procedure. Purification fractions were analyzed by SDS-PAGE and immunoblotting (**Figure 2.6**). The lower intensity of the plant-produced *Os*NifB^{Mi} band was shown in Coomassie staining (**Figure 2.6**). Thus, the quantitative method was optimized by reducing standard protein loading and using immunoblots to detect protein signals. The yield of isolated

*Os*NifB^{Mi} from plants was decreased to 1% compared to callus, ca: 1.4 µg per 100 g fresh weight rice leaves (**Figure 2.7**).

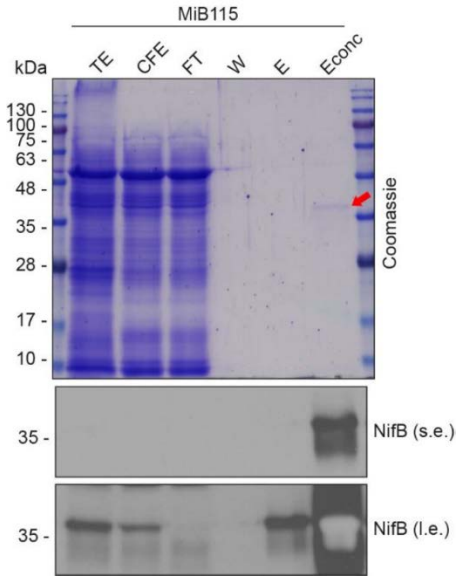

Figure 2.6 STAC purification of *Os*NifB^{Mi} from MiB115 plants. TE, total extract; CFE, soluble cell-free extract; FT, flow-through fraction; W, wash fraction; and E, elution fraction. Fractions were analyzed by SDS-PAGE, followed by Coomassie gel staining or immunoblot analysis using antibodies detecting the NifB^{Mi}. The red arrow indicates the signal from the plant-produced *Os*NifB^{Mi} detected with antibodies against NifB^{Mi}.

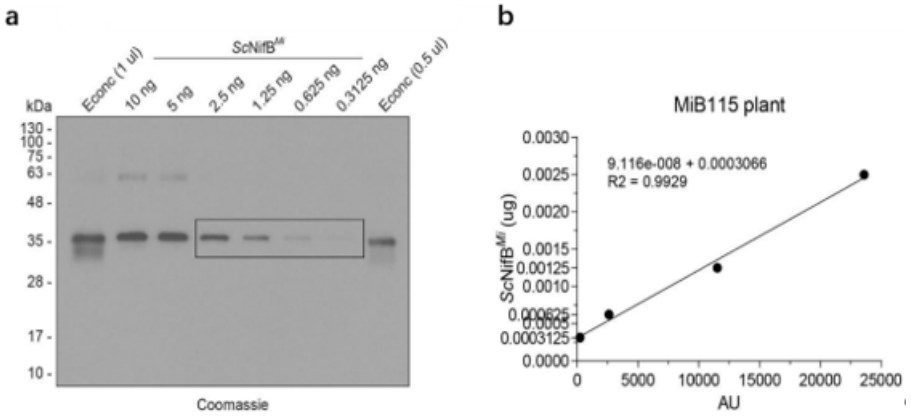

Figure 2.7 Immunoblot and quantification of plant-produced *Os*NifB^{Mi} purified from line MiB115. Loading a known amount of *Sc*NifB^{Mi} (a) was used to quantify the concentration of the concentrated *Os*NifB^{Mi} (b) using linear regression. The slope, intercept and the R^2 for the linear regression analyses are indicated. Black boxes indicate the immunoblot signal for *Sc*NifB proteins used for the analyses.

2.4.5. *Os*NifB^{Mi} and *Os*NifB^{Mt} catalyze FeMo-co synthesis *in vitro*

The minimal protein components for FeMo-co synthesis *in vitro* are NifB, NifEN and NifH (Curatti et al. 2007). The isolated *Os*NifB^{Mi} or *Os*NifB^{Mt} protein was mixed *with*

[4Fe-4S] cluster-loaded NifU^{Av} purified from *E. coli* (*Ec*NifU^{Av}, as the source of [4Fe-4S] precursor clusters for NifB-co biosynthesis), *A. vinelandii* NifEN with the permanent [4Fe-4S] clusters but lacking FeMo-co precursor cluster (apo-NifEN^{Av}), *A. vinelandii* NifH protein (NifH^{Av}), and *A. vinelandii* NifDK with the P-cluster but lacking FeMo-co (apo-NifDK^{Av}). Molybdate, *R*-homocitrate and SAM were added as they are the required substrates for NifB-dependent *in vitro* FeMo-co synthesis.

The as-isolated *Os*NifB^{Mi} and *Os*NifB^{Mt} proteins were colorless and inactive in FeMo-co synthesis. However, when loaded with [4Fe-4S] clusters from NifU, the *Os*NifB^{Mi} and *Os*NifB^{Mt} proteins were functional in the FeMo-co synthesis assay, in which the *Os*NifB-dependent activation of apo-NifDK was measured by *in vitro* acetylene reduction activity of the reconstituted enzyme (**Figure 2.8**). This shows that the NifB-co produced by *Os*NifB matured into FeMo-co at the NifEN/NifH complex, which was then transferred to apo-NifDK^{Av}. The OsNifB^{Mi}-dependent activation of apo-NifDK resulted in nitrogenase activities of 35 ± 0.95 and 19 ± 0.94 nmol C_2H_4 min^{-1} mg^{-1} using *Os*NifB^{Mi} isolated from lines MiB32 and MiB115, respectively, while FeMo-co synthesis using OsNifB^{Mt} isolated from MtB35 resulted in four-fold higher nitrogenase activities (137 nmol C_2H_4 min^{-1} mg^{-1}) (**Figure 2.8**).

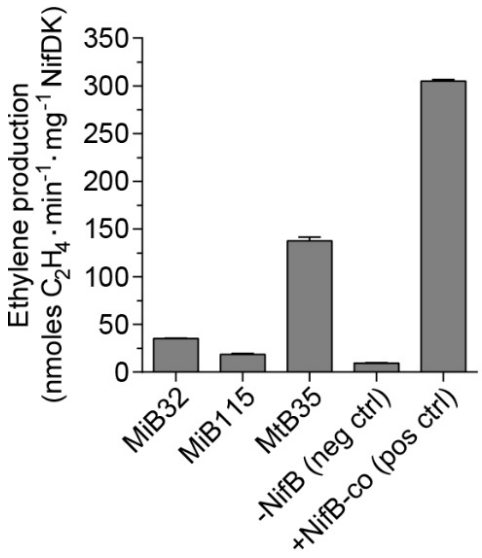

Figure 2.8 *In vitro* FeMo-co synthesis and apo-NifDK reconstitution using the as-isolated *Os*NifB^{Mi} and *Os*NifB^{Mt} proteins supplemented with [4Fe-4S] cluster substrates. Activity is represented as nanomoles of ethylene produced per minute and milligram of NifDK. The activity of the positive control reaction for FeMo-co synthesis (containing pure NifB-co instead of NifB) was 305 ± 2 units, and the activity of the ATP-mix control reaction (containing holo-NifDK) was 1506 ± 95 units. MiB32, MiB115 and MtB35 denote *Os*NifB^{Mi} or *Os*NifB^{Mt} isolated from three independent lines. Data are means ± SD (n = 2).

To rule out that this higher *Os*NifB^{Mt} activity was affecting cell growth and development thus precluding the generation of plants expressing *Os*NifB^{Mt} (**Figures 2.3**), I generated more lines expressing *Os*NifB^{Mt}. The MtB15 line expressed *Os*NifB^{Mt} at high levels (in addition to *Os*FdxN^{Av}) in callus **(Figure 2.9a)**, but also in leaves of the corresponding regenerated plants **(Figure 2.9b)**, which indicates that expression of NifB^{Mt} is likely not detrimental to the plants. Similarly, NifB^{Mi} expression was shown to be stable in T1 plants of the MiB115 line (**Figure 2.10**).

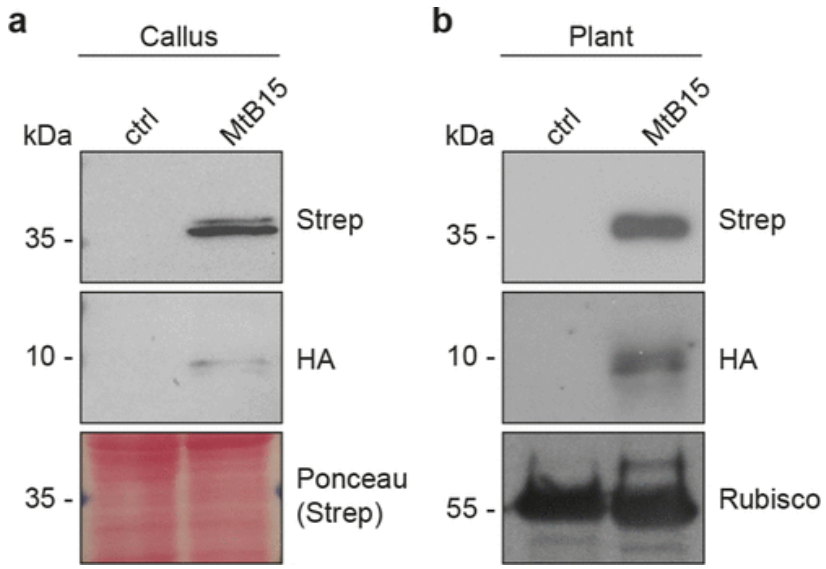

Figure 2.9 Accumulation of *Os*NifB^{Mt} and *Os*FdxN^{Av} in rice callus (a) and plants (b). *Os*NifB^{Mt} was detected with antibodies against the N-terminal TS tag. *Os*FdxN^{Av} was detected with antibodies against the C-terminal HA tag. *Os*NifB^{Mt}, *O. sativa*-derived *M. thermautotrophicus* NifB; *Os*FdxN^{Av}, *O. sativa*-derived *A. vinelandii* FdxN. MtB15 is a line accumulating *Os*NifB^{Mt} in callus and leaves.

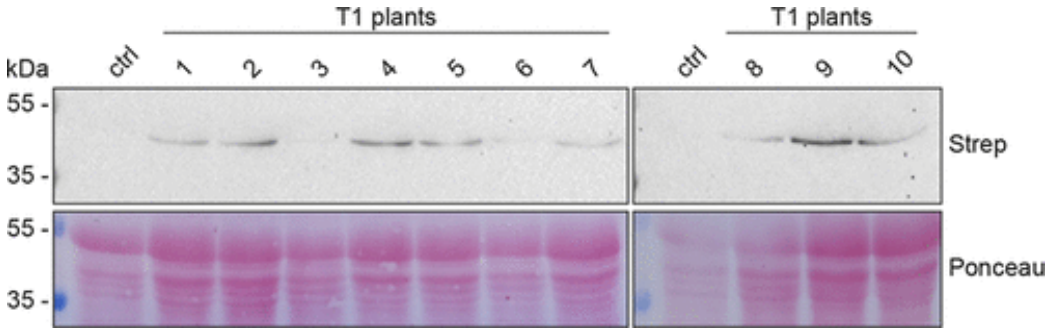

Figure 2.10 Expression of *Os*NifB^{Mi} in 10 different T1 plants from the MiB115 line. Immunoblot analysis was performed using leaf soluble protein extracts and antibodies detecting the TS tag. The control (ctrl) lane was loaded with leaf soluble protein extracts obtained from wild-type *O. sativa* plants.

2.5. Discussion

The engineering of staple crops to fix nitrogen has been an important goal of plant biotechnology for several decades. If successful, this approach offers the potential to reduce or even abolish our dependence on nitrogen fertilizers while maintaining the nitrogen content of soils. Natural BNF occurs only in some prokaryotes, and is catalyzed by a nitrogenase complex with assistance from various accessory proteins required to assemble and incorporate metal cofactors into nitrogenase (Burén et al. 2020). Many of these components are extremely sensitive to O_2, which is an additional challenge when transferring the trait to plants (Shah and Brill 1973). One solution is to express the nitrogenase and its accessory proteins in the plant mitochondria, a strategy that would reduce O_2 exposure and supply energy for nitrogenase activity and a ready source of [Fe-S] clusters generated by proteins similar to the bacterial NifUS system (Curatti and Rubio 2014).

Although many *Nif* genes are involved in the assembly and activity of nitrogenase and its metal cofactors in bacteria, not all are expected to be required to reconstitute nitrogenase activity in plants because some of the accessory functions can be fulfilled by endogenous proteins (Yang et al. 2017). The minimal gene set that must be transferred to plants includes NifD, NifK and NifH, which form the nitrogenase enzyme, and NifE, NifN and NifB, which catalyze essential reactions in the biosynthesis of FeMo-co, the active-site cofactor of nitrogenase (Burén et al. 2020). These six genes have been expressed in the mitochondria of *S. cerevisiae* and, in some cases, also transiently in *N. benthamiana* (Allen et al. 2017; Burén et al. 2017a; Burén et al. 2017b; Burén et al. 2019; Allen et al. 2020; Okada et al. 2020; Jiang et al. 2021), but have yet to be expressed in any staple crop.

The solubility of NifB is a prerequisite for its activity. Previous work on *S. cerevisiae* showed that the NifB proteins from both *M. thermautotrophicus* and *M. infernus* were soluble in the mitochondria and accumulated at high levels (Burén et al. 2019). I therefore generated rice plants co-expressing *Os*NifB^{Mi}, or *Os*NifB^{Mt}, with *Os*FdxN^{Av}. In initial experiments, the expression frequency and levels of *Os*NifB appeared to depend on the NifB variant. While I could detect *Os*NifB^{Mi} in three out of four callus lines I analyzed, *Os*NifB^{Mt} was only detectable in one out of four lines. At the plant level, protein accumulation could only be measured in *Os*NifB^{Mi} lines. I hypothesized that *Os*NifB^{Mt} might be detrimental to cell growth and development, thus limiting the number of plants able to regenerate, when expressing the protein. I therefore initiated new transformation

experiments aiming to generate more lines expressing *Os*NifB^{Mt}. Indeed, it proved to be difficult to generate additional lines expressing this protein. It is possible that NifB activity might interfere with other essential developmental processes in the cells or compete for essential precursors in metabolic processes sharing common precursors. However, I was able to obtain one line that accumulated *Os*NifB^{Mt} in both callus and regenerated plants (**Figure 2.9)**. While moving forward with the nitrogenase engineering process, it would be desirable to circumvent this problem, for example, by using tissue-specific or regulated promoters.

Likely reasons to explain the different outcomes when expressing NifB^{Mi} and NifB^{Mt} in rice plants are not clear at present. An overlay of NifB^{Mi} and NifB^{Mt} structures (**Figure 2.11**) shows that secondary structure elements and relevant residues in both structures match, with only two differences : (1) the NifB^{Mi} H^{22} residue, which appears to stabilize the K-cluster (Jenner et al. 2021), has been modelled by AlphaFold in the rotated position compared to its equivalent H^{24} residue in NifB^{Mt}; and (2) the C-terminal stretch of NifB proteins. This region was well resolved in the NifB^{Mt} structure containing the K-cluster, but it was shown to be disordered before K-cluster formation (in the absence of K2-cluster) in the crystal structure of another archaean NifB homolog (Fajardo et al. 2020). It was proposed that the short C-terminal stretch acted as a strap closing the side of the NifB β-barrel structure and stabilizing the K2-cluster (Jenner et al. 2021). It should also be noted that the confidence of the AlphaFold model for this region of NifB^{Mi} was low.

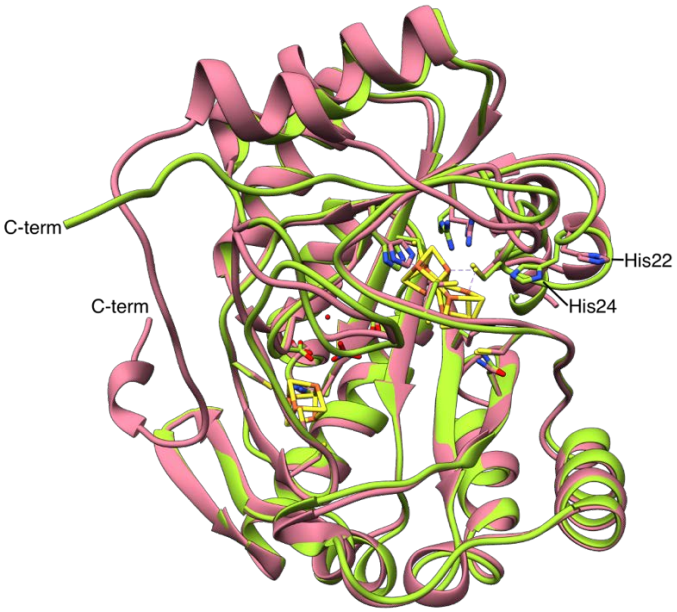

Figure 2.11 Overlay of NifB^{Mi} and NifB^{Mt} structures shown as ribbon diagrams. The NifB^{Mt} structure (green) was solved by X-ray crystallography (Jenner et al. 2021). The NifB^{Mi} structure (magenta) is an AlphaFold prediction of UniProt D5VRM1. The RS cluster and the [8Fe-8S] K-

cluster, formed by the fusion of K1- and K2-clusters, are shown in sticks together with residues important for their coordination or stability (Jenner et al. 2021). Modelling carried out by Dr Xi Jiang (Universidad Politecnica de Madrid Centro de Biotecnologia y Genomica de Plantas).

Earlier studies involved the co-expression of NifB with NifU, NifS and FdxN in *S. cerevisiae* (Burén et al. 2017a; Burén et al. 2019). Analysis of the *Sc*NifB^{Mi} protein isolated from different *S. cerevisiae* strains showed that while NifUS was important for providing [4Fe-4S] clusters, FdxN was more important for NifB activity (Burén et al. 2019). Earlier studies had shown that FdxN was required for efficient NifB-co biosynthesis (Jimenez-Vicente et al. 2014), but its exact role is still unknown. Several non-exclusive roles in NifB [Fe-S] cluster acquisition or maturation, or as a participant of the NifB reaction have been proposed (Burén et al. 2020). For example, NifB produced recombinantly in *S. cerevisiae* required FdxN to acquire the EPR signatures of its three clusters (Burén et al. 2019). FdxN could also promote the reductive coupling of K1- and K2-clusters to form the [8Fe-8S] K-cluster, a reaction intermediate of NifB-co synthesis (Jenner et al. 2021). Finally, FdxN could serve as an electron donor to the NifB RS-cluster for the reductive cleavage of SAM and release of the dA·radical.

In contrast to NifB, NifH showed similar activity (400 nmol of C_2H_4 min $^{-1}$ mg $^{-1}$ NifDK) when co-expressed with either NifM alone or with NifM, NifS, and NifU in *S. cerevisiae* (Lopez-Torrejon et al. 2016). This may reflect the distinct mechanisms used to incorporate [4Fe-4S] clusters into the NifB and NifH proteins, or different requirements for these clusters. While NifH contains a permanent [4Fe-4S] cluster only required for catalysis, NifB requires a [4Fe-4S] cluster for catalysis and two additional [4Fe-4S] clusters as substrates for NifB-co biosynthesis.

The host organism can also influence the activity of Nif proteins. For example, when NifH was co-expressed with NifM, NifU and NifS, it was functional in *S. cerevisiae* but not in *N. benthamiana*, and in the latter case reconstitution *in vitro* was necessary to restore activity (Jiang et al. 2021). Although NifU is the major provider of [4Fe-4S] clusters for nitrogenase *in vivo*, the *Klebsiella pneumoniae NifUS* double mutant still synthesizes NifB-co, albeit at lower levels compared to the wild-type stain (Zhao et al. 2007). This shows that [4Fe-4S] clusters for NifB-co biosynthesis can be provided by other sources, such as the iron-sulfur cluster assembly or sulfur mobilization systems (Zhao et al. 2007; Braymer et al. 2021). The IscS protein purified from an *A. vinelandii* strain with deleted *NifS* catalyzes the same reaction as NifS (Zheng et al. 1998). I have been unable to obtain detectable protein accumulation of NifU^{Av} and NifS^{Av} in

transformed rice, and this study was therefore limited to the expression of NifB and FdxN. The identification of NifU and NifS variants suitable for expression in rice should therefore be the focus of future studies.

2.6. Conclusions

In conclusion, I was able to express the *M. infernus* NifB, *M. thermautotrophicus* NifB and *A. vinelandii* FdxN proteins in rice, using the Cox4 peptide to ensure the efficient targeting of all three proteins to the mitochondria, where they were correctly processed. The enzymatic activity of purified *Os*NifB proteins was confirmed by using the *in vitro* FeMo-co synthesis assay. I therefore show that both NifB proteins fulfil the requirements for a functional NifB in the rice mitochondria, such as stability, solubility, and competence to acquire [4Fe-4S] cluster substrates, but further research is required to demonstrate active NifB *in planta*. This may require the co-expression of additional *Nif* genes. Nevertheless, the expression of functional NifB in this study is a critical step toward the engineering of nitrogenase activity in cereals.

2.7. References

Allen RM, Chatterjee R, Ludden PW, Shah VK. 1995. Incorporation of iron and sulfur from NifB cofactor into the iron-molybdenum cofactor of dinitrogenase. *J. Biol. Chem.* **270**: 26890-26896.

Allen RS, Gregg CM, Okada S, Menon A, Hussain D, Gillespie V, Johnston E, Devilla R, Warden AC, Taylor M et al. 2020. Plant expression of NifD protein variants resistant to mitochondrial degradation. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 23165-23173.

Allen RS, Tilbrook K, Warden AC, Campbell PC, Rolland V, Singh SP, Wood CC. 2017. Expression of 16 nitrogenase proteins within the plant mitochondrial matrix. *Front. Plant Sci.* **8**: 287.

Arragain S, Jiménez-Vicente E, Scandurra AA, Burén S, Rubio LM, Echavarri-Erasun C. 2017. Diversity and functional analysis of the FeMo-cofactor maturase NifB. *Front. Plant Sci.* **8**: 1947.

Baysal C, Perez-Gonzalez A, Eseverri A, Jiang X, Medina V, Caro E, Rubio LM, Christou P, Zhu C. 2020. Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Res.* **29**: 37-52.

Boyd ES AA, Miller S, Hamilton TL, Lavin M, Peters JW. . 2011. A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* **9**: 221-232.

Braymer JJ, Freibert SA, Rakwalska-Bange M, Lill R. 2021. Mechanistic concepts of iron-sulfur protein biogenesis in Biology. *Biochim. Biophys.Acta-Mo.l Cell Res.* **1868**: 118863.

Burén S, Jiang X, López-Torrejón G, Echavarri-Erasun C, Rubio LM. 2017a. Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NifB. *Front. Plant Sci.* **8**: 1567.

Burén S, Jiménez-Vicente E, Echavarri-Erasun C, Rubio LM. 2020. Biosynthesis of nitrogenase cofactors. *Chem. Rev.* **120**: 4921-4968.

Burén S, Pratt K, Jiang X, Guo Y, Jimenez-Vicente E, Echavarri-Erasun C, Dean DR, Saaem I, Gordon DB, Voigt CA et al. 2019. Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 25078-25086.

Burén S, Young EM, Sweeny EA, Lopez-Torrejón G, Veldhuizen M, Voigt CA, Rubio LM. 2017b. Formation of Nitrogenase NifDK Tetramers in the Mitochondria of Saccharomyces cerevisiae. *ACS Synth. Biol.* **6**: 1043-1055.

Burgess BK, Lowe DJ. 1996. Mechanism of Molybdenum Nitrogenase. *Chem. Rev.* **96**: 2983-3012.

Christiansen J, Goodwin PJ, Lanzilotta WN, Seefeldt LC, Dean DR. 1998. Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a *nifB*-deletion mutant of *Azotobacter vinelandii*. *Biochem.* **37**: 12611-12623.

Creissen GP, Mullineaux PM. 1995. Cloning and characterisation of glutathione reductase cDNAs and identification of two genes encoding the tobacco enzyme. *Planta* **197**: 422-425.

Curatti L, Hernandez JA, Igarashi RY, Soboh B, Zhao D, Rubio LM. 2007. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 17626-17631.

Curatti L, Ludden PW, Rubio LM. 2006. NifB-dependent i*n vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 5297-5301.

Curatti L, Rubio LM. 2014. Challenges to develop nitrogen-fixing cereals by direct *nif*-gene transfer. *Plant Sci.* **225**: 130-137.

Davarinejad H. 2015. Quantifications of western blots with ImageJ. *University of York*.

Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genom.* **13**: 162.

Fajardo AS, Legrand P, Paya-Tormo LA, Martin L, Pellicer Marti Nez MT, Echavarri-Erasun C, Vernede X, Rubio LM, Nicolet Y. 2020. Structural insights into the mechanism of the Radical SAM carbide synthase NifB, a key nitrogenase cofactor maturating enzyme. *J. Am. Chem. Soc.* **142**: 11006-11012.

Fay AW, Wiig JA, Lee CC, Hu Y. 2015. Identification and characterization of functional homologs of nitrogenase cofactor biosynthesis protein NifB from methanogens. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 14829-14833.

Geier BM, Schagger H, Ortwein C, Link TA, Hagen WR, Brandt U, Von Jagow G. 1995. Kinetic properties and ligand binding of the eleven-subunit cytochrome-c oxidase from *Saccharomyces cerevisiae* isolated with a novel large-scale purification method. *Eur. J. Biochem.* **227**: 296-302.

Guo Y, Echavarri-Erasun C, Demuez M, Jimenez-Vicente E, Bominaar EL, Rubio LM. 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.

Halbleib CM, Ludden PW. 2000. Regulation of Biological Nitrogen Fixation. *J. Acad. Nutr. Diet.* **130**: 1081-1084.

Harwood CS. 2020. Iron-only and vanadium nitrogenases: fail-safe enzymes or something more? *Annu. Rev. Microbiol.* **74**: 247-266.

Hawkes TR, Smith BE. 1983. Purification and characterization of the inactive MoFe protein (NifB-Kp1) of the nitrogenase from nifB mutants of Klebsiella pneumoniae. *Biochem. J.* **209**: 43-50.

Hu Y, Corbett MC, Fay AW, Webber JA, Hedman B, Hodgson KO, Ribbe MW. 2005. Nitrogenase reactivity with P-cluster variants. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 13825-13830.

Hu Y, Fay AW, Lee CC, Ribbe MW. 2007. P-cluster maturation on nitrogenase MoFe protein. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 10424-10429.

Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR. 1989. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. *Mol. Gen. Genet.* **219**: 49-57.

Jenner LP, Cherrier MV, Amara P, Rubio LM, Nicolet Y. 2021. An unexpected P-cluster like intermediate en route to the nitrogenase FeMo-co. *Chem. Sci.* **12**: 5269-5274.

Jiang X, Coroian D, Barahona E, Echavarri-Erasun C, Castellanos-Rueda R, Eseverri Á, Aznar-Moreno Jose A, Burén S, Rubio Luis M. 2022. Functional Nitrogenase Cofactor Maturase NifB in Mitochondria and Chloroplasts of Nicotiana benthamiana. *mBio* **13**: e00268-00222.

Jiang X, Paya-Tormo L, Coroian D, Garcia-Rubio I, Castellanos-Rueda R, Eseverri A, Lopez-Torrejon G, Buren S, Rubio LM. 2021. Exploiting genetic diversity and gene synthesis to

identify superior nitrogenase NifH protein variants to engineer N_2-fixation in plants. *Commun. Biol.* **4**: 4.

Jimenez-Vicente E, Navarro-Rodriguez M, Poza-Carrion C, Rubio LM. 2014. Role of *Azotobacter vinelandii* FdxN in FeMo-co biosynthesis. *FEBS Lett.* **588**: 512-516.

Jimenez-Vicente E, Yang ZY, Martin Del Campo JS, Cash VL, Seefeldt LC, Dean DR. 2019. The NifZ accessory protein has an equivalent function in maturation of both nitrogenase MoFe protein P-clusters. *J. Biol. Chem.* **294**: 6204-6213.

Jin X, Bai C, Bassie L, Nogareda C, Romagosa I, Twyman RM, Christou P, Zhu C. 2019. ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (Zea mays) β-carotene hydroxylase 2 gene. *New Phytol.* **222**: 793-804.

Joerger RD, Bishop PE. 1988. Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. *J. Bacteriol.* **170**: 1475-1487.

Lopez-Torrejon G, Jimenez-Vicente E, Buesa JM, Hernandez JA, Verma HK, Rubio LM. 2016. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat. Commun.* **7**: 11426.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**: 292.

Okada S, Gregg CM, Allen RS, Menon A, Hussain D, Gillespie V, Johnston E, Byrne K, Colgrave ML, Wood CC. 2020. A synthetic biology workflow reveals variation in processing and solubility of nitrogenase proteins targeted to plant mitochondria, and differing tolerance of targeting sequences in a bacterial nitrogenase assay. *Front. Plant Sci.* **11**: 552160.

Peters JW, Stowell MHB, Soltis SM, Finnegan MG, Johnson MK, Rees DC. 1997. Redox-Dependent Structural Changes in the Nitrogenase P-Cluster. *Biochem.* **36**: 1181-1187.

Rubio LM, Ludden PW. 2008. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu. Rev. Microbiol.* **62**: 93-111.

Seefeldt LC, Yang ZY, Lukoyanov DA, Harris DF, Dean DR, Raugei S, Hoffman BM. 2020. Reduction of substrates by nitrogenases. *Chem. Rev.* **120**: 5082-5106.

Shah VK, Allen JR, Spangler NJ, Ludden PW. 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**: 1154-1158.

Shah VK, Brill WJ. 1973. Nitrogenase. IV. Simple method of purification to homogeneity of nitrogenase components from *Azotobacter vinelandii*. *Biochim. Biophys. Acta.* **305**: 445-454.

Shah VK, Imperial J, Ugalde RA, Ludden PW, Brill WJ. 1986. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 1636-1640.

Smith AD, Jameson GN, Dos Santos PC, Agar JN, Naik S, Krebs C, Frazzon J, Dean DR, Huynh BH, Johnson MK. 2005. NifS-mediated assembly of [4Fe-4S] clusters in the N- and C-terminal domains of the NifU scaffold protein. *Biochem.***44**: 12955-12969.

Soboh B, Igarashi RY, Hernandez JA, Rubio LM. 2006. Purification of a NifEN protein complex that contains bound molybdenum and a FeMo-co precursor from an *Azotobacter vinelandii* Δ*nifHDK* strain. *J. Biol. Chem.* **281**: 36701-36709.

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P. 1998. An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. *Transgenic Res.* **7**: 289-294.

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P. 1998. Transgenic Central American, West African and Asian Elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. *Ann. Bot.* **82**: 795-801.

Wiig JA, Hu Y, Chung Lee C, Ribbe MW. 2012. Radical SAM-dependent carbon insertion into the nitrogenase M-cluster. *Science* **337**: 1672-1675.

Wilcoxen J, Arragain S, Scandurra AA, Jimenez-Vicente E, Echavarri-Erasun C, Pollmann S, Britt RD, Rubio LM. 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.

Yang X, Chen L, He J, Yu W. 2017. Knocking out of carotenoid catabolic genes in rice fails to boost carotenoid accumulation, but reveals a mutation in strigolactone biosynthesis. *Plant Cell Rep.* **36**: 1533-1545.

Zhao D, Curatti L, Rubio LM. 2007. Evidence for *nifU* and *nifS* participation in the biosynthesis of the iron-molybdenum cofactor of nitrogenase. *J. Biol. Chem.* **282**: 37016-37025.

Zhao Y, Bian S, Zhang C, Zhou H, Wang H, Zhao J, Huang J. 2005. Characterization of a FeMo cofactor-deficient MoFe protein from anifE-deleted strain (DJ35) ofAzotobacter vinelandii. *Chinese Sci. Bull. n* **50**: 2305-2310.

Zheng L, Cash VL, Flint DH, Dean DR. 1998. Assembly of iron-sulfur clusters. Identification of an *iscSUA-hscBA-fdx* gene cluster from *Azotobacter vinelandii*. *J. Biol. Chem.* **273**: 13264-13272.

Chapter III.
Purification and *in vitro* activity of an engineered nitrogenase Fe protein from transgenic rice

3.0 Abstract

Incorporation of biological nitrogen fixation into cereal crops is a long-term goal of plant biotechnology that will allow the direct use of atmospheric nitrogen, rather than synthetic or natural fertilizers. However, this is hindered by the complexity of the machinery needed to synthesize nitrogenase and its metal clusters. Many of the protein components are O_2 sensitive and only sparingly soluble. To address these challenges, I expressed Fe protein (NifH) from the *Azotobacter vinelandii* and *Hydrogenobacter thermophilus* in rice mitochondria. Both recombinant proteins were formed as soluble proteins. Purified *H. thermophilus* NifH was able to carry out the fundamental roles of the Fe protein required to engineer nitrogen fixation, including electron transfer to the nitrogenase component iron-molybdenum （MoFe） protein and the biosynthesis of the nitrogenase iron-molybdenum cofactor (FeMo-co). The stable expression of a biologically active major nitrogenase component in a staple crop represents the critical first step toward the expression of a functional nitrogen fixation (Nif) complex, necessary to achieve biological nitrogen fixation *in planta*.

3.1 Introduction

3.1.1 The function and importance of Fe protein

Biological nitrogen fixation (BNF) is the reduction of N_2 gas to ammonia by the enzyme nitrogenase (Bulen and LeComte 1966). BNF is widespread in prokaryotes (bacteria and archaea), but no eukaryotic species is yet known to directly convert N_2 into a biologically useful form (Mus et al. 2018). The direct transfer of nitrogenase genes from prokaryotes to staple crops is one of the most ambitious strategies to achieve BNF in plants (Curatti and Rubio 2014; Burén et al. 2018). However, developing this technology is hindered by the genetic and biochemical complexity of nitrogenase biosynthesis. Nitrogenase is composed of two component proteins and contains three distinct metal clusters. The catalytic enzyme component MoFe protein loaded with P-cluster ([8Fe7S]) and FeMo-co [7Fe-9S-C-Mo-*R* homocitrate] has been described in Chapter II. The reductase component Fe protein contains the [4Fe4S] clusters and nucleotide binding sites (Georgiadis et al. 1992). The Fe protein is encoded by *NifH*, an essential metalloenzyme for the catalysis of nitrogenase. Fe protein is a multi-purpose component of nitrogenase, participating in catalysis and metal cluster assembly (Jasniewski et al. 2018).

NifH is the unique electron provider for MoFe protein to fix N_2 (Burgess and Lowe 1996). The process of electron transfer referred to as the ¨Fe protein cycle¨, is initiated by the binding of two MgATP to the Fe protein (Burgess and Lowe 1996). The reduced Fe protein transiently associates with MoFe protein to form a complex. Once Fe protein is bound to the MoFe protein, the hydrolysis of MgATP is activated. The two MgATP are hydrolyzed to two MgADP and a single electron is transferred from the Fe protein [4Fe-4S] cluster into the MoFe protein P-cluster and then into the FeMo-co (Danyal et al. 2010). After MgATP hydrolyzation, the oxidized Fe protein dissociates from the MoFe protein. The released oxidized Fe protein re-reduces and the two MgADP replaced by MgATP, to complete one Fe protein cycle (Yang et al. 2016).

In addition to the function in catalysis, the Fe protein is also essential for the assembly of both the FeMo-co and P-cluster. *NifH* deletion strains of *A. vinelandii* produced inactive MoFe protein that not only lacked the FeMo-co but also formed irregular conformation P-clusters (Ribbe et al. 2002). NifH is the entry point of molybdenum into the FeMo-co biosynthetic pathway as it supplies Mo to FeMo-co synthesis *in vitro* (Rangaraj and Ludden (2002). NIfH was confirmed as a homocitrate insertase in FeMo-co maturation. (Hu et al. 2006). Fe protein delivers Mo and homocitrate simultaneously to the Mo-free precursor. In this process, Mo is directly attached to Fe protein and homocitrate is indirectly attached through Mo in an arrangement of Fe protein–Mo–homocitrate (Hu et al. 2006; Fay et al. 2010). The MoFe proteins isolated from Δ*NifB* and Δ*NifH A. vinelandii* showed near identical metal content, with 60% of the iron content compared to the wild-type and no detectable Mo (Ribbe et al. 2002). However, the MoFe protein isolated from Δ*NifB A. vinelandii* can be directly activated by adding isolated FeMo-co because it contains two intact P-clusters, while the Δ*NifH* MoFe protein also required incubation with Fe protein extracts and MgATP for activation (Tal et al. 1991). This strict ATP- and NifH-dependence for reconstitution points out that NifH plays an important role in P-cluster assembly. However, the exact role of NifH in P-cluster maturation has not been elucidated yet. The prevailing hypothesis is that NifH regulates the redox potential of the two [4Fe-4S] and places them in the correct orientation for efficient coupling (Ribbe et al. 2014).

3.1.2 NifH protein and its maturation

The *A. vinelandii* NifH protein is an α_2 dimer , and the binding site for MgATP is located at the dimer interface (Georgiadis et al. 1992). The [4Fe-4S] is located at the interface of

the two subunits of the dimer and is coordinated by two cysteine residues on each subunit (residues α-Cys97, α-Cys132, β-Cys97, and β-Cys132) (Rubio and Ludden 2005).

Fe protein maturation is divided into two processes: the correct folding and formation of NifH homodimers, and the insertion of [4Fe-4S] clusters (Burén et al. 2020). The importance of NifM has been demonstrated in *K. pneumoniae* and *A. vinelandii NifM* mutants (Roberts et al. 1978; Jacobson et al. 1989). NifM deletion resulted in negligible Fe protein activity even though Fe protein polypeptide accumulated in substantial amounts (Roberts et al. 1978; Jacobson et al. 1989). NifM is a putative peptidyl-prolyl *cis-trans* isomerase that improves protein folding by catalyzing the *cis-trans* isomerization of proline imidic peptide bonds (Gavini et al. 2006). The requirement of NifM to produce active Fe protein has also been demonstrated by heterologous expression in yeast. Fe protein purified from yeast co-expressing *A. vinelandii* NifH and NifM was active, but no activity was detectable when the same NifH was purified from yeast cells lacking NifM (López-Torrejón et al. 2016). The initial cluster of the [4Fe-4S] cluster for NifH maturation is synthesized by NifU and NifS, where NifU provides the scaffold for the assembly of [Fe-S] clusters and NifS mobilizes sulfur from cysteine for [Fe-S] cluster synthesis on NifU (Jacobson et al. 1989; Smith et al. 2005). Subsequently, the [4Fe-4S] cluster is transferred to cluster-less apo-Fe protein without additional accessory proteins (Burén et al. 2020).

3.1.3 Heterologous expression of NifH in yeast and plants

Given that nitrogenase is sensitive to oxygen and the requirement for high energy input, reducing power and metal cofactors, mitochondria and chloroplasts have been tested as suitable sub-cellular cites for the expression of nitrogenase components (Ivleva et al. 2016; López-Torrejón et al. 2016; Burén et al. 2017; Burén et al. 2019; Eseverri et al. 2020; Jiang et al. 2021). This is because both organelles are energy-conversion organelles in eukaryotic cells and they are thus able to provide reducing power and ATP for nitrogenase synthesis. The low-O_2 environment of mitochondria also protects nitrogenase from O_2 damage (Yang et al., 2017).

The first heterologous expression of active Fe was in aerobic-grown yeast, requiring only co-expression of *A. vinelandii* NifH and NifM in the mitochondrial matrix (López-Torrejón et al. 2016). In contrast, active Fe protein in the cytosol requires both anoxic growth conditions and co-expression of NifH and NifM with NifU and NifS from *A. vinelandii* (López-Torrejón et al. 2016). Under low oxygen concentration (10% oxygen),

Klebsiella pneumoniae NifH produced in tobacco plastids showed low Fe protein reduction activity after [Fe-S] cluster reconstitution (Ivleva et al. 2016). In transiently expressed tobacco, either co-expression *A. vinelandii* NifH with NifM, NifU, and NifS within the mesophyll cell matrix or co-expression *H. thermophilus* NifH with *A. vinelandii* NifM, NifU, and NifS in mitochondria, NifH proteins were only active after reconstitution (Eseverri et al. 2020). Direct Fe protein activity (as isolated, without [Fe-S] cluster reconstitution) was only reported from yeast co-expressing transiently NifH and NifM either alone or with NifS and NifU (López-Torrejón et al. 2016; Jiang et al. 2021).

Here, I engineered rice plants to express *NifH* from *A. vinelandii* and *H. thermophilus*, targeting the corresponding proteins to the mitochondrial matrix to minimize O_2 damage. Both recombinant rice-produced NifH proteins formed as soluble proteins and purified from rice tissue by strep-tag affinity chromatography (STAC). The as-isolated *H. thermophilus* NifH exhibited limited functionality in electron donation to the MoFe protein and in FeMo-co synthesis, two fundamental activities required to engineer nitrogenase. *In vitro* transfer of [4Fe-4S] clusters from NifU (donor) to NifH was required to achieve maximum Fe protein activity. This indicates that even though NifH incorporated some endogenous rice mitochondrial [4Fe-4S] clusters leading to activity, much of it accumulated as apo-protein, thereby identifying [Fe-S] cluster biosynthesis, insertion, and stability as areas that should be the focus of future research.

3.2 Aims and objectives

The aims of this chapter were to express nitrogenase Fe protein in transgenic rice and analyze the function of recombinant rice Fe protein *in vitro*. The specific objectives were to: 1) introduce the *A. vinelandii NifH* combined with *A. vinelandii NifM, NifS and NifU*, and *H. thermophilus NifH* combined with *A. vinelandii NifM* into rice, respectively; 2) analyze mRNA expression of *A. vinelandii NifH*, *H. thermophilus NifH*, *A. vinelandii NifM*, *NifS* and *NifU* in transgenic rice callus and the corresponding regenerated plants; 3) identify the corresponding protein accumulation in transgenic rice callus and plants; 4) purify *A. vinelandii* and *H. thermophilus* NifH from rice callus and plants; 5) determine the enzymatic activity of rice-produced *A. vinelandii* and *H. thermophilus* NifH; 6) determine the function of rice-produced *H. thermophilus* NifH in FeMo-co synthesis.

3.3 Materials and methods

3.3.1 Expression constructs for rice transformation

The gene sequences of *A. vinelandii NifH* (*NifH*^{Av}), *H. thermophilus nifH* (*NifH*^{Ht}), *A. vinelandii NifM* (*NifM*^{Av}), *A. vinelandii NifS* (*NifS*^{Av}) and *A. vinelandii NifU* (*NifU*^{Av}), the mitochondrial targeting peptide sequences of *S. cerevisiae* cytochrome c oxidase subunit IV (Cox4) and *Neurospora crassa* mitochondrial ATPase subunit 9 (Su9) and the sequence of Twinstrep (TS) tag were codon optimized for *S. cerevisiae* using the GeneOptimizer tool (Thermo Fisher Scientific) and synthesized by Thermo Fisher Scientific as part of the Engineering Nitrogen Symbiosis for Africa project (https://www.ensa.ac.uk/). The empty vector pUC57 (GenScript Biotech, Piscataway, NJ, USA) was digested with Acc65I and SalI, allowing the insertion of the *ZmUbi1*+1sti promoter. The synthetic Cox4-TS-*NifH*^{Av}-t*Nos* sequence was introduced to the intermediate vector pUC57- Acc65I-*ZmUbi1*+1sti-SalI at SalI and SphI sites to generate pAvNifH. The empty vector pUC57 (GenScript) was digested with BamHI and PstI to insert *ZmUbi1*+1sti promoter. The synthetic Cox4-TS-*NifH*^{Ht}-t*Nos* sequence was introduced into the intermediate vector pUC57-BamHI-*ZmUbi1*+1sti-PstI at the PstI and SphI sites to produce pHtNifH. The empty vector pUC57 (GenScript) was digested with Acc65I and XbaI to insert the *OsActin* promoter. The pUC57-*OsActin* plasmid was then digested with XbaI and SalI to insert the synthetic Su9-*NifM*^{Av}- t*Nos* sequence to produce pAvNifM. The empty vector pUC57 (GenScript) was digested with Acc65I and BamHI to introduce *ZmUbi1*+1sti promoter to control the expression of *NifS*^{Av}. The synthetic Su9-*nifS*^{Av}-t*Adh1* sequence was introduced to the intermediate vector pUC57-Acc65I-*ZmUbi1*+1sti-BamHI at the BamHI and HindIII sites to produce pAvNifS. The empty vector pUC57 (GenScript) was digested with EcoRI and SacI to insert *ZmUbi1*+1sti promoter for driving the expression of *NifU*^{Av}. The synthetic Su9-*nifU*^{Av}-t*Cyc1* was inserted into the intermediate vector pUC57-EcoRI-*ZmUbi1*+1sti-SacI at the SacI and BamHI sites to produce pAvNifU. All restriction enzymes and T4 DNA ligase were obtained from Promega (Madison, WI, USA) or New England Biolabs (Ipswich, MA, USA). The plasmids were amplified in *E. coli* DH5α cells grown at 37°C in LB medium supplemented with 100 μg/mL ampicillin. The final constructs were verified by Sanger sequencing (Stabvida, Caparica, Portugal). The genetic elements and cloning primers for the rice expression constructs are listed in **Tables 3.1 and 3.2**.

Table 3.1 Vectors and constructs used (a) Plant expression vectors. (b) Sequences of genetic constructs.

(a)

Plasmids	Expressed proteins	Promoter / terminator	Size (kDa, Processed with MPP)
PUC57	Cox4-TS-NifH^{Av}	p*ZmUbi1*+1st i / t*Nos*	35
PUC57	Cox4-TS-NifH^{Ht}	p*ZmUbi1*+1st i / t*Nos*	33
PUC57	Su9-NifM^{Av}	p*OsActin*/ t*Nos*	33
PUC57	Su9- NifS^{Av}	p*ZmUbi1*+1sti / t*Adh1*	43
PUC57	Su9- NifU^{Av}	p*ZmUbi1*+1sti / t*Cyc1*	33

MPP, mitochondrial processing peptidase; Cox4, mitochondrial targeting peptide of *S. cerevisiae* cytochrome c oxidase subunit IV, Su9, mitochondrial targeting peptide of *Neurospora crassa* mitochondrial ATPase subunit 9, TS, Twinstrep tag; *NifH^{Av}, A. vinelandii NifH; NifH^{Ht}, H. thermophilus NifH*; *NifM^{Av}, A. vinelandii NifM*; *NifS^{Av}, A. vinelandii NifS;* and *NifU^{Av}, A.* vinelandii *NifU.*

(b)

Genetic elements	Nucleotide sequence Amino acid sequence										
Cox4											
	ATG	CTT	TCA	CTT	AGA	CAA	TCT	ATT	AGA	TTT	TTC
	M	**L**	**S**	**L**	**R**	**Q**	**S**	**I**	**R**	**F**	**F**
	AAG	CCA	GCT	ACA	AGA	ACT	TTG	TGT	TCT	TCT	AGA
	K	**P**	**A**	**T**	**R**	**T**	**L**	**C**	**S**	**S**	**R**
	TAT	CTT	CTT	CAG	CAA	AAA	CCT				
	Y	**L**	**L**	**Q**	**Q**	**K**	**P**				
Su9											
	ATG	CTT	TCA	CTT	AGA	CAA	TCT	ATT	AGA	TTT	TTC
	M	**L**	**S**	**L**	**R**	**Q**	**S**	**I**	**R**	**F**	**F**
	AAG	CCA	GCT	ACA	AGA	ACT	TTG	TGT	TCT	TCT	AGA
	K	**P**	**A**	**T**	**R**	**T**	**L**	**C**	**S**	**S**	**R**
	TAT	CTT	CTT	CAG	CAA	AAA	CCT				
	Y	**L**	**L**	**Q**	**Q**	**K**	**P**				
TS											
	TGG	AGT	CAT	CCT	CAG	TTT	GAG	AAA	GGT	GGA	GGT
	W	**S**	**H**	**P**	**Q**	**F**	**E**	**K**	**G**	**G**	**G**
	TCA	GGT	GGT	GGA	AGC	GGT	GGA	TCT	GCT	TGG	TCA
	S	**G**	**G**	**G**	**S**	**G**	**G**	**S**	**A**	**W**	**S**
	CAT	CCA	CAA	TTT	GAA	AAA					
	H	**P**	**Q**	**F**	**E**	**K**					
NifH^{Av}											
	ATG	GCT	ATG	AGA	CAA	TGT	GCC	ATC	TAT	GGT	AAA
	M	**A**	**M**	**R**	**Q**	**C**	**A**	**I**	**Y**	**G**	**K**
	GGT	GGT	ATT	GGT	AAA	TCT	ACC	ACC	ACT	CAA	AAT
	G	**G**	**I**	**G**	**K**	**S**	**T**	**T**	**T**	**Q**	**N**
	TTG	GTT	GCT	GCT	TTG	GCT	GAA	ATG	GGT	AAA	AAG

L	**V**	**A**	**A**	**L**	**A**	**E**	**M**	**G**	**K**	**K**
GTT	ATG	ATC	GTT	GGT	TGT	GAT	CCA	AAG	GCT	GAT
V	**M**	**I**	**V**	**G**	**C**	**D**	**P**	**K**	**A**	**D**
TCT	ACT	AGA	TTG	ATC	TTG	CAT	TCC	AAG	GCT	CAA
S	**T**	**R**	**L**	**I**	**L**	**H**	**S**	**K**	**A**	**Q**
AAC	ACC	ATT	ATG	GAA	ATG	GCT	GCT	GAA	GCT	GGT
N	**T**	**I**	**M**	**E**	**M**	**A**	**A**	**E**	**A**	**G**
ACT	GTT	GAA	GAT	TTG	GAA	TTG	GAA	GAT	GTT	TTG
T	**V**	**E**	**D**	**L**	**E**	**L**	**E**	**D**	**V**	**L**
AAG	GCT	GGT	TAT	GGT	GGT	GTT	AAG	TGT	GTT	GAA
K	**A**	**G**	**Y**	**G**	**G**	**V**	**K**	**C**	**V**	**E**
TCT	GGT	GGT	CCA	GAA	CCA	GGT	GTT	GGT	TGC	GCT
S	**G**	**G**	**P**	**E**	**P**	**G**	**V**	**G**	**C**	**A**
GGT	AGA	GGT	GTT	ATT	ACT	GCT	ATT	AAC	TTC	TTG
G	**R**	**G**	**V**	**I**	**T**	**A**	**I**	**N**	**F**	**L**
GAA	GAA	GAG	GGC	GCT	TAC	GAA	GAT	GAT	TTG	GAT
E	**E**	**E**	**G**	**A**	**Y**	**E**	**D**	**D**	**L**	**D**
TTT	GTT	TTC	TAC	GAC	GTT	TTG	GGT	GAT	GTT	GTT
F	**V**	**F**	**Y**	**D**	**V**	**L**	**G**	**D**	**V**	**V**
TGT	GGT	GGT	TTT	GCT	ATG	CCA	ATC	AGA	GAA	AAC
C	**G**	**G**	**F**	**A**	**M**	**P**	**I**	**R**	**E**	**N**
AAG	GCA	CAA	GAA	ATC	TAC	ATC	GTT	TGC	TCT	GGT
K	**A**	**Q**	**E**	**I**	**Y**	**I**	**V**	**C**	**S**	**G**
GAA	ATG	ATG	GCT	ATG	TAT	GCT	GCT	AAC	AAC	ATC
E	**M**	**M**	**A**	**M**	**Y**	**A**	**A**	**N**	**N**	**I**
TCT	AAG	GGT	ATC	GTT	AAG	TAC	GCT	AAC	TCT	GGT
S	**K**	**G**	**I**	**V**	**K**	**Y**	**A**	**N**	**S**	**G**
TCT	GTT	AGA	TTA	GGT	GGT	TTG	ATC	TGC	AAC	TCT
S	**V**	**R**	**L**	**G**	**G**	**L**	**I**	**C**	**N**	**S**
AGA	AAC	ACC	GAT	AGA	GAA	GAT	GAA	CTG	ATT	ATT
R	**N**	**T**	**D**	**R**	**E**	**D**	**E**	**L**	**I**	**I**
GCT	TTG	GCC	AAC	AAG	TTG	GGT	ACT	CAA	ATG	ATT
A	**L**	**A**	**N**	**K**	**L**	**G**	**T**	**Q**	**M**	**I**
CAT	TTC	GTC	CCA	AGA	GAT	AAC	GTT	GTT	CAA	AGA
H	**F**	**V**	**P**	**R**	**D**	**N**	**V**	**V**	**Q**	**R**
GCC	GAA	ATT	AGA	AGA	ATG	ACC	GTT	ATC	GAA	TAC
A	**E**	**I**	**R**	**R**	**M**	**T**	**V**	**I**	**E**	**Y**
GAC	CCA	AAA	GCT	AAA	CAA	GCT	GAT	GAG	TAT	AGA
D	**P**	**K**	**A**	**K**	**Q**	**A**	**D**	**E**	**Y**	**R**
GCC	TTG	GCT	AGA	AAA	GTT	GTC	GAT	AAC	AAG	TTG
A	**L**	**A**	**R**	**K**	**V**	**V**	**D**	**N**	**K**	**L**
CTG	GTT	ATC	CCA	AAT	CCA	ATC	ACC	ATG	GAT	GAA
L	**V**	**I**	**P**	**N**	**P**	**I**	**T**	**M**	**D**	**E**
CTA	GAA	GAG	TTG	TTG	ATG	GAA	TTC	GGT	ATC	ATG
L	**E**	**E**	**L**	**L**	**M**	**E**	**F**	**G**	**I**	**M**
GAA	GTC	GAA	GAT	GAA	TCT	ATC	GTT	GGT	AAG	ACT
E	**V**	**E**	**D**	**E**	**S**	**I**	**V**	**G**	**K**	**T**
GCT	GAA	GAA	GTC	TGA						
A	**E**	**E**	**V**	**-**						

NifH^{Ht}

ATG	AGA	CAA	ATT	GCC	ATC	TAC	GGT	AAA	GGT	GGT
M	**R**	**Q**	**I**	**A**	**I**	**Y**	**G**	**K**	**G**	**G**
ATT	GGT	AAA	TCT	ACC	ACT	ACT	CAA	AAC	ACT	GTT
I	**G**	**K**	**S**	**T**	**T**	**T**	**Q**	**N**	**T**	**V**
GCT	GCT	TTG	GCT	GAA	GCT	GGT	AGA	AAA	TGT	TTT
A	**A**	**L**	**A**	**E**	**A**	**G**	**R**	**K**	**C**	**F**
ATC	GTT	GGT	TGT	GAT	CCA	AAG	GCC	GAT	TCT	ACC
I	**V**	**G**	**C**	**D**	**P**	**K**	**A**	**D**	**S**	**T**
AGA	TTG	ATA	TTG	CAT	GTT	AAG	GCC	CAA	TCT	ACC
R	**L**	**I**	**L**	**H**	**V**	**K**	**A**	**Q**	**S**	**T**
GTT	ATG	CAT	TTG	GCT	GCT	GAA	AGA	GGT	GCT	GTT
V	**M**	**H**	**L**	**A**	**A**	**E**	**R**	**G**	**A**	**V**
GAA	GAT	TTG	GAT	TTG	GAT	GAA	GTT	ATG	TTG	GTT
E	**D**	**L**	**D**	**L**	**D**	**E**	**V**	**M**	**L**	**V**
GGT	TTC	GGT	GGT	ATT	AAG	TGT	GTT	GAA	TCT	GGT
G	**F**	**G**	**G**	**I**	**K**	**C**	**V**	**E**	**S**	**G**
GGT	CCA	GAA	CCA	GGT	GTT	GGT	TGC	GCT	GGT	AGA
G	**P**	**E**	**P**	**G**	**V**	**G**	**C**	**A**	**G**	**R**
GGT	GTT	ATT	ACT	GCT	ATT	AAC	TTC	TTG	GAA	GAA
G	**V**	**I**	**T**	**A**	**I**	**N**	**F**	**L**	**E**	**E**
AAC	GGT	GCC	TTC	GAT	GAT	GAT	TTG	GAC	TAC	GTT
N	**G**	**A**	**F**	**D**	**D**	**D**	**L**	**D**	**Y**	**V**
TTT	TAC	GAT	GTT	TTG	GGT	GAT	GTT	GTT	TGT	GGT
F	**Y**	**D**	**V**	**L**	**G**	**D**	**V**	**V**	**C**	**G**
GGT	TTT	GCT	ATG	CCA	ATT	AGA	GAA	GGT	AAA	GCC
G	**F**	**A**	**M**	**P**	**I**	**R**	**E**	**G**	**K**	**A**
CAA	GAA	ATC	TAC	ATC	GTT	ACA	TCT	GGT	GAA	ATG
Q	**E**	**I**	**Y**	**I**	**V**	**T**	**S**	**G**	**E**	**M**
ATG	GCT	ATG	TAC	GCT	GCT	AAC	AAT	ATC	TCC	AAG
M	**A**	**M**	**Y**	**A**	**A**	**N**	**N**	**I**	**S**	**K**
GGT	ATT	TTG	AAG	TAC	GCT	CAT	TCT	GGT	GGT	GTT
G	**I**	**L**	**K**	**Y**	**A**	**H**	**S**	**G**	**G**	**V**
AGA	TTA	GGT	GGT	TTG	ATC	TGT	AAC	TCC	AGA	AAC
R	**L**	**G**	**G**	**L**	**I**	**C**	**N**	**S**	**R**	**N**
GTT	GAT	AAC	GAG	AGG	GAA	TTG	ATT	GAA	GCT	TTG
V	**D**	**N**	**E**	**R**	**E**	**L**	**I**	**E**	**A**	**L**
GCA	GAA	AAA	TTG	GGC	ACC	CAA	ATG	ATT	CAT	TTC
A	**E**	**K**	**L**	**G**	**T**	**Q**	**M**	**I**	**H**	**F**
TTG	CCT	AGA	AAC	AAC	ATT	GTC	CAA	GAA	GCC	GAA
L	**P**	**R**	**N**	**N**	**I**	**V**	**Q**	**E**	**A**	**E**
TTG	AGA	AGA	ATG	ACC	GTT	ATT	GAA	TAT	GCC	CCA
L	**R**	**R**	**M**	**T**	**V**	**I**	**E**	**Y**	**A**	**P**
GAT	CAT	CCA	ATG	GCT	GAT	GAG	TAT	AGA	ACT	TTG
D	**H**	**P**	**M**	**A**	**D**	**E**	**Y**	**R**	**T**	**L**
GCC	AAG	AAG	ATC	GAA	GAG	AAC	AGA	AAG	TTG	TCT
A	**K**	**K**	**I**	**E**	**E**	**N**	**R**	**K**	**L**	**S**
ATC	CCA	ACT	CCA	TTG	ACC	ATG	GAT	GAA	TTG	GAA
I	**P**	**T**	**P**	**L**	**T**	**M**	**D**	**E**	**L**	**E**
CAA	CTG	TTG	GTT	GAA	TAC	GGT	ATC	ATG	AAG	CCA

	Q	**L**	**L**	**V**	**E**	**Y**	**G**	**I**	**M**	**K**	**P**
	GAA	GAA	GTT	GCT	TGA						
	E	**E**	**V**	**A**	**-**						
NifM^{Av}											
	ATG	GCC	TCA	GAA	AGA	TTA	GCT	GAT	GGT	GAC	TCC
	M	**A**	**S**	**E**	**R**	**L**	**A**	**D**	**G**	**D**	**S**
	AGA	TAT	TAC	TTG	TTA	AAA	GTT	GCC	CAT	GAA	CAA
	R	**Y**	**Y**	**L**	**L**	**K**	**V**	**A**	**H**	**E**	**Q**
	TTT	GGT	TGT	GCT	CCT	GGT	GAA	TTA	TCA	GAA	GAA
	F	**G**	**C**	**A**	**P**	**G**	**E**	**L**	**S**	**E**	**E**
	CAA	TTG	CAA	CAA	GCA	GAT	AGA	ATT	ATA	GGT	AGA
	Q	**L**	**Q**	**Q**	**A**	**D**	**R**	**I**	**I**	**G**	**R**
	CAA	AGA	CAC	ATA	GAA	GAT	GCA	GTT	TTA	AGA	TCA
	Q	**R**	**H**	**I**	**E**	**D**	**A**	**V**	**L**	**R**	**S**
	CCA	GAC	GCC	ATA	GGT	GTT	GTC	ATC	CCA	CCT	TCC
	P	**D**	**A**	**I**	**G**	**V**	**V**	**I**	**P**	**P**	**S**
	CAA	TTG	GAA	GAA	GCT	TGG	GCA	CAT	ATT	GCT	TCA
	Q	**L**	**E**	**E**	**A**	**W**	**A**	**H**	**I**	**A**	**S**
	AGA	TAT	GAA	TCC	CCT	GAA	GCC	TTG	CAA	CAA	GCT
	R	**Y**	**E**	**S**	**P**	**E**	**A**	**L**	**Q**	**Q**	**A**
	TTA	GAT	GCC	CAA	GCT	TTG	GAC	GCT	GCT	GGT	ATG
	L	**D**	**A**	**Q**	**A**	**L**	**D**	**A**	**A**	**G**	**M**
	AGA	GCA	ATG	TTG	GCC	AGA	GAA	TTA	AGA	GTT	GAA
	R	**A**	**M**	**L**	**A**	**R**	**E**	**L**	**R**	**V**	**E**
	GCT	GTC	TTA	GAT	TGT	GTC	TGC	GCA	GGT	TTG	CCA
	A	**V**	**L**	**D**	**C**	**V**	**C**	**A**	**G**	**L**	**P**
	GAA	ATT	AGT	GAT	ACA	GAC	GTA	TCT	TTG	TAC	TAC
	E	**I**	**S**	**D**	**T**	**D**	**V**	**S**	**L**	**Y**	**Y**
	TTC	AAC	CAT	GCT	GAA	CAA	TTC	AAG	GTA	CCA	GCA
	F	**N**	**H**	**A**	**E**	**Q**	**F**	**K**	**V**	**P**	**A**
	CAA	CAT	AAA	GCC	AGA	CAC	ATA	TTG	GTT	ACT	ATA
	Q	**H**	**K**	**A**	**R**	**H**	**I**	**L**	**V**	**T**	**I**
	AAT	GAA	GAT	TTT	CCT	GAA	AAC	ACA	AGA	GAA	GCC
	N	**E**	**D**	**F**	**P**	**E**	**N**	**T**	**R**	**E**	**A**
	GCT	AGA	ACC	AGA	ATC	GAA	ACT	ATC	TTG	AAG	AGA
	A	**R**	**T**	**R**	**I**	**E**	**T**	**I**	**L**	**K**	**R**
	TTG	AGA	GGT	AAA	CCA	GAA	AGA	TTC	GCT	GAA	CAA
	L	**R**	**G**	**K**	**P**	**E**	**R**	**F**	**A**	**E**	**Q**
	GCA	ATG	AAA	CAC	TCT	GAA	TGT	CCT	ACA	GCT	ATG
	A	**M**	**K**	**H**	**S**	**E**	**C**	**P**	**T**	**A**	**M**
	CAA	GGT	GGT	TTG	TTA	GGT	GAA	GTA	GTT	CCA	GGT
	Q	**G**	**G**	**L**	**L**	**G**	**E**	**V**	**V**	**P**	**G**
	ACC	TTG	TAT	CCT	GAA	TTA	GAT	GCA	TGC	TTG	TTT
	T	**L**	**Y**	**P**	**E**	**L**	**D**	**A**	**C**	**L**	**F**
	CAA	ATG	GCC	AGA	GGT	GAA	TTA	TCA	CCA	GTT	TTG
	Q	M	A	R	G	E	L	S	P	V	L
	GAA	**TCC**	**CCT**	**ATT**	**GGT**	**TTC**	**CAT**	**GTT**	**TTA**	**TAC**	**TGT**
	E	S	P	I	G	F	H	V	L	Y	C
	GAA	**TCC**	**GTC**	**AGT**	**CCA**	**GCA**	**AGA**	**CAA**	**TTG**	**ACC**	**TTG**
	E	S	V	S	P	A	R	Q	L	T	L

	GAA	**GAA**	**ATC**	**TTG**	**CCT**	**AGA**	**TTG**	**AGA**	**GAT**	**AGA**	**TTG**
	E	E	I	L	P	R	L	R	D	R	L
	CAA	**TTG**	**AGA**	**CAA**	**AGA**	**AAG**	**GCT**	**TAC**	**CAA**	**AGA**	**AAG**
	Q	L	R	Q	R	K	A	Y	Q	R	K
	TGG	**TTG**	**GAA**	**TCT**	**TTG**	**TTG**	**CAA**	**CAA**	**AAT**	**GCT**	**ACC**
	W	L	E	S	L	L	Q	Q	N	A	T
	TTG	**GAA**	**AAC**	**TTA**	**GCA**	**CAT**	**GGT**	**TAA**			
	L	E	N	L	A	H	G	-			
NifS^{Av}											
	TGG	GAC	TAC	TCT	GAA	AAG	GTT	AAG	GAA	CAT	TTC
	W	**D**	**Y**	**S**	**E**	**K**	**V**	**K**	**E**	**H**	**F**
	TAC	AAT	CCA	AAG	AAC	GCC	GGT	GCT	GTA	GAA	GGT
	Y	**N**	**P**	**K**	**N**	**A**	**G**	**A**	**V**	**E**	**G**
	GCA	AAC	GCC	ATT	GGT	GAC	GTT	GGT	TCA	TTA	TCC
	A	**N**	**A**	**I**	**G**	**D**	**V**	**G**	**S**	**L**	**S**
	TGT	GGT	GAC	GCT	TTG	AGA	TTA	ACA	TTG	AAA	GTT
	C	**G**	**D**	**A**	**L**	**R**	**L**	**T**	**L**	**K**	**V**
	GAC	CCT	GAA	ACC	GAT	GTC	ATC	TTG	GAC	GCA	GGT
	D	**P**	**E**	**T**	**D**	**V**	**I**	**L**	**D**	**A**	**G**
	TTT	CAA	ACT	TTC	GGT	TGC	GGT	TCT	GCT	ATT	GCA
	F	**Q**	**T**	**F**	**G**	**C**	**G**	**S**	**A**	**I**	**A**
	TCT	TCA	TCC	GCT	TTG	ACT	GAA	ATG	GTT	AAG	GGT
	S	**S**	**S**	**A**	**L**	**T**	**E**	**M**	**V**	**K**	**G**
	TTG	ACA	TTG	GAT	GAA	GCA	TTG	AAA	ATC	TCA	AAC
	L	**T**	**L**	**D**	**E**	**A**	**L**	**K**	**I**	**S**	**N**
	CAA	GAT	ATC	GCT	GAC	TAT	TTG	GAT	GGT	TTG	CCA
	Q	**D**	**I**	**A**	**D**	**Y**	**L**	**D**	**G**	**L**	**P**
	CCT	GAA	AAG	ATG	CAT	TGT	TCC	GTC	ATG	GGT	AGA
	P	**E**	**K**	**M**	**H**	**C**	**S**	**V**	**M**	**G**	**R**
	GAA	GCC	TTA	CAA	GCT	GCA	GTA	GCT	AAC	TAC	AGA
	E	**A**	**L**	**Q**	**A**	**A**	**V**	**A**	**N**	**Y**	**R**
	GGT	GAA	ACC	ATT	GAA	GAT	GAC	CAC	GAA	GAA	GGT
	G	**E**	**T**	**I**	**E**	**D**	**D**	**H**	**E**	**E**	**G**
	GCA	TTG	ATA	TGT	AAA	TGC	TTT	GCC	GTT	GAT	GAA
	A	**L**	**I**	**C**	**K**	**C**	**F**	**A**	**V**	**D**	**E**
	GTT	ATG	GTC	AGA	GAT	ACC	ATA	AGA	GCA	AAT	AAG
	V	**M**	**V**	**R**	**D**	**T**	**I**	**R**	**A**	**N**	**K**
	TTA	AGT	ACT	GTA	GAA	GAT	GTT	ACT	AAC	TAC	ACA
	L	**S**	**T**	**V**	**E**	**D**	**V**	**T**	**N**	**Y**	**T**
	AAA	GCT	GGT	GGT	GGT	TGT	TCT	GCT	TGC	CAT	GAA
	K	**A**	**G**	**G**	**G**	**C**	**S**	**A**	**C**	**H**	**E**
	GCA	ATA	GAA	AGA	GTT	TTG	ACA	GAA	GAA	TTG	GCC
	A	**I**	**E**	**R**	**V**	**L**	**T**	**E**	**E**	**L**	**A**
	GCT	AGA	GGT	GAA	GTA	TTC	GTT	GCA	GCC	CCA	ATT
	A	**R**	**G**	**E**	**V**	**F**	**V**	**A**	**A**	**P**	**I**
	AAA	GCC	AAA	AAG	AAA	GTC	AAG	GTA	TTG	GCT	CCA
	K	**A**	**K**	**K**	**K**	**V**	**K**	**V**	**L**	**A**	**P**
	GAA	CCT	GCC	CCA	GCT	CCT	GTT	GCA	GAA	GCC	CCA
	E	**P**	**A**	**P**	**A**	**P**	**V**	**A**	**E**	**A**	**P**
	GCT	GCA	GCC	CCT	AAG	TTG	TCA	AAT	TTG	CAA	AGA

	A	**A**	**A**	**P**	**K**	**L**	**S**	**N**	**L**	**Q**	**R**
	ATT	AGA	AGA	ATC	GAA	ACA	GTC	TTG	GCT	GCA	ATA
	I	**R**	**R**	**I**	**E**	**T**	**V**	**L**	**A**	**A**	**I**
	AGA	CCT	ACC	TTG	CAA	AGA	GAC	AAA	GGT	GAC	GTC
	R	**P**	**T**	**L**	**Q**	**R**	**D**	**K**	**G**	**D**	**V**
	GAA	TTA	ATT	GAT	GTA	GAC	GGT	AAA	AAT	GTT	TAC
	E	**L**	**I**	**D**	**V**	**D**	**G**	**K**	**N**	**V**	**Y**
	GTC	AAA	TTG	ACC	GGT	GCT	TGT	ACT	GGT	TGC	CAA
	V	**K**	**L**	**T**	**G**	**A**	**C**	**T**	**G**	**C**	**Q**
	ATG	GCA	TCC	ATG	ACA	TTA	GGT	GGT	ATA	CAA	CAA
	M	**A**	**S**	**M**	**T**	**L**	**G**	**G**	**I**	**Q**	**Q**
	AGA	TTG	ATC	GAA	GAA	TTG	GGT	GAG	TTC	GTC	AAA
	R	**L**	**I**	**E**	**E**	**L**	**G**	**E**	**F**	**V**	**K**
	GTT	ATC	CCA	GTC	TCC	GCT	GCC	GCA	CAC	GCC	CAA
	V	**I**	**P**	**V**	**S**	**A**	**A**	**A**	**H**	**A**	**Q**
	ATG	GAA	GTC	TGA							
	M	**E**	**V**	**-**							
NifU^{Av}											
	TGG	GAC	TAC	TCT	GAA	AAG	GTT	AAG	GAA	CAT	TTC
	W	**D**	**Y**	**S**	**E**	**K**	**V**	**K**	**E**	**H**	**F**
	TAC	AAT	CCA	AAG	AAC	GCC	GGT	GCT	GTA	GAA	GGT
	Y	**N**	**P**	**K**	**N**	**A**	**G**	**A**	**V**	**E**	**G**
	GCA	AAC	GCC	ATT	GGT	GAC	GTT	GGT	TCA	TTA	TCC
	A	**N**	**A**	**I**	**G**	**D**	**V**	**G**	**S**	**L**	**S**
	TGT	GGT	GAC	GCT	TTG	AGA	TTA	ACA	TTG	AAA	GTT
	C	**G**	**D**	**A**	**L**	**R**	**L**	**T**	**L**	**K**	**V**
	GAC	CCT	GAA	ACC	GAT	GTC	ATC	TTG	GAC	GCA	GGT
	D	**P**	**E**	**T**	**D**	**V**	**I**	**L**	**D**	**A**	**G**
	TTT	CAA	ACT	TTC	GGT	TGC	GGT	TCT	GCT	ATT	GCA
	F	**Q**	**T**	**F**	**G**	**C**	**G**	**S**	**A**	**I**	**A**
	TCT	TCA	TCC	GCT	TTG	ACT	GAA	ATG	GTT	AAG	GGT
	S	**S**	**S**	**A**	**L**	**T**	**E**	**M**	**V**	**K**	**G**
	TTG	ACA	TTG	GAT	GAA	GCA	TTG	AAA	ATC	TCA	AAC
	L	**T**	**L**	**D**	**E**	**A**	**L**	**K**	**I**	**S**	**N**
	CAA	GAT	ATC	GCT	GAC	TAT	TTG	GAT	GGT	TTG	CCA
	Q	**D**	**I**	**A**	**D**	**Y**	**L**	**D**	**G**	**L**	**P**
	CCT	GAA	AAG	ATG	CAT	TGT	TCC	GTC	ATG	GGT	AGA
	P	**E**	**K**	**M**	**H**	**C**	**S**	**V**	**M**	**G**	**R**
	GAA	GCC	TTA	CAA	GCT	GCA	GTA	GCT	AAC	TAC	AGA
	E	**A**	**L**	**Q**	**A**	**A**	**V**	**A**	**N**	**Y**	**R**
	GGT	GAA	ACC	ATT	GAA	GAT	GAC	CAC	GAA	GAA	GGT
	G	**E**	**T**	**I**	**E**	**D**	**D**	**H**	**E**	**E**	**G**
	GCA	TTG	ATA	TGT	AAA	TGC	TTT	GCC	GTT	GAT	GAA
	A	**L**	**I**	**C**	**K**	**C**	**F**	**A**	**V**	**D**	**E**
	GTT	ATG	GTC	AGA	GAT	ACC	ATA	AGA	GCA	AAT	AAG
	V	**M**	**V**	**R**	**D**	**T**	**I**	**R**	**A**	**N**	**K**
	TTA	AGT	ACT	GTA	GAA	GAT	GTT	ACT	AAC	TAC	ACA
	L	**S**	**T**	**V**	**E**	**D**	**V**	**T**	**N**	**Y**	**T**
	AAA	GCT	GGT	GGT	GGT	TGT	TCT	GCT	TGC	CAT	GAA
	K	**A**	**G**	**G**	**G**	**C**	**S**	**A**	**C**	**H**	**E**

GCA	ATA	GAA	AGA	GTT	TTG	ACA	GAA	GAA	TTG	GCC
A	**I**	**E**	**R**	**V**	**L**	**T**	**E**	**E**	**L**	**A**
GCT	AGA	GGT	GAA	GTA	TTC	GTT	GCA	GCC	CCA	ATT
A	**R**	**G**	**E**	**V**	**F**	**V**	**A**	**A**	**P**	**I**
AAA	GCC	AAA	AAG	AAA	GTC	AAG	GTA	TTG	GCT	CCA
K	**A**	**K**	**K**	**K**	**V**	**K**	**V**	**L**	**A**	**P**
GAA	CCT	GCC	CCA	GCT	CCT	GTT	GCA	GAA	GCC	CCA
E	**P**	**A**	**P**	**A**	**P**	**V**	**A**	**E**	**A**	**P**
GCT	GCA	GCC	CCT	AAG	TTG	TCA	AAT	TTG	CAA	AGA
A	**A**	**A**	**P**	**K**	**L**	**S**	**N**	**L**	**Q**	**R**
ATT	AGA	AGA	ATC	GAA	ACA	GTC	TTG	GCT	GCA	ATA
I	**R**	**R**	**I**	**E**	**T**	**V**	**L**	**A**	**A**	**I**
AGA	CCT	ACC	TTG	CAA	AGA	GAC	AAA	GGT	GAC	GTC
R	**P**	**T**	**L**	**Q**	**R**	**D**	**K**	**G**	**D**	**V**
GAA	TTA	ATT	GAT	GTA	GAC	GGT	AAA	AAT	GTT	TAC
E	**L**	**I**	**D**	**V**	**D**	**G**	**K**	**N**	**V**	**Y**
GTC	AAA	TTG	ACC	GGT	GCT	TGT	ACT	GGT	TGC	CAA
V	**K**	**L**	**T**	**G**	**A**	**C**	**T**	**G**	**C**	**Q**
ATG	GCA	TCC	ATG	ACA	TTA	GGT	GGT	ATA	CAA	CAA
M	**A**	**S**	**M**	**T**	**L**	**G**	**G**	**I**	**Q**	**Q**
AGA	TTG	ATC	GAA	GAA	TTG	GGT	GAG	TTC	GTC	AAA
R	**L**	**I**	**E**	**E**	**L**	**G**	**E**	**F**	**V**	**K**
GTT	ATC	CCA	GTC	TCC	GCT	GCC	GCA	CAC	GCC	CAA
V	**I**	**P**	**V**	**S**	**A**	**A**	**A**	**H**	**A**	**Q**
ATG	GAA	GTC	TGA							
M	**E**	**V**	**-**							

Table 3.2 Primers used for vector construction.

Genetic element	Primer sequence (5′→3′)	Restriction enzymes
p*ZmUbi1*+1sti	F: TAAGCAGGATCCGGAGTGCAGTGCAGCGTGA	*Acc*65I
	R: TGCTTACTGCAGAAGTAACACCAAACAACAG	*Sal*I
p*ZmUbi1*+1sti	F: TAAGCAGGTACCTGCAGTGCAGCGTGACCCG	BamHI
	R: TGCTTAGTCGACCTGCAGAAGTAACACCAAA	PstI
p*ZmUbi1*+1sti	F: TAAGCAGAATTCTGCAGTGCAGCGTGACCCG	Acc65I
	R: TGCTTAGGATCCCTGCAGAAGTAACACCAAA	BamHI
p*ZmUbi1*+1sti	F: TAAGCAGAATTCTGCAGTGCAGCGTGACCCG	EcoRI
	R: TGCTTAGAGCTCCTGCAGAAGTAACACCAAA	SacI
p*OsActin*	F: TAAGCAGGTACCTAGCTAGCATACTCGAGGT	Acc65I
	R: TGCTTATCTAGACTTCTACCTACAAAAAAGC	XbaI
Cox4-*NifH*^{Av}-t*Nos*	F: TAAGCAGTCGACATGCTTTCACTTAGACAAT	SalI
	R: TGCTTAGCATGCGATCTAGTAACATAGATGA	SphI
Cox4-*NifH*^{Ht}-t*Nos*	F: TAAGCAGTCGACATGCTTTCACTTAGACAA	PstI
	R: TGCTTAGCATGCACATACAAATGGACGAACG	SphI
Su9-*NifM*^{Av}-t*Nosv*	F: TAAGCATCTAGAATGGCCTCCACTCGTGTCC	XbaI
	R: TGCTTAGTCGACGATCTAGTAACATAGATGA	SalI
Su9-*NifS*^{Av}-t*Adh1*	F: TAAGCAGGATCCATGGCCTCCACTCGTGTCC	BamHI
	R: TGCTTAAAGCTTGAGCGACCTCATGCTATAC	HindIII
Su9-*NifU*^{Av}-t*Cyc1*	F: TAAGCAGAGCTCATGGCCTCCACTCGTGTCC	SacI

R: TGCTTAGGATCCCTTCGAGCGTCCCAAAACC BamHI

p*ZmUbi1*+1st I, maize ubiquitin promoter and first intron; p*OsActin*, rice actin promoter; Cox4, mitochondrial targeting peptide of *S. cerevisiae* cytochrome c oxidase subunit IV; Su9, mitochondrial targeting peptide of *Neurospora crassa* mitochondrial ATPase subunit 9; TS, Twinstrep tag; $NifH^{Av}$, *A. vinelandii NifH;* $NifH^{Ht}$, *H. thermophilus NifH*; $NifM^{Av}$, *A. vinelandii NifM*; $NifS^{Av}$, *A. vinelandii NifS;* and $NifU^{Av}$, *A.* vinelandii *NifU*; t*Nos*, nopaline synthase terminator; t*Cyc1*, cytochrome-c oxidase terminator; t*Adh1*, alcohol dehydrogenase terminator.

3.3.2 Transformation and recovery of transgenic rice plants

Seven-day-old mature rice embryos (*O. sativa* cv. Nipponbare or cv. Bomba) were used as explants for transformation using particle bombardment (Sudhakar et al. 1998; Valdez et al. 1998). Ten mg gold particles were coated with transgene constructs (pAvNifH, pAvNifM, pAvNifS, and pAvNifU, or pHtNifH and pAvNifM) and the selectable marker hygromycin phosphotransferase (*Hpt*) vector at a molar ratio of 3:3:3:3:1 or 3:3:1. The generation of transgenic rice callus and regeneration and hardening of transgenic plantlets were as described in Chapter II.

3.3.3 Gene expression analysis by real-time quantitative reverse transcription PCR

Total RNA was isolated from rice callus and leaves of the corresponding regenerated plants using the LiCl method (Creissen and Mullineaux 1995). The cDNA was synthesized from 1 µg of total RNA using Maxima™ H minus cDNA Synthesis Master Mix with dsDNase kit (Thermo Fisher Scientific). The cDNA products were diluted 5-fold prior to use in real-time quantitative reverse transcription PCR (RT-qPCR). RT-qPCR was carried out as previously described (Chapter II) using the gene-specific primers listed in **Table 3.3**. Primer specificity was confirmed by melting curve analysis of the final PCR products and the identity of the PCR products was confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA.

Table 3.3 Primers used for real-time quantitative reverse transcription PCR analysis.

Gene	Primer sequence (5′→3′)	Product size (bp)
$Actin^{Os}$	F: TCATGTCCCTCACAATTTCC	181
	R: GACTCTGGTGATGGTGTCAGC	
$NifH^{Av}$	F: CGACGTTTTGGGTGATGTTGT	187
	R: AGAGTTGCAGATCAAACCACCT	
$NifH^{Ht}$	F: ACGCTCATTCTGGTGGTGTT	198
	R: CCATTGGATGATCTGGGGCA	
$NifM^{Av}$	F: TGCAAGGTGGTTTGTTAGGTG	156
	R: GTCTTGCTGGACTGACGGAT	

NifS^{Av}	F: GCAGACAAAATGGCGGTTGA	154
	R: ACCCTTCCAGTTTGCATTCCT	
NifU^{Av}	F: CGAAACAGTCTTGGCTGCAA	206
	R: CAGCGGAGACTGGGATAACT	

Actin^{Os}*, O. sativa Actin; NifH*^{Av}*, A. vinelandii NifH; NifH*^{Ht}*, H. thermophilus NifH*; *NifM*^{Av}*, A. vinelandii NifM*; *NifS*^{Av}*, A. vinelandii NifS;* and *NifU*^{Av}*, A.* vinelandii *NifU.*

3.3.4 Protein extraction and immunoblot analysis

Total rice protein extracts for initial expression analysis from callus and leaves were prepared as described earlier (chapter II).

Soluble rice protein extracts were prepared by grinding ~50 mg leaf tissue (snap-frozen in liquid N_2) in 2 mL Eppendorf tubes using 3 mm BeadBug steel-balls and a microtube homogenizer (Benchmark Scientific, Sayreville, NJ, USA) operating at 400 rpm for 20s. Leaf powder was resuspended in 7 volumes (v/w) of extraction buffer comprising 100 mM Tris-HCl (pH 8.6), 200 mM NaCl and 10% glycerol, supplemented with 1 mM PMSF, 1 μg mL^{-1} leupeptin and 5 mM EDTA. After three rounds of homogenization, cell debris was removed by centrifugation at 20,000 × *g* for 5 min at 4 °C, and the supernatant was collected and stored at −80°C. Soluble rice callus extracts were prepared using a blender (see below 3.3.5 section).

Rice proteins were separated by SDS-PAGE and then immunoblotted to Protran Premium 0.45 μm nitrocellulose membranes (GE Healthcare, Chicago, IL, USA) using a semidry transfer apparatus (Bio-Rad, Hercules, CA, USA) at 20 V for 45 min. Similar loading was confirmed by staining polyacrylamide gels with Coomassie Brilliant Blue or staining of nitrocellulose membranes with Ponceau S. The membranes were blocked with 5% non-fat milk in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.02% Tween-20 (TBS-T) for 1 h at room temperature before incubation with primary antibodies overnight at 4 °C. Primary polyclonal antibodies against NifH^{Av}, NifM^{Av}, NifS^{Av}, and NifU^{Av} (generated at the Centre for Plant Biotechnology and Genomics, CBGP) or monoclonal antibodies against strep-tag II (2-1507-001, IBA Lifesciences, Göttingen, Germany), or Rubisco were diluted 1: 2,000-1: 5,000 in TBS-T supplemented with 5% bovine serum albumin (BSA). Secondary antibodies (Invitrogen, Thermo Fisher Scientific) were diluted 1: 20,000 in TBS-T supplemented with 2% non-fat milk and incubated for 2 h at room temperature. Membranes were developed on medical X-ray films (AGFA, Mortsel, Belgium) using enhanced chemiluminescence.

3.3.5 Purification of OsNifH by strep-tag affinity chromatography

Preparation of *Os*NifH^{Av} (hereafter *Os*NifH^{Av} to indicate expression in rice) and *Os*NifH^{Ht} (hereafter *Os*NifH^{Ht}) protein extracts for strep-tag affinity chromatography (STAC) purifications was performed at O_2-levels below 1 ppm in anaerobic (Coy Laboratory Products, Grass Lake, MI, USA, or MBraun, Garching, Germany). The purification was carried out at Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)- Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus Montegancedo UPM, Madrid, Spain） Rice callus was snap-frozen in liquid nitrogen, transferred inside the glovebox and then resuspended in lysis buffer comprising 100 mM Tris-HCl (pH 8.5), 200 mM NaCl, 10% glycerol, 3 mM sodium dithionite (DTH), 5 mM 2-mercaptoethanol (2-ME), 1 mM PMSF, 1 μg mL^{-1} leupeptin, 10 μg mL^{-1} DNAse I, and 1:200 (v/v) BioLock solution (2-0205-050, IBA Lifesciences) at a ratio of 1:2 (w/v). Above-ground tissue from Fe fertilized plants was harvested at the end of the 12 h dark cycle (before the onset of light), snap-frozen in liquid nitrogen, transferred inside the glovebox, and then resuspended in the lysis buffer at a ratio of 1:4 (w/v). The amounts of callus and plant tissue used for purifications are reported as fresh-weight. Callus and plant total extracts were prepared by mechanical disruption under an anaerobic atmosphere using a blender (Oster 4655) modified with a water-cooling system operating at full speed in four cycles of 2 min at 4 °C. The total extract was transferred to centrifuge tubes equipped with sealing closures (Beckman Coulter, Brea, CA, USA) and centrifuged at 50,000 × *g* for 1.5 h at 4 °C using an Avanti J-26 XP device (Beckman Coulter). The supernatant was passed through filtering cups with a pore size of 0.2 μm (Nalgene, Thermo Fisher Scientific) to produce a cell-free extract containing soluble proteins. This was loaded at a flow rate of 2.5 mL min^{-1} onto a 5 mL Strep-Tactin XP column (IBA LifeSciences) attached to an ÄKTA FPLC system (GE Healthcare). The column was washed at 16 °C with 150 mL washing buffer comprising 100 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10% glycerol, 2 mM DTH, and 5 mM 2-ME. Bound proteins were eluted with 15-20 mL washing buffer supplemented with 50 mM biotin (IBA LifeSciences). The elution fraction was concentrated using an Amicon Ultra centrifugal filter with a cut-off pore size of 10 kDa (Merck Millipore, Burlington, MA, USA). Biotin was removed by passing the protein through PD-10 desalting columns (GE Healthcare) equilibrated with washing buffer. Desalted eluate was further concentrated and snap-frozen in cryovials (Nalgene) and stored in liquid nitrogen.

3.3.6 Quantification of purified OsNifH proteins

The yields of purified *Os*NifH^{Av} and *Os*NifH^{Ht} were quantified by the densitometric analysis of Coomassie gel titration. Images of the Coomassie SDS-PAGE gels were analysed using ImageJ (Davarinejad 2015). The standards were the purified *S. cerevisiae* NifH^{Av} (*Sc*NifH^{Av}) or NifB^{Mi} (*Sc*NifB^{Mi}) proteins. The standard loading amount range was 0.025-2 µg. The concentration of purified *Os*NifH was determined by the comparison of the optical density of bands among *Os*NifH proteins and standard.

3.3.7 N-terminal sequencing and mass spectrometry analysis of purified OsNifH proteins

N-terminal amino acid sequences were determined by Edman degradation (Centro de Investigaciones Biológicas, Madrid, Spain). *Os*NifH^{Ht} protein (~250 pmol) was separated by SDS-PAGE, transferred to a Sequi-Blot PVDF membrane with a 0.2 µm pore size (Thermo Fisher Scientific) in 50 mM borate buffer (pH 9.0), stained with freshly prepared Coomassie R-250 (Sigma-Aldrich, St Louis, MO, USA) (0.1% in 40% methanol and 10% acetic acid), and then destained using 50% methanol.

Mass spectrometry analysis was performed at the Universidad Complutense de Madrid, Spain. *Os*NifH^{Ht} protein (~60 pmol) was separated by SDS-PAGE followed by Coomassie staining and destaining as described above. Protein excised from the gel was analyzed by MS.

3.3.8 *In vitro* NifH activity analysis

The activity of *Os*NifH was determined in an anaerobic chamber as described by (Jiang et al. 2021). The enzymatic activity assay was carried out at Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)- Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus Montegancedo UPM, Madrid, Spain. Purified *Os*NifH^{Av} and *Os*NifH^{Ht} were analyzed using an acetylene reduction assay (ARA) following the addition of NifDK^{Av} and ATP-regenerating mixture (1.23 mM ATP, 18 mM phosphocreatine, 2.2 mM $MgCl_2$, 3 mM DTH, 46 µg mL^{-1} creatine phosphokinase, 22 mM Tris-HCl pH 7.5) in a final volume of 400 µL in 9 mL serum vials containing 500 µL acetylene under an argon atmosphere. The assay was performed at 30 °C in a shaking water bath for 20 min. Reactions were stopped by adding 100 µL 8 M NaOH. Positive control reactions for acetylene reduction were carried out with NifH^{Av}. The ethylene formed in the reaction was measured in 50-µL gas-phase samples using a

Porapak N 80/100 column in a gas chromatograph (Shimadzu, Duisburg, Germany). The ethylene peak from reactions devoid of both NifH and NifDK proteins (no nitrogenase component added) in each experiment was subtracted from the corresponding reported ARA activities.

3.3.9 *In vitro* [Fe-S] cluster reconstitution and reconstituted NifH activity

The [Fe-S] cluster reconstitutions of *Os*NifH^{Av} and *Os*NifH^{Ht} were carried out by incubating with *Ec*NifU^{Av} (*A. vinelandii* NifU expressed and purified from *E. coli* cells) previously loaded with [4Fe-4S] clusters in anaerobic chambers. First, 20 μM NifU dimer was added to 100 mM MOPS (pH 7.5) supplemented with 8 mM 1,4-dithiothreitol (DTT), and the reaction was incubated at 37 °C for 30 min. Then, 1 mM L-cysteine, 1 mM DTT, 300 μM $(NH_4)_2Fe(SO_4)_2$, and 225 nM NifS^{Av} purified from *E. coli* (*Ec*NifS^{Av}) were added to the reduced NifU and incubated the reaction on ice for 3 h. Finally, the proteins were diluted 40,000-fold in 100 mM MOPS (pH 7.5) and concentrated using Amicon centrifugal filters with a 30-kDa cut off to remove excess reagents. Purified *Os*NifH^{Av} or *Os*NifH^{Ht} protein was mixed with the [4Fe-4S] cluster-reconstituted *Ec*NifU^{Av} (NifU-mediated reconstitution) and then immediately used for the ARA. The enzymatic activity assay was carried out at Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)- Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus Montegancedo UPM, Madrid, Spain.

3.3.10 FeMo-co synthesis and apo-NifDK reconstitution *in vitro*

NifB-dependent FeMo-co synthesis assays were carried out within anaerobic chambers as described by (Jiang et al. 2021). The activity assay was carried out by Dr S. Buren at Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM)- Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus Montegancedo UPM, Madrid, Spain. Each 100 μL reaction contained 3 μM recombinant *Os*NifH^{Ht} with *Sc*NifB^{Mt}, 1 μM *Sc*NifB^{Mt}, 9 μM [4Fe-4S] cluster-loaded *Ec*NifU^{Av}, 1.5 μM apo-NifEN^{Av}, 0.3 μM apo-NifDK^{Av}, 125 μM SAM, 17.5 μM Na_2MoO_4, 175 μM R-homocitrate, 1 mg mL^{-1} BSA, and ATP-regenerating mixture (1.23 mM ATP, 18 mM phosphocreatine, 2.2 mM $MgCl_2$, 3 mM DTH, 46 μg mL^{-1} creatine phosphokinase, 100 mM MOPS pH 7.5) at 30 °C for 1 h. In the positive control reaction, the assay was carried out by using *Sc*NifH^{Ht} instead of the rice purified variant, and NifB-co (20.4 μM Fe) instead of *Sc*NifB^{Mt} and *Ec*NifU^{Av}. In the negative control reaction, the reaction mixture contained all the components except for any source of NifB-co or *Sc*NifB^{Mt} and

*Ec*NifU^{Av}. Following the *in vitro* synthesis and insertion of FeMo-co, 17.5 μM $(NH_4)_2MoS_4$ was added to prevent further FeMo-co incorporation into apo-NifDK^{Av}, and the reaction was incubated for 20 min at 30 °C, shaking at 600 rpm. Activation of apo-NifDK^{Av} was analyzed by adding 500 μL ATP-regenerating mixture and NifH^{Av} (2.0 μM final concentration) in 9-mL vials containing 500 μL acetylene under argon gas. The acetylene reduction assay was performed at 30 °C for 20 min, and the resulting ethylene was measured in 50 μL gas-phase samples using a Porapak N 80/100 column as described above.

3.3.11 Statistical analysis

For relative gene expression data, standard deviation (SD) was calculated based on three technical replicates. The SD for plant activity assay data was calculated based on two biological replicates (two technical replicates each). The SD for callus activity assay data was calculated based on three biological replicates (three technical replicates each).

3.4 Results

3.4.1 Expression constructs for rice transformation

In contrast to the native *O. sativa* codon usage table, the *S. cerevisiae* codon-optimized *NifH*^{Av}, *NifM*^{Av}, *NifS*^{Av}, *NifU*^{Av}, and *NifH*^{Ht} sequences showed high relative adaptiveness (Nakamura et al. 2000). There were no low or very low frequency codons in these sequences (Tran et al. 2017). Therefore, the above *S. cerevisiae* codon-optimized *Nif* gene sequences were used for rice transformation. *Nif* gene expression was controlled by the strong constitutive *ZmUbi1*+1sti promoter except *NifM*^{Av}, which was controlled by the *OsActin* promoter (also constitutive). The N-terminal mitochondrial targeting peptides Cox4 or Su9 were added to direct the Nif proteins to rice mitochondria (Baysal et al. 2020). The TS tag was added to the N-terminal of NifH to facilitate the detection and purification of *Os*NifH. This tag does not affect NifH activity in *S. cerevisiae* or *N. benthamiana* (Eseverri et al. 2020; Jiang et al. 2021). The *NifH*^{Av} and *NifH*^{Ht} constructs were used in separate experiments, with the former combined with *NifM*^{Av}, *NifS*^{Av} and *NifU*^{Av} and the latter with only *NifM*^{Av}. An additional construct carrying the selectable marker *Hpt* was also used. Transgenic rice callus was transformed by direct DNA transfer as described (Sudhakar et al. 1998; Valdez et al. 1998). Plantlets were regenerated from the corresponding callus lines under hygromycin selection and grown to maturity as described (Sudhakar et al. 1998; Valdez et al. 1998).

3.4.2 mRNA expression analysis of *Nif* genes in transgenic rice callus and regenerated plants

mRNA expression of *NifH*^{Av}, *NifM*^{Av}, *NifS*^{Av}, *NifU*^{Av}, and *NifH*^{Ht} in callus and the corresponding regenerated plants was analyzed by RT-qPCR. Two independent *Os*NifH^{Av} lines (AvH147 and AvH182) were confirmed to co-express *NifH*^{Av}, *NifM*^{Av}, *NifS*^{Av}, *NifU*^{Av} at mRNA in callus (**Figure 3.1a**). mRNA expression was also confirmed in the corresponding regenerated plants of line AvH182 (**Figure 3.1b**). Three independent *Os*NifH^{Ht} rice callus lines (HtH189, HtH200 and HtH202) and their corresponding regenerated plants co-expressed of *NifH*^{Ht} and *NifM*^{Av} at the mRNA level (**Figure 3.1c, d**). In both *Os*NifH^{Av} and *Os*NifH^{Ht} lines, all five *Nif* genes were expressed at higher levels in plants than in the callus (**Figure 3.1**). The relative mRNA expression level of *NifM*^{Av} was lower than the other *NifH*^{Av}, *NifS*^{Av}, *NifU*^{Av}, and *NifH*^{Ht}.

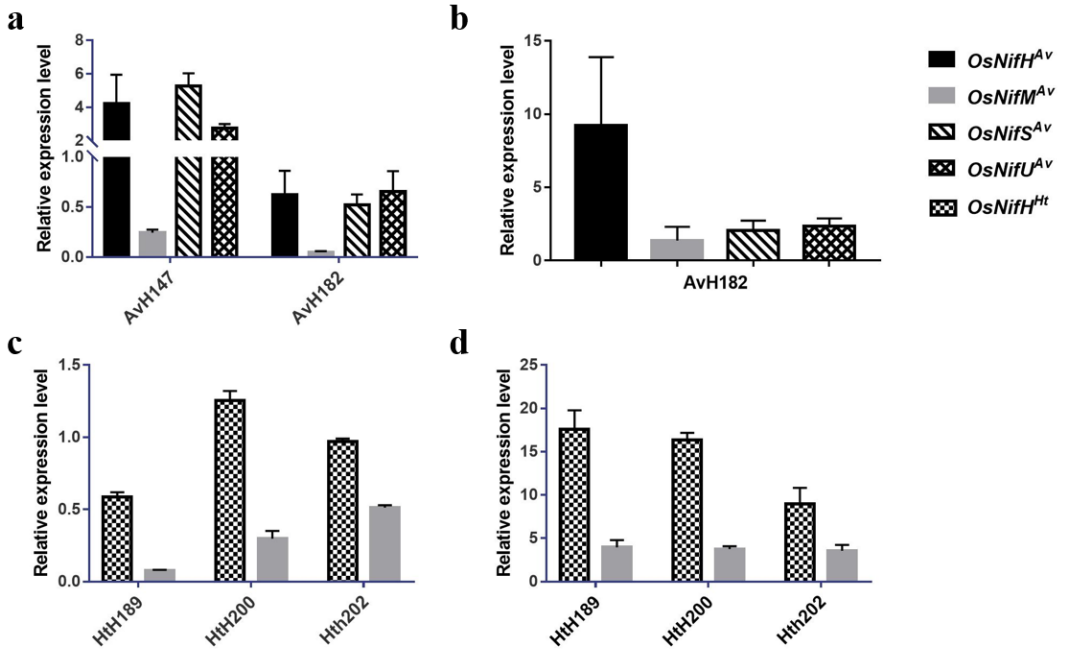

Figure 3.1 mRNA expression analysis of *NifH*^{Av}, *NifM*^{Av}, *NifS*^{Av}, *NifU*^{Av}, and *NifH*^{Ht} in representative *OsNifH*^{Av} and *OsNifH*^{Ht} callus lines and the corresponding regenerated plants. Relative mRNA expression levels of *NifH*^{Av}, *NifM*^{Av}, *NifS*^{Av}, *NifU*^{Av} in two independent *Os*NifH^{Av} callus lines (a) and line AvH182 corresponding plants (b). Relative mRNA expression levels of *NifH*^{Ht} and *NifM*^{Av} in three independent *Os*NifH^{Ht} callus lines (c) and the corresponding plants (d). Data (normalized to *OsActin* mRNA) are means ± SD (n = 3 technical replicates). *OsNifH*^{Av}, *O. sativa*-derived *NifH*^{Av}; *OsNifM*^{Av}, *O. sativa*-derived *NifM*^{Av}; *OsNifS*^{Av}, *O. sativa*-derived *NifS*^{Av}; *OsNifU*^{Av}, *O. sativa*-derived *NifU*^{Av}; *OsNifH*^{Ht}, *O. sativa*-derived *NifH*^{Ht}. AvH147, AvH182 are two independent *Os*NifH^{Av} lines, HtH189, HtH200 and HtH202 are three independent *Os*NifH^{Ht} lines.

3.4.3 Transgenic rice callus and plants accumulate *Os*NifH and *Os*NifM proteins

Accumulation of *Os*NifH^{Av}, *Os*NifH^{Ht} and *Os*NifM^{Av} proteins in callus and the corresponding regenerated plants was confirmed by immunoblot analysis using antibodies specific for the NifH^{Av}, TS tag and NifM^{Av}, respectively (**Figure 3.2**). The major *Os*NifH^{Av} moiety on SDS-gel appeared at ~35 kDa which indicated that *Os*NifH^{Av} was mostly correctly processed in rice mitochondria. The *Os*NifH^{Ht} gave rise to a single band of the expected size of ~35 kDa based on SDS-gel migration, indicating that the proteins were correctly processed and stable in rice mitochondria (**Figure 3.2**).

Although I was unable to obtain clear western blot signal for *Os*NifM^{Av} (**Figure 3.2a, c**), the fact that *Os*NifH^{Av} and *Os*NifH^{Ht} were soluble indicated that *Os*NifM^{Av} was expressed at sufficient levels to facilitate NifH polypeptide maturation. I was not able to detect *Os*NifS^{Av} and *Os*NifU^{Av} protein accumulation in callus and regenerated plants despite the high levels of mRNA accumulation of the transgenes (**Figure 3.1**). There was no difference among the 12 HtH200 T1 plants, the *Os*NifH^{Ht} expressing T1 progeny showing normal growth and development (**Figure 3.3a**). Analysis of *Os*NifH^{Ht} protein expression in the T1 generation confirmed that *Os*NifH^{Ht} was stably inherited and expressed in progeny (**Figure 3.3b**).

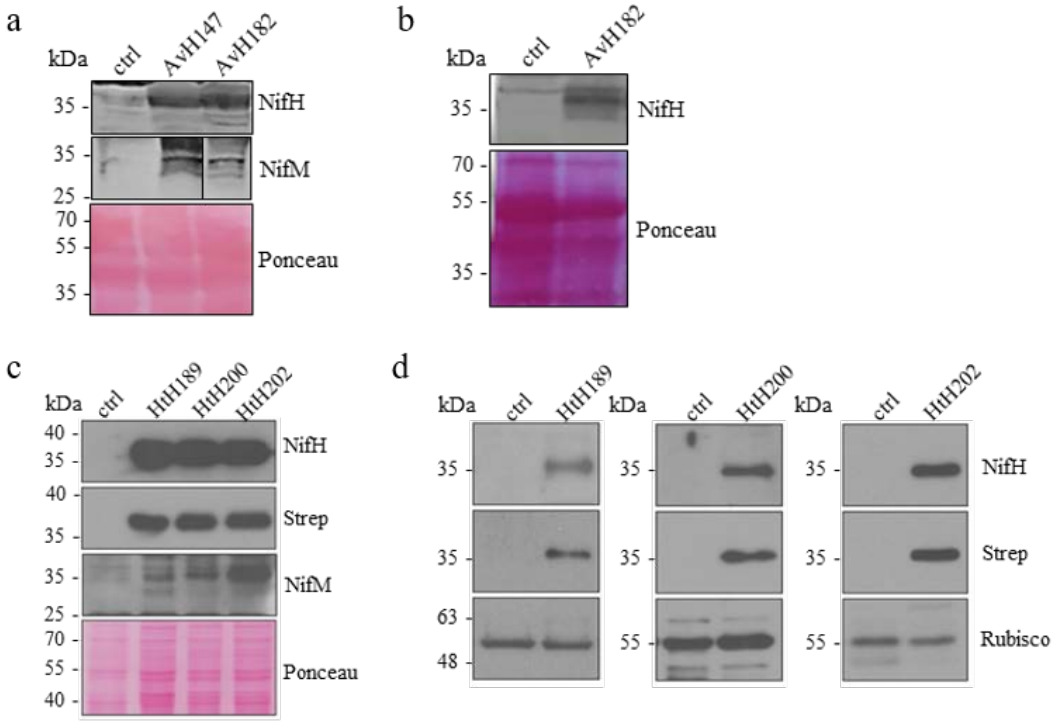

Figure 3.2 Protein accumulation of *Os*NifH^{Av}, *Os*NifH^{Ht} and *Os*NifM^{Av} in callus and plants. Immunoblot analysis of cell-free extracts prepared from callus (a) *Os*NifH^{Av} line; (c) *Os*NifH^{Ht} lines and plants (b) *Os*NifH^{Av} lines; (d) *Os*NifH^{Ht} lines probed with antibodies against NifH, NifM, and the Strep-tag. Ponceau staining or antibodies against RuBisCO were used as loading control.

Ctrl lane shows non-transformed callus and plant lines. AvH147, AvH182 are two independent *Os*NifH^{Av} lines, HtH189, HtH200 and HtH202 are three independent *Os*NifH^{Ht} lines.

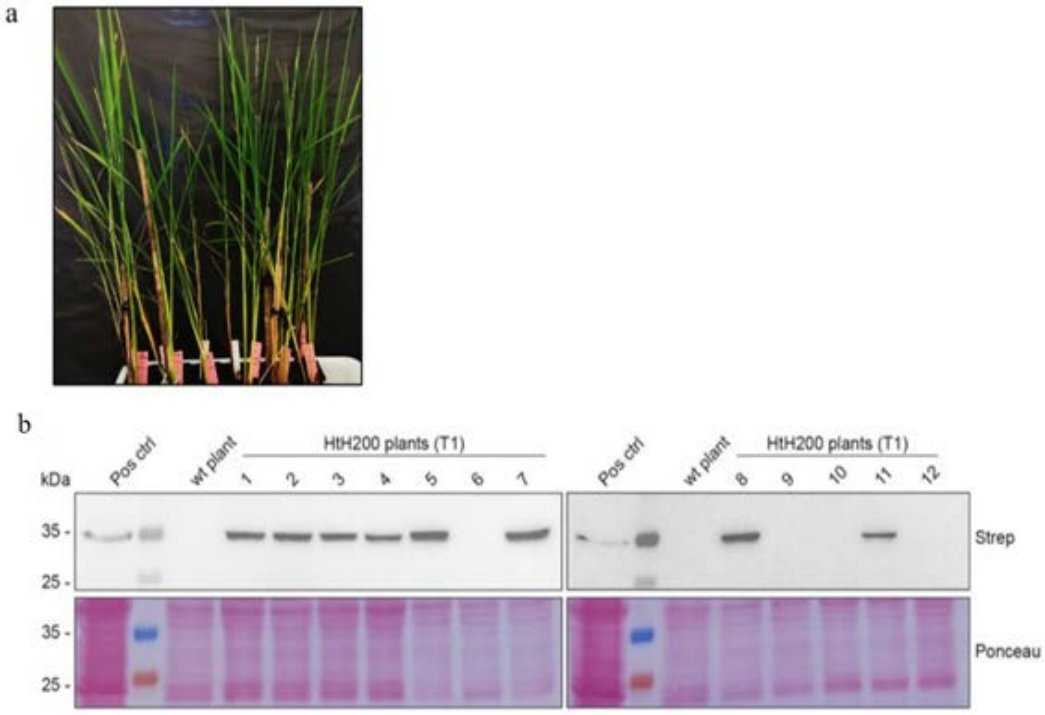

Figure 3.3 Phenotype and NifH^{Ht} protein accumulation in *Os*NifH^{Ht} T1 plants. (a) Phenotype of *Os*NifH^{Ht} expressing T1 progeny showing normal growth and development. (b) Stable expression of *Os*NifH^{Ht} in the T1 segregating generation of plant line HtH200. Immunoblot analysis was performed using leaf soluble protein extracts and antibodies for the TS tag. Protein extract from callus expressing *Os*NifH^{Ht} (line HtH206, an additional independent line) was used as the positive control (Pos ctrl).

3.4.4 Purification of *Os*NifH^{Av} and *Os*NifH^{Ht}

*Os*NifH^{Av} and *Os*NifH^{Ht} were purified from rice callus lines (AvH147, AvH182, HtH189, HtH200 and HtH 202) and the corresponding regenerated plant HtH200 by STAC **(Figure 3.4**). To minimize oxygen exposure produced during photosynthesis and maximize the isolation of functional *Os*NifH, plants were Fe-fertilized and harvested before the onset of light at the end of the dark period, following a procedure previously shown to be successful for active NifH^{Av} and NifH^{Ht} from tobacco chloroplasts and mitochondria (Aznar-Moreno et al. 2021; Jiang et al. 2021). No significant amount of protein was lost during centrifugation and filtration of the cell extract, confirming that *Os*NifH was accumulated in a soluble form (**Figure 3.4a, b**). Although *Os*NifH^{Av} was purified near homogeneity (**Fig. 3.4a, c**), SDS-PAGE analysis showed additional slower migrating co-eluting proteins. Isolated *Os*NifH^{Av} yields were 90 and 50 μg 100mg^{-1} fresh weight callus

from lines AvH147 and AvH189, respectively (**Figure 2.5**). Purified *Os*NifH^{Ht} was mostly soluble (**Figure 3.4b**) and could be isolated from plant HtH200 with yields of 50 µg 100 mg^{-1} fresh weight (1-month-old plants) or 25 µg 100mg^{-1} fresh weight (2-month-old plants) (**Figure 3.6g, h**). The yield of *Os*NifH^{Ht} from three independent callus lines was 250 µg 100mg^{-1} (line HtH189), 700 µg 100mg^{-1} (line HtH200) and 650 µg 100mg^{-1} (line HtH202) (**Figure 3.6a-f**). The comparison of migration of *Sc*NifH and *Os*NifH (both targeted to the mitochondria) was also consistent with the correct processing of both *Os*NifH^{Av} and *Os*NifH^{Ht} in the rice mitochondrial matrix, resulting in proteins with molecular weights of ~35 kDa (**Figure 3.7**). Moreover, N-terminal sequencing of purified *Os*NifH^{Ht} revealed a major product with the anticipated N-terminal residues QKP, following the cleavage of the Cox4 peptide, together with three minor variants representing alternative processing events (**Figure 3.8a**). Mass spectrometry identified the *Os*NifH^{Ht} protein with 63% sequence coverage and confirmed that the C-terminus of the protein was intact (**Figure 3.8b**).

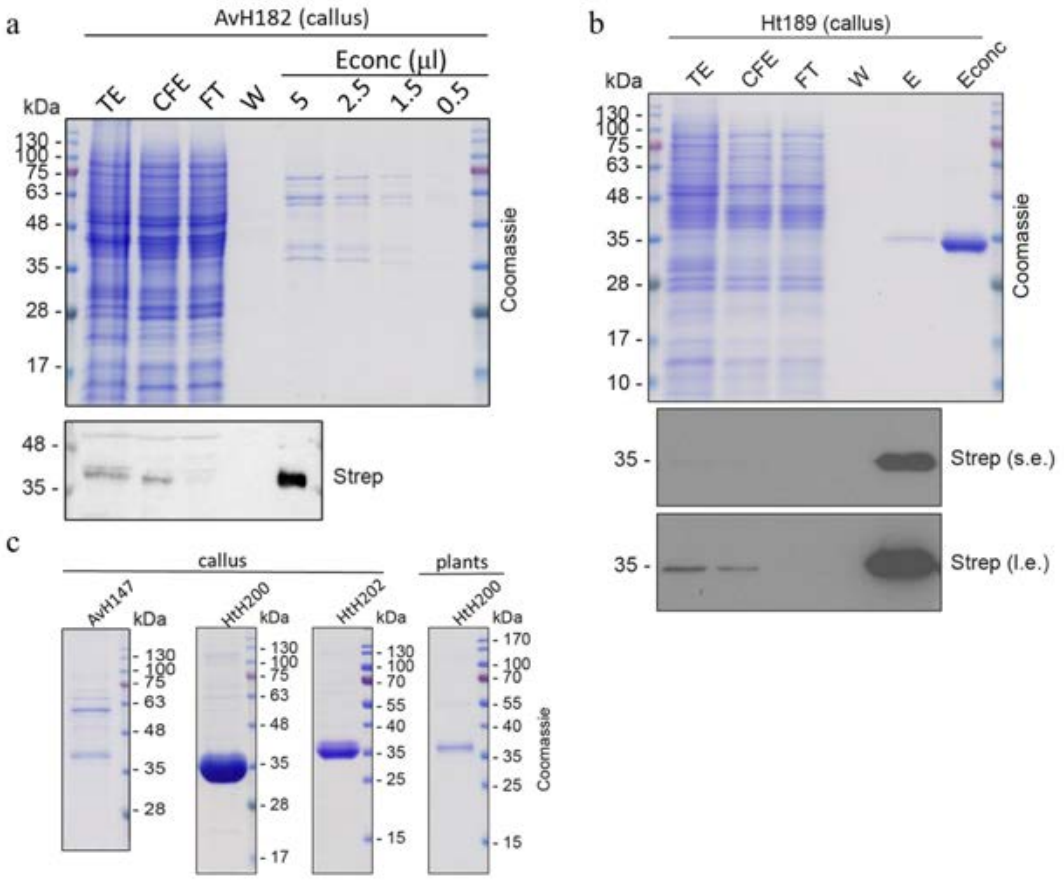

Figure 3.4 STAC purification of *Os*NifH^{Av} and *Os*NifH^{Ht} proteins. (a) Purification of *Os*NifH^{Av} from AvH182 callus. (b) Purification of *Os*NifH^{Ht} from HtH189 callus. Fractions were analyzed by SDS-PAGE followed by Coomassie staining or immunoblotting against antibodies of Strep tag. TE, total extract; CFE, soluble cell-free extract; FT, flow-through fraction; W, wash fraction;

E, elution fraction; Econc, concentrated elution fraction (final sample collected). s.e., short exposure; l.e., long exposure. (c) Final *Os*NifH^{AV} protein purified from AvH147 callus line and *Os*NifH^{Ht} sample isolated from Ht200 and Ht202 callus lines, and Ht200 rice plants.

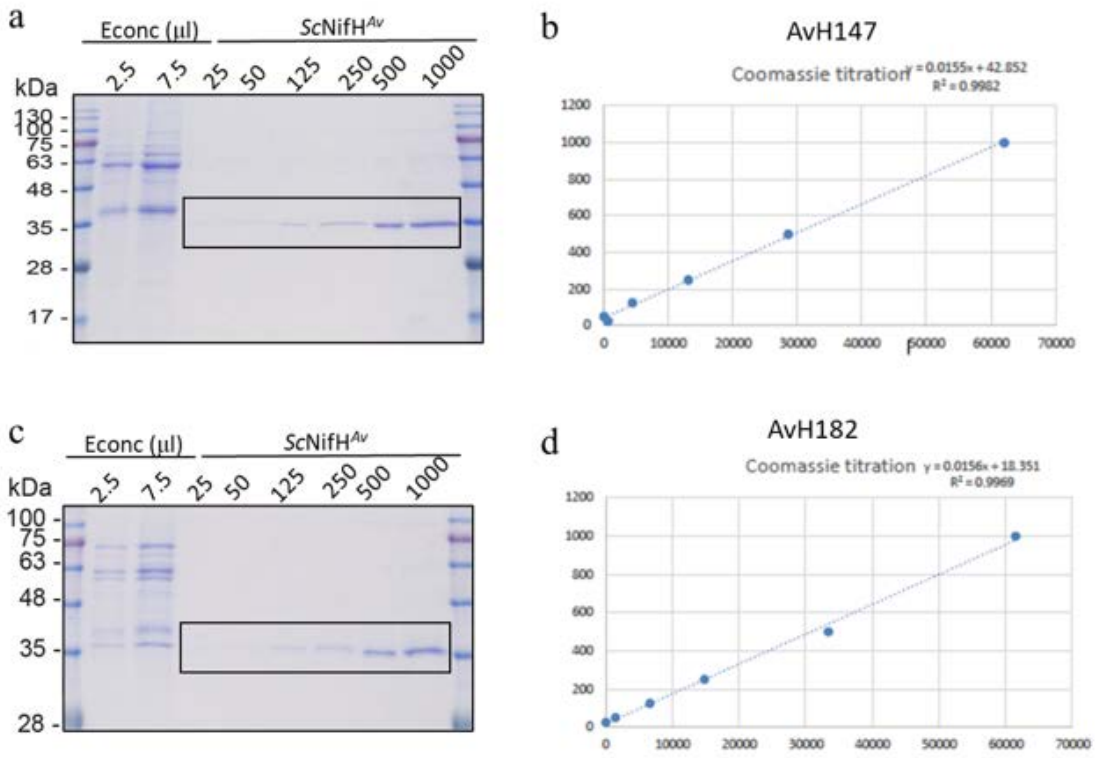

Figure 3.5 Quantification of *Os*NifH^{Av} isolated from rice callus. SDS-PAGE and Coomassie staining of the final *Os*NifH^{Av} protein samples (Econc, concentrated elution fraction) isolated from (a) AvH147 and (c) AvH189, together with the indicated amounts of purified *Sc*NifH^{Av} protein (quantified using the BCA protein assay) (Burén et al. 2019). The *Sc*NifH^{Av} samples enclosed by black squares were used to create a standard curve for quantification of the *Os*NifH^{Av} proteins isolated from (b) AvH147 and (d) AvH189. The intercept and slope for the *Sc*NifH^{Av} standard curve and the R^2 fit to the curve are indicated.

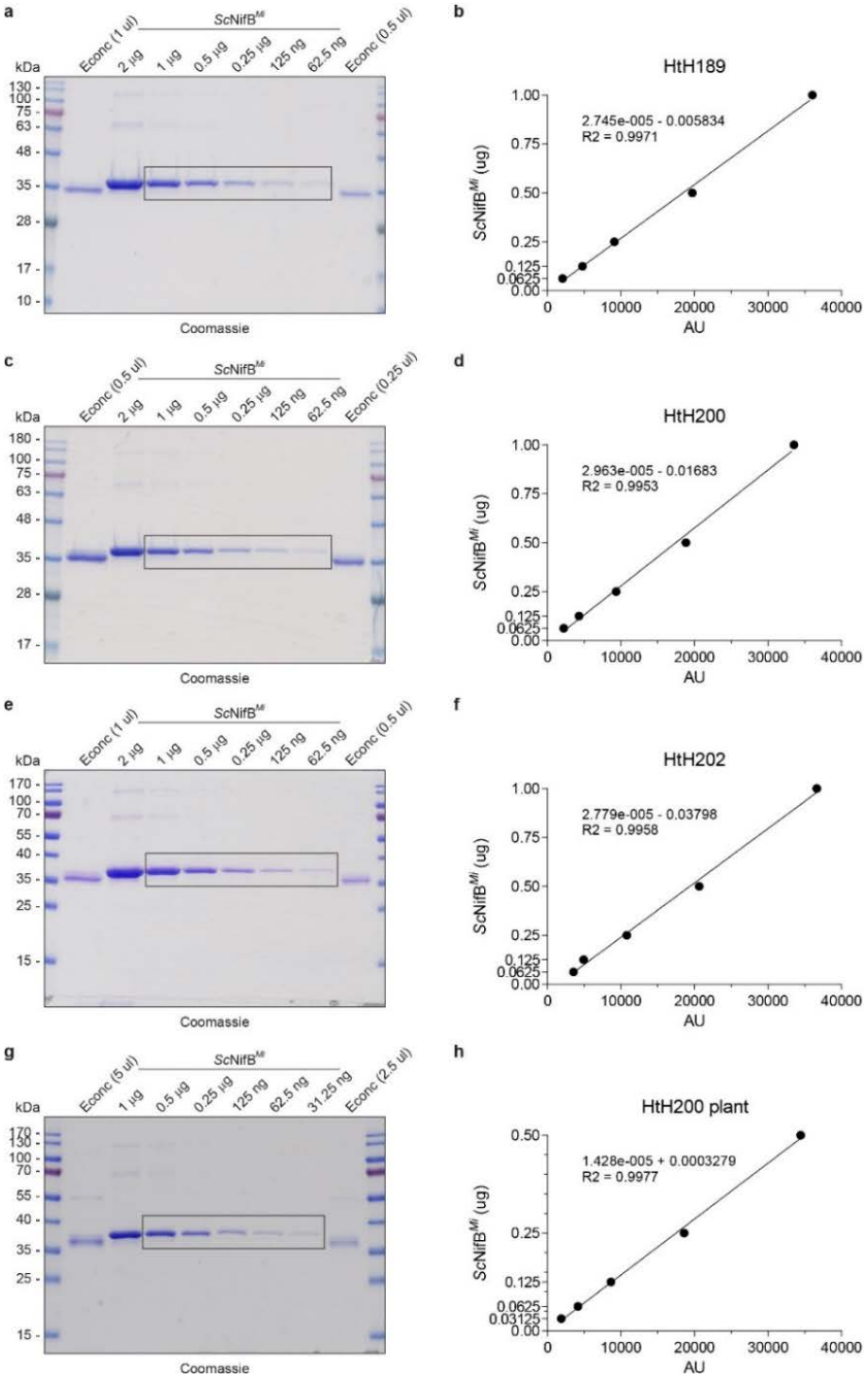

Figure 3.6 Quantification of *Os*NifH^{Ht} isolated from rice callus and plants. SDS-PAGE and Coomassie staining of the final *Os*NifH^{Ht} protein samples (Econc, concentrated elution fraction) isolated from (a) HtH189 callus, (c) HtH200 callus, (e) HtH202 callus, and (g) HtH200 plants together with the indicated amounts of purified *Sc*NifB^{Mi} protein (quantified using the BCA protein assay) (Burén et al. 2019). The *Sc*NifB^{Mi} samples enclosed by black squares were used to create a standard curve for quantification of the *Os*NifH^{Ht} proteins isolated from (b) HtH189 callus, (d) HtH200 callus, (f) HtH202 callus, and (h) HtH200 plants. The intercept and slope for the *Sc*NifB^{Mi} standard curve and the R^2 fit to the curve are indicated.

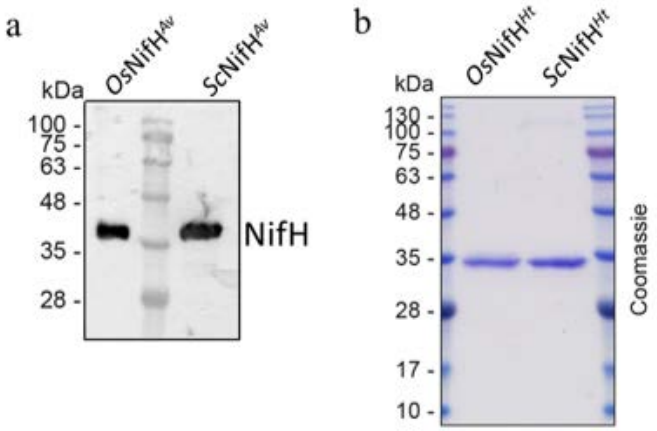

Figure 3.7 Migration of isolated *Os*NifH and *Sc*NifH. (a) Isolated NifH^{Ht} in immunoblotting probed with antibodies against NifH from AvH182 (*Os*NifH^{Av}) or yeast (*Sc*NifH^{Av}). (b)Side-by-side comparison of NifH^{Ht} migration in Coomassie gel when isolated from HtH189 callus (*Os*NifH^{Ht}) and yeast (*Sc*NifH^{Ht}).

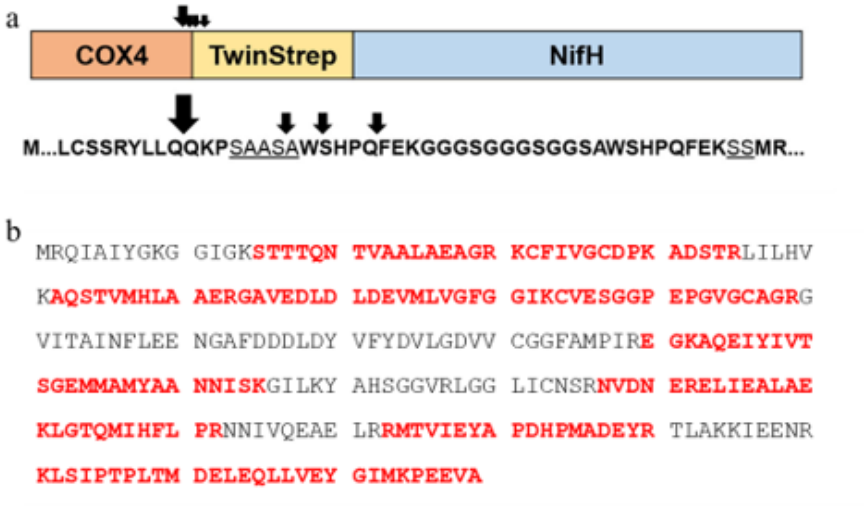

Figure 3.8 N-terminal sequencing and mass spectrometry analysis of isolated *Os*NifH^{Ht}. (a) N-terminal sequencing of *Os*NifH^{Ht} isolated from HtH189 callus. The deduced Cox4 mitochondrial targeting peptide processing sites are indicated with arrows (the major processing cite is indicated by a larger arrow). The underlined sequences indicate peptide linkers between the Cox4 signal and the TS-tag (SAASA), and between the TS-tag and *Os*NifH^{Ht} (SS). (b) Mass spectrometry analysis of purified *Os*NifH^{Ht} protein. Bold red letters indicate the peptides from *Os*NifH^{Ht} (isolated from HtH189 callus) that were identified by mass spectrometry.

3.4.5 *Os*NifH^{Av} and *Os*NifH^{Ht} enzymatic substrate reduction activity *in vitro*

The Fe protein activity of purified *Os*NifH^{Ht} was determined *in vitro* using an ARA after mixing with NifDK purified from *A. vinelandii* (NifDK^{Av}). The *Os*NifH^{Ht} protein isolated from rice leaves showed low activity (~7 nmol ethylene min^{-1} mg^{-1} NifDK^{Av}) (**Figure 3.9a**), probably limited by the relatively low yield of NifH in the plant tissue and hence the low amount of NifH (~8.9 µg) and low ratio of NifH:NifDK (~10:1) in the assay. To

confirm the functionality of the as-isolated rice Fe protein, we performed titration experiments using increasing amounts of *Os*NifH^{Ht} purified from three independent callus lines. As anticipated for a functional Fe protein, higher activity was observed with an increasing NifH:NifDK ratio (**Figure 3.9b**). Importantly, the activity of the *Os*NifH^{Ht} protein isolated from rice leaves was about half that of the callus protein at a similar ratio, which may reflect the use of more diluted plant *Os*NifH^{Ht} protein in the assay (we typically observed slight inhibition caused by large amounts of buffer).

To better understand the effect of NifM on NifH activity, we compared the activities of *Os*NifH^{Ht} proteins purified from three independent rice callus lines in the same experiment at the same NifH:NifDK ratio of 40:1 (**Figure 3.9c**). We observed that the activities were almost identical despite the varying expression level of NifM (**Figure 3.1**), indicating that *Os*NifM^{Av} expression was sufficient in all three rice lines, and agreeing with our observation that soluble *Os*NifH^{Ht} isolated from yeast expressed either alone or with NifM^{Av} presented identical specific activity (**Figure 3.10**).

We then supplemented the reaction mixture with pure *A. vinelandii* NifU expressed in *Escherichia coli* and loaded with [4Fe-4S] clusters (*Ec*NifU^{Av}). The reconstitution of *Os*NifH^{Ht} by the *in vitro* transfer of [4Fe-4S] from NifU increased its activity 9-fold (compare activities in **Figure 3.9c, d**). Importantly, the activity of *Os*NifH^{Ht} and *Sc*NifH^{Ht} proteins was identical after reconstitution (**Figure 3.9d**), suggesting that *Os*NifH^{Ht} was correctly folded but not fully mature, probably due to insufficient insertion of the [4Fe-4S] cluster into *Os*NifH^{Ht} in the rice mitochondria, or perhaps due to instability of the clusters as previously suggested in *N. benthamiana* (Jiang et al. 2021). There was no detectable acetylene reduction from purified *Os*NifH^{Av} after the reconstitution with pure *Ec*NifU^{Av}, probably limited by the low amount of *Os*NifH^{Av} and the low ratio of NifH:NifDK (~10:1) in the assay.

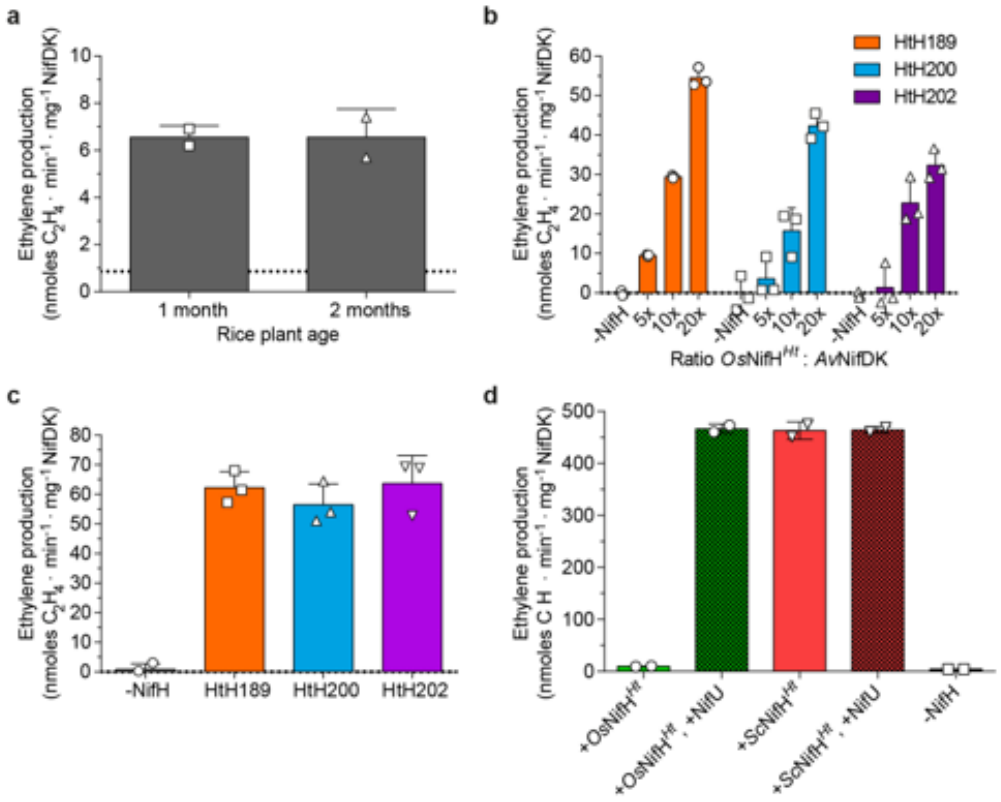

Figure 3.9 Acetylene reduction assay (ARA) with purified *Os*NifH^{Ht} and *Os*NifH^{Av}. (a) ARA using *Os*NifH^{Ht} isolated from HtH200 rice plants (10:1 ratio of *Os*NifH^{Ht}:NifDK^{Av}) (data are means ± SD, *n* = 2 technical replicates). Dotted line indicates the negative control ARA in the absence of NifH (mean of *n* =3 technical replicates). The positive control ARA (40:1 ratio of NifH^{Av}:NifDK^{Av}) generated 1285 ± 225 units (mean ± SD, *n* = 3 technical replicates). (b) ARA with increasing amounts of *Os*NifH^{Ht} isolated from three different rice callus lines (0, 5:1, 10:1 and 20:1 ratio of *Os*NifH^{Ht}:NifDK^{Av}). The positive control ARA (20:1 ratio of NifH^{Av}:NifDK^{Av}) generated 1426 ± 43 units for Ht189, 1127 ± 65 units for Ht200 and 1074 ± 153 units for Ht202 (data are means ± SD, *n* = 3 technical replicates). (c) ARA using *Os*NifH^{Ht} isolated from three different rice callus lines at a 40:1 ratio of *Os*NifH^{Ht}:NifDK^{Av}. The positive control ARA (40:1 ratio of NifH^{Av}:NifDK^{Av}) generated 1285 ± 225 units (data are means ± SD, *n* = 3 technical replicates). (d) ARA using *Os*NifH^{Ht} and *Sc*NifH^{Ht} before and after reconstitution with [4Fe-4S] cluster-loaded *Ec*NifU^{Av}. The positive control ARA (using NifH^{Av}) generated 1553 ± 82 units. All reactions were performed with a 20:1 ratio of NifH:NifDK^{Av} (data are means ± SD, *n* = 2 technical replicates). All activities are reported as nmol ethylene formed per min and mg of NifDK^{Av}.

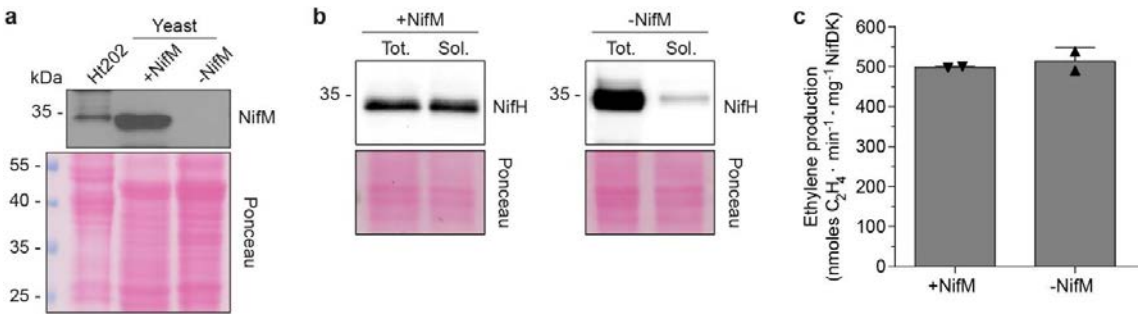

Figure 3.10 NifM^{Av} dependent NifH^{Ht} solubility and activity. (a) Immunoblot analysis to compare the migration of NifM^{Av} produced in rice callus (line Ht202) and yeast (+NifM). Yeast not expressing NifM^{Av} (–NifM) was used as a control for antibody specificity. (b) Accumulation of *Sc*NifH^{Ht} in total protein extracts (Tot.) and soluble protein extracts (Sol.) in yeast either co-

expressing (+NifM) or not expressing (–NifM) *Sc*NifM^{Av}. Ponceau S staining was used as a loading control. (c) ARA using *Sc*NifH^{Ht} isolated from yeast either co-expressing (+NifM) or not expressing (–NifM) *Sc*NifM^{Av} using a 40:1 ratio of *Sc*NifH^{Ht}:NifDK^{Av}. The positive control ARA (40:1 ratio of NifH^{Av}:NifDK^{Av}) yielded 1961 ± 88 units (data are means ± SD, n = 2 technical replicates).

3.4.6 *Os*NifH^{Ht} supports the FeMo-co synthesis *in vitro*

The isolated *Os*NifH^{Ht} also supported *in vitro* FeMo-co synthesis, which is another fundamental role of NifH required for BNF. FeMo-co synthesis and the activation of apo-NifDK^{Av} (containing P-clusters but devoid of FeMo-co) were measured *in vitro* by combining *Methanothermobacter thermautotrophicus* NifB expressed in yeast (*Sc*NifB^{Mt}), *Ec*NifU^{Av} (provider of [4Fe-4S] clusters for NifB-co biosynthesis), *S*-adenosyl methionine (SAM), molybdate, homocitrate, apo-NifEN^{Av} (isolated from a Δ *NifB* strain containing permanent [4Fe-4S] clusters but lacking the FeMo-co precursor), purified *Os*NifH^{Ht} and apo-NifDK^{Av}. Following the FeMo-co synthesis and insertion reactions, tetrathiomolybdate was added to prevent further FeMo-co incorporation into apo-NifDK^{Av} during the acetylene reduction assay. *Os*NifH^{Ht} and *Sc*NifB^{Mt} jointly catalyzed the NifB-dependent synthesis of FeMo-co *in vitro*. FeMo-co synthesis confirmed the compatibility of *Os*NifH^{Ht} and *Sc*NifB^{Mt} as demonstrated by the *in vitro* maturation of NifDK (**Figure 3.11**). It also demonstrated the interspecies compatibility of *Os*NifH^{Ht}, *Sc*NifB^{Mt}, NifEN^{Av} and NifDK^{Av}, collectively establishing the conserved biochemical core of nitrogenase.

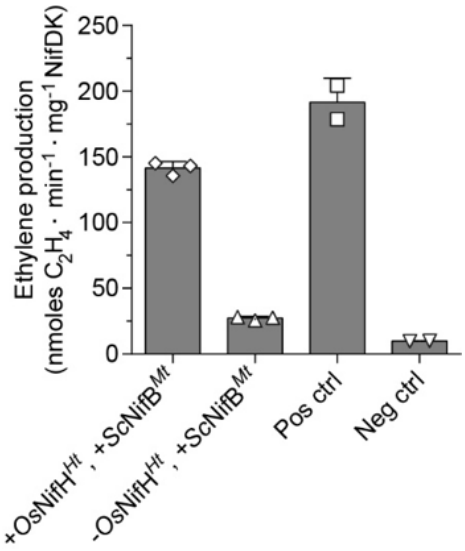

Figure 3.11 NifB-dependent *in vitro* FeMo-co synthesis using *Os*NifH^{Ht} and *Sc*NifB^{Mt}. *In vitro* FeMo-co synthesis was performed by combining *M. thermautotrophicus* NifB purified from yeast (*Sc*NifB^{Mt}) with *Os*NifH^{Ht} (bar 1). The subsequent ARA was performed using a 20:1 molar ratio

of NifH^{Av} to NifDK^{Av}. To inhibit apo-NifDK^{Av} activation by FeMo-co produced during the ARA, tetrathiomolybdate was added to all vials following *in vitro* FeMo-co synthesis and insertion into apo-NifDK^{Av}. Bar 2 represents the background activity observed from using *Sc*NifB^{Mt} in the absence of *Os*NifH^{Ht}. As control reactions for the functionality of apo-NifEN^{Av} and apo-NifDK^{Av}, FeMo-co synthesis was performed using *Sc*NifH^{Ht} in the presence (Pos ctrl) or absence (Neg ctrl) of purified NifB-co. Data are means ± SD, n = 2 technical replicates.

3.5 Discussion

Engineering BNF in plants is a major longstanding goal of plant biotechnology. Earlier strategies to reduce global dependence on N fertilizers included the use of diazotrophic bacteria to colonize the rhizosphere (Stoltzfus et al. 1997; Mueller et al. 2014; Fox et al. 2016; Mus et al. 2016; Van Deynze et al. 2018; Ryu et al. 2020). The introduction of bacterial *Nif* genes into cereals to increase productivity offers a more direct approach in which the plants fix their own N. However, the O_2 sensitivity of nitrogenase and its accessory proteins, the complexity of the machinery that provides metal clusters to nitrogenase components, and the intricate regulation of the nitrogenase system hinder the engineering of BNF plants.

Many *Nif* genes are involved in the assembly and activity of nitrogenase and its metal cofactors in bacteria. The essential bacterial genes that must be transferred to cereals include at least *NifH*, *NifD*, *NifK NifB*, *NifE* and *NifN* because there are no corresponding endogenous proteins. In contrast, some or all of the accessory genes, such as *NifU*, *NifS*, *NifQ* and *NifV* (which provide the FeMo-co building blocks: [Fe-S] clusters, molybdenum, and homocitrate), as well as *NifJ* and *NifF* (which are involved in electron transfer to NifH) show at least partial functional redundancy with plant proteins (Curatti and Rubio 2014; Yang et al. 2017). Plant counterparts of the NifU/NifS system are found in the mitochondria, which could therefore provide in principle a ready source of [Fe-S] clusters as well as energy and the near anoxic environment necessary to protect the heterologous nitrogenase (Balk and Lobréaux 2005). Importantly, [4Fe-4S] cluster availability and incorporation efficiency differ among different species, e.g., *S. cerevisiae* and *N. benthamiana* (López-Torrejón et al. 2016; Jiang et al. 2021).

In this study, I generated rice plants co-expressing NifH^{Av}, NifM^{Av}, NifU^{Av} and NifS^{Av}, or co-expressing NifH^{Ht} and NifM^{Av}. The proper expression kinetics of *Nif* genes may be important in engineering nitrogen fixation (Poza-Carrión et al. 2014) To achieve the appropriate stoichiometric ratio for Nif protein production in yeast, the *Nif* genes were expressed at a constant ratio, with *NifH* and *NifUS* designed to express at higher levels than *NifM* (Burén et al. 2017). Here, two different constitutive promoters *ZmUbi* and

OsActin were selected to obtain lower *NifM* expression. Although both constitutive promoters are highly active in rice, β-glucuronidase expressed under the control of *ZmUbi1* showed 7-fold higher activity compared the that of the *OsActin* promoter (Cornejo et al. 1993).

It is not yet possible to directly transform the plant mitochondrial genome, but proteins encoded by nuclear transgenes can be directed to the mitochondria if augmented with the appropriate N-terminal targeting peptide. I selected the N-terminal mitochondrial targeting peptides Cox4 or Su9 to direct the Nif proteins to rice mitochondria, because both had been shown to be effective in directing the eGFP to the mitochondria in earlier experiments (Baysal et al. 2020).The correctly mitochondrial processing peptidase (MPP) processed *Os*NifH^{Av} and *Os*NifH^{Ht} were purified from stably transformed rice tissues by STAC. Despite the enhanced solubility of *Os*NifH^{Av} (**in Figure 3.4a**) compared to tobacco, it was inactive even after [Fe-S] cluster reconstitution. The probable reasons could be the reduction activity was limited to below detection level due to the low amount of *Os*NifH^{Av} and the low ratio of NifH:NifDK (~10:1) in the assay. Since *Os*NifH^{Av} showed the proper mitochondria targeting and MPP processing, another reason might be that the cysteine residues (Cys97 and Cys132) which interact with [4Fe-4S] cluster in *Os*NifH^{Av} might be altered and resulting in steric hindrance around the [4Fe-4S] insertion site. Consistent with the previous study in *S. cerevisiae* and *N. benthamiana* (Jiang et al. 2021), *Os*NifH^{Ht} exhibited much higher protein accumulation, with the purified yields being 22-fold higher than *Os*NifH^{Av}. The isolated *Os*NifH^{Ht} protein was colorless, but this was expected given its low concentration in rice extracts compared to those from bacteria or yeast. Most importantly, the *Os*NifH^{Ht} protein was soluble and stable in rice mitochondria, although not fully equipped with its [4Fe-4S] cluster, and could readily be activated when complemented with [4Fe-4S] cluster-loaded NifU *in vitro*. This is essential because the Fe protein must be abundant, stable and soluble for successful nitrogenase engineering in plants. The accumulation of an unstable NifH protein could induce cellular stress and damage if exposed to O_2, or if [4Fe-4S] clusters were limited for other reasons. The efficient *in vitro* reconstitution of *Os*NifH^{Ht} protein indicated that future research should focus on better Fe transport, [Fe-S] cluster biosynthesis, delivery to Nif proteins and protection. Despite high mRNA expression levels of NifU^{Av} and NifS^{Av} was observed, the detection of protein accumulation for *Os*NifU^{Av} and *Os*NifS^{Av} in rice callus and regenerated plants was not achieved. Alternatively, [Fe-S] cluster delivery to Nif proteins

could be enhanced by manipulating the host Fe homeostasis (using Fe fortified rice lines) and endogenous mitochondrial pathways.

3.6 Conclusions

I have demonstrated the expression of *NifH*^{Av}, *NifH*^{Ht}, *NifM*^{Av}, *NifS*^{Av} and *NifU*^{Av} at the mRNA and/or protein levels in stably transformed rice callus and regenerated plants. In collaboration with colleagues at CBGP (Madrid, Spain) I purified the correctly processed mitochondria NifH^{Av} and NifH^{Ht} proteins from rice callus and plants. Rice-produced NifH^{Ht} was capable of electron transfer to NifDK, confirming the incorporation of endogenous rice mitochondrial [4Fe-4S] clusters, albeit at low levels. The formation of FeMo-co is one of the most important outcomes of this study, confirming the functionality of rice-produced NifH^{Ht} and its compatibility with the other proteins essential for nitrogenase biogenesis (NifB^{Mt} and NifDK^{Av}) despite their diverse origins. Active Fe protein expressed in and purified from rice, a major staple food security crop, represents a critical step toward the expression of a complete functional Nif complex as required to achieve BNF in cereals.

3.7 References

Aznar-Moreno JA, Jiang X, Burén S, Rubio LM. 2021. Analysis of nitrogenase Fe protein activity in transplastomic tobacco. *Front. Agron.* **3**: 657227.

Balk J, Lobréaux S. 2005. Biogenesis of iron–sulfur proteins in plants. *Trends Plant Sci.* **10**: 324-331.

Baysal C, Pérez-González A, Eseverri Á, Jiang X, Medina V, Caro E, Rubio L, Christou P, Zhu C. 2020. Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Res.* **29**: 37-52.

Bulen W, LeComte J. 1966. The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **56**: 979-986.

Burén S, Jiménez-Vicente E, Echavarri-Erasun C, Rubio LM. 2020. Biosynthesis of Nitrogenase Cofactors. *Chem. Rev.* **120**: 4921-4968.

Burén S, López-Torrejón G, Rubio LM. 2018. Extreme bioengineering to meet the nitrogen challenge. *Proc. Natl. Acad. Sci. U. S. A.* **115**: 8849-8851.

Burén S, Pratt K, Jiang X, Guo Y, Jimenez-Vicente E, Echavarri-Erasun C, Dean DR, Saaem I, Gordon DB, Voigt CA. 2019. Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 25078-25086.

Burén S, Young EM, Sweeny EA, Lopez-Torrejón G, Veldhuizen M, Voigt CA, Rubio LM. 2017. Formation of Nitrogenase NifDK Tetramers in the Mitochondria of Saccharomyces cerevisiae. *ACS Synth. Biol.* **6**: 1043-1055.

Burgess BK, Lowe DJ. 1996. Mechanism of molybdenum nitrogenase. *Chem Rev* **96**: 2983-3012.

Cornejo M-J, Luth D, Blankenship KM, Anderson OD, Blechl AE. 1993. Activity of a maize ubiquitin promoter in transgenic rice. *Plant Mol. Biol.* **23**: 567-581.

Creissen GP, Mullineaux PM. 1995. Cloning and characterisation of glutathione reductase cDNAs and identification of two genes encoding the tobacco enzyme. *Planta* **197**: 422-425.

Curatti L, Rubio LM. 2014. Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Science* **225**: 130-137.

Danyal K, Mayweather D, Dean DR, Seefeldt LC, Hoffman BM. 2010. Conformational Gating of Electron Transfer from the Nitrogenase Fe Protein to MoFe Protein. *J. Am. Chem. Soc*. **132**: 6894-6895.

Davarinejad H. 2015. Quantifications of western blots with ImageJ. *University of York*.

Eseverri Á, López-Torrejón G, Jiang X, Burén S, Rubio LM, Caro E. 2020. Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnol. J.* **18**: 1882-1896.

Fay AW, Blank MA, Yoshizawa JM, Lee CC, Wiig JA, Hu Y, Hodgson KO, Hedman B, Ribbe MW. 2010. Formation of a homocitrate-free iron-molybdenum cluster on NifEN: implications for the role of homocitrate in nitrogenase assembly. *Dalton Trans.* **39**: 3124-3130.

Fox AR, Soto G, Valverde C, Russo D, Lagares Jr A, Zorreguieta Á, Alleva K, Pascuan C, Frare R, Mercado-Blanco J. 2016. Major cereal crops benefit from biological nitrogen fixation when inoculated with the nitrogen-fixing bacterium Pseudomonas protegens Pf-5 X940. *Environ. Microbiol.* **18**: 3522-3534.

Gavini N, Tungtur S, Pulakat L. 2006. Peptidyl-prolyl cis/trans isomerase-independent functional NifH mutant of Azotobacter vinelandii. *J. Bacteriol* **188**: 6020-6025.

Georgiadis MM, Komiya H, Chakrabarti P, Woo D, Kornuc JJ, Rees DC. 1992. Crystallographic Structure of the Nitrogenase Iron Protein from *Azotobacter vinelandii*. *Science* **257**: 1653-1659.

Hu Y, Corbett Mary C, Fay Aaron W, Webber Jerome A, Hodgson Keith O, Hedman B, Ribbe Markus W. 2006. Nitrogenase Fe protein: A molybdate/homocitrate insertase. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 17125-17130.

Ivleva NB, Groat J, Staub JM, Stephens M. 2016. Expression of active subunit of nitrogenase via integration into plant organelle genome. *PLoS One* **11**: e0160951.

Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR. 1989. Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii. *Mo.l Genet.* **219**: 49-57.

Jasniewski AJ, Sickerman NS, Hu Y, Ribbe MW. 2018. The Fe protein: An unsung hero of nitrogenase. *Inorganics* **6**: 25.

Jiang X, Payá-Tormo L, Coroian D, García-Rubio I, Castellanos-Rueda R, Eseverri Á, López-Torrejón G, Burén S, Rubio LM. 2021. Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N2-fixation in plants. *Commun. Biol.* **4**: 4.

López-Torrejón G, Jiménez-Vicente E, Buesa JM, Hernandez JA, Verma HK, Rubio LM. 2016. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat. Commun.* **7**: 11426.

Mueller ND, West PC, Gerber JS, MacDonald GK, Polasky S, Foley JA. 2014. A tradeoff frontier for global nitrogen use and cereal production. *Environmental Research Letters* **9**: 054002.

Mus F, Alleman AB, Pence N, Seefeldt LC, Peters JW. 2018. Exploring the alternatives of biological nitrogen fixation. *Metallomics* **10**: 523-538.

Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GE, Poole PS. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl. Environ. Microbiol.* **82**: 3698-3710.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**: 292-292.

Poza-Carrión C, Jiménez-Vicente E, Navarro-Rodríguez M, Echavarri-Erasun C, Rubio LM. 2014. Kinetics of nif gene expression in a nitrogen-fixing bacterium. *J. Bacteriol* **196**: 595-603.

Rangaraj P, Ludden PW. 2002. Accumulation of 99Mo-containing Iron-Molybdenum Cofactor Precursors of Nitrogenase on NifNE, NifH, and NifX of *Azotobacter vinelandi*i. *J. Biol. Chem.* **277**: 40106-40111.

Ribbe MW, Hu Y, Guo M, Schmid B, Burgess BK. 2002. The FeMoco-deficient MoFe protein produced by a nifHDeletion strain of Azotobacter vinelandii shows unusual P-cluster features. *J. Biol. Chem.* **277**: 23469-23476.

Ribbe MW, Hu Y, Hodgson KO, Hedman B. 2014. Biosynthesis of Nitrogenase Metalloclusters. *Chem. Rev.* **114**: 4063-4080.

Roberts GP, MAcNEIL T, MAcNEIL D, Brill WJ. 1978. Regulation and characterization of protein products coded by the nif (nitrogen fixation) genes of Klebsiella pneumoniae. *J. Bacteriol* **136**: 267-279.

Rubio LM, Ludden PW. 2005. Maturation of Nitrogenase: a Biochemical Puzzle. *J. Bacteriol* **187**: 405-414.

Ryu M-H, Zhang J, Toth T, Khokhani D, Geddes BA, Mus F, Garcia-Costas A, Peters JW, Poole PS, Ané J-M. 2020. Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* **5**: 314-330.

Smith AD, Jameson GN, Dos Santos PC, Agar JN, Naik S, Krebs C, Frazzon J, Dean DR, Huynh BH, Johnson MK. 2005. NifS-mediated assembly of [4Fe-4S] clusters in the N- and C-terminal domains of the NifU scaffold protein. *Biochem.* **44**: 12955-12969.

Stoltzfus J, So R, Malarvithi P, Ladha J, De Bruijn F. 1997. Isolation of endophytic bacteria from rice and assessment of their potential for supplying rice with biologically fixed nitrogen. *Plant Soil* **194**: 25-36.

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P. 1998. An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. *Transgenic Res.* **7**: 289-294.

Tal S, Chun TW, Gavini N, Burgess BK. 1991. The delta nifB (or delta nifE) FeMo cofactor-deficient MoFe protein is different from the delta nifH protein. *J. Biol. Chem.* **266**: 10654-10657.

Tran A-M, Nguyen T-T, Nguyen C-T, Huynh-Thi X-M, Nguyen C-T, Trinh M-T, Tran L-T, Cartwright SP, Bill RM, Tran-Van H. 2017. Pichia pastoris versus Saccharomyces cerevisiae: a case study on the recombinant production of human granulocyte-macrophage colony-stimulating factor. *BMC Res. Notes* **10**: 148.

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P. 1998. Transgenic Central American, West African and Asian Elite rice varieties resulting from particle bombardment of foreign DNA into mature seed-derived explants utilizing three different bombardment devices. *Ann. Bot.* **82**: 795-801.

Van Deynze A, Zamora P, Delaux P-M, Heitmann C, Jayaraman D, Rajasekar S, Graham D, Maeda J, Gibson D, Schwartz KD. 2018. Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. *PLoS Biol.* **16**: e2006352.

Yang J, Xie X, Yang M, Dixon R, Wang Y-P. 2017. Modular electron-transport chains from eukaryotic organelles function to support nitrogenase activity. *Proc. Natl. Acad. Sci. U.S.A.* **114**: E2460-E2465.

Yang Z-Y, Ledbetter R, Shaw S, Pence N, Tokmina-Lukaszewska M, Eilers B, Guo Q, Pokhrel N, Cash VL, Dean DR. 2016. Evidence that the Pi release event is the rate-limiting step in the nitrogenase catalytic cycle. *Biochem.* **55**: 3625-3635.

Chapter IV.
Toward engineering of the nitrogenase component NifD in rice

4.0 Abstract

The expression of nitrogenase subunits is a necessary requirement to engineer biological nitrogen fixation in plants. Mitochondria minimize recombinant protein exposure to oxygen and have thus been considered an appropriate subcellular organelle for the biosynthesis of the oxygen-sensitive nitrogenase and its components. *NifD* and *NifK* encode the α and β subunits of nitrogenase, respectively. The expression of NifD has been difficult in heterologous eukaryotic expression systems compared to NifK, which was expressed in *Nicotiana benthamiana* at detectable levels. The degradation of NifD when it was imported into *N. benthamiana* and *Saccharomyces cerevisiae* mitochondria is a bottleneck for the expression of functional nitrogenase in eukaryotes. Following a strategy that was reported for expressing NifD, resistant to mitochondrial degradation, in *N. benthamiana*, I assessed three *Azotobacter vinelandii* NifD variants ($NifD^{Av}$-99K, $NifD^{Av}$-99Q, $NifD^{Av}$-99Q100T) with one or two amino acid substitutions for their ability to resist mitochondrial degradation. I introduced the variants separately into rice callus to assess the stability of the corresponding recombinant proteins. Expression of the three variants was confirmed at the transcriptional level, and no significant differences in mRNA accumulation were observed among them. However, all variants only accumulated as truncated proteins. The instability of the three variants in rice is not consistent with the behaviour reported in yeast or tobacco, suggesting differences in the processing of the protein in mitochondria of different species. Further work is therefore required to understand the reasons and underlying mechanism of NifD instability in rice.

4.1 Introduction

4.1.1 NifD protein and its role in nitrogenase function

NifD encodes the alpha subunits of molybdenum–iron (MoFe) protein (Orme-Johnson 1985). The MoFe protein is a tetramer ($\alpha_2\beta_2$) composed of two identical halves, each containing an α- and a β-subunit (encoded by NifK). In *A. vinelandii*, NifD and NifK are synthesized separately in equimolar ratios and subsequently assembled as a tetramer (Poza-Carrión et al. 2014). The MoFe protein is loaded with two metal clusters: the FeMo cofactor (FeMo-co) [7Fe-9S-C-Mo-R homocitrate] and the P-cluster [8Fe-7S] (Burén et al. 2020). Both clusters are essential for nitrogenase function. During catalysis, the P-cluster transfers electrons from a [4Fe4S] cluster within the Fe protein to FeMo-co (Burgess and Lowe 1996; Peters et al. 1997). The transferred electrons then activate the FeMo-co cluster to reduce N_2 to ammonia (Hawkes and Smith 1983). Functionality of an

inactive MoFe protein in a *NifD* deletion mutant was restored by providing a NifD protein extract from *A. vinelandii*, thus confirming that NifD is essential for nitrogenase function (Robinson et al. 1986)

NifD contains 491 amino acids (in *A. vinelandii*) with a molecular weight of ~55 kDa (Georgiadis et al. 1992) and provides specific histidine and cysteine residues (α-Cys^{275} and α-His^{442} in *A. vinelandii* NifD) that covalently attach to the catalytic cofactor FeMo-co (**Figure 4.1**). The latter is responsible for binding and reducing N_2 (Hu et al. 2008)The α-His^{274} maintains the coordination of FeMo-co with MoFe proteins and helps correct FeMo-co localization to bind and reduce N_2 (Kim et al. 1995). Additionally, α-Lys^{315}, α-Lys^{426}, α-Arg^{96}, α-Arg^{97}, α-Arg^{277}, α-Arg^{359}, α-Arg^{361}, α-His^{274}, α-His^{362}, α-His^{442}, and α-His^{451} help FeMo-co to attach correctly to MoFe protein (Dos Santos et al. 2004). α-His^{274}, α-His^{362} and α-His^{451} have been shown to be specifically involved in FeMo-co insertion. Mutations of these histidine residues result in substantially reduced FeMo-co accumulation in MoFe protein (Fay et al. 2007; Hu et al. 2007). Moreover, three cysteine residues contributed by NifD namely α-Cys^{62}, α-Cys^{88}, andα-Cys^{154} (in *A. vinelandii* NifD) co-bridge the P-cluster with the other three cysteine residues from NifK (β-Cys^{70}, β-Cys^{95}, and β-Cys^{153} in the *A. vinelandii* NifK) at the NifD and NifK interface (**Figure 4.1**) (Hu et al. 2008).

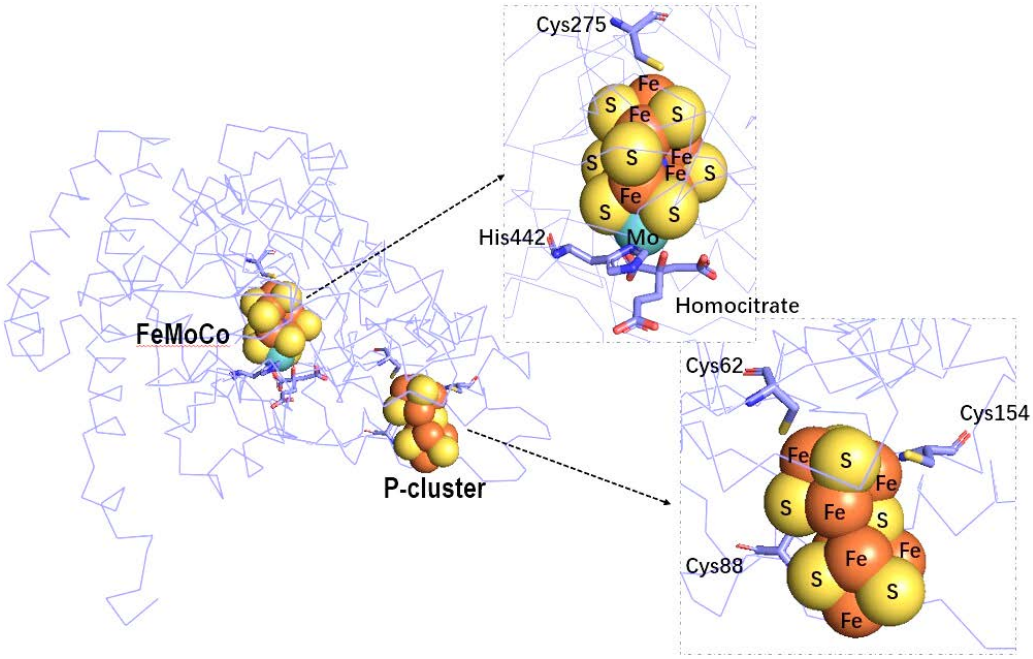

Figure 4.1 Ribbon diagram of A. vinelandii NifD subunit drawn from 1M1N (PDB, https://www.rcsb.org). The metal clusters FeMo-co and P-cluster are shown as spheres and the key residues are represented by sticks. Residues Cys^{275} and His^{442} are covalently attached to FeMo-co and residues Cys^{62}, Cys^{88}, and Cys^{154} are covalently attached to P-cluster. Atoms are

colored as follows: orange, iron; yellow, sulfur; cyan, molybdenum; slate, carbon; blue, nitrogen; red, oxygen.

4.1.2 Heterologous expression of NifD in *S. cerevisiae* and *N. benthamiana*

Expression of NifD in eukaryotes has been reported in *S. cerevisiae* and *N. benthamiana* (transiently). To minimize the oxygen damage, NifD was imported into the mitochondria using targeting peptides in both systems. However, the resulting instability of the recombinant protein during import into mitochondria represents a bottleneck in obtaining functional MoFe protein in eukaryotic organelles.

NifD was shown to be the most challenging protein to express among 16 *Klebsiella pneumoniae* Nif proteins (NifB, D, E, F, H, J, K, M, N, Q, S, U, V, X, Y, and Z) tested (transiently) in *N. benthamiana* mitochondria (Allen et al. 2017). Accumulation of *K. pneumoniae* NifD protein could not be detected in *N. benthamiana*, although high levels of NifD RNA accumulation were observed, in addition to a detectable *K. pneumoniae* NifD signal at the expected size when the same genetic construct was introduced to *E. coli* (Allen et al. 2017). The *K. pneumoniae* NifD-K fusion protein resulted in increased NifD accumulation. However, most of the *recombinant protein* accumulated in an unprocessed form (Allen et al. 2017). Similarly, when *A. vinelandii* NifD and NifK were expressed in *S. cerevisiae* the resulting *recombinant* NifDK complex was not identical to the native *A. vinelandii* as shown in an experiment in which the proteins were separated on anoxic native gels. The unexpected size of yeast-produced NifDK was shown to be caused by incorrect accumulation of NifD, because SDS-PAGE gels showed that the yeast-produced NifD was smaller (~ 50kDa) compared to the *A. vinelandii* NifD (~55 kDa) (Burén et al. 2017b). Production of the truncated protein was later shown to be the result of secondary cleavage of NifD upon import into mitochondria (Allen et al. 2020; Xiang et al. 2020). The highly conserved sequence RRNYYT (R, N, Y and T: Arg, Asn, Tyr and Thr, respectively) within NifD was predicted as a mitochondrial processing peptidase (MPP) cleavage site ($R^{96}R^{97}N^{98}\downarrow Y^{99}Y^{100}T^{101}$) (↓ indicates the cleavage site in *A. vinelandii*) that is responsible for the degradation of the NifD protein (Allen et al. 2020; Xiang et al. 2020). The completely conserved Arg^{97} and Arg^{98} (in *Klebsiella oxytoca* NifD) are in proximity to FeMo-co. These two positively charged Arg residues help to steer negatively charged FeMo-co to His^{443} (in *K. oxytoca* NifD) for docking and they are required for enzymatic activity. Variants, resistant to MPP, were created by replacing the residues Tyr^{100} and/or Tyr^{101} (Allen et al. 2020). The substitutions at the Tyr^{100} residue (Y100Q and Y100K) and at Tyr^{100} and Tyr^{101} (Y100KY100T) did not produce the

secondary cleavage products in transiently expressed *N. benthamiana* mitochondria and these NifD mutations retained functionality similar to wild-type levels produced from a nitrogen-fixing *E. coli* strain (Allen et al. 2020).

In this chapter, I describe the generation of engineered rice plants expressing three *A. vinelandii* NifD variants (NifD^{Av}-Y99K, NifD^{Av}-Y99Q, NifD^{Av}-Y99QY100T) targeted to the mitochondria. The three *NifD*^{Av} variants were integrated into the rice genome and accumulated mRNA at similar levels. However, all three NifD^{Av} variants were still susceptible to cleavage by mitochondrial processing peptidase, resulting in the formation of only truncated proteins.

4.2 Aims and objectives

The aim of the work described in this chapter was to express the nitrogenase component NifD protein in rice using a transgenic callus system, before embarking on experiments at the whole plant level. The specific objectives were to: 1) introduce the *A. vinelandii NifD* (three *A. vinelandii NifD* variants, separately) into rice; 2) confirm integration of the *A. vinelandii NifD* gene in transgenic callus; 3) measure *NifD* mRNA accumulation in transgenic callus; 4) determine the integrity of NifD protein in transgenic callus.

4.3 Materials and methods

4.3.1 Transformation constructs

The sequence of *NifD*^{Av} variants (*NifD*^{Av}-Y99K, *NifD*^{Av}-Y99Q, *NifD*^{Av}-Y99QY100T), the *S. cerevisiae* cytochrome c oxidase subunit IV (Cox4) mitochondrial leader sequence were codon-optimized for *S. cerevisiae* using the GeneOptimizer tool (Thermo Fisher Scientific, Waltham, MA, USA), and synthesized by Thermo Fisher Scientific. Constructs pAvNifD-Y99K (*ZmUbi1*+1sti-Cox4-TS-*NifD*^{Av}-Y99K-Nos), pAvNifD-99YQ (*ZmUbi1* +1sti-Cox4-TS-*NifD*^{Av}-Y99Q-Nos), and pAvNifD-Y99QY100T (*ZmUbi1*+1sti-Cox4-TS-*NifD*^{Av}-Y99QY100T -Nos) were generated with the MoClo cloning system in the pUC57 (GenScript Biotech, Piscataway, NJ, USA) vector backbone by Dr Á. Eseverri Sabaté at the Centre for Plant Biotechnology and Genomics (CBGP, Madrid, Spain). The sequences of *S. cerevisiae* codon-optimized DNA constructs for expression in rice are listed in **Table 4.1.**

Table 4.1 Sequences of genetic constructs.

Genetic elements	Nucleotide sequence Amino acid sequence
Cox4	ATG CTT TCA CTT AGA CAA TCT ATT AGA TTT TTC **M L S L R Q S I R F F** AAG CCA GCT ACA AGA ACT TTG TGT TCT TCT AGA **K P A T R T L C S S R** TAT CTT CTT CAG CAA AAA CCT **Y L L Q Q K P**
NifD^{Av}	ATG TCC CGC GAA GAA GTT GAA AGT TTG ATA CAA **M S R E E V E S L I Q** GAA GTC TTG GAA GTT TAT CCA GAA AAA GCT AGA **E V L E V Y P E K A R** AAG GAC AGA AAT AAG CAT TTG GCA GTA AAC GAT **K D R N K H L A V N D** CCT GCC GTC ACC CAA TCT AAA AAG TGT ATC ATC **P A V T Q S K K C I I** TCT AAC AAA AAG TCA CAA CCA GGT TTA ATG ACT **S N K K S Q P G L M T** ATT AGA GGT TGC GCC TAT GCT GGT TCA AAA GGT **I R G C A Y A G S K G** GTT GTA TGG GGT CCA ATT AAG GAT ATG ATA CAT **V V W G P I K D M I H** ATC TCC CAC GGT CCT GTT GGT TGT GGT CAA TAC **I S H G P V G C G Q Y** AGT AGA GCA GGT AGA AGA AAC TAC TAC ATA GGT **S R A G R R N Y Y I G**
NifD^{Av}-Y99K	AAA TAC **K Y**
NifD^{Av}-Y99Q	CAA TAC **Q Y**
NifD^{Av}-Y99QY100T	CAA ACT **Q T**
	ACT ACA GGT GTT AAT GCC TTC GTA ACT ATG AAC **T T G V N A F V T M N** TTC ACA TCA GAT TTC CAA GAA AAG GAC ATC GTT **F T S D F Q E K D I V** TTT GGT GGT GAC AAA AAG TTG GCT AAG TTG ATC **F G G D K K L A K L I** GAC GAA GTA GAA ACA TTG TTC CCA TTG AAC AAA **D E V E T L F P L N K** GGT ATC TCC GTC CAA AGT GAA TGC CCT ATT GGT **G I S V Q S E C P I G** TTG ATA GGT GAC GAC ATT GAA TCC GTA AGT AAA **L I G D D I E S V S K** GTC AAG GGT GCA GAA TTA TCC AAG ACC ATA GTT **V K G A E L S K T I V**

CCA	GTA	AGA	TGT	GAA	GGT	TTC	AGA	GGT	GTT	TCC
P	**V**	**R**	**C**	**E**	**G**	**F**	**R**	**G**	**V**	**S**
CAA	AGT	TTG	GGT	CAT	CAC	ATT	GCA	AAT	GAT	GCC
Q	**S**	**L**	**G**	**H**	**H**	**I**	**A**	**N**	**D**	**A**
GTC	AGA	GAC	TGG	GTT	TTA	GGT	AAA	AGA	GAT	GAA
V	**R**	**D**	**W**	**V**	**L**	**G**	**K**	**R**	**D**	**E**
GAT	ACC	ACT	TTT	GCA	TCT	ACT	CCT	TAT	GAT	GTT
D	**T**	**T**	**F**	**A**	**S**	**T**	**P**	**Y**	**D**	**V**
GCC	ATC	ATT	GGT	GAC	TAC	AAC	ATT	GGT	GGT	GAC
A	**I**	**I**	**G**	**D**	**Y**	**N**	**I**	**G**	**G**	**D**
GCT	TGG	TCG	TCA	AGA	ATC	TTG	TTG	GAA	GAA	ATG
A	**W**	**S**	**S**	**R**	**I**	**L**	**L**	**E**	**E**	**M**
GGT	TTG	AGA	TGT	GTT	GCT	CAA	TGG	TCT	GGT	GAC
G	**L**	**R**	**C**	**V**	**A**	**Q**	**W**	**S**	**G**	**D**
GGT	TCC	ATC	TCA	GAA	ATA	GAA	TTG	ACA	CCA	AAG
G	**S**	**I**	**S**	**E**	**I**	**E**	**L**	**T**	**P**	**K**
GTT	AAG	TTG	AAT	TTG	GTA	CAT	TGC	TAC	AGA	TCT
V	**K**	**L**	**N**	**L**	**V**	**H**	**C**	**Y**	**R**	**S**
ATG	AAC	TAC	ATC	TCA	AGA	CAC	ATG	GAA	GAA	AAG
M	**N**	**Y**	**I**	**S**	**R**	**H**	**M**	**E**	**E**	**K**
TAC	GGT	ATC	CCA	TGG	ATG	GAA	TAC	AAT	TTC	TTT
Y	**G**	**I**	**P**	**W**	**M**	**E**	**Y**	**N**	**F**	**F**
GGT	CCT	ACA	AAA	ACC	ATT	GAA	TCT	TTG	AGA	GCC
G	**P**	**T**	**K**	**T**	**I**	**E**	**S**	**L**	**R**	**A**
ATA	GCT	GCA	AAG	TTC	GAT	GAA	TCA	ATC	CAA	AAG
I	**A**	**A**	**K**	**F**	**D**	**E**	**S**	**I**	**Q**	**K**
AAA	TGT	GAA	GAA	GTT	ATC	GCA	AAG	TAT	AAG	CCA
K	**C**	**E**	**E**	**V**	**I**	**A**	**K**	**Y**	**K**	**P**
GAA	TGG	GAA	GCT	GTC	GTT	GCA	AAA	TAC	AGA	CCT
E	**W**	**E**	**A**	**V**	**V**	**A**	**K**	**Y**	**R**	**P**
AGA	TTG	GAA	GGT	AAA	AGA	GTT	ATG	TTG	TAC	ATC
R	**L**	**E**	**G**	**K**	**R**	**V**	**M**	**L**	**Y**	**I**
GGT	GGT	TTA	AGA	CCT	AGA	CAT	GTA	ATT	GGT	GCT
G	**G**	**L**	**R**	**P**	**R**	**H**	**V**	**I**	**G**	**A**
TAC	GAA	GAT	TTG	GGT	ATG	GAA	GTA	GTC	GGT	ACT
Y	**E**	**D**	**L**	**G**	**M**	**E**	**V**	**V**	**G**	**T**
GGT	TAT	GAA	TTC	GCA	CAT	AAT	GAT	GAC	TAC	GAC
G	**Y**	**E**	**F**	**A**	**H**	**N**	**D**	**D**	**Y**	**D**
AGA	ACA	ATG	AAA	GAA	ATG	GGT	GAC	TCC	ACC	TTG
R	**T**	**M**	**K**	**E**	**M**	**G**	**D**	**S**	**T**	**L**
TTA	TAC	GAT	GAC	GTC	ACT	GGT	TAC	GAA	TTC	GAA
L	**Y**	**D**	**D**	**V**	**T**	**G**	**Y**	**E**	**F**	**E**
GAA	TTC	GTT	AAG	AGA	ATT	AAG	CCA	GAT	TTG	ATC
E	**F**	**V**	**K**	**R**	**I**	**K**	**P**	**D**	**L**	**I**
GGT	TCT	GGT	ATC	AAA	GAA	AAG	TTT	ATC	TTC	CAA
G	**S**	**G**	**I**	**K**	**E**	**K**	**F**	**I**	**F**	**Q**
AAG	ATG	GGT	ATT	CCT	TTT	AGA	CAA	ATG	CAT	TCC

K	**M**	**G**	**I**	**P**	**F**	**R**	**Q**	**M**	**H**	**S**
TGG	GAT	TAT	AGT	GGT	CCT	TAC	CAC	GGT	TTT	GAT
W	**D**	**Y**	**S**	**G**	**P**	**Y**	**H**	**G**	**F**	**D**
GGT	TTC	GCC	ATT	TTT	GCT	AGA	GAT	ATG	GAC	ATG
G	**F**	**A**	**I**	**F**	**A**	**R**	**D**	**M**	**D**	**M**
ACT	TTG	AAT	AAC	CCT	TGT	TGG	AAA	AAG	TTA	CAA
T	**L**	**N**	**N**	**P**	**C**	**W**	**K**	**K**	**L**	**Q**
GCC	CCT	TGG	GAA	GCC	AGT	GAA	GGT	GCC	GAA	AAA
A	**P**	**W**	**E**	**A**	**S**	**E**	**G**	**A**	**E**	**K**
GTT	GCC	GCA	TCC	GCC	TGA					
V	**A**	**A**	**S**	**A**	**-**					

Cox4, mitochondrial targeting peptide of *S. cerevisiae* cytochrome c oxidase subunit IV; $NifD^{Av}$, *A. vinelandii* NifD; The underlined sequence is the variant sequence sites. The mutant sites of NifD-Y99K, NifD-Y99Q and NifD-Y99QY100T are indicated in red, blue, and green, respectively.

4.3.2 Transformation of rice explants and recovery of transgenic callus

Seven-day-old mature rice embryos (*Oryza sativa* cv. Nipponbare) were used as explants for transformation using particle bombardment (Sudhakar et al. 1998; Valdez et al. 1998). The 0.9μm gold particles were coated separately with plasmids pAvNifD-Y99K, pAvNifD-Y99Q, or pAvNifD-Y99QY100T, and the hygromycin phosphotransferase (*Hpt*) selectable marker in binary combinations at a molar ratio of 3:1. Selection of transgenic rice callus was performed as described in Chapter II.

4.3.3 DNA extraction and PCR analysis

Genomic DNA was extracted from rice callus as described by Creissen and Mullineaux (1995). Callus (0.1-0.2 g) were ground in liquid nitrogen and the powder was transferred into 600 μL DNA extraction buffer (500 mM NaCl, 100 mM Tris–HCl pH 8 and 50 mM EDTA pH 8) and 75 μL 20% (w/v) sodium dodecyl sulfate (SDS). The mixture was homogenized by vortexing for 10 min at room temperature and then incubated at 65 °C for 15 min. Phenol-chloroform-isoamyl alcohol (25:24:1, by volume) was then added to the above mixture (1:1, v/v), and agitated well by inversion. Centrifugation at 13000 rpm for 5 min was done to separate the phases. The supernatant was transferred into an Eppendorf tube containing 6 μL RNAase (10 mg mL^{-1}) and incubated for 30 min at 37 °C. Phenol extraction was repeated one more time. The supernatant was transferred into a new Eppendorf tube and an equal volume of isopropanol was added to precipitate the DNA. The supernatant was removed after centrifugation at 13000 rpm at 4°C for 15 min. The pellet was washed with 1 mL of 70% (v/v) ethanol and centrifuged again for 5 min.

The samples were dried at 37 °C to remove the ethanol and the pellet was resuspended in 100 µL Millipore water. DNA quality and yield were determinedwith a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific).

The PCR reaction was carried out in a total volume of 20 µl and included 1 µl of extracted DNA (500 ng μl^{-1}). Amplification was performed on Doppio Gradient 2 × 48 well thermocycler (VWR®, Radnor, PA, USA) with an initial denaturation at 95 ºC/5 min, followed by 35 cycles at 95 °C /30 s denaturation, 56°C/30 s annealing, extension at 72 ºC/1 min 15 s; and final extension at 72 °C/10 min. The three *NifD*^{Av} variants were detected by using forward primer ZmUbi F: 5'-AGCCCTGCCTTCATACGC-3' and reverse primer NifD R: 5'-TGGCGAAACCATCAAAACC-3'. The predicted amplification product size was 1649 bp. The PCR products were resolved using agarose gel (1%) electrophoresis with 1 × Tris-acetate buffer (TAE) and visualized on Gel Doc XR (Bio-Rad Laboratories, Hercules, CA, USA). The positive control and blank reactions were performed by replacing the genomic DNA with plasmid and Millipore water. Genomic DNA from a line transformed with *Hpt* was used as a negative control. Positive, negative, and blank controls were included in each batch for all tests.

4.3.4 Gene expression analysis by real-time quantitative reverse transcription PCR

Total RNA extraction, cDNA synthesis and real-time quantitative reverse transcription PCR (RT-qPCR) were carried out as described in Chapter II. The gene-specific primers used for RT-qPCR analysis are listed in **Table 4.2**. The PCR products were confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA.

PCR (RT-qPCR) were carried out as described in Chapter 2. The gene-specific primers used for RT-qPCR analysis are listed in **Table 4.2**. The PCR products were confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA.

Table 4.2 Primers used for real-time quantitative reverse transcription PCR analysis.

Gene	Primer sequence (5′→3′)	Product size (bp)
Actin^{Os}	F: TCATGTCCCTCACAATTTCC R: GACTCTGGTGATGGTGTCAGC	181
NifD^{Av}	F: ATGGGTGACTCCACCTTGTT R: TGGCGAAACCATCAAAACC	193

Actin^{Os}*, O. sativa Actin; NifD*^{Av}*, A. vinelandii NifD.*

4.3.5 Protein extraction and immunoblot analysis

Protein extraction and immunoblot were performed as described at Chapter II. After confirming equal loading of total protein extracts on nitrocellulose membranes, the membranes were incubated with primary polyclonal rabbit antibodies against *A. vinelandii* NifD (generated at the Centre for Plant Biotechnology and Genomics, CBGP) with a1: 2,000 dilution in TBS-T supplemented with 5% bovine serum albumin (BSA). Secondary antibodies (Thermo Fisher Scientific) were diluted 1: 20,000 in TBS-T which was supplemented with 2% non-fat milk and incubated for 2 h at room temperature. Membranes were developed using BCIP/NBT (Thermo Fisher Scientific) as substrate.

4.3.6 Statistical analysis

The standard deviation (SD) of relative mRNA expression data was calculated based on three technical replicates. The difference in mRNA expression between the three *NifD*^{Av} variants was analysis by one-way analysis of variance (ANOVA). Significant difference was determined using Fisher's least significant difference (LSD) test ($p < 0.05$).

4.4 Results

4.4.1 Expression constructs for rice transformation

Based on the native *O. sativa* codon usage table, we determined that there were no rare codons in the *S. cerevisiae* codon-optimized *NifD*^{Av} variant (Y99K, Y99Q, and Y99QY100T) sequences. The codon-optimized *NifD*^{Av} variant sequences showed high relative adaptiveness (Nakamura et al. 2000). Therefore, the *S. cerevisiae* codon-optimized *NifD*^{Av} variant sequences were used for rice transformation directly. Expression of the *NifD*^{Av} variants was under the control of the *ZmUbi1*+1sti promoter and *Nos* terminator. The mitochondrial targeting peptide Cox4 was added to the N-terminal of the NifD proteins. Cox4 had been shown earlier, to correctly direct the nuclear-encoded enhanced green fluorescent protein (eGFP) to rice mitochondria (Baysal et al. 2020). An additional construct carrying the selectable marker *Hpt* was also used. Transgenic rice callus was transformed by direct DNA transfer as described earlier.

4.4.2 Presence of NifD^{Av} in transgenic callus

The presence of *NifD*^{Av} in the transgenic rice callus was confirmed by the production of a PCR product with a size of 1649 bp (**Figure 4.2**). This fragment corresponds to the end

of the *ZmUbi1*+1st i promoter and a segment of *NifD*^{Av}. No amplification product could be detected from genomic DNA extracted from the callus transformed with *Hpt*, alone. Among the transformants, 17 independent callus lines had the expected size band (6 lines from NifD^{Av}-Y99K **Figure 4.2A**;5 lines from NifD^{Av}-Y99Q **Figure 4.2B** and 6 lines from NifD^{Av}-Y99QY100T **Figure 4.2C**)

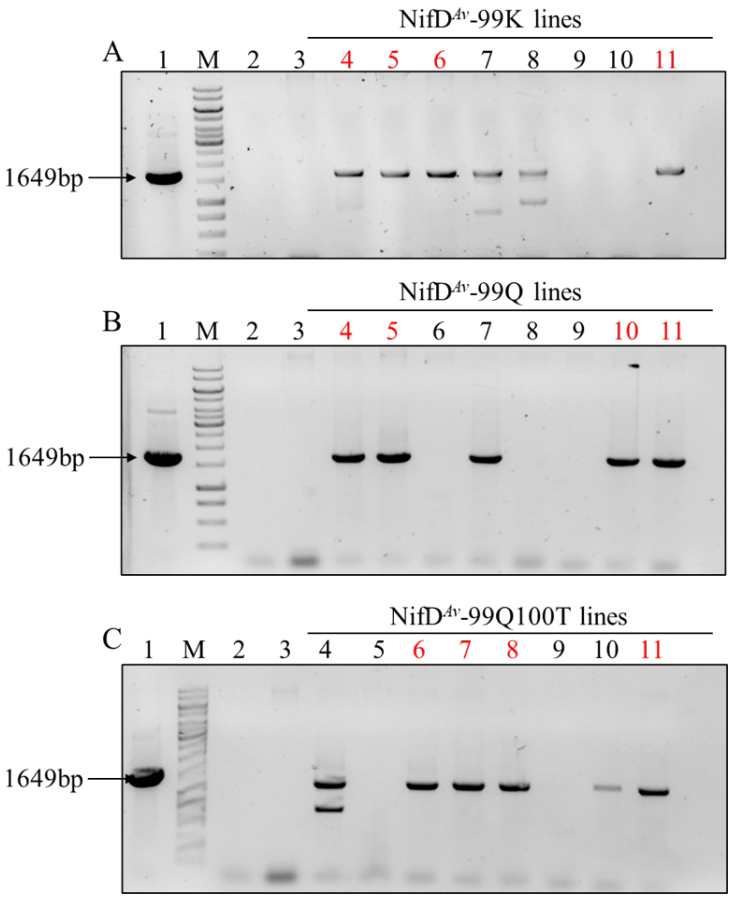

Figure 4.2 PCR amplification of *NifD*^{Av} from putative transgenic callus lines. The PCR product (1649 bp) is shown with an arrow. Lane 1, PCR amplification of pAvNifD-Y99K (A), NifD^{Av}-Y99Q (B) and NifD^{Av}-Y99QY100T (C) plasmid DNA, used for rice transformation; lane M, GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific); lane 2, PCR amplification of millipore water (blank); lane 3, PCR amplification of the genomic DNA extracted from a line transformed with *Hpt*; lanes 4-11, independent NifD^{Av}-Y99K (A), NifD^{Av}-Y99Q (B) and NifD^{Av}-Y99QY100T lines (C). The callus lines used for gene expression analysis are indicated in red.

4.4.3 Gene expression in transgenic callus

The expression of *NifD*^{Av} in callus was confirmed by RT-qPCR analysis with mRNA levels normalized against *OsActin* mRNA. Nine of the twelve independent lines analyzed accumulated *NifD*^{Av} mRNA (**Figure 4.3**). The highest mRNA accumulation was observed in line 5 NifD^{Av}-Y99Q which was at least four times higher than the other lines. There

was no significant difference ($p < 0.05$) in relative mRNA expression levels between the three variants overall.

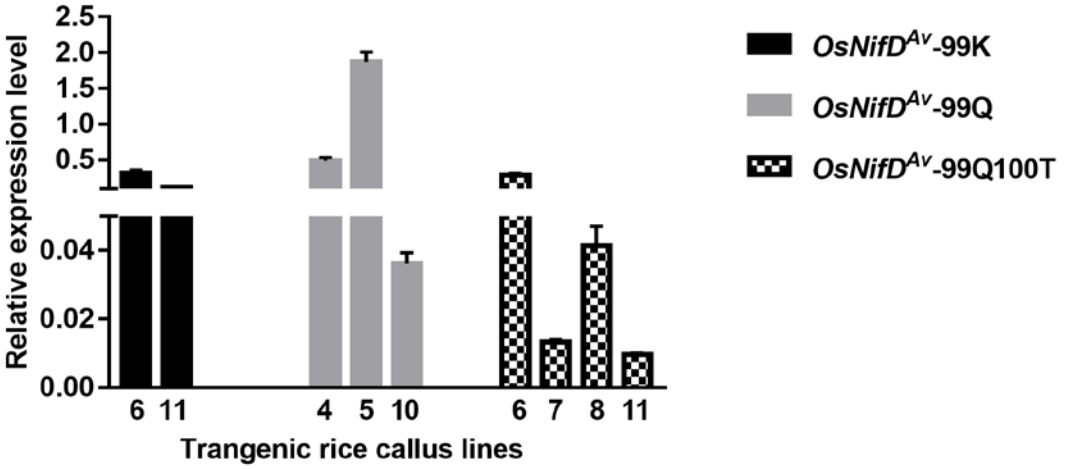

Figure 4.3 mRNA expression analysis of *NifD* in selected *Os*NifD^{Av}-Y99K, *Os*NifD^{Av}-Y99Q and *Os*NifD^{Av}-Y99QY100T callus lines. Relative mRNA expression levels (normalized to *OsActin* mRNA) are means ± SD (n = 3 technical replicates). *OsNifD*^{Av}-Y99K, *O. sativa*-derived *NifD*^{Av}-Y99K; *OsNifD*^{Av}-Y99Q, *O. sativa*-derived *NifD*^{Av}-Y99Q; *OsNifD*^{Av}-Y99KY100T, *O. sativa*-derived *NifD*^{Av}-Y99KY100T.

4.4.4 *Os*NifD^{Av} protein accumulation in transgenic callus

Accumulation of *Os*NifD^{Av} in callus was analyzed by immunoblot using antibodies specific for NifD^{Av}, (**Figure 4.4**). The expected size of NifD^{Av} is ~55 kDa. However, a non-specific band was also observed in the negative control lane at ~55 kDa (a line transformed with only *Hpt*), which resulted in strong interference and difficulties in confirming the accumulation of correctly processed *Os*NifD^{Av} variants. Line A1-1 **(Figure 4.4**) is a line which was generated in earlier experiments in our lab. In those experiments, the native *A. vinelandii NifH, M, D, K* genes were used for rice transformation. Line A1-1 is a represented line that contains all four input genes. However, two 45-48 kDa lower-molecular-mass products of *A. vinelandii* NifD were observed in line A1-1 and also in all the other independent lines derived from the earlier experiments (data not shown). I used Line A1-1 as a control to indicate the migration of the native *A. vinelandii* NifD protein targeted to rice mitochondria. The same two truncated proteins (45-48 kDa) were also present in the plants containing all *Os*NifD^{Av} variants I generated.

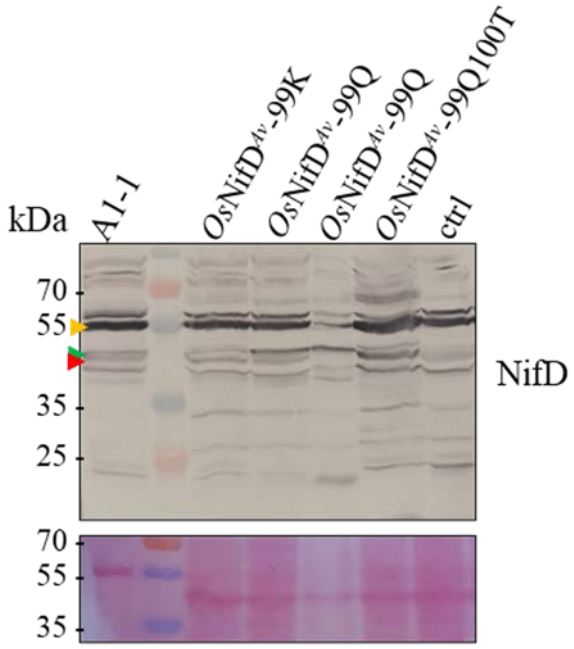

Figure 4.4 Accumulation of *Os*NifD^{Av} in rice callus. *Os*NifD^{Av}-Y99K, *Os*NifD^{Av}-Y99Q and *Os*NifD^{Av}-Y99QY100T were detected with antibodies against NifD^{Av}. A1-1 is an additional line which was generated in earlier experiments in our lab containing the native *A. vinelandii NifD*, shown to accumulate a truncated version of the recombinant protein. Ctrl lane is a line transformed only with *Hpt*. Ponceau staining was used as loading control. The yellow triangle indicates the expected size of correctly-processed NifD^{Av}; the green and red triangles indicate the size of the truncated *Os*NifD^{Av}.

4.5 Discussion

Biological nitrogen fixation is catalyzed by the nitrogenase enzyme complex, which is extremely sensitive to O_2 (Shah and Brill 1973). To limit oxygen damage to nitrogenase components, one strategy to engineer BNF in plants is to direct nitrogenase and its accessory proteins to mitochondria. (Curatti and Rubio 2014). Mitochondria were confirmed as a suitable subcellular compartment accumulation of functional NifH and NifB in aerobic growing yeast (López-Torrejón et al. 2016; Burén et al. 2017a). However, a remarkable degree of variability in protein abundance was observed among 16 *K. pneumoniae* nitrogenase proteins expressed transiently in *N. benthamiana* mitochondria (Allen et al. 2017). NifD was the most difficult protein to express in plant mitochondria with low abundance and instability, compared to NifB, NifH, NifK, NifS, and NifY which accumulated in much higher levels (Allen et al. 2017). In addition, incorrect processing of *A. vinelandii* NifD resulted in NifDK tetramers accumulating in an inactive form in yeast mitochondria (Burén et al. 2017b). Full-length NifD expressed in the cytoplasm confirmed that the observed degradation of NifD was the result of incorrect processing when NifD was imported into mitochondria (Allen et al. 2020).

The MPP plays an essential role in mitochondrial protein import. It is responsible for cleaving the N-terminal targeting signals of nuclear-encoded mitochondrial proteins upon their transport into the organelle (Gakh et al. 2002). Although the primary amino acid sequence in mitochondrial target peptides is not conserved, nearly all of them have an overall positive net charge and are all cleaved at a specific site by MPP (Abe et al. 2000). The cleavage sites of MPP in plants are grouped into three classes according to the position of arginine: R-2 motif (R-X↓X), R-3 motif (R-X-(Y/F/ L)↓(S/A/X)) and R-none motifs (X↓X-(S/X)) (where X indicates any amino acid, R, Y, F, S and A indicate Arg, Tyr, Phe and Ser, respectively and ↓ indicates the cleavage site) (Mossmann et al. 2012). The RRNYYT motif within NifD was predicted to be a cleavage site as $R^{97}R^{98}N^{99}\downarrow Y^{100}Y^{101}T^{102}$ (N and T indicate Asn and Thr, respectively, ↓ shows the cleavage site in *K. pneumoniae* and *K. oxytoca*) using the MitoFates prediction program, which can be considered as an R-2 motif (Allen et al. 2020; Xiang et al. 2020). Arg is found mostly at the second or third residue at the amino side of the cleavage recognition site (Mossmann et al. 2012). Between Arg^{97} and Arg^{98}, Arg^{98} was shown to be responsible for the degradation of *K. oxytoca* NifD in *S. cerevisiae* mitochondria, but variants with direct substitution of Arg^{98} (R98P, R98H and R98N) resulted in reduced nitrogenase activity (54.1 - 68.7%) compared to the activity of the wild-type *K. oxytoca* NifD (Xiang et al. 2020). However, the variants with substitution of Arg^{98} (R98A and R98K) were still prone to degradation in *N. benthamiana* (Allen et al. 2020). The degradation-resistant NifD variants in *N. benthamiana* were mostly with substitution of Tyr^{100} alone or together with Tyr^{101} and the variants Y100Q, Y100Q/Y101T and Y100K retained function similar to wild-type NifD (Allen et al. 2020).

In order to determine if results reported for *N behtamiana* using transient expression were more generally applicable to other pants, in this case rice, I generated transgenic rice callus containing the *A. vinelandii NifD* variants as a first step in assessing the stability and performance of NifD in differentiated tissue, prior to embarking on the more time consuming and labor intensive process of recovering intact plants expressing the variants. Similarly to the experiments in *N benthamiana*, the *A.vinelandii* NifD ($R^{96}R^{97}N^{98}\downarrow Y^{99}Y^{100}T^{101}$) mutants exhibited one or two substitutions in Tyr^{99} alone or together with Tyr^{100} (Y99K, Y99Q and Y99QY100T) (Allen et al. 2020). I used the N-terminal mitochondrial targeting peptide Cox4 to direct the variant proteins to rice mitochondria (Baysal et al. 2020). However, all three variants were still susceptible to MPP and only accumulated as truncated proteins, similarly to the native *A.vinelandii* NifD

expressed in rice. Degradation of the NifD variants in rice mitochondria was surprising as these three variants had been shown to prevent secondary cleavage in *N. benthamiana* (Allen et al. 2020). Notably, the same NifD variant Y100Q was also inconsistent in terms of resistance to degradation between *S. cerevisiae* and *N. benthamiana.* NifD variant Y100Q were resistant to degradation in *S. cerevisiae* and *N. benthamiana* mitochondria, to different extent, because degraded products were only observed in yeast (Allen et al. 2020). Additionally, the variants with substitution of Arg^{98} were stable in yeast but degraded in *N. benthamiana*. Xiang et al. (2020) reported that when the wild type NifD was co-expressed with MPPs in *E. coli*, plant MPPs (from rice, *Nicotiana tabacum*, and *Arabidopsis thaliana*) were less efficient in cleaving wild type NifD proteins compared to their yeast MPP counterparts. Thus, differences in cleavage efficiency of MPPs from different species, does not explain differences in stability of recombinant proteins targeted to mitochondria in different species. A more in depth understanding of rice mitochondrial precursor protein processing and identification of the specific cleavage sites in NifD are therefore, needed to remove this bottleneck.

4.6 Conclusions

I generated transgenic rice callus containing *A. vinelandii NifD* variants targeted to mitochondria in order to determine whether the variants might be resistant to degradation, as previously shown in *N. benthamiana*. I demonstrated similar levels of mRNA accumulation for the three variants. Surprisingly, all three variant recombinant proteins accumulated in truncated form in transgenic rice callus, suggesting differences in protein import into mitochondria in different species. Expression of the structural protein NifD is a necessary requirement for engineering BNF in plants. Therefore, further studies need to focus on understanding the basis of this instability and producing intact NifD in rice.

4.7 References

Abe Y, Shodai T, Muto T, Mihara K, Torii H, Nishikawa S-i, Endo T, Kohda D. 2000. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* **100**: 551-560.

Allen RS, Gregg CM, Okada S, Menon A, Hussain D, Gillespie V, Johnston E, Devilla R, Warden AC, Taylor M et al. 2020. Plant expression of NifD protein variants resistant to mitochondrial degradation. *Proc. Natl. Acad. Sci. U.S.A.* **117**: 23165-23173.

Allen RS, Tilbrook K, Warden AC, Campbell PC, Rolland V, Singh SP, Wood CC. 2017. Expression of 16 nitrogenase proteins within the plant mitochondrial matrix. *Front. Plant Sci.* **8**: 287.

Baysal C, Pérez-González A, Eseverri Á, Jiang X, Medina V, Caro E, Rubio L, Christou P, Zhu C. 2020. Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Res.* **29**: 37-52.

Burén S, Jiang X, López-Torrejón G, Echavarri-Erasun C, Rubio LM. 2017a. Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NifB. *Front. Plant Sci.* **8**: 1567.

Burén S, Jiménez-Vicente E, Echavarri-Erasun C, Rubio LM. 2020. Biosynthesis of nitrogenase cofactors. *Chem. Rev.* **120**: 4921-4968.

Burén S, Young EM, Sweeny EA, Lopez-Torrejón G, Veldhuizen M, Voigt CA, Rubio LM. 2017b. Formation of Nitrogenase NifDK Tetramers in the Mitochondria of Saccharomyces cerevisiae. *ACS Synth. Biol.* **6**: 1043-1055.

Burgess BK, Lowe DJ. 1996. Mechanism of Molybdenum Nitrogenase. *Chem. Rev.* **96**: 2983-3012.

Creissen GP, Mullineaux PM. 1995. Cloning and characterisation of glutathione reductase cDNAs and identification of two genes encoding the tobacco enzyme. *Planta* **197**: 422-425.

Curatti L, Rubio LM. 2014. Challenges to develop nitrogen-fixing cereals by direct *nif*-gene transfer. *Plant Sci.* **225**: 130-137.

Dos Santos PC, Dean DR, Hu Y, Ribbe MW. 2004. Formation and Insertion of the Nitrogenase Iron−Molybdenum Cofactor. *Chem. Rev.* **104**: 1159-1174.

Fay AW, Hu Y, Schmid B, Ribbe MW. 2007. Molecular insights into nitrogenase FeMoco insertion – The role of His 274 and His 451 of MoFe protein α subunit. *J. Inorg. Biochem.* **101**: 1630-1641.

Gakh O, Cavadini P, Isaya G. 2002. Mitochondrial processing peptidases. *Biochim. Biophys. Acta Mol. Cell Res.* **1592**: 63-77.

Georgiadis MM, Komiya H, Chakrabarti P, Woo D, Kornuc JJ, Rees DC. 1992. Crystallographic Structure of the Nitrogenase Iron Protein from *Azotobacter vinelandii*. *Science* **257**: 1653-1659.

Hawkes TR, Smith BE. 1983. Purification and characterization of the inactive MoFe protein (NifB-Kp1) of the nitrogenase from nifB mutants of Klebsiella pneumoniae. *Biochem. J.* **209**: 43-50.

Hu Y, Fay AW, Lee CC, Yoshizawa J, Ribbe MW. 2008. Assembly of Nitrogenase MoFe Protein. *Biochem.* **47**: 3973-3981.

Hu Y, Fay AW, Ribbe MW. 2007. Molecular insights into nitrogenase FeMo cofactor insertion: the role of His 362 of the MoFe protein α subunit in FeMo cofactor incorporation. *J. Biol. Inorg. Chem.* **12**: 449-460.

Kim CH, Newton WE, Dean DR. 1995. Role of the MoFe protein alpha-subunit histidine-195 residue in FeMo-cofactor binding and nitrogenase catalysis. *Biochem.* **34 9**: 2798-2808.

López-Torrejón G, Jiménez-Vicente E, Buesa JM, Hernandez JA, Verma HK, Rubio LM. 2016. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat. Commun.* **7**: 1-6.

Mossmann D, Meisinger C, Vögtle FN. 2012. Processing of mitochondrial presequences. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**: 1098-1106.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**: 292-292.

Orme-Johnson W. 1985. Molecular basis of biological nitrogen fixation. *Annu. Rev. Biophys.* **14**: 419-459.

Peters JW, Stowell MHB, Soltis SM, Finnegan MG, Johnson MK, Rees DC. 1997. Redox-Dependent Structural Changes in the Nitrogenase P-Cluster. *Biochem.* **36**: 1181-1187.

Poza-Carrión C, Jiménez-Vicente E, Navarro-Rodríguez M, Echavarri-Erasun C, Rubio LM. 2014. Kinetics of *nif* Gene Expression in a Nitrogen-Fixing Bacterium. *J. Bacteriol.* **196**: 595-603.

Robinson AC, Burgess BK, Dean DR. 1986. Activity, reconstitution, and accumulation of nitrogenase components in Azotobacter vinelandii mutant strains containing defined deletions within the nitrogenase structural gene cluster. *J. Bacteriol.* **166**: 180-186.

Shah VK, Brill WJ. 1973. Nitrogenase IV. Simple method of purification to homogeneity of nitrogenase components from Azotobacter vinelandii. *Biochim. Biophys. Acta - Bioenerg* **305**: 445-454.

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P. 1998. An efficient rice transformation system utilizing mature seed-derived explants and a portable, inexpensive particle bombardment device. *Transgenic Res.* **7**: 289-294.

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P. 1998. Transgenic Central American, West African and Asian Elite Rice Varieties Resulting from Particle Bombardment of Foreign DNA into Mature Seed-derived Explants Utilizing Three Different Bombardment Devices. *Ann. Bot.* **82**: 795-801.

Xiang N, Guo C, Liu J, Xu H, Dixon R, Yang J, Wang Y-P. 2020. Using synthetic biology to overcome barriers to stable expression of nitrogenase in eukaryotic organelles. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 16537-16545.

Chapter V.
Toward enhancing the photosynthetic efficiency of rice by engineering an ectopic oxygen scavenging pathway

5.0 Abstract

Increasing crop productivity to address the growing demand for food is critical. Given that carbon assimilation by photosynthesis is the source of plant productivity, improving the efficiency of photosynthesis has been a major strategy to increase crop productivity. In C_3 crops, photorespiration reduces overall photosynthetic efficiency because it results in a 20-50% loss of carbon fixed by photosynthesis (Jordan and Ogren 1981; Mann 1999). Therefore, reducing photorespiration in C_3 crops is an important strategy to improve photosynthetic efficiency. Here, I described preliminary experiments aiming to construct a novel oxygen scavenging pathway in rice towards minimizing photorespiration thus enhancing photosynthesis. I introduced *Lactobacillus buchneri* lactate oxidase (*Lb*LOX), *Escherichia coli* lactate dehydrogenase (*Ec*LDH), and *Oryza sativa* catalase (*Os*CAT) into rice and generated a number of transgenic lines. Expression of the introduced transgenes was confirmed at the mRNA level. Accumulation of *Ec*LDH protein was confirmed in callus and regenerated plants but *Lb*LOX and *Os*CAT protein could not be detected. Further experiments are needed to confirm the accumulation of both recombinant proteins and to ensure the functionality of the entire pathway in rice. Once achieved, further experiments should focus on efforts to ascertain if the oxygen scavenging pathway can reduce photorespiration sufficiently to permit substantial increases in photosynthesis, without compromising plant development and fertility.

5.1 Introduction

5.1.1 Rubisco and its catalytic properties

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is present in most prokaryotic autotrophs, including photosynthetic and chemoautotrophic bacteria, cyanobacteria and archaea and eukaryotes such as various algae and higher plants (Ellis 1979). Rubisco is a multimeric enzyme containing two different types of subunits: large (L, 50–55 kDa), and small subunits (S, 12–18 kDa). There are four forms of Rubisco in nature, differing in their quaternary structure. Form I comprises eight L subunits and eight S subunits (L_8S_8). Form II contains only 2, 4, 6 or 8 L subunits ($L_{2\text{-}8}$). Form III has either 2 or 10 L subunits [L_2 or $(L_2)_5$], and form IV contains two L subunits (Tabita et al. 2008). Although the absence of the S subunit in forms II and III appears to suggest that this subunit is not absolutely necessary for the activity, it is important for holoenzyme stability and maximal catalytic activity in form I (Mao et al. 2022). Form I Rubisco is the most abundant, and it is present mainly in chemoautotrophic bacteria, cyanobacteria, algae, and

all higher plants, accounting for ca: 50% of the total soluble protein in plant leaves (Baker et al. 1977; Ellis 1979). In the context of the work described in this chapter, I focus mainly on form I. Although the amino acid sequence identity among different Rubisco forms is 25-30%, the secondary structure (overall fold) of the L subunit is extremely conserved in all forms (Andersson and Backlund 2008). The L subunit consists of a smaller N-terminal domain and a larger C-terminal domain (as exemplified by the *Arabidopsis thaliana* Rubisco L subunit, **Figure 5.1**). The 150 amino acids in the N-terminal domain fold into a four-stranded antiparallel β sheet with two α helices (**Figure 5.1**). The larger C-terminal domain contains 329 amino acids forming an eight-stranded parallel α/β barrel structure and the loops located at the end of the β-strands are the catalytic and substrate binding sites (**Figure 5.1**) (Andersson and Taylor 2003; Valegård et al. 2018).

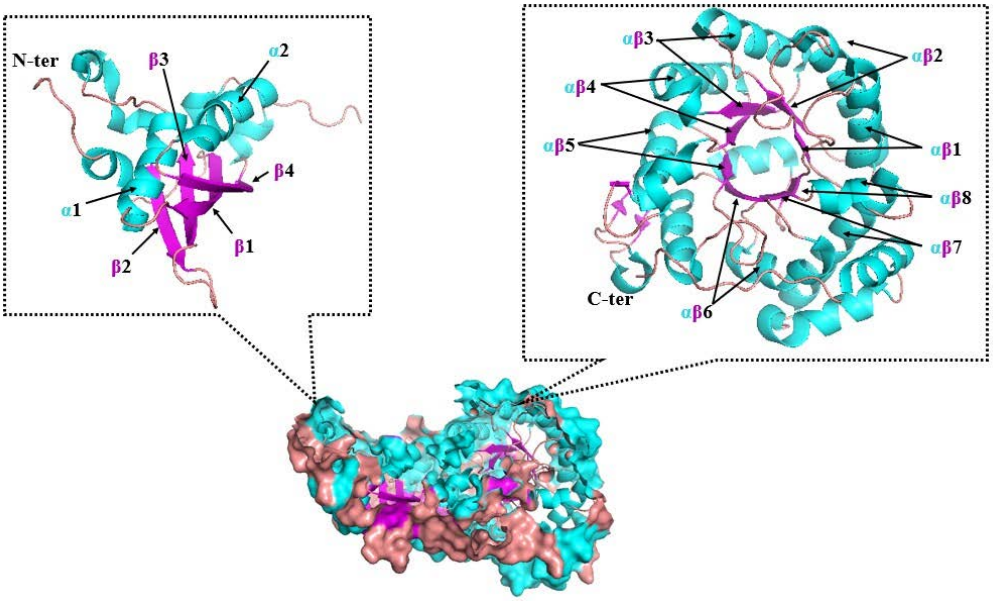

Figure 5.1 Structure of the *A. thaliana* Rubisco large subunit. (PDB: 5IU0. https://www.rcsb.org). The blue, purple and pink colors indicate the α helix, β sheet and random coil, respectively. The N-terminal domain is comprised of a four-stranded β-sheet and two α-helices. The C-terminal domain contains an eight-stranded α/β-barrel structure.

Rubisco catalyzes the initial step of inorganic carbon fixation in the Calvin cycle, the assimilation of CO_2 by carboxylation of ribulose 1,5-bisphosphate (RuBP) to form two molecules of 3-phosphoglyceric acid (3PGA) (Kelly et al. 1976). This reaction predominates the inorganic to organic carbon transfers entering the biosphere (Rothschild 2008). However, Rubisco as the central enzyme in photosynthetic carbon assimilation, is a bifunctional enzyme, as it catalyzes a second competing oxygenation reaction at the same active site (Li et al. 2003).

Since, both the carboxylase and oxygenase reactions occur at the same active site, the efficiency of Rubisco carboxylase is mainly affected by the specificity ratio between O_2 and CO_2 and the relative concentration of O_2 and CO_2 in the vicinity of the enzyme (Parry et al. 2007). There are also many environmental factors which affect carboxylase relative to oxygenase, including temperature and drought (Parry et al. 2007). The CO_2/O_2 specificity of Rubisco (referred to as Ω) represents the ratio of catalytic efficiencies for carboxylation (Vc/Kc) to oxygenation (Vo/Ko), VcKo/VoKc. Vc and Vo are maximal velocities for carboxylation and oxygenation, respectively, and Kc and Ko are the Michaelis constants for CO_2 and O_2, respectively. A negative correlation between Ω and maximum catalytic rates for carboxylation per active site exists in photosynthetic organisms. For example, higher plants have higher Ω values but lower carboxylation catalytic rates (Jordan and Ogren 1981). There is a ca: 30% variation in Rubisco I in different plants, cyanobacteria, and algae. However, among C_3 plants, the variation in Ω values is less than 10% (Flamholz et al. 2019). In addition, the intracellular relative concentrations of CO_2 and O_2 clearly influence the ratio of carboxylase and oxygenase reactions. Microalgae and cyanobacteria have evolved carbon concentrating mechanisms (CCM) to increase the CO_2 concentration in the vicinity of Rubisco. In contrast, higher plants have evolved C_4 photosynthesis and crassulacean acid metabolism (CAM) to achieve the same objective (Dehigaspitiya et al. 2019). In C_3 plants, atmospheric CO_2 diffuses to the chloroplast through stomata and the air spaces between cells, where it is then fixed by Rubisco (**Figure 5.2**). C_3 photosynthesis is more common and ca: 85% of all higher plants are C3 plants, including major crops such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), soybean (*Glycine max*), and potato (*Solanum tuberosum*) (Yamori et al. 2014). In C_4 and CAM plants, atmospheric CO_2 entering the cells, is hydrated to bicarbonate (HCO_3^-) by carbonic anhydrase (CA). HCO_3^- and phosphoenolpyruvate (PEP) are converted to 4-carbon acids by phosphoenolpyruvate carboxylase (PEPC) (**Figure 5.2**). The main difference between C_4 and CAM is the special or temporal separation of Rubisco and PEPC (**Figure 5.2**). In C_4 plants, such as maize (*Zea mays*) and sugarcane (*Saccharum officinarum*), Rubisco and PEPC are active in bundle sheathe and mesophyll cells, respectively. In CAM plants, Rubisco and PEPC operate in different periods of a diurnal cycle. PEPC is active at night and Rubisco is active during the day. This temporal separation of photosynthesis is common in succulents (Yamori et al. 2014).

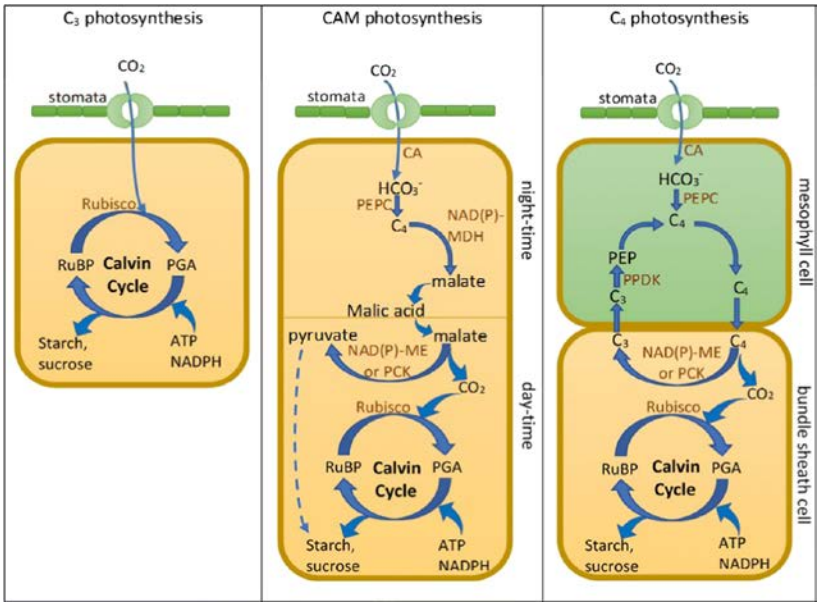

Figure 5.2 Schematic representation of C_3, CAM and C_4 photosynthesis. Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; RuBP, ribulose 1,5-bisphosphate; PGA, phosphoglyceric acid; CA, carbonic anhydrase; HCO_3^-, bicarbonate, NADP-MDH, NADP-malate dehydrogenase; NADPME, NADP-dependent malic enzyme; PCK, phosphoenolpyruvate carboxykinase; PEPC, phosphorenolpyruvate carboxylase; PPDK, pyruvate phosphate dikinase; PEP, phosphorenolpyruvate (Dehigaspitiya et al. 2019).

5.1.2 Photorespiration and impact on plant metabolism

Rubisco oxygenates RuBP to produce one molecule of 3PGA and one molecule of 2-phosphoglycolate (2PG). The latter, not only cannot be used for carbon assimilation, but is also a toxic metabolite inhibiting triose-phosphate isomerase (TPI) and sedoheptulose 1,7-bisphosphate phosphatase (SBPase) in the Calvin cycle, and alters the allocation of photosynthates between RuBP regeneration and starch synthesis (Flügel et al. 2017). Metabolite damage caused by 2PG necessitates an energy-requiring salvage pathway to recycle 2PG back to 3PGA. This energy-requiring repair process is known as photorespiration. In energy terms 3.25 mol ATP and 2 mol NADPH are consumed per oxygenation reaction which is about 30% of the total energetic cost of CO_2 fixation (Wingler et al. 2000). Photorespiration involves 16 enzymes and more than six transporters distributed over the chloroplast, peroxisome, and mitochondria (Douce and Neuburger 1999). There are multiple enzymatic reactions in the photo-respiratory pathway, including de-phosphorylation of 2PG, glycolate metabolism in the peroxisome, conversion of glycine to serine in mitochondria, formation of glycerate in the peroxisome, regeneration of phosphoglycerate, and ammonia re-assimilation (Peterhansel et al. 2010).

Photorespiration has a multitude of effects on plant metabolism. Firstly, it is closely integrated with nitrogen assimilation. NH_3 released from photorespiration is assimilated

by the GS/GOGAT system composed of glutamate synthase (GOGAT) and glutamine synthase (GS) in the chloroplast. The rate of refixation of NH_3 from photorespiration by the GS/GOGAT system is ten times faster than the assimilation of primary NH_3 (derived from nitrate reduction) (Leegood et al. 1995). Furthermore, the photorespiration pathway contributes to the biosynthesis of certain amino acids, for example serine and glycine in photosynthetic tissues (Gerbaud and Andre 1979). The photorespiratory pathway also interacts with C_1 metabolism because the conversion of glycine to serine produces 5,10-methylenetetrahydrofolate (CH2-THF), the precursor of C_1 metabolism (Peterhansel et al. 2010). CH_2-THF is converted to several derivatives of THF (10-formyl-THF,5,10-methenyl-THF, and 5-methyl-THF). These C_1 units are used in a wide range of primary and secondary metabolic reactions, including protein synthesis, methylation of DNA, RNA, proteins, phospholipids and the biosynthesis of many secondary products such as glycine betaine, nicotine, and lignin (Busch 2020). Photorespiration further intersects with sulfur metabolism because serine is used to synthesize cysteine by serine O-acetyltransferase (Abadie and Tcherkez 2019). Cysteine is the primary product of S-assimilation and is important to methionine synthesis and glutathione metabolism (Rausch and Wachter 2005).

5.1.3 Optimization of photorespiration in plants through genetic engineering.

Although photorespiration is ubiquitous in plants, the photorespiration rate of C_3 plants is higher than that of their C_4 counterparts, because of the absence of the CO_2 concentration mechanism (which as explained above) (Ogren 1984). Therefore, to improve the efficiency of light energy conversion of C_3 plants, efforts have focused on reducing photorespiration through three approaches: 1) modification of the specificity of Rubisco through protein engineering, 2) introducing C_4 photosynthesis components to C_3 plants 3) creating a photorespiratory bypass.

Rubisco specificity varies widely among plants, cyanobacteria, and algae. In plants Rubisco specificity is higher, but carboxylation catalytic rates are lower. In contrast, Rubisco in cyanobacteria and algae exhibits faster carboxylation catalytic rates with lower specificity (Jordan and Ogren 1981). The complexity of Rubisco functions makes its modification challenging. A hybrid enzyme comprising *Nicotiana tabacum* S subunits and *Synechococcus elongatus* L subunits was assembled in plants, but significantly reduced Rubisco activity compared with the wild-type tobacco enzyme (Orr et al. 2019). The gene encoding the tobacco Rubisco L subunit was replaced by *S. elongatus* Rubisco

L and S subunit genes, in combination with either the corresponding assembly chaperone, RbcX, or an internal carboxysomal protein, CcmM35. The resulting recombinant proteins showed a 1.5-3-fold increase in the CO_2 fixation rate per enzyme unit (the amount of CO_2 fixed per second per mole of active sites) compared to the wild-type tobacco enzyme (Lin et al. 2014).

Photosynthesis in C_4 plants significantly reduces photorespiration by increasing the CO_2 concentration near Rubisco (Sharpe and Offermann 2014). Enzymes related to C_4 plant photosynthesis include carbonic anhydrase (CA), phosphoenolpyruvate carboxylase (DEPC), NADH-malate dehydrogenase (NADH-MDH), and pyruvate orthophosphate kinase (PPDK). Genes for these enzymes are also present in C_3 plants, but they express at very low levels (Brown et al. 2005; Hibberd et al. 2008). Due to the specific Kranz structure of C_4 plants, the distribution and functions of key photosynthetic enzymes have obvious spatiotemporal characteristics. PEPC is active in the mesophyll tissue, and 4-carbon acids have to be transported intercellularly to Rubisco. Kranz structure consists of two types of photosynthetic cells: bundle sheath and mesophyll cells in which the later cells cluster around bundle sheath cells in the form of a wreath. The inner compartment of the Kranz structure is formed by bundled sheath cells surrounding the leaf veins, to where Rubisco is restricted (Yamori et al. 2014). Introducing C_4 photosynthetic genes into C_3 plants to form a C_4 single-cell type may not be sufficient to achieve the high light-efficiency function as in C_4 plants. This is because an essential anatomical configuration for an efficient C_4 cycle is a high bundle sheath to mesophyll ratio such as Kranz structure in photosynthetic tissue (Zhu et al. 2010; Edwards 2019). A maize GOLDEN2-LIKE gene (a transcription factor regulating plant chloroplast development, GLK), expressed in rice generated a transitional Kranz anatomy (Wang et al. 2017). As a result, engineered plants had bigger bundle sheath cells and plasmodesmata connectivity at the bundle sheath/mesophyll cells interface was increased. Similarly, Rubisco and Rubisco activase accumulation increased in the transitional Kranz rice. However, photosynthetic rate did not increase over that of the wild type (Wang et al. 2017). Similarity, Li et al. (2020) observed small differences (not statistically significant) in net photosynthetic rate between GLK transgenic and wild type plants when the light intensity was < 200 μmol m^{-2} s^{-1}. However, GLK transgenic rice plants showed a significant increase (48% higher) in net photosynthetic rate compared to the wild type at 1500 μmol m^{-2} s^{-1} light intensity (Li et al. 2020). The results were attributed to the fact that GLK upregulates the expression

of nuclear photosynthetic genes involved in chlorophyll biosynthesis and light harvesting (Waters et al. 2009).

As discussed above, photorespiration is a complex process requiring energy, involving multiple enzymatic reactions distributed in chloroplasts, peroxisomes, and mitochondria. The main photorespiratory bypass pathways tested in C_3 plants are summarized in **Figure 5.3** (Kebeish et al. 2007; Maier et al. 2012; Wang et al. 2020). All three bypass pathways were designed to reduce the loss of energy and fixed carbon and enhance CO_2 concentration in chloroplasts by establishing an additional metabolic glycolate pathway.

In the first bypass, glycolate is converted directly to glycerate by a sequence of enzymatic reactions using chloroplast-targeted *E. coli* genes encoding glycolate dehydrogenase (*Ec*GDH), glyoxylate carboligase (*Ec*GCL) and tartronic semialdehyde reductase (*Ec*TSR) (Kebeish et al. 2007). In energy terms 2.75 mol ATP and 1 mol NAD(P)H are consumed which is 0.5 mol ATP and 2 mol NAD(P)H less compare to plant photorespiration (Peterhansel et al. 2013). Resulting *A. thaliana* plants expressing the *E. coli* GDH subunits D, E, and F, either alone (DEF) or in combination with *Ec*GCL and *Ec*TSR (DEF-GT) showed a significant increase in biomass (as measured by rosette diameter, shoot and root dry weight) and CO_2 release from glycolate in chloroplast extracts (Kebeish et al. 2007). The apparent rate of CO_2 assimilation in DEF and DEF-GT transgenic *A. thaliana* plants increased by 22% and 45%, respectively, compared to wild type (Kebeish et al. 2007). Expression of *Ec*GDH (DEF) in potato (*Solanum tuberosum*) chloroplasts boosted the apparent photosynthetic rate by 37% and the tuber yield by 2.3-fold (Nölke et al. 2014).

In an alternative bypass, *Arabidopsis thaliana* glycolate oxidase (*At*GLO), *E. coli* catalase (*Ec*CAT), and *Cucurbita maxima* malate synthase (*Cm*MS) were targeted to *A. thaliana* chloroplasts. The premise underpinning this bypass is that glycolate is converted to glyoxylate by *At*GlO, which is further converted to malate by *Cm*MS. Malate is subsequently decarboxylated to pyruvate by chloroplastic NADP-malic enzyme (NADP-ME) and chloroplastic pyruvate dehydrogenase (PDH) converts pyruvate to Acetyl-CoA. *Ec*CAT detoxifies excess H_2O_2, produced by the oxidation of glycolate. Decarboxylation of malate to pyruvate and oxidation of pyruvate to acetyl-CoA result in the production of one molecule of CO_2 in each of the two reactions (Maier et al. 2012). This bypass is very demanding energetically requiring 2.75 mol ATP and 1 mol NAD(P)H more than plant photorespiratory pathway (Peterhansel et al. 2013). Transgenic plants expressing the input transgenes (*AtGlo, EcCat* and *CmMs*) showed an increase in rosette dry and fresh weights,

by 36% and 28%, respectively (Maier et al. 2012). The plants also exhibited a trend toward higher CO_2 assimilation rates, even though results were not statistically significant compared to wild type (Maier et al. 2012).

A third bypass was engineered into rice chloroplasts by introducing *O. sativa* glycolate oxidase (*Os*GLO), *Ec*CAT、*Ec*GCL and *Ec*TSR (Wang et al. 2020).The first reaction in this bypass is the same as in the second bypass (described above), in that glycolate is converted into glyoxylate and resulting H_2O_2 is removed by *Ec*CAT. The pathway then diverges, with glyoxylate being converted to glycerate through *Ec*GCL and *Ec*TSR, similarly to the first bypass. This bypass consumes 0.5 mol ATP less in terms of energy compared to the photorespiratory pathway (Wang et al. 2020). The transgenic plants expressing all transgenes (*OsGlo*, *EcCat*, *EcGcl* and *EcTst*, GCGT) showed 27% increase in grain yield per plant and a 24% increase in stem and leaf dry weight. However, the seed setting rate was decreased by 21% (Wang et al. 2020). GCGT plants exhibited a 16% increase in the net photosynthetic rate while the photorespiration rate decreased by 19% (Wang et al. 2020).

In my work, I established a novel oxygen scavenging pathway to reduce photorespiration by introducing *Lactobacillus buchneri* lactate oxidase (*Lb*LOX), *E. coli* lactate dehydrogenase (*Ec*LDH) and *O. sativa* catalase (*Os*CAT) into rice chloroplasts (**Figure 5.3**). The basis for this relies on the creation of an O_2 consumption cycle through *Lb*LOX and *Ec*LDH. The cycle is initiated by the consumption of O_2 and the oxidation of lactate to pyruvate by *Lb*LOX, and completed by the conversion of pyruvate to lactate by *Ec*LDH. *Os*CAT serves to remove the H_2O_2 produced by the oxygen consumption cycle.

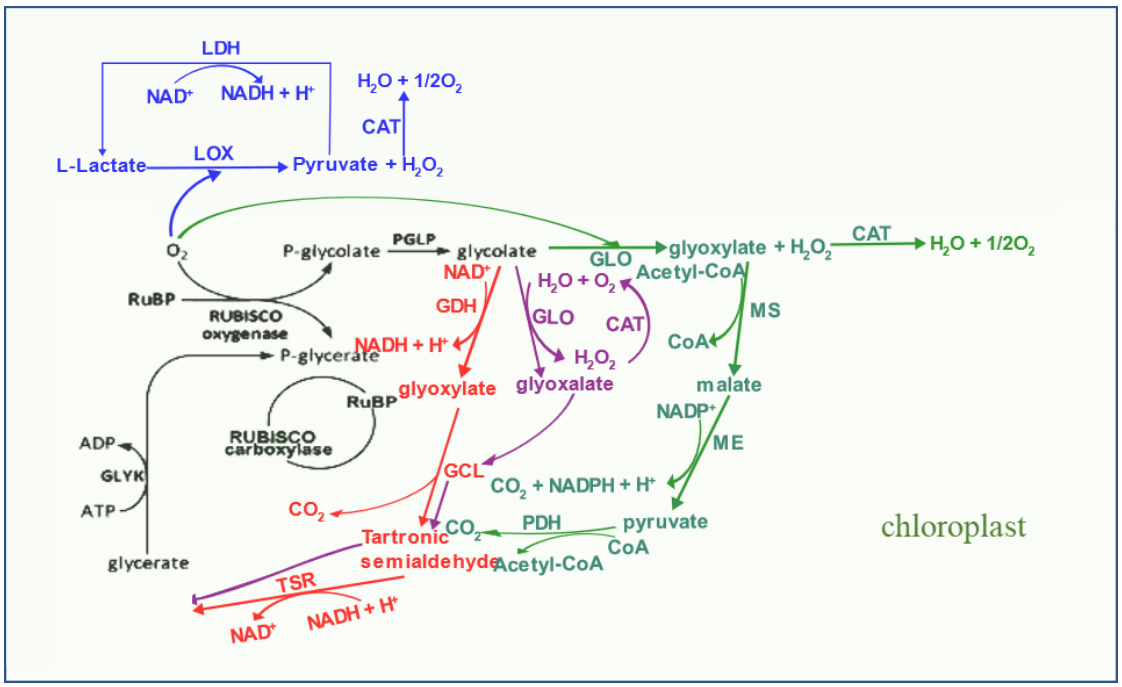

Figure 5.3 Photorespiratory bypass pathways introduced in C_3 plants and a novel oxygen scavenging pathway (under investigation in this work). The plant photorespiratory pathway is shown in black and the various engineered photorespiration bypass pathways are indicated in different colors, representing the approaches used in the different strategies. The first bypass (red) includes glycolate dehydrogenase (GDH) glyoxylate carboligase (GCL) and tartronic semialdehyde reductase (TSR). The second (green) includes glycolate oxidase (GLO), malate synthase (MS), malic enzyme (ME), pyruvate dehydrogenase (PDH), and CAT. The third (pink) comprises GLO, catalase (CAT), GCL; and TSR. The novel bypass (blue) involves lactate oxidase (LOX), lactate dehydrogenase (LDH) and CAT.

5.2 Aims and objectives

The aim of the work described in this chapter was toward introduce a novel oxygen scavenging pathway into rice comprising LOX, LDH, and CAT to enhance photosynthetic efficiency by reducing photorespiration. The specific objectives were to: 1) introduce *Lactobacillus buchneri Lox*, *E. coli Ldh*, and *O. sativa Cat* into rice chloroplasts; 2) confirm the presence of the transgenes in rice callus; 3) measure mRNA and protein accumulation in transgenic callus and plants.

5.3 Materials and methods

5.3.1 Constructs for rice transformation

The *Solanum tuberosum* codon-optimized sequences of *Lactobacillus buchneri subsp. Silage* lactate oxidase (*LbLox*, AFR99121.1), *E. coli* lactate dehydrogenase (*EcLdh*, AAA03585.1), chloroplast stromal targeting peptide *S. tuberosum* Rubisco small subunit (CTP), Myc and Tag54 tags were cloned into a potato multigene expression vector

(pL2_Oxy_KanP, *At*RbcS1_RbcS15'UTR_CTP_*LbLox*_tag54_ 3'UTR HSP18.2_*St*LS1_ PVX 5'UTR_CTP_*EcLdh*_ Myc_*At*Ac2) by Dr G. Nölke (Department of Plant Biotechnology, Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Aachen, Germany). The *At*Rbcs2B 5'UTR-CTP-*LbLox*-Tag54-3'UTR *At*HSP18.2 cassette was generated by PCR using pL2_Oxy_KanP as the template and introduced into the rice expression vector pAL76 containing the maize *Ubiquitin-1* promoter plus first intron and *Nos* transcriptional terminator at the BamHI and EcoRI sites to produce pLOX. The PVX 5'UTR-CTP-*EcLdh*-6xHis was amplified by PCR using pL2_Oxy_KanP as the template and introduced into pAL76 at the EcoRI and HindIII sites to produce pLDH-6xHis. In addition, the sequence *EcLdh*-Myc was amplified using the template pL2_Oxy_KanP with restriction sites EcoRI and HindIII, and inserted into pUC57mimi carrying the maize *Ubiquitin-1* promoter plus first intron following the *O. sativa* codon optimized chloroplast stromal targeting peptide *O. sativa* Rubisco small subunit (RC2) and *Nos* transcriptional terminator, to generate pLDH-Myc. The coding region of *O. sativa* catalase (*OsCat*, C_Q10S82) was synthesized and cloned into the above pUC57mini by Genscript (Piscataway, USA), named the final construct as pCAT. All restriction enzymes were obtained from Promega (Madison, USA), and T4 DNA ligase was obtained from Roche (Basel, Switzerland). The plasmids were amplified in *E. coli* DH5α grown at 37 °C in LB medium supplemented with 100 μg/mL ampicillin. The final constructs were verified by Sanger sequencing (Stabvida, Caparica, Portugal). Primers used for plasmid construction are shown in **Table 5.1.**

Table 5.1 Primers used for vector construction.

Genetic element	Primer sequence (5′→3′)	Restriction enzymes
*At*Rbcs2B 5'UTR-CTP-*LbLox*-Tag54-3'UTR *At*HSP18.2	F:CGCGGATCCGGGCTTTTTGCCTCTTACGGTTCTC R:CCGGAATTCCATAGTCCATACCATAGCACATACAGTAG	BamHI EcoRI
PVX 5'UTR-CTP-*EcLdh*-6xHis	F:CCGGAATTCACACCACCAACACAACCAAACCCAC R:CCCAAGCTTTCAATGGTGATGGTGATGGTGAGCAGCATTGCCTTTTGCCATTG	EcoRI HindIII
EcLdh-Myc	F:CCGGAATTCATGATCATCAGCGCTGCATCCGATT R:CCCAAGCTTAAAACTCATCTCAGAAGAGGATCTG	EcoRI HindIII

*At*Rbcs2B 5'UTR and PVX 5'UTR, 5'UTR of *A. thaliana* Ribulose bisphosphate carboxylase small subunit 2B and potato virus X, 3'UTR *At*HSP18.2, 3'UTR of *A. thaliana* heat shock protein 18.2; CTP, chloroplast stromal targeting peptide *S. tuberosum* Rubisco small subunit; *LbLox, Lactobacillus buchneri* lactate oxidase; *EcLdh, E. coli* lactate dehydrogenase.

5.3.2 Transformation of rice explants, callus recovery and regeneration of transgenic plants

Six-day-old mature rice embryos (*O. sativa* cv. Bomba) were used as explants for transformation using particle bombardment (Sudhakar et al. 1998; Valdez et al. 1998). The 0.9 μm gold particles (Bio-Rad, California, USA) were coated with either pLDH-6xHis or pLDH-Myc, together with pCAT, pLOX and the hygromycin phosphotransferase (HPT) selectable maker at a molar ratio of 3:3:3:1. The generation of transgenic rice callus and regeneration and hardening of transgenic plantlets were as described in chapter II.

5.3.3 DNA extraction and PCR analysis

Genomic DNA extraction and PCR proceeded as described in Chapter IV. Primers used for PCR analysis are shown in **Table 5.2**. The PCR for *LbLox* follows a protocol: initial denaturation at 95 ºC/5 min, followed by 35 cycles at 95 °C/30 s denaturation, 63 °C/ 30 s, extension at 72 ºC/1 min 30 s; and final extension at 72 °C/10 min. The conditions for *EcLdh* and *OsCat* genes were similar to those of *LbLox* except that the annealing temperature was changed to 54 °C and 60 °C, respectively. The PCR products were resolved using agarose gel (1%) electrophoresis with 1 × Tris-acetate buffer (TAE) and visualized on Gel Doc XR (Bio-Rad Laboratories, Hercules, CA, USA).

Table 5.2 Primers used for PCR analysis

Gene	Primer sequence (5′→3′)	Product size (bp)
OsCat	F: CCGGAATTCATGGATCCCTACAAGCACCG R:ATGCTCCCAGTCCTTGAGGTGCTTCGCCAGTTTCTGAC	1476
LbLox	F: AGGATCCGGGCTTTTTGCCTCTTACGGTTCTC R:TGATATCGAATTCCATAGTCCATACCATAGCACCAGTAG	1679
EcLdh	F: CGTAACGTGGAGGACTTATCT R: CTTGAACTAGGCTATCTTGAGTAA	1024

OsCat, O. sativa catalase; *LbLox, Lactobacillus buchneri* lactate oxidase; *EcLdh, E. coli* lactate dehydrogenase.

5.3.4 Gene expression analysis by real-time quantitative reverse transcription PCR

Total RNA extraction, cDNA synthesis and real-time quantitative reverse transcription PCR (RT-qPCR) were carried out as described in Chapter II. The gene-specific primers

used for RT-qPCR analysis are listed in **Table 5.3**. The PCR products were confirmed by sequencing. Expression levels were normalized against *OsActin* mRNA.

Table 5.3 Primers used for real-time quantitative reverse transcription PCR analysis.

Gene	Primer sequence (5′→3′)	Product size (bp)
OsActin	F: TCATGTCCCTCACAATTTCC	181
	R: GACTCTGGTGATGGTGTCAGC	
OsCat	F: CTCGCTCCAAGGCGTTTCTCATA	125
	R: ACCAAACTACCTGCTGCTCC	
LbLox	F: GCGGTATCTTTCTCTCCCTGA	91
	R: ATGTCTTTGGAATCCCCCTAA	
EcLdh	F: CTTTCCACCGTTTCTGTTTGTCC	118
	R: CTCGCTCCAAGGCGTTTCTCATA	

OsActin, O. sativa Actin; OsCat, O. sativa catalase; *LbLox, Lactobacillus buchneri;*lactate oxidase; *EcLdh, E. coli* lactate dehydrogenase.

5.3.5 Protein extraction and immunoblot analysis

Total protein was extracted from rice tissues by grinding 50-100 mg callus or leaves in liquid nitrogen and homogenizing in 0.2–0.4 mL of protein extraction buffer (7 M Urea, 2 M thiourea, 30 Mm Tris-HCl, 4% (w/v) CHAPS and proteinase inhibitor (Sigma, P9590)). The mixture was incubated on ice for 30 min, with vortexing every 10 min. After incubation, the debris was removed by centrifugation at 17,000 g for 20 min at 4 °C, the supernatant was collected and kept at 4°C. The protein concentration in the supernatant was determined using the Bradford method (AppliChem). For immunoblot analysis, 50 µg of total protein was boiled in 5x Laemmli Buffer (62.5 mM Tris–HCl pH 6.8, 25 mM DTT, 4 % (w/v) SDS, 30 % (w/v) glycerol, 0.05 % (w/v) bromophenol blue) for 4 min at 95 °C. The extracted protein was separated by 10% SDS polyacrylamide gel electrophoresis and then immunoblotted to Protran Premium 0.45 mm nitrocellulose membranes (Amersham) using a semidry transfer apparatus (Bio-Rad). Equal loading was confirmed by Ponceau staining of nitrocellulose membranes. Primary monoclonal antibodies against Tag54 (Rasche et al. 2011), polyclonal antibodies Myc tag (C3956, Sigma-Aldrich, St Louis, MO, USA) or 6x-His Tag (PA1-983B, Thermo Fisher Scientific) were diluted at 1: 1500-1: 2500 in TBS-T which was supplemented with 2% milk. Secondary antibodies (Thermo Fisher Scientific) were diluted at 1: 20,000 in TBS-T which was supplemented with 2% non-fat milk and incubated for 1 h at room temperature. Detection was done with Immobilon Crescendo Western HRP substrate (WBLUR0100, Millipore Corporation, Massachusetts, USA) and the Molecular Imager ChemiDoc XRS+ System (Bio-Rad).

5.3.6 Statistical analysis

The standard deviation (SD) of relative expression data was calculated based on three technical replicates.

5.4 Results

5.4.1 Expression constructs for rice transformation

I established a novel oxygen scavenging pathway encompassing *Lb*LOX, *Ec*LDH and *Os*CAT in rice, by direct DNA transfer. *Os*CAT scavenges the H_2O_2 produced by the oxygen consumption step catalyzed by *Lb*LOX and was used in the experiments because it was reported to work effectively in rice (Shen et al. 2019). *Lb*LOX and *Ec*LDH were selected based on the enzymatic activity in the native organism. *Lb*LOX with oxygen as acceptor, catalyzes the reaction

L-lactate + O_2 = H_2O_2 + pyruvate

in the forward direction (Heinl et al. 2012). *Ec*LDH specifically interconverts pyruvate and L-lactate, while the other two *E. coli* lactate dehydrogenases (D-lactate dehydrogenase encoded by *EcDld* and fermentative D-lactate dehydrogenase encoded by *EcLdhA*) are specific for D-lactate (Tarmy and Kaplan 1968; Kohn and Kaback 1973; Dong et al. 1993). The *S. tuberosum* codon-optimized *LbLox* and *EcLdh* sequences were used for rice transformation directly because there was no rare codon base on *O. sativa* codon usage (Nakamura et al. 2000). The *LbLox*, *EcLdh*, and *OsCat* were introduced into separate vectors for rice transformation. Expression of these genes was under the control of the maize *Ubiquitin-1* promoter plus first intron and the *Nos* transcriptional terminator. The AtRbcs2 5'UTR with AtHSP18.2 3'UTR, and the potato virus X 5'UTR were used to enhance the stability of transcription of *LbLox* and *EcLdh*, respectively. The transit peptide of Rubisco small subunit from *S. tuberosum* (CTP) or *O. sativa* (RC2) was fused to the N-terminal of *Lb*LOX, *Ec*LDH, and *Os*CAT to enable targeting to the chloroplast stroma. The protein tags Tag54, Myc or 6xHis were add to the C-terminal of *Lb*LOX, *Ec*LDH, and *Os*CAT for protein detection. An additional construct carrying the selectable marker *Hpt* under the control of the CaMV 35S promoter was also used.

5.4.2 DNA analysis in transgenic callus

To confirm the presence of *LbLox*, *EcLdh*, and *OsCat* in rice callus, twelve putative transgenic callus lines were analysed by PCR (**Figure 5.4**). Six independent callus lines (lines 1, 2 4, 5, 6, 11) had specific 1679 bp, 1024 bp and 1476 bp amplicons corresponding to *LbLox*, *EcLdh*, and *OsCat*, respectively, confirming the presence of the three transgenes in these lines.

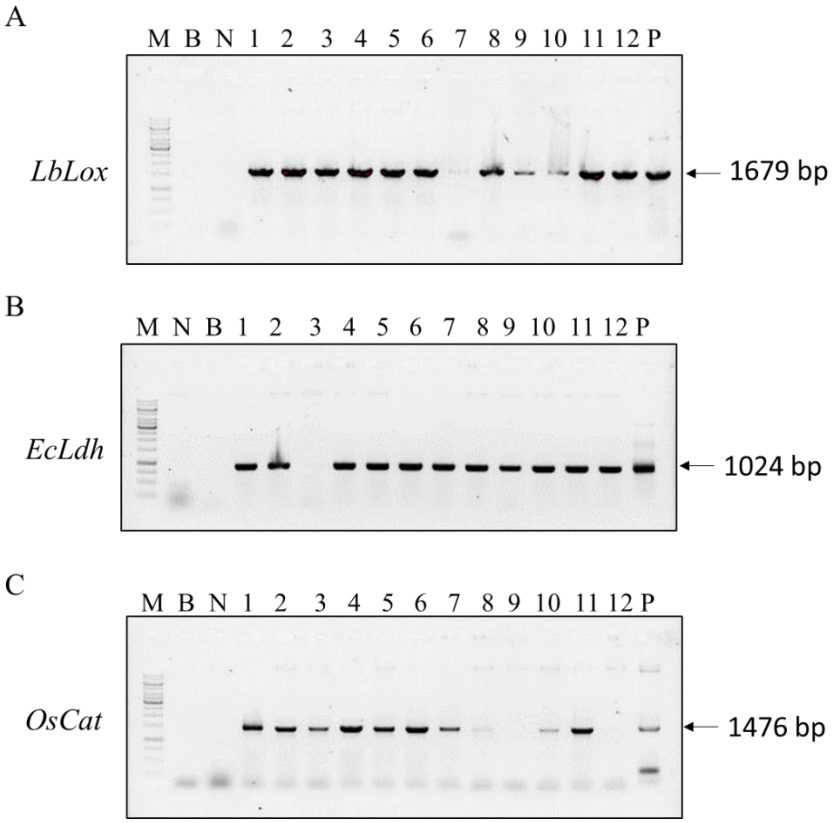

Figure 5.4 PCR amplification of *LbLox*, *EcLdh*, and *OsCat* in transgenic callus lines. (A) *LbLox* amplification product (1679 bp); (B) *EcLdh* amplification product (1024 bp); and (C) *OsCat* amplification product (1476 bp). Lane M, GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific); lane B, millipore water (blank); lane N, genomic DNA extracted from wildtype callus (negative control); lane P, plasmid DNA (positive control); lanes 1-12, independent transgenic lines.

5.4.3 mRNA accumulation in transgenic rice callus and plants

The expression of *LbLox*, *EcLdh*, and *OsCat* in callus lines and plants was confirmed by RT-qPCR analysis with mRNA levels normalized against *OsActin* mRNA (**Figure 5.5 and 5.6**). mRNA accumulation of all three transgenes *LbLox*, *EcLdh*, and *OsCat* was detected at callus Lines 1, 2, 4, 6 and 11. The abundance of *OsCat* mRNA was higher in all five transgenics lines, compared to *LbLox* and *EcLdh*. There was no significant difference in the relative expression levels of *LbLox* and *EcLdh* except in callus line 4. A

similar trend was noted in transgenic plants regenerated from the callus lines. Higher mRNA levels were measured for *OsCat* in all six plants with comparable expression levels for *LbLox* and *EcLdh* in four plants (**Figure 5.6**).

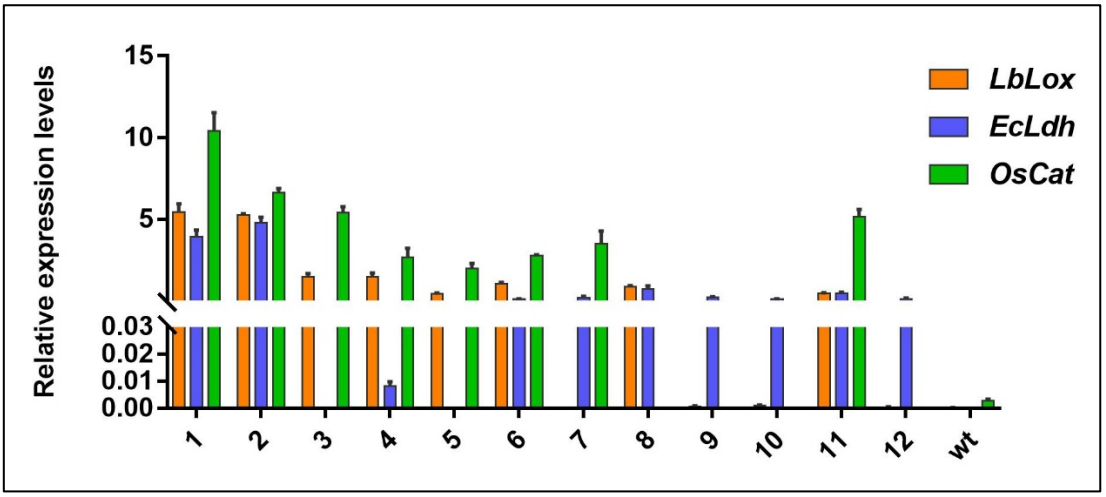

Figure 5.5 Relative expression levels of *LbLox*, *EcLdh*, and *OsCat* in transgenic callus lines. Relative mRNA expression levels (normalized to *OsActin* mRNA) are means ± SD (*n* = 3 technical replicates). 1-12, independent transgenic lines; wt, wildtype callus.

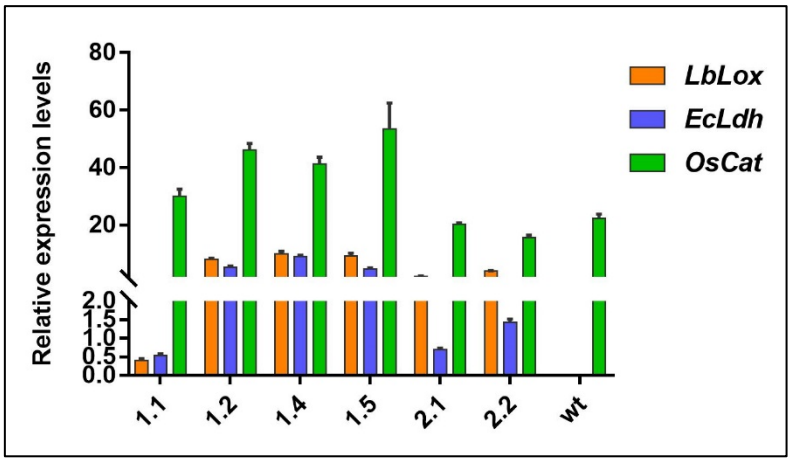

Figure 5.6 Relative expression levels of *LbLox*, *EcLdh*, and *OsCat* in transgenic plants. Relative mRNA expression levels (normalized to *OsActin* mRNA) are means ± SD (*n*=3 technical replicates). 1.1, 1.2, 1.4 and 1.5 are four siblings from callus Line 1; 2.1 and 2.2 are two siblings from callus Line 2; wt, wildtype plant.

5.4.4 Analysis of recombinant protein accumulation in transgenic rice callus and plants

Accumulation of *Ec*LDH protein in callus and the corresponding regenerated plants was confirmed by immunoblot analysis using antibodies specific for the detection tags (**Figure 5.7** and **5.8**). The molecular weights of the precursor forms of two different versions (different chloroplast transit peptides and detection tags) of the *Ec*LDH

recombinant protein were 50 kDa (CTP-*Ec*LDH-6xHis) and 53 kDa (RC2-EcLDH-Myc). The correctly processed mature form in chloroplasts is 44 kDa. Signals at the expect size (44 kDa) were observed for both CTP-*Ec*LDH-6xHis and RC2-EcLDH-Myc (**Figure 5.7** and **5.8**). The chloroplast stroma-targeting peptide RC2 had been shown to effectively target nuclear-encoded recombinant proteins into rice chloroplasts (Shen et al. 2017). These results indicated that chloroplast stroma-targeting peptide of the Rubisco small subunit from either *S. tuberosum* (CTP) or rice (RC2) correctly targeted the *Ec*LDH protein to the rice chloroplast. However, I was not able to detect *Lb*LOX, *and Os*CAT using antibodies against the Tag54, even in transgenic lines which had high accumulation of mRNA (callus lines 1, 2 and 11).

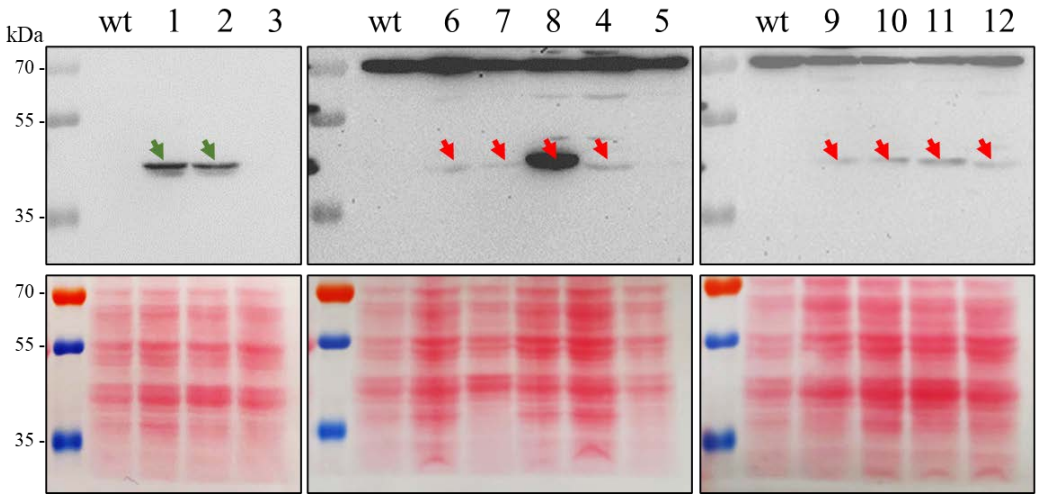

Figure 5.7 Accumulation of *Ec*LDH protein in callus. *Ec*LDH accumulation were detected with antibodies against 6xHis tag (indicated by green arrows) or Myc tag (indicated by red arrows). 1-12, independent transgenic lines; wt, wildtype callus.

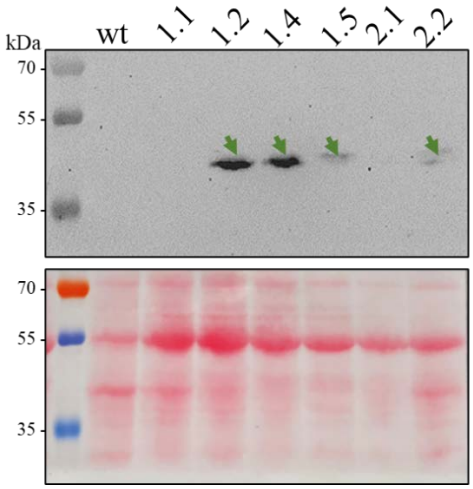

Figure 5.8 Accumulation of *Ec*LDH protein in rice plants. *Ec*LDH accumulation were detected with antibodies against 6xHis tag (indicated by green arrows). 1.1, 1.2, 1.4 and 1.5 are four siblings from callus Line 1; 2.1 and 2.2 are two siblings from callus Line 2; wt, wildtype plant.

5.5 Discussion

To address the increasing demand for food, increasing crop yields is an absolute priority. By increasing the rate of photosynthesis, crop yields can be significantly improved (Zhu et al. 2010). Reduction of photorespiration rates has been viewed as a useful strategy to increase crop carbon fixation and in turn yields in major crops. Reported heterologous photorespiratory bypasses in C3 plants recycle photorespiratory intermediate products, such as glycolate, thereby increasing the efficiency of carbon fixation and reducing energy loss.

In rice, two different synthetic photorespiration bypasses (GOC and GCGT) have been reported. GOC comprises three rice genes (*OsGlo*3, *OsCatC* and *OsOxo3*), which completely oxidize glycolate to two molecules of CO_2 (Shen et al. 2019). The GCGT bypass, similarly to the GOC bypass, also starts with glycolate as the initial substrate but differs in the conversion of two molecules of glycolate to one molecule of glycerate and one molecule of CO_2 through *Os*GLO1, *Ec*CAT, *Ec*GCL, and *Ec*TSR (Wang et al. 2020). Although the introduction of both photorespiratory bypasses in rice resulted in higher yields (up to 27%) in the transgenic plants compared to the wild type, reduced seed setting rates were observed (Shen et al. 2019; Wang et al. 2020).

In this study, a novel oxygen scavenging pathway comprising *Lb*LOX, *Ec*LDH and *Os*CAT was designed and introduced into rice to reduce photorespiration. This oxygen scavenging pathway differs from the earlier photorespiratory bypasses in that it is initiated by oxygen consumption. Photorespiration and the Rubisco oxygenase reaction were

strongly reduced when the O_2 concentration at the site of Rubisco was reduced to between 1% and 2% (Wingler et al. 2000). The O_2 scavenging cycle is initiated by the consumption of O_2 by *Lb*LOX and the oxidation of lactate to pyruvate. The cycle is completed by the conversion of pyruvate to lactate by *Ec*LDH. *Os*CAT removes H_2O_2 generated by this pathway. In terms of energy cost, the oxygen scavenging pathway requires two molecules of NADH (equivalent to five molecules of ATP) to scavenge one molecule of oxygen by producing two molecules of H_2O, avoiding the Rubisco oxygenation reaction and photorespiration. This is considerably less energy than that required for the photorespiration and its bypass pathways GCGT and GOC, which have been reported in rice, previously. The energy cost of one Rubisco oxygenation reaction and plant photorespiration is 12.25 molecules of ATP (Peterhansel et al. 2010): 5.75 molecules of ATP are used to reduce 0.5 molecules of CO_2 and refix 0.5 molecules of NH_3 released from the decarboxylase of glycine; 0.5 molecules of ATP is used to phosphorylate 0.5 molecules of glycerate to form 0.5 molecules of PGA. The latter, together with one molecule of PGA formed by the Rubisco oxygenation reaction requires six molecules of ATP in the Calvin cycle for the regeneration of ribulose-1,5 bisphosphate. The energy cost of the reported rice photorespiratory bypass pathways is 11.75 (GCGT) and 20 (GOC) molecules of ATP (Peterhansel et al. 2013; Wang et al. 2020). Both bypass pathways avoid NH_3 release and therefore, do not require NH_3 refixation, which save 1.75 molecules of ATP. The energy cost of the GCGT pathway is 1.25 molecules of ATP to convert glycolate to 0.5 molecules of CO_2 and 0.5 molecules of glycerate. The latter is phosphorylated to 0.5 molecules of PGA using 0.5 molecules of ATP. The energy cost to reduce 0.5 molecules of CO_2 and 1.5 molecules of PGA (0.5 molecules formed by GCGT bypass and one molecule formed by the Rubisco oxygenation reaction) is four and six molecules of ATP, respectively (Wang et al. 2020). In the GOC pathway, glycolate is completely oxidized to two molecules of CO_2 and the energy cost of refixing CO_2 is 16 molecules of ATP. The energy cost for reducing one PGA molecule (formed during the Rubisco oxygenation reaction) is four molecules of ATP (Shen et al. 2019).

I generated transgenic rice plants co-expressing *LbLox, EcLdh* and *OsCat* at the mRNA. Although expression of the three genes was controlled by the same constitute promoter (maize *Ubiquitin-1* promoter plus first intron), different mRNA abundances was observed for *OsCat*, *LbLox* and *EcLdh*. At the protein level, I was only able to detect *Ec*LDH. Two forms of this recombinant protein, differing in the targeting peptide to direct protein accumulation in the chloroplasts were equally effective (Shen et al. 2017). Preliminary

data in transgenic potato plants show that accumulation of *Ec*LDH protein and *OsCat* and *LbLox* mRNA resulted in 22-95% increase in tuber yield, even though the latter two proteins were not detectable (Dr Greta Nölke, personal communication). Further experiments are required to determine *Os*CAT and *Lb*LOX enzymatic activity and the replacement of the assay tags (such as Myc or 6xHis tags) to verify the functionality of the complete oxygen scavenging pathway in rice study.

5.6 Conclusions

A novel oxygen scavenging pathway was constructed and introduced into rice to minimize photorespiration thus enhancing photosynthesis. I generated transgenic rice plants containing and expressing (mRNA level) *LbLox, EcLdh* and *OsCat*. I also confirmed *Ec*LDH protein accumulation (targeted to the chloroplast) but no other recombinant protein in transgenic plants. The *Lb*LOX and *Os*CAT accumulation will be confirmed by further enzymatic assay and replacement of detection tags to ensure the full pathway introduction into rice. Subsequently, further analysis will focus on determining whether the oxygen scavenging pathway reduces photorespiration without affecting plant development.

5.7 References

Abadie C, Tcherkez G. 2019. Plant sulphur metabolism is stimulated by photorespiration. *Commun. Biol.* **2**: 1-7.

Andersson I, Backlund A. 2008. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**: 275-291.

Andersson I, Taylor TC. 2003. Structural framework for catalysis and regulation in ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Arch. Biochem.* **414**: 130-140.

Baker TS, Eisenberg D, Eiserling F. 1977. Ribulose bisphosphate carboxylase: a two-layered, square-shaped molecule of symmetry 422. *Science* **196**: 293-295.

Brown NJ, Parsley K, Hibberd JM. 2005. The future of C4 research–maize, Flaveria or Cleome? *Trends Plant Sci.* **10**: 215-221.

Busch FA. 2020. Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism. *Plant J.* **101**: 919-939.

Dehigaspitiya P, Milham P, Ash GJ, Arun-Chinnappa K, Gamage D, Martin A, Nagasaka S, Seneweera S. 2019. Exploring natural variation of photosynthesis in a site-specific manner: evolution, progress, and prospects. *Planta* **250**: 1033-1050.

Dong J, Taylor J, Latour D, Iuchi S, Lin E. 1993. Three overlapping lct genes involved in L-lactate utilization by Escherichia coli. *J. Bacteriol.* **175**: 6671-6678.

Douce R, Neuburger M. 1999. Biochemical dissection of photorespiration. *Curr. Opin. Plant Biol.* **2**: 214-222.

Edwards EJ. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytol.* **223**: 1742-1755.

Ellis RJ. 1979. The most abundant protein in the world. *Trends Biochem. Sci.* **4**: 241-244.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D, Milo R. 2019. Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochem.* **58**: 3365-3376.

Flügel F, Timm S, Arrivault S, Florian A, Stitt M, Fernie AR, Bauwe H. 2017. The Photorespiratory Metabolite 2-Phosphoglycolate Regulates Photosynthesis and Starch Accumulation in Arabidopsis. *Plant Cell* **29**: 2537-2551.

Gerbaud A, Andre M. 1979. Photosynthesis and Photorespiration in Whole Plants of Wheat. *Plant Physiol.* **64**: 735-738.

Heinl S, Wibberg D, Eikmeyer F, Szczepanowski R, Blom J, Linke B, Goesmann A, Grabherr R, Schwab H, Pühler A. 2012. Insights into the completely annotated genome of Lactobacillus buchneri CD034, a strain isolated from stable grass silage. *J. Biotechnol.* **161**: 153-166.

Hibberd JM, Sheehy JE, Langdale JA. 2008. Using C4 photosynthesis to increase the yield of rice—rationale and feasibility. *Curr. Opin. Plant Biol.* **11**: 228-231.

Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* **291**: 513-515.

Kebeish R, Niessen M, Thiruveedhi K, Bari R, Hirsch H-J, Rosenkranz R, Stäbler N, Schönfeld B, Kreuzaler F, Peterhänsel C. 2007. Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana. *Nat. Biotechnol.* **25**: 593-599.

Kelly GJ, Latzko E, Gibbs M. 1976. Regulatory aspects of photosynthetic carbon metabolism. *Annu. Rev. Plant Physiol.* **27**: 181-205.

Kohn LD, Kaback HR. 1973. Mechanisms of Active Transport in Isolated Bacterial Membrane Vesicles: XV. PURIFICATION AND PROPERTIES OF THE MEMBRANE-BOUND d-LACTATE DEHYDROGENASE FROM ESCHERICHIA COLI. *Biol. Chem.* **248**: 7012-7017.

Leegood RC, Lea PJ, Adcock MD, Häusler RE. 1995. The regulation and control of photorespiration. *J. Exp. Bot* **46**: 1397-1414.

Li G, Mao H, Ruan X, Xu Q, Gong Y, Zhang X, Zhao N. 2003. An improved equation and assay for determining the CO2/O2 specificity for Rubisco. *Photosynth. Res.* **75**: 287-292.

Li X, Wang P, Li J, Wei S, Yan Y, Yang J, Zhao M, Langdale JA, Zhou W. 2020. Maize GOLDEN2-LIKE genes enhance biomass and grain yields in rice by improving photosynthesis and reducing photoinhibition. *Commun. Biol.* **3**: 151.

Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**: 547-550.

Maier A, Fahnenstich H, Von Caemmerer S, Engqvist MK, Weber AP, Flügge U-I, Maurino VG. 2012. Transgenic introduction of a glycolate oxidative cycle into A. thaliana chloroplasts leads to growth improvement. *Front. Plant Sci.* **3**: 38.

Mann CC. 1999. Genetic Engineers Aim to Soup Up Crop Photosynthesis. *Science* **283**: 314-316.

Mao Y, Catherall E, Díaz-Ramos A, Greiff GR, Azinas S, Gunn LH, McCormick AJ. 2022. The small subunit of Rubisco and its potential as an engineering target. *J. Exp. Bot.* **72**: 543-561.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**: 292-292.

Nölke G, Houdelet M, Kreuzaler F, Peterhänsel C, Schillberg S. 2014. The expression of a recombinant glycolate dehydrogenase polyprotein in potato (S olanum tuberosum) plastids strongly enhances photosynthesis and tuber yield. *Plant Biotechnol. J.* **12**: 734-742.

Ogren WL. 1984. Photorespiration: pathways, regulation, and modification. *Annu. Rev. Plant Physiol.* **35**: 415-442.

Orr DJ, Worrall D, Lin MT, Carmo-Silva E, Hanson MR, Parry MAJ. 2019. Hybrid Cyanobacterial-Tobacco Rubisco Supports Autotrophic Growth and Procarboxysomal Aggregation1 [CC-BY]. *Plant Physiol.* **182**: 807-818.

Parry M, Madgwick P, Carvalho J, Andralojc P. 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci.* **145**: 31.

Peterhansel C, Blume C, Offermann S. 2013. Photorespiratory bypasses: how can they work? *J. Exp. Bot.* **64**: 709-715.

Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kurkcuoglu S, Kreuzaler F. 2010a. *Arabidopsis Book*: e0130.

Rasche S, Martin A, Holzem A, Fischer R, Schinkel H, Schillberg S. 2011. One-step protein purification: use of a novel epitope tag for highly efficient detection and purification of recombinant proteins. *Open Biotechnol. J.* **5**.

Rausch T, Wachter A. 2005. Sulfur metabolism: a versatile platform for launching defence operations. *Trends Plant Sci.* **10**: 503-509.

Rothschild LJ. 2008. The evolution of photosynthesis… again? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **363**: 2787-2801.

Sharpe RM, Offermann S. 2014. One decade after the discovery of single-cell C4 species in terrestrial plants: what did we learn about the minimal requirements of C4 photosynthesis? *Photosynth. Res.***19**: 169-180.

Shen B-R, Wang L-M, Lin X-L, Yao Z, Xu H-W, Zhu C-H, Teng H-Y, Cui L-L, Liu EE, Zhang J-J et al. 2019. Engineering a New Chloroplastic Photorespiratory Bypass to Increase Photosynthetic Efficiency and Productivity in Rice. *Mol. Plant* **12**: 199-214.

Shen B-R, Zhu C-H, Yao Z, Cui L-L, Zhang J-J, Yang C-W, He Z-H, Peng X-X. 2017. An optimized transit peptide for effective targeting of diverse foreign proteins into chloroplasts in rice. *Sci. Rep.* **7**: 46231.

Sudhakar D, Duc LT, Bong BB, Tinjuangjun P, Maqbool SB, Valdez M, Jefferson R, Christou P. 1998. An Efficient Rice Transformation System Utilizing Mature Seed-derived Explants and a Portable, Inexpensive Particle Bombardment Device. *Transgenic Res.* **7**: 289-294.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**: 1515-1524.

Tarmy EM, Kaplan NO. 1968. Chemical Characterization of D-Lactate Dehydrogenase from Escherichia coli B. *J. Biol. Chem.* **243**: 2579-2586.

Valdez M, Cabrera-Ponce JL, Sudhakar D, Herrera-Estrella L, Christou P. 1998. Transgenic Central American, West African and Asian Elite Rice Varieties Resulting from Particle Bombardment of Foreign DNA into Mature Seed-derived Explants Utilizing Three Different Bombardment Devices. *Ann. Bot.* **82**: 795-801.

Valegård K, Hasse D, Andersson I, Gunn LH. 2018. Structure of Rubisco from Arabidopsis thaliana in complex with 2-carboxyarabinitol-1, 5-bisphosphate. *Acta Crystallogr. D* **74**: 1-9.

Wang L-M, Shen B-R, Li B-D, Zhang C-L, Lin M, Tong P-P, Cui L-L, Zhang Z-S, Peng X-X. 2020. A Synthetic Photorespiratory Shortcut Enhances Photosynthesis to Boost Biomass and Grain Yield in Rice. *Mol. Plant* **13**: 1802-1815.

Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA. 2017. Re-creation of a key step in the evolutionary switch from C3 to C4 leaf anatomy. *Curr. Biol.* **27**: 3278-3287. e3276.

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009. GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis *Plant Cell* **21**: 1109-1128.

Wingler A, Lea PJ, Quick WP, Leegood RC. 2000. Photorespiration: metabolic pathways and their role in stress protection. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **355**: 1517-1529.

Yamori W, Hikosaka K, Way DA. 2014. Temperature response of photosynthesis in C3, C4, and CAM plants: temperature acclimation and temperature adaptation. *Photosynth. Res.* **119**: 101-117.

Zhu X-G, Long SP, Ort DR. 2010. Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* **61**: 235-261.

GENERAL DISCUSSION

General discussion

Rice is one of the most important cereal crops in the world and the primary food staple for over half of the world's population. To meet the projected shortfall in rice productivity, the global N fertilizer demand is approximately 110.8 million metric tons, with ca: 65% of all applied N fertilizer being lost to the environment with severe environmental problems such as groundwater contamination and soil acidification (Omara et al. 2019; Fernández 2022). My thesis focused on two different critical aspects of agricultural sustainability, namely to engineer rice with the major components of nitrogenase as a first step towards engineering biology nitrogen fixation in plants and to address the negative contribution of the process of photorespiration on photosynthesis through the engineering of an oxygen scavenging pathway to improve photosynthetic efficiency.

The two major bottlenecks in engineering BNF in plants are the O_2 sensitivity of nitrogenase and its accessory proteins, and the complexity of the metal cluster biosynthesis mechanism of nitrogenase (Shah and Brill 1973; Burén et al. 2020). I was able to make major progress towards addressing the constraint of O_2 sensitivity. I generated transgenic rice plants which accumulated active nitrogen fixation proteins for the first time. My experiments focussed on the engineering of NifB, NifH and NifD in transgenic rice. NifB catalyzes the first committed step in the biosynthesis of FeMo-co (the active-site cofactor of nitrogenase) and the latter two genes encode the nitrogenase reductase Fe protein and nitrogenase MoFe protein α-subunit. The *M. thermautotrophicus* and *M. infernus* NifB, *A. vinelandii* and *H. thermophilus* NifH and *A. vinelandii* NifD all accumulated as soluble proteins in transgenic rice. The NifB and NifH proteins were correctly targeted and processed in rice mitochondria by the N-terminal mitochondrial targeting peptides Cox4 which had been used earlier to direct Nif proteins to *S. cerevisiae* and *N. benthamiana* mitochondria and eGFP to rice mitochondria (Burén et al. 2019; Baysal et al. 2020; Jiang et al. 2021). Cox4 also targeted *A. vinelandii* NifD to mitochondria but protein degradation was observed which prevented the formation of active protein. Although NifD degradation was prevented in *N. behtamiana* by substituting one or two amino acids (generated variants NifD Y100Q, Y100Q/Y101T and Y100K) within the predicted mitochondrial processing peptidase (MPP) cleavage site ($R^{97}R^{98}N^{99}\downarrow Y^{100}Y^{101}T^{102}$), my results showed that the same variants remained susceptible to MPP and accumulated only as truncated proteins in rice (Allen et al. 2020). Interestingly, the same NifD variant Y100Q also behaved differently between *S. cerevisia*e and *N. benthamiana* in terms of resistance to degradation (Allen et al. 2020).

These differences can probably be attributed to the different cleavage efficiency of MPPs in different species (Xiang et al. 2020), but a deeper understanding of the processing of rice mitochondrial precursor proteins is still needed to produce functional NifD.

The entire *Nif* cluster comprises *NifH, NifD, NifK, NifT, NifY, NifE, NifN, NifX, NifU, NifS, NifV, NifW, NifZ, NifM*, and *NifF* which are involved in the assembly and activity of nitrogenase and its metal cofactors in *A. vinelandii* (Jacobson et al. 1989). To simplify the complexity of reconstituting functional nitrogenases in plants, *NifD, NifK, NifH, NifE, NifN* and *NifB* were postulated as the minimum set of genes required for the formation of functional nitrogenase (Curatti and Rubio 2014; Burén et al. 2020). This is because some of the components, including NifU and NifS which provide the [4Fe-4S] clusters for the biosynthesis of the metal cofactors appear to have counterparts in plant mitochondria (Balk and Lobréaux 2005). However, the as-isolated rice-produced NifB proteins without co-expressing NifU and NifS were colourless and inactive in FeMo-co synthesis, and function was restored by loading [4Fe-4S] clusters *in vitro*. Similarly, the as-isolated *H. thermophilus* NifH protein produced in rice plants co-expressing NifH and NifM was colourless with low activity. However, activity increased 9-fold after reconstitution with [4Fe-4S] from NifU *in vitro*. My results indicate that rice-produced Nif proteins were correctly fold but not (fully) matured with their [4Fe-4S] cluster. This is likely due to the insufficient insertion of the [4Fe-4S] cluster into Nif proteins by the endogenous rice mitochondrial cluster assembly machinery, or instability of the clusters caused by O_2 (Jiang et al. 2021). Future research should focus on improving Fe transport, [Fe-S] cluster biosynthesis, delivery to Nif proteins and protection.

Producing a functional nitrogenase was predicted to affect the energy balance of plants significantly (Hardy and Havelka 1975). Analysis of *H. thermophilus* NifH protein expression in the T1 generation confirmed that *H. thermophilus* NifH was stably inherited and expressed in the progeny. There were no differences between T1 progeny expressing *H. thermophilus* NifH and negative segregants in terms of growth and development, which was normal. Thus, functional Fe protein accumulation does not compromise plant growth and development. It was more difficult to generate transgenic plants expressing *M. thermautotrophicus* NifB compared to its *M. infernus* counterpart. The reasons for this difference remain unclear but my results suggest that NifB might interfere with other essential developmental processes in the cells or compete for essential precursors in metabolic processes with common precursors. While moving forward with the

nitrogenase engineering process, it would be desirable to circumvent this problem, for example, by using tissue-specific or regulated promoters.

Rubisco is a bifunctional enzyme. It catalyses two competing reactions, carboxylation and oxygenation of ribulose 1,5-diphosphate, at the same active site (Li et al. 2003). The carboxylation reaction converts atmospheric CO_2 and ribulose 1,5-diphosphate to 3-phosphoglyceric acid, and this is the predominant mechanism that inorganic carbon is converted to organic molecules, entering the biosphere (Kelly et al. 1976). However, the oxygenation of ribulose 1,5-diphosphate produces 2-phosphoglycolate (2PG) which is not part of the Calvin cycle. In addition, 2PG inhibits triose-phosphate isomerase and sedoheptulose 1,7-bisphosphate phosphatase in the Calvin cycle (Flügel et al. 2017). Photorespiration recycles 2PG to 3-phosphoglyceric acid (3PGA), which then enters the Calvin cycle. However, at the same time photorespiration is wasteful because the assimilated carbon and nitrogen are lost as CO_2 and NH_3. The balance of the carboxylation to the oxygenation reaction depends on the ratio of CO_2 and O_2 concentration around Rubisco (Parry et al. 2007). Reduction of photorespiration rates is viewed as a useful strategy to increase carbon fixation in crops because photorespiration results in a loss of 20-50% of the carbon fixed by photosynthesis in C_3 plants (South et al. 2019). I engineered a novel oxygen scavenging pathway comprising *LbLox*, *EcLdh* and *OsCat* into rice to reduce photorespiration and thus increase photosynthetic efficiency. The O_2 scavenging cycle is initiated by the consumption of O_2 by *Lb*LOX and the oxidation of lactate to pyruvate, and is completed by the conversion of pyruvate to lactate by *Ec*LDH. The reported novel photorespiration bypasses (GOC and GCGT) engineered in rice chloroplasts commence with glycolate as the substrate. Both bypasses result in the release of CO_2 from the metabolism of glycolate in the vicinity of Rubisco, without the loss of fixed N (in plant photorespiration process, in mitochondria) (Shen et al. 2019; Wang et al. 2020). The oxygen scavenging pathway has a low energy cost, requiring two molecules of NADH (equivalent to five molecules of ATP) to scavenge one molecule of oxygen. In contrast, the GCGT and GOC bypasses require 11.75 and 20 molecules of ATP, respectively, to metabolite one molecule of 2PG (Peterhansel et al. 2013; Wang et al. 2020). Although the *LbLox*, *EcLdh*, and *OsCat* mRNA and *Ec*LDH protein accumulation confirmed the successful introduction of the O_2 scavenging pathway in transgenic rice plants, challenges still remain. Further experiments are needed to confirm accumulation of *Lb*LOX and *Os*CAT to ensure functionality of the entire pathway in rice. Once achieved, further experiments are needed to ascertain if the oxygen scavenging pathway can reduce

photorespiration sufficiently to permit substantial increases in photosynthesis, without compromising plant development and fertility.

References

Allen RS, Gregg CM, Okada S, Menon A, Hussain D, Gillespie V, Johnston E, Devilla R, Warden AC, Taylor M et al. 2020. Plant expression of NifD protein variants resistant to mitochondrial degradation. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 23165-23173.

Balk J, Lobréaux S. 2005. Biogenesis of iron–sulfur proteins in plants. *Trends Plant Sci.* **10**: 324-331.

Baysal C, Pérez-González A, Eseverri Á, Jiang X, Medina V, Caro E, Rubio L, Christou P, Zhu C. 2020. Recognition motifs rather than phylogenetic origin influence the ability of targeting peptides to import nuclear-encoded recombinant proteins into rice mitochondria. *Transgenic Res.* **29**: 37-52.

Burén S, Jiménez-Vicente E, Echavarri-Erasun C, Rubio LM. 2020. Biosynthesis of nitrogenase cofactors. *Chem. Rev.* **120**: 4921-4968.

Burén S, Pratt K, Jiang X, Guo Y, Jimenez-Vicente E, Echavarri-Erasun C, Dean DR, Saaem I, Gordon DB, Voigt CA et al. 2019. Biosynthesis of the nitrogenase active-site cofactor precursor NifB-co in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 25078-25086.

Curatti L, Rubio LM. 2014. Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. *Plant Sci.* **225**: 130-137.

Fernández L. 2022. Global fertilizer demand by nutrient 2011-2022. *Statista*.

Flügel F, Timm S, Arrivault S, Florian A, Stitt M, Fernie AR, Bauwe H. 2017. The Photorespiratory Metabolite 2-Phosphoglycolate Regulates Photosynthesis and Starch Accumulation in Arabidopsis. *Plant Cell* **29**: 2537-2551.

Hardy R, Havelka U. 1975. Nitrogen Fixation Research: A Key to World Food? Investigators in a variety of disciplines are searching for new technologies for producing fixed nitrogen. *Science* **188**: 633-643.

Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR. 1989. Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. *J. Bacteriol.* **171**: 1017-1027.

Jiang X, Paya-Tormo L, Coroian D, Garcia-Rubio I, Castellanos-Rueda R, Eseverri A, Lopez-Torrejon G, Buren S, Rubio LM. 2021. Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N_2-fixation in plants. *Commun. Biol.* **4**: 4.

Kelly GJ, Latzko E, Gibbs M. 1976. Regulatory aspects of photosynthetic carbon metabolism. *Annu. Rev. Plant Physiol.* **27**: 181-205.

Li G, Mao H, Ruan X, Xu Q, Gong Y, Zhang X, Zhao N. 2003. An improved equation and assay for determining the CO2/O2 specificity for Rubisco. *Photosynth. Res.* **75**: 287-292.

Omara P, Aula L, Oyebiyi F, Raun WR. 2019. World Cereal Nitrogen Use Efficiency Trends: Review and Current Knowledge. *Agrosyst. Geosci. Environ.* **2**: 180045.

Parry M, Madgwick P, Carvalho J, Andralojc P. 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci* **145**: 31.

Peterhansel C, Blume C, Offermann S. 2013. Photorespiratory bypasses: how can they work? *J. Exp. Bot.* **64**: 709-715.

Shah VK, Brill WJ. 1973. Nitrogenase. IV. Simple method of purification to homogeneity of nitrogenase components from *Azotobacter vinelandii*. *Biochim. Biophys. Acta.* **305**: 445-454.

Shen B-R, Wang L-M, Lin X-L, Yao Z, Xu H-W, Zhu C-H, Teng H-Y, Cui L-L, Liu EE, Zhang J-J et al. 2019. Engineering a New Chloroplastic Photorespiratory Bypass to Increase Photosynthetic Efficiency and Productivity in Rice. *Mol. Plant* **12**: 199-214.

South PF, Cavanagh AP, Liu HW, Ort DR. 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* **363**: eaat9077.

Wang L-M, Shen B-R, Li B-D, Zhang C-L, Lin M, Tong P-P, Cui L-L, Zhang Z-S, Peng X-X. 2020. A Synthetic Photorespiratory Shortcut Enhances Photosynthesis to Boost Biomass and Grain Yield in Rice. *Mol. Plant* **13**: 1802-1815.

Xiang N, Guo C, Liu J, Xu H, Dixon R, Yang J, Wang Y-P. 2020. Using synthetic biology to overcome barriers to stable expression of nitrogenase in eukaryotic organelles. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 16537-16545.

GENERAL CONCLUSIONS

General conclusions

1. *M. infernus* NifB (NifB^{Mi}), *M. thermautotrophicus* NifB (NifB^{Mt}), *A. vinelandii* NifH (NifH^{Av}) and *H. thermophilus* NifH (NifH^{Ht}) accumulated in a soluble form in engineered rice callus and regenerated plants. Protein accumulation was stable at least to homozygosity.
2. Nif proteins were purified effectively from rice callus and regenerated plants by strep-tag affinity chromatography. The yields of NifB^{Mi} and NifH^{Ht} from fresh leaves were 1.4 and 50 µg per 100 g, respectively. Yields of NifB^{Mi}, NifB^{Mt}, NifH^{Av} and NifH^{Ht} were 44, 87, 90 and 700 µg per 100 g fresh callus, respectively.
3. The *S. cerevisiae* mitochondrial targeting peptide cytochrome c oxidase subunit IV was correctly processed and effectively directed NifB^{Mi}, NifB^{Mt}, NifH^{Av} and NifH^{Ht} to rice mitochondria.
4. Purified NifB^{Mi} and NifB^{Mt} were functional after loading with [4Fe-4S] clusters demonstrating that the two recombinant proteins were correctly folded and able to acquire the metal cluster.
5. As-isolated NifH^{Ht} produced in rice plants was functional confirming that the endogenous mitochondrial [Fe-S] cluster assembly machinery was able to transfer [4Fe-4S] clusters to NifH^{Ht}.
6. Purified NifH^{Ht} exhibited nine-fold higher activity compared to its as-isolated counterpart, after loading with [4Fe-4S] clusters indicating that NifH^{Ht} was correctly folded but not fully mature in rice mitochondria.
7. The NifD^{Av} variants were susceptible to cleavage by mitochondrial processing peptidase, resulting in the formation of only truncated proteins.
8. NifD^{Av} variants appear to be processed differently in the mitochondria in different species.
9. A novel oxygen scavenging pathway comprising *L. buchneri* lactate oxidase (*Lb*LOX), *E. coli* lactate dehydrogenase (*Ec*LDH) and *O. sativa* catalase (*Os*CAT) was introduced into rice.
10. The chloroplast stroma-targeting peptide of the Rubisco small subunit from either *S. tuberosum* (CTP) or rice (RC2) correctly targeted the *Ec*LDH protein to the rice

chloroplast. Further experiments are required to determine if CTP and/or RC2 are also effective in targeting *Lb*LOX and *Os*CAT to the chloroplast.