Design and engineering of a cyanobacterial photoautotrophic chassis and parts for Synthetic Biology applications

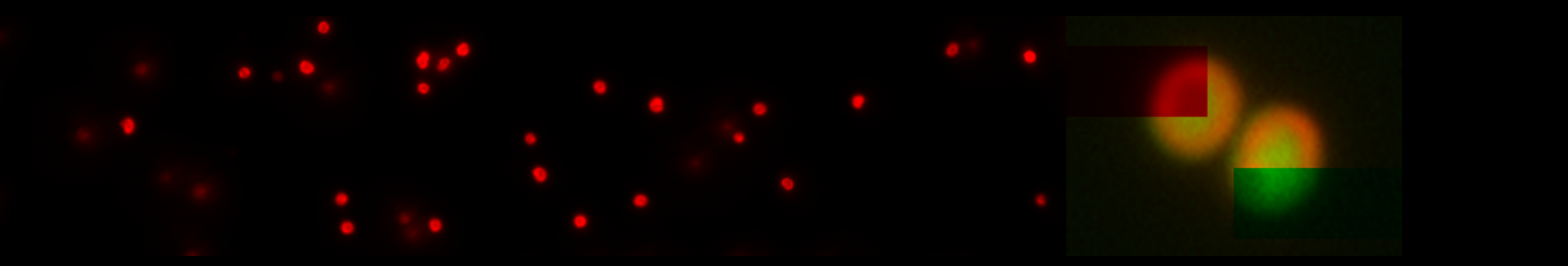
Filipe José Vilarinho Pinto

Departamento de Biologia

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Porto, 2012



Filipe Pinto

Design and engineering of a cyanobacterial photoautotrophic chassis and parts for Synthetic Biology applications

Faculdade de Ciências da Universidade do Porto
Departamento de Biologia



Porto 2012

Filipe José Vilarinho Pinto

Design and engineering of a cyanobacterial photoautotrophic chassis and parts for Synthetic Biology applications

Tese submetida à Faculdade de Ciências da Universidade do Porto para a obtenção do grau de Doutor em Biologia

Orientador

Paula Tamagnini

Professora Associada, Faculdade de Ciências da Universidade do Porto, Departamento de Biologia, Portugal

Co-orientador

Alfonso Jaramillo

Group leader, Institute of Systems and Synthetic Biology (ISSB), CNRS, France

2012

Jo sóc geòmetra, que vol dir sintètic.
(I am a geometrician, which is to say, synthetic.)

Antoni Gaudí

The creation continues incessantly through the media of man. But man does not create... he discovers. Those who look to the laws of Nature for support for their new works collaborate with the Creator. Copiers do not collaborate. Because of this, originality consists in returning to the origin.

Antoni Gaudí
(on the sleeve of 'A Design For Life')

Aos meus pais (sempre...)

Acknowledgements

This work was financially supported by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011), FCOMP-01-0124-FEDER-009003 (PTDC/BIA-MIC/100370/2008) and grants SFRH/BD/36378/2007, SFRH/BPD/64095/2009, SFRH/BPD/74894/2010, EU FP6-NEST-2005-Path-SYN project BioModularH_2 (contract no. 043340), and Programa PESSOA – FCT/DREBM Proc. 4.1.1 França.

I would also like to acknowledge several people that directly or indirectly contributed to this thesis:

To my supervisor, Prof. Paula Tamagnini, thank you very much for this opportunity. Thank you for your scientific guidance and critic evaluation of my work and results. Thank you for believing in my capabilities. Thank you for the chances of presenting our work all over the world, at places I would probably never travel to. It was an amazing scientific and personal experience. Thank you for your care and friendship over these last 10 years (only 4 of PhD ☺). It has been a long and not always easy path, and I'm honored to have you as mentor.

To my co-supervisor Alfonso Jaramillo, thank you for everything. It has been a pleasure to meet and work with you. Your enthusiastic visions of life and of science are unique. Thank for that wonderful month at your lab, and at Paris. It was special!

To Catarina Pacheco... a huge thank you! Thank for you precious help, for your friendship, for sharing your knowledge, for you patience, for sometimes being organized for me, for your precious help (I already said this ☺), for brainstorming my crazy ideas, for giving me strength, and I could continue... Thank you 'Cat. P.'.

To Paulo Oliveira, welcome back, and thanks again... for you teachings back in 2002, and for your teachings now. Thanks you your interest in my ideas and my work and for your help. Your dedication, motivation and good mood are an example for anyone.

Special thanks to Marta Mendes for your precious teachings at the bench, especially on the RNA world, and for your friendship.

To Paula Magalhães, for her precious help on the RT-qPCR experiments, availability, friendship and laughs, thank you very much.

To Tatiana Moutinho, a special thanks for introducing me to the fluorescence microscopy.

To the present and former members of the MicroBioSyn, MCA and ReCeS groups, thank you for your friendship and support over these years.

To everyone involved in the BioModularH_2 project (in random order): María Suárez, Daniel Camsund, Arnau Montagud, Thomas Landrain, Boris Kirov, Javier Carrera, Guillermo Rodrigo, Emilio Navarro, Héctor García, Javier Urchueguía, Peter Lindblad, Thorsten Heidorn, Adrian Hugenmatter, Meritxell Notari, Karin van Elburg, Phillip Wright, Josselin Noirel, *et al*. Thank you very much for you friendship, scientific sharing and fruitful conversations. To some of you: *viva la fiesta!*

To those random people (that will probably never ready this) that I met over and over again at scientific conferences, and that I hope to keep meeting, thanks for the good times spent together. Science is also about that.

To all my friends (too many to name you all), thank you for all the good moments spent together: drinks, parties, laughs, dinners, walks, movies, talks, games, hugs... It made possible to keep standing. Thank you for being there!

Special thanks to my colleagues and friends Daniela, Pedro and Zsófi for making the lab a comfortable place to be. The text above is also for you ;)

Thank *you* for being by my side! Thanks for all the friendship, support, care and special moments shared.

To all the IBMC staff, it has been a pleasure to know you and share so many laughs over these years. Thank you for your help and friendship, and for making bureaucracies much easier: the hardest of the tasks...

À D. Júlia, Bárbara e Tiago, muito obrigado pela sempre boa disposição, amizade, carinho e comidinha ;).

A toda a minha família mais próxima, um muito obrigado por todo o apoio e amizade. À minha irmã e ao meu cunhado muito obrigado pela “prenda” chamada Rodrigo. Acreditem que muito mudou em mim…

Finalmente, aos meus pais. Não tenho palavras para descrever a minha gratidão pelo vosso apoio moral (e económico!!! ☺), compreensão, preocupação e amor incondicional em todos estes anos de luta constante. Muito obrigado por me apoiarem *nesta* minha opção de vida!

Thank you all! / Obrigado a todos!

List of publications

(Ao abrigo do disposto no número 2 do Artigo 8º do Decreto-Lei nº 388/70)

This dissertation is based on the following publications, and some unpublished results:

I. **Pinto F.**, van Elburg K.A., Pacheco C.C., Lopo M., Noirel J., Montagud A., Urchueguía J.F., Wright P.C., Tamagnini P. (2012) "Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC 6803 mutant lacking a functional bidirectional hydrogenase" *Microbiology*, **158**: 448-464

II. **Pinto F.**, Pacheco C.C., Ferreira D., , Moradas-Ferreira P., Tamagnini P. (2012) "Selection of suitable reference genes for RT-qPCR analyses in cyanobacteria" *PLoS One*, **7**:e34983

III. **Pinto F.**, Pacheco C.C., Oliveira P., Montagud A., Urchueguía J.F. and Tamagnini P. (2012) "Characterization of neutral sites to use *Synechocystis* sp. PCC 6803 as a photoautotrophic chassis" (manuscript in preparation)

Other publications:

i. Ferreira D., **Pinto F**., Moradas-Ferreira P., Mendes M.V., Tamagnini P. (2009) "Transcription profiles of hydrogenases related genes in the cyanobacterium *Lyngbya majuscula* CCAP 1446/4" *BMC Microbiology*, **9**:67

ii. Fraga H., Faria T.Q., **Pinto F.**, Almeida A., Brito R.M., Damas A.M. (2010) "FH8 – a small EF-hand protein from *Fasciola hepatica*" *FEBS Journal*, **277**:5072-5085

Abstract

Synthetic biology applies the engineering principles of decoupling, standardization and hierarchical abstraction to design and construct new biological parts, devices and systems that do not exist in Nature, or to re-design natural biological systems for useful purposes. Synthetic devices are molecular constructs assembled from well characterized parts, such as promoters, regulators, coding regions, and terminators. The biological systems that are used to accommodate synthetic devices are defined as chassis, and should be well studied organisms, easy to engineer, minimal in terms of the subset of genes that will allow to retain viability and easy to grow in standard laboratorial conditions. The bacterium *Escherichia coli* has been extensively used as chassis, since it is one of the most well studied biological systems. Nonetheless, other prokaryotic and eukaryotic organisms have been successfully used as chassis, and among them cyanobacteria constitute promising "low-cost" microbial cell factories. Cyanobacteria are photosynthetic prokaryotes with simple nutritional requirements, high metabolic plasticity, and several tools are available for their genetic manipulation. The unicellular non-nitrogen-fixing *Synechocystis* sp. PCC 6803 is suitable to be used as a photoautotrophic chassis, since it is naturally transformable with foreign DNA, it has a medium-size genome and it is the best studied cyanobacterial strain with a vast amount of physiological and molecular data available that allowed the construction of genome-scale metabolic models.

In order to use *Synechocystis* as a chassis for hydrogen production, and since a heterologous hydrogenase was going to be introduced, a mutant was produced by deleting the two genes (*hoxY* and *hoxH*) encoding the hydrogenase part of the native bidirectional hydrogenase. The Δ*hoxYH* mutant, lacking an active bidirectional hydrogenase, was characterized at different levels: physiological, proteomic and transcriptional. The growth experiments, performed under different cultivation conditions, showed that the mutant exhibits growth rates similar to the wild-type. The proteomics studies were performed in conditions favoring hydrogenase activity and revealed that from the 210 identified proteins 17 had significant differential fold changes, comparing the mutant to the wild-type. Most of these proteins are related to the redox and energy state of the cell, supporting the role(s) previously attributed to the bidirectional hydrogenase. Transcriptional studies showed that only 6 genes encoding those proteins exhibited significant differences in transcript levels. Overall, this study revealed that the *Synechocystis* Δ*hoxYH* mutant is robust and can be used as a photoautotrophic chassis for the integration of synthetic devices.

For the stable integration of synthetic devices into *Synechocystis*, several chromosomal neutral sites were identified, matching the following criteria: (i) length of the putatively encoded proteins ≤301 amino acids, (ii) encoded proteins with no predicted transmembrane domains, (iii) no reported interaction with other proteins, (iv) no relevant similarities found at the protein sequence level, and (v) very low or undetectable transcription levels of those chromosomal sites. The neutrality of the sites was assessed by producing and analyzing mutants with each locus replaced by a kanamycin resistance cassette. The viability of the mutants was not affected and no distinguishable phenotype was observed. Moreover the neutrality of the sites was evaluated by producing deletion/insertion mutants and analyzing their growth under different cultivation conditions. The possibility of implementing a functional synthetic device in a neutral site was also evaluated using the reporter green fluorescence protein (GFP). *Synechocystis* mutants were produced, introducing a GFP-encoding synthetic device in each site. Green fluorescence was detected in all mutants by confocal microscopy, which confirmed the functionality of the sites. Moreover, all the integrative vectors constructed include a multiple cloning site (compatible with the BioBrick RFC[10] standard) insulated by transcription terminators. The robustness of this cloning interface was tested producing and analyzing mutants carrying a promoterless GFP-encoding synthetic device. Since it was not possible to detect green fluorescence in any of these mutants, it can be concluded that the interface is robust and there is no leakiness from backbone DNA sequences that could putatively act as promoter regions.

In this work, three different synthetic hydrogen producing devices were designed and constructed, all based on hydrogenases heterologous to *Synechocystis*. For this purpose, the [FeFe] hydrogenases from *Chlamydomonas reinhardtii* (HydA1 and HydA2) and from *Clostridium acetobutylicum* ATCC 824 (HydA) were fused to their primary electron donors (ferredoxins), through a glycine-rich peptide, in order to improve electron transfer between the two proteins. Three-dimensional models of two of these fusion proteins (HydA1:ferredoxin and HydA2:ferredoxin) were also constructed and may constitute important tools to improve enzymatic activity by protein engineering. Moreover, synthetic devices for the maturation of the hydrogenases were also designed and constructed to ensure the production of active enzymes. In addition, synthetic promoters and respective regulatory proteins were designed and constructed to modulate gene transcription. The promoter libraries are based on well characterized *E. coli* promoters and the transcription factors are composed of arrays of DNA binding domains of zinc finger proteins. Operator sequences of 12 bp can be specifically recognized by zinc finger arrays composed of 4 zinc finger domains. The binding of the zinc finger array to the operator, placed in the core

region of each promoter, would repress gene transcription by blocking the recognition of the promoter by the RNA polymerase holoenzyme via steric hindrance. Therefore, to prevent the interaction of the new synthetic transcription factors with the DNA from the chassis, only operator sequences that are not present in the genome of *E. coli* were selected. These synthetic promoter libraries and corresponding regulators (zinc finger arrays) will allow a more tight regulation of synthetic genetic circuits, and they can be used or redesigned to work in other organisms, namely *Synechocystis*.

The full characterization of synthetic parts and devices is required to predict systems behavior. The quantification of gene transcripts allows the characterization of synthetic parts, like promoters, transcription factors and terminators, and real-time quantitative polymerase chain reaction (RT-qPCR) is the standard technique used. However, to obtain reliable RT-qPCR results the use and validation of reference genes is mandatory. Towards this goal twelve candidate reference genes from three morphologically distinct cyanobacteria were selected and analyzed in cultures grown under routinely used cultivation conditions. The six genes exhibiting less variation in each organism were evaluated in terms of their expression stability using three available algorithms: geNorm, NormFinder and BestKeeper. In addition, the minimum number of reference genes required for normalization was determined. Based on the three algorithms, a list of genes for cyanobacterial RT-qPCR data normalization was elaborated, constituting a valuable starting point for future works.

Overall, the results obtained in this work demonstrate the potential of *Synechocystis* to be used as a photoautotrophic chassis in synthetic biology applications, and contributed to the field of synthetic biology by providing *Synechocystis* mutants prepared to receive synthetic devices. Moreover, BioBrick™-compatible vectors for the stable integration of those devices into novel chromosomal neutral sites were developed, and new parts and devices for synthetic biology purposes were designed and constructed. Furthermore, reference genes that can be used for normalization of transcription data in cyanobacteria were identified.

Resumo

A biologia sintética aplica os princípios de padronização, dissociação e abstração hierárquica da engenharia para desenhar e construir novas partes biológicas, módulos e sistemas que não existem na Natureza, ou para redesenhar sistemas biológicos naturais de modo a que estes desempenhem novas funções. Os módulos sintéticos são construídos a partir de partes bem caracterizadas, tais como promotores, reguladores, regiões codificantes, e terminadores. Os sistemas biológicos usados para acomodar os módulos sintéticos são designados *chassis*, e devem ser organismos bem estudados, fáceis de manipular geneticamente, com um conjunto diminuto de genes que permitam reter a viabilidade e fáceis de cultivar em condições normalmente usadas em laboratório. A bactéria *Escherichia coli* tem sido muito utilizada como *chassis*, uma vez que é um dos sistemas biológicos mais bem estudados. Contudo, outros organismos procariotas e eucariotas também têm sido utilizados com sucesso e, entre eles, destacam-se as cianobactérias que constituem promissoras "fábricas celulares", com custos reduzidos. As cianobactérias são procariontes fotossintéticos com requisitos nutricionais simples e elevada plasticidade metabólica, existindo várias ferramentas moleculares que tornam a sua manipulação genética possível e fácil de executar. A cianobactéria *Synechocystis* sp. PCC 6803 é um organismo unicelular não-fixador de azoto, sendo a estirpe mais estudada até hoje. A grande quantidade de dados fisiológicos e moleculares, um genoma de tamanho médio e os modelos metabólicos disponíveis à escala genómica, fazem desta cianobactéria um organismo adequado para ser utilizado como um *chassis* fotoautotrófico.

Tendo em vista a utilização de *Synechocystis* como *chassis* para a produção de hidrogénio, e tendo em conta a introdução de uma hidrogenase heteróloga, foi produzido um mutante em que foram eliminados dois genes (*hoxY* e *hoxH*) que codificam duas subunidades da hidrogenase bidirecional nativa. O mutante Δ*hoxYH*, cuja hidrogenase bidirecional não é funcional, foi caracterizado a nível fisiológico, proteómico e transcricional. As experiências de crescimento realizadas sob diferentes condições de cultura mostraram que o mutante tem um comportamento semelhante ao da estirpe selvagem. Estudos de proteómica, realizados em condições que favorecem a atividade da hidrogenase nativa, revelaram que das 210 proteínas identificadas, 17 tinham alterações significativas dos seus níveis, comparando o mutante com a estirpe selvagem. A maior parte destas proteínas estão relacionados com o estado energético e redox da célula, o que está de acordo com as funções previamente atribuídas à hidrogenase bidirecional de cianobactérias. Estudos de transcrição mostraram que apenas 6 dos

genes que codificam estas proteínas apresentam diferenças significativas nos níveis de transcritos. De uma forma geral, este estudo revelou que o mutante Δ*hoxYH* é robusto, podendo ser utilizado como *chassis* para a integração de módulos sintéticos.

Para a integração estável de módulos sintéticos em *Synechocystis*, foram identificados vários locais neutros no cromossoma, de acordo com os seguintes critérios: (i) tamanho das proteínas que poderiam codificar ≤301 aminoácidos, (ii) produtos génicos sem domínios transmembranares previstos, (iii) sem interações descritas com outras proteínas, (iv) sem semelhanças relevantes com outras proteínas a nível da sequência, e (v) níveis de transcrição muito baixos ou indetetáveis. A neutralidade dos locais selecionados foi avaliada através da produção e análise de mutantes em que a sequência original foi substituída por uma cassete de resistência à canamicina. Os mutantes foram produzidos sem que a sua viabilidade fosse afetada e não foi observado nenhum fenótipo que os distinguisse da estirpe selvagem. Foram feitas, ainda, análises comparativas do crescimento dos mutantes e da estirpe selvagem sob diferentes condições de cultura. A possibilidade de implementar um módulo sintético funcional num local neutro foi também avaliada usando como repórter a proteína verde fluorescente GFP (*green fluorescence protein*). Para este efeito foram produzidos mutantes de *Synechocystis* introduzindo em cada local neutro um módulo sintético para a expressão da GFP. Utilizando microscopia confocal, foi possível detetar fluorescência verde em todos os mutantes analisados, o que confirmou a funcionalidade dos locais. Todos os vetores integrativos construídos incluem um local de clonagem (compatível com o padrão RFC[10] do sistema BioBrick) isolado por terminadores de transcrição. A robustez deste interface de clonagem foi testada produzindo e analisando mutantes contendo o módulo sintético para a expressão da GFP, mas sem qualquer região promotora. Uma vez que não foi possível detetar fluorescência verde em nenhum destes mutantes, é possível concluir que o interface é robusto, evitando que possíveis regiões promotoras existentes nas sequências flanqueantes interfiram com os módulos sintéticos inseridos.

Neste trabalho, foram desenhados e construídos três módulos sintéticos para a produção de hidrogénio, tendo como base hidrogenases heterólogas a *Synechocystis*. Para este efeito, as hidrogenases de ferro ([FeFe]) de *Chlamydomonas reinhardtii* (HydA1 e HydA2) e de *Clostridium acetobutylicum* ATCC 824 (HydA) foram acopladas aos seus principais dadores de eletrões (ferredoxinas), através de um péptido rico em glicina, de forma a melhorar a transferência de eletrões entre as duas proteínas. Foram também construídos modelos tridimensionais de duas destas proteínas de fusão (HydA1: ferredoxina e HydA2: ferredoxina), que constituem ferramentas importantes para uma possível melhoria da atividade enzimática através da engenharia de proteínas. De forma a garantir que as

enzimas produzidas são ativas foram ainda desenhados e construídos módulos sintéticos contendo as proteínas de maturação destas hidrogenases.

Adicionalmente, foram desenhados e construídos promotores e respetivas proteínas reguladoras, para tornar possível a modulação da transcrição de genes. As bibliotecas de promotores são baseadas em promotores de *E. coli* bem caracterizados e os fatores de transcrição são compostos por vários domínios de ligação ao DNA de proteínas denominadas *zinc finger proteins*. Sequências operadoras de 12 nucleótidos podem ser especificamente reconhecidas por conjuntos de 4 domínios *zinc finger*. A ligação dos fatores de transcrição à sequência operadora colocada na região central de cada promotor (entre os elementos -10 e -35) reprime a transcrição de genes, bloqueando o reconhecimento do promotor pela RNA polimerase através de impedimento estérico. De forma a evitar a interação dos novos fatores de transcrição sintéticos com o DNA nativo do *chassis*, foram selecionadas sequências operadoras únicas e que não existem no genoma de *E. coli*. Estas bibliotecas de promotores sintéticos, juntamente com os respetivos reguladores (conjuntos de *zinc fingers*), vão permitir uma regulação mais precisa de circuitos genéticos sintéticos, podendo ser utilizados ou redesenhados para funcionar noutros organismos, nomeadamente em *Synechocystis*.

A caracterização precisa de partes e módulos sintéticos pode ser obtida pela quantificação dos níveis de um determinado transcrito (RNA mensageiro), sendo essencial para se poder prever o comportamento dos sistemas. A reação da transcriptase reversa, seguida de reação em cadeia da polimerase com deteção em tempo real (RT-qPCR) é a técnica padrão utilizada para a quantificação de transcritos. No entanto, para se obterem resultados de RT-qPCR fiáveis é necessária a utilização e validação de genes de referência. Assim sendo, doze possíveis genes de referência de três cianobactérias morfologicamente distintas foram selecionados e a sua transcrição analisada em células cultivadas em condições normalmente usadas no laboratório. A estabilidade dos seis genes que exibiram menor variação em termos de transcrição em cada organismo foi avaliada utilizando três algoritmos diferentes: NormFinder, geNorm e BestKeeper. O número mínimo de genes de referência necessários para a normalização foi também determinado. Com base nestes algoritmos, foi elaborada uma lista de genes para a normalização dos dados de RT-qPCR em cianobactérias, constituindo um valioso ponto de partida para trabalhos futuros.

De uma forma geral, os resultados obtidos neste estudo demonstram o potencial de *Synechocystis* para ser usado como um *chassis* fotoautotrófico em aplicações de biologia sintética, e contribuiu para esta área do conhecimento, disponibilizando mutantes de *Synechocystis* preparados para receber módulos sintéticos, desenvolvendo vetores

compatíveis com o sistema BioBrick para a integração estável de módulos em locais neutros identificados no cromossoma, desenhando e construindo novas partes e módulos para serem utilizadas em projetos de biologia sintética, e identificando genes de referência que podem ser utilizados para a normalização de dados em estudos de transcrição em cianobactérias.

Table of contents

ACKNOWLEDGEMENTS *vii*

LIST OF PUBLICATIONS *ix*

ABSTRACT *xi*

RESUMO *xv*

TABLE OF CONTENTS *xix*

LIST OF ABBREVIATIONS *xxiii*

CHAPTER I | GENERAL INTRODUCTION 1

I.1 Synthetic biology 3

I.1.1 Parts, devices and circuits 5

I.1.2 Chassis 6

I.1.3 Minimal genome, artificial cells and xenobiology 7

I.1.4 Synthetic biology applications and possibilities 8

I.1.5 Societal and ethical issues 9

I.2 The potential of cyanobacteria to be used as photoautotrophic chassis 10

I.3 Natural transformation and recombination in *Synechocystis* sp. PCC 6803 12

I.4 Tools for the design and characterization of synthetic parts, modules and circuits in cyanobacteria 15

I.4.1 Design and assembly of parts 15

I.4.2 Characterization and implementation of synthetic parts and devices 16

I.5 Aims of this work 20

CHAPTER II | CONSTRUCTION OF A CHASSIS FOR HYDROGEN PRODUCTION: PHYSIOLOGICAL AND MOLECULAR CHARACTERIZATION OF A *SYNECHOCYSTIS* SP. PCC 6803 MUTANT LACKING A FUNCTIONAL BIDIRECTIONAL HYDROGENASE 21

ABSTRACT 23

INTRODUCTION 23

METHODS 24
Organisms and standard growth conditions 24
DNA extraction and recovery 24
Plasmids construction 24
Generation of the *Synechocystis* sp. PCC 6803 *hoxYH* deletion mutant 25
Growth experiments 26
Statistical analysis of the specific growth 26
Sample collection, protein extraction and iTRAQ labeling 27
LC-MS and MS analysis 27
Pathway-level analysis of the proteomic data 27
Sample collection and RNA extraction for transcriptional studies 27
Transcription analysis by RT-qPCR 28
RESULTS AND DISCUSSION 28
Construction and physiological characterization of *Synechocystis* sp. PCC 6803 Δ*hoxYH* deletion mutant 28
Comparison of the proteomes of wild-type *Synechocystis* and of the Δ*hoxYH* mutant 29
Pathway analysis 32
Transcription analysis of genes encoding proteins with different abundance in *Synechocystis* wild-type and Δ*hoxYH* mutant 32
CONCLUSIONS 35
ACKNOWLEDGEMENTS 36
REFERENCES 36
SUPPLEMENTARY MATERIAL I 41
Supplementary methods 43
Supplementary tables 47
Supplementary figures 49
Supplementary references 53

CHAPTER III | CHARACTERIZATION OF NEUTRAL SITES TO USE *SYNECHOCYSTIS* SP. PCC 6803 AS A PHOTOAUTOTROPHIC CHASSIS 55
ABSTRACT 57
INTRODUCTION 57
RESULTS 59
Selection of putative neutral sites 59
Transcription analysis and final selection 59
Generation of mutants 62
Statistical analysis of cyanobacteria specific growth 65
Synthetic interface robustness and functionality of the neutral sites 67

DISCUSSION 69
MATERIALS AND METHODS 72
Organisms and maintenance culture conditions 72
Bioinformatics 72
DNA and RNA extraction 73
RT-PCR 73
Homologous regions amplification and fusion by PCR 74
Plasmid construction 74
Generation of *Synechocystis* mutants 75
Growth experiments 76
Statistical analysis of cyanobacteria specific growth 77
Confocal microscopy 77
Western blot 77
ACKNOWLEDGEMENTS 78
AUTHOR CONTRIBUTIONS 78
CONFLICT OF INTEREST 78
REFERENCES 79
SUPPLEMENTARY MATERIAL II 85
Supplementary figures 87
Supplementary tables 90

CHAPTER IV | DESIGN OF PARTS FOR SYNTHETIC BIOLOGY APPLICATIONS 97
IV.1 Design of a BioBrick-compatible cloning interface 99
IV.2 Design of a synthetic constitutive promoter 101
IV.3 Design of a synthetic hydrogen producing device (HPD) 102
IV.4 Development of orthogonal regulation networks 107
SUPPLEMENTARY MATERIAL III 111
Supplementary figures 113
Sequences 116

CHAPTER V | SELECTION OF SUITABLE REFERENCE GENES FOR RT-qPCR ANALYSES IN CYANOBACTERIA 123
ABSTRACT 125
INTRODUCTION 125
RESULTS AND DISCUSSION 126
Choice of candidate reference genes 126
Amplicon specificity, RT-qPCR efficiencies and analysis of Cq values 126

Assessing gene-stability and minimum number of reference genes required 126
Choice of reference genes 127
CONCLUSIONS 128
MATERIALS AND METHODS 128
Strains, culture conditions and sample collection 128
Candidate reference genes, primer design and amplicon specificity 130
RNA extraction and cDNA synthesis 130
Transcription analysis by Real-time RT-PCR 130
Analyses of gene stability and number of optimal reference genes 130
Choice of reference genes 131
MIQE guidelines 132
SUPPORTING INFORMATION 132
ACKNOWLEDGMENTS 132
AUTHOR CONTRIBUTIONS 132
REFERENCES 132
SUPPLEMENTARY MATERIAL IV 135
Supplementary figures 137
Supplementary tables 138

CHAPTER VI | GENERAL DISCUSSION AND FUTURE PROSPECTIVES 143
VI.1 Model organism 145
VI. 2 Preparation of a photoautotrophic chassis 145
VI.2.1 Characterization of *Synechocystis* Δ*hoxYH* mutant 145
VI.2.2 Chromosomal neutral sites: platforms for the integration of synthetic devices 148
VI.3 Design of orthogonal regulatory networks 150
VI.4 Normalization of transcription data 150
VI.5 Beyond the work developed 151
VI.6 Conclusions 151

GENERAL REFERENCES 153

List of abbreviations

μ – Specific growth

AES – Atomic emission spectrometry

AHL – Homoserine lactone

Amp – Ampicillin

ANOVA – Analysis of variance

asRNA – Antisense RNA

ATCC – American Type Culture Collection

BCA – Bicinchoninic acid

bp – Base pairs

CCAP – Culture Collection of Algae and Protozoa

CCY – Culture collection Yerseke

cDNA – Complementary DNA

CL – Continuous light

DNA – Deoxyribonucleic acid

dNTP – Deoxyribonucleotide

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

EPR – Electron paramagnetic resonance

ESR – Electron spin resonance

EU – European Union

Fd – Ferredoxin

FDR – False discovery rate

G+C – Percentage of guanine or cytosine bases on a DNA molecule

G0 – Medium not supplemented with glucose

G5 – Medium supplemented with 5 mM glucose

gDNA – Genomic DNA

GFP – Green fluorescent protein

HEPES – 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid

HPD – Hydrogen producing device

HR – Homologous region

IS – Insertion elements

iTRAQ – Isobaric tag for relative and absolute quantitation

kb – Kilobase pairs

kDa – Kilodaltons

Km – Kanamycin

Km^R – Resistance to kanamycin

LB – Lysogeny broth

LD – Light/dark

LL – Long linker

Mb – Megabase pairs

MCS – Multiple cloning site

MFE – Minimum free energy

mRNA – Messenger RNA

mt – Mutant

MW – Molecular weight

ncDNA – Noncoding DNA

NGS – Next-generation sequencing

OD – Optical density

OP – Operator sequence

ORF – Open reading frame

PCC – Pasteur culture collection

PCR – Polymerase chain reaction

Pf – BioBrick prefix

PNA – Peptide nucleic acid

RBS – Ribosome binding site

rDNA – Ribosomal DNA

RNA – Ribonucleic acid

RT – Reverse transcription

RT-PCR – Reverse transcriptase PCR

RT-qPCR – Quantitative real time PCR

Sc – BioBrick scar

SDM – Site directed mutagenesis

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sf – BioBrick suffix

SL – Short linker

Suc – Sucrose

Suc^S – Sensitivity to sucrose

TAE – Buffer solution containing a mixture of Tris base, acetic acid and EDTA

TBS – Tris buffered saline

TES – 2-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]ethanesulfonic acid

TF – Transcription factor

Tp – Transposase

Tris – 2-Amino-2-hydroxymethyl-propane-1,3-diol

TSP – Transcription start point

TT – Double transcriptional terminator

TTBS – TBS supplemented with Tween 20

WT/wt – Wild-type

XNA – Xeno nucleic acid

ZF – Zinc finger

ZFA – Zinc finger arrays

ZFP – Zinc finger proteins

CHAPTER I

General introduction



General introduction

I.1 Synthetic biology

Synthetic [sɪn'θɛtɪk]

adj.

1. Relating to, involving, or of the nature of synthesis.
2. *Chemistry* Produced by synthesis, especially not of natural origin.
3. a. Not natural or genuine; artificial or contrived: *"counterfeit rhetoric that flourishes when passions are synthetic" (George F. Will).*
   b. Prepared or made artificially: *synthetic leather*.
4. *Linguistics* Relating to or being a language, such as Latin or Russian, that uses inflectional affixes to express syntactic relationships.
5. *Logic & Philosophy* Relating to or being a proposition that attributes to a subject a predicate not inherent in the subject and that does not result in a contradiction if negated.

n.

A synthetic chemical compound or material.

[Greek sunthetikos, skilled in putting together, component, from sunthetos, combined, from suntithenai, to put together.]

in The American Heritage® Dictionary of the English Language (Hmc, 2000).

The term *synthetic* was applied for the first time to Biology by Stéphane Leduc (1910, 1912) in the sense that life could be synthesized through (organic) chemistry and explained by physics:

"*When trying to reproduce the substances or phenomena that we study, science has become synthetic*" (Leduc, 1910).

The term has evolved and in 1974 Wacław Szybalski, in view of the advances in the field of molecular biology, has given a new meaning to synthetic biology as the "next phase" of molecular biology:

"*But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with the unlimited expansion potential and hardly any limitations to building "new better control circuits" and ... finally other "synthetic" organisms, like a "new better mouse". ... I am not concerned that we will run out of exciting and novel ideas... in the synthetic biology, in general*" (Szybalski, 1974).

In 1980, Barbara Hobom used the term *synthetic* to describe bacteria that had been genetically engineered using recombinant DNA technology (Hobom, 1980), and in this context synthetic biology was essentially the same as genetic engineering: living systems modified by human intervention.

The major breakthrough in synthetic biology occurred in 2003 with the foundation of The Registry of Standard Biological Parts, at the Massachusetts Institute of Technology. This registry is not a mere database, but also a collection/repository of standardized genetic parts that can be assembled into devices and circuits (http://partsregistry.org).

Nowadays, synthetic biology stands as a new and emerging field of biology (http://syntheticbiology.org), aiming at designing and constructing new biological parts, devices and systems that do not exist in Nature, and re-designing natural biological systems for useful purposes (Fu, 2006; Heinemann and Panke, 2006; Agapakis and Silver, 2009). Applying the engineering principles of decoupling, standardization and abstraction hierarchy to biological systems, synthetic biology aims at shaping biology into a predictable discipline, where design principles govern rational engineering projects (Endy, 2005). Decoupling and standardization are the principles applied in order to design and construct biological parts (BioBricks™) that should be fully characterized (Canton *et al.*, 2008), and similarly to electronics, these parts can be assembled into more complex modules (devices and circuits). Then, the synthetic modules can be introduced into a biological chassis and interact with each other (Figure 1). The design of synthetic circuits follows hierarchical layers of increasing complexity, and abstraction allows designers to focus on one layer of biological complexity at a time. Thereby, biological systems complexity is reduced, biological functions standardized, and diversity leveled.

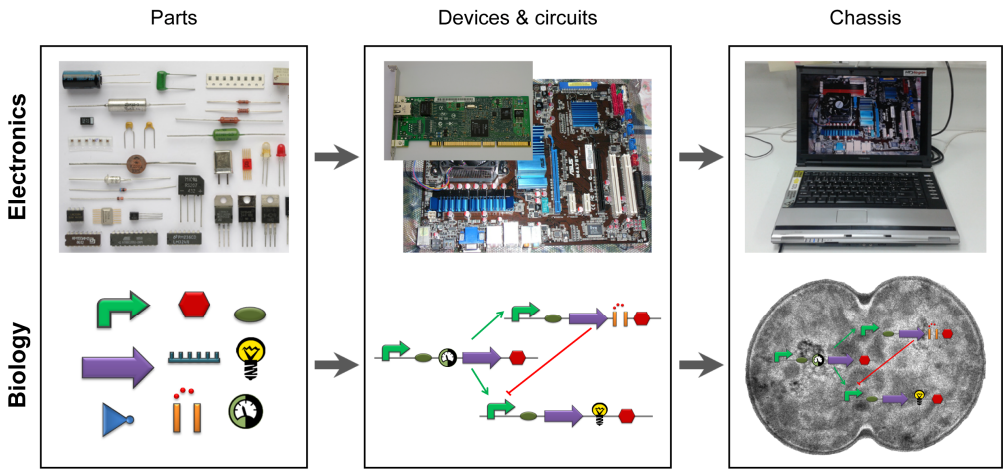


Figure 1 | Analogy between electronics and synthetic biology. Similarly to electronics, individual genetic elements (parts) can be assembled into more complex devices and circuits. Ultimately, these modules are introduced into a chassis to perform a designated task.

The advances in synthetic biology are mainly due to the existence of vast genomic and proteomic databases, the advances in the areas of nanotechnology, systems biology, bioinformatics and mathematical modeling, and to developments in DNA synthesis technology, which will make possible to synthesize ever larger DNA fragments at reduced costs (Figure 2). Unlike classical genetic engineering, synthetic biology allows to (re)write the genetic code at the nucleotide level, making possible to optimize a gene sequence, improving, for example, protein expression in a heterologous system (Tang *et al.*, 2011; Hutterer *et al.*, 2012; Mellitzer *et al.*, 2012; Sarduy *et al.*, 2012), or to design and synthesize new genes and gene networks (Clancy and Voigt, 2010; Liang *et al.*, 2011; Nandagopal and Elowitz, 2011) and, ultimately, entirely new genomes (Gibson *et al.*, 2008; Carrera *et al.*, 2009; Gibson *et al.*, 2010; Wang, 2010). The possibility of designing new genetic circuits from scratch, based on fully characterized DNA parts with predictable behavior, allows developing and implementing more efficient strategies, shortening the period of time necessary to produce genetically modified organisms.

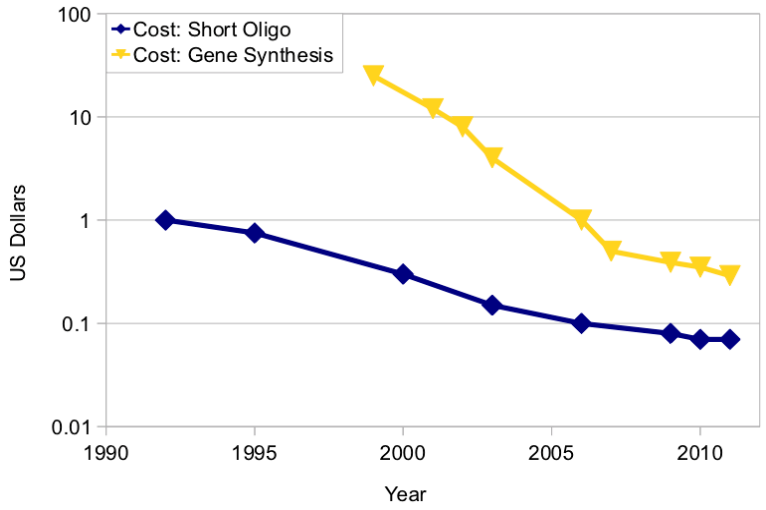


Figure 2 | Cost per base of synthetic DNA. The costs of synthesis of DNA fragments are decreasing over the years (short oligonucleotides in blue and larger sequences in yellow), due to developments in DNA synthesis technology. (Adapted from Rob Carlson, June 2011, www.synthesis.cc).

I.1.1 Parts, devices and circuits

A biological part is a DNA sequence that follows the standards of synthetic biology and can be, for instance, a promoter, a ribosome binding site (RBS), a sequence encoding a protein or a protein domain, or a transcription terminator. Since decoupling implies that each individual part is independent of any other and that its behavior is fully predictable, parts characterization is a requisite in synthetic biology (Boyle and Silver, 2009; Ellis *et al.*, 2009; Purnick and Weiss, 2009; Huang *et al.*, 2010). The assembly standards of BioBricks were designed in such way to make possible to assemble any parts with simple *cut &*

paste DNA molecular biology methodologies. Since there are several standards, additional efforts have been made to develop plasmids compatible with the BioBrick assembly system (Shetty *et al.*, 2008). The assembly standards are defined by the use of a prefix and a suffix (flanking each part), containing unique restriction sites that allow the assembly of two or more parts (Figure 3). Consequently, new longer and more complex parts can be constructed, while still maintaining the structural elements of the assembly standard. In this way, any newly composed parts will become themselves a part with the same assembly standard specifications that can be used in future assemblies. When a complex part has a defined output, such as the production of a protein or the transcription of an interference RNA, it is defined as a device. Moreover, synthetic devices can be unrelated to each other, or they can interact constituting a synthetic circuit.

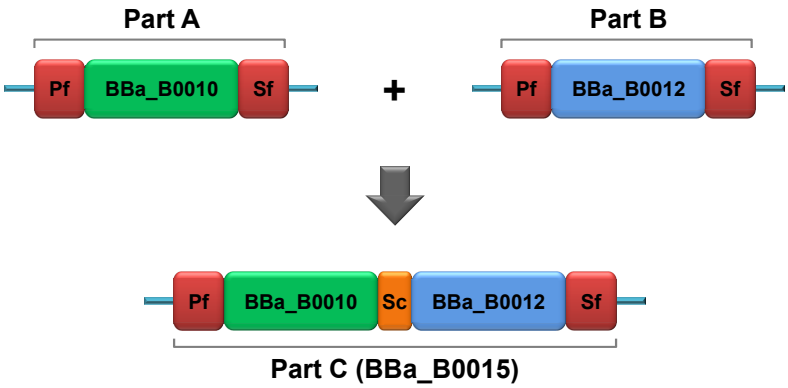


Figure 3 | Structure and specifications of BioBricks. BioBricks are standardized DNA sequences (parts) flanked by a prefix (Pf) and a suffix (Sf) containing unique restriction sites that allow the simple assembly of two or more parts. Composite parts maintain the standards specifications (the flanking prefix and suffix) and a "scar" (Sc) resulting from the assembly may remain between the parts.

I.1.2 Chassis

Eventually, the synthetic parts/devices must be integrated into a cell/organism to perform the designated task they were design to accomplish – the chassis. Model organisms to be used as chassis should be well studied, with high throughput genomic and proteomic data available, easy to engineer with the available molecular tools, minimal in terms of the subset of genes that will allow to retain viability and easy to grow in standard laboratorial conditions (ETC_Group, 2007; de Lorenzo, 2011). The Gram-negative bacterium *Escherichia coli* has been the most commonly used chassis in biotechnological applications, but other organisms are emerging, such as the Gram-positive bacterium *Bacillus subtilis* (Griffith and Grossman, 2008; Itaya *et al.*, 2008; Morimoto *et al.*, 2008; e.g. Nielsen *et al.*, 2009; Singh *et al.*, 2009; Rojas Contreras *et al.*, 2010; Li *et al.*, 2011b; Yamaoka *et al.*, 2011; Baek *et al.*, 2012; Biswas *et al.*, 2012; Jia *et al.*, 2012), the

eukaryotic yeast *Saccharomyces cerevisiae* (e.g. Cantone *et al.*, 2009; Hwang *et al.*, 2009; Hou, 2010; Ohto *et al.*, 2010; Yu *et al.*, 2010; Zelle *et al.*, 2010; Albertsen *et al.*, 2011; Tavares *et al.*, 2011; Kim *et al.*, 2012; Kondo *et al.*, 2012; Martínez *et al.*, 2012; Zhou *et al.*, 2012) and even partially synthetic bacteria like *Mycoplasma laboratorium* (derived from *M. genitalium*) (Reich, 2000), a project that is being developed at the J. Craig Venter Institute (http://www.jcvi.org/). Chassis can also be cells deprived of their own genetic material, in order to receive completely synthetic chromosomes (Gibson *et al.*, 2010) or even totally artificial cells (Solé *et al.*, 2007; Jewett and Forster, 2010; Noireaux *et al.*, 2011).

I.1.3 Minimal genome, artificial cells and xenobiology

The ultimate aim of synthetic biology would be a self-replicating chassis with a totally artificial chromosome, having the minimal number of proteins possible, and inside of which synthetic pathways could be built without cross-talk or interference from the natural metabolism (Gibson *et al.*, 2010; Jewett and Forster, 2010). Therefore, a minimal genome can constitute a totally predictable and controllable system, being advantageous for the construction of biological chassis orthogonal (*i.e.* context independent) to the synthetic modules implemented (Klasson and Andersson, 2010). The concept of a minimal genome can be defined as a genome possessing the minimal necessary elements that would allow the survival of an organism under specific conditions (Mushegian, 1999; DeWall and Cheng, 2011; Juhas *et al.*, 2011). Understanding the genomic structure and function, as well as the cellular dynamics, is crucial for designing a useful synthetic organism (Cavalier-Smith, 2001; A. O'Malley *et al.*, 2008; Murtas, 2009; Walde, 2010; Shuler, 2012). Such organism should maintain metabolic homeostasis, undergo cellular reproduction, and possess Darwinian evolution capabilities (Pohorille and Deamer, 2002; Rasmussen *et al.*, 2004; Loakes and Holliger, 2009). The design and construction of an artificial cell from simpler artificial precursors could revolutionize biotechnological processes and technologies (Forster and Church, 2007; Foley and Shuler, 2010). Going from quasi-cellular systems (Roodbeen and van Hest, 2009; Walde, 2010; Stano, 2011), to cell-free molecular translation systems (the PURE system) (Shimizu *et al.*, 2001), numerous advances are being made towards the *synthetic cell*. Moreover, engineered synthetic modules should be made orthogonal in respect to existing counterparts, decreasing the possibility of creating emergent properties (de Lorenzo, 2010a). In this context, research is being made towards the expansion of the genetic code by using four-based codons for protein synthesis, increasing the number of codons from 64 to 256 (Magliery *et al.*, 2001), or by the development of molecules capable of mimicking nucleic acids, with the capability of storing genetic information and replicate. The latter case counts with the development

of new nucleotides (termed XNA for *xeno* nucleic acids) and new DNA peptide-based backbones (termed PNA for *peptide* nucleic acid) (Schmidt, 2010). Moreover, synthesis and incorporation of non-canonical amino acids has also been achieved by redesigning several components of the translational machinery (Xie and Schultz, 2005). These new molecules, at least in theory, are safer in terms of biological risk contention (biosafety) and could be used in the future in synthetic biology approaches.

I.1.4 Synthetic biology applications and possibilities

Using synthetic biology principles to (re)design, (re)write and (re)model biological systems makes virtually possible to "create" anything from scratch. Ranging from fundamental to more applied purposes, the use o synthetic biology is only limited by imagination and biological unpredictability. Some examples of synthetic biology applications on different scientific areas will be subsequently presented:

(*i*) The design and construction of a synthetic oscillating network (the "repressilator"), using three transcriptional repressor systems that are not part of any natural biological clock. The implementation of this network in *E. coli* periodically induces the synthesis of green fluorescent protein (GFP), resulting in oscillations with typical periods of hours. Despite the noisy behavior of this artificial clock, which may be a consequence of the stochastic fluctuations of its components, this work shows that it is possible to design and construct artificial genetic networks with new functional properties, from generic components that naturally occur in other contexts (Elowitz and Leibler, 2000).

(*ii*) The synthetic metabolic engineering of *S. cerevisiae* to produce the immediate precursor of the antimalarial drug artemisinin (artemisinic acid), by using an engineered mevalonate pathway, and introducing the amorphadiene synthase and a novel cytochrome P450 monooxygenase, both from *Artemisia annua*. This engineered yeast is able to produce artemisinic acid at a significantly higher specific productivity than *A. annua*, and yield optimization and industrial scale-up will reduce production costs and increase drug quality (Ro *et al.*, 2006).

(*iii*) The development of synthetic gene switches responding to stimuli, within the physiological limits, to be used in combination with biomaterials like biohybrid hydrogels. These new stimulus-sensitive materials can be used *in vivo* to release a therapeutic growth factor for controlling cell fate and function in a time- and space-resolved manner (Kämpf and Weber, 2012).

(*iv*) The use of multilamellar liposomes loaded with a bacterial *in vitro* transcription and translation system together with gene constructs encoding specific antigens for

vaccination. These antigen-expressing immunostimulatory liposomes (AnExILs) represent a new platform for DNA-based vaccines combining antigen production, adjuvanticity and delivery (Amidi *et al.*, 2011).

(*v*) The implementation of self-regulatory protein production systems, coupling the control of recombinant protein production to a stress-induced negative feedback mechanism. The introduction of such self-regulatory synthetic circuits may increase the soluble/insoluble ratio of recombinant proteins, avoiding significant stress in the host organism (Dragosits *et al.*, 2012).

(*vi*) The production of biofuels, as an alternative to the limited fossil fuel resources. Synthetic biology approaches are being used to produce biohydrogen (Johnson and Schmidt-Dannert, 2008; Agapakis and Silver, 2010), alcohols, isoprenoids and lipids (Connor and Atsumi, 2010; Lindberg *et al.*, 2010; Colin *et al.*, 2011), and even to increase microbes tolerance to next-generation biofuels (Dunlop, 2011).

Furthermore, next-generation sequencing (NGS) technologies allow the inexpensive production of large volumes of sequence data (Metzker, 2010), even from microorganisms that can not be cultivated in isolation (Schloss and Handelsman, 2005; Kwon *et al.*, 2011). New genes can, therefore, be identified and the development of sophisticated bioinformatics analysis tools for genome mining will allow novel protein functions and new metabolic pathways to be predicted, increasing synthetic biology possibilities at a tremendous rhythm.

I.1.5 Societal and ethical issues

Going from agriculture (Orzaez *et al.*, 2010), pharmaceutics (Chang and Keasling, 2006), medical/therapeutic strategies (Aubel and Fussenegger, 2010; Shankar and Pillai, 2011), bioremediation (Looger *et al.*, 2003; Porcar and Moya, 2010), to the energy field (Fortman *et al.*, 2008; Waks and Silver, 2009), synthetic biology potential is endless, and, in this sense, this discipline raises societal and ethical issues like biosafety, intellectual property rights, regulation and legal framework (de Lorenzo, 2010b). Recently, an updated and expanded version of a guide reviewing how existing regulations apply to the emerging field of synthetic biology has been produced (Bar-Yam *et al.*, 2012). The authors conclude that weighting the costs and benefits of each rule is the challenge faced by regulators, since burdensome regulations impede legitimate research and positive advances in synthetic biology. Also in Europe, several documents have been published regarding the potential and use of synthetic biology. For instance, The European Group on Ethics in Science and New Technologies has issued an Opinion (No. 25) on the ethical, legal and social implications that may derive from synthetic biology, entitled *Ethics in Synthetic*

Biology. The Portuguese National Council of Ethics for the Life Sciences in collaboration with the Bioethics Committee of Spain also issued an Opinion on Synthetic Biology (61/CNECV/2011), which reflects the global impact and concerns regarding this field.

I.2 The potential of cyanobacteria to be used as photoautotrophic chassis

For biotechnological applications, heterotrophic organisms like *E. coli* have been preferably used, but the need to supplement the medium with carbon sources increases production costs. Therefore, the use of autotrophic organisms as photosynthetic prokaryotes (Abed *et al.*, 2009; Huang *et al.*, 2010; Skizim *et al.*, 2011; Subashchandrabose *et al.*, 2011; Xu *et al.*, 2011; Keskin and Hallenbeck, 2012) and eukaryotes like microalgae (Xu *et al.*, 2006; Wijffels, 2008; Li *et al.*, 2011a; Malcata, 2011; Yan *et al.*, 2011) is preferable. In this context, cyanobacteria emerge as promising organisms to become “low-cost” cell factories, mainly due to their minimal nutritional requirements (photosynthetic and in some cases nitrogen fixers) and cellular simplicity (prokaryotes). These organisms are capable of synthesizing several photosynthetic pigments, such as chlorophyll *a* and/or *b*, and are able to perform oxygenic photosynthesis, using water as electron donor. The ancestry of cyanobacteria is well documented and fossil traces have been found in sedimentary rocks formed 3500 million years ago (Schopf, 2006). It is generally believed that cyanobacteria were the first major group of phototrophs to arise with the ability to oxidizing water to produce molecular oxygen that caused, in the Precambrian Era, the transition in the Earth's atmosphere from its primordial anaerobic state to its current aerobic condition (Schopf, 2000). Cyanobacteria constitute one of the largest and more important groups of bacteria on Earth and, probably due to their long evolutionary history and metabolic plasticity, are found in almost any ecological niche, going from salt to fresh water, to soil and even to extreme environments (Whitton and Potts, 2000). Some cyanobacterial strains have the ability to fix atmospheric nitrogen, contributing significantly to the input of combined nitrogen in nitrogen-poor soils and oceans (Vitousek *et al.*, 2002; Zehr and Ward, 2002). The ecological significance of these organisms goes beyond being primary producers, since many cyanobacteria can modify their habitats through the synthesis of biologically active natural products, including antibiotics, toxins, iron-chelating siderophores and immunosuppressants (Ehrenreich *et al.*, 2005). Moreover, cyanobacteria are unique in their capacity to form symbiotic associations with a wide range of eukaryotic organisms, including plants, fungi, algae, protists and even animals, providing combined nitrogen and occasionally carbon to the host (Adams, 2002).

Cyanobacteria exhibit a large morphological diversity, including unicellular, colonial and filamentous forms (Figure 4) (Castenholz, 2001; Berman-Frank *et al.*, 2003). Some filamentous cyanobacteria evolved a degree of cell differentiation, producing specialized cells where nitrogen fixation occurs (heterocysts), characterized by the presence of additional cell wall layers and modified thylakoids lacking a functional photosystem II, and resting cells, that are able to endure environmental stress (akinetes) (Tomitani *et al.*, 2006). Only recently, cyanobacteria have been recognized as a potent source for numerous biologically active natural products (Falch *et al.*, 1995; Burja *et al.*, 2001; Tan, 2007; Berry *et al.*, 2008; Blunt *et al.*, 2009; Esquenazi *et al.*, 2011). To date, about 800 molecules of cyanobacterial origin are known, including pharmacologically interesting compounds, such as substances with anticancer, antimicrobial and hypertension lowering activities (Moore, 1996; Schaeffer and Krylov, 2000; Takamatsu *et al.*, 2003; Dembitsky, 2006; Devillers *et al.*, 2007; Simmons *et al.*, 2008). Overall, cyanobacteria not only play key roles in the global earth cycles of carbon and nitrogen, but they are becoming increasingly important economically.

In the last years, the number of cyanobacterial genomes sequenced and high-throughput genomics and proteomics studies increased significantly, making possible to integrate these data in ever more accurate genetic-to-metabolic computer models. As result it is possible to use these models to predict the behavior of biological systems to changes (e.g. the introduction of a synthetic module) (Marchisio and Stelling, 2008; Bentley, 2009;

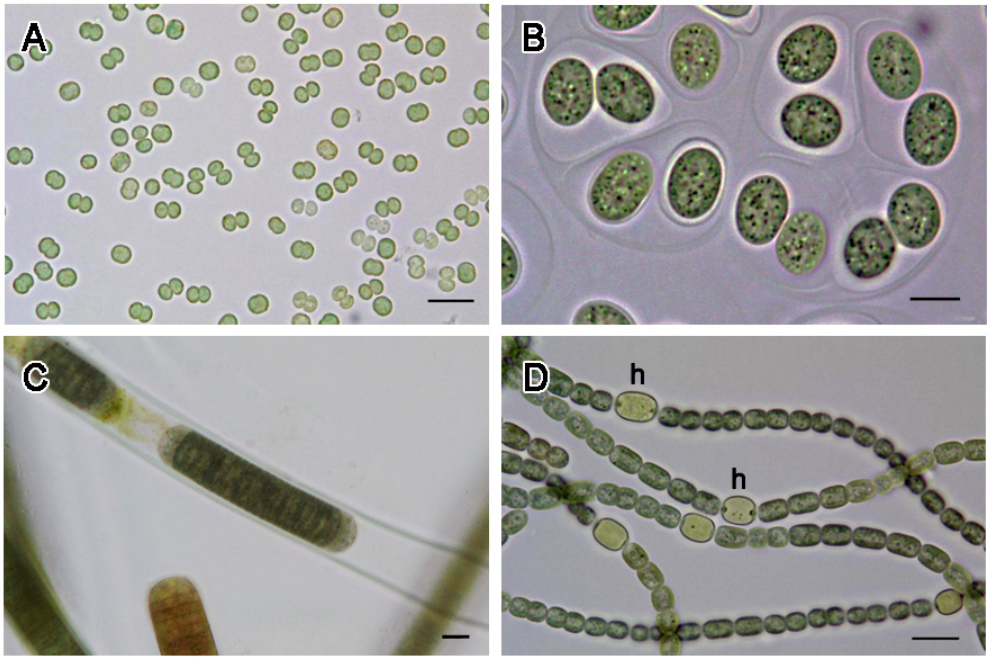


Figure 4 | Cyanobacteria morphological diversity. (A) *Synechocystis* sp. PCC 6803 (unicellular); (B) *Gloeothece* sp. PCC 6909 (unicellular, colonial); (C) *Lyngbya majuscula* CCAP 1446/4 (filamentous, non-heterocystous); (D) *Nostoc punctiforme* ATCC 29133 (filamentous, heterocystous; h – heterocyst). Bars: 5 µm.

Endler *et al.*, 2009; Marchisio and Stelling, 2009; MacDonald *et al.*, 2011; Marchisio, 2012; Medema *et al.*, 2012). Nonetheless, behaviors predicted *in silico* have to be validated *in vivo*, which requires the development of a toolbox for synthetic biology applications using cyanobacteria.

The best studied cyanobacterium is the unicellular non-N_2-fixing *Synechocystis* sp. PCC 6803 (henceforth referred to as *Synechocystis*), since its genome was the first to be sequenced among photosynthetic organisms (Kaneko *et al.*, 1996b; Kaneko *et al.*, 2001; Kaneko *et al.*, 2003). As consequence, numerous studies using this cyanobacterium were performed over the last years, generating a large amount of genomic, proteomic and physiological information, as well as well defined metabolic profiles (Fulda *et al.*, 2006; Huang *et al.*, 2006; Suzuki *et al.*, 2006; Sato *et al.*, 2007; Ow and Wright, 2009; Mitschke *et al.*, 2011; Lopo *et al.*, 2012). These studies also allowed the construction of genome-scale metabolic models of *Synechocystis* (Fu, 2009; Montagud *et al.*, 2010; Montagud *et al.*, 2011), which are important tools to predict systems responses to changes. *Synechocystis* can grow both autotrophically or mixotrophically and, as most cyanobacteria, has simple nutritional requirements, which makes it a proper candidate for biotechnological applications. This cyanobacterium possesses a medium-size genome, comprising a 3.6 Mb chromosome (Kaneko *et al.*, 1996a) and 7 plasmids (from 2.3 kb to 120 kb) (Yang and McFadden, 1993, 1994; Xu and McFadden, 1997; Kaneko *et al.*, 2003), it is naturally transformable with exogenous DNA, having a natural tendency towards homology-directed recombination (Williams, 1988; Kufryk *et al.*, 2002; Zang *et al.*, 2007a). In addition, there are plenty of molecular tools available for its genetic manipulation (Vermaas, 1996; Koksharova and Wolk, 2002) and several synthetic biology tools have already been developed for *Synechocystis*, as well as for other cyanobacteria (Huang *et al.*, 2010; Heidorn *et al.*, 2011).

I.3 Natural transformation and recombination in *Synechocystis* sp. PCC 6803

As stated above, several molecular tools were developed to genetically manipulate *Synechocystis*, and transformation is a limiting step to successfully produce deletion mutants and to introduce synthetic devices. This is particularly important in cyanobacteria, since their doubling time is considerably higher than that of other bacteria like *E. coli*, and because many cyanobacteria possess multiple copies of the chromosome per cell (Herdman *et al.*, 1979; Schneider *et al.*, 2007; Watanabe *et al.*, 2012).

Similarly to other bacteria, *Synechocystis* is naturally transformable with exogenous DNA

(Grigorieva and Shestakov, 1982; Stevens and Porter, 1986; Williams, 1988; Yura *et al.*, 1999). DNA uptake is mediated by Type IV pili (Yoshihara *et al.*, 2001; Okamoto and Ohmori, 2002; Yoshihara *et al.*, 2002; Nakasugi *et al.*, 2006) and the incorporation into the genome is achieved by homologous recombination (Williams, 1988; Labarre *et al.*, 1989), allowing both the generation of deletion/disruption mutants and the integration of genes/genetic circuits. Single integrative crossover is a phenomenon with very low frequency, occurring rarely in *Synechocystis*, and recombination occurs by means of double homologous recombination (Figure 5). In this sense, a heterologous DNA fragment, such as an antibiotic resistance cassette, can be introduced into *Synechocystis* if it is flanked on both sides by DNA sequences homologous to the chromosome (Williams, 1988; Labarre *et al.*, 1989). Deletion mutants carrying no selection cassettes can also be obtained in two transformation steps using positive and negative selection markers: (*i*) the target gene is replaced by a double selection cassette, selecting for the positive marker, and (*ii*) the cassette is removed or replaced by other gene(s), selecting for the negative marker (Figure 5).

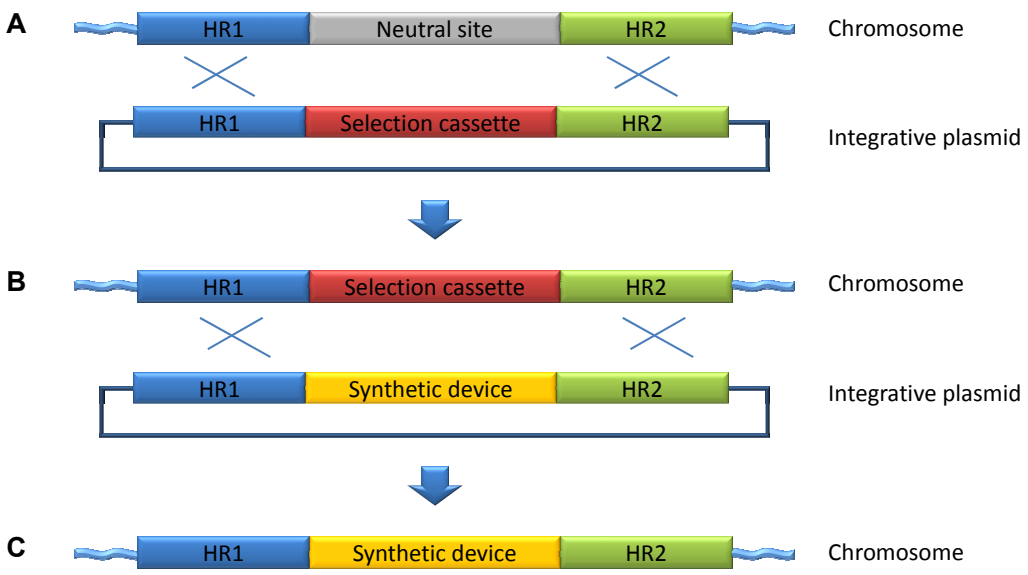


Figure 5 | Markerless integration of a synthetic device into a *Synechocystis* chromosomal neutral site by double homologous recombination. Recombination occurs at the two homologous regions (HR1 and HR2) at the same time. The chromosomal region between these two regions is replaced by the selection cassettes from the vector (A). A second recombination event allows replacing the selection cassettes by a synthetic construct present in another plasmid containing the same homologous recombination regions (B). Through two recombination events is possible to create mutants with no selection markers (C).

Since *Synechocystis* possesses multiple copies of the genome per cell, and to guarantee that all the copies are mutated, it is necessary to segregate the mutants, which is a time

consuming task (Figure 6). This is usually achieved by consecutively increasing the selection pressure (e.g. antibiotic concentration) and full segregation should be confirmed using molecular techniques like PCR and Southern blot.

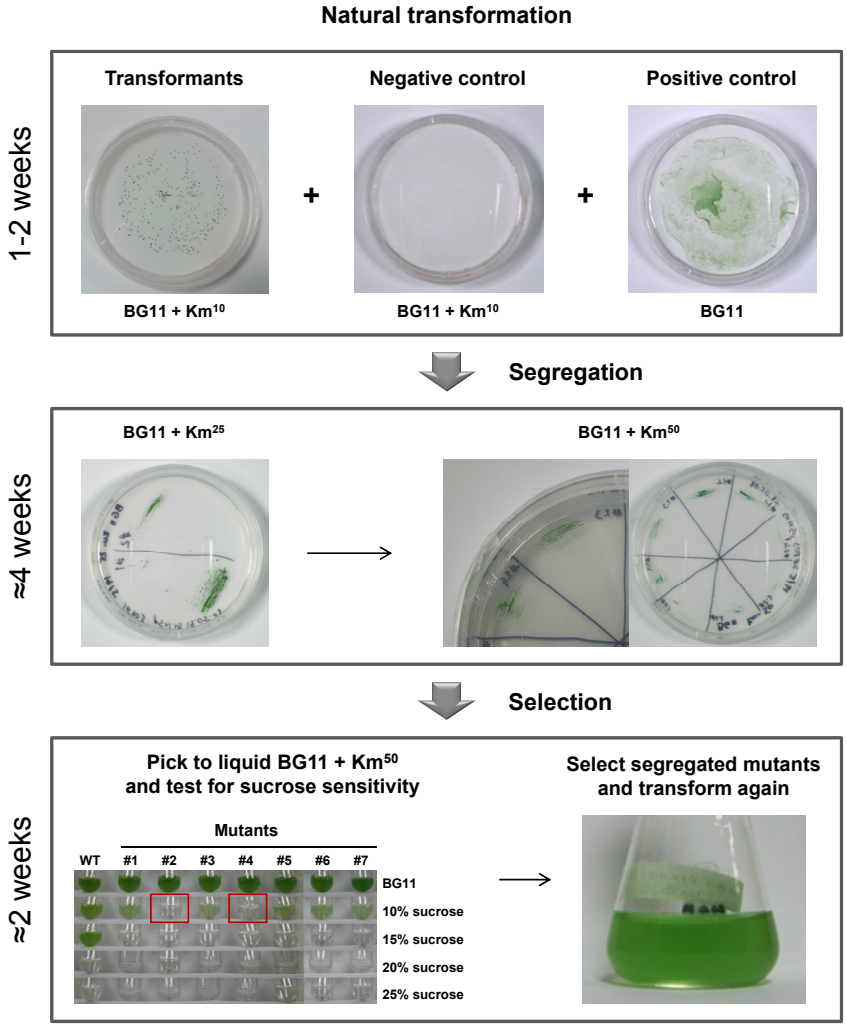


Figure 6 | **Natural transformation of *Synechocystis* sp. PCC 6803 with a plasmid carrying a double selection cassette.** *Synechocystis* can be naturally transformed incubating the cells with exogenous DNA. Transformant colonies appear in selective medium in 1-2 weeks and increasing the selective pressure leads to full segregation of the mutants (≈4 weeks). The final selection is based on the sensitivity conferred by the second marker (≈2 weeks), and after confirmation of full segregation by molecular techniques, the mutants can be used for further transformations.

Several studies have been carried out in order to optimize natural transformation in *Synechocystis*. Cell density (growth phase), DNA concentration, incubation time and conditions, size of homologous regions used, among others, were described as being important parameters (Kufryk *et al.*, 2002; Zang *et al.*, 2007b). Moreover, a *Synechocystis* deletion mutant of the gene *sll1354* (encoding the single-stranded-DNA-specific exonuclease RecJ) was shown to have the transformation efficiency increased by two

orders of magnitude (Kufryk *et al.*, 2002). This mutant with enhanced transformability can be used for difficult transformations where the number of transformants is very low using the parental strain. Moreover, some chromosomal loci have been used to introduce heterologous DNA into *Synechocystis* (e.g. Williams, 1988; Burnap *et al.*, 1994; Aoki *et al.*, 1995; Aoki *et al.*, 2011), namely to test promoters and transcription regulation or to introduce new functionalities into the organism, although their neutrality has not been extensively evaluated (see below).

The introduction of replicative plasmids into cyanobacteria is also possible through conjugation or electroporation (Matsunaga and Takeyama, 1995; Vioque, 2007). The type of plasmids used are usually not specific for cyanobacteria, and are typically based on the broad-host-range *E. coli* IncQ plasmid RSF1010 (Scholz *et al.*, 1989; Sode *et al.*, 1992; Marraccini *et al.*, 1993; Mermet-Bouvier *et al.*, 1993; Mermet-Bouvier and Chauvat, 1994; Huang *et al.*, 2010). Although presenting some advantages over integrative plasmids, since they are easier to construct and their introduction into the cell is faster, broad-host-range replicative plasmids can exhibit some instability (David *et al.*, 1983; Priefer *et al.*, 1985; Becker and Meyer, 1997; Meyer, 2009). Therefore, the prediction and control of the systems behavior may be compromised. In this sense, chromosomal recombination may constitute a better alternative for the integration of synthetic devices into cyanobacterial hosts, namely *Synechocystis*.

I.4 Tools for the design and characterization of synthetic parts, modules and circuits in cyanobacteria

I.4.1 Design and assembly of parts

The design of new synthetic parts and devices must follow the assembly standard specifications, and special attention must be paid to the chassis were they are going to function. Biological parts like promoters, RBSs, riboswitches and terminators do not always exhibit the same behavior in different biological backgrounds. For instance, the well characterized *E. coli* promoter P_{tetR} (BioBrick BBa_R0040) does not work in *Synechocystis* (Huang *et al.*, 2010). Therefore, the innumerous synthetic modules of the Registry of Standard Biological Parts containing the P_{tetR} promoter can not be used in this cyanobacterium. The same is true for proteins/protein modules DNA-coding sequences. Heterologous protein expression can be compromised due to differences in the frequency of occurrence of synonymous codons, in DNA of different organisms – the codon usage bias (Comeron and Aguadé, 1998; Ermolaeva, 2001; Suzuki *et al.*, 2008). For example, enhanced expression of the heterologous plant enzyme isoprene synthase has been

achieved in *Synechocystis* by codon optimization, leading to a 10 fold increase, comparing to the expression of the original plant gene (Lindberg *et al.*, 2010). Nonetheless, optimization should be wisely performed to avoid new sequences that greatly differ from the G+C content of the host, which could dramatically alter the nucleic acids conformation and melting temperatures, or even to prevent the potential depletion of high usage tRNAs. Codon optimization is therefore an important issue to take into consideration when inserting protein coding sequences into a heterologous organism (Ikemura, 1981; Yadava and Ockenhouse, 2003; Fox and Erill, 2010), and that can be assisted by computer programs like GeneDesigner (Villalobos *et al.*, 2006) and OPTIMIZER (Puigbò *et al.*, 2007).

Each new synthetic part should be designed following the BioBrick standards specifications in order to be assembled with existing ones: new DNA sequences have to be flanked by a prefix (upstream) and a suffix (downstream), and recognition sites for the restriction enzymes used in the assembly procedure should not be present. Several BioBrick assembly standards have been developed to fulfill the emerging needs of synthetic biologists (http://partsregistry.org), but the assembly standard RFC[10] is currently the most commonly used: the prefix contains recognition sites for *Eco*RI (E), *Not*I and *Xba*I (X) and the suffix for *Spe*I (S), *Not*I and *Pst*I (P). The assembly is performed by simply digesting a vector containing *part A* (donor) with E and S and a vector containing *part B* (recipient) with E and X for upstream cloning, or digesting a vector containing *part A* (recipient) with S and P and a vector containing *part B* (donor) with E and S for downstream cloning (Figure 7). Since the overhangs produced by X and S are compatible, producing a “scar” without restoring any of the sites, the final assembled *partA+B* has the same specifications in terms of prefix and suffix and can be further integrated into more complex parts/modules. Several bioinformatics tools like GenoCAD (Cai *et al.*, 2010) have also been developed to simplify the design of complex genetic constructs and artificial gene networks.

I.4.2 Characterization and implementation of synthetic parts and devices

The accurate characterization of individual genetic parts is mandatory to use synthetic biology tools to model and predict systems behavior in response to the implementation of complex synthetic circuits. If, for instance, fully characterized promoters make possible to fine-tune the transcription of a particular gene in the synthetic network, other parts contribute to the proper regulation of protein synthesis (translation). In this sense, Mitschke *et al*. (2011) used differential RNA sequencing to establish a genome-wide map of transcriptional start points (TSPs) of *Synechocystis*, enhancing the information on

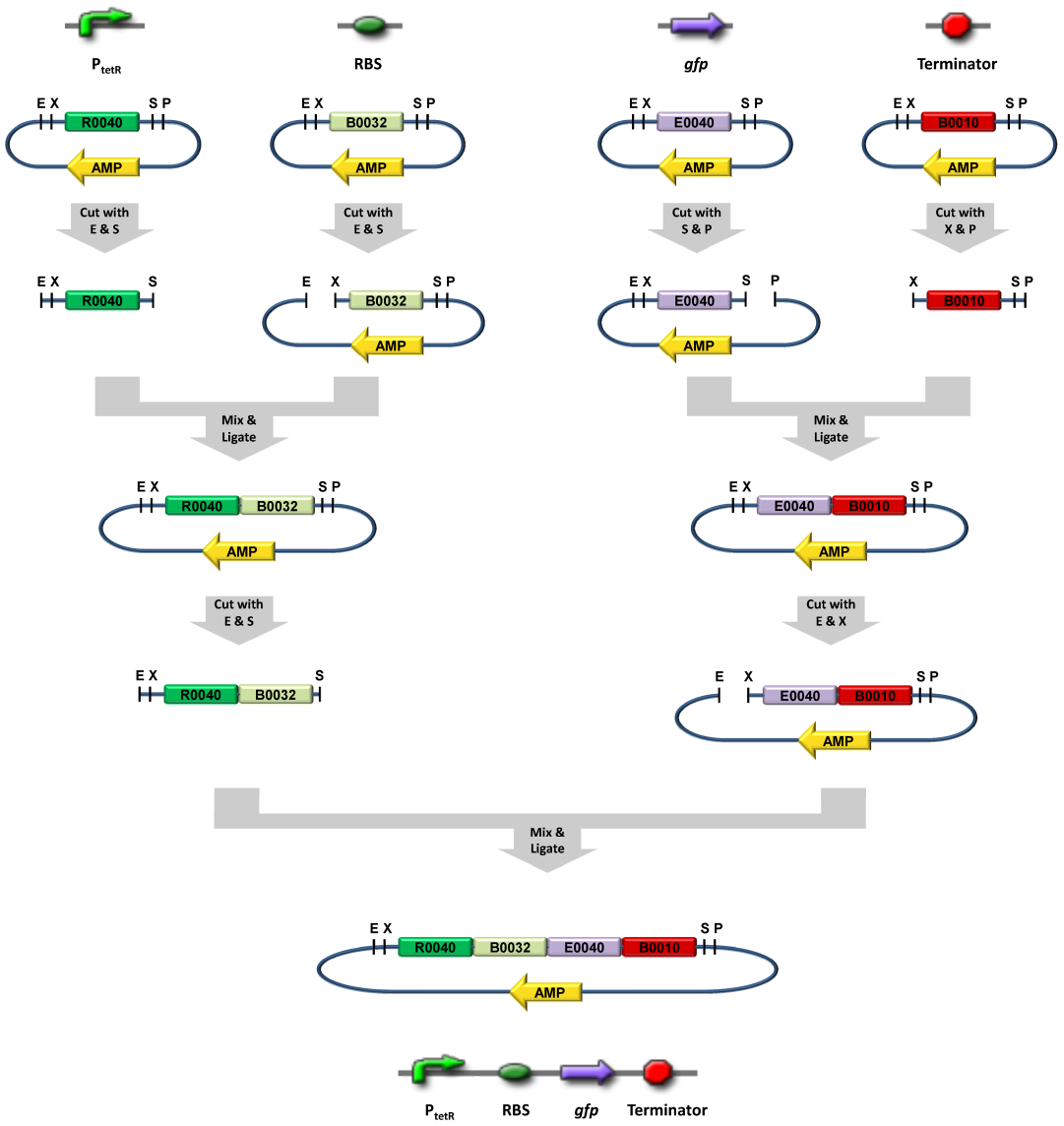


Figure 7 | Standard assembly of individual parts to construct a module expressing GFP. Standardized biological parts can be simply assembled by *cut & paste* molecular techniques into more complex/composite parts, using enzymes capable of recognizing the prefix and suffix restriction sites (E - *Eco*RI, X - *Xba*I, S – *Spe*I, P - *Pst*I). The assembly is possible due to the compatibility between X and S overhangs that after DNA ligation form a "scar" between two parts, but do not restore any recognition site. In the end, the module assembled from several individual parts (e.g. a promoter, a RBS, a GFP-coding gene and a transcription terminator) has the same structural elements of the assembly standard, and can be used in further assemblies.

promoters by a factor of 40. Knowing the "canonical" structure of *Synechocystis*' promoters makes possible to design new promoter libraries with different transcription strengths to control transcript levels or regulators to switch gene transcription on or off (Figure 8A). Subsequently, vectors like those developed by Marraccini *et al.* (1993) and Kunert *et al.* (2000) can be used to test the synthetic promoters. Similarly, RBS libraries can be designed based on the work of Hayes and Borodovsky (1998) and further tested

using the several reporter genes available for cyanobacteria, like *gfp* or the *luxAB* system (Friedberg, 1988; Koksharova and Wolk, 2002). Additionally, computational methods like TransTermHP (Ermolaeva *et al.*, 2000; Kingsford *et al.*, 2007) allow the prediction of transcription terminators on the genome of a specific organism, again enabling the design of terminator structures with different strengths. Further characterization of such structures can be performed placing terminator sequences between two reporter genes, assessing the transcription of the latter (Figure 8B).

In cyanobacteria, the characterization of synthetic parts and devices can be performed using reporter proteins (Friedberg, 1988; Koksharova and Wolk, 2002), and/or evaluating gene transcription using, for instance, quantitative real-time polymerase chain reaction (RT-qPCR) (Nolan *et al.*, 2006). Although more expensive, other technologies like

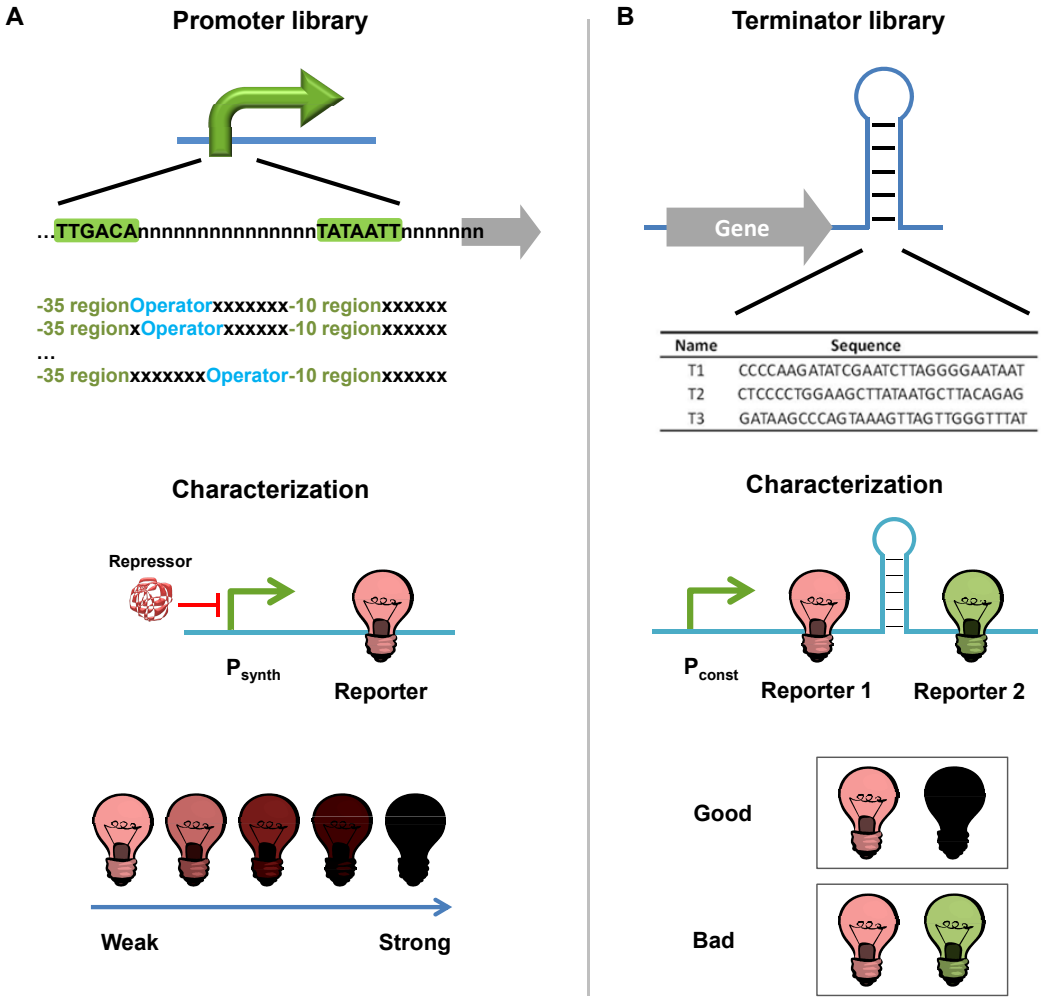


Figure 8 | Parts design and characterization. (A) The design and characterization of promoter libraries allow the fine tuning of gene transcription. Transcription factors can bind to operator sequences present in the promoter and regulate gene transcription. (B) The proper insulation of each transcription unit is provided by strong Rho-independent transcription terminators. A good terminator prevents transcription override from one module to the other.

microarrays (Hihara *et al.*, 2001; Singh *et al.*, 2003) and deep-sequencing techniques (Mitschke *et al.*, 2011) may also be used. Furthermore, the effects at the metabolic level can be assessed using high-throughput methodologies, like isobaric tag for relative and absolute quantitation (iTRAQ) proteomics (Ow and Wright, 2009).

Reporter genes usually originate proteins with physico-chemical properties that allow an easy assessment of protein expression, but these properties may be influenced by several factors such as temperature or redox balance, affecting the detectability (Tsien, 1998). Thus, inferring transcription data from reporter proteins signal may, sometimes, lead to erroneous results and mischaracterization of parts like promoters and RBSs. Therefore, the comprehensive evaluation of gene expression patterns is particularly important in the characterization process and, at present, RT-qPCR is considered the gold standard for the measurement of transcripts abundance (Bustin, 2010). This technique allows transcription analysis by the simultaneous amplification and quantification of a target amplicon by measuring the fluorescence increment in each PCR cycle in a fast, extremely sensitive and accurate way (VanGuilder *et al.*, 2008). Additionally, the data normalization strategy used to eliminate technical or experimentally induced variation is very important. Normalization against internal control genes is frequently used, yet the expression of housekeeping genes can vary significantly depending on the samples and experimental conditions (Thellin *et al.*, 1999; Bustin, 2000; Suzuki *et al.*, 2000; Dheda *et al.*, 2004). To minimize the influence of individual variation experimental validation of each reference gene is required for a particular sample and condition, and the use of three or more reference genes is now standard practice (Bustin *et al.*, 2009; Taylor *et al.*, 2010).

To characterize and implement synthetic parts and modules, it is necessary to introduce the DNA into the chassis. As referred above, heterologous DNA can be inserted into *Synechocystis* through double homologous recombination and, to date, several integration sites have been used (e.g. Williams, 1988; Burnap *et al.*, 1994; Aoki *et al.*, 1995; Aoki *et al.*, 2011). Although their neutrality has not been extensively addressed, insertions in those sites were reported as “not appear to be deleterious”, “did not show any detectable phenotype”, or being “a neutral site of the chromosome”. *Neutral sites* are chromosomal locations where integration can be performed without impairing the viability and fitness of the recombinant clones. Therefore, it is essential to identify and characterize additional neutral sites in *Synechocystis*, and the development of BioBrick-compatible integrative vectors would be useful for a straightforward integration of synthetic devices and circuits into the cyanobacterial chassis.

I.5 Aims of this work

The majority of the work presented in this thesis was related to the EU-funded project BioModularH_2 (FP6-NEST Pathfinder Synthetic Biology project number 043340) that aimed at producing a cyanobacterial chassis and synthetic modules, primarily designed for hydrogen production.

The specific aims of this work were:

(*i*) *Construction and characterization of a cyanobacterial chassis primarily designed for the integration of a synthetic hydrogen producing device* – Chapter II (Paper I).

(*ii*) *Identification and characterization of genomic "neutral sites" in* Synechocystis, *and construction of BioBrick-compatible integrative vectors* – Chapter III (Manuscript in preparation).

(*iii*) *Design of parts for synthetic biology applications* – Chapter IV (Unpublished results).

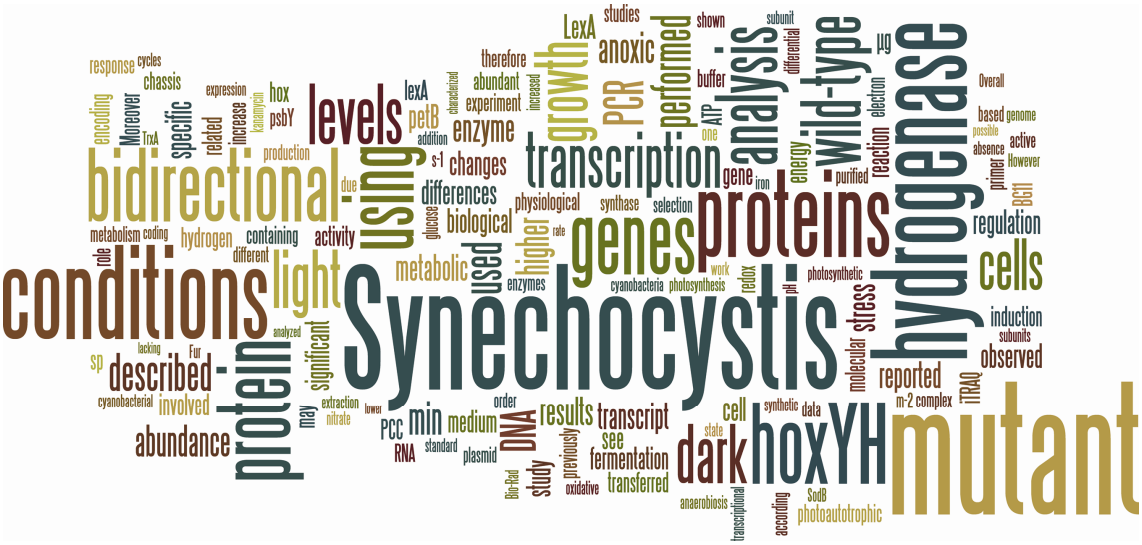
(*iv*) *Validation of reference genes for RT-qPCR analyses in cyanobacteria* – Chapter V (Paper II).

CHAPTER II

Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC 6803 mutant lacking a functional bidirectional hydrogenase

Pinto F, van Elburg KA, Pacheco CC, Lopo M, Noirel J, Montagud A, Urchueguía JF, Wright PC and Tamagnini P

Microbiology (2012) **158**:448-64



Microbiology (2012), **158**, 448–464 DOI 10.1099/mic.0.052282-0

Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC 6803 mutant lacking a functional bidirectional hydrogenase

Filipe Pinto,[1,2] Karin A. van Elburg,[3] Catarina C. Pacheco,[1] Miguel Lopo,[1] Josselin Noirel,[3] Arnau Montagud,[4] Javier F. Urchueguía,[4] Phillip C. Wright[3] and Paula Tamagnini[1,2]

Correspondence
Paula Tamagnini
pmtamagn@ibmc.up.pt

[1]IBMC – Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal

[2]Faculdade de Ciências, Departamento de Biologia, Universidade do Porto, Rua do Campo Alegre, Edifício FC4, 4169-007 Porto, Portugal

[3]Biological and Environmental Systems Group, ChELSI Institute, Department of Chemical and Biological Engineering, University of Sheffield, Mapping Street, Sheffield S1 3JD, UK

[4]Instituto Universitario de Matemática Pura y Aplicada, Universidad Politécnica de Valencia, Camí de Vera, E-46071 Valencia, Spain

Received 22 June 2011
Revised 10 October 2011
Accepted 10 November 2011

Cyanobacteria are photosynthetic prokaryotes that are promising 'low-cost' microbial cell factories due to their simple nutritional requirements and metabolic plasticity, and the availability of tools for their genetic manipulation. The unicellular non-nitrogen-fixing *Synechocystis* sp. PCC 6803 is the best studied cyanobacterial strain and its genome was the first to be sequenced. The vast amount of physiological and molecular data available, together with a relatively small genome, makes *Synechocystis* suitable for computational metabolic modelling and to be used as a photoautotrophic chassis in synthetic biology applications. To prepare it for the introduction of a synthetic hydrogen producing device, a *Synechocystis* sp. PCC 6803 deletion mutant lacking an active bidirectional hydrogenase (Δ*hoxYH*) was produced and characterized at different levels: physiological, proteomic and transcriptional. The results showed that, under conditions favouring hydrogenase activity, 17 of the 210 identified proteins had significant differential fold changes in comparisons of the mutant with the wild-type. Most of these proteins are related to the redox and energy state of the cell. Transcriptional studies revealed that only six genes encoding those proteins exhibited significant differences in transcript levels. Moreover, the mutant exhibits similar growth behaviour compared with the wild-type, reflecting *Synechocystis* plasticity and metabolic adaptability. Overall, this study reveals that the *Synechocystis* Δ*hoxYH* mutant is robust and can be used as a photoautotrophic chassis for the integration of synthetic constructs, i.e. molecular constructs assembled from well characterized biological and/or synthetic parts (e.g. promoters, regulators, coding regions, terminators) designed for a specific purpose.

INTRODUCTION

Cyanobacteria are photosynthetic prokaryotes that are able to perform oxygenic photosynthesis, using water as an electron donor. They constitute one of the largest and most remarkable groups of bacteria on Earth, playing a key role in both the carbon and the nitrogen cycle (Bothe *et al.*, 2010b; Moisander *et al.*, 2010). Cyanobacteria are a prolific source of bioactive compounds (Burja *et al.*, 2001; Dembitsky, 2006), they have been used as biofertilizers and food supplements, and their strong potential in the fields of energy, bioremediation, bioplastics and medical diagnostics has been reported (Abed *et al.*, 2009; Devillers *et al.*, 2007). The simple nutritional requirements of these organisms, combined with their autotrophy and metabolic plasticity, as well as the availability of molecular tools for their genetic manipulation, makes them promising 'low-cost' microbial

Abbreviations: FDR, false discovery rate; RT-qPCR, real time quantitative PCR.

A set of supplementary methods, four supplementary figures and two supplementary tables are available with the online version of this paper.

052282 © 2012 SGM *Printed in Great Britain*

cell factories (Abed *et al.*, 2009; Thajuddin & Subramanian, 2005).

Synthetic biology applies the standardization and hierarchical abstraction of engineering to biological systems, in order to redesign existing systems or design entirely new ones (Bashor *et al.*, 2010; Endy, 2005; Khalil & Collins, 2010; Yadav & Stephanopoulos, 2010). Following this philosophy, the BioModularH_2 project (www.biomodularh2.org) aimed at designing and constructing molecular parts/modules/devices that, once integrated into a cyanobacterial chassis, will redirect metabolic fluxes towards efficient hydrogen production. For this purpose, the unicellular non-N_2-fixing *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis*) was chosen as the chassis. Biological chassis should be robust and well characterized organisms preferably capable of growing on minimal, inexpensive carbon sources considering their use for laboratory- or large-scale production of the desired molecule(s). Nonetheless, even in most robust hosts/chassis, foreign DNA instability can pose a problem, due to their ability to inactivate or dispose of the foreign genes (Keasling, 2008). In *Escherichia coli* this issue has been addressed by removing large sections of the chromosome, some containing mobile elements and cryptic virulence genes (Pósfai *et al.*, 2006; Sharma *et al.*, 2007a, b). Additionally, the presence of genes encoding proteins that may interfere and/or compete with those introduced should be eliminated. In this context, *Synechocystis* is a photoautotrophic organism with minimal nutritional requirements that uses CO_2 and sunlight as sources of carbon and energy. Its genome, the first to be sequenced among cyanobacteria, comprises a 3.6 Mb chromosome and seven plasmids (Kaneko *et al.*, 1996, 2003; Xu & McFadden, 1997; Yang & McFadden, 1993, 1994). This strain became the most studied cyanobacterium, for which plenty of physiological and molecular information is now available (Barrios-Llerena *et al.*, 2006; Bhaya *et al.*, 2000; Koksharova & Wolk, 2002; Ow & Wright, 2009; Schmitt & Stephanopoulos, 2003; Suzuki *et al.*, 2006; Zhang *et al.*, 2008). Moreover, it is naturally transformable with exogenous DNA, and a variety of molecular tools have been developed for its genetic manipulation. The vast amount of data, together with a relatively small genome, makes *Synechocystis* suitable for computational metabolic modelling and to be used as a photoautotrophic chassis.

Taking into account the aim of the BioModularH_2 project, and since a heterologous hydrogenase (hydrogen producing device) was going to be introduced, the native bidirectional hydrogenase of *Synechocystis* became a redundant part, and therefore should be removed. It has been reported that the bidirectional hydrogenase of *Synechocystis* could be involved in dark fermentation, act as an electron valve during photosynthesis, or even be part of the respiratory complex I (Antal & Lindblad, 2005; Appel & Schulz, 1996; Appel *et al.*, 2000; Gutthann *et al.*, 2007; Troshina *et al.*, 2002). The bidirectional hydrogenase of *Synechocystis* has been recently characterized as a truly bidirectional enzyme with a bias to H_2 production (McIntosh *et al.*, 2011); however, further investigation is needed to clarify the role of this type of enzyme (Angermayr *et al.*, 2009; Bothe *et al.*, 2010a; Carrieri *et al.*, 2011; McIntosh *et al.*, 2011; Oliveira, 2008; Tamagnini *et al.*, 2002, 2007). In this work, a *Synechocystis* deletion mutant (Δ*hoxYH*), lacking an active bidirectional hydrogenase and without any selection marker, was produced and characterized. In addition, a strategy to construct vectors compatible with the BioBrick system (http://partsregistry.org/Main_Page) was developed to efficiently remove redundant genes and/or introduce DNA-based synthetic parts and circuits into the *Synechocystis* genome. The molecular tools (BioBrick compatible integrative vectors) and the photoautotrophic chassis produced in this work can be easily extended to other biotechnological applications aiming to exploit a light-driven biological system.

METHODS

Organisms and standard growth conditions. The unicellular non-N_2-fixing cyanobacterium *Synechocystis* sp. PCC 6803 (obtained from the Pasteur Culture Collection, Paris, France) was maintained in BG11 medium (Stanier *et al.*, 1971) at 25 °C and 12 h light (20 μE m^{-2} s^{-1} with Osram L18W/765 cool white daylight bulbs)/12 h dark cycles, unless stated otherwise. Cosine-corrected irradiance was measured by using a quantum meter (Skye SKP 216 Quantum sensor with SKP 200 Measuring unit, Hansatech Instruments). For solid medium, BG11 was supplemented with 1.5 % Noble agar (Difco), 0.3 % sodium thiosulfate and 10 mM TES–KOH buffer (pH 8.2). For the selection of the mutants, BG11 medium was supplemented with kanamycin (Km, 10–50 μg ml^{-1}) or sucrose [10 % (w/v)]. For the proteomics analysis, standard growth conditions were used, with a scalar irradiance of 75 μE m^{-2} s^{-1} (measured with a QSL-2100 Radiometer, Biospherical Instruments, and equivalent to the standard growth light intensity described above). Anaerobic conditions were induced by sparging with argon for 30 min in the dark. Cells were then kept in the dark for 2 h before being harvested. *E. coli* DH5α (Stratagene) was used for cloning purposes. *E. coli* transformants were cultivated at 37 °C in Luria–Bertani (LB) medium supplemented with 100 μg ampicillin ml^{-1} or 50 μg kanamycin ml^{-1}.

DNA extraction and recovery. Cyanobacterial genomic DNA was extracted using the phenol/chloroform method described previously (Tamagnini *et al.*, 1997). Agarose gel electrophoresis was performed by standard protocols using 1× TAE buffer (Sambrook & Russell, 2001), and the DNA fragments were isolated from gels using the GFX PCR DNA and Gel Band Purification kit (GE Healthcare), according to the manufacturer's instructions.

Plasmid construction. The plasmids used in this work, pGDYH and pGDYH.NS, were based on pGEM-T Easy (Promega) and contain the *Synechocystis* chromosomal regions flanking the *hoxYH* genes (encoding subunits of the bidirectional hydrogenase). The regions upstream from *hoxY* and downstream from *hoxH* were amplified by PCR using primers with tags containing restriction sites for cloning purposes (Table 1). Additionally, the primers 5-I and 3-I contain sites compatible with the BioBrick system: *Mun*I, *Xba*I, *Spe*I and *Pst*I. Each PCR mixture (50 μl) contained 1.5 U *Pfu* DNA polymerase (Fermentas), 1 × reaction buffer [10 × reaction buffer: 200 mM Tris/HCl pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 mg BSA ml^{-1}, 1 % (v/v) Triton X-100, 20 mM $MgSO_4$], 250 μM of each deoxyribonucleotide triphosphate, 200 nM each primer (primer pairs 5-O/5-I or 3-O/3-I) and 7 ng *Synechocystis* genomic DNA. The PCR profile was 2 min at 95 °C followed by 35 cycles of 30 s at 95 °C, 45 s at 60 °C and

Table 1. Oligonucleotide primers used in this study

Name	Sequence (5′–3′)	Purpose
5-O*	GAAGCCCGGGTTCCCATTCCCACCCTTTGCCATTTAG	Mutant construction/Southern probe
5-I*	CTGCAGACTAGTACGACCTTCTAGACAATTGGAGCCA TTCGTCCATATCAAGGAAGGACATATGACAGC	Mutant construction/Southern probe
3-O*	TAGTAAGCTTGCCTTTGATAACCACAGTGCCCAG	Mutant construction
3-I*	CAATTGTCTAGAAGGTCGTACTAGTCTGCAGCC TGCCTTAGTTGTTCTACCCATGCAGCGGGACAAATG	Mutant construction
FO	AGAATATTGGATTGTGGATGTGAAGGC	Mutant confirmation
RO	CAGGGAGTGGAGTTAATTAGAACAGG	Mutant confirmation
NeosacB2F*	GCTGGAATTCAGGAAGCGGAACACGTAGAAAG	Selection cassette amplification
NeosacB3R*	CTACCAATTGCGTAACAGATGAGGGCAAGCGGATGG	Selection cassette amplification
U.atpC_2F	CTAGTGGCTCTATTCGTATC	RT-qPCR
U.atpC_2R	ATAACTGGTTGGTGTTGTAC	RT-qPCR
U.lexA_F	GGTTACGCAATAAGG	RT-qPCR
U.lexA_R	AGAACGGAGAATCAG	RT-qPCR
U.lrtA_F	GAAGACAAGCCAGTG	RT-qPCR
U.lrtA_R	TGAATAACGCCATAGC	RT-qPCR
U.petB_2F	TCAATGGTTCAATGATCGTC	RT-qPCR
U.petB_2R	CATGATGTATTGGACAGAGG	RT-qPCR
U.psbY_F	ATTGGCGTGTAATTGTAG	RT-qPCR
U.psbY_R	AACGTCCTGTAACTGC	RT-qPCR
U.rplK_F	GATGATTATTCCTGTGG	RT-qPCR
U.rplK_R	GTCGTTGGCATTAAG	RT-qPCR
U.sodB_F	CGCTGACTTTGGTAG	RT-qPCR
U.sodB_R	TCGGCAATGTAATCG	RT-qPCR
U.trxA_F	TTAGACAGCGAGTTAC	RT-qPCR
U.trxA_R	CACCACCATATCCAC	RT-qPCR
U.tufA_F	GAATTGGTGGAACTG	RT-qPCR
U.tufA_R	GTGATGGAGAATACG	RT-qPCR
D.ppa_2F	TAAATACGAGTTCGACAAAG	RT-qPCR
D.ppa_2R	TCAATCATTTCCAACATACC	RT-qPCR
D.pgk_2F	GAATTACAGTTCCTCCAAGG	RT-qPCR
D.pgk_2R	AGTCCAATTTGTCTTCTTCC	RT-qPCR
D.slr2025_F	CAAGATAACGACCAG	RT-qPCR
D.slr2025_R	GAGAGCCACTAAATG	RT-qPCR
BD16SF1†	CACACTGGGACTGAGACAC	RT-qPCR
BD16SR1†	CTGCTGGCACGGAGTTAG	RT-qPCR

*Restriction enzyme recognition sites are underlined.
†Ferreira *et al.* (2009).

2.5 min at 72 °C, and a final extension at 72 °C for 7 min. The two purified PCR fragments were fused by 'overlap-PCR', and the reaction mixture (50 µl) contained 1.25 U *Taq* polymerase (GE Healthcare), 1 × reaction buffer (10 × reaction buffer: 100 mM Tris/HCl pH 9.0, 15 mM $MgCl_2$ and 500 mM KCl), 250 µM of each deoxyribonucleotide triphosphate, 0.2 µg BSA μl^{-1}, 130 nM each outer primer (5-O and 3-O) and 80 ng each purified DNA fragment. The PCR profile was 5 min at 95 °C followed by 35 cycles of 30 s at 95 °C, 45 s at 60 °C and 1.75 min at 72 °C, and a final extension at 72 °C for 10 min. The resulting product was purified and cloned into the vector pGEM-T Easy, according to the manufacturer's instructions, producing the pGDYH plasmid. A selection cassette, containing the *nptII* gene (conferring resistance to neomycin and kanamycin) and the *sacB* gene (sucrose sensitivity), was amplified by PCR from the plasmid pK18mobsacB (Schäfer *et al.*, 1994), using the primer pair NeosacB2F and NeosacB3R. The PCR mixture (20 µl) contained 0.5 U *Taq* polymerase (GE Healthcare), 1 × reaction buffer (10 × reaction buffer is 100 mM Tris/HCl pH 9.0, 15 mM $MgCl_2$ and 500 mM KCl), 200 µM of each deoxyribonucleotide triphosphate, 1 µM of each primer and 3 ng of each purified DNA fragment. The PCR profile was: 2 min at 95 °C followed by 40 cycles of 30 s at 95 °C, 30 s at 60 °C and 3 min at 72 °C and a final extension at 72 °C for 7 min. The purified PCR fragment was cloned into pGEM-T Easy, and the resulting vector was digested with *Eco*RI and *Mun*I restriction enzymes (Fermentas), according to the manufacturer's instructions. Subsequently, the purified DNA fragment containing the selection cassette was cloned in the *Mun*I restriction site of pGDYH, using the T4 DNA ligase (Fermentas) to form the plasmid pGDYH.NS. All the constructs were confirmed by DNA sequencing (STAB VIDA).

Generation of the *Synechocystis* sp. PCC 6803 *hoxYH* deletion mutant. *Synechocystis* was transformed using a method based on a procedure described by Williams (1988). Briefly, *Synechocystis* cultures were grown in BG11 medium at 30 °C, under continuous

light (20 μE m^{-2} s^{-1}), to OD_{730} ~0.5. Cells were harvested by centrifugation and resuspended in 1/10 volume of BG11. A 100 μl sample of these cells was incubated with the purified pGDYH.NS plasmid for 5 h, in light at room temperature, with a final plasmid DNA concentration of 6–20 μg ml^{-1}. Cells were then spread onto Immobilon-NC membranes (0.45 μm pore size, 82 mm, Millipore) resting on solid BG11 plates, grown at 30 °C under continuous light, and were transferred to selective plates containing 10 μg kanamycin (Km) ml^{-1} after 24 h. Transformants were observed after 1–2 weeks. For complete segregation, Km-resistant colonies were grown at increasing Km concentrations (25 and 50 μg ml^{-1}) and finally transferred into liquid medium. Mutants were then tested for sucrose sensitivity, and confirmed by PCR and Southern blot with a probe labelled with digoxigenin using the DIG DNA labelling kit (Roche Molecular Biochemicals) (Fig. 1). Subsequently, to remove the selection markers from the disruption mutant, cells were transformed as described above with the vector pGDYH, and the mutants were selected in solid BG11 containing 10 % sucrose (w/v). These mutants were also screened for Km-sensitivity, and the full segregation was confirmed by PCR using primers external to the insertion sites (FO and RO, Table 1) and by Southern blot (Fig. 1). Southern blot was performed using genomic DNA that was digested with *Hin*dIII or *Hin*cII, and a probe covering the 5′ flanking region of *hoxYH* genes (amplified by PCR using the primer pair 5-O and 5-I). A representative diagram of the mutant construction is depicted in Fig. 2. Moreover, the presence/absence of an active bidirectional hydrogenase in the wild-type and Δ*hoxYH* mutant was assessed by gas chromatography after placing the cells in dark and replacing air with argon.

Growth experiments. Cultures of *Synechocystis* wild-type and Δ*hoxYH* mutant were grown to OD_{730} ~1.0, and 24 h before the beginning of the experiment, were transferred to fresh BG11 medium. Inocula of 200 μl were then added to sterile 4.5 ml cuvettes containing 1800 μl medium with or without 5 mM glucose and given concentrations of nitrate (0.1 or 1.5 g l^{-1}). Each experiment was performed in triplicate and under aseptic conditions. The cuvettes were closed with sterilized Parafilm and placed in acrylic racks specially designed for this experiment. The racks were placed in a chamber with constant temperature (25 °C), under continuous light (20 μE m^{-2} s^{-1}) or a 12 h light (20 μE m^{-2} s^{-1})/12 h dark regimen, and the OD_{730} was recorded daily.

Statistical analysis of the specific growth. The quantitative analysis of *Synechocystis* wild-type and Δ*hoxYH* mutant growth in various conditions was based on the optical density measurements (see above). For each bin, *i*, a specific growth parameter, representing the exponent of (assumed) exponential growth within the given time interval was calculated as

$$\mu_i = \frac{\log(x_{i+1}/x_i)}{t_{i+1} - t_i}$$

where t_i is time, in units of days running from day 0 to $t_{i\max}$, and x_i is the corresponding measured optical density.

In order to work with a statistic that does not depend rather arbitrarily on test duration, particularly if under some particular conditions stationary growth is reached sooner than in others, we chose not to work with the mean specific growth rate but rather the maximum specific growth rate:

$$\mu_{\max} = \max(\mu_i)|_{i=0,\ldots,i_{\max}}$$

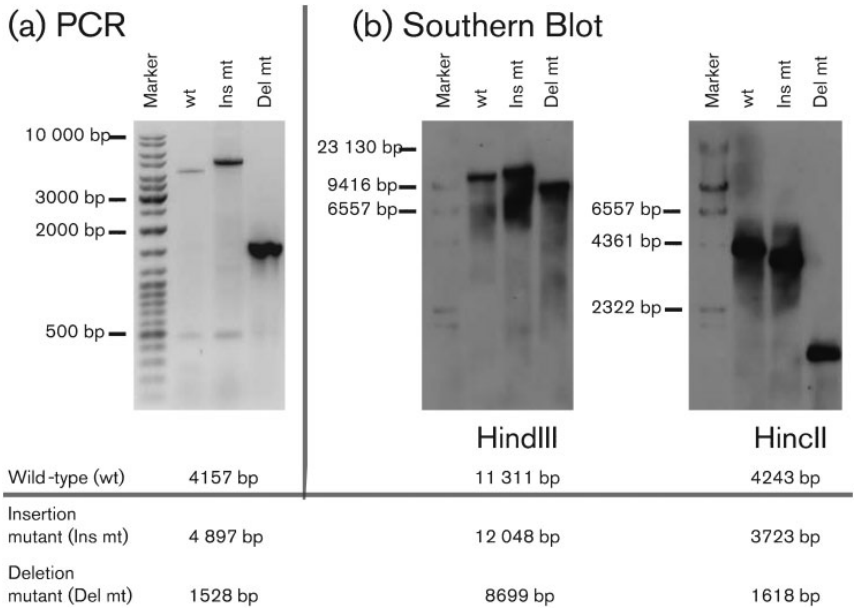


Fig. 1. Confirmation of the segregation of *Synechocystis* insertion (Ins mt) and deletion (Del mt) mutants by PCR amplification using a primer upstream of *hoxU* and another within *fabF* (a) and by Southern blot hybridization (b). Genomic DNA was digested with *Hin*dIII or *Hin*cII and hybridized with a probe covering the 5′ flanking region of *hoxYH* genes. The expected sizes of the PCR products and restriction fragments are shown below the figure.

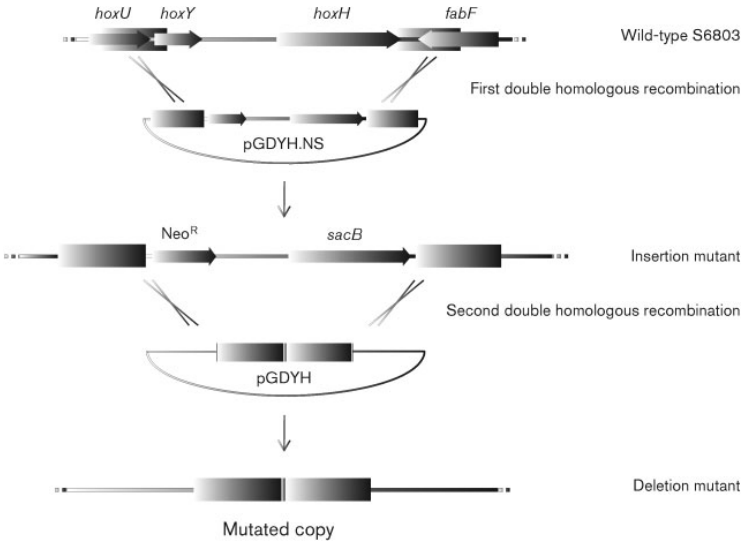


Fig. 2. Schematic representation of the experimental design used to generate the *Synechocystis* Δ*hoxYH* mutant.

In addition, we can obtain the time ($t_{i\max}$) at which maximum specific growth is reached and the corresponding optical density recorded ($x_{i\max}$).

From the three replicas of any of the experimental samples with a given level of light, nitrate and glucose, it was possible to calculate confidence intervals for the sample growth parameters, corrected using Student's *t*-test. The usual confidence level of 95 % was selected and used to represent mean and error bars in Fig. 3.

Sample collection, protein extraction and iTRAQ labelling. Cells were collected by centrifugation, the supernatant (medium) was removed, the pellets were snap-frozen in liquid nitrogen and preserved at −80 °C prior to protein extraction. The protein extraction and quantification was performed as described by Gan *et al.* (2007). Protein samples (100 µg) were precipitated, prepared and digested as described by Chong *et al.* (2006). iTRAQ 8-plex labelling (ABSciex) was performed as described by Ow *et al.* (2009).

LC-MS and MS analysis. Two biological *Synechocystis* wild-type replicates were labelled with the 117 and 118 reagents and two biological Δ*hoxYH* mutant replicates were labelled with the 119 and 121 reagents (ABSciex). Peptide pre-fractionation of complex iTRAQ-labelled lysate was performed as described by Gan *et al.* (2007). Peptide detection was performed using LC–MS-MS iTRAQ compatible settings (see Gan *et al.*, 2007). Protein quantification and identification was carried out using Phenyx Software version 2.6 (Genebio). Spectra were searched against the Uniprot database (accessed August 2009), containing 3576 protein sequences of *Synechocystis*. The false discovery rate (FDR) was calculated using a reverse database generated with Phenyx (Reidegeld *et al.*, 2008), as described by Elias & Gygi (2007). Both the true and the reversed database were searched in the same run. The FDR was below 5 % (Ow & Wright, 2009). Searches were performed in four steps in order to increase proteome coverage (see Supplementary Methods, available with the online version of this paper). Mass tolerances were set to 0.4 Da MS and 0.4 Da MS–MS. The peptide modifications were the same for each search (MMTS, fixed cysteine modification; oxidation, variable; iTRAQ_8plex_K, fixed; iTRAQ_8plex_N, fixed). The peptide lists generated by Phenyx were analysed as described by Ow *et al.* (2009). This list was used for further analysis as described by Pham *et al.* (2010) to produce a list of differentially expressed proteins. Protein function and annotations were assigned according to CyanoBase (Nakamura *et al.*, 2000).

Pathway-level analysis of the proteomic data. The network modelling approach 'mixture model on graphs' (MMG) (Noirel *et al.*, 2008; Sanguinetti *et al.*, 2008) was used on the metabolic network of *Synechocystis* as given by the Kyoto Encyclopedia of Genes and Genomes (KEGG; accessed September 2009). This provides a pathway-level detection of changes based on the proteomics dataset, even when the changes are too small to pass the statistical test used to detect differential expression (for details see Noirel *et al.*, 2009). Briefly, this method classifies the genes' products into three categories: upregulated, downregulated or unregulated. This is based on assumed distribution of the log ratios (Gaussian distribution for unregulated proteins, exponential distribution for differentially regulated proteins) and the topology of the network: if an enzyme is surrounded in the metabolic network by other upregulated enzymes, it will be more likely to be classified as upregulated. The details are described by Sanguinetti *et al.* (2008).

An improvement with respect to the original implementation has been introduced here: instead of using a single variance parameter σ^2 for the unregulated proteins' Gaussian distribution, a protein-specific variance σ^2 is introduced. It is calculated as

$$\sigma^2=\sigma_B{}^2+(\sigma_{117}{}^2+\sigma_{118}{}^2+\sigma_{119}{}^2+\sigma_{121}{}^2)/n$$

where the biological variation $\sigma_B{}^2$ is estimated from all eight-plex values; $\sigma_{117}{}^2$ etc. represent the peptide-level, technical variation which is adjusted by the number of MS–MS spectra, *n*. Proteins whose quantifications are more reliable will therefore be more reliably classified.

Sample collection and RNA extraction for transcriptional studies. The sample collection for RNA extraction is depicted in Fig. 4. Briefly, three independent cultures of *Synechocystis* wild-type and Δ*hoxYH* mutant cells were grown under standard conditions (see above) and at the middle of the light phase, they were transferred to the dark and sparged with argon to induce anaerobiosis. Samples were collected 0, 60 and 120 min after anaerobiosis induction (t_0, t_{60} and t_{120}, respectively). In addition, extra cultures were put back in light

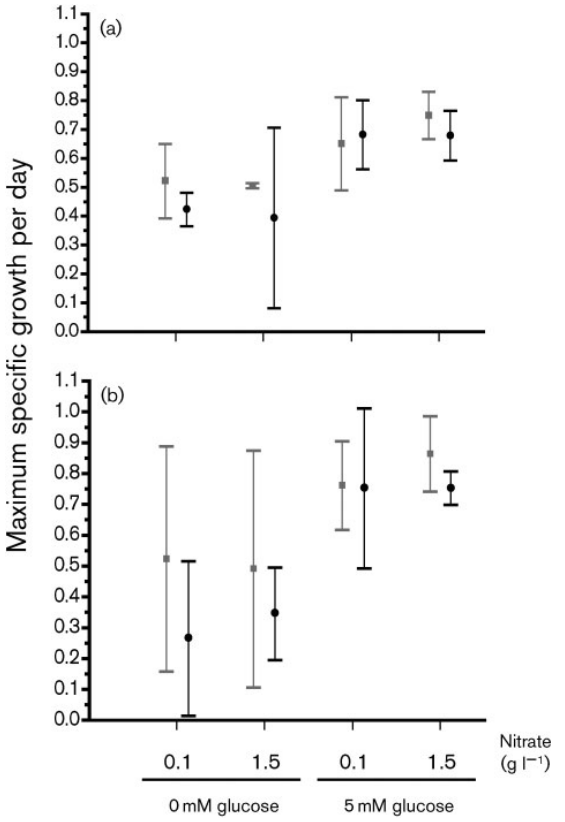


Fig. 3. Mean of the maximum specific growth of *Synechocystis* wild-type (■) and Δ*hoxYH* mutant (●), under different nitrate, glucose and light regimens: (a) 12 h light (20 μE m^{-2} s^{-1})/12 h dark regimen; (b) continuous light (20 μE m^{-2} s^{-1}). Whiskers represent confidence intervals. Mean and confidence intervals limits were calculated and corrected using the Student's *t*-distribution.

60 min after anaerobiosis induction, and samples were collected 60 min later [$t_{120(60L)}$]. To inhibit RNA transcription, rifampicin was added to each sample upon harvesting at a final concentration of 300 μg ml^{-1}. Cells were centrifuged (8 min at 4500 ***g***, 4 °C) and washed once with fresh BG11, and the cell pellets were frozen at −80 °C. RNA was extracted using the TRIZOL Reagent (Invitrogen) according to the method described previously (Leitão *et al.*, 2006). RNA was quantified on a NanoDrop 2000 spectrophotometer (Thermo Scientific) and the quality was checked using the Experion RNA StdSens Analysis kit (Bio-Rad).

Transcription analysis by real time quantitative PCR (RT-qPCR). For cDNA synthesis, 1 μg total RNA was transcribed with the iScript Select cDNA Synthesis kit (Bio-Rad), using the random primers supplied and following the manufacturer's instructions. The primer pairs used for the gene transcription analysis are listed in Table 1 and Supplementary Table S1 (available with the online version of this paper), and were designed using Beacon Designer 6 software (PREMIER Biosoft International). For each analysis, the 16S rRNA was used for normalization. The real-time PCRs

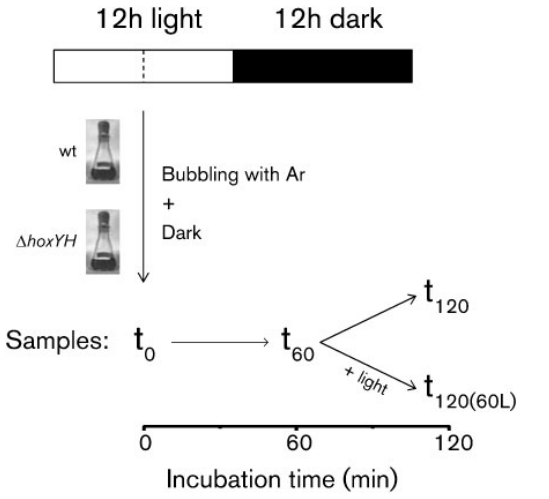


Fig. 4. Schematic representation of the experimental procedure for the collection of the samples for the RNA extraction used for RT-qPCR.

were performed using 0.25 μM each primer, 10 μl iQ SYBR Green Supermix (Bio-Rad) and 2 μl template cDNA, and technical triplicates were performed. The PCR profile was: 3 min at 95 °C followed by 40 – 60 cycles of 30 s at 95 °C, 30 s at 51 °C and 30 s at 72 °C; exceptionally the annealing temperature for the study of *atpC* and *ppa* was 53 °C and 54 °C for *pgk*. Standard dilutions of the cDNA were used to check the relative efficiency and quality of primers. Negative controls (no template cDNA) were included in all assays. A melting curve analysis was performed at the end of each PCR program, to exclude the formation of non-specific products. Real time quantitative PCRs (RT-qPCRs) were carried out in the iCycler iQ5 Real-Time PCR Detection System (Bio-Rad). The data obtained were analysed using the Bio-Rad iQ5 Optical System Software v2.1 (Bio-Rad). The two-way ANOVA statistical analysis of the data obtained in the RT-qPCR experiment was performed using the software GraphPad Prism v5 (GraphPad Software), comparing *Synechocystis* Δ*hoxYH* mutant with the wild-type, under dark anoxic conditions (t_0, t_{60} and t_{120}) and after cells were transferred back to light [t_{60} and $t_{120(60L)}$].

RESULTS AND DISCUSSION

Construction and physiological characterization of *Synechocystis* sp. PCC 6803 Δ*hoxYH* deletion mutant

The genes coding for the two subunits corresponding to the hydrogenase part of the bidirectional hydrogenase (*hoxYH*) were deleted from the *Synechocystis* genome in a two-step procedure: a first transformation was performed using the plasmid pGDYH.NS in order to replace *hoxYH* with a selection cassette conferring kanamycin resistance and sucrose sensitivity, followed by a second transformation, using the plasmid pGDYH, in order to remove the selection cassette. The complete segregation of both mutants was confirmed by resistance/sensitivity to kanamycin/sucrose,

and by PCR and Southern blot (Fig. 1). The absence of H_2 evolution (bidirectional hydrogenase activity) by the Δ*hoxYH* mutant was confirmed by gas chromatography (data not shown). To elucidate the possible effects of this mutation on the fitness of *Synechocystis*, the growth of both the wild-type and the Δ*hoxYH* mutant was compared under several physiological conditions (see Methods). In Fig. 3(a) the sample mean of maximum specific growth, μ_{max}, at different levels of nitrate and glucose, under 12 h light (20 μE m^{-2} s^{-1})/12 h dark cycles for *Synechocystis* wild-type and Δ*hoxYH* mutant are shown. As can be observed, in the presence of 1.5 g nitrate l^{-1} and with glucose, the wild-type μ_{max} increases significantly compared with the same conditions in the absence of glucose. In Fig. 3(b), with continuous 20 μE m^{-2} s^{-1} irradiance, glucose increases maximum specific growth, while nitrate decreases the variability of the mutant. In conclusion, these analyses show no significant difference in terms of specific growth for *Synechocystis* wild-type and Δ*hoxYH* mutant under the different levels of nitrate, glucose and irradiance used. These results are in agreement with what was previously reported for a disruption mutant (with a kanamycin resistance cassette within *hoxH*) growing in continuous light (100 μE m^{-2} s^{-1}) and bubbled with air (Appel *et al.*, 2000). However, differences in growth have been detected when cultures were bubbled with CO_2-enriched air [2 % (v/v) CO_2], suggesting that Δ*hoxH* cells need more time than the wild-type to reach the maximal growth rate (Appel *et al.*, 2000).

The role of the bidirectional hydrogenase in cyanobacterial metabolism is not fully understood (Bothe *et al.*, 2010a; Carrieri *et al.*, 2011; Tamagnini *et al.*, 2007). However, it was hypothesized by Ludwig *et al.* (2006) that the ancestor of cyanobacteria would have vertically transmitted the two hydrogenases genes (*hox* – bidirectional hydrogenase, *hup* – uptake hydrogenase), and that none, one or both enzymes were subsequently lost through evolution. Thus, the current distribution would be the result of cyanobacterial adaptation to various ecological niches. About 50 % of the N_2-fixing strains do not possess the bidirectional hydrogenase (Ludwig *et al.*, 2006; Schütz *et al.*, 2004; Tamagnini *et al.*, 1997, 2000), and although most of the non-N_2-fixing strains contain it, this enzyme does not seem to be essential for cell survival (Appel *et al.*, 2000; Masukawa *et al.*, 2002). Accordingly, our results suggest that, under the conditions tested, the bidirectional hydrogenase has no significant effect on the fitness of *Synechocystis*. Altogether, and taking also into account the results obtained by Appel *et al.* (2000), the absence of this enzyme seems to have only a marginal effect.

Comparison of the proteomes of wild-type *Synechocystis* and of the Δ*hoxYH* mutant

The proteomes of *Synechocystis* wild-type and of the Δ*hoxYH* mutant were analysed using the iTRAQ methodology, under conditions in which the hydrogenase activity is favoured – dark anoxic conditions – and the effect of the mutation should be easily observed (Cournac *et al.*, 2002, 2004). In total, 210 proteins were identified using Phenyx v2.6, of which there were 145 with two or more MS/MS spectra; false positive discovery rate was estimated to be 1.7 %. None of the bidirectional hydrogenase subunits was detected, probably due to their low abundance in the cells. To date, only Gan *et al.* (2005) have been able to identify a bidirectional hydrogenase subunit in proteomics studies. Out of the 210 proteins identified, 192 proteins were quantified, of which there were 131 with two or more MS–MS spectra. The methodology for quantification was described by Ow *et al.* (2009). The peptide matches and the protein identifications and quantifications are included in Supplementary Table S2, as recommended by Noirel *et al.* (2011). The quantified proteins were further analysed to determine which ones had their levels significantly altered by the Δ*hoxYH* mutation, according to the method used by Pham *et al.* (2010). According to this analysis, 17 proteins were differentially expressed, with a 95 % confidence interval, with five being less abundant in the mutant (Fig. 5 and Table 2). Sixteen of these proteins

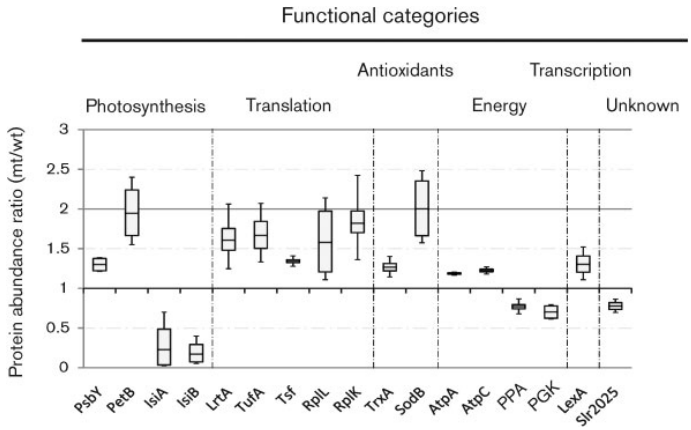


Fig. 5. Box-and-whisker plot of the abundance ratio between proteins showing significant differential fold changes in *Synechocystis* Δ*hoxYH* mutant and wild-type. The proteins are grouped according to their functional categories.

Table 2. List of proteins exhibiting significant differences in their abundance in *Synechocystis* Δ*hoxYH* mutant compared with the wild-type strain, after induction of dark anoxic conditions

Functional categories: PR, photosynthesis and respiration; Tl, translation; AD, antioxidant defences; EM, energy metabolism; Tr, transcription; U, unknown.

ID	Gene*	Protein	Putative function/characteristics of the encoded protein*	Category
sll0247	*isiA*	IsiA	Iron-stress chlorophyll-binding protein, homologous to *psbC* (CP43)	PR
sll0248	*isiB*	IsiB	Flavodoxin	PR
sll0947	*lrtA*	LrtA	Light repressed protein A homologue	Tl
sll1099	*tufA*	TufA	Elongation factor Tu	Tl
sll1261	*tsf*	Tsf	Elongation factor Ts	Tl
sll1326	*atpA*	AtpA	ATP synthase alpha chain	EM
sll1327	*atpC*	AtpC	ATP synthase gamma chain	EM
sll1626	*lexA*	LexA	Transcription factor	Tr
sll1743	*rpl11*	RplK	50S ribosomal protein L11	Tl
sll1746	*rpl12*	RplL	50S ribosomal protein L12	Tl
slr0342	*petB*	PetB	Cytochrome b_6	PR
slr0394	*pgk*	PGK	Phosphoglycerate kinase	EM
slr0623	*trxA*	TrxA	Thioredoxin	AD
slr1516	*sodB*	SodB	Iron superoxide dismutase	AD
slr1622	*ppa*	PPA	Soluble inorganic pyrophosphatase	EM
slr2025	*slr2025*	Slr2025	Hypothetical protein of unknown function DUF1821	U
sml0007	*psbY*	PsbY	Photosystem II protein Y	PR

*According to CyanoBase annotation.

fit into five functional categories: photosynthesis and respiration (IsiA, IsiB, PetB and PsbY), translation (LrtA, RplL, RplK, Tsf and TufA), antioxidant defences (SodB and TrxA), energy metabolism (AtpA, AtpC, PGK and PPA) and transcription (LexA). Additionally, a protein of unknown function was identified (Slr2025). The differential expression of each one of these proteins will be discussed in light of the putative roles of the bidirectional hydrogenase.

The association of the bidirectional hydrogenase activity with the photosynthetic and respiratory processes has been extensively discussed previously (e.g. Appel & Schulz, 1996, 1998; Appel *et al.*, 1999, 2000; Ludwig *et al.*, 2006; Schmitz & Bothe, 1996). In both cases, the bidirectional hydrogenase would serve as an important element for a proper redox balance of the electron-transport chains in transition states (Appel & Schulz, 1998). Our results revealed the higher abundance of a photosystem II (PSII) protein (PsbY) and of PetB in the Δ*hoxYH* mutant, indicating that both photosynthetic and respiratory machineries are being affected by the loss of hydrogenase activity. Moreover, the photosynthesis-associated proteins IsiA and IsiB, which are expressed under iron-deficiency conditions, were less abundant in the Δ*hoxYH* mutant, suggesting a higher bioavailability of this metal in the mutant. Since the cyanobacterial hydrogenases are metalloproteins possessing iron–sulfur clusters and nickel–iron active sites, it is likely that the absence of this enzyme could contribute to higher iron content in the mutant cells. It has been suggested that, in *Synechocystis*, the *isiAB* operon is under the control of the ferric uptake regulator (Fur) (Kunert *et al.*, 2003; Vinnemeier *et al.*, 1998). The importance of Fur is not limited to its participation as a modulator in all the iron-dependent-regulation genes; it has been shown that Fur is involved in the oxidative stress response (Nunoshiba *et al.*, 1999) and in overcoming acidic stress situations (Hall & Foster, 1996). Therefore, the lower abundance of IsiA and IsiB in the Δ*hoxYH* mutant could be related to a stress condition that the cells have to cope with.

The higher abundance of the five proteins involved in translation in the Δ*hoxYH* mutant indicates that protein synthesis is altered. Two of these proteins are structural components of the ribosome (RplK and RplL) and two others are involved in the process of translation itself (TufA and Tsf), indicating a reinforcement of the translational machinery in the mutant. Additionally, LrtA belongs to the σ^{54} modulation protein/ribosomal protein S30EA family and is also more abundant in the Δ*hoxYH* mutant. Although no clear function has been attributed to this protein, its homologue in *E. coli* – the ribosome binding protein Y (YfiA) – putatively works as a stress-response protein that binds the ribosomal subunit interface and arrests translation (Agafonov *et al.*, 2001; Vila-Sanjurjo *et al.*, 2004). The increased levels of LrtA in the mutant suggest a tighter control of translation, as a response to the stress condition caused by the absence of the bidirectional hydrogenase. Moreover, according to CyanoBase, this protein is assigned to the functional category of 'adaptations and atypical conditions'. Furthermore, a large-scale analysis of *Synechocystis* protein–protein interaction (Sato

et al., 2007) reported that LrtA interacts with a hypothetical protein encoded by *sll1773*, which is a Pirin-like protein induced under stress conditions (Hihara *et al.*, 2004).

Interestingly, even after oxygen has been removed from the medium, *Synechocystis* Δ*hoxYH* mutant cells still display increased levels of the enzymes SodB and TrxA, which act as antioxidant defences. SodB, the iron superoxide dismutase (Fe–SOD), catalyses the elimination of superoxide radicals ($O_2^{-\cdot}$). The higher levels of this protein may indicate a more reductive intracellular environment, leading to an increased concentration of toxic oxygen reduction products, and therefore, the necessity for higher levels of antioxidant defences. Furthermore, Nefedova *et al.* (2003) reported that mutants lacking SodB are non-viable under standard photoautotrophic conditions, emphasizing the importance of the Fe–SOD for providing resistance of *Synechocystis* cells to oxidative stress. Thioredoxins constitute a large family of redox-active enzymes, capable of reducing protein disulphides, and participate in a variety of processes, such as enzyme modulation, donation of reducing equivalents and signal transduction. Lindahl & Florencio (2003) have identified several proteins in *Synechocystis* likely to be thioredoxin (TrxA) substrates, mainly enzymes participating in anabolic processes. Among these is the elongation factor EF-Tu (TufA), which has been reported to interact with chloroplast thioredoxins (Motohashi *et al.*, 2001). Both TrxA and TufA levels are increased in the Δ*hoxYH* mutant compared with the wild-type. It has been suggested that TrxA is regulated by the photosynthetic electron transport (Navarro *et al.*, 2000; Pérez-Pérez *et al.*, 2009); therefore, the differential expression of TrxA protein may be due to a higher redox imbalance in the mutant. Again, significant changes in proteins related to defence mechanisms suggest the mutant is coping with a stress generated by the lack of the bidirectional hydrogenase, and this stress is likely to be related to the redox state of the cell.

Regarding the energy-metabolism-related proteins, two subunits of the ATP synthase (AtpA and AtpC) were more abundant in the Δ*hoxYH* mutant, while the phosphoglycerate kinase (PGK) and soluble inorganic pyrophosphatase (PPA) were less abundant. The ATP synthase is a ubiquitous membrane enzyme complex that functions to couple ATP synthesis to proton translocation across a membrane and, in cyanobacteria, this complex is present in both thylakoid and plasma membranes (Sherman *et al.*, 1994). The higher abundance of the two ATP synthase subunits suggests a higher requirement for this enzyme, under the conditions tested. Since the mutant lacks an active bidirectional hydrogenase, which is responsible for the reduction of protons to molecular hydrogen under anoxic conditions, one can hypothesize that the higher abundance of the ATP synthase is a cellular mechanism to cope with the excess of H^+ in order to maintain intracellular pH homeostasis. This mechanism has already been described for cell adaptations to changes in pH (Padan *et al.*, 2005; Summerfield & Sherman, 2008). PGK is the enzyme catalysing the reversible phosphorylation of 3-phosphoglycerate to 1,3-diphosphoglycerate. It is involved in the Calvin cycle, in glycolysis and in the glycolytic degradation of glycogen (fermentation). In the filamentous cyanobacterium *Arthrospira* (*Spirulina*) *maxima*, H_2 evolution represents a major pathway for energy (ATP) production during fermentation by regenerating the NAD^+ essential for the glycolytic degradation of glycogen and the catabolism of other substrates (Ananyev *et al.*, 2008). The lower abundance of PGK in the mutant under dark anoxic conditions can, therefore, be related to a decrease in the glycogen catabolism, due to a lower availability of NAD^+. PPA is an essential and ubiquitous metal-dependent enzyme that is thought to be involved in the removal of inorganic pyrophosphate (PPi), which is a by-product of many vital anabolic reactions (Kornberg, 1962). It has been shown that PPA function is essential for cell viability and that this protein is more abundant in *Synechocystis* cells subjected to phosphate deprivation (Gómez-García *et al.*, 2003). Accordingly, the lower abundance of PPA in the mutant suggests a higher availability of phosphate within the cells.

The transcription factor LexA is encoded in almost every bacterial group with a wide range of evolutionary distances and plays an important role in the SOS response in *E. coli* (Fernández de Henestrosa *et al.*, 2000; Little & Mount, 1982). Still, its precise functions in each group/species are largely unknown. It has been shown that in several cyanobacteria this protein interacts with the promoter region of the bidirectional hydrogenase (Antal *et al.*, 2006; Gutekunst *et al.*, 2005; Oliveira & Lindblad, 2005, 2009; Sjöholm *et al.*, 2007) and it has been suggested that LexA is a mediator of the redox state in *Synechocystis* cells (Antal *et al.*, 2006; Patterson-Fortin *et al.*, 2006). Interestingly, the LexA regulator was more abundant in the Δ*hoxYH* mutant. Since it was proposed that LexA acts as a transcriptional activator of the *hox* genes in *Synechocystis* (Gutekunst *et al.*, 2005), the higher abundance of LexA could be a cellular response to the lack of a functional bidirectional hydrogenase. It can be postulated that a feedback loop leads to an increased quantity of activator that in the wild-type would promote the transcription of the *hox* genes. Moreover, a predictive analysis of LexA binding sites on 26 sequenced cyanobacterial genomes (Li *et al.*, 2010) showed that this protein could regulate the expression of some of the genes reported in this study: *psbY*, *ppa*, *pgk*, *petB*, *isiB*, *tsf*, *trxA*, *rpl12*, *rpl11* and *lexA* itself.

According to a similarity BLAST analysis (Johnson *et al.*, 2008), the protein Slr2025 belongs to the DUF1821 protein family, which comprises uncharacterized proteins principally found in cyanobacteria with subunit and dimer structures similar to the 'Type III secretory system chaperone' from proteobacteria. In a large-scale protein–protein interaction analysis in *Synechocystis* (Sato *et al.*, 2007) using the yeast two-hybrid (YTH) system it was revealed that Slr2025 interacts with four other proteins: Slr2037 (unknown protein), Slr1444 (hypothetical protein), Sll1875 (HO2, haem oxygenase) and notably Sll1226

(HoxH, hydrogenase large subunit of the bidirectional hydrogenase).

Overall, the iTRAQ results indicate that, under dark anoxic conditions, *Synechocystis* cells lacking a functional bidirectional hydrogenase have to cope with a more stressful intracellular environment. Nonetheless, the cellular responses triggered seem to be sufficient to avoid a major interference in its metabolism as a whole. These conclusions are consistent with our growth experiments that showed no significant differences between the wild-type and Δ*hoxYH* mutant, under the conditions tested. Moreover, the proteins with altered levels are mainly related to the redox and energy state of the cells, corroborating the previously attributed role of the bidirectional hydrogenase in avoiding an overload of low potential electrons in the electron transport chains (Appel & Schulz, 1998; Appel *et al.*, 2000; Carrieri *et al.*, 2011; Cournac *et al.*, 2002, 2004).

Pathway analysis

The analysis based on 'mixture model on graphs' (MMG) provided some insight into the biological pathways that may undergo subtler, yet wider regulation patterns (see Supplementary Fig. S1 and Supplementary Table S2, available with the online version of this paper). This reveals and predicts changes in the central metabolism that had not been detected using the traditional protein-to-protein analysis of the previous section. Although individual protein quantifications are questionable, overall the evidence at the pathway level seems to point at a general downregulation of the pentose phosphate pathway (PPP), of glycolysis/gluconeogenesis and of the glyceraldehyde-3-phosphate-to-pyruvate route (see Fig. 6). This is of particular interest if we consider the putative involvement of the bidirectional hydrogenase in fermentation. It has been proposed that during fermentation in *Cyanothece* sp. PCC 7822, protons may serve as terminal electron acceptors, being reduced to molecular hydrogen in a hydrogenase-catalysed reaction (van der Oost *et al.*, 1989). In addition, two possible fermentative pathways are: (i) the heterofermentative lactate fermentation, via the pentose phosphate pathway; and (ii) the CoA-dependent decarboxylation of pyruvate, via glycolysis (Stal & Moezelaar, 1997; van der Oost *et al.*, 1989). In a recent study, McNeely *et al.* (2010) reported the successful improvement of hydrogen production in *Synechococcus* sp. PCC 7002, by redirecting the reductant flux, through metabolic engineering. In their study, the gene encoding D-lactate dehydrogenase (*ldhA*), responsible for the formation of the main fermentation product (lactate) was inactivated, resulting in the increase of other metabolic by-products, namely hydrogen. Interestingly, the *ldhA* mutant did not exhibit significant differences in terms of photoautotrophic growth in aerobic respiration either. In our study, in the absence of an active hydrogenase, capable of producing hydrogen to remove the excess reducing equivalents produced during the fermentative degradation of carbohydrates, it is reasonable to expect (i) a downregulation of the pathways originating this reducing power (namely glycolysis and PPP) and/or (ii) an increase in other metabolic by-products. This has to be further investigated, through experiments targeting the fermentative metabolism of the wild-type and the Δ*hoxYH* mutant.

Transcription analysis of genes encoding proteins with different abundance in *Synechocystis* wild-type and Δ*hoxYH* mutant

To assess if the changes observed at the protein levels were due to transcriptional regulation, RT-qPCR analyses were carried out. This experiment covers all gene clusters encoding the proteins showing different relative levels reported above, with the exception of *isiAB* and *tsf*, for which the PCR conditions could not be optimized (secondary amplification products were always detected in the melting curve analysis). For this experiment, three biological and three technical replicates were used (see Methods for further details). Fig. 7 shows the transcriptional analysis of the genes encoding proteins representative of each functional category (full analysis is shown in Supplementary Fig. S2, available with the online version of this paper; graphics were constructed based on one biological replicate representative of the biological variability observed). In all cases, a two-way ANOVA analysis was performed (Table 3), revealing significant differences between the wild-type and the Δ*hoxYH* mutant in the transcription of *psbY*, *petB*, *slr2025*, *ppa*, *pgk* and *lexA*. The transcription patterns of the genes *ppa*, *pgk*, *lexA* and *slr2025* seem to be in agreement with the differences found in the iTRAQ experiment, while for the other two genes (*psbY* and *petB*), the changes observed in protein abundance do not seem to be directly related to their transcriptional regulation and, most likely, the regulation is exerted at post-transcriptional and/or protein synthesis/degradation levels.

Additionally, this experiment revealed significant variations in the transcript levels of all genes under the physiological conditions tested (dark-anaerobiosis/light-aerobiosis). Overall, the genes exhibited one of the following transcription patterns: (i) induction under dark anoxic conditions and repression by light (*sodB*); (ii) repression under dark anoxic conditions and induction by light (*atpC*, *psbY* and *trxA*); (iii) induction under dark anoxic conditions and unaffected by light (*lexA*); (iv) repression under dark anoxic conditions and unaffected by light (*petB*, *pgk*, *ppa*, *slr2025* and *tufA*); and (v) transient induction under dark anoxic conditions and unaffected by light (*lrtA* and *rplK*). The transcription patterns of the genes grouped in (i), (ii) and (iii), which presented greater transcript levels variations, and *petB* that presented an inverse correlation with the proteomics analysis, will be discussed.

Interestingly, *sodB* transcript levels increased after anaerobiosis induction and decreased when cells were transferred back to the light. These results are unexpected, since it has been previously shown that, in *Synechocystis*, *sodB* expression is light-dependent, being downregulated in the

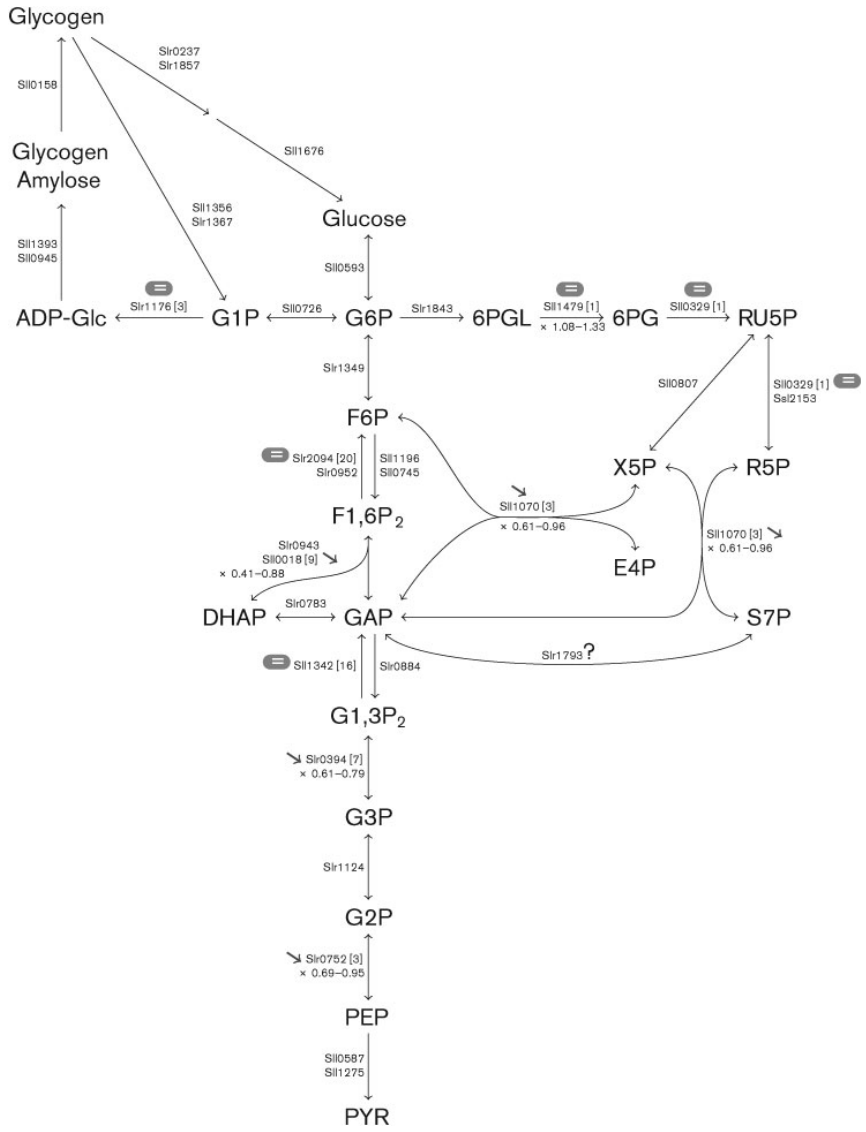


Fig. 6. Protein changes in the glycolysis/gluconeogenesis, the PPP pathway and glycogen metabolism in *Synechocystis* Δ*hoxYH* mutant according to the MMG analysis. The arrows indicate underexpression of enzymes in the mutant and the equals signs indicate expression levels that are not significantly different. The number of MS/MS spectra used for quantification is indicated in brackets (adapted from Osanai *et al.*, 2005). Abbreviations: G1P, glucose-1-phosphate; G6P, glucose 6-phosphate; 6PGL, 6-phosphogluconolactone; 6PG, 6-phosphogluconate; F6P, fructose-6-phosphate; $F1,6P_2$, fructose 1,6-bisphosphate; GAP, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; $G1,3P_2$, 1,3-bisphosphoglycerate; G3P, 3-phosphoglycerate; G2P, 2-phosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; RU5P, ribulose-5-phosphate; R5P, ribose 5-phosphate; S7P, sedoheptulose 7-phosphate; X5P, xylulose 5-phosphate; E4P, erythrose 4-phosphate.

darkness (Kim & Suh, 2005; Ushimaru *et al.*, 2002). Still, an increase in the transcript levels of *sodB* induced by anaerobiosis has already been described for other organisms, although SodB is known to act as an antioxidant defence (Dubrac & Touati, 2000; Kanematsu & Sato, 2008). The increase in *sodB* transcript levels was attributed to the Fur regulator, which may play an important role in the oxidative stress response, as referred to above. This may be a mechanism for the cell to prepare and protect itself from an eventual O_2 burst and the concomitant production of

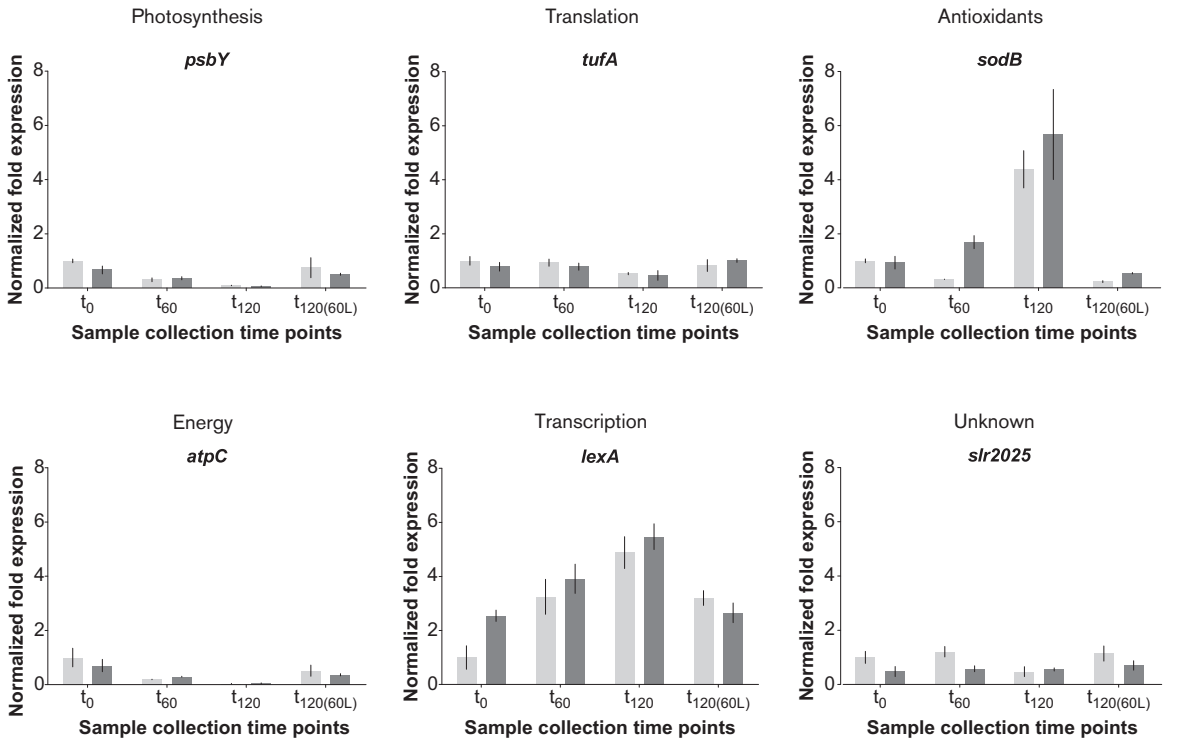


Fig. 7. Transcription profiles of selected genes encoding proteins (one of each functional category, see Fig. 5) showing significant differential fold changes in *Synechocystis* wild-type (light grey) and Δ*hoxYH* mutant (dark grey) and evaluated by RT-qPCR. The mean normalized expression was calculated relative to the transcription of the 16S rRNA gene and the reaction internal normalization was performed using the sample t_0 from *Synechocystis* wild-type. Error bars represent SD.

Table 3. Statistical significance of the data obtained in the RT-qPCR analysis, comparing *Synechocystis* Δ*hoxYH* mutant (mt) with the wild-type (wt) under dark anoxic conditions (Dark) and after transferring cells back to light (Light)

Gene	*P*-values mt/wt	
	Dark*	Light†
psbY	<0.05/<0.001	NS/<0.05
petB	<0.05/<0.01	NS/NS
lrtA	NS/<0.01	NS/NS
tufA	NS/<0.01	NS/NS
rplK	NS/<0.001	NS/NS
trxA	NS/<0.05	NS/<0.01
sodB	NS/<0.001	<0.001/<0.01
atpC	NS/<0.01	NS/<0.05
ppa	<0.01/<0.001	NS/NS
pgk	<0.05/<0.001	NS/NS
lexA	<0.05/<0.001	NS/NS
slr2025	<0.01/<0.05	<0.01/NS

*Two-way ANOVA analysis of t_0, t_{60} and t_{120}.
†Two-way ANOVA analysis of t_{60} and $t_{120(60L)}$.

reactive oxygen species. However, SodB activity did not present major differences before and after light transition, as it was observed in both gel and spectrophotometric superoxide dismutase activity assays (data not shown).

Cyanobacteria switch immediately from photoautotrophy to fermentation when exposed to dark anoxic conditions and, under these conditions, it is expected that the ATP synthase transcription is downregulated. It is generally assumed that the energy yield of fermentation is minimal, thus sustaining only survival (Stal & Moezelaar, 1997). This implies that biosynthetic processes are switched off during anaerobic dark metabolism in order to decrease energy demand. When the cells are transferred back to light, photosynthesis is reactivated and the requirement for ATP synthase may induce the transcription of the genes coding for its subunits. Moreover, the gene *atpC* encodes the gamma subunit of the ATP synthase, and AtpC is believed to be involved in the regulation of the ATPase activity and the flow of protons through the membrane-embedded domain of the enzyme (CF_0 complex) (Krenn *et al.*, 1997). Similarly, when cells are transferred into dark conditions, there is a decrease in the transcription of *psbY*, which encodes a core component of PSII (Kawakami *et al.*, 2007).

This is probably due to the cessation of the photosynthesis and downregulation of the transcription of components of the photosystems. A small increase in *psbY* transcript levels can be observed when cells were exposed back to light. In accord with what has been previously described by Navarro *et al.* (2000) and by Pérez-Pérez *et al.* (2009), *trxA* transcription slightly decreases in the dark and increases when cells are transferred to light, exhibiting photosynthetic dependence.

LexA seems to have a broad effect on gene regulation in *Synechocystis*, and although its precise function is unknown, LexA does not seem to be a regulator of the SOS response in this organism, as has been proposed for other cyanobacteria (Li *et al.*, 2010). Several studies reported that *lexA* transcription patterns do not always follow the same trend as those of *hox* genes (Oliveira & Lindblad, 2008; Zhang *et al.*, 2008), indicating that additional factors may be involved in the regulation of the transcription of the bidirectional hydrogenase genes of *Synechocystis* (Oliveira & Lindblad, 2008, 2009). However, our results are in agreement with previous studies that reported an increase in *lexA* transcript levels under microaerobic conditions (Kiss *et al.*, 2009; Summerfield *et al.*, 2011). In fact, in our study, after induction of dark anoxic conditions, an increase in the transcription of *lexA* can be observed (see Fig. 7). The fact that the transcription of both *lexA* and *hox* genes increases under anaerobiosis (Oliveira & Lindblad, 2008) and that LexA interacts with the promoter region of the *hox* genes (Antal *et al.*, 2006; Gutekunst *et al.*, 2005; Oliveira & Lindblad, 2005, 2009; Sjöholm *et al.*, 2007), suggests that LexA activates the transcription of the *hox* genes particularly under dark anoxic conditions.

The *petB* gene encodes the cytochrome b_6 subunit of the cytochrome b_6f complex that mediates electron transfer between PSII and PSI, cyclic electron flow around PSI and state transitions (Cramer *et al.*, 1994; Kruip *et al.*, 1994; Stroebel *et al.*, 2003). As has been previously reported for *Synechococcus* sp. PCC 7002 (Brand *et al.*, 1992), *petB* transcripts decrease when cells are placed in dark conditions. Overall, the transcript levels of *petB* are lower in the Δ*hoxYH* mutant, in contrast with the results obtained at the protein level (which was approximately twofold higher in the mutant), again indicating post-transcriptional regulation. Our results on the transcription of both *psbY* and *petB* are in disagreement with a study performed using another hydrogenase mutant (Δ*hoxH*), where the amounts of the two transcripts were slightly higher than in the wild-type (Appel *et al.*, 2000). However, in that work, Northern blot hybridizations were performed using RNA isolated from 7-day-old cultures, grown under light and bubbled with air; therefore a direct comparison with the present study is not possible. Nonetheless, both studies concur that the hydrogenase has an effect on the regulation of the *petB* gene.

Interestingly, a recent study on *Synechocystis* gene expression under low oxygen conditions, revealed changes in a number of genes coding for proteins putatively involved in assembling or stabilizing PSII, and an upregulation of the *hox* operon and *lexA* (Summerfield *et al.*, 2011). These responses, mostly focused on PSII and overall redox control, suggest an involvement of the bidirectional hydrogenase in the cell homeostasis under specific conditions or transition states.

A peculiar observation, regarding the genes coding for the proteins showing differential abundance (iTRAQ analysis), is that most of them are located in the same half of the *Synechocystis* chromosome, surrounding the *hox* cluster (see Supplementary Fig. S3, available with the online version of this paper), with a binomial test giving a *P*-value <0.013. However, the biological relevance (if any) of this occurrence needs to be further investigated.

In addition, the transcription of genes encoding proteins related to stress response and iron levels was studied: two ferric uptake regulation proteins – Fur (encoded by *sll0567* and *sll1937*) and the Fur-like peroxide-responsive repressor PerR (*slr1738*) (Kobayashi *et al.*, 2004), the putative glutamate-cysteine ligase GshA, which is the enzyme responsible for the first step in the glutathione biosynthesis pathway (*slr0990*), a small heat-shock protein, HspA, that stabilizes the thylakoid membrane proteins under oxidative stress (*sll1514*) (Sakthivel *et al.*, 2009) and the ferritin-type storage complex MrgA (*slr1894*), suggested to play a role in the coordination of iron homeostasis and oxidative stress response in *Synechocystis* (Shcolnick *et al.*, 2009). The transcription of *mrgA* is suggested to be under the control of the PerR regulator and, therefore, induced under oxidative stress (Li *et al.*, 2004). Overall, the RT-qPCR results showed no significant differences between the wild-type and the Δ*hoxYH* mutant, with the exception of the Fur regulator genes. Significantly ($P<0.05$) lower levels of *sll0567* transcript (in the dark) and a transient decrease in *sll1937* transcript levels (60 min after dark anoxic condition induction) could be observed in the mutant (see Supplementary Fig. S4, available with the online version of this paper). Since *sll0567* and *sll1937* encode transcription factors involved in iron homeostasis, these results suggest differences in iron pools/availability in the wild-type and the Δ*hoxYH* mutant.

Conclusions

In conclusion, a *Synechocystis* mutant lacking an active bidirectional hydrogenase was produced by deleting the two genes – *hoxY* and *hoxH* – encoding the hydrogenase part of this enzyme. Physiological studies were performed by growing the wild-type strain and the Δ*hoxYH* mutant under different physiological conditions and assessing and comparing their growth behaviour. From this analysis, no significant differences were observed, and therefore it can be assumed that the mutant exhibits similar growth behaviour to the wild-type. Moreover, changes in protein content were analysed, in conditions where the hydrogenase activity is favoured, revealing that 17 proteins had significant differential fold changes. These proteins are

mainly related to the redox and energy state of the cells. Twelve proteins were more abundant in the mutant, including one of the possible transcription regulators of the bidirectional hydrogenase of *Synechocystis*. Overall, the functions of the proteins exhibiting different relative levels correlate with the proposed roles of the bidirectional hydrogenase in dark fermentation and photosynthesis (acting as an electron valve), as well as a role in the maintenance of cell homeostasis under specific conditions (transition states). In addition, transcription studies were performed to assess the patterns of 12 genes coding for proteins showing differential abundance. From this analysis, six genes exhibited significant differences in transcript levels following the same pattern as the protein abundance, with the exception of *psbY* and *petB* that showed an inverse trend. Altogether, these results indicate an adjustment in the metabolism of the Δ*hoxYH* mutant allowing growth at similar rates to the wild-type and reflecting the plasticity and metabolic adaptability of *Synechocystis*. Finally, this study suggests that the *Synechocystis* Δ*hoxYH* mutant is robust and can be used as a photoautotrophic chassis for the integration of synthetic devices to fulfil specific purposes, namely hydrogen production.

ACKNOWLEDGEMENTS

This work was financially supported by EU FP6-NEST-2005-Path-SYN project BioModularH2 (contract no. 043340); Fundação para a Ciência e a Tecnologia (SFRH/BD/36378/2007, SFRH/BPD/64095/2009, PTDC/BIA-MIC/100370/2008); European Science Foundation (III Quadro Comunitário de Apoio), COMPETE – Programa Operacional Factores de Competitividade na sua componente FEDER; Acções Integradas Luso-Britânicas, Treaty of Windsor Programme 2010-11 (Processo B18/10); Engineering and Physical Sciences Research Council (EP/E036252/1), under the ChELSI initiative; Generalitat Valenciana grant BFPI/2007/283; and MICINN project ArtBioCom (TIN2009-12359). The authors are grateful to Dr Alan Dunbar (Chemical and Biological Engineering, University of Sheffield) and Dr Saw Ow Yen (Chemical and Biological Engineering, University of Sheffield) for helpful discussion. The authors would also like to thank the 'National BioResource Project (NIG, Japan): *E. coli*' for providing the plasmid pK18mobsacB.

REFERENCES

Abed, R. M., Dobretsov, S. & Sudesh, K. (2009). Applications of cyanobacteria in biotechnology. *J Appl Microbiol* **106**, 1–12.

Agafonov, D. E., Kolb, V. A. & Spirin, A. S. (2001). Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage. *EMBO Rep* **2**, 399–402.

Ananyev, G., Carrieri, D. & Dismukes, G. C. (2008). Optimization of metabolic capacity and flux through environmental cues to maximize hydrogen production by the cyanobacterium "*Arthrospira* (*Spirulina*) *maxima*". *Appl Environ Microbiol* **74**, 6102–6113.

Angermayr, S. A., Hellingwerf, K. J., Lindblad, P. & Teixeira de Mattos, M. J. (2009). Energy biotechnology with cyanobacteria. *Curr Opin Biotechnol* **20**, 257–263.

Antal, T. K. & Lindblad, P. (2005). Production of H_2 by sulphur-deprived cells of the unicellular cyanobacteria *Gloeocapsa alpicola* and *Synechocystis* sp. PCC 6803 during dark incubation with methane or at various extracellular pH. *J Appl Microbiol* **98**, 114–120.

Antal, T. K., Oliveira, P. & Lindblad, P. (2006). The bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Int J Hydrogen Energy* **31**, 1439–1444.

Appel, J. & Schulz, R. (1996). Sequence analysis of an operon of a NAD(P)-reducing nickel hydrogenase from the cyanobacterium *Synechocystis* sp. PCC 6803 gives additional evidence for direct coupling of the enzyme to NAD(P)H-dehydrogenase (complex I). *Biochim Biophys Acta* **1298**, 141–147.

Appel, J. & Schulz, R. (1998). Hydrogen metabolism in organisms with oxygenic photosynthesis: hydrogenases as important regulatory devices for a proper redox poising? *J Photochem Photobiol B* **47**, 1–11.

Appel, J., Phunpruch, S. & Schulz, R. (1999). Hydrogenase(s) in *Synechocystis* – tools for Photohydrogen Production? In *BioHydrogen*, pp. 189–196. Edited by O. R. Zaborsky, J. R. Benemann, T. Matsunaga, J. Miyake & A. San Pietro. US: Springer.

Appel, J., Phunpruch, S., Steinmüller, K. & Schulz, R. (2000). The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Arch Microbiol* **173**, 333–338.

Barrios-Llerena, M. E., Chong, P. K., Gan, C. S., Snijders, A. P. L., Reardon, K. F. & Wright, P. C. (2006). Shotgun proteomics of cyanobacteria–applications of experimental and data-mining techniques. *Brief Funct Genomics Proteomics* **5**, 121–132.

Bashor, C. J., Horwitz, A. A., Peisajovich, S. G. & Lim, W. A. (2010). Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* **39**, 515–537.

Bhaya, D., Vaulot, D., Amin, P., Takahashi, A. W. & Grossman, A. R. (2000). Isolation of regulated genes of the cyanobacterium *Synechocystis* sp. strain PCC 6803 by differential display. *J Bacteriol* **182**, 5692–5699.

Bothe, H., Schmitz, O., Yates, M. G. & Newton, W. E. (2010a). Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiol Mol Biol Rev* **74**, 529–551.

Bothe, H., Tripp, H. J. & Zehr, J. P. (2010b). Unicellular cyanobacteria with a new mode of life: the lack of photosynthetic oxygen evolution allows nitrogen fixation to proceed. *Arch Microbiol* **192**, 783–790.

Brand, S. N., Tan, X. & Widger, W. R. (1992). Cloning and sequencing of the *petBD* operon from the cyanobacterium *Synechococcus* sp. PCC 7002. *Plant Mol Biol* **20**, 481–491.

Burja, A. M., Banaigs, B., Abou-Mansour, E., Burgess, J. G. & Wright, P. C. (2001). Marine cyanobacteria – a prolific source of natural products. *Tetrahedron* **57**, 9347–9377.

Carrieri, D., Wawrousek, K., Eckert, C., Yu, J. & Maness, P.-C. (2011). The role of the bidirectional hydrogenase in cyanobacteria. *Bioresour Technol* **102**, 8368–8377.

Chong, P. K., Gan, C. S., Pham, T. K. & Wright, P. C. (2006). Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J Proteome Res* **5**, 1232–1240.

Cournac, L., Mus, F., Bernard, L., Guedeney, G., Vignais, P. & Peltier, G. (2002). Limiting steps of hydrogen production in *Chlamydomonas reinhardtii* and *Synechocystis* PCC 6803 as analysed by light-induced gas exchange transients. *Int J Hydrogen Energy* **27**, 1229–1237.

Cournac, L., Guedeney, G., Peltier, G. & Vignais, P. M. (2004). Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. *J Bacteriol* **186**, 1737–1746.

Cramer, W. A., Martinez, S. E., Furbacher, P. N., Huang, D. & Smith, J. L. (1994). The cytochrome b_6f complex. *Curr Opin Struct Biol* **4**, 536–544.

Dembitsky, V. M. (2006). Anticancer activity of natural and synthetic acetylenic lipids. *Lipids* **41**, 883–924.

Devillers, J., Doré, J. C., Guyot, M., Poroikov, V., Gloriozova, T., Lagunin, A. & Filimonov, D. (2007). Prediction of biological activity profiles of cyanobacterial secondary metabolites. *SAR QSAR Environ Res* **18**, 629–643.

Dubrac, S. & Touati, D. (2000). Fur positive regulation of iron superoxide dismutase in *Escherichia coli*: functional analysis of the *sodB* promoter. *J Bacteriol* **182**, 3802–3808.

Elias, J. E. & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207–214.

Endy, D. (2005). Foundations for engineering biology. *Nature* **438**, 449–453.

Fernández de Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H. & Woodgate, R. (2000). Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol Microbiol* **35**, 1560–1572.

Ferreira, D., Pinto, F., Moradas-Ferreira, P., Mendes, M. V. & Tamagnini, P. (2009). Transcription profiles of hydrogenases related genes in the cyanobacterium *Lyngbya majuscula* CCAP 1446/4. *BMC Microbiol* **9**, 67.

Gan, C. S., Reardon, K. F. & Wright, P. C. (2005). Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of *Synechocystis* sp. PCC 6803. *Proteomics* **5**, 2468–2478.

Gan, C. S., Chong, P. K., Pham, T. K. & Wright, P. C. (2007). Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res* **6**, 821–827.

Gómez-García, M. R., Losada, M. & Serrano, A. (2003). Concurrent transcriptional activation of *ppa* and *ppx* genes by phosphate deprivation in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Biochem Biophys Res Commun* **302**, 601–609.

Gutekunst, K., Phunpruch, S., Schwarz, C., Schuchardt, S., Schulz-Friedrich, R. & Appel, J. (2005). LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. *Mol Microbiol* **58**, 810–823.

Gutthann, F., Egert, M., Marques, A. & Appel, J. (2007). Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta* **1767**, 161–169.

Hall, H. K. & Foster, J. W. (1996). The role of Fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition. *J Bacteriol* **178**, 5683–5691.

Hihara, Y., Muramatsu, M., Nakamura, K. & Sonoike, K. (2004). A cyanobacterial gene encoding an ortholog of Pirin is induced under stress conditions. *FEBS Lett* **574**, 101–105.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res* **36** (Web Server issue), W5–W9. doi:10.1093/dnares/10.5.221

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M. & other authors (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**, 109–136.

Kaneko, T., Nakamura, Y., Sasamoto, S., Watanabe, A., Kohara, M., Matsumoto, M., Shimpo, S., Yamada, M. & Tabata, S. (2003). Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Res* **10**, 221–228.

Kanematsu, S. & Sato, S. (2008). Cloning of Fe-superoxide dismutase gene from the diazotroph *Azotobacter vinelandii* and the stimulation of its expression under anaerobic conditions in *Escherichia coli*. *Bull Minamikyushu Univ* **38A**, 7–18.

Kawakami, K., Iwai, M., Ikeuchi, M., Kamiya, N. & Shen, J.-R. (2007). Location of PsbY in oxygen-evolving photosystem II revealed by mutagenesis and X-ray crystallography. *FEBS Lett* **581**, 4983–4987.

Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3**, 64–76.

Khalil, A. S. & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367–379.

Kim, J.-H. & Suh, K. H. (2005). Light-dependent expression of superoxide dismutase from cyanobacterium *Synechocystis* sp. strain PCC 6803. *Arch Microbiol* **183**, 218–223.

Kiss, É., Kós, P. B. & Vass, I. (2009). Transcriptional regulation of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* 6803. *J Biotechnol* **142**, 31–37.

Kobayashi, M., Ishizuka, T., Katayama, M., Kanehisa, M., Bhattacharyya-Pakrasi, M., Pakrasi, H. B. & Ikeuchi, M. (2004). Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol* **45**, 290–299.

Koksharova, O. A. & Wolk, C. P. (2002). Genetic tools for cyanobacteria. *Appl Microbiol Biotechnol* **58**, 123–137.

Kornberg, A. (1962). On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions. In *Horizons in Biochemistry: Albert Szent-Györgyi Dedicatory Volume*, pp. 251–264. Edited by M. Kasha & B. Pullman. New York: Academic Press, Inc.

Krenn, B. E., Strotmann, H., Van Walraven, H. S., Scholts, M. J. C. & Kraayenhof, R. (1997). The ATP synthase γ subunit provides the primary site of activation of the chloroplast enzyme: experiments with a chloroplast-like *Synechocystis* 6803 mutant. *Biochem J* **323**, 841–845.

Kruip, J., Nixon, P. J., Osiewacz, H. D. & Rögner, M. (1994). Nucleotide sequence of the *petB* gene encoding cytochrome b_6 from the mesophilic cyanobacterium *Synechocystis* PCC 6803: implications for evolution and function. *Biochim Biophys Acta* **1188**, 443–446.

Kunert, A., Vinnemeier, J., Erdmann, N. & Hagemann, M. (2003). Repression by Fur is not the main mechanism controlling the iron-inducible *isiAB* operon in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* **227**, 255–262.

Leitão, E., Pereira, S., Bondoso, J., Ferreira, D., Pinto, F., Moradas-Ferreira, P. & Tamagnini, P. (2006). Genes involved in the maturation of hydrogenase(s) in the nonheterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4. *Int J Hydrogen Energy* **31**, 1469–1477.

Li, H., Singh, A. K., McIntyre, L. M. & Sherman, L. A. (2004). Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp. strain PCC 6803. *J Bacteriol* **186**, 3331–3345.

Li, S., Xu, M. & Su, Z. (2010). Computational analysis of LexA regulons in Cyanobacteria. *BMC Genomics* **11**, 527.

Lindahl, M. & Florencio, F. J. (2003). Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different. *Proc Natl Acad Sci U S A* **100**, 16107–16112.

Little, J. W. & Mount, D. W. (1982). The SOS regulatory system of *Escherichia coli*. *Cell* **29**, 11–22.

Ludwig, M., Schulz-Friedrich, R. & Appel, J. (2006). Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. *J Mol Evol* **63**, 758–768.

Masukawa, H., Mochimaru, M. & Sakurai, H. (2002). Disruption of the uptake hydrogenase gene, but not of the bidirectional hydrogenase

gene, leads to enhanced photobiological hydrogen production by the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120. *Appl Microbiol Biotechnol* **58**, 618–624.

McIntosh, C. L., Germer, F., Schulz, R., Appel, J. & Jones, A. K. (2011). The [NiFe]-hydrogenase of the cyanobacterium *Synechocystis* sp. PCC 6803 works bidirectionally with a bias to H_2 production. *J Am Chem Soc* **133**, 11308–11319.

McNeely, K., Xu, Y., Bennette, N., Bryant, D. A. & Dismukes, G. C. (2010). Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. *Appl Environ Microbiol* **76**, 5032–5038.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P. & Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. *Science* **327**, 1512–1514.

Motohashi, K., Kondoh, A., Stumpp, M. T. & Hisabori, T. (2001). Comprehensive survey of proteins targeted by chloroplast thioredoxin. *Proc Natl Acad Sci U S A* **98**, 11224–11229.

Nakamura, Y., Kaneko, T. & Tabata, S. (2000). CyanoBase, the genome database for *Synechocystis* sp. strain PCC6803: status for the year 2000. *Nucleic Acids Res* **28**, 72.

Navarro, F., Martín-Figueroa, E. & Florencio, F. J. (2000). Electron transport controls transcription of the thioredoxin gene (*trxA*) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol Biol* **43**, 23–32.

Nefedova, L. N., Mel'nik, V. A. & Babykin, M. M. (2003). Mutants of cyanobacterium *Synechocystis* sp. PCC 6803 with insertional inactivation of the *sodB* gene encoding Fe-superoxide dismutase. *Russ J Genet* **39**, 386–389.

Noirel, J., Sanguinetti, G. & Wright, P. C. (2008). Identifying differentially expressed subnetworks with MMG. *Bioinformatics* **24**, 2792–2793.

Noirel, J., Ow, S. Y., Sanguinetti, G. & Wright, P. C. (2009). Systems biology meets synthetic biology: a case study of the metabolic effects of synthetic rewiring. *Mol Biosyst* **5**, 1214–1223.

Noirel, J., Evans, C., Salim, M., Mukherjee, J., Ow, S. Y., Pandhal, J., Pham, T. K., Biggs, C. A. & Wright, P. C. (2011). Methods in quantitative proteomics: setting iTRAQ on the right track. *Curr Proteomics* **8**, 17–30.

Nunoshiba, T., Obata, F., Boss, A. C., Oikawa, S., Mori, T., Kawanishi, S. & Yamamoto, K. (1999). Role of iron and superoxide for generation of hydroxyl radical, oxidative DNA lesions, and mutagenesis in *Escherichia coli*. *J Biol Chem* **274**, 34832–34837.

Oliveira, P. (2008). *Regulation of the Cyanobacterial Bidirectional Hydrogenase*, pp. 63. Uppsala: Acta Universitatis Upsaliensis.

Oliveira, P. & Lindblad, P. (2005). LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* **251**, 59–66.

Oliveira, P. & Lindblad, P. (2008). An AbrB-Like protein regulates the expression of the bidirectional hydrogenase in *Synechocystis* sp. strain PCC 6803. *J Bacteriol* **190**, 1011–1019.

Oliveira, P. & Lindblad, P. (2009). Transcriptional regulation of the cyanobacterial bidirectional Hox-hydrogenase. *Dalton Trans* (45), 9990–9996.

Osanai, T., Kanesaki, Y., Nakano, T., Takahashi, H., Asayama, M., Shirai, M., Kanehisa, M., Suzuki, I., Murata, N. & Tanaka, K. (2005). Positive regulation of sugar catabolic pathways in the cyanobacterium *Synechocystis* sp. PCC 6803 by the group 2 σ factor sigE. *J Biol Chem* **280**, 30653–30659.

Ow, S. Y. & Wright, P. C. (2009). Current trends in high throughput proteomics in cyanobacteria. *FEBS Lett* **583**, 1744–1752.

Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009). iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **8**, 5347–5355.

Padan, E., Bibi, E., Ito, M. & Krulwich, T. A. (2005). Alkaline pH homeostasis in bacteria: new insights. *Biochim Biophys Acta* **1717**, 67–88.

Patterson-Fortin, L. M., Colvin, K. R. & Owttrim, G. W. (2006). A LexA-related protein regulates redox-sensitive expression of the cyanobacterial RNA helicase, *crhR*. *Nucleic Acids Res* **34**, 3446–3454.

Pérez-Pérez, M. E., Martín-Figueroa, E. & Florencio, F. J. (2009). Photosynthetic regulation of the cyanobacterium *Synechocystis* sp. PCC 6803 thioredoxin system and functional analysis of TrxB (Trx x) and TrxQ (Trx y) thioredoxins. *Mol Plant* **2**, 270–283.

Pham, T. K., Roy, S., Noirel, J., Douglas, I., Wright, P. C. & Stafford, G. P. (2010). A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen *Tannerella forsythia*. *Proteomics* **10**, 3130–3141.

Pósfai, G., Plunkett, G., III, Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S. & other authors (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046.

Reidegeld, K. A., Eisenacher, M., Kohl, M., Chamrad, D., Körting, G., Blüggel, M., Meyer, H. E. & Stephan, C. (2008). An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. *Proteomics* **8**, 1129–1137.

Sakthivel, K., Watanabe, T. & Nakamoto, H. (2009). A small heat-shock protein confers stress tolerance and stabilizes thylakoid membrane proteins in cyanobacteria under oxidative stress. *Arch Microbiol* **191**, 319–328.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*. New York, USA: Cold Spring Harbor Laboratory Press.

Sanguinetti, G., Noirel, J. & Wright, P. C. (2008). MMG: a probabilistic tool to identify submodules of metabolic pathways. *Bioinformatics* **24**, 1078–1084.

Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y. & Tabata, S. (2007). A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res* **14**, 207–216.

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.

Schmitt, W. A., Jr & Stephanopoulos, G. (2003). Prediction of transcriptional profiles of *Synechocystis* PCC6803 by dynamic autoregressive modeling of DNA microarray data. *Biotechnol Bioeng* **84**, 855–863.

Schmitz, O. & Bothe, H. (1996). The diaphorase subunit HoxU of the bidirectional hydrogenase as electron transferring protein in cyanobacterial respiration? *Naturwissenschaften* **83**, 525–527.

Schütz, K., Happe, T., Troshina, O., Lindblad, P., Leitão, E., Oliveira, P. & Tamagnini, P. (2004). Cyanobacterial H_2 production – a comparative analysis. *Planta* **218**, 350–359.

Sharma, S. S., Blattner, F. R. & Harcum, S. W. (2007a). Recombinant protein production in an *Escherichia coli* reduced genome strain. *Metab Eng* **9**, 133–141.

Sharma, S. S., Campbell, J. W., Frisch, D., Blattner, F. R. & Harcum, S. W. (2007b). Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains. *Biotechnol Bioeng* **98**, 1056–1070.

Shcolnick, S., Summerfield, T. C., Reytman, L., Sherman, L. A. & Keren, N. (2009). The mechanism of iron homeostasis in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 and its relationship to oxidative stress. *Plant Physiol* **150**, 2045–2056.

Sherman, D. M., Troyan, T. A. & Sherman, L. A. (1994). Localization of membrane proteins in the cyanobacterium *Synechococcus* sp. PCC7942 (radial asymmetry in the photosynthetic complexes). *Plant Physiol* **106**, 251–262.

Sjöholm, J., Oliveira, P. & Lindblad, P. (2007). Transcription and regulation of the bidirectional hydrogenase in the cyanobacterium *Nostoc* sp. strain PCC 7120. *Appl Environ Microbiol* **73**, 5435–5446.

Stal, L. J. & Moezelaar, R. (1997). Fermentation in cyanobacteria. *FEMS Microbiol Rev* **21**, 179–211.

Stanier, R. Y., Kunisawa, R., Mandel, M. & Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriol Rev* **35**, 171–205.

Stroebel, D., Choquet, Y., Popot, J.-L. & Picot, D. (2003). An atypical haem in the cytochrome b_6f complex. *Nature* **426**, 413–418.

Summerfield, T. C. & Sherman, L. A. (2008). Global transcriptional response of the alkali-tolerant cyanobacterium *Synechocystis* sp. strain PCC 6803 to a pH 10 environment. *Appl Environ Microbiol* **74**, 5276–5284.

Summerfield, T. C., Nagarajan, S. & Sherman, L. A. (2011). Gene expression under low-oxygen conditions in the cyanobacterium *Synechocystis* sp. PCC 6803 demonstrates Hik31-dependent and -independent responses. *Microbiology* **157**, 301–312.

Suzuki, I., Simon, W. J. & Slabas, A. R. (2006). The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics. *J Exp Bot* **57**, 1573–1578.

Tamagnini, P., Troshina, O., Oxelfelt, F., Salema, R. & Lindblad, P. (1997). Hydrogenases in *Nostoc* sp. strain PCC 73102, a strain lacking a bidirectional enzyme. *Appl Environ Microbiol* **63**, 1801–1807.

Tamagnini, P., Costa, J. L., Almeida, L., Oliveira, M. J., Salema, R. & Lindblad, P. (2000). Diversity of cyanobacterial hydrogenases, a molecular approach. *Curr Microbiol* **40**, 356–361.

Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wünschiers, R. & Lindblad, P. (2002). Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol Mol Biol Rev* **66**, 1–20.

Tamagnini, P., Leitão, E., Oliveira, P., Ferreira, D., Pinto, F., Harris, D. J., Heidorn, T. & Lindblad, P. (2007). Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiol Rev* **31**, 692–720.

Thajuddin, N. & Subramanian, G. (2005). Cyanobacterial biodiversity and potential applications in biotechnology. *Curr Sci* **89**, 47–57.

Troshina, O., Serebryakova, L., Sheremetieva, M. & Lindblad, P. (2002). Production of H_2 by the unicellular cyanobacterium *Gloeocapsa alpicola* CALU 743 during fermentation. *Int J Hydrogen Energy* **27**, 1283–1289.

Ushimaru, T., Nishiyama, Y., Hayashi, H. & Murata, N. (2002). No coordinated transcriptional regulation of the *sod-kat* antioxidative system in *Synechocystis* sp. PCC 6803. *J Plant Physiol* **159**, 805–807.

van der Oost, J., Bulthuis, B. A., Feitz, S., Krab, K. & Kraayenhof, R. (1989). Fermentation metabolism of the unicellular cyanobacterium *Cyanothece* PCC 7822. *Arch Microbiol* **152**, 415–419.

Vila-Sanjurjo, A., Schuwirth, B.-S., Hau, C. W. & Cate, J. H. D. (2004). Structural basis for the control of translation initiation during stress. *Nat Struct Mol Biol* **11**, 1054–1059.

Vinnemeier, J., Kunert, A. & Hagemann, M. (1998). Transcriptional analysis of the *isiAB* operon in salt-stressed cells of the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* **169**, 323–330.

Williams, J. G. K. (1988). Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol* **167**, 766–778.

Xu, W. D. & McFadden, B. A. (1997). Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* **37**, 95–104.

Yadav, V. G. & Stephanopoulos, G. (2010). Reevaluating synthesis by biology. *Curr Opin Microbiol* **13**, 371–376.

Yang, X. Y. & McFadden, B. A. (1993). A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC 6803 encodes a Rep protein and replicates by a rolling circle mechanism. *J Bacteriol* **175**, 3981–3991.

Yang, X. Y. & McFadden, B. A. (1994). The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid* **31**, 131–137.

Zhang, Z., Pendse, N. D., Phillips, K. N., Cotner, J. B. & Khodursky, A. (2008). Gene expression patterns of sulfur starvation in *Synechocystis* sp. PCC 6803. *BMC Genomics* **9**, 344.

Edited by: C.-C. Zhang

Supplementary material I

SUPPLEMENTARY METHODS

For the proteomics analysis, standard growth conditions were used, with a scalar irradiance of 75 µE m^{-2} s^{-1} (measured with a QSL-2100 Radiometer, Biospherical Instruments Inc., and equivalent to the standard growth light intensity described in Methods section). Anaerobic conditions were induced by sparging argon for 30 min in the dark. Cells were then kept in the dark for 2 h before being harvested. Cells were collected by centrifugation, the supernatant (medium) removed, the pellets were snap-frozen in liquid nitrogen and preserved at -80 ºC prior to protein extraction. Cells were thawed at room temperature and resuspended in 0.5 M TEAB (triethylammonium bicarbonate), pH 8.5 and the proteins were extracted with mechanical cracking, using a mortar and pestle, and liquid nitrogen. The crude extract was centrifuged at 21000 *g* for 30 min at 4 °C. The supernatant, containing the soluble proteins, was transferred to protein LoBind tubes (Eppendorf) and stored at -80 °C. Proteins were quantified using the Bradford Assay (Bio-Rad) and checked by SDS-PAGE.

100 µg of proteins were diluted in 1 M TEAB, pH 8.5 to a final volume of 20 µL. Proteins were denatured using 2% SDS and reduced with 20 mM TCEP (tris(2-carboxyethyl)phosphine) for 1 h at 60 °C. Cysteines were blocked using 200 mM MMTS (methyl methanethiosulfonate) for 10 min at room temperature. The reduced and blocked proteins were digested overnight at 37 °C, using trypsin 1:10 (Promega) in 1 mM HCl. iTRAQ, 8-plex reagents (ABSciex) were resuspended in 100 µL isopropanol and added to the appropriate digested proteins. The peptides and labels were incubated at room temperature for 2 h. After incubation, all labeled peptide fractions were combined and split in two separate tubes, with each tube containing 400 µg of protein in total. These protein samples were dried using a vacuum concentrator and stored at -20 °C prior to fractionation. In this study, two biological *Synechocystis* wild-type replicates were labeled with the 117 and 118 reagents and two biological Δ*hoxYH* mutant replicates were labeled with the 119 and 121 reagents.

Dried peptides resuspended in 200 µL of Buffer A were fractionated using a PolySULFOETHYL A Column (PolyLC Inc.) of 100 x 4.6 mm, with a 3 µm particle size and 300 Å pore size, on a BioLC HPLC, with a GS50 gradient pump (Dionex). Peptides were eluted using two buffers in a linear gradient. Buffer A: 20% ACN (acetonitrile), 0.1% FA (formic acid) in water. Buffer B: 20% ACN, 0.1% FA, 0.5 M KCl in water. The gradient consisted of 16-19% B over 19 minutes, 19-25% B over 10 min, 25-100% B over 10 min at a flow rate of 0.4 $mL.min^{-1}$. Peptide elution was monitored using a UVD170U detector at a wavelength of 214 nm and with Chromeleon software, version 6.50 (Dionex/LC packings). Fractions were collected using a Foxy Junior (Dionex) fraction collector and dried using a vacuum concentrator. The dried fractions were stored at -20 °C prior to mass spectrometry analysis.

Each peptide fraction was resuspended in Switchos buffer (0.1% FA and 3% ACN) and three injections were made in the nano-LC-ESI-MS/MS system. In this system an Ultimate, Farnos and Switchos nano-LC (Dionex), containing a 0.075 x 150 mm analytical column (3 µm i.d., 100 Å, C18, Dionex-LC Packings) was coupled to a QSTAR XL ESI-Q-TOF-MS/MS mass spectrometer (ABSciex). A 90 min gradient elution was performed using buffer A (0.1 % FA and 3% ACN) and buffer B (0.1% FA and 97% ACN), during which buffer B increased from 3% to 35% at a flow rate of 400 $nL.min^{-1}$, followed by a 5 min rapid ramp to 95% of elution buffer B. MS survey scans were acquired at 1 Hz with a m/z range of 350-2000. During the TOF-MS scan, two dynamically selected precursors with a +2 or +3 charge state were isolated for CID fragmentation (scan range 65-2000 m/z) and accumulated 3 sec for product ion scans. The collision energy range was increased with 20% compared to the standard settings. Data was acquired and processed using Analyst QS version 2.0 (ABSciex).

Protein quantification and identification was carried out using Phenyx Software version 2.6 (Genebio). Spectra were searched against the Uniprot database (accessed August 2009), containing 3576 protein sequences of *Synechocystis*. The false discovery rate (FDR) was calculated using a reverse database generated with Phenyx (Reidegeld *et al.*, 2008), as

described by Elias and Gygi (2007). Both the true and the reversed database were searched in the same run. The FDR was below 5%, as recommended by Ow & Wright (2009). Searches were performed in 4 steps in order to increase proteome coverage.

- In step A, original spectra were analyzed (maximum missed cleavages 2, cleaved mode “normal”, minimal peptide length 6, minimal score 7, maximal p-value $1x10^{-6}$, AC-score 7);
- In step B, unmatched spectra were analyzed (minimal score 5, maximal p-value $1x10^{-4}$, AC-score 5);
- In step C, unmatched spectra were analyzed (cleaved mode “half”, minimal score 7, maximal p-value $1x10^{-6}$, AC-score 7);
- In step D, unmatched spectra were analyzed (minimal score 6, maximal p-value $1x10^{-4}$, AC-score 6).

Mass tolerances were set to 0.4 Da MS and 0.4 Da MS/MS. The peptide modifications were the same for each search (MMTS, fixed cysteine modification; oxidation, variable; iTRAQ_8plex_K, fixed; iTRAQ_8plex_N, fixed).

The peptide lists generated by Phenyx were analyzed as described in Ow *et al.* (2009). This list was used for further analysis as described in Pham *et al.* (2010) to produce a list of differentially expressed proteins. Protein function and annotations were assigned according to Cyanobase (Nakamura *et al.*, 2000).

Table S1. Oligonucleotide primers used in this study to perform additional RT-qPCR analyses of the transcription of genes related to stress response and iron levels

Gene ID	Protein encoded	Primer name	Sequence (5' → 3')
sll1514	Small 16.6 kDa heat shock protein HspA	sll1514F[a]	AACTGAAGAAGCCTATGTGC
		sll1514R[a]	CAAAGTCAAAGTTAGGATACCG
slr1894	Ferritin-type storage complex MrgA	slr1894F[a]	ACTGTTAGCCGATACCTACACC
		slr1894R[a]	GAATGCCGTCCACTTCTTGC
sll0567	Ferric uptake regulator Fur	sll0567F[b]	TCTCAGTGCCGAGGAGTTAC
		sll0567R[b]	CGATGGTGCGATTACATTGC
sll1937	Ferric uptake regulator Fur	sll1937F[a]	TAGAGTCCCTCACAGTCAAC
		sll1937R[a]	GGTTTGGTAAAGCAATTCCC
slr1738	Fur-like the Peroxide-responsive repressor PerR	slr1738F[a]	TCTATGCCAATCTACTCCAAC
		slr1738R[a]	CCAATCCACATCCTCTATCG
slr0990	Glutamate-cysteine ligase GshA	slr0990F[a]	GTAACAGGCTACCATTCCAG
		slr0990R[a]	AATCGCAGATCCGTAATTCC

a – Annealing temperature used: 51 ºC.
b – Annealing temperature used: 55 ºC.

Table S2. List of proteins extracted from the MMG analysis comparing *Synechocystis* Δ*hoxYH* mutant with the wild-type (see Fig. S1), along with quantification, and posterior probabilities (Noirel *et al.*, 2008; Sanguinetti *et al.*, 2008).

Node	Data (log)	Data	P(+)	P(0)	P(-)	Description
Down regulation						
56	N/A	N/A	0,299386	0,297596	0,403018	phosphoglucomutase
78	-0,26618	1,20262	0	0	1	transketolase
266	-0,07967	1,056777	0	0,002062	0,997938	glucose-1-phosphate adenylyltransferase
392	N/A	N/A	0,271578	0,298708	0,429714	ribose-5-phosphate isomerase A; [...] phosphate isomerase B
60	N/A	N/A	0,2514	0,278954	0,469646	ribulose-phosphate 3-epimerase
111	-0,10452	1,075133	0	0,19344	0,80656	phosphoribulokinase
344	-0,50774	1,421821	0	0,000718	0,999282	fructose-bisphosphate aldolase
188	-0,36143	1,284699	0	0	1	phosphoglycerate kinase
174	N/A	N/A	0,289628	0,303952	0,40642	bifunctional GMP synthase/glutamine amidotransferase protein
192	N/A	N/A	0,284224	0,310196	0,40558	dihydroxy-acid dehydratase
268	N/A	N/A	0,302512	0,291828	0,40566	cytosine deaminase
306	-0,03053	1,021384	0	0,159342	0,840658	inositol-5-monophosphate dehydrogenase
348	-0,47622	1,391095	0	0,025104	0,974896	pyrimidine regulatory protein PyrR; uracil phosphoribosyltransferase
362	N/A	N/A	0,289406	0,287426	0,423168	glycogen synthase
Up regulation						
17	N/A	N/A	0,422342	0,29656	0,281098	uroporphyrin-III synthase
30	0,127884	0,915173	0,808878	0,191122	0	citrate synthase
272	N/A	N/A	0,401182	0,29627	0,302548	gamma-glutamyltranspeptidase
325	0,004128	0,997143	0,884574	0,115426	0	porphobilinogen deaminase
339	0,071562	0,951607	1	0	0	GlpX protein
402	0,071562	0,951607	1	0	0	fructose-1,6-bisphosphatase; GlpX protein
151	N/A	N/A	0,426816	0,292182	0,281002	delta-aminolevulinic acid dehydratase
425	0,341645	0,789141	0,949182	0,050818	0	glutathione S-transferase; hypothetical protein
76	0,191752	0,875542	1	0	0	polyribonucleotide nucleotidyltransferase
427	0,542586	0,686539	1	0	0	DNA-directed RNA polymerase beta subunit; [...] ' subunit; [...] alpha subunit; [...] gamma chain; [...] omega subunit
158	N/A	N/A	0,416066	0,285382	0,298552	bacteriochlorophyll a synthase
198	0,174172	0,886276	0,99956	0,00044	0	protochlorophyllide oxidoreductase

N/A – Non-available

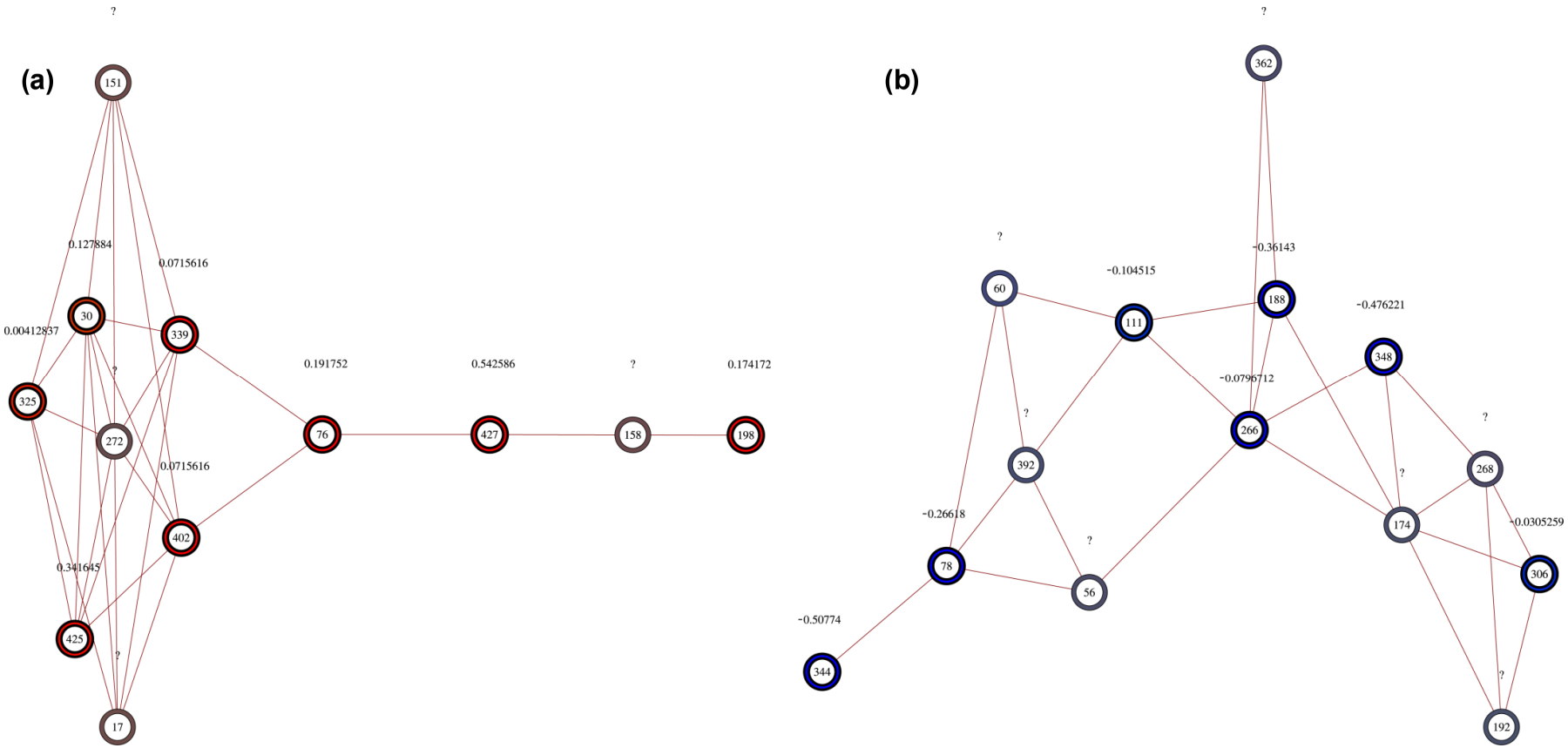


Fig. S1. Mixture model on graphs was performed as described in section 'Experimental Procedures'; the resulting probabilities were used for extracting graph components that would be (a) up- and (b) down-regulated using simple graph percolation. The parameters used for graph percolation were (1) probability of being differentially expressed > 0.80 or (2) > 0.40 if non-quantified for both the down- and up-regulated graphs. The list of corresponding enzymes is given in supplementary Table S2.

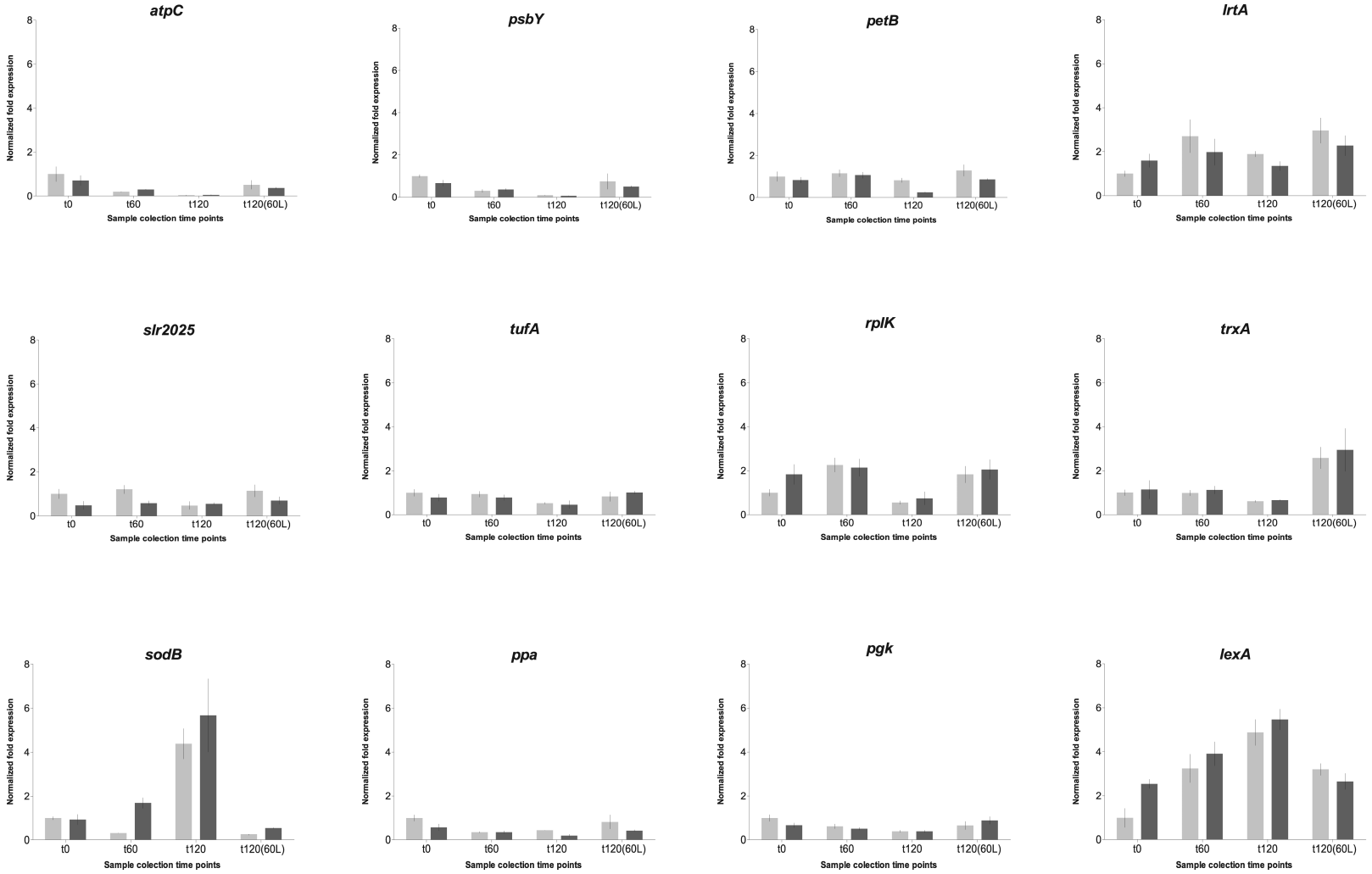


Fig. S2. Transcription profiles of the genes encoding proteins showing significant fold changes in *Synechocystis* wild-type (light gray) and Δ*hoxYH* mutant (dark gray) and evaluated by RT-qPCR. The mean normalized expression was calculated relative to the transcription of the 16S rRNA gene and the reaction internal normalization was performed using the sample t_0 from *Synechocystis* wild-type. Error bars represent one standard deviation.

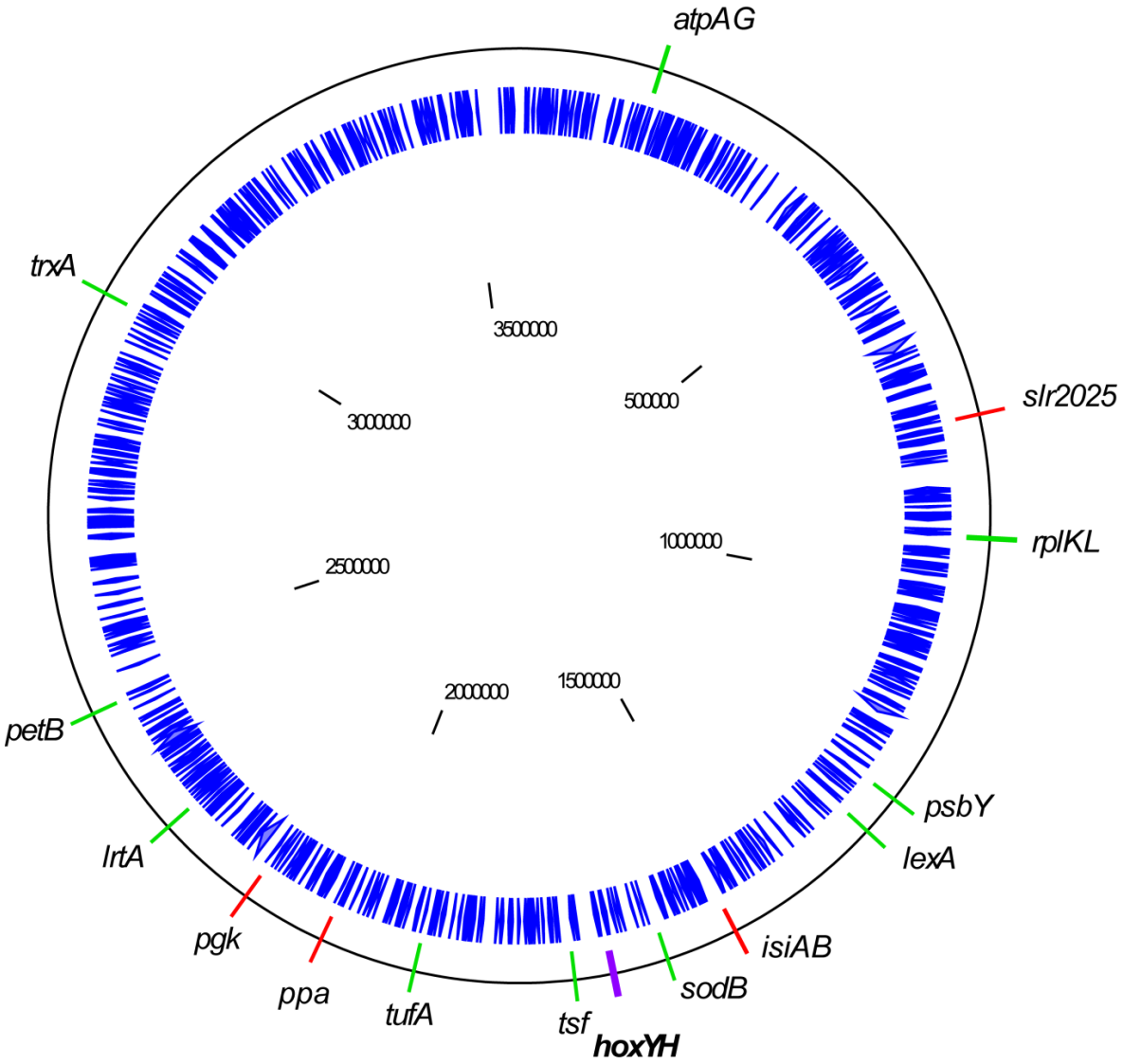


Fig. S3. Genome mapping of the genes encoding proteins showing significant fold changes between *Synechocystis* wild-type and Δ*hoxYH* mutant. The genes encoding the proteins more abundant in the mutant are mapped in green, and those encoding less abundant proteins are mapped in red. The position of the *hox* cluster is indicated by a purple bar.

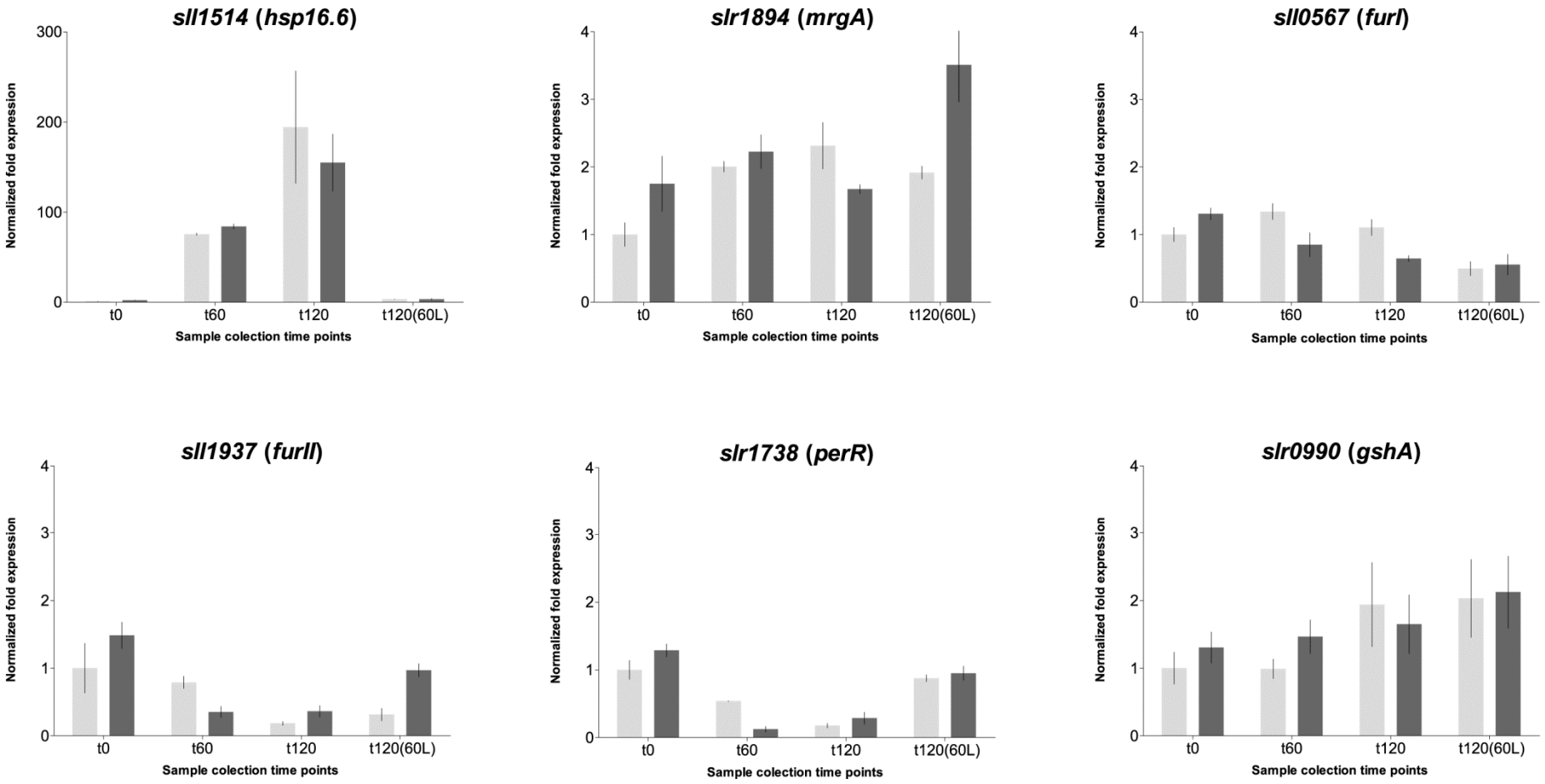


Fig. S4. Transcription profiles of the genes encoding proteins related to stress response and iron levels, evaluated by RT-qPCR (the primers used are listed in Table S1, and the annealing temperatures used are indicated). Light gray bars relate to *Synechocystis* wild-type and dark gray bars relate to Δ*hoxYH* mutant. The mean normalized expression was calculated relative to the transcription of the 16S rRNA gene and the reaction internal normalization was performed using the sample t_0 from *Synechocystis* wild-type. Error bars represent one standard deviation.

SUPPLEMENTARY REFERENCES

Elias, J. E. & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214.

Nakamura, Y., Kaneko, T. & Tabata, S. (2000). CyanoBase, the genome database for *Synechocystis* sp. strain PCC6803: status for the year 2000. *Nucleic Acids Res* **28**, 72-72.

Noirel, J., Sanguinetti, G. & Wright, P. C. (2008). Identifying differentially expressed subnetworks with MMG. *Bioinformatics* **24**, 2792-2793.

Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I. & Wright, P. C. (2009). iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **8**, 5347-5355.

Ow, S. Y. & Wright, P. C. (2009). Current trends in high throughput proteomics in cyanobacteria. *FEBS Lett* **583**, 1744-1752.

Pham, T. K., Roy, S., Noirel, J., Douglas, I., Wright, P. C. & Stafford, G. P. (2010). A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen *Tannerella forsythia*. *Proteomics* **10**, 3130-3141.

Reidegeld, K. A., Eisenacher, M., Kohl, M., Chamrad, D., Körting, G., Blüggel, M., Meyer, H. E. & Stephan, C. (2008). An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. *Proteomics* **8**, 1129-1137.

Sanguinetti, G., Noirel, J. & Wright, P. C. (2008). MMG: a probabilistic tool to identify submodules of metabolic pathways. *Bioinformatics* **24**, 1078-1084.

CHAPTER III

Characterization of neutral sites to use *Synechocystis* sp. PCC 6803 as a photoautotrophic chassis

Pinto F, Pacheco CC, Oliveira P, Montagud A, Urchuegía JF and Tamagnini P

Manuscript in preparation



Characterization of neutral sites to use *Synechocystis* sp. PCC 6803 as a photoautotrophic chassis

Filipe Pinto[1,2], Catarina C. Pacheco[1], Paulo Oliveira[1], Arnau Montagud[3], Javier F. Urchueguía[3] and Paula Tamagnini[1,2]

[1] IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal, [2] Faculdade de Ciências, Departamento de Biologia, Universidade do Porto, Porto, Portugal, [3] Instituto Universitario de Matemática Pura y Aplicada, Universidad Politécnica de Valencia, Camí de Vera, E-46071 Valencia, Spain.

Abstract

Cyanobacteria are promising "low-cost" cell factories since they have minimal nutritional requirements and high metabolic plasticity. The unicellular non-N_2-fixing cyanobacterium *Synechocystis* sp. PCC 6803 is the best studied strain and the vast amount of physiological and molecular data available already allowed the construction of computational genome-scale metabolic models. Within this work, five neutral sites were identified and characterized to stably integrate ectopic DNA into the chromosome, foreseeing the use of *Synechocystis* as a photoautotrophic chassis in biotechnological applications. The neutrality of the sites was evaluated by producing and analyzing deletion/insertion mutants, and their functionality was assessed introducing a synthetic device expressing the reporter green fluorescent protein. Moreover, integrative vectors including a multiple cloning site compatible with the BioBrick RFC[10] standard and insulated by transcription terminators were constructed, constituting robust cloning interfaces for synthetic biology purposes. *Synechocystis* mutants (chassis) ready to receive purpose-built synthetic modules/circuits are also available.

Keywords: chassis; neutral site; *Synechocystis*; synthetic biology

Introduction

Microorganisms have been widely used in biotechnological applications to produce added-value products for human use (e.g. Huang *et al*, 2008; Krauter *et al*, 2012; Westfall *et al*, 2012; Zeng, 2011; Zhou *et al*, 2012). The Gram-negative bacterium *Escherichia coli* is the most commonly used, but others such as the Gram-positive bacterium *Bacillus subtilis* (Biswas *et al*, 2012; Jia *et al*, 2012; Li *et al*, 2011b; Yamaoka *et al*, 2011), and the eukaryotic yeast *Saccharomyces cerevisiae* (Albertsen *et al*, 2011; Cantone *et al*, 2009; Hou, 2010; Kim *et al*, 2012; Martínez *et al*, 2012), have also been successfully exploited.

The use of heterotrophic organisms requires supplementation of the growth medium with external carbon sources increasing production costs. Therefore, the use of autotrophic organisms emerges as a valid alternative (Abed *et al*, 2009; Chen *et al*, 2011; Keskin & Hallenbeck, 2012; Li *et al*, 2011a; Malcata, 2011). In this context, cyanobacteria are promising "low-cost" cell factories since they can use CO_2 as a carbon source, water as reducing power, light as an energy source and in some cases atmospheric N_2 as nitrogen source. These organisms are found in almost any ecological niche on Earth (Whitton & Potts, 2000), which reflects their high degree of metabolic plasticity. Among cyanobacteria, the unicellular non-N_2-fixing *Synechocystis* sp. PCC 6803 (henceforth referred to as *Synechocystis*) is the most well studied strain, and the large amount of genomic, proteomic and physiological information generated over the years allowed the construction of genome-scale models, enabling the prediction of the system behavior (Fu, 2009; Montagud *et al*, 2010; Montagud *et al*, 2011). Moreover, it was the first photosynthetic organism to have its genome sequenced, comprising a 3.6 Mb chromosome (Kaneko *et al*, 1996) and 7 plasmids (from 2.3 kb to 120 kb) (Kaneko *et al*, 2003; Xu & McFadden, 1997; Yang & McFadden, 1993; Yang & McFadden, 1994), it is naturally transformable, and there are plenty of molecular tools available for its genetic manipulation (Heidorn *et al*, 2011; Huang *et al*, 2010; Koksharova & Wolk, 2002; Vermaas, 1996). Altogether, these characteristics allow the exploitation of *Synechocystis* as a photoautotrophic chassis for biotechnological applications. In contrast to *E. coli*, for which several replicative plasmids from different compatibility groups are available (Velappan *et al*, 2007), for *Synechocystis* these tools are limited (Koksharova & Wolk, 2002). Instead, suicidal vectors (*i.e.* non-replicating) are usually used to implement new functionalities integrating ectopic DNA into the genome, through homologous recombination (Kufryk *et al*, 2002; Williams, 1988; Zang *et al*, 2007a). A few integration sites in *Synechocystis* genome have previously been used, for example, to complement knockout mutants (Aoki *et al*, 2011; Burnap *et al*, 1994; Williams, 1988), to test promoters and study transcription regulation (Aoki *et al*, 1995; Kunert *et al*, 2000), or for the heterologous expression of proteins (Zang *et al*, 2007b). Although those loci have been described as "not appear to be deleterious" (Aoki *et al*, 1995), "does not otherwise affect any observable functions" (Burnap *et al*, 1994), "did not show any detectable phenotype" (Aoki *et al*, 2011), or being "a neutral site of the chromosome" (Kunert *et al*, 2000), they have not been extensively characterized as *neutral sites* – loci that can be disrupted without affecting cellular viability or cause any distinguishable phenotype (Clerico *et al*, 2007). Therefore, a comprehensive and systematic strategy to identify and characterize genomic neutral sites would greatly improve the ability to introduce new functionalities into *Synechocystis*.

In this work, a set of neutral sites in the chromosome of *Synechocystis* have been identified and characterized. To the best of our knowledge, this is the first time that a systematic approach has been used to map neutral sites in the genome of an organism, foreseeing its use as a photoautotrophic chassis for biotechnological applications. Moreover, the new integrative vectors were designed to be compatible with the BioBrick™ RFC[10] standard (http://www.biobricks.org/), which will make easier to use *Synechocystis* in synthetic biology approaches. Finally, the availability of several integration loci, spread throughout the cyanobacterium chromosome, will allow the implementation of complex synthetic circuits into this chassis.

Results

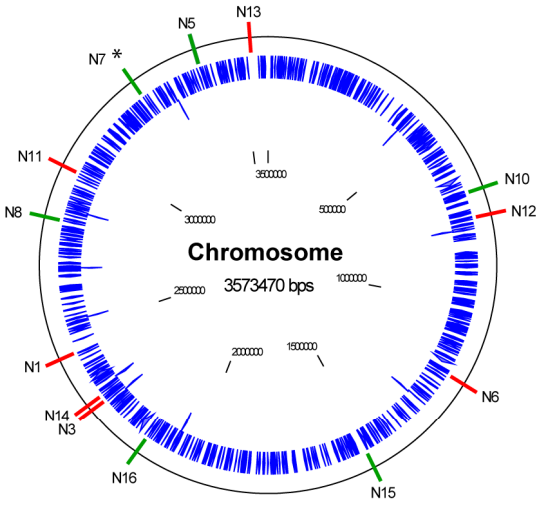
Selection of putative neutral sites

To identify putative neutral sites in the chromosome of *Synechocystis*, the list of annotated open reading frames (ORFs) with the respective encoded proteins was retrieved from CyanoBase (http://genome.kazusa.or.jp/cyanobase). From the 3264 ORFs listed, 1532 were annotated as being unknown or hypothetical and a subsequent bioinformatics analysis was performed to identify those matching the following criteria: (*i*) length of the putatively encoded proteins ≤301 amino acids, (*ii*) encoded proteins with no predicted transmembrane domains, (*iii*) no reported interaction with other proteins, and (*iv*) no relevant similarities found at the protein sequence level. This analysis led to the selection of 16 candidate neutral sites – termed N1 to N16 (see Figure 1 and Table I). Subsequently, a qualitative analysis of the genomic context of these ORFs was performed (Figure 1B), revealing that N2 site superimposes the 5' region of an ORF (*slr0369*) putatively encoding a cation or drug efflux system protein, N4 site superimposes the 5' region of an ORF (*slr0415*) putatively encoding a Na^+/H^+ antiporter, and N9 site corresponds to the ORF located immediately upstream and in the same direction as an ORF (*sll0182*) putatively encoding an ABC transporter. Therefore, these 3 sites were disregarded, and transcription analyses were performed for the other 13 loci.

Transcription analysis and final selection

To assess transcription, RT-PCRs were initially performed using RNA extracted from cells collected at the exponential growth phase (OD_{730}≈0.8). The results revealed high amplification levels for N3, N11, N12 and N13 sites, low amplification levels for N6, N8, N14 and N15 sites, low primers efficiency for N1 site, and no detectable amplification for the remaining sites (N5, N7, N10 and N16) (see Supplementary Figure 1).

A



B

Site	Genomic context
N1	'slr0348 (lytB) slr0350 (tp) **ssl0606** slr0351 sll0318 slr0352 (tp) sll0317 (tp)
N2	'slr0366 **slr0368** slr0369 (envD)
N3	'slr0338 slr0168 **ssl0318** slr0169 slr0171 (ycf37) slr0172 (guaB) 'sll0177
N4	'sll0397 sll0396 (ompR) sll0395 (gpm) sll0394 **ssr0680** slr0415 (nhaS5) slr0416
N5	'sll1477 slr1583 **sll1476** sll1475 (hik32) 'sll1474 (tp)
N6	'sll1775 ssr3154 sll1774 **slr1869** ssr3159 slr1870 ssl3389
N7	sll0495 (asnS) trnY-GUA trnT-GGU **sll0494** sll0493 sll0902 (argF)
N8	'ssl1047 slr0572 sll0543 ssl1046 **slr0573** sgl0001 ssl1045 slr0574 (cyp)
N9	sll0182 (abc) **sll0181** ssl0350 sll0180
N10	sll1286 slr1395 **slr1396** slr1397 slr1398 'sll1285
N11	'slr0615 (abc) slr0616 (mvrA) **ssr1038** slr0617 'sll0590
N12	'slr2027 sll1913 sll1912 **ssl3615** sll1911 'sll1910 (zam)
N13	slr0585 (argG) slr0586 **slr0587** sll0578 (purK) sll0577 'slr0588
N14	'sll0169 (ftn2) sll0168 **sll0167** sll0166 (cobA/hemD) 'sll0163
N15	'slr0269 slr0270 ssl0453 (nblA2) ssl0452 (nblA1) **slr0271** slr0272 slr0273 'sll0243
N16	'sll0361 sll0360 sll0359 (cyAbrB) **slr0397** slr0398 slr0399 (ycf39) 'slr0400 (ppnK)

Figure 1 | Chromosomal location and genomic context of the identified putative neutral sites. (A) The sites chosen for mutants construction are in green and the ones disregarded are in red (* - site disregarded due to cloning difficulties). (B) Representation of the genomic context of each putative neutral site (5000 bp). The ORFs corresponding to the selected sites are in orange and ORFs encoding proteins with assigned functions and immediately downstream of the sites are in red. tp – transposase (putative); ‘ – partial ORFs.

Table I | List of the putative neutral sites identified

Site name	ORF ID	Chromosome position	Orientation*	Length of the putatively encoded protein (amino acids)
N1	*ssl0606*	2441925 - 2442083	c	52
N2	*slr0368*	2365848 - 2366120	d	90
N3	*ssl0318*	2302457 - 2302645	c	62
N4	*ssr0680*	2684766 - 2684990	d	74
N5	*sll1476*	3398409 - 3398717	c	102
N6	*slr1869*	1212609 - 1213358	d	249
N7	*sll0494*	3224429 - 3225334	c	301
N8	*slr0573*	2816517 - 2816960	d	147
N9	*sll0181*	2737929 - 2738552	c	207
N10	*slr1396*	707437 - 708054	d	205
N11	*ssr1038*	2950107 - 2950343	d	78
N12	*ssl3615*	776058 - 776342	c	94
N13	*slr0587*	3536090 - 3536422	d	110
N14	*sll0167*	2317536 - 2318030	c	164
N15	*slr0271*	1524568 - 1525086	d	172
N16	*slr0397*	2146801 - 2147376	d	191

* d – direct sequence, c – complement sequence.

Further transcriptional studies were performed using cells collected at different growth phases (early exponential, exponential and stationary), to evaluate a possible differential transcription of sites N5, N6, N7, N8, N10, N14, N15 and N16 (see Figure 2). The results showed that there was residual transcription for N6 and N14 sites throughout the various growth phases and, therefore, these two sites were disregarded. Concerning the other sites, it was possible to detect residual transcription for N8 and N15 sites, but only on samples collected at the early exponential phase. Accordingly, six sites (N5, N7, N8, N10, N15 and N16) were selected to generate mutants in order to validate their “neutrality”. Due to recurring cloning difficulties found constructing N7-related plasmids, this site was also disregarded. The final target sites (N5, N8, N10, N15 and N16) are evenly distributed throughout the chromosome, with intervals in the order of 60-90º (72º corresponds to 1/5 of the chromosome, see Figure 1A).

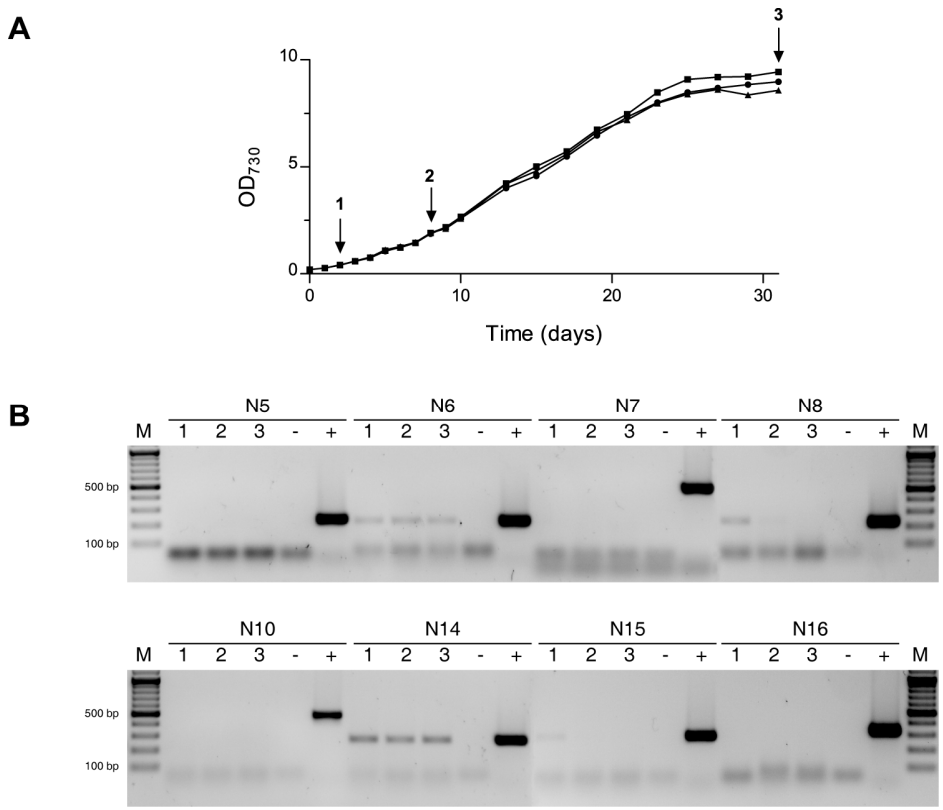


Figure 2 | Transcription analysis of eight selected loci. (A) Growth curves of three independent cultures of *Synechocystis* and sample collection points for RNA extraction (indicated by arrows). (B) Gene transcription analysis by RT-PCR, using total RNA extracted from cells collected at three different growth phases. These results are representative of the three biological replicates and technical duplicates. Samples from different collection points (1, 2 and 3), negative control (-, absence of template) and positive control (+, genomic DNA) are indicated.

Generation of mutants

The integrative plasmids (Table II and Supplementary Table 1) used to transform *Synechocystis* and characterize the neutral sites were constructed following the design depicted in Figure 3 (for details see 'Materials and Methods'). *Synechocystis* mutants were obtained by natural transformation with each plasmid (Table II and Supplementary Table 2) and full segregation was confirmed by PCR, using primers within (N*n*F and N*n*R, Supplementary Table 3) and external to each site (N*n*.5O and N*n*.3O, Supplementary Table 4), and by Southern blots. All the mutants produced in this study were fully segregated, with all chromosomal copies mutated (Figure 4 and Supplementary Figures 2 and 3). The mutants produced using the pSN*n*KS plasmid series (containing the double selection cassette) exhibited the expected phenotypes of kanamycin resistance (Km^R) and sucrose sensitivity (Suc^S) (Supplementary Figure 4), and mutants deriving from these, without the double selection cassette, are being produced and characterized (Table II and Supplementary Table 2).

Table II | Organisms and plasmids

Organisms

Name*	Description	Source
Escherichia coli DH5α	Transformation/cloning strain	Invitrogen
Synechocystis sp. PCC 6803	Wild-type strain	Pasteur Culture Collection
SN*n*K	*Synechocystis* mutants with replacement of the neutral site N*n* by a kanamycin resistance cassette.	This work
SN*n*KS	*Synechocystis* mutants with replacement of the neutral site N*n* by a kanamycin resistance/sucrose sensitivity cassette.	This work
SN*n*K.gfp	*Synechocystis* mutants with replacement of the neutral site N*n* by a kanamycin resistance cassette and the promoterless GFP encoding gene.	This work
SN*n*K.Cgfp	*Synechocystis* mutants with replacement of the neutral site N*n* by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	This work
SN*n*.gfp	*Synechocystis* mutants derived from SN*n*KS mutants with replacement of the double selection cassette on neutral site N*n* by the promoterless GFP encoding gene.	This work
SN*n*.Cgfp	*Synechocystis* mutants derived from SN*n*KS mutants with replacement of the double selection cassette on neutral site N*n* by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	This work

Plasmids

Name*	Description	Source
pGEM®-T easy	T/A cloning vector	Promega
pK18mobsacB	Small mobilizable multi-purpose cloning vector derived from the *E. coli* plasmids pK18, conferring resistance to kanamycin (*nptII*), and sensitivity to sucrose (*sacB*).	National BioResource Project (NIG, Japan): E.coli
pSB1A3-E0240	BioBrick™ vector containing the part BBa_E0240 (GFP generator).	Registry of Standard Biological Parts, MIT
pJ201-P_{psbA2*}	Synthesized minimal *psbA2* promoter region, meeting the BioBrick RFC[10]standard specifications, cloned into the delivery vector pJ201	DNA 2.0 Inc., this work
pBSK-FP300	Synthetic interface containing a BioBrick compatible multiple cloning site(*Mun*I, *Not*I, *Xba*I, *Spe*I, *Not*I, *Pst*I) insulated with the transcriptional terminators BBa_B1006 and BBa_B0062 (upstream), and BBa_B0053 and BBa_B020 (downstream).	Epoch Life Science Inc., this work
pSN*n*	pGEM-T easy- based plasmid series containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site *n*.	This work
pSN*n*K	pSN*n* plasmid series with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.	This work
pSN*n*KS	pSN*n* plasmid series with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.	This work
pSN*n*.gfp	pSN*n* plasmid series with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.	This work
pSN*n*.Cgfp	pSN*n* plasmid series with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.	This work
pSN*n*K.gfp	pSN*n*K plasmid series with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.	This work
pSN*n*K.Cgfp	pSN*n*K plasmid series with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.	This work

**n*=5, 8, 10, 15 or 16

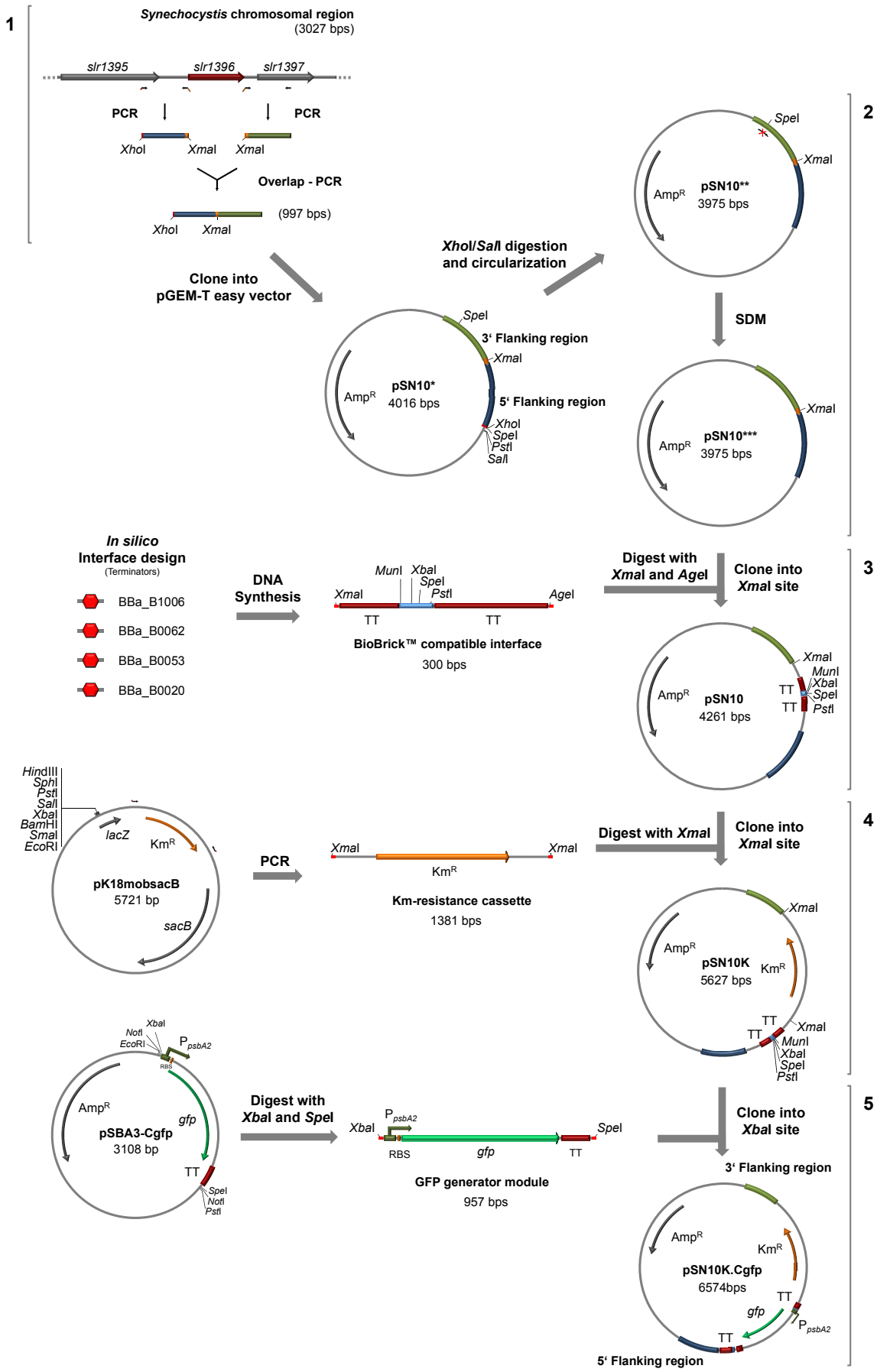

1
Synechocystis chromosomal region
(3027 bps)
slr1395
slr1396
slr1397
PCR
PCR
XhoI
XmaI
XmaI
Overlap - PCR
(997 bps)
XhoI
XmaI
Clone into
pGEM-T easy vector
XhoI/SalI digestion
and circularization
2
SpeI
XmaI
AmpR
pSN10**
3975 bps
SDM
SpeI
3' Flanking region
XmaI
AmpR
pSN10*
4016 bps
5' Flanking region
XhoI
SpeI
PstI
SalI
XmaI
AmpR
pSN10***
3975 bps
In silico
Interface design
(Terminators)
BBa_B1006
BBa_B0062
BBa_B0053
BBa_B0020
DNA
Synthesis
MunI
XbaI
SpeI
XmaI
PstI
AgeI
TT
TT
BioBrick™ compatible interface
300 bps
3
Digest with
XmaI and AgeI
Clone into
XmaI site
XmaI
MunI
XbaI
SpeI
PstI
AmpR
pSN10
4261 bps
TT
TT
HindIII
SphI
PstI
SalI
XbaI
BamHI
SmaI
EcoRI
lacZ
KmR
pK18mobsacB
5721 bp
sacB
PCR
XmaI
XmaI
KmR
Km-resistance cassette
1381 bps
4
Digest with XmaI
Clone into
XmaI site
XmaI
AmpR
pSN10K
KmR
5627 bps
TT
TT
XmaI
MunI
XbaI
SpeI
PstI
NotI
EcoRI
XbaI
PpsbA2
RBS
AmpR
gfp
pSBA3-Cgfp
3108 bp
TT
SpeI
NotI
PstI
Digest with
XbaI and SpeI
XbaI
PpsbA2
SpeI
RBS
gfp
TT
GFP generator module
957 bps
5
Clone into
XbaI site
3' Flanking region
AmpR
KmR
pSN10K.Cgfp
6574bps
TT
gfp
PpsbA2
TT
5' Flanking region

Figure 3 | Schematic representation of pSN10K.Cgfp construction. Homologous regions flanking the putative neutral site were amplified by PCR (primers indicated by arrows) and fused by overlap-PCR (1). The fragment containing the two flanking regions was cloned into the pGEM-T easy vector; the restriction sites incompatible with the BioBrick standard RFC[10] were removed by digestion with *Xho*I and *Sal*I and vector re-circularization, and by site directed mutagenesis (SDM, mutagenic primers indicated by a crossed arrow) (2). The BioBrick-compatible cloning interface, flanked by two double BioBrick transcription terminators (TT), was synthesized and cloned into the *Xma*I site of the vector originating the pSN10 plasmid (3). A selection cassette conferring resistance to kanamycin was amplified by PCR (primers indicated by arrows) from the plasmid pK18mobsacB and cloned into the *Xma*I site (pSN10K) (4). Subsequently, the module containing the GFP encoding gene under the control of the minimal *psbA2* promoter (P_{psbA2*}) was excised from the BioBrick vector pSBA3-Cgfp and cloned into the cloning interface, originating the integrative plasmid pSN10K.Cgfp (5).

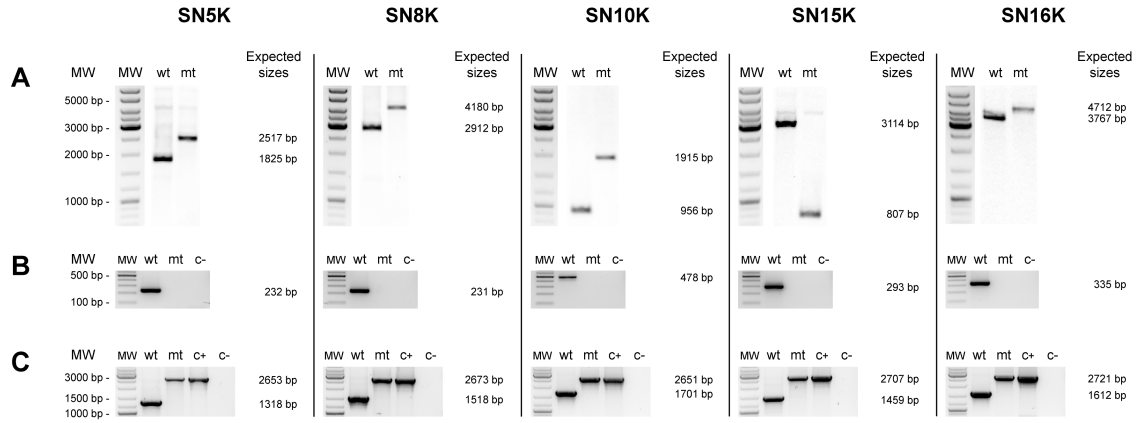


Figure 4 | Confirmation of the segregation of the *Synechocystis* disruption mutants. Full mutant segregation was confirmed by Southern blot (A) and PCRs (B and C): using primers within the ORF corresponding to each neutral site (B) and with the primers external to the neutral sites (C). The results depicted correspond to each mutant (mt), using the wild-type (wt) as reference; the plasmids used to produce the mutants were used in PCRs as positive controls (c+) and no template controls were also included (c-).

Statistical analysis of cyanobacteria specific growth

To assess the neutrality of the sites, *Synechocystis* mutants were produced exchanging each putative neutral site by a kanamycin resistance cassette, using the pSN*n*K plasmids series (Table II and Supplementary Table 2). The successful generation of fully segregated mutants in the five sites showed that the mutations did not affect cell viability under standard growth conditions. The growth of the wild-type and the five mutants was evaluated under conditions in which *Synechocystis* is usually cultivated: continuous light (20 µmol photons/m^2/s; CL) or 12h light (20 µmol photons/m^2/s)/12h dark cycles (LD), and in autotrophy (no glucose; G0) or mixotrophy (5 mM glucose; G5). The experiments were performed in 4.5 mL cuvettes containing 2 mL of the cultures (at 28 ºC and 30 ºC), and in 25 mL flasks containing 10 mL of the cultures (at 30 ºC). The optical density of the cultures (OD_{730}) was monitored daily for 10 days and the resulting growth curves for each

organism and cultivation conditions are depicted in Figure 5. The statistical analysis of variance (ANOVA) of the growth data was performed to compare the wild-type and the mutant organisms, using the specific growth rate parameter – μ (see 'Materials and Methods'). The use of this parameter is reliable since the test duration was the same in all conditions and stationary growth was reached at similar moments. The ANOVA was four-way in the cuvettes set and three-way in the flasks set, and values of up to second order interactions can be found in Supplementary Table 5. Regarding the effect of each mutation in comparison to the wild-type, highly significant *P*-values were obtained for the mutants SN5K (in both experimental sets) and SN8K (in flasks), while for SN10K the

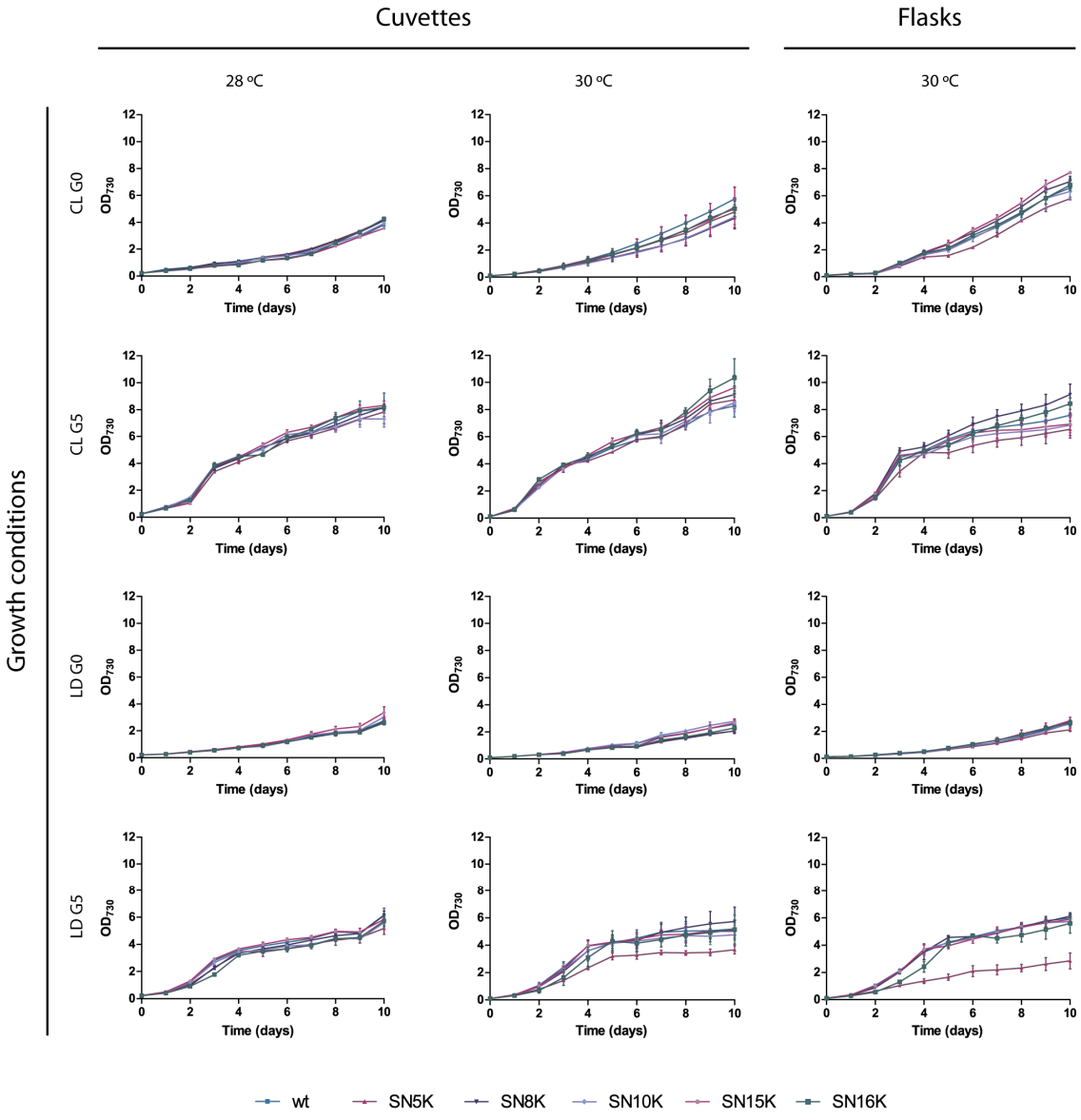


Figure 5 | Growth curves of *Synechocystis* wild-type and mutants cultivated under different conditions. OD_{730} values were recorded daily for 10 consecutive days in continuous light (CL) or 12 h light/12 h dark cycles (LD), in autotrophy (G0) or mixotrophy (5 mM glucose, G5), and in 4.5 mL cuvettes at 28 ºC and 30 ºC, or 25 mL flasks at 30 ºC.

P-values are approximately 0.05, and for SN15K and SN16K the *P*-values are not statistically significant (Supplementary Table 5). Looking into the growth curves (Figure 5) it can be seen that the mutant SN5K overall grows less than the wild-type (especially in the presence of glucose), while the mutant SN8K grows slightly better. Taking these results into consideration, the neutrality of the chromosomal loci may be qualitatively ordered as N15≈N16>N10>N8>>N5, from the most to the less similar to the wild-type. For for the rest of the variables considered in the ANOVA, irradiance and glucose presence were the major parameters affecting growth (highly significant *P*-values).

Synthetic interface robustness and functionality of the neutral sites

To use the constructed vectors as platforms for the integration of synthetic devices, a BioBrick-compatible cloning interface was designed and synthesized. Therefore, the *Mun*I, *Not*I, *Xba*I, *Spe*I, *Not*I and *Pst*I cloning sites, compatible with the RFC[10] assembly standard, were flanked by two double transcription terminators (BioBricks BBa_B1006 and BBa_B0062 upstream, and BBa_B0053 and BBa_B0020 downstream), to prevent any potential promoter activity from backbone DNA sequences. The interface was synthesized and cloned between the homologous recombination sequences of each vector. To assess the proper transcriptional insulation of the interface, a promoterless green fluorescent protein generator (GFP; BioBrick composite part BBa_E0240) was cloned in each pSN*n* and pSN*n*K plasmids series, and *Synechocystis* mutants were produced (mutants SN*n*(K).gfp, see Table II and Supplementary Table 2). Additionally, to assess the functionality of the neutral sites, *Synechocystis* mutants carrying the GFP generator under the control of a constitutive synthetic promoter (minimal *psbA2* promoter, P_{psbA2^*}) were constructed (mutants SN*n*(K).Cgfp, see Table II and Supplementary Table 2). GFP expression was analyzed by Western blot (Figure 6) and confocal microscopy (Figure 7).

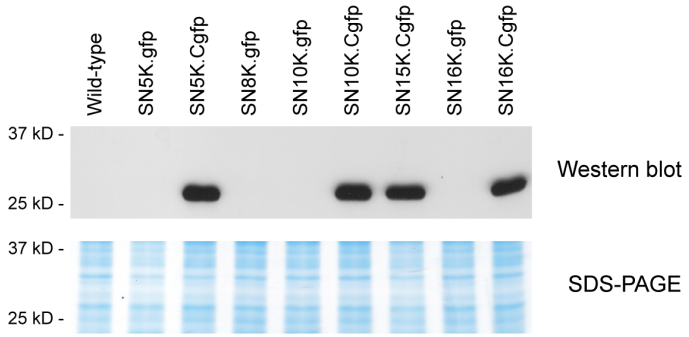


Figure 6 | Evaluation of the synthetic BioBrick-compatible interface leakiness and functionality. Protein extracts were obtained from *Synechocystis* wild-type and mutants bearing the promoterless GFP generator (SN*n*K.gfp) or under the control of the constitutive minimal P_{psba2^*} synthetic promoter (SN*n*K.Cgfp), and separated on SDS-polyacrylamide gels. Proteins were transferred to a nitrocellulose membrane and GFP (MW≈27 kDa) was detected by a specific antibody (Western blot). A Comassie blue stained gel (SDS-PAGE) is shown as control of the protein amount loaded.

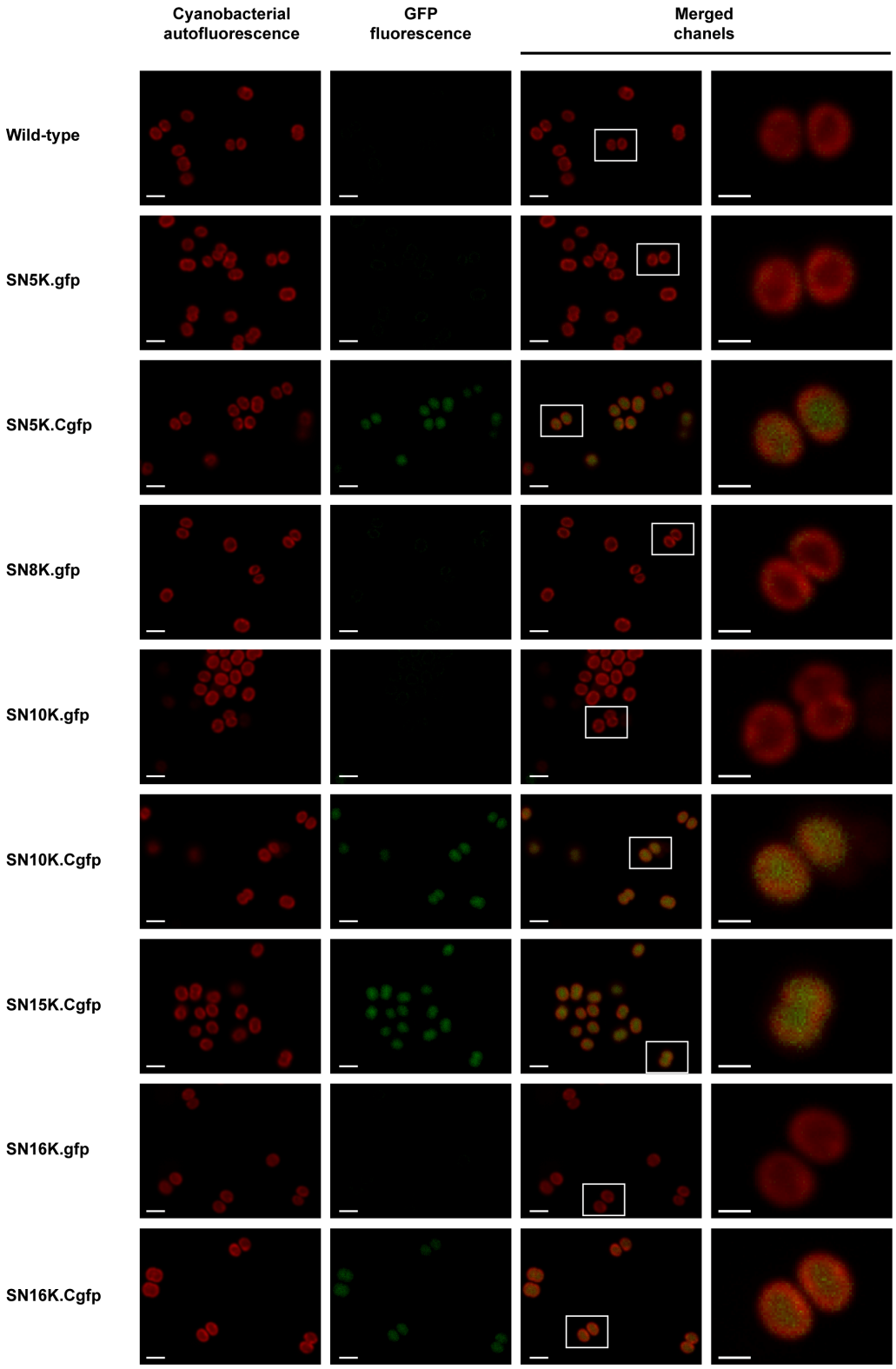


Figure 7 | Confocal micrographs of *Synechocystis*. *Synechocystis* wild-type and insertion mutants containing a promoterless GFP generator (SN*n*K.gfp) or the GFP generator under the control of the constitutive minimal P_{psba2^*} synthetic promoter (SN*n*K.Cgfp) were analyzed. The cyanobacterial autofluorescence is depicted in the first column from left, while the GFP signal is shown in the second column. The merging of the signals from both channels (cyanobacterial autofluorescence and GFP) is shown in the third column. The first three columns are at the same scale (bar= 2.5 μm) and the fourth column represents the magnification of the area within the white rectangles (bar= 1 μm).

The results show that GFP antigens (Western blot) and fluorescence (confocal microscopy) could only be detected in mutants where *gfp* was under the control of the constitutive promoter. These results confirm the proper insulation of the BioBrick-compatible cloning interface and show that the integration platforms are indeed functional.

Discussion

Synechocystis is a well studied cyanobacterium with plenty of molecular tools developed for its genetic manipulation. Replicative plasmids have been used to introduce foreign DNA into this organism, which is achieved by conjugation or electroporation (Matsunaga and Takeyama, 1995; Vioque, 2007). These plasmids are usually based on the broad-host-range *E. coli* IncQ plasmid RSF1010 (Scholz *et al.*, 1989; Sode *et al.*, 1992; Marraccini *et al.*, 1993; Mermet-Bouvier *et al.*, 1993; Mermet-Bouvier and Chauvat, 1994; Huang *et al.*, 2010), and usually are not specific for cyanobacteria. Although presenting some advantages, such as being relatively easy to construct and quickly introduced into the cell, broad-host-range replicative plasmids can exhibit some instability (David *et al.*, 1983; Priefer *et al.*, 1985; Becker and Meyer, 1997; De Gelder *et al.*, 2007; Meyer, 2009). Moreover, (*i*) the use of large native DNA fragments may lead to the recombination of the replicative vectors with the chromosome, (*ii*) the absence of selective pressure may lead to the loss of the vector and (*iii*) DNA structure of the vectors may have a strong influence on transcription, which altogether may compromise system predictions and control. In addition, a recently published work shows that *Synechocystis*' endogenous plasmids copy number per chromosome vary with nutritional conditions and growth phase (Berla and Pakrasi, 2012). Taking into consideration that the number of chromosomes also vary throughout growth (Griese *et al.*, 2011), plasmids copy number calculations are even more complex. Instead, chromosomal recombination may be a more viable alternative for the implementation of synthetic devices into the cyanobacterial chassis and the availability of fully characterized chromosomal loci is crucial for this purpose.

To date, few chromosomal loci have been used to introduce foreign DNA into *Synechocystis*, namely to test promoters and transcription regulation, to introduce new functionalities or to complement knockout mutants (e.g. Williams, 1988; Burnap *et al.*, 1994; Aoki *et al.*, 1995; Aoki *et al.*, 2011). Although some of those loci have been described as "not appear to be deleterious", "does not otherwise affect any observable functions", "did not show any detectable phenotype", or being "a neutral site of the chromosome", they have not been extensively characterized, and some of their features will be subsequently discussed on the light of recent findings. The locus described by Williams (1988) disrupts the gene *slr0168* that encodes a protein secreted by

Synechocystis (Sergeyenko and Los, 2000). Although its function is unknown, Slr0168 possesses a DUF4114 protein domain. This is a repeated domain found towards the C-terminal of many different types of bacterial proteins, and the highly conserved glutamate and aspartate residues suggest that this domain may carry enzymatic activity (Marchler-Bauer *et al.*, 2009). Burnap *et al.* (1994) describes a locus in a "nonessential, nontranscribed" region. In contrast, the recent work of Mitschke *et al.* (2011) shows that the genes flanking that insertion locus are transcribed. Moreover, this locus is located in the intergenic region upstream of the hypothetical gene *ssr0761* that putatively encodes a "prevent-host-death" protein (antitoxin of toxin-antitoxin stability system) (Marchler-Bauer *et al.*, 2009). The locus described by Aoki *et al.* (1995) disrupts the *ssl0410* gene, which is located downstream of a gene encoding a subunit of the type I NAD(P)H dehydrogenase complex (*ndhB*) that is important for CO_2 uptake in *Synechocystis* (Ohkawa *et al.*, 2000). Moreover, *ssl0410* and *ndhB* are differentially transcribed under inorganic carbon limiting conditions (Wang *et al.*, 2004; Mitschke *et al.*, 2011). The locus described by Aoki *et al.* (2011) is located in the intergenic region upstream of *slr2031* that encodes the regulatory protein RsbU. This gene is constitutively transcribed and mutants in *slr2031* showed a defect in regenerating growth after nitrogen- and sulfur-starvation-induced stationary phase (Huckauf *et al.*, 2000). The estimated length of the main transcripts for *slr2031* is about 1.7 kb (Huckauf *et al.*, 2000), which is in agreement with the transcription start site mapped inside the immediately upstream gene *slr2030* (Mitschke *et al.*, 2011). Consequently, insertions in that intergenic region may disrupt the transcriptional unit of *slr2031* and may lead to the same phenotype described for disruption mutants on this gene. Since the integration sites described previously do not match at least one of the selection criteria established in this study, they were not tested/assessed here.

Within this work five putative neutral sites for the stable integration of synthetic modules/circuits into *Synechocystis* chromosome were identified following an unbiased systematic approach, and characterized. Although non-coding DNA (ncDNA) regions could be used for this purpose they may possess *cis*-regulatory control sequences, thus affecting the transcription profiles of several genes (Castillo-Davis, 2005; Janga & Collado-Vides, 2007; Morris, 2012; Zhou *et al*, 2010). Therefore, only ORFs putatively encoding unknown or hypothetical proteins were taken into consideration for this work. The initial selection of the putative neutral sites was performed prior to the publication of the recent work of Mitschke *et al.* (2011). In that work, the transcription of sites N5 and N10 was detected by the "Whole Transcriptome Shotgun Sequencing" technique, and potential antisense transcripts (asRNA), originated from the reverse complementary strand of these two sites, were described. Regarding the other sites, the number of

readings was ≤2. Nonetheless, in our work transcription of the identified putative neutral sites was assessed throughout various growth phases, and appeared to be negligible. The successful generation of fully segregated mutants, with each selected neutral site replaced by a kanamycin resistance cassette, showed that they do not seem to encode proteins essential to the viability of the cyanobacterium, at least under standard growth conditions. Several studies reported that genes crucial to the viability of *Synechocystis* cannot be successfully and completely inactivated, resulting in heteroploid mutant cells, containing both chromosomes harboring the wild-type gene and the knockout locus (García-Domínguez *et al.*, 2000; Domain *et al.*, 2004; Ishii and Hihara, 2008; Oliveira and Lindblad, 2008). Moreover, the neutrality of the each site was assessed by analyzing the growth of site-replacement mutants cultivated under different conditions. The results showed that the mutants on the sites N15 and N16 exhibit growth patterns similar to those of the wild-type, while the mutants on the sites N5, N8 and N10 present some statistically significant differences. Nonetheless, the mutant on the site N8 exhibited growth rates slightly higher than the parental strain, a feature that may be advantageous in biotechnological applications. As for the rest of the variables considered in the ANOVA, and similarly to the results obtained by Lopo *et al.* (2012), irradiance and the presence of glucose were the major parameters affecting growth (highly significant *P*-values), while the effect of the temperature was not so obvious in our analysis. Although the experiments in flasks were not carried out at 28 ºC, one can assume that the growth patterns of each organism would be similar to that of the cuvettes, according to the 'scale-up' experiments performed in the work of Lopo *et al.* (2012) that showed no significant differences between growth in cuvettes and flasks. Additionally, the statistical influence of temperature in growth is not correlated to the mutations, since two-way interaction of these two parameters ('Temperature + Mutation', Supplementary Table 5) does not have significant *P*-values. This means that they are not co-variables and affect growth independently. Thus, it is expected that the SN15K and SN16K mutants have the closest growth behavior to the wild-type when grown in flasks at 28 ºC.

All the integrative plasmids constructed in this work possess a BioBrick-compatible synthetic interface for cloning purposes. The robustness of this synthetic interface was assessed using the promoterless GFP generator (BioBrick BBa_E0240), which showed no leakiness from the chromosomal backbone for each site (no GFP expression). Moreover, all the mutants generated bearing the GFP generator under the control of a constitutive synthetic promoter based on the photosensitive promoter of the *psbA2* gene of *Synechocystis*, were able to express the GFP protein. *gfp* transcription studies must also be performed to confirm the results obtained at the protein level and to evaluate if

transcription is affected by the position of each neutral sites in the chromosome.

Altogether, these results show that the selected sites are functional and liable to receive synthetic modules/devices. In addition, the vectors constructed in this work followed a rational design that allows the stable integration of synthetic devices/modules into *Synechocystis* chromosome. Moreover, the possibility to generate markerless secondary mutants enables to use the vectors to introduce more than one synthetic device/module in a single organism without any resistance marker. Therefore, the existence of more than one integration site, spread throughout the chromosome of *Synechocystis*, makes possible to implement more complex synthetic circuits. The vectors design also considered the possibility to use different resistance cassettes making the implementation of these complex circuits faster. Finally, this work contributes to the field of synthetic biology by making possible to use *Synechocystis* as a photoautotrophic chassis, providing not only standard BioBrick-compatible vectors to integrate synthetic modules/devices into fully characterized chromosomal neutral sites, but also with chassis prone to receive a synthetic module/device without selective markers.

Materials and Methods

Organisms and maintenance culture conditions

Wild-type and mutants of the cyanobacterium *Synechocystis* sp. PCC 6803 were maintained in BG11 medium (Stanier *et al.*, 1971) at 25 ºC, under a 16 h light (20 µmol photons/m^2/s)/8 h dark regimem. For solid medium, BG11 was supplemented with 1.5% (w/v) agar noble (Difco), 0.3% (w/v) sodium thiosulfate and 10 mM TES-KOH buffer, pH 8.2. For the selection of mutants, BG11 medium was supplemented with kanamycin (Km, 10-300 µg/mL). For cloning purposes *Escherichia coli* strain DH5α (Stratagene) was used. *E. coli* cells were grown at 37 ºC on LB medium (Sambrook and Russell, 2001), supplemented with ampicillin (100 µg/mL).

Bioinformatics

The list of the predicted/annotated ORFs of *Synechocystis* was retrieved from CyanoBase (Distribution of Sequence and Annotated Data Files, ftp://ftp.kazusa.or.jp/pub/CyanoBase/Synechocystis) (Nakamura *et al.*, 1998; Nakamura *et al.*, 2000). From the 3264 ORFs listed, only those putatively encoding unknown or hypothetical proteins were selected. From these, the final selection was based on the following criteria: (*i*) length of the putatively encoded proteins ≤301 amino acids, (*ii*)

encoded proteins with no transmembrane domains predicted by the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), (*iii*) no interaction with other proteins in the two-hybrid system (CyanoBase data), and (*iv*) no relevant similarities found at the protein sequence level, when comparing with other sequences using BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The genomic context of the ORFs was also taken into consideration and sites were disregarded when in the vicinity of genes with assigned putative functions. To assess the transcription of the selected ORFs, primers were designed automatically from the DNA sequence (Kaneko *et al.*, 1996) using the Beacon Designer 6 software (PREMIER Biosoft International). For other purposes, primers were designed manually and analyzed using the Integrated DNA Technologies web resource OligoAnalyzer 3.1 (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/).

DNA and RNA extraction

Cyanobacterial genomic DNA (gDNA) extraction was carried out according to the procedure described previously (Tamagnini *et al.*, 1997). For RNA extraction *Synechocystis* cells were collected by centrifugation (10 min at 3850 *g*), and frozen at -80 ºC. RNA was extracted using the TRIzol® Reagent (Invitrogen) according to the method described previously (Leitão *et al.*, 2006).

RT-PCR

To assess target ORFs transcription, RNA was extracted from *Synechocystis* cells grown to an OD_{730} of 0.8-0.9, under continuous light at 30 ºC. Further transcriptional studies were performed with RNA extracted from samples collected at three different growth phases (OD_{730} of approximately 0.4, 2 and 9, see Figure 2A), from three independent cultures grown under the same conditions. RT-PCRs were carried out as described in Ferreira *et al.* (2009). Briefly, cDNA was synthesized in duplicate from 1 µg of total RNA using the iScript™ Select cDNA Synthesis Kit (Bio-Rad) and the random primers supplied, following the manufacturer's instructions. The two RT reactions from each sample were pooled together and each PCR reaction contained 1.5 µL of cDNA, 0.1 µM of each primer (see Supplementary Table 3), 200 µM dNTPs mix, 1X PCR reaction buffer, and 0.025 U of illustra™ *Taq* DNA polymerase (GE Healthcare). The PCR profile was: 2 min at 94 ºC followed by 40 cycles of 30 s at 94 ºC, 30 s at 50 ºC, and 30 s at 72 ºC, concluding with a 7 min extension step at 72 ºC. RNA contamination by gDNA prior to cDNA synthesis was assessed by PCR using primers specific for the *Synechocystis* 16S rDNA, and the RT reaction efficiency was assessed using the same primers pair. All the RT-PCRs included a negative control (omission of template) and a positive control (*Synechocystis* gDNA).

Homologous regions amplification and fusion by PCR

PCRs to amplify the regions flanking the target loci (see below) contained 7 ng of gDNA, 0.2 µM of each primer, 250 µM dNTPs mix, 1X *Pfu* reaction buffer with $MgSO_4$, and 0.025 U of *Pfu* DNA polymerase (Fermentas). The PCR profile was: 1 min at 94 ºC followed by 35 cycles of 30 s at 94 ºC, 45 s at 48 ºC, and 90 s at 72 ºC, concluding with a 7 min extension step at 72 ºC. Agarose gel electrophoresis was performed by standard protocols using 1X TAE buffer (Sambrook and Russell, 2001), DNA fragments were purified using the GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare), according to the manufacturer's instructions. "Overlap-PCRs" (see below) contained 80 ng of each purified DNA fragment, 0.125 µM of each primer, 250 µM dNTPs mix, 1X PCR reaction buffer, and 0.025 U of illustra™ *Taq* DNA polymerase (GE Healthcare). The PCR profile was: 5 min at 95 ºC followed by 10 cycles of 30 s at 95 ºC, 45 s at 48 ºC, and 90 s at 72 ºC, and 30 cycles of 30 s at 95 ºC, 45 s at 55 ºC, and 90 s at 72 ºC, concluding with a 7 min extension step at 72 ºC. The oligonucleotide primers used to amplify the homologous regions are listed in Supplementary Table 4.

Plasmid construction

The plasmids used in this work were constructed following the design represented in Figure 3. Briefly, the chromosomal regions up and downstream of the target loci (Table I) were amplified by PCR using the *Pfu* polymerase and primers containing tags with restriction sites for cloning purposes (N*n*.5O/N*n*.5I and N*n*.3O/N*n*.3I, Supplementary Table 4). The corresponding up and downstream fragments of each locus were fused by "overlap-PCR", using the *Taq* polymerase (see above). The resulting products were cloned into the pGEM®-T easy vector (Promega), according to the manufacturer's instructions, originating the pSN*n** plasmids series (*n* for 'neutral site number'). The pGEM®-T easy vector restriction sites incompatible with the BioBrick RFC[10] standard were eliminated by digesting the vector with *Xho*I and *Sal*I and recircularizing the vector (originating the pSN*n*** series). The BioBrick RFC[10] standard incompatible restriction sites within the amplified chromosomal regions of pSN5** and pSN10** were eliminated using the QuikChange® Multi Site-Directed Mutagenesis Kit (Stratagene), according to the manufacturer's instructions, using the primers N5_SDM and N10_SDM, respectively (Supplementary Table 4), originating the pSN5*** and the pSN10*** plasmids. A BioBrick-compatible cloning interface was synthesized (Epoch Life Science Inc.), containing the *Mun*I, *Xba*I, *Spe*I and *Pst*I restriction sites flanked by two double transcription terminators (based on the registry parts BBa_B1006, BBa_B0062, BBa_B0053 and BBa_B0020). This interface was excised from the delivery vector (pBSK-FP300) using *Xma*I and *Age*I and

subsequently cloned into the *Xma*I site of the pSN*n*** or pSN*n**** plasmids, originating the pSN*n* plasmid series (see Figure 3). Afterwards, two selection cassettes, one containing the *nptII* gene (encoding a neomycin phosphotransferase that confers resistance to neomycin and kanamycin) and the *sacB* gene (encoding a levansucrase that confers sucrose sensitivity), and another containing only the *nptII* gene, were amplified by PCR from the plasmid pK18mobsacB (Schäfer *et al.*, 1994), using the primer pairs Km.KmScFwd/KmScRev and Km.KmScFwd/KmRev, respectively (Supplementary Table 4). These selection cassettes were cloned into the pGEM-T easy vector, digested with *Xma*I and cloned into the *Xma*I site of the pSN*n* plasmid series. The final plasmid series containing only the Km-resistance cassette was named pSN*n*K and the plasmid series containing the cassette with the two selective markers was named pSN*n*KS. Additionally, based on the works of Ericksson *et al.* (2000), Agrawal *et al.* (2001) and Shibato *et al.* (2002) the minimal promoter sequence required for *Synechocystis psbA2* basal transcription (-38 to +14 bp; +1 corresponding to the transcription start site) was synthesized (P_{psbA2*}), meeting the BioBrick RFC[10] standard specifications (DNA 2.0 Inc.). This promoter region was cloned into the BioBrick vector pSB1A3 containing the part BBa_E0240 (composed by the RBS part BBa_B0032, the GFP-coding gene part BBa_E0040 and the double transcription terminator part BBa_B0015), according to the BioBrick standard cloning procedure. The BioBrick composite part BBa_E0240 (promoterless GFP generator) and the part with the synthetic constitutive promoter P_{psbA2*} were digested with *Xba*I and *Spe*I and cloned into the *Xba*I site of both pSN*n* and pSN*n*K series, originating the pSN*n*.gfp, pSN*n*K.gfp, pSN*n*.Cgfp, pSN*n*K.Cgfp plasmid series (C for 'constitutive promoter'). These plasmids were used to assess proper cloning interface transcriptional insulation (promoterless *gfp*) and functionality of the neutral sites (*gfp* under the control of the constitutive promoter). The vectors were named as follows: p for 'plasmid', S for '*Synechocystis*', N for 'neutral site', *n* for the 'number of the neutral site' (5, 8, 10, 15 or 16), K for 'kanamycin resistance cassette', KS for the 'double selection cassette (Km^R/Suc^S)', C for 'constitutive P_{psbA2*} promoter', and gfp for the 'GFP generator BioBrick part BBa_E0240'. Plasmids DNA was isolated from overnight grown *E. coli* cultures, using the GenElute^{TM} Plasmid Miniprep Kit (Sigma-Aldrich), according to the manufacturer's instructions. All the BioBrick parts can be found at the Registry of Standard Biological Parts website (http://partsregistry.org/) and GenBank files of the final plasmids are available at the supplementary material in the CD-ROM attached.

Generation of *Synechocystis* mutants

Synechocystis was transformed based on the procedure described previously (Williams, 1988). Briefly, *Synechocystis* was grown in standard conditions to an OD_{730}≈0.5. Cells

were harvested by centrifugation (10 min at 3850 *g*) and resuspended in BG11 to a final OD_{730}≈2.5. The purified plasmids were incubated for 5h with 500 µL of those cells (final plasmid DNA concentration of 20 µg/mL), in light and at 25 ºC. Cells were then spread onto Immobilon™-NC membranes (0.45 µm pore size, 82 mm, Millipore) resting on solid BG11 plates, grown in standard growth conditions, and transferred to selective plates after 24 h (Km, 10 µg/mL). Transformants were observed after 1-2 weeks. For complete segregation, Km-resistant colonies were grown at increasing Km concentrations (25 and 50 µg/mL) and finally transferred into liquid medium. The mutant full segregation was confirmed by PCR, using primers within (N*n*F and N*n*R, Supplementary Table 3) and external to each site (N*n*.5O and N*n*.3O, Supplementary Table 4), and by Southern blot with probes labeled with digoxigenin using the DIG DNA labeling kit (Roche Molecular Biochemicals) and following the manufacturer's instructions. The probes were amplified with primers N5.3O and N5.3I for mutants in N5, N8.3O and N8.3I for mutants in N8, N10.3O and N10.3I for mutants in N10, N15.5O and N15.5I for mutants in N15, and N16.3O and N16.3I for mutants in N16 (see Supplementary Table 4). gDNA digestions were performed with *Mun*I, with the exception of N8 mutants for which the gDNA was digested with *Spe*I. When necessary antibiotic concentration was increased up to 300 µg/mL to fully segregate the mutants. Second mutants were produced by transforming SN*n*KS mutants with pSN*n*.(C)gfp plasmids as described above, with the difference that cells were grown on solid BG11 plates for 48 h and then transferred to solid BG11 plates containing 5% (w/v) sucrose. Transformants were observed after 2-3 weeks. The mutants were named as follows: S for '*Synechocystis*', N for 'neutral site', *n* for the 'number of the neutral site' (5, 8, 10, 15 or 16), K for 'kanamycin resistance cassette', KS for the 'double selection cassette (Km^R/Suc^S)', C for 'constitutive P_{psbA2*} promoter', and gfp for the 'GFP generator BioBrick part BBa_E0240'.

Growth experiments

For the growth experiments, 50 mL of BG11 medium were inoculated with *Synechocystis* wild-type or SN*n*K mutants to OD_{730}≈0.3. The cultures were grown to OD_{730}≈1.0, under the experimental light regimen (continuous light (20 µmol photons/m^2/s) or 12 h light (20 µmol photons/m^2/s)/12 h dark cycles) and subsequently diluted to a final OD_{730}≈0.1-0.2 in fresh BG11. 2 mL of each diluted culture were added to 4.5 mL sterile cuvettes (see Lopo *et al.*, 2012) and, when necessary, glucose was added to a final concentration of 5 mM. The cuvettes were closed with sterilized Parafilm® and placed in acrylic racks, in a chamber with constant temperature (28 ºC or 30 ºC) and irradiance (20 µmol photons/m^2/s). Similarly, 10 mL of each diluted culture was added to sterile borosilicate glass flasks, closed with cellulose stopper, and placed in a chamber with constant temperature (30 ºC)

and irradiance (20 µmol photons/m^2/s). All experiments were performed at least in triplicate and the optical density of the cultures was recorded daily at 730 nm for 10 days, on a UV-Vis Spectrophotometer UVmini 1240 (Shimadzu).

Statistical analysis of cyanobacteria specific growth

The quantitative analysis of *Synechocystis* wild-type and neutral sites mutant growth in various conditions was based on the OD measurements (see above). For each bin, i, a specific growth parameter, representing the exponent of exponential growth within the given time interval was calculated as:

$$\mu_i = \frac{\log[x_{i+1}/x_i]}{t_{i+1} - t_i}$$

where t_i is time, in units of days running from day 0 to $t_{i_{\max}}$ and x_i is the corresponding measured OD. As test duration was 10 days in all strains and stationary growth was reached at similar moments, the specific growth rate of all the exponential phase of each replica of each organism was considered to be a reliable parameter to assess growth patterns. ANOVA statistical analysis of the growth data was performed to compare the wild-type to each mutant organism, using the software Wolfram's Mathematica©. This ANOVA was four-way in cuvettes set and three-way in flasks set, where a model with all the effects and interactions up to the third order was tested. Values of up to second order interactions can be studied in Supplementary Table 5.

Confocal microscopy

Cells of wild-type *Synechocystis* and mutants harboring the *gfp* gene (SN*n*K.gfp and SN*n*K.Cgfp mutants series), were grown in liquid BG11 medium under the maintenance growth conditions described above. Samples (5 µL) were loaded on a glass slide, followed by 12.5 µL of 1% (w/v) low-melting-point agarose dissolved in BG11 medium, and this mixture was finally covered with a coverslip. The GFP emission (collected between 500 nm and 540 nm) was observed when the cells were exposed to a laser beam at 488 nm, using a Leica TCS SP5 II confocal microscope (Leica Microsystems). Cyanobacterial autofluorescence was collected between 640nm and 680 nm after excitation at 633 nm. Images were analyzed with the LAS AF Lite v2.6 software (Leica Microsystems).

Western blot

Total protein extractions were carried out using a lysis buffer composed of 10 mM HEPES pH 8.0, 10 mM EDTA pH 8.0, 0.5% (v/v) Triton X-100, 2 mM DTT supplemented with a cocktail of protease inhibitors (cOmplete Mini, EDTA free, Roche). Cells were collected by

centrifugation, resuspended in lysis buffer and disrupted by sonication on ice. Cell debris was separated by two rounds of 10 min centrifugation at 10 000 *g* and 4 ºC. Protein extracts were saved at -20 ºC until further analysis. Protein quantification was determined by the BCA method (Thermo Scientific) and 15 µg of each extract was separated by electrophoresis on 12% (w/v) SDS-polyacrylamide gels. Protein extracts were visualized with colloidal Coomassie blue staining (Sigma) or transferred to a Hybond™-C extra nitrocellulose membrane (GE Healthcare). After the transfer, membranes were blocked for 3 h with 5% (w/v) milk powder in TTBS (Tris buffered saline (TBS) with 0.05% (v/v) Tween 20) and then incubated overnight at 4 ºC with monoclonal mouse anti-GFP (Roche) in fresh blocking buffer. After washing with TTBS, blots were incubated for 1 h with goat-anti-mouse IgG (Invitrogen) linked to horseradish peroxidase. Membranes were washed with TTBS before immunodetection, which was performed by chemiluminescence using ECL Western Blotting Analysis System detection reagents.

Acknowledgements

This work was financially supported by EU FP6-NEST-2005-Path-SYN project BioModularH2 (contract no. 043340), by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011) and grants SFRH/BD/36378/2007, SFRH/BPD/64095/2009, SFRH/BPD/74894/2010.

The authors are grateful to Dr Carol Harley for reading the manuscript, Miguel Lopo for the preliminary growth experiments, and the 'National BioResource Project (NIG, Japan): E. coli' for providing the plasmid pK18mobsacB.

Author Contributions

FP and PT conceived and designed the study. FP, CP and PO carried out the experiments, and AM and JU performed the statistical analysis of the data. FP and PT wrote the manuscript. PT coordinated the work and manuscript. All authors gave intellectual input and comments on the manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Abed RMM, Dobretsov S & Sudesh K (2009) Applications of cyanobacteria in biotechnology. *Journal of Applied Microbiology* **106**: 1-12.

Agrawal GK, Kato H, Asayama M & Shirai M (2001) An AU-box motif upstream of the SD sequence of light-dependent *psbA* transcripts confers mRNA instability in darkness in cyanobacteria. *Nucleic Acids Research* **29**: 1835-1843.

Albertsen L, Chen Y, Bach LS, Rattleff S, Maury J, Brix S, Nielsen J & Mortensen UH (2011) Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. *Applied and Environmental Microbiology* **77**: 1033-1040.

Aoki R, Goto T & Fujita Y (2011) A heme oxygenase isoform is essential for aerobic growth in the cyanobacterium *Synechocystis* sp. PCC 6803: modes of differential operation of two isoforms/enzymes to adapt to low oxygen environments in cyanobacteria. *Plant & Cell Physiology* **52**: 1744-1756.

Aoki S, Kondo T & Ishiura M (1995) Circadian expression of the *dnaK* gene in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* **177**: 5606-5611.

Becker EC & Meyer RJ (1997) Acquisition of resistance genes by the IncQ plasmid R1162 is limited by its high copy number and lack of a partitioning mechanism. *Journal of Bacteriology* **179**: 5947-5950.

Berla BM & Pakrasi HB (2012) Up-regulation of plasmid-encoded genes during stationary phase in *Synechocystis* sp. PCC 6803, a cyanobacterium. *Applied and Environmental Microbiology* **78**: 5448-5451.

Biswas R, Yamaoka M, Nakayama H, Kondo T, Yoshida K-i, Bisaria V & Kondo A (2012) Enhanced production of 2,3-butanediol by engineered *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **94**: 651-658.

Burnap RL, Qian M, Shen JR, Inoue Y & Sherman LA (1994) Role of disulfide linkage and putative intermolecular binding residues in the stability and binding of the extrinsic manganese-stabilizing protein to the photosystem II reaction center. *Biochemistry* **33**: 13712-13718.

Cantone I, Marucci L, Iorio F, Ricci MA, Belcastro V, Bansal M, Santini S, di Bernardo M, di Bernardo D & Cosma MP (2009) A yeast synthetic network for *in vivo* assessment of reverse-engineering and modeling approaches. *Cell* **137**: 172-181.

Castillo-Davis CI (2005) The evolution of noncoding DNA: how much junk, how much func? *Trends in Genetics* **21**: 533-536.

Chen C-Y, Yeh K-L, Aisyah R, Lee D-J & Chang J-S (2011) Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: a critical review. *Bioresource Technology* **102**: 71-81.

Clerico EM, Ditty JL & Golden SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Circadian Rhythms,* Vol. 362 (Rosato Ezio), pp. 155-171. Humana Press, Totowa, USA.

David M, Vielma M & Julliot J-S (1983) Introduction of IncQ plasmids into *Rhizobium meliloti*. Isolation of a host-range mutant of RSF1010 plasmid. *FEMS Microbiology Letters* **16**: 335-341.

De Gelder L, Ponciano JM, Joyce P & Top EM (2007) Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology* **153**: 452-463.

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. *Molecular Microbiology* **53**: 65-80.

Eriksson J, Salih GF, Ghebramedhin H & Jansson C (2000) Deletion mutagenesis of the 5' *psbA2* region in *Synechocystis* 6803: identification of a putative cis element involved in photoregulation. *Molecular Cell Biology Research Communications* **3**: 292-298.

Ferreira D, Pinto F, Moradas-Ferreira P, Mendes M & Tamagnini P (2009) Transcription profiles of hydrogenases related genes in the cyanobacterium *Lyngbya majuscula* CCAP 1446/4. *BMC Microbiology* **9**: 67.

Fu PC (2009) Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion. *Journal of Chemical Technology and Biotechnology* **84**: 473-483.

García-Domínguez M, Reyes JC & Florencio FJ (2000) NtcA represses transcription of *gifA* and *gifB*, genes that encode inhibitors of glutamine synthetase type I from *Synechocystis* sp. PCC 6803. *Molecular Microbiology* **35**: 1192-1201.

Griese M, Lange C & Soppa J (2011) Ploidy in cyanobacteria. *FEMS Microbiology Letters* **323**: 124-131.

Heidorn T, Camsund D, Huang H-H, Lindberg P, Oliveira P, Stensjö K & Lindblad P (2011) Synthetic biology in cyanobacteria: engineering and analyzing novel functions. *Methods in Enzymology,* Vol. Volume 497 (Chris Voigt), pp. 539-579. Academic Press.

Hou L (2010) Improved production of ethanol by novel genome shuffling in *Saccharomyces cerevisiae*. *Applied Biochemistry and Biotechnology* **160**: 1084-1093.

Huang B, Guo J, Yi B, Yu X, Sun L & Chen W (2008) Heterologous production of secondary metabolites as pharmaceuticals in *Saccharomyces cerevisiae*. *Biotechnology Letters* **30**: 1121-1137.

Huang H-H, Camsund D, Lindblad P & Heidorn T (2010) Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Research* **38**: 2577-2593.

Huckauf J, Nomura C, Forchhammer K & Hagemann M (2000) Stress responses of *Synechocystis* sp. strain PCC 6803 mutants impaired in genes encoding putative alternative sigma factors. *Microbiology* **146**: 2877-2889.

Ishii A & Hihara Y (2008) An AbrB-like transcriptional regulator, Sll0822, is essential for the activation of nitrogen-regulated genes in *Synechocystis* sp. PCC 6803. *Plant Physiology* **148**: 660-670.

Janga SC & Collado-Vides J (2007) Structure and evolution of gene regulatory networks in microbial genomes. *Research in Microbiology* **158**: 787-794.

Jia X, Li S, Xie S & Wen J (2012) Engineering a metabolic pathway for isobutanol biosynthesis in *Bacillus subtilis*. *Applied Biochemistry and Biotechnology* **168**: 1-9.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M & Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research* **3**: 109-136.

Kaneko T, Nakamura Y, Sasamoto S, Watanabe A, Kohara M, Matsumoto M, Shimpo S, Yamada M & Tabata S (2003) Structural analysis of four large Plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Research* **10**: 221-228.

Keskin T & Hallenbeck PC (2012) Hydrogen production from sugar industry wastes using single-stage photofermentation. *Bioresource Technology* **112**: 131-136.

Kim I-K, Roldão A, Siewers V & Nielsen J (2012) A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Research* **12**: 228-248.

Koksharova OA & Wolk CP (2002) Genetic tools for cyanobacteria. *Applied Microbiology and Biotechnology* **58**: 123-137.

Krauter H, Willke T & Vorlop K-D (2012) Production of high amounts of 3-hydroxypropionaldehyde from glycerol by *Lactobacillus reuteri* with strongly increased biocatalyst lifetime and productivity. *New Biotechnology* **29**: 211-217.

Kufryk GI, Sachet M, Schmetterer G & Vermaas WFJ (2002) Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiology Letters* **206**: 215-219.

Kunert A, Hagemann M & Erdmann N (2000) Construction of promoter probe vectors for *Synechocystis* sp. PCC 6803 using the light-emitting reporter systems Gfp and LuxAB. *Journal of Microbiological Methods* **41**: 185-194.

Leitão E, Pereira S, Bondoso J, Ferreira D, Pinto F, Moradas-Ferreira P & Tamagnini P (2006) Genes involved in the maturation of hydrogenase(s) in the nonheterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4. *International Journal of Hydrogen Energy* **31**: 1469-1477.

Li P, Miao X, Li R & Zhong J (2011a) *In situ* biodiesel production from fast-growing and high oil content *Chlorella pyrenoidosa* in rice straw hydrolysate. *Journal of Biomedicine and Biotechnology* **2011**: 141207.

Li S, Wen J & Jia X (2011b) Engineering *Bacillus subtilis* for isobutanol production by heterologous Ehrlich pathway construction and the biosynthetic 2-ketoisovalerate precursor pathway overexpression. *Applied Microbiology and Biotechnology* **91**: 577-589.

Lopo M, Montagud A, Navarro E, Cunha I, Zille A, de Córdoba PF, Moradas-Ferreira P, Tamagnini P & Urchueguía JF (2012) Experimental and modeling analysis of *Synechocystis* sp. PCC 6803 growth. *Journal of Molecular Microbiology and Biotechnology* **22**: 71-82.

Malcata FX (2011) Microalgae and biofuels: a promising partnership? *Trends in Biotechnology* **29**: 542-549.

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N & Bryant SH (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research* **37**: D205-D210.

Marraccini P, Bulteau S, Cassier-Chauvat C, Mermet-Bouvier P & Chauvat F (1993) A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera *Synechococcus* and *Synechocystis*. *Plant Molecular Biology* **23**: 905-909.

Martínez J, Liu L, Petranovic D & Nielsen J (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Current Opinion in Biotechnology* **23**: 965-971.

Matsunaga T & Takeyama H (1995) Genetic engineering in marine cyanobacteria. *Journal of Applied Phycology* **7**: 77-84.

Mermet-Bouvier P, Cassier-Chauvat C, Marraccini P & Chauvat F (1993) Transfer and replication of RSF1010-derived plasmids in several cyanobacteria of the genera *Synechocystis* and *Synechococcus*. *Current Microbiology* **27**: 323-327.

Mermet-Bouvier P & Chauvat F (1994) A conditional expression vector for the cyanobacteria *Synechocystis* sp. strains PCC6803 and PCC6714 or *Synechococcus* sp. strains PCC7942 and PCC6301. *Current Microbiology* **28**: 145-148.

Meyer R (2009) Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62**: 57-70.

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. PCC6803. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 2124-2129.

Montagud A, Navarro E, Fernandez de Cordoba P, Urchueguia J & Patil K (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology* **4**: 156.

Montagud A, Zelezniak A, Navarro E, de Córdoba PF, Urchueguía JF & Patil KR (2011) Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal* **6**: 330-342.

Morris KV (2012) *Non-coding RNAs and the epigenetic regulation of gene expression: drivers of natural selection*. Caister Academic Press, Norfolk, UK.

Nakamura Y, Kaneko T, Hirosawa M, Miyajima N & Tabata S (1998) CyanoBase, a www database containing the complete nucleotide sequence of the genome of *Synechocystis* sp. strain PCC6803. *Nucleic Acids Research* **26**: 63-67.

Nakamura Y, Kaneko T & Tabata S (2000) CyanoBase, the genome database for *Synechocystis* sp. strain PCC6803: status for the year 2000. *Nucleic Acids Research* **28**: 72.

Ohkawa H, Price GD, Badger MR & Ogawa T (2000) Mutation of *ndh* genes leads to inhibition of CO_2 uptake rather than HCO_3^- uptake in *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* **182**: 2591-2596.

Oliveira P & Lindblad P (2008) An AbrB-like protein regulates the expression of the bidirectional hydrogenase in *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* **190**: 1011-1019.

Pinto F, van Elburg KA, Pacheco CC, Lopo M, Noirel J, Montagud A, Urchueguía JF, Wright PC & Tamagnini P (2012) Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC 6803 mutant lacking a functional bidirectional hydrogenase. *Microbiology* **158**: 448-464.

Priefer UB, Simon R & Pühler A (1985) Extension of the host range of *Escherichia coli* vectors by incorporation of RSF1010 replication and mobilization functions. *Journal of Bacteriology* **163**: 324-330.

Sambrook J & Russell DW (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York, USA.

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.

Scholz P, Haring V, Wittmann-Liebold B, Ashman K, Bagdasarian M & Scherzinger E (1989) Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* **75**: 271-288.

Sergeyenko TV & Los DA (2000) Identification of secreted proteins of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *FEMS Microbiology Letters* **193**: 213-216.

Shibato JS, Agrawal GA, Kato HK, Asayama MA & Shirai MS (2002) The 5'-upstream *cis*-acting sequences of a cyanobacterial *psbA* gene: analysis of their roles in basal, light-dependent and circadian transcription. *Molecular Genetics and Genomics* **267**: 684-694.

Sode K, Tatara M, Ogawa S & Matsunaga T (1992) Maintenance of broad host range vector pKT230 in marine unicellular cyanobacteria. *FEMS Microbiology Letters* **99**: 73-78.

Stanier RY, Kunisawa R, Mandel M & Cohenbaz.G (1971) Purification and properties of unicellular blue-green algae (order *Chroococcales*). *Bacteriological Reviews* **35**: 171-205.

Tamagnini P, Troshina O, Oxelfelt F, Salema R & Lindblad P (1997) Hydrogenases in *Nostoc* sp. strain PCC 73102, a strain lacking a bidirectional enzyme. *Applied and Environmental Microbiology* **63**: 1801-1807.

Velappan N, Sblattero D, Chasteen L, Pavlik P & Bradbury ARM (2007) Plasmid incompatibility: more compatible than previously thought? *Protein Engineering Design and Selection* **20**: 309-313.

Vermaas W (1996) Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: Principles and possible biotechnology applications. *Journal of Applied Phycology* **8**: 263-273.

Vioque A (2007) Transformation of cyanobacteria. Vol. 616 (León Rosa, Galván Aurora & Fernández Emilio), pp. 12-22. Springer New York.

Wang H-L, Postier BL & Burnap RL (2004) Alterations in global patterns of gene expression in *Synechocystis* sp. PCC 6803 in response to inorganic carbon limitation and the inactivation of *ndhR*, a LysR family regulator. *Journal of Biological Chemistry* **279**: 5739-5751.

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD & Paddon CJ (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E111-E118.

Whitton BA & Potts M (2000) Introduction to the Cyanobacteria. *The Ecology of Cyanobacteria - Their Diversity in Time and Space,* (Whitton Brian & Potts Malcolm), pp. 1-11. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Williams JGK (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods in Enzymology* **167**: 766-778.

Xu WD & McFadden BA (1997) Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* **37**: 95-104.

Yamaoka M, Osawa S, Morinaga T, Takenaka S & Yoshida K-i (2011) A cell factory of *Bacillus subtilis* engineered for the simple bioconversion of myo-inositol to scyllo-inositol, a potential therapeutic agent for Alzheimer's disease. *Microbial Cell Factories* **10**: 69.

Yang XY & McFadden BA (1993) A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC 6803 encodes a Rep protein and replicates by a rolling circle mechanism. *Journal of Bacteriology* **175**: 3981-3991.

Yang XY & McFadden BA (1994) The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid* **31**: 131-137.

Zang XN, Liu B, Liu SM, Arunakumara K & Zhang XC (2007a) Optimum conditions for transformation of *Synechocystis* sp. PCC 6803. *Journal of Microbiology* **45**: 241-245.

Zang XN, Liu B, Liu SM, Sun PN, Zhang XQ & Zhang XC (2007b) Transformation and expression of *Paralichthys olivaceus* growth hormone cDNA in *Synechocystis* sp. PCC 6803. *Aquaculture* **266**: 63-69.

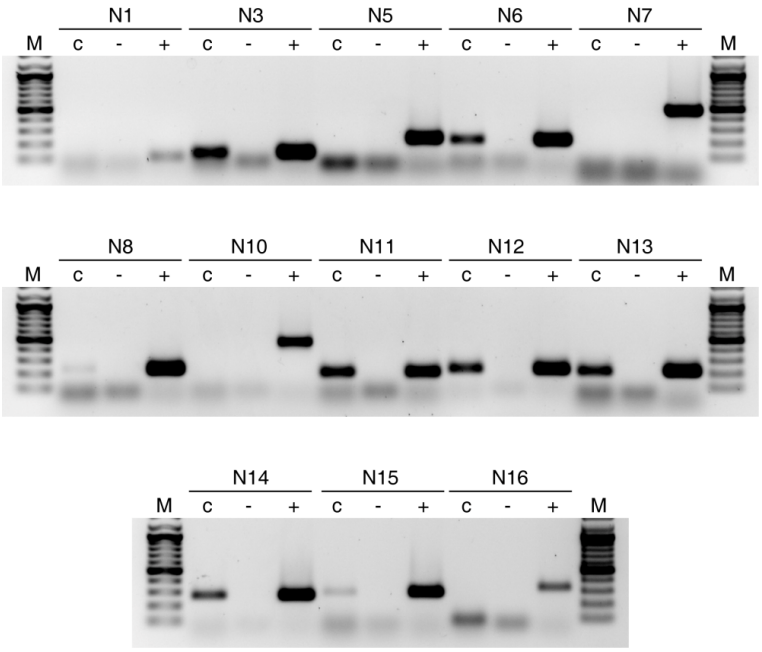
Zeng A-P (2011) Meeting report: New biotechnological production systems. *Biotechnology Journal* **6**: 907-908.

Zhou H, Hu H & Lai M (2010) Non-coding RNAs and their epigenetic regulatory mechanisms. *Biology of the Cell* **102**: 645-655.

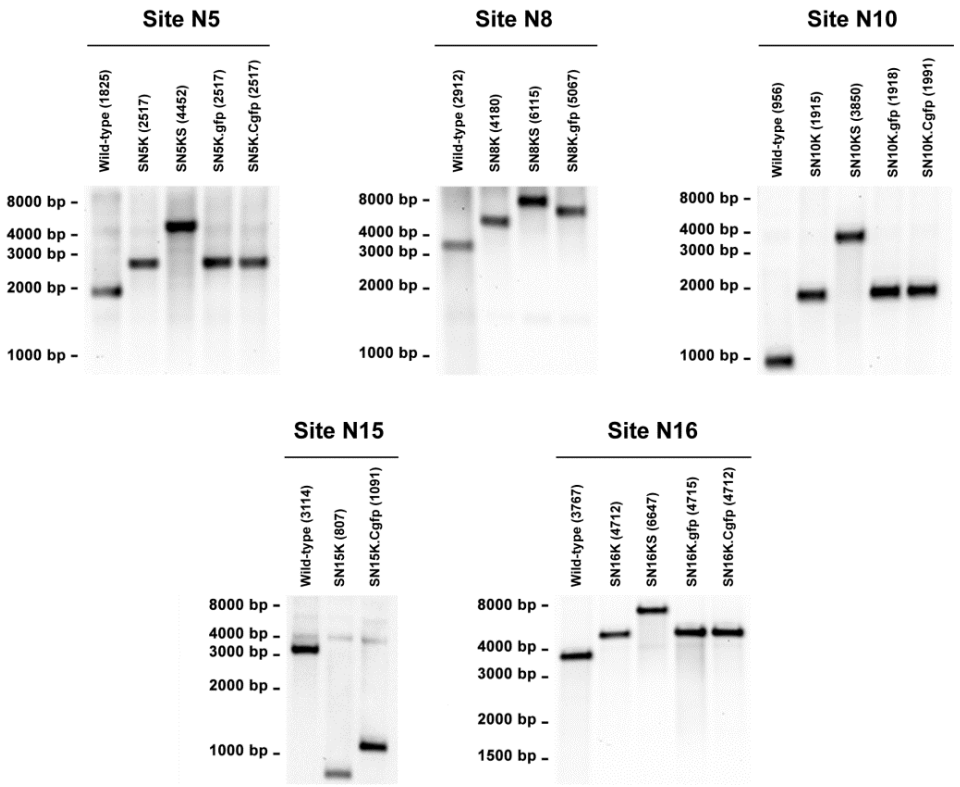
Zhou Y, Ma X, Hou Z, Xue X, Meng J, Li M, Jia M & Luo X (2012) High cell density cultivation of recombinant *Escherichia coli* for prodrug of recombinant human GLPs production. *Protein Expression and Purification* **85**: 38-43.

Supplementary material II

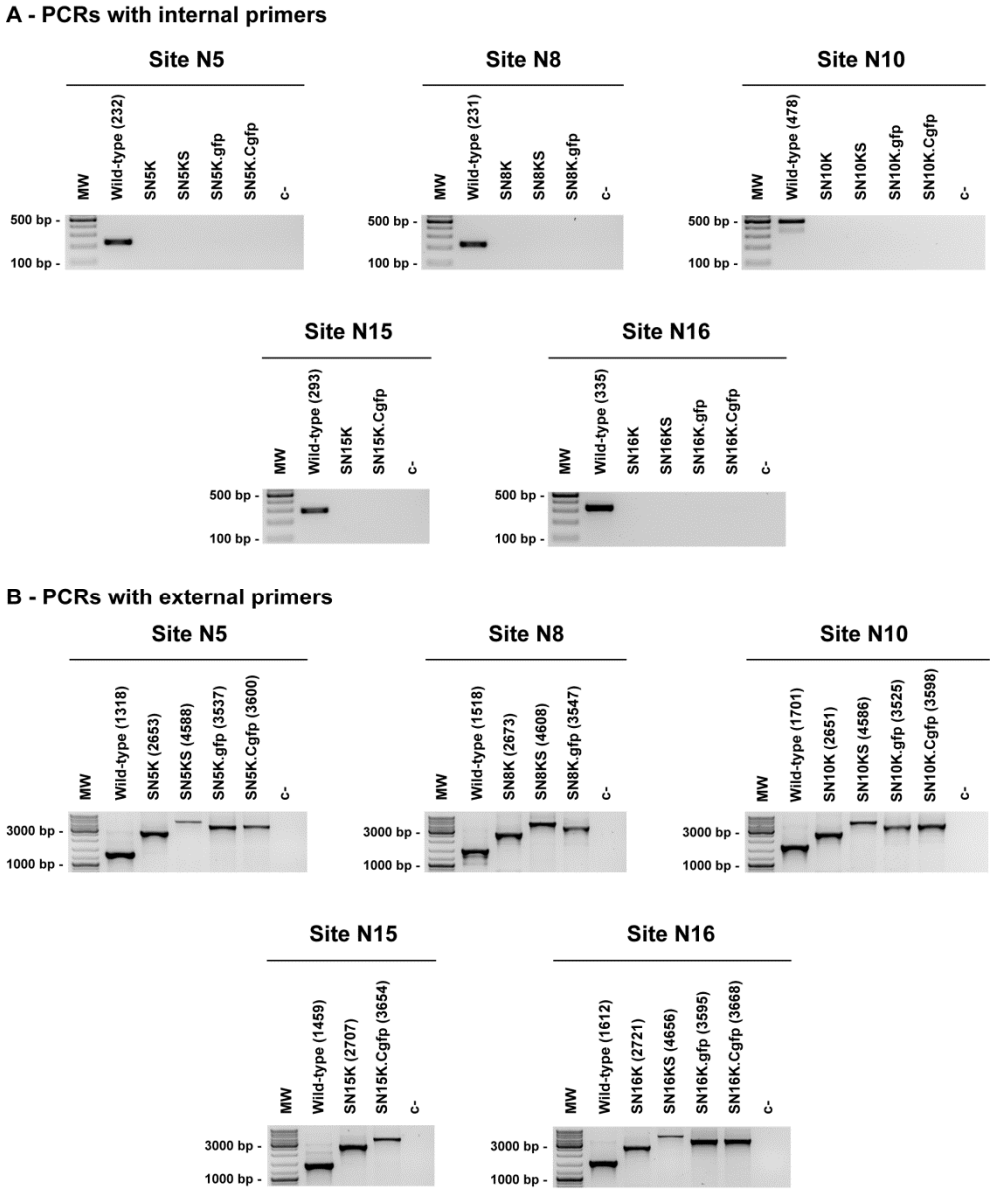
Supplementary figures



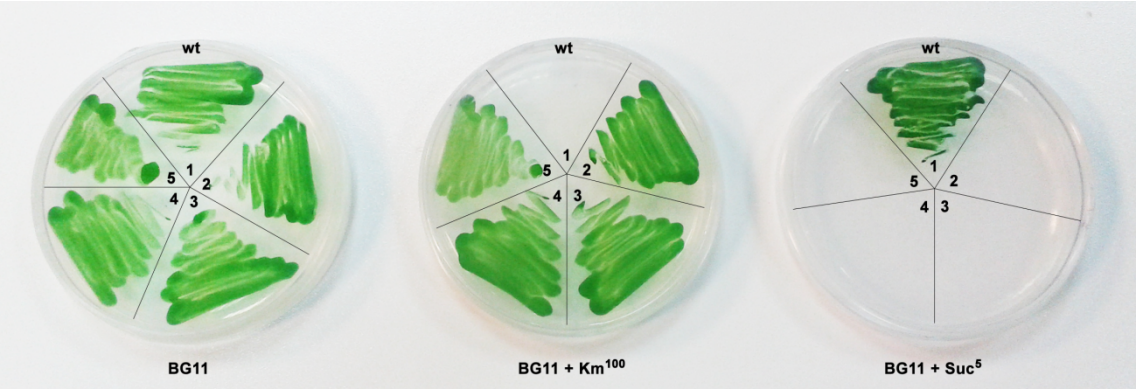
Supplementary Figure 1 | Transcription analysis of the thirteen loci selected by RT-PCR. *Synechocystis* wild-type samples for RNA extraction were collected at OD_{730} ≈ 0.8-0.9. These results are representative of the three biological triplicates and technical duplicates. cDNA (c), negative control (-, absence of template) and positive control (+, genomic DNA) are indicated.



Supplementary Figure 2 I Confirmation of the full segregation of the *Synechocystis* mutants by Southern blot. Expected bands sizes are within brackets (in base pairs).



Supplementary Figure 3 I Confirmation of the full segregation of the *Synechocystis* mutants by PCR. The PCR reactions were performed using primers within the ORF corresponding to each neutral site (A) and with primers external to each site (B). No template controls were always included (c-) and the expected bands sizes are within brackets (in base pairs). MW – GeneRuler™ DNA Ladder Mix (Fermentas).



Supplementary Figure 4 | Phenotypic characterization of mutants harboring the double selection cassette conferring kanamycin resistance and sucrose sensitivity (SN*n*KS). *Synechocystis* wild-type and mutants were grown in BG11 medium (BG11), BG11 medium supplemented with 100 µg/mL kanamycin (BG11 + Km^{100}) and BG11 medium supplemented with 5% (m/v) sucrose (BG11 + Suc^{5}). 1 – wild-type, 2 – *Synechocystis* mutant SN5KS, 3 – *Synechocystis* mutant SN8KS, 4 – *Synechocystis* mutant SN10KS, 5 – *Synechocystis* mutant SN16KS.

Supplementary tables

Supplementary Table 1 | List of plasmids constructed in this work

Plasmid name	Description
pSN5	pGEM-T easy vector containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site N5.
pSN8	pGEM-T easy vector containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site N8.
pSN10	pGEM-T easy vector containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site N10.
pSN15	pGEM-T easy vector containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site N15.
pSN16	pGEM-T easy vector containing the BioBrick compatible interface flanked by the two regions for double homologous recombination on neutral site N16.
pSN5K	pSN5 plasmid with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.
pSN8K	pSN8 plasmid with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.
pSN10K	pSN10 plasmid with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.
pSN15K	pSN15 plasmid with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.
pSN16K	pSN16 plasmid with the kanamycin resistance cassette cloned upstream of the BioBrick compatible interface.
pSN5KS	pSN5 plasmid with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.
pSN8KS	pSN8 plasmid with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.
pSN10KS	pSN10 plasmid with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.
pSN15KS	pSN15 plasmid with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.
pSN16KS	pSN16 plasmid with the kanamycin resistance/sucrose sensitivity cassette cloned upstream of the BioBrick compatible interface.
pSN5.gfp	pSN5 plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN8.gfp	pSN8 plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN10.gfp	pSN10 plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN15.gfp	pSN15 plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN16.gfp	pSN16 plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN5.Cgfp	pSN5 plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN8.Cgfp	pSN8 plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN10.Cgfp	pSN10 plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.

pSN15.Cgfp	pSN15 plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN16.Cgfp	pSN16 plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN5K.gfp	pSN5K plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN8K.gfp	pSN8K plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN10K.gfp	pSN10K plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN15K.gfp	pSN15K plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN16K.gfp	pSN16K plasmid with the GFP generator BioBrick BBa_E0240 cloned in the BioBrick compatible interface.
pSN5K.Cgfp	pSN5K plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN8K.Cgfp	pSN8K plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN10K.Cgfp	pSN10K plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN15K.Cgfp	pSN15K plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.
pSN16K.Cgfp	pSN16K plasmid with the GFP generator BioBrick BBa_E0240 under the control of a constitutive promoter (P_{psbA2*}), cloned in the BioBrick compatible interface.

Supplementary Table 2 | List of mutants constructed in this work

Mutant name	Description	Status
SN5K	*Synechocystis* mutant with replacement of the neutral site N5 by a kanamycin resistance cassette.	Segregated
SN8K	*Synechocystis* mutant with replacement of the neutral site N8 by a kanamycin resistance cassette.	Segregated
SN10K	*Synechocystis* mutant with replacement of the neutral site N10 by a kanamycin resistance cassette.	Segregated
SN15K	*Synechocystis* mutant with replacement of the neutral site N15 by a kanamycin resistance cassette.	Segregated
SN16K	*Synechocystis* mutant with replacement of the neutral site N16 by a kanamycin resistance cassette.	Segregated
SN5KS	*Synechocystis* mutant with replacement of the neutral site N5 by a kanamycin resistance/sucrose sensitivity cassette.	Segregated
SN8KS	*Synechocystis* mutant with replacement of the neutral site N8 by a kanamycin resistance/sucrose sensitivity cassette.	Segregated
SN10KS	*Synechocystis* mutant with replacement of the neutral site N10 by a kanamycin resistance/sucrose sensitivity cassette.	Segregated
SN15KS	*Synechocystis* mutant with replacement of the neutral site N15 by a kanamycin resistance/sucrose sensitivity cassette.	Under construction
SN16KS	*Synechocystis* mutant with replacement of the neutral site N16 by a kanamycin resistance/sucrose sensitivity cassette.	Segregated
SN5K.gfp	*Synechocystis* mutant with replacement of the neutral site N5 by a kanamycin resistance cassette and the promoterless GFP encoding gene.	Segregated
SN8K.gfp	*Synechocystis* mutant with replacement of the neutral site N8 by a kanamycin resistance cassette and the promoterless GFP encoding gene.	Segregated
SN10K.gfp	*Synechocystis* mutant with replacement of the neutral site N10 by a kanamycin resistance cassette and the promoterless GFP encoding gene.	Segregated
SN15K.gfp	*Synechocystis* mutant with replacement of the neutral site N15 by a kanamycin resistance cassette and the promoterless GFP encoding gene.	Under segregation
SN16K.gfp	*Synechocystis* mutant with replacement of the neutral site N16 by a kanamycin resistance cassette and the promoterless GFP encoding gene.	Segregated
SN5K.Cgfp	*Synechocystis* mutant with replacement of the neutral site N5 by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	Segregated
SN8K.Cgfp	*Synechocystis* mutant with replacement of the neutral site N8 by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	Under construction
SN10K.Cgfp	*Synechocystis* mutant with replacement of the neutral site N10 by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	Segregated
SN15K.Cgfp	*Synechocystis* mutant with replacement of the neutral site N15 by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	Segregated
SN16K.Cgfp	*Synechocystis* mutant with replacement of the neutral site N16 by a kanamycin resistance cassette and the GFP encoding gene under the control of a constitutive promoter (P_{psbA2*}).	Segregated
SN5.gfp	*Synechocystis* mutant derived from SN5KS mutant with replacement of the double selection cassette on neutral site N5 by the promoterless GFP encoding gene.	Under confirmation
SN8.gfp	*Synechocystis* mutant derived from SN8KS mutant with replacement of the double selection cassette on neutral site N8 by the promoterless GFP encoding gene.	Under confirmation

SN10.gfp	*Synechocystis* mutant derived from SN10KS mutant with replacement of the double selection cassette on neutral site N10 by the promoterless GFP encoding gene.	Under confirmation
SN15.gfp	*Synechocystis* mutant derived from SN15KS mutant with replacement of the double selection cassette on neutral site N15 by the promoterless GFP encoding gene.	To be constructed
SN16.gfp	*Synechocystis* mutant derived from SN16KS mutant with replacement of the double selection cassette on neutral site N16 by the promoterless GFP encoding gene.	Under confirmation
SN5.Cgfp	*Synechocystis* mutant derived from SN5KS mutant with replacement of the double selection cassette on neutral site N5 by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2^*}).	Under confirmation
SN8.Cgfp	*Synechocystis* mutant derived from SN8KS mutant with replacement of the double selection cassette on neutral site N8 by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2^*}).	Under construction
SN10.Cgfp	*Synechocystis* mutant derived from SN10KS mutant with replacement of the double selection cassette on neutral site N10 by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2^*}).	Under confirmation
SN15.Cgfp	*Synechocystis* mutant derived from SN15KS mutant with replacement of the double selection cassette on neutral site N15 by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2^*}).	To be constructed
SN16.Cgfp	*Synechocystis* mutant derived from SN16KS mutant with replacement of the double selection cassette on neutral site N16 by the GFP encoding gene under the control of a constitutive promoter (P_{psbA2^*}).	Under construction

Supplementary Table 3 | Primers used for RT-PCR

Primer name	Sequence 5' → 3'	Amplicon size (bp)
N1F	CATGGATTGAAGGAAGGG	100
N1R	TTTGGTGAACAATTAATGAAC	
N3F	GCAGTCCCTCCTCCAAAC	129
N3R	AAGCCGATATTCAGGTAAGATG	
N5F	GGAAGTGTTGTTGCTGTC	232
N5R	GTCCTTCGCCGTAGATTG	
N6F	AAGGTAATCAAGCAGAAATG	217
N6R	ATTGGTAACAGCACTTCC	
N7F	TGACTGTAAGCAATCTACC	500
N7R	ACTAACGATGGAATCTTGG	
N8F	AACTTAACTTCTATTCGTGAG	231
N8R	GAAACCTTGATAAGCAGTC	
N10F	GATATTGATGTGTATTGC	478
N10R	TTATTATCTTCTGTCTCC	
N11F	TGAATACCACATCTCCCATTGAC	200
N11R	TTTAACCTGCTGGCGTGAC	
N12F	TTCTCACCGCTACTGTTAC	227
N12R	CACTTCTCGCACTAATTCG	
N13F	CCCAAATGATGACCGAAG	216
N13R	CTAGTAACTGCTGTTCCTC	
N14F	AAGTGTCCAAAGCCAAAC	270
N14R	TCAACAATGCCATCTTCC	
N15F	TTAATTGCCACTGCCTTAC	293
N15R	TAGAGACAGCGGATATGC	
N16F	TGATGTGGACGAGGATAC	335
N16R	GGGCTAAGGGAATATGGG	

Supplementary Table 4 | Other primers used in this work

Primer name	Sequence 5' → 3'*	Purpose
Km.KmScFwd	CTGACCCCGGGTGAATGTCAGCTACTGG	Selection cassettes amplification
KmRev	CAAACCCGGGCGATTTACTTTTCGACCTC	
KmScRev	ACAGACCCGGGCAAGCGGATGGCTGATG	
N5.5O**	GCGGCCTCGAGTGGGGGCATCTGCTAGGCAATATTTG	Amplification of the DNA fragments flanking the neutral sites
N5.5I	GATTACACCCGGGTGTAATCGGTTGATTATTTCAGTGGCCCGGCCG ATGGATAATTAC	
N5.3O**	GTCCAAATCAGCATTGCTCTGCCAGGTGAGAC	
N5.3I	GATTACACCCGGGTGTAATCCTGAAGCCTTGACAATCCCCGTTGGT GATTTATACTTAG	
N8.5O**	GAACACTCGAGTGCCAGCAACGACAATG	
N8.5I	GATTACACCCGGGTGTAATCCTTCATTCCCCTATTGTTATTAATAGTG CTC	
N8.3O**	GATGACCGCTGGCGGAGTTTAGTCCAG	
N8.3I	GATTACACCCGGGTGTAATCTTGTCACCAATTTTTGTAGGGATGTTG GGCTAAATTG	
N10.5O**	GATAACTCGAGTCGCCCGGTGGATTTAATGC	
N10.5I	GATTACACCCGGGTGTAATCCCCATACTCAGCCTTCTAATGCTGAGA GAATG	
N10.3O**	GGACTGACCCAAGATACAGTGGTAG	
N10.3I	GATTACACCCGGGTGTAATCCGAGCGCAAACTACAATGCGCTTCGT TTGC	
N15.5O**	GGTTCTCGAGCCGCTGAATTTAGTCAACATCG	
N15.5I	GATTACACCCGGGTGTAATCCATTGGGCACGAGAGTTAGTAAGGCA GTG	
N15.3O*	CTAAACTTACGGCATTGGCATCAACGGGAG	
N15.3I	GATTACACCCGGGTGTAATCTCTTTACAATGGCCAGGTCTTTAGGGA GCGGTGAC	
N16.5O**	GAACTCGAGTAGTAACCACAGGCTTTTG	
N16.5I	GATTACACCCGGGTGTAATCGCTGGAGGCGAACTGGGTGAGAACCA T	
N16.3O**	GTGAGCTTGATGGTGATGGTGGGTAAAG	
N16.3I	GATTACACCCGGGTGTAATCTGCTGGCTTTGTTGCCCTGTCAACCAA AGTTC	
N5_SDM	CGAGGCGATCGCCCAGTTGGAAGAATTGGCC	SDM***
N10_SDM	CTAAAAAGACAAGTCTGTGGCTAGTTACTATGACGAGGC	SDM***
GFP.F	TCTTGTTGAATTAGATGGTG	RT-PCR
GFP.R	TGTGAGTTATAGTTGTATTCC	

*Restriction sites underlined
**Primers used to confirm genomic integrations and mutants full segregation.
***SDM – Site directed mutagenesis

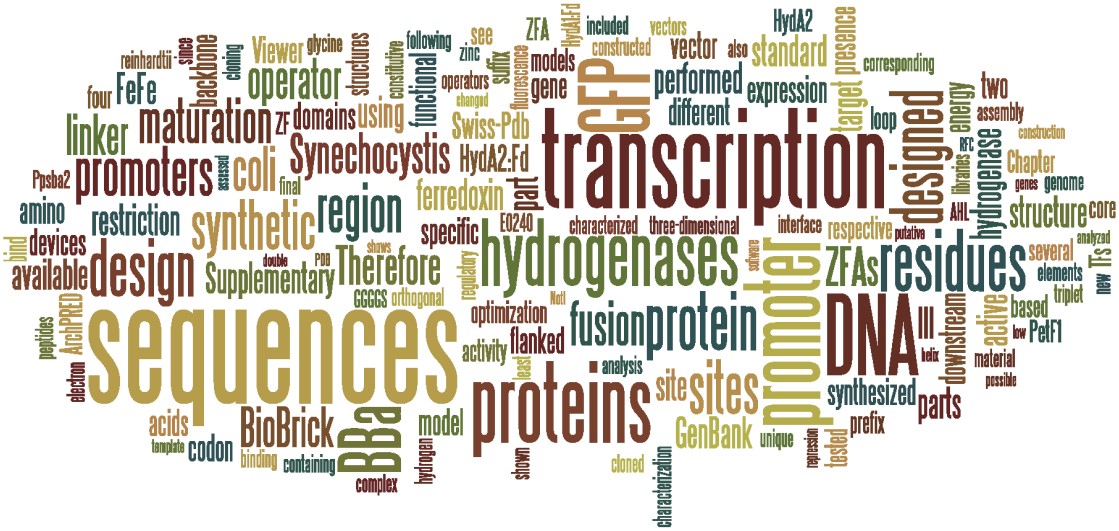
Supplementary Table 5 | *P*-values of up to second order interactions produced by a four-way (cuvettes set) and a three-way (flasks set) ANOVA analysis.

	Cuvettes					Flasks				
Parameters \ Analysis*	WT vs. SN5K	WT vs. SN8K	WT vs. SN10K	WT vs. SN15K	WT vs. SN16K	WT vs. SN5K	WT vs. SN8K	WT vs. SN10K	WT vs. SN15K	WT vs. SN16K
Light	$6.17x10^{-19}$	$3.52x10^{-17}$	$2.84x10^{-16}$	$8.30x10^{-17}$	$3.94x10^{-17}$	$3.44x10^{-15}$	$1.41x10^{-14}$	$3.57x10^{-12}$	$7.08x10^{-14}$	$1.23x10^{-13}$
Temperature	0.2208	0.0709	0.0284	0.0086	0.0044	-	-	-	-	-
Glucose	$9.44x10^{-21}$	$6.61x10^{-23}$	$7.97x10^{-21}$	$3.68x10^{-21}$	$1.44x10^{-20}$	$5.96x10^{-10}$	$9.15x10^{-13}$	$1.79x10^{-10}$	$2.71x10^{-10}$	$2.19x10^{-11}$
Mutation**	0.0034	0.8412	0.0473	0.4647	0.4582	$1.99x10^{-09}$	0.0004	0.0367	0.2777	0.2196
Light + Temperature	$5.48x10^{-07}$	$2.29x10^{-05}$	$4.97x10^{-06}$	$2.23x10^{-07}$	$3.64x10^{-06}$	-	-	-	-	-
Light + Glucose	0.0007	0.0387	0.0026	0.0001	0.0043	$6.93x10^{-05}$	$9.70x10^{-07}$	$1.59x10^{-07}$	$7.67x10^{-10}$	$4.51x10^{-06}$
Light + Mutation	0.0221	0.9084	0.2704	0.6799	0.0393	0.0015	0.0033	0.1832	0.5288	0.0252
Temperature + Glucose	0.0079	0.1660	0.0422	0.0200	0.2948	-	-	-	-	-
Temperature + Mutation	0.1771	0.4462	0.7520	0.7511	0.4342	-	-	-	-	-
Glucose + Mutation	0.3445	0.0082	0.4825	0.3051	0.0737	$7.53x10^{-06}$	0.0459	0.1993	0.0006	0.7635

*Analysis performed comparing *Synechocystis* wild-type (WT) to the SN*n*K mutants (*n*=5, 8, 10, 15, 16).
**The statistical analysis of the effect produced by each mutations introduced is highlighted in grey.

CHAPTER IV

Design of parts for synthetic biology applications


sequences
proteins
transcription
GFP
hydrogenases
promoter
protein
DNA
residues
designed
synthetic
design
region
maturation
operator
linker
promoters
Synechocystis
BioBrick
BBa
sites
fusion
Therefore
Supplementary
ferredoxin
HydA2
vector
standard
performed
expression
GenBank
parts
synthesized
ZFAs
structure
hydrogenase
Swiss-Pdb
HydA2-Fd
coli
FeFe
backbone
Viewer
restriction
devices
available
amino
acids
codon
model
activity
flanked
optimization
downstream
active
based
two
energy
Chapter
core
several
unique
tested
prefix
cloned
shown
binding
containing
complex
template
ArchPRED
peptides
detection
Ppsba2
four
linker
part
specific
region
target
loop
assembly
construction
genome
triplet
PetF1
material
possible
analyzed
TFs
new
low
III
helix
site
GGGGS
orthogonal
regulatory
PDB
double
shows
software
interface
respective
putative
characterized
three-dimensional
E0240
AHL
RFC
libraries
genes
elements
corresponding
presence
included
constructed
vectors
also
HydA2
ZFA
HydA1Fd
see
zinc
suffix
fluorescence
operators
changed
models
gene
different
Viewer
glycine
structures
constitutive
following
functional
using
domains
ZF
assessed
final
since
reinhardtii
cloning
characterization
repression
hydrogen
template
codon
acids
BioBrick
BBa
model
complex
containing
binding
cloned
shown
unique
tested
prefix
GenBank
site
sites
synthesized
parts
DNA
III
downstream
active
based
PetF1
material
possible
analyzed
TFs
new
low
elements
triplet
genes
genome
core
structure
hydrogenase
AHL
RFC
libraries
two
energy
assembly
construction
Chapter
several
respective
putative
interface
target
loop
presence
corresponding
standard
also
vector
vectors
included
constructed
HydA2
ZFA
HydA1Fd
see
zinc
suffix
fluorescence
operators
changed
models
gene
different
performed
expression
Swiss-Pdb
HydA2-Fd
coli
FeFe
backbone

Design of **p**arts for **s**ynthetic **b**iology **a**pplications

Synthetic biology applications require the construction of functional devices/circuits with predictable behavior. Therefore, the design and construction of DNA parts such as promoters, RBSs and terminators, as well as their characterization on the chassis, are important requirements to their successful implementation. Nonetheless, the number of available parts characterized in cyanobacteria is scarce (Huang *et al.*, 2010), and an effort should be done towards the design and characterization of parts in these organisms.

This chapter illustrates the design and construction of new synthetic parts, some that were already used in Chapter III and others that were developed and will be characterized and used in the future.

IV.1 Design of a BioBrick-compatible cloning interface

BioBrick standards are on the basis of DNA parts modularity, allowing the assembly of synthetic devices. This standardization is achieved by constructing each DNA part flanked by a prefix and a suffix containing unique and compatible restriction sites. Therefore, to use the integrative vectors constructed (Chapter III) as platforms for the integration of synthetic devices into *Synechocystis*'s chromosome, a robust BioBrick-compatible interface was designed and cloned into every vector (Figure 9).

The plasmid backbone used to construct the integrative vectors (pGEM-T easy vector), and some of the homologous recombination sequences used, had several restriction sites incompatible with the BioBrick RFC[10] standard. This was overcome by eliminating the unique sites through digestion and re-circularization of the vectors and by site directed mutagenesis, when the sites were within the homologous regions. The presence of two *Eco*RI sites in pGEM-T easy was circumvented by the use of the *Mun*I restriction site that produces overhangs compatible with those of *Eco*RI. Thus, the multiple cloning site comprised the *Mun*I, *Not*I, *Xba*I, *Spe*I, *Not*I and *Pst*I restriction sites and was flanked by four transcription terminators (BioBricks BBa_B1006, BBa_B0062, BBa_B0053 and BBa_B0020). The restriction sites *Xma*I and *Age*I were included at each sequence terminus for cloning purposes.

It was expected that the flanking terminators would prevent any potential promoter activity from backbone DNA sequences, and the proper insulation of the cloning interface was tested and confirmed in Chapter III. The promoterless GFP generator (BioBrick composite

part BBa_E0240) was cloned into the integrative vectors and mutants of *Synechocystis* were produced. GFP expression was analyzed by Western blot and confocal microscopy and the absence of both GFP hybridization signal and green fluorescence indicates that the interface is indeed robust (see Chapter III, Figures 6 and 7).

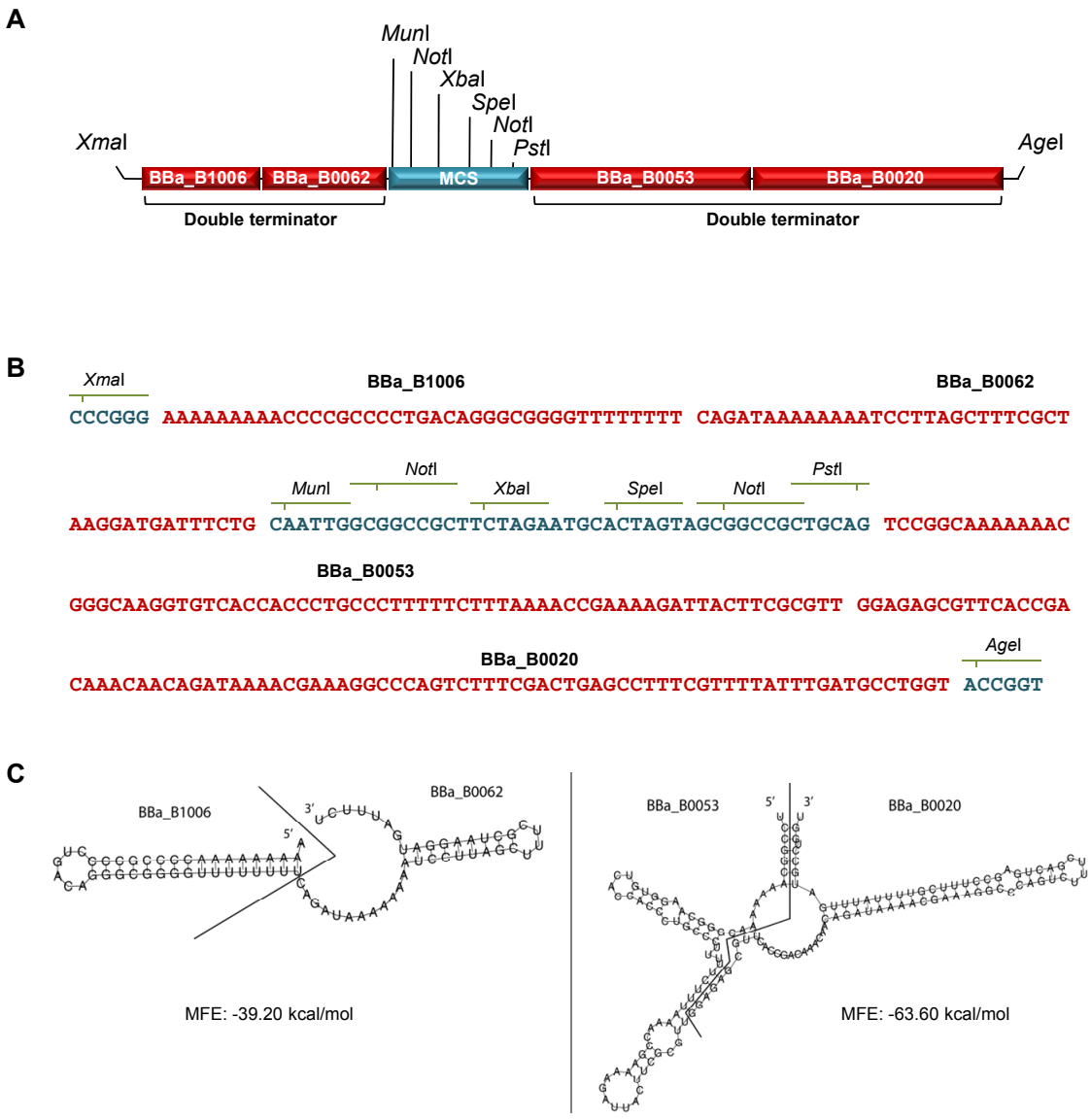


Figure 9 | Synthetic BioBrick-compatible cloning interface. Schematic representation (A) and sequence features (B) of the synthetic interface with the multiple cloning site (MCS) flanked by two double transcription terminators (BioBrick parts BBa_B1006, BBa_B0062, BBa_B0053 and BBa_B0020). The RNA fold of the two double terminators (C) was predicted using the RNAfold server (http://rna.tbi.univie.ac.at/). MFE – minimum free energy.

IV.2 Design of a synthetic constitutive promoter

The majority of the DNA building blocks available at the Registry of Standard Biological Parts (http://partsregistry.org) derive from *Escherichia coli*, and they have not been tested in cyanobacteria. Although in some cases orthology is desirable, it can be problematic for instance concerning promoter sequences. RNA polymerase may not recognize orthologous promoters, since σ-factors may be different between species, thus compromising the assembly of a functional transcriptional unit.

To overcome this issue, a synthetic constitutive promoter was designed based on the light-responsive promoter of *Synechocystis psbA2* gene. It has been shown that there is a minimal sequence capable of promoting the light-independent gene transcription in *Microcystis aeruginosa* K-81 (Agrawal *et al.*, 2001; Shibato *et al.*, 2002). By sequence comparison, the *Synechocystis psbA2* promoter sequence corresponding to that region was synthesized from bases -38 to +14 (P_{psbA2*}; +1 corresponds to the transcription start site), flanked by the BioBrick standard RFC[10] prefix and suffix (Figure 10). This minimal promoter region does not contain the AU-box motif associated to mRNA instability in the dark (Agrawal *et al.*, 2001), nor the putative *cis* elements involved in photoregulation (Eriksson *et al.*, 2000).

Therefore, this region should be sufficient to constitutively transcribe downstream genes, and the potential promoter activity of P_{psbA2*} was tested and confirmed in Chapter III. The GFP generator (BioBrick part BBa_E0240) was cloned downstream of P_{psbA2*} and GFP expression was analyzed by Western blot and confocal microscopy (see Chapter III, Figures 6 and 7). The detection of both GFP signal and green fluorescence indicates that P_{psbA2*} is functional in *Synechocystis*.

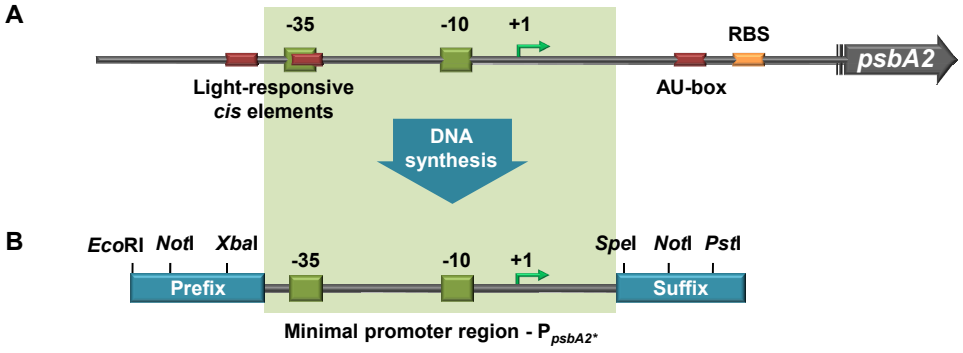


Figure 10 | Design of a synthetic constitutive promoter for *Synechocystis*. The native DNA region upstream of *Synechocystis psbA2* is depicted, showing the putative light-responsive *cis* regulatory elements and the AU-box motif (red), the RNA polymerase recognition elements (green, -35 and -10), the ribosome binding site (RBS, orange) and the start of the *psbA2* gene (green arrow, +1) (A). The minimal promoter region (bases -38 to +14) was adapted to the BioBrick RFC[10] standard by the incorporation of a prefix and a suffix (B). The synthetic promoter sequence is available at Supplementary material III.

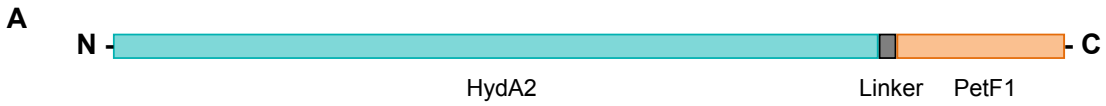
IV.3 Design of a synthetic hydrogen producing device (HPD)

Hydrogenases are metalloproteins capable of catalyzing the reversible oxidation of dihydrogen (H_2), a promising future energy vector, as an alternative to the limited fossil fuel resources. They constitute a heterogeneous group of enzymes with complex metallic active sites ([NiFe], [FeFe] and [Fe]) and complex maturation pathways (Casalot and Rousset, 2001; Leitão *et al.*, 2006; Forzi and Sawers, 2007; Tamagnini *et al.*, 2007). [FeFe] hydrogenases differ from the others by their unique prosthetic group (H-cluster) in the active centre, which results in higher enzyme activity compared to the other type of hydrogenases (Adams, 1990; Happe *et al.*, 2002). Moreover, while the maturation of [NiFe] hydrogenases depends on the activity of at least seven core proteins, the maturation of [FeFe] hydrogenases is simpler, requiring three core proteins (Mulder *et al.*, 2011). The reduced [2Fe-2S] ferredoxin (Fd), which mediates electron transfer from the photosynthetic electron transport chain to the reductive cellular metabolism, is the primary electron donor for proton reduction by the [FeFe] hydrogenases (Roessler and Lien, 1984; Vignais *et al.*, 2001; Long *et al.*, 2009).

To design and produce an effective synthetic hydrogen producing device (HPD) three [FeFe] hydrogenases were taken into consideration: HydA1 and HydA2 from *Chlamydomonas reinhardtii*, and HydA from *Clostridium acetobutylicum* ATCC 824. Hydrogen production by these organisms and their respective hydrogenases have been extensively studied (Roessler and Lien, 1984; Happe and Naber, 1993; Gorwa *et al.*, 1996; Podestá *et al.*, 1996; Forestier *et al.*, 2003; Girbal *et al.*, 2005; Demuez *et al.*, 2007; Oh *et al.*, 2009; Meuser *et al.*, 2012). Moreover, to improve electron transfer between the ferredoxin and the hydrogenase, fusion proteins were designed, in order to link the hydrogenases to their ferredoxin partners. Three-dimensional models of the fusion proteins from *C. reinhardtii* were also generated, based on the models of HydA2 and PetF1 ([FeFe] hydrogenase:ferredoxin) proposed by Chang *et al*. (2007). From the two models proposed in that work, complex 16 was used as template structure, since it better reflects a more relevant functional conformation (Chang *et al.*, 2007). Structural analysis of HydA2:Fd complex shows that a loop of at least 7 residues is required to join the C-terminus of HydA2 to the N-terminus of the PetF1. It has been shown that glycine-rich peptides confer flexibility allowing the specific engineering of hinge regions into proteins (Wriggers *et al.*, 2005), and such soft glycine hinges can be designed as stretches of amino acids where at least four of every six residues are glycines (Mancebo *et al.*, 1990). Therefore, the 10 residues sequence [$(GGGGS)_2$] was used to link the two proteins.

Both the hydrogenases and the ferredoxin from *C. reinhardtii* have N-terminal transit peptides that route these proteins into the chloroplast stroma (Chang *et al.*, 2007). These

peptides are cleaved by a protease and are not part of the mature hydrogenase. Since *Synechocystis* is a prokaryote, the transit peptides were not considered for the design of the fusion proteins, which contained only residues 63 to 498 from HydA2 (GenBank: AAL23573.1), the ten residues linker and residues 31 to 126 from PetF1 (GenBank: AAC49171.1), resulting in a 544 amino acids protein (HydA2:Fd) (Figure 11). To improve loop flexibility, PetF1 starting methionine was changed to a glycine (M447G) and this sequence was initially submitted to SWISS-MODEL (Guex and Peitsch, 1997) to retrieve a three-dimensional model, subsequently used to predict the linker structure by ArchPRED (Fernandez-Fuentes *et al.*, 2006).



B

ATATDAVPHWKLALEELDKPKDGGRKVLIAQVAPAVRVAIAESFGLAPGAVSPGKLATGLRALGFDQV
FDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMFTSCCPGWVAMMEKSYPELIPFVSSCKSPQ
MMMGAMVKTYLSEKQGIPAKDIVMVSVMPCVRKQGEADREWFCVSEPGVRDVDHVITTAELGNIFKER
GINLPELPDSDWDQPLGLGSGAGVLFGTTGGVMEAALRTAYEIVTKEPLPRLNLSEVRGLDGIKEASV
TLVPAPGSKFAELVAERLAHKVEEAAAAEAAAAVEGAVKPPIAYDGGQGFSTDDGKGGLKLRVAVANG
LGNAKKLIGKMVSGEAKYDFVEIMACPAGCVGGGGQPRSTDKQITQKRQAALYDLDERNTLRRSHENE
AVNQLYKEFLGEPLSHRAHELLHTHYVP<u>GGGGSGGGGS</u>MAYKVTLKTPSGDKTIECPADTYILDAAEE
AGLDLPYSCRAGACSSCAGKVAAGTVDQSDQSFLDDAQMGNGFVLTCVAYPTSDCTIQTHQEEALY

Figure 11 | HydA2:Fd fusion protein. Schematic representation (A) and protein sequence features (B) showing the HydA2 hydrogenase (blue) and the PetF1 ferredoxin (orange) linked by a ten amino acids glycine-rich polypeptide (black, underlined).

Initially, four candidate structures were used by ArchPRED to model the loop (Supplementary figure 1), and the analysis of the models with Swiss-Pdb Viewer (Guex and Peitsch, 1997) revealed an abnormal rotation of the linker backbone, which was caused by proline 436. This residue was therefore changed to a glycine (P436G) and remodeled with ArchPRED. Only one candidate template structure was retrieved to model the new linker [G $(GGGGS)_2$G] (PDB ID:1rp4, Chain A, amino acids 181 to 192; Supplementary figure 2), and the structural analysis of the new model did not show an abnormal rotation of the linker backbone. Finally, the initial adenine was changed to a methionine (A1M), setting the fusion protein start, and remodeled with ArchPRED. The result obtained was similar to the previous one (Supplementary figure 3).

The tree-dimensional model retrieved from ArchPRED was analyzed with the Swiss-Pdb Viewer v4.0 and the *Clostridium pasteurianum* Fe-only hydrogenase structure (CpI; PDB ID:1FHE) was used as template for optimization. Sequence alignment and structure superimposition shows that HydA2:Fd has an additional sequence (residues 276 to 329) that exhibits an abnormal torsion of the backbone (Figure 12A). This region was optimized

using the Swiss-Pdb Viewer loop database (Figure 12B) and several steps of energy minimization were applied to the whole structure. The total energy computed after optimization was -15 880.921 kJ/mol. Energy computations were performed with GROMOS96 implementation of Swiss-Pdb viewer, *in vacuo* using the GROMOS96 43B1 parameters set, and without reaction field.

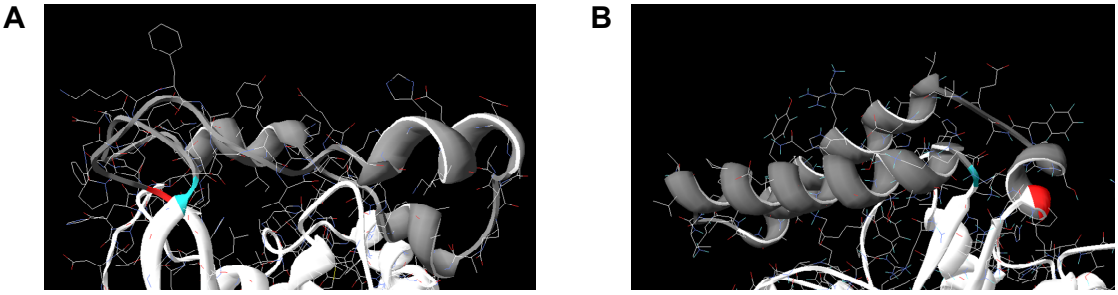


Figure 12 | Loop optimization of HydA2:Fd. An abnormal backbone torsion in the sequence between amino acids P276 (red) and G329 (blue) was detected in the initial model (A). Using the Swiss-Pdb viewer loop database this region was remodeled and the torsion corrected (B).

Similarly, a second model of the fusion protein HydA1:Fd was constructed based on HydA2:Fd and CpI, using the Swiss-Pdb Viewer software. The fusion protein contained the residues 56 to 487 from HydA1 (GenBank: AAL23572.1), the linker [$(GGGGS)_2$] and residues 31 to 126 from PetF1 (GenBank: AAC49171.1), resulting in a 540 amino acids protein (HydA1:Fd). Again, the following mutations were performed: A1M, A432G and M443G. An extra region (between residues P278 and G322) did not align with CpI and had a different sequence size comparing to the same region of HydA2. This region was also remodeled using the Swiss-Pdb Viewer loop database. Several steps of energy minimization were applied to the whole structure and the total energy computed after optimization was -16 343.063 kJ/mol.

The three-dimensional structures of both fusion proteins HydA1:Fd and HydA2:Fd are depicted in Figure 13 (PDB files for HydA1:Fd and HydA2:Fd are available available at the supplementary material in the CD-ROM attached). The superimposition of both structures shows a high degree of similarity, and the main difference resides in the region corresponding to the sequence stretch with low homology (see above). These models constitute important tools to address activity improvement by protein engineering.

The additional fusion protein was designed based on the sequences of HydA (GenBank: AAK78015.1) and ferredoxin (GenBank: AAK78284.1) from *C. acetobutylicum*. Several linkers were introduced and sequence alignments were performed with the previous designed sequences. The best linker was the same 10 residues long [$(GGGGS)_2$], and the

best alignment was achieved if the last 9 residues from the hydrogenase C-terminus were deleted. The length of the final fusion protein was 632 amino acids.

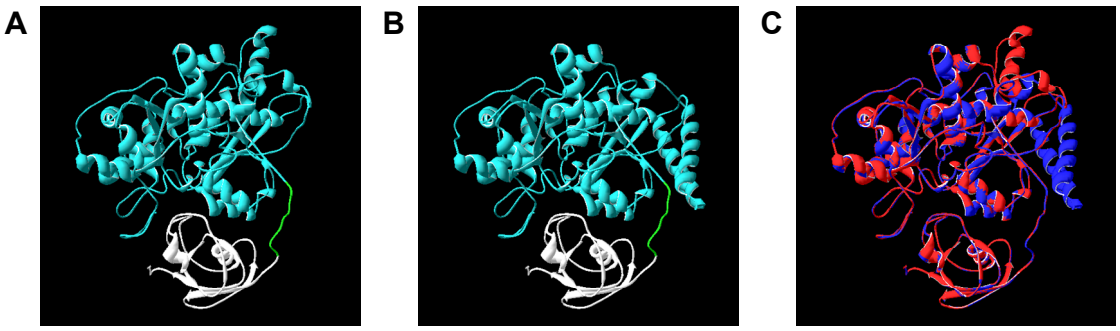


Figure 13 | Three-dimensional models of the fusion proteins. The fusion protein models for HydA1:Fd (A) and HydA2:Fd (B) are depicted. The hydrogenase domains are depicted in light-blue, the ferredoxin domain is in white and the linker is in green. The high structural similarity between the HydA1:Fd (Blue) and HydA2:Fd (red) model is shown by structure superimposition (C).

As stated above, maturation is a key issue to achieve mature and functional hydrogenases. Therefore, maturation devices were also designed to express the hydrogenase maturation proteins identified in *C. reinhardtii* (HydEF and HydG) and *C. acetobutylicum* (HydE, HydF and HydG). It is expected that the combination of the maturation proteins with the respective fusion proteins triggers the assembly of active hydrogenases.

All the protein sequences (see Supplementary material III) were back translated using the GeneDesigner v1.1.4.1 software (DNA 2.0 Inc.), and codon optimization was performed following a codon usage table in which the frequency is the mean of *E. coli* K12 and *Synechocystis* codon frequencies. The software makes random choices for each codon used but with preference for the most frequent, avoiding new sequences that greatly differ from the G+C content of the hosts, and preventing the potential depletion of high usage tRNAs. In addition, an extra stop codon was included in all coding sequences, following the Registry regulations for Protein-Coding Parts (http://partsregistry.org). The synthetic modules were designed following the BioBrick guidelines and the standard prefix and a suffix were included in all the final sequences. Moreover, a ribosome binding site (BBa_B0034) was added upstream of each coding sequence, and a double terminator (BBa_B0015) was added downstream of each transcriptional unit (see Figure 14). To be able to test the heterologous hydrogenases without the fused ferredoxin domain, a *Nhe*I restriction site was included in the linker sequence, changing the protein sequence to [GGGASGGGGS], and repeated just before the stop codons (Figure 14). This way, cutting the final plasmid with *Nhe*I will eliminate the ferredoxin part and after re-ligation, the site

will be restored and the hydrogenase gene will be in frame with the stop codons. This substitution of a glycine by an alanine in the linker does not affect the three-dimensional structures.

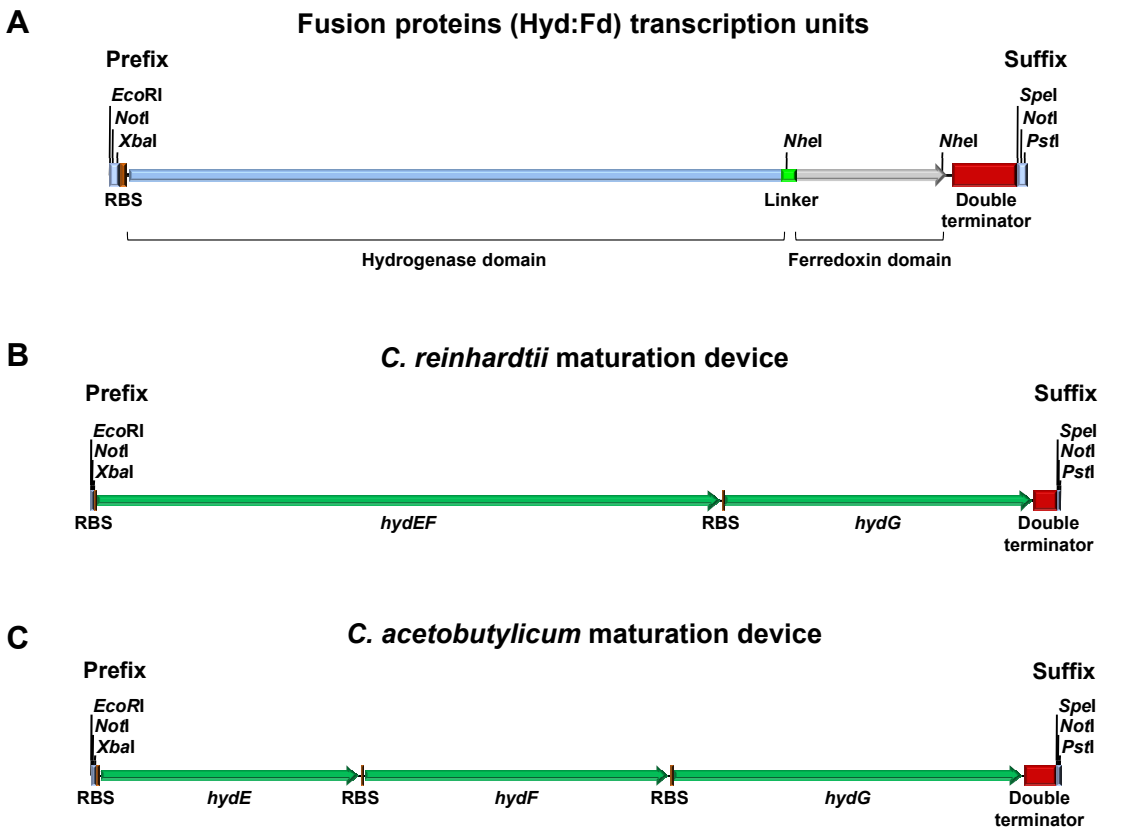


Figure 14 | Schematic representation of the hydrogenase-related synthetic devices. The sequence features of the fusion proteins from *C. reinhardtii* and from *C. acetobutylicum* are depicted in (A), the maturation device from *C. reinhardtii* is depicted in (B) and the maturation device from *C. acetobutylicum* is depicted in (C). DNA sequence lengths for the fusion proteins (Hyd:Fd) transcription units are: 1824 bp for HydA1:Fd and 1836 bp for HydA2:Fd from *C. reinhardtii*; and 2124 bp for HydA:Fd from *C. acetobutylicum*. DNA sequence lengths for the maturation devices are 5390 bp from *C. reinhardtii* and 3967 bp from *C. acetobutylicum*. The DNA sequences are available at Supplementary material III.

The designed hydrogen producing and maturation devices were synthesized and were tested at the laboratory of Professor Peter Lindblad (Uppsala University, Sweden). Preliminary results indicate that these devices are functional in *E. coli* BL21: proteins are expressed and hydrogenases, co-expressed with respective maturation proteins, are active. The system is being implemented in *Synechocystis* wild-type and in a *Synechocystis* mutant lacking an active native bidirectional hydrogenase (Pinto *et al.*, 2012).

IV.4 Development of orthogonal regulation networks

Gene transcription can be modulated by the use of regulatory proteins (transcription factors – TFs) that bind specific DNA sequences (operators), allowing the fine tuning of genetic circuits in order to meet design specifications. These TFs can either activate or repress the transcription of target genes, and their full characterization improves systems predictability.

When designing synthetic circuits, it should be taken into consideration that the regulatory networks introduced in the system/chassis must, preferably, be orthogonal. This means that the transcription regulation of the synthetic devices shall not interfere with the endogenous gene expression, and vice-versa. Therefore, in this work, a library of promoters and their associated TFs were initially designed to be orthogonal to *E. coli*, but later on they could be used or redesigned to work in other organisms, namely *Synechocystis*.

The design of the regulatory elements (TFs) was based on the eukaryotic Cys_2His_2 zinc finger protein (ZFP) regulatory domains. ZFPs constitute a large superfamily of DNA binding domains and each zinc finger (ZF) binds a DNA sequence of 3 nucleotides (Pavletich and Pabo, 1991; Klug, 2010). ZF motifs consist of two beta sheets and an alpha helix, stabilized by coordination of a zinc ion by pairs of conserved cysteine and histidine residues (Figure 15A). ZFs bind target DNA sites with amino acids of the recognition alpha helix contacting consecutive nucleotides in the 3' to 5' direction (Fu *et al.*, 2009): residues -1, +3 and +6 of the alpha-helix (numbered relative to the start of the helix) contact nucleotides at the 3′-end, middle and 5′-end of the DNA triplet, respectively (Figure 15B). Since each ZF is specific for a triplet, longer target DNA sequences can be recognized joining several of these domains (Figure 15C). The resulting multi-finger array recognizes a specific and preferably unique sequence in double-stranded DNA and,

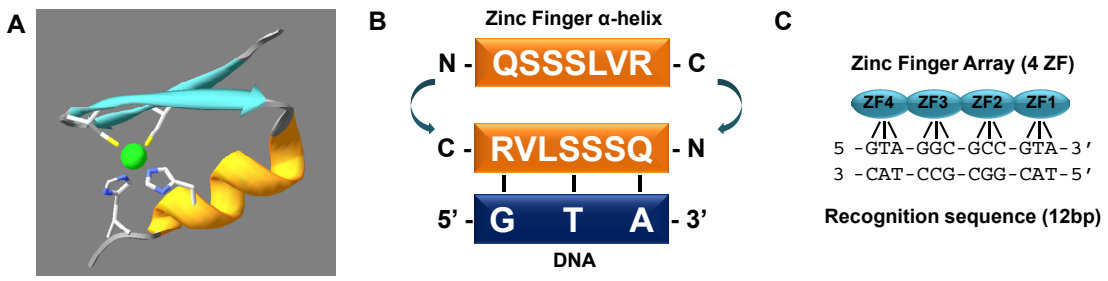


Figure 15 | Zinc finger domains (ZFs) recognize DNA sequences of 3 nucleotides. Cartoon representation of a zinc finger motif of the Krueppel-like factor 4 (Klf4, PDB ID: 2WBS), consisting of two antiparallel beta sheets (blue) and an alpha helix (orange), coordinating a zinc ion (green) by two cysteine and two histidine residues (Cys_2His_2) (A). Residues of the ZF alpha helix recognize specific triplet DNA sequences in the inverse orientation (3' to 5') (B). Arrays of several ZF domains specifically recognize longer DNA sequences (C).

thereby, target transcriptional activation domains, repressors, or nucleases to specific locations in the genome. Therefore, DNA-binding specificities of ZF domains can be re-engineered and can theoretically be constructed to target nearly any nucleotide sequence (Wright *et al.*, 2006). Such modular proteins can be design *in silico* (www.zincfingers.org), and constructed using modular assembly ZF kits (available from Addgene), or they can be synthesized *de novo*. It has been previously shown that zinc fingers arrays (ZFA) can be expressed in *E. coli* and specifically bind to target DNA sequences (Nagai *et al.*, 1988; Joung, 2001; Park *et al.*, 2005; Giesecke and Joung, 2007).

The ZF-based transcription factors were design to act as repressors by binding a specific 12 bp sequence (operator) placed between the -35 and -10 promoter elements. Recognition of the promoter by the RNA polymerase holoenzyme would be blocked by steric hindrance caused by the bound repressor (Browning and Busby, 2004). In order to make the system orthogonal to the chassis, only sequences not present in the genome of *E. coli* were chosen as putative operators. Sequences of the form GNN-GNN-GNN-GNN present higher recognition specificity by ZFAs (Dreier *et al.*, 2000; Liu *et al.*, 2002), and all possible sequence combinations were generated (4^8 sequences). *E. coli*"s genome was screened for the presence of these sequences and all possible substitution, deletion and insertion mutations on the sequences were also checked, keeping only those that were not found in the genome. Moreover, sequences with GGGG motifs and DNA methylation sites were avoided. From the initial 65536 possible 12 bp operators, only 17 sequences remained and alignments were performed between them to guarantee othogonality. ZFAs were then designed by finding the modules for each triplet of four operator sequences (Figure 16A), using the ZiFiT web-based tool (accessed in 2008) (Sander *et al.*, 2007; Sander *et al.*, 2010). The human transcription factor Sp1 was used as protein backbone to assemble four-finger arrays (Zhang *et al.*, 2000; Liu *et al.*, 2001; Liu *et al.*, 2002), consisting of two two-finger DNA binding domains (2x2), joined by linkers of different sizes (Moore *et al.*, 2001) (Figure 16B). Codon optimization was performed for ZFAs expression in *E. coli* at Epoch Life Science Inc., avoiding the BioBrick restriction sites (*Eco*RI, *Xba*I, *Spe*I, *Pst*I and *Not*I), DNA methylation sites, secondary structures and repetitive sequences. The ZFAs were synthesized with an upstream RBS sequence (BBa_B0032) and a downstream double transcription terminator (BBa_B0015), and the sequence was flanked by the standard RFC[10] prefix and suffix (sequences available at Supplementary material III). The low to medium copy BioBrick standard vector pSB3C5 (chloramphenicol resistance marker) containing the part BBa_F2620 was used to clone the ZFAs coding sequences. Genes cloned downstream of BBa_F2620 are regulated by the transcription factor LuxR, that is active in the presence of cell-cell signaling molecules (namely acyl-

homoserine lactones – AHL). This way, in the presence of AHL, LuxR will promote ZFAs transcription.

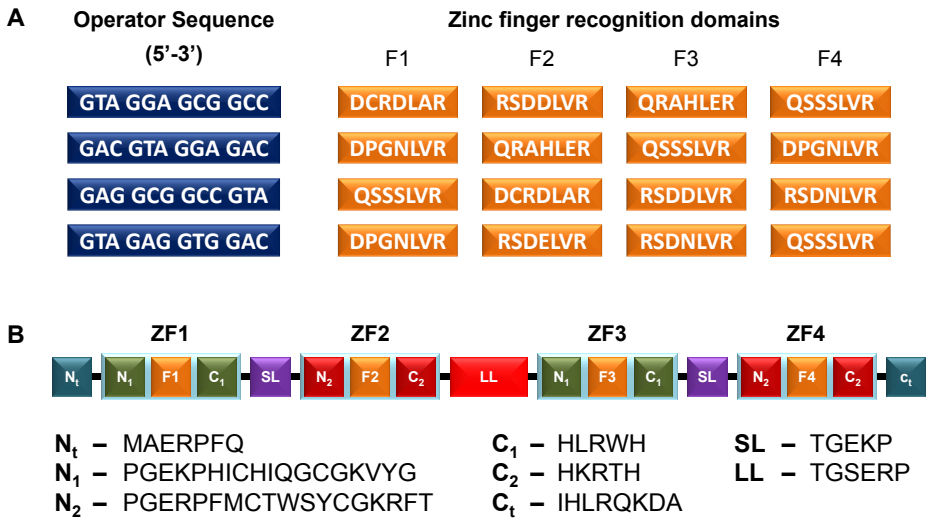


Figure 16 | Design of zinc finger arrays recognizing 12 bp operators orthogonal to *E. coli*. Four operator sequences (left) and their corresponding zinc fingers recognition domains (right) are depicted (A). Schematic representation of the modular design of a four-finger array showing the external N- and C-terminal sequences (N_t and C_t), the N- and C- terminal sequences of each zinc finger (N_n and C_n), the zinc finger recognition domains and the short (SL) and long (LL) linkers (B).

To design the promoter libraries, three *E. coli* promoters were taken into consideration: the consensus promoter (P_{cons}) (Shultzaberger *et al.*, 2007), P_{tetR}, and P_{lacI} (Cox *et al.*, 2007). The first promoter is constitutive and the last two are known to be operational in the absence of their respective repressors. Therefore, libraries were designed by placing a putative 12 bp operator sequence at different positions between the -35 and -10 promoters elements (Figure 17) (Cox *et al.*, 2007). These combinatorial promoters (different promoters with combinations of different operators) will serve to determine the operator position that exhibits higher repression levels by the corresponding ZFAs. For this initial analysis, only four operator sequences were considered (see below).

The promoter libraries sequences were synthesized flanked by the standard RFC[10] prefix and suffix, and the DNA sequences are available at Supplementary material III. In this case, the low to medium copy BioBrick standard vector pSB3K3 (kanamycin resistance marker), containing the part BBa_E0240, was used to clone the promoters. The part BBa_E0240 comprises a GFP-encoding gene flanked by an RBS (upstream), and a double transcription terminator (downstream). *gfp* transcription should be constitutive with all promoters from the libraries.

Future characterization must be performed to assess the capacity of a ZFA to bind the corresponding operator sequence within the core promoter region and repress *gfp*

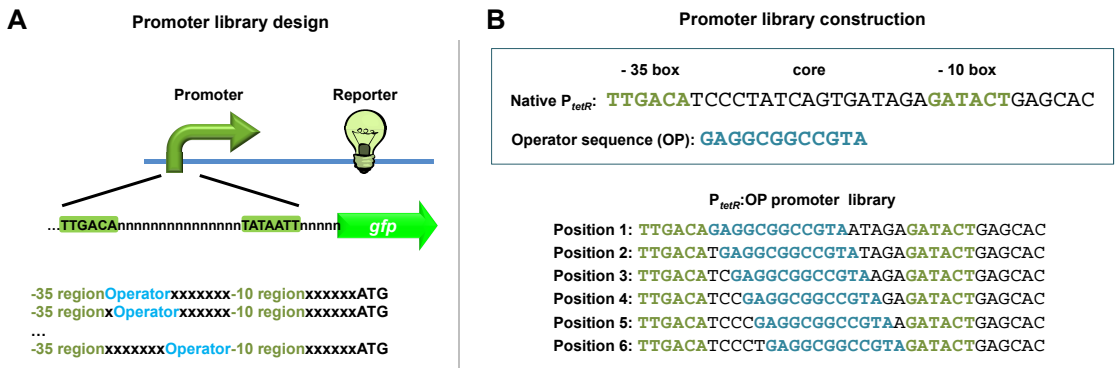


Figure 17 | Design and construction synthetic promoter libraries. The promoter libraries were designed by placing operator sequences (OP) at different positions between the -35 and -10 boxes of different promoters, controlling the transcription of the reporter *gfp* gene (A). Construction of a library based on the native sequence of the *tetR* promoter, consisting of six sequences with different operator positions within the promoter core sequence (B).

transcription. To fulfill this purpose each of the promoters containing an operator sequence will be transformed in an *E. coli* strain carrying the vector expressing the counterpart ZFAs proteins. Transcription repression by the ZFAs can be assessed by complementing the growth medium with AHL and measuring GFP fluorescence intensity. In the presence of the AHL, the transcription factor LuxR is active and the ZFA is transcribed. If the ZFA is able to bind the operator sequence within the core promoter region, *gfp* transcription will be impaired and a decrease in GFP fluorescence will occur. The best ZFA repression at a determined operator position at the core promoter region corresponds to the lowest fluorescence intensity (Figure 18). This will allow obtaining a library of characterized promoters (and respective TFs) with varying strengths.

All the DNA sequences were synthesized and assembled at Epoch Life Science Inc. and are being characterized at the laboratory of Dr. Alfonso Jaramillo (Université d'Évry Val d'Essonne, France).

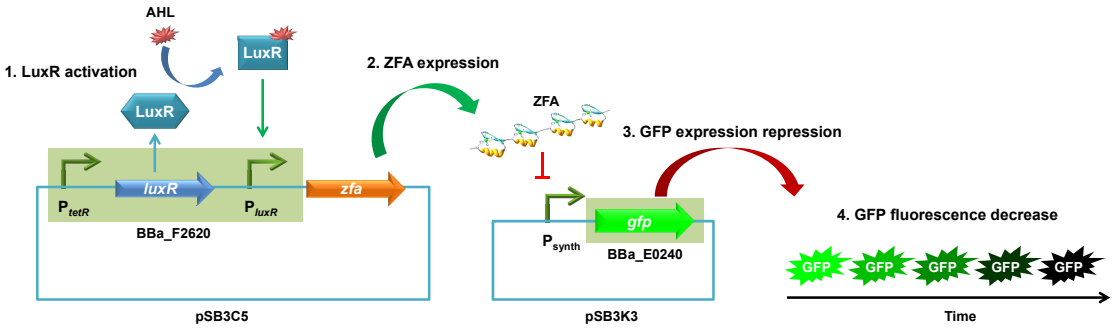
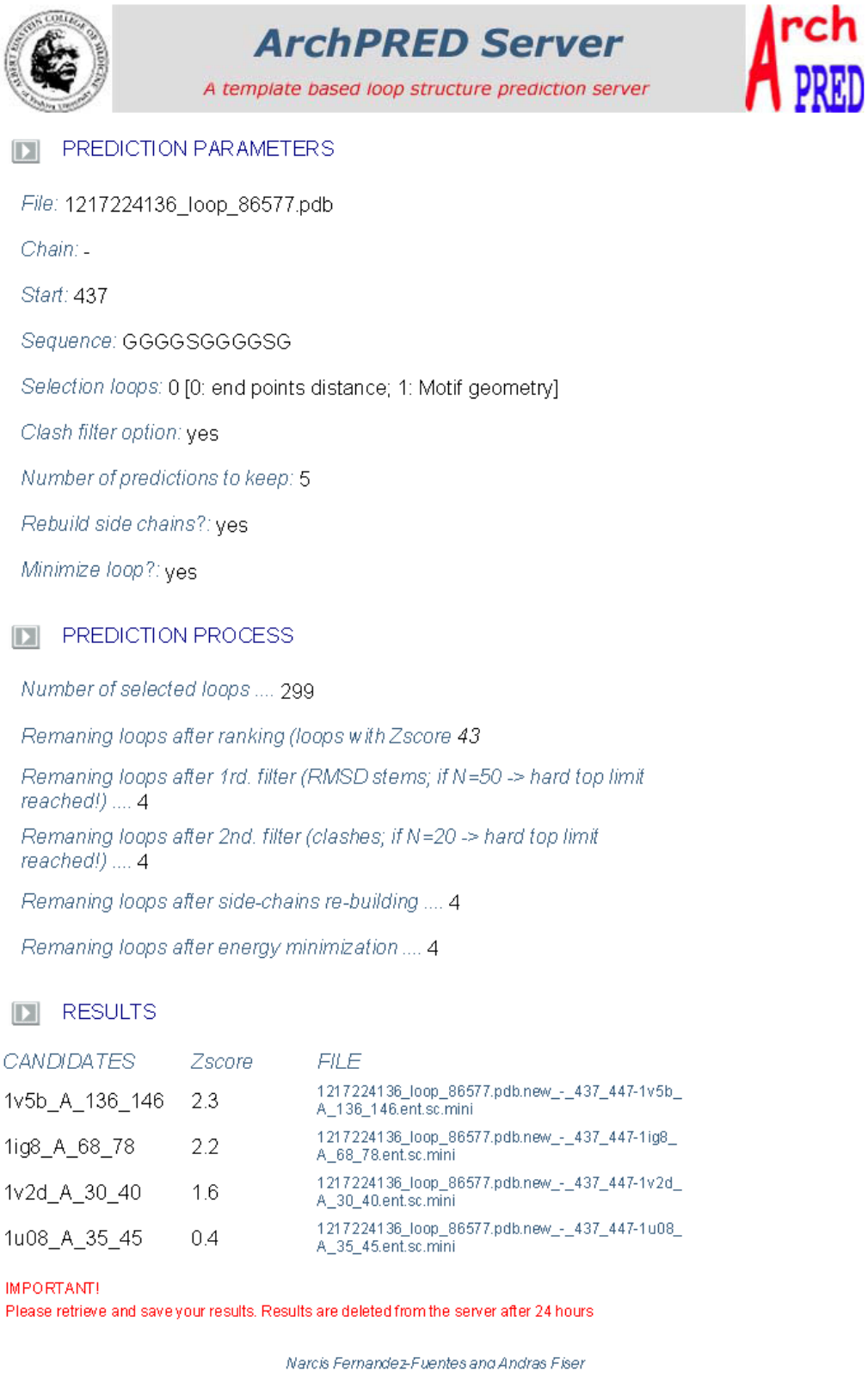


Figure 18 | Characterization of the synthetic promoter libraries and ZFA counterparts. LuxR is expressed from BBa_F2620 and activated by AHL (1). Active LuxR promotes the transcription of the ZFA (2), which will interact with the synthetic promoters (P_{synth}) and impair *gfp* transcription from BBa_E0240 (3). The effective transcription repression leads to the decrease o GFP fluorescence signal over time (4).

Supplementary material III

Supplementary figures



ArchPRED Server

A template based loop structure prediction server

PREDICTION PARAMETERS

File: 1217224136_loop_86577.pdb

Chain: -

Start: 437

Sequence: GGGGSGGGGSG

Selection loops: 0 [0: end points distance; 1: Motif geometry]

Clash filter option: yes

Number of predictions to keep: 5

Rebuild side chains?: yes

Minimize loop?: yes

PREDICTION PROCESS

Number of selected loops 299

Remaning loops after ranking (loops with Zscore 43

Remaning loops after 1rd. filter (RMSD stems; if N=50 -> hard top limit reached!) 4

Remaning loops after 2nd. filter (clashes; if N=20 -> hard top limit reached!) 4

Remaning loops after side-chains re-building 4

Remaning loops after energy minimization 4

RESULTS

CANDIDATES	Zscore	FILE
1v5b_A_136_146	2.3	1217224136_loop_86577.pdb.new_-_437_447-1v5b_A_136_146.ent.sc.mini
1ig8_A_68_78	2.2	1217224136_loop_86577.pdb.new_-_437_447-1ig8_A_68_78.ent.sc.mini
1v2d_A_30_40	1.6	1217224136_loop_86577.pdb.new_-_437_447-1v2d_A_30_40.ent.sc.mini
1u08_A_35_45	0.4	1217224136_loop_86577.pdb.new_-_437_447-1u08_A_35_45.ent.sc.mini

IMPORTANT!

Please retrieve and save your results. Results are deleted from the server after 24 hours

Narcis Fernandez-Fuentes and Andras Fiser

Supplementary figure 1 | ArchPRED output 1



ArchPRED Server

A template based loop structure prediction server

PREDICTION PARAMETERS

File: 1217225775_loop_97246.pdb

Chain: -

Start: 436

Sequence: GGGGGSGGGGSG

Selection loops: 0 [0: end points distance; 1: Motif geometry]

Clash filter option: yes

Number of predictions to keep: 5

Rebuild side chains?: yes

Minimize loop?: yes

PREDICTION PROCESS

Number of selected loops 158

Remaning loops after ranking (loops with Zscore 22

Remaning loops after 1rd. filter (RMSD stems; if N=50 -> hard top limit reached!) 5

Remaning loops after 2nd. filter (clashes; if N=20 -> hard top limit reached!) 1

Remaning loops after side-chains re-building 1

Remaning loops after energy minimization 1

RESULTS

CANDIDATES	*Zscore*	*FILE*
1rp4_A_181_192	1.5	1217225775_loop_97246.pdb.new_-_436_447-1rp4_A_181_192.ent.sc.mini

IMPORTANT!
Please retrieve and save your results. Results are deleted from the server after 24 hours

Narcis Fernandez-Fuentes and Andras Fiser
Center of Bioinformatics
Albert Einstein College of Medicine

© Copyright 2005 Fiser Lab, All Rights Reserved

Supplementary figure 2 | ArchPRED output 2



ArchPRED Server

A template based loop structure prediction server

PREDICTION PARAMETERS

File: 1217227268_loop_41517.pdb

Chain: -

Start: 436

Sequence: GGGGGSGGGGSG

Selection loops: 0 [0: end points distance; 1: Motif geometry]

Clash filter option: yes

Number of predictions to keep: 5

Rebuild side chains?: yes

Minimize loop?: yes

PREDICTION PROCESS

Number of selected loops 158

Remaning loops after ranking (loops with Zscore 18

Remaning loops after 1rd. filter (RMSD stems; if N=50 -> hard top limit reached!) 4

Remaning loops after 2nd. filter (clashes; if N=20 -> hard top limit reached!) 1

Remaning loops after side-chains re-building 1

Remaning loops after energy minimization 1

RESULTS

CANDIDATES	*Zscore*	*FILE*
1rp4_A_181_192	1.4	1217227268_loop_41517.pdb.new_-_436_447-1rp4_A_181_192.ent.sc.mini

IMPORTANT!
Please retrieve and save your results. Results are deleted from the server after 24 hours

Narcis Fernandez-Fuentes and Andras Fiser
Center of Bioinformatics
Albert Einstein College of Medicine

© Copyright 2005 Fiser Lab, All Rights Reserved.

Supplementary figure 3 | ArchPRED output 3

Sequences

P_{psbA2*} sequence
AGCTTTACAAAACTCTCATTAATCCTTTAGACTAAGTTTAGTCAGTTCCAATCTG

HydA1_Fd (*Chlamydomonas reinhardtii*)
MAAPAAEAPLSHVQQALAELAKPKDDPTRKHVCVQVAPAVRVAIAETLGLAPGATTPKQLAEGLRRLGFDEVF
DTLFGADLTIMEEGSELLHRLTEHLEAHPHSDEPLPMFTSCCPGWIAMLEKSYPDLIPYVSSCKSPQMMLAAM
VKSYLAEKKGIAPKDMVMVSIMPCTRKQSEADRDWFCVDADPTLRQLDHVITTVELGNIFKERGINLAELPEG
EWDNPMGVGSGAGVLFGTTGGVMEAALRTAYELFTGTPLPRLSLSEVRGMDGIKETNITMVPAPGSKFEELLK
HRAAARAEAAAHGTPGPLAWDGGAGFTSEDGRGGITLRVAVANGLGNAKKLITKMQAGEAKYDFVEIMACPAG
CVGGGGQPRSTDKAITQKRQAALYNLDEKSTLRRSHENPSIRELYDTYLGEPLGHKAHELLHTHYVGGGGASG
GGGSGAYKVTLKTPSGDKTIECPADTYILDAAEEAGLDLPYSCRAGACSSCAGKVAAGTVDQSDQSFLDDAQM
GNGFVLTCVAYPTSDCTIQTHQEEALYAS

hydA1_fd
ATGGCTGCTCCTGCTGCTGAAGCTCCTTTAAGTCATGTACAACAAGCATTAGCAGAGTTAGCTAAACCTAAAG
ATGATCCCACACGCAAGCACGTGTGCGTCCAAGTCGCTCCCGCAGTGCGAGTAGCTATCGCTGAAACATTAGG
TTTGGCTCCCGGAGCCACCACCCCTAAGCAATTAGCCGAAGGTTTACGTCGTCTCGGTTTCGATGAAGTTTTT
GATACCTTATTTGGTGCTGACCTCACTATCATGGAAGAAGGTTCTGAGTTATTACATCGATTGACCGAACATT
TAGAAGCCCATCCCCATAGTGATGAACCTCTCCCCATGTTTACTTCCTGTTGCCCCGGTTGGATCGCTATGCT
CGAAAAATCTTATCCCGACCTCATTCCTTATGTAAGCTCTTGCAAATCTCCCCAAATGATGTTAGCCGCGATG
GTTAAAAGCTATTTGGCCGAGAAGAAGGGCATTGCTCCCAAAGACATGGTGATGGTCAGTATTATGCCCTGTA
CTCGCAAACAAAGTGAAGCTGATCGAGACTGGTTTTGCGTTGATGCTGATCCCACTCTCCGCCAATTGGACCA
TGTAATCACTACCGTAGAACTCGGCAACATCTTTAAAGAACGAGGTATCAATTTAGCAGAATTACCCGAGGGT
GAATGGGATAATCCCATGGGAGTAGGTAGTGGCGCCGGTGTCTTGTTTGGAACCACTGGAGGCGTCATGGAAG
CTGCTTTACGAACAGCTTATGAACTCTTTACTGGAACCCCCTTACCCCGATTATCTTTATCCGAGGTCCGCGG
CATGGATGGCATCAAAGAAACAAATATTACTATGGTCCCTGCGCCTGGCTCCAAATTCGAAGAGTTACTCAAA
CATCGCGCGGCTGCTCGAGCCGAAGCCGCCGCTCATGGCACTCCTGGCCCCCTCGCTTGGGATGGTGGAGCAG
GCTTTACATCCGAAGATGGACGCGGCGGAATTACTTTGCGCGTCGCTGTGGCCAACGGCTTAGGTAATGCTAA
GAAATTGATCACCAAGATGCAAGCGGGTGAAGCCAAATACGATTTCGTCGAGATCATGGCATGTCCTGCCGGT
TGTGTTGGTGGTGGTGGTCAACCCCGAAGTACTGATAAAGCGATCACGCAAAAACGCCAAGCAGCGTTATACA
ATTTAGACGAAAAATCTACGTTGCGTCGCAGTCATGAGAATCCCTCTATCCGAGAATTATATGACACTTATTT
AGGAGAACCCTTAGGTCATAAAGCGCATGAACTCTTACATACTCATTACGTCGGTGGCGGTGGTGCTAGCGGC
GGTGGAGGCAGTGGCGCTTATAAGGTCACTTTGAAAACCCCTTCCGGAGATAAAACGATTGAATGTCCCGCAG
ATACTTATATCTTAGATGCCGCTGAAGAGGCAGGATTAGACTTGCCCTATAGTTGTCGAGCAGGAGCATGCAG
CAGCTGTGCAGGTAAGGTCGCAGCCGGTACCGTAGATCAATCCGATCAAAGTTTCTTGGATGATGCACAAATG
GGTAACGGATTCGTATTAACCTGCGTAGCTTACCCTACATCCGATTGCACTATCCAAACTCACCAAGAAGAAG
CATTGTACGCTAGCTAATAA

HydA2_Fd (*Chlamydomonas reinhardtii*)
MTATDAVPHWKLALEELDKPKDGGRKVLIAQVAPAVRVAIAESFGLAPGAVSPGKLATGLRALGFDQVFDTLF
AADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMFTSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMGAMVKTY
LSEKQGIPAKDIVMVSVMPCVRKQGEADREWFCVSEPGVRDVDHVITTAELGNIFKERGINLPELPDSDWDQP
LGLGSGAGVLFGTTGGVMEAALRTAYEIVTKEPLPRLNLSEVRGLDGIKEASVTLVPAPGSKFAELVAERLAH
KVEEAAAAEAAAAVEGAVKPPIAYDGGQGFSTDDGKGGLKLRVAVANGLGNAKKLIGKMVSGEAKYDFVEIMA
CPAGCVGGGGQPRSTDKQITQKRQAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHTHYVGGG
GASGGGGSGAYKVTLKTPSGDKTIECPADTYILDAAEEAGLDLPYSCRAGACSSCAGKVAAGTVDQSDQSFLD
DAQMGNGFVLTCVAYPTSDCTIQTHQEEALYAS

hydA2_fd
ATGACTGCAACTGATGCTGTGCCGCATTGGAAATTGGCTTTAGAAGAATTAGACAAACCTAAAGATGGCGGTC
GCAAGGTGTTGATCGCCCAAGTGGCACCCGCGGTTCGTGTCGCCATCGCGGAGTCCTTTGGTTTGGCACCCGG
TGCCGTGAGCCCCGGCAAACTGGCAACTGGGTTGCGCGCGTTGGGTTTCGACCAGGTGTTTGATACTTTATTT
GCTGCCGATTTGACGATCATGGAGGAAGGTACTGAACTGCTGCACCGCCTGAAAGAGCACTTAGAAGCCCATC
CGCATAGCGACGAACCGTTGCCGATGTTTACCAGCTGTTGTCCCGGCTGGGTGGCAATGATGGAAAAATCCTA
TCCGGAACTGATTCCCTTTGTTAGCTCCTGTAAATCCCCGCAGATGATGATGGGTGCGATGGTGAAAACTTAC
CTGAGTGAAAAGCAGGGTATTCCCGCCAAAGACATTGTGATGGTGTCCGTTATGCCGTGTGTTCGCAAGCAGG
GTGAAGCTGATCGTGAATGGTTCTGCGTGAGCGAACCCGGCGTTCGCGATGTCGATCATGTGATCACCACTGC

TGAATTAGGGAACATCTTTAAAGAGCGCGGCATTAACCTGCCGGAGCTGCCGGACAGCGATTGGGACCAACCG
CTGGGTTTAGGGTCCGGCGCGGGTGTGTTGTTTGGTACCACCGGCGGTGTTATGGAGGCCGCCTTGCGGACGG
CATATGAGATCGTGACCAAAGAACCCCTGCCCCGTCTGAACCTGAGTGAAGTGCGTGGCTTGGACGGCATCAA
AGAAGCGAGCGTTACCCTGGTGCCGGCTCCGGGGTCCAAATTCGCGGAATTAGTTGCAGAACGCCTGGCACAT
AAGGTCGAGGAGGCCGCGGCTGCCGAAGCGGCTGCGGCAGTGGAAGGGGCGGTGAAACCCCCGATTGCCTATG
ACGGCGGGCAGGGGTTCAGTACTGACGATGGTAAGGGTGGTTTGAAGCTGCGCGTCGCCGTCGCGAACGGTCT
GGGTAATGCTAAGAAACTGATCGGTAAAATGGTGAGCGGTGAAGCAAAATACGATTTTGTTGAGATCATGGCC
TGCCCCGCAGGCTGTGTTGGCGGTGGCGGTCAACCCCGTTCCACCGATAAACAGATTACCCAGAAACGGCAGG
CTGCCCTGTATGATCTGGATGAACGGAACACCCTGCGTCGTTCCCATGAAAATGAGGCAGTTAATCAACTGTA
TAAAGAATTTTTAGGCGAGCCCTTGAGTCATCGCGCGCATGAATTACTGCATACCCATTACGTGGGTGGCGGT
GGTGCTAGCGGCGGTGGAGGCAGTGGCGCTTATAAGGTCACTTTGAAAACCCCTTCCGGAGATAAAACGATTG
AATGTCCCGCAGATACTTATATCTTAGATGCCGCTGAAGAGGCAGGATTAGACTTGCCCTATAGTTGTCGAGC
AGGAGCATGCAGCAGCTGTGCAGGTAAGGTCGCAGCCGGTACCGTAGATCAATCCGATCAAAGTTTCTTGGAT
GATGCACAAATGGGTAACGGATTCGTATTAACCTGCGTAGCTTACCCTACATCCGATTGCACTATCCAAACTC
ACCAAGAAGAAGCATTGTACGCTAGCTAATAA

HydA_Fd (*Clostridium acetobotylicum* ATCC 824)

MKTIILNGNEVHTDKDITILELARENNVDIPTLCFLKDCGNFGKCGVCMVEVEGKGFRAACVAKVEDGMVINT
ESDEVKERIKKRVSMLLDKHEFKCGQCSRRENCEFLKLVIKTKAKASKPFLPEDKDALVDNRSKAIVIDRSKC
VLCGRCVAACKQHTSTCSIQFIKKDGQRAVGTVDDVCLDDSTCLLCGQCVIACPVAALKEKSHIEKVQEALND
PKKHVIVAMAPSVRTAMGELFKMGYGKDVTGKLYTALRMLGFDKVFDINFGADMTIMEEATELLGRVKNNGPF
PMFTSCCPAWVRLAQNYHPELLDNLSSAKSPQQIFGTASKTYYPSISGIAPEDVYTVTIMPCNDKKYEADIPF
METNSLRDIDASLTTRELAKMIKDAKIKFADLEDGEVDPAMGTYSGAGAIFGATGGVMEAAIRSAKDFAENKE
LENVDYTEVRGFKGIKEAEVEIAGNKLNVAVINGASNFFEFMKSGKMNEKQYHFIEVMACPGGCINGGGQPHV
NALDRENVDYRKLRASVLYNQDKNVLSKRKSHDNPAIIKMYDSYFGKPGEGLAHKLLHVKYTGGGASGGGGSA
YKITDACVSCGSCASECPVSAISQGDTQFVIDADTCIECGNCANVCPVGAPVQE

hydA_fd

ATGAAAACTATTATCTTAAATGGAAATGAAGTGCACACGGACAAAGATATTACCATTTTAGAATTAGCACGTG
AAAATAATGTAGATATCCCTACTTTGTGCTTCTTAAAAGATTGCGGCAACTTCGGCAAATGTGGCGTCTGTAT
GGTCGAAGTTGAAGGTAAAGGTTTCCGCGCTGCGTGTGTCGCCAAAGTAGAAGATGGAATGGTCATTAATACT
GAGTCTGATGAAGTTAAAGAACGCATCAAAAAGCGCGTCAGCATGTTATTAGATAAACATGAATTTAAGTGCG
GACAATGTTCCCGCCGTGAAAATTGTGAGTTTCTCAAATTGGTTATCAAAACTAAAGCTAAAGCAAGTAAACC
CTTTTTACCTGAAGATAAAGATGCATTAGTGGATAATCGTTCTAAAGCAATCGTTATCGATCGATCTAAATGT
GTATTATGCGGTCGCTGTGTAGCGGCTTGTAAACAACATACTAGCACGTGTTCTATTCAATTTATCAAGAAAG
ACGGACAACGCGCCGTCGGTACTGTTGATGACGTGTGTTTAGACGACAGTACTTGTCTCTTATGTGGTCAATG
TGTAATCGCTTGTCCCGTGGCGGCCTTAAAAGAAAAATCTCACATTGAAAAAGTCCAAGAAGCTTTAAATGAT
CCTAAAAAGCATGTAATTGTCGCTATGGCACCTTCTGTACGTACCGCGATGGGTGAACTCTTTAAAATGGGAT
ATGGTAAAGATGTAACTGGCAAATTGTACACAGCGTTACGCATGTTAGGCTTTGACAAAGTCTTTGATATCAA
TTTTGGTGCAGATATGACTATCATGGAAGAAGCCACAGAATTGTTAGGCCGCGTCAAAAACAACGGCCCCTTT
CCTATGTTTACTTCCTGTTGTCCTGCTTGGGTACGCTTAGCCCAAAATTATCATCCCGAACTCTTAGACAATT
TAAGTTCCGCAAAAAGTCCCCAACAAATCTTTGGCACTGCAAGTAAAACTTATTATCCCTCCATCTCTGGCAT
TGCTCCCGAGGATGTATATACAGTAACTATCATGCCCTGTAACGATAAGAAATATGAAGCAGATATCCCTTTT
ATGGAAACTAATTCTTTACGAGATATCGATGCCTCCTTAACTACCCGAGAATTAGCAAAGATGATCAAAGATG
CAAAAATCAAATTTGCTGACTTAGAAGATGGTGAAGTAGATCCCGCCATGGGCACTTATTCTGGCGCGGGTGC
GATTTTTGGAGCTACTGGTGGTGTAATGGAAGCTGCTATTCGCTCCGCGAAAGATTTCGCGGAAAACAAAGAA
TTAGAAAATGTAGATTATACCGAAGTCCGAGGTTTTAAAGGAATTAAAGAAGCGGAGGTAGAAATTGCGGGTA
ATAAATTAAACGTTGCGGTGATTAATGGCGCAAGTAATTTCTTTGAATTTATGAAATCTGGAAAAATGAATGA
AAAACAATATCACTTTATTGAAGTAATGGCCTGTCCCGGTGGATGTATCAATGGAGGAGGTCAACCTCATGTA
AACGCTTTAGATCGCGAAAATGTAGACTATCGCAAATTACGAGCGAGTGTCCTCTATAATCAAGATAAAAATG
TTTTGAGCAAACGAAAAAGTCACGATAATCCCGCTATCATTAAAATGTATGACAGTTATTTCGGCAAACCCGG
TGAAGGATTAGCTCATAAATTGTTACATGTGAAATATACAGGCGGTGGTGCTAGCGGCGGTGGAGGCAGTGCC
TATAAGATCACTGACGCTTGTGTAAGTTGTGGAAGTTGTGCCTCCGAGTGCCCCGTCAGCGCGATCTCCCAAG
GTGACACTCAATTTGTCATCGATGCTGACACTTGTATTGAATGCGGTAATTGTGCAAATGTGTGCCCTGTCGG
CGCCCCCGTGCAAGAAGCTAGCTAATAA

HydEF (*Chlamydomonas reinhardtii*)

MAHSLSAHSRQAGDRKLGAGAASSRPSCPSRRIVRVAAHASASKATPDVPVDDLPPAHARAAVAAANRRARAM
ASAEAAAETLGDFLGLGKGGLSPGATANLDREQVLGVLEAVWRRGDLNLERALYSHANAVTNKYCGGGVYYRG
LVEFSNICQNDCSYCGIRNNQKEVWRYTMPVEEVVEVAKWALENGIRNIMLQGGELKTEQRLAYLEACVRAIR
EETTQLDLEMRARAASTTTAEAAASAQADAEAKRGEPELGVVVSLSVGELPMEQYERLFRAGARRYLIRIETS

NPDLYAALHPEPMSWHARVECLRNLKKAGYMLGTGVMVGLPGQTLHDLAGDVMFFRDIKADMIGMGPFITQPG
TPATDKWTALYPNANKNSHMKSMFDLTTAMNALVRITMGNVNISATTALQAIIPTGREIALERGANVVMPILT
PTQYRESYQLYEGKPCITDTAVQCRRCLDMRLHSVGKTSAAGVWGDPASFLHPIVGVPVPHDLSSPALAAAAS
ADFHEVGAGPWNPIRLERLVEVPDRYPDPDNHGRKKAGAGKGGKAHDSHDDGDHDDHHHHHGAAPAGAAAGKG
TGAAAIGGGAGASRQRVAGAAAASARLCAGARRAGRVVASPLRPAAACRGVAVKAAAAAAGEDAGAGTSGVGS
NIVTSPGIASTTAHGVPRINIGVFGVMNAGKSTLVNALAQQEACIVDSTPGTTADVKTVLLELHALGPAKLLD
TAGLDEVGGLGDKKRRKALNTLKECDVAVLVVDTDTAAAAIKSGRLAEALEWESKVMEQAHKYNVSPVLLLNV
KSRGLPEAQAASMLEAVAGMLDPSKQIPRMSLDLASTPLHERSTITSAFVKEGAVRSSRYGAPLPGCLPRWSL
GRNARLLMVIPMDAETPGGRLLRPQAQVMEEAIRHWATVLSVRLDLDAARGKLGPEACEMERQRFDGVIAMME
RNDGPTLVVTDSQAIDVVHPWTLDRSSGRPLVPITTFSIAMAYQQNGGRLDPFVEGLEALETLQDGDRVLISE
ACNHNRITSACNDIGMVQIPNKLEAALGGKKLQIEHAFGREFPELESGGMDGLKLAIHCGGCMIDAQKMQQRM
KDLHEAGVPVTNYGVFFSWAAWPDALRRALEPWGVEPPVGTPATPAAAPATAASGV

hydEF

ATGGCACATAGTTTAAGTGCACATAGTCGACAAGCAGGAGATCGTAAATTAGGCGCCGGTGCGGCTTCTTCTC
GCCCCAGCTGTCCCAGCCGTCGAATTGTACGAGTGGCCGCTCATGCCAGTGCGTCTAAAGCAACCCCCGACGT
GCCTGTTGACGATTTACCCCCTGCACATGCACGAGCAGCTGTCGCAGCTGCTAATCGTCGAGCTCGAGCTATG
GCAAGTGCTGAAGCCGCAGCGGAAACCTTGGGAGACTTTTTAGGCCTCGGTAAAGGCGGTTTAAGCCCCGGTG
CTACGGCAAATTTAGACCGCGAGCAAGTCTTGGGCGTATTAGAAGCGGTATGGCGACGTGGTGACTTAAACTT
AGAACGTGCTTTATATAGTCATGCTAATGCTGTGACTAACAAATATTGCGGAGGTGGTGTCTATTATCGCGGT
TTGGTAGAATTTAGCAACATCTGCCAAAACGATTGCAGTTATTGCGGTATTCGCAATAATCAAAAAGAAGTAT
GGCGCTATACCATGCCTGTGGAAGAAGTTGTCGAAGTCGCAAAATGGGCATTAGAAAATGGTATCCGTAACAT
CATGCTCCAAGGAGGTGAGCTCAAAACAGAACAACGTTTGGCCTATTTAGAAGCTTGCGTGCGCGCGATTCGC
GAAGAAACAACCCAATTAGACTTAGAGATGCGAGCACGAGCCGCTAGCACTACAACAGCCGAAGCAGCCGCCT
CTGCCCAAGCTGACGCTGAAGCTAAACGTGGTGAACCTGAATTAGGAGTGGTCGTCTCCTTAAGCGTGGGCGA
ATTGCCCATGGAGCAATATGAACGTTTATTTCGCGCTGGTGCTCGCCGATATCTCATCCGCATCGAAACAAGT
AATCCCGACTTATATGCCGCATTACACCCCGAACCTATGAGCTGGCACGCCCGCGTCGAGTGTTTACGCAACT
TAAAGAAAGCTGGATACATGTTAGGCACTGGTGTTATGGTCGGCCTCCCCGGCCAAACTCTCCACGATCTCGC
AGGCGATGTAATGTTCTTTCGTGATATTAAAGCTGACATGATTGGTATGGGCCCCTTCATTACTCAACCCGGT
ACACCCGCTACCGATAAATGGACGGCATTATATCCCAATGCAAATAAGAATAGTCATATGAAATCCATGTTTG
ATTTAACCACTGCAATGAATGCGTTGGTACGTATCACCATGGGTAATGTCAATATTTCTGCTACCACAGCATT
ACAAGCAATTATTCCTACTGGCCGCGAAATCGCGTTAGAACGTGGTGCGAATGTCGTAATGCCTATCTTAACC
CCTACTCAATACCGCGAGAGTTATCAATTATACGAAGGTAAACCCTGTATTACCGACACTGCTGTACAATGCC
GTCGCTGTTTAGATATGCGCCTCCATAGTGTGGGAAAAACCAGTGCCGCAGGTGTGTGGGGAGATCCCGCATC
CTTTTTGCATCCTATTGTCGGTGTTCCCGTGCCCCATGACTTATCCAGCCCCGCACTCGCCGCTGCTGCAAGT
GCCGATTTTCATGAAGTGGGTGCGGGCCCTTGGAACCCCATTCGCTTAGAGCGATTAGTAGAGGTACCCGATC
GCTATCCCGATCCCGATAACCATGGTCGCAAGAAAGCGGGCGCAGGAAAAGGCGGAAAAGCACATGACTCTCA
TGATGATGGTGATCATGACGATCATCACCATCACCATGGCGCAGCTCCCGCCGGTGCCGCGGCAGGCAAAGGA
ACCGGCGCTGCTGCTATCGGAGGTGGTGCAGGTGCCTCTCGCCAACGCGTTGCTGGCGCGGCAGCAGCTAGTG
CTCGCTTATGTGCTGGAGCCCGACGTGCTGGTCGCGTGGTTGCAAGCCCCTTACGCCCCGCAGCGGCCTGCCG
AGGCGTGGCAGTGAAAGCTGCCGCAGCAGCTGCTGGAGAAGATGCTGGTGCCGGTACATCTGGTGTAGGTTCT
AATATCGTTACTAGCCCTGGTATTGCGAGCACCACTGCGCATGGCGTACCTCGTATCAATATCGGTGTCTTTG
GCGTGATGAATGCTGGTAAAAGCACCTTAGTGAATGCCTTGGCTCAACAAGAAGCCTGTATTGTAGATAGCAC
CCCTGGTACTACCGCAGATGTAAAAACAGTTTTATTAGAGTTGCATGCTTTAGGACCTGCCAAGTTATTAGAT
ACGGCTGGTTTGGACGAAGTAGGTGGCCTCGGAGATAAGAAACGCCGTAAAGCACTCAATACTTTAAAAGAGT
GCGATGTTGCCGTTTTGGTTGTTGATACCGATACTGCTGCTGCTGCCATTAAAAGCGGTCGATTAGCCGAAGC
ATTAGAATGGGAGTCTAAAGTAATGGAGCAAGCTCACAAATACAACGTGAGTCCCGTGTTGTTGTTAAATGTA
AAATCCCGCGGCTTACCTGAAGCACAAGCGGCCAGTATGTTAGAAGCAGTGGCCGGCATGTTAGATCCTAGCA
AACAAATCCCCCGAATGAGCTTGGACCTCGCAAGTACTCCCCTCCATGAGCGCAGTACAATTACCAGTGCATT
TGTAAAAGAGGGAGCGGTTCGCAGCAGTCGTTATGGCGCCCCTTTACCCGGTTGCCTCCCTCGCTGGTCCCTC
GGCCGCAACGCTCGCTTATTAATGGTTATCCCCATGGATGCGGAGACTCCTGGTGGCCGTTTATTACGCCCTC
AAGCACAAGTGATGGAGGAAGCCATTCGCCACTGGGCAACCGTCTTAAGCGTGCGCTTAGATTTAGATGCAGC
TCGCGGTAAATTAGGCCCCGAAGCCTGCGAAATGGAACGCCAACGCTTTGATGGCGTGATTGCCATGATGGAA
CGTAATGACGGACCCACTCTCGTCGTAACTGACAGTCAAGCAATTGATGTGGTGCACCCCTGGACATTAGATC
GTAGCAGCGGCCGTCCTTTGGTCCCTATCACTACATTTTCTATCGCCATGGCGTATCAACAAAATGGCGGTCG
CTTGGATCCCTTTGTGGAAGGTTTAGAAGCGTTAGAAACCCTCCAAGATGGTGACCGCGTTTTGATCAGTGAA
GCGTGTAATCACAATCGTATTACTTCTGCGTGTAACGACATCGGTATGGTCCAAATCCCTAACAAATTAGAAG
CGGCTTTAGGCGGTAAAAAGCTCCAAATCGAACACGCATTCGGCCGCGAATTTCCCGAGTTAGAGTCTGGAGG
TATGGATGGACTCAAATTGGCAATCCACTGTGGTGGCTGTATGATTGACGCTCAAAAGATGCAACAACGTATG
AAAGATTTACATGAAGCGGGCGTCCCCGTTACGAATTATGGTGTGTTCTTTAGTTGGGCAGCATGGCCTGATG

CTTTACGCCGTGCCTTAGAACCCTGGGGTGTCGAACCTCCCGTAGGTACTCCCGCAACACCCGCGGCTGCTCC
CGCTACCGCTGCAAGTGGCGTGTAATAA

HydG *Chlamydomonas reinhardtii*)

MSVPLQCNAGRLLAGQRPCGVRARLNRRVCVPVTAHGKASATREYAGDFLPGTTISHAWSVERETHHRYRNPA
EWINEAAIHKALETSKADAQDAGRVREILAKAKEKAFVTEHAPVNAESKSEFVQGLTLEECATLINVDSNNVE
LMNEIFDTALAIKERIYGNRVVLFAPLYIANHCMNTCTYCAFRSANKGMERSILTDDDLREEVAALQRQGHRR
ILALTGEHPKYTFDNFLHAVNVIASVKTEPEGSIRRINVEIPPLSVSDMRRLKNTDSVGTFVLFQETYHRDTF
KVMHPSGPKSDFDFRVLTQDRAMRAGLDDVGIGALFGLYDYRYEVCAMLMHSEHLEREYNAGPHTISVPRMRP
ADGSELSIAPPYPVNDADFMKLVAVLRIAVPYTGMILSTRESPEMRSALLKCGMSQMSAGSRTDVGAYHKDHT
LSTEANLSKLAGQFTLQDERPTNEIVKWLMEEGYVPSWCTACYRQGRTGEDFMNICKAGDIHDFCHPNSLLTL
QEYLMDYADPDLRKKGEQVIAREMGPDASEPLSAQSRKRLERKMKQVLEGEHDVYL

hydG

ATGTCTGTACCCTTACAATGCAATGCCGGTCGCTTATTAGCAGGTCAACGCCCCTGCGGAGTTCGCGCACGCT
TAAATCGCCGCGTATGTGTACCTGTAACTGCTCACGGTAAAGCATCTGCAACTCGTGAATATGCAGGTGATTT
CCTCCCTGGTACCACCATCAGCCATGCGTGGTCCGTAGAACGCGAAACTCACCATCGCTATCGTAATCCCGCC
GAATGGATTAATGAAGCTGCTATTCACAAAGCTTTAGAAACCTCTAAGGCAGACGCTCAAGATGCAGGACGCG
TTCGCGAAATCTTAGCTAAAGCGAAAGAAAAAGCCTTTGTTACTGAGCATGCGCCCGTGAATGCAGAAAGTAA
AAGCGAATTTGTCCAAGGTTTAACCTTAGAAGAATGCGCTACCTTAATTAACGTTGATTCTAATAATGTGGAA
TTAATGAATGAGATTTTCGACACCGCTTTGGCTATCAAAGAACGCATCTATGGTAACCGCGTGGTTTTGTTTG
CTCCCTTATATATCGCTAATCATTGCATGAATACCTGCACGTATTGCGCTTTTCGTTCTGCGAATAAAGGCAT
GGAGCGTTCTATCTTAACTGATGATGATTTACGTGAAGAAGTCGCGGCCTTACAACGACAAGGTCACCGCCGC
ATTTTAGCCTTAACTGGCGAACATCCCAAATATACCTTTGATAACTTTTTACATGCGGTTAACGTTATCGCCT
CCGTTAAAACTGAGCCCGAAGGAAGTATCCGCCGCATCAATGTGGAGATTCCCCCTTTATCTGTGAGTGATAT
GCGTCGCTTGAAAAATACTGATTCCGTCGGCACATTTGTGTTGTTTCAAGAAACTTACCACCGCGATACTTTT
AAAGTCATGCATCCTAGTGGACCTAAAAGTGATTTTGATTTCCGCGTTCTCACCCAAGACCGCGCCATGCGAG
CCGGATTGGATGATGTTGGTATTGGAGCGTTATTTGGTTTGTATGATTATCGATACGAAGTGTGCGCTATGTT
AATGCATTCCGAACATTTGGAACGAGAATATAATGCAGGTCCCCATACCATTAGTGTACCTCGAATGCGTCCT
GCCGATGGAAGTGAATTATCTATCGCACCCCCTTATCCCGTTAACGATGCAGACTTCATGAAATTAGTGGCTG
TCTTACGCATTGCGGTGCCTTATACGGGCATGATTCTCAGTACTCGAGAAAGCCCCGAAATGCGCAGTGCCTT
ACTCAAATGCGGAATGTCCCAAATGAGCGCTGGTAGTCGTACCGATGTGGGCGCCTATCATAAGGATCATACG
TTATCTACTGAAGCAAACCTCAGTAAATTAGCAGGTCAATTTACATTACAAGACGAACGCCCCACCAATGAAA
TTGTTAAATGGTTAATGGAAGAAGGATACGTGCCCTCCTGGTGCACCGCCTGTTATCGACAAGGACGCACTGG
AGAAGACTTTATGAATATTTGTAAAGCGGGCGACATTCATGACTTCTGTCATCCTAACTCTTTATTAACTTTA
CAAGAGTATCTCATGGATTATGCGGATCCCGATTTGCGCAAAAAGGGCGAACAAGTAATCGCTCGTGAAATGG
GTCCCGATGCGAGCGAGCCCTTGTCCGCCCAAAGTCGTAAACGATTAGAGCGCAAAATGAAACAAGTCTTAGA
GGGAGAGCATGATGTTTATTTATAATAA

HydE (*Clostridium acetobutylicum* ATCC 824)

MDNIIKLINKAEVTHDLTKDELVTLLKDDTHNEEIYKAADRVREKYVGEEVHLRGLIEFSNICKRNCMYCGLR
RDNKNIKRYRLEPDEIIHLAKSAKNYGYQTVVLQSGEDDYYTVEKMKYIVSEIKKLNMAITLSIGEKTFEEYE
EYRKSGADRYLIRIETTDKELYEKLDPKMSHENRINCLKNLRKLGYEVGSGCLVGLPNQTIESLADDILFFKE
IDADMIGVGPFIPNEDTPLGEEKGGEFFMSVKVTALIRLLLPDINIPATTAMESLYPNGRSIALTSGANVVMP
NVTEGEYRKLYALYPGKICVNDTPGHCRQCISLKINKINRKVSATKGFRKKSYKESIG

hydE

ATGGATAATATCATCAAACTCATCAATAAAGCAGAAGTAACTCACGATTTAACGAAAGATGAATTAGTCACCT
TATTGAAAGATGATACTCATAATGAGGAAATCTATAAAGCCGCGGATCGTGTACGTGAGAAATACGTTGGTGA
AGAAGTACATCTCCGTGGCCTCATCGAATTTTCTAACATCTGTAAACGTAATTGTATGTATTGCGGATTGCGA
CGCGACAACAAAAATATTAAACGCTATCGATTGGAGCCTGATGAGATCATTCACTTAGCTAAAAGTGCTAAAA
ACTACGGATACCAAACGGTCGTCTTACAAAGTGGAGAAGATGATTATTACACTGTAGAGAAAATGAAATATAT
CGTTAGTGAAATCAAGAAACTCAATATGGCGATTACTTTAAGTATCGGTGAGAAAACGTTCGAGGAATACGAG
GAGTATCGTAAATCTGGTGCTGATCGCTATTTAATCCGTATCGAAACTACTGATAAAGAATTATATGAGAAAT
TAGATCCTAAAATGTCTCATGAAAACCGAATTAATTGCTTAAAAAACTTGCGTAAATTAGGTTACGAAGTTGG
TAGCGGATGTTTAGTCGGTTTACCCAACCAAACTATTGAAAGCCTCGCTGATGATATTTTGTTCTTTAAAGAA
ATTGATGCTGATATGATCGGTGTCGGTCCTTTTATTCCCAATGAGGACACCCCCTTGGGCGAAGAAAAAGGCG
GCGAGTTCTTTATGAGTGTAAAAGTGACCGCCCTCATCCGCTTATTATTGCCCGACATCAACATCCCCGCAAC
TACCGCTATGGAAAGTTTATATCCTAATGGCCGCTCTATCGCTTTAACATCTGGTGCTAATGTCGTTATGCCC
AACGTGACTGAAGGAGAATATCGCAAATTATACGCCCTCTATCCCGGTAAAATCTGTGTGAACGACACTCCCG

GTCATTGCCGCCAATGTATTTCCTTAAAAATCAACAAAATCAATCGTAAAGTATCCGCCACCAAAGGCTTTCG
TAAAAAGTCCTACAAAGAATCTATTGGTTAATAA

HydF (*Clostridium acetobutylicum* ATCC 824)

MNELNSTPKGERLHIALFGKTNVGKSSVINALTSQEIALVSNVKGTTTDPVYKAMELLPLGPVMLIDTAGLDD
ISDLGELRRGKTLEVLSKTDVAILVFDVESGITEYDKNIYSLLLEKKIPLIGVLNKIDKKDYKLEDYTSQFKI
PIVPISALNNKGINNLKDELIRLAPENDDKFKIVGDLLSPGDIAVLVTPIDKAAPKGRLILPQQQTIRDILES
DAIAMVTKEFELRETLDSLRKKPKIVITDSQVFLKVAADTPKDILMTSFSILMARHKGDLIELARGARAIEDL
KDGDKILIAEACTHHRQSDDIGKVKIPRWLRQKTGKKLEFDFSSGFSFPPNIEDYALIVHCAGCMLNRRSMLH
RIESSVKKQIPIVNYGVLIAYVQGILPRALKPFPYADRIFNQSSRN

hydF

ATGAATGAGTTAAATAGTACGCCCAAGGGAGAACGATTACATATCGCATTATTTGGTAAAACTAACGTAGGTA
AAAGTTCTGTCATTAATGCTTTAACCTCCCAAGAAATTGCGTTAGTTTCCAATGTGAAAGGAACAACTACCGA
TCCCGTTTATAAAGCCATGGAATTATTACCTTTAGGTCCTGTTATGTTAATTGATACTGCGGGATTAGATGAT
ATCAGTGATTTAGGAGAATTGCGTCGCGGTAAAACCTTAGAGGTATTATCTAAAACCGACGTTGCCATTTTAG
TATTTGATGTTGAAAGTGGTATTACGGAATATGATAAAAACATCTATTCCCTCTTATTAGAGAAAAAGATTCC
TCTCATTGGAGTCTTAAACAAAATTGATAAGAAAGACTATAAACTCGAAGATTATACAAGTCAATTCAAAATT
CCCATCGTACCCATCAGTGCTTTAAACAATAAAGGCATTAACAACTTAAAAGATGAATTAATTCGCTTAGCGC
CCGAAAATGATGACAAATTCAAAATTGTCGGTGATTTGTTATCTCCTGGCGATATTGCTGTCTTGGTCACGCC
CATTGATAAAGCTGCACCCAAAGGACGCCTCATTTTACCCCAACAACAAACCATTCGCGATATTTTAGAATCC
GATGCTATCGCAATGGTTACGAAGGAATTTGAATTACGCGAAACTTTGGATAGTCTCCGAAAAAAGCCCAAGA
TTGTTATCACTGATAGTCAAGTCTTTTTGAAAGTCGCAGCAGATACACCCAAAGATATCTTAATGACCTCTTT
TAGTATCTTAATGGCTCGACATAAAGGTGATTTAATCGAGCTCGCTCGTGGTGCCCGTGCTATCGAAGATCTC
AAAGATGGTGACAAAATTCTCATCGCAGAAGCGTGTACTCATCACCGTCAATCCGACGACATTGGCAAAGTCA
AAATCCCTCGCTGGTTGCGCCAAAAAACTGGTAAGAAATTGGAATTTGACTTCAGTTCCGGTTTTAGCTTTCC
TCCCAACATTGAAGATTATGCTTTAATCGTGCACTGCGCTGGCTGCATGTTGAACCGCCGCAGTATGTTACAT
CGCATCGAATCCAGCGTGAAGAAACAAATCCCTATCGTGAACTATGGCGTTTTAATCGCGTATGTACAAGGCA
TTTTACCTCGAGCATTAAAACCCTTTCCCTATGCTGATCGCATTTTTAACCAAAGCAGTCGTAATTAATAA

HydG (*Clostridium acetobutylicum* ATCC 824)

MYNVKSKVATEFISDEEIIDSLEYAKQNKSNRELIDSIIEKAKECKGLTHRDAAVLLECDLEDENEKMFKLAR
EIKQKFYGNRIVMFAPLYLSNYCVNGCVYCPYHHKNKHIARKKLSQEDVKRETIALQDMGHKRLALEAGEDPV
NNPIEYILDCIKTIYSIKHKNGAIRRVNVNIAATTVENYKKLKDAGIGTYILFQETYNKKSYEELHPTGPKHD
YAYHTEAMDRAMEGGIDDVGIGVLFGLNMYKYDFVGLLMHAEHLEAAMGVGPHTISVPRIRPADDIDPENFSN
AISDEIFEKIVAIIRIAVPYTGMIVSTRESKKTRERVLELGISQISGGSSTSVGGYVESEPEEDNSSQFEVND
NRTLDEIVNWLLEMNYIPSFCTACYREGRTGDRFMSLVKSGQIANCCQPNALMTLKEYLEDYASSNTQKNGEA
LIASEVEKIPNEKVKSIVKKHLTELKEGQRDFRF

hydG

ATGTATAACGTCAAAAGCAAAGTCGCAACTGAGTTTATTAGTGATGAAGAAATCATCGATAGCTTAGAATACG
CAAAACAAAACAAAAGTAATCGCGAATTGATTGATTCTATTATTGAAAAGGCTAAAGAATGCAAAGGCTTAAC
TCATCGCGATGCCGCTGTATTGTTAGAATGTGATTTGGAAGATGAAAACGAAAAGATGTTTAAGTTAGCACGC
GAAATCAAACAAAAGTTTTATGGTAACCGCATCGTCATGTTTGCTCCCTTATATCTCTCTAACTATTGTGTAA
ATGGTTGTGTTTATTGTCCCTATCATCACAAAAACAAACATATCGCACGCAAGAAATTATCTCAAGAAGATGT
GAAACGCGAAACTATTGCATTACAAGACATGGGTCATAAACGATTAGCCTTAGAAGCCGGTGAAGATCCCGTA
AATAATCCCATCGAATATATCTTGGATTGTATCAAGACTATTTATTCCATCAAACATAAAAATGGTGCTATTC
GCCGCGTTAATGTTAATATCGCTGCGACTACCGTAGAAAACTATAAAAAATTGAAAGATGCAGGTATTGGCAC
CTACATCTTATTTCAAGAAACCTATAATAAGAAATCTTATGAAGAATTGCACCCCACTGGTCCTAAACACGAC
TATGCGTATCATACGGAAGCAATGGACCGTGCTATGGAAGGTGGCATCGACGACGTAGGTATTGGTGTTTTAT
TTGGTTTGAATATGTACAAATATGACTTTGTTGGCCTCTTGATGCACGCCGAACATTTAGAAGCCGCTATGGG
AGTTGGTCCCCATACAATTTCTGTACCCCGCATTCGACCCGCTGATGACATCGATCCCGAAAACTTTTCTAAT
GCTATCTCTGATGAAATTTTTGAAAAAATCGTCGCTATTATTCGCATTGCTGTTCCTTATACCGGAATGATCG
TAAGTACTCGAGAATCTAAAAAGACTCGCGAACGAGTCTTAGAACTCGGCATCTCCCAAATTTCCGGTGGCAG
CTCCACCTCTGTTGGCGGTTATGTCGAAAGCGAACCCGAGGAGGATAATAGTTCTCAATTTGAGGTTAACGAC
AACCGCACTCTCGACGAGATCGTAAATTGGCTCTTAGAAATGAATTATATCCCCAGTTTTTGTACCGCATGCT
ATCGCGAAGGACGAACCGGCGATCGTTTTATGAGCCTCGTTAAAAGTGGTCAAATTGCAAATTGTTGTCAACC
TAACGCTTTAATGACCTTAAAAGAATACTTGGAAGATTATGCTAGTAGTAATACTCAAAAAAATGGTGAAGCA

TTGATTGCTTCCGAAGTAGAAAAGATCCCCAACGAAAAAGTCAAAAGCATCGTAAAGAAACATCTCACTGAAT
TAAAAGAAGGACAACGCGATTTTCGATTTTAATAA

ZFA1
MAERPFQCRICMRNFSQSSSLVRHIRTHTGEKPFACDICGRKFADCRDLARHTKIHTGSERPFQCRICMRNFS
RSDDLVRHIRTHTGEKPFACDICGRKFARSDNLVRHTKIHLRQKDA

zfa1
ATGGCAGAACGTCCTTTTCAATGCCGTATTTGTATGCGCAATTTCAGTCAGAGCAGTAGCCTGGTGCGTCATA
TTCGTACGCACACTGGCGAGAAACCGTTTGCGTGTGATATTTGTGGCCGCAAGTTTGCTGACTGTCGCGATCT
GGCACGTCATACGAAAATTCATACCGGTAGCGAGCGCCCGTTCCAGTGTCGCATCTGCATGCGTAACTTTTCT
CGCAGCGACGATCTGGTTCGCCACATCCGCACACATACCGGTGAGAAGCCATTCGCTTGCGACATCTGCGGGC
GTAAATTCGCCCGTTCTGATAACCTGGTTCGTCACACCAAGATTCACCTGCGCCAGAAGGATGCCTAATAA

ZFA2
MAERPFQCRICMRNFSDPGNLVRHIRTHTGEKPFACDICGRKFAQRAHLERHTKIHTGSERPFQCRICMRNFS
QSSSLVRHIRTHTGEKPFACDICGRKFADPGNLVRHTKIHLRQKDA

zfa2
ATGGCAGAACGTCCTTTTCAATGCCGTATTTGTATGCGCAATTTCAGTGACCCAGGGAATCTGGTGCGTCATA
TTCGTACGCACACTGGCGAGAAACCGTTTGCGTGTGATATTTGTGGCCGCAAGTTTGCTCAGCGTGCTCACCT
GGAACGTCATACGAAAATTCATACCGGTAGCGAGCGCCCGTTCCAGTGTCGCATCTGCATGCGTAACTTTTCT
CAGTCTAGTAGTCTGGTTCGCCACATCCGCACACATACCGGTGAGAAGCCATTCGCTTGCGACATCTGCGGGC
GTAAATTCGCCGATCCTGGTAATCTGGTTCGTCACACCAAGATTCACCTGCGCCAGAAGGATGCCTAATAA

ZFA3
MAERPFQCRICMRNFSDCRDLARHIRTHTGEKPFACDICGRKFARSDDLVRHTKIHTGSERPFQCRICMRNFS
QRAHLERHIRTHTGEKPFACDICGRKFAQSSSLVRHTKIHLRQKDA

zfa3
ATGGCAGAACGTCCTTTTCAATGCCGTATTTGTATGCGCAATTTCAGTGATTGCCGTGATCTGGCCCGTCATA
TTCGTACGCACACTGGCGAGAAACCGTTTGCGTGTGATATTTGTGGCCGCAAGTTTGCTCGCAGCGACGACCT
GGTTCGTCATACGAAAATTCATACCGGTAGCGAGCGCCCGTTCCAGTGTCGCATCTGCATGCGTAACTTTTCT
CAGCGTGCGCACCTGGAGCGCCACATCCGCACACATACCGGTGAGAAGCCATTCGCTTGCGACATCTGCGGGC
GTAAATTCGCCCAATCTAGCAGCCTGGTTCGTCACACCAAGATTCACCTGCGCCAGAAGGATGCCTAATAA

ZFA4
MAERPFQCRICMRNFSDPGNLVRHIRTHTGEKPFACDICGRKFARSDELVRHTKIHTGSERPFQCRICMRNFS
RSDNLVRHIRTHTGEKPFACDICGRKFAQSSSLVRHTKIHLRQKDA

zfa4
ATGGCAGAACGTCCTTTTCAATGCCGTATTTGTATGCGCAATTTCAGTGATCCGGGTAACCTGGTGCGTCATA
TTCGTACGCACACTGGCGAGAAACCGTTTGCGTGTGATATTTGTGGCCGCAAGTTTGCTCGCAGTGACGAACT
GGTACGTCATACGAAAATTCATACCGGTAGCGAGCGCCCGTTCCAGTGTCGCATCTGCATGCGTAACTTTTCT
CGTAGCGATAATCTGGTTCGCCACATCCGCACACATACCGGTGAGAAGCCATTCGCTTGCGACATCTGCGGGC
GTAAATTCGCCCAGTCTTCTTCTCTGGTTCGTCACACCAAGATTCACCTGCGCCAGAAGGATGCCTAATAA

Promoter 1 (P_{cons})
TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC

Promoter 2 (P_{tetR})
TGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCC

Promoter 3 (P_{lacI})
TTGACATCCCTATCAGTGATAGAGATACTGAGCAC

Synthetic operator 1
GAGGCGGCCGTA

Synthetic operator 2
GACGTAGGAGAC

Synthetic operator 3
GTAGGAGCGGCC

Synthetic operator 4
GTAGAGGTGGAC

Promoter libraries synthesized (-35 and -10 elements in gray, operators in bold)

TTGACA**GAGGCGGCCGTA**CTAGGTATAATGCTAGC
TTGACAG**GAGGCGGCCGTA**TAGGTATAATGCTAGC
TTGACAGC**GAGGCGGCCGTA**AGGTATAATGCTAGC
TTGACAGCT**GAGGCGGCCGTA**GGTATAATGCTAGC
TTGACAGCTA**GAGGCGGCCGTA**GTATAATGCTAGC
TTGACAGCTAG**GAGGCGGCCGTA**TATAATGCTAGC

TTGACATCCCTATCAGTGATAGAGATACTGAGCAC
TTGACA**GAGGCGGCCGTA**ATAGAGATACTGAGCAC
TTGACAT**GAGGCGGCCGTA**TAGAGATACTGAGCAC
TTGACATC**GAGGCGGCCGTA**AGAGATACTGAGCAC
TTGACATCC**GAGGCGGCCGTA**GAGATACTGAGCAC
TTGACATCCC**GAGGCGGCCGTA**AGATACTGAGCAC
TTGACATCCCT**GAGGCGGCCGTA**GATACTGAGCAC

TGGTGCAA**GACGTAGGAGAC**TATGGCATGATAGCGCC
TGGTGCAAA**GACGTAGGAGAC**ATGGCATGATAGCGCC
TGGTGCAAAA**GACGTAGGAGAC**TGGCATGATAGCGCC
TGGTGCAAAAC**GACGTAGGAGAC**GGCATGATAGCGCC
TGGTGCAAAACC**GACGTAGGAGAC**GCATGATAGCGCC
TGGTGCAAAACCT**GACGTAGGAGAC**CATGATAGCGCC

TTGACA**GTAGGAGCGGCC**CTAGGTATAATGCTAGC
TTGACAG**GTAGGAGCGGCC**TAGGTATAATGCTAGC
TTGACAGC**GTAGGAGCGGCC**AGGTATAATGCTAGC
TTGACAGCT**GTAGGAGCGGCC**GGTATAATGCTAGC
TTGACAGCTA**GTAGGAGCGGCC**GTATAATGCTAGC
TTGACAGCTAG**GTAGGAGCGGCC**TATAATGCTAGC

TGGTGCAA**GTAGAGGTGGAC**TATGGCATGATAGCGCC
TGGTGCAAA**GTAGAGGTGGAC**ATGGCATGATAGCGCC
TGGTGCAAAA**GTAGAGGTGGAC**TGGCATGATAGCGCC
TGGTGCAAAAC**GTAGAGGTGGAC**GGCATGATAGCGCC
TGGTGCAAAACC**GTAGAGGTGGAC**GCATGATAGCGCC
TGGTGCAAAACCT**GTAGAGGTGGAC**CATGATAGCGCC

CHAPTER V

Selection of suitable reference genes for RT-qPCR analyses in cyanobacteria

Pinto F, Pacheco CC, Ferreira D, Moradas-Ferreira P and Tamagnini P

PLoS One (2012) **7**:e34983



OPEN ACCESS Freely available online

PLoS one

Selection of Suitable Reference Genes for RT-qPCR Analyses in Cyanobacteria

Filipe Pinto[1,2], Catarina C. Pacheco[1], Daniela Ferreira[1,2¤], Pedro Moradas-Ferreira[1,3], Paula Tamagnini[1,2]*

1 IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal, **2** Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal, **3** ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

Abstract

Cyanobacteria are a group of photosynthetic prokaryotes that have a diverse morphology, minimal nutritional requirements and metabolic plasticity that has made them attractive organisms to use in biotechnological applications. The use of these organisms as cell factories requires the knowledge of their physiology and metabolism at a systems level. For the quantification of gene transcripts real-time quantitative polymerase chain reaction (RT-qPCR) is the standard technique. However, to obtain reliable RT-qPCR results the use and validation of reference genes is mandatory. Towards this goal we have selected and analyzed twelve candidate reference genes from three morphologically distinct cyanobacteria grown under routinely used laboratory conditions. The six genes exhibiting less variation in each organism were evaluated in terms of their expression stability using geNorm, NormFinder and BestKeeper. In addition, the minimum number of reference genes required for normalization was determined. Based on the three algorithms, we provide a list of genes for cyanobacterial RT-qPCR data normalization. To our knowledge, this is the first work on the validation of reference genes for cyanobacteria constituting a valuable starting point for future works.

Citation: Pinto F, Pacheco CC, Ferreira D, Moradas-Ferreira P, Tamagnini P (2012) Selection of Suitable Reference Genes for RT-qPCR Analyses in Cyanobacteria. PLoS ONE 7(4): e34983. doi:10.1371/journal.pone.0034983

Editor: Brett Neilan, University of New South Wales, Australia

Received February 2, 2012; **Accepted** March 12, 2012; **Published** April 4, 2012

Copyright: © 2012 Pinto et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was financially supported by the Fundação para a Ciência e a Tecnologia (FCT) (SFRH/BD/36378/2007 (FP), SFRH/BD/16954/2004 (DF), SFRH/BPD/64095/2009 (CCP), PTDC/BIA-MIC/100370/2008), European Science Foundation (ESF) (III Quadro Comunitário de Apoio) and Programa Operacional Factores de Competitividade (COMPETE) in it's FEDER (Fundo Europeu de Desenvolvimento Regional) component. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pmtamagn@ibmc.up.pt

¤ Current address: Department of Microbiology, University of California Davis, Davis, California, United States of America

Introduction

Cyanobacteria are a multifaceted group of photosynthetic prokaryotes, with a large morphological diversity including unicellular, colonial and filamentous forms. Some filamentous strains exhibit cellular differentiation, with cells specialized in photosynthesis (vegetative cells) and in nitrogen fixation (heterocysts) [1–2]. Besides the morphological diversity, cyanobacteria are versatile microorganisms with simple growth requirements that are able to produce several secondary metabolites of commercial interest [3–5]. To exploit this biotechnological potential and use cyanobacteria as cell factories it is necessary to understand their physiology and metabolism at a systems level. The comprehensive evaluation of gene expression patterns is an important part of this characterization and, at present, the quantitative real-time polymerase chain reaction (RT-qPCR) is considered the gold standard for measurement of transcript abundance [6]. This technique allows the simultaneous amplification and quantification of a target amplicon by measuring the fluorescence increment in each PCR cycle in a fast, extremely sensitive and accurate way [7]. However, it has been shown that the poor design of RT-qPCR experiments can lead to erroneous or biologically irrelevant data [6]. A number of parameters such as the sample origin and amount, RNA quality and integrity, PCR efficiency, qPCR protocol and validation have been shown to compromise the quality of the final RT-qPCR experiment [8–10]. Additionally, the data normalization strategy used to eliminate technical or experimentally induced variation is very important. Normalization against internal control genes is frequently used, yet the expression of housekeeping genes can vary significantly depending on the samples and experimental conditions [11–14]. To minimize the influence of individual variation experimental validation of each reference gene is required for a particular sample and condition, and the use of three or more reference genes is now standard practice [10,15]. Moreover, several algorithms have been developed to allow the evaluation of candidate reference genes in terms of their expression stability and determination of the minimum number of reference genes to be used [16–19].

In this work we have validated candidate reference genes for accurate normalization of RT-qPCR data in cyanobacteria. Several genes were selected and their transcription analyzed in three morphological distinct organisms: the unicellular *Synechocystis* sp. PCC 6803 (*Synechocystis*), the filamentous *Lyngbya aestuarii* CCY 9616 (*Lyngbya*), and the filamentous heterocystous *Nostoc* sp. PCC 7120 (*Nostoc*). Our analysis provides a list of reference genes that can be used in RT-qPCR experiments with these cyanobacteria, grown in conditions routinely used in the laboratory. Although validation is always mandatory, our work constitutes a valuable starting point for the selection of reference genes, even for other cyanobacterial strains.

Results and Discussion

Choice of candidate reference genes

Two approaches were followed to choose the candidate reference genes to be validated: (*i*) genes that have been used as controls in previous RT-qPCR studies and (*ii*) genes whose transcription did not vary significantly in two microarray studies [20–21] available in the CyanoBIKE server [22]. Twelve genes from independent pathways were selected to minimize the effects of co-regulation (Table 1): seven based on the bibliography (*i*) and five from CyanoBIKE (*ii*). A preliminary RT-qPCR analysis was performed using cells grown under a 16 h light/8 h dark regimen (LD) and collected in the middle of the phases (L8 and D4), in medium with combined nitrogen (N^+) or without combined nitrogen (N^-) for *Lyngbya* and *Nostoc*, while in *Synechocystis* only N^+ medium was used (non-N_2-fixing strain). The six candidate reference genes with lowest Cq variation were selected for further studies in each organism: *rrn16S* (16S), *rnpB* and *secA* (tested in all organisms), *ilvD* (*Nostoc*), *petB* (*Nostoc* and *Synechocystis*), *ppc* (*Lyngbya* and *Synechocystis*), *purC* (*Lyngbya*), *rnpA* (*Lyngbya* and *Nostoc*) and *rpoA* (*Synechocystis*). In the second RT-qPCR analysis, samples collected in continuous light (CL) were included and four time points were tested in all conditions ($CL.N^+$, $CL.N^-$, $LD.N^+$ and $LD.N^-$).

Amplicon specificity, RT-qPCR efficiencies and analysis of Cq values

The size of the amplicons was confirmed by agarose gel electrophoresis (one sample per primer pair per organism, see Fig. S1) and their specificity confirmed by DNA sequencing. In all the RT-qPCR runs, standard curves were included (using cDNA serial dilutions) and the amplification efficiencies (E) obtained were in the recommended range (90–110%), except for the *rpoA* gene (*Synechocystis*) in which E was slightly higher (117.4%, see Table 2). The candidate reference genes exhibited desirable E values with acceptable R^2 values demonstrating linearity and reproducibility of the reactions across organisms and variety of conditions tested. The melting curves showed single peaks corresponding to unique amplicons, and in the no template controls (NTCs) no amplification or non-specific products (unstable at the specific amplicon melting temperature) were detected (Table 2). The parameters derived from the RT-qPCR analysis required by the MIQE guidelines are listed in Table 2.

The raw quantification cycle (Cq) values were extracted from the iQ5 Optical System Software v2.1 (Bio-Rad) and represented by box-and-whiskers plots (Fig. S2). The median Cq value for the 16S was the lowest in all organisms ($Cq \approx 20$), which was expected since it is one of the most abundant transcripts. For the other genes, similar median Cq values were observed in *Lyngbya* ($Cq \approx 30$), while for *Nostoc* and *Synechocystis* larger variation was found ($23 < Cq \leq 36$). The use of 16S as reference gene is controversial since the degradation machinery may not affect rRNA and mRNA in the same manner [23–24], and it is also recommended that reference and target genes should have similar Cq values [25]. The latter issue can be circumvented by ensuring that the calculations are performed within the linear range of the amplification curves of reference and target genes [25].

Assessing gene-stability and minimum number of reference genes required

Gene expression stability was analyzed using three distinct algorithms: geNorm [16], NormFinder [18] and BestKeeper [17]. These algorithms rank the candidate reference genes according to the calculated gene expression stability values (*M*, geNorm and NormFinder) or Pearson's correlation coefficients (*r*, BestKeeper) (see Material and Methods for further details). Two types of analyses were performed to assess the stability: (*i*) under each growth condition ($CL.N^+$, $CL.N^-$, $LD.N^+$ and $LD.N^-$) pooling the data corresponding to four collection time points, or (*ii*) under each light regimen (CL and LD) pooling the data corresponding to eight collection time points (four in each medium). For each time point, data from each biological and technical replicates were considered. The analysis of reference genes stability using the different algorithms revealed similar gene rankings for all organisms under the different conditions (Tables S1, S2 and S3). Furthermore, the optimal number of reference genes for each organism in each specific experimental condition was determined using $geNorm^{PLUS}$, which calculates the pair-wise variation – V_n/

Table 1. Candidate reference genes evaluated in this study.

Approach	Gene	Description	Reference
Bibliographic search	*rrn16S*	16S ribosomal RNA	[20,31–37]
	petB	Cytochrome b_6, involved in electron transport and ATP generation	[38]
	ppc	Phospho*enol*pyruvate carboxylase (PEPC), central enzyme in carbon concentrating mechanism	[39]
	rnpA	Protein subunit of ribonuclease P (RNase P)	[38]
	rnpB	RNA subunit of ribonuclease P	[40–44]
	rps1B	Small subunit ribosomal protein S1	[42,44]
	rpoA	RNA polymerase, alpha subunit	[45–46]
CyanoBIKE*	*ilvD*	Dihydroxyacid dehydratase, participates in the biosynthetic process of branched chain amino acid family	This study
	mrp	Multiple drug resistance protein (MRP) homolog	This study
	prsA	Ribose-phosphate pyrophosphokinase, involved in the pentose phosphate pathway	This study
	purC	SAICAR synthetase, involved in the *de novo* purine biosynthetic pathway	This study
	secA	Part of the Sec protein translocase complex	This study

*Search based on the works by Ehira and Ohmori [20] and Hihara *et al.* [21].

doi:10.1371/journal.pone.0034983.t001

Table 2. Amplicon sizes and parameters derived from RT-qPCR data analysis.

Organism	Genes	Amplicon size (bp)	Amplicon T_m (°C)	Median Cq	NTC* (Cq)	Amplification efficiency	R^2	Slope	y interception
***Lyngbya aestuarii* CCY 9616**	*rrn16Sa.b*	191/192	86.50	21.035	31.12	101.3	0.972	−3.292	17.859
	ppc	90	82.00	30.93	43.51**	100.4	0.968	−3.313	29.319
	purC	112	80.00	32.72	35.30**	103.4	0.921	−3.243	29.339
	rnpA	133	81.50	30.895	35.34**	105.3	0.990	−3.201	28.23
	rnpB	189	86.00	32.79	N.d.	101.3	0.987	−3.291	29.104
	secA	131	83.00	31.92	53.68**	98.6	0.912	−3.355	26.701
***Nostoc* sp. PCC 7120**	*rrn16Sa.b.c.d*	191	83.50	17.865	31.83	105.4	0.981	−3.198	16.516
	ilvD	77	79.50	29.21	N.d.	99.7	0.978	−3.329	25.359
	petB	157	82.50	30.235	N.d.	94.3	0.988	−3.466	30.284
	rnpA	204	82.50	33.37	37.69**	97.5	0.980	−3.383	30.194
	rnpB	131	84.00	23.9	N.d.	100.3	0.994	−3.314	20.060
	secA	131	81.50	30.785	N.d.	99.5	0.988	−3.333	28.202
***Synechocystis* sp. PCC 6803**	*rrn16Sa.b*	190	86.00	18.425	28.84	100.4	0.981	−3.312	14.654
	ppc	108	83.50	34.4358	47.57**	95.9	0.960	−3.425	30.680
	petB	179	84.50	29.63	N.d.	101.5	0.958	−3.287	28.052
	rnpB	136	81.50	26.38	42.43**	103.8	0.982	−3.233	22.366
	rpoA	198	84.00	32.875	41.41**	117.4	0.991	−2.965	28.836
	secA	113	82.00	35.04	N.d.	102.1	0.958	−3.271	31.765

*No template control.
**Non-specific amplicons (unstable at the specific amplicon melting temperature).
N.d. – Not detected.
doi:10.1371/journal.pone.0034983.t002

V_{n+1}. This analysis showed that, in most of the conditions, the value of $V_{2/3}$ was below 0.15 indicating that the minimum number of reference genes required for normalization is two (Fig. 1). For *Lyngbya*, under the conditions CL, LD.N$^-$ and LD the pair-wise variation was always above the threshold and therefore the use of three genes for the first two conditions and five for latter is necessary (Fig. 1A). Taking into consideration the ranking of the three algorithms as well as the pair-wise analysis of geNorm^{PLUS}, a comprehensive selection of adequate reference genes is presented in Table 3. In all the conditions tested and in the three organisms, a minimum of three stable reference genes was always considered since it is more appropriate for a reliable normalization of data [16]. In this study the 16S gene was shown to be a stable reference gene in all organisms, and can be used for normalization in all the conditions tested in *Nostoc* (six conditions) and in *Synechocystis* (two conditions). The *rnpB* gene was stable in the three cyanobacteria, but the conditions in which it can be used are more limited. This first study on the validation of cyanobacterial reference genes identified 16S and *rnpB* as the most stable in the majority of the conditions tested; by coincidence these two genes were the more commonly used for normalization in previous cyanobacterial RT-qPCR studies (Table 1). The *rnpA* and *secA* genes can also be considered for data normalization in *Lyngbya* and *Nostoc* in a wide range of conditions. All the other reference genes listed in Table 3 are more strain-dependent: *ilvD* can be used in *Nostoc*, while *ppc* and *purC* are specific for *Lyngbya*, and *petB* is suggested for *Synechocystis* and *Nostoc*, but in the latter only in one condition (LD.N$^-$).

Choice of reference genes

To demonstrate the effects of the choice of a non-optimal reference gene, an expression analysis was simulated using data from *Nostoc* grown under a light/dark regimen. In this particular condition, 16S was identified as the most stable gene, followed by *secA*, and *rnpB* was identified as the least stable. Therefore, the simulation was performed considering *secA* as target and its expression was normalized against 16S or *rnpB* (Fig. 2A). The relative fold expression of *secA* was calculated and normalized to the lowest value: 1.88±0.98 in N$^+$ and 1.00±0.61 in N$^-$ medium. This normalization resulted in a significant difference in relative fold expression when comparing the two media ($P<0.0001$). The expression level of *secA* was calculated relative to 16S, and it was found to be stable in both media, $P=0.7157$ (1.08±0.77 in N$^+$ medium and 1.00±1.00 in N$^-$ media). In contrast, when *rnpB* was used as reference the relative expression of *secA* varied between 1.00±0.66 and 1.77±1.53 in N$^+$ and N$^-$ media, respectively, with a significant P value of 0.0112. Consequently, the difference between gene expression levels of *secA* is an artifact caused by the wrong choice of reference gene. This finding is supported by the results of the gene expression stability for *secA* (see above). A second analysis was performed taking *rnpB* as target and 16S and *secA* as reference genes (Fig. 2B). The *rnpB* relative fold expression was also calculated and normalized to the lowest value: 3.32±1.33 in N$^+$ and 1.00±0.61 and N$^-$ medium ($P<0.0001$). Subsequently, *rnpB* expression was normalized to 16S, *secA* or both genes generating similar results: *rnpB* expression was always significantly lower in N$^-$ medium. Overall, this analysis confirms the expression stability of 16S and *secA* genes and the inaccuracy of

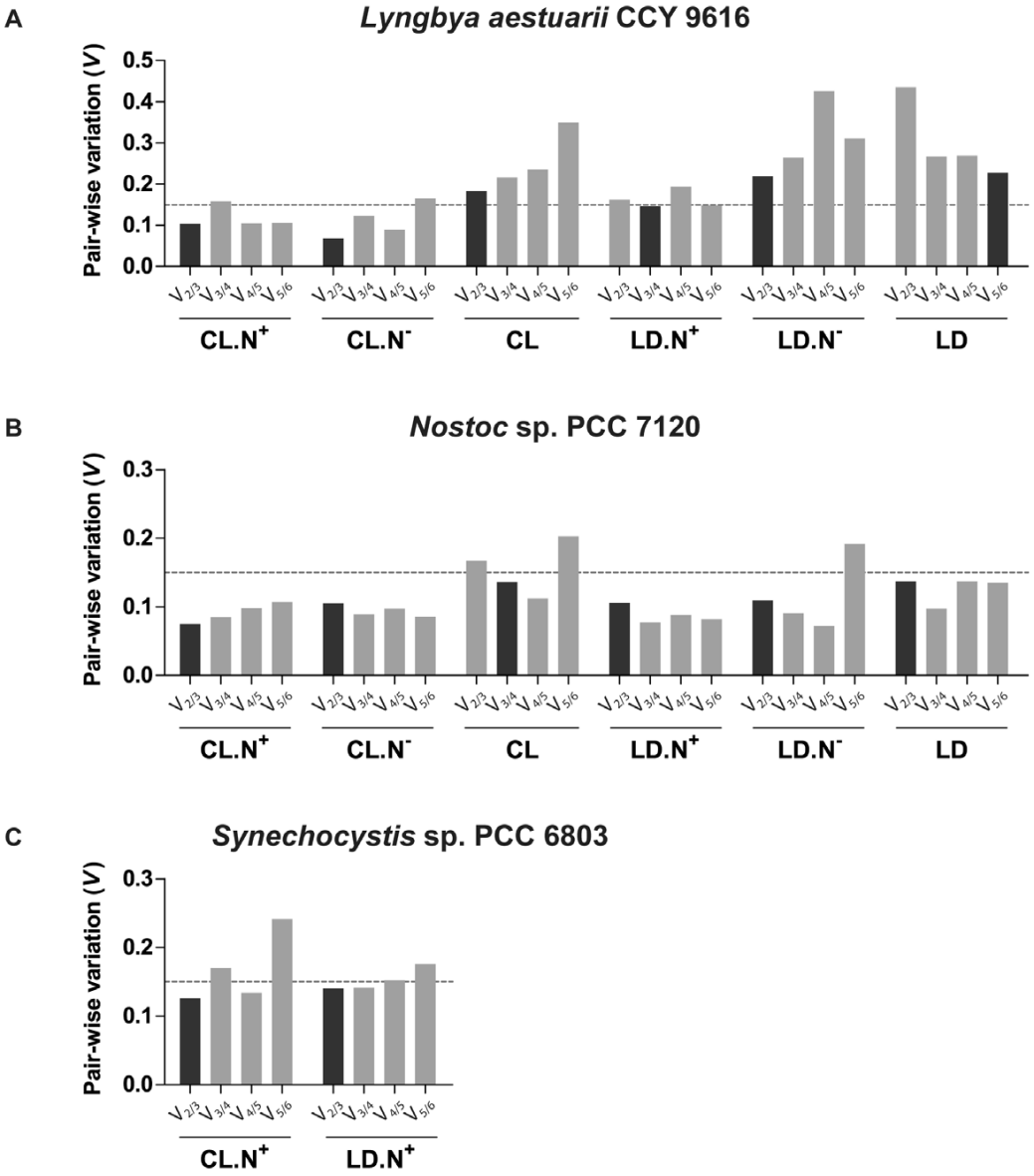


Figure 1. Pair-wise variation V_n/V_{n+1} calculated by geNorm^{PLUS} for *Lyngbya* (A), *Nostoc* (B) and *Synechocystis* (C). Dark-grey bars indicate the minimum number of reference genes required for accurate normalization.
doi:10.1371/journal.pone.0034983.g001

using the *mpB* gene as a reference, emphasizing that the choice of non-optimal reference genes leads to data misinterpretation.

Conclusions

This is the first study on validation of reference genes from cyanobacteria and, although a "universal" set could not be identified, a list of recommended genes is provided for the normalization of RT-qPCR data. The expression of a set of candidate reference genes was analyzed in three morphologically distinct cyanobacteria grown under different conditions. The different algorithms used to assess expression stability did not rank the candidate genes in the same order, but in general identified the same as the most stable. This reinforces not only the need to use more than one algorithm for accurate validation, but also the requirement of normalizing data against three or more reference genes. The array of experimental conditions tested makes this work a valuable starting point in the choice of reference genes when studying cyanobacteria.

Materials and Methods

Strains, culture conditions and sample collection

Three cyanobacterial strains, representing different sections among cyanobacteria, were used: the unicellular *Synechocystis* sp. PCC 6803 (Pasteur Culture Collection of Cyanobacteria, Paris, France), the filamentous non-heterocystous *Lyngbya aestuarii* CCY

Table 3. Recommended reference genes for *Lyngbya*, *Nostoc* and *Synechocystis*, under the conditions tested.

Organism	Condition*	Candidate reference genes**								
		16S	***ilvD***	***petB***	***ppc***	***purC***	***rnpA***	***rnpB***	***rpoA***	***secA***
***Lyngbya aestuarii* CCY 9616**	$CL.N^+$	−	ns	ns	+	+	−	−	ns	+
	$CL.N^-$	−	ns	ns	+	−	+	−	ns	+
	CL***	−	ns	ns	−	+	+	−	ns	+
	$LD.N^+$	+	ns	ns	+	−	+	−	ns	+
	$LD.N^-$	+	ns	ns	−	+	+	−	ns	−
	LD***	−	ns	ns	+	+	+	+	ns	+
***Nostoc* sp. PCC 7120**	$CL.N^+$	+	−	−	ns	ns	+	−	ns	+
	$CL.N^-$	+	−	−	ns	ns	+	+	ns	−
	CL***	+	+	−	ns	ns	−	+	ns	+
	$LD.N^+$	+	−	−	ns	ns	+	−	ns	+
	$LD.N^-$	+	+	+	ns	ns	−	−	ns	+
	LD***	+	−	−	ns	ns	+	−	ns	+
***Synechocystis* sp. PCC 6803**	$CL.N^+$	+	ns	+	−	ns	ns	+	−	−
	$LD.N^+$	+	ns	+	−	ns	ns	+	−	−

*CL – continuous light; LD – light/dark regimen; N^+ – medium with combined nitrogen; N^- – medium without combined nitrogen.
**Genes recommended (+), not recommended (−) or not selected after preliminary analysis (ns).
***Calculation performed pooling data from cells grown in both media and in the same light regimen.
doi:10.1371/journal.pone.0034983.t003

9616 (Culture Collection Yerseke, NIOO-KNAW, The Netherlands; co-identity *Lyngbya* sp. PCC 8106), and the filamentous heterocystous *Nostoc* sp. PCC 7120 (Pasteur Culture Collection of Cyanobacteria, Paris, France). *Lyngbya* was grown in artificial seawater ASN-III or ASN_0-III (ASN-III devoid of nitrate) [26], *Nostoc* was grown in BG11 or $BG11_0$ and *Synechocystis* only in BG11

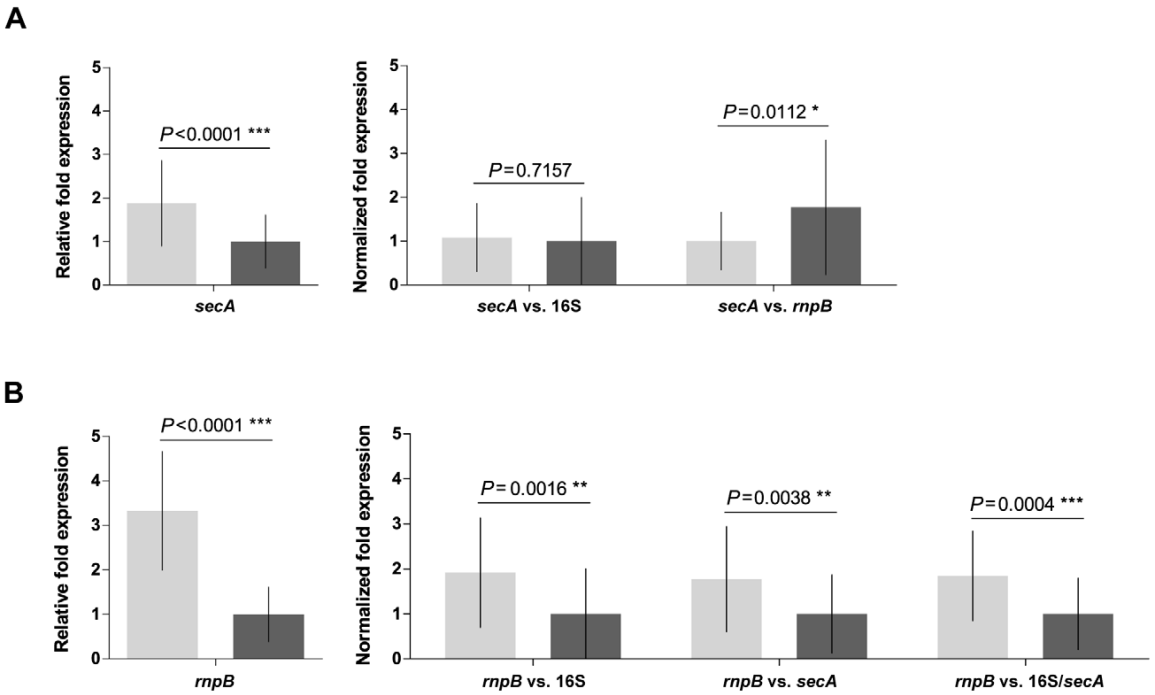


Figure 2. Effect of choice of reference gene(s). Relative (left) and normalized (right) fold expression of *secA* (A) and *rnpB* (B). Expression normalization was performed using data from the three biological replicates of *Nostoc* grown under a light/dark regimen, in N^+ (light-grey) or N^- (dark-grey) medium, using different genes as calibrators. Error bars indicate one standard error of the mean.
doi:10.1371/journal.pone.0034983.g002

[27]. The cyanobacterial cultures were kept at 25°C with agitation, on a continuous light (20 μmol photons m^{-2} s^{-1}, CL) or 16 h light (15 μmol photons m^{-2} s^{-1})/8 h dark (LD) regimen. For sample collection, three independent cultures of the organisms were grown for three to four days (under the different media and light conditions), at this stage the cells were changed to new medium and sample collection started after 24 h. For cultures in light/dark regimen, samples were collected in the transitions between the dark and the light phases (L0), in the middle of the light phase (eight hours into the light phase, L8), in the transition between the light and the dark phases (D0) and in the middle of the dark phase (four hours into the dark phase, D4). For continuous light, the first sample was collect 24 h after medium change (L0), and the remaining samples were collected after 8 h (L8), 16 h (L16) and 20 h (L20), time-points equivalent to the light/dark regimen sample collection. The mass or volumes of samples collected were 0.1–0.3 g for *Lyngbya*, 20 mL for *Nostoc* and 40 mL for *Synechocystis*. Cells were spun down (8 min at 4500 *g*, 4°C), resuspended in 0.5 mL medium and mixed with 1 mL RNAprotect® Bacterial reagent (Qiagen) and processed according to manufacturer's instructions, the cell pellets were frozen at −80°C until further use.

Escherichia coli strain DH5α (Stratagene) was used for cloning purposes. *E. coli* transformants were cultivated at 37°C in Luria-Bertani (LB) medium supplemented with 100 μg mL^{-1} ampicillin.

Candidate reference genes, primer design and amplicon specificity

The selection of the candidate reference genes was performed by bibliographic search, and using the CyanoBIKE server (http://biobike.csbc.vcu.edu:8003, accessed 2008) [22] selecting genes whose relative transcript levels varied between 0.7 and 1.3 but their difference in all conditions tested was below 0.5.

The sequences of the twelve selected genes were obtained from GenBank (accession number and location in Table S4). The primers were designed using the Beacon Designer 6 software (PREMIER Biosoft International) and synthesized (with cartridge purification) by STAB Vida (Lisbon). For each primer pair, the annealing temperature was optimized through gradient PCR, in reaction mixtures (20 μL) containing: 0.25 μM of each primer, 10 μL of iQ™ SYBR® Green Supermix (Bio-Rad) and 2 μL cDNA (1/9 dilution). The PCR profile was: 3 min at 95°C followed by 60 cycles of 30 s at 95°C, 30 s at 51, 53 or 56°C (Table 4 and Table S5) and 30 s at 72°C. A melting curve analysis was performed at the end of each PCR program, to exclude the formation of nonspecific products. After optimization, the size of the PCR products (for each gene and in each organism) was checked by agarose gel electrophoresis performed by standard protocols, using 1× TAE buffer [28]. These products were also cloned into the pGEM-T® Easy (Promega) vector according to the manufacturer's instructions, and the amplicon specificity was confirmed by DNA sequencing (STAB Vida).

RNA extraction and cDNA synthesis

For RNA extraction, the TRIzol® Reagent (Ambion) was used in combination with the Purelink™ RNA Mini Kit (Ambion). Briefly, the cells were disrupted in TRIzol containing 0.2 g of 0.2 mm-diameter glass beads (acid washed, Sigma) using a Mini-Beadbeater (Biospec Products) or a FastPrep®-24 (MP Biomedicals) and the following extraction steps were performed according to the manufacturer's instructions. RNA was quantified on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc.) and the quality and integrity was checked using the Experion™ RNA StdSens Analysis Kit (Bio-Rad). The RNA samples (≈300 ng) were treated with 1 U of RQ1 RNase-free DNase (Promega) according to manufacturer's instructions. The absence of genomic DNA contamination was checked by PCR, in reaction mixtures containing: 0.5 U of GoTaq Flexi DNA Polymerase (Promega), 1× GoTaq Flexi buffer, 200 μM of each deoxyribonucleotide triphosphate (dNTP), 1 μM of each BD16S primer (Table 4), and 2 μL total RNA. The PCR profile was: 2 min at 95°C followed by 25 cycles of 30 s at 95°C, 30 s at 51°C and 30 s at 72°C, and a final extension at 72°C for 7 min. The PCR reactions were run on agarose gel electrophoresis.

For cDNA synthesis, 150 ng of total RNA was transcribed with the iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad) in a final volume of 20 μL, following the manufacturer's instructions. A control PCR was performed using 1 μL of cDNA as template and the same reaction conditions and PCR program described above. Three-fold standard dilutions of the cDNAs were made (1/3, 1/9 and 1/27) and stored at −20°C.

Transcription analysis by Real-time RT-PCR

The RT-qPCRs were performed on iQ™ 96-well PCR plates covered with Optical Sealing Tape (Bio-Rad). Reactions were manually assembled and contained 0.25 μM of each primer, 10 μL of iQ™ SYBR® Green Supermix (Bio-Rad) and 2 μL of template cDNA (dilution 1/9). The PCR profile was: 3 min at 95°C followed by 60 cycles of 30 s at 95°C, 30 s at 51 or 56°C (Table 4) and 30 s at 72°C. Standard dilutions of the cDNA were used to check the relative efficiency and quality of primers. Negative controls (no template cDNA) were included and a melting curve analysis was performed in all assays. RT-qPCRs were performed with three biological replicates and technical triplicates/duplicates of each cDNA sample in the iCycler iQ5 Real-Time PCR Detection System (Bio-Rad). The data obtained were analyzed using the iQ5 Optical System Software v2.1 (Bio-Rad). Efficiency values were calculated and the Cq values for each data set were exported to a Microsoft Office Excel file, and imported into the qbase^{PLUS2} software (Biogazelle). The relative quantities of each sample were calculated using the gene-specific efficiency acquired from the dilutions series and normalized to the mean Cq value.

GraphPad Prism v5 (GraphPad Software, Inc.) was used to create the box-and-whisker plots. For each gene in a given condition, the Cq values of all technical replicates corresponding to the four collection time points were pooled together. Boxes correspond to Cq values within the 25th and 75th percentiles and the median is represented by an horizontal line. Whiskers include Cq values within the 10th and the 90th percentiles and Cq values outside this range (outliers) are represented as dots (Fig. S2).

Analyses of gene stability and number of optimal reference genes

Gene stability was assessed using three different mathematical algorithms: geNorm^{PLUS} [16], NormFinder [18], and BestKeeper [17]. The geNorm^{PLUS} algorithm determines internal control gene stability measure (M) as the average pair-wise variation of each reference gene with all the other reference genes, and enables the elimination of the least stable gene and the recalculation of the M values, resulting in the ranking of the most stable genes. The lower the M value, the higher is the gene stability; a good reference gene should have an M below 0.5 in homogeneous sample sets [15,29]. Furthermore, the minimum number of genes required for the calculation of a reliable normalization factor is determined, calculating the pair-wise variation V_n/V_{n+1} between two sequential normalization factors NF_n and NF_{n+1} (geometric mean of the expression levels of n reference genes). The stepwise inclusion of reference genes is performed until V_n/V_{n+1} drops below the recommended threshold of 0.15, when the benefit of using an

Table 4. Primer nucleotide sequences and annealing temperatures (T_a) used in RT-qPCR.

Organism	Genes	Primer name	Primer Sequence (5′→3′)	T_a (°C)
***Lyngbya aestuarii* CCY 9616**	*rrn16Sa.b*	BD16SF1	CACACTGGGACTGAGACAC	56
		BD16SR1	CTGCTGGCACGGAGTTAG	
	ppc	ppcF	TATCGCCAAGAACCCTATCG	56
		LppcR	GCTATTGTAAAGTCGCCAG	
	purC	purCF	GATACCTGTCGTTTGTGG	56
		LpurCR	GAACTTGTTGATAAGCCG	
	rnpA	rnpAF1	AACGTCAAATTCGAGCCG	56
		rnpAR1	AACAACTGCTCTAATTCTTGC	
	rnpB	LrnpBF	ATCATCTGTTGTTGGTAAGG	51
		LrnpBR	GAGGGCAATTATCTATCTGG	
	secA	secAF	AACCTACTACTACGACATCC	51
		secAR	ACTTAATCACCTGTTCTTTCAA	
***Nostoc* sp. PCC 7120**	*rrn16Sa.b.c.d*	BD16SF1	CACACTGGGACTGAGACAC	51
		BD16SR1	CTGCTGGCACGGAGTTAG	
	ilvD	ilvDF	GGCTGTGATAAGAATATGC	56
		ilvDR	CCACCGTAAACAAAGATAG	
	petB	petBF1	AGTTTGCCACTGGATTCG	56
		petBR1	AGAATCATCATCAACACCATC	
	rnpA	NrnpAF	TACGCTCATTGGTGTCTCG	56
		rnpAR1	AACAACTGCTCTAATTCTTGC	
	rnpB	NrnpBF2	TAGGGAGAGAGTAGGCGTTG	56
		NrnpBR2	TTCTGTGGCACTATCCTCAC	
	secA	secAF	AACCTACTACTACGACATCC	56
		secAR	ACTTAATCACCTGTTCTTTCAA	
***Synechocystis* sp. PCC 6803**	*rrn16Sa.b*	BD16SF1	CACACTGGGACTGAGACAC	51
		BD16SR1	CTGCTGGCACGGAGTTAG	
	ppc	ppcF	TATCGCCAAGAACCCTATCG	56
		SppcR	CATCGTTTGCCGTTCTTCTG	
	petB	SpetB1F	CCTTCGCCTCTGTCCAATAC	56
		SpetB1R	TAGCATTACACCCACAACCC	
	rnpB	rnpBF1	CGTTAGGATAGTGCCACAG	56
		rnpBR1	CGCTCTTACCGCACCTTTG	
	rpoA	SrpoAF	TTGGACTAATGGCAGTATTCA	56
		rpoAR1	CGCTTGAGACAGTTATAGGC	
	secA	secAF	AACCTACTACTACGACATCC	51
		SsecAR	TTAAATCCAAACCTTCCAG	

doi:10.1371/journal.pone.0034983.t004

extra gene (n+1) is limited [16,30]. Gene stability determination by geNorm^{PLUS} was performed using the qbase^{PLUS2} software v2.2 (Biogazelle).

NormFinder focuses on the expression variations both between different groups (inter-group variation) and inside one group (intra-group variation), in a model-based approach of mixed linear effect modeling. This algorithm determines inter- and intra-group variations and combines both results in a stability value for each investigated gene. According to NormFinder, genes with lowest M value will be top ranked. For the analysis with this algorithm, expression values were calculated using qbase^{PLUS2} and first exported to a Microsoft Office Excel file for use with the NormFinder applet.

BestKeeper calculates standard deviations (*SD*) and the coefficient of variance (*CV*) based on the Cq values of all candidate reference genes, considering the genes with *SD* lower than 1 as stably expressed. The algorithm calculates the BestKeeper index (*BI*) as the geometric mean of the stable reference genes Cq values and then establishes pair-wise correlations between each gene and the *BI*. The genes with highest Pearson correlation coefficient ($r \approx 1$) and probability (p) value lower than 0.05 are considered the most stable. The average Cq value of each technical triplicate/duplicate was calculated (without conversion to relative quantity) and imported to the Microsoft Excel based application BestKeeper to analyze gene stability.

Choice of reference genes

The gene expression simulation was performed using data from the three biological replicates of *Nostoc* grown under a light/dark

regimen, pooling the data from the different time collection points ($n = 32$). Relative and normalized fold expression values were calculated using the iQ5 Optical System Software v2.1 (Bio-Rad). Expression data were imported to GraphPad Prism v5 (GraphPad Software, Inc.) and the statistical significance of the results was assessed through the Student's t test.

MIQE guidelines

In this work, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [10] were followed to promote the effort for experimental consistency and transparency, and to increase the reliability and integrity of the results obtained.

Supporting Information

Figure S1 Confirmation of amplicon sizes for the selected candidate reference genes studied in each cyanobacteria. Agarose gel electrophoresis showing specific PCR products of the expected sizes for each candidate reference gene in *Lyngbya* (A), *Nostoc* (B) and *Synechocystis* (C). MM – GeneRuler^{TM} 50 bp DNA Ladder (Fermentas).
(TIF)

Figure S2 Box-and-whiskers plots of candidate reference gene Cq values in *Lyngbya* (A), *Nostoc* (B) and *Synechocystis* (C). Boxes correspond to Cq values within the 25th and 75th percentiles and the median is represented by an horizontal line. Whiskers include Cq values within the 10th and the 90th percentiles and Cq values outside this range (outliers) are represented as dots.
(TIF)

Table S1 Ranking of the candidate reference genes according to their stability value (M) calculated by geNorm.
(DOC)

Table S2 Ranking of the candidate reference genes according to their stability value (M) calculated by NormFinder.
(DOC)

Table S3 Ranking of the candidate reference genes according to the Pearson's correlation coefficient (r) calculated against the BestKeeper index and probability values (p).
(DOC)

Table S4 Candidate reference genes accession number and location.
(DOC)

Table S5 Primer nucleotide sequences of genes tested only in preliminary RT-qPCR analysis.
(DOC)

Acknowledgments

We acknowledge Paula Magalhães for technical support, Marta Mendes for excellent advice and suggestions and Carol Harley for reading the manuscript.

Author Contributions

Conceived and designed the experiments: FP CCP DF PT. Performed the experiments: FP CCP DF. Analyzed the data: FP CCP DF PT. Contributed reagents/materials/analysis tools: PMF PT. Wrote the paper: FP CCP DF PT.

References

1. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T (2006) The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proc Natl Acad Sci USA 103: 5442–5447.
2. Castenholz RW (2001) Phylum BX. Cyanobacteria. General characteristics of the cyanobacteria. In: Garrity GM, Boone DR, Castenholz RW, eds. Bergey's Manual of Systematic Bacteriology, Volume 1: The Archaea and the Deeply Branching and Phototropic Bacteria. 2nd ed. ed. New York: Springer-Verlag. pp 474–487.
3. Ducat DC, Way JC, Silver PA (2011) Engineering cyanobacteria to generate high-value products. Trends Biotechnol 29: 95–103.
4. Angermayr SA, Hellingwerf KJ, Lindblad P, Teixeira de Mattos MJ (2009) Energy biotechnology with cyanobacteria. Curr Opin Biotechnol 20: 257–263.
5. Burja AM, Dhamwichukorn S, Wright PC (2003) Cyanobacterial postgenomic research and systems biology. Trends Biotechnol 21: 504–511.
6. Bustin SA (2010) Why the need for qPCR publication guidelines? - The case for MIQE. Methods 50: 217–226.
7. VanGuilder HD, Vrana KE, Freeman WM (2008) Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44: 619–626.
8. Bustin S, Beaulieu J-F, Huggett J, Jaggi R, Kibenge F, et al. (2010) MIQE précis: practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol Biol 11: 74.
9. Derveaux S, Vandesompele J, Hellemans J (2010) How to do successful gene expression analysis using real-time PCR. Methods 50: 227–230.
10. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611–622.
11. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, et al. (1999) Housekeeping genes as internal standards: use and limits. J Biotechnol 75: 291–295.
12. Suzuki T, Higgins PJ, Crawford DR (2000) Control selection for RNA quantitation. BioTechniques 29: 332–337.
13. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25: 169–193.
14. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, et al. (2004) Validation of housekeeping genes for normalizing RNA expression in real-time PCR. BioTechniques 37: 112–114, 116, 118–119.
15. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M (2010) A practical approach to RT-qPCR - Publishing data that conform to the MIQE guidelines. Methods 50: S1–S5.
16. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: research0034.
17. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnol Lett 26: 509–515.
18. Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64: 5245–5250.
19. Silver N, Best S, Jiang J, Thein S (2006) Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol Biol 7: 33.
20. Ehira S, Ohmori M (2006) NrrA directly regulates expression of *hetR* during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. J Bacteriol 188: 8520–8525.
21. Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. Plant Cell 13: 793–806.
22. Elhai J, Taton A, Massar J, Myers JK, Travers M, et al. (2009) BioBIKE: a Web-based, programmable, integrated biological knowledge base. Nucleic Acids Res 37: W28–W32.
23. Solanas M, Moral R, Escrich E (2001) Unsuitability of using ribosomal RNA as loading control for Northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. Anal Biochem 288: 99–102.
24. Ferguson BS, Nam H, Hopkins RG, Morrison RF (2010) Impact of reference gene selection for target gene normalization on experimental outcome using real-time qRT-PCR in adipocytes. PLoS ONE 5: e15208.
25. Hruz T, Wyss M, Docquier M, Pfaffl M, Masanetz S, et al. (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 12: 156.
26. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111: 1–61.
27. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G (1971) Purification and properties of unicellular blue-green algae (order Chroococcales). Bacteriol Rev 35: 171–205.
28. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. New York, USA: Cold Spring Harbor Laboratory Press.

29. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19.
30. Biogazelle website. geNorm^{PLUS} manual available: http://www.biogazelle.com/files/manual/genormPLUS.pdf. Accessed 2011 Aug 20.
31. Doblin MA, Coyne KJ, Rinta-Kanto JM, Wilhelm SW, Dobbs FC (2007) Dynamics and short-term survival of toxic cyanobacteria species in ballast water from NOBOB vessels transiting the Great Lakes - implications for HAB invasions. Harmful Algae 6: 519–530.
32. El-Shehawy R, Lugomela C, Ernst A, Bergman B (2003) Diurnal expression of *hetR* and diazocyte development in the filamentous non-heterocystous cyanobacterium *Trichodesmium erythraeum*. Microbiology 149: 1139–1146.
33. Higo A, Katoh H, Ohmori K, Ikeuchi M, Ohmori M (2006) The role of a gene cluster for trehalose metabolism in dehydration tolerance of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. Microbiology 152: 979–987.
34. Huang X, Dong Y, Zhao J (2004) HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. Proc Natl Acad Sci USA 101: 4848–4853.
35. Imamura S, Tanaka K, Shirai M, Asayama M (2006) Growth phase-dependent activation of nitrogen-related genes by a control network of group 1 and group 2 σ factors in a cyanobacterium. J Biol Chem 281: 2668–2675.
36. Rinta-Kanto JM, Ouellette AJA, Boyer GL, Twiss MR, Bridgeman TB, et al. (2005) Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. Environ Sci Technol 39: 4198–4205.
37. Schäfer L, Sandmann M, Woitsch S, Sandmann G (2006) Coordinate up-regulation of carotenoid biosynthesis as a response to light stress in *Synechococcus* PCC 7942. Plant, Cell Environ 29: 1349–1356.
38. Price GD, Woodger FJ, Badger MR, Howitt SM, Tucker L (2004) Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proc Natl Acad Sci USA 101: 18228–18233.
39. Woodger FJ, Badger MR, Price GD (2003) Inorganic carbon limitation induces transcripts encoding components of the CO_2-concentrating mechanism in *Synechococcus* sp. PCC 7942 through a redox-independent pathway. Plant Physiol 133: 2069–2080.
40. Fujimori T, Higuchi M, Sato H, Aiba H, Muramatsu M, et al. (2005) The mutant of *sll1961*, which encodes a putative transcriptional regulator, has a defect in regulation of photosystem stoichiometry in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Physiol 139: 408–416.
41. Holtzendorff J, Marie D, Post AF, Partensky F, Rivlin A, et al. (2002) Synchronized expression of *ftsZ* in natural *Prochlorococcus* populations of the Red Sea. Environ Microbiol 4: 644–653.
42. Mary I, Tu C-J, Grossman A, Vaulot D (2004) Effects of high light on transcripts of stress-associated genes for the cyanobacteria *Synechocystis* sp. PCC 6803 and *Prochlorococcus* MED 4 and MIT 9313. Microbiology 150: 1271–1281.
43. Mary I, Vaulot D (2003) Two-component systems in *Prochlorococcus* MED 4: genomic analysis and differential expression under stress. FEMS Microbiol Lett 226: 135–144.
44. Tu C-J, Shrager J, Burnap RL, Postier BL, Grossman AR (2004) Consequences of a deletion in *dspA* on transcript accumulation in *Synechocystis* sp. strain PCC 6803. J Bacteriol 186: 3889–3902.
45. Lemeille S, Geiselmann J, Latifi A (2005) Crosstalk regulation among group 2-Sigma factors in *Synechocystis* PCC 6803. BMC Microbiol 5: 18.
46. Lemeille S, Latifi A, Geiselmann J (2005) Inferring the connectivity of a regulatory network from mRNA quantification in *Synechocystis* PCC 6803. Nucleic Acids Res 33: 3381–3389.

Supplementary material IV

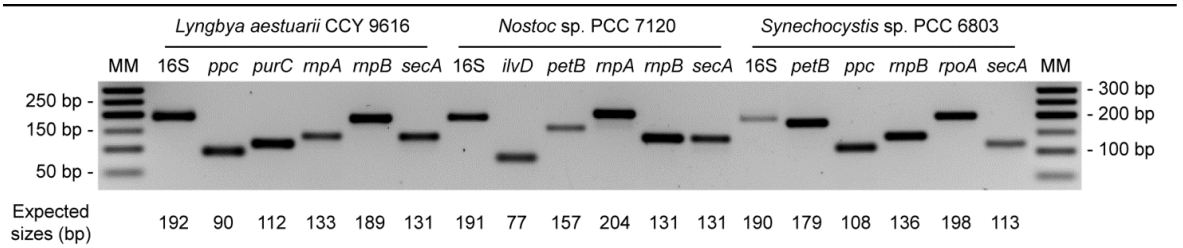


Figure S1. Confirmation of amplicon sizes for the selected candidate reference genes studied in each cyanobacteria. Agarose gel electrophoresis showing specific PCR products of the expected sizes for each candidate reference gene in Lyngbya (A), Nostoc (B) and Synechocystis (C). MM – GeneRuler™ 50 bp DNA Ladder (Fermentas).

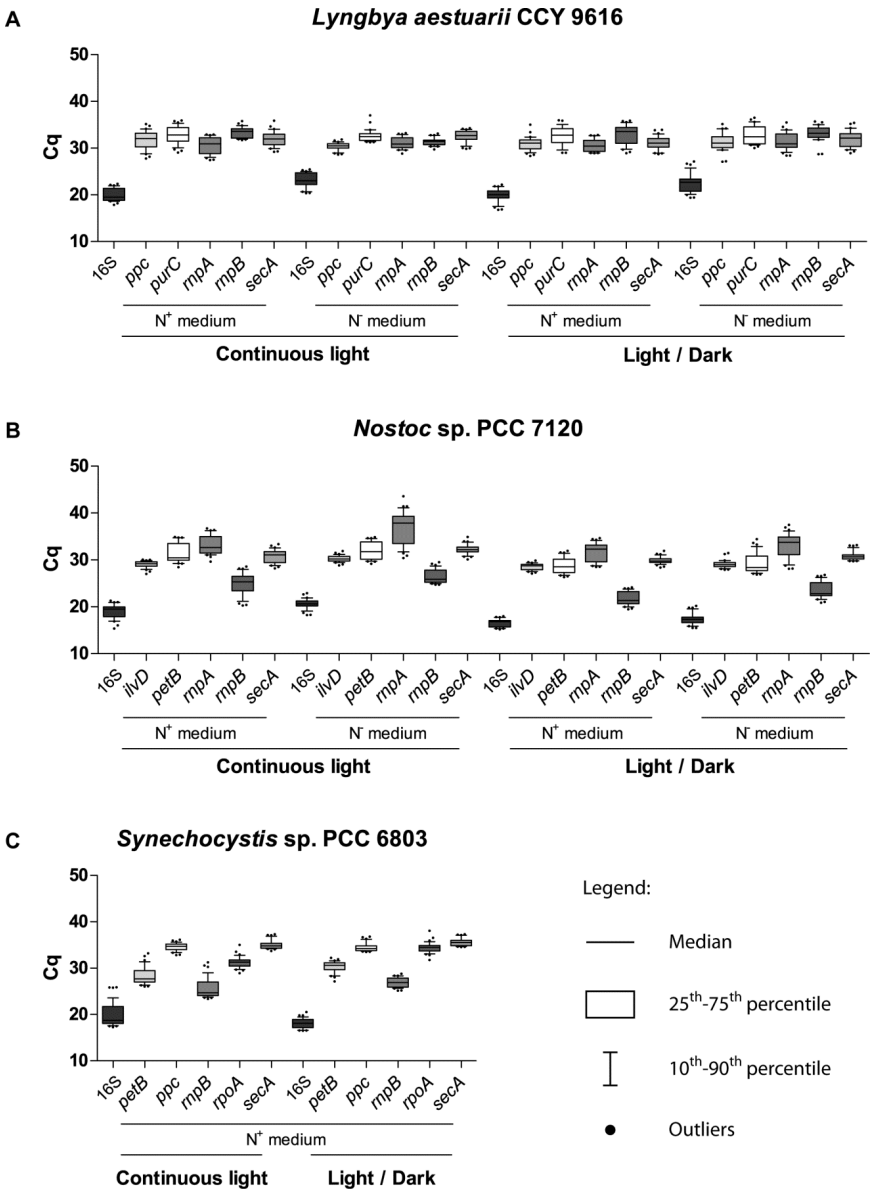


Figure S2. Box-and-whiskers plots of candidate reference gene Cq values in Lyngbya (A), Nostoc (B) and Synechocystis (C). Boxes correspond to Cq values within the 25th and 75th percentiles and the median is represented by an horizontal line. Whiskers include Cq values within the 10th and the 90th percentiles and Cq values outside this range (outliers) are represented as dots.

Table S1. Ranking of the candidate reference genes according to their stability value (M) calculated by geNorm.

Organism	Condition*		Ranking (less stable to more stable)**					
			6	5	4	3	2	1
***Lyngbya aestuarii* CCY 9616**	CL.N$^+$		*rnpB*	*rnpA*	16S	*purC*	***ppc***	***secA***
		M	0.572	0.512	**0.453**	**0.260**	**0.231**	**0.208**
	CL.N$^-$		16S	*rnpB*	*purC*	*rnpA*	***secA***	***ppc***
		M	0.609	**0.405**	**0.337**	**0.171**	**0.153**	**0.137**
	CL***		16S	*rnpB*	*ppc*	***purC***	***rnpA***	***secA***
		M	1.329	0.929	0.728	0.545	0.521	0.505
	LD.N$^+$		*rnpB*	*purC*	*ppc*	**16S**	***rnpA***	***secA***
		M	0.789	0.700	**0.487**	**0.376**	**0.318**	**0.316**
	LD.N$^-$		*ppc*	*rnpB*	*secA*	**16S**	***rnpA***	***purC***
		M	1.572	1.375	0.854	0.604	0.561	0.543
	LD***		16S	***secA***	***purC***	***rnpA***	***rnpB***	***ppc***
		M	1.377	1.292	1.188	1.148	1.045	1.024
***Nostoc* sp. PCC 7120**	CL.N$^+$		*rnpB*	*petB*	*ilvD*	*secA*	**16S**	***rnpA***
		M	**0.475**	**0.377**	**0.282**	**0.212**	**0.200**	**0.193**
	CL.N$^-$		16S	*secA*	*rnpB*	***petB***	***rnpA***	***ilvD***
		M	**0.437**	**0.433**	**0.377**	**0.350**	**0.346**	**0.343**
	CL***		*rnpA*	*petB*	***ilvD***	***rnpB***	**16S**	***secA***
		M	0.799	0.569	0.520	**0.438**	**0.399**	**0.363**
	LD.N$^+$		*petB*	*ilvD*	*rnpB*	***secA***	**16S**	***rnpA***
		M	**0.434**	**0.380**	**0.318**	**0.285**	**0.264**	**0.226**
	LD.N$^-$		*rnpA*	*rnpB*	**16S**	***secA***	***ilvD***	***petB***
		M	0.647	**0.385**	**0.361**	**0.324**	**0.310**	**0.300**
	LD***		*rnpA*	*rnpB*	*ilvD*	***petB***	***secA***	**16S**
		M	0.661	0.552	**0.439**	**0.427**	**0.415**	**0.395**
***Synechocystis* sp. PCC 6803**	CL.N$^+$		*secA*	*ppc*	*rpoA*	16S	***rnpB***	***petB***
		M	0.893	0.600	0.510	**0.331**	**0.300**	**0.298**
	LD.N$^+$		*rpoA*	*secA*	*ppc*	*rnpB*	***petB***	**16S**
		M	0.787	0.636	0.520	**0.436**	**0.424**	**0.400**

*CL – continuous light; LD – light/dark regimen; N$^+$ – medium with combined nitrogen; N$^-$ – medium without combined nitrogen.

**The best combination of genes obtained in the pair-wise analysis and the *M* values below the 0.5 threshold are in bold.

***Calculation performed pooling data from cells grown in both media and in the same light regimen.

Table S2. Ranking of the candidate reference genes according to their stability value (*M*) calculated by NormFinder.

Organism	Condition*		Ranking (less stable to more stable)**					
			6	5	4	3	2	1
***Lyngbya aestuarii* CCY 9616**	CL.N$^+$		16S	*rnpB*	*purC*	*rnpA*	*secA*	ppc
		M	1.797	0.662	0.249	0.131	0.068	0.068
	CL.N$^-$		*rnpB*	*ppc*	*purC*	16S	*rnpA*	*secA*
		M	0.612	0.579	0.348	0.329	0.084	0.040
	CL***		16S	*rnpB*	*ppc*	***purC***	***rnpA***	*secA*
		M	1.812	1.417	0.744	0.300	0.197	0.181
	LD.N$^+$		*rnpB*	*purC*	*ppc*	*rnpA*	16S	*secA*
		M	0.593	0.534	0.371	0.339	0.305	0.282
	LD.N$^-$		*secA*	*ppc*	*rnpB*	*purC*	*rnpA*	16S
		M	1.805	1.501	0.979	0.232	0.128	0.128
	LD***		16S	*ppc*	*secA*	*rnpB*	***purC***	***rnpA***
		M	0.980	0.688	0.669	0.503	0.397	0.258
***Nostoc* sp. PCC 7120**	CL.N$^+$		*rnpB*	16S	*petB*	*secA*	*ilvD*	*rnpA*
		M	0.325	0.207	0.195	0.179	0.170	0.166
	CL.N$^-$		*secA*	*ilvD*	*rnpA*	16S	*petB*	*rnpB*
		M	0.116	0.079	0.068	0.062	0.044	0.027
	CL***		*rnpA*	*petB*	*rnpB*	**16S**	***secA***	*ilvD*
		M	0.380	0.286	0.186	0.182	0.167	0.098
	LD.N$^+$		*petB*	*rnpB*	*ilvD*	*secA*	16S	*rnpA*
		M	1.311	1.027	0.795	0.581	0.493	0.170
	LD.N$^-$		*rnpA*	*ilvD*	*petB*	16S	*secA*	*rnpB*
		M	1.314	0.658	0.576	0.531	0.385	0.254
	LD***		*rnpB*	*ilvD*	*petB*	***rnpA***	***secA***	16S
		M	1.227	0.923	0.864	0.702	0.634	0.385
***Synechocystis* sp. PCC 6803**	CL.N$^+$		*rpoA*	*secA*	*petB*	*ppc*	*rnpB*	16S
		M	1.866	1.317	0.812	0.274	0.244	0.084
	LD.N$^+$		16S	*rpoA*	*secA*	*ppc*	*rnpB*	*petB*
		M	0.499	0.453	0.411	0.319	0.264	0.145

*CL – continuous light; LD – light/dark regimen; N$^+$ – medium with combined nitrogen; N$^-$ – medium without combined nitrogen.
**The best pair of genes comparing two different groups (N$^+$ vs. N$^-$) are in bold.
***Calculation performed pooling data from cells grown in both media and in the same light regimen.

Table S3. Ranking of the candidate reference genes according to the Pearson's correlation coefficient (r) calculated against the BestKeeper index and probability values (p).

Organism	Condition*		Ranking (less stable to more stable)**					
			6	5	4	3	2	1
***Lyngbya aestuarii* CCY 9616**	CL.N$^+$		*rnpB*	*secA*	*rnpA*	*purC*	16S	*ppc*
		r	0.732	0.802	0.871	0.883	0.905	0.957
		p-Value	**0.007**	**0.002**	**0.001**	**0.001**	**0.001**	**0.001**
	CL.N$^-$		*rnpB*	*purC*	*rnpA*	16S	*secA*	*ppc*
		r	0.112	0.423	0.584	0.693	0.699	0.713
		p-Value	0.726	0.170	**0.047**	**0.012**	**0.011**	**0.009**
	CL***		*rnpB*	16S	*purC*	*ppc*	*secA*	*rnpA*
		r	0.147	0.651	0.731	0.761	0.781	0.809
		p-Value	0.491	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	LD.N$^+$		*rnpA*	*purC*	*secA*	16S	*rnpB*	*ppc*
		r	0.323	0.455	0.755	0.780	0.847	0.861
		p-Value	0.306	0.136	**0.005**	**0.003**	**0.001**	**0.001**
	LD.N$^-$		*secA*	*rnpB*	*ppc*	*purC*	16S	*rnpA*
		r	0.370	0.485	0.606	0.727	0.727	0.826
		p-Value	0.236	0.111	**0.037**	**0.007**	**0.007**	**0.001**
	LD***		*purC*	*secA*	*rnpB*	*rnpA*	*ppc*	16S
		r	0.569	0.579	0.614	0.656	0.665	0.772
		p-Value	**0.004**	**0.003**	**0.001**	**0.001**	**0.001**	**0.001**
***Nostoc* sp. PCC 7120**	CL.N$^+$		*ilvD*	*petB*	*rnpB*	*rnpA*		16S/*secA*
		r	0.917	0.940	0.947	0.950		0.991
		p-Value	**0.001**	**0.001**	**0.001**	**0.001**		**0.001**
	CL.N$^-$		*petB*	*secA*	*ilvD*	*rnpA*	*rnpB*	16S
		r	0.198	0.684	0.797	0.813	0.846	0.942
		p-Value	0.537	**0.014**	**0.002**	**0.001**	**0.001**	**0.001**
	CL***		*petB*	*rnpA*	*ilvD*	*secA*	*rnpB*	16S
		r	0.568	0.643	0.810	0.861	0.906	0.970
		p-Value	**0.004**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	LD.N$^+$		*petB*	*ilvD*	*rnpB*	*rnpA*	16S	*secA*
		r	0.491	0.547	0.548	0.653	0.936	0.951
		p-Value	0.105	0.065	0.065	**0.021**	**0.001**	**0.001**
	LD.N$^-$		*rnpB*	*petB*	*rnpA*	*ilvD*	*secA*	16S
		r	0.895	0.954	0.958	0.958	0.973	0.994
		p-Value	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	LD***		*rnpB*	*petB*	*ilvD*	*rnpA*	*secA*	16S
		r	0.850	0.865	0.890	0.927	0.941	0.985
		p-Value	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
***Synechocystis* sp. PCC 6803**	CL.N$^+$		*secA*	*ppc*	*rpoA*	*petB*	*rnpB*	16S
		r	-0.356	0.702	0.894	0.952	0.987	0.989
		p-Value	0.258	0.011	0.001	0.001	0.001	0.001
	LD.N$^+$		*secA*	*rpoA*	*ppc*	*petB*	16S	*rnpB*
		r	0.149	0.192	0.725	0.899	0.923	0.935
		p-Value	0.642	0.549	**0.008**	**0.001**	**0.001**	**0.001**

*CL – continuous light; LD – light/dark regimen; N$^+$ – medium with combined nitrogen; N$^-$ – medium without combined nitrogen.
**Significant *p*-Values are in bold.
***Calculation performed pooling data from cells grown in both media and in the same light regimen.

Table S4. Candidate reference genes accession number and location.

Organism	Genes	Accession number (nucleotides, locus tag)
***Lyngbya aestuarii* CCY 9616**	*ilvD*	NZ_AAVU01000015.1 (64391 to 66073*, L8106_26847)
	mrp	NZ_AAVU01000008.1 (141808 to 142878, L8106_17697)
	petB	NZ_AAVU01000005.1 (107734 to 108402*, L8106_09696)
	ppc	NZ_AAVU01000013.1 (80928 to 84044, L8106_19036)
	prsA	NZ_AAVU01000001.1 (11840 to12856*, L8106_04291)
	purC	NZ_AAVU01000001.1 (122582 to 123316, L8106_04766)
	rnpA	NZ_AAVU01000059.1 (974 to 1381, L8106_29250)
	rnpB	NZ_AAVU01000017.1 (76166 to 76621*, N/A)
	rpoA	NZ_AAVU01000003.1 (146690 to 147634*, L8106_15090)
	rps1B	NZ_AAVU01000005.1 (123246 to 124133*, L8106_09766)
	rrn16Sa	NZ_AAVU01000008.1 (780 to 2269*, N/A)
	rrn16Sb	NZ_AAVU01000018.1 (81295 to 82783, N/A)
	secA	NZ_AAVU01000009.1 (97343 to 100138*, L8106_16729)
***Nostoc* sp. PCC 7120**	*ilvD*	NC_003272.1 (3367200 to 3368891, alr2771)
	mrp	NC_003272.1 (754691 to 755761,alr0652)
	petB	NC_003272.1 (4130387 to 4131034, alr3421)
	ppc	NC_003272.1 (5788301 to 5791249*,all4861)
	prsA	NC_003272.1 (5570574 to 5571566,alr4670)
	purC	NC_003272.1 (2730420 to 2731157,alr2268)
	rnpA	NC_003272.1 (4122806 to 4123228, alr3413)
	rnpB	NC_003272.1 (4950355 to 4950814*, allrs04)
	rpoA	NC_003272.1 (5038362 to 5039309*,all4191)
	rps1B	NC_003272.1 (1261663 to 1262580,alr1078)
	rrn16Sa	NC_003272.1 (2375734 to 2377222, allrr01)
	rrn16Sb	NC_003272.1 (2500525 to 2502013, allrr04)
	rrn16Sc	NC_003272.1 (4918283 to 4919771*, allrr09)
	rrn16Sd	NC_003272.1 (5945700 to 5947188*, allrr12)
	secA	NC_003272.1 (5778711 to 5781503, alr4851)
***Synechocystis* sp. PCC 6803**	*ilvD*	NC_000911.1 (3498903 to 3500588, slr0452)
	mrp	NC_000911.1 (2596787 to 2597848, slr0067)
	petB	NC_000911.1 (2428010 to 2428678, slr0342)
	ppc	NC_000911.1 (1994147 to 1997251*, sll0920)
	prsA	NC_000911.1 (2938630 to 2939631*, sll0469)
	purC	NC_000911.1 (1030195 to 1030989, slr1226)
	rnpA	NC_000911.1 (1826911 to 1827285, slr1469)
	rnpB	NC_000911.1 (153057 to 153655, 6803s01)
	rpoA	NC_000911.1 (829141 to 830085*, sll1818)
	rps1B	NC_000911.1 (1815454 to 1816371, slr1984)
	rrn16Sa	NC_000911.1 (2452187 to 2453675*, 6803r03)
	rrn16Sb	NC_000911.1 (3325053 to 3326541, 6803r04)
	secA	NC_000911.1 (2647386 to 2650184*, sll0616)

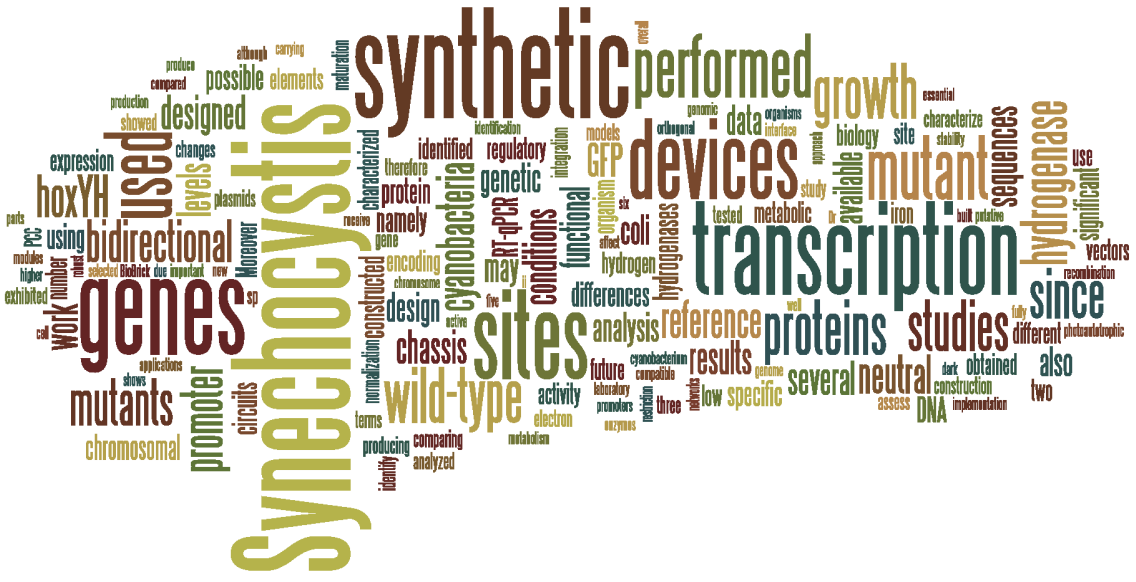
*Complement sequence.
N/A – not-available.

Table S5. Primer nucleotide sequences of genes tested only in preliminary RT-qPCR analysis.

Organism	Genes	Primer name	Primer Sequence (5' → 3')	T_a (ºC)
***Lyngbya aestuarii* CCY 9616**	*ilvD*	ilvDF	GGCTGTGATAAGAATATGC	51
		ilvDR	CCACCGTAAACAAAGATAG	
	mrp	mrpF	GAACTGAATATGATTCGCAACG	51
		LmrpR	ACAGGCGGGAGTGGTGAG	
	petB	LpetB1F	GTATGATGGTATTGATGATGATTC	51
		LpetB1R	AGGTATGTAAGGTGTAGAAGC	
	prsA	prsAF	TTAAATGTGATTGGTGATGT	56
		LprsAR	AGTCGCACAAGCATAAAC	
	rpoA	LrpoAF	ATATCTGGACGAATGGTA	51
		LrpoAR	CGTTTGAGACAGTTATAGG	
	rps1B	Lrps1F	TTGAATATACGAGCGATGG	56
		Lrps1R	TGAGAGGAGAACTTGACC	
***Nostoc* sp. PCC 7120**	*mrp*	mrpF	GAACTGAATATGATTCGCAACG	51
		NmrpR	TGGCAATCTTCAACAATAAATTCC	
	ppc	ppcF	TATCGCCAAGAACCCTATCG	51
		NppcR	CCTGTTTCCGTCAAGTTATGC	
	prsA	prsAF	TTAAATGTGATTGGTGATGT	56
		NprsAR	ATCGTTCAACAGCAGGAG	
	purC	purCF	GATACCTGTCGTTTGTGG	51
		NpurCR	CTCCTGATAAGCATCTTCTAC	
	rpoA	NrpoAF	ACTCCGACATTCCAGACG	51
		rpoAR1	CGCTTGAGACAGTTATAGGC	
	rps1B	Nrps1F	GGTGGATGTGAAGGTGAC	51
		Nrps1R	CAGCCGTGAGAGTATGAC	
***Synechocystis* sp. PCC 6803**	*ilvD*	ilvDF2	ATGGGAACCGAGGGGATG	51
		LilvDR	TTATCACAACCGCCAATCG	
	mrp	SmrpF	GGCATTATCCGTCAGTTCC	56
		SmrpR	GGCACCGATTGAGTTAGG	
	prsA	SprsAF	CGTCGTCAGAGCCATAAC	51
		SprsAR	GAGCAAACGGGAGCCTTC	
	purC	purCF	GATACCTGTCGTTTGTGG	56
		SpurCR	TTCCACTGTTTGATAAGCG	
	rnpA	rnpAF1	AACGTCAAATTCGAGCCG	56
		rnpAR1	AACAACTGCTCTAATTCTTGC	
	rps1B	Srps1F	TTGGCAGAGTTGGAGGAG	56
		Srps1R	CGAGGATATGGGCTTTGAG	

CHAPTER VI

General discussion and future prospectives



General discussion and future prospectives

In this final section the main achievements of this thesis will be recapitulated and discussed in general. Additional work and future prospectives will also be addressed.

VI.1 Model organism

In this study, the cyanobacterium *Synechocystis* sp. PCC 6803 was used as base to generate a photoautotrophic chassis for synthetic biology applications. This organism was chosen since it is the best studied cyanobacterium, and a vast amount of transcriptomics, proteomics and metabolomics data is available. Using this information, genome-scale metabolic models of *Synechocystis* were built (Fu, 2009; Montagud *et al.*, 2010; Montagud *et al.*, 2011), which are important tools required for rational metabolic engineering approaches. These computer models allow the prediction of systems response to changes, such as the implementation of synthetic devices and circuits. This would not be possible for other cyanobacteria, since the number of high-throughput studies available is low and the existing information is insufficient to built robust metabolic models.

VI. 2 Preparation of a photoautotrophic chassis

Synthetic biology is being extensively used for biotechnological purposes, namely for the production of biofuels (e.g. Lee *et al.*, 2008; Service, 2008; Berry, 2010; Connor and Atsumi, 2010; Dellomonaco *et al.*, 2010; Colin *et al.*, 2011; Du *et al.*, 2011; Dunlop, 2011; Lamsen and Atsumi, 2012). In this context, the BioModularH_2 project aimed at producing a cyanobacterial chassis prone to receive synthetic devices primarily designed to produce hydrogen. These modules comprise synthetic heterologous hydrogenases, and since *Synechocystis* harbors a native bidirectional hydrogenase, this enzyme had to be inactivated to avoid electron competition. Therefore, a *Synechocystis* mutant was produced by deleting the two genes (*hoxY* and *hoxH*) encoding the hydrogenase part of the native bidirectional hydrogenase, and characterized. In addition, novel chromosomal neutral sites for the integration of synthetic devices were indentified.

VI.2.1 Characterization of *Synechocystis* Δ*hoxYH* mutant

The mutant lacking an active bidirectional hydrogenase (Δ*hoxYH*) was characterized at several levels: physiological, proteomic and transcription. The physiological studies were performed growing the wild-type strain and the Δ*hoxYH* mutant under different levels of

nitrate, glucose and irradiance, and the results obtained showed that there are no statistically significant differences comparing the growth of both organisms. These results are consistent with those recently published by Eckert *et al.* (2012) for similar continuous and light/dark conditions, and suggest that the absence of the bidirectional hydrogenase only has a marginal and not significant effect on the growth and fitness of *Synechocystis*, under the conditions tested. Although not fully understood, it has been proposed that the bidirectional hydrogenase has a role in dark fermentation and photosynthesis (acting as an electron valve), as well as in the maintenance of cell homeostasis under specific conditions (transition states) (see Appel and Schulz, 1996; Schmitz and Bothe, 1996; Appel and Schulz, 1998; Appel *et al.*, 1999; Appel *et al.*, 2000; Ludwig *et al.*, 2006). About 50% of the N_2-fixing strains do not possess the bidirectional hydrogenase (Tamagnini *et al.*, 1997; Tamagnini *et al.*, 2000; Schütz *et al.*, 2004; Ludwig *et al.*, 2006), and although most of the non-N_2-fixing strains contain it, this enzyme does not seem to be essential for cell survival (Appel *et al.*, 2000; Masukawa *et al.*, 2002).

Moreover, to assess changes in protein content, the proteomes of *Synechocystis* wild-type and Δ*hoxYH* mutant were compared under conditions that favor the bidirectional hydrogenase activity (dark and anaerobiosis) (Cournac *et al.*, 2002; Cournac *et al.*, 2004), using the iTRAQ methodology. In this experiment it was possible to identify 210 proteins and from these, 17 exhibited significant differential fold changes: twelve with higher levels in the mutant and five with higher level in the wild-type. Most of these proteins are related to the redox and energy state of the cell, showing a correlation to the proposed roles of the bidirectional hydrogenase. Interestingly, LexA was one of the proteins more abundant in the Δ*hoxYH* mutant than in the wild-type. This protein is one of the possible transcription regulators of *Synechocystis*' bidirectional hydrogenase genes (Gutekunst *et al.*, 2005; Oliveira and Lindblad, 2005; Antal *et al.*, 2006; Sjöholm *et al.*, 2007; Oliveira and Lindblad, 2009) and it was also suggested that it is a mediator of the redox state in *Synechocystis* cells (Antal *et al.*, 2006; Patterson-Fortin *et al.*, 2006). The number of identified proteins obtained in this proteomics approach is low, however, this is a recurrent problem for unicellular cyanobacteria probably due to the technical interference of highly abundant proteins, mainly phycobiliproteins (Gan *et al.*, 2005; Barrios-Llerena *et al.*, 2006; Pereira *et al.*, 2011). Nonetheless, very stringent criteria were used to reduce the number of false positives, and thus the proteins that were quantified are robust and have excellent statistical certainty due to the *P*-value of quantitation. Despite the low number of quantifiable proteins the network modeling approach 'mixture model on graphs' (MMG) (Noirel *et al.*, 2008; Sanguinetti *et al.*, 2008) was used on the metabolic network of *Synechocystis*, allowing to infer and conclude about overall differences in the metabolism

of the mutant as compared to wild-type.

Transcription studies were also performed to assess the patterns of (*i*) twelve genes encoding proteins showing significant differential abundance in the preceding study and (*ii*) six additional genes, encoding proteins related to stress response and iron levels. From the first group, 6 genes exhibited significant differences in transcript levels following the same pattern as the protein abundance, with the exception of two genes that showed an inverse trend. From the second group, only the two genes encoding transcription factors involved in iron homeostasis (Fur) exhibited significant changes comparing the wild-type to the mutant.

Both proteomic and transcription analysis suggest differences in iron pools/availability in the wild-type and the Δ*hoxYH* mutant. Since these results represent only indirect evidences, future studies should be performed to address iron availability in the two organisms using established protocols to measure intracellular free iron, namely by whole-cell electron paramagnetic resonance (EPR) spectroscopy (Yegorov *et al.*, 1993; Keyer and Imlay, 1996), electron spin resonance (ESR) (Yamamoto *et al.*, 2004), or atomic emission spectrometry (AES) (Ouyang and Isaacson, 2006).

Altogether, the previous results indicate that the lack of a functional bidirectional hydrogenase leads to an adjustment in the metabolism of the Δ*hoxYH* mutant that allows growth rates similar to the wild-type, reflecting *Synechocystis*' plasticity and metabolic adaptability. The *Synechocystis* Δ*hoxYH* mutant is therefore a robust chassis that and can be used for the integration of synthetic devices to fulfill specific purposes, namely hydrogen production.

To pursue this aim, several hydrogen producing devices (HPD) were designed and constructed, based on heterologous [FeFe] hydrogenases, namely HydA1 and HydA2 from *Chlamydomonas reinhardtii* (eukaryotic) and HydA from *Clostridium acetobutylicum* ATCC 824 (prokaryotic), fused to their ferredoxin counterparts through an artificial glycine-rich peptide. [FeFe] hydrogenases were taken into consideration since they exhibit higher enzymatic activity compared to the other type of hydrogenases (Adams, 1990; Happe *et al.*, 2002), and hydrogen production by *C. reinhardtii* and *C. acetobutylicum* and the respective hydrogenases have been extensively studied (Roessler and Lien, 1984; Happe and Naber, 1993; Gorwa *et al.*, 1996; Podestá *et al.*, 1996; Forestier *et al.*, 2003; Girbal *et al.*, 2005; Demuez *et al.*, 2007; Oh *et al.*, 2009; Meuser *et al.*, 2012). Linking the hydrogenases to the ferredoxins was performed to improve electron transfer between the two proteins. Three-dimensional models of the eukaryotic fusion enzymes were also built, constituting important tools to address activity improvement by protein engineering.

Moreover, since maturation is important to achieve functional enzymes, synthetic modules to express the specific maturation proteins were also designed and constructed.

Preliminary tests have already been carried out in *E. coli* containing both the HPD and maturation devices and the results obtained showed hydrogen evolution only in the presence of the latter (work performed at Prof. Peter Lindblad's laboratory, at the University of Uppsala, Sweden). The results obtained showed that, at least in *E. coli*, the fusion hydrogenases are active and that the maturation proteins are essential to produce functional hydrogenases. The introduction of the HPDs into *Synechocystis* wild-type and Δ*hoxYH* mutant has already been initiated at Prof. Peter Lindblad's laboratory, using a broad-host-range BioBrick-compatible shuttle vector (replicative), developed by the group (Huang *et al.*, 2010).

VI.2.2 Chromosomal neutral sites: platforms for the integration of synthetic devices

Complex synthetic genetic circuits may require not only the implementation of several devices designed to perform a given task (e.g. production of a metabolite), but also other devices to control gene transcription and regulate the genetic circuit. In *E. coli* this is usually achieved by the use of several compatible replicative plasmids, but in cyanobacteria the availability of such vectors is limited and that approach may not always be viable. To date several sites have been used to insert DNA sequences into *Synechocystis* chromosome through double homologous recombination (e.g. Williams, 1988; Burnap *et al.*, 1994; Aoki *et al.*, 1995; Aoki *et al.*, 2011), but a close analysis revealed that they may present some drawbacks. Therefore, an effort was done in trying to identify and characterize novel chromosomal neutral sites foreseeing the integration of synthetic devices into *Synechocystis*. An initial bioinformatics analysis of the predicted unknown or hypothetical chromosomal ORFs of *Synechocystis* was performed, following several established criteria for the selection of such sites. Sixteen putative sites were identified and after the analysis of their genomic context, the transcription of thirteen sites was assessed by RT-PCR at the exponential growth phase. From these, eight sites exhibited low or undetectable transcript levels and further transcription studies were performed at other stages of growth. The six sites exhibiting very low or undetectable transcription levels were selected and integrative vectors were designed to produce mutants by homologous recombination. Due to constant cloning difficulties found during vectors construction, one of the sites was disregarded, and a final set of five putative neutral sites was characterized. Moreover, to be able to use this sites for the stable introduction of synthetic devices, a cloning platform carrying restriction sites compatible with the BioBrick assembly standard RFC[10] was synthesized and insulated by double

transcription terminators. This BioBrick-compatible interface was cloned in each integrative vector and series of plasmids were constructed. Initially, a series of plasmids was constructed to substitute the neutral sites chromosomal region by a kanamycin resistance cassette and fully segregated mutants were produced. The successful generation of fully segregated mutants in each selected sites shows that they do not seem to encode proteins essential to the viability of the cyanobacterium under standard growth conditions. These mutants were used in growth experiments to assess the "neutrality" of the sites. *Synechocystis* wild-type and each of the mutants were grown under different cultivation conditions and the growth patterns were analyzed using the ANOVA test. The results obtained allowed to rank the selected neutral sites. Additionaly, experiments were carried out to assess the robustness of the synthetic BioBrick-compatible interface, constructing plasmids carrying a promoterless green fluorescent protein generator (GFP; BBa_E0240) and producing *Synechocystis* mutants. These mutants were analyzed by molecular and live imaging techniques and it was not possible to detect GFP expression, which shows that the synthetic interface is properly insulated and any possible promoter activity from backbone DNA sequences is prevented. In addition, the functionality of each site was evaluated producing mutants carrying the same GFP generator under the control of a synthetic constitutive promoter, designed based on the light-responsive promoter of the *psbA2* gene (Eriksson *et al.*, 2000; Agrawal *et al.*, 2001; Shibato *et al.*, 2002). The detection of GFP (protein expression and fluorescence detection) shows not only that the promoter is functional, but also that the sites can be used to introduce synthetic devices. RT-PCR studies will be performed to confirm the results obtained at the protein level. Furthermore, since the available genetic parts for cyanobacteria are scarce, the synthetic promoter constructed constitutes an additional tool that can be used in the design and construction of new devices and circuits, and the BioBrick-compatible vectors constructed can be used to integrate synthetic devices previously assembled in *E. coli*, into the chromosome of *Synechocystis*.

Structural conformation of the DNA or unknown regulatory *cis* elements upstream of the introduced devices may affect promoter strength and future work shall be performed to evaluate if the position of each neutral site in the chromosome may affect transcription. This will be performed analyzing *gfp* transcription by RT-qPCR, in the mutants that constitutively express GFP. Since all these mutants have the same promoter upstream of the *gpf* gene the differences found in transcript level, if detected, will reflect the possible effect of the genomic context in transcription at each neutral site. Moreover, the identification of new neutral sites shall be pursued, maybe focusing in pseudogenes that have been defined as nonfunctional sequences of genomic DNA originally derived from

functional genes, bearing in mind that they may exhibit functional roles, such as gene expression and regulation, generation of genetic diversity, or even be involved in gene conversion or recombination with functional genes (Balakirev and Ayala, 2003).

VI.3 Design of orthogonal regulatory networks

In order to increase the number of parts available for synthetic biology, regulatory networks, based on well characterized *E. coli* promoters (Cox *et al.*, 2007; Shultzaberger *et al.*, 2007) and the DNA binding domains of zinc finger proteins (Wright *et al.*, 2006), were designed and constructed. The core region of each promoter included a 12 bp sequence (operator), orthogonal to *E. coli*, and that is specifically recognized by a zinc finger array (ZFA). Binding of the ZFA to the operator sequence will block the recognition of the promoter by the RNA polymerase holoenzyme by steric hindrance. These synthetic regulatory networks are being tested at Dr. Alfonso Jaramillo's laboratory, and eventually, a library of characterized promoters with varying strengths, together with the associated transcription factors, will be available to design new synthetic devices and regulate genetic circuits. The promoters and regulators design has been made orthogonal to *E. coli*, but later on these elements could be used or redesigned to work in other organisms, namely *Synechocystis*.

VI.4 Normalization of transcription data

An issue that was particularly evident, when the transcription studies were performed to characterize the *Synechocystis* Δ*hoxYH* mutant, was the lack of studies addressing reference genes to normalize transcription data using RT-qPCR in cyanobacteria. Several studies have been performed to identify and validate suitable reference genes in other organisms (e.g. Teste *et al.*, 2009; Reiter *et al.*, 2011; Chechi *et al.*, 2012; Liu *et al.*, 2012), but in cyanobacteria the choice was limited to the genes used in other publications. Since recent publications recommend the use of a minimum of three reference genes for RT-qPCR data normalization (Bustin *et al.*, 2009; Taylor *et al.*, 2010), a comprehensive study was performed to identify and validate suitable reference genes for three different cyanobacteria, representative of three distinct morphological groups: *Synechocystis* sp. PCC 6803 (unicellular), *Lyngbya aestuarii* CCY 9616 (filamentous) and *Nostoc* sp. PCC 7120 (filamentous and heterocystous). The study was performed assessing the transcription of (*i*) genes that have been used as controls in previous RT-qPCR studies and (*ii*) genes that did not exhibit significant variation of transcript levels in two microarray studies (Hihara *et al.*, 2001; Ehira and Ohmori, 2006) available in the CyanoBIKE server (Elhai *et al.*, 2009). Six candidate reference genes were selected and their transcription was analyzed in each cyanobacterium, grown under conditions routinely used for

cultivation. Gene expression stability was assessed using three distinct validation algorithms (Vandesompele *et al.*, 2002; Andersen *et al.*, 2004; Pfaffl *et al.*, 2004) and the most stable genes were identified for each organism in each growth condition. The recommended reference genes for RT-qPCR data normalization in *Synechocystis* are the 16S rRNA, *petB* and *rnpB*. Although reference genes validation is always mandatory for every experiment, the array of experimental conditions tested makes this work a valuable starting point in the choice of reference genes when studying cyanobacteria. Future transcription studies performed in the continuation of the work presented in this thesis will consider the genes identified to serve as reference in RT-qPCR data normalization.

VI.5 Beyond the work developed

Future work shall be performed to improve chassis functionality, namely in terms of genetic stabilization. *Synechocystis*' genome is full of insertion elements (IS) like transposases (Kaneko *et al.*, 1996b; Kaneko *et al.*, 2001; Kaneko *et al.*, 2003), including an active homing nuclease (Orlowski *et al.*, 2007; Zhao *et al.*, 2009). In this organism, the sequences assigned to putative transposases represent at least 76 Mbp, without considering intergenic sequences. Elimination of these sequences would not only provide genetic stability but will also reduce the genome size. Nonetheless, sequences that are putatively recognized by these enzymes should be avoided in sequences designed *de novo* and special attention shall be taken regarding the existing ones. In addition, several genes putatively encoding restriction and modification systems were identified in the genome of *Synechocystis* (Roberts *et al.*, 2010). Preliminary analysis of those genes indicate that they are indeed transcribed (this work, data not shown), although no biochemical evidence of their activity has ever been described in the literature. This subject should also be addressed in future studies, not only to improve genetic stability, but possibly to increase transformation efficiency, since restriction systems are known to impair or even abolish the capacity of an organism to receive foreign DNA (Elhai *et al.*, 1997; Zhang and Blaser, 2012).

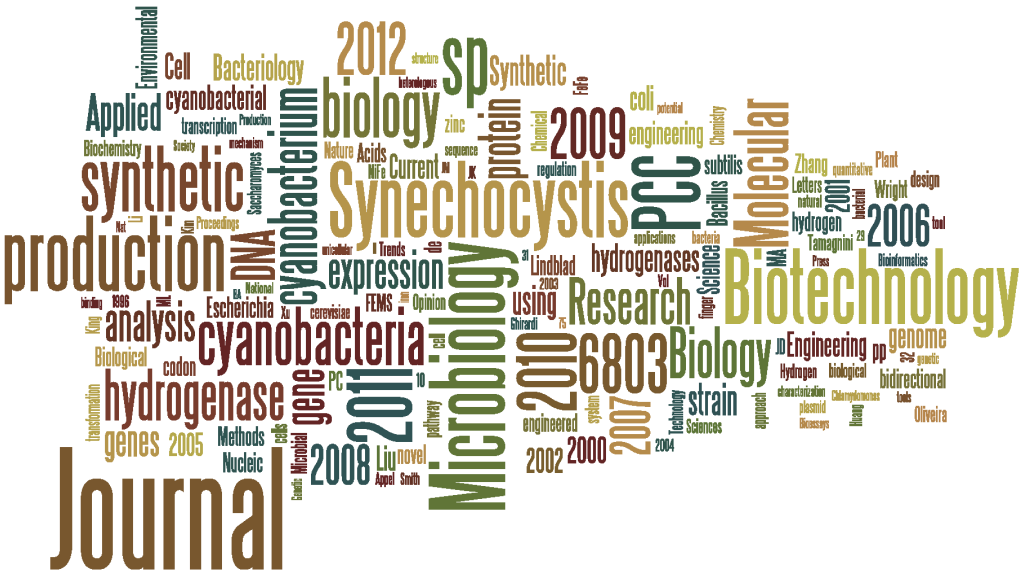
VI.6 Conclusions

In this work, the cyanobacterium *Synechocystis* sp. PCC 6803 was used to produce photoautotrophic chassis prepared to accommodate synthetic devices. To stably integrate these devices into the chassis, novel chromosomal neutral sites have been identified and characterized, and integrative vectors compatible with BioBrick standards were developed. Moreover, new synthetic parts and devices were designed and constructed to be used in synthetic biology applications, namely in hydrogen production and in the regulation of genetic circuits. Furthermore, several reference genes to be used for RT-

qPCR data normalization in cyanobacteria were validated. Good reference genes are essential for the accurate interpretation of transcription data and, consequently, for the characterization of synthetic devices and circuits implemented into a chassis.

Overall, the results obtained demonstrate the potential of *Synechocystis* to be employed in biotechnological approaches requiring a solar-driven microbial factory.

General references



General references

A. O'Malley M, Powell A, Davies JF & Calvert J (2008) Knowledge-making distinctions in synthetic biology. *Bioessays* **30**: 57-65.

Abed RMM, Dobretsov S & Sudesh K (2009) Applications of cyanobacteria in biotechnology. *Journal of Applied Microbiology* **106**: 1-12.

Adams D (2002) Symbiotic interactions. *The Ecology of Cyanobacteria,* (Whitton BA & Potts M), pp. 523-561. Kluwer Academics Publishers, Dordrecht, The Netherlands.

Adams MWW (1990) The structure and mechanism of iron-hydrogenases. *Biochimica et Biophysica Acta - Bioenergetics* **1020**: 115-145.

Agapakis CM & Silver PA (2009) Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems* **5**: 704-713.

Agapakis CM & Silver PA (2010) Modular electron transfer circuits for synthetic biology: insulation of an engineered biohydrogen pathway. *Bioengineered Bugs* **1**: 413-418.

Agrawal GK, Kato H, Asayama M & Shirai M (2001) An AU-box motif upstream of the SD sequence of light-dependent *psbA* transcripts confers mRNA instability in darkness in cyanobacteria. *Nucleic Acids Research* **29**: 1835-1843.

Albertsen L, Chen Y, Bach LS, Rattleff S, Maury J, Brix S, Nielsen J & Mortensen UH (2011) Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. *Applied and Environmental Microbiology* **77**: 1033-1040.

Amidi M, de Raad M, Crommelin D, Hennink W & Mastrobattista E (2011) Antigen-expressing immunostimulatory liposomes as a genetically programmable synthetic vaccine. *Systems and Synthetic Biology* **5**: 21-31.

Andersen CL, Jensen JL & Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**: 5245-5250.

Antal TK, Oliveira P & Lindblad P (2006) The bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *International Journal of Hydrogen Energy* **31**: 1439-1444.

Aoki R, Goto T & Fujita Y (2011) A heme oxygenase isoform is essential for aerobic growth in the cyanobacterium *Synechocystis* sp. PCC 6803: modes of differential operation of two isoforms/enzymes to adapt to low oxygen environments in cyanobacteria. *Plant and Cell Physiology* **52**: 1744-1756.

Aoki S, Kondo T & Ishiura M (1995) Circadian expression of the *dnaK* gene in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* **177**: 5606-5611.

Appel J & Schulz R (1996) Sequence analysis of an operon of a NAD(P)-reducing nickel hydrogenase from the cyanobacterium *Synechocystis* sp. PCC 6803 gives additional evidence for direct coupling of the enzyme to NAD(P)H-dehydrogenase (complex 1). *Biochimica et Biophysica Acta* **1298**: 141-147.

Appel J & Schulz R (1998) Hydrogen metabolism in organisms with oxygenic photosynthesis: hydrogenases as important regulatory devices for a proper redox poising? *Journal of Photochemistry and Photobiology B: Biology* **47**: 1-11.

Appel J, Phunpruch S & Schulz R (1999) Hydrogenase(s) in *Synechocystis* - Tools for Photohydrogen Production? *BioHydrogen,* (Zaborsky Oskar R., Benemann John R., Matsunaga Tadashi, Miyake Jun & San Pietro Anthony), pp. 189-196. Springer US.

Appel J, Phunpruch S, Steinmüller K & Schulz R (2000) The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Archives of Microbiology* **173**: 333-338.

Aubel D & Fussenegger M (2010) Mammalian synthetic biology - from tools to therapies. *BioEssays* **32**: 332-345.

Baek JO, Seo JW, Kwon O, Park SM, Kim CH & Kim IH (2012) Production of human papillomavirus type 33 L1 major capsid protein and virus-like particles from *Bacillus subtilis* to develop a prophylactic vaccine against cervical cancer. *Enzyme and Microbial Technology* **50**: 173-180.

Balakirev ES & Ayala FJ (2003) Pseudogenes: are they "junk" or functional DNA? *Annual Review of Genetics* **37**: 123-151.

Bar-Yam S, Byers-Corbin J, Casagrande R, Eichler F, Lin A, Oesterreicher M, Regardh P, Turlington RD & Oye KA (2012) *The regulation of synthetic biology: a guide to United States and European Union regulations, rules and guidelines*. World-wide electronic publication, http://synberc.org/sites/default/files/Concise%20Guide%20to%20Synbio%20Regulation%20OYE%20Jan%202012_0.pdf.

Barrios-Llerena ME, Chong PK, Gan CS, Snijders APL, Reardon KF & Wright PC (2006) Shotgun proteomics of cyanobacteria - applications of experimental and data-mining techniques. *Briefings in Functional Genomics & Proteomics* **5**: 121-132.

Becker EC & Meyer RJ (1997) Acquisition of resistance genes by the IncQ plasmid R1162 is limited by its high copy number and lack of a partitioning mechanism. *Journal of Bacteriology* **179**: 5947-5950.

Bentley PJ (2009) Methods for improving simulations of biological systems: systemic computation and fractal proteins. *Journal of The Royal Society Interface* **6**: S451-S466.

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

Berry DA (2010) Engineering organisms for industrial fuel production. *Bioengineered* **1**: 303-308.

Berry JP, Gantar M, Perez MH, Berry G & Noriega FG (2008) Cyanobacterial toxins as allelochemicals with potential applications as algaecides, herbicides and insecticides. *Marine Drugs* **6**: 117-146.

Biswas R, Yamaoka M, Nakayama H, Kondo T, Yoshida K-i, Bisaria V & Kondo A (2012) Enhanced production of 2,3-butanediol by engineered *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **94**: 651-658.

Blunt JW, Copp BR, Hu W-P, Munro MHG, Northcote PT & Prinsep MR (2009) Marine natural products. *Natural Product Reports* **26**: 170-244.

Boyle PM & Silver PA (2009) Harnessing nature's toolbox: regulatory elements for synthetic biology. *J R Soc Interface* **6 Suppl 4**: S535-546.

Browning DF & Busby SJW (2004) The regulation of bacterial transcription initiation. *Nat Rev Micro* **2**: 57-65.

Burja AM, Banaigs B, Abou-Mansour E, Grant Burgess J & Wright PC (2001) Marine cyanobacteria - a prolific source of natural products. *Tetrahedron* **57**: 9347-9377.

Burnap RL, Qian M, Shen JR, Inoue Y & Sherman LA (1994) Role of disulfide linkage and putative intermolecular binding residues in the stability and binding of the extrinsic manganese-stabilizing protein to the photosystem II reaction center. *Biochemistry* **33**: 13712-13718.

Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**: 169-193.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J & Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**: 611-622.

Bustin SA (2010) Why the need for qPCR publication guidelines? - The case for MIQE. *Methods* **50**: 217-226.

Cai Y, Wilson ML & Peccoud J (2010) GenoCAD for iGEM: a grammatical approach to the design of standard-compliant constructs. *Nucleic Acids Research* **38**: 2637-2644.

Canton B, Labno A & Endy D (2008) Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26**: 787-793.

Cantone I, Marucci L, Iorio F, Ricci MA, Belcastro V, Bansal M, Santini S, di Bernardo M, di Bernardo D & Cosma MP (2009) A yeast synthetic network for *in vivo* assessment of reverse-engineering and modeling approaches. *Cell* **137**: 172-181.

Carrera J, Rodrigo G & Jaramillo A (2009) Towards the automated engineering of a synthetic genome. *Molecular Biosystems* **5**: 733-743.

Casalot L & Rousset M (2001) Maturation of the [NiFe] hydrogenases. *Trends in Microbiology* **9**: 228-237.

Castenholz RW (2001) Phylum BX. Cyanobacteria. General characteristics of the cyanobacteria. *Bergey's Manual of Systematic Bacteriology, Volume 1: The Archaea and the Deeply Branching and Phototropic Bacteria,* (Garrity George M., Boone David R. & Castenholz Richard W.), pp. 474-487. Springer-Verlag, New York.

Cavalier-Smith T (2001) Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *Journal of Molecular Evolution* **53**: 555-595.

Chang CH, King PW, Ghirardi ML & Kim K (2007) Atomic resolution modeling of the ferredoxin:[FeFe] hydrogenase complex from *Chlamydomonas reinhardtii*. *Biophysical Journal* **93**: 3034-3045.

Chang MCY & Keasling JD (2006) Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol* **2**: 674-681.

Chechi K, Gelinas Y, Mathieu P, Deshaies Y & Richard D (2012) Validation of reference genes for the relative quantification of gene expression in human epicardial adipose tissue. *PLoS ONE* **7**: e32265.

Clancy K & Voigt CA (2010) Programming cells: towards an automated 'Genetic Compiler'. *Current Opinion in Biotechnology* **21**: 572-581.

Colin VL, Rodríguez A & Cristóbal HA (2011) The role of synthetic biology in the design of microbial cell factories for biofuel production. *Journal of Biomedicine and Biotechnology* **2011**: 601834.

Comeron JM & Aguadé M (1998) An evaluation of measures of synonymous codon usage bias. *Journal of Molecular Evolution* **47**: 268-274.

Connor MR & Atsumi S (2010) Synthetic biology guides biofuel production. *Journal of Biomedicine and Biotechnology* **2010**: 541698.

Cournac L, Mus F, Bernard L, Guedeney G, Vignais P & Peltier G (2002) Limiting steps of hydrogen production in *Chlamydomonas reinhardtii* and *Synechocystis* PCC 6803 as analysed by light-induced gas exchange transients. *International Journal of Hydrogen Energy* **27**: 1229-1237.

Cournac L, Guedeney G, Peltier G & Vignais PM (2004) Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis* sp. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. *Journal of Bacteriology* **186**: 1737-1746.

Cox RS, Surette MG & Elowitz MB (2007) Programming gene expression with combinatorial promoters. *Molecular Systems Biology* **3**: 145.

David M, Vielma M & Julliot J-S (1983) Introduction of IncQ plasmids into *Rhizobium meliloti*. Isolation of a host-range mutant of RSF1010 plasmid. *FEMS Microbiology Letters* **16**: 335-341.

de Lorenzo V (2010a) Synthetic biology: something old, something new…. *Bioessays* **32**: 267-270.

de Lorenzo V (2010b) Environmental biosafety in the age of Synthetic Biology: Do we really need a radical new approach? *BioEssays* **32**: 926-931.

de Lorenzo V (2011) Beware of metaphors: chasses and orthogonality in synthetic biology. *Bioengineered Bugs* **2**: 3-7.

Dellomonaco C, Fava F & Gonzalez R (2010) The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microbial Cell Factories* **9**: 3.

Dembitsky V (2006) Anticancer activity of natural and synthetic acetylenic lipids. *Lipids* **41**: 883-924.

Demuez M, Cournac L, Guerrini O, Soucaille P & Girbal L (2007) Complete activity profile of *Clostridium acetobutylicum* [FeFe]-hydrogenase and kinetic parameters for endogenous redox partners. *FEMS Microbiology Letters* **275**: 113-121.

Devillers J, Doré JC, Guyot M, Poroikov V, Gloriozova T, Lagunin A & Filimonov D (2007) Prediction of biological activity profiles of cyanobacterial secondary metabolites. *SAR and QSAR in Environmental Research* **18**: 629 - 643.

DeWall MT & Cheng DW (2011) The minimal genome - a metabolic and environmental comparison. *Briefings in Functional Genomics* **10**: 312-315.

Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G & Zumla A (2004) Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *BioTechniques* **37**: 112-114, 116, 118-119.

Dragosits M, Nicklas D & Tagkopoulos I (2012) A synthetic biology approach to self-regulatory recombinant protein production in Escherichia coli. *Journal of Biological Engineering* **6**: 2.

Dreier B, Segal DJ & Barbas Iii CF (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *Journal of Molecular Biology* **303**: 489-502.

Du J, Shao Z & Zhao H (2011) Engineering microbial factories for synthesis of value-added products. *Journal of Industrial Microbiology & Biotechnology* **38**: 873-890.

Dunlop M (2011) Engineering microbes for tolerance to next-generation biofuels. *Biotechnology for Biofuels* **4**: 32.

Eckert C, Boehm M, Carrieri D, Yu J, Dubini A, Nixon PJ & Maness P-C (2012) Genetic analysis of the Hox hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 reveals subunit roles in association, assembly, maturation, and function. *Journal of Biological Chemistry.* DOI: 10.1074/jbc.M112.392407.

Ehira S & Ohmori M (2006) NrrA directly regulates expression of *hetR* during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Journal of Bacteriology* **188**: 8520-8525.

Ehrenreich IM, Waterbury JB & Webb EA (2005) Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. *Applied and Environmental Microbiology* **71**: 7401-7413.

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, Taton A, Massar J, Myers JK, Travers M, Casey J, Slupesky M & Shrager J (2009) BioBIKE: a Web-based, programmable, integrated biological knowledge base. *Nucleic Acids Research* **37**: W28-W32.

Ellis T, Wang X & Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* **27**: 465-471.

Elowitz MB & Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335-338.

Endler L, Rodriguez N, Juty N, Chelliah V, Laibe C, Li C & Le Novère N (2009) Designing and encoding models for synthetic biology. *Journal of The Royal Society Interface* **6**: S405-S417.

Endy D (2005) Foundations for engineering biology. *Nature* **438**: 449-453.

Eriksson J, Salih GF, Ghebramedhin H & Jansson C (2000) Deletion mutagenesis of the 5' *psbA2* region in *Synechocystis* 6803: identification of a putative cis element involved in photoregulation. *Molecular Cell Biology Research Communications* **3**: 292-298.

Ermolaeva MD, Khalak HG, White O, Smith HO & Salzberg SL (2000) Prediction of transcription terminators in bacterial genomes. *Journal of Molecular Biology* **301**: 27-33.

Ermolaeva MD (2001) Synonymous codon usage in bacteria. *Current Issues in Molecular Biology* **3**: 91-97.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC & Gerwick WH (2011) Temporal dynamics of natural product biosynthesis in marine cyanobacteria. *Proceedings of the National Academy of Sciences* **108**: 5226-5231.

ETC_Group (2007) *Extreme Genetic Engineering - An Introduction to Synthetic Biology*. World-wide electronic publication, www.etcgroup.org/upload/publication/602/01/synbioreportweb.pdf.

Falch B, König G, Wright A, Sticher O, Angerhofer C, Pezzuto J & Bachmann H (1995) Biological activities of cyanobacteria: evaluation of extracts and pure compounds. *Planta Medica* **61**: 321-328.

Fernandez-Fuentes N, Zhai J & Fiser A (2006) ArchPRED: a template based loop structure prediction server. *Nucleic Acids Research* **34**: W173-W176.

Foley PL & Shuler ML (2010) Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnology and Bioengineering* **105**: 26-36.

Forestier M, King P, Zhang L, Posewitz M, Schwarzer S, Happe T, Ghirardi ML & Seibert M (2003) Expression of two [Fe]-hydrogenases in *Chlamydomonas reinhardtii* under anaerobic conditions. *European Journal of Biochemistry* **270**: 2750-2758.

Forster AC & Church GM (2007) Synthetic biology projects in vitro. *Genome Research* **17**: 1-6.

Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E & Keasling JD (2008) Biofuel alternatives to ethanol: pumping the microbial well. *Trends in biotechnology* **26**: 375-381.

Forzi L & Sawers R (2007) Maturation of [NiFe]-hydrogenases in *Escherichia coli*. *Biometals* **20**: 565-578.

Fox JM & Erill I (2010) Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Research* **17**: 185-196.

Friedberg D (1988) Use of reporter genes in cyanobacteria. *Methods in Enzymology,* Vol. Volume 167 (Lester Packer Alexander N. Glazer), pp. 736-747. Academic Press.

Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, Dobbs D, Miller L & Voytas DF (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Research* **37**: D279-D283.

Fu P (2006) A perspective of synthetic biology: assembling building blocks for novel functions. *Biotechnology Journal* **1**: 690-699.

Fu PC (2009) Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion. *Journal of Chemical Technology and Biotechnology* **84**: 473-483.

Fulda S, Mikkat S, Huang F, Huckauf J, Marin K, Norling B & Hagemann M (2006) Proteome analysis of salt stress response in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Proteomics* **6**: 2733-2745.

Gan CS, Reardon KF & Wright PC (2005) Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of *Synechocystis* sp. PCC 6803. *Proteomics* **5**: 2468-2478.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA & Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi Z-Q, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO & Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52-56.

Giesecke AV & Joung JK (2007) The bacterial two-hybrid system as a reporter system for analyzing protein-protein interactions. *Cold Spring Harbor Protocols* **2007**: pdb.prot4672.

Girbal L, von Abendroth G, Winkler M, Benton PMC, Meynial-Salles I, Croux C, Peters JW, Happe T & Soucaille P (2005) Homologous and heterologous overexpression in *Clostridium acetobutylicum* and characterization of purified clostridial and algal Fe-only hydrogenases with high specific activities. *Applied and Environmental Microbiology* **71**: 2777-2781.

Gorwa MF, Croux C & Soucaille P (1996) Molecular characterization and transcriptional analysis of the putative hydrogenase gene of *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* **178**: 2668-2675.

Griffith KL & Grossman AD (2008) Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Molecular Microbiology* **70**: 1012-1025.

Grigorieva G & Shestakov S (1982) Transformation in the cyanobacterium *Synechocystis* sp. 6803. *FEMS Microbiology Letters* **13**: 367-370.

Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

Gutekunst K, Phunpruch S, Schwarz C, Schuchardt S, Schulz-Friedrich R & Appel J (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. *Molecular Microbiology* **58**: 810-823.

Happe T & Naber JD (1993) Isolation, characterization and N-terminal amino acid sequence of hydrogenase from the green alga *Chlamydomonas reinhardtii*. *European Journal of Biochemistry* **214**: 475-481.

Happe T, Hemschemeier A, Winkler M & Kaminski A (2002) Hydrogenases in green algae: do they save the algae's life and solve our energy problems? *Trends in Plant Science* **7**: 246-250.

Hayes WS & Borodovsky M (1998) Deriving ribosomal binding site (RBS) statistical models from unannotated DNA sequences and the use of the RBS model for N-terminal prediction. *Pacific Symposium on Biocomputing* **1998**: 279–290.

Heidorn T, Camsund D, Huang H-H, Lindberg P, Oliveira P, Stensjö K & Lindblad P (2011) Synthetic biology in cyanobacteria: engineering and analyzing novel functions. *Methods in Enzymology,* Vol. Volume 497 (Chris Voigt), pp. 539-579. Academic Press.

Heinemann M & Panke S (2006) Synthetic biology - putting engineering into biology. *Bioinformatics* **22**: 2790-2799.

Herdman M, Janvier M, Rippka R & Stanier RY (1979) Genome size of cyanobacteria. *Journal of General Microbiology* **111**: 73-85.

Hihara Y, Kamei A, Kanehisa M, Kaplan A & Ikeuchi M (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *The Plant Cell Online* **13**: 793-806.

Hmc O (2000) The American Heritage Dictionary of the English Language.), Houghton Mifflin Company.

Hobom B (1980) Gene surgery: on the threshold of synthetic biology. *Medizinische Klinik* **75**: 834-841.

Hou L (2010) Improved production of ethanol by novel genome shuffling in *Saccharomyces cerevisiae*. *Applied Biochemistry and Biotechnology* **160**: 1084-1093.

Huang F, Fulda S, Hagemann M & Norling B (2006) Proteomic screening of salt-stress-induced changes in plasma membranes of *Synechocystis* sp. strain PCC 6803. *Proteomics* **6**: 910-920.

Huang H-H, Camsund D, Lindblad P & Heidorn T (2010) Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Research* **38**: 2577-2593.

Hutterer KM, Zhang Z, Michaels ML, Belouski E, Hong RW, Shah B, Berge M, Barkhordarian H, Le E, Smith S, Winters D, Abroson F, Hecht R & Liu J (2012) Targeted codon optimization improves translational fidelity for an Fc fusion protein. *Biotechnology and Bioengineering* **109**: 2770-2777.

Hwang YC, Lin CC, Chang JY, Mori H, Juan HF & Huang HC (2009) Predicting essential genes based on network and sequence analysis. *CORD Conference Proceedings* **5**: 1672-1678.

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology* **151**: 389-409.

Itaya M, Fujita K, Kuroki A & Tsuge K (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nature Methods* **5**: 41-43.

Jewett MC & Forster AC (2010) Update on designing and building minimal cells. *Current Opinion in Biotechnology* **21**: 697-703.

Jia X, Li S, Xie S & Wen J (2012) Engineering a metabolic pathway for isobutanol biosynthesis in *Bacillus subtilis*. *Applied Biochemistry and Biotechnology* **168**: 1-9.

Johnson ET & Schmidt-Dannert C (2008) Light-energy conversion in engineered microorganisms. *Trends in Biotechnology* **26**: 682-689.

Joung JK (2001) Identifying and modifying protein–DNA and protein–protein interactions using a bacterial two-hybrid selection system. *Journal of Cellular Biochemistry* **84**: 53-57.

Juhas M, Eberl L & Glass JI (2011) Essence of life: essential genes of minimal genomes. *Trends in Cell Biology* **21**: 562-568.

Kämpf MM & Weber W (2012) Synthetic gene networks as blueprint for smart hydrogels. *Synthetic Gene Networks,* Vol. 813 (Weber Wilfried & Fussenegger Martin), pp. 377-389. Humana Press.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M & Tabata S (1996a) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research* **3**: 109-136.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M & Tabata S (1996b) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**: 109-136.

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 Research* **8**: 205-213.

Kaneko T, Nakamura Y, Sasamoto S, Watanabe A, Kohara M, Matsumoto M, Shimpo S, Yamada M & Tabata S (2003) Structural analysis of four large Plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Research* **10**: 221-228.

Keskin T & Hallenbeck PC (2012) Hydrogen production from sugar industry wastes using single-stage photofermentation. *Bioresource Technology* **112**: 131-136.

Keyer K & Imlay JA (1996) Superoxide accelerates DNA damage by elevating free-iron levels. *Proceedings of the National Academy of Sciences, USA* **93**: 13635-13640.

Kim I-K, Roldão A, Siewers V & Nielsen J (2012) A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Research* **12**: 228-248.

Kingsford C, Ayanbule K & Salzberg S (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology* **8**: R22.

Klasson L & Andersson SGE (2010) Research on small genomes: implications for synthetic biology. *Bioessays* **32**: 288-295.

Klug A (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry* **79**: 213-231.

Koksharova OA & Wolk CP (2002) Genetic tools for cyanobacteria. *Applied Microbiology and Biotechnology* **58**: 123-137.

Kondo T, Tezuka H, Ishii J, Matsuda F, Ogino C & Kondo A (2012) Genetic engineering to enhance the Ehrlich pathway and alter carbon flux for increased isobutanol production from glucose by *Saccharomyces cerevisiae*. *Journal of Biotechnology* **159**: 32-37.

Kufryk GI, Sachet M, Schmetterer G & Vermaas WFJ (2002) Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiology Letters* **206**: 215-219.

Kunert A, Hagemann M & Erdmann N (2000) Construction of promoter probe vectors for *Synechocystis* sp. PCC 6803 using the light-emitting reporter systems Gfp and LuxAB. *Journal of Microbiological Methods* **41**: 185-194.

Kwon YM, Ricke SC, Hongoh Y & Toyoda A (2011) Whole-genome sequencing of unculturable bacterium using whole-genome amplification. *High-throughput next generation sequencing: methods and applications,* Vol. 733 (Kwon Young Min & Ricke Steven C.), pp. 25-33. Humana Press.

Labarre J, Chauvat F & Thuriaux P (1989) Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *Journal of Bacteriology* **171**: 3449-3457.

Lamsen EN & Atsumi S (2012) Recent progress in synthetic biology for microbial production of C3-C10 alcohols. *Frontiers in Microbiology* **3**: 196.

Leduc S (1910) *Théorie physico-chimique de la vie et générations spontanées*. A. Poinat, Paris.

Leduc S (1912) *La biologie synthétique*. A. Poinat, Paris.

Lee SK, Chou H, Ham TS, Lee TS & Keasling JD (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology* **19**: 556-563.

Leitão E, Pereira S, Bondoso J, Ferreira D, Pinto F, Moradas-Ferreira P & Tamagnini P (2006) Genes involved in the maturation of hydrogenase(s) in the nonheterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4. *International Journal of Hydrogen Energy* **31**: 1469-1477.

Li P, Miao X, Li R & Zhong J (2011a) *In situ* biodiesel production from fast-growing and high oil content *Chlorella pyrenoidosa* in rice straw hydrolysate. *Journal of Biomedicine and Biotechnology.* DOI: 10.1155/2011/141207.

Li S, Wen J & Jia X (2011b) Engineering *Bacillus subtilis* for isobutanol production by heterologous Ehrlich pathway construction and the biosynthetic 2-ketoisovalerate precursor pathway overexpression. *Applied Microbiology and Biotechnology* **91**: 577-589.

Liang J, Luo Y & Zhao H (2011) Synthetic biology: putting synthesis into biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **3**: 7-20.

Lindberg P, Park S & Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. *Metabolic Engineering* **12**: 70-79.

Liu C, Wu G, Huang X, Liu S & Cong B (2012) Validation of housekeeping genes for gene expression studies in an ice alga *Chlamydomonas* during freezing acclimation. *Extremophiles* **16**: 419-425.

Liu P-Q, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li P-X, Chen B, Mendel MC, Zhong X, Lee Y-L, Eisenberg SP, Spratt SK, Case CC & Wolffe AP (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. *Journal of Biological Chemistry* **276**: 11323-11334.

Liu Q, Xia Z & Case CC (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *Journal of Biological Chemistry* **277**: 3850-3856.

Loakes D & Holliger P (2009) Darwinian chemistry: towards the synthesis of a simple cell. *Molecular Biosystems* **5**: 686-694.

Long H, King PW, Ghirardi ML & Kim K (2009) Hydrogenase/ferredoxin charge-transfer complexes: effect of hydrogenase mutations on the complex association *The Journal of Physical Chemistry A* **113**: 4060-4067.

Looger LL, Dwyer MA, Smith JJ & Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. *Nature* **423**: 185-190.

Lopo M, Montagud A, Navarro E, Cunha I, Zille A, de Córdoba PF, Moradas-Ferreira P, Tamagnini P & Urchueguía JF (2012) Experimental and modeling analysis of *Synechocystis* sp. PCC 6803 growth. *Journal of Molecular Microbiology and Biotechnology* **22**: 71-82.

Ludwig M, Schulz-Friedrich R & Appel J (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: Implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. *Journal of Molecular Evolution* **63**: 758-768.

MacDonald JT, Barnes C, Kitney RI, Freemont PS & Stan G-BV (2011) Computational design approaches and tools for synthetic biology. *Integrative Biology* **3**: 97-108.

Magliery TJ, Anderson JC & Schultz PG (2001) Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of “shifty” four-base codons with a library approach in Escherichia coli. *Journal of Molecular Biology* **307**: 755-769.

Malcata FX (2011) Microalgae and biofuels: a promising partnership? *Trends in Biotechnology* **29**: 542-549.

Mancebo R, Lo PC & Mount SM (1990) Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. *Molecular and Cellular Biology* **10**: 2492-2502.

Marchisio MA & Stelling J (2008) Computational design of synthetic gene circuits with composable parts. *Bioinformatics* **24**: 1903-1910.

Marchisio MA & Stelling J (2009) Computational design tools for synthetic biology. *Current Opinion in Biotechnology* **20**: 479-485.

Marchisio MA (2012) *In silico* implementation of synthetic gene networks. *Methods in Molecular Biology,* Vol. 813 (Weber Wilfried & Fussenegger Martin), pp. 3-21. Humana Press.

Marraccini P, Bulteau S, Cassier-Chauvat C, Mermet-Bouvier P & Chauvat F (1993) A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera *Synechococcus* and *Synechocystis*. *Plant Molecular Biology* **23**: 905-909.

Martínez J, Liu L, Petranovic D & Nielsen J (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Current Opinion in Biotechnology* **23**: 965–971.

Masukawa H, Mochimaru M & Sakurai H (2002) Disruption of the uptake hydrogenase gene, but not of the bidirectional hydrogenase gene, leads to enhanced photobiological hydrogen production by the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120. *Applied Microbiology and Biotechnology* **58**: 618-624.

Matsunaga T & Takeyama H (1995) Genetic engineering in marine cyanobacteria. *Journal of Applied Phycology* **7**: 77-84.

Medema MH, van Raaphorst R, Takano E & Breitling R (2012) Computational tools for the synthetic design of biochemical pathways. *Nat Rev Micro* **10**: 191-202.

Mellitzer A, Weis R, Glieder A & Flicker K (2012) Expression of lignocellulolytic enzymes in *Pichia pastoris*. *Microbial Cell Factories* **11**: 61.

Mermet-Bouvier P, Cassier-Chauvat C, Marraccini P & Chauvat F (1993) Transfer and replication of RSF1010-derived plasmids in several cyanobacteria of the genera *Synechocystis* and *Synechococcus*. *Current Microbiology* **27**: 323-327.

Mermet-Bouvier P & Chauvat F (1994) A conditional expression vector for the cyanobacteria *Synechocystis* sp. strains PCC6803 and PCC6714 or *Synechococcus* sp. strains PCC7942 and PCC6301. *Current Microbiology* **28**: 145-148.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews: Genetics* **11**: 31-46.

Meuser JE, D'Adamo S, Jinkerson RE, Mus F, Yang W, Ghirardi ML, Seibert M, Grossman AR & Posewitz MC (2012) Genetic disruption of both Chlamydomonas reinhardtii [FeFe]-hydrogenases: Insight into the role of HYDA2 in H2 production. *Biochemical and Biophysical Research Communications* **417**: 704-709.

Meyer R (2009) Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62**: 57-70.

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. PCC6803. *Proceedings of the National Academy of Sciences* **108**: 2124-2129.

Montagud A, Navarro E, de Córdoba PF, Urchueguía JF & Patil KR (2010) Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology* **4**: 156.

Montagud A, Zelezniak A, Navarro E, de Córdoba PF, Urchueguía JF & Patil KR (2011) Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal* **6**: 330-342.

Moore M, Klug A & Choo Y (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proceedings of the National Academy of Sciences* **98**: 1437-1441.

Moore RE (1996) Cyclic peptides and depsipeptides from cyanobacteria: A review. *Journal of Industrial Microbiology and Biotechnology* **16**: 134-143.

Morimoto T, Kadoya R, Endo K, Tohata M, Sawada K, Liu S, Ozawa T, Kodama T, Kakeshita H, Kageyama Y, Manabe K, Kanaya S, Ara K, Ozaki K & Ogasawara N (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Research* **15**: 73-81.

Mulder DW, Shepard EM, Meuser Jonathan E, Joshi N, King Paul W, Posewitz Matthew C, Broderick Joan B & Peters John W (2011) Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure* **19**: 1038-1052.

Murtas G (2009) Artificial assembly of a minimal cell. *Molecular Biosystems* **5**: 1292-1297.

Mushegian A (1999) The minimal genome concept. *Current Opinion in Genetics & Development* **9**: 709-714.

Nagai K, Nakaseko Y, Nasmyth K & Rhodes D (1988) Zinc-finger motifs expressed in *E. coli* and folded in vitro direct specific binding to DNA. *Nature* **332**: 284-286.

Nakasugi K, Svenson CJ & Neilan BA (2006) The competence gene, comF, from *Synechocystis* sp. strain PCC 6803 is involved in natural transformation, phototactic motility and piliation. *Microbiology* **152**: 3623-3631.

Nandagopal N & Elowitz MB (2011) Synthetic biology: integrated gene circuits. *Science* **333**: 1244-1248.

Nielsen DR, Leonard E, Yoon S-H, Tseng H-C, Yuan C & Jones Prather K (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metabolic Engineering* **11**: 262-273.

Noireaux V, Maeda YT & Libchaber A (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proceedings of the National Academy of Sciences* **108**: 3473-3480.

Noirel J, Sanguinetti G & Wright PC (2008) Identifying differentially expressed subnetworks with MMG. *Bioinformatics* **24**: 2792-2793.

Nolan T, Hands RE & Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nat Protocols* **1**: 1559-1582.

Oh S-E, Zuo Y, Zhang H, Guiltinan MJ, Logan BE & Regan JM (2009) Hydrogen production by *Clostridium acetobutylicum* ATCC 824 and megaplasmid-deficient mutant M5 evaluated using a large headspace volume technique. *International Journal of Hydrogen Energy* **34**: 9347-9353.

Ohto C, Muramatsu M, Obata S, Sakuradani E & Shimizu S (2010) Production of geranylgeraniol on overexpression of a prenyl diphosphate synthase fusion gene in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* **87**: 1327-1334.

Okamoto S & Ohmori M (2002) The cyanobacterial PilT protein responsible for cell motility and transformation hydrolyzes ATP. *Plant and Cell Physiology* **43**: 1127-1136.

Oliveira P & Lindblad P (2005) LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiology Letters* **251**: 59-66.

Oliveira P & Lindblad P (2009) Transcriptional regulation of the cyanobacterial bidirectional Hox-hydrogenase. *Dalton Transactions* 9990-9996.

Orlowski J, Boniecki M & Bujnicki JM (2007) I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics* **23**: 527-530.

Orzaez D, Monforte AJ & Granell A (2010) Using genetic variability available in the breeder's pool to engineer fruit quality. *GM Crops* **1**: 120-127.

Ouyang Z & Isaacson R (2006) Identification and characterization of a novel ABC iron transport system, *fit*, in *Escherichia coli*. *Infection and Immunity* **74**: 6949-6956.

Ow SY & Wright PC (2009) Current trends in high throughput proteomics in cyanobacteria. *FEBS Letters* **583**: 1744-1752.

Park K-S, Jang Y-S, Lee H & Kim J-S (2005) Phenotypic alteration and target gene identification using combinatorial libraries of zinc finger proteins in prokaryotic cells. *Journal of Bacteriology* **187**: 5496-5499.

Patterson-Fortin LM, Colvin KR & Owttrim GW (2006) A LexA-related protein regulates redox-sensitive expression of the cyanobacterial RNA helicase, *crhR*. *Nucleic Acids Research* **34**: 3446-3454.

Pavletich NP & Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**: 809-817.

Pereira SB, Ow SY, Barrios-Llerena ME, Wright PC, Moradas-Ferreira P & Tamagnini P (2011) iTRAQ-based quantitative proteomic analysis of *Gloeothece* sp. PCC 6909: Comparison with its sheathless mutant and adaptations to nitrate deficiency and sulfur limitation. *Journal of Proteomics* **75**: 270-283.

Pfaffl MW, Tichopad A, Prgomet C & Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**: 509-515.

Pinto F, van Elburg KA, Pacheco CC, Lopo M, Noirel J, Montagud A, Urchueguía JF, Wright PC & Tamagnini P (2012) Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC 6803 mutant lacking a functional bidirectional hydrogenase. *Microbiology* **158**: 448-464.

Podestá JJ, Estrella CN & Esteso MA (1996) Evaluation of the absorption on mild steel of hydrogen evolved in glucose fermentation by pure cultures of *Clostridium acetobutylicum* and *Enterobacter*. *Sensors and Actuators B: Chemical* **32**: 27-31.

Pohorille A & Deamer D (2002) Artificial cells: prospects for biotechnology. *Trends in Biotechnology* **20**: 123-128.

Porcar M & Moya A (2010) Paving the way for synthetic biology-based bioremediation in Europe. *Microbial Biotechnology* **3**: 134-135.

Priefer UB, Simon R & Pühler A (1985) Extension of the host range of *Escherichia coli* vectors by incorporation of RSF1010 replication and mobilization functions. *Journal of Bacteriology* **163**: 324-330.

Puigbò P, Guzmán E, Romeu A & Garcia-Vallvé S (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research* **35**: W126-W131.

Purnick PE & Weiss R (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **10**: 410-422.

Rasmussen S, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF & Bedau MA (2004) Transitions from nonliving to living matter. *Science* **303**: 963-965.

Reich KA (2000) The search for essential genes. *Research in Microbiology* **151**: 319-324.

Reiter L, Kolstø A-B & Piehler AP (2011) Reference genes for quantitative, reverse-transcription PCR in *Bacillus cereus* group strains throughout the bacterial life cycle. *Journal of Microbiological Methods* **86**: 210-217.

Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R & Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**: 940-943.

Roberts RJ, Vincze T, Posfai J & Macelis D (2010) REBASE - a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research* **38**: D234-D236.

Roessler PG & Lien S (1984) Purification of hydrogenase from *Chlamydomonas reinhardtii*. *Plant Physiology* **75**: 705-709.

Rojas Contreras JA, Pedraza-Reyes M, Ordoñez LG, Estrada NU, Barba de la Rosa AP & De León-Rodríguez A (2010) Replicative and integrative plasmids for production of human interferon gamma in *Bacillus subtilis*. *Plasmid* **64**: 170-176.

Roodbeen R & van Hest JCM (2009) Synthetic cells and organelles: compartmentalization strategies. *Bioessays* **31**: 1299-1308.

Sander JD, Zaback P, Joung JK, Voytas DF & Dobbs D (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research* **35**: W599-W605.

Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK & Dobbs D (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research* **38**: W462-W468.

Sanguinetti G, Noirel J & Wright PC (2008) MMG: a probabilistic tool to identify submodules of metabolic pathways. *Bioinformatics* **24**: 1078-1084.

Sarduy ES, Muñoz AC, Trejo SA & Chavéz Planes MdlA (2012) High-level expression of Falcipain-2 in *Escherichia coli* by codon optimization and auto-induction. *Protein Expression and Purification* **83**: 59-69.

Sato S, Shimoda Y, Muraki A, Kohara M, Nakamura Y & Tabata S (2007) A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Research* **14**: 207-216.

Schaeffer DJ & Krylov VS (2000) Anti-HIV Activity of Extracts and Compounds from Algae and Cyanobacteria. *Ecotoxicology and Environmental Safety* **45**: 208-227.

Schloss P & Handelsman J (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biology* **6**: 229.

Schmidt M (2010) Xenobiology: A new form of life as the ultimate biosafety tool. *Bioessays* **32**: 322-331.

Schmitz O & Bothe H (1996) The diaphorase subunit HoxU of the bidirectional hydrogenase as electron transferring protein in cyanobacterial respiration? *Naturwissenschaften* **83**: 525-527.

Schneider D, Fuhrmann E, Scholz I, Hess W & Graumann P (2007) Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *BMC Cell Biology* **8**: 39.

Scholz P, Haring V, Wittmann-Liebold B, Ashman K, Bagdasarian M & Scherzinger E (1989) Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* **75**: 271-288.

Schopf JW (2000) The fossil record: tracing the roots of the cyanobacterial lineage. *The Ecology of Cyanobacteria,* pp. 13-35.

Schopf JW (2006) Fossil evidence of Archaean life. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**: 869-885.

Schütz K, Happe T, Troshina O, Lindblad P, Leitão E, Oliveira P & Tamagnini P (2004) Cyanobacterial H_2 production - a comparative analysis. *Planta* **218**: 350-359.

Service RF (2008) Eyeing oil, synthetic biologists mine microbes for black gold. *Science* **322**: 522-523.

Shankar S & Pillai MR (2011) Translating cancer research by synthetic biology. *Molecular BioSystems* **7**: 1802-1810.

Shetty R, Endy D & Knight T (2008) Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* **2**: 5.

Shibato JS, Agrawal GA, Kato HK, Asayama MA & Shirai MS (2002) The 5'-upstream *cis*-acting sequences of a cyanobacterial *psbA* gene: analysis of their roles in basal, light-dependent and circadian transcription. *Molecular Genetics and Genomics* **267**: 684-694.

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K & Ueda T (2001) Cell-free translation reconstituted with purified components. *Nat Biotech* **19**: 751-755.

Shuler M (2012) Modeling life. *Annals of Biomedical Engineering* **40**: 1399-1407.

Shultzaberger RK, Chen Z, Lewis KA & Schneider TD (2007) Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Research* **35**: 771-788.

Simmons L, Coates C, Clark B, Engene N, Gonzalez D, Esquenazi E, Dorrestein P & Gerwick W (2008) Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 4587-4594.

Singh AK, McIntyre LM & Sherman LA (2003) Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiology* **132**: 1825-1839.

Singh M, Patel S & Kalia V (2009) *Bacillus subtilis* as potential producer for polyhydroxyalkanoates. *Microbial Cell Factories* **8**: 38.

Sjöholm J, Oliveira P & Lindblad P (2007) Transcription and regulation of the bidirectional hydrogenase in the cyanobacterium *Nostoc* sp. strain PCC 7120. *Applied and Environmental Microbiology* **73**: 5435-5446.

Skizim NJ, Ananyev GM, Krishnan A & Dismukes GC (2011) Metabolic pathways for photobiological hydrogen production by nitrogenase- and hydrogenase-containing unicellular cyanobacteria *Cyanothece*. *Journal of Biological Chemistry* **287**: 2777-2786.

Sode K, Tatara M, Ogawa S & Matsunaga T (1992) Maintenance of broad host range vector pKT230 in marine unicellular cyanobacteria. *FEMS Microbiology Letters* **99**: 73-78.

Solé RV, Munteanu A, Rodriguez-Caso C & Macía J (2007) Synthetic protocell biology: from reproduction to computation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**: 1727-1739.

Stano P (2011) Minimal cells: relevance and interplay of physical and biochemical factors. *Biotechnology Journal* **6**: 850-859.

Stevens SE & Porter RD (1986) Heterospecific transformation among cyanobacteria. *Journal of Bacteriology* **167**: 1074-1076.

Subashchandrabose SR, Ramakrishnan B, Megharaj M, Venkateswarlu K & Naidu R (2011) Consortia of cyanobacteria/microalgae and bacteria: Biotechnological potential. *Biotechnology Advances* **29**: 896-907.

Suzuki H, Brown CJ, Forney LJ & Top EM (2008) Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Research* **15**: 357-365.

Suzuki I, Simon WJ & Slabas AR (2006) The heat shock response of *Synechocystis* sp. PCC 6803 analysed by transcriptomics and proteomics. *Journal of Experimental Botany* **57**: 1573-1578.

Suzuki T, Higgins PJ & Crawford DR (2000) Control selection for RNA quantitation. *BioTechniques* **29**: 332-337.

Szybalski W (1974) *In vivo* and *in vitro* initiation of transcription. *Control of Gene Expression,* (Shatkay A. Kohn and A.), pp. 23-24. Plenum Press, New York.

Takamatsu S, Hodges TW, Rajbhandari I, Gerwick WH, Hamann MT & Nagle DG (2003) Marine natural products as novel antioxidant prototypes. *Journal of Natural Products* **66**: 605-608.

Tamagnini P, Troshina O, Oxelfelt F, Salema R & Lindblad P (1997) Hydrogenases in *Nostoc* sp. strain PCC 73102, a strain lacking a bidirectional enzyme. *Applied and Environmental Microbiology* **63**: 1801-1807.

Tamagnini P, Costa JL, Almeida L, Oliveira MJ, Salema R & Lindblad P (2000) Diversity of cyanobacterial hydrogenases, a molecular approach. *Current Microbiology* **40**: 356-361.

Tamagnini P, Leitao E, Oliveira P, Ferreira D, Pinto F, Harris DJ, Heidorn T & Lindblad P (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiol Rev* **31**: 692-720.

Tan LT (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* **68**: 954-979.

Tang L, Jiang R, Zheng K & Zhu X (2011) Enhancing the recombinant protein expression of halohydrin dehalogenase HheA in *Escherichia coli* by applying a codon optimization strategy. *Enzyme and Microbial Technology* **49**: 395-401.

Tavares S, Grotkjær T, Obsen T, Haslam RP, Napier JA & Gunnarsson N (2011) Metabolic engineering of *Saccharomyces cerevisiae* for production of Eicosapentaenoic Acid, using a novel Δ5-Desaturase from *Paramecium tetraurelia*. *Applied and Environmental Microbiology* **77**: 1854-1861.

Taylor S, Wakem M, Dijkman G, Alsarraj M & Nguyen M (2010) A practical approach to RT-qPCR - Publishing data that conform to the MIQE guidelines. *Methods* **50**: S1-S5.

Teste M-A, Duquenne M, Francois J & Parrou J-L (2009) Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Molecular Biology* **10**: 99.

Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A & Heinen E (1999) Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology* **75**: 291-295.

Tomitani A, Knoll AH, Cavanaugh CM & Ohno T (2006) The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *PNAS* **103**: 5442-5447.

Tsien RY (1998) The green fluorescent protein. *Annual Review of Biochemistry* **67**: 509-544.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A & Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**: research0034.

VanGuilder H, Vrana K & Freeman W (2008) Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* **44**: 619-626.

Vermaas W (1996) Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: Principles and possible biotechnology applications. *Journal of Applied Phycology* **8**: 263-273.

Vignais PM, Billoud B & Meyer J (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews* **25**: 455-501.

Villalobos A, Ness J, Gustafsson C, Minshull J & Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7**: 285.

Vioque A (2007) Transformation of cyanobacteria. Vol. 616 (León Rosa, Galván Aurora & Fernández Emilio), pp. 12-22. Springer New York.

Vitousek PM, Cassman K, Cleveland C, Crews T, Field CB, Grimm NB, Howarth RW, Marino R, Martinelli L, Rastetter EB & Sprent JI (2002) Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* **57**: 1-45.

Waks Z & Silver PA (2009) Engineering a synthetic dual-organism system for hydrogen production. *Appl Environ Microbiol* **75**: 1867-1875.

Walde P (2010) Building artificial cells and protocell models: Experimental approaches with lipid vesicles. *Bioessays* **32**: 296-303.

Wang HH (2010) Synthetic genomes for synthetic biology. *Journal of Molecular Cell Biology* **2**: 178-179.

Watanabe S, Ohbayashi R, Shiwa Y, Noda A, Kanesaki Y, Chibazakura T & Yoshikawa H (2012) Light-dependent and asynchronous replication of cyanobacterial multi-copy chromosomes. *Molecular Microbiology* **83**: 856-865.

Whitton BA & Potts M (2000) Introduction to the cyanobacteria. *The ecology of cyanobacteria - Their diversity in time and space,* (Whitton Brian & Potts Malcolm), pp. 1-11. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wijffels RH (2008) Potential of sponges and microalgae for marine biotechnology. *Trends in Biotechnology* **26**: 26-31.

Williams JGK (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetc engineering methods in *Synechocystis* 6803. *Methods in Enzymology* **167**: 766-778.

Wriggers W, Chakravarty S & Jennings PA (2005) Control of protein functional dynamics by peptide linkers. *Peptide Science* **80**: 736-746.

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF & Joung JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protocols* **1**: 1637-1652.

Xie J & Schultz PG (2005) Adding amino acids to the genetic repertoire. *Current Opinion in Chemical Biology* **9**: 548-554.

Xu H, Miao X & Wu Q (2006) High quality biodiesel production from a microalga *Chlorella protothecoides* by heterotrophic growth in fermenters. *Journal of Biotechnology* **126**: 499-507.

Xu WD & McFadden BA (1997) Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* **37**: 95-104.

Xu Y, Alvey RM, Byrne PO, Graham JE, Shen G, Bryant DA & Carpentier R (2011) Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in *Synechococcus* sp. PCC 7002. *Methods in Molecular Biology,* Vol. 684 (Walker John M.), pp. 273-293. Humana Press.

Yadava A & Ockenhouse CF (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infection and Immunity* **71**: 4961-4969.

Yamamoto Y, Fukui K, Koujin N, Ohya H, Kimura K & Kamio Y (2004) Regulation of the intracellular free iron pool by Dpr provides oxygen tolerance to *Streptococcus* mutans. *Journal of Bacteriology* **186**: 5997-6002.

Yamaoka M, Osawa S, Morinaga T, Takenaka S & Yoshida K-i (2011) A cell factory of *Bacillus subtilis* engineered for the simple bioconversion of myo-inositol to scyllo-inositol, a potential therapeutic agent for Alzheimer's disease. *Microbial Cell Factories* **10**: 69.

Yan D, Lu Y, Chen Y-F & Wu Q (2011) Waste molasses alone displaces glucose-based medium for microalgal fermentation towards cost-saving biodiesel production. *Bioresource Technology* **102**: 6487-6493.

Yang XY & McFadden BA (1993) A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC 6803 encodes a Rep protein and replicates by a rolling circle mechanism. *Journal of Bacteriology* **175**: 3981-3991.

Yang XY & McFadden BA (1994) The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid* **31**: 131-137.

Yegorov DY, Kozlov AV, Azizova OA & Vladimirov YA (1993) Simultaneous determination of Fe(III) and Fe(II) in water solutions and tissue homogenates using desferal and 1,10-phenanthroline. *Free Radical Biology and Medicine* **15**: 565-574.

Yoshihara S, Geng X, Okamoto S, Yura K, Murata T, Go M, Ohmori M & Ikeuchi M (2001) Mutational analysis of genes involved in pilus structure, motility and transformation competency in the unicellular motile cyanobacterium *Synechocystis* sp. PCC6803. *Plant and Cell Physiology* **42**: 63-73.

Yoshihara S, Geng X & Ikeuchi M (2002) *pilG* gene cluster and split *pilL* genes involved in pilus biogenesis, motility and genetic transformation in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant and Cell Physiology* **43**: 513-521.

Yu KO, Kim SW & Han SO (2010) Engineering of glycerol utilization pathway for ethanol production by *Saccharomyces cerevisiae*. *Bioresource Technology* **101**: 4157-4161.

Yura K, Toh H & G−o M (1999) Putative mechanism of natural transformation as deduced from genome data. *DNA Research* **6**: 75-82.

Zang XN, Liu B, Liu SM, Arunakumara K & Zhang XC (2007a) Optimum conditions for transformation of *Synechocystis* sp. PCC 6803. *Journal of Microbiology* **45**: 241-245.

Zang XN, Liu B, Liu SM, Sun PN, Zhang XQ & Zhang XC (2007b) Transformation and expression of *Paralichthys olivaceus* growth hormone cDNA in *Synechocystis* sp. PCC 6803. *Aquaculture* **266**: 63-69.

Zehr JP & Ward BB (2002) Nitrogen cycling in the ocean: New perspectives on processes and paradigms. *Applied and Environmental Microbiology* **68**: 1015-1024.

Zelle RM, de Hulster E, Kloezen W, Pronk JT & van Maris AJA (2010) Key process conditions for production of C4-dicarboxylic acids in bioreactor batch cultures of an engineered *Saccharomyces cerevisiae* strain. *Applied and Environmental Microbiology* **76**: 744-750.

Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP & Case CC (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. *Journal of Biological Chemistry* **275**: 33850-33860.

Zhang X-S & Blaser MJ (2012) Natural transformation of an engineered *Helicobacter pylori* strain deficient in type II restriction endonucleases. *Journal of Bacteriology* **194**: 3407-3416.

Zhao L, Pellenz S & Stoddard BL (2009) Activity and specificity of the bacterial PD-(D/E)XK homing endonuclease I-Ssp6803I. *Journal of Molecular Biology* **385**: 1498-1510.

Zhou YJ, Gao W, Rong Q, Jin G, Chu H, Liu W, Yang W, Zhu Z, Li G, Zhu G, Huang L & Zhao ZK (2012) Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *Journal of the American Chemical Society* **134**: 3234-3241.

