Declaration of Originality

I confirm that all materials in this thesis are my own work, and any sources of information used therein have been appropriately cited.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

First of all, I am grateful to my supervisor, Dr Patrik Jones for providing the opportunity to work on this exciting project in an inspiring environment in the heart of London. He was always open to new ideas, guiding me through my path to becoming an independent scientist. I also would like to thank the European Union's FP7 People Programme (Marie Curie Actions, 317184) and the BBSRC sLoLa grant (BB/N003608/1) for providing funding for this project.

I thank the entire PHOTO.COMM consortium for the excellent training events and the great community, especially Dr Kristine Groth Kirkensgaard and Ms Karin Norris for arranging the project transfer from Turku to London.

I am thankful to Professor Enrique Flores at University of Seville for kindly providing the *Anabaena* sp. PCC 7120 base strains and Mr Anthony Riseley, a fellow PhD student at the University of Cambridge for his essential contribution to investigating the potential of the nitrogen excretion strains. I am also grateful to Dr Dennis Nürnberg at Imperial College London for all his help and endless patience in teaching me the conjugative transformation of cyanobacteria.

I would like to thank the members of the MME group at Imperial College for the friendly atmosphere and their contribution to my work in one way or another. I especially thank Ms Phoebe Tickell for her great support in the laboratory by making my everyday life so much easier.

I am grateful to Dr András Pásztor and Zsu Horváth for supporting us in many different ways while in Finland, and also afterwards in spite of the time difference.

Finally, I would like to thank my lab buddies, Ms Marine Valton, Dr Paulina Bartasun and Dr John Rowland for their presence; the amazing conversations, the laughs, as well as the invaluable discussions on many different matters of my thesis. It would not have been the same without them.

Lastly, I am thankful to the SCR at Imperial College London for the unmatched fish and chips. Some might even say it is the best one in London.

To Lilla, Luca and Laura

for supporting me all the way on this great endeavour;

and to Máté

for patiently waiting until it has been completed.

Modelling and Engineering *Anabaena* sp. PCC 7120 for Nitrogen Excretion

David Malatinszky

Imperial College London

Department of Life Sciences

Submitted for the Degree of Doctor of Philosophy

2017

Abstract

Nitrogen is an essential element for every organism on Earth. Modern agricultural activity depletes soil in nitrogen much faster than it is naturally replenished. Therefore, fertilization is key for feeding a fast-growing population. However, the use efficiency of fertilizer nitrogen is only about 60%. The rest of reactive nitrogen leaches to the environment, and the pollution caused demands modern societies tens of billions of euros annually as remediation costs. Rationalisation of current agricultural practices is essential, including a more targeted application of fertilizers, to tackle the nitrogen crisis. One way is the use of nitrogen-fixing organisms as biofertilizer in close association to agricultural crops. In this thesis, a stoichiometric model was reconstructed for the heterocystous nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120 to understand the nature of metabolite exchange between its photosynthetic and diazotrophic cell types, and design metabolic engineering strategies for nitrogen excretion. Using flux balance analysis of diazotrophically grown filaments, excretion of ammonia followed by urea achieved the highest molar nitrogen flux. To achieve a similar effect experimentally, glutamine synthetase (GS) was inhibited using L-methionine sulfoximine. It was possible to accumulate about 760 µM ammonia in a 7-day assay with stagnating growth. For stable excretion, GS has been replaced for an active-site mutant exhibiting decreased specific activity for ammonia. The metabolic changes have been implemented in both wild type and an ammonium uptake transporter mutant (Δ*amt*). The resulting strains displayed increased ammonia excretion up to about 8-fold compared to the wild type. Furthermore, IF7A, a small oligopeptide controlling the activity of GS, has been overexpressed under four different promoters. The constructs driven by two of these promoters, P_{nifHDK} and P_{petE} enabled the growth of the non-diazotrophic alga *Chlorella vulgaris* at 68% of that of the algal monoculture on combined nitrogen. Overall, the ammonia-excreting strains provided an important proof-of-principle for the development of more efficient biofertilizers in future agriculture.

Table of Contents

Abstract 9

List of Figures 14

List of Tables 17

Abbreviations 18

1 Introduction 22

1.1 Synopsis 22

1.2 Essential nitrogen 24

1.2.1 Nitrogen for life 24

1.2.2 The nitrogen cycle 25

1.2.3 Haber–Bosch ammonia 28

1.3 The Nitrogen Crisis 30

1.3.1 Environmental and health effects 30

1.3.2 Costs and measures 31

1.4 Metabolic modelling of entire organisms 32

1.4.1 Systems biology 32

1.4.2 Flux Balance Analysis 33

1.5 Cyanobacteria 36

1.5.1 Nitrogen assimilation 38

1.5.1.1 Heterocysts and nitrogen fixation 38

1.5.1.2 Ammonia assimilation and excretion 41

1.5.1.3 Nitrogenase enzyme complex 44

1.5.2 Importance of iron 46

1.5.3 Natural and synthetic communities 47

1.6 Challenges in metabolic engineering of *Anabaena* sp. PCC 7120 49

1.6.1 Developing synthetic biology 49

1.6.2 Homologous recombination favours single recombination 51
1.6.3 Oligoploid cyanobacteria 51
1.6.4 Multicellularity in filamentous cyanobacteria 52
1.7 Aims and objectives 53
2 Materials and methods 56
2.1 Chemicals and reagents 56
2.2 General protocols 56
2.2.1 Incubators and sterile work 56
2.2.2 Laboratory centrifuges 57
2.3 Strains and plasmids 57
2.3.1 Cryopreservation of bacteria 58
2.3.2 Cryopreservation of cyanobacteria 58
2.4 Media and culturing conditions 59
2.4.1 Liquid LB (Luria-Bertani) medium 59
2.4.2 LB agar medium 59
2.4.3 Liquid BG-11 medium 59
2.4.4 BG-11 agar medium 60
2.4.5 Nitrate-free medium 60
2.4.6 Bacterial cultivation 60
2.4.7 Cultivation of cyanobacteria 61
2.5 Molecular biology methods 61
2.5.1 Amplification of DNA fragments 61
2.5.2 Agarose gel electrophoresis 62
2.5.3 DNA extraction and purification 63
2.5.4 Purification of genomic DNA from cyanobacteria 63
2.5.5 Quantification of DNA concentration 64
2.5.6 Site-directed mutagenesis 65
2.5.7 Overlap extension PCR 65
2.5.8 Restriction cloning 66
2.5.9 Gibson assembly 66
2.6 Genetic manipulation techniques 67
2.6.1 Preparation of chemically competent cells 67
2.6.2 Heat-shock transformation of bacteria 67
2.6.3 Triparental conjugation of *Anabaena* sp. PCC 7120 68

2.6.4 Isolation of cyanobacterial strains 69
2.6.5 Colony PCR of cyanobacteria 69
2.6.6 Filament fragmentation by sonication 70
2.7 Analytical methods 70
2.7.1 Microscopic analysis of cyanobacterial cultures 70
2.7.2 Peptide preparation for LC-MS/MS from cyanobacteria 71
2.7.3 Triple quadrupole mass spectrometry analysis of proteotypic tryptic peptides 72
2.7.4 Glutamine synthetase (GS) bioactivity assay 74
2.7.5 Ammonia quantification by the Willis method 75
2.7.6 Ammonia quantification by a commercial kit 76
2.7.7 Detection of siderophore activity by chrome azurol S assay 76
2.8 Computational methods 77
2.8.1 Metabolic reconstruction 77
2.8.2 Flux Balance Analyses 80
2.8.3 Experimental carbon source evaluation 81
2.8.4 Sequence analysis 82
3 Modelling the metabolism of *Anabaena* sp. PCC 7120 83
3.1 Introduction 83
3.2 Results 85
3.2.1 Reconstruction of the metabolic network 85
3.2.2 Filament representation and the biomass model 86
3.2.3 Characterization of the reconstruction 90
3.2.3.1 Autotrophic growth 90
3.2.3.2 Comparison of nitrogen sources 94
3.2.3.3 Mixotrophic growth by the single-cell model 97
3.2.4 Predictive ability of the single-cell model 99
3.2.5 Metabolite exchange by the two-cell model 100
3.2.6 Prediction of gene essentiality 107
3.2.7 Evaluation of nitrogen excretion 109
3.3 Discussion 110
4 Engineering *Anabaena* sp. PCC 7120 for ammonia excretion 119
4.1 Introduction 119
4.2 Results 122
4.2.1 Glutamine synthetase inhibition 122

4.2.2 Strategy A – Replacement of the *glnA* gene (alr2328) by an active-site mutant 125
4.2.2.1 Cloning strategies for replacing wild type *glnA* for *glnA*[p.D52S] 126
4.2.2.2 Assembly of constructs 129
4.2.2.3 Isolation of *glnA*[p.D52S] active-site mutants 133
4.2.2.4 Genetic segregation 135
4.2.2.5 Ammonia excretion 142
4.2.3 Strategy B – Overexpression of IF7A from a self-replicative vector 143
4.2.3.1 Isolation of IF7A overexpression strains 144
4.2.3.2 Evaluation of strain growth 148
4.2.3.3 Glutamine synthetase bioactivity in nitrogen-depleted and replete conditions 151
4.2.3.4 Ammonia production 159
4.2.3.5 Co-cultivation with microalgae 161
4.2.3.6 Long-term stability of the ammonia producing strains 164
4.2.4 Strategy C – Overexpression of *gifA* from a neutral site and knockout of *nsiR4* 166
4.2.4.1 Cloning strategies and assembly of constructs 167
4.2.4.2 Isolation of double recombinants 171
4.3 Discussion 174
5 Systematic study of the schizokinen operon and its genes 185
5.1 Introduction 185
5.2 Results 187
5.2.1 Bioinformatics study of the schizokinen cluster 188
5.2.2 Structure of the cluster 193
5.2.3 Knockout and complementation strategies 195
5.2.4 Isolation of schizokinen mutants 202
5.2.4.1 Schizokinen knockout 202
5.2.4.2 Complementation strains 205
5.2.5 Characterization of the *sch* operon mutants 207
5.2.5.1 Chrome azurol S assay for siderophore detection 207
5.2.5.2 Growth characterization of the schizokinen mutant strains 208
5.3 Discussion 212
6 Conclusions 218
Bibliography 224
Appendices 244

List of Figures

Figure 1.1. Elemental composition of microorganisms by Bowen (1966). 25

Figure 1.2. The nitrogen cycle. 27

Figure 1.3. Simplified regulatory network of heterocyst development. 40

Figure 1.4. Interplay between metabolic and regulatory pathways of nitrogen assimilation in *Anabaena* sp. PCC 7120. 43

Figure 2.1. Comparison between *Anabaena* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 (Knoop et al., 2013) stoichiometric models and their improvement over the KEGG database (Kanehisa et al., 2004). 79

Figure 3.1. Compartments considered in the two-cell model. 88

Figure 3.2. Predicted optimal growth rates of *Anabaena* sp. PCC 7120 as a function of light absorption and bicarbonate uptake. 91

Figure 3.3. Growth rates predicted as the function of different transport reactions under diazotrophic conditions. 92

Figure 3.4. Comparison of nitrogen sources in phototrophic growth. 95

Figure 3.5. Predicted growth rates on different carbon sources under mixo- and heterotrophic conditions. 98

Figure 3.6. Correlation between experimental and predicted mixotrophic growth rates. 99

Figure 3.7. Predicted growth rates in response to the number of intercellular exchange reactions in the two-cell model. 102

Figure 3.8. Main metabolic fluxes for the exchange of sucrose, glutamine, glutamate and 2-oxoglutarate. 104

Figure 3.9. Main metabolic fluxes for the exchange of sucrose, fructose, alanine and ammonia. 106

Figure 3.10. Distribution of genes in the genome of *Anabaena* sp. PCC 7120. 108

Figure 3.11. Comparison of metabolic capacity for the excretion of nitrogen compounds. 109

Figure 3.12. Response of growth rate to changes in the biomass fractional composition. 112

Figure 3.13. Comparison of glucose and glycerol assimilatory pathways. 115

Figure 3.14. Sucrose exchange as a function of light harvesting by the HCSC. 117

Figure 4.1. Equilibrium of ammonia and ammonium as a function of pH. 120
Figure 4.2. Nitrogen metabolism in *Anabaena* sp. PCC 7120. 121
Figure 4.3. Ammonia excretion of wild type and Δ*amt* strains due to treatment with 55 µM MSX. 123
Figure 4.4. Overview of three metabolic engineering strategies for extracellular ammonia accumulation. 124
Figure 4.5. DNA constructs assembled to replace *glnA* for *glnA*[p.D52S]. 127
Figure 4.6. Replacement of *glnA* via single recombination followed by intrachromosomal recombination. 129
Figure 4.7. Isolation steps of single recombinants for construct C6. 134
Figure 4.8. Binding site of the allele-specific primer glnA.D52S*-F. 136
Figure 4.9. Colony PCR results at different stages of genetic segregation. 137
Figure 4.10. Effect of sonication on fragment length of segregating *glnA* mutants 138
Figure 4.11. Colony PCR results after addition of sucrose to the selective medium. 140
Figure 4.12. Free ammonia in the supernatants of GlnA active site-mutant isolates. 142
Figure 4.13. Evaluation of *gifA* overexpression isolates by colony PCR. 146
Figure 4.14. Growth characteristics of the IF7A overexpression strains under non-diazotrophic conditions. 149
Figure 4.15. Short-term growth test of IF7A overexpression strains in BG-11_0 under diazotrophic conditions. 150
Figure 4.16. Glutamine synthetase activity of different IF7A overexpression strains under diazotrophic conditions. 152
Figure 4.17. Levels of GS (GlnA) in cell-free extracts of IF7A overexpression strains under diazotrophic conditions. 154
Figure 4.18. Level of GlnA in cell-free extracts of IF7A overexpression mutants under different growth conditions. 155
Figure 4.19. Level of GlnA in the IF7A overexpression strains in response to spiking with ammonia. 157
Figure 4.20. Level of NtcA in the IF7A overexpression strains in response to spiking with ammonium. 158
Figure 4.21. Ammonia production by the IF7A overexpression strains under diazotrophic conditions. 160
Figure 4.22. Evaluation of selected IF7A overexpressing strains in co-cultivation with *Chlorella vulgaris*. 163
Figure 4.23. Long-term stability and ammonia excretion by the IF7A overexpression strains. 165
Figure 4.24. DNA constructs for *gifA* overexpression and nsiR4 knockout. 167
Figure 4.25. Colony PCR results of chromosomal *gifA* overexpression candidates. 171
Figure 4.26. Colony PCR results of *nsiR4* knockout candidates. 172

Figure 4.27. Appearance of single recombinants in sucrose counter-selection. 173
Figure 5.1. Conserved genes in the putative schizokinen pathway. 188
Figure 5.2. Putative pathway for the biosynthesis of schizokinen. 193
Figure 5.3. Structure of the proposed schizokinen operon in *Anabaena* sp. PCC 7120. 194
Figure 5.4. DNA constructs designed and assembled for studying the schizokinen operon. 200
Figure 5.5. Results of overlap extension PCR (SOE) for the assembly of C10, C11 and C12 constructs. 201
Figure 5.6. Colony PCR results for consecutive segregation rounds of the Δ*sch* mutant. 204
Figure 5.7. Colony PCR results of Δ*sch* complementation strains bearing C10, C11 and C12. 206
Figure 5.8. Results of a typical CAS assay on the schizokinen mutant strains. 208
Figure 5.9. Comparison of wild type and Δ*sch* knockout strains' growth under iron limiting conditions 209
Figure 5.10. Growth of schizokinen mutant strains under iron limitation. 211
Figure H-1. Mixotrophic growth of *Anabaena* sp. PCC 7120 on single carbon sources and nitrate as nitrogen source. 265
Figure I-1. Reaction network of the *Anabaena* sp. PCC 7120 two-cell model. 267

List of Tables

Table 2.1. Strains, plasmids and primers acquired from commercial or academic sources. 57

Table 2.2. Program of a typical PCR amplification. 62

Table 2.3. Signature peptides of *Anabaena* sp. PCC 7120 proteins analysed in this work. 73

Table 3.1. Major differences between the two cell types (super-compartments) in the *Anabaena* sp. PCC 7120 model 87

Table 3.2. Biomass composition of *Anabaena* sp. PCC 7120, as represented in the model. 90

Table 4.1. Oligonucleotide primers, DNA constructs and strains used in *glnA* gene replacement. 131

Table 4.2. List of primers, DNA fragments and constructs used to generate IF7A overexpression mutants. 144

Table 4.3 Summary of the IF7A overexpression strains 147

Table 4.4. List of oligonucleotides, constructs and strains for *gifA* genomic overexpression and *nsiR4* knockout. 168

Table 5.1. Similarity scores of homologous genes in Figure 5.1 to genes in the putative schizokinen operon. 190

Table 5.2. Pfam motifs detected in the homologues of *rhbF* and *all0390*. 191

Table 5.3. Pairwise similarity of the last two ligases in the schizokinen and aerobactin pathways. 192

Table 5.4. List of oligonucleotides, DNA constructs and strains 196

Table A-I. Orphan reactions not included in the *Anabaena* sp. PCC 7120 model. 244

Table B-I. Simplifications and assumptions in the *Anabaena* sp. PCC 7120 model 247

Table C-I. Differences of the two super-compartments in the *Anabaena* sp. PCC 7120 model 248

Table D-I. Reactions added to the *Anabaena* sp. PCC 7120 model based on biochemical or bioinformatics evidence in the literature. 249

Table E-I. List of newly annotated genes in the *Anabaena* sp. PCC 7120 model. 252

Table F-I. Transport reactions in the *Anabaena* sp. PCC 7120 model 256

Table F-II. Intracellular exchange reactions in the *Anabaena* sp. PCC 7120 model 261

Table G-I. Reaction equations for biomass components and the biomass objective function 263

Abbreviations

2-og, 2OG	2-oxoglutarate
aacC1	gentamicin 3-N-acetyltransferase conferring resistance to gentamicin
aadA	aminoglycoside resistance protein (confers resistance to spectinomycin and streptomycin)
AAT	aspartate transaminase
ADP	adenosine diphosphate
ala, L-ala	L-alanine
AMP	adenosine monophosphate
AMP	adenosine monophosphate
amt14B	ammonium uptake transporter genes
aphA	acid phosphatase conferring kanamycin resistance
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
*Bam*HI	*Bam*HI restriction endonuclease
BB	backbone
BCA	bicinchoninic acid
BG-11	BG-11 medium
$BG\text{-}11_0$	nitrogen-free BG-11 medium
$BG\text{-}11_{Fe}$	iron-free BG-11 medium
BLAST[X/P/N]	Basic Local Alignment Search Tool
BM_G	biomass flux (growth-associated)
BM_M	biomass flux (maintenance-associated)
BNF	biological nitrogen fixation
BSA	bovine serum albumin
*Bsa*I	*Bsa*I restriction endonuclease
CAS	chrome azurol S
Cb	carbenicillin
CDS	coding sequence
CE	collision energy
CFU	colony forming unit
CI	chloroform-isoamyl alcohol
Cm	chloramphenicol
CO_2	carbon dioxide
COBRA	Constraint-Based Reconstruction and Analysis Toolbox
CS	complementation strain
CSΔ*all0395*	single-gene knockout complementation strains (Δ*all0395–all0396*)
cyp	cyanophycin monomer (L-β-aspartyl-L-arginine)
DHAP	dihydroxyacetone phosphate
DNA	deoxyribonucleic acid
DP	declustering potential
DR	double recombinant
DTT	dithiothreitol
DW	dry weight
E. coli	*Escherichia coli*
EC	Enzyme Commission number
EDTA	ethylenediaminetetraacetic acid
EU	European Union
EV	e-value
FAME	Flux Analysis and Modeling Environment
FBA	flux balance analysis
fdxN	ferredoxin gene
Fe(III)	ferric ion
fhuCDB	ferric hydroxamate transporter operon
FNR	ferredoxin-$NADP^+$-reductase
fru	fructose
fru1,6bP	fructose 1,6-bisphosphate

fru6P	fructose 6-phosphate
fur	ferric uptake regulator genes
GA3P	glyceraldehyde 3-phosphate
GDH	glutamate dehydrogenase
gifA	glutamine synthetase inactivation factor (IF7A) gene
gln, L-gln	L-glutamine
glnA	glutamine synthetase gene
*glnA**	truncated version of the *glnA* gene
glnA[p.D52S]	active-site mutant *glnA* (Asp → Ser at the 52nd amino acid position)
glnN	*Synechocystis* sp. PCC 6803 glutamine synthetase gene
glu, L-glu	L-glutamate
gluc, glc	glucose
gluc6P	glucose 6-phosphate
glyc	glycerol
GOGAT	glutamine oxoglutarate aminotransferase
GPR	gene-protein-reaction association
GS	glutamine synthetase
HC	heterocyst
HCSC	heterocyst super-compartment
HDTMA	hexadecyltrimethylammonium
hepA	late-stage heterocyst development regulator gene
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer
HetF	heterocyst differentiation factor
HetL	early-stage heterocyst differentiation protein
HetN	heterocyst pattern regulator protein
HetR	heterocyst differentiation regulator
HF	high fidelity
his, L-his	L-histidine
HPLC	high-pressure liquid chromatography
hupL	uptake hydrogenase large subunit gene
ID	identifier
ID, %	identity
IDH	isocitrate dehydrogenase
IF7A	glutamine synthetase inactivation factor protein (*gifA*)
iuc	aerobactin operon in *Escherichia coli*
iucA	aerobactin synthase gene (*Escherichia coli*)
iucC	aerobactin synthase gene (*Escherichia coli*)
kb	kilobase(s)
KEGG	Kyoto Encyclopedia of Genes and Genomes
Km	kanamycin
KO	knockout
KOH	potassium hydroxide
LB	Luria-Bertani medium
LC	liquid chromatography
LED	light emitting diode
MFS	major facilitator superfamily
MJ	megajoule
mob	mobilization genes on RSF1010-based vectors
MOPS	3-(N-morpholino)propanesulfonic acid buffer
MRM	multiple reaction monitoring
MS	mass spectrometry
MSX	L-methionine sulfoximine
MWCO	molecular weight cutoff
NAD^+	nicotinamide adenine dinucleotide (deprotonated form)
NADH	nicotinamide adenine dinucleotide (protonated form)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (deprotonated form)
NADPH	nicotinamide adenine dinucleotide phosphate (protonated form)
NCBI	National Center for Biotechnology Information
NCO^-	cyanate ion
NEB	New England BioLabs Ltd., UK
NH_3	ammonia
NH_3–N	ammonia nitrogen (including both forms)
NH_4^+	ammonium ion
nifD	nitrogenase α-subunit gene
Nm	neomycin
NO	nitric oxide

NO_2^-	nitrite ion
NO_3^-	nitrate ion
NO_x	nitric oxides
nptII	neomycin phosphotransferase II gen
N_r	reactive nitrogen
nsiR4	nitrogen-stress induced sRNA 4
NtcA	global nitrogen regulator transcription factor
nucA	sugar-non-specific nuclease gene
nuiA	sugar-non-specific nuclease inhibitor gene
OD	optical density
OD_{730}	optical density measured at 730 nm
OGDH	2-oxoglutarate dehydrogenase complex
ORF	open reading frame
ori	origin of replication
oriT	origin of transfer
orn	L-ornithine
patB	late-stage heterocyst development regulator gene
PatS	heterocyst inhibition-signalling peptide
PCC	Pasteur Culture collection of Cyanobacteria
PCI	phenol-chloroform-isoamyl alcohol
PCR	polymerase chain reaction
PEG	polyethylene glycol
P_{gifA}	promoter of the *gifA* gene
P_{glnA}	promoter of the *glnA* gene
P_i	orthophosphate
P_{nifHDK}	promoter of the *nifHDK* operon
P_{petE}	promoter of the *petE* gene
PQ	plastoquinone
P_{rbcLS}	promoter of the *rbcLS* operon
P_{sch}	schizokinen operon and *schT* bidirectional promoter
PSI	photosystem I
PSII	photosystem II
pvsC	vibrioferrin inner membrane exporter (*Vibrio parahaemolyticus*)
pyr	pyruvate
QC	query coverage
RBS	ribosomal binding site
rep	replicase gene on RSF1010-based vectors
rhb	rhizobactin 1021 operon in *Sinorhizobium meliloti* 1021
rhbF	rhizobactin siderophore biosynthesis gene (*Sinorhizobium meliloti* 1021)
rhrA	rhizobactin 1021 transcriptional activator (*Sinorhizobium meliloti* 1021)
rhtA	outer membrane receptor gene for rhizobactin 1021 (*Sinorhizobium meliloti* 1021)
rhtX	rhizobactin 1021 uptake permease (*Sinorhizobium meliloti* 1021)
RNA	ribonucleic acid
pRSF1010	broad-host-range vector
RuBisCO	ribulose 1,5-bisphosphate carboxylase
rxn	reaction
sacB	*Bacillus subtilis* levansucrase gene
sch	schizokinen operon in *Anabaena* sp. PCC 7120
schT	schizokinen outer membrane transporter (*alr0397*)
SD	standard deviation
SDM	site-directed mutagenesis
Sm	streptomycin
sn-glyc3P	sn-glycerol 3-phosphate
SOE	splicing by overlap extension (overlap extension PCR)
Sp	spectinomycin
sp.	species
SR	single recombinant
suc	sucrose
TAE	tris(hydroxymethyl)aminomethane-acetic acid-EDTA buffer
TAP	tris-acetate-phosphate medium
TCA	tricarboxylic acid cycle
TES	N-[Tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid
Tg	teragrams (10^{12}, one trillion)
T_{gifA}	*gifA* transcriptional terminator
T_{glnA}	*glnA* transcriptional terminator
T_{oop}	*oop* transcriptional terminator (heterologous)

TQMS	triple quadrupole mass spectrometer
TS	total score
TSS	transcription start site
UK	United Kingdom
USA	United States of America
UV	ultraviolet
VC	vegetative cell
VCSC	vegetative cell super-compartment
WT	wild type
*Xba*I	*Xba*I restriction endonuclease
Δ*amt*	*amt* cluster knockout
Δ*sch*	schizokinen operon knockout
Δ*sch*::*sch*	complementation strain
Δ*sch*::*sch*Δ*all0390*	single-gene knockout complementation strains (Δ*all0390*–*all0396*)

1 Introduction

1.1 Synopsis[1]

Cyanobacteria are ubiquitous photosynthetic organisms found in almost every habitat on Earth, including hot springs and Antarctic rocks , as well as the fur of some sloths (Aiello, 1985). Cyanobacteria are highly diverse in terms of morphology: some species are filamentous, others are unicellular or can form aggregates, several species are capable of nitrogen fixation in differentiated heterocysts, and some form motile hormogonia or spore-like akinetes (Flores and Herrero, 2010; Singh and Montgomery, 2011). In their natural environment, cyanobacteria are often an integral part of complex ecosystems with other species from all three domains of life (Stewart et al., 1983; Adams, 2000; Adams and Duggan, 2008). Several species build up thick microbial mats in extreme environments (Reysenbach et al., 1994), or composite with fungal filaments to form lichens (Rikkinen et al., 2002), while others live inside their symbiotic plant hosts (Adams, 2000). In case of the aquatic *Azolla caroliniana*, a small water fern, a filamentous, heterocyst-forming cyanobacterium is found within the ovoid cavities in the plant's leaves, maintaining a mutually beneficial symbiotic relationship with the plant. This symbiont, *Anabaena azollae* provides fixed nitrogen to the fern and, in return, receives carbon sources and a protected environment from *Azolla* (Hill, 1977; Lechno-Yossef and Nierzwicki-Bauer, 2002). The highly-productive *Azolla-Anabaena* symbiosis has long been recognized as a cheap and effective biofertilizer of tropical rice paddies, and more recently it has been successfully applied in temperate climate as well (Wagner, 1997; Bocchi and Malgioglio, 2010). Outside of its plant host, the free-living form of *Anabaena azollae* has significant contribution to the carbon and nitrogen

[1] Section 1.1 is a reproduction of the introduction in Malatinszky et al. (2017) including minor modifications, with permission from the American Society of Plant Biologists under the license number 4165351266912 (27/07/2017, Appendix J).

economy of tropical soils as well, forming microbial communities with other nitrogen-fixing cyanobacteria (Singh, 1950). When living freely, however, *Anabaena azollae* only develops 5 to 10% of its cell to heterocysts. This frequency increases up to 25 to 30%, when the symbiosis is extended to also include rice. This higher rate of nitrogen-fixation is the result of an adjustment to provide sufficient nitrogen for all three species, i.e. the cyanobacterium, the fern and the co-cultivated rice (de Macale and Vlek, 2004). *Anabaena* sp. PCC 7120, an isolated strain with very high sequence similarity to *Anabaena azollae*, shows the same developmental pattern of a single heterocyst for every 10 to 20 vegetative cells (Kumar et al., 2010; Ehira, 2013), and acts as a representative model organism of the free-living cyanobacterium. To mimic the productivity of the symbiotic form, *Anabaena* sp. PCC 7120 has recently been modified to increase the expression of the HetR protein controlling heterocyst frequency and thus to enhance the organism's potential as a nitrogen biofertilizer. The resulting mutant strain has been reported for its ability to provide rice seedlings with beneficial levels of nitrogen in short-term hydroponic experiments (Chaurasia and Apte, 2011). In order to utilise such biochemical traits in designed applied processes, it becomes important to understand community behaviour and metabolic interactions in natural and simple ecosystems where these feature.

In fact, *Anabaena* sp. PCC 7120 can alone be argued to form such a very simple yet incompletely understood "community" of cells with multiple metabolic states and interdependent metabolic exchange. Under diazotrophic conditions, approximately every tenth vegetative cell irreversibly transforms into a heterocyst to provide a low-oxygen environment for the nitrogenase enzyme to function (Golden and Yoon, 2003). This enzyme is responsible for the conversion of atmospheric molecular nitrogen into ammonia in a highly energy-expensive reaction, consuming chemical energy stored in 16 molecules of ATP and 8 electrons carried by ferredoxin molecules for every molecule of nitrogen assimilated. Furthermore, the nitrogenase is irreversibly inactivated by oxygen which makes oxygenic photosynthesis and nitrogen fixation incompatible processes (Fay, 1992). Therefore, these specialized heterocyst cells undergo a series of changes to minimize the level of internal oxygen, including the deposition of two additional envelope layers around the cell and the degradation of photosystem II and carboxysomes (Wolk et al., 2004; Nicolaisen et al., 2009; Awai et al., 2010). As a result, heterocysts are dependent on vegetative cells as a source of electrons and carbon (Kumar et al., 2010). In return, vegetative cells obtain fixed nitrogen (Meeks and Elhai, 2002). Heterocysts and vegetative cells are therefore mutually interdependent, showing the features of a very simple "ecosystem". This "ecosystem" is a suitable and simple model to simulate and elucidate community metabolism. In addition, *Anabaena* spp. has a long recognized evolutionary record as a plant symbiont (Hill, 1977; Lechno-Yossef and Nierzwicki-Bauer, 2002) that has been exploited successfully in agriculture (Wagner, 1997; Bocchi and Malgioglio, 2010). It may be possible to rationally extend this

community behaviour by partially redirecting the flow of fixed nitrogen from heterocysts to the surrounding medium. Such strains will represent an important first step towards the systematic development of engineered crop (endo)symbionts that, in the end, may revolutionize nitrogen fertilization in modern agriculture (Colnaghi et al., 1997; Mus et al., 2016; Ambrosio et al., 2017).

1.2 Essential nitrogen

1.2.1 Nitrogen for life

Nitrogen is the fifth most abundant element in our solar system (Anders and Grevesse, 1989; Croswell, 1995) and by mass, the third most abundant constituent of anything living (Figure 1.1). Not surprisingly, nitrogen is one of the primary nutrients critical for the survival of all living organisms, by being essential for the synthesis of the two most important polymers of life, nucleic acids and proteins, as well as chlorophyll and many other biomolecules. Indeed, nitrogen requirements of the biosphere are enormous: depending on the life form, between 2 and 20 atoms of nitrogen is incorporated into cells for every 100 carbon atoms (Sterner and Elser, 2002).

Despite its overwhelming abundance on the planet, nitrogen is a scarce resource and often limiting primary productivity in both marine and terrestrial ecosystems (Falkowski et al., 2008; Bernhard, 2010). The reason being, over 98% of the nitrogen in global biogeochemical reservoirs is stored as molecular nitrogen (N_2) either in the atmosphere or in solution in the ocean (Kinzig and Socolow, 1994; Mackenzie, 1998). The strong triple bond of the N≡N molecule makes dinitrogen virtually inert and therefore, largely inaccessible in this form to most organisms.

Together with N_2, nitrogen exists in many different forms including both inorganic and organic molecules, which can be divided into two major classes. Unreactive nitrogen (N_2) makes up 78% of Earth's atmosphere. Reactive nitrogen (N_r) includes every other form of the element, such as nitrogen oxides (NO_x), nitrous oxide (N_2O), ammonia (NH_3), nitrite (NO_2^-) and nitrate (NO_3^-). All biological systems rely on these reactive nitrogen species, but historically these forms have been in short supply. Until the end of the nineteenth century, the main agricultural source was fixation of N_2 by symbiotic bacteria in legumes planted for that purpose, combined with careful recycling of the limited amount of nitrogen in manure (Sutton et al., 2011b).

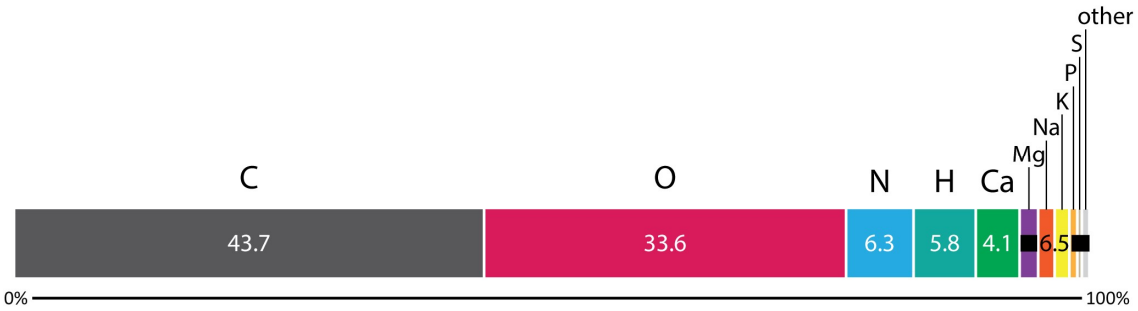


Figure 1.1. Elemental composition of microorganisms by Bowen (1966). Preceded by carbon (C) and oxygen (O), nitrogen (N) is the third most abundant biogenic element in microorganisms by mass. Hydrogen (H) closely follows the contribution by nitrogen, while calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), phosphorus (P), sulphur (S) and the other microelements contribute altogether about 10%.

Nitrogen undergoes many different transformations in the ecosystem, changing from one form to another as organisms use it for growth and, in some cases, energy (Bernhard, 2010). The major transformations of nitrogen are nitrogen fixation, nitrification, denitrification, anaerobic ammonium oxidation and ammonification (Figure 1.2). The transformation of nitrogen into its many oxidation states is key to productivity in the biosphere and is highly dependent on the activities of a diverse group of microorganisms, such as bacteria, archaea and fungi, composing a complex cycle of all forms of nitrogen (Bernhard, 2010).

The earliest nitrogen cycle on Earth was controlled by atmospheric reactions and slow geological processes. The diverse microbial processes evolved about 2.7 billion years ago to form the modern nitrogen cycle with robust natural feedbacks and controls (Canfield et al., 2010).

1.2.2 The nitrogen cycle

The nitrogen cycle is one of the most important nutrient cycles in terrestrial ecosystems. Nitrogen cycling involves four microbiological processes: nitrogen fixation, mineralization (ammonification), nitrification and denitrification. Only a few species of aquatic and terrestrial bacteria and cyanobacteria can convert N_2 into ammonium (NH_4^+) through the process of biological nitrogen fixation (BNF). This biological pathway provides the dominant natural flow of fixed nitrogen on the planet (Kinzig and Socolow, 1994). Lightning, however, provides a non-biological pathway for nitrogen fixation, oxidizing N_2 to nitric oxide (NO), which is rained out as nitrate (NO_3^-) within days (Lawrence et al., 1995). The product, with a fixation rate of 3 Tg of nitrogen per year (Tg N yr^{-1}), is only a few

percent of the rate of BNF, though nitrogen was fixed exclusively by lightning and in the shock waves of meteors before the origin of life on this planet (Schlesinger and Bernhardt, 2013).

Biological nitrogen fixation is currently estimated to provide a global annual input of 258 Tg N yr^{-1} to the biosphere (Fowler et al., 2013) making it the largest single global input of N_r, although there are significant uncertainties about the magnitude and spatial distribution of fluxes. In terrestrial ecosystems BNF is on the order of 90–130 Tg N yr^{-1} (Galloway et al., 1995) and stratospheric lightning produces N_r in the order of 3–5 Tg N yr^{-1} (Lawrence et al., 1995; Galloway, 1998). Marine ecosystems fix N_2 in the order of 40–200 Tg N yr^{-1}, and possibly more due to a likely enhancement of BNF from atmospheric iron deposition (Michaels et al., 1996). The reverse process, i.e. the return of reactive nitrogen from terrestrial and marine ecosystems to the atmosphere is governed by three major fluxes: First, a small amount (about 10%) of the N_r fixed by natural terrestrial ecosystems is emitted to the atmosphere as NO_x primarily from soils and, to a lesser extent from biomass burning, at a rate of about 5 Tg N yr^{-1}. Second, about 8 Tg N yr^{-1} is emitted to the atmosphere as NH_3 from soils, animal waste and biomass burning; and third, about 13 Tg N yr^{-1} of NH_3 is emitted to the atmosphere by marine ecosystems (Galloway et al., 1995). There is also a considerable hydrologic distribution of N_r, with about 35 Tg N yr^{-1} transported via rivers to coastal systems, which does not explicitly involve emission to the atmosphere (Galloway et al., 1995).

Ammonia produced by BNF is oxidized by ammonia-assimilating bacteria in a process called nitrification. Nitrification is carried out primarily by two chemolithoautotrophic bacterial groups, the ammonia-oxidizing bacteria (*Nitrosomonas*, *Nitrosospira* and *Nitrosococcus* (Head et al., 1993)) and the nitrite-oxidizing bacteria (*Nitrobacter*, *Nitrospina*, *Nitrococcus* and *Nitrospira* (Teske et al., 1994)), converting ammonia in the soil to nitrate via nitrite (Figure 1.2). At a much smaller rate than in case of chemolithoautotrophs, heterotrophic nitrification is carried out by a wide phylogenetic range of bacteria and fungi that can oxidize ammonia or reduced nitrogen from organic compounds to hydroxylamine, nitrite and nitrate (Verstraete and Focht, 1977). The excess of nitrate produced by auto- and heterotrophic nitrification causes the contamination of ground water, and the gaseous by-products of nitrification (NO and N_2O) are two of the most potent greenhouse gases (Prosser, 1989; Parton et al., 1996).

The nitrogen cycle is completed by gradual reduction of nitrate and nitrite. Denitrification is defined as the dissimilatory reduction of NO_3^- or NO_2^- to N_2O and N_2, respectively, by microbial reactions (Figure 1.2). Besides several anaerobic bacterial phyla including *Pseudomonas* and *Paracoccus* (Zumft, 1997; Hayatsu et al., 2008), various fungi including ascomycota (such as *Cylindrocarpon tonkinense* and *Gibberella fujikuroii*) and basidiomycota (e.g. *Trichosporon cutaneum*) have been found to exhibit

denitrifying activity (Shoun et al., 1992). Furthermore, several archaea, such as the hyperthermophile *Pyrobaculum aerophilum* and the halophile *Haloferax denitrificans*, are capable of denitrification (Philippot, 2002; Cabello et al., 2004). In the eukaryotic domain, many denitrifying fungi combine nitrogen atoms from nitrite and other nitrogen compounds (cosubstrates) under denitrifying conditions, producing hybrid N_2 or N_2O. The cosubstrates, such as ammonium and azide, are denitrified in a process called co-denitrification, induced by nitrite or nitrate. In the absence of the inducers the cosubstrates are incapable of activating co-denitrification (Shoun et al., 1992; Tanimoto et al., 1992).

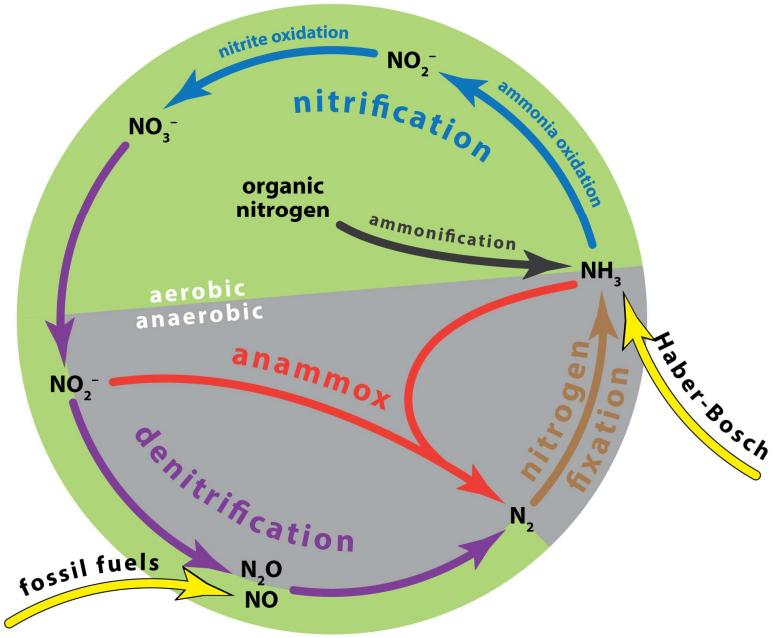


Figure 1.2. The nitrogen cycle. Atmospheric dinitrogen (N_2) is fixed to ammonia (NH_3) in an anaerobic process. Ammonia is oxidized to nitrite (NO_2^-) and further oxidized to nitrate (NO_3^-) during aerobic nitrification. Denitrification follows both aerobically and anaerobically (by different microorganisms), and nitrate is converted first to nitrite and finally back to dinitrogen through nitrogen oxide (NO) and nitrous oxide (N_2O), closing the nitrogen cycle. Ammonification (mineralization) and anaerobic ammonium oxidation (anammox) interconnect distinct parts of the cycle. Anthropogenic nitrogen input to the nitrogen cycle is depicted by yellow arrows (Haber–Bosch process and fossil fuel burning).

The fourth major biological process interconverting the reactive forms of nitrogen is anaerobic ammonium oxidation (anammox). Based on theoretical thermodynamic calculations, the possibility of anaerobic ammonium oxidation in biological systems has been postulated (Broda, 1977), but the reaction was only discovered in 1995 (Mulder et al., 1995). Anammox bacteria belonging to the

planctomycete phylum convert ammonium and nitrite into dinitrogen gas under anaerobic conditions (Mulder et al., 1995; Strous et al., 1999; Schmid et al., 2005).

Some of the fungi (e.g. *Fusarium solani*) are able to produce N_2 through fungal co-denitrification from nitrite and ammonia (Shoun et al., 1992). In addition, another type of dissimilatory nitrate metabolism, ammonia fermentation, has also been found in the denitrifying *Fusarium oxysporum* (Zhou et al., 2002; Takasaki et al., 2004). In ammonia fermentation, nitrate is reduced to ammonia and ethanol is simultaneously oxidized to acetate to generate energy in the form of ATP.

In addition to the above natural processes, humans began to have an enormous impact on the global nitrogen cycle in the twentieth century, by developing industrial processes to reduce N_2 to NH_4^+, by implementing new agricultural practices that boost crop yields, and by burning fossil fuels (Vitousek et al., 1997). The increased need for fertilizer nitrogen to feed a fast-growing population in the early 1900s demanded an efficient and large-scale process to industrially fix atmospheric dinitrogen to ammonia. The new process called the Haber–Bosch synthesis revolutionized fertilizer production and expressed an immense impact on both humanity and our planet. Among all inventions of the last hundred years the most important is unquestionably the extremely efficient Haber–Bosch synthesis providing the world with a reliable source of fertilizer nitrogen (Erisman et al., 2008).

1.2.3 Haber–Bosch ammonia

In 1909, Fritz Haber, a German chemist discovered how ammonia, the chemically reactive, highly usable form of nitrogen, could be synthesized by reacting atmospheric dinitrogen with hydrogen gas in the presence of an iron catalyst at high pressures and temperatures (Haber and van Oordt, 1905). For his invention Haber was awarded the Nobel Prize in Chemistry in 1918. The process was later developed at industrial scale by Carl Bosch in 1913, for which he was awarded with a Nobel Prize in 1931 (Smil, 2004). Today, this process is known as the Haber–Bosch synthesis, providing fertilizer nitrogen at about 100 Tg per year worldwide (Erisman et al., 2008). Although the process has undergone remarkable engineering improvements over the decades, the original solid iron catalyst is still in use (Vojvodic et al., 2014), and elemental synthesis of ammonia is performed at similar temperatures (650–750 K) and pressures (50–200 bar) as in the original invention (Haber and van Oordt, 1905; Tamaru, 1991; Nielsen, 1995; Schlögl, 2008).

The importance of the Haber–Bosch synthesis for today's human population cannot be overestimated. Nearly 80% of the nitrogen found in human tissues can be originated from the Haber-Bosch process

(Howarth, 2008). In addition, it has been estimated that at the end of the twentieth century, about 40% of the world's population depended on fertilizer inputs to produce food (Smil, 2002, 2004). Moreover, another analysis, based on long-term experiments and national statistics, concluded that about 30–50% of the crop yield increase was due to nitrogen application through mineral fertilizer (Stewart et al., 2005). Furthermore, another estimation suggested that nitrogen fertilizer has supported approximately 27% of the world's population over the past century, equivalent to around 4 billion people born (Erisman et al., 2008). In other words, more than half of today's human population would not even exist without the inventions by Haber and Bosch. During 2015 alone, the Haber–Bosch process supplied 140 Tg N (Ober, 2017).

If we assume that the global nitrogen cycle was in an approximate equilibrium prior to industrialization, BNF would have been balanced by the reductive processes of denitrification returning N_2 to the atmosphere, with estimates of around 260 Tg N yr^{-1} arising from terrestrial and oceanic sources (Galloway et al., 2004). Over the past century, however, the development of new agricultural practices to satisfy a growing global demand for food has drastically disrupted the balance of the nitrogen cycle (Canfield et al., 2010). Human activity has approximately doubled the global rate of nitrogen fixation compared to pre-industrial times (Canfield et al., 2010; Schlesinger and Bernhardt, 2013). About 80% of the total nitrogen manufactured by the Haber–Bosch process is used in the production of agricultural fertilizers (Galloway et al., 2008). However, a large proportion of this nitrogen is lost to the environment: in 2005, approximately 100 Tg N from the Haber–Bosch process was used in global agriculture, whereas only 17 Tg N was consumed by humans in crop, dairy and meat products (Braun, 2007). This highlights extremely low nitrogen-use efficiency in agriculture. A recent study suggested that approximately 40% of fertilizer nitrogen lost to the environment is denitrified back to unreactive atmospheric dinitrogen (Galloway et al., 2004). In principle this loss is environmentally benign, although it represents a waste of the energy used in the Haber–Bosch process, equivalent to at least 32 MJ kg^{-1} N fixed, or about 1% of the global primary energy supply (Erisman et al., 2008). The input to the nitrogen cycle (in addition to the contribution by fertilizer ammonia as shown by yellow arrows in Figure 1.2) from agriculture alone reaches about 33.6 Tg N yr^{-1} because of cultivation-induced nitrogen fixation, primarily due to the use of fodder legumes as a side crop (Galloway et al., 2004). In 2008, fossil fuel combustion generated another 25.2 Tg (Gruber and Galloway, 2008). Together, anthropogenic sources contribute to the nitrogen cycle around 45% of the total fixed nitrogen produced annually on Earth (Canfield et al., 2010), posing an ever-increasing environmental and health problem to humanity.

1.3 The Nitrogen Crisis

1.3.1 Environmental and health effects

Humans may have produced the largest impact on the nitrogen cycle since the major pathways of the modern cycle originated some 2.5 billion years ago. Natural feedbacks driven by microorganisms will likely produce a new steady state over time scales of many decades. At the end of this process excess nitrogen added from human sources will no longer accumulate, but will be removed at rates equivalent to rates of addition (Canfield et al., 2010). However, because of the projected increase in human population through at least 2050 (United Nations, 2017), there will be demand for a concomitant rise in fixed nitrogen for crops to feed this population.

In fact, the emission of N_r continues to increase every year. From 1860 to 1995, energy and food production increased steadily on both an absolute and per capita basis; similarly, N_r creation also increased from about 15 Tg N in 1860 to 156 Tg N in 1995. The change was enormous, over an order of magnitude, and it increased further in the following decade reaching up to 187 Tg N yr^{-1} in 2005 (Galloway et al., 2008). In another study it was estimated that, between 1960 and 2000, the use of nitrogen fertilizers increased by about 800% (Fixen and West, 2002), with wheat, rice, and maize accounting for about 50% of fertilizer use (Canfield et al., 2010). For these crops, however, the nitrogen use efficiency is typically below 40%, meaning that most applied fertilizer either washes out of the root zone or is lost to the atmosphere by denitrification before it is assimilated into crop biomass (Canfield et al., 2010). Denitrification forms mainly N_2 under anoxic conditions, but also forms N_2O (Figure 1.2, purple arrows), a fraction of which is lost to the atmosphere and increasingly contributes to the rise in atmospheric N_2O concentrations (Forster et al., 2007). As a greenhouse gas, N_2O has 300 times (per molecule) the warming potential of CO_2, and it also reacts with and destroys ozone in the stratosphere (Ravishankara et al., 2009).

Nitrifying bacteria can also convert the unused nitrogen fertilizer NH_4^+, forming the highly mobile NO_3^-. The excess nitrate can leach into rivers, lakes and aquifers (Canfield et al., 2010), ultimately leading to eutrophication of coastal waters and detrimental impacts on water quality, as well as creating huge hypoxic zones around the world (Seitzinger et al., 2002; Diaz and Rosenberg, 2008). Nitrification also produces N_2O as an intermediate; therefore, agricultural systems represent huge sources of N_2O to the atmosphere, accounting for about one-quarter of global N_2O emissions (Mosier et al., 1998; Canfield et al., 2010).

Once N_r enters the environment, its effects on terrestrial, aquatic, and atmospheric realms can influence human health and welfare in several ways (Galloway et al., 2008). The excess N_r in the environment poses direct health threats to humans including an increased risk of cancer and reproductive disorder due to elevated nitrate in drinking water (Ward et al., 2005), exacerbated pulmonary disease due to tropospheric ozone and fine-particle formation (Townsend et al., 2003), and increased prevalence of important infectious diseases, such as malaria, West Nile virus, cholera and schistosomiasis (McKenzie and Townsend, 2007). Although the threats and challenges posed by the excessive and inefficient use of nitrogen fertilizers is better understood by every year, the actual scale of the associated costs and intervention strategies have started being fully realised only recently.

1.3.2 Costs and measures

An analysis published in 2011 calculated that excess nitrogen in the environment costs the European Union (EU) between €70 billion and €320 billion per year (Sutton et al., 2011a; Sutton et al., 2011b). It is the first time that an economic value has been placed on the threats posed by nitrogen pollution, including contributions to climate change and biodiversity loss. On one hand, manufactured fertilizer produces a direct benefit to European farmers in terms of crops grown, topping at €25 billion to €130 billion per year, when the long-term benefits are included (Sutton et al., 2011a). On the other hand, however, this total benefit is only less than half the value that nitrogen fertilizers are estimated to cost the EU every year as remediation costs of a severe pollution, due to the loss of about half of the nitrogen in fertilizers and manures to the surrounding environment (Galloway et al., 2004). In economic terms, this converts to a loss of potential benefits to farmers of €13 billion to €65 billion per year. On these grounds alone, there is a strong case for using nitrogen more efficiently (Sutton et al., 2011b), not to mention the associated health and environmental effects discussed earlier.

Of the total cost of damage from N_r emissions, 75% comes from the effects of NO_x and NH_3 on human health and ecosystems. Although N_2O has recently been heralded as the main cause of stratospheric ozone depletion (Ravishankara et al., 2009), this represents only about 1% of the damage costs (Sutton et al., 2011b). Climate change and ozone thinning are important, but the threats to health and ecosystems are an even stronger argument for taking action on nitrogen. Clearly nitrogen is one of the major environmental challenges of the twenty-first century, especially to manage nitrogen better in agriculture and to moderate the developed world's consumption of animal protein (Sutton et al., 2011b). An improved nitrogen management would include controlling NO_x emissions from fossil-fuel combustion, increasing nitrogen-uptake efficiency of crops and improving animal-management

processes (Galloway et al., 2008). According to estimations by Galloway et al. (2008) such measures may contribute to the potential decrease of N_r at a 53 Tg N yr^{-1} rate, or about 28% of the annual N_r production in 2005 (Galloway et al., 2008). Canfield et al. (2010) on the other hand, described more focussed intervention strategies to tackle the increasing N_r emission: (1) introducing of systematic crop rotation (Peoples et al., 2009), (2) optimizing the timing and amounts of fertilizer applied to increase the efficiency of their use by crops (Raun et al., 2002), (3) developing genetically engineered varieties for improved nitrogen use efficiency (Tester and Langridge, 2010), (4) improving the ability of economically important varieties of wheat, barley, and rye to produce nitrification inhibitors (Subbarao et al., 2009; Tester and Langridge, 2010), and (5) further developing cereals and other crops with (endo)symbiotic nitrogen-fixing microorganisms to supply their nitrogen needs (Colnaghi et al., 1997; Iniguez et al., 2004; Bocchi and Malgioglio, 2010; Ran et al., 2010; Mus et al., 2016; Ambrosio et al., 2017). The latter approach (i.e. engineering of symbiotic nitrogen-fixing microorganisms) is also the topic of this thesis. Such projects altering the core metabolism through genetic and metabolic engineering, two essential aspects in synthetic biology, cannot succeed without systemic understanding of the organism's biochemistry.

1.4 Metabolic modelling of entire organisms

1.4.1 Systems biology

Systems biology is characterized by synergistic integration of theory, computational modelling, and experimental observation (Kitano, 2002). It investigates the behaviour and relationships of all the elements in a particular biological system while it is functioning. These data can then be integrated, graphically displayed, and ultimately modelled computationally. According to this definition, biological systems can be seen as being composed of two types of information: genes, encoding the molecular machinery that executes all cellular functions, and networks of regulatory interactions, specifying how genes are expressed (Ideker et al., 2001; Kitano, 2002). The purpose of systems biology is to comprehensively gather information from all levels (gene, reaction, regulation and phenotype) for individual biological systems, and to integrate these data to generate predictive mathematical models of the system (O'Malley and Soyer, 2012). These models are highly diverse in terms of scope (metabolic or regulatory), system size (quantitative or genome-scale) and the implementation of the studied system by the model. The choice of analytical method used depends on the availability of biological knowledge to incorporate into the model. Smaller models integrate more detail and provide a comprehensive kinetic description of the corresponding biological system. A wide range of

computational tools have been developed for such representations in the past decade (Le Novere and Shimizu, 2001; Hoops et al., 2006). In contrast, a steady-state analysis can be done using only the network structure, without knowing the rate constants for a particular reaction. These representations use the constraint-based approach and describe the biological system as a static entity (Price et al., 2003; Orth et al., 2010). In between the two methods are dynamic constraint-based tools for the simulation of dynamic biological systems, by assuming organisms reach steady-state rapidly in response to changes in the extracellular environment (Antoniewicz, 2013; Gomez et al., 2014). Typically, genome-scale models are represented either by static or dynamic constraint-based approaches. These large-scale models are powerful tools allowing for the prediction of cellular growth, flux profiles, and mutant strain phenotypes (Orth et al., 2010). In recent years, with the development of new computational algorithms, genome-scale models have been used to guide the design of strains for biochemical production, such as biofuels and commodity chemicals (Curran and Alper, 2012).

1.4.2 Flux Balance Analysis

Mathematical modelling of metabolism and simulation of cellular processes has long been recognized as an excellent tool to understand the organization and behaviour of biological systems (Fell and Small, 1986; Savinell and Palsson, 1992). The predictive ability of these models has evolved remarkably in the past few years. The quality of a stoichiometric model and its predictions is directly dependent on the experimental data incorporated into them. However, there is a reverse relationship between the level of detail to which a model can predict the behaviour of the described system and the actual size of the system (Steuer et al., 2012). In this sense, predictive kinetic models require the highest level of detail (quantitative models); whereas topological models usually contain very large networks (genome-scale or qualitative networks).

The data incorporated into stoichiometric network reconstructions include detailed genetic and biochemical information as well as kinetic data of intracellular processes (Price et al., 2003). Although more and more complete genomes are being sequenced every year, biochemical and kinetic information is sparse or missing entirely for many organisms. Kinetic models provide precise and essential information on the studied system (Loew and Schaff, 2001; Hoops et al., 2006); however, it is difficult to reconstruct kinetic models for genome-scale networks because of the large number of parameters needed and their computational complexity (Steuer et al., 2012). In contrast, even genome-scale metabolic networks (easily containing several hundred or thousand individual reactions) can be analysed with the constraint-based modelling approach (Orth et al., 2010). The exact

reconstruction steps of a genome-scale metabolic network are detailed in exhaustive protocols (Feist et al., 2009; Thiele and Palsson, 2010).

In short, the underlying metabolic network of biochemical reactions is reconstructed first from biochemical textbooks, primary literature and public databases, based primarily on an annotated genomic sequence (Palsson, 2006). The network reconstruction contains all known metabolic functions (reactions) of an organism, including the genes and proteins associated with each biochemical reaction.

These genome-scale reconstructions of entire organisms are, however, largely underdetermined, due to inevitable simplifications of the system (containing only genetic information). To counter the lack of kinetic data, constraints are placed on every single reaction in the network to comply with directionality and thermodynamics. The collection of these reaction constraints essentially determine the rules under which the reconstructed network can operate (Price et al., 2003). Applying these constraints the flow (flux) of metabolites can be calculated through the network, thereby making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite (Orth et al., 2010). The approach is called flux balance analysis (FBA) that is a widely used mathematical tool for studying biochemical networks, at genome-scale in particular.

In FBA, metabolic reactions are represented by a stoichiometric matrix (Palsson, 2006; Becker et al., 2007; Orth et al., 2010). The entries in the matrix are stoichiometric coefficients describing the underlying reaction network. Each column represents a reaction, and each row represents the participation (stoichiometric coefficient) of a given metabolite in every reaction throughout the network (Palsson, 2006; Orth et al., 2010). Mathematically, the stoichiometric matrix is a linear transformation of the flux vector to the time derivative of the concentration vector. The flux vector represents the fluxes through all reactions in the network, whereas the concentration vector describes the concentrations of all metabolites (Palsson, 2006). In steady state, where the time derivative of the concentration vector is zero (no net accumulation or consumption of any compound), the metabolic fluxes through the entire network can be optimized (see Equation 1).

Equation 1. Mathematical description of a metabolic network in steady state. Concentrations of all metabolites are represented by vector ***x***. Matrix ***S*** represents the stoichiometric matrix and vector ***v*** denotes for fluxes through all reactions in the network.

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{S}\boldsymbol{v} = \mathbf{0}$$

S is a sparse matrix (of *m* columns and *n* rows) because most biochemical reactions involve only a few different metabolites. Flux distributions (***v*** of length *n*) satisfying Equation 1 define a solution space of multiple solutions, spanned by the constraints of the system. Within the range of solutions it is possible to identify and analyse single points that meet certain mathematical criteria. Such points may be for example the maximum growth rate of an organism or optimal production of a certain metabolite, called the objective function. FBA seeks to optimize (maximize or minimize) an objective function within the constrained solution space. The output of such optimzation is a particular flux distribution (***v***) at the maximum or minimum of the objective function. Several powerful tools have been developed to construct, curate and optimize genome-scale models, including FluxAnalyzer, the COBRA toolbox and FAME (Klamt et al., 2003; Becker et al., 2007; Schellenberger et al., 2011; Boele et al., 2012).

For growth rate prediction, an extra reaction is added to the stoichiometric matrix describing biomass formation. The biomass equation is an artificial lumped reaction consuming energy and growth precursors (e.g. DNA, RNA and proteins), and converting the building blocks into one gram of biomass (Feist and Palsson, 2010). In bacteria, the steady-state criterion has been successfully applied to cultures in their exponential growth phase for the description of growth (Varma and Palsson, 1994; Edwards et al., 2001). However, for eukaryotes, especially multicellular organisms, the maximal sum of metabolic exchange fluxes has been proposed as the predictor of growth instead of the biomass equation (Zarecki et al., 2014).

The optimization of the objective function in FBA is carried out by linear programming. In linear programming, an optimization problem may be defined as the problem of maximizing or minimizing a linear function subject to linear constraints (Ferguson, 2000). The constraints are the restrictions or limitations on the decision variables (or in other words, the reaction rates), and are usually given as linear inequalities (upper and lower bounds). The objective function can be any linear combination of the decision variables, in a form that defines the weights of each factor. These weights indicate how much each variable contributes to the objective function (Bertsimas and Tsitsiklis, 1997). In standard FBA, the objective function contains only one variable (the weights for all other variables are zero), usually the biomass reaction (hence biomass objective function, BOF) (Feist and Palsson, 2010; Orth et al., 2010).

The BOF is a key determinant of model outcome. It describes the rate at which all of the biomass precursors are made in the correct proportions. The formulation of a detailed BOF requires information on the molecular composition of the cell and energetic requirements necessary to generate biomass content from metabolic precursors (Feist and Palsson, 2010). The energetic

requirements may include contributions from growth-associated (GAM) and non-growth-associated (NGAM) maintenance. The GAM details the necessary energy that the cell has to make to drive its biosynthetic processes generating biomass precursors. The NGAM, on the other hand, defines the energy limit for producing biomass (and growth) by describing the energy requirements of cell homeostasis (Feist and Palsson, 2010).

From the modelling point of view, the BOF has the highest impact on model predictions. Eventually, the BOF determines elemental composition, C/N ratio, cellular yields and growth rate. Therefore, formulation of a high-quality BOF is very important, as well as one of the key "pain points" in model reconstruction (Thiele and Palsson, 2010). The steps towards the generation of a BOF involve acquisition of high-quality datasets of detailed biomass composition and growth data for the prediction of energetic parameters; sources of data that are usually not available or very difficult to acquire in case of genome-scale models. Moreover, the BOF does not take into account physiological aspects like pH, membrane permeability or enzyme inhibition. These limitations are characteristic to most genome-scale (stoichiometric) models, and are usually dealt with by predefined assumptions and simplifications.

Several manually curated, genome-scale reconstructions are available in online databases, while uncurated reconstructions can be automatically generated from genetic data with the help of sophisticated algorithms (Overbeek et al., 2005; Schellenberger et al., 2010). These genome-scale metabolic reconstructions have been used to place high-throughput OMICS data in context, make hypothesis-driven discoveries and understand multi-species relationships, as well as a guidance in metabolic engineering (Oberhardt et al., 2009). Recently, the scope of genome-scale reconstructions has been extended to photosynthetic microorganisms as well, including cyanobacteria (Nogales et al., 2012; Saha et al., 2012; Vu et al., 2012; Knoop et al., 2013; Malatinszky et al., 2017).

1.5 Cyanobacteria

Cyanobacteria are among the most abundant organisms on Earth, occupying diverse ecological niches and exhibiting enormous diversity in terms of their habitats, physiology, morphology and metabolic capabilities (Beck et al., 2012). Cyanobacteria are tremendously relevant for primary carbon fixation in many ecosystems, being the only known prokaryotes capable of oxygenic photosynthesis. Indeed, the majority of photosynthetic genes have first appeared in the cyanobacterial lineage, and thus, ancestors of modern cyanobacteria have been suggested for being the first phototrophs to appear

some 3.4 billion years ago (Mulkidjanian et al., 2006). Just like their modern descendants, ancient cyanobacteria used water as the ultimate source of electrons for the generation of reductant in photosynthesis. The concomitant release of free oxygen was one of the most significant events in Earth's history, gradually transforming the primordial reducing atmosphere to an oxidizing atmosphere about 2.3 billion years ago (Fay, 1992; Bekker et al., 2004). The formation of a new atmosphere triggered the evolution of a complex oxygenic animal life, while it may have caused a cataclysm in those organisms that were first exposed to reactive oxygen species (Sessions et al., 2009; Shields-Zhou and Och, 2011).

Due to their numerical abundance, most notably in marine environments, cyanobacteria are fundamental for oceanic primary production, but have profound impact on almost all biochemical cycles. They are major players in global oxygen supply and CO_2 sequestration, introduce atmospheric nitrogen into the sea and are therefore indispensable for proper functioning of the entire biosphere (Peschek, 1999; Flores and Herrero, 2010). Cyanobacteria are found in numerous symbiotic relationships with diatoms, sponges, ferns, fungi and even higher plants (Stewart et al., 1983; Rai et al., 2002; Adams and Duggan, 2008).

Cyanobacteria are an ancient bacterial group that can be traced back 2.5–3.4 billion years in the Earth's history (Altermann and Kazmierczak, 2003; Mulkidjanian et al., 2006). Such a long evolutionary time scale may have contributed to the high genomic diversity and vast metabolic and adaptive capability observed among different cyanobacteria. Genome sizes range from a minimum of 1.44 Mb for the marine cyanobacterium UCYN-A (Tripp et al., 2010) to a maximum of 9.05 Mb in the facultative symbiont *Nostoc punctiforme* ATCC 29133 (Meeks et al., 2001). Most cyanobacteria possess a circular chromosome and a small number of additional plasmids, which can be up to several hundred kilobases (kb) in size. Some marine picocyanobacteria, *Prochlorococcus* and *Synechococcus* are exceptions, lacking any plasmids accompanying their chromosome (Hess, 2011).

Chromosomal copy numbers were also reported to be largely different among the many clades of cyanobacteria (Griese et al., 2011). Experimentally determined ploidy levels suggested marine picocyanobacteria are mainly monoploid or diploid. However, some cyanobacteria are oligoploid (between 3 and 10 chromosome copies per cell, for example the filamentous, nitrogen-fixing *Anabaena* sp. PCC 7120), and some are highly polyploid (Griese et al., 2011). Among them, the popular model organism *Synechocystis* sp. PCC 6803 may possess a maximum of 218 chromosome copies per cell, although actual copy numbers may vary significantly based on growth phase and some environmental conditions (Griese et al., 2011; Zerulla et al., 2016).

Cyanobacteria are generally considered as being a promising resource for diverse biotechnological applications (Ducat et al., 2011; Gangl et al., 2015) for two main reasons. First of all, they are the only known prokaryotes to carry out oxygenic photosynthesis, and second, cyanobacteria are an extremely rich source of novel, bioactive secondary metabolites with extensive pharmaceutical and biotechnology potential (Hess, 2011). Indeed, until recently, over 700 different compounds have been isolated and characterized in a variety of different cyanobacteria (Tidgewell et al., 2010). Some of these metabolites develop substantial environmental toxicity and pose threats to humans, wildlife and livestock, such as the toxins produced by *Microcystis aeruginosa*, *Raphidiopsis brookii* and other species belonging to various genera (Carmichael, 1992; Portmann et al., 2008; Stucken et al., 2010; Shao et al., 2011). However, many of these secondary metabolites received the immediate attention of pharmaceutical and pesticide industries due to their potential benefit in those areas. For example the filamentous *Lyngbya majuscule*, for which putative biosynthesis pathways of the anticancer agent curacin A and the molluscicide barbamide have been identified (Gerwick et al., 1994; Orjala and Gerwick, 1996; Osborne et al., 2001). Furthermore, the filamentous non-nitrogen-fixing cyanobacterium, *Arthrospira platensis* has become known for its nutritive value (*Spirulina*), but *Arthrospira* species also serve as a source of beta-carotene and as an industrial organic material (Chung et al., 1978; Abdulqader et al., 2000).

More recently, proof of principle has been delivered for cyanobacteria as being a promising platform for the production of hydrogen (Bandyopadhyay et al., 2010), ethanol (Deng and Coleman, 1999), isobutyraldehyde, isobutanol (Atsumi et al., 2009), ethylene (Takahama et al., 2003), isoprene (Lindberg et al., 2010), and even alkanes (Schirmer et al., 2010). The availability of well-annotated genome information is absolutely essential for metabolic engineering, all synthetic and systems biology approaches and the development of suitable metabolic models (Vu et al., 2012; Knoop et al., 2013; Malatinszky et al., 2017).

1.5.1 Nitrogen assimilation

1.5.1.1 Heterocysts and nitrogen fixation

Many cyanobacterial species are capable of utilizing atmospheric nitrogen. However, oxygenic photosynthesis and nitrogen fixation are incompatible processes because nitrogenase is irreversibly inactivated by oxygen (Fay, 1992). Cyanobacteria have developed two distinct strategies to separate these activities (Kumar et al., 2010; Muro-Pastor and Hess, 2012). Some maintain a biological circadian

clock to separate them temporally by alternating between photosynthesis during the daytime and nitrogen fixation at night (Pfreundt et al., 2012). Other species evolved multicellularity for spatial separation. These microorganisms, in the presence of a combined nitrogen source such as nitrate or ammonium, grow in long filaments consisting of exclusively photosynthetic vegetative cells. In the absence of combined nitrogen, however, some cells transform into nitrogen-fixing heterocysts (Kumar et al., 2010; Muro-Pastor and Hess, 2012). These highly specialized cells are terminally differentiated (Yoon and Golden, 2001) and provide a microoxic environment for nitrogen fixation (Golden and Yoon, 2003; Ehira, 2013). Heterocysts differentiate at semiregular intervals, forming a developmental pattern of a single heterocyst for every 10 to 20 vegetative cells along filaments (Kumar et al., 2010; Ehira, 2013).

Regulation of heterocyst differentiation involves several genes and proteins. The whole process is very tightly regulated and most likely coupled to the regulation of nitrogen fixation genes in response to nitrogen deprivation. The global transcriptional regulatory protein NtcA is found in all cyanobacteria (Herrero et al., 2001; Herrero et al., 2004). It has been shown to mediate nitrogen control in response to the C/N balance of the cells with 2-oxoglutarate as the effector molecule (Zhao et al., 2010). In short, the increase of intracellular 2-oxoglutarate levels is transduced by P_{II}, a signal transduction protein. Effectively, P_{II} is transmitting a signal of nitrogen starvation via an interaction with the PipX protein, leading to the co-activation of NtcA expression (Espinosa et al., 2014).

Along with NtcA, HetR is a key regulator of heterocyst development. Proteins HetC and HetL are required during early development, although HetL it is not essential for normal heterocyst development (Liu and Golden, 2002). While HetC may be involved at early stage (Golden and Yoon, 2003), together with HetP it is also required for the progress of differentiation (Buikema and Haselkorn, 2001; Muro-Pastor and Hess, 2012). PatS and HetN, on the other hand,are both controlling the developmental pattern by suppressing the formation of new heterocysts adjacent to previously formed heterocysts. The genes *patB* and *hepA* are thought to be responsible for the regulation of late developmental stage, with *hepA* being also involved in the synthesis of the two-layer heterocyst envelope, together with *hepC* (Zhu et al., 1998; Golden and Yoon, 2003; Flores and Herrero, 2010; Muro-Pastor and Hess, 2012; Ehira, 2013). The major players in the regulation of heterocyst development are shown in Figure 1.3.

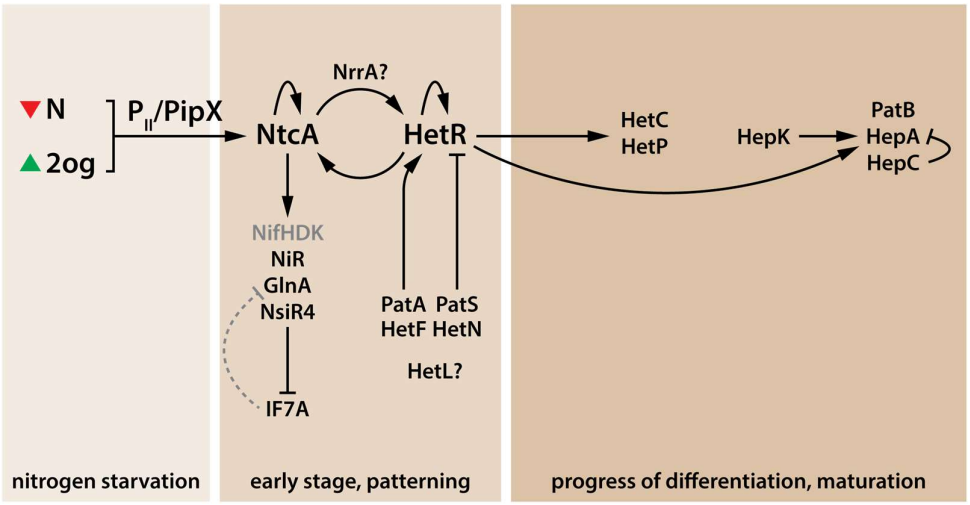


Figure 1.3. Simplified regulatory network of heterocyst development. Nitrogen starvation: NtcA is activated by the action of P_{II} and PipX in response to elevated levels of 2-oxoglutarate. **Early stage, patterning:** NtcA and HetR mutually upregulate each other's expression, while also exhibit positive feedback on their own activation. PatA and HetF has a positive effect on HetR, while also influencing development of heterocyst pattern. PatS and HetN are primarily responsible for pattern formation by inhibiting HetR and suppressing initiation of heterocyst development in neighbouring cells. HetL is an accessory protein not essential for heterocyst differentiation. **Progress of differentiation, maturation:** HetC and HetP are upregulated by HetR, and are required for the progress of differentiation, although HetC may also be involved in early development. PatB is essential at the late stage, as well as HepA and HepC. The latter two are also responsible for deposition of the heterocyst envelope during maturation. Lines ending in arrows and bars indicate positive and negative interaction, respectively. NifHDK is coloured grey as it is expressed in mature heterocysts.

During their development from vegetative cells under diazotrophic conditions (in the absence of any combined nitrogen), heterocysts undergo a series of morphological changes. Most prominently, two additional envelope layers are deposited around the pre-existing cell wall to limit the influx of air (Nicolaisen et al., 2009; Awai et al., 2010). The inner layer is composed of glycolipids and possibly reduces the cell wall permeability to oxygen. The other layer consists of polysaccharides and is thought to physically protect the glycolipid layer (Muro-Pastor and Hess, 2012). In addition to their thicker envelope, heterocysts dismantle their water-splitting photosystem II, disabling oxygen evolution (Wolk et al., 2004). Thylakoid membranes re-organize at the cell poles into the so-called honeycomb formation (Muro-Pastor and Hess, 2012). The increased respiration rate further decreases the amount of oxygen entering from the neighbouring vegetative cells. Moreover, heterocysts lack carboxysomes and the ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzyme complex, and therefore do not photosynthetically fix carbon dioxide (Flores and Herrero, 2010; Kumar et al., 2010; Muro-Pastor and Hess, 2012). Nonetheless, vegetative cells continue to perform photosynthetic CO_2 fixation during diazotrophic growth and provide heterocysts with fixed carbon; in return, heterocysts supply

fixed nitrogen (Wolk et al., 2004). Several carbohydrates including fructose, erythrose and sucrose were suggested as exchange compounds between vegetative cells and heterocyst for the transport of carbon (Privalle and Burris, 1984; Schilling and Ehrnsperger, 1985). Sucrose, in particular, has been shown to participate in intercellular transfer via the septal junctions between vegetative cells and heterocysts (Nürnberg et al., 2015). Furthermore, the different expression patterns of its metabolic genes in response to nitrogen availability have also been shown, suggesting an important role in supplying carbon to heterocysts (Cumino et al., 2007). Interestingly, however, heterocysts lack a functional GS–GOGAT cycle (see below in 1.5.1.2) and may therefore rely on vegetative cells for a source of glutamate as well (Martin-Figueroa et al., 2000). Due to the lack of the GOGAT enzyme in heterocysts, an excess of 2-oxoglutarate is thought to be transported to vegetative cells (Böhme, 1998). In *Anabaena cylindrica*, vegetative cells obtain fixed nitrogen in the form of glutamine or other amino acids (Wolk et al., 1976; Thomas et al., 1977; Picossi et al., 2005). For the intercellular exchange of these metabolites, two routes have been proposed: either via a continuous periplasm constructed by a continuous outer membrane (Flores et al., 2006; Mariscal et al., 2007) or by diffusion from cytoplasm to cytoplasm involving septal proteins (Flores et al., 2007; Merino-Puerto et al., 2010; Merino-Puerto et al., 2011). These proteins have been proposed to form septal junctions at cell–cell connections comprising the so-called microplasmodesmata (Giddings and Staehelin, 1981; Flores and Herrero, 2010). Deletion of their genes showed the importance of these proteins for filament integrity, diazotrophic growth and intercellular communication, consistent with their proposed involvement in intercellular exchange of cytoplasmic commodities (Mullineaux et al., 2008).

1.5.1.2 Ammonia assimilation and excretion

Amino acids can serve as an organic source of nitrogen for most organisms, with glutamine being the most favourable of all (Flores and Herrero, 2004). Nevertheless, ammonium is the preferred nitrogen source for most bacteria, since all forms of nitrogen (nitrate, nitrite, urea, cyanate and molecular nitrogen) are converted to ammonia first (Herrero et al., 2001).

Ammonia can freely diffuse through the cell wall and membranes of cyanobacteria; however, specific transporters are required for the translocation of ammonium, the ionic form of ammonia. The active transport of ammonium is performed by three different ammonia transporters, *amt1*, *amt4*, and *amtB* in *Anabaena* sp. PCC 7120 (Paz-Yepes et al., 2008). In a knockout mutant, removal of the complete *amt14B* cluster resulted in elevated levels of external ammonium, although the change did not affect diffusion-driven uptake (Paz-Yepes et al., 2008).

Nonetheless, high levels of ammonium are toxic to most cyanobacteria (Dai et al., 2014). Therefore, the synthesized ammonia normally does not build up within the cell; rather, it is rapidly incorporated into carbon skeletons mainly through the sequential action of two enzymes, glutamine synthetase (GS) and glutamate synthase (glutamine oxoglutarate aminotransferase, GOGAT) (Muro-Pastor et al., 2005). GS is the primary assimilator of ammonia, converting glutamate to glutamine (Luque and Forchhammer, 2008). In the presence of L-methionine sulfoximine (MSX), a glutamate analogue, GS is irreversibly inactivated (Ronzio et al., 1969), leading to the build-up and eventually the release of intracellular ammonia (Thomas et al., 1990; Singh and Tiwari, 1998). A phenotypically similar effect was reported due to the rational modification of the active site of GS. The aspartate residue within the conserved region of the enzyme's active site has been shown to have the highest impact on the binding of ammonia, and therefore on the overall activity of GS in *Anabaena azollae* (Crespo et al., 1999). The lost ammonia can, nevertheless, re-enter either via diffusion (in the form of ammonia) or through one of the transporters (in the form of ammonium), depending on the ambient pH. The most active of the three transporters, *amt1* is thought to be responsible for the recapture of ammonium leaked out from the cells in *Synechococcus* sp. PCC 7942 (Vázquez-Bermúdez et al., 2002).

The other protein involved in nitrogen assimilation is the GOGAT enzyme, supplying glutamate for GS by combining glutamine and 2-oxoglutarate. The GS–GOGAT cycle this way constitutes the link between nitrogen and carbon metabolism (Martin-Figueroa et al., 2000). Cyanobacteria may possess two types of GOGAT different in the source of the reductant they require: NADH-GOGAT utilises reduced NAD coenzyme, whereas Fd-GOGAT depends on reduced ferredoxin (Muro-Pastor et al., 2005). In *Anabaena* sp. PCC 7120 only the ferredoxin-dependent enzyme could be identified (Martin-Figueroa et al., 2000). The operation of the GS–GOGAT cycle requires two direct photosynthetic products (ATP and reducing power), as well as a supply of carbon skeletons in the form of 2-oxoglutarate. The main producer of 2-oxoglutarate is isocitrate dehydrogenase (IDH), an enzyme in the TCA cycle. Cyanobacteria lack the 2-oxoglutarate dehydrogenase complex (OGDH), the primary consumer of 2-oxoglutarate in the TCA cycle. The predominant role of 2-oxoglutarate in these organisms may therefore be to supply carbon skeleton for ammonia assimilation, even though recent studies showed the presence of alternative bypasses to complete the cyanobacterial TCA cycle (Zhang and Bryant, 2011; Steinhauser et al., 2012; Knoop et al., 2013).

Glutamate and glutamine are the major nitrogen donors for the synthesis of other nitrogen-containing molecules. Glutamine is consumed in many different transamination reactions to produce the majority of amino acids. The GS–GOGAT pathway is tightly regulated in cyanobacteria by adjusting the activity of GS. When carbon is abundant, nitrogen deficiency results in high level of GS activity. In

contrast, when nitrogen becomes abundant, GS activity is downregulated (Luque and Forchhammer, 2008).

A secondary route of ammonia assimilation might occur via glutamate dehydrogenase (GDH) that is present in various cyanobacteria. However, physiological studies could not demonstrate any role of GDH in ammonia assimilation (Muro-Pastor et al., 2005; Luque and Forchhammer, 2008). The primary routes of nitrogen assimilation in nitrogen-fixing cyanobacteria and the interplay between metabolic and regulatory networks is summarized in Figure 1.4.

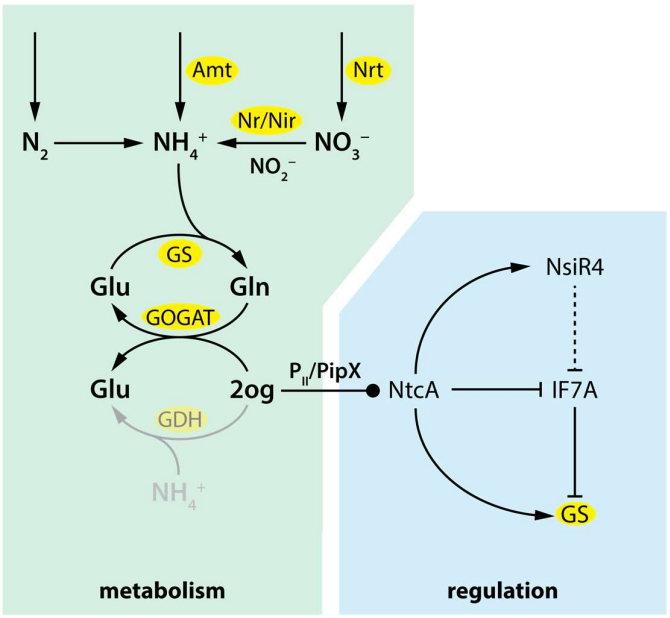


Figure 1.4. Interplay between metabolic and regulatory pathways of nitrogen assimilation in *Anabaena* sp. PCC 7120. Nitrogen is acquired in different forms via diffusion or specific transporters (Amt and Nrt). GS is the primary assimilator of nitrogen, acting in a tightly controlled cycle with GOGAT. An increase in the 2og pool activates the expression of NtcA. NtcA upregulates GS and NsiR4, while represses the expression of IF7A. NsiR4 may also downregulate IF7A. When 2og returns to normal level, NtcA deactivates, and IF7A expression starts. Consequently, GS gets inactivated. Lines with arrowheads indicate metabolic reactions (green area) or upregulation (blue area). Lines ending in a bar show repression. GDH is greyed out as it is not the primary ammonia assimilator. Regulation by NsiR4 (dashed line) is hypothetical. Enzyme names are highlighted in yellow.

Regulation of GS activity in *Synechocystis* sp. PCC 6803 is performed by two inactivation factors, IF7 and IF17 (García-Domínguez et al., 1999). IF7 and IF17 are small polypeptides that inhibit the activity of GS by direct protein-protein interaction (García-Domínguez et al., 1999; Galmozzi et al., 2010). In

Anabaena sp. PCC 7120, the system operates with only one factor called IF7A (Galmozzi et al., 2010). The expression of the inactivation factors is under the control of the transcriptional regulator NtcA (Muro-Pastor et al., 2005) that also regulates the expression of GS and IDH, and in this way influences the carbon skeleton supply and the GS–GOGAT pathway function (Herrero et al., 2001; Herrero et al., 2004). Thus, NtcA is a global transcriptional regulator of nitrogen-stress-induced processes in all cyanobacteria (Herrero et al., 2001). It belongs to the cyclic AMP receptor protein family of regulators and binds to specific DNA sites found in the promoter regions of multiple genes involved in nitrogen assimilation and related functions (Luque et al., 1994; Flores and Herrero, 2010).

Under nitrogen deprivation in *Anabaena* sp. PCC 7120, NtcA is expressed and the *gifA* gene encoding for IF7A is repressed, increasing the activity of GS. Following the addition of ammonium to the medium, expression of *gifA* is derepressed, leading to the synthesis of IF7A, and consequently, GS is inactivated (Galmozzi et al., 2010). Recently, it has been discovered that NtcA is not the only regulator of *gifA* expression, but small RNA transcripts may also be involved (Mitschke et al., 2011). NsiR4, a small non-coding RNA (sRNA) has been identified in *Synechocystis* sp. PCC 6803 and *Anabaena* sp. PCC 7120 as well, and it has been found targeting the *gifA* mRNA in both organisms. Moreover, in *Synechocystis* sp. PCC 6803 NsiR4 has decreased the abundance of the GS inactivation factors, and this way affected the intracellular GS pool as well (Klähn et al., 2014).

1.5.1.3 Nitrogenase enzyme complex

Nitrogenase is a highly conserved enzyme complex. It is found in a diverse group of prokaryotes from the bacteria and archaea (Zehr et al., 2000; Zehr et al., 2001), but is not encoded in any eukaryotic genome (Berman-Frank et al., 2003). Phylogenetic analyses suggest all extant nitrogenases sequenced so far are derived from a single common ancestor and that the catalytic subunits of the enzyme complex indicate the enzyme existed prior to oxygenation of Earth's atmosphere (Broda and Peschek, 1983); which may also explain the deleterious effects of oxygen on nitrogenase activity (Fay, 1992).

In many diazotrophs, nitrogenase comprises about 10% of total cellular proteins and consists of two components, an iron protein (Fe-protein) and an iron-molybdenum protein (MoFe-protein) (Berman-Frank et al., 2003). However, alternative nitrogenases have also been found that are homologous to the described enzyme, yet have vanadium or iron substituting for molybdenum (Thiel, 1993; Eady, 1996). The catalytic efficiency of these alternative nitrogenases is lower than that of the MoFe-nitrogenase; the specific activity of the VFe-nitrogenase is about 1.5 times lower than that of the MoFe-nitrogenase at 30 °C (Miller and Eady, 1988). It is nonetheless common to all nitrogenases that

their activity is destroyed by oxygen (Fay, 1992). Therefore, to keep the otherwise tightly controlled and highly complex nitrogenase intact, cyanobacteria evolved two distinct strategies to protect the enzyme from the harmful action by oxygen (Peschek et al., 1991).

First, a temporal separation without the need for a microaerobic environment is observed for non-heterocystous cyanobacteria. These organisms fix nitrogen at night and nitrogenase is typically found in all cells (Berman-Frank et al., 2003; Pfreundt et al., 2012). In the unicellular diazotroph *Cyanothece* ATCC 51142 and other members of the same genus, nitrogen is fixed during the dark (or during the subjective dark phase when grown under continuous light), controlled by cues from the diurnal cycle and strain-specific intracellular metabolic signals (Bandyopadhyay et al., 2013). High nitrogenase activity coincides with high respiration rates, and with a phase difference of 12 h from the peak of photosynthetic activity. This pattern is also reflected at the transcriptional level, and observed under either continuous light or darkness, implicating circadian control (Colón-López et al., 1997; Schneegurt et al., 2000).

On the other hand, a highly refined specialization is found in heterocystous cyanobacteria. In these organisms, nitrogenase is confined to a microaerobic cell, the heterocyst, which differentiates completely and irreversibly 12–20 h after combined nitrogen sources are removed from the medium (Golden and Yoon, 2003; Kumar et al., 2010). These cells are characterized by a thick multilayer cell wall that slows the diffusion of oxygen; high photosystem I (PSI) activity, the absence of water-splitting photosystem II (PSII) and carboxysomes, and loss of division capability (Wolk et al., 2004; Awai et al., 2010; Flores and Herrero, 2010). As a consequence of the missing PSII function in their heterocysts, these organisms cannot obtain reductant directly from non-cyclic photosynthetic electron flow, and rely on the supply of fixed carbon from adjacent vegetative cells for reducing equivalents (Böhme, 1998; Meeks and Elhai, 2002; Kumar et al., 2010). Similarly, the high ATP requirement of nitrogenase is met by cyclic electron flow around PSI, rather than being supplied by the linear electron transport chain (Wolk et al., 2004). However, the importance of PSI activity in these specialized cells is twofold. Besides supplying ATP for nitrogen fixation, PSI in the heterocyst effectively acts as a photon-catalyzed oxydase, consuming molecular oxygen through pseudocyclic electron transport (Asada, 1999; Milligan et al., 2007). The resulting Mehler reaction (Mehler, 1951) produces reactive oxygen species (e.g. superoxide anion) that are rapidly converted by the sequential action of superoxide dismutase and either catalase or peroxidase to minimize oxidative stess (Zhao et al., 2007a; Zhao et al., 2007b). Thus, PSI has an essential role also in protecting nitrogenase from oxygen in heterocysts.

The expression of nitrogenase is activated during the late stages of heterocyst differentiation, by completion of three developmentally regulated DNA rearrangements (Golden et al., 1985). These

chromosomal rearrangements involve the excision of elements from *nifD* (encoding the α-subunit of nitrogenase), *fdxN* (encoding a bacterial-type ferredoxin) and *hupL* (encoding the large subunit of an uptake hydrogenase) genes (Golden et al., 1988; Matveyev et al., 1994; Carrasco et al., 1995). The three excision events are required for the restoration of function to the respective operons that are necessary for the emergence of nitrogenase activity.

Under anaerobic conditions some heterocystous cyanobacteria, such as *Anabaena variabilis*, can synthesize a different Mo-dependent nitrogenase (Nif2) in the vegetative cells (Thiel et al., 1995). Nif2 is expressed shortly after nitrogen depletion but prior to heterocyst formation, and can support the fixed nitrogen needs of the filaments independently of Nif1 nitrogenase in the heterocysts (Thiel and Pratte, 2001).

1.5.2 Importance of iron

The nitrogenase enzyme complex is an iron-rich metalloprotein consisting of two components: a MoFe protein, called dinitrogenase and a Fe-containing protein, dinitrogenase reductase. The reduction of dinitrogen by nitrogenase involves three basic types of electron transfer steps: first, the reduction of Fe-protein by electron carriers such as ferredoxin or flavodoxin; second, the transfer of single electrons from Fe-protein to MoFe-protein; and third, electron and proton transfer to the substrate at the surface of the MoFe-protein (Georgiadis et al., 1992). The component proteins of nitrogenase are themselves quite complex.

Dinitrogenase is a tetramer composed of two pairs of different subunits, harbouring four (4Fe-4S) clusters and two molecules of MoFe cofactor. The MoFe cofactor is an essential component of dinitrogenase, it contains eight Fe and seven or eight S atoms per Mo atom, without forming 4Fe-4S clusters. Dinitrogenase reductase, on the other hand, is a dimer composed of two identical subunits with a single bridging 4Fe-4S cluster (Peters et al., 1995). Thus, nitrogenase itself can contain up to 38–50 iron atoms per complex (Kirn and Rees, 1992; Anand, 1998; Tuit et al., 2004). In addition, the expression of iron-containing electron transport proteins (e.g. cytochromes and ferredoxin), which provide reductants for the nitrogenase system, is also increased under diazotrophic conditions (Rueter, 1988), contributing to a higher demand for iron under diazotrophic conditions. Indeed, both marine and freshwater cyanobacteria have been shown to require up to 5–10 times as much iron for nitrogen-fixing cultures to grow at the same rate as those grown on nitrate (Lammers, 1982; Kustka et al., 2003a; Kustka et al., 2003b). Therefore, the availability of iron largely influences N_2 fixation in

cyanobacteria from its direct effect on iron-rich protein synthesis, to effects on photosynthesis, growth and also global productivity (Falkowski, 1997; Berman-Frank et al., 2001).

In fact, limitation by iron is one of the most common stress factors in cyanobacterial communities that restricts nitrogen fixation (Mills et al., 2004). Not surprisingly, FurA, the global transcriptional regulator of iron homeostasis in cyanobacteria, has been found influencing the heterocyst differentiation regulatory cascade as well, and even the expression of the nitrogenase (*nifHDK*) operon via NtcA (López-Gomollón et al., 2007a; López-Gomollón et al., 2007b; González et al., 2013). By modulating the expression of NtcA, FurA creates a key connection between iron homeostasis and nitrogen fixation during iron limitation.

To alleviate the limitation of iron, many marine heterotrophic bacteria and some cyanobacteria produce siderophores, small organic molecules that tightly bind iron and thereby increase its solubility (Vraspir and Butler, 2009). The bacteria then take up the siderophores via outer-membrane transporters that are specific for different groups of siderophores. By contrast, eukaryotic phytoplankton are not known to produce siderophores or to directly take up bacterially derived iron(III)-siderophore complexes (Amin et al., 2009). Instead, many eukaryotic phytoplankton release porphyrins, another type of iron-complexing metabolites relatively unavailable for uptake to cyanobacteria (Hutchins et al., 1999). Porphyrins are tetrandentate ligands stabilizing with iron(III) in four coordinate bonds (Hutchins et al., 1999), whereas bacterial siderophores typically form hexadentate ligands (Goldman et al., 1983). The competition between marine prokaryotes and eukaryotes for organically-bound iron may therefore depend on the chemical nature of available iron complexes, with consequences for ecological niche separation (Bailey and Taub, 1980; Geider, 1999; Hutchins et al., 1999).Eukaryotic fungi, on the other hand, may produce a variety of chelators in iron-limited environments, similar to bacterial siderophores (Haas et al., 2008; Hopkinson and Morel, 2009).

1.5.3 Natural and synthetic communities

In nature, no organism lives in isolation, but rather as part of an ecosystem of varying diversity. Yet, cultivation of microorganisms for biotechnology generally focuses on single species, particularly where a product is required at high purity (Kazamia et al., 2014). Such biological systems are, however, unstable and prone to perturbation, causing high risk of contamination by invasive species (Kazamia et al., 2012). In comparison, natural microbial communities have been recognized for their higher robustness against fluctuations in environmental conditions, although the mechanisms underlying

their stability have long been debated (Goodman, 1975; McNaughton, 1977; McCann, 2000; Girvan et al., 2005; Keith et al., 2010; de Vries and Shade, 2013). Furthermore, microbial communities exhibit emergent biochemical properties not found in clonal monocultures, such as resistance to toxins (Pelz et al., 1999), production of uncommon secondary metabolites (Pettit, 2009) or cross-feeding of essential metabolites to complement each other's auxotrophy (Wintermute and Silver, 2010). Understanding of microbial communities is therefore of great importance in order to design and exploit features relevant for biotechnology (Ortiz-Marquez et al., 2013).

The cost reduction of sequencing technologies and the development of metagenomics approaches (Handelsman, 2005) has led to the characterization of species diversity in different natural communities (Raes and Bork, 2008). However, natural microbial communities characterized to date contain tens to more than thousands of species (Curtis et al., 2002). Therefore, it is usually not possible to verify which species are actively exhibiting a community feature or performing key functions (Jessup et al., 2004), significantly limiting our understanding of the molecular and ecological bases of community-level functions, such as diversity, structure and size (de Vries and Shade, 2013). A promising way to overcome the difficulties associated with studying natural communities is to create artificial communities that retain the key features of their natural counterparts (Großkopf and Soyer, 2014). Such simple synthetic communities can act as model systems to assess mechanisms that control the dynamic behaviour of the more complex natural consortia (Widder et al., 2016). Moreover, synthetic communities may exhibit novel features that single species alone do not, allowing for the establishment of rational engineering strategies for biotechnological applications (Ortiz-Marquez et al., 2013; Kazamia et al., 2014).

Efficiency and stability of naturally occurring communities rely on the mutual benefit of all members in the relationship (Hay et al., 2004). Similarly, mutualism within synthetic communities is inevitable and can be driven by the reciprocal interchange of key metabolites (Klitgord and Segrè, 2010). Therefore, assembly of a synthetic community generally relies on the identification of these key metabolites as exchange commodities. As a next step, candidates that are potentially able to supply the exchange commodities have to be identified by comprehensive screening of metabolic data (Freilich et al., 2011) or designed by metabolic engineering to excrete the compound (Ortiz-Marquez et al., 2013). In microalgal-bacterial consortia, the commodities may include cross-feeding of essential vitamins (Croft et al., 2005; Kazamia et al., 2012), iron (Beliaev et al., 2014) or ammonium (Ortiz-Marquez et al., 2012) in exchange for photosynthetically fixed carbon. In fact, the exchange of nitrogen (ammonium) for carbon is not uncommon in biology. Similar phenomenon occurs in the aforementioned *Azolla-Anabaena* symbiosis or at a smaller scale, within the filaments of *Anabaena*

sp. PCC 7120 (Hill, 1977; Lechno-Yossef and Nierzwicki-Bauer, 2002; Kumar et al., 2010). Under nitrogen starvation, a few vegetative cells at approximately equal distance within the filament terminally differentiate into heterocysts harbouring the oxygen-sensitive nitrogenase (Golden and Yoon, 2003). In the microoxic environment provided in a mature heterocyst, nitrogenase converts atmospheric dinitrogen to ammonia and actively distributes the essential commodity to adjacent photosynthetic vegetative cells (Meeks and Elhai, 2002). Heterocysts in return, rely on vegetative cells for carbon skeletons and other photosynthates (Kumar et al., 2010). The comprising metabolic interactions between the two cell types render vegetative cells and heterocysts mutually interdependent on one another, providing a very simple "ecosystem" for the elucidation of community metabolism. Furthermore, heterocysts' ability to supply fixed nitrogen to closely associated cells makes *Anabaena* sp. PCC 7120 a potential candidate for the development of nitrogen-independent synthetic communities.

1.6 Challenges in metabolic engineering of *Anabaena* sp. PCC 7120

1.6.1 Developing synthetic biology

Cyanobacteria have gathered recently the attention of metabolic engineering efforts as carbon-neutral organisms, potentially capable of producing biofuels or other valuable compounds. Their key advantage over other bacteria is their ability to use photosynthesis to harness energy from sunlight and convert CO_2 into the product of interest. As compared with eukaryotic algae and especially plants, cyanobacteria are easier to manipulate genetically and grow considerably faster (Berla et al., 2013). This is not the case, however, in comparison with bacteria and yeast. In fact, while cyanobacteria feature the unique advantages of autotrophs and diazotrophs in biotechnology applications, development of synthetic biology tools has lagged way behind that for *Saccharomyces cerevisiae* and *Escherichia coli*. In particular, there are only a handful of controllable promoters, expression vectors or neutral genomic sites available for cyanobacteria, posing a challenge to routine metabolic engineering. Most of the available inducible promoters respond to ambient concentrations of transition metals. The dynamic range of these promoters can be quite high, reaching from a few-hundred-fold to thousand-fold expression upon induction, while some of them are known to be leaky in the absence of the environmental trigger (Wang et al., 2012; Berla et al., 2013). The commonly used cyanobacterial inducible promoters are P_{petE} (copper-induced), P_{psbA2} (light-induced), P_{rbcLS} (light-induced), P_{nrsB} (nickel-induced), P_{nir} (nitrate-induced) and P_{nifHDK} (nitrogen-starvation-induced).

The most studied among all cyanobacteria are the unicellular freshwater *Synechocystis* sp. PCC 6803, the unicellular marine *Synechococcus* sp. PCC 7002 and the filamentous nitrogen-fixing *Anabaena* sp. PCC 7120. Many cyanobacteria are capable of natural transformation and homologous recombination, but almost all can be modified by conjugal gene transfer (Koksharova and Wolk, 2002). The currently available synthetic biology tools for engineering cyanobacteria are described exhaustively in recent reviews (Koksharova and Wolk, 2002; Heidorn et al., 2011; Wang et al., 2012; Berla et al., 2013). For the creation of markerless chromosomal mutations the most common method is the use of a conditionally toxic gene linked to an antibiotic resistance cassette. In the first step, both markers are inserted into the chromosome, with selection for antibiotic-resistant mutants. Next, due to a second transformation or induction of intrachromosomal recombination the resistance cassette and toxin gene are deleted, and markerless mutants are selected which have lost the toxic gene (Cai and Wolk, 1990). This principle has been used in cyanobacteria with the *Bacillus subtilis* levansucrase synthase gene *sacB*, which confers sucrose sensitivity (Cai and Wolk, 1990; Eaton-Rye, 2004), as well as with *Escherichia coli mazF*, a general protein synthesis inhibitor expressed under a nickel-inducible promoter (Cheah et al., 2013). Both methods allow the reuse of a single selectable marker for introducing multiple successive changes to the cyanobacterial chromosome. Recombinase-based techniques have also been developed to engineer mutants that lack a selectable marker, including the Cre-LoxP system in *Anabaena* sp. PCC 7120 (Zhang et al., 2007), or the flipase-based FLP/FRT system in *Synechocystis* sp. PCC 6803 (Tan et al., 2013). However, these methods leave a short scar in the sequence that may potentially lead to undesirable crossover events or other unexpected results in case of multiple mutations introduced to the same strain lineage.

Although a few shuttle (expression) vectors have been developed for cyanobacteria, there has been little characterization of their copy numbers in cyanobacterial hosts. The most commonly used shuttle vectors include pDU1 in *Anabaena* spp., and the broad-host-range vector pRSF1010 (Wang et al., 2012). Plasmids derived from pRSF1010 appear to have a copy number of 10–30 in *Synechocystis* sp. PCC 6803 (Huang et al., 2010), but copy numbers of other broad host-range plasmids have not been quantified until recently (Berla et al., 2013). Endogenous cyanobacterial plasmids with copy numbers ranging between 1 and 8 per chromosome may also be used as target sites for gene expression (Xu et al., 2011; Berla and Pakrasi, 2012). The remarkable advantage of expression vectors over chromosomal integration is their higher rate of successful transformation and the independence from homologous recombination as well as laborious genetic segregation.

1.6.2 Homologous recombination favours single recombination

Gene transfer in cyanobacteria can be performed by transformation and conjugation, but not viral transduction (Koksharova and Wolk, 2002). In addition, only a few cyanobacteria are naturally transformable by exogenous DNA, such as *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002 (Shestakov and Khyen, 1970; Stevens and Porter, 1980). The filamentous *Anabaena* sp. PCC 7120 on the other hand, can be transformed via conjugation and, to a lesser efficiency, by electroporation. Plasmids to be transferred via conjugation can be made mobilizable by provision of an *oriT* fragment (origin of transfer) of RP4. An RP4-encoded enzyme nicks its own double stranded DNA at *oriT*, providing single-stranded DNA for transfer (Koksharova and Wolk, 2002). Once inside the host strain, foreign DNA bearing homologous sequences to a genomic locus can participate in homologous recombination with the host strain's genome (Golden et al., 1987). This site-directed modification of the chromosome has been applied successfully to many unicellular cyanobacteria because of the ease with which double recombinants can be isolated from these organisms (Cai and Wolk, 1990). In contrast, isolation of double recombinants from filamentous cyanobacteria such as *Anabaena* species has been difficult for two reasons. First, when homologous sequences of DNA within plasmids are transferred to *Anabaena* spp. by conjugation, single-crossover events (integration recombinations) occur far more frequently (by about two orders of magnitude) than double-crossover (replacement recombination) events (Golden and Wiest, 1988; Cai and Wolk, 1990). Indeed, exhaustive screens have failed to isolate double recombinants when the size of the homologous region was below 4 kb (Cai and Wolk, 1990). Second, because cells of *Anabaena* spp. have multiple genomic equivalents (similar to other cyanobacteria), isolation of recessive mutants requires both segregation of mutant and wild-type forms of the genome and physical detachment of adjacent cells of a filament bearing the two genomic forms (Flores and Herrero, 2010; Griese et al., 2011).

1.6.3 Oligoploid cyanobacteria

Ploidy of an organism determines the copy number of homologous chromosomes in the genome of a single cell. There are many truly monoploid bacteria that carry each of their chromosomes in a single genomic copy (Pecoraro et al., 2011), although multicopy chromosomes are also very common. In fact, the number of known oligoploid and polyploid prokaryotic species outnumbers the monoploid species and it seems that monoploidy is not typical for prokaryotes, in contrast to the general belief (Griese et al., 2011). Similarly, in case of cyanobacteria monoploid species are somewhat exceptional, rather than being the majority. Most of confirmed monoploid cyanobacteria belong to the *Synechococcus*

genus (Armbrust et al., 1989; Binder and Chisholm, 1995), whereas *Anabaena* spp. are typically oligoploid and *Synechocystis* spp. were found highly polyploid (Griese et al., 2011). In particular, *Anabaena* sp. PCC 7120 contains on average eight copies of its chromosomal DNA (Hu et al., 2007). The ploidy level is therefore highly variable among cyanobacteria, even within the same genus. There is no obvious correlation between the number of genome copies, genome size, growth conditions and doubling time (Griese et al., 2011). However, various evolutionary advantages of oligo- and polyploidy exist for cyanobacteria. These advantages include resistance against double stranded breaks due to higher copy numbers (Domain et al., 2004), the lack of stringent control of equal chromosome segregation (Hu et al., 2007) and gene redundancy, opening the possibility that under unfavourable conditions, mutations are induced in some genome copies, whereas the wild type information is retained in others (Takahama et al., 2004; Nodop et al., 2008). Especially this last phenomenon, i.e. some genome copies in a mutant strain retain the original genotype, makes the isolation of clean mutant genotypes a laborious, time consuming process. Genetic segregation is therefore a crucial step in the genetic engineering of multiploid cyanobacteria, essentially required following every successful transformation event.

1.6.4 Multicellularity in filamentous cyanobacteria

Cyanobacteria have greatly diversified through evolution, producing both unicellular and multicellular forms. The multicellular forms consist of hair-like trichomes (filaments) that, in some strains, can contain hundreds of cells (Rippka et al., 1979). In the filaments of most heterocyst-forming cyanobacteria, such as *Anabaena* spp. and *Nostoc* spp., the cells divide in only one plane, but in the filaments of some other species, such as *Fischerella* spp., the cells can divide in more than one plane, producing branched filaments (Rippka et al., 1979). Filaments mainly contain photosynthetic vegetative cells, which, upon certain environmental triggers, differentiate into highly specialized cell types. Certain strains produce motile hormogonia, spores called akinetes or nitrogen-fixing heterocysts (Rippka et al., 1979). The structure of the filament bears a thick Gram-negative type of cell envelope containing an outer membrane encapsulating the cytoplasmic membrane of each cell in the filament. In case of heterocyst-forming cyanobacteria the outer membrane does not enter the septum between adjacent cells (Ris and Singh, 1961). Rather, the outer membrane appears to be continuous in these species. The presence of a continuous outer membrane suggests that the periplasmic space, which lies between the cytoplasmic membrane and the outer membrane, is also continuous and may be involved in intracellular communication and commodity exchange (Flores et al., 2006). In *Anabaena* sp. PCC 7120, adhesion of adjacent cells and filament integrity involves the

function of at least three genes, *sepJ*, *fraC* and *fraD* (Flores et al., 2007; Merino-Puerto et al., 2010; Merino-Puerto et al., 2011). Considering the fact that reproduction in heterocyst-forming cyanobacteria takes place by random filament breakage, the above two genes may be important in maintaining filament length and thus, colony structure. The filament, which is viewed as a string of cells encapsulated by the outer membrane, can be considered as an organismic unit in these bacteria (Flores and Herrero, 2010).

In a multicellular, oligoploid cyanobacterium, such as *Anabaena* sp. PCC 7120 a confirmed transformation is followed by the disjunction of transformed and untransformed cells by mechanical disruption of the filament (e.g. via sonication). True positive transformants (single or multiplet cells that actually carry the desired new genotype) are segregated to isolate homozygous strains that lack any wild type chromosome and only carry the modified DNA in a copy number that is characteristic to the organism's ploidy level. In *Anabaena* sp. PCC 7120 in particular, single recombination (insertion) is far more common than double recombination (replacement) and therefore, markerless mutations through double recombination require strong positive selection and screening. Despite the challenges in isolating a particular engineered strain of *Anabaena* sp. PCC 7120, its metabolic potential, simple cultivation and relatively rich set of available molecular tools make this organism a promising target of synthetic biology endeavours.

1.7 Aims and objectives

The aim of this thesis was to investigate whether diazotrophic cyanobacteria can be genetically modified for the excretion of highly-available fixed nitrogen through systemic understanding of the organism's metabolism. The rationale for the development of such engineered photodiazotrophs is the urgent need for more efficient fertilizer distribution in modern agriculture. This thesis attempts to provide a proof-of-concept example for strategies that may tackle present day's nitrogen crisis by achieving the following objectives:

(1) Identification of an ideal host organism for nitrogen excretion. Many soil bacteria, both anaerobic (e.g. *Klebsiella penumoniae*) and aerobic (e.g. *Azotobacter vinelandii*) are capable of nitrogen fixation. However, fertilization on soil surface prefers an aerobic organism that is also efficient in aquatic environments (e.g. as a side crop in rice paddies). The aerobic *Azotobacter vinelandii* could be an option, although its very high respiratory activity to prevent oxygen damage of nitrogenase may be undesired (Poole and Hill, 1997). Cyanobacteria, on

the other hand, are aerobic and also carbon neutral that, instead of high respiratory rates, prevent damage of nitrogenase by temporal (circadian) or spatial separation from their oxygenic processes. The former group fixes nitrogen only at night (e.g. *Cyanothece* sp. PCC 7822), whereas the latter develops a confined microoxic environment for nitrogenase, allowing continuous nitrogen fixation (e.g. *Anabaena* sp. PCC 7120). The filamentous, heterocyst-forming *Anabaena* sp. PCC 7120 is one of the most studied cyanobacteria and also a model organism for photosynthetic nitrogen fixation. Its biochemical properties and potential in metabolic engineering are discussed in Chapter 1.

(2) Reconstruction of the host organism's metabolic network in order to:

a. Understand the factors that may influence nitrogen metabolism from a broader perspective. The stoichiometric model was used to simulate metabolic states of *Anabaena* sp. PCC 7120 under non-diazotrophic conditions as a single vegetative cell, and under diazotrophic conditions as a filament of mutually interdependent alternating cell types, represented by a two-cell association of a heterocyst and a vegetative cell. The two modes (i.e. single-cell and two-cell) were compared for the utilisation of different carbon and nitrogen sources; metabolite exchange between the two cell types in the two-cell model was also studied. Details are given in Chapter 3. Due to the essentiality of iron as an active component of the nitrogenase enzyme, iron acquisition under limiting conditions was also investigated in Chapter 5.

b. Determine the ideal candidate metabolite for the release of fixed nitrogen. Several different nitrogen-containing metabolites were tested as potential candidates for the release of combined nitrogen using the stoichiometric model. Both urea and ammonia were found very promising in terms of yield; however, ammonia was selected as the primary target due to its preference as a nitrogen source over urea for most microorganisms (Herrero et al., 2001) and its carbon neutral nature. The yield of the different nitrogen compounds is discussed in Section 3.2.7.

c. Identify metabolic engineering strategies that promote nitrogen excretion. Using predictions and the visual representation of the model (Appendix I), as well as literature data, glutamine synthetase (GS) was identified as the main target for metabolic engineering. The effect of a repressed GS activity on ammonia excretion was demonstrated by administering a specific inhibitor. For a similar, but stable phenotypic effect, three engineering strategies were designed that decrease GS activity one way or another. The strategies are explained in Section 4.1 and 4.2.

(3) Evaluate the feasibility of engineering strategies *in vivo*. Wild type and ammonium transporter knockout (Δ*amt*) strains of *Anabaena* sp. PCC 7120 were modified genetically to excrete ammonium. Results are collected in Chapter 4. Difficulties in isolating the expected genotypes are also discussed.

(4) Provide strains for testing in co-cultivation. The most promising strains that may excrete ammonia under diazotrophic conditions were sent for evaluation in co-cultivation with a microalga. The comprising synthetic community and the related encouraging results are described in Section 4.2.3.5 and discussed in Section 4.3.

2 Materials and methods

2.1 Chemicals and reagents

Chemicals and reagent components, unless otherwise noted, were ordered form Sigma-Aldrich (Sigma-Aldrich Co. Ltd., Dorset, UK), VWR (VWR International Ltd., Lutterworth, UK) or Thermo Fisher Scientific (Fisher Scientific Ltd., Loughborough, UK).

2.2 General protocols

2.2.1 Incubators and sterile work

Cyanobacterial liquid cultures were incubated in an AlgaeTron AG 230 light incubator (Photon Systems Instruments, spol. s r.o., Czech Republic) at 1% elevated CO_2 controlled by an Evolution horticultural carbon dioxide controller (Ecotechnics UK Ltd., UK). Agar plates of cyanobacteria were incubated in a Sanyo MIR-253 chamber (Panasonic Biomedical B.V., UK) at ambient CO_2.

Bacterial liquid cultures were grown in a New Brunswick shaker incubator (New Brunswick 44R, Eppendorf, UK). Bacterial agar plates were incubated in a Labnet Mini Incubator (ST2185; Appleton Woods Ltd., UK).

Sterile work with both bacteria and cyanobacteria was performed in class I or class II biosafety cabinets under continuous laminar flow.

2.2.2 Laboratory centrifuges

Microcentrifuge tubes (1.5–2-ml) were centrifuged in an Eppendorf 5424R (5404000669, Eppendorf, UK) or a Heraeus Fresco 21 (Thermo Fisher Scientific, UK) microcentrifuge. Centrifuge tubes (15 and 50-ml) were centrifuged in a Heraeus Megafuge 16R benchtop centrifuge (Thermo Fisher Scientific, UK). Unless otherwise noted, centrifugation was performed at room temperature.

2.3 Strains and plasmids

Strains and plasmids that have been provided by others are listed in Table 2.1. All other DNA constructs, plasmids and oligonucleotide primers, unless otherwise noted, were generated during the course of this work and are collected in Table 4.1, Table 4.2, Table 4.4 and Table 5.4.

Table 2.1. Strains, plasmids and primers acquired from commercial or academic sources.

Name	Description[a]	Source
Strains		
Anabaena sp. PCC 7120		
WT	wild-type	Prof Enrique Flores[b]
CSP19	Δ*amt*::C.K3; Amt4^{-}, Amt1^{-}, AmtB^{-}, Nm^{R}	(Paz-Yepes et al., 2008)
Escherichia coli		
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Invitrogen (Thermo Scientific, UK)
HB101	*recA13 mcrBC*, host for pRL623 (helper strain)	(Elhai et al., 1997)
ED8654	host for pRL443 (conjugative strain)	(Elhai et al., 1997)
Plasmids		
pRL443	conjugative plasmid; Cb^{R}	(Elhai and Wolk, 1988)
pRL623	helper plasmid; DNA methylases (M.*Ava*I, M.*Eco*47II, M.*Eco*T22I), Cm^{R}	(Elhai et al., 1997)

Name	Description[a]	Source
pK19mob*sacB*	cloning vector; Nm^R, Suc^S	(Schäfer et al., 1994)
pDF-lac	template for *aadA* cassette (Sm^R, Sp^R)	(Guerrero et al., 2012)
pVZ322	template for the self-replicative pRSF1010 derivatives	Prof Wolfgang Hess[c]
Primers		
5'*glnA-Xba*I-F (40)	5'–GATCTCTAGATTCCTTCTTCTCCCAATCTTTG–3'	Mr Anthony Riseley[d]
3'*glnA-Bam*HI-R (41)	5'–GATCGGATCCAACTCGGTCTTCCTGCTGAA–3'	Mr Anthony Riseley[d]

[a] Abbreviations: Nm: neomycin/kanamycin, Sp: streptomycin/spectinomycin, Cb: carbenicillin, Cm: chloramphenicol; ^{R} denotes resistance to the corresponding antibiotic. Suc^S: sucrose sensitivity.

[b] Professor Enrique Flores is affiliated with IBVF, CSIC and Universidad de Sevilla, Spain.

[c] Professor Wolfgang Hess is affiliated with Genetics & Experimental Bioinformatics, University of Freiburg, Germany.

[d] Mr Anthony Riseley is affiliated with Department of Biochemistry, University of Cambridge, United Kingdom.

2.3.1 Cryopreservation of bacteria

Overnight cultures of *Escherichia coli* grown at 37 °C in 50-ml centrifuge tubes in 5 ml liquid LB supplemented with the appropriate antibiotics were centrifuged at 4,500 × g in a laboratory centrifuge for 5 min at room temperature. The supernatant was discarded and the cell pellet was resuspended in 1 ml fresh LB medium. Eight hundred microliters of the resulting cell concentrate was mixed with 300 µl sterile 80% (w/v) glycerol (autoclaved) in a 2-ml Nalgene cryogenic vial (Thermo Scientific, UK). Cryogenic tubes were placed immediately in paper boxes at -80 °C.

2.3.2 Cryopreservation of cyanobacteria

Cyanobacterial cultures were grown in 100-ml Erlenmeyer flasks until late exponential phase ($0.7 < OD_{730} \leq 1$ measured in a 1-cm cuvette) in 20 ml BG-11 liquid medium supplemented with the appropriate antibiotics, under standard growth conditions. The culture was harvested by centrifugation in 50-ml sterile centrifuge tubes at 3,000 × g in a laboratory centrifuge for 10 min at room temperature. The supernatant was discarded and the cell pellet was resuspended in 1 ml fresh BG-11. The resulting cell concentrate was transferred to a 2-ml Nalgene cryogenic vial (Thermo Scientific, UK) and gently mixed with 230 µl 80% (w/v) sterile glycerol (autoclaved) by inverting the tube a couple of times. The cryotube was immediately placed in a cooling block (pre-cooled at -20 °C

for 1 h) and transferred to -80 °C as quickly as possible. Cryotubes were moved to labelled paper boxes at -80 °C after 3 h in the cooling block.

2.4 Media and culturing conditions

2.4.1 Liquid LB (Luria-Bertani) medium

Standard liquid LB medium was prepared by dissolving 20 g LB Broth (Lennox formulation, L3022, Sigma-Aldrich, UK) to 1 l deionized water in a 1-l Schott screw cap bottle (VWR, UK). The mix was autoclaved at 121 °C for 15 min. Antibiotics were supplemented at room temperature as required.

2.4.2 LB agar medium

Solid LB agar medium was prepared similar to liquid LB medium by the addition of 15 g agar powder (20767.298, VWR, UK) into 1 l liquid LB medium. The mix was aliquoted in 333 ml portions in 500-ml Schott screw cap bottles and autoclaved at 121 °C for 15 min. Solid LB agar was melted by gentle boiling in a laboratory microwave oven. Antibiotics, when necessary, were supplemented at temperatures below 60 °C. Twenty millilitres of liquefied LB agar were distributed into 9-cm plastic Petri plates, solidified in sterile laminar flow and stored upside down at 4 °C for up to 3 weeks.

2.4.3 Liquid BG-11 medium

Liquid BG-11 medium was prepared according to (Rippka et al., 1979). A 100-times concentrated stock solution (100× BG-11) was prepared by dissolving 149.6 g $NaNO_3$, 7.49 g $MgSO_4 \cdot 7\ H_2O$, 3.6 g $CaCl_2 \cdot 2\ H_2O$, 0.89 g Na- citrate $\cdot 2\ H_2O$, 1.12 ml 0.25 M sodium EDTA pH 8.0 and 100 ml 1,000× trace minerals solution. The 1,000-times concentrated trace minerals solution was made as follows: 2.86 g H_3BO_3, 1.81 g $MnCl_2 \cdot 4\ H_2O$, 0.22 g $ZnSO_4 \cdot 7\ H_2O$, 0.39 g $Na_2MoO_4 \cdot 2\ H_2O$, 0.079 g $CuSO_4 \cdot 5\ H_2O$ and 0.0494 g $Co(NO_3)_2 \cdot 6\ H_2O$ were dissolved in 1 l deionized water and stored at 4 °C until being used. Standard BG-11 liquid medium was prepared by mixing 10 ml 100× BG-11 stock solution with 1 ml 1,000× ferric ammonium citrate (0.6 g ferric ammonium citrate dissolved in 100 ml deionized water), 1 ml 1,000× Na_2CO_3 (2 g Na_2CO_3 dissolved in 100 ml deionized water) and 1 ml 1,000× K_2HPO_4 (3.05 g K_2HPO_4 dissolved in 100 ml deionized water), and adjusting the volume to 1 l with deionized water. The mix

was autoclaved at 121 °C for 15 min in a 1-l Schott screw cap bottle and kept at room temperature until being used. Antibiotics, when appropriate, were added at room temperature. The bottle was mixed thoroughly prior to use to re-dissolve an orange-brown iron precipitation.

2.4.4 BG-11 agar medium

Solid BG-11 agar medium was prepared in 500-ml Schott screw cap bottles in 300-ml batches as follows: 3 ml 100× BG-11 stock solution, 0.3 ml 1,000× ferric ammonium citrate, 0.3 ml 1,000× Na_2CO_3, 0.3 ml 1,000× K_2HPO_4 and 5 ml 1 M TES-NaOH buffer pH 8.2 were mixed with 2.35 g $Na_2S_2O_3 \cdot 5\ H_2O$ and 4.5 g Difco Bacto agar (BD, Plymouth, UK), and the volume adjusted to 300 ml with deionized water. One molar TES-NaOH buffer was made by dissolving 57.3 g TES-Na (T0772, Sigma-Aldrich, UK) in 200 ml deionized water, adjusting the pH to 8.2 with 10 N NaOH (16 g NaOH dissolved in 30 ml deionized water and the final volume brought to 40 ml) and making the final volume 40 ml with deionized water. The buffer was filter sterilized using 0.2 µm syringe filters and stored at 4 °C. Bottles of pre-mixed BG-11 agar medium were autoclaved at 121 °C for 15 min and stored at room temperature. Solid BG-11 agar was melted gently by boiling in a laboratory microwave oven. Antibiotics, when appropriate, were added to liquefied agar at 60 °C or below. Twenty millilitres of liquefied BG-11 agar were distributed into 9-cm plastic Petri plates, solidified in sterile laminar flow and stored upside down at 4 °C for up to 3 weeks.

2.4.5 Nitrate-free medium

Liquid and solid nitrate-free BG-11_0 was prepared similar to BG-11 standard medium by leaving out sodium nitrate from the 100× stock solution and making a 100× BG-11_0 stock instead. The rest of the components were added as described for standard BG-11. Nitrate-free media were used to perform diazotrophic cultivation of strains of *Anabaena* sp. PCC 7120.

2.4.6 Bacterial cultivation

Strains of *Escherichia coli* were grown and maintained on LB agar plates at 37 °C in a benchtop Labnet Mini Incubator (ST2185; Appleton Woods Ltd., UK). Liquid cultures were inoculated by picking individual colonies using a 1-mm plastic inoculation needle into 5 ml liquid LB medium in 50-ml

centrifuge tubes. Cultures were grown at 37 °C with 200 rpm orbital shaking for 16 h (overnight) in a shaker incubator (New Brunswick 44R, Eppendorf, UK). Tube caps were loosened to allow proper aeration of the culture. Culturing media were supplemented with appropriate antibiotics at the following final concentrations: 100 µg/ml carbenicillin (Cb), 30 µg/ml chloramphenicol (Cm), 50 µg/ml kanamycin (Km) or 20 µg/ml streptomycin (Sm) combined with 50 µg/ml spectinomycin (Sp) was used in both solid and liquid media.

2.4.7 Cultivation of cyanobacteria

Plate cultures of *Anabaena* sp. PCC 7120 were grown on BG-11 agar plates in a Sanyo MIR-253 light incubator (Panasonic Biomedical B.V., UK) at 30 °C, under continuous illumination by fluorescent light at about 60 $\mu Em^{-2}s^{-1}$. Liquid cultures were inoculated from plates by collecting cells with a 1-mm plastic inoculation loop and washing them into 1 ml fresh BG-11. After about 8 days 20 ml fresh BG-11 was inoculated with the 1-ml pre-culture in a 100-ml Erlenmeyer flask covered with aluminium foil, and incubated at 30 °C in an AlgaeTron AG 230 light incubator (Photon Systems Instruments, spol. s r.o., Czech Republic) on a Titramax 1000 vibrational shaker (544-12200-00, Heidolph Instruments GmbH, Germany) at 180 rpm, under continuous illumination by cool white LED panels at 40 $\mu Em^{-2}s^{-1}$, in a 1-% enriched CO_2 atmosphere. For diazotrophic cultivation BG-11_0 was used, lacking sodium nitrate (and any other form of combined nitrogen). Solid and liquid media were supplemented with the appropriate antibiotics when necessary, at concentrations of 100 µg/ml for neomycin (Nm) or 2.5 µg/ml for Sm and Sp in case of plates, and 30–100 µg/ml Nm or 1–1.5 µg/ml Sm and Sp combined for any liquid medium.

2.5 Molecular biology methods

2.5.1 Amplification of DNA fragments

DNA fragments, inserts and plasmid backbones were amplified in polymerase chain reaction (PCR) using Q5 high-fidelity DNA polymerase (New England BioLabs, UK) and oligonucleotide primers (synthesized by Integrated DNA Technologies, UK) in a thermocycler (Biometra Professional Trio, Biometra, Germany). Reaction mixtures were assembled in PCR tubes on ice from the following reagents in a final 25 µl volume: 14.75 µl nuclease-free water, 5 µl 5× Q5 reaction buffer, 0.5 µl deoxynucleotide solution mix (dNTP), 1.25 µl product-specific forward primer, 1.25 µl product-specific

reverse primer, 2 µl template DNA and 0.25 µl Q5 high-fidelity DNA polymerase. The thermocycler was set to run the following program (Table 2.2):

Table 2.2. Program of a typical PCR amplification.

Denaturation of DNA	98 °C (30 sec)
Amplification, 35 cycles	98 °C (10 sec) T_a (20 sec) 72 °C (t_{ext})
Final extension	72 °C (10 min)
Pause	4 °C (forever)

Annealing temperature (T_a) was calculated for each reaction individually, based on the sequence of the primers used by an online tool (http://tmcalculator.neb.com) recommended for Q5 high-fidelity DNA polymerase. Extension time (t_{ext}) was calculated from the length of the desired DNA fragment and the speed of the polymerase (20–30 seconds/kilobase for the Q5 enzyme).

2.5.2 Agarose gel electrophoresis

Agarose gels were prepared by dissolving 1 g molecular biology grade agarose in 100 ml TAE buffer (40 mM Tris, 20 mM acetic acid and 1mM EDTA at pH 8.3) by boiling in a microwave oven until no undissolved crystals of agarose were visible. SYBR Safe DNA gel stain (S33102, Thermo-Fisher, UK) was added to the solution in a 1:10,000 dilution (10 µl to a 100 ml solution), mixed and poured into a gel casting tray with a plastic cone inserted. After solidification the cone was removed and the gel tray was transferred to an electrophoresis chamber (1704405, Bio-Rad Laboratories Ltd., UK) and covered completely in TAE buffer. Samples were mixed with a 6-times concentrated purple gel loading dye (B7024S, New England BioLabs, UK) to get a 6-fold dilution of the dye, and loaded to the bottom of a gel pocket. The chamber was set to 110 V for 40 min by a PowerPac Basic Power Supply unit (1645050, Bio-Rad). Resolved bands were visualized under blue light on a non-UV transilluminator (Dark Reader, DR88M, Clare Chemical Research, USA) or in a UV gel imager (GelDoc-It, UVP, Cambridge, UK).

2.5.3 DNA extraction and purification

Routine DNA extraction from bacterial cultures was performed using the QIAprep Spin Miniprep Kit (27106, Qiagen, Crawley, UK) and following the manufacturer's instructions. In short, bacterial cells were collected by centrifugation at 4,500 × g in a benchtop centrifuge and the supernatant was carefully discarded. Centrifuge tubes were dried upside down on paper towel for 10 min at room temperature. Cells were resuspended in 250 µl Buffer P1 containing RNase A and transferred to a 1.5-ml Eppendorf microcentrifuge tube. Samples were lysed by the addition of 250 µl lysis buffer (Buffer P2) and proteins precipitated by adding the neutralisation buffer (Buffer N3). Cell debris and other precipitates were removed by centrifugation at 17,900 × g in a table-top centrifuge. The resulting supernatant was filtered on the provided QIAprep spin columns to bind DNA to the column. Columns were washed with 750 µl Buffer PE (containing ethanol) and purified DNA was eluted to fresh 1.5-ml Eppendorf tubes by the addition of 30–50 µl nuclease-free water (New England BioLabs, UK) on the top of the column and centrifugation at 17,900 × g for 1 min.

Purification of DNA fragments following PCR amplification (liquid purification) or gel electrophoresis (gel purification) was carried out using the QIAquick Gel Extraction Kit (28706, Qiagen, Crawley, UK) and following the manufacturer's provided protocol. In short, DNA fragments were excised from agarose gel using a scalpel and the gel slice was weighed on an analytical balance in a 1.5-ml Eppendorf microcentrifuge tube. Three volumes of Buffer QG was added to 1 volume of gel (where 100 mg gel equals 100 µl). Similarly, Buffer QG was added to liquid samples from PCR amplification in a 3-fold excess to the sample volume. Tubes were incubated at 50 °C for 10 min or until the gel slice has completely dissolved. The solution was mixed with 1 gel/sample volume of isopropanol and transferred on the top of a QIAquick spin column. The column was centrifuged at 17,900 × g in a microcentrifuge to bind DNA to the column and remove the soluble components. The column was washed with 750 µl Buffer PE and centrifuged for 1 min at 17,900 × g. Purified DNA was eluted to clean 1.5-ml Eppendorf tubes by the addition of 30 µl nuclease-free water (New England BioLabs, UK) on the top of the column and centrifuging at 17,900 × g.

Purified DNA was quantified using NanoDrop or Tecan's NanoQuant plate (see details in section 2.5.5).

2.5.4 Purification of genomic DNA from cyanobacteria

Strains of *Anabaena* sp. PCC 7120 were grown in 20–80 ml liquid BG-11 medium (in 100 or 250-ml Erlenmeyer flasks, depending on the volume) supplemented with the appropriate antibiotics until late

exponential phase (OD_{730} = 0.8). Cultures were pelleted in 50-ml centrifuge tubes at 4,700 × g for 15 min at 4 °C. The pellet was resuspended in 0.5 ml extraction buffer (20 mM Tris-HCl, pH 8.0 and 5 mM EDTA pH 8.0) and thoroughly mixed with 10 µl of 0.1 M dithiothreitol (DTT, freshly prepared). For cell lysis, the suspension was transferred to a microcentrifuge tube containing 0.5 ml acid-washed glass beads of the size 425-600 µm (G8772, Sigma-Aldrich). The mixture was vortexed six times for 30 s with 2-minute breaks on ice, and the lysate transferred to a clean microcentrifuge tube. Beads were washed with 0.5 ml extraction buffer and the supernatant combined with the cell lysate. A phenol–chloroform–isoamyl alcohol (PCI) extraction was performed by adding 500 µl PCI mix (prepared the previous day by combining phenol:chloroform:isoamyl alcohol in 25:24:1 in solvent-clean glassware and stored at 4 °C overnight) as follows: the lysate–PCI mix was well vortexed, centrifuged at 16,000 × g for 5 min and the aqueous (upper) phase transferred to a clean microcentrifuge tube. The PCI extraction was repeated multiple times until a clear interface was formed (complete removal of proteins). Residual phenol was removed from the sample by adding an equal volume of 24:1 chloroform:isoamyl alcohol (CI) and performing extraction as follows: the protein-free aqueous (upper) phase from the PCI step was mixed well with CI by vortexing, centrifuged at 16,000 × g for 5 min and the upper phase transferred to a fresh microcentrifuge tube. Ribonucleic acids were removed by enzymatic degradation from 20 µl 20 mg/ml RNase A (R6513, Sigma-Aldrich, UK) combined with the addition of 2.5 µl 1 M $MgCl_2$ at 37 °C for 30 min. A subsequent PCI extraction (repeated until reaching a clear interface, but at least two times) followed by CI extraction was carried out for the removal of RNase A. DNA was precipitated by adding one-tenth volume of 3 M sodium acetate (pH 4.0) and 2.5-times volume of absolute ethanol. The solution was well mixed and incubated at -20 °C overnight (a white precipitate was formed). The next day the sample was centrifuged at 16,000 × g for 10 min and the supernatant discarded. The resulting pellet was washed in 1 ml ice-cold 70% ethanol, centrifuged at 16,000 × g for 1 min and the supernatant carefully removed. The washed pellet was air-dried for 10 min and resuspended in 50 µl nuclease-free water.

2.5.5 Quantification of DNA concentration

The approximate DNA concentration of sampels was determined either by a NanoDrop Lite Spectrophotometer (ND-LITE; Thermo Scientific Ltd., UK) or using a NanoQuant plate with a Tecan Infinite M200 Pro plate reader (Tecan AG, Switzerland), following the manufacturers' protocols. In both cases 2 µl nuclease-free water as blank was loaded first in the sample well, measured and wiped off using low-lint Kimwipes (Kimtech Science Kimwipes; Fisher Scientific, UK). Samples were loaded in 2 µl and DNA concentration was calculated by the instrument software. Ratios of absorbances at 260

nm vs. 280 nm and 260 nm vs. 230 nm were calculated to determine purity of the DNA sample. Sample purity was accepted for 260/280 ratios between 1.8 and 2.0 and between 2.0 and 2.2 for 260/230.

2.5.6 Site-directed mutagenesis

The Quick Change II site-directed mutagenesis (SDM) protocol (Agilent Technologies, USA) was modified and used with molecular biology grade enzymes from NEB (New England BioLabs Inc., UK). A 50-µl reaction mixture contained 10 µl 5× Phusion HF reaction buffer (NEB), 2 µl DNA plasmid template at concentrations 2.5, 10 and 25 ng/µl, 2.5 µl of each mutagenesis primer (*glnA*.D52S-F and *glnA*.D52S-R) at 10 µM concentration, 1 µl 10 mM dNTP mix (NEB), 0.5 µl Phusion DNA polymerase (NEB) and 31.5 µl nuclease-free water (NEB). The reaction mixture was assembled on ice in PCR tubes and transferred to a thermocycler (Biometra Professional Trio, Biometra, Germany) running the same temperature program described in Table 2.2. Following the reaction in the thermocycler, PCR tubes were placed on ice for 2 min and 1 µl *Dp*nI (NEB) was added to digest away methylated DNA (template background). Tubes were further incubated in the thermocycler at 37 °C for 2 h and the restriction enzyme was heat-inactivated at 80 °C for 2 min. Tubes were placed back on ice and 2 µl of each was transformed into *E. coli* DH5α chemically competent cells by heat-shock transformation.

2.5.7 Overlap extension PCR

Fragments of DNA amplified by standard PCR with compatible overhangs, which contained complex DNA secondary structures or were otherwise too large for Gibson assembly, were joined by overlap extension PCR (Bryksin and Matsumura, 2010). Individual fragments of the insert were amplified using the appropriate primers at an annealing temperature (T_a) of 60 °C (extension PCR). Purified fragments (up to four at a time) were mixed in equimolar quantites and run for 15 cycles in the absence of oligonucleotide primers (adjacent fragments primed each other) using standard PCR settings at T_a = 60 °C (overlap PCR). The reason for this step was to prepare a variety of fused fragments, potentially also including a fully assembled sequence. In the final step the outermost oligonucleotide primers of two terminal fragments were added to the overlap PCR mix and incubated for additional 20 PCR cycles at a T_a of 72 °C (purification PCR) to specifically amplify the fully assembled sequence. Assemblies of the correct size were confirmed on agarose gel and purified.

2.5.8 Restriction cloning

Restriction cloning was performed on plasmid and insert with compatible sticky ends created by double digestion with the appropriate restriction enzymes provided by NEB (New England BioLabs Inc., UK). A general reaction contained the following reagents: 2 µg purified DNA (X µl), 1 µl of each restriction enzyme (NEB CutSmart compatible, for a total of Y µl), 3 µl 10× CutSmart buffer (NEB) and 27-X-Y µl nuclease-free water (NEB) in 30 µl final volume. Each enzyme was added alone as well to separate samples to serve as controls, along with a negative control to which no enzyme was added. The reaction was incubated at 37 °C for 1h, run on 1% agarose gel at 100 V for 30 min; fragments of the right size were excised and purified using a QIAquick Gel Extraction Kit (Qiagen Ltd., Crawley, UK).

The plasmid backbone was dephosphorylated after restriction digestion to prevent self-ligation. A 30-µl reaction mix contained the following reagents: 20 µl purified linear plasmid (restriction digested), 3 µl 10× FastAP buffer (Thermo Fisher Scientific, UK), 1 µl FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, UK) and 6 µl nuclease-free water (NEB). The reaction was assembled on ice and incubated at 37 °C for 10 min in a Biometra Professional Trio thermocycler (Biometra, Jena, Germany) and the enzyme inactivated at 75 °C for 5 min.

The digested parts with compatible sticky ends were ligated together in a 20-µl reaction made up of 1 µ T4 DNA ligase (NEB), 2 µl 10× T4 DNA ligase reaction buffer, X µl digested insert and Y µl digested vector in a 3:1 molar ratio, and 17-X-Y µl nuclease-free water (NEB). The reaction was incubated at room temperature for an hour and 5 µl of the ligation mix was transformed into chemically competent *E. coli* DH5α by heat-shock transformation. Colonies were screened on LB-agar plates supplemented with the appropriate antibiotics.

2.5.9 Gibson assembly

Constructs of DNA were assembled using the one-step isothermal DNA assembly protocol modified from Gibson et al. (2009) and Gibson (2011). In detail, 6 ml of 5× isothermal reaction buffer was prepared by combining 3 ml of 1 M Tris-HCl pH 7.5, 150 µl of 2 M $MgCl_2$, 60 µl of 100 mM dGTP, 60 µl of 100 mM dATP, 60 µl of 100 mM dTTP, 60 µl of 100 mM dCTP, 300 µl of 1 M DTT, 1.5 g PEG-8000 and 300 µl of 100 mM NAD. The buffer was aliquoted and stored at -20 °C. An assembly master mixture was prepared by combining 320 µl 5× isothermal reaction buffer, 0.64 µl of 10 U/µl T5 exonuclease, 20 µl of 2 U/µl Phusion DNA polymerase, 160 µl of 40 U/µl Taq DNA ligase and nuclease-free water up to a final volume of 1.2 ml. Fifteen microliters of this reagent-enzyme mix were aliquoted into 200-µl PCR tubes and stored at -20 °C. Frozen 15 µl assembly mixture aliquots were thawed and then kept

on ice until ready to be used. Five microliters of the DNA to be assembled were added to the master mixture in equimolar amounts for each fragment up to a total of 200 ng DNA. Between 10 and 100 ng of each ≤ 6 kb DNA fragment was added. Larger DNA segments were split to two by PCR. Incubation of the assembly reaction was performed at 50 °C for 60 min. Enzymes were purchased from New England BioLabs, UK.

Usually 20–100 colonies were formed on the next day of which 10 were picked in a test cycle until a clone with the desired genotype was found by colony PCR. Candidates were confirmed by sequencing the purified plasmid DNA (Source BioScience, Nottingham, UK).

2.6 Genetic manipulation techniques

2.6.1 Preparation of chemically competent cells

Commercial strains of *E. coli* (DH5α or HB101) were grown overnight in 5 ml liquid LB media supplemented with antibiotics, where necessary (100 µg/ml carbenicillin in case of HB101). An overnight culture was subcultured in a 50-ml centrifuge tube by inoculating 25 ml liquid LB (supplemented with the appropriate antibiotics, where necessary) with 250 µl culture and grown at 37 °C and 200 rpm until reaching an optical density (OD) of 0.3–0.5 at 600 nm. The tube was buried in ice for 15 min, the contents were centrifuged at 3,000 × g and 4 °C for 5 min and the resulting supernatant was discarded. Cells were resuspended very gently in cold Tfb1 (100 mM RuCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$ and 15% glycerol at pH 5.8 adjusted by acetic acid and filter sterilized), 0.4-times of the original volume (10 ml Tfb1 for a 25-ml cell culture), rested on ice for 15 min, pelleted at 3,000 × g and 4 °C for 5min, and the supernatant discarded. The cell pellet was resuspended very gently in cold TfbII (10 mM MOPS, 10 mM RuCl, 75 mM $CaCl_2$ and 15% glycerol at pH 6.5 adjusted by KOH and filter sterilized), 0.04-times of the original volume (1 ml Tfb2 for a 25-ml cell culture), placed on ice for 15 min, aliquoted into sterile Eppendorf tubes in 50 µl volumes and stored at -80 °C.

2.6.2 Heat-shock transformation of bacteria

Aliquots of chemically competent cells of *E. coli* (DH5α or HB101) were removed from -80 °C, thawed on ice for 5 min and mixed very gently by tapping the tubes. Plasmid DNA was added in 5 µl volume

(up to 100 ng), mixed very gently by tapping the tube and kept on ice for 30 min. Cells were transformed by heat shock at 42 °C for 45 sec in a thermoshaker (PCMT; Grant Instruments, UK), rested on ice for 2 min and recovered by the addition of 500 µl liquid LB media and incubation at 37 °C in a thermoshaker (with shaking) for an hour. Recovered cells in 100 µl were plated on LB-agar supplemented with the appropriate antibiotics. The rest of the cells were spun down, the majority of the supernatant removed, the cells resuspended in the remaining liquid and spread on LB-agar plates containing the appropriate antibiotics. Agar plates were incubated at 37 °C overnight in a small incubator (Labnet Mini Incubator, ST2185; Appleton Woods Ltd., UK).

2.6.3 Triparental conjugation of *Anabaena* sp. PCC 7120

Triparental mating of *Anabaena* sp. PCC 7120 was performed with *E. coli* strains HB101 (carrying pRL623 helper plasmid) and ED8654 (carrying pRL443 conjugative plasmid) (Elhai and Wolk, 1988; Elhai et al., 1997). The helper strain carrying the helper plasmid was first transformed with a mobilizable cargo plasmid bearing the gene of interest. For the purpose of chromosomal integration the suicide plasmid pK19mobsacB was used; for plasmid expression a self-replicative pRSF1010-derivative plasmid was transformed to HB101. Both *E. coli* strains were grown overnight in 5 ml LB liquid medium, supplemented with the respective antibiotics at 37 °C, 180 rpm, to maintain the plasmids. Strain HB101 (pRL623 and the cargo plasmid) was grown in the presence of 34 µg/ml chloramphenicol, in combination with 20 µg/ml streptomycin and 50 µg/ml spectinomycin or 50 µg/ml kanamycin for the cargo plasmid. Strain ED8654 (pRL443) was grown in 100 µg/ml carbenicillin. Overnight cultures of the above strains were subcultured in 10 ml LB supplemented with the appropriate antibiotics at 37 °C and 180 rpm for 2.5 hours. The subcultures were inoculated with 350 and 250 µl for HB101 and ED8654, respectively. Cells were harvested by centrifugation at 3,000 × g, room temperature for 3 minutes, and washed three-times in antibiotics-free fresh LB medium. The leftover supernatant was removed, cells were resuspended in 60 µl fresh LB medium and combined at room temperature for 2 hours.

The recipient strains of *Anabaena* sp. PCC 7120 were grown up to 3–5 days in 20 ml BG-11 liquid medium on a rotary shaker in 100-ml Erlenmeyer flask at 30 °C, 1% enriched CO_2 with continuous illumination from cool-white LED lamps at 40 $\mu Em^{-2}s^{-1}$. Cyanobacterial culture was concentrated by centrifugation at 3,000 × g and room temperature for 10 minutes, washed in 20 ml and resuspended in 500 µl fresh BG-11 medium. Chlorophyll *a* content was determined by measuring the absorbance of 1 ml methanolic extract at 665 nm in a 1-cm cuvette by mixing 0.9 ml methanol with 100 µl cell

concentrate (in a Tecan Infinite M200 Pro multimode reader), vortexing the mixture for 1 minute and centrifuging at full speed (17,000 × g in a benchtop centrifuge) for a minute. The approximate chlorophyll *a* concentration was calculated using the formula below:

$$\text{Chl } a \text{ [µg/ml]} = \text{[absorbance at 665 nm]} \cdot 13.43 \cdot 10$$

A 9-cm Millipore nitrocellulose membrane of 0.45-µm pore size (HATF09025; Merck-Millipore, Watford, UK) was placed on a BG-11-agar plate containing 5% LB medium, slowly soaked into the agar and adjusted until no bubbles were present under the filter. A volume of the cyanobacterial concentrate corresponding to 10 µg Chl *a* was gently mixed to the *E. coli* suspension containing both bacterial strains, and the triparental mix was carefully spread over the nitrocellulose membrane using a pipette tip. The agar plate was placed under low light (20 $\mu Em^{-2}s^{-1}$) for 3 hours and then brought to normal light conditions (40 $\mu Em^{-2}s^{-1}$) for 21 hours, when the filter membrane was transferred to a fresh BG-11 agar plate (no LB). The next day the filter was transferred to selective agar medium (BG-11 supplemented with selective antibiotics for the cargo plasmid) and the transfer was repeated every second day until the appearance of transformant colonies (after about 2 weeks).

2.6.4 Isolation of cyanobacterial strains

Individual colonies arising on conjugation plates were picked from the filter membrane with toothpicks onto selective agar plates and maintained in line cultures, eight clones per plate (see panels A–D in Figure 4.7 as an example). Once a week single colonies were picked from each section of a plate and streaked onto fresh selective agar plates. Colony PCR was performed from every second re-streaking to monitor the status of segregation. Re-streaking was continued until a fully segregated isolate was detected by colony PCR (section 2.6.5). To improve the efficiency of segregation colonies growing on plates were picked, cultivated in liquid BG-11 to an OD_{730} value between 0.5 and 0.8 and filament length was decreased by cavitation. Fragmented cultures were washed in fresh BG-11 and spread on BG-11 plates for re-isolation. Genetic homogeneity and completion of segregation was confirmed by sequencing of high-quality genomic DNA prepared from candidate isolates.

2.6.5 Colony PCR of cyanobacteria

Individual isolates of transformed cyanobacterial strains were collected with a 1-mm inoculation loop and suspended into 20 µl nuclease-free water in PCR tubes. PCR tubes were frozen at -80 °C for 10

min and then thawed in a thermocycler (Biometra Professional Trio, Biometra, Germany) at 60 °C for 2 min. The freeze-thaw cycle was repeated five times, and PCR tubes were centrifuged for 1 min in a benchtop mini centrifuge (Sprout, Heathrow Scientific, USA). Two microliters of the supernatant was brought to a PCR as template, using primers specific to the test. Products were visualized on 1% agarose gel running at 100 V for 30–40 min (depending on the size of the product) and photographed in a gel imager (GelDoc-It, UVP, Cambridge, UK).

2.6.6 Filament fragmentation by sonication

Samples were treated with sonication to break up long filaments and facilitate segregation of genotypes. Candidate colonies were inoculated into 1 ml liquid B-11 medium supplemented with the appropriate antibiotics, and grown under standard conditions (40 µE m^{-2} s^{-1}, 30 °C in 1% enriched CO_2) to an OD_{730} between 0.5 and 0.8. Samples were fragmented via cavitation in a sonication bath (F5 Minor, Decon Ultrasonics Ltd., UK). Undiluted cultures (500 µl in sterile 2 ml glass vials) were washed in fresh BG-11 medium before cavitation and treated for 4×2, 2×4, 4×4 and 2×8 min at room temperature. Status of filament fragmentation was monitored by microscope (Zeiss Axioskop 2 Plus, Carl Zeiss Ltd., Cambridge, UK). A treatment for 2×8 min was found to be the most efficient to introduce filament breakage at every 3–6 cells, and this setting was used to routinely sonicate filaments. Fragmented samples were washed again in fresh BG-11 and the supernatant was transferred on BG-11 agar plates supplemented with the appropriate antibiotics. Appearing colonies were picked, streaked onto fresh plates and thereby returned to the segregation cycle.

2.7 Analytical methods

2.7.1 Microscopic analysis of cyanobacterial cultures

Liquid cultures of cyanobacteria were sampled for 200 µl homogenous samples into 1.5-ml Eppendorf microcentrifuge tubes under sterile conditions, and the culture was returned to incubation. The sample was serial diluted 3×, 10× and 50×, and 10–20 µl of each dilution step was transferred on a glass microscope slide. Sample droplets were covered with a glass cover slip and the edges of the slip were painted with transparent nail polish to avoid quick drying of the culture. A fluorescent microscope (Zeiss Axioskop 2 Plus, Carl Zeiss Ltd., Cambridge, UK) was used in bright field mode to study samples. Hundred-times magnification was used for overviewing the sample and 400× (with a drop of immersion oil between the objective and the cover slip) for analysis and imaging. Colour

photos were auto corrected for contrast and tone using Adobe Photoshop CS6 (Adobe Systems Inc., USA).

2.7.2 Peptide preparation for LC-MS/MS from cyanobacteria

Cells of well-grown 20-ml cyanobacterial cultures (at an approximate OD_{730} = 1) were harvested in 50-ml centrifuge tubes by centrifugation at 4,700 × g and 4 °C for 10 min. The supernatant was discarded and cells were resuspended in 500 µl extraction buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0 and 2 mM DTT). Samples were kept on ice throughout the procedure. Lysis tubes were prepared as follows: 500 µl 425–600 µm acid washed glass beads (G8772; Sigma-Aldrich, Dorset, UK) were measured into 2-ml microcentrifuge tubes and washed 3 times in extraction buffer to equilibrate them prior to lysis. The extraction buffer was removed from the beads and the resuspended cells were transferred to the lysis tube. The cell sample tube was rinsed with an additional 500 µl extraction buffer and this additional 500 µl was combined with the rest of the cells already in the lysis tube. Tube holder racks of TissueLyzer (TissueLyzer II; Qiagen Ltd., Crawley, UK) were pre-cooled at -20 °C for 5 min and lysis tubes were inserted into the wells. Cells were disrupted by two 5-min programs in the TissueLyzer at 30 Hz. The lysate was recovered by transferring the liquid to a fresh microcentrifuge tube, washing the beads in 500 µl fresh extraction buffer and combining the wash with the rest of the lysate (crude lysate). Samples were centrifuged at 12,000 × g for 10 min in a pre-cooled centrifuge at 4 °C to remove unbroken cells from the lysate, and the supernatant was transferred to fresh microcentrifuge tubes without disturbing the pellet (cell-free extract). Total protein concentration was determined from the cell-free extract by BCA protein assay (Pierce BCA Protein Assay Kit; Thermo Fisher Scientific, UK) against a BSA (bovine serum albumin) standard curve, following the kits protocol. In short, reagent A and B from the kit were mixed in a 50:1 ratio, and 200 µl of this working reagent was added to 25 µl of each sample and calibration standard in different wells of a microtitre plate. The plate was covered with lid and incubated for 30 min in a plate incubator (PHMP; Grant Instruments, UK) at 37 °C with vibrational shaking at 800 rpm. Absorbance of samples was measured at room temperature at 562 nm in a microplate reader (Infinite M200 Pro; Tecan AG, Switzerland) and protein concentrations were calculated in gram per litre using the two parameters of the calibration curve. The above cell-free protein extracts were diluted by 0.1 M ammonium bicarbonate to get 100 µg protein in 300 µl, reduced by adding 3 µl 1 M DTT and 15 µl 1 M ammonium bicarbonate and incubating for an hour at 56 °C in a thermoshaker (PHMT; Grant Insturments, UK), and alkylated by the addition of 30 µl 0.5 M iodoacetamide (prepared fresh in 0.1 M ammonium bicarbonate) at 37 °C, 500 rpm in the dark for 30 min. The alkylated sample was digested by proteomics-grade trypsin

(V5117; Promega, UK) by adding 5 µl of 0.4 µg/µl trypsin to 100 µg protein and incubating at 37 °C overnight. Tryptic digestion was stopped by quenching the reaction with 98% formic acid at a final acid concentration of 1% for 30 min at 37 °C, 500 rpm in a thermoshaker. Acidified samples were centrifuged at 17,000 × g in a benchtop centrifuge to pellet water immiscible degradation products; the supernatant was transferred to a fresh microcentrifuge tube and stored at -80 °C until analysis by mass spectrometry.

2.7.3 Triple quadrupole mass spectrometry analysis of proteotypic tryptic peptides

Analytical measurements of *Anabaena* sp. PCC 7120 proteome and specific peptides were performed by Mr Mark H Bennett (Mass Spectrometry Service Manager; Department of Life Sciences, Imperial College London, UK). Pre-processing of measurement data was also done by Mr Mark H Bennett; sample preparation and data evaluation were carried out by the author of this thesis. The LC-MS/MS instrument comprised of an Agilent 1100 LC system and an ABSciex 6500 Qtrap MS. Chromatography was performed on a Phenomenex Luna C18(2) column (100 mm × 2 mm × 3 µm) at a temperature of 50 °C utilising a gradient solvent system of A (94.9% H_2O, 5% CH_3CN and 0.1% formic acid), and B (94.9% CH_3CN, 5% H_2O and 0.1% formic acid). A gradient from 0% B to 35% B over 30 min at a flow rate of 250ml/min was used. The column was then washed with 100% B for 3 min and then re-equilibrated with 100% A for 6 min. Typically 40 µl injections were used for the analysis. The MS was configured with an Ion Drive Turbo V source; Gas 1 and 2 were set to 40 and 60 respectively, the source temperature to 500 °C and the ion spray voltage to 5,500 V. The MS, configured with high mass enabled, was used in "Trap" mode to acquire Enhanced Product Ion (EPI) scans for peptide sequencing and "Triple Quadrupole" mode for Multiple Reaction Monitoring (MRM). Data acquisition and analysis was performed with SCIEX software Analyst 1.6.1 and MultiQuant 3.0. Signature peptides for GlnA (Alr2328), GifA (Asl2329), NtcA (Alr4392), RpoC1 (Alr1595), RpoC2 (Alr1596) and AtpB (All5039) proteins were determined from trial MRM-MS runs. RpoC1, RpoC2 and AtpB were measured for reference. However, expression level of AtpB was found highly dependent on growth, and therefore, only RpoC1 and RpoC2 were used as internal standards for protein normalisation. The typical work flow to select the best signature peptides was to analyse samples using transitions generated by an *in silico* analysis with the open-source Skyline Targeted Mass Spec Environment (MacLean et al., 2010; Abbatiello et al., 2013). The identity of candidate peptides was then confirmed by Enhanced Product Ion scans. Background proteome of *Anabaena* sp. PCC 7120 (http://genome.kazusa.or.jp/cyanobase) was used to check for uniqueness of candidate peptides. Typically 3–5 transitions per peptide were used. In the final method 2–4 peptides per proteins were used for identification and quantification of

the corresponding protein. Signature peptides for each protein are listed in Table 2.3. Protein quantification was performed based on relative peak intensities of the analysed protein and normalized to the relative peak intensities of the RpoC1 and RpoC2 native standards.

Table 2.3. Signature peptides of *Anabaena* sp. PCC 7120 proteins analysed in this work.

Protein	t_R (s)[a]	Peptide sequence	DP (V)[b]	CE (eV)[c]
All5039	16.3	VVDLLTPYR	70.4	28.2
All5039	18.1	FVQAGSEVSALLGR	83.4	34.7
All5039	25.1	FLSQPFFVAEVFTGSPGK	102.5	44.1
Alr1595	15.6	FATSDLNDLYR	79.1	32.5
Alr1595	19.9	LQEILAPEIIVR	82	34
Alr1595	22.1	LGIQAFEPILVEGR	87.4	36.6
Alr1595	12.9	VTTNEDGSR	66.8	26.5
Alr1596	17.7	SLLEAAEEEIR	77.1	31.5
Alr1596	16	LVDVSQDVIIR	77	31.5
Alr1596	9	FAGVEVQK	63.1	24.7
Alr1596	14.1	SVEGVELLR	67.7	26.9
Alr1596	21.7	GDNLVLLVFER	77.6	31.8
Alr1596	14	TGDIIQGLPR	70.1	28.1
Alr1596	12.6	IEELLEAR	66.6	26.4
Alr1596	13.2	VVYGDGDEAIAIK	80.4	33.2
Alr1596	17.8	AQYTPVLLGITK	78.7	32.3
Asl2329	19.3	SAQELGLPAEELSHYWNPTQGK	120.6	60
Alr2328	16.5	IELIDLK	61.9	24.1
Alr2328	8.1	TGEWYNR	64.9	25.5
Alr2328	10.8	LGVPIEK	58.7	22.5
Alr2328	7.7	IPLSGTNPK	64.9	25.6
Alr2328	16.6	NIYELSPEELAK	82.4	34.2

Protein	t_R (s)[a]	Peptide sequence	DP (V)[b]	CE (eV)[c]
Alr4392	11.7	ALANVFR	60	23.1
Alr4392	16.2	TIFFPGDPAER	76.7	31.4
Alr4392	20.1	VYEAGEEITVALLR	88.1	37
Alr4392	10.4	LSHQAIAEAIGSTR	66.5	24

[a] Retention time of the corresponding peptide in seconds.

[b] Declustering potential applied to prevent clustering of ions, expressed in volts.

[c] Collision energy applied for fragmentation of peptides, expressed in electron volts.

2.7.4 Glutamine synthetase (GS) bioactivity assay

GS activity of exponentially grown *Anabaena* sp. PCC 7120 cultures was tested combining similar protocols by (Orr et al., 1981; Bressler and Ahmed, 1984; Mérida et al., 1991). First a cell-free protein extract was made as follows. Exponentially grown cultures were pelleted by centrifugation and washed in fresh liquid medium (BG-11 or BG-11_0). In a 2-ml microcentrifuge tube 500 ul acid washed 425–600 µm glass beads (G8772; Sigma-Aldrich, Dorset, UK) were loaded (lysis tubes). Beads were washed two times in a buffer made of 50 mM HEPES and 0.2 mM EDTA at pH 7.3. Cell pellets were resuspended in the same buffer and transferred onto the equilibrated beads. Cell pellet tubes were washed in additional 500 µl buffer to transfer most cells. Cell lysis was performed in pre-cooled tube blocks of TissueLyzer (TissueLyzer II, Qiagen Ltd., Crawley, UK) at 30 Hz for 5 min. The crude lysate was decanted into a fresh microcentrifuge tube (collection tube); the beads were washed in additional 500 µl buffer and combined with the rest of the crude lysate. Collection tubes were centrifuged at 4 °C, full speed (21,000 × g) for 10 min to pellet cell debris. The clear, dark blue cell-free lysate was transferred to clean microcentrifuge tubes and analysed immediately.

Prior to further analysis, cell-free protein extracts were assayed for total protein content using the BCA protein assay (Pierce BCA Protein Assay Kit, ThermoFisher). Exactly 50 µg total protein was loaded into a reaction consisting of the following solutions: 200 µl 1M imidazole-HCl buffer pH 7.0 (adjusted with cc. HCl), 100 µl 10mM NH_4Cl, 50 µl 60 mM Na-ATP · H_2O pH 7.0 (prepared on ice, pH adjusted with 1 M KOH), 50 µl 1.67 M $MgCl_2$ · 6 H_2O, 50 µl 1 M sodium glutamate pH 7.0 (adjusted with cc. HCl) 100-X µl deionized water and X µl cell-free extract (for 50 µg total protein per reaction). The resulting 550 µl reaction mix was loaded into clean wells of a clear bottom 24-deepwell polystyrene plate and incubated at 30 °C for 15 min in a plate incubator (PHMP, Grant Instruments, UK). The reaction was quenched by the addition of 1.8 ml 0.8% w/v $FeSO_4$ · 7 H_2O in 0.015 N H_2SO_4 (prepared by dissolving

0.8 g $FeSO_4 \cdot 7\ H_2O$ in 25 ml deionized water, adding 0.042 ml concentrated H_2SO_4 and adjusting the volume to 100 ml). Plates were mixed on a Titramax 1000 plate shaker (544-12200-00, Heidolph Instruments) every 5 minutes and placed on ice prior to quenching. To develop colour (proportional to the phosphate content freed up from ATP by the activity of GS) 150 µl 6.6% w/v ammonium heptamolybdate in 7.5 N H_2SO_4 was added (prepared by dissolving 0.66 g $(NH_4)_6Mo_7O_{24} \cdot 4\ H_2O$ in 2.5 ml deionized water, adding 2.1 ml cc. H_2SO_4 and adjusting the volume to 10 ml with deionized water) and plates were placed back on ice for 5 min. A precipitation formed on the addition of the molybdate solution that disappeared over the 5-min incubation. Clear solutions were homogenized by pipetting the mixture up and down and colour intensity was measured immediately after mixing at 850 nm in a Tecan Infinite M200 Pro multimode reader. To improve comparability of different strain samples in the assay each extracted sample was diluted to three different concentrations (1×, 3× and 10×) and a substrate blank was also included on every plate (50 µl deionized water added instead of sodium glutamate).

2.7.5 Ammonia quantification by the Willis method

The original colorimetric ammonia/ammonium quantification protocol by Willis et al. (1996) was modified to fit into wells of a 96-well microtitre plate. The main reagent contained 1.28 g sodium salicylate, 1.6 g $Na_3PO_4 \times 12\ H_2O$ and 0.02 g sodium nitroprusside as the active reactant, dissolved in 40 ml deionized water. The orange solution was prepared under a chemical fume hood in a 50-ml centrifuge tube. The tube was covered with aluminium foil and stored at 4 °C. Hypochlorite solution (40×) was prepared fresh from 1 ml bleach containing 10–15% active chlorine and 39 ml deionized water. A series of standards were prepared from ammonium chloride at the following concentrations: 0.01, 0.02, 0.04, 0.05, 0.08, 0.10, 0.15, 0.20, 0.50, 0.80, 1.00, 5.00 and 10.00 mM by diluting a 1-M stock solution with deionized water. The 1-M standard stock solution was prepared by measuring 0.535 g NH_4Cl into a 10-ml volumetric flask using an analytical balance, dissolving the crystals in 30 ml deionized water and adjusting the volume to the graduation line with deionized water. Biological samples were filtered through 10-kDa cutoff MWCO filters (516-0227; VWR International Ltd., UK) to remove any protein contamination. The procedure was performed in a 96-well clear bottom polystyrene plate (734-0954, VWR International Ltd., UK). Standards (including deionized water as blank) and samples were loaded carefully at the bottom of the wells in 10 µl by single-channel pipette. Two-hundred microliters of the orange reagent was transferred from a reservoir using a multichannel pipette on top of the standards and samples. Hypochlorite solution (40×) was added immediately in 50 µl to each well using a multichannel pipette and mixed by pipetting the reaction mix up and down

a few times. The plate was covered with lid and incubated at room temperature for 15 min in a plate shaker (PHMP; Grant Instruments, UK) set to 900 rpm. Absorbance of wells was read at 685 nm in a Tecan Infinite M200 Pro multimode reader (Tecan AG, Switzerland). Ammonia concentration in the samples was calculated by parameters of the standard curve (in the linear range).

2.7.6 Ammonia quantification by a commercial kit

Ammonia measurements, where specifically noted, were performed using the Abcam Ammonia Assay Kit (modified Berthelot, ab102509; Abcam Plc., Cambridge, UK), following the manufacturer's protocol. In short, biological samples were first filtered using a 10 kDa spin column (516-0227; VWR International Ltd., UK). A dilution series was prepared from the provided Ammonium Chloride Standard by diluting 10 µl of 100 mM NH_4Cl with 990 µl deionized water resulting in a 1-mM standard stock solution. A series of wells in a 96-well clear bottom polystyrene microtitre plate (734-0954, VWR International Ltd., UK) were filled with 0, 2, 4, 6, 8, and 10 µl of the 1-mM NH_4Cl standard and the volume adjusted to 100 µl in each well. Hundred microliters of samples were added to different wells and 80 µl Reagent 1 was dispensed on top of all wells (including samples and standards) using a multichannel pipette. Forty microliters of Reagent 2 was added to initiate the reaction and the plate was incubated at 37 °C for 30 min a plate incubator (PHMP; Grant Instruments, UK) shaking at 900 rpm. Absorbance of wells was read at 670 nm in a Tecan Infinite M200 Pro multimode reader (Tecan AG, Switzerland). Ammonia concentration in the samples was calculated by parameters of the standard curve (in the linear range).

2.7.7 Detection of siderophore activity by chrome azurol S assay

A protocol described by Nicolaisen et al. (2010) was used to specifically detect the presence and activity of siderophores. In short, the blue dye chrome azurol S (10511831, Fisher Scientific, UK) was added to BG-11 agar plates as follows: 60.5 mg chrome azurol S (CAS) was dissolved in 50 ml deionized water and mixed with 10 ml 1 mM $FeCl_3$ solution (1 mM $FeCl_3 \cdot 6\ H_2O$ in 10 mM HCl). The mixture was continuously stirred while 40 ml hexadecyltrimethylammonium (HDTMA) solution (72.9 mg HDTMA dissolved in 40 ml water) was added, resulting in 100 ml working CAS reagent. This working reagent was added to liquid BG-11 media in a 1:10 ratio (CAS:BG-11) and solidified by 1% Bacto agar.

Fully grown cyanobacterial cultures were streaked on the CAS-BG-11 agar plates as round patches and incubated under standard growth conditions (40 μE m^{-2} s^{-1}, 30 °C in 1% enriched CO_2). Halo formation was evaluated after two weeks.

2.8 Computational methods

2.8.1 Metabolic reconstruction

The metabolic network of *Anabaena* sp. PCC 7120 was reconstructed based on the protocol by Thiele and Palsson (2010) for high-quality genome-scale models. Some assumptions and simplifications were necessary to create a working model; these are collected in Appendix B and are referenced at their first mention in the text.

The gene annotation for *Anabaena* sp. PCC 7120 was downloaded from several databases and the information contained therein was merged (Kaneko et al., 2001; Peterson et al., 2001; Kanehisa et al., 2004; Nakao et al., 2010; Tatusova et al., 2014). The metabolic function of proteins inferred from genomic data was collected from biochemical repositories (Kanehisa et al., 2004; Magrane and Consortium, 2011; Caspi et al., 2012) and primary literature (Appendix D). A systematic, automated algorithm predicting novel gene-protein-reaction (GPR) associations for cyanobacteria was also considered (Krishnakumar et al., 2013) but not used here due to the large number of contradictions to the other sources and experimental data. Instead, GPR associations were manually constructed based on literature as a logical expression to handle isozymes and multi-domain proteins. A GPR string was generated for every reaction in the model. The string contained the identifiers of all genes associated with that particular reaction. Genes encoding for proteins of the same multi-domain enzyme were connected with “AND”. Genes encoding for isozymes (different proteins catalysing the same reaction) were connected with an “OR” logical operator. The GPR string was used to determine gene essentiality.

All the information gathered above was mapped onto general metabolic pathways drawn for the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004) and compared with the data therein. The most recent metabolic reconstruction of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 (Knoop et al., 2013) was also mapped for comparison and to serve as a template during the identification of critical metabolic functions and gaps. The majority of these gaps were resolved either by finding genes in primary literature (Appendix D) or by identifying novel gene/protein candidates based on sequence homology (light and dark pink bands on Figure 2.1,

respectively). Homology-based searches were performed using the BLASTX engine (Altschul et al., 1990) on the NCBI *Anabaena* sp. PCC 7120 proteome against gene sequences of reviewed protein entries from the UniProt Knowledgebase (Magrane and Consortium, 2011) to identify best hits. The best hits were verified by BLASTX search against NCBI RefSeq protein database filtered for cyanobacterial entries (Appendix E). A few gaps, otherwise unresolved, were resolved by artificially adding the corresponding metabolic reaction to the model to allow the biosynthesis of key metabolites (e.g. L-methionine and L-asparagine, light green bands on Figure 2.1) or complete missing steps of close-to-compete pathways (dark green bands). Gaps with over two missing reaction steps were treated differently. In most cases such gaps indicate orphan reactions, and appear in sparsely represented pathways due to e.g. misannotation. Lacking any connection to the rest of the network these orphan reactions do not carry metabolic flux in an FBA solution, and were therefore removed from the reconstruction (dark grey bands on Figure 2.1). Dead-end metabolites not consumed by any reaction in the network were resolved by sink reactions artificially added to the model (yellow bands on Figure 2.1, also including biomass equations).

The equation for each metabolic reaction was adopted either from the KEGG and MetaCyc databases (Kanehisa et al., 2004; Caspi et al., 2012) or from the *Synechocystis* sp. PCC 6803 model (Knoop et al., 2013). The general mechanisms of light harvesting, photosynthesis and cell respiration in Nogales et al. (2012) have been substantially improved by Knoop et al. (2013), and adapted here for *Anabaena* sp. PCC 7120. A detailed map of photosynthetic and respiratory electron chains can be found in the supplementary material of Malatinszky et al. (2017).

Reaction thermodynamics, in terms of reaction directionality, were either acquired from the MetaCyC database (Caspi et al., 2012) or calculated using eQuilibrator (Flamholz et al., 2012), an online calculator to estimate reactions' Gibbs free energy change. Also, every reaction was evaluated for mass and charge balance using eQuilibrator, and adjusted if necessary. Those reactions not found in public resources were set bidirectional. Coenzyme dependencies (NAD^+, $NADP^+$ and quinones) in the KEGG database were adapted for *Anabaena* sp. PCC 7120 wherever primary biochemical evidence was available in the literature or left unchanged (light grey band on Figure 2.1).

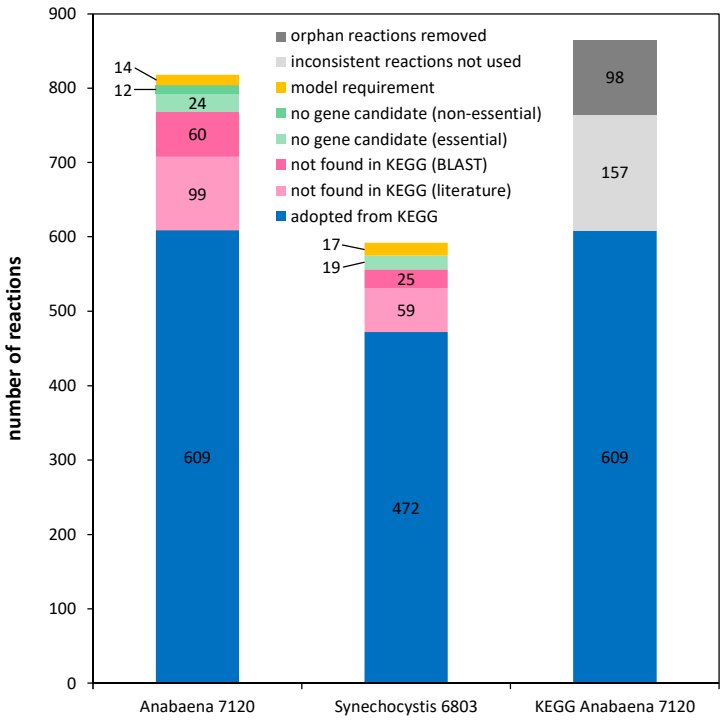


Figure 2.1. Comparison between *Anabaena* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 (Knoop et al., 2013) stoichiometric models and their improvement over the KEGG database (Kanehisa et al., 2004). Metabolic gaps were resolved by either adapting reactions from literature (light pink bands) or by identifying new gene candidates (dark pink bands). A number of gaps could not be associated with any gene in *Anabaena* sp. PCC 7120 (light and dark green bands). Some reactions in the KEGG database were omitted due to inconsistent coenzyme usage or incomplete reaction formula (light grey band), or due to the lack of connection to the rest of the network (orphan reactions, dark grey band).

The stoichiometric model for *Anabaena* sp. PCC 7120 was generated in two formats: as a single-cell model including photosynthesis and carbon-concentrating reactions for non-diazotrophic growth in a vegetative cell, and as a two-cell model setting every reaction to the appropriate super-compartment (either a heterocyst or a vegetative cell) under diazotrophic conditions. Simulations described in section 3.2, unless otherwise noted, were performed on the two-cell model considering six intracellular compartments (cytoplasm, cytoplasmic membrane, thylakoid lumen, thylakoid membrane, carboxysome and periplasmic space) and the external medium (Figure 3.1).

2.8.2 Flux Balance Analyses

All simulations were run using the COBRA toolbox version 2 with Gurobi Optimizer 5.6.0 as solver in the MATLAB R2013b environment (MATLAB, 2011; Schellenberger et al., 2011; Gurobi Optimization, 2013). All FBA optimizations were calculated minimising the taxicab norm using linear programming (Schellenberger et al., 2011). In other words, FBA optimizations were performed to find the minimum of the modulus of the flux vector, thus eliminating internal cycles in the final solution (Smallbone and Simeonidis, 2009).

Unconstrained reactions were allowed to carry fluxes between –1000 and 1000 mmol g DW^{-1} h^{-1} (millimoles per gram dry cell weight per hour), a standard setting for FBA calculations (Becker et al., 2007; Schellenberger et al., 2011). Furthermore, bicarbonate uptake rate was constrained to of the arbitrary 10 mmol g DW^{-1} h^{-1} at the upper bound in phototrophic conditions for both cell types. Simulations with the single-cell model were run on nitrate as the nitrogen source up to 10 mmol g DW^{-1} h^{-1}, unless otherwise noted. The two-cell model was set to solely use molecular nitrogen at an upper limit of 10 mmol g DW^{-1} h^{-1}. The choice of the 10 mmol g DW^{-1} h^{-1} as rate constraints for the above reactions is arbitrary; any (positive) value within the allowed range (between –1000 and 1000 mmol g DW^{-1} h^{-1}) could have been used without essentially altering the outcome of the FBA simulations.

When comparing single-cell growth on combined nitrogen to the two-cell model under diazotrophic conditions, both models were constrained to an equivalent total photon uptake of 10 mmol g DW^{-1} h^{-1}. The optimal distribution of the 10 mmol g DW^{-1} h^{-1} photons among the two super-compartments was 7 and 3 mmol g DW^{-1} h^{-1} in the vegetative cell and the heterocyst, respectively. In all other simulations using the two-cell model photon uptake in both super-compartments was constrained to 10 mmol g DW^{-1} h^{-1}. For the evaluation of exchange metabolites between the heterocyst and the vegetative cell unconstrained bidirectional diffusion was included to the model for each metabolite.

ATP-driven transport reactions were added to the single-cell model for each carbon and nitrogen source except for molecular nitrogen and ammonia, which were exchanged via simple diffusion (the actual transport reactions can be found in 0). Since actual cyanobacterial transporters are unknown for the vast majority of the investigated nutrients, ATP-dependence was postulated for comparative reasons (row 3 in Appendix B). For mixo- and heterotrophic simulations, nitrogen was supplied solely by nitrate (in section 3.2.3.3), whereas nitrogen sources were compared on bicarbonate as carbon source in autotrophic conditions (in section 3.2.3.2). In heterotrophic simulations both photon and bicarbonate uptake rates were set to zero. Under mixotrophic conditions photon and bicarbonate uptake reactions were constrained at 10 mmol g DW^{-1} h^{-1}. Both carbon and nitrogen uptake fluxes

were constrained to carry a maximum of 10 mmol g DW^{-1} h^{-1} carbon and nitrogen source, respectively. For example, upper bound of the bicarbonate transport reaction (a single carbon atom) was set to 10 mmol g DW^{-1} h^{-1}, whereas glutamine uptake was set to 2 mmol g DW^{-1} h^{-1} (five carbon atoms).

The biomass objective function (BOF) was formulated based on BOFs in Nogales et al. (2012) and Knoop et al. (2013) for *Synechocystis* sp. PCC 6803. The fractional composition described in Nogales et al. (2012), as well as the composition of individual fractions were adapted to *Anabaena* sp. PCC 7120 using genetic, proteomic and pigment data. In particular, composition of DNA, RNA and protein fractions were updated using the *Anabaena* sp. PCC 7120 genome sequence and transcriptome data (Kaneko et al., 2001; Flaherty et al., 2011). The composition of the pigment fraction was updated using analysis results in Mochimaru et al. (2008). For the lipid composition of *Anabaena* sp. the data sources found were scarce (Sallal et al., 1990; Rathore et al., 1993), and therefore the composition in Knoop et al. (2013) was used. For the rest of the fractions (cell wall, inorganic ion and metabolite pool) the formulae in Knoop et al. (2013) were applied. Further details to the *Anabaena* sp. PCC 7120 BOF are given in Appendix G and Table 3.2.

Assumptions and simplifications by the model are listed in Appendix B.

2.8.3 Experimental carbon source evaluation

Wild-type *Anabaena* sp. PCC 7120 was grown in triplicates at 30°C, with continuous illumination from cool white LED panels at 60 µE m^{-2} s^{-1}, on a rotary shaker at 200 rpm. Cultures were inoculated into sterile 100-ml Erlenmeyer flasks containing 30 ml BG-11 medium (Rippka et al., 1979) to an approximate OD_{730} of 1, measured in a 1-cm cuvette using a Tecan Infinite M200 Pro plate reader (Tecan AG, Switzerland). Culture health was evaluated by recording absorbance spectrum between 300 and 800 nm. The spectra indicated no differences in pigment composition of the biological replicates. The three replicates were then mixed to minimize biological variation. Cells from the mixed wild-type culture were harvested by centrifugation at 3000 × g, washed in fresh BG-11, and resuspended to the original volume in fresh BG-11.

Each of the $BG\text{-}11_C$ media containing one of the organic carbon sources were prepared from the same BG-11 standard medium described by Rippka et al. (1979), by replacing bicarbonate for the corresponding organic substrate encapsulating equimolar carbon with 5 mM glucose, and filter sterilized. For example, fructose (six carbon atoms per molecule) and glycerol (three carbon atoms per molecule) were set to a final concentration of 5 mM and 10 mM, respectively. The washed and

resuspended cyanobacterial culture was diluted 10-times in 3 ml of each BG-11_C media and dispensed into 3 wells of untreated 6-well flat-bottom plates in a pre-randomized fashion. Plates were covered with a sterile lid, wrapped into Parafilm (291-1212, VWR International Ltd., UK) and incubated under the same conditions as the original shake flask cultures. Cyanobacterial growth and health was observed for up to 8 days by optical density measurements using the same plate reader at 730 nm for culture density, and at 440 nm (absorbance maximum of chlorophyll *a*) for cellular health. The two sets of growth curves acquired at the different wavelengths showed excellent correlation; therefore, only readings at 730 nm were evaluated thereafter. Growth rates were determined for the exponential phase of the growth curves provided in Figure H-1 in Appendix H.

2.8.4 Sequence analysis

DNA and protein sequences of *Anabaena* sp. PCC 7120 for analysis were derived from the curated NCBI RefSeq databases (Tatusova et al., 2014) and the KEGG database (Kanehisa et al., 2004). Sequences of cyanobacterial proteins used for the identification of novel *Anabaena* sp. PCC 7120 enzyme candidates (Appendix E) were derived from the UniProt protein database (Magrane and Consortium, 2011). Biochemical functions for enzyme candidates were confirmed using searches against the Pfam protein domain database (Finn et al., 2016). Sequence homology was evaluated using the BLASTX and BLASTP tools of NCBI (Altschul et al., 1990).

Computer manipulations of DNA sequences, including the design of constructs and the alignment of sequencing results were carried out using the SnapGene software (from GSL Biotech; available at snapgene.com)

3 Modelling the metabolism of *Anabaena* sp. PCC 7120

Parts of this chapter were adapted and reworked from Malatinszky et al. (2017) with permission from the American Society of Plant Biologists under the license number 4165351266912 (27/07/2017, Appendix J). Affected sections are: 3.2.1, 3.2.2, 3.2.3, 3.2.4 and 3.2.5. The stoichiometric model is available as Supplemental File S1 and S2 in (Malatinszky et al., 2017).

3.1 Introduction

Metabolic reconstructions of genome-scale biochemical networks are powerful tools for system-level analysis and prediction of metabolic states of an organism in response to changes in environmental conditions and genetic perturbations (Oberhardt et al., 2009). In addition, stoichiometric models allow the prediction of gene essentiality as well as the calculation of the theoretical yield for a particular metabolite (Joyce and Palsson, 2008; Pásztor et al., 2015). By applying specific FBA tools, stoichiometric models can be used to identify gene/protein targets e.g. for knockout or overexpression, in order to achieve a certain metabolic engineering goal (Burgard et al., 2003; Patil et al., 2005; Tervo and Reed, 2014). The availability of detailed genomic information for *Anabaena* sp. PCC 7120 provides the means to construct its genome-scale, constraint-based, metabolic model, which may then enable the systemic understanding of this organism's metabolism. Highly-curated models have been developed for a handful of photosynthetic microorganisms, including *Synechocystis* sp. PCC 6803 (Nogales et al., 2012; Knoop et al., 2013), *Cyanothece* sp. ATCC 51142 (Saha et al., 2012; Vu et al., 2012) and *Chlamydomonas reinhardtii* (Boyle and Morgan, 2009; Kliphuis et al., 2012), but

not *Anabaena* sp. PCC 7120. Metabolic modelling in oxygenic photoautotrophs may be quite challenging due to the complexity of photosynthetic and respiratory electron transport chains, and the potential effects of two distinct photosystems upon the generation and fate of the reductants and the energy that drives the remainder of metabolism. The inclusion of oxygenic nitrogen fixation further adds to this complexity, by requiring either a precise circadian rhythm placed on the model or a separate nitrogen-fixing compartment with restricted intercellular metabolite exchange.

The two major computational challenges in metabolic network reconstruction are the identification of missing reactions in a metabolic network (gap-filling), and the association of genes with network reactions. Gap-filling is commonly performed based on a pre-defined metabolic capability that the network is expected to be able to fulfil. For example, the ability to generate essential biomass components under various genetic and environmental conditions, or the synthesis of specific secondary metabolites identified by metabolomics or other analytical methods (Vitkin and Shlomi, 2012). The association of genes with biochemical reactions in the network (gene-protein-reaction association, GPR) can be done based on sequence similarity to known genes and proteins or through the experimental determination of enzyme activity – an extremely laborious process in the case of large networks.

In the metabolic reconstruction of *Anabaena* sp. PCC 7120, gap-filling and GPR associations were carried out manually using related genetic and biochemical information in public databases such as the KEGG, MetaCyc, NCBI RefSeq and CyanoBase repositories (Kanehisa et al., 2004; Nakao et al., 2010; Caspi et al., 2012; Tatusova et al., 2014) and primary literature, as detailed in the following sections. The metabolic reconstruction enabled exhaustive computational analysis of possible exchange metabolites and the ranking of these exchange commodities according to optimal growth of the filament. To benchmark the reconstruction, growth states of *Anabaena* sp. PCC 7120 were compared under both photoautotrophic and mixotrophic conditions, either consuming a combined nitrogen source or growing diazotrophically. The stoichiometric model is represented using the systems biology markup language (Hucka et al., 2003) and can be analysed by a constraint-based optimisation approach (Price et al., 2003; Steuer et al., 2012). To date, this reconstruction is the first extensively curated, genome-scale model for *Anabaena* sp. PCC 7120. It is also the first complete reconstruction for heterocyst metabolism and among the first attempts to simulate a simple multicellular organism at genome-scale.

3.2 Results

3.2.1 Reconstruction of the metabolic network

The genome-scale metabolic network of *Anabaena* sp. PCC 7120 was reconstructed using the protocol by Thiele and Palsson (2010), with details provided in section 2.8.1. Genome annotations from Kaneko et al. (2001) and Peterson et al. (2001) were merged with available biochemistry data in reaction databases (Kanehisa et al., 2004; Caspi et al., 2012) to establish gene-protein-reaction (GPR) associations. Biochemical reactions were revised for elemental and charge balance using eQuilibrator (Flamholz et al., 2012), and sorted into six intracellular compartments (cytosol, thylakoid lumen, carboxysome, cytoplasmic membrane, thylakoid membrane and periplasmic space) in order to simulate the growth of vegetative cells on a combined nitrogen source (single-cell model).

The single-cell reconstruction contains a total of 777 metabolites interconnected via 804 enzymatic and 14 spontaneous reactions, as well as 79 transport reactions (0) between intracellular compartments or the external space (Figure 3.1). Ninety-nine metabolic reactions have not previously been reported in databases for *Anabaena* sp. PCC 7120 (Kanehisa et al., 2004; Caspi et al., 2012), but were acquired from the primary literature for this organism. In addition, 56 reactions were associated with candidate genes here for the first time, based on sequence similarity to homologous genes in related organisms. However, no candidate genes could be identified for a total of 36 reactions of which 24 are essential for growth (see section 2.8.1 for more details).

The reconstruction was examined for orphan reactions, that is, reactions that have been annotated for *Anabaena* sp. PCC 7120 in databases, but are disconnected from the rest of the network. Such reactions (or reaction subsets) indicate misannotation (overannotation) or scarcely known pathways. Orphan reactions involve dead-end metabolites, and are therefore not part of any FBA solution. In order to decrease the size of the network, a total of 98 orphan reactions were removed from the model (Appendix A). Such reactions were identified as being separated by more than two enzymatic steps from the closest element within the reference pathway in the KEGG database (Kanehisa et al., 2004). Gaps spanning over two or less reaction steps were resolved by finding gene associations to the missing steps, wherever possible. The gap-filling approach is further detailed in section 2.8.1.

Biosynthesis routes of the major biomass building blocks in current databases (Kanehisa et al., 2004) were revised and updated from primary literature if necessary. For example, metabolic pathways for the synthesis of carotenoids (Albrecht et al., 1996; Takaichi et al., 2005; Takaichi and Mochimaru, 2007; Mochimaru et al., 2008; Graham and Bryant, 2009), phycobilin, thiamine and molybdopterin

(Schluchter and Glazer, 1997; Gutzke et al., 2001; Ruiz et al., 2010; Biswas, 2011) were updated in the current model, as well as was sucrose metabolism (Cumino et al., 2007; Marcozzi et al., 2009; Du et al., 2013). In addition, the reconstruction includes a proposed pathway for the iron (III)-siderophore schizokinen based on the biosynthesis route of a similar compound, rhizobactin, from *Shinorhizobium meliloti* 1021 (Lynch et al., 2001; Nicolaisen et al., 2008; Malatinszky and Jones, unpublished). In order to enable the visualisation and analysis of predicted fluxes, a visual representation of the network was prepared in Cytoscape (Shannon et al., 2003), including tools in Matlab (MATLAB, 2011) for the mapping of flux values (Supplemental File S6 and S7 in Malatinszky et al. (2017)).

3.2.2 Filament representation and the biomass model

In order to reflect the multicellular structure of the *Anabaena* sp. PCC 7120 filament under diazotrophic conditions, the single-cell model was extended to contain two super-compartments. This so-called two-cell model consists of a vegetative cell (VCSC) and a heterocyst super-compartment (HCSC), and represents the multicellular filament as a "symbiotic association" of the two cell types. The two-cell model contains a total of 1797 reactions including exchange between the two super-compartments (see Figure 3.1 and Table 3.1 for differences of VCSC and HCSC). Transport reactions across compartments and exchange between super-compartments were assumed bidirectional, independent of ATP and unconstrained; in contrast, transport reactions to the external space were defined as ATP-driven, unless evidence for a different driving mechanism could be found in the literature (see rows 1, 3 and 8 in Appendix B for model assumptions and 0 for a list of the actual transport reactions).

The vast majority of the reactions can be found in both the VCSC and the HCSC, although there are characteristic differences between the two super-compartments. Most importantly, only the vegetative cell is able to perform oxygenic photosynthesis via linear photophosphorylation, whereas only the heterocyst is capable of performing nitrogen fixation, using ATP from cyclic photophosphorylation at PSI. In addition, PSI in the heterocyst may also act as an O_2 reductant via the Mehler reaction, and thus protect nitrogenase from oxygen (Milligan et al., 2007).

Table 3.1. Major differences between the two cell types (super-compartments) in the *Anabaena* sp. PCC 7120 model

vegetative cell	heterocyst
photosynthesis (PSI and PSII)[a]	cyclic photophosphorylation (PSI only)[a] (Wolk et al., 2004)
carboxysomes and RuBisCO[b]	neither carboxysomes nor an active RuBisCO[b] (Madan and Nierzwicki-Bauer, 1993; Valladares et al., 2007)
no nitrogen fixation	nitrogen fixation (nitrogenase)
GS-GOGAT cycle[c]	Fd-GOGAT missing[c] (Martin-Figueroa et al., 2000)
FNR at PSI produces NADPH[d]	FNR produces red. FdxH for nitrogenase[d] (Razquin et al., 1996)
cox1 cytochrome c oxidase	contains *cox2* and *cox3* only (Valladares et al., 2003)

[a] PSI: photosystem I; PSII: photosystem II

[b] RuBisCO: ribulose-1,5-bisphosphate carboxylase/oxygenase

[c] GS: glutamine synthetase; GOGAT: glutamine-oxoglutarate aminotransferase; Fd-: Ferredoxin-dependent

[d] FNR: ferredoxin-$NADP^+$ reductase; FdxH: heterocyst-specific ferredoxin

Reactions responsible for oxygen evolution in PSII were deleted from the HCSC. Similarly, the inactive RuBisCo-dependent carbon fixation was removed from the HCSC, although other carbon fixation mechanisms may still be active (see below). In addition, nitrogen metabolism in the HCSC lacks the expression of the GOGAT enzyme, but it may have an active nitrogenase in place. The physiological differences between the two super-compartments are listed in Table 3.1; Appendix C lists the resulting differences in terms of metabolic reactions.

Figure 3.1 summarizes the main fluxes concerning carbon and nitrogen metabolism in the two-cell model predicted under diazotrophic growth conditions. The VCSC fixed carbon via the Calvin cycle driven by photosynthesis, and produced an excess of sucrose from glyceraldehyde 3-phosphate and an excess of glutamate synthesized by the GOGAT enzyme. The primary source of glutamate was internally recycled 2-oxoglutarate and glutamine of heterocyst origin. The excess sucrose and glutamate were exchanged for glutamine and 2-oxoglutarate from the heterocyst. Glutamine was derived from glutamate by incorporating ammonia from heterocystous nitrogen fixation to glutamate from the VCSC. Energy (ATP) and electron (reduced ferredoxin) requirements of the nitrogenase reaction were mainly covered by cyclic photophosphorylation at PSI. The rest of the energy was provided by degrading sucrose from the VCSC, and spending its carbon content on cellular maintenance (Figure 3.1).

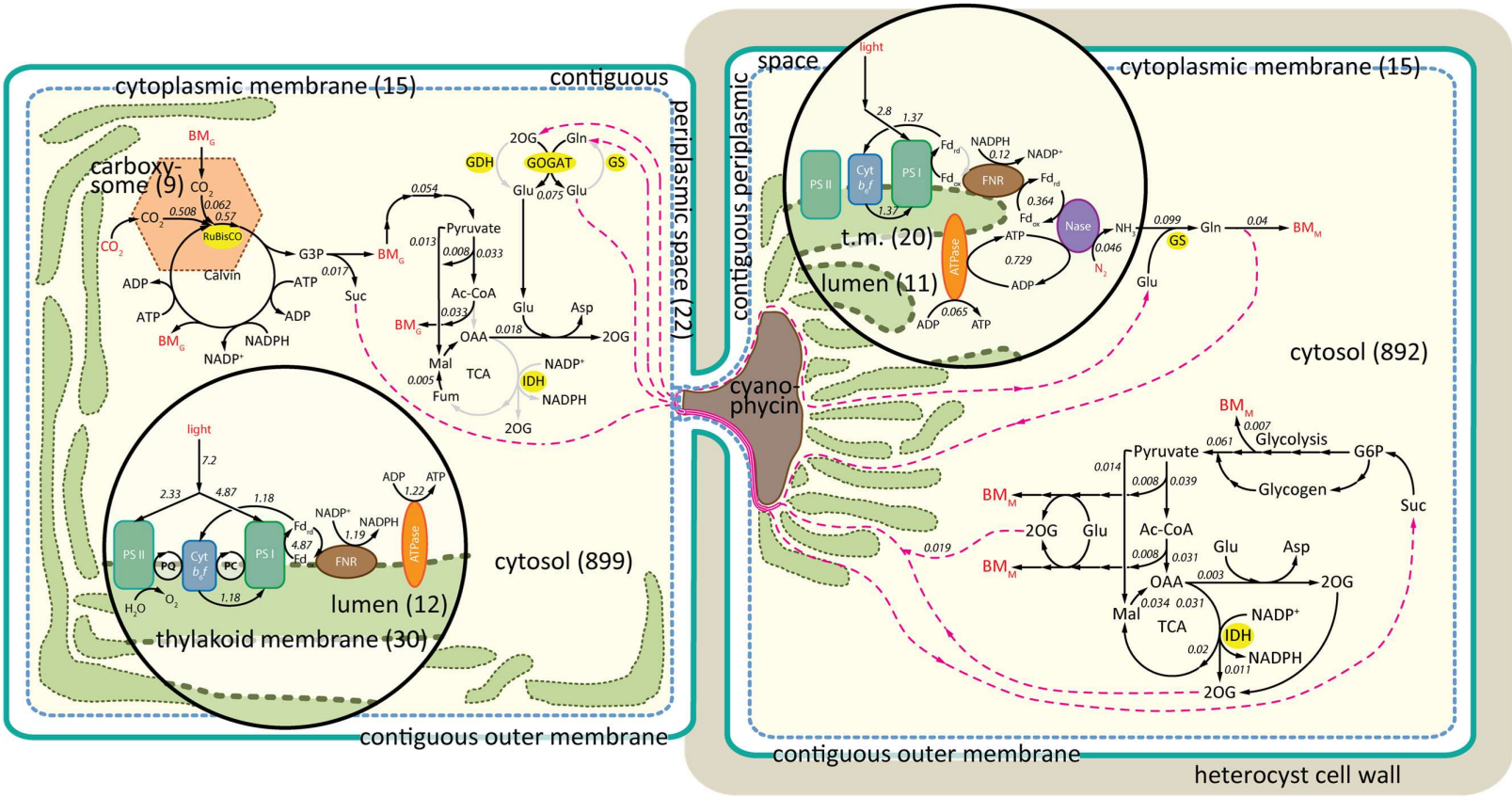


Figure 3.1. Compartments considered in the two-cell model. The filamentous structure of *Anabaena* sp. PCC 7120 under diazotrophic conditions is represented by two super-compartments (a vegetative cell and a heterocyst) sharing certain metabolites via exchange reactions (magenta dashed line). Black arrows and numbers in italic indicate main fluxes of diazotrophic growth. Super-compartments are divided into (sub)-compartments: both cells contain a cytosol (light yellow body), a thylakoid lumen (green bands), a thylakoid membrane (dark green dashes) and a cytoplasmic membrane (blue dashes) and share a contiguous periplasmic space (white space delimited by the cytoplasmic and outer membranes). The vegetative cell also carries carboxysomes (orange hexagon). The brown body in the heterocyst is the nitrogen storage cyanophycin, not a separate compartment on its own. Numbers in parentheses indicate the number of reactions associated with that compartment (including transport). Cell types and compartments are not to scale.

The terminally differentiated heterocyst does not grow or undergo cell division; therefore, the objective function of the two-cell model is defined as the biomass equation of the vegetative cell (row 5 in Appendix B). To account for macromolecular turnover in the HCSC (row 6 in Appendix B), the biomass reaction in this super-compartment (essentially the same reaction as in the VCSC) was constrained to a lower bound equal to 10% of the maximum biomass production in the VCSC, following a figure suggested by van Bodegom (2007) as the approximation of the maintenance requirement of a microbial cell. Moreover, both super-compartments include an artificial ATP hydrolysis reaction to account for the energy requirement of growth-independent cell maintenance at a fixed flux rate (row 7 in Appendix B). This flux rate is equal to 10% of total ATP consumption at maximum growth rate, similar to that of previous stoichiometric models (Feist et al., 2007; Nogales et al., 2012; Knoop et al., 2013). In the initial two-cell model, the two super-compartments were allowed to exchange four metabolites: sucrose (Schilling and Ehrnsperger, 1985; Cumino et al., 2007; Nürnberg et al., 2015), glutamine (Wolk et al., 1976; Thomas et al., 1977; Picossi et al., 2005), glutamate (Martin-Figueroa et al., 2000) and 2-oxoglutarate (Böhme, 1998) (row 12, Appendix B). Transport reactions for these metabolites were unconstrained, bidirectional and provided direct exchange between super-compartments without the involvement of other compartments or the external space (Table F-II in Appendix F). In addition, any dilution occurring due to the size difference of the two cell types was not taken into account (rows 2 and 9 in Appendix B). At optimal growth, the HCSC was found to supply sufficient fixed nitrogen for the growth of exactly 7.6 new vegetative cells, based on the nitrogen content of the vegetative cell in the biomass equation. This calculation sets the minimum value for the number of growing vegetative cells a single heterocyst can support. However, already existing vegetative cells adjacent to the heterocyst require less nitrogen to remain functional, increasing this ratio to the range that can be observed experimentally (Kumar et al., 2010; Ehira, 2013). At such ratios the predicted growth rate dropped gradually, reaching a point at about 46% of the maximal rate where a single heterocyst sustained exactly twenty vegetative cells. It is worth noting that the growth rate predicted by the model does not predict actual cell number, but biomass accumulation rate. Therefore, in the reconstruction the *Anabaena* sp. PCC 7120 filament can be represented as a single VCSC and a single HCSC, denoted as the two-cell model in the following (Figure 3.1 and row 4 in Appendix B).

There is very little information on the exact composition of *Anabaena* sp. PCC 7120 biomass in the literature. Therefore, the biomass equation constructed for *Synechocystis* sp. PCC 6803 by Nogales et al. (2012) and adjusted by Knoop et al. (2013) was used and adapted to *Anabaena* sp. PCC 7120, complemented by sparsely available analytical data (Table 3.2). Vegetative cells and heterocysts were assumed to share the same fractional composition comprising DNA, RNA, proteins, pigments, lipids,

cell wall, inorganic ions and the metabolic pool (row 10, Appendix B). The fractional composition given in Nogales et al. (2012) was applied; however, some of the fractions were recalculated with available data (Table 3.2 and row 11 in Appendix B). The actual reaction equations describing the fractions and the biomass objective function are collected in Appendix G.

Table 3.2. Biomass composition of *Anabaena* sp. PCC 7120, as represented in the model. Fractional composition was adopted from the *Synechocystis* sp. PCC 6803 model (Nogales et al., 2012; Knoop et al., 2013). The definition and the reaction formula of the different fractions were adapted for *Anabaena* sp. PCC 7120 and updated with specific analysis data, where available.

Fraction	Source of data and reference
Pigments	Analysis of carotenoid composition (Takaichi et al., 2005; Takaichi and Mochimaru, 2007; Mochimaru et al., 2008; Graham and Bryant, 2009)
DNA	Base abundance calculated from genomic sequence (Kaneko et al., 2001)
RNA	Base abundance calculated for annotated genes and weighted by RNAseq abundance (Kaneko et al., 2001; Flaherty et al., 2011)
Proteins	Amino acid abundance calculated and weighted by RNAseq abundance (Kaneko et al., 2001; Flaherty et al., 2011)
Lipids	Adopted from *Synechocystis* sp. PCC 6803 (Nogales et al., 2012; Knoop et al., 2013)
Cell wall	Adopted from *Synechocystis* sp. PCC 6803
Inorganic ions	Adopted from *Synechocystis* sp. PCC 6803
Pool fraction	Adopted from *Synechocystis* sp. PCC 6803[e]

[e] No genes could be identified for the biosynthesis of spermidine, although the compound has been detected in *Anabaena* sp. PCC 7120 (Jantaro et al., 2003; Incharoensakdi et al., 2010).

3.2.3 Characterization of the reconstruction

3.2.3.1 Autotrophic growth

The metabolic model of *Anabaena* sp. PCC 7120 was tested under photoautotrophic conditions, to evaluate the basic behaviour of the reconstruction. Super-compartments (VCSC and HCSC) were allowed to exchange glutamate, glutamine, 2-oxoglutarate and sucrose, while bicarbonate, dinitrogen and photons were set as the only external substrates. The uptake of external substrates was constrained to upper bounds of 10 mmol g dry weight $(DW)^{-1}$ h^{-1} each. Figure 3.2 shows the

relationship between light intensity and bicarbonate uptake. The maximal growth rate was reached at 10 mmol g DW^{-1} h^{-1} photon flux (the upper bound) and at a bicarbonate uptake rate of 0.68 mmol g DW^{-1} h^{-1}.

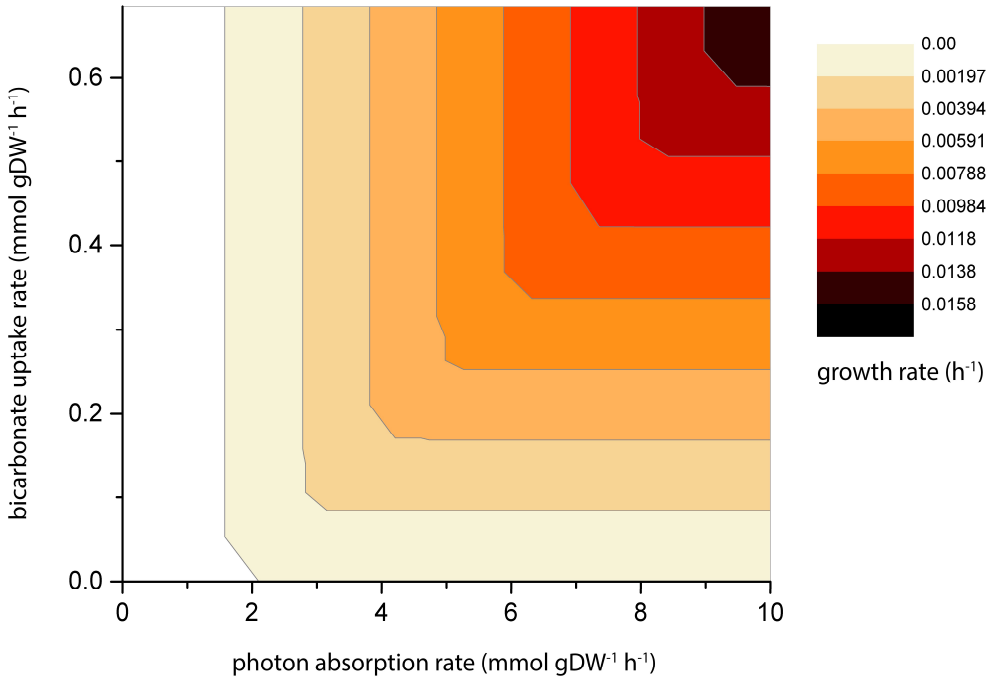


Figure 3.2. Predicted optimal growth rates of *Anabaena* sp. PCC 7120 as a function of light absorption and bicarbonate uptake. Darker colours represent higher growth rates (see legend to the right). Photon requirement of cell maintenance is represented by a white area on the left side of the contour plot.

As expected, growth rate gradually increased by the increasing uptake rate of both photons and bicarbonate. The minimum light requirement was about 2 mmol photons gDW^{-1} h^{-1}. At this photon uptake rate and below the light energy was entirely utilised for cell maintenance (simulated as an artificial ATP hydrolysis reaction at a fixed rate; see row 7 in Appendix B), as indicated by the white area in Figure 3.2.

To obtain insight into the properties of the two-cell model, and to test to what extent model-based predictions coincide with known metabolic exchange fluxes, the two-cell model was evaluated under photodiazotrophic conditions as well (Figure 3.3). Photons, bicarbonate and nitrogen uptake, unless

otherwise noted, were constrained at 10 mmol gDW^{-1} h^{-1}. An uptake reaction was varied in each case (indicated at the top of the corresponding panel), while the rest of the reactions were optimized within the given constraints.

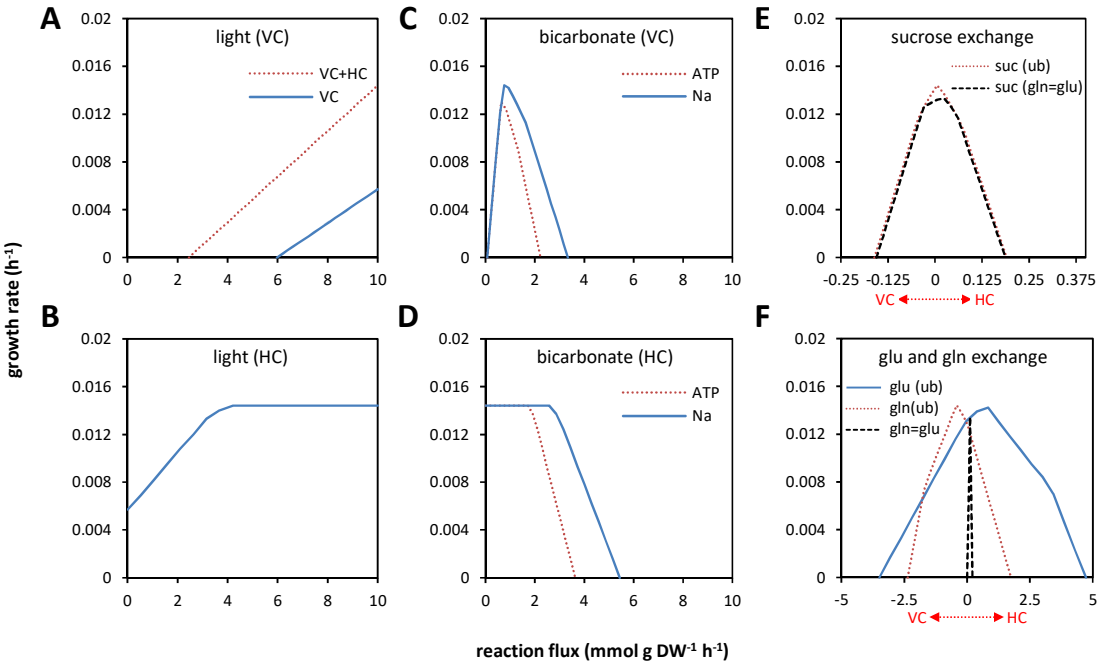


Figure 3.3. Growth rates predicted as the function of different transport reactions under diazotrophic conditions. (A) Impact of light availability on growth rate when both cell types (dotted red line) or only the VCSC (solid blue line) harvested photons. **(B)** Photon uptake by the HCSC in combination with optimal light harvesting in the VCSC. **(C)** Bicarbonate uptake by the two transport reactions (sodium symport, solid blue line; active transport, dotted red line) in the VCSC. **(D)** Bicarbonate uptake by the HCSC. **(E)** Exchange of sucrose in case when the glutamine-to-glutamate ratio was unbound (red dotted line) or when it was fixed to 1 (black dashed line). **(F)** Exchange of glutamate and glutamine at ratios fixed to 1 (black dashed line) or left unbound (solid blue and dotted red lines). Red double arrows on E and F show the direction of exchange. VC: vegetative cell super-compartment (VCSC); HC: heterocyst super-compartment (HCSC).

The minimum photon requirement to cover non-growth associated maintenance costs, without supporting growth (x-intercept), is shown in Figure 3.3A. This requirement was equal to a photon flux of 6 mmol g DW^{-1} h^{-1}, when light was harvested by the vegetative cell only (solid blue line). At the upper bound of the VCSC's photon uptake (10 mmol g DW^{-1} h^{-1}) the growth rate was 0.006 h^{-1} (Figure 3.3A, solid blue line and y-intercept on panel B) that increased up to the maximum (0.0144 h^{-1}) by the HCSC's contribution to light harvesting. In the case when both super-compartments harvested light (dotted red line) the energy contribution by the HCSC lowered the requirement from the VCSC by

about 2.3 mmol g DW^{-1} h^{-1}. This contribution from the HCSC via cyclic photophosphorylation saturated at approx. 4.2 mmol g DW^{-1} h^{-1}, over which the proton gradient through the thylakoid membrane was replenished by secondary reactions in the electron transport chain without the synthesis of additional ATP (Figure 3.3B).

In contrast, light or carbon uptake by the HCSC alone could not support growth of the VCSC (data not shown). On the other hand, any of the two bicarbonate transporters in the VCSC provided sufficient carbon for growth, although the maximum rate was lower for the ATP-driven transport (Figure 3.3C, dotted red line). The optimal bicarbonate uptake in the VCSC was 0.76 and 0.63 mmol gDW^{-1} h^{-1} for the symport and active transport, respectively.

When forcing bicarbonate uptake over the optimum, however, both transport reactions exhibited a negative effect on growth as light became limiting (Figure 3.3C, both lines). According to Figure 3.3D, maximal growth was achieved even at zero bicarbonate uptake by the HCSC, suggesting that the VCSC alone was able to fix sufficient amount of carbon (via RuBisCO in the carboxysome) and reach the maximum growth rate. Moreover, if bicarbonate uptake in the HCSC was enforced (Figure 3.3D, both curves), the carbon was not utilised for growth (straight horizontal lines on the left side of each curve), but rather recycled via the C_4 dicarboxylic acid cycle and released as carbon dioxide (not shown here). The recycling capacity depleted around 1.8 and 2.6 mmol g DW^{-1} h^{-1} bicarbonate over the active transport and the symport reactions, respectively, resulting in gradual decrease of growth rate (Figure 3.3D, both curves).

Notably, experimental evidence suggests that the main source of carbon for heterocysts is likely to be sucrose (Böhme, 1998; Martin-Figueroa et al., 2000; Kumar et al., 2010; Nürnberg et al., 2015). In contrast, the model predicted a very low flux for sucrose at the optimal growth rate (Figure 3.3E, red dotted line). In the same simulation, glutamate was transferred to the heterocyst at a two-fold higher rate than glutamine, and in the reverse direction. This is possible because glutamine and glutamate exchange were optimized as independent reactions and suggests that glutamate was partially utilised as a carbon source in the HCSC. However, sucrose became the primary source of carbon in the HCSC (also increasing its flux by 4-times) for cases when the glutamine-glutamate exchange ratio was fixed to 1; at the same time, the growth optimum decreased by only about 7% (Figure 3.3E and F, dashed black lines; row 13 in Appendix B).

3.2.3.2 Comparison of nitrogen sources

Several nitrogen sources including inorganic compounds and amino acids were investigated *in silico* for biomass yield under phototrophic conditions using the single-cell model (except for dinitrogen; in that case the two-cell model was used). ATP-driven transport reactions were set for every compound, except for molecular nitrogen and ammonia that were set to freely diffuse across membranes. The uptake fluxes for each nitrogen source were set to an upper bound of equimolar nitrogen (10 mmol nitrogen atom g DW^{-1} h^{-1}). Photon and bicarbonate uptake were either constrained to a maximum of 10 mmol g DW^{-1} h^{-1} or set to unconstrained (panel A or B on Figure 3.4, respectively). Calculations were carried out using the single-cell model and compared to diazotrophic growth on molecular nitrogen using the two-cell model. Results were sorted in an ascending order by growth rate on panel A and in a descending order based on light and bicarbonate consumption on panel B.

As expected, growth on molecular nitrogen achieved the lowest rate (panel A), and required the highest bicarbonate and photon uptake rates (panel B), due to the high energy demand of the nitrogenase reaction in the HCSC. Nitrate, nitrite and ammonia, the other inorganic nitrogen sources, consumed equal amounts of bicarbonate. However, the energy demand of their respective reactions required to reach the same metabolic state increased in that order, reflecting variations in electron requirement of the respective pathways (panel B). In addition, there was 14% difference in growth rate between ammonia and the other two inorganic nitrogen sources (nitrite and nitrate), in favour for the former. Indeed, ammonia is the preferred inorganic nitrogen source for most microorganisms (Herrero et al., 2001) and is also the most reduced form. The growth rate on nitrate was slightly lower than on nitrite, and about 44% higher than under diazotrophic conditions, showing a very good correlation with literature data (Pernil et al., 2010). In fact, each of these inorganic nitrogen sources end up as ammonia within the cell, and thereafter become incorporated into glutamate, while forming glutamine via the GS-GOGAT cycle. Not surprisingly, glutamate was an excellent source of nitrogen (and carbon) in these simulations, requiring no other source of carbon to achieve maximal growth (panel B), at a rate which was more than 17-times higher than under diazotrophic conditions (panel A).

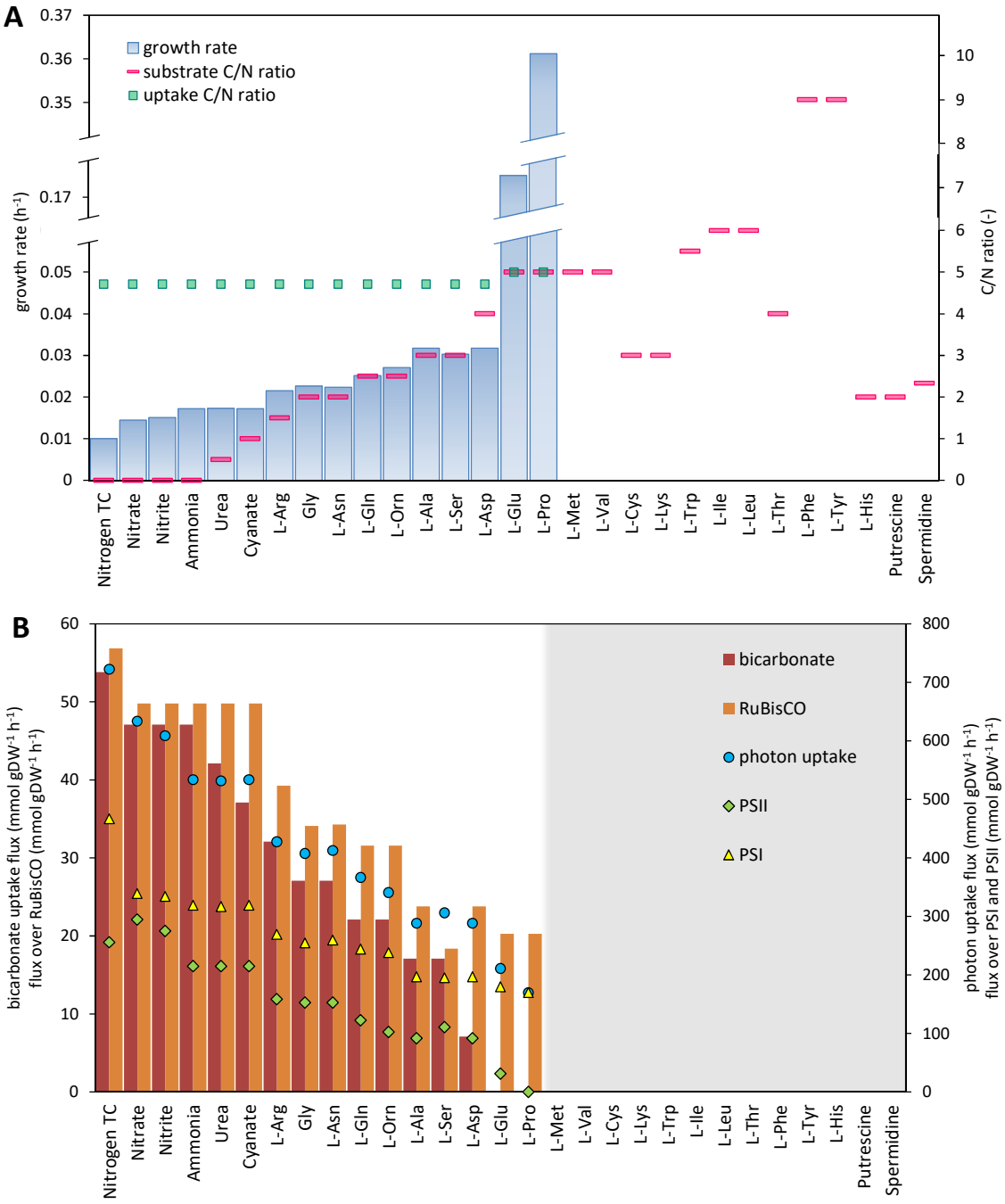


Figure 3.4. Comparison of nitrogen sources in phototrophic growth. (A) Predicted growth rates on different nitrogen sources under photon and bicarbonate uptake constrained at an upper bound of 10 mmol g DW^{-1} h^{-1} and an upper bound of equimolar nitrogen uptake. Uptake C/N ratio: all uptake reactions combined; substrate C/N ratio: ratio in the corresponding nitrogen source. **(B)** Carbon and energy requirement of nitrogen assimilation in case of the different nitrogen sources. Photon and bicarbonate uptake were unconstrained and nitrogen uptake was set to an upper bound of equimolar nitrogen. Growth rates were the same for cases with white background and no growth was observed for those with grey. PSI: flux over photosystem I; PSII: flux over photosystem II. Results for diazotrophic growth (Nitrogen TC) on both A and B were calculated using the two-cell model. In all other cases the single-cell model was used.

Growth characteristics on proline were similar to that on glutamate, involving an equal RuBisCO rate assimilating the internally generated bicarbonate from NADPH dehydrogenase. The dehydrogenase transferred electrons to plastoquinone (PQ) by pumping protons through the thylakoid membrane for both nitrogen sources, and therefore reduced the contribution of PSII (panel B). Indeed, assimilation routes for glutamate and proline are essentially the same as proline is converted to glutamate first via 1-pyrroline-5-carboxylate and glutamate-5-semialdehyde. Yet, proline achieved nearly two times higher growth rate than glutamate (panel A) and consumed 20% less photons, whilst PSII was found completely inactive (panel B). Interestingly, the first step in the conversion of proline towards glutamate donates additional electrons to the PQ pool rendering linear electron transport through PSII unnecessary. Furthermore, the C/N ratio of glutamate and proline were found optimal, as no bicarbonate was used in the simulation, in contrast to all the other nitrogen sources (green squares and red dashes on panel A). The stoichiometric model could optimally convert these compounds into biomass, whereas uptake fluxes had to be adjusted accordingly in all other cases to compensate for the non-optimal C/N ratio of the nutrients.

It is worth noting that methionine and valine have the same molecular C/N ratio as glutamate and proline, but no degradation route has been identified so far in cyanobacteria to connect valine to central carbon metabolism (Bothe and Nolteernsting, 1975; Serrano, 1992). Methionine degradation in *Anabaena* sp. PCC 7120 to 2-oxobutanoate via L-cystathionine is also unclear. Therefore, their carbon and nitrogen atoms cannot be assimilated by the metabolic network and accordingly, no growth was possible on neither of these substrates as sole nitrogen sources.

Among the remaining amino acids tested, alanine, aspartate, serine, ornithine, glutamine, glycine, asparagine and arginine were able to sustain growth at rates descending in that order. In the case of glutamine, asparagine and arginine, utilization as sole nitrogen source has previously been shown in some cyanobacteria (Neilson and Larsson, 1980; Flores and Herrero, 2004). In addition, ornithine and histidine (along with glutamine and glutamate) have been suggested to act as nitrogen exchange compounds via the periplasmic space of *Anabaena* sp. PCC 7120, originated most likely from heterocysts, and consumed primarily by vegetative cells under diazotrophic conditions (Montesinos et al., 1995).

Amino acids in general are also potential carbon sources for the cell, and thus growth on these compounds resulted in consequently higher growth rates than on inorganic compounds, cyanate and urea (panel A). These latter two were predicted to have a very similar growth rate as ammonia, showing the same 72% increase over diazotrophic conditions. In addition, photon uptake and RuBisCO reaction rates were also similar in unconstrained growth (panel B). Experimentally, cyanate has been

shown to serve well as the sole nitrogen source allowing growth at a near-maximal rate in marine cyanobacteria (Kamennaya et al., 2008). Cyanate can also accumulate internally as a by-product of the urea cycle or via the degradation of carbamoyl phosphate. Intracellular cyanate is highly toxic, and was shown to hinder growth of several species of cyanobacteria (Harano et al., 1997; Kamennaya and Post, 2011). In the case of *Anabaena* sp. PCC 7120, a putative cyanase gene was identified (Kaneko et al., 2001), although its role in urea-cyanate metabolism has not been evaluated yet. The detrimental effect of urea on the growth rate of *Anabaena* sp. PCC 7120 (Figure H-1) may indicate the toxicity of cyanate for this organism.

3.2.3.3 Mixotrophic growth by the single-cell model

To further investigate metabolic interactions between the two cell types, heterotrophic and mixotrophic growth were simulated on various carbon sources, with nitrate as the sole nitrogen source. Simulations were run on the single-cell model using one of the following as carbon source: bicarbonate, urea, sugars (glucose, fructose, sucrose and maltose), a sugar alcohol (glycerol), fermentation products (pyruvate, formate and acetate), amino acids (the 20 common L-amino acids and L-ornithine) or polyamines (putrescine and spermidine).

For comparative reasons, each uptake reaction was assumed to hydrolyse one molecule of ATP and was constrained to transport equal moles of carbon with each substrate at the upper bound. Light harvesting, bicarbonate and nitrate uptake were constrained to an upper bound of 10 mmol g DW^{-1} h^{-1}. Predictions by the model are shown on Figure 3.5.

The four sugars supported growth at high rates under both heterotrophic (red columns) and mixotrophic (blue columns) conditions (Figure 3.5). Growth on glucose, fructose and the disaccharides (sucrose and maltose) occurred at the exact same rate, at about 8 and 10-times the rate of autotrophic growth (purple column) for heterotrophic and mixotrophic conditions, respectively.

Acetate and pyruvate sustained growth *in silico* at rates 4.3 and 6.1-times higher than the control under heterotrophic conditions (red columns), respectively; the same carbon sources showed 5.7 and 7.7-times higher growth rate in mixotrophic conditions (blue columns), respectively. Interestingly, glycerol was found an outstanding carbon source. This compound achieved the overall highest growth rate with 11.4 and 9.7-fold increase over the control under mixotrophic and heterotrophic conditions, respectively.

In darkness under heterotrophic conditions, only glutamate and proline were predicted to sustain growth among the tested amino acids. Similarly, these two carbon sources showed the highest growth rate among amino acids under mixotrophic conditions. In addition, similar to sugars, acetate, pyruvate and glycerol, no additional bicarbonate was required for maximum growth on L-glu and L-pro (green bars in Figure 3.5). Moreover, nitrate uptake was also zero suggesting that glutamate and proline are well suited substrates for growth, containing the essential elements (carbon and nitrogen) at an optimal molar ratio. In contrast, glycine, alanine, serine, asparagine, aspartate, glutamine, ornithine and arginine were less favoured substrates, contributing to biomass production to a variable, but lower extent than glutamate and proline.

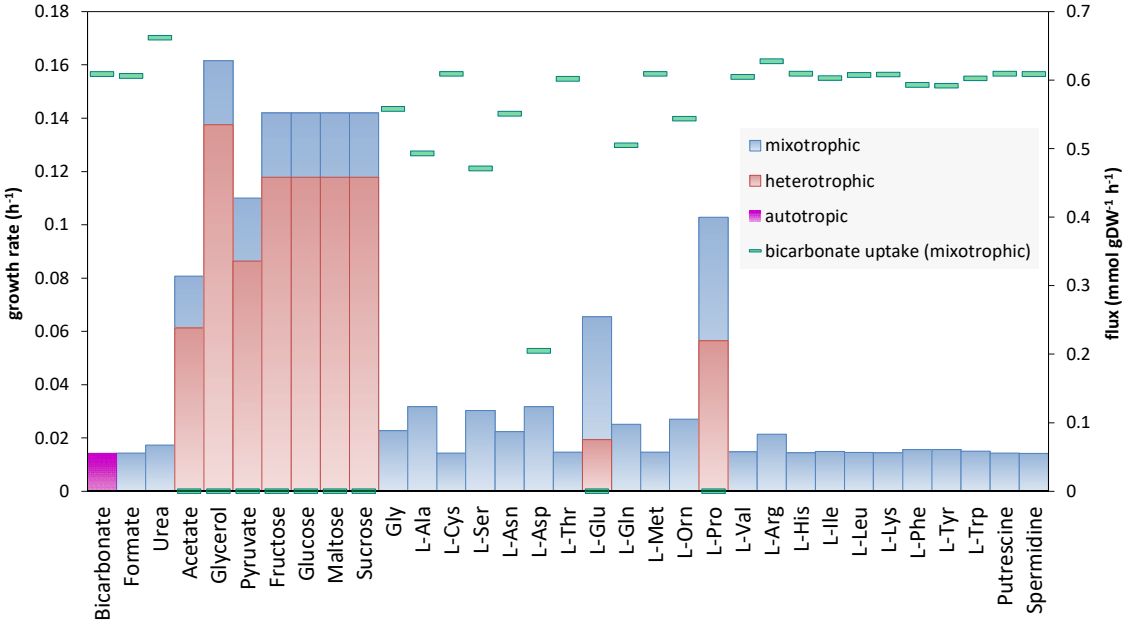


Figure 3.5. Predicted growth rates on different carbon sources under mixo- and heterotrophic conditions. Simulations were run using the single-cell model on nitrate as nitrogen source. Carbon source uptake reactions were set to hydrolyse one molecule of ATP for every molecule transported; their upper bound was constrained to transport carbon atoms equivalent to 10 mmol g DW^{-1} h^{-1} bicarbonate flux. Photoautotrophic growth on bicarbonate was also calculated as a reference (purple column).

Histidine, phenylalanine, tyrosine and tryptophan were utilised at very low rates. The rest of the amino acids, formate, urea and polyamines showed the same or very similar growth rates than the control autotrophic growth on bicarbonate.

3.2.4 Predictive ability of the single-cell model

Experimental (Figure H-1) and predicted (Figure 3.5) mixotrophic growth rates on a variety of carbon sources were compared relative to autotrophic growth on bicarbonate, using nitrate as the sole nitrogen source in the single-cell model. For comparative reasons, each uptake reaction was assumed to hydrolyse one molecule of ATP and was constrained to transport equal moles of carbon with each substrate (row 14 in Appendix B), as detailed in 3.2.3.3. Light harvesting, bicarbonate and nitrate uptake were constrained to an upper bound of 10 mmol g DW^{-1} h^{-1}. The two datasets of relative growth rates in exponential phase are shown in Figure 3.6.

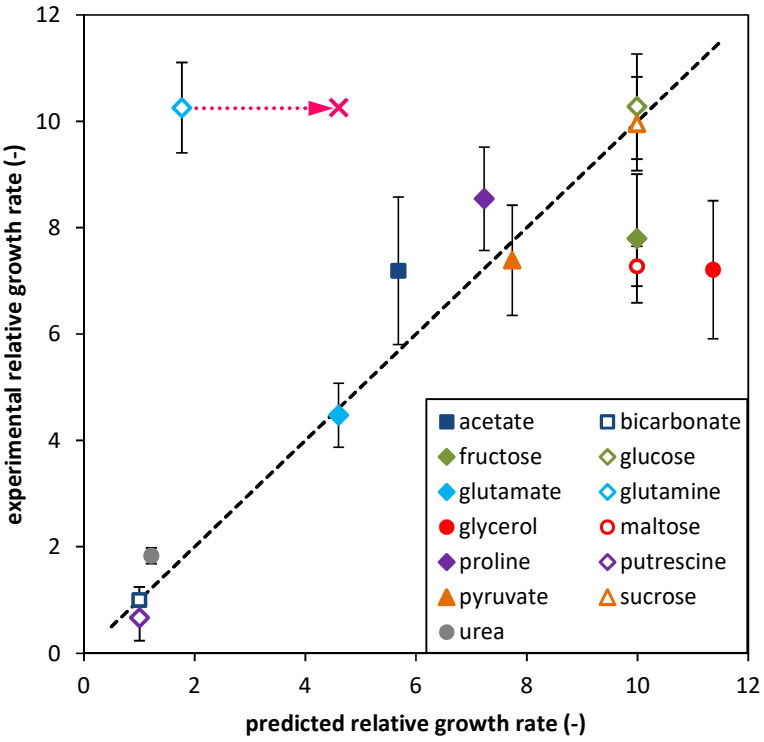


Figure 3.6. Correlation between experimental and predicted mixotrophic growth rates. Experimental growth was assessed in the exponential phase, relative to autotrophic growth on bicarbonate (Figure H-1). Error bars depict ± standard deviation of three biological replicates. The pink cross and dotted arrow show the shift in predicted growth rate on glutamine if ammonia is being excreted. The black dashed line highlights the correlation between experimental and computational datasets.

Most data points show a good fit to the trend, except for glycerol, some sugars and glutamine. The latter amino acid supported the second highest growth rate in the experiments, exceeded only by

glucose. In contrast, simulated growth rate on glutamine was just above the control bicarbonate. Interestingly, the model predicted the same growth rate for glutamine and glutamate when ammonia excretion was allowed, resulting in a 2.6-fold increase of growth rate on glutamine. Among the sugar compounds, datasets for glucose and sucrose showed good correlation, whilst fructose and maltose underperformed in wet-lab experiments. In fact, the four sugars showed essentially the same growth rate in simulations, due to high stoichiometric similarities between the metabolic pathways of these substrates. Surprisingly, *in silico* growth on glycerol resulted in the highest growth rate, whereas under experimental conditions glucose (and glutamine) achieved the highest rate.

The prediction of glycerol as the best carbon source was somewhat surprising. In fact, glucose and glycerol are both metabolised to glyceraldehyde 3-phosphate (GA3P), but via two different pathways (see further details in the discussions of this chapter, section 3.3). It is worth noting that sub-optimal growth on glycerol as sole carbon source, as compared to predictions using an FBA-based model, has also been previously observed for *Escherichia coli* K-12 (Ibarra et al., 2002). For *E. coli*, adaptive evolution resulted in an increased growth rate on glycerol, in good agreement with model-predicted values (Ibarra et al., 2002). For the present model, however, a variety of reasons might be responsible for the discrepancy, such as the thermodynamic properties of the dehydrogenation step to dihydroxyacetone phosphate (Figure 3.13), or lack of appropriate NAD(P)(H) balancing. Nonetheless, the results in Figure 3.6 show that the constraint-based model provides a reasonable prediction of mixotrophic growth rates on different carbon sources, thereby justifying the use of this stoichiometric model to evaluate the feasibility and optimality of potential exchange reactions in more detail.

3.2.5 Metabolite exchange by the two-cell model

Following the characterization of the single-cell model, stoichiometric optimality of metabolic exchange between vegetative cells and heterocysts was assessed using the two-cell model. Notably, a systematic analysis of the stoichiometric and energetic implications of different metabolites in intra-species cellular exchange is challenging to carry out experimentally, whilst being feasible *in silico*.

From previous studies, sucrose was proposed to act as the sole source of electrons and carbon for heterocysts (Curatti et al., 2002; Golden and Yoon, 2003; Cumino et al., 2007) and glutamine was suggested to serve as a nitrogen carrier, and glutamate as carbon skeleton for ammonia incorporation (Flores and Herrero, 2010; Kumar et al., 2010). In addition, the lack of GOGAT in heterocysts was postulated to result in the accumulation and subsequent transport of 2-oxoglutarate into vegetative cells (Böhme, 1998; Martin-Figueroa et al., 2000). These compounds and other central carbon

metabolites including some amino acids that may be involved in intercellular nitrogen exchange (Montesinos et al., 1995) were included as potential exchange metabolites in the two-cell model. In addition to amino acids, ammonia was also investigated as an alternative carrier of nitrogen. Therefore, in total, twelve different metabolites were considered as exchange compounds.

All possible combinations were comprehensively evaluated with respect to the predicted maximal growth rate, resulting in a total of 4096 combinations. Earlier simulations suggested that the ratio of glutamine and glutamate exchange may be constrained to unity (Figure 3.3) and therefore this case (as an additional constraint) was also included. Results were plotted as a distribution chart against the number of exchange reactions involved in each solution (Figure 3.7A). Selected metabolite combinations (Figure 3.7B) showing the highest growth rates by involving the least number of exchange metabolites were further evaluated on Figure 3.8 and Figure 3.9, highlighting the major metabolic pathways and reactions involved in the solution.

None of the exchange metabolites investigated here, including ammonia and glutamine, were able to individually allow growth of the filament (Figure 3.1). In addition, none of the two-metabolite and three-metabolite combinations resulted in maximal growth rate. As the best-ranked combination of two exchange reactions, export of ammonia and import of alanine (from the HCSC's perspective) resulted in 70% of the maximal growth rate (Figure 3.7B, case L). Moreover, although a total of twelve metabolites were allowed to exchange, no more than ten were ever chosen in any combination by the model to provide a feasible solution and non-zero growth. The absolute maximal growth rate was reached by combinations of seven reactions, although only subtle improvements could be observed over combinations of four reactions. Flux distributions were, to variable extent, different in each case.

At maximal growth rates, regardless of the number of exchange reactions involved or the glutamine-glutamate ratio, alanine was consumed by the HCSC and nitrogen was transferred as ammonia. Furthermore, sucrose was produced and excreted, rather than being consumed, by the HCSC. When sucrose was consumed (blue and purple areas), glutamine, glutamate and 2-oxoglutarate were transported as suggested by literature. However, the growth rate was on average 30% higher if sucrose was exported, independent of the glutamine-glutamate ratio.

Ammonia was favoured over glutamine for the transfer of nitrogen throughout the simulations, increasing growth rates by 5–7% when the glutamine to glutamate ratio was fixed (Figure 3.7A, yellow area) and by 3–7% when it was unbound (red area). When only four reactions were allowed, inclusion of ammonia increased growth rate by 38% over the reference (blue square) if glutamine to glutamate ratio was fixed, and by 34% over the reference (red circle) if unbound.

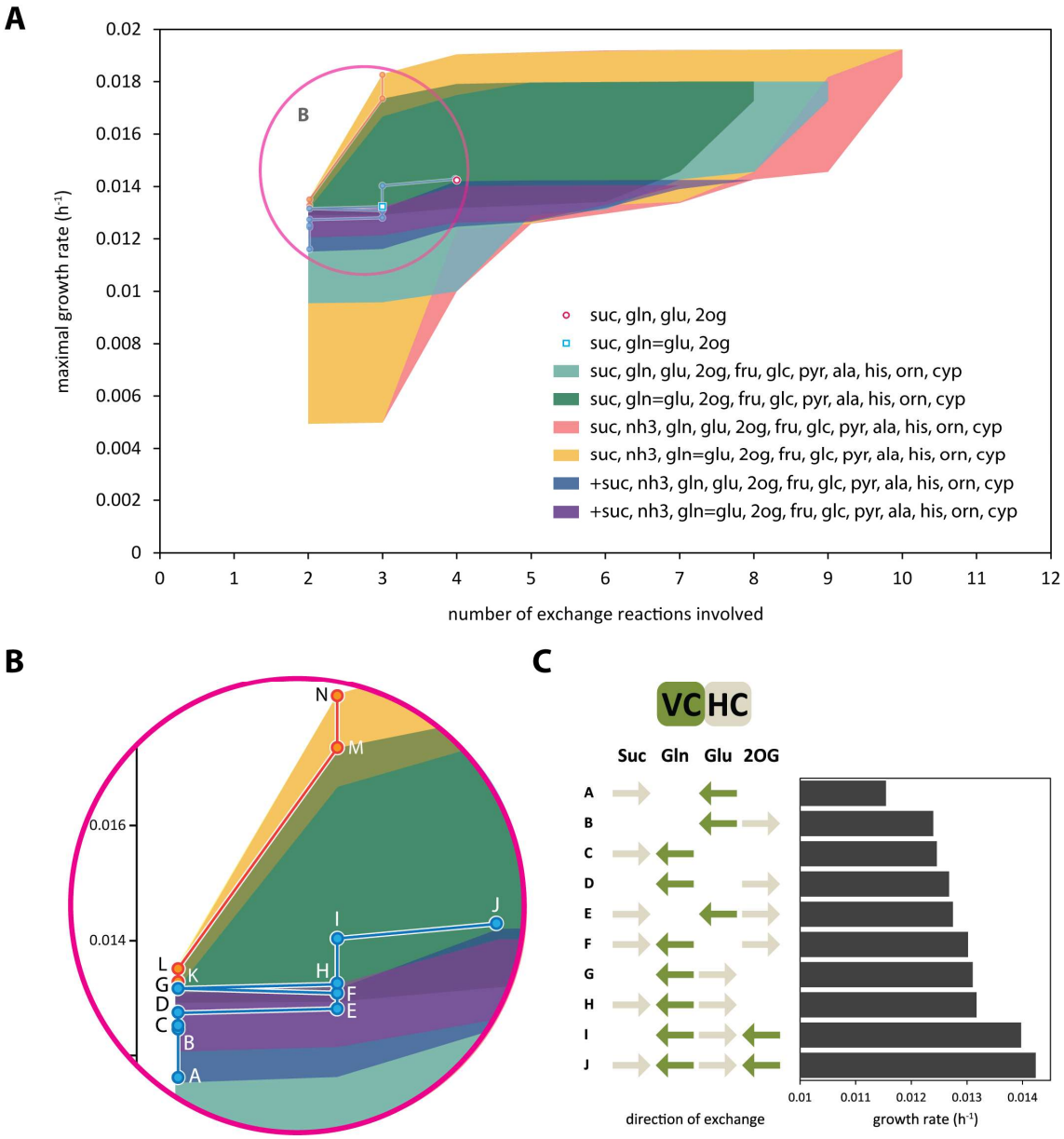


Figure 3.7. Predicted growth rates in response to the number of intercellular exchange reactions in the two-cell model. **(A)** Each coloured area represents the distribution of non-zero solutions with a different set of exchange reactions. In case of yellow and green glutamine to glutamate ratio was fixed at 1 and it was unbound for red and teal. Yellow and red areas evaluate the effect of ammonia exchange. Blue and purple areas highlight solutions where sucrose was consumed by the HCSC. Growth rate of four exchange metabolites (i.e. sucrose, 2-oxoglutarate, glutamate and glutamine) is highlighted by a red circle (glutamine to glutamate ratio unbound) and a blue square (ratio fixed to 1). All simulations were performed using the two-cell model. Abbreviations: suc: sucrose, 2og: 2-oxoglutarate, fru: fructose, glc: glucose, pyr: pyruvate and cyp: cyanophycin monomer. **(B)** Zoomed-in section from panel A showing growth rates for selected exchange metabolite combinations. Letters refer to cases on panel C and the corresponding panels on Figure 3.8 and Figure 3.9. **(C)** Combinations of a maximum of four reactions exchanging sucrose (Suc), glutamine (Gln), glutamate (Glu) or 2-oxoglutarate (2OG). Arrows indicate participating metabolites and direction of the exchange (uptake by the VCSC and the HCSC is represented by green and beige colours, respectively). Black bars show growth rate in each case.

Those four exchange metabolites suggested in literature, i.e. sucrose (Suc), glutamine (Gln), glutamate (Glu) and 2-oxoglutarate (2OG) were evaluated in more detail by looking at the flux distribution of each possible combination individually (Figure 3.7B blue line and panel C).

Only ten out of the sixteen exchange combinations allowed growth, with +Suc and –Glu giving the lowest growth rate (in the following, + and – signs preceding metabolite names denote uptake and excretion by the HCSC, respectively). In this case (Figure 3.8 case A), an incomplete TCA cycle was driven by sucrose originated from the VCSC to produce 2-og by IDH. The product 2-og was then converted to glutamate by GDH, incorporating ammonia fixed by nitrogenase. Eventually, electrons required by nitrogenase were also derived partially from sucrose via pyruvate. The glutamate produced was transferred back to the VCSC to serve as a source of assimilable nitrogen. This nitrogen was directly incorporated into different amino acids and finally, biomass (Figure 3.8 case A).

Growth rate increased in ascending order when –Glu +2-og (Figure 3.7C case B), +Suc –Gln (case C) and –Gln +2-og (case D) were combined. Interestingly, pairing glutamine with glutamate (–Gln +Glu, case G) improved growth rate the most significantly among all the combinations of two reactions. About 78% of glutamate was used to incorporate ammonia into glutamate by GS and was sent back to the VCSC as glutamine. The rest of the glutamate fueled the second half of the TCA cycle via aspartate transaminase (AAT). In the VCSC, GOGAT instead of GS became active and produced Glu for filament growth and for the HCSC.

The 2-og required by GOGAT was the product of a broken TCA cycle and AAT, converting oxaloacetate to aspartate in the VCSC (Figure 3.8 case G). Inclusion of 2-og exchange increased growth rate by another 7% (Figure 3.7C) allowing the HCSC to recycle some of the glutamate and return it to the VCSC as 2-og, independent of the exchange of nitrogen (Figure 3.8 case I). This carbon transfer from the HCSC in the form of 2-og was higher when Suc was also allowed to exchange (case J). In such case, sucrose was used to run only the second half of the TCA cycle without consuming any 2-og in the pathway, and sending the majority of 2-og to the VCSC. In the absence of an active first half of the TCA cycle the primary source of electrons for nitrogenase became malate dehydrogenase, also involving transhydrogenase (Figure 3.8 case J, orange arrow in the upper cell). Transhydrogenase shuffled electrons from NADH, produced in the malate dehydrogenase reaction, to NADPH. In all cases discussed above (Figure 3.8 case A, G, I and J), reduced ferredoxins for nitrogenase were mainly provided by ferredoxin-$NADP^+$ reductase (FNR) transferring electrons from NADPH to ferredoxin.

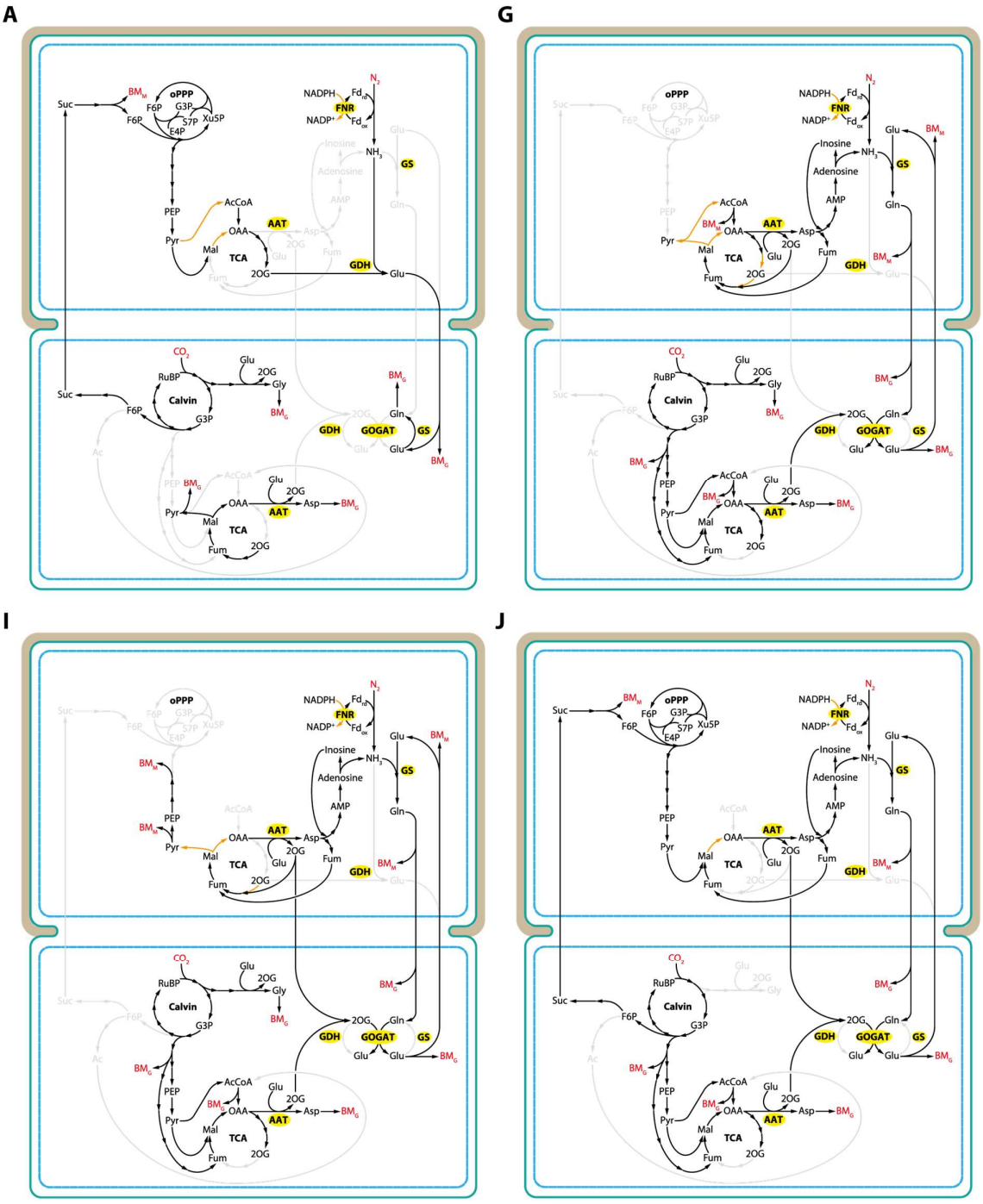


Figure 3.8. Main metabolic fluxes for the exchange of sucrose, glutamine, glutamate and 2-oxoglutarate. Panels are labelled according to cases on Figure 3.7B (blue dots) and C, and show increasing growth rates from (A) to (J). Exchange reactions are as follows: **(A)** +Suc –Glu; **(G)** +Glu –Gln; **(I)** +Glu –Gln –2-og; **(J)** +Suc +Glu –Gln –2-og. Ratio of Glu to Gln was unbound. Compounds in red indicate a sink or uptake of an external metabolite. Orange arrows in the HCSC highlight reactions providing electrons (NADPH) for nitrogenase. BM_G: biomass (growth), BM_M: biomass (maintenance). Enzyme names are highlighted in yellow. Upper cells: HCSC, lower cells: VCSC.

The highest increase in growth rate due to the addition of one more reaction occurred from two to three (Figure 3.7B, red line and circles), using ammonia as the carrier of fixed nitrogen. The lowest growth rate among those with only two reactions was achieved when ammonia exchanged for fructose (Fru, Figure 3.9 case K). Ammonia in the VCSC was assimilated solely by GS and the produced Gln was incorporated into other amino acids and biomass. The source of Fru was the Calvin cycle and Fru was converted to pyruvate in the HCSC. Pyruvate eventually formed oxaloacetate to maintain the TCA cycle that was providing NADPH for the reverse reaction at FNR, producing reduced ferredoxins for nitrogenase (Figure 3.9 case K). As discussed above, exchange of ammonia for alanine (Figure 3.9 case L) gave the highest growth rate among all the combinations of only two reactions. Ammonia in this case was assimilated by both GS and GDH in the VCSC with significantly higher contribution from the first enzyme, while GOGAT remained inactive. In addition, ammonia was also incorporated to pyruvate forming alanine that was then transferred back to the HCSC. The nitrogen carried by alanine was converted back to ammonia via aspartate and adenosine.

Allowing the exchange of sucrose, growth rate improved by about 30% (case M, Figure 3.7B) compared to the case when only +Fru and $-NH_3$ were exchanging (case K). Also, the model increased Fru exchange 100-times and set the flux of Suc accordingly. In other words, Fru and Suc exchange fluxes were closely equivalent for carbon content, except for some of the Fru consumed by the HCSC to run its TCA cycle. Stoichiometrically, the net outcome of this cycling of Fru to Suc in the HCSC and Suc to Fru in the VCSC is ultimately one mole of ATP for every mole of Fru, by the reverse reaction of fructokinase in the VCSC. It is unclear, however, whether such metabolite concentrations to shift the fructokinase reaction kinetics to the direction of ATP generation can ever occur in a living vegetative cell.

It is also worth noting, in most cases the model favoured, when available, the exchange of carbon sources other than sucrose, or transported sucrose towards the VCSC. In case of the fastest growing among those using only three reactions (case N), the increase in growth rate due to the addition of Suc exchange was about 36% compared to case L (Figure 3.7B), which involved only two reactions (+Ala $-NH_3$). Similar to that, case N had no active GOGAT in the VCSC, while actively assimilated ammonia via GS, GDH and alanine dehydrogenase producing alanine. The higher exchange rate of alanine compared to case L increased the flux over the second half of the TCA cycle as well (generating more reducing equivalents), allowing higher nitrogen fixation rate and ultimately, more vegetative cell biomass. The carbon content of Ala was recycled through pyruvate and fructose 6-phosphate, and sent back to the VCSC in the form of sucrose.

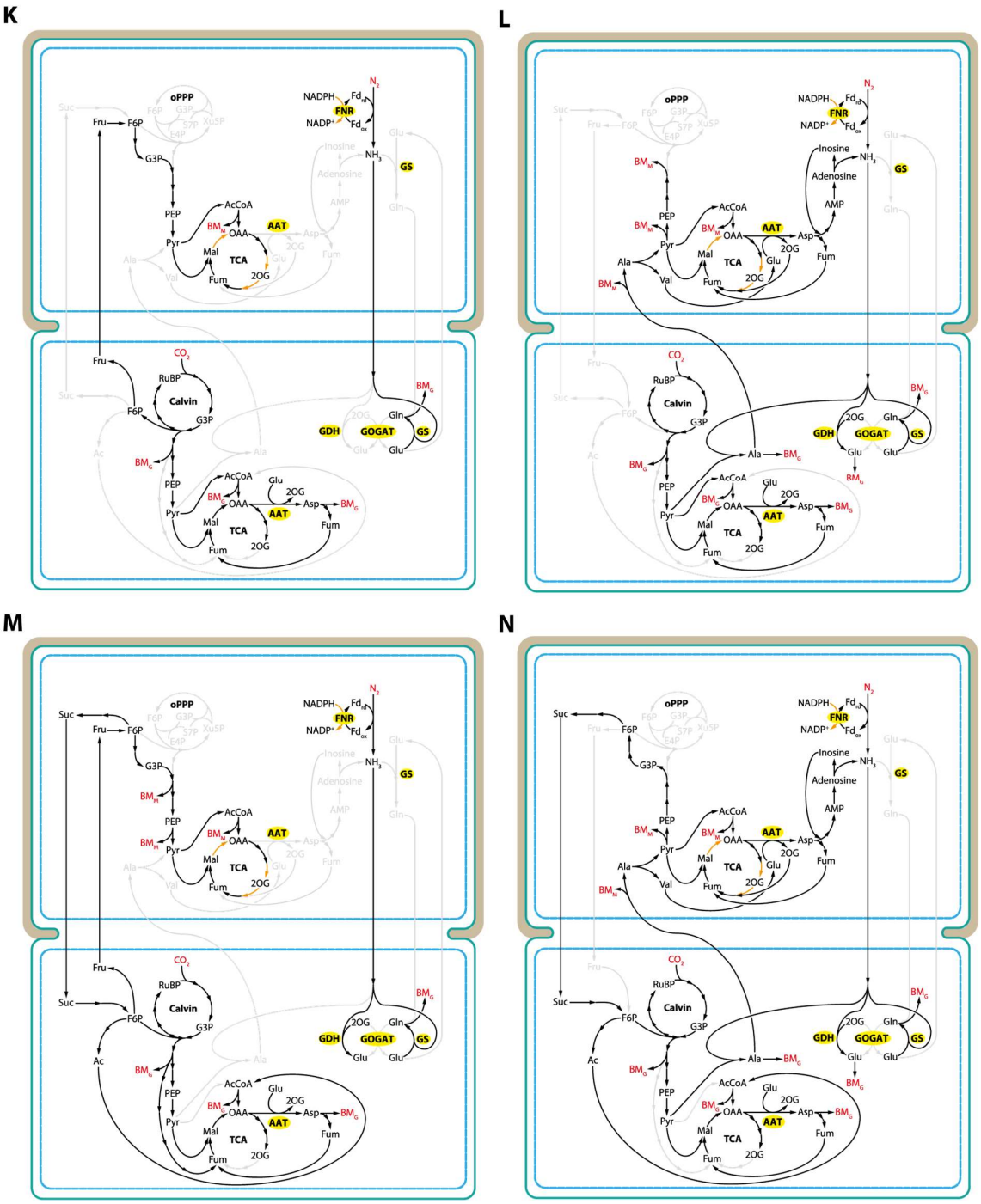


Figure 3.9. Main metabolic fluxes for the exchange of sucrose, fructose, alanine and ammonia. Panels are labelled according to cases on Figure 3.7B (red dots), and show increasing growth rates from (K) to (N). Exchange reactions are as follows: **(K)** +Fru –NH_3. **(L)** +Ala –NH_3. **(M)** +Fru –Suc –NH_3. **(N)** +Ala –Suc –NH_3. Ratio of Glu to Gln was unbound. Compounds in red indicate a sink or uptake of an external metabolite. Orange arrows in the HCSC highlight reactions providing electrons (NADPH) for nitrogenase. BM_G: biomass (growth), BM_M: biomass (maintenance). Enzyme names are highlighted in yellow. Upper cells: HCSC, lower cells: VCSC.

3.2.6 Prediction of gene essentiality

There is very little known about essential genes in cyanobacteria. While a vast body of knowledge exists for essential genes in other organisms (Chen et al., 2012; Cherry et al., 2012; Zhou and Rudd, 2013; Luo et al., 2014), only specific genes and gene clusters have been studied in cyanobacteria (Fukuzawa et al., 1992; Fujita et al., 1996; Fiedler et al., 1998; Porankiewicz et al., 1998; Katoh et al., 2001; Shibata et al., 2002). In addition, there are no comprehensive lists or databases of essential genes that would assist the design of engineering strategies for cyanobacteria; rather, the available information is extremely scattered. Stoichiometric models, however, may be used to compile such resources based on certain computational criteria (Joyce and Palsson, 2008). In recent years, metabolic models have been successfully employed to probe the metabolic capabilities of a number of organisms, to generate and test experimental hypotheses, and to predict accurately metabolic phenotypes and evolutionary outcomes (Becker and Palsson, 2005; Deutscher et al., 2006; Raghunathan et al., 2009; Suthers et al., 2009; Orth et al., 2011). Similar to these studies, gene essentiality was determined *in silico* for *Anabaena* sp. PCC 7120, using the stoichiometric model described here and available in Malatinszky et al. (2017).

In the following context, gene essentiality was defined as the loss of ability to produce biomass in a simulation, upon deletion of the metabolic reaction(s) the corresponding gene is associated with. As part of the reconstruction process (detailed in section 2.8.1), GPR associations were determined for every reaction in the model. Based on these GPR associations, all reactions a certain gene is required for (i.e. no alternative gene for that reaction with “OR” relationship in the GPR string) were listed for that gene. To determine gene essentiality, genes were taken one by one and the reactions they are required for were systematically removed from separate instances of the stoichiometric model. The resulting models, lacking all reactions a particular gene is required for, were optimized for biomass production rate (i.e. growth). A gene that, upon removal of the associated reactions, prevented growth was termed “essential”. The distribution of essential genes in the genome of *Anabaena* sp. PCC 7120 is displayed on Figure 3.10. Each column spans over 64 kilobases in the genome and represents the number of genes found in that region.

The distribution of metabolic genes in the genome of *Anabaena* sp. PCC 7120 showed high variations (blue columns). Some regions were found considerably richer in genes of biochemical function than others. In addition, close to the middle of the chromosome the frequency of metabolic genes was particularly low, with a single 64-kilobase bin harbouring no genes at all. Indeed, this region was found the least populated by any genes, as shown by the green columns representing all annotated ORFs.

The distribution of essential genes (red columns) somewhat followed the trend of metabolic gene distribution (blue columns), and no particular pattern could be observed.

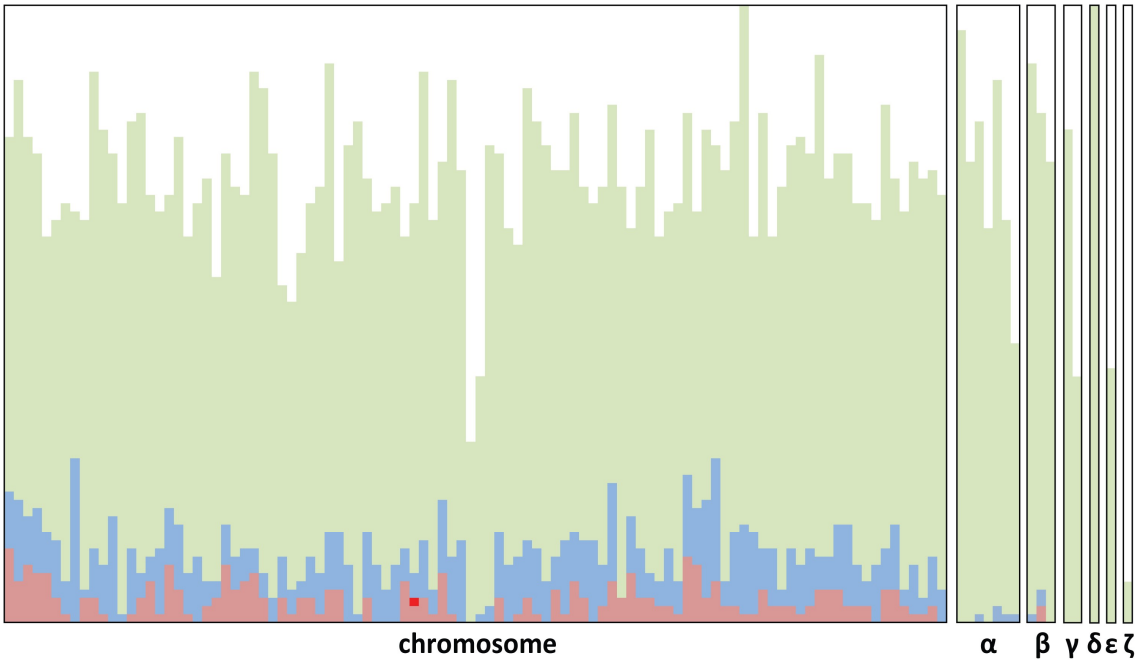


Figure 3.10. Distribution of genes in the genome of *Anabaena* sp. PCC 7120. Boxes represent the chromosome and the natural α–ζ megaplasmids. Green columns display the distribution of all annotated genes, whereas blue columns show the distribution of metabolic genes based on their genomic location (bin size: 64 kilobases). Red columns represent essential genes. Location of the essential *glnA* gene encoding glutamine synthetase is highlighted by a dark red square.

Interestingly, only the α and β megaplasmids were found to carry metabolic genes among the natural plasmids of *Anabaena* sp. PCC 7120. In addition, only two essential genes were found on plasmids (both of them on the β megaplasmid): *alr7622* and *all7623*. The gene *alr7622* encodes for a Cd^{2+}/Zn^{2+} exporting ATPase, the only membrane protein responsible for the transport of zinc(II) ions in the model. The other one, *all7623* encodes for a phosphatidate phosphatase involved in glycerolipid biosynthesis, and was identified as a novel gene candidate based on homology to cyanobacterial genes of similar functionality during the reconstruction process (Appendix E).

It is worth noting that *glnA*, the only gene encoding for GS in *Anabaena* sp. PCC 7120 (Reyes and Florencio, 1994) was also determined as essential by the above process. This information was very important in designing metabolic engineering strategies for the excretion of ammonia in Chapter 4.

However, the schizokinen cluster and its genes, *all0390–all0396* have been predicted as non-essential and therefore, a systematic knockout study could be considered as a viable approach (Chapter 5).

3.2.7 Evaluation of nitrogen excretion

After challenging the stoichiometric model to use a variety of carbon and nitrogen sources, and evaluating growth rates under diazotrophic (two-cell) and non-diazotrophic (single-cell) conditions, the model was used to compare a selection of nitrogen-containing metabolites in terms of their maximal possible excretion rate under the same growth conditions. The stoichiometric model was constrained at a constant 10% biomass production (growth) rate and at an upper bound of 10 mmol g DW^{-1} h^{-1} bicarbonate and dinitrogen uptake rate. Light harvesting was also constrained at an upper bound of 10 mmol g DW^{-1} h^{-1}. Transport reactions to excrete each compound were temporarily added to separate instances of the model. These transport reactions were set to require no ATP or other compounds for excretion, but the corresponding nitrogen-containing metabolite.

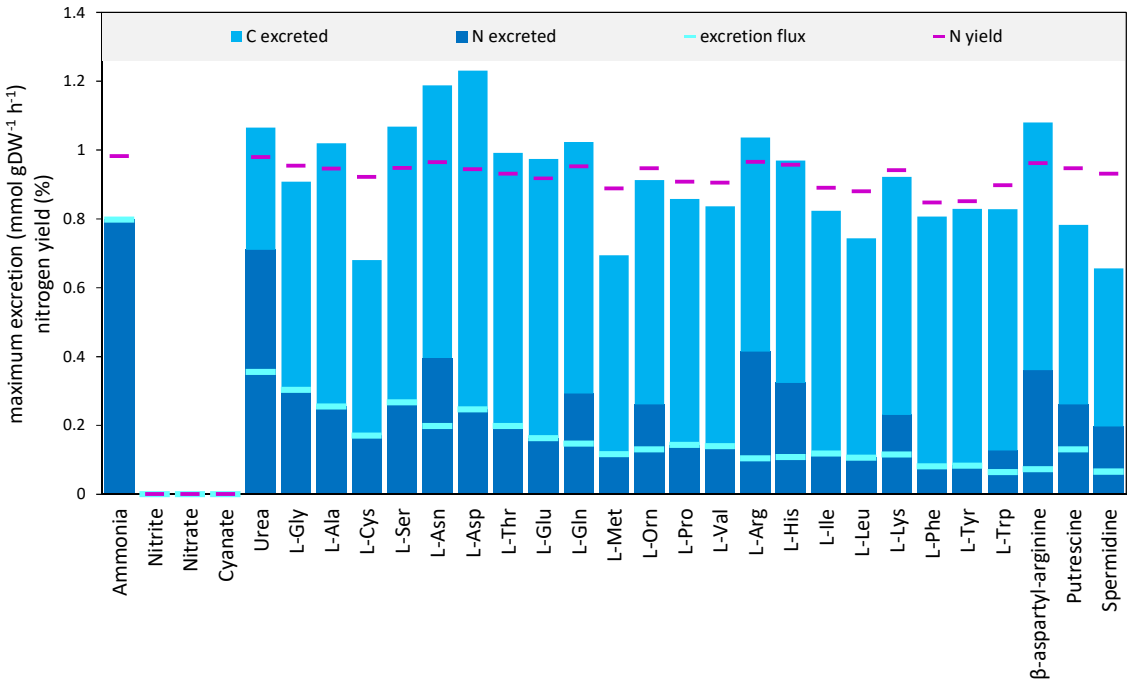


Figure 3.11. Comparison of metabolic capacity for the excretion of nitrogen compounds. The model was constrained to 10% of optimal growth rate and production of each nitrogen compound was optimized. Cyanophycin storage was simulated by the excretion of its monomer, β-aspartyl-arginine. Light blue bars display excretion flux in mmol gDW^{-1} h^{-1}. Purple bars show nitrogen yield based on excreted nitrogen and N_2 uptake (diazotrophic conditions). Stacked columns represent the molar carbon (lighter blue) and nitrogen (darker blue) content of the nitrogen-carrying molecule in the excretion flux.

Simulations were set to optimize the excretion of each nitrogen compound in Figure 3.11. The solutions of these simulations (maximum achievable excretion of the corresponding compound under the defined conditions) were compared for their delivery of nitrogen and carbon on a molar basis.

Stacked columns in Figure 3.11 show the carbon (light blue columns) and nitrogen (dark blue columns) content of the excretion flux (bands of a different light blue colour). Among all the nitrogen compounds tested, ammonia was found to excrete nitrogen at the highest molar rate reaching 0.8 mmol gDW^{-1} h^{-1}. The second best metabolite for nitrogen excretion was urea, with 0.71 mmol gDW^{-1} h^{-1}. The rest of the nitrogen compounds achieved significantly lower excretion rates. However, all organic nitrogen compounds including urea excreted considerable amounts of carbon as well. Interestingly, nitrogen excretion yield based on total dinitrogen uptake and nitrogen release by excretion was relatively similar for all metabolites (purple bars). In case of ammonia and urea the nitrogen yield was about 98%, and varied between 96% and 85% for the rest of the compounds, with L-phe being the lowest. Regarding the 21 amino acids including L-orn, excretion of L-asn and L-arg allowed the highest rate of nitrogen release. The cyanophycin monomer β-aspartyl-arginine showed similar excretion characteristics and nitrogen yield as L-asn and L-arg, most likely due to high similarities between the underlying bioconversion pathways of these three compounds. At the same time, L-glu and L-pro, the two amino acids that were found excellent substrates for growth in Figure 3.4 and Figure 3.5 were not particularly efficient in nitrogen excretion.

Notably, ammonia was the only inorganic nitrogen species that could be excreted. The reason is no enzymatic reactions can be found in the model to convert N_2 to NO_2^-, NO_3^- or NCO^-. In fact, such reactions may not exist in other organisms either due to the large shift in the reactions' chemical equilibrium to the opposite direction (calculated by eQuilibrator).

Although both ammonia and urea were found promising candidates for the excretion of nitrogen under diazotrophic conditions, ammonia was selected as the preferred candidate for two reasons: (1) nitrogen release in the form of ammonia is independent of carbon excretion, and (2) urea was found potentially detrimental to growth in *Anabaena* sp. PCC 7120 (Figure H-1).

3.3 Discussion

The metabolic model of *Anabaena* sp. PCC 7120 was reconstructed to contain a total of 865 genes and 923 unique reactions in two super-compartments. Gene annotations and associated biochemical reactions were collected from public databases (Kanehisa et al., 2004; Nakao et al., 2010; Caspi et al.,

2012; Tatusova et al., 2014) and curated manually to incorporate recent advances in literature and to be mass and charge balanced (Appendix D). Notably, the current annotation of the *Anabaena* sp. PCC 7120 genome contains a total of 6223 genes (Kaneko et al., 2001) in contrast to the 865 genes of metabolic function included in the metabolic reconstruction (Malatinszky et al., 2017), of which 259 were found to be essential for growth *in silico* (Figure 3.10). More than half (about 58%) of these annotated genes are, however, unknown or hypothetical. In addition, a large number of proteins and their associated reactions were found to be isolated from other parts of the network. These associations were not part of any known cyanobacterial metabolic pathway, and have therefore been removed from the reconstruction (orphan reactions in Appendix A). At the same time, novel genes were identified to resolve essential gaps in the metabolism of *Anabaena* sp. PCC 7120 (Appendix E).

Due to the lack of comprehensive data on the composition of the *Anabaena* sp. PCC 7120 biomass, a biomass model originally developed for *Synechocystis* sp. PCC 6803 was used (Nogales et al., 2012; Knoop et al., 2013). The ratio of the major biomass fractions (DNA, RNA, proteins, lipids, cell wall, pigments, inorganic ions, glycogen and metabolite pool) was adopted without modification. It is worth noting, the two cyanobacteria exhibit significant differences in their cellular physiology as well as their metabolism under certain environmental conditions. Most importantly, *Synechocystis* sp. PCC 6803 is a unicellular, non-diazotrophic cyanobacterium in contrast to the filamentous, heterocyst-forming, diazotrophic *Anabaena* sp. PCC 7120. In addition to the expectedly larger contribution of cell wall components to the total biomass in *Anabaena* sp. PCC 7120 (due to the thick heterocyst envelope), *Synechocystis* sp. PCC 6803 is well known for its higher DNA content, maintaining a few hundred copies of its chromosome under some conditions (Zerulla et al., 2016). Furthermore, composition of the heterocyst biomass is different from that of the vegetative cell (Kumar et al., 2010). It was therefore important to investigate the implications by an “improperly” formulated biomass objective function.

The relative ratio of the individual biomass fractions was varied to study the effects of biomass composition on FBA model predictions. The contribution by each fraction was tested at 80% and 120% where 100% is the contribution of that fraction in the original BOF. The contribution of the rest of the fractions was scaled up or down accordingly, while keeping their relative ratios. The resulting change in growth rate compared to the growth rate using the original BOF is given in Figure 3.12.

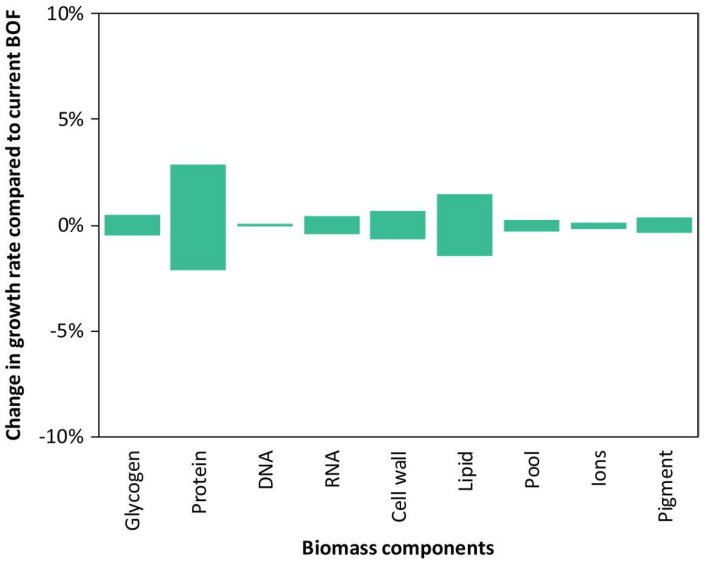


Figure 3.12. Response of growth rate to changes in the biomass fractional composition. The individual components of the BOF were varied at ±20% one at a time, so that the rest of the components were scaled up or down, while keeping their original ratios to each other. Green columns display the percentage difference plus and minus at optimal growth rate compared to using the original BOF.

The response of growth rate to variations of biomass composition showed significant difference in case of the individual biomass components (green columns in Figure 3.12). Clearly, varying the protein fraction displayed the highest impact on growth rate, followed by that of the lipid fraction. However, this effect, even in case of varying the protein fraction, translated to less than 3% change in growth rate compared to optimal growth using the original BOF. At the same time, the effect of varying DNA and cell wall fractions changed the growth rate below 1% in case of both, although these two fractions have been hypothesized as the most different components of the *Synechocystis* sp. PCC 6803 and the *Anabaena* sp. PCC 7120 biomass. This test confirmed that the fractional composition of the model's biomass may change quite dynamically without significant impact on optimal growth. It is important to note, however, variations in biomass composition may affect the individual reaction fluxes in the network more significantly. Consequently, a precisely formulated BOF is a prerequisite of a highly predictive model (see issues with glycerol and glutamine below).

The organism was represented by two different modes of the stoichiometric model according to environmental conditions. Growth on combined nitrogen sources (e.g. nitrate or ammonia) was tested using the single-cell model that contained the representation of a vegetative cell. Under diazotrophic conditions, however, the filamentous structure of *Anabaena* sp. PCC 7120 was simulated by the association of two cell types, a vegetative cell represented by the VCSC and a heterocyst represented

by the HCSC (Figure 3.1). The two super-compartments were connected by exchange reactions transferring glutamine, glutamate, sucrose and 2-oxoglutarate. Both cell types contained the same biomass equation comprising of DNA, RNA, proteins, lipids, pigments, cell wall, inorganic ions and the free metabolite pool (Table 3.2 and Appendix G). In the case of the two-cell model, growth was simulated by the growth of the VCSC, while the HCSC was also running the incorporated biomass reaction at a constant rate (10% of the optimal growth rate) to account for its macromolecular turnover and energy requirement. The majority of known physiological and metabolic differences between the two cell types were implemented in the model (Table 3.1 and Appendix C). Most importantly, the HCSC was incapable of oxygen evolution at PSII, and consequently, linear photophosphorylation. Instead of oxygenic photosynthesis, the energy requirement of nitrogen fixation and the maintenance of the HCSC itself were covered by cyclic photophosphorylation at PSI, as described previously (Razquin et al., 1996; Wolk et al., 2004; Vu et al., 2012). In order to be able to compare the two modes of the stoichiometric model (i.e. single-cell and two-cell) under the same light regimes, optimal distribution of light harvesting between the two super-compartments of the two-cell model was determined. As a result, it was found that the optimal distribution of 10 mmol g DW^{-1} h^{-1} photons among the two super-compartments was 7 and 3 mmol g DW^{-1} h^{-1} (scaling proportionally) in the VCSC and the HCSC, respectively. Nonetheless, in all other simulations where such comparison of the two modes was not necessary, photon uptake by both super-compartments of the two-cell model was constrained to 10 mmol g DW^{-1} h^{-1}.

The model correctly predicted the vegetative cell to heterocyst ratio under diazotrophic conditions showing that the HCSC (a single heterocyst) can supply the formation of 7.6 vegetative cells at maximum growth rate. This ratio was increased at the expense of VCSC growth rate, decreasing growth rate to 46% of the maximum to sustain 20 vegetative cells by a single heterocyst. Nitrogen uptake rate in the HCSC was not limited by the upper bound of its current constraint (i.e. 10 mmol g DW^{-1} h^{-1}) or by light, but instead, it was limited by the amount of carbon skeleton (glutamate) the VCSC was able to provide for the incorporation of ammonia. From a stoichiometric point of view, the HCSC can be forced to produce even more fixed nitrogen from more glutamate, resulting in further decrease of VCSC growth rate.

The two-cell model was tested under photodiazotrophic conditions to evaluate the responses of the network to changes in light and nutrient availability. Growth on bicarbonate and dinitrogen required a minimum of 2 mmol photons gDW^{-1} h^{-1} for cell maintenance (Figure 3.2). Over this photon uptake rate, growth rate gradually increased. Light harvesting by the VCSC alone sustained growth at 0.006 h^{-1}, about 42% of the theoretical maximum that was achieved when both the VCSC and the HCSC were

allowed to capture photons (Figure 3.3A). In contrast to light, bicarbonate uptake was only beneficial for growth when assimilated by the VCSC. The optimal rate for bicarbonate uptake was 0.76 and 0.63 mmol gDW^{-1} h^{-1} for the symport and active transport, respectively (Figure 3.3C). In case of the HCSC, however, enforced bicarbonate uptake posed a negative effect on growth rate. Instead of being used for growth, the "unwanted" carbon in the HCSC was recycled via the C_4 dicarboxylic acid cycle and released as carbon dioxide, reaching its capacity limit at 1.8 and 2.6 mmol g DW^{-1} h^{-1} bicarbonate through the active transport and the symport reactions, respectively (Figure 3.3D). Nonetheless, it is still unclear whether in reality heterocysts utilise this C_4 route to fix their own carbon (Popa et al., 2007); although cyanobacteria have been described to assimilate about 20% of the total fixed CO_2 in the form of C_4 acids (Owttrim and Colman, 1988; Luinenburg and Coleman, 1992, 1993).

The stoichiometric model was further challenged under mixotrophic conditions on nitrate as the nitrogen source in single-cell mode. Together with bicarbonate, thirteen different carbon sources were compared in simulations for growth rate; standard BG-11 medium was also supplemented with the same thirteen carbon sources and their impact on growth was evaluated experimentally. The two datasets showed good correlation for most data points (Figure 3.6). As expected, bicarbonate sustained growth at the lowest rate under both *in silico* and experimental conditions. The second highest growth rate was observed for glutamine (L-gln) experimentally, slightly behind the growth rate on glucose. In contrast, simulated growth rate on L-gln was just above the control bicarbonate.

However, the model predicted the same growth rate for glutamine and glutamate when ammonia excretion was allowed, resulting in a 2.6-fold increase of growth rate on glutamine (Figure 3.6, pink cross). This suggests that the two-fold higher molar nitrogen content of L-gln makes this amino acid a stoichiometrically unbalanced substrate compared to glutamate, which can only be overcome by the removal of excess nitrogen (in the form of ammonia in this case). By excreting nitrogen the model compensated for the suboptimal C/N ratio of glutamine and shifted the intracellular C/N ratio to the same level as with glutamate (Figure 3.4A, green squares). Nevertheless, the adjustment via ammonia excretion brought predicted growth on glutamine only halfway to experimental levels and the contrasting difference between the two datasets therefore remained incompletely explained. In addition, excretion of ammonia under diazotrophic conditions has not yet been observed experimentally for *Anabaena* sp. PCC 7120. However, it may be possible that the excess nitrogen is being deposited to nitrogen storage (i.e. cyanophycin), rather than being lost via excretion, which scenario cannot be tested due to limitations of FBA. Moreover, the remarkable difference between experimental and predicted datasets might indicate that importance of L-gln as a component of the *Anabaena* sp. PCC 7120 biomass is underestimated. In such a case, re-formulation of the

corresponding objective function in the stoichiometric model by precise measurements of the contribution by glutamine may help improving the fit of *in silico* data to experimental values for this particular compound.

While glutamine was the best substrate for growth under experimental conditions, glycerol resulted in the highest growth rate in simulations using the single-cell model. The 13% higher biomass production rate achieved on glycerol relative to glucose indicates a slightly more efficient stoichiometric solution for the underlying network compared to that in case of glucose. The metabolic efficiency suggested by simulations for these two substrates, however, contrasts with the experimental data (Figure H-1 in Appendix H). One explanation may be the differences in the pathways metabolising both glucose and glycerol to glyceraldehyde 3-phosphate (GA3P).

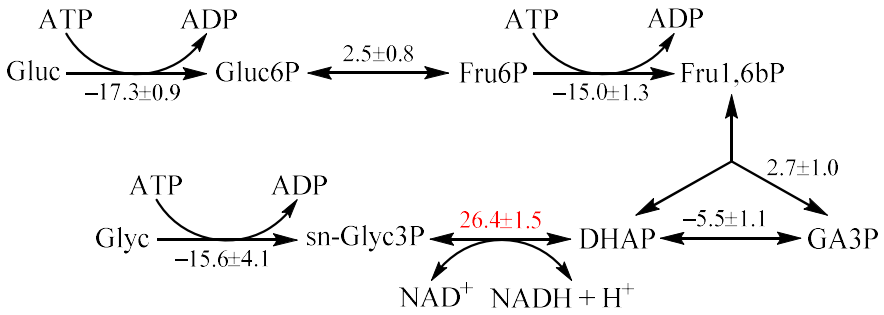


Figure 3.13. Comparison of glucose and glycerol assimilatory pathways. Numbers over or under arrows represent eQuilibrator (Flamholz et al., 2012) predictions for the Gibbs free energy change ($\Delta_r G'^m$, kJ/mol) during a reaction from left to right. Abbreviations: Gluc: glucose, Gluc6P: glucose 6-phosphate, Fru6p: fructose 6-phosphate, Fru1,6bP: fructose 1,6-bisphosphate, Glyc: glycerol, sn-Glyc3P: sn-glycerol 3-phosphate, DHAP: glycerone phosphate, GA3P: glyceraldehyde 3-phosphate.

In the case of glucose, glycolysis requires five consecutive enzymatic steps and two moles of ATP for each mole of glucose to produce two moles of GA3P (Figure 3.13). In the fourth step, fructose 1,6-bisphosphate is converted to dihydroxyacetone phosphate (DHAP) and GA3P. The second mole of GA3P is derived from DHAP in the last step. In contrast, one mole of glycerol requires only 3 enzymatic steps for conversion to one mole of GA3P; however, it also expends one mole of ATP and produces one mole of NADH. Therefore, for an equimolar amount of GA3P, two extra moles of NADH are generated in cells fed with glycerol compared to glucose. At the same time, the predicted Gibbs free energy change for the dehydrogenation step to DHAP in the glycerol route (Figure 3.13, numbers highlighted in red) suggests that the reaction in this direction is thermodynamically unfavourable. As

a consequence, either very high initial substrate concentration or rapid removal of product would be required in order for the reaction to proceed (Noor et al., 2014). As a property of standard genome-scale FBA models, this reconstruction also lacks such biophysical constraints. This may be one reason, among many possible, for the large discrepancy between predicted and observed growth rates with glycerol. Another potential issue is the lack of flexible NAD(P)(H) balancing since metabolism of glycerol is expected to result in the reduction of NAD^+, a likely minor electron carrier with unclear routes for metabolic homeostasis in cyanobacteria (Peschek, 1999). In other words, the FBA model can flexibly juggle electrons between the NAD and NADP electron carriers using the unconstrained transhydrogenase reaction, and therefore predict high growth rates on glycerol (involving the reduction of NAD^+). However, such transhydrogenase activity is not likely to be present in a living cyanobacterium (Kämäräinen et al., 2017).

Despite some disagreements of simulated and measured data for some carbon sources, the single-cell model predicted the growth of *Anabaena* sp. PCC 7120 for most substrates fairly well. These tests also showed the flexibility and robustness of the stoichiometric model in describing the various metabolic states of this organism under non-diazotrophic conditions. In order to predict the metabolic behaviour of diazotrophically grown *Anabaena* sp. PCC 7120, intercellular metabolite exchange was studied using the two-cell model. In total, 4096 exchange metabolite combinations were evaluated between the VCSC and HCSC, from single metabolite translocation to the complex barter of twelve metabolites. Surprisingly, no more than ten metabolites were ever active simultaneously, and the growth rate increased up to seven metabolites involved in nutrient exchange (Figure 3.7A).

A minimum of two exchange metabolites were always required for growth; in the case of the best pair (ammonia and alanine), the growth rate was 70% of the achievable maximum. In general, when ammonia was transferred, growth rates were consistently higher by about 3–7% than the growth rates obtained using glutamine as the nitrogen carrier. In the case of four reactions, inclusion of ammonia increased growth rates by 34–38%. In addition to that, the overall maximum growth rate was about 38% higher than that predicted for the four exchange metabolites (L-gln, L-glu, sucrose and 2-og) suggested in literature.

Interestingly, for most metabolite combinations overperforming the literature case, carbon was transferred in forms other than sucrose. In addition, even when sucrose was transported, it was excreted by the HCSC, rather than consumed, suggesting that stoichiometrically this direction of sucrose exchange is preferred. In other cases when both fructose and sucrose were exchanged (among other metabolites), the model found a way to ultimately transfer ATP from the HCSC to the VCSC. This transfer utilizing the reverse reaction of fructokinase may be thermodynamically feasible at product

concentrations 100-times higher than that of fructose, although there is no evidence that such concentrations occur in a living cell.

Nevertheless, it is important to evaluate the consequences of different exchange metabolite combinations computationally, although not all predictions will necessarily match experimental observations. For example, to understand the role of sucrose, the four-reaction case (exchange of glutamine, glutamate, sucrose and 2-oxoglutarate) was further investigated by systematically changing the parameter settings of the two-cell model.

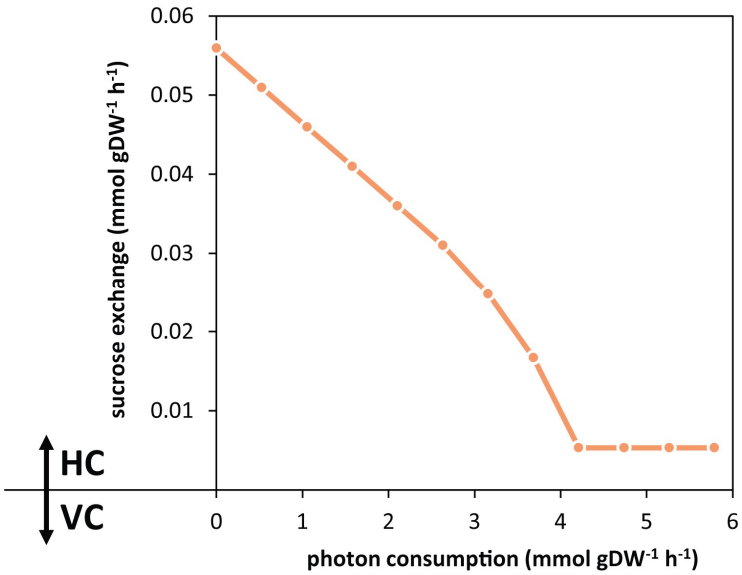


Figure 3.14. Sucrose exchange as a function of light harvesting by the HCSC. Photon consumption by the HCSC was varied at fixed VCSC light harvesting under diazotrophic conditions. In case of positive values, sucrose is transported to the HCSC, as indicated by the black labels and arrows to the left (HC: heterocyst super-compartment, VC: vegetative cell super-compartment).

Upon varying photon uptake by the HCSC, sucrose exchange between the two super-compartments responded remarkably (Figure 3.14). This response was inversely proportional to light availability, decreasing sucrose uptake by the HCSC as forced photon consumption increased. This behaviour suggested that sucrose coming from the VCSC was used as a source of energy and electrons in the HCSC. As a consequence, in cases when sucrose was not exchanged in Figure 3.7A and yet growth rates were higher than that of the literature case, the model has presumably selected exchange metabolites that could have been converted stoichiometrically more efficiently than sucrose. When the HCSC used sucrose as the source of electrons and carbon, however, glutamine, glutamate and 2-

oxoglutarate were selected by the model as the best combination for growth, in a good agreement with the literature (Wolk et al., 1976; Thomas et al., 1977; Schilling and Ehrnsperger, 1985; Böhme, 1998; Martin-Figueroa et al., 2000; Picossi et al., 2005; Cumino et al., 2007). Notably, the rate of sucrose transport was closer to reported figures (Nürnberg et al., 2015) when the exchange of glutamine to glutamate was fixed to a 1:1 ratio.

The two-cell model was also used to evaluate the yield of different nitrogen-containing compounds when excreted under diazotrophic conditions. The goal was to determine which compound would be best for excretion *in vivo*. Interestingly, nitrogen yield based on nitrogen uptake and release fluxes was quite similar for every compound. In contrast, the molar ratio of carbon and nitrogen released was highly variable based on the chemical composition of each compound. On a molar basis, ammonia and urea were found by far the highest yielding, although one mole of carbon was also lost for every two moles of nitrogen excreted in case of urea (Figure 3.11).

4 Engineering *Anabaena* sp. PCC 7120 for ammonia excretion

4.1 Introduction

The potential utilisation of modified cyanobacteria as efficient biofertilizers makes these organisms promising candidates for diazotrophic nitrogen excretion. In particular, the filamentous heterocyst-forming *Anabaena* sp. PCC 7120 has been studied at great detail as the model of photosynthetic nitrogen fixation. In addition, several molecular tools have been developed for or tested with this organism, a very important aspect for metabolic engineering. Moreover, *Anabaena* sp. PCC 7120 expresses a single GS, the enzyme responsible for rapid assimilation of ammonia. Thus, metabolic engineering efforts need to affect only this one enzyme, in order to increase levels of free ammonia inside the cell. As suggested by predictions using the stoichiometric model (Figure 3.11), ammonia is the best form of combined nitrogen for excretion, as well as the preferred source of nitrogen for most microorganisms (Herrero et al., 2001). However, ammonia, but not ammonium, can cross the cell membrane via simple diffusion, and therefore, the rapid action of GS has been suggested to prevent the loss of this precious compound, especially under diazotrophic conditions (Luque and Forchhammer, 2008). Furthermore, cyanobacteria express the high-affinity ammonium uptake transporter Amt proteins to scavenge for leaked ammonia, to further decrease the loss of fixed nitrogen (Vázquez-Bermúdez et al., 2002; Paz-Yepes et al., 2008).

All the work below was carried out in both wild type *Anabaena* sp. PCC 7120 and an *amt41B* knockout strain (Paz-Yepes et al., 2008) kindly provided by Professor Enrique Flores. The Δ*amt* mutant strain

lacks the complete *amt* cluster responsible for the uptake of ammonium[2] at low concentrations that is also necessary for the recapture of ammonia lost via diffusion. Thus, such strains cannot transport the protonated form; although ammonia can still diffuse freely through their cell membrane driven by the concentration gradient (Kleiner, 1981). The equilibrium of the two forms of ammonia nitrogen (NH_3-N) is governed by the respective pH, on either side of the cell boundary (Figure 4.1).

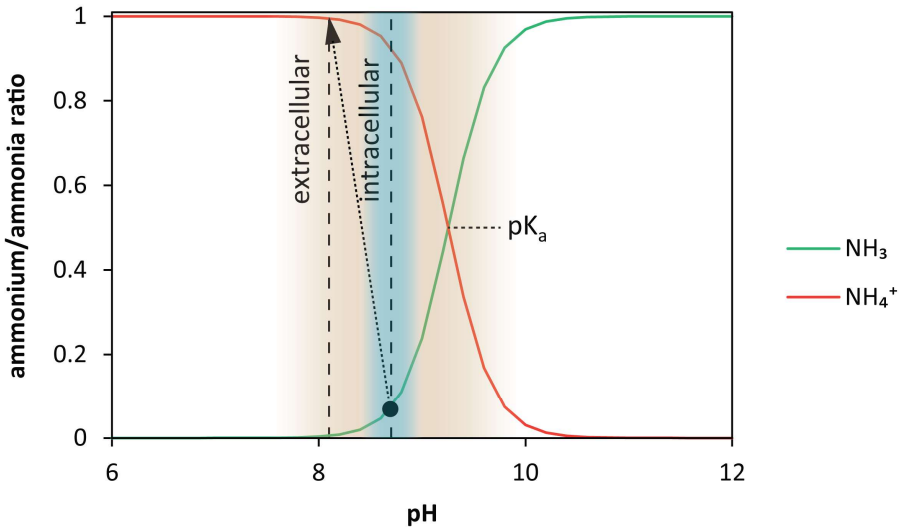


Figure 4.1. Equilibrium of ammonia and ammonium as a function of pH. Dashed vertical lines represent typical values of intracellular and extracellular pH in a batch culture in standard BG-11 medium, whereas blue and brown areas display possible ranges of fluctuating intracellular and extracellular pH, respectively. The dotted arrow indicates the protonation of leaked ammonia at lower extracellular pH.

Between pH 7.5 and 10, the pH optima of the majority of cyanobacteria (Brock, 1973), the protonated form is dominant at the lower end of the pH range, with gradual transition to the dissociated form (NH_3) through pH 9.25, the pK_a of ammonia. Since, below this point, the internal pH of a cyanobacterial cell is higher (more alkaline) than the external by values between 0.6 and 1.1 (±0.1) pH units (Giraldez-Ruiz et al., 1997), any intracellular ammonia that escapes the cell converts ultimately to ammonium (Figure 4.1). At the elevated ambient pH of 8.1, at least 7% of intracellular NH_3-N will be in the deprotonated form. Some of this 7% can leak out driven by diffusion where it will become protonated at the external side of the cell membrane due to more acidic pH conditions. At this lower pH, more

[2] In this work, the *amt* cluster proteins were considered to transport only the protonated form NH_4^+ as indicated by a large body of molecular data. However, a systems biology analysis suggested that an alternative mechanism may be operative in prokaryotes (Boogerd et al., 2011).

than 99% of the leaked NH_3-N is protonated and therefore, remains inaccessible to a Δ*amt* transporter mutant. Raising the extracellular pH increases the ratio of free ammonia to ammonium as well, up to a point where the deprotonated form becomes dominant above the pK_a (pH 9.25). Under such pH conditions removal of the *amt* cluster is not effective anymore in "locking out" NH_3-N of the cell. Therefore, a more viable strategy for ammonia excretion at higher external pH is to raise the overall intracellular pool of free ammonia. To design such strategies it is necessary to know the players affecting the concentration of intracellular ammonia.

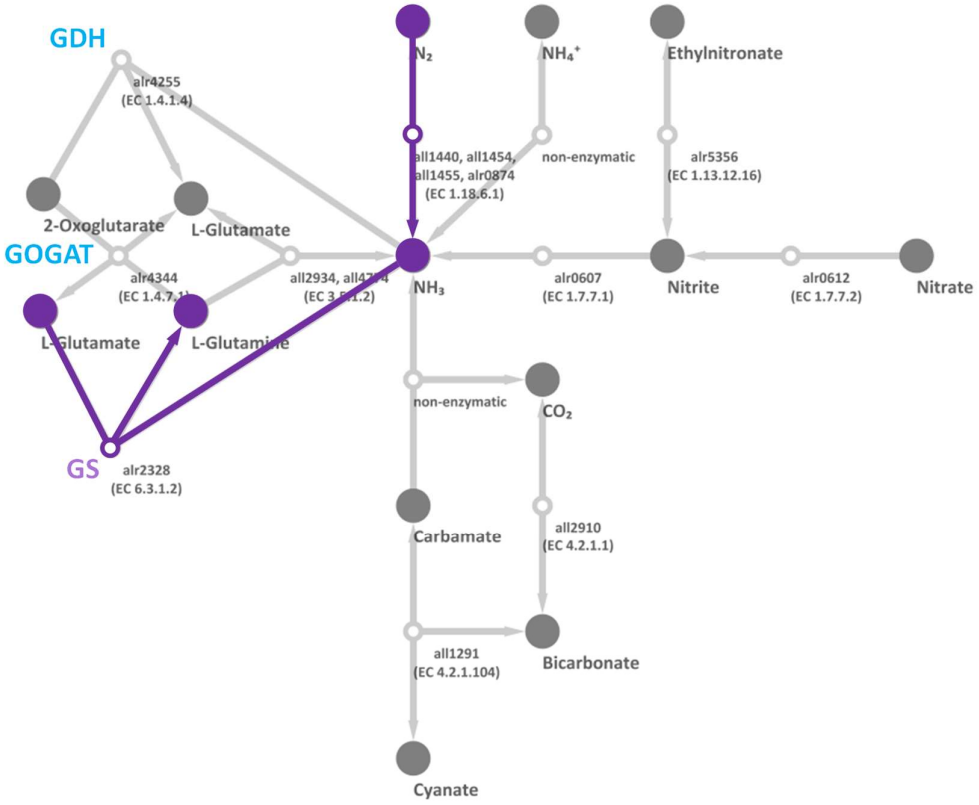


Figure 4.2. Nitrogen metabolism in *Anabaena* sp. PCC 7120. The central role of ammonia is indicated by the many metabolic interactions (grey lines representing biochemical reactions) it is involved in with other compounds. Large dark grey disks represent metabolites, small light grey circles and connecting light grey lines represent biochemical reactions. Arrowheads on the lines indicate reaction direction *in vivo*. Ammonia concentration is governed by three enzymes labelled with blue and purple capital letters. The primary assimilator of ammonia, GS, and the metabolic route of nitrogen fixation is highlighted in purple. The image is an extract of the graphical map of the complete metabolic network of *Anabaena* sp. PCC 7120 (Figure I-1).

Figure 4.2 summarizes the nitrogen metabolism of *Anabaena* sp. PCC 7120 with relevant enzymes and the genes encoding them. The network clearly shows the importance of ammonia as the central hub of nitrogen metabolism. Basically, all inorganic nitrogen sources are converted to ammonia first, before being incorporated into carbon skeleton. Three main enzymes govern the intracellular concentration of ammonia, GDH, GOGAT and GS. It is worth noting, that only GS and GDH are expressed in the heterocyst, and therefore this cell type lacks a functional GS–GOGAT cycle (Martin-Figueroa et al., 2000). Although GDH may also convert ammonia under certain environmental conditions, the primary assimilator of ammonia is GS (Luque and Forchhammer, 2008). The inhibition of this enzyme has been shown to result in increased intracellular ammonia and excretion of nitrogen in many organisms (Shanmugam et al., 1978). A similar effect was expected for *Anabaena* sp. PCC 7120, and therefore, effect of partial inhibition of GS in this cyanobacterium was investigated.

4.2 Results

4.2.1 Glutamine synthetase inhibition

Intracellular ammonia is rapidly incorporated into glutamate by glutamine synthetase in cyanobacteria (Luque and Forchhammer, 2008). In the presence of L-methionine DL-sulfoximine (MSX), a GS-specific glutamate analogue, however, the enzyme can be irreversibly inactivated (Ronzio et al., 1969), leading to the build-up, and eventually, the release of intracellular ammonia in *Anabaena* species (Thomas et al., 1990; Singh and Tiwari, 1998). Panel A in Figure 4.3 shows the results of MSX treatment of photodiazotrophically grown *Anabaena* sp. PCC 7120 strains. Cultures were grown to high density (OD_{730} > 1) in standard BG-11 medium, harvested by centrifugation, washed three times in BG-11_0 nitrogen-free medium and incubated further in BG-11_0 for up to 5 days in the presence and absence of MSX. As shown by the purple bars it was possible to produce considerable amounts of NH_3-N by the treatment of both the wild type and the Δ*amt* mutant with 55 µM MSX. In the absence of the GS inhibitor, however, there was no ammonia detected in the cultures, not even in case of the Δ*amt* strain. In addition, the total growth of untreated cultures measured as culture density at 730 nm was about 3 to 4-times higher than that of the treated samples. Oddly enough, growth of the Δ*amt* strain was less affected by MSX than the wild type, while ammonia produced by the mutant at the end of the 5 days was nearly twice as much as for the wild type.

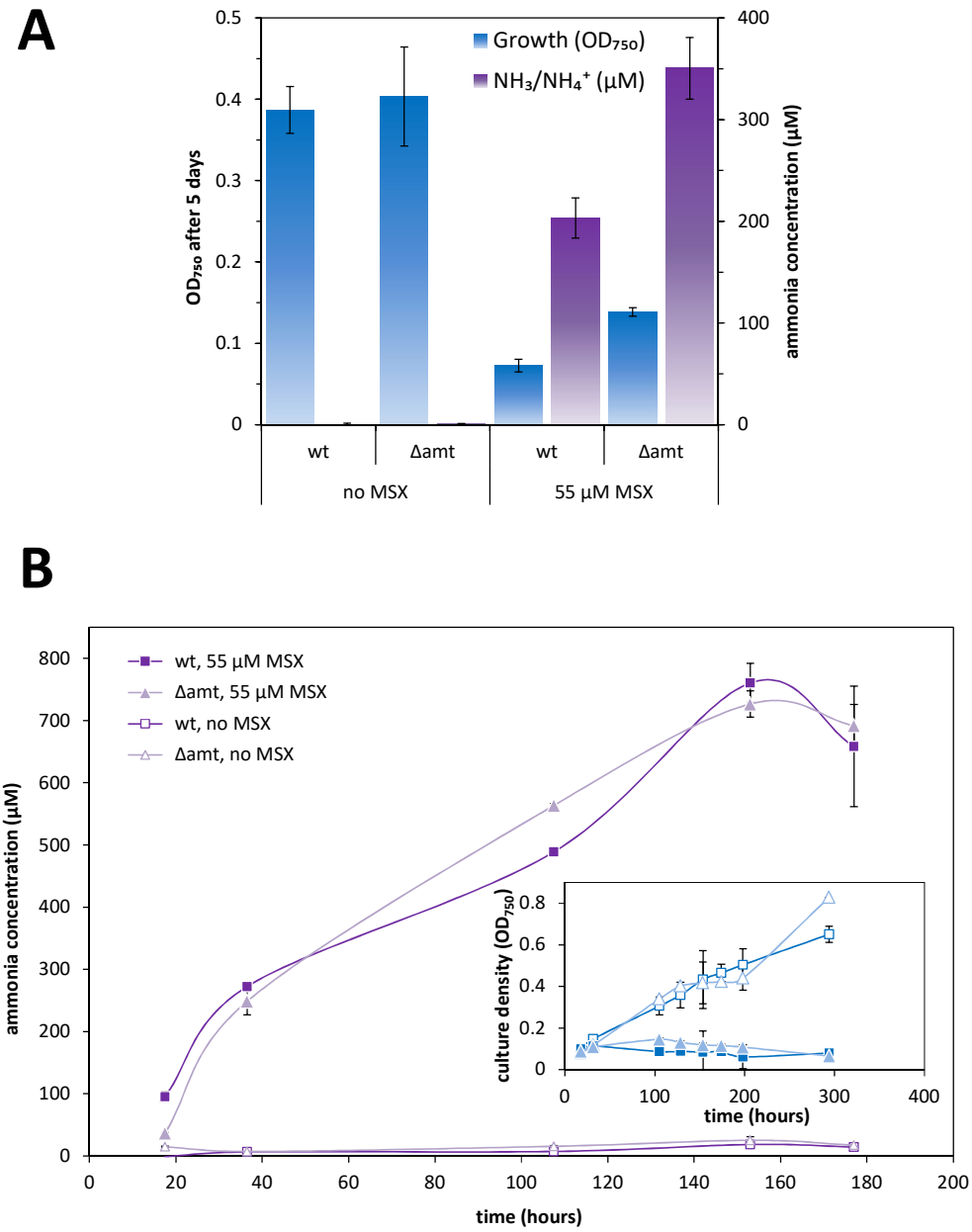


Figure 4.3. Ammonia excretion of wild type and Δ*amt* strains due to treatment with 55 µM MSX. (A) Blue bars show culture density and purple bars represent total ammonia concentration after 5 days of incubation under diazotrophic conditions. **(B)** Time course of ammonia accumulation and culture density on nitrate in the presence and absence of 55 µM MSX. Filled squares and triangles mark wild type and Δ*amt* in the presence of 55 µM MSX, respectively. Empty markers represent the controls with no MSX added. Purple and blue colours denote for ammonia concentration and OD_{730}, respectively. Error bars depict ±1 standard deviation of 3 biological replicates. Ammonia was assayed from sample supernatants using the Abcam Ammonia Assay Kit (modified Berthelot test, ab102509, Abcam Plc., Cambridge, UK).

These results confirmed that reduction of GS activity allows the production of extracellular NH_3-N at the expense of growth. However, MSX levels may drop over time due to possible degradation of the inhibitor. Continuous expression of GS may deplete the medium of the inhibitor and recapture of the lost ammonia eventually overcomes diffusion-driven leakage, as suggested by panel B in Figure 4.3. After 153 hours the concentration of external ammonia started to decline in the case of the wild type, while the culture density also increased. A similar effect was observed for the mutant strain as well, except that external ammonia was consumed at a much lower rate, probably due to the lack of ammonium transporters. In order to maintain GS activity at a lower-than-normal level, three different genetic engineering strategies were considered and evaluated.

Common to all three strategies was the idea to use the native regulatory system of *Anabaena* sp. PCC 7120 and increase the amount of free ammonia by decreasing the enzymatic activity of GS. Figure 4.4 summarizes the three methods and their connection to the regulatory network around GS.

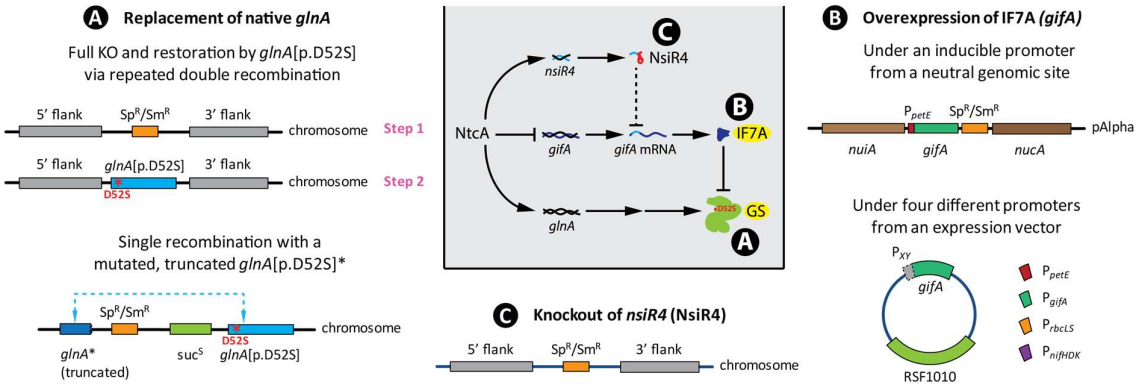


Figure 4.4. Overview of three metabolic engineering strategies for extracellular ammonia accumulation. (A) The active-site mutant GS (encoded by *glnA*) was introduced to the host strains either by two repeated double recombinations or by single recombination followed by an intrachromosomal rearrangement. In both cases a suicide vector was used. **(B)** The GS inactivation factor, IF7A (encoded by *gifA*) was expressed from either a chromosomal neutral site or a low copy expression vector under the effect of four different promoters. **(C)** The posttranscriptional regulator of IF7A, the sRNA NsiR4 was knocked out by double recombination using a suicide vector. The regulatory network connecting the three strategies with GS are summarized in the grey box.

First, GS was mutated by introducing a single amino acid change to its active site, and rendering it less active in the assimilation of ammonium. The change was performed on an aspartate residue at the fifty-second amino acid position (D52) for serine (Ortiz-Marquez et al., 2014). The D52 residue is responsible for binding and deprotonation of ammonia and a mutation to serine (D52S) has been

shown to decreases GS activity by 75% (Crespo et al., 1999). As a second strategy, the native inactivation factor of GS, IF7A was selected for overexpression to decrease the activity of GS at the protein level (Galmozzi et al., 2010). Third, a recently discovered small regulatory RNA (sRNA), *nsiR4* was considered as a knockout target. The sRNA *nsiR4* has been shown to repress the expression of GS inactivation factors (IF7 and IF17) in *Synechocystis* sp. PCC 6803 (Klähn et al., 2015). Although not yet explicitly proven, due to the relatively high homology between the *Synechocystis* IF7 and IF7A it was assumed that *nsiR4* in *Anabaena* will have a similar effect on IF7A, and therefore a knockout of the sRNA will (partially) remove the control over IF7A and thereby decrease GS activity. Figure 4.4 provides an overview of the three approaches discussed (Figure 1.4 on page 43 also summarizes the interplay between the corresponding metabolic and regulatory networks).

In the following sections ammonia and ammonium will be used interchangeably, unless it is essential to make distinction of the two forms. In such cases the distinction will be clearly stated.

4.2.2 Strategy A – Replacement of the *glnA* gene (*alr2328*) by an active-site mutant

It has been shown in Figure 4.3A that cultures supplemented with the GS drug MSX can achieve considerable ammonia excretion. The effect of MSX, however, fades away over time probably due to slow consumption of the drug, and the level of free extracellular ammonia eventually begins to drop (Figure 4.3B). MSX binds to GS irreversibly by forming a covalent bond with the enzyme (Ronzio et al., 1969). However, expression of GS is only dependent on the nitrogen status of the cell (regulated by the transcription factor NtcA), and thus, it may reach a state where all MSX is covalently bound to GS molecules synthesized previously. In such case, all newly formed GS will be uninhibited.

Technically, a strain with an unaltered expression level but decreased overall activity due to irreversible inactivation of GS is expected to display similar characteristics in terms of GS activity to a mutant strain with impaired specific activity of the same enzyme. In the case of the highly conserved GS it has been shown that residues D50, D49 or D51 (an aspartate at closely the same amino acid position) confer the ammonium ion binding ability of GS in *Escherichia coli* (Liaw et al., 1995), *Azotobacter vinelandii* (Ortiz-Marquez et al., 2014) or *Anabaena azollae* (Crespo et al., 1999), respectively. Moreover, strains bearing a mutant GS (D49S) that exhibits decreased specific activity to ammonia have been shown to natively accumulate extracellular ammonia in *Azotobacter vinelandii* (Ortiz-Marquez et al., 2014). Among the five mutant enzymes (D51S, D51A, D51E, D51N and D51R) described for *Anabaena azollae*, replacement of Asp51 for Ser51 (D51S) showed the highest specific activity, about 25% of the wild type GS (Crespo et al., 1999). In this work, the equivalent of Asp51 in

Anabaena azollae and Asp49 in *Azotobacter vinelandii* was mutated in *Anabaena* sp. PCC 7120 by site-directed mutagenesis to a serine residue, in order to retain GS activity at a relatively high level, and omit severe growth deprivation due to the lack of essential enzyme activity. Based on a BLASTX sequence similarity search (Altschul et al., 1990) against *Anabaena azollae* and *Azotobacter vinelandii* GS sequences the Asp52 residue of the *Anabaena* sp. PCC 7120 *glnA* gene was identified as the target mutagenesis site. The desired mutant genotype containing a serine at the 52nd amino acid position will be henceforth denoted as *glnA*[p.D52S].

4.2.2.1 Cloning strategies for replacing wild type glnA *for* glnA*[p.D52S]*

Glutamine synthetase in cyanobacteria acts as the primary assimilator of nitrogen (Luque and Forchhammer, 2008). Thus, its metabolic function converting L-glu and ammonia into L-gln is essential and most cyanobacteria carry more than one gene to encode for the same function (*glnA* and *glnN* in *Synechocystis* sp. PCC 6803 for example). In *Anabaena* sp. PCC 7120, however, *glnA* is the only gene (Reyes and Florencio, 1994) and therefore, a mutant strain lacking the functionality of GS may not be viable unless supplemented with appropriate amounts of glutamine. Indeed, gene essentiality simulations using the *Anabaena* sp. PCC 7120 stoichiometric model revealed the extreme importance of this enzyme for growth (darker red dot in Figure 3.10). In this work, two different cloning strategies were designed and evaluated for the replacement of wild type *glnA* for the active-site mutant *glnA*[p.D52S]. Figure 4.5 describes the three constructs used in the two approaches to introduce the desired genetic modification.

The first approach involves complete disruption of GS activity by knocking in the *aadA* cassette (brown arrow, Figure 4.5A). Although the plasmid contained the *nptII* gene (conferring resistance to Km and Nm, dark blue arrow) selection was strictly performed on appropriate amounts of Sm and Sp (*aadA*). In the first step the host strains (*Anabaena* sp. PCC 7120 wild type and Δ*amt*) were transformed with plasmid P_{glnA}-5'*glnA*[p.D52S]-*aadA*-3'*glnA*-T_{glnA}::pK19mob*sacB* (construct C7). The insert in this plasmid contains approximately 1-kb homologous regions flanking the Sm/Sp resistance cassette (*aadA*). In case of a successful double recombination event the whole insert will be introduced to the chromosome at the *glnA* site replacing the native gene. The resulting strain (denoted as *glnA*::*aadA*) lacks any essential GS expression and therefore, requires continuous supplementation with glutamine. It is worth noting that at this stage the disrupted *glnA* is already the mutant (yet dysfunctional) version (pink line in 5'glnA* fragment in Figure 4.5A). The second step transforms the *glnA*::*aadA* strains with the construct in Figure 4.5B (construct C8) and restores GS activity via a second double recombination.

The restored GS contains the desired D52S mutation and thus exhibits lower enzymatic activity towards ammonium.

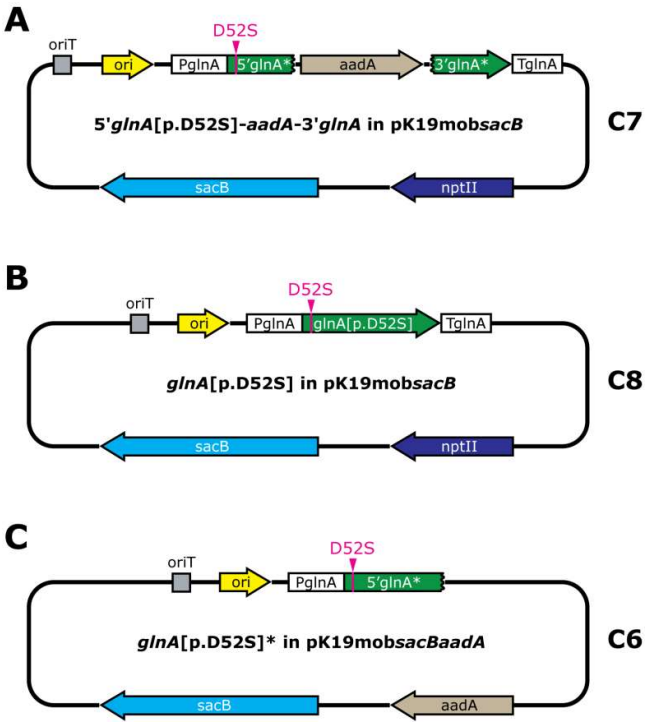


Figure 4.5. DNA constructs assembled to replace *glnA* for *glnA*[p.D52S]. (A) Knock-in construct for *glnA* disrupting its function by inserting a streptomycin/spectinomycin resistance cassette via double recombination. The construct contained both *aadA* and *nptII* for Sm^R/Sp^R and Km^R/Nm^R, respectively, but only *aadA* was used. **(B)** The construct restores the disrupted *glnA* gene to the active-site mutant *glnA*[p.D52S] by a second transformation and double recombination building on the outcome of **(A)**. **(C)** Single recombination construct inserting a truncated version of the mutant *glnA*[p.D52S] gene to the chromosome. Pink line shows the location of the D52S mutation in *glnA* and an asterisk (*) denotes for a truncated version of the gene. All constructs contained the *sacB* counter-selection marker, an *oriT* region required for mobilization and a replication origin (*ori*). A short name for each construct can be found to the right of their drawing.

The underlying difficulty of the above approach is the subsequent segregation cycles for both double recombinations. Complete genetic segregation of oligoploid *Anabaena* sp. PCC 7120 may be challenging on its own as the organism carries 8–10 copies of its genome, depending on the physiological status of the cell (Griese et al., 2011). In addition, disruption of an essential gene rendering the resulting strain a glutamine auxotroph may be a burden that a strain so heavily optimized for nitrogen assimilation cannot carry. Furthermore, the efficiency of double recombination is $< 10^{-7}$ in contrast to 10^{-5} for a single recombination (Cai and Wolk, 1990; Elhai et al., 1997), in other

words, less than 1% of all successful crossover events will result in a double recombinant. Thus, an alternative strategy involving a single recombination followed by an optional intrachromosomal recombination was also designed. In the single recombination approach host strains were transformed with a plasmid containing only the truncated and therefore dysfunctional version of *glnA* (Figure 4.5C, construct C6). In a single crossover event the complete plasmid was integrated into the genome of *Anabaena* sp. PCC 7120 retaining a functional *glnA* throughout the process. The clear advantage of this approach over the two subsequent double recombinations above is the need for laborious genetic segregation only once and that no auxotrophy is generated. The method is explained in more detail in Figure 4.6 below.

In case of successful transformation the plasmid P_{glnA}-*glnA*[p.D52S]*::pK19mob*sacBaadA* (construct C6) containing a truncated version of *glnA*[p.D52S] will target the homologous *glnA* site and integrate the whole plasmid in the first crossover (top of Figure 4.6). The resulting strain may contain a functional *glnA*[p.D52S] with the desired mutation (case A) or retain its wild type phenotype (case B). In both cases a truncated version of *glnA* (green boxes with an asterisk in their label) followed by all the other features on the original plasmid will be inserted upstream of the functional gene. It is possible to screen for the favoured scenario (i.e. case A) by colony PCR and sequencing using the allele-specific primer glnA.D52S*-F (76) and isolate the strain. Nevertheless, a spontaneous intrachromosomal recombination may remove the obsolete truncated version and the plasmid parts, leaving only one *glnA* of either version (mutant or wild type) seamlessly in the genome (bottom of Figure 4.6).

The second crossover to loop out the unnecessary plasmid parts can be induced by the addition of sucrose in a negative selection. A single copy of the *sacB* gene (blue arrow) converts the supplemented sucrose to a toxic product. Only those strains losing the plasmid and the *sacB* gene along with it can grow under such conditions. Regardless of their original genotype (case A or B) the resulting recombinants are left with either a mutant version of *glnA* (*glnA*[p.D52S] for case C) or revert back to wild type (case D). Genome sequencing over the affected region and colony PCR using the aforementioned allele-specific primer glnA.D52S*-F (76) can distinguish between the two outcomes. Primers and DNA constructs used are listed in Table 4.1.

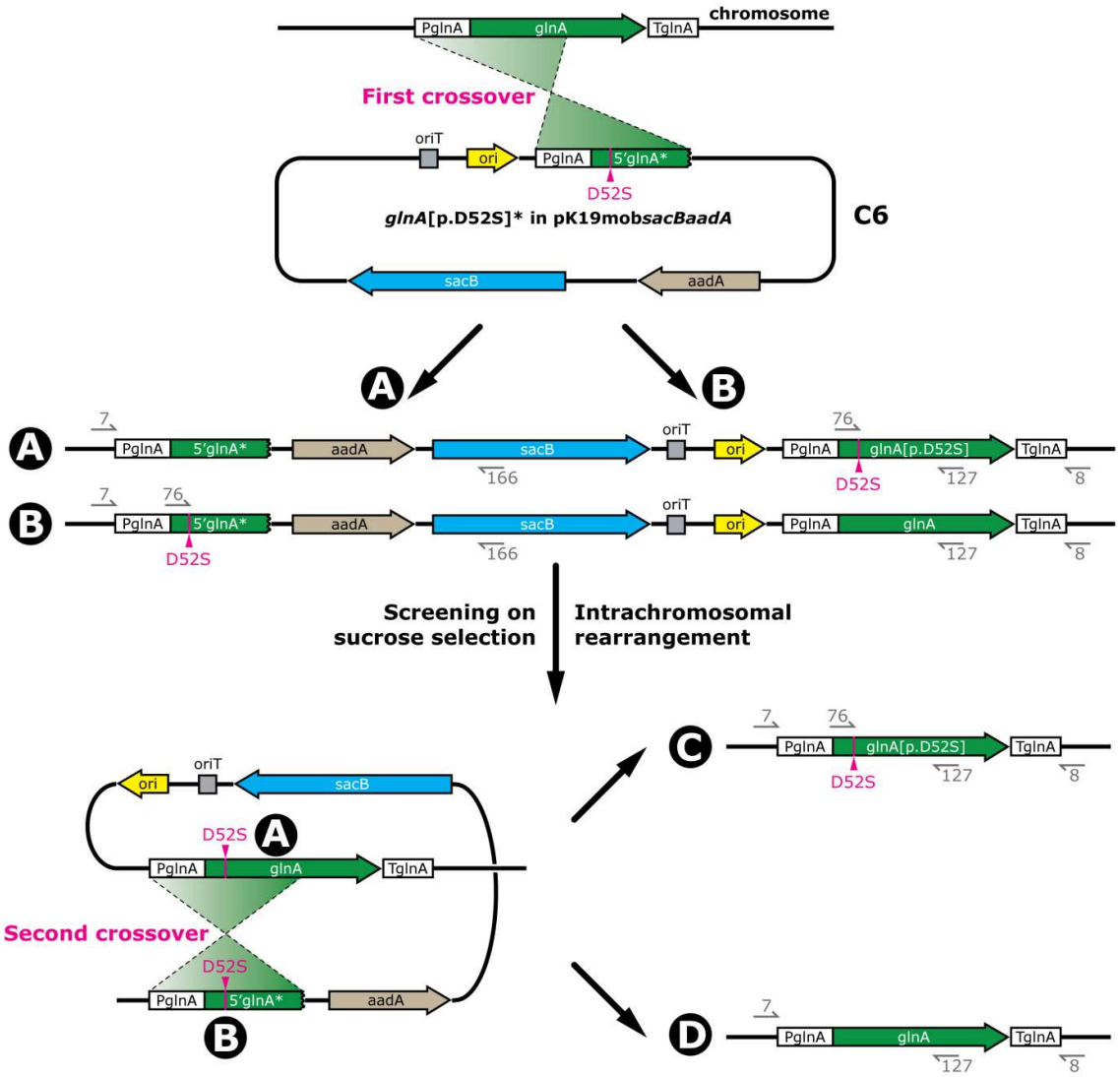


Figure 4.6. Replacement of *glnA* via single recombination followed by intrachromosomal recombination. In the first crossover the plasmid *glnA*[p.D52S]::pK19mob*sacBaadA* is integrated into the chromosome by recombining with wild type *glnA*. The resulting strains carry the full plasmid flanked by either a dysfunctional (truncated) wild type *glnA* and a functional *glnA*[p.D52S] (case A) or a truncated mutant *glnA* and an untouched wild type gene (case B). In a second crossover induced by negative selection on sucrose (mediated by the expression of *sacB*) both case, A and B can rearrange to either a fully functional *glnA*[p.D52S] (case C) or revert back to wild type (case D) by looping out the plasmid parts. The *glnA* mutation site is marked in pink. Truncated versions of *glnA* are marked with an asterisk (*). Grey arrows and numbers display binding sites for colony PCR primers.

4.2.2.2 Assembly of constructs

For technical reasons the construct in Figure 4.5B (construct C8) was prepared first as follows. Glutamine synthetase gene (*glnA* or *alr2328*) was amplified by PCR from high-purity genomic DNA of

Anabaena sp PCC 7120 wild type strain using primers 5'*glnA*-*Xba*I-F and 3'*glnA*-*Bam*HI-R (primers 40 and 41, respectively; kindly provided by Mr Anthony Riseley, University of Cambridge). Primers were designed to include regions 360-bp upstream and 272-bp downstream the coding sequence of *glnA* to serve as promoter and terminator, respectively. The PCR product was gel purified using a QIAquick Gel Extraction Kit (Qiagen Ltd., Crawley, UK), restriction digested by *Xba*I and *Bam*HI (NEB) and purified again by the same kit. The high copy number vector pUC18 was linearized using the same restriction enzymes and dephosphorylated to prevent self-ligation. The digested PCR product was ligated into the linearized vector and transformed to *E. coli* DH5α competent cells on selective LB-agar plates containing 100 μg/ml carbenicillin (Cb). Resulting colonies were picked into 5 ml liquid LB medium supplemented with the same antibiotics as above, grown overnight at 37 °C with shaking and plasmid purified using a QIAprep Spin Miniprep Kit (Qiagen Ltd., Crawley, UK).

The purified plasmid (P_{glnA}-*glnA*-T_{glnA}::pUC18) was used as a template in a site-directed mutagenesis reaction with primers *glnA*.D52S-F and *glnA*.D52S-R to introduce a single amino acid change form aspartate to serine at the amino acid position 52. Candidates were sequence-verified and the *glnA*[p.D52S] cassette was subcloned into the mobilizable, moderate copy number plasmid pK19mob*sacB* carrying *nptII* (conferring resistance to kanamycin and neomycin) and the *sacB* counter-selection marker (encoding for levansucrase, a sucrose sensitisation enzyme from *Bacillus subtilis*). Both donor and acceptor plasmids were digested with *Xba*I and *Bam*HI and digestion products were purified on 1% agarose gel. In addition, the donor pUC18 plasmid was also digested with *Bsa*I to make fragment identification by size easier on gel. The 2.6-kb pUC18 vector backbone was split to two 1.3-kb fragments to avoid confusion of the full length backbone with the 2.1-kb *glnA*[p.D52S] insert. The linearized acceptor plasmid pK19mob*sacB* was dephosphorylated, ligated with the insert and transformed to *E. coli* DH5α. Transformants were selected on LB-agar plates containing 50 μg/ml kanamycin, picked into 5 ml liquid LB medium supplemented with the same antibiotics (37 °C, overnight with shaking) and plasmid purified using a QIAprep Spin Miniprep Kit (Qiagen Ltd., Crawley, UK). The sequence of the purified plasmid (P_{glnA}-*glnA*[p.D52S]-T_{glnA}::pK19mob*sacB*) was verified via sequencing over the cloning site (insert).

Table 4.1. Oligonucleotide primers, DNA constructs and strains used in *glnA* gene replacement. Sequences are written from 5' to 3' direction. Restriction enzyme cut sites are highlighted in bold. Letters in small case denote for overhangs. All strains are based on *Anabaena* sp. PCC 7120.

Primers

ID	Name	Sequence
3	SmR-3'pK19mobsacB-F	atgaggatcgtttcgcATGAGGGAAGCGGTGATC
4	SmR-5'pK19mobsacB-R	accccagagtcccgcTTATTTGCCGACTACCTTGG
5	pK19mobsacB-3'SmR-F	GCGGGACTCTGGGGTTCG
6	pK19mobsacB-5'SmR-R	GCGAAACGATCCTCATCCTGTC
7	glnA_ext-F	TAAGCTTGTTACTGCATCGC
8	glnA_ext-R	AATCTATGCTGATGCGTTCC
127	glnA-cPCR-R	TGCCAAATAAGGGTTAGAGG
40	5'glnA-*Xba*I-F	GATC**TCTAGA**TTCCTTCTTCTCCCAATCTTTG
41	3'glnA-*Bam*HI-R	GATC**GGATCC**AACTCGGTCTTCCTGCTGAA
70	pK19mobsacB-BB-F	CACTGGCCGTCGTTTTAC
71	pK19mobsacB-BB-R	ATCATGTCATAGCTGTTTCCTG
74	glnA*-trunc-F	AGGAAGACCGAGTTGGATCC
75	glnA*-trunc-R	AACTTCGTGGTGATGTTTTTCG
76	glnA.D52S*-F	TCTGATGGCGTACCTTTTTC
77	5'glnA*-pK19mobsacB-F	ggaaacagctatgacatgatATTCCTTCTTCTCCCAATC
78b	5'glnA*-SmR-R	agccatgaaaTGGTGATGTTTTTCGATGG
79b	3'glnA-SmR-F	gtgagaatccCGAAGTTGCTACTGGTGGTC
80b	3'glnA-pK19mobsacB-R	tgtaaaacgacggccagtgcAGGAAAGATCGATCCCACC
81b	SmR-5'glnA*-F	aacatcaccaTTTCATGGCTTCTTGTTATGAC
82b	SmR-3'glnA-R	agcaacttcgGGATTCTCACCAATAAAAAACG

DNA fragments

ID	Description	Primers	Template
F13	pK19mob*sacB* backbone	70+71	pK19mob*sacB*
F14	*glnA*[p.D52S]* in pK19mob*sacBaadA*	74+75	pK19mob*sacB*::*glnA*[p.D52S] (C8)
F15	P_{glnA}-5'*glnA*[p.D52S]*	77+78b	pK19mob*sacB*::*glnA*[p.D52S] (C8)
F16	3'*glnA*-T_{glnA}	79b+80b	pK19mob*sacB*::*glnA*[p.D52S] (C8)
F17	Sm^R cassette (*aadA*)	81b+82b	pDF-lac
F18	*glnA*[p.D52S] in pK19mob*sacB*	40+41	wild type *Anabaena* sp. PCC 7120

Constructs

ID	Insert	Assembled from	Genotype
C6	*glnA*[p.D52S]*	F14	*glnA*[p.D52S]*::pK19mob*sacBaadA*
C7	5'*glnA*[p.D52S]-*aadA*-3'*glnA*	F15, F16, F17, F13	pK19mob*sacB*::5'*glnA*[p.D52S]-*aadA*-3'*glnA*
C8	*glnA*[p.D52S]	F18	pK19mob*sacB*::*glnA*[p.D52S]

Strains

Name	Purpose	Genotype
6w	replace *glnA* for *glnA*[p.D52S] in wild type strain	Δ*glnA*::*glnA*[p.D52S], Sm^R, Sp^R
6d	replace *glnA* for *glnA*[p.D52S] in Δ*amt* strain	Δ*amt*, Δ*glnA*::*glnA*[p.D52S], Nm^R, Sm^R, Sp^R

In order to generate construct *glnA*[p.D52S]*::pK19mob*sacBaadA* in Figure 4.5C (construct C6) the construct in Figure 4.5B (construct C8) was prepared in vector pK19mob*sacBaadA* as well by restriction cloning, the same way as described above. The plasmid pK19mob*sacBaadA* was created from pK19mob*sacB* by replacing the *nptII* cassette for the *aadA* coding sequence (CDS, conferring resistance to streptomycin and spectinomycin) in a Gibson assembly reaction. The *aadA* CDS was amplified from plasmid pDF-lac (Table 2.1) using primers SmR-3'pK19mobsacB-F (3) and SmR-5'pK19mobsacB-R (4), overlapping with primers pK19mobsacB-3'SmR-F (5) and pK19mobsacB-5'SmR-R (6), used for the linearization of the pK19mob*sacB* backbone without the *nptII* cassette (Table 4.1). The two fragments were fused in an isothermal assembly reaction (see section 2.5.8 for details); the

resulting construct was cloned to *E. coli* DH5α and purified. The construct in Figure 4.5C was amplified from the relevant section of the purified P_{glnA}-*glnA*-T_{glnA}::pK19mob*sacBaadA* plasmid using primers *glnA**-trunc-F (74) and *glnA**-trunc-R (75). The linear product was phosphorylated in the following reaction: in a 50-µl final volume (adjusted by nuclease-free water) 5 µl sample DNA (up to 300 pmol 5' ends) was mixed with 5 µl 10× T4 DNA ligase buffer (NEB, contains ATP) and 1 µl T4 polynucleotide kinase (NEB). The reaction was incubated at 37 °C for 30 min, the enzyme was heat-inactivated at 65 °C for 20 min and the phosphorylated product was self-ligated (by the addition of T4 DNA ligase) in a 1-hour reaction at room temperature. The resulting construct can be described as P_{glnA}-*glnA*[p.D52S]*::pK19mob*sacBaadA* with an asterisk (*) denoting that it is a truncated version lacking the 3' end of the *glnA* gene and its terminator T_{glnA}.

Finally, construct P_{glnA}-5'*glnA*[p.D52S]-*aadA*-3'*glnA*-T_{glnA}::pK19mob*sacB* (construct C7 in Figure 4.5A) was assembled from four compatible fragments in a Gibson isothermal reaction. First, the 5' and 3' targeting arms flanking the *aadA* Sm^R cassette were amplified from genomic DNA of *Anabaena* sp. PCC 7120 using primer pairs 5'glnA*-pK19mobsacB-F (77) plus 5'glnA*-SmR-R (78b) and 3'glnA-SmR-F (79b) plus 3'glnA-pK19mobsacB-R (80b), respectively. Second, the *aadA* cassette was amplified with overlapping flanks from plasmid pDF-lac (Table 2.1) using primers SmR-5'glnA*-F (81b) and SmR-3'glnA-R (82b). Third, the plasmid pK19mob*sacB* was linearized to contain compatible ends for the 5' and 3' targeting arms using primers pK19mobsacB-BB-F (70) and pK19mobsacB-BB-R (71). PCR products were gel purified, assembled and transformed to *E. coli* DH5α. The desired plasmid was isolated from liquid cultures of the resulting colonies, and the assembly was verified by Sanger sequencing (Source BioScience, Nottingham, UK) over the insertion site.

4.2.2.3 Isolation of glnA*[p.D52S] active-site mutants*

The three constructs (C6, C7 and C8) in Figure 4.5 to replace wild type *glnA* for the active-site mutant *glnA*[p.D52S] were assembled as described above. In the first round constructs C6 and C7 were transformed to *Anabaena* sp. PCC 7120 wild type and the Δ*amt* strain by triparental conjugation. Exconjugants became visible on the filter after approximately two weeks of screening. Every third day the conjugation filter was transferred on a fresh BG-11 agar plate supplemented with the appropriate antibiotics. The C6 plates were supplemented with 2.5 µg/ml Sm and Sp, whereas C7 transformants were screened on 30 µg/ml Nm in the presence of 0.2% glutamine[3]. Plates for C7 also contained 5%

[3] Glutamine supplementation was originally set to 0.2%, a level used with *E. coli* glutamine auxotrophs (Bloom et al., 1978; Rothstein et al., 1980). The amount was later increased to 0.8% used with *Synechocystis* sp. PCC 6803 (Mérida et al., 1992).

sucrose (final concentration) from the third day to apply evolutionary pressure for the development of double recombinants (Cai and Wolk, 1990). Figure 4.7 summarizes the stages of mutant isolation in *Anabaena* sp. PCC 7120 wild type and Δ*amt* strains for construct C6, but the process was essentially the same for C7 as well. Green transformant colonies arisen (panel A and B) were picked from the filter with toothpicks and streaked onto selective agar plates for segregation (panels C and D). The segregation status was regularly monitored by colony PCR following subsequent rounds of sub-culturing.

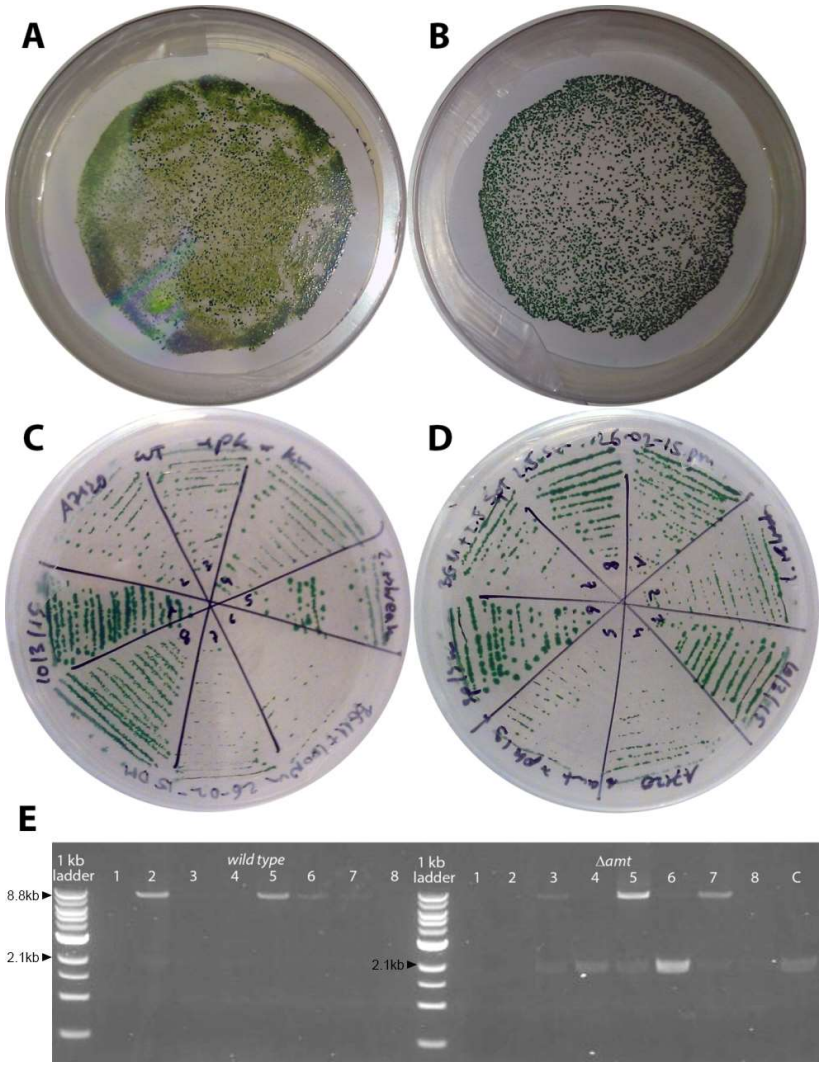


Figure 4.7. Isolation steps of single recombinants for construct C6. (A) and **(B)** show conjugation plates with growing exconjugants of *Anabaena* sp. PCC 7120 wild type and Δ*amt* strains, respectively. **(C)** and **(D)** illustrate passaging of wild type and Δ*amt*, respectively, for segregation. Panel **(E)** displays results of colony PCR for the 6th round of segregation. An 8.8-kb band represents successful genomic integration, whereas a 2-kb band indicates the presence of the wild type chromosome. The appearance of both bands suggests a mixed genotype. Fragment size was predicted using a 1-kb DNA ladder (NEB).

A couple of hundred green transformants appeared on conjugation filters for C6 after 3 weeks of selection, with the first clones becoming visible in about 8 days (Figure 4.7A and B for the wild type and Δ*amt*-based strains, respectively). Interestingly, the rate of background decay was considerably slower for the wild type, visible as a yellow-green background in panel A. In contrast, the non-conjugant background of the Δ*amt* strain died back gradually, disappearing almost completely by the time of the photo in panel B was taken. In the case of C7, only a few dozen exconjugants arose from the background. However, these initial colonies disappeared in the first two weeks of screening on antibiotic selection in the presence of glutamine. The conjugation process was repeated in the presence of four times increased glutamine (0.8%), but no transformants could be isolated. Due to the failure to isolate C7 double recombinants with a disrupted *glnA*, construct C8 could not be used to restore GS activity. Instead, the alternative approach for genomic integration of C6 via single recombination (Figure 4.6) was continued with segregation to clean the mutant genotype of any wild type contamination.

4.2.2.4 Genetic segregation

Conjugant green colonies arisen on the conjugation filter were picked with toothpicks and streaked as straight lines onto one of eight sections on a BG-11 selective agar plate supplemented with 2.5 µg/ml Sm and Sp (Figure 4.7C and D). A loopful of green biomass was collected from multiple lines of each section, washed into 20 µl nuclease-free water and freeze-thawed five times at alternating -80 °C and 60 °C, using an ultra-low temperature freezer (-80 °C) and a thermocycler, respectively. The lysate was spun down in a benchtop mini centrifuge (Sprout, Heathrow Scientific, USA) and colony PCR was carried out from 2 µl of the supernatant as template. Primers glnA_ext-F (7) and glnA_ext-R (8) amplifying from genomic locations closely flanking the insertion site were used to detect a single recombinant (8.8 kb) over the wild type (2.1 kb). An allele-specific primer glnA.D52S*-F (76, Figure 4.8) with glnA-cPCR-R (127) was also used to amplify either a 1-kb (case A in Figure 4.6) or a 7.7-kb product (case B in the same figure). Similar oligonucleotide designs to distinguish between wild type and mutant genotype have been used successfully by others (Gaudet et al., 2009; Ortiz-Marquez et al., 2014). Binding sites for colony PCR primers are indicated as grey arrows in Figure 4.6.



Figure 4.8. Binding site of the allele-specific primer glnA.D52S*-F. The oligonucleotide (76) is specific to the mutation site due to the two bases at its 3' end not priming with the wild type sequence (red circle). Mutation site is highlighted in pink on the mutant DNA (top) and in green on the wild type DNA (bottom). Genomic DNA sequence is written in 3'–5' direction, whereas primer sequence is oriented in 5'–3' direction.

Results of typical colony PCR tests over the course of about 8 months are shown in Figure 4.9. Products of PCR were visualized on 1% agarose gels in an electrophoresis cell at 110 V for 30–40 min and photographed in a gel imager under UV light (GelDoc-It, UVP, Cambridge, UK). Altogether sixteen transformants were picked from both conjugation plates (for wild type and Δ*amt*) and all tested in colony PCR.

Panel A in Figure 4.9 collects the results of the first four candidate colonies for *glnA*[p.D52S] in both wild type (6w1–4) and Δ*amt* (6d1–4). Bands on lanes 1–8 were amplified using the mutant-specific primer set (76+127). The extension time was set to a value optimal for a 1-kb product specific to the favoured orientation as in case A, in Figure 4.6. In this scenario the integration of the plasmid disrupts wild type *glnA*, but at the same time also restores another copy to a mutant *glnA*[p.D52S]. The desired 1-kb product was visible for all candidate colonies with stronger bands for the Δ*amt*-based samples, most likely due to more optimal parameters for PCR (more DNA template, fewer inhibitors and cell debris). The right side of panel A displays bands characteristic to the amplification of the entire insertion site. The absence of any 2.1-kb fragment for 6w1–4 (lanes 9–12) may indicate failure of the PCR, although there was enough DNA for the amplification with 76+127 on the left side. Another option is that these strains were fully segregated single recombinants (band size would be 8.8 kb), although this was unlikely at that stage. In addition, wild type bands of 2.1 kb appeared for all Δ*amt* candidates, as expected (lanes 13–16). Several rounds of restreaking were carried out with no significant improvement of the results in colony PCR. In order to enhance the time-consuming segregation process samples were treated with sonication to break up long filaments and facilitate separation of the two genotypes (Cai and Wolk, 1990).

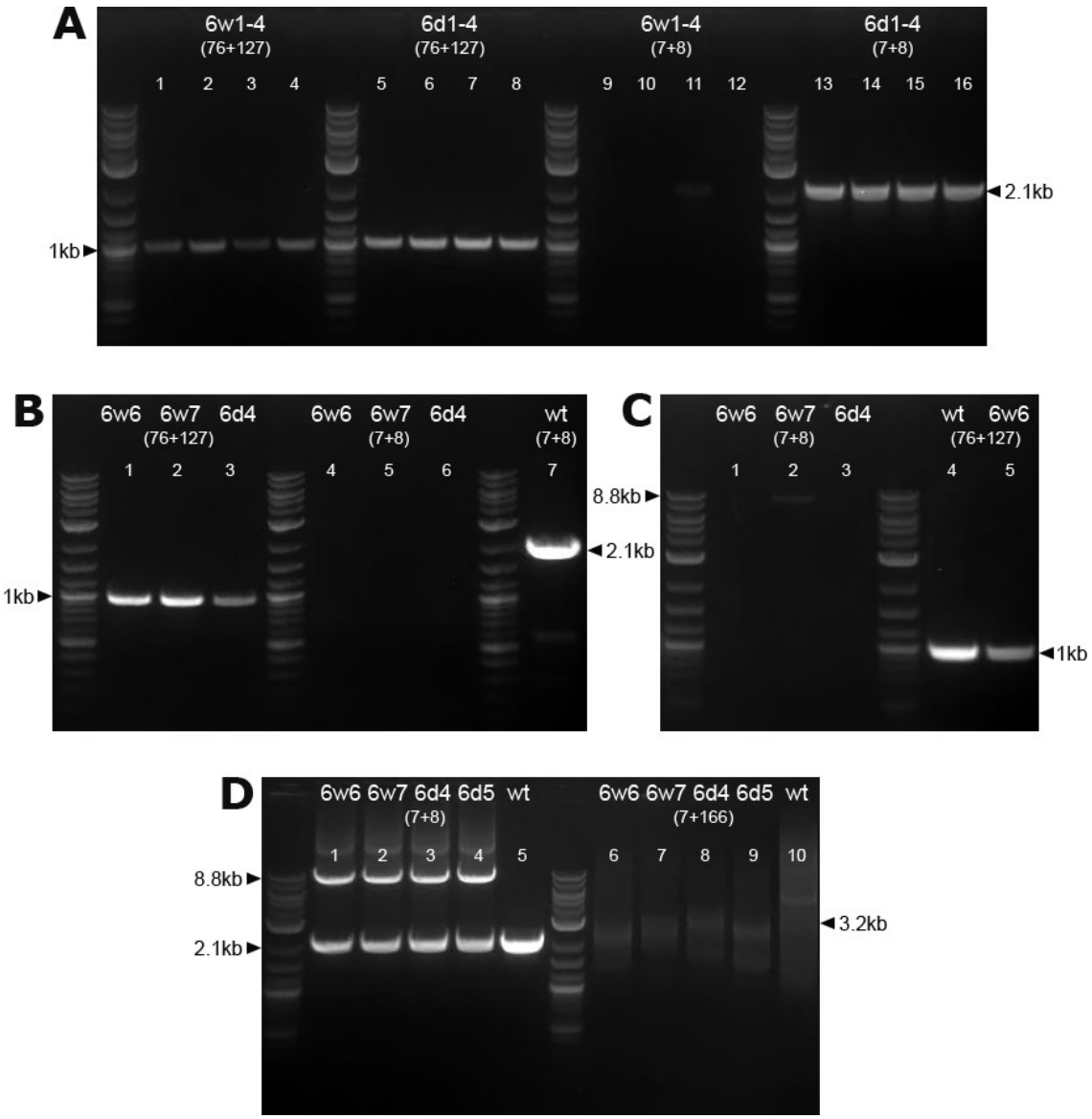


Figure 4.9. Colony PCR results at different stages of genetic segregation. (A) Four candidate colonies of both C6 strains (wild type and Δ*amt*-based) were evaluated using a set of primers specific for the active-site mutation (76+127) and with another set amplifying from just outside the insertion site (7+8). For primers 76+127 a 0.99 kb product is expected from a mutant with no band from the wild type. In case of primers 7+8 an 8.8-kb product indicates an insertion (single recombinant), whereas the wild type genotype results in a 2.1-kb fragment. **(B)** Three isolates of the C6 transformation in wild type (6w6 and 6w7) and in Δ*amt* (6d4) were tested. The same primer sets were used as in panel A (76+127 for lanes 1–3 and primers 7+8 for lanes 4–6). High-purity wild type genomic DNA was used as a control. **(C)** Repetition of the colony PCR in panel B with primers 7+8. Wild type genomic DNA was also included as a negative control with mutant-specific primers (76+127). Isolate 6w6 was added with primers 76+127 as a positive control. **(D)** Results of colony PCR after filament fragmentation by sonication. Isolates were tested with primers 7+8 and 7+166, the latter being specific to parts of the plasmid backbone (a 3.2-kb product indicates presence of *sacB* at the insertion site). In all cases the 2-log DNA ladder (NEB) was used as a marker to predict and compare the size of PCR products. The exact locations of colony PCR primers' binding sites are indicated as grey arrows and numbers in Figure 4.6.

Candidate colonies were inoculated into 1 ml liquid B-11 medium supplemented with 1.5 µg/ml Sm and Sp antibiotics, and grown under standard conditions (40 µE m^{-2} s^{-1}, 30 °C in 1% enriched CO_2) to an OD_{730} between 0.5 and 0.8. Samples were fragmented via cavitation in a sonication bath (F5 Minor; Decon Ultrasonics Ltd., UK). Undiluted cultures (500 µl in sterile 2 ml glass vials) were washed in fresh BG-11 medium before cavitation and treated for different lengths at room temperature. Status of filament fragmentation was monitored under a bright-field microscope (Zeiss Axioskop 2 Plus, Carl Zeiss Ltd., Cambridge, UK), as documented in Figure 4.10. Samples were treated for 4×2, 2×4, 4×4 and 2×8 min of which sonication for 2×8 min was found to be the most efficient to introduce filament breakage at every 3–6 cells. Fragmented samples were washed again in fresh BG-11 and the supernatant was transferred to BG-11 agar plates supplemented with the appropriate antibiotics. Appearing colonies were picked, streaked onto fresh plates and thereby returned to the segregation cycle.



Figure 4.10. Effect of sonication on fragment length of segregating *glnA* mutants. Bright-field microscopic images of *Anabaena* sp. PCC 7120 wild type transformed with construct C6. Samples from left to right were treated with 4×2 min, 2×4 min, 4×4 min and 2×8 min cavitation, respectively, in a sonication bath at room temperature. Faint, yellow-green cells have been disrupted by sonication. Larger cells are heterocysts. Magnification: 400×.

Surprisingly, a few of heterocysts could be observed in all samples treated with sonication. The appearance of heterocysts may have occurred due to depletion of the culture medium in combined nitrogen (nitrate in this case). It is also possible that occasional heterocysts are phenotypic indicators of a mutant *glnA*[p.D52S] due to a disturbed nitrogen homeostasis. Nevertheless, the heterocyst frequency was lower than normal (more than 20 vegetative cells between two heterocysts) and appeared to be irregular (Figure 4.10).

Panel B in Figure 4.9 shows results of a colony PCR following the re-isolation of *glnA*[p.D52S] candidates after sonication. Similar to panel A, PCR products could be detected with primers 76+127

only (lanes 1–3), but not with primers 7+8 (lanes 4–6). This is somewhat unexpected, as 7+8 should give a band from both wild type and mutants. Colony PCR was repeated on the same isolates (6w6, 6w7 and 6d4) from fresh lysate and a negative control was also included (wild type genomic DNA, wt, Figure 4.9C). A faint band for isolate 6w7 appeared at 8.8 kb (lane 2), which indicates a single recombinant with the full insert in place. However, the negative control (lane 4) also produced a band at 1 kb, which was unexpected.

On lane 5 isolate 6w6 was loaded as a positive control for 76+127, showing that the band for the negative control has the exact same size than that expected for the mutant-specific primers (76+127). Clearly, the band produced by the negative control (wt, lane 4) was of equal size. As indicated in Figure 4.8, however, PCR amplification from wild type *glnA* with primer glnA.D52S*-F (76) should not be possible, as the 3' end of this oligonucleotide does not anneal to the template. The reaction was repeated two more times at stringent annealing temperature (62 °C, NEB Tm Calculator v 1.9.7) with similar results. To avoid further confusion future segregation cycles were evaluated using primers 7+8 and 7+166 (Figure 4.9D).

Liquid cultures of candidates were sonicated again and strains were re-isolated on plates. After 13 cycles of segregation following the initial isolation from conjugation filters the 8.8 kb band indicating explicitly a single recombination appeared in colony PCR. Figure 4.9D displays bands for four isolates (lanes 1-4) and a wild type control (lane 5). All isolates produced bands at both 8.8 kb and 2.1 kb, characteristic to the single recombinant and the wild type, respectively. However, no bands were detected with the other set of oligonucleotides. Primers 7 and 166 are supposed to amplify a 3.2-kb fragment from just upstream the insertion site into the *sacB* cassette on the cargo plasmid. Although it was not possible to specifically detect parts of the integrated plasmid in colony PCR, the 8.8-kb product on lanes 1–4 clearly indicate the presence of the entire construct. In addition to the above, there was no amplification using 7+166 from the wild type control either, as expected. In conclusion, despite the difficulties to specifically detect the active-site mutation or parts of the cargo plasmid, it was possible to repeatedly confirm single recombinants in both base strains by colony PCR. In addition, filament fragmentation was required to improve segregation efficiency.

Although no fully segregated mutants had been detected at that point, candidates in Figure 4.9D were exposed to sucrose counter-selection to speed up the isolation process. Single colonies were picked and streaked onto BG-11 agar selective plates supplemented with the appropriate antibiotics and 5% sucrose as well. The concentration of the antibiotics was increased to 3.5 µg/ml for both Sm and Sp in order to enhance segregation. After one cycle candidates were evaluated in a colony PCR test

presented in Figure 4.11. Panel A collects results of an amplification using primers 7 and 8. In total, seven clones were assayed of wild type (lanes 1–7) and eight of Δ*amt* origin (lanes 8–15).

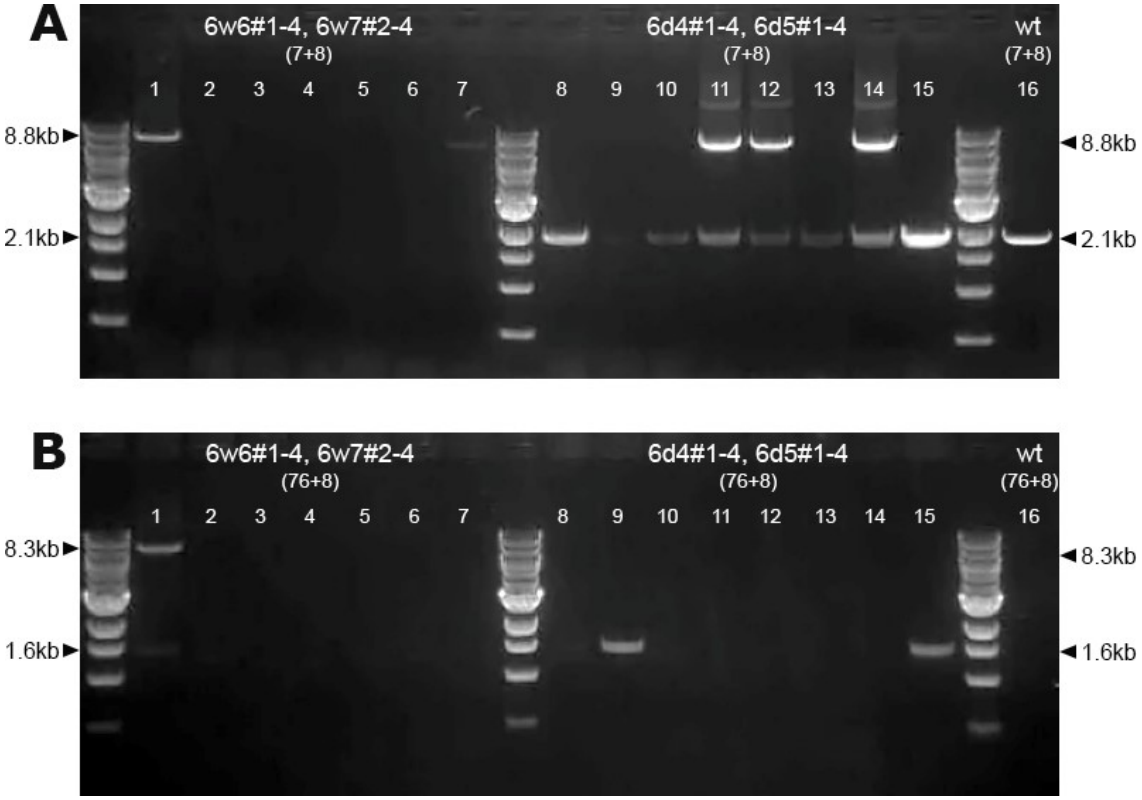


Figure 4.11. Colony PCR results after addition of sucrose to the selective medium. Candidates were grown on selective BG-11 agar plates supplemented with 3.5 μg/ml Sm and Sp, in the presence of 5% sucrose. After addition of sucrose colonies were restreaked twice over about a month and tested in colony PCR. **(A)** Four clones of the isolates in Figure 4.9B–D were evaluated using primers 7+8 amplifying from just outside the insertion site. A product of 8.8 kb indicates a single recombinant and a 2.1 kb product implies presence of wild type. **(B)** The same clones of the isolates in panel A were assayed using allele-specific primers 76+8. Both an 8.3-kb and a 1.6-kb band would suggest the presence of the desired mutation, although a 1.6-kb band indicates the preferred scenario (case A in Figure 4.6). For both panels wild type genomic DNA was used with the appropriate primers as control (lane 16). The exact locations of colony PCR primers' binding sites are indicated as grey arrows in Figure 4.6. Fragment sizes were predicted by comparing their bands to a DNA ladder (1 kb DNA Ladder, NEB).

High-purity genomic wild type DNA was used as control, giving a single band at 2.1 kb as expected. Candidates, on the other hand, produced all combinations of the two bands. In case of the wild type transformants a single 8.8-kb band was observed for isolates 6w6#1 and 6w7#4 (lanes 1 and 7, respectively), suggesting a fully segregated single recombinant genotype for these clones, although the band was faint for 6w7#4. The rest of the clones (6w6#2–4 and 6w7#2–3) gave no bands at all,

most likely due to the failure of colony PCR. All the other strains originating from Δ*amt* on lanes 8–15 produced a 2.1-kb band, even 6d4#2 on lane 9, for which the band was very faint. Moreover, the band at 8.8 kb, characteristic to single recombinants, was also detected for clones 6d4#4, 6d5#1 and 6d5#3. The presence of the 8.8-kb band for these clones together with 6w6#1 and 6w7#4 (on lanes 1 and 7, respectively) suggest that these strains retained the integrated plasmid, regardless of sucrose counter-selection. However, a single 2.1-kb band for 6d4#1–3, 6d5#2 and 6d5#4 suggests that the insertion site was restored in these strains by looping out the plasmid parts as explained for genotypes C and D in Figure 4.6.

The clones were further analysed using primers 76+8 in another colony PCR (Figure 4.11B). This oligonucleotide pair is expected to yield an 8.3-kb product from a strain possessing the genotype of case B in Figure 4.6, or give a 1.6-kb product in any other case (cases A, C and D in Figure 4.6). Although genotype D does not contain a binding site for primer 76 it has been shown earlier that an amplicon of the right size could be detected even from wild type DNA (Figure 4.9D). Interestingly, the dubious primer 76 this time resulted no band from wild type genomic DNA in opposition to Figure 4.9D. This time, however, primer 76 was used in pair with primer 8 that has only one binding site on a single recombinant template, in opposition to primer 127 (Figure 4.6). In addition, the calculated annealing temperature (62 °C) was increased to 65 °C for these amplifications to improve specificity of the primer pair 76+8 compared to 76+127. Regarding the mutant strains only clones 6w6#1, 6d4#1, 6d4#2 and 6d5#4 produced any bands (lanes 1, 8, 9 and 15 in panel B, respectively). The 8.3-kb product for 6w6#1 revealed that this clone most likely contains the full insert with a truncated *glnA*[p.D52S] and a functional wild type *glnA* (genotype B in Figure 4.6). There is also a very faint 1.6-kb band visible on lane 1 for this strain, suggesting that traces of either genotype C or genotype A from Figure 4.6 may be present as well. In contrast, a solid 1.6-kb band was amplified from 6d4#2 and 6d5#4. Although barely visible, the same band was present for 6d4#1 as well. Together with the results in panel A this strongly suggests that 6d4#2 and 6d5#4 (and also 6d4#1) are possessing the desired mutant genotype (genotype C in Figure 4.6) and can therefore be described as *glnA*[p.D52S] active-site mutants.

In summary, segregated single recombinants were successfully isolated for both 6w6 (6w6#1) and 6w7 (6w7#4), following the counter-selection on sucrose, although 6w6 may be of mixed genotype. Moreover, segregation may be complete for 6d4#1, 6d4#2, 6d4#3, 6d5#2 and 6d5#4 as well, of which 6d4#2 and 6d5#4 are likely to bear the desired mutation. All clones above were further assessed by determining ammonia accumulation in a 3-day growth assay.

4.2.2.5 Ammonia excretion

Candidate colonies for *glnA*[p.D52S] mutation in Figure 4.11 were picked from sucrose counter-selective plates into 1 ml BG-11 liquid medium supplemented with 1.5 µg/ml Sm and Sp and incubated up to four days under standard growth conditions (40 µE m^{-2} s^{-1}, 30 °C in 1% enriched CO_2) in 24-well microtitre plates. Samples were harvested by centrifugation at 4,000 × g for 10 min. Ammonia concentration was measured from the supernatant using the Willis micro-scale method. Figure 4.12 shows the results for the isolated strains and *Anabaena* sp. PCC 7120 wild type as control.

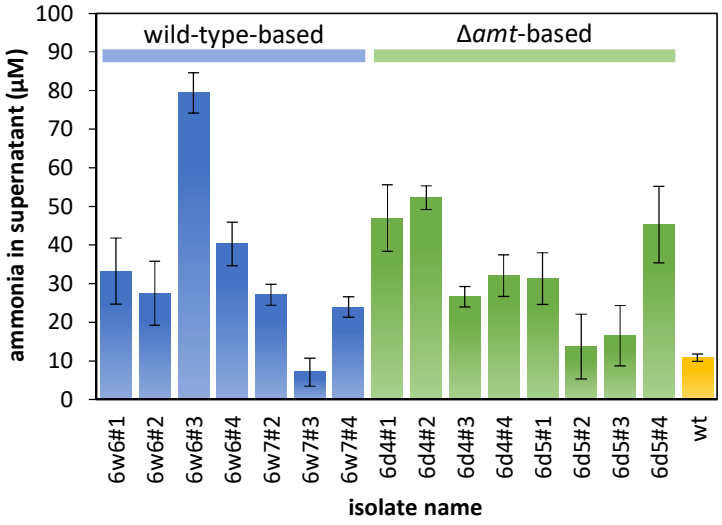


Figure 4.12. Free ammonia in the supernatants of GlnA active site-mutant isolates. Strains were grown in normal BG-11 containing $NaNO_3$ as the source of nitrogen for four days under continuous illumination by 40 µE m^{-2} s^{-1} cool white light in 30°C and 1% CO_2 atmosphere. Ammonia was measured from cell-free supernatants by the micro-scale Willis method. Error bars display ±1 standard deviation of the mean of two technical replicates.

It is worth noting that culture density was not measured this time and ammonia levels could not be normalized to the growth variations of the different strains. Thus, absolute concentration values in Figure 4.12 cannot be directly compared in order to select the best strain. However, the control culture was prepared at the same starting OD as the mutants and therefore, any improvement over this control represents a real difference. In the light of the above, all clones exhibited higher rates of ammonia accumulation than the control wild type, except for 6w7#3, 6d5#2 and 6d5#3. Concentration of ammonia was significantly higher in the supernatant of 6w6#3 (tallest blue column) than in any other sample. Although no bands could be detected for this clone in neither of the colony PCR tests in

Figure 4.11, it may be that this strain bears genotype C (Figure 4.6) similar to 6d4#2 and 6d5#4. In fact, these latter two showed the highest level of ammonia accumulation among the Δ*amt*-based strains (green columns). In addition, extracellular ammonia in the sample of 6w4#1 was measured at a similar level, suggesting that this clone may be genotype C, although the specific band in Figure 4.11B could not be detected. These results on ammonia accumulation provided further evidence that replacement of *glnA* for *glnA*[p.D52S] has been successfully completed in clones 6d4#2, 6d5#4 and maybe in 6w4#1, 6w6#3 and 6d4#1 as well.

The above isolates (6w6#1–4, 6w7#2–4, 6d4#1–4 and 6d5#1–4) were grown in 3 ml BG-11 containing 1.5 µg/ml Sm and Sp in 6-well microtitre plates under standard growth conditions for up to 8 days. Cells were harvested by centrifugation at 4,000 × g for 10 min, resuspended in 3 ml fresh BG-11 containing the appropriate antibiotics and further incubated up to 6 days. Cells were harvested again as above, resuspended in 1 ml fresh BG-11 and cryopreserved in the presence of 18% glycerol.

4.2.3 Strategy B – Overexpression of IF7A from a self-replicative vector

Due to the discussed difficulties during the isolation of segregated double recombinants of genomic integrations and the instability of the *sacB* counter-selection marker, a new strategy was designed for plasmid-based expression of the GS inactivation factor IF7A. A plasmid backbone that is replicative in both *E. coli* and cyanobacteria was selected for carrying the insert consisting of one of four promoters, the *gifA* gene (*asl2329*) and its native terminator, T_{gifA} from *Anabaena* sp. PCC 7120. Four overexpression constructs were assembled different in their promoter region driving the *gifA* cassette. The four promoters chosen are as follows (Figure 4.4): P_{petE} (copper-inducible promoter of plastocyanin), P_{gifA} (native for IF7A), P_{rbcLS} (promoter for the large and small subunit of RuBisCO) and P_{nifHDK} (promoter for the nitrogenase operon). Promoters P_{gifA} and P_{rbcLS} are expected to express only under non-diazotrophic conditions. In contrast, P_{nifHDK} is only active under nitrogen starvation and P_{petE} may be active regardless of the nitrogen status of the cell. Activity of P_{petE} mainly depends on the availability of copper, and therefore, it is assumed to express even in normal BG-11 (the normal media contains 0.3 µM $CuSO_4$). While P_{gifA} and P_{petE} may be active in both cell types, P_{rbcLS} is expected to express only in the vegetative cell and P_{nifHDK} only in the heterocyst. Altogether eight strains have been generated via triparental conjugation from two base strains (wild type *Anabaena* sp. PCC 7120 and a Δ*amt* mutant lacking the complete *amt* cluster responsible for the uptake of ammonium at low concentrations) and four constructs carrying a *gifA* cassette under the effect of one of the above promoters.

4.2.3.1 Isolation of IF7A overexpression strains

The *gifA* gene, also known as *asl2329* from *Anabaena* sp. PCC 7120 was cloned into a pRSF1010-based broad-host-range shuttle vector (Table 4.2) derived from plasmid pVZ322 (Table 2.1). The template pVZ322 plasmid contains two antibiotic resistance cassettes, *aphA* for kanamycin (Km^R) and *aacC1* for gentamicin resistance (Gm^R) which were both cut out using primers RSF1010-bb-F (133) and RSF1010-bb-R (134) in a PCR linearization, resulting in only the RSF1010 segment of the plasmid. The vector backbone was combined with an Sm^R cassette conferring resistance to streptomycin and spectinomycin antibiotics, amplified from the plasmid pDF-lac (Table 2.1) using primers SmR-TgifA-F and SmR-RSF1010-R (primers 164 and 163, respectively; see also in Table 4.2). The *gifA* cassette was amplified from high-purity genomic DNA of *Anabaena* sp. PCC 7120 to include its native promoter, CDS and terminator for construct C3, but to include only its CDS and terminator for constructs C2, C19 and C20 (bearing promoters P_{petE}, P_{rbcLS} and P_{nifHDK}, respectively). Primers, DNA fragments and constructs used here are listed in Table 4.2.

Table 4.2. List of primers, DNA fragments and constructs used to generate IF7A overexpression mutants. Sequences are written from 5' to 3' direction. Overlapping sequences are in small case.

Primers		
ID	**Name**	**Sequence**
27	gifA-R	gaagccatgaTTAACACCTGTTAGCAATTTATAG
133	RSF1010-bb-F	CTGAAAGCGACCAGGTGC
134	RSF1010-bb-R	GCAGGAGCAGAAGAGCATAC
135	PpetE-RSF1010-F	gtatgctcttctgctcctgcCAGTACTCAGAATTTTTTGCTGAGG
136	PpetE-gifA-R	gaatagacatGCGTTCTCCTAACCTGTAG
137	gifA-PpetE-F	aggagaacgcATGTCTATTCAAGAAAAATCTCG
141	gifA-RSF1010-F	gtatgctcttctgctcctgcTTGCAGTGTTCTGTTGCTG
149	RSF1010-Psch-seq-F	ATACCATGCTCAGAAAAGGC
155	PrbcLS-RSF1010-F	gtatgctcttctgctcctgcGGAAGTAAAGAAGAATGACTATGGAC
156	PrbcLS-gifA-R	gaatagacatATCTATCCTTCCAAGATGTC
157	gifA-PrbcL-F	aaggatagatATGTCTATTCAAGAAAAATCTCG

Primers (continued)

ID	Name	Sequence
158	PnifHDK-RSF1010-F	gtatgctcttctgctcctgcGTTGCGGTTCCTGTTTAGC
159	PnifHDK-gifA-R	gaatagacatTGTTCTCTTTTCCTGCAATTG
160	gifA-PnifH-F	aaagagaacaATGTCTATTCAAGAAAAATCTCG
163	SmR-RSF1010-R	gagcacctggtcgctttcagATTCTCACCAATAAAAAACGC
164	SmR-TgifA-F	caggtgttaaTCATGGCTTCTTGTTATGAC
168	PSmr-seq-R	ACATCAAACATCGACCCACG

DNA fragments

ID	Description	Primers	Template
F6b	P_{petE} promoter	135+136	wild type *Anabaena* sp. PCC 7120
F7c	*gifA*-T_{gifA} cassette (C2)	137+27	wild type *Anabaena* sp. PCC 7120
F7d	*aadA* cassette (Sm^R)	164+163	pDF-lac
F8c	P_{gifA}-*gifA*-T_{gifA} cassette	141+27	wild type *Anabaena* sp. PCC 7120
F43	pRSF1010 backbone	133+134	pVZ322
F45	P_{rbcLS}	155+156	wild type *Anabaena* sp. PCC 7120
F46b	*gifA*-T_{gifA} cassette (C19)	157+27	wild type *Anabaena* sp. PCC 7120
F47	P_{nifHDK}	158+159	wild type *Anabaena* sp. PCC 7120
F48b	*gifA*-T_{gifA} cassette (C20)	160+27	wild type *Anabaena* sp. PCC 7120

Constructs

ID	Insert	Assembled from	Genotype
C2	P_{petE}-*gifA*-T_{gifA}	F6b, F7c, F7d, F43	pRSF1010-SmR:: P_{petE}-*gifA*-T_{gifA}, Sm^R, Sp^R
C3	*gifA* cassette (wild type)	F8c, F7d, F43	pRSF1010-SmR:: P_{gifA}-*gifA*-T_{gifA}, Sm^R, Sp^R
C19	P_{rbcLS}-*gifA*-T_{gifA}	F45, F46, F7d, F43	pRSF1010-SmR:: P_{rbcLS}-*gifA*-T_{gifA}, Sm^R, Sp^R
C20	P_{nifHDK}-*gifA*-T_{gifA}	F47, F48, F7d, F43	pRSF1010-SmR:: P_{nifHDK}-*gifA*-T_{gifA}, Sm^R, Sp^R

Strains
See in Table 4.3.

DNA fragments consisting of the vector backbone, the Sm^R cassette, one of the four promoters and the *gifA* cassette were seamlessly cloned via Gibson assembly into chemically competent cells of *E. coli* DH5α on Sm/Sp selective agar plates. The resulting colonies were picked, grown in 5 ml liquid LB medium (supplemented with the appropriate antibiotics), plasmid purified, and the isolated plasmids sequenced (Source BioScience, Nottingham, UK) using sequencing primers RSF1010-Psch-seq-F (149) or PSmR-seq-R (168). An isolated construct with confirmed sequence was transformed to both *Anabaena* sp. PCC 7120 wild type and the Δ*amt* mutant using triparental conjugation. Resulting exconjugants were isolated on BG-11 agar plates supplemented with 2.5 µg/ml Sm and Sp, and assayed by colony PCR. Primers 149+168 were used to amplify a specific product from a successful transformant. The product size varied between 0.8 and 1.3 kb based on the size of the different promoters. Fragments of the right size were gel purified and sequenced (Source BioScience, Nottingham, UK) for verification of the desired genotype. Results of colony PCR are shown on Figure 4.13 for four isolates of each strain.

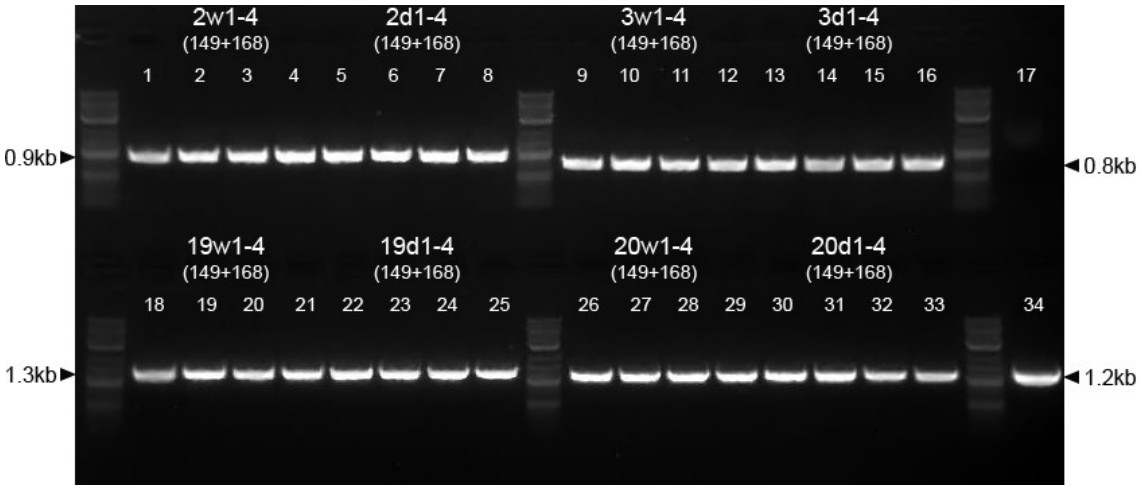


Figure 4.13. Evaluation of *gifA* overexpression isolates by colony PCR. Fragment lengths were predicted by comparison to a DNA ladder (2-log DNA Ladder, NEB).

All isolates tested by colony PCR produced a band using primers 149+168. Band intensity was similarly high for all lanes, suggesting that the amount of plasmid was similar in these samples. Sizes of bands

were as expected, 0.9 kb, 0.8 kb, 1.3 kb and 1.2 kb for constructs C2 (lanes 1–8), C3 (lanes 9–16), C19 (lanes 18–25) and C20 (lanes 26–33), respectively. The wild type negative control on lane 17 produced no band, as expected. On the other hand, the positive control using purified C20 plasmid as template resulted in the appropriate band at 1.2 kb (lane 34). It was therefore concluded that all isolates expressed the appropriate construct. Isolates were then picked into 1 ml BG-11 containing the relevant antibiotics and grown under standard conditions for a week. Culture density and health was visually observed and the best growing and greenest cultures of each strain were selected for further study (Table 4.3).

Cyanobacterial isolate cultures were subcultured in 20 ml liquid BG-11 medium for two weeks and cryopreserved in 18% glycerol. Whenever cultivated in liquid or solid media, the IF7A overexpression strains were grown in the presence of Sm and Sp antibiotics to maintain the expression plasmid. Their respective genotypes and specific growth conditions are listed in Table 4.3.

Table 4.3 Summary of the IF7A overexpression strains.

Isolate	Isolate name[a]	Description of genotype	Parent strain	Expression type
165	2w4	P_{petE}-*gifA*-T_{gifA}::pRSF1010-SmR	wild type	copper inducible[c]
167	2d2	P_{petE}-*gifA*-T_{gifA}::pRSF1010-SmR	Δ*amt*[b]	copper inducible[c]
164	3w5	P_{gifA}-*gifA*-T_{gifA}::pRSF1010-SmR	wild type	native
161	3d4	P_{gifA}-*gifA*-T_{gifA}::pRSF1010-SmR	Δ*amt*[b]	native
166	19w2	P_{rbcLS}-*gifA*-T_{gifA}::pRSF1010-SmR	wild type	VC-specific
162	19d1	P_{rbcLS}-*gifA*-T_{gifA}::pRSF1010-SmR	Δ*amt*[b]	VC-specific
160	20w6	P_{nifHDK}-*gifA*-T_{gifA}::pRSF1010-SmR	wild type	HC-specific[d]
163	20d5	P_{nifHDK}-*gifA*-T_{gifA}::pRSF1010-SmR	Δ*amt*[b]	HC-specific[d]

[a] Naming logic: construct number + parent strain (w: wild type, d: Δ*amt*) + isolate number.

[b] Strain contains the deletion of the *amt* cluster (*alr0990*, *alr0991* and *alr0992*) involved in ammonium uptake (Paz-Yepes et al., 2008).

[c] Expression induced by the presence of copper (e.g. 3 μM Cu^{2+}).

[d] Expression under diazotrophic conditions only.

4.2.3.2 Evaluation of strain growth

Isolated colonies for each strain in Table 4.3 were tested in a 5-day kinetic growth assay and compared to the wild type strain to identify any growth deficiency possibly caused by the presence of the overexpression construct. Cultures were also imaged under a bright-field microscope to see whether the mutant strains possess any different morphological features compared to their parental strain (higher filament fragmentation or an unusual heterocyst pattern for example).

Isolates of strains in Table 4.3 were grown in 1 ml liquid BG-11 supplemented with 1.5 µg/ml Sm and 1.5 µg/ml Sp for 3 days in a 24-well microplate (734-0020, VWR International Ltd., UK), under continuous illumination by cool white LED panels at 40 µE m^{-2} s^{-1}, 30 °C in 1% enriched CO_2 on a plate shaker at 190 rpm. Cultures were subcultured into 20 ml at a starting OD_{730} of approx. 0.05, grown for up to 5 days under the same conditions and sampled once every day by drawing 100 µl homogenous sample to different wells of a 96-well clear bottom polystyrene plate (734-0954, VWR International Ltd., UK). Optical density was recorded as absorbance at 730 nm (OD_{730}) in a Tecan Infinite M200 Pro multimode reader (Figure 4.14A). Due to the quick settling of cells to the bottom of the wells the contents were mixed thoroughly prior to reading by pipetting the sample up and down a couple of times.

All strains were able to grow in standard BG-11 on nitrate as the nitrogen source. There was significant difference between the individual OD values of the strains at each time point, however, the overall growth characteristics of the mutants showed a very similar pattern to the control (Figure 4.14A). Growth rate was highly dependent on the initial culture density (e.g. in case of strain 19w2, green markers), although strains with slower initial growth were able to catch up with the rest by the end of the 5 days. Strains of different parental origin bearing the same construct behaved similarly (empty and filled markers of each colour), reaching similar final ODs. Strain 3w5 (grey triangles) carrying the P_{gifA}-driven expression cassette remarkably underperformed its Δ*amt*-based strain pair, mostly due to the unexpected change in its growth rate following 72 hours. It is worth noting, all constructs except those under the effect of P_{nifHDK} were expected to express under the tested conditions.

Microscopic images of the filaments, on the other hand, showed no structural variations among the different strains including the control, and no heterocyst formation was observed in any of the samples at the end of the 5-day growth assay (Figure 4.14B). In conclusion, the above data suggest the expression system introduced to the IF7A overexpression strains may represent a slight burden (observed as slower growth in some cases), but has no considerable impact on growth characteristics and filament health when grown on combined nitrogen. In addition, the introduced constructs, under

the conditions tested, did not alter the C/N balance of the cells so radically that could have caused unusual heterocyst formation.

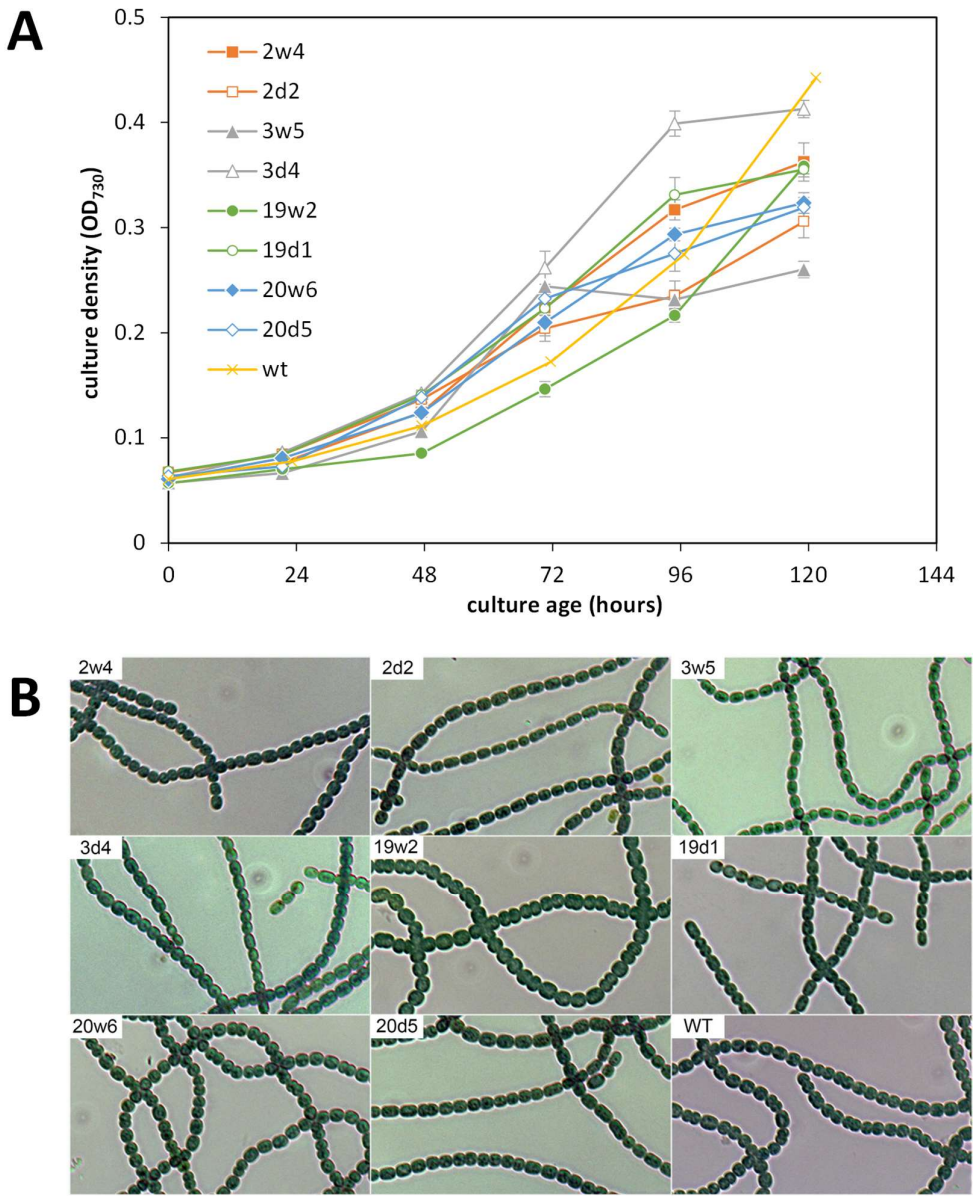


Figure 4.14. Growth characteristics of the IF7A overexpression strains under non-diazotrophic conditions. (A) Growth curves of eight mutants and a control (wild type) strain. Data points are the average of two biological replicates, with ±1 standard deviation displayed as error bars. **(B)** Bright-field microscopic images of cyanobacterial filaments at 400× magnification. Strains were cultivated in standard BG-11 under continuous illumination and 1% enriched CO_2 for up to 5 days, and sampled every day. Biological replicates had been pooled prior to microscopic imaging. Only one of several images per strain is shown.

However, the primary purpose of the IF7A overexpression strains is to maintain growth in nitrogen starvation while also excreting ammonia. Therefore, mutant strains were also tested under diazotrophic conditions in BG-11_0, in the absence of any combined nitrogen (Figure 4.15A).

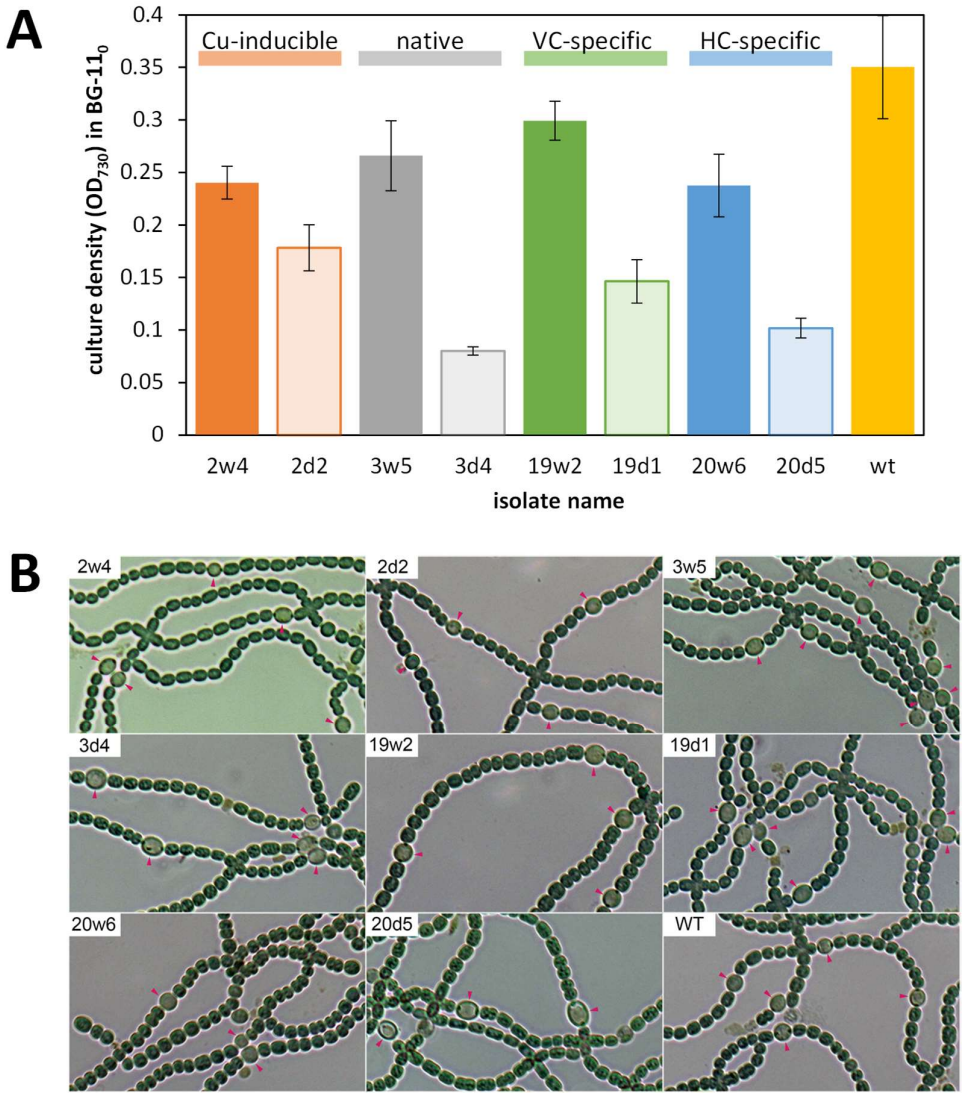


Figure 4.15. Short-term growth test of IF7A overexpression strains in BG-11_0 under diazotrophic conditions. (A) Culture density at 730 nm. Columns of the same colour represent strains overexpressing *gifA* under the same promoter. The type of expression determined by the different promoters is indicated above each column pair. Error bars depict ±1 standard deviation of two biological replicates. **(B)** Bright-field microscopic images of cyanobacterial filaments at 400x magnification. Pink arrowheads point at heterocysts. Strains were cultivated diazotrophically in BG-11_0 under continuous illumination and 1% enriched CO_2 for up to 5 days. Biological replicates had been pooled prior to microscopic imaging. Only one of several images per strain is shown.

First, exponentially growing non-diazotrophic cultures were harvested by centrifugation at 4,000 × g for 10 min, washed in BG-11_0 three times and diluted to a starting OD_{730} of 0.05 in 20 ml BG-11_0. Cultivation was performed under standard conditions for 5 days, as described above. Culture density was measured at 730 nm from 100 µl homogenous sample in a 96-well microplate (Figure 4.15).

In contrast to the OD_{730} values on combined nitrogen (Figure 4.14A), diazotrophic growth was about 40% lower for all strains including the control in endpoint measurements of a 5-day growth assay. In addition, nitrogen-starved mutants have significantly underperformed the diazotrophically grown wild type as well (yellow column, Figure 4.15A). Those of wild type origin (darker columns) showed considerably higher growth than the strains lacking the *amt* cluster (lighter columns), with 19w2 (green column) being the highest among the mutants and only slightly lower than the control. Mutants based on the Δ*amt* strain (lighter columns) were highly deprived in their growth, especially strains 3d4 and 20d5 (grey and blue columns, respectively). All strains including the control showed normal fragment health and heterocyst pattern in the 5-day growth assay (Figure 4.15B). The frequency of heterocysts along the filament was in the range of one heterocyst for every 8–16 vegetative cells for the IF7A overexpression mutants and the wild type control strain as well that is a typical value for *Anabaena* sp. PCC 7120 (Kumar et al., 2010; Ehira, 2013). Based on these observations the wild type-based strains are only moderately growth deprived under nitrogen-depleted conditions, whereas their Δ*amt*-based pairs experience a pronounced growth difficulty. Nonetheless, heterocyst pattern and frequency was not found to be affected by the introduced overexpression constructs.

In order to determine whether the mutant strains possess the desired phenotype of a lowered GS activity, a specific bioassay was carried out on cell-free protein extracts of each IF7A strain and their respective controls.

4.2.3.3 Glutamine synthetase bioactivity in nitrogen-depleted and replete conditions

The direct measure for the efficiency of the individual IF7A overexpression constructs is the total GS activity in the strain. The higher the overexpression of the inactivation factor, the lower overall GS activity is anticipated. A lower-than-normal GS activity is expected to result in excretion of intracellular ammonia, as concluded by the proof-of-concept experiment in Figure 4.3.

Activity of GS was determined from cell-free extracts of 3-day-old diazotrophic cultures of the IF7A overexpression strains. Cultures were grown in BG-11_0 under normal conditions (30 °C, 40 µE m^{-2} s^{-1} continuous illumination from cool white light and 1% enriched CO_2), harvested by centrifugation,

washed in fresh medium and the total protein was extracted. Residual cell debris was removed by centrifugation. Exactly 50 µg total protein was loaded into an enzymatic bioactivity assay to determine activity of GS over 15 minutes. The principle of the assay is the following reaction:

$$L - Glu + NH_3 + ATP \rightleftharpoons L - Gln + ADP + P_i$$

Activity of the enzyme is directly proportional to the amount of inorganic phosphate produced in the reaction. To develop colour (bluish grey) ammonium heptamolybdate (colourless) was added and absorbance recorded at 850 nm in a Tecan Infinite M200 Pro plate reader. A substrate blank was prepared for every sample by replacing sodium glutamate in the reaction mix for deionized water in order to eliminate variations due to the initial blue colour of samples. Figure 4.16 shows the bioactivity of GS determined from cell-free protein extracts.

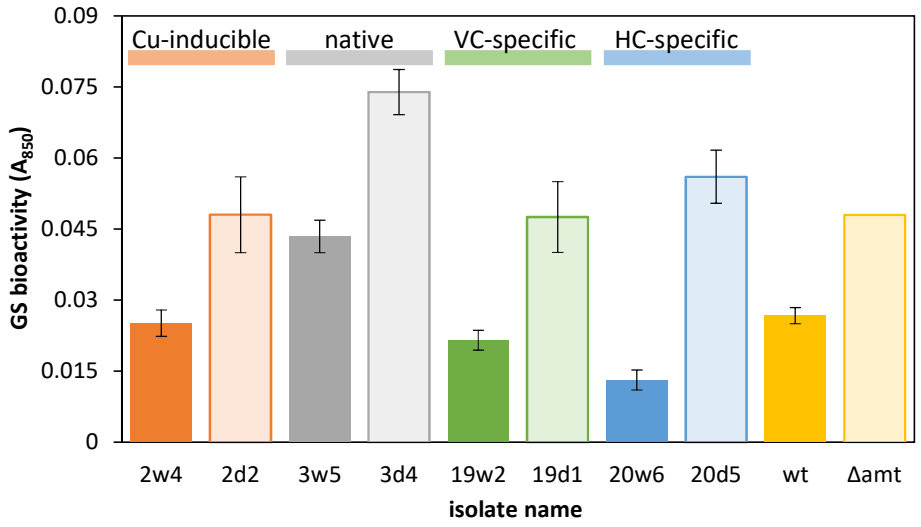


Figure 4.16. Glutamine synthetase activity of different IF7A overexpression strains under diazotrophic conditions. Activity values are normalized to 50 µg total protein. Columns of the same colour depict strains overexpressing *gifA* under the same promoter. Darker and lighter colours represent strains of wild type and Δ*amt* origin, respectively. The type of expression determined by the different promoters is indicated above each column pair. Error bars represent ±1 standard deviation of three biological replicates.

Remarkably, Δ*amt* strains in all cases possessed considerably higher GS activity than their wild type-based pair (a strain carrying the same construct). The difference between the pairs was fourfold in case of strains 20w6 and 20d5, bearing the construct driven by the heterocyst-specific P_{nifHDK} promoter. Even for the control pair the Δ*amt* strain showed two times higher GS activity compared to

the wild type. These results suggest that removal of the *amt* cluster has significant effect on the activity of GS *in vitro*. It is also worth noting that GS activity for 19w2 and 19d1 (green columns) was slightly below the controls possibly indicating an increased level of IF7A, although the difference was not significant statistically. In contrast, strains 3w5 and 3d4 (grey columns), as well as 20w6 and 20d5 (blue columns) showed significant difference from their respective controls. This is not surprising as these strains contain the two overexpression constructs (bearing P_{gifA} and P_{nifHDK}) expected to be active under these conditions. It was assumed that an active overexpression construct will lower GS activity. However, 3w5 and 3d4 reacted the opposite way, and exhibited an activity 1.6 and 1.3-times higher than the wild type and Δ*amt* controls, respectively. Similarly, strain 20d5 also showed higher GS activity than its strain control, and only 20w6 resulted in a decrease of GS activity. The fact that 3w5, 3d4 and 20d5, despite their supposedly active overexpression construct, possessed an increased GS bioactivity is surprising. A possible explanation may be that these strains enhanced their GS expression to compensate for the elevated levels of IF7A and therefore retained their normal ammonia conversion capacity. To investigate this hypothesis GS was quantified from cell-free protein extracts of diazotrophically cultivated samples.

Cell-free protein extracts for the quantification of GS and other proteins were prepared as the samples in Figure 4.16. The analysis was carried out on proteotypic tryptic peptides of the target proteins in an LC-MS/MS system configured with a triple quadrupole mass analyser (TQMS). Proteotypic peptides were generated in an overnight digestion by proteomics grade trypsin and analysed as described in section 2.7.3. Several parent ion–daughter ion transitions were compared, outliers identified and discarded, and peak areas normalized to decrease noise and improve comparability of the different peptides. Target proteins were identified from their respective proteotypic peptides based on amino acid sequence information derived from species-specific proteomics databases. Along with GlnA (GS) levels of IF7A, NtcA, AtpB, RpoC1 and RpoC1 were quantified. The latter three proteins (ATP synthase beta subunit, RNA polymerase gamma subunit and RNA polymerase beta subunit, respectively) were included as quality controls and were also used as internal reference.

Due to its small size, quantification of IF7A was difficult using this method. There have been two proteotypic peptides identified for IF7A: a shorter one consisting of 6 amino acids (peptide sequence: IDPTTR) and a longer one comprising of 22 amino acid residues (peptide sequence: SAQELGLPAEELSHYWNPTQGK). To improve detection of IF7A, both proteotypic peptides were chemically synthesized (Biomatik Corp., Canada). In addition the analytical method was optimized to detect more reliable and reproducible transitions for these peptides. Even after method development, only the longer proteotypic peptide could be detected in some samples, close to the limit of detection

for the applied method. Therefore, instead of directly quantifying IF7A from the overexpressing cultures its hypothetical effect on the level of GS (GlnA) was evaluated. A similar effect to that of *amt* knockout in Figure 4.16 was observed on GS. The level of GlnA in the Δ*amt*-based strains was a fewfold higher than that of the wild types for all pairs. Nevertheless, the presence of the overexpression constructs showed no significant impact on GlnA expression, except for 19d1 (light green column) and 20w6 (blue column, Figure 4.17).

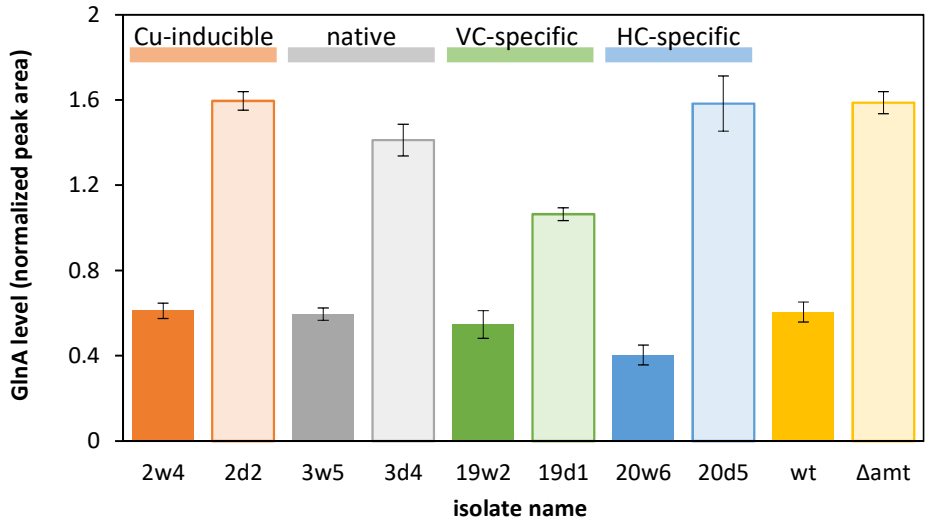


Figure 4.17. Levels of GS (GlnA) in cell-free extracts of IF7A overexpression strains under diazotrophic conditions. Protein measurements were performed in a TQMS instrument from tryptic peptides. Columns of the same colour denote for strains overexpressing *gifA* under the same promoter. The type of expression determined by the different promoters is indicated above each column pair. Error bars represent ±1 standard deviation over the average of normalized peak area of all transitions.

In fact, GlnA in both 19w2 and 19d1 (green columns) was lower than in the respective controls, although the difference in case of 19w2 was not statistically significant. The construct present in these strains (under the effect of P_{rbcLS}) is assumed to be expressed constitutively that may serve as an explanation to the subtly lowered GS activity in Figure 4.16. It is an open question, however, whether a GS molecule inactivated by the covalently binding IF7A is detectable in the TQMS method. Surprisingly enough, GlnA concentrations in 3w5, 3d4 (grey columns) and 20d5 (light blue column) were at the respective controls' level, even though the measured bioactivity was above the controls in Figure 4.16. Therefore, the hypothesis that the elevated GS activity in Figure 4.16 may be a response to the cell's higher IF7A concentration does not hold. It is unlikely that these strains would compensate

for IF7A overexpression by GS upregulation. At the same time, higher GS activity in Figure 4.16 for 3w5 and 3d4 (grey columns) did not coincide with higher GlnA level in Figure 4.17. This would mean the same amount of protein exhibits higher specific activity, which is very unlikely. To clarify the effect of the overexpression constructs it was necessary to assess strains side-by-side under conditions that clearly activate or deactivate each construct.

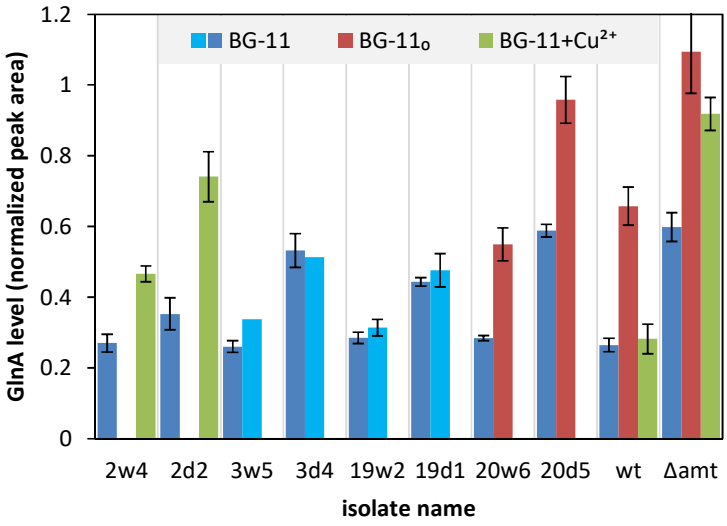


Figure 4.18. Level of GlnA in cell-free extracts of IF7A overexpression mutants under different growth conditions. Cultures were grown in BG-11 liquid medium (containing nitrate) for 5 days, then under promoter-specific conditions inducing the overexpression of IF7A. Strains 2w4 and 2d2 were grown in the presence of 3 µM copper, and strains 20w6 and 20d5 were cultivated under diazotrophic conditions for 3 days, together with their respective controls. Strains 3w5, 3d4, 19w2 and 19d1 were grown under standard photoautotrophic conditions throughout the experiment (lighter blue bars represent the 3-day incubation). Columns display the average of normalized peak area of all transitions of 2 replicates; error bars display ± standard deviation. GlnA levels are relative to 50 µg total protein.

In order to further investigate some of the discrepancies observed for IF7A overexpression, each construct was induced specific to the promoter of their overexpression cassette. First, strains were grown in 20 ml BG-11 supplemented with 1.5 µg/ml spectinomycin and streptomycin antibiotics for up to 5 days in continuous illumination (40 µE m^{-2} s^{-1}) by cool white light at 30°C, harvested by centrifugation at 4,500 × g, room temperature for 10 min, washed twice in 20 ml, and one half of the volume resuspended in 3 ml fresh BG-11_0 medium. The other half was immediately frozen at -80°C for future protein extraction (uninduced sample). The remainder was inoculated into 20 ml specific culture media (BG-11, BG-11 + Cu^{2+} or BG-11_0) supplemented with the appropriate antibiotics.

According to the overexpressing promoter, strains 3w5, 3d4, 19w2 and 19d1 were inoculated into normal BG-11, 2w4 and 2d2 were inoculated to BG-11 + Cu^{2+}, and 20w6 and 20d5 were further cultivated in $BG\text{-}11_0$. After two days cultures were harvested as above and stored at -80°C until further analysis. Protein extraction was performed as described above on exactly 50 µg protein (determined by a BCA protein assay) and tryptic peptides analysed by LC-MS/MS (Figure 4.18).

It is worth noting that strains carrying expression cassettes with the *gifA* promoter (3w5 and 3d4) and the vegetative-cell specific P_{rbcLS} (19w2 and 19d1) are not inducible. Theoretically, the overexpression cassettes in strains 19w2 and 19d1 are already activated under non-diazotrophic conditions and continuous light regimes, as the RuBisCO promoter P_{rbcLS} assures a relatively constant expression level under such conditions. However, P_{gifA} in 3w5 and 3d4 is only active under nitrogen starvation and otherwise repressed by NtcA, the global nitrogen regulator of cyanobacteria. To avoid diazotrophic cultivation masking the impact of IF7A overexpression by other effects in response to nitrogen deprivation, strains 3w5 and 3d4 were cultivated in normal BG-11 throughout the experiment. As expected for 3w5, 3d4, 19w2 and 19d1, there was no change in the level of GlnA from a 5-day culture (last day of the uninduced period for the other cultures) to an 8-day culture (end of induction). This indicates that GlnA expression remains constant in a growing culture relative to total protein, as shown by the two shades of blue columns for these two strains in Figure 4.18. Similarly, GlnA level was constant for the uninduced control strains as well (grown in BG-11 for 8 days; only 5-day values are included here). Notably, strains 3w5, 3d4, 19w2, 19d1 and the controls all expressed GS at the same level over the course of the experiment (blue columns of both shades). Those strains that have been induced increased their GS expression remarkably compared to the non-induced state (blue columns versus green and red columns). The addition of copper to the cultures of 2w4 and 2d2 increased GlnA expression by about twofold in these strains (green columns). Interestingly, this increase was only higher than that of the control in case of 2w4, while GlnA in 2d2 remained slightly below its induced control (Δ*amt*, green column). In fact, GlnA in 2d2 was lower than the control also before the induction by copper, although all samples were diluted to the same total protein concentration prior to analysis. The significant response of Δ*amt* to the addition of copper was unexpected. In addition, any effect observed for the copper-induced 2d2 (green column) that is not bigger than the induced Δ*amt* control may be explained by a complex cellular response to copper, rather than by successful induction of the IF7A overexpression construct.

Cultures of 20w6 and 20d5 reacted to nitrogen stepdown very similar to the control strains (blue and red columns), suggesting that any effect causing the nearly twofold jump in GlnA level is due to the change in culturing conditions, rather than the induced expression of IF7A itself. These results strongly

suggest that expression of GS is not influenced by the overexpression of IF7A, in other words, strains do not compensate for the elevated levels of IF7A by also overexpressing GS. Moreover, the observed independence of GS expression of conditions that theoretically induce the IF7A overexpression constructs may indicate that the *gifA* genes are not expressed at all. Furthermore, direct measurement of IF7A by mass spectrometry was not feasible from cell-free protein extracts (as mentioned above); therefore, conditions that activate native IF7A expression were tested as well.

In wild type *Anabaena* sp. PCC 7120 expression of GS and IF7A is primarily influenced by ammonium. In the presence of this nitrogen compound GS gets inactivated due to the upregulated expression of IF7A; however, in the absence of ammonium IF7A is repressed and activity of GS is at its native level (Galmozzi et al., 2010). In order to investigate whether IF7A overexpression can be activated upon addition of ammonium, diazotrophically grown cultures of the IF7A mutant strains were spiked with NH_4Cl. First, 20 ml of a 3-day old diazotrophic culture was harvested for each mutant strain by centrifugation at 4,500 × g, room temperature for 10 min, washed twice in 20 ml and resuspended in 1 ml fresh BG-11_0 liquid medium. Half of the concentrated sample was prepared for protein extraction as described above (nitrogen-depleted conditions, purple columns in Figure 4.19); the rest of the culture was inoculated into 20 ml BG-11_0 supplemented with the appropriate antibiotics, spiked with NH_4Cl to 500 µM final concentration and cultivated for an additional 3 days (nitrogen replete conditions, blue columns in Figure 4.19).

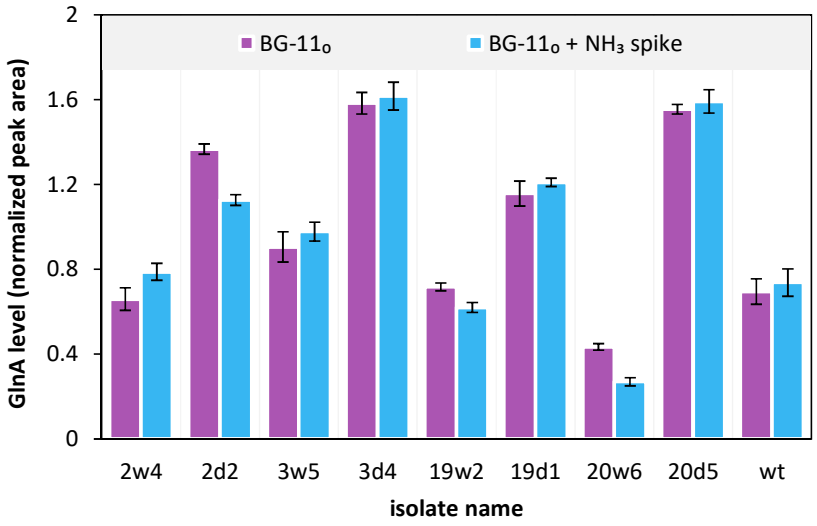


Figure 4.19. Level of GlnA in the IF7A overexpression strains in response to spiking with ammonia. Cultures were grown under nitrogen-depleted conditions for 3 days (purple columns), spiked with 500 µM NH_4Cl and cultivated for another 3 days (blue columns). Level of GlnA was determined by the analysis of tryptic peptides derived from extracted proteome. Columns represent the average of normalized peak areas of all transitions in exactly 50 µg total protein. Error bars depict ±1 standard deviation of 3 replicates.

Level of GS in diazotrophically grown cultures in Figure 4.19 (purple columns) showed a similar pattern to that of non-diazotrophic cultivation of the same strains in Figure 4.17. The same difference in GlnA concentration between the wild type and Δ*amt*-based strains could be observed for each mutant. Furthermore, GS in strain 3w5 possessed significantly higher concentration than in the rest of the mutants of identical parental origin and in the control. The elevated GS in 3w5 may also be indicative of the increased GS activity of this strain in Figure 4.16. Oddly enough, spiking the cultures with ammonium had no significant effect on the level of GlnA (Figure 4.19, blue columns), or at least, nothing was detectable after 3 days of the addition of NH_4Cl. However, other proteins characteristic to the nitrogen status of the cell truly reflected the change in the availability of fixed nitrogen. Figure 4.20A below collects expression levels for NtcA under diazotrophic cultivation and following spiking with 500 µM NH_4Cl. Protein concentrations for NtcA were determined from the same samples as GlnA in Figure 4.19, but in a different run on the TQMS. Although the lack of change in GS levels in Figure 4.19 postulated no difference in cellular nitrogen status following the perturbation, the nitrogen status has indeed changed upon addition of ammonium. Green and brown columns in Figure 4.20A show the original diazotrophic and the perturbed state of NtcA expression, respectively. For all strains including the control NtcA level responded to the addition of ammonia with a significant drop in concentration, as it was expected (Herrero et al., 2001). The lower level of NtcA following spiking with ammonium should result in higher IF7A expression. Indeed, IF7A has become detectable upon addition of NH_4Cl (Figure 4.20B, brown columns), although only in a subset of the samples.

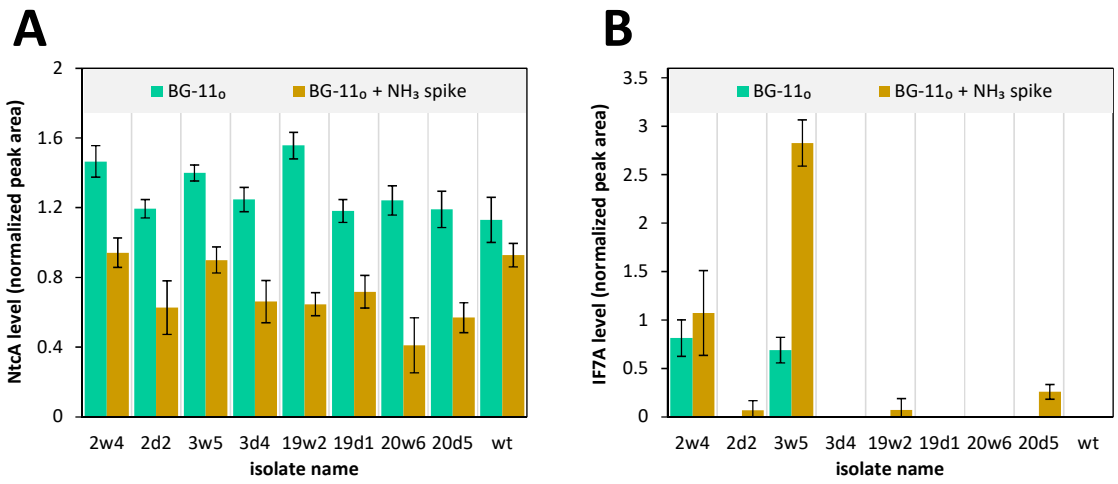


Figure 4.20. Level of NtcA in the IF7A overexpression strains in response to spiking with ammonium. NtcA **(A)** and IF7A **(B)** concentrations in cell-free extracts of diazotrophically grown cultures prior to spiking with 500 µM NH_4Cl (green columns) and within three days following the perturbation (brown columns).

Although visible for half of the mutants, IF7A level remained below detection limit for the other half of the overexpression strains and the control. The highest IF7A concentration was measured for 3w5 (bearing the native *gifA* promoter in the overexpression construct), for which strain the increase upon spiking reached up to a threefold boost (brown column) over the original diazotrophic level (green column). In contrast, the second highest IF7A expressing strain, 2w4 showed only a modest 20% increase over its original expression level.

In conclusion, the above analyses of selected proteins and the measurement of GS bioactivity revealed that removal of the *amt* gene cluster significantly affected the behaviour of all strains in response to nitrogen deprivation, and in other cases, in response to increased copper concentration. Although strains 2w4, 2d2, 20w6 and 20d5 reacted to these perturbations as expected, the effect by which the induction of the IF7A overexpression constructs contributed to the changes in GlnA expression profile was not distinguishable from the impact of the *amt* deletion. In some cases the effect of an active IF7A overexpression was clearly visible (19d1 in Figure 4.16, 2d2 and 19w2 in Figure 4.18 and 20w6 in Figure 4.19). In addition, upon spiking with ammonium, increased levels of IF7A for some of the mutant strains became obvious, despite the sensitivity issues of the applied analytical method for the detection of IF7A. As a final attempt to determine the potential of the new strains to excrete NH_3–N, free ammonia was quantified from the supernatants of diazotrophic cultures and compared to their respective controls.

4.2.3.4 Ammonia production

Intracellular ammonia concentration in the native cell is tightly controlled by the rapid action of GS. It was shown that inactivation of the enzyme by a glutamate analogue (MSX) results in the excretion of ammonia (Figure 4.3A and B), while non-treated cultures produce no detectable NH_3-N. It was therefore hypothesized that overexpression of the native regulator of GS activity, IF7A will result in a similar phenotype. Direct measurement of IF7A was found difficult due to the small size of this oligopeptide and indirect measurements by determining activity and concentration of GS were also inconclusive. Therefore, free extracellular ammonia was quantified from culture supernatants to assess the capability of each overexpression construct to excrete NH_3-N compared to their parental strains, while under nitrogen starvation. Diazotrophic cultures were grown in 20 ml BG-11_0 at 30 °C under continuous illumination from 40 µE m^{-2} s^{-1} of cool white light in a 1-% enriched CO_2 atmosphere in 100-ml Erlenmeyer flasks. Cultures were sampled after 3 days following nitrogen deprivation by drawing 500 µl supernatant into a microcentrifuge tube. Tubes were centrifuged at 4,500 × g to pellet

any remaining cells and a clear supernatant was transferred into a clean well of a microtitre plate. Concentration of free ammonium was quantified using either the Abcam Ammonia Assay Kit (modified Berthelot test, ab102509; Abcam Plc., Cambridge, UK) or a modified, micro-scale Willis method (Willis et al., 1996). The two assay methods produced very similar results, and only values for the Willis method are shown. Colorimetric readings were converted to micromoles per litre (µM) using an NH_4Cl calibration curve and have also been normalized to culture density (Figure 4.21A). For technical reasons no biological replicates could be included in the experiment, although the two assay methods discussed above may be treated as technical replicates.

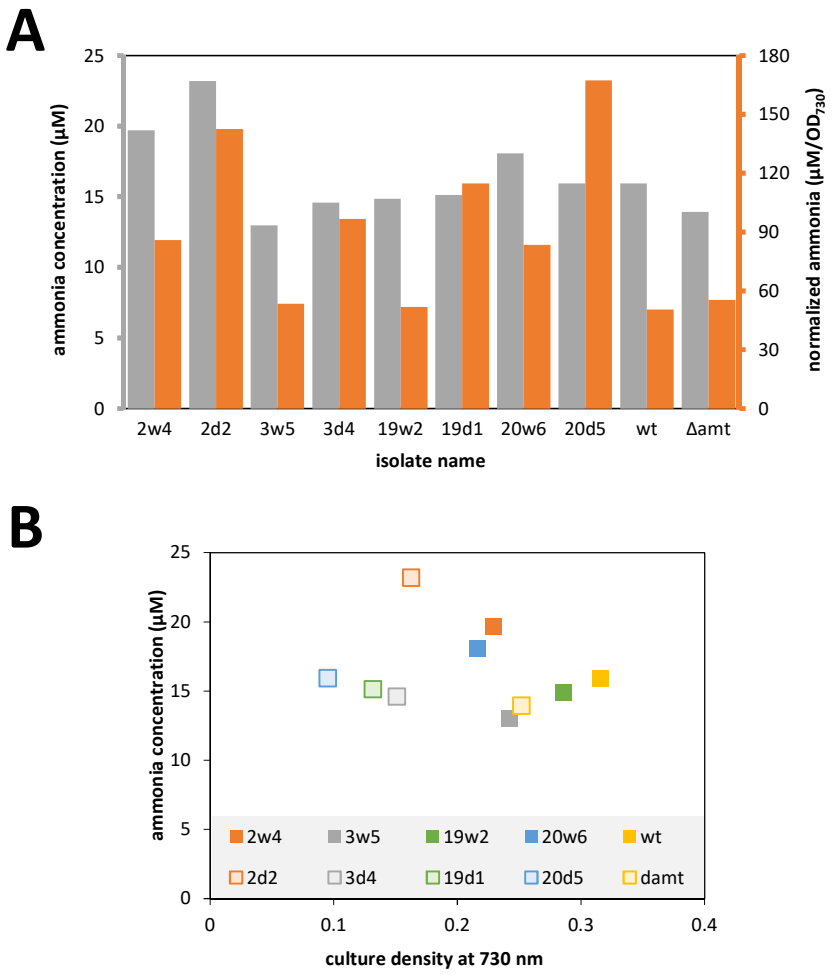


Figure 4.21. Ammonia production by the IF7A overexpression strains under diazotrophic conditions. (A) Grey columns show ammonia concentrations (left y-axis) in micromoles per litre. Orange columns represent ammonia productivity (right y-axis) normalized to culture density for each strain in a 3-day assay. **(B)** Correlation between extracellular ammonia and culture density. Dark and light colour squares represent strains of wild type and Δ*amt* origin, respectively. Identical colours indicate the presence of the same overexpression construct. The experiment included no biological replicates.

Some ammonia was detectable in all strains' supernatants, including the controls. Most cultures produced ammonia between 12 and 15 μM, similar to the controls. Mutants 2w4 and 2d2 (P_{petE}) showed the highest ammonia accumulation with 20 and 23 μM, respectively. For most of the mutant pairs the Δ*amt*-based strain achieved slightly higher ammonia titres than their wild type-based pair, except for 20w6, 20d5 and the two controls (grey columns, Figure 4.21A). The difference in extracellular ammonia by the different strains was independent of culture density, as shown in Figure 4.21B. In general, wild type-based strains showed a more active growth, while their ammonia production was similar or lower than that of the Δ*amt*-based strains.

As expected, the excretion profile for ammonia was remarkably different when endpoint concentrations were normalized to growth (measured as OD at 730 nm). According to the calculated dataset (orange columns, Figure 4.21A), the strains' ability to excrete ammonia is highly dependent on their parental origin, with Δ*amt*-based strains producing significantly more ammonia than their respective wild type-based pairs. The effect of the different promoters on ammonia accumulation was somewhat lower than that of strain origin, but still significant. Strains overexpressing IF7A under the copper-inducible P_{petE} (2w4 and 2d2) and the heterocyst-specific P_{nifHDK} (20w6 and 20d5) showed considerably higher ammonia values than the strains under the effect of the other two promoters, P_{gifA} and P_{rbcLS} (orange columns in Figure 4.21A). In general, all but 3w5 and 19w2 were better ammonia producers than the controls. Mutants 20d5 and 2d2 achieved the highest levels of extracellular ammonia, followed by 19d1. Regarding the mutants of wild type origin, only strains 2w4 and 20w6 accumulated more ammonia than the controls (orange columns, Figure 4.21A). The data above clearly shows that the majority of mutant strains excrete ammonia under diazotrophic conditions, at levels significantly above the production by the control strains. Together with the results in Figure 4.16, Figure 4.18 and Figure 4.19 ammonia accumulation in the supernatants of the mutant strains in Figure 4.21A strongly suggest that the mutants possessed the desired phenotype of ammonia excretion. In addition, the significantly higher levels of ammonia excretion in the transformed strains suggest that this new phenotype may be the result of IF7A overexpression.

4.2.3.5 Co-cultivation with microalgae

In addition to the above experiments, IF7A overexpressing strains in Table 4.3 were assessed in co-cultivation with a eukaryotic microalga. The purpose of these tests was to investigate whether the NH_3-N accumulated by the mutant strains is accessible to non-diazotrophs. The co-cultivation experiments were performed by Mr Anthony Riseley at the University of Cambridge (Prof Chris Howe's

Laboratory, Department of Biochemistry; Cambridge, United Kingdom). The below paragraphs were compiled from experimental data and observations by Mr Anthony Riseley, and have been re-evaluated by the author of this thesis. Further details about the experiment can be found in the corresponding PhD dissertation titled "Exploring the potential of algae-bacteria communities for biotechnology" (Riseley, 2017).

In order to test whether the nitrogen-excreting IF7A overexpression mutants of *Anabaena* sp. PCC 7120 could promote the growth of algal strains, algal co-cultures were prepared with selected mutant strains in BG-11 medium. Due to its ability to grow in BG-11 medium, the green alga *Chlorella vulgaris* was selected as the algal partner in these experiments (Feng et al., 2011; Belotti et al., 2014). *Chlorella vulgaris* is a unicellular freshwater green alga that is widely studied and grown for its biotechnological applications, such as lipid accumulation and bioremediation (Converti et al., 2009; Lim et al., 2010; Heredia-Arroyo et al., 2011). Strains 2w4, 2d2, 20w6, 20d5 and wild type *Anabaena* sp. PCC 7120 were selected for co-cultivation with *Chlorella vulgaris* in an 8-day assay under standard growth conditions (30 °C, 40 µE m^{-2} s^{-1}), in 50ml BG-11_0 medium bubbled with air at a constant volumetric rate. The medium was supplemented with the appropriate antibiotics to maintain the IF7A overexpression constructs. In case of co-cultivation with mutant strains 2w4 and 2d2, 3 µM $CuSO_4$ was also added to the medium to induce the P_{petE} promoter in these strains. The alga was also grown alone in BG-11 (containing nitrate) and BG-11_0 (no nitrogen source added) to serve as positive and negative controls in the experiment, respectively. Culture density (OD_{750}) was measured at the end of the 8-day assay. In addition, 8-day cultures were harvested, and exactly 1 ml was spread on TAP agar plates containing 100 µg/ml ampicillin. These plates were used to specifically count colony forming units (CFUs) of *Chlorella vulgaris*. This was possible because previous test showed that *Anabaena* sp. PCC 7120 cannot survive under these conditions and therefore, *Chlorella vulgaris* can be separated from the co-cultivated cyanobacterium (Riseley, 2017).

Figure 4.22A displays culture density at 750 nm. All co-cultures including the wild type showed significant, healthy growth. The algal monoculture in BG-11 (positive control, light grey column) was also growing well (reaching an OD of 2.7 at 750 nm), but not the negative control in BG-11_0 (dark grey column, growing to OD_{750} = 0.1), as expected. These OD values corresponded to 116 × 10^6 and close to zero *Chlorella vulgaris* CFUs per ml for the positive and negative controls, respectively (Figure 4.22B). The co-culture with wild type *Anabaena* sp. PCC 7120 strain was also growing at high density, reaching the OD_{750} value of 2.1 (Figure 4.22A, yellow column). This density value of the wild type–*Chlorella vulgaris* co-culture corresponds to about 80% of the non-diazotrophic *Chlorella vulgaris* monoculture (positive control). It is worth noting, however, that the high OD_{750} value measured for

the wild type co-culture was almost entirely the optical density of *Anabaena* sp. PCC 7120 cells, as indicated by the practically zero *Chlorella vulgaris* CFUs per ml in Figure 4.22B (yellow column).

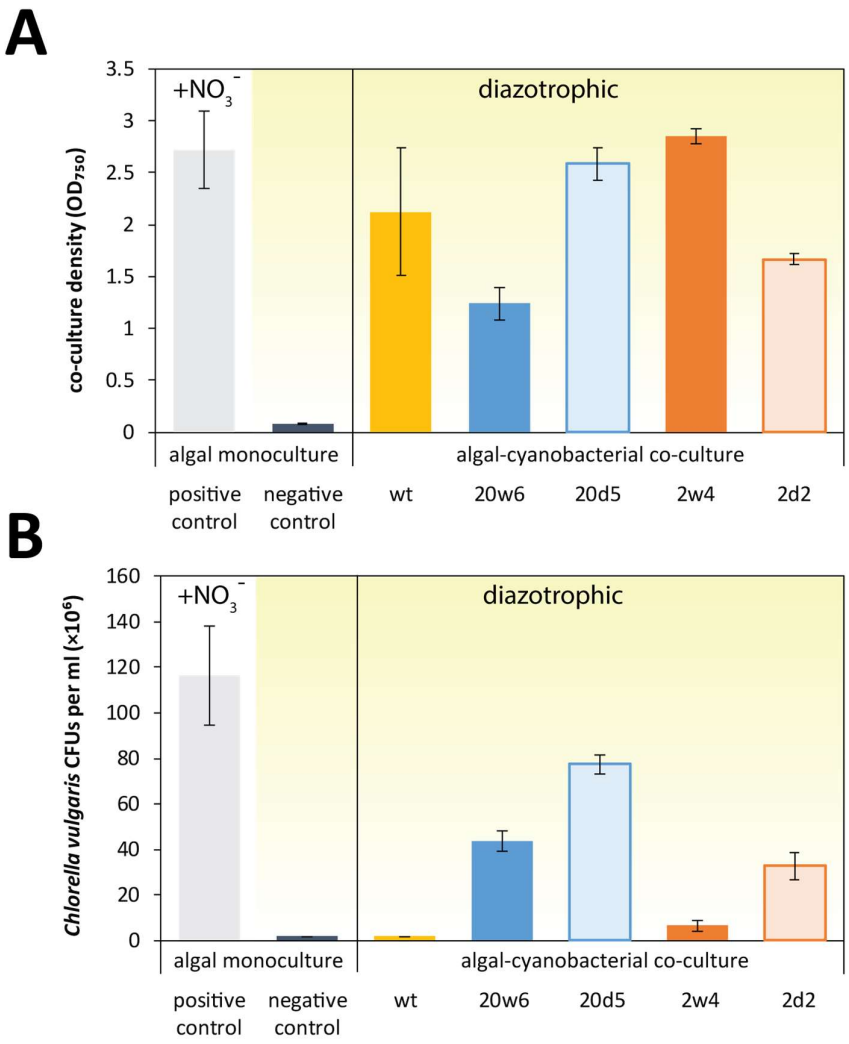


Figure 4.22. Evaluation of selected IF7A overexpressing strains in co-cultivation with *Chlorella vulgaris*. (A) Culture density measured at 750 nm. **(B)** Number of colony forming units (CFUs) of *Chlorella vulgaris* in 1 ml culture on TAP agar plates. Cultures were bubbled with air in BG-11_0 for 8 days and supplemented with 3 µM $CuSO_4$ in case of 2w4 and 2d2. Negative control contained *Chlorella vulgaris* only. Positive control (white background) contained *Chlorella vulgaris* grown in standard BG-11 (containing nitrate). Results in yellow background were prepared in diazotrophic cultivation. Error bars display standard deviation from three biological replicates. The figure was prepared based on Figure 4.9A–B in Riseley (2017).

Among the IF7A overexpression strains, mutant 2w4 achieved the highest culture density among the mutant co-cultures by reaching an OD_{750} of 2.8, followed by strain 20d5 at an OD_{750} of 2.6. The other

two mutant co-cultures displayed somewhat lower, but still remarkable growth ability, achieving the final OD_{750}'s of 1.6 and 1.2 for strains 2d2 and 20w6, respectively. The ability of the mutant strains to enable growth of *Chlorella vulgaris* followed, however, a different pattern. In case of the 2w4 co-culture, the number of *Chlorella vulgaris* CFUs per ml was very low, about 6.2×10^6. The other three co-cultures resulted in higher CFU numbers by several fold, enabling growth of the alga at 32×10^6, 44×10^6 and 77×10^6 *Chlorella vulgaris* CFUs per ml, when co-cultivated with strains 2d2, 20w6 and 20d5, respectively. Thus, the most productive diazotrophic co-culture in terms of *Chlorella vulgaris* CFU number (in case of mutant 20d5) reached about 70% of the CFU number of a *Chlorella vulgaris* monoculture (positive control, grey column in Figure 4.22B) growing on nitrate.

4.2.3.6 Long-term stability of the ammonia producing strains

In Figure 4.14 it was shown that maintenance of the overexpression constructs does not place an extraordinary burden on the growth of the mutant strains, when combined nitrogen is available. Under nitrogen deprivation, however, overexpression strains lacking the *amt* cluster displayed significantly slower growth compared to their wild type-based pairs (Figure 4.15). In industrial co-cultivation it may be necessary to maintain strains for a longer period. To further assess the viability of the IF7A overexpression strains, all mutants were grown in 20 ml BG-11 medium for 5 days under standard growth conditions (40 µE m^{-2} s^{-1} continuous illumination, 30 °C, 1% enriched CO_2 and shaking at 190 rpm). The resulting cultures were washed in 20 ml BG-11_0 medium twice at 4,500 × g, room temperature for 10 min, subcultured in 20 ml BG-11_0 in an 11-times dilution (2 ml of the original biomass was inoculated into 20 ml fresh liquid medium) in 100-ml Erlenmeyer flasks and further incubated under continuous illumination (40 µE m^{-2} s^{-1}), 1% enriched CO_2 at 30°C and 190 rpm. Subculturing was repeated every week for up to a month and optical density of the final culture was measured using a Tecan Infinite M200 Pro multimode reader. Results are collected in Figure 4.23A.

Three of the mutant strains of wild type origin showed fairly normal growth compared to their control, similar to that described earlier (Figure 4.15). Only strain 20w6 among the wild type-based strains exhibited a considerably slower growth than the wild type. Mutants derived from the Δ*amt* strain, however, were severely growth deprived at the end of the one-month incubation in contrast to their respective Δ*amt* control. These cultures looked green or yellow-green and somewhat healthy just like their wild type-based pairs, although maintained themselves at a remarkably lower culture density. The slower growth may indicate that these strains cannot recapture the lost ammonium, as expected from their lack of the *amt* cluster of transporter genes. In contrast, the wild type-based strains,

although potentially also leaking ammonium, may more efficiently recover the lost commodity. In order to investigate this hypothesis, free ammonia was measured from culture supernatants of the above samples by the Willis micro-scale method, and results were plotted in Figure 4.23B.

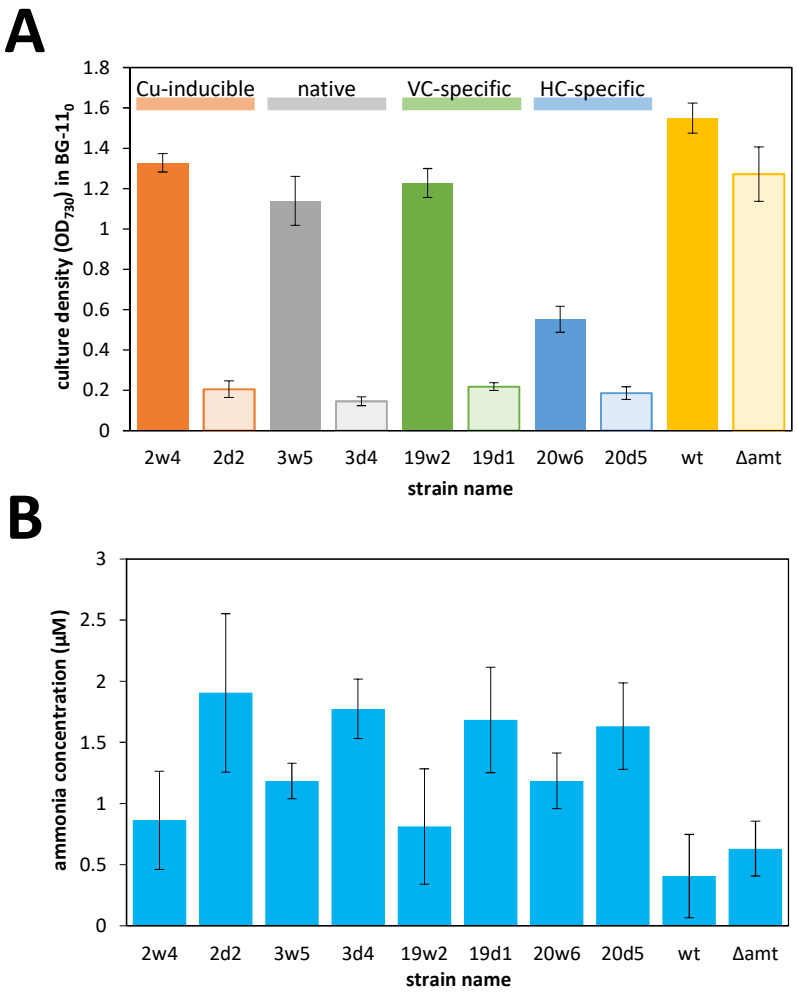


Figure 4.23. Long-term stability and ammonia excretion by the IF7A overexpression strains. Cultures were grown under diazotrophic conditions and continuous illumination. Samples were subcultured every week for up to a month. **(A)** Endpoint culture density was determined at 730 nm. **(B)** Ammonium was measured colorimetrically by the Willis micro-scale assay from cell-free supernatants. Colours encode for the different promoters driving the overexpression cassettes. Error bars display ±1 standard deviation of two biological replicates.

Although standard deviation of replicates was relatively high (large error bars in Figure 4.23B), a clear trend could be observed with statistical significance. According to this trend, concentration of extracellular ammonia was higher in all Δ*amt* strains compared to their wild type-based pairs, as

hypothesized. The elevated extracellular ammonia in case of the Δ*amt* strains may indicate that a larger portion of fixed nitrogen is lost to these strains, and therefore these strains experience the severe growth deprivation discussed for Figure 4.23A. Overall, ammonia was detected at very low levels for all strains, ranging between 1.9 and 0.4 µM for strains 2d2 and the wild type control (Figure 4.23B), respectively. In comparison to the values recorded at the end of the 3-day growth assay in Figure 4.21A, ammonia concentrations were 10-times lower on average in this experiment. The reason may be that ammonia has been lost due to the longer incubation period. Actually, because of regular subculturing, the majority of ammonia from previous rounds has been washed off by the time of the measurement. Therefore, measurement values reflected ammonia accumulation of one week. In addition, the week-old cultures in Figure 4.21 were inoculated with healthy cyanobacterial cells grown on nitrate. In contrast, cells during the long-term test had been growing in diazotrophic conditions for 3 weeks prior to the last week of ammonia accumulation.

Further to the above, at higher pH levels (pH 9 and above), the volatile ammonia, which is the dominant form of NH_3–N at that pH (Figure 4.1), may have gradually gassed out from the aqueous phase. Such pH levels may easily be the consequence of biological activity in a week-old culture. Indeed, cyanobacterial growth has been reported contributing to an increase to pH 9 or above (Ward, 1985), and thus gradually remove dissolved ammonia from the medium. Nevertheless, any ammonia that remained soluble in the medium under those conditions was also biologically accessible, potentially enabling the growth of a co-cultivated alga, as shown in Figure 4.22.

4.2.4 Strategy C – Overexpression of *gifA* from a neutral site and knockout of *nsiR4*

Stable ammonia accumulation by IF7A overexpression strains above can be assured through continuous selection pressure. These strains express *gifA* from an expression construct that requires undisturbed maintenance by supplementation of the appropriate antibiotics to the culture medium. Antibiotics, however, may potentially hamper the growth of a partner, like a non-diazotroph cyanobacterium or microalga. In addition to the extra costs of culturing, any long-term or large-scale incubation may become unstable due to fluctuations in antibiotic concentrations. It was therefore investigated whether *gifA* can be stably and efficiently overexpressed from a genomic locus in *Anabaena* sp. PCC 7120. In parallel, the non-coding RNA gene *nsiR4* was knocked out (Mitschke et al., 2011). In *Synechocystis* sp. PCC 6803 it has been shown that NsiR4 targets the mRNA of *gifA* and modulates the expression of IF7 (Klähn et al., 2015); however, a similar effect in *Anabaena* sp. PCC 7120 has not been investigated. Although a homologous gene has been identified in *Anabaena* sp.

PCC 7120 as well (Klähn et al., 2015), the pronounced differences between the two NsiR4 forms in the two cyanobacteria has left the question unanswered. In case there was a similar function for NsiR4 in *Anabaena* sp. PCC 7120, ammonia accumulation by an IF7A overexpressing strain would benefit from the removal of *nsiR4* as well.

4.2.4.1 Cloning strategies and assembly of constructs

The gene encoding IF7A was integrated to a neutral site and expressed under the *petE* promoter for two reasons. First, P_{petE} is inducible and thus can be switched on and off as needed. Second, this promoter has been shown to successfully accumulate ammonia as part of an overexpression construct (Figure 4.22 and Figure 4.23B). The P_{petE}-*gifA*-T_{gifA} insert was fused with the *aadA* cassette conferring resistance to Sm and Sp antibiotics and flanked with homologous sequences upstream and downstream the target genomic location. Figure 4.24A shows details of the construct designed to integrate the copper-inducible *gifA* cassette to the *nuiA-nucA* neutral site (Olmedo-Verd et al., 2006), located on the α megaplasmid of *Anabaena* sp. PCC 7120.

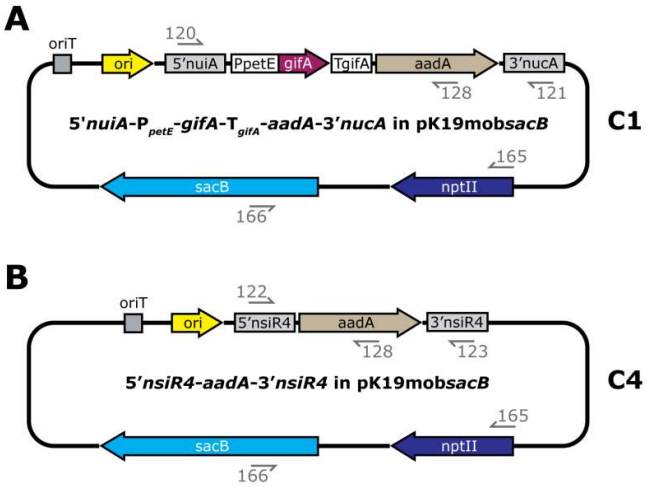


Figure 4.24. DNA constructs for *gifA* overexpression and *nsiR4* knockout. Constructs for *gifA* overexpression **(A)** and removal of *nsiR4* **(B)** were prepared by Gibson isothermal assembly using primers with compatible overhangs. Inserts contained the *aadA* cassette (encoding Sm^R and Sp^R), and they were flanked with 1-kb homologous regions to the target site (grey boxes). The mobilizable plasmid backbone (non-replicative in cyanobacteria) also carried the *sacB* marker for sucrose counter-selection (light blue arrow) to allow evolutionary pressure for double recombinants. Binding sites for primers used in colony PCR are indicated by grey arrows. The *nptII* cassette (dark blue arrow) encoding Km^R and Nm^R was used only in *E. coli* to maintain the construct.

Two 1-kb regions flanking the target site on the α megaplasmid were amplified from high-purity genomic DNA of *Anabaena* sp. PCC 7120 using primers 72+21 for 5'*nuiA* and primers 22+73 for 3'*nucA* (grey boxes in Figure 4.24A). The promoter region of *petE* was amplified using primers 24+25, and the *gifA*-T_{gifA} sequence was produced using primers 26+27 from genomic DNA as template (white boxes). Finally, the Sm^R cassette was amplified from plasmid pDF-lac with primers 28+29 (*aadA*, brown arrow). All resulting fragments contained an 8–10-nucleotide (nt) compatible flank to their neighbouring fragments, creating overlaps of 16–20 nt in size. In addition, the mobilizable pK19mob*sacB* vector was linearized using primers 70+71, and the six fragments above were assembled via Gibson isothermal reaction to create construct C1 (Figure 4.24A). Table 4.4 lists all primers and DNA fragments used in the assembly.

Table 4.4. List of oligonucleotides, constructs and strains for *gifA* genomic overexpression and *nsiR4* knockout. Sequences are given in direction from 5' to 3'. Lower-case letters indicate overlapping sequences.

Primers

ID	Name	Sequence
21	5'nuiA-R	ctgagtactgTCAAGTTTCCACAACTTTAGTAG
22	3'nucA-F	tggtgagaatCCTATTTACTAATTATCAACTTTACTCTC
24	PpetE-F	ggaaacttgaCAGTACTCAGAATTTTTTGCTGAG
25	PpetE-R	gaatagacatGGCGTTCTCCTAACCTGTAG
26	gifA-F	ggagaacgccATGTCTATTCAAGAAAAATCTCG
27	gifA-R	gaagccatgaTTAACACCTGTTAGCAATTTATAG
28	SmR-F (pDF-lac)	caggtgttaaTCATGGCTTCTTGTTATGAC
29	SmR-R (pDF-lac)	agtaaataggATTCTCACCAATAAAAAACGC
64	5'nsiR4-pK19mobsacB-F	ggaaacagctatgacatgatTAGCAGAATACGCCAGTTAG
65	5'nsiR4-SmR-R	gaagccatgaAATCTCTAGCCTATTCTGC
66	3'nsiR4-SmR-F	gtgagaatccGGTTACTGTCTACCGTAGGC
67	3'nsiR4-pK19mobsacB-R	tgtaaaacgacggccagtgcCAACAGTCAACAGTCAATAGC
68	SmR-5'nsiR4-F	gctagagattTCATGGCTTCTTGTTATGAC
69	SmR-3'nsiR4-R	gacagtaaccGGATTCTCACCAATAAAAAACG

Primers (continued)

ID	Name	Sequence
70	pK19mobsacB-BB-F	CACTGGCCGTCGTTTTAC
71	pK19mobsacB-BB-R	ATCATGTCATAGCTGTTTCCTG
72	5'nuiA-pK19mobsacB-F	ggaaacagctatgacatgatCTTTGTTGTGTTGTTCATCG
73	3'nucA-pK19mobsacB-R	tgtaaaacgacggccagtgcAGAGCGTAGCTGATCGTC
120	nuiA-cPCR-F	ctgttgtctgaccaaaatgc
121	nucA-cPCR-R	tctatgtactctgggagtgg
122	5'nsiR4-cPCR-F	aactttggttttcagtacgg
123	3'nsiR4-cPCR-R	aaaagtcaaatcccataccc
128	SmR-cPCR-R	ttagcgcctcaaatagatcc
165	NmR-cPCR-F	gaacaagatggattgcacgc
166	sacB-cPCR-R	atccgcatttttaggatctccg

DNA fragments

ID	Description	Primers	Template
F1	5'*nuiA* targeting flank	72+21	wild type *Anabaena* sp. PCC 7120
F2	3'*nucA* targeting flank	22+73	wild type *Anabaena* sp. PCC 7120
F3	P_{petE} promoter	24+25	wild type *Anabaena* sp. PCC 7120
F4	*gifA*-T_{gifA}	26+27	wild type *Anabaena* sp. PCC 7120
F5	*aadA* cassette (used in C1)	28+29	pDF-lac
F9	*aadA* cassette (used in C4)	68+69	pDF-lac
F11	5'*nsiR4* targeting flank	64+65	wild type *Anabaena* sp. PCC 7120
F12	3'*nsiR4* targeting flank	66+67	wild type *Anabaena* sp. PCC 7120
F13	pK19mobsacB backbone	70+71	pK19mob*sacB*

Constructs

ID	Insert	Assembled from	Genotype
C1	5'*nuiA*-P_{petE}-*gifA*-T_{gifA}-*aadA*-3'*nucA*	F1, F2, F3, F4, F5, F13	pK19mob*sacB*:: 5'*nuiA*-P_{petE}-*gifA*-T_{gifA}-*aadA*-3'*nucA*
C4	5'*nsiR4-aadA*-3'*nsiR4*	F9, F11, F12, F13	pK19mob*sacB*:: 5'*nsiR4-aadA*-3'*nsiR4*

Strains

Name	Purpose	Genotype
1w	*gifA* overexpression from a neutral site in wild type	*nuiA*::P_{petE}-*gifA*-T_{gifA}::*nucA*, Sm^R, Sp^R
1d	*gifA* overexpression from a neutral site in Δ*amt*	Δ*amt*, *nuiA*::P_{petE}-*gifA*-T_{gifA}::*nucA*, Nm^R, Sm^R, Sp^R
4w	*nsiR4* knockout in wild type	Δ*nsiR4*::*aadA*, Sm^R, Sp^R
4d	*nsiR4* knockout in Δ*amt*	Δ*amt*, Δ*nsiR4*::*aadA*, Nm^R, Sm^R, Sp^R

Construct C4 for the knockout of *nsiR4* through double recombination (Figure 4.24B) was prepared via similar steps as construct C1. First, targeting flanks homologous to 1-kb regions upstream and downstream of *nsiR4* (grey boxes) were amplified from the wild type chromosome of *Anabaena* sp. PCC 7120 using primers 64+65 (for 5'*nsiR4*) and 66+67 (for 3'*nsiR4*). The *aadA* cassette was amplified with overlapping regions to the flanks from plasmid pDF-lac using primers 68+69 (brown arrow). Three fragments of the insert were assembled with pK19mob*sacB* plasmid backbone linearized with the same primers as for C1 (Table 4.4). In addition to C4, a complementation construct on an RSF1010-based expression vector for the restoration of *nsiR4* in a knockout mutant was also prepared (construct C5) but later has not been used (therefore its structure is not shown here).

Amplification products were purified from 1% agarose gel using a QIAquick Gel Extraction Kit (Qiagen Ltd., Crawley, UK) and quantified on a NanoQuant Plate (Tecan AG, Switzerland). Fragments were loaded into a Gibson isothermal assembly reaction in equimolar quantities and transformed to *E. coli* DH5α. Colonies were screened on LB-agar plates containing 50 μg/ml Km. Constructs C1 and C4 were purified from liquid cultures using a QIAprep Spin Miniprep Kit (Qiagen Ltd., Crawley, UK) and their sequence was confirmed by sequencing (Source BioScience, Nottingham, UK).

4.2.4.2 Isolation of double recombinants

Exponentially growing cultures of *Anabaena* sp. PCC 7120 wild type and Δ*amt* strains were transformed with both constructs, C1 for *gifA* neutral site overexpression and C4 for *nsiR4* knockout. Triparental conjugation was performed as described in 2.6.3. Resulting transformants were screened on BG-11 plates supplemented with 2.5 µg/ml Sm and Sp. The counter-selection on sucrose was applied as described in Cai and Wolk (1990) to provide selective pressure for double recombinants. In short, early exconjugants (after approximately two weeks on the conjugation filter) were grown in 1 ml liquid BG-11 containing 1.5 µg/ml Sm and Sp and sonicated for 2 × 8 min in a sonication bath, as described for C6 in 4.2.2.4 and Figure 4.10. Broken filaments were plated on BG-11 agar plates (supplemented with 2.5 µg/ml Sm and Sp) containing 5% sucrose. Single colonies were streaked onto fresh plates every 1–2 weeks and segregation status was monitored by colony PCR.

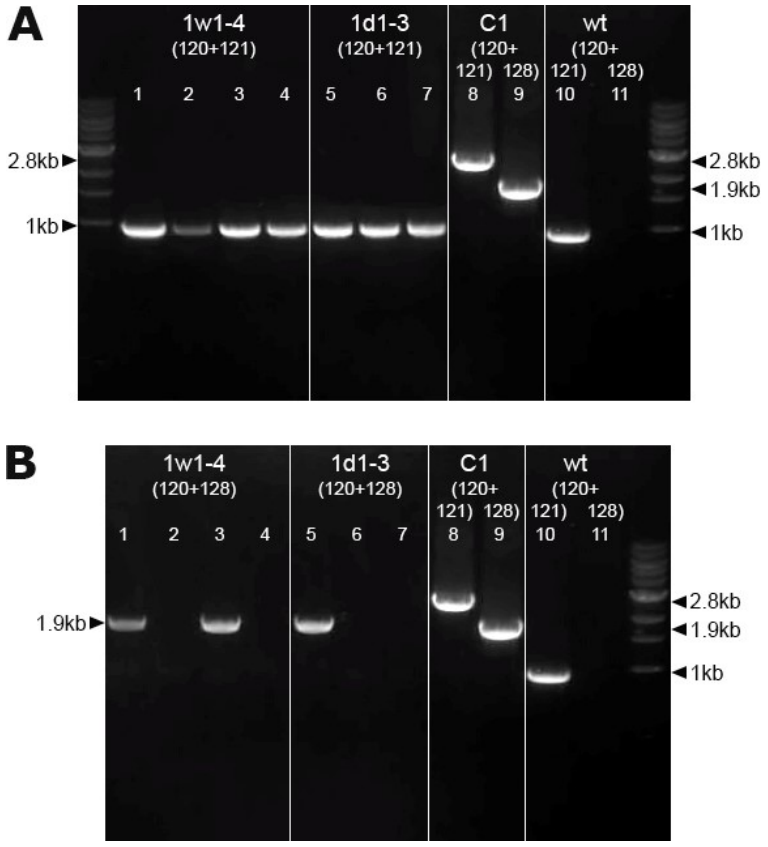


Figure 4.25. Colony PCR results of chromosomal *gifA* overexpression candidates. (A) Primers 120+121 amplify a 1-kb product from the wild type and a 2.8-kb product from a mutant. **(B)** Primers 120+128 amplify a 1.9-kb band specific to the mutant. Plasmid C1 and wild type genomic DNA were used as controls for mutant and wild type, respectively. Letter w or d in the names of isolates denote for their parental strain wild type or Δ*amt*, respectively.

A total of 48 isolates were picked from the original conjugation filter in groups of eight, brought to segregation and assayed in colony PCR. Figure 4.25 shows a typical gel image of a colony PCR test for strains transformed with construct C1. The 1-kb product matching the wild type control (lane 10) in panel A indicates that the integration of C1 to the *nuiA-nucA* site was either unsuccessful for all isolates (lanes 1–7) or the ratio of double recombinants was below detection limit for colony PCR. In order to prove the success of transformation, isolates were analysed using primers 120+128, a set specific to the mutation, as well (panel B). Isolate 1w1, 1w3 and 1d1 produced a band at 1.9 kb similar to the positive control C1 (lane 9) indicating that these isolates actually contained the desired insertion (lanes 1, 3 and 5 in panel B). At the same time, there was no product visible for the wild type control on lane 11, as expected.

Similar to the above, 48 isolates were screened for double recombinants of construct C4 as well. Primers 122 and 123 were used to detect both wild type and mutant genotype by amplifying a 1.1-kb and a 2.1-kb product, respectively. However, the 2.1-kb band indicating double recombination could not be detected in any of six cycles of segregation (over about three months).

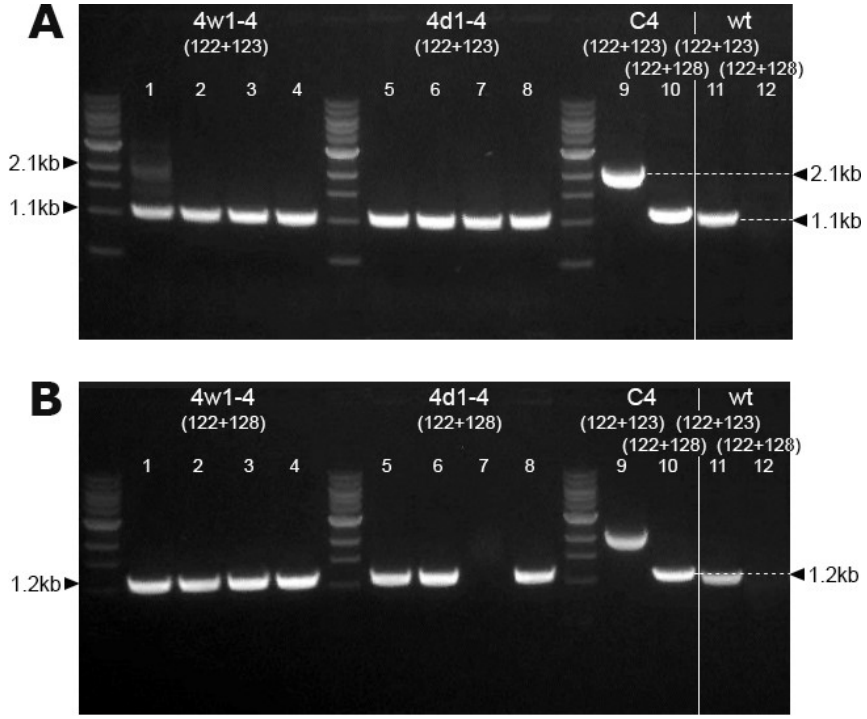


Figure 4.26. Colony PCR results of *nsiR4* knockout candidates. (A) Primers 122+123 amplify a 1.1-kb product from the wild type and a 2.1-kb product from a mutant. **(B)** Primers 122+128 amplify a 1.2-kb product specific to the mutant. Plasmid C4 and wild type genomic DNA were used as controls. Letter w or d in the names of clones denote for their parental strain wild type or Δ*amt*, respectively.

Figure 4.26 summarizes the results of the final segregation cycle for C4 transformants. The 1.1-kb band indicating presence of the wild type (or the lack of insert) was detected in all samples (lanes 1–8), including the control on lane 11. A very faint secondary band in case of isolate 4w1 could also be observed at the same height as for the mutant positive control (C4 plasmid, lane 9). Despite the lack of a secondary band at 2.1 kb for the other seven isolates, the mutant-specific product with primers 122+128 at 1.2 kb was detected in all samples, except 4d3 (panel B). These results suggest that wild type and mutant insertion sites coexisted in these strains, although their exact ratio could not be determined from these colony PCRs. Notably, a single recombinant would be indistinguishable from a successful double recombinant in a fully segregated strain with these primers. It was therefore hypothesized after several segregation rounds unsuccessful in the isolation of the desired double recombinants, that strains may actually possess the single recombinant genotype. Indeed, when tested with a primer set specific to the cargo plasmid (165+166) the single recombinant genotype could immediately be detected. Primers 165+166 amplify a 1.6-kb product on the plasmid backbone (Figure 4.24), and indicate the presence of the *nptII* and *sacB* cassettes. Figure 4.27 shows colony PCR results for both C1 and C4 isolates.

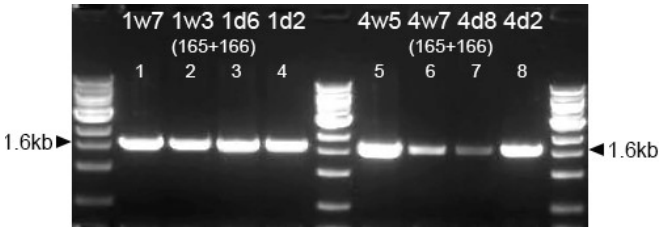


Figure 4.27. Appearance of single recombinants in sucrose counter-selection. Two isolates of each strain were tested for the absence of plasmid backbone after transformation. A 1.6-kb band indicates the presence of neomycin resistance (*nptII*) and sucrose sensitivity (*sacB*) cassettes.

Although it was counter-selected against in the presence of sucrose, all isolates contained the 1.6-kb sequence characteristic to single recombinants. One explanation may be that the *sacB* cassette has undergone inactivation in the recipient strain during conjugation. In any case, helper *E. coli* cells used to donate the cargo plasmid in conjugation showed over 99% lethality when grown on 5% sucrose in combination with the appropriate antibiotics (tested experimentally). In other words, the *sacB* cassette was intact until the point of conjugation with the cyanobacterium. However, the level of lethality of *sacB* in sucrose-supplemented *Anabaena* sp. PCC 7120 is unknown (Cai and Wolk, 1990). It might be that a small number of *sacB*-containing cells (in this case: single recombinants) can survive

when surrounded by healthy cells (in the filament, for example). In addition, *sacB* inactivation has also been reported in several cases (Pelicic et al., 1996; Wu and Kaiser, 1996; Lalioti and Heath, 2001; Choi and Schweizer, 2006; Marx, 2008). It is therefore critical to distinguish single recombinants from double recombinants from the very beginning. To this end, in case of construct C9 (for the removal of the entire schizokinen gene cluster; see Chapter 5 for details) a different method was applied using primers that amplify from just outside the affected site. Using such primers, wild type, single recombinant and double recombinant genotypes can be distinguished from one another by band size in a single colony PCR. The method is described in more detail in 5.2.4.

In conclusion, no double recombinants could be isolated for the genomic overexpression of *gifA* and the knockout of *nsiR4* due to difficulties in sucrose counter-selection. Transformants that contain the desired insert, however, have been isolated, most likely bearing the genomic arrangement typical for a single recombinant. Supernatants of these candidates were also tested for ammonia accumulation in a 5-day growth assay under diazotrophic conditions; however, no ammonia could be detected in any of the samples (results not shown).

4.3 Discussion

The metabolic model of *Anabaena* sp. PCC 7120 highlighted the central importance of ammonia in the nitrogen metabolism of this organism (Figure 4.2). Concentration of ammonia in the intracellular space is tightly controlled for two reasons: (1) reduction of ambient nitrogen (nitrite, nitrate, but especially dinitrogen) consumes large amount of energy (Luque and Forchhammer, 2008) that is wasted if ammonia escapes the cell, and (2) high levels of free ammonia pose a risk of PSII photodamage (Drath et al., 2008; Dai et al., 2014). The pool of free ammonia is governed by the synchronised action of three major enzymes, GS, GOGAT and GDH (Figure 4.2). The primary assimilator of ammonia in cyanobacteria is the GS–GOGAT cycle, in particular, the activity of GS (Muro-Pastor et al., 2005). Although GDH also assimilates ammonia, the ratio at which GDH incorporates nitrogen into carbon skeleton only becomes comparable to that of GS at later stages of culture growth, when light starts limiting (Chávez et al., 1999; Luque and Forchhammer, 2008). In the heterocyst under diazotrophic conditions, however, the GS–GOGAT cycle is not operative due to the lack of GOGAT expression (Martin-Figueroa et al., 2000). Thus, the L-glu consumption of GS is met by a transport from neighbouring photosynthetic vegetative cells.

The continuous enzymatic action of GS in the heterocyst assures the rapid incorporation of ammonia produced by nitrogen fixation. It has been shown, however, that reduction of GS activity results in elevated levels of free ammonia and ultimately, excretion of fixed nitrogen to the environment (Kannaiyan et al., 1994; Colnaghi et al., 1997; Srivastava and Amla, 2002; Zhang et al., 2007; Chaurasia and Apte, 2011; Grizeau et al., 2016). Ambient pH has crucial impact on the availability of leaked ammonia in such scenarios. The reason being ammonia is a weak base and rapidly receives a proton at pH values below 9.25 (Figure 4.1). While the basic form, ammonia can freely diffuse, biological membranes present an impenetrable barrier against the translocation of the protonated form, ammonium (Boogerd et al., 2011). Cyanobacteria have evolved the *amt*-family transporters to scavenge for lost ammonium. Similar to AmtB in other bacteria, the *amt* cluster in cyanobacteria is responsible for the high-affinity active transport of ammonium at low concentrations (Vázquez-Bermúdez et al., 2002; Paz-Yepes et al., 2008; Boogerd et al., 2011). However, the removal of this cluster alone did not result in detectable ammonia excretion in diazotrophically grown cultures of *Anabaena* sp. PCC 7120 (Figure 4.3A, purple columns).

Inhibition by MSX

The role of MSX as an inhibitor of GS activity is well documented in many organisms (Ronzio et al., 1969; Shanmugam et al., 1978). Several studies observed the impact of MSX on enzyme activity and growth of cyanobacteria (Thomas et al., 1990; Singh and Tiwari, 1998). In the case of *Anabaena* sp. it has also been shown that GS inhibition can be gradually increased by administration of MSX, up to the point at about 500 µM where the inhibition by MSX reaches lethality (Thomas et al., 1990). Under diazotrophic conditions the effect of MSX is even more severe to growth, constraining culture density at about 20–25% of that of an untreated culture (Figure 4.3A, blue columns). Supplementing the medium with MSX at an initial 55 µM, however, it was possible to culture both wild type and Δ*amt* strains of *Anabaena* sp. PCC 7120 in a nitrogen-free medium (BG-11_0). Both strains excreted considerable amounts of NH_3–N at the expense of culture density (Figure 4.3A, purple and blue columns). Surprisingly, growth of the wild type strain was more affected than that of the Δ*amt* mutant, by a factor of two. This is in good agreement with the finding that Δ*amt* mutants exhibit increased GS expression, by about the same factor (Figure 4.16, all column pairs of the same colour).

However, the lower sensitivity of the Δ*amt* strain to MSX may not be the only reason. Toxicity of free ammonium in the medium may also be different for the two strains. In wild type *Synechocystis* sp. PCC 6803, 450 µM extracellular free ammonium has been shown to decrease PSII activity by 37% due to

accelerated photodamage (Drath et al., 2008). Concentration of free ammonium in the wild type strain in Figure 4.3A was detected at a comparable level (~200 µM) that is expected to have a proportional impact on PSII activity. In contrast, the Δ*amt* strain is effectively incapable of ammonium uptake and therefore, it may be less affected by the toxicity of elevated levels of free ammonium.

Although it was possible to continuously produce ammonia by the administration of MSX to cultures growing on nitrate (Figure 4.3B), the increase of extracellular ammonia concentration plateaued after about 6 days. During this time the density of inhibited cultures showed no significant change, suggesting that the level of MSX was high enough to allow only cell maintenance, but not growth (Figure 4.3B, blue curves). It is unclear, however, whether the slow elimination of the inhibitory effect of MSX observed between day 6 and 7 (Figure 4.3B) is due to slow enzyme turnover or the lack of transfer of the inhibitor from a degraded enzyme to a new molecule (Lamar, 1968). Nevertheless, continuous supplementation of drug-like MSX is not a viable strategy in long-term cultivation for two reasons. First, it is very difficult to keep MSX concentration constant (a prerequisite for constant ammonia excretion), and second, any co-cultivated strain relying on the excreted ammonia would also be affected by MSX.

Recently, the soil-dwelling aerobic bacterium, *Azotobacter vinelandii* has been engineered successfully for ammonia excretion by specific repression of the organism's GS activity (Ortiz-Marquez et al., 2014; Ambrosio et al., 2017). In this work, three parallel, but distinct attempts were made to achieve the same goal (Figure 4.4A–C). The first strategy (strategy A, section 4.2.2) altered the activity profile of GS in *Anabaena* sp. PCC 7120, by mutating *glnA*. The active site of GlnA was targeted at the 52nd amino acid position to decrease its specific activity to ammonia (Crespo et al., 1999). The D52 residue (L-asp) of the encoded protein was switched to a serine amino acid by site-directed mutagenesis of the *glnA* gene.

Modification of the GlnA active site

Two molecular biology approaches were performed to replace wild type *glnA* for the active-site mutant version, *glnA*[p.D52S]. The first approach required disruption of the *glnA* cassette resulting in glutamine auxotrophy, and subsequent restoration of GS activity by the active-site mutant gene in a second transformation. The second approach was designed to directly replace wild type *glnA* for the mutant version in two subsequent homologous recombinations, of which the second is a *sacB*-induced intrachromosomal rearrangement. This approach may result in the desired mutant genotype only in 50% of the cases (as explained in Figure 4.6). The advantage of the first approach over the second is

that, even though glutamine auxotrophs are generated in the first step, all isolated prototrophs are expected to be *glnA*[p.D52S] mutants.

However, there have been difficulties isolating GS knockouts in *Anabaena* sp. PCC 7120. As predicted by the stoichiometric model (Figure 3.10, red dot), *glnA* was found essential under experimental conditions, and no knockout candidates could be isolated. Only a few exconjugants appeared on the conjugation plates following the transformation with C7 (a knockout construct for the entire *glnA* cassette, Figure 4.5A), and these colonies were lost over the course of segregation. This observation suggests, the gradual loss of *glnA* function was severely detrimental, and eventually, lethal to transformants of *Anabaena* sp. PCC 7120.

It is also possible the 0.2% L-gln supplementation reported for the screening of *E. coli* glutamine auxotrophs (Bloom et al., 1978; Rothstein et al., 1980) was insufficient for these strains. Therefore, in a repeated conjugation, L-gln supplementation was increased to 0.8%, a more reasonable level for cyanobacteria (Mérida et al., 1992). Nonetheless, no transformants could be isolated that carried construct C7. *Anabaena* sp. PCC 7120 is known to express several amino acid transporters for the translocation of L-gln over the membrane (Montesinos et al., 1995; Picossi et al., 2005; Pernil et al., 2008), and thus it is unlikely the (un)availability of glutamine has hindered the isolation of *glnA* mutants. In fact, this may suggest that *glnA* in *Anabaena* sp. PCC 7120 is not only essential for GS activity (Figure 3.10), but there are other reasons for its importance to growth that are unknown at this time.

The second approach generated a good number of initial transformants of single recombinant genotype (Figure 4.6, genotype A and B). Initially, it was problematic to use the allele-specific primer (primer 76, Figure 4.8) for selective detection of the favoured genetic arrangement (Figure 4.6, case A). The reaction was incapable of differentiating wild type and mutant genotypes even at optimized annealing temperature, although such allele-specific primers have been used with success previously (Gaudet et al., 2009; Ortiz-Marquez et al., 2014). It is also worth noting that commercial DNA polymerases (Q5 High-Fidelity DNA Polymerase from NEB in this case) possess strong proofreading activity that can remove the 3' unpaired overhanging nucleotides. It may be that the polymerase itself 'repaired' the wobbly primer and therefore amplification became immediately possible, even from wild type DNA (Figure 4.9C). To avoid such scenarios, some sources recommend the inclusion of a destabilizing mismatch within five bases of the 3' end of the allele-specific primer, also to increase specificity (Gaudet et al., 2009), which was not implemented for primer 76.

Instead of further investigating the unexpected behaviour of primer 76, a less error-prone primer pair was designed (7+8). These primers were used to simultaneously detect wild type and mutant genotypes, and segregation was continued until wild type was not detectable by colony PCR. The entire segregation process turned out to be extremely laborious and time consuming, taking about 8 months. Over this period candidates of *glnA*[p.D52S] were sonicated several times in liquid culture to enhance genetic segregation. Among the four different treatment conditions evaluated (Figure 4.10), two times 8-min bursts of sonication were found the most efficient to fragment *Anabaena* sp. PCC 7120 filaments to pieces containing 3–6 cells. Candidates showing the band pattern of (nearly) segregated single recombinants (similar to Figure 4.9D) were picked for selection on sucrose-containing medium, while the antibiotic concentration was also increased.

Although a selective medium supplemented with 5% sucrose should be lethal for strains containing the *sacB* gene, single recombinants were detected even after counter-selection on sucrose (Figure 4.11A, lanes 1, 11, 12 and 14). A single recombinant strain of C6 is expected to carry the entire cargo plasmid and therefore, the counter-selective *sacB* gene as well. For this particular marker, however, partial or complete loss of activity has been reported previously due to alleged point mutations in the *sacB* CDS, altered gene orientation or inactivation of its promoter (Pelicic et al., 1996; Wu and Kaiser, 1996; Lalioti and Heath, 2001; Choi and Schweizer, 2006; Marx, 2008). Nevertheless, the *sacB* cassette was repeatedly confirmed as being active in the conjugation helper strain (*E. coli* HB101) carrying the C6 cargo plasmid at the time of conjugative transformation. Therefore, it remains an open question, whether the *sacB* gene has suffered any mutation disrupting its activity in the single recombinant isolates of the C6 strain lines. Remarkably, the *sacB*-mediated isolation of double recombinants for constructs C1, C4 and C9 experienced similar difficulties. These challenges may suggest the existence of an unknown factor in sucrose counter-selection of *Anabaena* sp. PCC 7120, not dealt with by current protocols (Cai and Wolk, 1990; Marx, 2008).

To aid the isolation of *glnA* mutants for *Anabaena* sp. PCC 7120, it may be worthwhile applying L-gln supplementation at the conjugation step and onwards. The strong disadvantage of a mutation in an essential gene (such as in case of the *glnA*[p.D52S]) may result in an evolutionary pressure to keep copies of the more efficient wild type version. Under such circumstances a mutant strain will remain of mixed genotype and contain wild type contamination as long as the respective copy number of the antibiotic cassette introduced with the mutation is sufficient to cope with the applied selective pressure (antibiotic concentration). However, during L-gln supplementation such evolutionary advantage of the wild type version would not be relevant, and genetic segregation may proceed at a larger pace.

Despite the described challenges in segregating *glnA* mutant strains, it was possible to isolate promising candidates in both wild type and *amt* knockout *Anabaena* sp. PCC 7120. These candidates on Figure 4.11 are either fully segregated single recombinants (6w6#1, 6w7#4), or fully segregated double recombinants (6d4#1, 6d4#2, 6d4#3, 6d5#2 and 6d5#4). The segregated single recombinants (Figure 4.6, genotype A and B) would require further rounds of counter-selection on sucrose to undergo the desired intrachromosomal rearrangement and lose the sequence of the cargo plasmid. Among the segregated double recombinants 6d4#2 and 6d5#4 have been confirmed to contain *glnA*[p.D52S] (Figure 4.6, genotype C) and are thus expected to display lowered GS activity under standard growth conditions. The rest of the double recombinants (i.e. 6d4#1, 6d4#3 and 6d5#2) may either express *glnA*[p.D52S] or be wild-type revertants (Figure 4.6, genotype C and D). Sequencing the corresponding region could explicitly determine whether these strains are successful *glnA* active-site mutants. Nonetheless, reisolated colonies of 6w6#1, 6w7#4, 6d4#2 and 6d5#4 were found to exhibit a fair level of ammonia excretion in a 4-day diazotrophic growth assay, up to 80 µM ammonia in case of a clone of 6w6. Moreover, all but three of the evaluated strains displayed a fewfold improvement over the wild type strain in terms of ammonia excretion (Figure 4.12). These strains are promising candidates for the excretion of fixed nitrogen, and would therefore be worth testing in co-cultivation with non-diazotrophs.

IF7A overexpression

The second engineering strategy (strategy B, section 4.2.3) focused on the inducible and cell-specific overexpression of the *gifA* gene. The small oligopeptide encoded by the *gifA* gene (IF7A) is responsible for the highly specific inactivation of GS when fixed nitrogen (e.g. ammonia or nitrate) is available, in order to prevent depletion of the intracellular glutamate pool due to the activity of GS (Galmozzi et al., 2010). The expression of IF7A is under the control of the transcription factor NtcA (Herrero et al., 2001; Muro-Pastor et al., 2005) that binds to the promoter of IF7A, P_{gifA}. Under nitrogen starvation, IF7A expression is repressed by NtcA. Therefore, three additional promoters were evaluated for the overexpression of IF7A and the consequent action of GS inactivation and ultimately, ammonia excretion. Two of these promoters are either unaffected or induced by the binding of NtcA (P_{petE} and P_{nifHDK}, respectively), whereas the third one (P_{rbcLS}) is repressed by NtcA under certain conditions (Herrero et al., 2004).

The *gifA* overexpression constructs were assembled in a self-replicative plasmid derived from the broad-host-range vector pRSF1010 maintained in both bacteria (*E. coli* in this case) and most

cyanobacteria, including *Anabaena* sp. PCC 7120. Altogether four constructs were created, different only in the promoter the IF7A expression cassette (*gifA* CDS and terminator) is driven by. Although each construct contained a 366-bp (base pair) homologous region to the *Anabaena* sp. PCC 7120 chromosome, no genomic integration was expected due to the fact that such events are very unlikely for homology regions shorter than 1000 bp (Cai and Wolk, 1990). Each plasmid carried an antibiotic resistance cassette; successful transformants were isolated under selective pressure of the respective antibiotic. Due to the lack of genomic integration, antibiotic selection was maintained continuously to retain the expression vector in the IF7A overexpression strains. Such strains requiring continuous supplementation of the appropriate antibiotics are not genetically stable. In addition, the use of antibiotics restricts the scope of potential partner organisms, as these symbionts are also required to tolerate the selective agent. The plasmid-based overexpression of IF7A has, however, a clear advantage over genomic integration: strain isolation does not require the extremely time-consuming segregation of genotypes, but a single-step antibiotic selection. The strategy was therefore used as a relatively quick, proof-of-principle study to evaluate the effect of different promoters on IF7A overexpression and, ultimately, on ammonia excretion via GS inhibition.

In total, 8 strains were isolated for the overexpression of IF7A. The isolated strains were different in the promoter under which IF7A was overexpressed and in the parental origin of the isolate (wild type or *amt* knockout strain). The overexpression strains showed similar growth characteristics, filament structure and culture health in a 5-day growth assay on nitrate (Figure 4.14A–B). Although actual culture density data showed significant variation, the difference could be tracked back to density variations of the original inoculum cultures. Strain pairs bearing the same overexpression construct but being of different parental origin behaved fairly similarly. However, in case of 3w5 (*gifA* under P_{gifA} in wild type *Anabaena* sp. PCC 7120) for example, the growth rate was considerably lower than in the other cases throughout the assay. In addition, the density of the 3w5 inoculum culture alone cannot account for the ~64% difference from its strain pair, 3d4 (Δ*amt*-based strain, identical expression cassette, grey data series in Figure 4.14A).

It is worth noting, under the conditions tested (i.e. in the presence of nitrate) the P_{gifA} promoter is derepressed by NtcA. Under these conditions the 3w5 and 3d4 strains are expected to display enhanced IF7A expression from the plasmid construct, as well as the native gene, leading to decreased GS activity compared to the wild type level. Interestingly, however, strain 3w5 and its pair 3d4 were not affected the same way, most likely due to the increased GS expression in 3d4 that was observed for all Δ*amt*-based strains (Figure 4.17). Among the rest of the overexpression constructs 2w4 and 2d2 may have been active due to the copper content of standard BG-11 medium. Strains bearing the P_{rbcLS}-

gifA construct (19w2 and 19d1) may have also experienced increased IF7A expression, but not strains 20w6 and 20d5 (with promoter P_{nifHDK} that is only activated in mature heterocysts). Nevertheless, growth characteristics of strain pairs of this group were very similar within pairs; suggesting that maintenance of the expression construct is not a significant burden when cultivated on a combined nitrogen source.

In contrast, all mutant strains showed obvious deprivation of growth after 5 days of diazotrophic cultivation (Figure 4.15A). Although filament health, heterocyst frequency and pattern for the mutants were indistinguishable from those of the wild type (Figure 4.15B), culture densities at the end of the 5 days were significantly lower in case of the overexpression strains, especially for those of Δ*amt* origin. The difference between the wild type and wild type-based mutants may be credited to the extra costs associated with the maintenance of the expression construct, including coping with the supplemented antibiotics. It is very difficult, however, to discover any significant pattern in the order of wild type-based mutants. Strain 19w2 showed the highest growth among these strains, in good agreement with the fact that P_{rbcLS} is inactive in heterocysts, where ammonia accumulation could take place under diazotrophic conditions. The other three strains behaved similarly to each other, although P_{nifHDK} in 20w6 is expected to be active, also P_{petE} in 2w4 to some extent, but not P_{gifA} in 3w5. Although the actual order of culture densities corresponded to the anticipated state of IF7A overexpression (i.e. active or inactive), the statistical difference between the measured values was not significant.

IF7A overexpression mutants were further characterized by determining GS activity and quantifying the enzyme from cell-free extracts of diazotrophically grown cultures. Both activity (Figure 4.16) and concentration data (Figure 4.17) for GS confirmed the elevated level of GlnA in all Δ*amt* strains that was previously unknown. The difference between wild type and Δ*amt* pairs was 2–4-fold, with the P_{gifA} strains being the least and the P_{nifHDK} strains being the most distinct. This remarkable difference may provide partial explanation to why the Δ*amt* strains showed severe growth deficiency in the 5-day growth assay (Figure 4.15A). In these strains the enhanced GS activity may deplete the intracellular glutamate pool, leading to carbon limitation and shifted C/N balance. In fact, under these circumstances, IF7A overexpression could be beneficial in retaining the C/N balance of the cell. However, GS and IF7A are not the only factors affecting C/N ratio, and it is more likely that the increased GS expression in these strains is already a sign of the organism's response to a disturbed nutrient balance. The higher GlnA concentration in the Δ*amt* strains was also observed under conditions that specifically activated the different overexpression constructs: addition of extra copper for the induction of P_{petE}, cultivation on nitrate for P_{rbcLS} and P_{gifA}, and diazotrophic growth for P_{nifHDK} (Figure 4.18). However, not even these conditions could increase the concentration of IF7A so that it

was detectable by triple quadrupole LC-MS/MS. The main reasons are the small size and low abundance of IF7A in the cell, as well as the unfortunate properties of the tryptic fragments generated from the oligopeptide. Spiking diazotrophic cultures with ammonium (nitrogen replete conditions) enabled the detection of IF7A from protein extracts of some strains (Figure 4.20B), but overall, the inactivation factor remained elusive in the applied analytical methods. It may be that the IF7A overexpression constructs require an initial boost by ammonium to overcome the barrier of continuous repression by NtcA under diazotrophic conditions. Once overexpression becomes feasible and IF7A starts inhibiting GS, the excess ammonium is assumed to leak out the cells. The continuous availability of ammonium in the medium may be sufficient to manage the expression of IF7A both from the native site and the overexpression construct.

Direct measurement of IF7A was very difficult by LC-MS. In addition, indirect measurements of the inactivation factor by determining GS activity and concentration were also inconclusive. Therefore, two additional methods were applied to assess the capability of each overexpression constructs to excrete ammonia under nitrogen starvation. First, free extracellular ammonia was quantified from culture supernatants. Ammonia production was highly dependent on culture density and it ranged between 12 and 24 µM. On a normalized scale, strains 20d5 and 2d2 excreted the most NH_3–N, whereas 3w5 and 19w2 were comparable to the parental strains (Figure 4.21A). Similarly, ammonia production by 20d5 and 2d2 following a 1-month long-term incubation was the highest (Figure 4.23B). The actual concentration values were, however, about 10-times lower than in case of the 3-day experiment (Figure 4.21A). There was a clear trend observed for the Δ*amt* strains, all of them being 30–240% more productive than their wild type pairs. Although loss of ammonia from the aqueous phase may be an important issue, it is also possible the excreted ammonia can be recaptured by organisms possessing high affinity ammonium transporters. This way ammonia is almost instantly captured and assimilated, preventing permanent loss to the gas phase at pH 9.25 and above.

The hypothesis that compatible microorganisms (that can tolerate *Anabaena* sp. PCC7120 and the antibiotics used to maintain the IF7A overexpression constructs) may be able to use the excreted NH_3–N was tested with the microalga *Chlorella vulgaris*. The 8-day diazotrophic co-cultures of this alga and selected IF7A mutants clearly showed that *Chlorella vulgaris* could very efficiently use excreted ammonia. Selected mutant strains 20d5, 20w6, 2d2 and 2w4 supported the growth of the alga at 66%, 38%, 28% and 6% of the control monoculture growing on nitrate, respectively (Figure 4.22). Moreover, the total biomass production of the mutant–*Chlorella vulgaris* co-cultures measured as optical density at 750 nm reached 105%, 95%, 61% and 45% of that of the algal monoculture (positive control) grown on nitrate for strains 2w4, 20d5, 2d2 and 20w6, respectively. Culture density of the wild type

Anabaena sp. PCC 7120–*Chlorella vulgaris* consortium was also very high (about 78% of the positive control); however, the corresponding *Chlorella vulgaris* CFUs per ml was very low, similar to the case of 2w4. These results indicated that wild type *Anabaena* sp. PCC 7120 was unable to support the growth of the alga, and that both the wild type and mutant 2w4 were dominating their corresponding co-culture. Although mutant 2w4 was feeding some fixed nitrogen to *Chlorella vulgaris*, this effect was much lower than in case of the other three co-cultures. Moreover, the inability of wild type *Anabaena* sp. PCC 7120 to support the growth of *Chlorella vulgaris* in contrast to the success of the selected mutant strains clearly indicated that survival of the alga under diazotrophic conditions is due to the genetic modification of the mutant strains. Furthermore, the ability to supply nitrogen was observed for mutant strains of both wild type and Δ*amt* origin, although more alga grew in co-cultivation with those lacking the high-affinity ammonium transporters (strains 2d2 and 20d5 in Figure 4.22B). Therefore, the overexpression constructs under the control of P_{nifHDK} and, to lesser extent, P_{petE} have been successfully confirmed as ammonia excreting strains that can provide fixed nitrogen to a co-cultivated partner.

The relationship of the two organisms can be considered as commensal, due to the lack of (known) mutualistic cross-feeding by the non-diazotrophic algal partner. For the sake of robustness and long-term stability of such a synthetic co-culture it may be necessary to introduce some sort of mutualism. Synthetic communities without the partners cross-feeding each other have been shown to suffer from instability, mainly by large alternations of partner ratio and community structure (Kazamia et al., 2012; Kazamia et al., 2014). In addition, prolonged co-cultivation for any practical application (e.g. production of a valuable metabolite) will require the elimination of antibiotics in the expression system. The third engineering strategy to modify the total activity of GS is aimed partially at this goal.

Chromosomal overexpression of IF7A and disruption of NsiR4

The third strategy (strategy C, section 4.2.4) involved the overexpression of IF7A at a neutral genomic locus and the removal of the putative repressor of *gifA* mRNA, NsiR4, in two separate mutant strains. The rationale behind a genomic IF7A overexpression strain is its considerably higher genetic stability and independence from the continuous use of selective antibiotics. For the co-cultivation experiments in Figure 4.22, a photosynthetic partner was specifically chosen for its insensitivity towards the antibiotics applied to maintain the IF7A overexpression constructs in the *Anabaena* sp. PCC 7120 mutant strains. Eliminating the need for selective antibiotics would, however, open the possibility to

use a wide range of partner organisms (including other cyanobacteria, heterotrophic bacteria and eukaryotic algae) at cultivation scales that would otherwise be impractical.

Based on preliminary data the construct driven by the P_{petE} promoter was chosen for overexpression from a demonstrated neutral site (Figure 4.24A). In parallel, a strain lacking the *nsiR4* gene was also generated for the alleged phenotype of increased IF7A expression. Although such an impact on GS inactivation factors in other cyanobacteria has previously been demonstrated (Klähn et al., 2015), the case was not trivial due to the somewhat different inactivation factor (IF7A) in *Anabaena* sp. PCC 7120. In a future strain, combination of the two mutations would be preferred under different inducible promoters, for achieving maximal ammonia excretion.

Promising candidates were successfully isolated for both strains, although it was not possible to complete genetic segregation. Accordingly, both strains showed characteristics of a mixed genotype containing wild type and mutant (single recombinant) DNA as well (Figure 4.25 and Figure 4.26), even after numerous rounds of segregation. In addition, counter-selection for double recombinants by the *sacB* gene was very inefficient. Following counter-selection on sucrose, single recombinants (containing the *sacB* gene) still dominated, and no double recombinants could be detected (Figure 4.27). Although a single recombinant may theoretically overexpress IF7A, no significant ammonia excretion could be observed for the candidate strains. It is possible, however, that continued segregation of the single recombinant type would increase extracellular ammonia to relevant levels.

5 Systematic study of the schizokinen operon and its genes

5.1 Introduction

By mass, iron is the second most abundant metal on Earth (Lutgens and Tarbuck, 2003), preceded only by aluminium. It is an essential element for the growth of most microorganisms and the fifth most abundant metal in their biomass (Chinnici et al., 1998). Intracellular iron is a cofactor in many biochemical processes, including nucleic acid synthesis, photosynthesis, cell respiration and even nitrogen fixation. Especially in the marine environment, however, iron is scarce and this resulted in the evolution of different metabolic strategies to capture this essential nutrient (Rudolf et al., 2015). Such strategies include the emergence of important microbial interactions, primarily in oceanic environments (Amin, 2010), but also elsewhere (Cuiv et al., 2006). These interactions are mainly driven by small organic compounds called siderophores. Under iron limitation, many microorganisms excrete siderophores to scavenge for iron in their environment (Neilands, 1995). Siderophores are high affinity chelators of the ferric ion (Fe^{3+}) and are produced in a wide range of chemical structures. Depending on the characteristic functional group, siderophores are divided into three main families: hydroxamates, catecholates and carboxylates. More than 500 different types of siderophores are known, of which 270 have been structurally characterized (Ahmed and Holmstrom, 2014). Their primary biological role is to scavenge iron, but many siderophores can form complexes with other essential elements (e.g. molybdenum, manganese, cobalt and nickel) in the environment, and make

them available for microbial cells (Bellenger et al., 2008). The ability of siderophores to bind a variety of metals in addition to iron has gained particular interest in environmental research (Braud et al., 2009). Their versatile chemical properties further increase the significance of siderophores as efficient bioremediation and chelation agents, biosensors and biocontrol compounds against pathogens (Ahmed and Holmstrom, 2014). Ferric ion chelates formed by siderophores are characterized by their exceptional thermodynamic stability. The formation constants of iron(III)-siderophores range from 10^{22} to 10^{49} (Albrecht-Gary and Crumbliss, 1998). At the same time, the affinity of siderophores towards complexing ferrous ions (Fe^{2+}) is relatively weak (Neilands, 1995). The most well-known siderophore is aerobactin, first isolated from *Aerobacter aerogenes* (Gibson and Magrath, 1969).

Several cyanobacteria excrete siderophores to scavenge for iron (III) ions in the surrounding medium. Iron uptake is of vital importance in cyanobacteria, because they generally require higher Fe:C ratios than non-photosynthetic bacteria (Keren et al., 2004; Nicolaisen and Schleiff, 2010). Indeed, iron starvation severely impacts cyanobacterial functioning and thus iron uptake is crucial especially under low iron conditions (Shcolnick and Keren, 2006). A variety of siderophores are synthesized in *Anabaena* sp. in response to iron deprivation (Goldman et al., 1983). However, schizokinen is the only siderophore that has been structurally characterized so far (Simpson and Neilands, 1976; Stevanovic et al., 2012). Siderophores use specific transporters to move across membranes and the cell wall. In *Anabaena* sp. PCC 7120, the TonB-dependent outer membrane protein, Alr0397 has been found responsible for the transport of schizokinen (Nicolaisen et al., 2008; Stevanovic et al., 2012). The genomic context of its gene, *alr0397* was proposed to harbour genes involved in the biosynthesis of this siderophore, however, the actual pathway has not been characterized yet (Jeanjean et al., 2008; Nicolaisen et al., 2008).

The biosynthesis and excretion of schizokinen, as well as the other siderophores produced by *Anabaena* sp. PCC 7120, are under the control of FurA (ferric uptake regulator), the global transcriptional regulator of iron homeostasis (Jeanjean et al., 2008; González et al., 2014). The negatively autoregulated FurA is induced under oxidative stress or iron deprivation (González et al., 2014). However, the modulator (repressor or activator) role of FurA is not limited to genes involved in oxidative stress response and iron homeostasis. Several major physiological processes of cyanobacteria have been suggested to depend on modulation by FurA. Most importantly, FurA downregulates the expression of NtcA, and thus influences heterocyst differentiation and nitrogenase expression. At the same time, NtcA upregulates FurA in early-stage heterocysts (López-Gomollón et al., 2007b; González et al., 2013). Simply on the basis of iron availability influencing nitrogen metabolism, it is worth investigating the impact of iron uptake on the potential of nitrogen excretion

under diazotrophic conditions. Further to this, iron has twofold importance as cofactor in both the photosynthetic and the nitrogen-fixing apparatus of *Anabaena* sp. PCC 7120.

Considering the above, iron may serve as a potent cross-feeding commodity between the nitrogen-excreting cyanobacterium and its commensal partner organism (see discussion in page 183). One possible scenario is the combination of a nitrogen-excreting strain deprived in its ability to scavenge iron with a eukaryotic algal partner providing chelated iron in an accessible form to the cyanobacterium. This approach requires the cyanobacterium to lack the excretion of any siderophore, while the alga needs to produce an iron chelator the cyanobacterium can use. Optimally, the siderophore to be removed from the cyanobacterium and to be added to the alga may be the same compound, schizokinen for example. Therefore, schizokinen-mediated iron acquisition in *Anabaena* sp. PCC 7120 was studied in further detail. In particular, functions of the biosynthetic genes and structure of the putative schizokinen cluster were investigated.

5.2 Results

The cluster incorporating genes *all0390–all0396* on the chromosome of *Anabaena* sp. PCC 7120 has been proposed to be involved in the synthesis of a siderophore (Jeanjean et al., 2008). Moreover, *alr0397* (*schT*), a gene exactly downstream to the hypothetical schizokinen cluster has been found to encode for an outer membrane transporter of schizokinen (Nicolaisen et al., 2008). Further sequence similarities of the genes in the cluster (except for *all0391*) to the biosynthetic genes of aerobactin in *E. coli* (de Lorenzo et al., 1986) and rhizobactin 1021 in *Sinorhizobium meliloti* 1021 (Lynch et al., 2001) suggest that these genes may be responsible for the biosynthesis of a hydroxamate-type siderophore, just like schizokinen. On the other hand, *all0391* shares similarity with *pvsC*, an inner membrane exporter of the siderophore vibrioferrin in *Vibrio parahaemolyticus* (Tanabe et al., 2006). The putative schizokinen operon consisting of seven genes was analysed by bioinformatics methods to predict the function of the encoded proteins. Furthermore, the corresponding 10.6-kb genomic region on the chromosome of wild type *Anabaena* sp. PCC 7120 was completely removed, and different versions of the cluster were expressed from an expression vector, each of them lacking another gene in the cluster. Mutant strains were analysed under iron limitation for changes in iron-stress response.

5.2.1 Bioinformatics study of the schizokinen cluster

As part of the reconstruction of the *Anabaena* sp. PCC 7120 metabolic network (Chapter 2.8), gene candidates have been identified for the production of schizokinen from aspartate-semialdehyde in cyanobacteria, based on sequence similarities to genes involved in the synthesis of aerobactin and rhizobactin 1021 siderophores in other organisms (de Lorenzo et al., 1986; Lynch et al., 2001). Nucleotide sequences from *Sinorhizobium meliloti* 1021 were searched against the *Anabaena* sp. PCC 7120 translated genome in the Reference Sequence (RefSeq) database of NCBI (National Center for Biotechnology Information) using the BLASTX program (Altschul et al., 1990). Preliminary BLASTX hits were used to identify and confirm the *all0390–alr0397* gene cluster in *Anabaena* sp. PCC 7120 (Figure 5.1, top row). Notably, no candidates were identified for *rhtX* and *rhrA*, a putative siderophore receptor and a transcriptional activator in *Sinorhizobium meliloti* 1021, respectively (Figure 5.1, grey arrows in the bottom row).

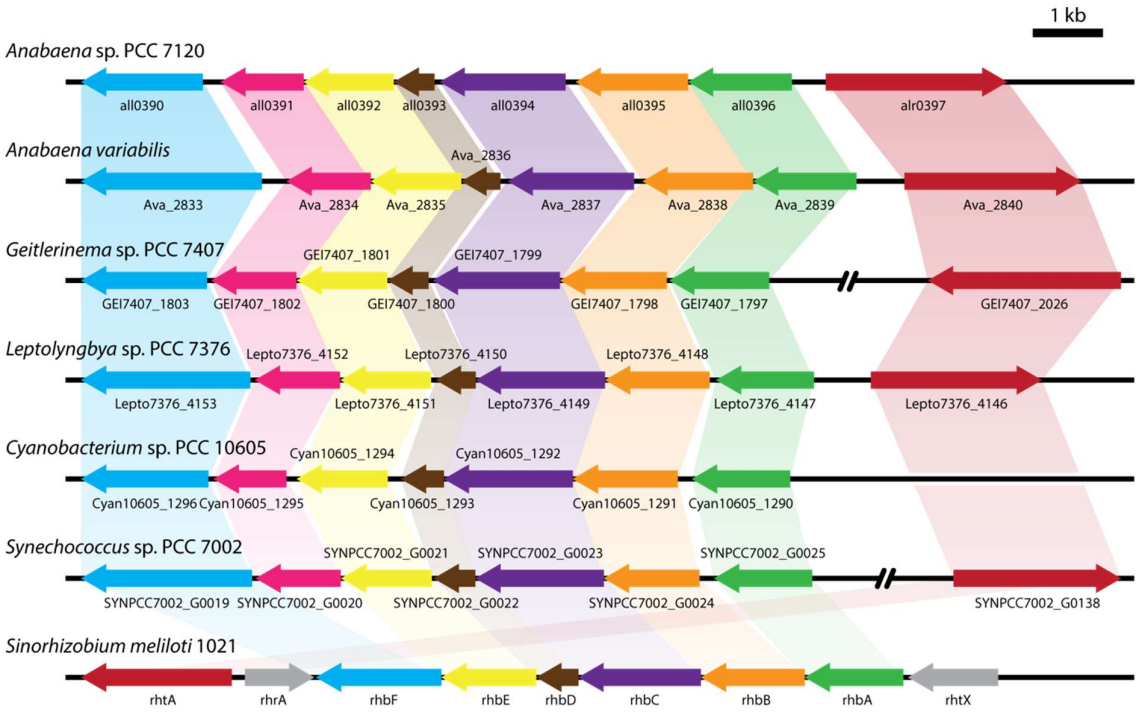


Figure 5.1. Conserved genes in the putative schizokinen pathway. Six cyanobacterial genomes were compared to the *rhb* gene cluster of the soil bacterium *Sinorhizobium meliloti* 1021, and similar genes were identified based on sequence homology. Genes and genomic distances are drawn to scale; a 1-kb marker is displayed in the top right corner.

Genes found in *Anabaena* sp. PCC 7120 were then submitted to reciprocal BLAST to assess conservation status of the newly identified cluster in other cyanobacteria. The genome of *Sinorhizobium meliloti* 1021 was also included as a reference. Reciprocal BLAST hits were evaluated based on their coverage of the query, statistical significance (e-value) and identity to the query sequence (Table 5.1). Together with *alr0397* (*schT*), eight genes related to schizokinen have been identified in many cyanobacterial genomes. The structure of the cluster, the genes and their putative protein products were found to be highly similar in most cyanobacteria. The first five organisms with highest similarity scores to *Anabaena* sp. PCC 7120 are displayed in Table 5.1 and also in Figure 5.1.

As expected, genes from *Anabaena variabilis*, a close relative of *Anabaena* sp. PCC 7120, showed the overall highest similarity to the query (over 90% identity with at least 97% coverage for all genes; Table 5.1, first block). Also not surprisingly, *Sinorhizobium meliloti* 1021 was found to be the least similar, lacking any homologous gene to *all0391*, which is otherwise present in all cyanobacterial genomes evaluated here. In general, candidate genes in *Synechococcus* sp. PCC 7002 and *Leptolyngbya* sp. PCC 7376 were found to be fairly similar to their *Anabaena* sp. PCC 7120 homologues, of which hits for *all0396* and *all0390* were the nearest to their queries. Interestingly, the corresponding gene cluster is located on a naturally occurring plasmid in case of *Synechococcus* sp. PCC 7002 and *Sinorhizobium meliloti* 1021, whereas the cluster is chromosomal in the other organisms investigated here.

The arrangement of the different genes within each cluster was fairly similar in all organisms. However, homologues of *alr0397* (*schT*) were found downstream of the biosynthesis cluster in the cyanobacterial genomes, in contrast to *rhtA* (encoding for a rhizobactin transporter in *Sinorhizobium meliloti* 1021), located upstream of the rhizobactin 1021 genes (Figure 5.1, red arrows). Moreover, in case of *Synechococcus* sp. PCC 7002 and *Geitlerinema* sp. PCC 7407, homologues of *schT* were identified 274 kb and 124 kb downstream of the cluster, respectively. Furthermore, the same sequence showed no significant similarity to any gene in the genome of *Cyanobacterium* sp. PCC 10605, even though candidates were identified for the rest of the cluster.

Table 5.1. Similarity scores of homologous genes in Figure 5.1 to genes in the putative schizokinen operon. Nucleotide sequences of the genes from *Anabaena* sp. PCC 7120 (queries) were searched in the indicated genomes using the NCBI BLASTX service. Hits are evaluated based on their coverage of (QC, %) and identity to (ID, %) the query sequence, and their statistical significance predicted by the software (e-value, EV). Capital E in e-values stands for scientific E-notation of the numeric value.

***Anabaena* sp. PCC 7120**	***Anabaena variabilis***			***Geitlerinema* sp. PCC 7407**			***Leptolyngbya* sp. PCC 7376**		
	QC	**ID**	**EV**	**QC**	**ID**	**EV**	**QC**	**ID**	**EV**
alr0397	97%	97%	0	92%	54%	0	93%	44%	0
all0396	99%	96%	0	94%	72%	0	91%	72%	0
all0395	99%	91%	0	91%	62%	0	90%	58%	0
all0394	99%	93%	0	99%	60%	0	96%	58%	0
all0393	98%	95%	8E-138	93%	69%	4E-91	97%	61%	6E-86
all0392	99%	93%	0	98%	63%	0	99%	59%	0
all0391	99%	92%	0	96%	62%	3E-153	94%	47%	1E-104
all0390	98%	94%	0	98%	74%	0	98%	69%	0

***Anabaena* sp. PCC 7120**	***Cyanobacterium* sp. PCC 10605**			***Synechococcus* sp. PCC 7002**			***Sinorhizobium meliloti* 1021**		
	QC	**ID**	**EV**	**QC**	**ID**	**EV**	**QC**	**ID**	**EV**
alr0397	n/a	n/a	n/a	93%	51%	0	80%	29%	3E-77
all0396	93%	71%	0	91%	74%	0	88%	60%	0
all0395	91%	58%	0	77%	62%	4E-177	88%	40%	2E-106
all0394	96%	59%	0	96%	58%	0	85%	39%	3E-114
all0393	97%	65%	6E-88	96%	59%	5E-83	93%	44%	2E-56
all0392	99%	64%	0	99%	58%	0	99%	46%	7E-139
all0391	92%	60%	4E-142	94%	47%	1E-105	n/a	n/a	n/a
all0390	99%	73%	0	98%	70%	0	97%	44%	4E-170

It is also worth noting that homologues of *rhbF* (including *all0390* in *Anabaena* sp. PCC 7120) displayed significant length variability. In fact, two versions of this gene were identified in the different organisms: a shorter one in *Anabaena* sp. PCC 7120, *Geitlerinema* sp. PCC 7407, *Cyanobacterium* sp. PCC 10605 and *Sinorhizobium meliloti* 1021, and a longer sequence in *Anabaena variabilis*, *Leptolyngbya* sp. PCC 7376 and *Synechococcus* sp. PCC 7002. Specific searches in the Pfam database (Finn et al., 2016) revealed that both versions contain a FhuF-like domain (from the ferric iron reductase family) and an IucA/IucC-like domain (for the biosynthesis of aerobactin in *E. coli*). In addition, the longer version also contains an Acetyltransferase-8 domain (Table 5.2). The rest of the genes in the cluster were found approximately equal in size in all organisms tested.

Table 5.2. Pfam motifs detected in the homologues of *rhbF* and *all0390*. Sequences were analysed using the Pfam database and search engine at http://pfam.xfam.org.

	Gene length	IucA/IucC family	FhuF family	Acetyltransf-8 domain
***Anabaena* sp. PCC 7120**	short	+	+	–
Anabaena variabilis	long	+	+	+
***Geitlerinema* sp. PCC 7407**	short	+	+	–
***Leptolyngbya* sp. PCC 7376**	long	+	+	+
***Cyanobacterium* sp. PCC 10605**	short	+	+	–
***Synechococcus* sp. PCC 7002**	long	+	+	+
***Sinorhizobium meliloti* 1021**	short	+	+	–

Predictions to the function of each gene in the schizokinen cluster were made from searching the Pfam database as above. Similar domains were detected for the rhizobactin 1021 genes in *Sinorhizobium meliloti* 1021 and the putative schizokinen cluster in *Anabaena* sp. PCC 7120. Therefore, it was postulated the two siderophores are synthesized via similar pathways in the two organisms, involving essentially the same intermediate metabolites. In *Sinorhizobium meliloti* 1021, it has been suggested the process starts from aspartate-semialdehyde (Lynch et al., 2001). In case of *Anabaena* sp. PCC 7120, aspartate-semialdehyde is produced from aspartate as an intermediate in the lysine biosynthesis pathway. Aspartate-semialdehyde also serves as a precursor in the biosynthesis of β-alanine. The putative steps of schizokinen biosynthesis pathway are drawn in Figure 5.2. In order to produce

schizokinen, aspartate-semialdehyde is first converted to diaminopropane by a transaminase (All0396, Pfam: aminotransferase class-III domain), followed by decarboxylation (All0395, Pfam: pyridoxal-dependent decarboxylase domain) and then hydroxylation by a monooxygenase (All0392, Pfam: L-lysine 6-monooxygenase, NADPH-requiring). The product hydroxyl-aminopropane is thereafter acetylated by an acetyltransferase (All0393, Pfam: acetyltransferase domain). The last two steps in the pathway are catalysed by the subsequent action of two ligases on the two *O*-acyl side-chains of a citrate backbone.

Table 5.3. Pairwise similarity of the last two ligases in the schizokinen and aerobactin pathways. Protein sequences were compared using the NCBI BLASTP service. TS: total score; QC: query coverage; EV: e-value; ID: identity.

	IucA				**IucC**			
	TS	**QC**	**EV**	**ID**	**TS**	**QC**	**EV**	**ID**
All0390	77.4	20%	0.001	24%	297	98%	2E-97	31%
All0394	231	98%	4E-72	29%	83.2	51%	2E-15	22%

Searching for Pfam motifs in *all0394* and *all0390* predicted the presence of an IucA/IucC family domain in both genes. In *E. coli* the *iucA* and *iucC* genes are responsible for the last two ligations in the biosynthesis of aerobactin, attaching two N^6-acetyl-N^6-hydroxy-L-lysine moieties to a citrate molecule (de Lorenzo and Neilands, 1986). In the case of schizokinen, it is proposed that two N^4-acetyl-N^4-hydroxy-1-aminopropane molecules are attached to citrate in two distinct condensation reactions, similar to those forming aerobactin. In addition, pairwise alignments of All0394 and All0390 to IucA and IucC by BLASTP showed that All0394 shares significantly higher similarity with IucA than with IucC. On the other hand, All0390 was found homologous to IucC (Table 5.3). Therefore, it was concluded that the putative All0394 may be responsible for the first ligation in the pathway forming N^2-citryl-N^4-acetyl-N^4-hydroxy-1-aminopropane, and the putative All0390 may catalyse the final synthesis reaction forming schizokinen. The entire process consumes two molecules of ATP, one molecule of NADPH and an acetyl-group (Figure 5.2).

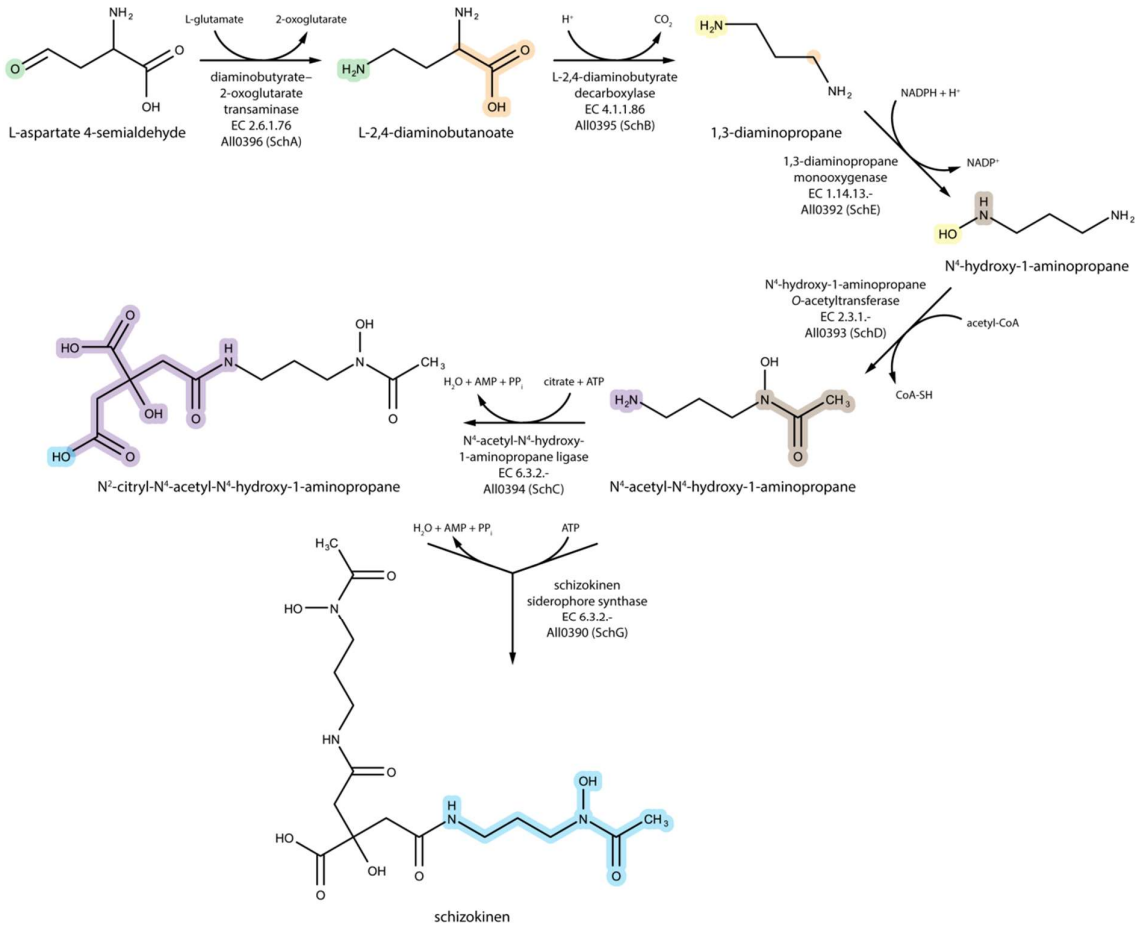


Figure 5.2. Putative pathway for the biosynthesis of schizokinen. Candidate proteins from the putative schizokinen operon were inferred based on homology to proteins in the *rhb* and *iuc* operons in *Sinorhizobium meliloti* 1021 and *E. coli*, respectively. Reacting groups and moieties are highlighted in each reaction using the colour coding of genes in Figure 5.1. Enzyme Commission (EC) numbers could not be completed for All0390, All0392, All0393 and All0394, as these reactions have not been confirmed in any organism yet.

5.2.2 Structure of the cluster

The schizokinen cluster was further studied to understand its genetic structure and also to find evidence to the existence of an operon in the region. To this end, the genomic context of the gene cluster on Figure 5.1 was analysed for transcriptional start sites (TSS), terminators and regulatory elements. First, a massive dataset originally prepared for the detection of nitrogen-stress induced TSSs was evaluated for any TSSs in the region (Mitschke et al., 2011). Interestingly, all three *fur* genes, *all1691*, *all2473* and *alr0957* (Hernandez et al., 2004) showed intensive transcriptional activity in the published dataset, significantly exceeding the non-stressed wild type level. Although the ferric uptake

regulator Fur family proteins in filamentous cyanobacteria are constitutively transcribed, their expression is only enhanced under iron limitation. Thus, it was concluded that the dataset can be used to detect TSSs that are only active under iron starvation. The experimentally determined TSSs by Mitschke et al. (2011) are shown in Figure 5.3A for the 19-kb genomic neighbourhood of the schizokinen cluster (black arrows). Two TSSs were detected on opposite strands in the intergenic region between genes *all0396* and *alr0397*. In addition, the 469-bp sequence also contains a FurA box for the binding of the iron-dependent master transcriptional regulator FurA (González et al., 2014), suggesting that the region is responsible for driving the transcription of both *all0396* and *alr0397* under iron-limited conditions. The region also harbours a putative IsaR1 binding site predicted by IntaRNA v2.0.2 (Busch et al., 2008). IsaR1 is a small regulatory RNA playing an essential role in the acclimation of cyanobacteria to low-iron conditions (Georg et al., 2017). Its presence in the corresponding genetic region provides further evidence in addition to the above, indicating that the sequence between *all0396* and *alr0397* may encompass a bidirectional promoter (P_{sch}) for the schizokinen cluster and the schizokinen outer-membrane transporter *alr0397* (*schT*).

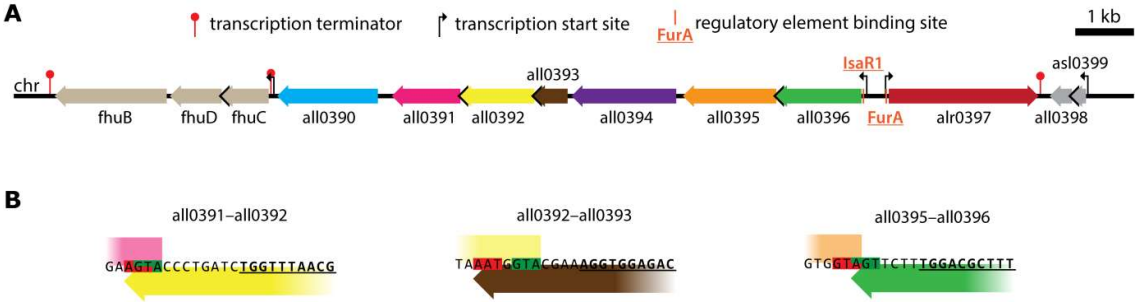


Figure 5.3. Structure of the proposed schizokinen operon in *Anabaena* sp. PCC 7120. (A) Transcriptional start sites (black arrows) and terminators (red symbols) are shown. The putative promoter region (P_{sch}) between *all0396* and *alr0397* contains binding sites for the FurA and IsaR1 regulatory elements. Features with a black cap overlap with the ORF downstream. Features are drawn to scale; a 1-kb marker is displayed in the top right corner. **(B)** Overlapping genes in panel A are detailed. Red and green boxes highlight stop and start codons of the adjacent genes, respectively. Bold underlined letters indicate sequence of RBS.

Transcriptional terminators were predicted using the WebGeSTer service (Mitra et al., 2011) and the ARNold tool (Naville et al., 2011). In total, three terminator sites were identified (Figure 5.3A, red symbols). The putative terminators together with the experimentally determined TSSs split the studied 19-kb region into four transcripts. Genes *fhuC*, *fhuD* and *fhuB* span the first transcript directly upstream of the schizokinen cluster. The *fhuCDB* cluster encodes for a ferric hydroxamate transporter

system, likely to be involved in the uptake of ferric-schizokinen, together with *alr0397* (Nicolaisen et al., 2008; Stevanovic et al., 2012). Transcript of the transporter *alr0397* shares a terminator with the hypothetical two-gene system of *all0398* and *asl0399*, both genes of unknown function. Finally, transcript of the putative schizokinen cluster stretches continuously between genes *all0390* and *all0396*. The entire region is very poor in intervening sequences between ORFs. Indeed, genes *fhuD*, *all0391*, *all0392*, *all0395* and *all0398* overlap with the gene downstream (indicated by black caps for the affected upstream genes in Figure 5.3A). The overlapping genes were further investigated in the case of the schizokinen cluster. Panel B in Figure 5.3 shows the detailed structure of the *all0391–all0392*, *all0392–all0393* and *all0395–all0396* junctions. Ribosomal binding sites (RBS) were detected in each junction region based on the consensus sequence TAGTGGAGGT in *Synechocystis* sp. PCC 6803 (Heidorn et al., 2011). In all three cases RBSs were found in the upstream gene for *all0391*, *all0392* and *all0395*, respectively (Figure 5.3B, bold underlined letters). Moreover, as a consequence of the overlap, start codons of the downstream gene in each studied junction precedes the stop codons of the corresponding upstream gene, as displayed by green and red boxes (for start and stop codons, respectively). All these details together with the identification of a single TSS, terminator and promoter for the cluster strongly suggest that the seven putative biosynthetic genes of schizokinen are transcribed into a single transcript, and thus, comprise an operon. The cluster was therefore proposed to be named as the schizokinen operon encompassing genes *schA* (*all0396*), *schB* (*all0395*), *schC* (*all0394*), *schD* (*all0393*), *schE* (*all0392*), *schF* (*all0391*) and *schG* (*all0390*). The putative metabolic function of each gene is indicated in Figure 5.2. The only exception is *all0391* that was found homologous to MFS-1 (Major Facilitator Superfamily) proteins, sharing similarity with the vibrioferrin inner membrane exporter, *pvsC* in *Vibrio parahaemolyticus* (Tanabe et al., 2006). The exact function of seven genes in the *schABCDEFG* operon was further investigated by removing the entire operon in *Anabaena* sp. PCC 7120 and complementing the knockout strain with systematic single-gene knockouts of the operon.

5.2.3 Knockout and complementation strategies

The schizokinen operon was knocked out via double recombination with a plasmid carrying an Sm/Sp resistance cassette flanked by 1-kb targeting sequences homologous to 5' (upstream) and 3' (downstream) regions of the operon (Figure 5.4A). Upstream and downstream targeting flanks were amplified from high-purity genomic DNA of *Anabaena* sp. PCC 7120 using primer sets 87+88 and 91+92, respectively (Table 5.4). It is worth noting the 5' flank was designed so that the knockout site excluded the bidirectional promoter P_{sch}. In this way P_{sch} was retained for unaltered expression of

alr0397 even in the absence of the *sch* operon. An *aadA* (Sm^R) cassette for the selection of Δ*sch* knockout mutants was amplified from plasmid pDF-lac using primers 89+90, overlapping the targeting arms. Fragments were confirmed by length and purified using a QIAquick Gel Extraction Kit (Qiagen Ltd., Crawley, UK) from 1% agarose gel (in 1 × TAE buffer consisting of 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA at pH 8.3) running at 110 V for 30–40 min. The three fragments (two targeting flanks and a selection marker) were assembled in an isothermal Gibson reaction with a pK19mob*sacB* plasmid backbone produced by PCR using primers 70+71. The comprising construct (C9 in Figure 5.4A) was maintained in *E. coli* DH5α, plasmid purified using a QIAprep Spin Miniprep Kit (Qiagen Ltd., Crawley, UK) and sequenced (Source BioScience, Nottingham, UK). Construct C9 carried the *sacB* gene (light blue arrow) to allow counter-selection of double recombinants on sucrose; it also contained an Nm^R cassette in addition to Sm^R which was not used (dark blue arrow). Oligonucleotides, DNA constructs and strains generated are collected in Table 5.4.

Table 5.4. List of oligonucleotides, DNA constructs and strains used in the schizokinen operon study. Sequences are written from 5' to 3' direction. Lower-case letters indicate overlapping sequences. Strains are based on wild type *Anabaena* sp. PCC 7120. Sequence of T_{oop} bacterial terminator is highlighted in bold.

Primers

ID	Name	Sequence
48	Psch-5'all0395-R	aatcaccatTGTGAAAATCCTCTAGGCG
49	all0395-3'Psch-F	gattttcacaATGGTGATTATTTCGCGC
50	all0396-5'all0394-R	tgtcgagcTCAAGAAACCTGCGAAAGAACCG
51	all0394-3'all0396-F	gtttcttgaGCTCGACACTCTCATTCTC
52	all0393-5'all0391-R	atgcttcatTTACCATGCTTTCCACCTC
53	all0391-3'all0393-F	gcatggtaaATGAAGCATCGGCTACCC
54	all0394-5'all0392-R	aatttaccatAAACTTCAGCCTCAAATTCCTC
55	all0392-3'all0394-F	ctgaagtttATGGTAAATTGTGTTTATGACTTGATTGG
58	3'all0391-Toop-NmR-R	aattccggttcgcttgctgt**aataaaaaacgcccggcggcaaccgagcgaatt**TTTGGCTAGGAGTTTTTGGCTAACTG
59	all0392-5'all0390-R	gaaacagagTCATGGGACTAGACCAAATTGCTG

Primers (continued)

ID	Name	Sequence
60	all0390-3'all0392-F	tcccatgaCTCTGTTTCTCTTAAATCTGATTTTTCAC
61	sch_mid-all0394-F	GGAGTTGCGAAACATCTTAGGC
62	sch_mid-all0394-R	GCCTAAGATGTTTCGCAACTCC
70	pK19mobsacB-BB-F	CACTGGCCGTCGTTTTAC
71	pK19mobsacB-BB-R	ATCATGTCATAGCTGTTTCCTG
87	5'sch-pK19mobsacB-F	ggaaacagctatgacatgatGGAATCTTGAGCTACTTCAG
88	5'sch-SmR-R	gaagccatgaTGTGAAAATCCTCTAGGC
89	SmR-5'sch-F	gattttcacaTCATGGCTTCTTGTTATGAC
90	SmR-3'sch-R	ttgagaattaATTCTCACCAATAAAAAACGC
91	3'sch-SmR-F	tggtgagaatTAATTCTCAATATTGAGGAAAACAC
92	3'sch-pK19mobsacB-R	tgtaaaacgacggccagtgcGTTAAAAAAGCGCGTCTAG
124	5'sch-cPCR-F	TATGGCTAGTGACACAATCC
125	3'sch-cPCR-R	GTCCATAACCAATCAATTCCC
129	Psch-RSF1010-F	tcctggctttgcttccagatgtatgctcttctgctcctgcCGCCTATTATTATTGACTTGCATTAG
130	all0390-Toop-NmR-R	aattccggttcgcttgctgt**aataaaaaacgcccggcggcaaccgagcgaatt**ATTATTAGCGCTCCATTGGC
131	NmR-Toop-F	gccgccgggcgtttttttattACAGCAAGCGAACCGGAA
132	NmR-RSF1010-R	gctgcgagtcttgccacgccgagcacctggtcgctttcagTTATAGTTTCTGTTGCATGGGC
133	RSF1010-bb-F	CTGAAAGCGACCAGGTGC
134	RSF1010-bb-R	GCAGGAGCAGAAGAGCATAC
149	RSF1010-Psch-seq-F	ATACCATGCTCAGAAAAGGC
161	all0394-all0394-F	ttcactaaagaatgttcagaatatttgcc
162	all0394-all0394-R	ttgtagatgaggaattactcttagggg
165	NmR-cPCR-F	GAACAAGATGGATTGCACGC
166	sacB-cPCR-R	ATCCGCATTTTTAGGATCTCCG

DNA fragments

ID	Description	Primers	Template
F13	pK19mob*sacB* backbone	70+71	pK19mob*sacB*
F19	Sm^R cassette	89+90	pDF-lac
F20	5'*sch* flank	87+88	wild type *Anabaena* sp. PCC 7120
F21	3'*sch* flank	91+92	wild type *Anabaena* sp. PCC 7120
F22	*sch* operon 1/2	129+161	wild type *Anabaena* sp. PCC 7120
F23	*sch* operon 2/2	130+162	wild type *Anabaena* sp. PCC 7120
F24	5'*schΔall0396*	129+48	wild type *Anabaena* sp. PCC 7120
F25	3'*schΔall0396* 1/2	49+62	wild type *Anabaena* sp. PCC 7120
F26	5'*schΔall0395*	129+50	wild type *Anabaena* sp. PCC 7120
F27	3'*schΔall0395* 1/2	51+62	wild type *Anabaena* sp. PCC 7120
F28	5'*schΔall0392*	129+52	wild type *Anabaena* sp. PCC 7120
F29	3'*schΔall0392*	53+130	wild type *Anabaena* sp. PCC 7120
F30	5'*schΔall0393*	129+54	wild type *Anabaena* sp. PCC 7120
F31	3'*schΔall0393*	55+130	wild type *Anabaena* sp. PCC 7120
F32	5'*schΔall0394*	129+56	wild type *Anabaena* sp. PCC 7120
F33	3'*schΔall0394*	57+130	wild type *Anabaena* sp. PCC 7120
F34	*schΔall0390* 2/2	58+162	wild type *Anabaena* sp. PCC 7120
F35	5'*schΔall0391* 2/2	59+162	wild type *Anabaena* sp. PCC 7120
F36	3'*schΔall0391*	60+130	wild type *Anabaena* sp. PCC 7120
F43	pRSF1010 backbone	133+134	pVZ322
F44	Nm^R cassette	131+132	pK19mob*sacB*

Constructs

ID	Insert	Assembled from	Genotype
C9	5'*sch*-Sm^R-3'*sch*	F19, F20, F21, F13	pK19mob*sacB*::5'*sch-aadA*-3'*sch*
C10	*sch* operon	F22, F23, F44, F43	pRSF1010-Nm^R::P_{sch}-*sch*-T_{oop}

Constructs (continued)

ID	Insert	Assembled from	Genotype
C11	*schΔall0396*	F24, F25, F23, F44, F43	pRSF1010-Nm^R::P_{sch}-*schΔall0396*-T_{oop}
C12	*schΔall0395*	F26, F27, F23, F44, F43	pRSF1010-Nm^R::P_{sch}-*schΔall0395*-T_{oop}
C13	*schΔall0394*	F32, F33, F40	pRSF1010-Nm^R::P_{sch}-*schΔall0394*-T_{oop}
C14	*schΔall0393*	F30, F31, F40	pRSF1010-Nm^R::P_{sch}-*schΔall0393*-T_{oop}
C15	*schΔall0392*	F28, F29, F40	pRSF1010-Nm^R::P_{sch}-*schΔall0392*-T_{oop}
C16	*schΔall0391*	F22, F35, F36, F40	pRSF1010-Nm^R::P_{sch}-*schΔall0391*-T_{oop}
C17	*schΔall0390*	F22, F34, F40	pRSF1010-Nm^R::P_{sch}-*schΔall0390*-T_{oop}

Strains

Name	Purpose	Genotype
9w	*sch* operon knockout (KO)	Δ*sch*, Sm^R, Sp^R
10w	*sch* operon KO full complementation	Δ*sch*::*sch*, Sm^R, Sp^R, Nm^R
11w	*sch* operon KO single-gene KO complementation	Δ*sch*::*schΔall0396*, Sm^R, Sp^R, Nm^R
12w	*sch* operon KO single-gene KO complementation	Δ*sch*::*schΔall0395*, Sm^R, Sp^R, Nm^R

Eight complementation constructs were designed comprising of a pRSF1010-based broad-host-range plasmid backbone, fragments of the schizokinen operon and a well-characterized bacterial terminator (T_{oop}, Figure 5.4B). Construct C10 contained the entire schizokinen operon, whereas C11, C12, C13, C14, C15, C16 and C17 were designed to carry single-gene knockouts for genes *all0396*, *all0395*, *all0394*, *all0393*, *all0392*, *all0391* and *all0390*, respectively. Furthermore, the expression vector backbone contained the *nptII* cassette (Nm^R, dark blue arrow) to be used in a Δ*sch* knockout strain already resistant to Sm and Sp. The other genes (*rep* and *mob*) are responsible for the mobilization and self-replication of the vector and were derived from the original pRSF1010 plasmid (purple and dark red arrows in Figure 5.4B).

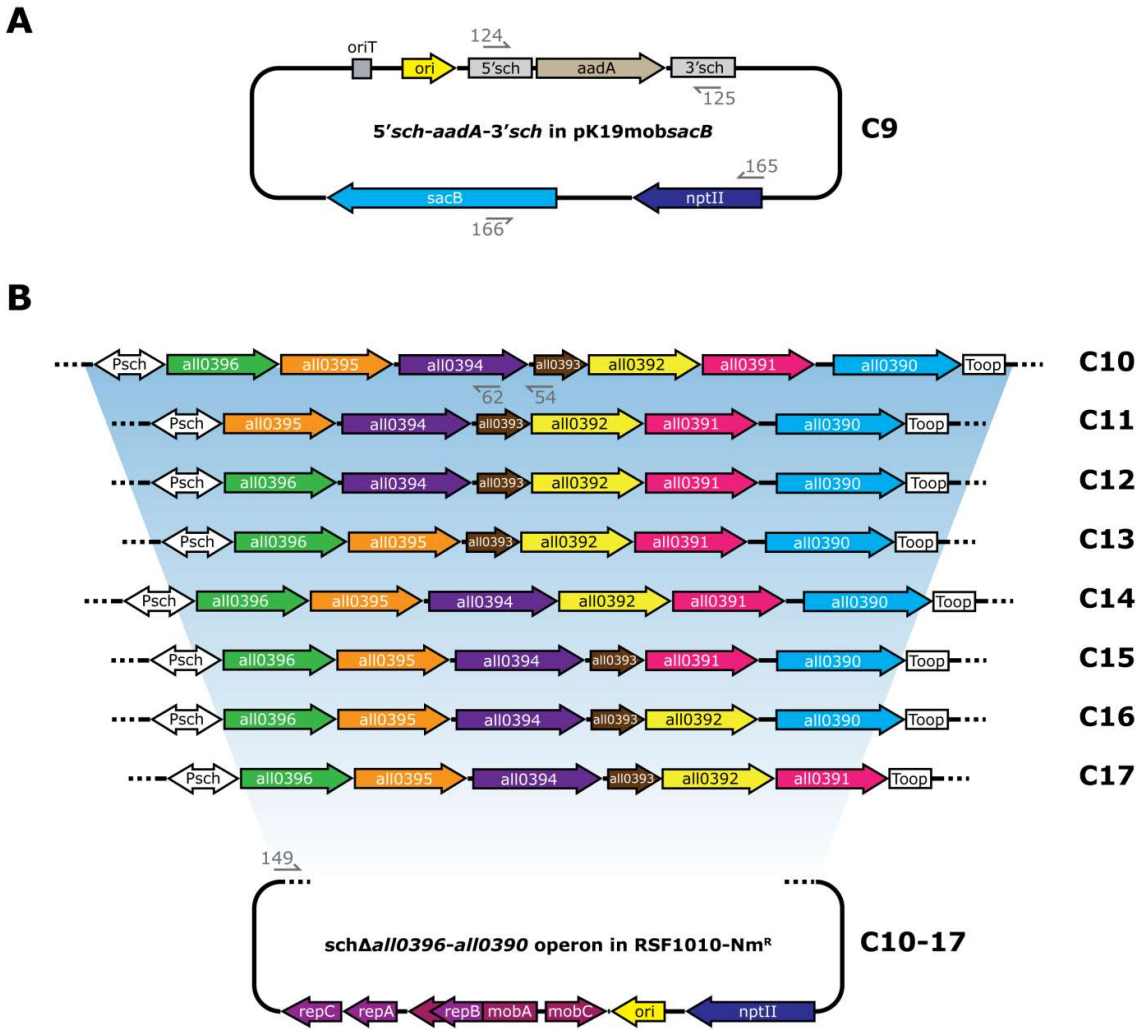


Figure 5.4. DNA constructs designed and assembled for studying the schizokinen operon. (A) Wild type *Anabaena* sp. PCC 7120 was transformed with a knockout construct carrying the *aadA* resistance cassette for Sm^R and Sp^R, flanked by 1-kb sequences (grey boxes labelled as 5'sch and 3'sch) homologous to upstream and downstream regions of the operon site. The plasmid also contained the *sacB* cassette for sucrose counter-selection. Colony PCR primer binding sites are shown as grey arrows. **(B)** Eight complementation constructs were assembled bearing the complete *sch* cluster (C10) or single-gene knockouts of it (C11–C17) lacking a different one of the seven genes in each case. All variants were under the control of their native P_{sch} promoter and a well-characterized transcriptional terminator (T_{oop}), inserted into the broad-host-range expression vector pRSF1010-Nm^R. The vector also contained the *nptII* cassette for Km^R and Nm^R. Constructs are also listed in Table 5.4.

Fragments for the complementation constructs were amplified from *Anabaena* sp. PCC 7120 genomic DNA using primers in Table 5.4. All fragments were confirmed by their length on 1% agarose gel run at 110 V for 30–40 min. Purified DNA fragments were brought to a Gibson isothermal assembly reaction and transformed to *E. coli* DH5α. However, efficiency of the assembly was remarkably lower than in earlier experiments. First, number of transformant colonies appearing on selective plates was

in the range of 0–10 in contrast to 50–100 colonies of other assemblies under the same transformation conditions. In addition, the fewer colonies returned false positives in restriction enzyme analysis and sequencing. To overcome this issue the assembly protocol was optimized to the amount and ratio of DNA fragments as described previously (Gibson et al., 2009; Gibson, 2011) with little or no positive effect on transformant numbers. In addition, size of overlaps between fragments was increased from the original 20 nucleotides to 40 nucleotides as recommended for larger (> 10 kb) assemblies (Gibson, 2011). It was also hypothesized that the strong DNA secondary structure of the T_{oop} terminator at the junction of two adjacent fragments may hinder the assembly of parts into one circular plasmid. Therefore a long primer enclosing T_{oop} and also containing a 20-nt overlap undisturbed by the terminator was also designed. Nevertheless, no constructs stitching together all fragments could be assembled.

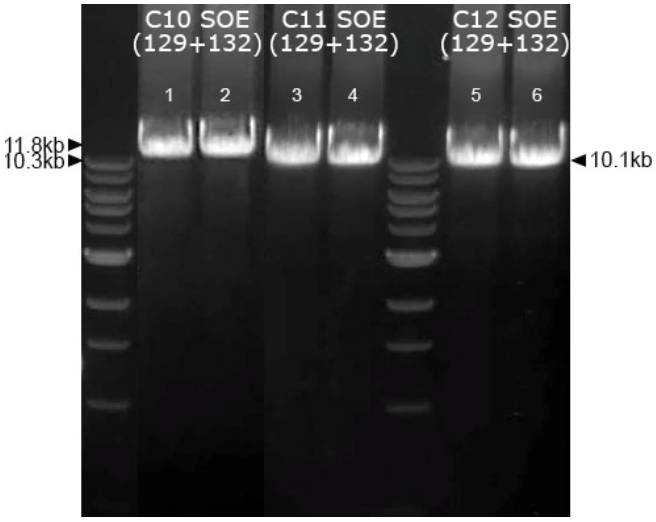


Figure 5.5. Results of overlap extension PCR (SOE) for the assembly of C10, C11 and C12 constructs. The first step of SOE was performed in the absence of oligonucleotide primers. In the second step terminal primers 129+132 were used to amplify the full insert from the background. Reactions were run in duplicates and pooled prior to purification to increase the concentration of the large product. Fragment sizes are indicated by black triangles next to the image.

Finally, redesigned primers with fourty-nucleotide overlaps were brought into an overlap extension PCR (Bryksin and Matsumura, 2010) using adjacent fragments as megaprimers and joined together using the outermost PCR primers as described in section 2.5.7. Individual fragments of the insert were amplified using the appropriate primers in Table 5.4. In the final step (purification PCR) the outermost oligonucleotide primers of two terminal fragments were also used to specifically amplify the fully

assembled sequence (splicing by overlap extension, SOE). Assemblies of the correct size were confirmed on agarose gel and purified. Results of overlap extension PCR for constructs C10, C11 and C12 are shown in Figure 5.5 in duplicates. Insert for C10 was assembled from three fragments spanning over 11.8 kb (lanes 1 and 2). The C11 SOE insert contained four fragments with 10.3 kb in size (lanes 3 and 4). Finally, C12 SOE was assembled from four fragments with a total length of 10.1 kb (lanes 5 and 6). The inserts were introduced to the plasmid backbone (linearized by standard PCR using primers 133+134) by standard Gibson isothermal reaction. The rest of the eight constructs on Figure 5.4B were prepared similarly.

All constructs including C9 (Figure 5.4A) were maintained in *E. coli* DH5α and confirmed by sequencing over each fragment junction in the corresponding construct.

5.2.4 Isolation of schizokinen mutants

Construct C9 was transformed to *Anabaena* sp. PCC 7120 wild type resulting in a strain denoted as Δ*sch* (9w in Table 5.4). Complementation constructs on a self-replicative broad-host-range vector were transformed into the Δ*sch* knockout mutant. Double recombinants with C9 were segregated until homogeneity of the desired genotype over the course of several weeks. Complementation strains, on the other hand, only required confirmation that the corresponding expression construct is stably present. In case of both methods candidate colonies were screened by colony PCR as described in the sections below.

5.2.4.1 Schizokinen knockout

Wild type *Anabaena* sp. PCC 7120 was transformed with construct C9 to knockout the entire *sch* operon via double recombination. Triparental conjugation to introduce the DNA construct to the cyanobacterium was performed on an exponentially growing culture as detailed in section 2.6.3. Exconjugants were selected on BG-11 agar plates supplemented with 2.5 µg/ml Sm and Sp antibiotics. Single recombinants were counter-selected against by the addition of 5% sucrose to the culture medium. Steps for the isolation of double recombinants were essentially the same as described in section 4.2.4.2. In addition to that, a primer pair amplifying from outside the affected genetic site was used to distinguish all possible scenarios (wild type, single recombinant and double recombinant) by colony PCR. Thus, segregating colonies were regularly tested using primers 169+170 amplifying the

entire integration site, producing a fragment of 12.2 kb, 21 kb or 3.2 kb for wild type, single recombinant or double recombinant, respectively. Moreover, primer sets 124+125 and 124+126 were also used to allow for specific detection of a knockout and the wild type, respectively. Primer 124 and 125 annealed to the middle of the 5' and 3' targeting flanks, respectively, amplifying an 11.1-kb product from wild type and a 2-kb product from a Δ*sch* mutant. In contrast, primer 126 possessed a binding site only in the wild type (inside *all0396*), and not in a Δ*sch* mutant. It is worth noting, however, that these latter two sets were not selective for single recombinants. Therefore, primers 165+166 were used to specifically screen for single recombinants. Figure 5.4A maps the colony PCR primers used with construct C9 (grey arrows). Examples of consecutive colony PCR rounds are collected in Figure 5.6.

Panel A and B display results for six isolates following filament fragmentation by sonication. All four strains in panel A (lanes 5–8) and both in panel B (lanes 5 and 6) produced the 2-kb band indicative of the double recombinant (DR) genotype. However, isolates 9w2 and 9w4 (on lanes 4 and 3, respectively) showed only the wild type (WT) band. In the same segregation round, 9w1 (lane 1 in panel A) did not give a band with primers 169+170 (amplifying the entire insertion site), which may suggest a single recombinant (SR) genotype. The reason is colony PCR was performed with an extension time optimal for the 12.2-kb wild type fragment that is insufficient for the amplification of the 21.1-kb SR product. Similarly, candidates 9w3 and 9w5 (panel B) contained the DR-indicative 2-kb band with primers 124+125, but no product was formed with primers 169+170 (lanes 1 and 2). Thus, these two were also assayed using primers 165+166. The resulting 1.6-kb band is an amplification of *nptII* and *sacB* cassettes from the plasmid (Figure 5.4), a typical product in case of a SR (lanes 3 and 4). These results suggested that 9w1, 9w3 and 9w5 contained the undesired SR genotype, and therefore, these isolates were excluded from further rounds of segregation.

Candidates 9w6 and 9w2, on the other hand, were selected for subsequent rounds of segregation including sonication in liquid and re-isolation in line cultures on selective BG-11 agar plates. Later, although still exhibiting the DR band with primers 124+125 (panel D, lanes 5 and 6), clones of 9w2 developed the SR-specific band with primers 165+166 in panel C (lanes 12 and 13). This is not surprising as only a single DR is expected for approximately every one hundred SR events (see section 4.2.2.1 for further explanation). Thus, it is likely 9w2 contained the SR genotype already in the previous round, although the wild type version was dominant. Nonetheless, three clones of 9w6 out of four produced the expected 3.2-kb band for a DR (lanes 1–3 in panel C) with no signs of the SR-specific 1.6-kb band with primers 165+166 (lanes 8–11). In the same test wild type *Anabaena* sp. PCC 7120 resulted in a 12.2-kb amplicon with primers 169+170, as expected for a WT.

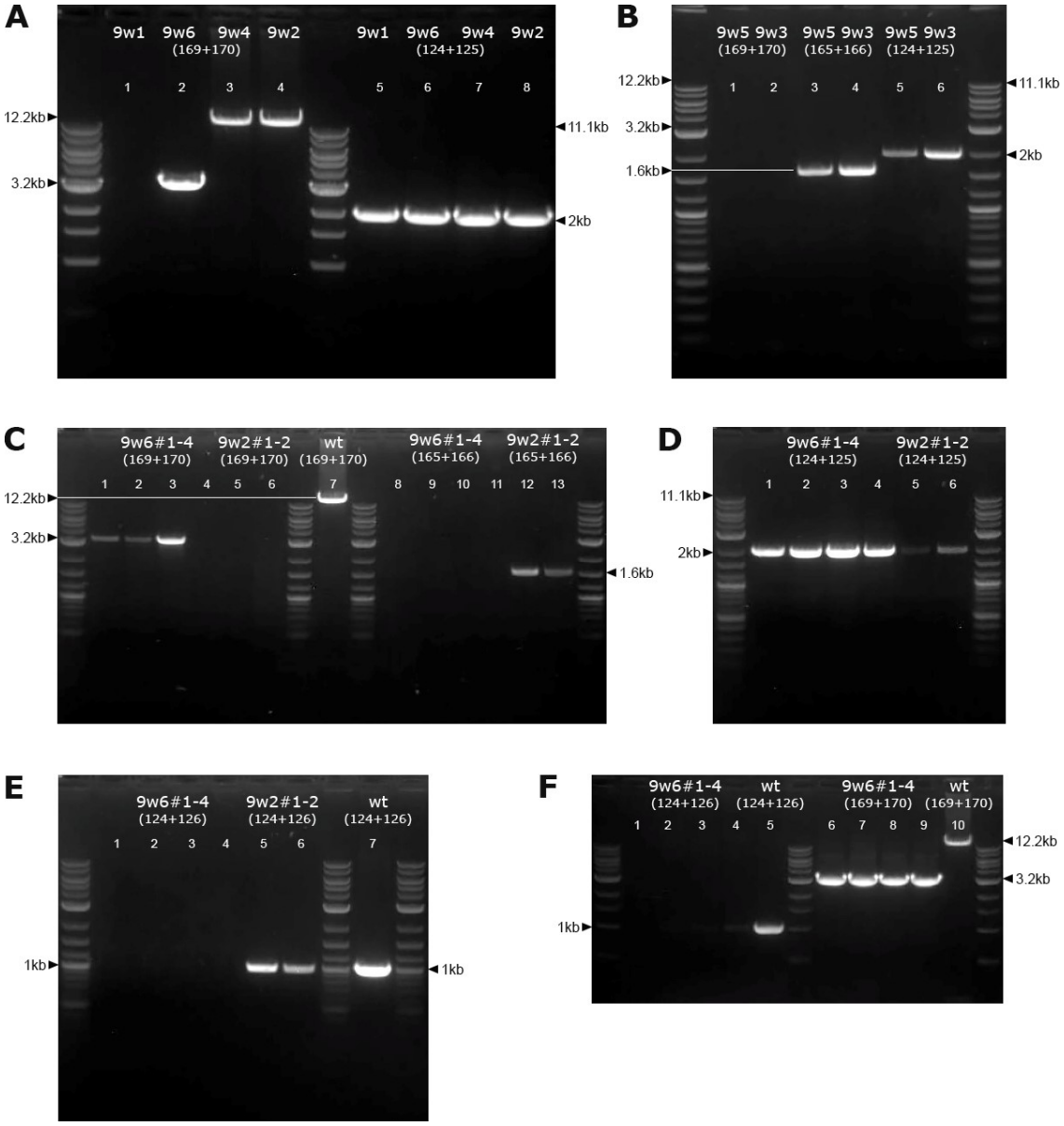


Figure 5.6. Colony PCR results for consecutive segregation rounds of the Δ*sch* mutant. Candidate colonies were tested with primers 169+170 for a 12.2-kb band or a 3.2-kb band in case of wild type (WT) or a double recombinant (DR), respectively. A single recombinant (SR) would theoretically result in a 21.1-kb band, but extension time was optimal for the 12.2-kb WT product. Amplification using primers 124+125 may result in an 11.1-kb band for a WT or a 2-kb band for a DR. A SR is expected to produce both bands. Primers 165+166 amplify a 1.6-kb band from SR only. Primer pair 124+126 is specific to WT producing a 1-kb band. Six isolates were tested in the first round in **(A)** and **(B)**. Two promising candidates (9w6 and 9w2) were further segregated in **(C)**, **(D)** and **(E)**. Isolate 9w6 was also confirmed from high-purity genomic DNA by PCR **(F)**. Further details are given in the main text. Band sizes were predicted by comparing to the 1-kb and 2-log DNA ladders from NEB.

The promising strain 9w6 was also analysed using primers 124+126 for the amplification of a WT band, together with 9w2 and a wild type control (panel E). As expected, wild type *Anabaena* sp. PCC 7120 as well as 9w2 produced the 1-kb WT band (lanes 5, 6 and 7 in panel E). However, no band was visible for any clone of 9w6 indicating that segregation of the DR genotype may be complete in these samples. In order to acquire further proof, genomic DNA was purified from all four clones of 9w6, and colony PCR was repeated on high-purity genomic DNA. In contrast to the strong 1-kb band in case of the wild type control in panel F (lane 5), signs of a WT genotype in the 9w6 clones remained undetectable (lanes 1–4). At the same time, the DR-indicative 3.2-kb band with primers 169+170 was successfully produced (lanes 6–9). Therefore, it was concluded that 9w6 is a double recombinant of construct C9 and genetic segregation is complete. Furthermore, since the *sch* operon was undetectable in 9w6, the strain was considered as a Δ*sch* mutant. The mutant was cryopreserved in 18% glycerol at -80 °C and used as a base strain in a subsequent complementation study.

5.2.4.2 Complementation strains

Complementation constructs for partial (single knockout) or complete restoration of the schizokinen operon were introduced to the full knockout Δ*sch* strain of *Anabaena* sp. PCC 7120. Originally, a total of eight strains complementing the schizokinen knockout were planned and designed (Figure 5.4B). However, only three strains were actually prepared as proof of concept for the complete method. More precisely, constructs C10, C11 and C12 were transformed into the Δ*sch* strain. Construct C10 intended full restoration of the schizokinen pathway, whereas C11 and C12 complemented Δ*sch* to a single-gene knockout of the first two genes in the pathway (*all0396* and *all0395*, respectively). Triparental conjugation was performed similarly to that generating the Δ*sch* strain, except for selection conditions. Exconjugants of complementation strains were screened on 100 μg/ml neomycin (Nm) plates also supplemented with 2.5 μg/ml Sm and Sp antibiotics, with no sucrose added. Green transformant colonies appeared on the filter after about 3 weeks of conjugation. Eight isolates were picked from each conjugation filter and streaked onto fresh BG-11 plates supplemented with the above antibiotics. Isolates were restreaked onto fresh plates once every 2 weeks for about 2 months to stabilize expression of the complementation construct. Candidates were confirmed by colony PCR as shown on Figure 5.7.

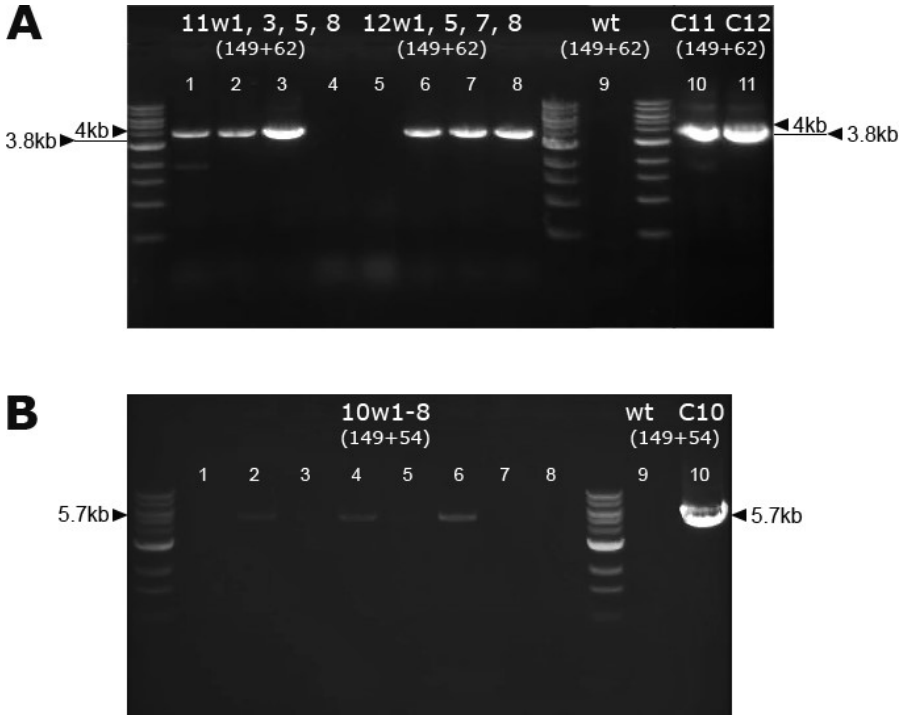


Figure 5.7. Colony PCR results of Δ*sch* complementation strains bearing C10, C11 and C12. Candidates for single-knockouts **(A)** and full complementation **(B)** were screened with primer sets 149+62 and 149+54, respectively. Oligonucleotide primer 149 was designed to anneal to the plasmid backbone upstream of the insert and amplify into the complementation cassette. The other primer in both pairs annealed to the insert. Approximate binding sites are indicated on Figure 5.4B (grey arrows). Band sizes were predicted by comparing to the 1-kb DNA ladders from NEB.

Four candidates of the Δ*all0396* and Δ*all0395* single-knockout complementation strains were tested with primers 149+62 in Figure 5.7A. Isolate 11w1, 11w3 and 11w5 on lanes 1–3 showed the desired band at 4 kb, similar to the positive control (C11) on lane 10. In addition, a 3.8-kb product was detected in samples 12w5, 12w7, 12w8 (lanes 6–8) and the plasmid control (C12 on lane 11), indicating the presence of the complementation construct. As expected, no amplicon was produced from wild type genomic DNA (lane 9). Band intensity was strongest for 11w5, although the difference between the bands on lanes 1–3 may only indicate variations in initial colony PCR conditions. Moreover, samples 11w8 and 12w1 (lanes 4 and 5) produced no band, indicating either the absence of the construct, or more likely, a failure to extract enough DNA or remove PCR inhibitors for colony PCR to work properly. Candidates for the full complementation of the *sch* operon were analysed using a different primer set, 149+54 (Figure 5.7B). This primer pair amplifies a 5.7-kb fragment from a mutant bearing the desired expression system, but no band from a wild type. Although bands were faint in general, presence of the complementation construct C10 could be clearly confirmed in candidates 10w2, 10w4 and 10w6 (lanes 2, 4 and 6, respectively).

Confirmed isolates of C10, C11 and C12 were inoculated into 1 ml BG-11 supplemented with 1.5 µg/ml Sp/Sm and 75 µg/ml Nm and grown for up to 8 days under standard growth conditions. Resulting liquid cultures were subcultured in 20 ml fresh BG-11 containing the appropriate antibiotics for an additional 8 days, and cryopreserved. Developed strains are listed in Table 5.4. Following revival from cryostocks complementation strains were studied under normal conditions and iron deprivation for the production of siderophores, in particular schizokinen. Single-knockout complementation strains (carrying C11 and C12) were designed with the intention to be compared to the wild type and assess the accumulation of dead-end metabolites related to schizokinen biosynthesis, and the lack of schizokinen in the supernatant. At the end, however, only growth characteristics in response to changes in iron availability could be compared.

5.2.5 Characterization of the *sch* operon mutants

Characterization of all strains related to the investigation of the schizokinen operon was performed by Mr Peter Wellham (MRes student, Microbial Metabolic Engineering Group, Imperial College London, UK), supervised by Ms Marine Valton (PhD student, Microbial Metabolic Engineering Group, Imperial College London, UK) and to a lesser extent, by the author of this thesis. In the following sections (5.2.5.1 and 5.2.5.2), all laboratory work was performed by Mr Peter Wellham including the generation of data plots (Figure 5.8, Figure 5.9 and Figure 5.10), and the information contained therein was interpreted and evaluated by the author of this thesis.

5.2.5.1 Chrome azurol S assay for siderophore detection

In order to compare siderophore production capability of the different schizokinen mutants, wild type (WT), Δ*sch* full knockout (KO), Δ*sch*::*sch* full complementation (CS) and the single-knockout complementation strains (Δ*sch*::*sch*Δ*all0395* and Δ*sch*::*sch*Δ*all0396*, see Table 5.4) were grown on BG-11 plates supplemented with the appropriate antibiotics (or without any antibiotics in case of WT) under standard growth conditions. Colonies were grown for about two weeks and streaked onto a BG-11 agar plate containing the blue dye chrome azurol S (CAS), forming little round patches of cyanobacterial biomass. The patches were cultivated for additional two weeks and the plate was photographed (Figure 5.8). The original blue colour of the CAS-iron complex changes to the native yellow-orange colour of CAS upon removal of the iron (III) from the complex. The CAS dye itself is a

mild iron chelator, and only stronger chelators (e.g. siderophores) can release the iron (III) ion from the CAS-iron complex.

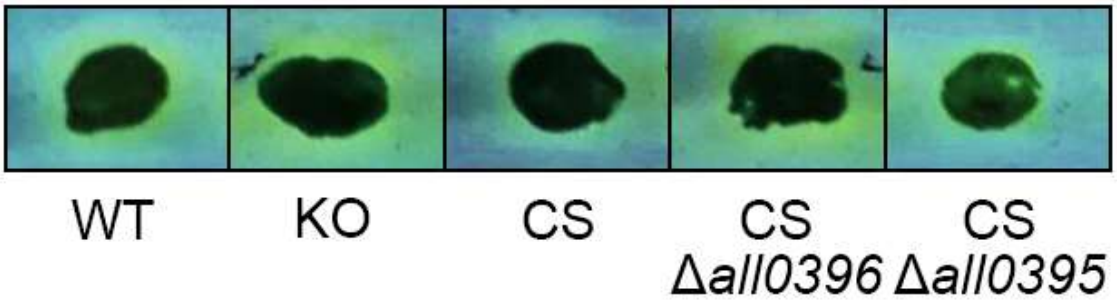


Figure 5.8. Results of a typical CAS assay on the schizokinen mutant strains. Mutants of the schizokinen operon together with wild type *Anabaena* sp. PCC 7120 were grown on CAS-containing plates for two weeks in round patches. Any halo formation around the patch cultures was evaluated. From left to right, pictures show halo formation by the wild type (WT), Δ*sch* operon KO (KO), the full complementation (CS) and the single-gene complementation strains (CS Δ*all0396* and CS Δ*all0395*).

As shown in Figure 5.8, there was a halo observed around all strain patches. This halo had a clear, distinctive yellow colour from the blue background (close to the edges of each small image). The growth and density of the patches were very similar in all cases, as well as the radii of the halos. Quite surprisingly, even wild type and Δ*sch* strains were undistinguishable from each other by this assay. A likely explanation is the excretion of iron chelators other than schizokinen, also by the knockout and complementation strains. The strains were further tested in growth assays to evaluate any influence on growth characteristics by the modifications related to schizokinen.

5.2.5.2 Growth characterization of the schizokinen mutant strains

The schizokinen knockout mutant was grown in parallel with the wild type strain in a multicultivator (MC1000-OD, Photon Systems Instruments, spol. s r.o., Drasov, Czecz Republic) to compare growth characteristics of the two strains at high resolution. The multicultivator was assembled following the manufacturers protocol (Photon-Systems-Instruments, 2017). Inoculum cultures were grown under standard growth conditions in 25 ml BG-11, supplemented with the appropriate antibiotics. After two weeks of growth, inoculum cultures were harvested by centrifugation, washed in fresh BG-11 three times and resuspended in one of the following growth media: standard BG-11, BG-11 supplemented

with 50 µM 2,2'-bipyridyl (D216305, Sigma-Aldrich Company Ltd., Dorset, UK) – a mild iron chelator, or BG-11 without added iron (the ammonium ferric citrate component was left out). Cultures in the multicultivator were diluted to a starting OD_{720} of 0.02 in 50 ml of the appropriate culture medium and incubated at 30 °C and 60 µE m^{-2} s^{-1} for two weeks, with automatic OD measurements at 720 nm every 30 minutes. The resulting growth curves are collected in Figure 5.9.

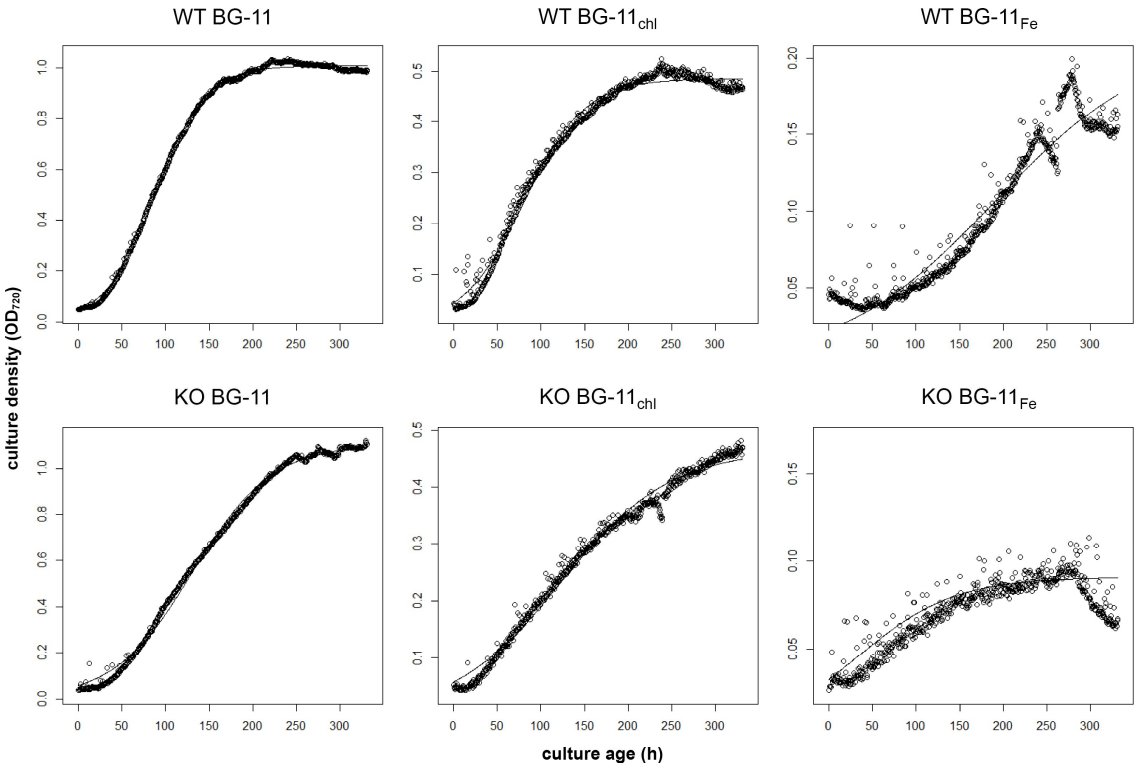


Figure 5.9. Comparison of wild type and Δ*sch* knockout strains' growth under iron limiting conditions. Strains were grown for two weeks in an MC1000-OD multicultivator at 30 °C, atmospheric CO_2 and 60 µE m^{-2} s^{-1} under continuous illumination. Three different liquid media were used: standard BG-11, BG-11 supplemented with a mild iron chelator ($BG\text{-}11_{chl}$) and BG-11 without added iron ($BG\text{-}11_{Fe}$). Culture density was recorded every 30 minutes at 720 nm. WT: wild type *Anabaena* sp. PCC 7120; KO: Δ*sch* strain. Data points are averages of four biological replicates with a relative error ≤ 11% for all points. Error bars are not shown to improve readability.

Significant level of growth could be observed in all six multicultivator tubes, although rate of growth and final density of the cultures were significantly different. In standard BG-11 the two strains showed similar growth characteristics, both reaching stationary phase at the OD_{720} value of 1 (WT BG-11 and KO BG-11 in Figure 5.9). The Δ*sch* strain (KO BG-11) exhibited slower growth rate than the wild type (WT BG-11), and yet reached the same final OD_{720} at the 10^{th} day of incubation. In the case when free

iron in the medium was chelated by 2,2'-bipyridyl (BG-11_{chl} middle column in Figure 5.9), both strains displayed similar growth, although the knockout strain was somewhat slower than the wild type. The final density of these cultures was about half of that of the "unconstrained" cultures in BG-11, suggesting that cells required extra energy and effort to capture iron from this environment. Such efforts may have included the excretion of siderophores. Surprisingly, however, there was no remarkable difference in terms of growth between the two strains (WT BG-11_{chl} and KO BG-11_{chl}). This result was similar to what the CAS plate assay showed in Figure 5.8, indicating that the schizokinen knockout strain was able to free up the essential iron from the chelator complex. This observation suggests the involvement of siderophores even in case of the KO strain, although, for that strain in particular, this siderophore could not be schizokinen.

The third set of growth curves showed signs of severe iron limitation in the iron-poor medium (BG-11_{Fe}, third column in Figure 5.9). The final culture density at the end of the two-week assay was about one-fifth of that of "unconstrained" growth in BG-11. In addition, growth was gradually slowing down in case of the Δ*sch* mutant compared to the wild type, which was growing at a fairly constant rate (WT BG-11_{Fe} and KO BG-11_{Fe}). In fact, it is quite surprising that growth was possible at all in a medium lacking ferric ammonium citrate (the iron-containing component of BG-11).

Overall, the growth rate of the knockout strain was about 30–40% slower than that of the wild type in all media types including standard BG-11, indicating that removal of the schizokinen cluster may have negatively affected this strain, beyond the expected impact under iron limitation. This possible impact was further investigated by the juxtaposition of all four schizokinen mutant strains (Δ*sch* full KO, Δ*sch* complementation and two of the Δ*sch* single-gene knockout complementation strains) and *Anabaena* sp. PCC 7120 wild type, in a two-week growth assay in 6-well microtitre plates. Inoculum cultures were grown in 25 ml BG-11 under standard incubation conditions in continuous light, harvested by centrifugation after two weeks, washed in fresh BG-11 three times and resuspended in 3 ml of one of the following media: standard BG-11, BG-11 supplemented with 50 µM 2,2'-bipyridyl, or BG-11 without added iron. Starting density of the cultures was set to 0.02 at 720 nm. The 6-well plates were shaken at 30 °C, and 60 µE m^{-2} s^{-1} continuous illumination, in 1% enriched CO_2. The optical density of the cultures was measured using a Tecan M200 Pro multimode plate reader at 720 nm once a day, by drawing 20 µl homogenous culture from each well. Results of the growth assay are collected in Figure 5.10. Measurement values of three biological replicates were aggregated by averaging, and a logistic growth curve was fitted on the data points using the Grofit R package (Kahm et al., 2010).

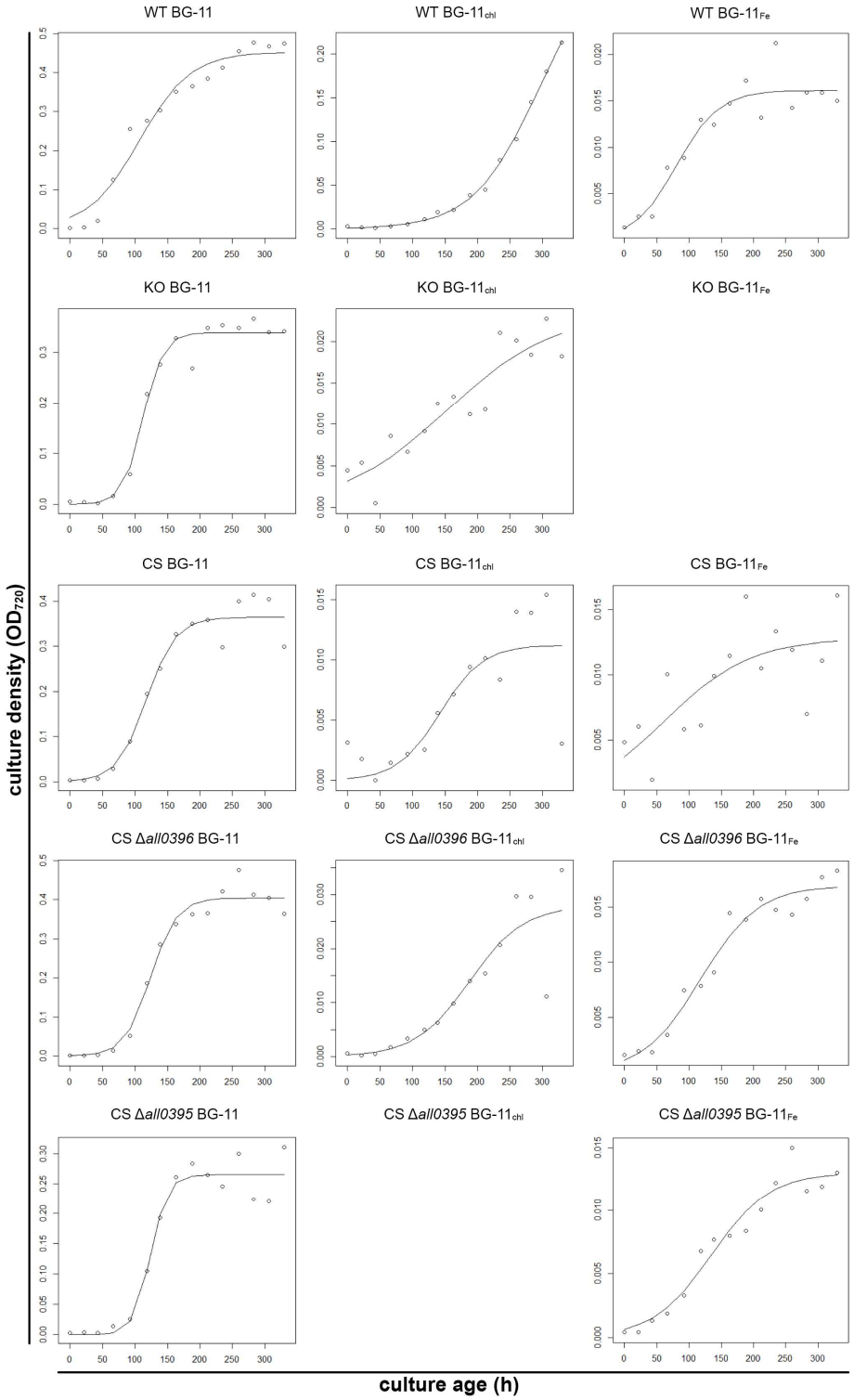


Figure 5.10. Growth of schizokinen mutant strains under iron limitation. Strains were grown in 3 ml BG-11, BG-11 supplemented with an iron chelator (BG-11_{chl}) and BG-11 without added iron (BG-11_{Fe}) for two weeks. Plates were incubated at 30 °C, 1% CO_2 and 60 µE m^{-2} s^{-1} continuous illumination. Culture density was measured daily at 720 nm. WT: wild type *Anabaena* sp. PCC 7120; KO: Δ*sch* strain; CS: complementation strain (Δ*sch*::*sch*); CS Δ*all0395* and CS Δ*all0396*: single-knockout complementation strains. Data points are averages of three biological replicates with a relative error ≤ 18% for all points. Error bars are not shown to improve readability.

In standard growth medium all strains showed balanced growth, although growth rates were slightly different for each strain (first column of charts in Figure 5.10). Interestingly, the growth rate difference experienced between WT and KO strains in the multicultivator was much less pronounced in 6-well plates (WT BG-11 and KO BG-11 in Figure 5.10). At the same time, the complementation strains (CS BG-11, CS Δ*all0395* BG-11 and CS Δ*all0396* BG-11) displayed very similar growth characteristics to the WT and the KO.

The second dataset (middle column in Figure 5.10) compared the different strains in BG-11 supplemented with a mild iron chelator (bipyridyl). As expected, growth rate and final OD_{720} at the end of the two weeks were both significantly lower for the WT (WT BG-11_{chl}) than the same parameters in standard BG-11 (without the chelator, WT BG-11). Indeed, these parameters indicated remarkably decreased growth in case of all strains, although the effect was more severe for the schizokinen mutants (KO, CS, CS Δ*all0395* and CS Δ*all0396*). For the CS Δ*all0395* single-knockout complementation mutant, growth was not visible in BG-11_{chl} at all, and a growth curve could not be fitted. The rest of the mutant strains were only barely growing, and yet a clear trend of culture growth was visible. However, due to the low rate and high noise in the optical density measurements these data curves could not be used to reliably evaluate the strains. Similarly, the iron-limited BG-11_{Fe} medium severely repressed growth for all strains, including the wild type. This effect was, however, expected, due to the lack of essential iron in the culture medium.

5.3 Discussion

Previous studies suggested the involvement of seven genes upstream of *alr0397* (*schT*) in the biosynthesis of schizokinen, based on sequence similarities to the genes involved in the production of rhizobactin 1021 siderophore in a rhizobium species (Nicolaisen et al., 2008). In addition, the region has been shown to be expressed only under iron limiting conditions (Jeanjean et al., 2008; Stevanovic et al., 2012), and the *all4025* (*schE*) gene has been identified as an exporter for this siderophore (Nicolaisen and Schleiff, 2010). Although the ability of *Anabaena* sp. PCC 7120 to excrete schizokinen upon iron starvation has long been recognized (Goldman et al., 1983), there is still very little known about the biosynthetic pathway of this siderophore. Lynch et al. (2001) has suggested a route for the synthesis of rhizobactin 1021 that involves schizokinen as an intermediate metabolite. The genes that may be associated with this pathway in *Anabaena* sp. PCC 7120 were identified and the structure of the corresponding gene cluster was analysed using bioinformatics approaches.

Based on sequence similarity to the *rhbA–rhbF* genes involved in the biosynthesis of rhizobactin 1021 in *Sinorhizobium meliloti* 1021, a cluster of seven genes was identified in several cyanobacterial genomes (Figure 5.1). The hypothetical genes were found highly conserved among the six cyanobacterial strains tested, with *Anabaena variabilis* showing the highest level of homogeneity to the closely related *Anabaena* sp. PCC 7120 (Table 5.1). The schizokinen outer membrane transporter (*alr0397* or *schT*) was also found in all the cyanobacterial species except one, *Cyanobacterium* sp. PCC 10605. This was somewhat surprising, because homologues of the rest of the genes were successfully identified, also in this organism. A likely explanation is the lack of a "dedicated" schizokinen transporter in this cyanobacterium. Stevanovic et al. (2012) has characterized a few siderophore-mediated iron uptake systems in *Anabaena* sp. PCC 7120, of which more than one has been shown to actively transport iron-schizokinen. If not *schT*, then one of these uptake systems may be operative in *Cyanobacterium* sp. PCC 10605, explaining the missing outer membrane transporter in the close genomic neighbourhood of the schizokinen cluster.

Interestingly, homologues of the *all0391* gene were found in all tested cyanobacterial genomes, although no biochemical function could be associated with this gene in relation to the biosynthesis of schizokinen. Nonetheless, *all0391* showed similarity to a siderophore inner membrane exporter (*pvsC* in *Vibrio parahaemolyticus*), suggesting its involvement in the complex membrane transport system of schizokinen.

One of the seven genes in the studied cluster, *all0390* in *Anabaena* sp. PCC 7120 and its homologues in the other cyanobacteria, showed remarkable size variation. Based on sequence length, two versions of this gene product could be distinguished: a shorter one consisting of two Pfam domains and a longer one containing an additional acetyltransferase domain (Table 5.2). The encoded protein is proposed to be responsible for one of the last ligation steps in the schizokinen pathway (Figure 5.2); it remained an open question, however, how the presence or absence of the additional acetyltransferase domain alters the functionality of the encoded enzyme.

The *all0390–all0396* genomic region and its 19-kb context was analysed for TSSs and transcription terminators. In total, four TSSs and three terminators were identified in the locus in question, splitting the site into four transcribed regions (Figure 5.3). Gene *alr0397* was found to be independently transcribed from the *all0390–all0396* region. This region (the schizokinen cluster) on its own comprises a single transcript, providing strong evidence for an operon (thereby called: *sch* operon) at this location. The existence of such a continuous transcript can be inferred from the TSS and the transcriptional terminator surrounding the region. In addition, the start codon of three genes (*all0391*, *all0392* and *all0395*) in this region can be found in the CDS of the preceding gene, of which two start

codons are actually overlapping with the stop codon of the upstream CDS (Figure 5.3). A putative bidirectional promoter (P_{sch}) was also identified for the *sch* operon, driving the transcription of *alr0397* (*schT*) as well. The promoter harbours a binding box for the iron uptake regulator FurA (González et al., 2014) and a predicted recognition site for the iron-stress induced sRNA, IsaR1 – providing further evidence for a promoter activated under iron limitation.

The proposed biochemical function of the six catalytic genes in the *sch* operon was predicted by the detection of protein domains and associated enzymatic activity by bioinformatics analysis (Figure 5.2). For further assessment of their role in schizokinen biosynthesis, all genes in the *sch* operon (including the non-catalytic *all0391*) have been knocked out in wild type *Anabaena* sp. PCC 7120. A Δ*sch* schizokinen KO strain was isolated first, and single-gene knockout mutants were generated by systematic complementation of this KO strain (Figure 5.4). The full complementation strain (called CS or Δ*sch*::*sch*) was derived similarly from the Δ*sch* strain, by transformation with a shuttle vector carrying the entire *sch* operon (Figure 5.4A). However, molecular assembly of the 18-kb complementation construct was challenging using Gibson assembly. It was either the sheer size of the construct or the strong secondary structure of one of the fragments' overlapping regions that interfered with the success of the assembly reaction; nevertheless, no optimization of the reaction conditions could overcome these difficulties. Therefore, construct assemblies involving the fusion of large and complex DNA sequences were carried out via overlap extension PCR (Figure 5.5). This approach was found particularly efficient in making large constructs, without increasing the ratio of false positives.

Although the isolation of fully segregated Δ*sch* mutants of *Anabaena* sp. PCC 7120 wild type was very slow due to the low efficiency of *sacB*-mediated counter-selection, it was possible to confirm a schizokinen operon KO strain (KO or Δ*sch*), and subsequently, three compensatory strains (CS or Δ*sch*::*sch*, CS Δ*all0395* and CS Δ*all0396*) derived from it. The strains were characterized and compared in growth assays under standard conditions and iron starvation. In a CAS assay, strains were compared based on their ability to free up iron from the CAS-iron complex. The CAS-iron complex had a distinctive blue colour that changed upon removal of iron (III) ions from the complex to yellow or orange. The colour change in this assay is an indication of siderophore production, as siderophores are stronger chelators of iron than CAS by several orders of magnitude, and are thus able to release iron from the CAS complex (Alexander and Zuberer, 1991; Louden et al., 2011). The yellow halo forming around wild type and full complementation strains was according to expectations. There was no observable difference in growth between these two, suggesting that none of them was limited by the availability of iron or in any other way (Figure 5.8). Quite surprisingly, the Δ*sch* full knockout (KO)

and the single-knockout complementation strains (CS Δ*all0395* and CS Δ*all0396*) showed very similar halo formation around their colonies. It is, however, not evident, whether this halo around the full KO and single-KO strains was the indication of schizokinen excretion. In fact, there is at least one more, otherwise uncharacterized siderophore *Anabaena* sp. PCC 7120 is known to be producing under iron limitation (Jeanjean et al., 2008; Nicolaisen and Schleiff, 2010). This other siderophore has been shown to be expressed and transported independent of the schizokinen release and uptake systems (Nicolaisen and Schleiff, 2010). This is not the case for other siderophores *Anabaena* sp. PCC 7120 is capable of utilizing, such as aerobactin and ferrioxamine B (Goldman et al., 1983; Rudolf et al., 2016). It is unknown, whether *Anabaena* sp. PCC 7120 can synthetize these siderophores on its own. Nevertheless, due to the existence of other siderophores in this cyanobacterium the CAS assay could not be used to specifically, and not even to quantitatively detect schizokinen in the tested strains.

It may be possible to develop a semi-quantitative CAS assay for the detection of schizokinen in the future, by spotting the structures at equal starting OD or CFU number in a dilution series. Also, application of an optimized version of the bioassay suggested by others (Lynch et al., 2001; Jeanjean et al., 2008; Amin, 2010) might also enable the detection of such subtle differences in overall siderophore excretion due to the loss of the putative schizokinen biosynthesis genes.

The five strains were further evaluated in growth assays under standard and iron-limiting conditions. In standard BG-11 medium the KO strain showed slightly slower growth compared to the wild type in both a multicultivator and a 6-well microplate (Figure 5.9 and Figure 5.10, respectively). In the microplate assay, two of the complementation strains, CS and CS Δ*all0396* displayed subtly higher culture density than the KO and the other single-gene knockout complementation strain, CS Δ*all0395* (Figure 5.10). The difference was only minor, and therefore it was hard to tell whether it had been caused by the interruption and restoration of the schizokinen pathway in the KO and the CS strains, respectively. However, similarity of the three complementation strains was somewhat expected, despite the lack of either *all0395* or *all0396* in the single-knockout complementation strains. The enzymes encoded by these two genes are involved in the β-alanine pathway leading to pantothenic acid and eventually coenzyme A. Although the ability to synthesize coenzyme A is essential, an alternative route is operative in *Anabaena* sp. PCC 7120. The enzyme All3569 (EC 4.1.1.11) may cover for the loss of biochemical functionality encoded by *all0395* and *all0396*, providing a bypass directly from L-asp to β-alanine. In such a case, no phenotypic difference is detectable between a full and a partial (i.e. single-knockout) complementation strain. In contrast, the difference between wild type and schizokinen mutant strains was remarkable in the microplate assay under iron limitation.

The effect of iron starvation was assessed in BG-11_{chl} medium supplemented with the mild iron chelator 2,2'-bipyridyl and also in BG-11_{Fe} medium lacking the iron component (ferric ammonium citrate) of the standard medium. Growth of wild type *Anabaena* sp. PCC 7120 was significantly delayed in the chelated medium compared to BG-11, however, after about 8 days growth characteristics showed similar pattern to BG-11 in BG-11_{chl} as well. This observation suggests that the organism required a few days to adapt to the conditions and produce siderophores to free up iron from the other chelator. In contrast, the modified strains hardly showed any growth, suggesting severe growth deprivation due to low iron accessibility, even for the full complementation strain. The growth of these strains was similarly low in BG-11_{Fe}, and therefore difficult to evaluate. However, it was not the case for this medium type in the multicultivator. Both wild type and the Δ*sch* strains showed some growth, although severely deprived compared to the chelated and the standard media (Figure 5.9). Still, their ability to grow in that medium in the absence of any added iron may indicate other sources of the essential ion. Indeed, in case of this experiment glassware has not been cleaned specifically of any residual iron from previous uses. In addition, normal deionized water was used instead of e.g. HPLC-grade water for the preparation of the base BG-11_{Fe} medium that may have contained a sufficient amount of iron to maintain growth at the observed rate in the multicultivator. Regardless, the lower growth rate of the Δ*sch* knockout strain compared to the wild type under iron starvation but not under standard conditions (Figure 5.9) indicated that the introduced genetic modification affected the efficiency of this strain to cope with iron limitation. Therefore, it can be concluded that the operon is involved in the iron-response of *Anabaena* sp. PCC 7120 and that it may be responsible for siderophore production.

In future studies the schizokinen knockout strain (Δ*sch*) may serve as a starting point in the development of cross-feeding in a synthetic community, as discussed in the introduction of this chapter (last paragraph on page 187). The (algal) partner could be genetically engineered (for e.g. excretion of a siderophore) to provide accessible iron for the cyanobacterium, releasing fixed nitrogen (ammonia) in return under diazotrophic conditions. Since in this case both the engineered alga and the Δ*sch* strain will excrete some iron chelators, the induction of cross-feeding would require the addition of a weaker iron chelator to the culture medium (e.g. 2,2'-bipyridyl).

A promising candidate siderophore for cross-feeding iron may be ferrioxamine B. Ferrioxamine B is a hydroxamate-type bacterial siderophore, recognized and transported as a heterologous iron carrier by a variety of organisms (Allnutt and Bonner, 1987a; Wang et al., 1993; Lesuisse et al., 1998; Llamas et al., 2006). In case of the synthetic community described in section 4.2.3.5, both partners, *Chlorella vulgaris* and the nitrogen-excreting *Anabaena* sp. PCC 7120 mutant can natively use this siderophore.

The alga may free up iron from ferrioxamine B by reductive release (Allnutt and Bonner, 1987a, 1987b), whereas *Anabaena* sp. PCC 7120 is known to uptake this siderophore via an unknown transporter (Goldman et al., 1983; Rudolf et al., 2015). Thus, in such synthetic community it may be possible to promote cross-feeding via iron by expressing ferrioxamine B in the engineered algal partner. In the end, organic chelation of the essential iron may restrict the ability of invasive species to penetrate a synthetic community, improving the stability of the community.

6 Conclusions

The industrial production of cheap ammonia revolutionized the twentieth century and the following decades in many ways. Most importantly, more than half of the human population would not exist without the benefits of the Haber-Bosch process. The cheap fertilizer produced by this single process provides the basis of agricultural food production for billions of people today. Modern agricultural practices, however, have drastically disrupted the balance of the nitrogen cycle. The reason is about 40% of fertilizer nitrogen is lost to the environment due to the low nitrogen use efficiency of common crops, such as wheat, rice and maize. Thus, the nitrogen crisis is not a question anymore. Several million tonnes of reactive nitrogen leaches to the environment every year, causing severe effects on terrestrial, aquatic and atmospheric systems and that also influence human health and welfare. A study in 2011 estimated the environmental costs associated with tackling the negative effects of the nitrogen crisis to be tens of billions of euros per annum for the European Union only (Sutton et al., 2011a). Although important measures have been taken to decrease N_r in the environment, the problem is enormous and requires focussed intervention strategies from all major societies. In addition, given the complexities of N_r use, its environmental mobility, and differences among regions, no single strategy will suffice. Instead, a global strategy reducing N_r emissions by all major sources is necessary. Controlling NO_x emissions from fossil-fuel combustion, improving animal management, extending access to sewage treatment in cities worldwide and increasing nitrogen-uptake efficiency of crops would decrease annual N_r creation by about one-quarter according to Galloway et al. (2008).

The improvement of current fertilization practices in agriculture assumes better conversion of fertilizer nitrogen; one way is the development of symbiotic (preferably carbon-neutral) nitrogen-fixing organisms that provide nitrogen to agricultural crops more efficiently. To this end, the filamentous heterocyst-forming cyanobacterium, *Anabaena* sp. PCC 7120 was selected as the

platform for metabolic engineering towards diazotrophic nitrogen excretion. The biofertilizer potential of *Anabaena* species, in particular *Anabaena azollae*, a free-living cyanobacterium displaying high similarity to *Anabaena* sp. PCC 7120, has long been recognized, mainly as a side crop in rice paddies. The utilization of this cyanobacterium as a source of fixed nitrogen, however, has been restricted to the encapsulated form inside its natural host *Azolla caroliniana*. It was therefore necessary to understand the biochemical traits of *Anabaena* sp. PCC 7120 at the system level as an isolated "single cell" as well as a functional "filament".

A highly curated, genome-scale metabolic model was developed for this organism here for the first time. As yet, stoichiometric reconstructions of cyanobacterial metabolism have mainly focused on unicellular non-diazotrophs (Saha et al., 2012; Knoop et al., 2013) or non-heterocystous diazotrophs (Resendis-Antonio et al., 2007; Saha et al., 2012; Vu et al., 2012). The reconstruction presented here is a high-quality stoichiometric model of a filamentous heterocystous nitrogen-fixing cyanobacterium with two distinct cell-types. During the reconstruction process fifty-six genes and encoded proteins have been newly annotated and associated with a metabolic function based on sequence homology and, in a few cases, experimental observations. These new annotations represent an important contribution to the genetic information currently available for *Anabaena* sp. PCC 7120. It is important to note, the level of evidence for these new genes and proteins is low in most cases. Despite the low evidence, it was necessary to include these proteins (and the related GPR associations) to fill in critical gaps in the metabolic network of *Anabaena* sp. PCC 7120, or complete pathways for known non-essential metabolites.

Most of these annotatinons are based on homology of the particular (translated) gene sequence with protein sequences from closely related cyanobacteria. However, these homologous proteins themselves are of varying confidence, with data for confirmed enzymatic activity available only in a handful of cases. In fact, the vast majority of proteins annotated for *Anabaena* sp. PCC 7120 are the results of automated workflows that, similarly to the proteins suggested here, face the issue of low biological evindence. Out of the 6070 proteins listed for this organism in the UniProt database (Magrane and Consortium, 2011), about 80% is based on gene prediction algorithms, 18% is inferred from sequence homology and less than 2%, some 100 proteins, are confirmed at the transcript or protein level. The lack of compelling evidence for biochemical activity exhibited by a predicted protein directly affects the quality of a metabolic model and the predictions it can make. Therefore, although immensely time-consuming, it would be crucial to improve the quality of protein information in the future by systematically isolating proteins and characterizing their primary activity. As a first step on this path, the proposed pathway for schizokinen biosynthesis was investigated in more detail.

Six genes in the *all0390–all0396* genomic region have been identified as schizokinen-related, based on sequence homology and genetic conservation among different cyanobacteria (this work), phenotypic evidence for the presence of schizokinen in the *Anabaena* sp. PCC 7120 excretome (Simpson and Neilands, 1976), and earlier findings that the genomic region is actively expressed under iron limitation (Stevanovic et al., 2012; González et al., 2014). In addition, it has been shown, based on experimental and bioinformatics evidence, that the cluster comprises a single transcript, driven by a bidirectional promoter activated under iron limitation. The region was named *sch* operon for its likely role in the biosynthesis of schizokinen. Indeed, removal of the operon under iron-limited experimental conditions resulted in pronounced growth deprivation compared to the wild type.

A total of thirty-six gene candidates have been proposed to fill essential and non-essential metabolic gaps in the biochemical network of *Anabaena* sp. PCC 7120. Moreover, the extensive manual curation of every reaction in the reconstruction and the design of a detailed interactive network map enabled the identification and elimination of inconsistent reactions that represented roughly 30% of the total reactions found in current metabolic databases for this organism. This way it was possible to substantially improve the publicly available gene-protein-reaction information for *Anabaena* sp. PCC 7120. The stoichiometric model predicted the vegetative cell to heterocyst ratio in the range of experimental observations, and could extend the number of vegetative cells a single heterocyst can support to twenty by decreasing growth rate to 46% of the maximum value. Growth simulations on thirteen different carbon sources revealed that glucose and sucrose are among the highest yielding substrates, followed by pyruvate, proline, acetate and glutamate. These results were confirmed by growth rate experiments which showed a surprisingly good overall correlation between the two datasets. Therefore, the model not only provided a comprehensively curated blueprint for the genome-scale metabolic network of *Anabaena* sp. PCC 7120, but it colud also serve as an important computational tool enabling the design of engineering strategies for this widely studied nitrogen-fixing cyanobacterium. Even though *Anabaena* sp. PCC 7120 may primarily be suitable for laboratory research, it is a highly suited first target organism to assess proof-of-principle engineering strategies towards the sustainable production of combined nitrogen (Chaurasia and Apte, 2011) or other important bio-products (Heyer and Krumbein, 1991).

In a future work the stoichiometric model would greatly benefit from the inclusion of a more precise biomass equation, specific for *Anabaena* sp. PCC 7120 both under diazotrophic and non-diazotrophic conditions. Such an update of the current biomass equation to reflect the true elemental composition of the organism's biomass may resolve the discrepancy between the computationally predicted and experimentally observed growth with glutamine. Nevertheless, even an improved biomass equation

would not solve the issue observed with glycerol as a carbon source. Most likely due to an unconstrained transhydrogenase reaction, glycerol showed the highest growth in mixotrophic simulations, in disagreement with experimental results. In addition, the applied modelling framework, FBA is limited in accurately predicting specific physiological states, by not considering e.g. membrane permeability and pH, two critical factors that otherwise greatly affect microbial growth.

Based on computational insight to nitrogen excretion efficiency provided by the model, three metabolic engineering strategies have been carried out to produce ammonia under diazotrophic conditions. Two of these strategies, the replacement of native GS for an active-site mutant and the overexpression of the GS inactivator IF7A resulted in considerable amounts of ammonia excretion. Strains carrying the mutant version of GS with decreased specific activity for ammonia (denoted as *glnA*[p.D52S]) have been confirmed genetically and also by their moderate ammonia excretion. The IF7A mutants, on the other hand, were found to be successful in supporting a synthetic community as well. In the case of these mutant strains, it has been shown that promoters P_{nifHDK} and P_{petE} overexpressed IF7A successfully in diazotrophic co-cultivation, leading to excretion of ammonia, and ultimately, the maintenance of non-diazotrophic algal growth. Thus, ammonia excretion has been achieved for the first time by *Anabaena* sp. PCC 7120 under diazotrophic conditions, without supplementation of a GS inhibitor or severe disturbance of filament health (Grizeau et al., 2016). Similar results have been reported recently in other organisms (Ambrosio et al., 2017), even in *Anabaena* sp. PCC 7120 by CRISPRi-mediated gene silencing of *glnA* (Higo et al., 2018) or by the overproduction of heterocysts at the expense of culture health (Chaurasia and Apte, 2011). In fact, enhanced heterocyst frequency and the concomitant increase of nitrogen fixation rate due to *hetR* overexpression, in combination with IF7A overexpression may further enhance the biofertilizer potential, and therefore, the ecological benefit of such strains. Besides, it may be necessary to develop genomic integration of IF7A overexpression for added genetic stability. Moreover, although confirmed as a regulator of IF7 expression in *Synechocystis* sp. PCC 6803 (Klähn et al., 2015), the role of NsiR4 in *Anabaena* sp. PCC 7120 requires further investigation. The target mRNAs in the two organisms are very similar; however, NsiR4 homologues exhibit a remarkable difference. The sRNA in *Synechocystis* sp. PCC 6803 contains a 20-nt hairpin at the 5' end, likely to be involved in the binding of IF7 mRNA. NsiR4 in *Anabaena* sp. PCC 7120, on the other hand, lacks the short hairpin structure and therefore it is unclear, whether a similar regulation of IF7A expression by NsiR4 is present in this organism.

It would also be important to introduce mutualism to synthetic communities comprising of the ammonia-excreting cyanobacterium and a non-diazotroph partner to improve robustness of the co-culture (Kazamia et al., 2014). A logical choice for a cross-feeding commodity may be iron for two

reasons. First, iron is a cofactor in both the photosynthetic and nitrogen-fixnig apparatus, and thus essential for *Anabaena* sp. PCC 7120. Second, the control of iron homeostasis and the modulation of diazotrophic metabolism are interconnected (via FurA), therefore the external trigger for nitrogen excretion can be the starvation for iron. The role of chelated iron as the promoter of mutualism is not unknown. In case of marine microalga and some heterotrophic bacteria, iron and the photosensitive siderophore, vibrioferrin were found responsible for creating such relationships (Amin et al., 2009). Although structurally quite different from vibrioferrin, the two hydroxamates, schizokinen and ferrioxamine B might serve a similar purpose in engineered synthetic communities. But involvement in iron acquisition is not the only aspect of the cross-feeding potential of siderophores.

Another siderophore, the peptide-like azotobactin from *Azotobacter vinelandii* has been shown to provide a rich nitrogen source for green algae, supporting their growth under diazotrophic conditions (Villa et al., 2014). Notably, one molecule of azotobactin may contain up to 14 nitrogen atoms in a ten-amino-acid peptide chain, in contrast with 4 and 6 nitrogen atoms per molecule of schizokinen and ferrioxamine B, respectively (Neilands, 1995; Palanche et al., 2004). Regardless of their lower molar nitrogen content, the smaller siderophores are also worth investigating as potential sources of nitrogen. In this context ferrioxamine B may be a more suitable candidate, not just because of the slightly higher nitrogen content, but also because of the native uptake systems available for this siderophore in *Anabaena* sp. PCC 7120 and several potential partner organisms, including eukaryotic algae.

This work provides a proof-of-principle example for emerging community traits by engineering nitrogen excretion under diazotrophic conditions. A simple synthetic community comprising of a nitrogen-fixing cyanobacterium and a commensal eukaryotic alga was evaluated for growth and total biomass production. The co-cultivated alga was able to maintain its growth under diazotrophic conditions at 70% of the level the monoculture was growing in the presence of a combined nitrogen source. In spite the multiple uses of the co-cultivated alga, *Chlorella vulgaris* (Keffer and Kleinheinz, 2002; Scragg et al., 2003; Converti et al., 2009; Lim et al., 2010; Heredia-Arroyo et al., 2011), the impact of nitrogen-excreting cyanobacterial strains as synthetic symbionts of agricultural crops would be much more remarkable. As discussed above, nitrogen crisis and overfertilization are problems of the twenty-first century that cannot be ignored anymore. Among the intervention strategies to reduce the scale of annual N_r production, improvement of crop plants, fertilizer distribution and use efficiency of fertilizer nitrogen are ranking high. However, symbiotic nitrogen fixation is largely limited to legumes in agricultural systems, whilst non-legumes produce the bulk of human food (Mus et al., 2016). Encouraged by some recent works (Bocchi and Malgioglio, 2010; Chaurasia and Apte, 2011) it

would be interesting to follow up on the nitrogen provision potential of the engineered *Anabaena* sp. PCC 7120 strains in co-cultivation with crop plants like rice and wheat. From a broader perspective, the results presented here may provide the basis for future metabolic engineering strategies aiming at the excretion of fixed nitrogen by cyanobacteria or heterotrophic bacteria. The modification of GS activity by chemical inhibition or by active-site mutation have been performed with success in bacteria, e.g. *Azotobacter vinelandii* as well. In contrast, the use of GS-specific inactivating factors (IFs) to control the activity of GS is unique to cyanobacteria. The relative simplicity of these interacting peptides and the reasonable homology of bacterial GSs to the cyanobacterial enzyme, however, envisages the potential for heterologous expression of modified cyanobacterial IFs in engineered nitrogen-fixing bacteria. Mutualism would be important in these synthetic communities as well. Instead of simply cross-feeding essential commodities, the plant may be modified to encapsulate the microbial partner similar to an "endosymbiont", as it happens to rhizobia and legumes (Long, 1989) or to *Anabaena azollae* and the fern *Azolla*. This way the exchange of valuable nitrogen would become extremely efficient, by minimizing the loss of fixed nitrogen due to leaching, while the host plant also provides a protected environment, and possibly fixed carbon to its partner. Such mutualistic plant-bacterial symbioses are expected to be stable and highly efficient, even addressing the nitrogen crisis to some extent. Genetically modified crops, however, pose an additional challenge on top of the technological difficulties already implicated by the development of these organisms. Due to the generally negative public perception of genetically modified organisms (GMOs) and the stringent regulations on their applications (Bartsch, 2014; The Non-GMO Project, 2018), the introduction of these crops to agricultural practice cannot be immediate. Nonetheless, at some point in the near future, the question may simplify to comparing the risks associated with GMOs and the nitrogen crisis. Ultimately, their impact on human health and the ecosystem will tip the scales in favour for one or another.

Bibliography

Abbatiello SE, Mani DR, Schilling B, MacLean B, Zimmerman LJ, Feng X, Cusack MP, Sedransk N, Hall SC, Addona T, Allen S, Dodder NG, Ghosh M, Held JM, Hedrick V, Inerowicz HD, Jackson A, Keshishian H, Kim JW, Lyssand JS, Riley CP, Rudnick P, Sadowski P, Shaddox K, Smith D, Tomazela D, Wahlander A, Waldemarson S, Whitwell CA, You J, Zhang S, Kinsinger CR, Mesri M, Rodriguez H, Borchers CH, Buck C, Fisher SJ, Gibson BW, Liebler D, MacCoss M, Neubert TA, Paulovich A, Regnier F, Skates SJ, Tempst P, Wang M, Carr SA (2013) Design, implementation, and multi-site evaluation of a system suitability protocol for the quantitative assessment of instrument performance in LC-MRM-MS. Molecular & Cellular Proteomics

Abdulqader G, Barsanti L, Tredici MR (2000) Harvest of *Arthrospira platensis* from Lake Kossorom (Chad) and its household usage among the Kanembu. Journal of Applied Phycology **12:** 493-498

Adams DG (2000) Symbiotic interactions. *In* BA Whitton, M Potts, eds, The Ecology of Cyanobacteria: Their Diversity in Time and Space. Kluwer Academic Publishers, Dordrecht, pp 523-561

Adams DG, Duggan PS (2008) Cyanobacteria–bryophyte symbioses. Journal of Experimental Botany **59:** 1047-1058

Ahmed E, Holmstrom SJ (2014) Siderophores in environmental research: roles and applications. Microb Biotechnol **7:** 196-208

Aiello A (1985) Sloth hair: unanswered questions. *In* GG Montgomery, ed, The evolution and ecology of armadillos, sloths, and vermilinguas. Smithsonian Institution Press, Washington, D.C., pp 213–218

Albrecht-Gary AM, Crumbliss AL (1998) Coordination chemistry of siderophores: thermodynamics and kinetics of iron chelation and release. Met Ions Biol Syst **35:** 239-327

Albrecht M, Linden H, Sandmann G (1996) Biochemical characterization of purified zeta-carotene desaturase from *Anabaena* PCC 7120 after expression in *Escherichia coli*. Eur J Biochem **236:** 115-120

Alexander DB, Zuberer DA (1991) Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. Biology and Fertility of Soils **12:** 39-45

Alexandrova AN, Jorgensen WL (2007) Why urea eliminates ammonia rather than hydrolyzes in aqueous solution. The journal of physical chemistry. B **111:** 720-730

Allnutt FCT, Bonner WD (1987a) Characterization of iron uptake from ferrioxamine B by *Chlorella vulgaris*. Plant Physiology **85:** 746-750

Allnutt FCT, Bonner WD (1987b) Evaluation of reductive release as a mechanism for iron uptake from ferrioxamine b by *Chlorella vulgaris*. Plant Physiology **85:** 751-756

Altermann W, Kazmierczak J (2003) Archean microfossils: a reappraisal of early life on Earth. Res Microbiol **154:** 611-617

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Ambrosio R, Ortiz-Marquez JCF, Curatti L (2017) Metabolic engineering of a diazotrophic bacterium improves ammonium release and biofertilization of plants and microalgae. Metabolic Engineering **40:** 59-68

Amin SA (2010) The role of siderophores in algal-bacterial interactions in the marine environment. University of California, San Diego and San Diego State University

Amin SA, Green DH, Hart MC, Küpper FC, Sunda WG, Carrano CJ (2009) Photolysis of iron–siderophore chelates promotes bacterial–algal mutualism. Proceedings of the National Academy of Sciences **106:** 17071-17076

Anand BN (1998) Iron sulfur proteins and their synthetic analogues: Structure, reactivity and redox properties. Resonance **3:** 52-61

Anders E, Grevesse N (1989) Abundances of the elements: Meteoritic and solar. Geochimica et Cosmochimica Acta **53:** 197-214

Antoniewicz MR (2013) Dynamic metabolic flux analysis-tools for probing transient states of metabolic networks. Curr Opin Biotechnol **24:** 973-978

Armbrust EV, Bowen JD, Olson RJ, Chisholm SW (1989) Effect of light on the cell cycle of a marine *Synechococcus* strain. Applied and Environmental Microbiology **55:** 425-432

Asada K (1999) The water-water cycle in chloroplasts: scavenging of active oxygens and dissipation of excess photons. Annual Review of Plant Biology **50:** 601-639

Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat Biotech **27:** 1177-1180

Awai K, Kakimoto T, Awai C, Kaneko T, Nakamura Y, Takamiya K-i, Wada H, Ohta H (2006) Comparative genomic analysis revealed a gene for monoglucosyldiacylglycerol synthase, an enzyme for photosynthetic membrane lipid synthesis in cyanobacteria. Plant Physiology **141:** 1120-1127

Awai K, Lechno-Yossef S, Wolk CP (2010) Heterocyst envelope glycolipids. *In* H Wada, N Murata, eds, Lipids in Photosynthesis, Vol 30. Springer Netherlands, pp 179-202

Awai K, Watanabe H, Benning C, Nishida I (2007) Digalactosyldiacylglycerol is required for better photosynthetic growth of *Synechocystis* sp. PCC6803 under phosphate limitation. Plant Cell Physiol **48:** 1517-1523

Bailey KM, Taub FB (1980) Effects of hydroxamate siderophores (strong Fe(III) chelators) on the growth of algae. Journal of Phycology **16:** 334-339

Bandyopadhyay A, Elvitigala T, Liberton M, Pakrasi HB (2013) Variations in the rhythms of respiration and nitrogen fixation in members of the unicellular diazotrophic cyanobacterial genus *Cyanothece*. Plant Physiology **161:** 1334-1346

Bandyopadhyay A, Stockel J, Min H, Sherman LA, Pakrasi HB (2010) High rates of photobiological H_2 production by a cyanobacterium under aerobic conditions. Nat Commun **1:** 139

Bartsch D (2014) GMO regulatory challenges and science: a European perspective. Journal für Verbraucherschutz und Lebensmittelsicherheit **9:** 51-58

Battchikova N, Eisenhut M, Aro EM (2011) Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. Biochim Biophys Acta **1807:** 935-944

Beck C, Knoop H, Axmann IM, Steuer R (2012) The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. BMC Genomics **13:** 56

Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc **2:** 727-738

Becker SA, Palsson BØ (2005) Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. BMC Microbiology **5:** 8

Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ (2004) Dating the rise of atmospheric oxygen. Nature **427:** 117-120

Beliaev AS, Romine MF, Serres M, Bernstein HC, Linggi BE, Markillie LM, Isern NG, Chrisler WB, Kucek LA, Hill EA, Pinchuk GE, Bryant DA, Steven Wiley H, Fredrickson JK, Konopka A (2014) Inference of interactions in cyanobacterial-heterotrophic co-cultures via transcriptome sequencing. ISME J **8:** 2243-2255

Bellenger JP, Wichard T, Kustka AB, Kraepiel AML (2008) Uptake of molybdenum and vanadium by a nitrogen-fixing soil bacterium using siderophores. Nature Geosci **1:** 243-246

Berla BM, Pakrasi HB (2012) Upregulation of plasmid genes during stationary phase in *Synechocystis* sp. strain PCC 6803, a cyanobacterium. Appl Environ Microbiol **78:** 5448-5451

Berla BM, Saha R, Immethun CM, Maranas CD, Moon TS, Pakrasi HB (2013) Synthetic biology of cyanobacteria: unique challenges and opportunities. Frontiers in Microbiology **4:** 246

Berman-Frank I, Cullen JT, Shaked Y, Sherrell RM, Falkowski PG (2001) Iron availability, cellular iron quotas, and nitrogen fixation in *Trichodesmium*. Limnology and Oceanography **46:** 1249-1260

Berman-Frank I, Lundgren P, Falkowski P (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Research in Microbiology **154:** 157-164

Bernhard A (2010) The nitrogen cycle: processes, players and human impact. Nature Education Knowledge **3:** 25

Bertsimas D, Tsitsiklis J (1997) Introduction to Linear Optimization. Athena Scientific

Binder BJ, Chisholm SW (1995) Cell cycle regulation in marine *Synechococcus* sp. Strains. Applied and Environmental Microbiology **61:** 708-717

Biswas A (2011) Identification and characterization of enzymes involved in the biosynthesis of different phycobiliproteins in cyanobacteria. University of New Orleans, University of New Orleans Theses and Dissertations

Bloom FR, Levin MS, Foor F, Tyler B (1978) Regulation of glutamine synthetase formation in *Escherichia coli*: characterization of mutants lacking the uridylyltransferase. J Bacteriol **134:** 569-577

Bocchi S, Malgioglio A (2010) *Azolla-Anabaena* as a biofertilizer for rice paddy fields in the Po Valley, a temperate rice area in northern Italy. International Journal of Agronomy **2010**

Boele J, Olivier B, Teusink B (2012) FAME, the Flux Analysis and Modeling Environment. BMC Systems Biology **6:** 8

Böhme H (1998) Regulation of nitrogen fixation in heterocyst-forming cyanobacteria. Trends in Plant Science **3:** 346-351

Boogerd FC, Ma H, Bruggeman FJ, van Heeswijk WC, Garcia-Contreras R, Molenaar D, Krab K, Westerhoff HV (2011) AmtB-mediated NH_3 transport in prokaryotes must be active and as a consequence regulation of transport by GlnK is mandatory to limit futile cycling of NH_4^+/NH_3. FEBS Lett **585:** 23-28

Bothe H, Nolteernsting U (1975) Pyruvate dehydrogenase complex, pyruvate: Ferredoxin oxidoreductase and lipoic acid content in microorganisms. Archives of Microbiology **102:** 53-57

Bowen HJM (1966) Trace elements in biochemistry. Academic Press, London-New York

Boyle N, Morgan J (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. BMC Systems Biology **3:** 4

Braud A, Jezequel K, Bazot S, Lebeau T (2009) Enhanced phytoextraction of an agricultural Cr- and Pb-contaminated soil by bioaugmentation with siderophore-producing bacteria. Chemosphere **74:** 280-286

Braun E (2007) Reactive nitrogen in the environment: too much or too little of a good thing. UNEP DTIE, Sustainable Consumption and Production (SCP) Branch, Paris

Bressler SL, Ahmed SI (1984) Detection of glutamine synthetase activity in marine phytoplankton: optimization of the biosynthetic assay. Mar. Ecol. Prog. Ser **14:** 207-217

Brock TD (1973) Lower pH limit for the existence of blue-green algae: Evolutionary and ecological implications. Science **179:** 480-483

Broda E (1977) Two kinds of lithotrophs missing in nature. Zeitschrift für allgemeine Mikrobiologie **17:** 491-493

Broda E, Peschek GA (1983) Nitrogen fixation as evidence for the reducing nature of the early biosphere. Biosystems **16:** 1-8

Bryksin AV, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. BioTechniques **48:** 463-465

Buikema WJ, Haselkorn R (2001) Expression of the Anabaena hetR gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. Proceedings of the National Academy of Sciences of the United States of America **98:** 2729-2734

Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng **84:** 647-657

Busch A, Richter AS, Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics **24:** 2849-2856

Cabello P, Roldan MD, Moreno-Vivian C (2004) Nitrate reduction and the nitrogen cycle in archaea. Microbiology **150:** 3527-3546

Cai YP, Wolk CP (1990) Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. Journal of Bacteriology **172:** 3138-3145

Canfield DE, Glazer AN, Falkowski PG (2010) The Evolution and Future of Earth's Nitrogen Cycle. Science **330:** 192-196

Carmichael WW (1992) Cyanobacteria secondary metabolites-the cyanotoxins. J Appl Bacteriol **72:** 445-459

Carrasco CD, Buettner JA, Golden JW (1995) Programmed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts. Proc Natl Acad Sci U S A **92:** 791-795

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research **40:** D742-D753

Chatterjee A, Majee M, Ghosh S, Majumder AL (2004) *sll1722*, an unassigned open reading frame of *Synechocystis* PCC 6803, codes for L-myo-inositol 1-phosphate synthase. Planta **218:** 989-998

Chaurasia AK, Apte SK (2011) Improved eco-friendly recombinant *Anabaena* sp. strain PCC 7120 with enhanced nitrogen biofertilizer potential. Applied and Environmental Microbiology **77:** 395-399

Chávez S, Lucena JM, Reyes JC, Florencio FJ, Candau P (1999) The presence of glutamate dehydrogenase is a selective advantage for the cyanobacterium *Synechocystis* sp. strain PCC 6803 under nonexponential growth conditions. Journal of Bacteriology **181:** 808-813

Cheah YE, Albers SC, Peebles CA (2013) A novel counter-selection method for markerless genetic modification in *Synechocystis* sp. PCC 6803. Biotechnol Prog **29:** 23-30

Chen W-H, Minguez P, Lercher MJ, Bork P (2012) OGEE: an online gene essentiality database. Nucleic Acids Research **40:** D901-D906

Cheng Z, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D (2003) Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. Plant Cell **15:** 2343-2356

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Res **40:** D700-705

Chinnici JP, Audesirk TE, Wadkowski SM, Audesirk GJ (1998) Biology: Life on Earth. *In*. Prentice Hall

Choi KH, Schweizer HP (2006) mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. Nat Protoc **1:** 153-161

Chung P, Pond WG, Kingsbury JM, Walker EF, Krook L (1978) Production and nutritive value of *Arthrospira platensis*, a spiral blue-green alga grown on swine wastes. Journal of Animal Science **47:** 319-330

Colnaghi R, Green A, He L, Rudnick P, Kennedy C (1997) Strategies for increased ammonium production in free-living or plant associated nitrogen fixing bacteria. Plant and Soil **194:** 145-154

Colón-López MS, Sherman DM, Sherman LA (1997) Transcriptional and translational regulation of nitrogenase in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium *Cyanothece* sp. strain ATCC 51142. Journal of Bacteriology **179:** 4319-4327

Converti A, Casazza AA, Ortiz EY, Perego P, Del Borghi M (2009) Effect of temperature and nitrogen concentration on the growth and lipid content of *Nannochloropsis oculata* and *Chlorella vulgaris* for biodiesel production. Chemical Engineering and Processing: Process Intensification **48:** 1146-1151

Crespo JL, Guerrero MG, Florencio FJ (1999) Mutational analysis of Asp51 of *Anabaena azollae* glutamine synthetase. D51E mutation confers resistance to the active site inhibitors L-methionine-DL-sulfoximine and phosphinothricin. Eur J Biochem **266:** 1202-1209

Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG (2005) Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. Nature **438:** 90-93

Croswell K (1995) The alchemy of the heavens: searching for meaning in the Milky Way. Anchor Books

Cuiv PO, Clarke P, O'Connell M (2006) Identification and characterization of an iron-regulated gene, *chtA*, required for the utilization of the xenosiderophores aerobactin, rhizobactin 1021 and schizokinen by *Pseudomonas aeruginosa*. Microbiology **152:** 945-954

Cumino AC, Marcozzi C, Barreiro R, Salerno GL (2007) Carbon cycling in *Anabaena* sp. PCC 7120. Sucrose synthesis in the heterocysts and possible role in nitrogen fixation. Plant Physiol **143:** 1385-1397

Curatti L, Flores E, Salerno G (2002) Sucrose is involved in the diazotrophic metabolism of the heterocyst-forming cyanobacterium *Anabaena* sp. FEBS Lett **513:** 175-178

Curran KA, Alper HS (2012) Expanding the chemical palate of cells by combining systems biology and metabolic engineering. Metabolic Engineering **14:** 289-297

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences **99:** 10494-10499

Dai G-Z, Qiu B-S, Forchhammer K (2014) Ammonium tolerance in the cyanobacterium *Synechocystis* sp. strain PCC 6803 and the role of the *psbA* multigene family. Plant, Cell & Environment **37:** 840-851

de Lorenzo V, Bindereif A, Paw BH, Neilands JB (1986) Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. J Bacteriol **165:** 570-578

de Lorenzo V, Neilands JB (1986) Characterization of *iucA* and *iucC* genes of the aerobactin system of plasmid ColV-K30 in *Escherichia coli*. Journal of Bacteriology **167:** 350-355

de Macale M, Vlek PG (2004) The role of *Azolla* cover in improving the nitrogen use efficiency of lowland rice. Plant and Soil **263:** 311-321

de Vries FT, Shade A (2013) Controls on soil microbial community stability under climate change. Frontiers in Microbiology **4:** 265

Deng M-D, Coleman JR (1999) Ethanol synthesis by genetic engineering in cyanobacteria. Applied and Environmental Microbiology **65:** 523-528

Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat Genet **38:** 993-998

Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine ecosystems. Science **321:** 926-929

Domain F, Houot L, Chauvat F, Cassier-Chauvat C (2004) Function and regulation of the cyanobacterial genes *lexA*, *recA* and *ruvB*: LexA is critical to the survival of cells facing inorganic carbon starvation. Mol Microbiol **53:** 65-80

Drath M, Kloft N, Batschauer A, Marin K, Novak J, Forchhammer K (2008) Ammonia triggers photodamage of photosystem II in the cyanobacterium *Synechocystis* sp. strain PCC 6803. Plant Physiology **147:** 206-215

Du W, Liang F, Duan Y, Tan X, Lu X (2013) Exploring the photosynthetic production capacity of sucrose by cyanobacteria. Metabolic Engineering **19:** 17-25

Ducat DC, Way JC, Silver PA (2011) Engineering cyanobacteria to generate high-value products. Trends Biotechnol **29:** 95-103

Eady RR (1996) Structure-function relationships of alternative nitrogenases. Chem Rev **96:** 3013-3030

Eaton-Rye JJ (2004) The construction of gene knockouts in the cyanobacterium *Synechocystis* sp. PCC 6803. *In* R Carpentier, ed, Photosynthesis Research Protocols. Humana Press, Totowa, NJ, pp 309-324

Edwards JS, Ibarra RU, Palsson BO (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol **19:** 125-130

Ehira S (2013) Transcriptional regulation of heterocyst differentiation in *Anabaena* sp. strain PCC 7120. Russian Journal of Plant Physiology **60:** 443-452

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proceedings of the National Academy of Sciences **105:** 17199-17204

Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, Wolk CP (1997) Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **179:** 1998-2005

Elhai J, Wolk CP (1988) Conjugal transfer of DNA to cyanobacteria. Methods Enzymol **167:** 747-754

Erisman JW, Sutton MA, Galloway J, Klimont Z, Winiwarter W (2008) How a century of ammonia synthesis changed the world. Nature Geosci **1:** 636-639

Espinosa J, Rodríguez-Mateos F, Salinas P, Lanza VF, Dixon R, de la Cruz F, Contreras A (2014) PipX, the coactivator of NtcA, is a global regulator in cyanobacteria. Proceedings of the National Academy of Sciences **111:** E2423-E2430

Falkowski PG (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature **387:** 272-275

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science **320:** 1034-1039

Fay P (1992) Oxygen relations of nitrogen fixation in cyanobacteria. Microbiological Reviews **56:** 340-373

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson B (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol **3:** 121

Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol **7:** 129-143

Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol **13:** 344-349

Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochemical Journal **238:** 781-786

Ferguson TS (2000) Linear programming: A concise introduction. Website. Available at https://www.math.ucla.edu/~tom/LP.pdf

Fiedler G, Arnold M, Hannus S, Maldener I (1998) The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. Molecular Microbiology **27:** 1193-1202

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research **44:** D279-D285

Fixen PE, West FB (2002) Nitrogen fertilizers: meeting contemporary challenges. Ambio **31:** 169-176

Flaherty B, Van Nieuwerburgh F, Head S, Golden J (2011) Directional RNA deep sequencing sheds new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. BMC Genomics **12:** 332

Flamholz A, Noor E, Bar-Even A, Milo R (2012) eQuilibrator—the biochemical thermodynamics calculator. Nucleic Acids Research **40:** D770-D775

Flores E, Herrero A (2004) Assimilatory nitrogen metabolism and its regulation. *In* D Bryant, ed, The Molecular Biology of Cyanobacteria, Vol 1. Springer Netherlands, pp 487-517

Flores E, Herrero A (2010) Compartmentalized function through cell differentiation in filamentous cyanobacteria. Nat Rev Micro **8:** 39-50

Flores E, Herrero A, Wolk CP, Maldener I (2006) Is the periplasm continuous in filamentous multicellular cyanobacteria? Trends Microbiol **14:** 439-443

Flores E, Pernil R, Muro-Pastor AM, Mariscal V, Maldener I, Lechno-Yossef S, Fan Q, Wolk CP, Herrero A (2007) Septum-localized protein required for filament integrity and diazotrophy in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. J Bacteriol **189:** 3884-3890

Forchhammer K (1999) The PII protein in *Synechococcus* PCC 7942 senses and signals 2-oxoglutarate under ATP-replete conditions. *In* G Peschek, W Löffelhardt, G Schmetterer, eds, The Phototrophic Prokaryotes. Springer US, pp 549-553

Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schultz M, Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. *In* S Solomon, D Qin, M Manning, Z Chen, M Marquis, KB Averyt, M Tignor, HL Miller, eds, Climate change 2007: The physical science basis. Cambridge University Press, Cambridge, United Kingdom, pp 129-234

Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, Vitousek P, Leach A, Bouwman AF, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M, Voss M (2013) The global nitrogen cycle in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **368**

Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E (2011) Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun **2:** 589

Fujita Y, Takagi H, Hase T (1996) Identification of the *chlb* gene and the gene product essential for the light-independent chlorophyll biosynthesis in the cyanobacterium *Plectonema boryanum*. Plant and Cell Physiology **37:** 313-323

Fukuzawa H, Suzuki E, Komukai Y, Miyachi S (1992) A gene homologous to chloroplast carbonic anhydrase (*icfA*) is essential to photosynthetic carbon dioxide fixation by *Synechococcus* PCC 7942. Proceedings of the National Academy of Sciences **89:** 4437-4441

Galloway JN (1998) The global nitrogen cycle: changes and consequences. Environmental Pollution **102:** 15-24

Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW, Seitzinger SP, Asner GP, Cleveland CC, Green PA, Holland EA, Karl DM, Michaels AF, Porter JH, Townsend AR, Vöosmarty CJ (2004) Nitrogen cycles: Past, present, and future. Biogeochemistry **70:** 153-226

Galloway JN, Schlesinger WH, Levy H, Michaels A, Schnoor JL (1995) Nitrogen fixation: Anthropogenic enhancement-environmental response. Global Biogeochemical Cycles **9:** 235-252

Galloway JN, Townsend AR, Erisman JW, Bekunda M, Cai Z, Freney JR, Martinelli LA, Seitzinger SP, Sutton MA (2008) Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. Science **320:** 889-892

Galmozzi CV, Saelices L, Florencio FJ, Muro-Pastor MI (2010) Posttranscriptional regulation of glutamine synthetase in the filamentous cyanobacterium *Anabaena* sp. PCC 7120: differential expression between vegetative cells and heterocysts. J Bacteriol **192:** 4701-4711

Gangl D, Zedler JA, Rajakumar PD, Martinez EM, Riseley A, Wlodarczyk A, Purton S, Sakuragi Y, Howe CJ, Jensen PE, Robinson C (2015) Biotechnological exploitation of microalgae. J Exp Bot **66:** 6975-6990

García-Domínguez M, Reyes JC, Florencio FJ (1999) Glutamine synthetase inactivation by protein–protein interaction. Proceedings of the National Academy of Sciences **96:** 7161-7166

Gaudet M, Fara A-G, Beritognolo I, Sabatti M (2009) Allele-specific PCR in SNP genotyping. *In* AA Komar, ed, Single Nucleotide Polymorphisms: Methods and Protocols. Humana Press, Totowa, NJ, pp 415-424

Geider RJ (1999) Complex lessons of iron uptake. Nature **400:** 815

Georg J, Kostova G, Vuorijoki L, Schon V, Kadowaki T, Huokko T, Baumgartner D, Muller M, Klahn S, Allahverdiyeva Y, Hihara Y, Futschik ME, Aro EM, Hess WR (2017) Acclimation of oxygenic photosynthesis to iron starvation is controlled by the sRNA IsaR1. Curr Biol **27:** 1425-1436.e1427

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

Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL (1994) Structure of curacin A, a novel antimitotic, antiproliferative and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. The Journal of Organic Chemistry **59:** 1243-1245

Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. Methods Enzymol **498:** 349-361

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods **6:** 343-345

Gibson F, Magrath DI (1969) The isolation and characterization of a hydroxamic acid (aerobactin) formed by *Aerobacter aerogenes* 62-1. Biochimica et Biophysica Acta (BBA) - General Subjects **192:** 175-184

Giddings TH, Staehelin LA (1981) Observation of microplasmodesmata in both heterocyst-forming and non-heterocyst forming filamentous cyanobacteria by freeze-fracture electron microscopy. Archives of Microbiology **129:** 295-298

Giraldez-Ruiz N, Mateo P, Bonilla I, Fernandez P, xf, as F (1997) The relationship between intracellular pH, growth characteristics and calcium in the cyanobacterium *Anabaena* sp. strain PCC 7120 exposed to low pH. The New Phytologist **137:** 599-605

Girvan MS, Campbell CD, Killham K, Prosser JI, Glover LA (2005) Bacterial diversity promotes community stability and functional resilience after perturbation. Environmental Microbiology **7:** 301-313

Golden JW, Carrasco CD, Mulligan ME, Schneider GJ, Haselkorn R (1988) Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **170:** 5034-5041

Golden JW, Robinson SJ, Haselkorn R (1985) Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. Nature **314:** 419-423

Golden JW, Wiest DR (1988) Genome rearrangement and nitrogen fixation in *Anabaena* blocked by inactivation of *xisA* gene. Science **242:** 1421-1423

Golden JW, Yoon HS (2003) Heterocyst development in *Anabaena*. Curr Opin Microbiol **6:** 557-563

Golden SS, Brusslan J, Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods Enzymol **153:** 215-231

Goldman SJ, Lammers PJ, Berman MS, Sanders-Loehr J (1983) Siderophore-mediated iron uptake in different strains of *Anabaena* sp. J Bacteriol **156:** 1144-1150

Gomez JA, Hoffner K, Barton PI (2014) DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis. BMC Bioinformatics **15:** 409

González A, Angarica VE, Sancho J, Fillat MF (2014) The FurA regulon in *Anabaena* sp. PCC 7120: in silico prediction and experimental validation of novel target genes. Nucleic Acids Research **42:** 4833-4846

González A, Valladares A, Peleato ML, Fillat MF (2013) FurA influences heterocyst differentiation in *Anabaena* sp. PCC 7120. FEBS Letters **587:** 2682-2690

Goodman D (1975) The theory of diversity-stability relationships in ecology. The Quarterly Review of Biology **50:** 237-266

Graham JE, Bryant DA (2009) The biosynthetic pathway for myxol-2' fucoside (myxoxanthophyll) in the cyanobacterium *Synechococcus* sp. strain PCC 7002. J Bacteriol **191:** 3292-3300

Griese M, Lange C, Soppa J (2011) Ploidy in cyanobacteria. FEMS Microbiology Letters **323:** 124-131

Grizeau D, Bui LA, Dupre C, Legrand J (2016) Ammonium photo-production by heterocytous cyanobacteria: potentials and constraints. Crit Rev Biotechnol **36:** 607-618

Großkopf T, Soyer OS (2014) Synthetic microbial communities. Current Opinion in Microbiology **18:** 72-77

Gruber N, Galloway JN (2008) An Earth-system perspective of the global nitrogen cycle. Nature **451:** 293-296

Guarino LA, Cohen SS (1979) Mechanism of toxicity of putrescine in *Anacystis nidulans*. Proc Natl Acad Sci U S A **76:** 3660-3664

Guerrero F, Carbonell V, Cossu M, Correddu D, Jones PR (2012) Ethylene synthesis and regulated expression of recombinant protein in *Synechocystis* sp. PCC 6803. PLOS ONE **7:** e50470

Güler S, Essigmann B, Benning C (2000) A cyanobacterial gene, *sqdX*, required for biosynthesis of the sulfolipid sulfoquinovosyldiacylglycerol. Journal of Bacteriology **182:** 543-545

Gurobi Optimization I (2013) Gurobi Optimizer reference manual. *In*,

Gutzke G, Fischer B, Mendel RR, Schwarz G (2001) Thiocarboxylation of molybdopterin synthase provides evidence for the mechanism of dithiolene formation in metal-binding pterins. Journal of Biological Chemistry **276:** 36268-36274

Haas H, Eisendle M, Turgeon BG (2008) Siderophores in fungal physiology and virulence. Annual Review of Phytopathology **46:** 149-187

Haber F, van Oordt G (1905) Über die Bildung von Ammoniak den Elementen. Zeitschrift für anorganische Chemie **44:** 341-378

Handelsman J (2005) Sorting out metagenomes. Nat Biotech **23:** 38-39

Harano Y, Suzuki I, Maeda S, Kaneko T, Tabata S, Omata T (1997) Identification and nitrogen regulation of the cyanase gene from the cyanobacteria *Synechocystis* sp. strain PCC 6803 and *Synechococcus* sp. strain PCC 7942. Journal of Bacteriology **179:** 5744-5750

Hay ME, Parker JD, Burkepile DE, Caudill CC, Wilson AE, Hallinan ZP, Chequer AD (2004) Mutualisms and aquatic community structure: The enemy of my enemy is my friend. Annual Review of Ecology, Evolution, and Systematics **35:** 175-197

Hayatsu M, Tago K, Saito M (2008) Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. Soil Science & Plant Nutrition **54:** 33-45

Head IM, Hiorns WD, Embley TM, McCarthy AJ, Saunders JR (1993) The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16S ribosomal RNA gene sequences. J Gen Microbiol **139 Pt 6:** 1147-1153

Heidorn T, Camsund D, Huang H-H, Lindberg P, Oliveira P, Stensjö K, Lindblad P (2011) Synthetic biology in cyanobacteria. Methods in Enzymology **497:** 539-579

Heredia-Arroyo T, Wei W, Ruan R, Hu B (2011) Mixotrophic cultivation of *Chlorella vulgaris* and its potential application for the oil accumulation from non-sugar materials. Biomass and Bioenergy **35:** 2245-2253

Hernandez JA, Lopez-Gomollon S, Bes MT, Fillat MF, Peleato ML (2004) Three *fur* homologues from *Anabaena* sp. PCC7120: exploring reciprocal protein-promoter recognition. FEMS Microbiol Lett **236:** 275-282

Herrero A, Muro-Pastor AM, Flores E (2001) Nitrogen control in cyanobacteria. Journal of Bacteriology **183:** 411-425

Herrero A, Muro-Pastor AM, Valladares A, Flores E (2004) Cellular differentiation and the NtcA transcription factor in filamentous cyanobacteria. FEMS Microbiology Reviews **28:** 469-487

Hess WR (2011) Cyanobacterial genomics for ecology and biotechnology. Curr Opin Microbiol **14:** 608-614

Heyer H, Krumbein W (1991) Excretion of fermentation products in dark and anaerobically incubated cyanobacteria. Archives of Microbiology **155:** 284-287

Higo A, Isu A, Fukaya Y, Ehira S, Hisabori T (2018) Application of CRISPR interference for metabolic engineering of the heterocyst-forming multicellular cyanobacterium *Anabaena* sp. PCC 7120. Plant and Cell Physiology **59:** 119-127

Hill DJ (1977) The role of *Anabaena* in the *Azolla-Anabaena* symbiosis. New Phytologist **78:** 611-616

Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) COPASI--a COmplex PAthway SImulator. Bioinformatics **22:** 3067-3074

Hopkinson BM, Morel FMM (2009) The role of siderophores in iron acquisition by photosynthetic marine microorganisms. BioMetals **22:** 659-669

Howarth RW (2008) Coastal nitrogen pollution: A review of sources and trends globally and regionally. Harmful Algae **8:** 14-20

Hu B, Yang G, Zhao W, Zhang Y, Zhao J (2007) MreB is important for cell shape but not for chromosome segregation of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. Mol Microbiol **63:** 1640-1652

Huang HH, Camsund D, Lindblad P, Heidorn T (2010) Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. Nucleic Acids Res **38:** 2577-2593

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Forum: atrotS, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics **19:** 524-531

Hutchins DA, Witter AE, Butler A, Luther Iii GW (1999) Competition among marine phytoplankton for different chelated iron species. Nature **400:** 858

Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature **420:** 186-189

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet **2:** 343-372

Incharoensakdi A, Jantaro S, Raksajit W, Mäenpää P (2010) Polyamines in cyanobacteria: biosynthesis, transport and abiotic stress response. *In* A Méndez-Vilas, ed, Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, Vol Vol 1. Formatex, Spain, pp 23-32

Iniguez AL, Dong Y, Triplett EW (2004) Nitrogen fixation in wheat provided by *Klebsiella pneumoniae* 342. Mol Plant Microbe Interact **17:** 1078-1085

Islam MR, Aikawa S, Midorikawa T, Kashino Y, Satoh K, Koike H (2008) *slr1923* of *Synechocystis* sp. PCC6803 is essential for conversion of 3,8-divinyl(proto)chlorophyll(ide) to 3-monovinyl(proto)chlorophyll(ide). Plant Physiol **148:** 1068-1081

Jantaro S, Mäenpää P, Mulo P, Incharoensakdi A (2003) Content and biosynthesis of polyamines in salt and osmotically stressed cells of *Synechocystis* sp. PCC 6803. FEMS Microbiology Letters **228:** 129-135

Jeanjean R, Talla E, Latifi A, Havaux M, Janicki A, Zhang CC (2008) A large gene cluster encoding peptide synthetases and polyketide synthases is involved in production of siderophores and oxidative stress response in the cyanobacterium *Anabaena* sp. strain PCC 7120. Environ Microbiol **10:** 2574-2585

Jessup CM, Kassen R, Forde SE, Kerr B, Buckling A, Rainey PB, Bohannan BJM (2004) Big questions, small worlds: microbial model systems in ecology. Trends in Ecology & Evolution **19:** 189-197

Joyce AR, Palsson BØ (2008) Predicting gene essentiality using genome-scale in silico models. *In* AL Osterman, SY Gerdes, eds, Microbial Gene Essentiality: Protocols and Bioinformatics. Humana Press, Totowa, NJ, pp 433-457

Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, Kschischo M (2010) grofit: Fitting biological growth curves with **R**. 2010 **33:** 21

Kämäräinen J, Huokko T, Kreula S, Jones PR, Aro EM, Kallio P (2017) Pyridine nucleotide transhydrogenase PntAB is essential for optimal growth and photosynthetic integrity under low-light mixotrophic conditions in *Synechocystis* sp. PCC 6803. New Phytologist **214:** 194-204

Kamennaya NA, Chernihovsky M, Post AF (2008) The cyanate utilization capacity of marine unicellular cyanobacteria. Limnology and Oceanography **53:** 2485-2494

Kamennaya NA, Post AF (2011) Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. Applied and Environmental Microbiology **77:** 291-301

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Research **32:** D277-D280

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M, Tabata S (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res **8:** 205-213; 227-253

Kannaiyan S, Rao KK, Hall DO (1994) Immobilization of *Anabaena azollae* from *Azolla filiculoides* in polyvinyl foam for ammonia production in a photobioreactor system. World Journal of Microbiology and Biotechnology **10:** 55-58

Katoh H, Hagino N, Grossman AR, Ogawa T (2001) Genes essential to iron transport in the cyanobacterium *Synechocystis* sp. strain PCC 6803. Journal of Bacteriology **183:** 2779-2784

Kazamia E, Aldridge DC, Smith AG (2012) Synthetic ecology – A way forward for sustainable algal biofuel production? Journal of Biotechnology **162:** 163-169

Kazamia E, Czesnick H, Nguyen TTV, Croft MT, Sherwood E, Sasso S, Hodson SJ, Warren MJ, Smith AG (2012) Mutualistic interactions between vitamin B_{12}-dependent algae and heterotrophic bacteria exhibit regulation. Environmental Microbiology **14:** 1466-1476

Kazamia E, Riseley AS, Howe CJ, Smith AG (2014) An engineered community approach for industrial cultivation of microalgae. Industrial Biotechnology **10:** 184-190

Keffer J, Kleinheinz G (2002) Use of *Chlorella vulgaris* for CO_2 mitigation in a photobioreactor. Journal of Industrial Microbiology and Biotechnology **29:** 275-280

Keith AR, Bailey JK, Whitham TG (2010) A genetic basis to community repeatability and stability. Ecology **91:** 3398-3406

Keren N, Aurora R, Pakrasi HB (2004) Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. Plant Physiology **135:** 1666-1673

Kinzig AP, Socolow RH (1994) Human impacts on the nitrogen cycle. Physics Today **47:** 24-31

Kirn J, Rees DC (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. Nature **360:** 553-560

Kitano H (2002) Systems biology: a brief overview. Science **295:** 1662-1664

Klähn S, Schaal C, Georg J, Baumgartner D, Knippen G, Hagemann M, Muro-Pastor AM, Hess WR (2015) The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. Proceedings of the National Academy of Sciences **112:** E6243-E6252

Klähn S, Schaal C, Georg J, Baumgartner D, Knippen G, Muro-Pastor AM, Hess WR (2014) The sRNA nsiR4 is involved in controlling nitrogen assimilation by posttranscriptional regulation of glutamine synthetase inactivation factor IF7. *In* 9th European Workshop on the Molecular Biology of Cyanobacteria. Royal Netherlands Institute of Sea Research Department of Marine Microbiologie, Texel, the Netherlands, p 29

Klamt S, Stelling J, Ginkel M, Gilles ED (2003) FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. Bioinformatics **19:** 261-269

Kleiner D (1981) The transport of NH_3 and NH_4^+ across biological membranes. Biochim Biophys Acta **639:** 41-52

Klemke F, Beyer G, Sawade L, Saitov A, Korte T, Maldener I, Lockau W, Nürnberg DJ, Volkmer T (2014) All1371 is a polyphosphate-dependent glucokinase in *Anabaena* sp. PCC 7120. Microbiology **160:** 2807-2819

Kliphuis AM, Klok AJ, Martens DE, Lamers PP, Janssen M, Wijffels RH (2012) Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. J Appl Phycol **24:** 253-266

Klitgord N, Segrè D (2010) Environments that induce synthetic microbial ecosystems. PLoS Comput Biol **6:** e1001002

Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, Steuer R (2013) Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. PLoS Comput Biol **9:** e1003081

Koksharova O, Wolk C (2002) Genetic tools for cyanobacteria. Applied Microbiology and Biotechnology **58:** 123-137

Krishnakumar S, Durai DA, Wangikar PP, Viswanathan GA (2013) SHARP: genome-scale identification of gene-protein-reaction associations in cyanobacteria. Photosynth Res

Kumar K, Mella-Herrera RA, Golden JW (2010) Cyanobacterial heterocysts. Cold Spring Harb Perspect Biol **2:** a000315

Kustka AB, Sañudo-Wilhelmy SA, Carpenter EJ, Capone D, Burns J, Sunda WG (2003a) Iron requirements for dinitrogen- and ammonium-supported growth in cultures of *Trichodesmium* (IMS 101): Comparison with nitrogen fixation rates and iron: carbon ratios of field populations. Limnology and Oceanography **48:** 1869-1884

Kustka AB, Sañudo-Wilhelmy SA, Carpenter EJ, Capone DG, Raven JA (2003b) A revised estimate of the iron use efficiency of nitrogen fixation, with special reference to the marine cyanobacterium *Trichodesmium* spp. (cyanophyta). Journal of Phycology **39:** 12-25

Lalioti MD, Heath JK (2001) A new method for generating point mutations in bacterial artificial chromosomes by homologous recombination in *Escherichia coli*. Nucleic Acids Research **29:** e14-e14

Lamar C (1968) The duration of the inhibition of glutamine synthetase by methionine sulfoximine. Biochemical Pharmacology **17:** 636-640

Lammers PJ (1982) Iron acquisition by cyanobacteria: siderophore production and iron transport by *Anabaena*. Portland State University, Portland

Lawrence MG, Chameides WL, Kasibhatla PS, II HL, Moxim W (1995) Lightning and atmospheric chemistry: The rate of atmospheric NO production. *In* H Volland, ed, Handbook of Atmospheric Electrodynamics, Vol 1. CRC Press, Boca Raton, Florida

Le Novere N, Shimizu TS (2001) STOCHSIM: modelling of stochastic biomolecular processes. Bioinformatics **17:** 575-576

Lechno-Yossef S, Nierzwicki-Bauer S (2002) *Azolla-Anabaena* symbiosis. *In* A Rai, B Bergman, U Rasmussen, eds, Cyanobacteria in Symbiosis. Springer Netherlands, pp 153-178

Lesuisse E, Simon-Casteras M, Labbe P (1998) Siderophore-mediated iron uptake in *Saccharomyces cerevisiae*: the SIT1 gene encodes a ferrioxamine B permease that belongs to the major facilitator superfamily. Microbiology **144:** 3455-3462

Liaw SH, Kuo I, Eisenberg D (1995) Discovery of the ammonium substrate site on glutamine synthetase, a third cation binding site. Protein Sci **4:** 2358-2365

Lim S-L, Chu W-L, Phang S-M (2010) Use of *Chlorella vulgaris* for bioremediation of textile wastewater. Bioresource Technology **101:** 7314-7322

Lin S, Cronan JE (2011) Closing in on complete pathways of biotin biosynthesis. Mol Biosyst **7:** 1811-1821

Lindberg P, Park S, Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. Metab Eng **12:** 70-79

Liu D, Golden JW (2002) *hetL* overexpression stimulates heterocyst formation in *Anabaena* sp. strain PCC 7120. J Bacteriol **184:** 6873-6881

Llamas MA, Sparrius M, Kloet R, Jiménez CR, Vandenbroucke-Grauls C, Bitter W (2006) The heterologous siderophores ferrioxamine b and ferrichrome activate signaling pathways in *Pseudomonas aeruginosa*. Journal of Bacteriology **188:** 1882-1891

Loew LM, Schaff JC (2001) The Virtual Cell: a software environment for computational cell biology. Trends Biotechnol **19:** 401-406

Long SR (1989) Rhizobium-legume nodulation: Life together in the underground. Cell **56:** 203-214

López-Gomollón S, Hernández JA, Pellicer S, Angarica VE, Peleato ML, Fillat MF (2007a) Cross-talk between iron and nitrogen regulatory networks in *Anabaena* (*Nostoc*) sp. PCC 7120: identification of overlapping genes in FurA and NtcA regulons. J Mol Biol **374:** 267-281

López-Gomollón S, Hernández JA, Wolk CP, Peleato ML, Fillat MF (2007b) Expression of *furA* is modulated by NtcA and strongly enhanced in heterocysts of *Anabaena* sp. PCC 7120. Microbiology **153:** 42-50

Louden BC, Haarmann D, Lynne AM (2011) Use of blue agar CAS assay for siderophore detection. J Microbiol Biol Educ **12:** 51-53

Luinenburg I, Coleman JR (1992) Identification, characterization and sequence analysis of the gene encoding phosphoenolpyruvate carboxylase in *Anabaena* sp. PCC 7120. J Gen Microbiol **138:** 685-691

Luinenburg I, Coleman JR (1993) Expression of *Escherichia coli* phosphoenolpyruvate carboxylase in a cyanobacterium. Functional complementation of *Synechococcus* PCC 7942 *ppc*. Plant Physiol **101:** 121-126

Luo H, Lin Y, Gao F, Zhang C-T, Zhang R (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Research **42:** D574-D580

Luque I, Flores E, Herrero A (1994) Molecular mechanism for the operation of nitrogen control in cyanobacteria. The EMBO Journal **13:** 2862-2869

Luque I, Forchhammer K (2008) Nitrogen assimilation and C/N balance sensing. *In* A Herrero, E Flores, eds, The Cyanobacteria: Molecular Biology, Genomics and Evolution. Caister Academic Press, pp 336-337

Lutgens FK, Tarbuck EJ (2003) Essentials of geology. *In*, Ed 8th. Prentice Hall

Lynch D, O'Brien J, Welch T, Clarke P, Cuiv PO, Crosa JH, O'Connell M (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. J Bacteriol **183:** 2576-2585

Mackenzie FT (1998) Our changing planet: An introduction to Earth system science and global environmental change. Prentice Hall

MacLean B, Tomazela DM, Abbatiello SE, Zhang S, Whiteaker JR, Paulovich AG, Carr SA, MacCoss MJ (2010) Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. Analytical Chemistry **82:** 10116-10124

Madan AP, Nierzwicki-Bauer SA (1993) In situ detection of transcripts for ribulose-1,5-bisphosphate carboxylase in cyanobacterial heterocysts. J Bacteriol **175:** 7301-7306

Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database **2011**

Malatinszky D, Jones PR (unpublished) Systematic study of biosynthetic genes in the schizokinen operon of *Anabaena* sp. PCC 7120.

Malatinszky D, Steuer R, Jones PR (2017) A comprehensively curated genome-scale two-cell model for the heterocystous cyanobacterium *Anabaena* sp. PCC 7120. Plant Physiol **173:** 509-523

Marcozzi C, Cumino AC, Salerno GL (2009) Role of NtcA, a cyanobacterial global nitrogen regulator, in the regulation of sucrose metabolism gene expression in *Anabaena* sp. PCC 7120. Arch Microbiol **191:** 255-263

Mariscal V, Herrero A, Flores E (2007) Continuous periplasm in a filamentous, heterocyst-forming cyanobacterium. Molecular Microbiology **65:** 1139-1145

Marolewski AE, Mattia KM, Warren MS, Benkovic SJ (1997) Formyl phosphate: a proposed intermediate in the reaction catalyzed by *Escherichia coli* PurT GAR transformylase. Biochemistry **36:** 6709-6716

Martin-Figueroa E, Navarro F, Florencio FJ (2000) The GS-GOGAT pathway is not operative in the heterocysts. Cloning and expression of *glsF* gene from the cyanobacterium *Anabaena* sp. PCC 7120. FEBS Lett **476:** 282-286

Marx CJ (2008) Development of a broad-host-range *sacB*-based vector for unmarked allelic exchange. BMC Research Notes **1:** 1

MATLAB (2011) MATLAB and Statistics Toolbox Release R2011b. *In*. The MathWorks Inc., Natick, Massachusetts

Matsuda N, Kobayashi H, Katoh H, Ogawa T, Futatsugi L, Nakamura T, Bakker EP, Uozumi N (2004) Na^+-dependent K^+ uptake Ktr system from the cyanobacterium *Synechocystis* sp. PCC 6803 and its role in the early phases of cell adaptation to hyperosmotic shock. Journal of Biological Chemistry **279:** 54952-54962

Matveyev AV, Rutgers E, Söderbäck E, Bergman B (1994) A novel genome rearrangement involved in heterocyst differentiation of the cyanobacterium *Anabaena* sp. PCC 7120. FEMS Microbiology Letters **116:** 201-207

McCann KS (2000) The diversity-stability debate. Nature **405:** 228-233

McKenzie VJ, Townsend AR (2007) Parasitic and infectious disease responses to changing global nutrient cycles. EcoHealth **4:** 384-396

McNaughton SJ (1977) Diversity and stability of ecological communities: A comment on the role of empiricism in ecology. The American Naturalist **111:** 515-525

Meeks JC, Elhai J (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. Microbiology and Molecular Biology Reviews **66:** 94-121

Meeks JC, Elhai J, Thiel T, Potts M, Larimer F, Lamerdin J, Predki P, Atlas R (2001) An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. Photosynth Res **70:** 85-106

Mehler AH (1951) Studies on reactions of illuminated chloroplasts: I. Mechanism of the reduction of oxygen and other Hill reagents. Archives of Biochemistry and Biophysics **33:** 65-77

Mérida A, Candau P, Florencio FJ (1991) Regulation of glutamine synthetase activity in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803 by the nitrogen source: effect of ammonium. Journal of Bacteriology **173:** 4095-4100

Mérida A, Flores E, Florencio FJ (1992) Regulation of *Anabaena* sp. strain PCC 7120 glutamine synthetase activity in a *Synechocystis* sp. strain PCC 6803 derivative strain bearing the *Anabaena glnA* gene and a mutated host *glnA* gene. Journal of Bacteriology **174:** 650-654

Merino-Puerto V, Mariscal V, Mullineaux CW, Herrero A, Flores E (2010) Fra proteins influencing filament integrity, diazotrophy and localization of septal protein SepJ in the heterocyst-forming cyanobacterium *Anabaena* sp. Molecular Microbiology **75:** 1159-1170

Merino-Puerto V, Schwarz H, Maldener I, Mariscal V, Mullineaux CW, Herrero A, Flores E (2011) FraC/FraD-dependent intercellular molecular exchange in the filaments of a heterocyst-forming cyanobacterium, *Anabaena* sp. Molecular Microbiology **82:** 87-98

Mi H, Endo T, Schreiber U, Ogawa T, Asada K (1992) Electron donation from cyclic and respiratory flows to the photosynthetic intersystem chain is mediated by pyridine nucleotide dehydrogenase in the cyanobacterium *Synechocystis* PCC 6803. Plant and Cell Physiology **33:** 1233-1237

Michaels AF, Olson D, Sarmiento JL, Ammerman JW, Fanning K, Jahnke R, Knap AH, Lipschultz F, Prospero JM (1996) Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. Biogeochemistry **35:** 181-226

Miller RW, Eady RR (1988) Molybdenum and vanadium nitrogenases of *Azotobacter chroococcum.* Low temperature favours N_2 reduction by vanadium nitrogenase. Biochemical Journal **256:** 429-432

Milligan AJ, Berman-Frank I, Gerchman Y, Dismukes GC, Falkowski PG (2007) Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection. Journal of Phycology **43:** 845-852

Mills MM, Ridame C, Davey M, La Roche J, Geider RJ (2004) Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. Nature **429:** 292-294

Mitra A, Kesarwani AK, Pal D, Nagaraja V (2011) WebGeSTer DB—a transcription terminator database. Nucleic Acids Research **39:** D129-D135

Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, Voß B, Steglich C, Wilde A, Vogel J, Hess WR (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC 6803. Proceedings of the National Academy of Sciences **108:** 2124-2129

Mitschke J, Vioque A, Haas F, Hess WR, Muro-Pastor AM (2011) Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. Proceedings of the National Academy of Sciences **108:** 20130-20135

Mochimaru M, Masukawa H, Maoka T, Mohamed HE, Vermaas WFJ, Takaichi S (2008) Substrate specificities and availability of fucosyltransferase and β-carotene hydroxylase for myxol 2'-fucoside synthesis in *Anabaena* sp. strain PCC 7120 compared with *Synechocystis* sp. strain PCC 6803. Journal of Bacteriology **190:** 6726-6733

Montesinos ML, Herrero A, Flores E (1995) Amino acid transport systems required for diazotrophic growth in the cyanobacterium *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **177:** 3150-3157

Mosier A, Kroeze C, Nevison C, Oenema O, Seitzinger S, van Cleemput O (1998) Closing the global N_2O budget: nitrous oxide emissions through the agricultural nitrogen cycle. Nutrient Cycling in Agroecosystems **52:** 225-248

Mulder A, van de Graaf AA, Robertson LA, Kuenen JG (1995) Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. FEMS Microbiology Ecology **16:** 177-183

Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, Dufresne A, Partensky F, Burd H, Kaznadzey D, Haselkorn R, Galperin MY (2006) The cyanobacterial genome core and the origin of photosynthesis. Proceedings of the National Academy of Sciences **103:** 13126-13131

Mullineaux CW, Mariscal V, Nenninger A, Khanum H, Herrero A, Flores E, Adams DG (2008) Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. The EMBO Journal **27:** 1299-1308

Muro-Pastor AM, Hess WR (2012) Heterocyst differentiation: from single mutants to global approaches. Trends Microbiol **20:** 548-557

Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. Photosynth Res **83:** 135-150

Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu MH, Oldroyd GE, Poole PS, Udvardi MK, Voigt CA, Ane JM, Peters JW (2016) Symbiotic nitrogen fixation and challenges to extending it to non-legumes. Appl Environ Microbiol

Nakao M, Okamoto S, Kohara M, Fujishiro T, Fujisawa T, Sato S, Tabata S, Kaneko T, Nakamura Y (2010) CyanoBase: the cyanobacteria genome database update 2010. Nucleic Acids Research **38:** D379-D381

Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol **8:** 11-13

Neilands JB (1995) Siderophores: Structure and function of microbial iron transport compounds. Journal of Biological Chemistry **270:** 26723-26726

Neilson AH, Larsson T (1980) The utilization of organic nitrogen for growth of algae: physiological aspects. Physiologia Plantarum **48:** 542-553

Nicolaisen K, Hahn A, Schleiff E (2009) The cell wall in heterocyst formation by *Anabaena* sp. PCC 7120. J Basic Microbiol **49:** 5-24

Nicolaisen K, Hahn A, Valdebenito M, Moslavac S, Samborski A, Maldener I, Wilken C, Valladares A, Flores E, Hantke K, Schleiff E (2010) The interplay between siderophore secretion and coupled iron and copper transport in the heterocyst-forming cyanobacterium *Anabaena* sp. PCC 7120. Biochimica et Biophysica Acta (BBA) - Biomembranes **1798:** 2131-2140

Nicolaisen K, Moslavac S, Samborski A, Valdebenito M, Hantke K, Maldener I, Muro-Pastor AM, Flores E, Schleiff E (2008) Alr0397 is an outer membrane transporter for the siderophore schizokinen in *Anabaena* sp. strain PCC 7120. J Bacteriol **190:** 7500-7507

Nicolaisen K, Schleiff E (2010) Iron dependency of and transport by cyanobacteria. *In* S Andrews, P Cornelis, eds, Iron Uptake in Microorganisms. Caister Academic Press, Norfolk, UK, pp 203-229

Nielsen A (1995) Ammonia synthesis. Springer, New York

Nodop A, Pietsch D, Höcker R, Becker A, Pistorius EK, Forchhammer K, Michel K-P (2008) Transcript profiling reveals new insights into the acclimation of the mesophilic fresh-water cyanobacterium *Synechococcus elongatus* PCC 7942 to iron starvation. Plant Physiology **147:** 747-763

Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proceedings of the National Academy of Sciences

Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput Biol **10:** e1003483

Nürnberg DJ, Mariscal V, Bornikoel J, Nieves-Morión M, Krauß N, Herrero A, Maldener I, Flores E, Mullineaux CW (2015) Intercellular diffusion of a fluorescent sucrose analog via the septal junctions in a filamentous cyanobacterium. mBio **6**

O'Malley MA, Soyer OS (2012) The roles of integration in molecular systems biology. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences **43:** 58-68

Ober JA (2017) Mineral commodity summaries 2017. *In* Mineral Commodity Summaries, Reston, VA, p 202

Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Molecular Systems Biology **5:** 320-320

Ogawa T, Mi H (2007) Cyanobacterial NADPH dehydrogenase complexes. Photosynthesis Research **93:** 69-77

Olmedo-Verd E, Muro-Pastor AM, Flores E, Herrero A (2006) Localized induction of the *ntcA* regulatory gene in developing heterocysts of *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **188:** 6694-6699

Orjala J, Gerwick WH (1996) Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula*. Journal of Natural Products **59:** 427-430

Orr J, Keefer LM, Keim P, Nguyen TD, Wellems T, Heinrikson RL, Haselkorn R (1981) Purification, physical characterization, and NH_2-terminal sequence of glutamine synthetase from the cyanobacterium *Anabaena* 7120. Journal of Biological Chemistry **256:** 13091-13098

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular Systems Biology **7**

Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotech **28:** 245-248

Ortiz-Marquez JC, Do Nascimento M, Curatti L (2014) Metabolic engineering of ammonium release for nitrogen-fixing multispecies microbial cell-factories. Metab Eng **23:** 154-164

Ortiz-Marquez JC, Do Nascimento M, Dublan MeL, Curatti L (2012) Association with an ammonium-excreting bacterium allows diazotrophic culture of oil-rich eukaryotic microalgae. Appl Environ Microbiol **78:** 2345-2352

Ortiz-Marquez JC, Do Nascimento M, Zehr JP, Curatti L (2013) Genetic engineering of multispecies microbial cell factories as an alternative for bioenergy production. Trends Biotechnol **31:** 521-529

Osborne NJT, Webb PM, Shaw GR (2001) The toxins of *Lyngbya majuscula* and their human and ecological health effects. Environment International **27:** 381-392

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res **33:** 5691-5702

Owttrim GW, Colman B (1988) Phosphoenolpyruvate carboxylase mediated carbon flow in a cyanobacterium. Biochemistry and Cell Biology **66:** 93-99

Palanche T, Blanc S, Hennard C, Abdallah MA, Albrecht-Gary AM (2004) Bacterial iron transport: coordination properties of azotobactin, the highly fluorescent siderophore of Azotobacter vinelandii. Inorg Chem **43:** 1137-1152

Palsson BO (2006) Systems biology: Properties of reconstructed networks. Cambridge University Press

Park J-J, Lechno-Yossef S, Wolk C, Vieille C (2013) Cell-specific gene expression in *Anabaena variabilis* grown phototrophically, mixotrophically, and heterotrophically. BMC Genomics **14:** 759

Parton WJ, Mosier AR, Ojima DS, Valentine DW, Schimel DS, Weier K, Kulmala AE (1996) Generalized model for N_2 and N_2O production from nitrification and denitrification. Global Biogeochemical Cycles **10:** 401-412

Pásztor A, Kallio P, Malatinszky D, Akhtar MK, Jones PR (2015) A synthetic O2-tolerant butanol pathway exploiting native fatty acid biosynthesis in Escherichia coli. Biotechnology and Bioengineering **112:** 120-128

Patil KR, Rocha I, Forster J, Nielsen J (2005) Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics **6:** 308

Paz-Yepes J, Merino-Puerto V, Herrero A, Flores E (2008) The amt gene cluster of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. J Bacteriol **190:** 6534-6539

Pecoraro V, Zerulla K, Lange C, Soppa J (2011) Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. PLOS ONE **6:** e16392

Pelicic V, Reyrat J-M, Gicquel B (1996) Positive selection of allelic exchange mutants in *Mycobacterium bovis* BCG. FEMS Microbiology Letters **144:** 161-166

Pelz O, Tesar M, Wittich RM, Moore ER, Timmis KN, Abraham WR (1999) Towards elucidation of microbial community metabolic pathways: unravelling the network of carbon sharing in a pollutant-degrading bacterial consortium by immunocapture and isotopic ratio mass spectrometry. Environ Microbiol **1:** 167-174

Peoples MB, Brockwell J, Herridge DF, Rochester IJ, Alves BJR, Urquiaga S, Boddey RM, Dakora FD, Bhattarai S, Maskey SL, Sampet C, Rerkasem B, Khan DF, Hauggaard-Nielsen H, Jensen ES (2009) The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. Symbiosis **48:** 1-17

Pernil R, Herrero A, Flores E (2010) Catabolic function of compartmentalized alanine dehydrogenase in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **192:** 5165-5172

Pernil R, Picossi S, Mariscal V, Herrero A, Flores E (2008) ABC-type amino acid uptake transporters Bgt and N-II of *Anabaena* sp. strain PCC 7120 share an ATPase subunit and are expressed in vegetative cells and heterocysts. Molecular Microbiology **67:** 1067-1080

Peschek GA (1999) Photosynthesis and respiration of cyanobacteria. *In* GA Peschek, W Löffelhardt, G Schmetterer, eds, The Phototrophic Prokaryotes. Springer US, Boston, MA, pp 201-209

Peschek GA, Villgrater K, Wastyn M (1991) 'Respiratory protection' of the nitrogenase in dinitrogen-fixing cyanobacteria. Plant and Soil **137:** 17-24

Peters JW, Fisher K, Dean DR (1995) Nitrogenase structure and function: A biochemical-genetic perspective. Annual Review of Microbiology **49:** 335-366

Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O (2001) The Comprehensive Microbial Resource. Nucleic Acids Res **29:** 123-125

Pettit RK (2009) Mixed fermentation for natural product drug discovery. Appl Microbiol Biotechnol **83:** 19-25

Pfreundt U, Stal LJ, Vos B, Hess WR (2012) Dinitrogen fixation in a unicellular chlorophyll d-containing cyanobacterium. ISME J **6:** 1367-1377

Philippot L (2002) Denitrifying genes in bacterial and archaeal genomes. Biochim Biophys Acta **1577:** 355-376

Photon-Systems-Instruments (2017) Multi-Cultivator MC 1000-OD instruction manual and user's guide for cultivation. *In*, https://www.psi.cz/download/document/manuals/multi-cultivator/Multi-Cultivator_Manual.pdf

Picossi S, Montesinos ML, Pernil R, Lichtle C, Herrero A, Flores E (2005) ABC-type neutral amino acid permease N-I is required for optimal diazotrophic growth and is repressed in the heterocysts of *Anabaena* sp. strain PCC 7120. Mol Microbiol **57:** 1582-1592

Poole RK, Hill S (1997) Respiratory protection of nitrogenase activity in A*zotobacter vinelandii* — roles of the terminal oxidases. Bioscience Reports **17:** 303-317

Popa R, Weber PK, Pett-Ridge J, Finzi JA, Fallon SJ, Hutcheon ID, Nealson KH, Capone DG (2007) Carbon and nitrogen fixation and metabolite exchange in and between individual cells of *Anabaena oscillarioides*. ISME J **1:** 354-360

Porankiewicz J, Schelin J, Clarke AK (1998) The ATP-dependent Clp protease is essential for acclimation to UV-B and low temperature in the cyanobacterium *Synechococcus*. Molecular Microbiology **29:** 275-283

Portmann C, Blom JF, Gademann K, Jüttner F (2008) Aerucyclamides A and B: Isolation and synthesis of toxic ribosomal heterocyclic peptides from the cyanobacterium *Microcystis aeruginosa* PCC 7806. Journal of Natural Products **71:** 1193-1196

Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol **21:** 162-169

Privalle LS, Burris RH (1984) D-erythrose supports nitrogenase activity in isolated *Anabaena* sp. strain 7120 heterocysts. J Bacteriol **157:** 350-356

Prosser JI (1989) Autotrophic nitrification in bacteria. Adv Microb Physiol **30:** 125-181

Quintero MJ, Montesinos ML, Herrero A, Flores E (2001) Identification of genes encoding amino acid permeases by inactivation of selected ORFs from the *Synechocystis* genomic sequence. Genome Res **11:** 2034-2040

Raes J, Bork P (2008) Molecular eco-systems biology: towards an understanding of community function. Nat Rev Micro **6:** 693-699

Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Systems Biology **3:** 38

Rai AN, Bergman B, Rasmussen U (2002) Cyanobacteria in symbiosis. Springer, Netherlands

Ran L, Larsson J, Vigil-Stenman T, Nylander JAA, Ininbergs K, Zheng W-W, Lapidus A, Lowry S, Haselkorn R, Bergman B (2010) Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium. PLOS ONE **5:** e11486

Rathore DS, Kumar A, Kumar HD (1993) Lipid content and fatty acid composition in N_2-fixing cyanobacterium *Anabaena doliolum* as affected by molybdenum deficiency. World J Microbiol Biotechnol **9:** 508-510

Raun WR, Solie JB, Johnson GV, Stone ML, Mullen RW, Freeman KW, Thomason WE, Lukina EV (2002) Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application. Agronomy Journal **94:** 815-820

Ravishankara AR, Daniel JS, Portmann RW (2009) Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. Science **326:** 123-125

Razquin P, Fillat MF, Schmitz S, Stricker O, Bohme H, Gomez-Moreno C, Peleato ML (1996) Expression of ferredoxin-$NADP^+$ reductase in heterocysts from *Anabaena* sp. Biochem J **316 (Pt 1):** 157-160

Resendis-Antonio O, Reed JL, Encarnación S, Collado-Vides J, Palsson BØ (2007) Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. PLoS Comput Biol **3:** e192

Reyes JC, Florencio FJ (1994) A new type of glutamine synthetase in cyanobacteria: the protein encoded by the *glnN* gene supports nitrogen assimilation in *Synechocystis* sp. strain PCC 6803. Journal of Bacteriology **176:** 1260-1267

Reysenbach AL, Wickham GS, Pace NR (1994) Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Applied and Environmental Microbiology **60:** 2113-2119

Rice D, Mazur BJ, Haselkorn R (1982) Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* 7120. J Biol Chem **257:** 13157-13163

Rikkinen J, Oksanen I, Lohtander K (2002) Lichen guilds share related cyanobacterial symbionts. Science **297:** 357

Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Journal of General Microbiology **111:** 1-61

Ris H, Singh RN (1961) Electron microscope studies on blue-green algae. J Biophys Biochem Cytol **9:** 63-80

Riseley AS (2017) Exploring the potential of algae-bacteria communities for biotechnology. University of Cambridge, Cambridge, UK

Ronzio RA, Rowe WB, Meister A (1969) Mechanism of inhibition of glutamine synthetase by methionine sulfoximine. Biochemistry **8:** 1066-1075

Rothstein DM, Pahel G, Tyler B, Magasanik B (1980) Regulation of expression from the *glnA* promoter of *Escherichia coli* in the absence of glutamine synthetase. Proc Natl Acad Sci U S A **77:** 7372-7376

Rudolf M, Kranzler C, Lis H, Margulis K, Stevanovic M, Keren N, Schleiff E (2015) Multiple modes of iron uptake by the filamentous, siderophore-producing cyanobacterium, *Anabaena* sp. PCC 7120. Mol Microbiol **97:** 577-588

Rudolf M, Stevanovic M, Kranzler C, Pernil R, Keren N, Schleiff E (2016) Multiplicity and specificity of siderophore uptake in the cyanobacterium *Anabaena* sp. PCC 7120. Plant Mol Biol **92:** 57-69

Rueter JG (1988) Iron stimulation of photosynthesis and nitrogen fixation in *Anabaena* 7120 and *Trichodesmium* (*Cyanophyceae*). Journal of Phycology **24:** 249-254

Ruiz M, Bettache A, Janicki A, Vinella D, Zhang C-C, Latifi A (2010) The *alr2505* (*osiS*) gene from *Anabaena* sp. strain PCC 7120 encodes a cysteine desulfurase induced by oxidative stress. FEBS Journal **277:** 3715-3725

Sadre R, Pfaff C, Buchkremer S (2012) Plastoquinone-9 biosynthesis in cyanobacteria differs from that in plants and involves a novel 4-hydroxybenzoate solanesyltransferase. Biochem J **442:** 621-629

Saha R, Verseput AT, Berla BM, Mueller TJ, Pakrasi HB, Maranas CD (2012) Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. PLoS One **7:** e48285

Sallal AK, Nimer NA, Radwan SS (1990) Lipid and fatty acid composition of freshwater cyanobacteria. Microbiology **136:** 2043-2048

Savinell JM, Palsson BO (1992) Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. J Theor Biol **154:** 421-454

Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, Pühler A (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene **145:** 69-73

Schellenberger J, Park JO, Conrad TM, Palsson B (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics **11:** 213

Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson B (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc **6:** 1290-1307

Schilling N, Ehrnsperger K (1985) Cellular differentiation of sucrose metabolism in *Anabaena variabilis*. Z. Naturforsch. **40:** 776-779

Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. Science **329:** 559-562

Schlesinger WH, Bernhardt ES (2013) Chapter 2 - Origins. *In* Biogeochemistry (Third Edition). Academic Press, Boston, pp 15-48

Schlögl R (2008) Handbook of heterogeneous catalysis, Ed 2nd. Wiley-VCH Verlag, GmbH & Co. KGaA, Weinheim

Schluchter WM, Glazer AN (1997) Characterization of cyanobacterial biliverdin reductase: Conversion of biliverdin to bilirubin is important for normal phycobiliprotein biosynthesis. Journal of Biological Chemistry **272:** 13562-13569

Schmid MC, Maas B, Dapena A, van de Pas-Schoonen K, van de Vossenberg J, Kartal B, van Niftrik L, Schmidt I, Cirpus I, Kuenen JG, Wagner M, Sinninghe Damsté JS, Kuypers M, Revsbech NP, Mendez R, Jetten MSM, Strous M (2005) Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. Applied and Environmental Microbiology **71:** 1677-1684

Schneegurt MA, Tucker DL, Ondr JK, Sherman DM, Sherman LA (2000) Metabolic rhythms of a diazotrophic cyanobacterium, *Cyanothece* sp. strain ATCC 51142, heterotrophically grown in continuous dark. Journal of Phycology **36:** 107-117

Schriek S, Ruckert C, Staiger D, Pistorius E, Michel K-P (2007) Bioinformatic evaluation of L-arginine catabolic pathways in 24 cyanobacteria and transcriptional analysis of genes encoding enzymes of L-arginine catabolism in the cyanobacterium *Synechocystis* sp. PCC 6803. BMC Genomics **8:** 437

Scragg A, Morrison J, Shales S (2003) The use of a fuel containing *Chlorella vulgaris* in a diesel engine. Enzyme and Microbial Technology **33:** 884-889

Seitzinger SP, Kroeze C, Bouwman AF, Caraco N, Dentener F, Styles RV (2002) Global patterns of dissolved inorganic and particulate nitrogen inputs to coastal systems: Recent conditions and future projections. Estuaries **25:** 640-655

Serrano A (1992) Purification, characterization and function of dihydrolipoamide dehydrogenase from the cyanobacterium *Anabaena* sp. strain PCC 7119. Biochemical Journal **288:** 823-830

Sessions AL, Doughty DM, Welander PV, Summons RE, Newman DK (2009) The continuing puzzle of the Great Oxidation Event. Current Biology **19:** R567-R574

Shanmugam KT, O'Gara F, Andersen K, Valentine RC (1978) Biological nitrogen fixation. Annual Review of Plant Physiology **29:** 263-276

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research **13:** 2498-2504

Shao J, Xu Y, Wang Z, Jiang Y, Yu G, Peng X, Li R (2011) Elucidating the toxicity targets of β-ionone on photosynthetic system of *Microcystis aeruginosa* NIES-843 (Cyanobacteria). Aquatic Toxicology **104:** 48-55

Shcolnick S, Keren N (2006) Metal homeostasis in cyanobacteria and chloroplasts. Balancing benefits and risks to the photosynthetic apparatus. Plant Physiology **141:** 805-810

Shestakov SV, Khyen NT (1970) Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. Mol Gen Genet **107:** 372-375

Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T (2002) Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: Function and phyolgenetic analysis. Journal of Biological Chemistry **277:** 18658-18664

Shields-Zhou G, Och L (2011) The case for a Neoproterozoic Oxygenation Event: Geochemical evidence and biological consequences. GSa Today **21:** 4-11

Shoun H, Kim DH, Uchiyama H, Sugiyama J (1992) Denitrification by fungi. FEMS Microbiol Lett **73:** 277-281

Simpson FB, Neilands JB (1976) Siderochromes in *Cyanophyceae*: isolation and characterization of schizokinen from *Anabaena* sp. Journal of Phycology **12:** 44-48

Singh AP, Tiwari DN (1998) Phenotypic expression of ammonia-excreting mutants of *Anabaena* 7120 under nitrogen limitation. World Journal of Microbiology and Biotechnology **14:** 591-593

Singh RN (1950) Reclamation of "usar" lands in India through blue-green algae. Nature **165:** 325-326

Singh SP, Montgomery BL (2011) Determining cell shape: adaptive regulation of cyanobacterial cellular differentiation and morphology. Trends Microbiol **19:** 278-285

Smallbone K, Simeonidis E (2009) Flux balance analysis: A geometric perspective. Journal of Theoretical Biology **258:** 311-315

Smil V (2002) Nitrogen and food production: proteins for human diets. Ambio **31:** 126-131

Smil V (2004) Enriching the Earth: Fritz Haber, Carl Bosch, and the transformation of world food production. MIT Press

Srivastava R, Amla DV (2002) Molecular characteristics of *glnA* linked mutations in the nitrogen-fixing cyanobacterium *Nostoc muscorum*. Current Microbiology **44:** 94-101

Steinhauser D, Fernie AR, Araujo WL (2012) Unusual cyanobacterial TCA cycles: not broken just different. Trends Plant Sci **17:** 503-509

Sterner RW, Elser JJ (2002) Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere. Princeton University Press, Princeton

Steuer R, Knoop H, Machne R (2012) Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth. J Exp Bot **63:** 2259-2274

Stevanovic M, Hahn A, Nicolaisen K, Mirus O, Schleiff E (2012) The components of the putative iron transport system in the cyanobacterium *Anabaena* sp. PCC 7120. Environ Microbiol **14:** 1655-1670

Stevens SE, Porter RD (1980) Transformation in *Agmenellum quadruplicatum*. Proc Natl Acad Sci U S A **77:** 6052-6056

Stewart WD, Rowell P, Rai AN (1983) Cyanobacteria-eukaryotic plant symbioses. Ann Microbiol (Paris) **134b:** 205-228

Stewart WM, Dibb DW, Johnston AE, Smyth TJ (2005) The contribution of commercial fertilizer nutrients to food production. Agronomy Journal **97:** 1-6

Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, van de Pas-Schoonen KT, Webb R, Kuenen JG, Jetten MSM (1999) Missing lithotroph identified as new planctomycete. Nature **400:** 446-449

Stucken K, John U, Cembella A, Murillo AA, Soto-Liebe K, Fuentes-Valdes JJ, Friedel M, Plominsky AM, Vasquez M, Glockner G (2010) The smallest known genomes of multicellular and toxic cyanobacteria: comparison, minimal gene sets for linked traits and the evolutionary implications. PLoS One **5:** e9235

Subbarao GV, Kishii M, Nakahara K, Ishikawa T, Ban T, Tsujimoto H, George TS, Berry WL, Hash CT, Ito O (2009) Biological nitrification inhibition (BNI) - Is there potential for genetic interventions in the *Triticeae*? Breeding Science **59:** 529-545

Suthers PF, Zomorrodi A, Maranas CD (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. Molecular Systems Biology **5**

Sutton MA, Howard CM, Erisman JW, Billen G, Bleeker A, Grennfelt P, van Grinsven H, Grizzetti B (2011a) The European Nitrogen Assessment. *In* MA Sutton, CM Howard, JW Erisman, G Billen, A Bleeker, P Grennfelt, H van Grinsven, B Grizzetti, eds. Cambridge University Press

Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W (2011b) Too much of a good thing. Nature **472:** 159-161

Takahama K, Matsuoka M, Nagahama K, Ogawa T (2003) Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbAI* locus. J Biosci Bioeng **95:** 302-305

Takahama K, Matsuoka M, Nagahama K, Ogawa T (2004) High-frequency gene replacement in cyanobacteria using a heterologous *rps12* gene. Plant Cell Physiol **45:** 333-339

Takaichi S, Mochimaru M (2007) Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. Cell Mol Life Sci **64:** 2607-2619

Takaichi S, Mochimaru M, Maoka T, Katoh H (2005) Myxol and 4-ketomyxol 2'-fucosides, not rhamnosides, from *Anabaena* sp. PCC 7120 and *Nostoc punctiforme* PCC 73102, and proposal for the biosynthetic pathway of carotenoids. Plant Cell Physiol **46:** 497-504

Takasaki K, Shoun H, Yamaguchi M, Takeo K, Nakamura A, Hoshino T, Takaya N (2004) Fungal ammonia fermentation, a novel metabolic mechanism that couples the dissimilatory and assimilatory pathways of both nitrate and ethanol. Role of acetyl CoA synthetase in anaerobic ATP synthesis. J Biol Chem **279:** 12414-12420

Tamaru K (1991) Catalytic ammonia synthesis: Fundamentals and practice. Plenum Press, New York

Tan X, Liang F, Cai K, Lu X (2013) Application of the FLP/FRT recombination system in cyanobacteria for construction of markerless mutants. Appl Microbiol Biotechnol **97:** 6373-6382

Tanabe T, Nakao H, Kuroda T, Tsuchiya T, Yamamoto S (2006) Involvement of the *Vibrio parahaemolyticus pvsC* gene in export of the siderophore vibrioferrin. Microbiol Immunol **50:** 871-876

Tanimoto T, Hatano K-i, Kim D-h, Uchiyama H, Shoun H (1992) Co-denitrification by the denitrifying system of the fungus *Fusarium oxysporum*. FEMS Microbiology Letters **93:** 177-180

Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I (2014) RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res **42:** D553-559

Tervo CJ, Reed JL (2014) Expanding metabolic engineering algorithms using feasible space and shadow price constraint modules. Metabolic Engineering Communications **1:** 1-11

Teske A, Alm E, Regan JM, Toze S, Rittmann BE, Stahl DA (1994) Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. J Bacteriol **176:** 6623-6630

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science **327:** 818-822

The Non-GMO Project (2018). *In* www.nongmoproject.org, accessed: 14-04-2018

Thiel T (1993) Characterization of genes for an alternative nitrogenase in the cyanobacterium *Anabaena variabilis*. Journal of Bacteriology **175:** 6276-6286

Thiel T, Lyons EM, Erker JC, Ernst A (1995) A second nitrogenase in vegetative cells of a heterocyst-forming cyanobacterium. Proceedings of the National Academy of Sciences **92:** 9358-9362

Thiel T, Pratte B (2001) Effect on heterocyst differentiation of nitrogen fixation in vegetative cells of the cyanobacterium *Anabaena variabilis* ATCC 29413. J Bacteriol **183:** 280-286

Thiele I, Palsson B (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc **5:** 93-121

Thomas J, Meeks JC, Wolk CP, Shaffer PW, Austin SM (1977) Formation of glutamine from [^{13}N]ammonia, [^{13}N]dinitrogen, and [^{14}C]glutamate by heterocysts isolated from *Anabaena cylindrica*. J Bacteriol **129:** 1545-1555

Thomas SP, Zaritsky A, Boussiba S (1990) Ammonium excretion by an L-methionine-DL-sulfoximine-resistant mutant of the rice field cyanobacterium *Anabaena siamensis*. Appl Environ Microbiol **56:** 3499-3504

Tidgewell K, Clark BR, Gerwick WH (2010) The natural products chemistry of cyanobacteria. *In* Comprehensive Natural Products II. Elsevier, Oxford, pp 141-188

Townsend AR, Howarth RW, Bazzaz FA, Booth MS, Cleveland CC, Collinge SK, Dobson AP, Epstein PR, Holland EA, Keeney DR, Mallin MA, Rogers CA, Wayne P, Wolfe AH (2003) Human health effects of a changing global nitrogen cycle. Frontiers in Ecology and the Environment **1:** 240-246

Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, Affourtit JP, Zehr JP (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature **464:** 90-94

Tuit C, Waterbury J, Ravizza G (2004) Diel variation of molybdenum and iron in marine diazotrophic cyanobacteria. Limnology and Oceanography **49:** 978-990

United Nations DoESA, Population Division (2017) Key findings and advance tables. *In* World Population Prospects: The 2017 Revision. United Nations

Valladares A, Herrero A, Pils D, Schmetterer G, Flores E (2003) Cytochrome c oxidase genes required for nitrogenase activity and diazotrophic growth in *Anabaena* sp. PCC 7120. Molecular Microbiology **47:** 1239-1249

Valladares A, Maldener I, Muro-Pastor AM, Flores E, Herrero A (2007) Heterocyst development and diazotrophic metabolism in terminal respiratory oxidase mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. Journal of Bacteriology **189:** 4425-4430

van Bodegom P (2007) Microbial maintenance: A critical review on its quantification. Microbial Ecology **53:** 513-523

Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol **60:** 3724-3731

Vázquez-Bermúdez MaF, Paz-Yepes J, Herrero A, Flores E (2002) The NtcA-activated *amt1* gene encodes a permease required for uptake of low concentrations of ammonium in the cyanobacterium *Synechococcus* sp. PCC 7942. Microbiology **148:** 861-869

Verstraete W, Focht DD (1977) Biochemical ecology of nitrification and denitrification. *In* M Alexander, ed, Advances in Microbial Ecology. Springer US, Boston, MA, pp 135-214

Villa JA, Ray EE, Barney BM (2014) *Azotobacter vinelandii* siderophore can provide nitrogen to support the culture of the green algae *Neochloris oleoabundans* and *Scenedesmus* sp. BA032. FEMS Microbiology Letters **351:** 70-77

Vitkin E, Shlomi T (2012) MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. Genome Biol **13:** R111

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of Earth's ecosystems. Science **277:** 494-499

Vojvodic A, Medford AJ, Studt F, Abild-Pedersen F, Khan TS, Bligaard T, Nørskov JK (2014) Exploring the limits: A low-pressure, low-temperature Haber–Bosch process. Chemical Physics Letters **598:** 108-112

Vraspir JM, Butler A (2009) Chemistry of marine ligands and siderophores. Ann Rev Mar Sci **1:** 43-63

Vu TT, Stolyar SM, Pinchuk GE, Hill EA, Kucek LA, Brown RN, Lipton MS, Osterman A, Fredrickson JK, Konopka AE, Beliaev AS, Reed JL (2012) Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. ATCC 51142. PLoS Comput Biol **8:** e1002460

Wagner G (1997) *Azolla*: A review of its biology and utilization. The Botanical Review **63:** 1-26

Walsby AE (2007) Cyanobacterial heterocysts: terminal pores proposed as sites of gas exchange. Trends Microbiol **15:** 340-349

Wang B, Wang J, Zhang W, Meldrum DR (2012) Application of synthetic biology in cyanobacteria and algae. Frontiers in Microbiology **3:** 344

Wang Y, Brown HN, Crowley DE, Szaniszlo PJ (1993) Evidence for direct utilization of a siderophore, ferrioxamine B, in axenically grown cucumber. Plant, Cell & Environment **16:** 579-585

Ward B (1985) Control of pH and inorganic carbon in batch cultures of cyanobacteria. Biotechnology Letters **7:** 87-92

Ward MH, deKok TM, Levallois P, Brender J, Gulis G, Nolan BT, VanDerslice J (2005) Workgroup report: Drinking-water nitrate and health—recent findings and research needs. Environmental Health Perspectives **113:** 1607-1614

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, Kettle H, Flint HJ, Haas AF, Laroche B, Kreft J-U, Rainey PB, Freilich S, Schuster S, Milferstedt K, van der Meer JR, Grokopf T, Huisman J, Free A, Picioreanu C, Quince C, Klapper I, Labarthe S, Smets BF, Wang H, Isaac Newton Institute F, Soyer OS (2016) Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J **10:** 2557-2568

Willis RB, Montgomery ME, Allen PR (1996) Improved method for manual, colorimetric determination of total Kjeldahl nitrogen using salicylate. Journal of Agricultural and Food Chemistry **44:** 1804-1807

Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol **6:** 407

Wolk CP, Ernst A, Elhai J (2004) Heterocyst metabolism and development. *In* D Bryant, ed, The Molecular Biology of Cyanobacteria, Vol 1. Springer Netherlands, pp 769-823

Wolk CP, Thomas J, Shaffer PW, Austin SM, Galonsky A (1976) Pathway of nitrogen metabolism after fixation of ^{13}N-labeled nitrogen gas by the cyanobacterium, *Anabaena cylindrica*. J Biol Chem **251:** 5027-5034

Wu B, Zhang B, Feng X, Rubens JR, Huang R, Hicks LM, Pakrasi HB, Tang YJ (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology **156:** 596-602

Wu F, Yang Z, Kuang T (2006) Impaired photosynthesis in phosphatidylglycerol-deficient mutant of cyanobacterium *Anabaena* sp. PCC 7120 with a disrupted gene encoding a putative phosphatidylglycerophosphatase. Plant Physiology **141:** 1274-1283

Wu SS, Kaiser D (1996) Markerless deletions of pil genes in *Myxococcus xanthus* generated by counterselection with the *Bacillus subtilis sacB* gene. Journal of Bacteriology **178:** 5817-5821

Xu Y, Alvey RM, Byrne PO, Graham JE, Shen G, Bryant DA (2011) Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in *Synechococcus* sp. PCC 7002. Methods Mol Biol **684:** 273-293

Yoon HS, Golden JW (2001) PatS and products of nitrogen fixation control heterocyst pattern. J Bacteriol **183:** 2605-2613

Zarecki R, Oberhardt MA, Yizhak K, Wagner A, Shtifman Segal E, Freilich S, Henry CS, Gophna U, Ruppin E (2014) Maximal sum of metabolic exchange fluxes outperforms biomass yield as a predictor of growth rate of microorganisms. PLoS ONE **9:** e98372

Zehr JP, Carpenter EJ, Villareal TA (2000) New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. Trends in Microbiology **8:** 68-73

Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM (2001) Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. Nature **412:** 635-638

Zerulla K, Ludt K, Soppa J (2016) The ploidy level of *Synechocystis* sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. Microbiology **162:** 730-739

Zhang L-C, Chen Y-F, Chen W-L, Zhang C-C (2008) Existence of periplasmic barriers preventing green fluorescent protein diffusion from cell to cell in the cyanobacterium *Anabaena* sp. strain PCC 7120. Molecular Microbiology **70:** 814-823

Zhang S, Bryant DA (2011) The tricarboxylic acid cycle in cyanobacteria. Science **334:** 1551-1553

Zhang Y, Pu H, Wang Q, Cheng S, Zhao W, Zhang Y, Zhao J (2007) PII is important in regulation of nitrogen metabolism but not required for heterocyst formation in the cyanobacterium *Anabaena* sp. PCC 7120. Journal of Biological Chemistry **282:** 33641-33648

Zhao M-X, Jiang Y-L, He Y-X, Chen Y-F, Teng Y-B, Chen Y, Zhang C-C, Zhou C-Z (2010) Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. Proceedings of the National Academy of Sciences **107:** 12487-12492

Zhao W, Guo Q, Zhao J (2007b) A membrane-associated Mn-superoxide dismutase protects the photosynthetic apparatus and nitrogenase from oxidative damage in the cyanobacterium *Anabaena* sp. PCC 7120. Plant and Cell Physiology **48:** 563-572

Zhao W, Ye Z, Zhao J (2007a) RbrA, a cyanobacterial rubrerythrin, functions as a FNR-dependent peroxidase in heterocysts in protection of nitrogenase from damage by hydrogen peroxide in *Anabaena* sp. PCC 7120. Molecular Microbiology **66:** 1219-1230

Zhou J, Rudd KE (2013) EcoGene 3.0. Nucleic Acids Res **41:** D613-624

Zhou Z, Takaya N, Nakamura A, Yamaguchi M, Takeo K, Shoun H (2002) Ammonia fermentation, a novel anoxic metabolism of nitrate by fungi. Journal of Biological Chemistry **277:** 1892-1896

Zhu J, Kong R, Wolk CP (1998) Regulation of *hepA* of *Anabaena* sp. strain PCC 7120 by elements 5′ from the gene and by *hepK*. Journal of Bacteriology **180:** 4233-4242

Zumft WG (1997) Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev **61:** 533-616

Appendices

Appendix A Orphan reactions removed from the *Anabaena* sp. PCC 7120 model

Table A-I. Orphan reactions not included in the *Anabaena* sp. PCC 7120 model.

Rxn ID	Definition
R00238	2 Acetyl-CoA <=> CoA + Acetoacetyl-CoA
R00408	Succinate + FAD <=> FADH2 + Fumarate
R00623	Primary alcohol + NAD+ <=> Aldehyde + NADH + H+
R00631	Aldehyde + NAD+ + H2O <=> Fatty acid + NADH + H+
R00702	2 trans,trans-Farnesyl diphosphate <=> Diphosphate + Presqualene diphosphate
R01185	Inositol 1-phosphate + H2O <=> myo-Inositol + Orthophosphate
R01186	myo-Inositol 4-phosphate + H2O <=> myo-Inositol + Orthophosphate
R01280	ATP + Hexadecanoic acid + CoA <=> AMP + Palmitoyl-CoA + Diphosphate
R01421	Benzoyl phosphate + H2O <=> Benzoate + Orthophosphate
R01498	beta-D-Glucosyl-(1<->1)-ceramide + H2O <=> D-Glucose + N-Acylsphingosine
R01687	(5-L-Glutamyl)-peptide + Taurine <=> Peptide + 5-L-Glutamyl-taurine
R02170	Penicillin + H2O <=> Carboxylate + 6-Aminopenicillanate
R02382	Tyramine + H2O + Oxygen <=> 4-Hydroxyphenylacetaldehyde + Ammonia + Hydrogen peroxide
R02404	ATP + Itaconate + CoA <=> ADP + Orthophosphate + Itaconyl-CoA
R02670	2 3-Hydroxyanthranilate + 4 Oxygen <=> Cinnavalininate + 2 O2.- + 2 Hydrogen peroxide + 2 H+
R02671	3-Hydroxyanthranilate + S-Adenosyl-L-methionine <=> 3-Methoxyanthranilate + S-Adenosyl-L-homocysteine
R02678	Indole-3-acetaldehyde + NAD+ + H2O <=> Indole-3-acetate + NADH + H+
R02782	2,4,6/3,5-Pentahydroxycyclohexanone <=> 3D-(3,5/4)-Trihydroxycyclohexane-1,2-dione + H2O
R02872	Presqualene diphosphate + NADPH + H+ <=> Diphosphate + Squalene + NADP+
R02912	Serotonin + 5-Methyltetrahydrofolate <=> 5-Methoxytryptamine + Tetrahydrofolate
R02957	D-Glucuronolactone + NAD+ + 2 H2O <=> D-Glucarate + NADH + H+
R03024	4-Nitrophenyl phosphate + H2O <=> 4-Nitrophenol + Orthophosphate
R03096	(Indol-3-yl)acetamide + H2O <=> Indole-3-acetate + Ammonia
R03371	Phytic acid + H2O <=> D-myo-Inositol 1,2,4,5,6-pentakisphosphate + Orthophosphate
R03596	Hydrogen selenide + 3 NADP+ + 3 H2O <=> Selenite + 3 NADPH + 5 H+

Rxn ID	Definition
R03599	L-Selenocysteine + Reduced acceptor <=> Hydrogen selenide + L-Alanine + Acceptor
R03867	Leukotriene C4 + Amino acid <=> Leukotriene D4 + 5-L-Glutamyl amino acid
R03893	cis-4-Carboxymethylenebut-2-en-4-olide + H2O <=> 2-Maleylacetate
R03943	N-Methyltyramine + S-Adenosyl-L-methionine <=> Hordenine + S-Adenosyl-L-homocysteine
R03955	Xanthurenic acid + S-Adenosyl-L-methionine <=> 8-Methoxykynurenate + S-Adenosyl-L-homocysteine
R03966	2-Hydroxymuconate <=> gamma-Oxalocrotonate
R03970	3-Cyano-L-alanine + L-Glutamate <=> gamma-Glutamyl-beta-cyanoalanine + H2O
R03971	3-Aminopropiononitrile + L-Glutamate <=> gamma-Glutamyl-beta-aminopropiononitrile + H2O
R04300	Dopamine + H2O + Oxygen <=> 3,4-Dihydroxyphenylacetaldehyde + Ammonia + Hydrogen peroxide
R04773	ATP + L-Selenomethionine + tRNA(Met) <=> AMP + Diphosphate + Selenomethionyl-tRNA(Met)
R04880	3,4-Dihydroxyphenylethyleneglycol + NAD+ <=> 3,4-Dihydroxymandelaldehyde + NADH + H+
R04903	5-Hydroxyindoleacetaldehyde + NAD+ + H2O <=> 5-Hydroxyindoleacetate + H+ + NADH
R04910	5-(3'-Carboxy-3'-oxopropenyl)-4,6-dihydroxypicolinate + NADPH + H+ <=> 5-(3'-Carboxy-3'-oxopropyl)-4,6-dihydroxypicolinate + NADP+
R04929	ATP + Selenate <=> Diphosphate + Adenylylselenate
R05233	trans-3-Chloro-2-propene-1-ol + NAD+ <=> trans-3-Chloroallyl aldehyde + NADH + H+
R05234	cis-3-Chloro-2-propene-1-ol + NAD+ <=> cis-3-Chloroallyl aldehyde + NADH + H+
R05237	trans-3-Chloroallyl aldehyde + H2O <=> trans-3-Chloroacrylic acid + 2 H+
R05238	cis-3-Chloroallyl aldehyde + H2O <=> cis-3-Chloroacrylic acid + 2 H+
R05265	4-Nitrocatechol + Oxygen + 3 H+ <=> Benzene-1,2,4-triol + Nitrite + H2O
R05286	Chloroacetaldehyde + NAD+ + H2O <=> Chloroacetic acid + NADH + H+
R05287	Chloroacetic acid + H2O <=> Glycolate + Hydrochloric acid
R05496	Acetylene + Reduced ferredoxin + 2 H+ + ATP + H2O <=> Ethylene + Oxidized ferredoxin + ADP + Orthophosphate
R05510	Protoanemonin + H2O <=> cis-Acetylacrylate
R05511	cis-2-Chloro-4-carboxymethylenebut-2-en-1,4-olide + H2O <=> 2-Chloromaleylacetate
R05551	Acrylic acid + Ammonia <=> Acrylamide + H2O
R05590	Benzamide + H2O <=> Benzoate + Ammonia
R05657	1-Phenanthrol + CH3-R <=> 1-Methoxyphenanthrene + R
R06363	Penicillin + H2O <=> Penicilloic acid
R06366	Perillyl aldehyde + H2O + NAD+ <=> Perillic acid + NADH + H+
R06401	alpha-Pinene + Reduced acceptor + Oxygen <=> Myrtenol + H2O + Acceptor
R06404	alpha-Pinene + Oxygen + 2 H+ + 2 e- <=> Pinocarveol + H2O
R06835	2,5-Dichloro-carboxymethylenebut-2-en-4-olide + H2O <=> 2,5-Dichloro-4-oxohex-2-enedioate
R06838	trans-4-Carboxymethylenebut-2-en-4-olide + H2O <=> 2-Maleylacetate
R06883	Bisphenol A + NADH + H+ + Oxygen <=> 1,2-Bis(4-hydroxyphenyl)-2-propanol + NAD+ + H2O
R06888	2,2-Bis(4-hydroxyphenyl)-1-propanol + NADH + H+ + Oxygen <=> 2,3-Bis(4-hydroxyphenyl)-1,2-propanediol + NAD+ + H2O
R06917	1-Hydroxymethylnaphthalene + NAD+ <=> 1-Naphthaldehyde + NADH + H+
R06927	(2-Naphthyl)methanol + NAD+ <=> 2-Naphthaldehyde + NADH + H+
R07322	Squalene <=> Diploptene
R07323	Squalene + H2O <=> Diplopterol
R07687	Anthracene + Oxygen + 2 H+ + 2 e- <=> Anthracene-9,10-dihydrodiol
R07697	Phenylboronic acid + Oxygen <=> Phenol + Boric acid

Rxn ID	Definition
R07700	Aniline + Oxygen <=> Catechol + Ammonia
R07706	Nitrobenzene + Oxygen <=> Catechol + Nitrite
R07766	Octanoyl-[acp] + Apoprotein <=> Protein N6-(octanoyl)lysine + Acyl-carrier protein
R07767	Protein N6-(octanoyl)lysine + 2 Sulfur donor + 2 S-Adenosyl-L-methionine <=> Protein N6-(lipoyl)lysine + 2 L-Methionine + 2 5'-Deoxyadenosine
R07768	Octanoyl-[acp] + 2 Sulfur donor + 2 S-Adenosyl-L-methionine <=> Lipoyl-[acp] + 2 L-Methionine + 2 5'-Deoxyadenosine
R07769	Lipoyl-[acp] + Apoprotein <=> Protein N6-(lipoyl)lysine + Acyl-carrier protein
R07770	ATP + Lipoate <=> Diphosphate + Lipoyl-AMP
R07771	Lipoyl-AMP + Apoprotein <=> Protein N6-(lipoyl)lysine + AMP
R07804	(+)-(3S,4R)-cis-3,4-Dihydroxy-3,4-dihydrofluorene <=> 3,4-Dihydroxyfluorene + 2 H+
R08014	Trinitrotoluene + 2 NADPH + 2 H+ <=> 4-Hydroxylamino-2,6-dinitrotoluene + 2 NADP+ + H2O
R08017	Trinitrotoluene + 2 NADH + 2 H+ <=> 4-Hydroxylamino-2,6-dinitrotoluene + 2 NAD+ + H2O
R08034	2,4-Diamino-6-nitrotoluene <=> 2,4-Diamino-6-hydroxylaminotoluene
R08042	Trinitrotoluene + 2 NADH + 2 H+ <=> 2-Hydroxylamino-4,6-dinitrotoluene + 2 NAD+ + H2O
R08105	4-Fluorocyclohexadiene-cis,cis-1,2-diol + NAD+ <=> 4-Fluorocatechol + NADH + H+
R08120	5-Fluoromuconolactone + H2O <=> 2-Maleylacetate + Hydrofluoric acid
R08121	4-Fluoromuconolactone + H2O <=> 2-Maleylacetate + Hydrofluoric acid
R08871	Demethylphosphinothricin + Acetyl-CoA <=> N-Acetyldemethylphosphinothricin + CoA
R09102	Bis(4-hydroxyphenyl)methane <=> Bis(4-hydroxyphenyl)methanol
R09104	4,4'-Dihydroxybenzophenone <=> 4-Hydroxyphenyl-4-hydroxybenzoate
R09132	cis-Chlorobenzene dihydrodiol + NAD+ <=> Tetrachlorocatechol + NADH + H+
R09133	cis-Chlorobenzene dihydrodiol + NADP+ <=> Tetrachlorocatechol + NADPH + H+
R09136	2,3,5-Trichlorodienelactone + H2O <=> 2,3,5-Trichloromaleylacetate
R09177	1-Hydroxypyrene + S-Adenosyl-L-methionine <=> 1-Methoxypyrene + S-Adenosyl-L-homocysteine
R09180	1-Hydroxy-6-methoxypyrene + CH3-R <=> 1,6-Dimethoxypyrene + R
R09220	3-Chloro-2-methyldienelactone + H2O <=> 3-Chloro-2-methylmaleylacetate
R09222	2-Chloro-5-methyl-cis-dienelactone + H2O <=> 2-Chloro-5-methylmaleylacetate
R09365	Selenohomocysteine + 5-Methyltetrahydropteroyltri-L-glutamate <=> L-Selenomethionine + Tetrahydropteroyltri-L-glutamate
R09372	2 NADPH + 2 H+ + Methylselenic acid <=> 2 NADP+ + 2 H2O + Methaneselenol
R10185	ATP + Methylphosphonate <=> alpha-D-Ribose 1-methylphosphonate 5-triphosphate + Adenine
R10186	alpha-D-Ribose 1-methylphosphonate 5-triphosphate + H2O <=> alpha-D-Ribose 1-methylphosphonate 5-phosphate + Diphosphate
R10204	alpha-D-Ribose 1-methylphosphonate 5-phosphate <=> alpha-D-Ribose 1,2-cyclic phosphate 5-phosphate + Methane

Appendix B Simplifications and assumptions in the *Anabaena* sp. PCC 7120 model

Table B-I. Simplifications and assumptions in the *Anabaena* sp. PCC 7120 model

#	Assumptions and simplifications	References
1	Transport reactions between compartments were assumed to be bidirectional, independent of ATP and unconstrained.	
2	Volume differences of intracellular compartments (cytosol, carboxysome, thylakoid membrane, thylakoid lumen, cell membrane, periplasmic space) were neglected, any dilution occurring due to metabolite transport between these compartments was not taken into account.	
3	Transport reactions with the extracellular space were assumed to be ATP-driven, unless evidence for a different mechanism (e.g. proton symport) was found in the literature.	
4	Diazotrophic filament of Anabaena sp. PCC 7120 is represented by a two-cell association of a single vegetative cell and a single heterocyst (super-compartments, Figure 3.1).	
5	The objective function of the two-cell model was defined as the sole growth of the vegetative cell.	(Yoon and Golden, 2001)
6	The heterocyst has constant flux over its biomass equation to account for its macromolecular turnover, equal to 10% of the maximum photodiazotrophic growth.	(van Bodegom, 2007)
7	Both super-compartments harbour an ATP hydrolysis reaction at fixed flux rate equal to 10% of the total ATP consumption at maximum growth rate, to account for non-growth associated maintenance.	(Feist et al., 2007; Nogales et al., 2012; Knoop et al., 2013)
8	Exchange reactions transferring metabolites between super-compartments were assumed to be bidirectional, unconstrained, proceed independently of ATP and provide direct exchange from cytosol to cytosol, without the involvement of other compartments or the external space.	(Flores et al., 2006; Walsby, 2007; Zhang et al., 2008; Flores and Herrero, 2010)
9	Any dilution occurring during the exchange of metabolites due to the size difference of the two super-compartments was neglected.	
10	The two super-compartments were assumed to share the same fractional composition comprising DNA, RNA, proteins, pigments, lipids, cell wall, inorganic ions and the pool fraction.	
11	The biomass composition of *Anabaena* sp. PCC 7120 is based on values originally set for *Synechocystis* sp. PCC 6803, and adapted to *Anabaena* wherever such dataset was available in the literature (see Table 3.2 and section 3.2.2 for details).	
12	In the initial two-cell model, the two super-compartments were allowed to exchange only four metabolites: sucrose, glutamine, glutamate and 2-oxoglutarate.	
13	For the fixed Glu:Gln ratio (Figure 3.3E and F) a special antiport-type exchange reaction was included while closing down the single-metabolite exchange for Glu and Gln, respectively.	
14	For comparative reasons carbon source uptake reactions were assumed to consume one mole of ATP for every mole of substrate transported (Figure 3.5). It was also assumed that, unless otherwise noted, ammonia excretion was not possible.	

Appendix C Differences of the two super-compartments in the *Anabaena* sp. PCC 7120 model

Table C-I. Differences of the two super-compartments in the *Anabaena* sp. PCC 7120 model

Rxn ID	Description of difference	Reference
R00024m	not in HCSC (RuBisCO)	(Madan and Nierzwicki-Bauer, 1993; Valladares et al., 2007)
R00021	not in HCSC (GOGAT)	(Martin-Figueroa et al., 2000)
R03140	not in HCSC (RuBisCO)	(Madan and Nierzwicki-Bauer, 1993; Valladares et al., 2007)
R10092	not in HCSC (carbonic anhydrase in carboxysome)	(Flaherty et al., 2011; Park et al., 2013)
RA0444	not in HCSC (light harvesting at PSII)	(Wolk et al., 2004)
RA0482, RA0484, RA0485, RA0486, RA0487, RA0488, RA0489, RA0490, RA0491	not in HCSC (electron transport at PSII)	(Wolk et al., 2004)
RT0065, RT0066, RT0067, RT0068, RT0069, RT0070	not in HCSC (cytosol-carboxysome transport)	
hRT0079	not in VCSC (nitrogen transport for nitrogenase)	
hRT0080	not in VCSC (hydrogen transport for nitrogenase)	
hR05185	not in VCSC (nitrogen fixation)	(Rice et al., 1982)

Appendix D List of reactions in the model included from primary literature

Table D-I. Reactions added to the *Anabaena* sp. PCC 7120 model based on biochemical or bioinformatics evidence in the literature.

Rxn ID	Description	References
RN0017	4 H+ + Plastoquinone-9 + 2 Reduced ferredoxin -> Plastoquinol-9 + 2 Oxidized ferredoxin + 4 H+	(Mi et al., 1992; Razquin et al., 1996; Ogawa and Mi, 2007; Battchikova et al., 2011)
RN0019	NADPH + 1-Hydroxy-1,2-dihydrolycopene -> H+ + NADP+ + 1'-Hydroxy-gamma-carotene	(Mochimaru et al., 2008; Graham and Bryant, 2009)
RN0020	H2O + 1'-Hydroxy-gamma-carotene -> 2 H+ + Plectaniaxanthin	(Mochimaru et al., 2008; Graham and Bryant, 2009)
RN0021	H2O + Plectaniaxanthin -> 2 H+ + Myxol	(Mochimaru et al., 2008; Graham and Bryant, 2009)
RN0022	GDP-L-fucose + Myxol -> GDP + (3R,2'S)-Myxol 2'-alpha-L-fucoside	(Mochimaru et al., 2008; Graham and Bryant, 2009)
RT0003	H+ + ATP + H2O -> ADP + Pi + schizokinen-Fe(III) + H+	(Nicolaisen et al., 2008)
RN0064	[MoaD]-COOH + ATP -> [MoaD]-CO-AMP + PPi	(Gutzke et al., 2001)
RN0065	[IscS]-S-sulfanylcysteine + [MoaD]-CO-AMP -> [MoaD]-COSH + [IscS]-cysteine + AMP	(Gutzke et al., 2001)
RT0025	Formate ->	(Heyer and Krumbein, 1991)
RN0069	[BioB]-COOH + ATP -> [BioB]-CO-AMP + PPi	(Lin and Cronan, 2011)
RN0070	[IscS]-S-sulfanylcysteine + [BioB]-CO-AMP -> [BioB]-COSH + [IscS]-cysteine + AMP	(Lin and Cronan, 2011)
R00013	2 Glyoxylate -> CO2 + 2-Hydroxy-3-oxopropanoate	(Eisenhut et al., 2008)
R00021	2 H+ + 2-Oxoglutarate + L-Glutamine + 2 Reduced ferredoxin -> 2 L-Glutamate + 2 Oxidized ferredoxin	(Martin-Figueroa et al., 2000)
R00178	H+ + S-Adenosyl-L-methionine -> CO2 + S-Adenosylmethioninamine	(Jantaro et al., 2003; Incharoensakdi et al., 2010)
R00248	H+ + NADPH + 2-Oxoglutarate + NH3 -> H2O + NADP+ + L-Glutamate	(Muro-Pastor et al., 2005; Luque and Forchhammer, 2008)
R00272	2-Oxoglutarate -> CO2 + Succinate semialdehyde	(Zhang and Bryant, 2011)
R00311m	Heme + 4 Reduced ferredoxin + 3 O2 -> Biliverdin + CO + Fe2+ + 4 Oxidized ferredoxin + 3 H2O	(Schluchter and Glazer, 1997; Biswas, 2011)
R00342	NAD+ + (S)-Malate <=> H+ + NADH + Oxaloacetate	(Eisenhut et al., 2008)
R00372	Glyoxylate + L-Glutamate -> 2-Oxoglutarate + Glycine	(Eisenhut et al., 2008)
R00466	H2O + O2 + Glyoxylate -> Hydrogen peroxide + Oxalate	(Eisenhut et al., 2008)
R00475	O2 + Glycolate -> Hydrogen peroxide + Glyoxylate	(Eisenhut et al., 2008)
R00476m	NAD+ + Glycolate -> H+ + Glyoxylate + NADH	(Eisenhut et al., 2008)
R00551	H2O + L-Arginine <=> Urea + L-Ornithine	(Schriek et al., 2007)
R00552	H2O + L-Arginine -> NH3 + L-Citrulline	(Schriek et al., 2007)
R00566	L-Arginine -> CO2 + Agmatine	(Schriek et al., 2007)
R00588	Glyoxylate + L-Serine <=> Glycine + Hydroxypyruvate	(Eisenhut et al., 2008)
R00667	2-Oxoglutarate + L-Ornithine <=> L-Glutamate + L-Glutamate 5-semialdehyde	(Schriek et al., 2007)

Rxn ID	Description	References
R00707	2 H2O + NAD+ + (S)-1-Pyrroline-5-carboxylate -> H+ + L-Glutamate + NADH	(Schriek et al., 2007)
R00714	H2O + NADP+ + Succinate semialdehyde -> H+ + NADPH + Succinate	(Schriek et al., 2007)
R00945	H2O + Glycine + 5,10-Methylenetetrahydrofolate <=> L-Serine + Tetrahydrofolate	(Eisenhut et al., 2008)
R01078m	Dethiobiotin + [BioB]-COSH + 2 S-Adenosyl-L-methionine -> Biotin + 2 L-Methionine + 2 5'-Deoxyadenosine + [BioB]-COOH	(Lin and Cronan, 2011)
R01221	NAD+ + Glycine + Tetrahydrofolate -> H+ + CO2 + NH3 + NADH + 5,10-Methylenetetrahydrofolate	(Eisenhut et al., 2008)
R01334	H2O + 2-Phosphoglycolate -> Pi + Glycolate	(Eisenhut et al., 2008)
R01398	L-Ornithine + Carbamoyl phosphate <=> Pi + L-Citrulline	(Schriek et al., 2007)
R01648	2-Oxoglutarate + 4-Aminobutanoate <=> L-Glutamate + Succinate semialdehyde	(Schriek et al., 2007)
R01747	H+ + NADPH + 2-Hydroxy-3-oxopropanoate -> NADP+ + D-Glycerate	(Eisenhut et al., 2008)
R01920	Putrescine + S-Adenosylmethioninamine -> H+ + 5'-Methylthioadenosine + Spermidine	(Jantaro et al., 2003; Incharoensakdi et al., 2010)
R01990	H2O + 4-Guanidinobutanoate -> Urea + 4-Aminobutanoate	(Schriek et al., 2007)
R02029m	H2O + Phosphatidylglycerophosphate_PG -> Pi + Phosphatidylglycerol_PG	(Wu et al., 2006)
R02185	D-Glucose + Polyphosphate -> D-Glucose 6-phosphate	(Klemke et al., 2014)
R04325	10-Formyltetrahydrofolate + 5'-Phosphoribosylglycinamide -> Tetrahydrofolate + 5'-Phosphoribosyl-N-formylglycinamide	(Marolewski et al., 1997)
R04472m	2 3-beta-D-Galactosyl-1,2-diacylglycerol_DGDG -> Digalactosyl-diacylglycerol_DGDG + 1,2-Diacylglycerol_DGDG	(Awai et al., 2006; Awai et al., 2007)
R05692	NADP+ + GDP-L-fucose <=> H+ + NADPH + GDP-4-dehydro-6-deoxy-D-mannose	(Mochimaru et al., 2008)
R05817m	4 Reduced ferredoxin + Biliverdin <=> 4 Oxidized ferredoxin + (3Z)-Phycocyanobilin	(Biswas, 2011)
R06867m	UDP-6-sulfoquinovose + 1,2-Diacylglycerol_SQDG -> UDP + Sulfoquinovosyldiacylglycerol_SQDG	(Güler et al., 2000)
R06896	H+ + NADPH + Divinylprotochlorophyllide -> NADP+ + Protochlorophyllide	(Islam et al., 2008)
R07324	D-Glucose 6-phosphate -> 1D-myo-Inositol 3-phosphate	(Chatterjee et al., 2004)
R07399	H2O + Acetyl-CoA + Pyruvate -> CoA + (R)-2-Methylmalate	(Wu et al., 2010)
R07501	S-Adenosyl-L-methionine + 2-Methyl-6-phytylquinol -> S-Adenosyl-L-homocysteine + 2,3-Dimethyl-5-phytylquinol	(Sadre et al., 2012)
R07512	7,9,7',9'-tetracis-Lycopene <=> Lycopene	(Mochimaru et al., 2008)
R07519	H2O + Lycopene -> 1-Hydroxy-1,2-dihydrolycopene	(Mochimaru et al., 2008; Graham and Bryant, 2009)
R07557m	O2 + NADPH + (3R,2'S)-Myxol 2'-alpha-L-fucoside -> 2 H+ + H2O + NADP+ + (3S,2'S)-4-Ketomyxol 2'-alpha-L-fucoside	(Mochimaru et al., 2008)
R07563m	Echinenone + O2 -> Canthaxanthin + H2O	(Mochimaru et al., 2008)
R08676	ADP + Sucrose -> D-Fructose + ADP-glucose	(Cumino et al., 2007; Marcozzi et al., 2009)

Rxn ID	Description	References
R09395m	Precursor Z + 2 [MoaD]-COSH -> Molybdopterin + 2 [MoaD]-COOH	(Gutzke et al., 2001)
R09656m	H+ + O2 + NADPH + 9,9'-dicis-zeta-Carotene -> 2 H2O + 7,9,9'-tricis-Neurosporene	(Albrecht et al., 1996)
R09658m	H+ + O2 + NADPH + 7,9,9'-tricis-Neurosporene -> 2 H2O + 7,9,7',9'-tetracis-Lycopene	(Albrecht et al., 1996)
RA0102	O2 + NADPH + beta-Carotene -> 2 H+ + H2O + NADP+ + Echinenone	(Mochimaru et al., 2008)
R10709	S-Adenosyl-L-methionine + 2-Methyl-6-solanyl-1,4-benzoquinol -> S-Adenosyl-L-homocysteine + Plastoquinol-9	(Cheng et al., 2003; Sadre et al., 2012)
RA0177	UDP-glucose + 1,2-Diacylglycerol_MGDG -> UDP + 3-D-Glucosyl-1,2-diacylglycerol_MGDG	(Awai et al., 2006)
RA0178	UDP-glucose + 1,2-Diacylglycerol_DGDG -> UDP + 3-D-Glucosyl-1,2-diacylglycerol_DGDG	(Awai et al., 2006)
RA0179	3-D-Glucosyl-1,2-diacylglycerol_MGDG <=> 3-beta-D-Galactosyl-1,2-diacylglycerol_MGDG	(Awai et al., 2006)
RA0180	3-D-Glucosyl-1,2-diacylglycerol_DGDG <=> 3-beta-D-Galactosyl-1,2-diacylglycerol_DGDG	(Awai et al., 2006)
RA0054	H2O + L-Arginine + Plastoquinone-9 -> NH3 + Plastoquinol-9 + 5-Guanidino-2-oxopentanoate	(Schriek et al., 2007)
RA0055	H2O + L-Arginine + Plastoquinone-9 -> NH3 + Plastoquinol-9 + 5-Guanidino-2-oxopentanoate	(Schriek et al., 2007)
RA0056	Hydrogen peroxide + 5-Guanidino-2-oxopentanoate -> H2O + CO2 + 4-Guanidinobutanoate	(Schriek et al., 2007)
RA0562	ATP + Formate + 5'-Phosphoribosylglycinamide -> ADP + Pi + 5'-Phosphoribosyl-N-formylglycinamide	(Marolewski et al., 1997)
RA0600	4 H+ + 3-Nonaprenyl-4-hydroxybenzoate -> H2O + 2-Methyl-6-solanyl-1,4-benzoquinol	(Sadre et al., 2012)
RT0032	-> Bicarbonate + Na+	(Shibata et al., 2002)
RT0036	-> L-Glutamate + Na+	(Quintero et al., 2001)
RT0040	-> NH4+	(Muro-Pastor et al., 2005)
RT0047	Na+ -> K+	(Matsuda et al., 2004)

Appendix E List of newly annotated genes and associated reactions in the *Anabaena* sp. PCC 7120 model

Table E-I. List of newly annotated genes in the *Anabaena* sp. PCC 7120 model. EC numbers, gene-protein-reaction associations and detailed BLAST results with similarity scores are shown.

Rxn ID	EC	GPR	BLAST score
R00009	1.11.1.6	alr0998 OR alr3090	BLASTX: alr0998, Mn-containing catalase [Nostoc sp. PCC7524], score: 535, e-value: 0, identity 89%; alr3090, manganese containing catalase [Anabaena variabilis ATCC 29413], score: 475, e-value: 2e-168, identity: 99%
R00025	1.13.12.16	alr5356	BLASTX: alr5356, 2-nitropropane dioxygenase [Anabaena variabilis ATCC 29413], score: 898, e-value: 0, identity: 98%
R00230	2.3.1.8	all4665	BLASTX: all4665, phosphate acetyltransferase [Anabaena sp. 90], score: 550, e-value: 0, identity: 76%
R00272	4.1.1.71	all3555	BLASTX: all3555, acetolactate synthase [Synechococcus sp. PCC 7002], score: 888, e-value: 0, identity: 77%
R00366	1.4.3.3	all3519 OR all2776 OR all1354	BLASTX: all3519, glycine oxidase [Nostoc sp. PCC 7524], score: 999, e-value: 0, identity: 86%; all2776, glycine/D-amino acid oxidase, deaminating [Nostoc sp. PCC 7524], score: 610, e-value: 0, identity: 79%; all1354, glycine/D-amino acid oxidase, deaminating [Nostoc sp. PCC 7524], score: 601, e-value: 0, identity: 81%
R00462	4.1.1.18	all4887	BLASTX: all4887, arginine/lysine/ornithine decarboxylase [Nostoc sp. PCC 7524], score: 745, e-value: 0, identity: 80%
R00509	2.7.1.25, 2.7.7.4	alr4787	BLASTX: alr4787, adenylylsulfate kinase [Anabaena cylindrica 7122], score: 835, e-value: 0, identity: 77%
R00582	2.7.1.39, 3.1.3.3	all1758	BLASTX: all1758, serine phosphatase [Anabaena variabilis ATCC 29413], score: 953, e-value: 0, identity: 99%
R00669	3.5.1.14, 3.5.1.16	all2102 OR alr4934	BLASTX: all2102, N-acyl-L-amino acid amidohydrolase [Richelia intracellularis], score: 615, e-value: 0, identity: 73%; alr4934, N-acyl-L-amino acid amidohydrolase [Microcystis aeruginosa NIES-843], score: 636, e-value: 0, identity: 75%
R00761	4.1.2.22	all1483 OR alr1850	BLASTX: all1483, xylulose-5-phosphate phosphoketolase/fructose-6-phosphate phosphoketolase [Crocosphaera watsonii WH 0402], score: 1340, e-value: 0, identity: 77%; alr1850, D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase [Nostoc punctiforme PCC 73102], score: 1338, e-value: 0, identity: 88%
R00766	2.4.1.14	alr3370 OR all4376	BLASTX: alr3370, sucrose-phosphate synthase [Gloeobacter violaceus PCC 7421], score: 450, e-value: 2e-152, identity: 56%; all4376, sucrose-phosphate synthase [Gloeobacter violaceus PCC 7421], score: 449, e-value: 6e-152, identity: 56%
R00805	3.1.3.24	all0376	BLASTX: all0376, sucrose-phosphate phosphatase [Anabaena variabilis ATCC 29413], score: 356, e-value: 5e-122, identity: 94%
R00883	2.7.7.22	alr0188	BLASTX: alr0188, GDP-mannose pyrophosphorylase [Anabaena sp. CH1], score: 689, e-value: 0, identity: 97%
R00894	6.3.2.2	alr3351	BLASTX: alr3351, glutamate--cysteine ligase [Anabaena variabilis ATCC 29413], score: 726, e-value: 0, identity: 98%
R01041	1.1.1.2, 1.1.1.21, 1.1.1.72	alr2054	BLASTX: aldo/keto reductase [Anabaena variabilis ATCC 29413], score: 603, e-value: 0, identity: 96%

Rxn ID	EC	GPR	BLAST score
R01090	2.6.1.42, 2.6.1.6	alr1260	BLASTX: alr1260, branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase [Nostoc sp. PCC 7524], score: 406, e-value: 4e-140, identity: 74%
R01214	2.6.1.42, 2.6.1.6	alr1260	BLASTX: alr1260, branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase [Nostoc sp. PCC 7524], score: 406, e-value: 4e-140, identity: 74%
R01224	1.5.1.20	all0783	BLASTX: all0783, methylenetetrahydrofolate reductase [Anabaena variabilis ATCC 29413], score: 616, e-value: 0, identity: 96%
R01286	4.4.1.8	alr1600	BLASTX: alr1600, cystathionine beta-lyase family protein [Nostoc PCC 7524], score: 769, e-value: 0, identity: 94%
R01302	4.1.3.40	all0938	BLASTX: all0938, chorismate lyase [Nostoc 7107], identity: 91%
R01392	1.1.1.81	alr1890 OR all8087	BLASTX: alr1890, D-3-phosphoglycerate dehydrogenase [Anabaena variabilis ATCC 29413], score: 982, e-value: 0, identity: 99%; all8087, 2-hydroxyacid dehydrogenase [Nodularia spumigena], score: 604, e-value: 0, identity: 88%
R01728	1.3.1.12, 1.3.1.43, 5.4.99.5	all1141 OR all0418	BLASTX: all1141, prephenate dehydrogenase [Nostoc sp. PCC 7524], score: 459, e-value: 1e-160, identity: 82%; all0418, prephenate dehydrogenase [Nostoc punctiforme PCC 73102], score: 414, e-value: 1e-141, identity: 73%
R01825	1.2.1.72	all2566 OR all5062 OR alr1095	BLASTX: alr1095, glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase [Prochlorococcus marinus str. MIT 9301], score: 450, e-value: 2e-148, identity: 60%
R02029m	3.1.3.27	alr1715	BLASTX: alr1715, phosphatidylglycerophosphatase B [Stenotrophomonas sp. SKA14], score: 110, e-value: 9e-25, identity: 39%
R02199	2.6.1.42	alr1260	BLASTX: alr1260, branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase [Nostoc sp. PCC 7524], score: 406, e-value: 4e-140, identity: 74%
R03222	1.3.3.4	alr0115	BLASTX: alr0115, protoporphyrinogen oxidase [Nodularia spumigena CCY9414], score: 396, e-value: 8e-135, identity: 71%
R04148	2.4.2.21	alr0297	BLASTX: alr0297, nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase [Nostoc punctiforme PCC 73102], score: 520, e-value: 0, identity: 75%
R04210	1.1.1.290	alr1890	BLASTP: low BLASTP homology for alr1890 to heterologous enzymes
R04594	3.1.3.73	alr3338 OR alr5200 OR alr1107	BLASTX: alr3338, alpha-ribazole phosphatase [Lyngbya aestuarii], score: 565, e-value: 0, identity: 61%; alr5200, alpha-ribazole phosphatase [Lyngbya aestuarii], score: 292, e-value: 4e-53, identity: 46%; alr1107, alpha-ribazole phosphatase [Microcystis aeruginosa SPC777], score: 280, e-value: 6e-92, identity: 60%
R05217	1.14.13.83	all0456 OR alr0607	BLASTX: all0456, precorrin-3B synthase [Nostoc punctiforme PCC 73102], score: 662, e-value: 0, identity: 67%; alr0607, precorrin-3B synthase [Lyngbya aestuarii], score: 818, e-value: 0, identity: 75%
R05218	1.16.8.1	all1864	BLASTP: all1864, cob(II)yrinic acid a,c-diamide reductase [Thalassiobium sp. R2A62], score: 55.4, e-value: 4e-09, identity: 27%
R05219	2.1.1.152	all2847 OR all0455 OR all4698	BLASTX: all2847, cobalt-precorrin-6A synthase [Anabaena variabilis ATCC 29413], score: 735, e-value: 0, identity: 96%

Rxn ID	EC	GPR	BLAST score
R06530	4.1.1.81	alr3936	BLASTX: alr3936, threonine-phosphate decarboxylase [Anabaena variabilis ATCC 29413], score: 627, e-value: 0, identity: 95%
R06867m	2.4.1.-	alr2265	BLASTX: alr2265, sulfolipid sulfoquinovosyldiacylglycerol biosynthesis protein [Synechococcus elongatus PCC 7942], score: 530, e-value: 0, identity: 73%
R06896	1.3.1.75	all1601	BLASTX: all1601, hypothetical protein slr1923 [Synechocystis sp. PCC 6803], score: 693, e-value: 0, identity: 82%
R07280	3.1.3.-	alr0221	BLASTX: alr0221, phosphohistidine phosphatase SixA [Anabaena variabilis ATCC 29413], score: 320, e-value: 2e-109, identity: 94%
R07324	5.5.1.4	all5196	BLASTX: all5196, hypothetical protein sll1722 [Synechocystis sp. PCC 6803], score: 411, e-value: 5e-141, identity: 53%
R07392	4.2.1.109	all1723	BLASTX: methylthioribulose-1-phosphate dehydratase [Chroococcidiopsis thermalis PCC 7203], score: 73.5, e-value: 2e-15, identity: 29%
R07399	2.3.1.182	alr3522	BLASTX: alr3522, (R)-citramalate synthase [Crocosphaera watsonii], score: 815, e-value: 0, identity: 70%
R07501	2.1.1.295	all2121	BLASTX: all2121, 2-methyl-6-solanesyl-1,4-benzoquinol methyltransferase [Stanieria cyanosphaera PCC 7437], score: 551, e-value: 0, identity: 84%
R08639	5.4.2.2	all3964 OR all5089	BLASTX: all3964, phosphoglucomutase/phosphomannomutase [Anabaena variabilis ATCC 29413], score: 869, e-value: 0, identity: 99%; all5089, phosphoglucomutase/phosphomannomutase [Anabaena variabilis ATCC 29413], score: 855, e-value: 0, identity: 98%
R09083m	1.13.11.79	all1864	BLASTP: all1864, nitroreductase [Anabaena variabilis ATCC 29413], score: 441, e-value: 1e-154, identity: 98%
R09655	5.2.1.12	alr3954	BLASTX: alr3954, zeta-carotene isomerase nnrU [Thermosynechococcus sp. NK55], score: 343, e-value: 5e-116, identity: 72%
R10709	2.1.1.295	all2121	BLASTX: all2121, 2-methyl-6-solanesyl-1,4-benzoquinol methyltransferase [Stanieria cyanosphaera PCC 7437], score: 551, e-value: 0, identity: 84%
RA0166	3.1.3.4	all7623	BLASTX: all7623, phosphatidic acid phosphatase [Microcoleus vaginatus], score: 161, e-value: 3e-45, identity: 48%
RA0167	3.1.3.4	all7623	BLASTX: all7623, phosphatidic acid phosphatase [Microcoleus vaginatus], score: 161, e-value: 3e-45, identity: 48%
RA0168	3.1.3.4	all7623	BLASTX: all7623, phosphatidic acid phosphatase [Microcoleus vaginatus], score: 161, e-value: 3e-45, identity: 48%
RA0373	1.14.19.3	all1597	BLASTX: all1597, linoleoyl desaturase [Ricinus communis], score: 371, e-value: 5e-122, identity: 52%
RA0375	1.14.19.3	all1597	BLASTX: all1597, linoleoyl desaturase [Ricinus communis], score: 371, e-value: 5e-122, identity: 52%
RA0416	2.4.1.129	alr0718	BLASTX: alr0718, peptidoglycan glycosyltransferase [Calothrix sp. PCC 7507], score: 858, e-value: 0, identity: 73%
RN0019	CruA or CruP	alr3524	BLASTX: lycopene cyclase, CruA type [Richelia intracellularis], score: 1061, e-value: 0, identity: 73%
RN0022	CruG	all0143	BLASTX: 2'-O-glycosyltransferase CruG [Nodularia spumigena CCY9414], score: 581, e-value: 0, identity: 86%
RN0005	RhbE	all0392	BLASTX: all0392, RhbE rhizobactin siderophore biosynthesis protein [Sinorhizobium meliloti 1021], score: 407, e-value: 7e-139, identity: 46%
RN0006	RhbD	all0393	BLASTX: all0393, RhbD rhizobactin siderophore biosynthesis protein [Sinorhizobium meliloti 1021], score: 177, e-value: 2e-56, identity: 44%
RN0007	RhbC	all0394	BLASTX: all0394, RhbC rhizobactin biosynthesis protein [Sinorhizobium meliloti 1021], score: 354, e-value: 3e-114, identity: 39%

Rxn ID	EC	GPR	BLAST score
RN0008	RhbF	all0390	BLASTX: all0390, RhbF rhizobactin siderophore biosynthesis protein RhsF [Sinorhizobium meliloti 1021], score: 498, e-value: 4e-170, identity: 44%

Appendix F Transport and exchange reactions in the *Anabaena* sp. PCC 7120 model

Table F-I. Transport reactions in the *Anabaena* sp. PCC 7120 model

Rxn ID	Description	Bounds	Reference
RT0001	<=> Mn2+[vc] + H+[vc]	-1000, 1000	
RT0002	ATP[vc] + H2O[vc] -> ADP[vc] + Fe(III)dicitrate[vc] + Pi[vc]	0, 1000	
RT0003	H+[vc] + ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + schizokinen-Fe(III)[vc] + H+[vp]	0, 1000	(Nicolaisen et al., 2008)
RT0005	<=> NH3[vc]	0, 0	
RT0006	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Nitrite[vc]	0, 0	
RT0007	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Cyanate[vc]	0, 0	
RT0008	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Cysteine[vc]	0, 0	
RT0009	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Tyrosine[vc]	0, 0	
RT0010	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Tryptophan[vc]	0, 0	
RT0011	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Isoleucine[vc]	0, 0	
RT0012	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Valine[vc]	0, 0	
RT0013	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Aspartate[vc]	0, 0	
RT0014	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Threonine[vc]	0, 0	
RT0015	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Serine[vc]	0, 0	
RT0016	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Proline[vc]	0, 0	
RT0017	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Phenylalanine[vc]	0, 0	
RT0018	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Leucine[vc]	0, 0	
RT0019	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Methionine[vc]	0, 0	
RT0020	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Glycine[vc]	0, 0	
RT0021	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Alanine[vc]	0, 0	
RT0022	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Asparagine[vc]	0, 0	
RT0023	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Lysine[vc]	0, 0	
RT0024	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Ornithine[vc]	0, 0	
RT0077	Dialurate[vc] ->	0, 1000	
RT0078	Methanol[vc] ->	0, 1000	
RT0025	Formate[vc] ->	0, 1000	(Heyer and Krumbein, 1991)

Rxn ID	Description	Bounds	Reference
RT0004	schizokinen[vc] <=>	-1000, 1000	
RT0026	-> Photon[vt]	0, 10	
RT0027	<=> H+[vc]	-1000, 1000	
RT0028	<=> O2[vc]	-1000, 1000	
RT0029	CO2[vc] ->	0, 1000	
RT0030	<=> H2O[vc]	-1000, 1000	
RT0031	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Bicarbonate[vc]	0, 10	
RT0032	-> Bicarbonate[vc] + Na+[vc]	0, 10	(Shibata et al., 2002)
RT0033	-> H+[vc] + D-Glucose[vc]	0, 0	
RT0034	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Glutamine[vc]	0, 0	
RT0035	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Arginine[vc]	0, 0	
RT0036	-> L-Glutamate[vc] + Na+[vc]	0, 0	(Montesinos et al., 1995; Quintero et al., 2001)
RT0037	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + L-Histidine[vc]	0, 0	
RT0038	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Urea[vc]	0, 0	
RT0039	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Nitrate[vc]	0, 0	
RT0040	-> NH4+[vc]	0, 0	(Muro-Pastor et al., 2005)
RT0041	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Sulfate[vc]	0, 1000	
RT0042	-> Sulfate[vc]	0, 1000	
RT0043	ATP[vc] + H2O[vc] -> ADP[vc] + 2 Pi[vc]	0, 1000	
RT0044	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Iron chelate[vc]	0, 1000	
RT0045	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Mn2+[vc]	0, 1000	
RT0046	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + K+[vc]	0, 1000	
RT0047	Na+[vc] -> K+[vc]	0, 1000	(Matsuda et al., 2004)
RT0048	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Zn2+[vc]	0, 1000	
RT0049	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Ni2+[vc]	0, 1000	
RT0050	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Co2+[vc]	0, 1000	
RT0051	H2O[vc] + GTP[vc] -> Pi[vc] + Fe2+[vc] + GDP[vc]	0, 1000	
RT0052	-> Mg2+[vc]	0, 1000	
RT0054	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Ca2+[vc]	0, 1000	
RT0055	H+[vc] <=> Ca2+[vc]	-1000, 1000	

Rxn ID	Description	Bounds	Reference
RT0056	H+[vc] <=> Na+[vc]	-1000, 1000	
RT0057	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Cu2+[vc]	0, 1000	
RT0058	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Molybdate[vc]	0, 1000	
RT0059	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Fe3+[vc]	0, 1000	
RT0060	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Putrescine[vc]	0, 0	
RT0061	ATP[vc] + H2O[vc] -> ADP[vc] + Pi[vc] + Spermidine[vc]	0, 0	
RT0062	5-Deoxy-D-ribose [vc] ->	0, 1000	
RT0063	S-Methyl-5-thio-D-ribulose 1-phosphate[vc] ->	0, 1000	
RT0064	CO[vc] ->	0, 1000	
RT0065	H+[vc] <=> H+[vx]	-1000, 1000	
RT0066	O2[vc] -> O2[vx]	0, 1000	
RT0067	Bicarbonate[vc] -> Bicarbonate[vx]	0, 1000	
RT0068	D-Ribulose 1,5-bisphosphate[vc] -> D-Ribulose 1,5-bisphosphate[vx]	0, 1000	
RT0069	3-Phospho-D-glycerate[vx] -> 3-Phospho-D-glycerate[vc]	0, 1000	
RT0070	2-Phosphoglycolate[vx] -> 2-Phosphoglycolate[vc]	0, 1000	
RT0071	H2O[vc] <=> H2O[vl]	-1000, 1000	
RT0072	O2[vl] <=> O2[vc]	-1000, 1000	
RT0073	Plastoquinone-9[vc] -> Plastoquinone-9[vt]	0, 1000	
RT0074	H2O[vp] <=> H2O[vc]	-1000, 1000	
RT0075	O2[vp] <=> O2[vc]	-1000, 1000	
RT0076	Plastoquinone-9[vc] -> Plastoquinone-9[vm]	0, 1000	
hRT0001	<=> Mn2+[hc] + H+[hc]	-1000, 1000	
hRT0079	<=> N2[hc]	-1000, 10	
hRT0080	H2[hc] <=>	-1000, 1000	
hRT0002	ATP[hc] + H2O[hc] -> ADP[hc] + Fe(III)dicitrate[hc] + Pi[hc]	0, 1000	
hRT0003	H+[hc] + ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + schizokinen-Fe(III)[hc] + H+[hp]	0, 1000	(Nicolaisen et al., 2008)
hRT0005	<=> NH3[hc]	0, 0	
hRT0006	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Nitrite[hc]	0, 0	
hRT0007	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Cyanate[hc]	0, 0	
hRT0008	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Cysteine[hc]	0, 0	

Rxn ID	Description	Bounds	Reference
hRT0009	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Tyrosine[hc]	0, 0	
hRT0010	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Tryptophan[hc]	0, 0	
hRT0011	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Isoleucine[hc]	0, 0	
hRT0012	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Valine[hc]	0, 0	
hRT0013	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Aspartate[hc]	0, 0	
hRT0014	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Threonine[hc]	0, 0	
hRT0015	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Serine[hc]	0, 0	
hRT0016	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Proline[hc]	0, 0	
hRT0017	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Phenylalanine[hc]	0, 0	
hRT0018	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Leucine[hc]	0, 0	
hRT0019	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Methionine[hc]	0, 0	
hRT0020	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Glycine[hc]	0, 0	
hRT0021	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Alanine[hc]	0, 0	
hRT0022	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Asparagine[hc]	0, 0	
hRT0023	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Lysine[hc]	0, 0	
hRT0024	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Ornithine[hc]	0, 0	
hRT0077	Dialurate[hc] ->	0, 1000	
hRT0078	Methanol[hc] ->	0, 1000	
hRT0025	Formate[hc] ->	0, 1000	(Heyer and Krumbein, 1991)
hRT0004	schizokinen[hc] <=>	-1000, 1000	
hRT0026	-> Photon[ht]	0, 10	
hRT0027	<=> H+[hc]	-1000, 1000	
hRT0028	<=> O2[hc]	-1000, 1000	
hRT0029	CO2[hc] ->	0, 1000	
hRT0030	<=> H2O[hc]	-1000, 1000	
hRT0031	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Bicarbonate[hc]	0, 10	
hRT0032	-> Bicarbonate[hc] + Na+[hc]	0, 10	(Shibata et al., 2002)
hRT0033	-> H+[hc] + D-Glucose[hc]	0, 0	
hRT0034	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Glutamine[hc]	0, 0	
hRT0035	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Arginine[hc]	0, 0	
hRT0036	-> L-Glutamate[hc] + Na+[hc]	0, 0	(Quintero et al., 2001)

Rxn ID	Description	Bounds	Reference
hRT0037	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + L-Histidine[hc]	0, 0	
hRT0038	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Urea[hc]	0, 0	
hRT0039	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Nitrate[hc]	0, 0	
hRT0040	-> NH4+[hc]	0, 0	(Muro-Pastor et al., 2005)
hRT0041	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Sulfate[hc]	0, 1000	
hRT0042	-> Sulfate[hc]	0, 1000	
hRT0043	ATP[hc] + H2O[hc] -> ADP[hc] + 2 Pi[hc]	0, 1000	
hRT0044	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Iron chelate[hc]	0, 1000	
hRT0045	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Mn2+[hc]	0, 1000	
hRT0046	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + K+[hc]	0, 1000	
hRT0047	Na+[hc] -> K+[hc]	0, 1000	(Matsuda et al., 2004)
hRT0048	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Zn2+[hc]	0, 1000	
hRT0049	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Ni2+[hc]	0, 1000	
hRT0050	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Co2+[hc]	0, 1000	
hRT0051	H2O[hc] + GTP[hc] -> Pi[hc] + Fe2+[hc] + GDP[hc]	0, 1000	
hRT0052	-> Mg2+[hc]	0, 1000	
hRT0054	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Ca2+[hc]	0, 1000	
hRT0055	H+[hc] <=> Ca2+[hc]	-1000, 1000	
hRT0056	H+[hc] <=> Na+[hc]	-1000, 1000	
hRT0057	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Cu2+[hc]	0, 1000	
hRT0058	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Molybdate[hc]	0, 1000	
hRT0059	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Fe3+[hc]	0, 1000	
hRT0060	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Putrescine[hc]	0, 0	
hRT0061	ATP[hc] + H2O[hc] -> ADP[hc] + Pi[hc] + Spermidine[hc]	0, 0	
hRT0062	5-Deoxy-D-ribose [hc] ->	0, 1000	
hRT0063	S-Methyl-5-thio-D-ribulose 1-phosphate[hc] ->	0, 1000	
hRT0064	CO[hc] ->	0, 1000	
hRT0071	H2O[hc] <=> H2O[hl]	-1000, 1000	
hRT0072	O2[hl] <=> O2[hc]	-1000, 1000	
hRT0073	Plastoquinone-9[hc] -> Plastoquinone-9[ht]	0, 1000	
hRT0074	H2O[hp] <=> H2O[hc]	-1000, 1000	

Rxn ID	Description	Bounds	Reference
hRT0075	O2[hp] <=> O2[hc]	-1000, 1000	
hRT0076	Plastoquinone-9[hc] -> Plastoquinone-9[hm]	0, 1000	

Table F-II. Intracellular exchange reactions in the *Anabaena* sp. PCC 7120 model

Rxn ID	Description	Bounds	Reference
hRX0001	Sucrose[vc] <=> Sucrose[hc]	-1000, 1000	(Nürnberg et al., 2015)
hRX0002	NH3[vc] <=> NH3[hc]	0, 0	
hRX0003	L-Glutamine[vc] <=> L-Glutamine[hc]	-1000, 1000	
hRX0004	L-Glutamate[vc] <=> L-Glutamate[hc]	-1000, 1000	
hRX0005	2-Oxoglutarate[vc] <=> 2-Oxoglutarate[hc]	-1000, 1000	
hRX0006	Maltose[vc] <=> Maltose[hc]	0, 0	
hRX0007	D-Glucose[vc] <=> D-Glucose[hc]	0, 0	
hRX0008	D-Fructose[vc] <=> D-Fructose[hc]	0, 0	
hRX0009	D-Glucose 6-phosphate[vc] <=> D-Glucose 6-phosphate[hc]	0, 0	
hRX0010	Glycerone phosphate[vc] <=> Glycerone phosphate[hc]	0, 0	
hRX0011	D-Glyceraldehyde 3-phosphate[vc] <=> D-Glyceraldehyde 3-phosphate[hc]	0, 0	
hRX0012	D-Fructose 1,6-bisphosphate[vc] <=> D-Fructose 1,6-bisphosphate[hc]	0, 0	
hRX0013	Phosphoenolpyruvate[vc] <=> Phosphoenolpyruvate[hc]	0, 0	
hRX0014	Pyruvate[vc] <=> Pyruvate[hc]	0, 0	
hRX0015	D-Ribulose 1,5-bisphosphate[vc] <=> D-Ribulose 1,5-bisphosphate[hc]	0, 0	
hRX0016	3-Phospho-D-glycerate[vc] <=> 3-Phospho-D-glycerate[hc]	0, 0	
hRX0017	2-Phosphoglycolate[vc] <=> 2-Phosphoglycolate[hc]	0, 0	
hRX0018	D-Fructose 6-phosphate[vc] <=> D-Fructose 6-phosphate[hc]	0, 0	
hRX0019	D-Erythrose 4-phosphate[vc] <=> D-Erythrose 4-phosphate[hc]	0, 0	
hRX0020	D-Ribose 5-phosphate[vc] <=> D-Ribose 5-phosphate[hc]	0, 0	
hRX0021	D-Ribulose 5-phosphate[vc] <=> D-Ribulose 5-phosphate[hc]	0, 0	
hRX0022	D-Xylulose 5-phosphate[vc] <=> D-Xylulose 5-phosphate[hc]	0, 0	
hRX0023	Sedoheptulose 7-phosphate[vc] <=> Sedoheptulose 7-phosphate[hc]	0, 0	

Rxn ID	Description	Bounds	Reference
hRX0024	Sedoheptulose 1,7-bisphosphate[vc] <=> Sedoheptulose 1,7-bisphosphate[hc]	0, 0	
hRX0025	D-Glucono-1,5-lactone 6-phosphate[vc] <=> D-Glucono-1,5-lactone 6-phosphate[hc]	0, 0	
hRX0026	6-Phospho-D-gluconate[vc] <=> 6-Phospho-D-gluconate[hc]	0, 0	
hRX0027	L-Alanine[vc] <=> L-Alanine[hc]	0, 0	
hRX0028	L-Histidine[vc] <=> L-Histidine[hc]	0, 0	
hRX0029	L-Ornithine[vc] <=> L-Ornithine[hc]	0, 0	
hRX0030	Cyanophycin[vc] <=> Cyanophycin[hc]	0, 0	

Appendix G Biomass objective function in the *Anabaena* sp. PCC 7120 model

Table G-I. Reaction equations for biomass components and the biomass objective function in both super-compartments.

Rxn ID	Description
ATP maintenance (non-growth associated):	
RB0001	ATP + H2O -> ADP + Pi
Protein fraction:	
RB0002	0.2318 L-Glutamine + 0.3349 L-Glutamate + 0.6371 L-Methionine + 0.4239 L-Serine + 0.2054 L-Aspartate + 0.4486 Glycine + 0.4459 L-Alanine + 0.2764 L-Lysine + 0.4004 L-Asparagine + 0.5348 L-Arginine + 0.4956 L-Tryptophan + 0.6753 L-Phenylalanine + 0.415 L-Tyrosine + 0.1863 L-Threonine + 0.0852 L-Cysteine + 0.39 L-Leucine + 0.2453 L-Histidine + 0.4378 L-Valine + 0.6802 L-Proline + 0.822 L-Isoleucine -> 8.3719 H2O + Biomass protein fraction
DNA fraction:	
RB0003	0.9682 dATP + 0.7149 dGTP + 0.9393 dTTP + 0.6232 dCTP -> 3.2456 PPi + Biomass DNA fraction
RNA fraction:	
RB0004	0.8377 ATP + 0.6 UTP + 1.0076 GTP + 0.6415 CTP -> 3.0868 PPi + Biomass RNA fraction
Cell wall fraction:	
RB0005	0.5182 Lipid A disaccharide + 0.5182 Peptidoglycan -> Biomass cell wall fraction
Lipid fraction:	
RB0006	0.1 Phosphatidylglycerol_PG + 0.265 Digalactosyl-diacylglycerol_DGDG + 0.255 Sulfoquinovosyldiacylglycerol_SQDG + 0.38 3-beta-D-Galactosyl-1,2-diacylglycerol_MGDG -> Biomass lipid fraction
Soluble pool fraction:	
RB0007	0.012 NADPH + 0.004 NADP+ + 0.01 Acetyl-CoA + 0.006 CoA + 0.008 Thiamin diphosphate + 0.008 Riboflavin + 0.002 NADH + 0.062 NAD+ + 0.008 Glutathione + 0.008 FAD + 0.008 S-Adenosyl-L-methionine + 0.008 Pyridoxal phosphate + 0.008 Heme + 0.003 Succinyl-CoA + 0.001 Malonyl-CoA + 0.008 Tetrahydrofolate + 0.008 10-Formyltetrahydrofolate + 0.008 5-Methyltetrahydrofolate + 0.008 Chorismate + 1.147 Putrescine + 0.008 5,10-Methenyltetrahydrofolate + 0.002 di-trans,poly-cis-Undecaprenyl diphosphate + 0.008 Cobamide coenzyme + 0.008 Heme O + 0.008 Heme A + 0.233 Spermidine -> Biomass soluble pool fraction
Inorganic ion fraction:	
RB0008	0.379 Pi + 0.379 Sulfate + 0.682 Fe2+ + 0.682 Fe3+ + 0.757 Mg2+ + 0.303 Co2+ + 1.136 NH4+ + 0.379 Na+ + 0.303 Mn2+ + 17.04 K+ + 0.303 Zn2+ + 0.454 Ca2+ + 0.303 Cu2+ + 0.303 Molybdate -> Biomass inorganic ion fraction
Pigment fraction:	
RB0009	0.1389 beta-Carotene + 0.8686 Chlorophyll a + 0.0237 (3R,2'S)-Myxol 2'-alpha-L-fucoside + 0.0406 (3S,2'S)-4-Ketomyxol 2'-alpha-L-fucoside + 0.1253 Echinenone + 0.0135 Canthaxanthin -> Biomass pigment fraction
Biomass objective function:	
RB0010	53.35 ATP + 53.35 H2O + 0.0341 Glycogen + 0.51 Biomass protein fraction + 0.031 Biomass DNA fraction + 0.17 Biomass RNA fraction + 0.0715 Biomass cell wall fraction + 0.12 Biomass lipid fraction + 0.029 Biomass soluble pool fraction + 0.01 Biomass inorganic ion fraction + 0.0244 Biomass pigment fraction -> 53.35 ADP + 53.35 Pi

Appendix H Experimental evaluation of carbon sources in mixotrophic growth

To compare differences between model predictions and that observed in real life, the growth rate with 13 different carbon sources (including bicarbonate) was evaluated in the laboratory under mixotrophic conditions. Cultivation was performed in BG-11 medium (Rippka et al., 1979) supplemented with the corresponding carbon source under constant illumination (60 µE m^{-2} s^{-1}) and elevated CO_2 (1 %), with nitrate as a combined nitrogen source. Growth was observed for up to 8 days by optical density measurements at 730 nm for culture density and at 440 nm (absorbance maximum of chlorophyll *a*) for cellular health. The two sets of growth curves acquired at the different wavelengths showed good correlation, and therefore only readings at 730 nm were evaluated thereafter (Figure H-1).

All substrates except for urea and putrescine sustained cyanobacterial growth at higher rates than bicarbonate (Figure H-1). Growth on urea was normal within the first 48 hours after which culture density started to decline. Furthermore, photosynthetic green cells detected at the absorption maximum of chlorophyll *a* (440 nm) disappeared over the next day. This may have been caused by the decomposition of carbamide into toxic compounds (cyanate and ammonium) or the highly shifted C/N ratio of the cell due to the relatively high nitrogen content of urea (Alexandrova and Jorgensen, 2007; Luque and Forchhammer, 2008). Even though some cyanobacteria express a cyanase to detoxify cyanate (Harano et al., 1997; Kamennaya and Post, 2011), enzymatic activity of the putative cyanate lyase All1291 in *Anabaena* sp. PCC 7120 has not yet been confirmed.

Putrescine, on the other hand, is reported to be highly toxic even at low levels such as that normally formed in microorganisms. The uncharged diamine putrescine can rapidly diffuse across membranes and get trapped and accumulated inside cells due to protonation. In the cyanobacterium *Anacystis nidulans*, intracellular concentration of putrescine has increased 600-fold over the endogenous level within the first few hours of exposure to 150 µM external putrescine (Guarino and Cohen, 1979). In this experiment putrescine was administered at 50-times higher concentration (7.5 mM) that may have easily led to the observed toxicity in *Anabaena* sp. PCC 7120.

The growth on pyruvate was similar to that observed for sugars, but the cells experienced a pronounced lag period extending over 48 hours. This may be related to a need to adapt metabolism (for gluconeogenesis) and/or metabolite uptake. In contrast, all the other substrates except for putrescine showed approximately a 24-hour lag phase (Figure H-1). The four sugars (glucose, fructose, maltose and sucrose), glycerol and acetate sustained growth rates at least 3-fold over the control bicarbonate in the exponential phase. Cultures growing on all substrates entered a stationary phase after approximately 6 days, including pyruvate. The highest growth rate and total biomass production was observed for glucose, immediately followed by glutamine and sucrose. Glutamate, however, appeared to be a less efficient nutrient for mixotrophic growth, probably due to its acidic character and/or because of a negative impact on nitrogen metabolism. For example, higher concentrations of glutamate may decrease the efficiency of the GS-GOGAT cycle and lead to elevated levels of 2-oxoglutarate, the effector molecule for C/N balance sensing (Forchhammer, 1999; Herrero et al., 2001). The other two amino acids, glutamine and proline are significantly better sources of carbon (and nitrogen) allowing approx. two-fold higher growth rate during the exponential phase. Both amino acids have been shown to suffice as the sole source of nitrogen for cyanobacteria (Luque and Forchhammer, 2008).

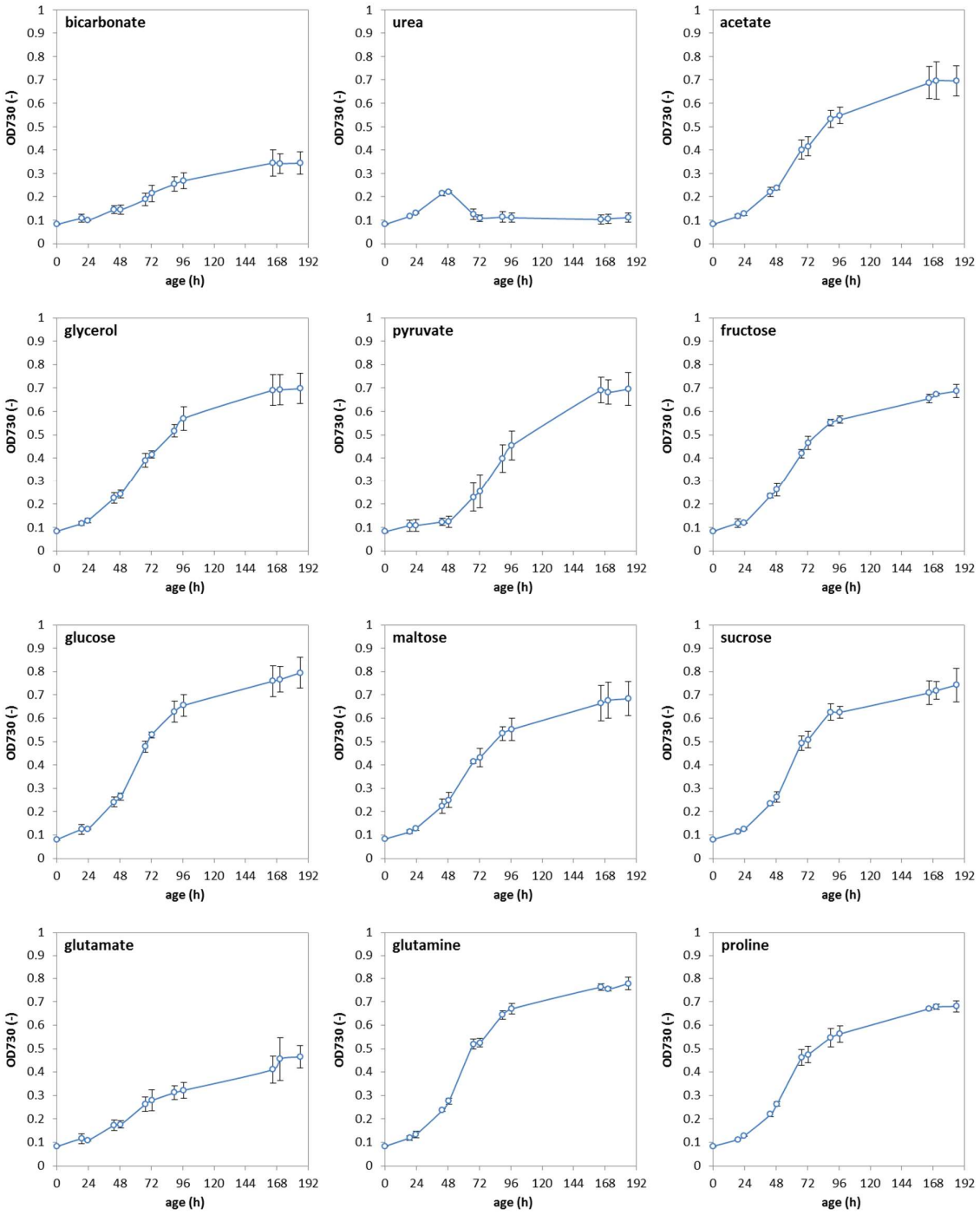


Figure H-1. Mixotrophic growth of *Anabaena* sp. PCC 7120 on single carbon sources and nitrate as nitrogen source. BG-11 medium was supplemented with 13 different substrates in molar equivalent to carbon contained in 5 mM glucose and growth was evaluated by optical density readings at 730 nm. Microbial health was monitored at the absorption maximum of chlorophyll *a* showing good correlation with OD_{730} (data not shown). Error bars: ±SD, n = 3.

Appendix I Visual representation of the two-cell model

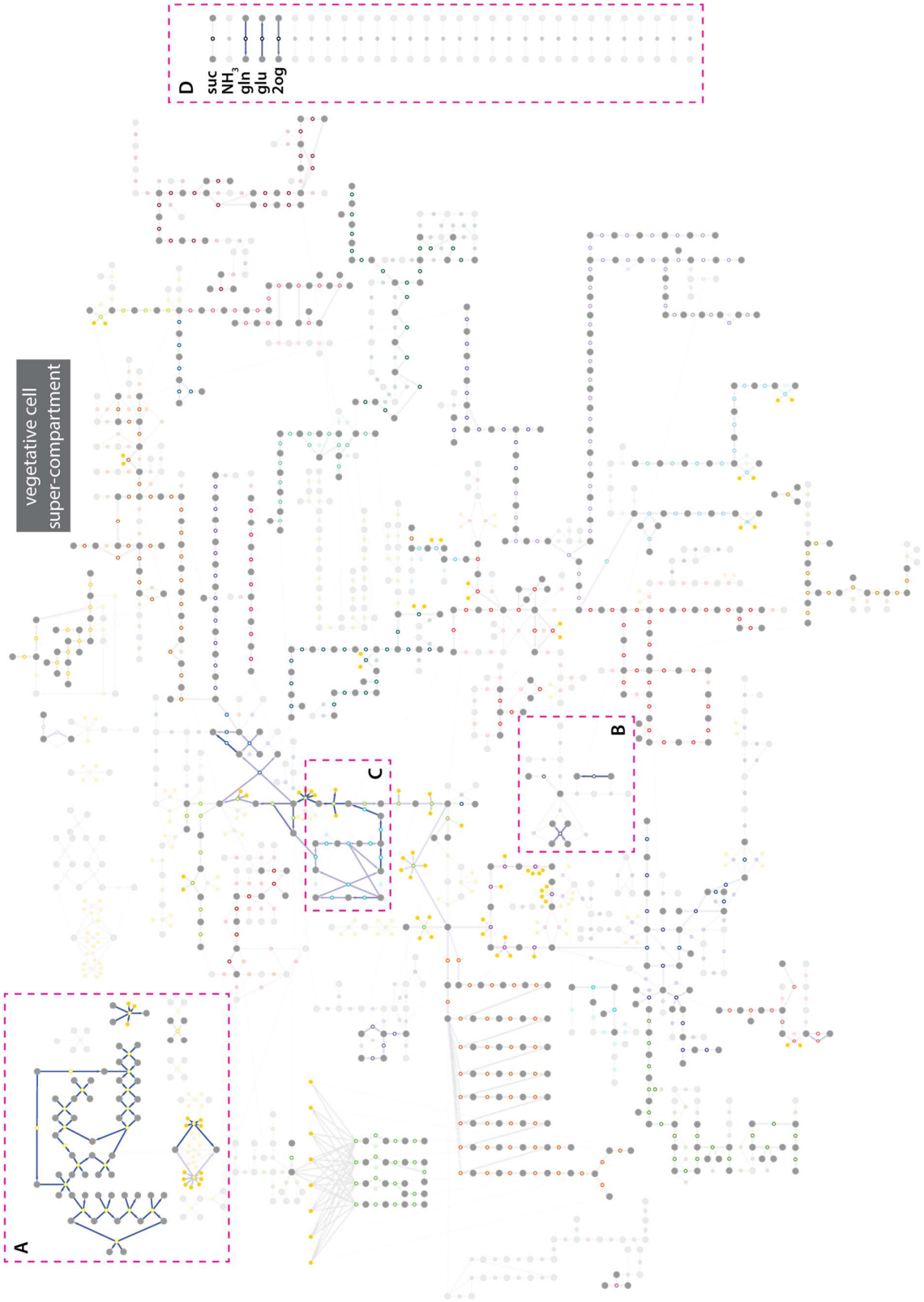




Figure I-1. Reaction network of the *Anabaena* sp. PCC 7120 two-cell model. Grey and orange nodes represent metabolites,coloured rings between metabolite nodes represent reactions and lines connecting nodes display reaction fluxes colour coded by a gradient between blue and light grey. Some important pathways are highlighted in both super-compartments: photosynthesis (A and D), nitrogen metabolism (B and F) and carbon fixation reactions (C and E). Exchange reactions (D) are also indicated. An interactive version of this map is available in Cytoscape format in Malatinszky et al. (2017).

Appendix J Permission to reproduce Malatinszky et al. (2017) in part or entirely

American Society of Plant Biologists LICENSE TERMS AND CONDITIONS

Aug 10, 2017

This is a License Agreement between David Malatinszky ("You") and American Society of Plant Biologists ("American Society of Plant Biologists") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Society of Plant Biologists, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	4165351266912
License date	Jul 27, 2017
Licensed content publisher	American Society of Plant Biologists
Licensed content title	Plant physiology
Licensed content date	Jan 1, 1998
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Number of pages in chapter/article	13
Title or numeric reference of the portion(s)	Entire work including figures, tables and supplementary data
Title of the article or chapter the portion is from	A Comprehensively Curated Genome-Scale Two-Cell Model for the Heterocystous Cyanobacterium Anabaena sp. PCC 7120
Editor of portion(s)	N/A
Author of portion(s)	David Malatinszky
Volume of serial or monograph.	N/A
Page range of the portion	509-521
Publication date of portion	Jan 2017
Rights for	Main product and any product related to main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	Up to 499
Made available in the following markets	Spiral repository, Imperial College London
Specified additional information	I would like to use the specified article, of which I am an author, in my PhD thesis at Imperial College London. I may need to perform

	minor editing of the text and figures. The requested material will be part of a larger work (a thesis).
The requesting person/organization is:	David Malatinszky
Order reference number	
Author/Editor	David Malatinszky
The standard identifier of New Work	not available
The proposed price	0
Title of New Work	Modelling and Engineering Anabaena sp. PCC 7120 for Nitrogen Excretion
Publisher of New Work	David Malatinszky
Expected publication date	Nov 2017
Estimated size (pages)	150
Total (may include CCC user fee)	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

The following terms are individual to this publisher:

Permission is granted for the life of the current edition and all future editions, in all languages and in all media.

Other Terms and Conditions:

ASPB grants to authors whose work has been published in Plant Physiology® or The Plant Cell the royalty-free right to reuse images, portions of an article, or full articles in any book, book chapter, or journal article of which the author is the author or editor. Reproductions must bear the full citation, the journal URL (www.plantphysiol.org or www.plantcell.org), and the following notice: "Copyright American Society of Plant Biologists." ASPB further grants to authors the permission to make digital or hard copies of part or all of a work published in Plant Physiology® or The Plant Cell without fee for personal or classroom use. This permission DOES NOT APPLY to figures/artwork credited to non-ASPB sources.

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in

these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. **Miscellaneous.**

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:
http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order

Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.