Modelling and analysis of biological systems to obtain biofuels

PhD Thesis by

Arnau Montagud Aquino

Advisors:

Pedro Fernández de Córdoba Castellá
Kiran Raosaheb Patil
Javier F Urchueguía Schölzel

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

For comments, questions and requests of the additional files please send word to arnau.montagud@gmail.com.

In memoriam

Joan Montagud Senon

Sollana, 18 Novembre 1954 – El Saler, 11 Febrer 2009

Pujant als cims de boira m'acompanyes,
també descendint a abismes sense llum.

Pep Laguarda i Tapineria, *Cims i abismes*, *Lluna pirata*, 1977

Running over the same old ground,
and how we found, the same old fears.
Wish you were here.

Pink Floyd, *Wish you were here*, *Wish you were here*, 1975

Abstract

This thesis is focused on the construction and uses of genome-scale metabolic models to efficiently obtain biofuels, such as ethanol and hydrogen. As a target organism, cyanobacterium *Synechocystis* sp. PCC6803 was chosen. This organism has been studied as a potential photon-fuelled production platform, for its ability to grow only from carbon dioxide, water and photons. This dissertation verses about methods to model, analyse, estimate and predict the metabolic behaviour of cells. Principal goal is to extract knowledge from the different biological aspects of an organism in order to use it for an industrial relevant objective.

This dissertation has been structured in chapters accordingly organized as the successive tasks that end up building an *in silico* cell that behaves as the carbon-based one. This process usually starts with the genome annotation files and ends up with a genome-scale metabolic model able to integrate –*omics* data. First objective of present thesis is to reconstruct a model of this cyanobacteria's metabolism that accounts for all the reactions present in it. This reconstruction had to be flexible enough as to allow growth under the different environmental conditions under which this organism grows in nature as well as to allow the integration of different levels of biological information. Once this requisite was met, environmental variations could be simulated and their effect studied under a system-wide perspective. Up to five different growth conditions were simulated on this metabolic model and differences were evaluated.

Following assignment was to define production strategies to weigh this organism's viability as a production platform. Genetic perturbations were simulated to design strains with an enhanced production of three industrially-relevant metabolites: succinate, ethanol and hydrogen. Resulting sets of genetic modifications for the overproduction of those metabolites are, thus, proposed. Moreover, functional reactions couplings were studied and weighted to their metabolite production importance. Finally, genome-scale metabolic models allow establishing integrative approaches to include different types of data that help to find regulatory hotspots that can be targets of genetic modification. Such regulatory hubs were identified upon light/dark shifts and general metabolism operational principles inferred. All along this process, blind spots in *Synechocystis* sp. PCC6803 metabolism, and more importantly, blind spots in our understanding of it, are revealed.

Overall, the work presented in this thesis unveils the industrial capabilities of cyanobacterium *Synechocystis* sp. PCC6803 to evolve interesting metabolites as a clean production platform.

Resum

Esta tesis es centra en la construcció i els usos del models metabòlics a escala genòmica per a obtenir eficientment biocombustibles, com etanol i hidrogen. Com a organisme diana, s'elegí el cianobacteri *Synechocystis* sp. PCC6803. Aquest organisme ha segut estudiat com una plataforma de producció nodrida per fotons, per la seva habilitat per créixer a partir únicament de diòxid de carboni, aigua i fotons. Aquesta tesi versa sobre mètodes per a modelitzar, analitzar, estimar i predir el comportament metabòlic de cèl·lules. La principal meta és extreure coneixement del diferents aspectes biològics d'un organisme de manera que s'usen per a un objectiu industrial rellevant.

La tesi ha segut estructurada en capítols organitzats d'acord a les successives tasques que acaben construint una cèl·lula *in silico* que es comporta, idealment, com la que està basada en carboni. Aquest procés generalment comença amb els arxius de l'anotació del genoma i acaba amb un model metabòlic a escala genòmica capaç d'integrar dades – *òmiques*. El primer objectiu de la present tesi és la reconstrucció d'un model del metabolisme d'aquest cianobacteri que tinga en compte totes les reaccions que hi estan presents. Esta reconstrucció havia de ser prou flexible com per permetre la simulació del creixement en les diferents condicions ambientals en les quals aquest cianobacteri creix en la natura, així com permetre la integració de diferents nivells d'informació biològica. Una vegada que aquest requisit fou assolit, es pogueren simular variacions ambientals i estudiar els seus efectes amb una perspectiva de sistema. S'han simulat fins a cinc condicions de creixement en este model metabòlic i les seves diferències han segut avaluades.

La següent tasca fou definir estratègies de producció per a valorar la viabilitat d'aquest organisme com a plataforma de producció. Es simularen pertorbacions genètiques per al disseny de soques amb producció millorada de metabòlits de rellevància industrial: succinat, etanol i hidrogen. Així, es proposen conjunts de modificacions genètiques per a la sobreproducció d'aquests metabòlits. També s'han estudiat reaccions acoblades funcionalment i s'ha ponderat la seva importància en la producció de metabòlits. Finalment, els models metabòlics a escala genòmica permeten establir criteris per integrar diferents tipus de dades que ens ajuden a trobar punts importants de regulació. Eixos centres reguladors, que poden ser objecte de modificacions genètiques, han segut investigats baix canvis dràstics d'il·luminació i s'han inferit principis operacionals del metabolisme. Al llarg d'aquest procés, s'han revelat punts cecs al metabolisme de *Synechocystis* sp. PCC6803 i, el més important, punts cecs en la nostra comprensió d'aquest metabolisme.

En general, el treball presentat en aquesta tesi dona a conèixer les capacitats industrials del cianobacteri *Synechocystis* sp. PCC6803 per a produir metabòlits d'interès, tot sent una plataforma de producció neta i sostenible.

Resumen

Esta tesis se centra en la construcción y usos de los modelos metabólicos a escala genómica para obtener biocombustibles de manera eficiente, como etanol e hidrógeno. Como organismo objetivo, se ha elegido a la cianobacteria *Synechocystis* sp. PCC6803. Este organismo ha sido estudiado como una potencial plataforma de producción alimentada por fotones, dada su capacidad de crecer solamente a partir de dióxido de carbono, agua y fotones. Esta tesis versa acerca de los métodos para modelar, analizar, estimar y predecir el comportamiento del metabolismo de las células. La principal meta es extraer conocimiento de los diferentes aspectos biológicos de un organismo con el fin de utilizarlo para un objetivo industrial pertinente.

Esta tesis ha sido estructurada en capítulos organizados de acuerdo con las sucesivas tareas que terminan con la construcción de una célula *in silico* que se comporta, idealmente, como la que está basada en el carbono. Este proceso suele comenzar con los archivos de anotación del genoma y termina con un modelo metabólico a escala genómica capaz de integrar datos *–ómicos*. El primer objetivo de la presente tesis es la reconstrucción de un modelo del metabolismo de esta cianobacteria que tenga en cuenta todas las reacciones presentes en la misma. Esta reconstrucción tenía que ser lo suficientemente flexible como para permitir el crecimiento en las distintas condiciones ambientales bajo las cuales este organismo crece en la naturaleza, así como permitir la integración de diferentes niveles de información biológica. Una vez que se cumplió este requisito, se pudieron simular variaciones ambientales y estudiar sus efectos desde una perspectiva de sistema. Se han estudiado hasta cinco diferentes condiciones de crecimiento en este modelo metabólico y sus diferencias han sido evaluadas.

La siguiente tarea fue definir estrategias de producción para sopesar la viabilidad de este organismo como una plataforma de producción. Se simularon perturbaciones genéticas para el diseño de cepas con producción mejorada de tres metabolitos de relevancia industrial: succinato, etanol e hidrógeno. Así, se proponen conjuntos de modificaciones genéticas para la sobreproducción de estos metabolitos. Por otra parte, se han estudiado reacciones acopladas funcionalmente y se ha ponderado su importancia en la producción de metabolitos. Finalmente, los modelos metabólicos a escala genómica permiten establecer criterios para integrar diferentes tipos de datos que ayuden a encontrar puntos importantes de regulación que pueden ser objeto de modificación genética. Estos centros reguladores han sido investigados bajo cambios drásticos de iluminación y se han inferido principios operacionales del metabolismo. A lo largo de este proceso, se han revelado puntos ciegos en el metabolismo de *Synechocystis* sp. PCC6803 y, lo más importante, los puntos ciegos en nuestra comprensión de dicho metabolismo.

En general, el trabajo presentado en esta tesis presenta las capacidades industriales de la cianobacteria *Synechocystis* sp. PCC6803 para producir metabolitos de interés, siendo una plataforma de producción limpia y sostenible.

Contents

Preface

If we had to observe your university under the same limitations that bind us in our observation of the Lusitanian aborigines, we would no doubt conclude that humans do not reproduce, do not form kinship groups, and devote their entire life cycle to the metamorphosis of the larval student into the adult professor. We might even suppose that professors exercise noticeable power in human society. A competent investigation would quickly reveal the inaccuracy of such conclusion.

¬Joao Figueira Alvarez, published posthumously in *Xenological Studies*, **22**(4):49-193

Orson Scott Card, *Speaker for the dead*, Tor Books Publisher, 1986

———————————

If you want to travel fast, go alone; but if you want to travel far, go together.

Traditional African saying

Esta és la part de la tesi a la que, sent sincers, més vegades havia desitjat aplegar. Primerament, perquè és senyal de que la feina comença a estar acabada, que aquest camí llarg, ple d'esforços i farcit d'alegries, comença a veure la seva fi. Segonament, perquè és emotiu girar el cap, observar les petjades en la sorra de la platja i donar-te compte del camí que has fet, dels motius pel qual l'has fet i de qui t'acompanya en el viatge. Et permet recordar gent i situacions que tan marcat i gràcies a les quals eres qui eres, com eres i on eres. Com dedicà Pat aquell curt, "*a nuestros recuerdos, porque sin ellos la vida sólo sería futuro*".

Ni esta tesi i ni la meva persona s'haguera pogut desenvolupar sense una família al voltant. No hauria passat res, de res, si dos persones no s'hagueren trobat fa molts anys i decidit continuar el camí juntes. Estes persones són mon pare i ma mare. Tinc la estranya sensació que no hauria tingut interès en la biologia, i per tant no haguera entrat en la Facultat de Biologia, si no fos pels contes de Gerald Durrell o per les xarrades sobre *animalets* de la meva infància. Ells dos, deixant-me tremenda llibertat per les eleccions, varen causar el fet de que un camí comencés, el meu. Així que res d'açò tindria sentit sense ells dos.

Gràcies a mon pare, Joan, que faltà a meitat camí de la realització d'esta tesi i a la memòria del qual voldria dedicar-la. Gràcies per haver sabut donar-me una base humanista i moral, per haver-me encaminat, molt intel·ligentment, a ser la persona que vull ser. Que espere es parega, en alguna cosa, a la persona que a tu t'haguera desitjat que jo fora. Gràcies a ma mare Lidia, per sempre estar ahí, per sempre haver-me donat una espenta quan ho necessitava, per sempre preguntar socràticament sobre les coses, per la llibertat i per la responsabilitat. Gràcies per ésser i estar, este treball en molta part és teu. Continuant amb els aforismes africans, hi ha un que m'encantà en el moment en que el vaig sentir: "*És feina de tota una tribu educar un nen*". La meva tribu, a banda de mons pares, han segut ells: Gràcies a Hernando, Maria Dolores, Alba i Àlex. Gràcies a Elisa, Paco, Angela, Laura i Àngela. Gràcies per haver-me aguantat estos anys i per haver confiat en mi quasi més del que confiava jo. Gràcies també a un membre de la família d'una altra espècie que, de tant en tant, em reconeixia com un igual, Onyx. Tindrà pau allà on estiga.

Recorde haver-ho tingut ben clar des de ben petit: sóc una persona curiosa, m'agrada saber, aprendre. Si la via més llarga de l'aprenentatge era fer una carrera i fer un doctorat, allà aniria. Per això, trobe que els primers agraïments des del punt de vista científic han d'anar a les persones que, estant a l'altra banda de la porta em digueren "*endavant, passa*". Vaig entrar, i fins avui hem compartit àrea científica. Estes persones son Pedro Fernández de Córdoba i Javier Urchueguía. Han segut els meus directors de tesi, els que trobaven temps per a animar-me en moments d'ànims baixos i ficar-me els peus a terra en moments de vols per les altures. Sense ells jo no haguera començat aquest treball ni molt probablement entrat en este camp d'investigació.

I feel the need to deeply thank Kiran Raosaheb Patil. And I think that *deeply* barely scratches the level of acknowledgement I owe to him. It is not an exaggeration to say that without him this thesis would not have seen an end. I thank him for the efforts, the time and thoughts that he altruistically dedicated to these works. I am grateful to him for welcoming me in his group in Copenhagen, and later to bring me the opportunity to tag along him as he moved to Heidelberg. You have truly showed me that science is the place where I want to be and opened new horizons in my scientific career. I am grateful that I can call you friend. In this line, I would also like to thank his colleagues and students Ana Rita, Ana Paula, Sarah, Aleksej, Sergej, and Thomas who, since my stay in Copenhagen, have been my friends and workmates. All of them have built up a working

atmosphere where it's easy to see days go by, enjoy your time, feel at home and be productive at the same time.

Uns agraïments força especials han d'anar a Emilio Navarro. Ell em tutelà, de forma més o menys conscient, en la meva entrada a l'aventura de la tesi doctoral. Em va veure i assistir en eixir de l'ou, créixer i desenvolupar-me com a científic. He tingut el plaer de ser el seu *padawan*. He aprés moltes coses d'ell en molts àmbits de la vida (i de la no científica, també) i tinc la estranya sensació de que hi havia més classes a les que ja no podré assistir. Gràcies per ésser i estar, per les xarrades científiques, ocioses, futbolístiques i banals. Gràcies pel temps compartit en els diferents llocs del món on ens han portat les conferències i estàncies: eixe cigarret a 45°C en Doha, eixa matinà a Hong Kong, eixa pizza a Mumbai, eixa crêpe a París, eixe Celtic-Lakers al TD Garden a Boston, eixa llagosta amb dominó de 9 a la platja cubana. Hi ha tants moments, que em donaria per un llibre. Parlant de llibres, de fet, encara pense que el llibre sobre sociologia de la ciència seria tot un èxit. La vida és curta, però prou llarga com per a que ens tornem a trobar algun dia pel camí, company.

Continuant l'àmbit professional, he d'agrair a la resta de gent d'InterTech. D'entre tots ells, crec que li dec més favors a la meva benvolguda i apreciada Minerva Báguena. No recorde les vegades que ha aconseguit salvar-me d'entre les poderoses ones de la burocràcia. No recorde quantes xarrades hem passat Emilio, ella i jo intentant solucionar el món a la vora d'un got de cafè. Gràcies per sempre estar ahí, a l'hora que calgués, el dia que calgués, per ficar els teus coneixements al servei d'un *novato* com jo. Ademés vull agrair les xarrades als fotònics Albert, Miguel Angel, Mario i, especialment, Carles per intentar explicar-me una i deu vegades els vostres articles sense que els vostres esforços tingueren massa fruit. Gràcies per ensenyar-me que la ciència es tan vasta que puga incloure les vostres investigacions i les meves. Tampoc vull oblidar-me del temps i els riures passats amb Pilar. A demés, una forta abraçada ha d'anar a la gent d'ixa illa on el *realisme màgic* pren cos: Cuba. Raymari, Vinelia, Julián, Garrido, Ramón, Juan Carlos, Dago, Castañeda, Falcón i Pérez. Allí passaren coses que per molt que les explique, la gent pensa que són ficció. De vegades, inclús jo pense que foren ficció. Salud hermanos!

Entrar al món universitari no sol ser un pas senzill. Amb els canvis que solen venir en la tardo-adolescència es junta el canvi d'ambient, d'estil d'ensenyament i de grup de gent. Per sort, hi ha vegades que et trobes a gent que et fa eixe canvi molt més fàcil, com el meus amics i companys de classes de la Facultat de Biologia: Núria, Regina, Caterina, Jessica, Guillem, Òscar, Adrià, Pau, Pepo i Carlos

Edo. Entre tots féreu que les classes més avorrides i les esperes més llargues es feren molt menys pesades al so de les "copetes" i amb converses interessantíssimes. Gràcies per la vostra amistat i confiança. Gràcies per les eixides, per les converses, pels concerts, per intentar explicar-me la diferència entre arbres o entre *pardalets*. Continuant en la Facultat, trobe que les meves ganes de seguir endavant un camí ple de pedres afilades i parets empinades vingueren animades, conscient o inconscientment, per una sèrie de professors: Isabel Fariñas, Juli Peretó, Amparo Latorre, Sari Gil, Mercè Pamblanco, Eladio Barrio, Ignacio Marín, Ismael Mingarro, Joaquín Moreno, i segur que me deixe a algú... Gràcies a tots pels granets d'arena.

Agraïments també han d'anar a la competició estudiantil del iGEM, ja que gràcies a ella vaig conèixer a gent amb una (de)formació comparable a la meva però diferent en direcció: físics, químics, matemàtics, biotecnòlegs, enginyers de vàries escoles. Li dec a Jesús Salgado haver entrat en aquesta aventura. Gràcies Jesús, per aquella conversa en la cafeteria de Farmàcia, sens dubte un dels moments claus de creuers de camins en la meva vida, gràcies per ensenyar-me on estava la porta i lo fàcil que era obrir-la. A més a més, no puc oblidar-me de tots els estudiants dels projectes Valencia iGEM que han passat estius sencers currant. Gràcies per les vostres idees, bogeries, esperances i somnis. Gràcies pels caps de setmana sense dormir a Boston; per eixe "*Hotel California*", versió Gipsy Kings, en l'autobús; per les llagostes, ai, les llagostes... He aprés molt de vosaltres, força més del que heu aprés de mi, em tem. Molt especialment, vull agrair l'haver compartit les primeres passes en el camí científic a la gent del iGEM 2006, a la *vella guardia*, molts d'ells ja doctors, tots ells amics: Diana, Cristina, Caterina, Alberto, Gus, Chevi, Carl, Edo. Foren uns mesos molt intensos, en molts aspectes; ens divertírem molt.

Finalment i no menys que la gent que els ha precedit, voldria donar les meves més sinceres gràcies als meus amics de tota la vida. A eixe grup de germans, eixa *band of brothers*, que ens juntarem sent adolescents i que hem crescut paral·lelament. Gràcies a Jose, Dani, Pat, Luis, Vicente, Pablo, Roberto, Anaïs, Ferran i Ana. Gràcies per ficar els peus a terra, per volar per les altures, pels somnis, pels desitjos, pels fets, pels records, per la confiança, per les crítiques, pels riures. Gràcies per permetre'm usar-vos com a espills i companys en el joc dialèctic de la forja i el creixement personal. Gràcies per haver segut els meus camells culturals i haver-me deixat ser el vostre. Gràcies per acompanyar-me a la recerca de Comala, Macondo i Nova Tanelorn. Per compartir temps i, fent-ho, celebrar que el pas del temps hem reafirma en allò que digué Jacques Bergier:

"*Amb els amics, estic en desacord en tot, menys en lo essencial*". Sou part d'este viatge que ens durà lluny.

Per últim, vull agrair a Beorn Khaariss i a Roark Bumaye (entre altres) ésser ahí, ja que no seria qui sóc sense vosaltres. Gràcies a Tamriel, a Nova Tanelorn i al Regne de Takhisis pel temps passat, per les ensenyances, per la diversió, per l'aprenentatge, per l'evasió, per la vida.

A tots vosaltres, mil gràcies. Espere que us haja marcat tant com vosaltres m'heu marcat a mi. I no us tindre massa en compte si deixeu de llegir esta tesi ací, ja que com digué Machado (i molt ben glosà Robe), a continuació,

"*Veréis llanuras bélicas y páramos de asceta*
–no fue por estos campos el bíblico jardín–:
son tierras para el águila, un trozo de planeta
por donde cruza errante la sombra de Caín."
–Por tierras de España, *Campos de Castilla*, 1912

Aims and Objectives

Interactions of scientific and technological fields are a sensitive matter, but of high importance to society as they have generated much of the advances we have in our world today. In fact, interactions between Mathematical modelling and Biology have already shown their potential and will be of critical importance for our future. This kind of interactions will not be safe from the struggles that regular interfaces have; neither will they have less potential than these have. The following work will study topics at this crossroads: the use of cyanobacteria in order to obtain biofuels, and understand their metabolism as a whole using mathematical models.

Mathematical models in Biology have eased the way researchers treat multi-functionality, non-linearity and complexity, three properties that emerge from even simple biological systems. They are no more, and no less, than tools that allow scientists to gather knowledge out of the tons of information that new high throughput biotechnological techniques produce. Furthermore, the construction of a model is intimately bound to the generation of hypothesis, that drive researchers to perform certain experiment that, in turn, can be used as proof to change parameters and variables in order to have improved models. This iterative process is of critical importance in order to discover and comprehend cellular mechanisms. Models have been used in biotechnology since its very beginning, solving problems in areas like biomedicine, chemicals and food industry. Models, as we will see in present dissertation, can be used to assess, explore and design production strategies for industrially relevant metabolites, such as biofuels.

Metabolism of an organism can be modelled into a network of metabolites and enzymes. This should integrate all biochemical reactions known on that organism in order to have a production platform, a cell factory. This process will start with the genotyping and annotation of a desired genome. These efforts will be the founding stones of future model that will have this information crosschecked with a variety of databases and deep literature surveys. In order to be able to model this set of reactions, it is important to recover and organize information on the boundary parameters that will affect organism growth, influx of substrates and efflux of products.

Unfortunately, having detailed, dynamical simulations is complicated. Considering all the metabolism mechanics would lead to quantitative predictions on cellular dynamics, but lack of knowledge on the intracellular reactions and its parameters makes this approach unfeasible. Constraint-based stoichiometric

models can by-pass this problem assuming intracellular reactions are much faster than organism's ones, *i.e.* growth. Then, internal metabolites would rapidly reach steady state, allowing researchers to ignore the dynamics of intracellular metabolism. Metabolism's state can then be studied through the set of fluxes these intracellular reactions have, later called flux landscape in the dissertation. Comparison of flux landscapes between different environmental perturbations is of high importance in order to infer metabolic knowledge of organisms. Furthermore, study of the organism resilience to genetic changes at the metabolism level is important to evaluate the impact on a given strategy of a knock out, knock in, knock down, etc.

Constraint-based stoichiometric models can also be used to study transcriptomic data, the set of RNA present in the cell upon some environmental or genetic condition, under a new light. The identification of regulatory hubs in the metabolic network outstands as potential design target of a biotechnological strategy. This thesis is devoted to the reconstruction and use of such model aimed at improving biofuels producing strategies in cyanobacterium *Synechocystis* sp. PCC6803.

Objectives.

The principal objectives of this dissertation are the following:

a) *Reconstruct a genome-scale metabolic model for* Synechocystis *sp. PCC6803.*

Cyanobacterium *Synechocystis* sp. PCC6803 has been targeted as a potential photon-fuelled production platform. Genome-scale metabolic models are a pre-requisite to study metabolism potentials as well as perturbations.

b) *Study environmental and genetic variations under a system-wide perspective.*

Cyanobacterium *Synechocystis* sp. PCC6803 will not be a desirable production platform if researchers do not know its behaviour under perturbations. Genome-scale metabolic model allows the integrative study of the entire metabolism under such variations, like the different growth modes of this cyanobacterium or mutations performed on its genome. This may allow detecting which variations are critical to the well-being of this organism.

c)   Define production strategies.

Three metabolites have been identified as desirable products from this organism: succinate, ethanol and hydrogen. Their theoretical production limits and functional coupled reactions need to be assessed. Enhanced-production mutants need to be studied and discussed under a system-wide perspective.

d)   Integrate different levels of information

Finally, strategies need to be performed in order to efficiently integrate different levels of biological information: genome, transcriptome, metabolome and fluxome. Genome-scale metabolic models allow establishing integrative approaches to such include different data and infer novel conclusions for the preceding step.

Thesis outline.

The first pages serve as introduction to the three scientific fields on which this PhD thesis stands on: systems biology, cyanobacterial biology and energy engineering. In the interphase of these three scientific fields is where this work is located: *systems biology* offers us techniques that will be applied to *cyanobacteria* in order to obtain *biofuels*. The following chapters will encompass different consecutive aspects of this project.

Chapter 2 reviews the efforts for the reconstruction of a genome-scale metabolic model of *Synechocystis* sp. PCC6803. Reconstruction is explained in detail, two versions of our model are presented and connectivity analyses are done. Biomass composition is described in detail as well as some ideas on how to improve it. Furthermore, our model is located among the state of the art of *Synechocystis* sp. PCC6803 metabolic models.

Chapter 3 is devoted to the studies of the flux landscapes of *Synechocystis* sp. PCC6803 and their variance upon environmental conditions changes. Flux balance analysis is used in order to have these flux simulations. Functional constraints are explained, simulations are described and variances among different environmental situations are clarified. Flux variability analysis is presented on changes in growth condition.

Genetic perturbations are studied in **Chapter 4**, where essential genes are evaluated as well as mutations that lead *Synechocystis* sp. PCC6803 to be a production platform of value-added metabolites, such as succinate, ethanol and

hydrogen. Single, double and triple knock out strategies are studied and theoretical production limits are assessed in the light of these overproducing strains.

Chapter 5 is devoted to the study of functional coupled reactions. Couplings across growth conditions, in the photosynthesis pathway and in the biofuels production pathways are studied. For these we have taken advantage of flux coupling analysis that studies the potentiality of reactions in pairs. Blocked reactions, as reactions that do not carry flux under a given set of conditions, are also identified.

In a first step to leap the simulation of constraint-based metabolic models to the chemostat conditions of *Synechocystis* sp. PCC6803 growth, **Chapter 6** bears an analysis on growth conditions of this cyanobacterium. Five variables that define different aspects of three main growth conditions were assessed and statistically evaluated after 4699 data growth points were recovered.

Integration of transcriptomic data on the metabolic network is the scope of **Chapter 7**. This data is analysed under the light of metabolic-reactions connectivity and of flux coupling connectivity. Regulatory hubs upon shift of light regime are identified and explained in a system-wide integrative manner.

Finally, **Chapter 8** gathers conclusions among all chapters and draws some milestones that need to be considered if we expect to have a *Synechocystis* sp. PCC6803 production platform for biofuels, specifically hydrogen.

Scientific contributions

- ***Synechocystis* sp. PCC6803 metabolic models.**

Navarro, Emilio, Arnau Montagud, Pedro Fernández de Córdoba, and Javier F Urchueguía: **Metabolic flux analysis of the hydrogen production potential in *Synechocystis* sp. PCC6803**. *International Journal of Hydrogen Energy* 2009, **34**:8828-8838.

Montagud, Arnau, Emilio Navarro, Pedro Fernández de Córdoba, Javier F Urchueguía, and Kiran Raosaheb Patil: **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium.** *BMC Systems Biology* 2010, **4**:156.

Montagud, Arnau, Aleksej Zelezniak, Emilio Navarro, Pedro Fernández de Córdoba, Javier F Urchueguía, and Kiran Raosaheb Patil: **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

- **Other *Synechocystis* sp. PCC6803 systems biology works.**

Pinto, Filipe, Karin A van Elburg, Catarina C Pacheco, Miguel Lopo, Josselin Noirel, Arnau Montagud, Javier F Urchueguía, Phillip C Wright, and Paula Tamagnini: **Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC6803 mutant lacking a functional bidirectional hydrogenase.** *Microbiology (Reading, England)* 2011, **158**:448-464.

Lopo, Miguel, Arnau Montagud, Emilio Navarro, Isabel Cunha, Andrea Zille, Pedro Fernández de Córdoba, Pedro Moradas-Ferreira, Paula Tamagnini, and Javier F. Urchueguía: **Experimental and modelling analysis of *Synechocystis* sp. PCC6803 growth.** *Journal of Molecular Microbiology and Biotechnology* 2012, XXX.

Montagud, Arnau, Pedro Fernández de Córdoba, and Javier F Urchueguía: ***Synechocystis* sp. PCC6803 metabolic models study for the enhanced production of biofuels.** *Manuscript in preparation.*

Montagud, Arnau, Gilberto Reynoso-Meza, Pedro Fernández de Córdoba, Javier Sanchis, and Javier F Urchueguía: **Multiobjective evolutionary algorithm allows genome-scale fluxes simulations without biological constraints.** *Manuscript in preparation.*

Reynoso-Meza, Gilberto, Arnau Montagud, Pedro Fernández de Córdoba, Javier F Urchueguía, and Javier Sanchis: **Simulation of the *Synechocystis* sp. PCC6803**

metabolic behaviour using stoichiometric representations and multi-objective evolutionary algorithms. *Manuscript in preparation.*

- **Synthetic biology works.**

Rodrigo, Guillermo, Arnau Montagud, Alberto Aparici, Maria Cristina Aroca, Minerva Baguena, Javier Carrera, Carlos Edo, et al.: **Vanillin cell sensor**. *IET Synthetic Biology* 2007, **1**:74.

Delás, Joaquina, Meritxell Notari, Jaume Forés, Joaquín Pechuan, Manuel Porcar, Emilio Navarro, Arnau Montagud, et al.: **Yeast cultures with UCP1 uncoupling activity as a heating device**. *New Biotechnology* 2009, **26**:300-6.

Vilanova, Cristina, Angeles Hueso, Carles Palanca, Guillem Marco, Miguel Pitarch, Eduardo Otero, Juny Crespo, et al.: **Aequorin-expressing yeast emits light under electric control.** *Journal of Biotechnology* 2011, **152**:93-5.

Gamermann, Daniel, Arnau Montagud, Pablo Aparicio, Emilio Navarro, Julián Triana, Francisco R Villatoro, Javier F Urchueguía, and Pedro Fernández de Córdoba. **A modular synthetic device to calibrate promoters.** *Journal of Biological Systems* 2012, **20**:1-24.

- **Book.**

Montagud, Arnau, Emilio Navarro, Pedro Fernández de Córdoba, and Javier F Urchueguía: *Introduction to Synthetic Biology*. Valencia: PoliCLICK; 2009

- **Bachelor thesis supervised.**

Siurana, Maria. **Strategies for hydrogen optimization in photosynthetic bacteria**. Universitat Politècnica de València. 2011

- **Other biotechnological works.**

Belda, Eugeni, Laia Pedrola, Juli Peretó, Juan F Martínez-Blanch, Arnau Montagud, Emilio Navarro, Javier F Urchueguía, Daniel Ramón, Andrés Moya, and Manuel Porcar: **Microbial diversity in the midguts of field and lab-reared populations of the European corn borer** *Ostrinia nubilalis*. *PLoS ONE* 2011, **6**:e21751.

Pedrola, Laia, Arnau Montagud, Eugeni Belda, Juan F Martínez-Blanch, Emilio Navarro, Manuel Porcar, Juli Peretó, Andrés Moya, Daniel Ramón, and Javier F Urchueguía: **Media optimization for ethanol production with designed yeast mutants**. *Manuscript in preparation.*

"Remember, the enemy's gate is *down*."
Molo, Soup, Vlad, Dumper, and Crazy Tom all
laughed. They remembered, too.
And Ender also laughed. It was funny. The adults
taking all this so seriously, and the children
playing along, playing along, believing it too until
suddenly the adults went too far, tried too hard,
and the children could see through their game.

Orson Scott Card, *Ender's game*, Tor Books
Publisher, 1985

1

Introduction

Where PhD applicant uses first pages of this dissertation to describe the scientific fields he has roamed around and on the fields' interaction this thesis is located on: metabolic models, cyanobacteria and biofuels.

Contents of this chapter are based on the following journal article:

- Montagud et al **Synechocystis sp. PCC6803 metabolic models study for the enhanced production of biofuels.** *Manuscript in preparation.*

1.1 Systems biology, metabolic models and fluxes.

Systems biology is the scientific area of Biology that tries to exploit, gather, organize and, finally, understand as many information as possible from a given biological system. Systems biology has been defined as *"the study of the interactions between the components of biological systems, and how these interactions give rise to the function and behaviour of that system"* (Snoep and Westerhoff, 2005). In taking this holistic point of view, this scientific area goes against the reductionist tide that has engulfed Molecular Biology since its explosion when DNA was first modelled (Watson and Crick, 1953) and the Central Dogma of Molecular Biology was published (Crick, 1970). In fact, Sauer, Heinemann and Zamboni stated the importance of studying the cell metabolism under a system-level approach in 2007: *"Rather than a reductionist viewpoint (that is, a deterministic genetic view), the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously, and by rigorous data integration with mathematical models. Such a system-wide perspective (so-called systems biology) on component interactions is required so that network properties, such as a particular functional state or robustness, can be quantitatively understood and rationally manipulated"* (Sauer et al., 2007).

Systems biology works trying to integrate as much information as possible, prefers all-inclusive explanations rather than local knowledge. Researchers can think of the complexity, hard work and emerging properties that this may lead to, but citing Hiroaki Kitano: *"It is often said that biological systems, such as cells, are* complex systems*. A popular notion of complex systems is of very large numbers of simple and identical elements interacting to produce* complex *behaviours. The reality of biological systems is somewhat different. Here large numbers of functionally diverse, and frequently multifunctional, sets of elements interact selectively and nonlinearly to produce coherent rather than complex behaviours"* (Kitano, 2002). This idea of *simplicity from complexity* or *coherence from complexity* was not new in literature, see Palsson (2000), and is a recurring explanation for experimental results. In addition, this idea differentiates biological processes from other complex systems.

In present dissertation, we have worked on the integrative view of metabolism. We had the need to understand metabolic physiology quantitatively in order to design and optimize biofuel-production bioprocesses. To this end, we used mathematical models. A model is a simplified description, especially a mathematical one, of a system or process, to assist calculations and predictions

(Soanes and Stevenson, 2010). This simplification is achieved defining relevant factors that will be included and others that will be de-emphasised. The criteria of which factors are selected must be chosen in accordance to the purpose of the model, because as the saying goes, *a model is neither good nor bad, a model is either useful or useless*. Obviously, different purposes will need different models and, thus, different factors may be emphasized. This richness has widened the field of mathematical modelling of cells and cell populations.

Our goals were to understand metabolism and how this could help us develop strategies for enhanced biofuels production. Hence, considerable attention has been devoted towards the mathematical description of metabolic function or metabolic models.

Metabolic model reconstruction.

Metabolism of an organism can be modelled into a network of metabolites and enzymes. This should integrate all biochemical reactions for which we have proof and/or evidence of presence in the desired cell. This information can be retrieved from different databases, genomic annotations and literature surveys. First studies, like the ones from Fell and Small (1986), Mavrovouniotis et al. (1992) and Savinell and Palsson (1992b), did not take into account all reactions in the metabolism, mostly because in the previous years of the genome sequencing boom having a whole genome sequenced was somewhat closer to a dream than to reality. This caused researchers to denote as *genome-scale metabolic models* when all the reactions from a genome annotation are included in a model. Additionally, the advent of metabolic studies of *Escherichia coli* (Edwards and Palsson, 2000b; Varma and Palsson, 1993a, 1993b), *Haemophilus influenzae* Rd (Edwards and Palsson, 1999) and *Saccharomyces cerevisiae* (Förster et al., 2003) as well as the dissemination of *genome-sequencing projects* (and the huge cut on sequencing prices) established a context where it was feasible to build genome-scale metabolic model of a desired organism (Figure 1.1).
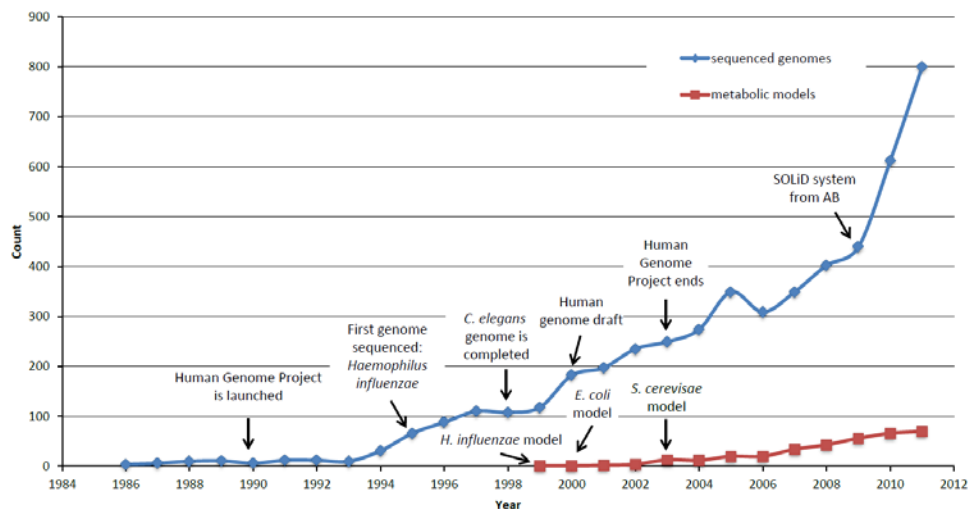
Figure 1.1 - Sequenced genomes versus genome-scale metabolic models. Hits count in Scopus database of terms "genome sequencing" and "genome-scale metabolic model OR network". Major milestones are depicted. Note that between 2008 and end of 2011 genome sequencing costs per megabase of DNA dropped 1000 times (Wetterstrand, 2012). For a complete review see Metzker (2010).

A proper metabolic reconstruction project should start with the annotated genome. This information should be processed gathering up all reactions that are documented for this organism, for a very instructive work on this see Förster et al. (2003). This draft has to be iteratively corrected with alternative information from several databases and from the study of the network in order to avoid the presence of false positive (and false negative) reactions. Recently, some works have described this correction process in detail, trying to establish a common protocol (Feist et al., 2009; Thiele and Palsson, 2010). This process ends up with a network of metabolites and enzymes that encode biochemical reactions taking place within the cell. Nodes usually represent metabolites and edges usually represent reaction rates or metabolic fluxes. We usually differentiate between internal fluxes as reactions occurring within cells and exchange fluxes as exchanges between cells and their environment (like drains of substrates and formation of products). It is important to gather information on these exchange fluxes, which act as boundary parameters, in order to be able to model this set of reactions, as we will see in the following section. Therefore, physiological requirements, under which there is relevant growth of the organism, had to be gathered. This information usually comes from deep diving into the bibliomic information of this organism, that is, searching for experiments where growth substrate is controlled and, ideally, where cell's products are measured. For instance, metabolites that are ingested or secreted by the organism are of crucial

knowledge, as they hint that some pathway are present that uptake or produce that metabolite or a metabolite derived from it. **Chapter 2** is devoted to the reconstruction of genome-scale metabolic models.

Once the metabolic model is completed and curated, this network can be the departing point of several other studies, such as:

- *Network connectivity.*

- *Comparative evolutionary studies to find patterns among organisms.*

- *Phenotypic phase plane analyses.*

- *Strain improvement: gene deletions and additions*

- *Network robustness studies.*

- *Study regulatory constraints.*

- *Reactions' flux analysis and variability.*

In present work, we were interested in simulate the organism's metabolic behaviour, that is to know each reaction's flux. The set of flux values, later defined as *flux vector*, characterizes *metabolic state of cells*, its *phenotype* or *flux landscape*, *i.e.* the metabolism's behaviour at a given time. Metabolic flux is a fundamental determinant of cell physiology and the most critical parameter of a metabolic pathway (Orth et al., 2010; Stephanopoulos and Stafford, 2002). Accurate quantification of pathway fluxes therefore is an important goal in metabolic engineering, especially where the aim is to convert as much substrate to desired metabolic product via strain improvement.

Perfect scenario would be to have detailed and dynamical simulations that would consider overall cell mechanics and would lead to quantitative predictions. Sadly, lack of knowledge on the intracellular reactions and its parameters hampers efforts to have a cell-wide dynamical model, even though some efforts have been focused on that direction (Gerdtzen et al., 2004; Nielsen and Villadsen, 1992; Rizzi et al., 1997).

In order to have a genome-scale metabolic model able to simulate flux landscapes attention has been drawn on *constraint-based stoichiometric models* (Llaneras and Picó, 2008). These models can obtain complete flux information making some assumptions. One of the algorithm researchers use to work with constraint-based stoichiometric models is *flux balance analysis* (FBA), described hereafter.

Constraint-based stoichiometric description of metabolism is not new, since 1986 (Fell and Small, 1986) and 1988 (Clarke, 1988), and peaking in the mid-1990s (Mavrovouniotis et al., 1992; Savinell and Palsson, 1992a, 1992b; Schuster and Schuster, 1993; Stephanopoulos et al., 1998; Varma et al., 1993a, 1993b; Varma and Palsson, 1993a, 1993b, 1994a), several studies paved the way to describe metabolic flux distributions and cell growth. This approach has yielded accurate and valuable information about how microbial cells utilize their metabolic fluxes and optimize their growth rates. There have been several initiatives towards development of the metabolic flux model in order to gain information about metabolic physiology of the culture in a quantitative manner (Bonarius et al., 1997; Schilling et al., 1999; Stephanopoulos et al., 1998; Varma and Palsson, 1994b). In fact, in 2007, a metabolic flux analysis technique was reported to be applied to hydrogen production using growth of *Escherichia coli* on glucose (Manish et al., 2007).

Flux balance analysis.

Flux balance analysis (FBA) is a widely used approach for studying biochemical networks, in particular the genome-scale metabolic network reconstructions. In present dissertation, we have used FBA in order to study cell growth and metabolite productivity. FBA retrieves information on reactions' fluxes and is, thus, very insightful in locating pathways and reactions where flux has changed upon a genetic or environmental variation. It is, hence, a valuable tool to study flux landscapes.

We will start considering metabolites' concentration changes in time:

$$\frac{dX_i}{dt} = S_{ij} \cdot v_j, \qquad \forall i \in M, \forall j \in N$$

The concentrations of all metabolites are represented by the vector X_i, with length m. The flux through all of the reactions in a network is represented by the vector v_j, which has a length of n. Metabolic reactions are represented as a stoichiometric matrix (S_{ij}) of size $m \times n$. Systems being composed of m compounds (rows) and n reactions (columns), entries in each column are the stoichiometric coefficients of the metabolites participating in a reaction. There is a negative coefficient for every metabolite consumed and a positive coefficient for every metabolite that is produced. A stoichiometric coefficient of zero is used for every metabolite that does not participate in a particular reaction. S_{ij} is a sparse matrix because most biochemical reactions involve only a few different

metabolites (Orth et al., 2010). Typically, the number of reactions (n) exceeds the number of metabolites (m).

In order to solve this underdetermined set of equations, which results in an infinite number of solutions of a continuous variety, we make an assumption that allows this dynamic problem to be a static one: we consider steady state of the system. Consequently, intracellular metabolites' concentrations are not allowed to change in time, and are consequently *balanced*. On the contrary, extracellular metabolites can change in time as they are not balanced and these will be the uptake of substrates and formation of products. We have now reached the point where the phrase *flux balance analysis* makes sense (Figure 1.2).

Steady state assumption is widely accepted in systems biology as it allows researchers to avoid the need of detailed dynamic descriptions of metabolism that account for kinetics and regulation of individual enzymes, which has proven difficult to obtain. Additionally, steady state can be justified by the fact that metabolic transients are more rapid than both cellular growth rates and the dynamic changes in the organism's environment. Metabolism typically has transients that are shorter than a few minutes and thus metabolic fluxes are in a quasi-steady state relative to growth and typical process transients (Varma and Palsson, 1994a).

Thus, our system would be:

$$\sum_{i=1}^{M}\sum_{j=1}^{N} S_{ij} \cdot v_j = 0,$$

As the stoichiometry matrix S_{ij} is known from the genome annotation, we treat the metabolic reaction fluxes, v_j, as the unknown quantities that need to be determined.

Researcher can constrain the elements of the vector v_j for the definition of irreversible ($v_{j,irr}$) and reversible ($v_{j,rev}$) reactions, adding further information to the problem:

$$v_{j,irr} \in R^+$$

$$v_{j,rev} \in R$$

Again, further assumptions have to be made if we want to solve this underdetermined set of equations. As typically $n > m$, a plurality of solutions exist and researchers can find an infinite number of combinations on the

metabolic fluxes distribution. Two factors affect this *space of solutions*. On one hand, the stoichiometric matrix *S*: if a reaction is added or deleted the *space of solutions* is affected. On the other hand, some reactions are constrained to certain biologically relevant values in order to build up the *allowable solution space*, such as:

$$v_{j,const} \in R, \qquad v_{min} < v_{j,const} < v_{max}$$

$$v_{j,uptake} \in R, \qquad v_{min} < v_{j,uptake} < v_{max}$$

Where two set of equations have been established, $v_{j,const}$, constrained metabolic reactions, and $v_{j,uptake}$, uptake reactions, known metabolic requirements for growth, which were bound by experimentally determined values from the literature. This set of constraints is the reason behind some researchers name this kind of modelling *constraint-based metabolic models* (Llaneras, 2010). The result of these two effects is that the *allowable space of solutions* is now an *n*-dimensional polytope. We have defined the *possible space of solutions* –with the help of the stoichiometry matrix– and reduced it to the *biologically feasible space of solution* –thanks to the set of constraints.
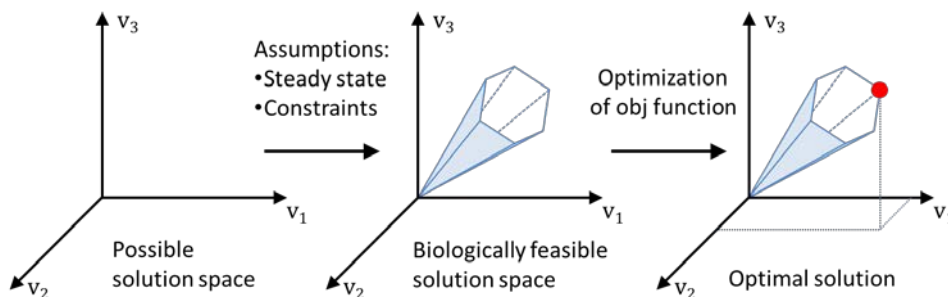
Figure 1.2 - The conceptual basis of constraint-based modelling. After Figure 1 of Orth et al. (2010). Without constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints are imposed by the stoichiometric matrix S and biological constraints are applied to a network, it defines a biologically feasible solution space, an *n*-dimensional polytope. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space.

The last step on this algorithm is to make such a rule that we can identify a single relevant result inside this polytope that makes sense in our work. This

objective will be to *optimize for an objective function*. Through this, FBA can identify a single optimal flux distribution that lies on the allowable solution space, on an edge of the n-dimensional polytope. First studies on the matter were conducted by chemical engineers whose goal was to know a given *product yield*: the maximum amount of product that can be generated per unit of substrate, $Y_{p/s}$. Given that reaction's stoichiometry was already used in generating matrix *S*, it was reasonable that *stoichiometric yields*, $Y_{X/s}$, were used instead as output of the system. Stoichiometric yields differ from product yields in that multiple biomass components (*e.g.* lipids) and biomass precursors (*e.g.* amino acids) have to be quantified in proportion to each other (Feist and Palsson, 2010).

In the early years of flux studies in metabolism several objective functions were used and discussed (Pramanik and Keasling, 1997; Savinell and Palsson, 1992a, 1992b; Varma and Palsson, 1994a) until the term *biomass synthesis,* or *composition* or *equation*, (and, more importantly, the idea behind) became standard. Biomass composition is a *stoichiometric yield for biomass*, a theoretical construct that is defined by the molar composition of a mole of biomass. In FBA context, it is considered as a drain of precursors or building blocks into a hypothetical biomass component. Flux through biomass composition reaction, being the biomass formation rate, is directly related to growth of the modelled organism (Stephanopoulos et al., 1998). Thus, the goal of our optimization is cellular growth: researchers consider that cells will tend to grow as much as resources allow them. According to Varma and Palsson (1994a), there are three reasons behind choosing optimal metabolic behaviour.

First, optimization of growth usually confers a growth advantage that is selected for by natural processes. Thus, the determined flux distribution can be used to describe experimental results and to predict how cells will respond to changes in their environment.

Second, one can determine the maximum allowable production capability of a particular strain and appropriate metabolic engineering strategies can be developed for strain design.

Third, a bioprocess engineer can seek to optimize process design and control, for instance, through optimal medium formulation.

It can be argued that, from a pure Darwinian perspective, organisms will tend to adapt to a given environment, not *just* grow. Nevertheless, it is no less true that in a closed environment under controlled conditions, small experimental times –in

the order of hours or days— and little room for environmental perturbations, fight for resources will be quite limited to competition in terms of growth rates.

Nowadays there are a set of software tools that ease the work with this algorithm. In present work we have extensively used OptGene software, that can be used offline (Patil et al., 2005; Rocha et al., 2010) and also online through BioMet Toolbox (Cvijovic et al., 2010). Other software are COBRA Toolbox (Becker et al., 2007), and latest version of Pathway Tools (Latendresse et al., 2012) that allows the solving of linear programming problems such as *flux balance analysis*.

The reconstruction and the *flux balance analysis* allow us to have a metabolic network that incorporates all information of cell's metabolism, a growth simulation of this network and a flux landscape that depends on the network architecture and the biological constraints that we have applied to it. Now we will see how we can study perturbations on this network and its flux landscape.

Method of minimization of metabolic adjustment.

Once we have this metabolic network and its growth and fluxes descriptions, we can go further into the study of its environmental and genetic variations. Environmental changes can be simulated with the study and variations of the constraints imposed on the system as well as changing the nutrients of the system, *i.e.* alternative carbon sources, as we study in **Chapter 3**. This has been investigated for a handful of reasons, namely media composition optimization (Diamant et al., 2009). On the same scope, researchers have used metabolic models in order to investigate on knock outs, and knock ins, of genes studying their flux variations. These works have been used for several objectives, for instance to study knock outs with increased production of a given industrially relevant product.
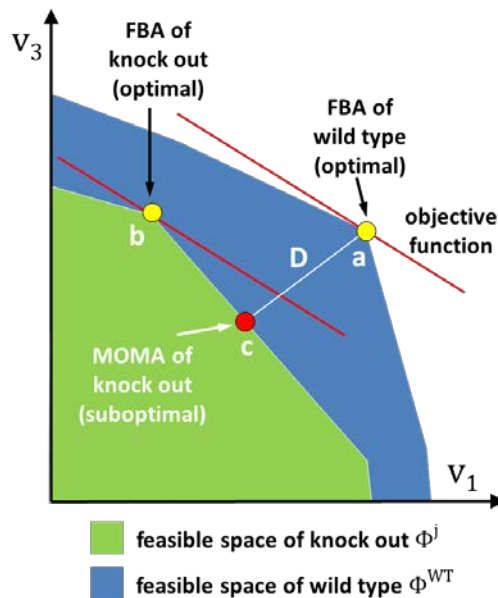
Figure 1.3 - The optimization principles underlying FBA and MOMA. Inspired in Figure 1A from Segrè et al. (2002). A schematic 2D depiction of the feasible space for the wild type (Φ^{WT}) and for mutant of flux j (Φ^{j}) is represented by the blue and superimposed green polygons. The coordinates are two arbitrary representative fluxes, an extremely simplified version of the multidimensional flux space. The solution to the FBA problem is the point that maximizes the objective function (red line). An optimal FBA prediction can be computed both for the wild type (a) and for a knock out (b). The alternative MOMA knockout solution (c), calculated through quadratic programming, can be thought of as a projection of the FBA optimum onto the feasible space of the mutant (Φ^{j}). The mutant FBA optimum and the corresponding MOMA solution are, in general, distinct.

First knock out studies were done altering the stoichiometric matrix (deleting the reaction column) according to the reaction whose deletion was simulated. FBA was subsequently simulated with the new matrix and with the assumptions this algorithm needs in order to find a single, meaningful solution. Segrè, Vitkup and Church were concerned on the assumption that mutant organisms will tend to grow optimally forced by genetic evolution. Hence, they addressed this point by introducing the *method of minimization of metabolic adjustment* (MOMA) as the same argument may not be valid for genetically engineered knock outs or other bacterial strains that were not exposed to long-term evolutionary pressure, although the assumption of optimality for a wild-type bacterium is justifiable (Segrè et al., 2002). With this method, they claim to have a better approximation of mutants' flux states than with FBA simulations.

MOMA is based on the same stoichiometric constraints as FBA, but relaxes the assumption of optimal growth flux for the mutants. MOMA tests the hypothesis that the corresponding mutant flux distribution is better approximated by the flux minimal response to the perturbation than by the optimal one, as can be seen in Figure 1.3.

This algorithm searches for a point in the feasible space of the solutions space of the knock out (Φ^j) that has minimal distance from a given flux vector w. The goal is to find the vector x $\in \Phi^j$ such that the Euclidean distance

$$D(w, x) = \sqrt{\sum_{i=1}^{N} (w_i - x_i)^2}$$

is minimized.

Chapter 4 is devoted to the use of these algorithms on our genome-scale metabolic models of *Synechocystis* sp. PCC6803.

1.2 Metabolic models of cyanobacteria.

We have centred our work in cyanobacterium *Synechocystis* sp. PCC6803. An organism whose first strain was sampled in a freshwater lake in Wisconsin, USA in 1949 and is part of Pasteur Culture Collection, previously known as Berkeley Culture Collection (Allen and Smith, 1969; Gugger and Biological Resource Center of Institut Pasteur, 2011a, 2011b). Cyanobacteria are commonly accepted to have played a crucial role in the Precambrian phase by contributing oxygen to the atmosphere (Schopf, 2000). This strain of a somewhat diverse genus (Swingley and Blankenship, 2008) has been a good candidate for biotechnology uses and molecular biology studies not only for its ability to be naturally transformed by exogenous DNA (Kufryk et al., 2002), but also for its ability to grow photoautotrophically (out of light and CO_2) or its potential as hydrogen producer (Houchins, 1984; Tamagnini et al., 2007), among other reasons. Interestingly, cyanobacteria are held to be the evolutionary ancestors of chloroplasts under the endosymbiont hypothesis (Douglas, 1998; Raven and Allen, 2003) and since these organisms are believed to be the ones that changed the ancient anoxygenic environment to oxygenic by photosynthesis (Kim et al., 2008; Schopf, 2000), many scientists have also used cyanobacteria as an ideal model organism to study adaptation to various abiotic environmental conditions (Douglas, 1998).

Studies on this organism had a boost when molecular tools started to be developed in the 1980s. In scarcely ten years extensive strain designations, restriction endonucleases, and cyanobacterial plasmids and cloning vectors had been developed (Houmard and Tandeau de Marsac, 1988) and allowed the components to be ready for assembly into sophisticated systems for genetic analysis of cyanobacteria (Thiel, 1994). Since then, *Synechocystis* sp. PCC6803 has become a cyanobacterial model organism for its robust growth characteristics. At present, there is probably no other cyanobacterium that has been investigated in such detail, making it an interesting organism for biotechnological applications (Gutthann et al., 2007) such as heterologous production of metabolites like isoprene (Lindberg et al., 2010), poly-beta-hydroxybutyrate (Wu et al., 2001), alcohols (Angermayr et al., 2009), bio-hydrogen (Tamagnini et al., 2007) and others biofuels (Liu et al., 2010). *Synechocystis* sp. PCC6803 is thus an attractive candidate for developing a clean and sustainable platform for biotechnological processes aimed at value-added products formation (Ducat et al., 2011b).

Synechocystis sp. PCC6803 genome was sequenced and annotated in 1995 by Kazusa's laboratory (Kaneko et al., 1996, 1995). Genome-wide transcriptional microarrays, first used to study stress responses in *Saccharomyces cerevisiae* in 2000 (Gasch et al., 2000), were developed in 2001 for this cyanobacterium (Hihara et al., 2001; Suzuki et al., 2001) and works have been published since then using this transcriptomic technique (Foster et al., 2007; Gill et al., 2002; Hihara et al., 2003; Huang et al., 2002b; Hübschmann et al., 2005; Kanesaki et al., 2002; Schmitt and Stephanopoulos, 2003; Singh et al., 2009, 2003; Summerfield and Sherman, 2008; Suzuki et al., 2006; Wang et al., 2004; Yamaguchi et al., 2002; Zhang et al., 2008). Finally, and following the work of pioneering metabolic studies of *Haemophilus influenzae* Rd (Edwards and Palsson, 1999), *Escherichia coli* (Edwards and Palsson, 2000b; Varma and Palsson, 1993a, 1993b) and *Saccharomyces cerevisiae* (Förster et al., 2003), several groups have developed metabolic models for this organism (Knoop et al., 2010; Kun et al., 2008; Montagud et al., 2010, 2011; Navarro et al., 2009; Shastri and Morgan, 2005; Yang et al., 2002b; Yoshikawa et al., 2011). **Chapter 2**, devoted to the reconstruction of genome-scale metabolic models, bears a comparison of all the metabolic models developed up to date for *Synechocystis* sp. PCC6803.

These metabolic models have been used for several goals, principally, database building, growth simulation and production yield evaluation. They have broadened the knowledge on *Synechocystis* sp. PCC6803 metabolism and have helped researchers focus on this organism's potential. Nowadays it is unthinkable

to consider a metabolic engineering project that is related to *Synechocystis* sp. PCC6803 and fails to use one, or more, of these metabolic models.

1.3 Biofuels and Biohydrogen.

In present dissertation, we will focus on using *Synechocystis* sp. PCC6803 metabolic models in order to find and evaluate strategies to design strains that have an enhanced production of a set of interesting metabolites. These metabolites will be related to biofuels, mainly ethanol and hydrogen. Motivation for this comes from the involvement of our research group in a hydrogen-producing consortium in the frame of a EU FP6 project: BioModularH2 (2005).

Biofuels overview.

Today, 83% of the United States of America's energy mix (U.S. Energy Information Administration, 2011b) and 78% of Europe's (European Commission, 2011b) comes from carbon-rich fossil fuels: oil, natural gas, and coal. With demand increasing worldwide, existing oil reserves could peak within 20 years (Zucchetto and National Research Council, 2006), followed by natural gas and coal sources. This limitation on the availability of fossil fuels does not come alone; energy dependence among countries as well as political turmoil in geographical areas with such fossil fuels reserves has compelled science-funding agencies from developed countries to sponsor research into alternative energies. In addition to this scenario, exploration of such alternatives has been spurred on by several international agreements aiming at the reduction of CO_2 emissions (United Nations, 1998, 2009), as 82% of emission of greenhouse gases are related to energy production solely in USA (U.S. Energy Information Administration, 2011a).

As it was outlined in the *Clean Development Mechanisms* of the *Kyoto Protocol* (United Nations, 1998), several alternatives of CO_2 emission-free energy production are already available, for instance, nuclear and renewable energies, as can be seen in figures 1.4 and 1.5. These solutions have been greatly developed in industrial and residential sectors, but not as much in transportation where energy is required to be in most cases in the form of a fuel. The transition from our dependence on *oil* for *transportation* to a cleaner and more efficient energy source is an issue of great importance. In fact, studies show that oil consumption in the transportation sector accounts for up to 66% of the net oil consumption in USA (U.S. Energy Information Administration, 2011a) with similar figures in Europe (European Commission, 2011a).
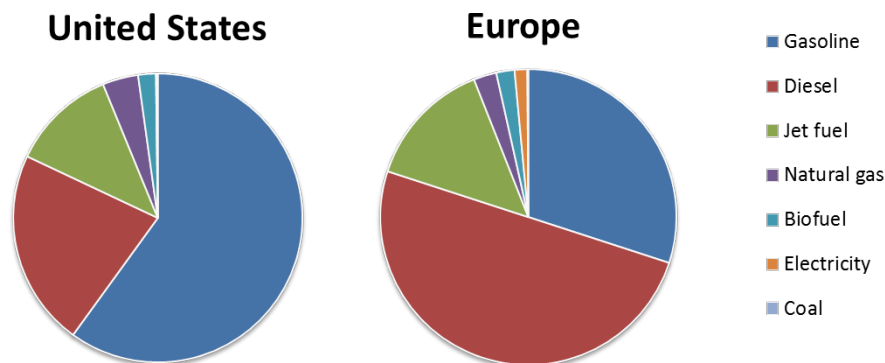
United States   **Europe**

Legend:
- Gasoline
- Diesel
- Jet fuel
- Natural gas
- Biofuel
- Electricity
- Coal

Figure 1.4 - Distribution of energy consumption in the transportation sector. Data from USA and European countries members of OECD from *BP Review of World Energy* (BP p.l.c., 2011).

It makes sense to turn to such a promising area as biotechnology in order to try to find a clean, wireless, transportation sector-friendly source of energy. Biotechnology has boosted its fruits in recent years due to a set of factors, like consolidation of genome-scale sequencing, high-throughput techniques, increased applied focuses and multidisciplinary education programs, like international Genetically Engineered Machine competition (iGEM) (2012). Thus, bearing in mind the advances that biotechnology has driven in health (Hockemeyer et al., 2011; Ro et al., 2006), bioremediation (Hong-Bo et al., 2008; de Lorenzo, 2006) or food (Prust et al., 2005), it is not strange that several biotechnological candidates have been proposed to partially substitute oil as energy demand. In fact, biofuels like biodiesel or ethanol seem to be in the short-term suitable candidates to substitute partially the oil demand. They share distribution systems with conventional fuel and even current engines are compatible to them (Shinnar and Citro, 2006). Both alternatives have advantages but even if all USA's corn and soybean production were dedicated to biofuels production, it would only meet 12% of gasoline demand and 6% of diesel demand (Hill et al., 2006). Active projects in the production processes are under development, for instance, trying to widen the spectrum of usable biomass, thus extending the source to cellulose in order to enhance ethanol production (Abril and Abril, 2009; Hamelinck et al., 2005).

Efficiency of the different production processes of fuels could be improved with these energy vectors. Nevertheless, if we are to satisfy the increasing energy demand and to be able to produce energy at a lower environmental impact, even more resources have to be invested in biotechnology and alternatives have to be found in a medium or long term. At this point, hydrogen appears as a very

promising candidate as a future energy vector (Momirlan and Veziroglu, 2005). Hydrogen is relevant to all of the energy sectors, like transport, residential, and industrial. It can provide storage options for base-load (geothermal), seasonal (hydroelectric) and intermittent (photovoltaic and wind) renewable resources and when combined with emerging decarbonisation technologies, can reduce the climate impacts of continued fossil fuel utilization (Elam et al., 2003).

USA and Europe have invested, and are investing, large amounts of resources on this alternative. In fact, the *Committee on Alternatives and Strategies for Future Hydrogen Production and Use* of the USA's *National Research Council* recommended in 2004 increased funding emphasis on the *Carbon dioxide-free energy technologies* section, with aims like the increase exploratory and fundamental research on hydrogen production by photobiological, photoelectrochemical, thin-film solar, and nuclear heat processes (Committee and National Research Council, 2004). Hydrogen can be produced from renewable sources and used in a wireless manner. Additionally, it has a clean combustion and an extraordinary energy density (142 MJ kg^{-1} for H_2 against 42 MJ kg^{-1} for oil) which allows weight reduction per heat unit when compared to other biofuels.

In spite of that, reports set the horizon of a clean hydrogen-based economy in 2015 at the soonest, that is to start introducing production facilities, adapting distribution networks and developing engines. The goal is to have a fully developed market at 2025 (United States Government, 2011). This is mainly due to two factors: on one side, the lack of competitiveness of fuel cell vehicles and hydrogen compared to conventional (*e.g.* gasoline and diesel) fuel vehicles and hybrid gasoline electric vehicles. On the other side, fuel cells cost that are still a factor of 10 to 20 times too expensive, have short durability, and have low energy efficiency for light-duty-vehicle applications (Committee and National Research Council, 2004). An additional technical problem is also the fact that hydrogen gas is one of the lightest gasses known in nature (0.08988 g L^{-1} at 0°C and 101.325 kPa), making it difficult to store large volumes of it.
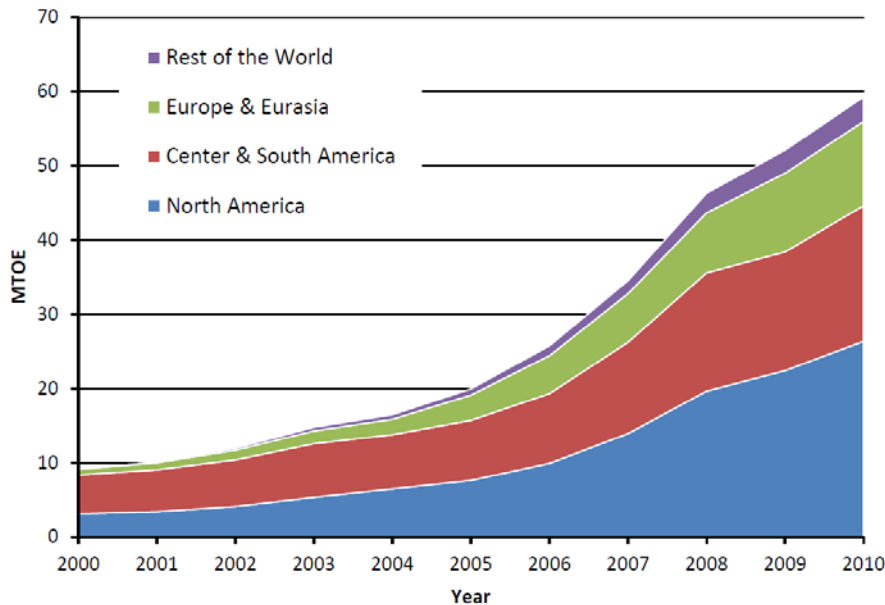
Figure 1.5 - Biofuels production in millions of tonnes oil equivalent. (MTOE). Data from USA and European countries members of OECD from *BP Review of World Energy* (BP p.l.c., 2011).

Hydrogen production.

Work is under way to overcome problems and, in fact, hydrogen is being produced nowadays at industrial level, mainly due to steam reforming of natural gas (Das and Veziroglu, 2008). This process is costly efficient and is neither renewable nor clean, as it has associated sulphur and CO_2 emissions. In fact, this is a good example of hydrogen production using fossil fuels as source, as most of the processes from fossil fuels generate approximately twice as much CO_2 per amount of hydrogen produced, in moles (Agrawal et al., 2007). Therefore, society would benefit if we could be able to use renewable and clean sources to produce hydrogen. In order to reach that primary goal several alternatives are under development:

- *Electrolysis from a renewable source of energy.*

Electrolysis using carbon-free electricity sources like wind or solar is one of the simplest ways of producing hydrogen and is currently the only way to produce large quantities of hydrogen without emitting the traditional undesirable by-products associated to fossil fuels. Its main limitation factor is the electricity cost for its production from renewable sources (Department of Energy, 2007).

43

- *Reforming biomass and wastes.*

Reforming biomass and wastes is another way to produce hydrogen efficiently (Kroposki et al., 2006). It is done mainly by three processes (gasification, pyrolysis/reforming and high-pressure water). It has been demonstrated that it is costly effective, in fact it is the most promising economic route for converting *syngas* into transportation fuels (Milne et al., 2002). Nevertheless, major limitations of this application are the availability of feedstock, efficient and durable catalysts for gas conditioning and efficient integration of processes.

- *Solar thermal water splitting.*

This production process probably represents the most efficient way to produce hydrogen as solar energy is directly used to split water molecules (Greene et al., 2009; Perret, 2005). The main drawback of this technology is that it requires very high temperatures (*circa* 2500 *℃*). This temperature need can be reduced using some additional chemical transformations that retrieves minimal residues. Currently commercial plants of this kind are under development in order to make it costly efficient (Greene et al., 2009).

- *Photoelectrochemical water splitting.*

This strategy allows hydrogen production from light in a one-step process splitting the water with an illuminated semiconductor immersed in an aqueous solution (Carty et al., 1981; Fujishima and Honda, 1972). The potential efficiency of this process is between 10 and 20% of the light irradiation on the semiconductor but the main disadvantage is that the required combination of physical, chemical, structural and economic properties is so restricted that no known material satisfies all of them in order to trigger an industrial production.

- *Photobiological water splitting.*

This biotechnological alternative is based on the light absorption and charge separation reactions in the photosynthesis performed by some organisms, with an absorbance capacity of 40 - 45% of solar energy (BioModularH2, 2005; Das and Veziroglu, 2001, 2008). Such an efficiency is relevant as theoretical estimations states that cultivation surface of 500 km^2 would be enough to produce enough energy to fulfil the world's transportation needs (Turner et al., 2008). Nowadays, works have obtained transient efficiencies of up to 13% by using a combination of processes (Melis, 2007). If this production could be sustained in time, this hydrogen production strategy would be cost efficient. Therefore, the reliability of this method of hydrogen production strongly depends on the future development

of this area, which is where this thesis comes to play. A more extended review of the different biological hydrogen production can be found in (Hallenbeck, 2002).

Thus, change to a clean economy based on hydrogen implies addressing several questions related to its production strategy and methods. Furthermore, hydrogen also raises technical issues related with its transportation and storage that are neither covered in this chapter nor in this thesis, but that have to be addressed and partially solved (Committee on Alternatives and Strategies for Future Hydrogen Production and Use and National Research Council, 2004; Turner et al., 2008). Most probably, the final global solution will come from a combination of technical solutions from different technologies.

Biological hydrogen production.

Biological hydrogen production is a known capacity of some microorganism since 1970 (Postgate, 1970) that involves two kind of enzymes: hydrogenases, refer to Tamagnini et al. (2002) for a detailed classification of them, and nitrogenases, involved in the nitrogen fixation metabolism. The main bio-hydrogen production mechanisms can be grouped in two categories:

a) Fermentation

Many anaerobic organisms like *Enterobacter aerogenes* (Fabiano and Perego, 2002), *Bacillus licheniformis* (Kalia et al., 1994), *Rhodopseudomonas palustris* P4 (Oh et al., 2002) or *Clostridium* sp. (Taguchi et al., 1996) are able to produce hydrogen as a by-product of the dark fermentation of sugars, amino acids and fatty acids. Process can be direct, anaerobically consuming sugars, or stepwise, first generating sugar complexes from other carbon sources like CO_2 and then fermenting them anaerobically. Fermentation has the advantage that it can use a wide range of substrates and that can be coupled to other industrial processes, like agricultural activities. This strategy is being successfully used industrially for hydrogen production (Maniatis, 2010; Solazyme Inc., 2012).

However, this mechanism has two drawbacks that need to be addressed. The first one is related to the cost of the substrate used as energy source for hydrogen. The most energetically suitable one is glucose, but its cost is a limiting factor to be a cost-efficient alternative. Nevertheless, this problem can be tackled in the same way that it has been solved for ethanol production: extending suitable sources to agricultural residues like lignocelluloses, which could provide a sustainable feedstock, would cut down prices, and would not divert stocks from human food chain. The second problem is related to the stoichiometry of

fermentation metabolism. Fermentation has a very low efficiency in terms of hydrogen production from glucose, in fact from a metabolic analysis of the reactions involved in the process each molecule of glucose can produce theoretically 4 molecules of hydrogen, while 2 - 3 molecules are produced under lab conditions (Thauer, 1976). This represents that just 40% (in terms of standard enthalpy of combustion) of the glucose energy input can be recovered in the form of hydrogen as theoretical maximum, and only 20 - 30% has been achieved experimentally. The rest of the energy is stored in cell's biomass (meaning growth of the organism) and other fermentation products like acetic acid, alcohols or lactic acid.

b)   *Photon-fuelled hydrogen production*

Photosynthetic organisms are able to use solar energy to make their own feedstock to live. Among them, cyanobacteria and some algae are able to use this energy to produce hydrogen as a transient by-product. The process of hydrogen production is based on the photosynthetic transport chain, which is shown in Figure 1.6. Schematically, the process consists of the absorption of photons by the photosystem II (PSII), which energy is used to split the water molecules into two protons and oxygen releasing two electrons to the electron transport chain. This electron transport chain, through a set of redox reactions, leads these electrons to the photosystem I (PSI) which can route them back through plastoquinone (PQ) into the *cyclic electron flow* or can drive the electrons to a NADPH dehydrogenase (FNR) into the *non-cyclic electron flow*. In several steps of the photosynthesis, namely PSII, PQ and cytochrome *b6-f* complex, protons are pumped from cytoplasm to the thylakoid lumen, where they can be transported back to cytoplasm, dissipating the proton gradient, through ATP synthase, which uses this energy to phosphorylate ADP to ATP. The cyclic process generates a proton gradient (thus ATP) and the non-cyclic process generates proton gradient and NADPH. Furthermore, depending on the actual cell requirements of NADPH (also known as redox potential of the cell) and ATP (major energy vector of the organism) the ratio of cyclic-to-non-cyclic photosynthesis is controlled by the cell metabolic regulation.

Figure 1.6 - Photosynthetic transport chain. In cyanobacteria, photosynthetic transport chain share several enzymes with the respiratory transport chain (Figure1.7), this fact allows the possibility of using electrons from glucose or other substrates in order to produce hydrogen

Cyanobacteria have a [Ni-Fe] hydrogenase, which produces hydrogen from NADPH according to the reaction

$$NADPH + H^+ \rightarrow NADP^+ + H_2$$
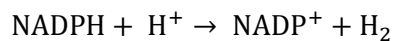
Green algae commonly have a [Fe-Fe] hydrogenase, which is linked to a ferredoxin enzyme of PSI taking directly the electrons from the electron transport chain and producing hydrogen. Recently there have been some efforts on heterologous hydrogenases' cloning in cyanobacteria. For instance, bacterial [FeFe] hydrogenase from *Clostridium acetobutylicum* has been expressed in the cyanobacterium *Synechococcus elongatus* together with the necessary maturation systems, and, under anoxic conditions, the resulting light-dependent H_2 evolution increased over 500 times compared to H_2 evolution from the endogenous [NiFe] hydrogenase (Ducat et al., 2011a).

However, the main limitation in order to have a cost-effective hydrogen production by photosynthesis is the fact that hydrogenases' activity is highly repressed by the oxygen, which is an inherent product of photosynthesis. Several projects (Barstow et al., 2011; Stapleton and Swartz, 2010) deal with this problem trying to find a naturally-occurring oxygen-tolerant hydrogenase, like the one from *Ralstonia eutropha* H16 (Burgdorf et al., 2005), or to design a synthetic one and expressing it in cyanobacteria or green algae, as in many projects like BioModularH2 (2005).

In nature, this problem has been addressed by some cyanobacteria separating in space the region in which oxygen is produced from the area in which hydrogen is produced. For instance, in cyanobacteria genus *Nostoc*, hydrogen is produced as a residual product of nitrogen fixation by nitrogenase in the heterocyst region. In order to use this cyanobacterium for industrial use, the main problem is that efficiencies are limited as nitrogen fixation is energetically expensive. Nevertheless, a review of works in this line could be found in Lopes Pinto et al. (2002).

An alternative to face this problem has been to temporally separate the hydrogen production and the oxygen production. One of the most successful strategies following this alternative has been developed by Melis et al. (2000). Their strategy is based on sulphur deprivation of algae, in such a way that the organism will not be able to normally repair the structural units of photosystem II, namely D1, which is a protein that requires a fast turnover (approximately 30 minutes). This effect produces a decrease of water splitting, and, thus, a reduction in the O_2 production associated to photosynthesis up to a level that is lower than (or, at least, much closer to) the oxygen consumed by the respiration system, oxidative phosphorylation, Figure 1.7. This mechanism is able to produce a (micro)anaerobic environment which activates the hydrogenase enzyme producing hydrogen during a relatively long time span if we are to compare it with previous results. This mechanism, however, relies on the availability of a substrate, such as succinate, NADH, acetate and the like, to be able to reduce the oxygen produced by photosynthesis and provide electrons to hydrogen production. In fact, we can say that the biomass produced during the active photosynthesis phase (without sulphur deprivation) is then fed to the electron transport chain to produce hydrogen, turning this hydrogen evolution process much more industrially interesting.

There is another possibility consisting on a combination of both basic strategies, in what is called *photofermentation*. In this alternative the residues from fermentation are used to feed photosynthetic bacteria in an anaerobic media, resulting in a continuous hydrogen production (Melis and Melnicki, 2006). We could also consider that sulphur deprivation strategy is closer in design to photofermentation, as one could consider this approach as a stepwise coupling of biomass generation and hydrogen production.
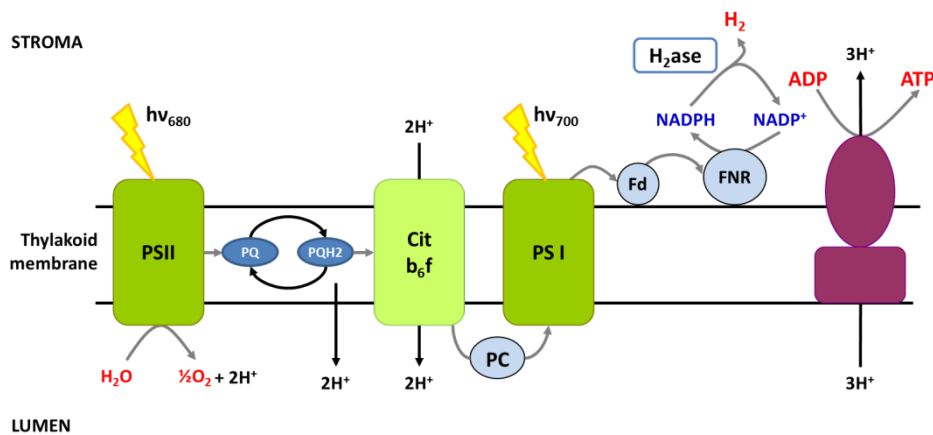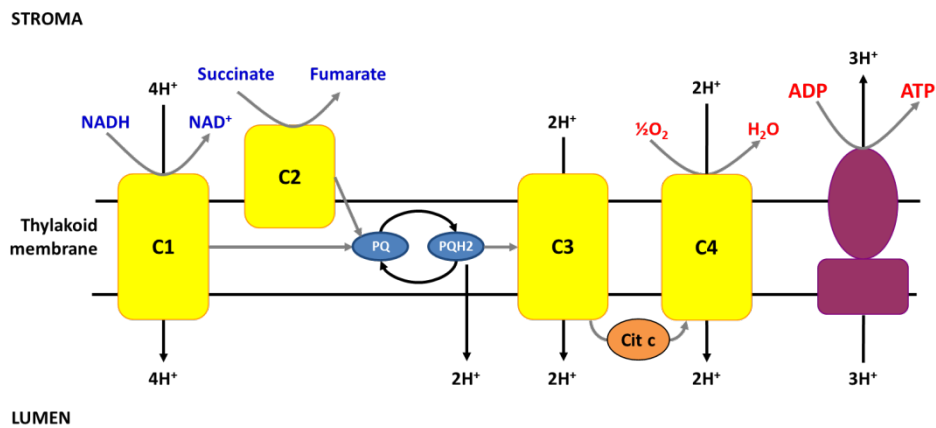
Figure 1.7 - Oxidative phosphorylation. In cyanobacteria, photosynthetic transport chain share several enzymes with the respiratory transport chain, such as *PQ* and *cit* b_6f that performs C3 functions.

Microbial pathway engineering has been mainly applied to industrial processes for biosynthesis of products of high economic value, which is not yet the case for hydrogen, and is one of the motivations of this thesis. Mathematical modelling of hydrogen metabolism is therefore important to evaluate maximum theoretical product yield and to understand the interactions between biochemical energy, carbon fixation and assimilation pathways from a system-wide perspective.

50

Science is knowledge which we understand so well
that we can teach it to a computer; and if we don't
fully understand something, it is an art to deal
with it.
Donald Knuth, 1974 Turing Award Lecture,
Communications of the ACM **17**, 12:668

2

Reconstruction of cyanobacterial metabolism

Where PhD applicant explains how he faced the disperse, divergent, non-structured data sources that the selected organism has among the literature, and how he gathered this information up to get the first genome-scale metabolic model of this organism.

Parts of the contents of this chapter are based on parts of the following journal articles:

- Montagud et al. **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

- Montagud et al. **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium.** *BMC Systems Biology* 2010, **4**:156.

2.1 Introduction.

In recent years, the need for a clean, sustainable and efficient chemical production platform for biofuels has trapped considerable interest of the society. Future energy requirements demand a sustainable alternative for the use of fossil fuels, to restrict further global warming and pollution. For instance, advances have been achieved in the design and implementation of microbial processes for the first, second and third generation biofuels. However, the design of a clean industrial production platform stands as a cornerstone for the next step in green biotechnology, focused on getting energy from light source and relieving air of CO_2. Such a strategy offers the much-needed avoidance of diversion of energy from human food chain or the recollection of vegetal leftover from the fields.

One of the best places to search for such alternative solutions is naturally occurring microorganisms, where the goal of trapping light and CO_2 has already been achieved several millennia back. Photoautotrophs harvest energy from photons through the photosynthesis and shape carbon skeletons out of atmospheric CO_2 through Calvin cycle's carbon fixation. Nowadays, the capability of taming such organisms to the production of energy-rich compounds for use as biofuels, *e.g.*, alcohols or hydrogen, is a goal that is within our reach (Tamagnini et al., 2007). In the search for such an ideal system, focus has been set on algae and cyanobacteria. These unicellular organisms have been targeted due to their ease of cultivation, little nutritional demands and tolerance – its wide range of habitats include aquatic (saltwater and freshwater), terrestrial and extreme environments (including frigid lakes of the Antarctic or hot springs). *Synechocystis* sp. PCC6803 is a cyanobacterial model organism for its robust growth characteristics and is naturally transformable (Kufryk et al., 2002). *Synechocystis* sp. PCC6803 genome was sequenced, annotated and made publicly available in 1995 (Kaneko et al., 1996, 1995) and has been the target of some metabolic modelling effort, as we will later discuss.

Metabolic models at the genome-scale are one of the pre-requisites for rational metabolic engineering approaches as well as for eventually developing synthetic cell factories (Barrett et al., 2006; Patil et al., 2004). Towards this end, a variety of tools/algorithms are available (Patil et al., 2004) and can be applied on metabolic models. Among them, we can find *flux balance analysis* (FBA) (Edwards et al., 1999; Varma and Palsson, 1993a), *minimization of metabolic adjustments* (MOMA) (Segrè et al., 2002), *regulatory on-off minimization* (ROOM) (Shlomi et al., 2005), *flux variability analysis* (FVA) (Mahadevan and Schilling, 2003) and *metabolic control analysis* (MCA) (Kacser and Burns, 1973; Rapoport et al., 1974).

Genome-scale metabolic network reconstruction is, in essence, a systematic assembly and organization of all the reactions, which build up the metabolism of a given organism. This work has been of great interest in the post-genomic era as it offers an opportunity to systematically analyse *omics* datasets in the context of cellular metabolic phenotypes and thereby allows researchers to gain insights into the operational principles of the cell factories. Genome-scale metabolic modelling approach has been applied to a diversity of organisms in a variety of conditions in the context of metabolic engineering, pathway re-routing and systems biology in general. This chapter presents such manually curated genome-scale reconstruction for *Synechocystis* sp. PCC6803 and its comparison to other works on *Synechocystis* metabolic models.

2.2 Genome-scale metabolic models of *Synechocystis* sp. PCC6803.

Reconstruction process.

A complete literature examination, including databases, biochemistry textbooks and the annotated genome sequence, was needed in order to extract the current state of the art on known metabolic reactions within the metabolic network of *Synechocystis* sp. PCC6803. For a thorough overview of the process of metabolic model reconstruction, refer to very instructive work by Forster et al. (2003) as well as review by Feist et al. (2009).

In detail, the reconstruction started with download of *Synechocystis* sp. PCC6803 genome and annotation files (Kaneko et al., 1996, 1995) from NCBI Entrez Genome repository as of date 10 of September of 2008 (NCBI, 2011). These files were organized with Pathway Tools software (Karp et al., 2002) in order to build an object-oriented database of all the genes, proteins and metabolites presents in the organism. Proteins supposed to mediate reactions were retrieved from this database and, if possible, connected together. We complemented this draft list of reactions with queries to a set of public databases, most of them specific on genomes (Karp et al., 2005), pathways (Kanehisa et al., 2008), enzymes (Bairoch, 2000; Chang et al., 2009) or proteins (The UniProt Consortium, 2007), some even specific to cyanobacteria (Nakao et al., 2010). See Figure 2.1 for an overview of the process. However, the lack of quality had to be considered as a major drawback of some of the databases: false positives, false negatives as well as wrongly annotated objects may hinder efforts of collecting accurate data (Weise et al., 2006). Consequently, manual reconstruction by detailed inspection of each and every reaction, ideally relating to published articles, biomass equation

based on metabolic building blocks (such as amino acids and nucleotides), consistency and integrity of the network is a pre-requisite for creating a high quality and useful metabolic model (Feist et al., 2009).
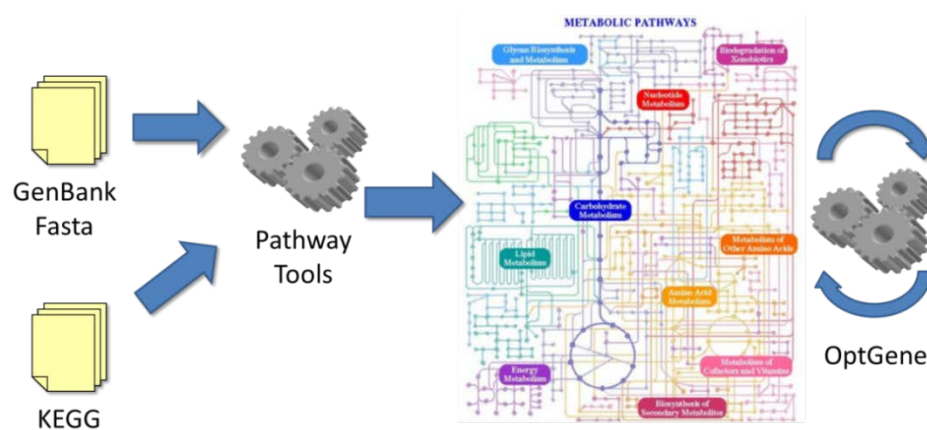
Figure 2.1 - Overview of the process of reconstruction of the genome-scale metabolic model. Metabolic map from KEGG database (www.genome.jp/kegg).

Reactions' EC numbers and stoichiometry were checked and verified with the help of the Enzyme nomenclature database (Bairoch, 2000) and KEGG pathway database (Kanehisa et al., 2008). No lumped reactions were made and photosynthesis was described as a set of 19 reactions, thus enabling the tracing of the corresponding fluxes. Reactions were elementally balanced except for protons, so that chemical conversions were coherent. In some of the reactions present in these databases, metabolites were reported in a non-specific form (*e.g. an alcohol*). This is insufficient for metabolic model simulation and, so, corresponding organism-specific metabolites had to be identified (Förster et al., 2003). Additionally, in a large number of reactions cofactors were not completely clarified: an enzyme being capable of using NADH or NADPH or both. In the latter, two reactions were included in the reconstructed metabolic network. Determination of reversibility of the reactions was assisted by specific enzyme databases, like BRENDA (Chang et al., 2009). If no conclusive evidence was reported, reactions were set to be reversible.

Parts that characterize *Synechocystis* network, like the incomplete TCA cycle (Pearce and Carr, 1967; Vazquez-Bermudez et al., 2000) [recently found to be complete through succinic semialdehyde by Zhang and Bryant (2011)], the presence of the glyoxylate shunt (Yang et al., 2002c), the interconnected photosynthesis and oxidative phosphorylation (Löffelhardt and Schmetterer,

1999), the amino acid transport reactions (Labarre et al., 1987) or the cyclic and non-cyclic electron transport related to these latter processes (Albertsson, 2001; Allen, 2002; Camacho Rubio et al., 2003), were accounted for in detail. In fact, in the reconstruction of the metabolic model, many reactions were found to be necessary for the production of the monomers, precursors or building blocks, that are considered in the biomass equation but which have no corresponding enzyme coding gene assigned (see Table 2.1). In consequence, many genes that were not annotated before should be considered, as they code for enzymes that should be present to allow the formation of biomass. For instance, enzymes malyl-CoA lyase and isocitrate lyase were not allocated in the annotation of the genome albeit their activities have been measured (Pearce and Carr, 1967; Yang et al., 2002c) and their presence is necessary to complete the glyoxylate shunt; consequently, they were included in the model. This was an example of information that was absent from databases, but was present in literature and was vital for the correct simulation of metabolism.

As transcriptomic analysis was one of the tasks we wanted to do with this model, gene presence and correspondence to a cognate reaction were also accounted for. At the end of the reconstruction process, four kinds of relationships were present in the database: reaction with cognate genes, reactions that needed to be included in the model in order to have metabolic precursors in the network (with no assigned genes), non-enzymatic reactions that have no related gene, and genes described in the annotations but with no assigned function.

Versions.

The product of this reconstruction process was a set of reactions that encompass all the known metabolite conversions that take place in *Synechocystis* sp. PCC6803. We have performed two reconstructions of this cyanobacterium's metabolism, *i*Syn669 in late 2009 and *i*Syn811 in mid-2010.

- *iSyn669*

The resulting network, *i*Syn669, consists of 882 metabolic reactions and 790 metabolites. A total of 669 genes were included, to which 639 reactions were assigned (see Additional file 1.1 for details); the difference between the number of genes and assigned reactions is due to the presence of considerable number of protein complexes (*e.g.* photosynthetic or respiratory activities) and isoenzymes. Interestingly, when bidirectional reactions are converted to unidirectional, the number of reactions is extended to 1045. Reactions with no cognate genes are

also present in *i*Syn669, 20 passive transport reactions and 39 chemical conversions (not mediated by enzymes) were included. Additionally, 79 reactions were included on the basis of biochemical evidence or physiological considerations, but currently with no annotated Open Reading Frame (ORF). *i*Syn669 genome-scale metabolic model is available in Additional file 1.2 (in OptGene (Patil et al., 2005) format). Table 2.1 shows, at a glance, this information.

*i*Syn669 spans all the biologically relevant flux nodes in the *Synechocystis* metabolism. Pyruvate, phosphoenolpyruvate (PEP), 3-phosphoglycerate, erythrose-4-phosphate and 2-oxoglutarate are main flux nodes for amino acids biosynthesis. Acetyl-CoA is an important flux node for fatty acids production, with high relevance for metabolic engineering towards biofuel production. Biosynthesis of nucleic acids comes from different metabolites, namely, ribose-5-phosphate, 5-phospho-beta-D-ribosyl-amine, L-histidine and L-glutamine. Moreover, with the information publicly available on databases, it was plausible to conclude that *Synechocystis* sp. PCC6803 bears an incomplete tricarboxylic acid cycle (TCA cycle), as it lacks 2-ketoglutarate dehydrogenase (EC 1.2.4.2). It has been published that glyoxylate shunt completes this cycle (Yang et al., 2002c), permitting the recycling of TCA metabolites. Alternatively, aspartate transaminase (reaction *2.6.1.1a* in *i*Syn669) can interconvert 2-ketoglutarate and oxaloacetate, thus bridging the gap of 2-ketoglutarate dehydrogenase, but short-circuiting TCA cycle. More recently Zhang and Bryant (2011) have successfully discussed an alternative hypothesis of completing cyanobacterial TCA cycle: through succinic semialdehyde and show solid proofs of it.

- *iSyn811*

Our previous genome-scale metabolic model was updated to *i*Syn811. We included and corrected many isoenzymes and complexes that were incomplete in *i*Syn669. We also incorporated several pathways that were believed to be in *Synechocystis* but that lacked many reactions (*i.e.* set of reactions that may have been wrongly annotated and are disconnected to the rest of the network). *i*Syn811 bears 976 reactions, 866 of them with cognate genes, 922 metabolites and 811 genes (see Additional file 1.3 for details). Again, the difference between 866 reaction catalysed by genes and the set of 811 genes that catalyse them are due to the multiple correspondence between genes and reactions in isoenzymes and complexes. As a matter of curiosity, when bidirectional reactions are converted to unidirectional, the number of reactions is extended to 1245. Reactions with no cognate genes are still present, 10 passive transport reactions and 28 chemical conversions (not mediated by enzymes) were included.

Additionally, proper databases searches and studies allowed us to assign reactions based on biochemical evidence or physiological considerations to ORF. Thus, we could identify 68 genes to those reactions, leaving 11 of the without a cognate gene. Table 2.1 can be used for this comparison.

Table 2.1 - Distribution of the model reactions as per cognate genes.

	*i*Syn669	*i*Syn811
Number of genes	669	811
Number of metabolites	790	922
Number of reactions	882	976
- With assigned genes	639	866
· Protein-mediated transport	78	78
- With no cognate gene	221	108
· Chemical conversion	39	28
· Transport reactions	20	10
· EC reactions not annotated	79	11
· Needed for biomass simulation	75	59

*i*Syn811 genome-scale metabolic model is available in Additional file 1.4 (in OptGene (Patil et al., 2005) format). Both versions have been compiled in OptGene/BioOpt format (Patil et al., 2005) in order to be readily usable with this software and is available at BioMet Toolbox webpage (http://www.sysbio.se/BioMet) (Cvijovic et al., 2010).

Biomass equation.

As we have discussed in the previous chapter, in order to be able to simulate metabolic fluxes under steady state, what we know as *flux balance analysis* (FBA), an objective function needs to be optimized. We have already seen the causes and consequences of biomass being the objective function and why that is important to mimic biomass growth. The present model features a detailed biomass equation which encompasses all the building blocks that are needed for a flux distribution simulation that reflects observed phenotype.

In present work, biomass equation needed to relate to molecular building blocks, such as amino acids (Feist et al., 2007), deoxyribonucleotides (Herdman et al., 1979), ribonucleotides (Allen and Smith, 1969), lipids (Tasakal et al., 1996),

carbohydrates (Burrows et al., 2008) and antenna chromophores (Miao et al., 2003). Table 2.2 encompasses biomass composition description with references where the information was retrieved from. All these building blocks with their respective stoichiometric coefficient is converted into one gram of dry cell weight (denoted as g_{DW}^{-1}). Biomass equation is reaction *Biomass* in Additional files 1.2, 1.4, 2.1 and 2.2. Some of the references used for this do not come from *Synechocystis* sp. PCC6803, like the one from amino acids that comes from *Escherichia coli* model (Feist et al., 2007), and others are from cyanobacteria but are so old that this organism was assigned to a different domain, such as blue-green algae, like the one from nucleic acids in Allen and Smith (1969). The process of experimentally retrieving this information is not trivial and most of the time is buried in articles that aim otherwise or higher in terms of scientific impact. This makes the process of collecting this information a challenging task. That is why we hope Table 2.2 will be useful to many people.

Table 2.2 - *i*Syn669 and *i*Syn811 biomass *composition*. Units in mmol g_{DW}^{-1}.

Metabolite	mmol / g_{DW}	Metabolite	mmol / g_{DW}
Amino acids (Feist et al., 2007)		**Deoxyribonucleotides** (Herdman et al., 1979)	
Alanine	0.499149	dATP	0.0241506
Arginine	0.28742	dTTP	0.0241506
Aspartate	0.234232	dGTP	0.02172983
Asparagine	0.234232	dCTP	0.02172983
Cysteine	0.088988	**Ribonucleotides** (Allen and Smith, 1969)	
Glutamine	0.255712	AMP	0.14038929
Glutamate	0.255712	UMP	0.14038929
Glycine	0.595297	GMP	0.12374585
Histidine	0.092056	CMP	0.12374585
Isoleucine	0.282306	**Lipids** (Tasakal et al., 1996)	
Leucine	0.437778	16C-lipid	0.20683718
Lysine	0.333448	(9Z)16C-lipid	0.01573412
Methionine	0.149336	18C-lipid	0.00351776
Phenylalanine	0.180021	(9Z)18C-lipid	0.03188596
Proline	0.214798	(9Z,12Z)18C-lipid	0.03568367
Serine	0.209684	(9Z,12Z,15Z)18C-lipid	0.01797109
Threonine	0.246506	(6Z,9Z,12Z)18C-lipid	0.05031906
Tryptophan	0.055234	(6Z,9Z,12Z,15Z)18C-lipid	0.01448179
Tyrosine	0.133993	**Antenna chromophores** (Miao et al., 2003)	
Valine	0.411184	Chlorophyll a	0.02728183
Carbohydrates (Burrows et al., 2008)		Carotenoids	0.00820225
Glycogen	0.014506		

2.3 Connectivity analysis of genome-scale metabolic models.

Before going into the metabolic flux analysis, we studied the network topology of the reconstructed models. This approach has been widely used in systems biology, with many works from many researchers from different fields, like Mathematics and Statistics. One of the major players in this field has been Professor Barabási, with pioneering works (Albert et al., 1999; Barabási and Albert, 1999) and reviews (Barabási and Bonabeau, 2003). This interdisciplinary approach, as it usually happens, has caused debate about ostensibly grand conclusions, like the one that construction of biological systems networks, as well as World-Wide Web network, follow a *power law distribution*, termed *"scale-free" distribution* (Keller, 2005, 2007).

From the network topology perspective, *i*Syn669 and *i*Syn811 display the connectivity distribution pattern similar to that of the other microbial genome-scale networks, *e.g.* yeast (Förster et al., 2003) and *Escherichia coli* (Feist et al., 2007) (Table 2.3 and 2.4). While most of the metabolites have few connections, few metabolites are involved in very many reactions and are often referred to as metabolic hubs. Homeostasis of such highly connected metabolites will affect globally the metabolic phenotype (as reflected in metabolite levels and fluxes) and therefore of interest for studying the organization of regulatory mechanisms on the genome-wide scale. Most connected metabolites include those related to energy harvesting (*e.g.* ATP, NADP+, oxygen), a key metabolite in the porphyrin and chlorophyll metabolism (S-adenosyl methionine), a couple of amino acids and its precursors (L-glutamate, L-glutamine and glutathione) and a key metabolite in the lipid biosynthesis pathway (malonyl-ACP). High connectivity of these metabolites hints to their potential central role in the adjustments of fluxes following environmental perturbations. In order to discover the corresponding regulatory mechanisms, additional studies should be done that are beyond the scope of this chapter –*e.g.* putative regulatory sequence motifs associated with the neighbours of these highly connected metabolites (Zelezniak et al., 2010).

Table 2.3 - Most connected metabolites in metabolic networks.

Metabolite	Neighbours in *i*Syn669	Neighbours in *i*Syn811	Neighbours in *E. coli*	Neighbours in yeast
H$_2$O	213	219	697	-
ATP	144	136	338	166
phosphate	108	112	81	113
ADP	103	111	253	131
diphosphate	97	84	28	-
H+	74	153	923	188
CO$_2$	72	72	53	66
NADP+	64	68	39	61
NADPH	63	68	66	57
NAD+	46	52	79	58
L-glutamate	45	44	52	56
NADH	42	48	75	52
AMP	36	21	86	48
oxygen O$_2$	36	40	40	31
ammonia	28	28	22	-
S-adenosyl-L-methionine	25	28	18	19
glutathione	25	26	17	10
a malonyl-ACP	23	23	15	10
L-glutamine	22	21	18	23
coenzyme A	21	23	71	39

Table 2.4 - Most connected metabolites (cofactors filtered) in metabolic networks.

Metabolite	Neighbours in *i*Syn669	Neighbours in *i*Syn811	Neighbours in *E. coli*	Neighbours in yeast
L-glutamate	45	44	52	56
S-adenosyl-L-methionine	25	28	18	19
glutathione	25	26	17	10
a malonyl-ACP	23	23	15	10
L-glutamine	22	21	18	23
S-adenosyl-L-homocysteine	21	24	12	14
cysteinylglycine	21	21	5	-
pyruvate	20	20	61	20
5-oxoproline	20	20	-	-
acetyl-CoA	15	15	34	24
L-aspartate	15	15	23	20
tetrahydrofolate	14	14	10	13
2-ketoglutarate	13	13	27	29
D-glyceraldehyde-3P	12	12	14	13
fructose-6-phosphate	10	10	18	18
phosphoenolpyruvate	10	10	26	12
L-methionine	10	10	15	15
L-serine	9	8	26	17
5-phosphoribosyl 1-pyroP	9	9	2	2
isopentenyl diphosphate	9	9	6	4

2.4 Our metabolic models and the state-of-the-art.

When we started working on *Synechocystis* sp. PCC6803 metabolic model reconstruction some central carbon metabolism models were published for this cyanobacterium. Hence, we had the chance to compare our advances to these works, but from that moment (mid-2008) until now, many efforts have been gathered around the metabolic reconstruction of this organism. In a sense, this makes researchers feel part of a community, but has also some downsides, like competition towards publication.

Anyhow, as a PhD candidate, I feel that our models have to be framed with the rest of these works in order to see which models owe efforts to which ones. Therefore, *Synechocystis* sp. PCC6803 metabolic models are hereby gathered and described. We have classified the models between *genome-scale metabolic models* and *central carbon metabolic models.*

a) *genome-scale metabolic models* are the ones that (1) include all reactions annotated in the genome and (2) a biomass equation that encompasses all building blocks in the cell metabolism.

b) *central carbon metabolic models* are the ones that (1) include reactions from glycolysis, tricarboxylic acid (TCA) cycle and pentose phosphate pathway and, in some cases, an additional set of reactions that connect some metabolites to amino acids and/or (2) a biomass equation without all building blocks.

We have described all metabolic models published to this date on this cyanobacterium and outlined their characteristics. For further details, refer to the correspondent paper.

- *Yang et al. central model* (2002c)

Seminal work published in 2002, it was the first time ever a metabolic model was developed for *Synechocystis* sp. PCC6803. Researchers build a metabolic network and simulated it under heterotrophic and mixotrophic growth conditions. Photosynthesis (only working under mixotrophy) was represented as one lumped reaction. This model bears 20 reactions and 15 metabolites and was used as a central-carbon scaffold for transcriptomic, metabolomic and fluxomic data produced by the same group (Yang et al., 2002a, 2002c).

- *Shastri and Morgan central model* (2005)

This model, published in 2005, extended the simulations to autotrophic growth and compared their results to the ones from Yang et al. (2002c). A list of reactions is presented that connects amino acid production to central carbon metabolites and photosynthesis is represented as one lumped reaction. This model bears 70 reactions (plus 23 for amino acids formation) and 50 metabolites.

- *Kun et al. genome-scale model* (2008)

This model, published in 2008, has its main sources on the automated reconstruction generated from the annotated genome and deposited in the MetaCyc database (Caspi et al., 2006; Krieger et al., 2004) and on the previous Shastri and Morgan central model (2005). Authors were interested in the study of sets of autocatalytic molecules, not on the analysis of flux behaviour or the integration of different levels of biological information. This model bears 916 reactions and 879 metabolites.

- *Navarro et al. central model* (2009)

Published in 2009, this central carbon metabolic model was developed independently by our group and simulates all three growth conditions on the same metabolic network. Obviously we compared our reconstruction and results to the ones from Yang et al. (2002c) and Shastri and Morgan (2005) and found improvements in oxygen evolution (of high importance in a photosynthetic bacterium) and flux distributions. Among other functions, we used this model in order to find out maximum hydrogen production titters and propose improved hydrogen producing strategies. This model has 90 reactions and 56 metabolites.

- *Fu central model* (2008)

Published in August 2008, this model claims to be genome-scale but fails in having a proper biomass equation. Its list of reactions are certainly from all the genome, meaning not only from a selected set of pathways, but this list is uncurated, featuring lumped key reactions (like the ones for the generation of NADPH and ATP, reaction 732 in Supplementary materials) and missing transport reactions. This model, however, is not suitable for genome-scale simulations due to lack of proper biomass equation, as a drain of cellular building blocks. Researcher takes biomass equation formulation from Shastri and Morgan (2005), so simulation results do not give more information than the ones already published by them. Despite that, phenotypic phase planes were performed to study usage of substrates to optimize cell growth and ethanol production strategies were assessed. This model has 831 reactions and 704 metabolites.

- i*Syn669 genome-scale model* (Montagud et al., 2010)

Our first genome-scale metabolic model sent to BMC Systems Biology on 5 February 2010 and accepted on 17 November 2010. Apparently, process of revision took so long as one of the reviewers that accepted to peer-review the manuscript failed to answer editors in due time, and several times. As we have detailed in this chapter, this metabolic model was gathered from public databases and literature in order to be able to simulate all possible growth modes under which this organism can grow. This metabolic model is one of the main characters of present chapter. Additionally in following chapters, we will detail other analyses done on top of this model: *flux balance analysis* in **Chapter 3**, transcription data integration in **Chapter 7**, productivity analysis in **Chapter 4**, etc. This model has 882 reactions (1045 unidirectional) and 790 metabolites

- *Knoop et al. genome-scale model* (2010)

This model, published in 2010, reconstructs *Synechocystis* sp. PCC6803 mainly from annotation files and KEGG database information, even though a selected

core set of reaction is selected to produce biomass. Authors study flux behaviour under different growth conditions using FBA algorithm. Interestingly, they shed light on the functioning of the oxidative reaction of the RuBisCO enzyme under photorespiration. This model has 380 reactions and 291 metabolic compounds.

- i*Syn811 genome-scale model* (Montagud et al., 2011)

Our update of *i*Syn669, sent to Biotechnology Journal on 23 June 2010 and accepted on 7 November 2010. Available online right away, it was published on a special issue of Biotechnology Journal on Applications in Biotechnology issued on February 2011. Update was based on enlarging the reaction set with isoenzymes and complexes as well as reactions that were disconnected to the rest of the network. Major changes were also made on the gene-reaction correspondence. As with *i*Syn669, this metabolic model is one of the main characters of present chapter, but you will also find it in following ones. Over this model we studied and applied an interesting study of flux capabilities, termed *flux coupling analysis* (Burgard et al., 2004), which is the main topic of **Chapter 5**. This model has 976 reactions (1245 unidirectional) and 922 metabolites.

- *Yoshikawa et al. genome-scale model* (2011)

This work presents yet another genome-scale metabolic model of *Synechocystis* sp. PCC6803 published in mid-2011. Authors compare their model to Knoop et al. (2010) and to *i*Syn669 (Montagud et al., 2010) as to their accuracy to experimental fluxes obtained by Yang et al. (2002c). This model has 493 reactions and 465 metabolites.

2.5 Conclusions.

We have explained the reconstruction of a genome-scale metabolic network for *Synechocystis* sp. PCC6803, which allows simulating production of all the metabolic precursors of the organism. This reconstruction has two versions: *i*Syn669 and *i*Syn811. The metabolic reconstruction represents an up-to-date database that encompasses all knowledge available in public databases, scientific publications and textbooks on the metabolism of this cyanobacterium.

These models are, up to our knowledge, the most complete and comprehensive work for *Synechocystis* sp. PCC6803 to date, which has its potential as the photosynthetic model organism, as we can appreciate in following chapters.

In order to facilitate the use of the *Synechocystis* sp. PCC6803 model, we have compiled our work as a Pathway Tools Tier 2 Pathway-Genome Database (PGDB). Searches for genes, reactions, proteins, pathways and regulations can be performed on this service. Our *Synechocystis* sp. PCC6803 database consists of 3622 genes, 3572 proteins, 58 transport reactions, 701 compounds, 43 tRNAs and 889 enzymatic reactions distributed among 180 MetaCyc pathways. Metabolic model have also been compiled in OptGene and SBML formats in order to be readily usable with different software and are available at BioMet Toolbox (Cvijovic et al., 2010) webpage (http://www.sysbio.se/BioMet).

Altogether, the genome-scale metabolic network of *Synechocystis* sp. PCC6803 is a valuable tool for the applied and fundamental research of *Synechocystis* sp. PCC6803, as well as for the broad field of metabolic systems biology. Applicability of *i*Syn669 and *i*Syn811 metabolic models is demonstrated by using a variety of computational analyses in following chapters.

Work from this chapter has been updated and has been used as founding stone to a (semi-)automated process developed by colleagues Gamermann and Reyes.

Socrates: Heracleitus [of Ephesus] says, you know, that
all things move and nothing remains still, and he likens
the universe to the current of a river, saying that you
cannot step twice into the same stream.
[origin of the aphorism *everything flows* (*pantha rhei*)]

Plato, Cratylus, Paragraph 402 section a line 8, in Harold
N. Fowler, *Plato in Twelve Volumes*, Vol.12, Harvard
University Press, 1921

3

Flux landscapes of *Synechocystis* sp. PCC6803

Where PhD applicant uses the model developed previously to have virtual snapshots of different fluxes landscapes that correlate to experimental snapshots gotten from people in other labs, in other continents.

Parts of the contents of this chapter are based on parts of the following journal articles:

- Montagud et al. **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

- Montagud et al. **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium.** *BMC Systems Biology* 2010, **4**:156.

- Pinto et al. **Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC6803 mutant lacking a functional bidirectional hydrogenase.** *Microbiology (Reading, England)* 2011, **158**:448-464.

3.1 Introduction.

As we said in previous chapter, genome-scale metabolic models have been used in many works to gain insights into the metabolic behaviour of the cell. If a genome-scale metabolic model is formulated properly, it is expected to allow simulating environmental and genetic perturbations in the metabolic network. Thus, together with appropriate constraints, a metabolic model would partially represent a virtual organism –an *in silico* model that allows studying possible flux distributions inside the cell under different environmental conditions and for a given genetic make-up. Studying how metabolic homeostasis buffers environmental perturbations is the main goal of present chapter. We will leave the study of how metabolism deals with genetic changes for following chapter.

In order to study how metabolic behaviour deals with environmental perturbations we will use *flux balance analysis* (FBA) described in **Chapter 1** as in many other works (Orth et al., 2010; Stephanopoulos et al., 1998). As a quick remainder, let me draft that this analysis needs to work with the stoichiometric information of the organism's metabolism, a set of boundary constraints and an objective function. In present case, stoichiometric information is given by *Synechocystis* sp. PCC6803 genome annotation, set of boundary constraints were uncovered in tons of bibliography and some personal communications with experimentalists (mainly Tamagnini's group at IBMC in Porto and Wright's group at University of Sheffield, England) and, finally, objective function as it is common in this kind of analysis has been selected to be biomass growth, *i.e.* cell growth.

Synechocystis sp. PCC6803 is capable of growing under three different growth conditions as marked by the utilized carbon source or sources (Herrero and Flores, 2008). This causes that three distinct modes of operation are interweaved over the same metabolic network, *viz*.:

- *photoautotrophy*, where energy comes from light and carbon from CO_2
- *heterotrophy*, where energy and carbon source is a saccharide, for instance glucose
- *mixotrophy*, a combination of the above two, where light is present as well as a combination of two carbon sources: glucose and CO_2

Other authors like Vermaas (1996) consider three more growth conditions according to light and oxygen presence: *anaerobiosis* (with glucose or other fixed-carbon as source), and split heterotrophy into *photoheterotrophy*, *light-activated heterotrophy* and *dark heterotrophy*.

Furthermore, different analyses are performed by using this metabolic reconstruction, including reaction knock out simulations, flux variability analysis and identification of transcriptional regulatory hotspots. Overall, *i*Syn669 and *i*Syn811 are valuable tools towards the development of a photo-biological production platform. These models have also contributed to the existing set of genome-scale models with the virtue of being one of the first stoichiometric models that completely accounts for photosynthesis.

3.2 *Synechocystis* sp. PCC6803 fluxes landscape.

*i*Syn669 and *i*Syn811, together with appropriate physiological constraints, were used as a stoichiometric simulation model by using FBA algorithm (Stephanopoulos et al., 1998). This model simulates steady state behaviour by enforcing mass balances constraints for all metabolic intermediates (again, please refer to **Chapter 1** for further information).

FBA simulation is solved by optimizing an objective function, which most of the time is cell growth. Biomass synthesis, a theoretical abstraction for cellular growth, is considered as a drain of some of the intermediates, *i.e.* building blocks, into a general biomass component. Presence of photosynthesis allows these metabolic models to "grow" (*i.e.*, FBA results in a feasible solution) under four metabolic states: photoautotrophy, mixotrophy and two modes on heterotrophy, as we will further down. Different studies have reported that the simulation results do not usually vary drastically when using a common biomass equation for different growth condition (Shastri and Morgan, 2005; Varma and Palsson, 1993a). Nevertheless, experimental efforts should be directed at the depiction of the best precursors and composition that could characterize, at least, the three main growth modes, *viz.*, autotrophy, heterotrophy and mixotrophy, in the scope of recent results (Schuetz et al., 2007). Due to the lack of such data, the present work uses one single biomass equation in the simulations of all three metabolic states (described in Table 2.2).

Growth under *pure* heterotrophy, or *dark* heterotrophy (in the absence of light) is a subject under study (Anderson and McIntosh, 1991; Carr and Whitton, 1982), being the regular experimental design to give a short light pulse prior to the *pure* heterotrophic phase (termed *light-activated* heterotrophy). Nevertheless, the theoretical flux distribution under heterotrophic conditions is interesting by itself –especially in comparison with the flux distribution in a light-fed energy metabolism. Moreover, fluxes in the heterotrophy mode may help in

obtaining insight into the variations under the mixotrophic condition, which is of high relevance for industrial applications (Navarro et al., 2009).

Constraints used for flux balance simulations.

Some capacity constraints had to be added in order to have a feasible solution for the linear programming problem. All FBA simulations were carried out under the appropriate constraints to match an autotrophic specific growth rate of 0.09 h^{-1}. As some examples, maximum carbon uptake rate was found to be 3.4 mmol g_{DW}^{-1} h^{-1} into the cell, with HCO_3^- and CO_2 as carbon sources (Shastri and Morgan, 2005). Additionally, we fixed the maintenance requirement for the heterotrophic case to be 1.67 ATP moles per mole of glucose consumed as was determined by Shastri and Morgan (2005), and was maintained for autotrophic and mixotrophic growth. For the sake of comparison across the different conditions, uptake rates for the corresponding carbon sources were matched based on normalization per number of carbon atoms. These parameters were maintained for comparative purposes in the heterotrophic condition, glucose being the sole carbon source and blocking photons input in the *dark* heterotrophy. Some of the constraints were altered in the mixotrophic conditions in order to simulate the characteristic flux distribution of this growth mode. Phosphate, water, sulphate, nitrate, ammonia as well as carbon monoxide and hydrogen peroxide transport across the membrane were considered and properly bounded. Some of the reversible reactions involving NADH and NADPH were constrained to be irreversible so that spurious transhydrogenation was controlled. Main constraints across different growth conditions can be seen in Table 3.1.

Table 3.1 - Major constraints across different growth conditions. Units in mmol g_{DW}^{-1} h^{-1}.

Simulated conditions	Autotrophy	Mixotrophy	Photo-heterotrophy	Dark heterotrophy
Light input in PSI	0.8	0.8	0.8	0
Light input in PSII	0.8	0.8	0.8	0
Glucose feeding	0	0.2835	0.567	0.567
CO_2 feeding	1.7	0.85	0	0
HCO_3^- feeding	1.7	0.85	0	0
RuBisCO carboxylase	unbound	unbound	0	0

70

Results of the subsequent FBA simulations for the three different growth conditions are presented in the following. Some of the reactions that are physiologically relevant for each of the conditions are summarized in Table 3.2 and Figure 3.1. Flux values for the rest of the reactions, including the upper and lower bounds are provided in Additional file 2.1 and 2.2.
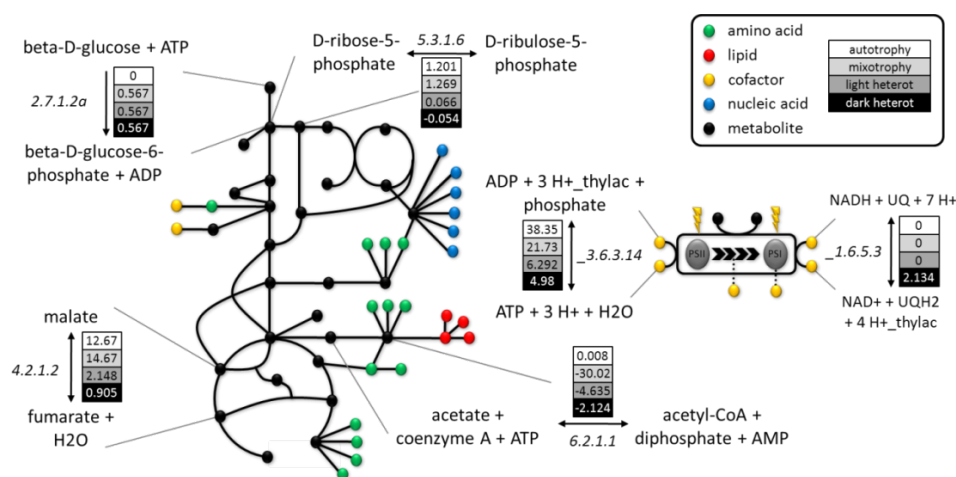

Figure 3.1 - Selected reactions in *i*Syn669 and *i*Syn811 network that display flux changes across the three basic growth modes. Flux values (in mmol g_{DW}^{-1} h^{-1}) for selected reactions in the *Synechocystis* sp. PCC6803 metabolism. These reactions mark changes across the three basic growth modes, *viz.*, autotrophy, heterotrophy and mixotrophy. Corresponding flux ranges can be found in Table 3.2 and in Additional file 2.1 and 2.2 for all the reactions in *i*Syn669 and *i*Syn811, respectively.

Simulations of the three metabolic modes.

- *Heterotrophy*

Heterotrophy was simulated by considering glucose as the sole carbon source with uptake rate of 0.567 mmol g_{DW}^{-1} h^{-1} (3.4 mmol divided by 6 carbon atoms), entering the system through *glcP* glucose transporter (reaction *TRANS-RXN59G-152*). With the purpose of having a *pure* heterotrophic state, photon uptake rate was constrained to 0; this caused photosynthesis fluxes to be shut down. In this case, glucose will be the source for the formation of carbon backbones for the building blocks of the cell, depicted in the biomass equation. The glycolytic and the oxidative mode of the pentose phosphate pathway were found to be active. Oxidative pentose phosphate pathway is the major pathway for glucose catabolism as was reported years ago (Pelroy et al., 1972).

Table 3.2 - Comparison of selected fluxes across different growth conditions. Units in mmol g_{DW}^{-1} h^{-1}. *2.7.1.2a*, glucokinase, is the reaction that phosphorylates beta-D-glucose upon entrance in the cell, marking the start of the glycolysis. The flux direction changes can be seen in reaction *4.2.1.2*, fumarate hydratase, from TCA cycle and *5.3.1.6*, ribose-5-phosphate isomerase, from the pentose phosphate pathway. *_UQ* and *_1.6.5.3* are reactions that reduce UQH2 from photosystem II or NADH oxidation, respectively, causing a pumping of protons to the thylakoid. *_3.6.3.14* is the ATP synthase that forms ATP shuttling protons from the thylakoid to the cytosol. *6.2.1.1*, acetate-CoA ligase, is the reaction that generates acetyl-CoA from acetate and coenzyme A, that would be a major flux hub in an ethanol-producing strain, standing as the first step of fermentation.

Reaction name	Autotrophy	Minimum flux	Maximum flux	Mixotrophy	Minimum flux	Maximum flux	Dark Heterotrophy	Minimum flux	Maximum flux	Photo-heterotrophy	Minimum flux	Maximum flux	Reaction description
2.7.1.2a	0	0	0	0.567	0.566	0.567	0.567	0.566	0.567	0.567	0.566	0.567	beta-D-glucose + ATP → beta-D-glucose-6-phosphate + ADP
4.2.1.2	12.67	12.667	+∞	14.67	14.657	+∞	0.905	0.884	+∞	2.148	1.836	+∞	malate ↔ fumarate + H2O
5.3.1.6	1.201	1.2	+∞	1.269	1.269	+∞	-0.054	-0.051	-0.055	0.066	0.067	+∞	D-ribose-5-phosphate ↔ D-ribulose-5-phosphate
_UQ	0.8	0	0.8	0.8	0	0.8	0	0	0	0.8	0	0.8	PSII* + UQ + 2 H+ → PSII + UQH2
_1.6.5.3	0	0	+∞	0	0	+∞	2.134	0	+∞	0	0	+∞	NADH + UQ + 7 H+ → NAD+ + UQH2 + 4 H+_peribac
_3.6.3.14	38.348	15.7	+∞	21.727	21.7	+∞	4.98	4.95	+∞	6.292	6.281	+∞	3 H+_peribac + phosphate O4P + ADP ↔ 3 H+ + H2O + ATP
6.2.1.1	0.008	-∞	+∞	-30.017	-∞	+∞	-2.124	-∞	+∞	-4.635	-∞	+∞	coenzyme A + acetate + ATP ↔ acetyl-CoA + diphosphate + AMP

Phosphoenolpyruvate carboxylase (reaction *4.1.1.31*) is the main anaplerotic flux to the Tricarboxylic (TCA) cycle. Carbon fixation efficiency is around 60 %, the rest being released in the form of CO_2, as reported in a previous work from our group (Navarro et al., 2009).

In contrast to *dark* heterotrophy, if a *photoheterotrophy* simulation is run, light enters the system and RuBisCO enzyme is active (reaction *4.1.1.39*), fixing all the CO_2 that was released in *dark* heterotrophy, boosting carbon efficiency to a theoretical 100%. In this case, global flux distribution as well as flux ranges resemble that of autotrophy more than that of the *dark* heterotrophy. Carbon skeletons are still produced through glycolysis and NAD(P)H is reduced along the glycolysis, pyruvate metabolism and TCA cycle. On the other hand, pentose phosphate pathway has shifted to the reductive mode due to RuBisCO activation, and the corresponding flux is increased in magnitude. Carbon fixation happens at the RuBisCO level, thereby assimilating the CO_2 produced by the glucose metabolism, and the production of ATP and NADPH through photosynthesis relieves the oxidative phosphorylation from draining NADPH to generate ATP.

We have also researched into a special subset of conditions: *anoxic dark heterotrophy*. Under these conditions, cell drains carbon reserves under anoxic conditions performing fermentation and generating products like ethanol. Global flux distribution as well as flux ranges are identical to that of *dark* heterotrophy, with the exception of ethanol production whose flux is enhanced and there is a consequent decrease in biomass production. Oxidative pentose phosphate pathway is still the major pathway for glucose catabolism and carbon efficiency is also 60% (carbon is now released, as CO_2 as before, and also *fermented* in the form of ethanol).

- *Autotrophy*

Photoautotrophy was initially simulated considering an illumination of 0.15 mE $m^{-2} s^{-1}$. Assuming that the mass of a typical *Synechocystis* sp. PCC6803 cell is 0.5 pg (Loferer-Krössbacher et al., 1998) and its radius is 1.75 μm (Lawrence et al., 1998), we estimated that the theoretical maximum illumination is 41563.26 mE $g_{DW}^{-1} h^{-1}$. Nevertheless, in order to estimate physiologically meaningful photon uptake values that were closer to the experimental measurements an additional optimization step was performed. First, carbon uptake rate was found that resulted in a specific growth rate of 0.09 h^{-1} (Shastri and Morgan, 2005), while the light intake was unconstrained. Next, the carbon uptake rate was constrained to this value and the second optimization problem was solved where light uptake

was minimized. This minimization resulted in photon uptake for photosystem I (reaction _lightI_) and photosystem II (reaction _lightII_) being 0.8 mE g_{DW}^{-1} h^{-1}. Carbon sources used in simulating photoautotrophy conditions were carbon dioxide and carbonic acid, and its entrance to the metabolism was mediated by RuBisCO (reaction _4.1.1.39_) and carbonic anhydrase (reaction _4.2.1.1b_) respectively.

As our models' biomass equation encompasses all essential metabolite precursors, these will be the sinks of our network, while photons, carbon dioxide and/or carbonic acid will be the sources. Thus, autotrophic fluxes will flow in the gluconeogenic direction and through the Calvin cycle, which is the reductive mode of the pentose phosphate pathway. Phosphoenolpyruvate carboxylase is the main anaplerotic flux to the TCA cycle and glyoxylate shunt is inactive.

- _Mixotrophy_

Photons, carbon dioxide and glucose are independent feed fluxes in this simulation. These fluxes entered the system through the same reactions as described for the previous growth modes. Carbon source presents, in this case, one more degree of freedom than in the rest of the conditions. In order to keep a comparative criterion across conditions, we normalized CO_2 and glucose inputs to the same carbon uptake flux as in the case of the autotrophy and the heterotrophy. Photon uptake rates were also normalized in a similar manner to match the autotrophic state.

Having the same metabolic sinks as the two previous modes and the sources from the both of them, it is logical to think that the resulting flux distribution will be equidistant to the autotrophic and heterotrophic simulations. Indeed, we observed that the mixotrophic flux distribution lies in-between the previous two states, being a bit closer to the heterotrophy. Glycolysis is present and glyoxylate shunt is shut down, an active photosynthesis is present, oxidative phosphorylation is less stressed than in heterotrophy as the energy can be produced from the photon uptake and Calvin cycle is active, as carbon sources are CO_2 and glucose.

3.3 Flux variability analysis.

Flux balance analysis presented above guarantees to find the optimal objective function value (in this case, biomass formation rate or cell growth). However, the predicted intra-cellular flux distribution is not necessarily unique due to the presence of multiple pathways that are equivalent in terms of their overall

stoichiometry. Thus, often the system exhibits multiple optimal solutions and further elucidation requires additional constraints based on experimental evidences (*e.g.* carbon labelling data). Alternatively, physiological insight can be still obtained by studying the variability at each flux node given the objective function value –a procedure referred to as *flux variability analysis* (Mahadevan and Schilling, 2003).
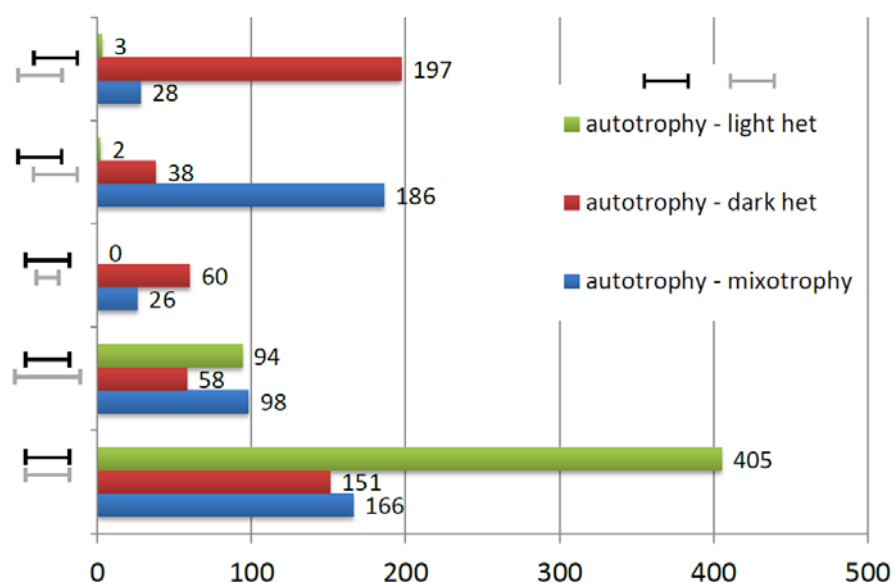

Figure 3.2 - Overview of the flux adjustments between different growth conditions.
Comparison of flux variability between autotrophy vs. mixotrophy, autotrophy vs. *dark* heterotrophy and autotrophy vs. photoheterotrophy (*light het*). Minimum and maximum flux ranges were compared for each reaction.

In order to gain insight into the flux changes underlying the changes in the *Synechocystis* sp. PCC6803 metabolism due to the (un)availability of light, we have compared the autotrophic growth with the other two growth modes by using flux variability analysis (Figure 3.2). Interestingly, autotrophy permits an overall broader flux landscape than heterotrophy (let it be *dark* or *light-activated*). On the other hand, autotrophic flux ranges are in general narrower than the mixotrophic ranges. Figure 3.1 and Table 3.2 depict some of the physiologically relevant reactions for which the feasible flux range differs across conditions. These include glucokinase from glycolysis, fumarate hydratase from TCA cycle, ribose-5-phosphate isomerase from pentose phosphate pathway, NADH dehydrogenase from oxidative phosphorylation or photosystem II oxidation. These reactions identify the key nodes in the metabolic network that must be

regulated in order to adapt to a change in the available energy/carbon source. Mechanisms underlying such changes will be of particular interest not only for biotechnological applications but also from the biological point of view. As a glimpse of the detailed flux (re-)distributions in each of the studied growth conditions, Figure 3.3 describes fluxes in the pyruvate metabolism.
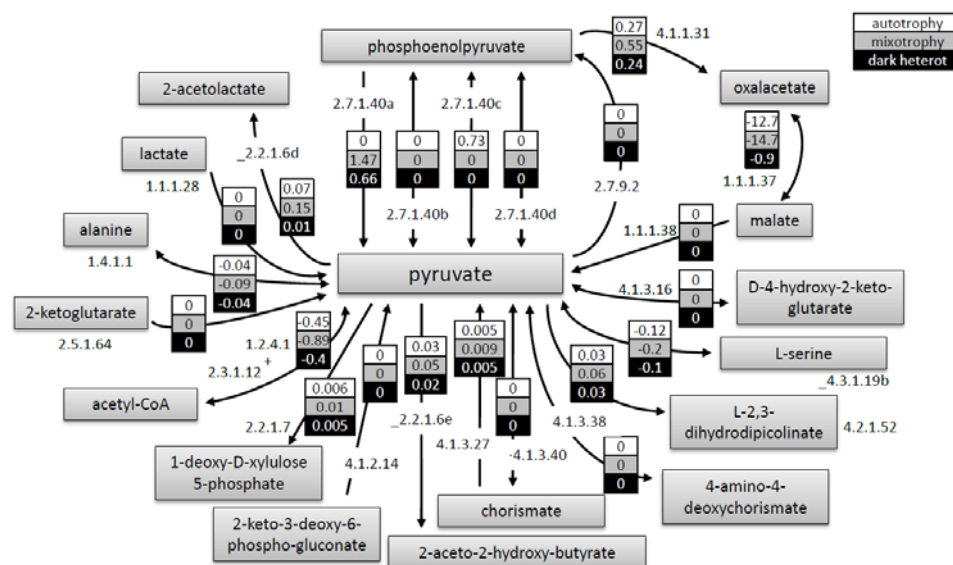

Figure 3.3 - Fluxes of reactions around pyruvate. Flux values (in mmol g_{DW}^{-1} h^{-1}) for reactions that produce or drain pyruvate in *Synechocystis* sp. PCC6803 metabolism. Negative sign in bidirectional reactions means pyruvate consumption. Reactions names can be traced in reaction list in Additional files 1.2 and fluxes can be found in Additional file 1.4.

3.4 Conclusions.

We have demonstrated the use of a genome-scale metabolic network reconstruction to simulate *Synechocystis* sp. PCC6803 metabolic behaviour under different environmental conditions. Suitability of the presented model for performing *in silico* metabolic engineering analysis was demonstrated by using OptGene software framework (Patil et al., 2005) and this suitability will be further developed in following chapters.

This is the first chapter where we demonstrate applicability of *Synechocystis* metabolic models by using a variety of computational analyses. In this chapter, *flux balance analysis* was applied in order to simulate the three physiologically

important growth conditions of cyanobacteria, *viz.*, heterotrophic, mixotrophic and autotrophic. Our metabolic model was capable of simulating the production of the monomers or building blocks that build up the cells, in the range that is in agreement with the reported growth experiments. Flux split reactions and major pathway *directions* are in accordance with experimental fluxes (Yang et al., 2002b). Our photosynthetic metabolic model includes all of the central metabolic pathways that previous works (Shastri and Morgan, 2005; Yang et al., 2002b) considered. Regarding the parts from our model that overlap with the previous works (part of the central carbon metabolism), the predictions for the flux directionality changes following light shift match between those models and *i*Syn669 and *i*Syn811. In fact, our models expand the flux study to all the pathways described in the *Synechocystis* sp. PCC6803 genome annotation.

Further insight into metabolic engineering of *Synechocystis* sp. PCC6803 are presented in following chapter, where we study mutants with improved production of a set of industrially relevant metabolites.

3.5 Methods.

Linear programming for flux balance analysis

In this chapter, we have extensively used linear programming to solve an FBA problem. The set of biochemical reactions of the genome-scale metabolic model were formulated as a steady state stoichiometric model:

$$\sum_{i=1}^{M}\sum_{j=1}^{N} S_{ij} \cdot v_j = 0,$$

The details are described elsewhere, for example in **Chapter 1** or Stephanopoulos *et al* (Stephanopoulos et al., 1998). This model describes cellular behaviour under pseudo steady-state conditions, where *S* is stoichiometric matrix that contains the stoichiometric coefficients corresponding to all internal (balanced) metabolites. *v* is flux vector that corresponds to the columns of *S*. Given a set of experimentally-driven constraints, former equation was solved by using linear programming, the approach known as *flux balance analysis*, or FBA (Edwards et al., 1999).

Since the number of reactions is typically larger than the number of metabolites, the system becomes underdetermined. In order to obtain a feasible

solution for the intracellular fluxes, an optimization criterion on metabolic balances has to be imposed. This can be formulated by maximizing one of the biochemical reactions, *e.g.* biomass equation, subject to the mass balance and the capacity constraints.

For instance,

$$\text{Max} \, (v_j) \qquad\qquad \text{subject to} \; \sum_{i=1}^{M} \sum_{j=1}^{N} S_{ij} \cdot v_j = 0,$$

$$v_{j,irr} \in R^+$$

$$v_{j,rev} \in R$$

$$v_{j,const} \in R, v_{min} < v_{j,const} < v_{max}$$

$$v_{j,uptake} \in R, v_{min} < v_{j,uptake} < v_{max}$$

where v_j is the rate of the jth reaction. The elements of the flux vector *v* were constrained for the definition of reversible and irreversible reactions, $v_{j,rev}$ and $v_{j,irr}$, respectively. Additionally, two set of equations were established, $v_{j,const}$, constrained metabolic reactions, and $v_{j,uptake}$, uptake reactions, which were bound by experimentally determined values from the literature. Biomass synthesis was considered as a drain of precursors or building blocks into a hypothetical biomass component. Flux through biomass synthesis reaction, being the biomass formation rate, is directly related to growth of the modelled organism (Stephanopoulos et al., 1998). Table 2.2 shows the biomass composition that was considered in the *i*Syn669 and *i*Syn811 metabolic model.

Simulations were performed with the OptGene software (Patil et al., 2005), nowadays available online at Biomet Toolbox (Cvijovic et al., 2010) (http://www.sysbio.se/BioMet).

Viver não é necessário; o que é necessário é criar
[Living is not necessary, what is needed is to create]

Fernando Pessoa, *Navegar é preciso*

4

Genetic variations of *Synechocystis* sp. PCC6803 metabolism

Where PhD applicant takes advantage of simulation's abilities to experiment with the organism's genome in order to have the metabolism behaving as he desires, or close to the desired scenario.

Parts of the contents of this chapter are based on parts of the following journal articles:

- Montagud et al. **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

- Montagud et al. **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium.** *BMC Systems Biology* 2010, **4**:156.

4.1 Introduction.

Microorganisms are widely used for producing pharmaceutical, food and energetic products. Currently there is an increasing trend to replace chemical synthesis processes with biotechnological routes. In order to economically produce desired compounds from microbial cell factories it is, however, generally necessary to reorient the metabolism's goal, since microorganisms are typically evolved for maximizing growth in their natural habitat.

In later years, rational design strategies based on genetic engineering have been applied with an increasing success. These strategies, often termed metabolic engineering include many experimental and mathematical tools that have been developed for introducing directed genetic modifications that will lead to desirable metabolic phenotypes resulting in improved production of desirable compounds (Nielsen, 2001; Patil et al., 2005).

Microbial metabolism is often subjected to tight regulation and is constrained by mass and energy conservation laws on a large number of intracellular metabolites, and this makes it difficult to predict the effects of introducing genetic modifications in a given cell. Moreover, as metabolic pathways and related regulatory processes form complex molecular and functional interaction networks (Ideker et al., 2001; Patil and Nielsen, 2005), it is only through analysis of the metabolism as a whole in an integrative systems approach (Stephanopoulos et al., 2004) that one may evaluate the effect of specific genetic modifications.

One of the goals of present thesis is the study of the construction of a photon-fuelled cell factory. We search an ideal system where energy is harvested from photons through the photosynthesis and carbon skeletons are shaped out of atmospheric CO_2 through Calvin cycle's carbon fixation. As a starting point, focus has been set on algae and cyanobacteria (Tamagnini et al., 2007). These organisms, apart from being of easy cultivation and little nutritional demands, would not suffer from the disadvantages encountered with the current strategies, being capable of producing biofuels in an essentially continuous process (Angermayr et al., 2009).

Synechocystis sp. PCC6803 is a cyanobacteria model organism interesting for biotechnological applications (Gutthann et al., 2007) such as heterologous production of metabolites like isoprene (Lindberg et al., 2010), poly-beta-hydroxybutyrate (Tyo et al., 2006; Wu et al., 2001), alcohols (Angermayr et al., 2009), biofuels (Liu et al., 2010) and bio-hydrogen (Tamagnini et al., 2007). *Synechocystis* is thus an attractive candidate for developing a clean and

sustainable platform for biotechnological processes aimed at value-added products formation.

Metabolic models at the genome-scale are needed for rational metabolic engineering approaches targeted at designing synthetic cell factories (Barrett et al., 2006; Morange, 2009; Patil et al., 2004). As we have seen in previous chapters, genome-scale metabolic models can be reconstructed and their metabolic behaviours uncovered with the aid of several mathematical tools, in this chapter we will focus on the capacity of targeting possible changes in the genome landscape according to a given desired metabolic behaviour, like manufacture of a value-added metabolic product. That is, we will use tools such as *flux balance analysis* (Edwards et al., 1999; Varma and Palsson, 1993a) and *minimization of metabolic adjustments* (Segrè et al., 2002) in order to design our own photosynthetic bio-refinery.

4.2 Construction of a photon-fuelled cell factory.

Towards studying the behaviour of metabolism to genetic fluctuations, we will use our metabolic models *i*Syn669 and *i*Syn811. In the reconstruction phase, we developed a database of gene-reaction relationship that will be of high interest for present task.

The comprehensive set of reconstructed biochemical equations of *i*Syn669 and *i*Syn811 and simulations done in **Chapter 3** enabled us to further analyse the characteristics and potential of the *Synechocystis* metabolic network. Firstly, we studied reactions (and thereby the corresponding genes) that are necessary for growth, identifying a set of reactions that should not be deleted and, secondly, to identify *in silico* proper targets for maximization of a given metabolite of socio-economic interest.

Essential genes.

*i*Syn669 network consists of 790 metabolites and 882 reactions. Among these, 350 genes (36% of the total, Figure 4.1) were found to be necessary for the formation of the biomass under the mixotrophic growth conditions by using FBA and MOMA algorithms, and performing it in triplicate in order to check for possible multiple solutions. This set of genes can be divided in to two categories: 1) *essential genes*, deletion of which completely inhibits biomass growth (304 genes, 34% of the total, with FBA); and 2) *genes deletion of which causes a reduced growth rate* (46 genes, 2% of the total, with FBA). The set of 304 essential genes can be understood as the core of the metabolism, as deleting them would produce an unviable organism. The results based on MOMA algorithm essentially tally these numbers: 311 essential genes, 35% of the total, and 45 that cause a reduced growth rate, 5% of the total, (Additional file 4.1).
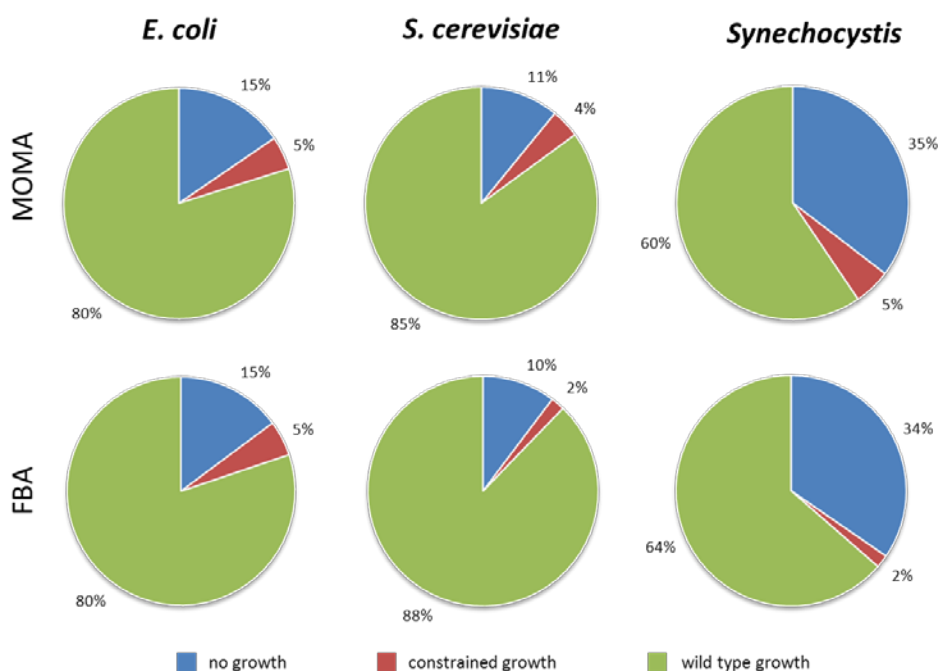

Figure 4.1 - Essential genes in *Synechocystis* sp. PCC6803. Distribution of gene knock out results for three model organisms, simulated by using FBA and MOMA algorithm, classified as wild-type growth, constrained growth and no growth.

Interestingly, if we compare the proportion of the essential genes under FBA simulation in the metabolic networks of *E. coli* (187 genes, 15% of the total) (Feist et al., 2007) and *Saccharomyces cerevisiae* (148, 10% of the total) (Förster et al.,

2003) with *i*Syn669, we find that *Synechocystis* has a significantly larger fraction of metabolic genes whose deletion obliterates biomass formation (304 genes, 34% of the total). One possible explanation for the difference in the relative proportion of essential genes in these three organisms would be an incomplete or incorrect annotation of the genome of *Synechocystis* sp. PCC6803. For example, if only one of the isoenzymes corresponding to a reaction is annotated, the corresponding *in silico* knock out will result in a false negative prediction. It is also important to note that the computational predictions of gene essentiality based on FBA are highly dependent on the growth medium used for the simulations. Thus, the comparison across different species may not be straightforward. Moreover, the natural growth conditions of *Synechocystis* may have dictated selection for a relatively high proportion of essential genes. Such hypotheses need careful consideration of several factors and are beyond the scope of this chapter.

Production of value-added compounds.

Synechocystis sp. PCC6803 is considered as a candidate photobiological production platform —it can potentially produce molecules of interest by using CO_2 and light (Lindberg et al., 2010). To this end, *i*Syn669 and *i*Syn811 can be used to perform simulations, not only for assessing the feasibility of producing a given compound, but also to identify potential metabolic engineering targets towards improved productivity. For example, flux simulations can help in estimating maximum theoretical yields for the products and intermediates of interest. As an example, we will present studies on productivity of different target metabolites: succinate, ethanol and hydrogen. These results are extended in Additional file 4.2.

   a) Succinate.

Succinate is an important metabolite for its biotechnological applications as well as for being a metabolite that bridges the TCA cycle with the electron transfer chain. As an example of the usefulness of the present metabolic model we have designed an *in silico* metabolic engineering strategy to improve the production of succinate. The underlying idea is to design a succinate over-producing metabolic network (through reaction knock out simulations), whereas the intracellular fluxes are distributed so as to maximize the biological objective function (*e.g.* growth) (Stephanopoulos et al., 2004). To this end, OptGene algorithm (Patil et al., 2005) was used together with *minimization of metabolic adjustment* (MOMA) algorithm. MOMA (Segrè et al., 2002) has been reported to provide better description of flux distributions in mutants or under non-natural growth conditions as opposed to

FBA. A *design objective function* which copes with the metabolite of interest, succinate, has been determined maintaining the *biological objective function* as the biomass formation. Additionally, simulations were performed in triplicates in order to check for possible multiple solutions.

OptGene simulations for single, double and triple knock out strategies were performed to obtain solutions with improved succinate production, but without drastically diminishing the biomass production. A schematic view can be found in Figure 4.2. We used mixotrophic conditions on *i*Syn669, for which wild type optimal growth rate was 0.17909 mmol g_{DW}^{-1} h^{-1}. Titters can be compared to closest case scenario found in literature: a recombinant *Escherichia coli* bearing carbonic anhydrase from cyanobacterium *Anabaena* sp. 7120 that increase its succinate yield from 0.19 mol mol glucose^{-1} to 0.39 mol mol glucose^{-1} (Wang et al., 2009). Resulting strain had, then, a production of 0.195 mmol g_{DW}^{-1} h^{-1}.

Table 4.1 - Proposed single knock outs for an improved succinate production. Units for *objective function* (biomass production) and *design function* (succinate production) in mmol g_{DW}^{-1} h^{-1}.

Reaction in model	Gene	% of wild type objective function	Design function	Times improvement on design function
2.7.1.40d	*pykF*	95.8123	0.591745	3858
2.7.4.6d	*ndkR*	95.9174	0.396922	2588
4.1.1.31	*ppc*	52.9414	0.191257	1247
2.6.1.1a	*aspC*	94.9190	0.086242	562
3.6.1.1	*mazG*	99.2886	0.0572139	373
_1.9.3.1	*ctaCDE*	84.8552	0.0325218	212
4.2.1.2	*fumC*	99.7160	0.0163946	106
4.3.2.2a	*purB*	99.8532	0.0111579	72
6.3.4.4	*purA*	99.8532	0.0111579	72
3.5.2.3	*pyrC*	95.9204	0.0110457	72

The best single knock out (Table 4.1) under mixotrophy was found to be the mutant of pyruvate kinase (reaction *2.7.1.40c* in *i*Syn669 and genes *sll0587* and *sll1275*) that has a succinate evolution of 0.591745 mmol g_{DW}^{-1} h^{-1} with a growth rate of 0.1716 mmol g_{DW}^{-1} h^{-1}. Blocking this reaction, preventing pyruvate and phosphoenolpyruvate from using GTP and GDP would drive a high increase in

succinate production. The flux between pyruvate and phosphoenolpyruvate can still be accomplished with reactions *2.7.1.40a* and *2.7.9.2*, but using ATP and ADP as cofactors. Best double deletion strategy under mixotrophy was a combination of best single deletion candidate, pyruvate kinase (reaction *2.7.1.40c* in *i*Syn669 and genes *sll0587* and *sll1275*) and third best single deletion candidate, phosphoenolpyruvate carboxylase (reaction *4.1.1.31* in *i*Syn669 and gene *sll0920*) that has a succinate evolution of 0.9476 mmol g_{DW}^{-1} h^{-1} with a growth rate of 0.0847 mmol g_{DW}^{-1} h^{-1}.

Table 4.2 - Proposed double knock outs for an improved succinate production. Units for *objective function* (biomass production) and *design function* (succinate production) in mmol g_{DW}^{-1} h^{-1}.

Pair of reactions	Pair of genes	% of wild type objective function	Design function	Times improvement on design function
2.7.1.40d, 4.1.1.31	pykF, ppc	47.3270	0.947654	6180
2.7.1.40d, 2.7.1.40c	pykF, pykF	94.1968	0.853623	5566
2.7.1.40d, 2.6.1.1a	pykF, aspC	92.0526	0.792371	5167
2.7.1.40d, 1.1.1.38	pykF, me	96.3736	0.684984	4467
2.7.1.40d, 2.7.4.6c	pykF, ndkR	95.3242	0.639151	4168
2.7.1.40d, 4.2.1.2	pykF, fumC	95.5193	0.637601	4158
2.7.1.40d, 2.2.1.2	pykF, talB	86.5027	0.635146	4142
2.7.1.40d, 2.7.7.6b	pykF, rpoABC1C2	95.3640	0.624174	4070
2.7.1.40d, 3.6.1.1	pykF, ppa	98.4841	0.621069	4050
2.7.1.40d, 2.7.7.7b	pykF, dnaEX	95.7289	0.618392	4033

Furthermore, the best triple knock out under autotrophy was found to be the combination of dihydrodipicolinate reductase (either be reaction *_1.3.1.26a*, that uses NADPH, or reaction *_1.3.1.26b*, that uses NADH, in *i*Syn811 and gene *sll1058*), cytochrome c oxidase (either be subunit I, reaction *_cit c* in *i*Syn811 and gene *slr1137*, or subunit II, reaction *_1.9.3.1* in *i*Syn811 and gene *sll0813*), and CysQ protein homolog that pumps ammonia to the cell using NADPH (reaction *ammonia TRANS-RXN59G-178* in *i*Syn811 and gene *sll0895*). In fact, the combinations of these genes make up top four triple knock out strategies. Best strategy has a succinate evolution of 0.3478 mmol g_{DW}^{-1} h^{-1} (68361 times higher than wild type strain) with a growth rate of 0.0482 mmol g_{DW}^{-1} h^{-1} (53.85% of wild type growth).

These studies on knock outs are reaction centred, even though the *in vivo* knock out building will ultimately be through gene manipulations. This is the reason underlying the fact that we found *2.7.1.40c* knock out as the best result. This design would hint at the idea of selection of a mutated pyruvate kinase protein specific for ATP cofactor. This may be difficult to achieve on the bench, but has high biotechnological expectations.
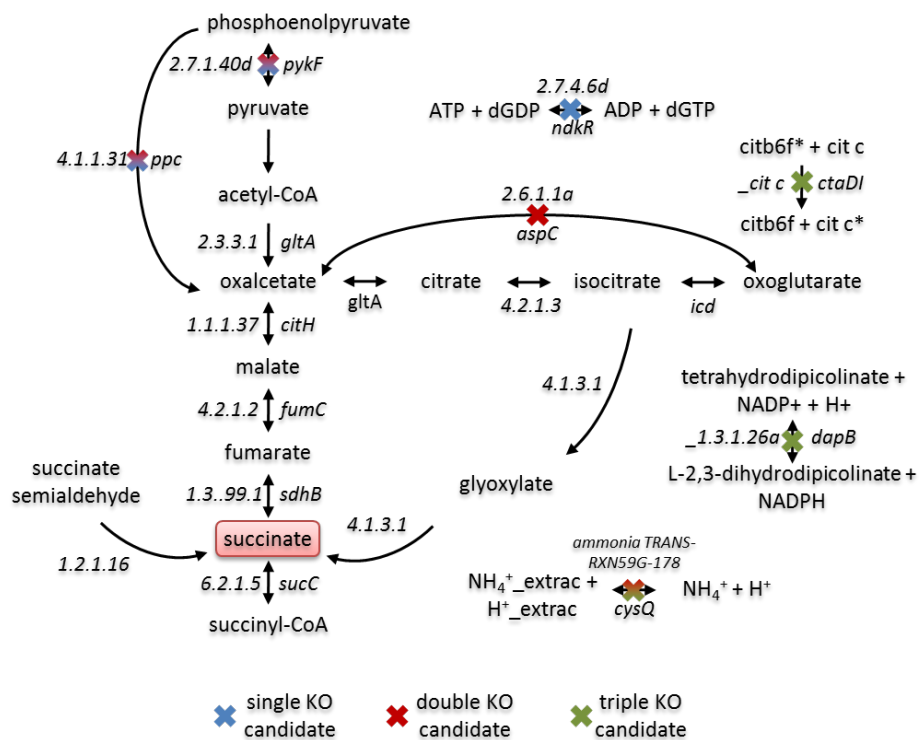

Figure 4.2 - Schematic view of candidate knock outs for succinate production. Single, double and triple knock out candidates are depicted on the pathways that produce succinate. Note that reaction *2.7.1.40d* uses dGTP and has been annotated as reversible

Additionally, theoretical productivity of succinate was probed by solving a series of linear optimization problems. Using autotrophic, mixotrophic and *dark* heterotrophic growth conditions, and fixing carbon and light feeds, theoretical succinate production rate *vs.* growth was studied (Figure 4.3).
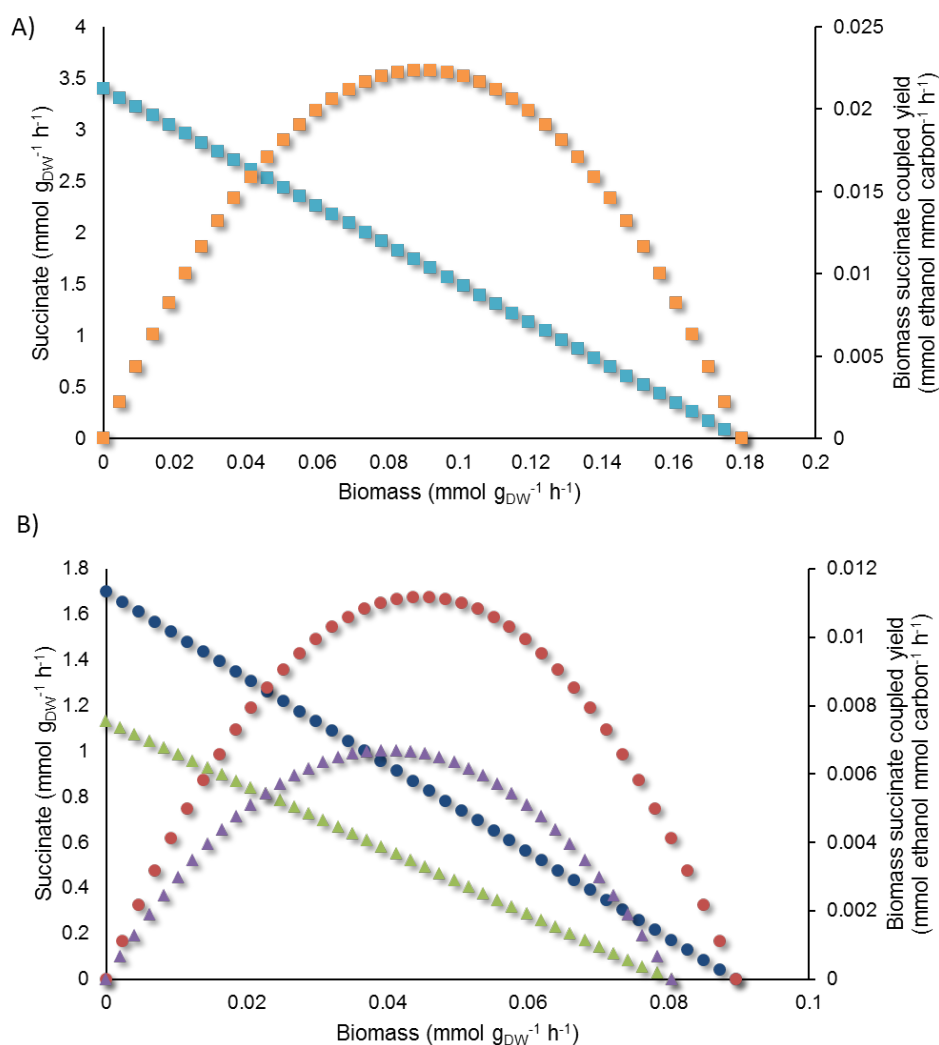
Figure 4.3 - Theoretical productivity of succinate as predicted by using *i*Syn811. Maximum ethanol production rate and biomass-product coupled yield are shown as a function of minimal demand on biomass formation under A) mixotrophic growth and B) autotrophic and heterotrophic growth. A) Blue squares represent succinate formation rate. Orange squares represent biomass succinate coupled yield (a measure of productivity). B) Dark blue circles represent the succinate production rate as a function of growth under autotrophy, green triangles the same under dark heterotrophy. Red circles represent biomass succinate coupled yield under autotrophy. Violet triangles represent productivity of succinate under dark heterotrophy.

Under autotrophy, maximum succinate productivity was reached at specific growth rate of 0.04476 h^{-1} with 0.85 mmol g_{DW}^{-1} h^{-1} of succinate production flux. Under *dark* heterotrophy, maximum succinate productivity was reached at 0.0403

h^{-1} with 0.567 mmol g$_{DW}$$^{-1}$ h^{-1} yield of succinate. Values of *photoheterotrophy* are comparable to dark heterotrophy. Under mixotrophy, maximum succinate productivity was reached at 0.0895 h^{-1} with production rate of 1.7 mmol g$_{DW}$$^{-1}$ h^{-1}.

b) Ethanol.

Another interesting compound of socio-economic importance that can be a potential product of *Synechocystis* cell factory is ethanol. As we did with succinate, theoretical productivity of ethanol was probed using FBA, performing it in triplicate in order to check for possible multiple solutions on the design function.

OptGene simulations for single, double and triple knock out strategies were performed to obtain solutions with improved ethanol production, but without drastically diminishing the biomass production. A schematic view can be found in Figure 4.4. We used autotrophic conditions on *i*Syn811, for which wild type optimal growth rate was 0.08909 mmol g$_{DW}$$^{-1}$ h^{-1}.

Table 4.3 - Proposed single knock outs for an improved ethanol production. Units for *objective function* (biomass production) and *design function* (ethanol production) in mmol g$_{DW}$$^{-1}$ h^{-1}.

Reaction in model	Gene	% of wild type objective function	Design function	Times improvement on design function
2.7.4.6d	ndkR	24.0458	0.0120599	1940
_1.6.5.3	ndhF4	82.1790	0.00401518	646
3.6.1.1	ppa	6.5486	0.00245396	395
2.7.4.3a	adk	98.8572	0.00117039	188
3.5.1.2a	glutaminase	96.1825	0.00108895	175
Phosphate TRANS-RXN59G-90	sll0681	99.1500	0.00069941	113
2.1.3.2	pyrB	97.9768	0.00045113	73
4.1.1.23	pyrF	97.9768	0.00045113	73
3.5.2.3	pyrC	97.9771	0.00045064	73
1.3.3.1	pyrD	97.9771	0.00045064	73

The best single knock out was found to be the mutant of nucleoside-diphosphate kinase (reaction *2.7.4.6d* in *i*Syn811 and gene *sll1852*) that has an

ethanol evolution of 0.0120599 mmol g_{DW}^{-1} h^{-1} with a growth rate of 0.0215255 mmol g_{DW}^{-1} h^{-1} (Table 4.3). Blocking this reaction, preventing dGDP from using ATP to get dGTP would drive a high increase in ethanol production. Interestingly, second best strategy would be to knock out NADH dehydrogenase that starts oxidative phosphorylation pathway.

Table 4.4 - Proposed double knock outs for an improved ethanol production. Units for *objective function* (biomass production) and *design function* (ethanol production) in mmol g_{DW}^{-1} h^{-1}.

Pair of reactions	Pair of genes	% of wild type objective function	Design function	Times improvement on design function
·2.3.3.9, _cit c	malate synthase, ctaDI	47.6568	0.00873415	2930
·4.1.3.1, _1.9.3.1	isocitrate lyase, ctaCDE	47.6567	0.00873414	2930
_1.3.99.1, _cit c	sdhB, ctaDI	47.6568	0.00873413	2930
_1.3.99.1, _1.9.3.1	sdhB, ctaCDE	47.6568	0.00873411	2930
·2.3.3.9, _1.9.3.1	malate synthase, ctaCDE	47.6569	0.0087341	2930
·4.1.3.1, _cit c	isocitrate lyase, ctaDI	47.6569	0.00873409	2930
2.2.1.2, _1.9.3.1	talB, ctaCDE	45.8546	0.00842123	2825
2.2.1.2, _cit c	talB, ctaDI	45.8534	0.00842067	2825
_1.9.3.1, Phosphate TRANS-RXN59G-90	ctaCDE, sll0681	47.7421	0.00798751	2680
_cit c, Phosphate TRANS-RXN59G-90	ctaDI, sll0681	47.7418	0.00798738	2680

Double deletion improved the results from the single knock out strain, evolving 50% more ethanol with 20% more growth rate (Table 4.4). This strain would have two genes knocked out: malate synthase from glyoxylate shunt (reaction *·2.3.3.9* in *i*Syn811) and cytochrome c oxidase (subunit I, reaction *_cit c* in *i*Syn811 and gene *slr1137*) in the thylakoid membrane. This would direct biomass flux to ethanol production reducing growth in half.
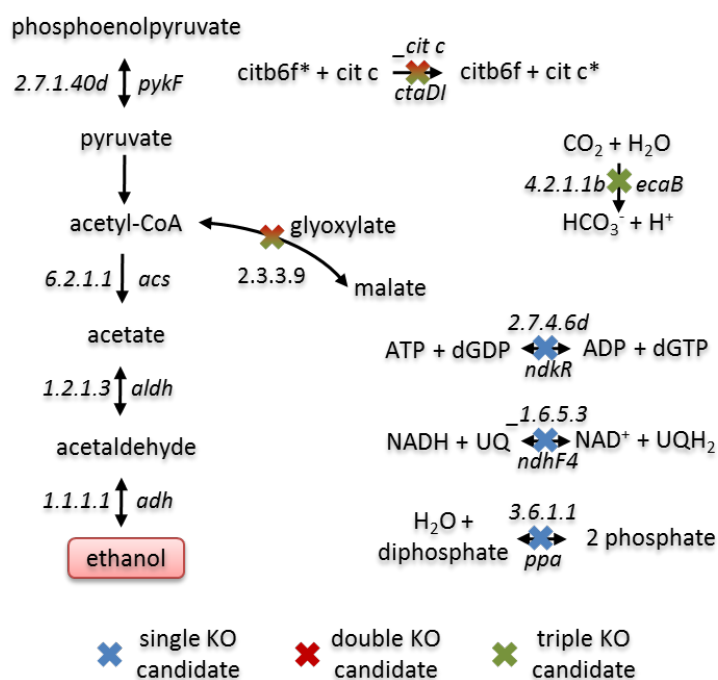
Figure 4.4 - Schematic view of candidate knock outs for ethanol production. Single, double and triple knock out candidates are depicted on the pathways that produce ethanol.

Furthermore, the best triple knock out was found to be the combination of malate synthase (reaction ·2.3.3.9 in *i*Syn811), cytochrome c oxidase (reaction _*cit c* in *i*Syn811 and gene *slr1137*) and periplasmic beta-type carbonic anhydrase (reaction *4.2.1.1b* in *i*Syn811 and gene *slr0051*). This simulated strain has an ethanol evolution of 0.0134116 mmol g_{DW}^{-1} h^{-1} (4500 times higher than wild type strain) with a growth rate of 0.0524567 mmol g_{DW}^{-1} h^{-1} (59% of wild type growth). This design combines the blocking of glyoxylate shunt and the cytochrome c flux with the deletion of the interconversion of CO_2 to HCO_3^-. This would lead to a situation where acetate is going to TCA cycle *only* what is strictly needed for biomass with the rerouting of flux in thylakoid membrane. Second best triple knock out is the same as the best, substituting gene *slr1137* with gene *sll0813*, both part of the cytochrome c oxidase complex (Additional file 4.2).
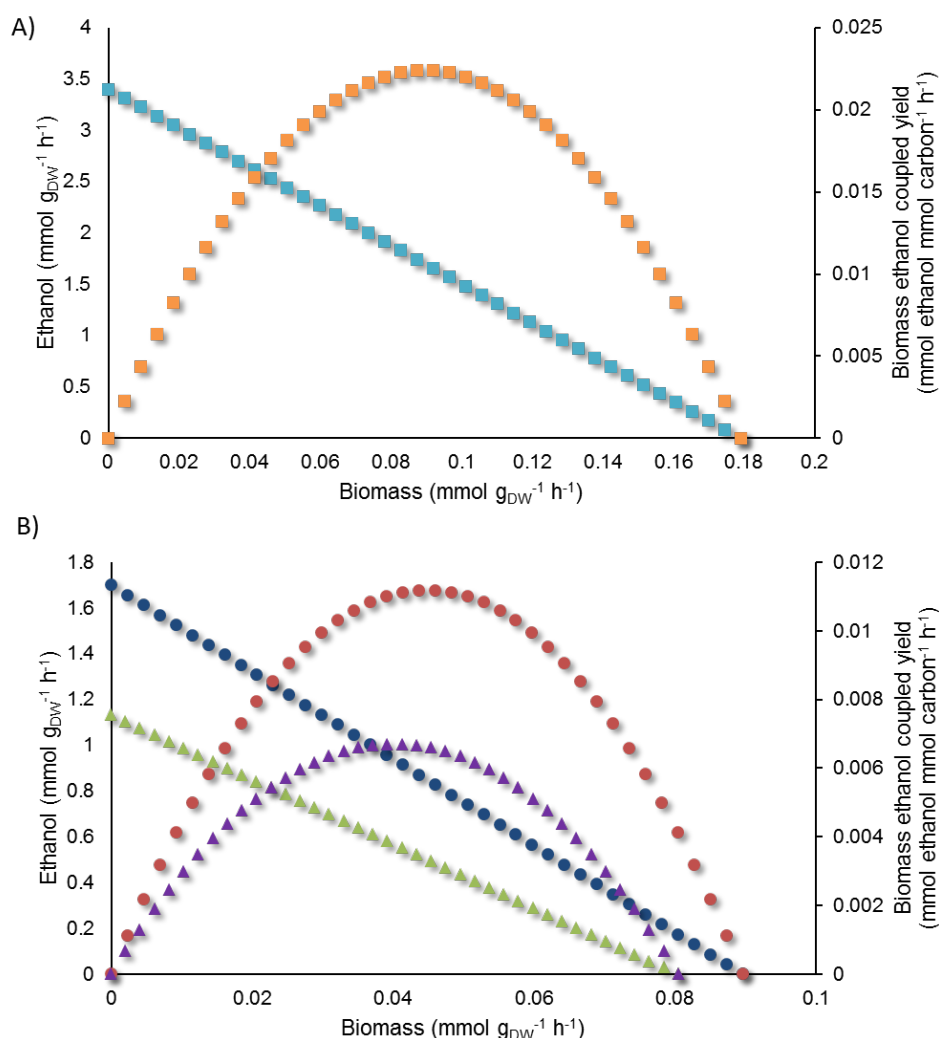
Figure 4.5 - Theoretical productivity of ethanol as predicted by using *i*Syn811. Maximum ethanol production rate and biomass-product coupled yield are shown as a function of minimal demand on biomass formation under (A) mixotrophic growth and (B) autotrophic and heterotrophic growth. A) Blue squares represent ethanol formation rate. Orange squares represent biomass ethanol coupled yield. B) Dark blue circles represent the ethanol production rate as a function of growth under autotrophy, green triangles the same under dark heterotrophy. Red circles represent biomass ethanol coupled yield under autotrophy. Violet triangles represent productivity of ethanol under dark heterotrophy.

Under autotrophy, maximum ethanol productivity was reached at specific growth rate of 0.0448 h^{-1} with 0.85 mmol g_{DW}^{-1} h^{-1} of ethanol production flux (Figure 4.5). Under dark heterotrophy, maximum ethanol productivity was reached at 0.0402 h^{-1} with 0.567 mmol g_{DW}^{-1} h^{-1} yield of ethanol. Values of

photoheterotrophy are comparable to dark heterotrophy. Under mixotrophy, maximum ethanol productivity was reached at 0.0895 h^{-1} with production rate of 1.7 mmol g_{DW}^{-1} h^{-1}. Note that theoretical limits of succinate and ethanol are very similar due to their closeness in the metabolism, sharing many precursors that behave similarly when over-producing these metabolites.

Even though *Synechocystis* sp. PCC6803 has negligible production of this metabolite in wild type strains (P. Wright, personal communication), recombinant strains that have genes from other cyanobacteria can be used instead for comparative purposes. Joule Unlimited (www.jouleunlimited.com) has reported in the patent literature a cyanobacterium that secretes ethanol at a rate of 1 mg L^{-1} h^{-1} (Devroe et al., 2010), and academic literature report typical levels of 0.2 mg L^{-1} day^{-1} (Deng and Coleman, 1999).

c) *Hydrogen.*

A product of obvious interest is *hydrogen*. Hydrogen production drains energy from NADPH pool (if we consider native *Synechocystis* hydrogenase) and, thus, is coupled to photosynthesis, if we pretend to produce it in an autonomous manner. To do things worse, oxygen, a by-product of photosynthesis, competes with protons for hydrogenase, inhibiting the production of hydrogen. As a matter of fact, oxygen also competes for electrons, further reducing the hydrogen production capacity. Previously, we estimated maximum theoretical hydrogen production values that are far from the current state of experimental reports (Navarro et al., 2009). We propose here, and in **Chapters 5 and 8**, *in silico* studies that can direct the efforts and counsel the scientists towards a hydrogen producing cyanobacterium that could be of impact.

OptGene simulations for single, double and triple knock out strategies were performed to obtain solutions with improved hydrogen production, but without drastically diminishing the biomass production. Simulations were performed in triplicates in order to check for possible multiple solutions on the design function. We used autotrophic conditions on *i*Syn811, for which wild type optimal growth rate was 0.08909 mmol g_{DW}^{-1} h^{-1}. It is important to consider that this knock out simulations work with a steady state behaviour of the metabolism. At present, hydrogen production in cyanobacteria has been reported to be a transient, and almost residual, phenomenon. Hence, knock out results should be taken as potential modifications of a continuous hydrogen producer.
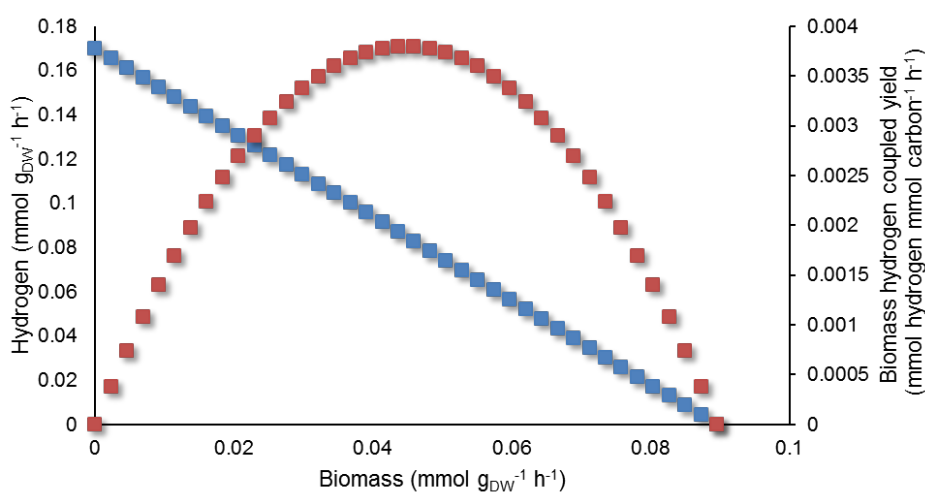
Figure 4.6 - Theoretical productivity of hydrogen as per *i*Syn811. Productivity of hydrogen was studied by maximizing H$_2$ evolution at different minimal demand constraints on biomass formation under autotrophic growth condition. Blue squares represent the value of hydrogen evolution, while red squares represent biomass hydrogen coupled yield –a measure of productivity of hydrogen, the product of hydrogen and biomass production rates.

The best single knock out was found to be the mutant of a component of methyleneTHF enzyme, specifically the enzyme responsible of the methylation of 10-formyl-tetrahydrofolate (reaction *3.5.4.9a* in *i*Syn811 and gene *sll0753*) that has a hydrogen evolution of $1.66 \cdot 10^{-6}$ mmol g_{DW}^{-1} h^{-1} with a growth rate of 0.0894655 mmol g_{DW}^{-1} h^{-1}. Blocking this reaction prevents the recycling of oxidized redox cofactors. Best second mutation would involve blocking glucose-6-phosphate isomerase, inhibiting such a reaction from the upper part of glycolysis should not have effect on growth titters, but increases hydrogen evolution. Anyway, there is not a single knock out candidate that increases hydrogen evolution more than twice the wild type's titters.

Double deletion greatly increased results from the single knock out strains. Best double knock out is the combination of deletions of transaldolase and carbamate kinase (reactions *2.2.1.2* and *2.7.2.2a* and genes *talB* and *arc*). This strain had an hydrogen evolution of $5.80 \cdot 10^{-5}$ mmol g_{DW}^{-1} h^{-1}, 81 times higher than wild type, and with a growth rate of 0.0884473 mmol g_{DW}^{-1} h^{-1}. This strain would have a disabled pentose phosphate pathway and the fixation of carbon dioxide into carbamoyl-phosphate. This would optimize the carbon incorporation through RuBisCO, having more NADPH to direct electrons to hydrogen.

Table 4.5 - Proposed single knock outs for an improved hydrogen production. Units for *objective function* (biomass production) and *design function* (hydrogen production) in mmol g_{DW}^{-1} h^{-1}.

Reaction in model	Gene	% of wild type objective function	Design function	Times improvement on design function
3.5.4.9a	*folD*	99.9407	1.66E-06	2
5.3.1.9c	*pgi*	99.9602	1.42E-06	2
_2.3.2.2p	*ggt*	99.9640	1.34E-06	2
magnesium TRANS-RXN59G-53	*rfrO*	99.9618	1.23E-06	2
_1.1.1.86	*ilvC*	99.8987	1.20E-06	2
1.4.1.4	*gdhA*	99.9649	1.16E-06	2
_2.4.1.21b	*glgA*	99.9682	1.11E-06	2
_2.4.1.21a	*glgA*	99.9632	1.10E-06	2
_2.3.2.2m	*ggt*	99.9722	1.09E-06	2
2.7.4.1	*ppk*	99.9703	1.08E-06	2

Furthermore, the best triple knock out was found to be the combination of transaldolase (reaction *2.2.1.2* in *i*Syn811and gene *slr1793*), cytochrome c oxidase (reaction *_cit c* in *i*Syn811and gene *slr1137*), and aldehyde dehydrogenase (reaction *1.2.1.3c* in *i*Syn811 and gene *slr0091*). This simulated strain has a hydrogen evolution of 0.014248 mmol g_{DW}^{-1} h^{-1} (1788 times higher than wild type strain) with a growth rate of 0.040269 mmol g_{DW}^{-1} h^{-1} (45% of wild type growth). This design combines shutting down a critical enzyme for the pentose phosphate pathway, such as transaldolase, and the production of acetate from acetaldehyde and the inhibition of cytochrome c oxidase. This would increase the available NADPH for hydrogen production (Additional file 4.2).

Table 4.6 - Proposed double knock outs for an improved hydrogen production.
Units for *objective function* (biomass production) and *design function* (hydrogen production) in mmol g_{DW}^{-1} h^{-1}.

Pair of reactions	Pair of genes	% of wild type objective function	Design function	Times improvement on design function
2.2.1.2, 2.7.2.2a	sll0573, talB	98.8033	5.80E-05	81
3.5.4.9a, ·1.3.1.75b	divinyl chlorophyllide a reductase, folD	99.9395	3.05E-06	4.28
3.5.4.9a, _3.6.3.27	sll0683, folD	99.9419	2.99E-06	4.19
3.5.4.9a, _3.6.3.5	slr0797, folD	99.9411	2.98E-06	4.17
3.5.4.13a, 3.5.4.9a	sll0753, dcd	99.9409	2.97E-06	4.17
3.5.4.9a, 3.4.11.2a	sll1343, folD	99.9351	2.96E-06	4.14
_4.2.1.20b, 3.5.4.9a	sll0753, trpB	99.9435	2.95E-06	4.14
5.3.1.9c, 3.5.4.9a	sll0753, pgi	99.9407	2.94E-06	4.13
3.5.4.9a, 3.1.3.5k	sll1108, folD	99.9420	2.93E-06	4.11
3.5.4.9a, ·2.7.1.1	hexokinase type IV, folD	99.9420	2.93E-06	4.11

As in case of ethanol, theoretical limit on the productivity of hydrogen was simulated with FBA. Using autotrophic growth conditions, and fixing carbon and light feeds, H_2 evolution was studied as a function of minimal demand on biomass formation (Figure 4.6). Maximum hydrogen productivity was observed at specific growth rate 0.0448 h^{-1} with corresponding maximum H_2 production rate 0.085 mmol g_{DW}^{-1} h^{-1}. Wild type production of hydrogen is highly dynamic in nature, being present mostly at the beginning of photosynthetic activity, this production usually reaches values of $3.149 \cdot 10^{-4}$ mmol g_{DW}^{-1} h^{-1} even though recent genetic engineering has been able to increase it to 0.004985 mmol g_{DW}^{-1} h^{-1}, both after 48h cultivation (Baebprasert et al., 2011).

As it is known, usually wild-type *Synechocystis* produces hydrogen in a transient manner; some authors have proposed that the organism may use hydrogen as temporal electron storage (Tamagnini et al., 2007). This dynamic behaviour and the fact that its production is coupled with NADPH pool, makes it terribly difficult to give a realistic list of knock outs with improved hydrogen yield. Further thoughts and works on this can be found in **Chapter 8**.

4.3 Conclusions.

We have here demonstrated the use of *Synechocystis* as a photon-fuelled production platform in order to generate metabolite of interest such as succinate, ethanol and hydrogen. Among these, hydrogen is the one that has gathered more interest. Evaluation of the theoretical potential of this organism to produce hydrogen was assessed, in support of the efforts directed to this direction from several groups (Lopes Pinto et al., 2002; Navarro et al., 2009; Pinto et al., 2011; Tamagnini et al., 2007) and scientific council initiatives (BioModularH2, 2005). Present hydrogen production projects are far from the theoretical potential, but efforts in this field can trigger a very significant increase of the present hydrogen evolution rates in *Synechocystis* sp. PCC6803 or other photobiological production platforms candidates, *e.g. Chlamydomonas reinhardtii*, *Nostoc punctiforme* and *Synechococcus* species.

Single reaction/gene knock out simulations revealed 311 genes that are essential for the survival. Bearing in mind the distance from the efforts taken in the annotation of the genome of the bacteria and yeast models to that of the cyanobacterium, our study shows that *Synechocystis* sp. PCC6803 has a larger fraction of genes that are essential for producing biomass, as opposed to *Escherichia coli* and *Saccharomyces cerevisiae*. Further investigation of the causes for this difference will be of definite interest in understanding the genome annotation and/or the evolution of the metabolic network of *Synechocystis* sp. PCC6803.

4.4 Methods.

Minimization of metabolic adjustment algorithm

Segrè et al. (2002) introduced the method of *minimization of metabolic adjustment* (MOMA) to better understand the flux states of mutants. MOMA is based on the same stoichiometric constraints as FBA, but relaxes the assumption of optimal growth flux for the mutants, testing the hypothesis that the corresponding flux distribution is better approximated by the flux minimal response to the perturbation than by the optimal one.

MOMA algorithm searches for a point in the feasible space of the solutions space of the knock out (Φ^j) that has minimal distance from a given flux vector w. The goal is to find the vector $x \in \Phi^j$ such that the Euclidean distance

$$D(w, x) = \sqrt{\sum_{i=1}^{N}(w_i - x_i)^2}$$

is minimized. For details, please address to **Chapter 1** and to Segrè et al (2002).

We have used *minimization of metabolic adjustment* algorithm with OptGene (Patil et al., 2005), nowadays available at BioMet Toolbox (Cvijovic et al., 2010).

All knock out simulations using FBA or MOMA were done in triplicate in order to check for multiple solutions on the design functions.

En la vida, igual que en la literatura, no se descubre
tierra nueva sin acceder a perder de vista primeramente, y
por largo tiempo, toda costa.
[In life, as in literature, new land is not found unless one
consents firstly to lose sight, and for a long time, all
coastlines]

Enrique Vila-Matas, citing André Gide, Llavaneres, 28
June 2009

5

Flux coupling analysis of *Synechocystis* sp. PCC6803

Where PhD applicant digs into fluxes behaviour and finds functional correlations among reactions, some of them being straight-forward to explain, while others are more complex to elucidate.

Parts of the contents of this chapter are based on parts of the following journal article:

- Montagud et al. **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

5.1 Introduction.

In the postgenomic era, each cellular function, biological actor or physiological event can be seen in the context of a complex network of interactions. Variations in cellular components, or environmental fluctuations, will have local, but also system-wide effects, as we have seen in previous chapter. Cells are open systems that affect, and are affected by, the environment. This chapter studies one of the tools that cells use in order to cope with environment changes and genetic fluctuations.

One of the interesting approaches for understanding the operating principles and capabilities of a microorganism is to analyse the reactions functional relationships at steady state –an approach termed *flux coupling analysis* and presented by Burgard et al. (2004). Quoting this work, "*An overarching attribute of metabolic networks is their inherent robustness and ability to cope with ever-changing environmental conditions. Despite this flexibility, network stoichiometry and connectivity do establish limits/barriers to the coordination and accessibility of reactions*." *Flux coupling analysis* focuses in finding these limits where network's flexibility does not reach and does not allow feasible flux behaviour. This algorithm has been used to study flux capabilities of *Saccharomyces cerevisiae* and *Escherichia coli* (Notebaart et al., 2008) and has facilitated metabolic flux analysis (Suthers et al., 2010). It allows the identification of blocked reactions and functional reactions subset as other works (Kholodenko, 1995; Klamt et al., 2003; Pfeiffer et al., 1999; Rohwer et al., 1996; Schilling et al., 2000; Schuster et al., 1994), but it circumvents the problems that these algorithms have when handling large networks, such as genome-scale metabolic networks (Golub and Van Loan, 1996).

We first presented in **Chapter 1** an updated genome-scale metabolic model that has been thoroughly tested for the use with the steady state metabolic simulation algorithms (see **Chapter 3 and 4**). This model has been the basis of *flux coupling analysis* which uncovers coupling potential among reactions. That is, the algorithm points to reactions whose flux is affected upon change in a given reaction's flux, uncovering behavioural relationships that may not be straight-forward to guess and see.

Genome-scale metabolic modelling approach has been applied to a diversity of organisms in a variety of conditions in the context of metabolic engineering and systems biology. In this sense, *Synechocystis* metabolic models, like the ones from

Chapter 2 are a good start to use these *in silico* design approaches toward creating a resilient photosynthetic bio-refinery.

5.2 Blocked reactions.

Reactions that could not carry steady state flux for a given set of environmental constraints (autotrophic, mixotrophic, *dark* heterotrophic, or, *light-activated* heterotrophic, see Methods] were identified as blocked reactions.

*i*Syn811 blocked reactions range between 39.45% under autotrophy to 41.25% under *dark* heterotrophic, some more than *Escherichia coli* (28%) (Edwards and Palsson, 2000a) and similar to *Saccharomyces cerevisiae* (39.2%) (Förster et al., 2003) both under aerobic heterotrophy (Table 5.1). Focusing on the different growth conditions of *Synechocystis*, mixotrophy stands as the condition with the least number of blocked reactions, but not far from the other conditions. Autotrophy has the import of glucose and the phosphorylation of glucose blocked. Many of the folate biosynthesis reactions, photosynthesis reactions, and Calvin cycle (CO_2 fixation) can have non-zero steady state flux only under autotrophic and mixotrophic conditions. *Dark* heterotrophic has most of the photosynthesis pathway blocked and, as *light-activated* heterotrophic and mixotrophy, allows flux in the upper part of the glycolysis and the import of glucose. Mixotrophy, as the condition with least blocked reactions, stands in between autotrophy and heterotrophy: flux is feasible through folate biosynthesis, Calvin cycle, photosynthesis as well as all the glycolysis; additionally, many coenzyme A biosynthesis reactions have potential non-zero flux only in this condition.

Blocked reactions in *i*Syn811 are in comparable numbers to that of *E. coli* and *S. cerevisiae* (Table 5.1). Altogether, almost 40% of the known metabolic reactions in those genome-scale models cannot carry flux under commonly used growth conditions. Among the main reasons for this apparent dispensability are: incomplete description of the biomass composition and medium composition used in simulation. With that said, it is not awkward to consider that a large number of metabolic reactions are possibly active only under environmental conditions not typically used in laboratory studies. Differences observed between the different growth modes of *Synechocystis* sp. PCC6803 are in accordance with the physiological information and previous modelling studies (Navarro et al., 2009): autotrophy has glucose transport and catabolism blocked and in heterotrophy, blocked reactions are among photons usage and carbon fixation. Mixotrophy stands as the growth mode in which less blocked reaction are

present. In fact, this mode has more uptake freedom degrees than the previous: glucose and CO_2 can be used as carbon source and glucose and photons as energy source.

Table 5.1 - Flux blocked reactions across different species. Data from *Escherichia coli* and *Saccharomyces cerevisiae* were retrieved from Burgard et al (2004). LH stands for photoheterotrophy and DH stands for *dark* heterotrophy.

		Blocked	Total
Escherichia coli		207	740
	%	28	100
Saccharomyces cerevisiae		460	1173
	%	39.2	100
Synechocystis **autotrophic**		503	1275
	%	39.45	100
Synechocystis **mixotrophic**		497	1275
	%	38.9	100
Synechocystis **LH**		524	1275
	%	41.09	100
Synechocystis **DH**		526	1275
	%	41.25	100

5.3 Flux coupling study.

We adapted the flux coupling finder procedure developed by Burgard et al. (2004) to analyse the functional associations between the reactions of the genome-scale metabolic network of *Synechocystis* sp. PCC6803 across the four different simulated growth conditions. This constraint-based modelling approach relies on minimization and maximization of the intracellular flux ratios to determine the extent of the dependency between any two reactions within the network, given the mass-balance constraints and the exchange fluxes with the environment, as Figure 5.1 summarizes. A detailed description of the algorithm can be found in Methods.

As reported by Burgard et al. (2004), an imposed constant stoichiometry biomass composition leads to the generation of one large coupled reaction set that is mostly fully coupled. This biomass reaction serves to the simulation purpose of draining the compounds necessary for cell growth (*e.g.* amino acids

and nucleotides) in a pre-specified stoichiometry. We have seen that the coupling sets of networks with and without biomass-coupled reaction are essentially the same in terms of diversity and relative fraction of different types of couplings (data not shown). In fact, only one set of reactions, which is the drain of monomers to the biomass reaction, is additionally present in the latter (set that represents 67 reactions, 5.25% of the total reactions in *i*Syn811). As the biomass reaction is an abstraction of growth and has purely simulation purposes, we here present only the results of networks with independent biomass monomer drains under the four studied growth modes.
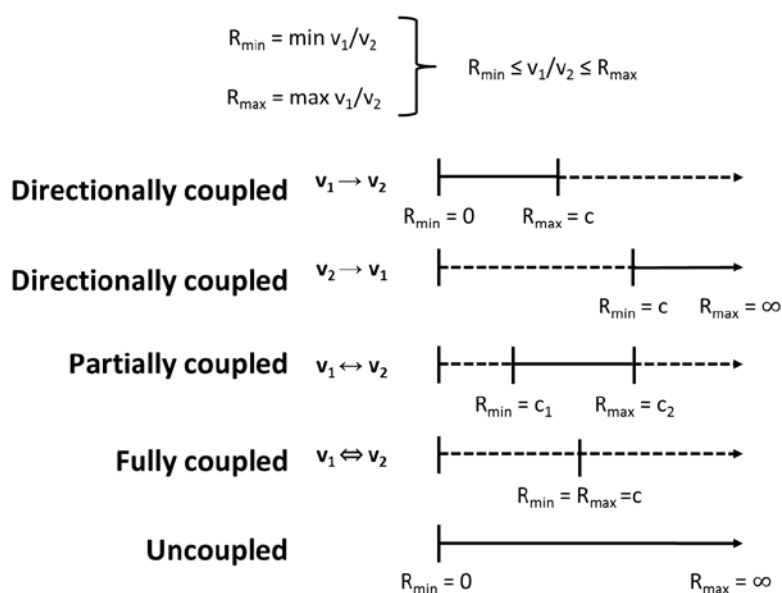

Figure 5.1 - Graphical representation of flux couplings. For details, refer to Methods and to work from Burgard et al. (2004).

*i*Syn811 presents similar relative distribution of coupling patterns as *E. coli* and *S. cerevisiae*, with twice more directionally coupled sets than fully coupled and almost one fifth of partially coupled groups (Table 5.2). Furthermore, the *Synechocystis* model portrays less fully coupled reactions and two-fold more partially coupled reactions than *S. cerevisiae* model. The distribution of coupled reactions is shown in Table 5.2 and the complete sets of coupled reactions are provided in Additional file 5.1.

Table 5.2. Relative distribution of reactions among different flux coupling relationships. Percentages of reactions in each coupling type are depicted in shaded cells. A reaction can participate in more than one coupling set and hence the number of reactions in a coupled group do not necessarily sum up to the total number of reactions. LH stands for photoheterotrophy and DH stands for *dark* heterotrophy.

		Directionally coupled	Fully coupled	Partially coupled	Reactions in model
Escherichia coli		421	353	—	1176
	%	35.8	30		
Saccharomyces cerevisiae		473	265	44	1204
	%	39.3	22	3.6	
***Synechocystis* autotrophic**		527	214	113	1275
	%	41.3	16.8	8.8	
***Synechocystis* mixotrophic**		517	226	111	1275
	%	39.4	16.4	8.7	
***Synechocystis* LH**		517	227	111	1275
	%	40	16	8.7	
***Synechocystis* DH**		512	213	111	1275
	%	39.6	15.7	8.7	

Couplings of the three metabolic modes.

- *Autotrophy*

In this growth condition, light is the source of energy and molecular CO_2 of carbon skeletons. Photosynthesis is essential in this mode in order to fulfil the cellular needs of ATP and redox potential. Looking at the effect of growth conditions on *flux coupling analysis*, we observed that the autotrophy has slightly less fully coupled sets and a few more directionally coupled reactions (Figure 5.2 and Table 5.2). An interpretation of the different coupling cases from a biological perspective can be found in Figure 5.3.

Most coupled reactions are directionally coupled, with the corresponding reactions spanning all functional pathways in the *Synechocystis* metabolic model. Directionally coupled reactions are relationships where activity in one reaction obliges another to have activity but not the other way around. The carbon fixation and the entrance of CO_2 to the cell stand in the centre of the network together with ATP production in the thylakoid membrane (which is the node with the highest degree, 254), leaving the rest of the coupling sets to stem from them.

Notably, almost all the fatty acid biosynthesis is directionally coupled to many central pathways as glycolysis, pentose phosphate pathway and Calvin cycle. Almost all reactions from photosynthesis are directionally coupled (only 5 couplings are left to be fully coupled) and in such a way that follows the photosynthesis pathway structure, as they are coupled also with the oxidative phosphorylation (Figure 5.4).
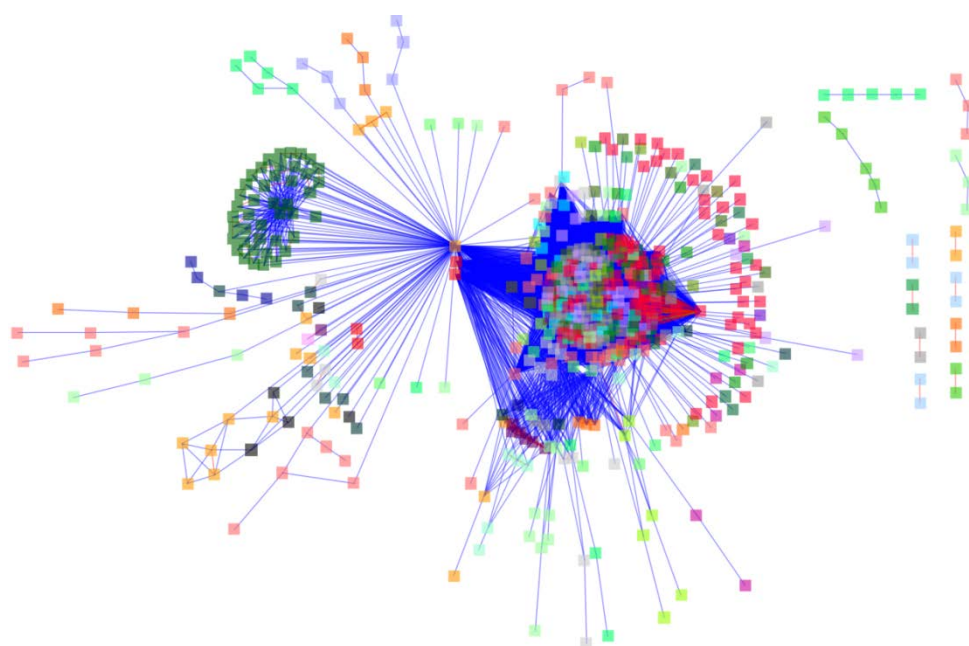

Figure 5.2 - Flux coupling network for autotrophic growth condition. Nodes represent reactions and are coloured according to the pathway they belong to (for instance, orange is photosynthesis, dark grey is oxidative phosphorylation, dark pink is Calvin cycle, dark green is glutathione metabolism, red is fatty acids and different grades of green are different amino acids syntheses). Edges represent flux coupling and are coloured according to the type of interaction (blue - directionally, red - fully and green - partially coupled).

In the fully coupled reaction network there is one big cluster, several medium size clusters (with four or more reactions) and many 2- and 3-mer sets. Fully coupled reactions are reactions whose flux ratios have a constant value, thus activity of one reaction forces another one to have a specific, unique, value. Interestingly, the large cluster (highly interconnected with 102 reactions and 5151 connections) is made up of most of the reactions (11 out of 13) of the chlorophyll pathway interconnected between them and connected with reactions from fatty

acid biosynthesis, from carotenoid biosynthesis, and several amino acids (aspartate, histidine and glutamate, among others). Medium-sized clusters are an 8-mer set for the of NAD(P) metabolism, an 8-mer set for biosynthesis of steroids and a 5-mer for purine metabolism.
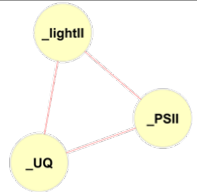

Analysis	Remarks	Example	
Fully coupled reactions	Suitable for the application of metabolic control analysis for the identification of rate controlling step (/s).	_lightII _PSII _UQ	If photosystem II could be engineered to be more efficient in using photons (_PSII), this would foster the initial step of linear photosynthesis (_UQ).
Directionally coupled reactions	Engineering of a target flux should be co-ordinated with the source flux.	4.1.1.39 4.1.1.39oxi _3.6.3.14_forward	ATP synthase is needed to have RuBisCo carboxylase and oxidase activities, but then increased ATP synthase activity may not direct an increased carboxylase activity.
Partially coupled reactions	Metabolic control analysis, as well as independent manipulation of nodes, may be necessary	4.1.1.39 2.7.1.19	Phosphoribulokinase activity has to have flux in a window of values compatible to that of RuBisCO carboxylase. Upper limit is fixed by other constraints, like CO_2 supply.
Reporter pair	Transcriptional co-regulation of the corresponding enzymes under the studied perturbation response. May involve (known or unknown) transcription factors.	1.2.4.1_forward 2.5.1.32	Provides clues for the discovery of new regulatory mechanisms. Engineering of the corresponding fluxes should take into account the co-regulatory patterns (esp. under non-steady state conditions)

Figure 5.3 - Examples of potential applications of *i*Syn881 *flux coupling analysis* for metabolic engineering.

Finally, partially coupled reactions are the reactions that are mutually needed, but whose fluxes can have a range of values. In partially coupled reactions sets, we have found two noteworthy groups. First, a complex of 111 reactions made up of a core of 11 reactions part of the porphyrin and chlorophyll metabolism (specifically, *S*-adenosyl methionine, or SAM, formation) and 17 reactions part of the fatty acid biosynthesis (malonyl-CoA formation and its union with acyl carrier protein are coupled with the oligomerization reaction of lipids). The second group is made of the carboxylative reaction of the ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) and the phosphoribulokinase reaction, this is, the reaction that fixes the atmospheric carbon and the one that generates the

substrate in which CO_2 will be fixed. These sets are a good example of partially coupled reactions. For example, flux ratios between SAM formation and chlorophyll synthesis have a minimum value (as chlorophyll *a* needs SAM and is also needed for the cell composition), but the activity on reaction *2.5.1.6* (formation of SAM) is not the only reaction where SAM can be produced and, thus, has a range of possible values in order to fulfil the need of chlorophyll.

- *Heterotrophy*

In this growth condition, energy and carbon come from glucose or other carbohydrates. Cyanobacteria researchers identify this condition as "*dark heterotrophy*", in order to differentiate it from a "photoheterotrophy" condition where photons are permitted to enter the system, hence activating photosynthesis, but with no CO_2 fixation. Little differences in the coupling distribution or network clustering have been found across these two conditions (Table 5.2). The only difference comparing both conditions is that the *light-activated* heterotrophic network has a complete photosynthesis cluster, which embeds the reactions from the oxidative phosphorylation, while the *dark* heterotrophic network has only the photosynthetic reactions needed for a proper oxidative phosphorylation pathway, as both share some elements in the thylakoid membrane.

Directional coupling is a predominant characteristic in this network, even though as a major difference to autotrophic condition, glucose entering the cell and its phosphorylation are centrally located in the network together with ATP synthase, which is again the most connected reaction with a degree of 251. This change alters the networks in such a way that all other groups and cluster are arranged around these reactions. Nonetheless, as in the autotrophic case, almost all of the fatty acid biosynthesis, NAD(P) metabolism and porphyrin and chlorophyll metabolism reactions are directionally coupled between themselves and connected to these two reactions. Moreover, purine and pyrimidine biosynthesis and glutamate and proline pathways are coupled in a linear way, but independent of the core of the network.

Fully coupled reactions sets in *dark* heterotrophic are analogous to autotrophic. The highly-connected big cluster is also found (with 102 reactions and 5151 connections inter and intra pathways), made out of chlorophyll pathway, fatty acid biosynthesis, carotenoid biosynthesis and several amino acids reactions. The similarity in the smaller sets is notable: medium clusters are present, like the ones with reactions of the steroids biosynthesis, NAD(P)

metabolism, and fatty acid biosynthesis and several smaller like most of the glutathione metabolism (forming 2-mers) as well as groups of reactions from fatty acid biosynthesis and amino acids. Nevertheless, some small sets stand out as different, like some reactions from pyruvate metabolism and pentose phosphate pathway, a few reactions from photosynthesis, as well as the appearance of the oxidative reaction of the RuBisCO enzyme that works as a source of glycolate, needed for the glyoxylate shunt.

Heterotrophic partially coupled reactions are different from the ones in the autotrophic: SAM coupled to all the reactions of chlorophyll metabolism and fatty acid biosynthesis are present, but the partial coupling of ribulose-1,5-bisphosphate production and use is no more present. If there is no carbon fixation, it is clear that ribulose-1,5-bisphosphate will not be drained at RuBisCO carboxylase reaction.

- *Mixotrophy*

This condition is a blend of the previous two conditions. Glucose, molecular CO_2 and photons are present in the system; hence, carbon and energy have more than one source. Interestingly, we see a different network topology of coupling networks from the previous conditions: fewer clusters and less connectivity are found in this condition, due to the fact that the mixotrophic condition has more feeding sources and is, thus, more flexible in the flux distribution, *i.e.*, with more degrees of freedom. Molecular carbon fixation and glucose transport and phosphorylation are coupled, but only connected to the ATP production, which is the reaction at the core of the network. This coupling permits the mixed flux behaviour that was observed in the previous works (Montagud et al., 2010; Navarro et al., 2009; Shastri and Morgan, 2005). Photosynthesis cluster is present analogously to the autotrophic condition. In our mixotrophic coupling network, ATP synthase is the most connected reaction with a degree of 256. Furthermore, mixotrophy is the condition with fewer couplings, which is due to the increase in the degrees of freedom of the flux distribution.

As it happens with the other two conditions, directionally coupled reactions are the largest set of coupling pairs in the network. In this case, though, carbon entering the system is detached from the nucleus of the network, leaving that place to the ATP production from which the biomolecular-building pathways stem. Fatty acid biosynthesis, NAD(P) metabolism, pyrimidine and purine biosynthesis, porphyrin and chlorophyll metabolism, sucrose metabolism and TCA

cycle reactions are directionally coupled inter pathways, and some of them also intra pathway.

Fully coupled reactions follow the trend of the other conditions. The highly connected big cluster is present with 102 reactions (and 5151 connections inter and intra pathways): fatty acid biosynthesis, porphyrin and chlorophyll metabolism are included, as well as several amino acids. As in the previous cases, the medium-sized sets include clusters for the biosynthesis of steroids, NAD(P) metabolism and purine metabolism. Smaller sets include 2-mers from Calvin cycle, glycolysis, pentose phosphate pathway, photosynthesis, porphyrin and chlorophyll metabolism and almost all the glutathione metabolism.

Mixotrophic partially coupled reactions are the same as in the heterotrophic case. Feeding of ribulose-1,5-bisphosphate is not found in partial couplings and has gone to the less restrictive directional coupling set due to the increase in the degrees of freedom of the mixotrophic network.

Coupling networks across growth conditions.

In all three conditions, directionally coupled reactions are dominant. Fully coupled reactions follow and partially coupled reactions are far behind (Table 5.2). Even though there are significant differences, the network of coupled reactions has quite similar topology: metabolic precursors needed for cell growth stem out of the reactions responsible of ATP production and carbon transport and metabolism (Calvin cycle in autotrophy, upper part of glycolysis in heterotrophy). Apart from the central nodes of the network, topology is almost identical. This suggests the idea that, once carbon has been metabolized by the cell and energy has been drained from it, the coupling of the anabolic part of the metabolism is independent of the growth condition.

Metabolic flux coupling of *Synechocystis* sp. PCC6803 has evolved around the need to be flexible in the acquisition of skeletons for building blocks, but is conservative on the construction of these building blocks (fatty acids, amino acids, nucleic acids, etc.) as well as on the ATP production. Following the "bow tie" description of metabolism (Csete and Doyle, 2004), *Synechocystis* appears flexible to growth conditions on the first half of the metabolism (catabolism of environmental molecules to construct precursors), but conservative on the anabolism of polymers and complex assemblies. This characteristic can help to explain the wide range of living conditions in which cyanobacteria is found in nature and the scarce needs it has for its growth on the bench (Baebprasert et al., 2011; Whitton and Potts, 2000).

Photosynthesis couplings.

Apart from direct biotechnological designs, *i*Syn881 model and present analyses can also be of use for the study of biochemical questions, like the functioning of the photosynthesis pathway (Figure 5.4). Photosynthesis, present in autotrophy, mixotrophy and photoheterotrophy, is mostly directionally coupled intra pathway and, with the reactions from oxidative phosphorylation, embedding four fully coupled clusters following the entrance of photons to photosystem I and II, the ADP recycling and the NADP+ reduction to NADPH. Photosynthesis and oxidative phosphorylation clusters are directionally coupled to the rest of the network through only one reaction –ATP synthase, which plays a central role in the network, as it is the most connected one in all the conditions and is coupled to reactions from almost all the pathways present in the cell metabolism.
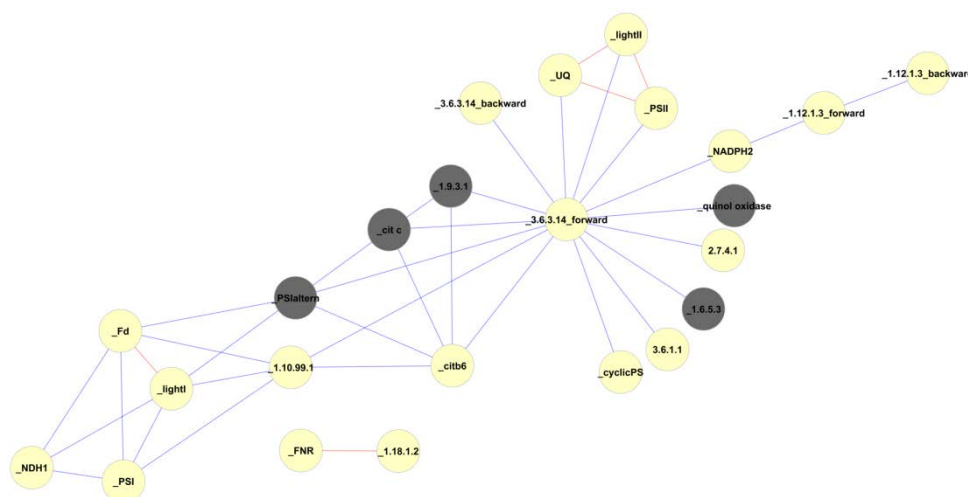

Figure 5.4 - Flux coupling network of photosynthesis and oxidative phosphorylation. Photosynthesis is depicted in light yellow and oxidative phosphorylation in dark grey. Directional couplings are blue edges and full couplings are red.

Photosynthesis can work in a linear manner (that includes photosystem I and II) or in a cyclic manner (around photosystem I). These different routes produce different products: linear evolves ATP and NADPH and cyclic produces solely ATP. In our coupling network, *_PSI* and *_PSII* (reactions for photosystem I and II excitation, respectively) are directionally coupled to ATP synthase, independently to each other. Thus, photosystem I and II are uncoupled between them, meaning

that there is no fixed flux ratio, or fixed ratio values window, between the two reactions. This ratio can be calculated as a result of the photons that actually enter each photosystem (in fact, usually studies look to the easier-to-measure electrons that flow on each photosystem), but the potentialities of the network do not constraint the values of that ratio. Additionally, neither photosystem I, II or ATP synthase are coupled to NADPH reduction in thylakoid membranes (reaction _1.18.1.2_, Figure 5.4). This reaction is fully coupled to _FNR_, reaction that links ferredoxin oxidation to NADPH reduction. The scarce couplings of _1.18.1.2_ can be interpreted as that there are many reactions that reduce NADPH among the metabolism or that reduce NADH and then this is converted to NADPH.

Biofuels and flux coupling.

An obvious interest from these coupling analyses is their effects on biofuel-producing strain design. Here, we have focused on the flux couplings of two industrially-relevant metabolites like *hydrogen* and *ethanol*. For FBA simulations on productivity of those metabolites please refer to **Chapter 4**, section 2.

Hydrogen production is fully coupled to NADPH formation as well as directionally coupled to different pathways linking to the metabolic potential of the cell, like carbon fixation through RuBisCO reaction, 1,3-diphosphateglycerate production in glycolysis and NADP metabolism. This indicates that those reactions may be bottlenecks for boosting the H_2 production in this organism and should be considered as genetic targets when aiming at increased hydrogen production strains.

Another interesting compound of socio-economic importance that can be a potential product of *Synechocystis* cell factory is ethanol. Ethanol production is directionally coupled to the pyruvate metabolism and to ATP synthase, as well as to some reactions from the central carbon metabolism, *e.g.* glycolysis and pentose phosphate pathway. Additionally, it is partially coupled to Calvin cycle under autotrophic condition. In order to increase the ethanol yield, the central carbon metabolism must be engineered –fluxes around pyruvate must be re-routed so as to increase alcohol production. Uptake fluxes of carbon and photons are also predicted targets for interesting genetic modifications. These findings will be useful in exploiting the photo-fermentative metabolism –so called *photanol* strategy (Angermayr et al., 2009).

5.4 Conclusions.

In summary, we have demonstrated the use of steady-state *flux coupling analysis* to gain insight into the metabolic potential of *Synechocystis* sp. PCC6803. The first steps of carbon acquisition and catabolism have been identified as the versatile centre of the coupling network, having a stable core of biological building blocks built around, which can explain the relative plasticity of this organism in terms of growth conditions and habitats.

Photosynthesis was not found to have physical constraints to be in a given PSII/PSI ratio window, in terms of flux capabilities. This ratio has no stoichiometric limitation in order to have other values than the ones found in the steady state flux analysis of **Chapter 3**. Hydrogen production is entangled to NADPH production and ethanol to the central carbon metabolism. From that couplings, potential bottlenecks for hydrogen and ethanol optimised strains production were identified.

Mutant generations can be hindered by many effects, some cannot be predicted, but a set of others, can. *Flux coupling analysis* tries to uncover part of the latter, giving researchers the idea that mutants may not only affect neighbouring reactions but also some others that we do not realise, because we did not have prior knowledge on them or just because their connectivity was not plausible. The identification of flux couplings can help metabolic engineers perform mutants that have a better probability of real-world success.

This work, together with previous chapter, will be useful for direct biotechnological applications as well as identifying misunderstandings in the model or finding weaknesses in knowledge annotation. These results will be valuable for designing and implementing fine-tuned *Synechocystis* sp. PCC6803 strains and will thereby help toward building of an economically viable and environment-friendly biofuel production platform.

5.5 Methods.

Constraints used for flux coupling analysis studies simulations

We were consistent with the constraints taken in **Chapter 3** when doing the FBA simulations, to match an autotrophic specific growth rate of 0.09 h^{-1}. As some examples, maximum carbon uptake rate was found to be 3.4 mmol g$_{DW}^{-1}$ h^{-1} into the cell, with HCO$_3^-$ and CO$_2$ as carbon sources (Shastri and Morgan, 2005). For the

sake of comparison across the different conditions, uptake rates for the corresponding carbon sources were matched based on normalization per number of carbon atoms. Main constraints across different growth conditions can be seen in Table 3.1.

Flux coupling analysis

We adapted the flux coupling finder procedure developed by Burgard et al. (2004) to analyse the functional associations between the reactions of the genome-scale metabolic network of *Synechocystis* sp. PCC6803. The difference between our implementation and the original algorithm (Burgard et al., 2004) is that we did not create coupled reaction sets; instead, we examined each of the reaction pairs for the type of flux coupling relationship. Computational requirements were on the order of minutes for the complete genome-scale model. The algorithm was implemented in Matlab® (MathWorks®) by using GLPK as linear programming solver (http://www.gnu.org/software/glpk) and is available upon request.

Firstly, the upper and lower limits of all flux ratios need to be determined (*i.e.*, $R_{max} = \max {v_1}/{v_2}$, $R_{min} = \min {v_1}/{v_2}$). Burgard et al. transformed this non-linear optimization problem to a linear one by performing the variable transformation ($\hat{v} = v \cdot t$) (see Burgard et al. (2004) for an extensive proof of equivalency). Thus, problem is reduced to:

Maximize $R_{max} = \hat{v}_1$ (or minimize $R_{min} = \hat{v}_1$)

subject to $\sum_{j=1}^{M} S_{ij}\hat{v}_j = 0$ $\forall\, i \in N$

$\hat{v}_2 = 1$

$\hat{v}_j^{uptake} \leq v_j^{uptake\_max} \cdot t, \quad \forall\, j \in M_{transport}$

$\hat{v}_j \geq 0, \qquad\qquad\qquad \forall\, j \in M$

$t \geq 0$

Here the variables $\hat{v}$ are the metabolic fluxes normalized by v_2. The above linear program can be interpreted in biological terms, as responses of metabolic networks to the perturbation of particular fluxes. Constraint $\hat{v}_2 = 1$ sets a reference flux to a unit value, whereas the optimization criteria are used to probe

flux variability for each tested reaction. Uncoupled fluxes are not affected by flux perturbations, whereas fluxes through coupled reactions decrease or increase in accordance with the encountered type of coupling (Burgard et al., 2004).

Four types of flux coupling relationships were considered (Figure 5.4 bears their graphical interpretation):

(i) *fully coupled*: non-zero flux in one reaction implies non-zero and fixed flux through the other reaction, and vice versa;

(ii) *partially coupled*: non-zero flux in one reaction implies non- zero but variable flux in the other reaction;

(iii) *directionally coupled*: non-zero flux in one reaction implies non-zero flux in the other, but not necessarily the reverse;

(iv) *uncoupled*: presence of flux through one reaction does not bound flux through the other and vice versa (reactions are stoichiometrically independent at steady-state).

The flux coupling results shown here are from the calculations run without including a biomass formation reaction (description of the constraints used can be found in Table 3.1, and coupled reactions can be found in Additional file 5.1). The biomass equation was excluded in order to avoid coupling of a large number of fluxes to the biomass formation reaction. However, all biomass components were allowed to be drained independent of one another. Calculations performed with the biomass formation reaction present retrieved similar coupling patterns within the rest of the reactions (data not shown).

Colonel Saul Tigh: It all traces back to us.
Tory Foster: No! … Go back far enough it's always them!
Colonel Saul Tigh: You point a finger back far enough,
and some germ gets blamed for splitting in two. No. We
share the blame…

*Battlestar Galactica, Episode 4.15, "No Exit", script
coordinated by Ryan Mottesheard, 2009*

6

Growth studies of *Synechocystis* sp. PCC6803

Where PhD applicant performs a wide study on different variables that, when properly tuned, should build a perfect environment for the growth of this organism.

6.1 Introduction.

We have seen in previous chapters how we can construct an *in silico* model of the metabolic behaviour of a photosynthetic microorganism. However, we were also interested in studying and modelling a system where we could take into account all the inputs that are usually considered when growing cyanobacteria. With efforts from **Chapters 2 and 3**, model responded to environmental changes in carbon source and amount of light and presence or absence of oxygen, nitrate and water. We considered important to leap the distance between this and the environmental conditions usually considered when growing unicellular non-nitrogen fixing *Synechocystis* sp. PCC6803. Consequently, we started by studying and experiencing on the factors that affect cyanobacterial growth at the batch level.

Cyanobacteria are known to survive a wide spectrum of environmental stresses, such as temperature shock (Waterbury, 2006), photo-oxidation, nutrient deficiency, pH changes (Blanco-Rivero et al., 2005), salinity, osmotic stress, and ultraviolet light (Mullineaux, 2008). As it has already been mentioned in **Chapter 1**, according to Vermaas (1996), *Synechocystis* sp. PCC6803 can potentially grow under *photoautotrophy* (CO_2, light, and active photosystem II and photosystem I), *photoheterotrophy* (glucose and active and constant photosystem II or photosystem I), *mixotrophy* (glucose, CO_2, and active photosystem II and photosystem I), *light-activated heterotrophy* (glucose and 5 minutes of light per day), and *anaerobiosis* (glucose or other fixed-carbon as source). Additionally, there is some controversy whether *Synechocystis* sp. PCC6803 can, in fact, grow under strict *dark* heterotrophic conditions, *i.e.* without the 5 minutes shot of light (Anderson and McIntosh, 1991; Bricker et al., 2004).

Many physiological studies have addressed the effects of environmental factors affecting cyanobacteria growth and *Synechocystis* sp. PCC6803 in particular:

  a) Light

This factor plays a crucial role in all photosynthetic organisms by regulating growth, as we saw in **Chapter 3**, altering gene expression, as we will see in **Chapter 7**, and resetting circadian rhythms, among others (Gill et al., 2002; Mullineaux, 2008), but it can also be harmful to the photosynthetic machinery. It has been observed in *Synechocystis* sp. PCC6803 that visible light might play a double role, inducing damage to photosystem II when it is strong and inducing

repair of the photodamaged photosystem II when it is weak (Allakhverdiev and Murata, 2004).

b) Temperature

Cyanobacteria are found in environments with different temperature ranges. Most cyanobacteria are mesophilic and live on environments where temperature may range from freezing to 40°C. They typically have growth optima between 20 and 35°C; however, organisms that can grow up to 75°C have been found (Castenholtz, 1969).

c) Nitrogen

Nitrate is the most common nitrogen source used by cyanobacteria, it is widely utilized for their growth and its limitation induces a well-characterized set of cellular responses such as: visible chlorosis or *yellowing* (Allen and Smith, 1969), degradation of phycobiliproteins (Collier and Grossman, 1992), alteration of the ratio of phycocyanin to allophycocyanin (Yamanaka and Glazer, 1980), degradation of thylakoid membranes, a decrease in chlorophylls, an increase in carotenoid content or carotenoid/chlorophyll ratio, as well as an increase in glycogen content (Stevens et al., 1981). Most of these alterations are related to light harvesting antenna physiology, thus critical in photosynthetic organisms.

d) pH

Under laboratory conditions, cyanobacteria have generally been reported to prefer neutral to slightly alkaline media (Kratz and Myers, 1955). In natural environments, however, cyanobacteria extend their distribution to pH values as low as 4 (Kurian et al., 2006; Steinberg et al., 1998). Kallas and Castenholz (1982) analysed cytoplasmic pH homeostasis in cyanobacteria and observed that the growth rate of *Synechococcus* sp. was inhibited at pH 7.0 and below, and no sustained growth took place at pH 6.0. However, even when these cells were exposed to pH 4.8, they retained a higher intracellular pH, suggesting that there are other factors involved in the acid tolerance mechanisms of these photosynthetic microorganisms to maintain homeostasis within the cell (Summerfield and Sherman, 2008). It has also been shown that acid-tolerant cyanobacteria maintain a neutral cytoplasmic pH, although how they keep a strong transmembrane pH gradient is still unknown (Steinberg et al., 1998). As well as being acid tolerant, cyanobacteria are among the most alkaliphilic

microbes and frequently dominate alkaline environments such as soda lakes and microbial mats (Summerfield and Sherman, 2008).

All the factors mentioned above affect *Synechocystis* sp. PCC6803 metabolism in a variety of manners and to different extents, which are *only* partially understood. Therefore, we have performed a holistic analysis in order to take into account the complete metabolism and the factors that perturb it.

In this chapter, done in close collaboration with researchers headed by Paula Tamagnini in IBMC Porto (Portugal), we have studied different environmental parameters such as temperature, irradiance, nitrate concentration, pH, and an external carbon source to unveil their influence in *Synechocystis* sp. PCC6803 growth. Porto group was the experimental part and used a high throughput system equivalent to batch cultures. Valencia group (mainly myself) performed a statistical analysis of this set of non-homogeneous data to compare these time series and to retrieve hidden governing factors.

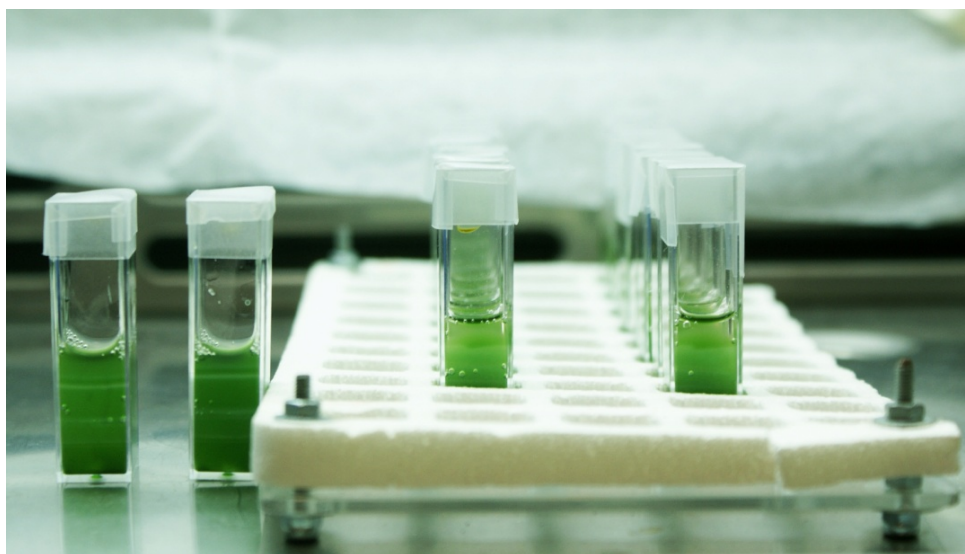## 6.2 Growth conditions study.

Figure 6.1 - Experimental design of growth study. Four clear-sides cuvettes with 2 mL of *Synechocystis* sp. PCC6803 culture on acrylic rack inside growth chamber. Photograph taken from Paula Tamagnini group.

We performed Optical Density (OD) measurements of cyanobacteria growing in cuvettes (Figure 6.1), as a proxy of a batch-like growth study. Three growth conditions were considered: photoautotrophic, mixotrophic and heterotrophic. Table 1 describes all the possible combinations of which we did relevant subsets that characterized each condition.

Table 6.1 - Experimental matrix with conditions/parameters to be tested on *Synechocystis* sp. PCC6803 growth. Units are light in $\mu E\ m^{-2}\ s^{-1}$, temperature in °C, glucose in mM and nitrate in $g\ L^{-1}$.

Condition	States				
	light	temperature	glucose	nitrate	pH
Heterotrophic	0	25, 30, 33	2, 5, 7	0.1, 0.75, 1.5	7.5, 8, 8.5, 9, 10, 10.5, 11
	light	temperature	glucose	nitrate	pH
Autotrophic and Mixotrophic	20, 40, 60	25, 27, 30, 33, 35	0, 2, 5, 7	0.1, 0.75, 1.5	7.5, 8, 8.5, 9.0, 10, 10.5, 11

- *Autotrophic growth*

Out of all the possible combinations described in Table 1, a subset of 72 autotrophic conditions was selected. 2221 Optical Density measurements were performed for the 254 samples. At least three independent experiments (biological replicates) were performed. Each experiment run in average during 8 days and the measurements were performed daily. The data were statistically analysed with a four-way ANOVA. In this statistical study a model with all the effects and interactions up to the third order was tested. Results showed that initial pH, temperature, and level of irradiance are the main factors affecting maximum autotrophic growth, with highly significant p-values ($p < 10^{-16}$). Nitrate levels initially present in the culture showed p-values close to 0.4, which clearly discards any influence in growth results. Among the possible two- and three-way interaction terms in the model, there are two that showed a clear statistical influence in growth (p-value $< 10^{-16}$): the interaction between initial pH and illumination level and between temperature and pH. In other words, illumination and temperature modify the way in which pH affects growth.
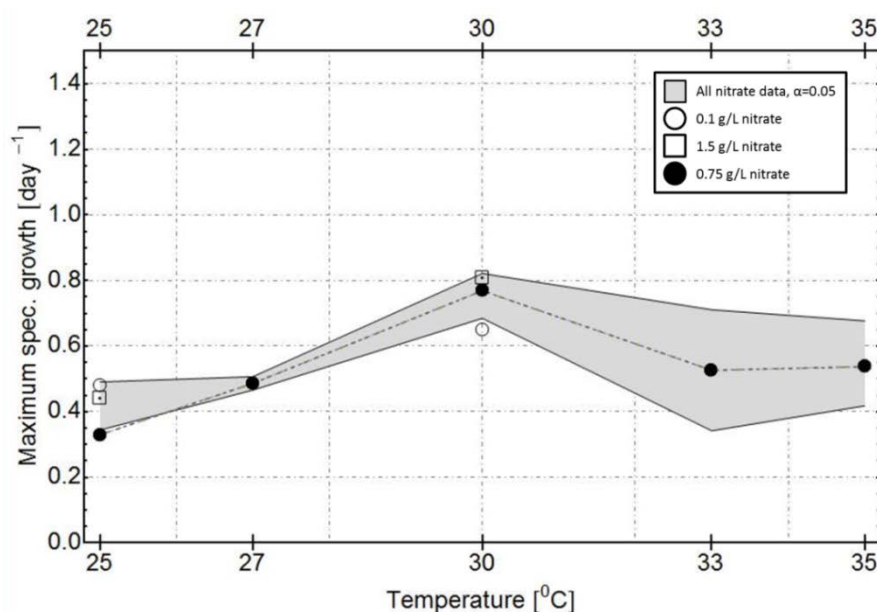
Figure 6.2 - Temperature profile of maximum specific autotrophic growth with different initial nitrate concentrations. The grey shaded area corresponds, at 95% confidence level, to the region of maximum growth if all nitrate levels (0.1, 0.75, and 1.5 g L^{-1}) are averaged. Data markers show the result for specific levels of nitrate: hollow circles correspond to 0.1 g L^{-1}, hollow squares correspond to 1.5 g L^{-1} and black circles (and dashed line) to 0.75 g L^{-1}. In all tests irradiance was 20 μE m^{-2} s^{-1} and HEPES buffer was used for cultivation with an initial pH of 8.5.

To reveal the pairwise differences between the experimental sets obtained at different levels of the main parameters, a Student *t*-distribution analysis was performed to represent the confidence regions of maximum specific growth at the different conditions (using a confidence level of α=0.05). Figure 6.2 depicts experimental results representing the sample mean of maximum specific growth, μ_{max}, at different levels of nitrate, initial pH of 8.5 (HEPES buffer) and 20 μE m^{-2} s^{-1} irradiance. The grey shaded area represents the range of the confidence region (α=0.05) of the maximum specific growth at the different temperatures after data are averaged over the different levels of nitrate. The dashed curve and isolated data markers represent maximum specific growth results at particular values of initial nitrate (0.1, 0.75 and 1.5 g L^{-1}). As it can be observed, growth reaches a maximum value around 30°C and decreases at higher temperatures whilst results at particular levels of nitrate fall within the confidence region of the nitrate-averaged curve. Again, this shows that, from the statistical point of view, nitrate concentrations did not have a noticeable influence on *Synechocystis* sp. PCC6803 growth (Figure 6.4), not even with values as low as 0.1 g L^{-1} (1.2 mM) which is reported in the literature as N-limited conditions (Schmitz et al., 2002). However,

before grand conclusions on this are drawn, one should bear in mind, as we do in following **Discussion** section, that the endogenous nitrogen reserves build up during the maintenance growth in BG11 medium might fulfil the organism needs.
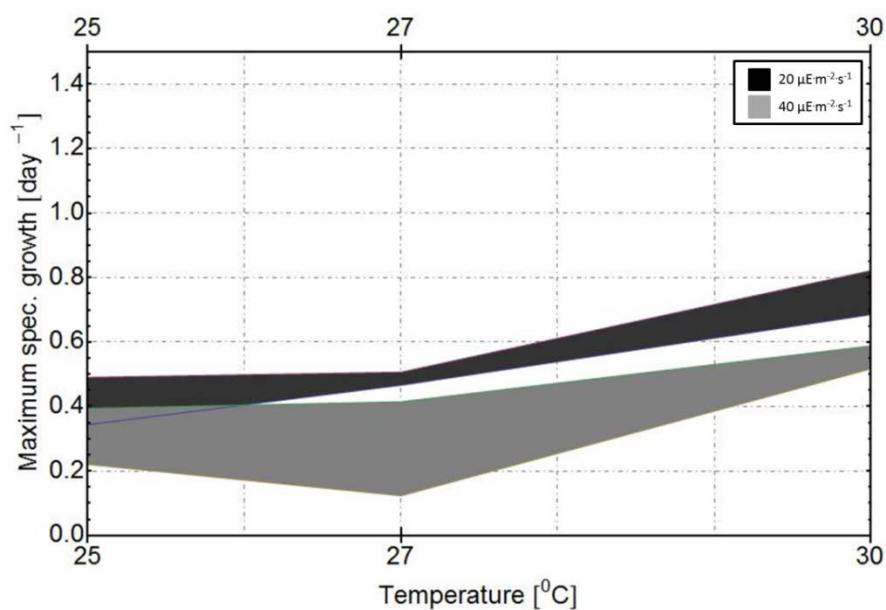
Figure 6.3 - Temperature profiles of the 95% level confidence regions of maximum specific autotrophic growth at different irradiance levels. The upper black filled area corresponds to experiments with a level of irradiance of 20 µE m^{-2} s^{-1}, and the lower grey area to experiments with a level of 40 µE m^{-2} s^{-1}. Results of experiments at different nitrate concentrations were averaged and HEPES buffer was used for cultivation with an initial pH of 8.5.

Figure 6.4 - Growth of *Synechocystis* sp. PCC6803 under different nitrate conditions. Cells were grown autotrophically in BG11 (A), heterotrophically (B) and mixotrophically (C) in BG11 supplemented with 5 mM glucose at 30°C, pH 8.5 and 20 $\mu E\ m^{-2}\ s^{-1}$.
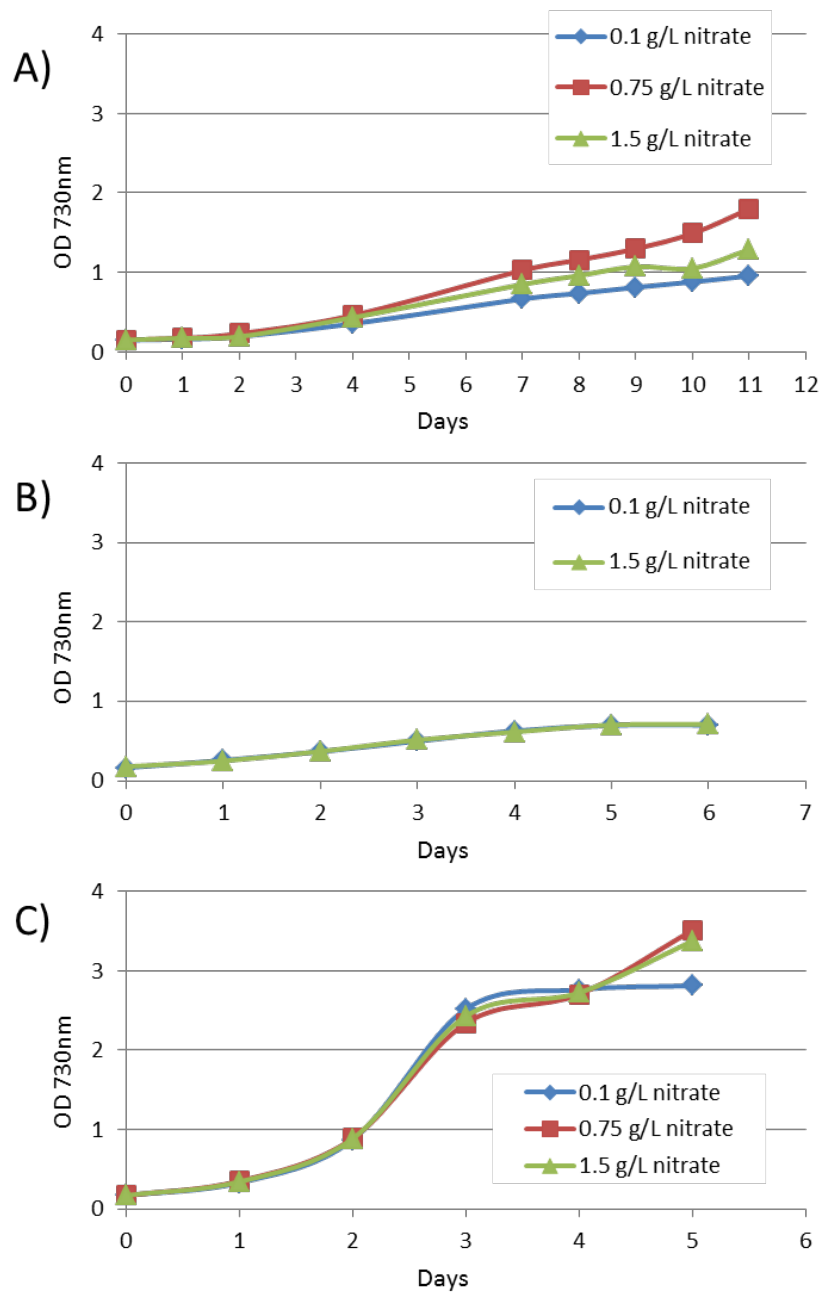
Figure 6.3 compares autotrophic maximum specific growth at the two tested illumination levels (20 and 40 μE m^{-2} s^{-1}) with an initial pH of 8.5. Results obtained at the lower level of illumination, 20 μE m^{-2} s^{-1}, are significantly above those measured at 40 μE m^{-2} s^{-1}, indicating that, in the case of such small cuvette volumes, light saturating conditions were probably reached already at 20 μE m^{-2} s^{-1}, with bleaching becoming noticeable at 40 μE m^{-2} s^{-1}. This effect is not so clear at lower initial pH values, when maximum growth is also lower. Additional growth profiles are depicted in Figure 6.5.
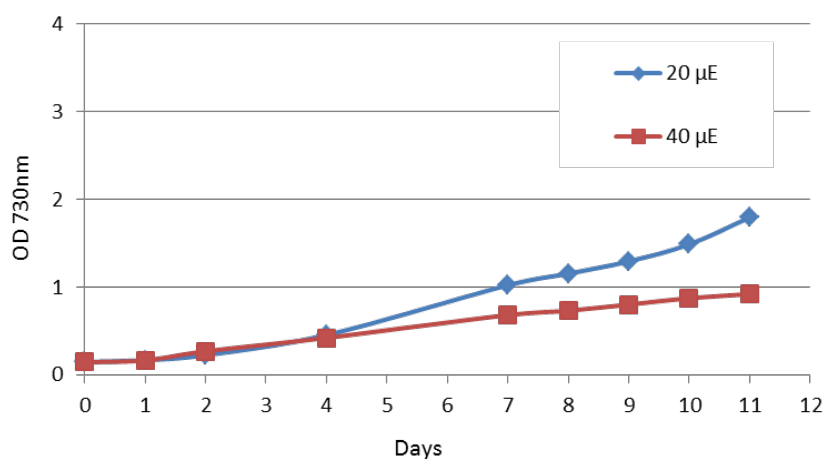

Figure 6.5 - Growth of *Synechocystis* sp. PCC6803 under different light intensities (20 and 40 μE m^{-2} s^{-1}). Cells were grown autotrophically in BG11 at pH 8.5 and 30°C.

From these observations, it was rather apparent that initial pH is strongly affecting culture growth. pH values rose considerably during cultivation time in the buffered conditions with lower pHs, as well as in the "no buffer" conditions, stabilizing around pH 10 (data not shown). As expected, the increase in pH was faster in the absence of any buffer, but occurred nonetheless in HEPES buffered medium reaching similar equilibrium values. This observation brought about the idea of using different buffers to sustain higher values of pH. Representative results of these measurements are shown in Figure 6.6, where the maximum growth results, $\langle \mu_{max} \rangle$, are compared for different types of buffers and initial pH conditions. A very significant increase in maximum growth was obtained at initial pHs of 9 and 10 in CHES buffered medium and 33°C. Results obtained with CAPS buffered medium and without buffer, also exposed in Figure 6.6, clearly confirm this trend in growth and increase in the optimum growth temperatures. Values of

growth reach up to about 1.8 day^{-1} in optimum conditions (CAPS buffer, pH 11 and 33°C) compared to a typical value of 0.7 day^{-1} in CHES buffered conditions taken as reference (Figure 6.7).
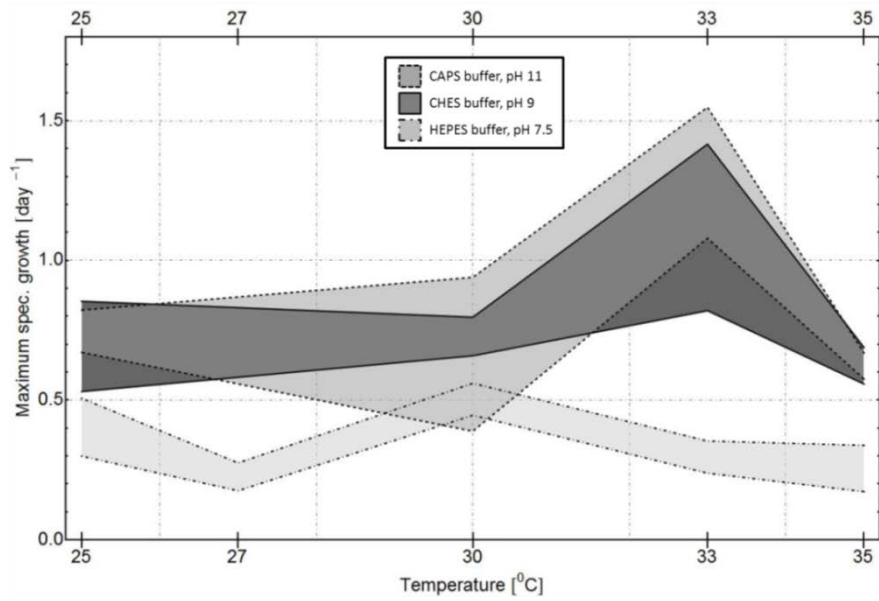

Figure 6.6 - Temperature profiles of the 95% level confidence regions of maximum specific autotrophic growth at different levels of initial pH. Cultures grown with different buffers: CAPS (initial pH 11) upper dashed limited grey area, CHES (pH 9) dark grey area, and HEPES (pH 7.5) lower light grey area. In all tests irradiance was 20 µE m^{-2} s^{-1}.

Figure 6.7 - Growth of *Synechocystis* sp. PCC6803 under different temperatures. Cells were grown autotrophically in BG11 at pH 11 (A) and mixotrophically in BG11 supplemented with 5 mM glucose (B) at pH 10.5 and 20 µE m^{-2} s^{-1}.

To complete the description of pH related behaviour, in Figure 6.8 the relative maximum growth vs. relative cell density evolution of three representative samples has been compared. The curve with filled circles corresponds to high growth conditions (CHES buffer, initial pH 9), whereas the curves with hollow squares and circles represent HEPES growth conditions at pH 8.5 and 7.5, respectively. With CHES buffer, maximum growth is reached already in the 0 - 1st day of cultivation and decreases rapidly afterwards. At a relative cell density of about 70% (with respect to cell density at the end of cultivation) a second growth maximum arises in what appears to be a diauxic type of growth profile. This type of profile is also found in CAPS buffered conditions, where even episodes of cell death (OD decrease) show up (data not shown). This phenomenon can also be

observed in some mixotrophic conditions with HEPES or CHES buffer. In contrast, HEPES buffered growth curve displays maximum growth in the 3rd - 4th day of cultivation. There, diauxic growth features seem to appear, but only at the very end of the experiment.
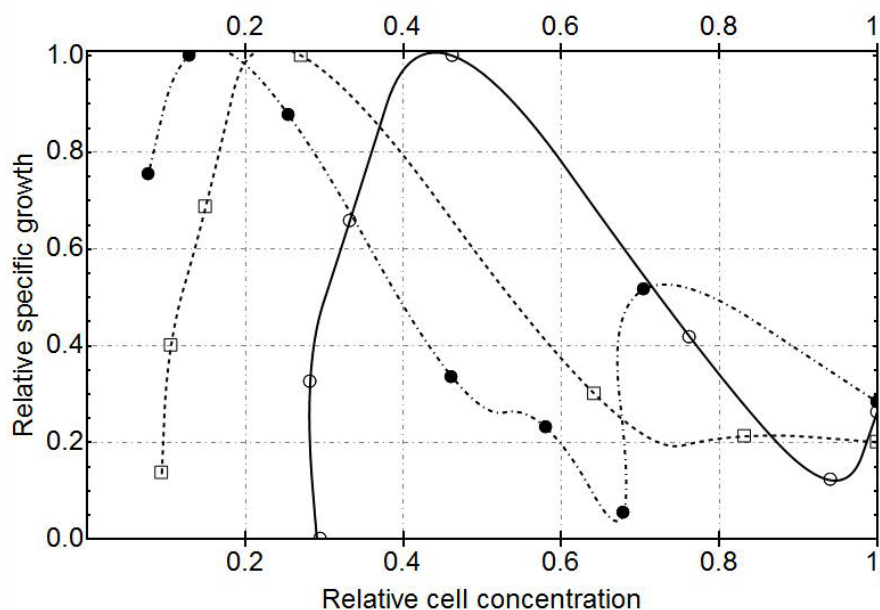
Figure 6.8 - Relative growth vs. relative cell density in autotrophic growth samples in different buffer/initial pHs. The horizontal axis represents cell concentration relative to maximum cell concentration (at the end of batch growth) and the vertical axis growth values relative to maximum growth. In all tests irradiance was 20 µE m^{-2} s^{-1}, temperature 30°C, and initial nitrate concentration 0.75 g L^{-1}. Dark circles correspond to CHES buffer pH 9, hollow squares and hollow circles represent HEPES at pH 8.5 and 7.5, respectively.

Summarizing, it seems clear from these data that cells find their optimum growth conditions at high pH values (between 9 and 11) and at temperatures close to 33°C (Figure 6.9). In media with higher starting pHs, with CHES or CAPS buffer, optimum growth conditions are reached in an earlier stage of cultivation, followed by decay in growth and then another growth burst at higher cell density values (Figure 6.8). This diauxic growth dynamic, points to some metabolic shift which, as will be explained in following section, could be connected to the carbon dioxide and carbonate assimilation mechanisms of the cell.
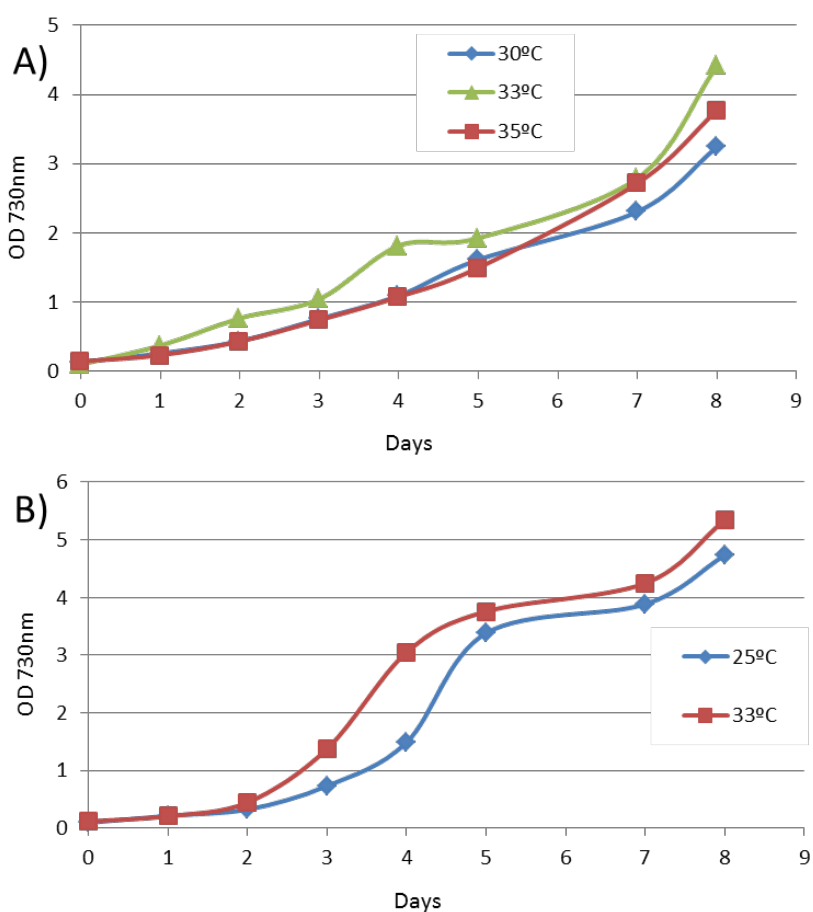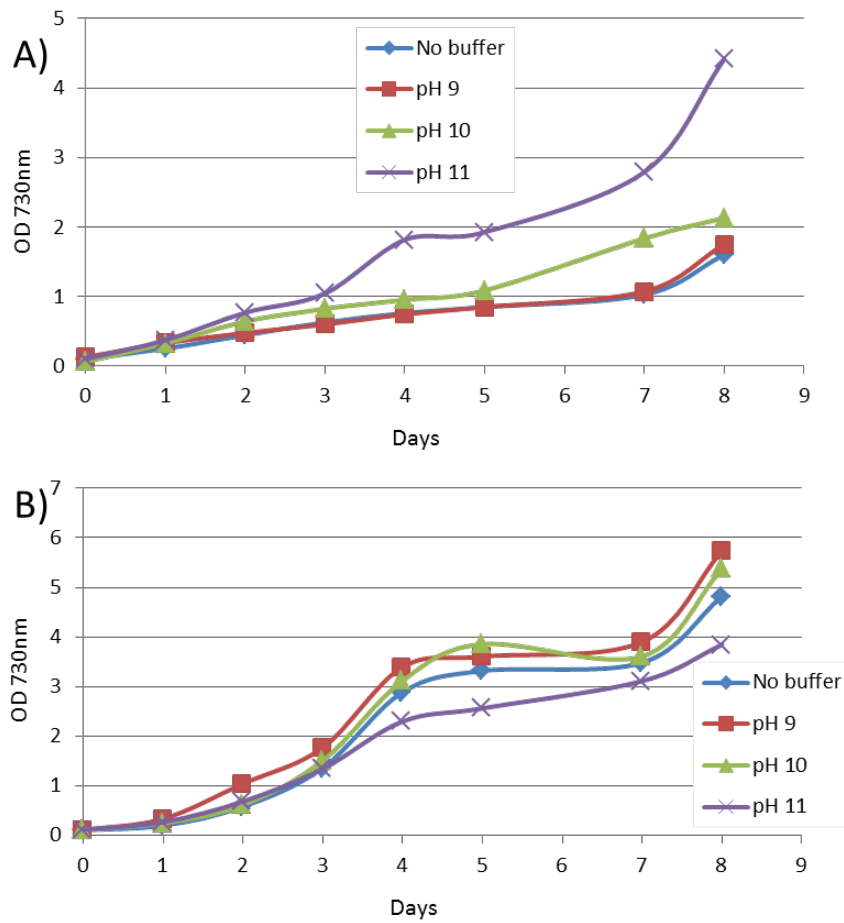
Figure 6.9 - Growth of *Synechocystis* sp. PCC6803 under different pHs. Cells were grown autotrophically in BG11 at 33°C (A) and mixotrophically in BG11 supplemented with 5 mM glucose at 33°C (B) and 20 µE m^{-2} s^{-1}.

Furthermore, results from a scale-up showed no significant differences between growth in the cuvettes and 100 mL cultures in Erlenmeyer flasks showing that our cuvettes are reliable system to evaluate *Synechocystis* sp. PCC6803 growth (see Figures 6.10a and 6.11a).
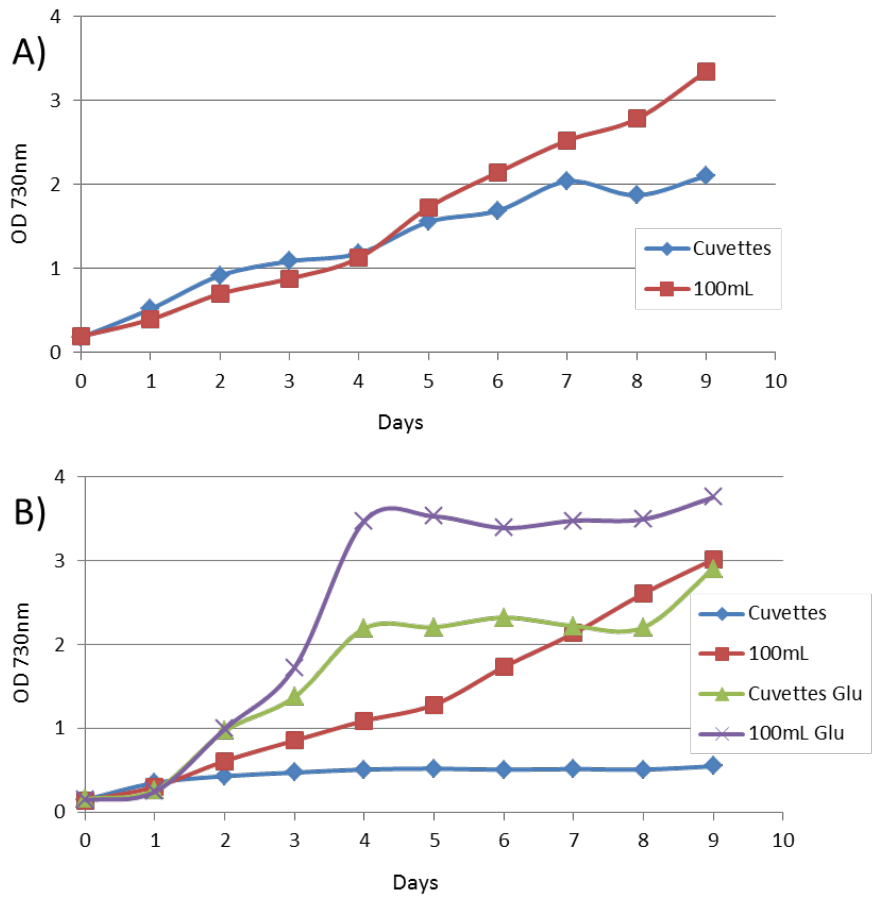
Figure 6.10 - Growth of *Synechocystis* sp. PCC6803 in different volumes. Cells were grown autotrophically in BG11 at 40 µE m^{-2} s^{-1} (A), and both autotrophically and mixotrophically (B) in BG11 supplemented with 5 mM glucose at 20 µE m^{-2} s^{-1} with orbital shaking (100 rpm) at 30°C.
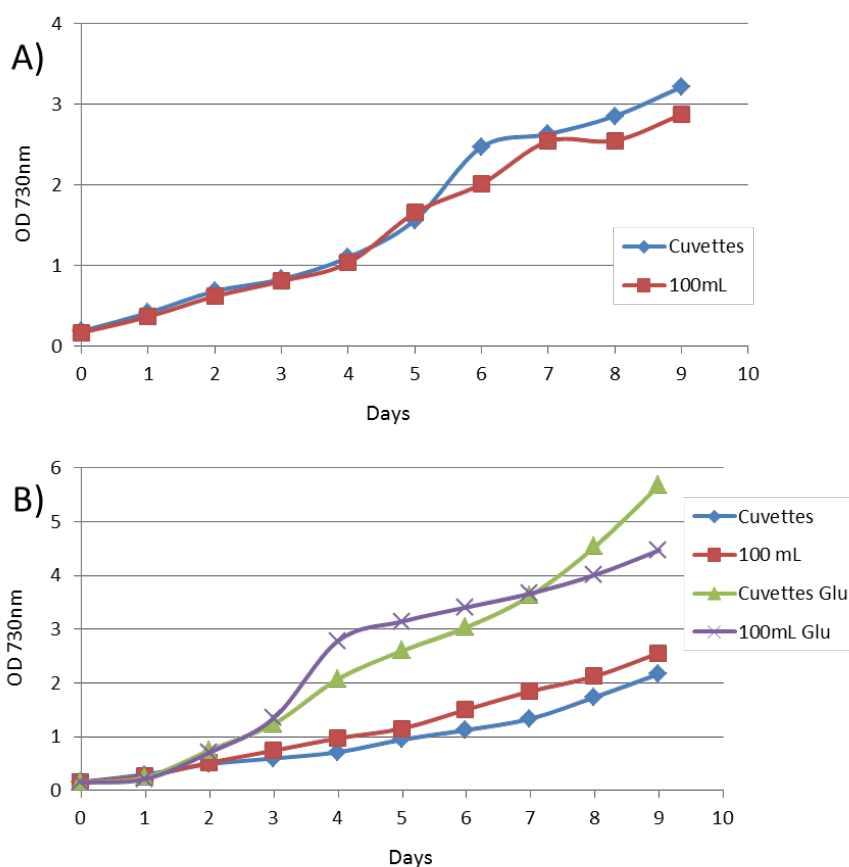
Figure 6.11 - Growth of *Synechocystis* sp. PCC6803 in different volumes. Cells were grown autotrophically in BG11 at 40 µE m^{-2} s^{-1} (A), and both autotrophically and mixotrophically (B) in BG11 supplemented with 5 mM glucose at 20 µE m^{-2} s^{-1} under static conditions at 30°C.

- *Mixotrophic growth*

For mixotrophic samples, irradiation was held constant at 20 µE m^{-2} s^{-1}. A subset of 86 conditions was selected giving a total of 258 samples and 2154 OD measurements. As before, four-way ANOVA results showed that temperature, initial pH, and glucose concentration are the main factors affecting maximum mixotrophic growth. Of all possible interactions, the one with a clearer significant statistical influence (p-value < 10^{-3}) on growth is the interaction between the glucose concentration and temperature. As in autotrophic growth, nitrate

concentration initially present in the culture can be discarded as having an effect on growth (with a non-significant p-value = 0.32).
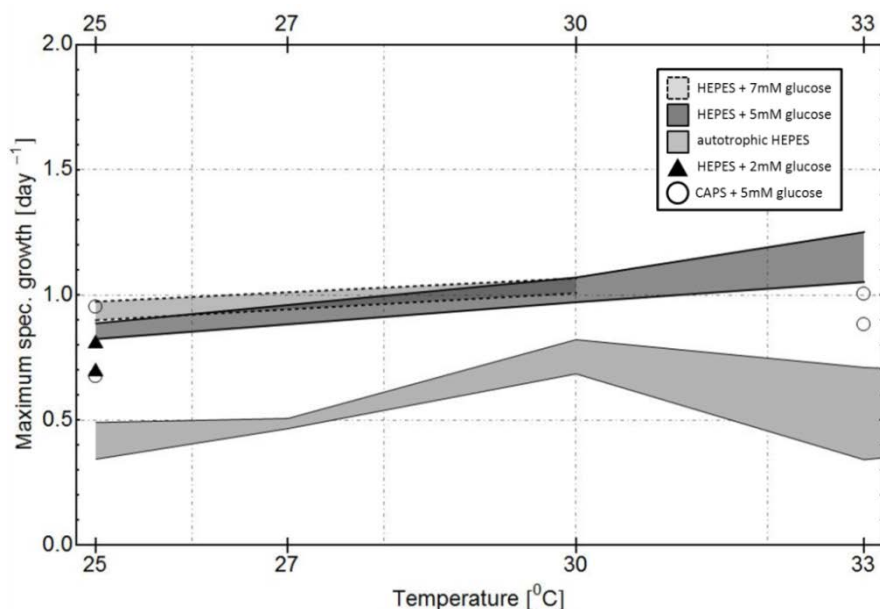
Figure 6.12 - Maximum specific growth vs. temperature profile of 95% confidence regions in mixotrophic vs. autotrophic growth. Irradiance was 20 μE m^{-2} s^{-1}, glucose concentration 2, 5, or 7 mM, and initial pH 8.5 or 10. Light grey area with dashed lines depicts maximum specific mixotrophic growth with 7 mM glucose, dark grey area corresponds to 5 mM glucose, and filled triangles correspond to 2 mM glucose, all with HEPES buffer pH 8.5. Hollow circles correspond to mixotrophic growth with 5 mM glucose and CAPS buffer pH 10, the lower light grey area shows autotrophic growth with HEPES buffer pH 8.5 (used as reference).

Results regarding mixotrophic maximum growth are shown in Figure 6.12, where maximum specific growth results at different temperature levels for the three different glucose concentrations (2, 5 and 7 mM) are presented. Autotrophic maximum growth values, with the same initial pH and buffers, are shown for comparison. Significantly higher growth values are always observed in mixotrophic growth. At 25°C, maximum specific growth is 2 - 2.5 times larger in mixotrophic than in autotrophic conditions and there is a steady increase with higher glucose concentration. The results at 30°C depict a smaller relative difference between growths in mixotrophic vs. autotrophic conditions (1.5 fold) and also between records with different glucose concentrations.
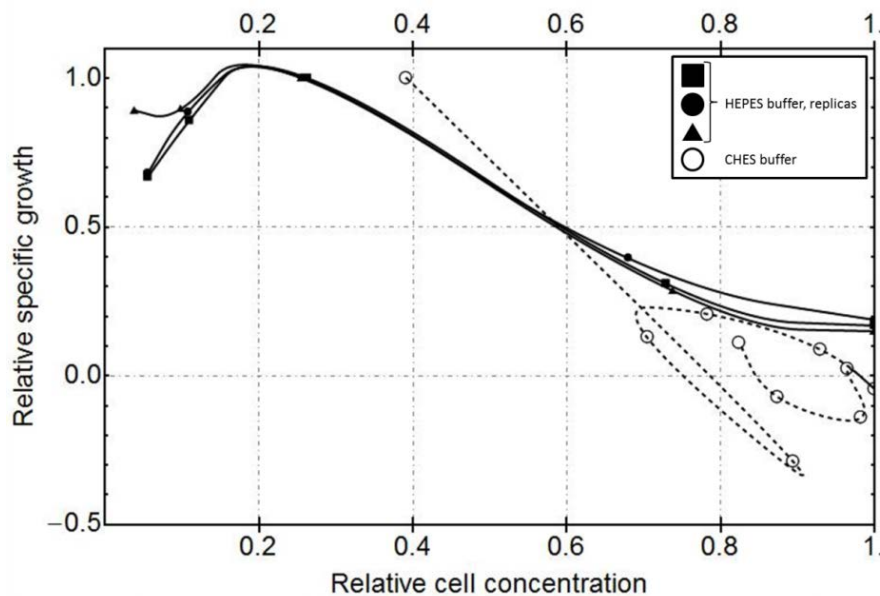
Figure 6.13 - Relative growth vs. relative cell concentrations curves of mixotrophic growth (5mM glucose) at 30°C and initial pH 8.5 with two different buffers. The three replicates of the same sample with HEPES buffer are represented by filled symbols: squares, triangles, and circles. The dashed curve with hollow circles corresponds to one replica with CHES buffer.

The use of CHES buffer, with higher initial pH, offered what seemed to be at first sight an unexpected result. It was observed that, irrespective of pH, differences between growths of different replicas with CHES buffer became systematically quite large, *i.e.* the confidence intervals of maximum specific growth broadened substantially. Another noticeable effect was the shortening of the lag phase, whereby the number of days to reach maximum growth shifted from 3 - 4 days with HEPES buffer to 0 - 1 day with CHES buffer. This behaviour is exemplified in Figure 6.13, in which curves displaying relative growth vs. relative cell concentrations in mixotrophic growth under 30°C and initial pH of 8.5, are presented for three replicas of one HEPES buffer culture (full symbols) and one replica of a CHES buffered culture (hollow circles). The curves corresponding to HEPES cultivation show the same general profile, which is completely different from the CHES curve. In the latter, phases of cell death (relative cell concentration drops from one day to the other) and diauxic growth can be observed. The growth profile of the rest of CHES buffered cultures consistently displays this type of irregularities with large quantitative differences between replicas grown under the same conditions. These observations suggest that, in conditions in which light-based and glucose-based cell growth mechanisms have a similar potential (as in

CHES conditions near 30°C) the resulting growth may become unstable. On the contrary, if one mechanism is predominant, culture growth becomes regular.

Additionally, a scale-up experiment was performed revealing that growth in these mixotrophic conditions is consistently higher than in the aforementioned autotrophic conditions (see Figures 6.10b and 6.11b).

- *Heterotrophic growth*

Finally, to study heterotrophic growth, 12 different conditions were selected totalizing 36 samples and 324 OD measurements. Glucose level was held constant at 5 mM and temperature was kept at 30°C in all tests. ANOVA results show that neither nitrate concentration nor initial pH substantially affect maximum specific growth within the confidence level chosen.

The overall averaged maximum heterotrophic growth value of 0.42 day^{-1} can be compared to the value of around 0.8 day^{-1} obtained for autotrophic growth and around 1.1 day^{-1} in mixotrophic growth under equivalent conditions. Furthermore, the analysis of growth profiles indicates that, systematically, after reaching the stationary phase (about the 5th - 6th day), cell cultures undergo a noticeable decrease in population with cell death rates that reach up to 50% of maximum specific growth rate (data not shown).

6.3 Discussion.

The wide and multifactorial set of experiments shown in this study was developed to highlight the importance and effect of the main variables relevant for *Synechocystis* sp. PCC6803 metabolism and to establish the optimum growth conditions for the use of this organism as a model for future molecular biology developments.

Following the results from previous authors (Anderson and McIntosh, 1991; Bricker et al., 2004; Vermaas, 1996), and in the line of our results, *Synechocystis* sp. PCC6803 can grow *photoautotrophically* (a condition requiring both photosystems to be functional), *mixotrophically*, *photoheterotrophically* (under constant light), and *light-activated heterotrophically* (using glucose as carbon source and with a daily pulse of light). We did not achieve a pure *dark* heterotrophic growth of significance (data not shown).

a) Light

Reports from the literature are arguably consensual regarding optimum light intensity for *Synechocystis* sp. PCC6803 or cyanobacteria in general. Depending on the culture volume, flask geometry and inoculum, *Synechocystis* sp. PCC6803 is able to withstand higher light intensities. For instance, Anderson and McIntosh (1991) exposed 75 mL cultures to 40 μE m^{-2} s^{-1}, whereas Bricker et al. (2004) grew 150 mL cultures at the same light intensity.

Some studies point out that *Synechocystis* sp. PCC6803 cells can grow under a wide range of light intensities, up to approximately 1000 μE m^{-2} s^{-1} or even more. Hihara et al. (2001) exposed 50 mL cultures of this cyanobacterium to a shift of light intensity from what they considered low light: 20 μE m^{-2} s^{-1}, to high light: 300 μE m^{-2} s^{-1}, and concluded that the higher intensity was sufficient to induce large changes in the gene expression profile. In their environmental study, Allakhverdiev and Murata (2004) grew *Synechocystis* sp. PCC6803 cells in 120 mL glass tubes at 70 μE m^{-2} s^{-1}. Subsequently, to assess the photodamage-repair cycle of photosystem II, the cells were exposed to various light intensities in 4.5 mL cuvettes for short periods, never exceeding 90 minutes. These light intensities were: 250, 500, 750, 1000, 1500 and 2000 μE m^{-2} s^{-1} and the authors considered values up to 1000 μE m^{-2} s^{-1} as weak light and from then onwards strong light. They concluded that the rate of photodamage was proportional to light intensity and that the rate of repair also depended on light intensity: it was high under weak light and reached a maximum at 300 μE m^{-2} s^{-1}. However, according to Vermaas (1996), *Synechocystis* sp. PCC6803 cells grow best at 40 - 70 μE m^{-2} s^{-1} and are photoinactivated easily at higher light intensity.

Our results show that for small volumes (2 mL) and our cuvette's geometry, 20 μE m^{-2} s^{-1} is closer to the optimum light intensity than the second irradiance chosen (40 μE m^{-2} s^{-1}), an indication that saturated light condition regime may have been reached at this higher intensity. It seems clear that the debate regarding optimum light intensity for *Synechocystis* sp. PCC6803 is still open and researchers need to realize how important it is to clarify, describe and consider all the parameters involved such as: culture volume, flask geometry, cell density and time of exposure, to name just a few.

b) Temperature

Previous observations have shown that *Synechocystis* sp. PCC6803 cells seem to suffer from low- and high-temperature stress below 25°C and above 40°C respectively, an effect associated to photosystem II sensitivity, with a complete arrest of growth at 15°C and 45°C (Berry and Bjorkman, 1980). In another study, regarding growth temperature acclimation and high-temperature effects on photosystem II of *Synechocystis* sp. PCC6803, Inoue et al. (2001) observed that maximum growth was obtained at 30°C, but that growth rates were similar in temperatures ranging from 25 to 40°C, which agrees with the rate of photosynthesis at these temperatures.

While growth limitation at high temperatures seems to be due to limitation of photosystem II activity, acclimation to low temperatures appears to take place in parallel with an increase in the desaturation level of thylakoid membranes fatty acids (Los et al., 1993). It has been shown that a rise in the desaturation level of fatty acids in the thylakoid galactolipids increases the resistance of the cyanobacterial cells to low temperatures (Wada et al., 1994). However, Sakamoto and Bryant (1998) saw that *Synechocystis* sp. PCC6803 cells grown at 15°C aggregate and undergo chlorosis, even though this may be due to nitrogen limitation.

Our results indicate that *Synechocystis* sp. PCC6803 growth holds critical temperature dependence: in autotrophic growth with HEPES buffered cultures (pH between 7.5 and 9), maximum growth is observed around 30°C; while with CHES buffer (pH between 9 and 11) it occurs at 33°C. Also temperature seems to be an asymmetric factor: growth rate of a given HEPES-buffered culture, with an optimum temperature of 30°C, dropped 50% when grown at 25°C, but only decreased 25% when grown under 35°C.

c) Carbon source

The presence of an exogenous carbon source (*i.e.* glucose) has proven to be a very important factor in *Synechocystis* sp. PCC6803 growth. In our experiments, the exact same culture was grown in media where the only difference was the amount of glucose concentration: autotrophic (without glucose) and mixotrophic (2, 5 and 7 mM glucose). Mixotrophic growth was approximately 2 - 2.5 times larger than autotrophic growth at 30°C at a concentration of 5mM. A further increase of concentration between 5 mM and 7 mM, does show a further effect, though non-significant, on growth (Figure 6.12). Yoo et al. (2007) obtained similar

significant results when growing *Synechocystis* sp. PCC6803 cells with and without 5 mM glucose. Experiments by other authors show the same type of enhanced growth when glucose is added (Williams, 1988).

d) Nitrogen

In our set of experiments the influence of nitrate on specific growth was below the statistical significance threshold. This behaviour may be explained by the fact that during maintenance growth in BG11 medium (1.5 g L^{-1} of $NaNO_3$) the organism may have acquired sufficient nitrogen reserves to survive the experiment's time-span (of about one week), in which low nitrate concentrations were maintained. It is known that growth under nitrogen-limiting conditions favours glycogen production in *Synechocystis* sp. PCC6803. Yoo et al. (2007) observed, using transmission electron microscopy, that in these conditions glycogen particles accumulate in large amounts and fill the cytosol of the cells.

e) pH

Cyanobacteria are among the most alkalotolerant or alkalophilic microbes and preferably grow at pHs ranging from neutral to 11 (Langworthy, 1978; Lopez-Archilla et al., 2004). Additionally, there is a well-documented connection between the action and effects of the photosynthesis machinery of the cells and their ability to increase the surrounding pH (Becking et al., 1960). In normal laboratory conditions, it is known that *Synechocystis* sp. PCC6803 cells increase their surrounding pH by 1 - 2 units over several days. In fact, when placed in acid stress situation at a tolerable pH (4.4 or above), this organism's cells are able to increase their surrounding pH to 6 and above within a few minutes. *Synechocystis* sp. PCC6803 exhibits a predictable physiological response to acid stress and is incapable of growing in media with a pH lower than 4.4. However, this capability is cell density dependent, as high cell concentration cultures have been able to withstand pH shock as low as 3.5 (Huang et al., 2002a).

The ability of cyanobacteria to grow at alkaline pH implies the presence of mechanisms for the maintenance of an intracellular pH more acidic than that of the environment (Miller et al., 1984). Indeed, an increase in the external pH of 2 units results only in an increase of 0.2 units in both the cytosol and the thylakoid lumen. Such changes alter the $[CO_2]/[HCO_3^-]$ ratio within the cell, and regulation of this ratio is essential for maintaining the carboxylase activity of RuBisCO, vital for the cell as the carbon fixing enzyme (Summerfield and Sherman, 2008). Additionally, there is a common pH-buffering mechanism in freshwater systems

that uses these two carbon molecules: the carbon dioxide/bicarbonate/carbonate equilibrium. Halophilic or halotolerant cyanobacteria are thought to possess a membrane-associated Na^+/H^+ antiport system that could participate in pH homeostasis. This could be the case of *Synechocystis* sp. PCC6803, since its genome contains information for such potential pumps (Maestri and Joset, 2000).

In the present work *Synechocystis* sp. PCC6803 cultures (initial pH 8.5) were grown in continuous light under pH-unstressed conditions, *i.e.* without buffer. The pH increase observed (to up to 10.5) in the buffered (after 10 days) and non-buffered (after 5 days) conditions may be attributable to carbon fixation through photosynthesis, which can shift the aforementioned pH-buffering system, the carbon dioxide/bicarbonate/carbonate equilibrium. This light-dependent increase of external pH can be observed in nature (Pierson et al., 1999) and is particularly remarkable in hypereutrophic systems, as a consequence of a very high primary production (Lopez-Archilla et al., 2004). Besides, photosynthesis is favoured under more alkaline conditions, since alkaline systems act as a trap for atmospheric carbon dioxide (Imhoff et al., 1979) which, in turn, is more available for primary production. Moreover, at high pHs nutrients such as phosphorous become more soluble (Talling and Talling, 1965). Evidence that *Synechocystis* sp. PCC6803 cells must photosynthesize to increase external pH is provided by the fact that cells placed in the dark do not grow and do not increase the external pH, even with a carbon source (such as glucose) added to the media as it was observed in our experiments. In addition, if an autotrophic culture is switched to dark, pH tends to decrease to the initial level. This effect is not observed in heterotrophic conditions. Overall, all these evidences –carbonates equilibrium, carboxylase activity of RuBisCO, favoured photosynthesis, phosphorous solubility, pH changes dependent on light in mixo- and autotrophy– point to the hypothesis that *Synechocystis* sp. PCC6803, when contained in a closed environment such as a batch culture, are favoured to grow in an alkaline medium and tend to transform their growth media to reach alkaline values.

We have also considered an alternative explanation for the external pH increase: ammonia may be excreted into the medium by pH-stressed *Synechocystis* sp. PCC6803 cells (Huang et al., 2002a). However, in the present work, no ammonia was detected in the high pH medium (data not shown), indicating that if there was any trace of ammonia, it would be consumed by the cells within a short time (Sakamoto et al., 1999), undoing its effect over pH changes.

6.4 Conclusions.

A multifactorial set of experiments, with over 3000 individual measurements, was conducted to highlight the effect of the main variables relevant for *Synechocystis* sp. PCC6803 growth and to establish the optimum conditions for its use as a model chassis for biotechnological applications. Irradiance, glucose presence, and temperature were found to be the major governing factors. In our experiments the influence of nitrate was too small to be detected within our statistical significance threshold and growth was meaningfully larger at a light intensity of 20 μE m^{-2} s^{-1} than for 40 μE m^{-2} s^{-1}, where, for our cuvette's volume and geometry, damaging effects appeared.

In autotrophic growth, there is a significant effect and interaction of pH and temperature. In HEPES buffered cultures (pH between 7.5 and 9), maximum growth is observed around 30°C, while CHES or CAPS buffered (pH between 9 and 11) and non-buffered cultures show a very marked peak in growth at 33°C. Maximum growth of 1.8 day^{-1} was obtained in the latter conditions, being 2.5 - 3 times larger than growth in lower pH media.

In mixotrophic conditions maximum growth is around 1 day^{-1} and growth is much more insensitive to pH and temperature. Opposite to autotrophy, mixotrophic growth is smaller in high pH media compared to lower pH media.

As a future work, it is of interest to consider bridging our metabolic models that describe the metabolite landscape in detail to these factors that affect cell growth at the batch level. Amount of light and nitrate are already included, but factors like pH and temperature should be considered and their molecular effects taken into consideration.

6.5 Methods.

Organisms and standard growth conditions

The unicellular non-N_2-fixing cyanobacterium *Synechocystis* sp. PCC6803 (obtained from the Pasteur Culture Collection, Paris, France) was maintained in BG11 medium (Stanier et al., 1971) at 25°C, and 16 h light (20 $\mu E^. m^{-2}. s^{-1}$) / 8 h dark cycles.

Growth experiments

Inoculums of 200 µL from the maintenance culture grown under standard conditions (transferred 24h before to fresh BG11 medium) with an $OD_{730} \approx 1.0$, were added to sterile 4.5 mL cuvettes containing 1800 µL of medium with given concentrations of glucose, nitrate and pH values. Each experiment was performed in triplicate and under aseptic conditions. The cuvettes were closed with sterilized Parafilm^{TM}, and placed in acrylic racks specially designed for this experiment (Figure 6.1). This constitutes a high throughput system, equivalent to batch cultures with the Parafilm^{TM} cover allowing gas exchanges. The racks were placed in a chamber with constant temperature and irradiance (20 or 40 μE m^{-2} s^{-1} with Osram L18W/765 cool white daylight bulbs). For pH control, several buffers (HEPES, CHES and CAPS) were used, according to their pKa, with a concentration of 10mM (or 10, 20, 30, 50 and 100 mM for HEPES, see Figure 6.14). Growth in normal BG11 media without controlled pH was also tested. Optical density at 730 nm was recorded daily.

Table 1 shows the initial experimental matrix with all the conditions to be tested. Not all the combinations of this matrix were tested: as the work flow moved along and new results were obtained, some combinations were abandoned and more relevant ones were introduced, notably other temperatures (33 and 35°C), and pHs (9, 10, 10.5 and 11).
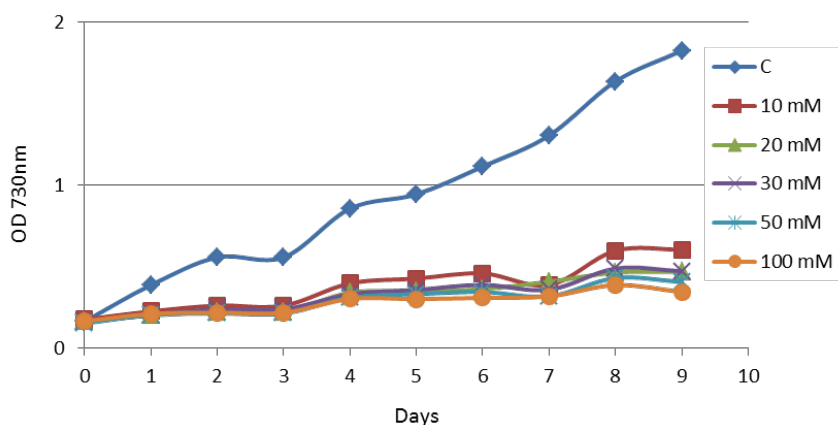
Figure 6.14 - Growth of *Synechocystis* sp. PCC6803 with different HEPES buffer concentrations (10, 20, 30, 50 and 100 mM), pH 7.5, and C-control without buffer. Cells were grown autotrophically in BG11 at 20 µE m^{-2} s^{-1} and 30°C under static conditions.

Evaluation and statistical analysis of the specific growth and other culture-characteristic parameters

To enable the quantitative analysis of the growth characteristics of the different cultures in the various tested conditions, each culture growth sample measurement is expressed as an array of Optical Density (OD) vs. time logs, $\{t_i - x_i\}(i = 0, \dots i_{max})$, 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 record. For each bin, i, a specific growth parameter, representing the exponent of the (assumed) exponential growth within the given time interval is calculated as:

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

Note that, if the time interval is the same between all sample records, the mean specific growth of the culture during the test period, $\bar{\mu} = \langle \mu_i \rangle$, is simply given by $\bar{\mu} = \frac{\log\left[x_{i_{max}}/x_0\right]}{t_{i_{max}} - t_0}$. It should be observed that this mean exponential growth parameter, $\bar{\mu}$, may depend rather arbitrarily on test duration, particularly if in some particular conditions stationary growth is reached sooner than in

others. Instead, maximum specific growth, defined as the maximum value of specific growths,

$$\mu_{max} = \text{Max}\{\mu_i\}|_{i=0,\ldots i_{max}}$$

allows for better comparison between the growth potential of the different cultures regardless of the duration of tests. In addition we can obtain the time t_{max} at which maximum specific growth is reached and the corresponding OD record, x_{max}.

The usefulness of these parameters can be appreciated if we compare the experimentally found values with theoretical estimates based on the different types of known growth kinetics which are found to describe the growth of microorganisms of many species (Bailey and Ollis, 1986). For instance, one of the simplest forms of batch growth model is given by the logistic curve, in which population evolution is sigmoidal shaped and given by:

$$x_{\log(t;\vec{c})} = \frac{pE^{kt}}{1 - \beta p(1 - E^{kt})}$$

$x_{\log(t;\vec{c})}$ being the culture density at time t and $\vec{c} = (p, \beta, k)$ a culture-characteristic set of parameters. This type of kinetics assumes the presence of some inhibiting factor with concentration $\propto x^2$. Exponential growth starts at $t = 0$ at a maximum specific growth rate of value k. The stationary phase is reached at a cell density of $x = \frac{1}{\beta}$. The relationship between specific growth and cell density is given by the simple expression:

$$\mu_{\log(x)} = k(1 - \beta_x)$$

A more general form of batch growth kinetics, in which a *lag phase* or delay to reach maximum specific growth may appear is given by the kinetics proposed by Konak (1975), in which the relation between cell density and specific growth is given by:

$$\mu_{ext} = c^{a+b} \frac{k\left(\frac{k}{c}\right)^a \left(1 - \frac{x}{c}\right)^b}{x}$$

In Figures 6.15 and 6.16 the resulting $\mu(x)$ curves are shown for a logistic and generalized type of kinetics. Both graph axes are normalized with respect to maximum growth and largest cell density respectively. The shape and position of maximum growth may vary considerably depending of the value of the (a, b, c, k) chosen in the case of generalized kinetics.

140

For each of the samples it is eventually possible to find, by means of nonlinear regression, a *best fit* between the experimental dataset and a given logistic-type or generalized-type growth kinetic function. It will be found that a reasonable fit is not always possible indicating that either large experimental uncertainty is present or that the proposed kinetics are not a good representation of the experimental situation.

From the three replicas of any of the experimental samples with a given level of control factors, $\vec{s}$ –in our case: light, nitrate, glucose and pH– it was possible to calculate confidence intervals for the sample growth parameters, such as $\mu_{max(\vec{s})}$ corrected by means of the Student t-distribution. The usual confidence level of 95% was selected. Furthermore, a univariate Analysis of Variance (ANOVA) to examine the differences between groups of means was used to determine to which extent each of the control factors affects the observed variability in growth.

Figure 6.15 - Theoretical specific growth kinetic curves for a logistic type (full curve) and generalized type of kinetics (dot-dashed curve). The shape and position of maximum growth of generalized kinetics may vary considerably as a function of the selected (a, b, c, k) parameters, following work by Konak (1975). The data shown here correspond to: a = 1.4, b = 4 and c = 2.

Finally, the time variation of specific growth can itself be represented (Figures 6.8 and 6.13) and compared to well-known theoretical growth kinetics, as those presented in Figure 6.15, in order to derive more qualitative conclusions about

changes in growth due to the different cultivation conditions. In Figure 6.15, the horizontal axis represents cell concentration relative to maximum cell concentration (at the end of batch growth) and in the vertical axis growth relative to maximum growth is depicted. In the case of generalized kinetics (dot-dashed curve), the shape and position of maximum growth may vary considerably as a function of the selected (a, b, c, k) parameters, following work by Konak (1975). As it can be seen from Figure 6.16, similar types of temporal behaviour can undercover very different growth logistics, and initial amount of cells governs the *shape* of the temporal curve, which does not happens on Figure 6.15-kind of graphs. As we were interested in cell growth (how the culture drains substrates and produces biomass), effects like initial amount of initial inoculum (that are extremely hard to standardize) had to be circumvented in this analysis. Those are the reasons behind the decision of using graphs like Figure 6.15 in order to depict Figures 6.8 and 6.13.
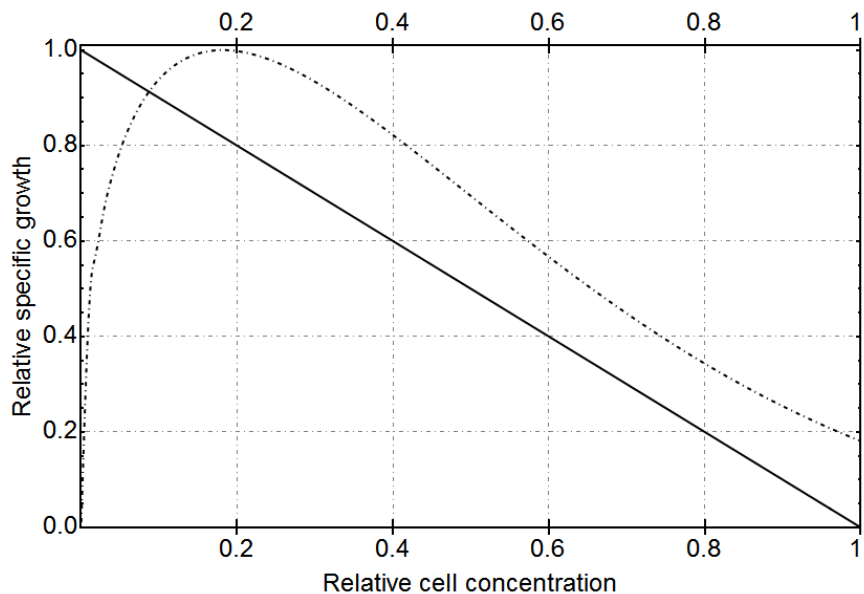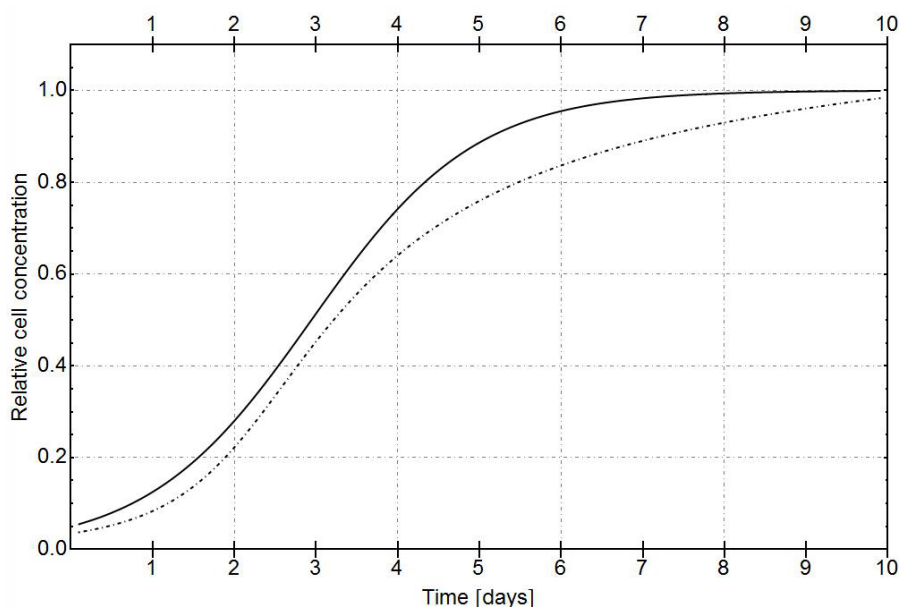

Figure 6.16 - Theoretical temporal specific growth kinetic curves for a logistic type (full curve) and generalized type of kinetics (dot-dashed curve). Temporal visualization of the graph in Figure 6.15, following work by Konak (1975). We assumed an initial relative cell concentration (inoculum) of 0.05.

The most merciful thing in the world, I think, is the
inability of the human mind to correlate all its contents.

HP Lovecraft, *The Call of Cthulhu*, in *Weird Tales*, Wildside
Press, February 1928

7

Reporter features upon *light-regime* perturbations

Where PhD applicant is able to connect the dots among different levels of biological
entities, and realizes that complexity sometimes has to be isolated and abstracted, in
order to retrieve knowledge from it.

Parts of the contents of this chapter are based on parts of the following journal articles:

- Montagud et al. **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal* 2011, **6**:330-342.

- Montagud et al. **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium.** *BMC Systems Biology* 2010, **4**:156.

7.1 Introduction.

After the turn of the millennium, great expectations were placed upon high-throughput data production with the advent of modern analytical techniques. Advances in medicine (Hood et al., 2004; Segal et al., 2005), molecular biology (Grünenfelder and Winzeler, 2002) and systems biology (Ideker et al., 2002) paved the way to new promises and challenges of modern biology. One of the major challenges resides on how to analyse and extract knowledge from the vast amounts of *omics* data being generated. Another major challenge is the ability to integrate these data together and retrieve new knowledge from the bundle.

Apart from the flux simulations, another important problem in the field of metabolic systems biology that can be addressed by using reconstructed genome-scale models is the integration of the different genome-wide bio-molecular abundance datasets, *i.e. omics* datasets, such as transcriptome and metabolome. Methods have been proposed to help revealing cellular transcriptional regulatory programs by using transcriptomic data, which is the most common and, presently, the only truly genome-wide type of quantitative *omics*. Most of them use gene-set enrichment methods reported in literature, analysing gene expression data in a biologically constrained way (Boyle et al., 2004; Doniger et al., 2003; Draghici et al., 2003; Maere et al., 2005; Subramanian et al., 2005).

Usually, transcriptome analysis assumes an all-to-all interaction among studied genes, which causes a huge dimensional search space, generating false positives and failing to enlighten underlying principles that may explain the observed behaviour. The dimensionality of the data analysis problem can be considerably reduced if biological information (*e.g.* physical or functional interactions between bio-molecules) is used in order to constrain the solution space (*i.e.* the number of possible regulatory hypotheses explaining the observed *omics* data), hence enhancing the possibility of uncovering the biological dimensions of the data (Ideker et al., 2002; Oliveira et al., 2008; Patil and Nielsen, 2005).

An example of algorithms for carrying out such an integrative analysis through the use of genome-scale metabolic networks is Reporter Features (Oliveira et al., 2008; Patil and Nielsen, 2005). Reporter Features algorithm allows integration of *omics* data with bio-molecular interaction networks, thereby allowing identification of cellular regulatory focal points (*i.e. reporter features*), for instance *reporter metabolites* or *reporter couplings* as regulatory hubs in the metabolic network. Cells respond to perturbations by changing the expression pattern of several genes involved in the specific part of the metabolism in which a

perturbation is introduced. Thus, the identification of these hubs in the metabolic network would outstand as potential design target of a biotechnological strategy, as their potential as major regulatory players would hint at their importance in the flux control and distribution.

Transcriptomic studies have been done in *Synechocystis* sp. PCC6803 since Suzuki et al. studied cold shock response in 2001 (Suzuki et al., 2001). These studies have been targeted to a wide range of goals: acclimation to high light (Hihara et al., 2001), salt and hyperosmotic stress (Kanesaki et al., 2002), irradiation with UV-B and white light (Huang et al., 2002b), light-to-dark transitions (Gill et al., 2002), knock out libraries of histidine kinases (Yamaguchi et al., 2002), building dynamic models of transcriptional behaviour (Schmitt and Stephanopoulos, 2003), redox-responsive genes (Hihara et al., 2003), iron deficiency and reconstitution (Singh et al., 2003), inorganic carbon limitation and the inactivation of *ndhR* (Wang et al., 2004), response to red and far-red light (Hübschmann et al., 2005), heat shock response (Suzuki et al., 2006), growth-phase differential gene expression (Foster et al., 2007), response to a pH 10 environment (Summerfield and Sherman, 2008), sulphur starvation (Zhang et al., 2008) and effects of light quality (Singh et al., 2009).

Following the scope of studying the characteristics of *Synechocystis* sp. PCC6803 as an industrially relevant production platform, in this chapter we have explored system-wide variations upon changes in light regime. Transcriptome data of these light shifts were acquired from a work first authored by Gill and done in 2002 at Stephanopoulos laboratory (Gill et al., 2002).

7.2 *i*Syn669 and *i*Syn811 as data integration scaffolds.

Reporter Features software was used to integrate transcriptional information over the reconstructed *Synechocystis* sp. PCC6803 network allowing us to infer regulatory principles underlying metabolic flux changes following shifts in growth mode. In particular, we analysed the data from a work (Gill et al., 2002) that reports the transcriptional changes caused in *Synechocystis* sp. PCC6803 by shifts from darkness to illumination conditions and back. As it can be understood from the rationale beneath the metabolic capabilities of this cyanobacterium discussed in previous chapters, the presence or absence of light drives big changes in the flux distribution through the network. We have focused our study on the relationship between the transcription of *Synechocystis* sp. PCC6803 genes and the reactions of the metabolic network. Associations between genes and

reactions (and genes and flus couplings) were identified, listing all the genes that performed or were involved in a specific reaction or coupling. With this information and the metabolic model and coupling analysis, Reporter Features analysis was carried out. In brief, the analysis helped to identify metabolites and couplings around which the transcriptional changes are significantly concentrated. These metabolites, termed *reporter metabolites,* represent key regulatory nodes in the network, as well as *reporter couplings* represent key regulatory nodes in the coupling landscape of the metabolism.

Gill et al. (2002) designed the experiment so that *Synechocystis* was grown to mid-exponential phase (A_{730} = 0.6 to 0.8). Then, the lights were extinguished and RNA samples were taken after 24 hours in the dark (*full dark*). Illumination was then turned back on for 100 minutes (*transient light*), followed immediately by an additional 100 minutes in the dark (*transient dark*).

Table 7.1 - KEGG orthology groups for the metabolic genes altered with the light shift.

	All time points		Dark to Light		Light to Dark	
	Number of genes	%	Number of genes	%	Number of genes	%
Energy Metabolism	128	60.38	128	51.82	127	61.65
Amino Acid Metabolism	25	11.79	31	12.55	24	11.65
Carbohydrate Metabolism	24	11.32	28	11.33	23	11.16
Metabolism of Cofactors and Vitamins	13	6.13	26	10.53	12	5.83
Nucleotide Metabolism	12	5.66	23	9.32	12	5.83
Lipid Metabolism	7	3.3	5	2.02	6	2.91
Membrane Transport	3	1.42	4	1.63	2	0.97
Biosynthesis of Secondary Metabolites	0	0	1	0.4	0	0
Biosynthesis of Polyketides and Nonribosomal Peptides	0	0	1	0.4	0	0
Total	212	100	247	100	206	100

In this chapter, we were interested in two aspects of this study: to identify metabolites and couplings around which regulation is centred during the light regime transitions; and to find the metabolic genes that were collectively significantly co-regulated across these transitions (Patil and Nielsen, 2005). The analysis was divided in three parts: an analysis of the data arrays from the whole experimental profile ("all time points"), an analysis of the shift from darkness to a light environment ("dark to light") and from light back to dark ("light to dark"). For a study of the overall genome and its light regulation, refer to Gill et al. (2002). In this chapter, as the relationship between the metabolism and this regulation was investigated, genes with no direct relationship to a metabolic reaction were not considered. Distributions of the genes across KEGG Orthologies related to the metabolism altered with the light shift are depicted in Table 7.1.

- *All time points*

When all seven arrays were used, *reporter metabolites* were found to be quite scattered across the metabolism spanning several metabolic pathways, and thus offering a global view of the transcriptional response in the metabolic network (see Figure 7.1 and Table 7.2a). Presence of some amino acids (*L-tyrosine, L-isoleucine*), nucleic acids and its precursors (*GTP, dihydroorotate*), carbon metabolism metabolites (*D-ribulose-5-phosphate, succinyl-CoA*), lipids precursors (*myo-inositol, D-myo-inositol 3-monophosphate*), cofactors (*thioredoxin, p-aminobenzoate*) and photosynthesis metabolites (*plastocyanin*) pictures a scenario of global regulation throughout the different metabolic pathways.

By using the metabolic sub-network search algorithm, we found 212 genes that have their expression changed across the arrays and that have a relationship with the metabolites of *i*Syn669 network. Furthermore, 50 genes were identified that are strongly co-regulated all along the profile of the experiment (Additional File 7.1, section a). This set of genes is characterized in two groups. The first set consists of the genes from photosynthesis (93.85%) and oxidative phosphorylation (6.15%). The second set is representative of a variety of genes from different pathways such as amino acid metabolism (39%), carbohydrate metabolism (22%), nucleotide metabolism (13%), nitrogen metabolism (13%) and metabolism of cofactors (9%) that globally regulates the entire metabolic network (see Table 7.1 for further details).
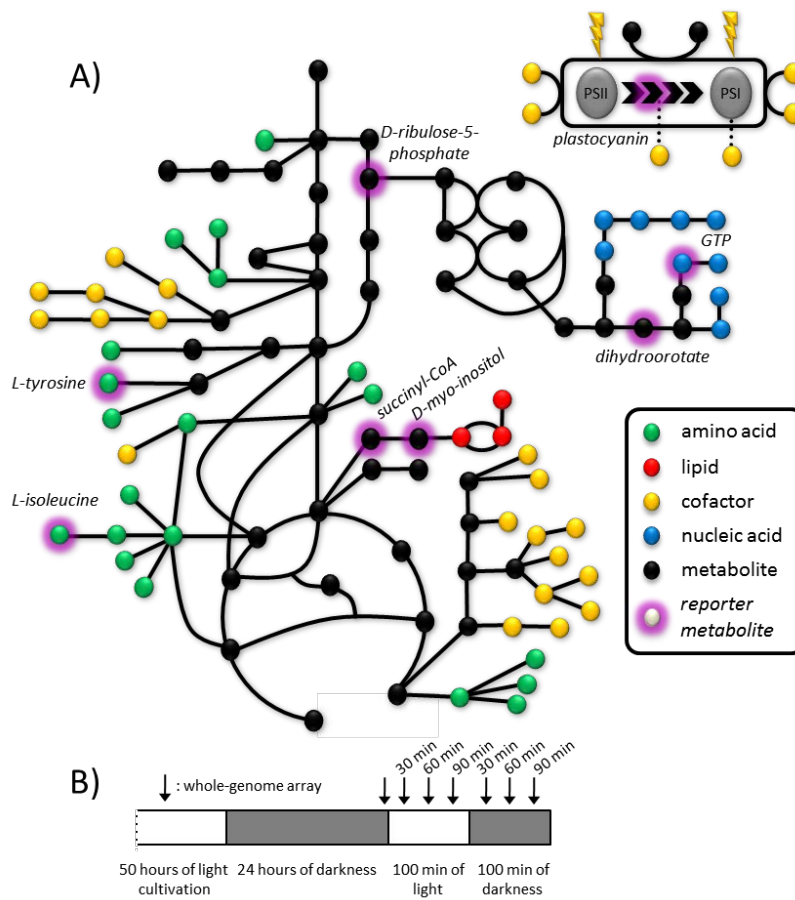
Figure 7.1 - *Reporter metabolites* under light/dark regime. A) *Reporter metabolites* for *all time points* set of arrays. B) Light/dark-shift profiles and localization of the genome arrays for the work from Gill et al. (2002).

It can be expected that an experimental design like the one we have based our work on, which combines a shift from dark to light with a shift back to darkness, will encompass an important part of the regulatory changes the cell is undergoing in its natural habitat. In a glucose-deficient environment, the presence or absence of light is the main condition around which the *Synechocystis* metabolism gravitates (Navarro et al., 2009). Indeed, one of the co-regulated sets consists of the genes coding for the proteins that work on, and around, the thylakoid membrane, let it be photosynthesis or oxidative phosphorylation genes.

Table 7.2 - *Reporter metabolites* for the light shift experiment. *Reporter metabolites* for each set of arrays analysed with Reporter Features software. Count means number of neighbours

A)

B)

Metabolite	Count	Metabolite	Count
All time points		**Dark to Light**	
L-tyrosine	4	N-carbamoyl-L-aspartate	3
N-carbamoyl-L-aspartate	3	dihydroorotate	3
dTDP	4	5-phosphoribosyl 1-pirophosphate	9
L-isoleucine	3	L-valine	3
D-ribulose-5-phosphate	4	5-phospho-ribosyl-glycineamide	3
D-myo-inositol (3)-monophosphate	2	O-phospho-L-homoserine	2
myo-inositol	2	peptidylproline (omega = 180)	4
L-valine	3	peptidylproline (omega = 0)	4
succinyl-CoA	3	indole-3-glycerol-phosphate	2
adenosine	2	5-aminoimidazole ribonucleotide	3
GTP	13	tetrahydrofolate cofactors	8
thioredoxin	11	GTP	13
thioredoxin disulphide	11	L-glutamate gamma-semialdehyde	2
p-aminobenzoate	2	inosine-5'-phosphate	5
acetylphosphate	2	pantetheine 4'-phosphate	2
glycine	7	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate	2
succinate	7	phytoene	2
dihydroorotate	3	thioredoxin	11
PC	12	thioredoxin disulphide	11

Table 7.2 - *Reporter metabolites* for the light shift experiment. (continued)

C)

Metabolite	Count
Light to Dark	
5-phosphoribosyl-N-formylglycineamidine	3
diphosphate	76
a 1,4-alpha-D-glucan_n	2
a 1,4-alpha-D-glucan_n1	2
UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminoheptanedioate	2
pyridoxine-5'-phosphate	2
(E,E)-farnesyl diphosphate	3
GMP	6
phosphoribosylformiminoAICAR-phosphate	2
L-aspartyl-4-phosphate	2
pantothenate	2
undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	2
MurAc(oyl-L-Ala-D-gamma-Glu-L-Lys-D-Ala-D-Ala)-diphospho-undecaprenol	2
undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine	2
L-aspartate-semialdehyde	2
5-phospho-ribosyl-glycineamide	3
5'-phosphoribosyl-N-formylglycineamide	4
sulphur	2
glycine	7

- *Dark to light*

Next, we considered the arrays that represent the shift from darkness to light, the first three arrays (from "24 hours of darkness" array to "60 minutes of light" array). *Reporter metabolites* were found to be largely within the nucleotide and amino acid metabolism (Table 7.2b). Some cofactors were also identified as regulation hubs like *tetrahydrofolate*, *thioredoxin* and *adenosylcobinamide*, which can be seen in Figure 7.2a.

Sub-network search yielded set of 247 genes that have their expression changed across the first three arrays and that are related with *i*Syn669 reactions. Furthermore, 84 genes were identified that are strongly co-regulated across the three arrays (Additional File 7.1, section b). This set of genes cover photosynthesis (25%), oxidative phosphorylation (24%), amino acid metabolism (11%), carbohydrate metabolism (11%), nucleotide metabolism (10%) and metabolism of cofactors (10%).

This set of data arrays are indeed a good example of a cell's metabolic machinery starting up. After a 24 hour period in darkness where cell density did not change (see Figure 1 in Gill et al. (2002)), light enters the system and the cell starts to synthesize new bio-molecules, mostly nucleotides so it can copy its genetic material and amino acids to build up proteins.

- *Light to dark*

Finally, we considered the arrays that represent the shift from light to dark, data from "90 minutes of light" array to "60 minutes of dark" array. Similar to the previous case study, *reporter metabolites* were found to be focused on the nucleotide and amino acid metabolism (Table 7.2c). Additionally, the presence of metabolite *a 1,4-alpha-D-glucan_n* and its cognate *a 1,4-alpha-D-glucan_n1* also stands out as they are involved in carbon reserves catabolism and anabolism, that can be seen in Figure 7.2b.

With the help of the sub-network search, 133 genes were identified as being significantly co-regulated across those three arrays (Additional File 7.1, section c). This set comprises of the genes from photosynthesis (34%), oxidative phosphorylation (26%), amino acid metabolism (12%), carbohydrate metabolism (12%), nucleotide metabolism (7.5%) and metabolism of cofactors (4.5%).
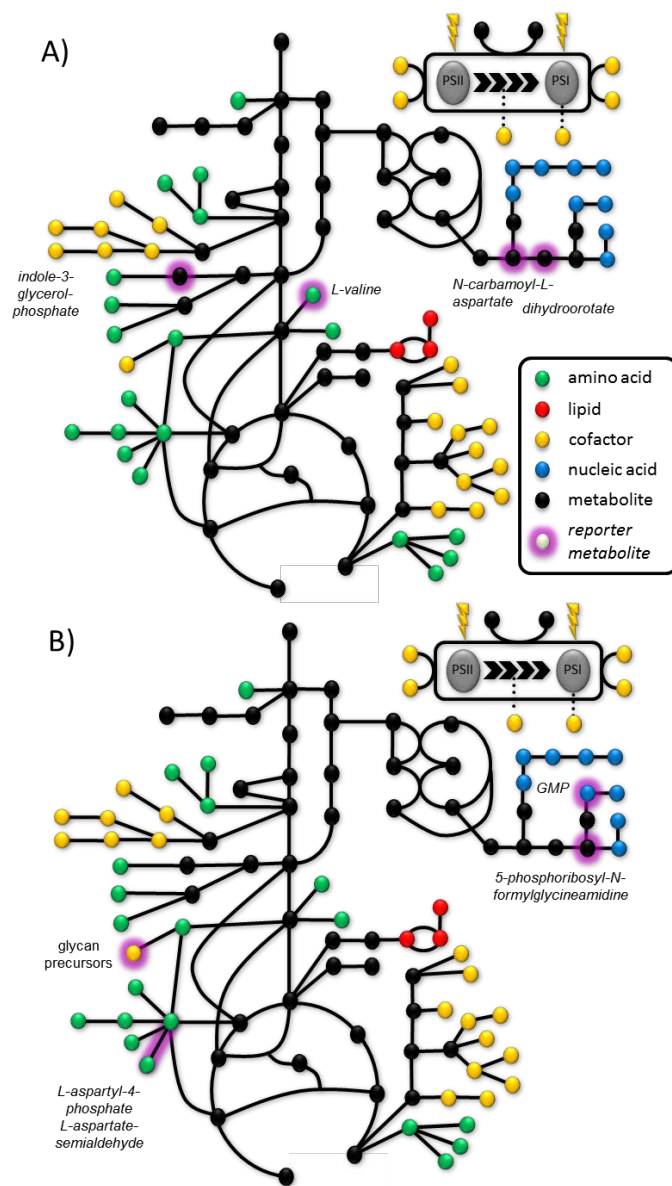
Figure 7.2 - *Reporter metabolites* under light/dark regime divided by shifts. A) *Reporter metabolites* for *dark to light* set of arrays. B) *Reporter metabolites* for *light to dark* set of arrays.

This last set of data array is a scenario where metabolism is being shut down, because of darkness and lack of carbohydrate sources. Without light, photosynthesis is blocked and carbon fixation is nearly obliterated. Cells strive to build up carbon reserves (hence the presence of *a 1,4-alpha-D-glucan_n* as a reporter metabolite) and oxidative phosphorylation is the main energy pathway that remains present. Regulation is centred on the energy metabolism shift (60%

of the total co-regulated sub-network), withholding amino acids and nucleotide precursors and keeping the cofactors available in a low-profile metabolism.

7.3 Reporter couplings study.

We also applied Reporter Features algorithm to the coupling network in order to identify *regulatory hubs –reporter flux coupling pairs* and *reporter flux coupling groups*. As before, we address the task of identification of flux coupling reaction pairs or groups that are differentially regulated during light shifts: from 24 hours of darkness to 100 minutes of light to 100 minutes of darkness (see Figure 7.1b). Genome-wide transcription data was taken from the study by Gill et al. (2002).

Reporter coupling pairs.

A coupling characterises functional relationship between the two reactions, *viz*. directional, full or partial coupling (see **Chapter 5**). Each coupling pair was independently studied across the transcription data arrays of a light-shift experiment discovering regulatory hubs –*reporter coupling pairs* (Additional file 7.2).

- *All time points*

Using all seven transcription arrays as input, we covered two light shifts: turning on the photosynthesis and metabolic machinery and shutting it down. We looked for couplings with significant co-regulation among all seven arrays. Most statistically significant pairs were the ones covering the connection between pathways from the central carbon metabolism (*purine metabolism, coenzyme A biosynthesis*) or biological building blocks (*fatty acid biosynthesis, porphyrin and chlorophyll metabolism, biosynthesis of steroids*) to *amino acid pathways* (*glutamate, threonine*).

- *Dark to light*

In this analysis, we focused on the shift from darkness to the growth in the presence of light. RuBisCO (reaction *4.1.1.39* in the model) stands as the most influential regulatory hub in this set and most significant pairs bear this vital reaction. Additionally, the core carbon metabolism (*Calvin cycle, TCA cycle*) is significantly paired to building block synthesis pathways (*carotenoid biosynthesis,*

porphyrin and chlorophyll metabolism, fatty acid biosynthesis) as well as to amino acid biosynthesis (*histidine, aspartate*). Photosynthesis reactions are scarcely present in seven pairs, which are not the most significant.

- *Light to dark*

Finally, we analysed the shift from an illuminated condition to darkness. Most statistically significant pairs cover pathways from the central carbon metabolism (*sucrose metabolism, glycolysis, pentose phosphate pathway*) links to biosynthesis pathways (*biosynthesis of steroids, purine metabolism, porphyrin and chlorophyll metabolism, fatty acid biosynthesis*) and to amino acids (*aspartate, isoleucine*). Interestingly, photosynthesis reactions are predominant in this set (93 out of 586), even though they are not present in the most significant pairs. Additionally, carboxylative reaction from RuBisCO is not significant.

Reporter coupling groups.

Groups of coupled reactions were compiled by connecting reactions coupled to each other. This way, we were able to identify *higher order regulatory hubs – reporter coupling groups* for the light shift dataset (Additional file 7.3).

- *All time points*

We looked for couplings groups with significant co-regulation among all the seven arrays, which covered both light shifts. Glutamate directionally coupled metabolism (*dir2*), purine fully coupled biosynthesis (*ful4*) and pyrimidine fully coupled set (*ful16*) were identified as *reporter coupling groups* (their knitting over the *i*Syn811 coupling network can be seen in Figure 7.3).

- *Dark to light*

We found three *reporter coupling groups* that were significant during the start-up of metabolism following the availability of light. Reporter couplings were glutamate directionally coupled metabolism (*dir2*) and two fully coupled sets: pyrimidine (*ful16*) and NAD(P) metabolism (*ful3*).
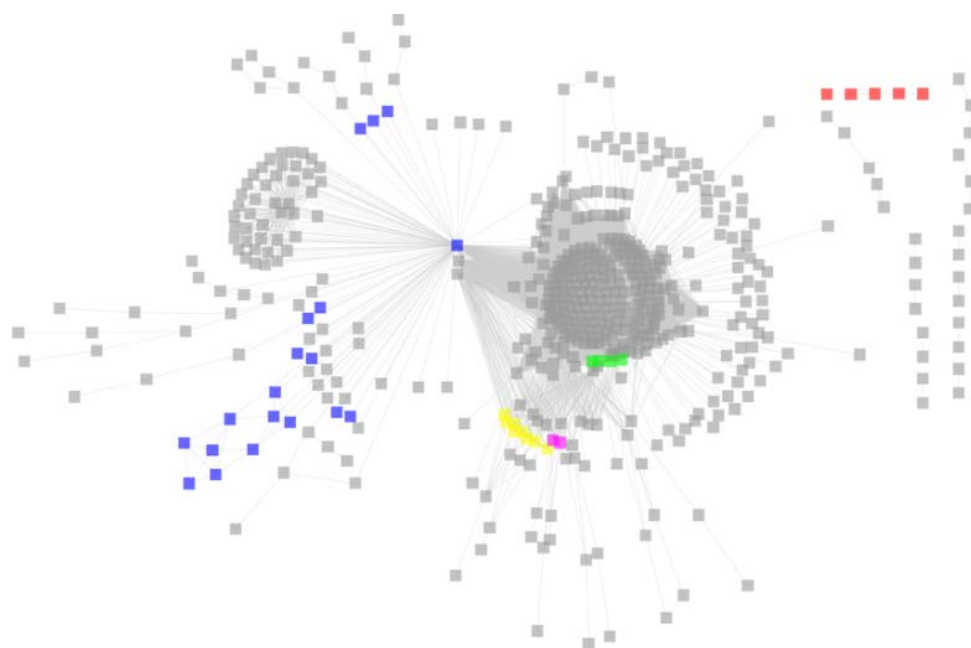
Figure 7.3 - Reporter flux coupling network for autotrophic growth condition. Pyrimidine fully coupled set (*ful16*) in light purple, blue colour represents the photosynthesis and oxidative phosphorylation cluster of dir1 coupling group. These groups are reporter couplings in all the studied conditions: all time points and in the light shifts from dark to light and light to dark. Purine fully coupled biosynthesis (*ful4*) in green colour highlights regulation in all time points and during the shift from light to dark. Glutamate directionally coupled metabolism (*dir2*) in red is identified to be a regulatory hub on all time points. Fully coupled set of NAD(P)H turnover (*ful3*), depicted in yellow, is a regulatory centre in dark to light environmental shift.

- *Light to dark*

Two fully coupled groups were revealed as *reporter coupling groups*: pyrimidine (ful16) and purine biosynthesis (ful4).

Finally, as *dir1* coupling group conveys 388 reactions (32.7% of the total) we have used clustering methods from Bader and Hogue (2003), downloadable from http://clusterviz.sourceforge.net, to identify significantly co-regulated subsets. The cluster made up of all directionally coupled reactions from the photosynthesis and the oxidative phosphorylation from *dir1* set was discovered as such in all three data sets, even though the 388-mer *dir1* set, as a whole, is not.

7.4 Conclusions.

We also show that metabolic models can be used as a scaffold to integrate system-wide *omics* data. As a case study, we identified key *reporter metabolites* and *reporter coupling pairs* and *groups* around which regulation during light shifts is organized, as well as gene sub-networks that were co-regulated during environmental variability in light.

As *reporter metabolites*, we have identified several metabolites that play a role as regulatory hubs when metabolism is being turned on ("dark to light") and shut down ("light to dark"). Similarly, we have identified the regulatory potential (and its principal actors) of this organism in terms of metabolism control ("all time points").

We have also integrated the transcriptome analysis on top of the coupling analysis of **Chapter 5**. The first reactions of *pyrimidine biosynthesis* and the *photosynthesis* and *oxidative phosphorylation* were identified as *reporter couplings*, regulatory hubs around which transcriptional changes are organized.

If researchers aim to design and build mutants with improved production of a given industrially relevant metabolite, focus should be laid upon the regulatory hubs that drive the production of this metabolite and to the energetic metabolic pathways that fuel up the cell. Else, mutants can be undermined by a lack of precursors or biomass potential.

This chapter is a nice proof of concept on how different layers of biological information can be interwoven together and knowledge useful for biotechnological uses can be retrieved from it. From our results, it seems that the hypothesis that cellular response to a perturbation can be modularized and characterized by using network topology information is not contrasted and, thus, algorithms such as Reporter Features can be applied to a great extent to biological networks.

7.5 Methods.

Transcriptome data analysis

Reporter Features algorithm (Oliveira et al., 2008; Patil and Nielsen, 2005) was used to integrate transcriptomic data in the metabolic reconstruction and in the flux coupling networks. This algorithm works with three kinds of information:

(i) *p-values* for genes, resulting from, for example, Student's *t*-test run on transcriptomic data,

(ii) *interaction file*, where genes/reactions are connected to the corresponding features, in this case the corresponding substrates and products as well as the sets of coupled reactions, and

(iii) *association file*, where genes are linked to the corresponding reactions, either by coding for the enzyme or by regulating the gene that codes for the enzyme.

In brief, Reporter algorithm converts the *p-value* for a given node (p_{gene_i}) to a z-score by using the inverse normal cumulative distribution function (cdf^{-1}).

$$z_{gene_i} = \text{cdf}^{-1}(1 - p_{gene_i})$$

After scoring each non-feature node in this fashion, we need to calculate the score of each feature *j*, $z_{feature_j}$. We used the scoring method based on distribution of the means, which is a test for the null hypothesis "genes adjacent to feature *j* display their normalized average response by chance". In particular, the score of each feature *j* is defined as the average of the scores of its neighbour N_j nodes (genes), *i.e.*:

$$z_{feature_j} = \frac{1}{N_j} \sum_{k=1}^{N_j} z_{gene_k}$$

To evaluate the significance of each $z_{feature_j}$, this value should be corrected for the background distribution of z-scores in the data, by subtracting the mean (m_N) and dividing by the standard deviation (s_N) of random aggregates of size N.

$$z_{feature_j}^{corrected} = \frac{\left(z_{feature_j} - m_N\right)}{s_N}$$

Transcriptomic data was retrieved from a contribution that studied clusters of genes differentially expressed with and without light (Gill et al., 2002). In this work, *Synechocystis* sp.PCC6803 was grown under 24 h of darkness, followed by 100 min of light and 100 min of darkness. Seven genome-wide transcriptional analyses were performed over the length of the experiment, see Figure 7.1b and corresponding reference (Gill et al., 2002).

"Forty-two!" yelled Loonquawl. "Is that all you've got to show for seven and a half million years' work?"
"I checked it very thoroughly," said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."
"But it was the Great Question! The Ultimate Question of Life, the Universe and Everything!" howled Loonquawl.
"Yes," said Deep Thought with the air of one who suffers fools gladly, "but what actually is it? [...] once you do know what the question actually is, you'll know what the answer means."

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*, Del Rey Publisher, 1995

8
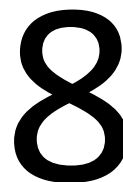
Conclusions and closing remarks

Where PhD applicant looks back at all the work done, draws a set of conclusions and suggests some improvements in his work and the other's that are not meant to be considered *ex cathedra*, but humble proposals.

8.1 Conclusions of the dissertation.

This thesis focused on the construction, uses and applications of constraint-based metabolic models of cyanobacteria oriented towards the construction of a photon-fuelled production platform of socioeconomic importance. The scope of this work is, thus, to build bridges between different scientific areas like biotechnology, systems biology, metabolic engineering, synthetic biology, energy engineering, and cyanobacterial biology. Our final goal is a biotechnological one, as we are trying to assess how to tame biology from an industrial perspective. However it seems foolish to pretend to tinker some system of which one understands little. That's the reason behind the efforts throughout this thesis focused on acquiring knowledge from a system with a small research community (in comparison to other model organisms), and with high-throughput information from a handful of experiments.

Additionally, biotechnological advances in metabolism are mainly based upon mutant generation. These mutants, let them be knock outs, knock ins, knock downs, etc., have a holistic effect on different parts of the metabolism. Researchers are not blind on this fact, and their habit of ignoring this has more to do with the lack of tools to cope with this complexity and uncertainty of data than with their willingness to overlook these systemic effects. Thus, it is of critical importance to pave the way to a situation where holistic analyses are possible and quantitative studies are conceivable. Simplifications, like identifying proteins to nuts and bolts, abstractions, like clustering gene expression around *reporter metabolites*, use tools from foreign areas, like Bayesian probability in transcriptomics; all of those are valid and useful if they allow us to get closer to a world where quantitative, all-inclusive approaches are possible and let us infer more efficient ways to *use* biology to our needs.

When first entering this systems biology/biotechnology field and interacting with colleagues and partners, I drew a rough comparison to geographical orientation skills: biologists drawing the map where the research underwent and engineers having the compass pointing to the goal of the project. I still think that this comparison stands due to the logic behind scientific fields: biology is a science to answer *how* and *why* questions, while engineering is technique field to know *what for* something could be used. Biotechnology combines both, as you cannot get oriented without a map and a compass. Hence, **Chapter 2** focused on the construction of this map, on top of which further developments will be based upon: the set of pathways and routes that enable researchers to be aware of what are the routes that nature has taken to allow some reactants to become products.

We used updated data from different sources to gather the information necessary to have the set of reactions described in *Synechocystis* sp. PCC6803. Alas, we concluded in this chapter that this was an iterative process with the rest of the works of the present thesis as, due to uncertainty of data and on-going research there is no such thing as a completed genome-scale metabolic model. Furthermore, we laid the founding stone of an automated and high throughput process of genome-scale metabolic model reconstructions done in my group, mainly by PhD student Reyes and Dr Gamermann.

The knowledge of these pathways lead us to dig one level down, to the flux level, for this in **Chapter 3** we used a common algorithm used in the field termed *flux balance analysis* (FBA) (Stephanopoulos et al., 1998). This allowed us to have **a computer-based model of *Synechocystis* sp. PCC6803 that behaved as close as data could do to a real life bacterium**. Here, the verb *behaves* is an elegant way to define *transfer function*: a function that, given an input, retrieves an output in a linear time-invariant system. In our case, inputs and outputs are mass: substrates and cell biomass; thus, our transfer function is one that involves energy flow. **Our model represents the possibility of *growth*, as it *generates biomass* or *allows energy fluxes*, in the environmental conditions where the bacterium could potentially grow**: photoautotrophy, photoheterotrophy, mixotrophy, and *dark* heterotrophy. **Not only it grows, but also its flux landscape is similar to the one seen experimentally.**

From the knowledge we had gathered in pathway presence and flux estimations, it seemed feasible to study the simulation of probable mutants and estimate the effects of mutations in growth and in the production of industrially-relevant metabolites. Therefore, in **Chapter 4** we focused in evaluating different mutants and seeing their effect in terms of growth and production of ethanol, succinate, and hydrogen. We identified different sets of mutants that could evolve optimized titters of these metabolites. This study, and the potential of *Synechocystis* **sp. PCC6803 to grow photoautotrophically, allowed us to present this organism as a good photon-fuelled production platform candidate**. Furthermore, using lethality studies we also spotted that this organism has more lethal genes (genes whose absence causes metabolism to be non-functional) than other model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. Nonetheless, as we defend in **Chapter 4** these results have to be handled with care as the amount of experiments and evidences in these two organisms are orders of magnitude larger than the ones regarding our cyanobacterium.

Additionally to the flux and knock outs studies, we felt that it was necessary to analyse the relationships among reaction functionalities and the potentialities of *Synechocystis* sp. PCC6803. In **Chapter 5** we addressed this issue with the *flux coupling analysis* (FCA) algorithm (Burgard et al., 2004) that studies the coupling capacity of the reaction network. Interestingly, we accomplished to make sense out of the huge connectivity degree that *Synechocystis* sp. PCC6803 metabolic network possesses. We found that catabolic reactions are connected to a reduced set of reactions that generate the biological building blocks and energetic metabolites; anabolic reactions, in turn, stem from this reduced set. This drove us to think about the *bow tie* metabolic organization theorized several years ago (Csete and Doyle, 2004). This **network configuration makes it difficult to modify core reactions away from the organism wild type behaviour without diminishing its potential** (like mutating its NADPH hydrogenases), **but allows tinkering of exterior pathways without drastically affecting cell growth** (like adding cellulose breakdown enzymes that connect to glycolysis). Additionally, we focused on the production of biofuels, hydrogen and ethanol, and found out that energy production pathways are intimately connected to hydrogen in such a way that obliges researchers to think of alternatives others than *just* knock an heterologous hydrogenase in, as these *core* reactions may be bottlenecks for boosting the H_2 production in this organism. Ethanol, instead, seems simpler, as its production is coupled with pyruvate usage, its substrate, and ATP production, like most of the pathways. Modifications on these reactions could be very interesting in order to have an enhanced production of ethanol in *Synechocystis* sp. PCC6803.

Most of the work in this thesis was framed in an FP6 European project called BioModularH2 (BioModularH2, 2005). My group collaborated in this project along with five other research groups of diverse backgrounds and abilities. In this three-year long project among other collaborations, one project appeared between Paula Tamagnini group at IBMC in Porto and my group in Valencia regarding growth data processing and analysis. One of the fruits of this collaboration, among others like Pinto et al. (2011), has been the work of **Chapter 6**. As we could not find a complete medium composition study, we decided to go ahead and perform one by ourselves. We concluded that out of the considered factors that affect growth (carbon source, nitrogen, temperature, pH, and light) the **governing factors are light irradiance, carbon source and temperature**. Interestingly, we found that **in autotrophic growth, there is a significant effect and interaction of pH and temperature**, and *Synechocystis* sp. PCC6803 has better growth under alkaline conditions.

Finally, we worked in a part of systems biology that I find very interesting and that, in my opinion, can facilitate researchers considering holistic studies as I discussed above. This part is the interaction and integration of different levels of information. In **Chapter 7** we present the use of Reporter Features algorithm (Oliveira et al., 2008; Patil and Nielsen, 2005) with different biological data (transcriptomic, network connectivity and flux coupling data) employing genome-scale metabolic reconstruction *i*Syn669 (Montagud et al., 2010) and *i*Syn811 (Montagud et al., 2011) as data scaffolds. We successfully identified *reporter metabolites* in a light-shift experiment from Gill et al. (2002): metabolites that play a role as regulatory hubs when metabolism is being turned on ("dark to light" data array) and shut down ("light to dark" data array). Similarly, we have identified the regulatory potential (and its principal actors) of this organism in terms of metabolism control ("all time points" data array). These *reporter metabolites* were not a few highly-connected metabolites that governed the whole metabolism; instead, they are a handful of metabolites, each of one connecting a different part of the metabolism. In a sense, we can draw the conclusion that **metabolism does not regulate its flux shifts through a few major players, but through a set of medium-sized players**. Additionally, we used flux coupling data to identify *reporter couplings*, around which transcriptional changes are organized: energetic reactions, like *photosynthesis* and *oxidative phosphorylation*, were identified as such. This is connected to the aforementioned coupling analysis, where almost all reactions were connected to ATP synthase and other energy pathways: **potential of cell production is intimately related to energy acquisition and handling**.

8.2 *Synechocystis* sp. PCC6803 as a production platform.

We have demonstrated the feasibility of building a reliable and complete metabolic model at the genome-scale level that includes all reactions annotated in the genome and, thus, known to be present in the organism. In order to start this work, *only* annotation files are needed. Identification and description of genes and, therefore, of cellular activities are done through similarities studies among already annotated genomes. High-fidelity sequencing and better identification is a request if we want to have a perfect metabolic model. These techniques have greatly evolved since 1995 (when *Synechocystis* sp. PCC6803 genome was first sequenced) or 2008 (when I started this work), but we are confident that in the future these initial steps will be streamlined and error-safe, at least to some extent.

Synechocystis sp. PCC6803 metabolic model can be used to study flux capabilities and potentials, as well as to study possible mutants to have an enhanced production of a given metabolite. In this sense, we have demonstrated the use of *flux balance analysis* to estimate metabolic behaviour. One of the bottlenecks of these simulations is the need to know several biological constraints of substrate drain and by-products generation. Nonetheless, the amount of information generated with this workflow, that only needs the network configuration and these constraint-based flux simulations, is order of magnitude higher than the information needed as input.

As for mutant studies, we have spotted a set of knock out mutants with increased production of succinate, ethanol and hydrogen. This set of mutants points to design strategies that range from the more classical, like shutting down flux diverging routes from objective metabolite, like the triple knock out proposal for improved production of ethanol, to the more synthetic biology ones like, for an improved succinate production, designing a new pyruvate kinase protein specific for ATP cofactor.

Furthermore, we identified theoretical maxima of production for succinate, ethanol and hydrogen. These values are far from what is currently produced in literature. For succinate, we have estimated 0.5695 mmol g_{DW}^{-1} h^{-1} with a growth rate of 0.0714 mmol g_{DW}^{-1} h^{-1} to 0.195 mmol g_{DW}^{-1} h^{-1} measured in a recombinant *Escherichia coli* with a cyanobacterial carbonic anhydrase (Wang et al., 2009). For hydrogen, estimated specific growth rate 0.0448 h^{-1} with corresponding maximum H_2 production rate 0.085 mmol g_{DW}^{-1} h^{-1} can be compared to typical 3.149 10^{-4} mmol g_{DW}^{-1} h^{-1} measured (Baebprasert et al., 2011). As for ethanol, even though *Synechocystis* sp. PCC6803 genome bears the genes to produce it, there is a negligible production of this metabolite in wild type strains (P. Wright, personal communication). As for genetically engineered strains, Joule Unlimited (www.jouleunlimited.com) has reported in the patent literature to secrete ethanol at a rate of 1 mg L^{-1} h^{-1} (Devroe et al., 2010), and academic literature report typical levels of 0.2 mg L^{-1} day^{-1} (Deng and Coleman, 1999). For a nice review on cyanobacteria production of interesting metabolites, please refer to Ducat et al. (2011).

Flux landscapes studies made us wonder about the capabilities of this network. We performed *flux coupling analysis* in order to uncover non-straight forward functional links among fluxes. *Flux coupling analysis* tries to uncover part of the effects that hinder mutant generation, allowing researchers to explain mutant phenotypes that were not considered because no prior knowledge was known or

because connectivity was not straight-forward. We found out that in terms of feasibility, it seemed easier to tinker flux couplings so to have an increased ethanol production than an increased hydrogen production. Hydrogen is interwoven together with photosynthesis and oxidative phosphorylation, making it difficult to design strategies that increase hydrogen, but don't affect growth capabilities much.

Additionally, we have seen that metabolism regulation is organized around hubs that outstand as *reporter metabolites*. These hubs can be very informative when studying a mutant with a selected enhanced production. In our analysis, considering a complete light-regime shift, *reporter metabolites* were identified as cofactors, amino and nucleic acid precursors, central carbon metabolism, lipids precursors and photosynthesis metabolites. Photosynthesis and oxidative phosphorylation were identified as *reporter couplings*. Reactions that should be taken into account when thinking of altering *Synechocystis* sp. PCC6803 metabolism.

Researchers were seduced by the idea of using this cyanobacterium, due to its capability of producing metabolites autonomously from carbon dioxide and photons, thus photoautotrophically. In this growth mode, photosynthesis provides electrons and oxygen from water, using photons as energy donors, pumping protons that can be used to generate ATP. Oxygen can be converted back to water by oxidative phosphorylation, mainly draining NADH and generating proton potential that can also be used to generate ATP. Hydrogen feeds from NADPH that, in turn, feeds from these electrons. Additionally, hydrogen is inhibited by oxygen. The more photosynthesis, the more electron flow and the more hydrogen-inhibiting oxygen are produced. Hence, photosynthesis, oxidative phosphorylation and hydrogen production are metabolic pathways that we have to engineer system-wide if we want to be successful.

Researchers have usually followed a two-steps strategy where biomass was generated in the first one, and hydrogen produced transiently in the second one. Melis et al. (2000), working with *Chlamydomonas reinhardtii*, reduced photosynthetic capability in order to have prolonged hydrogen production, showing that this mutant had matching titters of photosynthesis and oxidative phosphorylation, and allowing to the latter the consumption of the oxygen the former produced.

In the course of my thesis research, several strategies have been drafted for future, probable, increased hydrogen evolving projects. Most of them are related with the coordination of oxidative phosphorylation and photosynthesis levels. If

Melis et al. work tried to tinker photosynthesis, we have focused in tinkering oxidative phosphorylation. Through a set of mutants, we would like to increase the levels of oxidative phosphorylation so they could match photosynthetic levels. Examples of these would be inclusion of terminal oxidases that drain oxygen and tinkering of ATPase P/O ratio, through faulty ATPases in order to change proton gradient to ATP generation ratio. Alas, computer-based work has pointed at the fact that there would be little room for *free* electrons for hydrogen evolution. Thus, a mixotrophic strategy is being considered, which would increase hydrogen titters, but would ruin the efforts of having an autonomous carbon- and photon-based production platform. In addition, experiments to false these hypotheses are being conducted at present time.

Finally, if we are to live in a hydrogen economy several years from now, there are a series of milestones that need to be accomplished. In this direction, the US National Research Council commissioned in 2004 a study to identify these efforts. This Committee on Alternatives and Strategies for Future Hydrogen Production and Use (Committee on Alternatives, 2004) considered that, first of all, the hydrogen system must be cost-competitive, it must be safe and appealing to the consumer, and it would preferably offer advantages from the perspectives of energy security and CO_2 emissions. Specifically for the transportation sector, dramatic progress in the development of fuel cells, storage devices, and distribution systems is especially critical.

The Committee on Alternatives and Strategies for Future Hydrogen Production and Use concluded that, in order to have a hydrogen-fuelled transportation as a beachhead of a clean hydrogen economy, the four most fundamental technological and economic challenges are these:

- To develop and introduce cost-effective, durable, safe, and environmentally desirable fuel cell systems and hydrogen storage systems.

- To develop the infrastructure to provide hydrogen for the light-duty-vehicle user

- To reduce sharply the costs of hydrogen production from renewable energy sources, over a time frame of decades

- To capture and store ("sequester") the carbon dioxide by-product of hydrogen production from coal

In present thesis we have assessed the suitability of *Synechocystis* sp. PCC6803 as a production platform and completed the first steps of that design, completing a metabolic model that integrates different levels of biological information, which is useful to study production potentialities of industrially-relevant metabolites. Nonetheless, this would *only* improve hydrogen production from renewable energy sources and, up to some point, solve the carbon dioxide capture generated by other energy strategies –like coal plants. But there are many other fields where R&D efforts are needed in order to be able to live in a clean hydrogen economy.

168

Afterword

At the break of the day, the refugees' fleet spotted the shore of what would become our new home, our new hope –Nueva Tanelorn. We were called to deck, and there we admired the view: Roark, Eora, Dao Long, Hennit, Aldor, Kromar, Darcyl, and myself, among others. This was it, at the end, a new beginning.

J. Bruegel, *A new hope*, Heroes' Feast Publisher

(Left) XKCD, Overestimulated (http://xkcd.com/602)

Jack comes upon a penguin. The penguin looks at him, cocks his head to signal Jack forward.
Penguin: Slide.
The penguin jumps onto a patch of ice and slides away.

Fight Club, screenplay by Jim Uhls, novel by Chuck Palahniuk, 1999

Ja està, ací acaba esta feina. I per a ser algú que no disfruta amb els comiats, ho estic allargant força. Així i tot, permeteu-me afegir unes darreres paraules a este text. I és que tot este treball, totes estes hores de picar pedra i córrer amunt la muntanya no hagueren pogut ésser possibles sense, no ja una persona, sinó un grup de persones, una generació.

En tota família mitjanament humil hi ha una generació que fa lo impossible perquè els seus fills apleguen a cotes acadèmiques que ells no podien albirar. En la meva família, eixa generació fou la dels meus avis. Curiós, o no tant, fou la saviesa que dóna la reflexió el que espurnejà que les meves àvies

facilitaren el pas dels meus pares a poder estudiar una carrera. Elles foren les que veieren que el futur no estava en la terra, en *desllomar-se* més que el veí, sinó en conèixer més que el veí. Elles envejaven el coneixement i, reconeguent-ho, insistiren en obrir una porta on altres veien murs. Persones profundament catòliques, en el sentit místic del catolicisme, no tant en l'eclesiàstic, no m'haguera estranyat que em digueren algo similar a "*el reialme terrenal serà el de la ment*". No m'ho digueren ni m'ho podran dir, ambdues passaren a l'altra banda de l'espill temps ençà, una quan jo passava del món de l'institut al món de la universitat, l'altra quan abandonava el món de la universitat i m'endinsava en aquesta tesis. Els seus consorts, els meus avis, tampoc pogueren veure este treball acabat, ambdós estan junt a elles. Em recorde a sovint de tots quatre. Igual que em recorde sovint de mon pare.

Òbviament, el que una generació òbriga la porta a la següent no fa que la següent la traspasse. Així, foren els continuats esforços dels meus pares els que els permeteren acabar una formació acadèmica alhora que no abandonaren, no podien, els quefers dels meus avis. Era el que *calia* fer, punt. Hi hagueren moltes discussions en eixes cases a l'època: de vegades comprometent temps, altres comprometent diners. He d'admetre que gràcies, també, a que els meus pares aguantaren la pluja a la intempèrie puc jo, ara, gaudir d'esta calor netament burgesa.

El record dels esforços d'una generació i de la lluita contracorrent de l'altra fan que quan tinc la temptació de deixar-me anar en eixa còmoda calor, s'òbriga una finestra de vent fred de realisme que impedeix adormir-me. Amb tot, gràcies per estar ahí, ja siga de cos present o amb records. Gràcies, a uns, per obrir la porta i gràcies, a altres, per traspassar-la.

Gràcies per permetre'm, per acció o omissió, començar, continuar i acabar "*allò que jo vullguera fer*". Al cap i a la fi, la llibertat no és altra cosa que elegir, i elegir implica responsabilitat i confiança en triar l'opció correcta.

I sempre ens quedarà poder mirar a allò fet i exclamar, com Fede Montagud aquell estiu de 1993, en guanyar el seu segon Campionat de Galotxa "Trofeu El Corte Inglés": *Som els més pintxos de València!*

<div align="right">

Arnau Montagud Aquino
Montolivet, València,
Dia de Sant Josep de l'any 2012

</div>

Appendices

Additional file 1.1 - *i*Syn669 reactions to gene connections

Excel file with the list of *i*Syn669 reactions and its cognate list of genes.

Additional file 1.2 - *i*Syn669 genome-scale metabolic model in OptGene format

Text file with the stoichiometric model, in OptGene (Patil et al., 2005) format, with all the constraints needed for its simulation with FBA algorithm.

Additional file 1.3 - *i*Syn811 reactions to gene connections

Excel file with the list of *i*Syn811 reactions and its cognate list of genes.

Additional file 1.4 - *i*Syn811 genome-scale metabolic model in OptGene format

Text file with the stoichiometric model, in OptGene (Patil et al., 2005) format, with all the constraints needed for its simulation with FBA algorithm.

Additional file 2.1 - *i*Syn669 metabolic fluxes simulated under four conditions

Excel file with all the reactions simulations and resulting flux ranges from the model simulated under four growth conditions: autotrophy, dark o pure heterotrophy, photoheterotrophy and mixotrophy.

Additional file 2.2 - *i*Syn811 metabolic fluxes simulated under four conditions

Excel file with all the reactions simulations and resulting flux ranges from the model simulated under four growth conditions: autotrophy, dark o pure heterotrophy, photoheterotrophy and mixotrophy.

Additional file 4.1 - FBA and MOMA simulation values for biomass growth in *Synechocystis* sp. PCC6803, *Escherichia coli* and *Saccharomyces cerevisiae* genome-scale metabolic models

Excel file with the growth values under MOMA simulation for *Synechocystis* sp. PCC6803, *Escherichia coli* and *Saccharomyces cerevisiae*. Data for *Synechocystis* is original from present work, data for *Escherichia coli* has been obtained from metabolic model from reference 18 and data for *Saccharomyces cerevisiae* is from reference 30.

Additional file 4.2 - XLS file with values of single, double and triple mutants for succinate, ethanol and hydrogen.

Excel file with growth values and design objective titters of single, double and triple mutants for enhanced producing strains.

Additional file 5.1 - XLS file with description of coupled reactions in all four growth modes.

Each Excel sheet can be easily converted to a SIF file to view network in Cytoscape (http://www.cytoscape.org). Finally, an Excel sheet with the reactions differently coupled across different growth conditions is provided.

Additional file 7.1 - *i*Syn669 groups of correlated genes in the three sets of arrays of light shift experiments.

Word file with the list of *i*Syn669 correlated genes in "All time points", "Dark to light" and "Light to dark" analyses.

Additional file 7.2 - XLS file with *reporter coupling pairs* analysis results under *all time points*, *dark to light* and *light to dark* conditions.

Additional file 7.3 - XLS file with coupling groups that have been identified as *reporter couplings groups*, under *all time points*, *dark to light* and *light to dark* conditions.

Bibliography

Abril D, Abril A. (2009) **Ethanol from lignocellulosic biomass**. *Ciencia e investigación agraria*, **36**:177-190.

Agrawal R, Singh NR, Ribeiro FH, Delgass WN. (2007) **Sustainable fuel for the transportation sector.** *Proceedings of the National Academy of Sciences of the United States of America*, **104**:4828-33.

Albert R, Jeong H, Barabási A-L. (1999) **Diameter of the World-Wide Web.** *Nature*, **401**:398-399.

Albertsson P. (2001) **A quantitative model of the domain structure of the photosynthetic membrane.** *Trends in Plant Science*, **6**:349-58.

Allakhverdiev SI, Murata N. (2004) **Environmental stress inhibits the synthesis de novo of proteins involved in the photodamage-repair cycle of photosystem II in *Synechocystis* sp. PCC6803**. *Biochimica et Biophysica Acta*, **1657**:23-32.

Allen JF. (2002) **Photosynthesis of ATP-electrons, proton pumps, rotors, and poise.** *Cell*, **110**:273-6.

Allen MM, Smith AJ. (1969) **Nitrogen chlorosis in blue-green algae**. *Archives of microbiology*, **69**:114-120.

Anderson SL, McIntosh L. (1991) **Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC6803: a blue-light-requiring process.** *Journal of Bacteriology*, **173**:2761-7.

Angermayr SA, Hellingwerf KJ, Lindblad P, Mattos MJT de. (2009) **Energy biotechnology with cyanobacteria.** *Current Opinion in Biotechnology*, **20**:257-63.

BP p.l.c. (2011) *BP Statistical Review of World Energy June 2011*.

Bader GD, Hogue CWV. (2003) **An automated method for finding molecular complexes in large protein interaction networks**. *BMC Bioinformatics*, **4**:1-27.

Baebprasert W, Jantaro S, Khetkorn W, Lindblad P, Incharoensakdi A. (2011) **Increased H2 production in the cyanobacterium *Synechocystis* sp. strain PCC6803 by redirecting the electron supply via genetic engineering of the nitrate assimilation pathway.** *Metabolic engineering*, **13**:610-616.

Bailey J, Ollis D. (1986) *Biochemical engineering fundamentals*. Second. New York: McGraw-Hill.

Bairoch A. (2000) **The ENZYME database in 2000.** *Nucleic acids research*, **28**:304-5.

Barabási A-L, Albert R. (1999) **Emergence of Scaling in Random Networks**. *Science (New York, N.Y.)*, **286**:509-512.

Barabási A-L, Bonabeau E. (2003) **Scale-free networks**. *Scientific American*, **288**:50-59.

Barrett CL, Kim TY, Kim HU, Palsson BØ, Lee SY. (2006) **Systems biology as a foundation for genome-scale synthetic biology**. *Current Opinion in Biotechnology*, **17**:488-492.

Barstow B, Agapakis CM, Boyle PM, et al. (2011) **A synthetic system links FeFe-hydrogenases to essential *E. coli* sulfur metabolism.** *Journal of Biological Engineering*, **5**:7.

Becker SA, Feist AM, Mo ML, et al. (2007) **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.** *Nature protocols*, **2**:727-38.

Becking LGMB, Kaplan IR, Moore D. (1960) **Limits of the natural environment in terms of ph and oxidation-reduction potentials**. *Journal of Geology*, **68**:243-284.

Berry J, Bjorkman O. (1980) **Photosynthetic response and adaptation to temperature in higher plants.** *Annual Review of Plant Physiology*, **31**:491-543.

BioModularH2. (2005) *Engineered Modular Bacterial Photoproduction of Hydrogen (BioModularH2), FP6-2005-NEST-PATH-SYN, PATHFINDER STREP, contract n°043340*.

Blanco-Rivero A, Leganés F, Fernández-Valiente E, Calle P, Fernández-Piñas F. (2005) ***mrpA*, a gene with roles in resistance to Na+ and adaptation to alkaline pH in the cyanobacterium *Anabaena* sp. PCC7120.** *Microbiology (Reading, England)*, **151**:1671-82.

Bonarius HPJ, Schmid G, Tramper J. (1997) **Flux analysis of underdetermined metabolic networks: the quest for the missing constraints**. *Trends in Biotechnology*, **15**:308-314.

Boyle EI, Weng S, Gollub J, et al. (2004) **GO::TermFinder -open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes.** *Bioinformatics (Oxford, England)*, **20**:3710-5.

Bricker TM, Zhang S, Laborde SM, et al. (2004) **The malic enzyme is required for optimal photoautotrophic growth of *Synechocystis* sp. strain PCC6803 under continuous light but not under a diurnal light regimen.** *Journal of Bacteriology*, **186**:8144-8.

Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. (2004) **Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions**. *Genome Research*, **14**:301-312.

Burgdorf T, Lenz O, Buhrke T, et al. (2005) **[NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation.** *Journal of molecular microbiology and biotechnology*, **10**:181-96.

Burrows E, Chaplen F, Ely RL. (2008) **Optimization of media nutrient composition for increased photofermentative hydrogen production by *Synechocystis* sp. PCC6803**. *International Journal of Hydrogen Energy*, **33**:6092-6099.

Camacho Rubio F, García Camacho F, Fernández Sevilla JM, et al. (2003) **A mechanistic model of photosynthesis in microalgae.** *Biotechnology and bioengineering*, **81**:459-73.

Carr NG, Whitton BA. (1982) *The Biology of cyanobacteria*. Berkeley: University of California Press.

Carty RH, Mazumder MM, Schreiber JD, Pangborn JB. (1981) *Thermochemical Hydrogen Production, Final Report*.

Caspi R, Foerster H, Fulcher CA, et al. (2006) **MetaCyc: a multiorganism database of metabolic pathways and enzymes.** *Nucleic acids research*, **34**:D511-6.

Castenholtz R. (1969) **Thermophilic blue-green algae and the thermal environment.** *Bacteriological reviews*, **33**.

Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. (2009) **BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009.** *Nucleic acids research*, **37**:D588-92.

Clarke BL. (1988) **Stoichiometric network analysis.** *Cell biophysics*, **12**:237-53.

Collier JL, Grossman AR. (1992) **Chlorosis induced by nutrient deprivation in *Synechococcus* sp. Strain PCC7942: Not all bleaching is the same**. *Journal of Bacteriology*, **174**:4718-4726.

Committee on Alternatives and Strategies for Future Hydrogen Production and Use, National Research Council. (2004) *The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs*.

Crick FH. (1970) **Central Dogma of Molecular Biology**. *Nature*, **227**:561-563.

Csete M, Doyle J. (2004) **Bow ties, metabolism and disease.** *Trends in Biotechnology*, **22**:446-50.

Cvijovic M, Olivares-Hernández R, Agren R, et al. (2010) **BioMet Toolbox : genome-wide analysis of metabolism**. *Nucleic acids research*, **38**:144-149.

Das D, Veziroglu TN. (2008) **Advances in biological hydrogen production processes**. *International Journal of Hydrogen Energy*, **33**:6046-6057.

Das D, Veziroglu TN. (2001) **Hydrogen production by biological processes: a survey of literature**. *International Journal of Hydrogen Energy*, **26**:13-28.

Deng MD, Coleman JR. (1999) **Ethanol synthesis by genetic engineering in cyanobacteria.** *Applied and environmental microbiology*, **65**:523-8.

Department of Energy. (2007) *Hydrogen production - Hydrogen, Fuel Cells and Infrastructure Technologies Program*. 1-47.

Devroe EJ, Kosuri S, Berry DA, et al. (2010) **Hyperphotosynthetic organisms**. *US Patent 7,785,861*.

Diamant I, Eldar YC, Rokhlenko O, Ruppin E, Shlomi T. (2009) **A network-based method for predicting gene-nutrient interactions and its application to yeast amino-acid metabolism.** *Molecular BioSystems*, **5**:1732-9.

Doniger SW, Salomonis N, Dahlquist KD, et al. (2003) **MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data.** *Genome biology*, **4**:R7.

Douglas SE. (1998) **Plastid evolution: origins, diversity, trends.** *Current Opinion in Genetics & Development*, **8**:655-61.

Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA. (2003) **Global functional profiling of gene expression.** *Genomics*, **81**:98-104.

Ducat DC, Sachdeva G, Silver PA. (2011a) **Rewiring hydrogenase-dependent redox circuits in cyanobacteria.** *Proceedings of the National Academy of Sciences of the United States of America*, **108**:3941-6.

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

Edwards JS, Palsson BØ. (2000a) **Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions.** *BMC Bioinformatics*, **1**:1.

Edwards JS, Palsson BØ. (1999) **Systems properties of the *Haemophilus influenzae* Rd metabolic genotype**. *Journal of Biological Chemistry*, **274**:17410-17416.

Edwards JS, Palsson BØ. (2000b) **The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities**. *Proceedings of the National Academy of Sciences of the United States of America*, **97**:5528-5533.

Edwards JS, Ramakrishna R, Schilling CH, Palsson BØ. (1999) **Metabolic flux balance analysis**. In *Metabolic Engineering*. edited by Lee S, Papoutsakis E New York: Marcel Dekker Inc.

Elam CC, Gregoire-Padró C, Sandrock G, et al. (2003) **Realizing the hydrogen future: the International Energy Agency's efforts to advance hydrogen energy technologies**. *International Journal of Hydrogen Energy*, **28**:601-607.

European Commission. (2011a) *Energy and Transport factsheets 2010*.

European Commission. (2011b) **Europe 2020 initiative [http://ec.europa.eu/energy/strategies/2010/2020_en.htm]**.

Fabiano B, Perego P. (2002) **Thermodynamic study and optimization of hydrogen production by *Enterobacter aerogenes*.** *International Journal of Hydrogen Energy*, **27**:149-156.

Feist AM, Henry CS, Reed JL, et al. (2007) **A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Molecular systems biology*, **3**:121.

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. (2009) **Reconstruction of biochemical networks in microorganisms.** *Nature reviews. Microbiology*, **7**:129-43.

Feist AM, Palsson BØ. (2010) **The biomass objective function.** *Current Opinion in Microbiology*, **13**:344-9.

Fell DA, Small JR. (1986) **Fat synthesis in adipose tissue. An examination of stoichiometric constraints.** *The Biochemical journal*, **238**:781-6.

Foster JS, Singh AK, Rothschild LJ, Sherman LA. (2007) **Growth-phase dependent differential gene expression in *Synechocystis* sp. strain PCC6803 and regulation by a group 2 sigma factor.** *Archives of microbiology*, **187**:265-79.

Fu P. (2008) **Genome-scale modeling of *Synechocystis* sp. PCC6803 and prediction of pathway insertion.** *Journal of Chemical Technology & Biotechnology*, **84**:473-483.

Fujishima A, Honda K. (1972) **Electrochemical photolysis of water at a semiconductor elctrode.** *Nature*, **238**:37-38.

Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. (2003) **Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network.** *Genome Research*, **13**:244-253.

Gasch AP, Spellman PT, Kao CM, et al. (2000) **Genomic expression programs in the response of yeast cells to environmental changes.** *Molecular biology of the cell*, **11**:4241-57.

Gerdtzen ZP, Daoutidis P, Hu W-S. (2004) **Non-linear reduction for kinetic models of metabolic reaction networks.** *Metabolic engineering*, **6**:140-54.

Gill RT, Katsoulakis E, Schmitt W, et al. (2002) **Genome-Wide Dynamic Transcriptional Profiling of the Light-to-Dark Transition in *Synechocystis* sp. Strain PCC6803.** *Journal of Bacteriology*, **184**:3671-3681.

Golub GH, Loan CF Van. (1996) *Matrix computations*. Baltimore, MD: Johns Hopkins University Press.

Greene SR, Flanagan GF, Borole AP. (2009) *Integration of Biorefineries and Nuclear Cogeneration Power Plants – A Preliminary Analysis*.

Grünenfelder B, Winzeler EA. (2002) **Treasures and traps in genome-wide data sets: case examples from yeast.** *Nature reviews. Genetics*, **3**:653-61.

Gugger M, Biological Resource Center of Institut Pasteur. (2011a) **Catalogue data sheet of *Synechocystis* PCC6308 [http://www.crbip.pasteur.fr/fiches/fichecata.jsp?crbip=PCC%206308]**.

Gugger M, Biological Resource Center of Institut Pasteur. (2011b) **Pasteur Culture Collection [http://www.pasteur.fr/ip/easysite/pasteur/en/research/collections/crbip /general-informations-concerning-the-collections/iv-the-open-collections/iv-iii-pasteur-culture-collection-of-cyanobacteria]**.

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. PCC6803.** *Biochimica et Biophysica Acta*, **1767**:161-9.

Hallenbeck P. (2002) **Biological hydrogen production; fundamentals and limiting processes**. *International Journal of Hydrogen Energy*, **27**:1185-1193.

Hamelinck CN, Hooijdonk G Van, Faaij A. (2005) **Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term**. *Biomass and Bioenergy*, **28**:384-410.

Herdman M, Janvier M, Waterbury JB, et al. (1979) **Deoxyribonucleic Acid Base Composition of Cyanobacteria**. *Journal of general microbiology*, **111**:63-71.

Herrero A, Flores E. (2008) *The cyanobacteria: molecular biology, genomics, and evolution*. Horizon Scientific Press:484.

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*, **13**:793-806.

Hihara Y, Sonoike K, Kanehisa M, Ikeuchi M. (2003) **DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC6803.** *Journal of Bacteriology*, **185**:1719-25.

Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. (2006) **Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels.** *Proceedings of the National Academy of Sciences of the United States of America*, **103**:11206-10.

Hockemeyer D, Wang H, Kiani S, et al. (2011) **Genetic engineering of human pluripotent cells using TALE nucleases.** *Nature biotechnology*, **29**:731-734.

Hong-Bo S, Li-Ye C, Ming-An S, Shi-Qing L, Ji-Cheng Y. (2008) **Bioengineering plant resistance to abiotic stresses by the global calcium signal system.** *Biotechnology advances*, **26**:503-10.

Hood L, Heath JR, Phelps ME, Lin B. (2004) **Systems biology and new technologies enable predictive and preventative medicine.** *Science (New York, N.Y.)*, **306**:640-3.

Houchins JP. (1984) **The physiology and biochemistry of hydrogen metabolism in cyanobacteria.** *Biochimica et Biophysica Acta Reviews on Bioenergetics*, **768**:227-255.

Houmard J, Tandeau de Marsac N. (1988) **Cyanobacterial genetic tools: current status.** *Methods in enzymology*, **167**:808-47.

Huang JJ, Kolodny NH, Redfearn JT, Allen MM. (2002a) **The acid stress response of the cyanobacterium *Synechocystis* sp. strain PCC6308.** *Archives of microbiology*, **177**:486-493.

Huang L, McCluskey MP, Ni H, LaRossa RA. (2002b) **Global gene expression profiles of the cyanobacterium *Synechocystis* sp. strain PCC6803 in response to irradiation with UV-B and white light.** *Journal of Bacteriology*, **184**:6845-58.

Hübschmann T, Yamamoto H, Gieler T, Murata N, Börner T. (2005) **Red and far-red light alter the transcript profile in the cyanobacterium *Synechocystis* sp. PCC6803: impact of cyanobacterial phytochromes.** *FEBS Letters*, **579**:1613-8.

Ideker T, Ozier O, Schwikowski B, Siegel AF. (2002) **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics (Oxford, England)*, **18 Suppl 1**:S233-40.

Ideker T, Thorsson V, Ranish JA, et al. (2001) **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science (New York, N.Y.)*, **292**:929-34.

Imhoff JF, Sahl HG, Soliman GSH, Truper HG. (1979) **The Wadi Natrun: Chemical composition and microbial mass developments in alkaline brines of Eutrophic Desert Lakes**. *Geomicrobiology Journal*, **1**:219-234.

Inoue N, Taira Y, Emi T, et al. (2001) **Acclimation to the growth temperature and the high-temperature effects on photosystem ii and plasma membranes in a mesophilic cyanobacterium, *Synechocystis* sp. PCC6803**. *Plant & Cell Physiology*, **42**:1140-1148.

International Genetically Engineered Machine competition (iGEM). (2012) **webpage [www.igem.org]**.

Kacser H, Burns JA. (1973) **The control of flux.** *Symposia of the Society for Experimental Biology*, **27**:65-104.

Kalia VC, Jain SR, Kumar A, Joshi AP. (1994) **Frementation of biowaste to H2 by *Bacillus licheniformis***. *World Journal of Microbiology & Biotechnology*, **10**:224-227.

Kallas T, Castenholz R. (1982) **Internal pH and ATP-ADP pools in the cyanobacterium *Synechococcus* sp. during exposure to growth-inhibiting low pH.** *Journal of Bacteriology*, **149**:229-236.

Kanehisa M, Araki M, Goto S, et al. (2008) **KEGG for linking genomes to life and the environment.** *Nucleic acids research*, **36**:D480-4.

Kaneko T, Sato S, Kotani H, et al. (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 (supplement).** *DNA research*, **3**:185-209.

Kaneko T, Tanaka A, Sato S, et al. (1995) **Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome.** *DNA research*, **2**:153-66, 191-8.

Kanesaki Y, Suzuki I, Allakhverdiev SI, Mikami K, Murata N. (2002) **Salt stress and hyperosmotic stress regulate the expression of different sets of genes in** *Synechocystis* **sp. PCC6803.** *Biochemical and biophysical research communications*, **290**:339-48.

Karp PD, Ouzounis C a, Moore-Kochlacs C, et al. (2005) **Expansion of the BioCyc collection of pathway/genome databases to 160 genomes.** *Nucleic acids research*, **33**:6083-9.

Karp PD, Paley S, Romero P. (2002) **The Pathway Tools software.** *Bioinformatics (Oxford, England)*, **18 Suppl 1**:S225-32.

Keller EF. (2007) **A clash of two cultures.** *Nature*, **445**:603.

Keller EF. (2005) **Revisiting "scale-free" networks.** *BioEssays*, **27**:1060-8.

Kholodenko B. (1995) **Composite control of cell function: metabolic pathways behaving as single control units**. *FEBS Letters*, **368**:1-4.

Kim W-Y, Kang S, Kim B-C, et al. (2008) **SynechoNET: integrated protein-protein interaction database of a model cyanobacterium** *Synechocystis* **sp. PCC6803.** *BMC Bioinformatics*, **9 Suppl 1**:S20.

Kitano H. (2002) **Computational systems biology.** *Nature*, **420**:206-10.

Klamt S, Stelling J, Ginkel M, Gilles ED. (2003) **FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps**. *Bioinformatics (Oxford, England)*, **19**:261-269.

Knoop H, Zilliges Y, Lockau W, Steuer R. (2010) **The metabolic network of** *Synechocystis* **sp. PCC6803: Systemic properties of autotrophic growth.** *Plant physiology*, **154**:410-422.

Konak AR. (1975) **Equation for batch bacterial growth.** *Biotechnology and bioengineering*, **17**:271-272.

Kratz WA, Myers J. (1955) **Photosynthesis and respiration of three blue-green algae**. *Plant physiology*, **30**:275-280.

Krieger CJ, Zhang P, Mueller LA, et al. (2004) **MetaCyc: a multiorganism database of metabolic pathways and enzymes.** *Nucleic acids research*, **32**:D438-42.

Kroposki B, Levene J, Harrison K, Sen PK, Novachek F. (2006) *Electrolysis : Information and Opportunities for Electric Power Utilities*.

Kufryk GI, Sachet M, Schmetterer G, Vermaas WFJ. (2002) **Transformation of the cyanobacterium *Synechocystis* sp. PCC6803 as a tool for genetic mapping: optimization of efficiency.** *FEMS Microbiology Letters*, **206**:215-9.

Kun A, Papp B, Szathmáry E. (2008) **Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks.** *Genome biology*, **9**:R51.

Kurian D, Phadwal K, Mäenpää P. (2006) **Proteomic characterization of acid stress response in *Synechocystis* sp. PCC6803.** *Proteomics*, **6**:3614-3624.

Labarre J, Thuriaux P, Chauvat F. (1987) **Genetic analysis of amino acid transport in the facultatively heterotrophic cyanobacterium *Synechocystis* sp. strain 6803.** *Journal of Bacteriology*, **169**:4668-73.

Langworthy T. (1978) *Microbial life in extreme environments*. London: Academic Press.

Latendresse M, Krummenacker M, Trupp M, Karp PD. (2012) **Construction and completion of flux balance models from pathway databases**. *Bioinformatics (Oxford, England)*:btr681.

Lawrence BA, Suarez C, DePina A, et al. (1998) **Two internal pools of soluble polyphosphate in the cyanobacterium *Synechocystis* sp. strain PCC6308: an in vivo 31P NMR spectroscopic study.** *Archives of microbiology*, **169**:195-200.

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**:70-9.

Liu X, Brune D, Vermaas W, Curtiss R. (2010) **Production and secretion of fatty acids in genetically engineered cyanobacteria.** *Proceedings of the National Academy of Sciences of the United States of America*, **6803**:2-7.

Llaneras F, Picó J. (2008) **Stoichiometric modelling of cell metabolism.** *Journal of bioscience and bioengineering*, **105**:1-11.

Llaneras F. (2010) **Interval and Possibilistic Methods for Constraint-Based Metabolic Models**. *PhD thesis*:1-266.

Loferer-Krössbacher M, Klima J, Psenner R. (1998) **Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis.** *Applied and environmental microbiology*, **64**:688-94.

Lopes Pinto F, Troshina O, Lindblad P. (2002) **A brief look at three decades of research on cyanobacterial hydrogen evolution**. *International Journal of Hydrogen Energy*, **27**:1209-1215.

Lopez-Archilla AI, Moreira D, Lopez-Garcia P, Guerrero C. (2004) **Phytoplankton diversity and cyanobacterial dominance in a hypereutrophic shallow lake with biologically produced alkaline ph**. *Extremophiles*, **8**:109-115.

Lorenzo V de. (2006) **Blueprint of an oil-eating bacterium.** *Nature biotechnology*, **24**:952-3.

Los D, Horvath I, Vigh L, Murata N. (1993) **The temperature-dependent expression of the desaturase gene desa in** *Synechocystis* **PCC6803**. *FEBS Letters*, **318**:57-60.

Löffelhardt W, Schmetterer G. (1999) *The phototrophic prokaryotes*. Kluwer Academic/Plenum.

Maere S, Heymans K, Kuiper M. (2005) **BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics (Oxford, England)*, **21**:3448-9.

Maestri O, Joset F. (2000) **Regulation by external ph and stationary growth phase of the acetolactate synthase from** *Synechocystis* **PCC6803.** *Molecular Microbiology*, **37**:828-838.

Mahadevan R, Schilling CH. (2003) **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models**. *Metabolic engineering*, **5**:264-276.

Maniatis K. (2010) *Biofuels from algae : Results of the 2010 FP7 Call*. 1-9.

Manish S, Venkatesh KV, Banerjee R. (2007) **Metabolic flux analysis of biological hydrogen production by** *Escherichia coli*. *International Journal of Hydrogen Energy*, **32**:3820-3830.

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. (1992) **Synthesis of biochemical production routes**. *Computers & Chemical Engineering*, **16**:605-619.

Melis A, Melnicki M. (2006) **Integrated biological hydrogen production**. *International Journal of Hydrogen Energy*, **31**:1563-1573.

Melis A, Zhang L, Forestier M, Ghirardi ML, Seibert M. (2000) **Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. *Plant physiology*, **122**:127-136.

Melis A. (2007) **Maximizing light utilization efficiency and hydrogen production in microalgal cultures**. *Department of Energy - Hydrogen Program*:182-184.

Metzker ML. (2010) **Sequencing technologies - the next generation.** *Nature reviews. Genetics*, **11**:31-46.

Miao X, Wu Q, Wu G, Zhao N. (2003) **Changes in photosynthesis and pigmentation in an agp deletion mutant of the cyanobacterium *Synechocystis* sp.** *Biotechnology letters*, **25**:391-6.

Miller AG, Turpin DH, Canvin DT. (1984) **Na+ requirement for growth, photosynthesis, and ph regulation in the alkalotolerant cyanobacterium *Synechococcus leopoliensis*. *Journal of Bacteriology*, **159**:100-106.

Milne TA, Elam CC, Evans RJ. (2002) *Hydrogen from Biomass - State of the Art and Research Challenges*.

Momirlan M, Veziroglu TN. (2005) **The properties of hydrogen as fuel tomorrow in sustainable energy system for a cleaner planet**. *International Journal of Hydrogen Energy*, **30**:795-802.

Montagud A, Navarro E, Fernández de Córdoba P, 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, et al. (2011) **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803.** *Biotechnology Journal*, **6**:330-342.

Morange M. (2009) **A new revolution? The place of systems biology and synthetic biology in the history of biology**. *EMBO reports*, **10**.

Mullineaux CW. (2008) **How do cyanobacteria sense and respond to light?** *Molecular microbiology*, **41**:965-971.

NCBI. (2011) **Entrez Genome for *Synechocystis* sp. PCC6803 [http://www.ncbi.nlm.nih.gov/genome/1018]**.

Nakao M, Okamoto S, Kohara M, et al. (2010) **CyanoBase: the cyanobacteria genome database update 2010.** *Nucleic acids research*, **38**:D379-81.

Navarro E, Montagud A, Fernández de Córdoba P, Urchueguía JF. (2009) **Metabolic flux analysis of the hydrogen production potential in *Synechocystis* sp. PCC6803**. *International Journal of Hydrogen Energy*, **34**:8828-8838.

Nielsen J, Villadsen J. (1992) **Modelling of microbial kinetics**. *Chemical Engineering Science*, **47**:4225-4270.

Nielsen J. (2001) **Metabolic engineering**. *Applied microbiology and biotechnology*, **55**:263-283.

Notebaart RA, Teusink B, Siezen RJ, Papp B. (2008) **Co-regulation of metabolic genes is better explained by flux coupling than by network distance.** *PLoS Computational Biology*, **4**:e26.

Oh Y-K, Seol E-H, Lee EY, Park S. (2002) **Fermentative hydrogen production by a new chemoheterotrophic bacterium *Rhodopseudomonas palustris* P4**. *International Journal of Hydrogen Energy*, **27**:1373-1379.

Oliveira AP, Patil KR, Nielsen J. (2008) **Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks.** *BMC Systems Biology*, **2**:17.

Orth JD, Thiele I, Palsson BØ. (2010) **What is flux balance analysis?** *Nature biotechnology*, **28**:245-8.

Palsson B. (2000) **The challenges of in silico biology.** *Nature biotechnology*, **18**:1147-50.

Patil KR, Akesson M, Nielsen J. (2004) **Use of genome-scale microbial models for metabolic engineering.** *Current opinion in Biotechnology*, **15**:64-9.

Patil KR, Nielsen J. (2005) **Uncovering transcriptional regulation of metabolism by using metabolic network topology.** *Proceedings of the National Academy of Sciences of the United States of America*, **102**:2685-9.

Patil KR, Rocha I, Förster J, Nielsen J. (2005) **Evolutionary programming as a platform for in silico metabolic engineering.** *BMC Bioinformatics*, **6**:308.

Pearce J, Carr NG. (1967) **The metabolism of acetate by the blue-green algae, *Anabaena variabilis* and *Anacystis nidulans*.** *Journal of general microbiology*, **49**:301-13.

Pelroy RA, Rippka R, Stanier RY. (1972) **Metabolism of glucose by unicellular blue-green algae.** *Archiv für Mikrobiologie*, **87**:303-22.

Perret R. (2005) *Solar Hydrogen Generation Research - High-Temperature Thermochemical*. 377-388.

Pfeiffer T, Sánchez-Valdenebro I, Nuño JC, Montero F, Schuster S. (1999) **METATOOL: for studying metabolic networks.** *Bioinformatics (Oxford, England)*, **15**:251-7.

Pierson B, Parenteau M, Griffin B. (1999) **Phototrophs in high-iron-concentration microbial mats: Physiological ecology of phototrophs in an iron-depositing hot spring.** *Applied and environmental microbiology*, **65**:5474-5483.

Pinto F, Elburg KA Van, Pacheco CC, et al. (2011) **Construction of a chassis for hydrogen production: physiological and molecular characterization of a *Synechocystis* sp. PCC6803 mutant lacking a functional bidirectional hydrogenase.** *Microbiology (Reading, England)*, **158**:448-464.

Postgate JR. (1970) **Biological nitrogen fixation.** *Nature*, **226**:25-7.

Pramanik J, Keasling JD. (1997) **Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements.** *Biotechnology and bioengineering*, **56**:398-421.

Prust C, Hoffmeister M, Liesegang H, et al. (2005) **Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*.** *Nature biotechnology*, **23**:195-200.

Rapoport TA, Heinrich R, Jacobasch G, Rapoport S. (1974) **A linear steady-state treatment of enzymatic chains. A mathematical model of glycolysis of human erythrocytes.** *European journal of biochemistry / FEBS*, **42**:107-20.

Raven JA, Allen JF. (2003) **Genomics and chloroplast evolution: what did cyanobacteria do for plants?** *Genome biology*, **4**:209.

Rizzi M, Baltes M, Theobald U, Reuss M. (1997) **In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model.** *Biotechnology and bioengineering*, **55**:592-608.

Ro D-K, Paradise EM, Ouellet M, et al. (2006) **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature*, **440**:940-3.

Rocha I, Maia P, Evangelista P, et al. (2010) **OptFlux: an open-source software platform for in silico metabolic engineering.** *BMC Systems Biology*, **4**:45.

Rohwer JM, Schuster S, Westerhoff HV. (1996) **How to recognize monofunctional units in a metabolic system.** *Journal of theoretical biology*, **179**:213-28.

Sakamoto T, Bryant D. (1998) **Growth at low temperature causes nitrogen limitation in the cyanobacterium *Synechococcus* sp. PCC7002.** *Archives of microbiology*, **169**:10-19.

Sakamoto T, Inoue-Sakamoto K, Bryant D. (1999) **A novel nitrate/nitrite permease in the marine cyanobacterium *Synechococcus* sp. Strain PCC7002.** *Journal of Bacteriology*, **181**:7363-7372.

Sauer U, Heinemann M, Zamboni N. (2007) **Getting closer to the whole picture.** *Science (New York, N.Y.)*, **316**:550-1.

Savinell JM, Palsson BO. (1992a) **Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism.** *Journal of theoretical biology*, **154**:421-454.

Savinell JM, Palsson BO. (1992b) **Network analysis of intermediary metabolism using linear optimization. II. Interpretation of hybridoma cell metabolism.** *Journal of theoretical biology*, **154**:455-73.

Schilling CH, Letscher D, Palsson BØ. (2000) **Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *Journal of theoretical biology*, **203**:229-48.

Schilling CH, Schuster S, Palsson BØ, Heinrich R. (1999) **Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era.** *Biotechnology progress*, **15**:296-303.

Schmitt WA, Stephanopoulos G. (2003) **Prediction of transcriptional profiles of *Synechocystis* PCC6803 by dynamic autoregressive modeling of DNA microarray data.** *Biotechnology and bioengineering*, **84**:855-63.

Schmitz O, Boison G, Salzmann H, et al. (2002) *Hoxe--a subunit specific for the pentameric bidirectional hydrogenase complex (hoxefuyh) of cyanobacteria*. *Biochimica et Biophysica Acta*, **1554**:66-74.

Schopf JW. (2000) **The fossil record: tracing the roots of the cyanobacterial lineage**. In *The ecology of cyanobacteria*. edited by Whitton BA, Potts M Kluwer Academic Publishers, Dordrecht, The Netherlands:13-35.

Schuetz R, Kuepfer L, Sauer U. (2007) **Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*.** *Molecular systems biology*, **3**:119.

Schuster R, Schuster S. (1993) **Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed**. *CABIOS*, **9**:79-85.

Schuster S, Hlgetag C, Hilgetag C. (1994) **On elementary flux modes in biochemical reaction systems at steady state**. *Journal of Biological Systems*, **2**:165-182.

Segal E, Friedman N, Kaminski N, Regev A, Koller D. (2005) **From signatures to models: understanding cancer using microarrays.** *Nature genetics*, **37 Suppl**:S38-45.

Segrè D, Vitkup D, Church GM. (2002) **Analysis of optimality in natural and perturbed metabolic networks.** *Proceedings of the National Academy of Sciences of the United States of America*, **99**:15112-7.

Shastri A, Morgan JA. (2005) **Flux balance analysis of photoautotrophic metabolism.** *Biotechnology progress*, **21**:1617-26.

Shinnar R, Citro F. (2006) **A road map to U.S. decarbonization.** *Science (New York, N.Y.)*, **313**:1243-4.

Shlomi T, Berkman O, Ruppin E. (2005) **Regulatory on/off minimization of metabolic flux changes after genetic perturbations.** *Proceedings of the National Academy of Sciences of the United States of America*, **102**:7695-700.

Singh AK, Bhattacharyya-Pakrasi M, Elvitigala T, et al. (2009) **A systems-level analysis of the effects of light quality on the metabolism of a cyanobacterium.** *Plant physiology*, **151**:1596-608.

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. PCC6803.** *Plant physiology*, **132**:1825-39.

Snoep JL, Westerhoff HV. (2005) **From isolation to integration , a systems biology approach for building the Silicon Cell**. *Topics in Current Genetics*, **13**:13-30.

Soanes C, Stevenson A. (2010) *The New Oxford Dictionary of English*. 3rd edition. Oxford: Oxford University Press.

Solazyme Inc. (2012) **webpage [http://www.solazyme.com/technology]**.

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

Stapleton JA, Swartz JR. (2010) **Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases.** *PLoS ONE*, **5**:e15275.

Steinberg C, Schafer H, Beisker W. (1998) **Do acid-tolerant cyanobacteria exist?** *Acta Hydrochimica Et Hydrobiologica*, **26**:13-19.

Stephanopoulos G, Alper H, Moxley J. (2004) **Exploiting biological complexity for strain improvement through systems biology.** *Nature biotechnology*, **22**:1261-7.

Stephanopoulos G, Aristidou A, Nielsen J. (1998) *Metabolic engineering: principles and methodologies*. San Diego: Academic Press.

Stephanopoulos G, Stafford DE. (2002) **Metabolic engineering: a new frontier of chemical reaction engineering**. *Chemical Engineering Science*, **57**:2595-2602.

Stevens SE, Balkwill DL, Paone DAM. (1981) **The effects of nitrogen limitation on the ultrastructure of the cyanobacterium *Agmeneilum quadruplicatum***. *Archives of microbiology*, **130**:204-212.

Subramanian A, Tamayo P, Mootha VK, et al. (2005) **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences of the United States of America*, **102**:15545-50.

Summerfield TC, Sherman LA. (2008) **Global transcriptional response of the alkali-tolerant cyanobacterium *Synechocystis* sp. strain PCC6803 to a pH 10 environment.** *Applied and environmental microbiology*, **74**:5276-84.

Suthers PF, Chang YJ, Maranas CD. (2010) **Improved computational performance of MFA using elementary metabolite units and flux coupling.** *Metabolic engineering*, **12**:123-8.

Suzuki I, Kanesaki Y, Mikami K, Kanehisa M, Murata N. (2001) **Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*.** *Molecular microbiology*, **40**:235-44.

Suzuki I, Simon WJ, Slabas AR. (2006) **The heat shock response of *Synechocystis* sp. PCC6803 analysed by transcriptomics and proteomics.** *Journal of experimental botany*, **57**:1573-8.

Swingley WD, Blankenship RE. (2008) **Insights into cyanobacterial evolution from comparative genomics**. In *The cyanobacteria: molecular biology, genomics, and evolution*. edited by Herrero A, Flores E:21-43.

Taguchi F, Yamada K, Hasegawa K, Taki-Saito T, Hara K. (1996) **Continuous hydrogen production by *Clostridium* sp. strain no. 2 from cellulose hydrolysate in an aqueous two-phase system**. *Journal of Fermentation and Bioengineering*, **82**:80-83.

Talling JF, Talling IB. (1965) **The chemical composition of african lake waters.** *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, **50**:421-463.

Tamagnini P, Axelsson R, Lindberg P, et al. (2002) **Hydrogenases and hydrogen metabolism of cyanobacteria**. *Microbiology and molecular biology reviews : MMBR*, **66**:1-20.

Tamagnini P, Leitão E, Oliveira P, et al. (2007) **Cyanobacterial hydrogenases: diversity, regulation and applications.** *FEMS Microbiology Reviews*, **31**:692-720.

Tasakal Y, Gombosl Z, Nishiyamal Y, et al. (1996) **Targeted mutagenesis of acyl-lipid desaturases in *Synechocystis*: evidence for the important roles of polyunsaturated membrane lipids in growth, respiration and photosynthesis.** *The EMBO journal*, **15**:6416-25.

Thauer R. (1976) **Limitation of microbial H2-formation via fermentation**. In *Microbial energy conversion*. edited by Schlegel HG, Barnea J E. Goltze KG, Göttingen:201-4.

The UniProt Consortium. (2007) **The Universal Protein Resource (UniProt).** *Nucleic acids research*, **35**:D193-7.

Thiel T. (1994) **Genetic analysis of cyanobacteria. Chapter 19.** In *Molecular Biology of Cyanobacteria*. 581-611.

Thiele I, Palsson BØ. (2010) **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nature protocols*, **5**:93-121.

Turner J, Sverdrup G, Mann MK, et al. (2008) **Renewable hydrogen production**. *International Journal of Energy Research*, **32**:379-407.

Tyo KE, Zhou H, Stephanopoulos GN. (2006) **High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. strain PCC6803.** *Applied and environmental microbiology*, **72**:3412-7.

U.S. Energy Information Administration. (2011a) *August 2011 Monthly Energy Review*.

U.S. Energy Information Administration. (2011b) **webpage [http://www.eia.gov/]**.

United Nations. (2009) *Copenhagen Accord to the United Nations framework on climate change*.

United Nations. (1998) *Kyoto Protocol to the United Nations framework convention on climate change*.

United States Government. (2011) **Hydrogen and Fuel Cells Interagency Working Group [http://www.hydrogen.gov/]**.

Varma A, Boesch BW, Palsson BØ. (1993a) **Biochemical production capabilities of *Escherichia coli*.** *Biotechnology and bioengineering*, **42**:59-73.

Varma A, Boesch BW, Palsson BØ. (1993b) **Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates.** *Applied and environmental microbiology*, **59**:2465-73.

Varma A, Palsson BØ. (1994a) **Metabolic Flux Balancing: Basic concepts, scientific and practical use**. *Nature*, **12**:994-998.

Varma A, Palsson BØ. (1993a) **Metabolic capabilities of *E. coli* II. Optimal growth patterns**. *Journal of theoretical biology*, **165**:503-522.

Varma A, Palsson BØ. (1993b) **Metabolic capabilities of *E. coli* I. Synthesis of biosynthetic precursors and cofactors**. *Journal of theoretical biology*, **165**:477-502.

Varma A, Palsson BØ. (1994b) **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.** *Applied and environmental microbiology*, **60**:3724-31.

Vazquez-Bermudez MF, Herrero A, Flores E. (2000) **Uptake of 2-oxoglutarate in *Synechococcus* strains transformed with the *Escherichia coli kgtP* gene**. *Journal Of Bacteriology*, **182**:211-215.

Vermaas WFJ. (1996) **Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC6803: Principles and possible biotechnology applications**. *Journal of Applied Phycology*, **8**:263-273.

Wada H, Gombos Z, Murata N. (1994) **Contribution of membrane lipids to the ability of the photosynthetic machinery to tolerate temperature stress**. *Proceedings of the National Academy of Sciences of the United States of America*, **91**:4273-4277.

Wang D, Li Q, Li W, Xing J, Su Z. (2009) **Improvement of succinate production by overexpression of a cyanobacterial carbonic anhydrase in *Escherichia coli***. *Enzyme and Microbial Technology*, **45**:491-497.

Wang H-L, Postier BL, Burnap RL. (2004) **Alterations in global patterns of gene expression in *Synechocystis* sp. PCC6803 in response to inorganic carbon limitation and the inactivation of *ndhR*, a *LysR* family regulator.** *The Journal of biological chemistry*, **279**:5739-51.

Waterbury J. (2006) **The prokaryotes: A handbook on the biology of bacteria: Bacteria: Firmicutes, cyanobacteria**. In *Cyanobacteria*. edited by Dworkin M, Falkow S, Rosenberg E, Stackebrandt E New York: Springer Science:1140.

Watson JD, Crick FH. (1953) **Molecular structure of nucleic acids, a structure for deoxyribose nucleic acid.** *Nature*, **171**:737-8.

Weise S, Grosse I, Klukas C, et al. (2006) **Meta-All: a system for managing metabolic pathway information.** *BMC Bioinformatics*, **7**:465.

Wetterstrand K. (2012) **DNA Sequencing Costs: Data from the NHGRI large-scale genome sequencing program.** *Available at: www.genome.gov/sequencingcosts.*:Accessed [21 January 2012].

Whitton BA, Potts M. (2000) *The ecology of cyanobacteria: their diversity in time and space.* Boston: Kluwer Academic Publishers.

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.

Wu GF, Wu QY, Shen ZY. (2001) **Accumulation of poly-beta-hydroxybutyrate in cyanobacterium *Synechocystis* sp. PCC6803.** *Bioresource technology*, **76**:85-90.

Yamaguchi K, Suzuki I, Yamamoto H, et al. (2002) **A two-component Mn_2-sensing system negatively regulates expression of the *mntCAB* operon in *Synechocystis*.** *The Plant Cell*, **14**:2901-2913.

Yamanaka G, Glazer AN. (1980) **Dynamic aspects of phycobilisome structure. Phycobilisome turnover during nitrogen starvation in *Synechococcus* sp.** *Archives of microbiology*, **124**:39-47.

Yang C, Hua Q, Shimizu K. (2002a) **Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*.** *Applied microbiology and biotechnology*, **58**:813-22.

Yang C, Hua Q, Shimizu K. (2002b) **Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose**. *Metabolic engineering*, **4**:202-216.

Yang C, Hua Q, Shimizu K. (2002c) **Quantitative analysis of intracellular metabolic fluxes using GC-MS and two-dimensional NMR spectroscopy.** *Journal of bioscience and bioengineering*, **93**:78-87.

Yoo S, Keppel C, Spalding M, Jane J. (2007) **Effects of growth condition on the structure of glycogen produced in cyanobacterium *Synechocystis* sp. PCC6803**. *International journal of biological macromolecules*, **40**:498-504.

Yoshikawa K, Kojima Y, Nakajima T, et al. (2011) **Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC6803.** *Applied microbiology and biotechnology*, **92**:1-5.

Zelezniak A, Pers TH, Soares S, Patti ME, Patil KR. (2010) **Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes.** *PLoS Computational Biology*, **6**:e1000729.

Zhang S, Bryant D. (2011) **The Tricarboxylic acid cycle in cyanobacteria**. *Science (New York, N.Y.)*, **334**:1551-1553.

Zhang Z, Pendse ND, Phillips KN, Cotner JB, Khodursky A. (2008) **Gene expression patterns of sulfur starvation in *Synechocystis* sp. PCC6803.** *BMC Genomics*, **9**:344.

Zucchetto J, National Research Council. (2006) *Trends in Oil Supply and Demand, Potential for Peaking of Conventional Oil Production, and Possible Mitigation Options: A Summary Report of the Workshop*.